DESIGNING RESILIENT VISUALIZATIONS TOWARD MORE ACCURATE DATA DISCOURSE

by

Maxim Lisnic

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

Doctor of Philosophy

in

Computing

Kahlert School of Computing

The University of Utah

August 2025

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Maxim Lisnic**
has been approved by the following supervisory committee members:

Marina Kogan ,	Co-Chair	**6/26/2025** Date Approved
Alexander Lex ,	Co-Chair	**6/24/2025** Date Approved
Katherine Eve Isaacs ,	Member	**6/24/2025** Date Approved
Vineet Pandey ,	Member	**6/24/2025** Date Approved
Crystal Lee ,	Member	**6/24/2025** Date Approved

by **Mary W. Hall** , Director of the Kahlert School of Computing
and by **Darryl P. Butt** , Dean of The Graduate School.

ABSTRACT

Data visualizations are a powerful medium for communicating complex information. Although previously mostly used in professional settings or broadcasted by the government or mainstream media, visualizations now play an increasingly prominent role in public discourse. Visualizations are central to online community building and sensemaking during developing events, such as a global pandemic. Yet when shared outside expert contexts—especially on social media—visualizations are frequently misinterpreted or even used to support harmful misleading claims. This dissertation investigates how lay audiences interpret, repurpose, and critique argumentative visualizations in online settings. Through a series of empirical studies, this dissertation explores the dynamics of data discourse on social media, analyzing both how misleading visualizations are originally presented and how communities respond to them and attempt to correct them. Drawing from the results of the analysis, I design and evaluate a set of interventions aimed at making visualizations more resilient to misinterpretation. This dissertation provides new insights into the life cycle of visualizations in public spaces and offers practical guidance for designing visualizations that anticipate and account for audience reasoning.

CONTENTS

ABSTRACT . . . . . **iii**

ACKNOWLEDGMENTS . . . . . **vi**

CHAPTERS

1. INTRODUCTION **1**

1.1 A Walk Through Data Discourse 1
1.2 Why Do We Need Resilient Visualizations? 4
1.3 Dissertation Overview 6

2. BACKGROUND AND RELATED WORK **8**

2.1 The Landscape of Online Misinformation 8
2.2 Challenges in Misinformation Intervention Design 9
2.3 Misleading Visualizations 11
2.4 Data Visualizations as Vehicles for Misinformation 12
2.5 Interventions for Visualizations 13

3. MISLEADING BEYOND VISUAL TRICKS: HOW PEOPLE ACTUALLY LIE WITH CHARTS **15**

3.1 Introduction 15
3.2 Methods 17
3.3 Findings 22
3.4 Discussion 31
3.5 Limitations and Future Work 38
3.6 Conclusion 38

4. "YEAH THIS GRAPH DOESN'T SHOW THAT": ANALYSIS OF ONLINE ENGAGEMENT WITH MISLEADING DATA VISUALIZATIONS **55**

4.1 Introduction 55
4.2 Study 1: Quantitative Analysis of Engagement 58
4.3 Study 2: Thematic Analysis of Engagement 62
4.4 Discussion and Implications 71
4.5 Limitations 75
4.6 Conclusion and Future Work 75

5. VISUALIZATION GUARDRAILS: DESIGNING INTERVENTIONS AGAINST CHERRY-PICKING IN INTERACTIVE DATA EXPLORERS **82**

5.1 Introduction 83
5.2 What Is Cherry-Picking? 86
5.3 A Threat Modeling Framework for Visualizations 87

5.4 Designing Guardrails 90
5.5 Prototype Implementation 97
5.6 Study 1: Production 98
5.7 Study 2: Reaction 102
5.8 Study 3: Controlled Reaction 104
5.9 Discussion and Design Recommendations 107
5.10 Conclusion and Future Work 108

6. "I DID NOT LET THE LABEL IMPAIR MY JUDGEMENT": INEFFECTIVENESS OF FACT-CHECKING LABELS ON VISUALIZATIONS 120

6.1 Introduction 120
6.2 What Should a Fact-Checking Label Look Like? 122
6.3 Case Studies with Misinterpreted Visualizations 124
6.4 User Evaluation: Do Labels Work? 132
6.5 Discussion 140
6.6 Limitations and Future Work 142
6.7 Conclusion 142

7. DISCUSSION 150

7.1 How to Design Resilient Visualizations 150
7.2 The Future Role of Visualization 153
7.3 Data (Mis)interpretation as a Critical Practice 154
7.4 Moving Forward to Post-Critique in Visualization 156

8. CONCLUSION 160

REFERENCES 161

ACKNOWLEDGMENTS

This work is, above all, evidence of how lucky I have been to be surrounded by incredibly supportive and inspiring people who helped me along.

I am grateful to my parents, who never hesitated to invest time and resources to help me pursue my interests and get the best education they could provide.

I have been fortunate to have had excellent mentors and collaborators who pushed me to keep going, especially when things felt most difficult.

I have also had the luck of working alongside generous, thoughtful labmates and peers who made this journey more enjoyable, and who I am confident will remain lifelong friends.

Most of all, I am grateful to my partner. Lynn, your support—in life and in this work—has meant everything. You have been by my side every single day: from the moment I voiced my wild idea to apply to graduate school to me drafting this acknowledgments section right now. I truly cannot thank you enough.

CHAPTER 1

INTRODUCTION

Data visualizations have become a popular way to share and interpret information. Charts, tables, and maps are no longer just tools for experts; they belong to the masses. Data are widely circulated in the media and play a central role in how people engage with important issues. Many of today's most debated topics—climate change, disease spread, economic policy, housing supply, social inequality—are deeply tied to numbers, and those numbers come to life through visual representations.

Social media, in particular, further encourages the sharing of visualizations. With virtually every platform either being predominantly visual or allowing users to attach a highly prominent image to a post, data visualizations are not just easy to share—they often form the main idea of a post or provide support for its accompanying text. Although platform feed-ranking algorithms are proprietary, there is evidence that visual content garners more attention [1], and it is likely that posts with visual content are prioritized over, for instance, external links that lead users off the platform. Consequently, data visualizations have become an important part of online collective sensemaking and community building.

Given their growing influence, it is important to recognize that data visualizations can also lead to harm or misinterpretation. When visualizations are widely shared and interpreted outside of expert contexts, as our examples below will show, they can shape public understanding in misleading or oversimplified ways. This makes it crucial to study how visualizations are interpreted, what kinds of misunderstandings they invite, and how we may adjust how we design visualizations, not just for clarity or accuracy, but with greater attention to how they will be read, shared, and used in real-world settings.

1.1 A Walk Through Data Discourse

One reason for the rich discourse around data visualizations online is simple: they are fun. Part of what makes visualizations so prominent in public discussion is that they are

enjoyable to create and to look at. Well-designed charts are eye-catching and can even become works of graphic art. For some, making visualizations is also a form of creative expression and many online communities are centered on sharing and discussing original infographics. There is even a competitive side: Tableau Community's annual IronViz competition challenges participants to build a dashboard in under 20 minutes, with entries judged by a panel of experts.

Not only experts judge charts—chart critique is another exciting topic that garners a lot of engagement online. Chart critique may range from suggesting minor chart construction improvements to simply finding and laughing at bad charts. Online communities on Reddit such as r/dataisugly offer a space specifically for critiquing poorly designed visuals. But even on r/dataisbeautiful, a subreddit primarily intended to celebrate well-crafted charts, discussions often turn into primarily critiques of design choices rather than insights. Many charts go viral with their engagement driven exclusively by the discussion around their bad design (↗). Numerous mainstream media articles and blog posts are written about design choices. All in all, people seem to love critiquing charts and finding examples of bad design.

But aesthetics and critique aside, people also do pay attention to what the data actually show. Visualizations help us make sense of the world by making the otherwise invisible things seen. Things like climate change may be (so far) imperceptible, yet deviations in temperature plotted on a chart may clearly show the trend. Similarly, during the COVID-19 pandemic and the associated lockdowns, data visualizations offered us a window into the invisible risks we were facing by aggregating cases worldwide, tracking hospital capacity, and conducting wastewater surveillance—things unnoticeable to the naked eye.

In these examples, the data analysis and presentation are most commonly done by professionals and government agencies, and the charts are primarily used as a narrative and to spread awareness. But the public conducts their original analyses, too. With democratized access to downloadable data and interactive data exploration platforms, online users now have tools to uncover and share evidence of phenomena and developing situations themselves. Data visualizations play an important role in revealing new insights, be it uncovering hidden trends, exposing wrongdoing, or making complex issues more understandable. Although previously data analysis for popular consumption has

primarily been the realm of data journalism, social media has enabled the public not only to wrangle data independently but also to share it with others in the manner of bona fide data journalists. As with much of the information online, these insights may be inaccurate, overlook the depth of analysis needed to support complex claims, or ignore crucial domain knowledge, sometimes even contributing to misinformation. However, they can also raise awareness of important issues.

In one example in 2023, ultrarunner Joasia Zakrzewski finished third at a 50-mile ultramarathon between Manchester and London. Shortly after the finish, however, online users turned to the runner's GPS data that she had automatically posted on Strava, a sports social media app (↗). The runner's data logs revealed surprising things: she seemed to have completed miles 38 and 39 of the race at a pace of 2 minutes per mile each, which is faster than the world record pace for a mile. This revelation prompted an investigation, and the runner finally confessed to having taken a ride in a car from a friend for that portion of the race, explaining the 30 miles per hour pace. Ms. Zakrzewski was then stripped of her medal [2].

Amateur data sleuthing does not always have a straightforward resolution yet is often central to a community's sensemaking of a developing event. In another example that rocked the chess world in 2022, world champion Magnus Carlsen surprisingly lost to a much less experienced opponent, Hans Niemann. Following the loss, Carlsen publicly accused Niemann of cheating and demanded an investigation. While the official investigation by FIDE—the main chess governing body—took time and was conducted in private, social media users could not and did not wait. For a while, the r/chess subreddit became chock full of charts and statistical analyses estimating the probability of cheating. Users compared the moves in Niemann's and Carlsen's games to the chess engine predictions to find evidence of high correlation (↗, ↗,↗). Other users analyzed Niemann's past games to come up with alternative calculations for his rating to reevaluate whether his win was surprising (↗), with the rest of the community auditing and re-analyzing each other (↗). Eventually, the lawsuits about the cheating allegations and defamation were settled without conclusive evidence shared with the public. In this example, community data discussions did not conclusively prove wrongdoing but fueled rumors and engagement for months.

Beyond rumors and amiguity, misguided data analysis in the public can also be misleading and outright harmful. One of the most notable examples of data misinformation originated well before social media: the false connection between the MMR vaccine and autism by all accounts emerged from a published scientific study in 1998 and gained extremely wide public attention [3]. The study likely did not fabricate the data, but the analysis was based on correlations on a small sample, did not include a control group, and made unsubstantiated claims in its summary and conclusion. The research world has a way to deal with such data fraud: the study was retracted, and the principal investigator lost his medical license. However, this misinformation not only persists but is supercharged by social media spread and continues to fuel vaccine hesitancy today. Online posts in support of vaccine skepticism use spurious data, such as the correlation between autism rates and the use of a certain chemical (↗), or link unrelated studies on vaccine effectiveness in rats (↗). This example highlights a key problem: data can be misused to support falsehoods, even when the underlying data are real. The data may be not fabricated yet still manipulated in ways that make them misleading or harmful.

1.2 Why Do We Need Resilient Visualizations?

The examples above show the spectrum of outcomes that can result from discussions around data. Data often yield undeniably useful insights and can serve as a focal point for online community-building; however, they are also not infallible. Even using factual and rigorously collected data is not enough as they can also lead to harmful conclusions. Although falsification does occur—academic studies have been retracted for relying on fabricated data, and researchers are developing better tools to detect it [4]—many cases of misinterpretation revolve around factual data and instead stem from spurious patterns and statistical fallacies. Conducting rigorous data analysis is difficult, and well-intentioned individuals may lack the expertise to recognize errors, amplifying non-causal correlations, or misapplying statistical methods. Human biases also play a role, as even trained analysts can overlook key caveats or fall prey to cognitive biases. Of course, misinterpretation can also be intentional. One of the important pillars of propaganda and misinformation campaigns is maintaining a connection to reality; thus, in the context of data-driven information, they often still rely on factual data.

Misinformation is not unique to visualizations or the internet era; throughout history, people have been exposed to harmful rumors and propaganda. However, rumors based on data more broadly and data visualizations specifically present unique challenges. Data visualization is the predominant medium by which the public consumes data. In a world where data and data-driven decisions are often perceived as objective and factual, distinguishing between rumors and scientific facts becomes more difficult. Additionally, data-supported misinformation is typically not completely false; as in all of the examples above, people typically base their insight on real data and thus any misinformation is still presented as a half-truth. This connection to reality helps misleading visualizations be not only more persuasive but also more resistant to common fact-checking efforts and thus harder to refute. Information operation campaigns commonly rely on unwitting agents to produce the initial misleading arguments and eventually spread the misinformation themselves [5]. As discussed above, data visualizations are central to many online communities, and collective data sleuthing produces lots of often spurious insights. Taken together, controlling and amplifying online data narratives is a potentially powerful tool for shaping public perception.

Consequently, visualization research should pay special attention to the way visualizations are shared and interpreted in public spaces such as social media. In research environments, visualizations are typically designed to help professionals and academic collaborators work with complex datasets in good faith. However, the life cycle of data visualizations and the conversations around them differ significantly between professional and public settings. When shared in the public sphere, visualizations take on a second life, they are being reused and reinterpreted. The general public is often less familiar with the specifics of data collection and processing, which is understandable, as most viewers did not collect the data or create the visualization they are sharing. Moreover, even when data collection and transformation procedures are thoroughly documented, this information may be not readily accessible to the general audience, or seem trustworthy enough to believe that the data has been interpreted in their best interest. More importantly, narratives around data and visualization in the public domain are influenced by factors such as political partisanship, social dynamics, and the limited time or capacity audiences have to engage with complex analyses. Ultimately, while visualizations are often created by and

for experts, it is now undeniable that they are consumed by broader audiences in a much more complex ecosystem where their meaning can shift significantly.

1.3 Dissertation Overview

This dissertation aims to deepen our understanding of how laypeople create, share, and consume visualizations that support arguments—whether misleading or not—and the resulting implications for visualization design. We examine visualizations shared on social media to investigate how they circulate and are interpreted once they move beyond professional and scientific contexts and into the public sphere. Throughout the chapters, we aim to tackle the fundamental challenges: How do people interpret data visualizations, and how do others react to it? What kinds of logical errors and corrections emerge in public discussions of data, and how persuasive are they? More importantly, is there potential to help crowds distinguish between sound data analysis and misleading claims? And can we learn general lessons about how to present data to encourage accurate interpretation while discouraging misuse?

To address these questions, we first seek to better understand the *data discourse*, and then use our findings to inform and test interventions aimed at making visualizations *resilient to misinterpretation*. Chapter 3—corresponding to a paper published in CHI 2023 [6]—examines the types of charts and logical arguments that circulated on social media during the COVID-19 pandemic. Our findings show that the majority of COVID-skeptic charts were not only well-constructed, but also reused from government reports or popular data explorer websites. Charts intentionally constructed to be misleading are a minority, and instead misinformation actors misuse existing visualizations to spread misinformation, using tactics like cherry-picking data and adding misleading annotations to reputable charts.

Next, Chapter 4—corresponding to a paper published at CHI 2024 [7]—analyzes how audiences engaged with these visualizations. We examine how visualizations spread, how audiences react to them, and what patterns emerge in the ways they were interpreted and debated. We find that COVID-skeptic visualizations attracted significantly larger discussions, in which the audience attempted to make sense of or debunk the premise of the original post. We describe the common types of responses—such as sharing counter-data

or drawing on relevant anecdotal experience—and find that, ultimately, data-driven misinformation is difficult to fully debunk but it successfully attracts increased engagement and discussions.

Building on these insights, the following chapters explore two strategies for making static and interactive data visualizations more resilient to misinterpretation.

We designed our interventions to reduce misinterpretations primarily when charts are shared on social media. To account for the possibility that charts may be shared as screenshots, we focused on visual and textual static augmentations that would remain intact even when the chart is viewed out of context.

First, Chapter 5—corresponding to a paper published at CHI 2025 [8]—focuses on using visualization design to safeguard data exploration platforms, such as OurWorldInData. We introduce a design space for guardrails, or visual interventions against cherry-picking, one of the common reasoning errors identified in Chapter 3. We then evaluate their effectiveness in countering both the creation of and the reaction to cherry-picked data. We find that guardrails successfully encourage skepticism, although they need to be simple and close to the original visual encoding to be noticed and understood.

Continuing our intervention work, in Chapter 6 we examine a different approach to safeguarding charts. To tackle the misinterpretation of static charts, we develop textual fact-checking labels and evaluate their impact on misinterpretations. Ultimately, we find that labels alerting viewers of possible misinterpretations are ineffective at preventing them. Our qualitative analysis reveals the reasons for this ineffectiveness, including the perceived neutrality and accuracy of data compared to the perceived biasedness of text.

Lastly, we present a discussion of our findings in Chapter 7, before offering a conclusion in Chapter 8.

CHAPTER 2

BACKGROUND AND RELATED WORK

In this chapter, we provide the background to online misinformation research in general as well as the previous work in data visualization that is related to misinterpretations, biases, and discussions. Throughout this chapter, we outline the gap in research on data-driven misinformation and discussion thereof, thus motivating the work presented in this dissertation.

2.1 The Landscape of Online Misinformation

Social media platforms enable rapid communication to broad audiences, allowing important information to spread almost instantly. However, this speed also creates opportunities (whether intentional or not) for the dissemination of inaccurate, misleading, or even harmful content. Conventional wisdom suggests that with enough users, the wisdom of crowds [9] would prevail, correcting misinformation as it spreads. This idea had originally been backed up by early research, with studies showing that online communities do identify and challenge rumors and false claims [10], [11]. Yet, more recent work begins to highlight the limitations of self-correction in online spaces, revealing how bad actors can exploit these weaknesses to spread misinformation [5], [12].

Although misinformation can gain traction through strategic information operations, its initial narratives often emerge organically as people attempt to make sense of unfolding events [5]. This collective sense-making is particularly common in situations characterized by high uncertainty and limited reliable information, such as man-made crises [10], [13], [14], natural disasters [15]–[17], and public health emergencies [18], [19].

In public health crises, such as viral disease outbreaks, people frequently turn to social media to discuss potential health risks [18]. Collective sense-making online became especially pronounced during the COVID-19 pandemic. Research by Pine et al. shows that individuals struggled to independently assess risks due to inconsistencies, gaps, and the

overwhelming scale of available data [19]. As a result, many relied on anecdotal evidence, which is often shaped by personal biases. This dissertation builds on these findings by examining how people engage with COVID-19 data on social media and identifying the ways in which data inconsistencies, cognitive biases, and other factors contribute to the spread of misinformation.

2.2 Challenges in Misinformation Intervention Design

Given the role of data inconsistencies and personal biases in the spread of misinformation, addressing misleading content requires carefully designed interventions. Researchers and social media platforms have explored a range of strategies to curb misinformation at scale, each with varying degrees of effectiveness. Aghajari et al.[20] categorize existing interventions into four broad types: content-, source-, user-, and community-oriented approaches. Among these, content-based interventions—those focused on evaluating the veracity or credibility of a post—have been the most widely implemented. Social media platforms such as Facebook and Twitter have adopted content-based strategies that involve removing, deprioritizing, or labeling content based on assessments by expert fact-checkers or algorithms[20]. Because this dissertation focuses on the content of social media posts, we primarily examine misinformation interventions through the lens of content-based approaches.

Fact-checking labels, one of the most common content-based interventions, have become a key tool in combating misinformation [21]. Both expert- and crowd-sourced fact checks reports have shown promise, effectively shifting beliefs [22], reducing the sharing of false information [23], and even lowering the willingness to pay for unproven treatments [24].

Despite these successes, research on the efficacy of fact-checking presents mixed findings [25]. The effectiveness of these interventions varies based on content, domain, and audience. Although corrections often succeed in changing beliefs, fact-checking can sometimes have unintended effects, such as reinforcing false beliefs (backfire effect) [26] or increasing hostility in discussions [27]. The implied truth effect further complicates matters, as it can lead users to assume that content lacking a fact-check label is inherently

accurate [28]. The source of the label also plays a crucial role: research suggests that people are more likely to trust fact-checks from a friend [29] or a familiar news outlet [30].

Political affiliation further influences the impact of fact-checking interventions. Studies have found that fact-checking labels are less effective for Republican voters [31], potentially due to the boomerang effect, where corrections reinforce pre-existing beliefs [32], or perceptions of bias in the labeling process [33]. These challenges are particularly relevant to crowdsourced fact-checking, which, while promising for scalability [34], faces obstacles such as users being less likely to fact-check those with aligned political views [35].

A key reason for these varied outcomes is the diversity of misinformation itself—misleading statements take many forms, and researchers are still working to classify and understand them [36]. The format of misinformation also influences the effectiveness of fact-checking labels. Research suggests that labels are less effective when misinformation is presented as a meme [37], and audiences often fail to recognize misleading native advertising even when explicitly labeled [38]. This suggests that fact-checking labels may be less effective for content that is not perceived as a straightforward statement or news story.

Even within news and statements, certain types of misinformation are more resistant to fact-checking. Walter and Salovich [36] find that audiences struggle to distinguish between factual reporting and opinion pieces, affecting the impact of fact-checking interventions. Other work highlights the difficulty of assessing content that falls into a "gray area" of misleading but factually accurate statements—such as opinions, misinterpretations of data, or satire [39], [40].

As this dissertation explores in greater detail, data-driven misinformation often falls into this gray area. Because it is based on real data but framed with potentially misleading interpretations, it requires distinct intervention strategies. Understanding how audiences engage with factual but misleading content presents an opportunity to refine misinformation interventions. Throughout this dissertation, we argue that data-driven misinformation is a distinct category that demands special consideration in intervention design.

2.3 Misleading Visualizations

We begin our exploration of data-driven misinformation by first discussing misleading visualizations more broadly. Early influential works on deceptive charts, such as Darrell Huff's *How to Lie with Statistics* [41] and Edward Tufte's *The Visual Display of Quantitative Information* [42], have shaped research and public discourse on misleading visualizations for decades [43]–[48]. Tufte [42] introduced the concepts of *graphical integrity* and the *lie factor*, emphasizing that the visual encoding of data should accurately reflect the magnitude of underlying values. Subsequent studies have demonstrated that common distortions—such as truncated, inverted, or rescaled axes—can influence perception [47]. However, more recent work suggests that the magnitude of these effects is limited and varies depending on context and the viewer's data literacy [49].

The visual distortions described by the lie factor have long been the dominant way both researchers and the general public conceptualize misleading visualizations. The annual "VisLies" (`www.vislies.org`) event at the IEEE VIS conference, for instance, formed a space for the community to gather, share, and discuss visual distortions, data errors, and other egregious plotting mistakes. The broader public perception's aligns with this view as well—Lo et al. [50] analyzed 1,143 charts labeled as misleading on social media and search engines, finding that the overwhelming majority exhibited visual errors, inappropriate encodings, or outright distortions. Only 7% of the collected charts were "faithfully plotted" but misleading due to suggestive framing, such as implying causation or omitting key context.

Yet, misleading visualizations are not always the result of intentional distortion. Charts can also mislead by playing into viewers' cognitive biases. Several frameworks have been proposed to examine how biases influence the interpretation of visualizations [51], [52], though research in this area remains limited. A large-scale literature review by Dimara et al. [51] found that most existing work focuses on biases affecting numerical estimation rather than higher-level cognitive processes. Recent discussions in the visualization community advocate for shifting beyond mere perceptual accuracy to broader considerations of how visualizations shape reasoning and decision-making [53]–[55]. However, key visualization tasks prone to bias—such as hypothesis assessment, causal attribution, and opinion formation—have yet to be systematically studied [51].

2.4 Data Visualizations as Vehicles for Misinformation

While misleading visualizations often result from poor design choices or misinterpretation, data visualizations have also been deliberately used to introduce or reinforce misinformation. Beyond merely distorting data representation, visualizations can serve as persuasive tools in misinformation campaigns, lending an air of credibility to misleading arguments.

Lee et al. [56] analyze COVID-skeptic visualizations shared on social media and find that many of these charts are well-designed, suggesting that misinformation is often driven by epistemological divides rather than a lack of visual literacy. Notably, both pro-mask and anti-mask communities used the same visualizations to support opposing claims, demonstrating that misleadingness is not always an inherent property of a visualization but can instead be shaped by viewer interpretation. These findings represent an important shift in research and highlight a previous blind spot in research on ways visualizations can mislead.

Although this is a novel way of looking at the problem of misleading visualizations, in retrospect, it does make sense in the context of many of the prior findings on the way people read charts. For instance, we know that several factors contribute to people's divergent readings: including the social context in which a viewer encounters a chart [57], biased framing and titles [58], individual cognitive differences [59], personal biases [60], and the *curse of knowledge*—the mistaken assumption that others interpret a visualization the same way as the creator or other viewers do [61].

Despite the growing body of work on misleading visualizations, research on misinformation supported through data visualization remains limited compared to textual misinformation. Recently, however, misinformation scholars outside of the visualization community have also begun to examine how visualizations contribute to misinformation ecosystems and how they fit into broader concerns about online information integrity. Weikmann and Lecheler [62] argue that visual disinformation constitutes a distinct category of falsehood, differing from textual misinformation in its greater modal richness and manipulative sophistication, making it particularly persuasive. Similarly, Matthew Hannah [63] presents a case study of QAnon conspiracy theories, contending that the

movement's success relies on its ability to leverage information visualizations and pattern-seeking as persuasive tools. He describes this as symptomatic of the broader challenge of combating misinformation that mimics the language and methods of data analysis.

2.5 Interventions for Visualizations

Similar to the broader research on conceptualizing visualization misinformation, research on visualization interventions for misinformation is particularly sparse. However, numerous related studies have explored techniques for more broadly mitigating biases in visual analytics and data interpretation. Much of this work has focused on professional and scientific settings, addressing challenges such as the forking paths problem [64], multiple comparisons [65], inaccurate model specifications [66], and imbalances in how variables are represented [67]. Many proposed technical solutions aim to encourage good-faith analysis, such as scoring the biasedness of exploratory decisions [67], detecting cherry-picking in large datasets [68]–[70], or visualizing an analyst's process [71]. Although these approaches are relevant, in the context of online misinformation, interventions cannot rely on analysts' self-regulation; instead, they must address how visualizations are perceived and used by broader audiences.

Beyond professional analytics, visualization interventions targeting audience-level biases and misinterpretations have been receiving increasing attention. Researchers have explored textual warnings against causal assumptions [72], multiple coordinated views to counteract misleading patterns [73], interactive linking between text and data [74], and design alternatives to highlight truncated vertical axes [75]. Although these interventions aim to address specific pitfalls, they do not fully account for the broader challenge of reasoning errors—how viewers integrate and interpret visual information. Zhang et al. [76] document concerns among COVID-19 dashboard designers about how their visualizations could be misinterpreted or misused, underscoring the difficulty of ensuring that data is read as intended.

One response to this challenge is to incorporate additional textual guidance into visualizations. Burns et al. [77] found that adding descriptive metadata improved perceptions of thoroughness, though it did not necessarily enhance interpretation accuracy. Similarly, textual prompts designed to encourage critical thinking have been studied in other domains,

such as AI-generated content [78], offering insight into how supplementary information might help viewers approach visualizations more thoughtfully. However, directly steering viewers toward a "correct" interpretation remains a complex task. Ge et al. [79] argue that addressing reasoning errors requires first clearly defining a visualization's primary intent. Yet, as discussed previously, even with a well-articulated goal, viewers' takeaways are shaped by framing [58], prior knowledge [61], storytelling techniques [80], individual cognitive differences [81], and personal biases [82]. The inherent variability in audience interpretation suggests that while designers can influence perception, they may not be able to fully dictate how a visualization will be understood post-hoc.

An alternative to navigating these challenges would be to design visualizations with misinterpretations in mind. An ante-hoc approach would depend on accurately anticipating what can go wrong and using narrative visualization techniques [83]. to design the chart accordingly. Adar and Lee-Robbins [84] propose an approach that explicitly defines learning objectives in visualization design. By clarifying what authors want audiences to take away from a visualization—and, crucially, identifying *anti-objectives* (interpretations to be avoided)—designers can better anticipate potential misreadings. This perspective is particularly relevant in the context of misinformation, where preemptively countering likely misinterpretations can help mitigate misuse. However, it may still often be impossible to foresee every potential misreading, making post-hoc interventions that clarify authorial intent remain an important area of study. The latter half of this dissertation builds on this idea, examining strategies for designing and implementing such interventions.

CHAPTER 3

MISLEADING BEYOND VISUAL TRICKS: HOW PEOPLE ACTUALLY LIE WITH CHARTS

Data visualizations can empower an audience to make informed decisions. At the same time, deceptive representations of data can lead to inaccurate interpretations while still providing an illusion of data-driven insights. Existing research on misleading visualizations primarily focuses on examples of charts and techniques previously reported to be deceptive. These approaches do not necessarily describe how charts mislead the general population in practice. We instead present an analysis of data visualizations found in a real-world discourse of a significant global event—Twitter posts with visualizations related to the COVID-19 pandemic. Our work shows that, contrary to conventional wisdom, violations of visualization design guidelines are not the dominant way people mislead with charts. Specifically, they do not disproportionately lead to reasoning errors in posters' arguments. Through a series of examples, we present common reasoning errors and discuss how even faithfully plotted data visualizations can be used to support misinformation.

This content of this chapter is drawn directly from M. Lisnic, C. Polychronis, A. Lex, and M. Kogan, "Misleading beyond visual tricks: How people actually lie with charts," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2023, pp. 1–21. doi: 10.1145/3544548.3580910 [6]. It is included in the dissertation with the permission of my coauthors, Cole Polychronis, Alexander Lex, and Marina Kogan.

3.1 Introduction

Data visualizations created for the general public can help explain and summarize complex phenomena and otherwise incomprehensible amounts of data. On social media sites, users actively share and comment on data visualization posts as a form of collective sense-making. The insights gained from these visualizations, however, can mislead people

by perpetuating misconceptions and misinformation.

The visualization community has been primarily defining *misleading visualizations* as charts that interfere with the viewer's ability to accurately read off and compare values. The terms "deceptive", "misleading", "lying" are typically used to describe visualizations with visual tricks, such as truncated or inverted axes, or the violation of visualization guidelines and best practices, such as the use of unjustified 3D or problematic color maps [47], [48], [85]. This type of visual deception is rooted in the gap between the true value of data points used as input for the chart and the different values perceived by the viewer. This discrepancy has been called the "Lie Factor" [42].

The visualization community has studied these deceptive design techniques in detail [50], and "VisLies" (`www.vislies.org`), a long-running satellite event of the IEEE VIS, showcases some of the worst examples. Consequently, conventional wisdom may suggest that visualizations that are used to spread misinformation online would predominantly be ones that employ such visual tricks, and may imply that the solution may lie in promoting visual literacy, so that the general public can spot these deceptive techniques. More recent research, however, demonstrates that, in many cases, people who mistrust the scientific establishment propose data-driven arguments in support of their ideas and use what Lee et al. describe as *counter-visualizations*: visualizations using orthodox methods to make unorthodox arguments [56].

The question of what makes a visualization misleading or deceptive and how this can be fixed, however, remains. In an attempt to provide an answer, we collected, categorized, and analyzed 9,958 posts shared on Twitter that contain data visualizations related to the COVID-19 pandemic. This chapter shows that common design violations in visualizations on social media are rare and occur at about the same rate in conspiracy- and misinformation-supporting posts as in neutral posts. We find that instead of using visualization design violations, actors who want to misinform draw attention to unexplained salient features of well-designed charts and assign meaning to them. We introduce the notion of *vulnerable visualizations*: visualizations created from accurate data with no clear intention to misinform but susceptible to supporting misinformation by not visualizing important context or not anticipating a biased reading. Through biased framing and annotations, vulnerable visualizations designed by authoritative sources can be misused

to create a type of counter-visualizations described by Lee et al. [56].

This chapter makes several contributions. Firstly, we analyze the prevalence of visualization misinformation techniques in a large-scale real-world data set. To our knowledge, this is the first work to describe misleading visualizations among a sample of charts on a given topic that a general audience may see online, as opposed to a sample of examples previously identified as deceptive. Secondly, we introduce a typology of attributes of visual misinformation on social media. We illustrate the typology by presenting examples of specific instances of visualization-supported misinformation. Thirdly, we provide a theoretical framework for understanding how data visualizations can be used to reinforce misinformation arguments through the lens of inductive reasoning. Lastly, based on our findings, we propose design recommendations to safeguard charts and prevent their misinterpretation.

3.2 Methods

In order to explore the ways in which individuals might deceive others using visualizations, we collected and processed social media posts related to the COVID-19 pandemic from Twitter, as illustrated in Figure 3.1 and described in the sections below.

We consider this data set to be important and appropriate for our analysis for several reasons. Firstly, it is a large-scale data set of social media posts that is collected in an unbiased way with respect to the deceptiveness of visualizations and reflects the online visualization discourse as is. Secondly, the topic of the posts is an important prolonged crisis that has drawn visualizations from novices and experts alike: the COVID-19 pandemic.

3.2.1 Data Collection

We collected data generated during the pandemic on Twitter using the official Twitter COVID-19 streaming endpoint. This endpoint provides the full collection of tweets containing at least one of the 585 hashtags and keywords[1] that Twitter internally identified as being relevant to conversations on their platform related to the pandemic. We started streaming data from this endpoint on May 15, 2020 and stopped streaming on September 6, 2021. In that time, we collected 2.2 billion tweets related to the pandemic, totaling

[1]Full list of filter terms available on the API documentation.

22 terabytes of data (see Figure 3.1a). We then filtered this full stream down to only English language tweets that contained images (although not explicitly visualizations) and removed any retweets or replies. From this set of tweets that potentially contain visualizations and meet the rest of our collection and scoping criteria, we extracted two data sets. First, we extracted the full first week of data May 15–22, 2020 (19,214 tweets), which we call our train-test data set. This data set was used to train machine learning models to be used for data processing (described in the following subsection). The first two authors manually labeled whether or not tweets in the train-test data set contained visualizations, resulting in a data set that had 8,170 positive examples (tweets that did share visualizations) and 11,044 negative examples (tweets that did not share visualizations). Second, we randomly sampled from the remaining data to create a data set of 162.8 million tweets (Figure 3.1b). It is this data set that we applied our data processing models to in order to arrive at our final data set, used to answer our research questions.

3.2.2 Data Processing

Retrieving the full set of over 162 million images from Twitter's photo storage system Blobstore [86] would take weeks of clock time. In order to scalably detect visualizations from our random sample of the data (Figure 3.1b), we applied two machine learning models, described in greater detail in Sections 3.2.2.1 and 3.2.2.2. First, we predicted if a tweet is likely to contain a visualization based on the content of its message. If this model predicted that the tweet has likely shared a visualization, we then retrieved that image from the Blobstore and applied a visualization detection model.

3.2.2.1 Predicting the Presence of a Visualization

In order to predict whether a tweet is likely to share a visualization, we trained a Multinomial Naive-Bayes classifier (scikit-learn v0.24.1) on the train-test data set. Specifically, we embedded the text of tweets using a bag-of-words approach and supplemented it with metadata features: the number of hashtags, emojis, user mentions, and capitalized words. This model had an accuracy of 0.75 on the train-test data (precision = 0.70, recall = 0.71). To better understand this accuracy, we also labeled all of the visualizations in our train-test data set to explore whether this model was more likely to miss certain kinds of visualizations. We found that common visualization types (point and bar charts, tables,

maps, etc.) are missed about 4.29% of the time. Rarer visualization types, such as isotype charts or network diagrams, are missed about 14.30% of the time. This suggests that our model performs better at detecting simple and more common charts, but this bias is negligible given our large sample size.

Applying this model as a filtering step reduced the number of tweets for which we had to retrieve images by about 80%, producing the data set shown in Figure 3.1c. We retrieved the image files between September 6, 2021 and September 8, 2021. Consequently, although some tweets may have since been removed or their authors may have been suspended, all posts in our analysis were public as of September 2021 and spent sufficient time on the platform to potentially be widely circulated.

3.2.2.2 Classifying Images as Visualizations

In order to determine whether or not images were visualizations, we used the Inception V3 Model, developed at Google [87] and pre-trained on ImageNet [88]. We used a transfer learning approach and fine-tuned this model with our train-test data set, resulting in an accuracy of 0.73 (precision = 0.71, recall = 0.65). Applying this model to the data set shown in Figure 3.1c and filtering any tweets that were not identified as sharing a visualization left us with a data set of 80,400 tweets that shared some kind of visualization (Figure 3.1d). We then qualitatively coded a subset of these as described in the section below.

3.2.3 Qualitative Coding

To arrive at our final corpus of tweets with visualizations, we first took a random sample (33,808 tweets) to reduce the amount of data to be manually coded, and then conducted multiple stages of qualitative coding. In the first stage, we manually examined the outputs of automated data processing to eliminate any irrelevant posts for future annotation. Two of the authors, with the help of a (paid) undergraduate student, reviewed the tweets to remove (1) posts that were incorrectly labeled as visualizations, which primarily consisted of images of text and diagrams, and (2) posts that contained visualizations that do not pertain to the COVID-19 pandemic, for instance charts showing stock performance without mentioning the pandemic. We held multiple discussions during this process to decide how to deal with any edge cases. This process yielded 9,958 posts that contained relevant visualizations for the next stage.

In the second stage, we developed a codebook and used it to annotate the 9,958 posts, as well as created a closely related typology derived from the codebook. First, the first author coded a sample of 400 visualization tweets through an open-coding process [89], and developed the initial set of codes. The codebook includes the following categories of codes: (a) source information, or whether the visualization was a screenshot of a chart from the government or a media outlet or an interactive dashboard, or was created by an unknown author, (b) the presence and opinion polarity of the tweet text, the chart title, native annotations (annotations added to chart by the visualization creator), and added annotations (annotations added to a screenshot of a chart by a third party), (c) any existing violations of common visualization guidelines, as well as whether the choice of visual encoding is appropriate for a given type of data, and whether it is generally possible to understand the chart given the available information, and (d) commonly occurring types of reasoning errors. In order to refine the codes, particularly those pertaining to the reasoning error category, two authors applied the codebook to a sample of 200 posts independently and held iterative discussions. At the end of this process, we achieved thematic saturation and no additional issues were identified. The finalized codebook, shown in Figure 3.2, was used by the first author to annotate the entire sample of visualizations. Then, we grouped related codes based on the common features to create a typology of visualization post attributes. We include the results of the annotation stage in the supplemental materials. In accordance with Twitter API Terms of Service[2], we provide tweet IDs that can be rehydrated to retrieve the contents of the post, as long as the post is still publicly accessible on Twitter.

3.2.4 Identifying Misleading Visualizations

In an attempt to present an unbiased analysis rather than one that reflects the authors' personal opinions about the deceptiveness of specific charts, we used the tweet text to identify deceptive charts by applying the following definition of misleading visualizations: visualizations that the public uses as the basis of opinionated conclusions that are entirely incorrect or contain significant reasoning errors. There are several reasons to this definition. Firstly, operationalizing deceptiveness in this way allows to provide actionable steps

[2] `https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases`

to combat deceptiveness by identifying specific reasoning errors. Secondly, we found that the decision of whether a certain visualization is misleading or not is *highly personal*, as it represents a judgement of whether a chart could theoretically deceive someone.

To test this hypothesis empirically, four annotators—two authors and two students affiliated with our academic institution—independently coded a sample of 400 posts, labeling them on a 5-point Likert scale [90] ranging from "not at all deceptive" to "extremely deceptive". To evaluate the inter-rater reliability, we calculated Krippendorff's α score [91] which had a value of 0.243. Converting our 5-point Likert scale results to a binary value of "not at all deceptive" and "deceptive" yielded an α of 0.351. Krippendorff recommends discarding results with $\alpha < 0.667$ [91], which indicates that our results had a very low degree of agreement. We held follow-up discussions and reviewed examples of disagreement and found that different annotators focus on different features of the chart, are biased by personal beliefs surrounding the pandemic, and have disagreements about the significance of common visualization design guidelines. The results of our experiment show that it is difficult to predict whether a visualization can mislead its audience. Instead, we used social media traces—such as tweet text—as evidence that a chart has been misinterpreted by the viewer.

Specifically, by applying the *text polarity* codes (shown in Figure 3.2), we can distinguish between COVID-19 data visualization posts on Twitter that are neutral—such as status updates and factual observations—and posts in which the author uses the visualization as the primary basis of support for a certain argument. For the purposes of this chapter, we refer to the latter group as *opinion tweets*. Opinion tweets often offer biased and incomplete interpretations of attached data visualizations. Therefore, we consider opinion tweets that contain reasoning issues in the way the author interprets the visualization (identified using the *reasoning error* codes shown in Figure 3.2) to be *misleading tweets*.

Although this approach allows us to identify visualizations that either mislead the tweet author or were used by the author in an attempt to mislead others, we cannot state that rest of the visualizations in our data set are unambiguously and universally *not* misleading. In particular, many visualizations that violate design guidelines still have the potential to deceive their viewers, as shown by prior research [47], even if we do not observe tweet text that misinterprets it. Figure 3.3 illustrates several examples of tweets

that share *potentially misleading visualizations*.

3.2.5 Ethical Considerations in Social Media Research

It is important to acknowledge the ethical implications of collecting, analyzing, and sharing social media posts in our work. Tweet authors—even those with public accounts—may not be aware of their data being collected and highlighted for research purposes. At the same time, we want to offer the reader transparency into our research methods by showing compelling examples of types of visualization posts and by providing the opportunity to further explore the discussion around them. As a solution, we present anonymized examples of posts appearing throughout the chapter but also offer the ability to follow the link and explore the tweet as it appears on the platform in its original context.

We consider it valuable also to include tweets that have since been deleted in our analysis. Many such tweets had been circulating on the platform for months before being removed, highlighting the importance of studying potential platform interventions and visualization design considerations to prevent the spread of misinformation. In an attempt to preserve the authors' right to remove their content, however, we only offer tweet IDs in our supplemental materials instead of the complete tweet data. Therefore, tweets deleted by the time of publication or removed later will not be available for future data collection.

3.3 Findings

Our annotated data enables us to offer descriptive statistics of the data set as well as a typology of attributes of tweets with misleading visualizations. In this section, we present our findings that are derived from analyzing the statistics, patterns, and specific examples of tweets after applying the codebook shown in Figure 3.2. An interactive browser of our data set and descriptive statistics is also available at `https://hashtag-misleading.netlify.app/`.

3.3.1 Quantitative Overview

In this section we present a quantitative overview of the data set in general and misleading tweets in particular, based on our qualitative coding. Figure 3.2 presents a breakdown of all codes in the data set.

3.3.1.1 Source of Visualization

Data visualizations in tweets come from a variety of sources, with the majority (55.3%) being screenshots of existing charts from media outlets, government websites and presentations, and data exploration websites. Data exploration websites—such as OurWorldInData [92] and Worldometer—and interactive dashboards that accompany COVID-related search results on Google offer interactive visualization dashboards that allow users to selectively plot variables and data points. The remaining 45% of tweets include data visualizations for which we were unable to unambiguously identify sources.

3.3.1.2 Text Polarity

In terms of the polarity of observations, 62% of all opinion tweets oppose conventional measures such as masks, lockdowns, and vaccination, or deny the existence or severity of the crisis. 34% of opinion tweets support measures and government intervention. The remaining 4% may not actively oppose measures but instead support alternative measures and medications, primarily hydroxychloroquine and ivermectin.

3.3.1.3 Visualization Design Violations

In our annotated data, we find that only 12% of all posts contain data visualizations that violate common visualization design guidelines. The prevalence of such posts across tweets that contain reasoning errors and those that do not is similar with 11% and 13%, respectively. Overall, the most common design guidelines violations among tweets with reasoning errors include the use of dual axes (5.4% of all posts), encoding of quantities as area or 3D shapes (5.0%), truncated axes (1.2%), inverted axes (0.6%), with other miscellaneous violations covering another 0.5% of posts.

These results can be contrasted with Lo et al.'s survey of charts tagged as "misleading", found through search engines and social media platforms [50]. Based on the authors' supplemental materials, 57% of charts have "visualisation design" or "perception" issues that roughly translate to our definition of visualization design violations. Among those, 18% of charts have a truncated axis or otherwise inappropriate axis range, 12% use area or 3D encoding, 7% have dual axes, and 2% have inverted axes. Not surprisingly, we see, overall, proportionally fewer visualization design violations, since we sampled visualizations from an online discourse not limited to misleading charts. As seen in Figure 3.4, however, the

prevalence of visualization design violations does not vary based on whether the post is neutral or opinionated, whether there are any reasoning errors in the interpretation, and what side of the argument the post supports. Using a Pearson's chi-squared test, we also find no statistically significant relationship between tweet polarity and presence of design violations ($\chi^2(3) = 3.2046$, $p = 0.3611$), or between presence of reasoning errors and presence of design violations ($\chi^2(1) = 1.5722$, $p = 0.4494$). Moreover, using a McNemar's chi-squared test we find no statistically significant difference in the marginal frequencies of design violations and reasoning errors across all tweets ($\chi^2(1) = 1.514$, $p = 0.2185$). This finding suggests that—relative to the prevalence of their use in support of biased and misinformation conclusions—the issue of design violations is overrepresented in research and in discussions of deceptive visualizations.

3.3.2 Typology of Misleading Visualization Attributes

In this section, we describe the tweets we can identify as misleading by offering a typology of visualization post features. The typology is derived by grouping related codes. We described each post in terms of two types of attributes: attributes of its argument's reasoning—such as cherry-picking or improper causal inferences—and attributes of the post's construction—like the use of annotations on the chart. The attributes are not mutually exclusive and each post can contain none, one, or many attributes. In fact, as discussed in more detail below, certain combinations of attributes can potentially interact with one another. Figure 3.5 provides a summary of the prevalence of individual attributes by polarity, and Figure 3.6 illustrates frequencies and common combinations of attributes in an UpSet plot [93]. As we introduce the various attributes of the typology throughout this section, we return to the figures to discuss relevant observations.

