Here's what you need to know about my data: Exploring Expert Knowledge's Role in Data Analysis

Haihan Lin, Maxim Lisnic, Derya Akbaba, Miriah Meyer, Alexander Lex

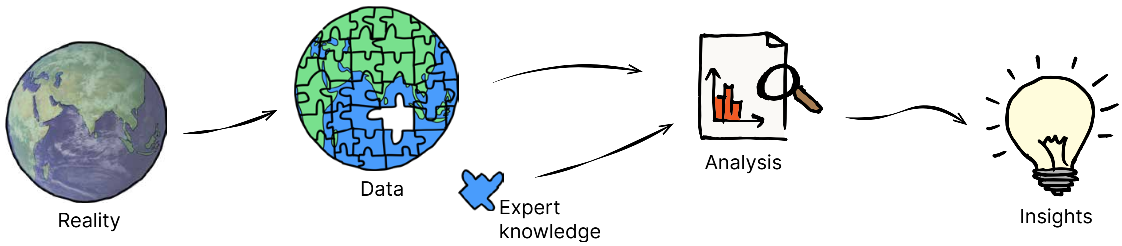

Fig. 1: An overview of the role of expert knowledge in data analysis workflows. Data is an imprecise and incomplete representation of reality. Expert knowledge helps with understanding the limitations of the data and may fill in the gaps between data and reality. Data analysis should leverage both knowledge and data to arrive at robust insights.

Abstract—Data-driven decision making has become a popular practice in science, industry, and public policy. Yet data alone, as an imperfect and partial representation of reality, is often insufficient to make good analysis decisions. Knowledge about the context of a dataset, its strengths and weaknesses, and its applicability for certain tasks is essential. Analysts are often not only familiar with the data itself, but also have data hunches about their analysis subject. In this work, we present an interview study with analysts from a wide range of domains and with varied expertise and experience, inquiring about the role of contextual knowledge. We provide insights into how data is insufficient in analysts' workflows and how they incorporate other sources of knowledge into their analysis. We analyzed how knowledge of data shaped their analysis outcome. Based on the results, we suggest design opportunities to better and more robustly consider both knowledge and data in analysis processes.

Index Terms—Human-Subjects Qualitative Studies

1 Introduction

On September 26, 1983, the Soviet Air Defense Forces' computers reported five missiles heading towards the Soviet Union from the United States, triggering a protocol that called for an immediate and compulsory nuclear counter-attack. Stanislav Petrov, the officer on duty, chose not to act. He determined that the incoming strike warning was more likely a system malfunction rather than a real attack. Petrov believed that if the US were to strike first, it would be a massive operation with hundreds of missiles, rather than only five missiles as the data indicated. He made the crucial decision to disregard the warning and not launch a nuclear attack, despite having no data to confirm his interpretation [31]. Later investigation revealed that the system had indeed malfunctioned due to a rare alignment of the detection satellite and the sun. Petrov's knowledge and experience enabled him to recognize the possibility of a false alarm and to interpret the data in the context of the political situation. Had he solely relied on the data provided by the warning system, the consequences could have been catastrophic.

While not all data (mis)interpretations lead to world-shattering consequences, data-driven decision-making has become an increasingly popular practice in fields like public policy, science, and industry, but also in making choices about our everyday lives. However, data alone is not sufficient to make good decisions. Data is an imperfect and partial representation of reality [17,50]; it can be misleading [19], hence acting solely based on data can be dangerous, as the story about Stanislav Petrov illustrates. Expert knowledge can provide essential context for the data in data analysis [41,56]. Experts know about relevant contexts based on their experience and domain knowledge, familiarity with the subject, and understanding of the data collection modalities.

Analysts who work with data often find themselves incorporating (their own or others') expert knowledge into their analysis, as illustrated in Fig 1. Expert knowledge provides analysts with context and caveats about the data and assures analysts of the soundness of their analysis. Previous works have recorded anecdotes of analysts applying experts' knowledge in their analysis [53,61]. However, there is a lack of focused work in the visualization research community on the details of how expert knowledge is integrated, documented, and communicated during an analyst's visualization workflow. We argue that current approaches to incorporating knowledge are ad-hoc, hampered by inefficient communication between stakeholders, and often not (sufficiently) documented; thereby leading to worse analysis results, a lack of reproducibility of analysis, and a lack of reusability of a dataset. Three panelists at an industry data science forum were recently asked: "if you had one wish to improve data analysis in your organization, what would it be?" Two of the three panelists responded that their wish would be to better understand what the data they have was originally used for, how it was collected, and what its appropriate use is now.[1]

In this work, we share the result of an interview study with 14 domain experts and analysts from a broad range of fields that investigates how they deal with data caveats in their workflow, how experts fill in the gap between data and reality using their knowledge from various sources, and how they document and communicate data caveats and knowledge in their work. Our primary contribution is an analysis of participants' current practices in documenting and communicating relevant knowledge in data analysis, and where current practices fall short. We discovered that the analysis output is not only influenced by data, the knowledge of data, but also the receiving audience of the output. Based on the interview analysis, we also contribute a discussion

- *Haihan Lin, Max Lisnic, and Alexander Lex are with the University of Utah E-Mail: alex@sci.utah.edu*
- *Derya Akbaba and Miriah Meyer are with Linköping University E-Mail: miriah.meyer@liu.se*

[1] Panel at Utah Data Science Day, Jan 12, 2024.

of design opportunities to better support analysts in documenting data caveats and expert knowledge in data analysis pipelines.

2 Positionality and Definitions

Within our team, we have a diversity of epistemic foundations that drive our research practices in individual fields. For some of us, there is a strong focus and attention toward learning about the world through building and deploying innovative visualization analysis tools. For others, our research attentions focus on bringing new epistemological lenses into visualization work to question commonly-held engineering assumptions and norms. In this work, our interests intersect through challenging assumptions prevalent in data analysis tooling that data are objective, authoritative inputs. Instead, we build our work on feminist and critical data studies perspectives that regard data as imperfect representations of reality.

This work builds on the concept of **data hunches** we previously defined as "an analyst's knowledge about how and why the data is an imperfect and partial representation of the phenomena of interest" [46]. Data hunches are the knowledge people have about the mismatch between reality and data, based on the understanding that data is a partial representation of the world, and that diverse perspectives are required to achieve a holistic view of reality. The concept of data hunches grew from feminist and critical epistemological theories [1]. We find the concept of data hunches interesting for the ways it removes the hierarchy and implicit power given to data over the subjective knowledge that experts bring to their analysis. While prior studies looked at how analysts work with imperfect, dirty data [12, 39, 53] and propose solutions for how to improve the representativeness of data through solutions such as better technologies, more measurements, and improved calibrations, the feminist situated knowledges perspective behind data hunches asserts that data "cannot fully represent the natural world" [46].

Follow-up work by Akbaba et al. emphasizes the "indispensable role of experts' knowledge about the entanglements of the data and their representational limitations" [1]. We embrace this position here by centering our interviews on the *relationship* between experts and their data. We use the term **expert** to refer to individuals with expertise in the domain they study (e.g., scientists, engineers, social workers, lawyers), as opposed to expertise in data analysis methods (statisticians, computer scientists, data scientists, etc.). This terminology draws on the visualization design study methodology that considers *domain experts* to be people with deep knowledge about a phenomenon in the world as well as the production of the data that is meant to represent it [72].

We distinguish data hunches from the narrower concept of **data caveats**. Data caveats warn about issues with the data but don't typically consider the relationship of the data to the phenomenon of interest. For the rest of this paper, we use the term "data hunches" when we want to express that analysts have a concrete hunch on what should be different, while we use "data caveats" to denote issues with the data, without a specific hunch on how the data could better represent a phenomenon of interest. For a more thorough discussion of the relationship of uncertainty and data hunches, refer to our previous work [46].

3 Related Work

We discuss previous visualization and HCI research that explores the existing practices of data analysts and how analysts document their data and analysis.

3.1 Practices of Data Analysts

Research is abundant from the visualization and HCI communities that provides insight into the current practices of data workers and analysts [4, 52, 57]. In a review of studies on data science workers, Crisan et al. synthesized the different processes performed by data workers: *preparation*, *analysis*, *deployment*, and *communication* [12]. They found that although visualizations touch all of the described processes, their actual use is limited. In contrast to prior work, we focus on the participants' personal experience in dealing with data imperfections in their work, how their data aligns with the phenomenon they study, and by which means they capture any discrepancies (if at all). Our interviews covered all these stages of analysis, but focused on preparation, analysis, and communication.

Prior works have studied how data analysts engage with data in various settings. Kandel et al. [40] interviewed 35 analysts from various fields and studied their processes and challenges in data analysis in enterprise settings. Their study participants reported using visualizations to identify outliers and artifacts in data. However, their study participants did not report usage of experts' knowledge in the data analysis pipeline, which is the focus of our study. Liu et al. [49] investigated analytical decision making and reporting in research settings and found that while multiple interpretations were commonly possible, there was a preference to report desired results corresponding to study participants' theories. Ruddle et al. [69] conducted interviews on data profiling—determining the quality of datasets and characterizing data for analyses. While both their and our work have found that analysts' "gut check" data for quality assurance, our work focuses more on using experts' knowledge to compensate for data issues and documenting and communicating experts' knowledge. Another key difference of our investigations compared to prior work is our focus on "domain experts", i.e., individuals trained in the subject matter they study or work in, as opposed to "data analysis experts". We investigated analysts from diverse backgrounds and focused on their practices of recording and communicating data caveats. In contrast, Kandel et al. [40] interviewed analysts with technical backgrounds and job descriptions such as "data scientist", "software engineer" or "chief technology officer".

