Intra, Extra, Read all about it! How Readers Interpret Visualizations with Intra- and Extratextual Information

Alyxander Burns
Claudia Gonzalez-Vazquez
ajburns@mtholyoke.edu
Mount Holyoke College
South Hadley, Massachusetts, USA

Abstract

A reader's interpretation of a visualization is informed by both intratextual information (the information directly represented in the visualization) and extratextual information (information not represented in the visualization but known by the reader). Yet, we do not know what kinds of intra- and extratextual information readers use or how they integrate it to form meaning. To explore this area, we conducted semi-structured interviews about four real-world visualizations. We used thematic analysis to understand the types of information that participants used and diffractive reading to reveal how participants blended intra- and extratextual information. Our thematic analysis showed that participants utilized a broad assortment of information from both expected and unexpected sources. Additionally, our diffractive reading exposed three ways that participants incorporated intra- and extratextual information: to decide what to look at, to make (in)accurate assumptions about what the visualization showed, and to discover insights beyond what was directly encoded.

CCS Concepts

• **Human-centered computing → Empirical studies in visualization**.

Keywords

Data Visualization, Critical Data Visualization, Interpretation, Intratextual Information, Extratextual Information

ACM Reference Format:
Alyxander Burns and Claudia Gonzalez-Vazquez. 2025. Intra, Extra, Read all about it! How Readers Interpret Visualizations with Intra- and Extratextual Information. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3706598.3713612

1 Introduction

It is well understood that a reader's interpretation of a visualization is not solely dependent on the design or content of that visualization. Instead, a reader's interpretation (what they think a visualization means or shows) is formed by both the information directly represented in the visualization and information which is not represented in the visualization but known by the reader. For instance, past work has observed how information such as personal insights into, and pre-existing beliefs about, the topic of a visualization can anchor readers' interpretations of what it shows and how much it matters [47, 53, 64, 75].

Yet, while we know *that* visualization readers are using both types of information when interpreting a visualization, we do not know which varieties readers frequently deploy or how they integrate those spheres of information to make meaning while reading a visualization. Communication and feminist theory teach us that peoples' existing knowledge and experience forms a lens through which they interpret the world [31, 32]. However, we also know that people do not apply every piece of information they know to every situation or in the same way. Better understanding both what information readers employ and how they do so can help us understand how even similar-seeming people can end up with divergent interpretations of the same visualization.

In textual analyses, the information which comes from inside of the text is known as **intratextual** information and the information which comes from outside is known as **extratextual** information. In that context, it is well understood that both intratextual and extratextual information are important components of how meaning is formed when a reader interprets a text [13]. Understanding the extent to which a person utilizes one or both types of information is one way to understand how people can end up with divergent interpretations of the same text (e.g., in Law [2, 73]). We use this intratextual/extratextual framing as a lens in this paper because, while texts and visualizations are certainly different mediums, they both are polysemic and have conventions which must be learned to be interpreted, and thus their interpretation may be similar.

To explore this area, we conducted semi-structured interviews with six college students. Our study investigated two central questions: What kinds of intra- and extratextual information do participants reference while interpreting each visualization? And, how do participants integrate intra- and extratextual information to form their interpretation? We showed participants about four real-world communicative visualizations.

To analyze our data, we interpreted our data with two epistemologically different analysis techniques. First, we conducted a thematic analysis [16, 17] of the interviews to reveal repeated categories of information mentioned by participants throughout their time with the visualizations. This method helped us to find and categorize points raised by participants which aligned with the broad themes we aimed to investigate in this study: intra- and extratextual information. Then, we occupied a feminist position and conducted a diffractive reading [10, 56] of the interview transcripts in order to

CHI '25, April 26-May 1, 2025, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713612

This is the authors' pre-print copy of the paper; see https://doi.org/10.1145/3706598.3713612 for the official ACM version.

reveal different ways that participants integrated intra- and extratextual information to form meaning. This technique allowed us to focus on and create knowledge about the details and contradictions which might otherwise be lost when centering broad, macro themes (as thematic analysis does) [10, 56].

Our thematic analysis revealed that participants favored global descriptions of what was in the visualization, used a broad assortment of extratextual information from both dominant and non-dominant sources of knowledge. Additionally, our diffractive reading highlighted three different ways that participants combined intra- and extratextual information: as a guide for what mattered most, as a means to skip text and fill in perceived gaps, and as a means of making conclusions beyond what was strictly visualized. While past work has observed the impact of prior information and experience on interpretation, the articulation of what is used and how they are integrated is unique to this work.

There are two main contributions of this work. First, we contribute an analysis of the intra- and extratextual information used by participants while interpreting unfamiliar visualizations. Second, we contribute three narratives of how participants integrated intra- and extratextual information: from deciding what to look at through making judgments about the message and worth of a visualization. As a whole, our work is an exploratory step toward better understanding how people make sense of visualizations and why similar-seeming people can end up with very different interpretations of the same visualization.

2 Background

Many people have observed the ways that intra- and extratextual information (collectively) inform visualization interpretation. It is well understood that visualization readers often come away with different interpretations of the same visualization (e.g., [12]). Communication theories (such as Hall's encoding-decoding model) describe that a reader's experiences form a lens through which they perceive media and ultimately derive a message [31]. One implication of this idea is that the message received by the reader is a combination of what is in that media and what the reader brings themselves [31]. A different way of viewing this same phenomena is through the lens of rhetoric: rhetorical choices made by the creator of a visualization about how to present intratextual information determine which interpretations are most likely to be received, while extratextual information informs how those rhetorical choices are understood and metabolized by the reader [35]. Therefore, to understand how visualization readers come to the conclusions they do, it is critical to carefully examine both the intra- and extratextual information involved.

A substantial body of visualization research has explored how intratextual information impacts interpretation. The intratextual information in a visualization is composed of both the data itself and the manner in which it is encoded. While different data obviously produce different interpretations, past work has also experimentally shown that a variety of design factors also can influence interpretation. Some of these factors include the content and style of annotation [24], the title and its alignment to the evidence in the visualization [46], the use and selection of pictographs [7, 21], and the colors used [34, 39, 53, 72], among others. To control for what intratextual information was available in our study, all of our participants saw the same four visualizations in the same format. However, as our results ultimately reflect, equal access to information does not necessarily mean equal *use* of that information.

While less well explored than intratextual information, existing work has also investigated the impact of extratextual information. One type of extratextual information which is comparatively well studied is readers' knowledge of how to decode specific visualization types. This work has noted, for example, how different amounts of knowledge change what strategies readers use to understand visualizations and what kinds of messages they extract [15, 48, 71].

Another type of extratextual knowledge which can impact the way that people interact with visualizations is prior knowledge about the underlying data set or domain. Past work has shown that when domain experts possess "data hunches" [50] or other types of knowledge about implicit errors of the dataset, they may choose to use different kinds of tools [70], or be hesitant to engage with visualization tools at all, when the visualization does not align with their own intuition [57]. However, these kinds of ambiguities are an inescapable part of many datasets and working with them is an unavoidable part of the sensemaking process [61]. Therefore, rather than try to remove this kind of ambiguity, a body of past work has explored different strategies for, and the impact of, making hunches, intuitions, and implicit errors visible (e.g., [50, 61, 63]).

More broadly, there has been substantial work exploring the impact of what readers expect the data to show on their interpretation of what a visualization means. For instance, Xiong et al. observed that priming individual readers with particular interpretations can anchor their attention to specific features of the visualization [75]. Viewing this phenomenon from a Bayesian reasoning perspective, readers, in aggregate, have been seen to update their beliefs in line with Bayesian reasoning predictions for small visualizations, though this effect diverges for larger ones [44]. Further, Kim et al. and Hullman et al. have explored how eliciting a reader's priors, or the priors of others, and making them visible increases recall accuracy [42, 43] and, in the context of visualizations of scientific results, enables readers to make more accurate predictions about potential replications [36].

We note that while factors like individual personality traits and cognitive abilities are extratextual, and may influence interpretation [40, 51, 64, 76], these factors are not the focus of this paper. Instead, in our study, we asked about and focused on *information* (i.e., knowledge or facts) known by readers because they can be identified in speech more concretely than emotion or personality, which may need to be inferred.

People have noted how intra- and extratextual information can be combined to both benefit and harm the reader. One area of research where this relationship has been observed is the study of how visualization readers approach unfamiliar visualizations. For instance, Rezaie et al. and Lee et al. observed that readers were sometimes able to use their existing knowledge to make helpful assumptions and overcome a state of "floundering" [48, 68]. Unfortunately, these attempts were not always successful and sometimes led readers to jump to conclusions unsupported by the visualization [68]. While we did not focus specifically on unfamiliar visualizations in this study, we did select visualizations which we expected would be of varying familiarity to our participants in order to see

This is the authors' pre-print copy of the paper; see https://doi.org/10.1145/3706598.3713612 for the official ACM version.

if they would employ different types of information while reading. Further, while we take a similar methodological approach to explore how intra- and extratextual information is integrated as a means to identify further opportunities for study, our focus is specifically on the information employed, rather than struggles and strategies for getting unstuck.

