THE PERSONAL INFORMATICS ANALYSIS GAP

by

James Moore

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computing

School of Computing
The University of Utah

May 2022

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **James Moore**

has been approved by the following supervisory committee members:

Miriah Dawn Meyer,	Co-chair	**February 18, 2022** Date Approved
Jason Wiese,	Co-chair	**February 17, 2022** Date Approved
Erik L. Brunvand,	Member	**February 17, 2022** Date Approved
Alexander Lex,	Member	**February 17, 2022** Date Approved
Wesley Willett,	Member	**February 18, 2022** Date Approved

by **Mary W. Hall**, Chair/Dean of

the Department/College/School of **Computing**

and by **David B. Kieda**, Dean of The Graduate School.

ABSTRACT

Personal informatics research helps people track personal data for the purposes of self-reflection and gaining self-knowledge. This field has predominantly focused on the data collection and insight-generation elements of self-tracking, however, with less attention paid to flexible data analysis. As a result, this inattention has led to inflexible analytic pipelines that do not reflect or support the diverse ways people want to engage with their data. This dissertation contributes a review of personal informatics and visualization research literature to expose a gap in research knowledge for designing flexible tools that assist people engaging with and analyzing personal data in personal contexts. This is identified as the *personal informatics analysis gap*. This dissertation explores this gap through a multiyear longitudinal study of how asthmatics engage with personal air quality data, and reports how this gap emerged through a series of challenges from attempting to design an improved visual analysis tool. This dissertation's primary contribution is the identification of the personal informatics analysis gap, with recommendations for how visualization researchers might bridge it: through designing for play, designing for entry points into personal data, and designing more collaborative or social systems.

For dad

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This dissertation would be nowhere without the guidance, encouragement, and patience — mostly patience — of my advisors, Miriah Meyer and Jason Wiese. To them, I offer my deepest gratitude and condolences for their work turning me into an independent researcher, sometimes in spite of myself. Thanks as well to my committee members, Erik Brunvand, Alex Lex, and Wes Willett, for your feedback along the way.

I would never have gotten this far without the love and support of my family. Thank you Mom, Dad, Chris, and Anna; as well as my family away from home: Lindsay and Carol. I could not have done this without you. I am also deeply grateful for the friendship of everyone who helped to make graduate school a worthwhile and less lonely place. This includes labmates, past and present: Derya Akbaba, Alex Bigelow, Kiran Gadhave, Pascal Goffin, Jochen Görtler, Aspen Hopkins, Kazi Sinthia Kabir, Michael Kern, Ethan Kerzner, Devin Lange, Haihan Lin, Max Lisnic, Nina McCurdy, Sean McKenna, Carolina Nobre, Sam Quinan, Jen Rogers, Jack Wilburn, and Dylan Wooten; friends near and far: Ryan Atkinson, Sarah Cotts, Nicole Davis, Mike & Lisa DeNight, Greg Furlich, Chris Hanney, David Huppman, Jackie Killian, Ryan & Mike Krisac, Nora Lang, Matt McGinnis, Conny Rasmussen, Steve Savage, Brandon Schlenker, Erika Schreiber, Paul & Greg Sookiasian, Elisa Tkach, and Cameron Waller; classmates: Praful Agrawal, Milinda Fernando, John Holmen, Ashok Jallepalli, Brian Kimmig, Michael Matheny, Karl Taht, Matt Territo, and Will Usher; JPL colleagues in the Human Centered Design and MDNav groups: Basak Alper, Kris Blackwood, John Carr, Scott Davidoff, Sonia Hernandez, Marijke Jorritsma, Michael Newcomb, Rohan Patel, Sarah Schnadt, Jeffery Stuart, and Ellen van Wyk; and, of course, participants P1, P2, P3, P4, P4a, P5, and P6. Each of you have helped to get me here in your own ways.

And to any unnamed people reading this now, I thank you as well. But don't you have better things to do than browse my dissertation?

CHAPTER 1

INTRODUCTION

This dissertation identifies and defines the *personal informatics analysis gap*. This gap describes the lack of focused research and design of systems that allow people to flexibly analyze their personal data, despite various research communities working toward these goals within personal informatics and visualization fields [1]. This gap reveals itself as a mismatch between what people may want to do with personal data, and the limited functionalities afforded by the tools at their disposal. For example, the motivation of this research wherein asthmatics log their indoor air quality, yet lack tools to aggregate or review this information for exploring how various indoor activities, outdoor conditions, or home improvements affect their immediate surroundings. We explore the personal informatics analysis gap in this dissertation through an initial longitudinal study of asthmatic families selected from a concurrent national research program to measure factors affecting pediatric asthma [2]. We discuss our study participants and selection criteria more in Section 3.1. Our study involves observing how these families live with an indoor air quality sensing system [3], and a follow-on multiyear qualitative study on engaging participants with their air quality data [1], [4]. We ground this study and its findings within a review of personal informatics and visualization literature to demonstrate why this gap exists, and identify potential solutions and implications for designing with the gap in mind [1].

1.1 Why the gap exists

The personal informatics analysis gap emerges from uncoordinated research efforts between the fields of personal informatics, visual analytics, and everyday visualization. Although each field works toward making data more actionable and accessible, their collective efforts remain largely limited to their individual communities. Personal informatics research focuses on tools and technologies that help people to collect and make sense of personally relevant data for the purposes of self-reflection and gaining self-knowledge [5]. This field is driven by the idea that data will help us make better, smarter decisions about how we live our lives, and studies how and why people engage with personal data across a broad range of application areas. This has led to the development of various tools or systems for tracking personal data in the areas of: physical fitness, with early work by Tsai et al. targeting caloric intake [6] or Consolvo et al. for developing design requirements to support[7] or sense [8] physical exercise; time management and online behavior, with work by Lindqvist et al. studying how and why people location share [9] or the *Pensieve* system developed by Peesapati et al. [10] to study personal reminiscence; and personal residential air quality monitoring, with *MAAV* developed by Moore et al. [3], *InAir* developed by Kim et al. [11]–[13], or *AirSense*, developed by Fang et al. [14].

A recent survey of over 500 personal informatics publications by Epstein et al. shows, however, that the field focuses more on designing tools that support data collection and reflection, and less on tools to support sense-making and analysis [15]. Existing personal informatics tools that *do* support analysis do so in a specific way: they remove the analysis burden of data processing and task operationalization by baking-in specific analysis workflows. These workflows are designed by personal informatics researchers and not the people who are collecting and analyzing their personal data [16]. This disconnect between designer and user results in tools that do not always reflect the ways people want to engage with or think about their data [17]. As a result, a gap exists between the capabilities personal informatics tools support for helping people to engage with personal data, and the range of tasks people want to perform on the data they collect.

This gap presents an opportunity for the visualization community to make an impact in this space by leveraging their experience with helping people engage with data. The visualization community, however, has primarily developed tools and techniques for helping domain experts work with data [1]. For example, whereas visual analytics research is dedicated to helping people answer diverse questions through the development of flexible analytics tools, this work is conducted for domain experts working in professional contexts [18], [19] who bring motivations, skills, and experiences to data analysis that cannot be assumed about people working in personal contexts [16]. On the other hand, personal visual analytics research has emerged to unify several threads of visualization research that "empower everyday people through exploring data" within personal contexts [16]. Despite consolidating several research areas focused on nonprofessional contexts, these approaches are not designed to support in-depth analysis of personal data. The subfields within everyday visualization are instead primarily oriented toward promoting data awareness, exploration, or social sharing [1].

We identify and define this lack of focused research and design of systems that allow people to flexibly analyze their personal data as the *personal informatics analysis gap*. This gap is a consequence of a collective oversight between individual fields that each contribute toward helping people engage with data, but that are yet to direct their individual strengths toward addressing this gap.

1.2 Contributions

This dissertation's primary contribution is identifying and defining the personal informatics analysis gap, and providing recommendations for how researchers can help bridge it [1]. We ground this gap within the personal informatics and visualization literature, and validate it through a multiyear longitudinal study on ways asthmatic families engage with personal indoor air quality data [1], [3], [4]. Our formative work began with a long-term deployment of several indoor air quality sensing systems to understand how asthmatic families would use them [3]. Afterwards, our attempts to design an improved visual analytics tool for indoor air quality stalled from having focused only on how participants used their system, rather than what participants wanted to do, or would know to do, with their data.

To gather this information, we developed a new interview method: the data engagement interview [4]. Conducting these interviews with our study participants allowed us to observe how they actually use their data to answer personal questions, rather than relying on post hoc rationalizations of their process, as with other methods. These observations gave us a better understanding of participants' needs and abilities for working with their air quality data, and the ways that surprise, exploration, and play can influence their workflows [1]. Analyzing these results in combination with an extensive literature review helped us to identify the personal informatics analysis gap as the source of our earlier design challenges, and motivates our recommendations for new design priorities within the visualization community as a first step to overcome these barriers. We recommend designers focus on supporting entry points into personal data, prioritizing play as a first-class design criteria, and creating more social and collaborative analytic systems. Our identification and recommendations for the personal informatics analysis gap form the core contribution of this dissertation, and stand to transform the way designers approach supporting people to flexibly analyze their personal data.

This dissertation also includes several secondary contributions. We offer insights on how asthmatic families use an indoor air quality sensing system, and recommendations for ways to improve their overall engagement [3]; we create the data engagement interview method as an effective way of engaging people with their personal data, both for helping researchers understand and identify design requirements for everyday visual analytics systems, and for helping participants gain a deeper understanding of their personal data [4]. We also present our multiyear qualitative study as its own research artifact, and as one possible template for how this kind of research can be reported within the visualization community.

1.3 Organization

The remainder of this dissertation is organized as follows: Chapter 2 identifies and defines the personal informatics analysis gap through a review of personal informatics and visualization research literature to highlight existing gaps for supporting people to engage with personal data. Chapter 3 introduces and describes our study participants and reflexively summarizes our broader longitudinal study timeline. Chapter 4 reports on the initial field deployment of a wireless indoor air quality sensing system to the study participants, prefaced with its contribution to the overall dissertation theme. Chapter 5 describes the data engagement interview, a novel interview method for helping us to better understand our participants' needs when engaging with personal air quality data, and a valuable tool for understanding the personal informatics analysis gap. Chapter 6 presents the outcomes from exploring the personal informatics analysis gap, and includes a discussion on implications and design opportunities in this space. Chapter 7 concludes this dissertation and provides suggestions for future research. This dissertation excerpts previously and soon-to-be published work [1], [3], [4].

CHAPTER 2

IDENTIFYING AND DEFINING THE PERSONAL INFORMATICS ANALYSIS GAP

This chapter reviews literature from the personal informatics and visualization communities to provide a theoretical grounding for the personal informatics analysis gap. The following sections describe how each field supports people to engage with data, and calls attention to a critical lack of design knowledge for supporting *personal* data analysis. These sections conclude with small summary observations on each field that we combine to identify and define the personal informatics analysis gap.

2.1 Personal informatics tools

Li et al. defines personal informatics as a research area in human-computer interaction that focuses on the "tools and technologies that help people to collect personally relevant data for the purposes of self-reflection and gaining self-knowledge" [5]. Interpreting and reflecting on these data in any meaningful way requires a deep, contextual awareness of people's lives, and is critical for drawing interesting conclusions or making insightful discoveries. Work by Tolmie et al. and Fischer et al. exploring how people make sense of personal data find that researchers who lack the situated knowledge of someone's social contexts, routines, and priorities are unable to correctly interpret others' logged personal data, much less meaningfully analyze or make recommendations from them [20], [21]. Thus, personal informatics researchers have prioritized ways of supporting people to capture and reflect on their own data [15].

Personal informatics emerged from the quantified self movement with early tools designed to help users gain self-insights by tracking single facets of their lives, such as their diet [22] or physical activity [23]. Work by Bentley et al., however, report that reflecting on single data streams limited the kinds of insights people could derive from their data [24].

Thus, research emerged for tracking multiple facets of people's lives to help sustain user engagement and improve insight generation, including Epstein et al.'s processing and visualizing personal life log data [17]. This shift, however, imposes a greater analytic complexity that users find unmanageable [25]–[27]. Jones et al. report how attempts to curb this complexity either seek to prioritize smaller, more manageable tracking tasks as a way to narrow the design and problem spaces, or to implement analytic pipelines that automatically detect and present potentially interesting correlations based on statistical analysis [26], [27]. Huang et al. comment how these solutions involve fixed processing pipelines that give the user little control over the kinds of analysis or types of insights afforded to them [16], with work by Bentley et al. finding the outcome of these decisions results in frustration about receiving obvious insights [24], and Jones et al. reporting on the information overload of having to review too many potential correlations [27].

One step toward supporting more interactive and exploratory personal data analysis involves a technique that Epstein et al. call data *cuts* [17]. Cuts refer to subsets of data, chosen by underlying features, that support detailed and potentially interesting comparisons. In building a tool to support the analysis of cuts, Epstein et al. developed a set of predefined cuts based on a survey of the kinds of questions users had of their personal data. User evaluations of these cuts found them to be effective at supporting *some* of their questions, but the predefined cuts failed to meet everyone's needs or preferences for engaging with their data. These findings echo those by Jones et al. and Bentley et al. that attempt to automate insights [24], [27], and led Epstein et al. to recommend that "designs do not attempt to limit cuts based on stated goals and instead offer a variety of cuts" [17]. What remains unclear, however, is how to design for arbitrary and flexible analysis, especially in the context of open-ended or ill-defined goals.

Alongside system design, personal informatics research has also developed models to describe the ways and reasons people track personal data [5], [28]. The stage-based model for personal informatics systems [5] classifies how people self-track in practice, and outlines an iterative five-stage process for users pursuing goal-oriented behavior change: *preparation*, *collection*, *integration*, *reflection*, and *action*. Later, the lived informatics model [28] revised the stage-based model to reflect a more inclusive classification of people's tracking behaviors and motivations (Figure 2.1). Specifically, it acknowledges users' decisions of

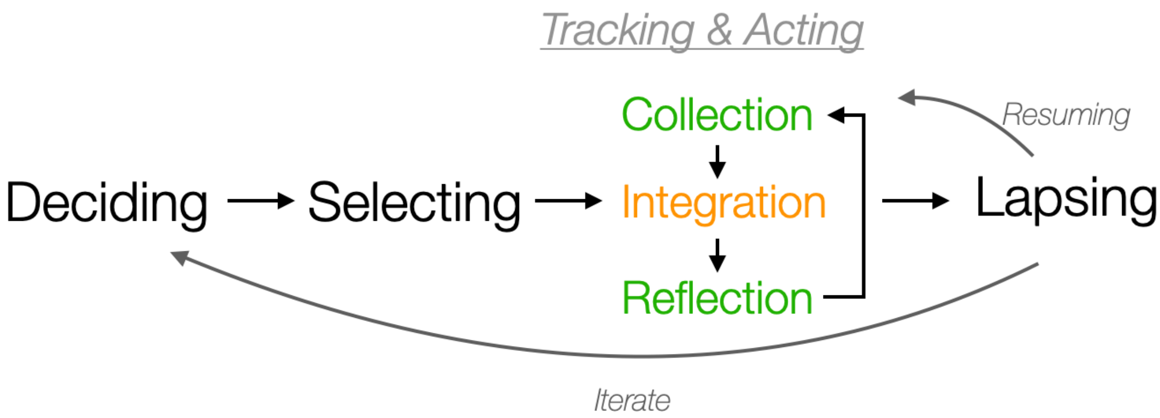


Figure 2.1. The lived informatics model of personal informatics [28]. A recent survey of over 500 personal informatics publications [15] found researchers focus primarily on the collection and reflection stages (green) of the tracking and acting cycle, and less on the integration stage (orange), where data are combined, transformed, and analyzed.

whether and how to track personal data, their ability to lapse and resume self-tracking practices, and interest to track or review personal data for non-goal-oriented motivations. These models are foundational in the personal informatics community, and guide how researchers develop and study systems that help people improve aspects of their lives.

Personal informatics research, however, is not uniformly distributed over these model stages. A recent retrospective survey of over 500 personal informatics publications by Epstein et al. found considerable research that addresses barriers in the collection and reflection stages, but with significantly less focus on integration [15]. The integration stage remaining understudied — the stage where data are combined, transformed, and analyzed — echoes findings by Jones et al. on the inherent difficulties with supporting these functionalities [27]. Tackling this challenge requires more attention and research efforts to raise greater awareness of these disparities; the work presented in this dissertation is a step in that direction.

Observation 1: Personal informatics leverages people's deep contextual knowledge to collect and reflect on personal data, but requires further study to develop flexible analysis tools that empower people to engage with their personal data in unique ways.

2.2 Flexible visualization tools

Visualization research has a long history of developing tools to help people productively engage with data, but it has largely targeted those working in professional contexts.

Developing visualization tools for use in professional settings stems from two core threads of research: the design of bespoke tools for domain experts, and the creation of powerful and flexible systems for data analysts.

A proliferation of bespoke tools for supporting rich visual analysis across a broad range of fields stems from the visualization research community's embrace of Munzer's call for "collaborating closely with domain experts who have appropriate driving tasks in data-rich fields to produce tools and techniques that solve clear real-world needs" [29]. The dominant research approach behind the design and development of these tools is *visualization design study*, an approach to problem-driven visualization research described by Sedlmair et al. that emphasizes designing visual analysis tools in close collaboration with domain experts [19]. Design study is now a standard method for conducting visualization research inquiry, informed by validation methods like Munzner et al.'s nested model [30], various process models including Sedlmair et al.'s design study methodology [19], Mckenna et al.'s design activity framework [31], and McCurdy et al.'s action design research [32], as well as Meyer et al.'s criteria for rigor [33], and Sedlmair et al.'s seven guiding scenarios for design study contributions [34]. Published research papers reporting on design studies cover a range of application areas, but they all focus on developing solutions for domain experts *working* with data. None report on collaborations outside a professional context.

Research into visualization systems for data analysts arises from the significant increase in the number of professionals whose primary task is data analysis. Various interview studies focus on how data analysts do their work, with Kandel et al. characterizing analysis processes within an enterprise organizational context [35]; Alspaugh et al. detailing patterns of exploratory data analysis [36]; Kandogan et al. reporting impediments to efficient data analysis [37]; Kim et al. outlining their roles within software development teams [38]; and Fisher et al. presenting their unique considerations when working with cloud architectures [39]. These studies extend and modernize earlier research on intelligence analysts' work practices by Pirolli & Card [40],[41], Cowley et al. [42] Patterson et al. [43], or Wright et al. [44] by seeking to understand the sense-making and information-foraging processes of data analysts, and what Russel et al. call the cost structure of sensemaking [45]. The results of these studies inform a growing ecosystem of tools for data wrangling like Wrangler [46] or Origraph [47] , widely used interactive visual analysis languages like D3.js [48] or Vega [49], [50], and research into visualization recommendation systems by Mackinlay et al. [51], Wongsuphawat et al. [52], [53] or Gadhave et al. [54].

Work by Amar et al. [55] and Chul et al. [56] report how these professional tools require users to translate their questions into accompanying analysis tasks and make appropriate decisions based on visualizations that they see. Chul et al., along with studies by Law et al. [57] and Grammel et al. [58] find, however, that people without these skills struggle to use visualization tools effectively. Generally speaking, people engaging with data in personal contexts tend to have less time, training, patience, motivation, capabilities, and crisply actionable tasks for their personal data than do professional analysts [16].

Observation 2: Visualization research excels at designing flexible data analysis tools for experts working in professional contexts, but does not target personal contexts, where existing tools do not easily transfer to the needs, skills, and motivations of people exploring their own data.

2.3 Everyday visualization

Visualization's increasing use and consumption in everyday contexts has spurred new research into making data accessible and understandable to everyone. Huang et al. define personal visualization and personal visual analytics subfields to help formalize this *personal context*, and call attention to the more everyday uses which include "different motivations, priorities, role expectations, environments, or time and resource budgets as compared to professional situations" [16]. This distinction was made to unite largely independent research communities within visualization and personal informatics that, by virtue of focusing on nonprofessional situations, extends to cover a broad range of use cases and data scopes.

Huang et al. acknowledge how this grouping "subsumes many related fields." However, applying this broad definition over several fields can make it difficult to articulate a common thread of research across these diverse areas, especially for use cases that do not actively include personal data. This distinction is especially important as design considerations and analytic processes underpinning visual analysis systems for engaging with personal data differ from developing systems that engaging with nonpersonal data in these personal contexts [20], [21]. A report by Lee et al. on reaching broader audiences with data visualization further identifies a lack of significant activity in this space, finding that "only a limited number of researchers have continued to work at the intersection of visualization and personal informatics"[59]. In an attempt to bring some clarity to this broad field, we introduce *everyday visualization* as a descriptive label to distinguish visualization research conducted within personal contexts, but that fails to leverage people's deep personal context in data exploration, or prioritize flexible analysis of personal data.

Everyday visualization encompasses multiple use-cases and research goals. Early work in this space explores ways to democratize visualization with *ManyEyes* [60],[61], *Vizster* [62], and *Sense.us* [63], engage people with new visualization techniques for familiar information like the stock market [64] or baby names [65], and help bring visualization to the people [66] with exploring data through constructive physicalizations [67], [68]. This community has since expanded its scope to investigating physicalizations for awareness or goal-setting [69]–[73], engagement with personal data in the home [74], and even emotional connection to personal data when mapped to living artifacts [75], [76]. Although work by Koytek et al. incorporates personal agency into visual interaction mechanisms [77] and Aseniero et al. explores proof-of-concept tools to support self-reflection in controlled lab studies [78], this work does not incorporate participants' personal data as a part of the analysis process. Instead, the overwhelming majority of research occurring within personal contexts remains rooted in exploration, awareness, or social sharing. Everyday visualization has yet to deploy truly flexible, scalable, or in-depth data analysis capabilities for personal data.

Observation 3: The democratized analysis goals of everyday visualization aim to empower people to explore data, but this field has yet to design for flexible, in-depth analysis of personal data.

2.4 The personal informatics analysis gap

We present the personal informatics analysis gap as the collective oversight emerging from how visual analytics, everyday visualization, and personal informatics fields focus their research efforts. This gap poses a multifaceted design challenge that individual fields have struggled to address with their singular design perspectives:

- Personal informatics approaches the gap with considerable experience building analysis tools for working with personal data. These tools, however, offer inflexible, designer-defined analysis pipelines that cannot be reactive to the diverse ways people may want to use their data.
- Visual analytics approaches the gap by applying methodologies developed for and alongside professional users to study how everyday people work with personal data. This approach, however, is complicated by the differing motivations, training, and resources available for users operating in personal contexts.
- Everyday visualization approaches the gap with its focus on how people engage with data outside of professional contexts. This focus, however, does not fully bridge the gap from a lack of significant activity at the intersection of visualization and personal informatics, and limited research that incorporates personal data and contextual knowledge to develop tools for flexible personal data analysis.

These limitations have so far prevented researchers in each field from addressing all individual strands of the personal informatics analysis gap. As a consequence, the personal informatics analysis gap emerges as a lack of focused research and design of systems that allow people to flexibly analyze their personal data. Any solution that overcomes this gap must acknowledge and incorporate people's personal and contextual knowledge, support flexible analysis that covers a variety of different circumstances and goals, and empower people to deeply and richly engage with their personal data. Enlisting each field's strength affords an opportunity for developing new approaches to learn how people engage with personal data, and for designing new tools and systems that will support them in doing so. Personal informatics' experience with collecting and acting on personal data can be augmented by visual analytics' background in customized analysis environments, and brought together with the everyday analysis goal of personal empowerment. Uniting these fields will help researchers consolidate existing design knowledge, experience, guidance, and methods to more effectively design visual analysis tools that support people engaging with their personal data. This dissertation reports on our explorations of the personal informatics analysis gap that drew from each of these fields' expertise over a multiyear longitudinal study of asthmatic families engaging with their personal indoor air quality data.

CHAPTER 3

LONGITUDINAL STUDY OVERVIEW

This chapter describes our journey toward identifying the personal informatics analysis gap. We tell this story through a reflexive summary of our multiyear study, and explain how our individual research activities inform this work (Figure 3.1). Although later chapters present more detail on these specific activities – field deployments in Chapter 4, data engagement interviews in Chapter 5, and the results and design implications from probing the personal informatics analysis gap in Chapter 6 – this chapter provides the necessary contextual knowledge for understanding how these individual contributions fit into this dissertation's broader narrative.

We start by introducing our longitudinal study participants in Section 3.1. Afterwards, we describe how our collaboration with these participants evolved from studying *them* in hopes of developing an improved visual analytics tool, to studying *ourselves*, and the lack

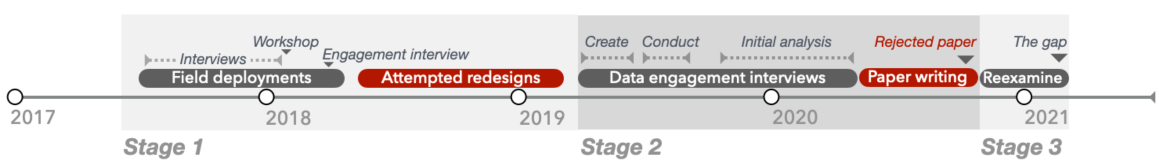


Figure 3.1. The longitudinal study timeline, grouped into three stages that describe our path toward identifying the personal informatics analysis gap. We separate stages by fundamental changes in how we understood and approached our underlying research goals: stage 1 describes our air quality system field deployments [3] and intent to redesign a visual analytics interface for processing air quality data. Stage 2 outlines our creation and application of the data engagement interview method [4] for capturing design insights into what participants wanted to do with their data, and continued attempts to characterize participants' questions for developing a revised interface (Appendix C). In stage 3 , we reanalyze participants' data engagement interviews and draw connections between existing research gaps within visualization and personal informatics communities, and why we needed a new interview method. We identified and defined the personal informatics analysis gap [1] in this stage.

of research on designing systems that help people to flexibly analyze personal data. Section 3.2 tells this story by summarizing our work in three research stages. These stages outline how our challenges and outcomes helped shift our focus away from developing a visual analysis tool, toward identifying and defining the personal informatics analysis gap.

3.1 Study participants

We recruited our study participants from a pool of families that were also participating in a concurrent national research program to measure factors affecting pediatric asthma [2]. This research program involved multiple universities, with the University of Utah acting as the informatics center. In collaboration with the College of Nursing, who ran and supervised the national asthma project, we recruited eight families that were local to the Salt Lake Valley area and that had at least one family member with moderate or severe asthma. To be eligible for our study, these families needed a high-speed Internet connection, a wireless home network, and the willingness to host an indoor air quality system. Two families dropped out prior to starting our study, citing measurement fatigue from the national asthma project, leaving us with six families that persisted throughout our longitudinal study. Figure 3.2 illustrates the individual family make-ups and distribution of our asthmatic participants.

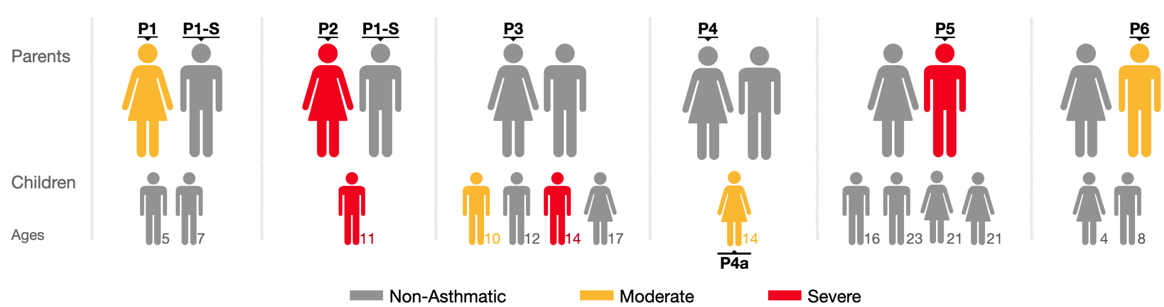


Figure 3.2. The six participant families enrolled in our longitudinal study. Labeled participants are those most engaged from each family, and from whom our study findings derive. Notable participants include P4a, the only adolescent participant in our study, and participant spouses P1-S and P2-S who participated in stage 2 only. Yellow and red figures represent moderate and severe asthmatics, respectively. Remaining figures are nonasthmatic.

Although we encouraged all family members to participate throughout this study, we did not require their complete involvement. As a result, each of the households settled into a pattern where a singularly motivated participant assumed the primary communication and feedback role. We denote these participants as *primary participants*, labeled P1 - P6 in Figure 3.2. We included participant P4's teenage daughter, P4a, as an additional primary participant given her significant engagement in this study. No other children actively participated. Spouses of participants P1 and P2, hereafter labeled P1-S and P2-S, also contributed feedback and suggestions during the data engagement interviews [4] in stage 2 but were otherwise not involved in the study and are not counted among the other primary participants. For brevity, we will often refer to individual primary participants simply as *participants*, or by their specific labels when attributing quotes, traits, characterizations, goals, and research outcomes throughout this dissertation.

Participants were themselves asthmatic (P1, P2, P4a, P5, P6), or primary caregivers to asthmatic children (P2, P3, P4). Participants P1, P5, and P6 are the only asthmatics in their household, whereas P2 was the only asthmatic parent with an asthmatic child. Of the primary participants in our study, 5 were female (P1, P2, P3, P4, P4a) and 2 male (P5, P6). All adult participants had received a high school diploma, with P2 completing some college education with no degree, and P1, P3, and P4 having a college degree. P5 and P6 had master's degrees. P1 and P2 are stay-at-home mothers, P3 is a web developer, P4 works as a nurse, P5 as a school administrator, and P6 on public policy.

Although the participants were alike in the fact that asthma had impacted their lives, the extent and degree to which they were affected was entirely personal. Participant P1, an asthmatic sensitive to pollen and other outdoor irritants, used her air quality system as a personal planning tool for deciding when to stay inside if outdoor particulate concentrations were high. Participant P2 is medically disabled as a result of her severe asthma and interested in how the data could improve her quality of life. Due to her health complications, however, she remained hesitant to interpret or act on any personal data. Instead, she preferred to receive personalized advice from medical professionals on ways she could improve her indoor air quality. Participants P3 and P4 are primary caregivers to asthmatic children and wanted to use their data to provide a healthy home environment for their family. Participant P4a, our youngest participant, was less concerned with the health impacts of air quality and more curious to see how air quality affected those around her at a community level. Participants P5 and P6 are adult asthmatics who already understood their symptoms and were interested to explore how personal and environmental factors affected their sensitivities. Table 3.1 summarizes how participants' asthma conditions influenced the way they engaged with their data, and what they sought from it.

3.2 Research stages

This section provides a reflexive summary of our longitudinal study stages, shown in Figure 3.1. For each stage, we describe our research goals at that point in time, the steps taken to achieve them within that stage, and then reflect on how subsequent challenges and outcomes helped to reshape our goals. These changes helped to shift our focus away from designing an improved visual analytics tool for indoor air quality, and toward better

Table 3.1. Participant characterizations.

P1	Stay-at-home mother and moderate adult asthmatic. Self-identifies as nontechnical, passive, and reactionary with regard to asthma management. Admitted she is not overly engaged with checking air quality conditions, despite sensitivities. Primarily interested in using MAAV for informing her family's outdoor activities.
P2	Medically disabled from asthmatic symptoms. Nontechnical, but engaged due to her and her child's severe asthma symptoms. Actively seeks air quality forecast information through a combination of phone app and local forecasts. Curious about what she could learn through her data, although prefers professional interpretation to her own.
P3	Work-from-home mother and nonasthmatic. Two of 4 children are asthmatic: one suffers from Common Variable Immune Deficiency and severe asymptomatic asthma. Interested in using her data to monitor indoor air quality to guard against potential triggers, and self-experimentation to find poor air quality sources.
P4, P4a	Nurse at a university hospital and caregiver to a teenage daughter with moderate asthma (P4a, student). Both were initially disengaged from air quality monitoring, but became more aware over the course of their deployment. Interested in drawing health correlations from visualizations.
P5	Public school administrator and severe asthmatic. Already aware of personal triggers, P5 was eager to use his data to identify spike sources and understand room activity levels.
P6	Public health employee and moderate asthmatic. Motivated to use data for characterizing living spaces and improving his sense control over indoor air quality.

understanding the challenges we experienced in pursuit of that goal. This shift ultimately led us to identify the personal informatics analysis gap.

3.2.1 Stage 1: Field deployments

Our study began with a year-long field deployment of prototype indoor air quality monitoring systems to six asthmatic families [3] (Figure 3.3). These families hosted their air quality systems to measure and understand the air quality in and around their homes. These systems worked by measuring and logging airborne particulate matter levels over time, with additional support to interactively review and annotate these data using a real-time interactive interface, text messaging, and a smart speaker. Chapter 4 provides more information on this system and deployment outcomes, with work by Lundrigan et al. [79] detailing the system's technical development.

Our goal for this stage was to translate observations of how our study participants interacted with their air quality system into design requirements for an improved visual analysis tool. We conducted multiple rounds of semistructured interviews over several months to collect feedback on what people knew about air quality (Phase I, Section 4.4.1.1); their expectations on how they might use their system (Phase II, Section 4.4.1.2); and how they engaged with the system after having lived with it for several weeks (Phase III,

Figure 3.3. Participants' air quality sensing system. From left to right, participants received an interactive interface for viewing their air quality data, a smart speaker for making annotations on their data, three wireless air quality monitors for detecting air quality in and around their home, and a Raspberry Pi gateway computer to collect and transfer data to a centralized server. More detailed information on this deployment can be found in Chapter 4.

Section 4.4.1.3). These interviews succeeded in capturing insights on *how* participants used their air quality systems (Section 4.5), but not *what* they wanted to do with the data their systems collected. Without knowing participants' underlying goals or motivations, we were unable to design an improved visual analysis tool for answering the kinds of questions that participants' existing air quality interface could not address.

In response, we conducted a creative visualization opportunities workshop [80] to capture these needs. These workshops are used in visualization research as a way to quickly identify potential design requirements for solving visual analytic challenges. Collecting feedback on what participants wanted to know, see, or do with their air quality data, however, revealed a broad range of high-level goals and motivations. Results from qualitatively analyzing participants' deployments (Appendix A) and workshop feedback (Appendix B) highlighted their diverse and personal goals, which included such questions as: "What type of vacuum cleaner should I buy?", "What's the air quality like when other people are sick?", and "Should we move?".

Although our analyses helped us identify specific questions that participants imagined using their data to help answer, it was less clear how they might go about finding solutions to these questions. Throughout this process, deeper questions began to emerge: What might our participants think to do with their data, and how would they think to operationalize these questions on their own, if they even could? Designing an improved visual analysis tool meant finding a way to understand what our participants knew to do with their data, and productive ways of engaging them to gathering this information. We needed a new approach.

3.2.2 Stage 2: Data engagement interviews

Whereas stage 1 focused on observing what participants *did* for the purposes of developing an improved interface, stage 2 is characterized by our shift toward investigating what participants *knew*, and how they might approach operationalizing their questions. We found ourselves in a double-bind, however, needing some kind of visual analytics tool to support us in observing how participants might engage with their data, while also needing to observe how participants engaged with their data to develop such a tool.

We looked to visualization and human-computer interaction literature for guidance on engaging people with their personal data to collect design requirements, and whether any such methods existed for using personal data in the process of developing visual analysis systems. We found no methods, however, for conducting this type of study with nonprofessional collaborators. This lack of existing guidance led us to develop our own method, the data engagement interview [4]. This interview method emerged as as an alternative to other analytic-minded interview techniques with two major changes: 1) it is specifically designed for people who are not domain experts; and 2) we incorporate an actual data analyst to conduct real-time visual analysis using participants' data within the interview process. The analyst removes any cognitive burdens on the interview participants from having to wrangle or program their data on their own, so they can more freely converse with the interviewers about their goals, motivations, questions, and thoughts. We conducted these interviews with the same study participants from our earlier field deployments, and incorporated their self-annotated air quality data in the interview process. Chapter 5 provides more information on the development and outcomes of participants' data engagement interviews . In short, this method was very successful at getting participants into and exploring their data, and allowed us to collect more ecologically valid observations on how they went about using personal data for answering their questions. This approach allowed us to gain insights that would not have been possible without directly engaging people with their data.

At the same time, we began wondering why we needed a *new* method to gather design requirements in the first place. Our growing awareness of the disconnect between what we needed and what existed in literature took shape as we analyzed our participants' data engagement interviews for an initial journal submission (Appendix C). This submission captures our goal-oriented design mindset for developing our improved analysis tool at that time, and includes our earliest conceptions of the personal informatics analysis gap. We identify the gap, however, purely through a tool-centric lens, defining it as the space between analytic systems that support answering a broad set of questions, and those that support people operationalizing questions into lower level analysis tasks. This goal-oriented analytic framing informed our entire submission, which also included a taxonomy of participants' questions for defining a preliminary task space for analyzing personal indoor air quality, and a proposed data model for how designers might support analytic workflows in this space. We also included an initial write up of our interview method, at this point named the *task elicitation interview*, further emphasizing our analytic and goal-oriented focus for developing a visual analysis tool.

In hindsight, we were too quick to overload our submission with underdeveloped concepts. Our reviewers felt similarly, and rejected our submission with feedback that its contributions were simultaneously overwhelming, unvalidated, and insufficient given its broad scope and framing. Reviewers remarked that our submission contained multiple papers' worth of content, and instead recommended we further develop these ideas as separate future submissions. This rejection and feedback helped us to shift our research focus toward more deeply exploring the reasons for creating the data engagement interview method [4], and to further articulate the nature and origins of the personal informatics analysis gap [1] as separate papers. This decision was a critical inflection point on our journey to identify the personal informatics analysis gap, and motivated our switch away from developing a visual analysis tool, toward investigating the source of these challenges.

3.2.3 Stage 3: The personal informatics analysis gap

Having abandoned our tool-centric design focus, stage 3 reflects our complete shift toward examining the source of the challenges that impacted our previous design attempts as the primary research goal. This stage unfolded over a 3-month period during which we revisited our participants' data engagement interviews from stage 2 to conduct a more thorough analysis of our interview transcripts. This analysis allowed us to formalize our interview framework (Section 5.4), explore how it unfolded in practice through a case study working with personal air quality data (Section 5.5), and identify the unique outcomes made possible by this method when engaging participants directly with their data (Section 5.6). This process also brought about the results we present in Chapter 6 on the personal informatics analysis gap, about how our participants engaged with their data (Section 6.1).

At the same time, we revisited our literature review with a deeper look into three fields that could help us understand why we needed a new method for engaging our participants, and where our difficulties with eliciting design requirements might come from (Chapter 2). We targeted personal informatics, a field focusing on tools to help people collect and reflect on personally relevant data for gaining self insights; visual analytics, and its history of developing flexible analysis tools for domain experts typically working with large and complex datasets in professional contexts; and a collection of fields we call everyday visual analysis, which includes research focused on engaging and empowering people using visualizations, and performing visual sensemaking outside of professional contexts.

Our deeper literature review helped us identify that each field has focused on helping people to engage with data under specific circumstances and contexts, yet no one field had prioritized supporting people to flexibly analyze personal data (Section 2.4). This discovery explained the necessity for creating the data engagement interview method, and spoke to a broader lack of focus at the community-level on supporting flexible personal data analysis. These events led to us defining the *personal informatics analysis gap* as a lack of focused research and design of systems that allow people to flexibly analyze their own data.

Realizing this gap — both through the literature and in our own experiences — motivated us to consider visualization design through a different lens. This perspective highlighted how the established methods we had relied upon were developed for domain experts, whose experiences and priorities did not reflect the circumstances of our own participants, and whose workflows conflicted with how we observed our participants engaging with their data (Section 6.1). This insight prompted a drastic rewrite of our earlier journal submission as two separate papers: a rigorous identification and definition of the personal informatics analysis gap [1], and a formalization of our data engagement interview method as a first step toward exploring this gap [4].

Revisiting our targeted literature review and participants' interview data also provided new insights into how we, as visualization design researchers, might design future visual analysis systems for engaging with personal data. We recommend alternative design goals that prioritize entry points into personal data, treat play as a first-class design criteria, and place more focus on designing collaborative analysis systems. Our core contribution and critical outcome of this dissertation is a call to action for rethinking how and what we design for engaging with personal data in the face of the personal informatics analysis gap . The following chapters provide more detail on how this longitudinal study began (Chapter 4), how we developed and conducted data engagement interviews with our study participants (Chapter 5), outcomes from using this interview method for exploring the personal informatics analysis gap (Chapter 6), and our recommendations for designing with the gap in mind (Section 6.2).

CHAPTER 4

MANAGING IN-HOME ENVIRONMENTS THROUGH SENSING, ANNOTATING, AND VISUALIZING AIR QUALITY DATA

This chapter describes the start of our longitudinal study, with a long-term field deployment of air quality sensing systems to six asthmatic families. These deployments were designed to serve two purposes: first, to help our study participants to better understand and manage their indoor air quality, and second, for us to observe how participants engaged with their systems to inform follow-on opportunities for designing improved visual analytics tools for working with air quality data. Our subsequent difficulties in designing this improved tool, however, initiated a multiyear qualitative study on the challenges visualization designers face when trying to develop visual analysis systems for personal data. This chapter is formative to this dissertation as the starting point to a deeper investigation into the difficulties of conducting visualization design research outside the bounds of professional environments.

We report how participants engaged with their deployed systems by analyzing system usage data and over 20 hours of interview audio from three rounds of semi-structured user interviews. This analysis provides design takeaways for other tool builders in this space. A secondary contribution of this study is the insight on ways tool builders can encourage and maintain high levels of self-tracking over a long-term deployment through supporting data annotation, using system-initiated notifications on mobile devices, and supporting users to interactively explore their data.

4.1 Introduction

The World Health Organization estimates approximately 3 million people die annually as a result of ambient air pollution [81]. Exposure to fine-particulate matter ($PM_{2.5}$) – particles with diameters smaller than 2.5 microns – has the greatest adverse health effects among air pollutants, linked with an increased incidence of cardiac arrhythmia, lung cancer, heart disease, and mortality [82]–[86]. Levels of $PM_{2.5}$ in urban areas are measured largely by a sparse distribution of expensive, government-run sensors that fail to capture known microenvironments of $PM_{2.5}$ [87]–[89]. Recent advances in sensor technology, however, have enabled motivated citizens, grass-roots organizations, and researchers to bring new, low-cost, real-time sensors online to address measurement gaps [11], [90]–[92]. These advances in sensor technology have also helped to improve our understanding of outdoor air quality conditions and pollution sources [93]–[96].

Although outdoor air quality is a growing concern for urban areas around the world, studies find poor correlations between outdoor $PM_{2.5}$ levels and personal exposure measurements due to the large percentage of time that people spend inside [97]. For example, most Americans are estimated to spend upwards of 90% of their time indoors, with about 70% of their day spent at home [98]. To empower residents to understand and modify their personal, indoor environments, recent studies deploy low-cost air quality monitors coupled with visualizations of the sensor data streams inside homes. Kim et al. developed their *inAir* system to help visualize and communicate indoor air quality levels to residents [11]–[13], Fang et al. developed *AirSense* to detect and classify indoor airborne pollutant sources [14], and the *MAQS* system developed by Jiang et al. [99] fielded a wearable sensor to track personalized exposure to indoor air quality pollutants. This work demonstrates the value of air quality monitors in residential environments, and participants in these studies reported being more aware of the air quality in their homes and more engaged in its management.

These same indoor studies, however, identify important deployment limitations. Studies deploying only one monitor per home [11]–[14], [99] cannot identify indoor and outdoor microenvironments, which requires multiple monitors in order to reliably detect and characterize $PM_{2.5}$ variability [100], [101]. Indoor and outdoor monitor placements are also needed to understand the effect of outdoor conditions on indoor air quality [102]–[105].

Furthermore, interpreting sensor data is often challenging without additional context [20], and user-driven labeling of air quality events is additionally complicated by air quality's invisibility. $PM_{2.5}$ levels are often not immediately or inherently apparent to residents in the moment, increasing the likelihood of forgetfulness when revisiting data to annotate after the fact [106]. These issues limit the ability of residents to effectively characterize and improve the air quality of their homes.

The goal of this field deployment is to capture the added value for residents when they have access to an air quality monitoring system that collects data from multiple monitors, supports proactive and in situ data annotation, and presents real-time air quality data and annotations in a interactive visualization. To accomplish this, we developed MAAV, a system to **M**easure **A**ir quality, **A**nnotate data streams, and **V**isualize real-time $PM_{2.5}$ levels. MAAV includes multiple air quality monitors placed both inside and outside a home to capture $PM_{2.5}$ microenvironments; three different annotation modalities to enable residents to contextualize data streams, including a system-initiated prompt; and a tablet-based interactive visualization for exploring measured $PM_{2.5}$ levels and annotations.

We deployed MAAV to six families over a period of 20-47 weeks (mean 37.7 weeks). Over this time, we conducted 34 interviews with participant families to understand their experience with MAAV. Results extracted from the interviews using qualitative analysis show that: 1) MAAV's multiple monitors enabled residents to observe variations in $PM_{2.5}$ activity throughout the home; 2) the availability of multiple annotation modalities led participants to generate many annotations throughout the deployment, sustaining long-term engagement; 3) the interactive visualization and annotations and supported participants to explore and draw insights from their data; and 4) participants remained engaged with MAAV over a long period of time, although with different patterns of engagement initially versus later in the study. These results yield insights about new types of nontemporal comparisons, the value of system-initiated annotation prompts, and the potential for a system interface that changes over time.

4.2 Related work

A significant amount of work explores the technical aspects of air quality sensing from various perspectives and application areas. For commercial contexts, Jin et al. [107], Jablonski et al. [108], and Chen et al. [91] study ways to monitor heating, ventilating, and air conditioning control. Work by Postolache et al. [109], Cheng et al. [91], Jiang et al. [99], and Kim et al. [110] contribute to system infrastructure and platform development, with Nikzad et al. [111] and Jiang et al. [99] focusing on personal exposure monitoring. Saad et al. [112] and Fang et al. [14] on source detection and classification. These studies, and others by Hasenfraztz et al. [94] and Huang et al. [113], primarily focus on a prototype's technical contributions or proof-of-concept systems architecture. Early work by Postolache et al. [109] proposes a multimonitor air quality sensing system, with significant work on modeling, calibrating, and processing sensor data to ensure an accurate system. Jiang et al.'s *MAQS* system presents CO_2 exchange rate and n-gram augmented Bayesian room localization models to estimate indoor air quality and personal exposure, respectively [99]. Both projects develop purpose-built systems for addressing research questions, but they do not explore the needs or questions of the end-users.

The increasing availability of low-cost, commercially available $PM_{2.5}$ sensors enables recent work to focus on human-centered aspects of air quality sensing in the home. Early work by Kim et al. [11]–[13] deploy the *inAir* system to support residents in understanding and managing potential air quality health threats. Using a single $PM_{2.5}$ monitor and ambient display, *inAir* succeeds in increasing participants' awareness of and reflection on indoor air quality, with an improved ability to make connections between in-home activities and air quality levels. Despite visualizing indoor air quality for the user, the system's ambient display does not support user-interactivity or direct data annotation, making it difficult for residents to contextualize events and reason about their causality. *inAir* also elicits feelings of powerlessness and frustration because residents cannot determine the sources of $PM_{2.5}$ spikes in their homes [11]. Furthermore, single-monitor deployments prevent participants from knowing the air quality within multiple indoor areas or the relationship between indoor and outdoor air quality in their homes [12].

Subsequent work demonstrates the ability to automatically classify $PM_{2.5}$ sources within a small set of detection categories. Fang et al. [14] address the issue of $PM_{2.5}$ source classification and targeted feedback with the use of a machine learning model. Their *AirSense* system is able to automatically detect and identify three sources of indoor pollution: cooking, smoking, and spraying pesticide. Their system also estimates personal exposure and provides actionable suggestions to help people improve their air quality [14]. However, similar to *inAir*, this system also fields only a single monitor and does not address issues related to source localization, outdoor air quality, air quality in other parts of the home, or contextualization of collected data.

For those collecting and consuming air quality sensor data, their lack of context can complicate its interpretation. Work by Tolmie et al. [20] identifies *why* annotation is important by demonstrating the large amount of articulation work that goes into interpreting raw sensor data. They conclude that "data generated through [a] networked sensing system is opaque when considered in isolation." These findings highlight the importance of supporting annotation in sensing systems not just to improve sensemaking, but also to reduce the amount of misplaced effort and resources related to incorrect data interpretation. Yet, no systems support this low-barrier creation, overlay, and review of annotation to contextualize the data. We incorporate these findings into MAAV and support multiple integrated annotation mechanisms to allow participants to label and contextualize their indoor air quality data in situ as they are collected.

Very limited work has been done to explore the nature of participant engagement with air quality sensing systems over longer term deployments. Kim et al. find participants' degree of engagement remained constant in separate 4-week [12] and 4-month [13] studies, although these results were based on self-report. Their broader work primarily focuses on the quantifiable air quality improvements via participants' behavior change – not the evolution of their interactions with the system over time. Fang et al. similarly find that hosting an air quality sensing system improves participants' sense of awareness and competence over 6- and 10-week deployments, but their study does not explore the nature of participants' engagement or use of the system beyond this period [14]. Our field deployment captures shifts in participant engagement over long-term deployments ranging between 20 - 47 weeks (average 37.7 weeks) relating to changes in curiosity, familiarity, goals, and seasonality.

Our focus on end-users and data legibility is in line with a variety of research conducted broadly in the space of eco-feedback, not just for air quality. For example, past work discusses the importance of *disaggregated* measurements [114] to support a finer granularity of resource usage akin to multimonitor deployments for measuring air quality in different parts of the home. Past work has also explored the importance of technology probes and display designs in the area of water conservation [115], specifically emphasizing views that support different levels of data and time granularity. This same work prioritizes data comparison for highlighting relative differences and deemphasizing absolute values and units. MAAV similarly supports participants to make relative comparisons between monitors and over different granularities of time. Finally, research shows that systems providing clear, specific, and frequent feedback via computerized and interactive tools successfully evoke lasting engagement [116]. Each of these pieces – interactivity, annotation, comparative views, and active feedback – come together in MAAV.

Building on this breadth of work, we deploy a flexible and open-ended technology probe utilizing multiple wireless $PM_{2.5}$ monitors, a host of annotation mechanisms, and an interactive tablet interface to identify participants' goals and use cases of such a system in the home. This study complements the previous *inAir* [11]–[13] and *AirSense* [14] systems by exploring the ways participants use a multimonitor annotation-enabled system, the questions they develop, and how these interaction mechanisms change over time. To the best of our knowledge, this is the first study combining these system features, and the longest running indoor air quality study with persistent deployments.