To better understand how analysts collaborate within their teams, Zhang et al. surveyed 183 data workers in machine learning and artificial intelligence and summarized their workflows in general and tools used in their workflows in particular [80]. The authors described the difference in patterns of communication exhibited by different roles within teams. Reflecting on the practices of analysts across different fields, we also find that communication patterns heavily depend on the role, and our work specifically looked into how experts communicate their domain knowledge in data analysis. Jung et al. ran an in-depth study into how domain experts work with data [37] and found that they put more value on their data being actionable than on the data having abstract qualities, such as high precision. They also discussed *conversations with the data*—procedures of working with data directly to better understand it [71]—as a critical part of the analysis process, similar to previous works [56, 76]. Dimara et al. [15] presented survey and interview results with a focus on the decision-making process and the tensions that arise from collaborations between decision makers and data analysts. They concluded that decision-making tools are lacking throughout different stages of decision-making, especially compared to the process of data analysis. Our study further explores different ways in which analysts make sense of and communicate data that come with caveats and how the HCI and visualization community can provide the support analysts need.

Prior work has also explored the meaning of caveats and uncertainty to data workers and data professionals, specifically with an eye on how uncertainty affects their analysis. Skeels et al. conducted an interview study with professionals from various domains and classified the types of uncertainty that domain experts encounter [74]. They reported that analysts used qualitative labels to describe uncertainty, but the labels were rarely stored along with the data. Boukhelifa et al. reported on the strategies that domain professionals employ to deal with uncertainty: understand, minimize, exploit, and ignore [7]. Hullman found more evidence as to why visualization practitioners actively choose to omit uncertainty in their visualizations, citing authors' concern of overwhelming their audience with too much information [33]. Nowak and Bartram [59] conducted studies with avalanche forecasters and found that they faced ambiguity in their work and used approaches such as reasoning based on personal knowledge and professional exchange to deal with those. Our work further explores how analysts view and communicate uncertainty: it highly depends on the situation and context, analysts' roles within their teams, and their perceived professional responsibilities.

3.2 Methods of Documenting Data and Analysis

Annotations are a common way to document data and analysis. They are often added to visualizations to provide context, highlight certain points or issues, or tell a story about the data [64]. Our interview investigates how analysts document their own data analysis and which tool they choose for the documentation. There has been much research on how to build tools to better support such annotations. Kim et al. reported that study participants in a laboratory setting recorded patterns on statistical distributions using textual annotations and identified trends and anomalies using graphical annotations [42]. Annotations can also help analysts to revisit and reflect on their findings, or help with contextualizing data [54]. Annotations appear in many forms in data visualization: text [26, 53], symbols [29], sketches [45, 67], or even audio recordings [20]. Annotation systems have also served as important tools in collaborative work. For instance, they can be used to capture insights for other users to see or to continue the analysis [81]. Sharing knowledge is another benefit that annotations can bring into collaborative settings, as demonstrated in McCurdy et al., where experts shared their tacit knowledge with peers working on the same data [53]. Rittenbruch et al. [66] described the need for more versatile visual analytic collaborative tools, because the professional analysis software they analyzed was limiting. Our work explores the practices of externalizing context about the analysis with their existing tools in much broader fields of applications.

Lab notebooks and field records are often used by analysts when conducting experiments and collecting data. These documents provide details on the process for subsequent analyses and a better understanding of the condition, quality, and caveats about the data [51]. Many lab notebooks have been transitioned into digital versions [68]. Computational notebooks, such as Jupyter [44] and R Markdown [6], are often discussed as a remedy for the issues we discuss here: they can be used to describe datasets and analysis steps, contain visualizations, and also contain executable code that (should) ensure reproducibility of analysis. Due to these advantages, significant research has been devoted to understanding how analysts use computational notebooks [28, 70] and to improve them [22, 23, 79]. Notes and records, in turn, often are transferred into method sections in publications and reports, where readers can find details about the data and analysis steps, helping them judge the validity and reproducibility of the study [38, 73]. Method sections, however, are space-limited, and details are omitted in favor of describing the main results of the publication. Additionally, detailed methods sections are the purview of scientific work—other data-driven domains often do not have a similarly rigorous culture.

Metadata, "data about data" [65] or "a statement about a potentially informative object" [63], is another medium to communicate the structure and information about the data. Metadata ensures the meaningfulness of the data [18] and provides critical information about the data [25]. Burns et al. compared differences of data visualizations shown with and without metadata and demonstrated that metadata imbues more trust and persuasiveness of the visualization [11]. In our interviews, we explored how participants utilize these established mediums to record caveats about their data; some of our participants selectively use these mediums to record caveats about their data, but more intuitive and integrated solutions are desirable. Existing solutions still present a barrier to understanding for collaborators with no statistical background and often do not provide intuitive visual information.

4 METHODS

Our interview study is inspired by our previous work on data hunches [46], where we conceptualized analysts' knowledge about how data partially represents the phenomenon of interest and proposed techniques to record data hunches. In this study, we wanted to better understand the role of expert knowledge in data analysis. To this end, we recruited a mix of academic and industry analysts to elucidate how domain experts and analysts apply data hunches in their workflows.

4.1 Participant Recruitment

We sought participants conducting data analysis. All participants are professional data analysts with degrees in their domain or fields such as mathematics and statistics. Notably, none have training in computer science. Most participants have extensive domain knowledge, and all have considerable experience in their domain. We recruited analysts through personal connections and used snowball sampling to identify additional participants. We contacted potential participants by e-mail. The participants (4 men, 10 women) have a range of experience (4 to 30+ years) and work in a variety of fields such as civil engineering, psychiatry, and policy making (see Table 1).

The interview protocol was submitted to the University of Utah IRB and deemed exempt from review. The participants gave informed consent to be in the study and to be audio-recorded before the interview. Participants were not compensated. We guaranteed anonymity, which helped us elicit unfiltered opinions on data—which was particularly important for participants in the public sector, since they could not speak for their organization. We prioritized in-person interviews because they are more conversational, and may provide richer responses [36]. 13 out of 14 participants were located in physical proximity to the interviewers.

4.2 Interviews

The goal of our interviews was to study *if* analysts use expert knowledge in their workflows, and if so, how. We conducted two pre-pilot interviews with lab members to test the interview script draft and solicit feedback on the procedure and structure. We then conducted two pilot interviews with collaborators who met the inclusion criteria to test the outcome and modality of the adjusted interview structure. We then conducted 14 semi-structured interviews, 12 in-person and 2 remote. We scheduled interviews as the project progressed and decided to stop recruiting new participants when we reached saturation, noticing that no new topics were brought up. The interviews were conducted by two authors using the pair interview approach [3]. Lin asked the prepared questions and guided the conversation, while Lisnic observed the conversation, took notes, and followed up with additional questions.

The interviews were scheduled for an hour and divided into three parts: *warm-up*, *current work practices* related to the role of knowledge in interviewees' data analysis, and *feedback on a technology demonstration*. We include the interview script in our supplementary materials. In the 15-minute *warm-up*, we first asked participants about their demographic background and relevant experiences, followed by a short activity where they were asked to write down the names of data-driven projects they had worked on. During the pilot interviews, we noticed that our participants got "stuck" with the project example they initially picked and tried to answer all the questions using the same example. Asking participants to refer to the list of projects helped us later in the interview to cover a broader range of topics.

We then asked participants to pick an example from the list they wrote down and give a high-level walk-through of their analysis process. The *warm-up* helped us familiarize ourselves with their domain and analysis flow, from obtaining the data, through processing and analysis, and to decisions and interpretations eventually made.

We then transitioned to the *current work practices* section, which lasted about 35 minutes and was the main part of the interview. We asked three questions: (1) Can you pick out an example where the data just "did not look right" to you or your colleagues? (2) What could be the reason for it? (3) What did you do about it? All participants were able to recollect an experience to answer these questions. We followed up with additional questions, such as how they dealt with situations themselves and within their team, and how data caveats affected their analysis deliverables. This part of the interview provided us with rich responses on how diverse problems surface in data analysis and different approaches participants take on mitigating these problems.

Finally, we transitioned to *feedback on a technology demonstration*, shown in Fig 2, where we presented our prior work on recording and communicating data hunches [46]. The demonstration uses a scenario where a faculty member reviews the number of enrolled students by research area in their department. It also shows how the faculty member can share their data hunches and collaborate with others, and how the data hunches are integrated into the existing visualizations when presenting to others. After the presentation, we asked whether participants were currently recording data hunches in their workflow, and how the participants might see usage of these or similar techniques in their work, if there were no technological limitations. The final segment of the interview helped us understand how to better support documenting data hunches, and served as a springboard to talk about tools and technological interventions the visualization community could develop to serve analysts working with data with caveats.