3 Methodology

To explore how visualization readers combine intra- and extratextual information while interpreting a visualization, we conducted semi-structured interviews with undergraduate students about four data visualizations.

3.1 Participants

For this study, we recruited six students from a small, residential, historically white undergraduate college in the United States. We decided to sample students as participants because they represent a fairly homogeneous group in terms of age and educational experience. As many scholars have observed, undergraduate students are not representative of almost any large "general" public (e.g., [18, 28]), but the goal of our qualitative study was not to produce generalizable results and we had not designed our study to do so. Because the relationship between intra- and extratextual information use has not been explored before, our aim was to describe the diverse information these participants employed and the ways that they put them together to produce unique insights, despite seeming to be a relatively homogeneous group. For this reason, we chose to survey a small convenience sample as a means to initially and deeply examine how this set of individuals engaged with visualizations. Though the number of participants is small, the data collected from each participant is information rich and involves an hour of discussion per participant for a total of 527 question/answer pairs. However, as a result of this design, our results may not be representative of more diverse populations or larger groups of students and discuss these possibilities further in the Discussion.

Though the call for participants was broad across campus, volunteers were predominantly first-years and sophomores and all had declared, or intended to declare, STEM majors (i.e., Computer Science, Neuroscience, Environmental Science, Psychology, and Biology). All six participants used she/her pronouns and were not asked to define their gender identities, though three disclosed during their interview that they identified as women. Participants were invited to provide their own pseudonyms to be used in the paper. Three participants provided their own pseudonyms. The authors selected pseudonyms for the remaining three participants in a style which matched the three selected by participants. A summary of information about each participant is provided in Table 1. Due to the size of the campus, we have not included each participant's major(s) in the table in order to better protect their privacy.

3.2 Procedure

We used a semi-structured interview format consisting of demographic questions and then discussions about four pre-selected communicative data visualizations. At the start of the interview, we told participants that they would be shown a series of four visualizations (some of which may be unfamiliar) and asked a series of questions about them. They were told that we were interested in their interpretations of each visualization and assured that there were not right or wrong answers to the questions. We then asked participants demographic questions about their age, major, and pronouns. Then, we asked participants to describe their relationship to data visualization. We had participants describe their relationship to visualizations, rather than rate it, in order to more holistically understand how these participants understood the role of visualizations in their lives. We have included short quotes from each participant describing their experiences in Table 1. In summary, all six described experience reading or creating visualizations with differing amounts of confidence, and five had done so as part of one or more of their college courses.

After answering the demographic questions, participants were presented with one of the four visualizations printed onto legal-sized paper and were invited to take as much time to look at it before providing initial thoughts. Participants were not given a specific task, but asked to "look at the visualization." They were invited to either think aloud immediately or to read quietly and indicate when they were ready to discuss. If they chose not to think aloud, they were prompted to talk through their initial thoughts when they were ready. Participants were asked the same three questions about every visualization. These questions were selected in order to collect different aspects of understanding from participants and were drawn from past work [11, 19, 74]. The three questions were:

- What is this visualization about?
- Do you think the author of this visualization is trying to communicate a message to you through it? If so, what?
- Is there anything that you find particularly interesting or surprising in this visualization? If so, what?

For each question, participants were asked follow-up questions to clarify statements made, explore interesting or unique points more deeply, and ask about intra- and extratextual information which the participant found relevant (if they had not mentioned it in their original answers). Specifically, participants were prompted to clarify or expand on ideas with phrases like "I heard you mention X, could you tell me more about that?," "When you say that the chart says X, what in the chart is telling you that?," and "Is there anything outside of this visualization, for example that you knew before seeing this chart, that you think informed your answer?" Once the participant had answered all of the questions, they were presented with the next visualization. Every participant saw all four visualizations and the order that the visualizations were presented was balanced using a Latin square design to obscure ordering effects.

After the interview, the audio was transcribed with a transcription software and then manually corrected by the interviewer (the first author). Participants were invited to see the transcript of their interview and provide any comments or corrections, prior to data analysis. Five participants opted to see and comment on their transcripts and either confirmed the accuracy of the transcript or provided no comments at all. All participants were compensated with $15 USD for their time at the conclusion of their interview.

3.3 Stimuli

During the interview, we showed each participant four real-world visualizations (Figure 1). Past work has observed that people with

This is the authors' pre-print copy of the paper; see https://doi.org/10.1145/3706598.3713612 for the official ACM version.

Participant	Age	Relationship to Visualization
Iman	21	*"I am familiar with [visualizations] as I've had to look at graphs for classes, but it's not intuitive to me."*
V	19	*"I'm currently taking [a Statistics class], so I'm working with data visualizations. I haven't gone farther."*
Noor	19	*"I'm taking a Data Science class ... We are learning to create simulations and generate graphs."*
Luna	20	*"I can appreciate the art of data visualization when it's not super complicated ... I can analyze a good graph."*
Orla	20	*"I mostly see data visualizations in Math classes and Statistics classes. I like [visualizations] because it's easier to interpret data when it's on a graph.'*
W	20	*"I definitely have used some [cartographic maps] before ... I can get something from the visualizations, even though I'm not a person who is very familiar with the topic before."*

Table 1: We interviewed six undergraduate students at a small, residential college in the United States. Three participants created their own pseudonyms and the authors assigned pseudonyms to the rest in a style that matched the participant-selected pseudonyms.

different amounts of familiarity with a visualization's topic, can receive different kinds of meaning from it [71]. Therefore, we selected the set of visualizations to differ along two dimensions: familiarity with the topic and familiarity with the visualization's encoding. While we selected visualizations that we believed would represent these differences, familiarity is subjective and our expectations did not always align with participants' experiences. In particular, while Post-Grad, Nobels, and Languages largely matched our predictions, we expected participants to be more familiar with the 24-hour clock visualizations used in Routines than they were.

All four visualizations were real-world, static, communicative visualizations. Three of the four visualizations were award-winning Information is Beautiful projects [26]. We selected Information is Beautiful projects to ensure that the visualizations used in the study were high quality. While we endeavored to select visualizations with differing levels of topic and encoding familiarity based on the authors' best judgment, we were not certain what our participants would know. Therefore, we selected a visualization published by the college attended by the participants as the fourth visualization to ensure that every participant saw a visualization with a very familiar topic. The four charts (shown in Figure 1) were:

- **Post-Grad** [59] visualizes the graduation outcomes of college alums six months after graduation, including their employment status, location, employment sector, and employer. The visualization uses a series of small charts including a table, proportional area chart, bubble map, and bar chart. The visualization that participants saw was a screenshot of the original interactive web page and was not modified to hide interactive features (e.g., by hiding available filters).
- **Languages** [52] features a large, circular area chart in the center that visualizes the size of languages with more than 50 million native speakers and the countries in which those native speakers live. There are also a series of smaller bar charts and a proportional area chart at the bottom with other information about world languages and their distributions.
- **Routines** [5] uses a series of 24-hour clocks to visualize the daily routines of 15 famous people. There is a photograph of each person in the middle of each clock and different types of activities are represented by color-coded arcs along each clock's edge including Primary Work, Sleep, and Exercise.
- **Nobels** [54] visualizes the demographics of Nobel laureates over the history of the six award categories. It has a central composite visualization made up of a series of line charts (of age and gender) connected to bar charts (about highest degree obtained) which lead into a Sankey diagram (about university affiliation). Below, there are stacked bar charts about the hometowns of laureates.

3.4 Data Analysis

We analyzed our results with two qualitative data analysis techniques: thematic analysis and diffractive reading. Our study aimed to answer two central questions: What kinds of intra- and extra-textual information do participants reference while interpreting each visualization? And, how do participants integrate those pieces of information during their interpretation? We explored each of these questions through a different qualitative analysis technique which best matched the kind of question. For the first, we used thematic analysis with open coding to explore broad, categorical patterns. For the second, we used diffractive reading to identify narratives and unique differences. These two analysis techniques are qualitative analysis approaches which have appeared in HCI and visualization literature before and occupy complementary epistemological positions. The choice to use these contrasting styles was an intentional one: different styles offer distinct insights and tools for answering questions. We selected approaches which had appeared in relevant literature before and that we felt would best let us answer our central research questions. We describe each analysis technique at the start of the section in which its results are discussed (Section 4 and 5, respectively).

This is the authors' pre-print copy of the paper; see https://doi.org/10.1145/3706598.3713612 for the official ACM version.

