From Invisible to Visible: Impacts of Metadata in Communicative Data Visualization

Alyxander Burns, Christiana Lee, Thai On, Cindy Xiong, Evan Peck, and Narges Mahyar

Abstract—Leaving the context of visualizations invisible can have negative impacts on understanding and transparency. While common wisdom suggests that recontextualizing visualizations with metadata (e.g., disclosing the data source or instructions for decoding the visualizations' encoding) may counter these effects, the impact remains largely unknown. To fill this gap, we conducted two experiments. In Experiment 1, we explored how chart type, topic, and user goal impacted which categories of metadata participants deemed most relevant. We presented 64 participants with four real-world visualizations. For each visualization, participants were given four goals and selected the type of metadata they most wanted from a set of 18 types. Our results indicated that participants were most interested in metadata which explained the visualization's encoding for goals related to understanding and metadata about the source of the data for assessing trustworthiness. In Experiment 2, we explored how these two types of metadata impact transparency, trustworthiness and persuasiveness, information relevance, and understanding. We asked 144 participants to explain the main message of two pairs of visualizations (one with metadata and one without); rate them on scales of transparency and relevance; and then predict the likelihood that they were selected for a presentation to policymakers. Our results suggested that visualizations with metadata were perceived as more thorough than those without metadata, but similarly relevant, accurate, clear, and complete. Additionally, we found that metadata did not impact the accuracy of the information extracted from visualizations, but may have influenced which information participants remembered as important or interesting.

Index Terms—Visualization, metadata, understanding, transparency, trust.

1 Introduction

Although designers and practitioners aim to create communicative visualizations which can be easily used by readers, visualizations do not always speak for themselves. For instance, a lack of transparency surrounding how visualizations with unusual or unfamiliar elements should be decoded can make successfully extracting information difficult for readers who have not encountered them before or do not use them regularly [1] (e.g., logarithmic scales [2], truncated y-axes [3]). The same lack of transparency may also impact how trustworthy or persuasive readers perceive visualizations to be (i.e., how much they believe the information communicated by visualizations is accurate and able to convince other readers of accuracy). For example, a lack of transparency surrounding where noise and uncertainty may be present in the data could sew doubt in the accuracy of the information and cause a reader to ignore the visualized information altogether [4].

Prior work has theorized that contextualizing visualizations with **metadata** (e.g., disclosing the creator, information about the encoding used, or the data source) may counter these effects and potentially increase understanding and perceived transparency (e.g., [5], [6]). For instance, providing metadata in the form of instructions for how to make sense of a visualization's encodings might enable readers to understand representations they otherwise would have difficulty with (e.g., as with visualizations viewed for the first time [7]). The practice of disclosing metadata may also increase the transparency of the processes underlying a visualization, helping the public build trust [8], [9], and encouraging people to use visualizations [10]. However, there is little empirical evidence of the impacts of providing metadata on visualization readers.

To address this gap, we conducted a pair of empirical experiments to investigate the impacts that metadata have on readers of static communicative visualizations. We build upon previous work by the authors that defines a taxonomy of **metadata**: *information that is not directly represented in a visualization which provides contextual information on the source of the data, the transformations applied to the data, the visualization elements, its purpose, the people involved in its creation, and its intended audience* [11]. This definition of metadata is broad and intentionally encompasses types of data that are typically considered metadata (e.g., about data collection or methods [12], [13]) as well as information that may not traditionally be thought of as "data" but are nonetheless important for establishing the social, cultural, or historical circumstances in which a visualization was made (e.g., authorial statements of positionality which elaborate on identities and experiences that inform how the authors relate to knowledge [14]) [11].

In our first study, we utilized an exploratory design to identify which of six categories of metadata readers wanted to have to accomplish user goals (e.g., better understand the topic, assess trustworthiness). We presented 64 participants from Prolific [15] with a set of four random visualizations from five award-winning data journalism projects. For each visualization, participants were given four goals and se-

- *Alyxander Burns is with Mount Holyoke College. Email: ajburns@mtholyoke.edu.*
- *Christiana Lee, Thai On, Cindy Xiong, and Narges Mahyar are with the University of Massachusetts Amherst. Emails: [christianale, ton]@umass.edu, [cindy.xiong, nmahyar] @cs.umass.edu.*
- *Evan Peck is with Bucknell University. Email: evan.peck@bucknell.edu.*

Manuscript received April 19, 2005; revised August 26, 2015.

lected the type of metadata they most wanted to have to accomplish each of them from a set of 18 types of metadata. Our results indicated that participants were most interested in metadata that explained the visualization's encoding (Encoding Explanation) for goals related to understanding (e.g., "to make you feel more confident that you understand the chart well) and metadata about the source of the data (Data Source) for assessing trustworthiness (i.e., "to determine if the chart is trustworthy").

Based on the results of Experiment 1, in the second experiment, we explored how these metadata (i.e., Encoding Explanation, Data Source) impact how participants perceive and understand visualizations. 144 participants from Prolific were shown pairs of visualizations, both with and without metadata. We collected open-response text related to the visualization's message and Likert-scales corresponding to accuracy, clarity, completeness, and thoroughness (operationalized to assess perceived transparency) and three variations of relevance (meaningfulness, relevance to self, relevance to others). Finally, participants compared two visualizations by predicting the likelihood that they would be selected by an organization for a presentation on the bases of trustworthiness and persuasiveness. Our results suggest that people view visualizations with metadata as more thorough than those without, but found no evidence that it contributes to perceived accuracy, clarity, completeness, or relevance. Additionally, our participants thought that visualizations with metadata were significantly more likely to be shown to policymakers than visualizations without metadata when told that the visualizations had been selected on the basis of trustworthiness and persuasiveness. Finally, we found that the presence of metadata did not impact how correctly participants described the main message and what they learned from a visualization, but may have influenced *which* information readers interpreted as important.

The contributions of this work are: 1) an exploratory study of what kinds of metadata people may perceive as most important to accomplish different goals (e.g., assess trustworthiness, understand perspectives); 2) quantitative results suggesting that metadata can increase perceptions of thoroughness, which has been identified by prior work as positively associated with perceived transparency; 3) quantitative results suggesting that the information in visualizations with metadata are not considered more relevant than in visualizations without; 4) quantitative and qualitative results implying that metadata did not impact the correctness of participant responses, but may have influenced *which* information readers paid attention to; and 5) a discussion of the possible implications of our results on trust, understanding, transparency, relevance, and the tension between disclosing information and providing more textual information.

2 Background

2.1 Potential Benefits of Metadata

Existing literature has posed that the practice of disclosing metadata may provide benefits for researchers, data users, and society as a whole. For example, past work has posed that disclosing metadata in research can help establish the accuracy of claims, prove authenticity, replicate work, and conduct meta-analyses [16], [17], [18], [19], [20], [21], [22], which can help establish the objectivity and context of research claims [23], [24]. This practice has also been argued to be beneficial for prospective users by helping them interpret data [18], judge whether it is appropriate to apply a dataset to a problem [25], and determine what kinds of conclusions can be responsibly inferred from data [26].

In addition, there may be other aspects of a visualization's history and context which may be important to divulge, but not typically considered data. For example, positionality statements which describe an author's worldview and the ways that they think about the social and political context of their research are commonplace made in social science [14], [27], [28]. Understanding the positionality of people is an integral component of understanding the visualizations they make because one's identities and experiences both frame and limit which knowledge they create [29], [30], [31], [32]. This kind of information is considered metadata by our definition but may have been left out by past work because it is not readily quantifiable and therefore not typically considered data. Past work in the visualization community has also demonstrated that a person's personal knowledge and experience are highly related to what they will perceive as useful [33], believe to be too difficult to understand [34], and think others will see in the visualization [35]. Finally, feminist scholars have emphasized the importance of understanding the social and historical contexts of data to challenge hegemonic power structures (e.g., [25], [29], [30]).

2.2 Increasing Perceived Transparency & Trustworthiness

One of our primary motivations for this work was to examine the potential for metadata to influence and increase perceived transparency and assessments of trustworthiness. Past research in visualization concluded that providing metadata like data provenance can be a tool for signaling transparency and trustworthiness to end-users [5], [6]. While increased transparency is often associated with increased trust (e.g., as in [36]), this relationship does not always hold. Instead, transparency increases trust to a critical point, at which increasing transparency further can counterintuitively decrease trust — especially when expectations are broken [9], [37]. This erosion of trust is theorized to occur because too much explanation confuses readers and directs their attention toward unexpected outcomes [37].

Although the relationship between transparency and trust is not often discussed in Computer Science-related fields, transparency is understood to influence trust (though is not necessarily a required antecedent of it) in other fields such as Economics and Management (e.g., as in [38]). In these contexts, transparency is said to be a combination of accuracy, clarity, and disclosure [38]. However, past work on transparency and trust in map-based visualizations found that elements of transparency from these areas (accuracy, clarity, disclosure, and thoroughness) did not predict participants' perceived level of trust [38], [39]. Instead, they found that only accuracy and disclosure significantly predicted participants' perceived levels of trust in visualizations, while clarity, disclosure, and thoroughness predicted which visualization participants selected for the experimental task [39]. Further, existing work suggests that the perceived value of transparency (as a component of trust in a resource) is higher for people with deeper relationships with an organization or resource [40]. Within the context of visualization, this result could indicate that individuals who engage with a visualization repeatedly or have more at stake if the visualization is incorrect may value transparency more than those who interact only once.

Trust, in general, is a critical aspect of successful visualization and has been identified as one of the field's most pressing challenges [10], [41]. Unfortunately, the study of trust is further complicated because there is no consensus on a definition for trust or how to measure it [42]. However, past work has established that evaluations of trust are based (in part) on stakeholders' expectations, which may make building trustworthy visualizations particularly challenging because they often have multiple stakeholders [43], [44]. Readers of a visualization can both assess trust through comprehensive processes (when readers have time or sufficient motivation) and through the use of mental shortcuts [10], [45]. For example, past work has shown that people use the trustworthiness of a source and how much the "raw" data had been processed as proxies for trustworthiness [45], [46], [47], [48]. It is conceivable that the presence of metadata might influence perceived trustworthiness for readers engaging in comprehensive processes (who want or need to look through all of the related information) and readers making quicker judgments using proxies (who could use metadata like the source of the data as a shortcut). However, little work has empirically investigated this relationship.

