# HERE'S WHAT YOU NEED TO KNOW ABOUT MY DATA: EXPLORING EXPERT KNOWLEDGE'S ROLE IN DATA ANALYSIS

**Haihan Lin**
University of Utah

**Maxim Lisnic**
University of Utah

**Derya Akbaba**
Linköping University

**Miriah Meyer**
Linköping University

**Alexander Lex**
University of Utah
alex@sci.utah.edu

### ABSTRACT

Data driven decision making has become the gold standard in science, industry, and public policy. Yet data alone, as an imperfect and partial representation of reality, is often insufficient to make good analysis decisions. Knowledge about the context of a dataset, its strengths and weaknesses, and its applicability for certain tasks is essential. In this work, we present an interview study with analysts from a wide range of domains and with varied expertise and experience inquiring about the role of contextual knowledge. We provide insights into how data is insufficient in analysts workflows and how they incorporate other sources of knowledge into their analysis. We also suggest design opportunities to better and more robustly consider both, knowledge and data in analysis processes.

***K*eywords** Data Visualization · Interview Study · Expert Knowledge

## 1 Introduction

On September 26, 1983, the Soviet Air Defense Forces' computers reported five missiles heading towards the Soviet Union from the United States, triggering a protocol that called for an immediate and compulsory nuclear counter-attack. 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 [1]. Later investigation revealed that the system had indeed malfunctioned due to a rare alignment of the detection satellite and the sun. Petrov's knowledge and experience enabled him to recognize the possibility of a false alarm and to interpret the data in the context of the political situation. Had he solely relied on the warning system, the consequences could have been catastrophic.

While not all data (mis)interpretations lead to world-shattering consequences, data-driven decision-making has become the gold standard in fields like public policy, science, and industry, but also in making choices about our everyday lives. However, data alone is not sufficient to make good decisions. Data is an imperfect and partial representation of reality [2, 3], it can be misleading [4], hence acting solely based on data can be dangerous, as the story about Stanislav Petrov illustrates. Expert knowledge, on the other hand, can provide essential context for the data and has a critical role in data analysis [5, 6]. Experts know about relevant contexts based on their experience and domain knowledge, familiarity with the subject, and understanding of the data collection modalities.

Analysts that work with data often find themselves incorporating (their own or others') expert knowledge into their analysis, as illustrated in Fig 1. Expert knowledge provides analysts with context and caveats about the data and assures analysts of the soundness of their analysis. However, there is yet much work in the visualization research community to explore the details of how expert knowledge is integrated throughout an analyst's workflow. In particular,

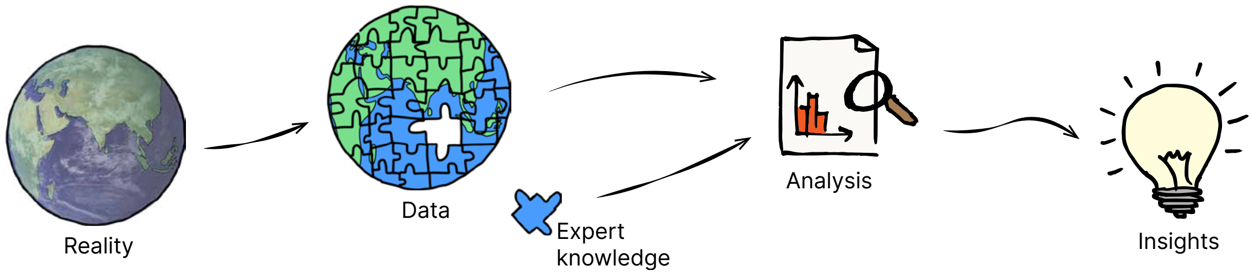


**Figure 1:** An overview of the role of expert knowledge in data analysis workflows. Data is an imprecise and incomplete representation of reality. Expert knowledge helps with understanding the limitations of the data and may fill in the gaps between data and reality. Data analysis should leverage both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

we argue that current approaches to incorporating knowledge are ad-hoc, hampered by inefficient communication between stakeholders, and often not (sufficiently) documented; thereby leading to worse analysis results, lack of reproducibility of analysis, and lack of reusability of a dataset.

In this work, we share the result of an interview study with 14 domain experts and analysts from a broad range of fields that investigates how they deal with data caveats in their workflow, how 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 primary contribution is an analysis of participants' current practices in capturing relevant knowledge and where the current practices fall short. We also contribute a discussion of design opportunities to better support analysts in documenting data caveats and expert knowledge in data analysis pipelines.