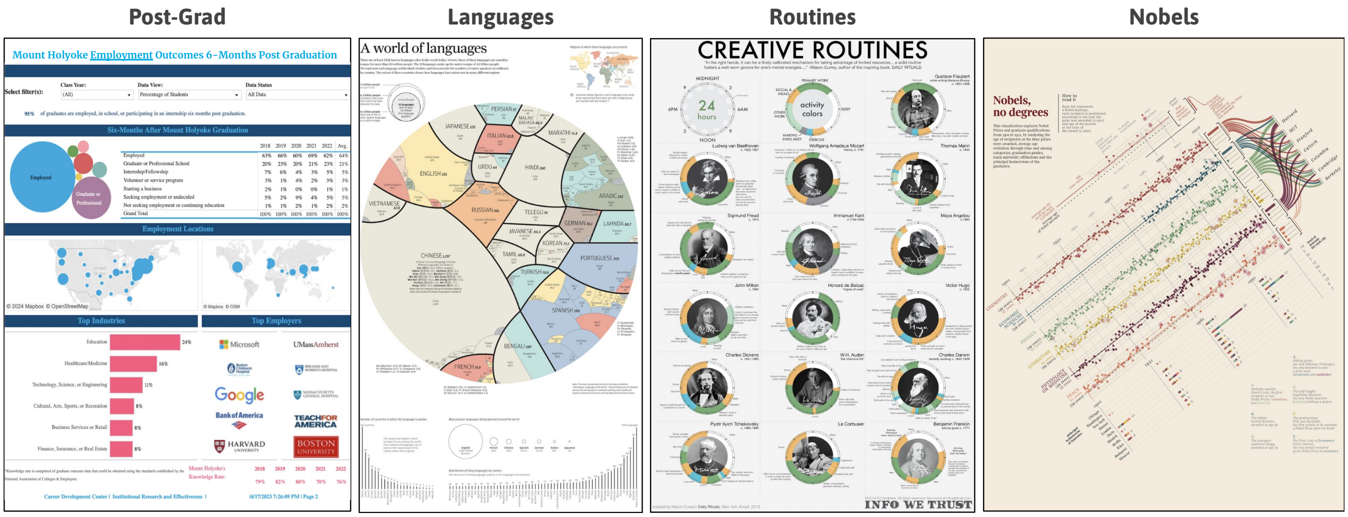

Figure 1: Participants were interviewed about four real-world visualizations. Based on their topics, we call these visualizations (from left to right): Post-Grad, Languages, Routines, and Nobels. Full-quality versions of the stimuli are available at: https://osf.io/fjznc/

4 What Kinds of Information Did Participants Use?

4.1 Analysis Approach: Thematic Analysis

To understand what kinds of knowledge participants utilized to interpret the visualizations, we conducted a thematic analysis of the interview transcripts. We divided each transcript up into question-answer pairs (to maintain context for the participants' answers) and applied codes to the participants' answers.

The codebook was developed and refined over a series of seven discussions between the two authors. We knew that we wanted to analyze both intra- and extratextual information, so we initially seeded the codebook with high-level themes for these two broad categories. Then, we iteratively generated and honed lower-level codes based on what emerged. We decided to use a style of thematic analysis which used top-down themes and bottom-up codes because it allowed us to flexibly generate codes that reflected what we saw in the data, while focusing our analysis on features of the data which were closely related to our research questions.

Codes were applied over four stages. First, the authors independently coded the same hour-long interview, then met in order to discuss the codes they had applied, analyze disagreements, and adjust the codebook as needed. Then, the authors coded different interviews and met again. These steps were repeated (i.e., the authors coded the same interview, and then separate interviews). The intercoder reliability over the two overlapping interview transcripts was 0.961 as calculated with an unweighted Cohen's Kappa. Table 2 includes a summary of the codes and themes. The entire codebook is available in our OSF repository: https://osf.io/fjznc/.

We initially imposed a top-down organization which divided our codes into two broad categories (Intratextual and Extratextual), but, after discussion, noticed that the Extratextual codes could be further divided based on the kind of epistemology that they represented: **dominant** epistemologies and **non-dominant** epistemologies. As a whole, epistemology describes how people know what they know [23, 45]. Dominant epistemologies are those which are valued and legitimized by dominant cultures (e.g., involving constructing knowledge through controlled experimentation) [23, 33]. In contrast, non-dominant epistemologies are those which are valued by marginalized or non-dominant groups such as women, non-European people, and indigenous people. These epistemologies are very diverse but could involve theories such as the construction of knowledge through community participation or embodied experience [23]. While our results in the following sections are organized by theme, we have included a heatmap visualizing the distribution of codes across each of the four visualizations used in the study (see Table 3).

4.2 Readers Mentioned Global Features More Often Than Specific Values (Category: Intratextual)

Participants referenced four different types of intratextual information when interpreting the visualizations. Of these, two represent (high-level) global features of the visualization (i.e., variables and comparisons) and two represent (low-level) specific values (i.e., encoded values and derived values). Participants much more frequently mentioned the high-level features than low-level ones.

By far, the most frequent type of intratextual information was in reference to one or more of the **variables** in the visualization (206 statements). When commenting on variables, participants often focused on a subset of the variables or levels and ignored the others. For instance, when discussing Nobels, participants rarely mentioned degrees at all (6) and instead focused heavily on university affiliation (28), prize categories (28), and the gender of laureates (24). We

Themes	**Codes** (# of responses code was applied to)	Total
Intratextual	Variable (206), Comparison (60), Derived Value (18), Encoded Value (16)	300
Extratextual – Dominant	General Knowledge (85), Statistical Principle (38), Historical (15), Commonness of Vis. Type (12), Location (10), Statistic (10)	170
Extratextual – Non-Dominant	Life Experience (47), Common Wisdom (29), Personal Identity (23)	99

Table 2: We generated 13 codes using open coding to understand what kinds of intra- and extratextual information participants mentioned while interpreting the visualizations. The themes of Intratextual and Extratextual Information were pre-selected, while the distinction between the use of Dominant and Non-Dominant knowledge emerged during analysis. The final column of this table indicates the total number of times that codes in each theme were applied.

observed this pattern repeated across all of the visualizations, with some variables discussed far more frequently than others. While some of this different amounts of focus may be attributable to design choices, we also saw evidence that differences in attention was also related to participants' existing knowledge. This was one of the central ways that participants integrated intra- and extratextual information (discussed in Section 5.2).

The second most frequent type of information drawn from the charts was a **comparison** between values in the visualization (60). Many of these comparisons represented statements about trends and often took the form of "X is more/less than Y," though these did not always map to maxima or minima. For instance, Orla made a series of comparisons while describing what she initially noticed about Nobels: *"It's clear that there's a lot more men winning the Nobel Prizes than women. And the age disparity also. It's mostly just people over 50."*

Participants also sometimes, but much less frequently, talked about specific, numeric values which were either an **encoded value** (16) directly from the visualization or a **derived value** (18) obtained from adding or subtracting two directly encoded values. Combined with the frequent mentions of variables, the large number of comparisons may suggest that participants often focused their attention on higher-level features instead of specific values when reading.

4.3 Readers Employed a Diverse Set of Extratextual Information (Theme: Extratextual – Dominant)

Perhaps unsurprisingly, **general knowledge** (85) was the most frequently employed type of extratextual information across all visualizations and participants. We used the term "general knowledge" to refer to descriptive and observable information which is reasonably available to participants but not present in the visualization (e.g., "Brazil was colonized by Portugal" and "Chemistry is a male-dominated field"). While this type of information was frequently employed, the specific pieces of information differed substantially. For instance, Iman brought up her knowledge of the names of languages and countries when reading Languages, Noor used the geographic location of prominent cities to interpret Nobels, and W used knowledge of the trades of the people in Routines.

The other five codes in this theme represent more granular types of information employed by participants. Two of these types of extratextual information were related to fields of social science (geographic **locations** (10) and **historical** events (15)), while the other three were related to familiarity with Statistics and visualization (information about **statistical principles** (38), the **commonness of visualization types** (12), and specific **statistics** (10)).

While most historical information, unsurprisingly, arose while participants interpreted Routines (7/15), we were surprised to see that participants most frequently mentioned geographic information while interpreting Nobels (7/10). Further, we observed that several participants used geographic information to identify the same trend among the hometowns of Nobel laureates: while laureates historically were from Europe, now most laureates are from the Americas. Counter-intuitively, participants almost never mentioned extratextual geographic information when interpreting Languages (1/10), which already contains lots of information about countries and continents. It is possible that because Languages already contains so much intratextual information about geography, participants may have not had more to add (or thought any more was necessary).