3.3.2.1 Reasoning Errors

Reasoning errors (RE) contain commonly occurring logical attributes that can form the basis of a misleading argument.

RE 1: Cherry-picking. Visualization posts are characterized by cherry-picking when the main conclusion is consistent with the incomplete evidence presented but likely would not be generalizable with more representative evidence.

We can distinguish between two types of cherry-picking: cherry-picking of the data

points and of the time frame. Figure 3.7a shows an example of data point cherry-picking, in which the author argues against the implementation of COVID-19 measures by providing a single data point as evidence. Sweden—a country that did not enforce strict lockdowns and reported fewer deaths and cases than some countries that did—is very commonly used as cherry-picked evidence in COVID-skeptic posts.

The post in Figure 3.7b is an example of time frame cherry picking. The post compares the case curves in the US in the Summers of 2020 and 2021, before and after vaccine availability. The limited time frame allows the author to omit the sharp fall in cases during the initial vaccine distribution in the first half of 2021.

Presenting one data point in support of an argument is not universally misleading. There is a balance between cherry-picking evidence for an argument and providing an illustrative example that condenses large-scale data into a zoomed-in and easy-to-understand chart. However, although an illustrative example is useful to *explain* a phenomenon, the post should offer a way to get additional evidence to prove its *validity*.

RE 2: Setting an Arbitrary Threshold. A common attribute of misleading tweets in our data set is the author setting an arbitrary threshold against which a certain phenomenon is judged. The threshold can be stated explicitly as a number, or be visual as an annotation in a chart.

In the context of public health crises, such as disease outbreaks like the COVID-19 pandemic, the lack of an official threshold provides opportunities for people to define their own. According to *A Dictionary of Epidemiology* [94], a pandemic is "an epidemic occurring worldwide, or over a very wide area, crossing international boundaries and usually affecting a large number of people", whereas an epidemic is "an illness... ...clearly in excess of normal expectancy." The terms "large number" and "excess of normal" in these definitions imply that the level of seriousness of a disease outbreak is to be determined holistically and leaves room for disagreement.

As seen in Figure 3.5, this type of reasoning error is especially common in COVID-skeptic posts, as they attempt to redefine the threshold of seriousness of the pandemic. For instance, the post in Figure 3.8a implies that COVID-19 is not a serious concern because "only" 1 in 23 deaths were caused by it. This post also shows data for a single day, in an example of time frame cherry-picking. The tweet from Figure 3.8b is part of a conversation

discussing that the ongoing pandemic does not warrant measures because the mortality rate in 2020 is "only" 15% higher than the previous five years and comparable to 2003. In both examples, the author takes advantage of the fact that in the context of personal health, people's level of risk aversion and cost-benefit calculations are highly personal. Additionally, the attached data visualizations are used to visually exaggerate the effect. One of the pie chart's largest sections is "All other causes", which makes the COVID-19 deaths appear relatively small. The bar chart includes data as far back as 1999, when the all-cause mortality was higher than in 2020.

RE 3: Causal Inference. Tweet authors often assign cause-and-effect relationships in an attempt to explain certain salient features of a chart and evaluate them. Causal relationships are typically evaluated either by themselves against an author-defined satisfactory threshold as discussed previously, or against another inferred causal relationship. Causality inferred from a visualization can be especially misleading in cases when the data are cherry-picked. This approach is used frequently: as seen from Figure 3.6, causal inference and cherry-picking are among the most commonly co-occurring attributes of a post.

The author of Figure 3.9a evaluates the effectiveness of vaccines by highlighting that the vaccination start date in Uruguay preceded a large spike in cases, implying a causal relationship. Although causal inference is common in posts with all types of arguments, Figure 3.5 shows that this reasoning error is especially common in tweets that promote alternative measures and attempt to prove their effectiveness. For instance, the author of Figure 3.9b provides two examples of cause-and-effect relationships: one of vaccines in Israel and one of ivermectin in Zimbabwe. Since the Zimbabwe case curve is lower, the author argues that ivermectin is more effective.

RE 4: Issues with Data Validity. During a fast-developing and novel crisis, issues may arise with the quality and consistency of data. In a developing situation, the lack of uncertainty communication may result in widespread confusion about what data can be used for inductive reasoning and how. If important caveats about data accuracy and data interpretation are not explicitly provided on charts [95], the viewer is left to trust the data to the level that supports their prior beliefs. The omission of such caveats results in two opposing strategies of using visualizations and data validity issues as basis of one's

argument: appealing to data issues when they are not present, and ignoring them when they are.

The first strategy, illustrated in Figure 3.10a, consists of the poster pointing out a salient feature of a chart, suggesting that it is caused by a data validity issue, and providing an explanation that supports a conspiracy. The author suggests that influenza being mistaken for COVID-19 is the primary reason for the small number of flu cases in 2020, the implication being that the pandemic is exaggerated.

On the other hand, one may reject verifiable data concerns if they provide support for their argument. In Figure 3.10b, the note on the chart explains that "[...] limited testing [...] means that the number of confirmed deaths may not be an accurate count." Nonetheless, the author points out that highly vaccinated Iceland has 119 times more cases than a low-vaccinated Nigeria and makes the argument that vaccination is not useful or potentially harmful. In this case, the data caveat likely has a large impact on the case chart. According to OurWorldInData [92], the source of the case chart, on the date highlighted by the author (August 11, 2021) Iceland performed over 200 times as many tests per capita as Nigeria did (8.93 versus 0.04 per 1,000 people).

As a result, data collection anomalies can be misused as evidence in support of misinformation, whereas explainable phenomena may be abused to sow potentially unwarranted distrust in data quality and reject an argument.

RE 5: Failure to Account for Statistical Nuance. Besides ignoring data issues, users often do not account for important statistical nuance in visualizations if doing so helps support their argument.

For instance, consider an experiment in which a given effect is evaluated by measuring outcomes for treatment and control groups. One common issue observed in our data set is users providing data showing the outcome of a single group in an experiment and judging the efficacy of the treatment against an arbitrary level of "goodness" rather than against the other group. In Figure 3.11a, the author argues that vaccines work by showing that among a population of 102,000 vaccinated, only one death has been reported. However, without knowing the death rate of the general population or of the nonvaccinated population, it is not possible to make conclusions about the efficacy of the treatment.

Another related issue is the *base rate fallacy*: charts providing outcome counts for both

treatment and control groups without information about their relative sizes or about the general population. For instance, the attachment from Figure 3.11b is a chart that often has accompanied COVID-skeptic arguments. The chart comes from a report published by the CDC [96]. The chart shows that the majority of COVID-19 cases in the surveyed population were among previously vaccinated people. The report generally argues for increased COVID prevention measures and highlights the idea that the then newly emerged Delta variant is highly transmissible. When taken out of context, however, this chart can be used to support the idea that people are more likely to get and spread the disease if vaccinated. The report lists multiple limitations to the study in the discussion section, including a note that "..., data from this report are insufficient to draw conclusions about the effectiveness of COVID-19 vaccines against SARS-CoV-2, including the Delta variant, during this outbreak. As population-level vaccination coverage increases, vaccinated persons are likely to represent a larger proportion of COVID-19 cases." This caveat is not communicated on the chart itself, which quickly became a shareable artifact on social media and was interpreted outside of this context.

RE 6: Misrepresentation of Scientific Studies. Promoting scientific literacy has long been discussed as a way to inoculate the general public against misinformation [97]. In an increasingly complex world, it is not feasible for everyone to develop scientific expertise in relevant domains, and therefore one of the main goals of increasing scientific literacy is to encourage the population to cautiously trust science. Science educators describe an ideal of scientific literacy among lay audiences as the right balance between the extremes of believing any form of scientific authority and believing nothing at all [98], [99].

Many misleading tweets in our data set lie close to one of the two extremes. Most of the reasoning attributes described above are characteristic of the users's tendency to believe nothing but their personal experiences and observations, but we also identify a type of reasoning in which users accept any scientific findings that align with their prior beliefs at face value and exaggerate their interpretation. For instance, users share figures from studies on the efficacy of certain types of medication that have not yet been peer-reviewed, reproduced, or otherwise scrutinized, e.g., by approval for use in most countries. Figure 3.12 illustrates examples of such tweets.

In this form of reasoning, instead of rejecting scientific authority in favor of pseudo-

science, users selectively exaggerate the importance of singular scientific results that confirm their beliefs.

RE 7: Incorrect Reading of Chart. In rare cases, visual distortions on the chart directly cause the viewers to arrive at inconsistent conclusions. Figure 3.13a illustrates an example that seems to show many fewer cases in Canada compared to the US, whereas the visual differences in the map are mostly caused by the inconsistent granularity of data between countries (province level for Canada and county level for the US). The author interprets the differences to be caused by COVID-19 restrictions.

In another example in Figure 3.13b, the author attaches a dual axis line chart of COVID-19 cases in counties with and without mask mandates. Whereas the case numbers are actually higher in mask mandate counties than in counties without a mandate, the unequal relative scales of the axes make the line appear lower, which was misinterpreted by the user.

Previous research on misleading visualizations has been primarily focused on this type of error [47]. However, we find incorrect reading of charts to be the least common reasoning error in our analysis, as seen in Figure 3.5.

3.3.2.2 Construction Attributes

Construction attributes (CA) are graphical and textual methods that communicate or emphasize the message of the post. In misleading visualization tweets, construction attributes describe channels used to introduce or exaggerate reasoning errors described previously.

CA 1: Use of Post Text. Prior research by Kong et al. shows that the framing of a visualization greatly influences the viewers' interpretations [58]. In another study, the authors find that the information from the title is more readily recalled by the viewers than the information from the chart itself [100]. This phenomenon provides an opportunity to deceive and mislead users by providing biased interpretations attached to the chart.

In the context of social media as a platform for sharing visualizations, including an interpretation via the post text is an intrinsic feature. In every tweet, authors add text to either provide their own interpretations of the chart or additional context that is important for their argument. In many examples, tweet text is used to frame a visualization in a way

that makes it deceptive.

CA 2: Use of Annotations on Chart. Another channel for providing interpretations and context for a visualization is the use of annotations directly on the chart. Among opinion tweets, 21% of visualizations include textual or graphical annotations. Lin et al. [95] introduce the concept of *data hunches* and describe a design space for adding important context about the data representativeness to the data visualization itself. For instance, annotations added to Figure 3.9a provide additional information about the start of the vaccination campaign, whereas annotations on Figure 3.10a highlight the lack of flu cases—an important salient feature of the chart.

Although annotations on a chart can be essential to its understanding and stem from expert knowledge, they can also be deceptive and suggest relationships and caveats that do not exist. In their paper, Lin et al. warn about the potential for harm and argue for the use of annotations only within "groups of experts that are supported by networks of trust" [95]. In the context of charts shared for general audiences online, our research shows that annotations are likely to be misused and mislead the audience.

CA 3: Reframing Screenshots of Existing Charts. We find that the majority of visualizations in opinion tweets (58%) are screenshots of existing charts from reputable sources. One the one hand, this points to the ubiquity and ease of availability of COVID-19 visualizations and data. As an attribute of a misinformation tweet, however, reframing a reputable chart could provide an illusion of impartiality and trust on the part of the author of the tweet, as well as plausible deniability in case their argument is proven false. Aside from static charts from reputable sources, users often repost charts from interactive dashboards. Such dashboards let the user individually select regions to be plotted on the same chart and compared against each other. Although this approach allows one to freely explore the data, such tools could encourage cherry-picking of data points and faulty comparisons between them. From Figure 3.6, we can see that the combination of the use of interactive dashboards, cherry-picking, and causal inference is the most common set of attributes among posts with any reasoning errors.

Researchers have expressed concern that with the increasing prevalence and complexity of interactive visualizations, nonexpert viewers are more likely to misinterpret visualizations created by experts and identify spurious correlations [101]. Recent work

also confirms that public health visualization dashboard designers should consider the possibility that their work can be used to mislead people [76]. The authors underscore the importance of expert knowledge in correct understanding of pandemic visualizations and seek ways to communicate important context.

CA 4: Violations of Common Visualization Design Guidelines. As seen from Figure 3.5, charts that violate common visualization design guidelines—for instance, those that use truncated axes or 3D figures—are not disproportionately used to support opinionated arguments. Moreover, results in Figure 3.5 show that explicit incorrect readings of charts caused by design violations are rare compared to other errors in visualization interpretation. Nonetheless, research shows that these techniques have the potential to affect the accuracy of viewers' perception of the chart [47].

Our findings suggest that even though design violations are not central to visualization-supported misinformation, they may help exaggerate the intended message. For instance, Figure 3.14a shows a post that attributes the drop in COVID-19 cases in the UK to the vaccination campaign. We would expect the perceived effect to be potentially stronger than the actual effect because the figure has a truncated axis [47]: upon first glance, a viewer might incorrectly infer that cases in the UK have almost completely disappeared. In another example, the post in Figure 3.14b assigns a cause-and-effect relationship to the rise of COVID-19 cases in India and increasing vaccinations. A dual axis chart is a common way to highlight a spurious correlation [102] that can make the association appear stronger and may even suggest a one-to-one relationship between cases and vaccines administered.

3.4 Discussion

In this section, we consider the implications of the results of our analysis on the study of deceptive visualizations. We provide recommendations for general-audience visualization design as well as for the direction of future research.

3.4.1 Visual Misinformation Beyond Design Violations

We were surprised to discover that widely studied common visual deception techniques are not the main driver of visualization-supported misinformation online. The vast

majority of both all charts in our data set in general (88%) and all charts with COVID-skeptic interpretations in particular (87%) do not have any features that violate common visualization design guidelines. The remaining 12% of visualizations violate visualization design guidelines by, for example, using truncated or inverted axes. However, these features are not typically used to support misinformation arguments and seem to occur at similar rates in opinion and neutral posts.

In posts that comment on the severity of the ongoing crisis, visualization design violations can be helpful to exaggerate the effect and help argue that, for instance, the impacts of the crisis are not sufficiently severe. However, the general argument of the post would still hold even if such techniques were not present, like in examples in Figure 3.14. Previous research suggests that design violations may exaggerate or diminish a message to a certain extent [47], but our analysis shows that they do not form the basis of visual misinformation and they are not disproportionately misinterpreted by tweet authors. In contrast, many data visualizations that fully conform to design guidelines can be used to support effective misinformation arguments.

Design violations are not more prevalent in online misinformation for several reasons. Firstly, whereas people may misinterpret the chart after a quick glance, engaging with it for a longer period of time required to attach it to a post and write text to go with it allows them to read the chart more carefully. Secondly, except for outright fabricated data, the information required to interpret the chart correctly is still present even in badly designed charts. We speculate that the possibility to read values accurately—albeit with difficulty—makes it easier for the audience to point out the mistake and to debunk the chart by leveraging their "collective intelligence" [14]. Consequently, it might be more difficult to make a deceptive argument with a mis-designed chart than a well-designed chart that contains more subtle reasoning flaws.

Most visualizations attached to our tweets of interest do not just conform to design guidelines but are also screenshots of charts from reputable sources, such as government and media outlets. Such visualizations are not intentionally created to be deceptive but rather are presented in a way that supports a common misinformation argument or conspiracy surrounding the crisis. This fact suggests that, depending on framing, conversation context, and anticipated audience, even some faithfully plotted visualizations are

vulnerable to misinterpretation.

3.4.2 Visual Misinformation as Weak Inductive Reasoning

One helpful way to think about how framing can make a visualization deceptive is to consider how most, if not all, arguments with reasoning errors are constructed. Typically, users posit an idea supported by a chart and context. The salient features of the visualization and user's context form premises that weakly support an *inductive argument*—in other words, factual observations are used to derive a more general principle. Figure 3.15 illustrates the structure of an inductive argument schematically.

Misinformation arguments of this form generally do not contain formal logical fallacies, as the conclusion always logically follows from the presented premises. Deductive arguments can be either true or not true, whereas inductive arguments are defined to either be "cogent" or "not cogent"—in other words, plausible and not plausible [103]. The strength of an inductive argument, or the measure of how plausible the conclusion is, depends on the completeness and strength of the premises. Most empirical knowledge is also derived through inductive reasoning and is almost never definitive, which philosopher David Hume identified as a problem many centuries ago [104]. As a solution, in order to distinguish between scientific and pseudo-scientific theories, Karl Popper introduced the standard of *falsifiability*: a determination of whether a theory can be logically contradicted [105]. Popper describes pseudo-science as a "confirmation bias machine" that provides theories that are very good at offering explanations for all phenomena but do not present conditions under which the theory could be refuted [106].

Modern misinformation largely resembles pseudo-scientific theories of the past and is not always amenable to fact-checking because the premises are typically based on true data. Cook et al. [107] present an analysis of climate change misinformation arguments through the lens of logic and reasoning and discuss that many climate denialist claims are plausible inductive arguments that are presented definitively. The authors discuss potential ways to invalidate such arguments through falsifiability, such as identifying hidden premises or implicit conditions necessary for the conclusion to hold. If shown to be not true or not plausible, the conclusion becomes not consistent or less plausible.

3.4.3 Hidden Premises Make Visualizations Vulnerable

In the context of visualization, we can consider potential hidden premises in posts in our analysis. As we have seen in examples throughout the chapter, a typical visualization-based weak inductive argument often hinges on various implicit assumptions: posts omit a larger sample size in favor of cherry-picked data points, omit notes about impactful events that are expected to have an effect on the data, and also omit important caveats about data quality and uncertainty. Existing charts often do not communicate any of such data uncertainties as salient features and present data with an illusion of precision or certainty [108]. This false sense of accuracy and certainty likely empowers users to make definitive deceptive claims using visualizations as "scientific evidence."

In many cases, support for a misinformation conclusion may be minimized by explicitly visualizing the conditions or caveats the author wrongly assumes to corroborate their reasoning. For instance, in the case of the chart from the CDC report in Figure 3.11b, showing the population-level vaccination coverage would indicate to the audience that infection is not more prevalent in the vaccinated population; in the case of the chart that compares Iceland and Nigeria in Figure 3.10b, introducing the uncertainty of death counts due to limited testing as a graphical property of the chart may warn users against making comparisons. Therefore, although most visualizations in our analysis are not themselves deceptive, they are *vulnerable to deception*: they do not visualize important context and do not anticipate a biased reading and thus have no defenses against misinterpretations and supporting existing common misconceptions.

Weak inductive reasoning is not unique to misinformation arguments, or in our case specifically COVID-skeptic arguments. 25% of tweets with reasoning errors in our data consist of posts that support commonly accepted methods of COVID-19 prevention and treatment and follow similar reasoning, as for instance the post in Figure 3.11a that attempts to prove the efficacy of vaccines. In this example, however, the conclusion is corroborated by multiple scientific studies about vaccine efficacy [109]–[112]. The omission of the control group outcomes on the chart in this case is not central to the argument. In fact, this omission has the opposite effect and makes the argument less convincing, which implies that understanding the problem of visual misinformation not only helps prevent misinformation but also improves the effectiveness of official crisis communication. By

visualizing a more complete set of premises for the anticipated conclusion, a chart can provide stronger support to the take-away and leave fewer avenues for vaccine-skeptic attacks.

The author or the audience of a visualization post may not consider these important caveats to their interpretations that are not explicitly shown in a vulnerable visualization for several reasons. One reason is confirmation bias, or the tendency to interpret information in a way that reinforces prior beliefs [113]. If the conclusion matches existing beliefs, there is little to no incentive for the viewer to challenge their conclusion. Another reason is the process of social influence, or conformity to the demands of one's social environment [114]. This process is especially relevant in the context of our analysis—social media—where due to filter bubbles and personalized suggestions, users often end up seeing posts from, and interacting with, only one side of the discussion [115]. In such a setting, challenging assumptions of a post's reasoning can be perceived as antagonistic or may fail to get a stranger's attention [116].

3.4.4 The Role of Data Exploration Websites

Screenshots of charts from data exploration tools—such as the OurWorldInData COVID explorer [92] and Worldometer—are common in our data set and in the examples throughout this chapter. On the one hand, their design provides easy access to COVID-19 data and allows even nonexpert users to freely explore and share their findings. At the same time, this freedom of exploration can lead to many of the reasoning errors we have discussed, such as cherry-picking, causal inference, and failure to account for data validity issues.

With their design centered on comparing COVID-19 data from different regions, data exploration tools encourage many types of inferences that are prone to misinterpretation. This type of visualization is useful to contextualize data and answer personally meaningful questions: for instance, it helps infer health risks in a travel destination by comparing it to a familiar local baseline. However, plotting any subset of data on one chart also encourages potentially inaccurate inferences related to evaluating the effects of restrictions and interventions. Such inferences can be both unintentional, as an attempt to make sense of salient differences in the chart, or an intentional way of spreading disinformation by

cherry-picking data.

Our results in Figure 3.5 show that more than 40% of all COVID-skeptic charts in the COVID-19 discourse on Twitter are screenshots of data exploration dashboards. This high prevalence may be explained by their popularity and ease of use. In the absence of explicit data caveats, the space of interactions offered in the dashboard is likely interpreted as the space of data representations that are valid for inference. As a result, users choose any subset of data that supports their existing beliefs or reinforces their (maybe not sincerely) held position.

In the case of OurWorldInData [92], data exploration tools are a relatively recent development: usually, the website provided in-depth data stories supported by interactive graphics, such as on the environmental impacts of food[3] or an analysis of the impact of vaccinations on COVID death rates[4]. We suspect that information presented in this way—as a guided narrative—is less susceptible to accidental misinterpretation. We highlight, however, the tension that platforms that support citizen science face. On the one hand, they inspire creativity, freedom of exploration, and democratize data. At the same time, such freedom can lead to a proliferation of conclusions based on weak premises and misinformation.

3.4.5 Design Recommendations

In this section, we summarize ways in which visualization designers can safeguard their charts or interactive dashboards and prevent them from being vulnerable to misinterpretations. Our design recommendations are practical methods for introducing hidden premises of potential interpretations of a visualization.

As we have briefly discussed before, in many cases it would be possible to **add a very salient feature to the chart that would render the misinformation argument not logical**. For example, the addition of population-level vaccination information to the CDC chart in Figure 3.11b would make it visually apparent that the proportion of COVID-19 patients who were vaccinated does not drastically exceed the proportion of the general population that is vaccinated. Even though the authors did consider the potential misinterpretation

[3] https://ourworldindata.org/environmental-impacts-of-food

[4] https://ourworldindata.org/covid-deaths-by-vaccination

of their report and added a text note in an attempt to prevent it, including it directly in the chart as a visual feature would help make the important context more portable and resilient to viral spread.

Charts should **make information about data caveats and accuracy more salient and dynamic**. For instance, many charts from data exploration websites in our analysis include a note in small font explaining that actual case numbers are likely much higher than reported, due to limited testing in some regions. However, this note is not a prominent feature of the chart and is not reflected in the visual design of the case curve. Additionally, the note is typically a generic statement that appears on all charts in the same form. The note could be more effective if the information about the testing limitations in a given region was chart-dependent and changed dynamically as viewers switch between regions.

Similarly, charts could **communicate the uncertainty of data estimates** to prevent viewers from assigning misleading meaning to data anomalies. Work by Hullman [108] discusses why visualization designers typically do not include uncertainty in their work products, including concerns that uncertainty can confuse the viewers, obfuscate the message of the plot, or undermine the designer's credibility. As our findings show, however, it is potentially beneficial to spread awareness about data imprecision to limit its support of pseudo-scientific arguments. Lin et al. [95] argue that "data hunches" should be explicitly yet distinctly communicated as part of the chart and should be considered in decision-making. Data exploration tools should prevent visualization of multiple items at the same time in cases when it is known that the items have a vastly different representativeness of the underlying phenomenon, as is, for example, the case when countries vary significantly in their testing strategies or data collection methodologies (Figures 3.9b and 3.10b).

In summary, we urge visualization designers to take into account potential misinterpretations of their charts and address them directly in the visual array. Our work shows that if important notes about data collection and use are not communicated as salient visual features in the chart, they are likely to be ignored by the general audience. Designers should consider what information is not shown in the chart and could be filled in by the viewers's beliefs and biases. Misinformation typically converges from many individual ideas to fewer, more common narratives [117], [118]. Therefore, visualization creators should be able to review common existing misconceptions on the topic and consider

whether the visualization could be used to support them.

3.5 Limitations and Future Work

Our research is subject to several important limitations. Firstly, our work is based only on data pertaining to the COVID-19 pandemic discourse on Twitter in English. However, we expect our findings to be generalizable to crisis and noncrisis situations where human behavior and data-driven policy choices have a significant effect on the economic and public health outcomes. Secondly, the specific affordances of a given social media platform may have an effect on the type of content that is shared and widely spread on it. In the case of Twitter, the strict character limit in tweet text may encourage users to post data visualizations, particularly ones with many annotations. The absence of replies in the feed view makes potentially important context added in the discussion not immediately available to the viewer. The fact that tweets are broadcasted in a general feed rather than common information spaces [119], make Twitter optimal for sharing information to reach a more diverse audience compared to, for instance, Facebook [120]. At the same time, lack of such common information spaces can also strip visualizations of their context, making them potentially more misleading. Thirdly, we are able to identify only explicit instances of chart misinterpretation as evidenced by social media trace data. Although the author of the tweet may not have explicitly misrepresent the chart, some of their audience might make inaccurate conclusions that were not shared, which could be valuable to explore in future research. Nonetheless, we believe studying misleading tweets that state their ideas explicitly is important, as this contributes to perpetuation of misinformation arguments.

We hope our contributions help direct future research in the spaces of deceptive visualization and online misinformation by offering a novel way of thinking about the notion of "lying with charts." Future work should formally analyze the replies to misleading social media posts, as well as explore the best ways to represent hidden premises into charts and evaluate the effectiveness of adding safeguards to vulnerable visualizations.

3.6 Conclusion

We collected, categorized, and organized thousands of data visualization posts from Twitter related to the COVID-19 pandemic to describe how people mislead with visualiza-

tions in practice. The results of our work show that visualization manipulation and the lie factor are not the main drivers of visual misinformation online. About 12% of charts we analyzed violate common visualization design guidelines, however, they are not typically used to support misinformation. Most COVID-skeptic data visualization posts on Twitter use faithfully plotted charts, accurate data, and make logically consistent arguments. Instead, tweet authors use salient features of visualizations as the premise for plausible inductive arguments that promote misinformation. In order to prevent a visualization from being vulnerable to such attacks, visualization designers should include safeguards in the form of important contextual information and uncertainty. Aside from optimizing the visualization design for its intended purpose, the designer should also anticipate how a biased viewer may use and misuse the chart.

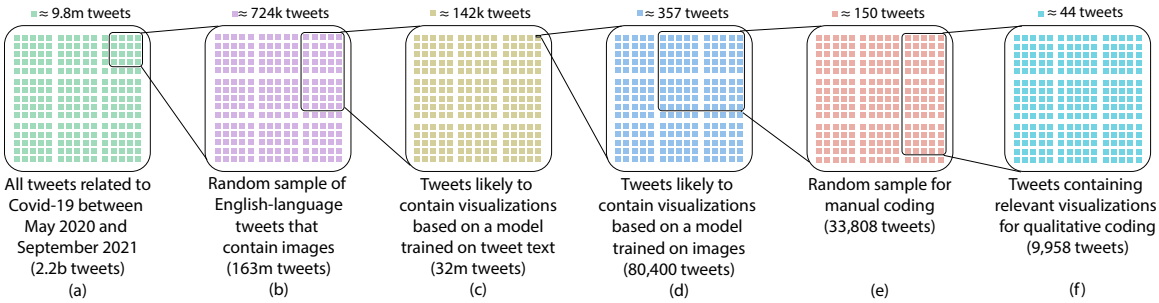
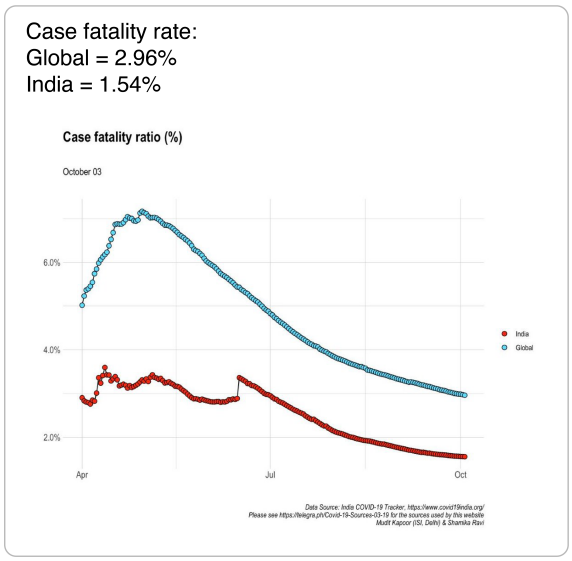


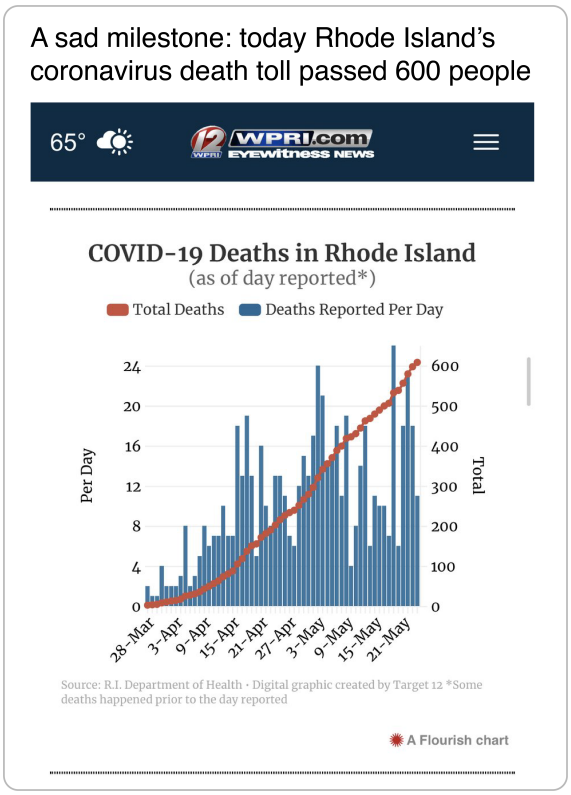
Figure 3.1: Illustration of the data collection and processing pipeline. (a) Initial set of all tweets related to COVID-19 collected between May 15, 2020 and September 6, 2021. (b) Random sample of original English language tweets that contained images. Original tweets indicate posts that are not retweets of other tweets of replies to existing tweets. (c) Tweets identified to contain visualizations based on analyzing tweet text using a Multinomial Naive Bayes classifier trained on tweet text. (d) Tweets further identified to contain visualizations based on a fine-tuned image recognition model. (e) Random sample taken from model outputs to reduce the size for manual qualitative coding. (f) Tweets identified to contain relevant data visualizations after removing images misclassified as visualizations and visualizations not pertaining to the COVID-19 pandemic through manual inspection. This set is used for qualitative coding and analysis.

Category	Code	Count	Percent
(a) Source of visualization	Unknown or created by author	4,453	44.7%
	Screenshot of a static chart from reputable source	2,646	26.6%
	Screenshot of an interactive dashboard from reputable source	2,860	28.7%
(b) Text polarity	Tweet text: neutral	8,419	84.5%
	Tweet text: support measures	517	5.2%
	Tweet text: oppose measures	954	9.6%
	Tweet text: promote alternative measures	68	0.7%
	Annotation native to chart: neutral	266	2.7%
	Annotation native to chart: support measures	45	0.5%
	Annotation native to chart: oppose measures	58	0.6%
	Annotation native to chart: promote alternative measures	17	0.2%
	Annotation added to chart: neutral	164	1.6%
	Annotation added to chart: support measures	44	0.4%
	Annotation added to chart: oppose measures	121	1.2%
	Annotation added to chart: promote alternative measures	16	0.2%
	Title on chart: neutral	8,816	88.5%
	Title on chart: support measures	120	1.2%
	Title on chart: oppose measures	29	0.3%
	Title on chart: promote alternative measures	4	0.0%
(c) Visualization design violation	Truncated axis	116	1.2%
	Dual axis	541	5.4%
	Value as area/volume	494	5.0%
	Inverted axis	57	0.6%
	Uneven binning	10	0.1%
	Unclear encoding	40	0.4%
	Inappropriate encoding	18	0.2%
(d) Reasoning errors	Cherry-picking data	514	5.2%
	Cherry-picking timeframe	69	0.7%
	Setting an arbitrary threshold	453	4.5%
	Causal inference	691	6.9%
	Suggesting data validity issues	80	0.8%
	Failure to account for data validity issues	65	0.7%
	Failure to account for statistical nuance	105	1.1%
	Misrepresentation of scientific studies	26	0.3%
	Incorrect reading of chart	10	0.1%
Total		**9,958**	**100.0%**

Figure 3.2: The codebook used to manually annotate the 9,958 relevant visualization tweets and the frequencies of codes. The codebook includes information about the (a) source of the visualization, (b) presence and polarity of textual components of the post and the chart, (c) presence of any visualization design violations, as well as (d) commonly occurring reasoning errors. Each post is described by one code from (a), and any number of codes from (b), (c), and (d).



(a) Tweet 1313069181873528834



(b) Tweet 1264592009219715072

Figure 3.3: Post with visualizations containing violations of common visualization design guidelines that are not identified as misleading. Importantly, we cannot state that they are unambiguously *not* misleading, as previous research has shown the deceptive potential of design violations [47]. (a) A chart with a truncated axis that a viewer could misinterpret to show an almost 0% case fatality rate in India but accurately described by the author. (b) A chart with a dual axis that a viewer could misinterpret to show that daily deaths are on the order of hundreds but accurately interpreted by the author.

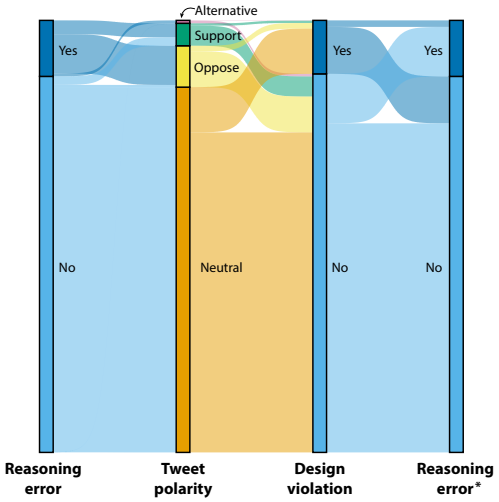


Figure 3.4: A flow diagram showing the relationship between tweet text polarity, presence of visualization design violations, and presence of reasoning errors. We observe similar proportions of tweets with and without design violations across text polarities. We also observe similar marginal frequencies of design violations and reasoning errors. The reasoning error column is shown twice, indicated by the asterisk.

Attribute	Opinion tweets						Neutral tweets	
	Support measures		Oppose measures		Alternative measures			
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Reasoning errors:								
Cherry-picking	144	27.9%	380	39.8%	51	75.0%	*Not applicable*	
Setting an arbitrary threshold	8	1.5%	445	46.6%	0	0.0%		
Causal inference	218	42.2%	415	43.5%	58	85.3%		
Issues with data validity	5	1.0%	136	14.3%	4	5.9%		
Failure to account for statistical nuance	76	14.7%	28	2.9%	1	1.5%		
Misrepresentation of scientific studies	6	1.2%	16	1.7%	8	11.8%		
Incorrect reading of chart	8	1.5%	2	0.2%	0	0.0%		
Any reasoning error	321	62.1%	904	94.8%	66	97.1%		
Construction attributes:								
Use of annotations on chart	87	16.8%	193	20.2%	34	50.0%	413	4.9%
Reframing screenshots of existing charts:								
Screenshot of static chart	192	37.1%	125	13.1%	9	13.2%	2,320	27.6%
Screenshot of interactive dashboard	141	27.3%	399	41.8%	20	29.4%	2,300	27.3%
Any screenshot	333	64.4%	523	54.8%	29	42.6%	4,620	54.9%
Violations of visualization design guidelines:								
Truncated axis	18	3.5%	19	2.0%	1	1.5%	78	0.9%
Dual axis	19	3.7%	49	5.1%	3	4.4%	470	5.6%
Value as area/volume	13	2.5%	33	3.5%	0	0.0%	448	5.3%
Inverted axis	2	0.4%	12	1.3%	2	2.9%	41	0.5%
Uneven binning	5	1.0%	2	0.2%	0	0.0%	3	0.0%
Unclear encoding	8	1.5%	14	1.5%	1	1.5%	17	0.2%
Inappropriate encoding	1	0.2%	11	1.2%	0	0.0%	6	0.1%
Any design violation	54	10.4%	129	13.5%	7	10.3%	1,043	12.4%
Total	**517**	**100.0%**	**954**	**100.0%**	**68**	**100.0%**	**8,419**	**100.0%**

Figure 3.5: The results of qualitative coding by polarity of the tweets. Neutral tweets are posts that do not offer explicit or implicit interpretations of data, and therefore reasoning error codes do not apply. Percent columns do not sum up to 100 as a tweet can have none, one, or multiple attributes. The prevalence of design violations is consistent across opinion tweets and neutral tweets. Moreover, potentially misleading design violations in opinion tweets are much less common than reasoning errors.

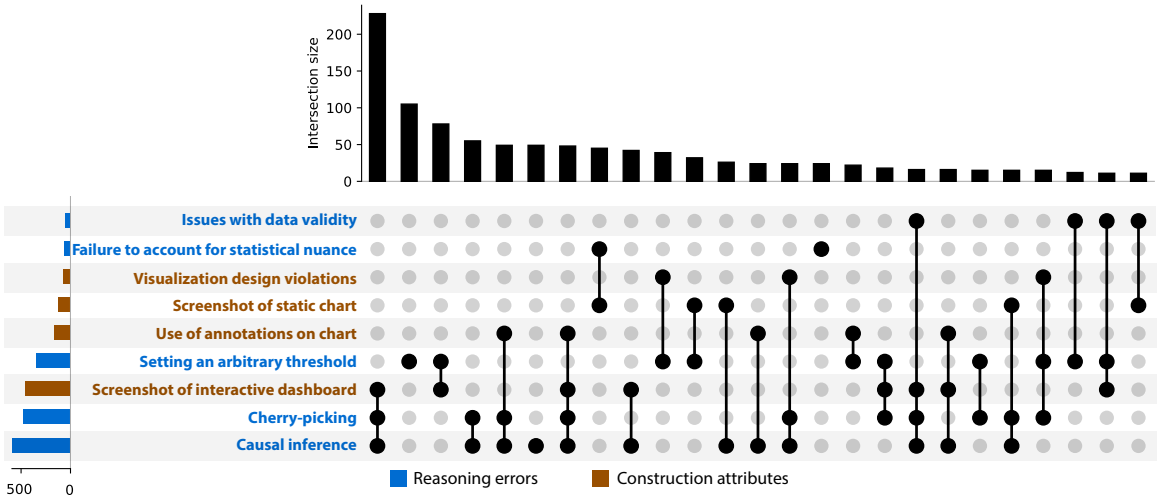
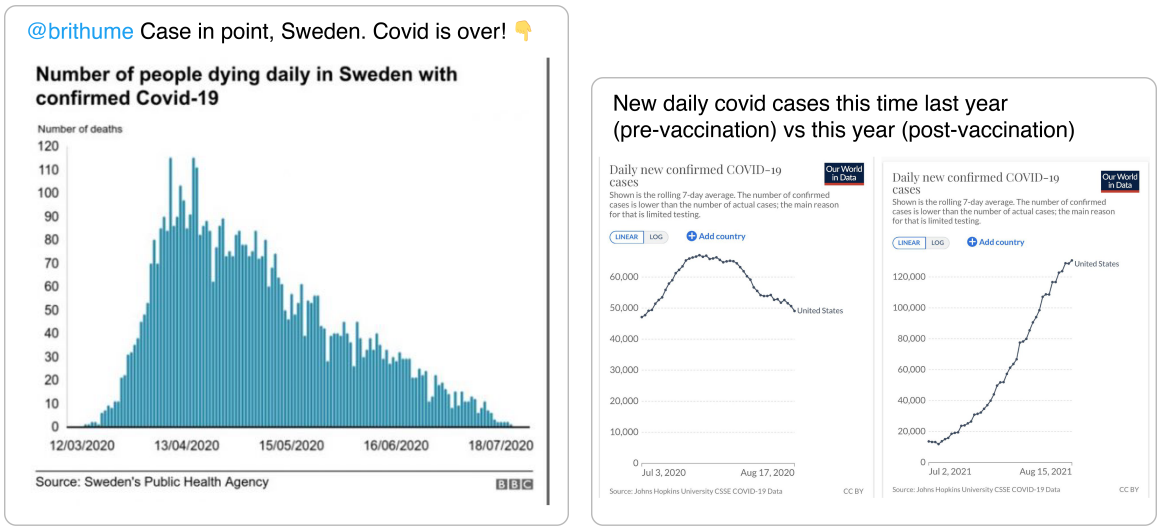


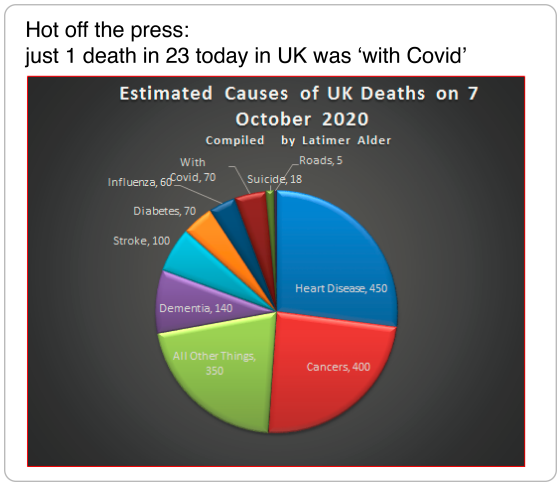
Figure 3.6: The prevalence and overlap of visualization tweet attributes among tweets with reasoning errors shown as an UpSet plot [93]. The horizontal bars show the total prevalence of a given attribute. The vertical bars reflect the number of tweets that have the exact combination of attributes highlighted by black dots directly below. We show the 25 most common overlapping attribute sets that describe 84% of the data. The remaining 16% are described by combinations that are more rare. For instance, reasoning errors such as Misrepresentation of Scientific Studies and Incorrect Reading of Chart are not part of the 25 most common attribute sets and are omitted.



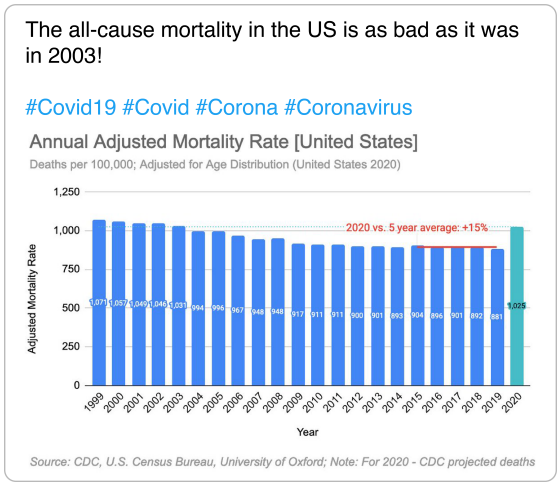
(a) Tweet 1289587370082234370 [removed]

(b) Tweet 1427307226264489991

Figure 3.7: Examples of posts that employ *cherry-picking*. (a) An author shares a screenshot from the BBC showing a COVID-19 death curve from one country that had fewer government-mandated restrictions relative to their neighbors (Sweden) approaching zero and arguing that COVID-19 restrictions are unnecessary altogether. In this post, the argument is more effective because of the omission of comparable neighbor countries that experienced fewer COVID-19 deaths, or other countries with limited restrictions that experienced more deaths. (b) A user shares COVID case curves for the United States for the periods of early July through mid August for the years 2020 (before the vaccination campaign) and 2021 (during the vaccination campaign). Because the number of cases in August 2021 is higher than in August 2020, the user suggests that the vaccination campaign failed. This example carefully selects the time frame that most effectively supports the argument, omitting a large drop in cases in Spring 2021.

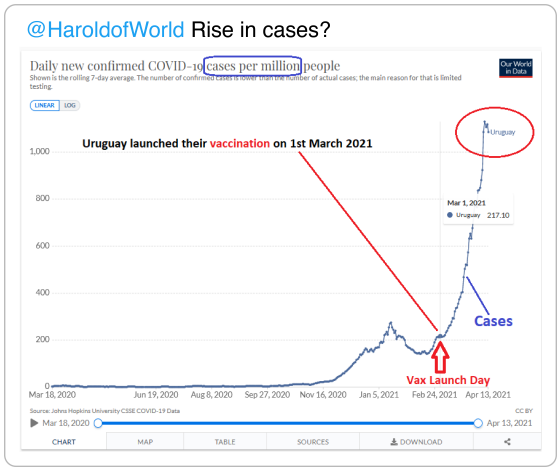


(a) Tweet 1313881187191054338

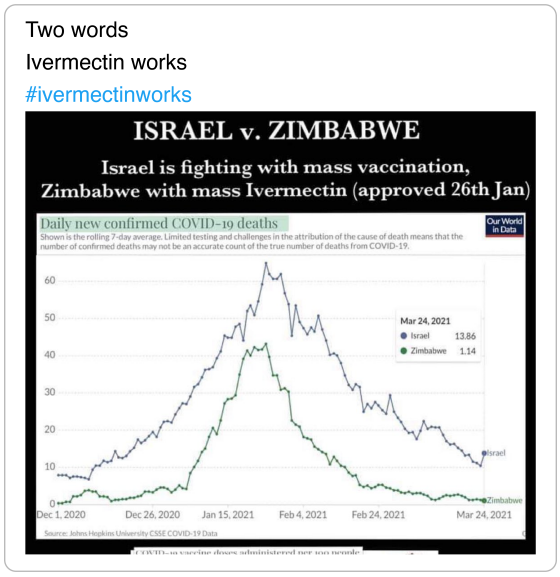


(b) Tweet 1375113309922291712

Figure 3.8: Examples of posts in which the user's argument hedges on the introduction of an *arbitrary threshold*. (a) A user shares data on causes of UK deaths and argues that since "only" 1 in 23 deaths was caused by COVID-19, it is not a significant problem. (b) An author shares a chart of annual mortality rate in the US and argues that an increase in deaths of "only" 15% is not significant enough.