While studies of metadata within visualization are limited, other related fields have investigated how metadata influences perceptions of transparency, credibility, and trust. For instance, past work in the Digital Humanities found that transparency was an important factor in social assessments of the trust for digital repositories [49] and educational resources [50]. Notably, the disclosure of metadata was specifically named as a sign of trustworthiness for many participants [49]. Recent work also found that metadata that indicated a software was reputable and would perform well increased perceived trustworthiness in the software [51]. Past work on credibility (the extent to which something is thought to be believable, trustworthy, accurate, and valid [52]) in this space has observed, for example, that the presence of sources believed to be experts increased the perceived credibility of memes [52], tweets [53], and social media posts [54]. Existing results on metadata, credibility, trust, and transparency may indicate that the disclosure of metadata alongside a visualization could increase perceptions of credibility, transparency, and trust among visualization readers. However, it is not yet clear whether these results translate to visualizations. For example, everyday readers of communicative visualizations may not expect that metadata is available or find it as important as researchers have in past work. Further, while metadata is mentioned frequently in the context of transparency, it is also unclear whether readers of visualizations *perceive* the disclosure of metadata as contributing additional information (that is, makes them seem more transparent).

2.3 Improving Understanding with Context

Finally, metadata may influence how accurately visualization readers are able to understand a visualization. For example, rhetorical choices made by visualization designers can make some interpretations more or less salient [6]. Therefore, metadata that include "instructions" for decoding a visualization could minimize confusion and misunderstanding, especially when a visualization is complex or when the rhetorical choices are novel. For instance, past work on providing instructions and tips in the form of cheat sheets [55] or slide shows [7] have both shown to help people who are new to visualization draw conclusions from unfamiliar data visualizations. However, providing additional information might not always be beneficial – an experiment in psychology found that while access to additional information made participants more confident in their responses, it also *decreased* their accuracy because they were biased by preconceived notions about the information they received [56].

Beyond influencing accuracy, it is also possible that metadata could influence *how* readers understand the visualization. The encoding/decoding theory model of communication suggests that when consuming new information, people use both their own knowledge and the new information to come to conclusions [57]. This theory is consistent with existing results in visualization which have shown that a reader's interpretation of a visualization may be affected by their own experience, knowledge, and perspectives (e.g., [6], [33], [34], [58], [59]). Consequently, providing more information about a visualization in the form of metadata could influence which conclusions are drawn by readers.

In summary, although metadata has been proposed to have positive impacts on people in a variety of contexts, there is little empirical evidence of these impacts. In this work, we explored the impacts of metadata on trust and understanding because they both can impact how much a visualization is used and have been theorized to be improved by the inclusion of metadata. Because the space of possible kinds of metadata to disclose is broad, we first needed to know what kinds of metadata visualization readers wanted to have in order to eventually study their impacts.

3 Experiment 1: Desirable Metadata

3.1 Methodology

In order to investigate the effects of metadata on visualization readers, we began with the high-level question *Which metadata do people **prefer** to see?*, by asking the following questions:

- **Q1:** Does the goal of a reader impact the category of metadata (*e.g.* data source, author, encoding explanation, etc.) that they prefer?
- **Q2:** Does the type of visualization and presented topic impact a person's metadata preferences?

Stimuli: Choosing Visualizations

Source of visualizations: We used a set of 32 visualizations collected from projects that won the 2022 Sigma Awards — an annual data journalism competition evaluated by an international panel of judges [60]. Sigma Awards are

Category of Metadata	Description of Metadata Category	Surveyed Examples of Metadata (Provided examples are highlighted)
(1) Data Source	Information about which dataset was used and details of its collection including people involved and methods used	**Link to the dataset, Name of the people or organization(s) that collected the data, Description of the data collection method**, Name of the data source (unlinked), Note about when the data were collected
(2) Cleaning & Processing	Methodological information about how data were cleaned and processed in order to produce the final visualization	**Description of how the data were processed, Description of a few impactful processing steps, Date when the data were last processed**, Link to external description of how the data were processed, Link to the tool used to create the chart
(3) Perceptual Challenges	Problems the reader might face while decoding the marks, channels, or other design elements	**Description of a potential misunderstanding, A warning against using the chart in some way, Description of the limits of the data**
(4) Encoding Explanation	Information that helps the reader make sense of the visualization through explaining key take-aways or how to read the visualization	**Explanation of how to read the chart, Description of the main message of the chart, Description of a few key insights**
(5) Creators	Information about creators of the visualization either directly (e.g., designers) or indirectly (e.g., donors)	**Names of the people who created the chart, Name of the funding organization(s), Name of the organization that made the chart**, Roles of the people who created the chart, Biographies of the people who created the chart, Photographs of the people who created the chart, Link to social media handles, Links to email addresses, Link to an external page with biographies and other work by the same author, Name of the creation team, Link to the funding organization(s) website(s)
(6) Intended Audience	Information about for whom the visualization was originally designed	**Description of the intended audience's demographics, Design choices made to cater to them, Location where it was displayed**

TABLE 1
We used six categories of metadata defined by the authors in prior work [11] (columns 1 & 2). We collected 28 examples of metadata in our survey of existing practices. Because there were too many examples to show participants, we selected a subset to show participants (3 per metadata category, bolded). Note: no examples for "Intended Audience" were collected during the survey and were instead generated by the researchers.

given to "projects" – bodies of work on one topic by the same organization or authors. The website is organized such that each project has a single page containing all of the resources submitted by the authors for review. We selected visualizations from the Sigma Awards because they were high-quality visualizations and were deployed in the wild.

Selection criteria: The 32 visualizations that we used were *every* English-language visualization linked from the winners' project pages.For the purposes of this search, photographs, illustrations, and street maps intended for navigation (e.g., Google Maps) were not considered data visualizations. There were five projects which were written in English and contained at least one visualization that met our selection criteria: *The COVID Tracking Project* at The Atlantic [61], *Mapping Makoko* [62], *Land-Grab Universities* [63], *Rough Justice* [64], and *Who Gets to Breathe Clean Air in New Delhi?* [65]. Within the set, there were a total of 12 maps, 10 bar charts, 8 line charts, 2 pie or donut charts, 2 area charts, and 1 infographic[1]. All 32 visualizations are provided in the Supplemental materials (available online: https://osf.io/mzgrp/?view_only=47f1e00053a14501babe93e45129e094).

Metadata

Past work defined six categories of metadata that could be provided to visualization readers (see Column 1 of Table 1 for the names and Column 2 for the definitions of all six categories) [11]. However, each of the six categories of metadata are broad. For example, there are many different pieces of information that could be considered metadata about a creator (e.g., their name, job title, photograph, the company they work for, etc). Even though all of these data points might be considered the same category of metadata, they may be useful in different scenarios and have different impacts on readers.

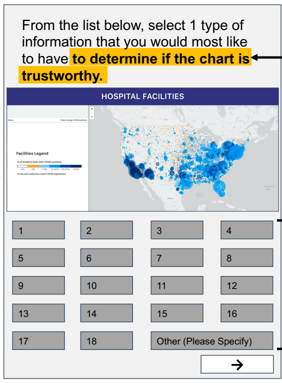
Since it may be difficult to select between broad metadata categories, we began by collecting concrete examples of disclosed metadata by news and journalism websites. Three of the authors compiled a set of examples disclosed in five recent articles published by the top five English-language news organizations, measured by the total number of website visits (New York Times, BBC, CNN, Daily Mail, and Fox News) and articles linked from the 2022 Sigma Awards winners' project pages. This inquiry resulted in a set of 28 discrete examples, listed in Column 3 of Table 1.

Selecting examples: From this set, we selected three examples for each of the six metadata categories for a total of $6 \times 3 = 18$ examples (highlighted in Table 1). The research team selected examples that were distinct from each other and used by multiple organizations (where possible). For example, two of the five examples of Data Source metadata were "Link to the dataset" and "Name of the data source (unlinked)." Both pieces of metadata disclose the exact source of the data, so we included the "Link to the dataset" because it appeared more frequently in our survey. At least three examples of metadata were collected for all of the metadata categories except for "Intended Audience," of which we found no explicit examples. In lieu of "Intended Audience" metadata, the research team generated three examples that represent ways in which intended audiences are described in fields like communication studies: by their demographics, as belonging to a place and time, and by the choices made to suit them (as discussed in [66]).

Goals

There were a total of eight goals that participants could encounter during the study (listed in Figure 1). The set of eight goals was generated to represent different types of outcomes that have been theorized to result from metadata access (e.g., in [11]) or were shown in existing literature to be a reason that people sought additional information (e.g., as in [67]). Each goal completed the phrase: "From the list below, please select 1 piece of information you most want to have..." For example, the prompt text for the goal "Increase Confidence with Chart" was "From the list below, please select 1 piece of information you most want to have to make you feel more confident that you understand the chart well."

1. Three visualizations contained two types of graphs in one image.



Participants were randomly assigned 4 goals from the table below for each of the 4 visualizations they viewed.

Goal:	Prompt Text:
1. Confidence in Chart	to make you feel more confident that you understand the chart well
2. Confidence in Topic	to make you feel more confident that you understand the topic of the chart well
3. Key Takeaways	to better understand the key takeaways of the chart
4. Assess Trustworthiness	to determine if the chart is trustworthy
5. Understand Method	to better understand how the chart was made
6. Understand Design	to better understand why the chart was designed this way
7. Understand Perspective	to better understand whose perspective the chart represents
8. Satisfy Interest	because the information seems interesting to you

These boxes represent the 19 options presented to participants (18 types of metadata + 1 option to self-specify)

Fig. 1. In Experiment 1, participants indicated which type of metadata they most wanted to accomplish a specified goal. Participants were randomly assigned 4 goals for each of the 4 visualizations they saw.

Participants

We recruited 64 participants who self-identified as fluent in English from Prolific [15]. Prolific is a crowdsourcing data collection platform that provides access to a diverse pool of vetted participants. In comparison to alternative platforms such as Amazon MTurk, Prolific provides more comprehensive and granular control over the selection of participants. We restricted our participant pool to only include individuals who identified as fluent in English to ensure that the instructions, visualization text, and metadata would be well understood. We decided to recruit 64 participants based on a small pilot study.

We collected demographic information about education and use of visualizations because they may be indicative of how fluently participants are able to use visualizations. A majority of our participants had completed some education beyond high school (71.19%), split fairly evenly among participants with some college education and completed a four-year degree. In addition, a majority of participants reported encountering or using visualizations at least once a week (53.13%) or once a month (26.56%). The entire experiment was implemented in Qualtrics [68] and participants were redirected back to Prolific after successful completion of the experiment. The study took about 10 minutes to complete and participants were paid $3.00.

Procedure

In each trial, participants were presented with a visualization and asked to indicate which piece of metadata they would most want to have to reach a specified goal. The question was posed as a multiple-choice question with 19 options: 18 examples of metadata (6 categories $\times$ 3 examples), plus one option to select "Other" and type in a type of metadata not listed (see Figure 1 for a simplified view of the study layout). Each participant completed a total of 16 trials (4 goals $\times$ 4 visualizations), where the order of the goals and visualizations were randomized. We used a 4-by-4 design for this experiment so that each participant would see a diverse set of visualizations and goals without getting fatigued from the repetition. Although this design means that not every participant saw visualizations from each project, the purpose of this study was to get a snapshot of which metadata categories participants were interested in. The number of times each goal was encountered varied between 115 and 144 due to random selection.