Regarding the use of statistics, participants mentioned many different kinds of statistical principles and processes over the course of the interview. Specific concepts employed included filtering and sorting data, correlations, outliers, and skew. We observed that participants often did not call attention to the act of using this kind of information directly, but integrated it directly into their speech naturally. For instance, when describing what she noticed about Nobels, Iman offered: "*They kindly put some outliers on the bottom for us to read about.*" Likewise, the (perceived) commonness of visualization types was often mentioned in order to contrast the visualization they were currently viewing against something else they had seen before. One example of this is the way that Noor contrasted the area chart in Languages against a pie chart:

> Noor: *It's definitely way more interesting than normal pie charts ... The pie chart is definitely more direct ... But for this, the longer time you look at it, the more information you can get from it by your own analysis. But for pie charts, they're all there, you can just read it. And it's like, 'yeah, so what?'*

It is worth noting that participants' perceptions of what was or was not a common visualization did not always match existing work (e.g., [66]). Instead, it is possible that the "commonness" mentioned

This is the authors' pre-print copy of the paper; see https://doi.org/10.1145/3706598.3713612 for the official ACM version.

	Intratextual				Extratextual - Dominant						Extratextual - Non-Dominant		
	Variable	Comparison	Derived Value	Encoded Value	General Knowledge	Statistical Principle	Historical	Common-ness	Location	Statistic	Life Experience	Common Wisdom	Personal Identity
Routines	52	16	8	4	26	8	7	2	0	2	11	14	1
Nobels	57	18	4	3	22	14	3	3	7	2	11	1	7
Post-Grad	48	13	6	4	21	7	2	2	2	0	22	10	9
Languages	49	13	0	5	16	9	3	5	1	6	3	4	6

Table 3: The distribution of codes was similar across the four visualizations, though there were some differences, particularly among codes in the Extratextual – Non-Dominant theme.

was a proxy for their own exposure: if they were familiar with that type of visualization, they assumed others were too.

4.4 Readers Often Used Information Drawn from their Own Lives (Theme: Extratextual – Non-Dominant)

In addition to the kinds of dominant information that one might obtain in a classroom, participants also drew lots of information from non-dominant sources: their **life experience** (47), **common wisdom** (29), and their **personal identities** (23).

The broadest category of information which was employed by participants which represents a non-dominant way of knowing was **common wisdom** (29). We applied the code for common wisdom to all statements which referenced commonly-accepted advice or opinions about how to live or be (e.g., "You need to go to college to get a good job" and "It is good to have a balance between work and rest"). While general knowledge was consistently mentioned across all four visualizations, common wisdom was most frequently employed in response to Routines (14/29). When interpreting that visualization, participants often mentioned ideas like work-life balance, what a "perfect" schedule looks like, and the relationship between a person's priorities and what they spend their time doing. Participants sometimes integrated both general knowledge and common wisdom together (11). For instance, Orla used general knowledge of the kind of work Charles Darwin did, combined with common wisdom, to reason about how his segmented routine might represent an "ideal" schedule:

> Orla: *When you look at Charles Darwin's– he seems to have been doing a lot. I think that's interesting because he studied human evolution, right? So he probably knew what he was doing ... People usually say we should have a little bit of everything ... And people are always saying you can't do a long stretch of something. You have to take breaks in between and you have to do other things. And that's what he seemed to have been doing.*

Participants most often referred to life experiences when interpreting Post-Grad (22/47), though they were mentioned across all visualizations. This is perhaps unsurprising given that Post-Grad was specifically selected because we knew that participants would have substantial experience with its topic. The information they attributed to life experience differed, as did the experiences themselves. For instance, participants talked about learning through internships, on-campus jobs, and experiences they had when deciding what college to attend.

Similarly, most of the references to personal identity came in response to Post-Grad (9/23), however they too were present across all four visualizations. For instance, Luna invoked both her gender identity and identity as a STEM student when asked what was most interesting or surprising about Nobels:

> Luna: *[I'm surprised by] how little women winners there are. And how most of them are in the Peace category, I believe. It's either Peace or Literature, so I'm just like, as a woman in STEM, I wish there was more women in STEM.*

Many of our participants mentioned their gender identities and connections to STEM while interpreting the visualizations, but also brought up an assortment of other identities such as identifying as an immigrant and a native speaker of languages such as English, Korean, and Chinese.

5 How Did Readers Integrate Intra- and Extratextual Information?

5.1 Analysis Approach: Diffractive Reading

To understand the ways that participants integrated intra- and extratextual information, we used diffractive reading. Diffractive reading is a feminist qualitative analysis technique which is used to notice and create knowledge about the details and contradictions which might otherwise be lost when centering broad, macro themes (as thematic analysis does) [10, 56]. Therefore, we decided to use diffractive reading as our central analysis technique to understand participants' approaches to using information because we were primarily interested in the nuanced, different ways that our participants approached the visualizations. We felt that diffractive reading would allow us to focus our discussions around the components which we found most interesting or insightful. Our use of this approach was inspired by Akbaba et al.'s project on collaboration and matters of care [1], which similarly uses diffractive reading to understand the differences among interviewees, situated in the unique experiences of each participant. For further examples of diffractive analyses in the visualization and human-computer interaction literature, see [9, 27, 67, 69].

The three narratives presented in the following section were selected after extensive discussion between the authors, rooted in our own experiences and positionalities. Over the course of seven meetings, the two authors spent over ten hours discussing what we noticed while coding each of the transcripts, bringing quotations to discuss and compare, and contrasting different approaches to the same visualizations. We found that analyzing each transcript from the different positions occupied by the two authors allowed new

This is the authors' pre-print copy of the paper; see https://doi.org/10.1145/3706598.3713612 for the official ACM version.

insights to emerge, which we then captured in notes, participant quotes, and mind-maps. The final three narratives provide examples of different ways that our participants integrated intra- and extratextual information while making sense of the visualizations throughout the reading process.

5.2 Using Extratextual Information to Highlight Intratextual Information

One way that participants used extratextual information was as a guide toward elements of intratextual information which may not have been otherwise emphasized by the design of the visualization. While participants employed this strategy with many types of extratextual information, they seemed most aware of it when employing their personal identities and life experiences.

For instance, many of our participants identified the gender of laureates as one of the most important components of Nobels, even though gender does not have a lot of visual weight. Visually, the (binary) gender of laureates is represented by either dots (for men) or circled dots (for women). Yet, every participant commented on the gender of laureates during their interview at some point and it was among the top three most frequently discussed topics for five of the six participants. For instance, Orla explored different aspects of gender across the chart including the increase in winners in the 1960s and the sole woman winner of the Economic Sciences prize. She speculated that her focus on gender might be related to her own identities:

> Orla: *Maybe this is just because of my demographic, but I think the thing that stands out to me the most is the gender disparities in the winners of the prizes. And I don't know what that attribute is caused by, so I won't read into it, but that's just what stands out to me.*

Participants focused on areas which connected directly with extratextual information across all of the the four visualizations. For example, W made a similar connection as Orla had to Nobels, when she explained the way that her personal experience impacted the way she looked at Languages:

> W: *I feel like I am able to build some connections as well because my mother tongue is also listed on this graph. And I would tend to look at that language at first. And I think that might be true for other people as well because they might find their most familiar language is listed on this map.*

Orla also made connections between her own experiences and areas of interest in Routines, but connected it instead to her life experiences of cultural norms in her community. For Orla, the first thing that popped out of Routines was the amount of sleep that the people in the visualization got. She expressed surprise that they slept as much as they did, explaining that this reaction was rooted in her experience as a STEM major:

> Orla: *Most of the culture now, at least in my department and other STEM majors, people always love to talk about how much sleep they didn't get. They're like, 'oh my God, I'm running on 3 hours of sleep today.' And then like when someone gets a really good test score, they're like, 'yeah, I pulled all nighter' or 'I didn't sleep, that's why.' But then also in this chart, you see these people are people who have accomplished a lot in their lives, but also it's a different time period. So I don't know if that says anything, it probably does.*

Similar to gender in Nobels, sleep is not particularly emphasized in Routines, but yet became a centrally important component for Orla, Noor, and Iman when combined with their own experiences.

5.3 Using Extratextual Information to Skip Intratextual Information

In opposition to using extratextual information to highlight intratextual information, participants also used it as a shortcut – as a means to skip some intratextual information entirely. In particular, we observed that participants both used this strategy productively and unproductively: some made accurate assumptions based on what they already knew, while others made inaccurate assumptions which led them to fall prey to visual mirages [58].