## 2 Data as a Tool

The typical reason why people collect data is to measure and record a phenomenon in reality. In practice, data is rarely a faithful and comprehensive representation of reality [7]. One of our participants recognized this, stating a variation on a common aphorism[1] from statistics: *"all observations are bad, but some are useful" (P2).* Since data is not a perfect depiction of reality, the usefulness of a particular dataset depends on how well it serves a task: **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. Like with any tool, knowledge about how and when to use the tool, and knowledge about its limitations is essential.

We previously defined such knowledge as **data hunches** [8]. Data hunches are the knowledge people have about the mismatch between reality and data. We refer to our previous work for a more detailed discussion of data hunches and its theoretical foundations. While 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 present with the data itself. To revisit the tool analogy: data hunches capture the knowledge on how to use the screwdriver effectively; data hunches help analysts use data, despite its imperfections, to serve their analysis needs. We argue that understanding the context and limitations of a dataset (i.e., knowing about the data hunches) is essential in the data analysis process [9, 10]. In this work, we attempt to understand *how* analysts develop, capture, utilize, and communicate data hunches.

[1] *All models are wrong, but some are useful*—often attributed to George Box.

## 3 Related Work

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

### 3.1 Practices of Data Workers

There is an abundance of research from the visualization and HCI communities that provides insight into the current practices of data workers and analysts [11, 12, 13, 14]. 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* [15]. They found that although visualizations touch all of 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 covered all these stages of analysis, with a focus on preparation, analysis, and communication.

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 [16]. The authors described the difference in patterns of communication exhibited by different roles within teams. Reflecting on the practices of analysts across different fields, our work also presents findings that the communication direction patterns heavily depend on the role, with an emphasis on the communication of domain expert knowledge. Jung et al. presented an in-depth study into how domain experts work with data [17] and found that they put 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 [18]—as a critical part of the analysis process, similar to previous works [5, 19]. Our study 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 [20]. 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 [21]. Hullman contributes to this research with more evidence as to why visualization practitioners actively choose to omit uncertainty in their visualizations, citing visualization authors' concern of overwhelming their audience with too much information and uncertainty [22]. Our work further explores how analysts view and communicate uncertainty: it highly depends on the situation and context, analysts' role within their team, and their perceived professional responsibilities.

### 3.2 Methods of Documenting Data and Analysis

Annotations are a common way to document data and analysis. They are often added to visualizations to provide context, highlight certain points or issues, or tell a story about the data. There has been much research on how to build tools to better support annotations. Kim et al. reported that study participants in a laboratory setting recorded patterns on statistical distributions using textual annotations and identified trends and anomalies using graphical annotations [23]. Annotations can also help analysts to revisit and reflect on their findings, or help with contextualizing data [24]. Annotations appear in many forms in data visualization: text [25, 26, 27], symbols [28], sketches [29, 30, 31], or even audio recordings [32]. Annotation systems have also served as important tools in collaborative work. For instance, they can be used to capture insights for other users to see or to continue the analysis [33]. Sharing knowledge is another benefit that annotations can bring into collaborative settings, as demonstrated in McCurdy et al., where experts shared their tacit knowledge with peers working on the same data [25]. Annotations can facilitate knowledge building and start conversations, similar to online communities [34, 28]. Our work explores the practices of externalizing context about the analysis with their existing tools.

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 better understanding of the condition, quality, and caveats about the data [35]. Many lab notebooks have been transitioned into digital versions [36, 37] with rich content like visualizations [38]. Even though maintaining lab notebooks is a standard practice in science communities, physical notebooks can be lost, and digital ones often lack flexibility [39]. Computational notebooks, such as Jupyter Notebooks [40] and R Markdown [41, 42] are often discussed as a remedy for the issues we discuss here: they can be used to describe datasets and analysis steps, contain visualizations, and also contain executable code that (in theory [43]) ensures reproducibility of analysis. Due to these advantages, significant research has been devoted to understanding how analysts use computational notebooks [44, 45, 46] and to improve them [47, 48, 49, 50].

Notes and records, in turn, often are transferred into method sections in publications and reports, where readers can find details about the data and analysis steps, helping them judge the validity and reproducibility of the study [51, 52]. Method sections, however, are often space-limited, and details are omitted or favor describing the main results of the publication. Additionally, detailed methods sections are the purview of scientific work.

Metadata is another medium to communicate the structure and information about the data. Metadata ensures the meaningfulness of the data [53, 54] and provides critical information about the data [9]. 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 [55]. In our interviews, we explored how the participants utilize these established mediums to record caveats about their data.

## 4 Methods

Our interview study is inspired by our previous work on data hunches [8]. In that paper, we conceptualized analysts' knowledge about how data partially represents the phenomenon of interest and proposed techniques to record data hunches. We drew inspiration from our collaboration with domain experts 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.

