Communicative Information Visualizations: How to make data more understandable by the general public

COMMUNICATIVE INFORMATION VISUALIZATIONS: HOW TO MAKE DATA MORE UNDERSTANDABLE BY THE GENERAL PUBLIC

A Dissertation Presented

by

ALYXANDER BURNS

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2022

Robert and Donna Manning College of
Information and Computer Sciences.

COMMUNICATIVE INFORMATION VISUALIZATIONS: HOW TO MAKE DATA MORE UNDERSTANDABLE BY THE GENERAL PUBLIC

A Dissertation Presented

by

ALYXANDER BURNS

Approved as to style and content by:

Narges Mahyar, Chair

Cindy Xiong, Member

Michelle Trim, Member

Steven Franconeri, Member

James Allan, Chair of the Faculty
Robert and Donna Manning College of
Information and Computer Sciences.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Narges Mahyar. You have taught me more about research, visualization, HCI, and writing than I could name. I am eternally grateful for your support, knowledge, and inspiration throughout our time working together. I would also like to thank my committee members, Cindy Xiong, Michelle Trim, and Steve Franconeri. To Cindy, Steve, & Evan Peck – thank you for your collaboration, mentorship, and support. It has been a joy to work with you all and have learned so much from you. To Michelle, thank you for pushing me to look at who and what isn't been said in my scholarship and my teaching; I am endlessly grateful for your support. I am grateful to my friends in CICS and to my labmates in the HCI-VIS lab including Mahmood Jasim, Hamza El-Hamdadi, Prateek Mantri, Emma Anderson, Hia Ghosh, and Emily Pruc. Our discussions have been an boundless inspiration for me in my life and my work and I hope that I can repay the support and encouragement that you have given me. I would also like to thank Emma Anderson, Erika Dawson-Head, the staff of the Office of Professional Development, and Kate Litterer for their guidance, wisdom, and encouragement. Finally, I would like to thank my family, my partner, and my friends outside of UMass for believing in me and for encouraging me throughout this experience. Mom, Dad, Zoë, Frau, Hannah, & Lillie, I could not have made it through this without you.

ABSTRACT

COMMUNICATIVE INFORMATION VISUALIZATIONS: HOW TO MAKE DATA MORE UNDERSTANDABLE BY THE GENERAL PUBLIC

SEPTEMBER 2022

ALYXANDER BURNS

B.A., MOUNT HOLYOKE COLLEGE

M.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Narges Mahyar

Although data visualizations have been around for centuries and are encountered frequently by the general public, existing evidence suggests that a significant portion of people have difficulty understanding and interpreting them. It might not seem like a big problem when a reader misreads a weather map and finds themselves without an umbrella in a rainstorm, but for those who lack the means, experience, or ability to make sense of data, misreading a data visualization concerning public health and safety could be a matter of life and death. However, figuring out how to make visualizations truly usable for a diverse audience remains difficult. In my thesis, I examined three areas where altering current practices may help make data visualizations more understandable and impactful in the future.

First, I conducted a critical analysis of the ways that audiences of data visualizations are defined in visualization research papers. Poorly defining the audience of a

visualization can have negative effects for both people and science. For example, it may lead to over-generalizing research results or guidelines for audiences that were never intended (or evaluated) in the original research or ignoring the (potentially unique) needs of different populations by aggregating them into majority groups. Therefore, in order to investigate current practices, I conducted a survey of every paper published in a major visualization venue that referred to "novices," "non-experts," the "general public," or "laypeople" in their title or abstract. I selected this set of audiences because they represent some of the words used to describe visualization's broadening audience. My analysis of 79 papers demonstrated that audiences were rarely explicitly defined. In place of explicit definitions, authors relied on implicit definitions (e.g., examples, counter-examples) to clarify who was in the audience. To improve how audiences are defined in future work, I draw on feminist theory and research methods to argue that authors should: (1) Explicitly define their audiences by specifying the dimensions which qualify an individual for inclusion, (2) Think of their audiences intersectionally by considering how multiple dimensions of identity and experience can produce unique perspectives on visualization, and (3) be careful not to reinforce existing oppressive power structures.