Analysis

To analyze our data, we grouped the 18 responses (all except "Other") back into the category of metadata to which they belonged. We counted the number of times that examples from each category were selected and compared them. "Other" responses were only given 7 times across the 1024 trials completed by participants and so are excluded from our statistical analysis.

3.2 Results

Our results suggested that among our participants, the goal given in the prompt impacted the category of metadata selected, but not the chart type or topic. We used a linear mixed-effects model to investigate if the category of metadata selected by participants was impacted by the chart type, chart topic, and goal. The fixed effects of our model were the chart type, the chart topic, and the goal. We also used a random intercept term to account for individual differences between participants. We observed a significant effect of goal ($\chi^2 = 17.5289$, $p < 0.05$), but not chart type or topic. Table 2 contains an overview of the distribution of participant responses for each goal. In the following subsections, we explore which kinds of metadata were selected when participants were presented with each goal.

For each goal, we utilized a Chi-squared test of independence to determine if there was a significant difference overall between the categories of metadata selected. If an overall difference was detected, we used a post-hoc analysis with Bonferroni adjustments [69] to explore pairwise differences.

Understanding-based Goals

Our results suggest that participants were most interested in accessing Encoding Explanation metadata (e.g., explanation of how to read the chart, the intended main message) for all three of the goals relating to understanding: **Increase Confidence with Chart**, **Increase Confidence with Topic**, and **Understanding Key Takeaways**. For all three goals, our Chi-squared test indicated an overall difference among the categories of metadata ($\chi^2 = [101.45, 237.11, 240.48]$, $p < 0.001$). Post hoc analysis with Bonferroni adjustments suggested that the number of requests for Encoding Explanation metadata was significantly higher than all other categories of metadata. Details of pairwise post-hoc comparisons can be found in the Supplemental Materials.

Assess Trustworthiness

Our results suggested that metadata about the Data Source (e.g., name of the data collector, link to the dataset) was selected most frequently by participants for the goal of **Assessing Trustworthiness**. Our Chi-squared test indicated that there was an overall difference among the categories of metadata requested by participants ($\chi^2 = 146.33$, $p < 0.001$). Post-hoc analysis suggested that the number of times metadata about the Data Source was selected was significantly higher than any other category of metadata.

Design-based Goals

The categories of metadata that were selected by participants for goals related to design (**Understand Method, Understand Design, Understand Perspectives**) did not follow

Goals	Data Sources	Cleaning & Processing	Perceptual Challenges	Encoding Explanation	Creators	Intended Audience	Other
Confidence in Chart	11% (13)	10% (12)	11% (13)	51% (59)	4% (4)	13% (15)	0% (0)
Confidence in Topic	14% (21)	6% (9)	7% (10)	58% (83)	1% (2)	13% (18)	1% (1)
Key Takeaways	11% (15)	8% (12)	4% (5)	65% (92)	3% (5)	9% (13)	0% (0)
Assess Trustworthiness	51% (62)	15% (18)	5% (6)	5% (6)	13% (16)	9% (11)	2% (2)
Understand Method	32% (40)	21% (27)	4% (6)	20% (25)	6% (7)	17% (21)	0% (0)
Understand Design	10% (12)	24% (27)	3% (4)	17% (19)	6% (7)	39% (45)	1% (1)
Understand Perspective	21% (25)	8% (9)	8% (9)	22% (27)	20% (24)	21% (26)	1% (1)
Satisfy Interest	26% (36)	11% (15)	7% (9)	24% (33)	7% (10)	25% (34)	1% (2)

Categories of Metadata

TABLE 2
In Experiment 1, participants selected which kind of metadata they would most want to see when provided one of the eight goals. This table provides the number and percentages of times each metadata category (columns) was selected per goal (rows). Participants most frequently requested in Encoding Explanation metadata, followed by Data Sources. Percentages are rounded up to the next integer; rows may not sum to 100.

a consistent pattern, unlike the group of goals related to understanding. Our Chi-squared tests for all three goals related to understanding indicated that there was an overall difference among the categories of metadata requested by participants ($\chi^2 = [39.714, 87.39, 39.19]$, $p < 0.001$). Post-hoc analysis suggested that for the goal of **Understand Method**, metadata about the Data Source, Cleaning & Processing, Encoding Explanation, and Intended Audience were the most requested, with no significant difference between the categories. For **Understand Design**, metadata about the Intended Audience (e.g., description of design choices) was selected significantly more than any other category of metadata except metadata about Cleaning & Processing (e.g., description of how data were processed). Finally, post-hoc analysis for **Understand Perspective** revealed that no category of metadata was selected significantly more than any other category.

Satisfy Interest

Our results suggested that when participants were given the goal of **Satisfying Interest**, there was an overall difference among the categories of metadata requested by participants ($\chi^2 = 59.971$, $p < 0.001$) and they selected metadata about the Encoding Explanation, Intended Audience, and Data Source significantly more often than the other categories. Among these three categories of metadata, none was selected more frequently.

3.3 Experiment 1: Summary of Results

In summary, we observed that our participants selected different categories of metadata across the goals they were presented with. Encoding Explanation metadata was selected most frequently for the three goals related to understanding. Additionally, participants most frequently selected metadata about the Data Source for the goal of Assessing Trustworthiness. Finally, the categories of metadata requested for Design-related goals and the goal related to personal interest did not indicate a strong preference for one category of metadata over others.

4 Experiment 2: Impacts of Metadata

The previous experiment identified metadata categories that participants were interested in accessing. This experiment investigates the impacts of those metadata on perceptions and understanding.

4.1 Methodology

Stimuli

We showed participants four visualizations used in Experiment 1. In Experiment 1, we found that the type and topic of the visualization did not have a strong effect on how participants responded to metadata. Therefore, for Experiment 2, we chose to sample all of the visualizations from a single source on a single topic: the *Land-Grab Universities* [63] project. This topic serves as a good testing bed because the project contained information about locations and universities that American participants might recognize, even if they were unfamiliar with the idea of land-grant universities prior to participating in the study. We instead varied the type of visualization shown and used a map, a bar chart, an infographic, and a series of pie charts (see Figure 2).

When participants were provided metadata with a visualization, they were told that there was additional information available and were directed to a screen containing only the textual metadata about the visualization and then a screen containing both the metadata and visualization (see the top of Figure 2 for screenshots of all three pages). We chose to direct the participants to designated screens to view the metadata to ensure that they were aware of its presence. The metadata presented to participants was always information about the Data Source and Encoding Explanation drawn from public sources related to the project. We always presented participants with both of these kinds of metadata because we wanted to identify the effects of metadata in general rather than a specific type of metadata. We selected these two categories of metadata based on our results from Experiment 1 because they were the most highly requested categories of metadata for goals related to understanding and trustworthiness. There were no goals from Experiment 1 which were explicitly related to relevance.

Perceived Transparency

The disclosure of metadata is often cited as a means to communicate transparency (e.g., in [5]), but do visualization readers see it that way? To date, there is no consensus as to what transparency is, how to measure transparency, or which factors contribute to it. However, previous work concluded that accuracy, clarity, and completeness[2] were essential components of transparency [38] and that, in a

2. In previous work, completeness is referred to as "disclosure."

visualization context, thoroughness captures the extent to which the data visualized represents the possible design space [39]. Therefore, instead of directly measuring transparency, we operationalized and measured these four more concrete dimensions. We asked participants to rate their agreement with four statements about the visualizations: accuracy, clarity, completeness, and thoroughness (see Table 3 for the statements provided to participants and our definitions for each of the dimensions). We then compared responses to these questions across participants to compare whether the presence of metadata had an effect on their perceptions of the dimensions.

Trustworthiness & Persuasiveness

Based on our goal to explore how disclosing metadata might influence readers' perceptions of visualizations, we asked participants to make a **prediction** about the probability that an organization would choose a visualization. After having seen two visualizations, participants were given the following scenario: *An organization selected one of the two charts you just saw to use in a presentation for local policymakers. They selected the chart that they believed was most persuasive and trustworthy.* and were prompted to "Use the sliders to indicate how probable you think it is that the organization selected each chart (out of a total of 100%)."

We asked participants to indicate a probability out of 100% to capture the magnitude of the difference between the two choices that participants felt (e.g., whether they felt strongly that one had been selected over the other). The prompt mentioned that the hypothetical organization had "used" trustworthiness and persuasiveness to select the visualization to ensure that participants used similar metrics to make judgments, rendering the predictions comparable between participants. Further, it mentioned both trustworthiness and persuasiveness because the two have been shown in past work to be closely linked (e.g., in [70]) and a visualization which is selected because it is trustworthy *and* persuasive may be believed to contain true information and possess some quality that convinces other readers of that truth. This may be desirable for communicative visualizations like the one in the scenario. We asked participants about how someone else would respond because prior work suggests that asking participants about how others will respond can result in responses that are more honest because participants are not providing answers that they think are more socially acceptable or desired by researchers [35]. The numerical predictions made were analyzed to determine whether participants assigned higher probabilities to the visualizations that they had seen with metadata.

Relevance

Participants indicated their agreement with three scales related to the relevance of the information in the chart. Past visualization literature by Peck et al. found that personal connections to visualizations (i.e., its perceived relevance) were more important than other design decisions in readers' judgments of which visualizations were useful [33]. Therefore, we tried to capture participant perceptions of relevance to see if we observed a similar effect with respect to accuracy, clarity, completeness, and thoroughness. Further, metadata may impact perceptions of relevance by revealing information about the creator or process with which they already have a connection [11]. For instance, revealing that a visualization was made by an organization that the reader is already interested in could make the visualization seem more relevant. We asked participants about three dimensions: Meaningfulness, Relevance to Self, and Relevance to Others (see Table 3 for the exact statements provided to participants). We compared responses to these questions across participants to compare whether the presence of metadata had an effect on how relevant participants found the information in the visualization.

Understanding

After seeing each visualization, participants were asked to describe its main message and anything else that they learned via an open-ended response. We used an open-ended question in order to better understand participants' thought processes (e.g., [71]). We analyzed participants' responses to these questions in two ways.

Understanding: Response Correctness. To establish whether the presence of metadata had any impact on participants' abilities to correctly extract information from the visualization, we measured the correctness of the information in each response. Two coders rated each response on a 4-point scale (0-3) as a measure of how well it was supported by the visualization or metadata. The scale and the process by which it was created and applied are described in depth in Section 4.6.

Understanding: Response Content. We also qualitatively analyzed the responses to capture what participants found important or memorable enough to comment on. Details about the codes assigned and the method used to generate codes can be found below in Section 4.7.