### 4.1 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. All our participants are professional data analysts with degrees in their domain, or fields such as mathematics and statistics. Notably, none have formal training in computer science.

| | 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 |

**Table 1:** Overview of the characteristics of the 14 participants across different fields in academia and industry.

Most participants have extensive domain knowledge based on their training (academics in the sciences, engineers), and all have considerable experience in their domain. 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) have a range of experience (4 to 30+ years) and work in a variety of fields such as civil engineering, legal services, atmosphere science, psychiatry, and policy-making (see Table 1 for details on the participants).

The interview protocol was submitted to the University of Utah IRB and deemed exempt from review. The participants gave informed consent to be in the study and to be audio-recorded before the interview. Participants were not compensated. We 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 [56, 57]. Hence, 13 out of 14 participants are based in Utah.

### 4.2 Interviews

The goal of our interviews was to study *if* analysts use expert knowledge in their workflows, and if so, how. We conducted two pre-pilot interviews with lab members to test the interview script draft and solicit feedback on the procedure and structure. We then conducted two pilot interviews with collaborators who met the inclusion criteria, to test the outcome and modality of the adjusted script and structure. We then conducted 14 semi-structured interviews, 12 in-person and 2 remote. We scheduled interviews as the project progressed and decided to stop recruiting new participants when we reached saturation, noticing that no new topics were brought up. The interviews were conducted by authors Lin and Lisnic using a two-to-one interview approach [58]. Lin asked the prepared questions and guided the conversation, while 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 [59].

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 where they were asked to write down titles of data-driven projects that they had worked on. The short activity was intended to help participants reflect on their past projects and to ensure that they had a list of possible topics to refer to throughout the interview. We then asked participants to pick an example 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, where 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. 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*, where we presented our previous work on recording and communicating data hunches [8] through a prepared slide deck (available in supplementary materials). After the presentation, we asked whether participants were currently recording data hunches in their workflow, and how the participants 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. We include the interview script in our supplementary materials.

### 4.3 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 [60]. Three authors (Lin, Lisnic, 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) 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. For readability, we tidied up reported quotes by correcting grammar and removing filler words (*like*, *yeah*, etc.). We include a table of the original, unedited versions in our supplementary materials.

## 5 Findings

We categorize our findings into four themes: the relationship between data and reality, how knowledge fills the gap between data and reality, current practices in dealing with imperfect data, and interventions for better communicating data hunches in analysis workflows. 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 using a yellow box.

### 5.1 The Relationship of Data and Reality

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

Data is not able to perfectly or completely represent the world [61, 62, 63], 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 analyzes 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:

> *P12: 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.*

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

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

Participants also work with data collected by equipment such as sensors (P6), satellites (P2, P8), and laser imagery (P5), however, they consistently noted that even sensor data is shaped by its context. For example, P6 installed sensors in various locations to collect air quality data 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.*

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 is filled with caveats, shaped by the contexts in which they were constructed. Nevertheless, participants, fully aware that data is shaped by context, often repurposed data to suit their analysis needs.

One of our participants (P4) used data collected by a foreign institute which used the data to study rainfall. P4, on the other hand, used the data to study snowfall models. However, 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.

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

Participants P1 and P3, who study suicide risks among certain populations, were using data labeled with ICD codes, collected in a clinical context, as proxies for patient diagnoses in their research. ICD codes were originally recorded for billing purposes, which results 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 clinicians with direct knowledge about the caveats on the billing codes. Even though the teams were aware that the data is 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. Participants employ different methods of working with caveats to fulfill their analysis. In the instance of snowfall modeling, knowledge about the way in which the data was processed allowed the analyst to use the data; yet this knowledge was not readily available. In the instance of working with patient data categorized with ICD codes, the analysts sought confirmation and aid from clinicians who understood the codes more expertly.

Participants know their data is imperfect.

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

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

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.

> *P1: 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: 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].*

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

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

**Summary** Our participants regularly use 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 repurpose datasets to fill their analysis needs but do so with attention to detail and the original contexts in which the data originates. Failing to account for these nuances has caused issues for participants before. Across participants, there is a sustained sentiment that data is simply an imperfect tool for the analyses that they are trying to do, rather than a representation of reality that is marked by uncertainty.