4.3 MAAV: A multimonitor air quality sensing and feedback system

Drawing insights from past work, we developed a system to **M**easure **A**ir quality, **A**nnotate data streams, and **V**isualize real-time $PM_{2.5}$ levels, which we call MAAV. MAAV consists of multiple low-cost air quality monitors, a gateway device to upload air quality sensor data to our back end server, mechanisms for users to annotate collected sensor data, and a tablet-based visualization that shows collected sensor data and annotations (Figure 4.1). This project is part of a larger multi-institutional effort to develop epidemiological tools for conducting research on the relationship between air quality and asthma [2]. The following subsections detail each of MAAV's components.

Figure 4.1. Elements of a MAAV deployment. From left to right: a wireless Dylos air quality monitor, tablet display, Google Home and Raspberry Pi gateway computer (front). Wireless connectivity allows distributed instrumentation throughout participants' homes. Each deployment received three air quality monitors but is capable of instrumenting many more.

4.3.1 Low-cost monitors and gateway computer

MAAV brings together air quality data from multiple $PM_{2.5}$ monitors. We use Dylos air quality monitors[1] for their sensitivity at low $PM_{2.5}$ levels and ability to detect the full range of various household spikes (Figure 4.2). These detectors operate via optical light scattering: ambient air is drawn in through the device and across a laser, where an optical sensor measures the amount of light scattered by suspended particles in the airflow. This detected level is output as a $PM_{2.5}$ count, averaged over a 60-second window, and logged each minute.

To integrate the stock air quality monitors within MAAV, we modified each with a BeagleBone Black embedded computer[2] for data formatting and wireless network connectivity. We also installed an improved RGB LCD for communicating system status and measurements to participants and technicians. Once running, the monitors log their

[1]Dylos DC1100 Pro - http://www.dylosproducts.com/dcproairqumo.html

[2]https://beagleboard.org/black

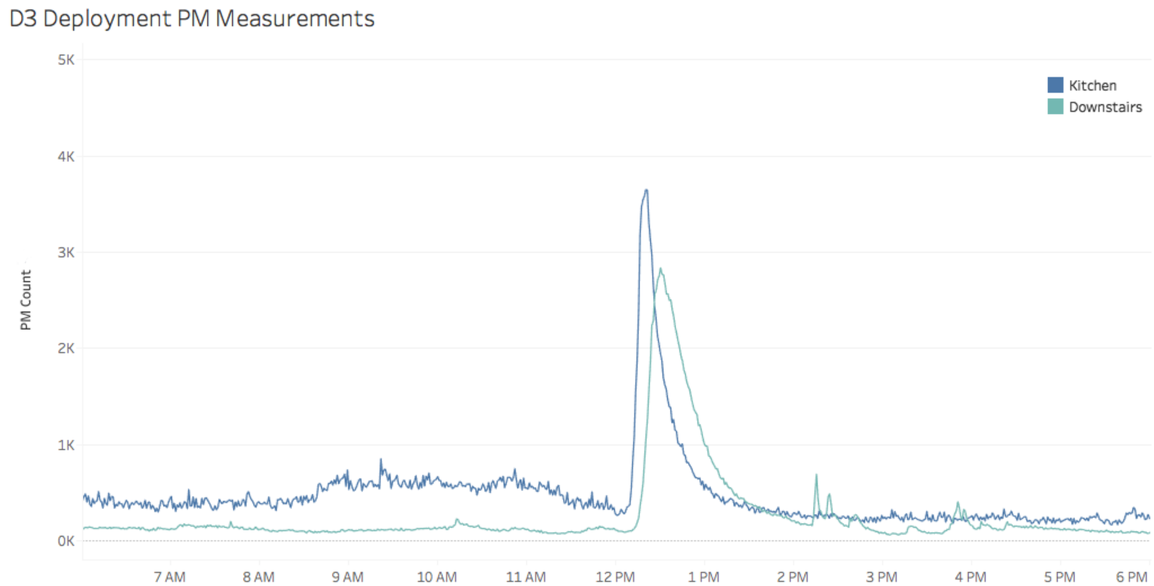


Figure 4.2. A $PM_{2.5}$ cooking spike in the kitchen of Deployment 3. Note its diffusion and registration in the downstairs monitor a short time after. Both rooms return to baseline levels after approximately 2 hours.

measurements to internal memory and transmit them to a Raspberry Pi running Home Assistant[3], which acts as the local gateway computer.

Dylos monitors come precalibrated from the factory and have been found to closely track laboratory-grade air quality instruments [117]. We validated this finding with a week-long pilot study involving 10 Dylos monitors colocated with several calibrated, laboratory - grade air quality monitors[4,5,6] in a researcher's home. All monitors were subjected to a host of domestic activities: lighting and extinguishing candles, vacuuming, changing bed sheets, and doing laundry. The results of these tests verify the monitors' measurement accuracy against calibrated air quality monitoring hardware [118]. During the study, a trained research assistant periodically cleaned participants' air quality monitors follow-

[3]https://www.home-assistant.io/

[4]GRIMM 11-B: https://www.grimm-aerosol.com/

[5]Dustrack: http://www.tsi.com/DUSTTRAK-DRX-Aerosol-Monitor-8533/

[6]minivol: http://www.airmetrics.com/products/minivol/index.html

ing the manufacturer's guidelines for maintaining the board-level air quality sensor but did not recalibrate. To ensure Dylos monitors remained within a satisfactory operational envelope throughout their deployment, we logged a variety of diagnostic and monitor health data, including case temperature, fan speed, and network statistics. Although the potential for sensor drift exists, any such long-term changes were overshadowed by $PM_{2.5}$ spike dynamics and other events examined in this work. Furthermore, the qualitative and human-centered approach of this study minimizes the need for quantitative metrics, instead focusing on the impact of significant air quality changes in the home.

All air quality measurements and participant annotations recorded with this system are saved on a HIPAA-compliant server. Back-end services further monitor, analyze, and alert participants based on their indoor air quality. More information on data aggregation and processing can be found in Lundrigan et al. [79], and our Dylos testing procedures are outlined in Hegde et al. [118].

4.3.2 Contextualizing air quality monitor data with annotation

Past work on data legibility has shown that "personal data generated through networked sensing systems are *opaque* when considered in isolation ... [and] that fine grain understandings of interaction cannot be 'read off' the data alone." [20]. In the context of indoor air quality systems, Kim and Paulos [12] report that a lack of contextual awareness can produce feelings of frustration and powerlessness in end-users and, ultimately, distrust in the system in severe cases. Annotation is therefore essential not only to improve data legibility but also to empower users and help legitimize deployed systems. Finally, the act of annotating can encourage users to reflect [119] on the connection between their daily activities and their indoor $PM_{2.5}$ levels.

The open-ended nature of daily activities that can affect air quality requires manual, user-provided annotation. Manual annotation, however, can introduce a major burden for users. In an attempt to increase the convenience and reduce the burden of manually annotating, MAAV supports multiple annotation modalities: a tablet-based mechanism incorporated into the visualization interface; a text messaging prompt based on automatic detection of $PM_{2.5}$ spikes; and voice transcription via a smart speaker. We also offered participants the ability to manually record their activities in Google Sheets as was done in previous studies [14], but all participants declined. Annotations are stored in their own database, along with annotation modality origin, logs of $PM_{2.5}$ spike activity, deployment spike notifications, and all relevant timestamps. Each annotation modality offers different trade-offs to support different situations where users may want to annotate their data.

4.3.2.1 Tablet

Users can record tablet annotations with a long-press at the desired location within the interface's main view area (Section 4.3.3). A pop-up window displays the selected date and time, along with a textbox for typing the annotation. When submitted, descriptions appear as interactive glyphs on the visualization.

Tablet annotation is effective in situations where users want to annotate an event when reviewing their air quality data in the tablet interface. This modality is not always convenient, however, such as when the tablet is not close at hand.

4.3.2.2 Text messaging

MAAV incorporates text messaging[7] to prompt users for an annotation when a $PM_{2.5}$ spike occurs. An online peak detection algorithm [120] evaluates incoming $PM_{2.5}$ measurements, and those surpassing independent rolling average and standard deviation thresholds are written to the database and generate a text message to the participant, including the spike's location and time: *"We detected a PM spike in the bedroom at 10:56 AM. Any idea what caused it?"*. Users' replies are logged as annotations and associated with the time the $PM_{2.5}$ spike was detected. Participants can also text the same phone number to record an unprompted annotation. Text prompts are disabled during participant-specified quiet hours or within a defined proximity to other detected peaks. For example, if a participant requests quiet hours from 9:00 PM - 9:00 AM, then MAAV will not send text notifications for $PM_{2.5}$ spikes within this time range. Similarly, participants requesting no more than three notifications per hour will have messages rate-limited to a 20-minute wait time. The system uses detection time to determine whether to actively prompt a participant to annotate.

[7] http://www.twilio.com

Text messages are an obvious choice for delivering a prompt to annotate air quality data. Daily activities that may have caused $PM_{2.5}$ to spike can be easily forgotten. The immediacy of receiving a text message and the convenience of responding to that message with a few short words offer users a high-value, low-burden solution. Prompts can be responded to in the moment or at a later, more convenient time. On the other hand, it can be useful to view $PM_{2.5}$ data while annotating, which is not easily supported by text messages. Additionally, past work has shown that people do not necessarily have their phones close at hand while at home [121].

4.3.2.3 Voice annotation

As a complement to tablet and text modalities, MAAV also supports voice annotation. MAAV leverages Google Cloud's Dialogflow[8] to support users in dictating their annotations with a custom command: "OK, Google ... Tell MAAV [annotation text]." Their annotation text is then transcribed and logged to the database.

Voice annotations are especially useful when users are unable to type – such as when preparing dinner or cleaning their home. Participants can dictate their annotations in these conditions, either in response to a text notification or as stand-alone annotations. Similar to tablet annotations, physical proximity is also necessary to engage with this modality, which can limit its utility and use-cases. Also, similar to text message annotation, it is not possible to review $PM_{2.5}$ data when providing voice annotations.

4.3.3 Tablet visualization

MAAV uses an interactive tablet interface to allow users to engage with their $PM_{2.5}$ data and annotations (Figure 4.3). We designed the interface to default to a glanceable, always-on visualization that shows the most recent 24 hours of data in a line chart (Figure 4.3b). The visualization is also interactive and designed to support users in reviewing, exploring, and annotating the data. Participants can change the main view (Figure 4.3b) by:

- zooming in or out on the data at predefined intervals by tapping a button on the left (Figure 4.3d)

[8] https://cloud.google.com/dialogflow

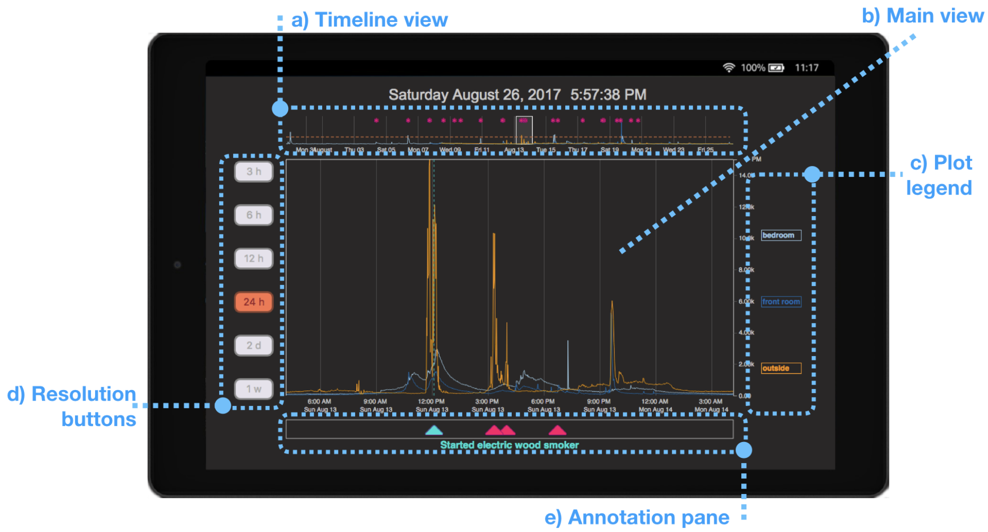


Figure 4.3. The tablet visualization interface. This image shows a number of indoor (blue) and outdoor (orange) spikes in the main view related to $PM_{2.5}$ measurements over a 24-hour time span. The annotation pane shows that the earliest and largest spike is annotated with the participant message: "Started electric wood smoker." These annotations lend context and legibility to the air quality dynamics shown on the interface. The timeline view shows that this is one of many such contextualizing annotations provided by the participant over their deployment.

- scrubbing across the top timeline view to look at data from the previous 30 days (Figure 4.3a)
- toggling the line for a monitor on or off by tapping that monitor's name in the legend (Figure 4.3c)

The visualization also shows annotations the user has provided for the data, represented by pink triangles underneath the main view (Figure 4.3e). The user can tap on a triangle to view the text of that annotation. The timeline view (Figure 4.3a) also shows pink asterisks at the top where there are annotations. As described in Section 4.3.2.1, the user can also long-press on the main view to provide a new annotation. Figure 4.3 illustrates the visualization's interactive components.

This interface was implemented on an Amazon Fire HD8 tablet for its low cost and customizability [122], using JavaScript and D3.js [48], and run in a kiosk browser[9] to pro-

[9]https://www.fully-kiosk.com/

vide a fixed, full-screen environment. We also incorporated Google Analytics to capture participants' interactions with the interface throughout their deployments.

4.4 Field deployments

Prior work shows that single-monitor indoor air quality systems with basic visualization capabilities improve participants' air quality awareness and increase their engagement [11]–[13]. Building on this, we sought to better understand how users would engage with MAAV's interactive visualization and rich annotation capabilities to support them in characterizing their home spaces. We deployed MAAV to six asthmatic families for a longitudinal study and conducted a series of interviews along the way to understand their experience hosting and interacting with this system. We label these individual household deployments D1 — D6 in accordance with study participants P1 — P6, introduced in Section 3.1, illustrated in Figure 3.2, and summarized in Table 3.1. By recruiting entire households, our goal was to observe ways the entire family's engagement in and awareness of their indoor air quality might change during the deployments. We staggered these deployments across Summer 2017 into Winter 2018 as participants were available and able to host MAAV.

4.4.1 Phased deployments

Participants received MAAV in a phased roll out to habituate them to hosting the system (Figure 4.4). These phases are delineated by participants' interaction with MAAV over the deployment: Phase I - passive deployment with no interaction (6-13 weeks), Phase II - active deployment with high interaction (2-6 weeks), and Phase III - long-term deployment with routine interaction (4 - 38 weeks), for a total operational time of 20-47 weeks (mean 37.7 weeks , Table 4.1). Each of the three phases began with user interviews to understand participants' air quality awareness (Phase I, predeployment interview), expectations for MAAV (Phase II, deployment interview), and experience with MAAV (Phase III, postdeployment interview). Although we encouraged all family members to participate, the majority of deployment families seemed disengaged and did not participate. These interviews consequently focused on the primary participants (Section 3.1). To better understand the (lack of) motivation for the remaining family members, we conducted a

Figure 4.4. The MAAV deployment timeline. Each phase corresponds to participants' level of interaction with the system: Phase I - no data provided (passive), Phase II - interactive feedback (active), Phase III - routine interaction (long-term deployment)

Table 4.1. Field deployment and interview lengths for individual households. Top: deployment D1—D6 durations by phase number, in weeks. Bottom: number and length of participant interviews for each deployment. Counts include predeployment, deployment, and postdeployment interviews with the primary participant, plus separate engagement interviews for both the primary and nonprimary participant(s), netting five interview sessions per deployment. Deployments D3 and D5 had an additional two interviews each from separate nonprimary engagement interviews with their teenage children.

Deployment Duration	D1	D2	D3	D4	D5	D6	**Avg.**
Phase I	6.3	7.3	7.1	9.7	12.7	5.9	**8.2**
Phase II	3.0	4.1	2.0	2.6	4.1	6.3	**3.7**
Phase III	37.9	35.7	34.9	34.7	4.0	7.9	**25.9**
Total Length (weeks)	47.1	47.1	44.0	47.0	20.9	20.0	**37.7**
Participant Interviews							**Total**
Number	5	5	7	5	7	5	**34**
Total Length (mins.)	139	192	160	245	230	258	**1224**

fourth round of engagement interviews. Here, we separately interviewed each parent per deployment, along with two additional interviews each with children from deployments D3 and D5. In all, we conducted 34 interviews with over 20 hours of interview audio (Table 4.1). All interview guides are available in Appendix A.

4.4.1.1 Phase I - Passive deployment (no interaction)

Phase I began with installing air quality monitors in participants' homes to habituate them to hosting the hardware. Each household received three air quality monitors and a small gateway computer for gathering, formatting, and packaging air quality measurements.

A trained research assistant placed the monitors following the parent study's established protocol to capture the majority of possible indoor activity: 1 outside, 1 in a common living space (kitchen or living room), and 1 in the asthmatic's bedroom. She placed indoor monitors between tabletop and head height to measure air quality within the breathing zone, approximately 30 - 60 inches, and away from vents, obstructions, or other environmental obstacles for accurate measurements. Outdoor monitors were mounted on the home exterior by the front door at shoulder height. These monitors were additionally placed inside a vented, non-temperature-controlled enclosure to permit sufficient airflow while keeping precipitation out. Outdoor monitors gave participants insights into the air quality around their home as well as providing a comparative data stream when understanding and assessing indoor air quality.

Once instrumented, we began collecting deployment-specific baseline data for use in the deployment interview. We disabled the monitor's on-board LCD display to not influence participants during this initial phase. Next we conducted the predeployment interview, with questions relating to the participants' awareness, understanding, and perception of air quality, both indoors and outdoors.

Phase I lasted 6-13 weeks (mean 8.2 weeks) and ended when we deployed the visualization and annotation modalities during the deployment interview. This deployment interview included an in-home demonstration of MAAV's interface, interactivity methods, and various annotation options. Participants saw the month of data that had been collected for the first time during the demonstration, which motivated a personalized explanation of their indoor air quality and exploration of the interactive tablet interface.

We also conducted a separate data literacy exercise to gauge participants' ability to interpret time series information. Each participant understood the data abstraction, and none had any difficulty using the tablet interface to answer questions about the data.

We conducted the interview in tandem with the system demonstration in order to capture the participants' reaction to seeing their data for the first time, along with their expectations of how they would use MAAV. When demonstrating the annotation modalities, we stressed that data annotation was completely optional and not a requirement for participating in the study. The Google Home and text message annotation modalities were also an optional deployment item; however, all participants elected to receive both modalities. We also purposefully did not provide any information on ways participants could improve their air quality to increase the likelihood that changes observed in subsequent interviews could be attributed to their use of the system.

4.4.1.2 Phase II - Active deployment (high interaction)

Phase II began after the deployment interview when participants were able to annotate and explore their air quality data. Phase II ran for a period of 2 - 6.3 weeks (average = 3.7 weeks), subject to participants' availability. During this phase, participants were free to use and interact with MAAV at their convenience. This deployment phase ended when we returned to conduct our postdeployment interview.

During the postdeployment interview, we focused on how participants used the system and on any insights or outcomes as a result of exploring their air quality data and making annotations. We asked additional questions for comparison against previous expectations relating to indoor air quality variability, system sensitivity, air quality awareness and engagement, and overall reflection on system utility. At the end of this interview, we gave each participant the option to return some or all of MAAV. All participants but P4 opted to keep their deployment hardware. Participant P4 chose to return their Google Home.

4.4.1.3 Phase III - Long-term deployment

Phase III began after the postdeployment interview, with the goal of leaving MAAV in place to capture the evolution of the participants' opinions and interactions after several months of familiarity. Phase III ran for a range of 4 - 38 weeks (average = 25.9 weeks). We conducted a follow-up engagement interview 4 - 35 weeks (average = 23.3 weeks) after the participants' postdeployment interview to investigate whether or not they were still using MAAV and how their use had evolved since they first received it. We also questioned the remaining nonprimary participants to gain insights into their perceptions of the system, its relevance to them, and mechanisms for their disengagement.

4.4.2 Qualitative data analysis

We transcribed and analyzed all interviews via case study analysis [123] and thematic analysis [124] to distill themes across the entire data set. This joint approach allowed us to identify and analyze within-interview-round themes between individual deployment phases using thematic analysis, and within-participant themes across all deployment phases using case study analysis.

We conducted the first round of qualitative analysis with data from participants P1-P4 and proceeded through each deployment phase. One researcher open-coded initial interviews to characterize participant replies, and two additional researchers then joined to collaboratively refine these codes, developing a set of closed codes. The original researcher then applied these closed codes to the remainder of the interviews in that phase and repeated this process for each phase. We met again at the end of coding each interview phase to iterate on codes and group them into higher level themes.

After completing our analyses with P1-P4, we enrolled P5 and P6 to validate our initial findings and verify thematic saturation. These additional participants received the same system through the same phases with identical interviews. This second round of qualitative analysis proceeded by coding P5 and P6 interviews using the closed codes identified in the first round, but with an eye to identifying anything new. No new codes were found, but the feedback and insights provided by these extra participants helped refine our findings. All resulting themes from this analysis are available in Appendix A.

4.5 Findings

Results from deploying MAAV highlight the benefits of a flexible multimonitor deployment and the diversity of tasks and questions it enabled participants to address. Findings in this section are deeply rooted in the particular context of each deployment and participant characterizations (see Table 3.1). Understanding participants' unique circumstances helps with interpreting their feedback and perspectives throughout this section.

4.5.1 Multimonitor deployments enable comparisons in and around the home

In addition to showing the benefits from deploying a single monitor [11]–[14], [99], MAAV supported participants in making comparisons between monitors. With three monitors, participants were able to observe and perform between-room and indoor/outdoor comparisons. These comparisons led to insights into home characteristics from observing $PM_{2.5}$ spike diffusivity (i.e., comparing two indoor monitors) and the ability to draw comparisons between indoor and outdoor monitors (Figure 4.5). Some participants found

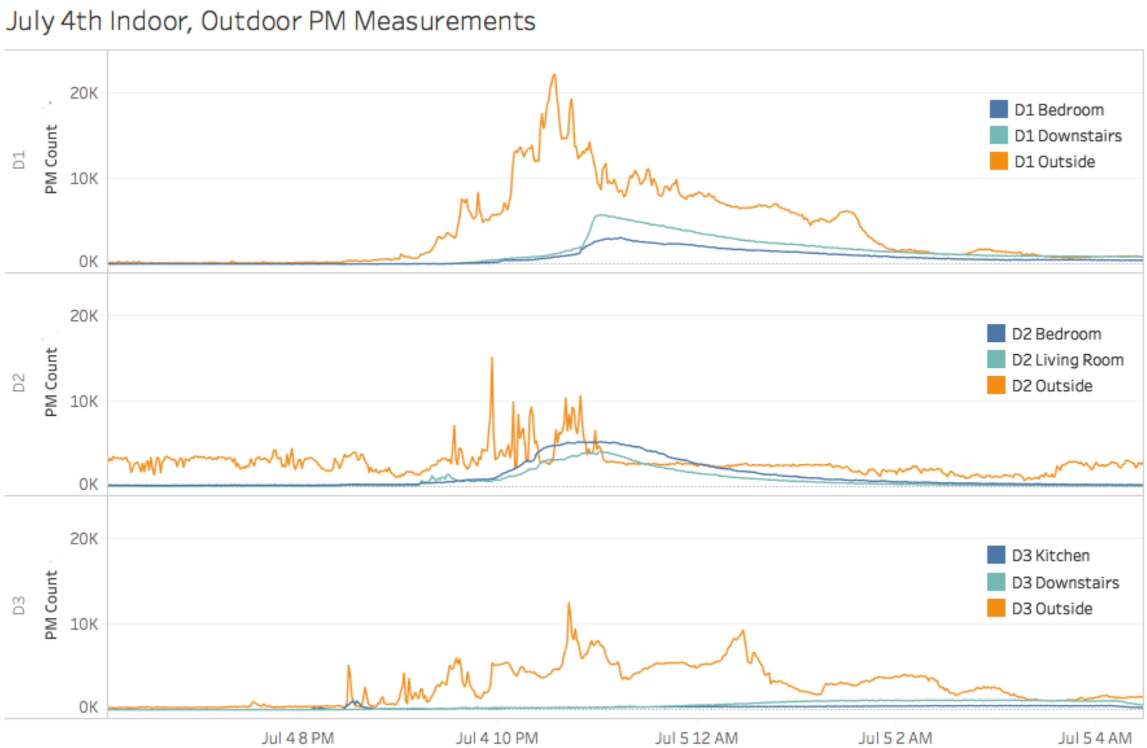


Figure 4.5. Impact of July 4th fireworks on outdoor and indoor air quality. Indoor $PM_{2.5}$ readings noticeably increase for deployments D1 and D2 whereas D3 shifts only slightly. In each case, it takes several hours for $PM_{2.5}$ readings to return to baseline levels. The readings from the outdoor monitors are different between deployments as well, highlighting the value of deploying monitors outdoors.

these comparisons so compelling that they asked if they could receive monitors for every room in their home. Similarly, when asked about their opinions on the adequacy of 3 monitors versus 1, participants unanimously preferred the flexibility of seeing multiple measurements at once.

> **P5:** It was nice to see there are differences, like what's going on in different rooms. If there's just one sensor for an entire house, it doesn't really tell you anything.

4.5.1.1 Between-room comparisons

When air quality changes within a home, it does not happen uniformly. Instead, it is subject to characteristics of the home itself, such as air exchange rates and airflow between rooms. During the deployment, participants used MAAV to reason about ways that the home itself might influence air quality. For example, P2 noticed differences in how $PM_{2.5}$ spikes diffused from one room to another:

> **P2:** It has surprised me to see that when [the living room] spikes, [the bedroom] tends to spike too. But when [the bedroom] has any activity [the living room] doesn't necessarily. I don't know if that's just the way the air flows in the house, or what.

Having multiple air quality monitors also enabled participants to detect air quality spikes that would otherwise be missed had there only been a single monitor.

> **P1:** I find it humorous that I can tell the days – when it's a Wednesday, my husband works from home and lights incense in the office downstairs, and there's a spike downstairs.

Likewise, P3 also drew connections between household activity and its impact on her family's indoor air quality.

> **P3:** I was definitely looking at the kitchen a lot because I think things change a lot more in the kitchen than the bedroom. Everything that happens in the house, happens in the kitchen. So it was always interesting when I saw a bedroom spike because ... what did they do? Seems really strange.

These situations where air quality events occur in localized spaces demonstrate microenvironments within the home, a phenomenon that can be observed only when there are at least two indoor monitors. Participants recognized these dynamics in their own living spaces, making observations and building a more nuanced understanding of how indoor air quality changes in the home.

> **P4a:** It's really interesting that we can compare [$PM_{2.5}$ spikes in different rooms] and see how some things that happen in different parts of the house can still affect other parts ... I think it's interesting that if there's big spikes outside, if you open the door – just a little bit – it affects the inside of the house. And it doesn't affect it just while the door's open, it kind of stays around for a while, because it's been let in. I think that's pretty interesting.

P4a's experience highlights the importance of multiple monitors. With MAAV, she was able to reflect on the source of an indoor spike by relating it to the air quality outside, something that would not have been apparent from a single-monitor deployment.

4.5.1.2 Indoor/outdoor comparisons

Past work indicates participants in single-monitor deployments have wanted to relate their indoor and outdoor air quality [11], and the primary method for doing so has been to pull data from third parties, such as local air quality agencies [13],[91],[125]. Unfortunately, these external sources are typically government-run sensors operating in fundamentally different ways, and providing $PM_{2.5}$ measurements at different time resolutions, spatial scales, and measurement units. It is therefore difficult to directly compare these third party measurements with a participant's own monitor. Moreover, many local activities (e.g., vehicular traffic, neighbors mowing their lawns, or barbecuing) do not show up at the municipal level, but they can directly influence air quality around a home. For example, Figure 4.5 shows very different outdoor readings across three different deployments during 4th of July fireworks.

MAAV's outdoor monitor lets participants directly compare indoor and outdoor measurements. They reported they were glad to have access to this information, and, in some cases, came to rely on these data in lieu of less representative public sources.

> **P6:** I wanna see what's happening with all three sensors. The outdoor is usually the one that stands out. Historically, that's where the higher [readings are].

Participants in this study all had some concerns about indoor air quality because someone in their home – either themselves or their child – had asthma. Before receiving MAAV, their only options for managing indoor air quality were to react quickly when symptoms manifest, guess about sources of poor air quality and manage them, or follow generalized medical advice. One common guideline for asthmatics recommends to stay indoors and keep windows closed, carrying the implicit assumption that the air quality indoors is generally better than outdoors. Despite this recommendation, past work has found indoor air quality can oftentimes be worse than outdoor [126].

Prior to receiving MAAV, participants had no way of verifying that indoor air quality was actually better at a given moment and whether the best course of action was to keep windows closed or air out their homes:

> **P2:** Sometimes you feel that people are like, oh, you should stay in because the [outdoor] air's bad. But, is it really that much worse than what's in my house?

After receiving MAAV, participants made use of the indoor and outdoor monitors to track fluctuations and trends between the two environments. MAAV empowered participants to evaluate the current situation by facilitating comparisons between indoor and outdoor $PM_{2.5}$ measures rather than depending on the generalized advice:

> **P3:** I have learned to pay a lot more attention to the difference between the air quality inside and the air quality outside. I think previously, I just kinda

> assumed that the air quality inside was always really great, because our windows are shut and our doors are shut and we don't open our windows ... and that's what I was told: just don't open your windows and use air conditioning and you're good. But I realized that's not really always the case. And that's been a big eye-opener for me.

Whereas P2 and P3 were relieved to find their indoor air quality was often better than the outdoors, P4 was relieved to see that their outdoor air quality was oftentimes just as good.

> **P4:** We open the windows all the time if the weather is nice, and I haven't noticed many spikes. That surprised me that there weren't that many.

P4's observation allowed her to relax from strictly following the generalized advice to always keep the windows closed. These findings indicate that having monitors both inside and outside of the home empowered participants to make informed, personalized decisions based on the unique environments in and around their homes. For many participants, the most compelling aspect of this comparison was the peace of mind in knowing that their indoor air quality was better than outdoors.

> **P3:** I loved that it showed that the inside didn't get really bad. That's one thing I was looking for – how was the outdoor air quality affecting indoor air quality.

> **P6:** The [greatest] value is knowing, "Hey, we're better than outside ... the air's better in here, guys", and checking that every day and having that reassurance. I don't know if I expected that that would be my best value, or my most favorite value.

Section 5.1 takeaway: *Participants made use of multiple monitors in their deployment to make comparisons between monitors, which supported new insights into the environment in and around their homes.*

4.5.2 Annotation improves user-experience, engagement

The ability to annotate in MAAV enabled participants to reflect and reason about the ways their behaviors impacted indoor air quality. Figure 4.6 illustrates participants' various use patterns and annotation modality preferences along with the evolution of those behaviors longitudinally over the deployments. All participants noted that annotation was essential for helping them recall past events, make connections, and find patterns.

> **P2**: I'm really glad those annotations are in there. It makes it easier to understand and put value on the data I'm seeing.

Figure 4.6. Long-term system interaction over phases II and III for deployments D1 - D6. This plot illustrates the daily distribution of system-initiated notifications (black), received text annotations (light blue), tablet annotations (blue green), and voice annotations (blue), along with participants' interface interactions (light purple). Interaction patterns for D2, D3, and D4 show varied cycles of engagement, and patterns for D3 and D6 exhibit re-engagement with the system immediately after the engagement interview. D3 also shows patterns of making voice annotations when experimentally determining spike sources. All participants came to rely on test messaging for their primary annotation mechanism as the study progressed. Additional deployment data is available in Appendix A.

For P3, annotations were necessary for engaging with the data.

> **P3:** I don't think this [visualization] would do me much good without annotations . . . You have to have the combination of [data and annotation]; otherwise [the visualization] is kind of useless.

P4a used her annotations as a common means to get into and explore the data.

> **P4a:** I look at annotations a lot. I like being able to see it. What I look at them for, I pay attention for if there's repeating things [patterns]. Today we played with the dog, and last week we played with the dog, but the spike is bigger.

Tolmie and Crabtree [20] found that raw sensor data are difficult to interpret without contextualizing annotations. By tightly integrating tablet, text, and voice annotation modalities into our system, participants could contextualize their data with personally meaningful annotations in whichever way suited their interaction preferences. Early versions of *inAir* [11], [12] provided a website for reviewing past data and logging activities via a calendar interface, although this manual form of journaling was not overlaid on the $PM_{2.5}$ data. Despite the effort to contextualize raw sensor data, participants reported feelings of powerlessness and frustration over a lack of context for the current data they were shown [11]. By tightly integrating annotation methods and interactive visualization, participants were able to use MAAV to alleviate this shortcoming and make annotations available for review as they were captured.

Annotations were facilitated through *user-initiated* and *system-initiated* annotations. User-initiated annotations relied on action from the user to engage, such as with the tablet interface or voice annotation, whereas system-initiated annotations were directly generated by MAAV to prompt participants for action.

4.5.2.1 User-initiated annotations: engaging through tablet and voice

Participants recorded *user-initiated annotations* through the tablet interface and voice modalities, often reviewing their data on the interface to explore the detected event, verify their annotation, and explore previous activities captured by past annotations. The combination of increased engagement and improved data legibility led participants to reason on the cause-effect relationships between their behaviors and the air quality in their homes.

During the deployment interview, P3 speculated about the potential for MAAV to support self-experimentation:

> **P3:** It would be nice to see if there's changes depending on when I have to use our giant oven versus the little convection oven and what difference that makes. I think I would definitely annotate those things as I start to make dinners.

Afterwards, P3 used MAAV to methodically test cooking behaviors and deduce sources of cooking spikes (Figure 4.2). When we returned for the postdeployment interview, P3 shared her process for annotating her cooking activities to determine which decisions influenced spikes – the combination of real-time sensing, longitudinal visualization, and annotation across multiple monitors enabled her to perform classic A/B experimental design and keep track of her test cases:

> **P3:** If my theory is that it went really high that day because I used the stove top and used something that made a lot of smoke, then the next time I'm gonna do that, I'd start that and [annotate] "I'm doing this again," and then see what happened.
>
> I figured out I get kitchen spikes when I cook with olive oil instead of avocado oil ... I don't use olive oil any more. I haven't had nearly as many kitchen spikes since then. Kinda cool.

P3's annotation led her to conclude that cooking with olive oil contributed to kitchen spikes and that using avocado oil eliminated these spikes. This is another case where participants performed comparisons and reached conclusions that would not have been possible without the use of integrated annotation and visualization capabilities with the sensing system. Of further note, P3 used voice annotations for self-experimentation, which allowed her to continue cooking without taking the time to type an annotation.

4.5.2.2 System-initiated annotations: push notifications engage participants and improve awareness in real-time

One important feature for facilitating annotation in MAAV was the inclusion of *system-initiated notifications* via text alerts. These alerts actively engaged the users, prompting them to annotate the spike with a returned text message. Motivated by work on *just in time interventions* [127], the prompt system notifies participants as spikes occur in order to maximize the likelihood that they will be able to accurately annotate the data. Whereas all participants used more than one annotation modality during their deployment, text message annotation in response to notifications was the most common and consistent mode across participants (Figure 4.6).

> **P2:** The thing I found most helpful were the texts. I *love* that it's "hey, we had a spike" and I'm just like "this is what it is." This is what I'm doing right now, or in the last 10 minutes. That's my favorite feature. I love being able to see how much and what is causing spikes.

Beyond presenting an opportunity to contextualize air quality data, text alerts also became proxies for other indicators. In postdeployment and engagement interviews, P4 and P4a talked about using MAAV's sensitivity to infer others' daily schedules, such as leaving for work or getting home from school.

> **P4a:** I'll wake up to a text: "at this time in the morning there was a spike ... " Oh, that's when my parents went to work! Or when I get home from school, the rooms will spike.

> **P4:** I have noticed a small pattern ... It's a little more common to get spikes around 3:30pm.

> **P4a:** That's when I come home, drop all my stuff, and get food. I actually dance more until parents get home. I don't dance when my parents are home.

4.5.2.3 Annotation modality influenced by review routine

Participants' personal schedules, asthmatic sensitivities, and goals helped influence the way they interacted with and annotated their air quality data (Table 3.1). Figure 4.6 and Table 4.2 show how participants used MAAV over time. P1's outdoor sensitivities motivated her to regularly check the tablet interface to keep informed of her immediate outdoor air quality, which, combined with her habit of reviewing her indoor air quality data at night, found her logging half of her annotations through the tablet interface over phase II. P2 adopted a weekly interface review regimen, leading her to be more reliant on text notifications to gauge indoor air quality. Therefore, P2 relied on the receipt of text notifications to know when her air quality became worse. Consequently, she annotated much more frequently using the texting modality. P3 engaged in self-experimentation to understand the impact of alternative behaviors on her in-home air quality, causing her to rely on voice annotation for annotating her self-experiments.

Deployment D4 was slightly unique from others in the study in that the primary participant (P4) was considerably less engaged with MAAV than other primaries in their deployments. This disengagement stemmed from two main factors. First, P4's job outside the home meant she was frequently away from the house when receiving text notifications. Her solution to this was to text P4a to determine what was going on at home. Second,

Table 4.2. Deployment annotations per modality over phases II and III. The annotation column is the sum of all annotation modalities. Participant annotations include those made in response to MAAV notifications as well as user-initiated annotations, exceeding 100% compliance rates for many deployments. Appendix A contains more deployment data.

	Phase II – Compliance						*Phase III – Compliance*					
	Text	Tablet	Voice	Anno.	Notif.	(%)	Text	Tablet	Voice	Anno.	Notif.	(%)
D1	4	4	1	9	8	112.5%	47	9	0	56	110	50.9%
D2	16	5	3	24	18	133.3%	81	7	3	91	103	88.3%
D3	5	4	12	21	5	420.0%	81	8	20	109	98	111.2%
D4	7	5	2	14	14	100.0%	61	2	1	64	382	16.8%
D5	19	9	4	32	19	168.4%	8	11	0	19	7	271.4%
D6	31	1	1	33	36	91.7%	24	0	0	24	33	72.7%

the tablet interface was set up in P4a's room, meaning P4 had comparatively less access and opportunity to review the data than the primaries in the other deployments. Deployment D4 experienced an equal number of text and tablet annotations within phase II, but as the deployment progressed, they came to rely on text messaging as the primary mode of annotation.

Both P5 and P6 adopted an after-work routine to review their data and annotations out of curiosity and were equally engaged with MAAV, primarily relying on text messaging to maintain high compliance rates with system notifications.

Section 5.2 takeaway: *Participants used different annotation modalities to annotate their data in different situations that seem to be tied to the affordances of that modality, indicating that the availability of multiple modalities led to more annotations.*

4.5.3 The interactive tablet interface and annotations enabled data exploration

To address prior findings that users wanted a larger snapshot of time when reviewing their air quality information [12], the tablet interface provides a one-month backlog of data to explore. Within phase II, all participants frequently engaged with their tablet interface for reviewing and exploring their air quality data (Figure 4.6). Participants appreciated the ability to scroll back to review their annotations and prior $PM_{2.5}$ spikes, although most primarily used the interface to explore the previous 1-2 days.

> **P2**: I'm more interested in what is happening right now, the last couple of hours, how it's affected (or might affect) what's going on today or tomorrow. I

> liked the idea of being able to go back a month or more, at first. I'm just finding I'm not using that.

> **P5:** I was more just [looking at] current data, like the last day or two.

The ability to look back farther also motivated some to explore the full extent of their data.

> **P4a:** I'd go back as far as it would let me. Most of the time I probably go 1-2 weeks back. Depends if I've looked at the spikes before ... Most of the time I'll just look over the past week. Especially the day that I'm looking at it and the day before, 'cause those ones are the most ... newer spikes that I haven't seen yet. Sometimes I'll go back and "what was I doing." Oh yeah, and I'll add an annotation.

Several participants reported not being interested in the value of the readings themselves and that seeing the relative dynamics was enough, echoing findings from design in eco-feedback [115]. Instead, they primarily focused their attention on the presence and distribution of spikes, along with their previous annotations.

> **P1:** I'm definitely more of a visual [person], so seeing the spikes, to me, is enough of an [indicator].

> **P4a:** Most of the time I'll just look at new spikes, but then I'll look at the little bar at top and "hey! there's a whole bunch of [annotations] here. I don't remember what this was" and I'll go back and look through it.

When reasoning about $PM_{2.5}$ spikes, some participants expressed interest in discovering and comparing patterns that MAAV's visualization does not make salient.

> **P1:** It would be kind of cool to see, "You've done these last 5 things cooking dinner. Here's how they overlap and compare." That would be an interesting thing.

This desire to view and compare groups of activities over time suggests that realizing the value of a longer data history might require alternative visualization and processing techniques beyond the pervasive time series view.

Section 5.3 takeaway: *Participants used the tablet interface to review their annotations and air quality data, drawing insights that would not have been possible without annotated data.*

4.5.4 Participants maintained long-term engagement but changed this use over time

Many facets of user engagement and system integration cannot be fully observed or validated over a short-term deployment. For air quality, long-term effects of the changing seasons can have a significant impact on the nature of data shown to participants and,

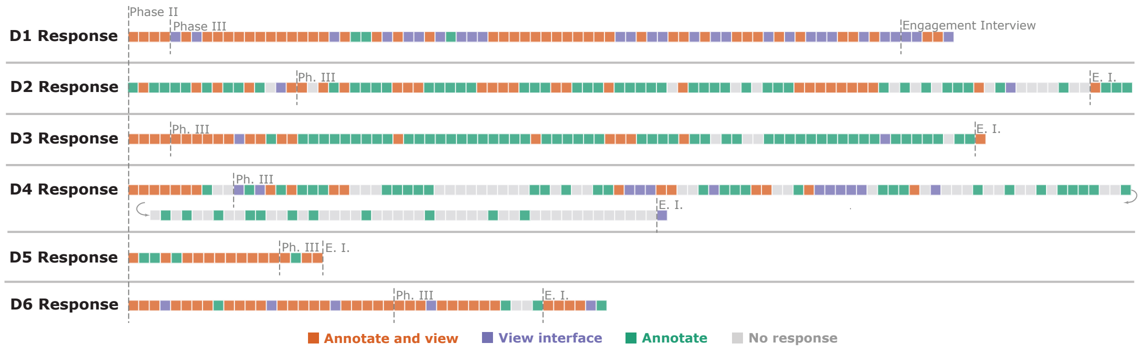


Figure 4.7. Engagement chart for deployments D1 - D6. Boxes represent individual days when a deployment received at least one notification from MAAV, indicating significant air quality changes in the home. We focus on these days as we do not expect that participants would necessarily interact with MAAV on days without notifications. Each box is color-coded according to the deployment's response to the notification on that day: both viewing the interface and annotating (orange), only viewing the interface (purple), only making an annotation (green), or not responding that day (gray). The sparsity of gray boxes illustrates participants' willingness to engage when we should expect it, serving as a proxy for their continued engagement. Outside of D4, which received significantly more notifications than other deployments, the remaining homes all had engagement rates above 80% (Table 4.3).

correspondingly, can affect their sense of engagement. In this context, our long-term deployment provides a rigorous probe into the notion of disengagement by avoiding short-term effects or confounds. Figure 4.7 and Table 4.3 provide evidence for deployments' continued involvement on the days where they received at least one text prompt, showing an 80% or greater compliance rate for 5 of 6 deployments. We did not track days without system notifications as these indicate days with few or no significant changes to indoor air quality in participants' homes. Consequently, we would not necessarily expect participants to interact with the system under these circumstances. By focusing on notification days, we can track the methods of participants' responses, which serve as a reasonable proxy for user engagement.

Although participants were primarily driven by curiosity to explore their data when first receiving the interface, their motivations evolved as deployments wore on. Mechanisms for this shift included an increase in familiarity with the data and a corresponding lack of surprises, but also evolving goals and seasonal impacts, as well as participants changing the questions they sought to answer.

Table 4.3. Engagement data from Figure 4.7. 'Notification Days' lists the number of boxes for each deployment and its corresponding percentage of deployment time. Participant response is further binned by 'Days Ignored' (gray boxes), and 'Days Engaged' (nongray boxes). The final column is the percentage of notification days receiving a deployment response. Total engagement with MAAV was above 80% for 5 of 6 total deployments for the entire study.

	Phase II & III Length (days)	Notification Days	% Days Notified	Days Ignored	Days Engaged	% Notification Days Engaged
D1	265	78	29.4%	0	78	100.0%
D2	250	95	36.0%	15	80	84.2%
D3	244	81	33.2%	4	77	95.1%
D4	243	144	59.3%	70	74	51.4%
D5	28	18	64.3%	0	18	100.0%
D6	55	45	81.8%	2	43	95.6%

> **P2:** I used it more in the winter when we have such bad air quality [...] I've not been using it as much [recently], but mostly because as the seasons have changed we're not having that terrible air anymore.

Participants also began to engage the data with more specific questions as their experiences grew, which our technology probe was not well suited to answer.

> **P2:** If it were to be something that I could have a more personalized overview that I could take to my doctor, I would probably use it more and be more interested in keeping track of a daily "how I'm feeling" [...] To be able to connect health with the data is important.

It is important to note that changes in participant use patterns were not always a result of disinterest. P3's interest in self-experimentation lent itself to periodic engagement, visible in her deployment D3 interaction data (Figure 4.6).

> **P3:** Once I figured some things out ... the kitchen thing ... I had a lot of fun trying to figure out why it would go off at various times in the kitchen. So I did a lot of experimenting early on, but once I had my questions answered it was good.

Regardless of these variations, participants' text response rate was largely unchanged throughout the deployment and had become the preferred annotation method for most participants by the time of the engagement interview (Figure 4.6).

> **P1:** Honestly ... I didn't think the text messages would be as easy to respond to ... It's a matter of, if it texts me right away, "hey this happened at this." I can either go "Oh yeah, I know what that is", annotate, done. That ease of

> [answering] even if I'm outside or [away from home] . . . honestly, that's what's changed how much I've annotated, is just that ease of the text message.

Similarly, for P5:

> **P5:** Phone texting was much more immediate and much more responsive. [texts] would draw out a response immediately that we would look into.

Not everyone remained engaged with text messaging, however. Both P4 and P4a reduced their annotations, but for different reasons.

> **P4:** It just got tedious. I figured out that everything in the house makes this room spike. It doesn't matter what we're doing. But I just got tired of repeating the same thing over and over again.

> **P4a:** When I first started [annotating], it was because I could annotate everything. I would still probably be doing that. I thought of it as "this is cool – this will help me realize what causes me to stop breathing as much". That was really exciting. I stopped annotating as much because [the tablet died].

Due to an uncommunicated hardware error, P4a was left without a functioning tablet at the end of their long-term deployment phase. P4a's feedback on how the ability to review their data motivated annotation points to a synergy between annotating and reviewing data not found in the extant literature.

For those who continued annotating, their messages became shorter as they defaulted to a subset of personally relevant keywords to describe tasks that set off text alerts.

> **P1:** I would say it's much more succinct. "Doing laundry," "cooked dinner," instead of being, "I cooked dinner and there was burnt pizza on the bottom and there was smoke and blah blah blah." They're kind of more "oh yeah" reminders than "oh 'this is this . . . ' " and I need to explain this whole situation. It's definitely condensed down to keywords.

Fang et al. [14] have developed an air quality sensing platform for detecting and classifying three common sources of indoor pollution: cooking, smoking, and spraying pesticide. Although we found a moderate amount of overlap in participants' keywords, we did not have instances of smoking or pesticide use. The remaining annotated activities, and participants' convergence to a set of abbreviated keywords, point to a more diverse set of in-home activities capable of generating $PM_{2.5}$ spikes. Moreover, arriving at these personalized, ecologically valid keywords takes time and would not necessarily be achieved within the shorter term deployments in past work.

4.5.4.1 Lack of family engagement

Maintaining deployments over many months allowed us to observe usage patterns and characteristics that would be challenging to discern over shorter term installations. One observation was how each deployment contained at most one engaged participant from the onset of the study.

When discussing their own senses of engagement, primary participants indicated a division of labor or a sense of responsibility very early in the deployment and postdeployment interviews.

> **P3:** Anyone in my family? Nooooooo. The reason is because I'm the one that takes care of the stuff. If they don't feel responsibility to take care of it, then they don't care.

This sentiment was shared by spouses across our interviews:

> **P6 Spouse:** I don't really get involved – it's his thing. He shares everything anyway, so I hear about it that way [...] my schedule is already pretty busy and this is just another thing.

> **P4 Spouse:** Yeah, [poor air quality] sucks for some people, but it doesn't affect me.

We were able to rule out insufficient communication or interview scheduling as engagement barriers over time, and conducted a dedicated fourth round engagement interview with the goal of understanding why certain participants were engaged and others were not.

When discussing engagement with nonasthmatic participants, we asked whether they could imagine circumstances where their level of engagement may have been different. Each cited their lack of asthma as a barrier. We also observed that age influenced a participant's willingness to engage; other than P4a, no other teenage or younger participant was aware of, or concerned with, air quality issues.

Engagement interviews also identified inconvenience and unintentional gate-keeping as other factors affecting people's ability to engage with the interface. By coordinating interviews with a single family contact, in some deployments it came to be seen as "*their thing*," or "*the study's*," which limited others' sense of responsibility.

> **P5:** Putting it over there so that my family didn't break it probably didn't help them to be more interactive with it.

Family dynamics are complex and vary between households. It is possible that with targeted effort we could have seen more engagement by multiple family members, but their reactions here suggest that, left to their own devices, they may also naturally settle into a routine where only one person is primarily concerned with their air quality monitoring system.

4.5.4.2 Participants converged to a single interaction, annotation modality over time

As tablet use decreased, participants relied more on text alerts to annotate and gauge their indoor air quality:

> **P6:** If it sucks enough for you to message me, that's going to indicate that something is up.

When asked what motivated them to continue annotating, primary participants cited text messaging's low barrier to responding and the importance of contextualizing their data. Some also mentioned a hope of receiving more accurate medical recommendations or individually tailored air quality advice as a future study benefit. Mobile interaction also proved popular with participants. Although earlier postdeployment interview feedback indicated that participants enjoyed having a dedicated tablet for the visualization, each primary participant – and even some nonprimary participants – identified that a mobile app would be more useful and would improve their involvement by the time of their engagement interviews.

> **P1 Spouse:** If it had been an app rather than in one place in the house, I could see being a lot more involved. Have the sensors running here [home], then I can pull it up at work, and see how things are at the house. It would become a thing like checking weather, I think. Before work, I check the weather on my phone. I check the air quality around the house just the same.