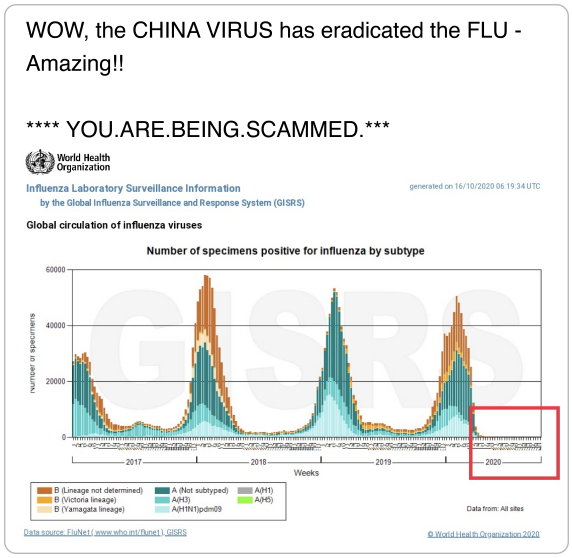


(a) Tweet 1382439566058065927 [removed]

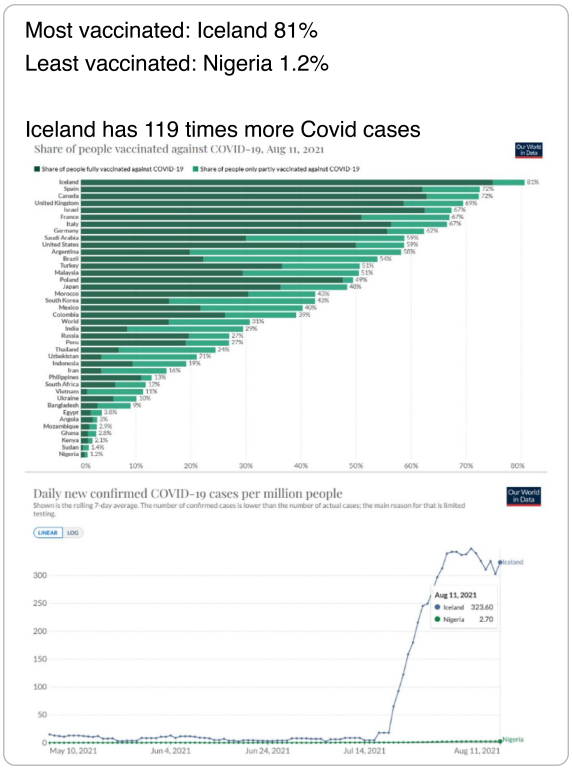


(b) Tweet 1375342512576077825 [removed]

Figure 3.9: Examples of posts implying *causal relationships* from limited and anecdotal data. (a) A tweet author suggests that the increase in COVID-19 cases in Uruguay—a prominent feature of the chart—was caused by the vaccination campaign. The user adds annotations to a COVID-19 dashboard screenshot to explain their reasoning. (b) An author shares a screenshot of a COVID-19 dashboard that shows that Israel is experiencing more COVID-19 deaths than Zimbabwe. The author states that the discrepancy is due to ivermectin being more effective than vaccinations.

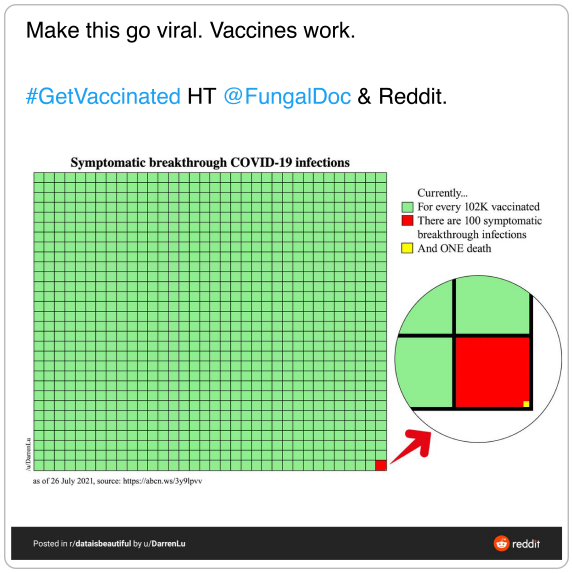


(a) Tweet 1317061948228460546

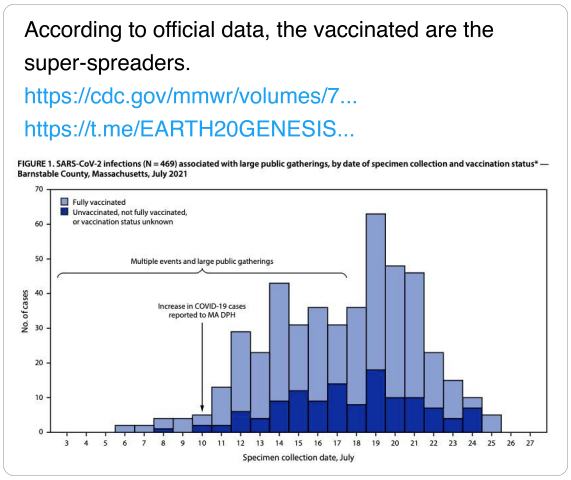


(b) Tweet 1425925153322635276

Figure 3.10: Examples of misleading arguments that suggest or do not account for *data validity*. (a) An author shares a WHO chart of flu cases and highlights a period of no new cases of flu. The author suggests that the COVID-19 pandemic is not real and the data are manipulated by miscounting flu cases as COVID-19. (b) A user shows two charts to highlight the fact that although Iceland is more vaccinated than Nigeria, it is experiencing more COVID-19 cases, implying that vaccines are not effective. The user fails to account for the fact that Iceland has a much higher testing rate, making it unreasonable to compare the two countries.

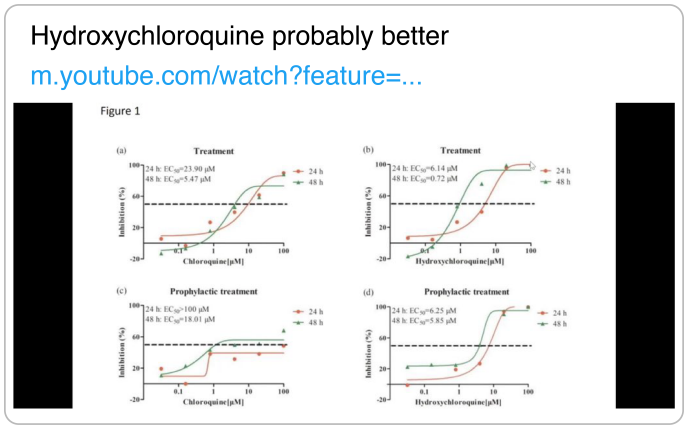


(a) Tweet 1420023807318843395

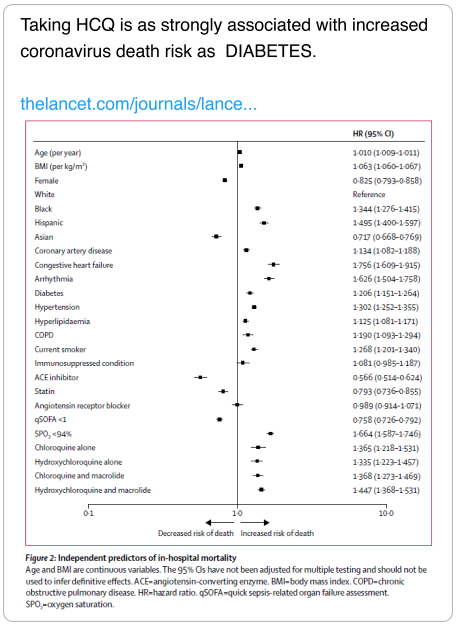


(b) Tweet 1422695341686677504

Figure 3.11: Examples of posts in which users *fail to account for statistical nuance* in their visualizations. (a) A user posts a chart showing the frequency of COVID-19 infections and deaths in the vaccinated population. The user concludes that vaccines are effective based on this chart alone, which is not possible without comparing the data to those from a control group. (b) An author reposts a chart from a CDC report [96] showing COVID-19 cases in Barnstable County, Mass. broken down between vaccinated and nonvaccinated. The author suggests that since there are more vaccinated cases, the vaccinated are "super-spreaders." The author fails to account for the high proportion of vaccinated in the general population. This caveat is highlighted in the text of the CDC report [96]: however, it is not communicated in the shared figure.

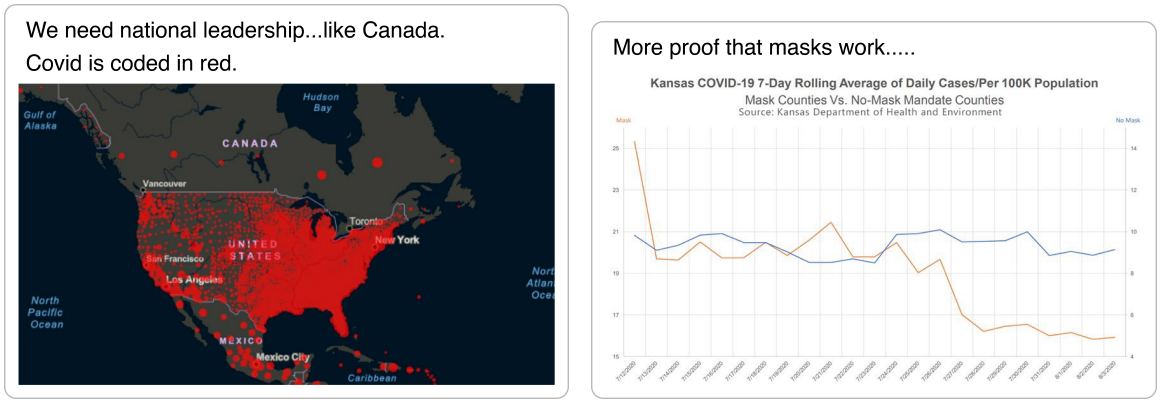


(a) Tweet 1239776019856461824



(b) Tweet 1264224112311844864

Figure 3.12: Examples of posts in which users *misrepresent scientific studies* through visualizations. (a) An author argues for the use of hydroxychloroquine in treatment of COVID-19. The shared figure is from an early 2020 in vitro study [121]. Later work by Lee et al. [122] notes that that although the evidence of the efficacy of hydroxychloroquine in vitro was promising, large-scale randomized clinical trials conducted afterwards have demonstrated low efficacy. The authors also discuss that the large number of hydroxychloroquine studies has been influenced largely by political pressure [122]. (b) A user argues against the use of hydroxychloroquine noting that it leads to an increased risk of mortality. The figure comes from a study that has since been retracted [123] due to concerns about the veracity of the data.



(a) Tweet 1322031826744020992 (b) Tweet 1291564622152183809

Figure 3.13: Examples of posts in which users *incorrectly read the charts* that violate common visualization design guidelines. (a) A user posts a map of COVID-19 cases in North America. The data in the map are not adjusted by population and is presented at different granularities: by county in the US and by province in Canada. As a result, the map looks much sparser in Canada than in the US. It is not possible to make precise readings from the map due to distortions, but the author nonetheless concludes that stricter COVID-19 measures in Canada are the cause of the visual discrepancy. (b) An author posts a dual axis chart of COVID-19 cases in Kansas broken down by counties with and without mask mandates. Both axes are truncated in such a way relative to each other (15–25 cases per 100k for counties with mask mandates, shown in orange, and 4–14 cases per 100k for counties without mask mandates), that even though counties with mask mandates have a higher number of cases, the line associated with them appears lower. The poster likely did not notice the vastly different scales and argues for the use of masks using this chart.

(a) Tweet 1373985666225213440 [removed]

(b) Tweet 1388578805636157440

Figure 3.14: Examples of posts in which violations of common visualization design guidelines potentially exaggerate the argument. (a) A user highlights the fall in COVID-19 hospitalizations in the UK and attributes it to vaccination. The perceived scale of the decrease is likely exaggerated by the truncated axis and by the vertically elongated scale of the chart. (b) An author shares a dual axis chart of vaccinations and deaths in India and suggests a causal relationship between the two, which is likely exaggerated by the use and scale of the dual axis.

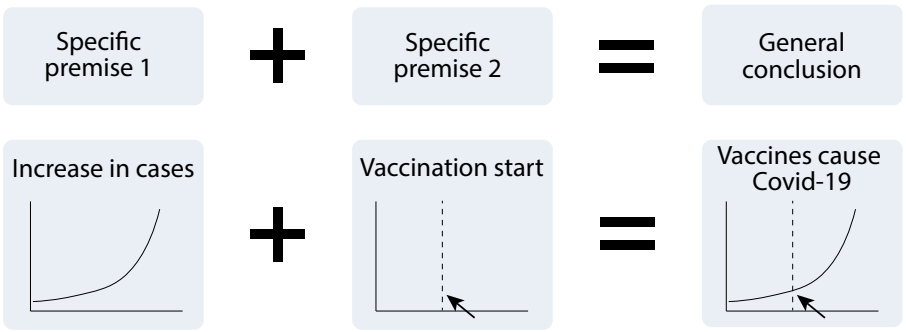


Figure 3.15: An example of inductive reasoning in a visualization tweet, similar to the tweet in Figure 3.9a. A user starts with a first premise: an existing chart showing an increase in COVID-19 cases. Through tweet text and chart annotations, the user provides a second premise: a vaccination campaign started around the same time the increase started. The user then suggests the conclusion: the vaccination start was the cause of the case increase. This attempt at creating a general conclusion from specific premises is logically consistent, but the conclusion is weakly supported by the limited premises.

CHAPTER 4

"YEAH THIS GRAPH DOESN'T SHOW THAT": ANALYSIS OF ONLINE ENGAGEMENT WITH MISLEADING DATA VISUALIZATIONS

Attempting to make sense of a phenomenon or crisis, social media users often share data visualizations and interpretations that can be erroneous or misleading. Prior work has studied how data visualizations can mislead, but do misleading visualizations reach a broad social media audience? And if so, do users amplify or challenge misleading interpretations? To answer these questions, we conducted a mixed-methods analysis of the public's engagement with data visualization posts about COVID-19 on Twitter. Compared to posts with accurate visual insights, our results show that posts with misleading visualizations garner more replies in which the audiences point out nuanced fallacies and caveats in data interpretations. Based on the results of our thematic analysis of engagement, we identify and discuss important opportunities and limitations to effectively leveraging crowdsourced assessments to address *data-driven misinformation*.

This content of this chapter is drawn directly from M. Lisnic, A. Lex, and M. Kogan, ""Yeah, this graph doesn't show that": Analysis of online engagement with misleading data visualizations," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2024, pp. 1–14. doi: 10.1145/3613904.3642448 [7]. It is included in the dissertation with the permission of my coauthors, Alexander Lex and Marina Kogan.

4.1 Introduction

Advances in data collection and data literacy and the rapid spread of information on social media have enabled us to use data visualizations to quickly discover and spread awareness about signs of otherwise invisible phenomena, such as climate change or viral disease epidemics. However, although data helps us uncover evidence of an event or make sense of it, an erroneous analysis may provide an illusion of evidence, lead to

false discoveries or false accusations, or trigger rumors. Whether intentional or stemming from misunderstanding, incorrect or incomplete interpretations of visualizations on controversial topics have the potential to cause harm by spreading misinformation. Indeed, research has documented that misleading data visualizations have been used in support of misinformation on a variety of topics, such as COVID-19 skepticism [6], [56], climate change denial [124], false claims of election fraud [125], and QAnon conspiracy theories [63].

Prior work has highlighted the ways in which data visualizations can deceive an audience due to visual tricks and mirages [47], [50], [126], [127]. However, charts that support misinformation arguments are most commonly well-designed and mislead viewers by being vulnerable to biased framing, misinterpretations, and logical fallacies. [6]. With the rise in popularity of interactive data exploration sites for COVID-19 data such as OurWorldInData [92] or Worldometer [128], the ability to create professional-looking data visualizations has become more democratized and accessible to non-expert users. Consequently, however, the problem of well-designed charts being vulnerable to misinterpretations has reached the scale of mass audiences and is used to fuel misinformation arguments on social media, with 42% of COVID-19-skeptic visualizations shared on Twitter[1] being screenshots of reputable data explorers with a recognizable style and branding [6]. For instance, the original post in Figure 4.1 attempts to promote a COVID-skeptic argument by sharing a data explorer chart showing a cherry-picked data selection, which was effectively countered by the analogous visualization in the reply.

However, can social media audiences always reliably point out such misleading tactics? Despite the fact that numerous studies have examined the spread [129], correction [25], [34], [35], and moderation [130], [131] of misinformation on social media, this research is mostly focused on text and has yet to examine how people share and react to visualization-supported misinformation. As a consequence, it is unclear whether existing findings on misinformation interventions also apply to misinformation supported by misleading visualizations. In their recent work, Weikmann and Lecheler discuss that visual disinformation, including misleading visualizations, is "its own type of falsehood

[1]Known as X since July 2023.

[that] differs from textual disinformation" because it allows for a higher level of manipulative sophistication [62]. All of the above points to the existence of a research gap in understanding the public engagement with, and the potential for mitigation of, data visualization-supported misinformation that opens the door for harmful rumors and conspiracies.

This chapter attempts to fill this research gap by presenting the results of a mixed-methods study of engagement with both misleading and accurate insights in COVID-19 data visualization posts on Twitter. We attempt to answer the following questions:

RQ1: Do misleading insights in a data visualization post have an effect on the count and duration of its engagement? **RQ2**: Do people identify and raise awareness about misleading data insights in their responses?

Based on the results of our work and a review of existing misinformation literature, we discuss the ways in which *data-driven misinformation* in visualization posts is distinct from factual forms of misinformation that are typically studied, such as misinformation based on text or deepfakes. We posit that **existing mitigation strategies may not be sufficient in supporting the verification of nuanced misinformative data interpretations** such as statistical fallacies or data collection caveats. Moreover, data visualizations are associated with credibility indicators that are distinct from those that apply to other types of misinformation, namely the source of the chart and the data, perceived data literacy and analytical expertise of the author, and perceived data integrity.

This chapter makes several contributions:

- Firstly, we conduct a quantitative study of engagement with posts containing data visualizations on social media. Our results show that posts offering interpretations of data are shared twice as frequently—regardless of the their accuracy. Misleading data interpretations garner an additional 60% more replies compared to accurate insights.
- Secondly, we present the results of a thematic analysis of replies to posts with interpretations of data visualizations through a series of case studies. Our findings show that the crowd has the potential to find and reason about nuanced caveats in misleading data-driven insights on social media.
- Thirdly, the results of our thematic analysis also describe important limitations of

the crowd's ability to effectively verify misleading data-driven insights using the existing platform affordances. We discuss approaches that could help tackle these limitations, such as meta-analyses, counter-analyses, and trust-building for data sources and analysts.

- Lastly, we describe the differences between data-driven misinformation and other forms of misinformation on social media and discuss important considerations in designing interventions to address it.

4.2 Study 1: Quantitative Analysis of Engagement

To address the question of whether accompanying a data visualization post with an insight—and, moreover, a misleading one—has an effect on audience's engagement with the post (**RQ1**), we conducted a quantitative analysis of engagement. Specifically, this analysis allows us to identify whether misleading data visualizations are associated with being discussed, shared, or liked more than other posts. In this section we describe our approach to data collection and regression analysis and summarize the results of our Study 1.

4.2.1 Methods

In order to quantitatively analyze the effects of visualization insights on engagement, we used our data to estimate regression model coefficients. In this section, we describe our approach in detail, from engagement data collection to considerations in model selection.

4.2.1.1 Data Collection and Processing

As the basis for our data collection, we used the publicly available data set and supplemental materials from Lisnic et al.'s study of misleading data visualizations on Twitter, which spans the time period between May 15, 2020 and September 6, 2021 [6]. In their data set, the authors provide tweet IDs and the corresponding descriptive variables, such as tweet polarity, presence of reasoning errors, or violations of visualization design guidelines. Of the 9,958 tweets from Lisnic et al. [6], 1,060 have been removed from the platform or made private by the authors, resulting in 8,898 original tweets used in our analysis.

In order to analyze engagement, we used Twitter API's full-archive search to collect

the complete engagement data associated with the original tweets: we collected 668,173 retweets, 229,764 replies, and 101,705 quote tweets for a total of 999,642 *engagement tweets*. To control for tweet author effects in our regression analyses, we additionally collected user data for all tweet authors in our data set to use as covariates, including follower count and verified (or "blue check") status. Our data collection occurred between February and March of 2023, and as such was not affected by the changes to Twitter's verification program from April 2023.

We minimally processed the data by merging engagement tweets and author data with the original tweet data. We provide our data processing scripts as well as tweet IDs of posts used in our analysis in the supplemental materials. To comply with Twitter's API policies, we are unable to provide full tweet data but it may be rehydrated using the IDs, as long as the tweet is still publicly accessible.

4.2.1.2 Regression Analysis

To **analyze the effects of providing accurate or erroneous insights in a data visualization post**, we conducted a regression analysis of count and duration of the main forms of engagement: replies, retweets, quotes, and likes. As our explanatory variables, we used the *opinion* and *reasoning error* data from Lisnic et al. [6]. In our analysis, we use the term *insight* to refer to Lisnic et al.'s opinion variable, which denotes tweets in which the author explicitly highlights or hints at observations, trends, or hypotheses in the data. Non-insight posts share data visualizations without interpretation, such as neutral status updates. Most insights are explicitly stated in the tweet text or added annotations, but some are inferred by holistically analyzing the tweet author's feed and follow-up replies [6].

To model the engagement count variables—the number of replies, quotes, retweets, and likes of a post—we fit Negative Binomial regression models. Negative Binomial regressions are a generalization of Poisson regressions that are commonly used to model count data. Negative Binomial models loosen the assumption of variance being equal to the mean used in Poisson models, and are thus more appropriate for our highly dispersed data, confirmed by the over-dispersion coefficient θ being highly statistically significant in our Negative Binomial regressions. Additionally, we confirmed that Negative Binomial

regressions outperformed Poisson on our data by various other model selection criteria, such as Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC), and Mean Absolute Error (MAE). As a robustness check, we provide the results of Poisson regressions and model selection tests in the supplemental materials, as well as the scripts used to generate them.

Social media engagement data generally tends to be highly right-skewed—with most posts receiving little to no engagement and few posts going viral [132],[133]—which is also the case with our data. One of the sources of high skewness we observe is the high number of zeros in the distribution of the reply counts, with 43% of tweets in our data set having no replies. It is possible that a post may receive zero replies via two mechanisms: structural zeros in posts that signify lack of interest in commenting on a post (or, being the first to comment on a particular post), and random zeros that stem from the fact that the post was not seen by enough people. To account for the excess zeros and model the two ways of generating such excess zeros in our reply data, we fit a Zero-Inflated Negative Binomial (ZINB) regression. A ZINB regression is a type of zero-augmented approach that models a mixture of two distributions: a logistic regression that models generation of zeros, and a Negative Binomial regression that estimates reply count. Zero-inflated regressions are a commonly used way to model social media engagement data [134]–[136].

Despite doubling the model complexity, as seen in Table 4.1, in our model selection tests the ZINB model for reply counts also outperformed the non-zero-augmented approach using Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC), which account for the additional model complexity of a ZINB. Table 4.1 also shows that for other metrics—retweets, quotes, and likes—the zero-inflated approach shows improvement in some metrics but not others, which is expected since their distributions, albeit still skewed and having excess zeros, contain fewer zeros than the replies. For consistency, we present the results of ZINB models for retweets, quotes, and likes as well; however, we note that the coefficients of corresponding non-zero-inflated models are similar and can be found in the supplemental materials.

In addition to engagement counts, we also investigated **the effect of data insights on the duration of the post's engagement**. Duration of engagement is calculated as time elapsed in hours between the original post and the latest reply, retweet, or quote tweet

as of February 2023. Since the Twitter API does not provide timestamps of individual like events, we are unable to make inferences about duration of likes for posts. To model engagement duration (a continuous variable rather than a count variable), we fit standard multiple linear regression models with log-transformed response variable, to account for the skewness.

The results of regressions presented in this chapter correspond to models that control for author-, visualization-, text-, and time-specific covariates. *Author features* include (log-transformed) number of followers and verified status. *Visualization features* describe whether the attached data visualization is a screenshot of an existing chart, has any author-added annotation, or has any violations of common visualization design guidelines (e.g., truncated, inverted, or dual axes). *Text features* control for the number of words in the tweet, as well as the number of mentions, emojis, hashtags, and external URLs. *Time features* include weekend and time-of-day fixed effects, separated into four six-hour segments. In the interest of robustness, we calculated the results excluding different sets of covariates and note that the statistical significance and magnitude of observed effects are consistent across model runs.

4.2.2 Results

Figures 4.2 and 4.3 show results of the Negative Binomial regressions of engagement counts as well as the logged duration of engagement regressions, respectively.

4.2.2.1 Engagement Count

Based on the results shown in Figure 4.2, we observe that **data visualization posts that provide an insight** by offering an interpretation or pointing out a specific aspect of the chart (as opposed to simply sharing a chart) **are associated with significantly higher levels of all forms of engagement.** Specifically, our results show that providing an insight is associated with, on average, 60% more replies, 147% more retweets, 65% more quotes, and 129% more likes.

As seen in Figure 4.3, an erroneous insight in a post is associated with an additional 60% more replies. The effect of errors on other types of engagement—such as retweets or likes—is limited in effect size and statistical significance. In other words, **an erroneous data interpretation attracts significantly *more discussion* while having *no effect on the***

breadth of spread.

4.2.2.2 Engagement Duration

We find that providing an insight is associated with longer-lasting engagement and conversations: as seen from Figure 4.2 our model with the complete set of covariates shows that posts with insights are associated with, on average, 131% longer duration of replies, 274% longer duration of retweets, and 313% longer duration of quotes.

The results of our duration regressions in Figure 4.3 also indicate that there is not a similar effect of reasoning errors on the longevity of engagement. We find, on average, slightly longer duration in replies and slightly shorter in retweets and quotes; however, the effect sizes and levels of significance are relatively low.

In summary, our results show that data visualization posts with insight remain relevant for a much longer time than those without. In the context of COVID-19, we speculate, based on these findings, that visualizations without an insight are used as status updates and provide the latest statistics that may be relevant for only one day (median of 14 hours). At the same time, posts with interpretations use the same data to tackle more fundamental questions, garnering discussions that last multiple days (median of 29 hours).

4.3 Study 2: Thematic Analysis of Engagement

Following the results of Study 1, we set out to explore the contents of replies to posts with erroneous insights and investigate whether online audiences are able to identify and raise awareness about misleading data interpretations as evidenced by the content of their replies (**RQ2**). To do so, we performed a thematic analysis of direct engagement in a subset of our data. In this section we describe our approach and summarize the results of Study 2.

4.3.1 Methods

With the goal of qualitatively analyzing the engagement with data visualization insights, we performed *template analysis* [137] to construct a hierarchical code book that describes the content of replies and quotes of posts in our data. This section outlines our process in detail, from selecting a sample of data for thematic analysis to performing quality and reflexivity checks.

4.3.1.1 Data Selection

To select a sample of data that is large enough to allow us to identify important themes yet small enough to be analyzed it in depth, we performed multistage stratified sampling. Firstly, we filtered our data set to posts that contain an *insight*—observations, trends, or hypotheses in the data highlighted by the author [138]. In the original data set, Lisnic et al. [6] use the term *opinion tweets* for this concept. These are the posts that are, by definition, amenable to being misleading and therefore are the focus of our engagement analysis.

Secondly, to limit our data to relevant engagement with the original post in question, we selected all first-level reply posts or quote posts, except for those authored by the same user as the original post. These posts form a set of all posts that *directly engage* with the original post, as opposed to replies or author's own follow-ups or threads. Thirdly, with the goal of reviewing a richer variety of responses, we excluded posts with fewer than 16 direct engagements, which is the median value among posts with any direct engagement. Lastly, to reduce our sample for thematic analysis, we randomly sampled 30 posts with a reasoning error and 30 without, for a total of 60 original posts with median-or-above engagement count. We then used all of their associated 3,806 first-level replies or quotes for our thematic analysis.

4.3.1.2 Template Analysis

Our approach to thematic analysis was guided by the template analysis techniques described by King [137]. In choosing a methodology for our thematic analysis of engagement with data visualizations, our goal was to strike a balance between the structure of "small q" qualitative methods that emphasize development of coding schemes, and a more contextual and reflexive analysis of themes offered by "Big Q" qualitative approaches, as described by Braun and Clarke [139]. In the context of this research, we wanted to acknowledge the participatory role of the researcher and our research goals, as well as our interpretation of the cultural and semantic context of social media discourse in our conceptualization of themes, while leveraging a structured code book to assist us in describing individual tweets—a relatively independent and small unit of analysis. At its core, template analysis involves developing a code book called a *template* in a way similar to more positivist and postpositivist approaches; however, the template is used as a tool

to help the researcher scaffold data and conceptualize themes rather than a way to convert qualitative into quantitative data [137], [139], [140].

The process of developing the coding template started with the first author reviewing a random sample of 500 first-level replies and noting an initial set of codes. Although we generated most of our code book inductively, in order to more efficiently process our large data set, we deductively defined a set of a priori codes [137] based on existing literature and our own domain knowledge. The lens through which we developed the initial set of codes was guided by the authors' interest in examining how social media audiences review or fact-check misleading data visualization posts. Consequently, our thematic analysis is influenced by the initial code book's direction and pays special attention to users' general analytical engagement with data and data insights, rather than specifics particular to the topic of the posts, COVID-19 data. In the next step, the first author reviewed the complete set of 3,806 direct engagement posts, iteratively revising the contents and structure of the code book. Lastly, the authors used subsets of the code book to conceptualize themes by highlighting and contrasting higher order categories of codes from the final template.

With the goal of validating and scrutinizing the analysis, we performed two iterations of quality and reflexivity checks, as described by King [137]. The first check occurred after development of an initial template and involved a coder independently coding 500 randomly selected posts using the initial template. The first author met with the coder to discuss whether the codes were straightforward to apply, whether the data was easily described by the codes, and whether the template failed to capture any relevant themes. As a result, a new theme related to audience's communication of trust was conceptualized and the template was adjusted for clarity. The second check occurred after the first author completed reviewing the full data set and developed an updated template. In the second check, two senior authors independently coded different subsets of 100 posts each. All the authors met twice, once in the middle of the check and once at the end, to discuss the clarity and richness of the template. Following the second quality check, no new themes were conceptualized, yet several template items were updated in name and definition to more broadly describe the data.

After conducting the second quality check, the authors agreed that the template provides a sufficiently good and rich representation of the themes we identified in the data.

The final coding template is presented in Figure 4.4. We provide an audit trail of the evolution of our template in the supplemental materials. The themes presented below were synthesized through interpreting the final template, noting insightful differences and similarities between individual codes or sets of codes.

4.3.2 Themes

In this section we present the results of our thematic analysis. For each theme we describe how it relates to specific codes or groups of codes from the template in Figure 4.4 and illustrate it with examples from our data. The examples of posts and replies presented throughout the chapter are minimally edited to fix typos and remove usernames to preserve anonymity. We then offer a discussion of the implications of the relevant findings of the theme in the context of designing interventions against data-driven misinformation. To conclude, we summarize our discussion by identifying the **opportunities** that the theme presents to effectively address misinformation and describing important **limitations** of the opportunity.

4.3.2.1 Analytical Wisdom of the Crowds

Based on our thematic analysis, we identify evidence that online crowds can and do reason about the accuracy or misleadingness of data visualization posts and analytically engage with the data and its interpretation. As seen from the subitems in code 2.1 in the final template in Figure 4.4, we observe six ways in which the audiences analytically assess the data interpretations in their response: sharing personal anecdotes or lived experiences that add context to the data (2.1.1. Anecdote), providing more data points of the same metric or a different variable (2.1.2. More data), highlighting important statistical or methodological caveats (2.1.3. Caveat), reinterpreting the original chart to underscore a different insight (2.1.4. Reinterpretation), raising awareness about the existence of more up-to-date and sufficiently different version of the data or the chart (2.1.5. Update), and generally cautioning against making strong conclusions based on limited data (2.1.6. General caution).

Notably, such analytical assessments not only serve to undermine and "fact-check" the original insight but also can be used to confirm or strengthen it, indicated by Direction codes 2.4.1. (uphold) and 2.4.2. (oppose). For instance, sharing a larger set of data points

may highlight the fact that the original visualization was cherry-picked if the trend is different, or it could provide evidence that it was not if the trend is consistent. Similarly, sharing a methodological caveat, e.g., that the recording of COVID-19 cases is delayed and thus undercounted, can weaken an insight that highlights a dip in cases but further strengthen an insight that highlights an increase.

4.3.2.1.1 Implications. Our findings describe the avenues that a crowd of nonexperts has to analytically assess the accuracy of a data interpretation on social media. In our data set we do not observe users sharing specialized domain knowledge stemming from their expertise, performing original research, or surveying existing research—which is to be expected of a majority nonexpert crowd in a fast-paced microblogging environment. Instead, users rely on their own lived experience and individual pieces of information or data already familiar to them to interpret or reinterpret the original conclusion. As a result users are likely biased by the information readily available to them.

A significant limitation is that individual lived experiences or counter-data cannot entirely disprove the original insight. Moreover, the crowd's assessments also cannot accurately estimate the extent to which a given caveat impacts the insight. For instance, the caveat that the vaccine adverse effects system (a web-platform to track adverse effects) allows unverified submission from anyone in Figure 4.5 suggests that cases of vaccine-related deaths and adverse effects are likely overcounted. However, since this caveat is merely directional and does not provide any information about by *how much* the cases are overcounted, we cannot know if the original insight still holds. Effectively, the audience's analytical assessments can be fruitful in sowing doubt and undermining trust in the original conclusion but cannot disprove it.

> **Opportunities:** Non-expert online audiences identify important and nuanced caveats in misleading data interpretations.
>
> **Limitations:** Caveats cannot fully disprove flawed data interpretations, only weaken them or sow doubt.

4.3.2.2 Debunking Is in the Eye of the Beholder

We identified an important difference between an audience agreeing with the *premise of the post* and agreeing with the *presented analysis or data interpretation*. Consequently, users

are able to find fault with the particulars of the data while still upholding the conclusion, with one reply stating: *"Yeah, this graph doesn't show that, but we get the point."* In the code book this difference is highlighted by groups of Codes 1.3. and 2.4. seen in Figure 4.4: codes in 1.3. describe the direction of trust, or whether the reply trusts the author's expertise and insight, whereas codes in 2.4. describe whether any analytical assessment strengthens or weakens this insight.

In another example, the audience proactively seeks to build on a flawed analysis they agree with by suggesting improvements: the post in Figure 4.6 attempts to highlight the effectiveness of vaccines against COVID-19 by sharing statistics of cases during a local outbreak. Numerous responses call attention to the fact that the interpretation is flawed due to base rate fallacy—the author did not share population-level statistics, only those pertaining to existing patients. Yet at the same time, most replies find it important to note that although they are pointing out this fallacy, they are in full support of vaccination and agree with the author's conclusion. One reply notes, *"I've been vaccinated. Just not one for misleading data."* At the same time, we observe explicit or implicit hesitation when commenters challenge an insight they agree with. As one respondent puts it, *"I can find holes in this graph but I won't because I want people to wear masks."*

4.3.2.2.1 Implications. This finding calls attention to an important difference between assessments of data-driven misinformation and factual statement-based misinformation. Previous work by Allen et al. finds that, in the context of factual statement-based misinformation, politically aligned users are unlikely to formally fact-check each other [35]. However, although a factual statement can be true or false, Lisnic et al. discuss that most misleading data visualization-supported arguments take the form of an inductive argument, which can be plausible or implausible [6]. As a result, it is possible to arrive at a correct conclusion even through a flawed analysis of data, and consequently it is possible to challenge the analysis without debunking the conclusion.

We still, however, observe evidence that like-minded users are sometimes hesitant to probe flawed data interpretations. This observation highlights a limitation in the crowd's ability to effectively evaluate the accuracy of data-driven insights: a large portion of a post's audience may forego their assessment of the *analysis* due to concerns about unintentionally convincing others that the *conclusion* is false. As a result, analytical assessments

are mostly submitted by users who disagree with the conclusion and attempt to attack it. Thus, submitting a flawed analysis to support a true conclusion may backfire and do a disservice to the conclusion: most replies are likely to be attacking the insight and inadvertently convincing others that it is wrong altogether.

> **Opportunities:** Users who agree with the conclusion still often point out that the analysis is misleading attempting to strengthen it.
>
> **Limitations:** Nonetheless, ideologically aligned users appear to be more hesitant to share their assessments.

4.3.2.3 What Cannot Be Fact-Checked Could Be Peer-Reviewed

Misleading data-driven insights leave few opportunities for audiences to share a statement that would, if true, prove the invalidity of the claim—or to "fact-check" it. Although fact-checking is common in cases of fact-based misinformation, visualizations insights typically take the form of data-supported hypotheses. In our analysis we identify limited cases in which audiences attempt to fact-check data-driven insights, listed as Codes 2.3.1. through 2.3.3. in Figure 4.4. In cases when the visualization is outdated, sharing new data could invalidate the original insight (2.3.1. Revisiting); in cases when the insight is true only with the addition of a nondata statement (for instance, a false claim that the FDA approved the use of a certain drug against COVID-19), that statement itself could be fact-checked (2.3.2. Fact-checking the nondata part); and lastly, some users attempt to invalidate a data-driven insight by sharing repudiating quotes and official statements from people in positions of authority, such as politicians or scientists (2.3.3. Redirect to authority figure).

Predominantly, however, misleading visualization insights in our data cannot be invalidated by a single response. As discussed in Section 4.3.2.1 and represented by Codes 2.1.1. through 2.1.6. in Figure 4.4, users attempt to contest misleading insights by sharing a single piece of counter-evidence or a caveat to the claim. In our analysis we note that although an individual user's response only provides *one piece of evidence* that often does not disprove the claim on its own, reviewing the entire conversation reveals a *variety* of independent pieces of evidence that form a consensus. For instance, the post in Figure 4.7 makes an argument that lockdowns are ineffective because Sweden—a country that did not have a strict lockdown—is experiencing a dip in cases. The responses point out

a variety of possible counterarguments: the caveat that Sweden allegedly undercounts deaths, additional data showing that Sweden has more cases than comparable Nordic countries and even than most other countries in the world, the caveat that death counts for recent dates are delayed, or personal anecdotes of locals reporting that they are still "cautious with sanitation and personal hygiene" despite a lack of formal lockdowns. Thus, a viewer is presented with vastly more evidence against the original insight than in support of it.

4.3.2.3.1 Implications. In Section 4.3.2.1 we discussed that it is typically not possible to estimate the extent to which an individual analytical assessment impacts the original insight. Evaluating the whole set of replies, however, may communicate a more complete assessment of the original claim: if multiple unrelated pieces of evidence point out the incompletenesses of the insight, it is likely that the insight is misleading. The process of individual users reviewing the accuracy of the original interpretation is akin to *crowd peer-review* or formation of a *crowd consensus* on the topic. A diverse crowd offers a wide variety of lived experience, domain knowledge, and data and statistical literacy, and contributes what they know best—usually only a single detail—to the conversation. Consequently, no single reply contains a complete assessment of the original post, but the entire conversation serves as the crowd's assessment.

Although our findings indicate an opportunity to leverage the hive-mind for a crowd peer review of misleading data interpretations, there are challenges. To be used effectively, the assessments from the entire conversation body need to be surveyed and synthesized into a *meta-review* that presents the diverse points of view. It is also necessary for the body of "reviewers" to be large and diverse, which is difficult to achieve for posts that do not go viral or authors with a highly partisan audience.

> **Opportunities:** Longer discussions of posts with misleading data interpretations cover a diverse set of caveats, counterexamples, and anecdotes.
>
> **Limitations:** To leverage the replies to (in)validate data insight, a large and diverse audience is required; and many individual assessments must be synthesized into a "meta-review" to present a complete picture.

4.3.2.4 Data Does Not Speak for Itself

Up to this point, our highlighted themes have focused on the audience's engagement with the analytical content of data interpretations. However, whereas analytical soundness of a data visualization insight is an important consideration of credibility brought up by the replies, we identify other credibility factors that exist independent of the insight itself. Codes grouped under Items 1.1. and 1.2. in Figure 4.4 describe a variety of explicit and implicit indications of trust and distrust of the author or insight shared by the replies, including trust or distrust in data integrity or data sources (1.1.3.), perceived level of data literacy or domain expertise of the original author (1.2.5.), or personal attitude about the author unrelated to the analysis (1.2.6.).

Examples in Figure 4.8 indicate that the lack of a source for the data or chart negatively affects its credibility (as one user noted sarcastically: "No source. Seems legit..."). At the same time, presence of a source a user disagrees with—whether it is "Florida and Texas" or "liberals"—can also lead to an insight being dismissed and distrusted. Furthermore, users often distrust some data visualization posts because they are aware of the fact that statistics can be presented in a misleading way, whereas others compare claims backed by data to "facts." Such replies do not analytically engage with the chart or the insight itself, pointing to the variety of credibility and trust factors beyond the content of the original post.

4.3.2.4.1 Implications. Data or its visual presentations do not exist in a vacuum but rather are entangled with the social media persona sharing it as well as the existing conspiracies and stereotypes concerning the topic of interest. Our results indicate that in many cases users exhibit such a strong sense of trust or distrust of the author or the data source that they do not feel the need to analytically engage with the data insight to decide whether they believe it.

Our findings highlight the flexible nature of using data as evidence of phenomena: although users often advocate for democratizing data, "doing one's own research," and compare data to "facts" (Code 1.1.5.), other examples indicate that being "a doctor [or] a scientist" is an important prerequisite for sharing data-driven insights. This consideration is important for effective scientific or public health communication: beyond sharing timely and accurate insights, there is a need for continuous trust-building and engagement with

the audience and transparency in data collection and processing methods.

> **Opportunities:** Users question unreliable sources and biased authors and recognize the potential for visualizations and statistics to mislead even with accurate data.
>
> **Limitations:** On the other hand, users may blindly accept flawed analyses posted by authors they trust.

4.4 Discussion and Implications

In this section we discuss our studies' findings and the implications in the context of existing work on data-driven misinformation and interventions. Based on our findings, we offer potential solutions to the limitations and challenges described in Section 4.3.2.

4.4.1 Data-Driven Misinformation

The results of our work underscore important affordances and challenges that *data-driven forms of misinformation* present. Misleading data insights fall under a category of posts that Walter and Salovich describe as a 'gray area' of statements that sound like fact-based claims but are actually unverifiable opinions" [36]. It would be unjust, however, to merely call a data-driven insight an opinion. Basing a claim in data offers a veneer of impartiality and scientific rigor, making it more believable than an opinion. And although a data-driven insight is not nearly as certain as a fact, herein lies yet another factor that makes it easier to spread misinformation: it is typically not completely verifiable.

The issues of confirmation or falsifiability of data-driven insights, such as claims of causal relationships, are of course not unique to conspiracy theories shared online. By and large, most of scientific advancements and policy decisions are based in similarly "useful-but-not-certain" data findings—albeit typically with more rigor, confirmatory experimentation, and, more importantly, an admission of uncertainty about the results. In their essay discussing the epistemology of fact-checking in the context of political science, Uscinski and Butler note that fact-checkers' attempts to assess the veracity of causal claims and predictions are futile because even after thorough research many "scientists would be hesitant to dichotomize [such claims] as true or false" [141].

In the world of scientific research, this ambiguity is typically resolved by the community of researchers reaching a scientific consensus. Before a consensus is reached, researchers merely accumulate what Kuhn describes as a "morass" of random facts and

unverified observations in hopes that something will show "significant promise for future problem-solving" [142]. Only when a community forms a *settled paradigm* can researchers perform "normal science": actually advancing the existing theory as opposed to challenging it [142]. Thus, the process of establishing a consensus is highly social and amounts to, through a period of debates, reaching an agreement that a given theory or opinion reflects a current best guess [143].

Although a best guess definitively does not equate with truth, it is useful to present it as highly certain, if not fact. In her exploration of the scientific consensus around climate change, Naomi Oreskes argues that we should treat opinions that the scientific literature largely agrees upon as facts [144]. Oreskes states that excessively communicating stipulations about the uncertainty of scientific findings—amplified by malicious actors who attempt to exaggerate the level of uncertainty—has resulted in general inaction on a variety of topics, such as anthropogenic climate change and the dangers of smoking [144], [145]. As a result, the public severely underestimates the high level of agreement among scientists on a number of seemingly controversial topics, resulting in policy paralysis: oftentimes, scientific consensus is followed by decades of inaction—and the associated costs of inaction—until an idea becomes publicly accepted as fact [146].

It follows that *data-driven misinformation* is most effective at exactly that: forming an illusion of scientific debate and sowing doubt in the existence of actual scientific consensus on a topic. The results of our quantitative study show evidence that data-driven insights with reasoning errors do incentivize such debates by attracting, on average, 60% more engagement that lasts 23% longer. Although these insights based on logical fallacies and spurious correlations may not always succeed in convincing the audience of their claim and forming a new dominant scientific paradigm, they may be persuasive enough in showing that science is not settled on a given topic.