### 5.2 Knowledge Fills the Gap Between Data and Reality

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

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 works 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 non-profit organizations to ensure that her interpretations were reasonable:

> *P7: I presented it and said, "I'm not a subject matter expert in evictions. Tell me what you see."*

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 works for a regional government, describes her experience of being the only data analyst in her office as being in a skill set "*silo*." Because of this, she often utilizes her connections to analysts in other departments and reaches out to double-check her analysis results:

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

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

Participants often seek the knowledge of individuals with situated or lived experience about the data. This includes individuals who reside in proximity to the reality that the data describes, for example, people who live close to a river that is being measured; or those whose lived experience is part of what is being analyzed, such as employees whose work output is reported on a dashboard. Their proximity to the data provides 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, hence not capturing the bottom of the river accurately. To remedy this, the team had to solicit the help of a local partner:

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

Similarly, P6, a chemical engineer, deployed air quality sensors in various communities and regularly monitored air quality through a central dashboard. In times of anomalous air quality readings, she would first email the local community to check for any special events that might have impacted the readings, such as controlled fires set off as part of forest management, before concluding that a sensor is faulty and may need replacement:

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

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, 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.*

> 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 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: You can give feedback, redirect, pivot, before you waste too much time making it perfect [...] 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 our introductory example about the threat of nuclear war. In another instance, P8, who works 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 shows 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?"*

**Summary** Our participants rely on expert 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.

## 5.3 Current Practices for Dealing with Imperfect Data

As discussed, 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.

### 5.3.1 Recording Data Hunches

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 participants do 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.

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

Participants often use email and chat apps, in which they post text (P1, P9, P13) and screenshots of visualizations (P5, P6, P7, P13), to elicit feedback from peers or domain experts. In return, the outcome of these communications becomes 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 is cited frequently as preventing properly documenting qualitative knowledge. One analyst (P11) described tracking caveats in cells next to the affected items in Excel, while another expressed hesitation about doing that because it might affect down-stream analysis tools:

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

Only one analyst (P8), who is part of a national weather forecast organization, uses an in-house tool with the capability for annotations built in to record any caveats in the weather forecast for shift handoff.

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 makes 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. When an intern joined the department temporarily, she verbally communicated to the intern how to treat the data because of all the data caveats, such as *"you can ignore that data from X, because I know they're wrong for various reasons"*. This knowledge exchange happened ad-hoc, and P13 was unsure whether she covered all the data caveats exhaustively. She also stated that such a lack of documentation has led to wasted effort before:

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

P4 also faced an issue related to a lack of documented data hunches, 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.

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. For example, P2 reported taking "abundant field notes" of the measuring equipment condition and weather context for weather data collection and then documenting the relevant aspects in a methods section in a research paper.

> *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.*

P5 prepares an appendix for his reports on flood disaster models, documenting uncertainty and other assumptions. This report with the appendix goes through a strict, multistage review process:

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

P14, the epidemiologist, uses metadata sheets to capture caveats in their analysis to make sure they have proper data context for the analysis down the pipeline, although she states that improving their internal documentation is a strategic goal.

**Summary** 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 ad-hoc records for their analysis.

#### 5.3.2 Communicating During the Analysis Process

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 is the main communication tool that participants use to elicit domain knowledge and feedback from the experts. However, they described being frustrated when dealing with text and screenshots. Data caveats that participants deal with can be complex, and participants find 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.

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

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: 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.*

Even though participants would draw and annotate on visualizations during these meetings, the drawings are not archived, hence the knowledge exchange is not preserved. Alternatively, participants use screenshots in chat, e-mail, or PowerPoint which they may or may not annotate. We found no instances of annotations happening directly in a visual analytic tool that participants use.

> *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 use simple graphics editors or do not annotate directly in the visualizations instead.

**Summary** During the analysis process, participants use tools like 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.

#### 5.3.3 Communicating Analysis Results and Data Hunches

Participants use a variety of mediums to report their analysis outcomes. These mediums include 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 takes different forms, which we will break down in this section.

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

Participants that are academics (P1, P2, P3, P4) most commonly use method sections in their publication to report data hunches. Similarly, participants that write up their analysis as reports (P5, P9, P11) document 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 [64].

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

Most participants use visualizations to communicate their data, as they value 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.

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

However, when asked about preparing and anticipating questions about the data, many participants answered that they would respond ad-hoc, rather than preemptively cover data imperfections.

> 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 see it as their professional responsibility to distill data into actionable items for decision-makers, and that they are trusted to correctly interpret the data to the best of their ability. Communicating more data caveats to leadership increases the complexity, and leadership and external audiences often are not interested in the details of the analysis.