4.3 Analysis

We used `Otter.ai` to transcribe the audio recordings of the interviews, followed by manual corrections. We employed a theoretical thematic

	Field	Title	Specialty	Exp. (Y)	Tools Used	Data Used	Deliverable
P1	Psychiatry	Professor	Suicide and Autism	10+	SAS, R, REDCap	Numerical data, Medical billing data	Research Manuscript
P2	Atmospheric Sciences	Professor	Snowfall Prediction	30+	Python	Meteorological measurement data	Research Manuscript
P3	Psychiatry	Professor	Genealogy and Suicide	30+	Diverse	Numerical, Genotype data, Medical billing data	Research Manuscript
P4	Atmospheric Sciences	Post-Doc	Rainfall Prediction	5	Python (Matplotlib)	Meteorological measurement data	Research Manuscript
P5	Civil Engineering	Engineer	Disaster Prevention Models	4	HEC-DSSVue, Excel	Geographical measurement data, Image, Map	Model Report Recommendation
P6	Chemical Engineering	Professor	Air Quality	20+	Visualization Dashboard	Sensor measurement data	Project Dashboard
P7	Government	Strategy Manager	Housing and Eviction Program	10	Power BI	Numerical data	Policy Recommendation
P8	Atmospheric Sciences	Science Officer	Weather Forecasting	10	Internal Tool	Meteorological measurement data, Images	Forecasts
P9	Environmental Econ	Consultant	Consulting for Legal Purposes	5	Excel, GIS Software, R	Text, Map, Numerical data	Reports
P10	Government	Politician	Public Health Legislation	14	Stata	Numerical data	Policy
P11	Government	Data Analyst	Human Services	5	Excel, Power BI	Numerical data	Dashboards, Reports
P12	Education	Specialist	CS Education	20	Power BI, LucidChart	Numerical data, Self-reported data	Policy Reports, Resource Allocation
P13	Government	Specialist	Public Defense Policy Analysis	10+	Excel, Tableau	Numerical data, Self-reported data	Reports, Recommendations
P14	Epidemiology	Program Manager	Infectious Disease Surveillance	4	Visualization Dashboard	Epidemiological numerical data	Dashboards, Healthcare Reports

Table 1: Characteristics of the 14 participants across different fields in academia and industry and the tools and data they use.

analysis approach to analyze the interview transcripts [10], guided by our research questions. Three authors (Lin, Lisnic, and Lex) reviewed each full transcript independently to label notable segments, then they met to discuss the labels they used on the interview scripts and the key findings they synthesized. On average, it took an hour to read and annotate a transcript and another hour to discuss the interview. We paid close attention to statements that provided new or surprising perspectives, especially on how participants dealt with or communicated data hunches in their workflow. The labels helped us notice overlapping themes from the transcript. Lin took notes and organized the notes and interview snippets into themes on a virtual whiteboard (available in supplementary materials) during and after each analysis session. We then went through the identified themes from the first round of analysis and categorized them into three groups: (1) the relationship between data and reality and how knowledge fills the gap between data and reality (Section 5), (2) current practices in dealing with imperfect data (Section 6), and (3) interventions for better communicating data hunches in analysis workflows (Section 7). These themes cover the full workflow of an analyst, from data collection and cleaning to analysis and interpretation, to finally delivering the analysis outcomes. Notably, our experiences engineering, working on design studies, and engaging with critical data studies influenced the ways that we read, categorized, and assessed the interview material.

5 The Role of Expert Knowledge in Data Analysis

We present the scenarios in which our participants found data insufficient and how they incorporated experts' knowledge in their analysis in this section. Additionally, we highlight how expert knowledge and input help bridge the gap between data and reality, enhancing the contextual accuracy and relevance of our participants' analyses.

5.1 The Relationship of Data and Reality

Participants are aware that data is shaped by socio-technical contexts and frequently repurposed. Understanding the context of data is a critical part of their analysis.

Data is not able to perfectly or completely represent the world [9, 13, 16], and all of our participants were acutely aware of the gaps between their data and the phenomena they were analyzing. Several of our participants described how socio-technical contexts—infrastructures, organizations, cultures, relationships, human behavior—shaped what information their data contained, and what it was missing.

For example, P7 described the ways that the US legal system dictated what data could and could not be collected about families affected by eviction court cases. P7 was studying how much the COVID-19 pandemic affected the local eviction rate and whether her agency could provide more support for people. P7 lamented that she was not able to have a good estimate or overall picture of the eviction issue due to the court not recording data about minors:

"*P7: Many of these cases are going to be families with children. And we have no idea how many kids there are. So think of this number as the low end.*"

Study participants also worked with data collected by equipment such as sensors (P6), satellites (P2, P8), and laser imagery (P5); however, they consistently noted that even sensor data is shaped by its context. For example, P6 installed sensors in various locations to collect air quality data and noted various environmental causes that impact the sensor measurements:

"*P6: So is it somebody smoking under the sensor? [...] Or is there a barbecue going on? Is there a fire? Or is it a malfunctioning sensor? Or did bugs [...] move into the sensors [...]? Those are just some of the issues that we deal with.*"

Across interviews, we heard stories about how data was filled with caveats, shaped by the contexts in which they were constructed. Nevertheless, participants often repurposed data to suit their analysis needs. One of our participants (P4) used meteorological data collected by a foreign institute to study rainfall. P4, on the other hand, used the data to study snowfall models. However, as he dug into the data, he failed to get meaningful results and finally realized that the data was not processed as expected, making it difficult to use for his purpose.

"*P4: So this is an auxiliary artifact of us trying to use the data for more than its original purpose.*"

Participants P1 and P3, who studied suicide risks, were using data labeled with disease codes (ICD codes), collected in a clinical context, to understand patient diagnoses. However, the codes were originally recorded for billing purposes, which results in instances where certain diagnoses may not represent the underlying truth. As Bowker and Star note, classification systems are embedded in social, cultural, and historical contexts [8]. To resolve this tension, P1 and P3 valued the input from clinicians with direct knowledge about the caveats:

"*P3: This is one reason to make sure that [...] your team includes some clinical folks who can tell you [...], 'This is a billing code, guys, remember, it's a billing code. This is how they can charge money for it. Or this is how they can access a certain class of drugs to treat a person. And so it's imperfect.'*"

The "perfect" dataset for a particular analysis project is often unobtainable or does not exist, which leads participants to seek datasets that are "good enough". The trade-off between accessibility and quality is often an issue that our participants face. Participants employed different methods of working with caveats to fulfill their analysis. In the instance of snowfall modeling, knowledge about the way in which the data was processed allowed the analyst to eventually use the data; yet this knowledge was not readily available.

Participants are well-versed with the caveats that come with the data.

In our interviews, we carefully posed our questions and avoided using the term *uncertainty*. Out of our 14 participants, only 2 participants (P2, P8), both working in weather forecasting, brought up uncertainty to describe the issues they faced with their data. Even though many data caveats that participants described could be labeled as qualitative or quantitative uncertainty, participants did not use these terms. We suspected that participants' expectations of data being imperfect and messy could be one reason for this:

"*P9: It's never perfect. I'm not convinced I'll ever find a data source that's like 100% perfect. I at least haven't yet.*"

One participant expressed great faith in data in the abstract sense, but then quickly acknowledged that her data did have issues:

"*P12: Numbers don't lie. Well, sometimes they did in my [use] case, but really, numbers don't lie.*"

5.2 Knowledge Fills the Gap Between Data and Reality

The primary way that participants try to fill the gaps between the available data and reality is by applying domain expertise or contextual knowledge about the data. Our participants often utilize their own prior experience or familiarity with the data and solicit input from domain experts, more senior and experienced colleagues, or from individuals who can provide crucial context, such as those in local communities.

Participants pull on diverse expertise to develop an appropriate interpretation of the data.

Soliciting the help of subject matter experts can uncover important caveats in the data that improve the analysis. The workflow of P9, a consulting analyst, provides an example of utilizing domain knowledge in analysis. P9 typically works in teams that hire external experts on the subject matter. In one instance, her team was tasked with calculating the monetary value of forests through history, and an academic with extensive knowledge of the history of land in this specific area joined their team of consulting analysts:

"*P9: We came up with a certain value for forestry in that time period. And [the land expert] said, 'Wait a second, there was this huge fire for multiple years in this area. You can't be attributing X dollars when there was no forestry activity happening because of this fire.'*"

This caveat was not known to P9, nor had it been documented in the data and resources available to her. However, this caveat was critical in P9's analysis and would not have been uncovered if not for the domain expert explicitly sharing the knowledge. Awareness of this single caveat in the data opened the door to investigating and uncovering more—the team researched other fire incidents in the area and adjusted the calculations accordingly.

Important input may also come from more senior colleagues who have experience working with a specific dataset or in the relevant context. P13, who works for a regional government, described her experience of being the only data analyst in her office as being in a "*skill set silo*." Because of this, she often contacts analysts in other departments to double-check her analysis results:

"*P13: I'll frequently do gut checks, like, 'Hey, my analysis says this. Does that make sense to you?' [...] Without that, I would be putting out a lot of very poor information.*"

We can categorize those inputs from colleagues and experts as tacit knowledge [77], "non-codified, disembodied know-how that is acquired via the informal take-up of learned behavior and procedures [32]." This definition aptly captures the experiential and procedural knowledge gained through one's professional duties. However, we discovered that this characterization falls short in comprehensively describing all types of knowledge that our participants utilized to bridge the gap between data and reality. Specifically, it inadequately accounts for **situated and lived experiences**—instances where individuals are physically close to the data or are the subjects that the data represents.