4.4.2 Designing Interventions Against Data-Driven Misinformation

Based on the above, we argue that, in designing interventions against data-driven misinformation, platforms should be especially cognizant of considerations about data-driven insights being presented as and treated as facts or opinions. In his article discussing the role of facts in modern data-driven discourse, Sun-ha Hong [147] argues that the term *fact*

is being overused and mythologized. Specifically, Hong identifies two common practices: *fact signaling*, or performative invocations of facts to discredit rivals and create an "evidence theatre" with data as props, and *fact nostalgia*, an imagined past when "facts were facts." Taken together, these two practices are commonly utilized by actors who spread misinformation to not only present data that support their arguments as facts but also through this process to evoke nostalgia for a mythologized past in which the society had a mutual understanding of what is true and what is false. Consequently, presenting caveats to data-driven insights as fact-checking may have the unintended effect of perpetuating fact signaling and endorsing a dichotomized world that lacks nuance and in which data is either true or false. Uscinski and Butler [141] similarly argue that "[fact-checking] practices share the tacit presupposition that there cannot be genuine political debate about facts, because facts are unambiguous and not subject to interpretation." By being a partial and imperfect representation of phenomena [95], data is often inherently ambiguous and requires contextual knowledge for an accurate interpretation. Hence, instead of presenting a rebuttal as fact, interventions against data-driven misinformation should communicate the ambiguous nature of data by highlighting the limitations of data-driven reasoning and the considerations in attempting to model complex real-world phenomena.

At the same time, if we avoid appealing to facts, we should be careful to not uphold the illusion of the existence of debate and lack of scientific consensus on many scientifically settled controversial topics, such as anthropogenic climate change and vaccine safety. This is a difficult balancing act that involves making a decision about which topics have or do not have scientific consensus. Ways of determining (and proving) the existence of consensus can range from examining literature surveys, consensus conferences [148], and publications such as Cochrane Reviews [149] to data-driven approaches that quantitatively estimate convergence in a network of scholarly literature [146]. We note that in our study we did not observe users attempting to appeal to scientific consensus. This finding could be, to an extent, influenced by the fact that COVID-19 is a novel virus, many aspects of which were, and still are, scientifically inconclusive. To our knowledge, however, existing credibility assessment interventions on social media platforms do not offer a way to raise awareness about scientific consensus, and instead confine the user to a dichotomy of factual correctness that may be confusing in this context. We argue that **the option to**

appeal to and cite scientific consensus should be a salient suggestion in the platform's misinformation reporting interface and not make a user decide whether, for instance, anthropogenic climate change is a fact or an opinion.

Our study shows evidence that online crowds do actively attempt to correct data-driven misinformation and are most effective at identifying and highlighting nuances and counter-examples to data insights. We argue that interventions against data-driven misinformation should leverage the strengths of the crowd, and to do so effectively they should address the limitations we outlined in Section 4.3.2. Specifically, to account for the fact that an individual caveat outlined in a reply is not sufficient to disprove a claim, platforms **should support the creation of meta-reviews of data insights** that summarize the multitude of nuances described by the entire audience. These reviews could be compiled manually by a moderator, by leveraging natural language processing techniques, or through interventions that assist collaborative judgements [150]. Additionally, platforms should **encourage users to share their suggestions for improvements of data interpretations they agree with** to counteract the potential of a backfire effect of flawed analyses in support of true claims. Platforms should also encourage users to share counter-analyses of data as a way of correcting misleading insights by **showing that the opposite conclusion is more strongly-supported**, and go beyond simply pointing out inconsistencies of the original insight.

Besides incentivizing "good data work" and disincentivizing "bad data work," we acknowledge the existence of important credibility indicators of data-driven insights that go beyond the accuracy of the analysis. Based on our findings, we argue that content creators—especially government- and domain-expert-run accounts—should actively work to **build trust in their data and presentation** by being transparent about data sources and collection methodologies and forthright about important data processing decisions. Since conversations surrounding posts with data-driven insights last more than twice as long as those for other visualization posts, expert accounts should communicate these details by continuously **engaging with the community** and directly addressing concerns raised about the trustworthiness of their insights.

In summary, our overarching recommendation for designing interventions is recognizing data-driven misinformation as a unique and nuanced threat to the integrity of

our information space. Misleading data-driven insights undermine the public's trust in scientific findings and promote harmful misinformation while—by the virtue of straddling the line between facts and opinions—remaining largely unaddressed. Through raising awareness about the nuanced spectrum of weak and strong evidence of phenomena, we can tackle the issue of false dichotomies that a claim can only be either fact or opinion or either true or false.

4.5 Limitations

Our work is subject to several limitations. Firstly, our data set consisted of content from one platform—Twitter—and thus our findings are influenced by the platform affordances. For instance, character length limits of posts and replies have the potential to limit the amount of detail users share in a single tweet. Additionally, Twitter does not have a variety of features common in message board-type social media sites that could be used to moderate caveats to data-driven insights, such as mega threads or reply pinning. Secondly, our analysis is limited to posts related to the COVID-19 pandemic. Although the initial outbreak of COVID-19 generated a large amount of rich data-driven discussions online, it is also a unique event that featured a lack of existing research on the topic and a high level of politicization. We believe that although such events happen rarely, studying the ways to mitigate the spread of misinformation during them is of utmost importance.

4.6 Conclusion and Future Work

In this chapter, we presented an analysis of the count, duration, and content of engagement with misleading data visualizations on social media. We hope our work inspires future research to formally study the distinct ways in which data-driven misinformation is generated, spread, and, we hope, corrected. Future work should investigate the impacts of platform affordances on the data-driven discourse by considering other social media sites, as well as the opportunities to address misinformation on various other data-driven topics, such as anthropogenic climate change and vaccine hesitancy. Additionally, future research should identify relevant factors that foster analytical assessments of data-driven insights in a post's discussion beyond the presence of a large and diverse audience.

Table 4.1: A summary of metrics used to evaluate and compare engagement count model specifications. We compared the fit of Zero-Inflated Negative Binomial (ZINB) and that of regular Negative Binomial (NB) using log-likelihood (LL), Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), and Mean Absolute Error (MAE) metrics. We highlight the most accurate performing model by each criterion (lower is better). As seen from the table, the metrics suggest that the Zero-Inflated version of the model provides a better fit for replies, but we observe mixed results for other metrics.

Model	df	LL	AIC	BIC	MAE
Replies					
Zero-Inflated NB	35	**-25,686**	**51,441**	**51,689**	**32.84**
NB	18	-25,908	51,852	51,980	33.32
Retweets					
Zero-Inflated NB	35	**-39,004**	**78,078**	78,327	98.13
NB	18	-39,078	78,193	**78,321**	**98.12**
Quotes					
Zero-Inflated NB	35	**-23,527**	47,124	47,372	**14.93**
NB	18	**-23,527**	**47,090**	**47,218**	**14.93**
Likes					
Zero-Inflated NB	35	**-127,049**	**254,168**	254,447	481.00
NB	18	-127,111	254,258	**254,402**	**480.87**

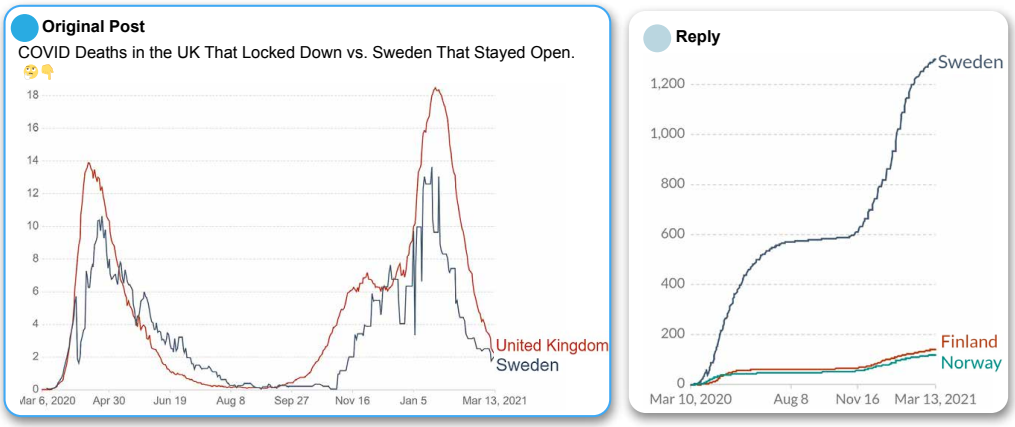


Figure 4.1: An example of a post with a data visualization-supported insight and an associated reply. The author of the original post shares a chart comparing COVID-19 deaths in the UK and Sweden, making an argument that lockdowns are not effective because the case curves of the two countries are similar. The reply makes a counter-argument (without the use of text) by sharing an analogous data visualization that supports the opposite conclusion, which shows that Sweden had many more cases than Finland and Norway, comparable Nordic countries.

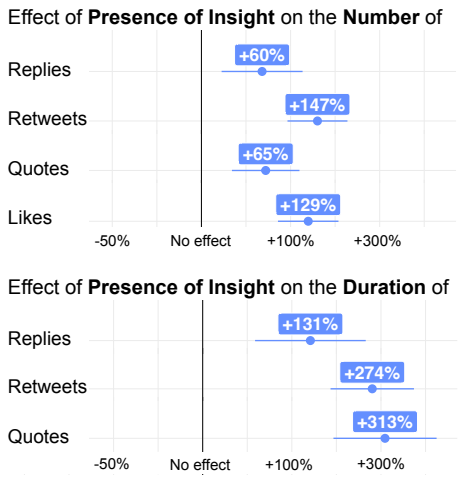


Figure 4.2: Average effect of presence of insight (compared to no insight in a post). Regression coefficients were estimated controlling for presence of reasoning error, effects of which are presented in Figure 4.3. We show 95% confidence intervals of estimated effect size of variable on count and duration. Estimated effects are calculated as $e^{\beta} - 1$, where β is the output regression coefficient. We observe that the presence of an insight in a post is associated with a higher number and longer duration of engagement.

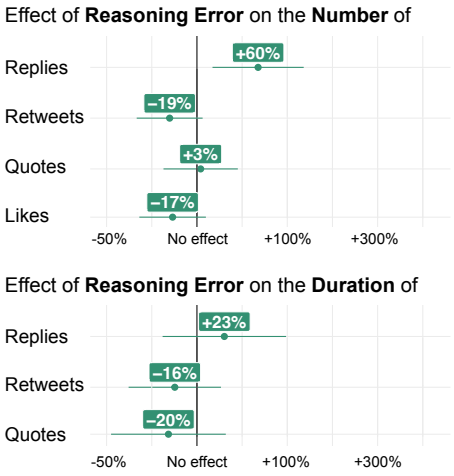


Figure 4.3: Average effect of presence of reasoning error in insight (compared to insight with no reasoning error). Regression coefficients were estimated controlling for presence of insight, effects of which are presented in Figure 4.2. We show 95% confidence intervals of estimated effect size of variable on count and duration. Estimated effects are calculated as $e^{\beta} - 1$, where β is the output regression coefficient.

1. **Sentiment**
   - **1.1. (Dis)trust in insight**
     - 1.1.1. Explicitly or implicitly agreeing
     - 1.1.2. Suggesting a conspiracy
     - 1.1.3. (Dis)trust of data/source
     - 1.1.4. (Dis)trust of statistics/visualization
     - 1.1.5. Appeal to facts
     - 1.1.6. Sharing (by quoting or tagging)
     - 1.1.7. Meme/joke
     - 1.1.8. Mocking caveats
   - **1.2. (Dis)trust in poster**
     - 1.2.1. Asking for advice/elaboration
     - 1.2.2. Asking for more/updated data
     - 1.2.3. Asking for source
     - 1.2.4. Gratitude/respect
     - 1.2.5. Lack/presence of expertise
     - 1.2.6. Personal attacks
   - **1.3. Direction**
     - 1.3.1. Trust
     - 1.3.2. Distrust
2. **Content**
   - **2.1. (Quasi)analytical**
     - 2.1.1. Anecdote
     - 2.1.2. More data
     - 2.1.3. Caveat
     - 2.1.4. Reinterpretation
     - 2.1.5. Update
     - 2.1.6. General caution
   - **2.2. Citations**
     - 2.2.1. None
     - 2.2.2. Visualization
     - 2.2.3. Raw data
     - 2.2.4. Article
     - 2.2.5. Authority figure
   - **2.3. Attempts to fact-check**
     - 2.3.1. Revisiting
     - 2.3.2. Fact-checking the non-data part
     - 2.3.3. Redirect to authority figure
   - **2.4. Direction**
     - 2.4.1. Uphold/strengthen insight
     - 2.4.2. Oppose/weaken insight

Figure 4.4: Final template used to describe the data and conceptualize themes. The codes under 1. Sentiment describe users' trust in the poster or in the general sentiment of the post. The codes under 2. Content describe the replies' analytical engagement with the data and visualization.

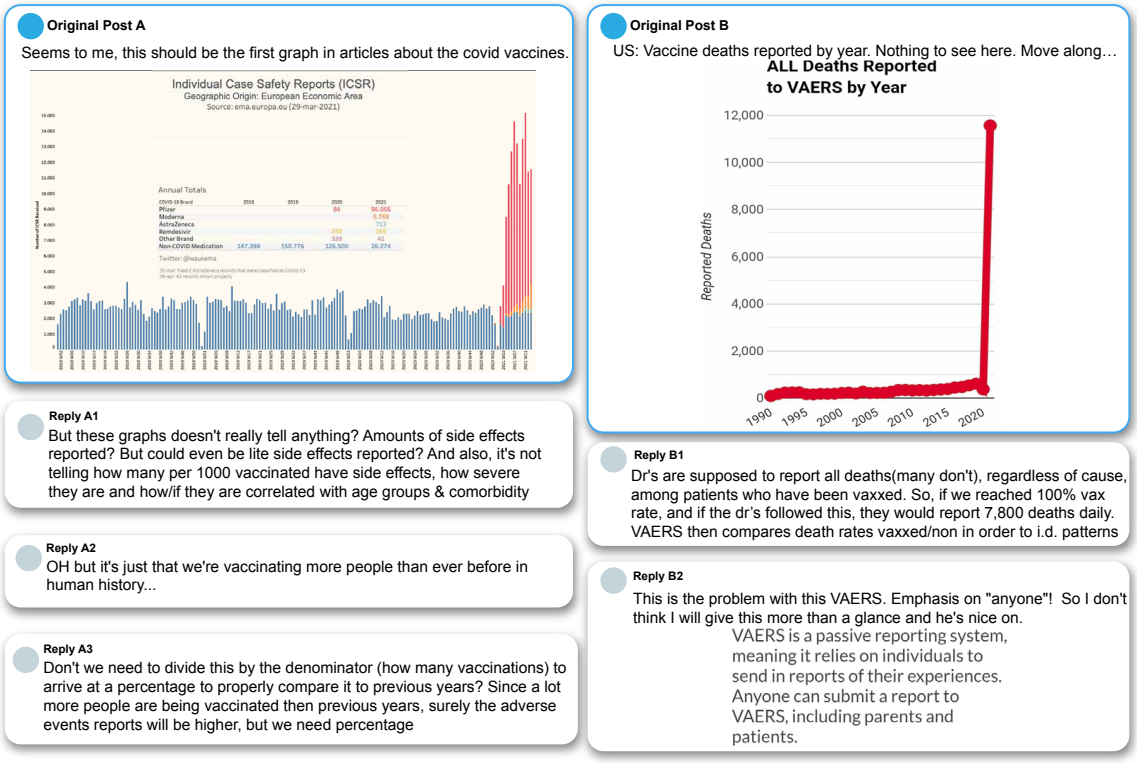


Figure 4.5: Example posts using similar data—EU's ICSA and US's VAERS adverse effects tracking systems—to spread skepticism around safety of vaccines. These databases have been widely misinterpreted by antivaccine activists to promote their views [151]. The responses point out caveats in the interpretation, such as the need to account for the fact that there are mass vaccinations underway, and data limitations like the lack of concrete definition of "adverse effect" and, most importantly, the fact that the submissions are not verified and can be submitted by anyone.

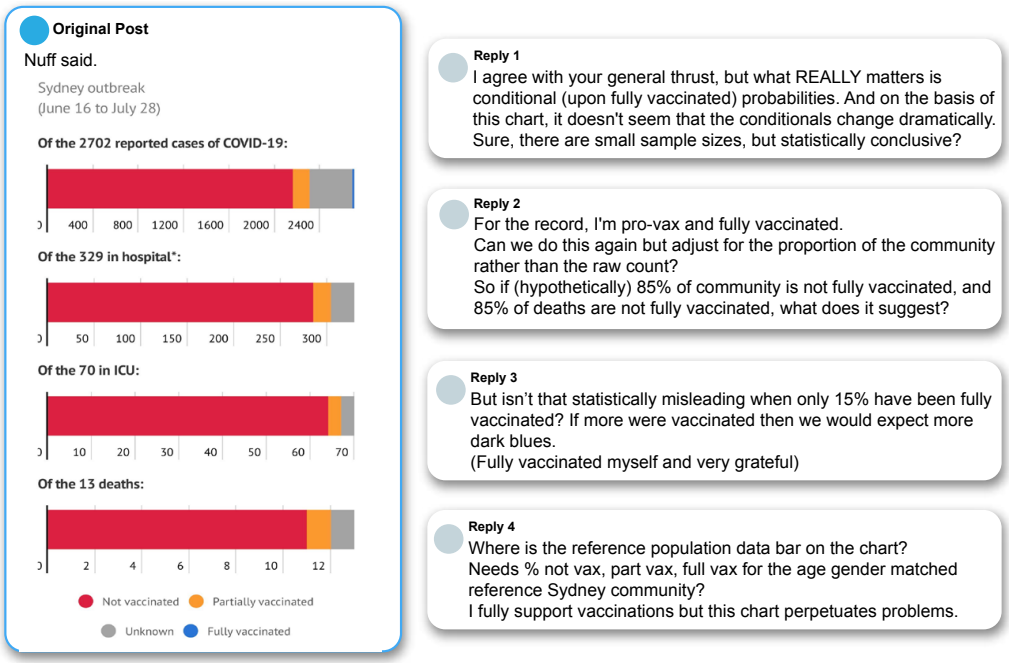


Figure 4.6: Example post with replies showing the types of analytical responses from the crowd. The responses are predominantly agreeing with the conclusion, yet still point out flaws in the data interpretation.

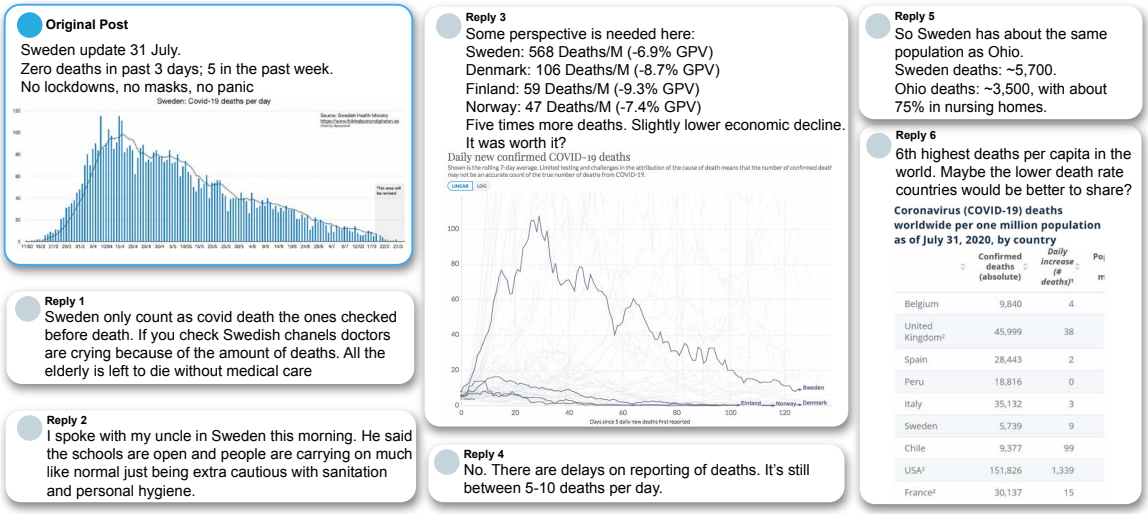


Figure 4.7: Example post where the author promotes the idea that COVID-19 containment measures, such as masking, are ineffective citing the data that shows death per day going down in Sweden. The replies to the post showcase the types of analytical responses from the crowd that challenge the accuracy and generalizability of the author's conclusion: sharing of more data, caveats, up-to-date data, and personal anecdotes.

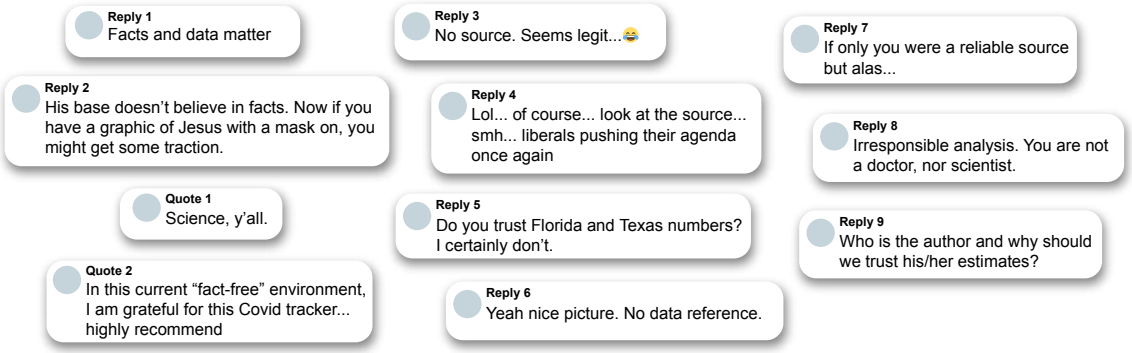


Figure 4.8: Examples of replies to and quotes about a variety of different posts with data-driven insights that indicate sentiment toward the author or the insight without analytically evaluating the insight. Examples include replies that trust data insights because they are based on "facts", or replies that distrust data insights because of their doubts about source validity or the author's expertise and credentials.

CHAPTER 5

VISUALIZATION GUARDRAILS: DESIGNING INTERVENTIONS AGAINST CHERRY-PICKING IN INTERACTIVE DATA EXPLORERS

The growing popularity of interactive time series exploration platforms has made data visualization more accessible to the public. However, the ease of creating polished charts with preloaded data also enables selective information presentation, often resulting in biased or misleading visualizations. Research shows that these tools have been used to spread misinformation, particularly in areas such as public health and economic policies during the COVID-19 pandemic. Post hoc fact-checking may be ineffective because it typically addresses only a portion of misleading posts and comes too late to curb the spread. In this work, we explore using visualization design to counteract cherry-picking, a common tactic in deceptive visualizations. We propose a design space of *guardrails*—interventions to expose cherry-picking in time-series explorers. Through three crowd-sourced experiments, we demonstrate that guardrails, particularly those superimposing data, can encourage skepticism, though with some limitations. We provide recommendations for developing more effective visualization guardrails.

This content of this chapter is drawn directly from M. Lisnic, Z. Cutler, M. Kogan, and A. Lex, "Visualization guardrails: Designing interventions against cherry-picking in interactive data explorers," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2025, pp. 1–19. doi: 10.1145/3706598.3713385 [8]. It is included in the dissertation with the permission of my coauthors, Zach Cutler, Marina Kogan, and Alexander Lex.

5.1 Introduction

Open data exploration platforms democratize access to data and visualizations of public importance. Examples include COVID-19 case dashboards on OurWorldInData [92], stock or cryptocurrency performance charts on Yahoo! Finance [152], or graphing platforms for various economic and policy indicators, such as Gapminder [153] or FRED [154]. Users not only create charts on these platforms for their own use, but also commonly share them on social media. While the popularity of data exploration platforms is a testament to their utility, the unconstrained and often unguided selection of data subsets and time periods that is commonly featured in such tools can lead to conclusions made based on *cherry-picked data*. Such data visualizations, when shared on social media, are often misrepresented in a way that supports wrong and, in the worst case, harmful arguments. Prior research has also shown that cherry-picking of items and time frames is an issue that contributed to misinformation arguments in over 40% of COVID-skeptic charts shared on Twitter, most of which were screenshots of data explorers [6].

It is important to acknowledge that the problem of cherry-picking in public-facing data explorers is a *wicked problem* [155]. Biased interpretations of data, and information in general, are entangled with the viewers' data and visualization literacy, data production quality, social and political movements, or the sense of belonging to a group with a strong ideology. This problem is hence wicked in the sense that, due to its complexity, it is resistant to any single solution. Rather, it requires continuous adaptation and innovation of approaches and interventions. In this work, we propose adding a novel type of intervention to the arsenal of tools against data-driven misinformation.

The most prevalent interventions for addressing misleading information focus on fact-checking, whether conducted by crowdsourcing or professional organizations [20]. However, these post hoc strategies come with significant limitations. Correcting every misleading post is nearly impossible, and unchecked misinformation can result in the implied truth effect, where uncorrected content is perceived as accurate [28]. Furthermore, because post hoc corrections and moderation inherently occur after the fact—sometimes delayed by several days [156]—the original misinformation often spreads widely before any correction can be made. Research also suggests that corrections are more effective when delivered by the same source that initially presented the misinformation [157] as op-

posed to other sources that may be dismissed as biased, indicating the value of enhancing the original content with contextual information. Prior research has found that data-driven misinformation arguments on social media, including cherry-picked data visualizations, often remain unaddressed by existing fact-checking interventions due to being rooted in factually accurate data and thus not being wrong, or viewed as personal opinions not worthy of debunking [7].

In this chapter, we set out to explore the possibility of ante hoc interventions tackling cherry-picking in time series explorers from the perspective of visualization design, attempting to curb the issue before it has a chance to spread. We asked ourselves: **is there anything we can do when designing and implementing data exploration platforms to minimize misuse, misunderstandings, and misinterpretations?** To answer this question, we propose looking at the problem of misinterpretation and misuse of data explorers through the lens of threat modeling [158], and explore the possible interventions against cherry-picking in interactive visualizations. As one such intervention, we introduce the concept of visualization guardrails. Analogous to protective highway guardrails, which are useful when losing control of a vehicle but no hindrance in regular operation, **visualization guardrails automatically show the contextual data that expose cherry-picking, but do not interfere with interpretation if no cherry-picking is present**. We describe the ways contextual data could be visualized and integrated into existing charts by outlining the design space of guardrails.

We break down the issue of misinformative charts into two distinct yet equally important problems. Firstly, the *production problem*: data exploration interfaces make it very easy to create (perhaps, even nudge authors toward creating) cherry-picked views. The circumstances that may lead an individual to creating a cherry-picked chart can vary from a desire to maliciously misinform, to innocent ignorance, or even to a genuine reason to be interested in zooming in on a certain subset of data. In any case, however, it should be more difficult to end up with a potentially misleading view.

The second problem is the *reaction problem*: when the resulting view is then shared with others through social media, it may end up spreading the incorrect insight and convincing others. Moreover, charts created with data exploration platforms are typically adorned with a veneer of impartiality and reputability offered by the data exploration platform's

logo and recognizable design, and hence may seem authoritative [7]. For an example, see the cherry-picked charts in the style of the Financial Times and Our World in Data in Figure 5.1, both reputable sources. Therefore, another major design goal for the guardrails is to introduce more nuance to views created with such reputable platforms.

To restate, our research questions are:

- What is the **design space** of guardrails against cherry-picking in data visualizations?
- Can we design guardrails that make cherry-picked charts **harder to produce**?
- Can we design guardrails that make cherry-picked charts less convincing and lead to a **more skeptical reaction** from the audience?

In this work, we describe the design space of visualization guardrails and implement a prototype data explorer with four distinct guardrail instances. We limit our designs to line charts showing time-series data, which is a type of chart that is often used in debates on controversial topics (such as impacts of policy decisions on metrics of interest). Additionally, time-series line charts have been commonly used for cherry-picking in a social media context [6]. In order to investigate both the production and the reaction problems, we conducted three crowd-sourced experiments using our prototype of guardrails that target item cherry-picking, in which we attempt to mimic the real-world adversarial process of designing and sharing misinformative charts. In the first experiment, we challenged participants to create cherry-picked views using control and guardrail interfaces, thus red teaming our prototype. We then showed these charts to another set of participants in the second experiment and asked them to make a behavioral trust decision based on the chart, evaluating the resulting cherry-picking. We then conducted a third experiment to confirm our findings in a more controlled scenario in which we regulated the effect of cherry-picking egregiousness and misleading captions on guardrails. Our findings show that the **guardrails achieved the strongest desired effect of encouraging skepticism when guardrails are closest to the original visual language of the chart**. At the same time, guardrails are less effective when merely providing a clue about cherry-picking by showing a statistical summary or being plotted beside the main canvas. In such cases, a large part of the audience ignored the guardrails and instead focused on the main features of the visualization and the attached caption. Based on our results, we outline recommendations for designing effective guardrails against cherry-picking in data explorers.

To summarize, this chapter makes several contributions:

- A novel conceptual framework for tackling issues of misuse and misinterpretations of data visualizations through the lens of threat modeling.
- A definition of the design space of visualization guardrails that protect against cherry-picking.
- A set of crowd-sourced user studies exploring the effects of guardrails on the production of and reaction to visualizations of cherry-picked data, and the resulting recommendations for implementing visualization guardrails in public-facing data explorers.

5.2 What Is Cherry-Picking?

In his testimony before the US House of Representatives, climate scientist Richard Somerville described cherry-picking as "[making] selective choices among competing evidence, so as to emphasize those results that support a given position, while ignoring or dismissing any findings that do not support it" [159]. By analogy, in the context of visualization we define cherry-picking as **plotting data that support a given position, while not plotting data that do not.**

Cherry-picking favorable data or results has long been acknowledged as a questionable research practice across the scientific community, alongside HARKing (hypothesizing after the results are known) and p-hacking [160], [161]. Cherry-picking, or selectively showing only information that supports a given argument, is an outcome that could be unintentional and point to ignorant or inattentive practices, or be intentional and reflect malicious intent to misinform. **Unintentional cherry-picking** is a common behavior in judgment under uncertainty and could stem from a variety of cognitive biases, such as availability bias (focusing only on information that is readily available) [162],confirmation bias (focusing on information that supports prior beliefs) [163], and anchoring bias (focusing on information presented first) [164]. **Intentional cherry-picking** is a strategy shown to support misinformation arguments about climate change denial [107], vaccine hesitancy [165], and is, more broadly, one of the hallmark tactics of denialism [166].

Cherry-picking is a practice that is not limited to data and visualizations. Quote-mining refers to cherry-picking a quote or a statement out of its original context [166]. Quoting

others out of context has been practiced for a variety of purposes over the course of history, ranging from political propaganda [167] to misusing critics' quotes in misleading advertisements [168]. Both cherry-picking data and quote-mining statements can be described as sharing half-truths, meaning these tactics are difficult to debunk since they are based in truth [169]. It then follows that in order to combat the issue of focusing on a subset of information, an intervention must either raise awareness about or explicitly show or summarize the missing context.

5.3 A Threat Modeling Framework for Visualizations

In this section, we propose adopting the concept of *threat modeling* from computer security as a way to approach the problem of data visualizations misused in support of misinformation. By applying the threat modeling framework to cherry-picking, we demonstrate that threat modeling is useful both as a mental model for surfacing vulnerabilities in data interfaces and as a practical guide for identifying appropriate interventions. We believe that this model is applicable to a variety of other problems as well, such as misinterpretations of statistics or incorrect causal inferences, and we urge researchers to explore such applications in future work.

5.3.1 What Is Threat Modeling?

The Threat Modeling Manifesto, put forward by a group of security researchers aiming to promote security and privacy during software development, defines threat modeling as "analyzing representations of a system to highlight concerns about security and privacy characteristics" [170]. Threat modeling allows the researcher to identify things that can go wrong, pinpoint design issues, and inform mitigation measures.

In his book *Threat Modeling: Designing for Security* [158], Adam Shostack proposes a four-step framework for threat modeling, which involves answering four questions: (1) What are you building?, (2) What can go wrong?, (3) What should you do about it?, and (4) Did you do a decent job?

Example outcomes of threat modeling in the security context could be identifying opportunities for hackers to access confidential information in a data base, or designing interventions against denial-of-service attacks. However, although a number of specific

computer security-related checklists or domain-specific heuristics threaten modeling itself, it is a value- and principle-driven approach. Therefore, it is highly adaptable to a wide variety of security and privacy issues, as well as (as we will argue below) issues of data and visualization misuse.

5.3.2 Applying Threat Modeling to Visualization Threats

Next, we go over the four steps of the framework with the goal of outlining the issue of misuse of interactive visualizations used by the general public to support misinformation, as well as motivating the application of threat modeling to this issue.

5.3.2.1 What Are You Building?

For the purposes of this work, we focus on general public-facing interactive time series exploration portals created by local government agencies such as state and county epidemiology tracking dashboards, news organizations such as Yahoo! Finance [152], or specialized data exploration platforms such as OurWorldInData [92]. In this step, the data explorer platform governance should outline the core functionality and values of the platform that should be protected and not compromised during the threat modeling exercise. We assume that examples of such core functionality could include instant access to data for everyone, freedom of exploration without major restrictions, ability to export and share views and data, among others.

5.3.2.2 What Can Go Wrong?

To answer this question, we can turn to previous work that outlined reasoning errors in social media users' interpretations of data visualizations [6]: cherry-picking favorable subsets of data, assigning causality to salient features of charts, or not accounting for common statistical fallacies. In this work, we focus on *cherry-picking*—one of the most often used tactics—as an illustrative example. Some of the core functionality in data explorers described in the previous section, such as access to data and freedom of exploration, result in the danger of making selective choices and emphasizing those results, pointing to a possibility of cherry-picking. Figure 5.1 shows two instances of tweets using cherry-picked data explorer charts to spread misinformation and to promote vaccine hesitancy.

5.3.2.3 What Should You Do About It?

Shostack outlines four possible paths of action: (1) *accept* that there is an issue and do nothing, (2) *eliminate* the feature causing an issue, (3) *transfer* the responsibility to the user, or (4) *mitigate* the issue [158]. All four strategies are feasible for data explorer platforms. An example of acceptance is simple: one could do nothing. Eliminating a feature could take the form of restricting problematic interactions by, for instance, not allowing plotting any two time series on the same chart. Transferring responsibility to the user may involve requiring data or visualization literacy evaluations, or a checklist that assists a user in evaluating whether their selection is misleading, but may be unrealistic to realize in platforms accessible to the general public.

Although we urge platform designers to consider all of the above strategies when performing threat modeling of their tools, in this study we will pursue the goal of *mitigation* of the cherry-picking threat and design *guardrails* against it. We argue that of the four possible actions, mitigation is the ultimate goal of misinformation interventions. Firstly, acceptance, or inaction against misinformation is not productive. Secondly, we believe that the benefits of the "problematic" features (i.e., unrestricted freedom of exploration and ability to take screenshots and share) are high, and therefore we opt to explicitly maintain such features and not pursue elimination. And lastly, in the adversarial context of misinformation, we posit that any transfer of responsibility strategy would be futile due to the high possibility of bad-faith actors purposefully seeking out cherry-picked views.

5.3.2.4 Did You Do a Decent Job?

The goal of this last step of threat modeling is to evaluate the success of the chosen intervention strategy. In order to answer this question, we conducted three rounds of evaluation studies, described in Sections 5.6, 5.7, and 5.8. Firstly, we conducted a study challenging participants to produce cherry-picked views with and without our guardrails in an approach similar to using a Red Team in cybersecurity. This enabled us to conduct a second experiment in which a different set of participants were asked to review the charts produced in the first study. Lastly, we conducted a third, more controlled experiment, modulating the severity of cherry-picking in the chart.

5.4 Designing Guardrails

In this section we describe our approach to characterizing the design space of visualization guardrails and outline the specifics and variations of the resulting design space.

5.4.1 Design Process

To develop a broad set of ideas, we engaged in a parallel prototyping process, where each of the authors independently developed multiple designs [171]. Before the start of the design process, the authors put together the design brief. We agreed that the main goal of the designs would be to provide missing context and promote skepticism in the viewer [72]. Additionally, the designs should be:

- **Nonobtrusive**: there should be no restrictions on exploration and selections, and every commonly allowed selection must be as visually salient with a guardrail as without. This goal stems from the fact that we target mitigation, rather than elimination of a feature (Step 3 of threat modeling).
- **Undemanding**: the guardrail should not directly slow down the user with pop-ups, questionnaires, or assessments. Making an author complete evaluations would more closely resemble transfer of responsibility (Step 3 of threat modeling).
- **Tamper-evident**: it should be difficult to get rid of the guardrail, for instance by cropping a screenshot. This goal is motivated by previous work [6] that showed that most of misleading charts shared on Twitter were screenshots using various levels of cropping.

In order to come up with initial guardrail designs, each of the authors independently created sketches based on the requirements. The sketches were either free-form or on top of examples of cherry-picked views, as identified by previous work on misleading visualizations [6]. We provide all of our sketches in *supplemental materials*. The first author then reviewed and organized the sketches into common themes and ideas, and all authors discussed the results and used them to describe the possible design space. Figure 5.2 presents an overview of the resulting design space along two dimensions: **context**, or *what* is shown, and **layout**, *where* it is placed.

5.4.2 Design Space

Given our design requirements and the problem at hand, we identified that the task of designing a guardrail against cherry-picking in a data exploration platform is related to presenting a helpful *visual comparison*. Gleicher et al. described three ways objects could be visually compared: by superimposition, by juxtaposition, or by explicit encoding of differences [172], [173]. Explicitly encoding a difference would, however, involve using a different representation of the selected items. For example, when comparing COVID cases in two countries, explicitly encoding differences could entail creating a derived dimension that subtracts the cases of the countries, and visualizing this derived dimension instead of the original data. As a result, this approach limits the saliency of selections (the original data is no longer shown) and violates our goal of nonobtrusiveness. However, the other two visual comparison strategies—*superimposition* and *juxtaposition*—fit our design requirements and describe where the guardrail could be placed.

The visual comparison strategies describe the **layout** of the guardrails, or where the guardrail is shown. The other dimension of the resulting design space is **context**, or what is being shown. We categorize our design sketches into two types of context: the guardrail can either show *primary data* in the same units, level of aggregation, and visual language as the main data, but potentially sampled to a small set of items or *visual summaries*—transformations and aggregations of the data, or additional data that provides a summary context (e.g., a market index for stock data).

5.4.2.1 Layout

We describe two main types of guardrail placement: superimposition and juxtaposition, as illustrated in Figure 5.2. **Superimposed** guardrails exist on the same canvas and scales as the main chart data. A designer of a Superimposed guardrail has two main decisions to make. The first one is defining what contextual data should be shown as the guardrail for the given scenario and domain. For instance, to tackle an instance of item cherry-picking, a COVID-19 data explorer showing Sweden's cases would probably plot other Nordic countries, such as Denmark, Norway, and Finland as well. The second design decision to consider is the treatment of the vertical scale: specifically, whether the axis should be adjusted or not in cases when the contextual data would go above or below the

original frame. Examples of this design variation can be seen in Figure 5.3. The axis could be zoomed out to include all contextual detail, thereby sacrificing saliency or detail of the main selection. Alternatively, the scale could be kept as is, and instead would include a visual indication that there is out-of-frame context.

One of the main advantages of the Superimposed guardrails is that, by virtue of being plotted together with the main data, they are not just tamper-evident but virtually tamper-proof: it would be very difficult to remove the guardrails from the view. Additionally, Superimposed guardrails exist on the same scale and units as the main data and offer an opportunity to directly compare the values of the main chart data and the contextual data.

The disadvantages of Superimposed guardrails include the danger of overplotting—a lot of contextual information in the main frame of the visualization may not scale with many items selected. This problem could be alleviated by dynamically adjusting the size of the comparison set as an author chooses more items. Another problem is that it might not always be obvious which data items or time frames are useful "important context." A system could leverage metadata (such as regions of the world for country data, or sectors for financial data) to make such a determination. A generic implementation that does not require additional data could leverage statistical information, i.e., by including a set of representative examples of the data set.

The idea of the **Juxtaposed** layout of guardrails is to leave the main canvas of the visualization unchanged. Instead, we provide contextual information in a separate, juxtaposed view to the side, above, or below the chart. When designing Juxtaposed guardrails, the decision of their placement depends on the underlying data and target issue. For instance, if the goal of the guardrail is to give an indication that a climate change-denying author chose a small fraction of the time frame of ocean temperature data, the guardrail could run along the "problematic" dimension—in this case, the horizontal axis right below the chart (as seen in the example on the left in Figure 5.4).

In Juxtaposed guardrails, the decision about what constitutes important context data is less central than in Superimposed guardrails. Specifically, juxtaposition allows the designer to show *all* of the data points that would fit into the frame. Similarly, a Juxtaposed guardrail is easy to combine with other guardrails, since, unlike with Superimposed guardrails, there is no issue of overplotting the main chart.

However, Juxtaposed approaches are croppable and thus not tamper-evident (a design goal). Implementations could use strategies to reduce the croppability, for example, by embedding a Summary directly into the axis, such that cropping out the guardrail would also involve cropping out part of the axis. Alternatively, the designer could wrap the chart and the guardrails with a frame that could be indicative of parts left out. In general, however, even croppable designs could provide a degree of protection, as it would be much easier for an online audience to highlight cherry-picking if they could point to the full chart in replies or community notes.

5.4.2.2 Context

We distinguish between two types of guardrails' contextual data: they could show Primary Data of the same type as the main chart data, or they could use aggregated or transformed data in the form of a Summary.

Primary Data guardrails provide contextual data by directly plotting the primary data—meaning data at the same level of granularity and of the same units as the main data in the chart. Primary Data guardrails are shown in the top row of Figure 5.2. The main advantage of Primary Data guardrails is their simplicity (both to implement and to understand): while they should be visually distinct from the main data, they still use the units as the chosen items and need virtually no explanation to be understood. This advantage is especially strong in a Superimposed Primary Dataguardrail: it utilizes the same visual encoding and exists on the same axes as the main data, and thus its meaning is self-explanatory.

Summary guardrails, on the other hand, condense the primary data into a summary form for the purposes of providing context. They are shown in the bottom row in Figure 5.2. There are many possibilities for the exact implementation of Summary guardrails, and, consequently, the designer should deliberate over what type of summary is the most meaningful, given data and domain. For instance, a simpler Summary could provide the average, interquartile range, or extrema of the data set. A more complicated Summary could show information about the variance or the shape of the distribution of the data.

Compared to Primary Data guardrails, the Summary guardrails are more compact and help mitigate the problem of overplotting by consolidating all of the contextual data into

a single visual representation. On the downside, however, Summary guardrails may be more complex to understand as they represent a departure from the visual encoding of the main chart and may involve an uncommon visualization type. Therefore, it would be advisable to train the viewer to read the chart using annotations or training modules.

5.4.2.3 Visual Techniques and Implementations

As alluded to in the previous sections, there are a number of design alternatives and decisions that one should consider when designing what the guardrails look like and what data they are composed of. Specifically, while our design space describes the general affordances of different guardrail types, each type of guardrail type could be realized using different visual encodings. In this section, we present an overview of specific implementation variations and related visualization techniques from prior work.

The **Superimposed Primary Data** guardrail is arguably the simplest: as it involves automatically plotting a subset of contextual data, there are few design choices to be made. The key decision in this guardrail design involves defining a contextual data set that is effective yet not overly large, akin to determining the optimal set of forecasts to display in a Multiple Forecast Visualization. [174]. We observe two simple implementations of this technique in data explorers online. First is the scatter plot on Gapminder [153] that automatically includes translucent data points for the entire available data set; however, the opacity of the context is adjustable and could be completely removed. The second example is Google Search: when looking up a macroeconomic or demographic metric for a given country, the resulting chart typically includes two other regions for reference. For instance, at the time of writing this chapter, looking up "population of the US" returns a chart with the populations of the US over the past century, as well as those of Russia and Mexico in fainter lines.

Designers of **Superimposed Summary** guardrails have more liberty in implementation, as there are many types of visual summaries to choose from. When tackling time frame cherry-picking, the Superimposed Summary could take form of lagged variance information. Similar designs have long been popular in financial trading with the purpose of encoding past volatility of a financial instrument alongside its value, and include Bollinger Bands [175] and Standard Deviation Channels, or Envelope Channels. In the context of

item cherry-picking, as discussed previously, Superimposed Summaries could take the form of a simple average, interquartile range, or extrema information of the contextual data using a representation similar to that of contour boxplots proposed by Whitaker et al. [176]. Aside from statistical summaries, a Superimposed Summary could aggregate the contextual data into a background heatmap using techniques such as DenseLines [177]. Superimposed Summaries could also simply draw on common domain-specific benchmarks, such as stock market indices in the financial sector.

The main goal of **Juxtaposed Primary Data** guardrails is to show omitted items or time periods of the main data next to the main chart. An example of previously proposed visualization technique that could serve as a guardrail for time cherry-picking is stack zooming [178]. In the context of data explorers, Juxtaposed Primary Data could benefit from being miniature as to not be too large to gist or end up distracting from the main view. To achieve this, the specific implementations could be chosen from the set of charts designed to take up little space, such as sparklines [179] or horizon charts [180] in instances where the sign of the data provides meaningful context. Additionally, although this type of guardrail does not require a precise definition of "context data set" as Superimposed Primary Data, space constraints may necessitate decisions about what data is seen immediately beside the visualization. Juxtaposed Primary Data guardrails often can play a dual role and take the form of scented widgets [181]. For instance, a sparkline can serve as the preview of an item's data next to each item's selection checkbox—a technique commonly used in financial data explorers. In guardrails against time cherry-picking, the horizontal axis selection slider could instead be a miniature chart of the entire period which the user could brush (as seen in the example on the left in Figure 5.4). An example of an existing implementation of this is the exchange rate explorer on CoinMarketCap.com [182]: a user could zoom in onto a short period of, for instance, Bitcoin price chart by brushing over a miniature view below the main chart.