> **P6:** If I had an app that would show me the exact same thing, I would use that app a ton.

Participants' convergence toward text notifications as their primary annotation and interaction modality lends strong support for the power of system-initiated prompts delivered with a mobile device for long-term engagement (Table 4.2).

Section 5.4 takeaway: *Participants changed how they engaged with MAAV over time, transitioning from an initial phase of deeper regular engagement to a maintenance mode, whereas their family members remained mostly disengaged throughout.*

4.6 Discussion

In our long-term deployment of MAAV, participants exhibited several behaviors not observed, or possible, in previous work. These behaviors included drawing insights from multimonitor comparisons, leveraging annotations to facilitate both awareness and sense-making, and changing – while still sustaining – their use of the system over time. These results have several important implications. First, annotations have the potential to facilitate new types of comparisons in users' data. Second, the surprising long-term adherence to our mobile phone push notifications offers a promising direction for data labeling and collection in future studies, and perhaps in different contexts. Third, participants' shifting usage of MAAV over time suggests opportunities for a user interface that supports different stages of system use. We discuss each of these implications in detail.

4.6.1 Beyond direct temporal comparison

Participants in this study were able to view synchronized and annotated data from multiple monitors and use that representation to draw insights that would not have been possible without between-monitor comparisons and contextualizing annotations. Integrating annotation into MAAV supported participants in making the data legible [20]. When reviewing their data, participants were less interested in scales, magnitudes, or baseline fluctuations communicated by the time series representation, but instead focused on the locations and distributions of $PM_{2.5}$ spikes. Reviewing annotated data, complete with personal activities and routines, allows participants to go beyond the numerical representation and reason about these trends and impacts at a higher level than would be possible without the context afforded by annotated data.

Over time, participants wanted to go beyond the comparison capabilities supported by MAAV. Through our interviews, we observed participants thinking across time and wanting to compare relationships between temporally disjoint events, e.g., compare the five most recent times I vacuumed. The fact that participants requested this type of feature demonstrates the value of annotations and of collecting long-term data logs. Supporting such user-driven *cuts* [17] requires re-thinking annotation logs. Currently, our annotation mechanism tags single points in time, although the tracked activities may describe events taking place over a duration. At a conceptual level, this type of comparison interface is straightforward to describe, but the implementation details much less so. Developing methods, for example, on how best to define, capture, and align a collection of $PM_{2.5}$ events, and the corresponding semantic rules or computational metrics for comparing event similarity is much more involved. Although neither MAAV nor other interfaces in the air quality literature currently support these temporal cuts, future systems should consider event tracking and processing techniques for supporting comparisons across time, place, and activity.

4.6.2 System-initiated notifications sustained engagement

From the perspective of sustained participant engagement, much of the success of MAAV seems to come from the system-initiated text message prompts. Figure 4.7 shows the majority of participants remained engaged with MAAV throughout their deployment, with majority engagement rates over 80% on days they received notifications. During interviews, participants indicated that the text messages were useful as a proxy for gauging indoor air quality and that the messages were lightweight and easy to respond to. Yet for P4, who received comparably more notifications than other deployments, these messages led to frustration over their repetition (Table 4.2). As an outlier, P4's experience points to a possible upper bound for user notifications, above which participant engagement may decline or drop off entirely (Figure 4.7). Based on P4's notification-fatigue, it may be preferable to allow participants to selectively mute specific indoor monitors or have control over their system's thresholding and rate-limiting parameters. Another possibility is to analyze spikes in the context of past spikes and user engagement to dynamically adjust which notifications to deliver. For example, multiple spikes that seem to be similar might be bundled together, or notifications might be reserved for spikes that appear to be sufficiently different from past spikes.

System notifications make people aware of changes in their environment in a way that is easy to miss with an ambient display. The next-longest longitudinal indoor air quality study [13] utilized ambient displays, requiring those participants to be in front of the monitor to witness spike activity, which even then showed only the most recent four hours of data. Thus, these results are new and novel for the way that they directly elicit and probe user engagement.

As an annotation prompt technique, text messaging was successful at keeping participants engaged with MAAV and thinking about air quality without requiring them to proactively and regularly check the monitor. The fact that our participants sustained their engagement, measured by response rate to text message prompts, over a long period of time suggests this level of engagement may be a reasonable steady state for an air quality monitoring system. Eventually, with enough labeled data, it might become possible to train a machine learning model to identify the highest frequency data labels, which could support labeling household-specific events, rather than focusing on a closed label set, e.g., Fang et al.'s classifier for cooking, smoking, and pesticide events [14].

4.6.3 Evolving system usage during long-term deployments

Whereas prior work focuses on the benefits of deploying residential air quality monitor systems and its impact on air quality [11]–[14], this field deployment focuses on how users engage with an air quality sensing system and how this engagement changes over time. It is notable that MAAV's tablet interface seems to have served participants well through their trajectory of system usage, starting at a high engagement level during an initial period of discovery and sensemaking, and then transitioning into a steady-state maintenance mode of sustained lower activity. Although the current interface supported this transition in some ways, future systems could design for these stages more directly. Toward the beginning, the system could take advantage of high engagement to learn more about the user's interests and relevant lifestyle factors. As data begin to accumulate, the system could begin to suggest comparisons, correlations, or views on the data. As the user transitions into a maintenance mode, the system could similarly scale back to a lighter weight visualization. It could even streamline annotation, such as asking the participant to confirm predicted annotations, rather than typing them out from scratch. This notion of an evolving interface could be more generally useful in other human-in-the-loop data collection domains, including eco-feedback and personal informatics.

Interpreting air quality monitoring within a personal informatics framework is compelling because we did not initially consider environmental measurements within this domain. Air quality, even indoors, is often perceived as a characteristic of an area. Yet, air quality's influence is clearly relevant to personal informatics when relating it to personal health impacts. Findings in this work also show $PM_{2.5}$ to be an unexpected proxy for human activity. Reflecting further on participants' evolving engagement, we began to consider this work in the context of personal informatics models [5], [28], with the initial high engagement similar to the *discovery* phase in Li's stage-based model for personal informatics systems [5]. Participants also exhibited changes in their engagement related to increased familiarity, seasonality, and new questions, leading to periodic interaction with the tablet (P2, P3, P4a, P6) and voice annotation modality (P3). Participants' periodic engagement with MAAV maps neatly to the lapsing and resuming cycles described in Epstein's lived informatics model [28]. Shorter term deployments involving P5 and P6 exhibited similar behavior, notably with occasional tablet use (P5) and re-engagement with the tablet interface (P6) immediately after the engagement interview.

The diversity in $PM_{2.5}$ activity between deployments offers opportunities to reflect on a residence's impact on indoor air quality and engagement. The combination of P4's low participant engagement and burden of high text notifications may account for their comparably low notification compliance, but we have no insight into mechanisms behind their increased $PM_{2.5}$ activity. Home permeability, furnace efficiency, and behavioral patterns can all contribute to better understanding these causes, but these factors were outside the scope of this work. Extrapolating or incorporating direct measurement of any of these factors could help account for what seemed an overly sensitive system for P4. Conversely, the most compliant participants (P3, P5) were those most curious to find out what was going on in their homes.

This innate curiosity, along with co-enrollment in an ongoing clinical study, helped sustain participants' engagement. Moving forward, the findings and observations from this field deployment will help us to design systems and interfaces to foster engagement and curiosity, enabling users to pose and answer their own questions through more sophisticated exploration, comparison, and hypothesis generation techniques. Implementing these features in an adaptive system would provide more relevant and actionable information to end-users.

4.7 Limitations

The findings from this field deployment should be interpreted in the context of specific limitations and study design decisions. Recognizing that multiple factors affect and influence indoor air quality, it is important to acknowledge that any self-labeled annotations and reasoning about $PM_{2.5}$ spikes represent a participant's best guess, and are not necessarily indicative of the reported source.

The air quality monitor's design – in particular the monitor's internal air quality sensor – was a source of some initial confusion. These sensors measure airborne particle concentration, and say nothing about chemical composition or its health impact. Therefore, an increase in $PM_{2.5}$ particle counts does not necessarily correlate to an unhealthy source. For example, humidity can create significant spikes. This distinction became clear, however, as participants recognized the monitor's responsiveness to a range of activities.

Our participants were concurrently enrolled in an ongoing medical study, which has both benefits and drawbacks. Advantageously, we engaged a highly motivated population already familiar with data collection and who were interested in participating. However, their involvement in a clinical study came with a number of caveats regarding the information we could disclose to avoid introducing confounds. Although clinicians are interested in participants self-labeling and contextualizing data, our findings show participants who engage with and annotate their own air quality data are also inclined to connect air quality to their own behaviors, reflect on that relationship, and take action to reduce their $PM_{2.5}$ exposure. To mitigate these broader concerns over how our interviews may influence participants and potentially confound the parent study's clinical findings, our interview questions were vetted by practitioners in the parent project to ensure we were not encouraging or influencing participants to change their behavior.

Further, our six household field deployment prevents us from generalizing our findings to a larger population, as does the asthmatic status of the participating families. This status also influences baseline levels of engagement, as asthma severity has been shown to be positively correlated with engagement of a technology probe [128].

We also encountered some technology limitations. Our voice transcription skill went offline from automated firmware updates early in phase II, preventing some participants from using this feature. This outage caused some participants to lose faith in this annotation modality. Furthermore, two participants experienced difficulty with Google Home recognizing and parsing their voice commands, further reducing its appeal. Lastly, staggered deployment scheduling resulted in deployments D5 and D6 starting later than the other participants, leading to a shorter overall deployment length.

4.8 Conclusions

Air quality and $PM_{2.5}$ exposure are important and often invisible aspects of personal health. We deployed MAAV, a multimonitor air quality measurement, annotation, and visualization system, in a longitudinal field study with six asthmatic families to understand how this combination of features could support participants in managing their air quality at home. The results of the deployment indicate that MAAV supported participants in gaining insights by comparing multiple monitor streams, recording and exploring annotations for contextualizing their air quality data, and remaining engaged over their long-term deployments by relying on low-burden text message alerts.

Following these deployments, we had planned to redesign an improved visual analytics interface for supporting the kinds of questions participants could not use their current interface to answer. Although participants' interviews provided feedback on what they imagined changing with their system and interface, they had not captured deeper insights on how they imagined using their data to answer their questions. Lacking this clear picture of participants' needs, abilities, and workflows, we were unable to develop a workable design to support them to flexibly analyze their air quality data. This next chapter describes a new interview method we created to help bridge this gap in our understanding.

CHAPTER 5

AN INTERVIEW METHOD FOR ENGAGING PERSONAL DATA

This chapter presents a novel interview method we call the *data engagement interview* [4]. We created the data engagement interview method to fill a gap in existing interview techniques that prevented researchers from directly observing how people engage with personal data. This method enables researchers to closely study the ways interview participants engage with their personal data for understanding their needs, abilities, and workflows. This method also strikes a balance with incorporating data in a design process. We situate data engagement interviews as a valuable middle ground approach between easily deployed methods – like data sketching [129] or physicalizations [72] – that do not scale to complexity of personal data, and the more time-intensive and design-heavy efforts behind custom tool development.

Data engagement interviews achieve this functionality by offloading data analysis responsibilities from the interview participant. Without the cognitive burden of wrangling or programming personal data on their own, participants are able to focus on applying their data to help answer a personal questions. Interviewers guide this process to observe participants' workflows and collect feedback through targeted engagement prompts that unfold as a natural conversation with the participant. This chapter outlines the data engagement interviews' development (Section 5.3) and includes a case study from performing this method with each of our study participants (Section 5.5). Our findings indicate data engagement interviews can be an important tool for helping researchers to engage participants with personal data and a valuable method for exploring the personal informatics analysis gap.

5.1 Introduction

Observing how people engage with their personal data offers a wealth of insights for researchers and practitioners. For example, understanding and identifying the kinds of questions people ask of their data, and the analysis strategies they employ to answer them, helps with designing new tools [130], [131]. Creating opportunities for people to learn new things from their personal data can also provide triggers for positive behavior changes [132], and showing participants the value of their personal data can help motivate continued self-tracking [25].

As the scope and scale of personal data increases — through improved sensor resolution and integrating multiple data sources — engaging with data increasingly requires the use of sophisticated analysis tools and methods. Lightweight approaches, such as sketching [133] or data physicalizations [72], can be quick to perform and require minimal design effort. These approaches, however, often involve abstract or incomplete data and do not scale for direct engagement with the complexity of many real-world self-tracked data sets. Examining raw data with these approaches may work for exploring a small amount of data at a time [20], but can quickly break down with larger data sets. These larger data sets generally require some amount of computation to support exploration and analysis, leading many personal informatics researchers to develop and deploy custom analysis tools in order to engage participants with their data. This heavyweight approach, however, requires significant design work and interpretation of what people might *actually* do from what they *say* they want to do, potentially leading to gaps in analysis support [17].

We propose a middle ground approach in this chapter that we call the *data engagement interview* [4]. The data engagement interview is a research method that sits between the lighter weight approaches involving minimal design effort, and the more heavyweight approaches that involve customized tool development. We developed data engagement interviews to help researchers better understand and identify what participants want from their personal data by observing participants ask and answer questions in real-time from their own data. This interview method includes a dedicated data analyst on the interview team to provide participants with a flexible toolbox of real-time analysis techniques. Using this method, interviewers can support participants as they explore their data to elicit and observe more authentic data engagements, while the data analyst takes direction on how to process or present participants' data to answer personal questions. Whereas data engagement interviews are more resource-intensive than other standard interview methods, this method strikes a balance between engagement strategies that fail to incorporate complex personal data and those requiring customized tool development prior to collecting any observations. This interview method can quickly help researchers with eliciting design requirements for potential future system development, while also helping participants use their data to flexibly answer unique and personal questions.

We developed the data engagement interview from our own research goals to design new visual analysis tools for asthmatic families living with indoor air quality sensors [3]. Through sensor deployments with six households, we collected various data sets for each family that included several months of quantitative and qualitative data, sampled over different timescales and measurement intervals, that require both personal annotations and contextual knowledge to productively interpret and analyze. These computational and contextual demands prevented us from using lightweight engagement methods. After developing the data engagement interview method, we conducted interviews with our participating families to observe how and why they engage with their personal indoor air quality data. In addition to extracting design requirements for a future analysis tool, our analysis of the interview transcripts showed that data engagement interviews can also yield a host of other insights and opportunities.

In this chapter, we present a framework for conducting data engagement interviews that allows researchers and practitioners to engage participants directly with their personal data without the need to develop custom data analysis tools. We also conduct a case study in which we apply the interview framework to characterize the motivations and analysis tasks of asthmatic families when working with personal air quality information. We observed evidence that this method can expose differences between what participants say they want to do with their data, and what they actually do; engage participants more readily than standard interview methods; teach participants new things about their data; teach researchers new things about design requirements; and benefit research outcomes by improving insights on study design and motivating participants to self-track. To support transferability, we have prepared an online guide[1] [134] that includes a sample interview protocol based on our experience of conducting data engagement interviews, along with other detailed suggestions, interview materials, and example data and processing scripts to help make this interview method more actionable for researchers and practitioners.

Our analysis of participants' data engagement interviews lends evidence that this method can be a promising approach for helping researchers and practitioners learn more about the goals and motivations of their target users. We further speculate that data engagement interviews can be a widely applicable research method, suitable across a broad range of personal informatics domains, and scalable to accommodate different types of personal data and high-resolution, multisource data sets. Although this method is not intended as a replacement for more traditional tools, its success at engaging our participants with their data suggests collaborative analysis via an analyst-in-the-loop is a viable alternative for insight generation compared to using customized tools, and an interesting direction for future work that we briefly discuss in this chapter, but detail more thoroughly in Chapter 6.

5.2 Background

The growth in technology for capturing data about people's everyday, lived experiences has led to an explosion of personal data and a wealth of new insights. The population of self-trackers are actively collecting data and learning things about their bodies using tools developed across a variety of application areas. Work by Consolvo et al. [7], [8] and Epstein et al. [17], [135] focus on fitness and activity trackers to understand how people self-track; Moore et al. [3], Kim et al. [11]–[13], and Fang et al. [14] develop systems for monitoring people's indoor air quality to help improve people's understanding and awareness of their immediate surroundings; Chung et al. [136],[137] and Tsai et al. [6] build tools to help people monitor their health through digital diaries and nutrition trackers; and research into reminiscience [10] and location sharing [9] explore how they spend their time through calendars and social-media trackers. For personal informatics researchers and practitioners, the explosion in available data sets has created myriad opportunities to learn about how and why people engage with personal data. This has led to formalized models

[1] https://vdl.sci.utah.edu/EngagementInterviews

of the self tracking process, including both Li et al.'s stage based personal informatics model [5], Epstein et al.'s Lived informatics model [28], along with research into the kinds of behavioral changes this engagement provokes [25],[138]. These opportunities, however, require engaging participants with their personal data. In this section, we describe the range of engagement methods researchers and practitioners have at their disposal, and argue for data engagement interviews as a middle ground approach.

5.2.1 Lightweight methods

Design literature provides various methods for informing researchers about what or how to build regarding interactive tools or interfaces. Participatory design [139] is a common approach that invites users to collaborate in the design process to help inform the final result. This technique can help identify commonly undertaken tasks, or solicit feedback on the ways they may be improved. These approaches, however, are tailored for collecting insights that inform design *outcomes* rather than deeply understanding ways to productively engage people with their personal data. Work by Tolmie et al. [20] describes how understanding how to engage with personal data requires a deep, situated knowledge of people's lives and routines to accurately interpret. Further work by Fischer et al. [21],[140] illustrates that collaboration between a data worker and participant are essential to derive insights or offer advice on personal data.

Existing tools that visualize personal data typically support data review through simplified interfaces with minimal interactivity. These tools are mostly designed to show data, not to thoroughly analyze it. Tolmie et al. talk homeowners through their personal data using a basic time series plot for displaying sensor measurements [20]. Other researchers provide similar interfaces to end users for exploring how to support people engaging with their personal air quality data [3], [11]–[14]. These interfaces help people gain a sense of what their data *are*, but not what it can *do*. Without the ability to easily modify or change the data's representation and visualization, these interfaces can support only a limited number of data analysis tasks.

Alternatively, data sketching provides a lightweight method that has people sketch their impressions of data with minimal design effort. Data sketching removes the barriers to how data can be organized and formatted to promote brainstorming and collaborative workflows [141], [142], storytelling [143], and communicating knowledge about data to others [133]. The process of sketching also improves thinking [129], supplements discussion [144], and helps clarify ideas about design [142]. Engaging people with sketching helps them externalize their thoughts and ideas about data organization, visualization goals, and any underlying trends or traits they suspect may live within their data [133], [144]. In this way, sketching can free people to more quickly communicate organizational goals or ideas, especially in the absence of formal design or analysis vocabulary. Sketching often does not incorporate real data, however, and efforts to encode this information, either by hand or through digital tools, can be slow or complicated [141]. Instead, sketching can be a useful design component for *imagining* personal data, but it does not suffice for concrete analysis tasks or questions that require engaging personal data directly.

Data physicalization, another lightweight method, helps people explore and communicate data through geometric or physical properties of an artifact [68]. Data physicalization has been successfully applied in workshops [145] and teaching environments [146] to engage people through prepared data sets. Work by Thudt et al. [72] extends this approach to personal contexts, and uses data physicalizations to bring people closer to their personal data in support of self-reflection. Whereas this approach succeeds at deeply engaging people with their personal data, it requires a significant manual effort, and limits the representational accuracy and scope due to its inherent physical constraints [72]. Consequently, the nature and scale of many personal data sources prevent physicalizations as a practical analysis strategy.

5.2.2 Heavyweight software

The messy and complex nature of many personal data sets requires some level of wrangling, formatting, and preprocessing, making it difficult to integrate into general purpose tools, many of which some people already find hard to use in personal contexts [25]. As an alternative, researchers, practitioners, and quantified self enthusiasts invest significant effort to design and build bespoke tools for people to engage with their data. These tools typically focus on a narrow set of specific or predefined questions, thereby eliminating the need for users to translate their questions into analysis tasks, or to wrangle their data into an appropriate representation [16], [17]. This approach, however, does not let users explore a broad set of personally relevant questions, nor does it leverage users' rich, situated, and extensive knowledge of what aspects of the data are personally interesting and insightful, and which are not [147]. The challenge for designers is that people who have never directly engaged deeply with their data may not be able to predict what they want to do. For example, Epstein et al. surveyed 139 people on common tracking goals, motivations, and influences for informing visual and data analysis criteria to evaluate lifelog data [17]. After developing and deploying a tool to support these goals, subsequent evaluations "did not find any correlations between valued cuts and the reported goals of participants," prompting guidance that users should receive several possible designs, versus "simply [generating] cuts corresponding to stated goals, as that could deprive trackers of potentially interesting discoveries in their data" [17]. Even when designing customized solutions, personal informatics tools may still struggle to provide flexible analytic capabilities that completely address or anticipate users' needs.

5.2.3 A middle ground approach

The data engagement interview proposed in this chapter takes a middle ground approach by helping researchers identify user needs through directly engaging these users with their personal data before expending the significant design effort to develop a custom tool. Data engagement interviews are an adaptation of the pair analytics research method that captures reasoning processes in visual analytics scenarios [148]. Pair analytics borrows from protocol analysis and pair programming techniques by joining a subject matter expert and visualization practitioner to collaboratively tackle a relevant analytical task. This approach avoids the cognitive and social loads reported in standard think-aloud applications [149]–[151] by capturing participants' analytical reasoning through a conversational and collaborative problem-solving process. This approach, however, requires that participants share equal analytical and computational skills to productively work through their given task, which may not always be the case in personal informatics contexts.

We build on the pair analytics approach and incorporate a dedicated data analyst role within the interview team. Whereas the interviewer role is responsible for engaging the participant and keeping discussion on topic, the data analyst takes analytic direction from the interview participant. Unlike the standard Wizard of Oz approach [152] where the interview participant unknowingly interacts with an analyst, the data engagement interview brings the analyst to the forefront to gain the collaborative and conversational benefits of pair analytics. The data engagement interviews also distinguishes itself from the operator-mediated analytic interview process described by Grammel et al. [58] by directly incorporating a participant's personal data in the interview process. Rather than using fictitious or nonpersonal data, data engagement interviews allow a participant to tailor their interview based on their motivations and experiences to more deeply engage them in the analysis process. This approach provides a more personalized analysis experience with the help of a dedicated data analyst to assist researchers and participants with data exploration using flexible analysis tools.

5.3 Developing the interview framework

This section outlines how we developed the data engagement interview framework. We describe the framework in Section 5.4, and give more detailed descriptions and recommendations for performing data engagement interviews in Section 5.5. Section 5.6 reports on the outcomes of conducting data engagement interviews with our participants.

5.3.1 Developing the interview protocol

In our search for guidance on how we might elicit design requirements from our participants, we found both the visualization and human computer interaction literature lacked any suitable research methods for directly engaging everyday users with their personal data. We developed data engagement interviews with the assumption that interview participants are *not* professional analysts, and therefore incorporated a data analyst as an active member in the interview process to offload analysis tasks from the participant. This change helps lower the barrier for engaging with personal data while still providing a rich suite of analysis capabilities. We also incorporated additional ways to elicit participants' analysis goals, such as reviewing physical data printouts and sketching, to help externalize their ideas. Recognizing the potential complexity of the interview dynamics, we further modified our draft protocol by splitting the interviewing responsibilities between two interviewers to maximize our likelihood for collecting and capitalizing on valuable research insights [153]. This pair interviewer approach has one interviewer lead the discussion, and the other track the conversational flow to help keep things on task.

We refined the interview protocol over two rounds of pilot interviews. The first round of piloting helped streamline and organize the interview structure. We recruited seven first-round pilot participants from our research lab, six of whom were computer science graduate students, and one computer science undergraduate student. These first-round pilot interviews did not incorporate a dedicated data analyst. Instead, we had pilot participants role-play as asthmatic self-trackers and sketch what they wanted to do with a set of representative air quality data.

In the second round of pilot interviews, we added our data analyst into the interview team. Pilot study participants were recruited from a convenience sampling of undergraduate students pursuing nonanalytic degrees and nonstudents from an online forum. The second-round pilot participants consisted of four dance majors, one physical therapist, one market researcher, and one social media influencer. These pilot interviews focused on evaluating the feasibility of performing real-time data analysis within the interview and improved how we introduced and presented data to participants. All second-round pilot participants were compensated with a $20 Amazon gift card.

5.3.2 Analyzing data engagement interviews

The interview framework emerged from analyzing seven data engagement interviews conducted with our study participants. We audio recorded each interview for 8.9 hours of interview audio, and maintained individual Jupyter notebooks in Python from each interview to create a self-documenting record of participants' analysis process and visualizations. These notebooks also allowed us to amass a store of reusable code that we could employ in subsequent interviews [154]. Reviewing and reflecting on these artifacts helped us build an understanding of what our participants wanted to do with their data, and to identify their goals and their overall approach to data analysis.

The interviewers also engaged in reflexive discussion after each data engagement interview, sharing their thoughts and reactions to what each found surprising, unexpected, frustrating, or insightful during an interview. One of the interviewers compiled reflexive notes on these experiences after each interview, which were further supplemented with a secondary summary after listening to the recorded interview audio. This process captured additional aspects of the interview mechanics, including participants' stated questions, goals, and motivations, while also providing an overall commentary on the interview process. We revisited the reflexive notes frequently throughout the analysis process.

Each interview audio recording was also professionally transcribed and imported into Google Sheets. One researcher then blocked and summarized individual interview sections of each interview to create a high-level overview summary for other researchers to review. Three researchers then read through and annotated participant transcripts, and then met to discuss aspects we found noteworthy from a methodological or research perspective for each of the seven interviews. These meetings were also audio recorded. A researcher listened to these meeting recordings to further summarize the main discussion points, and populated an affinity mapping board with meeting summary notes. The resulting affinity diagram was supplemented with direct evidence within the transcripts associated with the summaries. The three researchers involved in analyzing the transcripts then iteratively produced an additional affinity diagram of core interview themes over several days. These themes informed the framework we describe in Section 5.4, and Section 5.6 presents the results from applying this interview framework.

5.4 Interview framework

The data engagement interview is a novel interview method that elicits engagement with personal data to support a host of observations and insights about those engagements. This method differs from more traditional interview approaches in two ways: first, through the inclusion of a dedicated data analyst on the interview team who has access to a prepared analytic toolbox and the participant's personal data; and second, by structuring the interview to cycle between exploratory and goal-oriented analysis strategies. The participant directs the analysis in these interviews by communicating their desired analytic tasks to the data analyst, who then performs the requested analysis on the participant's data. This process allows the participant to engage with their data in a flexible and personally relevant way, and provides researchers and practitioners opportunities to observe what participants *actually* do with their data when given the freedom and resources to do so.

Data engagement interviews foster a conversational dynamic around personal data by offloading the analytic burden from the participant, so they can more readily share thoughts on their process, justifications, and reactions, in their own words, as part of a naturally unfolding conversation. Maintaining this conversational dynamic can avoid post hoc rationalizations that can be common to other concurrent or retrospective verbal reports [155]. This approach can also help researchers collect more rich and authentic insights into participants' motivations and problem-solving processes Whereas several think-aloud techniques exist in the interviewing literature [155]–[157], few are tailored for specifically engaging visual analytic tasks and processes [148], and none incorporate self-tracked personal data in the analytic process.

The following sections describe the interviewer and analyst roles and phases of the data engagement interview framework. Section 5.5 provides details on how we prepared and performed these interviews through a case study with asthmatic families living with an indoor air quality sensing system. Section 5.6 outlines our case study outcomes. We also provide an online guide [134] with materials for preparing data engagement interviews, recommendations for selecting an analyst, and a sample protocol for how we structured our own data engagement interviews.

5.4.1 The interview team

We performed our data engagement interviews by adding a dedicated data analyst to a pair interviewer team [153]. In this arrangement, we divided interviewing responsibilities between two interviewers and left real-time analysis tasks to the analyst. One interviewer was responsible for prompting participants to articulate what they want to do with their data, and the other interviewer kept track of the overall conversational flow. Depending on interviewer experience or subject matter complexity, other research teams may be able to perform data engagement interviews with a single interviewer and analyst. If performing data engagement interviews using a single interviewer, however, this sole interviewer will be simultaneously responsible for engaging the participant in conversation; tracking various opportunities, comments, or observations worth digging into; and making sure to keep the interview on track and on time. In our data engagement interviews, we found that pair interviewing reduced the cognitive burden on the individual interviewers, and allowed for more focused and productive data engagement interviews in line with the experience of other pair interviewer teams [153]. Regardless of the number of interviewers, anyone conducting a data engagement interview should be familiar with the study participants and associated research data. This foreknowledge is important to properly guide the data engagement interview, pose meaningful engagement prompts, and avoid lines of conversation or analysis they know to be unsuccessful. For these reasons, we recommend that the interviewer role not be outsourced to an external third-party researcher.

Our data analyst was responsible for implementing participants' directions for processing their data and communicating the analysis results back to them in an understandable way. We encouraged the analyst to interact with the participants and interviewers to gain any necessary clarifications or analytic details for completing their analyses, although we cautioned the analyst not to actively comment on or suggest analysis options so as not to steer participants' choices or behaviors. We prioritized candidates with strong analytic and interpersonal skills when evaluating potential analysts, in order to select someone comfortable with analyzing data on-the-fly during an interview, while also taking direction and communicating with both researchers and participants. Good candidate analysts ideally are: experienced working with data similar to what they will process during the data engagement interview; fluent in their preferred programming language and processing environment and able to exercise good visualization and interview techniques.

5.4.2 Interview materials

The interview team brings with them an enriched and formatted collection of the participant's personal data, and a suite of tools and devices for conducting real-time analysis during the data engagement interview. We used Jupyter notebooks for conducting our real-time data processing, but we speculate that other interactive platforms such as Power BI or Observable can be effective. Although what and how much data to prepare can depend on the broader project aims or goals, it was our experience that participants' questions required access to external data sources to more fully contextualize and support analysis goals. For our participants, this preparation meant adding local weather and outdoor air quality measures to help further contextualize their own personal indoor air quality data. We benefited from performing pilot interviews, brainstorming, and drawing insights from the literature prior to conducting interviews to help us identify what kinds of additional data sources were likely to be helpful in our context.

Labeling, organizing, formatting, and cleaning data are important steps for any analysis project, and can take an estimated 80% – 90% of the effort in data analysis work [46], [158]. It is important, however, to consider how these transformations may impact or slow down an analyst's ability to handle unanticipated analysis requests during the interview. Breaking data into separate files can make certain analyses more modular, although the choice to segment data versus maintaining one large data structure can affect its accessibility. For example, in our interviews we experienced that population-wide summaries and between-participant comparisons became more time consuming to compile when this information was separated across different files and directories. These decentralized data caused our analyst to spend time during some interviews reformatting and wrangling several disconnected sources in order to address unanticipated questions, ultimately exposing the limits of the interviews' real-time capabilities.

The tools and techniques used to process personal data may also impact data engagement interviews. We came to our interviews prepared with a variety of read-to-apply analytic techniques based on what we suspected our participants may request and what we knew their data could support, enabling our analyst to quickly perform a variety of common requests in our interviews. Based on our observations of how participants engaged with their data, we also recommend considering the kinds of entry points [159] that participants may take into their data. For example, in our interviews, the participants often wanted to jump into their time-series data at a particular season, month, day of week, hour of day, or combination of these conditions. Anticipating entry points that rely on aggregations or data cuts can aid in quickly addressing participants' analysis requests.

Performing the data engagement interview will require that the interview team come prepared with a laptop and external monitor for showing data to a participant. We also suggest that the analyst prioritize creating visualizations that participants can easily read and understand. For participants comfortable with processing numerical information, standard statistical charts such as line charts, bar charts, and scatterplots should be sufficient [160], [161]. We recommend Munzner's *Visualization Analysis & Design* [162] or Ilinsky and Steele's *Designing Data Visualizations* [163] as starting points for researchers interested in learning more about designing effective visualizations. Standard interview materials such as audio/video recording equipment, note-taking, or sketching supplies are also good practice.

5.4.3 Interview phases

We divided our data engagement interviews into three distinct phases: onboarding, the engagement cycle, and wrap-up. Both the onboarding and wrap-up phases align with traditional interview practices, whereas the engagement cycle is a unique and critical phase of data engagement interviews. Figure 5.1 illustrates an overview of these phases.

Our **onboarding phase** introduced the participants to the overall goals of the interview and the format the interview would take. This phase includes introducing and explaining the role of the data analyst, along with the scope and scale of the data they have access to during their interview. During this phase, we actively primed our participants for engaging with their data by having them reflect and expound on their personal data

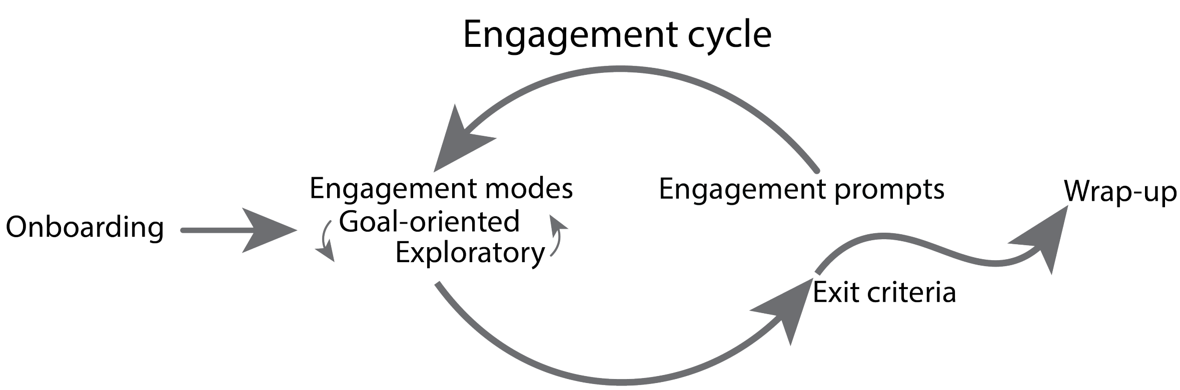


Figure 5.1. Outline of the data engagement interview phases. The data engagement interviews began with an introductory stage to remind participants of the goals and scope of the interview. Incorporating personal data into the interview transitions to an engagement cycle, where interviewers can guide the participant between engagement activities and research questions. Engagement activities are also cyclic, and can switch between exploratory and goal-oriented modes. The exploratory engagement mode starts with data, and relies on curiosity or surprise to determine what questions participants will use to direct the analysis process. Goal-oriented engagements use participants' prepared questions or goals for determining how they engage with data and the analysis tasks they undertake. Interviewers can pose various engagement prompts throughout the interview, either in response to participant comments and data engagements or while waiting for analysis results.

engagement goals, along with prompting them to further operationalize those goals into more specific and concrete analysis tasks. Next, the interview enters the **engagement cycle**. This interview phase cycles between the participants engaging with their data via the data analyst, and the interviewers prompting the participants with questions, comments, or observations that are meant to surface research insights and feedback. The participants can engage with their data using either an exploratory or goal-oriented strategy; these strategies are themselves cyclic, and provide the interviewers some flexibility during the interview for eliciting productive engagements.

Exploratory engagement is a bottom-up approach occurring serendipitously when the participants explore their data out of curiosity without a concrete goal in mind. We elicited this type of engagement by showing our participants some part of their data that we thought they might find interesting. Our participants also engaged in an exploratory approach when they inadvertently become distracted while reviewing data for some other goal. Distraction played a prominent role in our interviews, and came from participants' surprise and curiosity when they encountered unexpected features in their data. When participants engaged with their personal data with an exploratory approach, we used their curiosity and surprise as an opportunity to encourage them to generate a direct question about the data. This prompt often transitioned them to a more goal-oriented engagement.

Goal-oriented engagement is a top-down approach occurring when the participants engage with their data in a goal-oriented way by posing a question and directing the analyst to process the data in service of answering that question. We found that this mode of engagement pushed our participants to grapple with how to both operationalize their questions and interpret the results. During goal-oriented engagements, surprises often distracted our participants, and they reverted to an exploratory approach until they identified a new question or we guided them back to their original goal.

Once our participants began to productively engage with their data, we posed questions and made comments or observations that served our goals as researchers. These **engagement prompts** leveraged the participants' engagement with their data and had them answer questions about their analysis goals, preferences, and approaches; motivated the participants to improve compliance with self-tracking activities; or supported the participants in taking what they learned from their data to make positive changes in their lives. We injected an engagement prompt in response to a specific participant action or statement, or to fill space if there was a lull in the interview. These engagement prompts resemble the traditional semistructured interview prompts for questioning or clarifying participants' statements [164], [165].

Engagement strategies and prompts can feed into one another from the conversational dynamic that arises around engaging data [148]. The dynamics of our interviews shifted between engaging a participant with their data and engaging them with a prompt to observe *why* they wanted to engage their data, *what* their priorities or goals were in practice, and *how* this changed through access to flexible and personalized analysis. We ended our engagement cycles when the interviews reached the time or energy limit of the participant, a satisfying result for the participant, or saturation of insights and goals of the interviewers.

Finally, the interview enters the **wrap-up phase**. We thanked our participant and summarized the interview trajectory to provide closure, as well as re-state the study goals to explain how this interview fit into our broader research to further validate the participant's efforts.

5.5 Applying the interview framework: An illustrative case study with asthmatics

We conducted seven data engagement interviews with participants from a longitudinal study on how asthmatic families engaged with personal indoor air quality data. Our initial goal with conducting these interviews was to better understand and identify how our participants might go about analyzing their data to develop design requirements for a future visual analysis tool. The interviews were designed to be completed in 90 minutes to give participants enough time to deeply engage with one question. Some variation is to be expected, however, depending on participants' preparedness, overall engagement, or question complexity. For example, P4's lack of preparation led her to more quickly review different aspects of her data at a higher level. In practice, our data engagement interviews ran between 50 - 110 minutes (79 minutes average).

5.5.1 Recruiting and preparing the data analyst

We recruited our data analyst from prospective graduate and undergraduate student candidates within our university's computer science, mathematics, and physics departments. These candidates came from other researchers' direct recommendation, and in response to a $17/hr work study position for an interactive data analysis project. We briefed applicants on the nature and goals of the interview method and provided test data sets similar in scope and content to participants' data in preparation for a live-coding interview. The interview process involved the analyst using this test data set to work through several sample questions modeled after those participants had asked in their deployments. Our recruited analyst was a physics graduate student with extensive experience processing large cosmic ray data sets. This background allowed him to easily handle our time series sensor measurements and to repurpose signal processing scripts to help bootstrap filtering and aggregating our air quality data.

The data analyst is a vital component of data engagement interviews and requires both strong analytic and interpersonal skills. Our data analyst refined his analytic toolbox and interview skills through his experience participating in our second-round pilot interviews. In preparation for the primary interviews, they compiled their accumulated analysis scripts into a workbook that allowed him to quickly execute commonly requested data analysis tasks, as well as create and customize data visualizations using minimal commands. This preparation saved him time and helped make real-time data analysis a reality for the data engagement interviews.

5.5.2 Collecting and wrangling participants' data

In preparation for conducting our data engagement interviews, we collected, enriched, and formatted each participant's deployment data, and extracted representative subsets to present as physical printouts during each of their interviews. An example is shown in Figure 5.2. Previous work outlines the data collected through these deployments [3]. In summary, participants' deployment data consisted of three air quality monitors that logged measurements at 60-second intervals over several months; a table of algorithmically detected spikes [120] from each monitor data stream, including the time, location, monitor ID, and spike value; a table of outbound text message alerts sent to participants

Figure 5.2. Reviewing air quality data with the data engagement interview. Left: A screenshot of the data analyst's Jupyter notebook used to process and display participants' data. This view was mirrored between the analyst's laptop and external monitor when presenting data to participants. Right: A typical data engagement interview setup. Interviews included physical printouts of a participant's data to help them to understand what data were available. Some participants also used sketching to communicate how they imagined using their data to the interviewers and data analyst.

based on these detected spikes, including the message timestamp, content, and spike location; and a table of participants' replies to these messages, including the reply timestamp, content, and annotation source (text, tablet, or Google Home).

Based on earlier participant feedback, we further supplemented these data sources with daily EPA Air Quality Index classifications, [2] self-tracked respiratory health surveys collected through a parent medical study [2], and environmental data including ambient temperature and humidity, also measured by the air quality monitors. We integrated these additional sources to further contextualize participants' data, and to provide a richer set of analysis opportunities than would be possible from air quality data alone [20], [24], [26].

We formatted these data to support a number of anticipated data cuts [17], and prepared scripts to filter and facet participants' annotated air quality data according to

[2] https://www.airnow.gov/aqi/

various questions we had received throughout the study. Examples of these processing scripts and data files are available in our online guide [134]. Basic data processing allowed us to plot filter, group, and aggregate raw air quality measurements by individual participants, sensor locations, or time spans. Further processing also allowed us to cut this information by other temporal characteristics, such as assigned categorical labels like mornings, weekends, seasons, etc., or participants' own annotations that we knew would be relevant to how they thought of their indoor air quality. Other derived data, like the times and locations of detected spikes in participants' air quality data, provided more opportunities to partition and review this information on a spike-by-spike basis. We additionally coded participants' annotations with representative class labels to provide more categorical filter criteria such as supporting data comparisons between "cooking" and "cleaning" annotation types. The EPA Air Quality Index and participants' self-tracked asthma surveys provided more options for categorical and contextual processing. These parameters helped us expand the analysis space for flexibly reviewing participants' data and helped us prepare for a wide range of potential questions during the data engagement interviews. Figure 5.3 provides an example of the kinds of visualizations participants requested using their data.

5.5.3 Other materials

We conducted data engagement interviews in participants' homes, requiring us to come prepared with all necessary interview data, equipment, and supporting materials. Each interview included a laptop for analysis, an external monitor for sharing analysis results with participants, and an interviewing kit for each participant. This kit contained a worksheet to capture participants' analysis goals and physical printouts of their self-tracked annotations, detected air quality events, asthma control test scores [166], and representative subsets of any sensor measurements (Figure 5.2).

These materials acted as visual aids and physical tokens during the interview to help explain the scope and scale of participants' available personal data in the onboarding phase, and to help them reflect on their data engagement goals within the engagement cycle. We also brought pens and paper to support them with externalizing their analysis process through sketching if they preferred.

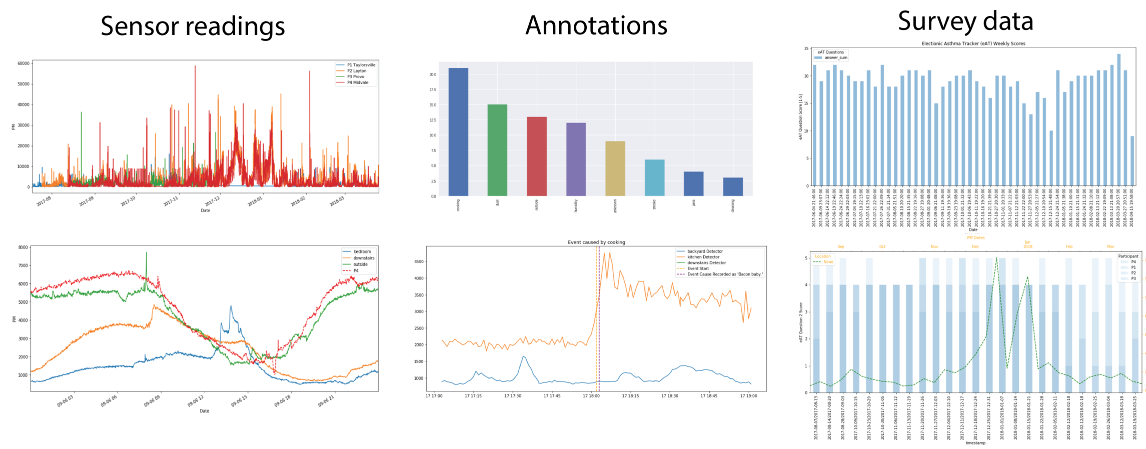


Figure 5.3. Typical visualizations generated during participants' data engagement interviews. Left: interviews started with showing participants the extent of their air quality data (top), which typically resulted on them zooming in to some feature (bottom). Center: Exploring air quality was facilitated by using participants' own annotation classes (top) to step through individual annotated spikes (bottom). Right: This information was usually contrasted against participants' asthma Control Test scores (top), which many asked to overlay against air quality dynamics in search of possible correlation (bottom).

5.5.4 Onboarding

In our data engagement interviews, the onboarding phase began ahead of the actual interview. Participants received an e-mail from the interviewers describing the interview's purpose and our request for them to come prepared with some questions to apply to their data: *"Imagine you've monitored and logged your home's indoor air quality for the past year. What would you want to do with it? What would you want to know?"* At the start of the interview, we began by revisiting this prompt and having the participant fill out a worksheet to capture their questions, motivations, and goals for engaging their data while we set up our processing environment. This worksheet was used to help focus the participant's thoughts and to serve as a visual reminder of their goals throughout the interview. In two of the seven interviews, participants neglected to prepare ahead of time. In anticipation of this possibility, we had prepared a collection of sample questions and offered them as options to chose from.

After participants completed their worksheet, we had them present their goals and questions, and asked them to explain why they made their choices. Finally, we reminded the participant of their previous year-long indoor air quality deployment [3] and the data that were collected using physical printouts of representative samples of the participants' own data, such as those shown in Figure 5.2. Using these printouts, we talked through what was available to them during the interview, and what each of these data sources contained.

Whereas some participants were ready at the end of the onboarding phase to direct the analyst on how they wanted to load and analyze their data, others were more hesitant or unsure. In these situations, we performed a short mock exercise between the interviewers and analyst with local weather data to role-play basic analysis tasks and representative participant/analyst interactions. We performed this exercise in an attempt to lower barriers or anxieties around engaging with the analyst so that participants might more freely take control of the interview once their data were loaded.

5.5.5 Data engagement

After the onboarding phase, the analyst loaded and presented the participant's data in an overview visualization, with interviewers encouraging the participant to direct the analyst on the ways they wanted to analyze or explore their data. Participants who were ready to dig into one of their prepared questions transitioned to goal-oriented engagement and began to direct the analyst to process data according to how they imagined approaching answering their question. We prompted these participants to explain their thought process and operationalization strategy. These discussions revealed details related to people's analytic approach and assumptions about their data. Other participants used the overview visualization to initiate exploratory engagements. For example, when initially viewing their air quality data, some participants asked to zoom in on prominent air quality spikes.

Outliers or deviations in air quality or health survey data frequently drew participants' attention while they engaged with their data. These features were a common source of distraction that shifted them from a goal-oriented to an exploratory engagement mode. Other distractions came when overlaying additional data sources, with many participants attempting to contextualize or correlate trends in one data source to features in another. In particular, several participants were interested to assign causation between their indoor air quality and health outcomes, which shifted their focus away from prior goals, and toward combining data sources to search for potential correlations. When participants became distracted, we prompted them to understand why, what was interesting, and whether they wanted to modify or change their question based on what they saw. Some participants took the opportunity to continue with a new focus, and others would shift back to their original question. Figure 5.4 outlines two representative data engagement interview timelines for Participants P3 and P4 showing how this process typically unfolded.

Throughout our data engagement interviews, we found that many participants quickly cycled between exploratory and goal-oriented engagement facilitated by the dynamic data analysis afforded by the interview framework. When participants were presented with new data, there was an immediate period of review and reflection, during which the interviewers prompted the participants to reflect on their path to answering a question, and to share how what they saw affected the way they thought. These reflexive prompts led to more and different questions that encouraged the participants to give the data analyst new directives, restarting the engagement cycle once again.

In all but one interview, the interviewers continued eliciting cycles of engagement until the participants reached at least one satisfying answer to one of their questions; in these interviews, the interviewers wrapped up the discussion once they reached the participants' time or energy limit. In the one interview that did not reach a satisfying answer on the part of the participant, the interviewers wrapped up once they reached saturation and were no longer extracting new information from their engagement prompts.

5.5.6 Engagement prompts

Our engagement prompts throughout the interviews served four purposes. First, we used prompts to encourage and elicit data engagements: *So, with the data we have, is there anything that you would want to see now? What do you want to see next? Does this data match what you thought you would see? What is important to you from this result?* Second, we prompted participants about their analysis strategies to gain insights for the design of a future visual analysis tool: *How confident are you that something like this gets at what you wanted to know? Does this satisfy the question you have? Based on all of the data you've seen, how do you feel it addressed the question you had at the beginning?*

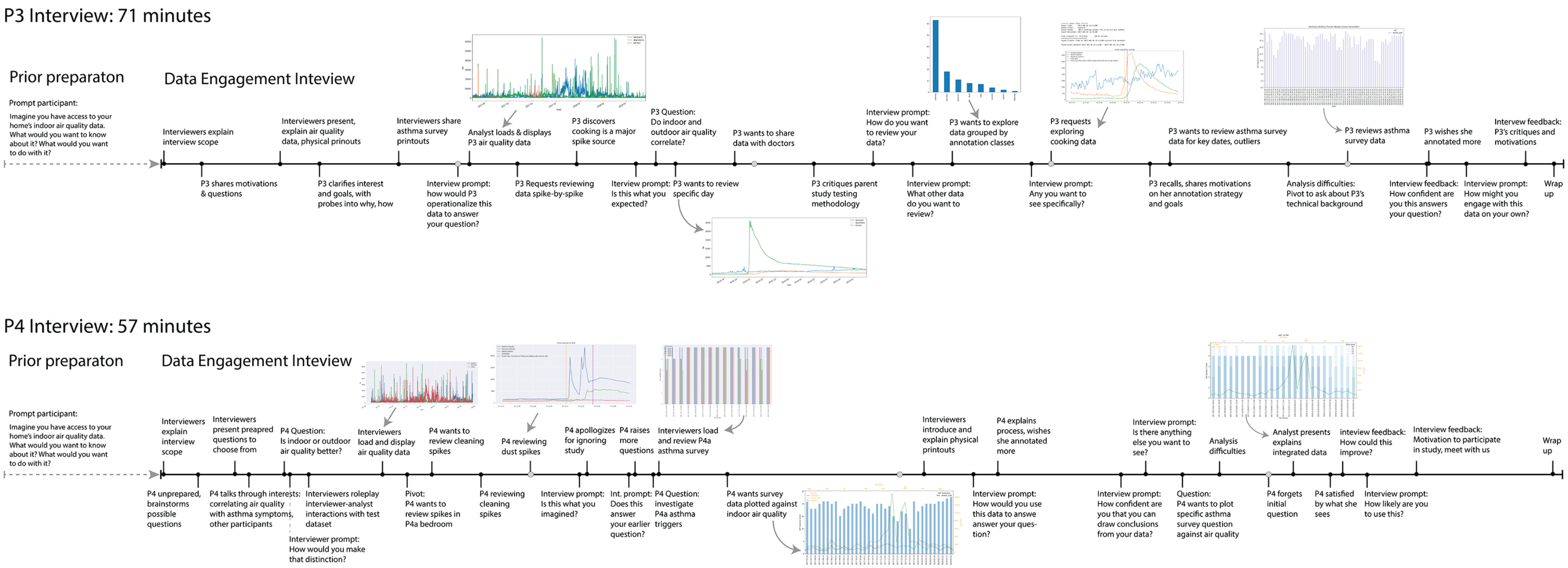


Figure 5.4. Examples of the data engagement interview timeline for P3 (top), and P4 (bottom). These timelines include visualizations created by the analyst, and summarize the conversational flow between the interviewers and participant. The gray dots are spaced at 15-minute intervals. Higher resolution versions are available on the data engagement interview website [134].