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

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

We discovered different approaches towards expressing caveats about the analysis depending on the format and the type of participants' jobs. For example, our academic participants are cautious about keeping their analysis choices transparent in research manuscripts (P1, P2, P3, P4), whereas other participants feel hesitant to express uncertainty explicitly in their analysis deliverables due to the reasons laid out above. This omission of caveats is even more amplified in verbal presentations, which our participants justified by being available for clarifying questions if needed. However, participants also adjust 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 add disclaimers to deliverables that are used asynchronously, such as dashboards. But even there, the focus is to avoid making the results confusing. For example, P14 would refrain 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:

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

However, she would add additional annotations to her dashboard if she would learn that an aspect was regularly misinterpreted, which she measured by the number of calls she received about an issue:

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

**Summary** 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.

### 5.4 Exploring Interventions for Recording Data Hunches

During the *"feedback on a technology probe"* section of the interview, we introduced our definition of data hunches to the participants and provided a brief demonstration (see supplementary material) of the techniques we proposed previously [8]. The demo included the basic workflow of recording and communicating a data hunch through the prototype 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.

> 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 great at verbally expressing their opinions and knowledge about the data. An interactive visual option, she commented, would be a good option for these collaborations:

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

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

> 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 might be a burden to experts that 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.*

**Summary** Participants showed preferences toward interactive visualization techniques to record and communicate knowledge about the data. A visual approach that can help bring analysts and experts on the same page is appealing. However, participants showed yet-another-tool fatigue and would prefer a solution integrated with their current tool suite to externalize and exchange knowledge.

## 6 Discussion

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

### 6.1 The Role of Expert's Knowledge in Data Analysis

Prior work characterizes data as an artifact of decisions: a culmination of the specific and situated contexts in which they were constructed [8, 2, 65, 10]. The construction of data leaves it with gaps and caveats such that for data to fully reflect reality, data requires context [65]. In the interviews, participants discussed many different ways that they understood and worked with the limitations of their data.

Expert knowledge often complemented the data, piecing together the spaces between data and reality. Surprisingly, analysts found expert knowledge outside of the traditional domain expert as conceptualized by the visualization community [66, 67]. 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. However, we heard many accounts of how recording this knowledge is brittle and unsystematic: scattered across ephemeral records like chat histories or one-off emails containing notes, or communicated in a meeting. And thus, the lack of documentation makes reanalysis and reproducibility challenging, creating a barrier for other analysts outside the discussion to join.

Furthermore, the participants never expected the data to be perfect. In fact, even though some participants had a strong faith in the numbers' accuracy, they still shared experiences where the numbers were an imperfect representation of what they were trying to study. The data were 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 [68], and ensemble plots [69] 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. The 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.

### 6.2 Design Opportunities

While the demonstration of our previously developed tool for externalizing and communicating data hunches [8] seemed to resonate with our participants and they could easily come up with opportunities where it would be beneficial, 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 [14] 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 [59, 70, 71] 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 at 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, this could be built-in annotation capabilities on top of screenshots for communication tools like 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, like Jupyter, were not mentioned once in all of our interviews, even though many of the issues discussed by our participants could be addressed using such tools. We also note that only two of our participants reported using programming languages within their analysis—R (P9) and Python (P4). These lead us to speculate that there is a (possibly large) number of data analysts whose analysis processes cannot meet standards for reproducibility laid out by various scientific bodies [72, 73]. We see this 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 **GUI-based visual analysis tools reproducible**. The GUI-based tools our participants use do not support annotated histories or workflows, unlike various research prototypes [74, 75, 76, 47]. 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.

Second, while data hunches are often expressed when viewing data through visualizations, we believe it is important to also **capture data hunches at the data level**—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 [9] to not only document the *what* that is in the data, but also the *why* and *how*. Ideally, datasets should be published or archived together with a reproducible analysis story that makes it clear how the data was used. 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, 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. It is notable that in our interviews we found that external pressures and established guidelines lead to better documentation practices.

## 7 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 believe that our results generalize to other analysts with similar characteristics.

## 8 Conclusion

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

## 9 Acknowledgements

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

# References

[1] David Hoffman. 'I Had A Funny Feeling in My Gut'. *The Washington Post*, page A19, February 1999.

[2] Johanna Drucker. Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, 5(1), 2011.

[3] Yanni Alexander Loukissas. *All Data Are Local: Thinking Critically in a Data-Driven Society*. MIT Press, April 2019.

[4] Mohamed K. Elhadad, Kin Fun Li, and Fayez Gebali. Detecting Misleading Information on COVID-19. *IEEE Access*, 8:165201–165215, 2020.

[5] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 126:1–126:15. Association for Computing Machinery, 2019.

[6] B. Karer, H. Hagen, and D. J. Lehmann. Insight Beyond Numbers: The Impact of Qualitative Factors on Visual Data Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1011–1021, February 2021.

[7] Melanie Feinberg. A Design Perspective on Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 2952–2963, New York, NY, USA, May 2017. Association for Computing Machinery.