Participants

We recruited 144 participants from Prolific. To derive our sample size, we conducted a power analysis based on data collected from a pilot of 24 participants that suggested 144 participants would yield 80% power to detect an overall difference between the subjective trust scores assigned for visualizations with and without metadata, at an alpha level of 0.05, assuming a medium effect size of 0.2. Following similar exclusion criteria as in Experiment 1, we filtered for participants that self-identified as fluent in English. We additionally excluded all participants that took part in Experiment 1. The study took about 15 minutes to complete and participants were paid $5.00.

Our participants were largely highly educated individuals who encountered or used visualizations semi-frequently. A majority of our participants had completed some education beyond high school (86.8%) and individuals who had completed a 4-year degree represented the largest portion of participants (38.19%). In addition, a majority of participants reported encountering or using visualizations at least once a week (42.36%) or once a month (31.25%).

Procedure

In Experiment 2, participants saw two pairs of visualizations (Step 1). Within each pair, one visualization was shown with metadata and the other without. After viewing

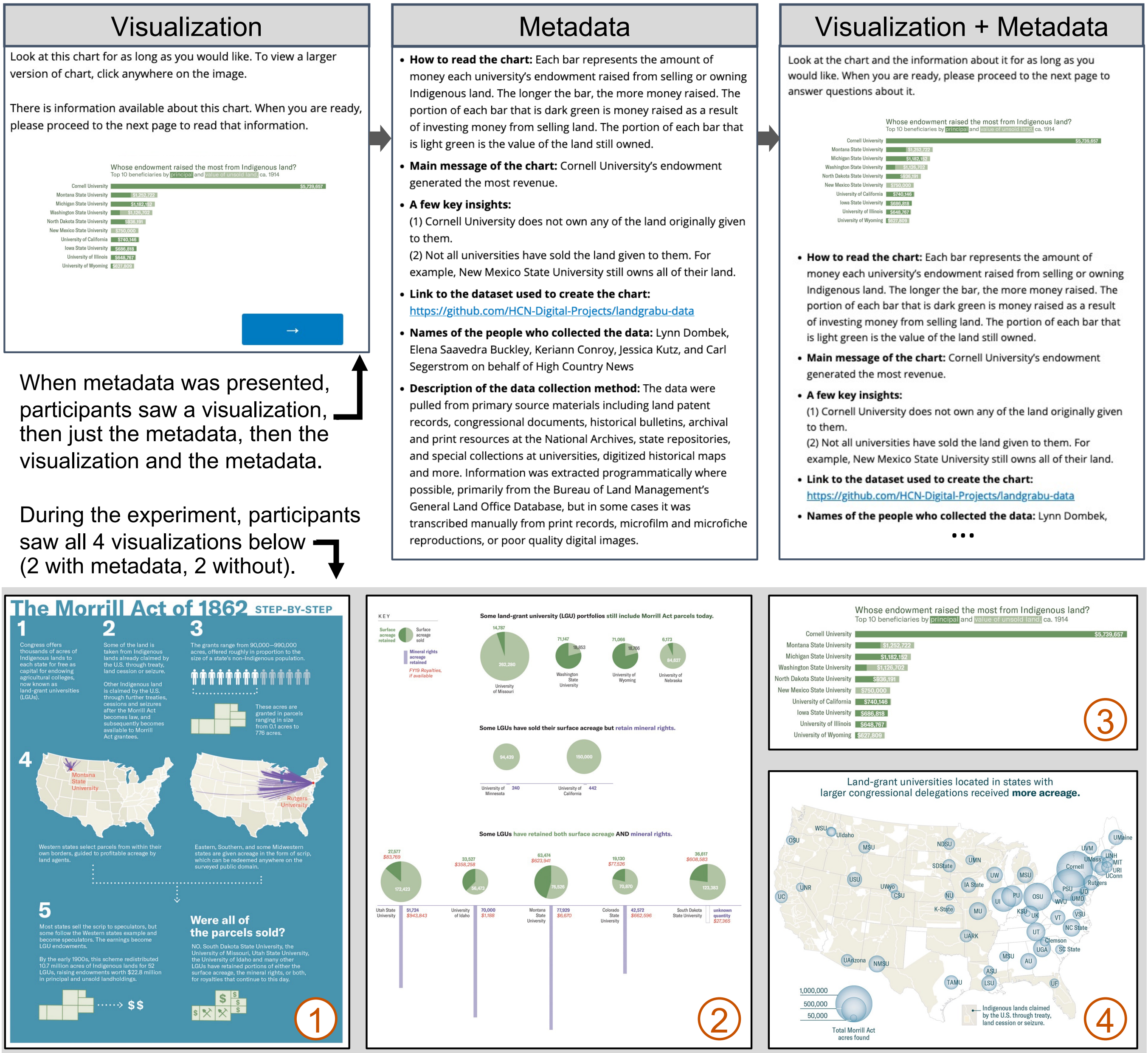


Fig. 2. When participants were shown metadata, they were presented with a screen containing just the visualization, a screen with only metadata, and a screen with both the visualization and metadata. Participants saw four visualizations from "Land-Grab Universities" [63] including an infographic about the Morrill Act (1), pie charts about land rights (2), a bar chart of endowments (3); and a map of land-grant universities and the amount of land given to each (4).

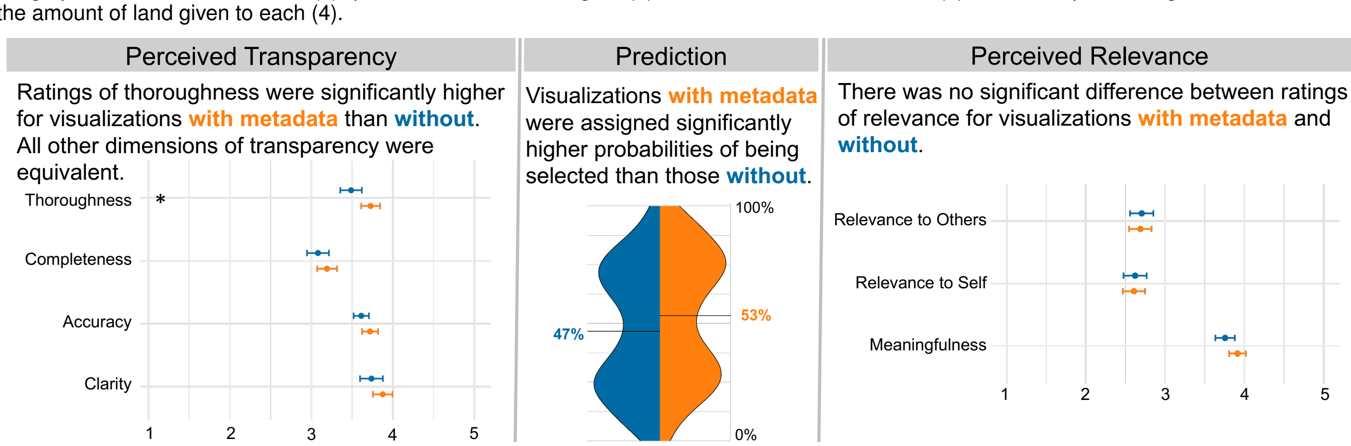


Fig. 3. Our results from Exp. 2 suggest that visualizations with metadata were perceived as more thorough, but similarly clear, accurate, complete, and relevant. Participants also assigned higher probabilities that visualizations with metadata were chosen for a presentation to policymakers. Error bars of ratings depict 95% Confidence Intervals.

Dimension	Definition	Likert-Scale Statement
Accuracy	The extent to which the information in the visualization is correct.	The chart was accurate.
Clarity	The extent to which the information in the visualization is understandable.	The chart was clear.
Completeness	The extent to which the visualization contains all of the possible components.	The chart told the whole story.
Thoroughness	The extent to which the visualization exhausts all possibilities.	The chart was thorough.
Meaningfulness	The extent to which the information is significant.	The information in the chart was meaningful.
Relevance to Self	The extent to which the information has a purpose for the reader.	The information in the chart was relevant to me.
Relevance to Others	The extent to which the information has a purpose for someone else.	The information in the chart was relevant to people I know.

TABLE 3
Participants rated each visualization by responding to seven 5-point Likert scales (1 - Strongly disagree to 5 - Strongly agree). Four scales assessed dimensions thought to be related to Perceived Transparency (Accuracy, Clarity, Completeness, Thoroughness), and three scales assessed Relevance (Meaningfulness, Relevance to Self, Relevance to Others).

each visualization, participants answered two open-ended understanding-based questions (Step 2) and then rated the visualization on a series of scales related to transparency and relevance (Step 3). After completing the tasks for each visualization in a pair, participants completed a prediction task in which they predicted the probability that the visualizations in the pair were selected by a hypothetical organization (Step 4). The order in which the visualizations were presented, accompanied with metadata, and compared was balanced using a Greco-Latin square design to prevent the ordering from biasing the results. All 24 permutations of the visualizations were used and the presence of metadata was balanced across conditions such that every chart appeared in each position in the order (e.g., first, second) three times with and without metadata. At the end of the study, the participants completed a demographic survey. The entire experiment was implemented in Qualtrics and participants were redirected back to Prolific after successful completion of the experiment.

4.2 Approach to Quantitative Analysis

In Experiment 2, participants rated each visualization on four dimensions related to Perceived Transparency (accuracy, clarity, completeness, and thoroughness) and three dimensions of Relevance. For both sets of measures, we analyzed our results using a multivariate analysis of variance (MANOVA), which is an analysis of variance (ANOVA) with two or more dependent variables [72]. Using MANOVAs allowed us to model each set of three or four ratings as dependent variables simultaneously, reducing the number of computations and the likelihood of Type I errors (falsely rejecting the true null-hypothesis) as opposed to conducting multiple independent ANOVAs [72]. In each MANOVA, we considered the effects of the presence of metadata, chart type, the order in which the charts were shown, the importance of the topic, and how frequently participants used charts on the sets of ratings. We report on the result of Pillai's trace (Pillai's value) because it is the most robust MANOVA test statistic [72], [73]. Pillai's value ranges from 0 to 1 and can be converted to an approximate F-statistic and then used to calculate a p-value [72]. MANOVAs can suggest an overall difference among values of some variable, but cannot determine which levels of that variable. For this, a post-hoc analysis must be conducted. Here, we conduct post-hoc analyses with Tukey's adjustment using the lme4 package [74] to correct p-values for multiple comparisons. The p-values generated in our analysis provides a sense of how unusual our data would be if all of the assumptions made when constructing the statistical model were correct [75], but given the complexity of our model, we caution against the dichotomous interpretation of these p-values (e.g., significant vs. not) using 0.05 as a threshold [76].