Next, I examined current visualization design techniques to test how more explicitly encoding data as an array of countable pictographs (instead of as solid abstract shapes) may impact understanding and experience. In existing research, infographics show positive empirical findings in terms of memory, engagement, and assessment of risk — particularly when they contain pictographs (simple, iconic pictures that represent a word or topic). However, there was little exploration of how pictographs affect and afford the general public's understanding of the underlying data or how the choice to use pictographs affects readers' personal experiences. Therefore, I conducted an experiment that utilized a novel method of producing questions that probe different aspects of a reader's understanding of 6 pairs of real-world visualizations which are

identical except for their use of pictograph arrays. My results indicated that the use of pictograph arrays does not directly impact understanding but can allow readers to more easily envision real-world connections.

Finally, I explored how accompanying a visualization with contextual information could impact understanding and experience. While visualizations are designed to present a multitude of data, they often are not accompanied by key metadata which provides background on the source of the data, the transformations applied to the data, the visualization elements, its purpose, the people involved in its creation, or its intended audience. Current conventions surrounding leaving the context of visualization invisible can have negative impacts for understanding, transparency, and ultimately trust. Common wisdom suggests that recontextualizing visualizations with metadata (e.g., disclosing who made the visualization, the data source, or instructions for decoding the visualizations' encoding) may counter these effects and increase understanding, transparency and, ultimately, trust. However, the impact of adding metadata to visualizations remains largely unknown. To fill this gap, I conducted two experiments. In Experiment 1, I explored what kinds of metadata participants valued the most and how chart type, topic, and user goal impacted their choices of metadata. My 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 trust. Based on the results of Experiment 1, in the second experiment, I explored how these two types of metadata impact trust, information relevance, and understanding. I 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 by an organization for a presentation to policy makers. My results suggested that among four dimensions of transparency, visualizations with metadata were perceived as more thorough but similarly accurate, clear, and

complete in comparison to those without. Additionally, visualizations with metadata were assigned higher probabilities of being selected by a hypothetical organization for a presentation for policy makers. However, participants did not perceive the information in visualizations with metadata as more relevant than those without. Finally, the presence of metadata did not impact the accuracy of extracting information from the visualizations, but may have influenced which information participants remembered as important or interesting.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

Members of the general public rely on data visualizations to make sense of the world. For example, individuals may consult a graph of COVID-19 health outcomes to decide whether to get vaccinated or examine a map of test positivity rates to determine the risk of traveling to see relatives. A majority of the visualizations that members of the general public encounter are considered **communicative data visualizations** because they are crafted to present, support, or explain conclusions or ideas [1].

Typically in data visualization, data-driven decision making processes are referred to as **sensemaking**. The most commonly used model of sensemaking, proposed by Pirolli and Card, contains both a foraging loop where data are sought, searched, filtered, and read as well as a sensemaking loop where those data are used to iteratively construct a mental model [292]. This process typically involves time-consuming analysis and re-analysis, with the intention of finding analytical "ah-ha" moments. It is also often conducted by expert analysts who have access to the original data and the resources to independently analyze it [292].

In contrast, members of the general public are conducting what may be considered **casual** sensemaking: understanding, reflecting upon, and making judgements or decisions based on information provided by casual (communicative) visualizations [297] without deep and time-consuming research using raw data [54]. Instead of being employed for the purpose deep analyses, casual sensemaking processes may be used to help people become more aware of issues relevant to their community or to self-reflect

[297]. This definition of casual sensemaking is therefore more similar to a definition of sensemaking from business and management in which the viewer uses a conceptual model of the world informed by available information "to comprehend, understand, explain, attribute, extrapolate, and predict" [12, 362].

Unfortunately, evidence suggests that a large proportion of the general public struggles to conduct casual sensemaking processes with visualizations. Past surveys have estimated that about 60% of Americans encounter significant difficulty using visualizations to draw accurate conclusions [148, 381]. Disentangling the factors that make visualizations particularly challenging for some people is difficult because it is deeply intersectional. For instance, while difficulty understanding visualizations is sometimes thought to correlate with education [148], highly educated populations (including doctors) also struggle [309]. Nonetheless, these results indicate that, at least for some portion of the population, data visualizations are often failing to provide the information that they were specifically designed to communicate.

To make matters worse, difficulty making sense of visualizations is not just an inconvenience — it can have very serious consequences. It may not seem like a big deal to misunderstand a weather prediction and end up without an umbrella, but can be far more serious when the visualization regards health and safety information. For instance, misunderstanding a visualization of where a hurricane will make landfall might lead a person to stay at home even when they are in grave danger. Further, being unable to understand data may prevent individuals from accessing social services that they are entitled to [418]. Therefore, it is critical that we find ways to create visualizations that allow more people to make sense data.