[8] Haihan Lin, Derya Akbaba, Miriah Meyer, and Alexander Lex. Data Hunches: Incorporating Personal Knowledge Into Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):504–514, September 2022.

[9] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. Datasheets for Datasets. *Communications of the ACM*, 64(12):86–92, December 2021.

[10] Catherine D'Ignazio and Lauren F Klein. Feminist Data Visualization. In *Workshop on Visualization for the Digital Humanities (VIS4DH)*, 2016.

[11] Danyel Fisher, Rob DeLine, Mary Czerwinski, and Steven Drucker. Interactions with big data analytics. *Interactions*, 19(3):50–59, May 2012.

[12] Yaoli Mao, Dakuo Wang, Michael Muller, Kush R. Varshney, Ioana Baldini, Casey Dugan, and Aleksandra Mojsilović. How Data Scientists Work Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question? *Proceedings of the ACM on Human-Computer Interaction*, 3(GROUP):237:1–237:23, 2019.

[13] Michael Muller and Angelika Strohmayer. Forgetting Practices in the Data Sciences. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–19, New York, NY, USA, April 2022. Association for Computing Machinery.

[14] Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Jin, and Marti A. Hearst. Futzing and Moseying: Interviews with Professional Data Analysts on Exploration Practices. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):22–31, January 2019.

[15] Anamaria Crisan, Brittany Fiore-Gartland, and Melanie Tory. Passing the Data Baton : A Retrospective Analysis on Data Science Work and Workers. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1860–1870, February 2021.

[16] Amy X. Zhang, Michael Muller, and Dakuo Wang. How do Data Science Workers Collaborate? Roles, Workflows, and Tools. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):022:1–022:23, May 2020.

[17] Ju Yeon Jung, Tom Steinberger, John L. King, and Mark S. Ackerman. How Domain Experts Work with Data: Situating Data Science in the Practices and Settings of Craftwork. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):58:1–58:29, April 2022.

[18] Donald A. Schon. Designing as reflective conversation with the materials of a design situation. *Research in Engineering Design*, 3(3):131–147, September 1992.

[19] Melanie Tory, Lyn Bartram, Brittany Fiore-Gartland, and Anamaria Crisan. Finding Their Data Voice: Practices and Challenges of Dashboard Users. *IEEE Computer Graphics and Applications*, pages 1–1, 2022.

[20] Meredith Skeels, Bongshin Lee, Greg Smith, and George Robertson. Revealing uncertainty for information visualization. *Information Visualization*, 9(1):70–81, 2010.

[21] Nadia Boukhelifa, Marc-Emmanuel Perrin, Samuel Huron, and James Eagan. How Data Workers Cope with Uncertainty: A Task Characterisation Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3645–3656. Association for Computing Machinery, 2017.

[22] Jessica Hullman. Why Authors Don't Visualize Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):130–139, January 2020.

[23] Yea-Seul Kim, Nathalie Henry Riche, Bongshin Lee, Matthew Brehmer, Michel Pahud, Ken Hinckley, and Jessica Hullman. Inking Your Insights: Investigating Digital Externalization Behaviors During Data Analysis. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, ISS

'19, pages 255–267. Association for Computing Machinery, 2019.

[24] Jimmy Moore, Pascal Goffin, Miriah Meyer, Philip Lundrigan, Neal Patwari, Katherine Sward, and Jason Wiese. Managing In-home Environments through Sensing, Annotating, and Visualizing Air Quality Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):128:1–128:28, September 2018.

[25] Nina Mccurdy, Julie Gerdes, and Miriah Meyer. A Framework for Externalizing Implicit Error Using Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):925–935, January 2019.

[26] Nitesh Goyal, Gilly Leshed, and Susan R. Fussell. Effects of Visualization and Note-taking on Sensemaking and Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2721–2724. Association for Computing Machinery, 2013.

[27] Qi Liu, Chong Chen, Enjian Shen, Fangqing Zhao, Zhongsheng Sun, and Jinyu Wu. Detection, annotation and visualization of alternative splicing from RNA-Seq data with SplicingViewer. *Genomics*, 99(3):178–182, March 2012.

[28] Jeffrey Heer, Fernanda B Viégas, and Martin Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM*, 52(1):87–97, January 2009.

[29] Primož Lavrič, Ciril Bohak, and Matija Marolt. Collaborative view-aligned annotations in web-based 3D medical data visualization. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 259–263, May 2017.

[30] Anna Wu, Gregorio Convertino, Craig Ganoe, John M. Carroll, and Xiaolong (Luke) Zhang. Supporting collaborative sense-making in emergency management through geo-visualization. *International Journal of Human-Computer Studies*, 71(1):4–23, January 2013.