The guardrail type with the largest set of implementation alternatives is **Juxtaposed Summary** , as there are virtually no restrictions on what could be shown, as long as it provides useful context. One subset of design alternatives consists of variations of focus-and-context visualizations, or designs that allow the user to see both detail and overview simultaneously [183]. This could include distorted views, such as Multistream from

Cuenca et al. [184]. Morrow et al.'s Periphery Plots [185] provide examples of nondistorted variations of this guardrail that could be helpful in the context of time cherry-picking (seen in the example on the right of Figure 5.4). Designs similar to periphery plots could similarly be used to tackle item cherry-picking—in this case, however, the Summary in the periphery would be calculated based on the omitted items rather than omitted time frame. The space of ways to show the distribution of contextual data is large and the choice of a specific visual encoding should depend on the features of the underlying distribution, as well as the expected audience's visual literacy, as these types of charts are typically less common in visualizations for general public. Correll provides a helpful analysis of advantages and disadvantages of distribution visualizations as well as their combinations as raincloud plots [186]. The example in Figure 5.2 as well as our prototype employ a vertical strip plot, but other designs we have considered include a box plot, a violin chart, or an inverted histogram (shown in the example on right in Figure 5.4).

5.4.2.4 Relatedness to Uncertainty Design Spaces

Many of the visual techniques that can be used as guardrails are also commonly used to visualize uncertainty, albeit with different goals and interpretations. The task of exposing cherry-picking is related to visualizing uncertainty: both intend to give a clue about, or cast doubt in, the completeness of the data displayed [187]. However, although uncertainty designs typically address questions like "How *reliable* is this estimate?" or "What is the spread of possible outcomes around *this point*?", guardrails shift the focus to "How *representative* is this value?" and "How does this data point compare to the *rest of the available data*?"

As a result of this difference in goals, there is only a partial overlap in visual techniques for guardrails and for uncertainty visualizations. The shared visual techniques consist of *distributional visualizations* [188] techniques: a box plot could be used to either visualize the uncertainty of an estimate or summarize the omitted data as a guardrail. Similarly, an ensemble visualization [189] of uncertainty could communicate the various forecasts of snowfall for the season, whereas a visually similar Superimposed Primary Dataguardrail could show the historical snowfall data for previous seasons as context. However, many guardrail techniques do not have equivalent uncertainty visualizations, such as the Jux-

taposed Primary Data visualization of the selected time period in Figure 5.4 that simply plots the entire available time series alongside the selection, rather than samples from a distribution. The reverse is also true: direct encodings of uncertainty in the visual channel are not applicable to the guardrail design space—such as modifying the mark of interest by blurring it proportionally to the uncertainty [190]—since there exists no quantifiable dimension of uncertainty. As discussed previously, guardrails do not communicate the uncertainty properties of the value of interest itself and have the goal of not modifying the existing mark.

Building on this distinction, our design space uniquely focuses on describing the opportunities for **augmenting existing visualizations**—a goal not addressed by previous related design spaces. Prior typologies of uncertainty visualization [188], [191], [192] primarily offer a categorization of the source of uncertainty (measurement error versus credibility) or specific visual properties of uncertainty visualization techniques (whether they are animated, 1- or 2-dimensional, have dichotomous boundaries or are "fuzzy"). In contrast, our proposed design space categorizes visualization techniques in a way that describes opportunities and trade-offs of incorporating them into an existing canvas in a nonobtrusive way.

5.5 Prototype Implementation

With the goal of evaluating the overall merits of guardrails, we implemented a prototype data explorer and therefore had to make decisions about what it would look like and what data it would use. As a starting point, we chose one simple visual technique as the implementation for each major guardrail type. In the designing of the prototype, we prioritized *ecological validity*: therefore, we used real-world data, and aimed to mimic the functionality and look of existing data explorer platforms, such as OurWorldInData's COVID-19 explorer [92]. Screenshots of our prototype implementation can be seen in Figures 5.5, 5.6, 5.7, as well as in *supplemental materials*. A sandbox version of our prototype is available *online*.

We chose to design for (and evaluate with) two different data sets that are commonly used in public-facing data exploration tools: COVID-19 death count data by country retrieved from OurWorldInData [92] and individual stock performance data retrieved from

Yahoo! Finance [152]. We anonymized the country and stock names to limit the influence of preconceived ideas about COVID-19 or particular stocks. We limited the number of items available to select to 15 or less to simplify the tasks in evaluation. We also chose to limit our guardrails prototypes to item-cherry-picking, excluding time-cherry-picking from our design and our study.

To create these prototypes, we needed to select a specific design instance of each guardrail. As discussed above in Section 5.4.2.3, the exact visual language and criteria for selecting contextual data are decisions the designer has to make depending on the domain and problem. In our case, in the **Superimposed Primary Data** condition we defined contextual items to be countries from the same region in the Viral scenario; and stocks from the same industry in the Stocks scenario. For **Superimposed Summary** , we showed the average of all items as a line and the interquartile range as a shaded area (Figure 5.6). For a more realistic presentation, we labeled the average line as "Market Index" in the Stocks scenario.

In the **Juxtaposed Primary Data** condition, we chose to provide a sparkline of each item next to its label (Figure 5.5). We filled in the area under the line chart in light color, which made it easier to compare magnitudes among the small multiples that are stacked vertically relative to just lines. Lastly, for the **Juxtaposed Summary** guardrail we implemented a stripplot showing data of all items across all time points, with each tick representing the value of one item on a given day. The ticks are shown with slight transparency to account for over-plotting, and we use color-coding to distinguish positive from negative values—a distinction relevant for the Stocks scenarios. A shaded funnel denotes which part of the global scale is currently shown on the main canvas, and also makes it more difficult to crop the plot without leaving evidence of tampering. For both of the Summary guardrails, we added a sentence explaining what the shaded area or the stripplot denote.

5.6 Study 1: Production

In order to evaluate whether the guardrails make it *more difficult to cherry-pick data*, we conducted a crowd-sourced experiment challenging our participants to use our prototype data exploration platform with and without guardrails in place. One goal of this

study was to evaluate whether participants find it more difficult to create cherry-picked visualizations with guardrails. The other goal of this study was to produce a data set of cherry-picked visualizations to be used in a subsequent reaction study (see Section 5.7). This section describes our methodology and results. Figure 5.8 presents a visual overview of the study procedure in the form of a data comic [193].

5.6.1 Methods

The experiment investigated how participants used our data explorer to create cherry-picked visualizations, both with and without guardrails.

During the study, participants were tasked with using our data explorer to create a visualization supporting a given proposition by selecting data in a line chart and writing a caption, mimicking the act of sharing a cherry-picked data explorer-based visualization on social media. Each participant was randomly assigned one guardrail condition, and they would perform half of their tasks with that guardrail and half without any guardrail, as controls. To prioritize ecological validity, the study employed anonymized real-world data and we realistic scenarios of malicious cherry-picking observed in recent events: such as downplaying the seriousness of a deadly viral disease or exaggerating the profitability of a specific security. Therefore, we constructed four data scenarios presented to the participants: two based on anonymized COVID-19 fatalities, reframed as a fictional viral disease (Viral A & B), and two using anonymized individual stock performance data (Stock A & B). In the Viral scenarios, participants acted as public health officials tasked with promoting not-so-effective policies. In the Stock scenarios, participants imagined themselves as financial advisors promoting underperforming funds at the direction of a supervisor. The prompt and user interface is shown in Figure 5.5.

The study was reviewed by our institution's IRB and deemed exempt from full board review. Before the main study, we conducted two pilots. After a first in-person pilot with two students, we made adjustments to clarify scenario descriptions, add a help button, and fix a visual bug. A second pilot with 10 Prolific participants confirmed the tasks were understandable and feasible for remote users, requiring no further changes. We recruited 130 participants from Prolific for the main study. The survey was conducted using the reVISit framework [194]. We logged the interaction data using the Trrack library [195]

and used the NASA TLX [196] questionnaire to assess participants' subjective workload after each task. At the experiment's conclusion, we debriefed the participants about the guardrails and asked whether they noticed them and felt they influenced the tasks. The median completion time was 15.5 minutes, and participants were compensated $5 (average hourly rate of $19). The study stimulus is available *online*.

The study produced 520 visualizations and captions. After the experiment, two authors independently reviewed the submitted visualizations and captions with the goal of removing unsuitable submissions for the follow-up study. They agreed on inclusion and exclusion for 395 evaluations, discussed and resolved disagreements on 125, and ultimately excluded 18% of submissions. Exclusions were due to irrelevant selections, unintelligible captions, or overly generic responses (e.g., "Chart"). The final data set consisted of 427 submissions.

Quantitative analysis involved paired t-tests to compare the number of clicks and NASA TLX scores between tasks with and without guardrails. Qualitative analysis examined captions and poststudy feedback to identify recurring themes and insights.

5.6.2 Findings

As a result of Study 1, our participants generated 520 cherry-picked visualizations and captions, of which 427 passed our quality review and were used in further analysis. The submissions spanned two scenarios and four guardrails, as well as the control condition. All resulting submissions—visualizations and captions—are available for review *online*.

5.6.2.1 Quantitative Results

Figures 5.9 and 5.10 summarize the within-subject differences in the number of clicks per task and the NASA TLX survey results and the subjective influence of guardrails on task difficulty. **We find that Superimposed conditions made cherry-picking more effortful, but Juxtaposed guardrails made navigating the data easier.**

Participants in the Juxtaposed Primary Data condition required significantly fewer clicks (21 versus 32 in Control; $T(38) = -4.00, p < 0.0001$) and less time to explore the data (106 seconds versus 143 seconds in Control; $T(38) = -2.12, p = 0.041$). They also reported better subjective performance (28.9 versus 39.7; $T(38) = -2.45, p = 0.019$) and lower mental demand (48.8 versus 55.8; $T(38) = -1.84, p = 0.075$). Similar "benefits"

were observed in the Juxtaposed Summary condition for performance (26.5 versus 31.7; $T(52) = -1.95, p = 0.055$). Additionally, Figure 5.10 shows that fewer than 5% of participants in the Juxtaposed Primary Data and Juxtaposed Summary conditions reported that the guardrails made their task "much more difficult." **This indicates that our juxtaposed guardrails actually make cherry-picking easier to achieve,** which is the opposite of our goal.

Those in the Superimposed Summary condition, however, reported higher mental demand (51.3 versus 43.5; T(47)=2.10, p=0.041). Figure 5.10 further illustrates that participants in Superimposed conditions more frequently rated the guardrails as making tasks "more difficult" or "much more difficult", indicating that **superimposed conditions support our goal of making cherry-picking more difficult**.

5.6.2.2 Qualitative Results

Consistent with our quantitative results, participants noted that the **Juxtaposed guardrails made cherry-picking easier**: *"It made it easier to visualize without constantly turning off and on each data set to conceptualize which ones would work best."* This aligns with the design of Juxtaposed Primary Data , which provides a clear overview beneficial for both authors and audiences of cherry-picked visualizations.

In contrast, **participants found Superimposed guardrails to be more obstructive**. Those in the Superimposed Summary condition expressed frustration in text responses and one participant remarked that the guardrails *"interfered with my ability to cherry-pick the data I needed to."* Another participant noted, *"I couldn't easily show that one investment was better than others without hiding some truth."* Two participants even refused to perform the task, with one caption simply stating: *"I can't. I'd be a liar. [Option A] is terrible."*

Despite these challenges, qualitative analysis of captions revealed participants' adaptability. When guardrails exposed context that debunked cherry-picking, **participants often shifted focus**. For instance, one caption downplayed the magnitude of infections under Policy A and emphasized trends: *"Policy A's peak comes fast but comes back down just as fast" (↗)*. Similarly, when promoting Airline stocks, a participant reframed average returns as a sign of stability: *"Over time, growth in stock prices in the airline industry has either been consistent with or outperformed the market average, thus being the most predictable"*

(↗). Others incorporated world knowledge to explain away an underperforming stock: *"Covid kind of ruined airlines but now that it's getting less and less prevalent we're going to see a resurgence"* (↗).

5.7 Study 2: Reaction

In order to evaluate whether the guardrails make *cherry-picked data less convincing*, we conducted a second crowd-sourced experiment asking another set of participants to review charts created by the participants of Study 1. This section describes our methodology and results. Figure 5.11 presents a visual overview of the study procedure in the form of a data comic [193].

5.7.1 Methods

The study was reviewed by our institution's IRB and deemed exempt from full board review. Similar to Study 1, this study was implemented using the reVISit framework [194], and full instructions are available in the *supplemental materials*. Before the main study, we conducted three pilot studies on Prolific with five participants each to test the clarity of the scenario and questionnaire. Based on their feedback, we made minor adjustments to the UI and task language.

For the main study, we recruited 160 English-speaking participants from Prolific. Participants were randomly assigned to one of four scenarios from Study 1 and shown five visualizations in random order—one from each condition (4 guardrails and 1 control), selected from 427 charts created in Study 1. The interface in Study 2 was the same as Study 1, with controls disabled and the addition of captions. Figure 5.6 shows the experimental setup.

To measure guardrails' effect on how convincing the cherry-picked claims are, we presented participants with a hypothetical decision-making scenario inspired by a trust game [197], an approach from behavioral economics. Trust game-based hypothetical decision-making and investment scenarios have been proposed to measure trust in visualizations [198], and more recently have been adapted in visualization work [199]–[203]. Participants in the Stock scenario were asked to imagine they were selecting an investment portfolio in a way that maximizes their profits. They were shown a visualization from

Study 1 that promoted a certain investment, and were asked to decide how much they would invest ($0 to $100) in the recommended industry based on the available information. In the Viral scenario, participants were asked to imagine they were traveling to a virus-affected area and had to buy insurance to minimize their financial risk. They were shown a visualization from Study 1 that downplayed the risk of virus and were asked to choose how much health insurance they would buy (also $0 to $100) based on the available information.

Participants were also asked to provided a brief rationale for their choices and completed Likert-scale questions assessing trustworthiness, persuasiveness, clarity, and likelihood of sharing the visualization. In the debrief, they were also asked whether they noticed the guardrails, and whether they understood their meaning. The median completion time was 10 minutes, and participants were compensated $2.50 (median hourly rate of $15/hr). The study stimulus is available *online*.

We analyzed the data quantitatively using repeated measures ANOVA and post hoc t-tests with the Benjamini-Hochberg [204] procedure for multiple comparisons. The scripts and the results of all statistical tests are available in the *supplemental materials*. We also qualitatively reviewed the justification and feedback text to identify key themes.

5.7.2 Findings

Figure 5.12 presents the monetary action results by guardrail and scenario. Guardrails had observable effects on skepticism in three of the four conditions. **Primary Data guardrails performed slightly more effectively than Summary guardrails**, with some participants finding the visual Summaries more difficult to interpret, leading them to ignore these guardrails. The Primary Data guardrails appeared to be more intuitive, as their visual encoding resembled the main data selections, and few participants expressed confusion.

5.7.2.1 Quantitative Results

Repeated measures ANOVA tests showed **significant differences in monetary decision between guardrails and control conditions in the Viral A scenario** ($F(4, 152) = 7.79, p < 0.001$). post hoc t-tests revealed that participants in the Superimposed Primary Datacondition spent, on average, $63 on insurance, compared to $45 in the Control con-

dition ($T(38) = 4.28, p = 0.001$). Superimposed Summary (mean \$55, $T(38) = 1.99, p = 0.076$) and Juxtaposed Primary Data (mean \$56, $T(38) = 2.72, p = 0.024$) also led to higher insurance purchases compared to Control. Despite this, the poststudy survey revealed that **most participants did not fully consider the guardrails**: 38% of participants did not notice them, and 23% noticed them but did not understand their meaning. As a result, we do not observe statistically significant results in the other scenarios.

5.7.2.2 Qualitative Results

Participants who did notice the guardrails directly referenced the guardrails in their rationales. For example, a participant in the Superimposed Primary Datacondition stated: *"[The chart] shows [option A] to be the worst one out of the lines shown."* Another noted in the Superimposed Summary condition: *"[the infection rate] is still considerably higher than the average..."* These findings suggest that the guardrails may blend into the visualization, depend on data literacy, or be overlooked due to added visual complexity. Even among those who understood the guardrails, some chose to focus on the main data instead, with one participant stating: *"I just followed the figures and how they were either rising or falling"*.

5.8 Study 3: Controlled Reaction

We conducted a third crowd-sourced experiment to further explore how guardrail effectiveness varies with the severity of cherry-picking, while also accounting for the influence of redeeming factors. In contrast to Study 2, in this experiment we prioritize internal validity by removing captions and controlling the data displayed in each scenario, which allowed us to isolate the impact of guardrails from other factors. Below, we describe our methodology and findings. Figure 5.7 presents a visual overview of the study procedure and screenshots of the conditions in the form of a data comic [193].

5.8.1 Methods

For Study 3, we used a simplified set of five stocks from the same Yahoo! Finance [152] data set as in Study 2, selecting stocks with nonoverlapping time series to highlight the highest return in any data subset. Unlike in previous studies, we removed the caption to isolate the effect of the guardrails. To measure guardrail effects, we asked a 5-point Likert question, "The visualization supports the idea that stock X yielded the highest returns in

2023", followed by an open-text rationale and an attention check.

Study 3 included three conditions: *Correct* (control), where all data was shown and participants were prompted about the actual top-performing stock; *Incorrect*, where the best-performing stock was hidden and participants were prompted about the second-best stock; and *More Incorrect*, where the top two stocks were hidden and participants were prompted about the the third-best stock, which performed below the average of all stocks. In all cases, the prompted stock was the highest performing "salient" stock, but the guardrails could be used to discover that the chart was cherry-picked. We conducted a first pilot with 30 participants, after which we rephrased the survey questions for clarity. We then conducted a second pilot with 150 participants and performed a power analysis to calculate the required sample size for detecting meaningful differences across conditions with a target power of 0.8 and an alpha level of 0.05.

For the main study, we recruited 675 Prolific participants, all fluent in English, who were randomly assigned to one of the three correctness conditions and one of five guardrail conditions, i.e., each participant saw only a single chart. The median completion time was 3 minutes, and participants were compensated $0.75 (average hourly rate of $15). The study stimulus is available *online*. We again analyzed Likert responses with ANOVA and post hoc t-tests, applying the Benjamini-Hochberg procedure [204] to adjust for multiple comparisons, and reviewed participants' text responses.

5.8.2 Findings

5.8.2.1 Quantitative Results

Figure 5.13 summarizes Study 3 results for participants' agreement with the statement that the cherry-picked item is the highest overall. Our ANOVA results showed statistically significant differences due to both cherry-picking severity ($F(2, 672) = 31.39, p < 0.0001$) and guardrail condition ($F(4, 670) = 43.18, p < 0.0001$), confirming that **guardrails influence skepticism toward cherry-picking**, with more severe cherry-picking yielding stronger effects. As expected, we observe no effects in the *Correct* condition, indicating that guardrails do not have an effect on correct interpretations.

Post hoc t-tests with Benjamini-Hochberg FDR correction revealed that, compared to Control, the Superimposed Primary Data guardrail had a large, significant effect in both

Incorrect and *More Incorrect* conditions ($T(88) = 10.88, p < 0.0001$ and $T(77) = 12.68, p < 0.0001$). The Superimposed Summary and Juxtaposed Primary Data guardrails also had significant effects but only in the *More Incorrect* scenario ($T(70) = 3.61, p = 0.002$ and $T(73) = 2.39, p = 0.058$, respectively). Similar to Study 2, the **Superimposed Primary Dataguardrail was the most effective against cherry-picking, while Juxtaposed and Summary guardrails had smaller effects**. Juxtaposed Summary , the most visually distinct guardrail, produced results closest to control. Notably, **the distribution of responses in all conditions is highly polarized**: almost all responses either strongly agreed or strongly disagreed with the statement. Several conditions exhibit a bimodal response distribution, suggesting that quardrails are likely effective when noticed and understood, but the challenge lies in capturing the viewer's attention and understanding.

5.8.2.2 Qualitative Results

We reviewed the open-text responses in the Superimposed Summary of *More Incorrect* condition, which was highly bimodal and allows us to better understand the source of an audience's disagreement. Participants' explanations reinforce the notion that guardrails are effective when they successfully capture both the viewer's attention and their understanding. We find that 13 of 15 responses that correctly disagreed (1 or 2 on the Likert scale) explicitly referenced the guardrail showing that the cherry-picked stock was below the industry average. One participant stated, *"Stock C was below the industry average at the end of the year, meaning it couldn't have possibly had the highest return in the industry."*

In contrast, only 9 of 31 responses that incorrectly agreed (4 or 5 on the Likert scale) referenced the average line, but none were able to integrate it into their reasoning. These participants merely acknowledged its presence and primarily focused on comparing the cherry-picked stocks, as illustrated by one explanation: *"I looked at the industry average. And then I also looked at all the other stocks, C looks like it's been doing better than any other one."* These results point to the fact that, in order to be effective, more complex guardrails should be both highly visually salient and directly cue the user about how to integrate them into higher level chart comprehension [205].

5.9 Discussion and Design Recommendations

Our findings indicate that guardrail designs have the potential to mitigate cherry-picking in data explorers, though they come with limitations and important considerations. In this section, we discuss these findings and offer design recommendations for effective guardrail implementation.

Recommendation 1: Prioritize simpler guardrails that maintain the original visual language. Our design procedure and crowdsourced studies uncover a tension between designing a guardrail that effectively summarizes contextual information and maintaining its alignment with the original chart's visual language, as summarized in Figure 5.14. The closer the guardrail is to the original chart visually, the easier it is for the audience to notice and understand it. At the same time, guardrails that use the same visual encoding for individual items can lead to overplotting. Our studies show that *Primary Data* guardrails are the most effective, primarily due to the fact that they are easier for audiences to notice and understand in both real-world and controlled scenarios. Additionally, they do not require viewers to independently extract high-level patterns, such as integrating the distributional statistics into their interpretation, a process highly affected by individual differences [205]. Therefore, we recommend prioritizing guardrails that closely match the visual encoding of the main data and using *Primary Data* guardrails as the default. However, in some cases the domain in question may require larger amounts of information to be used in the guardrail or could benefit from specific statistical summaries. When using a *Primary Data* guardrail is not an option, we suggest that *Summary* guardrails should always be paired with tutorials or detailed annotations to help users interpret and apply them correctly.

Recommendation 2: Carefully identify potential targets and contextual information. Across all types, constructing a guardrail demands a clear and careful definition of "contextual information" that would debunk cherry-picking, whether it is demographically comparable subsets, or countries with similar climates. Defining this context correctly may require a careful examination of the domain and consulting with a domain expert. Additionally, an evaluation of the domain and existing misinformation can help determine which types of cherry-picking should be addressed in the first place. For example, climate change misinformation often involves cherry-picking specific time frames [107], while

COVID-19 conspiracy theories typically cherry-pick data from items (e.g., countries) [6]. In this process, we recommend following the threat modeling steps outlined in Section 5.3: reflect on the system goals, identify vulnerabilities, propose and evaluate interventions.

Recommendation 3: Implement guardrails to deter cherry-picking and enhance platform usability. Our results from Study 1 show that *Superimposed* and *Juxtaposed* guardrails have different impacts on authors' experience using the data explorer. Similar to previous work evaluating visualization composition, we found that participants found it easier to integrate superimposed information [205]–[207]. *Superimposed* guardrails thus made it harder to create cherry-picked views and led to less convincing presentations. This could push misinformation actors toward less reputable platforms, while helping unintentional cherry-pickers recognize their biases. In contrast, participants cited that *Juxtaposed* guardrails made it easier to explore the data. Although this may aid careful cherry-picking, we found no evidence of exploration ease leading to more misleading visualizations. Since *Superimposed* and *Juxtaposed* guardrails can be effectively combined, we recommend using both to prevent cherry-picking while improving data exploration.

Recommendation 4: Continually adapt to ever-evolving misleading strategies. Importantly, we found that cherry-picking is an *adversarial process*. Creators adapt to effective guardrails and find new ways to cherry-pick. Previous research shows that people often crop visualization screenshots, add misleading text and annotations [6]. Our experiments confirm that authors shift focus to specific salient data features to influence readers' interpretations. As a result, it is essential to continually adapt to ever-evolving misinformation arguments and to iteratively update the designs to ensure their clarity. Analyzing usage logs from the data explorer could offer insights into the data subsets creators focus on, guiding future guardrail improvements.

5.10 Conclusion and Future Work

In this chapter, we describe an approach to designing technical interventions against the misuse of data visualizations in support of misinformation. We examine cherry-picking in visualization through the lens of threat modeling and describe the design space of *guardrails*: interventions incorporating contextual data that would expose cherry-picking if it is there, and not interfere with interpretation if not. Our experiments find

that guardrails make it more difficult to create cherry-picked charts and encourage viewer skepticism; however, the difficulty in implementing successful guardrails lies in drawing the audiences attention to them.

We are hopeful that data exploration platforms adopt similar interventions in their designs. A review of visualization flaws from designers' perspective by Lan and Liu [208] highlights multiple stages in the design workflow where such flaws can emerge, uncovering opportunities for more targeted interventions. Implementing guardrails and other visualization threat mitigation strategies in practice would allow future work to examine the role that they play in complex real-world contexts. As discussed, misinformation using data visualizations is adversarial and a wicked problem. Because of this, studying real-world adoption of guardrails would be especially important: while the results of our experiments show moderate effects of guardrails encouraging skepticism, it is challenging to predict the exact effects of guardrails on online data discourse.

We hypothesize that beyond influencing an individual in isolation, guardrails could have indirect effects in a world where the general public is familiar with their use. For instance, guardrails could provide evidence that triggers a fact-checking discussion on social media, while guardrails cropped out of a screenshot could alert the audience of tampering attempts. Guardrails surfaced in a chart, even if missed by the original poster, can be referenced by community notes and replies to fact-check the post without any external information. The use of guardrails against misinformation could also eventually be associated with reputable sources and serve as a trustworthiness indicator in and of itself [7]. Aside from adapting and evolving their tactics, it is also likely that malicious actors would migrate to other, nonguardrailed platforms, and create a demand for "alternative" data exploration sites.

In addition to studying the effects of guardrail adoption, future work should also examine strategies to make guardrails—in particular Juxtaposed designs—less amenable to cropping, as well as the effects of combining multiple forms of guardrails. Future research should also investigate automated methods of determining appropriate domain data or aggregation type to be used as context.

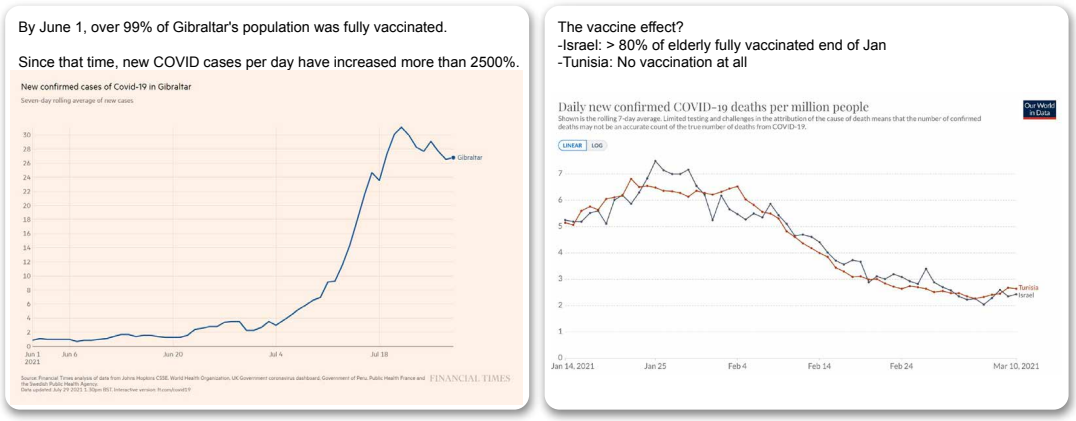


Figure 5.1: Examples of tweets that spread vaccine hesitancy using screenshots of time series explorers with cherry-picked data. The tweet on the left used the Financial Times explorer to show an increase in cases in a single highly-vaccinated country (Gibraltar), implying that vaccines are harmful. The tweet on the right used the OurWorldInData explorer [92] to show two countries of different vaccination levels with similar trends of deaths (Israel and Tunisia), implying that vaccines are ineffective.

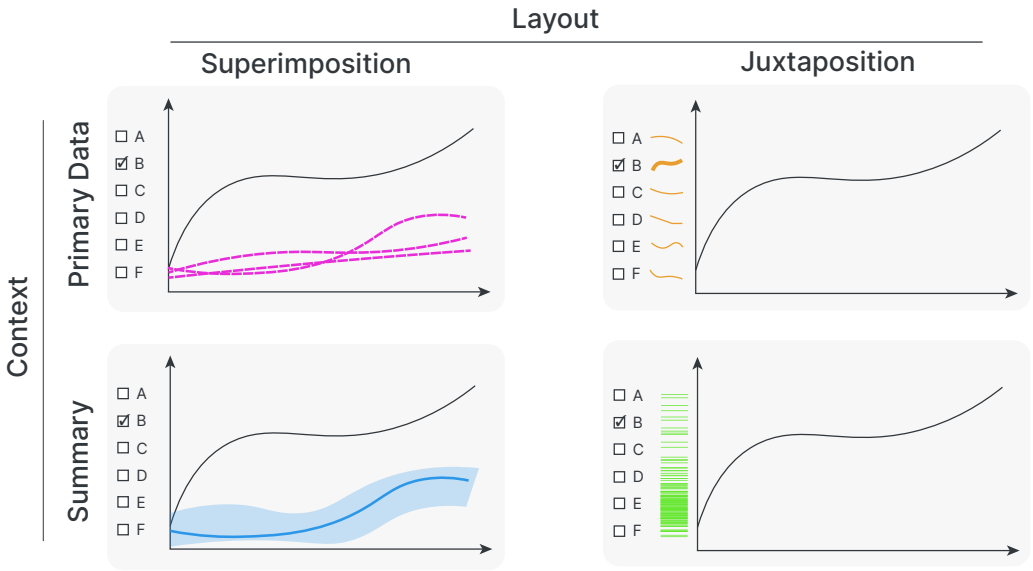


Figure 5.2: The design space of visualization guardrails against cherry-picking along two dimensions: what *context* is shown (primary data or a summary) and *layout*, or where it is shown (superimposed on or juxtaposed with the main chart). The figure shows the *main chart data* in black and the *context* in color. In all example charts, the user selects only one item (B) which, as shown by the guardrails, is an outlier in the data set.

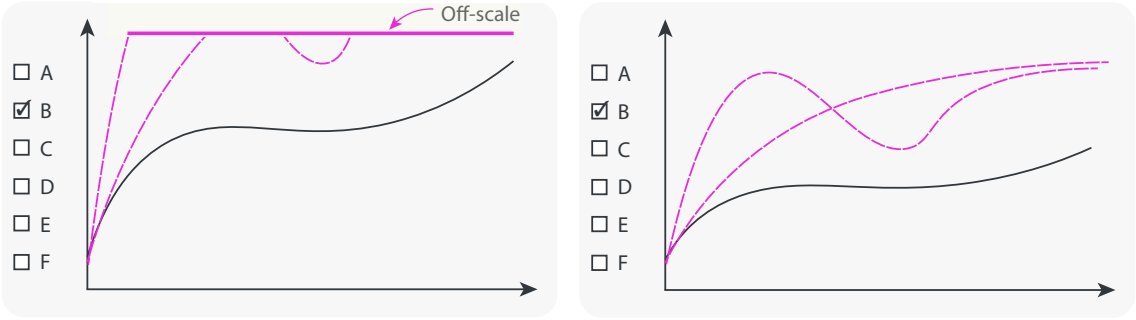


Figure 5.3: Two design variations of vertical scale treatment in Superimposed guardrails. The example on the left preserves the scale of the main chart data, emphasizing the original selection but truncating the off-scale context. The design on the right adjusts the scale to fit the guardrails.

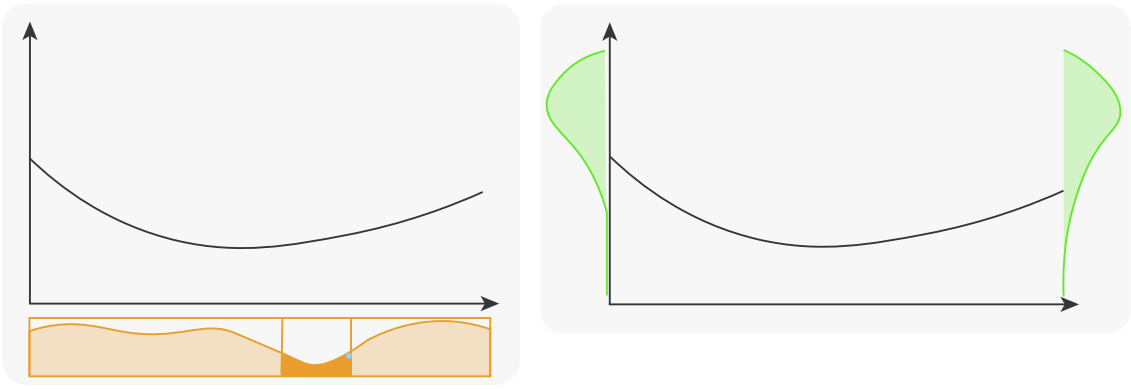


Figure 5.4: Examples of Juxtaposed guardrail variations for time frame cherry-picking. On the left is an example of Juxtaposed Primary Data : the miniature view below shows the entire time frame and highlights the cherry-picked period of a dip in value. On the right is a periphery plot [185]—an instance of Juxtaposed Summary . The peripheries show that the value of interest is much higher in the periods before and after the selection.

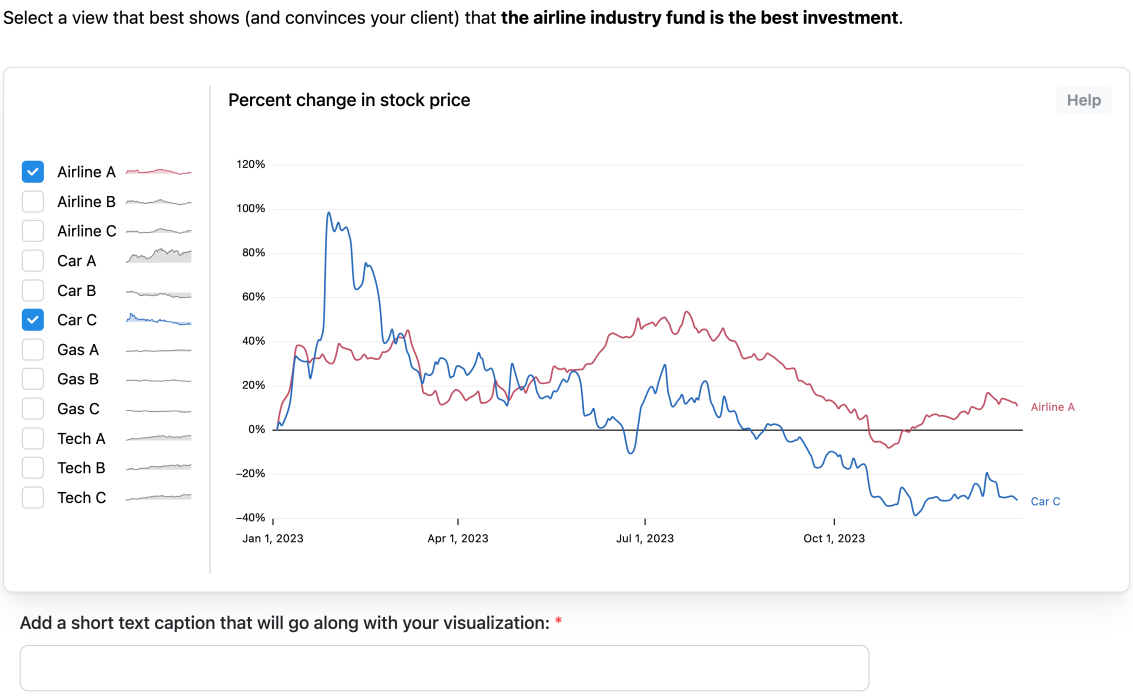


Figure 5.5: Screenshot of our prototype implementation of data explorer in the Study 1 experimental setup. Shown is the Stocks B scenario with a Juxtaposed Primary Data guardrail.

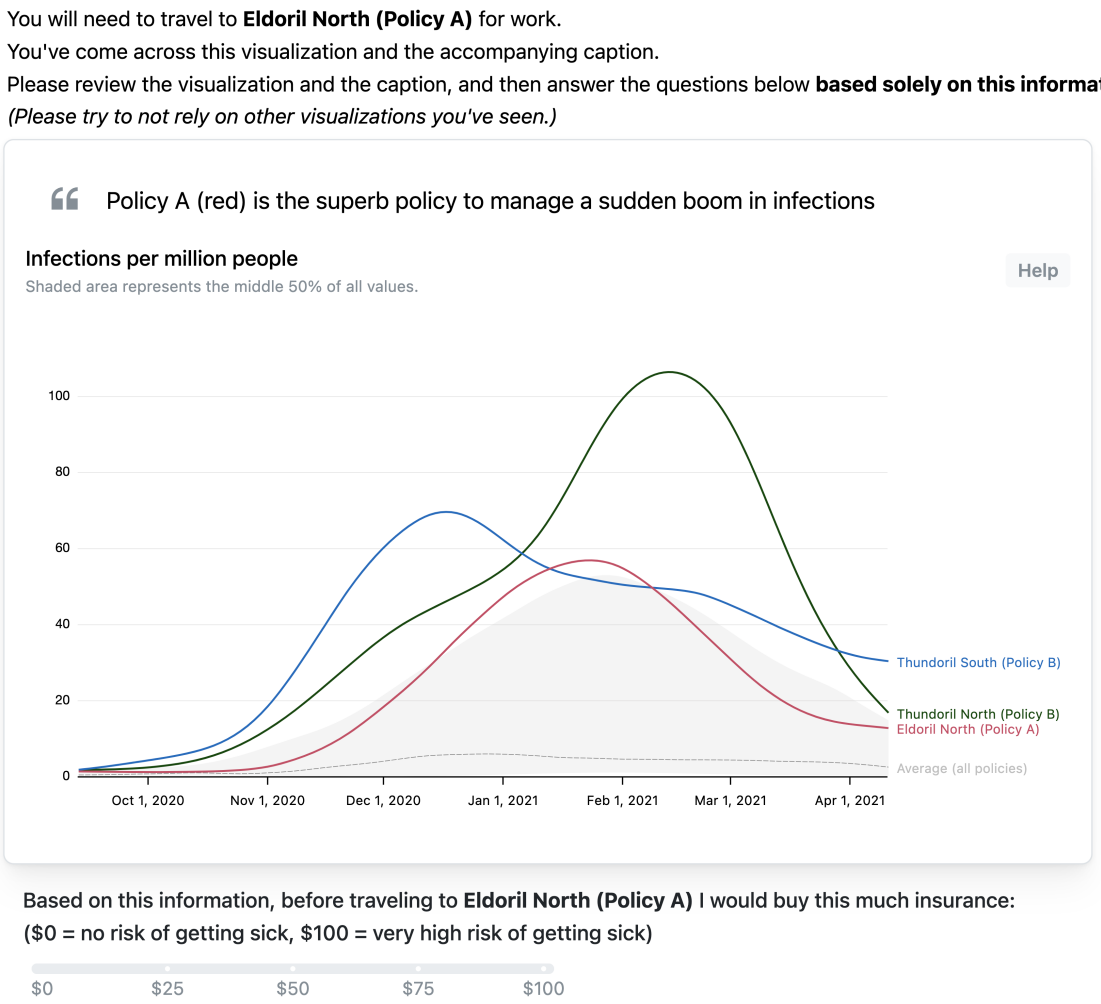


Figure 5.6: Screenshot of our prototype implementation of data explorer in the Study 2 experimental setup. Shown is the Viral A scenario with Superimposed Summary .

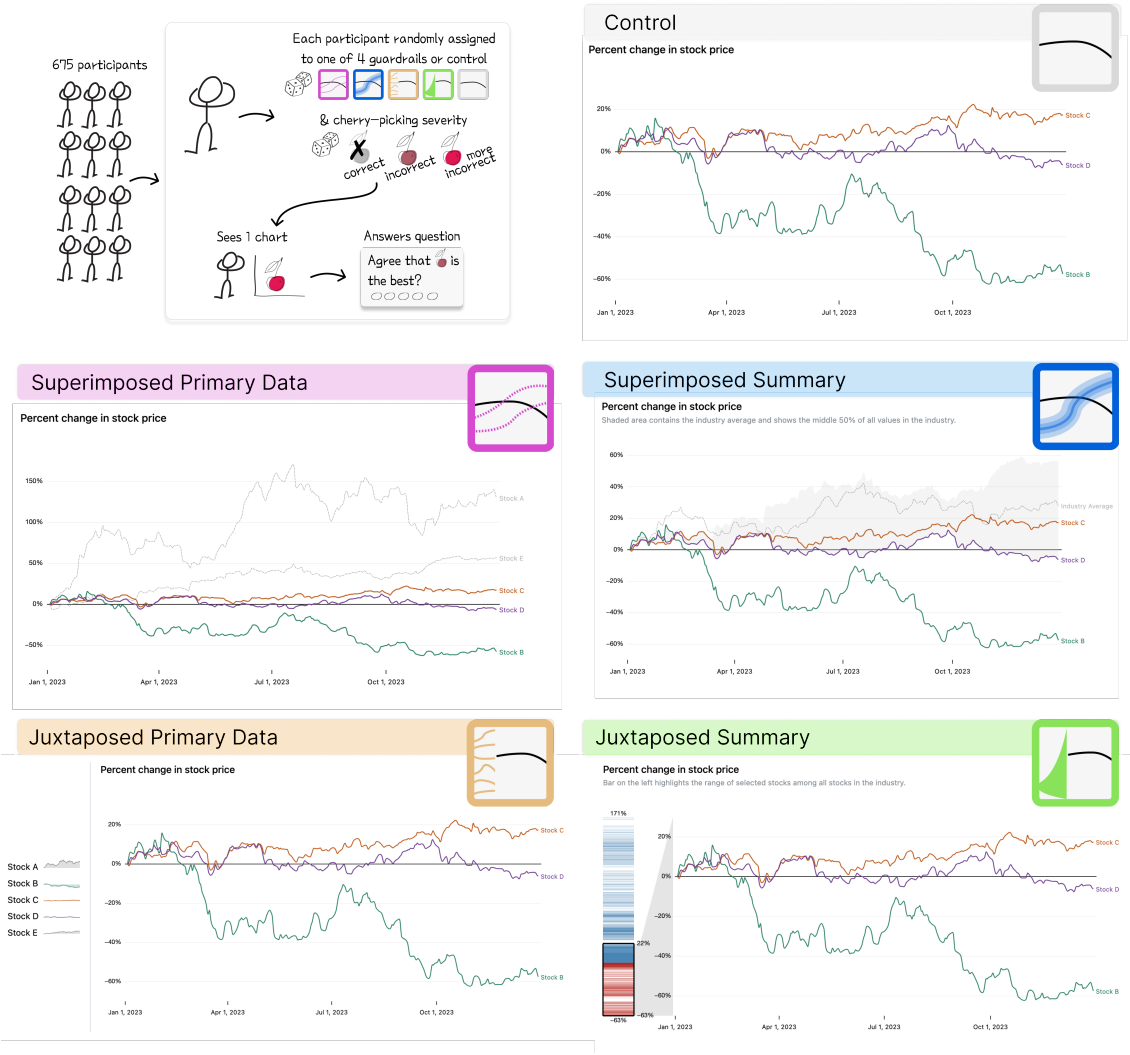


Figure 5.7: Data comic and conditions in Study 3: Controlled Reaction. Participants were asked to rate the following prompt on a 5-point Likert scale ('Strongly disagree' to 'Strongly agree') "The visualization supports the idea that stock C yielded the highest returns in 2023." Screenshots show the visualization seen by participants in the "More Incorrect" scenario, in which the top two true best stocks are hidden. The Incorrect scenario only hides the top stock, whereas the Correct scenario shows all stocks.

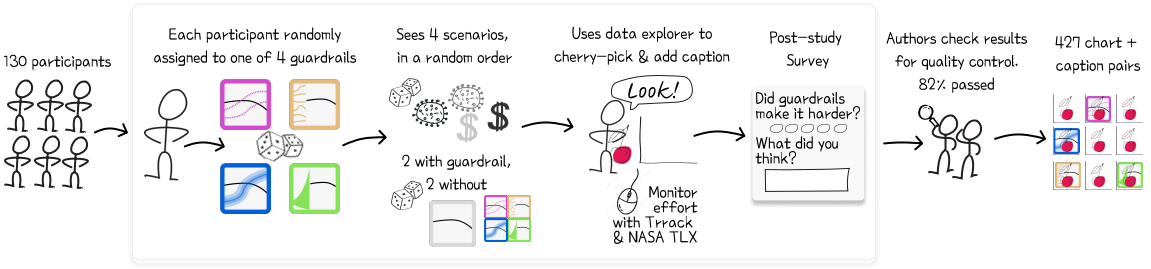


Figure 5.8: Data comic showing the study design in Study 1: Production. The study design employed a between-subjects approach for the guardrail condition and within-subjects approach for the task scenarios.

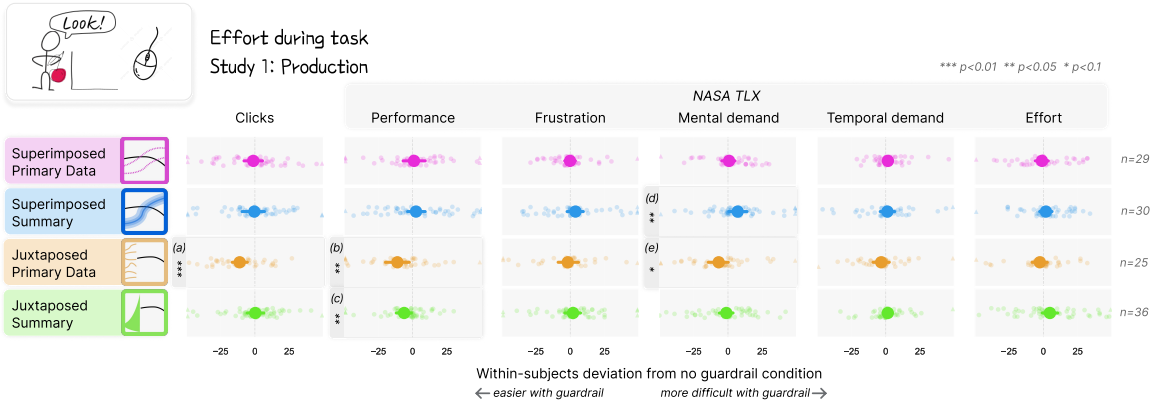


Figure 5.9: Juxtaposed guardrails make it easier to complete the cherry-picking task: participants require fewer clicks to complete the task (a), report better performance (b, c), and slightly lower mental demand (e). Superimposed Summary leads to a higher reported mental demand (d). Shown are the mean difference between the metrics for guardrail and control tasks for each participant, as well as 95% bootstrapped confidence intervals. For clarity, large outliers (¡5% of data) are shown as triangle markers on the edges. For the NASA TLX Performance metric, lower means better [196].

