DATA HUNCHES: EXPRESSING PERSONAL KNOWLEDGE IN DATA VISUALIZATIONS

by

Haihan Lin

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

Doctor of Philosophy

in

Computing

School of Computing
The University of Utah
December 2023

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Haihan Lin**

has been approved by the following supervisory committee members:

Alexander Lex,	Chair(s)	**Sep. 17, 2023** Date Approved
Miriah Meyer,	Member	**Sep. 18, 2023** Date Approved
Marina Kogan,	Member	**Oct. 24, 2023** Date Approved
Jason Stampfer Wiese,	Member	**Sep. 17, 2023** Date Approved
Eytan Adar,	Member	**Oct. 24, 2023** Date Approved

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

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

and by **Darryl P. Butt**, Dean of The Graduate School.

ABSTRACT

Data-driven decision-making has become the gold standard in science, industry, and public policy. The trouble with data is that it frequently provides only an imperfect and partial representation of a phenomenon of interest. The gap between data and reality makes data alone insufficient to make good analysis decisions and data interpretations, and, as a result, analysts and experts utilize personal knowledge from various sources to fill in the gap between data and reality. In practice, personal knowledge is typically not incorporated in analysis tools in a structured way, which is problematic if others who lack that knowledge interpret the data. This dissertation centers around the topic of data hunches, an analyst's knowledge about how and why data is an imperfect and partial representation of the phenomena of interest, and investigates how experts' knowledge about data is utilized in data analysis and how interactive data visualizations can facilitate the process of recording and communicating experts' knowledge for the analysis process. The dissertation makes three contributions to the topic: 1) an analysis of interview studies with analysts from a wide range of domains and with varied expertise and experience inquiring about the role of contextual knowledge and the process of incorporating various sources of knowledge into analyses; 2) defining, theorizing, and characterizing experts' knowledge about data and data as *data hunches*, and positioning data hunches in the existing understanding of uncertainty; 3) proposing a framework and guidelines to design visualizations that support recording and communicating data hunches through visualizations intuitively and effectively. This dissertation aims to elevate the role of experts' knowledge in data analysis and provides guidelines and techniques to design visualizations to support externalizing knowledge explicitly. Through the analysis and proposed guidelines, it is envisioned that data hunches will empower analysts to externalize their knowledge, facilitate collaboration and communication, support the ability to learn from others' data hunches, and ultimately, lead to better data-driven decision-making.

To my parents.

CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my gratitude towards my committee members, Alexander Lex, Miriah Meyer, Marina Kogan, Jason Wiese, and Eytan Adar. I am grateful to have Alex as my advisor for my PhD, and I deeply appreciate his support and guidance throughout my PhD. Working alongside him has been an absolute honor.

Thank you to the wonderful members and alumni of the Visualization Design Lab, who have helped me immensely throughout my journey. Thank you to my co-authors whom I have learned so much, Amanda Bakian, Ryan A. Metcalf, Derya Akbaba, Carolina Nobre, Jack Wilburn, Max Lisnic, Alex, and Miriah. I am forever grateful to my family and friends who have always been there for me. This dissertation would not have happened without all those mentioned above.

This dissertation extends published works for which we have permission to re-use. ©2021 SAGE. Reprinted [1], with permission, from Lin, Metcalf, Wilburn, and Lex, Sanguine: Visual Analysis for Patient Blood Management, *Information Visualization*, August 2021. ©2022 IEEE. Reprinted [2], with permission, from Lin, Akbaba, Meyer, and Lex, Data Hunches: Incorporating Personal Knowledge into Visualizations, *IEEE Transactions on Visualization and Computer Graphics*, September 2022.

This dissertation was supported by ARUP laboratories; the National Science Foundation (IIS 1751238, OAC 1835904); the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation; the National Institutes of Health (1S10OD021644-01A1).

CHAPTER 1

INTRODUCTION

This dissertation delves into the crucial role of personal knowledge in data analysis and introduces the concept of "data hunches." Data hunches describe an analyst's understanding of why and how data does not always accurately reflect reality. Our formative design study [1] revealed challenges in which the data did not align with the domain experts' expectations and experiences, resulting in a visualization interface that did not accurately represent reality for the experts. Building on our experience with our formative design study and other, prior work, we conducted an interview study with 14 domain experts and analysts from various fields, investigating how they account for data caveats in their workflows and how expert knowledge fills the gap between data and reality [3]. Additionally, we examined how analysts document and communicate data caveats and knowledge in their work. Based on these foundations, we explored the topic of personal knowledge and data imperfections in more depth, introducing the concept of "data hunches" as a formal way to describe an analyst's understanding of how and why data falls short in representing phenomena of interest. By defining data hunches, we elevate the role of personal knowledge in data analysis. Furthermore, we explore a range of techniques for recording and communicating data hunches through interactive visualizations.

1.1 Motivation and Overview

On September 26, 1983, the Soviet Air Defense Forces' computers reported five missiles heading toward the Soviet Union from the United States, triggering a protocol that called for an immediate and compulsory nuclear counterattack. However, Stanislav Petrov, the officer on duty, relied on his expert knowledge and determined that the incoming strike warning was more likely a system malfunction rather than a real attack. Petrov believed that if the US were to strike first, it would be massive, rather than just five missiles, as the data was suggesting. He made the crucial decision to disregard the warning and not launch a nuclear attack, despite having no data to confirm his interpretation [4]. Later investigation revealed that the system had indeed malfunctioned due to a rare alignment of the detection satellite and the sun. Petrov's knowledge and experience enabled him to recognize the possibility of a false alarm and to interpret the data in the context of the political situation. Had he relied solely on the warning system, the consequences could have been catastrophic.

While not all data (mis)interpretations lead to world-shattering consequences, data-driven decision-making [5] has become the gold standard in fields like public policy [6], science [7], and industry [8], but also in making choices about our everyday lives. However, data alone is not sufficient to make good decisions; data is an imperfect and partial representation of reality [9], [10]. In the formative design study of this dissertation presented in Chapter 4, we implemented an interactive data analysis interface for medical professionals to analyze transfusion practices with interventions and patient outcomes. The interactive data visualizations allowed our collaborators to explore and analyze transfusion practices and its associated medical outcomes. However, when we showed the data visualizations to the target audience (surgeons), they remarked that there seemed to be a discrepancy between the data behind the visualizations and what they did in practice. Upon further investigation, we did not discover any issue with the data compilation. Hence, we concluded that the data collected did not fully capture reality.

The phenomenon that data does not match personal knowledge has surfaced in previous visualization studies as well [11]–[14]. This is common in data analyses, but as researchers, we have a limited understanding of what kind of knowledge is often used, if the knowledge is incorporated into the analysis, and whether the analysts documented the process. To answer these questions, we set out to conduct an interview study with 14 analysts to investigate how they use knowledge to bridge data and reality in their own workflows and how the knowledge is reflected in their communication and analysis results. The results of the interview study are presented in Chapter 5. After the interviews and subsequent analyses, we had deeper insights into how analysts incorporated personal knowledge about the data into their own workflow, either from their own experience from working on the subject, or from experts with deeper domain understanding. We found that combining knowledge and data in the analysis was a common practice in various domains; we illustrate this analysis pipeline in Figure 1.1. The interviews also showed that the recording of knowledge in the analysis is limited and ad hoc. Most of our interviewees used commercially available data analysis tools like Excel and Tableau. Even though these tools have some annotation features, almost none of our participants used them, either worrying that the annotation would negatively impact their existing data structure or because the annotation process was not intuitive and expressive enough for our interviewees. We found that the existing processes made the knowledge about data personal and not available to others [11], [15], [16], and identified many design opportunities to better support analysts to record knowledge and make it available to colleagues and collaborators.

Through the formative design study and the interviews with analysts, we found that the existing research does not provide enough depth to understand how personal knowledge affects data analysis. The most related term used to describe similar concept is uncertainty. Uncertainty is often used to describe characteristics of the data itself [17], whereas, we found that knowledge is rarely part of the data itself — it is added by analysts, domain experts, or simply those with more living experience with what the data is describing. Uncertainty (or more specifically, epistemic uncertainty [18]) is also often used to describe something unknown due to a lack of knowledge or information. Based on our formative design study and interviews with analysts, relevant information about the data was known to some, such as domain and local experts, but can be unknown to the analysts without domain knowledge and the visualization designer who build tools to support the analysis. Therefore, we found the term uncertainty is not sufficient to describe how personal knowledge is separate from the data. We coined the term "data hunches" to describe the personal knowledge that domain and local experts have about how data does not perfectly represent the world, presented in Chapter 6. In our interviews, we also found that the existing approach of recording data hunches is limited to annotations, and there is potential for technological interventions to help with recording data hunches.

What if experts could record their hunches directly in a visualization tool in a way that allowed others to interpret the data alongside their hunches? What if a visualization tool supported—and encouraged—the explicit incorporation of expert knowledge during data analysis? Based on the interviews, we think there is a need for methods and tools to help with recording and presenting data hunches. In Chapter 6, we present a design space for data hunches and discuss techniques that are suitable to intuitively and expressively communicate personal knowledge about the data, including annotations, sketching, and manipulating the visualization, without jeopardizing the legibility and separation between the data hunches and the original visualizations. To consider data hunches in practical applications, we provide a set of design recommendations that designers should consider when incorporating proposed techniques or additional techniques into their applications. In addition, we implement a prototype to demonstrate how data hunches can be expressed on a simple bar chart.

In conclusion, this dissertation explores how personal knowledge is a critical part of data analysis and how data visualizations can facilitate this process to make the knowledge accessible and expressive. Through this dissertation, we seek to question the notion of only data being the gold standard for representing phenomena in the world and open up the potential to grow visualization research beyond constrained notions of data.

1.2 Contributions

The dissertation makes three contributions to the topic of personal knowledge in data analysis. (1) **We provide an analysis of an interview study** with analysts from a wide range of domains and with varied expertise and experience inquiring about the role of contextual knowledge and the process of incorporating various sources of knowledge into analyses. The interview analysis also leads us to identify design opportunities for possible technological interventions for expressing data hunches in applications. (2) **We define data hunches**, an analyst's knowledge about how and why data is an imperfect and partial representation of the phenomena of interest, and discuss data hunches' relationship to uncertainty. The recognition and theorization of data hunches elevate the role of knowledge in data analysis and provide a new perspective on how to design visualization tools for experts that allow sharing of knowledge about the ways in which data is imperfect. (3) **We propose an array of visualization techniques** to support recording and communicating data hunches through visualizations intuitively and effectively, as well as **a set of design recommendations** for incorporating data hunches in applications. Additionally, to demonstrate these techniques, we present a prototype to annotate data hunches with the techniques. These three contributions expand our current understanding of how personal knowledge aids data analysis and how data visualizations can facilitate and explicitly communicate the integration of knowledge in data analysis.

An additional contribution of this dissertation comes from the formative design study that inspired our work in data hunches and personal knowledge. Through the formative design study, we contribute an open-source visual analytic tool for analyzing patient blood management data, an analysis of domain goals, and a task abstraction for visualizing transfusion practices.

1.3 Organization

This dissertation presents a portion of the research work I conducted for my PhD. Chapter 2 provides background and related works on data hunches and interview studies regarding personal knowledge in data analysis. Chapter 2 also positions the data hunch definition with the existing understanding of uncertainty. Chapter 3 presents the definition of data hunches in detail and discusses the methodology we adopted to theorize and define data hunches. Chapter 4 describes the formative design study Sanguine, which focuses on an interactive visual analytic tool for visualizing transfusion practices, based on our Information Visualization journal publication [1]. Chapter 5 presents the interview study conducted with analysts and domain experts from a broad range of backgrounds, which sheds light on how they incorporate data and domain knowledge into their analysis practices. This Chapter 5 is based on a preprint currently under review [3]. In Chapter 6, we explore various techniques for recording and communicating data hunches through visualizations and propose guidelines for implementing data hunches in applications. Chapter 3 and 6 are based on my published work from IEEE VIS 2022 [2]. Chapter 7 explores the implications of recognizing the critical role of data hunches in data analysis and discusses how the visualization community can design interfaces that take data hunches into consideration. Finally, Chapter 8 concludes this dissertation by highlighting potential future directions to continue the work on externalizing knowledge through visualizations.

This dissertation does not follow the order in which the works were published. The interview study was conducted after the work on data hunches was published. However, the interview study offers a broader perspective on how analysts' knowledge is incorporated into the analysis. The work on data hunches is more focused on defining the knowledge of data as data hunches and how specific visualization techniques can be used to record and communicate data hunches. Therefore, the interview study is presented before the work on data hunches in this dissertation. All chapters use plural pronouns because all works were collaborative.

Several research projects that I worked on during my PhD are not included in this dissertation. These include the technique called Clipped Graphs [19] to visualize compact time series in tabular format, network visualizations on family geographical proximity and longevity using the Utah Population Database, and interactive visualizations on Computer Science enrollment in Utah with the Utah Board of Education [20].

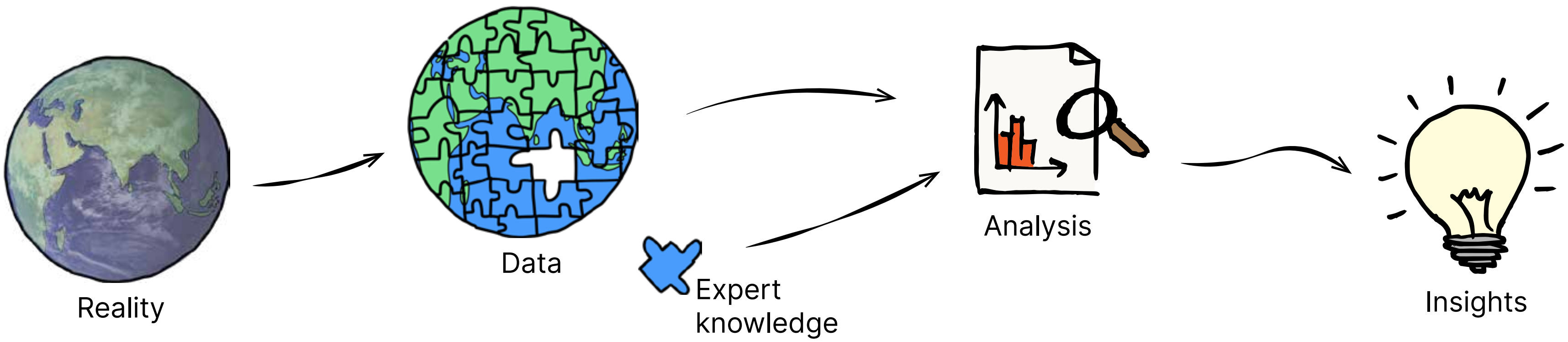


Figure 1.1: An overview of the role of expert knowledge in data analysis workflows. Data is an imprecise and incomplete representation of reality. Expert knowledge helps with understanding the limitations of the data and may fill in the gaps between data and reality. Data analysis should leverage both knowledge and data to arrive at robust insights.

CHAPTER 2

BACKGROUND AND RELATED WORK

In this chapter, we first discuss how uncertainty is defined in the visualization research community, and how uncertainty is problematic for describing personal knowledge of the data. Secondly, we present an overview of the existing techniques within the visualization research community from which we draw inspiration in our exploration of techniques to express data hunches. Finally, we outline the existing research into how data analysts collaborate and communicate with data caveats.

2.1 Uncertainty

In this section, we present an overview of how uncertainty is defined, primarily within the visualization research field, and how visualization designers deal with uncertainty in applications.

2.1.1 Defining Uncertainty

Measurement errors, modeling assumptions, heterogeneous data recording methods, and missing context are just some of the ways in which values stored in a dataset are neither a perfect nor complete view of a phenomenon. The visualization community has a long history of researching methods to characterize, quantify, and visualize the limitations of data under the heading of *uncertainty*. As all data is collected, quantified, measured, or simulated, our inability to perfectly measure the data translates into imperfections in data [21]. Almost all data has uncertainty to some degree; as Richard Feynman said, "What is not surrounded by uncertainty cannot be the truth."

Explicit definitions of uncertainty are sparse in the literature, with the definitions that do exist covering a range of interpretations. For example, Hullman refers to uncertainty as "the possibility that the observed data or model predictions could take on a set of possible values" [22], which suggests that the possible values can be known in theory, but, in practice, are often overlooked, or simply categorized as unknown. The review from Bonneau et al. stated that "Uncertainty is the lack of information" [21].

Uncertainty is not a unique problem within the data visualization community. Many experts from other fields, such as weather forecasting [23], [24], geography [25], [26], and data-driven policy making [27], [28] also deal with uncertainty in their research and analyses. Bhatt et al. emphasized the importance of disclosing uncertainty to increase transparency in existing machine learning pipelines [29]. Fields outside visualization also have added interpretations of the term, such as the Walker et al. general notion of uncertainty as "any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system" [30]; Ravetz and Funtowicz's characterization of uncertainty as "inexactness, unreliability and border with ignorance" [31]; and the Covitt et al. experiential definition of uncertainty as the "ways in which scientists recognize and analyze limits in their studies and conclusions" [32]. Regardless of the variations in definitions, the existing definitions of uncertainty across various fields heavily focus on how the data itself is limited, and thus the outcomes from analyses are limited by the data.

Researchers have openly acknowledged the ill-defined nature of uncertainty. Boukhelifa et al. [33] stated that "There is no unified single definition of uncertainty across all domains. The general consensus is that there are different meanings and that the term itself encapsulates many concepts." Brodlie et al. [34] went further and argued that the lack of a clear, consensus definition has held the field back: "The self-referential problem of uncertainty about uncertainty terminology has been a notable stumbling block in this avenue of inquiry."

The goal of this dissertation is not to attempt to rectify the uncertainty about uncertainty. In the following subsections, we examine the most often-used categorization for uncertainty, qualitative and quantitative, and point out the difficulty of precisely positioning theoretical perspectives of personal knowledge in relation to the existing uncertainty literature. Despite this challenge, we situate data hunches within the visualization uncertainty literature because this body of work has focused on understanding and characterizing how people, visualizations, and imperfect and partial data come together. More specifically, the uncertainty literature focuses on the ways that data is a limited representation of the world, and how people become aware of these limitations. The focus on data itself is in contrast to other visualization research threads that focus on expert knowledge more broadly, such as work on the insight that considers different types of experts' knowledge that impact insight generation [35].

2.1.1.1 Quantitative Uncertainty

Researchers provide a variety of characterizations of uncertainty through descriptions of the many ways that data can be uncertain. Potter et al. used a characterization from computational sciences that describes *epistemic* versus *aleatoric* uncertainty. Epistemic uncertainty describes the ways in which a lack of knowledge about and from the data induce a computationally unknowable uncertainty, and aleatoric uncertainty encompasses data limitations that can be assumed and modeled statistically [18]. Padilla et al. compared *direct quantitative* uncertainties and *indirect qualitative* uncertainties [17]. In this framing, direct uncertainties are quantifiable expressions such as confidence intervals and probability distributions, whereas indirect uncertainties can be expressed only qualitatively. Direct, quantifiable expressions of uncertainty are typically computed from sources of imperfections and partialities: limited resources, such as the limited computation power in making climate model assessment [36]; bounded by precision, such as the precision limitation of the measuring tool [37]; or limited knowledge, such as the unpredictability of weather forecasts [23].

The visualization community has historically focused on developing and testing methods for visualizing quantifiable uncertainty [21], [38], [39]. Some proposed techniques encode uncertainty through modifications of a data item's graphical mark using quantile dot plots [17], glyphs [38], [40], [41], or ensemble plots [38], shown in Figure 2.1. Other approaches have instead explored visual representations that directly display summary statistics [42]–[44] or use animations showing hypothetical outcomes [45]. Researchers have also developed uncertainty-specific evaluation techniques, such as eliciting users' internal models of probability distributions, recording the effects of uncertainty on decision-making, and assessing participants' sense of confidence after viewing uncertainty visualizations [46].

2.1.1.2 Qualitative Uncertainty

Qualitative uncertainty—also referred to as indirect or epistemic uncertainty [18]—has been described as "The quality of knowledge concerning how effectively facts, numbers, or hypotheses represent reality" [17]. Definitions of qualitative uncertainty make explicit reference to knowledge, shifting the emphasis from exploitable information about the data to inaccessible subjective knowledge. Van der Bles et al. defined qualitative uncertainty as "phenomena that we currently do not know but could, at least in theory, know or establish," and unlike quantitative uncertainty that can be clearly described in values, qualitative uncertainty is generally conveyed through "caveats about data" [47].

The sources of qualitative uncertainty stem from the imperfections and partialities that metrics for quantifying uncertainty pertain to. Boukhelifa et al. [33] classify these sources as imperfect, messy, and missing data; imperfect and limited models; approximate digital representations of data in a visualization interface; and cognitive differences between interpretations from individual analysts. In addition, McCurdy et al. identified implicit error, as the sources of qualitative uncertainty, "a type of measurement error that is inherent to a dataset but not explicitly recorded, yet is accounted for qualitatively by experts during analysis, based on their implicit domain knowledge" [11].

The predominant way that visualization designers encode qualitative expressions of uncertainty is through text-based annotations. In her interview study with visualization practitioners, Hullman [22] reported that:

> Uncertainty as a qualitative expression of a gap in knowledge came up in most interviews with interviewees as well as several survey responses. 62% of survey respondents had used text to warn their viewers of the potential for uncertainty in results.

Researchers have used visual approaches to communicate qualitative uncertainty, such as the use of perceptually imprecise visual encoding channels, e.g., sketchiness [48] and glyphs [15]. A different approach taken in both the visualization and machine learning communities is to explicitly expose information about the data collection process, providing analysts with contextual information that allows them to incorporate personal knowledge about potential shortcomings of the data during their interpretation [16], [49], [50].

Several visualization systems have explored ways to record expert knowledge about qualitative sources of uncertainty. One notable example is the work of McCurdy et al. [11], shown in Figure 2.2. In their tool for supporting public health experts, they designed a template with structured questions that enabled the experts to record what they know about implicit errors in the data. The recorded results were marked on the visualizations with glyphs that displayed annotations when interacted with. Similarly, Franke et al. [51] collected levels of confidence about data sources from historians through a web interface, and varying levels of confidence were then presented alongside other data in a hierarchical tree view, representing the distribution of confidence along different dimensions of the data source in question. These examples demonstrate how visualization systems can serve as effective tools for recording and presenting expert knowledge regarding qualitative sources of uncertainty. However, existing systems do not fully utilize the visual elements in visualization systems, such as using the existing channels and marks of the visualization to show personal knowledge juxtaposed with the data.

2.1.2 Trouble with Uncertainty

We face several challenges when using uncertainty to describe personal knowledge about the data. The visualization community's characterization of what types of uncertainty are quantifiable would seemingly place our blood-reuse example in Chapters 1 and 4 as something that could be quantified—a limitation of the measurement capabilities. Indeed, in principle, we could attempt to model and quantify this limitation, but any metric is likely to be grossly inaccurate because of the abstracted nature of the knowledge about the imperfections, a point raised by Thomson et al. [52]:

> In addition to uncertain measures, analysts are concerned with abstract uncertainties such as the credibility of a particular source or the completeness of a set of information. As the uncertainty becomes more abstract, it is more difficult to quantify, represent, and understand.

Qualitative uncertainty identifies uncertainty that is "due to lack of knowledge and limited data which could, in principle, be known, but in practice are not" [18], [21]. However, the focus on unknown information contradicts what we have observed: our collaborators knew about the reality of transfusion practices based on their experience. This contradiction leads us to ask the question: unknown by whom? The mismatch between their experience and the data was unknown to us, the visualization designers, but our collaborators were well aware of the issue. The focus on unknowable information also presents the challenge of designing visualizations to express and communicate personal knowledge about the data. If it is unknown to us, how can we begin to explore the design space to support expressing personal knowledge?

This dissertation presents a significant shift in how we think about designing for qualitative uncertainty. Data hunches shift the knowledge about sources of uncertainty away from the visualization tool designers, to the expert analysts who conduct the analyses. The concept of data hunches is an explicit acknowledgment that many sources of qualitative uncertainty are, in fact, known and certain—known to the experts who can articulate the knowledge when triggered by their interactions with a visualization tool. These experts' hunches may be about a more precise value or could be a hunch about the data collection process or agencies. This shift complements the visualization community's perspectives regarding uncertainty by focusing on knowledge about sources of uncertainty and thus opening up opportunities to design new ways that visualizations can support recording and communicating this knowledge during data analysis. Leveraging the new techniques, this dissertation also focuses on empowering users to record their own knowledge—data hunches—on the visualizations instead of being given the information by the visualization designers. The ultimate goal of the new techniques is to use interactive data visualizations as a medium to obtain information and share knowledge about the data.

2.2 Techniques for Expressing Knowledge in Visualizations

The techniques of expressing data hunches are based on the established understanding of interactions and communication in visualization community. Some of the techniques have been studied by the visualization community, and some have been used in different scenarios. This section provides an overview of the existing work on these techniques.

2.2.1 Annotations

In the visualization research community, annotations are records that people append on top of the visualizations; at the same time, the content is closely related to the visualizations. Annotations are widely used for various purposes, such as capturing insights [53], externalizing thought process and in situ conditions [54],[55], and facilitating collaboration and communications [56],[57]. We recognized annotations as a valuable tool for recording data hunches, given their versatile nature. The techniques to record data hunches use textual and graphical annotations based on the existing works. We provide an overview of previous works on textual annotations and graphical annotations.

Textual annotations have been used in many visualization systems [58], [59] and in the media, including the *New York Times* [60]. Liu et al. [61] proposed an annotation system that uses a structured format strictly according to the domain context. Goyal et al. [62] offered more freedom to users by allowing them to use a notepad for free-form notes during their experiment. McCurdy et al. [11] adopted a structured form format for experts to annotate their knowledge of the data, shown in Figure 2.2. Beyond annotating thoughts and opinions about the visualization, annotations can also contextualize the data visualizations by adding information that can explain the data pattern [63] or augment the data representation [64]. Furthermore, Lai et al. proposed using a convoluted neural network to automatically apply textual annotations to visualizations based on users' description input [65]. Visualization work [56] has also bridged storytelling and collaboration, aiming to communicate the analysis process through annotations.

Graphical annotations incorporate graphical elements into annotations, allowing the users to draw and describe ideas with more detail unconstrained by the limitations of text. Kim et al. [54] studied data analysts' preference between textual annotations and sketching and found that they preferred the freedom and intuitive tools provided by graphical annotations. With the aim to facilitate communications in business, Marasoiu et al. [66] implemented a prototype that uses graphical markups to clarify hypotheses in data analysis. For more casual users' scenarios, Ren et al. [67] developed a toolkit to annotate and augment charts to facilitate storytelling with visualizations, shown in Figure 2.3. Heer et al. [14] provided graphical annotations for users to accompany their textual comments with graphical annotations on the visualization and inspired useful detailed annotations to illustrate opinions in the comments as well as some playful annotations from the participants. Additionally, Romat et al. [68] presented a graphical markup system with automation to facilitate sensemaking.

Existing research in annotations also suggests challenges that we may face when using annotations for data hunches. For example, Kim et al. [54] indicated that multiple annotations can lead to visual clutter. To improve visual clarity, they suggested graphical annotations need to be distinct so the user can differentiate the original visualization and various annotations, and using visual cues can be helpful to reassure users that the annotations are saved. Visual clarity will remain a priority for designers to consider when implementing methods to record data hunches. Other researchers found that even when annotations are present in the system, users are unwilling to utilize them. Boukhelifa et al. [33] discussed that annotations (on uncertainty) are less valuable because they are not used extensively in data analysts' processes. Similarly, Kandel et al. [69] found that since the workflow of data analysts has no natural point where they would stop and annotate, their participants were not actively annotating their analysis process. These findings suggest that when annotations are not actively part of the analytical process, they do not provide much value and do not encourage users to annotate their opinions. However, we suggest that as we incorporate data hunches as part of the analytical process, they become a valuable source of context and information and will promote willingness to annotate and record.

2.2.2 Notebooks

Lab notebooks and field records are often used by analysts when conducting experiments and collecting data. These documents provide details on the process for subsequent analyses and a better understanding of the condition, quality, and caveats about the data [70]. Many lab notebooks have been transitioned into digital versions [71], [72] with rich content such as visualizations [73] enabling more collaborative maintenance practices [74]. Even though maintaining lab notebooks is standard practice in science communities, physical notebooks can be lost or unavailable to all interested parties, and digital ones often lack flexibility [75].

Notes and records, in turn, often are transferred into methods sections in publications and reports, where readers can find details about the data and analysis steps, helping them judge the validity and reproducibility of the study [76], [77]. Bruce [78] surveyed methods sections across different disciplines and found different patterns; social science publications use 'slow' method sections with descriptive structures, whereas physical science employ 'fast' Method sections with more discourse-like structure. The difference suggest that methods sections can be less approachable for readers from different disciplines. Additionally, method sections are often space-limited. Details are omitted or favor describing the main results of the publication, even though detailed methods sections are the purview of scientific work.

Computational notebooks, such as Jupyter Notebooks [79] and R Markdown [80], [81] are often discussed as a remedy for some of the issues we discuss here: they can be used to describe datasets and analysis steps and may include visualizations and executable code that (in theory [82]) ensures reproducibility of analysis. Due to these advantages, significant research has been devoted to understanding how analysts use computational notebooks [83]–[85] and to improve them [86]–[88], such as emphasizing on reusability of computation notebooks [89].

2.2.3 Metadata

Metadata is another medium to communicate structure and information about data. Metadata, "data about data" [90], has been long used by professionals in the library, museum, and information science communities to describe, retrieve, and manage information in their digital workspace [91]. Metadata relieves the data user (human and computer systems) from having to know the full characteristics of the data in order to utilize them [92]. For example, a library system has metadata within the system, so that a reader can easily find a book of their interest (the data) based on indexing information the library provides (the metadata). A significant effort has been made to standardize metadata with formats like XML Schema, so that software applications can understand and retrieve the required information [93]. Various frameworks have been proposed to help manage metadata for different fields [94]–[97], as well as using visualizations to achieve a better understanding of data structure and relationships within [98], [99]. Burns et al. compared differences in data visualizations shown with and without metadata and demonstrated that metadata imbues more trust and persuasiveness of the visualization [100]. Metadata ensures the meaningfulness of the data and provides critical information about the data [101], [102].

Datasheet for data, proposed by Gebru et al. [49], emphasizes accompanying datasets with the necessary background, data characteristics, and recommended usage for the data, in order to promote publishing and reusing data ethically. They also suggested having a standardized questionnaire for datasets, with questions such as why the dataset was created, whom the dataset was funded by, and if the dataset will be updated. These are promising steps toward more informed and ethical analysis practices. In this dissertation, however, we pivot the idea of a datasheet for data, proposing a "sheet of hunches" to be accompanied with data. We argue that knowledge about the caveats of the data, though not originated from the ones who compiled the data, is as important as the context provided by the datasheets. Extending on reporting metadata for data context, several works in the machine learning community have proposed a more standardized approach to model reporting [103] and data reporting [104] to achieve transparency in research.

2.3 Current Analysts' Practices of Working with Data

An abundance of research from the visualization and HCI communities has provided insight into the current practices of data workers and analysts [105]–[108]. In a review of prior studies on data science workers, Crisan et al. synthesized the different processes performed by data workers: preparation, analysis, deployment, and communication [109]. They found that although visualizations touch all the described processes, their actual use is quite limited. In order to investigate opportunities to better align visualization tools to data workers' practices, our interviews in Chapter 5 covered all these stages of analysis, with a focus on preparation, analysis, and communication. In this subsection, we present an overview of previous interview studies on how analysts collaborate with data caveats, as well as present and record them.