However, creating visualizations which work for the general public is complicated. In particular, past work has characterized the general problem of creating visualizations for non-expert audiences as a **wicked problem** – a problem without a defined solution, without even one way to define or test for success, which cannot be "solved"

with an algorithmic solution [317, 83, 151]. Therefore, by definition, the design of visualizations which enable the general public to equitably utilize the data they encode must be studied from multiple angles to find approaches and techniques which get closer to the goal.

Therefore, in my thesis, I examined three areas where changing current practices might help make data visualizations more understandable and impactful for the general public in the future: Research methodology, Visual design, and the text that accompanies visualizations. First, I took a critical look at the place where best practices and guidelines for creating visualizations are derived – visualization research – and examined how the ways that visualization audiences are defined may have unintended consequences (Chapter 3). In this work, I surveyed 79 published visualization papers which used one of four words for their audience (novice, non-expert, layperson, general public) and argued that these audiences are largely implicitly (and simply) defined which, in turn, can result in poor outcomes such as overaggregating audiences with differing needs into majority groups or concluding that visualizations "work" for people that they were never intended to be used by. Next, I studied how changing the design of visualizations could impact sensemaking. As a case study, I measured the impacts of converting abstract shapes into countable pictograph arrays on casual sensemaking. My analysis revealed that while pictographs did not impact aspects of sensemaking related to comprehension, they did impact experience (Chapter 4). Finally, I studied the practice of revealing implicit information about the history and context of a visualization and how providing that information may impact different sensemaking (Chapter 5). My analyses revealed that the kinds of implicit information readers want access to depends on what they will do with it and that providing readers with information that they *think* is relevant does not impact their ability to extract accurate information from the visualization or make the information seem

more relevant, but may have directed readers' attention to different features of the visualization.

Scope & Definitions

In this dissertation, I focus on three areas where changes to current practice could create **communicative visualizations** that enable **casual sensemaking** for members of the **general public**. I define the bolded keywords and phrases as follows:

- **Communicative visualizations** (or, visualizations for communication) refer to data visualizations which are designed to present, support, or explain conclusions or ideas. In particular, I focus on static visualizations which are relatively simple and display (at most) a few hundred datapoints and three dimensions.
- **Casual sensemaking** refers to understanding, reflecting upon, and making judgements or decisions based on information provided by casual (communicative) visualizations [297] without deep and time-consuming research using raw data [54].
- In this work, the **general public** refers to adults who typically do not have any formal training on how to read or use data visualizations but already encounter them in some form (e.g., at work, in the news). The portion of the general public who I focus on in this thesis are also able to visually perceive visualizations, can read and write in English, and have high-speed internet access.

Contributions

The contributions of this thesis are:

Chapter 3 — The Audience Matters! A Critical Analysis of Audiences in Visualization

- An analysis of how "novice," "non-expert," "layperson," and "general public" are defined in visualization research papers revealing that these audiences are rarely defined explicitly and definitions which are present rely on only one aspect of an individual's attitudes, experiences, or skills.
- A discussion of the unintended consequences of defining audiences implicitly and one-dimensionally.
- Suggestions for improving current practice in three areas which draw inspiration from fields such as critical race theory and feminist theory.

Chapter 4 — Designing with Pictographs: How does Design Impact Casual Sensemaking?

- A novel framework that provides a systematic way to evaluate levels of understanding in a visualization based on Bloom's Taxonomy.
- Empirical results that suggest that infographics with pictograph arrays are just as good as more traditional, geometric part-to-whole charts at helping people make sense of data.
- Empirical insights that show some participants view charts containing pictograph arrays as easier to envision, while others view them as unnecessarily cluttered and slower to understand.
- Qualitative results suggesting that perceptions of visual appeal may be impacted by ease of understanding, making the topic seem real may help readers envision the topic more easily, and feelings of urgency and importance may be influenced by real-world connections.

Chapter 5 — From Invisible to Visible: Impacts of Metadata in Communicate Data Visualization

- A comprehensive taxonomy of six types of metadata and discussion of the potential benefits and risks of disclosing each.
- Results from an exploratory study suggesting that what kinds of metadata people may perceive as most important varies according to their goals (e.g., assess trust, understand perspectives).
- Quantitative results suggesting that metadata can increase perceptions of thoroughness, which may positively impact the perceived transparency of a visualization.
- Quantitative results suggesting that the information in visualizations with metadata are not considered more relevant than in visualizations without
- 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.
- A discussion the possible implications of the results on trust, understanding, and the tension between disclosing information and providing more textual information.