[31] Hugo Romat, Nathalie Henry Riche, Ken Hinckley, Bongshin Lee, Caroline Appert, Emmanuel Pietriga, and Christopher Collins. ActiveInk: (Th)Inking with Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 42:1–42:13. Association for Computing Machinery, 2019.

[32] Sean E. Ellis and Dennis P. Groth. A Collaborative Annotation System for Data Visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, pages 411–414, New York, NY, USA, 2004. ACM.

[33] Jian Zhao, Michael Glueck, Petra Isenberg, Fanny Chevalier, and Azam Khan. Supporting Handoff in Asynchronous Collaborative Sensemaking Using Knowledge-Transfer Graphs. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):340–350, 2018.

[34] Fernanda B. Viegas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. ManyEyes: A Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.

[35] Joseph MacNeil and Renee Falconer. When Learning the Hard Way Makes Learning Easy: Building Better Lab Note-Taking Skills. *Journal of Chemical Education*, 87(7):703–704, July 2010.

[36] Michael Rubacha, Anil K. Rattan, and Stephen C. Hosselet. A Review of Electronic Laboratory Notebooks Available in the Market Today. *JALA: Journal of the Association for Laboratory Automation*, 16(1):90–98, February 2011.

[37] Philip J. Guo and Margo Seltzer. BURRITO: Wrapping Your Lab Notebook in Computational Infrastructure. In *Proceedings of the 4th USENIX Conference on Theory and Practice of Provenance*, TaPP'12, pages 7–7, Berkeley, CA, USA, 2012. USENIX Association.

[38] S. P. Callahan, Juliana Freire, E. Santos, C. E. Scheidegger, Claudio Silva, and H. T. Vo. *Visualization in Radiation Oncology: Towards Replacing the Laboratory Notebook (SCI Institute Technical Report, No. UUSCI-2006-17)*. University of Utah, 2006.

[39] Clemens Nylandsted Klokmose and Pär-Ola Zander. Rethinking Laboratory Notebooks. In Myriam Lewkowicz, Parina Hassanaly, Volker Wulf, and Markus Rohde, editors, *Proceedings of COOP 2010*, pages 119–139, London, 2010. Springer.

[40] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, Sylvain Corlay, et al. Jupyter Notebooks-a publishing format for reproducible computational workflows. In *ELPUB*, pages 87–90, 2016.

[41] Benjamin Baumer and Dana Udwin. R Markdown. *WIREs Computational Statistics*, 7(3):167–177, 2015.

[42] Yihui Xie, J. J. Allaire, and Garrett Grolemund. *R Markdown: The Definitive Guide*. CRC Press, July 2018.

[43] Souti Chattopadhyay, Ishita Prasad, Austin Z. Henley, Anita Sarma, and Titus Barik. What's Wrong with Computational Notebooks? Pain Points, Needs, and Design Opportunities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–12, New York, NY, USA, April 2020. Association for Computing Machinery.

[44] Adam Rule, Aurélien Tabard, and James D. Hollan. Exploration and Explanation in Computational Notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Montreal QC Canada, April 2018. ACM.

[45] April Yi Wang, Anant Mittal, Christopher Brooks, and Steve Oney. How Data Scientists Use Computational Notebooks for Real-Time Collaboration. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):39:1–39:30, November 2019.

[46] Alena Guzharina. We Downloaded 10,000,000 Jupyter Notebooks From Github – This Is What We Learned | The JetBrains Datalore Blog. https://blog.jetbrains.com/datalore/2020/12/17/we-downloaded-10-000-000-jupyter-notebooks-from-github-this-is-what-we-learned/, December 2020.

[47] Kiran Gadhave, Zach Tyler Cutler, and Alexander Lex. Reusing Interactive Analysis Workflows. *EuroVis (to appear)*, 2022.

[48] John Wenskovitch, Jian Zhao, Scott Carter, Matthew Cooper, and Chris North. Albireo: An Interactive Tool for Visually Summarizing Computational Notebook Structure. In *2019 IEEE Visualization in Data Science (VDS)*, pages 1–10, October 2019.

[49] Yifan Wu, Joseph M. Hellerstein, and Arvind Satyanarayan. B2: Bridging Code and Interactive Visualization in Computational Notebooks. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, UIST '20, pages 152–165, New York, NY, USA, October 2020. Association for Computing Machinery.

[50] Mary Beth Kery and Brad A. Myers. Interactions for Untangling Messy History in a Computational Notebook. In *2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 147–155, 2018.

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

[52] Vivian Siegel. Reproducibility in research. *Disease Models & Mechanisms*, 4(3):279–280, May 2011.