Prior works have studied how data analysts engage with data in various settings. Kandel et al. [69] interviewed 35 anlaysts from various fields and studied their process and challenges in data analysis in enterprise settings. Their study participants reported using visualizations to identify outliers and artifacts in data and made assumptions in their analyses. However, their study participants did not report usage of experts' knowledge in data analysis pipeline, which we focus more in this dissertation. Liu et al. [110] investigated the analytical decisions and reporting in research settings and found that because of multiple interpretation and outcomes available from the analytical decisions, there was a preference to report desired results corresponding to study participants' theory. Rittenbruch et al. also discovered the need for more versatile visual analytic collaborative tools, because their interview participants in biomedical field shared that analysis results becomes too complex to communicate on formats like PowerPoints and emails [111]. Ruddle et al. [112]

shared the results from a survey and follow-up interviews on data profiling–determining the quality of datasets and characterizing data for analyses. While both Ruddle et al. and our work have found that analysts "gut check" data for quality assurance, our work focuses more on using experts' knowledge to compensate data issues and documenting and communicating experts' knowlege down the stream.

To better understand how analysts collaborate within their teams, Zhang et al. surveyed 183 data workers in machine learning and artificial intelligence and summarized their workflows in general and tools used in their workflows in particular [113]. The authors described the difference in patterns of communication exhibited by different roles within teams, such as that domain experts are more engaged at every stage of the analysis process, while managers were often absent during more technical stages of a project. In addition, the authors also discovered little use of documentation in collaborative analysis processes, especially during the feature-extraction and feature-engineering phase, leading to human decisions invisibly incorporated into the analysis process, jeopardizing subsequent re-analysis and revisions.

Jung et al. presented an in-depth study into how domain experts work with data [114] and found that they place more value on their data being actionable than the data having abstract qualities, such as high precision. They also discussed *conversations with the data*—procedures of working with data directly to better understand it [115]—as a critical part of the analysis process, similar to previous works [116], [117]. Chapter 5 further explores different ways in which analysts make sense of and communicate data that come with caveats.

Prior work has also explored the meaning of caveats and uncertainty to data workers and data professionals, specifically with an eye on how uncertainty affects their analysis. Skeels et al. conducted an interview study with professionals from various domains and classified the types of uncertainty that domain experts encounter [118]. They reported that analysts used qualitative labels to describe uncertainty, but the labels were rarely stored along with the data. Boukhelifa et al. reported on the strategies that domain professionals employ in order to deal with uncertainty: understand, minimize, exploit, and ignore [33]. Hullman has contributed to this research with more evidence as to why visualization practitioners actively choose to omit uncertainty in their visualizations, citing visualization authors' concerns about overwhelming their audience with too much information and uncertainty [22]. Current approaches heavily focus on presenting the visualization consumers (data experts, analysts, and the general public) with uncertainty already embedded in the visualization. However, since visualization designers are always informed by those with personal knowledge of the data, the externalization of the knowledge is often expressed after it might have been lost at some point in the process.

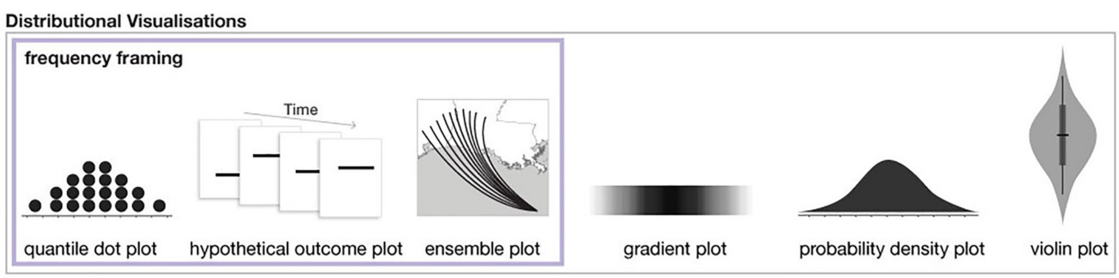


Figure 2.1: Examples of uncertainty visualization techniques; figure from Padilla et al. [17] © 2021 Padilla, Powell, Kay and Hullman. Reprinted under a CC-BY license.

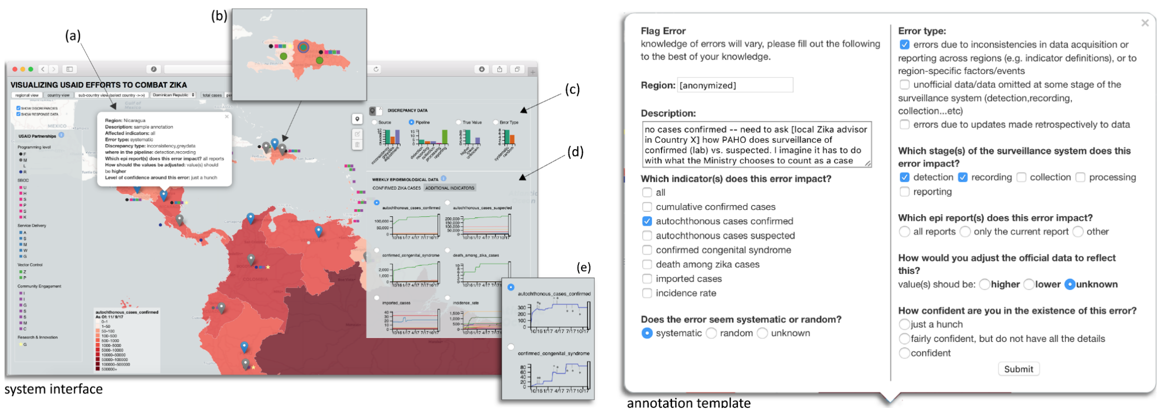


Figure 2.2: Prototype for externalizing implicit error on a interactive visualization application for Zika outbreak analysis, proposed by McCurdy et al. [11] © 2019 IEEE. Reprinted with permission.

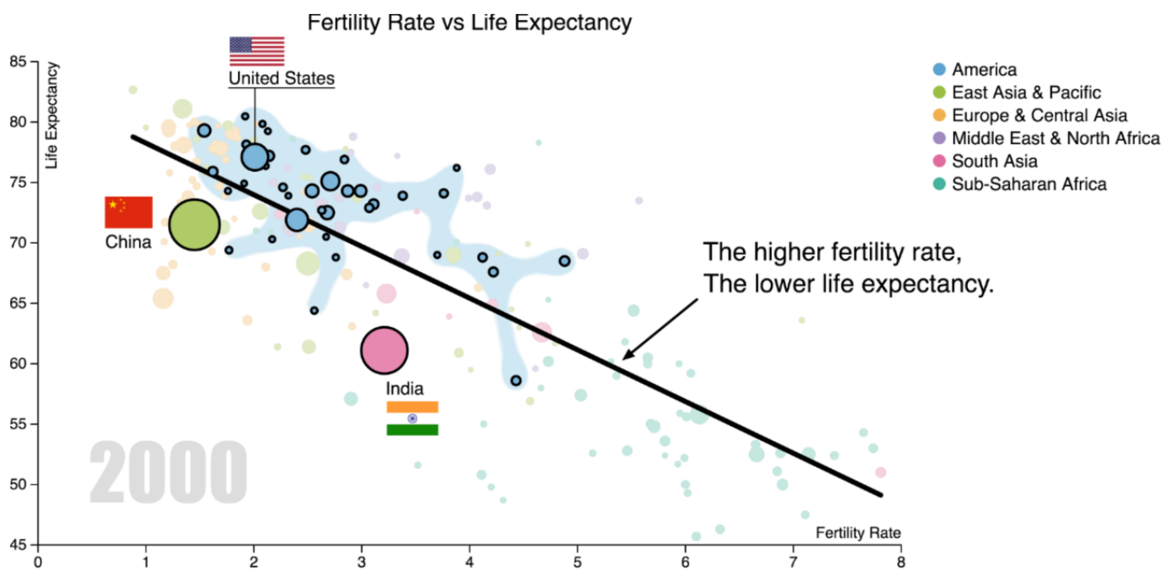


Figure 2.3: Annotated chart for storytelling; figure from Ren et al. [67] © 2017 IEEE. Reprinted with permission.

CHAPTER 3

DATA HUNCHES

In this chapter, we present our definition of data hunches with an in-depth discussion on the significance of recognizing personal knowledge in data analysis. We also provide our methodology for arriving at the definition and the design space for data hunches.

3.1 What Are Data Hunches?

In Chapter 2, we showed that our community's current framing of uncertainty implies that data workers and tool builders are responsible for identifying, characterizing, and quantifying sources of uncertainty from data. Data hunches acknowledge and incorporate experts who come to data analysis with deep knowledge about the limitations of their data. As a complementary perspective, data hunches focus on knowledge about data, capture that knowledge from a diverse set of stakeholders, and are embeddable in the analysis process. We argue that this perspective shift offers a breadth of new opportunities for recording and communicating data hunches in support of richer data analysis.

Through data hunches, we elevate the role that personal knowledge of the data plays in the process of understanding and analyzing it. To reiterate, we define *data hunches* as **an analyst's knowledge about how and why the data is an imperfect and partial representation of the phenomena of interest**. These hunches can range from the abstract—expressing concern about the validity of the data set—to the concrete—expressing a numerical value that is closer to the phenomena of interest than the measured data. The scope of a data hunch can be individual data points, a complete dataset, or anything in between. Data hunches emerge when an analyst interacts with the data, triggering reflection about the ways the data is imperfect and partial based upon their inherent knowledge about the data collection process, domain, and more. Analysts also reach out to experts to solicit data hunches for their projects. A data hunch can be based on the missing context necessary to fully comprehend the phenomenon, discrepancies between a mental model and data, opinions on the quality of the data generation process, and so on. Data hunches are knowledge about sources of qualitative uncertainty. During data analysis, data hunches influence an analyst's interpretation of the data, derived knowledge, and decisions made.

As we challenge the existing definition of uncertainty and define personal knowledge as data hunches, how does a data hunch relate to a dataset? The typical reason people collect data is to measure and record a phenomenon in reality. In practice, data is rarely a faithful and comprehensive representation of reality [119]. Since data is not a perfect depiction of reality, the usefulness of a particular dataset depends on how well it serves a task. In this dissertation, we consider data to be a tool, appropriate for some tasks but not as useful for others. By analogy, a screwdriver can be used to hammer a nail, but a hammer is more effective. As with any tool, knowledge about how and when to use the tool and knowledge about its limitations is essential.

Data hunches are the knowledge people have about the mismatch between reality and data. Whereas some data hunches may be useful independent of an application context, typically, data hunches are based on the intended usage of data. They do not exist independently nor are present with the data itself. To revisit the tool analogy, data hunches capture the knowledge about how to use the screwdriver effectively, and they help analysts use data, despite its imperfections, to serve their analysis needs.

We have shown that data hunches are prevalent, but often implicit, in data analysis, and as important as the data itself. The knowledge experts bring to data analysis is a vital component of data-driven decision-making [35], [120]; however, such knowledge is often recorded outside of visual analysis tools. This disconnect then requires mental gymnastics on the part of an analyst to incorporate back into the data analysis process, if it is not overlooked completely. In this chapter, we envision that a set of tools that fully utilize what visualization can offer will provide a more visual and intuitive representation of experts' knowledge in visualizations. Using data and data hunches in tandem supports a richer representation of a phenomenon, leading potentially to improved analysis. By acknowledging and naming data hunches, we aim to elevate the potential for personal knowledge to actively and explicitly contribute to data analysis.

We purposefully scope data hunches for use within collaborative, expert settings for both pragmatic and ethical reasons. Pragmatically, previous work on collaborative visualizations highlights the value of designing tools that support sharing of expert knowledge. In the context of collaborative analysis sessions, Mahyar et al. [121] showed the importance of recording visualizations and note taking in collaborative visual analysis. Similarly, Walny et al. [122] studied the use of data visualizations on whiteboards in corporate offices and found that visualizations as sketches promote team discussions. In another example of data science workflows that utilize computational notebooks, Wang et al. [84] found that data scientists annotated screenshots of visualizations when collaborating as a way to communicate limitations of a tool. Ethically, scoping data hunches to collaborative expert settings reduces potential harm as we expect experts collaborating with peers to record well-reasoned and nuanced hunches. However, experts can still be biased, and data hunches may help to reinforce biases even further. We believe that peer-review tools, such as comments or ratings, may be useful to mitigate the risk of reinforcing biases, yet acknowledge that more research on the topic is necessary. We discuss these and other issues further in Chapter 7.

By identifying data hunches as productive and insightful expert knowledge, we can re-interpret past work on collaborative visualization tools with this framing. For example, a visualization designer can incorporate commenting and discussion features to promote externalization of data hunches [11], [14]; apply provenance tracking to record actions they took based on hunches to wrangle the data [89], [123]–[125]; and use visualization techniques like linked views and visualization states to show a collection of data hunches [56], [58], [126]. We see a wealth of opportunities for incorporating data hunches into old and new ways of visually analyzing data.

3.2 Methodology

Our methodology for theorizing about data hunches and developing a framework for recording and communicating data hunches was based on reflective practices [127], [128]. We began by reflecting on our experiences working with a variety of domain experts who have rich knowledge about their data, knowledge that was not captured in their datasets. Through group discussions about our experiences, we recognized the missing formalization of personal knowledge and its impact on data analysis. We began mapping out the scope of data hunches, the relationship between data hunches and existing visualization concepts, and how hunches have been reported in the existing literature. This process included an interview study with domain analysts on their experience of utilizing domain knowledge, a literature search into data feminism, critical data studies, and uncertainty, as well as searching works on design studies and reviewing any reported sources of qualitative uncertainty in previous design studies. We will present more details on the methods for the design study in Section 4.4 and for the interview study in Section 5.1.

After investigating the landscape of data hunches, we iteratively developed our understanding of data hunches. The iterations critically reflected illustrative examples from our prior experiences and recollections from our interview participants. We additionally received feedback from our research lab and colleagues and made adjustments accordingly. We used the design space to re-imagine visualization systems presented in several design studies [1], [11], [14].

Initially, we considered our framework for recording data hunches as a medium for collecting input and knowledge about a dataset from a general audience. Our reasoning was that the crowd might have insights about datasets based on their own experiences, such as collective and local knowledge about a COVID-19 dataset. We came to appreciate that a key challenge of data hunches, however, is that they could be used to explain away inconvenient data points, or that they could exacerbate the problem of confirmation bias [129], [130]. Our interview results reinforced our concerns about individuals with personal agendas using data hunches for personal gains. We thus made the decision to argue for scoping data hunches to collaborative settings with groups of experts who are supported by networks of trust [131]. We discuss the potential benefits and harms that could be associated when data hunches are implemented for general audience systems in Chapter 7.

CHAPTER 4

FORMATIVE DESIGN STUDY – SANGUINE: VISUAL ANALYSIS FOR PATIENT BLOOD MANAGEMENT

In this chapter, we describe the formative work that inspired the core contribution of this dissertation. Sanguine [1] is an interactive visualization tool that helps physicians and experts analyze transfusion data with meaningful context. This work inspired us to investigate how data hunches exist in various projects and studies and how these hunches affect our engagement with the data. All figures mentioned in this chapter are placed at the end of the chapter.

4.1 Formative Aspect

The core of this design study was the design and implementation of Sanguine. We followed the design study methodology [132] in developing Sanguine, and because its main users are patient blood management experts, a secondary user group—clinicians—were not closely involved in the design and development of Sanguine beyond participating in the creative visualization-opportunities (CVO) workshop [133]. During the evaluation and reflection stage of the design study, we presented the interface to several clinicians to solicit their feedback on the interface. However, they reported that they did not think that some of the data shown in the visualizations correctly represented their practice. For example, the surgeons reflected how they used cell salvage (the recycling of a patient's own blood instead of using blood products during sugery) in almost all cases, yet our data visualizations show usage in only a fraction of cases. They suspected that the usage was not recorded in the database when the volume was low.

This work is formative to this dissertation in that it was an example of the data not perfectly matching the knowledge that experts have, making it a partial and imperfect depiction of reality. Existing research lacks recognition and the potential to utilize personal knowledge in analysis. The experience with domain experts and the lack of existing research inspired us to investigate further experts' knowledge in general data analysis workflows. We expanded this research goal into an interview study with analysts presented in Chapter 5 on how they use experts' knowledge to bridge the gap between data and reality. We define experts' knowledge of data imperfection as data hunches in Chapter 3 and present techniques and guidelines to support data hunches in interactive visualization applications in Chapter 6.

4.2 Motivation and Overview

Transfusion is the most commonly performed medical procedure during hospitalizations in the United States [134], with 17.2 million blood products transfused in 2015 [135]. Although transfusions are often medically necessary and even life-saving, they have also been identified as one of the most overused procedures in hospitals [136], and unnecessary transfusions can lead to adverse reactions and complications [137].

Management and acquisition of blood products are also expensive, creating financial incentives for hospitals to optimize their usage. To better address these problems, many hospitals have hired dedicated experts in patient blood management (PBM), and PBM has become an independent academic field. Through PBM, healthcare providers aim to reduce unnecessary transfusions and improve patients' outcomes using various methods, such as treating anemia (lack of red blood cells) prior to surgeries, providing guidelines on when transfusions are necessary, and using cell salvage ("recycling" blood) during surgery [138]. Successful PBM involves multiple stakeholders. PBM experts analyze the usage of blood products and give advice or develop guidelines on best practices. Clinicians, in turn, make decisions about when and how much to transfuse; thus, they need to be well informed about best practices, their individual performance regarding these best practices, and their performance relative to their peers. These analysis processes, however, are difficult with the existing tools available in hospitals. Electronic health records applications may provide static charts and reports on blood product usage, but they lack the nuance and sophistication necessary to make a holistic evaluation of the practice of individual clinicians. For example, when a surgeon uses significantly more blood products than their peers, they might rightly claim that they do so because they are treating the most difficult cases with many patient comorbidities. With current tools, analysis is often static or limited to only certain clinical contexts.

To tackle this issue, we created Sanguine, shown in Figure 4.1, an interactive visual analytics tool for PBM analysis. Sanguine was developed as a design study using an iterative design process involving PBM experts, surgeons, and anesthesiologists at the University of Utah Hospital. Sanguine enables analysts to view outcomes juxtaposed with blood product usage data, review patient blood management practices for different providers, and compare the effect of patient blood management in clinical settings. Sanguine is designed with two types of stakeholders in mind: PBM experts who conduct in-depth analysis and clinicians (primarily surgeons and anesthesiologists) who receive tailored and annotated interactive reports from the PBM experts. These two groups have radically different types of engagement with Sanguine. The PBM experts independently analyze data and ask a variety of analysis questions, creating charts in the process. The clinicians, in contrast, mostly consume the visualizations as reports that contain feedback on their practice. We designed Sanguine to facilitate communication around patient blood management between the two types of stakeholders.

The main contribution of this chapter is the design and development of Sanguine, an open-source visual analytics tool for the analysis of patient blood management data. We also contribute an analysis of domain goals and task abstraction. We evaluate the utility of Sanguine in case studies and through interviews with domain experts.

4.3 Related Work—Medical Visualization

Visualizing medical records has been an important area of research for analyzing medical data [139]. Caban et el. [140] described four major types of visual analytic applications in healthcare: clinicians analyzing patients' records, administrators making data-supported decisions, researchers working on large medical datasets, and patients understanding their own data. Our work falls into the first category. Shneiderman et al. [141] analyzed the challenges of implementing interactive visualizations for healthcare professionals, which include offering busy clinicians timely information in the right format. West et al. [142] encouraged creators of visualization systems to also consider the training time required. In Sanguine, we address these points by enabling a workflow where a PBM professional curates an annotated visual report for the clinicians.

In many clinical contexts, a timeline of symptoms and treatments is informative, and an abundance of work has focused on visualizing medical records in a timeline format [143]–[146]. We found that existing works focus on visualizing an individual patient or a group of patients on a timeline, and emphasize specific events happening in the sequence. Whereas event sequences can play a role in transfusions, Sanguine takes a provider- and practice-focused approach instead.

Cohort definition and creation is a common theme in medical visualization works, and previous works have used methods such as temporal visualization querying [147] and attributes-driven filtering [148], [149] to create cohorts. To compare patient cohorts, previous works adopted approaches such as set visualizations, sunburst graphs [150], and compact static dashboards [151]. Sanguine, instead, focuses on cohorts that relate to the practice of the providers: cohorts are either created temporally to make year-to-year comparisons, or by providers, to compare the performance of surgeons and anesthesiologists.

Another research topic on visualizing health records is bridging the communication gap between providers and patients [152]. As an example, Hakone et al. [153] designed an interface that helps patients understand the complex risk of cancer and the treatment plans available to them through visualizations. We are not aware of existing techniques explicitly designed for provider-to-provider communication, which is a key aspect of Sanguine.

Finally, various projects have applied statistical analysis to transfusion data. Gálvez et al. [154] developed a tool for decision-making support on blood product ordering at a pediatric hospital, which gives patient-specific blood order advice. This tool allows users to filter the data based on age groups, procedures, and conditions. After filtering the data, the tool visualizes all the transfusion data and other attributes of interest in previous procedures using bar charts. This approach offers a concrete solution to a patient-specific blood order schedule and shows how visualization can be effective for visualizing transfusion in a clinical setting. Although the tool shows all blood usage of cases from filtered results, Sanguine offers more personalized blood ordering preparation for providers, where they can view their history of transfusions on a particular procedure with specific conditions.

4.4 Methods

Our main collaborator, Dr. Ryan Metcalf, is the medical director of the Transfusion Medicine Service at the University of Utah Hospital and the PBM expert user for the end result tool. Dr. Metcalf was closely involved throughout the design study as the "domain expert" in the design study methodology [132]. The core development iterations took place over 12 months, and the publication was based on the result of the core development. After the publication, we followed up with additional maintenance support and additional feature implementations.

To understand the needs of a broader set of stakeholders involved in PBM, we conducted a creative visualization-opportunities (CVO) workshop [133] in July of 2019. We recruited participants from Dr. Metcalf's network at the Division of Cardiothoracic Surgery. Participants included two cardiac anesthesiologists, one cardiac surgery critical care physician, and one IT manager — all employees of the University of Utah Hospital. To accommodate the schedule of the clinicians, the workshop lasted three hours, which is shorter than a typical CVO workshop. After a brief introduction, we used the *Wishful Thinking*, *Visualization Analogies*, and *Barrier Identification and Removal* activities [133]. After the workshop, we did an initial analysis of the results we collected. One key insight that emerged was that clinicians believed that a tailored report to their practice would be much more useful and likely be more widely adopted than a general-purpose tool. The key insight, in turn, allowed us to design for these two user groups, PBM experts and clinicians. To also understand whether the results also match up with cardiac surgeons' workflows, we decided to present our intermediate results and elicit feedback from a large group of cardiac surgeons during the department's weekly all-hands meeting. The surgeons stated that a tool that can aid them in their blood management practice could lead to improved outcomes but cautioned that they would not want a system that would obstruct them from ordering blood or slow down their current practice.

The workshop and notes from the all-hands meeting generated rich artifacts in the form of thematically grouped sticky notes, audio recordings, photographs, and notes by the facilitators. To analyze all the outcomes from the workshop and the meeting, we first created a machine-readable corpus (transcribing sticky notes, etc.) and then grouped the concepts into thematic areas. We then based our domain goals, and subsequently, our design on these results.

4.5 Domain Goals and Data

Identifying problematic transfusion practices, and thereby, improving outcomes, is the primary goal that all our stakeholders share. Meeting this goal requires the analysis of transfusion data in the context of patient records and evidence-based guidelines from the literature. Comparisons of various aspects (between clinicians, between time intervals, etc.) also emerged as a way to achieve the goal. Even though real-time decision support for transfusion was brought up during the workshop, the technical and regulatory difficulties made it hard to achieve within the scope of this design study.

We also discovered barriers to sharing PBM experts' insights effectively with surgeons. Surgeons may not always have the time to analyze data closely. At the same time, they would like for analyses to account for variations in patient characteristics and other specific contexts to meaningfully capture the complexity of a particular case or group of cases. Because a well-designed visualization can communicate insights from the data with ease, we decided to design and implement a visual analysis tool for analyzing and reviewing patient blood management practice with outcomes of interest.

4.5.1 Domain Goals

We break down the high-level goal of identifying problematic transfusion practices into domain goals as follows:

- *G1: Comparing Transfusion Practice:* We identified comparisons of transfusion practice, accounting for the complexity and heterogeneity of surgical cases and other risk factors, such as the general health of the patients, as a central aspect of the analysis question. Our interviews and workshop revealed that providing this contextual information is critical for practitioners to trust the visualization.
- *G2: Analyzing Adherence to Best Practices and Standards:* The fields of medicine have developed a suite of evidence-based clinical guidelines and best practices for many procedures for patient blood management, such as usage of blood recycling and treating bleeding pharmaceutically instead of transfusions. Both patient blood management experts and clinicians need to analyze whether these standards are met.

- *G3: Analyzing Individual Patients:* In certain situations, aggregate or summary information is inadequate to make a holistic evaluation of a case; therefore, clinicians need to access individual patient records.
- *G4: Prepare for Surgery Using a "Patients Like Mine" Approach:* Ordering an appropriate amount of blood is important for both outcomes and the efficient use of resources. PBM professionals want to make good estimates of the amount of blood to prepare by exploring similar historical cases to see the amount and variability of blood needed, and then order blood products based on the exploration.
- *G5: Communication and Sharing:* Our visual analysis interface needs to be tailored to both the "power users" and the "consumers" of the visualizations: PBM professionals who want to interactively explore the data, create custom views, and create cohorts based on multiple attributes; clinicians who want to see their performance relative to their peers and to analyze whether they follow evidence-based standards.

These domain-specific goals map to a wide variety of visualization tasks. With Brehmer and Munzner's task typology [155], G1 to G3 map to the "discover" and "lookup" tasks on the "why" dimension, and the "encode," "select," "filter," and "aggregate" tasks on the "how" dimension. G1 is a comparison task, for which Gleicher has developed a set of considerations [156]. The characteristics of the data subsets to be compared (a handful of different items, such as the records of surgeons, and multiple complex distributions as data values) suggest "juxtaposition" as an appropriate design choice for the comparison task. G4—looking for "patients-like-mine" is, again using Brehmer and Munzner's typology, of type "browse" in the "search" category, as the characteristics are known, but the target is not. G5—communication and sharing—is different from the other tasks, as it is mostly about "producing" (why) based on "annotations" and "recordings" (how). The heterogeneity of the goals and tasks suggests that a flexible, multiview visualization will be necessary to address all of them.

4.5.2 Data

In cooperation with the data warehouse at the University of Utah Hospital, we compiled a dataset drawing from the electronic health records of over 4000 cardiac surgery patients, spanning the years 2014 to 2020, and covering 111 different cardiac surgery procedures, ranging from routine coronary bypass surgeries to heart transplants. The primary data is the number of units transfused, cell salvage (an alternative to transfusions, measured in milliliters), and laboratory measurements of hemoglobin level, which are the main indicators for anemia management and transfusion appropriateness. Depending on the research question, other data can provide relevant context: patient outcomes (death, need for long-term ventilation, etc.), PBM-related drug administration, and information about whether a surgery was planned or was an emergency are examples. These attributes are extracted directly from electronic health records and anesthesia flowsheets.

4.6 Sanguine

Sanguine consists of three main components, shown in Figure 4.1: a filtering and selection view, the main workspace where visualizations can be flexibly arranged, and a patient-specific detail view. A video demonstrating Sanguine is available at https://youtu.be/DhTNyvCJgtM. The source code of Sanguine is available at https://github.com/visdesignlab/Sanguine.

4.6.1 Comparing and Contextualizing

Sanguine uses heatmap as the main chart type in Sanguine to enable comparison using the juxtaposition strategy [156] (shown in Figure 4.1). Addressing the practice-focused comparison (*G1*), the heatmap shows transfusion data aggregated by surgeons, anesthesiologists, or years. To account for the difference in total case count between these facets, we use relative scales showing the percentage of cases that had 0–4 units of red blood cells or 5+ transfusions. The threshold varies depending on the blood component. The decision to aggregate values above a certain threshold into one bin was driven by the effect that rare outliers had on the overall visualization. For cell salvage data, which are measured in continuous values, we used bins with a dedicated bin for no usage.

As is evident from the heat map in Figure 4.1, a large percentage of cases received no transfusion. This aspect of the data is important to include in the visualization, but the heavily skewed distribution makes it hard to notice any difference in the nonzero transfusion cases. To address this problem, we provide a toggle to remove zero transfusion from the color scale in the visualization, i.e., the maximum value that maps to the darkest red color is taken from the counts of cases that have received at least one transfusion. Only mapping non-zero transfusion units to the red color scale makes differences in transfusion practice much more apparent, as shown in Figure 4.2. The heatmap still shows the data on zero transfusion, but we use a separate gray color scale encoding the percentage of zero transfused cases out of all cases of the row.

The resulting heatmap enables PBM experts and clinicians to compare their relative practice of transfusion (*G1*). Although Sanguine can be used to compare transfusions across all procedures, filtering by procedures is essential, because diverse procedures, such as heart transplants and bypass surgery, differ in the typical need for transfusions. However, even when narrowing cases down to specific procedures, there can be systematic differences in how complicated cases are, and hence what outcomes can be expected and how much will be transfused. To provide this context and show transfusion-relevant outcomes, Sanguine can visualize additional attributes to display along with the heatmap, such as lab values and patient outcomes (see Figure 4.2). The visual encoding used depends on the attribute types. For numerical/distribution values (e.g., hemoglobin values), we use a violin plot that morphs into a dot plot when there are few observations; for individual numerical values, we use bar charts (e.g., average transfusions per case) or labeled heatmap cells (e.g., mortality rates).

Finally, to enable outcome and intervention-focused comparisons, the heatmap in Sanguine can be divided by binary outcome variables such as mortality or by a time range. The heatmap shown in Figure 4.2 is divided by the need for long-term ventilation, as indicated by the green and blue fields, respectively. Every row is divided, showing the transfusion data and contextual information in separate rows. In Figure 4.2, for example, we can see that long-term ventilation appears to correlate with higher transfusion rates and higher risk scores.

4.6.2 Standards and Best Practices

Hemoglobin laboratory values are a key indicator in patient blood management. Clinical trials have led to the development of evidence-based guidelines that use hemoglobin values as red blood cell transfusion thresholds, and corresponding expectations for post-transfusion hemoglobin targets can thus be inferred. Hence, analyzing transfusion practice in this context is essential (*G2*). A low preoperative hemoglobin level indicates improper anemia management before surgery, whereas high postoperative hemoglobin levels imply excessive transfusion during surgery. To facilitate the analysis, Sanguine provides a dumbbell chart dedicated to hemoglobin level evaluation (see Figure 4.3). An individual dumbbell visualizes the preoperative (green) and the postoperative (blue) hemoglobin values of a single case as dots. A link connecting the dots shows the gap between these values. Cases and their corresponding dumbbells can be divided by all relevant attributes; Figure 4.3 shows dividing by surgeons. Within each division, solid horizontal lines show the medians for pre- and postoperative hemoglobin levels, respectively, and stippled lines provide clinically recommended values for pre- and postoperative hemoglobin levels. The dumbbells within each division can be sorted based on preoperative value, postoperative value, or the gap between the two.

4.6.3 Filters and Brushes

In addition to dynamic comparisons, filters and brushes are essential tools to create the visualization relevant to answer specific analysis questions. In addition to the procedure filter view (Figure 4.1), Sanguine provides views specifically designed for filtering based on data values. The dot plot (Figure 4.1) can visualize correlations and support rectangle brushes, which can be converted into filters in the filter manager view. The dot plot also visualizes mean values and confidence intervals. Additionally, we integrate a LineUp [157],[158] view to visualize all attributes of individual cases (*G3*) and complement Sanguine's filter system with the one that comes with LineUp. LineUp visualizes data in a tabular layout and applies different techniques to columns based on the attribute types.

4.6.4 Case View

An alternative way to view information about individual cases (*G3*) is the detail view, shown in Figure 4.1 on the right. All selected cases are shown in a list; the attributes for one case are shown, including transfusion records, surgery descriptions, relevant medicines administered, and outcomes. Using the detail view, analysts can study these cases closely, and practitioners can cross-reference cases with the medical record system for more information on the patients.