CHAPTER 2

BACKGROUND

In this section, I will outline existing work relevant to all three components of the thesis. First, I will talk about the study of visualization for communication. Then I will discuss how visualizations have historically been evaluated and the debates surrounding these practices. Finally, I will describe what sensemaking is, contrast the notion of sensemaking common in visualization to the broader definition used in this thesis, and describe how this too can be evaluated.

2.1 Designing Communicative Visualizations

There are countless books, web articles, and guides on how to create effective visualizations for communication (e.g., [57, 58, 205, 379, 255, 336]). Unfortunately, many of the design guidelines presented in these texts are not backed by evidence, but instead rely on convention and assumptions [207]. However, there has been recent work evaluating some of these long-held beliefs that have both found evidence for some of these guidelines (e.g., removing clutter and providing visual focus [4]) and against others (e.g., embellishment is always harmful [19, 14]). See Franconeri et al. and Schwabish [131, 336] for an exhaustive list of evidence-backed guidelines and best practices. Nonetheless, there is still work to be done to validate the effectiveness of many common practices.

There has also been recent work on tools for creating communicative visualizations. There are several tools that focus broadly on the creation of visualizations that can be used for communication which vary broadly from code-based tools intended

for use by tech-savvy individuals with coding experience (e.g., D3 [38], Vega [330]) to code-free tools aimed at the general public (e.g., Tableau [369]). Further, tools to broadly and easily create visualizations are embedded many popular presentation software systems (e.g., Microsoft Powerpoint [251], Google Slides [149], Prezi Present [300]). There are also visualization creation tools intended for more specialized uses such as creating visualizations to talk about climate change [146] and generating simple part-to-whole visualizations from natural-language phrases [96], and automatically creating infographic-like "fact sheets" from tabular data [391].

Although there seems to be considerable interest in studying visualizations for communication within the visualization research community, the pool of existing research on the design of communicative visualizations is historically limited but increasing in size.

2.2 Evaluating Visualizations

A critical component of designing visualizations that meet need of their audiences is effective and appropriate evaluation methods [207, 117]. While evaluation techniques which do not involve human participants (e.g., heuristic evaluations) are sometimes still conducted on visualizations or visualization algorithms, studies involving human participants (also referred to as user studies) have become increasingly common over time [185]. User studies are particularly critical to the evaluation of visualizations because they can help establish which visualization techniques are most appropriate for the context and establish where theory differs from reality [209].

Historically, the evaluation of visualizations with human participants was focused on capturing metrics like time and accuracy [326, 117]. These two metrics have both been used on their own (e.g., in [166, 353, 208]) and used as a proxy to measure other phenomena such as knowledge acquisition [264] and working memory [166]. While the relevance of the accuracy component is fairly straightforward, proponents

of measuring time have argued that if visualizations are intended to make content easier to understand, then "saving time" is a useful metric of measuring success [70]. Nonetheless, the reliance on time and accuracy as evaluation metrics has received fierce critiques.

Critics of focusing on time and accuracy argue that these methods are not rich enough to capture the breadth of human experience. For example, existing work has pointed out that time and accuracy alone cannot capture: how hard it was for a person to complete a task [178], what insights participants drew [264], or how much fun or enjoyable visualization was [326]. Nonetheless, under the right circumstances, all of these metrics could all be considered metrics of success [326].

Critics therefore suggest that novel techniques need to be used to supplement time and accuracy by capturing other critical aspects of the experience of using visualizations. For instance, North proposed new methods for measuring insights by asking difficult, open-ended questions [264] and Mahyar et al. proposed utilizing Bloom's taxonomy to evaluate the depth of users' engagement with visualizations [242]. The desire for new metrics and evaluation methods beyond time and accuracy have even spurred workshops specifically on the topic (most notably the biannual BELIV Workshop [338]).

In sum, there is a general agreement among researchers that time and accuracy are not the only ways to evaluate visualizations, in part due to a diversity of opinions on what the "right" way to evaluate a visualization is [386]. Further, it is clear that there is still a need for novel methodologies that can more holistically capture what makes particular visualizations valuable and appropriate in different circumstances.