Expertise is not limited to academic or professional credentials, but rather encompasses situated and lived experiences.

Participants often sought the knowledge of individuals with such situated or lived experience about the data. For example, P5, a civil engineer, used data collected using LiDAR (a laser mapping method) on the depths of riverbeds to develop a disaster mitigation model, but the data suffered from inaccuracies. To remedy this, his team had to solicit the help of a local partner:

"*P5: We have a local partner who says, 'The channel is 20 feet deep.' But our LiDAR is showing that this is 15 feet deep. We'll say, 'Okay, we know it should probably be 5 more feet.'*"

Similarly, P6, a chemical engineer, deployed air quality sensors in various communities and regularly monitored air quality through a central dashboard. In times of anomalous air quality readings, she would first email the local community to check for any special events that might have impacted the readings:

"*P6: We'll notice the levels are high and we'll be like, 'Hey, is anything going on?' And they're like, 'Yes, there's a controlled burn over here.'*"

In both scenarios, the efforts of P5 and P6 were directed towards supporting their local partners, leading them to place a significantly higher level of trust in the information supplied by these collaborators. Relying on the insights of their partner communities can also enhance the outcomes of their analyses, contributing to a more accurate depiction of the local environment they seek to represent through data.

Expert knowledge can also come from the lived experiences of the subjects of analysis. P11, an analyst in human services, discussed an example where input from workers about their working-hour data led to starkly opposite interpretations:

"*P11: A lot of staff were telling supervisors, 'We are being overworked, we have way more demand than usual, we are putting in a lot more hours.' [The supervisors] looked at the numbers and said, 'Well, your numbers looked exactly the same as the past few months.' And they ended up finding that [... the staff] were so busy that they were not entering their data.*"

In these instances, the knowledge from "non-experts" played pivotal roles in helping our participants grasp the intricacies of reality and identify the shortcomings in the data. These local communities and individuals emerged as the true experts when it came to understanding the nuances of the data in these specific cases.

Participants sought expert knowledge not to override data, but rather to augment it in their decision-making process.

Several participants brought up that the main goal of adding expert knowledge to their analysis is not to achieve perfect numbers, but rather to come up with better actionable recommendations. For example, when P9 was working with the land history expert, she would double-check the land value coming out of her analysis with the expert to verify her method's soundness, and the result was within a reasonable range:

"*P9: [The expert] would look at the numbers we came up with, and [see if] they seemed reasonable to him. It's all about ballparks, right? We were not arguing about individual dollars; it was like, 'Is this in the realm of the right number of millions of dollars that we'd be expecting?'*"

Participants also stated that their audiences typically understand that data is not a perfect representation of reality. Hence, the goal was to reach a satisficing rather than *the ideal* outcome, and not to *"waste too much time making it perfect"*(P7). As a result, P7 discussed that precision is less of a priority than finding actionable directions of work.

"*P7: I'm working with a reasonably sophisticated audience. People want data; they know that it's imperfect. People expect me to follow up with, 'Here's what we're not sure about.' Or, 'Here's what we haven't double-checked yet.'* "

The challenge of applying knowledge to analysis and making mental adjustments to data under practical time constraints is amplified in rapidly developing high-stakes scenarios. Similar to the story in the opening of this paper, P8, who works in weather forecasting, discussed the role of expert knowledge in using radar data to rapidly distinguish between benign hail and a potentially destructive tornado:

"*P8: Being able to identify when it's the real thing and when it's not is really important. [...] Putting out a tornado warning and alerting people [of a] damaging tornado coming is a really important decision to make, and you're making it under time pressure [...] You have to be able to quickly discern almost on the fly with just what you know about how the storm should work.*"

6 Current Practices for Documenting Data Hunches

As discussed, participants make various explicit and implicit adjustments in an attempt to compensate for data imperfections, as the producers of data hunches. These adjustments are recorded and communicated to a different extent and using different media by our participants. They tend to use tools that are readily available to them and have varied personal preferences on how much they record. Most of our participants do not directly act on the data they analyze themselves, but rather communicate the data and their insights to different audiences, ranging from peer analysts, managers, policymakers, the scientific community in a field, to the general public. We observed that the methods and extent of how data imperfections and caveats are communicated vary more based on the audience and their expectations than based on the extent or type of data issues, and our participants played both the producer and consumer roles when communicating data hunches with others.

6.1 Recording Data Hunches

> Participants often do not document their analysis decisions and the ways in which they account for hunches about data. Written records are often incidental and not accessible to others.

Few participants kept detailed records about data caveats and knowledge that was relevant to the analysis. The majority of participants did not document their analysis process at all. The exception was participants (P2, P5) who needed to submit a report or record of their analysis approach, such as methods sections in research manuscripts (P2) or formally reviewed reports (P5). These participants mentioned that they kept detailed documentation of their analysis. Other participants' analyses were often ad hoc: they analyzed the data as required in the workflow. The most common form of a permanent record was incidentally recorded conversations, such as e-mails or chat histories. However, these communication logs are only accessible and hard to archive due to limiting factors, such as chat history expiration.

"*P7: I would guess that my inbox ends up becoming a form of my notes, or we use WebEx chat. [...] But it's not documented in any sort of like institutional knowledge transfer way, which is bad.*"

Study participants often use email and chat apps, in which they post text (P1, P9, P13) and screenshots of visualizations (P5, P6, P7, P13), to elicit feedback from peers or domain experts, echoing Rittenbruch et al.'s finding [66]. In return, the outcome of these communications becomes their temporary documentation. Many participants identified the issue of a lack of long-term documentation, but found it hard to properly track knowledge input in their existing workflow due to resource or technical limitations. Particularly, a lack of support in tools is cited frequently as preventing proper documentation of qualitative knowledge. Additionally, participants relied on a variety of analysis tools, such as Excel and Python (see Table 1), and collaboration platforms, including email, Zoom, Microsoft Teams, and Slack (P1, P2, P3, P7, P9, P13). The separation of tools used to discuss and analyze the data further complicated the process of annotating and sharing insights and data hunches, as well as maintaining cohesive documentation.

Of 14 participants, only one analyst (P8), who worked at a national weather forecast organization, used the annotation tool in their existing analysis system to record annotations in the weather forecast for shift handoff. P8 identified some pitfalls with their internal system, where they deal with data that is constantly updating, and where their annotations are not preserved as the data changes:

"*P8: One being, if somebody just doesn't look [for it] to see that there are comments. [...] Or, conversely, maybe [the annotation] was a good change when they made it. And now new data and new forecast have come in, and have surpassed the [annotation] that they made. [...] Then you have, what was a good change, that's not as good as the new data now passed forward, you would want to actually use the new data that's better, rather than what the previous change was.*"

As described in Section 5.2, many of our participants expressed that the main goal of their analysis often exclusively consists of recommendations and synthesized outcomes. Much of their audience did not expect to see the detailed process of their analysis, but rather actionable items. We suspect that the expectations may have an influence on the level of detail in data hunch and analysis documentation practices.

> Participants are generally aware that their lack of recording hunches is problematic and have encountered problems caused by it.

Our participants reflected on their past projects and shared that perhaps a more thorough documentation of knowledge injected into analysis would have resulted in a smoother workflow. P13 identified herself as the lone analyst in her department. Instead of explicitly documenting caveats in her workflow, P13 made mental notes. However, the lack of recording becomes problematic when new members join or leave the analysis team or when the project hibernates. When an intern joined the department temporarily, she verbally communicated to the intern how to treat the data because of all the data caveats, such as *"you can ignore that data from X, because I know they're wrong for various reasons"*. This knowledge exchange happened ad hoc, and P13 was unsure whether she covered all the data caveats exhaustively. She also stated that such a lack of documentation has led to wasted effort before:

"*P13: I probably made some mistakes by re-analyzing data that I had forgotten I'd already sifted out.*"

P4 also faced an issue related to a lack of documented data hunches. He downloaded atmospheric data from a foreign institute, but the data did not seem to make sense in his context. He brought questions about this data issue to the foreign scientist, who stated that they had processed the data, but had not documented the processing steps. Because of the lack of proper recording of data caveats, P4 spent extra time and effort trying to make sense of the data.

> For some participants, external pressures lead to documentation of the data analysis process and data hunches.

We found that the scientists and engineers in our participant pool were more likely to document their processes and data hunches. For example, P2 reported taking "abundant field notes" of the measuring equipment condition and weather context for weather data collection, and then documenting the relevant aspects in a methods section in a research paper. Similarly, P5 prepared an appendix for his reports on flood disaster models, documenting uncertainty and other assumptions:

"*P5: It's reviewed by somebody within our district, and then it goes through an agency technical review [...] by somebody outside of the district. So there are a lot of quality checks that happen to make sure that [everything is] accounted for.*"

P14, the epidemiologist, used metadata sheets to capture caveats in their analysis to make sure they had proper data context for the analysis down the pipeline, although she stated that improving their internal documentation was a strategic goal. Overall, the practice of documentation appeared to be strongly influenced by the expectations of the target audience for the analysis deliverable. Participants were more inclined to maintain detailed notes when they anticipated that their analysis results might be closely examined alongside their documentation.