4.6.5 Communication and Sharing

Sanguine enables communication and sharing (*G5*) in two complementary ways. First, communication is facilitated through annotations: each visualization is accompanied by a text field, which can be used for notes or to record insights and conclusions. Second, sharing of findings is enabled through provenance tracking. Each change to the visualization, including visualization configurations, filter settings, and annotations, is saved as a state and recorded using the Trrack provenance tracking library [124]. Based on Trrack's functionality, Sanguine provides undo/redo, saving and loading the state of a workspace to/from a server, and sharing the state via a URL. URL-based sharing is convenient for distributing findings made by the PBM expert to clinicians, who can then review the visualizations and adjust their practice, if appropriate. As these "visualization consumers" typically do not want to leverage the full complexity of Sanguine, we introduced a "View Mode," which removes editing functionality and simplifies the interface.

4.6.6 Implementation and Deployment

Sanguine is developed in Typescript using the React framework. We use Semantic UI and D3 for the graphical interface and the MobX and Trrack library [124] for state control. The front-end is supported by a Django server. The server interfaces with an SQL database housed in the data warehouse of the University of Utah Hospital. It is deployed in a protected environment suitable for sensitive medical data and uses encryption and two-factor authentication to ensure security.

4.7 Evaluation

We evaluate Sanguine through case studies with Dr. Metcalf, the PBM expert, and through feedback from surgeons, thereby covering our two user groups of analysts who interact with the tool in depth and clinicians who consume curated reports. As our collaborators were closely involved in the development of Sanguine, experimental demand characteristics, which often arise in close collaborations [159], can bias responses; hence, we refrain from reporting subjective assessments and instead report factually on analysis scenarios and make comparisons of capabilities of a workflow with and without Sanguine.

4.7.1 Case Study

To demonstrate the utility of Sanguine, we present case studies of two scenarios where a PBM expert, Dr. Metcalf, using Sanguine to analyzing transfusion practices in this subsection. The main research question is if Sanguine can provide insights into transfusion data and can be used to answer the domain goals. Our collaborator had the opportunity to use the tool in practice for several months after we completed a prototype before reporting use cases.

4.7.1.1 PBM Review and Analysis

The first use case focuses on reviewing PBM practices among providers and against established medical guidelines using visualizations available in Sanguine.

• **Peer Group Benchmarking:** Context matters when comparing the transfusion practices of surgeons and anesthesiologists to their peers (*G1*). For example, if one surgeon transfuses more blood on average than others in the peer group, the surgeon may rightly complain that the comparison is not fair if they tend to see sicker or more complex patients. Previous work [160] has shown that diagnosis-related group (DRG) billing code weights are associated with transfusion volumes in surgical populations. In other words, more complex patients will have a higher DRG weight and would be expected to receive more blood on average.

To analyze differences in transfusion practice, the PBM expert created a heat map to visualize the blood utilization of the different surgeons, shown in Figure 4.4. He filtered to include only the most common complex cardiac surgery procedures. Blood use varies substantially, so to see whether the differences were due to the complexity of the cases and allow for fairer comparisons, he chose to visualize the distribution of DRG weights next to the heatmap and the bar chart of average transfusion per case. By adding the DRG weight distribution, it is easy to see which surgeons represent a reasonable peer group comparison and which surgeons do not. In the example shown in Figure 4.4, the two surgeons (indicated as (a) and (b)) seeing the most complex cases did not have the highest blood utilization, whereas Surgeon (c) mostly saw cases with low DRG weight but still frequently transfused their patients. The expert commented that this identifies surgeons performing above and below the context-appropriate peer group, which in turn can be used as a starting point for a discussion and intervention to improve practice.

- **Anemia Management:** Preoperative anemia management (*G2*) is an essential part of PBM. Dr. Metcalf started his analysis by studying whether anemic patients (preoperative hemoglobin level $<$ 10g/dl) were treated for it before surgery, as anemia can require additional transfusions that would otherwise not be necessary. To view the overall hemoglobin trend, our collaborator first created a dumbbell chart comparing surgeons. Since preoperative anemia management can be done only for elective surgeries, he then used the surgery urgency filter in the filter manager to remove emergent and urgent cases. He also selected a common procedure, coronary artery bypass grafting (CABG), and analyzed a subset of surgeons who have performed CABG. By comparing preoperative hemoglobin values to the stippled line for 13g/dL as the clinically recommended preoperative hemoglobin values, as shown in Figure 4.3, he noticed that several cases should have been treated for anemia before surgery, but were not. He also noticed that there appeared to be some variation in patient preoperative hemoglobin levels between surgeons.

- **Transfusion Appropriateness:** Using another dumbbell chart, Dr. Metcalf also reviewed transfusion appropriateness (*G2*). When a patient receives a red blood cell transfusion, the provider's target postoperative hemoglobin value is between 7g/dL and 9g/dL. To view only patients who received red blood cell transfusions, our collaborator added a dot plot of postoperative hemoglobin and units of red blood cells transfused intraoperatively. Appropriate transfusion is a practice applicable to all cases, elective or urgent; hence, he removed the elective case filter, and sorted the dumbbell chart to postoperative hemoglobin levels for easier comparison of the postoperative values. Using the brush feature of the dot plot, our collaborator could filter out patients who did not receive red blood cell transfusions during the surgery, resulting in the dumbbell chart shown in Figure 4.5. After applying the filter, our collaborator observed a significant amount of cases for which postoperative hemoglobin values were much higher than the target value, indicating over-transfusion in these cases. He then used this information to advise surgeons and anesthesiologists for better hemoglobin level targeting when performing transfusions.

- **Cell Salvage:** Another key point of patient blood management is the use of cell salvage. Cell salvage is recycling a patient's blood when they bleed during the surgery, an alternative to transfusions. Ideally, all surgeons should use cell salvage when performing transfusions (*G2*). Dr. Metcalf started with a heatmap of cell salvage, which he aggregated by surgeons. The heatmap shows that most surgeons used cell salvage, but the chart does not indicate how surgeons were using cell salvage when they were also transfusing.

After removing all cases without transfusions, our collaborator observed that even though cell salvage was used in most cases, there was still room for improvement, as shown in Figure 4.6. Most providers were using cell salvage properly; however, a few providers were performing transfusions without using cell salvage for over 20% of their cases. This information is valuable to PBM experts, who can now identify the providers and inform them about properly using cell salvage.

• **Drug Treatment:** Finally, Dr. Metcalf analyzed the usage of antifibrinolytic agents, in this case, aminocaproic acid. Antifibrinolytic agents help reduce bleeding in surgeries and, overall, the use of transfusion (*G2*). To visualize the use of aminocaproic acid, he added a comparison heatmap of red blood cells transfused, aggregated by surgeons, divided by aminocaproic acid usage, shown in Figure 4.7. The chart gave him an idea of how frequently providers were administering aminocaproic acid and how effective it was at reducing transfusions. By analyzing the heatmap and comparing the case counts, he concluded that the majority of cases did appropriately receive aminocaproic acid, which is indicated for patient blood management in cardiac surgeries.

Our collaborator identified a trend of fewer transfusions associated with the use of aminocaproic acid across individual surgeons To investigate the use of aminocaproic acid in more context, he added plots of the distribution of the DRG weights (risk scores). He noticed that despite similar DRG weights, cases using aminocaproic acid used less blood. This is consistent with what is known about the benefit of this drug in cardiac surgery (reduced need for transfusions) and further drives home the importance of this patient blood management modality.

Dr. Metcalf remarked that none of these analyses were possible before, at least not with the flexibility that Sanguine enables.

4.7.1.2 "Patients Like Mine" Decision Support

Dr. Metcalf explored using Sanguine as a decision support tool for blood component ordering when preparing for surgeries (*G4*). For example, a surgeon preparing for an open mitral valve replacement (MVR) can view the transfusion data for prior open MVR cases, using Sanguine's heatmap shown in Figure 4.8. To get a more specific picture, they can apply a filter using the dot plot based on their patient's preoperative hemoglobin value, and even filter based on the anesthesiologist with whom they will work in the surgery. The surgeon can view their own historic transfusion record for open MVR and order blood components based on their personal records. For example, a surgeon who performed 11 open MVR cases (first row in Figure 4.8) did not transfuse in 64% of cases, and used only 1 unit in 15% of all their cases. Hence, they can conclude that they should order one unit of red blood cells before surgery. The surgeon can also view the outcomes of cases from historic data, and know what they can expect for their upcoming case. For example, for the same surgeon in the first row, they can see that 55% of the cases required long-term ventilation after the surgery.

4.7.2 Feedback from Clinicians

Dr. Metcalf, the PBM expert, presented his findings, described in the previous section, to a cardiac surgeon and several anesthesiologists in virtual meetings, using a screen-share of Sanguine in "View Mode" (*G5*). As we designed Sanguine to be used by clinicians as visualization consumers, we considered the modality of the PBM expert presenting via screen-share and advising on blood product usage representative of a real-world use case.

After the presentation, the surgeon said, "This is what I have been looking into and could not find an answer anywhere. Besides just blood transfusion, there is a lot of other data in there too." The anesthesiologists discussed opportunities Sanguine could offer for implementing interventions, "[Sanguine] can guide our potential interventions...this is great to look at combinations of data...and for evaluating practices. I would use this tool every day."

We asked them to compare Sanguine with their current practice for analyzing PBM data, which consists mostly of reviewing the semiannual report provided by the Society of Thoracic Surgeons. The surgeon commented on the difference, "the STS reports we currently receive are about six months delayed, and you can update [Sanguine] much faster... This can be used for meetings to see where we can improve."

Speaking to a weakness, the surgeon raised a concern regarding the integrity of some of the data. First, they questioned the validity of the cell salvage values, because sometimes the usage is not properly captured in the anesthesia flowsheets, from which Sanguine derives the data. They also were surprised by and doubted the accuracy of the initial database results showing minimal usage of aminocaproic acid, which they expected everyone to use on every case for cardiac surgery. The mismatch between the visualization and their experience prompted us to go back to the database and identify an error in how the aminocaproic acid data was extracted and populated. This initial error was present in the data that Sanguine is built on, but was subsequently resolved after our investigation that spanned several month with the data warehouse. The observation that the data that is present in a database may not be a perfect reflection of reality, and that domain experts often have deep knowledge about such issues [11] spawned a line of work that this dissertation will present in Chapter 5 and Chapter 6, as we investigate what knowledge becomes critical in data analysis, and how such knowledge can be externalized and made explicit in the visualization so that it can be shared and reviewed.

4.8 Discussion and Limitations

Overall, Dr. Metcalf remarked that using Sanguine is a much more personalized and precise way for surgeons to prepare, compared to the guidelines currently in place. He commented that in this way, Sanguine can help reduce waste, while at the same time also reducing the need for emergency blood releases during the surgery. The COVID-19 pandemic has led to unprecedented blood shortages at hospitals. Each morning, our collaborator has been using Sanguine's "patient like mine" support to predict blood utilization for the most complex surgeries for the day. Using Sanguine, in turn, has enhanced preparation for individuals surgeries and to maintain the hospital blood bank inventory to meet critical patient needs.

4.8.1 Comparison to Prior Practice

Prior to the implementation of Sanguine, the most commonly used information about clinical practice quality at our collaborator's institution was reports from the Society of Thoracic Surgeons (STS). The STS database includes quality reporting data and has many participating sites across the United States. However, these data are static and not focused on patient blood management (PBM), and there is typically a time lag of several months to receive reports. Other institutions have published their standard PBM performance visualizations in the clinical literature [161]. These are useful for evaluating broad trends, but they lack detailed context and are not flexible or interactive like Sanguine. In contrast, Sanguine's dynamic and interactive nature allows the user to interact with the data and create highly contextualized visualizations. The filtering and context setting adds meaning because the clinical circumstances are important for understanding what an outcome means or why it may have occurred. However, a benefit of the STS reports over Sanguine is that the reports can also be used to make comparisons between different clinics. Hence, we believe that Sanguine and STS reports complement each other well: the former is useful for internal analysis and benchmarking, and the latter can be used to compare the overall performance of a site nationally.

4.8.2 Limitations of Evaluations

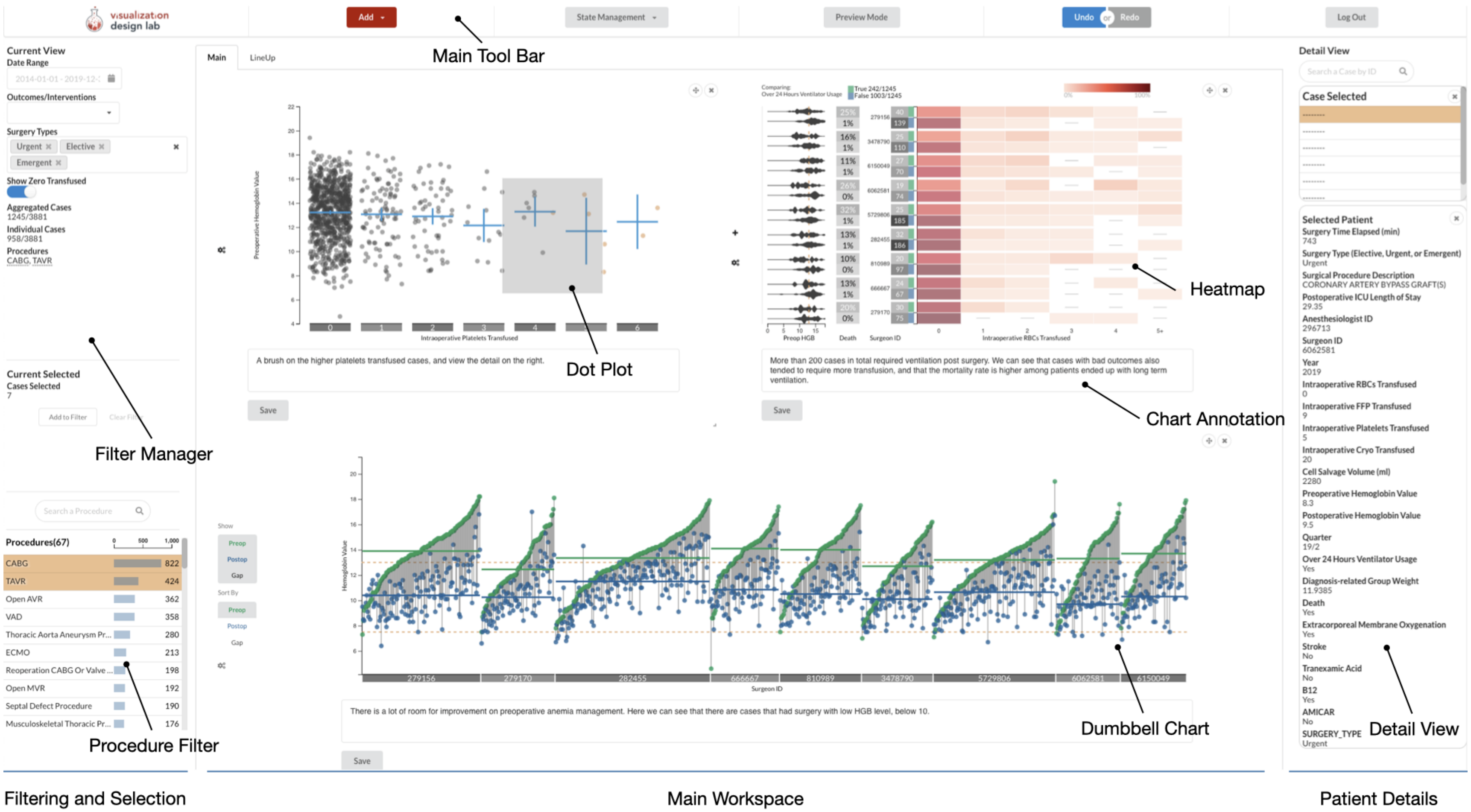
We demonstrate the utility of Sanguine through case studies and through feedback from clinicians and PBM experts. Case studies with real users, tasks, and data sets are the most common evaluation approach in design study research [132], and demonstrate rich, practical and impactful insights on the domain. Even though case studies tend to not significantly contribute to an understanding of the research contribution, they are considered necessary to ensure validity and trust [162]. Nevertheless, our validation is limited to our primary collaborator and a small group of clinicians. Alternative evaluation strategies, such as a comparative, quantitative evaluation, are difficult to achieve with specialized tools due to the lack of alternative systems that address the same task, and the small group of expert users that could meaningfully participate. We consider long-term observations on usage and adoption as the gold standard for evaluating design study tools. Sanguine has now been used in regular practice for about a year by Dr. Metcalf. However, adoption by outside, unaffiliated parties takes time, as striking agreements, getting IRB approval, and compiling the relevant databases in a different hospital is a complicated process. While we are currently discussing adoption in two unaffiliated medical centers, reporting on adoption in these centers is beyond the scope of this chapter.

4.9 Conclusion

In this chapter, we introduced Sanguine, a tool for analyzing patient blood management data. Sanguine is currently deployed and our collaborator has adopted it in his regular workflow. Going forward, we are in conversation to evaluate the long-term use of Sanguine and also study the influence it has on the decision-making process with independent experts at different institutions. We have taken first steps toward a dedicated, simplified interface for clinicians with the "View Mode," and we plan on ensuring compatibility of this mode with mobile interfaces, such as tablets and phones and explore automatic emailing feature to improve communication with surgeons with work-scale visualizations embedded [163]. Even though we focus on PBM data in this design study, the techniques we used can likely be applied to other areas of medicine, where analysts want to quickly view outcomes of interest and their relationships with variables of interest.

As mentioned in Section 4.7.2, the integrity of some of the data was questioned when we presented Sanguine to clinicians, which led to the identification of a problem in how the database was populated. However, no further problem was identified even though the clinicians had experience in their work that was inconsistent of the data. A tool like Sanguine can make the clinical team aware of deficiencies in the data collection process. However, the existing annotation system was limited in how expressive the clinicians could convey their knowledge to others. This problem sparked the following works in this dissertation, in order to explore more innovative techniques to externalize and communicate these data hunches.

Figure 4.1: The Sanguine interface for visualizing patient blood management data. The main workspace provides different types of charts that show the blood components transfused, test values, and patient outcomes. A heatmap shows intraoperative red blood cells (RBC) transfused, divided by patient outcomes; in this case, whether the patient survived the procedure. A dotplot shows RBC by preoperative hemoglobin value, and the dotplot allows analysts to use a brush to select a group of cases. The third chart is a dumbbell plot that shows preoperative and postoperative hemoglobin values over RBC units transfused. The vertical distance shows the difference in hemoglobin levels before and after surgeries, an indicator of patient blood management efficiency. The toolbox in the left side panel allows analysts to filter attributes including procedures, surgery types, blood components transfused, test value range, and case outcomes. In the figure, only the TAVR (transcatheter aortic valve replacement) procedure is selected. The detail view on the right panel shows selected case details, including surgery description, test values, and record numbers for cross-referencing with medical records in the database.



Main Tool Bar
Dot Plot
Heatmap
Chart Annotation
Filter Manager
Procedure Filter
Dumbbell Chart
Detail View
Filtering and Selection
Main Workspace
Patient Details

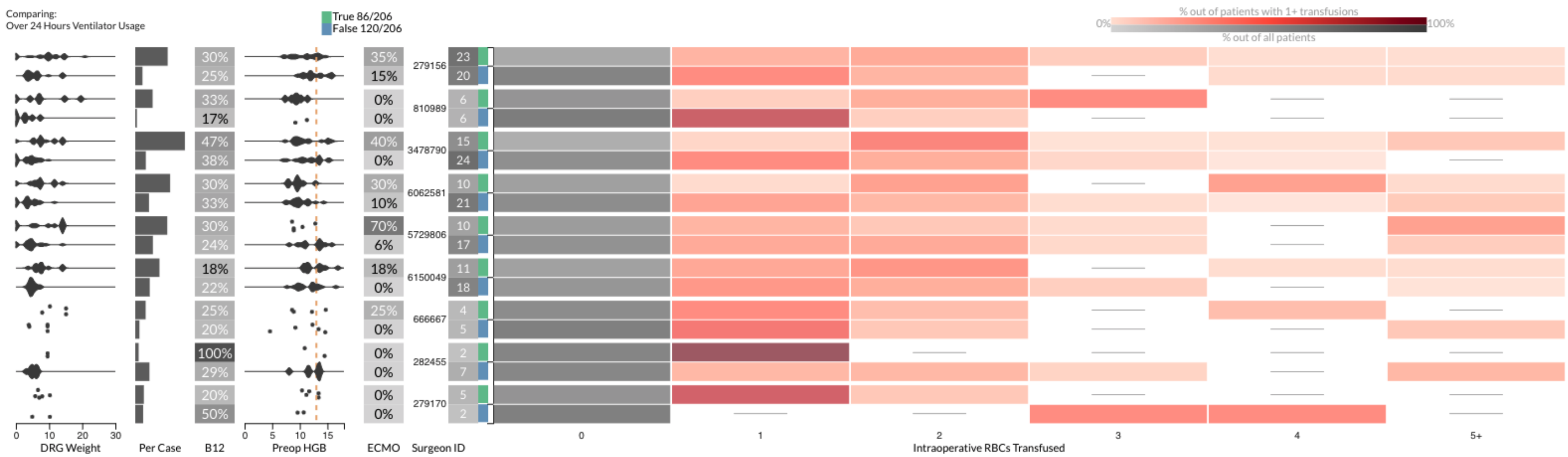


Figure 4.2: A heatmap showing transfusion data (red cells) for surgeons performing bypass procedures (CABG), subdivided by the need for long-term ventilation. Cases shown in the rows with a green indicator required ventilation for more than 24 hours, whereas cases in blue rows were removed from ventilation before that or did not require a ventilator. For context, five attributes are visualized on the left: distribution of DRG weights (risk scores), per case transfusion amount, B12 medicine usage, preoperative hemoglobin distributions, and ECMO usage rates. The red color map for red blood cell transfusion is set to exclude cases without transfusions, and the gray color map shows the percentage of zero transfusions in the total cases of each row.

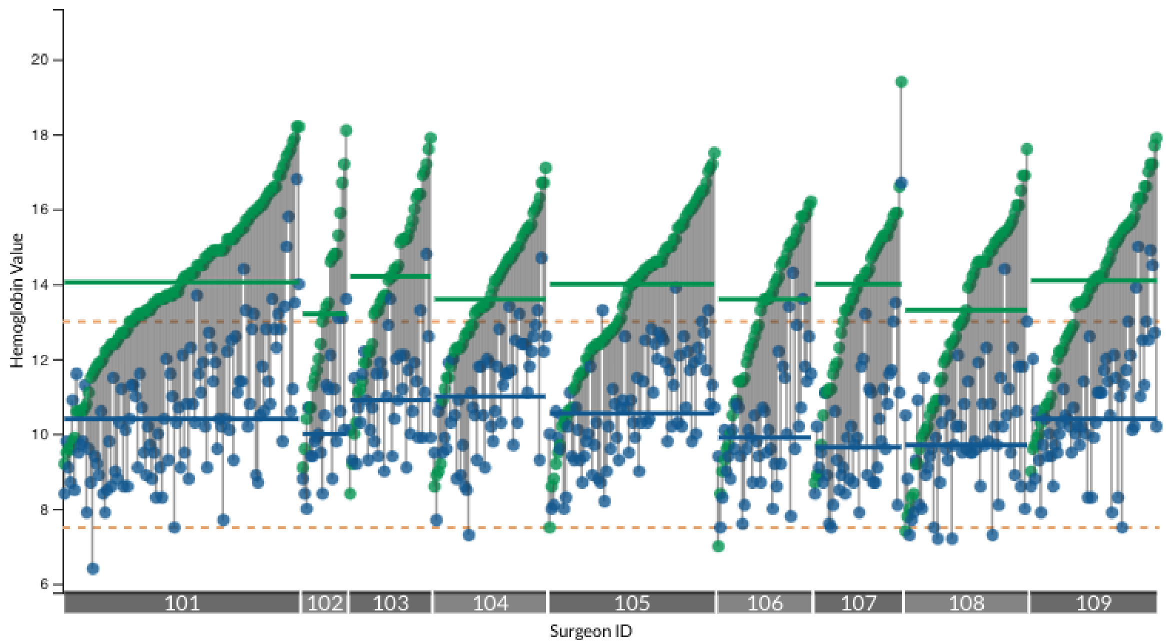


Figure 4.3: A dumbbell chart that shows pre- and postoperative hemoglobin values of elective surgery cases, grouped by surgeons and sorted by preoperative hemoglobin levels. Horizontal lines in each band represent the medians. The stippled lines are for references to preoperative hemoglobin level (13g/dL) and transfusion-trigger hemoglobin level (7.5g/dL). No RBC units filter is applied.

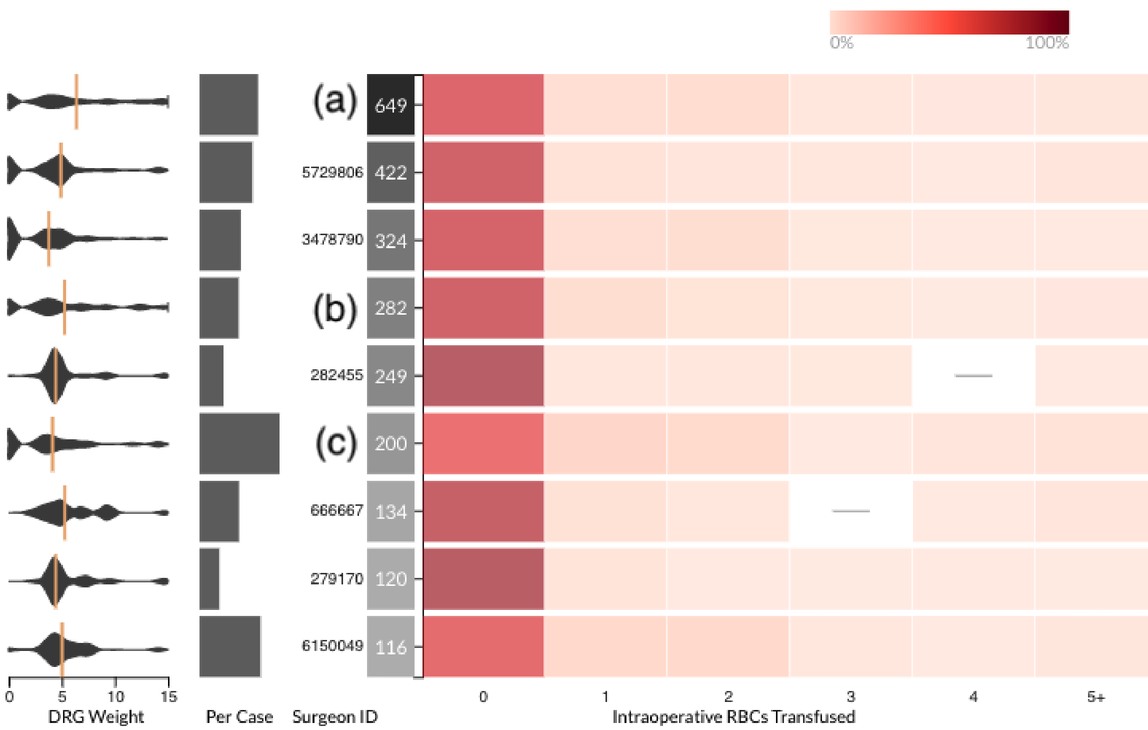


Figure 4.4: Heatmap of units of red blood cells transfused for the CABG, TAVR, Open AVR, VAD, and thoracic aortic aneurysm procedure. DRG weights and per case units transfused are added for peer group benchmarking. Surgeon (a) and Surgeon (b) have higher medium value for DRG values, 6.39 and 5.26, respectively, but their per case units transfused are 0.67 and 0.45, whereas cases treated by Surgeon (c) have relatively low average DRG weight (4.13), but a high transfusion rate at 0.93 units per case.

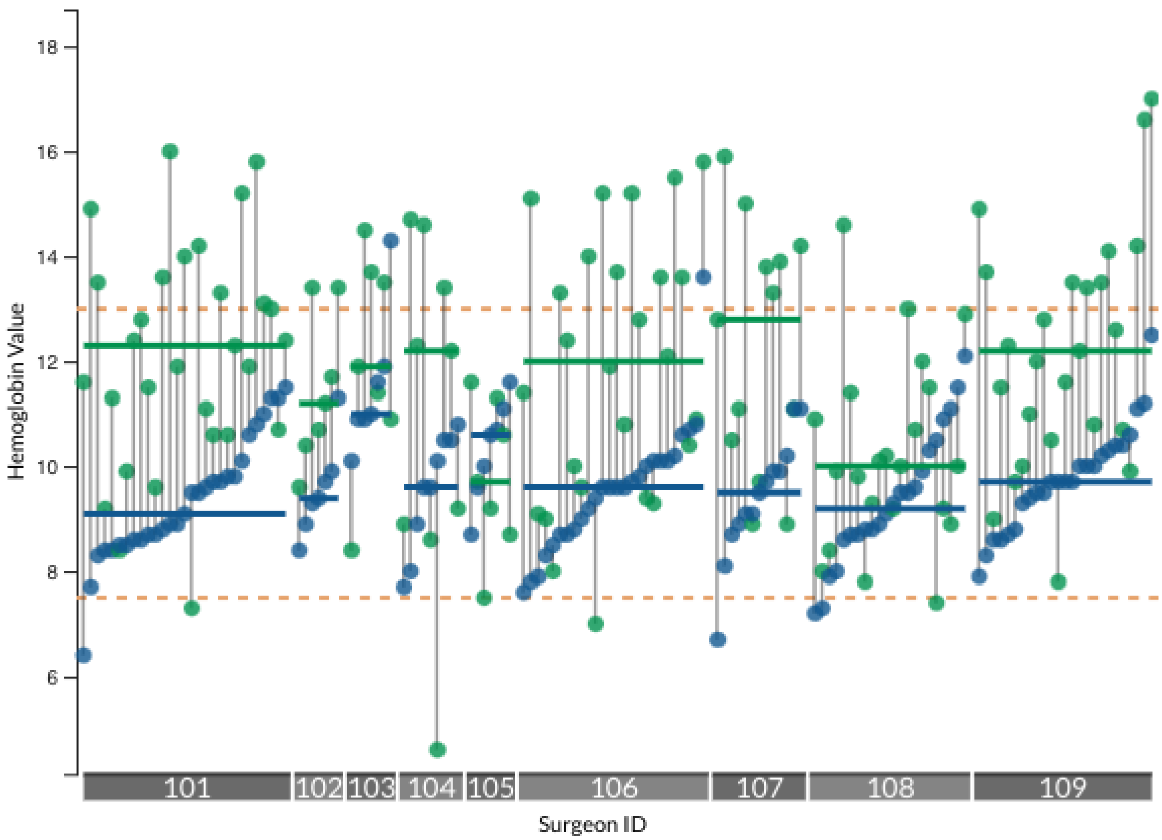


Figure 4.5: A dumbbell chart comparing surgeons, sorted by postoperative hemoglobin levels. Cases shown are with one or more units of red blood cell transfusions. The high postoperative hemoglobin values (blue line above the lower stippled line) indicate that over-transfusion is widespread.