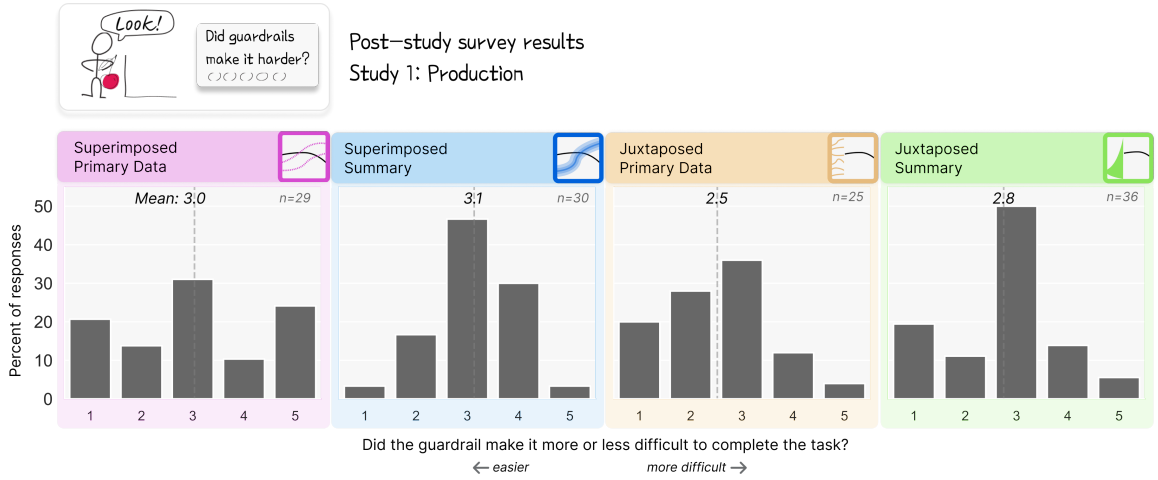


Figure 5.10: Participants in the Superimposed Primary Data condition described the guardrail to make cherry-picking "much more difficult" more often than those in any other condition; however, the responses are very split. On average, participants reported Juxtaposed guardrails to make it slightly easier to complete the task.

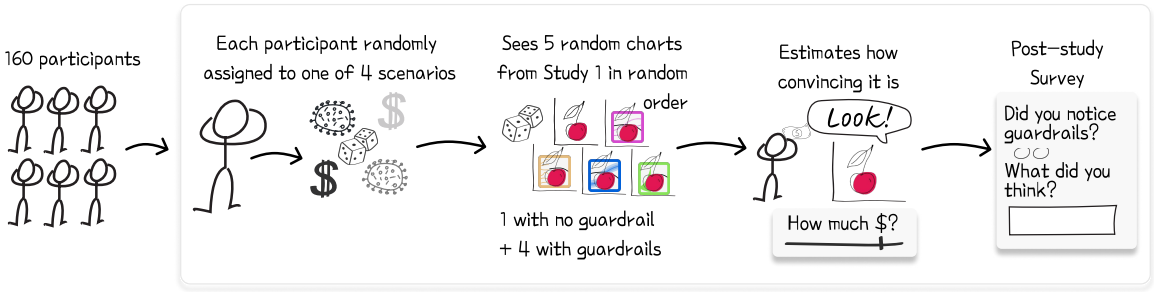


Figure 5.11: Data comic showing the study design in Study 2: Reaction. The study design employed a between-subjects approach for the task scenarios and within-subjects approach for the guardrail condition.

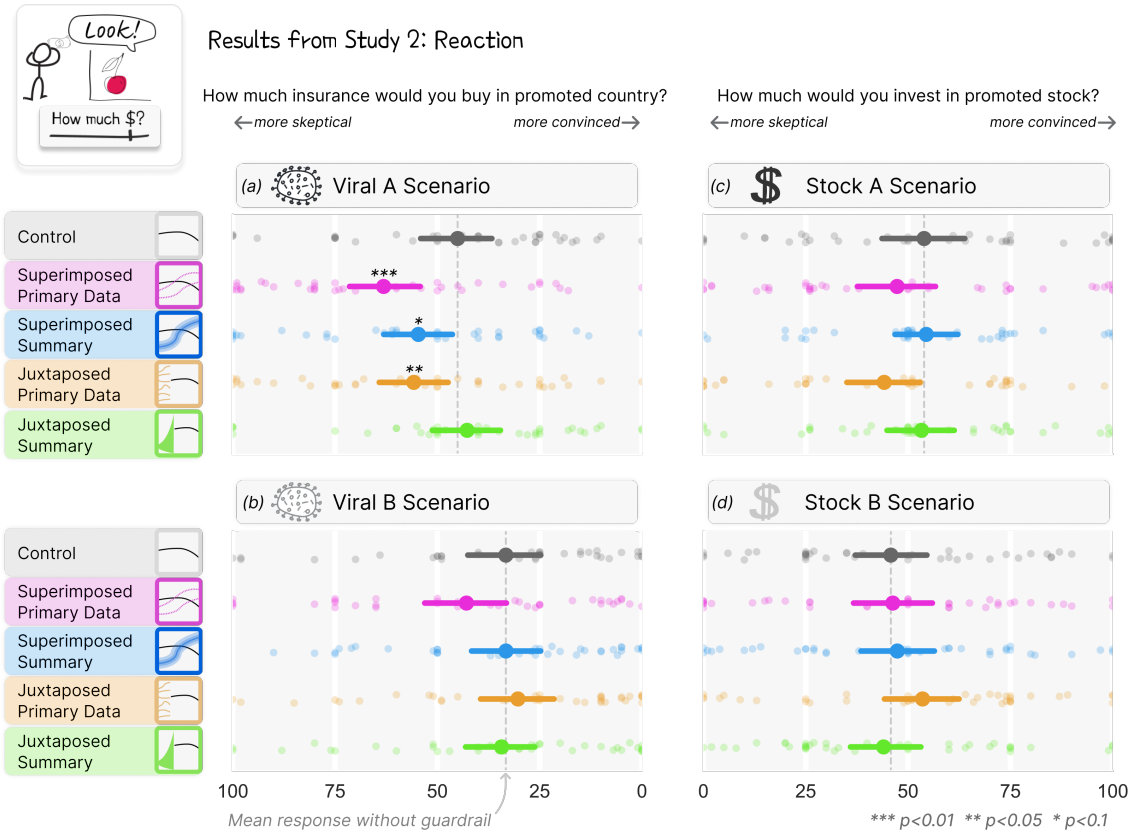


Figure 5.12: Most guardrails had a significant effect on making the viewers skeptical of the cherry-picked charts in the Viral A scenario (a). However, the effect was smaller in other scenarios (b–d). The effects vary highly by guardrail and scenario. Shown are average values and bootstrapped 95% CI ($n = 1000$) and individual data points in the background. Note the inverse scale between the Viral and Stock scenarios: the monetary action in the Viral scenarios involves making a decision about insurance purchase (less insurance = trusting cherry-picking), whereas in the Stock scenarios it involves making an investment (more investment = trusting cherry-picking).

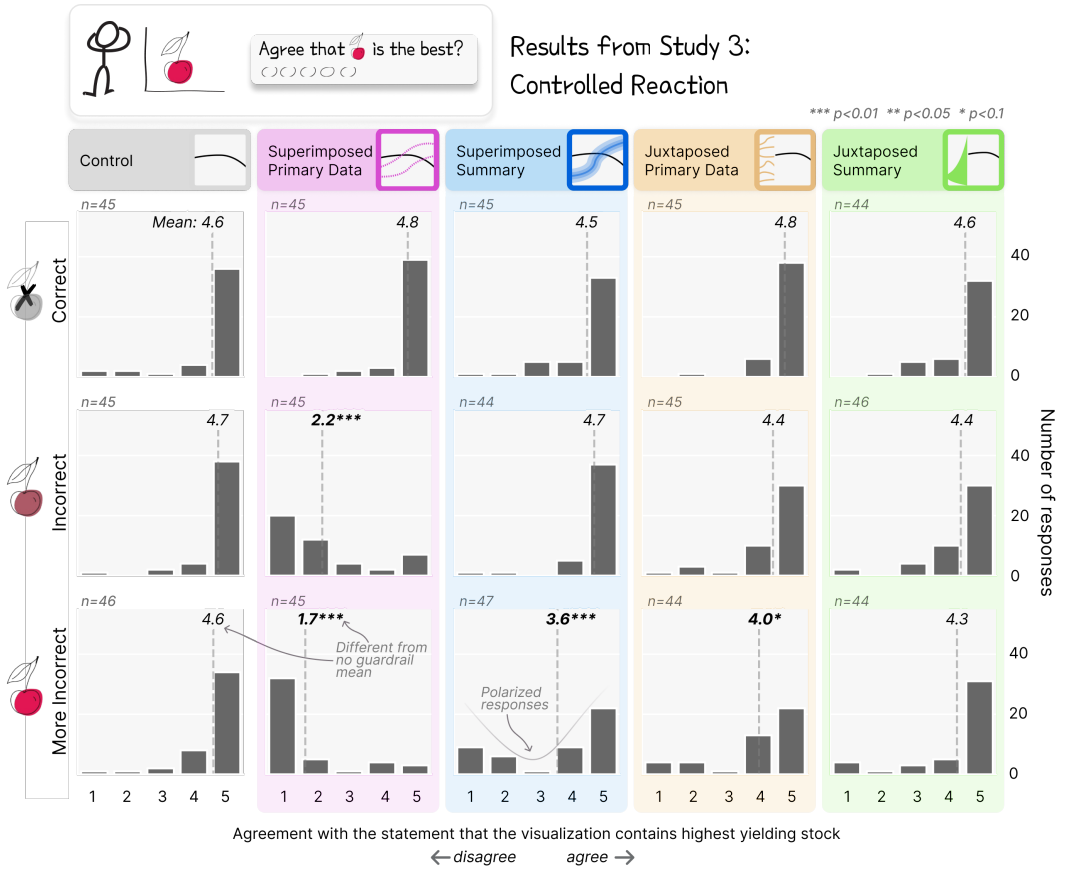


Figure 5.13: In Study 3, Superimposed Primary Data has a very strong effect of encouraging skepticism in both cherry-picked conditions compared to no guardrail. Superimposed Summary and Juxtaposed Primary Data have a significant effect in the More Incorrect condition. As expected, guardrails have no effect in the Correct, not cherry-picked condition. Note the highly polarized distributions of responses: participants either strongly agree or disagree.

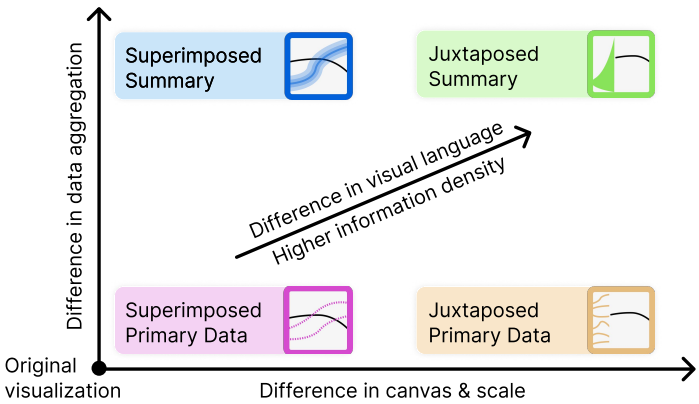


Figure 5.14: This overview highlights the design trade-offs in the guardrails space, as identified through our design process and evaluation studies. By moving beyond the data aggregation level and chart canvas of the original visualization, designers can incorporate more of the potentially omitted context and information, allowing for a richer guardrail representation. However, this shift also moves away from the visual language of the original chart, which may make it harder for the audience to accurately interpret the guardrail. Balancing this tension is crucial when designing guardrails in data explorers.

CHAPTER 6

"I DID NOT LET THE LABEL IMPAIR MY JUDGEMENT": INEFFECTIVENESS OF FACT-CHECKING LABELS ON VISUALIZATIONS

Recent work on misleading visualizations emphasizes that reasoning errors—such as misinterpreting causality or missing normalization—are a common problem in visual misinformation. While redesigning charts may help, reasoning errors arise in relation to specific data questions: a visualization may be well-designed for its intended message but still lend itself to misinterpretations. Despite fact-checking labels being common in other misinformation contexts, research on textual warnings in data visualizations is limited. In this chapter, we conduct a study to explore the effects of different types of labels on real-world charts commonly misinterpreted to support conspiracy theories. A crowdsourced study reveals that such labels do not consistently increase skepticism or trust. Our analysis identifies reasons for their ineffectiveness, including perceived bias and aversion to uncertainty, in contrast to the perceived neutrality and certainty of the visualization. As a consequence, we argue that mitigating misinterpretations requires anticipatory visual design, not just textual framing interventions.

6.1 Introduction

The same chart can support many interpretations, even opposing ones. A declining case rate in a COVID-19 chart might be taken as evidence that the virus was disappearing—when in reality, testing had simply decreased. A chart showing a recent uptick in economic growth might be presented as proof of recovery, while zooming out to a longer time frame could reveal stagnation. Without additional context, how do we know which interpretation aligns with reality? Was the chart designed with a particular narrative

in mind, or does it require domain knowledge, statistical nuance, or methodological details to interpret correctly? And if a viewer already has strong beliefs, do they just see what they want to see? Misleading interpretations—whether intentional or not—can reinforce false narratives, shape public opinion, and spread misinformation under the guise of objective data.

Misleading visualization interpretations such as these present a great challenge to correct. By relying on real data, the persuasive power of visual misinformation comes not from outright falsification but from how the data are framed, contextualized, or selectively emphasized and may lead its audience to make a nuanced analytical reasoning error. Consequently, designing against such misinterpretations is difficult and simply following common design guidelines is not enough. Prior research on COVID-skeptic visualizations shared on social media found that most of such charts were well-constructed [56] or even repurposed from government dashboards and other reputable sources [6]. Similarly, while Ge et al.'s V-FRAMER framework [79] provides guidance for designers on anticipating reasoning errors, authors emphasize that design choices must be tailored to a specific policy question. Consequently, it is unlikely that a single design fix can eliminate all potential misreadings.

These findings pose a dilemma: what if a chart is well-designed for policymakers and researchers but when pulled out of context, reframed, or selectively cropped, it becomes evidence for an entirely different narrative? Instead of endlessly refining charts to preempt every possible misinterpretation, an alternative approach is to intervene after the fact—providing context on the way the visualization should be interpreted. Textual framing of visualizations has been found to greatly influence people's takeaways [58],[209]. Given the popularity of fact-checking labels and community notes in other misinformation contexts, a textual label could present a promising avenue for addressing misleading visualizations.

Much of what we know about fact-checking labels comes from studies on conventional misinformation: fabricated statistics, false claims, and misattributed quotes. Research consistently finds that fact-checking labels reduce belief in myths, increase skepticism toward unreliable sources, and even prevent further sharing of falsehoods [21], [25]. However, misinformation is not always that explicit. Other formats—memes, AI-generated images, and misrepresented photos—rely more on suggestive framing than outright deception,

significantly reducing the effectiveness of fact-checking labels [37].

In this work, we conduct a mixed-methods study to investigate whether fact-checking labels can counter misinterpretations of charts. Drawing from misinformation research, we develop three types of labels designed to address reasoning errors in visualization interpretation. We develop labels for five charts in a case study of figures from real-world government reports and scientific publications that were widely reinterpreted online in ways that fueled misinformation. For each chart, we identify the intended factual interpretation (the *fact*) and the unintended misleading interpretation (the *myth*), borrowing terms from the misinformation literature [210]. Using a preregistered crowdsourced study with 100 participants, we test the impact of fact-checking labels on trust in the chart and agreement with both the fact and the myth.

Ultimately, counter to our preregistered hypothesis, we find that **fact-checking labels were ineffective at promoting skepticism in misleading readings of charts**. Through a qualitative analysis of participant responses, we reveal the diverse ways people internalize fact-checks on visualizations—and, crucially, the reasons they fail. Unlike text, visualizations and data are often perceived as inherently objective. Participants frequently expressed trust in the data while dismissing or downplaying the fact-checking label, seeing it as an editorial addition rather than a correction or important context. These findings suggest that fact-checking strategies that work in other misinformation contexts may not translate directly to visualizations, instead requiring visual and data-driven approaches to fact-checking.

6.2 What Should a Fact-Checking Label Look Like?

In this section, we identify the distinct types of information one could provide in a fact-checking label, in order to evaluate the merits of different approaches to fact-checking visualizations in our study. We draw on research from conventional misinformation fact-checking to help guide the types of information to include when addressing data-driven fallacies [210]. The ideal fact-checking label should clarify misinterpretations, while avoiding the risk of seeming biased or inadvertently reinforcing false narratives. Based on these goals, we developed three distinct types of labels that we term *Warning*, *Counter*, and *Guide*. Each label type has pros and cons informed by both visualization fallacies and

insights from debunking literature. The example labels in this section are based on the visualization in Figure 6.1 and discussed in further detail in Section 6.3.4.

6.2.1 Warning Labels: Highlighting the Fallacy

> Zooming out on the entire time series may obscure the important recent trends in climate.

Warning labels serve two key functions in addressing misinformation. First, they state the myth itself. While this may seem counterintuitive, research suggests that explicitly addressing the myth helps viewers recognize the error's existence in the first place [211], [212]. Second, a warning label directly explains the fallacy [107], such as cherry-picking or the base rate fallacy, helping debunk the myth [213].

One benefit of warning labels is that they are domain-agnostic: not only can they be applied to a wide range of misinterpretations but they can also educate the audiences about other, related myths [214]. However, this approach also has limitations. Although it addresses errors directly, it risks coming across as purely corrective or even adversarial. If not carefully constructed, a warning label may inadvertently strengthen the false interpretation by bringing additional attention to the myth [215]. This label may also alienate viewers, especially if they feel their interpretation being dismissed without further explanation.

6.2.2 Counter Labels: Offering an Alternative Narrative

> Although the climate has varied over the course of history, we are currently experiencing the fastest recorded increase in temperature, which is a concern.

Counter labels offer a corrective by presenting a substitute narrative. This approach is particularly valuable in ensuring the viewer not only understands that a myth is incorrect, but also grasps a more plausible, evidence-based explanation [216]. Research indicates that providing a counter-explanation is more effective than merely highlighting the falsehood, as it helps to solidify the corrected interpretation in the viewer's memory [217], [218]. A counter label helps replace the myth with a more accurate or nuanced understanding of the data, which can be essential when dealing with complex or ambiguous situations [219].

Offering the viewer an alternative explanation can be more memorable than simply negating a myth. However, if the alternative explanation is unconvincing, it may fail to correct the misinterpretation and could even leave the viewer more confused. Addition-

ally, counter labels can be challenging to construct, especially when the myth is nuanced or when there is no clear "correct" interpretation.

6.2.3 Guide Labels: Shaping the Interpretation from the Start

> This chart presents the most detailed reconstruction of Earth's temperatures over the past 485 million years, revealing a history of dramatic shifts in climate before humans existed. This research provides a deeper understanding of how past environments influenced life on Earth and opens new questions about the mechanisms driving climate stability and change in the current geological stage.

Guide labels take a different approach. Instead of correcting a misinterpretation, guide labels provide context and outline the intended message of the visualization [215]. In situations where a myth may not have been obvious or popularized yet, a guide label may be more effective by not perpetuating the myth and instead clarifying the intended insights. This approach has been found useful, especially with more complex scientific information [220].

Guide labels aim to prevent misinterpretation before it occurs. However, this also means that guide labels may not be as effective once a myth has already taken hold. If the viewer is already thinking it, the myth remains unaddressed.

6.2.4 The Case for Combining All Three

While each type of fact-checking label can be effective on its own, the most robust strategy for combating misinterpretation may involve combining all three approaches. Drawing from Lewandowsky et al.'s Debunking Handbook [210], effective labels are typically comprehensive rather than simple explanations. This approach helps address misinterpretation from multiple angles, potentially forming a stronger correction or, at a minimum, offering the reader to focus on a correction that is most relevant to them. At the same time, a significantly longer label may be harder to read or tempting to completely skip over.

6.3 Case Studies with Misinterpreted Visualizations

In this section, we introduce the visualizations we selected for our study. Each of the visualizations originated from a reputable source—such as the CDC or the WHO—and was created with no intention of promoting conspiracy theories, however has been widely

misinterpreted on social media. To select the charts, we primarily drew on the data set of COVID-19 visualizations shared on Twitter (X) collected as part of our previous work [6], additionally supplemented by more recently-posted charts by some of the prominent accounts from our data set.

For each of the charts, we use the original source to identify the associated intended interpretation, or the *fact*, and the surrounding discussion on social media to establish a common misleading interpretation, or the *myth*. The social media posts sharing the myth interpretation are linked from the text and listed in Table 6.1. Afterwards, we discuss the reasoning error that lead to the myth reading and the process of developing fact-checking labels for the visualization.

We identified these charts primarily for the purposes of our label evaluation study, described in Section 6.4. However, we also intend to use this section to highlight the origins of common visualization-supported misinformation arguments, the role that government agencies and news outlets may play in the spread of these myths, and the challenges of balancing multiple visual communication goals.

6.3.1 COVID-19 Outbreak in Barnstable County

The visualization, shown in Figure 6.2, was part of a CDC report published at the height of the COVID-19 pandemic in 2021 [96]. The bar chart showcases a rapid rise in COVID-19 positive tests following 4th of July festivities and gatherings, followed by a drop-off a couple of weeks later. Interestingly, the authors break out the data by vaccination status, making it clear that the vast majority of identified cases were among fully-vaccinated individuals.

As inferred from the original report [96], the main point of the report and its figures was to showcase the insurgence of the then-novel Delta variant of COVID-19. It spread quickly among communities, and researchers also felt that it is important to highlight the fact that previously-vaccinated individuals can get sick and be contagious. The researchers came to the conclusion that jurisdictions should impose "masking in indoor public settings regardless of vaccination status" and continuously promoted vaccination as an important measure against severe illness and death. The report also suggested that the researchers were aware at the time of writing that these data may be misinterpreted. In the limitation

section they describe that "data from this report are insufficient to draw conclusions about the effectiveness of COVID-19 vaccines against SARS-CoV-2." They also discuss that although we are unable to estimate the base rate of vaccination among the attendees, at the time at least 69% of Massachusetts residents were fully-vaccinated.

Alas, these warnings and explanations did not reach everyone. This chart got picked up by misinformation actors online and was actively shared on social media to promote the idea that "the vaccinated are the real super-spreaders" (#1 , #2), along with other figures from same report (#3). The sensationalized story was also picked up by many mainstream news outlets, such as The Washington Post [221] or CNBC [222] and lead the local officials to issue a statement on the fact that the outbreak has been quickly contained and resulted in no deaths (#4).

> **Fact:** Based on the chart, cases may quickly rise following large gatherings.
> **Myth:** Based on the chart, vaccinated people are more likely to get infected.

There are many plausible and useful explanations as to why there are so many vaccinated cases in this data set, including the fact that vaccines do not fully protect from catching the virus, or that these data do not show the severity of disease between the subgroups. But at the core of this myth lies the issue of the base rate fallacy: we simply do not know the vaccination prevalence of the underlying population. The report cites the figure of 69% of Massachusetts residents being vaccinated; however, it is highly likely that the in-person attendees of an event in Cape Cod—a highly affluent and educated region with a large elderly population—were vaccinated at even higher rates. And the higher the base rate of vaccination is, the more likely it is that many of the recorded cases would be among the vaccinated. In any case, the base rate is *unknowable*, making any relative comparisons meaningless.

Building on our guidelines discussed in Section 6.2, we developed the following labels for this chart:

> **Warning:** We can't determine the relative risk of contracting COVID-19 from this chart because it doesn't account for vaccination rates in the population.
> **Counter:** As more people get vaccinated, it is more likely that the majority of cases would be among the vaccinated.
> **Guide:** This chart tracks the surge of COVID-19 cases after big summer events in Barnstable County. It highlights how quickly the Delta variant spread and reinforces why precautions like masking were still important, even in highly

vaccinated communities.

6.3.2 Influenza Surveillance

The next chart was created by the Global Influenza Surveillance and Report System (GISRS), a network of laboratories providing influenza data to the World Health Organization [223]. The GISRS dashboard exists to enable global sharing of flu data, monitoring for early outbreak detection, and primarily for the "selection of suitable vaccine viruses, in relation to the manufacture, regulation and administration/distribution of influenza vaccines." [224]. The most common reason for the low effectiveness of flu vaccine campaigns in any given season is the mismatch between the targeted flu subtypes and the vaccine composition [225], making it invaluable to track which subtypes are emerging globally.

In other words, although the GISRS data do not account for every single case of flu in the world, its primary goal is to identify the composition of emerging subtypes of flu. The COVID-19 pandemic-related measures—such as compulsory masking, quarantines, and travel restrictions—have disrupted the behavior of the flu in 2020 on an unprecedented scale. Researchers found that, following this unprecedented "natural experiment" of COVID lockdowns, flu cases have virtually disappeared, giving us the first widespread evidence of the efficacy of nonpharmaceutical interventions against the flu virus [226] and even eliminating certain strains of the flu [227].

This causal effect of COVID-19 measures, however, is not something readily seen from the data and depends largely on thorough analyses and virus sequencing. Many COVID skeptics promoted the idea that the disappearance of the flu in 2020 is evidence of the fact that the flu is simply mistaken or even "rebranded" as COVID. It is highly unlikely that flu data have been purposefully or erroneously not collected—the lack of flu cases in 2020 lead to high uncertainty about which strains of flu will become dominant in the future and disrupted vaccine selection [228]. However, in isolation, the chart from the WHO shown in Figure 6.3 may look like a compelling argument that flu cases stopped getting counted. The specific chart from the WHO was widely shared and reposted on social media, promoting the conspiracy seemingly backed up by official WHO data (#5, #6), along with many other similar visualizations from either the WHO (#7, #8, #9) or other outlets (#10, #11, #12).

> **Fact:** Based on the chart, the vaccine for this season should include both A and B strains of flu as both are common.

Myth: Based on the chart, flu infections may have been erroneously misclassified as COVID.

The primary reasoning error that lead people to the myth interpretation of the chart is assigning incorrect causal relationship to a salient feature of the chart. The cases sharply dropped around the start of the COVID-19 pandemic, but the chart or the data provided no reason why, leaving the viewer to fill in the gaps themselves through the lens of own beliefs and biases. Conflating correlation with causation is a widely known and reported human bias [229], [230] that is difficult to debunk. It is especially difficult to debunk timely—during a developing situation, like an outbreak of a pandemic—as scientists could only speculate about the real reasons for certain data phenomena. Having the benefit of hindsight, we developed our counter fact-checking label to describe the now-prevailing hypothesis that COVID-related social distancing lead to the drop of flu cases. The other labels do not speculate on the real reasons: the warning reminds the audience to not make causal conclusions from this slice of data, while the guide reinforces the primary reason for the existence of this chart.

Warning: Avoid drawing causal conclusions from these data patterns, as multiple factors may influence the trends observed.

Counter: The observed drop in cases is likely a result of preventive measures in place due to COVID lockdowns.

Guide: The WHO surveillance data is collected in order to track the dominant strains of flu. This allows scientists to identify the most effective vaccine type for the season and allocate resources accordingly.

6.3.3 Improving Communication on Breakthrough Cases

The chart shown in Figure 6.4 was originally part of an internal presentation at the CDC titled "Improving communications around vaccine breakthrough and vaccine effectiveness" [231]. The misrepresentation of this chart is ironic: as implied by the title, the goal of the presentation was to strategize on the public communication of vaccine effectiveness. The slides show that breakthrough COVID-19 cases are becoming more prevalent as more people get vaccinated, which means that public health officials should preempt panic by reminding people that vaccines are safe and effective and these data can appear misleading.

This chart is part of a series of similar examples where internal, unpublished data from the CDC was obtained by journalists and sensationalized [232]. Similarly, prelim-

inary estimates on the contagiousness of COVID-19 [233] or the previously-discussed Massachusetts outbreak [221], [222] were reported on without due statistical or analytical nuance. The chart from Figure 6.4 also received particular attention on social media, leading to numerous posts arguing that the CDC internally agrees that vaccines do not work (*#13*, *#14*).

This phenomenon highlights a particularly difficult problem of chart misinterpretation: when such internal, unpublished data leaks into the public, it makes it seem that these data are being actively hidden and that its publication exposes a conspiracy. Additionally, charts from internal presentations are even more likely to not be designed with the general public in mind and to not include every single necessary caveat on every slide. In the context of the presentation, this chart was hardly misleading. For instance, the preceding slide in the presentation included results of a controlled study that showed that vaccination lead to a 25-fold reduction of the likelihood of hospitalization [231], contextualizing the data in Figure 6.4.

> **Fact:** Based on the chart, breakthrough cases become more common as more people get vaccinated.
>
> **Myth:** Based on the chart, vaccines may lead to more hospitalizations.

Similar to the Massachusetts outbreak chart in Figure 6.2, the risk of misinterpreting this chart stems from the lack of precise and visualized data on the base rate of vaccination in the underlying population. A reminder about this fallacy constitutes our warning label. An additional caveat mentioned in the original presentation is that the high rate of hospitalization among vaccinated is also an artifact of vaccination rollout: the elderly and sick were vaccinated first and are more likely to experience severe symptoms [231].

> **Warning:** We can't determine the relative risk of contracting COVID-19 from this chart because it doesn't account for vaccination rates in the population.
>
> **Counter:** The rise in vaccinated hospitalizations primarily reflects increases in vaccine coverage and higher coverage in older adults.
>
> **Guide:** This chart is from a CDC report on improving communication about vaccine effectiveness. As more people get vaccinated, it is expected that cases and hospitalizations among the vaccinated will rise despite no change in vaccine effectiveness. This highlights the need to update public communication on breakthrough cases and emphasize that they are not a reason for panic.

6.3.4 History of Climate on Earth

The chart showing the history of the Earth's climate (Figure 6.5) is based on a research paper published in Science [234] and popularized by an article by The Washington Post [235]. The main point of the study was to share the novel data set constructed by complex statistical climate models, as well as the insights we can glean from these new data. Overall, the paper does not directly touch on the modern climate change debate. And it doesn't have to—there is no debate on the reality and sources of climate change among climate scientists [144], by whom and for whom this paper was primarily published.

The full article that the chart in Figure 6.5 comes from did (in our view) a commendable job recapping the Science paper for a popular science audience and adds the much-needed context on climate change. Throughout the Washington Post article, the authors zoom in on various parts of the timeline that lead to mass extinction events, and specifically highlight how unprecedented the recent changes to climate are [235]. All of this, unfortunately, did not just not make it to the headline and the teaser—the full article is behind a paywall. As a result, the view in Figure 6.5 is all that non-subscribers can see.

Hence, it is unsurprising that a quick glance at the chart might suggest that not only is climate change overblown, but also that we are actually living through the coldest time in the last 485 million years. This specific visualization has since been repeatedly promoted by climate change skeptics on social media (*#15*, *#16*) and was subsequently featured in an episode of the Joe Rogan Experience podcast to promote climate change skepticism [236].

> **Fact:** Based on the chart, climate has varied a lot over the history of the Earth.
>
> **Myth:** Based on the chart, fears of climate change are exaggerated since it has been much hotter before.

Several reasoning errors may lead to the conclusion that this chart disproves climate change. We could interpret this as cherry-picking: the highly zoomed-out scale of the chart obscures a trend that would otherwise be obvious on a different scale, which is a common tactic in time series charts [8]. Another related issue is the naturalistic fallacy [237] and the "is-ought" problem [238]: the climate *has* drastically changed before, which means that climate change is *natural*, which means that climate change *ought* to happen, and maybe perhaps is even a good thing. Lastly, it is often unclear how "bad" a certain amount of temperature increase is, leading to the problem of an arbitrarily set threshold for

goodness [6]. Consequently, a chart showing large swings in Earth's temperature allows for this line of thinking: if it's been hotter before, then why is this increase alarming?

Other charts that showcase the results of long-term climate modeling shared by the NOAA [239] have been previously picked up by climate change skeptics. The scale of the time shown on the horizontal axis is so long that it doesn't just obscure recent climate change trends—it makes the entirety of human existence not noticeable. As such, our warning label alerts the audience of the issue of zooming out.

> **Warning:** Zooming out on the entire time series may obscure the important recent trends in climate.
>
> **Counter:** Although the climate has varied over the course of history, we are currently experiencing the fastest recorded increase in temperature, which is a concern.
>
> **Guide:** This chart presents the most detailed reconstruction of Earth's temperatures over the past 485 million years, revealing a history of dramatic shifts in climate before humans existed. This research provides a deeper understanding of how past environments influenced life on Earth and opens new questions about the mechanisms driving climate stability and change in the current geological stage.

6.3.5 Map of Reported COVID-19 Cases

The last chart we constructed labels for, seen in Figure 6.6, is an exemplar of a commonly-misinterpreted type of chart: simple thematic choropleth maps can very commonly mislead their audience by visualizing non-population-normalized data leading to conclusions about the relative rate of incidence. Previous work explored visualization techniques to de-bias such maps [240], however the majority of maps created for the general public use simple metrics: either raw counts or ones normalized by population in the region.

Such maps were popular during the COVID-19 pandemic, both normalized by population and those showing raw counts. Both normalized and non-normalized representations are perfectly valid, however should be used for distinct reasons. As discussed by Ge et al. in their framework for mitigating reasoning errors in visualizations, the problem of normalization is directly related to the policy question at hand: if you're interested in relative risk you are likely to make reasoning errors with a non-normalized map [79]. The chart in Figure 6.6 was published by the CDC for general transparency and reporting purposes. We have observed similar maps used both correctly (to discuss resource

allocation) [241], but also misinterpreted by social media responses (to estimate risk of contracting COVID-19) (#17, #18). COVID skeptics also commonly tie in other correlated factors when interpreting non-normalized case maps, such as what political party governs the state (#19) or whether a country had a large vaccination program (#20), leading to insights nearly entirely explainable by population numbers.

> **Fact:** Based on the chart, more medical resources should be allocated to states colored in purple than those in yellow.
>
> **Myth:** Based on the chart, you are more likely to contract COVID in states colored in purple than those in yellow.

Since the non-normalized map highly resembles a simple map of the US population, the main goal of the labels is to remind people to not make relative risk comparisons: you can't really make these estimates just by looking at this chart. The issue here is similar to the base rate fallacy in some of the previous figures.

> **Warning:** We cannot determine the relative risk of contracting COVID-19 based on this map, since the data is not normalized by population count.
>
> **Counter:** The states with most reported cases are those with higher population counts, while relative risk depends on cases per capita.
>
> **Guide:** This map helps CDC officials and responders quickly identify where case numbers are highest, guiding decisions on deploying medical personnel, medication, or testing kits. It's a tool for prioritizing support where the demand is greatest, ensuring that overwhelmed areas get the assistance they need.

6.4 User Evaluation: Do Labels Work?

In this section, we describe our user evaluation of fact-checking labels on visualizations. We first cover the methodology used, before presenting the numerical results of our study as well as a thematic analysis of participants' text responses.

6.4.1 Methods

This section outlines the methodology employed to first construct the study design and conduct the evaluation, and then to analyze the results both quantitatively and qualitatively.

6.4.1.1 Study Plan

To uncover the influence of our labels on people's insights and trust in the visualizations, we conducted a crowdsourced experiment using Prolific and the ReVISit study plat-

form [194]. The experiment was reviewed by the University of Utah IRB and deemed exempt from full board review (IRB 00175486). Before the main experiment, we conducted 2 pre-pilots with graduate students at our university, followed by a 20-person crowdsourced pilot. Following the pilots, we adjusted the language in some of the prompts, added an additional introductory tutorial page to the survey, estimated the average completion time to appropriately compensate the participants, and conducted a power analysis to estimate an appropriate sample size for the main study. For the main study, we recruited 100 crowd participants, each paid $5 for a median completion time of 26 minutes. Before conducting the main study, we submitted a preregistration of our analysis, available here.

In the study, each participant was shown the series of 5 charts described in Section 6.3 in random order. To improve consistency across chart conditions and standardize the existing non-label text, we made minor adjustments to the visualizations, such as adding a title to the map chart and removing an annotation from the internal CDC chart. One of conditions had no label, with the rest having one of 4 possible label conditions described in Section 6.2: *warning*, *counter*, *guide*, and *all three*. To make sure that participants consider both the data and the label together (but, at the same time, not explicitly ask them to do that as to not draw more attention to the label than they would naturally pay), we enclosed both in a visible frame and referred to them together as a "slide," as shown in Figure 6.6. For each of the 5 slides, we asked the participants a series of questions. First, they were tasked to describe all insights and observations they can make based on the information, similar to the approach in the Quadri et al. study of chart interpretations [205]. Next, participants were asked to rate the slide's trustworthiness on a 0–100 scale that included plain language labels along the scale, similar to the survey used in Padilla et al. [174]. They were also invited to explain their numerical choice in a free text response box. Lastly, participants were asked to rate their agreement with the statement that the slide supports the associated *fact* and *myth* we had identified for each of the charts, previously described in Section 6.3. Their agreement was measured on a 7-point Likert scale, ranging from Strongly Disagree to Strongly Agree. After performing these tasks for each of the 5 stimuli, we explained to the participants the various labels they had seen and asked whether they noticed them and whether they feel like the labels influenced their responses. Lastly, to account for the politicized and potentially controversial nature of some of the charts, we

collected the respondents' political affiliation using a 7-point ideology scale [242]. The whole experiment as well as the source code of the experiment are accessible here.

The preregistered hypotheses for our primarily analysis were:

- **H1:** The **presence** of a fact-checking label would affect the interpretation. Specifically, a label would:
  - **H1a:** increase trust
  - **H1b:** increase agreement with fact
  - **H1c:** increase disagreement with myth
- **H2:** The **different types** of fact-checking labels would affect interpretation differently. Specifically,
  - **H2a:** Label conditions (Warning, Counter, Guide, All) will significantly differ in levels of trust and agreements.
  - **H2b:** The All label condition will show higher magnitude (increased trust, agreement with fact, decreased agreement myth) effects than others.

6.4.1.2 Quantitative Analysis Methods

Following our preregistration, we analyzed the impact of label condition on the trustworthiness and agreement variables by fitting mixed-effects linear models. Our analysis scripts are available in the supplemental materials. We fit models with the label condition as fixed effect and participant and chart condition as random effects. This allowed us to account for the nested structure of our within-subjects data, where multiple measurements come from the same participant and across different chart conditions. By including participant and chart condition as random effects, we get robust estimates of the label effects and control for potential individual differences and variability between chart types, which is expected given highly different charts. We then compared the models using likelihood ratio tests to determine the significance of adding the label condition relative to the reduced model. Beyond the primary preregistered analysis, we also conducted an exploratory analysis, controlling for the ideology scale results.

6.4.1.3 Qualitative Analysis Methods

We performed an exploratory qualitative analysis by reviewing the contents of the text responses in the survey to uncover the reasons why labels do or do not affect interpreta-

tions. We followed the methodology of *template analysis* [137] to construct a codebook of codes to annotate the data. Our primary interest in uncovering the attitudes towards, and effects of, the labels, informed the lens through which we analyzed the data. As a result, we focused on the text relevant to the interpretation of the label, the attitude towards the label, or the upholding or debunking of the myth or reasoning error. Consequently, we did not assign codes that highlight unrelated data-driven insights from the chart. Additionally, because many of the codes reflect participants' self-reported thoughts or aspects of the labels we had not specifically queried in the survey, we do not report the prevalence of any particular label. For instance, even if only one participant mentioned a certain important theme it is highly likely that it applied to many others but was simply not mentioned. Instead, we consider all codes to be of equal importance and to reflect an informative dimension of people's perception of fact-checking labels on charts.

To construct the codebook, the first author conducted multiple iterations of assigning codes, consolidating similar codes, and reorganizing the hierarchy of the codebook. At the end, all authors met to discuss the final codebook and whether all codes are sufficiently clear and distinct. We include the audit changelog of the codebook that contains all of its iterations in the supplemental materials.

6.4.2 Numerical Results

Figure 6.7 shows the numerical responses of the survey by label condition, grouped across all participants and chart conditions. Nearly all of the hypotheses of our primary analysis yielded null results. Controlling for participant and chart effects, the likelihood ratio tests showed no effect of the presence of the label on trust ($\chi^2(1) = 0.11, p = 0.74$), agreement with the fact ($\chi^2(1) = 0.01, p = 0.93$) or the myth ($\chi^2(1) = 0.34, p = 0.56$). When breaking out the label into the conditions, the type of label similarly had no significant effect on trust ($\chi^2(4) = 2.73, p = 0.60$) and agreement with the fact ($\chi^2(4) = 5.17, p = 0.27$). Although above the common 5% α-level threshold, we observe a small effect of the type of label on agreement with the myth ($\chi^2(4) = 9.14, p = 0.058$).

To further explore the differences between label conditions and the agreement with myth, we used Tukey's HSD post hoc test to conduct pairwise comparisons. Although not statistically significant at the 5% level, we observed that the comparison between the

Guide resulted in an average 0.63 points higher agreement with the myth on a 7-point scale compared to *All* labels condition ($t(388) = 2.66, p = 0.062$). This small effect is also apparent in Figure 6.7. We interpret these results to be a sign of the *implied truth effect* [28]: as the *Guide* label did not explicitly address the myth, it likely lead participants to assume that the myth is more true since they were not warned about reasoning errors in the label like in the other stimuli. Although this result is partially an artifact of our within-subjects design, it is an ecologically-valid outcome as people are likely to see many various visualizations online.

We also conducted an exploratory analysis to examine the effects of participants' self-reported political leaning on a 7-point ideology scale. Figure 6.8 shows the results split up by ideology (1 or 2 being *Liberal*, 3–5 being *Moderate*, and 6 or 7 being *Conservative*). The likelihood ratio tests of additionally controlling for ideology type showed a weak effect on agreement fact ($\chi^2(2) = 5.63, p = 0.06$), and no effect on trust ($\chi^2(2) = 2.54, p = 0.28$). Importantly, we observe a very strong effect of ideology on agreement with the myth ($\chi^2(2) = 10.22, p = 0.006$). Pairwise post hoc Tukey HSD tests reveal that Liberal participants agree with the myth by 1.1 fewer Likert points compared to Conservatives ($t(97) = 3.14, p = 0.006$) and by 0.55 fewer points than Moderates ($t(97) = 2.14, p = 0.09$).

6.4.3 Thematic Analysis

We conducted a thematic analysis to better understand the underlying reasons for the inefficacy of labels against reasoning errors. The final codebook we used can be seen in Figure 6.9, while this section presents an overview of the most important themes. We distilled the themes by juxtaposing the ways in which labels were liked and disliked, considered or ignored, seen as helpful and not so. Throughout the text, we highlight quotes taken directly from our participants' responses. All of the text responses can be accessed in our supplemental materials.

6.4.3.1 Labels Provide Context but Seen as Lesser than the Data

Participants often referenced the labels as having helped them better understand the data in general or certain nuances about the data. The participants described that the context of the vaccination rates was "*further explanation that was needed*" and that the labels

offered "*omissions worth pointing out.*" Aside from adding information not seen in the graph, others found that the labels "*made the graphs more understandable*" and "*the data easier to interpret.*"

At the same time, others did not allow the text to override any observations they had made themselves by looking at the data. Many participants found the data (and consequently the conclusions they end up making from it) to be the more important source of information of the two. Some participants reported not looking at the labels and instead being "*more interested in the [...] logical interpretation of the charts*" and "*critiquing the data in a standard way taught in [their] degree.*" Apart from the hierarchy of importance, the chart is also seen as more prominent and noticeable than the label. As one participant put it plainly: "*visual data is just more interesting to look at.*"

6.4.3.2 Labels Add Transparency That May Undermine the Data

Labels have often been reported as having added transparency to the analysis and show care put in from the authors' end. Participants reported considering the information trustworthy because of the labels "*offering an explanation*", "*showing disclaimers*", "*pointing out limitations*". The transparency afforded by the labels positively reflected on the author and source of the data, it "*demonstrated the data provider's awareness of these issues*" and "*implies that they're being honest.*" Participants who praised transparency are those generally aware of possible limitations of data and its presentation, with one noting that the label "*made [them] trust the data more because it was guarding against interpretation of the data as an absolute.*"

Other participants, however, interpreted any data limitations to mean that the data are just plain useless. Participants who rated the visualizations as not trustworthy often described the limitations exposed in the label as the primary reason why. Others explicitly described that labels "*lead [them] to not rely on the visualization*" or to mean that "*this data cannot be relied on to measure anything accurately.*" And although all the labels have been limited to at most describing a potential statistical fallacy in interpretation, some participants misread or misinterpreted the labels to say that the "*data could be inaccurate or even 'ditched'*" and dismissed the information altogether.

6.4.3.3 Labels Spur Skepticism but Others May Not Read Them

Besides adding transparency, many participants found labels to be a helpful prompt to pause and think critically about the data and its presentation—perhaps even to consider other indicators of trustworthiness, such as the source. One participant noted that the label made them "*think about what might make a graph not accurately represent the whole picture even if it is accurately representing one aspect of the data.*" However, while the labels encouraged individual reflection, some participants remained skeptical that others would engage with them in the same way.