6.2 Communicating During Collaborative Analysis Processes

We found that, in addition to synchronous meetings, email exchange, and chat were the main communication tools that participants used to elicit domain knowledge and feedback from the experts, because the exchange of expert knowledge often happened through those mediums. However, they described being frustrated when dealing with text and screenshots: important data caveats were sometimes buried in long emails (P9); in other cases, emails were not exhaustive enough to describe the issue (P1, P9, P13). P9 described an unpleasant experience where the data provider did not communicate the assumptions that went into the data collection in their email exchange, and that the analyst in charge of interfacing with the provider was new to the job. This mixture of inexperience and ineffective communication led to a wrong analysis, undermining a high-stakes legal case. The common issue is the ineffectiveness of conveying data caveats through words.

"*P13: In general, I email people, and I feel like a lot is lost in translation in the emails.*"

> Participants use screenshots or screen-shares of visualizations to communicate about data and data caveats, but do not use annotation features of their visualization tools.

Study participants found visualizations to be effective when conversing about data, even helping with difficulties such as language barriers (P3) and data literacy (P11). Even though participants would draw and annotate on visualizations during these meetings, the drawings were not archived; hence, the knowledge exchange was not preserved. Alternatively, participants used screenshots in chat, e-mail, or PowerPoint, which they may or may not annotate. We found no instances of annotations happening directly in a visual analytics tool that participants use, except for the previously mentioned weather forecast tool (P8).

"*P1: I'm not sophisticated enough to have some program where [...] I guess I could, but I would do it in such a clunky way. I'd like to put it into PowerPoint [...], but instead, I would just send them the figure and say, 'This doesn't look right.'*"

While P1 finds using simpler tools for annotations less 'clunky', P13 described that adding hunches and annotations to their spreadsheet, primarily used for analysis, might have an unintended consequence of making it messy:

"*P13: Even just how to leave a better breadcrumb trail from Excel is something I'm not great at. Probably messing up my pretty sheet, you know?*"

Hence, even though tools like Excel, Tableau, and Power BI (which were among the tools used by our participants) support annotation, participants used simple graphics editors to annotate the visualization at the time of sharing or simply attempted to communicate using text.

6.3 Communicating Analysis Results and Data Hunches

Participants used a variety of media to report their analysis outcomes. These mediums included research manuscripts (P1-4), project reports (P5, P9, P1-4), live presentations with slides (P9, P12, P14), and dashboards on websites (P6, P11, P14). Because of the differences in the method of delivery, communicating data hunches to the intended audiences took different forms.

> How much about caveats is disclosed is highly situational, depending on the perceived stakes, but also on the reporting format.

Participants who were academics (P1-4) most commonly use method sections in their publications to report data hunches. Similarly, participants who wrote up their analysis as reports (P5, P9, P11) documented data hunches as an appendix or as part of their reports. However, these formats are commonly kept brief and may not provide enough detail to reproduce the analysis or reuse the data [35].

When communicating their results in meetings, participants found slides and handouts to be useful media. Participants frequently added bullet points in their slides next to their static visualizations to explain the context or caveats required to properly interpret the data.

"*P7: I have a little disclaimer at the bottom [of my slides ...] that notes that [this approach is] not going to catch duplicates.*"

However, when asked about preparing and anticipating questions about the data, many participants answered that they would respond to the audience's questions ad hoc, rather than preemptively cover data imperfections in their presentations.

We found that several participants (P5, P7, P11, P13) were hesitant to communicate caveats. Some participants saw it as their professional responsibility to distill data into actionable items for decision-makers, and that they were trusted to correctly interpret the data to the best of their ability. Communicating more data caveats to leadership would increase the complexity of their interpretations, and leadership and external audiences often were not interested in the details of the analysis.

"*P7: Part of our job is to synthesize down to the main points for leadership. When it's getting to the mayor as talking points or a policy memo, if it has too much [about caveats], it's just going to be a distraction.*"

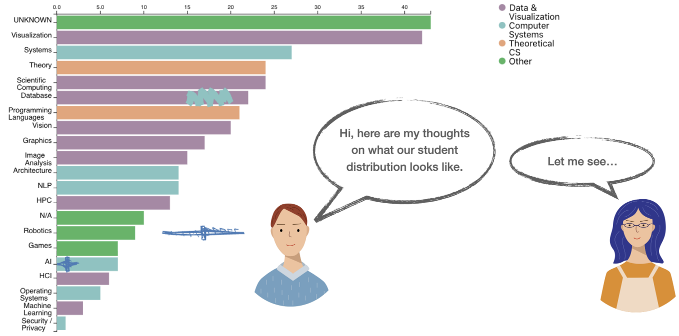

Fig. 2: A screenshot of the technology demonstration we used in our interview based on the data hunches implementation by Lin et al. [46]. The scenario illustrates how two faculty members can record their data hunches about enrollment numbers in different sub-fields of computer science directly on the visualization through interactive techniques.

We discovered different approaches towards expressing caveats about the analysis depending on the format and the type of participants' jobs. For example, our academic participants were cautious about keeping their analysis choices transparent in research manuscripts (P1-4), whereas other participants felt hesitant to express uncertainty explicitly in their analysis deliverables due to the reasons laid out above. This omission of caveats was even more amplified in verbal presentations, which our participants justified by being available for clarifying questions, if needed. However, participants also adjusted to the stakes and the certainty of their analysis:

"*P14: When we're communicating [COVID-19 related] data to them, because we knew that it would have big consequences in terms of policy recommendations and political action, we are very careful to present the limitations upfront. We'll generally provide a written copy of limitations [...], and we repeat it often throughout the presentation.*"

Participants more often added disclaimers to deliverables that were used asynchronously, such as dashboards. But even there, the focus was to avoid making the results confusing. For example, P14 would refrain from adding all data caveats to her public COVID-19 dashboard:

"*P14: There's no reason to put [data caveats] on the website if people aren't going to understand it, or if they're going to misinterpret it.*"

However, she would add additional annotations to her dashboard if she would learn that an aspect was regularly misinterpreted:

"*P14: If it's a broad misunderstanding, and we're getting a lot of public calls, I might add something to the dashboard that has something embedded in the figure, [such as] shading for when Delta started or when Omicron started.*"

7 Exploring Interventions for Recording Data Hunches

During the *"feedback on a technology demonstration"* section of the interview, we introduced the definition of data hunches to the participants and provided a brief demonstration as described in Section 4.2. The demo included the basic workflow of recording and communicating a data hunch and how the data hunch looked when being recorded in the visualization using sketchy rendering based on our previous work [46]. We presented a use case where a faculty member recorded their data hunches on a bar chart about students' research area distribution in their department, and then shared their data hunches with another faculty member. The demonstration illustrated both the concept of data hunches and possible technical solutions to recording and communicating data hunches during collaborative analysis scenarios.

> Annotations or other ways to quickly express data hunches on top of interactive visualizations can help stakeholders with various backgrounds get on the same page.

Participants liked the collaborative and visual aspect of the prototype (P2-5, P7-9, P11, P13). P9 commented on using the visualizations to make sure that everyone was on the same page on the project:

"*P9: So if you had something like [the data hunches prototype], where then the analyst was sharing their screen and making the ad-*

justment, [to show] what they think experts are talking about, and the expert could see it adjust in front of them, then everyone can make sure they're understanding. [...] I think that would kind of bridge the knowledge gap between the data people and the experts in a successful way."

The techniques we presented used interactions directly on a data visualization, which participants considered to be an easy way to express opinions. P11 reflected that most of the domain professionals she worked with were not great at verbally expressing their opinions and knowledge about the data. An interactive visual approach would be a good option for these collaborations:

"*P11: I think sometimes they just don't really know how to phrase what they want to see. [...] this to me seems useful [...] for people who aren't the [visualization] designers to be able to offer feedback.*"

P10, a state-level politician, was interested in how visually expressing data hunches could promote discussion among policymakers, rather than having them dismiss an opposing point outright:

"*P10: I think there's a great opportunity for [visually expressing and communicating data hunches], especially if it's a policy issue that people really do want to collaborate [on] and everybody agrees that something needs to happen. We just have to come to terms on how to get there, [and] then there's really good opportunity for a model like this. I think a feedback loop like this could do a lot of good.*"

Several participants commented on how the data hunch techniques could be helpful for asynchronous knowledge transfer. P4, a post-doc researcher, discussed how he could use sketches and annotations to provide his knowledge on the caveats of a dataset:.

"*P4: If I'm not around to point out the nuances, then there's no recording of [...] the issues. I think [recording data hunches visually] would be a useful way of archiving somebody's hunch on the data [...], [so it's available even after] I've been removed from my PhD work for a couple of years. So [...if] a new student [starts to] work on a similar project, and [my advisor can] say, "Hey, here's some issues with the data that we have recorded. Go check them out and why." That would be a good way of recording it.* "

Our participants, particularly ones who worked with public policies, also raised their concerns, worrying that if tools with data hunches were to be open to the general public, people would use those tools to serve their agenda. P13, for example, would refrain from opening her data report to the general public and allowing annotating data hunches on there, as she was concerned that people drew their data hunches without proper context or knowledge to back up their hunches (a common occurrence in visual misinformation on social media [47]). P10, a politician, was particularly aware of the bias that analysts may have due to their political background and analytical experiences. She used an example where a professor in economics and a policy analyst at a taxpayer association may have conflicting data hunches about a tax policy outcome. As a result, she discussed the importance of utilizing identities and reasons behind a data hunch to make a comprehensive interpretation:

"*P10: [To establish trustworthiness of the data hunch], a big part of it's going to be the credentials that they come with, and every person that provides information would have to have some explanation [of] who they are and where they come from.*"

However, the usefulness of annotating data hunches remained situational, as several study participants (P5, P7, P11, P13) agreed that data hunch techniques would be more suitable for discussion and knowledge sharing within the core team, rather than for presentations and dissemination.