We observed that participants sometimes used extratextual information to make *accurate assumptions* about the visualization and avoided reading the visualization in depth. For instance, Iman explained that she combined the information in Language's title with extratextual information to quickly infer what it was about:

> Iman: *Well, I did skim. For this one, I just kind of read the title and that kind of filled me in with enough context clues. The big bolded [labels], Chinese, Vietnamese, English, I know those are languages. Little smaller subdivisions like India, Bangladesh, Colombia, those are countries and the rest kind of pieced itself together real quick after that.*

One circumstance in which participants repeatedly made successful assumptions was while inferring what the people in Routines had in common (or why they had been included). None of our participants reported recognizing all fifteen of the people represented, but they used their knowledge of the people they knew in order to make guesses about the rest. For instance, V reported not recognizing most of the famous figures, but used the ones she did to infer that *"there's a relationship between them being either scientifics [sic] or artists from a while ago."*

On the other hand, participants also used extratextual information to make *inaccurate assumptions* which, in most cases, were not corrected by the participant. The most frequent example of this in our study was participants' (mis)interpretations of the universities in Nobels. Although our participants were all able to decode all of the visualizations quite accurately, there was one persistent error: on the right side of Nobels, there is a list of seven universities which is described in the chart as "Principle university affiliations of Nobel laureates at the moment the Prize was awarded." However, V and W erroneously believed that the institutions represented the places that the laureates had graduated from. In V's words, the chart described *"the university [the laureates] studied in"*. When asked what in the chart was telling her this, she replied:

> V: *Well, on the right side it's not specified but by seeing names of top colleges, I am assuming we're talking about the percentage between these people and these college,*

This is the authors' pre-print copy of the paper; see https://doi.org/10.1145/3706598.3713612 for the official ACM version.

> *which I'm assuming it means like this percentage of people went to these places.*

Here, V described seeing the university names and assumed, based on that information, that they must represent the places where the laureates graduated from (instead of where they were affiliated at the time of their award).

While some inaccurate assumptions went unchecked, a few participants were able to correct their inaccurate assumptions to gain new insights. When interpreting Nobels, Noor initially made the same error that V and W did; she assumed that the colleges represented where laureates graduated from. However, she corrected this mistake almost immediately. When asked about her correction, she responded:

> Noor: *Well, because I think based on my prior knowledge about this, most news, when they introduce a Nobel Prize winner, they usually say where they graduated from. At first I thought this was the same, it's about where do they graduate from. This graph actually shows who they affiliate with. This is a new perspective as saying it. Because now we can just see which school [has] a stronger teaching strength in which category because they have Nobel prizes.*

Although she too felt that her prior knowledge had led her to the false assumption, Noor came to a new insight when she was able to correct her error: she realized that what was plotted complemented what she thought she already knew.

5.4 Using Extratextual Information to Read Deeper into Intratextual Information

We also observed that participants used extratextual information in order to "read between the lines" of the visualizations and come up with interpretations beyond what was strictly visualized. They utilized different types of extratextual information and this behavior often generated the most unique interpretations of what the visualizations meant and how they were relevant to the participants' own lives.

The participants who came away with messages beyond what was strictly visualized often saw connections between the data and their past experiences. They then used these connections to extrapolate life advice from the visualization. For instance, Noor summarized the message of Routines as:

> Noor: *Yeah, it's trying to tell me that even for those celebrities, they all have a different style of life. It's trying to tell me like 'it's not necessary for you to follow the scientifically best for yourself routine.' But instead, you can just always follow your own pace and sleep whenever time you want.*

Noor often actively compared the data in the visualization to her own experiences, remarking, for example, that she felt like she had a connection to Gustave Flaubert because: *"This is also what my daily routine looks like."*

However, they also used information from dominant sources of knowledge to derive new insights. One example of this was the way that Orla used (extratextual) historical information to derive a unique message from Languages about colonization. Immediately upon reading the visualization, Orla remarked:

> Orla: *I guess this chart just takes into account who has what language as a first language in the country. And for stuff like in English, the countries that we have in the little box are mostly, if not all, English colonies or British colonies ... We have some African countries who are also colonized by the British, and they also speak English. Obviously, we also have the United Kingdom, they speak English. And then we can see the same patterns for Portuguese ... And then for Spanish, it's the same story.*

This story of colonization was only discussed by Orla, but it is well supported by the points in the visualization. It is possible that Orla's unique prior knowledge enabled her to identify this message which was invisible or unremarkable to other participants.

Participants also used the combination of intra- and extratextual information to question the visualization itself. When interpreting Post-Grad, both Luna and W took issue with the "Top Employers" section at the bottom of the page because it listed the logos of institutions which employed alumni but did not provide any information about the actual number of alumni employed there. For Luna, negotiating the balance between the potential impact of missing information and prior knowledge on the source of the information was complicated. In the case of Post-Grad, she noticed that information was missing but was quick to balance her reaction against her existing knowledge of the people who made the visualization, saying:

> Luna: *I know how hard they work to make sure that everyone meets their goals after graduation. I know how hard they work in terms of their research as well. They are very good at gathering data about alums. They're in constant communication with alums. It's really helping the reliability of this graph.*

We can contrast this response against her response to the unknown reliability of information in Routines where she had no prior experience with the authors:

> Luna: *I don't really see a source anywhere. I'm looking for it and there's this little part 'inspired by the book Daily Rituals.' But I have no information on the credibility of that source. Like where did he get this information from? So I do look at this a little skeptically, you know, like how sure can we be that Kant went to the pub for like 3 hours and that was his only meal of the day? You know what I mean? He lived in 1764. Did he write this in a diary? Like what happened here?*

Here, Luna used what she knew about the time period to question where the data had come from, but had no prior exposure to the organization who made the visualization and so based her reasoning on its name alone. By using different kinds of extratextual information, she changed how she viewed the visualization as a whole: first she questioned the reliability of the visualizations' message and then reasoned about how well-placed that judgment was.

This is the authors' pre-print copy of the paper; see https://doi.org/10.1145/3706598.3713612 for the official ACM version.

6 Discussion

In this paper, we presented a semi-structured interview study with six college students to understand what kinds of intra- and extratextual information they discussed while interpreting communicative visualizations and the ways that they integrated those spheres of information. We utilized two different types of qualitative data analysis to explore what kinds of intra- and extratextual information participants mentioned during their interview and how they integrated it while interpreting the visualization. Our thematic analysis of the kinds of information mentioned by participants revealed that participants focused on information regarding global features of the visualization over specific values and utilized information reflective of both dominant and non-dominant ways of knowing. Additionally, our diffractive reading generated three ways that participants integrated intra- and extratextual information while reading the visualizations: they used extratextual information to decide what intratextual information was most important, made (in)accurate assumptions about what the visualizations showed, and read messages beyond those strictly encoded by the data. In the following section, we reflect on possible implications of these results.

6.1 Types of Intra- and Extratextual Information Used are Numerous and Complicated.

In our study, participants mentioned a diverse set of intra- and extratextual information types. Because we utilized a bottom-up coding style in our thematic analysis, our codes do not directly correspond to those used in past work. However, some do bear a resemblance to factors observed in existing work. In particular, all four types of intratextual information that we observed appear in existing models of visualization comprehension (e.g., readers consult the textual components of the visualizations to differing degrees [55, 62]; and compare points of interest and, infrequently, extract individual values [6]). On the other hand, a majority of the extratextual codes do not have such clear analogs in existing work. Notable exceptions to this are our participants' use of life experiences (which may be similar to the use of personal connections observed by Peck et al. [64]), and our participants' use of statistical principles (which may be similar to the kinds of mathematical skills noted in [29, 41, 49]).

As a whole, our results emphasize the complexity of the web of information which may be used by visualization readers. Designing visualizations that take into consideration what readers already know is important because the resulting visualization is more likely to be understood without undue burden [30]. However, predicting what information readers are likely to know is a challenge on its own and is further complicated by the fact that knowing *that* someone knows something does not mean they will find it salient in a given circumstance. Instead, what information is used to interpret new information is highly context-dependent [31]. Within the context of our study, for instance, we suspect that all of our participants know where cities such as New York, Paris, and Berlin are located, but only a few of them used this information to infer a trend among laureates' hometowns in Nobels. While we know that patterns of what information is used might be influenced by phenomena such as cognitive dissonance when charts contain information that contradicts existing beliefs [53], future work with a larger group of participants and visualizations may further explore how visualization readers navigate how and when to apply different types of extratextual information.

Our results also emphasize how important information derived from non-dominant ways of knowing were to participants' interpretations. Feminist scholars have long emphasized the value of recognizing non-dominant epistemologies because they expand both who is considered a valid knower and the methods through which someone can come to know something [3]. In our study, we observed that participants frequently mentioned information which likely came from non-dominant sources (e.g., drawn from embodied life experiences). While our participants did not mention information from non-dominant sources as frequently as information from dominant sources (99 vs 170), the prevalence of non-dominant knowledge use emphasizes that readers *do* use non-dominant information when interpreting visualizations. This is an important insight because it shows one way that participants subvert hierarchies of power while interpreting a visualization. Namely, although dominant hegemony poses non-dominant epistemologies as invalid sources of information, participants still viewed information from these sources as helpful and valid enough to frequently mention in their responses. This is particularly interesting given all of our participants were students at a historically white [14] college, which are spaces which value very narrow, dominant epistemologies [4]. Our results suggest that future visualization studies collect and consider the impact of information derived from non-dominant sources such as personal identity, life experiences, emotion, and word of mouth. Further, they suggest that visualization designers who wish to factor in what is known by their readers also consider information from non-dominant sources as well.