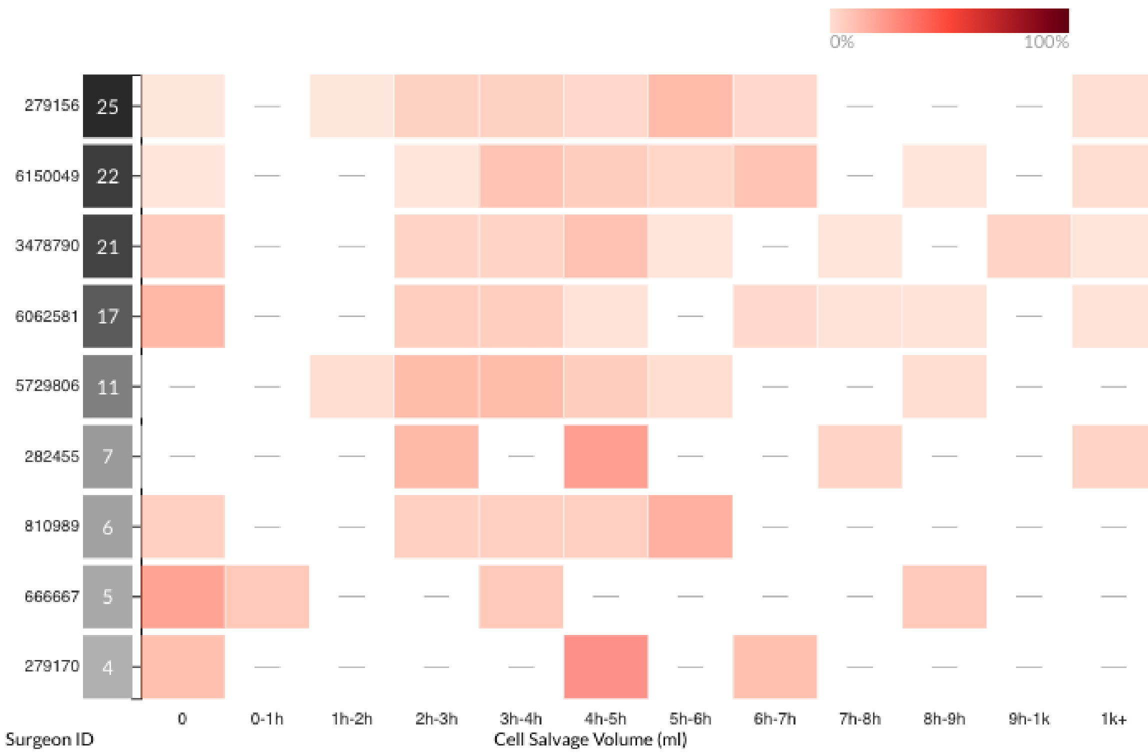


Figure 4.6: A heatmap of cell salvage usage for cases with one or more units of red blood cells transfused. Several surgeons do not or rarely use cell salvage when transfusing, not following best practices.

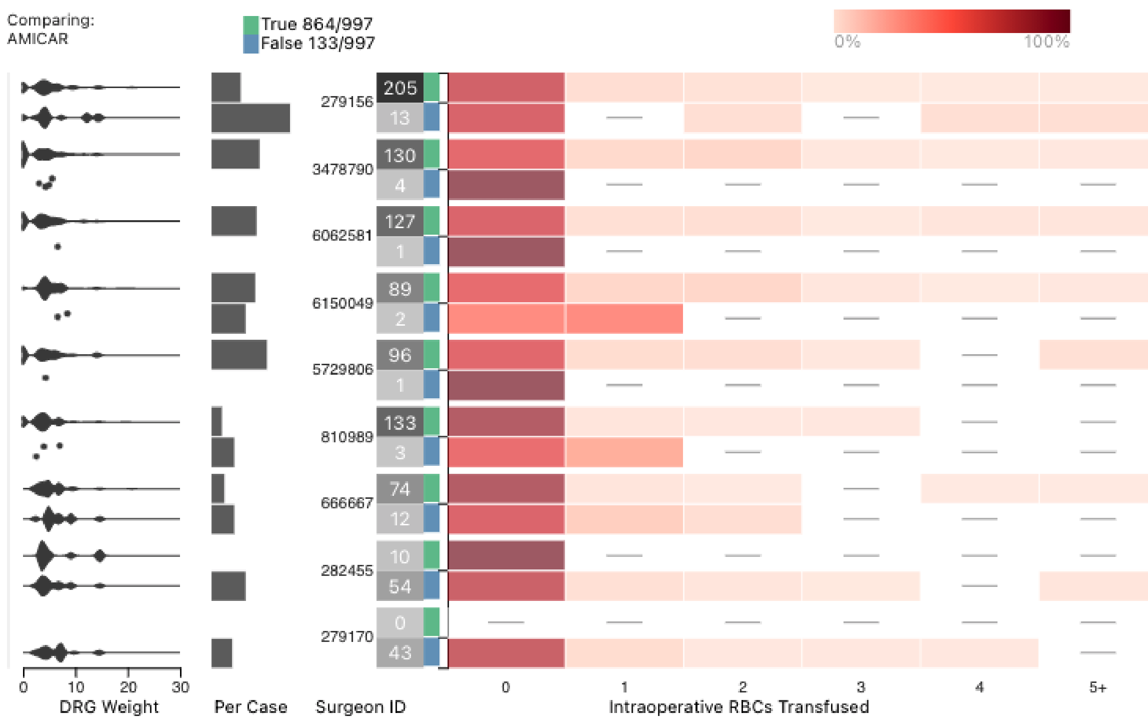


Figure 4.7: A heatmap of red blood cell transfusion, divided by the use of aminocaproic acid, a drug that can reduce bleeding. The violin plot shows DRG weights (risk scores), and the bar chart shows the units transfused per case. Cases using aminocaproic acid used fewer RBC units.

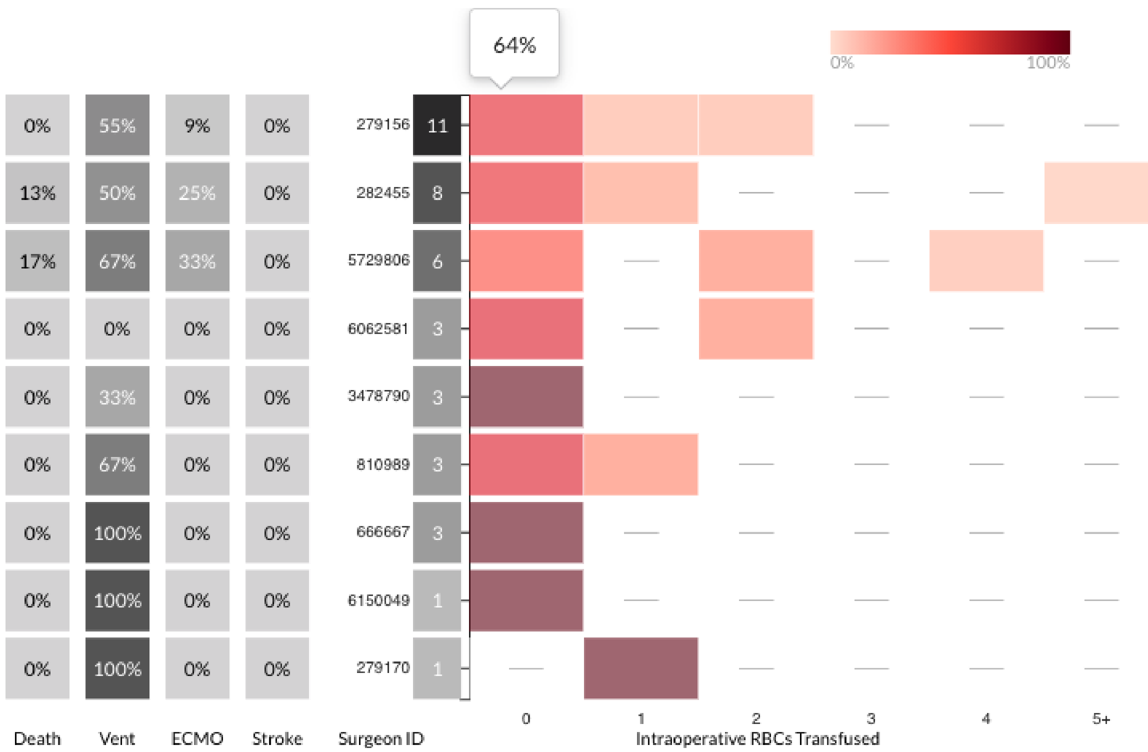


Figure 4.8: Heatmap of units of red blood cells transfused for the open MVR procedure and cases with preoperative hemoglobin levels of 10-12g/dL. By narrowing the records down to procedures and parameters specific to the case, surgeons can plan how many units of blood to order.

CHAPTER 5

INTERVIEW STUDY: EXPLORING THE ROLE OF EXPERT KNOWLEDGE IN DATA ANALYSIS

From our design study in Chapter 4, we had experts that used their working experience and voiced their concern about data not containing all the incidents that it was supposed to capture. This was not an isolated event. McCurdy et al. [11] also reported similar events where experts' knowledge was critical to have a more complete dataset for the analysis outcomes. Previous works have shown that in analysis pipelines, domain knowledge input was critical to achieving successful outcomes [69], [164]. Expert knowledge provides analysts with context and caveats about the data and assures analysts of the soundness of their analysis. However, there is yet much work in the visualization research community to systematically explore the details of how expert knowledge is integrated throughout an analyst's workflow. This lack of knowledge poses a challenge in providing extensive support to analysts in their data analysis tasks. To better understand the role of experts' knowledge in data analysis, we proceeded to undertake an empirical investigation through an interview-based study.

In this chapter, we share the results of an interview study with 14 domain experts and analysts from a broad range of fields. By recruiting a diverse group of analysts from academia and industry, we aimed to gain insights into the topic of data hunches in various scenarios. We investigate how they deal with data caveats in their workflow, how expert knowledge from various sources fills in the gap between data and reality, and how they currently practice documenting and communicating data caveats and knowledge in their work. Our analysis of the interview results provided us with insights into analysts' views on data imperfections and their strategy for incorporating various types of knowledge into their analyses. Furthermore, we expanded the scope of the study beyond data hunches.

We also investigated the tools that our interview participants use to communicate their analysis outcomes, in addition to how study participants present analysis results with caveats to different audiences in varied settings.

The primary contribution of this study is a comprehensive analysis of participants' current practices in capturing domain knowledge and an identification of the areas where these practices and their tools may fall short. By understanding the role of expert knowledge in data analysis, we can work towards developing better support systems and tools to enhance the analysis process and address the challenges posed by imperfect data and the need for effective communication of analysis outcomes.

5.1 Methods

The primary goal of the interview study was to investigate how analysts utilize, record, and communicate the usage of experts' knowledge in data analysis as a way to mitigate situations where data does not describe the reality or objective analysts' study perfectly. We drew inspiration from our collaboration with domain experts, including the design study in Chapter 4 and what we observed regarding the role of domain knowledge when they were using visualizations. In this study, we wanted to better understand the role of expert knowledge in data analysis. To this end, we recruited a mix of analysts from academia and industry to elucidate how domain experts and analysts apply data hunches in their own workflows.

5.1.1 Participant Recruitment

For our study, we sought participants conducting data analysis, i.e., those who actively use data to draw conclusions or inform decisions, as part of their work. We refer to our participants as *analysts* in this chapter. All our participants were professional data analysts with degrees in their domain or from "data fields" such as mathematics and statistics. Notably, none have formal training in computer science. Some participants had extensive domain knowledge based on their training (academics in the sciences, engineers), and hence, they were qualified as *domain experts* in the context of this chapter.

We recruited analysts through personal connections and used snowball sampling to identify additional participants. We recruited by e-mail; in our initial message, we disclosed that we were conducting interviews with analysts who work and collaborate on *messy* data, without any additional information about the interview topics or questions. The participants (4 men, 10 women) had a range of experience (4 to 30+ years) and had worked in a variety of fields such as civil engineering, legal services, atmospheric science, psychiatry, and policy-making (see Table 5.1 for details on the participants).

The interview protocol was submitted to the University of Utah Institutional Review Board (IRB) and deemed exempt from review. The participants gave informed consent to be in the study and to be audio-recorded before the interview. Participants were not compensated. We also discussed our anonymization protocol with participants, stating that their name or their organization's name would not appear in the publication. Ensuring anonymity helped us elicit unfiltered opinions on data—which was particularly important for participants in the public sector, since they did not wish to publicly speak for the organization they work for. We prioritized in-person interviews because they are more conversational, can help develop rapport, and may provide us with richer responses [165], [166]. Hence, 13 of 14 participants were based in Utah.

5.1.2 Interviews

The goal of our interviews was to study if analysts use personal knowledge in their workflows, and if so, how they utilize the knowledge, and what procedures they follow to document the process. We conducted two pre-pilot interviews with lab members to test the interview script draft and solicit feedback on the procedure and structure. We then conducted two pilot interviews with collaborators who met the inclusion criteria to test the outcome and modality of the adjusted script and structure. For the interviews that are reported in this study, we conducted 14 semi-structured interviews in total, 12 in-person and 2 remote. We scheduled interviews as the project progressed and decided to stop recruiting new participants when we reached saturation, noticing that no new topics were brought up. The interviews were conducted by two authors of the manuscript [3]. Lin and Lisnic using a two-to-one interview approach [167]. Lin asked the prepared questions and guided the conversation, whereas Lisnic observed the conversation, took notes, and followed up with additional questions. We used a two-to-one approach because we previously found it helpful in ensuring that interviews remained on track, while also lessening the burden of note-taking on the primary interviewer [168].

The interviews were scheduled for an hour and divided into three parts, *warm-up*, *current work practices* related to the role of knowledge in interviewees' data analysis, and *feedback on a technology probe*. In the *warm-up*, which lasted about 15 minutes, we first asked participants about their demographics and experiences, followed by a short activity, for which they were asked to write down titles of data-driven projects they had worked on. During the pilot interviews, we noticed that our participants got "stuck" with the project example they picked from the beginning and tried to fit their example to all the questions we asked. This observation led us to include an activity at the beginning of the interviews to help our participants reflect on a broader range of their past projects. The activity also ensured that our participants had a list of possible topics to refer to throughout the interview. We then asked participants to pick an example from the list they wrote down and give a high-level walk-through of their entire analysis process. The *warm-up* helped us familiarize ourselves with their domain and analysis flow, from obtaining the data, through processing and analysis, and to decisions and interpretations eventually made. It also helped us establish a good rapport to have a conversational and productive interview.

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

Finally, we transitioned to *feedback on a technology probe*, in which we presented a prepared slide deck with some techniques for recording and communicating data hunches [2]. After the presentation, we asked whether participants were currently recording any data hunches in their workflow, and how they might see the usage of these or similar techniques in their work if there were no technological limitations. The final segment of the interview helped us understand how to better support documenting data hunches, and served as a springboard to talk about tools and technological interventions the visualization community could develop to better serve analysts when working with data with caveats. The interview script and the technology probe presentation are available at https://osf.io/f89jp/.

5.1.3 Analysis

We used Otter.ai to transcribe the audio recordings of the interviews, followed by a manual quality check. We employed an inductive analysis approach to analyze the interview transcripts [169]. Three authors of the submission [3], Lin, Lisnic, and Lex, read, annotated, and labeled all interviews independently, and then met to discuss them. On average, it took an hour to read and annotate a transcript and another hour to discuss the interview. We paid close attention to statements that provided new or surprising perspectives, especially on how participants dealt with or communicated data hunches in their own workflow. Lin took notes and organized the notes and interview snippets into themes on a virtual whiteboard (available in supplementary materials at https://osf.io/f89jp/) during and after each analysis session. Following the initial analysis, we went through the identified themes from the first round of analysis and categorized them into groups that we present in Section 5.2. For readability, we tidied up reported quotes by correcting grammar and removing filler words (*like*, *yeah*, etc.).

5.2 Findings

We categorize our findings into three themes in this chapter: the relationship between data and reality, how knowledge fills the gap between data and reality, current practices in dealing with imperfect data. We cover our findings of interventions for better communicating data hunches in analysis workflows in Chapter 6 Section 6.6. These themes cover the full workflow of an analyst, from data collection and cleaning, to analysis and interpretation, and to finally delivering the analysis outcomes. We highlight key insights in italics and then share results from the interviews to support the insight.

5.2.1 The Relationship Between Data and Reality

Insight: *Data is shaped by socio-technical contexts. Understanding that context is critical for the analysis.*

Data is unable to perfectly or completely represent the world [170]–[172], and all of our participants were acutely aware of the gaps between their data and the phenomena they were analyzing. Several of our participants described how socio-technical contexts—infrastructures, cultures, relationships, human behavior—shaped what information their data contained, and what it was missing.

For example, P12, an education specialist, routinely analyzed student engagement with computer science in a local K-12 (primary and secondary education) school system. The elementary schools within this system do not have set courses for computer science, and thus there is no concrete way to track how much time a student is exposed to computer science material. Instead, teachers must self-report data on student engagement with technology. The pressure to meet requirements can induce over-reporting, "Some people feel like when they're self-reporting data, they don't want to have a zero. So then they say, well, [students] really got extra computer science in their science class; or however they want to justify it," as P12 said.

In another interview, P7 described the ways in which the US legal system dictated what data could and could not be collected about families affected by eviction court cases. P7 was studying how much the COVID-19 pandemic affected the local eviction rate, and if her agency could provide assistance for those in need. She described how it was impossible to know the actual number of people evicted due to the court not recording minors in their database. P7 lamented that she was not able to have a good estimate or overall picture of the eviction issue, "Many of these cases are going to be families with children. And we have no idea how many kids there are. So think of this number as the low end"(P7).

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

P6: So is it somebody smoking under the sensor? [...] Or is there a barbecue

> going on? Is there a fire? Or is it a malfunctioning sensor? Or did bugs [...] move into the sensors [...]? Those are just some of the issues that we deal with.

Insight: *Data is frequently repurposed, but repurposing is fraught and requires knowledge about the context of the dataset.*

Across interviews, we heard stories about how data was filled with caveats, shaped by the contexts in which the data was constructed. Nevertheless, study participants, fully aware that data is shaped by context, often repurposed data to suit their analysis needs. One participant (P4) used data collected by a foreign institute that used the data to study rainfall, however, he used the data to study snowfall models. As he was digging into the data, he failed to get meaningful results and finally realized that the data was not processed in the way that he expected. "So this is an auxiliary artifact of us trying to use the data for more than its original purpose," P4 said.

Participants P1 and P3, who studied suicide risks among certain populations, were using data labeled with International Classification of Diseases (ICD) codes, collected in a clinical context, as proxies for patient diagnoses in their research. Actual diagnoses can be extremely difficult to collect, and may even require handwriting recognition tools to compile diagnoses into a usable dataset. ICD codes were originally recorded for billing purposes, which resulted in instances where certain diagnoses may not represent the underlying truth.

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

Although P1 and P3 repurposed the data to suit their analysis, they stressed the importance of working with someone familiar with the data's original purpose. In this instance, P1 and P3 valued the input from their clinical colleagues with direct knowledge about the caveats on the billing codes. Even though the teams were aware that the data was an imperfect representation of patients' diagnoses, it was the best data they could get. The trade-off between accessibility and quality is often an issue that our participants face.

The perfect dataset for a particular analysis project is often unobtainable or does not exist, which leads participants to seek datasets that are good enough, though filled with their own caveats. Study participants employed different methods of working with caveats to fulfill their analysis. In the instance above of P4's snowfall modeling, knowledge about the way in which the data was processed allowed the analyst to use the data; yet this knowledge was not readily available. In the instance of working with patient data categorized with ICD codes, P1 and P3 sought confirmation and aid from clinicians who understood the codes from a professional context.

Insight: *Participants know their data is imperfect.*

In our interviews, we carefully posed our questions to avoid using the term *uncertainty*, to observe how participants would describe data imperfections and caveats. As a result, of our 14, only two participants (P2, P8), both working in weather forecasting, brought up uncertainty to describe the issues they faced with their data. Even though many data caveats that participants described could be labeled as qualitative or quantitative uncertainty, participants did not use these terms. We suspect that participants' expectations of data being imperfect and messy could be a reason for this. "It's never perfect. I'm not convinced I'll ever find a data source that's like 100% perfect. I at least haven't yet," P9 voiced her opinion on data imperfections.

Cleaning, sanity checking, and making sense of the data are part of participants' routine workflow. Furthermore, all participants responded to our questions about messy data by expressing views that their data was never perfect for their purposes. For example, P1 shared her experience with data preparations, "We've run into issues where the data didn't look right. [...] We always do data sanity checks, [they] are incredibly important."

We found that many participants were well-versed with the caveats that come with the data and brought in external expert knowledge in the analysis. P7, for example, reached out to others for more knowledge on the data. "But one thing I realized, as I started looking at this data is that the court doesn't do anything to clean [their data]. [...] I really needed subject matter knowledge [to process the data]"(P7). One participant(P12) expressed great faith in data in the abstract sense, but then quickly acknowledged that her data did have issues, "Numbers don't lie. Well, sometimes they did in my [use] case, but really, numbers don't lie"(P12).

Takeaway: Our participants regularly used data that is a limited and partial representation of reality. The data is shaped by socio-technical contexts and is accepted as imperfect. Many participants repurposed datasets to fill their analysis needs but did so with attention to detail and the original contexts in which the data originates. Failing to account for these nuances caused issues for some participants before. Across participants, there was a sustained sentiment that data was simply an imperfect tool for the analyses that they were trying to do, rather than a representation of reality that was marked by uncertainty.

5.2.2 Knowledge Fills the Gap Between Data and Reality

The primary way that our participants tried to fill the gaps between the available data and reality was by applying domain expertise or contextual knowledge about the data. This knowledge can come from an analyst's own prior experiences or familiarity with the data. Our participants also often solicited input from domain experts, more senior and experienced colleagues, or from individuals who could provide crucial context, such as those in local communities. In this section, we describe insights that pertain to applying personal knowledge external to the data in an attempt to paint a more accurate picture of reality.

Insight: *Diverse expertise is crucial for the appropriate interpretation of data.*

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

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

This caveat was not known to P9, nor had it been documented in the data and resources available to her. Awareness of this single caveat in the data opened the door to investigating and uncovering more—the team researched other fire incidents in the area and adjusted the calculations accordingly.

Similarly, P7, an analyst for a local government agency, studied eviction case data during the pandemic and worked closely with colleagues who have more domain expertise.

She discussed regularly presenting her dashboard to the group consisting of people from the city government and local nonprofit organizations to ensure that her data analyses were reasonable. "I presented it and said, 'I'm not a subject matter expert in evictions. Tell me what you see"'(P7).

Important input may come not only from subject matter experts, but also from more senior colleagues who either have more experience working with a specific dataset or simply can lend another pair of eyes. P13, who also worked for a regional government, described her experience of being the only data analyst in her office as being in a skill set "*silo*." Because of this, she often utilized her connections to analysts in other departments and reaches out to double-check her analysis results. "I'll frequently do gut checks, like, 'Hey, my analysis says this. Does that make sense to you?' [...] Without that I would be putting out a lot of very poor information," P13 added.

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

Participants often sought the knowledge of individuals with situated knowledge or lived experience about the data. These individuals included those who resided in proximity to the reality that the data describes, for example, people who lived close to a river that our participant was studying; or those whose lived experience was part of what is being analyzed, such as employees whose work output was reported on our participant's dashboard. Their proximity to the data provided additional expertise that is important to the analysis.

For example, P5, a civil engineer, used data on the depths of riverbeds to develop a disaster mitigation model. The data was originally collected using LiDAR (a method of mapping the terrain with lasers), but it suffered from inaccuracies. The laser could be reflected by the water, and, hence, might not capture the bottom of the river accurately. To remedy this, P5 told us about their collaboration with local partners, "We have a local partner who says, 'The channel is 20 feet deep.' But our LiDAR is showing that this is 15 feet deep. We'll say, 'Okay, we know it should probably be 5 more feet."'

Similarly, P6, a chemical engineer, deployed air quality sensors in various communities and regularly monitored air quality through a central dashboard. In times of anomalous air quality readings, she would email the local community first to check for any special events that might have impacted the readings, such as controlled fires set off as part of forest management, before concluding that a sensor is faulty and may need replacement. P6 recalled her experience, "We'll notice the levels are high and we'll be like, 'Hey, is anything going on?' And they're like, 'Yes, there's a controlled burn over here."'

Expertise that is important for a holistic understanding of data can also come from the lived experiences of the subjects of analysis. P11, an analyst in human services of the local government, discussed an example where input from workers about their working-hours data led to starkly opposite interpretations:

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

Insight: *Expert knowledge serves not to override data, but rather augment it for the purposes of decision-making.*

Several participants brought up the fact that the main goal of adding expert knowledge to their analysis is not to achieve precise value estimates, but rather to find more accurate and reasonable actionable recommendations. For example, P9 described her approach as "not striving for perfection, but for the most reasonable." When she was working with the land history expert, she would double-check the land value coming out of her analysis with the expert to verify her method's soundness and make sure that the result was within a reasonable range:

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

Participants also underscored the fact that they and their audiences typically understand that data is an estimate and not a perfect representation of reality. For instance, P7 discussed that data precision is less of a priority than finding actionable directions of work.

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

The challenge of making mental adjustments to data under practical time constraints is amplified in rapidly developing high-stakes scenarios, as illustrated by the example about the threat of nuclear war in Chapter 1. In another instance, P8, who worked in weather forecasting, discussed the role of expert knowledge in using radar data to rapidly distinguish between hail, a mostly harmless event, and debris lifted by a potentially destructive tornado:

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

This example also illustrates that balancing the need for (mentally) correcting data with the risk of disregarding evidence of legitimate signals requires careful consideration on the part of analysts. Expertise is especially important when distinguishing between unusual data values that stem from anomalies and those that reflect rare but important events. Since it is possible that the data provides a useful signal, our participants often do not overwrite or discard it. Instead, expert knowledge is typically embedded at the level of the final recommendation or interpretation. P2, an atmospheric scientist, expressed hesitation about discarding zero wind speed readings on a mountaintop, as it is rare but still possible for such a reading to occur naturally. He discussed how, although expertise is essential to adjust the interpretation of the numbers, the final interpretation cannot stray too far from the data:

> **P2**: We know we don't deal with the truth. We also try to make sure that conclusions are in line with what we've done, and that we're not stretching those too much. But that decision—that's a human decision. It's imperfect. [...] When you submit the paper, then the reviewers will also look at it from that standpoint, "Does what they did make sense?"

Takeaway: Our participants rely on personal knowledge to fill in the gap between data and reality when analyzing their data. This knowledge may come from domain expertise, professional experience, and proximity and familiarity with the data. The goal of analyses is to produce actionable outcomes, as opposed to precise numbers, and as such personal knowledge is especially important in interpretation and decision-making.

5.2.3 Current Practices for Dealing with Imperfect Data

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

5.2.3.1 Recording Data Hunches

Insight: *Participants often do not document their analysis decisions and the ways in which they adjust data. Written records are often incidental (e.g., e-mail) and not accessible to others.*

The majority of the participants did not document their analysis process at all. These participants' analyses are often adhoc: they analyze the data as required in the workflow. Few participants keep detailed records about data caveats and knowledge that is relevant to the analysis. Rather, the most common form of a permanent record is incidentally recorded conversations, such as e-mails or chat histories. However, these communication logs are only accessible internally and require knowledge about what to look for. They are also hard to archive due to limiting factors, such as chat history expiration. "I would guess that, my inbox ends up becoming a form of my notes, or we use WebEx chat. [...] But it's not documented in any sort of like, institutional knowledge transfer way, which is bad," P7 said.

Study participants often used email and chat apps, in which they posted text (P1, P9, P13) and screenshots of visualizations (P5, P6, P7, P13), to elicit feedback from peers or domain experts. In return, the outcome of these communications became analysts' temporary documentation. Many participants identified the issue of lack of long-term documentation, but found it hard to properly track knowledge input in their existing workflow due to resource or technical limitations. Particularly, a lack of support in tools was cited frequently as preventing properly documenting qualitative knowledge. One analyst (P11) described tracking caveats in cells next to the affected items in Excel, and P13 expressed hesitation about doing that because it might affect down-stream analysis tools, "Even just how to leave a better breadcrumb trail from Excel is something I'm not great at. Probably messing up my pretty sheet, you know?" Of 14 participants, only one analyst (P8), who worked at a national weather forecast organization, used an in-house tool with the capability for annotations built in to record any caveats in the weather forecast for shift hand-off.

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

Instead of explicitly documenting caveats in her workflow, P13 made mental notes. She also stated, however, that the lack of recording becomes problematic when new members join or leave the analysis team or when the project hibernates. P13 identified herself as the lone analyst in her department, which she stated, contributed to her lack of documentation. When an intern joined the department temporarily, she verbally communicated to the intern how to treat the data because of all the data caveats, such as, "You can ignore that data from X, because I know they're wrong for various reasons"(P13). This knowledge exchange happened adhoc, and P13 was unsure whether she covered all the data caveats exhaustively. She also stated that such a lack of documentation has led to wasted effort before, "I probably made some mistakes by re-analyzing data that I had forgotten I'd already sifted out"(P13).

P4 also faced an issue related to a lack of documented data hunches, already described previously. He downloaded atmospheric data from a foreign institute, but the data did not seem to make sense. He brought questions about this data issue to the foreign scientist, who stated that they had processed the data, but had not documented the processing steps. Because of the lack of proper recording of data caveats, P4 spent extra time and effort trying to make sense of the data. Additionally, he informed his peers about the data caveats through social media, yet the data hunch is still not officially documented with the dataset source.

Insight: *External pressures, such as strong community expectations or formal requirements lead to documentation of the data analysis process and data hunches.*

We found that scientists and engineers are more likely to document their processes and data hunches than study participants work in other sectors. For example, P2 reported taking "abundant field notes" about the measuring equipment condition and weather context for weather data collection and then documenting the relevant aspects in a methods section in a research paper.

> **P2**: It's very important, when you start analyzing field data, to have really good metadata, describing what was going on exactly where the system was, what the operating parameters are. [...] We tend to keep lots of notes, so we can go through and make sure that what [...] we think we're seeing in the data [is what] we're seeing.

To document uncertainty and other assumptions, P5 prepared an appendix for his reports on flood disaster models. This report, with an appendix, went through a strict, multistage review process. P5 described the process in details, "It's reviewed by somebody within our district, and then it goes through an agency technical review [...] by somebody outside of the district. So there are a lot of quality checks that happen to make sure that [everything is] accounted for"(P5). P14, an epidemiologist, commented on her organization's usage of metadata sheets. They used metadata to capture caveats in their analysis to make sure they have proper data context for the analysis down the pipeline, although she stated that improving their internal documentation is a strategic goal.

Takeaway: Participants sometimes but not always record data caveats and knowledge exchanges. Recordings are often not systematic and rely on incidental records in chat or email applications. The recording is situational, dependent on the profession and habit of the analyst. Engineers and researchers are more likely to document detailed caveats and analysis processes, whereas other participants keep more adhoc records for their analysis.

5.2.3.2 Communicating During the Analysis Process

Insight: *Generic communication tools like email and chat applications are the most common way to communicate data caveats during the analysis.*

Communication is critical when there is a mix of expertise on a team working on the same project. We found that, in addition to synchronous meetings, email exchange was the main communication tool that study participants used to elicit domain knowledge and feedback from the experts. However, they described being frustrated when dealing with text and screenshots. Data caveats that participants dealt with could be complex, and participants found text to be inadequate for effective communication. Sometimes, important data caveats were buried in long emails (P9), in other cases, emails were not exhaustive enough to describe the issue (P1, P9, P13). P9 described an unpleasant experience where the data provider did not clearly communicate the assumptions that went into the data collection in their email exchange, and that the analyst in charge of interfacing with the provider was new to the job. This mixture of inexperience and ineffective communication led to a wrong analysis, undermining a high-stakes legal case. The common issue is the ineffectiveness of conveying data caveats through words. "In general, I email people and I feel like a lot is lost in translation in the emails," P13 added regarding her dislike for email exchanges.

Insight: *Participants use screenshots or screenshares of visualizations to communicate about data and data caveats, but do not use annotation features of their visualization tools.*

Our participants commonly stated that visualizations are essential in their analysis and communication process, either in person or over virtual meetings. P9 explained about her process, "My preferred approach is to get on a Zoom call and share my screen or have them share their screen [..] which is much more efficient than a phone call or an email"(P9). Even though study participants drew and annotated on visualizations during these meetings, the drawings were not archived, and, hence the knowledge exchange was not preserved. Alternatively, participants used screenshots in chat, e-mail, or PowerPoint, which they might or might not annotate. We found no instances of annotations happening directly in a visual analytic tool that participants used.

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

Therefore, even though tools like Tableau or PowerBI (which are among the tools used by our participants) support annotation, participants used simple graphics editors or did not annotate directly on the visualizations instead.