Fatigue with tools is prominent. Participants do not want an additional tool in their already complicated workflow.

When speaking about using a tool to record data hunches, participants' desire for a separate solution along the lines of what we presented with the data hunch prototype was in tension with their aversion to adding another tool to their toolbox. Changing and juggling between tools can be challenging, especially during a meeting, as P9 described the process as "*scrambling between apps a million times*". Some participants were concerned that a new tool might be a burden for experts who have valuable insight but are not avid technology users.

"*P3: So it would be a little bit of a concern that, you know, you'd have somebody who's a busy clinician, and they have really great insight. And they are not going to use the tool, because you know, they essentially use Excel, and maybe Word. [...] The best insights are not necessarily [from] a very sophisticated tool user.*"

Our participants, especially those with 10+ years of experience, preferred solutions that fit their existing workflow. They already faced the *yet-another-tool* problem (explicitly mentioned by three participants, P2, P3, P9), and did not want to add more tools to record and communicate data caveats. Many participants stated that they would prefer an embedded solution over a new tool (P1, P3, P8, P9, P11). In contrast, two of the participants (P7, P12) expressed enthusiasm for a new tool.

"*P2: I think the biggest problem is [that] there are too many tools. [...] it's a pain in the ass to be perfectly honest. [...] And it is a huge suck on time having to juggle all these different applications. So yeah, what I would want is I want to embed it into something that works for everything.*"

8 Discussion

Based on our findings, we discuss the significance of expert knowledge in the analysis process and how the visualization community could provide interventions to support experts in documenting their knowledge of data hunches.

8.1 The Role of Expert's Knowledge in Data Analysis

Prior feminist and critically-oriented work characterizes data as an artifact of decisions: a culmination of the specific and situated contexts in which they were constructed [1, 14, 17, 43, 46]. The construction of data leaves it with gaps and caveats such that for data to reflect reality fully, data requires context [43]. In our interviews, participants discussed many different ways that they understood and worked with the limitations of their data.

Expert knowledge often complements the data, piecing together the spaces between data and reality. Surprisingly, analysts found expert knowledge beyond traditional domain experts as conceptualized by the visualization community [72]. Those experts ended up being anyone close to the data—aware of how data is constructed or of the environment from which data is derived. However, we heard many accounts of how recording this knowledge is brittle and unsystematic: scattered across ephemeral records like chat histories, one-off emails, or communicated in a meeting. Thus, the lack of documentation makes reanalysis and reproducibility challenging, creating a barrier for other analysts outside the discussion to join.

Furthermore, our participants never expected their data to be perfect. The data were imperfect for many reasons, including errors in measurement devices, human factors, the data being originally generated for different reasons, or better data being simply unattainable. Caveats about the data were often not communicated for a variety of reasons, most prominently because our participants felt that it was part of their professional responsibility to make easy-to-interpret and actionable analyses from the data. They were trusted to communicate what was necessary, and their delivery excluded many of the caveats that they worked with. This finding complements what Hullman found about why visualization authors do not communicate uncertainty: because showing uncertainty is difficult for the author, and reading charts with uncertainty is difficult for the audience [33]. We saw evidence in support of both, but the role of the expert as a trusted party that abstracts complexity was unique.

The literature on uncertainty addresses only part of the concern when it comes to visualizing the imperfections of data. Uncertainty expressions like confidence intervals, hypothetical outcome plots [34], and ensemble plots [60] focus on conveying the uncertain nature of the data and are well-studied within the visualization community. And yet, throughout our interviews, most participants never brought up uncertainty when describing their data. Instead, we found that most participants described how they adapted their workflows to account for data and its caveats. They were, in fact, *certain* about the data's limitations and were able to reduce the effects of the limitations through knowledge of the data's context.

Data is not perfect—our participants did not believe that it is, and neither should the visualization community. Across interviews, we saw the importance of context for our participants' understanding and handling imperfections in their data. Within analysis scenarios, knowledge about the data is more important to record for purposes of reanalysis than for communicating final results. In contexts of trust and expertise, there is a common understanding that the data is meaningless without knowledge of its context.

8.2 Design Opportunities

Our interviews show that the majority of participants acknowledge that they do not document their analysis processes or data caveats and hunches well. Many participants have reported suffering from a lack of documentation when they or someone else revisited a dataset or an analysis. This indicates that to improve documentation practices, improving tooling can only be part of the solution. The organizations or communities in which data analysts operate need to encourage best practices or even mandate documentation [25]. While such practices are already required or at least incentivized in many scientific and engineering disciplines, they may be less common in fields like government or social services.

While the demonstration of a previously developed tool for externalizing and communicating data hunches [46] seemed to resonate with our participants, the interviews made us doubt that a standalone tool could be successful with the analysts we interviewed. The skepticism about new tools and the fatigue resulting from the fragmented analysis tool space [4] was palpable. Furthermore, our participants did not use the annotation capabilities of the tools they already had at their disposal. Participants chose other approaches, citing the desire to keep their analysis work 'clean' and separate (P13) or the efficiency of more direct communication (P9) compared to annotations in an analysis platform.

Consequently, we join previous works [2, 5, 55] in the call for rethinking how we design and develop visualization interfaces, especially when the goal is real-world adoption. Instead of developing *yet-another-tool*, we argue for meeting analysts where they are in their analysis workflows. For example, we envision designs that lower the friction to annotate and record hunches in existing environments, such as built-in annotation capabilities on top of screenshots for communication tools like Slack. These lightweight methods for capturing hunches could also be designed to support annotations from many people, including field workers and others with close knowledge of the data and its context.

We were surprised that computational notebooks, like Jupyter, were not mentioned in all of our interviews, even though many of the issues discussed by our participants could be addressed using such tools. We also note that only three of our participants reported using programming languages within their analysis—R (P9) and Python (P2, P4). We speculate that a (possibly large) number of data analysts' processes cannot meet standards for reproducibility laid out by various scientific bodies [30, 58]. We see this as an opportunity to explicitly capture the ways expert knowledge shapes and impacts analysis processes in visual analysis tools and processes.

First, we need to make GUI-based visual analysis tools reproducible. The GUI-based tools our participants use do not support annotated histories or workflows, unlike various research prototypes [22, 24, 27], pointing to a potentially overlooked area in research on reproducible visualizations. For example, there is typically no way to comment on why some data was filtered out in the tools used by our participants. Hence, we call on commercial tool developers to consider making analytical provenance available and salient to users, and the scientific community to continue to innovate in that space.

Second, while data hunches are often expressed when viewing data through visualizations, we believe it is important to also capture data hunches at the data level. As a low-tech intervention, we encourage the extension of metadata files, data dictionaries, or data sheets [25] to not only document the *what* that is in the data, but also the *why* and *how*. Ideally, datasets should be published or archived together with a reproducible analysis story that makes it clear how the data was used.

Third, we argue that we need to develop guidelines and standards for documenting heterogeneous analysis processes, especially those that include interactive tools. These guidelines should detail best practices for acknowledging and capturing analysis steps and externalizing expert knowledge that goes into decision-making. In our interviews, we found that external pressures and established guidelines lead to better documentation practices. The visualization research community could use this opportunity to establish guidelines for submitting data stories alongside datasets, improving reproducibility by offering support such as templates and repository structures.

8.3 Trust, Biases, and Data Hunches

We reported that our participants have concerns about biases built into data hunches in applications. We envision that data hunches can enrich our comprehension of information, with all involved parties contributing positively based on their expertise and insights. Nevertheless, data hunches might be wielded to explain away inconvenient data points, bolster an analyst's preconceived notions, and serve as a premise to support existing misinformation arguments [48]. There is a risk that data hunches could amplify biases and subjectivities of experts, rather than providing a stronger objectivity through rich, diverse perspectives, such as "data hunches" that are rooted not in a deep understanding of the data but rather in biases or mistaken beliefs. The presence of biases in data hunches is compounded when other analysts encounter and subsequently base their interpretations on these potentially skewed data hunches, leading to an unjustified outcome based on the data [21].

To facilitate the identification of biases in the human-in-the-loop analysis, Wall et al. [78] proposed a quantification framework to indicate any underlying biases in the analysis. Such techniques may help analysts judge the potential unjust bias behind a data hunch. Our design framework for data hunches [46] mandates the inclusion of explanations and justifications for each recorded data hunch and suggests that they may only be trustworthy and effective within existing networks of trust [62], as also hypothesized by our participants (P5, P7, P11, P13). By providing reasoning and contextual information for these hunches, analysts can better assess them from a comprehensive standpoint.

Nonetheless, how biases can be both alleviated or reinforced by a data hunch remains an open avenue for further inquiry. It is important to investigate whether those consuming data hunches would interpret them with the same perceived objectivity typically afforded to quantitative data. Making the expert labor behind a data hunch visible could encourage a critical interpretation of the data and more transparently situate it within its social and historical context [43]. However, as discussed by Star and Strauss [75] and mentioned by P10 and P14, visibility could be highly political and also lead to resistance if the author's views are perceived as partisan or challenge existing dominant perspectives.