Third, we reserved our final engagement prompt to collect observational data on people's operationalization abilities: *Knowing what you know right now, how would you use your data if you had to go back and do this again?* Fourth, we prompted participants to give feedback on the data engagement interview method: *How was this process for you? How was interacting with the data analyst? Did anything feel too slow or rushed?*

5.6 Interview framework outcomes

This section reports on our observed data engagement interview outcomes across seven primary participants. These outcomes speak to the interview's strengths at engaging and teaching participants through data analysis, teaching researchers through observing participants, and supporting the design and outcomes of personal data studies.

5.6.1 What participants say versus what they do

There is often a (big) difference between what people say they want to do with technology, and what they actually do when given the opportunity [167]. In almost every data engagement interview we conducted, we observed participants asking unanticipated questions and approaching their analysis in unexpected ways once they began to actively engage with their data.

For example, at the start of P4's interview, she stated her interest to understand how her family's indoor air quality might have affected her daughter's asthma. She said she wanted to overlay periods of data that contained spikes with her daughter's self-tracked asthma data. The analyst began by pulling up an initial overview of her data:

> **Interviewer:** Based on what you wanted to see, how would you like to use this data to answer your question?
>
> **P4:** Actually, now that I'm looking at [the data], I would also like to see how [spikes in the bedroom] correlate to, for example, vacuuming or something.

P4 noticed spikes in the data stream from the sensor in her daughter's bedroom that were much larger than readings from the other sensors. This realization pivoted the interview toward exploring the sources of various spikes and how they correlated (or not) across the various sensors in her deployment. From this comparative analysis task, she learned that certain types of cleaning activities impacted her air quality more than others, which led to a broader conversation on alternative goals and more questions.

Curiosity and surprise also triggered other participants to alter their goals, exposing a possibly more authentic portrayal of what they wanted to do, and could do, with their data. At the start of P3's interview, she told us she wanted to get a sense of whether the air quality in her home was good or bad, and what changes she could make to improve her home's air quality. She initially stated that she wanted to look at large spikes and their annotations to find possible patterns. While exploring her data, however, an especially prominent spike captured P3's attention:

> **P3:** Okay. Wait, go back one more. That one, I want to see that one, because that's weird.

At this point, P3 switched from an exploratory engagement to one that was goal-oriented in attempt identify the source of this specific unannotated event. She went on to identify other interesting spikes that had gone unannotated and realized that answering her questions relied on richly annotated data, prompting frustration that she had not been more diligent in self-tracking events during the deployment.

Participants' interests occasionally even contradicted their own stated goals within the same interview:

> **Interviewer:** Would you ever look back to see when you were sick and look at the air quality then?
>
> **P5:** No, it would just be for right then, on demand.
>
> *[...Loads data showing previous spikes...]*
>
> **P5:** Okay, that's pretty interesting. That makes me wonder if it was all through the house [...] That's a question that I have, was that at that same time I was sick?

P5's stated interest to review his indoor air quality had always been motivated by wanting to know what was happening around him in the moment [3]. The process of reviewing his data, however, made him engage deeply with exploring the source and potential health impact of an earlier pattern of late-night spikes, going against his claim minutes earlier about his disinterest in retrospective analysis.

Providing feedback on data without actively engaging with it required our participants to *imagine* how their data may support their goals, or which aspects they suspected would be relevant or interesting to review. Counter to other research techniques that do not incorporate personal data, our observations from conducting data engagement interviews illustrate how directly engaging with personal data can refocus participants' attention, resulting sometimes in new and different goals.

In personal informatics studies, this difference between what people say and do is a critical one when trying to understand how access to data can impact people's lives. Without an understanding of what people *actually* do, researchers and practitioners run the risk of designing tools for the wrong tasks, misunderstanding people's relationship with their data, and missing opportunities to support people in making positive changes in their lives.

5.6.2 Data engagement interviews can be engaging

All seven of our data engagement interviews were successful in getting participants to actively and productively engage with their data. In most cases, participants readily hopped into the engagement cycle and began directing the data analyst. In two interviews, however, the participants required a significant amount of initial prompting by the interviewers to *really* engage with their data.

One example of this reluctance was the interview with P4, who had been much less engaged during her deployment than other participants [3], and did not come prepared with a question for her data engagement interview. When discussing her goals with P4a, who made a brief appearance in P4's interview, she described her overall detachment from the process:

> **P4:** I was not well prepared. You can do a better job. Here's the questions. You can see how much I wrote. [points to blank page]

The interviewers were undeterred and discussed aspects of P4's deployment that they knew were important to her. Despite her unpreparedness, reviewing her data and annotations reminded P4 of the challenges she faced with managing P4a's health. By the end of the interview, P4 was actively engaging with her data to explore what role her indoor air quality may have played in impacting P4a's health. During the wrap-up phase, she apologized for her initial reluctance at the start of the interview and said:

> **P4:** I know I wasn't really well prepared. [...] I kind of forgot our ultimate reasons for doing it. But looking over these things, I remember now what we were dealing with two years ago. Like I said, I had kind of forgotten about that, but looking at this, I do remember the struggles we were having and what we

> were trying to do to figure things out. And having this information back then probably would've been very helpful.

P4's distance from the initial deployment made her forget many of her goals for participating, but her ability to directly engage with her data surfaced several memories and quickly led to multiple questions, ultimately providing her with valuable insights. P4 even quipped at the end of her interview that she had lost track of which question we were working on from having posed so many.

The data engagement interviews were also successful in engaging new members of the participating families. During the deployment stage (S1), both the interactions with the air quality monitoring systems and the interviews were predominately led by a single, primary caretaker in the home [3]. Postdeployment interviews confirmed these disengagements:

> **Interviewer:** Could you describe your level of involvement with the [deployment]?
>
> **P1-S:** Minimally. If the app crashes, I'll turn it back on. If I see a big spike, I'll ask [P1] about it. But I'm not doing anything differently because of it.

Family members like P1-S attributed their indifference to not being personally affected by air quality, not having the interest or time to engage with their air quality system, or the distribution of labor within the home.

During the data engagement interview with P1, however, P1-S entered the room and stopped to look at the data:

> **P1-S**: So it looks like inside gets worse than outside.
>
> **P1:** It's hard to tell...
>
> **P1-S:** But everything spiked.

P1-S proceeded to pull up a chair and take part in the rest of the interview, directing the analyst to review indoor air quality spikes as a way of comparing their indoor and outdoor air quality. A similar engagement happened with P2's husband during her data engagement interview.

The involvement of P1-S and P2-S in the data engagement interviews surprised us, as we had made considerable efforts to engage multiple family members during the initial system deployments and again in postdeployment interviews, but with limited success [3]. Making the families' personal data accessible and analyzable through this interview framework succeeded in generating more interest and broader family engagement than our previous traditional deployment and interview practices. Effecting this level of engagement with participants and previously disengaged family members lends further evidence of data engagement interviews' promise, both as a research method and an outreach tool for generating interest with personal data.

5.6.3 Engagements can teach participants new things

During data engagement interviews, many of the participants learned new things about their home's air quality, acquired new data analysis insights, and became more confident in their analysis skills. For example, while reviewing data during the interview, P1-S — who had not engaged with the data during the initial deployment — voiced a strong opinion that his home's indoor air quality was quite poor. P1, looking at the same data, drew different conclusions. This disconnect between P1 and P1-S's interpretations led to a broader discussion between the two of them on how the deployed sensors measured air quality data, and how to interpret those measurements. P1-S directed the analyst throughout the conversation to process and display the data in a variety of ways, which P1 used to explain differences between how personal activities affect their indoor air quality, compared to how outside air quality conditions influence the air quality inside their home. Eventually, P1-S and P1 arrived at a shared understanding, backed up by the data:

> **P1-S:** It looks like ... Maybe I was wrong, okay. I was just curious, because we hear that all the time ... They're just like, "Stay indoors. It's bad outside." But then I've seen way more spikes indoors than I would've guessed. I'm just curious if, even if you're indoors with all the windows closed, do you still see a spike indoors on the bad [outdoor] days?

In this situation, the ability to jointly engage with data resulted in a productive conversation between participants, resulting in P1-S broadening his understanding of air quality, revising his position, and eventually articulating a more actionable question to explore.

Working closely with data led other participants to recognize analysis limitations when trying to correlate data collected over different timescales. P2-S directed the analyst to overlay air quality measurements from real-time sensors to P2's self-tracked and weekly aggregated asthma survey scores:

P2-S: Maybe you could see trailing [asthma results]. Like if there were high [air quality spikes] on Monday, maybe Tuesday your [asthma] values would just suck [...but these surveys are] weekly, so you couldn't really tell.

P3 describes a similar limitation:

P3: If he has an asthma attack and there was a spike in the kitchen, if I'm using weekly [asthma] data, I'm not sure I can correlate the spike with his asthma, because maybe it was when he was camping, but I might not know when I come back and look at the information.

Whereas P2-S identifies limitations of trying to attribute specific real-time events with a highly aggregated score, P3 identifies the problem of correlating two different data sources without contextual information that would indicate whether these sources do indeed have a dependency. Both of these realizations came from directly engaging with the data and were unprompted from the interviewers. Our previous participant interviews in S1 had recorded their frequent requests to compare and correlate air quality data to health data, yet it was not until participants directly attempted this comparison that they realized inherent challenges that come from working with diverse and independently collected data.

Other participants learned that data analysis was not as daunting as they assumed it would be. Given P4's general lack of engagement throughout the deployment stage of our longitudinal study, we were surprised to find her engaging and directing the analyst by the end of her data engagement interview. In spite of her initial lack of confidence or interest for engaging her data, she felt more comfortable in her ability to engage and analyze her own data by the end:

Interviewer: Seeing the data in this way, how confident do you feel being able to draw some of these conclusions?

P4: I think I could probably do it. For me it would be more a matter of how simple the program is. I am very bad with computers, so it would have to be very simple. But maybe just talking about it, just using this information and applying it, I could do that. It doesn't seem hard.

Despite P4's initial lack of confidence, working with an analyst helped her to develop a sense of competence that she may not have acquired through use of a custom system. We attribute this change of behavior to lowering the analysis barrier, and allowing P4 to focus on her data and what it could show her, rather than the analytic steps or perceived skill required to show it. In fact, all participants appreciated working directly with an analyst, and often remarked how this collaboration helped make the process, and their data, more understandable:

> **P2:** I think because we were working together it wasn't confusing, and you told me what was going to change. [...] it was helpful that you asked me to organize it, otherwise I would be like... this doesn't matter to me, why would I look at this, you know?

P4a felt similarly:

> **P4a:** If you just showed me all this stuff, I'd be very confused. But with [the analyst], I think it helps people understand it better [...] I think it's a very flexible way for you guys to cooperate with the participants, and say like, "Is that what you're asking about?" And then answering the questions about that.

The interview process also impressed our more analytically experienced P5:

> **P5:** Overall, being able to see [my data], especially interacting with the questions, was kind of cool. I thought it was kind of cool to be able to actually have a question, and look at [the data]. [The analyst] changed four words, and three numbers, and boom, boom, boom, and there [is the answer]. That was pretty awesome, actually.

These observations lend evidence that collaborations with a dedicated analyst can help participants to critically engage with their personal data in data engagement interviews, and that these engagements can benefit participants' sensemaking, problem solving, and overall sense of agency. These benefits also advance our own research goal to observe how to help people gain insights from their personal data.

5.6.4 Engagements can teach researchers new things

As a team of visualization and personal informatics researchers, we conducted our data engagement interviews with the research intent of acquiring design requirements for a future visual analysis system that could support people in analyzing their personal data. Analyzing these interviews revealed a wealth of new insights into the challenges and opportunities for visualization design. We report in detail on those insights in a companion paper [1]; here, we briefly summarize a few of our findings and their implications for visualization design to illustrate the usefulness of data engagement interviews for gaining research insights. We encourage readers to see our companion paper for more detailed validation of the utility of the data engagement interview method.

Given our study context of asthmatic families monitoring their indoor air quality, we approached our participants' data engagement interviews in a serious and goal-oriented way, expecting they would as well. Once our participants began engaging with their data, however, we observed that they became playful with how they recounted their deployments, and were quick to joke about previous annotations, many around cooking and burning food:

> **P1:** It's funny, I kind of did start putting in snarky remarks in some of the comments, you probably noticed. 'Cause there was a while that my oven had burnt pizza on the bottom and every time we would turn the oven on [the system] was like "HEY!! HEY!! HEY!!" and, yep, still haven't cleaned my oven. You wanna come clean my oven? 'Cause I still haven't cleaned my oven. [laughs]

Similarly, for P3:

> **P3:** We did the same thing over and over again. Bacon. [laughs] I think that like 90% of our annotations are probably bacon. I like bacon! [laughs]

During the interview, P3's playful interest in bacon transitioned into a broader exploration of her cooking habits. This exploration provided P3 with a more holistic view of what was causing poor indoor air quality in her home:

> **P3:** I remember making the connection between the olive oil and the [spikes]. And I also knew that it was kind of every time we cooked bacon there was a [spike]. But I guess I didn't realize how many of them, overall, were actually cooking episodes... Like "cooking pancakes," "cooking eggs," "[my daughter] burning the tortillas." It's all cooking.

P3 and other participants used play as a mechanism to dive more deeply into their data, which frequently led them to serendipitous discoveries. This type of playful and serendipitous engagement is understudied and undervalued in visualization research, perhaps due to a decades-long framing of visualization as a vehicle for cognitive amplification and insight generation [168]. Instead, most visual analysis tools are designed for goal-oriented behaviors [19], [169], [170]. Our findings suggest that prioritizing play and serendipity in the design of new systems could lead to innovative ways to support people in engaging with personal data.

As we discussed in Section 5.6.3, directly engaging with personal data during data engagement interviews helped all our participants learn new things, and in some cases increased their confidence for doing so. Yet, when we asked them directly at the end of the interviews how likely they would be to analyze their data on their own, all but P4a – our youngest participant – were reluctant to do so. Asthmatic parents, and parents of asthmatic children, live lives full of responsibilities. We witnessed this during our interview with P1, whose parenting and household routines left little time for exploring or analyzing her personal data:

> **P1:** As a busy mom with small children, having to just... [video game noises in the background]
>
> **Interviewer:** You're busy?
>
> **P1:** Yeah, busy! Obviously! With herding small people [laughs]. For me... it's interesting... I just don't have the time to sit down and look through all the numbers, and do all that stuff.

For P2, her reluctance stemmed from a concern about medical implications:

> **P2:** I don't know that I would ever just pull it up and look at it for data's sake, if that makes sense? I'm not a numbers person, I'm not a computers person. If I can look something up and say, "hey, I see this pattern," I can take it back to my doctor and talk about that there, and maybe that helps change a treatment plan, or whatever ... I could see myself doing that.

In this case, P2's hesitation came from the potential health risk of doing something wrong, and preferring instead to have her doctor interpret her data, rather than risk drawing those conclusions herself.

We were surprised by our participants' lack of enthusiasm for a visual analysis tool that would enable them to perform the same types of analysis they had engaged with during their interviews given the productive and positive outcomes of those interviews. Furthermore, if we were unable to motivate tool-usage by asthmatics living in an area that frequently experiences some of the worst air quality in the country [171], we wondered how hard it might be to motivate other people living in less dire circumstances. The success of the data engagement interviews with our participants, however, points to opportunities to focus visualization research efforts on designing collaborative social systems rather than just tools.

Chapter 6 expands on these two ideas – designing for play, and designing social systems – and presents more results from analyzing our data engagement interviews. The interviews provided us with new insights into how we might build future visual analysis tools and systems that help people to engage with their data. We provide the summary here as evidence for the efficacy of data engagement interviews as a research method.

5.6.5 Seeing the value of self-tracking

Like many personal informatics domains, our specific context — asthmatic families living with an indoor air quality monitoring system — includes self-tracked data. These data include participant-provided annotations of household activities, such as vacuuming and cooking, and a daily survey that tracked the respiratory health of the asthmatic family members. Although most of our participants maintained a high degree of self-tracking compliance throughout their deployments [3], several of them still found the quality of their data lacking when they tried to make use of it during the interview:

> **P3:** It kind of frustrates me. It would make me want to go back now, knowing that I could have it all, I would be more vigilant about [annotating]. Because I kind of get lax about it. Then I'd be [annotating more] so that I could use that to cross-reference stuff.

In this case, P3 realized how the lack of content in some of her annotations failed to support the kinds of correlations she was looking to make with her health data.

P6 experienced a similar issue when attempting to use his health survey data to determine whether his home's air quality affected his asthma, but was unable to do so because of his irregular and inaccurate survey responses:

> **P6:** You know a lot of [the value] is dependent on good sensors and then good data I'm inputting. Knowing how it can be done is going to motivate me to pay more attention to those [survey questions]. Because moving forward, if this is an opportunity to get my raw data, I'm not going to want to see just a bunch of fives. Like there's no way in January, December, I was that fine every week. That to me is just nothing but laziness on my part... Knowing that the information is retrievable motivates me to want to provide more accurate data.

Through engaging their data during the interview, P3 and P6 came to understand that their self-tracked data are valuable only when they commit to tracking regularly, accurately, and richly. Even the least compliant self-tracker among our participants came to understand the value of her data during the data engagement interview:

> **P4:** I remember we got a little tired of all the [annotating] and now I feel bad I didn't respond more because I can see how you use this.

Although we performed these interviews after the participants' deployments had ended, and thus cannot say that the motivation they exhibited during the interview would translate to better self-tracking, we speculate that conducting data engagement interviews early in a study could motivate participants to better comply with self-tracking. Motivating and maintaining self-tracking is a balance of lowering barriers to reduce capture burden [28], [172] or tracking fatigue [25] while maintaining enough engagement to not lose a sense of responsibility to the tracking process [173]. Motivation is directly related to people's willingness to track [174], making data engagement interviews a useful technique for engaging participants to improve or maintain their self-tracking habit once they see how useful their data can be.

5.6.6 Improving field deployments

Our field deployment design included an alert system that sent a text message to our primary participants when the air quality monitoring system detected a spike in their indoor air quality. We developed this alert system primarily as a mechanism to encourage annotation — the participants could respond to an alert-text with a short message about any potentially correlated activities occurring in their home. As the alerts were meant to elicit a response by the participants, we created rules that would turn off alerts during nighttime hours.

Ahead of the data engagement interview with P5, we knew that he had been using the air quality data to hold his family members accountable for impacting their home's indoor environment [3]. We learned during the interview that he had greater asthma symptoms at night, which motivated him to direct the data analyst to look at air quality sensor data and annotations during nighttime hours. He found, however, that nighttime annotations were missing:

> **P5:** Well, if we [had the data], I could say okay, who was out in the kitchen cooking because [my son] likes to cook late night snacks, sometimes. But, I guess we won't be able to get there.

Although it seemed reasonable to enact text alert rules based on our assumptions of normative family dynamics at the time of designing our field deployment, our assumptions denied P5 opportunities to sufficiently self-track. Worse still, our assumptions and decisions caused P5 to develop an incomplete awareness of his indoor air quality by not alerting him to many nighttime air quality spikes, which he was surprised to observe during the data engagement interview. Had we performed this data engagement interview early in the deployment, we could have modified our texting rules to better accommodate P5's family dynamics, and enabled him to be aware and make an effort to investigate their cause. We speculate that other field deployments would similarly benefit from using data engagement interviews early in a deployment to challenge normative assumptions built into deployed technology.

5.7 Discussion

Our case study outcomes lend evidence that data engagement interviews can engage people with their personal data in analytic contexts (Section 5.6.2;). teach participants new things through their engagements (Section 5.6.3); motivate the value of self-tracking by showing people how they can use their data (Section 5.6.5); and improve our own understanding of how to better design for users' needs (Sections 5.6.6, 5.6.4). Where our previous high-level participatory workshop feedback was related to things participants *imagined* wanting to do, our choice to incorporate personal data as a core design element in our interviews helped us to directly observe what our participants *actually did* with their data (Section 5.6.1). We speculate that these outcomes more broadly position this interview method as a viable interview technique for personal informatics and visual analytics research.

5.7.1 Flexible, to a point

During our data engagement interviews, we strove to create the illusion that any analysis was possible, that it could be done in real-time, and that it was made to order for our participants. In practice, however, data analysis is often not an instant-answer kind of endeavor [46], [175]. The analysis tasks within our data engagement interviews were often nearly real-time due to our data preprocessing and prepared scripts based on the kinds of questions we anticipated our participants might ask. In spite of this preparation, participants still managed to pose questions that surprised us. In some ways, this was a victory — despite all of our previous efforts to find out what they wanted from their data, data engagement interviews still yielded new kinds of questions. Yet, some of these questions also slowed our analysis down to a crawl and shattered the real-time data analysis illusion.

When these situations arose in our sessions — as they did with P2 P3, P4a, and P5 — we did not have a predefined plan for how to respond. These breakdowns emerged when attempting analysis tasks that required additional or unexpected data processing, such as aggregation in ways that underlying data organization made difficult, or attempting to answer questions that sat at the outer reaches of what was possible with the participants' available data. We speculate that these challenges stem from a need to reformat data on the fly, or a lack of access to relevant data in the limited time available for participants' interviews. When encountering these lulls, interviewers filled this downtime by leaning on interview engagement prompts to have the participants anticipate or reflect on what the answer *could* be, while the analyst wrangled with the data. In some of these cases, however, it became obvious that we were stalling, and some participants even apologized for asking the analyst to do something "hard." Based on our experiences, we encourage interviewers to have a plan for what to do when the data analysis requires time. One suggestion is to cut short certain analytic pathways or reject the question outright if interviewers suspect they will not be productive. If a participant is a long-term participant, another approach is to work on the analysis postinterview and bring it to the participant later. Having a plan ready for these circumstances can help reduce the chance of taking the interview participant out of a collaborative or analytic head space, or disrupting the data engagement interview flow.

5.7.2 Two interviewers are (probably) better

Our decision to include two interviewers on the interview team was rooted in our concerns about the complexities of data engagement interviews from the interviewer perspective. We anticipated that a single interviewer may find it difficult to remain deeply engaged in both the data analysis process and guidance of a participant, as well as keeping track of the larger interview direction. Adding a second interviewer to our interview team eased the interviewer burdens in our data engagement interviews and allowed one interviewer to lead the discussion and maintain engagement with the participant, with the second interviewer ensuring the goals of the interview were met and calling attention to anything interesting or surprising the first interviewer may have missed. Our pair of interviewers were also able to discuss their insights and reflections with each other postinterview, and did so as they drove back to their lab from each participant's home. We speculate this pair interviewer approach has the potential to not only increase the quality of findings, but also improve rapport with participants [153].

In some circumstances, however, a second interviewer may be considered a liability. Examples of this include instances where multiple interviewers may intimidate some participants, such as when engaging sensitive populations or research topics. Thus, the decision to utilize two interviewers hinges on whether these potential interpersonal effects outweigh the benefits of sharing interview tasks. For data engagement interviews, we suspect that the complexity of the interview protocol necessitates a pair interviewer, but that the interview team should consider ways to reduce negative effects such as intentionally diversifying the team.

5.7.3 Empowerment

Prior to conducting data engagement interviews, we had struggled to engage other family members throughout our data collection and deployment phases. We report in previous work that this lack of engagement was driven by a division of labor within the home and a general lack of interest from nonasthmatic family members [3]. Using data engagement interviews, however, we were able to motivate previously disengaged family members (P1-S, P2-S) to participate in our study. Based on these outcomes, we speculate that it may be possible to advantageously use data engagement interviews for purposefully motivating disengaged participants or family members.

Although we were able to use these interviews to motivate and engage more people than at previous times in our longitudinal study, not everyone was equally empowered. Many of our female participants expressed low confidence in their participation (P1, P2, P4, P4a), whereas male perspectives were more confident and outspoken (P1-S, P2-S, P5, P6). We see this shift as potentially rooted in traditional gender role stereotypes associating a feminine identity with caregiving and a masculine identity with math, data analysis, and computing. Similar dynamics exist in the context of smart homes, where incorporating a technological component into household maintenance shifts domestic attitudes toward "default to the 'household expert'," who is typically male [176]. Data engagement interviews with P1 and P2 reflected these dynamics by soliciting, or complying with, their husbands' priorities. This deference stands to alter the interview dynamics and may have obscured P1 and P2's perspectives. Being aware of these interpersonal dynamics can help researchers plan their own data engagement interviews and purposefully guard against collecting or propagating normative views in the design of new data analysis tools. In this vein, fielding a more diverse interview team may help researchers motivate and engage with traditionally disempowered perspectives, and offers another advantage for having multiple interviewers for this method.

5.7.4 A way for improving experimental design

The benefits of data engagement interviews have so far focused on its ability to explore what personal data can show or say. This interview method, however, can also be used to identify improvements for study or experimental designs. Performing data engagement interviews earlier in a design study can capture valuable observational data on what participants do and how they think to do it. These data can then allow researchers to more quickly learn about participants' habits apart from any assumptions or self-reports and further refine a candidate designs to better fit participants' lives and routines. In our case study, waiting to conduct our data engagement interviews ultimately prevented us from correcting a design assumption that kept P5 from receiving system alerts from unusual late night spike activity in his home deployment. Performing these interviews sooner would have allowed us to iterate on our messaging protocol, allow P5 to annotate these phenomena, and enable us to collect more detailed deployment data. Deciding to integrate data engagement interviews as part of an experimental design protocol can help researchers improve the quality and quantity of ecologically valid observational data over the course of a study.

5.7.5 Transferability

We conducted data engagement interviews with a small number of participants working with indoor air quality data, but speculate that this method can transfer more broadly to other personal informatics domains. This interview framework is not tied to any one kind of personal data, and so this approach can be tailored for various research objectives related to how people review and learn from their data. Data engagement interviews allow participants to apply their own lived experiences to identify, direct, and prioritize analytic tasks for their own particular needs. Free from this responsibility, researchers can focus on using the engagement cycle to strategically prompt participants and capture relevant research insights. We speculate this method can be applied in design contexts to capture more actionable and accurate feedback early in research design studies; for helping everyday people to better understand and interpret real-world data processing and interpretation by work through basic analysis tasks using their data; or as a hands-on educational tool for motivating self-trackers to improve or persist in their tracking regimen.

Data engagement interviews can excel at uncovering how people engage with and analyze data if they have some knowledge about the context of that data, regardless of their backgrounds or analysis expertise. This strength especially lends itself to personal informatics research, where people are experts on their own lives by definition, but often do not have analysis expertise. For the data engagement interview to work well, the data sets need to be rich enough that they lend themselves to analysis tasks, and the data should either be quantitative or easily quantifiable. Some quantitative data sets, however, will be too dense or complex for the analyst to work with in real time while the researcher keeps the participant engaged. For example, accelerometer measurements may be difficult to engage in this context but, if processed into step counts, these data are likely to be engaging.

Although our access to various data sources and manually annotated data helped us to flexibly process participants' data in our interviews, this is not a requirement for successfully performing data engagement interviews. Instead, we observed that participant interest plays a greater factor for successfully completing data engagement interviews. We speculate that this interview method is less likely to be useful in situations where the participant has limited or no familiarity with the context of the data, where they genuinely do not care about the data, or where the data do not lend itself to quantitative analysis. Our participants' engagement came from their curiosity to review annotated personal data, though motivation to review data in a specific application domain – annotated or not – may be sufficient provided the data is legible to the participant. Answering research questions, however, about whether or how participants would engage with the data on their own may also be difficult given the context of the data engagement interviews being so different from the circumstances they would encounter if they were engaging with their data alone.

Data engagement interviews are a new and different tool in the toolbox of HCI methods, including those typically used in personal informatics. The most common method for personal informatics research to this point has been more traditional semistructured interviews [15], occasionally also presenting collected personal data. With interviews, researchers and participants are constrained by the representations and analysis that are available, thus limiting the ability to engage with the data interactively in real time. Participatory design methods are another approach to eliciting needs. These methods share an aspect of interactive engagement, where the participant is an expert. However, they also tend to focus on the goal that there is a tool being designed, rather that the immediate task of engaging the data. In contrast, one takeaway from our results was that perhaps designing a tool is not the best approach for this specific user context. It is difficult to imagine arriving at a similar conclusion with a participatory design framing.

Data engagement interviews share some similarities with the think aloud method. Both ask participants to provide step by step explanations of their thought process for accomplishing a task — in our case exploring their data set. However, think alouds involve a participant using an interface without external support or intervention while a researcher observes. Data engagement interviews are distinct from this approach, which requires that participants interact with the analyst and the interviewer, rather than solely interacting with an interface. The standard Wizard of Oz [152] approach is another similar method that presents an interface to a participant, which is actually powered by a human "behind the curtain." In the case of data engagement interviews, no interface has been designed; the analyst *is* the interface, and no clear reason exists to hide them away. Data engagement interviews draw heavily on the success and value of pair analytics [148], but with the key difference that pair analytics requires both the domain expert and the analyst to have similar levels of analytic and computational expertise in order to productively work through a task. Work by Grammel et al. [58] describe a similar approach using commercial visualization software and a dedicated operator to study how visualization novices construct visualizations. Their work uses fictionalized sales data to reduce participants' familiarity or preconceived ideas about data while data engagement interviews are specifically designed to incorporate a participants' own data and leverage their lived experience for capturing how they engage with personal data. The addition of data engagement interviews to the HCI toolbox enables research and data collection with greater flexibility than adjacent methods by eliminating the need for a participant-facing interface or for participant analytic capabilities.

5.8 Limitations

Although we argue that data engagement interviews can afford a more authentic view of people's personal data engagements across a range of contexts, we acknowledge the potential limitations to the ecological validity of our observations. We speculate that participants' data engagement interviews will have lasting effects for how they think of their data and behaviors, yet more work is needed to explore how data engagement interviews might influence long-term behavior change. Furthermore, despite showing evidence that these interviews advanced our broader research goals to help participants learn more about their data, the observations of this work are limited to the specific contexts and circumstances from conducting seven data engagement interviews with a small and specialized user-group. Further validation of this interview framework will require broader application in other situations, and comparison to alternative research methods. We encourage others to use data engagement interviews and explore how this method can be applied in other personal informatics use cases.

Whereas we advocate for a three-person team of two interviewers and one data analyst to improve the efficacy of data engagement interviews, we recognize that this team size may pose challenges from an overemphasized power imbalance through disparities in gender, race, numeracy, or socioeconomic status. Examples in this work include our all-male interview team interviewing female participants on their own, or our participants with limited data analysis experiences directing a graduate-educated analyst. These circumstances may have caused some participants to feel less willing to openly discuss their thought processes or analysis ideas, potentially preventing our observational data from accurately reflecting how people might engage with data on their own, or with one another, outside our study context. We encourage other research teams conducting data engagement interviews to be mindful of the trade-off between interview efficacy and power dynamics, and to consider ways to diversify the interview team as a mechanism to reduce possible imbalances.

Finally, conducting data engagement interviews benefits from having a trained data analyst. This suggestion can pose challenges for conducting this method in more remote or underdeveloped locations where it may be difficult to find a suitable candidates to fill this role. As a result, this approach may not scale as broadly as other interview methods or deployed analysis tools. Further research could explore whether remote collaborations are a suitable replacement for an in-person analyst.

5.9 Conclusion

This chapter presents the data engagement interview: an interview method that supports people to deeply engage with personal data. The data engagement interview strikes a balance between the flexible, lightweight user engagement approaches that do not incorporate personal data, and the more custom, heavyweight analytic tools requiring significant design overhead. We outline a general framework for conducting these interviews, and present a case study from performing seven data engagement interviews with our study participants. We speculate that data engagement interviews can be extended beyond working with personal air quality data, and be applicable across many different personal informatics domains. For future work, we are interested in conducting data engagement interviews in other contexts, and to continue mining our rich interview results for designing future tools and techniques that support people in analyzing indoor air quality data.

This chapter also reflects the central theme of this dissertation. After finding a lack of guidance for engaging people with personal data, the absence of suitable methods within the personal informatics and visualization communities motivated us to develop and conduct the data engagement interview. Developing this interview method then led to a closer look into why these fields had not offered such guidance. This targeted literature review resulted in us identifying and defining the personal informatics analysis gap. The next chapter focuses on what we can learn when engaging people with personal data, and presents our results from analyzing participants' data engagement interviews in this context. In addition to these results, we also offer recommendations for conducting visualization design research with the personal informatics analysis gap in mind.

CHAPTER 6

RESULTS AND DISCUSSION

This chapter presents the results from observing how participants engaged with their indoor air quality data using the data engagement interview method. We report how participants shared diverse goals and questions (Section 6.1.1), exhibited similar patterns when exploring their data (Section 6.1.2), and engaged with their data in playful ways (Section 6.1.3) that frequently led to serendipitous discoveries (Section 6.1.4), yet remained reluctant to analyze data on their own (Section 6.1.5). The chapter ends with a discussion of what these results can mean for designing future visual analysis tools and systems (Section 6.2).

These results extend our data engagement interview findings (Section 5.6), and emerged from the same analysis (Section 5.3.2). We supplemented this process with critical feedback from our research group on the proposed implications for designing within the personal informatics analysis gap, collected during practice presentations and research pitch meetings. Based on this feedback, we conducted an additional round of affinity diagramming using participants' analyzed transcripts to arrive at the results and discussion points presented in this chapter. Our discussion points similarly evolved through several conversations between researchers while we began to formalize the personal informatics analysis gap [1].

In our discussion of these results, we propose three design recommendations that contribute to the core findings of this dissertation. We recommend that visualization designers focus on developing entry points into personal data (Section 6.2.1), prioritize play as a first-class design criteria (Section 6.2.2), and reconsider normative tool development in favor of more social and collaborative systems (Section 6.2.3). This shift away from goal-oriented and task-based design toward more playful and ludic systems can transform how the visualization community approaches supporting people to flexibly analyze their personal data.

6.1 Results

This section presents our results from analyzing participants' data engagement interviews with a focus on what new things we can learn about their goals and workflows using this method. These results illustrate the varied ways that our participants approached, analyzed, and engaged with their personal data.

6.1.1 Diverse goals and questions

All our participating families received identical hardware, collected the same types of measurements, and had the same interest in improving their families' respiratory health. Yet, despite these similarities, the specific questions our participants brought to their data engagement interviews, and the types of insights they acquired, were surprisingly diverse.

Participants' data analysis goals covered a myriad of topics:

> **P2:** What's going to be the best vacuum cleaner to help keep dust down?
>
> **P4a:** I'd like to compare my data with other people's data to see how people were affected by outdoor air quality, near me or in the valley.
>
> **P5**: Can I anticipate days I shouldn't go outside?
>
> **P6:** Should I move?

The analysis approaches the participants took were equally diverse. For example, P1's interest in reviewing her air quality data stemmed from wanting to understand whether periods of poor outdoor air quality impacted her indoor air quality. These discussions unfolded as she and her husband jointly explored data together, but became a broader conversation on indoor and outdoor air quality dynamics after witnessing how their personal activities impacted their living spaces. In contrast, P3 and P4 each chose to step through their air quality data spike-by-spike to see what annotations were associated with the most prominent outliers. These explorations evolved into discussions about the proportions of spikes associated with particular activities or locations. Reframing their air quality events by faceting on their underlying properties changed the way both participants came to see their indoor air quality. Finally, in a third example, P4a was less curious about her air quality measurements and preferred to compare her weekly health survey responses against those of other participants. This comparison served as a mechanism to determine whether her respiratory issues were more likely triggered by widespread environmental conditions – which she presumed would equally affect other participants – as opposed to personal behaviors, and affect only her.

Even though we collected questions and goals from only seven participants, their diverse personal interests cover a wide range of analysis needs. The questions require integrating data over different locations (*Q: How does outdoor air quality compare across the valley?*); timescales (*Q:What tends to be the worst time for my indoor air quality?*); participants (*Q: How does my indoor air quality compare to other participants?*); and data sources (*Q: Can I correlate my air quality data to my health surveys?*). This diversity in analysis needs poses a significant challenge for designing a tool that is capable of supporting them all.

6.1.2 Pattern of exploration

Despite the varied data streams and integrations that our participants' diverse questions demanded, we *did* see a generalized pattern in how they explored their data. This pattern was consistent regardless of their underlying motivations or enthusiasm, and often first appeared after we brought up an overview of their data early in the interview:

> **Interviewer**: What would you like to start with? What's the first thing you want to see?
>
> **P2:** I'd like to just ... okay, let's glance at this great, big, huge, orange [spike]. Did we mark what that was? If so, what was it?

Large and visibly prominent spikes were a common distraction when reviewing participant data, and one of the most frequently requested features to explore:

> **P3:** I think I would start with the bad peaks. I mean like this [spike] right here, this [high magnitude spike]. Some of these higher ones.

Other outliers in the data also piqued participants' interest, such as those in their self-tracked health surveys:

> **P6**: I'm really looking for any outliers in the [survey] results... Let's look at that big outlier on, I guess, in January.

or memories of notable air quality events:

> **P1**: It might be easy to jump to a day like that, where we know 4^{th} of July is going to have fireworks.

Seeing these outliers in their data motivated participants to pause or modify their initial interview goals in favor of investigating the underlying events that caused them. Throughout our interviews, participants would readily speculate about what might have led to the data feature. For P3, this involved hypothesizing the role that local external air quality events might play on her indoor air quality:

> **P3:** That [spike] was a huge fire in Spanish Fork. You could smell it everywhere, it was terrible.

P5 used his data to question whether the air quality impacts detected by separate indoor air quality sensors might have extended to other locations in his home:

> **P5:** That's pretty interesting. [These spikes] make me wonder if it was all through the house, the fact that they're pretty similar.

Participants then asked to contextualize the data features with additional data streams to help validate their speculations or better understand what they were seeing. For example, both P2 and P3 wanted to use their text annotations to understand which behaviors were influencing the indoor air quality spikes they noticed:

> **P2:** I'd be curious to see what our annotations are on the highest spikes... think you can do that?

> **P3:** Being able to match my annotations with the larger spikes would be helpful to find patterns.

For P5, however, reviewing his data uncovered unexpected late-night spikes that emerged during the last week of his deployment, and he requested a different type of contextualizing data stream:

> **P5:** Can you overlay the temperatures? Outside, like the outside temp?

P5's interest in incorporating his outside temperature measurements was a proxy for determining whether these periodic spikes may have been due to his furnace kicking on and off during cold temperatures.

As we analyzed participants' interviews, we began to recognize a recurring pattern regarding the sorts of data elements that caught their attention, and their method of investigating these elements. The participants would: 1) discover a prominent feature in their data, either through visual inspection or reflecting on their experiential knowledge; 2) determine if the feature warranted further inspection, and if so, attempt to correlate it with features in other data streams; and 3) speculate about underlying factors and potential behavior changes.

In her data engagement interview, participant P3 distinguished between health- or air quality-motivated workflows, yet ultimately described the same exploration pattern:

> **P3:** I mean this is an air quality thing, right? So if it were me, I would start with the air quality. When am I having a spike? I would then look at the spikes, and then from that I would try to correlate: On those days, what happened to cause a spike? Was anybody ill? Did we have an asthma attack? And try and do that. I think that's the direction I would go, because I'm thinking of it as an air quality thing. If it were sick thing [where] I'm trying to make [my son] healthier, I think I would start with his asthma data, and then go the other direction. So I think it just depends on which approach I would take.

This exploration pattern naturally emerged in all of our participant interviews, regardless of their prior preparation or motivation, and independent of whichever particular feature they chose to enter their personal data. Air quality measurement spikes, key dates, personal associations or memories, text annotations, and survey response outliers were the most frequent subjects that drew the participants' attention in their interviews.

6.1.3 Playful engagements

Our longitudinal study enrolled parents struggling to control the symptoms of severe asthma in their children and themselves. They participated because of their hopes that data about the air quality in their homes could improve the health of the asthmatics in their households — a serious undertaking with clear implications for the health and well-being of their families. Nonetheless, their engagements with the deployed system, and with their data, led to numerous moments of play.

For example, some of the participants' self-tracked annotations captured how repetitive data entry gave way to playful antagonism toward their deployments and data entry. In our interview with P4a, she admitted to her increasingly deprecating annotations about her mom:

> **P4a:** Towards the end I think I got a little more joke-y with it – "oh you know, mamma's been cooking again." Before it was "mom burnt the chicken." Now it's "nothing new is happening now!"

Similarly, P1 explained why some of her annotations personified the deployed system's alert mechanism:

> **P1:** We did start putting snarky remarks in some of the comments [laughs]. You probably noticed! There was a while where my oven had burnt pizza on the bottom, and every time we turned the oven on [the alert] was like ”HEY!! HEY!! HEY!!!” And I was just like “yep, still haven’t cleaned my oven! You want to come clean my oven? ’Cause I still haven’t cleaned my oven.” [laughs]

Cooking annotations were prominent in each participants’ dataset and occupied the majority of discussion around household behaviors, with a special focus on bacon:

> **P2:** That’s me, burning the bacon.

similarly, for P5:

> **P5:** Oh, and you have where I put in that we were cooking bacon: “Me cooking bacon.“

and also P3:

> **P3:** We did the same thing over and over again. Bacon. [laughs] I think that 90% of our annotations are probably bacon. I like bacon! [laughs]

P3 was especially amused to see how often bacon was present in her data, bringing it up multiple times throughout her data engagement interview. These amusements turned into deeper engagements with her data:

> **P3:** I’m looking at [my data] and there’s a ***lot*** of bacon ... If I notice that every single spike is because we’re cooking bacon, then I might think, is that a problem? What is it that cooking bacon puts in the air? Is that a bad thing, or, the smell’s a good thing, right? It’s a good thing.

Unlike most of the participants, P2 was much more difficult to engage in the data review process due her lack of confidence in her analytic abilities, and a stated preference for having medical professionals interpret her information. As the most severe asthmatic in our study, P2 was also overly cautious about making any behavioral changes due to the perceived consequences of incorrectly interpreting her data. These facts made it all the more surprising when P2 *also* got drawn into her data through exploring her spikes and cooking annotations. Like others, she, too, began cracking wry jokes about burning food and cooking bacon:

> **P2:** That’s me, burning the bacon ... I’m real good in the kitchen, I can tell you that!

Every one of our data engagement interviews captured moments of play. Participants’ interpretations of what they were seeing in their data, filtered through their self-deprecating, sarcastic humor, revealed insights into the challenges of balancing health with other priorities:

> **Interviewer:** What's interesting for you here?
>
> **Husband:** ...my painting
>
> **P1**: You were spraying [figurines] on the 23rd... Being a geek is hazardous to your health!
>
> **Husband**: Sorry, had a D&D coming up.

These exchanges illustrate the overall playfulness we witnessed when engaging people with their data. This sense of play that emerged when exploring personal data highlights a broader dynamic within each participant's deployment. Our existing rapport, plus the potential for seeing their data in new ways, made participants excited to dive into their data and learn new things. This excitement manifested itself differently for each participant, but a playful sense of curiosity helped pull people into their data and unpack what they had to show.

6.1.4 Serendipitous discoveries

Participants' concurrent enrollment in a national asthma study [2] and our own visualization design experience primed us to assume that they would be goal-oriented when it came to engaging their data at each stage of the longitudinal study. Much of what we observed in the data engagement interviews, however, was productive free-form exploration, often facilitated by playful engagements.

For example, P3 used her annotations as a way to self-experiment with her cooking habits to uncover which kind of cooking oil produced the fewest spikes during her deployment [3]. During her data engagement interview, however, her playful interest in bacon gave way to a broader exploration of her cooking habits. This exploration led P3 to a different view of activities that affected her indoor air quality:

> **P3:** I remember making the connection between the olive oil and the [spikes]. And I also knew that it was kind of every time we cooked bacon there was a [spike]. But I guess I didn't realize how many of them, overall, were actually cooking episodes... Like "cooking pancakes," "cooking eggs," "[my daughter] burning the tortillas." It's all cooking.

Most other participants also experienced similar realizations after idly exploring their data. P4 came unprepared for her interview without considering what she wanted to explore ahead of time. Yet, when reviewing her data during the interview, unexpected spikes caught her attention:

> **P4:** It seems like most of [the spikes] are actually in [P4a's] bedroom, which surprised me.

In contrast, P5 came to the interview with the goal of finding connections between his indoor air quality and respiratory health. However, unexpected spikes distracted him:

> **P5:** Oh, and I have no idea what would be in the room making it that high. Why would there be a spike in the bedroom, and not downstairs?

Following this discovery, he proceeded to spend over half of the interview attempting to find possible sources of his mysterious, nightly indoor air quality spikes. P2, who was the least willing to engage with her data, was also drawn in by reviewing spikes and their annotations. She saw that many of her air quality spikes were from cooking, and she became unexpectedly invested in understanding the extent of cooking spikes in her data:

> **P2:** I bet that [spike] is cooking, too. If it's not I'm going to be surprised.

The data engagement interviews gave participants the time and space to stumble upon unexpected and surprising observations, often leading to new insights. Despite our attempts to conduct goal-oriented analysis — by explicitly priming participants ahead of each interview to discuss their analysis goals — the most productive outcomes came from serendipitous discoveries.

6.1.5 Reluctance to personally analyze

Participants' levels of engagement during their interviews — and throughout the longitudinal study — were as varied as their questions. Some participants enthusiastically engaged with their data, some were reluctant but still tenacious, and others were difficult to motivate at all. Regardless of their level of engagement, however, every participant stopped short of advocating for a tool that would allow them to analyze their data in similar ways as the data engagement interview. We asked each participant how likely they would be to use an idealized tool that offered the same flexible features. Other than the teenage participant (P4a) who felt she "might use it," all other participants were pessimistic. P1 was too busy:

> **P1**: As a busy mom with small children, I just don't have the time.

P6 was less motivated:

> **P6**: I could do it, I just don't know how often I would.

and P2 preferred someone else do the interpretation, entirely:

> **P2**: I don't know that I would ever just pull it up and look at it for data's sake... I can take it back to my doctor.

Participants' reluctance to engage with personal data is consistent with findings in other informatics disciplines. P2's preference for having her personal health data interpreted by medical professionals echos similar findings in other chronic health management research [177], and recent work shows even Division 1 collegiate athletes, with access to vast stores of personal data and dedicated analysts, also resist engaging with their data for various reasons [178]. Both these contexts bear a resemblance to our own participants' circumstances as asthmatics living in a region that experiences some of the worst air quality days in the world [179], lending evidence to broader, more complex challenges.

6.2 Discussion

This section reflects on the results gathered from participants' data engagement interviews, and what these results can mean for future visualization design research. Based on these outcomes, we present three design recommendations for how researchers can further understand and bridge the personal informatics analysis gap: designing for entry points into personal data, prioritizing play as a first-class design criteria, and reconsidering normative analytic tool design in favor of more social or collaborative systems. We offer these ideas as a starting point for thinking differently about how we design for personal contexts, and to highlight the opportunities for the visualization community to make an impact in this space.

6.2.1 Design for entry points

In Section 6.1.2, we discuss the common exploration patterns that we observed throughout the data engagement interviews. Participants often fixated on the same kinds of visually prominent features within their plotted data — like spikes or outliers — as well as performing the same exploration tactics when engaging with their data, in ways that bear a resemblance to established sense-making process models of intelligence analysts [40]. Because participants shared these similarities regardless of their prior preparation or general interest in reviewing their data, designing for these tendencies could provide valuable insights into ways designers could make interfaces more engaging, especially in personal contexts.

These engagement and exploration patterns are examples of what the design community calls *entry points* [159], which are "a point of physical or intentional entry into a design" [180]. Rogers et al. identify the utility of entry points for interactive interfaces, recommending that designers incorporate them into their designs as a way to "think about the coordination and sequencing of actions and the kinds of feedback to provide in relation to how objects are positioned and structured at an interface" [181]. As an invitation to action, entry points bear a resemblance to affordances [182], [183]. In the same way that affordances indicate potential interaction mechanisms to users, entry points provide hints for what can be done with engaging data [184].

As a concept, entry points describe an intuitive way for inviting users into a system or interface. Google Maps, for example, uses a device's location to generate an initial view as a way to experience its interface. Recent visualization tools explicitly implement entry points for investigating how people explore interactive visualizations [185], as well to facilitate rich conversations and exploration, encourage engagement with data, and make visualizations or datasets feel relevant to a wider audience [186]. These approaches serve to help orient and engage people with unfamiliar tools or data sets [187], but the visualization community has largely overlooked entry points as a formalized construct to engage people with their personal data.

Existing theoretical work on entry points in the design community details how to design entry points into a system. The design literature recommends establishing "points of prospect" to give an overview of the different ways people can engage with their data, and "progressive lures" to incrementally bring them into their data [180]. This approach can lower barriers to entry and invite progressively deeper inspection. Data engagement interviews [4] can also benefit from considering entry points as a design element by helping to prioritize which analytic capabilities researchers should support.

Human-computer interaction research into entry points can also help with determining the kinds of entry points different types of personal data may support. Tailoring interactions that reflect how people think of a particular data source may help lure them in to freely explore, and improve their overall chances for finding interesting parts of their data. For example, Choe et al. support basic temporal cuts for reviewing logged physical activity data within their Visualized Self web application [147], which users in their study found helpful and compelling. Time Lattice [188] generalizes this concept to support interactive analysis and comparisons for large-scale and distributed sensor data over a broader set of user-selectable constraints.

Entry points like these can allow for more meaningful exploration of self-tracked data in the context of supporting comparisons for how personal data vary across specific constraints or conditions, whether these are temporal (*Q: What time of year is typically worst for air quality?*), geographic (*Q: What part of the state has the best air quality?*), relative (*Q:What's the air quality like when other people are sick*), absolute (*Q: How many days were above the red air quality cutoff?*), or a mixture of these, and more. Systems supporting these approaches stand to better reflect how people think of their data and their lives, especially when compared to the static, linear time series plots typically provided for reviewing personal data.