2.3 Sensemaking

Casual sensemaking, as defined in this thesis, is related to but distinct from the model of sensemaking proposed by Pirolli and Card which is often used in the visual-

ization community [292]. In the Pirolli and Card model, the sensemaking process is made up of the following activities: collecting information, creating a mental model of that information in a way that helps analysis, manipulating that model to produce an insight, and finally producing an idea or action based on that insight [292]. This process is not linear and instead involves iteratively looping between the steps until sense is made. In contrast, casual sensemaking does not necessarily involve the same kind of rigorous, analytic process. It can instead be thought of as an ongoing process [103] of using information provided by a visualization to update ones' mental model of the world [362, 12] in order to reflect and make decisions. By extension, sensemaking is the process by which understanding is obtained [103] and (in line with existing work on emotional design) is influenced both by the information received as well as the reader's experience [263]. Accordingly, in this work, I will define casual sensemaking as measurable through a combination of assessing both understanding and experience. While the visualization research community has many tools for evaluating the experiential aspects of these processes, it still lacks comprehensive methods for evaluating understanding.

As visualization is still a relatively new field, we can take inspiration from the effective methods used in other disciplines. For example, in visual literacy research, questionnaires comprised of multiple choice questions have been used to quantify how fluently people understand graphics in general (e.g. [227]) and in medical contexts (e.g. [136, 267]). We see evidence of similar use of questionnaires in Medicine and Chemistry to, for example, measure patients' understanding of Informed Consent documents [189] and assess the public's understanding of nanotechnology [334]. Further, benchmarks have been used to measure knowledge of topics important to the job performance of professional psychologists and to evaluate the way that this knowledge appears in on-the-job behavior [130]. In another approach, researchers used a set

of open-ended questions about Climate Change to establish the effect of combining refutation texts with graphics and analogies on understanding [99].

Additionally, there are several existing methods within visualization which in some way accomplish this goal. For example, one study asked participants to describe one thing they found interesting or surprising [387]. Collecting salient points is a very direct way of assessing what participants found important, but only requiring one insight may mean that the information reported is not necessarily indicative of all that they have learned. Further, while it measures what the perceived takeaway is, it is not able to assess other related levels of learning, such as how the knowledge could be applied to a new situation; a step critical to the learning process [286].

Another approach asked participants to describe everything that came to mind as they interacted with a visualization [265]. Due to the open-ended nature of this method, it is certainly thorough, but it may not be possible in remote settings and relies on the ability of the participant to verbalize their thoughts reliably. Additionally, one downside of the approach (as identified in the original paper) is the sheer volume of qualitative information that is generated [265]. Our approach considerably reduces this burden, condensing the amount of information collected into only six questions.

A third approach in this vein asked participants to describe different aspects of a visualization such as the overall message, trend, and sentiment [19]. By requiring participants to comment on each aspect of interest, this method is effective at assessing the breadth of knowledge obtained by participants, but it too misses the opportunity to assess other kinds of knowledge acquisition which might have revealed other differences in the stimuli.

As I have summarized, while some researchers are asking more difficult questions that assess particular aspects of understanding, **they all lack a system for comprehensively evaluating different levels of understanding**. Such concrete models can also help to measure affordances that are provided by visualizations to help read-

ers obtain knowledge. This will enable visualization designers to assess their readers beyond graph literacy and create more effective visualizations for their audience.

2.4 Summary

In summary, in this thesis, I will focus on the design and evaluation of communicative visualizations that enable members of the general public to more easily perform casual sensemaking. Communicative visualizations share the same perceptual building blocks as other kinds of visualizations, but are designed specifically to present, support, or explain ideas. There are countless resources that dispense advice on how to design effective visualizations, though much of it is not backed by evidence. One component of designing visualizations well is having thorough methods for evaluation. Visualization evaluation practices have faced critiques (in part because of their historical focus on "usability" metrics), but many researchers have shifted toward including more holistic metrics. One of these goal metrics is sensemaking, which, in this thesis, is considered to be a combination of both a reader's understanding and their experience. Understanding has been quantified in many ways, but also has faced critiques due to a historical reliance on time and accuracy. Therefore, in this thesis, I will utilize open-ended questioning to evaluate understanding using methods inspired by the field of education.