9 Limitations

Our study has several limitations that are common in interview-based research in the HCI communities. Firstly, our sample of participants was not randomly selected but recruited from our professional networks. This may have introduced biases into our sample, as those who are more closely connected to us or our network may have different perspectives or experiences from those who are not. Secondly, our preference for conducting in-person interviews in English limited the geographic and cultural diversity of our sample. Finally, all participants had at least a bachelor's degree, which may have limited the diversity of perspectives in our sample. Overall, we believe our study provides valuable insights into the experiences of the participants we interviewed and believe that our results generalize to other analysts with similar characteristics.

10 Conclusion

We conducted a series of interviews with analysts from various fields and levels of experience to investigate how expert knowledge influences their analysis. Our findings highlight the importance of including and documenting expert knowledge in data analysis, as well as the potential pitfalls of neglecting this information. We also collected feedback on potential interventions to support the recording and communication of data hunches more effectively. Our ultimate goal is to draw attention to how data is an incomplete representation of the reality it aims to depict, and that expert knowledge is crucial in making data a useful tool to answer analytical questions. We suggest future research directions for developing better methods to make analysis processes and data reproducible and reusable.

Acknowledgments

We would like to thank our interviewees for their time and participation in the study, and the Visualization Design Lab for the fruitful discussions and feedback. We gratefully acknowledge funding from the National Science Foundation (OAC 1835904) and from the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

[1] D. Akbaba, L. Klein, and M. Meyer. Entanglements for Visualization: Changing Research Outcomes through Feminist Theory. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):1279–1289, 2025. doi: 10.1109/TVCG.2024.3456171 2, 8

[2] D. Akbaba, D. Lange, M. Correll, A. Lex, and M. Meyer. Troubling Collaboration: Matters of Care for Visualization Design. In *ACM CHI Conference on Human Factors in Computing Systems*, 2023. doi: 10.1145/3544548.3581168 9

[3] D. Akbaba and M. Meyer. "Two Heads are Better than One": Pair-Interviews for Visualization. In *IEEE Visualization and Visual Analytics (VIS)*, pp. 206–210. IEEE, 2023. doi: 10.1109/VIS54172.2023.00050 3

[4] 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*, 25(1):22–31, 2019. doi: 10.1109/TVCG.2018.2865040 2, 9

[5] L. Bartram, M. Correll, and M. Tory. Untidy Data: The Unreasonable Effectiveness of Tables. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):686–696, 2022. doi: 10.1109/TVCG.2021.3114830 9

[6] B. Baumer and D. Udwin. R Markdown. *WIREs Computational Statistics*, 7(3):167–177, 2015. doi: 10.1002/wics.1348 3

[7] N. Boukhelifa, M.-E. Perrin, S. Huron, and J. Eagan. How Data Workers Cope with Uncertainty: A Task Characterisation Study. In *CHI Conference on Human Factors in Computing Systems*, pp. 3645–3656, 2017. doi: 10.1145/3025453.3025738 2

[8] G. C. Bowker and S. L. Star. *Sorting Things Out: Classification and Its Consequences*. MIT Press, 1999. 4

[9] d. boyd and K. Crawford. Critical Questions for Big Data. *Information, Communication & Society*, 15(5):662–679, 2012. doi: 10.1080/1369118X.2012.678878 4

[10] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. doi: 10.1191/1478088706qp063oa 4

[11] A. Burns, C. Lee, T. On, C. Xiong, E. Peck, and N. Mahyar. From Invisible to Visible: Impacts of Metadata in Communicative Data Visualization. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–16, 2022. doi: 10.1109/TVCG.2022.3231716 3

[12] A. Crisan, B. Fiore-Gartland, and M. Tory. Passing the Data Baton : A Retrospective Analysis on Data Science Work and Workers. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1860–1870, Feb. 2021. doi: 10.1109/TVCG.2020.3030340 2

[13] C. M. Dalton, L. Taylor, and J. Thatcher (alphabetical). Critical Data Studies: A dialog on data and space. *Big Data & Society*, 3(1):2053951716648346, 2016. doi: 10.1177/2053951716648346 4

[14] C. D'Ignazio and L. F. Klein. Feminist Data Visualization. In *Workshop on Visualization for the Digital Humanities (VIS4DH)*, 2016. 8

[15] E. Dimara, H. Zhang, M. Tory, and S. Franconeri. The Unmet Data Visualization Needs of Decision Makers Within Organizations. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4101–4112, 2022. doi: 10.1109/TVCG.2021.3074023 2

[16] M. Dörk, P. Feng, C. Collins, and S. Carpendale. Critical InfoVis: Exploring the politics of visualization. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, pp. 2189–2198, 2013. doi: 10.1145/2468356.2468739 4

[17] J. Drucker. Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, 5(1), 2011. 1, 8

[18] E. Duval and W. Hodgins. Metadata principles and practicalities. *D-Lib Magazine*, 8:2002, 2002. 3

[19] M. K. Elhadad, K. F. Li, and F. Gebali. Detecting Misleading Information on COVID-19. *IEEE Access*, 8:165201–165215, 2020. doi: 10.1109/ACCESS.2020.3022867 1

[20] S. E. Ellis and D. P. Groth. A Collaborative Annotation System for Data Visualization. In *Working Conference on Advanced Visual Interfaces*, AVI, pp. 411–414. New York, NY, USA, 2004. doi: 10.1145/989863.989938 3

[21] B. Friedman and H. Nissenbaum. Bias in computer systems. *ACM Transactions on Information Systems*, 14(3):330–347, July 1996. doi: 10.1145/230538.230561 9

[22] K. Gadhave, Z. Cutler, and A. Lex. Reusing Interactive Analysis Workflows. *Computer Graphics Forum*, 41(3):133–144, 2022. doi: 10.1111/cgf.14528 3, 9

[23] K. Gadhave, Z. Cutler, and A. Lex. Persist: Persistent and Reusable Interactions in Computational Notebooks. *Computer Graphics Forum*, 43(3):e15092, 2024. doi: 10.1111/cgf.15092 3

[24] K. Gadhave, J. Görtler, Z. Cutler, C. Nobre, O. Deussen, M. Meyer, J. M. Phillips, and A. Lex. Predicting intent behind selections in scatterplot visualizations. *Information Visualization*, 20(4):207–228, 2021. doi: 10.1177/14738716211038604 9

[25] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. Iii, and K. Crawford. Datasheets for Datasets. *Communications of the ACM*, 64(12):86–92, Dec. 2021. doi: 10.1145/3458723 3, 9

[26] N. Goyal, G. Leshed, and S. R. Fussell. Effects of Visualization and Note-taking on Sensemaking and Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pp. 2721–2724. Association for Computing Machinery, 2013. doi: 10.1145/2470654.2481376 3

[27] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit. From Visual Exploration to Storytelling and Back Again. *Computer Graphics Forum*, 35(3):491–500, 2016. doi: 10.1111/cgf.12925 9

[28] A. Guzharina. We Downloaded 10,000,000 Jupyter Notebooks From Github – This Is What We Learned, 2020. 3

[29] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM*, 52(1):87–97, Jan. 2009. doi: 10.1145/1435417.1435439 3

[30] B. J. Heil, M. M. Hoffman, F. Markowetz, S.-I. Lee, C. S. Greene, and S. C. Hicks. Reproducibility standards for machine learning in the life sciences. *Nature Methods*, 18(10):1132–1135, Oct. 2021. doi: 10.1038/s41592-021-01256-7 9

[31] D. Hoffman. 'I Had A Funny Feeling in My Gut'. *The Washington Post*, p. A19, Feb. 1999. 1

[32] J. Howells. Tacit knowledge. *Technology Analysis & Strategic Management*, 8(2):91–106, Jan. 1996. doi: 10.1080/09537329608524237 5

[33] J. Hullman. Why Authors Don't Visualize Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):130–139, Jan. 2020. doi: 10.1109/TVCG.2019.2934287 2, 8

[34] J. Hullman, P. Resnick, and E. Adar. Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering. *PLOS ONE*, 10(11):1–25, Nov. 2015. doi: 10.1371/journal.pone.0142444 8

[35] P. Hunter. The reproducibility "crisis". *EMBO reports*, 18(9):1493–1496, Sept. 2017. doi: 10.15252/embr.201744876 7

[36] D. R. Johnson, C. P. Scheitle, and E. H. Ecklund. Beyond the In-Person Interview? How Interview Quality Varies Across In-person, Telephone, and Skype Interviews. *Social Science Computer Review*, 39(6):1142–1158, Dec. 2021. doi: 10.1177/0894439319893612 3

[37] J. Y. Jung, T. Steinberger, J. L. King, and M. S. Ackerman. How Domain Experts Work with Data: Situating Data Science in the Practices and Settings of Craftwork. *Human-Computer Interaction*, 6:58:1–58:29, Apr. 2022. doi: 10.1145/3512905 2

[38] R. H. Kallet. How to Write the Methods Section of a Research Paper. *Respiratory Care*, 49(10):1229–1232, Oct. 2004. 3

[39] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. van Ham, N. H. Riche, C. Weaver, B. Lee, D. Brodbeck, and P. Buono. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4):271–288, Oct. 2011. doi: 10.1177/1473871611415994 2

[40] 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*, 18(12):2917–2926, Dec. 2012. doi: 10.1109/TVCG.2012.219 2

[41] B. Karer, H. Hagen, and D. J. Lehmann. Insight Beyond Numbers: The Impact of Qualitative Factors on Visual Data Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1011–1021, Feb. 2021. doi: 10.1109/TVCG.2020.3030376 1

[42] Y.-S. Kim, N. Henry Riche, B. Lee, M. Brehmer, M. Pahud, K. Hinckley, and J. Hullman. Inking Your Insights: Investigating Digital Externalization Behaviors During Data Analysis. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, ISS '19, pp.