The ability to seek or see one's self in data plays a significant role how people engage with and experience visualizations [189]. More generally, any guidelines for developing engaging visualization systems will require a better understanding of what compels someone to start exploring visualizations of their data. As a concept, entry points offer a promising approach to help address the shortcomings behind the personal informatics analysis gap. Understanding what users find interesting or engaging in their data can help designers lower barriers and improve usability by identifying compelling entry points into those datasets or visualizations. Furthermore, researching entry points in personal contexts can help formalize ways people engage with personal data, along with common priorities, interests, or data characteristics they find especially relevant across different use-cases.

6.2.2 Design for play

We approached participants' data engagement interviews with a goal-oriented mindset, expecting they would come prepared with their own goals and approaches for analyzing their data. We asked participants to think about what they wanted to know about their data ahead of their interviews, and we structured the start of the interview to probe their specific questions. Despite our attempts to prime our participants, some were disinterested in defining a goal and digging into their data, as detailed in Section 6.1.5.

This lack of enthusiasm may have posed problems for more traditional, retrospective think-aloud interview methods, but the opportunity to directly engage with their personal data in the data engagement interview led each participant — even those who *were* prepared with analytic goals — to quickly distract themselves in playful ways as we guided them to examine their data. As we discussed in Section 6.1.4, participants preferred to freely explore their data, but not in the ways the visualization literature describes data exploration. Brehmer and Munzner's task typology [170] defines exploration as "searching for characteristics," although explorations of our participants had less to do with clear searching objectives, and more to do with stumbling into serendipitous discoveries. We had attempted to promote a goal-oriented experience, but this playful and open-ended exploration was what kept people engaged with their data, and what motivated them to proceed through their interviews.

This type of playful engagement is traditionally overlooked in visualization research in favor of more goal-oriented behavior. The visualization community's focus on goals perhaps comes as a consequence of a decades-long framing of visualization as a vehicle for cognitive amplification and insight generation [168]. Fun and enjoyment are secondary considerations, if they are considered at all. Prior work on ways to design for fun further codify this prioritization with recommendations to consider fun and enjoyment only after "the functionality and usability have been accommodated in the design"[169]. Our findings provide evidence that people engage with their data in productive ways through play, and lends support for prioritizing play as a first-class design requirement.

Play, and the dynamics that give rise to or influence play, have been studied at length in human-computer interaction and design fields. Prior work shows people are likely to engage in activities seen as fun or playful, such as interactive filters on social media [190], or experiences that elicit curiosity from their ambiguous or unspecified outcomes [191],[192]. Designs that prioritize play and entertainment lead to more engaging experiences [193], as do interfaces that are designed through processes that include situated play design [194], ludic design [195], and playification [196]. Bertran et al. describe how *situated play design* can inform early aspects of a design process by identifying design targets and inspiration based on their capacity for playful experiences [194]. This design approach also recommends engaging users within these contexts to explore how their play is integrated with their activities, and to use these dynamics as design targets that can be further refined with iterative prototyping [197]. Elements of ludic design similarly recommend promoting curiosity and minimizing externally defined goals as a way to embrace ambiguity and break away from the requirements of meeting participants' every need [195]. Playification [196], [198] advocates designing for playful and intrinsically compelling experiences over more extrinsically gamified elements, like scores, that apply game mechanics to nonplayful activities [199]. Our observations that participants were naturally given to playful engagements and open-ended exploration when engaging their data, and reluctant to engage with an extrinsically goal-oriented analysis tool, resonate with the motivations for play-based design processes.

We speculate that characterizing people's naturally playful behaviors can help inform design elements for developing more engaging experiences in personal contexts. This focus, however, requires that the visualization community re-evaluate its goal-oriented design bias and explore what playfulness means within personal contexts. What is fun in the context of exploring personal data, and what does it mean to make something fun *enough* to engage users? This framing highlights alternative motivations that prioritize fun and enjoyment as first-class design criteria.

6.2.3 Reconsider designing tools

In our previous work, we provided evidence that data engagement interviews were productive and allowed our participants to learn new things from their data [4]. In Section 6.1.3 and Section 6.1.4, we further validate this interview method with additional evidence on how people were able to playfully engage with their data and explore in open-ended ways to arrive at serendipitous discoveries, regardless of their level of preparation or interest. We also described in Section 6.1.1 that the interviews exposed a set of diverse goals and questions that our participants had, complicating the design space for potential visual analysis tools. Furthermore, in Section 6.1.5, we detail how the majority opinion among our participants was that they were unlikely to analyze personal data on their own, from a lack of time (P1), interest (P6), relevance (P3, P4, P5), or confidence (P2). Finally, although we succeeded at engaging participants with their data, we were unsuccessful at focusing participants' data engagements toward goal-oriented data analysis.

Instead, efforts to encourage goal-oriented review fell apart, with participants engaging in open-ended exploration once they got into their data. This finding suggests that any tool designed for our participants would need to be very flexible, but not overly complicated; and even if we developed such a tool, it remains unclear whether our participants would be motivated to use it.

Designing tools to support people engaging with personal data is a difficult and time-intensive process, and requires deep, contextual knowledge for reaching reliable or meaningful interpretations [20]. These requirements raise doubts about how a designer can expect to know what is important within people's data or the ways it can relate to their diverse goals. These concerns are further compounded when any applied analysis may inadvertently obscure or remove information that users may find relevant or important [200]. Furthermore, if designing for personal data is difficult from its reliance on personal knowledge, and the people who have this personal knowledge are the same ones who – for various reasons – cannot or will not afford the time to engage with their data, where does this leave us? If asthmatics living in an area that experiences some of the worst air pollution in the world [179] cannot be motivated to engage with their personal air quality data, what hope do we have of mobilizing *anyone* in open-ended personal visual analytics?

We question whether it makes sense to expend our time and resources to develop potentially complex and design-intensive tools for small user groups that often struggle to translate these tools to other contexts. Taken to extremes, a truly generalized analysis tool capable of approaching these requirements runs the risk of reinventing Tableau or Excel — an overwhelming design proposition in its own right. Given the diverse breadth of participants' questions throughout our longitudinal study, we pose the argument that any standalone visual solution may be impractical or impossible to provide.

In spite of these challenges, our data engagement interviews were especially productive at engaging people in exploring their data. This observation leads us to ask: What if the solution is not another tool, but something wholly different? What if the antidote to increasingly sophisticated and customized visual analysis tools is an investment in communal systems or infrastructures that allows people to cooperatively share their data with dedicated data experts? These social systems could help offset the intellectual burdens of traditional visual analytics systems and enable those experienced with sophisticated tools to collaborate alongside people with personal data to leverage the strengths of each community: computation and context. This idea is not new, and previous work has outlined some logistics and challenges of this approach [201], yet more work is needed to understand ways to attract and sustain professional attention; motivate people to volunteer their time and expertise; and manage resources, systems, and structures that could make this a reality.

We cannot design new ways to support people engaging with their data using the same thinking that created the personal informatics analysis gap in the first place. Instead, this gap offers an opportunity for the visualization community to develop new methods for exploring the gap and new designed futures to bridge it.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

This dissertation identifies and defines the personal informatics analysis gap through a review of personal informatics and visual analytics literature, along with our own experiences working with asthmatics over a multiyear longitudinal study. The gap describes a lack of attention to flexible analytic systems for supporting people to engage with their personal data, and how it can be addressed by coordinating the individual research strengths of personal informatics, visual analytics, and everyday visualization fields. Raising awareness on this gap, and the ways it affects visualization and personal informatics researchers, identifies a previously untapped research space with a wide range of design opportunities for the visualization community, and one that stands to be approached from multiple perspectives.

To help researchers gather information on how to design within the personal informatics analysis gap, we introduce the data engagement interview. This interview method allows visualization design practitioners to better understand the motivations and abilities of end-users, and can help to streamline requirements elicitation within a design process. We also provide three recommendations for how designers can create solutions that help bridge this gap: designing for entry points into personal data, supporting playful engagements with personal data, and reconsidering analytic tools in favor of more social and collaborative systems.

As we move forward with designing for the personal informatics analysis gap, these next steps will require new thinking and approaches compared to existing methods. Broadening design thinking beyond rigid, goal-oriented tools to consider alternative priorities, such as designing for exploration, play, and more collaborative systems, can leverage the skills and strengths of analytic professionals to help people get the most from their data. Acknowledging the personal informatics analysis gap exists — and taking steps to explore it — is an important first step in addressing the present limitations for gaining deeper insights into what people want to do with their data. This dissertation calls to the visualization community to explore the personal informatics analysis gap through careful, qualitative work that unpacks personal data engagement, and to explore design opportunities that challenge normative visualization design.

The visualization community can learn from existing design knowledge and literature from communities already familiar with entry points [159], [184], [187]. One important first step can be to conduct a thorough survey of existing personal informatics tools to better understand and characterize the design space for incorporating entry points into personal data visualization. Coding for commonly supported interaction methods can identify common entry points into personal data and provide insights on understudied methods for inviting people into their data. More targeted research can also explore how specific or preferred entry points vary across different research domains, types of personal data, or whether traditional analytic workflows or could inform analogous entry points into personal data.

We present data engagement interviews [4] as our solution for gathering observational data on how people work with personal data in the context of logged indoor air quality. We speculate this method can transfer to other use-cases, however, and are excited to see how this interview can be applied in future work. In our own work, we are interested to apply the data engagement interview method within other research contexts, as well to continue mining our rich interview results toward designing future tools and techniques that support people with analyzing indoor air quality data. For other researchers, this interview method can be further validated by applying it within other research domains, use-cases, and sources of personal data. Examples of this work can include investigating how well data engagement interviews can engage different user populations, or adaptations for processing alternative data sources such as medical or financial data.

One broader adaptation may come from how visualization designers choose to develop future personal data analysis tools. Growing interest in value-sensitive design [202], [203] concerning explainable artificial intelligence [204]–[206] and human-centered data work [207]–[209] can provide opportunities to incorporate and leverage aspects of these fields that improve interactive analysis and data legibility, and offer productive, transparent, and collaborative personal data analysis for everyday people. This method of interdisciplinary work can further our understanding of ways applied visualization research can assist with developing next-generation solutions for bridging the personal informatics analysis gap.

APPENDIX A

FIELD DEPLOYMENT INTERVIEW PROTOCOLS AND THEMES

This appendix includes the semistructured interview protocols for each of our field deployment phases and later engagement interview (Sections A.1, A.2, A.3, A.4). We also include the list of themes from each deployment phase, derived from our qualitative analysis of participants' interviews (Section A.5).

A.1 Phase I interview protocol: Predeployment interviews

Phase I of our deployment interviews involved the following semistructured interview questions. This was our first meeting and introduction to our study participants and we were interested to get a baseline of what it was people knew and thought about air quality.

General Introductory Questions

1. How important is air quality in your life Is it something you have given much thought to? If so, describe.
2. Do you know of (or suspect) any activities or conditions that trigger respiratory problems for you or you child?

User-awareness of air quality and pollutants

1. Do you regularly check air quality?
   (a) Do you or other family members check publicly available air quality information resources (government air quality websites, Salt Lake City air quality index, phone apps, etc.)?
      i. Describe circumstances that would motivate you to check the air quality.
      ii. What information do you use (ratings, scientific measurements, warning levels, etc.)? How do you use this information?

iii. Have you ever modified an activity or taken precautionary measures as a result of air quality concerns?

A. *If PRISMS deployment has already occurred*: Other than the PRISMS sensors, do you have any other devices or strategies for monitoring the air quality in or around your home?

B. *If PRISMS deployment has not yet occurred*: Do you measure air quality in or around your home? If so, what devices or strategies do you use?

2. How would you characterize the air quality in different parts of your home?
   (a) What are the causes of these variations?
3. What do you think are the largest contributors to indoor air pollution in your home?
4. Are there any behaviors/circumstances/activities that you think these air quality monitors would help you learn more about?
5. How does air quality outside compare to the air quality inside your house? Is one always better or worse than the other?

User privacy, implications of data collection

1. Do you have privacy concerns about the collection of your air quality data?
   (a) As it relates to your personal, family, or residential privacy?
2. Let's think about how your air quality sensor data might be shared with others...
   (a) Do you consider information about the air quality in your home to be private information?
      i. How comfortable are you with sharing this information with PRISMS program researchers at other sites (including the National Institutes of Health)?
      ii. Are you comfortable sharing this information with researchers who are not part of the PRISMS program?
   (b) How comfortable would you be, if we asked you to share your air quality information with other families that are part of the PRISMS studies?
   (c) Would it be okay if your air quality information was posted publicly online?
      i. *Probe*: We assume you will want names removed - but what about address, or zip code (or something like a geocode - latitude/longitude) - is there any part of the location they would consider acceptable to share publicly?
   (d) Would it be OK if your air quality data were visible to visitors in your home?

A.2 Phase II interview protocol: In-home system deployment demonstration

Participants received their real-time interface in phase II, allowing them to see their indoor air quality data for the first time. This round of semistructured interviews served several purposes: a demonstration of our system and what it could do; a preliminary test of participants' data literacy for interpreting and exploring their data; and to collect feedback on how participants imagined they would use their air quality sensing system.

A.2.1 Protocol overview

Duration: 60 minutes

This site visit serves both as a technology demonstration of our visualization and annotation interfaces, as well as a standard participant interview in order for us to gauge participants' knowledge and comfort with the deployment. For this interview, we should have the primary participants present, along with any older children. Our visit is organized into the following categories:

1. **Interface workshop** - Introduce our tablet interface and show participants their previously collected data,with a focus on having the participants be the drivers (i.e. we demonstrate once and have them do the rest, but help them know what to do). We use this exploration activity as a means of teaching interaction techniques and educating users to navigate their data.
2. **Data Literacy** - While exploring data, use this opportunity to engage participants and discuss/determine their comfortability with and knowledge of what this data represents. Inform them that they should not interpret health risks from any recorded measurements as their system is equally sensitive to water vapor as other irritants. Further discuss time series curves, scales, etc.
3. **Reflection** - After participants are comfortable navigating the interface and are shown (or told) sufficient information on what the data means, proceed with a small exercise (series of data exploration tasks) to have them reinforce this knowledge.
4. **Annotation** - Raw time-series data is not very important without contextual knowledge of what may have contributed to this behavior. Annotation is key to providing

and preserving this context. Here, we introduce the importance of annotation, as well as discussing possible modalities:

(a) Pen & Paper: Offer to leave notebook to support activity capture, spike labeling.
(b) Tablet: Show how annotation works on tablet device.
(c) Voice Annotation: Using Google Home, show how families can annotate via voice commands
(d) Text Annotation: Allow users to send a text and have it show up on the tablet.
(e) Pre-emptive Annotation: Show mini-demo where we generate some type of PM spike and receive a text from PRISMS asking what happened.

Let participants choose which ones they want, and which ones to opt out of.

A.2.2 Interview questions

Our hardware demonstration lends itself to interjecting a handful of category-relevant questions for each of the above sections. These help capture participant's initial reactions, thoughts, and opinions which will be important in our future analysis.

A.2.2.1 Introduction

Today I'm here with some visualization equipment, and we're going to use this to explore your Dylos data and take the opportunity to explain the interface. The interview has a couple of purposes: One is to help you get comfortable with the visualization displays and practice using different settings, and the other purpose is to think through your preferences for data annotation and display. The interview usually takes about 30 to 45 minutes, and you can ask questions or revisit topics at any time, or slow down or skip any questions you don't want to answer.

Hand off tablet interface and together with the participant explain the layout and five main interaction methods:

- Layout: Timeline view, Main view, Legend, subset buttons, and annotation strip
- Interactions: Time-range buttons, click/dragging on the timeline view, toggling monitor time series display, and annotations.

As we explore each of these points, incorporate the following questions:

A.2.2.2 Interface workshop

Questions / Tasks:

1. Brush overview: How many day's worth of data is included in this visualization?
2. Resolution buttons: How would you show the previous week's worth of data, starting from today?
3. Subsetting and brushing: Show a 2-day period period Including Tuesday, July 4th.
4. Legend manipulation: How can you change the plot lines to show only the outdoor monitor?
5. For Interviewers: Demonstrate two ways of using the brush view: touch to re-center on a region vs. click and drag to scrub through overview

As we proceed with data exploration, add in the following questions:

A.2.2.3 Data literacy

Questions / Tasks:

1. What do the peaks in this graph represent?
2. How are the three monitors at your home represented on the graph?
3. What information does the plot line tell you for each monitor?
4. How is time displayed in this graph?

Now we'll have a quick exercise to pose some questions which could come to mind when using this interface:

A.2.2.4 Reflection

Questions / Tasks:

1. What is the highest peak value on the plot?
   (a) What is its value? When did it occur?
2. Can you find a PM spike which occurs outside, but not inside?
3. Can you find a PM spike that is only detected by one of the indoor monitors?
   (a) By all three monitors?
4. Do more peaks happen inside or outside?
5. Looking at the inside monitors, do you recall the cause of any of these spikes?
6. What is the oldest spike you can remember the source of?

Introduce the idea of annotation, explain how it can help aid recall and increase data usefulness. Show how to annotate data on the graph (if not discussed already), and introduce Each method we've prepared for the participant:

- Notebook, texting, tablet, google voice.

Now we have some questions to gauge your interest or preference for each of these methods.

A.2.2.5 Annotation

Questions / Tasks:

1. Notebook: How likely are you to write down your activities or sources of spikes after the fact? [5-point Likert Scale]
2. Which annotation modality seems the most convenient to you? [Alt: rank on likert scale]
3. Which seems the least convenient? [Alt: rank on likert scale]
4. Would you ever choose to annotate an event which did not have a corresponding air quality impact (i.e., washing dishes)?
   (a) Would you be more motivated to capture this information from activities which surprised you (i.e., you expected a strong response and there wasn't, or vice versa).
5. Compared to receiving a text, would you prefer an automated voice prompt from Google Home to tell you that a spike has occurred and to ask for your annotation?
   (a) What would you think about that? Would this be too intrusive? Annoying?
6. We have developed some annotation tools to assist with capturing events. If you chose to use them,how often would you want to be alerted to provide annotations for detected air quality spikes in your home?
7. Does your preferred frequency change if the alert mechanism was done via text vs. a voice alert? Are these different? If so, why?
8. What times of day would be appropriate to receive these alerts? Are they different for text or voice? Why?
9. Which annotation options would you like to receive? Are there any you would like to opt out of?

10. Currently, how likely are you to look at your old data? And how far back are you most likely to look?
11. Would using/having annotations affect this?

Remind Users that they should...

1. Only annotate if they want, and only in the ways that make the most sense / are most convenient for them.
2. Feel free to provide any additional comments on the annotations they make
   (a) expectations of the result, surprise, or any other details they feel are useful to capture for themselves, or for our understanding of the event.

A.2.2.6 Wrap-up

Questions / Comments:

1. Are you surprised by how your data looks? (Is this what you expected it to look like?)
2. Did any initial questions that come to mind when seeing your data this way? (For example: "what are these peaks about?" "Can I see when I cook?")
3. What aspect of your data seems the most interesting to you? Does anything stand out?
4. Retrospective: Now that you can see your past data, is there any time you can think of where you would like to re-visit to inspect? What were you doing and why would you like to see it? What do you expect to see? (return to annotation activity and have them make an annotation?)
5. Anticipatory: Now that you have this visualization tool, are there any future events you anticipate using this to monitor?
6. For you personally, how do you expect to use this system most? And how often?
7. How much old data do you wish to see from before this site visit, if any?

A.3 Phase III interview protocol: Postdeployment interview questions

We conducted Phase III interviews after participants had some time to use their deployments (2-6 weeks). We tailored our questions to understand how they used it, what they learned, and to capture general feedback from their long-term air quality system deployment (20 - 47 weeks).

A.3.1 Introductory questions

1. Have you seen anything interesting using the system?
2. Can you describe how you used the system?
3. Were you able to do what you wanted to do
   - If YES: What did you use it for?
   - If NO: What did you want to do but could not?
4. did you notice any patterns in your data

A.3.2 Interface usage

Data/Pattern Searches

1. Have you noticed any behaviors or activities which reliably produce spikes?
2. Did you ever compared two different sensors?
3. Did you ever refer to the outdoor monitor readings before making a decision indoors (like opening windows?)
4. Did seeing differences between indoor and outdoor air quality lead to any insights, realizations, or surprises?

Tablet/Visualization Interactions

1. When you used the tablet did you use it as an exploratory tool, or did you look out for specific features?
2. Did you find the tablet useful as a data browsing tool? Compared to e-asthma tracker? Would you use such an interface differently / more often if it would be a mobile application? item What time resolution did you use most often to look at your data?
3. What time of day did you tend to use the tablet most? How about time of the week?
4. When looking at data,

(a) Did you do any comparison between rooms?
(b) Did you do any comparison between events (in a single room)?

5. How far back did you typically look?
   (a) Did that change over the duration of the study?

Engagement & Impact

1. Has having the tablet changed your air quality checking habits?
2. Has having this device inspired other people to check air quality?
3. How often did you or others interact with the tablet?
4. Did you have any conversations with family over your air quality data?
5. Did having this tablet make you any more likely to check or engage with your air quality data?
6. Does being able to view sensor data from your home influence your decision on whether to participate in an asthma study (probe – more likely?
7. Less likely? No influence?)
8. Do you think that being able to view sensor data from your home (or "having an interactive visualization") would influence how long you remain in the study?
   (a) Would access to the data make you more likely to "stick with" a long term study?
9. Would your answers to the above be different if you could not see the data in real time (for example, you could see the data from yesterday, but not from today; or only once a week, once a month)?
10. Do you use eAT daily? What information is shown? Is it intelligible? has having access to this information affected your knowledge of indoor air quality or thoughts for this?
11. Did having access to your data, change how you think of your Indoor air quality?
12. Are there any rooms in your home which turned out to be more or less active (peaks) than you anticipated?
13. What events prompted you to check the interface
14. Did any activities have less of an observed impact than you expected?
15. Did you observe anything interesting or unexpected as a result of being able to see your indoor air quality?

16. Have you had a chance to share any of this data with guests? Did they have any reactions? Questions?
17. Have you changed any cleaning, cooking habits as a result of seeing this data?
18. Did you notice any sources or factors impacting air quality in the home that you did not expect?

Annotations: Have the user rate their annotation options again

1. How often did you annotate?
2. Did you annotate as often as you would have liked?
   - If not, what prevented you from annotating more often?
3. Did the annotation feature add value to the system?
4. Did you find the texting option convenient? (And why)
5. Did you prefer to annotate via text or tablet? (And Why)
6. Did you have any thoughts on the usefulness of google home vs. text? (And Why)
7. How did you use Google Home most during the deployment?
8. What annotation method do you think is most useful for you now that you have had time to use them (And why)?

Health Correlations

1. Did being able to see your indoor air quality make you feel more comfortable about your (or your child's) personal health?
2. Did you use the tablet to investigate air quality if you were experiencing asthma symptoms?
3. Did you use the tablet to help make any health decisions or behavior changes?
4. Did you use annotations to keep track of health symptoms for correlating indoor air quality with respiratory health?

Privacy Insights

1. Now that you can see your data for the past few weeks, Has your general feelings towards privacy changed with regard to this system?
2. Have your thoughts changed on what this system is able to detect?
3. Have your thoughts changed on what behaviors or activities you are able to identify through air quality data?
4. How would you define the term "personal data"?

5. How would you define the term “private data”?

User-Interface Feedback

1. Was there anything you wanted to do which you could not due to the current interface design?
2. Do you feel the values shown to you were accurate?
3. Do you feel the system was reliable?
4. Do You have any comments, critiques, or suggestions for the interface?
5. What changes (if any) would you like to make to the interface so that it’s better suited to help answer the questions you have about air quality?

Wrap-up

1. How did you gauge your air quality over this deployment?
2. Do you feel that 2-3 weeks is a long enough time to try this system to get the most out of it?
3. Did you use the system differently during the first week of your deployment versus now.
4. What was the most useful aspect of having this tablet system?
5. How did you end up using the interface most?
6. Do you have any requests or feedback on features or interactivity you would like to have after having a chance to use it for 2 weeks
7. Would you like to keep the system?

Demographics

1. How much of your waking time did you spend at home (on week vs. weekend)?
2. What are the ages and occupations of the people in the home?

A.4 Follow-up engagement interview

We conducted follow-up engagement interviews after noticing how disengaged other family members were compared to the primary participant in each household. For each deployment, We conducted separate interviews with the primary participant and the other family members in attempt to uncover the reasons for this general lack of engagement.

A.4.1 For the nonparticipants

These questions were for our engagement interviews with nonprimary participants.

1. In general, how much does air quality factor into your decision-making process?
   (a) Do you think about it often?
   (b) Do you think about for your personal health?
   (c) Do you think about for your daily activities?
2. How much do you think about the air quality in your home?
3. Could you describe your level of involvement with the PRISMS project?
4. Prior to receiving this deployment, did you feel you had much control over air quality (indoors or outdoors?)
   (a) Has this changed at all with the delivery of the air quality system and its interface?
5. Was the tablet interface easily accessible during its deployment?
6. Did having access to the tablet spark any personal interest or curiosity towards the air quality data?
   (a) If so: how?
   (b) If not: why not?
7. How interested were you to interact with the tablet interface?
8. How do you feel your interface usage differs from your partner's/parents?
   (a) Why do you think that is?
9. If you used it, what are some things you felt the interface did or did not communicate well?
10. Do you feel whether there were lifestyle factors, daily routines, or personal habits which influenced the way you interacted with the interface?
    (a) If so: how?
11. Can you imagine any circumstances under which your level of engagement would have been different?
12. Is there anything that could have been different about the interface that would have changed your engagement with it?

A.4.2 For primary participants

These questions were for our engagement interviews with primary participants.

1. It's been a while since our last interview session – how has your tablet use changed? Are you still using it?
   (a) If so: how?
   (b) If not: why not?
2. In the time since you've been using the system, do you think you've gotten as much out of is as you're going to get?
   (a) If so: what are the things you've learned as a result of using this system?
   (b) If not: What are the things you're still hoping to learn from this system?
3. What motivates you to continue annotating (in text and in general?)
   (a) When you annotate, are you primarily motivated to support your own understanding, or to provide information for the study?
   (b) (for themselves) How have your annotations changed or evolved over time?
      i. Has the language you've used to annotate events changed or stabilized over time?
      ii. Do you find yourself reusing annotations over time?
   (c) (for the study) How would you prefer/imagine your annotations being incorporated or used to improve the PRISMS project or this system, specifically?
      i. What impacts could these annotations have?
      ii. What do you think these annotations could provide to clinicians or anyone reviewing the data?
4. As an expert in your own life and data you've collected, if you were going to share it with someone else (a doctor, etc.), what are the things you would want to communicate and show?
   (a) What is it that you feel you would need help or support to communicate?
5. If this interface was re-designed to focus primarily on your annotations and various ways to explore them, would that be useful to you?

A.5 Field deployment themes and descriptions

This section outlines the derived themes from each of our field deployment phases. We thematically analyzed [124] participants' interviews to identify codes and themes from each phase's interview protocols.

A.5.1 Phase I themes and descriptions: Predeployment interviews

Phase I interviews focused on participants awareness of air quality, the role it plays in their lives, plus their attitudes or sense of agency toward managing their indoor air quality.

Personal identity with respect to air quality (e.g., agency): These high-level theme categories capture how people internalize both their relationship with air quality and any implications the data might connote, i.e.:

- Acceptance of status quo: the tendency of participants to accept circumstances borne either from social structure (family dynamics, social roles) or asthmatic symptoms.
  - Lack of agency: Captures the consequential lack of control that comes from relying on third parties to provide information, resources, or decision making.
  - Submission - Captures resignation some participants feel towards their own symptoms and necessary lifestyle changes.
- Air quality impacts on sense of self - This covers the participant's relationship with their data and how what the data says can become instrumental to how participants see themselves.
  - Air quality & identity - Captures how collected health data can impact participants self-esteem and identity.
  - Social status and health - Health tracking can be used as a status symbol for participants who choose to align their personal identity with their measurement
- Sentinel - A shared theme between "Acceptance of status quo" and "air quality impacts on sense of self", capturing some participant's identity as sole caregiver in the immediate family, either by active choice or lack of engagement from others.

Position towards data: These high-level theme categories capture participant's thoughts and feelings towards the air quality data collection process. This includes perceived implications of personal privacy, data ownership, and characterization.

- *Privacy*: Details participant's data collection associations regarding what is useful, applicable, personally identifiable, comfortable to share, etc.
  - Data are uninteresting for nonasthmatics: Feelings of anonymity arising from the belief that the specialized interest in their data limits the scope of its transmission
  - Privacy concerns: Listed concerns relating to personally identifiable information and distribution
  - Non-concerns: Listed non-concerns relating to what participants were willing to share or make available for others
  - Invasiveness: A distinction made which affects the public perception of a technology being 'intrusive'.
- *Ownership*: personal agency and stake-holding associations towards data collected through a scientific study.
  1. What is personal? - Distinction made on what qualities of data make it personal, but not necessarily private.
  2. What is private?: Distinction on what qualities of data make it private.
  3. Data ownership: Feelings of ownership over data collected as part of a scientific study.
- *What data says about me*: This captures participant's awareness/association with what the collected data represented.
  - Privacy literacy: Collected opinions of what participants thought was – or was not – possible to know from their collected data
  - Data abstraction: Opinions of what their collected data represented, ranging from personal activities, to more general notions of geographic variations.

Mental model / assumptions about air quality: These high-level theme categories capture participant's internalized views of air quality, ranging from ways outside factors influence air quality, to ways air quality influences participants

- *Experiential reflections* - this theme collects participant's air quality attitude as a result of their lived experiences
  - Realizations: Participant's air quality insights and associations.

- Questions: Open questions relating to lifestyle and behaviors as they impact Indoor air quality, or the health of immediate family members
- Awareness of air quality Impacts - Reflections on ways air quality affects health and lifestyle

- *Presumptions About air quality impacts* - Participant associations towards factors influencing air quality
  - Air Quality Associations: a list of characteristics which are suspected to imply causal air quality impacts (i.e., better airflow = better air quality, older homes = worse air quality, etc.)
  - Sources, Activities, Characteristics: a larger mid-level theme collecting factors participants felt influenced air quality, including: Sources (dust, pets, etc), Activities (cooking, opening windows, etc.), Characteristics (construction techniques, indoor air flow, etc.)

Tech feedback: These high-level theme categories represent participants' feedback on the air quality data logging system. This includes commentary on the Dylos monitors as well as requests for a variety of comparative data analysis techniques and visualization methods.

- *Interactions on data* - Requests for various interaction capabilities including comparison, exploration, automated active feedback, etc.
- *Visualization design* - Request for specific visualization attributes to improve data context and user understanding (legends, descriptions, etc.)
- *Hardware feedback* - Specific feedback on the Dylos monitors
- *Data of interest* - List of air quality measures which would be useful to include in future interface versions (pollen, mold, particulate matter, etc.)
- *Mobile devices* - Participants referred multiple times to interacting with visualizations, data through a mobile device.

Altruism: This theme captures a common motivation for participants to be involved in this study. The following mid-level themes detail facets of their participation:

- *Service of science* - Participants were eager to collect and share data in support of efforts to learn more about air quality and to help improve quality of life for other asthmatics.

- *Help others* - Participants were interested to share data in the hopes that it would be of use to other researchers, and their sharing could benefit the asthmatic community in general.
- *Accommodation* - Participants expressed the desire to use this data to help make life easier for family members and guests.
- *Social sharing* - Participants were interested to share and compare their own data with friends, family, and other participants.

How air quality impacts life: These high-level theme categories capture how asthma affects participants' lives. This ranges from precautionary regimens (daily air quality-based planning, symptom mitigation & management) to increased communication between family members and schools, over and beyond that of the general public.

- *Symptom management* - Methods used to prevent, mitigate, or treat asthmatic symptoms.
- *Daily checks* - Routine air quality checks to verify forecasts, inform participants, and motivate precautionary behavior.
- *Air quality-based planning* - Ways and reasons participants revised their daily schedules based on forecast or suspected air quality conditions.
- *School oversight* - Increased reliance on schools to serve in loco parentis to protect the health of school children.
- *Motivation to check air quality* - Circumstances in which participants are reminded/motivated to check air quality.
- *Motivation to Learn* - Circumstances motivating participants to learn more about air quality.
- *Motivation to participate* - Circumstances motivating participants to participate in this study.

How to get air quality information: These high-level theme categories compile the methods and resources participants use to gather their air quality data

- *Gathering methods*- Methods used to collect data
- *Air quality resources* - Resources used to collect data

A.5.2 Phase II themes and descriptions: In-home deployment interview

Phase II themes summarize participants' reactions to seeing their air quality data for the first time, what they imagined wanting to do with this information, and ways they imagine using their interface.

Reactions

- Seeing or reviewing air quality data validates existing participant beliefs
- Sharing indoor air quality expectations
- Participants surprised by their indoor air quality
- Annotations are valuable
- Annotations increase system usefulness
- Participants curious about system responsivity
- Participants eager to annotate
- High user engagement with data

Data meaning and value

- Outdoor air quality contextualizes indoor air quality
- Recorded data needs more context
- Memory fades over time
- Participants are mostly interested in spikes
- Memories, events are recalled relative to other things

System use driven by perceived problems

- Spike activity motivates system use
- Health issues motivate system use

Patterns

- Use data/ annotations to find patterns
- Use patterns to identify opportunities for change
- Use patterns to motivate/justify taking action
- Use patterns to validate interventions

Anticipated uses

- User(s) expect to use the system for...
  - Self-experimentation

  - Health management
  - Data review (daily/weekly/monthly)
  - Extracting event correlations
  - Investigating spike sources
  - Informing behavior change
  - Extracting health correlations
- Users want to...
  - Share data with doctors
  - Monitor air quality
  - Improve air quality
  - Compare indoor and outdoor air quality
  - Perform between-room comparisons
  - Use through Salt Lake City inversion season

Annotations

- Convenience & proximity affects annotation
  - Google Home, voice prompts, and tablet all require proximity to use
  - Active prompts increase annotation likelihood
  - Users need to remember, make time for hardware (interface)
- Participation: willingness to annotate influenced by...
  - Air quality anomalies (i.e., large spikes)
  - Participant availability (competing routines, activities, schedule)
- Expectations - how participants expected to annotate throughout their deployment:
  - Daily annotation habit
  - Preemptive annotation habits
- Barriers - what would get in the way of annotations
  - Physical proximity to hardware
  - Their own memory
  - Lack of habit
  - Requisite situational awareness to make an annotation
- Motivations - What would motivate participants to annotate
  - Identify spike sources

  - Improve data meaning
  - Uncover, track health correlations
  - Desire to "fix" air quality
- Modality feedback - what participants had to say about their annotation options
  - No one likes notebooks
  - Texting was preferred method
  - System-initiated voice prompts would require an immediate reply
  - Tablet was middle preference.

Participants were also interested to use their systems to:

- Correlate: observe and relate events which are detected with any number of dependent factors (health)
- Investigate: use the system to self-experiment to find spike sources.
- Validate: using the system to change behaviors

A.5.3 Phase III themes and descriptions: Postdeployment interview

Themes from Phase III reflect participants' longer-term relationship with their data, and experiences interacting with the interface and system-initiated annotations. We also report on privacy attitudes and a basic analysis for ways in-home deployments could improve medical studies, and the various things our participants wanted to do with their data.

Annotations

- Lower barriers to annotate: simplicity affects participants' annotation likelihood
- Likelihood to annotate directly related to being physically present to annotation modality
- Convenience plays an important role in annotating data. Proximity to annotation modality affects participants' preference for how to annotate.
- Annotation improves participants' memory and understanding
- Annotation adds value and meaning to logged data
- System-initiated prompts are essential for collecting annotations
- Visual references are important when annotating.
- Timely prompts improve participant recall, ability to identify potential sources.

Privacy attitudes

- Indifference:
  - Sharing air quality data
  - Sharing health data (fine if anonymized)
  - Sharing annotations on air quality data
- Slight caring:
  - Personal association to air quality data
  - Sharing anonymized data
  - Sharing publicly available information (names, phone numbers, etc.)
- Concerns:
  - Undisclosed surveillance
  - Sharing demographic data
  - Sharing/deriving personal schedules
  - Sharing health data (it's private)
- What is "personal" and what is "private":
  - Participants likely to conflate terms
  - What constitutes "private" data is subjective
  - Distinction made between what happens inside (private) vs. outside (public) the home
  - Personal data seen to be data that is about oneself
  - Private data is a subset of personal data
  - Whatever is deemed "private" should *not* be shared
- Distinction
  - Security in obscurity - air quality data considered too abstract, inconsequential to be of concern
  - Participants do not consider air quality to be personally identifiable information

Feedback for National Institutes of Health

- For asthmatics, real-time air quality updates motivate study recruitment
- Real-time updates enable behavior change
- Participants join study to improve lives: their own, and others helped by their contributions

- Happy users make happy doctors
- High participant engagement leads to better participant retention in study
- Sharing measured data improve sense of engagement with a study
- Sharing measured data empowers users to make changes
- Stale data alerts are less useful, timeliness is directly related to its usefulness.

e-AsthmaTracker comparisons: Tools are useful in different ways

- e-AsthmaTracker measures reactive behavior
- Our air quality interface supports proactive behavior
- e-AsthmaTracker seen as designed for clinicians, not user friendly
- Deployment air quality interfaces seen as more user-centric
- e-AsthmaTracker is difficult to use, interpret
- Air quality system interface is easy to use

System use-cases

- General
  - Exploratory use
  - Regular, routine use
- Data-centric
  - Finding patterns
  - Event comparisons
  - Monitor comparisons (between-streams)
  - Reviewing 'whats happening now'
- Spike-centric
  - Finding sources or origins of poor air quality events
  - Exploring properties and distributions of poor air quality events (spatially, temporally)
  - Understanding air quality diffusion
- Health-centric
  - Health journaling
  - Validating personal symptoms
  - Supporting, empowering behavior change to improve health
  - Determining personal asthma triggers

  - Supporting preparatory or preventative health planning (medication, etc.)
- Behavior-centric
  - Inform or justify participant behaviors
- Annotation-centric
  - Reading, reviewing, editing, searching, sorting, participant annotations.
- Air quality-centric
  - Understanding personal activity in the home and its impact on air quality
  - Understanding the influence outdoor air quality has on indoor air quality
- Self-experimentation

System Feedback

- Observations, usage
  - System must be accessible (both interactively and proximally) to be usable
  - Participants engage with system only when convenient (and nearby)
  - System deployments increased participant air quality awareness
  - System-initiated prompts are essential for engaging participants
- Praise
  - Participants had compliments, positive things to say about their deployments
  - The interface supported users tasks
  - Participants felt the data shown was accurate and reliable
- Criticisms
  - System deployment does not engage, inspire broader family
  - Participants were initially interested in using annotations for health journaling, but did not
  - System deployment time was too short
    * Air quality data was not shared with guests
    * System did not capture asthma events

A.6 Additional tables and figures

A.6.1 Annotation counts

A.6.1.1 Full deployment

Table A.1. The total number of notifications sent and annotation received throughout Phase I - III for each deployment. User-generated annotations are additionally broken out by annotation modality

Deployment	# Notifications	# Annotations	Compliance (%)	Text #	Tablet #	Voice #
D1	118	65	55.1%	51	13	1
D2	121	115	95.0%	97	12	6
D3	103	130	126.2%	86	12	32
D4	396	78	19.7%	68	7	3
D5	26	51	196.2%	27	20	4
D6	69	57	82.6%	55	1	1

Table A.2. This table lists the text-only compliance rate for MAAV notifications and participant text annotation counts. Entries per phase reflect values for that phase only.

	Phase II				*Phase III*			
	Duration (weeks)	Number Notifications	Text Annotations	Compliance	Duration (weeks)	Number Notifications	Text Annotations	Compliance (%)
D1	3.0	8	4	50.0%	37.9	110	47	42.7%
D2	4.1	18	16	88.9%	35.7	103	81	78.6%
D3	2.0	5	5	100%	34.9	98	81	82.7%
D4	2.6	14	7	50.0%	34.7	382	61	16.0%
D5	4.1	19	19	100%	4.0	7	8	114.3%
D6	6.3	36	31	86.1%	7.9	33	24	72.7%

A.6.2 Engagement proxy plot breakout counts

Table A.3. Deployment engagement counts. This table breaks down the number of response types for each deployment when prompted with a MAAV notification.

	Notification Days	Days with Annotation-only	Days with Interface-only	Days with Annotation & Interface
D1	78	3	28	47
D2	95	49	2	29
D3	81	54	2	21
D4	144	43	2	21
D5	18	4	0	14
D6	45	4	5	34

A.6.3 Participant interaction with tablet interface

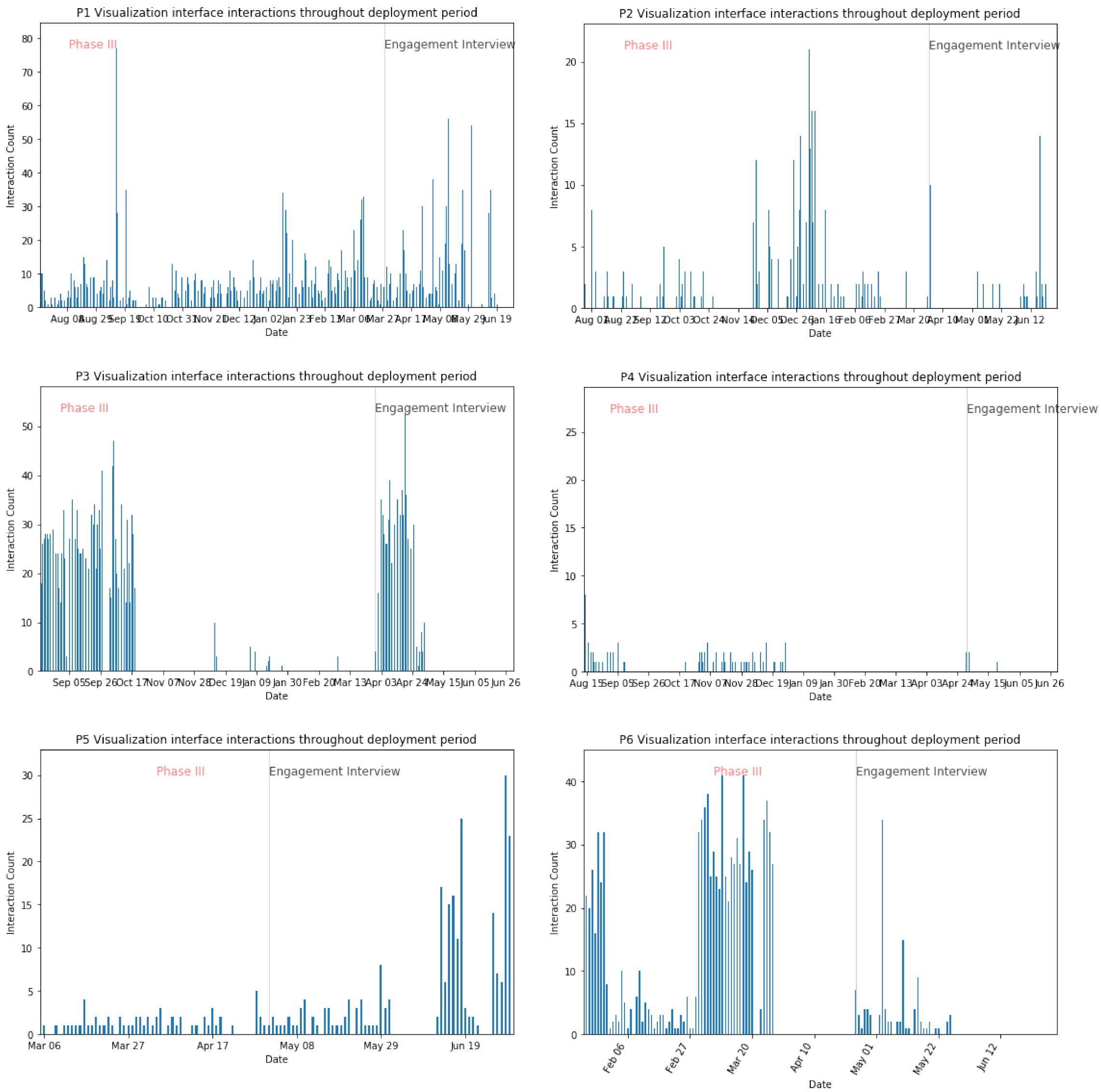


Figure A.1. Number of tablet interface interactions per day for each deployment.

A.6.4 Participant annotations per week

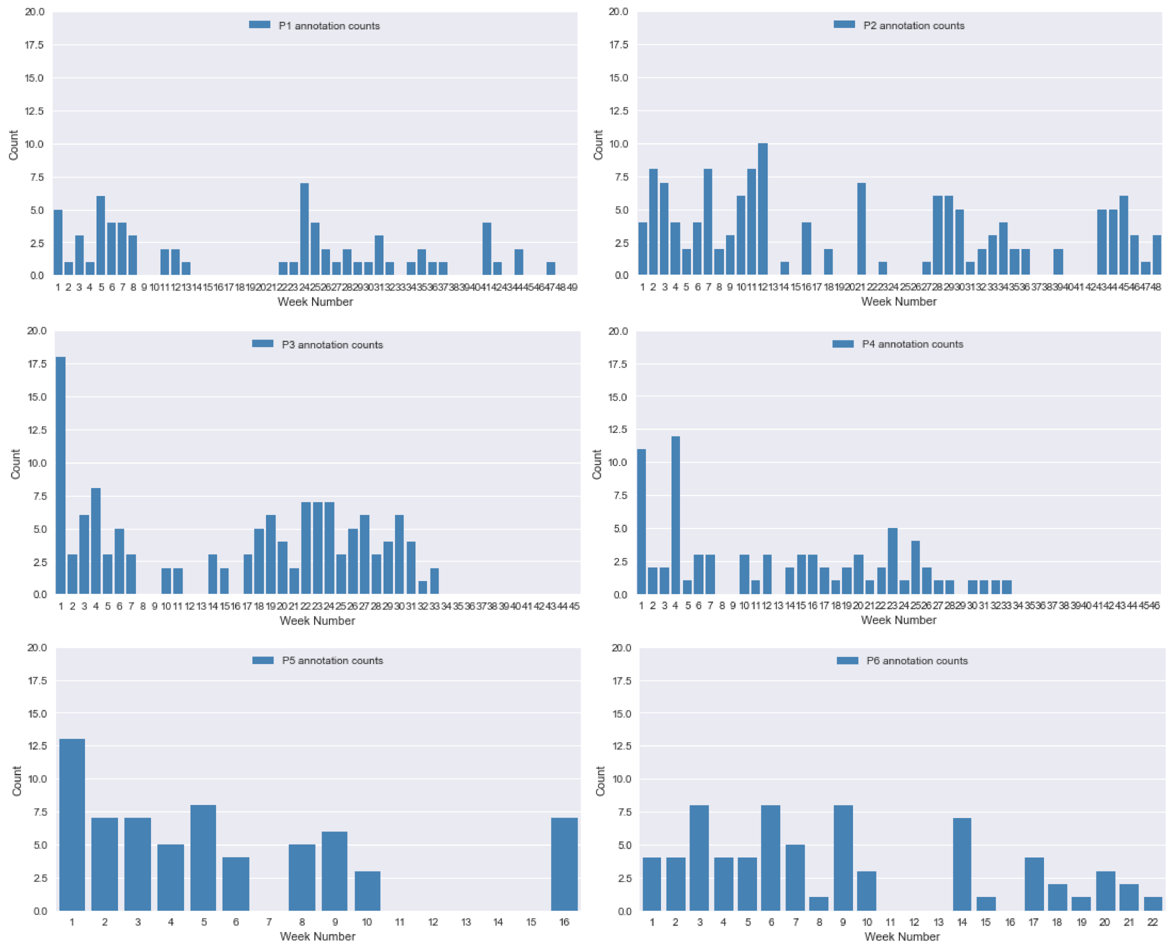


Figure A.2. Number of tablet interface interactions per week for each deployment.

APPENDIX B

CREATIVITY WORKSHOP OUTCOMES

This appendix outlines the high-level themes and question topics related to what participants wanted to know, see, and do with their tracked air quality data. These themes summarize the outcomes of our creative visualization design opportunities workshop [80] conducted toward the end of our field deployments.

B.1 Workshop Summary

A shortened summary of the high-level themes discussed during our creativity work shop with participants P3 and P6.

B.1.1 Actionable themes

Improving health: Correlate air quality data to physical health and identify ways to improve both.

Decision support: Workshop participants identified wanting to use their data to help make personal decisions regarding their health and well being, and to support prioritizing their decisions.

B.1.2 General themes

Greater education and awareness: Workshop participants wanted to know more about how the system and detectors worked in order to better interpret what the tablet interface was showing them. This includes on-screen explanations and access to more detailed external information.

Improving communication: Air quality data can be shared with friends, family, and clinicians to increase awareness of people's exposure to particulate matter.

Empowerment and control: At the highest level, workshop participants wanted to use their air quality data to help manage their family's health and asthma symptoms.

Personalized tools

- *Health journaling and tracking*: Support logging and tracking personal health information alongside air quality to help identify and track asthma triggers and sensitivity trends.
- *Self-experimentation*: Support users to perform their own mini-experiments to see how air quality changes when performing certain actions, or to analyze periods of time to determine whether the air quality has improved relative to some event (such as changing a filter, buying a new vacuum, replacing windows, etc.)

Data comparisons: Participants wanted the ability to compare historical measurements across a number of different conditions: time, activity, location, other participants, or 3rd party information.

Data summaries: Participants provided several ideas for generating and sharing summary statistics:

- *User-defined analysis windows*: Summarize data across different time ranges (weeks, months, seasons, years, etc).
- *Baseline metrics and statistics*: Examples include tracking the number of spikes, average peak intensity, time to recover, and/or room or house-wide averages for a given time period.
- *Highlight anomalous events*: Largest spikes, most active room, spikes occurring in one room but not others, etc.
- *Aggregations*: Characterize the air quality activity over selected time range. Spike activity by hour of day or days of week. Aggregate on activity or location, etc.
- *Contrasting summaries*: Compare the current summary statistics with the previous summary. For example: now vs. previous month, or this time last year. Highlight any changes over previous summaries.
- *Customizability*: Support users creating customized data summaries. Include only the data, measurements, or external information relevant to the person or family unit
- *Integrate third party information*: Include a news feed of local events to help contextualize AQ measurements (fire, accidents, holidays, etc.).