Takeaway: During the analysis process, participants use tools, such as email and chat applications to elicit and communicate data caveats. However, many found that using text is inefficient when it comes to complex data hunches. Therefore, screenshots or screenshares of visualizations are added to the conversation, which may be annotated.

5.2.3.3 Communicating Analysis Results and Data Hunches

Study participants used a variety of mediums to report their analysis outcomes. These mediums included research manuscripts (P1, P2, P3, P4), project reports (P5, P9, P11, P12, P13, P14), live presentations with slides (P9, P12, P14), and dashboards on websites (P6, P11, P14). Because of the differences in the method of delivery, communicating data hunches to the intended audiences took different forms, which we will break down in this section.

Insight: *Participants that write reports or papers use established textual formats to report on data hunches.*

Participants who were academics (P1, P2, P3, P4) most commonly used method sections in their publication to report data hunches. Similarly, participants who wrote up their analysis as reports (P5, P9, P11) documented data hunches as an appendix or as part of their reports. However, these formats are commonly kept brief and may not provide enough detail to reproduce the analysis or reuse the data without issues [173].

Insight: *Participants prefer static visualizations for presenting analysis results and use bullet points for data hunches if they consider the hunch essential.*

Most study participants used visualizations to communicate their data, as they valued the accessibility of visual representations.

> **P9**: I feel visuals are really helpful, throughout the process. So as we're doing summary statistics, we're always creating some sort of visual to go with it, especially when you're the person [that knows] the data, communicating with the person that's not [familiar with] the data. Having visuals is a great way to translate between those two.

When communicating their results in meetings, participants found slides and handouts to be useful mediums. Participants frequently added bullet points in their slides next to their static visualizations to explain the context or caveats required to properly interpret the data. "I have a little disclaimer at the bottom [of my slides ...] that notes that [this approach is] not going to catch duplicates," P7 elaborated on her work. However, when asked about preparing and anticipating questions about the data, many participants answered that they would respond adhoc, rather than preemptively cover data imperfections.

Insight: *Participants omit data caveats in their results because it adds complexity and they see* *it as their professional responsibility to synthesize data into an easily digestible format.*

We found that several participants (P5, P7, P11, P13) were hesitant to communicate caveats. Some participants saw it as their professional responsibility to distill data into actionable items for decision-makers, and that they were trusted to correctly interpret the data to the best of their ability. Communicating more data caveats to leadership increases the complexity of the presentation, and leadership and external audiences often are not interested in the details of the analysis. "Part of our job is to synthesize down to the main points for leadership. When it's getting to the mayor as talking points or a policy memo, if it has too much [about caveats], it's just going to be a distraction," P7 commented.

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

We discovered different approaches to expressing caveats about the analysis depending on the format and the type of participants' jobs. For example, our academic participants were cautious about keeping their analysis choices transparent in research manuscripts (P1, P2, P3, P4), whereas other participants felt hesitant to express uncertainty explicitly in their analysis deliverables due to the reasons laid out above. This omission of caveats was even more amplified in verbal presentations, which our participants justified by being available for clarifying questions if needed. However, participants also adjusted to the stakes and the certainty of their analysis. When making recommendations about COVID-19 policies, for example, P14 stated:

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

As we previously discussed, participants more often added disclaimers to deliverables that were used asynchronously, such as dashboards. Even here, however, the focus was to avoid making the results confusing. For example, P14 refrained from adding all data caveats to her public COVID-19 dashboard because she did not want to confuse the general public with the data complexity. "There's no reason to put [data caveats] on the website if people aren't going to understand it, or if they're going to misinterpret it," P14 commented.

However, she would add additional annotations to her public facing dashboards if she learnt that an aspect was regularly misinterpreted, which she measured by the number of calls she received about an issue. To elaborate on the practice, P14 said, "If it's a broad misunderstanding, and we're getting a lot of public calls, I might add something to the dashboard that has something embedded in the figure, [such as] shading for when Delta started or when Omicron started."

P14 also weighed how important the knowledge of data caveats was for the data she was presenting. She would repeatedly inform policy makers about caveats on hospital bed usage reporting—it was a snapshot of the actual usage, dependent on the time of the day it was reported, because that data caveat could be influential for the policy made from it and was potentially a much higher stake than a public dashboard.

Additionally, analysts can be restricted by the limited outlets to provide more context. P13's analysis outcomes were often presented as at most a single page report. Therefore, she needed to keep the outcome concise and the message clear. The extremely reduced reporting, professional responsibility, and the concern for distraction made analysts refrain from reporting more data caveats with their analysis results.

Takeaway: To communicate data caveats to others, our participants use method sections and metadata notes for reports, visualizations with notes for presentations, and footnotes for dashboards. The extent of the communication is situational, often dependent on the expectation of the stakeholders, the effects on the outcome, and the format of the communication.

5.3 The Role of Personal Knowledge in Data Analysis

Prior work has characterizes data as an artifact of decisions: a culmination of the specific and situated contexts in which they were constructed [9], [12], [174], as we have shown in Chapter 2. The construction of data leaves it with gaps and caveats such that for data to fully reflect reality, data requires context [12]. In the interviews, participants discussed many different ways that they understood and worked with the limitations of their data. To fully understand the reality, data must be placed into the appropriate context, and our interviews with analysts provided additional evidence that data is full of caveats that need to be considered.

Experts' personal knowledge often complement the data, piecing together the spaces between data and reality. These experts, in more traditional sense, are scientists, engineers, physicians, with an advanced degree and deep understanding of their domain. However, through our interview, we found analysts utilized knowledge outside of the traditional domain expert as conceptualized by the visualization community [132], [175]. The experts ended up being anyone close to the data—aware of how data is constructed or of the environment from which data is derived. In almost all interviews, we heard many accounts of how recording personal knowledge in analyses is brittle and unsystematic: scattered across ephemeral records like chat histories or one-off emails containing notes, or communicated in a meeting. Thus, the lack of documentation makes reanalysis and reproducibility challenging, creating a barrier for other analysts outside the discussion to join. The interview study provides evidence that more intuitive and integrated support is needed for expressing data hunches.

Furthermore, the participants never expected the data to be perfect. In fact, even though some participants had strong faith in the numbers' accuracy, they still shared experiences where the numbers were imperfect representations of what they were trying to study. The data was imperfect for many reasons, including errors in measurement devices, human factors, the data being originally generated for different reasons, or the data being simply unattainable.

Caveats about the data were often not communicated for a variety of reasons, most prominently because our participants felt that it was part of their professional responsibility to make easy-to-interpret and actionable analyses from the data. They were trusted to communicate what was necessary from their analysis and this excluded many of the caveats that they worked with. This finding complements what Hullman found about why authors do not communicate uncertainty: because showing uncertainty is difficult for the author, and reading charts with uncertainty is difficult for the audience [22]. We saw evidence in support of both, but the role of the expert as a trusted party that abstracts complexity was unique.

The literature on uncertainty addresses only part of the concern when it comes to visualizing the imperfections of data. Uncertainty expressions like confidence intervals, hypothetical outcome plots [45], and ensemble plots [38] focus on conveying the uncertain nature of the data and are well-studied within the visualization community. And yet, throughout our interviews, most participants did not bring up uncertainty when describing their data. Instead, we found that most participants described how they adapted their workflows to account for data and its caveats. They were, in fact, *certain* about the data's limitations and were able to reduce the effects of the limitations through knowledge of the data's context.

Data is not perfect—our participants did not believe that it is and neither should the visualization community. Across interviews, we saw the importance of context when it came to how our participants understood and handled imperfections in their data. Our participants turned to people who were close to the data to fill in those gaps and, in turn, made decisions on what aspects of the data they would present to decision-makers. Within analysis scenarios, knowledge about the data is more important to record for purposes of reanalysis than for communicating final results. In contexts of trust and expertise, there is a common understanding that the data is meaningless without knowledge of its context.

5.4 Limitations

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

5.5 Conclusion

We conducted a series of interviews with analysts from various fields and levels of experience to investigate how expert knowledge influences their analysis. Our findings provide evidence that most of the data used in analyses is imperfect, highlight the importance of including and documenting personal knowledge in data analyses, and emphasize the potential pitfalls of neglecting this information. Experts' personal knowledge, beyond traditional domain expertise, was instrumental in bridging the gaps between data and reality. However, the study revealed that the documentation of personal knowledge in analyses is often scattered and unsystematic, hindering reanalysis and reproducibility. We also collected feedback on potential interventions to support the recording and communication of data hunches more effectively. The findings highlight the need for intuitive and integrated support for expressing data hunches. We continue the exploration of techniques for expressing data hunches in Chapter 6.

Table 5.1: Overview of the characteristics of the 14 participants across different fields in academia and industry.

ID	**Field**	**Title**	**Specialty**	**Experience (Y)**	**Typical Deliverable**
P1	Psychiatry	Professor	Suicide and Autism	10+	Research Manuscript
P2	Atmospheric Sciences	Professor	Snowfall Prediction	30+	Research Manuscript
P3	Psychiatry	Professor	Genealogy and Suicide	30+	Research Manuscript
P4	Atmospheric Sciences	Post-Doc	Rainfall Prediction	5	Research Manuscript
P5	Civil Engineering	Engineer	Disaster Prevention Models	4	Model Report, Recommendation
P6	Chemical Engineering	Professor	Air Quality	20+	Project Dashboard
P7	Government	Strategy Manager	Housing and Eviction Program	10	Policy Recommendation
P8	Atmospheric Sciences	Science Officer	Weather Forecasting	10	Forecasts
P9	Environmental Economics	Consultant	Consulting for Legal Purposes	5	Reports
P10	Government	Politician	Public Health Legislation	14	Policy
P11	Government	Data Analyst	Human Services	5	Dashboards, Reports
P12	Education	Specialist	CS Education	20	Policy Reports, Resource Allocation
P13	Government	Specialist	Public Defense Policy Analysis	10+	Reports, Recommendations
P14	Epidemiology	Program Manager	Infectious Disease Surveillance	4	Dashboards, Healthcare Reports

CHAPTER 6

DESIGNING FOR DATA HUNCHES

The interview study in Chapter 5 confirmed our hypothesis that analysts frequently encounter data filled with caveats in their analyses, and often use experts' knowledge to complement the data caveats. The study provided evidence that data hunches play a crucial role in the analysis process and can provide valuable insights into the underlying data. In Chapter 6, we introduced the definition of data hunches and described the methodology we used for developing the framework to describe and design space to externalize data hunches. Recognizing the significance of data hunches, it is essential to classify and categorize them to better understand their nature and potential impact on data analysis. On the other hand, we found the existing tools that our interview participants used were inadequate for intuitively externalizing data hunches along with the data. The absence of reliable tools and, as a result, lack of proper documentation of data hunches, can lead to experts' knowledge being overlooked during the analysis process. This can result in data misinterpretation, and the reproducibility of an analysis may suffer. Therefore, establishing a design space and design guidelines are crucial to facilitate the sharing and documentation of experts' knowledge, ultimately enhancing the analytical process.

In this chapter, we introduce three types of data hunches, assessment hunches, structural hunches, and value hunches. The classification helps us to better understand and decide the appropriate techniques for different hunches. We then dive into ways of expressing data hunches through interactive visualizations. By providing the foundation to visually and intuitively express data hunches, we seek to ensure that crucial expert knowledge is not lost or forgotten during the analysis process. The proposed design space allows analysts to articulate their insights, concerns, and alternative perspectives, leading to a more comprehensive understanding of the data and its limitations. We also consider how data hunches can be used in collaborative settings, starting with simple features such as voting and sharing. Based on our interactions with the prototype for externalizing and communicating data hunches, we develop design guidelines for visualization designers to consider when incorporating data hunches in their visual applications.

Because we had established the design space prior to conducting the interview study, we were able to gather valuable feedback from our interview participants regarding our visualization techniques for data hunches. Their evaluation was based on their individual workflows and the specific audience they report to. In this chapter, we present the feedback we collected and reflect upon the validity of the proposed design space.

6.1 Types of Data Hunches

In our development of data hunch classifications, we listed all the data hunches we encountered in our previous design studies, brainstormed other forms of data hunches that may surface in data analysis, and generalized them based on whether they are qualitative or quantitative, the specificity of the hunches, and the forms they take if expressed. We identify three types of data hunches in support of determining suitable methods for reporting and communicating expert knowledge about limitations of data.

- **Assessment Hunches:** Assessment data hunches speak to the trustworthiness or quality of a dataset or individual data items, or simply provide context. Assessment hunches can take different forms, ranging from ratings (thumbs-up/down, numerical scores, etc.), to written comments about data items or datasets.
- **Structural Hunches:** Structural data hunches state that certain data points or relationships should not be included in the dataset (exclusion) or that a data item or relationship is missing (inclusion). In a network dataset, for example, an inclusion data hunch could be used to indicate that an edge is missing. For many data types, inclusion hunches should also be combined with an estimate of a value of the included hunch. For example, when indicating that an item is missing from a dataset, the data hunch could also contain an estimate of the value of the item.
- **Value Hunches:** Value hunches make a statement about how a specific data value should be different from what is recorded in the dataset. Value hunches equally apply to numerical, categorical, and textual/label data. For example, a value hunch for a category could state that an item should be in category A instead of category B.

For practical reasons, we found it useful to further distinguish three methods of expressing value hunches, reflecting different levels of "precision" about a value hunch. *Directional Hunches* express that values should be different (higher or lower) from the recorded value without giving a specific value. They are a middle ground between assessment hunches, which make no statement about directionality, and hunches that give estimates for actual values. *Specific Value Hunches* express exactly how values in a dataset should be different. For example, a value data hunch could state that the value encoded by a bar chart should be 20 instead of 12. *Value Range Hunches* acknowledge uncertainty about the value to be specified. Instead of expressing a specific value, it states that a value should be within a certain range. For example, a range hunch could state that the value of a bar should be between 18 and 22 instead of 12. We acknowledge that different sub-types of any of the higher-level types of data hunches could be useful in different contexts.

All these types of data hunches can be expressed for an individual data item, groups of data items, or whole datasets. For example, a value data hunch could apply to a single point (this should be twice as much), to a few data points (all of these items should be twice as high), or to the whole dataset (all data points should be twice as high). These different types of data hunches also establish that hunches may not always be precise, supporting analysts to record some data hunches across varying levels of knowledge. In Section 6.2 and 6.4, we how to design visual applications with data hunches in mind. With an emphasis on visual methods, knowledge about sources of qualitative uncertainty can be represented through graphical elements and interpreted context of the data.

6.2 Recording Data Hunches

A key aspect of data hunches is that they are expressed during analysis by a diverse set of stakeholders. Hence, visualizations of the data is the ideal medium to express, record, and consume data hunches. In this section, we explore the set of approaches that can be used for recording data hunches on top of a visualization.

6.2.1 Data Space

As analysts explore and conduct analysis on datasets, they often discover missing data, duplicates, or unusual patterns in the data. These hunches about the data come to fruition without many visual aids, and hence can be directly expressed through numbers and models. A basic method to record a data hunch is to manipulate the data in data space, before the data has been mapped to a visual element. We consider form-based manipulation and model-based manipulation as the two main methods for recording data hunches in data space. **Form-based manipulation**, shown in Figure 6.1a, is concerned with inputting a data value or an attribute of the data point through a form, a table, etc., and is suitable for data hunches for specific data items. **Model-based manipulation**, illustrated in Figure 6.1b, uses a model to bulk-input or edit data values, e.g., through a mathematical function. Note that manipulating the data in data space does not imply that the original data is overwritten.

Previous works have explored ways to express models and values to record knowledge in visualizations. Marasoiu et al. [66], for example, presented an interface that allows users to sketch models, which then generates data points based on the sketch, as a way to facilitate communication between customers and analysts. Romat et al. [68] included data editing in their digital ink externalization system, a functionality requested by participants. Although this functionality was added post facto, it illustrates a preference for editing data directly in systems. Kandel et al. [176] proposed an integrated statistical analysis tool to highlight data anomolies within the data space and incorporated visualizations to assist with assessing the data in context. Data space is not well suited to communicate the data hunches that have been recorded in the context of visualizations, as e.g., a tabular representation of data hunches would be detached from the visualization of the data. Instead, designers will have to consider methods to visualize data hunches provided in data space in visual space. Kandel et al. [177] explored another form of direction manipulation in form based transformation, but with more focus on using scripts to automate the process.

6.2.2 Visual Space

Recording data hunches in visual space on top of a visualization provides a direct connection between a data hunch and the visualization. Analysts can think about the data hunch in visual terms as they manipulate it, and consider other data points that are visualized while recording their hunch. Another key benefit of recording data hunches in visual space is that, to a large extent, the same encodings can be used for recording and communicating data hunches. We suggest two techniques for recording data hunches in visual space: free-form sketching and direct manipulation.

Free-form sketching refers to adding visual elements directly to a visualization, using approaches such as pen/mouse-based sketching, or adding elements to a visualization using functionality similar to a drawing program (Figure 6.1c). The technique provides freedom for analysts to express their data hunch in the way they see fit and can record a variety of data hunches, including structural exclusion through crossing out data points, value/directionality (e.g., a value should be higher in reality) through drawing an arrow, categorical value through shading an area in a color, and value range through drawing an area where a value is expected to be. Visual markups have been used for note taking and communicating though process in visualizations [54], [66], [68], as well as conversation starters [178]. The process of graphical externalization helps with the understanding of and reasoning about visual information [179] and supports reading and reflecting on the visualization [13],[180],[181]. A downside of markup is that it cannot (easily) be converted into structured data, and hence is connected only to the visualization, but not to the underlying data, making reuse of these hunches in other visualizations of the same dataset exceedingly difficult.

Direct manipulation, illustrated in Figure 6.1d, involves moving, resizing, removing, adding, or otherwise changing parts of the visualization that encode data. While restricting analysts to the marks and channels of the visualization, direct manipulations offer beneficial affordances: dragging a bar element is easier with a mouse than sketching a new bar, for example. Manipulated marks are also straightforward to translate into data space. Previous works have suggested direct manipulation on visual encodings is a viable way to edit data and provide visual demonstrations of thought processes. Baudel [182] presented editing single or groups of data items in a dataset using graphical manipulations in data visualizations. Saket et al. [183] used graphical manipulation, through repositioning, resizing, and recoloring marks in visualizations, to help users express their expected visualization with increments in direct manipulations, and in turn, the system suggests visual transformations. A drawback of direct manipulation is that each possible manipulation has to be designed and implemented for each chart type, in contrast to free-form sketching, which can be implemented once and re-used for all types of hunches and charts.

6.2.3 Abstract Space

Assessment data hunches can be expressed only through text, comments, or ratings. For example, an analyst might know that a data source is unreliable, but might not have a concrete idea on what the true data should be. To record such a hunch, they want to add comments to the data and the data visualization. Such hunches are recorded in "abstract space," as they do not directly suggest a different structure or value. We identify two methods through which hunches can be recorded in abstract space: **structured elicitation** and **textual annotations** (in addition to, e.g., hand-writing using free-form sketching features).

Structured elicitation (Figure 6.1e) uses form-based UI elements, ratings, and up/down votes, whereas textual annotations (Figure 6.1f) are free-formed notes. Previous works have explored the use of rating, structured form, and textual annotations in data visualizations. Quispel and Maes [184] used ratings to investigate preferences of visualization types between groups of people. McCurdy et al. [11] used structured forms to elicit data hunches from domain experts, and Goyal et al. [62] offered more freedom to users by allowing them to use a notepad for free-form notes during their experiment. Structured elicitation is different from form-based manipulation in data space: although both can use forms, structured elicitation is about assessment, whereas form-based manipulation is used to express concrete hunches in data space.

6.3 Prototype Implementation

We used an iterative design and development process for our prototype system, with the intent of embodying our ideas and evaluating their feasibility. Over a period of six months, we sketched alternative design ideas and implemented promising solutions, subjecting them to evaluation within our team of four authors and subsequent discussion within our research lab. We designed and implemented many variations that we subsequently abandoned for various reasons (see the supplementary material at https://osf.io/syjkf/ for examples). We describe the lessons learned from the entire process that stretched over more than a year as guidelines in Section 6.4.

Our eventual interactive web-based prototype (available at https://vdl.sci.utah.edu/data-hunch/) showcases visualization methods for all types of hunches we describe in Section 6.1. For specific value hunches, we also provide dedicated methods for larger numbers of data hunches. The prototype also implements various appropriate methods for recording data hunches (see Section 6.2) for each type. A specific hunch, for example, can be recorded in data space through form-based manipulation, through model-based manipulation, as well as in visual space through direct manipulation and free-form sketching.

A concern for designing hunches was scalability: how we can show many different data hunches reliably on top of a visualization. We address scalability in two ways: first, we implement dedicated encodings for hunches in case too many hunches are recorded. Figure 6.2 shows specific value hunches on top of three bars. We use sketchy bars for one or two hunches on top of a bar, but switch to sketchy ticks for more hunches, as shown on the top. However, switching from bars to ticks violates the guideline for making hunches appear similar to the original marks and hence is a trade-off we have to make. Second, we support collaborative features. If a data hunch is already present, another person sharing the hunch could endorse, reject, or comment on data hunches instead of logging a new data hunch, thereby reducing the number of hunches that need to be visualized simultaneously.

We use three example datasets: COVID-19 case counts in selected countries, greenhouse gas emissions across the food supply chain (both downloaded from OurWorldInData), and the size of research areas at the School of Computing at the University of Utah. Data is loaded from CSV files, where data hunches are stored in a Firebase database. Data hunches can be freely added by guests or after signing-in via Google. A tabular visualization (Figure 6.3) gives an overview of all data hunches and their meta-data, although all relevant information about data hunches is also available through the visualization interface.

Our prototype is open source, and the source code is available at https://github.com/visdesignlab/data-hunches-package. We use React and D3 for rendering the UI and the visualizations, and the RoughJS library (https://roughjs.com), which is based on the techniques for sketchy rendering developed by Wood et al. [185], to render visual elements for data hunches in sketchy patterns.

6.4 Guidelines for Designing for Data Hunches

To demonstrate the feasibility of the techniques we proposed in Section 6.2 and to explore possibilities of communicating and recording data hunches, we developed a prototype, shown in Figure 6.3, allowing users to record their data hunches on a simple bar chart, and we describe the implementation in details in Section 6.3. As described in Section 3.2, we used an iterative design process to develop the prototype. We describe insights gleaned from this process for designing visualizations that support recording and displaying data hunches in the form of design guidelines.

6.4.1 Do Not Change the Original Data

Techniques to express and communicate data hunches aim to enable analysts to express their knowledge about the data, but not to alter the dataset. Data hunches and original data are different entities and should be clearly separated, to both avoid confusion about the difference between data and data hunch and to retain the integrity of the original data. Data hunches are also valuable only in the context of the recorded data that they refer to. For example, for a value hunch expressed in Figure 6.3h, we use a sketchy font and arrows to indicate a different value from the original data, and place the value hunch next to the original value. Furthermore, while transparently expressing data hunches, such as doubts or ideas about what a data point should be, is valuable contribution to the analysis process, editing data can be considered deceptive or even fraudulent.

From a technical aspect, designers should treat recorded data hunches as a completely seperate dataset that only references (elements in) the original dataset. In our prototype, for example, data hunches are recorded as a separate dataset, which is shown in a table next to the visualization (Figure 6.3m). An unfortunate consequence of separating hunches from data is that off-the-shelf visualization systems and libraries are unlikely to make it easy to render data hunches in addition to the underlying data.

6.4.2 Make Data Hunches Distinct

Consistent with our arguments about not changing the original data, data hunches also need to be clearly distinguishable from the original data within the visualizations. Furthermore, the encoding used for data hunches should not only be distinct from the primary encoding, but also should clearly communicate that what is shown is not the original data.

In our prototype, we use sketchy rendering [185] for visual elements and handwriting-style fonts to make data hunches discernible from the crisp, clean lines of the original visualizations (see Figure 6.3). The goal of using sketchiness was to make it obvious, even to first-time users, that data hunches are not original data but that instead they represent people's thoughts and knowledge. Wood et al., for example, speculate that "sketchy [..] visualization has a role to play in constructing visualization narratives where an author's voice is important" [185]. In our first designs, shown in Figure 6.4, we attempted to render data hunches using a different color, and experimented with gradients and blur to communicate uncertainty and ranges. However, we abandoned this design as we realized that it could lead to confusion, as the marks could be interpreted as belonging to the original data. To avoid this confusion, we developed a rule we applied throughout our design process that all data hunches should look as if they were hand-sketched or written on top of a visualization, to emphasize the humanness of the hunches.

We do not argue that sketchiness is the only, or even the best, choice to communicate data hunches. Other designs, tailored to other visualization techniques and affordances, are conceivable. This reasoning applies to all the implemented features we describe in this section.

Another consideration is to be mindful of how disruptive data hunch encodings could be when placed over the original visualizations. The original visualizations must remain visible to properly interpret data hunches. For example, in an early design, we rendered a bar representing a larger data hunch over the original bar. We abandoned this idea in favor of hatched bars that ensure that the underlying data point remains visible.

6.4.3 Make Data Hunches Similar

Data hunches should use the same or similar visual encodings as the visualizations of the original data. While this guideline seems like a direct contradiction of the guideline on making data hunches distinct, we believe that data hunches and the original data should be read together without the need for mental conversion. For example, a value data hunch could be expressed as a written numerical value on top of a bar chart. However, we argue that this would make it difficult for an analyst to judge the relative differences between the original value and the data hunch. Comparisons are easier if both are expressed using the same visual channel (size/position in the case of a bar chart). Our prototype uses sketchy bars on top of regular bars for numerical specific value hunches (Figure 6.3i), and hatched color for categorical hunches (Figure 6.3a), in both cases using the same encoding channel as the original data. However, some hunches cannot be expressed using the same visual encoding. For example, a range value hunch (providing an estimate that a value should be in a certain range) is not compatible with a bar chart encoding used for the original data. To address this, we use a position encoding on the same scale as the bars, showing the middle and the extend of the range (Figure 6.3e). We use similar techniques for directional value hunches (Figure 6.3b) and hunches that do not fit on the scale of a chart (Figure 6.3h).

To reconcile this guideline with the guideline of making hunches distinct from visualizations of the original data, in our demo we relied on using an additional visual channel that was not used in the original visualization: sketchy texture.

6.4.4 Keep Data Hunches Close

As assessment data hunches (comments, ratings, etc.) are not expressed in data space, using the same visual encoding is not feasible. However, designers should ensure that assessment hunches can still be read easily together with the original data. For example, instead of showing assessment hunches in a table, they could be rendered next to the element they are referring to, illustrated for annotations in Figure 6.3d and for ratings in Figure 6.3g. If a textual hunch requires more space than is available in the chart, we truncate the comment and reveal the full text in a tooltip. We also considered what to do with assessment hunches referring to the whole dataset and opted for placing a note and an asterisk next to the chart title (Figure 6.3k); our reasoning is that analysts might read the title and caption as they are attempting to understand the visualization.

6.4.5 Use Direct Manipulation

Data hunches emerge when analysts explore the data and examine the data visualizations. Hence, the thinking and analysis process happens in visual and data space. While we lay out different approaches for recording data hunches in Section 6.2, we argue that recording of data hunches should be done as close to the way the data and the data hunch are presented as possible. In our prototype, we trigger recordings by right-clicking on a mark or legend whenever possible and provide methods to record a data hunch through direct manipulation of the data hunch as it will appear once it is recorded (Figure 6.5). When recording a hunch in data space, or when recording an assessment hunch, a visualization can provide visual feedback for the analysts (Figure 6.6). Also, in our prototype, we place the input forms for data hunches right next to the visual elements in the chart.

6.4.6 Design for Data Hunches

Not all visualization techniques are equally well suited to visualize data hunches. In our prototype, we have chosen a bar chart with categorical values because bar charts are an important class of visualizations, and because they have affordances that are compatible with data hunches. For example, an analyst could express a value data hunch on top of an individual bar without affecting a neighboring bar. Other visualization techniques, however, do not equally support such similarities. For example resizing a segment in a pie chart, or in a stacked bar chart, requires affecting the other data marks, or overplotting. While different approaches for designing data hunches for bar charts may exist (small multiples, for example), they are likely not as effective at integrating data hunches with the original data in one visualization.

Another consideration is the complexity of a chart: the more complex, and the more visual channels are used, the more difficult it will likely be to find a suitable design for data hunches. Similarly, visualizations that give overviews of large amounts of data, like cluster heatmaps, will require different approaches for data hunches, as the manipulation of individual data items is less relevant and the visualizing of the hunches more challenging.

Nevertheless, even when employing visualization techniques well-suited for representing data hunches, such as bar charts, line charts, or scatterplots (see Figure 6.1), we suggest some practices to better support data hunches. For example, our original design used vertical bars (see Figure 6.4). However, we quickly found that vertical bars are problematic for rendering longer comments (assessment hunches) next to the bar due to the text orientation, so we switched to horizontal bars. Likewise, our original design had a chart title and a subtitle at the top, which made it difficult to find a suitable place for comments on the whole chart. Hence, we moved the title below the chart, and reserved space below the subtitle for comments. Finally, we found it useful to leave white-space from the beginning, so that data hunches can be easily expressed and rendered. For example, our prototype has large margins to the right of the bar chart, so that larger value hunches and comments can be effectively rendered. The designer can also use binning and grouping to organize several data hunches and associate them with a data point, a technique that Badam et al. [186] adopted in FacetNotes. While these specific examples may not directly translate to other visualization techniques, the larger lesson of thinking about position, layout, and space for data hunches as a designer creates a visualization holds.

6.4.7 Design for Collaboration

Data hunches are predominantly a medium to communicate knowledge about data to others, and hence, data hunches are inherently collaborative. Data analysis activities are also commonly collaborative efforts in the first place. We argue that data hunches should be designed with collaboration in mind.

Our prototype acknowledges the importance of collaboration by allowing multiple people to log in and review data hunches. However, we also speculate that enabling multiple collaborators to record only data hunches might be insufficient, as collaborators might also want to endorse, reject, or comment on others' data hunches. To address this aspect of collaborating, we introduce features to endorse or reject a particular data hunch, using a thumbs-up or thumbs-down metaphor, illustrated in Figure 6.2. Another more expressive method is to add capability to comment on data hunches. This way, a team member can express their sentiment about a data hunch without having to re-specify a data hunch they endorse.

6.4.8 Elicit Context and Accountability

What a data hunch says about the data is different from *why* a person has the hunch. The context of a data hunch is as critical for its interpretation as the context of the data. Scholars in the field of critical data studies have asserted that data cannot be considered out of context [12],[171],[187], and identity can affect how trust is established in collaborations [188]. Similarly, the reasoning and identity of the data hunch author can effect how a data hunch is perceived and trusted.

We propose that along with recording data hunches, context is important for establishing trustworthiness. This may include reasoning about the data hunch or the identity of person making the hunch. As designers work with stakeholders to determine how data hunches are recorded, they should also explore how important contextual information can be recorded and shared. In our prototype, we require that analysts provide reasoning for and express their confidence in a data hunch when recording it (Figure 6.5). Additionally, the identity of the data hunch author is recorded as an attribute of the data hunch. We then visualize these attributes in a tool-tip (Figure 6.3), although other more salient approaches are conceivable. For example, it might be worth exploring the use of opacity to encode the confidence of a data hunch. These attributes not only enrich the recording of data hunches, but also allow for features such as filtering and sorting.

We acknowledge a tension between revealing identities, to ensure accountability and leverage networks of trust, and the desire to be anonymous to record inconvenient opinions or facts. In prior work, for example, we found that experts in an organization were unwilling to record hunches under their name due to tensions in the organization. In addition to logged-in recording of data hunches, our prototype also provides a "log-in as guest" option to record a hunch without revealing one's identity. We further discuss this issue in Chapter 7.

6.5 Usage Scenario

Here we illustrate how data hunches could be used in a scenario with real data about the size of research areas in the University of Utah's School of Computing, shown in Figure 6.7. To begin, a faculty member, also a co-author of this paper, first noticed that while the largest bar was for *UNKNOWN* students, there was another bar labeled *N/A*. He recorded a hunch that *N/A* should be removed (a structural hunch) and that *UNKNOWN* should be bigger (a value hunch).