We also tested for collinearity to evaluate if our sets of three or four ratings measured different qualities using pairwise correlation coefficients and variance of inflation factors (VIF). Correlation coefficients can range from -1 to 1 where values close to -1 suggest a strong negative correlation between two variables, close to 0 suggest no correlation, and close to 1 suggest a strong positive correlation [77]. For variables to be considered independent, correlation coefficients should be close to 0. On the other hand, VIF values indicate the extent to which a variable can be described by a combination of the other variables. There is no consensus on what an unacceptable VIF value is, but higher VIF values suggest more interdependence [78].

For both the Prediction and Response Correctness scores, we used a linear mixed-effect model. This allowed us to model the prediction percentage and correctness score as a function of variables that we had measured (fixed effects) and individual differences between individuals (through a random intercept term). We modeled the Prediction percentage as a function of the chart shown, whether it was shown with or without metadata, and the four transparency-related ratings (to determine if the prediction, based on trust and persuasion, was influenced by participants' perceptions of transparency). Because we did not expect the Response Correctness scores to interact with transparency, we modeled the score only as a function of the chart shown and whether it was shown with or without metadata.

4.3 Results: Perceived Transparency

In our second experiment, we asked participants to rate each visualization on a set of four dimensions thought to be related to perceived transparency: Accuracy, Clarity, Completeness, and Thoroughness. Our results indicated that the four dimensions were impacted differently by the presence of metadata, chart type, importance of the topic, and how frequently participants used charts (see Figure 3). Our MANOVA analysis revealed a trending effect of the presence of metadata ($F(1, 538) = 2.27, p = 0.06$, Pillai's value = 0.02) and a significant effect of the chart type ($F(3, 538) = 8.03, p < 0.001$, Pillai's value = 0.17), importance of the topic ($F(5, 538) = 1.98, p < 0.01$, Pillai's value = 0.07), and frequency that participants used charts ($F(5, 538) = 2.23, p < 0.01$, Pillai's value = 0.08). We will now discuss the observed significant effects of each modeled variable. Details of all pairwise analyses can be found in the Supplemental Materials.

Effect of Metadata: The only dimension which was significantly impacted by the presence of metadata was thoroughness ($F(1, 538) = 7.71, p < 0.01$). Post-hoc analysis with Tukey's adjustment suggested that visualizations with metadata were considered to be significantly more Thorough than those without ($Est = 0.25, p = 0.005$).

Effect of Chart Type: Clarity ($F(3, 538) = 15.75, p < 0.001$), Completeness ($F(3, 538) = 11.51, p < 0.001$), and Thoroughness ($F(3, 538) = 9.62, p < 0.001$) were significantly impacted by the chart type. Post-hoc analysis suggested that that the pie chart visualization was perceived as significantly *less* Clear and the infographic was perceived as significantly *more* Complete and Thorough than the other three visualizations.

Effect of Topic Importance: Accuracy ($F(5, 538) = 4.17, p = 0.001$), Completeness ($F(5, 538) = 2.98, p = 0.01$), and Thoroughness ($F(5, 538) = 3.59, p = 0.003$) scores were impacted by the importance of the topic. Although the model suggested an overall difference among ratings of topic importance, we observed no significant effects in post-hoc pair-wise comparisons.

Frequency Charts were Used: Completeness scores were the only dimension significantly impacted by the frequency participants used charts ($F(5, 538) = 2.90, p = 0.01$).

Collinearity: Our measures of Accuracy, Clarity, Completeness, and Thoroughness were distinct, but not entirely independent (see Table 4). We can conclude that there is a small to medium correlation between the measures because all of the VIF values are greater than 1 but below 4. The highest correlation seemed to be between measures of completeness and thoroughness (see Table 4).

	Clarity	Accuracy	Completeness	Thoroughness	VIF
Clarity	1	0.41	0.48	0.50	1.43
Accuracy		1	0.51	0.48	1.45
Completeness			1	0.72	2.27
Thoroughness				1	2.24

TABLE 4
The four dimensions of transparency were not entirely independent. There was a small to medium correlation between the measures.

4.4 Results: Prediction

We also asked participants to indicate how probable they thought it was that an organization selected each of two visualizations for a presentation to policymakers. The prompt indicated that the organization had selected the chart because they deemed it the most persuasive and trustworthy. We used a linear mixed-effects model to investigate whether the predictions assigned by participants were (1) impacted by whether a visualization had appeared with or without metadata and (2) correlated with the ratings for the four dimensions of transparency that they had assigned in a previous step.

(1) Effects of Metadata: We observed a significant effect of metadata on the predictions ($\chi^2(1) = 8.06$, $p = 0.004$). Further, **our results suggest that visualizations with metadata were assigned higher probabilities of being selected than those without**. Post-hoc analysis with Tukey adjustments suggested that visualizations with metadata were assigned significantly higher probabilities in comparison to those without ($Est = 5.99$, $p = 0.004$).

(2) Correlations with Accuracy, Clarity, Completeness, and Thoroughness: Our linear mixed-effect model suggested a significant effect of Clarity ($\chi^2(1) = 13.53$, $p < 0.001$), but not of Accuracy ($\chi^2(1) = 0.12$, $p = 0.72$), Completeness ($\chi^2(1) = 0.26$, $p = 0.61$), or Thoroughness ($\chi^2(1) = 1.48$, $p = 0.22$). These results suggest that the probabilities that were assigned by participants may have been informed, in part, by perceptions of Clarity.

We observed a significant effect on the predictions from the chart ($\chi^2(3) = 40.70$, $p < 0.001$). Post-hoc analysis suggests that the pie chart was given significantly *lower* predictions than the other three visualizations ($Est = [-16.70, -13.52, -16.89], p < 0.001$). This is consistent with our results in the previous section which revealed that the pie chart visualization was assigned lower Clarity scores.

4.5 Results: Relevance

We additionally asked participants to rate each visualization on three Likert scales related to relevance: Meaningfulness, Relevance to Self, and Relevance to Others. Our results indicate that the three measures of relevance were impacted by the chart type, the importance of the topic, and how frequently participants used charts, but not the presence of metadata. The analysis revealed a *non-significant* effect of the presence of metadata (Pillai's value = 0.01), but a significant effect for all other variables (see Figure 3). Additionally, the three measures of relevance were not entirely independent. Because all of the VIFs are above 1, we can conclude that there is some correlation between all three measures (see Table 5). The VIFs for relevance to self and to others are high (above 3) and seem to be highly correlated.

	Meaningfulness	Relevance (Self)	Relevance (Other)	VIF
Meaningfulness	1	0.49	0.41	1.31
Relevance (Self)		1	0.81	3.18
Relevance (Other)			1	2.93

TABLE 5
The three measures of relevance were not entirely independent. Relevance to self and relevance to others were highly correlated.

4.6 Results: Understanding – Response Correctness

We asked participants two open-ended questions about each visualization. In the first question, participants described the main message of the visualization. In the second, they describe anything else that they learned. We analyzed each of these responses with respect to correctness, how they talked about the data, and the topics they mentioned.

Analysis

To analyze how well participants were able to extract accurate information from the visualizations and metadata, we generated a scale to rate how well responses were supported by the visualization and metadata. The scale was iteratively created by two authors. Two annotators generated an initial version of the scale based on scales from past work (e.g., in [79]). Once the initial scale was generated, both annotators used it to independently code three sets of 25 responses from each question that were randomly sampled. After each iteration, the annotators compared the codes they had assigned, discussed the scale, and then made revisions where needed. After three iterations, the inter-rater agreement (as measured by weighted Cohen's Kappa) was 0.61. After the scale was finalized, one of the annotators applied the scale to all of the responses. The other author verified the codes after they were assigned. The final scale used was as follows:

- 3 points: The response is *entirely supported* by the chart or metadata.
- 2 points: The response is *partially supported* by the chart or metadata but contains some inaccuracies.
- 1 point: The response is *neither supported nor refuted* by the chart or metadata.
- 0 points: The response is *directly refuted* by the chart or metadata, has no relevant information, or states that the reader does not know.

Results

Overall, a majority of the responses provided to the question about the main message (90%) and anything else learned (86%) were entirely supported by the information in the visualization or metadata.

Our results indicated that the presence of metadata had no significant impact on the correctness of descriptions of the main message. We used a linear mixed-effects model to investigate whether the correctness of a response to the main message question was impacted by the presence or absence of metadata. We did not observe a significant effect of metadata on the correctness of reader interpretation ($\chi^2 = 0.28$, $p = 0.60$), nor did the chart ($\chi^2 = 3.80$, $p = 0.28$). Our results also indicate no significant interaction between the two ($\chi^2 = 2.95$, $p = 0.40$). Similarly, our results suggested that the presence of metadata had no effect on the correctness of responses about what participants learned. The presence of metadata did not have a significant effect on the correctness of what participants learned ($\chi^2 = 2.73$, $p = 0.10$), nor did the chart type ($\chi^2 = 1.30$, $p = 0.73$). Our results also indicate no significant interaction between the two ($\chi^2 = 2.69$, $p = 0.44$).

4.7 Results: Understanding – Topics Discussed

Analysis

In addition to measuring how the presence of metadata impacted how accurately participants understood the visualizations, we also wanted to qualitatively evaluate what participants talked about. To investigate this, we iteratively developed a codebook (see Table 6) using an iterative method similar to the one used in [80]. To create the initial codebook, two annotators skimmed the dataset and generated an initial set of codes. They then completed three iterations on the codebook in which they independently coded a set of 25 responses from each question, compared the codes assigned, and then altered, added, or removed codes based on discussion and disagreements between assigned codes. After three iterations, the inter-rater agreement (as measured by Cohen's Kappa) was 0.82. Once all three iterations were complete and the annotators agreed on the codes, the annotators used the final codebook to apply codes to different sections of the qualitative data. These codes were verified by the other annotator and any disagreements were discussed and resolved. The entire codebook, with definitions for each code, is available in the Supp. Materials.

Category	Codes	Multiple
Discussions of Data	Extrema, Specific Point, Relationship, Correlation/Causation, Comparison, Process, Other	Yes
Topics of Interest	Named State, Named University, Geographical Region, Land, Money, Government, Laws, People, History, Theft, Other	Yes

TABLE 6
We iteratively generated a codebook to analyze participants' responses to questions about the main message and what else they learned from visualizations. Our codebook coalesced around two areas: (1) description and discussion of data and (2) topics of interest.

Results

Our qualitative analysis of the responses to the question about the visualization's main message and anything else learned suggested that **the presence of metadata did not change what participants perceived as the most dominant themes of the visualization.** For example, most of the responses referenced the use or ownership of land as a central aspect of the main message of all four visualizations (88.76% $n = 511$). For the map visualization, discussions of the use of land appeared frequently with comments related to governmental representation such as *"States with large congressional delegations received more land overall."* Similarly, participants described relevant laws when describing the infographic explaining the Morrill Act (e.g., "*It gave step-by-step information on the Morrill Act of 1862*"), and mentioned money in their descriptions of the visualization on college endowments (e.g., "*It showed the top ten beneficiaries of sold indigenous land and the money raised from it.*"). The dominance of these topics is not surprising, and instead reflects the repetition of topics in the visualization, the textual layer, and the metadata.