Improved system feedback: Workshop participants highlighted opportunities to provide more actionable feedback. A number of suggestions can be implemented with specific engineering:

- *Send better texts*: Incorporate humidity and temperature information in text information to help determine whether the spike is caused by humidity. Indicate whether the spike size is normal or larger than usual given its time, location, etc.
- *Identify patterns*: assist with finding annotation or spike patterns.
- *Receive personalized Feedback*: Tell users how to improve their air quality based on their readings and spike activity. Indicate when and how they should take action.
- *Highlight differences*: Validate users' effort to improve air quality by showing when there's a difference (between rooms, between times, etc)

Integrating external devices and/or data sources: Adding weather information, mold spore, or pollen count forecasts. Providing additional context via newsfeed for local environmental events possibly affecting air quality. Other suggestions included:

- *Integrate e-AsthmaTracker results*: Show how indoor air quality relates to personal Asthma Control Test scores.
- *Incorporate wearables*: Deduce air quality exposure away from home
- *Integrate sentiment data*: mood, stress level, illness, quality of life, etc.
- *Personal exposures*: connect the interface to a community-level data tracking system

B.2 Workshop themes and descriptions

The complete listing of themes discussed in our creativity workshop.

B.3 What participants wanted to know

Education and awareness

- How the detectors work – Participants wanted a better understanding of the system. What generates spikes and why? What doesn't generate spikes and why?

- Better understandings of microclimates – Why do monitors respond the way they do? Why are some seemingly more connected than others?
- Answers to general health concerns and questions – the ability to correlate health questions with air quality data, whether the home air quality contributes to poor health, what the air quality is like at my child's school, is indoor air quality better than outdoor air quality? etc.

Comparisons

- How someone's air quality evolves in time – near term, long term, by location, etc.
- Community comparisons & Competition – How am I doing compared to other people in the study?
- Relational improvements – To know whether a lifestyle or home improvement has made a measurable impact on indoor air quality.
- Time characterization – the ability to summarize data for specific segments of time (months, seasons, years, etc.).

Finer details

- Spike information – to be told the the source of detected particulates and spikes, to distinguish between humidity vs. PM, and to know which spikes are "bad".
- Integrating a newsfeed to understand local influence (for outdoor air quality).
- Integrating other environmental measures (pollen, mold, etc.).
- Provide better context with text prompts (is this a big spike? Or normal?).
- Provide summary statistics for later review.

B.4 What participants wanted to see

Education and awareness

- Provide a greater emphasis on data explanations. Tooltips, and text summary of what it is people are seeing. Provide information on implications of levels, and links to explain what things mean.

Comparisons

- Integrate results from e-AsthmaTracker surveys. Show how indoor air quality relates to ACT score.

- Validate users' effort to improve air quality by showing when there's a difference (between rooms, between times, etc.).
- Display EPA outdoor air quality forecast classifications and nearby DAQ readings.
- View indoor air quality, outdoor air quality data compared against EPA scales.

Finer details

- Receive personalized feedback - how to improve my air quality based on my readings and spike activity? When should I take action? How?
- Provide more "intelligent" and detailed SMS messaging. Incorporate humidity and temperature traces.
- Integrating other environmental measures (pollen, mold, etc.)

B.5 What participants wanted to do

Empowerment and control

- Achieve some personalized notion of success – "absolute" (the best I can be) or "relative" (better than other participants)
- View data on the go.
- Make Educated decisions – use data to inform corrective behavior, and to justify preventative decisions.
- Share data – with family members, teachers, doctors, local officials, etc.
- Use data to improve family health.

Comparisons and summaries

- Compare against: other deployments, the local area, homes with similar characteristics, the most "air quality efficient", etc.
- Compare data across specific activities or activity types (cooking, cleaning, etc.)
- Summarize data across time : days, weeks, months, years.
- Analyze and compare basic air quality statistics across time ranges (number of spikes, baseline average, etc.).
- Create customized data summaries – include data, measurements, or external information relevant to the person or family unit.

Annotations

- Revision – the ability to edit, move, delete.

- Expressiveness – choosing graphical icons for activity type.
- Automation – suggest classes of annotations based on time, location, previous entries, dynamics, etc.
- Search and Filter – have the ability to highlight or link annotations (and underlying data) of the same class, text, date/time, etc.
- Annotate time ranges rather than data points.

Finer detail

- Self-experimentation – Support A/B tests.
- Health journaling and tracking – find sensitivity trends, identify and track triggers.
- Incorporate wearables - deduce air quality exposure away from home.
- Community-level data tracking and broadcast system.
- Collect and Integrate sentiment data (mood, stress level ,illness, quality of life, etc).
- Pattern flagging and discovery - assist with finding annotation or spike patterns.

B.6 What can be possible

Feeling empowered

- Combating helplessness
- Knowing what's in your control

Improving decision making

- The ability to make personal decisions from data
- The ability to prioritize personal interventions – let people perform their own cost-benefit analysis with the help of this data
- Motivating and justifying preventive behavior
- Improving health awareness and engagement
- Validating medical advice
- Supporting self-experimentation

Improving communication

- Between family and clinicians
- Between family and social circle (w/ summary reports)

- Automated feedback between system and user – incorporate community knowledge, guidelines, advice and known outcomes to suggest behaviors and alternatives for improving air quality practices.

B.7 Discussion on receiving summary statistics

Important to show:

- Baseline metrics and statistics – number of spikes, average peak intensity, time to recover, average home or room baseline
- Provide temporal self comparisons – this vs. previous month, now vs. last year, improvements over previous summaries, etc.
- Seasonality / temporal comparisons - support analyses over user-selected time ranges
- Include outdoor sensor readings - compare personal readings to DAQ, others
- Stack spike activity as detected spikes over time of day & spikes over days of week.
- Include a news feed of local events (fire, accidents, holidays, etc) to overlay with or contextualize corresponding air quality measurements.
- Highlight hyperlocal spikes – those which occur in one room, but not others
- Identify whole-home events – those which spike in both indoor sensors.

APPENDIX C

DATA MODEL PAPER

This appendix presents an initial paper submission covering very early articulations on the personal informatics analysis gap. This rejected paper illustrates our design-oriented thinking from our stated contributions, including: a taxonomy of our participants' questions to define a preliminary task space, a proposed data model to aid with designing tools for answering participants' questions, and our data engagement interview method — initially called the 'task elicitation interview' – further highlighting our analytic focus.

We present this work as-is, and acknowledge it as incomplete and unfinished research.

Exploring the Personal Informatics Analysis Gap Through an Indoor Air Quality Study

Personal informatics promises self-knowledge for all, but lacks the tools for users without previous technical background to analyze that data beyond basic developer-determined representations — this is the personal informatics analysis gap. Our participants collected a year's worth of indoor air quality data and had questions that the data could answer, but no way to get those answers on their own. We developed a novel interview method, the task elicitation interview, that incorporated elements of think-aloud, wizard-of-oz, and live coding to understand how participants would answer those questions were technical background not a barrier. Results showed a breadth of participants' goals and questions, which evolved as they explored their data during these interviews. We used these results to specify a data model that lays the groundwork for a tool to bridge the personal informatics analysis gap. These results and the task elicitation interview show promise for others seeking to bridge the gap in other domains of personal informatics.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing design and evaluation methods**; **Empirical studies in ubiquitous and mobile computing**; **Visualization design and evaluation methods**; *Empirical studies in HCI*; *Information visualization.*

Additional Key Words and Phrases: Indoor air quality, asthma, personal visual analytics, intermediate analysis tools, non-expert user, personal informatics

ACM Reference Format:
. 2020. Exploring the Personal Informatics Analysis Gap Through an Indoor Air Quality Study. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 0, 0, Article 0 (2020), 30 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

People use technology to capture data about their everyday, lived experience because they believe that these data will help them make better, smarter decisions about how they live their lives. They actively collect data about their bodies through fitness trackers and sleep devices [25, 26, 33, 35, 65]; about their environments through air quality monitors and utility usage sensors [11, 38, 45, 70, 71, 91]; about their health through digital diaries and nutrition trackers [22, 24, 121]; and about how they spend their time through calendars and social-media trackers [80, 98], to name just a few. Tools for exploring, analyzing, and understanding these data give users opportunities to reflect on their lives and consider ways to improve their health, well-being, and communities [53, 78]. Despite this proliferation of technology, however, personal data are not making the seismic changes in people's lives that technologists have imagined [37, 50, 127].

A significant challenge in utilizing personal data is that most people today lack adequate training, experience, and resources for asking and answering questions with data [44, 53]. Existing tools for making sense of personal data overcome this challenge by focusing on a narrow set of specific, predefined questions, eliminating the need for users to translate their questions into analysis tasks or to wrangle their data into an appropriate representation. These tools, however, do not let users explore a broad set of personally relevant questions, nor do they take

Author's address:

2474-9567/2020/0-ART0 $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

advantage of users' rich, situated, and extensive knowledge about which aspects of their data are personally interesting and insightful, and which are not [19]. On the other hand, visualization tools designed for rich, bespoke data exploration assume that users have the skills and resources to translate their domain questions into a form that is amenable to analysis [75, 129]. The result is a gap in the existing ecosystem of analysis tools that enable people without analytic experience to take full advantage of the extensive personal data that they are collecting: the personal informatics analysis gap.

This work takes steps toward bridging the personal informatics analysis gapthrough a study of residents and their indoor air quality data collected from their homes. We build on previous work that focuses on understanding *how* people used a deployed air quality monitoring in their homes [69, 91], and extend that focus here to understand *what* people want to get out of their data: what are the range of questions residents want to ask from their indoor air quality, and how we can design technology to support them in answering those questions. We engaged residents from the study in a workshop that pushed them to speculate on a broad and diverse set of questions they might ask of their data, and we conducted a novel form of data-centric interviews with them to understand how they approach exploring and analyzing their data to answer their questions. Our analysis of this workshop and interview data reveals a diverse set of personal questions about indoor air quality, as well as a data model that is sufficiently flexible for accommodating a broad range of questions and analysis tasks.

More specifically, this study offers three contributions:

- A new interview method for eliciting data analysis approaches and insights by engaging participants directly with their personal data.
- A characterization of the range of question our participants wanted to ask of their indoor air quality. These categories align well with recent work describing a goal-framework developed from personal informatics data about migraine control [110, 111].
- A data model for indoor air quality data that supports analyzing a rich and broad set of personally relevant questions. This model can serve as the backbone for visual analysis tools that enable residents to take full advantage of their indoor air quality data.

We speculate that both our proposed interview method and data model will generalize more broadly to other contexts within personal informatics. Additionally, our work shows that participants' motivations for engaging with personal air quality data are similar to migraine self-tracking goals [110, 111], and likely other PI contexts as well. Taken together, these results provide the foundations for designing new personal informatics analysis tools that enable people to ask a broad set of personally relevant questions from their indoor air quality data, as well as a roadmap for acquiring those foundations in other personal informatics domains.

We begin by outlining the motivation and theoretical backdrop for identifying the personal informatics analysis gapin Section 2. We explore this space through interviews and a workshop to understand how asthmatic families engage with their personal indoor air quality data (Section 3), and discuss a novel method for collecting user feedback (Section 4). We analyzed the collected data to produce a characterization of participants' goals and questions (Section 5) and a data model that can be used for developing a system that fills the personal informatics analysis gapfor indoor air quality (Section 6). We conclude with a discussion that outlines potential implications and techniques for designing for this population, speculation on how we might transfer this knowledge to other personal informatics domains, and ideas for future work (Section 7).

2 THE GAP IN ANALYSIS TOOLS FOR PERSONAL DATA

This section calls attention to a gap in the ecosystem of analysis tools for supporting people asking and answering a broad range of individualized questions from their personal data. We argue that personal informatics tools are designed to support users without extensive data analysis experience and skills at the expense of constraining them to a narrow set of questions, while the design of flexible visual analysis tools targets domain and data experts with the time, resources, and skills to dig into data. The space between these two classes of tools — systems that support a flexible and broad set of questions while accounting for the challenges of operationalizing questions into low-level data analysis tasks — is the *personal informatics analysis gap* and is underexplored in the HCI and visualization literature.

2.1 Analysis Tools for Personal Informatics

Personal informatics research emerged from the quantified-self movement and describes ways of supporting people in collecting and reflecting on personally relevant information for gaining self-insights [78]. As a community, quantified-selfers are considered an extreme user group since they are highly self-motivated and ambitious in their search for insights from their personal data, using a wide variety of analysis tools and techniques [17, 20]. Despite their motivation, a variety of challenges complicate quantified-selfers' ability to extract insights from complex data stemming from difficulties in deriving novel, non-obvious insights [5, 33], selecting appropriate visualizations [20], and navigating information overload from tracking [20] or analyzing too many variables [60, 61].

The availability and growing popularity of numerous self-tracking products has democratized self-tracking to a broader user-base and led to a growth of bespoke personal informatics tools for helping people make sense of their data [34]. Whereas early tools tracked or analyzed single facets of people's lives, such as sleep schedules [18] or physical activity [79], subsequent research shows that analyzing single data streams limits potential insights [5, 60, 61]. Systems for tracking multiple facets thus emerged in order to sustain engagement and improve insight generation [5, 33, 46, 78, 89, 127], with a focus on reducing the effort required to integrate and analyze diverse data streams through mechanisms such as the automatic detection of potentially interesting correlations based on statistical analysis [60, 61, 127]. An underlying assumption of these systems is that users are looking for serendipitous insights through exploration of the data, but people can have other goals for self-tracking [36], for example, looking for answers to high-level questions that they bring to their data analysis and that guide the data relationships they seek.

Work by Epstein et al. [33] proposes visual *cuts* as a technique for supporting interactive, exploratory analysis of personal data. Cuts are subsets of data, defined by any attribute in a data set, that support detailed comparisons of interesting parts of the data. This work paired predefined cuts with appropriate visual representations to support users in finding meaningful patterns and trends in their data; the cuts were determined based on formative work that characterized the kinds of questions users had of their personal data. A summative evaluation found that cuts were effective at supporting some of the users' questions as well as triggering serendipitous discoveries. These results, however, also highlight that predefined cuts fail to meet everyone's needs — a finding echoed in other studies that attempt to automate insights [5, 61] — leading the authors to recommend that "designs do not attempt to limit cuts based on stated goals and instead offer a variety of cuts" [33].

What remains unclear is how to design tools for arbitrary and flexible cuts, especially in the absence of explicitly prescribed goals, data attributes, or boundaries. Our work addresses these challenges by reframing cuts as *comparison targets* [42] and proposing a general data model for supporting a wide range of individualized questions about personal data, including questions for which comparison targets are fuzzy and ill-defined. We argue that this data model can serve as a basis for designing flexible, yet approachable, visual analysis tools for personal data.

2.2 Visualization for the Masses

As the use and consumption of visualizations has spread beyond professional contexts and into everyday lives, research into designing tools for casual users has become a prominent thread of inquiry [48, 53]. Early research into visualization tools for mass audiences experimented with effective ways to communicate data. Visualizations such as the Map of the Market [125] and the Baby Name Voyager [126] experimentally deployed relatively advanced techniques to audiences largely unfamiliar with data visualization. Other tools such as ManyEyes [30, 122], sense.us [49], and Vizster [47] looked to democratize visualization – moving visual data analysis from the realm of professionals and into the hands of everyone [54] – with a particular focus on sensemaking as a social process [48].

Following the wave of tools designed for casual contexts, visualization researchers expanded their investigations of visualizations for the masses. Initial research topics in this field include visualization literacy of novices [7, 9], the inclusivity of casual visualizations [27, 97], telling stories with data [74, 85, 113, 133], and the sensemaking benefits of data physicalization [54, 56, 117]. Researchers also explored how mass audiences create and use visualizations, and what gaps exist in what we know about casual use. This research often makes a binary distinction between professional and casual use of visualizations, with casual users characterized as analysis novices unfamiliar with visualization and visual analysis [44, 102].

In a study on casual users, researchers found that novices struggle to "translate questions into data attributes, construct visualizations that help to answer these questions from a set of data attributes, and interpret the visualizations... [which] impeded the overall analytics process significantly" [44]. Another study on people's motivations for engaging with casual visualization finds that personal interests shift these motivations [115]. Finally, a survey of casual visualization research highlights the role people's contextual knowledge plays in forming rich insights. The survey authors note that most tools do not support personalized analysis as "current designs are mostly devised by system designers, who seem to decide 'what information to present' and 'what metaphor should convey the message' without considering the unique perspectives of individuals" [53]. This critique motivates our work toward developing an interview method for eliciting analysis tasks from casual users to better ground design requirements in the needs and skills of end users, not system designers. It also highlights the need for flexible visual analysis tools targeted to the needs, skills, and motivations of casual users.

2.3 Flexible Visual Analysis Tools

Research and development of rich, flexible visual analysis tools has largely targeted professional contexts [112, 129]. Developing visualization tools for use in professional settings stems from two core threads of research. The first is the design of bespoke tools for domain experts through the process of visualization design study. The second is the creation of powerful and flexible systems for data analysts, informed by studies considering how these analysts work with data.

A proliferation of bespoke tools for supporting rich visual analysis in a broad range fields — from biology [86–88, 94] to poetry [51, 84], meteorology [105] and law enforcement [75] — stems from the visualization research community's embrace of a call to "collaborating closely with domain experts who have appropriate driving tasks in data-rich fields to produce tools and techniques that solve clear real-world needs" [92]. These tools provide insight and innovations in how visualization techniques are, and can be used for enabling data-driven insight by domain experts with deep and extensive domain knowledge [129]. The dominant research approach behind the design and development of these tools is *visualization design study*, an approach to problem-driven visualization research that emphasizes designing visual analysis tools in close collaboration with domain experts [112]. Published research papers that report on design studies cover a range of application areas, but all these design studies focus on developing analysis tools for domain experts *working* with data; none report on collaborations outside of a professional context.

Research into visual analysis systems for data analysts arises from the significant increase in the number of professionals whose primary task is data analysis. Visualization researchers have focused on understanding the work practices and visual analysis needs of *data analysts*, defined as "professionals who analyze data as a major part of their daily job" [63]. Data analysts use advanced skills and tools to explore data in order to "suggest hypotheses, assess assumptions, and support selection of further tools, techniques, and data sets" [1]. Various interview studies focus on how data analysts do their work, and offer characterizations of their overall analysis process within the organizational context [63]; their patterns of exploratory data analysis [1]; their impediments to efficient data analysis [64]; their roles within software development teams [68]; and their unique considerations when working with cloud architectures [39]. These studies extend and modernize earlier research on intelligence analysts' work practices [28, 96, 101, 132] by seeking to understand the sense-making and information-foraging processes of data analysts [100, 108] and providing insights to guide the design of sophisticated analysis tools. An growing ecosystem of visual analytic tools for enhancing these processes supports data wrangling [6, 62], interactive visual analysis [8, 82], and recommendations for effective views [83, 130, 131]. These systems require, however, that users are able to both translate their questions into accompanying analysis tasks and have the experience to make appropriate decisions based on visualizations that they see [2, 21]. Otherwise, users without these skills will struggle to use visualization tools effectively [21, 44, 75].

Sense-making and information foraging have not been similarly studied in a deep and systematic way for the visual analysis practices of casual users, leaving a gap in what we know about how to design tools for this user group. This gap motivates our work in studying the types of questions people ask of their personal data, the ways in which they might go about answering those questions, and a data model that can support rich and flexible visual analysis systems that can support them.

2.4 Personal Data Analysis with Indoor Air Quality Data

Indoor air quality data provide an ideal context to explore the personal informatics analytics gap. Most research on indoor air quality data is focused on sensing, and produces technical contributions related to system-level development [13], hardware [67], or modeling [38, 59]. Kim et al.'s 2009 *inAir* paper was the first to explore how effectively communicating indoor air quality data increases a household's attention and engagement with indoor air quality, but this and subsequent work focuses on ways of communicating data to improve awareness, but not *how* people want to use their data to answer personal questions [70–72]. Recent work by Kim et al. and others continues to find that few studies exist that focus on how people use indoor air quality systems or what they would like to do with the collected data [69, 91]. Current work investigating the preferences and practices of those engaging with air quality systems seek to bridge this gap, but still present a data-centric approach. Separate studies by Kim et al. [69] and Sakhnini et al. [109] provide high-level design recommendations for air quality sensing systems, but do not suggest how to process personal data beyond existing visual analytic advice within eco-feedback [41, 55] and ambient technology literature [90, 107].

To the best of our knowledge, this is the first work that characterizes the goals and motivations behind *why* people engage with personal indoor air quality data. We also go beyond existing high-level analytic recommendations to provide the first data model that supports analyzing a rich and broad set of personally relevant questions on indoor air quality. These contributions provide a strong foundation for future system designers to build flexible visual analytic tools that support users' broad and diverse personal air quality questions.

3 METHODS

We performed three research activities (Figure 1) to build an understanding of participants' broad range of indoor air quality questions. We designed these activities to help us understand how they used their air quality sensing system and identify associated design considerations for developing an effective visual analytics system for personal air quality data. Activity 1 (A1) tracks how participants engaged with an indoor air quality sensing system over a long-term deployment [91]. Activity 2 (A2) engaged participants in a participatory design workshop tailored to elicit a broad set of questions around what participants wanted to ask from the data collected from their homes. Activity 3 (A3) engaged participants with dynamic data-centric interviews to explore how they might operationalize questions from their personal air quality data into analysis tasks.

The information collected throughout these activities elucidate the motivations, goals, and needs of our participants in the context of engaging their personal indoor air quality data. Our qualitative, interpretivist analysis on these data offers more general insights from two perspectives: a characterization of the questions people want to answer with and from their indoor air quality data, and a data model for flexible analysis that can support answering those questions. The remainder of this section describes our participants, research activities, and analysis methods in more detail.

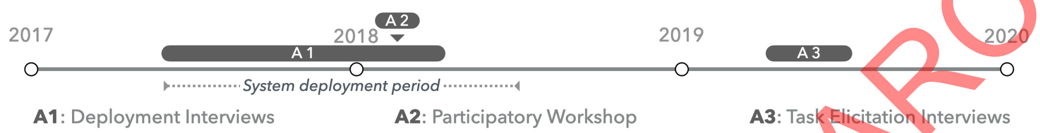


Fig. 1. We performed three separate research activities to understand how participants engaged with their personal air quality data. This timeline shows our initial deployment interviews (A1) to track how participants engaged with an indoor air quality sensing system [91]; a participatory workshop (A2) to capture what people wanted to know, see, and do with their indoor air quality data; and a series of task elicitation interviews (A3) for observing how participants approached operationalizing their questions and analyzing personal air quality data.

3.1 Participants

We retained our study participants from A1 [91] for their prior experience and engagement with reviewing personal indoor air quality data in the subsequent A2 and A3 research activities. These participants were initially selected from a concurrent university-run medical study involving asthmatic families [93] . Our participants were themselves asthmatic (P1, P2, P4a, P5, P6), or primary caregivers to asthmatic children (P3, P4). We include participant P4's teenage daughter, P4a, as an additional study participant given her significant engagement within this study. Other participant family members also contributed feedback and suggestions during A3 interviews but were otherwise not involved in A1 or A2.

Asthma motivated participants' initial engagement in our previous work; however, a variety of personal factors that arose during that work further shape how they wanted to engage with their air quality data. Participant P1 used her air quality system as a personal planning tool to stay inside when outdoor particulate concentrations were high, and participants P3, P4, P5, and P6 were interested to explore how air quality impacted physical and respiratory health more broadly. Participants P3 and P4 are primary caregivers to asthmatic children and wanted to use their data to provide a healthy home environment for their family. Participants P5 and P6 are adult asthmatics who already understood their symptoms and were interested to explore how environmental factors affected their sensitivities. Participant P4a was less concerned with personal health impacts and more curious to see how air quality affected those around her at a community level. Participant P2 is medically disabled as a result of her respiratory issues and was interested in ways to improve her quality of life, but wary of making random changes on her own. Rather than independently exploring or analyzing any data, she preferred receiving personalized advice from medical professionals on ways she could improve her living space. These different motivations shaped how each participant approached their data, its analysis, and the sorts of questions the participants asked within the study.

This study aggregates perspectives from each participant, collected within each research activity over this multiyear effort. With the exception of the design workshop conducted with participants P3 and P6, all research activities involved in-person interviews with every participant.

3.2 Research Activities

The first research activity **(A1) focuses on how participants engaged with an air quality sensing system over a long-term deployment** – a prior report provides extensive details about the deployment and system [91]. We staggered system deployments and participant interviews over a one-year period and engaged participants P1, P2, P3, P4, and P4a in the summer of 2017 and participants P5 and P6 in the winter of 2017-2018. We collected air quality measurements over the full deployment lengths for each participant with 3 air quality monitors per residence: 2 indoor and 1 outdoor. Participants also annotated the data in situ to describe activities occurring in their homes, and responded to daily surveys about the asthmatic family member's symptoms and behavior. During the deployment, we conducted three sequential in-person interviews at their homes to capture interview data on how participants engaged with their indoor air quality data, and a fourth engagement interview on broader family involvement. Our initial analysis of these interviews showed how an open-ended technology probe kept participants engaged over their long-term deployments and provided individual, personalized insights into the air quality in and around their homes [91]. Our analysis also showed that supporting participants in creating and reviewing contextualizing annotations on their data allowed them to appropriate the system in unique and occasionally unforeseen ways. These results point to the strength of fielding flexible platforms that allow people to engage with their data in a personal way.

Following this first research activity, we had planned to design a visual analysis system to support our participants in more fully engaging with their data. The interviews from A1 contained a significant amount of feedback on ways to improve the deployed system's visualization interface; however, further analysis revealed that the suggested improvements would not support the higher level goals participants shared at various points in their deployments. Upon reflection, we understood that, in the context of evaluating a field deployment, our deployment interviews were designed to gauge *how* participants used their air quality system, not *what* tasks they needed to perform in order to answer their personal questions. To address this shortcoming, we refocused our efforts toward better characterizing the range of questions our participants wanted to answer with their indoor air quality data and how they might go about doing so.

We conducted a **creative-visualization opportunities workshop [66] toward the end of the system deployment period (A2)** with the goal of collecting and characterizing participants' questions and motivations for hosting an air quality monitoring system in their homes. These participatory visualization workshops are designed to accelerate the problem definition and task analysis processes when developing visualization system design requirements. We conducted a half-day workshop with two participants (P3, P5) to understand how they wanted to use their air quality data. The workshop provided an outlet for the participants to externalize and share ideas about what they wanted to *know*, *see*, and *do* with their personal air quality data, organized around a series of guided exercises to solicit this information. We photographed and audio recorded all workshop exercises, and preserved exercise artifacts for future analysis.

Combining the data collected in A1 and A2, we again attempted to translate these data directly into design and task requirements for a visual analysis system. We surveyed the range of participants' stated goals, finding that they ranged from the direct and concrete – *What is the worst time of year for indoor air quality?* – to more abstract or out of scope – *Where's the best place to live in the valley?*. We suspected that seeing an answer to their posed questions would likely prompt follow-up questions or additional refinements that we would want any eventual system design to support. Once again, our efforts to characterize participant questions failed to provide insight into types of data analysis tasks that an effective system would need to support: what, specifically, would our participants want to do with their data?

To better understand participants' approach to working with their data, **we developed and conducted a new interview method (A3), the *task elicitation interview*, to directly observe their problem-solving approaches with their data**. This method began as a modified think-aloud protocol with hand-drawn data

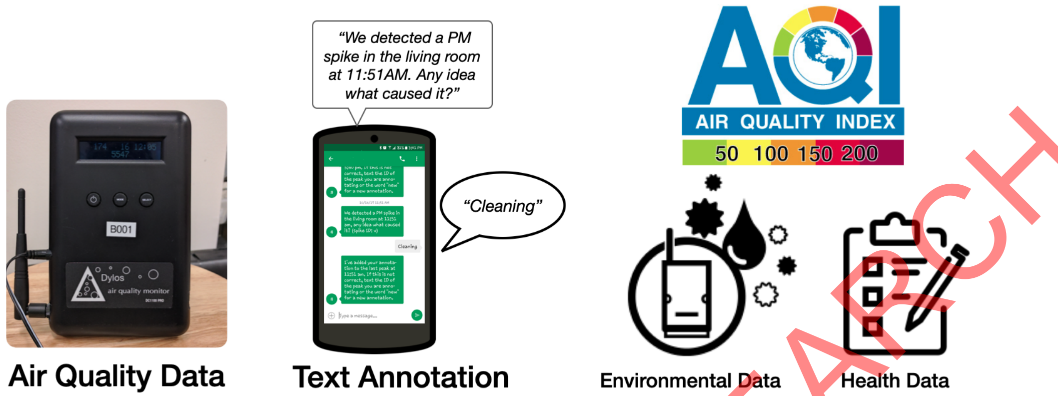


Fig. 2. Deployment data streams include sensor readings from multiple PM 2.5 monitors in and around users' homes and user-generated annotations for contextualizing air quality measurements. These are supplemented with external data streams such as daily EPA Air Quality Index classifications, environmental data (temperature, humidity, etc), and participant health survey data. Access to health information was granted by our IRB.

sketches based on what pilot participants wanted to see. This pilot work, which we conducted with visualization graduate students familiar with the project and other non-STEM recruited undergraduates, helped establish the kinds of preliminary views and analytic functionality we would need to employ with our primary participants. We then further developed the method to incorporate live coding by a dedicated data analyst who worked in concert with the interviewers to present a participant's data back to them in the format they requested. This method borrows techniques from visualization analytics interview methods [3] and the wizard-of-oz technique [29] to become a hybrid approach, which we described in detail in Section 4. The data we collected from this method during A3 provide a rich understanding of both the participants' motivations behind specific questions and their approaches and reactions to analyzing their own data. Our interviewers additionally engaged in reflexive journaling after each task elicitation interview, capturing thoughts and circumstances around what was surprising, unexpected, frustrating, or insightful during an interview. We revisited the journal frequently during our analysis of the task elicitation interview data.

We made use of multiple data streams within this study (Figure 2). These streams included participants' air quality measurements and user annotation data collected in A1. Based on participant feedback collected in A1 and A2, we incorporated daily EPA Air Quality Index classifications[1], IRB-approved access to participants' asthma health surveys collected through the parent medical study[93], and additional environmental data streams such as ambient temperature and humidity as external contextualizing data streams for the A3 s. Joining these primary and external data streams provided a rich set of analysis opportunities and greater participant engagement than would be possible with analyzing air quality data alone [5, 60, 119].

In total, we conducted 42 interviews across the 3 research activities, netting over 30 hours of interview audio. All participant interviews and workshop protocols proceeded from semi-structured scripts to ensure meeting primary goals while affording the flexibility to more deeply explore participants' motivations and thought processes as necessary. Interview protocols were cleared by our University's IRB and are available in supplemental materials. All interview participants were compensated with a $30 Amazon gift card after each research activity.

[1]https://www.airnow.gov/aqi/

3.3 Data Analysis

One researcher reviewed all interview and workshop data to find instances where participants shared their questions, goals, and motivations for working with their air quality data. In total, 160 of these excerpts were entered into AirTable, a cloud-based collaborative database, and two additional researchers assisted with independently clustering these entries. One pair of researchers coded these entries for high-level intents as a way of characterizing questions and motivations (Section 5). Another pair of researchers coded entries based on the visualization design patterns that would help support the questions and goals (Section 6). We performed several rounds of coding to explore different tagging and clustering strategies before developing our participant question characterization and settling on an organizational framework which became our data model. This process was informal, highly iterative, and represents what we found to be the most interesting interpretation of the available data.

4 TASK ELICITATION INTERVIEW METHOD

The task elicitation interview method is a technique for acquiring an understanding about how people operationalize questions about personal data into analysis tasks for reaching an answer. The technique has participants ask personally relevant questions and consider how they might go about answering those questions from their data. Next, the technique brings an experienced data analyst into the interview to perform on-the-fly analysis on participant's data as directed by the participant during the interview, allowing them to confirm or refute hunches. Real-time analysis allows participants to ask follow-up questions inspired by what they see and learn from their data. This method lowers the technical barrier to analysis for those without deep analytic skills by providing data analytics and visualization capabilities for engaging with personal data. The task elicitation interview provides researchers with insight into the tasks that a participant would want to perform with their own data without the developmental burden of designing a bespoke analytics tool first.

This interview protocol is an adaptation of the pair analytics research method [3], which was designed for capturing reasoning processes in visual analytics. Pair analytics borrows from protocol analysis and pair programming techniques by joining a subject matter expert and visualization practitioner to collaboratively tackle a relevant analytical task. This approach avoids the cognitive and social loads reported in standard think-aloud applications [32, 120, 128] by capturing participants' analytical reasoning through a conversational and collaborative problem-solving process. However, the benefits of this approach require that pair analytics participants share equal analytical and computational skills to productively work through their given task. The challenge we encountered was that our participants lacked any formalized analytic language or experience, but we still wanted them to articulate and operationalize their questions while exploring and analyzing their data.

To address this challenge, we build on the pair analytics approach and incorporate a dedicated data analyst role within the interview. Whereas the interviewer is responsible for engaging the participant and keeping discussion on-topic, the data analyst takes analytic direction from the interview participant. Unlike the standard Wizard of Oz methodology [29] where the interview participant unknowingly interacts with an analyst, the task elicitation interview brings the analyst to the forefront to gain the collaborative and conversational benefits of pair analytics. The participant drives the interview and is responsible for communicating their analytic goals and problem-solving process that the analyst then uses to process and present data, as directed. Using the analyst to augment their analytic ability allows participants to focus on operationalizing their questions into discrete steps toward reaching a solution.

Offloading the analytic burden also frees the participant to share thoughts on their process, justifications, and reactions in their own words as part of a naturally unfolding conversation. Whereas several think-aloud techniques exist in interviewing literature [40, 76, 103], few are tailored for specifically engaging visual analytic tasks and processes [3], and none incorporate self-tracked personal data in the analytic process. The task elicitation interview method enables researchers to collect more rich and ecologically valid insights on nonexpert users' motivations and problem-solving processes than other concurrent or retrospective verbal reports [40].

We process and visualize participants' data during the interview using an interactive computational notebook and used Jupyter Notebooks with Python for the task elicitation interviews we report on in this paper. An interactive computational environment allows for immediate feedback during the task elicitation interview and helps the analyst to easily communicate and iterate on these results. Computational notebooks are also an easy way to keep the participant engaged and informed on how their requests are implemented. Performing each interview within individual notebooks also creates a self-documenting record of participants' analysis process. Reviewing and reflecting on these artifacts helped us build our understanding of what participants want to do with their data and their overall approach to data analysis. It also helped us amass a store of reusable code that we could employ in subsequent interviews [73].

The following sections outline practical considerations for performing task elicitation interviews and outline interview results from six participants about their indoor air quality data. We also highlight the method's ability to provide insights into requirements for future visual analysis systems, as well as its ability to (re)engage participants with their data and the study. All resources related to the task elicitation interview method are available in supplemental materials.

4.1 Before: Preparing the Interview

Before conducting an task elicitation interview, researchers must acquire participant data and a skilled and attentive analyst. The following sections explore these considerations as well as preparing participants and establishing an effective protocol.

4.1.1 Gathering and Processing the Data. The efficacy of the task elicitation interview hinges on having access to a rich collection of data during the interview to ensure that participant questions can be sufficiently addressed. The collection must contain enough personal data to capture important patterns, as well as enough contextualizing data to make sense of patterns within the personal data. Incorporating and drawing from multiple data sets will significantly improve analytic flexibility and the likelihood of generating useful insights [5, 33, 61]. The data should be processed to facilitate on-the-fly analysis during the interview. Common preparation tactics, such as data integration, cleaning, transformation, and reduction, help organize and arrange data to aid the analyst with processing information more efficiently, and help catch or fix common errors with real-world data ahead of analysis [81, 104]. Separating data sources into organized tables or dataframes also helps make data analysis more flexible.

In the context of our air quality project, we incorporate participants' in-home air quality data collected during their system deployments[91], including measurement logs for each deployed air quality monitor around their home, and their entire text alert and response history over the extent of their deployment. These air quality data streams ranged from 5 to 12 months. We also included participant annotations of the air quality data, such as event labels like cooking and cleaning, along with a list of detected particulate matter spikes from our automated signal processing. Prior interviews and participant feedback also indicated an interest from the participants in contextualizing their indoor air quality data with official outdoor air quality classifications and their own self-reported respiratory health surveys[2] collected over the same time period. Access to this information was granted by an IRB-approved consent form as part of their accompanying medical study [93].

We faceted the collected air quality data by participant, and faceted individual participants data into separate tables that included their three air quality monitors, text alerts, text annotations, and particulate matter spike logs. Each source was labeled, organized, and formatted to help with easy search and filtering. This data processing step included splitting text alerts from user annotations, processing measurement timestamps to convert them

[2]https://asthmatracker.utah.edu

into participants' local time zones to clarify data review, assigning various temporal hierarchies [19, 89] such as days of the week, and manually classifying participant annotations by their activity, like cooking and cleaning. These labels made it possible for the analyst to filter and aggregate data across noncontiguous time spans and provide a greater number of possible data cuts [33].

4.1.2 Selecting an Analyst. Selecting a skilled and attentive data analyst is essential for fostering a collaborative problem-solving environment. The ideal analyst is someone who has an expert grasp of data processing techniques and workflows, can take direction well, and can clearly communicate their process to both researchers and participants. These traits help make the interview process informative, engaging, and enjoyable for the participant.

We selected our analyst from prospective graduate and undergraduate student candidates within our university's Computer Science, Mathematics, and Physics departments. These candidates came from other researchers' direct recommendation, and in response to a $17/hr work study position for an interactive data analysis project. We briefed applicants on the nature and goals of the interview method and provided test data sets similar in scope and content to participants' data in preparation for a live-coding interview where they answered sample questions using the test data set.

Finding the right analyst outside of an academic setting may be more difficult and require reaching out to local or professional data science groups to find candidates interested in working with researchers. Our chosen analyst was a senior physics PhD student with extensive experience analyzing large astronomy data sets. His ability to quickly filter, aggregate, and visualize sensor data made him an ideal fit for this project. His prior experience with Python, Pandas, and Jupyter Notebooks also enabled us to integrate preexisting code to help bootstrap our interactive analysis capabilities.

4.1.3 Participant and Researcher Preparation. To ensure a smooth interview process, participants should be informed of the sorts of questions and activities they will encounter. After establishing the communication contract [4] for on-boarding potential interview participants, we provided a conversation prompt to consider ahead of their interview: *Given access to your air quality data, what would you want to do with it? What would you want to know about it?* This preparatory step helps avoid downtime from participants searching for suitable tasks during the interview, and improves the chances that their stated goals are personally meaningful and relevant to their needs. In the event a participant lacks an actionable question involving their data, we offer a list of sample questions we are prepared to answer. Only two study participants asked to choose from the this list of questions (P2,P4)

Researchers must also come prepared ahead of the interview, having checked that all data, equipment, and materials are ready for use. Our mobile interview kit included the analyst's laptop and external monitor for sharing data with participants; an interview worksheet restating participants' conversational prompt with text boxes for participants to write in their questions and goals; assorted visual aides to assist with reflecting on their data or air quality system; and a sketching kit to support externalizing their analysis process. Each interview included physical printouts of the participant's data to provide a physical reference for exploring or motivating conversation around specific events. We found that placing information in the hands of participants and letting them browse and review aspects of their data, such as their text annotations, independent of analyst assistance, helped seed conversation or directives for what to explore or review. This benefit exists separately from having identical digital representations under analyst-control, and is a simple approach to further engage participants.

4.1.4 Developing an Interview Protocol. As with any resource-intensive user evaluation, it is important to pilot the task elicitation interview prior to engaging with participants. In total, we conducted 14 pilot interviews – 7 within our research group to collect feedback and suggestions on the interview protocol's high-level structure, and an additional 7 with several undergraduates to better understand how to integrate the analyst into the interview. The pilot study participants were not involved with our air quality deployments and thus were working with data that were not their own. During these pilots, we found that having participants write down their questions at the start gave a natural segue into exploring how and why their particular goals were important, and how introducing their data early on helped capture their attention and ground their analysis tasks in the available data. These pilot outcomes shaped our interview protocol to have participants articulate their goals and motivations from the outset and helps keep everyone informed, introduces personal data early to ground discussion over what is available, and utilizes iterative design cycles between participant and data analyst to rapidly generate and modify data in response to the participant's direction. This protocol structure is discussed in more detail in the next section.

4.2 During: Performing the Interview

The task elicitation interview method engages nonexpert users with a combination of problem identification and decomposition, sketching, and data review to help externalize and triangulate their needs (Figure 3). This approach offers researchers a rich and ecologically valid record of participants' thoughts and processes. The task elicitation interview is designed to be completed in 100 minutes with participants taking between 50 - 110 minutes (79 minutes average).

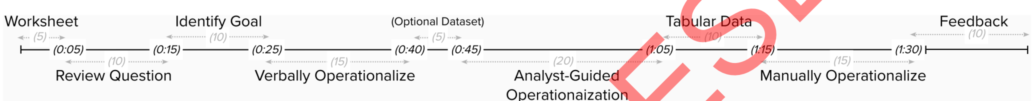


Fig. 3. Task elicitation interview timeline.

4.2.1 Collect, Discuss, and Verbally Operationalize Questions. The task elicitation interview starts with the participant filling out an interview problem solving-worksheet (5 minutes). The worksheet prompts them to identify what questions they want to answer, why they want answers, and what they plan to do with the answer. We capture this information to provide a conversational focus and a physical reminder of the interview goal to keep conversations on track. This step also helps put the participant in a problem-solving mindset and gives researchers time to set up and prepare for the analysis interview while the participant outlines their goals.

Depending on the length of time between the project data collection stage and performing the task elicitation interview, reviewing the project and participants' data may be worthwhile. Because we interviewed our participants several months after their initial deployments (Figure 1), we started their interviews with a short project overview of their in-home air quality system, the collected data, and their interactive interface. This review helped to refresh our participants' memories and frame the remainder of the interview.

After reviewing the project and their data, we have the participant review their question with the researchers as a way of seeding conversational points around participants' goals and motivations (10 minutes). Next, we encourage the participant to identify a specific goal they hope to achieve (10 minutes), and share how they imagine operationalizing their question to finding a solution (15 minutes). We allow the participant ample time to explain their approach, in their own words, and occasionally pose probing questions if they get stuck or omit any important analytic details. This problem description step helps participants externalize their thought process and provide insight on ways they think about their own problem solving process. This think-aloud approach also offers additional qualitative data on how and why they choose their question.

4.2.2 Analyst-Guided Operationalization. After the participant communicates their approach, we transition to loading and reviewing their data with the analyst. At the researchers' discretion, we optionally role-play a scripted exercise in which the interviewer directs the analyst to explore a local weather data set around key dates (5 minutes). This mini-exercise depends on the participant's technical background or comfort with operationalizing their question and is designed to illustrate the sorts of analytic tasks the analyst can provide.

Once the participant feels prepared to interact with the analyst, we introduce and review digital copies of their personal data. From here, the participant directs the analyst to prepare and present any necessary information they feel they need to answer their worksheet question (20 minutes). When exploring their data, we encourage participants to request specific views or regions that they feel are important, and to iterate with the analyst until they feel they have seen enough information to sufficiently address their original question. Throughout this process, we encourage participants to also sketch any ideas or imagined solutions as a way of further eliciting their analytic process. We found that encouraging participants to externalize their needs using different techniques led to organic explanations and clarifying details, which provided additional valuable qualitative data during the interview process. We collected all physical items generated during the interview and annotated them to capture the circumstances and motivations behind participants' choices.

4.2.3 Manual Operationalization with Tabular Data. Once the participant is satisfied with what they see, we reintroduce physical printouts of their data, including full text alerts, spike logs, and annotations, as well as a sample of their sensor measurements to explain the makeup of their personal data (10 minutes). Using their analyst-facilitated solution as a target, we ask how they would engage with each of these data sets for arriving at the same solution (15 minutes). Providing physical copies allows participants to directly mark and draw connections between multiple data sets and helps further externalize their thought process. This stage helps capture participants' approaches and reasonings through a series of guiding questions, and enables researchers to directly observe how participants work with data to combine personal information and draw connections and arrive at their earlier answer. This phase makes it possible to map participant actions to existing analytic taxonomies [10] and additionally outline, generalize, or compare their analytic approach versus that of other participants.

4.2.4 Reflection. At the end of the interview, we ask the participants to reflect on their experience. We ask for feedback on how the participant feels about their solution, how well it answers their question, and their overall confidence with the data they saw. We also ask for feedback on how comfortable they felt working with the analyst and the interview process in general (10 minutes).

4.3 Method Outcomes

We conducted seven task elicitation interviews with our participants to better understand the questions they would like to answer about their personal air quality data and how they would go about answering these questions. From these interviews, we extracted 75 questions, goals, motivations, and use cases that we analyzed with those from our previous research activities to develop our characterization of goals and questions in Section 5 and data model in Section 6. Additionally, the interviews revealed a diverse set of operationalization approaches and analysis tasks by our participants. These provide insights for designing a rich set of flexible analysis features within a future visual analysis tool. The interviews also re-engaged our participants with their data and caused them to reflect on their data collection practices during the indoor air quality deployments. In this section, we provide examples from our interviews for illustrating these two results.

4.3.1 Insights into design requirements. We developed the task elicitation interview method in order to better understand how people without deep analytical experience and skills would go about operationalizing their questions into analysis tasks. Our interviews revealed a host of strategies driven by the diversity of the questions our participants were interested in answering.

For example, as the primary caregiver for a child with asymptomatic asthma, P3 wanted to know whether any activities or conditions within her home contributed to her son's symptoms. During her task elicitation interview, P3 speculated about her primary analysis approach for addressing these concerns(verbal operationalization, Section 4.2.1):

> **P3**: If he's feeling really poorly, I can go and say you know what, we're having a lot of problems this week, or we've been having a really bad month. I'm going to go now and look back and look at the whole month and see the days he was feeling the worst, and I can go back and see what was going on now in the house, or in the air outside. What was happening in the environment around these times, and see if I can find correlations. And that's how I would use it.

This description specifies an approach of first faceting the temporal air quality data by week or month and using contextualizing recorded health data to identify specific facets of interests. Once P3 identifies specific weeks or months during which her son was experiencing poor health, she would look more closely at both the indoor and outdoor air quality to discover any correlations or trends.

During the task elicitation interview, P4 described a different approach to analyzing her data, sparked from curiosity after reviewing a list of her largest annotated spikes (analyst-guided operationalization, Section 4.2.2) :

> **P4**: On this one right here, I probably would first look at that big spike and say, 'well, what's going on there, what made it spike?' Then I'd go over and look at the part where it wasn't spiking and I'd say, 'Well, I was cooking dinner here and here. What was the difference?' That kind of thing, just to figure out what I can do to prevent as many spikes from happening. Especially if I notice it correlates to her breathing issues.

In this scenario, P4 would first identify large air quality spikes in her home and then facet the data on the event annotation associated with a large spike. She would then compare these facets to better understand if specific in-home events like cooking reliably lead to poor air quality. She additionally indicates that she would like to contextualize the facets with recorded health data to discover correlations and trends.

These two examples illustrate different types of *cuts*[33] through the data that our participants would want to make — cuts based on time and cuts based on annotated events. In both of these scenarios, contextualizing the cuts with additional data is vital for determining which cuts are interesting and why. Our task elicitation interviews revealed a range of cuts that our participants' questions would require, emphasizing the need for personal air quality analysis systems to move beyond current practices of presenting data as a fixed-time span line chart [13, 38, 67, 70–72, 91].

In another example, P3 describes a different faceting approach to the data, and speculates about how she would analyze the facets. When presented with the facets of her data by the analyst, however, she engages with the data in an unanticipated way:

> **P3**: I think I would start with the peaks. I mean this is an air quality thing, right? So if it were me, I would start with the air quality. When am I having a spike? I would then look at the spikes, and then from that I would try to correlate. On those days, what happened to cause a spike? Was anybody ill? Did we have an asthma attack? And try and do that.
> ***(Analyst loads data)***
> **P3**: I think what I would do if I saw this is I would want to find out what some of these are... anything above 3,000. Those look like kind of the outliers. What caused those?
> ***(Analyst tabbing through spikes one by one)***
> **P3**: I like seeing the correlation between the different detectors, because it is really interesting to me, because I don't think I realized previously. I knew we were causing spikes in the kitchen. I don't know that I recognized that we were also causing spikes downstairs. And that seems even more of a bigger deal, because I think normally I'd just be like, 'okay, well, I'm cooking, so why don't you guys go downstairs'. But maybe that's not really helpful. I guess I didn't realize how it affected the

> air quality in the whole house, and not just the kitchen area. So it's those kinds of things I think are really useful, just looking at this kind of stuff.

Once P3 was presented with the requested spike-facets of the data, she did not look for correlated events and health outcomes as she had anticipated, but instead saw things in the data that piqued her interest. First, she noticed very large spikes and wondered why those might be outliers. After reviewing individual spikes one-by-one she engaged in a closer comparative analysis about what different sensors within her home were reporting when a spike was detected.

This example illustrates the difference between what someone says they want to do, and what they actually do. Allowing our participants to engage directly with their data in flexible ways prompted unexpected analysis tasks as they actively participated in data analysis on the fly, learning new things about their data and prompting pivots in their analysis approaches. This scenario points to the need for flexible exploration in a visual analysis system to support serendipitous discoveries. The task elicitation interview method, through directly engaging participants with their personal data, provides rich insights into how unconstrained visual analysis may proceed in the wild and how we might design new tools to support those approaches, bridging the personal informatics analysis gap.

4.3.2 (Re)engaging participants with their data and the study. In line with other personal informatics projects [119], we found that annotations recorded by our participants during the system deployment were critical for sense-making and analysis [91]. At the end of the task elicitation interview with P4 she reflected on her experience of looking through her air quality data in conjunction with her annotations:

> **P4**: Annotations help you know what's going on because at this point I can't really remember what I was doing in 2017, but when you read the stuff, 'Oh yeah, that was the day that we tore out our front steps and I forgot to close the windows first.' And stuff like that. ...But especially when you're looking back this distance over this quantity [of data], it's hard to remember what it was doing. The annotations help you capture that.

In every task elicitation interview we conducted, the participants' annotations were vital for contextualizing the patterns and trends observed in the air quality data.

Despite the importance of annotation, we found during deployments that participants' annotation practices waxed and waned, occasionally ceasing completely until something new or unexpected occurred. During the task elicitation interviews, a number of our participants expressed frustration that they had not annotated more:

> **P3:** It kind of frustrates me. It would make me want to go back now, knowing that I could have it all. I would be more vigilant about making comments. Because I kind of get lax about it.