CHAPTER 3

THE AUDIENCE MATTERS! A CRITICAL ANALYSIS OF AUDIENCES IN VISUALIZATION

In the Introduction of my thesis, I described the problem that many people who need to make use of visualizations have difficulty understanding them. To approach solving this problem, I began by investigating where best practices for creating visualizations come from – the visualization research process. In particular, I investigated how researchers define and discuss the people who they imagine will use their visualization – the visualization's audience. By exploring practices for defining audiences, I analyzed how current practices may have unintended consequences which, ultimately, may lead to the creation of visualizations that do not meet the needs of the people who use them. This chapter represents the contents of [50].

3.1 Introduction

Poorly defining the audience of a visualization can have negative, unintended consequences for the ways that visualizations are designed and evaluated, the guidelines that are produced, and the people who will ultimately use them. For example, by using language which broadly defines an audience, researchers may risk ignoring the (potentially unique) needs of different populations by aggregating them into majority groups. The effects of aggregating divergent populations into majority ones is likely to disproportionately impact groups that have been historically left out of the visualization literature because their needs have not been identified as distinct from the majority within which they are contained (e.g., in the case of casual visualization users [297]).

Further, these cases can implicitly discourage the scientific community from continued research in some topics by incorrectly implying that the research has already been covered (e.g., that visualization for "novices" has been done already) and produce "one size fits all" designs which do not meet the needs of the people they are to be used by (e.g., as with accessibility guidelines [221]). One of the audiences which visualization has expanded in recent years to study and create visualizations for is "novices" [170]. Grammel and Heer et al. provide divergent definitions for the term. In one, novices are *"users who create visualizations to support their primary tasks, but who are typically not trained in data analysis, information visualization, and statistics"* [152], while in the other they are *"users who have experience operating a computer, but no experience with programming in general, let alone programming visualization techniques"* [170]. These divergent definitions demonstrate how a lack of specificity in who belongs to an audience could lead to over-generalizing research results or guidelines for audiences that were never intended (or evaluated) in the original research — had their explanations been less clear, readers of the two papers might generalize the results across their audiences even though they refer to entirely different groups of people.

Past work has emphasized the broadening audiences for visualizations and created visualizations for audiences including (but not limited to) domain experts [403], casual users [297], non-experts [68, 172], and the general public [54, 406]. In this work, we use the word "**audience**" to refer to a group of people who share similar attributes, experiences, perspectives, or interests and are expected to read or otherwise engage with a visualization. Namely, the audience is the group of people who the visualization designers have in mind when they design a visualization and the group of people from which they sample participants when they conduct human-subject studies.

To investigate how audiences are typically defined in visualization research, we conducted a survey of papers published in IEEE, ACM, and Eurographics visualiza-

tion venues which used the terms "novice," "non-expert," "layperson," or "general public" in their abstract or title ($n = 79$). We focused on this group of audiences because they are used to refer to the diverse set of individuals in the expanding audience of visualizations (as discussed in [39, 170]). Three coders collected data on how audience was defined explicitly and implicitly and how participants were selected for user studies. Taking the characteristics we observed in the survey as a case study, we argue that there are fundamental problems with the way that audiences are typically defined in visualization papers — namely, that explicit definitions of who belongs in an audience are minimal and that the implicit definitions which *are* provided are based on only one aspect of an individual's identities, interests, or experiences.

We take inspiration from research methodology and feminist theory (e.g., [222, 163]) to provide concrete suggestions for how researchers can define their audiences more inclusively and thoroughly in three areas. First, we suggest that authors **make the implicit explicit** by providing clear definitions and inclusion/exclusion criteria which clarify who the authors think belongs in the audience and what attributes are necessary for their inclusion. Second, we encourage authors to **think of their audiences intersectionally** by considering how multiple dimensions of identity and experience can produce unique perspectives on visualization. Finally, we advise that when defining an audience, authors **be aware of systems of power** so that the ways an audience is defined does not cause harm through reinforcing existing oppressive power dynamics. Our suggestions may better equip readers of visualization papers to understand to whom research applies and how it fits in with existing work. By clearly articulating who they are studying or for whom they are building a visualization, researchers may better align their experimental methods with their intentions. We do not argue that there is one "correct" way to talk about or define an audience. Instead, the purpose of this critique is to encourage visualization researchers to think critically about how they identify and talk about their audience.

There are three main contributions of this paper. First, we contribute an analysis of how "novice," "non-expert," "layperson," and "general public" are defined in visualization research papers revealing that these audiences are rarely defined explicitly and definitions which are present rely on only one aspect of an individual's attitudes, experiences, or skills. Next, we provide a discussion of the unintended consequences of defining audiences implicitly and one-dimensionally. Finally, we provide suggestions for improving current practice in three areas which draw inspiration from fields such as critical race theory and feminist theory.