Our results also suggest that **metadata may direct readers' attention toward aspects of the visualization which may be less obvious or emphasized in the visualization alone.** For example, there are several possible main messages afforded by the bar-chart visualization about endowments, such as: "*Cornell University generated the most revenue*," "*The chart shows how much money each university has made from selling or owning indigenous land*," and "*Top 10 beneficiaries from indigenous lands*." However, our results indicated that participants who saw metadata were twice as likely to conclude that Cornell's endowment was the main message of the chart (33.3%, $n = 24$) in comparison to participants who saw the same chart without metadata (16.6%, $n = 12$). In this case, the metadata specifically mentioned two universities by name (Cornell and New Mexico State University), which may suggest that it impacted participants' interpretations (see Figure 2 for the metadata provided to participants).

We can see the pattern repeated in the features of the data that are mentioned in comments: **all participants mentioned the same dominant themes, but differed in the details**. For example, participants frequently commented on the distribution of data in the map visualization, irrespective of whether they had seen (79.17%, $n = 57$) or had not seen the metadata (91.67%, $n = 66$). However, participants who had seen the metadata made a comparison in their description of the main message more often (31.94% $n = 23$) than participants who did not see the metadata (13.89% $n = 10$). For instance, one participant concluded that: "*...The chart showed that the amount of land was not equal and Eastern universities got more on average.*"

Although rare, we also observed that some participants explicitly mentioned information that was *only* present in the metadata in their responses. The strongest example of this behavior was when one participant referenced the organization which produced several of the visualizations that we used, commenting that "***High Country News*** *seems to be interested in this situation*" (emphasis added).

5 Discussion & Future Work

Our two experiments showed the following main results:

- Visualizations with metadata were perceived as more thorough but similarly accurate, clear, and complete in comparison to visualizations without metadata.
- Visualizations with metadata were also assigned higher probabilities of being selected by a hypothetical organization than visualizations without on the basis of trustworthiness and persuasiveness.
- Participants did not perceive the information in visualizations with metadata as more relevant than those without.
- Metadata did not impact the accuracy of extracting information from the visualizations, but our results suggest that metadata might have directed readers' attention toward aspects of the visualization which were less obvious or emphasized in the visualization alone.

5.1 How do metadata impact trustworthiness and perceived transparency?

Our motivation for this study was to investigate whether the disclosure of metadata alongside visualizations would impact the way that visualization readers perceive transparency. Due to a lack of consensus on how to measure transparency, we operationalized four dimensions that are thought to contribute to transparency: accuracy, clarity, completeness, and thoroughness. Our results suggested that participants rated visualizations with metadata higher on thoroughness, but we observed no evidence of an effect of metadata on accuracy, clarity, or completeness. Further, participants predicted that it was significantly more likely that an organization would select the visualization with metadata for a presentation on the basis of trustworthiness and persuasiveness. These predictions were, in turn, impacted by participants' perceptions of clarity. These results suggest that metadata may impact perceptions of thoroughness and clarity. But how well do these dimensions translate to perceived transparency and trustworthiness? While the link between transparency and trust (or trustworthiness) is not often discussed in Computer Science, it is well documented in fields such as economics (e.g., [38]). Within visualization, past work that investigated the relationship between trust and accuracy, clarity, completeness, and thoroughness found that neither thoroughness nor clarity was a good predictor of participants' reported levels of trust [39]. Additionally, while prediction scores were higher for visualizations with metadata, our experimental design did not disambiguate whether participants were motivated because of trustworthiness or persuasiveness. Therefore, we must conclude that our results show weak support for the idea that metadata improve perceived transparency or trustworthiness.

However, studying transparency, trustworthiness, and the mechanisms behind building trust are difficult [10], [42]. This is, in part, because there is not yet agreement within visualization on a definition of trust, what factors compose it, or how it should be measured [42]. The same may be said of perceived transparency, which (like trust) is amorphous [81], but may be more concrete, less contextual, and has seen less study within the visualization community. Although the methods we employed in our study to assess transparency and trust are consistent with prior work (e.g., as summarized in [42]), there are certainly alternative ways of measuring these variables. For instance, researchers could consider directly asking participants how transparent they think a visualization is. This approach could be beneficial because the factors which influence or make up perceived transparency are complicated and not well-understood [81], but it also may be difficult for participants to answer because there are so many different definitions of transparency. We hope that the visualization community will see this work as a call to investigate the effects of metadata further and explore additional ways of measuring transparency and trustworthiness. For example, it may be fruitful to investigate whether metadata could help distinguish between credible and suspect visualizations or increase trust in credible visualizations and decrease trust in misleading ones.

Additionally, there is a rich literature surrounding the visualization of uncertainty and its relationship to trust. Both visualizing uncertainty within the data and the practice of disclosing metadata aim to increase transparency [82]. Rather than contradictory, these two techniques of reducing ambiguity may build upon one another. Similar to the results presented here, previous work on uncertainty found that the act of disclosing the presence of uncertainty also raised trust [83]. While uncertainty visualizations may use visual encoding to communicate the quantifiable uncertainty within the data [82], the practice of disclosing metadata may be able to elucidate the *source* of uncertainty or aspects of uncertainty that cannot be quantified easily. On the other hand, it is yet unclear whether visualization creators face the same apprehensions about disclosing metadata that they face to communicating uncertainty, such as a lack of resources and worry that disclosing uncertainty will impair understanding [84]. The visualizations that we used in our experiments did not directly visualize uncertainty, but future researchers may explore how communicating metadata and visualizing uncertainty can support each other.

5.2 How do metadata impact information relevance?

We also investigated the effects of metadata on how relevant information seemed, which we operationalized through three scales: meaningfulness, relevance to oneself, and relevance to others. We found that metadata did not have an impact on how relevant the information seemed to participants. In existing literature, there are few studies that measure the perceived relevance of information in a visualization. One example of how it has been operationalized in past work, however, was as a component of how readers talked about their communities [71]. Outside of visualization, researchers have measured perceived relevance by measuring whether participants choose to engage in particular behaviors (e.g., signing a petition [85]). Future work may therefore draw inspiration from other fields and explore the space of perceived relevance further.

5.3 How do metadata impact understanding & recall?

In Experiment 2, we conceptualized understanding in two ways: (1) The correctness of the responses given and (2) the topics that were mentioned within those responses. Our results suggested that the accuracy of responses provided by participants was not impacted by the presence of metadata, but it might have influenced what participants remembered as interesting or important. For example, participants who were given metadata about the map visualization were more likely to mention specific regions of the map in their descriptions, even beyond the regions which were specifically mentioned in the metadata.

Past work on understanding and recall found that textual elements like the title can have a strong impact on whether they recall the visualization accurately [47], [86]. For example, the title of a visualization can bias descriptions of the main message of a visualization toward the conclusion described instead of what is shown in the visualization [47]. The ability of textual elements to bias what participants remember as the main message is consistent with our results regarding the topics mentioned by participants in their responses. It is therefore possible that participants utilized metadata in a similar way to other textual elements. Future work could explore the interplay between the textual elements of a visualization and the metadata such as exploring the impact of perceived misalignment between the information in the visualization and metadata, as has been done with other elements of the textual layer (e.g., [47]).

However, if the metadata was used by participants in a similar way to other textual elements, then why did we observe no impact of metadata on the correctness of responses? (as has been observed in past work including [86]). One possible explanation is the visualizations that we used as stimuli. Namely, that the stimuli we used did not require much explanation. Past literature characterized the three basic types of visualizations that we used in our experiments (maps, bar charts, and pie charts) as very familiar to members of the general public [87], [88]. Nearly all participants were able to extract information from the visualizations that we provided, which may suggest a ceiling effect. When there were misunderstandings, this often seemed to be a result of misreading or misinterpreting the historical context. For example, some participants interpreted the visualization about money that was raised by colleges through the sale of formerly Indigenous land as the money that was raised by colleges to give to Indigenous people. Therefore, it is possible that metadata may impact the correctness of responses when accompanying more complicated or less familiar types of visualizations. Additionally, the lack of observed impact on understanding may have been a result of the prompts we used. While asking about the main message of a visualization as a measure of understanding is common (e.g., as in [47], [79]), this kind of question only probes one aspect of understanding. Past work has defined means of assessing participants' understanding at different levels which range in complexity from retrieving a single value to providing evaluations with evidence [89], [90]. Therefore, it might be that metadata impacts a different kind of understanding than the one we evaluated. Similarly, we only considered the correctness of the information with regard to the amount it was supported by the chart and metadata. Although common, this technique does not allow us to distinguish between the complexity of the information that was retrieved or the difficulty of retrieving it. For example, it may be easier to directly copy a number from a chart than to draw an inference from those numbers. Future work may therefore explore more complex ways of measuring understanding such as correctness alongside the mental effort required to extract information (as has been done in past work on the impact of pictographs [71]).

5.4 Tension Between Providing Text vs. Visuals

While our experiment setup mirrored the metadata disclosure methods of data journalism projects (i.e., as long-form text), it is worth asking: Are there more interesting or engaging ways to disclose metadata? This question may be especially prescient because there may be a trade-off between the amount of text provided, its readability, and its efficacy [91]. By definition, visualizations are intended to provide a *visual* representation of data to help readers complete tasks [92]. However, despite this intention, textual information has proven to be a successful component of narrative visualizations [93], better for extracting critical information [94], and can be added to existing visualizations in order to add context, heighten empathy, and introduce temporal references [95]. Nonetheless, we can still ask: how much text is too much? What alternatives are there to representing the metadata solely as text? One alternative to presenting the metadata solely as text might be to visualize the metadata. Visualizing textual data can enable readers to fully understand insights from large amounts of text [96]. However, visualizing textual data is also well understood to be very difficult (see [97] for an extensive discussion of the challenges including high dimensionality and irregularity) and some categories of metadata resist quantification (e.g., names, sources, or personal socio-cultural identities). However, future work may be able to leverage some visualization techniques to deliver the information differently or reduce the amount of information that is immediately presented to the reader. For example, in situations where some amount of narrative text is necessary, storytelling techniques like infographics, data comics [98], or cheat sheets [55] could be effective for integrating metadata into visualizations. Recent work on the use of infographics to communicate methodology has also found that infographics increased both accuracy and trust when compared to text-only communications [99]. Future work could investigate whether using infographics or storytelling techniques to communicate metadata might improve upon the results we observed in our experiments with only text.