[53] Richard Gartner. What Metadata Is and Why It Matters. In Richard Gartner, editor, *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*, pages 1–13. Springer International Publishing, Cham, 2016.

[54] Erik Duval and Wayne Hodgins. Metadata principles and practicalities. *D-Lib Magazine*, 8:2002, 2002.

[55] Alyxander Burns, Christiana Lee, Thai On, Cindy Xiong, Evan Peck, and Narges Mahyar. From Invisible to Visible: Impacts of Metadata in Communicative Data Visualization. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–16, 2022.

[56] Rafid Rahman. Comparison of Telephone and In-Person Interviews for Data Collection in Qualitative Human Research. *Interdisciplinary Undergraduate Research Journal*, 1(1):10–13, April 2015.

[57] David R. Johnson, Christopher P. Scheitle, and Elaine Howard Ecklund. Beyond the In-Person Interview? How Interview Quality Varies Across In-person, Telephone, and Skype Interviews. *Social Science Computer Review*, 39(6):1142–1158, December 2021.

[58] Javier Monforte and Joan Úbeda-Colomer. Tinkering with the two-to-one interview: Reflections on the use of two interviewers in qualitative constructionist inquiry. *Methods in Psychology*, 5:100082, December 2021.

[59] Derya Akbaba, Devin Lange, Michael Correll, Alexander Lex, and Miriah Meyer. Troubling Collaboration: Matters of Care for Visualization Design. In *SIGCHI Conference on Human Factors in Computing Systems (CHI) (to Appear)*. ACM, 2023.

[60] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006.

[61] Marian Dörk, Patrick Feng, Christopher Collins, and Sheelagh Carpendale. Critical InfoVis: Exploring the politics of visualization. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA'13, pages 2189–2198. Association for Computing Machinery, 2013.

[62] danah boyd and Kate Crawford. Critical Questions for Big Data. *Information, Communication & Society*, 15(5):662–679, June 2012.

[63] Craig M Dalton, Linnet Taylor, and Jim Thatcher (alphabetical). Critical Data Studies: A dialog on data and space. *Big Data & Society*, 3(1):2053951716648346, June 2016.

[64] Philip Hunter. The reproducibility "crisis". *EMBO reports*, 18(9):1493–1496, September 2017.

[65] Lauren F. Klein and Catherine D'Ignazio. *Data Feminism*. The MIT Press, 2020.

[66] Y. L. Wong, K. Madhavan, and N. Elmqvist. Towards Characterizing Domain Experts as a User Group. In *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*, pages 1–10, October 2018.

[67] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics (InfoVis)*, 18(12):2431–2440, 2012.

[68] Jessica Hullman, Paul Resnick, and Eytan Adar. Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering. *PLOS ONE*, 10(11):1–25, November 2015.

[69] Lace M K Padilla, Sarah H Creem-Regehr, and William Thompson. The powerful influence of marks: Visual and knowledge-driven processing in hurricane track displays. *Journal of Experimental Psychology: Applied*, 26(1):1–15, 2020.

[70] Jimmy Moore, Pascal Goffin, Jason Wiese, and Miriah Meyer. Exploring the Personal Informatics

Analysis Gap: "There's a Lot of Bacon". *IEEE Transactions on Visualization and Computer Graphics*, 28(1):96–106, January 2022.

[71] Lyn Bartram, Michael Correll, and Melanie Tory. Untidy Data: The Unreasonable Effectiveness of Tables. *IEEE Transactions on Visualization and Computer Graphics*, 28(01):686–696, January 2022.

[72] National Academies of Sciences. *Reproducibility and Replicability in Science*. National Academies Press, September 2019.

[73] Benjamin J. Heil, Michael M. Hoffman, Florian Markowetz, Su-In Lee, Casey S. Greene, and Stephanie C. Hicks. Reproducibility standards for machine learning in the life sciences. *Nature Methods*, 18(10):1132–1135, October 2021.

[74] Jeffrey Heer, Jock Mackinlay, Chris Stolte, and Maneesh Agrawala. Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '08)*, 14(6):1189–1196, 2008.

[75] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Nicola Cosgrove, and Marc Streit. From Visual Exploration to Storytelling and Back Again. *Computer Graphics Forum (EuroVis '16)*, 35(3):491–500, 2016.

[76] Kiran Gadhave, Jochen Görtler, Zach Cutler, Carolina Nobre, Oliver Deussen, Miriah Meyer, Jeff M. Phillips, and Alexander Lex. Predicting intent behind selections in scatterplot visualizations. *Information Visualization*, 20(4):207–228, October 2021.