6.2 Integrating Intra- and Extratextual Information is Inevitable but Double-Edged.

During our study, we observed participants integrated information together in an assortment of (un)productive ways throughout the process of interpreting the four visualizations. While, we observed that our participants often integrated intra- and extratextular information very effectively, sometimes these attempts went awry and participants came to incorrect conclusions about the visualizations.

A recent study by Rezaie et al. which investigated struggles and problem-solving strategies of visualization readers calls this the "double-edged sword" of knowledge: while prior (extratextual) information can be a helpful guide, it can also lead readers to make incorrect assumptions [68]. As a corollary to this "double-edged sword," one question we can ask is whether readers with more extratextual information may fall for visualization mirages [58] more often (because they are quickly skimming or skipping components of the chart) or whether they experience more incidents of "mind-drift" while reading as they make connections between things in the visualization and what they already know [8]. Past work has explored the impact of exposing prior assumptions about what the trends of data on comprehension (e.g., [42]). However, the mistakes made by our participants seemed to be a different kind of mistake: they did not assume that they knew what trends the data would show — they assumed that the visualization was showing entirely different data. Drawing inspiration from studies on mind-wandering [8], it is possible that the kinds of mistakes we observed may have been circumstances where readers began looking at the chart, which then caused their mind to wander, and they mistook their wandering thought as what the visualization reflected. Understanding how these two situations relate to each other may be fruitful because they may better explain the landscape of assumptions visualization readers make and how those relate to readers' interpretations of those visualizations.

Despite the double-edged sword, integrating intratextual and extratextual information may be essential for obtaining "deeper" understanding or insights. We observed that one of the most common ways that participants combined intra- and extratextual information was to derive "deeper" messages, or "insights," from the visualization. While definitions of insights vary, insights can be thought of as "ah-ha" moments which reveal answers to "questions you didn't know you had" [22, 65] and have been a topic of high interest in the visualization community for some time [22, 60]. The pattern we observed regarding the integration of intra- and extratextual information to generate deep insights raises the question — do "deep" insights *require* integrating intra- and extratextual information? Or is it possible to generate insights with intratextual information alone? While some existing models of understanding define distinctive types with real-world connections(such as [20]), future work may further explore the relationship between extratextual information and insights more generally.

Further, for some kinds of data, the integration of intra- and extratextual information may be absolutely essential. For instance, Nowak and Bartram have written about the critical role that domain knowledge which is not present in the dataset plays in the daily analyses and decision making conducted by avalanche forecasters [61]. Nowak and Bartram's work is part of a group of literature related to the kinds of knowledge that domain experts or analysts have and use, but which is necessarily not present in the visualization such as their knowledge of implicit errors in the dataset (e.g., [57, 63]). Inspired by this line of work, another opportunity for future work exploring the interconnectedness of intra- and extratextual information may be to offer readers the opportunity to externalize the extratextual information they have onto the visualization itself (perhaps in the same spirit as Lin et al.'s work on visualizing "data hunches" [50]).

6.3 On the Impact of the Participant Pool

In this study, we interviewed a small sample of undergraduate students. Though it is true that small, student samples can be a problem because they are not representative of broader populations [18], we did not aim to produce generalizable results, but instead to describe existing behavior, rooted in its cultural context, as a means to begin to explore the unknown relationship between intra- and extratextual information use. Because examining the information used by participants through an intra- and extratextual lens is new, our work functions as an existence proof that, when we look for it, readers do use and integrate both types of information in interesting ways. However, we know that these results may neither be generalizable to all people or even reflective of all students. Nonetheless, we can still learn from the behavior of these student participants even if we do not know what may have motivated their approaches by considering the participants' cultural contexts which may have produced them.

One possibility is that the patterns we observed are reflective of our participants' ages and stage of life. All of our participants were between the ages of 19 and 21. This transition between adolescence and early adulthood is a key moment for identity formation in which people explore who they are and what their relationship is to others [25]. It is possible that our participants use of their own life experiences and personal identities as a lens through which to view the visualizations is a reflection of that identity formation stage. Future work which explores this possibility may contrast the approaches of individuals in this stage of life to those in different stages or examine the relationship between the identities which people are exploring and the ways they utilize those identities to understand visualizations.

Another possibility it that strategies reflect participants' role as STEM college students. Engaging in higher education is a social and cultural experience which impacts the personal and social identities of attendees [38]. Additionally, higher educational spaces heavily privilege dominant epistemologies [4]. Therefore, it is reasonable to believe that the strategies used by our participants may have been formed as a result of their experiences in higher education or the social roles that they have learned and inhabited as a result of that space. Further, STEM education scholars have written about the ways that engaging in the study of STEM disciplines requires agreeing to unspoken "agreements" which dictate how knowledge is created and valued [37]. Our participants' experiences in STEM spaces may have therefore helped to form their approaches and orientations to the visualizations, such as emphasizing or deemphasizing the relevance of statistics, statistical principles, and judgments of the commonnness of visualizations in their interpretations. Future work exploring this option further could consider the differences in perspectives of STEM and non-STEM students, individuals at different kinds of higher education institutions, and those with no experience with higher education, among others.

6.4 Limitations

There are several limitations to our work. First, the visualizations we selected may have some impact on the outcomes of our results. The four visualizations that we used were selected to try to vary in familiarity in terms of topic and encoding, but we did not control for many aspects of the visualizations' design and how their topics related to people. Future work may explore other topics, styles, and encodings. Finally, our analysis relied on what participants *said* – this makes it good because of its precision and reflects what they felt was most salient [22], but we might have missed things that the participants did not mention.

7 Conclusion

In this paper, we presented the results of a semi-structured interview study which investigated what kinds of intra- and extratextual information was used by participants and how they incorporated different types of information while interpreting four static, communicative visualizations. We analyzed our results with two qualitative methods: thematic analysis and diffractive reading. Our thematic analysis results indicated that our participants used a large assortment of different types of information from both dominant and non-dominant sources. We also described three ways that our participants integrated these disparate types of information throughout the process of reading the visualizations. Based in qualitative methodologies, our results expand on what we know about how visualization readers interpret visualizations and we hope that they offer opportunities for future research on the roles of intra- and extratextual information in the interpretation of visualizations.

Acknowledgments

The authors wish to thank our participant volunteers for sharing their time and experiences, Jordan Crouser for his revision feedback, and our anonymous reviewers for their thoughtful comments and perspectives on how to improve the work.

References

[1] Derya Akbaba, Devin Lange, Michael Correll, Alexander Lex, and Miriah Meyer. 2023. Troubling Collaboration: Matters of Care for Visualization Design Study. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* ACM, NY, USA, Article 812, 15 pages. https://doi.org/10.1145/3544548.3581168

[2] Akhil Reed Amar. 1998. Intratextualism. *Harvard Law Review* 112 (1998), 747.

[3] Elizabeth Anderson. 2024. Feminist Epistemology and Philosophy of Science. In *The Stanford Encyclopedia of Philosophy* (Fall 2024 ed.), Edward N. Zalta and Uri Nodelman (Eds.). Metaphysics Research Lab, Stanford University, CA, USA.

[4] Vanessa Andreotti, Cash Ahenakew, and Garrick Cooper. 2011. Epistemological Pluralism: Ethical and pedagogical challenges in higher education. *AlterNative: An International Journal of Indigenous Peoples* 7, 1 (2011), 40–50. https://doi.org/10.1177/117718011100700104

[5] RJ Andrews. 2014. Creative Routines. Retrieved 2024 from https://www.informationisbeautifulawards.com/showcase/483-creative-routines

[6] Natalia Andrienko and Gennady Andrienko. 2006. *Tasks.* Springer Berlin Heidelberg, Berlin, Heidelberg, 47–161. https://doi.org/10.1007/3-540-31190-4_3

[7] Adriana Arcia, Niurka Suero-Tejeda, Michael E Bales, Jacqueline A Merrill, Sunmoo Yoon, Janet Woollen, and Suzanne Bakken. 2016. Sometimes more is more: iterative participatory design of infographics for engagement of community members with varying levels of health literacy. *Journal of the American Medical Informatics Association* 23, 1 (2016), 174–183. https://doi.org/10.1093/jamia/ocv079

[8] Anjana Arunkumar, Lace Padilla, and Chris Bryan. 2025. Mind Drifts, Data Shifts: Utilizing Mind Wandering to Track the Evolution of User Experience with Data Visualizations. *IEEE Trans. Vis. Comput. Graph.* 31, 1 (2025), 1169–1179. https://doi.org/10.1109/TVCG.2024.3456344