5.5 Investigating Animation & Interactivity

While our experiments focused on static visualizations, interactivity and animation may afford new possibilities for communicating metadata. For example, past work demonstrated that animation can be used to effectively communicate steps taken in data analysis pipelines [100] or help readers better understand unfamiliar visualizations [101]. Future results could build upon these promising results to establish how these techniques impact understanding and perceptions of trustworthiness. Given that the value of different metadata varied significantly between user goals in Experiment 1, one could also consider using interactive technology like a conversational chatbot to deliver only the metadata that participants wanted. While it may be possible to guess some of the tasks readers want to complete, it may be advantageous to use interactive techniques to allow readers to directly access the metadata that they find most relevant. Recent work has outlined methods for the creation of chatbots for visualization which build and preserve trust(e.g., [102]). This work may be extended to explore how chatbots could communicate metadata and its impacts on other measures such as understanding and relevance.

Limitations

As with all studies, there are limitations to our work. First, our results may have been impacted by our participant pool. In both experiments, we recruited participants from Prolific who, while fairly diverse, were highly educated and may not be representative of visualization readers. Second, Experiment 1 used an exploratory design in order to reduce the complexity of the design space surrounding metadata, and inform the design of Experiment 2. However, the findings from that experiment would be reinforced by hypothesis-driven studies that interrogate their generalizability across diverse contexts. In Experiment 2, we used visualizations about a non-polarizing topic from sources that may not have been familiar to our participants. Future work may wish to examine how a polarizing topic or pre-existing levels of trust in a visualization source impacts the outcomes we observed in our study. Additionally, the design of our experiments specifically focused on assessing the impact of the metadata that participants *thought* would be most relevant on fairly simple, static visualizations. As a result, the design of our experiment pulled participants' attention toward the metadata and did not distinguish between the effects of different types of metadata. This was an intentional choice in order to ensure participants noticed the text, but it may not realistically reflect situations where people encounter metadata and revealed what the purpose of the experiment was to participants. Further, in Experiment 2, we did not provide definitions for any of the terms participants rated visualizations on (e.g., trust, accuracy), nor did we ask participants to provide their own definitions, which means that participants may have had different interpretations of these terms. As a result, it is unclear to what extent the results we observed generalize to specific types of metadata, visualizations, and situations where attention is not drawn to the metadata. Finally, we recognize that there are multiple ways to construct a statistical model to analyze data [103]. In our analysis of Experiment 2, we used a MANOVA model based on the justification provided in Section 4.2. We experimented with other analytical approaches (e.g., a Bayesian multivariate multilevel model [104], available in the Supplemental Materials) and found consistent results. Future work may explore alternative approaches to systematically identify new research directions in this space.

Several of our design choices for the Prediction component of Experiment 2 may have influenced our results. In our scenario, we did not provide a description of the organization that "selected" the visualizations. Participants' assumptions of what kind of organization this was may have influenced how they interpreted the prompt. Additionally, we told participants that the organization had selected the visualizations on the basis of trust and persuasion. While this choice allowed us to make sure all participants judged the visualizations on similar criteria, the design does not allow us to disambiguate whether the visualizations were selected because of trust or persuasion (or both). These criteria may also not be meaningful to all participants and a more open-ended prediction scenario may have been able to gather more insight into the characteristics participants valued as well as how metadata relates to those characteristics. Future work may therefore consider providing further information on the hypothetical organization and utilizing open-ended prediction scenarios.

6 Conclusion

There is a lot about visualization that goes unsaid. This ambiguity is a reflection of a lack of transparency and can have impacts on understanding. In this paper, we interrogated the claim that providing metadata alongside a visualization could be an effective way to increase transparency and understanding. We adopted a broad definition of metadata and an associated six-category taxonomy from the authors' past work and conducted two experiments. In the first, we investigated which categories of metadata readers of visualizations think are relevant to specific goals (e.g., assess trustworthiness, build confidence). Based on the results of our first experiment, we ran a second experiment to study the impacts of the metadata that our first group of participants *thought* was most relevant on four dimensions thought to be related to perceived transparency, three dimensions of information relevance, and understanding. We found that visualizations with metadata were perceived as more thorough, but similarly accurate, clear, complete, and relevant in comparison to visualizations without metadata. Additionally, we found that the presence of metadata did not impact the correctness of descriptions given by participants, but may have impacted what they saw as important enough to mention. Our results raise further questions about the potential role and impacts of metadata in visualization which we hope will inspire the visualization community to investigate further.

Acknowledgments

The authors would like to thank the anonymous reviewers for their thoughtful feedback, Steve Haroz for his feedback on the reporting of statistics, and members of the HCI-VIS lab for their feedback on early versions of the paper.

REFERENCES

[1] S. Lee *et al.*, "How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 499–508, 2015.

[2] A. Romano *et al.*, "The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences," *Health economics*, vol. 29, no. 11, pp. 1482–1494, 2020.

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

[4] L. G. Jaimes, T. Murray, and A. Raij, "Increasing trust in personal informatics tools," in *Design, User Experience, and Usability. User Experience in Novel Technological Environments*, A. Marcus, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 520–529.

[5] M. Dörk *et al.*, "Critical infovis: exploring the politics of visualization," in *CHI'13 Extended Abstracts on Hum. Factors in Comput. Sys.* ACM, 2013, pp. 2189–2198.

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

[7] Q. Wang *et al.*, "Narvis: Authoring narrative slideshows for introducing data visualization designs," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 779–788, 2019.

[8] K. C. Elliott and D. B. Resnik, "Science, policy, and the transparency of values," *Environmental health perspectives*, vol. 122, no. 7, pp. 647–650, 2014.

[9] R. Horvath and D. Katuscakova, "Transparency and trust: The case of the european central bank," *Applied Economics*, vol. 48, no. 57, pp. 5625–5638, 2016.

[10] E. Mayr, N. Hynek, S. Salisu, and F. Windhager, "Trust in information visualization," in *EuroVis Workshop on Trustworthy Vis. (TrustVis)*. The Eurographics Association, 2019.

[11] A. Burns *et al.*, "Making the invisible visible: Risks and benefits of disclosing metadata in visualization," in *2021 IEEE Workshop on Vis. for Social Good (VIS4Good)*, 2021, pp. 11–15.

[12] M. Couper, "Measuring survey quality in a casic environment," *Proc. of the Survey Research Methods Section of the ASA at JSM1998*, pp. 41–49, 1998.

[13] J. O'Reilly, "Paradata and blaise: A review of recent applications and research," in *Int. Blaise Users Conf. (IBUC), Latvia*, 2009.

[14] A. G. D. Holmes, "Researcher positionality–a consideration of its influence and place in qualitative research–a new researcher guide," *Shanlax Int. J. of Education*, vol. 8, no. 4, pp. 1–10, 2020.

[15] Prolific, "Prolific," 2022, https://www.prolific.co/.

[16] K. Xu *et al.*, "Survey on the analysis of user interactions and visualization provenance," *Comput. Graphics Forum*, vol. 39, no. 3, pp. 757–783, 2020.

[17] P. Buneman, S. Khanna, and T. Wang-Chiew, "Why and where: A characterization of data provenance," in *Int. Conf. on Database Theory*. Springer, 2001, pp. 316–330.

[18] Y. L. Simmhan, B. Plale, and D. Gannon, "A survey of data provenance in e-science," *ACM Sigmod Record*, vol. 34, no. 3, pp. 31–36, 2005.

[19] E. D. Ragan *et al.*, "Characterizing provenance in visualization and data analysis: an organizational framework of provenance types and purposes," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 31–40, 2015.

[20] S. Haroz, "Open practices in visualization research: Opinion paper," in *2018 IEEE Evaluation and Beyond-Methodological Approaches for Vis. (BELIV)*. IEEE, 2018, pp. 46–52.

[21] L. Battle *et al.*, "Evaluating visual data analysis systems: A discussion report," in *Proc. of the Workshop on Hum.-In-the-Loop Data Analytics*, 2018, pp. 1–6.

[22] L. C. Burgess, "Provenance in digital libraries: Source, context, value and trust," in *Building trust in information*. Springer, 2016, pp. 81–91.

[23] M. Borenstein *et al.*, *Introduction to meta-analysis*. John Wiley & Sons, 2021.

[24] P. A. Lamal, "On the importance of replication," *J. of Social Behavior and Personality*, vol. 5, no. 4, p. 31, 1990.

[25] T. Gebru *et al.*, "Datasheets for datasets," *Communications of the ACM*, vol. 64, no. 12, pp. 86–92, 2021.

[26] P. Fyfe, "An archaeology of victorian newspapers," *Victorian Periodicals Review*, vol. 49, no. 4, pp. 546–577, 2016.

[27] M. Q. Foote and T. Gau Bartell, "Pathways to equity in mathematics education: How life experiences impact researcher positionality," *Educ. Stud. in Math.*, vol. 78, no. 1, pp. 45–68, 2011.

[28] M. Savin-Baden and C. Howell-Major, "Qualititative research: The essential guide to theory and practice," *Qualitative Research: The Essential Guide to Theory and Practice. Routledge*, 2013.

[29] D. Haraway, "Situated knowledges: The science question in feminism and the privilege of partial perspective," *Feminist studies*, vol. 14, no. 3, pp. 575–599, 1988.

[30] C. D'Ignazio and L. F. Klein, *Data feminism*. MIT Press, 2020.

[31] C. Lee *et al.*, "Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online," in *Proc. CHI Conf. Hum. Factors Comput. Sys.*, 2021, pp. 1–18.

[32] W. E. Rowe, "Positionality," in *The SAGE encyclopedia of action research*, D. Coghlan and M. Brydon-Miller, Eds. Sage, 2014.

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

[34] H. Kennedy *et al.*, "Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness," *First Monday*, vol. 21, no. 11, 2016.

[35] C. Xiong, L. Van Weelden, and S. Franconeri, "The curse of knowledge in visual data communication," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 10, pp. 3051–3062, 2019.

[36] A. Munzel, "Assisting consumers in detecting fake reviews: The role of identity information disclosure and consensus," *J. of Retailing and Consumer Services*, vol. 32, pp. 96–108, 2016.

[37] R. F. Kizilcec, "How much information? effects of transparency on trust in an algorithmic interface," in *Proc. CHI Conf. Hum. Factors Comput. Sys.*, 2016, pp. 2390–2395.

[38] A. K. Schnackenberg and E. C. Tomlinson, "Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships," *J. of management*, vol. 42, no. 7, pp. 1784–1810, 2016.

[39] C. Xiong *et al.*, "Examining the components of trust in map-based visualizations," in *EuroVis Workshop on Trustworthy Vis. (TrustVis)*. The Eurographics Association, 2019.