> **P6**: You know a lot of it is dependent on good sensors and then good data I'm inputting. Knowing how it can be done is going to motivate me to pay more attention to those [annotation prompts]... It's one of those things moving forward, it's increased my confidence in how I can use the data to make better decisions... knowing that the information is retrievable helps me and motivates me to want to provide more accurate data.

During the interviews, participants were frequently confronted with the importance of their annotations again and again, particularly when they saw interesting patterns in the data that they could not contextualize due to missing annotations, or interpret without them. Through directly engaging participants in the analysis process, they became more aware of the necessary steps and data to answer their question, and saw increased value of annotations. We speculate that these participants would increase their annotation practices after participating in the task elicitation interviews.

Task elicitation interviews may be an opportunity for personal informatics studies in other application areas that require participants to manually track data to encourage and motivate productive data-tracking practices.

Furthermore, providing participants an opportunity to get their personal questions answered during the interview may further engage them in a personal informatics study on the whole, and provide opportunities for researchers to acquire more feedback and buy-in in later stages of tool design and deployment. This engagement with personal questions was confirmed by P5 when he was asked to reflect on his experience with the interview:

> **P5**: Just being able to see it, especially interacting with the questions, was kind of cool. I thought it was kind of cool to be able to actually have a question, and look at it.

We argue that task elicitation interviews can be useful in a personal informatics study not only for providing important insights to the research team, but also for encouraging active and productive involvement by participants.

5 CHARACTERIZING ASTHMATIC'S GOALS AND QUESTIONS

User-centered indoor air quality sensing research primarily focuses on ways air quality data can engage users [69–72]. No studies exist for characterizing people's goals, motivations, or use-cases for engaging with an air quality sensing system. One reason for these lack of studies is that participant needs are diverse and difficult to design for. To address this difficulty, we analyze the A1, A2, and A3 research activities to characterize participant motivations for engaging with self-tracked indoor air quality data. We build on our previous work exploring *how* people use indoor air quality monitoring systems [91] to better understand *what* they want to do. This section presents a characterization of our participants' goals and motivations for engaging their indoor air quality. As discussed below, this work complements findings by Schroeder et al., whose research on self-tracking goals of migraine sufferers [110, 111] reports similar motivations. These similarities point to the potential for generalizablilty to other health-motivated self-tracking communities.

5.1 Participant Motivations

We identified five themes that capture participants' questions and goals evident in the collected data.

5.1.1 Answering Questions about Air Quality. Although all participants were already thinking about air quality in the context of their daily lives prior to joining our study, none had previous experience with air quality sensing systems or with tracking personal air quality data. As a result, the process of familiarizing themselves with their air quality through the deployed system generated several high-level questions. Underlying motivations behind questions in this category included establishing a basic understanding of their data and how the system operates – *What affects my sensor readings?*, orienting their perceptions of what is normal – *What is the general baseline reading for my home?*, how that normal level changes – *What time of year is worst for indoor air quality? what activities impact air quality?*, and comparing personal measurements with external conditions – *How does my air quality compare to DAQ*[3] *measurements? Is my air quality indoors better than outdoors [in general]?*.

Some questions in this theme are general enough to be addressed through a FAQ – *What kind of information can I get from my data? Do particulates rise with hot air?* – whereas others would require more interactive or context-specific explanations:

> **P3:** I know that "30k" is high because I can see that everything else is low, but I don't really have a reference to anything else. I can see this spike is "21k" and this one is "2k". Is "2k" a big spike or little spike? Giving me some sort of idea – like when you get blood tests, it could be low, but barely low and it's not a big deal [would be nice].

Participants had an intuitive understanding of what air quality measurements represented when reviewing their data, but they were less confident about details related to scales and units. Systems that anticipate and

[3]Utah Division of Air Quality, https://air.utah.gov/

design for these kinds of questions, either by contextualizing data representations with informational overlays or more intuitive units [41], will help users to get the most out of their system, and their data.

Takeaway: Questions and motivations within this category involve learning about air quality as its own end, and frequently answerable with nonpersonal data.

5.1.2 What's in My Control? Many participants wanted to review their air quality data to determine what aspects of their indoor air quality they could control – *How can this system help me understand what is within my control?*, wanted to receive personalized advice on knowing what or when to change – *Do I need to replace my windows or furnace?*, or wanted to know how their behaviors impact air quality – *What activity generates the most PM spikes in my home?*. For P3, the ability to track and review data is also a means to combat anxiety and helplessness.

> **P3:** I feel like there's so many things in his life that I just can't control. And his medical stuff is all kinda a mess. If there's one thing that I can do that will make something better, that's a huge motivation to me.

As the primary caregiver for a child suffering from asymptomatic asthma, P3 found the potential of what this system could provide especially powerful. Control, and the absence of control, is a strong motivator for families [31, 99], and designing flexible and personalized analytics tools can return control to people in the form of increased decision-making power.

Agency and control play a similar role in Schroeder, et al.'s "predicting and preventing migraines" goal category [110]. The control-seeking behavior we observe is similarly reflected in migraine sufferers' self-tracking goals to learn – and ultimately avoid – factors that trigger their condition. Observing this goal in separate chronic symptom communities helps validate this finding.

Takeaway: Questions and motivations within this category focus on determining or (re)establishing personal agency for improving indoor air quality or avoiding exposure to poor air quality.

5.1.3 Managing Symptoms & Health. It is not surprising that our asthmatic participants wanted to use their air quality data for managing health and asthma symptoms. Participants were interested in connecting their air quality data with their, or their child's, physical health, either as a means to track symptoms and sensitivities – *How are my asthma symptoms related to my indoor air quality?* – or to inform health-motivated decisions – *Can I anticipate days where I shouldn't go outside?*.

> **P5**: It's all about overlaying additional data to see if things correlate, especially if I can't figure out [what's causing me to feel bad]. Oftentimes I can't understand what's the trigger? What's happening? If that's apparent, that would be [huge].

> **P6:** if it turns out my symptoms are more correlated with low humidity, ozone, PM2.5, etc, then it would be nice to know when to telecommute or work from home those days, assuming my indoor air quality is better.

These motivations are also reflected in "Managing and monitoring migraines over time" self-tracking goal reported by Schroeder et al. [110]. Again, We observe direct overlap between both participant group's shared motivation to track symptom-related data for helping identify correlations with symptom-related factors.

Multiple participants (P3,P4,P5,P6) remarked how reviewing or exploring correlations in air quality data would have been useful when they or their children were newly diagnosed asthmatics:

> **P4:** [...] I feel like we kind of have a handle on a lot of the problems. But I think something like this [system] would've helped us figure it out faster.

> **P5:** If I were a little kid, I would have liked my mom to have had this when I was a kid to say, "This is why [P5] is a mess."

Systems combining user annotations, labeled activities, environmental conditions, and personal symptoms could support correlative searching and help users more quickly understand their conditions, especially for new diagnoses. Air quality sensing literature has already shown that next-generation platforms containing multiple gas sensors could infer particulate sources and offer targeted advice [38].

Takeaway: Questions and motivations within this category involve exploring what aspects of air quality influence or relate to health outcomes.

5.1.4 Social Sharing & Comparing. Participants wanted to share and compare their indoor air quality data with different people: between participants to understand where they stood, with doctors or specialists to assist with treatment or diagnosis, with government officials to affect change, and with the broader community for the greater good. Participants saw social comparison as a way of understanding how localized their air quality problems were and how many other people experience similar things:

> **P3:** It would be nice to have some kind of an air quality comparison. [...] based on my situation, and what's happening in my home, and what's happening in the area that I live in, is there something that I can do about the quality of my house? Is it really worse than my neighbors? And if it is, what can I do about it?

> **P4a:** I'd like to see where I was in comparison to other people. But in general, just everybody in the study, was it this day everyone had trouble? And then why?

P5 also discussed sharing their air quality with immediate family members in their own home to establishing accountability:

> **P5:** [Sharing data] is mostly getting after the people and what they're doing in the house.

P5 and P6 both see value in using their air quality data to communicate their conditions and severity with employers or other family members, similar to the "social recognition" goal category from Schroeder et al. [110]. Participants were also interested in sharing their air quality data and circumstances preceding respiratory issues with their doctor in hopes that the doctor would know how to interpret that data, consistent with other self-tracking literature [23, 110, 111]:

> **P2:** If I were able to bring this into a doctor and have my doctor say 'okay, I see this pattern, let's try this change in your treatment plan', that could be helpful.

Other more local applications involved using personal air quality information to share with elected officials as a way of advocating change, what Sakhnini et al. call an *advocacy goal* [109].

> **P6**: I definitely would want to show [my data] off. I think everybody in the valley has taken air quality to heart and if they were to see something [in the data] it hopefully would encourage them to take action. Write their legislators. I mean, session started today. It's like, 'Are you gonna do something about it this year?'

More broadly, participants were interested in sharing knowledge for the greater good, and interested to give or receive advice for improving air quality practices. Giving back to the community was important for all participants, and they were interested to log their data in the hopes their contributions would improve the lives of others.

> **P2:** Even if I never make any use of any of this, I get joy in just knowing I helped [authors] move along their path, whatever they're doing.

Prior work has explored how social sharing motivates engagement with an air quality system [71], including as a tool for advocacy or community-scale change [52].

Takeaway: Questions and motivations within this category involve sharing or comparing data with others.

5.1.5 Curiosity. Not all people approach personal data with specific goals in mind[36], and our participants were no different. Even when reminded to have a question prepared for their task elicitation interview, participants P2 and P4 were less interested in analyzing their data than exploring what the data had to say.

P2: I'd just be curious to see what our annotations are on the highest spikes...think you can do that?

Different terms signaled participants' curiosity, such as "review", "explore", or "just looking", and were primarily encountered in participants' self-described usage behaviors. By nature of A2 and A3, no participant explicitly identified tracking for curiosity as a concrete goal. This theme instead came about through observing how participants engaged in their activities.

Takeaway: Motivations within this category are primarily driven by exploration and lack any concrete question or analysis goal.

These five themes reflect aspects of self-tracking motivation reported in the quantified self [20] and chronic symptom management [110, 111] literature, with tie-ins to personal informatics models [36]. Our strongest thematic overlaps occur with migraine sufferers' self-tracking goals [110, 111] with the exception of the additional curiosity motivation. This is less surprising, however, given the specific goal-directed focus of that work [111].

Even though each of these tracking communities can be characterized as health-motivated and driven to understand and improve physical condition in various ways, this diverse overlap lends support for the classification provided above, and suggests that the categories may generalize to other sensitive populations within personal informatics.

6 INDOOR AIR QUALITY DATA MODEL

The standard approach for visualizing univariate sensor data frequently involves plotting measurements within a fixed time-span line chart. In the context of indoor air quality, this visualization technique is the standard view for many commercially available systems and most research platforms [13, 38, 67, 70–72, 91], with few exceptions [14, 58, 118]. This approach is sufficient for immediate or narrow explorations, such as comparing multiple data streams over a small time span, but is otherwise poorly suited for reviewing noncontinuous events — *show me every time I vacuum* — or longitudinal summaries — *what's the monthly average air quality for the past year?* These types of questions, which reflect the timescales required for answering many of our participants' questions, stray from the linear evaluations that the standard visualization approaches support. In response, we add to a growing body of work rejecting one-dimensional personal informatics analytic tools [5, 60, 61] and rethink the underlying data model that governs how we design for personal data.

Prior work by Epstein et al. proposes visual *cuts* as a method for helping self-trackers identify more meaningful or actionable insights from self-tracked data [33]. A cut is a subset of the data that shares a common feature, faceting or filtering the data for supporting meaningful comparisons. This approach addresses the analytic limitations of linear, fixed-length views by using shared features within tracked data to group and process information for generating more engaging and informed analytic views. Epstein et al. report cuts to be useful and effective for supporting self-tracking goals, yet found no correlation between participants' preferred cuts and the goals the research team designed specific cuts to support. Participants instead preferred reviewing several cuts at once to receive a more complete picture of their data, leading the researchers to recommend *against* building interfaces solely around fixed design goals and to instead "present a swath of cuts rather than relying on one to summarize activity" [33]. One solution is to create flexible interfaces that support user-generated cuts[19, 89], but designing sophisticated analysis systems capable of supporting individualized self-tracking goals remains a difficult and open problem [61, 111].

To address this problem, we recast cuts into the more general class of *comparative targets* over which analytic tasks such as correlation and finding differences are performed [42]. In this way, cuts become a specific application of broader work within visualization research dealing with comparison [42, 43, 57, 75, 77, 95, 116, 123]. Using the visualization comparison framework of Gleicher [42], we deconstructed our participants' questions into comparative targets in order to abstract away specific themes (Section 5), instead focusing on grouping questions by their common target properties. For example, the question *"what is the air quality like on days I am sick?"* generalizes to a target set of 24-hour subsets of air quality data for each day with a "sick" annotation label. Applying concepts from the visualization comparison literature allows us to draw from research into perceptual underpinnings of visual comparison [57, 95], design principles for developing flexible and effective visual analysis systems that support comparison [43, 75, 77, 116], and the challenges of comparison at scale [42, 123].

Based on this comparison reframing, we propose a data model for personal air quality data that consists of three comparative time span categories (Table 1): **direct comparisons** taking place within a shared window of time –*Is my air quality indoors better than outdoors right now?*; **distributed comparisons** involving targets independently distributed through time – *What is my indoor air quality like on 'red air' days?*; and **longitudinal comparisons** requiring the entire time span to answer –*How is my air quality changing over time?*. These categories present multiple design challenges to managing the number and nature of diverse data streams, along with optimizing how information is organized and presented over each time span. The following sections describe these categories in greater depth and outline design perspectives for engaging personal data in each case. We also present a complementary data processing pipeline to provide preliminary suggestions on ways tool builders can improve visual analysis tools for self-tracked data.

6.1 Visualization Comparison Framework

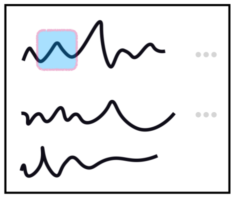
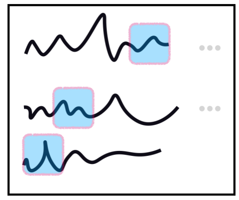
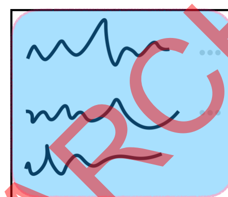
Our proposed data model derives from the high-level design considerations of Gleicher's visualization comparison framework [42]. In this work, comparison is defined as tasks that involve "a set of *targets* (i.e, the set of items being compared), and an *action* performed on the relationship (e.g. similarities or differences) among these targets". The core contributions of this framework provide a thorough exposition of the challenges inherent in supporting visual comparison tasks, principally those requiring careful matching of visual design techniques and analytic strategies for combating scalability challenges. These challenges can arise from comparing too many targets, overly large or complex targets, or overly large or complex target relationships. Broad strategies for mitigating these issues involve different ways of reducing the number of targets the user reviews. Suggested solutions include presenting data in a linear way for *sequential scanning*, reducing data through filtering or selection to define a *subset*, or abstracting the underlying data through *summarization* or other encoding approaches.

Applying good visual design principles can also reduce the cognitive burden associated with scalability challenges. Based on a survey of over 100 papers, Gleicher et al. identify three techniques for visualizing comparison [43]: a *juxtaposed* small multiple approach where elements are placed side by side; *superposition* of data to provide comparisons via overlays; and *explicit encoding* to visualize a relationship between targets rather than the targets themselves. These design considerations define a fundamental tradespace for informing potential design decisions when building tools for analyzing self-tracked and other personal informatics data. We will refer to these strategies and techniques when discussing our own considerations for designing around the direct, distributed, and longitudinal question categories.

6.2 Data Model

This section describes the three question categories of our proposed indoor air quality data model (Table 1). We examine individual design challenges and opportunities for each comparative category, and offer potential design considerations. We also outline a complementary visual analytics processing pipeline similar to the well-known data state reference model [12, 15]. Whereas many processing pipelines exist in the visualization literature [124], the pipeline we propose is developed in conjunction with the indoor air quality data model to outline a

Table 1. A summary of design considerations and visualization suggestions for the direct, distributed, and longitudinal comparison categories.

	Direct	Distributed	Longitudinal
			
Description	Evaluating data collected in the same moment in time	Evaluating data over separate or distinct moments	Evaluating data over the entire time span
Gist	What is happening in a moment?	How do things relate?	How do things look over time?
Time spans	Shared window	Independently sized subsets	Full time span
Targets	Data streams	Windowed subsets, groups, or summaries	Sequential, summarized time facets
Comparison	Between targets	Between targets	Across targets
Challenge	Number of targets and visual complexity	Number of targets, temporal scale	Temporal scale
Strategy	Good visual design principles	Grouping & summarization	Temporal facets, aggregation & summarization

system-level design for engaging personal air quality data from multiple perspectives. We speculate that both the data model and processing pipeline may transfer to other personal informatics application areas.

6.2.1 Direct Comparisons. The **direct comparison** category involves comparisons of different data streams collected within a short period of time. Most current visualization systems designed for indoor air quality projects support this category of comparison using a fixed-time span line chart [13, 38, 67, 70–72, 91]. For indoor air quality applications, comparative targets for this category can include any number of individual air quality data streams, plus other contextualizing data sources such as user annotations and asthma data within the time period of interest. Analytic tasks for questions in this category relate to comparing targets within the same instant in time, or over a short time period, for instance *At any given moment, is the inside air better, or is the outside air better?*. The gist of the questions in this category is the desire to know how things are in a moment.

Scalability challenges for this category primarily involve managing the complexity of comparing too many targets. Because direct comparison involves between-target analysis over a shared time span, all three visual design patterns for comparison can apply[43]. The choice of which to use will depend on the specific analytic goals and sound visualization design principles. Tool builders should consider supporting users to interactively load and overlay individual data streams in a shared view to support direct comparison tasks. Limiting views to 3-5 data streams at a time would help minimize visual clutter and avoid overloading plot axes.

6.2.2 Distributed Comparisons. The **distributed comparison** category includes a comparison of subsets of the data streams distributed over separate or distinct moments in time but that share some common features. This category is similar to personal informatics cuts [33]. For indoor air quality applications, comparative targets for this category can include windowed subsets of data, groups, or data summaries of air quality sensor measurements. Analytic tasks for questions in this category relate to analyzing target set items that match some selection criteria, such as *What is my indoor air quality like on 'red air' days?* The overall gist of questions in this category is exploring how separate items relate.

Scalability challenges primarily involve managing the complexity of the total number of targets items, the timescales spanned by the target items, or both. Because direct comparison involves between-target analysis across independent – and independently windowed – data subsets, design approaches that support juxtaposed views, with additional affordances to arrange, group, or summarize targets by derived or categorical properties, can help to minimize cognitive burden. Depending on the analytic task, target items may need to be further grouped and summarized, either to reduce complexity or provide additional information. For example, the question *How does air quality compare between different activities?* would involve filtering data according to annotated activity classes for cross-class comparison, and optionally further summarize each activity class to simplify comparisons between them. To support distributed comparisons, tool builders should design for users to interactively select and modify filtering criteria for defining comparative targets on a data set-by-data set basis. This approach generalizes the researcher-led design process of visual cuts [33] to empower the user to explore and review their data on their terms, without designer assistance.

6.2.3 Longitudinal Comparisons. The **longitudinal comparison** category includes comparisons with ill-defined targets. Like the direct comparison category, longitudinal comparison targets can include any number of individual air quality data streams, plus other contextualizing data sources. Analytic tasks for questions in this category require the entire data set to answer, either because questions have extensive scope — *what time of year is worst for indoor air quality?* — or are ill-defined — *what can I do to improve air quality in my house?*. Questions in this category have ill-defined target time-spans, and thus require users to dynamically evaluate the data to determine both the targets and the results of their comparison. We recommend providing access to the entire data set through an overview-first design approach [114] that meaningfully summarizes the data in the overview, such as using the number of air quality spikes per day or week. We found that questions from our study in this category relate to reviewing how things change over time.

Scalability challenges primarily involve managing temporal scale. Because longitudinal comparison involves longer time spans, design approaches should support sequential, summarized time facets with additional affordances to arrange, group, or summarize targets by derived or categorical properties. To support longitudinal comparisons, tool builders should design for users to interactively specify their date ranges and temporal faceting criteria [89]. Users should also have control over how data are returned, such as the ability to define facet arrangement, clustering, or summarization properties. For air quality data, this may include faceting and summarizing data to report daily indoor averages. This information can be presented as small multiples, joined sequentially in a calendar view, or additionally overlaid with other information to review correlations, such as reviewing daily indoor averages against outdoor EPA air quality classifications.

6.3 Visual analytics pipeline

Evaluating the various design challenges and strategies resulting from our indoor air quality data model led us to create a tailored data analysis pipeline for processing participants' air quality questions (Figure 4) that could be used as a system design backbone for developing new personal informatics visual analysis tools. This pipeline focuses on specific data handling considerations for developing user-friendly visual analysis tools. Beginning with data collection, we outline various stages of the air quality analytics process: data storage, target selection, target processing, and presentation. This pipeline is similar to the data state reference model [12], which is used extensively by visualization designers to inform them of the challenges and opportunities for crafting interactive visualizations. Just as the data state reference model helped articulate the transition of control from designer to practitioners, the pipeline we present here can help transition analytic control from practitioners to users.

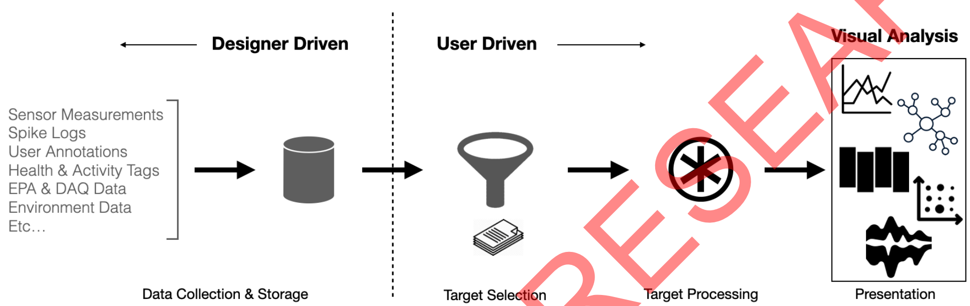


Fig. 4. Analytics pipeline for personal data.

6.3.1 Data collection & storage. Identifying and collecting relevant data streams is an essential first step in any analytics system, and a well-designed system will be capable of seamlessly integrating a variety of sources. Depending on the problem domain or use-case, these data may come from any number of sensor measurements, self-reports, survey data, and social media, or may be imported from other databases or external services. In the context of personal indoor air quality, this might include: individual deployment air quality data streams and user annotations; external streams from a variety environmental sources such as weather and pollen counts; personal health resources such as symptom reports, and asthma surveys; or even anonymized data streams from other study participants such as average readings and spike counts. Tool builders should also incorporate data preprocessing techniques to help catch and fix common errors with real-world data to help facilitate efficient analysis [106].

6.3.2 Target Selection. The first step toward offering a personalized analysis tool relies on allowing end-users to select the data they want to analyze. The interactive capabilities we outlined for the data model describe a variety of methods for choosing filter criteria to select comparative targets. The target selection stage should support users to select any number of available data streams, and choose from an interactive list of automatically populated filtering criteria or categorical parameters afforded by their data selection. The ability to flexibly define target selection criteria allows people to move beyond restrictive one-dimensional analytic systems and customize the data retrieval process to fit their needs and goals. Tool builders should design interactive mechanisms that support people to intuitively load and filter their data streams and dynamically populate additional filtering criteria for further refinement based on underlying target set properties. These filtering options combat scalability challenges by reviewing large or complex data sets, and empower nonexpert users to find relevant information more quickly, and on their own.

6.3.3 Target Processing. After selecting data and applying filtering criteria, the resulting target set may require additional processing. The target processing stage allows users to further group or summarize target data, either to combat scalabilty challenges (Section 6.2) or to prepare data for presentation. Designers can provide additional grouping options based on temporal hierarchies [19, 89] or underlying data features in the target set [33].

6.3.4 Presentation. After loading, filtering, and processing target data, the targets are presented using interactive visualization techniques. The previous data model discussion highlights three comparative categories designers can use to focus their design considerations. Each category's unique scalability challenges require different design strategies to effectively present target information. Providing generalized design recommendations for interactive interfaces is outside the scope of this work, but we include three interface prototypes in supplemental materials to illustrate the concepts discussed in this work.

7 DISCUSSION

Manual self-tracking is burdensome, and users are often reluctant to do it [16]. One observation during the task elicitation interviews was that participants became re-engaged with their data and expressed regret at not having recorded more annotations, or different annotations, during the data collection. This suggests several interesting opportunities for motivating better manual data collection. In the future, flexible analytic tools that do not exist yet, and are the motivation of this work, can help users see the value of annotating their data and can help them decide what kinds of things they want to annotate. User awareness of these tools and the ability to experiment with them early could thus motivate targeted manual data collection, such as users' annotations in this work. Tools can be designed to explicitly facilitate this use-case of deciding what to track, similar to the goal scaffolding idea by Schroeder, et al. [111]. However, until such applications are available, task elicitation interviews can be a useful tool for motivating annotations and other manual self-tracking data across a broad range of personal informatics contexts. For this reason, conducting a task elicitation interview in the middle of a data collection process could serve both to re-engage participants who may be losing interest, and to motivate data collection because they see what kind of questions they would be able to answer if they collect the right data.

Recasting the personal informatics notion of *cuts* [33] as the more general visualization notion of *comparative targets* [42] offered us a new perspective for designing a flexible data model that can address the diverse set of questions participants are interested in addressing with their personal air quality data. Although this data model comes from analyzing indoor air quality deployments, we speculate it can generalize to other personal informatics application areas that involve time-based data, such as chronic symptom management, personal fiance, energy consumption, and emotional wellness, to name a few. Furthermore, the visualization literature on comparison is rich and growing due to the fundamental nature of comparison in data analysis tasks [42]. This literature provides a wealth of insights into the perceptual processes behind visual comparison tasks [57, 95], and technique and system design for supporting a broad range of visual comparison tasks and data types [43, 75, 77, 116, 123]. There is an opportunity to address the personal informatics analysis gap through consideration of this literature for designing the next generation of visual analysis tools in a broad range of personal informatics application areas.

Before developing the task elicitation interview, we lacked a mental model for understanding how nonexpert users would approach solving their problems, and even what they would find understandable or intuitive to do. The tendency for domain expert collaborations in visualization research has overlooked the nonexpert user population in terms of how they engage with data, leaving them understudied and underserved as a population. Personal informatics' goal-oriented tool design ethos has overlooked opportunities for users who engage with data to freely explore rather than having concrete goals in mind. Both of these blind spots have given rise to the personal informatics analysis gap. The task elicitation interview method offers a first step for exploring this space by allowing us to reconstruct how nonexpert users approach the problem-solving process as it unfolds during the interview. We plan to analyze this information into individual analysis arcs for characterizing and comparing how participants approached operationalizing their problems. Constructing these arcs consists of a close read of the interview transcript to identify where participants engage with their data and making note of relative locations, analytic tasks undertaken or requested, and task durations. Analytic tasks could then be coded according to established visual analytics task typologies [10], and visualized as timelines for comparing results between participants.

Participants' goals and questions that emerged in Section 5 are similar to those reported in past work on self-tracking and personal informatics, suggesting that a domain-agnostic classification of these categories may exist. How general such a classification could be is unclear — would more generalized wordings for these categories also cover domains as broad as diet tracking, time management, and personal finance, or are they limited to contexts where users are self-tracking in the context of medical conditions? Are there other constraints? We have an opportunity to look across these domains and do a broad literature search and meta-analysis to explore the potential of a generalized classification. If such a classification does generalize across multiple domains, other important follow-up issues to explore in this direction include: whether the classification is useful for predicting the kinds of questions that people have in a particular domain, or is it only useful in categorizing data that has been collected? Can we design for these categories of questions and goals, or are the specific questions and goals that users have within a context necessary to elicit in order to develop tolls to bridge the personal informatics analysis gap? These questions, and the potential for a meta-analysis in this domain, are ready opportunities for future work.

8 CONCLUSION

We present a study that probes into the types of questions people want to ask from their personal air quality data, and how they might go about answering those questions from their data. Spanning three separate research activities, this study explores the analysis gap in the ecosystem of personal informatics tools for enabling people to ask personal and varied questions from their data. Our findings outline a characterization of the range of question types our participants pose concerning their indoor air quality, a new interview method for eliciting analysis approaches and insights from nonexpert users, and a data model for processing indoor air quality data that supports designing for a rich and broad set of participant questions. These results provide a foundation for designing new visual analysis tools for indoor air quality that can support a broad range of questions and tasks that people may ask. Although the results of this work come from studying and designing for indoor air quality and asthmatics, we speculate these themes will generalize more broadly to other personal informatics application areas. Areas for future work pointed to by this study include applying visualization comparison models to personal informatics design challenges, conducting meta-analyses of goals and questions across other self-tracking application areas, and using task elicitation interviews to develop a richer understanding of how people with limited data analysis experience and skills might answer their questions from data. We hope this work can inspire these exciting research directions

REFERENCES

[1] Alspaugh, S., Zokaei, N., Liu, A., Jin, C., and Hearst, M. A. Futzing and Moseying: Interviews with Professional Data Analysts on Exploration Practices. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 22–31.

[2] Amar, R. A., and Stasko, J. T. Knowledge precepts for design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics 11*, 4 (2005), 432–442.

[3] Arias-Hernandez, R., Kaastra, L. T., Green, T. M., and Fisher, B. Pair Analytics: Capturing Reasoning Processes in Collaborative Visual Analytics. In *2011 44th Hawaii International Conference on System Sciences* (jan 2011), IEEE, pp. 1–10.

[4] Bambino, D. Critical friends. *Educational leadership 59*, 6 (2002), 25–27.

[5] Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., Price, E., Catrambone, R., and Wilson, J. Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. *ACM Transactions on Computer-Human Interaction 20*, 5 (2013), 1–27.

[6] Bigelow, A., Nobre, C., Meyer, M., and Lex, A. Origraph: interactive network wrangling. In *2019 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2019), IEEE, pp. 81–92.
[7] Börner, K., Maltese, A., Balliet, R. N., and Heimlich, J. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization 15*, 3 (2016), 198–213.
[8] Bostock, M., Ogievetsky, V., and Heer, J. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics 17*, 12 (2011), 2301–2309.
[9] Boy, J., Rensink, R. A., Bertini, E., and Fekete, J.-D. A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics 20*, 12 (2014), 1963–1972.
[10] Brehmer, M., and Munzner, T. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics 19*, 12 (2013), 2376–2385.
[11] Campbell, T., Larson, E., Cohn, G., Froehlich, J., Alcaide, R., and Patel, S. N. Wattr: A method for self-powered wireless sensing of water activity in the home. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing* (New York, NY, USA, 2010), UbiComp '10, ACM, pp. 169–172.
[12] Card, S. K. Information Visualization. In *Readings in information visualization : using vision to think*, 1 ed. 1999, ch. 1, pp. 1–34.
[13] Chen, X., Zheng, Y., Chen, Y., Jin, Q., Sun, W., Chang, E., and Ma, W.-y. Indoor air quality monitoring for smart building. *UbiComp* (2014), 2–6.
[14] Cheng, Y., Li, X., Li, Z., Jiang, S., Li, Y., Jia, J., and Jiang, X. Aircloud: a cloud-based air-quality monitoring system for everyone. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems* (2014), ACM, pp. 251–265.
[15] Chi, E. H.-h. A taxonomy of visualization techniques using the data state reference model. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings* (2000), IEEE, pp. 69–75.
[16] Choe, E. K., Abdullah, S., Rabbi, M., Thomaz, E., Epstein, D. A., Cordeiro, F., Kay, M., Abowd, G. D., Choudhury, T., Fogarty, J., Lee, B., Matthews, M., and Kientz, J. A. Semi-Automated Tracking: A Balanced Approach for Self-Monitoring Applications. *IEEE Pervasive Computing 16*, 1 (jan 2017), 74–84.
[17] Choe, E. K., Lee, B., et al. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications 35*, 4 (2015), 28–37.
[18] Choe, E. K., Lee, B., Kay, M., Pratt, W., and Kientz, J. A. Sleeptight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), ACM, pp. 121–132.
[19] Choe, E. K., Lee, B., Zhu, H., Riche, N. H., and Baur, D. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (2017), ACM, pp. 173–182.
[20] Choe, E. K., Lee, N. B., Lee, B., Pratt, W., and Kientz, J. A. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2014), ACM, pp. 1143–1152.
[21] chul Kwon, B., Fisher, B., and Yi, J. S. Visual analytic roadblocks for novice investigators. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2011), IEEE, pp. 3–11.
[22] Chung, C.-F., Agapie, E., Schroeder, J., Mishra, S., Fogarty, J., and Munson, S. A. When personal tracking becomes social: Examining the use of instagram for healthy eating. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), pp. 1674–1687.
[23] Chung, C.-F., Dew, K., Cole, A., Zia, J., Fogarty, J., Kientz, J. A., and Munson, S. A. Boundary negotiating artifacts in personal informatics: Patient-provider collaboration with patient-generated data. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (2016), pp. 770–786.
[24] Chung, C.-F., Wang, Q., Schroeder, J., Cole, A., Zia, J., Fogarty, J., and Munson, S. A. Identifying and planning for individualized change: Patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies 3*, 1 (2019), 7.
[25] Consolvo, S., Everitt, K., Smith, I., and Landay, J. A. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), pp. 457–466.
[26] Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2008), pp. 1797–1806.
[27] Correll, M. Ethical dimensions of visualization research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), ACM, p. 188.
[28] Cowley, P., Nowell, L., and Scholtz, J. Glass box: An instrumented infrastructure for supporting human interaction with information. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (2005), IEEE, pp. 296c–296c.
[29] Dahlbäck, N., Jönsson, A., and Ahrenberg, L. Wizard of oz studies—why and how. *Knowledge-based systems 6*, 4 (1993), 258–266.
[30] Danis, C. M., Viegas, F. B., Wattenberg, M., and Kriss, J. Your place or mine? visualization as a community component. In *Proceedings*

of the SIGCHI conference on Human factors in computing systems (2008), pp. 275–284.

[31] Davidoff, S., Lee, M. K., Yiu, C., Zimmerman, J., and Dey, A. K. Principles of smart home control. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 4206 LNCS* (2006), 19–34.

[32] Dickson, J., McLennan, J., and Omodei, M. M. Effects of concurrent verbalization on a time-critical, dynamic decision-making task. *The Journal of general psychology 127*, 2 (2000), 217–228.

[33] Epstein, D., Cordeiro, F., Bales, E., Fogarty, J., and Munson, S. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 conference on Designing interactive systems* (2014), ACM, pp. 667–676.

[34] Epstein, D. A. Everyday Personal Informatics. *ProQuest Dissertations and Theses* (2018), 173.

[35] Epstein, D. A., Caraway, M., Johnston, C., Ping, A., Fogarty, J., and Munson, S. A. Beyond abandonment to next steps: understanding and designing for life after personal informatics tool use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 1109–1113.

[36] Epstein, D. A., Ping, A., Fogarty, J., and Munson, S. A. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2015), ACM, pp. 731–742.

[37] Estrin, D. Small data, where n= me. *Communications of the ACM 57*, 4 (2014), 32–34.

[38] Fang, B., Xu, Q., Park, T., and Zhang, M. Airsense: an intelligent home-based sensing system for indoor air quality analytics. In *Proceedings of the 2016 ACM International joint conference on pervasive and ubiquitous computing* (2016), pp. 109–119.

[39] Fisher, D., DeLine, R., Czerwinski, M., and Drucker, S. Interactions with big data analytics. *interactions 19*, 3 (2012), 50–59.

[40] Fonteyn, M. E., Kuipers, B., and Grobe, S. J. A description of think aloud method and protocol analysis. *Qualitative health research 3*, 4 (1993), 430–441.

[41] Froehlich, J., Patel, S., Landay, J. A., Findlater, L., Ostergren, M., Ramanathan, S., Peterson, J., Wragg, I., Larson, E., Fu, F., and Bai, M. The design and evaluation of prototype eco-feedback displays for fixture-level water usage data. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (2012), 2367.

[42] Gleicher, M. Considerations for visualizing comparison. *IEEE transactions on visualization and computer graphics 24*, 1 (2017), 413–423.

[43] Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., and Roberts, J. C. Visual comparison for information visualization. *Information Visualization 10*, 4 (2011), 289–309.

[44] Grammel, L., Tory, M., and Storey, M.-A. How information visualization novices construct visualizations. *IEEE transactions on visualization and computer graphics 16*, 6 (2010), 943–952.

[45] Gupta, S., Reynolds, M., and Patel, S. ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home. *Proceedings of the 12th ACM international conference on Ubiquitous computing* (2010), 139–148.

[46] Haddadi, H., and Brown, I. Quantified self and the privacy challenge. *Technology Law Futures 6* (2014).

[47] Heer, J., and Boyd, D. Vizster: Visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* (2005), IEEE, pp. 32–39.

[48] Heer, J., Van Ham, F., Carpendale, S., Weaver, C., and Isenberg, P. Creation and collaboration: Engaging new audiences for information visualization. In *Information visualization*. Springer, 2008, pp. 92–133.

[49] Heer, J., Viégas, F. B., and Wattenberg, M. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM 52*, 1 (2009), 87–97.

[50] Hekler, E. B., Klasnja, P., Chevance, G., Golaszewski, N. M., Lewis, D., and Sim, I. Why we need a small data paradigm. *BMC medicine 17*, 1 (2019), 1–9.

[51] Hinrichs, U., Forlini, S., and Moynihan, B. Speculative Practices: Utilizing InfoVis to Explore Untapped Literary Collections. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 429–438.

[52] Hsu, Y.-C., Dille, P., Cross, J., Dias, B., Sargent, R., and Nourbakhsh, I. Community-empowered air quality monitoring system. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), pp. 1607–1619.

[53] Huang, D., Tory, M., Aseniero, B. A., Bartram, L., Bateman, S., Carpendale, S., Tang, A., and Woodbury, R. Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics 21*, 3 (2014), 420–433.

[54] Huron, S., Carpendale, S., Thudt, A., Tang, A., and Mauerer, M. Constructive visualization. In *Proceedings of the 2014 conference on Designing interactive systems* (2014), ACM, pp. 433–442.

[55] Jain, R. K., Taylor, J. E., and Peschiera, G. Assessing eco-feedback interface usage and design to drive energy efficiency in buildings. *Energy and buildings 48* (2012), 8–17.

[56] Jansen, Y., Dragicevic, P., Isenberg, P., Alexander, J., Karnik, A., Kildal, J., Subramanian, S., and Hornbæk, K. Opportunities and Challenges for Data Physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (New York, New York, USA, 2015), ACM Press, pp. 3227–3236.

[57] Jardine, N., Ondov, B. D., Elmqvist, N., and Franconeri, S. The perceptual proxies of visual comparison. *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (2019), 1012–1021.

[58] Jiang, Y., Li, K., Tian, L., Piedrahita, R., Yun, X., Mansata, O., Lv, Q., Dick, R. P., Hannigan, M., and Shang, L. Maqs: A personalized mobile sensing system for indoor air quality monitoring. In *Proceedings of the 13th International Conference on Ubiquitous Computing*

(New York, NY, USA, 2011), ACM, pp. 271–280.
[59] Jin, M., Bekiaris-Liberis, N., Weekly, K., Spanos, C., and Bayen, A. Sensing by proxy: Occupancy detection based on indoor co2 concentration. *UBICOMM 2015 14* (2015).
[60] Jones, S. L. Exploring correlational information in aggregated quantified self data dashboards. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (2015), ACM, pp. 1075–1080.
[61] Jones, S. L., and Kelly, R. Dealing with information overload in multifaceted personal informatics systems. *Human–Computer Interaction 33*, 1 (2018), 1–48.
[62] Kandel, S., Paepcke, A., Hellerstein, J., and Heer, J. Wrangler: Interactive visual specification of data transformation scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 3363–3372.
[63] Kandel, S., Paepcke, A., Hellerstein, J. M., and Heer, J. Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2917–2926.
[64] Kandogan, E., Balakrishnan, A., Haber, E. M., and Pierce, J. S. From data to insight: work practices of analysts in the enterprise. *IEEE computer graphics and applications 34*, 5 (2014), 42–50.
[65] Kay, M., Choe, E. K., Shepherd, J., Greenstein, B., Watson, N., Consolvo, S., and Kientz, J. A. Lullaby: a capture & access system for understanding the sleep environment. In *Proceedings of the 2012 ACM conference on ubiquitous computing* (2012), pp. 226–234.
[66] Kerzner, E., Goodwin, S., Dykes, J., Jones, S., and Meyer, M. A Framework for Creative Visualization-Opportunities Workshops. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (jan 2019), 748–758.
[67] Kim, J.-Y., Chu, C.-H., and Shin, S.-M. Issaq: An integrated sensing systems for real-time indoor air quality monitoring. *IEEE Sensors Journal 14*, 12 (2014), 4230–4244.
[68] Kim, M., Zimmermann, T., DeLine, R., and Begel, A. The emerging role of data scientists on software development teams. In *Proceedings of the 38th International Conference on Software Engineering* (2016), ACM, pp. 96–107.
[69] Kim, S., and Li, M. Awareness, understanding, and action: A conceptual framework of user experiences and expectations about indoor air quality visualizations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–12.
[70] Kim, S., and Paulos, E. inair: measuring and visualizing indoor air quality. In *Proceedings of the 11th international conference on Ubiquitous computing* (2009), pp. 81–84.
[71] Kim, S., and Paulos, E. Inair: sharing indoor air quality measurements and visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), pp. 1861–1870.
[72] Kim, S., Paulos, E., and Mankoff, J. inair: a longitudinal study of indoor air quality measurements and visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), pp. 2745–2754.
[73] Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J. B., Grout, J., Corlay, S., et al. Jupyter notebooks-a publishing format for reproducible computational workflows. In *ELPUB* (2016), pp. 87–90.
[74] Kwon, B. C., Stoffel, F., Jäckle, D., Lee, B., and Keim, D. Visjockey: Enriching data stories through orchestrated interactive visualization. In *Poster Compendium of the Computation+ Journalism Symposium* (2014), vol. 3.
[75] Law, P.-M., Basole, R. C., and Wu, Y. Duet: Helping data analysis novices conduct pairwise comparisons by minimal specification. *IEEE transactions on visualization and computer graphics 25*, 1 (2018), 427–437.
[76] Lenzner, T., Neuert, C., and Otto, W. Cognitive pretesting. *GESIS Survey Guidelines* (2016), 3.
[77] Li, G., Zhang, Y., Dong, Y., Liang, J., Zhang, J., Wang, J., McGuffin, M. J., and Yuan, X. Barcodetree: Scalable comparison of multiple hierarchies. *IEEE transactions on visualization and computer graphics 26*, 1 (2019), 1022–1032.
[78] Li, I., Dey, A., and Forlizzi, J. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2010), ACM, pp. 557–566.
[79] Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G., and Strub, H. B. Fish'n'steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing* (2006), Springer, pp. 261–278.
[80] Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., and Zimmerman, J. I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2011), pp. 2409–2418.
[81] Liu, M.-J., Wang, X.-f., and Huang, Y.-l. Data preprocessing in data mining. *Computer science 27* (2000), 54–57.
[82] Mackinlay, J. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog) 5*, 2 (1986), 110–141.
[83] Mackinlay, J., Hanrahan, P., and Stolte, C. Show me: Automatic presentation for visual analysis. *IEEE transactions on visualization and computer graphics 13*, 6 (2007), 1137–1144.
[84] McCurdy, N., Lein, J., Coles, K., and Meyer, M. Poemage: Visualizing the sonic topology of a poem. *IEEE transactions on visualization and computer graphics 22*, 1 (2015), 439–448.
[85] McKenna, S., Henry Riche, N., Lee, B., Boy, J., and Meyer, M. Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences. *Computer Graphics Forum 36*, 3 (2017), 377–387.

[86] Meyer, M., Munzner, T., DePace, A., and Pfister, H. Multeesum: a tool for comparative spatial and temporal gene expression data. *IEEE transactions on visualization and computer graphics 16*, 6 (2010), 908–917.
[87] Meyer, M., Munzner, T., and Pfister, H. Mizbee: a multiscale synteny browser. *IEEE transactions on visualization and computer graphics 15*, 6 (2009), 897–904.
[88] Meyer, M., Wong, B., Styczynski, M., Munzner, T., and Pfister, H. Pathline: A tool for comparative functional genomics. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 1043–1052.
[89] Miranda, F., Lage, M., Doraiswamy, H., Mydlarz, C., Salamon, J., Lockerman, Y., Freire, J., and Silva, C. T. Time lattice: A data structure for the interactive visual analysis of large time series. In *Computer Graphics Forum* (2018), vol. 37, Wiley Online Library, pp. 23–35.
[90] Moere, A. V. Towards Designing Persuasive Ambient Visualization.
[91] Moore, J., Goffin, P., Meyer, M., Lundrigan, P., Patwari, N., Sward, K., and Wiese, J. Managing in-home environments through sensing, annotating, and visualizing air quality data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)(Ubicomp '18) 2*, 3 (Sept 2018).
[92] Munzner, T., Johnson, C., Moorhead, R., Pfister, H., Rheingans, P., and Yoo, T. S. Nih-nsf visualization research challenges report summary. *IEEE Computer Graphics and Applications 26*, 2 (2006), 20–24.
[93] National Institute of Biomedical Imaging Bioengineering. Pediatric Research Using Integrated Sensor Monitoring Systems, 2015.
[94] Nielsen, C. B., Jackman, S. D., Birol, I., and Jones, S. J. Abyss-explorer: visualizing genome sequence assemblies. *IEEE transactions on visualization and computer graphics 15*, 6 (2009), 881–888.
[95] Ondov, B., Jardine, N., Elmqvist, N., and Franconeri, S. Face to face: Evaluating visual comparison. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2018), 861–871.
[96] Patterson, E. S., Roth, E. M., and Woods, D. D. Predicting vulnerabilities in computer-supported inferential analysis under data overload. *Cognition, Technology & Work 3*, 4 (2001), 224–237.
[97] Peck, E. M., Ayuso, S. E., and El-Etr, O. Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), ACM, p. 244.
[98] Peesapati, S. T., Schwanda, V., Schultz, J., Lepage, M., Jeong, S.-y., and Cosley, D. Pensieve: supporting everyday reminiscence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), pp. 2027–2036.
[99] Pina, L. R., Sien, S.-W., Ward, T., Yip, J. C., Munson, S. A., Fogarty, J., and Kientz, J. A. From personal informatics to family informatics: Understanding family practices around health monitoring. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (2017), pp. 2300–2315.
[100] Pirolli, P., and Card, S. Information foraging. *Psychological review 106*, 4 (1999), 643.
[101] Pirolli, P., and Card, S. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (2005), vol. 5, McLean, VA, USA, pp. 2–4.
[102] Pousman, Z., Stasko, J., and Mateas, M. Casual information visualization: Depictions of data in everyday life. *IEEE transactions on visualization and computer graphics 13*, 6 (2007), 1145–1152.
[103] Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., and Singer, E. Methods for testing and evaluating survey questions. *Public opinion quarterly 68*, 1 (2004), 109–130.
[104] Pyle, D. *Data preparation for data mining*. morgan kaufmann, 1999.
[105] Quinan, P. S., and Meyer, M. Visually comparing weather features in forecasts. *IEEE transactions on visualization and computer graphics 22*, 1 (2015), 389–398.
[106] Rahm, E., and Do, H. H. Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull. 23*, 4 (2000), 3–13.
[107] Rodgers, J., and Bartram, L. Exploring Ambient and Artistic Visualization for Residential Energy Use Feedback. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (dec 2011), 2489–2497.
[108] Russell, D. M., Stefik, M. J., Pirolli, P., and Card, S. K. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (1993), pp. 269–276.
[109] Sakhnini, N., Yu, J. E., Jones, R. M., and Chattopadhyay, D. Personal air pollution monitoring technologies: User practices and preferences.
[110] Schroeder, J., Chung, C.-F., Epstein, D. A., Karkar, R., Parsons, A., Murinova, N., Fogarty, J., and Munson, S. A. Examining self-tracking by people with migraine: goals, needs, and opportunities in a chronic health condition. In *Proceedings of the 2018 Designing Interactive Systems Conference* (2018), pp. 135–148.
[111] Schroeder, J., Karkar, R., Murinova, N., Fogarty, J., and Munson, S. A. Examining opportunities for goal-directed self-tracking to support chronic condition management. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3*, 4 (2019), 1–26.
[112] Sedlmair, M., Meyer, M., and Munzner, T. Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics 18*, 12 (2012), 2431–2440.
[113] Segel, E., and Heer, J. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics 16*,

6 (2010), 1139–1148.
[114] Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization*. Elsevier, 1996, pp. 364–371.
[115] Sprague, D., and Tory, M. Exploring how and why people use visualizations in casual contexts: Modeling user goals and regulated motivations. *Information Visualization 11*, 2 (2012), 106–123.
[116] Srinivasan, A., Brehmer, M., Lee, B., and Drucker, S. M. What's the difference? evaluating variations of multi-series bar charts for visual comparison tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), pp. 1–12.
[117] Thudt, A., Hinrichs, U., Huron, S., and Carpendale, S. Self-reflection and personal physicalization construction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM, p. 154.
[118] Tian, R., and Paulos, E. MyPart : Personal , Portable , Accurate Airborne Particle Counting. 1338–1348.
[119] Tolmie, P., Crabtree, A., Rodden, T., Colley, J., and Luger, E. "this has to be the cats": Personal data legibility in networked sensing systems. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (2016), ACM, pp. 491–502.
[120] Trickett, S. B., Fu, W.-T., Schunn, C. D., and Trafton, J. G. From dipsy-doodle to streaming motions: Changes in representation in the analysis of visual scientific data. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (2000), vol. 22.
[121] Tsai, C. C., Lee, G., Raab, F., Norman, G. J., Sohn, T., Griswold, W. G., and Patrick, K. Usability and feasibility of pmeb: a mobile phone application for monitoring real time caloric balance. *Mobile networks and applications 12*, 2-3 (2007), 173–184.
[122] Viegas, F. B., Wattenberg, M., Van Ham, F., Kriss, J., and McKeon, M. Manyeyes: a site for visualization at internet scale. *IEEE transactions on visualization and computer graphics 13*, 6 (2007), 1121–1128.
[123] von Landesberger, T. Insights by visual comparison: The state and challenges. *IEEE computer graphics and applications 38*, 3 (2018), 140–148.
[124] Wang, X. M., Zhang, T. Y., Ma, Y. X., Xia, J., and Chen, W. A Survey of Visual Analytic Pipelines. *Journal of Computer Science and Technology 31*, 4 (2016), 787–804.
[125] Wattenberg, M. Visualizing the stock market. In *CHI'99 extended abstracts on Human factors in computing systems* (1999), pp. 188–189.
[126] Wattenberg, M. Baby names, visualization, and social data analysis. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* (2005), IEEE, pp. 1–7.
[127] Wiese, J., Das, S., Hong, J. I., and Zimmerman, J. Evolving the ecosystem of personal behavioral data. *Human Computer Interaction 32*, 5-6 (2017), 447–510.
[128] Wilson, T. D. The proper protocol: Validity and completeness of verbal reports. *Psychological Science 5*, 5 (1994), 249–252.
[129] Wong, Y. L., Madhavan, K., and Elmqvist, N. Towards characterizing domain experts as a user group. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)* (2018), IEEE, pp. 1–10.
[130] Wongsuphasawat, K., Moritz, D., Anand, A., Mackinlay, J., Howe, B., and Heer, J. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE transactions on visualization and computer graphics 22*, 1 (2015), 649–658.
[131] Wongsuphasawat, K., Moritz, D., Anand, A., Mackinlay, J., Howe, B., and Heer, J. Towards a general-purpose query language for visualization recommendation. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics* (2016), pp. 1–6.
[132] Wright, W., Schroh, D., Proulx, P., Skaburskis, A., and Cort, B. The sandbox for analysis: concepts and methods. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), ACM, pp. 801–810.
[133] Zhi, Q., Ottley, A., and Metoyer, R. Linking and Layout: Exploring the Integration of Text and Visualization in Storytelling. *Computer Graphics Forum 38*, 3 (2019), 675–685.