Upon reviewing the classification of research areas into larger fields, he expressed concern that *Databases* is classified as *Data and Visualization*, given his knowledge about the type of research conducted by the database groups at the University of Utah, and that it should rather be in the *Computer Systems* category — he recorded his hunch about the different classification of the bar (a value hunch). He expressed a similar data hunch for *Graphics*, which should be in the *Other* bin.

When reviewing the number of *Architecture* researchers, he was surprised by the seemingly high number. He provided a value hunch that he considered closer to reality, and added a comment speculating that some Electrical and Computer Engineering students advised by Computer Science faculty are included in this count. Finally, he realized that *Robotics* is likely shown smaller than it is, probably because some students are incorrectly classified as *AI* instead. He left a value range hunch also noting his reasoning. This faculty member then passed on the visualization to a second faculty member, another co-author of this paper, who reviewed his hunches, upvoted several, and left some additional hunches about her own thoughts on where the data did not reflect the makeup of the department.

6.6 Interview Feedback

Because we conducted the interview after proposing the techniques to visually express data hunches and building the prototype to showcase them, we solicited feedback during the *feedback on a technology probe* stage of the interview. We introduced our definition of data hunches to the participants and provided a brief demonstration of the techniques, including textual annotations, graphical manipulations, and collaborative viewing and voting. The demo included the basic workflow of recording and communicating a data hunch and how the data hunch looked when being recorded in the visualization using sketchy rendering. The demonstration illustrated both the concept of data hunches and possible technical solutions to recording and communicating data hunches during collaborative analysis scenarios.

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

Participants liked the collaborative and visual aspect of the prototype (P2, P3, P4, P5, P7, P8, P9, P11, P13). P9, for example, commented that being able to annotate and express data hunches efficiently would allow her to make sure that everyone was on the same page on the project:

> **P9**: So if you had something like [the data hunches prototype], where then the analyst was sharing their screen and making the adjustment, [to show] what they think experts are talking about, and the expert could actually see it adjust

> in front of them, then everyone can make sure they're understanding. [...] I think that would kind of bridge the knowledge gap between the data people and the experts in a successful way.

The techniques we presented used interactions directly on a data visualization, which participants considered to be an easy way to express opinions. P11 reflected that most of the domain professionals she worked with were not optimal at verbally expressing their opinions and knowledge about the data. An interactive visual option, she commented, would be a good option for these collaborations. "I think sometimes they just don't really know how to phrase what they want to see. [...] this to me seems useful [...] for people who aren't the [visualization] designers to be able to offer feedback, " P11 added.

Insight: *Interactive visualizations can help with the feedback loop and be more inviting and engaging for audiences without analytical backgrounds.*

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

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

Several participants commented on how the interactive visual techniques we presented could be helpful for asynchronous knowledge transfer. P4, a postdooctoral researcher, discussed how he could use sketches and annotations to provide his knowledge on the caveats of a dataset, and that those could be archived so that others could know about the caveats.

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

Insight: *Fatigue with tools is prominent. Another tool to add to the already complicated workflow is not desirable.*

On the other hand, participants' desire for a separate solution along the lines of what we presented with the data hunch prototype was in tension with their aversion to adding another tool to their toolbox. Changing and juggling between tools can be challenging, especially during a meeting, as P9 described the process as "scrambling between apps a million times." Some participants were concerned that a new tool that is separated from their existing workflow might be a burden to experts who have great insights into and knowledge of the data.

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

Our participants, especially the ones with 10+ years of experience, much prefer solutions that fit their existing workflow. The yet-another-tool problem has already been something that our participants were facing and they did not want to add more tools to record and communicate data caveats.

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

As a result, we found that our participants preferred interactive visualization techniques to record and communicate knowledge about the data. A visual approach that can help get analysts and experts on the same page is appealing. However, participants showed yet-another-tool fatigue and said they would prefer a solution integrated with their current tool suite to externalize and exchange knowledge.

6.7 Epistemology

The visualization community has characterized the imperfect and partiality of data as uncertainty; that if all errors and imprecision can be accounted for in the data, we can arrive at the perfect representation of reality that covers all perspectives, and as a result, there can exist the perfect dataset to describe the phenomenon. However, this understanding of data is not the only available classification. Within the fields of critical data studies and digital humanities [189], data is an artifact of decisions and situated contexts that reflect one captured slice of reality: "Data are capta, taken not given, constructed as an interpretation of the phenomenal world, not inherent in it" [9]. Data, then, as an object of decision-making practices, is one representation of many possibilities. This recognition is important because it highlights the non-neutrality of data [170], [174], [190]. These perspectives stem from different epistemological roots, in contrast to the positivist epistemology that most computer scientists are traditionally inclined toward. Feminist perspectives on data are based on the theory of "situated knowledges" [191]. The feminist theory posits that knowledge cannot be obtained from a single source but rather is best derived through a collection and collaboration across partial and overlapping perspectives. Data, similarly, cannot fully represent the natural world. From this theoretical grounding, data and data hunches would capture different perspectives of reality but contribute to a richer more complete picture. This dissertation asserts that data merely offers a partial portrayal of the world. With the recognition of personal knowledge, we can, therefore, have more integral representation and more holistic understanding of the world. This is only a musing, however, leading us to bigger questions, such as: In what ways could other epistemologies be productive in visualization research? Could we better characterize and describe the entangled relationship between data, visualizations, and people?

6.8 Conclusion

This chapter positions and theorizes data hunches and elevates the role of personal knowledge in visual data analysis. We presented the types of data hunches and the three spaces, abstract, data, and visual, to express data hunches. We envision analytical tools incorporating data hunches in an analyst's natural workflow, and, hence, we propose suitable techniques for each space and showcase a prototype to demonstrate data hunches in action. We collected feedback on the prototype during our interviews with analysts and present the result in this chapter. We also outline the eight design guidelines when realizing data hunches in applications.

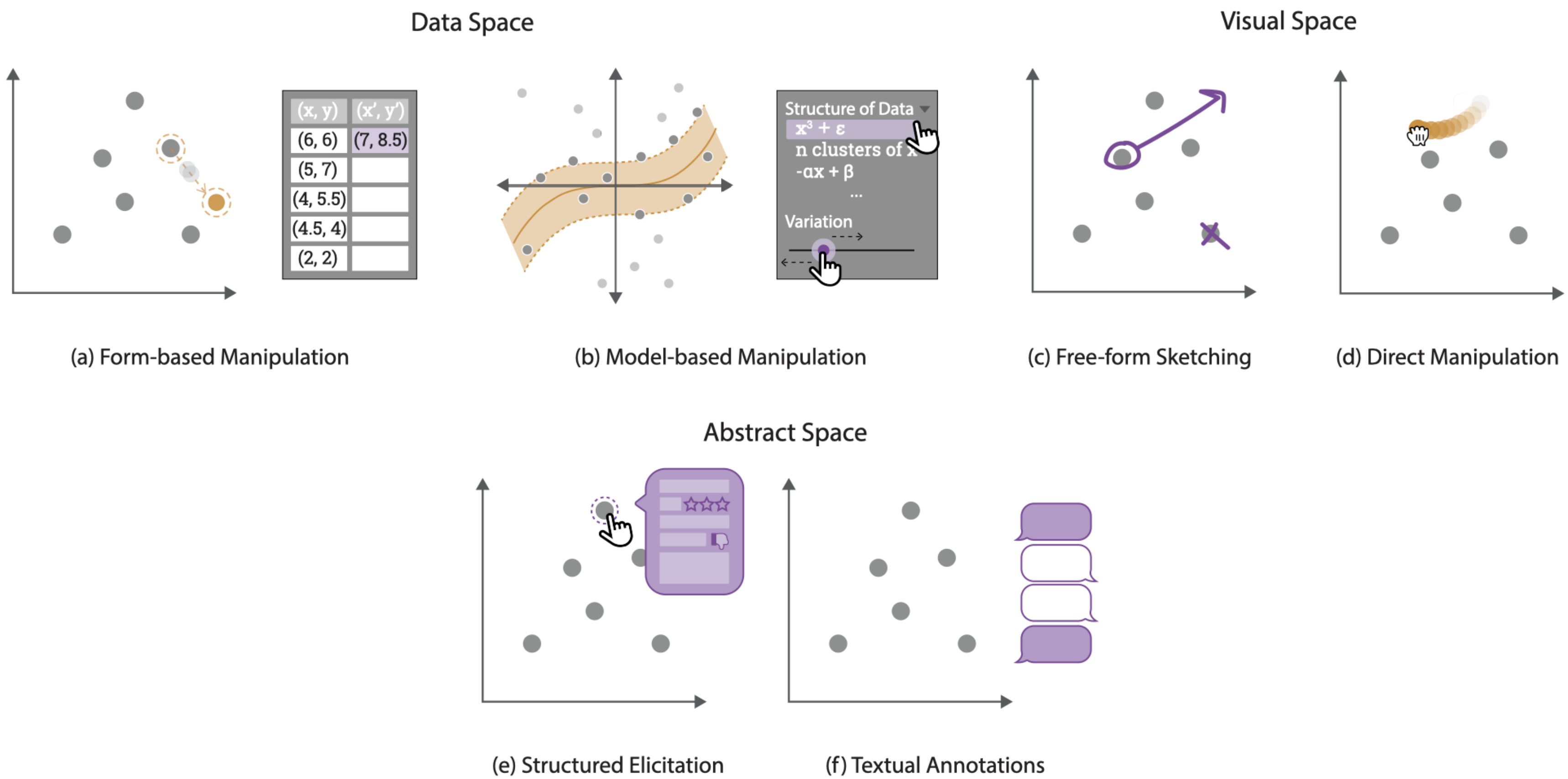


Figure 6.1: A set of abstract techniques to record data hunches. We distinguish three spaces: data, visual, and abstract. Recording in **data space**, such as through (a) form-based manipulation and (b) model-based manipulation, affords a basic technique where analysts can externalize hunches by either manually inputting items, or expressing them through models. Recording in **visual space**, such as through (c) free-form sketching and (d) direct manipulation, affords visual recording of data hunches, ranging from sketching to dragging graphical elements to represent an analyst's hunch. Recording in **abstract space**, such as through (e) structured elicitation and (f) textual annotations, enables analysts to record hunches that cannot be recorded in the data and visual spaces.

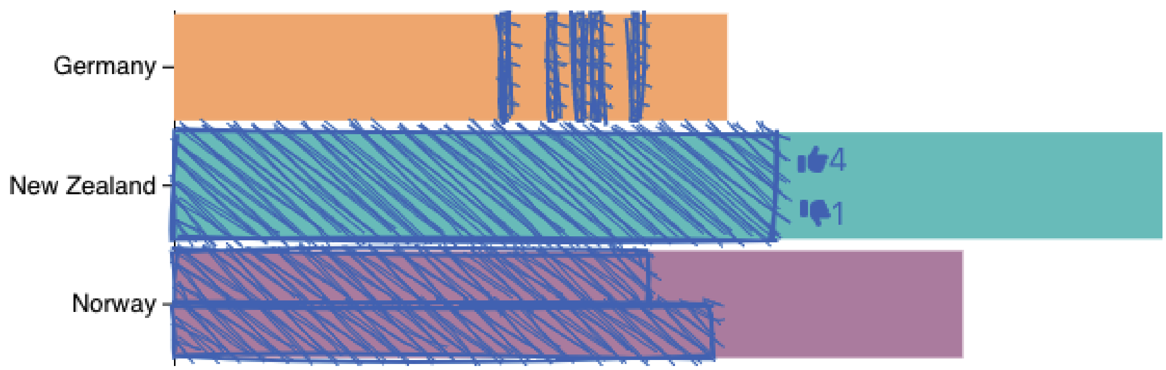
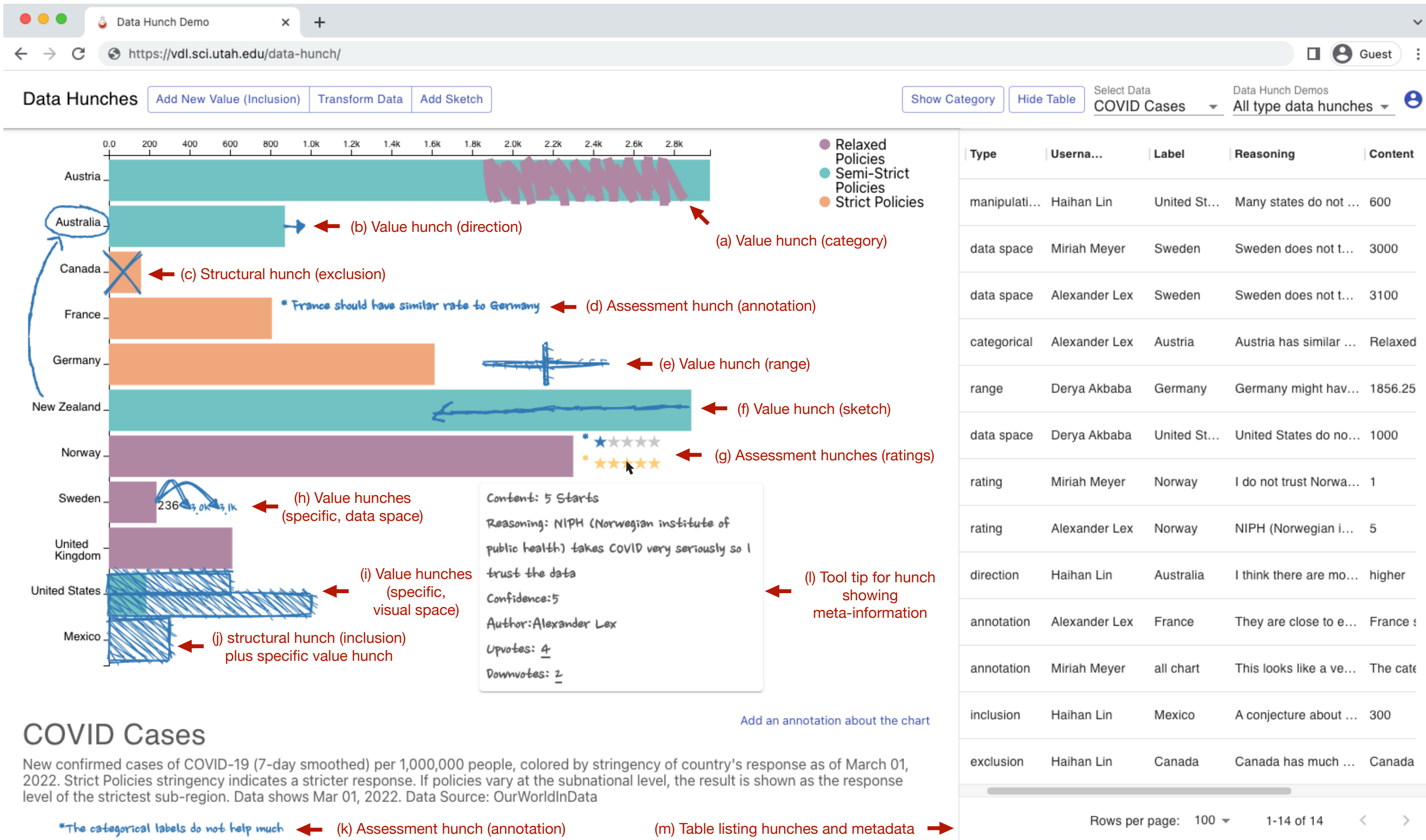


Figure 6.2: Collaboration features, such as endorsing, rejecting, and commenting on data hunches, help in reducing the need for logging similar data hunches. Different visual encodings account for scalability in value data hunches. If more than three data hunches are logged for a single bar, we replace the sketchy bars with sketchy ticks.

Figure 6.3: Our prototype showing a variety of data hunches expressed using tailored methods and designs. Note that red annotations have been added to explain the figure. Everything else is as it appears in the prototype. The main chart, showing COVID-19 data and data hunches, is on the left. In the chart, we demonstrate (a) a categorical value hunch that indicates Austria should be colored in purple as it has relaxed policies that are not reflected in the data. (b) A directional value hunch expresses that the data value should be higher, without specifying details. (c) A structural hunch indicates that Canada should not be part of this chart. (d) An assessment hunch is used to comment on the values associated with France. (e) A range value hunch indicates that German values should be higher within a certain range. (f) A value hunch expressed using free-form sketching indicates that values for New Zealand should be lower. (g) Two assessment hunches show diverging perspectives about the quality of Norway's data. (h) Two value hunches expressed in data space indicate that values in Sweden should be much higher. They are rendered as arrows/labels as they would break the scale. (i) Two specific value hunches show alternative opinions on how high the values for the United States should be. (j) A structural hunch indicates that Mexico should be included in the chart and gives an estimate for a value. (k) An assessment hunch comments on the whole of the dataset. (l) A tool-tip shows information for a selected hunch, including reasoning, confidence, endorsements, and rejections. (m) A (hide-able) table lists all hunches.


Data Hunch Demo
https://vdl.sci.utah.edu/data-hunch/
Guest
Data Hunches
Add New Value (Inclusion)
Transform Data
Add Sketch
Show Category
Hide Table
Select Data
COVID Cases
Data Hunch Demos
All type data hunches
0.0
200
400
600
800
1.0k
1.2k
1.4k
1.6k
1.8k
2.0k
2.2k
2.4k
2.6k
2.8k
Austria
Australia
Canada
France
Germany
New Zealand
Norway
Sweden
United Kingdom
United States
Mexico
Relaxed Policies
Semi-Strict Policies
Strict Policies
(a) Value hunch (category)
(b) Value hunch (direction)
(c) Structural hunch (exclusion)
* France should have similar rate to Germany
(d) Assessment hunch (annotation)
(e) Value hunch (range)
(f) Value hunch (sketch)
(g) Assessment hunches (ratings)
236
(h) Value hunches (specific, data space)
(i) Value hunches (specific, visual space)
(j) structural hunch (inclusion) plus specific value hunch
Content: 5 Starts
Reasoning: NIPH (Norwegian institute of public health) takes COVID very seriously so I trust the data
Confidence:5
Author:Alexander Lex
Upvotes: 4
Downvotes: 2
(l) Tool tip for hunch showing meta-information
Type
Userna...
Label
Reasoning
Content
manipulati... Haihan Lin United St... Many states do not ... 600
data space Miriah Meyer Sweden Sweden does not t... 3000
data space Alexander Lex Sweden Sweden does not t... 3100
categorical Alexander Lex Austria Austria has similar ... Relaxed
range Derya Akbaba Germany Germany might hav... 1856.25
data space Derya Akbaba United St... United States do no... 1000
rating Miriah Meyer Norway I do not trust Norwa... 1
rating Alexander Lex Norway NIPH (Norwegian i... 5
direction Haihan Lin Australia I think there are mo... higher
annotation Alexander Lex France They are close to e... France
annotation Miriah Meyer all chart This looks like a ve... The cate
inclusion Haihan Lin Mexico A conjecture about ... 300
exclusion Haihan Lin Canada Canada has much ... Canada
Rows per page: 100
1-14 of 14
Add an annotation about the chart
COVID Cases
New confirmed cases of COVID-19 (7-day smoothed) per 1,000,000 people, colored by stringency of country's response as of March 01, 2022. Strict Policies stringency indicates a stricter response. If policies vary at the subnational level, the result is shown as the response level of the strictest sub-region. Data shows Mar 01, 2022. Data Source: OurWorldInData
*The categorical labels do not help much
(k) Assessment hunch (annotation)
(m) Table listing hunches and metadata

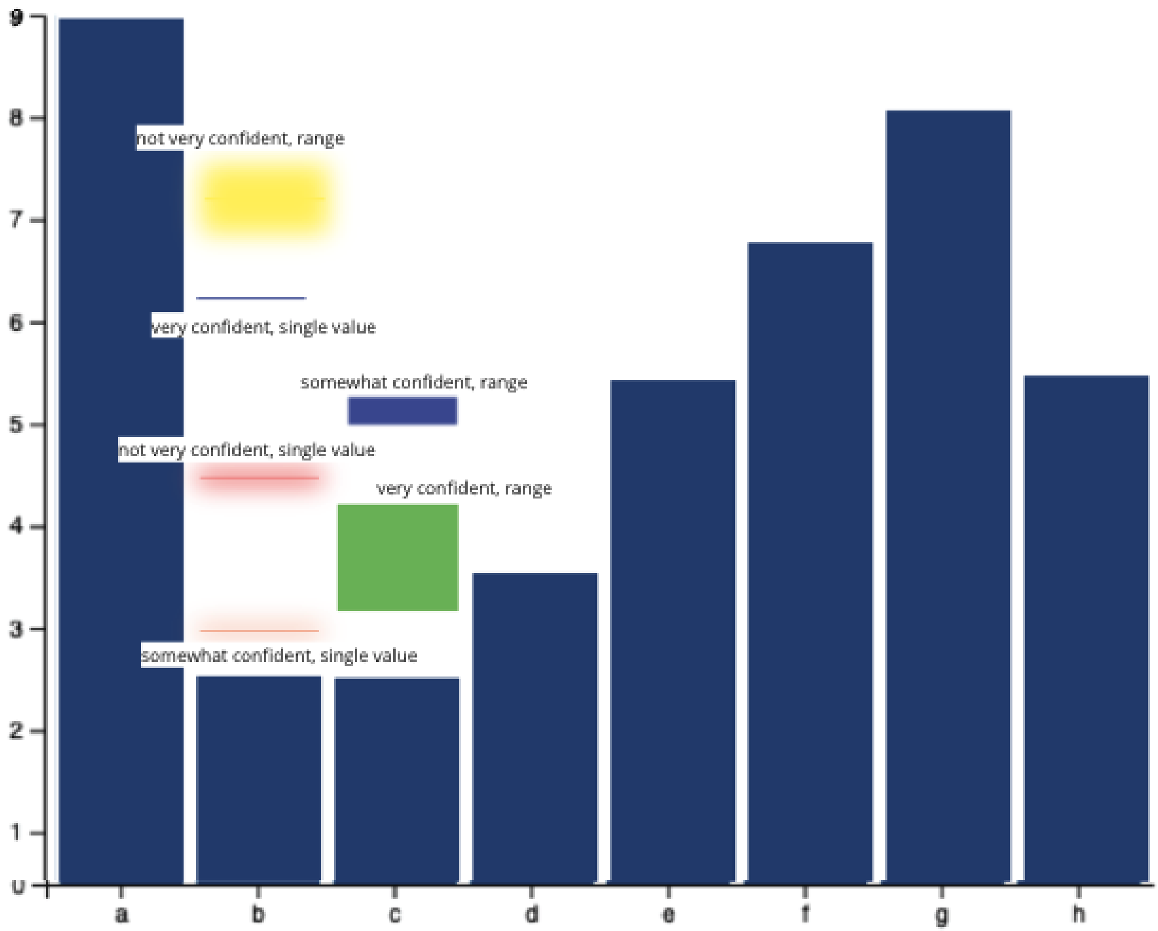


Figure 6.4: Original, non-sketchy design for data hunches. While the data hunches were distinguishable from the original data by color, the distinction was not immediately obvious and could be confused with additional data values that are part of the original data. Hence, we abandoned this design in favor of sketchy renderings.

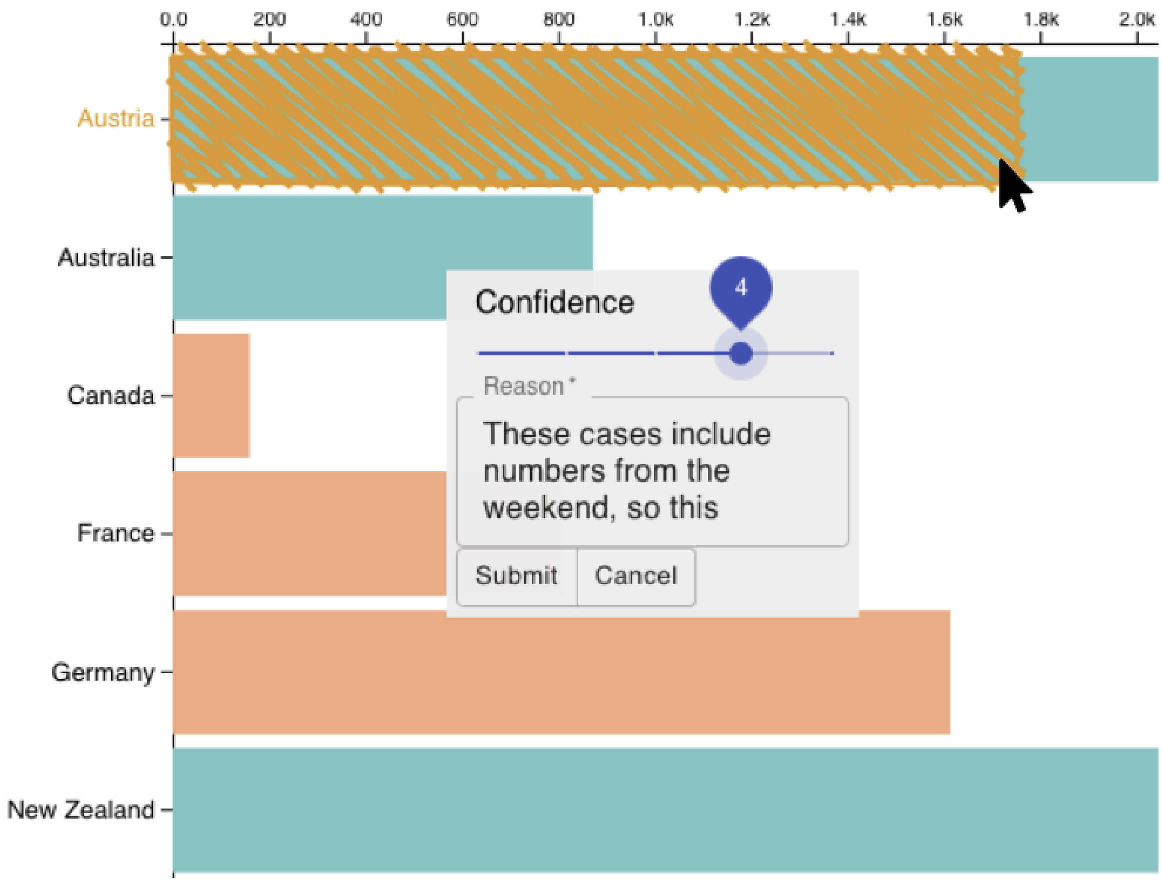


Figure 6.5: Recording a data hunch using direct manipulation of a bar, with a form for confidence and reason.

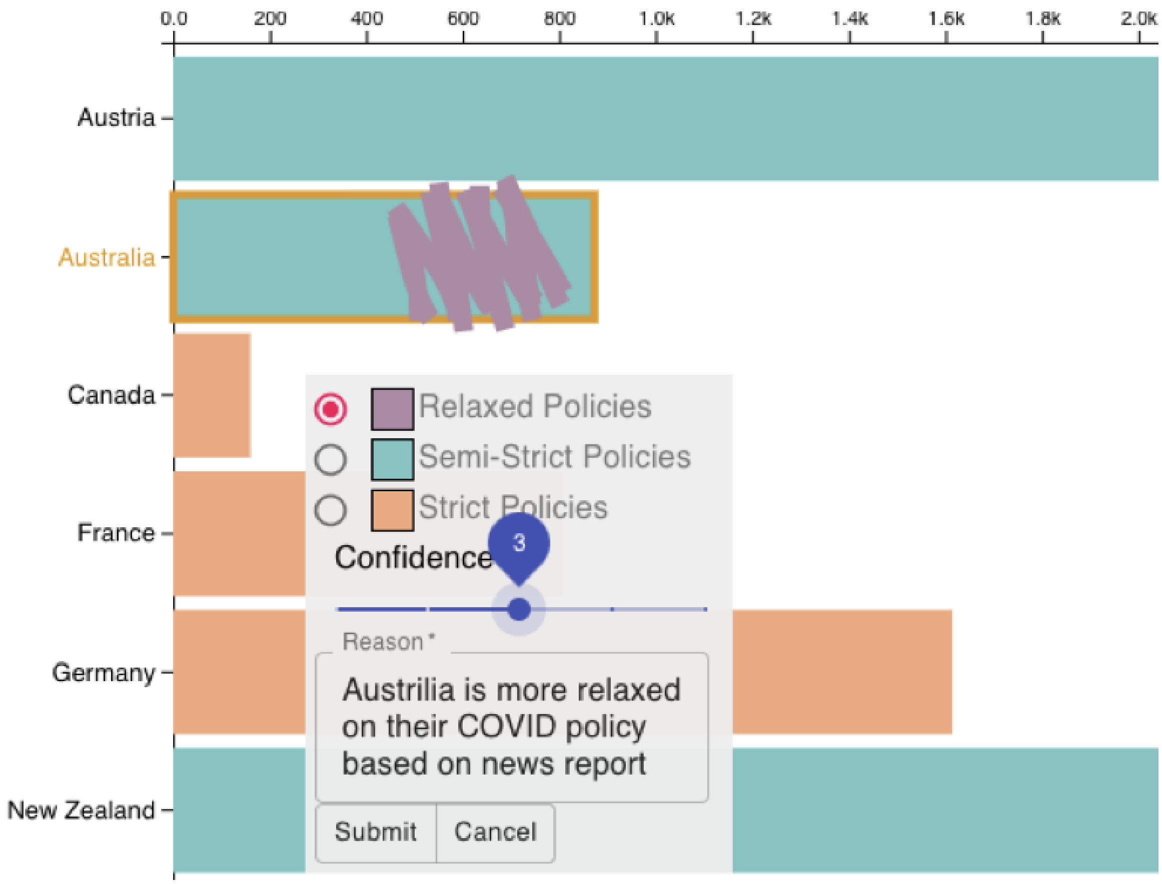


Figure 6.6: Form-based input for a categorical hunch. Selecting an option immediately shows a preview. The input forms are placed close to the selected data point for clear association with the data point.