Even when participants interpreted the data correctly, they still saw its potential to mislead, as the chart "*can be misunderstood by some.*" One summarized this concern: "*I feel like people may draw the wrong conclusions from this.*" Despite recognizing how labels shaped their own understanding, many still viewed the visualization as fundamentally untrustworthy—especially given how they had seen it used in broader discourse. As one participant put it, "*this visualisation is often used deliberately to downplay the effects of climate change,*" reinforcing their distrust regardless of the label's presence.

6.4.3.4 Labels Explain the Visuals but Also Conflict with Them

As previously described in Section 6.3, the data visualizations used in our study were deliberately chosen to be those that were not designed with the intention to mislead but amenable to incorrect interpretations. In other words, they were designed to show something but through reasoning errors may end up showing something else. As our labels attempted to warn against the erroneous interpretation they inevitably exposed the duality of the presented data: the fact that the same chart can be interpreted in very different ways.

Our study participants varied in their reaction to this fact. While some welcomed the explanation of the secondary visual elements they have observed, others felt the labels to be conflicting with the chart. Participants described the labels as not clarifying but rather including "*contradictory information.*" As a result, this conflict lead a subset of our participants to further dismiss the label. The perceived contradiction exacerbated the feeling that the data are more factual and superior to the text which "*contradicted real world data [and]*

gave no further evidence." Some participants disliked the ambiguity introduced by the label in contrast to, in their eyes, unambiguous patterns seen in the data: "*conclusions cannot say whether it's accurate or not despite a clear pattern of increased deaths [and] hospitalizations.*"

6.4.3.5 Labels Shape Thoughts but Also Are Perceived as Biased

Many individuals quoted the text from the labels verbatim when describing the charts or reflecting on their trustworthiness. This suggests that they relied on the labels not only to understand the charts but also to articulate their thoughts. While it could indicate a degree of anchoring on the labels, some participants explicitly stated that the labels helped them formalize their thoughts, making it "*easier to answer questions*" and "*giving immediate words to use*".

While some participants found the labels helpful in shaping their thoughts, many others felt uncomfortable by the text's influence. Moreover, they compared the biasedness of the labels with their perception that, in contrast, the data and the chart were *not* influencing them. Thus, unlike the (supposedly) infallible and naturally occurring data, the labels were seen as explicitly man-made and potentially agenda-driven: "*They want to put caveats and footnotes on the graph, but the data is there that this is a naturally occurring thing.*" In some cases, this influence was even perceived as antagonistic, with participants actively resisting the labels' potential bias: "*While I tried to take the label information into account, I do not think I let it impair my judgement.*"

6.4.3.6 Labels Align Beliefs but Some Prefer Own Explanations

While some participants were uncomfortable with the idea of data or labels being biased, others welcomed text that helped them align what they see in the data with their own beliefs. Some participants praised the ability of labels to explain away alarming observations they might have made upon their first glance and "*back up the initial hesitations with the data*". Labels were also reported to "*promote trust in conclusions [one has] already drawn*".

However, the explanation in the labels didn't resonate with everyone. Many participants directionally agreed with the label but cited completely different reasons. For

instance, while some remained unaware of base rate fallacy when assessing vaccine effectiveness, they described additional reasons for skepticism not mentioned in the labels. These included vaccines being "*not really made with the intention of giving you 100% immunity*", data being limited to "*one isolated event*", and lack of information on "*how many were tested*", as well as factors such as seasonality. Similarly, while some may have overlooked the nuance that the map was not normalized by population, they explained the higher numbers in some states by saying that blue states conduct more testing, larger states are "*less inclined to follow COVID rules*", are more "*popular for tourists to visit*", or speculating that "*the transmission can mostly be associated with immigrants*".

This theme strongly resonates with the findings of our previous analysis of social media responses to COVID-19 visualizations [7], which showed that people find a variety of anecdotal and methodological caveats in an attempt to counter the visualization they see. Similarly, here we observe participants drawing on their own knowledge and speculation to explain the data while brushing the label aside.

6.5 Discussion

In this section, we examine what our results reveal about preventing misinterpretation in data visualization. We also consider whether misinterpretation is always a problem.

6.5.1 How to Prevent Misinterpretation of Visualizations?

Our findings indicate that labels are ineffective at preventing misinterpretations, highlighting the broader challenge of guiding audience interpretations of data. Some participants described how they were taught to interpret data literally, as neutral information, reinforcing the need for broader educational efforts to address overconfidence in data——and the medium of visualization—as an objective truth. Changing these deeply ingrained assumptions is a long-term challenge. In the short term, however, it may be more effective to instead leverage the fact that data are convincing and proactively shape visual communication. Borrowing from learning objectives and anti-objectives [84], when making visualizations in good faith designers should anticipate and consider which interpretations they want to support or discourage when creating a chart.

The other finding—that labels are strongly perceived as human-written and agenda-

setting—may also present an opportunity. Rather than avoiding subjectivity, visualization designers may be able to embrace it, embedding their voice in annotations that engage audiences as narratives rather than directives. Interpretive writing principles, such as those used in museum wall texts, may offer useful guidance. Literature on interpretive writing [243], [244] echoes our findings and recognizes that audiences engage with text selectively and are aware of its human authorship—this awareness encourages seeing writing not as a medium for data caveats but rather one for reflection and storytelling.

Lastly, literature on adapting scientific figures for popular science communication explores how to reframe and contextualize academic figures, which often lack broader context [245]. Recontextualizing scientific knowledge involves intentional design choices such as simplifying, emphasizing, annotating, and putting them into context of the relevant political discourse. For example, the CDC could prepare their outbreak chart for public consumption by annotating or visualizing the estimated base rate of vaccination. Our findings suggest that, in an era of abundant information and the ethos of "doing one's own research," authors—whether or not targeting a popular audience—should consider the broader public when designing their figures.

6.5.2 Should We Prevent Misinterpretation of Visualizations?

The question of whether misinterpretation should be actively prevented at all is complex. Many participants felt uncomfortable with the idea that only one "correct" interpretation should be encouraged, while other (potentially factually incorrect) ones should be discouraged. This mirrors debates in art and literary criticism about whether an artist's intent matters when interpreting their works. If an author's intended message differs from what the audiences take away, is that a failure; a sign of a rich, important, and engaging material; or a natural consequence of people finding other uses for a material beyond its initial intent [246]?

Perhaps it is equally valuable to read "with the grain" to understand the author's intent and point of view, as it is to read "against the grain" to uncover the hidden meanings [247]. Adapting insights from feminist theory to visualization [248], [249] may highlight how reading against grain—perhaps even when factually incorrect—can expose underlying gaps or assumptions in our messaging or data collection. Misinformation narratives, for

example, do not simply reflect misunderstandings; they reveal what people care about, where knowledge gaps exist, and how communication strategies might be failing. For instance, do charts zooming out on climate change reflect people's unawareness of cherry-picking fallacies and a failure of data literacy? Or simply the fact that measuring climate change in degree Celsius deviations fails to resonate with the public and does not directly translate to their lived experiences? Instead of treating misreadings purely as errors, we could recognize them as signals of the audience's perspectives and learn from them.

6.6 Limitations and Future Work

Our study prioritized ecological validity by focusing on politicized topics which likely influenced responses, given the observed differences by ideology. Although this choice reflects real-world conditions where interventions are most needed, future work could explore whether stronger effects emerge in less controversial contexts. Additionally, while our measures capture people's agreement with statements, alternative evaluation metrics could be insightful, such as tracking belief shifts or recall. An interesting avenue for further research is to investigate how subtle variations in readability, tone, or wording could shape the effectiveness of labels.

6.7 Conclusion

Our findings highlight that data visualizations carry more rhetorical weight than the analytical nuance needed to interpret them accurately. Simply put, visualizations are eye-catching, easy to grasp, and seem objective. Meanwhile, fact-checking labels are perceived as extraneous, easy to disregard, and biased. This contrast points to the need to design visualizations that actively address and counter related data misconceptions. The perceived neutrality of data is not an inherent truth, but rather a powerful rhetorical device that can shape interpretations and support specific agendas.

Table 6.1: Social media posts referenced in the paper under Citation IDs, their original platforms, and links.

Citation ID	Platform	Post ID	Link
Post #1	Twitter / X	1422687139150577676	↗
Post #2	Twitter / X	1423405168670826506	↗
Post #3	Twitter / X	1425299466324647939	↗
Post #4	Twitter / X	1421139894257266691	↗
Post #5	Twitter / X	1426900435537960966	↗
Post #6	Twitter / X	1352365346548838405	↗
Post #7	Twitter / X	1317061948228460546	↗
Post #8	Twitter / X	1320745472957599746	↗
Post #9	Twitter / X	1323657390911356936	↗
Post #10	Twitter / X	1317134214073036800	↗
Post #11	Twitter / X	1319998103416655874	↗
Post #12	Twitter / X	1352992422943481857	↗
Post #13	Twitter / X	1421149764800946178	↗
Post #14	Twitter / X	1424949884470366209	↗
Post #15	Twitter / X	1875197327897698455	↗
Post #16	Twitter / X	1879227547705372872	↗
Post #17	Facebook	10157969471451026	↗
Post #18	Twitter / X	1260689978176286723	↗
Post #19	Twitter / X	1421794297876291587	↗
Post #20	Twitter / X	1421794297876291587	↗

[h]

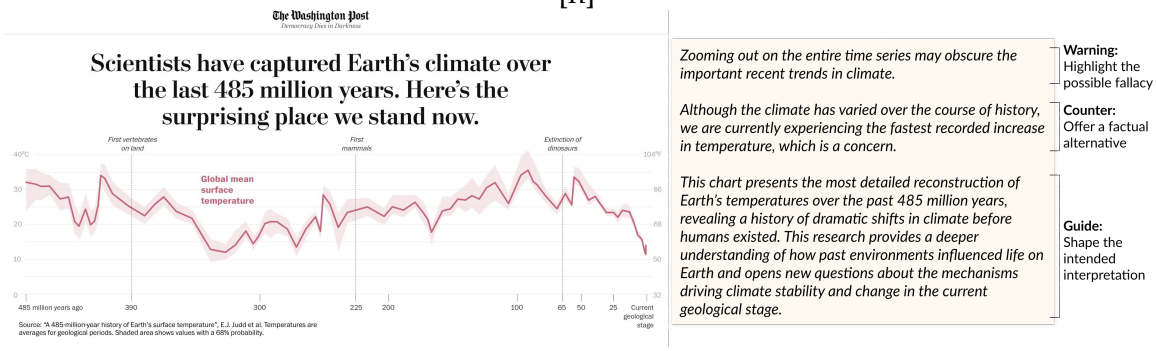


Figure 6.1: A climate modeling visualization, originally from an academic paper [234] and popularized by an article by The Washington Post [235], that has been commonly used as evidence by climate change denialists. We design and evaluate three main components of a fact-checking label intended to discourage misinterpretation of the data: warning against the fallacy, countering with a fact, and guiding the intended interpretation.

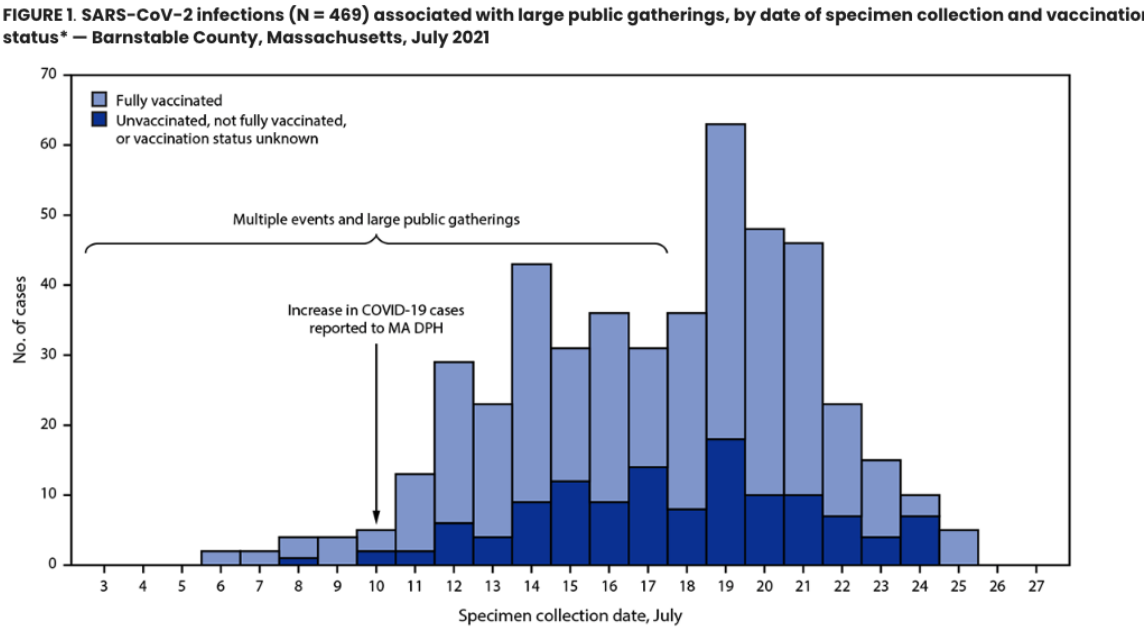


Figure 6.2: A figure published in a CDC report [96] that was widely used on social media to suggest that vaccinated people are more likely to get infected with COVID-19, thus promoting vaccine skepticism.

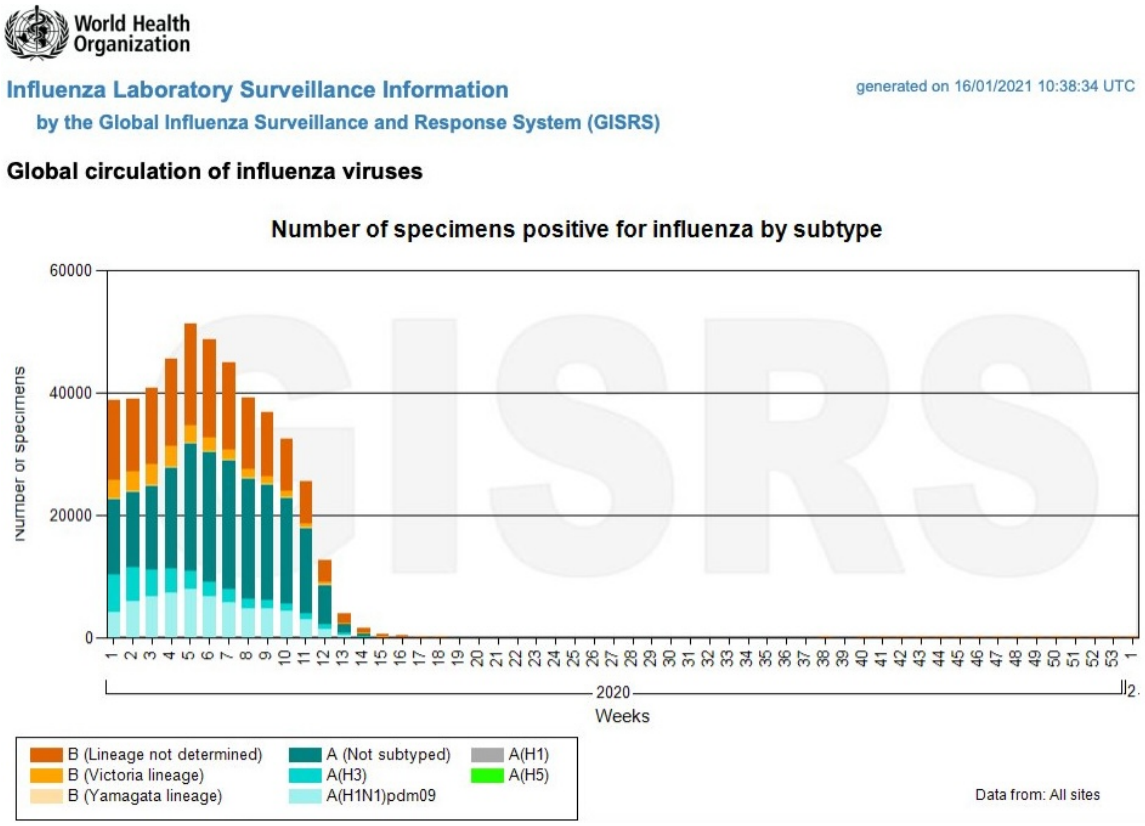


Figure 6.3: An influenza dashboard by the WHO [223] that was widely used on social media to argue that the flu was misclassified as COVID, furthering COVID-19 denialism.

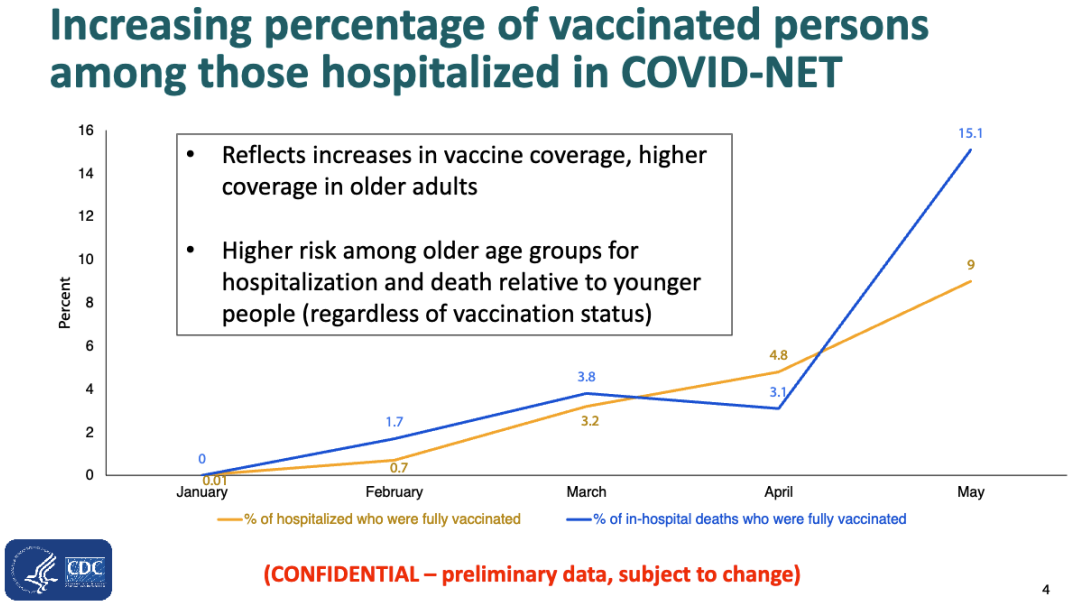


Figure 6.4: A figure from an internal CDC presentation [231] that was widely used on social media to suggest that vaccinated people are more likely to get hospitalized, thus promoting vaccine skepticism.

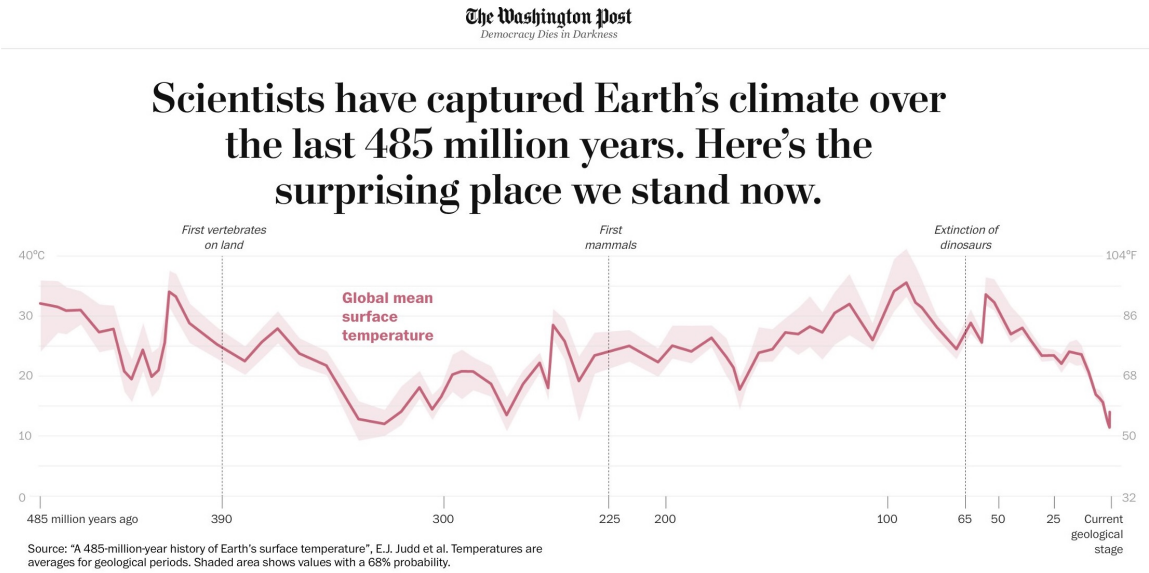
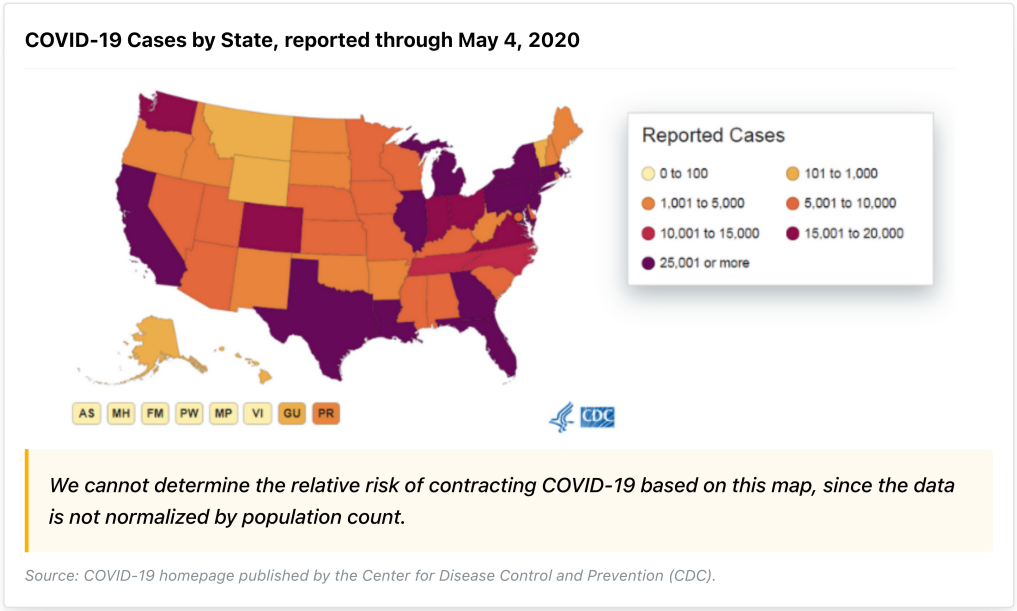


Figure 6.5: A figure from a scientific paper [234] and reproduced in a popular science article [235] that has been widely misinterpreted on social media to downplay the existence and the risks of anthropogenic climate change.



Consider the two statements below. For each, do you agree that the slide above supports the statement? *

Based on the above, more medical resources should be allocated to states colored in purple than those in yellow.

Strongly Disagree | Disagree | Somewhat Disagree | Neither | Somewhat Agree | Agree | Strongly Agree

Based on the above, you are more likely to contract COVID in states colored in purple than those in yellow.

Strongly Disagree | Disagree | Somewhat Disagree | Neither | Somewhat Agree | Agree | Strongly Agree

Figure 6.6: Screenshot of our experiment interface. The participant sees a map showing COVID-19 cases not normalized by population with a *Warning* label in the slide. The participant is asked to indicate their agreement with the *fact* and *myth* statements.

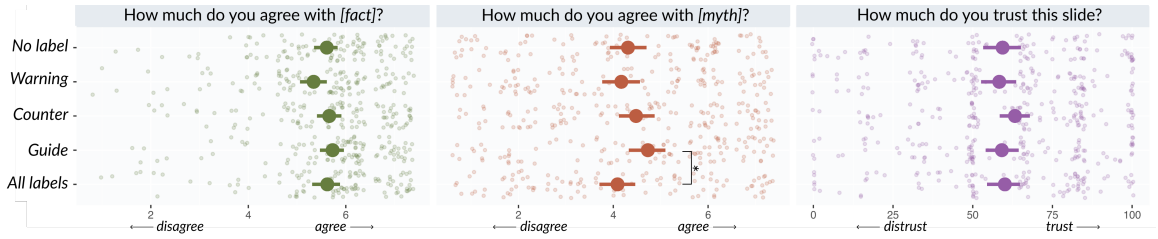


Figure 6.7: Labels do not substantially influence agreements with facts, myths, or influence trust. Shown are participants' responses to numerical questions in the study, with means and bootstrapped 95% confidence intervals ($n = 500, 100$ per condition). The results show little variation by condition, indicating no overall effect of labels. However, the *Guide* label may exhibit the implied truth effect: by not addressing a reasoning error it may inadvertently signal that the myth is correct [28]. As a result, we observe higher agreement with the myth statement when compared to *All Three* labels together ($p = 0.06$). Note that while our statistical models account for chart and participant effects, this figure presents results aggregated across all chart conditions.

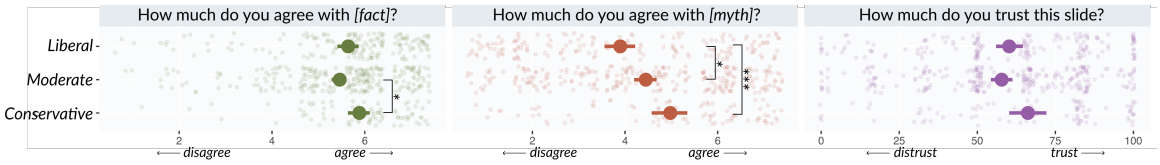


Figure 6.8: The level of agreement with the myth interpretation is strongly dependent on the respondents' political affiliation. Shown are participants' responses to numerical questions in the study broken down by their self-reported political leaning, with means and bootstrapped 95% confidence intervals. Controlling for label condition, we observe that Liberal participants ($n = 33$) are less likely to agree with the myth than Moderates ($n = 50, p = 0.09$) or Conservatives ($n = 17, p = 0.006$). Note that while our statistical models account for chart, participant, and label condition effects, this figure presents aggregated results.

1. Labels help by:
    - 1.1 Providing context
    - 1.2 Increasing transparency
    - 1.3 Sparking skepticism
    - 1.4 Reconciling existing beliefs
    - 1.5 Aligning with visuals
    - 1.6 Clarifying thoughts
2. Labels don't help by:
    - 2.1 Being less important than the data
    - 2.2 Highlighting data limitations
    - 2.3 Revealing misleading elements
    - 2.4 Competing with personal explanations
    - 2.5 Clashing with visuals
    - 2.6 Seeming biased
    - 2.7 Being ignored

Figure 6.9: Final codebook used to describe the data and conceptualize themes. We categorized responses based on reasons participants found the labels helpful or unhelpful, revealing the tensions that ultimately limited their effectiveness in preventing reasoning errors.

CHAPTER 7

DISCUSSION

In this section, we recap the findings of this dissertation to then offer a discussion of what the future may hold for both visualization interventions and, more generally, the role of data visualizations in the public sphere.

7.1 How to Design Resilient Visualizations

This dissertation confirms that visualization-driven misinformation is a real and pressing concern. Data visualizations are frequently used to support misleading arguments (Chapter 3), they tend to attract higher engagement on social media platforms (Chapter 4), and our user studies show that such visualizations are often persuasive and resistant to corrective interventions. From our evaluation results, we can distill several lessons about how to design visualizations that are more resilient to misinterpretation.

7.1.1 Show, Don't Hint

One key finding from Chapter 5 is that effective interventions must be visually simple and closely resemble the primary visualization. Users were able to understand and integrate guardrails that mirrored the form of the original chart, but struggled with more complex or visually distinct alternatives, such as distributional plots. Although we anticipated some confusion around unfamiliar formats, such as strip plots, it was unexpected that even a simple average line would be so frequently overlooked. These results suggest that showing even common summary statistics can still be significantly less compelling than presenting the raw data itself.

This finding is likely related to the reason why cherry-picked charts are compelling in the first place: they are easy to gist and straightforward to understand. If guardrails based on individual data points are more effective than those based on full distributions, then any corrective effort runs the risk of mimicking the very strategy it is trying to

counter—selectively presenting a subset of the data. Our Superimposed Primary Data design, which reintroduces key data points into the original chart, highlights the tension between clarity and representativeness: in order to prevent overplotting one should only visualize a handful of counter-data points. In practice, this means that visual corrections may end up engaging in a form of counter-cherry-picking.

The relative ineffectiveness of summary statistics or juxtaposed data is likely also indicative of the limits of human cognition. While the "hints" are often in essence as revealing as obviously-plotted guardrails, they require time and cognitive effort to process. As such, the results do not necessarily mean that less prominent interventions should not be used. There is considerable room for future work to devise strategies to make more subtle interventions more noticeable and study the effects. Examples of techniques to make subtle guardrails more noticeable may include animations, training, interactivity, or simply a more visually-prominent design.

7.1.2 Show, Don't Tell

Further findings from Chapter 6 underscore the limits of textual interventions. Although users generally read and considered the explanatory text provided alongside the visualizations, they tended to prioritize what they could directly see in the chart. When the visual message and the textual message conflicted, users most often trusted the visualization and not the text. This suggests that while text can provide context or tell a story, it cannot override what the data appears to show. For visualization designers, this emphasizes the need to work with—not against—the visual elements themselves when trying to design charts for misinterpretation resiliency.

Aside from perceived objectivity, the relative cognitive load of text can also be a factor leading to its ineffectiveness. Just as a subtle summary statistic is less effective than a visual plot, reading a paragraph requires more time and effort than glancing at a chart. This does not mean that visual elements are always immediately understood, but they do allow viewers to form a quick first impression—accurate or not. Text, by contrast, encodes information more abstractly and must be read carefully to be meaningful. As with our earlier discussion on making distributional or statistical interventions more noticeable, future work could explore whether increasing text's prominence, adjusting its placement,

or simplifying its phrasing might reduce cognitive load and improve the effectiveness of textual interventions.

7.1.3 Rethink What to Show

The relative power of visual elements is also a good reminder that it really matters what visualization you produce in the first place. Taken together, the findings of Chapter 5 and Chapter 6 illustrate the difficulty of intervening after a misleading interpretation has already taken hold. Both studies rely on strategies that attempt to counter or reframe an interpretation based on data that have already been shown; however, our results suggest that the efficacy of a reactive approach is limited. Once viewers have formed a conclusion from a chart, whether accurate or not, it becomes remarkably difficult to unsee or revise that interpretation.

This points to a broader lesson for visualization design: resilience may be more effectively achieved upstream, by anticipating misinterpretations during the initial design process rather than appending interventions after the fact, similar to the idea of learning anti-objectives [84]. Although this kind of foresight may be challenging in interactive data explorers when users control what is shown, static charts offer more opportunities for deliberate design choices. Designers should consider not only whether a particular visual element may be misinterpreted but also the overall framing of the chart: What assumptions are invited by visualizing these variables? Would adding another dimension or another data point clarify or confuse? Is a subset of data a compelling example, or could it lead viewers to overgeneralize? And, more fundamentally, is this visualization more likely to add insight or unintentionally cause harm in the first place? Considering these questions leads not only to preventing misinterpretation, but also to engaging more deeply with the ethical dimensions of creating a visualization and anticipating its role once released to the general public.

Although these insights instill hope for the future of visualization design and visual communication more broadly, anticipatory design goes beyond a simple corrective misinformation intervention. Unfortunately, we cannot redesign an already posted and already seen chart. Consequently, we advocate for anticipatory approaches not just as an answer to misinformation but as part of a broader shift in visualization practice that encourages

considering the ethics and the plurality of interpretations in visualization design.

7.2 The Future Role of Visualization

We personally find the findings of this dissertation to be both unsurprising and surprising. On the one hand, it is not surprising that data can be misinterpreted, that charts are compelling, and that nuance is hard to convey or absorb. We know that people can lie with data, and we know how difficult it is to engage with it deeply and analytically. At the same time, the broader findings are, if not surprising, then deeply unsatisfying: it is often difficult to distinguish truth from falsehood by just looking at a visualization.

How do we cope with the fact that we often cannot tell them apart? In the constrained and fast-moving context of social media, the difference practically disappears. Most misleading visual tactics have legitimate, rigorous counterparts: what looks like cherry-picking might be an illustrative example; what seems like a spurious correlation could actually reflect causation. And with the sheer volume of content shared online, there is little time or space to formally assess the accuracy of any single visualization. Each topic demands its own specific domain knowledge to interpret responsibly. Yet, our findings show that even when more context is added—whether through text or additional data—it often has little effect on how people interpret a chart.

Taken together, these findings suggest a sobering conclusion: the trustworthiness of a chart in isolation may be practically undecidable for most viewers. Instead, users have to rely on heuristics—Does the chart look polished? Do I trust the source? Does the conclusion align with what I already believe? These shortcuts are not always analytical but they are far from unreasonable. But if the public interprets data primarily through external heuristics, do the data themselves still matter? Outside of scientific and professional settings, do data retain meaning? One possible answer is that we are on a trajectory toward a *post-data* world.

Over the past decade, we have seen widespread discourse around *post-truth*—a term that describes the growing anxiety about what claims can be accepted as fact. It is not that truth once existed and now does not; rather, post-truth reflects a societal shift in which there is no longer a shared understanding of what truth is and how it can be known. Similarly, we may be losing a common understanding of what things we can or cannot

learn from data and what skills and knowledge one needs to perform meaningful analyses.

A similar shift is underway with other forms of evidence affected by technological advancements. For example, generative AI made artificial images nearly indistinguishable from real photographs, undermining our ability to use photos as reliable documentation. This represents a major loss in the way we record and verify events; however, it is not entirely unfamiliar. What is on the surface a technological advancement is also a regression into the older and perhaps more unreliable methods of knowing the world. Before photography, societies relied on paintings to depict events with varying degrees of accuracy. For example, Jacques-Louis David's famous early-1800s portrait *Napoleon Crossing the Alps* shows the emperor heroically ascending on horseback, while Paul Delaroche's version depicts him riding a mule, led by a guide. The latter is widely considered more historically accurate; however, without additional evidence or without preexisting knowledge of the historical context, either version would look equally compelling. It is obvious in this case that we simply cannot rely on the painting (or at least on the painting itself) to learn anything factual about the events, and we have to do extra work and turn to other forms of inquiry.

In a post-data future, it is similarly unclear what authority data visualizations alone will hold in public discourse. A dip in the stock market might be interpreted as a collapse or merely a correction—depending on who shared it and for what purpose. It is possible that we will eventually agree on common indicators of trustworthiness that are associated with yet still separate from the chart: such as our interventions—guardrails or descriptive notes—that people would expect to see on a rigorous chart but not on a misleading one. However, it is also possible that if we lose shared standards of evaluation, we may see people disengage altogether from considering data a source of truth. In either case, we find ourselves in a gap between primarily relying on data to understand the world and recognizing that it may no longer be enough.

7.3 Data (Mis)interpretation as a Critical Practice

These anxieties echo discussions that arose in the world of literary critique at the turn of the century. In 2004, Bruno Latour published his now famous essay "Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern." [250]. In the essay,

Latour lamented the fact that it seems as if the entire toolbox of literary critique—reading between the lines, finding hidden meaning in language and symbols, questioning and dissecting statements—has been coopted by conspiracy theorists. Latour expresses his shock to realize that the narratives used to dismiss the September 11 attacks as a false flag operation were eerily similar to the methods he and his colleagues use to expose power structures and inequalities.

It is not difficult to see parallels to the issues we observe with misleading data analysis. Perhaps for too long, we have depended on our data-driven methods to always lead us toward some approximation of truth. Collecting, transforming, analyzing, and presenting data is relied on to understand the world, expose unseen phenomena, and prepare for the future. The findings from this dissertation show that these methods in and of themselves do not lead us to the truth; they require at the very least deep domain knowledge, statistical accuracy, and good-faith intentions, whose presence or absence is not at all readily seen in a simple data presentation.

Another idea echoing the same concerns is Eve Kosofsky Sedgwick's notion of paranoid reading [251]. According to Sedgwick, paranoid reading is the normative way of conducting critical practice and it is rooted in constant suspicion—it always assumes that material is hiding something harmful or deceptive. This mode of reading is ultimately protective: the reader always outsmarts the material, anticipates bad outcomes, and refuses to be surprised. Sedgwick describes paranoid reading to be a "strong theory," meaning that through its anticipation it is endlessly adaptable and can offer an explanation for everything.

It may be tempting to view only the original misinterpretations of data as instances of paranoid reading; that misinformation actors refuse to be surprised by data and thus employ reasoning errors to interpret it according to their worldview. But in fact, our reactions to the reasoning errors themselves can be thought of as suspicious critical practices. Common responses to misinformation—fact-checking, correcting, exposing flaws—are similarly defensive. We anticipate reasoning errors in interpretations of data that we disagree with and then object through fact-checking. The dynamic of paranoid-on-paranoid reading was especially evident in out study in Chapter 4, where discussions around misleading visualizations spun into endless chains of caveats and counter-data,

rarely reaching a clear conclusion. In this loop of constant debunking, it becomes difficult to tell what is true. If every analysis comes with limitations, if every visualization is met with skepticism, what stands out as credible? Over time, this cycle of critique may erode trust in data altogether.

When data interpretation becomes a form of suspicious critique, interventions often fall flat—not because they fail to address the content, but because they enter the same game. Just as the original chart is picked apart for its flaws, so too are the responses, each offering their own caveats, corrections, and counterpoints. The result is not cumulative insight, but a recursive loop where critique gives rise to more critique. Suspicious critique can be seen as a social dynamic that ends up amplified on social media: Sedgwick describes paranoid reading as marked by "outsmarting" and "one-upmanship" [251], a sentiment echoed by Castiglia's critique of "smug knowingness" as the default critical stance [252]. Thus online data critiques often resemble less a collective effort to make sense of the data and more a competition over who can spot the most damning caveat in the chart, its caption, or the fact-check. With every interpretation appearing vulnerable to dismantling, the data itself lose credibility. In such a landscape, reading visualizations becomes less about understanding and more about having the last word.

7.4 Moving Forward to Post-Critique in Visualization

Resolving this tension is not easy. Perhaps the first step involves admitting at least a partial defeat: the data themselves are no longer bearers of facts. In fact, Latour advocates for a move away from facts altogether. As the title of his essay implies, Latour suggests shifting the focus from matters of fact to matters of concern: from debating over what has transpired to what has meaning [250]. In the context of critique, this involves moving away from negative, dissecting, and destructive readings towards ones that build instead: findings values and commonalities that we, as a society, share and focus on rebuilding our shared reality. Similarly, Sedgwick challenges us to consider whether our critical energy is best spent trying to force the text—or in this case, the data—to reveal some hidden truth [251]. Instead, Sedgwick advocates for what she calls a reparative reading: attempts to understand the emotions and concerns behind the (data) interpretation. This set of ideas, further developed by Rita Felski in her book "The Limits of Critique," is now widely

known as *post-critique* [253].

We can interpret the findings of our dissertation through the lens of post-critique. For example, in Chapter 6, we developed Guide labels to elaborate on the intended use of a visualization and observed a negative effect: people believed the myth more. Although we originally theorized that this is indicative of the implied truth effect (after seeing direct fact-checking, a chart without one is more trustworthy), it is also possible that the intended message of the chart simply did not resonate with viewers' matters of concern, and the guide label made that abundantly clear. People would look at a chart of COVID-19 cases to assess their own risk (their matter of concern), but the label redirected them toward an interpretation that felt disconnected from their priorities. Taken together, the relative ineffectiveness of visualization interventions we designed in Chapters 5 and 6 may be indicative of the ineffectiveness of applying an analytical approach to combating a fundamental mismatch between the information we present and the information people need or care about.

Applying post-critique to visualization invites us to expand our approach, shifting from solely debunking toward also seeking to understand. Interpretations that diverge from rigorous analysis may indeed reflect reasoning errors like the base rate fallacy or cherry-picking but they can also reveal deeper emotional undercurrents: fears, anxieties, or hopes that a purely analytical lens may overlook. For example, as we briefly discussed in Chapter 6, we may be able to learn from misinterpretations of data what people actually care about. Rather than dissecting and countering them, we can try to understand what our common values are and consider these ideas in future visualization design. Most of us can probably agree on shared goals surrounding the climate: fewer catastrophes, better predictions for crop yields, and livable outdoor temperatures. But when public discourse focuses on whether the mean surface temperature increased by one degree Celsius, whether the change is due to CO2 or El Nino, whether it is visible on a cherry-picked scale, or whether the data are current and precise, we risk losing sight of what actually matters. Perhaps what we learn from our findings is that a metric designed for scientific monitoring or policy decisions is not always a metric that resonates with the public or speaks to their values.

What post-critique looks like in practice is still an open question, but that ambiguity,

as Felski argues, is part of its power: it "serves as a placeholder for emerging ideas" [253]. Below, we speculate on what data visualization of matters of concern might look like.

One idea lies in shifting what we visualize in the first place and choosing the right metrics or their transformations. Instead of showing projected temperature increases, for example, it might be more meaningful to show the impacts on insurance rates or the probability of specific climate events—especially when these data are localized. This kind of framing may better reflect what actually matters to the people viewing the chart, rather than directly reusing data from scientific reports or government trackers that may not map to individuals' lived experiences.

Another way to think about visualizing matters of concern is to let go of the idea that visualizations are neutral. Even when the data is presented as objective, the visualization itself is created by someone, for a purpose. It can be helpful, then, to think of a chart as an utterance, a way of saying something. This idea is similar to literary scholar Toril Moi's framing of post-critique [254]: to consider any text as an action rather than an object. Asking what you want to say with a chart, or what kind of conversation you are inviting by showing it, can be a useful starting point for design. In this line of thought, we can see that visualizing the output of a climate modeling simulation says to the general public that "scientists predict that the ocean will be warmer" and not much more. It may then become painfully obvious that this utterance is not designed to convince the viewer about the risks of climate change and that we should not expect such a visualization to change minds.

This perspective can also shift how we study public interpretation of visualizations. Much of our work, especially in Chapter 3 and Chapter 4, focused on interpretation through a matters-of-fact lens: we looked at what kinds of analytical errors people made, or what statistical nuance they missed. This shaped both our conclusions and our ideas for future interventions. Looking through a matters-of-concern lens, though, might lead to a different approach. Instead of starting with what is right or wrong, we might start by asking what concerns people bring to the visualization. In our data, even "incorrect" interpretations often reflected shared and valid concerns—about health, safety, or economic security. From this reparative reading perspective, those divergences might not be analytical but instead could reflect issues of trust, weak public communication, or even flaws in the metrics and data used by institutions themselves.

In summary, rather than abandoning analytical thinking, we invite the readers to broaden our interpretive stance, to attend not only to what data gets wrong, but to how and why it resonates with people in the first place. In the context of data-driven misinformation, this might mean moving from fact-checking to fact-building, where the goal is not only to refute falsehoods but to construct meaningful and compelling narratives. It might involve shifting from visualizing data that are easily accessible and repurposeable to identifying new data and metrics that represent matters of concern. Above all, in order to maintain the importance of data in the public sphere, we encourage visualization to move from suspicion to sustained collective engagement, to interpret data not as adversaries, but as coinvestigators.

CHAPTER 8

CONCLUSION

The findings from this dissertation suggest that misleading data visualizations are both widespread and difficult to challenge. At the same time, they often spark intense partisan debates, with their persuasive power largely shaped by viewers' prior beliefs. Addressing this issue requires not only design interventions but also long-term efforts in education and continued observation of how people interact with data.

While much of this dissertation focused on visualizations that reinforce misinformation, the broader goal is to consider how publicly shared data visualizations can be made more useful, well-contextualized, and resistant to misuse. Ultimately, this research highlights the rhetorical power of data and charts—not as neutral conveyors of truth, but as tools that shape public discourse. Though no simple fix exists for data-driven misinformation, interventions may help shift how data is understood and used in society over time.

REFERENCES

[1] M. Bica, J. L. Demuth, J. E. Dykes, and L. Palen, "Communicating hurricane risks: Multi-method examination of risk imagery diffusion," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–13.

[2] V. Mather, "Elite runner who rode in a car during a race is banned for a year," *The New York Times*, Nov. 2023.

[3] C. Dyer, "Lancet retracts Wakefield's MMR paper," *BMJ*, vol. 340, no. 4, p. c696, Feb. 2010.

[4] D. Lange, S. Sahai, J. M. Phillips, and A. Lex, "Ferret: Reviewing tabular datasets for manipulation," *Comput. Graph. Forum*, vol. 42, no. 3, pp. 187–198, Jun. 2023.

[5] K. Starbird, A. Arif, and T. Wilson, "Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. CSCW, pp. 127:1–127:26, Nov. 2019.

[6] M. Lisnic, C. Polychronis, A. Lex, and M. Kogan, "Misleading beyond visual tricks: How people actually lie with charts," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2023, pp. 1–21.

[7] M. Lisnic, A. Lex, and M. Kogan, ""Yeah, this graph doesn't show that": Analysis of online engagement with misleading data visualizations," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2024, pp. 1–14.

[8] M. Lisnic, Z. Cutler, M. Kogan, and A. Lex, "Visualization guardrails: Designing interventions against cherry-picking in interactive data explorers," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2025, pp. 1–19.

[9] J. Surowiecki, *The Wisdom of Crowds*. New York, NY, USA: Knopf Doubleday Publ. Group, 2005.

[10] A. Arif, J. J. Robinson, S. A. Stanek, E. S. Fichet, P. Townsend, Z. Worku, and K. Starbird, "A closer look at the self-correcting crowd: Examining corrections in online rumors," in *Proc. ACM Conf. Comput. Supported Coop. Work Soc. Comput.*, Feb. 2017, pp. 155–168.

[11] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: Can we trust what we RT?" in *Proc. Workshop on Social Media Analytics (SOMA)*, Jul. 2010, pp. 71–79.