REFERENCES

[1] J. Moore, P. Goffin, J. Wiese, and M. Meyer, "Exploring the personal informatics analysis gap: 'There's a lot of bacon'," *IEEE Transactions on Visualization and Computer Graphics (VIS)*, pp. 95–105, 2021.

[2] National Institute of Biomedical Imaging Bioengineering, "Pediatric research using integrated sensor monitoring systems," https://www.nibib.nih.gov/research-funding/prisms, 2015, Accessed Mar 2017.

[3] J. Moore, P. Goffin, M. Meyer, P. Lundrigan, N. Patwari, K. Sward, and J. Wiese, "Managing in-home environments through sensing, annotating, and visualizing air quality data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)(Ubicomp '18)*, vol. 2, no. 3, Sept 2018.

[4] J. Moore, P. Goffin, J. Wiese, and M. Meyer, "An interview method for engaging personal data," *The Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 5, no. 4, pp. 1–28, dec 2021.

[5] I. Li, A. Dey, and J. Forlizzi, "A stage-based model of personal informatics systems," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 557–566. [Online]. Available: https://doi.org/10.1145/1753326.1753409

[6] C. Tsai, G. Lee, F. Raab, G. Norman, T. Sohn, W. Griswold, and K. Patrick, "Usability and feasibility of PmEB: A mobile phone application for monitoring real time caloric balance," *Mobile Networks and Applications*, vol. 12, pp. 173–184, 06 2007.

[7] S. Consolvo, K. Everitt, I. Smith, and J. A. Landay, "Design requirements for technologies that encourage physical activity," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 457–466. [Online]. Available: https://doi.org/10.1145/1124772.1124840

[8] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay, "Activity sensing in the wild: A field trial of UbiFit Garden," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 1797–1806. [Online]. Available: https://doi.org/10.1145/1357054.1357335

[9] J. Lindqvist, J. Cranshaw, J. Wiese, J. Hong, and J. Zimmerman, "I'm the mayor of my house: Examining why people use Foursquare - a social-driven location sharing application," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 2409–2418. [Online]. Available: https://doi.org/10.1145/1978942.1979295

[10] S. T. Peesapati, V. Schwanda, J. Schultz, M. Lepage, S.-y. Jeong, and D. Cosley, "Pensieve: Supporting everyday reminiscence," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 2027–2036. [Online]. Available: https://doi.org/10.1145/1753326.1753635

[11] S. Kim and E. Paulos, "InAir: Measuring and visualizing indoor air quality," p. 81–84, 2009. [Online]. Available: https://doi.org/10.1145/1620545.1620557

[12] ——, "InAir: Sharing indoor air quality measurements and visualizations," p. 1861–1870, 2010. [Online]. Available: https://doi.org/10.1145/1753326.1753605

[13] S. Kim, E. Paulos, and J. Mankoff, "InAir: A longitudinal study of indoor air quality measurements and visualizations," p. 2745–2754, 2013. [Online]. Available: https://doi.org/10.1145/2470654.2481380

[14] B. Fang, Q. Xu, T. Park, and M. Zhang, "AirSense: An intelligent home-based sensing system for indoor air quality analytics," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 109–119. [Online]. Available: https://doi.org/10.1145/2971648.2971720

[15] D. A. Epstein, C. Caldeira, M. C. Figueiredo, X. Lu, L. M. Silva, L. Williams, J. H. Lee, Q. Li, S. Ahuja, Q. Chen, P. Dowlatyari, C. Hilby, S. Sultana, E. V. Eikey, and Y. Chen, "Mapping and taking stock of the personal informatics literature," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 4, dec 2020. [Online]. Available: https://doi.org/10.1145/3432231

[16] D. Huang, M. Tory, B. Adriel Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury, "Personal visualization and personal visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 3, pp. 420–433, 2015.

[17] D. Epstein, F. Cordeiro, E. Bales, J. Fogarty, and S. Munson, "Taming data complexity in lifelogs: Exploring visual cuts of personal informatics data," in *Proceedings of the 2014 Conference on Designing Interactive Systems*, ser. DIS '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 667–676. [Online]. Available: https://doi.org/10.1145/2598510.2598558

[18] Y. L. Wong, K. Madhavan, and N. Elmqvist, "Towards characterizing domain experts as a user group," in *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*, 2018, pp. 1–10.

[19] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, 2012.

[20] P. Tolmie, A. Crabtree, T. Rodden, J. Colley, and E. Luger, ""This has to be the cats": Personal data legibility in networked sensing systems," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work amp; Social Computing*, ser. CSCW '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 491–502. [Online]. Available: https://doi.org/10.1145/2818048.2819992

[21] J. E. Fischer, A. Crabtree, T. Rodden, J. A. Colley, E. Costanza, M. O. Jewell, and S. D. Ramchurn, “”Just whack it on until it gets hot”: Working with IoT data in the home,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’16. New York, NY, USA: Association for Computing Machinery, 2016, p. 5933–5944. [Online]. Available: https://doi.org/10.1145/2858036.2858518

[22] K. A. Siek, K. H. Connelly, Y. Rogers, P. Rohwer, D. Lambert, and J. L. Welch, “When do we eat? An evaluation of food items input into an electronic food monitoring application,” pp. 1–10, 2006.

[23] J. J. Lin, L. Mamykina, S. Lindtner, G. Delajoux, and H. B. Strub, “Fish’n’Steps: Encouraging physical activity with an interactive computer game,” in *Proceedings of the 8th International Conference on Ubiquitous Computing*, ser. UbiComp’06. Berlin, Heidelberg: Springer-Verlag, 2006, p. 261–278. [Online]. Available: https://doi.org/10.1007/11853565_16

[24] F. Bentley, K. Tollmar, P. Stephenson, L. Levy, B. Jones, S. Robertson, E. Price, R. Catrambone, and J. Wilson, “Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change,” *ACM Trans. Comput.-Hum. Interact.*, vol. 20, no. 5, nov 2013. [Online]. Available: https://doi.org/10.1145/2503823

[25] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz, “Understanding quantified-selfers’ practices in collecting and exploring personal data,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’14. New York, NY, USA: Association for Computing Machinery, 2014, p. 1143–1152. [Online]. Available: https://doi.org/10.1145/2556288.2557372

[26] S. L. Jones, “Exploring correlational information in aggregated quantified self data dashboards,” p. 1075–1080, 2015. [Online]. Available: https://doi.org/10.1145/2800835.2800963

[27] S. L. Jones and R. Kelly, “Dealing with information overload in multifaceted personal informatics systems,” *Human–Computer Interaction*, vol. 33, no. 1, pp. 1–48, 2018. [Online]. Available: https://doi.org/10.1080/07370024.2017.1302334

[28] D. A. Epstein, A. Ping, J. Fogarty, and S. A. Munson, “A lived informatics model of personal informatics,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’15. New York, NY, USA: Association for Computing Machinery, 2015, p. 731–742. [Online]. Available: https://doi.org/10.1145/2750858.2804250

[29] T. Munzner, C. Johnson, R. Moorhead, H. Pfister, P. Rheingans, and T. Yoo, “NIH-NSF visualization research challenges report summary,” *IEEE Computer Graphics and Applications*, vol. 26, no. 2, pp. 20–24, 2006.

[30] T. Munzner, “A nested model for visualization design and validation,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 921–928, 2009.

[31] S. McKenna, D. Mazur, J. Agutter, and M. Meyer, “Design activity framework for visualization design,” *IEEE Transactions on Visualization and Computer Graphics (InfoVis ’14)*, vol. 20, no. 12, pp. 2191–2200, 2014.

[32] N. McCurdy, J. Dykes, and M. Meyer, "Action design research and visualization design," 2016, pp. 10–18.

[33] M. Meyer and J. Dykes, "Criteria for rigor in visualization design study," *IEEE Transactions on Visualization and Computer Graphics (InfoVis)*, vol. 26, no. 1, pp. 87–97, 2019.

[34] M. Sedlmair, "Design study contributions come in different guises: Seven guiding scenarios," in *Proc. of the IEEE VIS Workshop Beyond Time and Errors: Novel Evaluation Methods for Visualization (BELIV)*. ACM Press, October 2016. [Online]. Available: http://eprints.cs.univie.ac.at/4798/

[35] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer, "Enterprise data analysis and visualization: An interview study," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2917–2926, 2012.

[36] S. Alspaugh, N. Zokaei, A. Liu, C. Jin, and M. A. Hearst, "Futzing and moseying: Interviews with professional data analysts on exploration practices," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 22–31, 2019.

[37] E. Kandogan, A. Balakrishnan, E. M. Haber, and J. S. Pierce, "From data to insight: Work practices of analysts in the enterprise," *IEEE Computer Graphics and Applications*, vol. 34, no. 5, pp. 42–50, 2014.

[38] M. Kim, T. Zimmermann, R. DeLine, and A. Begel, "The emerging role of data scientists on software development teams," in *Proceedings of the 38th International Conference on Software Engineering*, ser. ICSE '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 96–107. [Online]. Available: https://doi.org/10.1145/2884781.2884783

[39] D. Fisher, R. DeLIne, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," *ACM Interactions*, May 2012, this is a pre-publication draft; do not distribute or reproduce. The official version may be found at http://dl.acm.org/citation.cfm?doid=2168931.2168943. [Online]. Available: https://www.microsoft.com/en-us/research/publication/interactions-with-big-data-analytics/

[40] P. Pirolli and S. Card, "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis," in *Proceedings of international conference on intelligence analysis*, vol. 5. McLean, VA, USA, 2005, pp. 2–4.

[41] ——, "Information foraging," *Psychological Review*, vol. 106, pp. 643–675, 10 1999.

[42] P. Cowley, L. Nowell, and J. Scholtz, "Glass box: An instrumented infrastructure for supporting human interaction with information," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 2005, pp. 296c–296c.

[43] E. Patterson, E. Roth, and D. Woods, "Predicting vulnerabilities in computer-supported inferential analysis under data overload," *Cognition, Technology & Work*, vol. 3, pp. 224–237, 12 2001.

[44] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort, "The sandbox for analysis: Concepts and methods," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 801–810. [Online]. Available: https://doi.org/10.1145/1124772.1124890

[45] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card, "The cost structure of sensemaking," in *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 1993, p. 269–276. [Online]. Available: https://doi.org/10.1145/169059.169209

[46] S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer, "Wrangler: Interactive visual specification of data transformation scripts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2011, p. 3363–3372. [Online]. Available: https://doi.org/10.1145/1978942.1979444

[47] A. Bigelow, C. Nobre, M. Meyer, and A. Lex, "Origraph: Interactive network wrangling," in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2019, pp. 81–92.

[48] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 data-driven documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.

[49] A. Satyanarayan, K. Wongsuphasawat, and J. Heer, "Declarative interaction design for data visualization," in *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 669–678. [Online]. Available: https://doi.org/10.1145/2642918.2647360

[50] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, "Vega-lite: A grammar of interactive graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 341–350, 2017.

[51] J. Mackinlay, P. Hanrahan, and C. Stolte, "Show me: Automatic presentation for visual analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1137–1144, 2007.

[52] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer, "Voyager: Exploratory analysis via faceted browsing of visualization recommendations," *IEEE transactions on visualization and computer graphics*, vol. 22, 09 2015.

[53] ——, "Towards a general-purpose query language for visualization recommendation," in *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, ser. HILDA '16. New York, NY, USA: Association for Computing Machinery, 2016. [Online]. Available: https://doi.org/10.1145/2939502.2939506

[54] K. Gadhave, J. Görtler, Z. Cutler, C. Nobre, O. Deussen, M. Meyer, J. Phillips, and A. Lex, "Predicting intent behind selections in scatterplot visualizations," *Information Visualization*, vol. 20, no. 4, p. 207–228, August 2021. [Online]. Available: https://doi.org/10.1177/14738716211038604

[55] R. Amar and J. Stasko, "Knowledge precepts for design and evaluation of information visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 4, pp. 432–442, 2005.

[56] B. c. Kwon, B. Fisher, and J. S. Yi, "Visual analytic roadblocks for novice investigators," in *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2011, pp. 3–11.

[57] P.-M. Law, R. C. Basole, and Y. Wu, "Duet: Helping data analysis novices conduct pairwise comparisons by minimal specification," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 427–437, 2019.

[58] L. Grammel, M. Tory, and M.-A. Storey, "How information visualization novices construct visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 943–952, 2010.

[59] B. Lee, E. K. Choe, P. Isenberg, K. Marriott, and J. Stasko, "Reaching broader audiences with data visualization," *IEEE Computer Graphics and Applications*, vol. 40, no. 2, pp. 82–90, 2020.

[60] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon, "ManyEyes: A site for visualization at internet scale," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1121–1128, 2007.

[61] C. M. Danis, F. B. Viegas, M. Wattenberg, and J. Kriss, "Your place or mine? Visualization as a community component," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 275–284. [Online]. Available: https://doi.org/10.1145/1357054.1357102

[62] J. Heer and D. Boyd, "Vizster: Visualizing online social networks," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 2005, pp. 32–39.

[63] J. Heer, F. B. Viégas, and M. Wattenberg, "Voyagers and voyeurs: Supporting asynchronous collaborative visualization," *Commun. ACM*, vol. 52, no. 1, p. 87–97, jan 2009. [Online]. Available: https://doi.org/10.1145/1435417.1435439

[64] M. Wattenberg, "Visualizing the stock market," in *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '99. New York, NY, USA: Association for Computing Machinery, 1999, p. 188–189. [Online]. Available: https://doi.org/10.1145/632716.632834

[65] ——, "Baby names, visualization, and social data analysis," in *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 2005, pp. 1–7.

[66] M. J. Danziger, "Information visualization for the people," Ph.D. dissertation, Massachusetts Institute of Technology, Department of Comparative Media Studies, 2008.

[67] S. Huron, S. Carpendale, A. Thudt, A. Tang, and M. Mauerer, "Constructive visualization," in *Proceedings of the 2014 Conference on Designing Interactive Systems*, ser. DIS '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 433–442. [Online]. Available: https://doi.org/10.1145/2598510.2598566

[68] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian, and K. Hornbæk, "Opportunities and challenges for data physicalization," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 3227–3236. [Online]. Available: https://doi.org/10.1145/2702123.2702180

[69] R. A. Khot, J. Lee, D. Aggarwal, L. Hjorth, and F. F. Mueller, "TastyBeats: Designing palatable representations of physical activity," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 2933–2942. [Online]. Available: https://doi.org/10.1145/2702123.2702197

[70] R. A. Khot, F. F. Mueller, and L. Hjorth, "SweatAtoms: Materializing physical activity," 2013. [Online]. Available: https://doi.org/10.1145/2513002.2513012

[71] S. Stusak, A. Tabard, F. Sauka, R. A. Khot, and A. Butz, "Activity sculptures: Exploring the impact of physical visualizations on running activity," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2201–2210, 2014.

[72] A. Thudt, U. Hinrichs, S. Huron, and S. Carpendale, "Self-reflection and personal physicalization construction," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–13. [Online]. Available: https://doi.org/10.1145/3173574.3173728

[73] A. Thudt, D. Baur, S. Huron, and S. Carpendale, "Visual mementos: Reflecting memories with personal data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 369–378, 2016.

[74] S. Houben, C. Golsteijn, S. Gallacher, R. Johnson, S. Bakker, N. Marquardt, L. Capra, and Y. Rogers, "Physikit: Data engagement through physical ambient visualizations in the home." in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, may 2016, pp. 1608–1619. [Online]. Available: https://dl.acm.org/doi/10.1145/2858036.2858059

[75] D. Holstius, J. Kembel, A. Hurst, P.-H. Wan, and J. Forlizzi, "Infotropism: Living and robotic plants as interactive displays," in *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, ser. DIS '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 215–221. [Online]. Available: https://doi.org/10.1145/1013115.1013145

[76] F. Botros, C. Perin, B. A. Aseniero, and S. Carpendale, "Go and grow: Mapping personal data to a living plant," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ser. AVI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 112–119. [Online]. Available: https://doi.org/10.1145/2909132.2909267

[77] P. Koytek, C. Perin, J. Vermeulen, E. André, and S. Carpendale, "Mybrush: Brushing and linking with personal agency," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 605–615, 2018.

[78] B. A. Aseniero, C. Perin, W. Willett, A. Tang, and S. Carpendale, “Activity river: Visualizing planned and logged personal activities for reflection,” 2020. [Online]. Available: https://doi.org/10.1145/3399715.3399921

[79] P. Lundrigan, K. T. Min, N. Patwari, S. Kasera, K. Kelly, J. Moore, M. Meyer, S. C. Collingwood, F. Nkoy, B. Stone, and K. Sward, “EpiFi: An in-home sensor network architecture for epidemiological studies,” *2018 IEEE 43rd Conference on Local Computer Networks Workshops (LCN Workshops)*, vol. 2, no. 3, pp. 30–37, Sept 2018.

[80] E. Kerzner, S. Goodwin, J. Dykes, S. Jones, and M. Meyer, “A framework for creative visualization-opportunities workshops,” *IEEE Transactions on Visualization and Computer Graphics (InfoVis)*, vol. 25, no. 1, pp. 748–758, 2019.

[81] World Health Organisation, “Ambient air pollution: A global assessment of exposure and burden of disease,” eSocialSciences, Working Papers, 2016. [Online]. Available: https://EconPapers.repec.org/RePEc:ess:wpaper:id:11368

[82] A. Peters, E. Liu, R. L. Verrier, J. Schwartz, D. R. Gold, M. Mittleman, J. Baliff, J. A. Oh, G. Allen, K. Monahan *et al.*, “Air pollution and incidence of cardiac arrhythmia,” *Epidemiology*, vol. 11, no. 1, pp. 11–17, 2000.

[83] A. Pope, R. T. Burnett, M. C. Turner, A. Cohen, D. Krewski, M. Jerrett, S. M. Gapstur, and M. J. Thun, “Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: Shape of the exposure-response relationships,” *Environmental Health Perspectives*, vol. 119, no. 11, pp. 1616–1621, 2011.

[84] J. Lepeule, F. Laden, D. Dockery, and J. Schwartz, “Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard six cities study from 1974 to 2009,” *Environmental Health Perspectives*, vol. 120, no. 7, pp. 965–970, 2012.

[85] American Lung Association, “State of the Air 2015,” American Lung Association, Tech. Rep., 2015.

[86] L. Shi, A. Zanobetti, I. Kloog, B. A. Coull, P. Koutrakis, S. J. Melly, and J. D. Schwartz, “Low-concentration $PM_{2.5}$ and mortality: Estimating acute and chronic effects in a population-based study,” *Environmental Health Perspectives*, vol. 124, pp. 46–52, 01 2016. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4710600/

[87] N. Duan, “Models for human exposure to air pollution,” *Environment International*, vol. 8, no. 1-6, pp. 305–309, 1982.

[88] M. L. Bell, K. Ebisu, and R. D. Peng, “Community-level spatial heterogeneity of chemical constituent levels of fine particulates and implications for epidemiological research,” *Journal of Exposure Science and Environmental Epidemiology*, vol. 21, no. 4, pp. 372–384, 07 2011. [Online]. Available: http://dx.doi.org/10.1038/jes.2010.24

[89] S. Steinle, S. Reis, and C. E. Sabel, “Quantifying human exposure to air pollution—Moving from static monitoring to spatio-temporally resolved personal exposure assessment,” *Science of The Total Environment*, vol. 443, pp. 184–193, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969712013988

[90] PurpleAir, “Purpleair air quality monitoring: An air quality monitoring network built on a new generation of ”internet of things” sensors,” http://map.purpleair.org/, 2017, Accessed Nov. 13, 2017.

[91] X. Chen, Y. Zheng, Y. Chen, Q. Jin, W. Sun, E. Chang, and W.-Y. Ma, “Indoor air quality monitoring system for smart buildings,” in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*. New York, NY, USA: ACM, 2014, pp. 471–475. [Online]. Available: http://doi.acm.org/10.1145/2632048.2632103

[92] P. Aoki, A. Woodruff, B. Yellapragada, and W. Willett, “Environmental protection and agency: Motivations, capacity, and goals in participatory sensing,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 3138–3150. [Online]. Available: https://doi.org/10.1145/3025453.3025667

[93] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, “Common sense: Participatory urban sensing using a network of handheld air quality monitors,” in *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys ’09. New York, NY, USA: Association for Computing Machinery, 2009, p. 349–350. [Online]. Available: https://doi.org/10.1145/1644038.1644095

[94] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, “Participatory air pollution monitoring using smartphones,” in *in Proceedings of the 1st International Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data*, 2012.

[95] Y.-C. Hsu, P. Dille, J. Cross, B. Dias, R. Sargent, and I. Nourbakhsh, “Community-empowered air quality monitoring system,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2017, p. 1607–1619. [Online]. Available: https://doi.org/10.1145/3025453.3025853

[96] S. Kuznetsov, S. E. Hudson, and E. Paulos, “A low-tech sensing system for particulate pollution,” in *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction*, ser. TEI ’14. New York, NY, USA: Association for Computing Machinery, 2014, p. 259–266. [Online]. Available: https://doi.org/10.1145/2540930.2540955

[97] Q. Y. Meng, B. J. Turpin, L. Korn, C. P. Weisel, M. Morandi, S. Colome, J. Zhang, T. Stock, D. Spektor, A. Winer *et al.*, “Influence of ambient (outdoor) sources on residential indoor and personal $PM_{2.5}$ concentrations: Analyses of RIOPA data,” *Journal of Exposure Science and Environmental Epidemiology*, vol. 15, no. 1, pp. 17–28, 2005.

[98] N. E. Klepeis, W. C. Nelson, W. R. Ott, J. P. Robinson, A. M. Tsang, P. Switzer, J. V. Behar, S. C. Hern, and W. H. Engelmann, “The national human activity pattern survey (NHAPS): A resource for assessing exposure to environmental pollutants,” *Journal of Exposure Science and Environmental Epidemiology*, vol. 11, no. 3, p. 231, 2001.

[99] Y. Jiang, K. Li, L. Tian, R. Piedrahita, X. Yun, O. Mansata, Q. Lv, R. P. Dick, M. Hannigan, and L. Shang, "MAQS: A personalized mobile sensing system for indoor air quality monitoring," in *Proceedings of the 13th International Conference on Ubiquitous Computing*, ser. UbiComp '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 271–280. [Online]. Available: https://doi.org/10.1145/2030112.2030150

[100] C. Monn, "Exposure assessment of air pollutants: A review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone," *Atmospheric Environment*, vol. 35, no. 1, pp. 1–32, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231000003307

[101] I. Salma, K. Dosztály, T. Borsós, B. Söveges, T. Weidinger, G. Kristóf, N. Péter, and Z. Kertész, "Physical properties, chemical composition, sources, spatial distribution and sinks of indoor aerosol particles in a university lecture hall," *Atmospheric Environment*, vol. 64, pp. 219–228, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231012009442

[102] National Academies of Sciences Engineering and Medicine, *Health Risks of Indoor Exposure to Particulate Matter: Workshop Summary*, D. A. Butler, G. Madhavan, and J. Alper, Eds. Washington, DC: The National Academies Press, 2016. [Online]. Available: https://www.nap.edu/catalog/23531/health-risks-of-indoor-exposure-to-particulate-matter-workshop-summary

[103] E. Abt, H. H. Suh, P. Catalano, and P. Koutrakis, "Relative contribution of outdoor and indoor particle sources to indoor concentrations," *Environmental Science & Technology*, vol. 34, no. 17, pp. 3579–3587, 2000.

[104] I. K. Koponen, A. Asmi, P. Keronen, K. Puhto, and M. Kulmala, "Indoor air measurement campaign in Helsinki, Finland 1999 – the effect of outdoor air pollution on indoor air," *Atmospheric Environment*, vol. 35, no. 8, pp. 1465–1477, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231000003381

[105] P. Kulkarni, P. A. Baron, and K. Willeke, *Aerosol measurement: Principles, techniques, and applications*. John Wiley & Sons, 2011.

[106] D. L. Schacter, "The seven sins of memory: Insights from psychology and cognitive neuroscience." *American psychologist*, vol. 54, no. 3, p. 182, 1999.

[107] M. Jin, N. Bekiaris-Liberis, K. Weekly, C. Spanos, and A. Bayen, "Sensing by proxy: Occupancy detection based on indoor CO_2 concentration," *UBICOMM 2015*, vol. 14, 07 2015.

[108] K. Jabłoński and T. Grychowski, "Fuzzy inference system for the assessment of indoor environmental quality in a room," *Indoor and Built Environment*, vol. 27, no. 10, pp. 1415–1430, 2018. [Online]. Available: https://doi.org/10.1177/1420326X17728097

[109] O. A. Postolache, J. M. Dias Pereira, and P. M. B. Silva Girao, "Smart sensors network for air quality monitoring applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 9, pp. 3253–3262, 2009.

[110] J.-Y. Kim, C.-H. Chu, and S.-M. Shin, "ISSAQ: An integrated sensing systems for real-time indoor air quality monitoring," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4230–4244, Dec 2014.

[111] N. Nikzad, N. Verma, C. Ziftci, E. Bales, N. Quick, P. Zappi, K. Patrick, S. Dasgupta, I. Krueger, T. v. Rosing, and W. G. Griswold, "CitiSense: Improving geospatial environmental assessment of air quality using a wireless personal exposure monitoring system," in *Proceedings of the Conference on Wireless Health*, ser. WH '12. New York, NY, USA: ACM, 2012, pp. 11:1–11:8. [Online]. Available: http://doi.acm.org/10.1145/2448096.2448107

[112] S. M. Saad, A. M. Andrew, A. Y. M. Shakaff, A. R. M. Saad, A. M. Y. Kamarudin, and A. Zakaria, "Classifying sources influencing indoor air quality (IAQ) using artificial neural network (ANN)," *Sensors*, vol. 15, no. 5, pp. 11 665–11 684, 2015. [Online]. Available: https://www.mdpi.com/1424-8220/15/5/11665

[113] Y. Huang, L. Hu, D. Yang, and H. Liu, "Air-Sense: Indoor environment monitoring evaluation system based on ZigBee network," in *IOP Conference Series: Earth and Environmental Science*, vol. 81, no. 1, 08 2017, p. 012208.

[114] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, and S. Patel, "Disaggregated end-use energy sensing for the smart grid," *IEEE Pervasive Computing*, vol. 10, no. 1, pp. 28–39, 2011.

[115] J. Froehlich, L. Findlater, M. Ostergren, S. Ramanathan, J. Peterson, I. Wragg, E. Larson, F. Fu, M. Bai, S. Patel, and J. A. Landay, "The design and evaluation of prototype eco-feedback displays for fixture-level water usage data," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2012, pp. 2367–2376. [Online]. Available: http://doi.acm.org/10.1145/2207676.2208397

[116] C. Fischer, "Feedback on household electricity consumption: A tool for saving energy?" *Energy Efficiency*, vol. 1, no. 1, pp. 79–104, 2008.

[117] R. J. Vercellino, D. K. Sleeth, R. G. Handy, K. T. Min, and S. C. Collingwood, "Laboratory evaluation of a low-cost, real-time, aerosol multi-sensor," *Journal of Occupational and Environmental Hygiene*, vol. 15, no. 7, pp. 559–567, 2018, pMID: 29683781. [Online]. Available: https://doi.org/10.1080/15459624.2018.1468565

[118] S. Hegde, K. T. Min, J. Moore, P. Lundrigan, N. Patwari, S. Collingwood, A. Balch, and K. E. Kelly, "Indoor household particulate matter measurements using a network of low-cost sensors," vol. 20, no. 2. Taiwan Association for Aerosol Research, 2020, pp. 381–394. [Online]. Available: https://doi.org/10.4209/aaqr.2019.01.0046

[119] E. Costanza, S. D. Ramchurn, and N. R. Jennings, "Understanding domestic energy consumption through interactive visualisation: A field study," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 216–225. [Online]. Available: https://doi.org/10.1145/2370216.2370251

[120] Brakel, J.P.G, "Robust peak detection algorithm using Z-Scores," *Stack Overflow*, 2014. [Online]. Available: https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data/22640362\#22640362

[121] A. K. Dey, K. Wac, D. Ferreira, K. Tassini, J.-H. Hong, and J. Ramos, "Getting closer: An empirical investigation of the proximity of user to their smart phones," in *Proceedings of the 13th International Conference on Ubiquitous Computing*, ser. UbiComp '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 163–172. [Online]. Available: https://doi.org/10.1145/2030112.2030135

[122] "How to install Google Play Store on FireOS(5.x)," https://forum.xda-developers.com/amazon-fire/general/how-to-install-google-play-store-fire-t3486603, October 24, 2016, Accessed: Jun 1, 2017.

[123] D. R. Hancock and B. Algozzine, *Doing case study research: A practical guide for beginning researchers*. Teachers College Press, 2016.

[124] V. Clarke and V. Braun, *Thematic Analysis*. New York, NY: Springer New York, 2014, pp. 1947–1952. [Online]. Available: https://doi.org/10.1007/978-1-4614-5583-7_311

[125] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia, and X. Jiang, "AirCloud: A cloud-based air-quality monitoring system for everyone," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, ser. SenSys '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 251–265. [Online]. Available: https://doi.org/10.1145/2668332.2668346

[126] L. Wallace, E. Pellizzari, T. Hartwell, C. Sparacino, R. Whitmore, L. Sheldon, H. Zelon, and R. Perritt, "The TEAM (Total Exposure Assessment Methodology) study: Personal exposures to toxic substances in air, drinking water, and breath of 400 residents of New Jersey, North Carolina, and North Dakota," *Environmental research*, vol. 43, no. 2, pp. 290–307, August 1987. [Online]. Available: https://doi.org/10.1016/S0013-9351(87)80030-0

[127] H. Sarker, M. Sharmin, A. A. Ali, M. M. Rahman, R. Bari, S. M. Hossain, and S. Kumar, "Assessing the availability of users to engage in just-in-time intervention in the natural environment," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 909–920. [Online]. Available: https://doi.org/10.1145/2632048.2636082

[128] T.-J. Yun, H. Y. Jeong, H. R. Lee, R. I. Arriaga, and G. D. Abowd, "Assessing asthma management practices through in-home technology probes," 2010, pp. 1–9. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs{_}all.jsp?arnumber=5482301

[129] B. Buxton, *Sketching User Experiences: Getting the Design Right and the Right Design*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007.

[130] J. van Kollenburg, S. Bogers, H. Rutjes, E. Deckers, J. Frens, and C. Hummels, "Exploring the value of parent tracked baby data in interactions with healthcare professionals: A data-enabled design exploration," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY,

USA: Association for Computing Machinery, 2018, p. 1–12. [Online]. Available: https://doi.org/10.1145/3173574.3173871

[131] J. Schroeder, C.-F. Chung, D. A. Epstein, R. Karkar, A. Parsons, N. Murinova, J. Fogarty, and S. A. Munson, "Examining self-tracking by people with migraine: Goals, needs, and opportunities in a chronic health condition," in *Proceedings of the 2018 Designing Interactive Systems Conference*, ser. DIS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 135–148. [Online]. Available: https://doi.org/10.1145/3196709.3196738

[132] L. M. Tang and J. Kay, "Harnessing long term physical activity data—how long-term trackers use data and how an adherence-based interface supports new insights," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, jun 2017. [Online]. Available: https://doi.org/10.1145/3090091

[133] M. Mărăşoiu, A. F. Blackwell, A. Sarkar, and M. Spott, "Clarifying hypotheses by sketching data," in *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization: Short Papers*, ser. EuroVis '16. Goslar, DEU: Eurographics Association, 2016, p. 125–129.

[134] J. Moore, "The data engagement interview guide," https://vdl.sci.utah.edu/EngagementInterviews/, 2021, Accessed: Oct. 15, 2021.

[135] D. A. Epstein, M. Caraway, C. Johnston, A. Ping, J. Fogarty, and S. A. Munson, "Beyond abandonment to next steps: Understanding and designing for life after personal informatics tool use," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1109–1113. [Online]. Available: https://doi.org/10.1145/2858036.2858045

[136] C.-F. Chung, Q. Wang, J. Schroeder, A. Cole, J. Zia, J. Fogarty, and S. A. Munson, "Identifying and planning for individualized change: Patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 1, mar 2019. [Online]. Available: https://doi.org/10.1145/3314394

[137] C.-F. Chung, E. Agapie, J. Schroeder, S. Mishra, J. Fogarty, and S. A. Munson, "When personal tracking becomes social: Examining the use of instagram for healthy eating." New York, NY, USA: Association for Computing Machinery, 2017, p. 1674–1687. [Online]. Available: https://doi.org/10.1145/3025453.3025747

[138] E. T. K. van Dijk, J. H. Westerink, F. Beute, and W. A. IJsselsteijn, "Personal informatics, self-insight, and behavior change: A critical review of current literature," *Human–Computer Interaction*, vol. 32, no. 5-6, pp. 268–296, 2017. [Online]. Available: https://doi.org/10.1080/07370024.2016.1276456

[139] D. Schuler and A. Namioka, *Participatory design: Principles and practices*. CRC Press, 1993.

[140] J. E. Fischer, A. Crabtree, J. A. Colley, T. Rodden, and E. Costanza, "Data work: How energy advisors and clients make iot data accountable," *Computer Supported*

Cooperative Work, vol. 26, no. 4–6, p. 597–626, dec 2017. [Online]. Available: https://doi.org/10.1007/s10606-017-9293-x

[141] J. Browne, B. Lee, S. Carpendale, N. Riche, and T. Sherwood, "Data analysis on interactive whiteboards through sketch-based interaction," in *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ser. ITS '11. New York, NY, USA: Association for Computing Machinery, 2011, pp. 154–157. [Online]. Available: https://doi.org/10.1145/2076354.2076383

[142] B. Buxton, "The anatomy of sketching," in *Sketching User Experiences*, ser. Interactive Technologies, B. Buxton, Ed. Burlington: Morgan Kaufmann, 2007, pp. 105–114. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123740373500547

[143] B. Lee, R. H. Kazi, and G. Smith, "SketchStory: Telling more engaging stories with data through freeform sketching," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2416–2425, 2013.

[144] J. Walny, S. Carpendale, N. H. Riche, G. Venolia, and P. Fawcett, "Visual thinking in action: Visualizations as used on whiteboards," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2508–2517, 2011.

[145] S. Huron, P. Gourlet, U. Hinrichs, T. Hogan, and Y. Jansen, "Let's get physical: Promoting data physicalization in workshop formats," *Proceedings of the ACM Conference on Designing Interactive Systems*, vol. 3, no. Umr 9217, pp. 1409–1422, 2017. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3064663.3064798

[146] W. Willett and S. Huron, "A constructive classroom exercise for teaching infovis," in *Pedagogy of Data Visualization Workshop at IEEE VIS 2016*, ser. Pedagogy of Data Visualization Workshop at IEEE VIS 2016, Baltimore, United States, Oct. 2016. [Online]. Available: https://hal.inria.fr/hal-01511020

[147] E. K. Choe, B. Lee, H. Zhu, N. H. Riche, and D. Baur, "Understanding self-reflection: How people reflect on personal data through visual data exploration," in *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, ser. PervasiveHealth '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 173–182. [Online]. Available: https://doi.org/10.1145/3154862.3154881

[148] R. Arias-Hernandez, L. T. Kaastra, T. M. Green, and B. Fisher, "Pair analytics: Capturing reasoning processes in collaborative visual analytics," in *2011 44th Hawaii International Conference on System Sciences*. IEEE, jan 2011, pp. 1–10. [Online]. Available: http://ieeexplore.ieee.org/document/5718616/

[149] S. B. Trickett, W.-T. Fu, C. D. Schunn, and J. G. Trafton, "From dipsy-doodle to streaming motions: Changes in representation in the analysis of visual scientific data," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 22, no. 22, 2000. [Online]. Available: https://escholarship.org/uc/item/1d2836rc

[150] J. Dickson, J. McLennan, and M. M. Omodei, "Effects of concurrent verbalization on a time-critical, dynamic decision-making task," *The Journal of General*

Psychology, vol. 127, no. 2, pp. 217–228, 2000. [Online]. Available: https://doi.org/10.1080/00221300009598580

[151] T. D. Wilson, "The proper protocol: Validity and completeness of verbal reports," *Psychological Science*, vol. 5, no. 5, pp. 249–252, 1994. [Online]. Available: https://doi.org/10.1111/j.1467-9280.1994.tb00621.x

[152] J. F. Kelley, "An iterative design methodology for user-friendly natural language office information applications," *ACM Trans. Inf. Syst.*, vol. 2, no. 1, p. 26–41, jan 1984. [Online]. Available: https://doi.org/10.1145/357417.357420

[153] J. Monforte and J. Úbeda Colomer, "Tinkering with the two-to-one interview: Reflections on the use of two interviewers in qualitative constructionist inquiry," *Methods in Psychology*, vol. 5, p. 100082, 10 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590260121000394

[154] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. E. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. B. Hamrick, J. Grout, S. Corlay *et al.*, "Jupyter notebooks-a publishing format for reproducible computational workflows." in *ELPUB*, 2016, pp. 87–90.

[155] M. E. Fonteyn, B. Kuipers, and S. J. Grobe, "A description of think aloud method and protocol analysis," *Qualitative Health Research*, vol. 3, no. 4, pp. 430–441, 1993. [Online]. Available: https://doi.org/10.1177/104973239300300403

[156] T. Lenzner, C. Neuert, and W. Otto, "Cognitive pretesting," *GESIS Survey Guidelines*, p. 3, 2016.

[157] S. Presser, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, J. M. Rothgeb, and E. Singer, "Methods for testing and evaluating survey questions," *Public Opinion Quarterly*, vol. 68, no. 1, pp. 109–130, 03 2004. [Online]. Available: https://doi.org/10.1093/poq/nfh008

[158] T. Rattenbury, J. M. Hellerstein, J. Heer, S. Kandel, and C. Carreras, *Principles of data wrangling: Practical techniques for data preparation*. O'Reilly Media, Inc., 2017.

[159] D. Kirsh, "The context of work," *Human–Computer Interaction*, vol. 16, no. 2-4, pp. 305–322, 2001. [Online]. Available: https://doi.org/10.1207/S15327051HCI16234_12

[160] S. Lee, B. C. Kwon, J. Yang, B. C. Lee, and S.-H. Kim, "The correlation between users' cognitive characteristics and visualization literacy," *Applied Sciences*, vol. 9, no. 3, p. 488, 2019. [Online]. Available: https://doi.org/10.3390/app9030488

[161] M. Galesic and R. Garcia-Retamero, "Graph literacy: A cross-cultural comparison," *Medical Decision Making*, vol. 31, no. 3, pp. 444–457, 2011. [Online]. Available: https://doi.org/10.1177/0272989X10373805

[162] T. Munzner, *Visualization analysis and design*. CRC press, 2014.

[163] N. Iliinsky and J. Steele, *Designing data visualizations: Representing informational relationships*. " O'Reilly Media, Inc.", 2011.

[164] G. McCracken, *The long interview*, ser. Qualitative Research Methods. Sage Publications, 1988, vol. 13. [Online]. Available: https://books.google.com/books?id=3N01cl2gtoMC

[165] B. L. Leech, "Asking questions: Techniques for semistructured interviews," *PS: Political science and politics*, vol. 35, no. 4, pp. 665–668, 2002.

[166] C. E. Jia, H. P. Zhang, Y. Lv, R. Liang, Y. Q. Jiang, H. Powell, J. J. Fu, L. Wang, P. G. Gibson, and G. Wang, "The asthma control test and asthma control questionnaire for assessing asthma control: Systematic review and meta-analysis," *Journal of Allergy and Clinical Immunology*, vol. 131, no. 3, pp. 695–703, 2013.

[167] H. Beyer and K. Holtzblatt, *Contextual design*. ACM New York, NY, USA, 1999, vol. 6, no. 1.

[168] M. Card, *Readings in information visualization: Using vision to think*. Morgan Kaufmann, 1999.

[169] B. Shneiderman, "Designing for fun: How can we design user interfaces to be more fun?" *Interactions*, vol. 11, no. 5, p. 48–50, Sep. 2004. [Online]. Available: https://doi.org/10.1145/1015530.1015552

[170] M. Brehmer and T. Munzner, "A multi-level typology of abstract visualization tasks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2376–2385, 2013.

[171] C. Neugebaur, "Salt Lake City has the worse air quality in the nation," https://kutv.com/news/local/salt-lake-city-has-the-worst-air-quality-in-the-nation, Accessed: Oct 15. 2021.

[172] F. Cordeiro, D. A. Epstein, E. Thomaz, E. Bales, A. K. Jagannathan, G. D. Abowd, and J. Fogarty, "Barriers and negative nudges: Exploring challenges in food journaling," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1159–1162. [Online]. Available: https://doi.org/10.1145/2702123.2702155

[173] I. Li, A. K. Dey, and J. Forlizzi, "Using context to reveal factors that affect physical activity," *ACM Trans. Comput.-Hum. Interact.*, vol. 19, no. 1, may 2012. [Online]. Available: https://doi.org/10.1145/2147783.2147790

[174] E. K. Choe, S. Abdullah, M. Rabbi, E. Thomaz, D. A. Epstein, F. Cordeiro, M. Kay, G. D. Abowd, T. Choudhury, J. Fogarty, B. Lee, M. Matthews, and J. A. Kientz, "Semi-automated tracking: A balanced approach for self-monitoring applications," *IEEE Pervasive Computing*, vol. 16, no. 1, pp. 74–84, jan 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7807194/

[175] T. Dasu and T. Johnson, *Exploratory data mining and data cleaning*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, 2003, vol. 479. [Online]. Available: https://books.google.com/books?id=HStUfJnCvT4C

[176] J. Kennedy, B. Nansen, M. Arnold, R. Wilken, and M. Gibbs, "Digital housekeepers and domestic expertise in the networked home," *Convergence*, vol. 21, no. 4, pp. 408–422, 2015. [Online]. Available: https://doi.org/10.1177/1354856515579848

[177] L. Mamykina, E. Mynatt, P. Davidson, and D. Greenblatt, "MAHI: Investigation of social scaffolding for reflective thinking in diabetes management," in *Proceedings of*

the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 477–486. [Online]. Available: https://doi.org/10.1145/1357054.1357131

[178] S. Kolovson, C. Liang, S. A. Munson, and K. Starbird, "Personal data and power asymmetries in u.s. collegiate sports teams," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. GROUP, Jan. 2020. [Online]. Available: https://doi.org/10.1145/3375185

[179] Utah physicians for a healthy environment, "SLC – Ranked among the worst globally for air quality," https://www.uphe.org/2020/09/14/slc-ranked-among-the-worst-globally-for-air-quality/, Accessed: Mar. 19, 2021.

[180] W. Lidwell, K. Holden, and J. Butler, *Universal principles of design*. Rockport Pub, 2010.

[181] Y. Rogers, "New theoretical approaches for human-computer interaction," *Annual review of information science and technology*, vol. 38, no. 1, pp. 87–143, '04.

[182] J. J. Gibson, *The ecological approach to visual perception*. Psychology Press, 1986.

[183] D. A. Norman, "Affordance, conventions, and design," *Interactions*, vol. 6, no. 3, pp. 38–43, 1999.

[184] P. Thorvald, "Triggers, entry points, and affordances: How to improve their cognitive congeniality," Master's thesis, University of Skövde.

[185] T. Blascheck, L. M. Vermeulen, J. Vermeulen, C. Perin, W. Willett, T. Ertl, and S. Carpendale, "Exploration strategies for discovery of interactivity in visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 2, pp. 1407–1420.

[186] J. Walny, S. Storteboom, R. Pusch, S. M. Hwang, S. Knudsen, S. Carpendale, and W. Willett, "PixelClipper: Supporting public engagement and conversation about visualizations," *IEEE Computer Graphics and Applications*, vol. 40, no. 2, pp. 57–70, 2020.

[187] E. Hornecker, P. Marshall, and Y. Rogers, "From entry to access: How shareability comes about," in *Proceedings of the 2007 Conference on Designing Pleasurable Products and Interfaces*, ser. DPPI '07. New York, NY, USA: Association for Computing Machinery, 2007, p. 328–342. [Online]. Available: https://doi.org/10.1145/1314161.1314191

[188] F. Miranda, M. Lage, H. Doraiswamy, C. Mydlarz, J. Salamon, Y. Lockerman, J. Freire, and C. T. Silva, "Time lattice: A data structure for the interactive visual analysis of large time series," *Computer Graphics Forum*, vol. 37, no. 3, pp. 23–35, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13398

[189] E. M. Peck, S. E. Ayuso, and O. El-Etr, "Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–12. [Online]. Available: https://doi.org/10.1145/3290605.3300474

[190] N. A. Dodoo and S. Youn, "Snapping and chatting away: Consumer motivations for and outcomes of interacting with snapchat AR ad lens," *Telematics and Informatics*, vol. 57, p. 101514, 2021.

[191] J. Tu and E. Durmanova, "Curioscape: A curiosity-driven escape room board game," in *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*, ser. CHI PLAY '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 94–97. [Online]. Available: https://doi.org/10.1145/3383668.3419925

[192] H. Slater. [Update] The curiosity cube diaries - volume 1. https://www.pocketgamer.com/articles/046406/update-the-curiosity-cube-diaries-volume-i/. Accessed: Mar 21, 2021.

[193] S. Deterding, "Situated motivational affordances of game elements: A conceptual model," in *ACM CHI Conference on Human Factors in Computing Systems*, July 2011, Gamification workshop.

[194] F. Altarriba Bertran, E. Márquez Segura, J. Duval, and K. Isbister, "Chasing play potentials: Towards an increasingly situated and emergent approach to everyday play design," in *Proceedings of the 2019 on Designing Interactive Systems Conference*, ser. DIS '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 1265–1277. [Online]. Available: https://doi.org/10.1145/3322276.3322325

[195] W. W. Gaver, J. Bowers, A. Boucher, H. Gellerson, S. Pennington, A. Schmidt, A. Steed, N. Villars, and B. Walker, "The drift table: Designing for ludic engagement," in *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 885–900. [Online]. Available: https://doi.org/10.1145/985921.985947

[196] A. Scott, "How playcentric research methods are contributing to new understanding and opportunities for design," in *The Routledge companion to design research*, P. Rodgers and J. Yee, Eds. Routledge, 2014, ch. 31, pp. 400–414.

[197] S. P. Dow, K. Heddleston, and S. R. Klemmer, "The efficacy of prototyping under time constraints," in *Proceedings of the Seventh ACM Conference on Creativity and Cognition*, ser. C'I&';C '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 165–174. [Online]. Available: https://doi.org/10.1145/1640233.1640260

[198] M. Thibault, "Play as a modelling system-a semiotic analysis of the overreaching prestige of games." in *GamiFIN*, 2017, pp. 105–110.

[199] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining "gamification"," in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, ser. MindTrek '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 9–15. [Online]. Available: https://doi.org/10.1145/2181037.2181040

[200] G. H. Orcutt, H. W. Watts, and J. B. Edwards, "Data aggregation and information loss," *The American Economic Review*, vol. 58, no. 4, pp. 773–787, 1968. [Online]. Available: http://www.jstor.org/stable/1815532

[201] J. Heer and M. Agrawala, "Design considerations for collaborative visual analytics," in *2007 IEEE Symposium on Visual Analytics Science and Technology*, vol. 7, no. 1, 2008, pp. 49–62. [Online]. Available: https://doi.org/10.1057/palgrave.ivs.9500167

[202] J. Van Der Hoven and N. Manders-Huits, *Value-sensitive design*. Routledge, 2020.

[203] A. J. Jones, A. Artikis, and J. Pitt, "The design of intelligent socio-technical systems," *Artif. Intell. Rev.*, vol. 39, no. 1, p. 5–20, jan 2013.

[204] U. Ehsan, Q. V. Liao, M. Muller, M. O. Riedl, and J. D. Weisz, "Expanding explainability: Towards social transparency in AI systems," 2021. [Online]. Available: https://doi.org/10.1145/3411764.3445188

[205] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins *et al.*, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020.

[206] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0004370218305988

[207] M. Kogan, A. Halfaker, S. Guha, C. Aragon, M. Muller, and S. Geiger, "Mapping out human-centered data science: Methods, approaches, and best practices," in *Companion of the 2020 ACM International Conference on Supporting Group Work*, ser. GROUP '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 151–156. [Online]. Available: https://doi.org/10.1145/3323994.3369898

[208] T. D. Anderson and N. Parker, "Keeping the human in the data scientist: Shaping human-centered data science education," *Proceedings of the Association for Information Science and Technology*, vol. 56, no. 1, pp. 601–603, 2019.

[209] C. Aragon, C. Hutto, A. Echenique, B. Fiore-Gartland, Y. Huang, J. Kim, G. Neff, W. Xing, and J. Bayer, "Developing a research agenda for human-centered data science," in *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, ser. CSCW '16 Companion. New York, NY, USA: Association for Computing Machinery, 2016, p. 529–535. [Online]. Available: https://doi.org/10.1145/2818052.2855518