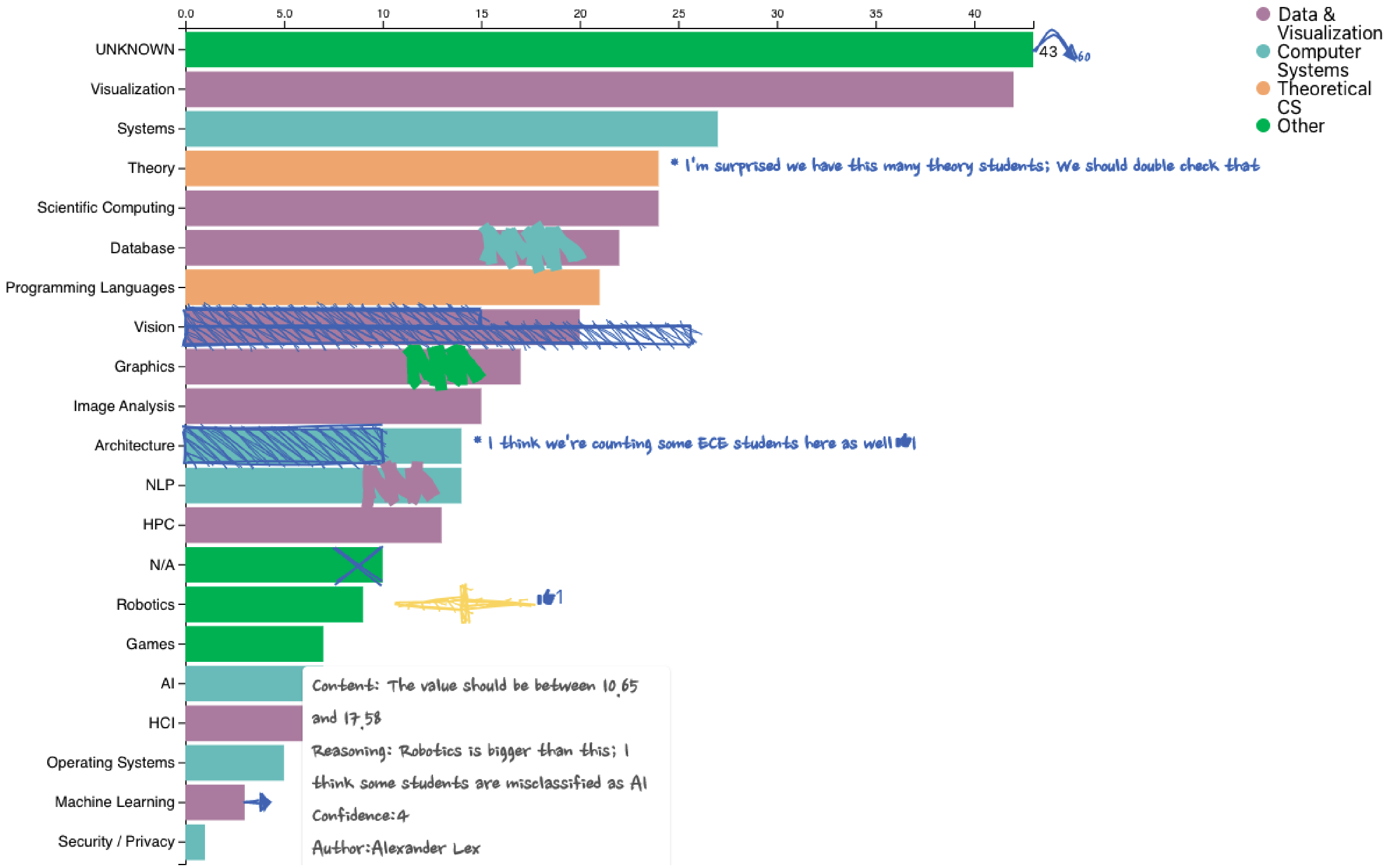


Figure 6.7: A usage scenario showing the data hunches of two faculty members about the size of research areas in a CS department. Demo at https://vdl.sci.utah.edu/data-hunch/?data=student&vol=1

CHAPTER 7

DISCUSSION

Inspired by the breadth of opportunities that data hunches open up, in this chapter, we present a series of discussion threads and future work possibilities on data hunches. Some of these threads reflect on ethical considerations for ensuring data hunches are used in productive and positive ways, and others consider a number of technical and design challenges for consideration. We end the chapter with a discussion on the epistemology of knowledge of data.

7.1 Design Opportunities

The demonstration of some interactive techniques for expressing and communicating data hunches through visualizations seemed to resonate with our interview participants, and they could easily come up with opportunities where using our proposed technique to document data hunches would be beneficial, but the interviews made us doubt that a standalone tool could be successful with the analysts we interviewed. The skepticism about new tools and the fatigue resulting from the fragmented analysis tool space [108] was palpable. Furthermore, our participants did not use the annotation capabilities of tools they already had at their disposal; both Tableau and PowerBI were used by participants, and both support sophisticated annotations.

Consequently, we join previous works [168], [192], [193] in calling for rethinking how we design and develop visualization interfaces, especially when the goal is real-world adoption. Instead of developing yet another tool, we argue for meeting analysts where they are in their analysis workflows. For example, we envision designs that lower friction to annotate and record hunches in the environments that are already being used. At the low-tech end, the interventions could be built-in annotation capabilities on top of screenshots for communication tools, such as Slack, MS Teams, or email. These lightweight methods for capturing hunches could also be designed to support annotations from many people, including field workers and others with close knowledge of the data and its context.

We were surprised that computational notebooks, such as Jupyter Notebooks, were not mentioned once in all our interviews, even though many of the issues discussed by our participants could be addressed using such tools. We also note that only two of our participants reported using programming languages within their analysis—R (P9) and Python (P4). The small usage of programming language and computation notebooks leads us to speculate that the analysis processes of a (possibly large) number of data analysts cannot meet standards for reproducibility laid out by various scientific bodies [194],[195]. We see this outcome as an opportunity for more visual analysis tools and processes to explicitly incorporate ways to capture the ways expert knowledge shapes and impacts analysis processes.

First, we need to make visual analysis tools reproducible and tools to record data hunches intuitive and accessible. The analytic tools our participants use do not support annotated histories or workflows, unlike various research prototypes [89],[125],[196],[197]. For example, there is typically no way to comment on why some data was filtered out in the tools used by our participants. Hence, we call on commercial tool developers to consider making analytical provenance available and salient to their users, and for the scientific community to continue to innovate in that space. Data hunches could be integrated into widely used off-the-shelf visualization libraries. For example, adding the capability to visualize and record data hunches to Excel and Tableau could make data hunches accessible and intuitive to a wide range of audiences.

On the one hand, we also anticipate the usage of data hunches in bespoke systems where the topic of the data shown is of shared interest among larger communities of experts. For example, a recent project elicited feedback from the scientific community on an animation of the SARS-CoV-2 protein structure [198]. Unlike visualization tools designed for an individual research lab, such applications target a wider audience with shared interests, where visualizing data hunches can lead to a deeper impact compared to casual visualizations. As designs for data hunches become more common and libraries to add data hunches to visualization become available, it might be feasible to integrate the recording and visualization of data hunches into a wider set of visualization tools.

Second, even though data hunches are often expressed when viewing data through visualizations, we believe it is important to also capture data hunches at the data level—if data is used as a tool, then the tool needs a manual. As a low-tech intervention, we encourage the extension of metadata files, data dictionaries, or data sheets [49] to document not only the *what* that is in the data, but also the *why* and *how*. Ideally, datasets should be published or archived together with a reproducible analysis story that makes it clear how the data was used. We realize that maintaining metadata is tedious, but we see this as an opportunity for new innovations. It is essential that tools such as Excel provide better support for clear and visible annotations about the data, without forcing users to destroy the clean structure of their data.

Third, with the existing practices, we argue that we need to develop guidelines and standards for documenting heterogeneous analysis processes, especially those that include interactive tools. These guidelines should detail best practices for acknowledging and capturing analysis steps and externalizing expert knowledge that goes into decision-making. In our interviews, we found that external pressures and established guidelines lead to better documentation practices. Promoting better guidelines can arrive at better documenting data hunches, and thus, having more efficient knowledge exchange.

7.2 Data Hunches as Structured Data

Depending on the type of hunch and the method for recording it, data hunches can also become structured data. Value hunches recorded via data space or direct manipulation, for example, are either directly available in the same space as the original data or can be easily translated back into data space. They are, hence, different from e.g., annotations provided on top of a figure, as they can be reused whenever a dataset is reused. For example, if a data hunch is recorded for a visualization of a dataset in a Jupyter Notebook, the data hunch could be shown not only in that one visualization, but could also be propagated to all subsequent and prior visualizations that are based on that dataset, thereby surfacing the hunch at all stages of the analysis process. Data hunches could also be preserved as datasets are updated, as long as the structure is preserved. For example, if a dataset is refined over time, possibly because of discrepancies recorded as data hunches, a new version of a dataset could be overlaid with data hunches recorded for the old version of the dataset, to see whether the hunches expressed still apply. Overlaying data hunches could be combined with an explicit comparison of datasets [89]. However, such a workflow would incur additional visual complexity and hence would require dedicated methods to manage that complexity.

7.3 Potential for Harm

In our advocacy for data hunches, we focus only on the use cases for analysts with rich knowledge about their field. Narrowing the target audience ensures that the data hunches are based on analysts' knowledge and experience of the field, and can provide a richer view of the phenomenon that the data represents. This limitation can also avoid misuse and misinformation in data hunches. Our view of data hunches implicitly assumes that all users are positively contributing to the visualizations when expressing their opinions and knowledge of the data. However, even with good intentions, there is a risk that data hunches could be used to explain away inconvenient data points or to reinforce an analyst's preconceived ideas. One interview participant (P10), a politician, was particularly aware of this potential risk. Her experience with people using data for their own agenda made us suspect that data hunches can have similar issues. Because of such risk, it is critical that data hunches are treated fundamentally differently from data, even if they are recorded in the same space, and that the data hunch consumers are made well aware of the difference between the data and the hunches. Designers that consider incorporating data hunches in their works should go to great lengths to avoid any confusion between the data and the data hunch. Such risk also makes it important that data hunches come with explanations and justifications. Using these techniques, analysts can evaluate a data hunch holistically and judge whether it is reasonable.

Much remains to be explored when considering how to support the public with the ability to record data hunches on visualizations. Data hunches can encourage open conversations about data and visualizations, similar to current online Q&A forums. Previous research concluded that identity-based trust, social feedback, and exposure all have positive effects on knowledge contribution [199]. Visualizations supporting data hunches are similar to online forums, where users can communicate their knowledge or opinions about the data through a set of techniques and receive feedback (such as upvotes, downvotes, and comments) from other users in the community, and the association with identity, feedback, and exposure can positively promote conversation and knowledge sharing about the data. P10, particularly, believed that techniques like data hunches would invite more conversations among people with opposing political affiliations. However, what if a hunch is wrong? Or worse, what if a hunch is maliciously intended as misinformation? Previous research has discovered that data can be used in contradicting ways, depending on how people understand the phenomenon [130], with the aid of visualizations [200]. It is possible that a system that supports the recording and visualization of data hunches can be used with malicious intent to fulfill a personal agenda. Moderating, allowing voting, and providing a reporting mechanism can potentially help with the issue, but it remains an open question that requires further investigation.

7.4 Trust, Privacy, and Biases

Even though we suggest that identities behind data hunches can play a big role in building trust in data hunches, **privacy** can become an issue and concern for analysts. In some settings, the politics of an organization or field could cause vulnerable people to remain silent about contradictory hunches, depriving others of important perspectives [201]. A similar situation was observed in our interviews, where P13 reported that their supervisor rejected data that contradicted their hypothesis. It can be argued that because data hunches are less concrete than a value point in the actual data, the likelihood of dismissing contradictory hunches is higher. On the other hand, anonymity can be equally caustic by invoking negative behavior toward others with opposite views [202]. How to ensure the value, credibility, trustworthiness, and transparency of data hunches is an important, yet open, question.

Data hunches may **reinforce biases** in applications. We envision that data hunches can enrich our comprehension of information, with all involved parties contributing positively based on their expertise and insights. Nevertheless, data hunches might be wielded to explain away inconvenient data points or bolster an analyst's preconceived notions. The occurrence of overlapping data hunches, rooted not in a true understanding of the data but rather in inherent biases against the subject matter or the data source, adds another layer of complexity. In these scenarios, the presence of such biases is compounded when other analysts encounter and subsequently base their interpretations on these potentially skewed data hunches, leading to an unjustified outcome based on the data [203]. To facilitate the identification of biases in the human-in-the-loop analysis, Wall et al. [204] proposed a quantification framework to indicate any underlying biases in the analysis. Such techniques may help analysts judge the potential unjust bias behind a data hunch. Presently, our design framework mandates the inclusion of explanations and justifications for each recorded data hunch. By providing reasoning and contextual information for these hunches, analysts are better equipped to assess them from a comprehensive standpoint. P10 was particularly aware of the bias that analysts may have due to their political background and analytical experiences. For example, a professor in economics and a policy analyst at a taxpayer association may have conflicting data hunches about a tax policy outcome. As a result, she discussed the importance of utilizing identities and reasons behind a data hunch to make a comprehensive interpretation. Nonetheless, the question of how biases can influence the interpretation of a data hunch remains an open avenue for further investigation.

Another interesting, open question is what happens to someone's **trust** in a visualization and the underlying data when data hunches are communicated in a tool. Previous work [205] has reported that *social information* can affect a user's trust and recall about the data visualization. We argue in this dissertation that data is an imperfect representation of reality, and making that imperfection visible is one goal of our work. However, if data hunches make people less trusting, will designers avoid including them, as they sometimes do with uncertainty [22]? Several study participants (P5, P7, P11, P13) agreed that data hunch techniques would be more suitable for discussion and knowledge sharing within the core team, rather than for presentations and dissemination. Therefore, the tension between presenting data caveats to provide context and the fear of confusing the audience remains an open question. A reader may trust the visualization more when data hunches are provided by experts, or when data hunches are highly rated. On the other hand, if too many data hunches disagree with the original data, the reader may trust the source of the visualization less. In the end, the goal of conceptualizing data hunches and proposing a design space for them is to formally recognize the role of personal knowledge in understanding data and empower users to express their views. Designers should fully consider the possible impacts of data hunches before committing to including or excluding them. To begin with, a designer can embrace a layered approach, maintaining the unchanged appearance of visualizations upfront, while granting readers the option to access a data hunch when they choose to do so. A reader can then offer feedback on the provided data hunch or even contribute their own. This approach allows the visualization designer to acquire more insights into the influence of a data hunch prior to full adoption.

7.5 Broader Applications

As a proof of concept, we used a bar chart to demonstrate how data hunches can be implemented in data visualizations. As we developed the prototype, we realized how quickly the additional layers of data hunches can complicate the visualizations. It was challenging to add data hunches in the chart while keeping the original visualization legible. For data hunches to become established, we need to **develop designs for a wide range of visualization types**. We believe that our guidelines apply broadly, but the specific design decisions in our prototype might not easily translate to certain types of space-filling visualizations, such as pie charts, treemaps, or icicle plots. Also, although our framework allows for structural hunches through inclusion and exclusion, tabular data is a simple case compared to network data, which has much more complex structural (topological) information. Hence, we believe that a good amount of design work remains to be done to make data hunches work well with such datasets.

The current scope of our explorations of data hunch is to use data hunches to express knowledge about the data on data visualizations to enhance data analysis and interpretations. Nonetheless, we believe that broader applications for data hunches remain to be explored. For example, considering the landscape of artificial intelligence and machine learning models, we wonder how the concept of data hunches can potentially benefit these areas. With transparency in AI and machine learning models gaining prominence, the drive to make these models interpretable and explainable has led to significant research endeavors [206]–[208]. Here, the notion of data hunches emerges as a potential tool to guide machine learning models and provide explainability. If a data hunch is expressed in the training data space that a machine learning model is using, users could employ their data hunches to influence the model's behavior in generating predictions that are more contextually aware. The success of such an application depends on a data hunch being accompanied by not only textual reasoning and context, but also more quantitative metrics such as suggested values and confidence levels. Implementing these metrics may ensure that a data hunch is not only reasoned and understandable, but also actionable in a machine-learning model. Additionally, if the machine-learning model is attuned to the presence of data hunches and, accordingly, makes predictions that incorporate a data hunch, the predictions can be shown with the data hunch, or even shown in a similar fashion to a data hunch. However, the issue of bias and trust discussed above still remains with these broader applications, and much more research effort to address these open questions is required.

CHAPTER 8

CONCLUSION

In this chapter, we reiterate the contributions of this dissertation, which arise from the interview study and reflective analysis of personal knowledge in data analysis and formative design studies. We conclude with interesting areas for future research into data hunches.

8.1 Summary

In this work, we presented several research projects that contributed to the topic of personal knowledge for data analysis. We first presented the formative design study Sanguine [1], in which we designed and implemented an interactive visual analytic tool for studying clinical transfusions. Based on the feedback from the design study, we learned that experts combine their own experience and knowledge when they interpret data. After the formative design study, we interviewed 14 analysts and experts on how they think about the data they work with and the strategies they use to incorporate experts' knowledge to fill in the gaps between data and reality [3]. Through the interviews, we provided abundant evidence that experts' knowledge is critical in data analyses. We also found that the technical support for recording and supporting knowledge exchange in analyses is currently lacking. We reflected on the design study and interview results, formalized the experts' knowledge about data as data hunches [2], and analyzed the implications of supporting data hunches in data analysis. We proposed techniques for recording and communicating data hunches in data visualizations, listed design guidelines, and implemented a prototype to demonstrate data hunches in action. The goal of this dissertation was to formalize and recognize the significant role that personal knowledge has in understanding data, which many works overlook, and elevate this personal knowledge into another form of information that can be explicitly recorded and utilized.

To summarize, this dissertation presented three contributions to the topic of personal knowledge of data in analysis. First, we presented an analysis of interview studies with analysts from a wide range of domains and with varied expertise and experience inquiring about the role of contextual knowledge and the process of incorporating various sources of knowledge into analyses. We then defined and characterized experts' knowledge about data and data as *data hunches*, and position data hunches in the existing understanding of uncertainty. Lastly, we proposed a framework and guidelines to design visualizations that support recording and communicating data hunches through visualizations intuitively and effectively. Ultimately, we seek to question the notion of data being the gold standard of representing phenomena in the world, and open up the potential to grow visualization research beyond constrained notions of data.

8.2 Future Work

In this dissertation, we presented the initial steps in delving into a vast realm that examines the inclusion of personal knowledge about data in visualizations, addressing the how, why, and when aspects of including personal knowledge explicitly. Although we acknowledge and emphasize the significance of data hunches, our primary focus has been on their application in professional settings between analysts and experts. As discussed in Section 7.3, data hunches can foster discussions about data but also possess the potential for misuse to serve personal agendas. Consequently, it is imperative to investigate how data hunches can positively contribute to the advancement of data literacy and facilitate productive data conversations.

An intriguing avenue for further research lies in the integration of support for data hunches into existing analysis tools employed by analysts, such as Tableau, PowerBI, and computation notebooks such as Jupyter. Our study participants did not extensively utilize the available annotation features in these tools. They might not be aware of the available features, or they did not find the existing features to fully support their needs. In our interviews, our participants expressed a desire for improved documentation. Thus, an open question emerges regarding how analysts can better be supported in this regard. Ideally, data hunches should be an integral part of analysts' daily routines and are systematically recorded alongside the relevant data operations. For instance, when analysts perform actions such as filtering or manipulating data, their hunches should be prominently shown and allowed to evolve as the data is refined. Additionally, more instruction on the available features might be beneficial. Achieving this goal requires further research and exploration into effective strategies for seamlessly achieving an ecosystem of data hunches and data.

There is a design opportunity to create and implement a library, such as a Python library, dedicated to facilitating adding data hunches non-invasively on top of notebooks and existing visualization libraries. By providing a streamlined and standardized approach to capturing and integrating data hunches, the library could significantly enhance the effectiveness and efficiency of the analytical process. Since data hunches exist in data space, we believe that most data hunches can be propagated into subsequent plots, such as in subsequent Jupyter notebook charts. With these potential features, data hunches can live in parallel with the data, not just in a single plot. Hence we could surface the hunch at all stages of the analysis process. By providing a streamlined and standardized approach to capturing and integrating data hunches, the library could significantly enhance the effectiveness and efficiency of the analytical process. Moreover, future research could explore the potential integration of data hunches into machine learning algorithms. These could make machine learning outputs more context-aware and provide more comprehensive predictions and suggestions. However, such a workflow would incur additional visual and computational complexity and hence would require dedicated methods to manage that complexity.

Furthermore, our prototype exclusively explored bar charts, leaving numerous other chart types awaiting exploration. Exploring more chart types, such as line charts and scatterplots, can expand our understanding of how visual analytic tools can use data hunch techniques to support analysts. In conclusion, there are numerous exciting avenues for future research to further enhance the integration of data hunches into existing analysis tools and visual applications, ultimately empowering analysts to leverage domain knowledge more effectively, enabling more informed data-driven decision-making, and fostering a deeper understanding of data.

REFERENCES

[1] H. Lin, R. A. Metcalf, J. Wilburn, and A. Lex, "Sanguine: Visual analysis for patient blood management," *Inf. Vis.*, vol. 20, no. 2–3, pp. 123–137, Aug. 2021.

[2] 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, Sep. 2022.

[3] H. Lin, M. Lisnic, D. Akbaba, M. Meyer, and A. Lex, "Here's what you need to know about my data: Exploring expert knowledge's role in data analysis," 2023, *OSF Preprints*. [Online]. Available: https://doi.org/10.31219/osf.io/dn32z.

[4] D. Hoffman, "'I had a funny feeling in my gut'," *Washington Post*, p. A19, Feb. 1999. [Online]. Available: https://www.washingtonpost.com/wp-srv/inatl/longterm/coldwar/soviet10.htm.

[5] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big Data*, vol. 1, no. 1, pp. 51–59, Mar. 2013.

[6] J. A. Marsh, J. Sloan McCombs, and F. Martorell, "How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools," *Educ. Policy*, vol. 24, no. 6, pp. 872–907, Nov. 2010.

[7] J. Cox and M. Mann, "Quantitative, high-resolution proteomics for data-driven systems biology," *Annu. Rev. Biochem.*, vol. 80, no. 1, pp. 273–299, Jul. 2011.

[8] E. Brynjolfsson and K. McElheran, "The rapid adoption of data-driven decision-making," *Am. Econ. Rev.*, vol. 106, no. 5, pp. 133–139, May 2016.

[9] J. Drucker, "Humanities approaches to graphical display," *DHQ*, vol. 5, no. 1, Mar. 2011.

[10] Y. A. Loukissas, *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA, USA: MIT Press, 2019.

[11] N. Mccurdy, J. Gerdes, and M. Meyer, "A framework for externalizing implicit error using visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 925–935, Jan. 2019.

[12] C. D'Ignazio and L. F. Klein, *Data Feminism*. Cambridge, MA, USA: MIT Press, 2020.

[13] Y.-S. Kim, K. Reinecke, and J. Hullman, "Explaining the gap: Visualizing one's predictions improves recall and comprehension of data," in *Proc. 2017 CHI Conf. Hum. Factors Comput. Syst.*, May 6–11, 2017, pp. 1375–1386.

[14] J. Heer, F. B. Viégas, and M. Wattenberg, "Voyagers and voyeurs: Supporting asynchronous collaborative visualization," *Commun. ACM*, vol. 52, no. 1, pp. 87–97, Jan. 2009.

[15] S. Nowak and B. Lyn, “I’m not sure: Designing for ambiguity in visual analytics,” in *Proc. Graph. Interface*, May 16–19, 2022, pp. 114–126.

[16] G. Panagiotidou, R. Vandam, J. Poblome, and A. Vande Moere, “Implicit error, uncertainty and confidence in visualization: An archaeological case study,” *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 12, pp. 4389–4402, Dec. 2022.

[17] L. M. K. Padilla, M. Powell, M. Kay, and J. Hullman, “Uncertain about uncertainty: How qualitative expressions of forecaster confidence impact decision-making with uncertainty visualizations,” *Front. Psychol.*, vol. 11, Art. no. 579267, Jan. 2021.

[18] K. Potter, P. Rosen, and C. R. Johnson, “From quantification to visualization: A taxonomy of uncertainty visualization approaches,” in *10th IFIP WG 2.5 Working Conf. WoCoUQ*, Aug. 1–4, 2012, pp. 226–249.

[19] H. Lin, C. Nobre, A. Bakian, and A. Lex, “Clipped graphs: A compact time-series encoding,” in *Poster Proc. IEEE Vis. Conf.*, Oct. 20–25, 2019, p. 2.

[20] H. Lin, “Utah CS,” https://haihan-lin.github.io/Utah-CS/ (accessed Apr. 11, 2023).

[21] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz, “Overview and state-of-the-art of uncertainty visualization,” in *Scientific Visualization*, C. D. Hansen, M. Chen, C. R. Johnson, A. E. Kaufman, and H. Hagen, Eds. London, UK: Springer, 2014, pp. 3–27.

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

[23] S. Joslyn and S. Savelli, “Communicating forecast uncertainty: Public perception of weather forecast uncertainty,” *Meteorol. Appl.*, vol. 17, no. 2, pp. 180–195, Jun. 2010.

[24] M. Williams, D. Cornford, L. Bastin, R. Jones, and S. Parker, “Automatic processing, quality assurance and serving of real-time weather data,” *Comput. Geosci.*, vol. 37, no. 3, pp. 353–362, Mar. 2011.

[25] G. Fusco, M. Caglioni, K. Emsellem, M. Merad, D. Moreno, and C. Voiron-Canicio, “Questions of uncertainty in geography,” *Environ. Plan A*, vol. 49, no. 10, pp. 2261–2280, Oct. 2017.

[26] J. D. Brown, “Knowledge, uncertainty and physical geography: Towards the development of methodologies for questioning belief,” *Trans. Inst. Br. Geogr.*, vol. 29, no. 3, pp. 367–381, Sep. 2004.

[27] M. Webster, “Communicating climate change uncertainty to policy-makers and the public,” *Clim. Change*, vol. 61, pp. 1–8, Nov. 2003.

[28] M. Booshehrian, T. Möller, R. M. Peterman, and T. Munzner, “Vismon: Facilitating analysis of trade-offs, uncertainty, and sensitivity in fisheries management decision making,” *Comput. Graph. Forum*, vol. 31, no. 3pt3, pp. 1235–1244, Jun. 2012.

[29] U. Bhatt, J. Antorán, Y. Zhang, Q. V. Liao, P. Sattigeri, R. Fogliato, G. Melançon, R. Krishnan, J. Stanley, O. Tickoo, L. Nachman, R. Chunara, M. Srikumar, A. Weller, and A. Xiang, “Uncertainty as a form of transparency: Measuring, communicating,

and using uncertainty," in *Proc. 2021 AAAIACM Conf. AI Ethics Soc.*, May 19–21, 2021, pp. 401–413.

[30] W. E. Walker, P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. K. von Krauss, "Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support," *Integr. Assess.*, vol. 4, no. 1, pp. 5–17, Mar. 2003.

[31] J. Ravetz and S. Funtowicz, *Uncertainty and Quality in Science for Policy*, 1st ed. Berlin, Germany: Springer, 1990.

[32] B. A. Covitt and C. W. Anderson, "Untangling trustworthiness and uncertainty in science," *Sci. Educ.*, vol. 31, no. 5, pp. 1155–1180, Feb. 2022.

[33] N. Boukhelifa, M.-E. Perrin, S. Huron, and J. Eagan, "How data workers cope with uncertainty: A task characterisation study," in *Proc. 2017 CHI Conf. Hum. Factors Comput. Syst.*, May 6–11, 2017, pp. 3645–3656.

[34] K. Brodlie, R. Allendes Osorio, and A. Lopes, "A review of uncertainty in data visualization," in *Expanding the Frontiers of Visual Analytics and Visualization*, J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, Eds. London, UK: Springer, 2012, pp. 81–109.

[35] B. Karer, H. Hagen, and D. J. Lehmann, "Insight beyond numbers: The impact of qualitative factors on visual data analysis," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 1011–1021, Feb. 2021.

[36] R. W. Katz, P. F. Craigmile, P. Guttorp, M. Haran, B. Sansó, and M. L. Stein, "Uncertainty analysis in climate change assessments," *Nature Clim. Change*, vol. 3, no. 9, pp. 769–771, Sep. 2013.

[37] W. D. Rowe, "Understanding uncertainty," *Risk Anal.*, vol. 14, no. 5, pp. 743–750, Oct. 1994.

[38] L. M. K. Padilla, S. H. Creem-Regehr, and W. Thompson, "The powerful influence of marks: Visual and knowledge-driven processing in hurricane track displays," *J. Exp. Psychol. Appl.*, vol. 26, no. 1, pp. 1–16, Mar. 2020.

[39] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman, "The science of visual data communication: What works," *Psychol Sci. Public Interest*, vol. 22, no. 3, pp. 110–161, Dec. 2021.

[40] C. R. Johnson and A. R. Sanderson, "A next step: Visualizing errors and uncertainty," *IEEE Comput. Grap. Appl.*, vol. 23, no. 5, pp. 6–10, Sep. 2003.

[41] C. M. Wittenbrink, A. T. Pang, and S. K. Lodha, "Glyphs for visualizing uncertainty in vector fields," *IEEE Trans. Vis. Comput. Graph.*, vol. 2, no. 3, pp. 266–279, Sep. 1996.

[42] K. Potter, J. Kniss, R. Riesenfeld, and C. Johnson, "Visualizing summary statistics and uncertainty," *Comput. Graph. Forum*, vol. 29, no. 3, pp. 823–832, Aug. 2010.

[43] M. Correll and M. Gleicher, "Error bars considered harmful: Exploring alternate encodings for mean and error," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2142–2151, Dec. 2014.

[44] K. Potter, R. M. Kirby, D. Xiu, and C. R. Johnson, "Interactive visualization of probability and cumulative density functions," *Int. J. Uncertain. Quantif.*, vol. 2, no. 4, pp. 397–412, Mar. 2012.

[45] J. Hullman, P. Resnick, and E. Adar, "Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering," *PLoS One*, vol. 10, no. 11, Art. no. e0142444, Nov. 2015.

[46] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay, "In pursuit of error: A survey of uncertainty visualization evaluation," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 903–913, Jan. 2019.

[47] A. M. van der Bles, S. van der Linden, A. L. J. Freeman, J. Mitchell, A. B. Galvao, L. Zaval, and D. J. Spiegelhalter, "Communicating uncertainty about facts, numbers and science," *R. Soc. Open Sci.*, vol. 6, no. 5, Art. no. 181870, May 2019.

[48] N. Boukhelifa, A. Bezerianos, T. Isenberg, and J. Fekete, "Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2769–2778, Dec. 2012.

[49] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. Daumé III, and K. Crawford, "Datasheets for datasets," *Commun. ACM*, vol. 64, no. 12, pp. 86–92, Dec. 2021.

[50] M. Arnold, R. K. E. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilović, R. Nair, K. N. Ramamurthy, A. Olteanu, D. Piorkowski, D. Reimer, J. Richards, J. Tsay, and K. R. Varshney, "FactSheets: Increasing trust in AI services through supplier's declarations of conformity," *IBM J. Res. Dev.*, vol. 63, no. 4/5, pp. 6:1–6:13, Jul. 2019.

[51] M. Franke, R. Barczok, S. Koch, and D. Weltecke, "Confidence as first-class attribute in digital humanities data," in *Proc. 4th VIS4DH Workshop*, Oct. 20, 2019, pp. 1–5.

[52] J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel, "A typology for visualizing uncertainty," in *Visualization and Data Analysis*, R. F. Erbacher, J. C. Roberts, M. T. Grohn, and K. Bomer, Eds., vol. 5669. Bellingham, WA, USA: SPIE, 2005, pp. 146–157.

[53] Y. Chen, S. Barlowe, and J. Yang, "Click2Annotate: Automated insight externalization with rich semantics," in *2010 IEEE Symp. Vis. Anal. Sci. Technol.*, Oct. 25–26, 2010, pp. 155–162.

[54] Y.-S. Kim, N. Henry Riche, B. Lee, M. Brehmer, M. Pahud, K. Hinckley, and J. Hullman, "Inking your insights: Investigating digital externalization behaviors during data analysis," in *Proc. 2019 ACM Int. Conf. Interact. Surf. Spaces*, Nov. 10–13, 2019, pp. 255–267.

[55] J. Moore, P. Goffin, M. Meyer, P. Lundrigan, N. Patwari, K. Sward, and J. Wiese, "Managing in-home environments through sensing, annotating, and visualizing air quality data," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 128:1–128:28, Sep. 2018.

[56] A. Mathisen, T. Horak, C. N. Klokmose, K. Grønbæk, and N. Elmqvist, “InsideInsights: Integrating data-driven reporting in collaborative visual analytics,” *Comput. Graph. Forum*, vol. 38, no. 3, pp. 649–661, Jun. 2019.

[57] J. Zhao, M. Glueck, P. Isenberg, F. Chevalier, and A. Khan, “Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs,” *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 340–350, Jan. 2018.

[58] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon, “ManyEyes: A site for visualization at internet scale,” *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 6, pp. 1121–1128, Nov.–Dec. 2007.

[59] A. Chhabra, S. R. S. Iyengar, P. Saini, and R. S. Bhat, “Presence of an ecosystem: A catalyst in the knowledge building process in crowdsourced annotation environments,” in *Proc. 2015 IEEEACM Int. Conf. Adv. Soc. Netw. Anal. Min.*, Aug. 25–28, 2015, pp. 286–291.

[60] A. Burns and N. Kulish, “Bloomberg’s billions: How the candidate built an empire of influence,” *New York Times*, Feb. 2020. [Online]. Available: https://www.nytimes.com/interactive/2020/02/15/us/politics/michael-bloomberg-spending.html.