[12] K. Starbird, J. Maddock, M. Orand, P. Achterman, and R. Mason, "Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston Marathon bombing," *iConference*, pp. 1–9, Mar. 2014.

[13] T. Heverin and L. Zach, "Use of microblogging for collective sense-making during violent crises: A study of three campus shootings," *J. Am. Soc. Inf. Sci. Technol.*, vol. 63, no. 1, pp. 34–47, Oct. 2012.

[14] L. Palen, S. Vieweg, S. B. Liu, and A. L. Hughes, "Crisis in a networked world: Features of computer-mediated communication in the April 16, 2007, Virginia Tech event," *Soc. Sci. Comput. Rev.*, vol. 27, no. 4, pp. 467–480, Nov. 2009.

[15] K. Starbird, L. Palen, A. L. Hughes, and S. Vieweg, "Chatter on the red: What hazards threat reveals about the social life of microblogged information," in *Proc. ACM Conf. Comput. Supported Coop. Work Social Comput.*, Feb. 2010, pp. 241–250.

[16] J. Sutton, L. Palen, and I. Shklovski, "Backchannels on the front lines: Emergent uses of social media in the 2007 Southern California wildfires," *Proc. 5th Int. ISCRAM Conf.*, pp. 1–9, May 2008.

[17] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: What Twitter may contribute to situational awareness," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2010, pp. 1079–1088.

[18] X. Gui, Y. Kou, K. H. Pine, and Y. Chen, "Managing uncertainty: Using social media for risk assessment during a public health crisis," in *Proc. ACM CHI Conf. Hum. Factors Comput. Syst.*, May 2017, pp. 4520–4533.

[19] K. H. Pine, M. Lee, S. A. Whitman, Y. Chen, and K. Henne, "Making sense of risk information amidst uncertainty: Individuals' perceived risks associated with the COVID-19 pandemic," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–15.

[20] Z. Aghajari, E. P. S. Baumer, and D. DiFranzo, "Reviewing interventions to address misinformation: The need to expand our vision beyond an individualistic focus," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CSCW1, pp. 87:1–87:34, Apr. 2023.

[21] C. Martel and D. G. Rand, "Misinformation warning labels are widely effective: A review of warning effects and their moderating features," *Curr. Opin. Psychol.*, vol. 54, p. 101710, Dec. 2023.

[22] D. Carnahan and D. E. Bergan, "Correcting the misinformed: The effectiveness of fact-checking messages in changing false beliefs," *Polit. Commun.*, vol. 39, no. 2, pp. 166–183, Mar. 2022.

[23] P. Mena, "Cleaning up social media: The effect of warning labels on likelihood of sharing false news on Facebook," *Policy Internet*, vol. 12, no. 2, pp. 165–183, Jun. 2020.

[24] D. MacFarlane, L. Q. Tay, M. J. Hurlstone, and U. K. H. Ecker, "Refuting spurious COVID-19 treatment claims reduces demand and misinformation sharing," *J. Appl. Res. Mem. Cogn.*, vol. 10, no. 2, pp. 248–258, Jun. 2021.

[25] N. Walter, J. Cohen, R. L. Holbert, and Y. Morag, "Fact-checking: A meta-analysis of what works and for whom," *Political Commun.*, vol. 37, no. 3, pp. 350–375, May 2020.

[26] I. Skurnik, C. Yoon, D. C. Park, and N. Schwarz, "How warnings about false claims become recommendations," *J. Consum. Res.*, vol. 31, no. 4, pp. 713–724, Mar. 2005.

[27] M. Mosleh, C. Martel, D. Eckles, and D. Rand, "Perverse downstream consequences of debunking: Being corrected by another user for posting false political news increases subsequent sharing of low quality, partisan, and toxic content in a Twitter field experiment," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–13.

[28] G. Pennycook, A. Bear, E. T. Collins, and D. G. Rand, "The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings," *Manag. Sci.*, vol. 66, no. 11, pp. 4944–4957, Nov. 2020.

[29] J. Turcotte, C. York, J. Irving, R. M. Scholl, and R. J. Pingree, "News recommendations from social media opinion leaders: Effects on media trust and information seeking," *J. Comput.-Mediat. Commun.*, vol. 20, no. 5, pp. 520–535, Sep. 2015.

[30] M. Karlsson, C. Clerwall, and L. Nord, "Do not stand corrected: Transparency and users' attitudes to inaccurate news and corrections in online journalism," *Journalism Mass Commun. Q.*, vol. 94, no. 1, pp. 148–167, Mar. 2017.

[31] M. W. a. J. Gottfried, "Republicans far more likely than democrats to say fact-checkers tend to favor one side," *Pew Research Center*, Jun. 2019.

[32] P. S. Hart and E. C. Nisbet, "Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies," *Commun. Res.*, vol. 39, no. 6, pp. 701–723, Dec. 2012.

[33] J. Shin and K. Thorson, "Partisan selective sharing: The biased diffusion of fact-checking messages on social media," *J. Commun.*, vol. 67, no. 2, pp. 233–255, Apr. 2017.

[34] J. Allen, A. A. Arechar, G. Pennycook, and D. G. Rand, "Scaling up fact-checking using the wisdom of crowds," *Sci. Adv.*, vol. 7, no. 36, p. eabf4393, Sep. 2021.

[35] J. Allen, C. Martel, and D. G. Rand, "Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2022, pp. 1–19.

[36] N. Walter and N. A. Salovich, "Unchecked vs. uncheckable: How opinion-based claims can impede corrections of misinformation," *Mass Commun. Soc.*, vol. 24, no. 4, pp. 500–526, Jul. 2021.

[37] A. Oeldorf-Hirsch, M. Schmierbach, A. Appelman, and M. P. Boyle, "The ineffectiveness of fact-checking labels on news memes and articles," *Mass Commun. Soc.*, vol. 23, no. 5, pp. 682–704, Sep. 2020.

[38] B. W. Wojdynski and N. J. Evans, "Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising," *J. Advert.*, vol. 45, no. 2, pp. 157–168, Apr. 2016.

[39] F. Jahanbakhsh, A. X. Zhang, A. J. Berinsky, G. Pennycook, D. G. Rand, and D. R. Karger, "Exploring lightweight interventions at posting time to reduce the sharing of misinformation on social media," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW1, pp. 18:1–18:42, Apr. 2021.

[40] F. Jahanbakhsh, A. X. Zhang, and D. R. Karger, "Leveraging structured trusted-peer assessments to combat misinformation," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, pp. 524:1–524:40, Nov. 2022.

[41] D. Huff, *How to Lie with Statistics*, 2nd ed. New York, NY, USA: W. W. Norton & Company, 1993.

[42] E. R. Tufte, *The Visual Display of Quantitative Information*, 1st ed. Cheshire, CT, USA: Graphics Press, 1983.

[43] A. Cairo, "Graphics lies, misleading visuals," in *New Challenges for Data Design*. London, UK: Springer, 2015, pp. 103–116.

[44] M. Correll and J. Heer, "Black hat visualization," in *Workshop on Dealing with Cognitive Biases in Visualisations (DECISIVe), IEEE VIS*, Oct. 2017.

[45] C. Lauer and S. O'Brien, "The deceptive potential of common design tactics used in data visualizations," in *Proc. 38th ACM Int. Conf. Design Commun.*, Oct. 2020, pp. 1–9.

[46] ——, "How people are influenced by deceptive tactics in everyday charts and graphs," *IEEE Trans. Prof. Commun.*, vol. 63, no. 4, pp. 327–340, Dec. 2020.

[47] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini, "How deceptive are deceptive visualizations? an empirical analysis of common distortion techniques," in *Proc. ACM Conf. Hum. Factors Comput. Syst.*, Apr. 2015, pp. 1469–1478.

[48] D. A. Szafir, "The good, the bad, and the biased: Five ways visualizations can mislead (and how to fix them)," *Interactions*, vol. 25, no. 4, pp. 26–33, Jun. 2018.

[49] J. E. P. Driessen, D. A. C. Vos, I. Smeets, and C. J. Albers, "Misleading graphs in context: Less misleading than expected," *PLOS ONE*, vol. 17, no. 6, p. e0265823, Jun. 2022.

[50] L. Y.-H. Lo, A. Gupta, K. Shigyo, A. Wu, E. Bertini, and H. Qu, "Misinformed by visualization: What do we learn from misinformative visualizations?" *Comput. Graph. Forum*, vol. 41, no. 3, pp. 515–525, Dec. 2022.

[51] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic, "A task-based taxonomy of cognitive biases for information visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 2, pp. 1413–1432, Feb. 2020.

[52] A. Calero Valdez, M. Ziefle, and M. Sedlmair, "Studying biases in visualization research: Framework and methods," in *Cognitive Biases in Visualizations*. Cham, Switzerland: Springer International Publishing, 2018, pp. 13–27.

[53] E. Bertini, M. Correll, and S. Franconeri, "Why shouldn't all charts be scatter plots? beyond precision-driven visualizations," in *Proc. IEEE Vis. Conf.*, Oct. 2021, pp. 206–210.

[54] M. Correll, "Are we making progress in visualization research?" in *Proc. IEEE BELIV Workshop*, Oct. 2022, pp. 1–10.

[55] L. Klein, "What data visualization reveals: Elizabeth Palmer Peabody and the work of knowledge production," *Harvard Data Sci. Rev.*, vol. 4, no. 2, pp. 1–34, Apr. 2022.

[56] C. Lee, T. Yang, G. D. Inchoco, G. M. Jones, and A. Satyanarayan, "Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online," in *Proc. 2021 CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2021, no. 607, pp. 1–18.

[57] J. Hullman, E. Adar, and P. Shah, "The impact of social information on visual judgments," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2011, pp. 1461–1470.

[58] H.-K. Kong, Z. Liu, and K. Karahalios, "Frames and slants in titles of visualizations on controversial topics," in *Proc. 2018 CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–12.

[59] C. Ziemkiewicz, A. Ottley, R. J. Crouser, K. Chauncey, S. L. Su, and R. Chang, "Understanding visualization by understanding individual users," *IEEE Comput. Graph. Appl.*, vol. 32, no. 6, pp. 88–94, Nov. 2012.

[60] E. M. Peck, S. E. Ayuso, and O. El-Etr, "Data is personal: Attitudes and perceptions of data visualization in rural Pennsylvania," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–12.

[61] C. Xiong, L. Van Weelden, and S. Franconeri, "The curse of knowledge in visual data communication," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 10, pp. 3051–3062, Oct. 2020.

[62] T. Weikmann and S. Lecheler, "Visual disinformation in a digital age: A literature synthesis and research agenda," *New Media Soc.*, pp. 3696–3713, Dec. 2022.

[63] M. N. Hannah, "A conspiracy of data: QAnon, social media, and information visualization," *Social Media Soc.*, vol. 7, no. 3, Jul. 2021.

[64] X. Pu and M. Kay, "The garden of forking paths in visualization: A design space for reliable exploratory visual analytics," in *Proc. IEEE BELIV Workshop*, Oct. 2018, pp. 37–45.

[65] E. Zgraggen, Z. Zhao, R. Zeleznik, and T. Kraska, "Investigating the effect of the multiple comparisons problem in visual analysis," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–12.

[66] A. Kale, Z. Guo, X. L. Qiao, J. Heer, and J. Hullman, "EVM: Incorporating model checking into exploratory visual analysis," *IEEE Trans. Vis. Comput. Graph.*, pp. 1–11, Jan. 2024.

[67] 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 *Proc. IEEE Conf. Vis. Anal. Sci. Technol.*, Oct. 2017, pp. 104–115.

[68] A. Asudeh, H. V. Jagadish, Y. W. Wu, and C. Yu, "On detecting cherry-picked trendlines," *Proc. VLDB Endow.*, vol. 13, no. 6, pp. 939–952, Feb. 2020.

[69] Y. Lin, B. Youngmann, Y. Moskovitch, H. V. Jagadish, and T. Milo, "On detecting cherry-picked generalizations," *Proc. VLDB Endow.*, vol. 15, no. 1, pp. 59–71, Sep. 2021.

[70] ——, "OREO: Detection of cherry-picked generalizations," *Proc. VLDB Endow.*, vol. 15, no. 12, pp. 3570–3573, Aug. 2022.

[71] E. Wall, A. Narechania, A. Coscia, J. Paden, and A. Endert, "Left, right, and gender: Exploring interaction traces to mitigate human biases," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 1, pp. 966–975, Jan. 2022.

[72] P.-M. Law, L. Y.-H. Lo, A. Endert, J. Stasko, and H. Qu, "Causal perception in question-answering systems," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–15.

[73] Q. Zhu, L. Y.-H. Lo, M. Xia, Z. Chen, and X. Ma, "Bias-aware design for informed decisions: Raising awareness of self-selection bias in user ratings and reviews," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, pp. 496:1–496:31, Nov. 2022.

[74] C. Zheng and X. Ma, "Evaluating the effect of enhanced text-visualization integration on combating misinformation in data story," in *Proc. IEEE PacificVis*, Apr. 2022, pp. 141–150.

[75] M. Correll, E. Bertini, and S. Franconeri, "Truncating the y-axis: Threat or menace?" in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–12.

[76] Y. Zhang, Y. Sun, J. D. Gaggiano, N. Kumar, C. Andris, and A. G. Parker, "Visualization design practices in a crisis: Behind the scenes with COVID-19 dashboard creators," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 1, pp. 1037–1047, Jan. 2023.

[77] 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 Trans. Vis. Comput. Graph.*, pp. 1–16, Jan. 2022.

[78] I. Drosos, A. Sarkar, Xiaotong, Xu, and N. Toronto, ""It Makes You Think": Provocations help restore critical thinking to ai-assisted knowledge work," *arXiv*, Feb. 2025.

[79] L. W. Ge, M. Easterday, M. Kay, E. Dimara, P. Cheng, and S. L. Franconeri, "V-FRAMER: Visualization framework for mitigating reasoning errors in public policy," *Proc. CHI Conf. Hum. Factors Comput. Syst.*, pp. 1–15, Apr. 2024.

[80] C. X. Bearfield, L. van Weelden, A. Waytz, and S. Franconeri, "Same data, diverging perspectives: The power of visualizations to elicit competing interpretations," *arXiv*, Sep. 2024.

[81] R. Davis, X. Pu, Y. Ding, B. D. Hall, K. Bonilla, M. Feng, M. Kay, and L. Harrison, "The risks of ranking: Revisiting graphical perception to model individual differences in visualization performance," *IEEE Trans. Vis. Comput. Graph.*, vol. 30, no. 3, pp. 1756–1771, Mar. 2024.

[82] E. Dimara and J. Stasko, "A critical reflection on visualization research: Where do decision making tasks hide?" *IEEE Trans. Vis. Comput. Graph.*, vol. 28, pp. 1128–1138, Feb. 2021.

[83] J. Hullman and N. Diakopoulos, "Visualization rhetoric: Framing effects in narrative visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2231–2240, Dec. 2011.

[84] E. Adar and E. Lee, "Communicative visualizations as a learning problem," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 946–956, Feb. 2021.

[85] A. Fan, Y. Ma, M. Mancenido, and R. Maciejewski, "Annotating line charts for addressing deception," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2022, pp. 1–12.

[86] T. Engineering, "Blobstore: Twitter's in-house photo storage system," 2012.

[87] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[88] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[89] A. Strauss and J. M. Corbin, *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Thousand Oaks, CA, USA: SAGE Publ. Ltd, 1990.

[90] R. Likert, "A technique for the measurement of attitudes," *Arch. Psychol.*, vol. 22, no. 140, pp. 55–55, Aug. 1932.

[91] K. Krippendorff, *Content analysis: An introduction to its methodology*. Thousand Oaks, CA, USA: SAGE Publ. Ltd, 2004.

[92] H. Ritchie, E. Mathieu, L. Rodés-Guirao, C. Appel, C. Giattino, E. Ortiz-Ospina, J. Hasell, B. Macdonald, D. Beltekian, and M. Roser, "Coronavirus pandemic (COVID-19)," *Our World in Data*, 2020.

[93] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister, "UpSet: Visualization of intersecting sets," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1983–1992, Dec. 2014.

[94] J. M. Last, *A Dictionary of Epidemiology*. New York, NY, USA: Oxford Univ. Press, USA, 2000.

[95] H. Lin, D. Akbaba, M. Meyer, and A. Lex, "Data hunches: Incorporating personal knowledge into visualizations," *IEEE Trans. Vis. Comput. Graph.*, vol. 29, no. 1, pp. 504–514, Jan. 2022.

[96] C. M. Brown, J. Vostok, and H. Johnson, "Outbreak of SARS-CoV-2 infections, including COVID-19 vaccine breakthrough infections, associated with large public gatherings - Barnstable county, Massachusetts, july 2021," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 70, Jul. 2021.

[97] A. J. Sharon and A. Baram-Tsabari, "Can science literacy help individuals identify misinformation in everyday life?" *Sci. Educ.*, vol. 104, no. 5, pp. 873–894, Sep. 2020.

[98] F. Hendriks, D. Kienhues, and R. Bromme, "Trust in science and the science of trust," in *Trust Commun. Digit. World: Models Concepts Trust Res.*, Feb. 2016, pp. 143–159.

[99] S. P. Norris, "Intellectual independence for nonscientists and other content-transcendent goals of science education," *Sci. Educ.*, vol. 81, no. 2, pp. 239–258, Apr. 1997.

[100] H.-K. Kong, Z. Liu, and K. Karahalios, "Trust and recall of information across varying degrees of title-visualization misalignment," in *Proc. 2019 CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–13.

[101] M. Oghbaie, M. J. Pennock, and W. B. Rouse, "Understanding the efficacy of interactive visualization for decision making for complex systems," in *Annu. IEEE Syst. Conf.*, Apr. 2016, pp. 1–6.

[102] T. Vigen, *Spurious correlations*. New York, NY, USA: Hachette Books, 2017.

[103] P. J. Hurley, *A Concise Introduction to Logic*. Boston, MA, USA: Cengage Learning, 2014.

[104] D. Hume, *An Inquiry Concerning Human Understanding. A Dissertation on the Passions. An Inquiry Concerning the Principles of Morals. The Natural History of Religion*. London, UK: T. Cadell, 1793.

[105] K. R. Popper, *The Logic of Scientific Discovery*. London, UK: Psychology Press, 2002.

[106] ——, *Conjectures and Refutations: The Growth of Scientific Knowledge*. London, UK: Psychology Press, 2002.

[107] J. Cook, P. Ellerton, and D. Kinkead, "Deconstructing climate misinformation to identify reasoning errors," *Environmental Research Letters*, vol. 13, no. 2, p. 024018, Feb. 2018.

[108] J. Hullman, "Why authors don't visualize uncertainty," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 1, pp. 130–139, Jan. 2020.

[109] A. M. Hause, "Safety monitoring of an additional dose of COVID-19 vaccine — United States, August 12–September 19, 2021," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 70, pp. 1379–1384, Sep. 2021.

[110] I. D. Plumb, "Effectiveness of COVID-19 mRNA vaccination in preventing COVID-19–associated hospitalization among adults with previous SARS-CoV-2 infection — United States, June 2021–February 2022," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 71, pp. 549–555, Apr. 2022.

[111] S. M. Olson, "Effectiveness of Pfizer-BioNTech mRNA vaccination against COVID-19 hospitalization among persons aged 12–18 years — United States, June–September 2021," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 70, pp. 1483–1488, Oct. 2021.

[112] M. W. Tenforde, "Effectiveness of mRNA vaccination in preventing COVID-19–associated invasive mechanical ventilation and death — United States, March 2021–January 2022," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 71, pp. 459–465, Mar. 2022.

[113] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," *Rev. Gen. Psychol.*, vol. 2, no. 2, pp. 175–220, Jun. 1998.

[114] R. B. Cialdini and N. J. Goldstein, "Social influence: Compliance and conformity," *Annu. Rev. Psychol.*, vol. 55, pp. 591–621, Feb. 2004.

[115] E. Pariser, *The filter bubble: What the internet is hiding from you*. London, UK: Penguin Books Limited, 2011.

[116] A. Hannak, D. Margolin, B. Keegan, and I. Weber, "Get back! You don't know me like that: The social mediation of fact checking interventions in Twitter conversations," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 8, no. 1, pp. 187–196, May 2014.

[117] A. E. Marwick and W. C. Partin, "Constructing alternative facts: Populist expertise and the QAnon conspiracy," *New Media Soc.*, pp. 1–21, Apr. 2022.

[118] L. Wu, F. Morstatter, X. Hu, and H. Liu, "Mining misinformation in social media," in *Big data in complex and social networks*. London, UK: Chapman and Hall/CRC, 2016.

[119] K. Schmidt and L. Bannon, "Taking CSCW seriously: Supporting articulation work," *Comput. Supported Coop. Work*, vol. 1, pp. 7–40, Mar. 1992.

[120] S. Phadke and T. Mitra, "Many faced hate: A cross platform study of content framing and information sharing by online hate groups," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–13.

[121] X. Yao, F. Ye, M. Zhang, C. Cui, B. Huang, P. Niu, X. Liu, L. Zhao, E. Dong, C. Song, S. Zhan, R. Lu, H. Li, W. Tan, and D. Liu, "In vitro antiviral activity and projection of optimized dosing design of hydroxychloroquine for the treatment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)," *Clin. Infect. Dis.*, vol. 71, no. 15, pp. 732–739, Aug. 2020.

[122] Z. Lee, C. R. Rayner, J. I. Forrest, J. B. Nachega, E. Senchaudhuri, and E. J. Mills, "The rise and fall of hydroxychloroquine for the treatment and prevention of COVID-19," *Am. J. Trop. Med. Hyg.*, vol. 104, no. 1, pp. 35–38, Jan. 2021.

[123] M. R. Mehra, F. Ruschitzka, and A. N. Patel, "Retraction—Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: A multinational registry analysis," *Lancet*, vol. 395, no. 10240, p. 1820, Jun. 2020.

[124] E. Zak, "The colors of a #climatescam: An exploration of anti-climate change graphs on Twitter," *J. Interdiscip. Sci.*, vol. 7, no. 1, pp. 13–28, May 2023.

[125] H. Matatov, M. Naaman, and O. Amir, "Stop the [image] steal: The role and dynamics of visual content in the 2020 U.S. election misinformation campaign," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW2, pp. 541:1–541:24, Nov. 2022.

[126] A. Cairo, *How Charts Lie: Getting Smarter about Visual Information*. New York, NY, USA: W. W. Norton & Company, 2019.

[127] A. McNutt, G. Kindlmann, and M. Correll, "Surfacing visualization mirages," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–16.

[128] Worldometers, "COVID — Coronavirus statistics — Worldometer," Sep. 2023.

[129] S. Sule, M. C. DaCosta, E. DeCou, C. Gilson, K. Wallace, and S. L. Goff, "Communication of COVID-19 misinformation on social media by physicians in the US," *JAMA Netw. Open*, vol. 6, no. 8, p. e2328928, Aug. 2023.

[130] O. Papakyriakopoulos and E. Goodman, "The impact of Twitter labels on misinformation spread and user engagement: Lessons from Trump's election tweets," in *Proc. ACM Web Conf.*, Apr. 2022, pp. 2541–2551.

[131] S. Zannettou, ""I won the election!": An empirical analysis of soft moderation interventions on Twitter," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 15, pp. 865–876, May 2021.

[132] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Rev.*, vol. 51, no. 4, pp. 661–703, Dec. 2009.

[133] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?" in *Proc. 19th Int. Conf. World Wide Web*, Apr. 2010, pp. 591–600.

[134] Y. Li and Y. Xie, "Is a picture worth a thousand words? an empirical study of image content and social media engagement," *J. Mark. Res.*, vol. 57, no. 1, pp. 1–19, Jan. 2020.

[135] M. Lichman and P. Smyth, "Prediction of sparse user-item consumption rates with zero-inflated Poisson regression," in *Proc. World Wide Web Conf.*, Apr. 2018, pp. 719–728.

[136] E. Sun, I. Rosenn, C. Marlow, and T. Lento, "Gesundheit! modeling contagion through Facebook news feed," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 3, no. 1, pp. 146–153, Jul. 2009.

[137] N. King, "Using templates in the thematic analysis of text," in *Essential Guide to Qualitative Methods in Organizational Research*. Thousand Oaks, CA, USA: SAGE Publ. Ltd, 2004, pp. 256–270.

[138] L. Battle and A. Ottley, "What do we mean when we say "insight"? a formal synthesis of existing theory," *IEEE Trans. Vis. Comput. Graph.*, vol. 30, no. 9, pp. 6075–6088, Sep. 2024.

[139] V. Braun and V. Clarke, *Thematic Analysis: A Practical Guide*. Thousand Oaks, CA, USA: SAGE Publ. Ltd, 2022.

[140] V. Braun, V. Clarke, N. Hayfield, and G. Terry, "Thematic analysis," in *Handbook of Research Methods in Health Social Sciences*. Singapore: Springer Singapore, 2019, pp. 843–860.

[141] J. E. Uscinski and R. W. Butler, "The epistemology of fact checking," *Crit. Rev.*, vol. 25, no. 2, pp. 162–180, Jun. 2013.

[142] T. S. Kuhn and I. Hacking, *The structure of scientific revolutions*, 4th ed. Chicago, IL, USA: Univ. Chicago Press, 2012.

[143] B. Latour, *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA, USA: Harvard Univ. Press, 2015.

[144] N. Oreskes, "The scientific consensus on climate change," *Science*, vol. 306, no. 5702, pp. 1686–1686, Dec. 2004.

[145] N. Oreskes and E. M. Conway, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*, 1st ed. New York, NY, USA: Bloomsbury Press, May 2010.

[146] U. Shwed and P. S. Bearman, "The temporal structure of scientific consensus formation," *Am. Sociol. Rev.*, vol. 75, no. 6, pp. 817–840, Dec. 2010.

[147] S.-h. Hong, "Fact signalling and fact nostalgia in the data-driven society," *Big Data Soc.*, vol. 10, no. 1, p. 20539517231164118, Jan. 2023.

[148] J. Cook, N. Oreskes, P. T. Doran, W. R. L. Anderegg, B. Verheggen, E. W. Maibach, J. S. Carlton, S. Lewandowsky, A. G. Skuce, S. A. Green, D. Nuccitelli, P. Jacobs, M. Richardson, B. Winkler, R. Painting, and K. Rice, "Consensus on consensus: A synthesis of consensus estimates on human-caused global warming," *Environ. Res. Lett.*, vol. 11, no. 4, p. 048002, Aug. 2016.

[149] Cochrane Library, "About the Cochrane Database of Systematic Reviews — Cochrane Library," 2023.

[150] Q. Z. Chen and A. X. Zhang, "Judgment sieve: Reducing uncertainty in group judgments through interventions targeting ambiguity versus disagreement," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CSCW2, pp. 1–26, Oct. 2023.

[151] M. Wadman, "Antivaccine activists use a government database on side effects to scare the public," *Science*, Apr. 2021.

[152] Yahoo!, "Yahoo! Finance," Jan. 1997.

[153] H. Rosling and Z. Zhang, "Health advocacy with Gapminder animated statistics," *J. Epidemiol. Glob. Health*, vol. 1, no. 1, p. 11, Sep. 2011.

[154] Federal Reserve Bank of St. Louis, "Fred, federal reserve economic data." St. Louis, MO, USA, 1997.

[155] H. W. J. Rittel and M. M. Webber, "Dilemmas in a general theory of planning," *Policy Sci.*, vol. 4, no. 2, pp. 155–169, Jun. 1973.

[156] N. Micallef, V. Armacost, N. Memon, and S. Patil, "True or false: Studying the work practices of professional fact-checkers," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW1, pp. 1–44, Mar. 2022.

[157] N. Walter and R. Tukachinsky, "A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it?" *Commun. Res.*, vol. 47, no. 2, pp. 155–177, Mar. 2020.

[158] A. Shostack, *Threat Modeling: Designing for Security*. Indianapolis, IN, USA: Wiley, Feb. 2014.

[159] R. Somerville, "Climate science and EPA's greenhouse gas regulations. Testimony before the US House of Representatives Committee on Energy and Commerce Subcommittee on Energy and Power." Mar. 2011.

[160] C. Andrade, "Harking, cherry-picking, p-hacking, fishing expeditions, and data dredging and mining as questionable research practices," *J. Clin. Psychiatry*, vol. 82, no. 1, p. 25941, Feb. 2021.

[161] F. Büttner, E. Toomey, S. McClean, M. Roe, and E. Delahunt, "Are questionable research practices facilitating new discoveries in sport and exercise medicine? the proportion of supported hypotheses is implausibly high," *British Journal of Sports Medicine*, vol. 54, no. 22, pp. 1365–1371, Nov. 2020.

[162] A. Tversky and D. Kahneman, "Availability: A heuristic for judging frequency and probability," *Cognit. Psychol.*, vol. 5, no. 2, pp. 207–232, Sep. 1973.

[163] S. Plous, *The psychology of judgment and decision making*. New York, NY, USA: McGraw-Hill Higher Education, 2007.

[164] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, Sep. 1974.

[165] È. Dubé, J. K. Ward, P. Verger, and N. E. MacDonald, "Vaccine hesitancy, acceptance, and anti-vaccination: Trends and future prospects for public health," *Annu. Rev. Public Health*, vol. 42, no. 1, pp. 175–191, Apr. 2021.

[166] M. Hoofnagle and C. J. Hoofnagle, "What is denialism?" *SSRN Electron. J.*, Apr. 2007.

[167] M. Mayer and R. J. Evans, *They thought they were free: The Germans, 1933–45*. Chicago, IL, USA: Univ. Chicago Press, Feb. 2017.

[168] L. Reina, "Why movie blurbs avoid newspapers.." *The Free Library. Duncan McIntosh Company, Inc.*, Jun. 1996.

[169] A. Barchetti, E. Neybert, S. P. Mantel, and F. R. Kardes, "The half-truth effect and its implications for sustainability," *Sustainability*, vol. 14, no. 11, p. 6943, Jun. 2022.

[170] Z. Braiterman, A. Shostack, J. Marcil, S. de Vries, I. Michlin, K. Wuyts, R. Hurlbut, B. S. Schoenfeld, F. Scott, M. Coles, C. Romeo, A. Miller, I. Tarandach, A. Douglen, and M. French, "Threat modeling manifesto," 2020.

[171] S. P. Dow, A. Glassco, J. Kass, M. Schwarz, D. L. Schwartz, and S. R. Klemmer, "Parallel prototyping leads to better design results, more divergence, and increased self-efficacy," *ACM Trans. Comput.-Hum. Interact.*, vol. 17, no. 4, pp. 1–24, Dec. 2010.

[172] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts, "Visual comparison for information visualization," *Inf. Vis.*, vol. 10, no. 4, pp. 289–309, Oct. 2011.

[173] M. Gleicher, "Considerations for visualizing comparison," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 413–423, Aug. 2018.

[174] L. Padilla, R. Fygenson, S. C. Castro, and E. Bertini, "Multiple forecast visualizations (MFVs): Trade-offs in trust and performance in multiple COVID-19 forecast visualizations," *IEEE Trans. Vis. Comput. Graph.*, pp. 1–11, Oct. 2022.

[175] J. Bollinger, *Bollinger on Bollinger Bands*. New York, NY, USA: McGraw-Hill, 2002.

[176] R. T. Whitaker, M. Mirzargar, and R. M. Kirby, "Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2713–2722, Dec. 2013.

[177] D. Moritz and D. Fisher, "Visualizing a million time series with the density line chart," *arXiv:1808.06019 [cs]*, Aug. 2018.

[178] W. Javed and N. Elmqvist, "Stack zooming for multi-focus interaction in time-series data visualization," in *Proc. IEEE Symp. Pacific Vis.*, Mar. 2010, pp. 33–40.

[179] E. R. Tufte, *The Visual Display of Quantitative Information*, 2nd ed. Cheshire, CT, USA: Graphics Press, 2001.

[180] J. Heer, N. Kong, and M. Agrawala, "Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2009, pp. 1303–1312.

[181] W. Willett, J. Heer, and M. Agrawala, "Scented widgets: Improving navigation cues with embedded visualizations," *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1129–1136, Nov. 2007.

[182] Binance, "Coinmarketcap," 2013.

[183] S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999.

[184] E. Cuenca, A. Sallaberry, F. Y. Wang, and P. Poncelet, "MultiStream: A multiresolution streamgraph approach to explore hierarchical time series," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 12, pp. 3160–3173, Dec. 2028.

[185] B. Morrow, T. Manz, A. E. Chung, N. Gehlenborg, and D. Gotz, "Periphery plots for contextualizing heterogeneous time-based charts," in *IEEE Vis. Conf.*, Oct. 2019, pp. 1–5.

[186] M. Correll, "Teru teru bōzu: Defensive raincloud plots," *Comput. Graph. Forum*, vol. 42, no. 3, pp. 235–246, Jun. 2023.

[187] N. Gershon, "Visualization of an imperfect world," *IEEE Comput. Graph. Appl.*, vol. 18, no. 4, pp. 43–45, Jul. 1998.

[188] L. Padilla, M. Kay, and J. Hullman, "Uncertainty visualization," in *Computational Statistics in Data Science*. Indianapolis, IN, USA: Wiley, Sep. 2022.

[189] J. Wang, S. Hazarika, C. Li, and H.-W. Shen, "Visualization and visual analysis of ensemble data: A survey," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 9, pp. 2853–2872, Sep. 2019.

[190] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan, "Visual semiotics & uncertainty visualization: An empirical study," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2496–2505, Dec. 2012.

[191] C. Stokes, C. Sanker, B. Cogley, and V. Setlur, "From delays to densities: Exploring data uncertainty through speech, text, and visualization," *Comput. Graph. Forum*, vol. 43, no. 3, pp. 1–14, Jun. 2024.

[192] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel, "A typology for visualizing uncertainty," in *Vis. Data Anal.*, vol. 5669, Jan. 2005, pp. 146–157.

[193] Z. Wang, J. Ritchie, J. Zhou, F. Chevalier, and B. Bach, "Data comics for reporting controlled user studies in human-computer interaction," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 967–977, Feb. 2021.

[194] Y. Ding, J. Wilburn, H. Shrestha, A. Ndlovu, K. Gadhave, C. Nobre, A. Lex, and L. Harrison, "reVISit: Supporting scalable evaluation of interactive visualizations," in *Proc. IEEE Vis. Conf.*, Oct. 2023, pp. 31–35.

[195] Z. T. Cutler, K. Gadhave, and A. Lex, "Trrack: A library for provenance tracking in web-based visualizations," in *Proc. IEEE Vis. Conf.*, Oct. 2020, pp. 116–120.

[196] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in *Adv. Psychol.*, 1988, vol. 52, pp. 139–183.

[197] J. Berg, J. Dickhaut, and K. McCabe, "Trust, reciprocity, and social history," *Games Econ. Behav.*, vol. 10, no. 1, pp. 122–142, Jul. 1995.

[198] H. Elhamdadi, A. Gaba, Y.-S. Kim, and C. Xiong, "How do we measure trust in visual data communication?" in *Proc. IEEE BELIV Workshop*, Oct. 2022, pp. 85–92.

[199] H. Elhamdadi, L. Padilla, and C. Xiong, "Using processing fluency as a metric of trust in scatterplot visualizations," *arXiv*, Sep. 2022.

[200] A. Gaba, Z. Kaufman, J. Cheung, M. Shvakel, K. W. Hall, Y. Brun, and C. X. Bearfield, "My model is unfair, do people even care? visual design affects trust and perceived bias in machine learning," *IEEE Trans. Vis. Comput. Graph.*, vol. 30, no. 1, pp. 327–337, Jan. 2024.

[201] T. J. Prestby, "Trust in maps: What we know and what we need to know," *Cartogr. Geogr. Inf. Sci.*, pp. 1–18, Nov. 2023.

[202] C. Xiong, L. M. K. Padilla, K. Grayson, and S. Franconeri, "Examining the components of trust in map-based visualizations," *EuroVis Workshop Trustworthy Vis. (TrustVis)*, pp. 1–5, Jun. 2019.

[203] F. Yang, C. R. Mortenson, E. Nisbet, N. Diakopoulos, and M. Kay, "In dice we trust: Uncertainty displays for maintaining trust in election forecasts over time," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2024, pp. 1–24.

[204] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc. B*, vol. 57, no. 1, pp. 289–300, Mar. 1995.

[205] G. J. Quadri, A. Z. Wang, Z. Wang, J. Adorno, P. Rosen, and D. A. Szafir, "Do you see what i see? a qualitative study eliciting high-level visualization comprehension," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2024, pp. 1–26.

[206] W. Javed, B. McDonnel, and N. Elmqvist, "Graphical perception of multiple time series," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 927–934, Nov. 2010.

[207] B. Ondov, N. Jardine, N. Elmqvist, and S. Franconeri, "Face to face: Evaluating visual comparison," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 861–871, Jan. 2019.

[208] X. Lan and Y. Liu, ""I came across a junk": Understanding design flaws of data visualization from the public's perspective," *IEEE Trans. Vis. Comput. Graph.*, vol. 31, no. 1, pp. 393–403, Jan. 2025.

[209] C. Stokes, V. Setlur, B. Cogley, A. Satyanarayan, and M. A. Hearst, "Striking a balance: Reader takeaways and preferences when integrating text and charts," *IEEE Trans. Vis. Comput. Graph.*, pp. 1–11, MONTH 2022.

[210] S. Lewandowsky, J. Cook, and D. Lombardi, "Debunking handbook 2020," 2020.

[211] U. K. H. Ecker, S. Lewandowsky, B. Swire, and D. Chang, "Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction," *Psychon. Bull. Rev.*, vol. 18, no. 3, pp. 570–578, Jun. 2011.

[212] U. K. H. Ecker, J. L. Hogan, and S. Lewandowsky, "Reminders and repetition of misinformation: Helping or hindering its retraction?" *J. Appl. Res. Mem. Cogn.*, vol. 6, no. 2, pp. 185–192, Jun. 2017.

[213] P. Schmid and C. Betsch, "Effective strategies for rebutting science denialism in public discussions," *Nat. Hum. Behav.*, vol. 3, no. 9, pp. 931–939, Sep. 2019.

[214] J. Cook, S. Lewandowsky, and U. K. H. Ecker, "Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence," *PLOS ONE*, vol. 12, no. 5, p. e0175799, May 2017.

[215] U. K. H. Ecker, S. Lewandowsky, and M. Chadwick, "Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect," *Cogn. Res. Princ. Implic.*, vol. 5, no. 1, p. 41, Aug. 2020.

[216] H. M. Johnson and C. M. Seifert, "Sources of the continued influence effect: When misinformation in memory affects later inferences," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 20, no. 6, pp. 1420–1436, Dec. 1994.

[217] U. K. H. Ecker, S. Lewandowsky, and D. T. W. Tang, "Explicit warnings reduce but do not eliminate the continued influence of misinformation," *Mem. Cognit.*, vol. 38, no. 8, pp. 1087–1100, Dec. 2010.

[218] C. M. Seifert, "The continued influence of misinformation in memory: What makes a correction effective?" in *Psychol. Learn. Motiv.* Academic Press, Jan. 2002, vol. 41, pp. 265–292.

[219] G. M. Sinatra and D. Lombardi, "Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments," *Educ. Psychol.*, vol. 55, no. 3, pp. 120–131, Jul. 2020.

[220] G. Lakoff, *Moral Politics: How Liberals and Conservatives Think*, 3rd ed. Chicago, IL, USA: Univ. Chicago Press, 2016.

[221] C. Y. Johnson, Y. Abutaleb, and J. Achenbach, "CDC study shows three-fourths of people infected in Massachusetts coronavirus outbreak were vaccinated but few required hospitalization," *The Washington Post*, Jul. 2021.

[222] B. L. Lovelace Jr, "CDC study shows 74% of people infected in Massachusetts Covid outbreak were fully vaccinated," Jul. 2021.

[223] WHO, "Global influenza surveillance and response system (GISRS)," 1952.

[224] A. J. Hay and J. W. McCauley, "The WHO global influenza surveillance and response system (GISRS)—a future perspective," *Influenza Other Respir. Viruses*, vol. 12, no. 5, pp. 551–557, Sep. 2018.

[225] B. Flannery, R. K. Zimmerman, L. V. Gubareva, R. J. Garten, J. R. Chung, M. P. Nowalk, M. L. Jackson, L. A. Jackson, A. S. Monto, S. E. Ohmit, E. A. Belongia, H. Q. McLean, M. Gaglani, P. A. Piedra, V. P. Mishin, A. P. Chesnokov, S. Spencer, S. N. Thaker, J. R. Barnes, A. Foust, W. Sessions, X. Xu, J. Katz, and A. M. Fry, "Enhanced genetic characterization of influenza a(H3N2) viruses and vaccine effectiveness by genetic group, 2014–2015," *J. Infect. Dis.*, vol. 214, no. 7, pp. 1010–1019, Oct. 2016.

[226] H. Takeuchi and R. Kawashima, "Disappearance and re-emergence of influenza during the COVID-19 pandemic: Association with infection control measures," *Viruses*, vol. 15, no. 1, p. 223, Jan. 2023.

[227] V. Dhanasekaran, S. Sullivan, K. M. Edwards, R. Xie, A. Khvorov, S. A. Valkenburg, B. J. Cowling, and I. G. Barr, "Human seasonal influenza under COVID-19 and the potential consequences of influenza lineage elimination," *Nat. Commun.*, vol. 13, no. 1, p. 1721, Mar. 2022.

[228] K. L. Laurie and S. Rockman, "Which influenza viruses will emerge following the SARS-CoV-2 pandemic?" *Influenza Other Respir. Viruses*, vol. 15, no. 5, pp. 573–576, Sep. 2021.

[229] A. Bleske-Rechek, K. M. Morrison, and L. D. Heidtke, "Causal inference from descriptions of experimental and non-experimental research: Public understanding of correlation-versus-causation," *J. Gen. Psychol.*, vol. 142, no. 1, pp. 48–70, Jan. 2015.

[230] T. E. Kida, *Don't Believe Everything You Think: The 6 Basic Mistakes We Make in Thinking*. Amherst, NY, USA: Prometheus Books, 2006.

[231] M. McMorrow, "Improving communications around vaccine breakthrough and vaccine effectiveness," Vaccine Effectiveness Team, CDC, Aug. 2021.

[232] J. Allsop, "How major media outlets screwed up the vaccine 'breakthrough' story," Aug. 2021.

[233] A. Mandavilli, "C.D.C. internal report calls delta variant as contagious as chickenpox," *The New York Times*, Jul. 2021.

[234] E. J. Judd, J. E. Tierney, D. J. Lunt, I. P. Montañez, B. T. Huber, S. L. Wing, and P. J. Valdes, "A 485-million-year history of Earth's surface temperature," *Science*, vol. 385, no. 6715, p. eadk3705, Sep. 2024.

[235] S. Kaplan and S. Ducroquet, "Scientists have captured earth's climate over the last 485 million years. here's the surprising place we stand now." *The Washington Post*, Sep. 2024.

[236] J. Rogan, "Joe Rogan Experience #2254 - Mel Gibson," Jan. 2025.

[237] W. K. Frankena, "The naturalistic fallacy," *Mind*, vol. 48, no. 192, pp. 464–477, Oct. 1939.

[238] W. H. Bruening, "Moore and "is-ought"," *Ethics*, vol. 81, no. 2, pp. 143–149, Jan. 1971.

[239] M. Scott and R. Lindsey, "What's the hottest Earth's ever been?" *NOAA Climate.gov*, Feb. 2025.

[240] M. Correll and J. Heer, "Surprise! bayesian weighting for de-biasing thematic maps," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 651–660, Jan. 2017.

[241] G. Mulvihill, "States with few COVID-19 cases get big haul of relief money," *Assoc. Press*, May 2020.

[242] M. Gabel, J. Gooblar, C. M. Roe, N. J. Selsor, and J. C. Morris, "Political ideology, confidence in science, and participation in alzheimer disease research studies," *Alzheimers Dis. Assoc. Disord.*, vol. 32, no. 3, pp. 179–184, Jul. 2018.

[243] L. Donnelly-Smith, E. Guarinello, B. Laurie, J. Olsson, L. Poster, J. Powell, A. Roberts Reeder, and J. Tasse, "The Smithsonian Institution's guide to interpretive writing for exhibitions," The Smithsonian Institution, Tech. Rep., 2021.

[244] P. Marincola, *What makes a great exhibition?* London, UK: Reaktion Books, 2014.

[245] S. Heekeren, "Popular science images: Reflections on visual practices in science communication," *Des. Issues*, vol. 37, no. 4, pp. 72–85, Sep. 2021.

[246] A. Burns, "Yes, yes in yellow: Memes and what they tell us about (queer) use of data visualization," *JoVI*, vol. 1, no. 1, Apr. 2025.

[247] J. Fetterley, *The Resisting Reader: A Feminist Approach to American Fiction*. Bloomington, IN, USA: Indiana Univ. Press, 1989.

[248] D. Akbaba, L. Klein, and M. Meyer, "Entanglements for visualization: Changing research outcomes through feminist theory," *IEEE Trans. Vis. Comput. Graph.*, vol. 31, no. 1, pp. 1279–1289, Jan. 2025.

[249] C. D'Ignazio and L. F. Klein, *Data Feminism*. Cambridge, MA, USA: MIT Press, 2020.

[250] B. Latour, "Why has critique run out of steam? from matters of fact to matters of concern," *Crit. Inq.*, vol. 30, no. 2, pp. 225–248, Jan. 2004.

[251] E. K. Sedgwick, *Touching Feeling: Affect, Pedagogy, Performativity*. Durham, NC, USA: Duke Univ. Press, 2003.

[252] C. Castiglia, "Critiquiness," *Engl. Lang. Notes*, vol. 51, no. 2, pp. 79–85, Sep. 2013.

[253] R. Felski, *The Limits of Critique*. Chicago, IL, USA: Univ. Chicago Press, 2015.

[254] E. S. Anker and R. Felski, *Critique and Postcritique*. Durham, NC, USA: Duke Univ. Press, 2017.