255–267. Association for Computing Machinery, Daejeon, Republic of Korea, 2019. doi: 10.1145/3343055.3359714 3

[43] L. F. Klein and C. D'Ignazio. *Data Feminism*. The MIT Press, 2020. 8, 9

[44] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, and Jupyter development team. Jupyter Notebooks – a publishing format for reproducible computational workflows. In *International Conference on Electronic Publishing*, pp. 87–90, 2016. doi: 10.3233/978-1-61499-649-1-87 3

[45] P. Lavrič, C. Bohak, and M. Marolt. Collaborative view-aligned annotations in web-based 3D medical data visualization. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 259–263, May 2017. doi: 10.23919/MIPRO.2017.7973430 3

[46] H. Lin, D. Akbaba, M. Meyer, and A. Lex. Data Hunches: Incorporating Personal Knowledge into Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):504–514, Jan. 2023. doi: 10.1109/TVCG.2022.3209451 2, 3, 7, 8, 9

[47] M. Lisnic, A. Lex, and M. Kogan. 'Yeah, this graph doesn't show that': Analysis of online engagement with misleading data visualizations. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pp. 1–14. ACM, Honolulu, HI USA, 2024. doi: 10.1145/3613904.3642448 8

[48] M. Lisnic, C. Polychronis, A. Lex, and M. Kogan. Misleading Beyond Visual Tricks: How People Actually Lie with Charts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–21. ACM, Hamburg Germany, 2023. doi: 10.1145/3544548.3580910 9

[49] Y. Liu, T. Althoff, and J. Heer. Paths Explored, Paths Omitted, Paths Obscured: Decision Points & Selective Reporting in End-to-End Data Analysis. In *ACM CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020. doi: 10.1145/3313831.3376533 2

[50] Y. A. Loukissas. *All Data Are Local: Thinking Critically in a Data-Driven Society*. MIT Press, Apr. 2019. 1

[51] J. MacNeil and R. Falconer. When Learning the Hard Way Makes Learning Easy: Building Better Lab Note-Taking Skills. *Journal of Chemical Education*, 87(7):703–704, July 2010. doi: 10.1021/ed100303y 3

[52] Y. Mao, D. Wang, M. Muller, K. R. Varshney, I. Baldini, C. Dugan, and A. Mojsilović. How Data Scientists Work Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question? *Proceedings of the ACM on Human-Computer Interaction*, 3(GROUP):237:1–237:23, 2019. doi: 10.1145/3361118 2

[53] N. Mccurdy, J. Gerdes, and M. Meyer. A Framework for Externalizing Implicit Error Using Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):925–935, Jan. 2019. doi: 10.1109/TVCG.2018.2864913 1, 2, 3

[54] 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*, 2(3):128:1–128:28, Sept. 2018. doi: 10.1145/3264938 3

[55] 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*, 28(1):96–106, Jan. 2022. doi: 10.1109/TVCG.2021.3114798 9

[56] M. Muller, I. Lange, D. Wang, D. Piorkowski, J. Tsay, Q. V. Liao, C. Dugan, and T. Erickson. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *ACM CHI Conference on Human Factors in Computing Systems*, pp. 126:1–126:15, 2019. doi: 10.1145/3290605.3300356 1, 2

[57] M. Muller and A. Strohmayer. Forgetting Practices in the Data Sciences. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pp. 1–19. Association for Computing Machinery, New York, NY, USA, Apr. 2022. doi: 10.1145/3491102.3517644 2

[58] National Academies of Sciences. *Reproducibility and Replicability in Science*. National Academies Press, Sept. 2019. 9

[59] S. Nowak and B. Lyn. I'm Not Sure: Designing for Ambiguity in Visual Analytics. In *Proceedings of Graphics Interface 2022*, pp. 114–126. Montreal QC Canada, May 2022. doi: 10.20380/gi2022.13 2

[60] L. M. K. Padilla, S. H. Creem-Regehr, and W. Thompson. The powerful influence of marks: Visual and knowledge-driven processing in hurricane track displays. *Journal of Experimental Psychology: Applied*, 26(1):1–15, 2020. doi: 10.1037/xap0000245 8

[61] G. Panagiotidou, R. Vandam, J. Poblome, and A. Vande Moere. Implicit Error, Uncertainty and Confidence in Visualization: An Archaeological Case Study. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4389–4402, June 2021. doi: 10.1109/TVCG.2021.3088339 1

[62] S. Passi and S. J. Jackson. Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):136:1–136:28, Nov. 2018. doi: 10.1145/3274405 9

[63] J. Pomerantz. *Metadata*. The MIT Press, Cambridge, Massachusetts ; London, England, Nov. 2015. 3

[64] M. D. Rahman, G. J. Quadri, B. Doppalapudi, D. A. Szafir, and P. Rosen. A Qualitative Analysis of Common Practices in Annotations: A Taxonomy and Design Space, July 2024. doi: 10.48550/arXiv.2306.06043 2

[65] J. Riley. *Understanding Metadata: What Is Metadata, and What Is It For?* NISO Press, Baltimore, MD, 2017. 3

[66] M. Rittenbruch, K. Vella, M. Brereton, J. M. Hogan, D. Johnson, J. Heinrich, and S. O'Donoghue. Collaborative Sense-Making in Genomic Research: The Role of Visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4477–4489, Dec. 2022. doi: 10.1109/TVCG.2021.3090746 3, 6

[67] H. Romat, N. Henry Riche, K. Hinckley, B. Lee, C. Appert, E. Pietriga, and C. Collins. ActiveInk: (Th)Inking with Data. In *CHI Conference on Human Factors in Computing Systems*, pp. 42:1–42:13, 2019. doi: 10.1145/3290605.3300272 3

[68] M. Rubacha, A. K. Rattan, and S. C. Hosselet. A Review of Electronic Laboratory Notebooks Available in the Market Today. *JALA: Journal of the Association for Laboratory Automation*, 16(1):90–98, Feb. 2011. doi: 10.1016/j.jala.2009.01.002 3

[69] R. A. Ruddle, J. Cheshire, and S. J. Fernstad. Tasks and Visualizations Used for Data Profiling: A Survey and Interview Study. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–12, Aug. 2023. doi: 10.1109/TVCG.2023.3234337 2

[70] A. Rule, A. Tabard, and J. D. Hollan. Exploration and Explanation in Computational Notebooks. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018. doi: 10.1145/3173574.3173606 3

[71] D. A. Schon. Designing as reflective conversation with the materials of a design situation. *Research in Engineering Design*, 3(3):131–147, Sept. 1992. doi: 10.1007/BF01580516 2

[72] M. Sedlmair, M. Meyer, and T. Munzner. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, Dec. 2012. doi: 10.1109/TVCG.2012.213 2, 8

[73] V. Siegel. Reproducibility in research. *Disease Models & Mechanisms*, 4(3):279–280, May 2011. doi: 10.1242/dmm.008037 3

[74] M. Skeels, B. Lee, G. Smith, and G. Robertson. Revealing uncertainty for information visualization. *Information Visualization*, 9(1):70–81, 2010. doi: 10.1057/ivs.2009.1 2

[75] S. L. Star and A. Strauss. Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work (CSCW)*, 8(1-2):9–30, Mar. 1999. doi: 10.1023/A:1008651105359 9

[76] M. Tory, L. Bartram, B. Fiore-Gartland, and A. Crisan. Finding Their Data Voice: Practices and Challenges of Dashboard Users. *IEEE Computer Graphics and Applications*, pp. 22–36, 2022. doi: 10.1109/MCG.2021.3136545 2

[77] K. Venkitachalam and P. Busch. Tacit knowledge: Review and possible research directions. *Journal of Knowledge Management*, 16(2):357–372, Jan. 2012. doi: 10.1108/13673271211218915 5

[78] E. Wall, L. M. Blaha, L. Franklin, and A. Endert. Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics. In *Visual Analytics Science and Technology*, pp. 104–115, 2017. doi: 10.1109/VAST.2017.8585669 9

[79] Y. Wu, J. M. Hellerstein, and A. Satyanarayan. B2: Bridging Code and Interactive Visualization in Computational Notebooks. In *User Interface Software and Technology*, UIST '20, pp. 152–165, Oct. 2020. doi: 10.1145/3379337.3415851 3

[80] A. X. Zhang, M. Muller, and D. Wang. How do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Human-Computer Interaction*, 4(CSCW1):022:1–23, May 2020. doi: 10.1145/3392826 2

[81] J. Zhao, M. Glueck, P. Isenberg, F. Chevalier, and A. Khan. Supporting Handoff in Asynchronous Collaborative Sensemaking Using Knowledge-Transfer Graphs. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):340–350, 2018. doi: 10.1109/TVCG.2017.2745279 3