[61] Q. Liu, C. Chen, E. Shen, F. Zhao, Z. Sun, and J. Wu, “Detection, annotation and visualization of alternative splicing from RNA-seq data with SplicingViewer,” *Genomics*, vol. 99, no. 3, pp. 178–182, Mar. 2012.

[62] N. Goyal, G. Leshed, and S. R. Fussell, “Effects of visualization and note-taking on sensemaking and analysis,” in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 27–May 2, 2013, pp. 2721–2724.

[63] J. Hullman, N. Diakopoulos, and E. Adar, “Contextifier: Automatic generation of annotated stock visualizations,” in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 27–May 2, 2013, pp. 2707–2716.

[64] G. Aisch and A. Cox, “A 3-D view of a chart that predicts the economic future: The yield curve,” *New York Times*, Mar. 2015. [Online]. Available: https://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html.

[65] C. Lai, Z. Lin, R. Jiang, Y. Han, C. Liu, and X. Yuan, “Automatic annotation synchronizing with textual description for visualization,” in *Proc. 2020 CHI Conf. Hum. Factors Comput. Syst.*, Apr. 25–30, 2020, pp. 1–13.

[66] M. Mărăşoiu, A. F. Blackwell, A. Sarkar, and M. Spott, “Clarifying hypotheses by sketching data,” in *Proc. Eurograph./IEEE VGTC Conf. Vis. Short Papers*, Jun. 6–10, 2016, pp. 125–129.

[67] D. Ren, M. Brehmer, Bongshin Lee, T. Höllerer, and E. K. Choe, “ChartAccent: Annotation for data-driven storytelling,” in *2017 IEEE Pac. Vis. Symp. PacificVis*, Apr. 18–21, 2017, pp. 230–239.

[68] H. Romat, N. Henry Riche, K. Hinckley, B. Lee, C. Appert, E. Pietriga, and C. Collins, "ActiveInk: (Th)inking with data," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 4–9, 2019, pp. 1–13.

[69] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer, "Enterprise data analysis and visualization: An interview study," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2917–2926, Dec. 2012.

[70] J. MacNeil and R. Falconer, "When learning the hard way makes learning easy: Building better lab note-taking skills," *J. Chem. Educ.*, vol. 87, no. 7, pp. 703–704, Jul. 2010.

[71] M. Rubacha, A. K. Rattan, and S. C. Hosselet, "A review of electronic laboratory notebooks available in the market today," *J. Lab. Autom.*, vol. 16, no. 1, pp. 90–98, Feb. 2011.

[72] P. J. Guo and M. Seltzer, "BURRITO: Wrapping your lab notebook in computational infrastructure," in *Proc. 4th USENIX Conf. Theory Pract. Proven.*, Jun. 14–15, 2012, p. 7.

[73] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. Silva, and H. T. Vo., "Visualization in radiation oncology: Towards replacing the laboratory notebook," SCI Inst., Salt Lake City, UT, USA, Tech. Rep. UUSCI-2006-17, Apr. 2006.

[74] A. Tabard, W. E. Mackay, and E. Eastmond, "From individual to collaborative: The evolution of prism, a hybrid laboratory notebook," in *Proc. 2008 ACM Conf. Comput. Support. Coop. Work*, Nov. 8–12, 2008, pp. 569–578.

[75] C. N. Klokmose and P.-O. Zander, "Rethinking laboratory notebooks," in *Proc. 9th Int. Conf. Des. Coop. Syst.*, May 18–21, 2010, pp. 119–139.

[76] R. H. Kallet, "How to write the methods section of a research paper," *Respir. Care*, vol. 49, no. 10, pp. 1229–1232, Oct. 2004.

[77] V. Siegel, "Reproducibility in research," *Dis. Model. Mech.*, vol. 4, no. 3, pp. 279–280, May 2011.

[78] I. Bruce, "Cognitive genre structures in methods sections of research articles: A corpus study," *J. English Acad. Purp.*, vol. 7, no. 1, pp. 38–54, Jan. 2008.

[79] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. E. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. B. Hamrick, J. Grout, S. Corlay *et al.*, "Jupyter notebooks—A publishing format for reproducible computational workflows," in *Proc. 20th Int. Conf. Electron. Publ.*, Jun. 7–9, 2016, pp. 87–90.

[80] B. Baumer and D. Udwin, "R Markdown," *WIREs Comput. Stat.*, vol. 7, no. 3, pp. 167–177, May–Jun. 2015.

[81] Y. Xie, J. J. Allaire, and G. Grolemund, *R Markdown: The Definitive Guide*. Boca Raton, FL, USA: CRC Press, 2018.

[82] S. Chattopadhyay, I. Prasad, A. Z. Henley, A. Sarma, and T. Barik, "What's wrong with computational notebooks? Pain points, needs, and design opportunities," in *Proc. 2020 CHI Conf. Hum. Factors Comput. Syst.*, Apr. 25–30, 2020, pp. 1–12.

[83] A. Rule, A. Tabard, and J. D. Hollan, "Exploration and explanation in computational notebooks," in *Proc. 2018 CHI Conf. Hum. Factors Comput. Syst.*, Apr. 21–26, 2018, pp. 1–12.

[84] A. Y. Wang, A. Mittal, C. Brooks, and S. Oney, "How data scientists use computational notebooks for real-time collaboration," *Proc. ACM Hum. Comput. Interact.*, vol. 3, no. CSCW, pp. 39:1–39:30, Nov. 2019.

[85] Alena Guzharina, "We downloaded 10,000,000 Jupyter notebooks from Github–This is what we learned," *JetBrains Datalore Blog*, Dec. 2020. [Online]. Available: https://blog.jetbrains.com/datalore/2020/12/17/we-downloaded-10-000-000-jupyter-notebooks-from-github-this-is-what-we-learned/.

[86] J. Wenskovitch, J. Zhao, S. Carter, M. Cooper, and C. North, "Albireo: An interactive tool for visually summarizing computational notebook structure," in *2019 IEEE Vis. Data Sci. VDS*, Oct. 20, 2019, pp. 1–10.

[87] Y. Wu, J. M. Hellerstein, and A. Satyanarayan, "B2: Bridging code and interactive visualization in computational notebooks," in *Proc. 33rd Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 20–23, 2020, pp. 152–165.

[88] M. B. Kery and B. A. Myers, "Interactions for untangling messy history in a computational notebook," in *2018 IEEE Symp. Vis. Lang. Hum.Centric Comput. VL/HCC*, Oct. 1–4, 2018, pp. 147–155.

[89] K. Gadhave, Z. Cutler, and A. Lex, "Reusing interactive analysis workflows," *Comput. Graph. Forum*, vol. 41, no. 3, pp. 133–144, Jun. 2022.

[90] A. J. Gilliland, "Setting the stage," in *Introduction to Metadata*, 3rd ed., M. Baca, Ed. Los Angeles, CA, USA: Getty Research Institute, 2016.

[91] S. S. Lazinger, *Digital Preservation and Metadata: History, Theory, Practice*. Englewood, CO, USA: Libraries Unlimited, 2001.

[92] L. Dempsey and R. Heery, "Metadata: A current view of practice and issues," *J. Doc.*, vol. 54, no. 2, pp. 145–172, Jan. 1998.

[93] J. Riley, *Understanding Metadata*. Baltimore, MD, USA: National Information Standards Organization, 2017.

[94] J. Greenberg, "Understanding metadata and metadata schemes," *Cat. Classif. Q.*, vol. 40, no. 3-4, pp. 17–36, Sep. 2005.

[95] M. Jones, C. Berkley, J. Bojilova, and M. Schildhauer, "Managing scientific metadata," *IEEE Internet Comput.*, vol. 5, no. 5, pp. 59–68, Sep. 2001.

[96] R. Devarakonda, B. Shrestha, G. Palanisamy, L. Hook, T. Killeffer, M. Krassovski, T. Boden, R. Cook, L. Zolly, V. Hutchison, M. Frame, A. Cialella, and K. Lazer, "OME: Tool for generating and managing metadata to handle BigData," in *2014 IEEE Int. Conf. Big Data Big Data*, Oct. 27–30, 2014, pp. 8–10.

[97] W. Yan, L. Jiajin, and H. Dongmei, "A metadata management framework for marine information based on XML," in *IET Int. Conf. Inf. Sci. Control Eng. 2012 ICISCE*, Dec. 7–9, 2012, pp. 1–4.

[98] P. Mutton and J. Golbeck, "Visualization of semantic metadata and ontologies," in *Proc. 7th Int. Conf. Inf. Vis. 2003 IV*, Jul. 18, 2003, pp. 300–305.

[99] P. Klein, H. Reiterer, F. Muller, and T. Limbach, "Metadata visualisation with VisMeB," in *Proc. 7th Int. Conf. Inf. Vis. 2003 IV*, Jul. 18, 2003, pp. 600–605.

[100] 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.*, advance online publication, pp. 1–16, Dec. 2022.

[101] R. Gartner, "What metadata is and why it matters," in *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*, R. Gartner, Ed. Cham, Switzerland: Springer International Publishing, 2016, pp. 1–13.

[102] E. Duval and W. Hodgins, "Metadata principles and practicalities," *D-Lib Mag.*, vol. 8, no. 4, Apr. 2002. [Online]. Available: https://www.dlib.org/dlib/april02/weibel/04weibel.html.

[103] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru, "Model cards for model reporting," in *Proc. Conf. Fairness Account. Transpar.*, Jan. 29–31, 2019, pp. 220–229.

[104] E. M. Bender and B. Friedman, "Data statements for natural language processing: Toward mitigating system bias and enabling better science," *Trans. Assoc. Comput. Linguist.*, vol. 6, pp. 587–604, Dec. 2018.

[105] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," *Interactions*, vol. 19, no. 3, pp. 50–59, May–Jun. 2012.

[106] Y. Mao, D. Wang, M. Muller, K. R. Varshney, I. Baldini, C. Dugan, and A. Mojsilović, "How data scientists work together with domain experts in scientific collaborations: To find the right answer or to ask the right question?" *Proc. ACM Hum. Comput. Interact.*, vol. 3, no. GROUP, pp. 237:1–237:23, Dec. 2019.

[107] M. Muller and A. Strohmayer, "Forgetting practices in the data sciences," in *Proc. 2022 CHI Conf. Hum. Factors Comput. Syst.*, Apr. 29–May 5, 2022, pp. 1–19.

[108] S. Alspaugh, N. Zokaei, A. Liu, C. Jin, and M. A. Hearst, "Futzing and moseying: Interviews with professional data analysts on exploration practices," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 22–31, Jan. 2019.

[109] A. Crisan, B. Fiore-Gartland, and M. Tory, "Passing the data baton: A retrospective analysis on data science work and workers," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 1860–1870, Feb. 2021.

[110] Y. Liu, T. Althoff, and J. Heer, "Paths explored, paths omitted, paths obscured: Decision points & selective reporting in end-to-end data analysis," in *Proc. 2020 CHI Conf. Hum. Factors Comput. Syst.*, Apr. 25–30, 2020, pp. 1–14.

[111] M. Rittenbruch, K. Vella, M. Brereton, J. M. Hogan, D. Johnson, J. Heinrich, and S. O'Donoghue, "Collaborative sense-making in genomic research: The role of visualisation," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 12, pp. 4477–4489, Dec. 2022.

[112] R. A. Ruddle, J. Cheshire, and S. J. Fernstad, "Tasks and visualizations used for data profiling: A survey and interview study," *IEEE Trans. Vis. Comput. Graph.*, Advance online publication, pp. 1–12, Jan. 2023.

[113] A. X. Zhang, M. Muller, and D. Wang, "How do data science workers collaborate? Roles, workflows, and tools," *Proc. ACM Hum. Comput. Interact.*, vol. 4, no. CSCW1, pp. 022:1–022:23, May 2020.

[114] J. Y. Jung, T. Steinberger, J. L. King, and M. S. Ackerman, "How domain experts work with data: Situating data science in the practices and settings of craftwork," *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. CSCW1, pp. 58:1–58:29, Apr. 2022.

[115] D. A. Schon, "Designing as reflective conversation with the materials of a design situation," *Knowl. Based Syst.*, vol. 3, no. 3, pp. 131–147, Sep. 1992.

[116] M. Muller, I. Lange, D. Wang, D. Piorkowski, J. Tsay, Q. V. Liao, C. Dugan, and T. Erickson, "How data science workers work with data: Discovery, capture, curation, design, creation," in *Proc. 2019 CHI Conf. Hum. Factors Comput. Syst.*, May 4–9, 2019, pp. 1–15.

[117] M. Tory, L. Bartram, B. Fiore-Gartland, and A. Crisan, "Finding their data voice: Practices and challenges of dashboard users," *IEEE Comput. Grap. Appl.*, vol. 43, no. 1, pp. 22–36, Jan.–Feb. 2022.

[118] M. Skeels, B. Lee, G. Smith, and G. G. Robertson, "Revealing uncertainty for information visualization," *Inf. Vis.*, vol. 9, no. 1, pp. 70–81, Jan. 2010.

[119] M. Feinberg, "A design perspective on data," in *Proc. 2017 CHI Conf. Hum. Factors Comput. Syst.*, May 6–11, 2017, pp. 2952–2963.

[120] M. G. Morgan, "Use (and abuse) of expert elicitation in support of decision making for public policy," *Proc. Natl. Acad. Sci.*, vol. 111, no. 20, pp. 7176–7184, May 2014.

[121] N. Mahyar, A. Sarvghad, and M. Tory, "Note-taking in co-located collaborative visual analytics: Analysis of an observational study," *Inf. Vis.*, vol. 11, no. 3, pp. 190–204, Jul. 2012.

[122] J. Walny, S. Carpendale, N. H. Riche, G. Venolia, and P. Fawcett, "Visual thinking in action: Visualizations as used on whiteboards," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2508–2517, Dec. 2011.

[123] M. Feng, C. Deng, E. M. Peck, and L. Harrison, "HindSight: Encouraging exploration through direct encoding of personal interaction history," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 351–360, Jan. 2017.

[124] Z. T. Cutler, K. Gadhave, and A. Lex, "Trrack: A library for provenance tracking in web-based visualizations," in *IEEE Vis. Conf. VIS*, Oct. 25–30, 2020, pp. 116–120.

[125] K. Gadhave, J. Görtler, Z. Cutler, C. Nobre, O. Deussen, M. Meyer, J. M. Phillips, and A. Lex, "Predicting intent behind selections in scatterplot visualizations," *Inf. Vis.*, vol. 20, no. 4, pp. 207–228, Oct. 2021.

[126] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen, "Collaborative visualization: Definition, challenges, and research agenda," *Inf. Vis.*, vol. 10, no. 4, pp. 310–326, Oct. 2011.

[127] R. Fleck and G. Fitzpatrick, "Reflecting on reflection: Framing a design landscape," in *Proc. 22nd Conf. Comput. Hum. Interact. Spec. Interest Group Austral. Comput. Hum. Interact.*, Nov. 22–26, 2010, pp. 216–223.

[128] D. A. Schön, *The Reflective Practitioner: How Professionals Think in Action*. Oxfordshire, UK: Routledge, 2017.

[129] F. Fischer, "Knowledge politics and post-truth in climate denial: On the social construction of alternative facts," *Crit. Policy Stud.*, vol. 13, no. 2, pp. 133–152, Apr. 2019.

[130] 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.*, May 8–13, 2021, pp. 1–18.

[131] S. Passi and S. J. Jackson, "Trust in data science: Collaboration, translation, and accountability in corporate data science projects," *Proc. ACM Hum. Comput. Interact.*, vol. 2, no. CSCW, pp. 136:1–136:28, Nov. 2018.

[132] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2431–2440, Dec. 2012.

[133] E. Kerzner, S. Goodwin, J. Dykes, S. Jones, and M. Meyer, "A framework for creative visualization-opportunities workshops," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 748–758, Jan. 2019.

[134] A. Pfuntner, L. M. Wier, and C. Stocks, "Most frequent procedures performed in U.S. hospitals, 2010: Statistical brief #149," in *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Rockville, MD, USA: Agency for Healthcare Research and Quality, 2013.

[135] K. D. Ellingson, M. R. P. Sapiano, K. A. Haass, A. A. Savinkina, M. L. Baker, K.-W. Chung, R. A. Henry, J. J. Berger, M. J. Kuehnert, and S. V. Basavaraju, "Continued decline in blood collection and transfusion in the United States–2015," *Transfusion*, vol. 57, no. Suppl 2, pp. 1588–1598, Jun. 2017.

[136] The Joint Commission and the American Medical Association-Convened Physician Consortium for Performance Improvement, "Proceedings from the National Summit on Overuse," Sep. 2012. [Online]. Available: https://www.parsemus.org/wp-content/uploads/2015/06/AMA-National_Summit_Overuse.pdf.

[137] M. Delaney, S. Wendel, R. S. Bercovitz, J. Cid, C. Cohn, N. M. Dunbar, T. O. Apelseth, M. Popovsky, S. J. Stanworth, A. Tinmouth, L. Van De Watering, J. H. Waters, M. Yazer, and A. Ziman, "Transfusion reactions: Prevention, diagnosis, and treatment," *Lancet*, vol. 388, no. 10061, pp. 2825–2836, Dec. 2016.

[138] I. Gross, B. Seifert, A. Hofmann, and D. R. Spahn, "Patient blood management in cardiac surgery results in fewer transfusions and better outcome," *Transfusion*, vol. 55, no. 5, pp. 1075–1081, May 2015.

[139] A. Rind, "Interactive information visualization to explore and query electronic health records," *Found. Trends Hum. Comput. Interact.*, vol. 5, no. 3, pp. 207–298, Feb. 2013.

[140] J. J. Caban and D. Gotz, "Visual analytics in healthcare–Opportunities and research challenges," *J. Am. Med. Inform. Assoc.*, vol. 22, no. 2, pp. 260–262, Mar. 2015.

[141] B. Shneiderman, C. Plaisant, and B. W. Hesse, "Improving healthcare with interactive visualization," *Computer*, vol. 46, no. 5, pp. 58–66, May 2013.

[142] V. L. West, D. Borland, and W. E. Hammond, "Innovative information visualization of electronic health record data: A systematic review," *J. Am. Med. Inform. Assoc.*, vol. 22, no. 2, pp. 330–339, Mar. 2015.

[143] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman, "LifeLines: Using visualization to enhance navigation and analysis of patient records," in *The Craft of Information Visualization*, B. B. Bederson and B. Shneiderman, Eds. San Francisco, CA, USA: Morgan Kaufmann, 2003, pp. 308–312.

[144] A. A. T. Bui, D. R. Aberle, and H. Kangarloo, "TimeLine: Visualizing integrated patient records," *IEEE Trans. Vis. Comput. Graph.*, vol. 11, no. 4, pp. 462–473, Jul. 2007.

[145] D. Gotz and H. Stavropoulos, "DecisionFlow: Visual analytics for high-dimensional temporal event sequence data," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 1783–1792, Dec. 2014.

[146] K. Wongsuphasawat and D. Gotz, "Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2659–2668, Dec. 2012.

[147] J. Krause, A. Perer, and H. Stavropoulos, "Supporting iterative cohort construction with visual temporal queries," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 91–100, Jan. 2016.

[148] J. Rogers, N. Spina, A. Neese, R. Hess, D. Brodke, and A. Lex, "Composer—Visual cohort analysis of patient outcomes," *Appl. Clin. Inform.*, vol. 10, no. 2, pp. 278–285, Mar. 2019.

[149] J. Bernard, D. Sessler, T. May, T. Schlomm, D. Pehrke, and J. Kohlhammer, "A visual-interactive system for prostate cancer cohort analysis," *IEEE Comput. Grap. Appl.*, vol. 35, no. 3, pp. 44–55, May 2015.

[150] M. Glueck, A. Gvozdik, F. Chevalier, A. Khan, M. Brudno, and D. Wigdor, "PhenoStacks: Cross-sectional cohort phenotype comparison visualizations," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 191–200, Jan. 2017.

[151] J. Bernard, D. Sessler, J. Kohlhammer, and R. A. Ruddle, “Using dashboard networks to visualize multiple patient histories: A design study on post-operative prostate cancer,” *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 3, pp. 1615–1628, Mar. 2018.

[152] Y. G. Rajwan and G. R. Kim, “Medical information visualization conceptual model for patient-physician health communication,” in *Proc. 1st ACM Int. Health Inform. Symp.*, Nov. 11–12, 2010, pp. 512–516.

[153] A. Hakone, L. Harrison, A. Ottley, N. Winters, C. Gutheil, P. K. J. Han, and R. Chang, “PROACT: Iterative design of a patient-centered visualization for effective prostate cancer health risk communication,” *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 601–610, Jan. 2017.

[154] J. A. Gálvez, L. Ahumada, A. F. Simpao, E. E. Lin, C. P. Bonafide, D. Choudhry, W. R. England, A. F. Jawad, D. Friedman, D. A. Sesok-Pizzini, and M. A. Rehman, “Visual analytical tool for evaluation of 10-year perioperative transfusion practice at a children’s hospital,” *J. Am. Med. Inform. Assoc.*, vol. 21, no. 3, pp. 529–534, May 2014.

[155] M. Brehmer and T. Munzner, “A multi-level typology of abstract visualization tasks,” *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2376–2385, Dec. 2013.

[156] M. Gleicher, “Considerations for visualizing comparison,” *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 413–423, Jan. 2018.

[157] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit, “LineUp: Visual analysis of multi-attribute rankings,” *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2277–2286, Dec. 2013.

[158] K. Furmanova, S. Gratzl, H. Stitz, T. Zichner, M. Jaresova, A. Lex, and M. Streit, “Taggle: Combining overview and details in tabular data visualizations,” *Inf. Vis.*, vol. 19, no. 2, pp. 114–136, Nov. 2020.

[159] B. Brown, S. Reeves, and S. Sherwood, “Into the wild: Challenges and opportunities for field trial methods,” in *Proc. 2011 Annu. Conf. Hum. Factors Comput. Syst.*, May 7–12, 2011, pp. 1657–1666.

[160] R. A. Metcalf, S. K. White, S. Potter, R. Barney, C. Hunter, M. White, T. Enniss, C. Galaviz, S. Reddy, N. Wanner, R. L. Schmidt, and R. Blaylock, “The association of inpatient blood utilization and diagnosis-related group weight: Implications for risk-adjusted benchmarking,” *Transfusion*, vol. 59, no. 7, pp. 2316–2323, Jul. 2019.

[161] S. M. Frank, L. M. Resar, J. A. Rothschild, E. A. Dackiw, W. J. Savage, and P. M. Ness, “A novel method of data analysis for utilization of red blood cell transfusion: Analysis of RBC utilization,” *Transfusion*, vol. 53, no. 12, pp. 3052–3059, Dec. 2013.

[162] J. Rogers, A. H. Patton, L. Harmon, A. Lex, and M. Meyer, “Insights from experiments with rigor in an EvoBio design study,” *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 1106–1116, Feb. 2021.

[163] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg, “Exploring the placement and design of word-scale visualizations,” *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2291–2300, Dec. 2014.

[164] J. Liu, N. Boukhelifa, and J. R. Eagan, “Understanding the role of alternatives in data analysis practices,” *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 1, pp. 66–76, Jan. 2020.

[165] A. Azad, E. Sernbo, V. Svärd, L. Holmlund, and E. Björk Brämberg, “Conducting in-depth interviews via mobile phone with persons with common mental disorders and multimorbidity: The challenges and advantages as experienced by participants and researchers,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 22, p. 11828, Nov. 2021.

[166] D. R. Johnson, C. P. Scheitle, and E. H. Ecklund, “Beyond the in-person interview? How interview quality varies across in-person, telephone, and Skype interviews,” *Soc. Sci. Comput. Rev.*, vol. 39, no. 6, pp. 1142–1158, Dec. 2021.

[167] J. Monforte and J. Úbeda-Colomer, “Tinkering with the two-to-one interview: Reflections on the use of two interviewers in qualitative constructionist inquiry,” *Methods Psychol.*, vol. 5, Art. no. 100082, Dec. 2021.

[168] D. Akbaba, D. Lange, M. Correll, A. Lex, and M. Meyer, “Troubling collaboration: Matters of care for visualization design,” in *Proc. 2023 CHI Conf. Hum. Factors Comput. Syst.*, Apr. 23–28, 2023, pp. 1–15.

[169] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, Jan. 2006.

[170] M. Dörk, P. Feng, C. Collins, and S. Carpendale, “Critical InfoVis: Exploring the politics of visualization,” in *CHI 13 Ext. Abstr. Hum. Factors Comput. Syst.*, Apr. 27–May 2, 2013, pp. 2189–2198.

[171] d. boyd and K. Crawford, “Critical questions for big data,” *Inf. Commun. Soc.*, vol. 15, no. 5, pp. 662–679, Jun. 2012.

[172] C. M. Dalton, L. Taylor, and J. Thatcher, “Critical data studies: A dialog on data and space,” *Big Data Soc.*, vol. 3, no. 1, pp. 1–9, Jun. 2016.

[173] P. Hunter, “The reproducibility ‘crisis’,” *EMBO Rep.*, vol. 18, no. 9, pp. 1493–1496, Sep. 2017.

[174] C. D’Ignazio and L. F. Klein, “Feminist data visualization,” in *Workshop Vis. Digit. Humanit. VIS4DH*, Oct. 24, 2016, pp. 1–5.

[175] Y. L. Wong, K. Madhavan, and N. Elmqvist, “Towards characterizing domain experts as a user group,” in *2018 IEEE Eval. Beyond Methodol. Approaches Vis. BELIV*, Oct. 21, 2018, pp. 1–10.

[176] S. Kandel, R. Parikh, A. Paepcke, J. M. Hellerstein, and J. Heer, “Profiler: Integrated statistical analysis and visualization for data quality assessment,” in *Proc. Conf. Adv. Vis. Interfaces AVI 12*, May 21–25, 2012, pp. 547–554.

[177] S. Kandel, A. Paepcke, J. Hellerstein, and J. Heer, “Wrangler: Interactive visual specification of data transformation scripts,” in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, May 7–12, 2011, pp. 3363–3372.

[178] W. Willett, P. Goffin, and P. Isenberg, "Understanding digital note-taking practice for visualization," *IEEE Comput. Grap. Appl.*, vol. 35, no. 4, pp. 38–51, Jul. 2015.

[179] M. Hegarty and K. Steinhoff, "Individual differences in use of diagrams as external memory in mechanical reasoning," *Learn. Individ. Differ.*, vol. 9, no. 1, pp. 19–42, Jan. 1997.

[180] J. Walny, S. Huron, C. Perin, T. Wun, R. Pusch, and S. Carpendale, "Active reading of visualizations," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 770–780, Jan. 2018.

[181] G. Aisch, A. Cox, and K. Quealy, "You draw it: How family income predicts children's college chances," *New York Times*, May 2015. [Online]. Available: https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html.

[182] T. Baudel, "From information visualization to direct manipulation: Extending a generic visualization framework for the interactive editing of large datasets," in *Proc. 19th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 15–18, 2006, pp. 67–76.

[183] B. Saket, H. Kim, E. T. Brown, and A. Endert, "Visualization by demonstration: An interaction paradigm for visual data exploration," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 331–340, Jan. 2017.

[184] A. Quispel and A. Maes, "Would you prefer pie or cupcakes? Preferences for data visualization designs of professionals and laypeople in graphic design," *J. Vis. Lang. Comput.*, vol. 25, no. 2, pp. 107–116, Apr. 2014.

[185] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby, "Sketchy rendering for information visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2749–2758, Dec. 2012.

[186] S. K. Badam, S. Chandrasegaran, and N. Elmqvist, "Integrating annotations into multidimensional visual dashboards," *Inf. Vis.*, vol. 21, no. 3, pp. 270–284, Jul. 2022.

[187] N. Seaver, "The nice thing about context is that everyone has it," *Media Cult. Soc.*, vol. 37, no. 7, pp. 1101–1109, Oct. 2015.

[188] Y. Zhang and C. Huxham, "Identity construction and trust building in developing international collaborations," *J. Appl. Behav. Sci.*, vol. 45, no. 2, pp. 186–211, Jun. 2009.

[189] R. Kitchin and T. Lauriault, "Towards critical data studies: Charting and unpacking data assemblages and their work," in *Thinking Big Data in Geography: New Regimes, New Research*, J. Eckert and J. Thatcher, Eds. Lincoln, NE, USA: University of Nebraska Press, 2018, pp. 3–20.

[190] M. Correll, "Ethical dimensions of visualization research," in *Proc. 2019 CHI Conf. Hum. Factors Comput. Syst.*, May 4–9, 2019, pp. 1–13.

[191] D. Haraway, "Situated knowledges: The science question in feminism and the privilege of partial perspective," *Fem. Stud.*, vol. 14, no. 3, p. 575, Sep. 1988.

[192] J. Moore, P. Goffin, J. Wiese, and M. Meyer, "Exploring the personal informatics analysis gap: "There's a lot of bacon"," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 1, pp. 96–106, Jan. 2022.

[193] L. Bartram, M. Correll, and M. Tory, "Untidy data: The unreasonable effectiveness of tables," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 1, pp. 686–696, Jan. 2022.

[194] National Academies of Sciences, *Reproducibility and Replicability in Science*. Washington, DC, USA: National Academies Press, 2019.

[195] B. J. Heil, M. M. Hoffman, F. Markowetz, S.-I. Lee, C. S. Greene, and S. C. Hicks, "Reproducibility standards for machine learning in the life sciences," *Nat. Methods*, vol. 18, no. 10, pp. 1132–1135, Oct. 2021.

[196] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala, "Graphical histories for visualization: Supporting analysis, communication, and evaluation," *IEEE Trans. Vis. Comput. Graph.*, vol. 14, no. 6, pp. 1189–1196, Nov. 2008.

[197] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit, "From visual exploration to storytelling and back again," *Comput. Graph. Forum*, vol. 35, no. 3, pp. 491–500, Jun. 2016.

[198] J. Iwasa, M. Meyer, A. Lex, J. Rogers, A. H. Liu, and M. Riggi, "SARS-CoV-2 visualization and annotation project," The Animation Lab. https://animationlab.utah.edu/cova (accessed Sep. 9, 2021).

[199] T. Guan, L. Wang, J. Jin, and X. Song, "Knowledge contribution behavior in online Q&a communities: An empirical investigation," *Comput. Hum. Behav.*, vol. 81, pp. 137–147, Apr. 2018.

[200] M. Lisnic, C. Polychronis, A. Lex, and M. Kogan, "Misleading beyond visual tricks: How people actually lie with charts," in *Proc. 2023 CHI Conf. Hum. Factors Comput. Syst.*, Apr. 23–28, 2023, pp. 1–21.

[201] S. Mohammed, "Toward an understanding of cognitive consensus in a group decision-making context," *J. Appl. Behav. Sci.*, vol. 37, no. 4, pp. 408–425, Dec. 2001.

[202] C. P. Barlett, "Anonymously hurting others online: The effect of anonymity on cyberbullying frequency," *Psychol. Pop. Media Cult.*, vol. 4, no. 2, pp. 70–79, Apr. 2015.

[203] B. Friedman and H. Nissenbaum, "Bias in computer systems," *ACM Trans. Inf. Syst.*, vol. 14, no. 3, pp. 330–347, Jul. 1996.

[204] 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 *2017 IEEE Conf. Vis. Anal. Sci. Technol. VAST*, Oct. 3–6, 2017, pp. 104–115.

[205] Y.-S. Kim, K. Reinecke, and J. Hullman, "Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 760–769, Jan. 2018.

[206] J. Wanner, L.-V. Herm, and C. Janiesch, "How much is the black box? The value of explainability in machine learning models," in *Proc. Eur. Conf. Inf. Syst.*, Jun. 15–17, 2020, pp. 1–14.

[207] R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke, "Explainable machine learning for scientific insights and discoveries," *IEEE Access*, vol. 8, pp. 42 200–42 216, Feb. 2020.

[208] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. F. Moura, and P. Eckersley, "Explainable machine learning in deployment," in *Proc. 2020 Conf. Fairness Account. Transpar.*, Jan. 27–30, 2020, pp. 648–657.