[9] Paulo Bala, Pedro Sanches, Vanessa Cesário, Sarah Leão, Catarina Rodrigues, Nuno Jardim Nunes, and Valentina Nisi. 2023. Towards Critical Heritage in the wild: Analysing Discomfort through Collaborative Autoethnography. In *Proc. CHI Conf. Hum. Factors Comput. Sys. (CHI '23)*. ACM, NY, USA, Article 771, 19 pages. https://doi.org/10.1145/3544548.3581274

[10] Karen Barad. 2014. Diffracting diffraction: Cutting together-apart. *parallax* 20 (2014), 168–187. Issue 3. https://doi.org/10.1080/13534645.2014.927623

[11] Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. 2010. Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* ACM, NY, USA, 2573–2582. https://doi.org/10.1145/1753326.1753716

[12] Cindy Xiong Bearfield, Lisanne van Weelden, Adam Waytz, and Steven Franconeri. 2024. Same Data, Diverging Perspectives: The Power of Visualizations to Elicit Competing Interpretations. *IEEE Trans. Vis. Comput. Graph.* 30, 6 (2024), 2995–3007. https://doi.org/10.1109/TVCG.2024.3388515

[13] Eric Berlatsky. 2009. Lost in the Gutter: Within and Between Frames in Narrative and Narrative Theory. *Narrative* 17, 2 (2009), 162–187. http://www.jstor.org/stable/25609361

[14] Eduardo Bonilla-Silva and Crystal E. Peoples. 2022. Historically White Colleges and Universities: The Unbearable Whiteness of (Most) Colleges and Universities in America. *American Behavioral Scientist* 66, 11 (2022), 1490–1504. https://doi.org/10.1177/00027642211066047

[15] Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. 2016. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization* 15, 3 (2016), 198–213. https://doi.org/10.1177/1473871615594652

[16] Richard E Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development.* Sage Publications, CA, USA.

[17] Virginia Braun and Victoria Clarke. 2022. Conceptual and design thinking for thematic analysis. *Qualitative psychology* 9, 1 (2022), 3.

[18] Alyxander Burns, Christiana Lee, Ria Chawla, Evan Peck, and Narges Mahyar. 2023. Who Do We Mean When We Talk About Visualization Novices?. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* ACM, NY, USA, 16. https://doi.org/10.1145/3544548.3581524

[19] Alyxander Burns, Christiana Lee, Thai On, Cindy Xiong, Evan Peck, and Narges Mahyar. 2024. From Invisible to Visible: Impacts of Metadata in Communicative Data Visualization. *IEEE Trans. Vis. Comput. Graph.* 30, 7 (2024), 3427–3443. https://doi.org/10.1109/TVCG.2022.3231716

[20] Alyxander Burns, Cindy Xiong, Steven Franconeri, Alberto Cairo, and Narges Mahyar. 2020. How to evaluate data visualizations across different levels of understanding. In *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*. IEEE, CA, USA, 19–28. https://doi.org/10.1109/BELIV51497.2020.00010

[21] Alyxander Burns, Cindy Xiong, Steven Franconeri, Alberto Cairo, and Narges Mahyar. 2022. Designing with Pictographs: Envision Topics without Sacrificing Understanding. *IEEE Trans. Vis. Comput. Graph.* 28, 12 (2022), 4515–4530. https://doi.org/10.1109/TVCG.2021.3092680

[22] Sheelagh Carpendale. 2008. Evaluating information visualizations. In *Information visualization: Human-centered issues and perspectives.* Springer, USA, 19–45. https://doi.org/10.1007/978-3-540-70956-5_2

[23] Erin A. Cech, Anneke Metz, Jessi L. Smith, and Karen deVries. 2017. Epistemological Dominance and Social Inequality: Experiences of Native American Science, Engineering, and Health Students. *Science, Technology, & Human Values* 42, 5 (2017), 743–774. https://doi.org/10.1177/0162243916687037

[24] Russell S Chun. 2020. Giving guidance to graphs: Evaluating annotations of data visualizations for the news. *Visual Communication Quarterly* 27, 2 (2020), 84–97. https://doi.org/10.1080/15551393.2020.1749842

[25] Elisabetta Crocetti. 2017. Identity formation in adolescence: The dynamic of forming and consolidating identity commitments. *Child Development Perspectives* 11, 2 (2017), 145–150. https://doi.org/10.1111/cdep.12226

[26] Data Visualization Society. 2024. Information is Beautiful Awards. Retrieved 2024 from https://www.informationisbeautifulawards.com/

[27] Laura Devendorf, Kristina Andersen, and Aisling Kelliher. 2020. Making Design Memoirs: Understanding and Honoring Difficult Experiences. In *Proc. CHI Conf. Hum. Factors Comput. Sys. (CHI '20)*. ACM, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376345

[28] James N Druckman and Cindy D Kam. 2011. *Students as experimental participants.* Cambridge University Press, Cambridge, GB, 41–57. https://doi.org/10.1017/CBO9780511921452.004

[29] Mirta Galesic and Rocio Garcia-Retamero. 2011. Graph literacy: A cross-cultural comparison. *Medical decision making* 31, 3 (2011), 444–457. https://doi.org/10.1177/0272989X10373805

[30] Phillip Gough, Xavier Ho, Kate Dunn, and Tomasz Bednarz. 2014. Art and Chartjunk: a Guide for NEUVis. In *Proc. of the 7th Int. Sym. on Visual Info. Comm. and Interaction (VINCI '14)*. ACM, NY, USA, 171–177. https://doi.org/10.1145/2636240.2636852

[31] Stuart Hall. 2019. Encoding and Decoding in the Television Discourse [originally 1973; republished 2007]. In *Essential Essays, Volume 1: Foundations of Cultural Studies.* Duke University Press, NC, USA, Chapter 8, 257 – 276.

[32] Donna Haraway. 1988. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist studies* 14, 3 (1988), 575–599.

[33] Sandra Harding. 1998. *Is science multicultural?: Postcolonialisms, feminisms, and epistemologies.* Indiana University Press, IN, USA.

[34] Rosemary Lucy Hill, Helen Kennedy, and Ysabel Gerrard. 2016. Visualizing junk: Big data visualizations and the need for feminist data studies. *Journal of Communication Inquiry* 40, 4 (2016), 331–350. https://doi.org/10.1177/0196859916666041

[35] Jessica Hullman and Nick Diakopoulos. 2011. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Trans. Vis. Comput. Graph.* 17, 12 (2011), 2231–2240. https://doi.org/10.1109/TVCG.2011.255

[36] Jessica Hullman, Matthew Kay, Yea-Seul Kim, and Samana Shrestha. 2018. Imagining Replications: Graphical Prediction & Discrete Visualizations Improve Recall & Estimation of Effect Uncertainty. *IEEE Trans. Vis. Comput. Graph.* 24, 1 (2018), 446–456. https://doi.org/10.1109/TVCG.2017.2743898

[37] Mays Imad, Michael Reder, and Madelyn Rose. 2023. Recasting the agreements to re-humanize STEM education. In *Frontiers in Education*, Vol. 8. Frontiers, Switzerland, 1193477. https://doi.org/10.3389/feduc.2023.1193477

[38] Peter Kaufman. 2014. The sociology of college students' identity formation. *New Directions for Higher Education* 2014, 166 (2014), 35–42. https://doi.org/10.1002/he.20093

[39] Tanja Keller, Peter Gerjets, Katharina Scheiter, and Bärbel Garsoffky. 2006. Information visualizations for knowledge acquisition: The impact of dimensionality and color coding. *Computers in Human Behavior* 22, 1 (2006), 43–65. https://doi.org/10.1016/j.chb.2005.01.006 Instructional Design for Effective and Enjoyable Computer-Supported Learning.

[40] Helen Kennedy, Rosemary Lucy Hill, William Allen, and Andy Kirk. 2016. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday* 21, 11 (2016), 20 pages. https://doi.org/10.5210/fm.v21i11.6389

[41] Natchaya Kijmongkolchai, Alfie Abdul-Rahman, and Min Chen. 2017. Empirically Measuring Soft Knowledge in Visualization. *Computer Graphics Forum* 36, 3 (2017), 73–85. https://doi.org/10.1111/cgf.13169

[42] Yea-Seul Kim, Katharina Reinecke, and Jessica Hullman. 2017. Explaining the Gap: Visualizing One's Predictions Improves Recall and Comprehension of Data. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* ACM, NY, USA, 1375–1386. https://doi.org/10.1145/3025453.3025592

[43] Yea-Seul Kim, Katharina Reinecke, and Jessica Hullman. 2018. Data Through Others' Eyes: The Impact of Visualizing Others' Expectations on Visualization Interpretation. *IEEE Trans. Vis. Comput. Graph.* 24, 1 (2018), 760–769. https://doi.org/10.1109/TVCG.2017.2745240

[44] Yea-Seul Kim, Logan A. Walls, Peter Krafft, and Jessica Hullman. 2019. A Bayesian Cognition Approach to Improve Data Visualization. In *Proc. CHI Conf. Hum. Factors Comput. Sys. (CHI '19)*. ACM, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300912

[45] Karin Knorr Cetina. 1999. *Epistemic cultures: How the sciences make knowledge.* Harvard University Press, MA, USA.