[40] M. Pirson and D. Malhotra, "Foundations of organizational trust: What matters to different stakeholders?" *Organization Science*, vol. 22, no. 4, pp. 1087–1104, 2011.

[41] J. Thomas and J. Kielman, "Challenges for visual analytics," *Information Vis.*, vol. 8, no. 4, pp. 309–314, 2009.

[42] H. Elhamdadi *et al.*, "How do we measure trust in visual data communication?" in *2022 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Vis. (BELIV)*, 2022.

[43] S. Boyd Davis, O. Vane, and F. Kräutli, "Can I believe what I see? data visualization and trust in the humanities," *Interdisciplinary Science Reviews*, vol. 46, no. 4, pp. 522–546, 2021.

[44] F. Windhager *et al.*, "Visualization of cultural heritage collection data: State of the art and future challenges," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 6, pp. 2311–2330, 2018.

[45] K. Kelton, K. R. Fleischmann, and W. A. Wallace, "Trust in digital information," *J. of the American Society for Information Science and Technology*, vol. 59, no. 3, pp. 363–374, 2008.

[46] E. Costante, J. den Hartog, and M. Petkovic, "On-line trust perception: What really matters," in *2011 1st Workshop on Socio-Technical Aspects in Security and Trust (STAST)*, 2011, pp. 52–59.

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

[48] N. Li *et al.*, "Communicating data: interactive infographics, scientific data and credibility," *J. of Science Communication*, vol. 17, no. 2, p. A06, 2018.

[49] E. Yakel *et al.*, "Trust in digital repositories," *The Int. J. of Digital Curation*, vol. 8, 2013.

[50] R. D. Frank *et al.*, "Trust in qualitative data repositories," *Proc. of the Association for Information Science and Technology*, vol. 54, no. 1, pp. 102–111, 2017.

[51] G. M. Alarcon *et al.*, "Trust perceptions of metadata in open-source software: The role of performance and reputation," *Systems*, vol. 8, no. 3, p. 28, 2020.

[52] B. Wasike, "Memes, memes, everywhere, nor any meme to trust: Examining the credibility and persuasiveness of COVID-19-related memes," *J. of Computer-Mediated Communication*, vol. 27, no. 2, p. zmab024, 2022.

[53] X. Lin, P. R. Spence, and K. A. Lachlan, "Social media and credibility indicators: The effect of influence cues," *Computers in Hum. behavior*, vol. 63, pp. 264–271, 2016.

[54] D. Sohn and S. Choi, "Social embeddedness of persuasion: effects of cognitive social structures on information credibility assessment and sharing in social media," *Int. J. of Advertising*, vol. 38, no. 6, pp. 824–844, 2019.

[55] Z. Wang *et al.*, "Cheat sheets for data visualization techniques," in *Proc. CHI Conf. Hum. Factors Comput. Sys.* New York, NY, USA: ACM, 2020, p. 1–13.

[56] C. C. Hall, L. Ariss, and A. Todorov, "The illusion of knowledge: When more information reduces accuracy and increases confidence," *Organizational Behavior and Hum. Decision Processes*, vol. 103, no. 2, pp. 277–290, 2007.

[57] S. Hall, "Encoding and decoding in the television discourse [originally 1973; republished 2007]," in *Essential Essays, Volume 1: Foundations of Cultural Studies*. Duke University Press, 2019, ch. 8, pp. 257 – 276.

[58] Y.-S. Kim *et al.*, "A bayesian cognition approach to improve data visualization," in *Proc. CHI Conf. Hum. Factors Comput. Sys.*, 2019, pp. 1–14.

[59] Y.-S. Kim, K. Reinecke, and J. Hullman, "Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 760–769, 2017.

[60] Sigma Awards, "About — the sigma awards," https://sigmaawards.org/about/, accessed: 2022-01-30.

[61] ——, "The COVID Tracking Project at The Atlantic," https://sigmaawards.org/the-covid-tracking-project-at-the-atlantic/, accessed: 2022-01-30.

[62] ——, "Mapping makoko," https://sigmaawards.org/mapping-makoko/, accessed: 2022-01-30.

[63] ——, "Land-grab universities: How expropriated indigenous land became the foundation of the land-grant university system," https://sigmaawards.org/land-grab-universities-how-expropriated-indigenous-land-became-the-foundation-of-the-land-grant-university-system/, accessed: 2022-01-30.

[64] ——, "Rough justice: How police are failing survivors of sexual assault," https://sigmaawards.org/rough-justice-how-police-are-failing-survivors-of-sexual-assault/, accessed: 2022-01-30.

[65] ——, "Who gets to breathe clean air in new delhi?" https://sigmaawards.org/who-gets-to-breathe-clean-air-in-new-delhi/, accessed: 2022-01-30.

[66] V. Nightingale, Ed., *The handbook of media audiences*. Wiley Online Library, 2011.

[67] H. K. Jach, C. G. DeYoung, and L. D. Smillie, "Why do people seek information? the role of personality traits and situation perception." *J. of Exp. Psychol.: General*, 2021.

[68] Qualtrics, "Qualtrics," 2022, https://www.qualtrics.com.

[69] M. Hervé, "Package 'RVAideMemoire'," 2020, https://CRAN.R-project.org/package=RVAideMemoire.

[70] M. Salem and K. Dautenhahn, "Evaluating trust and safety in hri: Practical issues and ethical challenges," *Emerging Policy and Ethics of Hum.-Robot Interaction*, 2015.

[71] A. Burns *et al.*, "Designing with pictographs: Envision topics without sacrificing understanding," *IEEE Trans. Vis. Comput. Graph.*, 2021.

[72] R. T. Warne, "A primer on multivariate analysis of variance (manova) for behavioral scientists." *Practical Assessment, Research & Evaluation*, vol. 19, 2014.

[73] H. Finch, "Comparison of the performance of nonparametric and parametric manova test statistics when assumptions are violated." *Methodology: European J. of Research Methods for the Behavioral and Social Sciences*, vol. 1, no. 1, p. 27, 2005.

[74] D. Bates *et al.*, "Package 'lme4'," *Convergence*, vol. 12, no. 1, p. 2, 2015.

[75] S. Greenland *et al.*, "Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations," *European J. of Epidemiology*, vol. 31, no. 4, pp. 337–350, 2016.

[76] S. Haroz, Email communication, 2022.

[77] R. Taylor, "Interpretation of the correlation coefficient: a basic review," *J. of diagnostic med. sonog.*, vol. 6, no. 1, pp. 35–39, 1990.

[78] C. G. Thompson *et al.*, "Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results," *Basic and Applied Social Psychol.*, vol. 39, no. 2, pp. 81–90, 2017.

[79] S. Bateman *et al.*, "Useful junk? the effects of visual embellishment on comprehension and memorability of charts," in *Proc. CHI Conf. Hum. Factors Comput. Sys.*, 2010, pp. 2573–2582.

[80] K. Mack *et al.*, "What do we mean by "accessibility research"? a literature survey of accessibility papers in chi and assets from 1994 to 2019," in *Proc. CHI Conf. Hum. Factors Comput. Sys.*, 2021, pp. 1–18.

[81] C. A. Van der Cruijsen and S. C. Eijffinger, "From actual to perceived transparency: The case of the european central bank," *J. of Economic Psychol.*, vol. 31, no. 3, pp. 388–399, 2010.

[82] L. Padilla, M. Kay, and J. Hullman, *Uncertainty Visualization*. John Wiley & Sons, Ltd, 2021, pp. 1–18.

[83] D. Sacha *et al.*, "The role of uncertainty, awareness, and trust in visual analytics," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 240–249, 2015.

[84] J. Hullman, "Why authors don't visualize uncertainty," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 1, pp. 130–139, 2019.

[85] Y. Luo and J. Zhao, "Motivated attention in climate change perception and action," *Frontiers in Psychol.*, p. 1541, 2019.

[86] M. A. Borkin *et al.*, "Beyond memorability: Visualization recognition and recall," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 519–528, 2015.

[87] A. Quispel, A. Maes, and J. Schilperoord, "Graph and chart aesthetics for experts and laymen in design: The role of familiarity and perceived ease of use," *Information Vis.*, vol. 15, no. 3, pp. 238–252, 2016.

[88] K. Börner *et al.*, "Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors," *Information Vis.*, vol. 15, no. 3, pp. 198–213, 2016.

[89] A. Burns *et al.*, "How to evaluate data visualizations across different levels of understanding," in *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Vis. (BELIV)*. IEEE, 2020, pp. 19–28.

[90] E. Adar and E. Lee, "Communicative visualizations as a learning problem," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 946–956, 2020.

[91] C. Stokes and M. A. Hearst, "Give text a chance: Advocating for equal consideration for language and visualization," in *NLViz Workshop*, 2021.

[92] T. Munzner, *Visualization analysis and design*. CRC press, 2014.

[93] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1139–1148, 2010.

[94] A. Ottley *et al.*, "The curious case of combining text and visualization." in *EuroVis (Short Papers)*, 2019, pp. 121–125.

[95] A. Figueiras, "Narrative visualization: A case study of how to incorporate narrative elements in existing visualizations," in *2014 18th Int. Conf. on Information Vis.*, 2014, pp. 46–52.

[96] N. Cao and W. Cui, *Overview of Text Visualization Techniques*. Paris: Atlantis Press, 2016, pp. 11–40.

[97] M. Alharbi and R. S. Laramee, "Sos textvis: An extended survey of surveys on text visualization," *Computers*, vol. 8, no. 1, p. 17, 2019.

[98] B. Bach *et al.*, "The emerging genre of data comics," *IEEE Computer Graphics and Applications*, vol. 37, no. 3, pp. 6–13, 2017.

[99] M. Kiernan *et al.*, "Effects of a methodological infographic on research participants' knowledge, transparency, and trust," *Health Psychol.*, vol. 37, no. 8, p. 782, 2018.

[100] X. Pu *et al.*, "Datamations: Animated explanations of data analysis pipelines," in *Proc. CHI Conf. Hum. Factors Comput. Sys.*, 2021, pp. 1–14.

[101] P. Ruchikachorn and K. Mueller, "Learning visualizations by analogy: Promoting visual literacy through visualization morphing," *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 9, pp. 1028–1044, 2015.

[102] V. Setlur and M. Tory, "How do you converse with an analytical chatbot? revisiting gricean maxims for designing analytical conversational behavior," *arXiv preprint arXiv:2203.08420*, 2022.

[103] P. Dragicevic, Y. Jansen, A. Sarma, M. Kay, and F. Chevalier, "Increasing the transparency of research papers with explorable multiverse analyses," in *Proc. CHI Conf. Hum. Factors Comput. Sys.*, 2019, pp. 1–15.

[104] P.-C. Bürkner, "brms: An r package for bayesian multilevel models using stan," *Journal of statistical software*, vol. 80, pp. 1–28, 2017.