[46] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and Slants in Titles of Visualizations on Controversial Topics. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* ACM, NY, USA, 12 pages. https://doi.org/10.1145/3173574.3174012

[47] Crystal Lee, Tanya Yang, Gabrielle D Inchoco, Graham M Jones, and Arvind Satyanarayan. 2021. Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* ACM, NY, USA, 1–18. https://doi.org/10.1145/3411764.3445211

[48] Sukwon Lee, Sung-Hee Kim, Ya-Hsin Hung, Heidi Lam, Youn-ah Kang, and Ji Soo Yi. 2015. How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE Trans. Vis. Comput. Graph.* 22, 1 (2015), 499–508. https://doi.org/10.1109/TVCG.2015.2467195

[49] Sukwon Lee, Bum Chul Kwon, Jiming Yang, Byung Cheol Lee, and Sung-Hee Kim. 2019. The Correlation between Users' Cognitive Characteristics and Visualization Literacy. *Applied Sciences* 9, 3 (2019), 20 pages. https://doi.org/10.3390/app9030488

[50] Haihan Lin, Derya Akbaba, Miriah Meyer, and Alexander Lex. 2023. Data Hunches: Incorporating Personal Knowledge into Visualizations. *IEEE Trans. Vis. Comput. Graph.* 29, 1 (2023), 504–514. https://doi.org/10.1109/TVCG.2022.3209451

[51] Zhengliang Liu, R Jordan Crouser, and Alvitta Ottley. 2020. Survey on individual differences in visualization. *Computer Graphics Forum* 39, 3 (2020), 693–712. https://doi.org/10.1111/cgf.14033

[52] Alberto Lucas López. 2015. A World of Languages. Retrieved 2024 from https://www.scmp.com/infographics/article/1810040/infographic-world-languages?page=all

[53] Yu Luo and Jiaying Zhao. 2019. Motivated attention in climate change perception and action. *Frontiers in Psychology* 10 (2019), 13 pages. https://doi.org/10.3389/fpsyg.2019.01541

[54] Giorgia Lupi, Simone Quadri, Gabriele Rossi, Davide Ciuffi, Federica Fragapane, and Francesco Majno. 2013. Nobels, No Degrees. Retrieved 2024 from https://www.informationisbeautifulawards.com/showcase/204-nobels-no-degrees

[55] Laura E. Matzen, Michael J. Haass, Kristin M. Divis, and Mallory C. Stites. 2017. Patterns of Attention: How Data Visualizations Are Read. In *Augmented Cognition: Neurocognition and Machine Learning*, Dylan D. Schmorrow and Cali M. Fidopiastis (Eds.). Springer International Publishing, USA, 176–191. https://doi.org/10.1007/978-3-319-58628-1_15

[56] Lisa A Mazzei. 2014. Beyond an Easy Sense: A Diffractive Analysis. *Qualitative Inquiry* 20, 6 (2014), 742–746. https://doi.org/10.1177/1077800414530257

[57] Nina Mccurdy, Julie Gerdes, and Miriah Meyer. 2019. A Framework for Externalizing Implicit Error Using Visualization. *IEEE Trans. Vis. Comput. Graph.* 25, 1 (2019), 925–935. https://doi.org/10.1109/TVCG.2018.2864913

[58] Andrew McNutt, Gordon Kindlmann, and Michael Correll. 2020. Surfacing visualization mirages. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* ACM, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376420

[59] Mount Holyoke College Career Development Center and Institutional Research and Effectiveness. 2023. Mount Holyoke Employment Outcomes 6-Months Post Graduation. Retrieved 2023 from https://www.mtholyoke.edu/academics/after-mount-holyoke/student-success-and-career-outcomes

[60] C. North. 2006. Toward measuring visualization insight. *IEEE Computer Graphics and Applications* 26, 3 (2006), 6–9. https://doi.org/10.1109/MCG.2006.70

[61] Stan Nowak and Lyn Bartram. 2022. I'm Not Sure: Designing for Ambiguity in Visual Analytics. In *Graphics Interface 2022*. Canadian Information Processing Society, Canada, 12 pages. https://openreview.net/forum?id=SHQU_yejZFv

[62] Alvitta Ottley, Aleksandra Kaszowska, R Jordan Crouser, and Evan M Peck. 2019. The Curious Case of Combining Text and Visualization. In *EuroVis (Short Papers)*. The Eurographics Association, Switzerland, 121–125. https://doi.org/10.2312/evs.20191181

[63] Georgia Panagiotidou, Ralf Vandam, Jeroen Poblome, and Andrew Vande Moere. 2022. Implicit Error, Uncertainty and Confidence in Visualization: An Archaeological Case Study. *IEEE Trans. Vis. Comput. Graph.* 28, 12 (2022), 4389–4402. https://doi.org/10.1109/TVCG.2021.3088339

[64] Evan M Peck, Sofia E Ayuso, and Omar El-Etr. 2019. Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* ACM, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300474

[65] Catherine Plaisant. 2004. The challenge of information visualization evaluation. In *Proc. of the Working Conf. on Adv. Visual Interfaces.* ACM, NY, USA, 109–116. https://doi.org/10.1145/989863.989880

[66] Annemarie Quispel, Alfons Maes, and Joost Schilperoord. 2016. Graph and chart aesthetics for experts and laymen in design: The role of familiarity and perceived ease of use. *Information Vis.* 15, 3 (2016), 238–252. https://doi.org/10.1177/1473871615606478

[67] Nina Rajcic, Maria Teresa Llano Rodriguez, and Jon McCormack. 2024. Towards a Diffractive Analysis of Prompt-Based Generative AI. In *Proc. CHI Conf. Hum. Factors Comput. Sys. (CHI '24)*. ACM, NY, USA, Article 844, 15 pages. https://doi.org/10.1145/3613904.3641971

[68] Maryam Rezaie, Melanie Tory, and Sheelagh Carpendale. 2024. Struggles and Strategies in Understanding Information Visualizations. *IEEE Trans. Vis. Comput. Graph.* 30, 6 (2024), 3035–3048. https://doi.org/10.1109/TVCG.2024.3388560

[69] Nicole Robson, Andrew McPherson, and Nick Bryan-Kinns. 2024. Thinking with Sound: Exploring the Experience of Listening to an Ultrasonic Art Installation. In *Proc. CHI Conf. Hum. Factors Comput. Sys. (CHI '24)*. ACM, NY, USA, Article 496, 14 pages. https://doi.org/10.1145/3613904.3642616

[70] Purvi Saraiya, Chris North, Vy Lam, and Karen A. Duca. 2006. An Insight-Based Longitudinal Study of Visual Analytics. *IEEE Trans. Vis. Comput. Graph.* 12, 6 (2006), 1511–1522. https://doi.org/10.1109/TVCG.2006.85

[71] Regina Schuster, Laura Koesten, Kathleen Gregory, and Torsten Möller. 2023. "The main message is that sustainability would help"' – Reflections on takeaway messages of climate change data visualizations. (2023), 5 pages. https://arxiv.org/pdf/2305.04030 Presented at the CHI '23 Workshop - HCI for Climate Change: Imagining Sustainable Futures.

[72] Samuel Silva, Beatriz Sousa Santos, and Joaquim Madeira. 2011. Using color in visualization: A survey. *Computers & Graphics* 35, 2 (2011), 320–333. https://doi.org/10.1016/j.cag.2010.11.015

[73] Adrian Vermeule and Ernest A Young. 1999. Hercules, Herbert, and Amar: The Trouble with Intratextualism. *Harv. L. Rev.* 113 (1999), 730.

[74] Jagoda Walny, Samuel Huron, and Sheelagh Carpendale. 2015. An exploratory study of data sketching for visual representation. *Computer Graphics Forum* 34, 3 (2015), 231–240. https://doi.org/10.1111/cgf.12635

[75] Cindy Xiong, Lisanne Van Weelden, and Steven Franconeri. 2019. The curse of knowledge in visual data communication. *IEEE Trans. Vis. Comput. Graph.* 26, 10 (2019), 3051–3062. https://doi.org/10.1109/TVCG.2019.2917689

[76] Caroline Ziemkiewicz, Alvitta Ottley, R Jordan Crouser, Krysta Chauncey, Sara L Su, and Remco Chang. 2012. Understanding visualization by understanding individual users. *IEEE CG&A* 32, 6 (2012), 88–94. https://doi.org/10.1109/MCG.2012.120