3.2 Background

3.2.1 What are Audiences?

The word "audience" has many different (and sometimes conflicting) meanings [397, 278]. Drawing on literature from marketing and communication, an audience may be broadly defined as "a group of individuals who are more similar to each other than to individuals in other groups" [354, 158]. While the word "audience" might evoke images of people sitting in chairs listening to a speech or performance, the word "audience" does not need to refer to a literal, physical group of people at all, but the collective set of knowledge and motivation shared by a group of people which will provide context to a work (for instance, as used in the phrase "know your audience") [278]. Further, audiences are not "observable, objective realities but instead constructed entities that emerge in popular memory, institutional practice, and academic research" [366, 9]. In other words, an individual is not *always or solely* a member of an audience — "audience member" is a role "that people temporarily perform" depending on the circumstances [55, 366]. In this paper, we will build upon these conceptions and use the word "audience" to refer to "a group of people who share similar attributes, experiences, perspectives, or interests and are expected to read or otherwise engage with a visualization." In this way, we focus on the role

that people play when they engage with a visualization — while engaging with the visualization's content, they are members of the visualization's audience.

Scholars in media studies have argued that there are three ways of thinking of the role of the audience: as a group of people who consume media ("as mass"), as people who are acted upon by media ("as outcome"), and as people who act upon media ("as agent") [397]. These three ways of thinking about an audience suggest different research direction: viewed as a mass, research questions could be about what media the audience consumes; as an outcome, how people are affected by the media; and as agents, what people do with media (e.g., make meaning of, make choices about) [397].

Within the context of visualization, prior works engaged with all three ways of viewing the audience's role. For example, visualization work that conceptualizes the audience as a mass has asked what kinds of visualization a specific group of people engaged with [37] or faced barriers to engaging with [123]. Viewed as an outcome, researchers have asked questions about how visualizations or visualization design choices impact people including their sensemaking [54], decision making [102, 275], and recall [35]. Finally, viewed as agents, visualization researchers have asked about what people do with visualizations such as investigating how visualizations are shared and discussed [224], constructed [152], and perceived [283, 269].

Our goal with this work is to examine the ways that audiences are defined in a visualization context, inspect the possible unintended consequences of those actions, and propose strategies for how future work might not repeat the challenges of the past.

3.2.2 Broadening Audiences in Visualization

In past years, the visualization community has discussed the expansion of visualization audiences. While visualization was once the realm of data analysts, the

creation and consumption of visualization has been adopted by broader audiences [170, 297, 223]. For example, existing work has characterized unique audiences such as "casual users" who engage with visualizations for personally meaningful, non-work purposes [297, 402] and "domain experts" who use visualizations to accomplish work-related tasks (e.g., data analyses) [403].

Along with the push for broader audiences, visualization research has observed many differences between groups of people. For example, a meta-analysis of eye tracking studies revealed that expert visualization users identified task-relevant information faster and returned to those areas more often than intermediate or novice users [142]. Work investigating the "Big Five" personality traits (Extraversion, Neuroticism, Openness, Conscientiousness, and Agreeableness) from the Five-Factor Model [147] have found differences such as introverts (people with low Extraversion) gathered more insights from the data and highly neurotic individuals completed tasks faster [155] (see [232] for a comprehensive review of cognitive traits investigated in visualization work).

To make visualizations that meet the needs and abilities of these distinct audiences, authors have argued that it is necessary to understand aspects of an audience such as their goals, needs, personality, knowledge, and the circumstances when a visualization will be used (e.g., [170, 111, 434]). For example, prior work has emphasized the importance of finding imagery and analogies which connect with the experiences of the audience (e.g.,[260, 100]) and adapting designs and interfaces to fit their skills and interests (e.g., [152, 170, 341]).

The broadening audiences of visualization and the wide array of differences between groups emphasizes the need to define audiences of visualizations and visualization research findings well. In this work, we employ a *critical analysis* to investigate the ways that these definitions are done and to question the assumption that current practices for defining audiences are sufficient.

3.2.3 Critical Analyses for Examining Assumptions

Broadly speaking, critical analyses examine and question the underlying assumptions and practices of a particular field or area [122]. For example, critical analyses (and critical theories more broadly) form the bases of areas such as Critical Race Theory and have been employed to examine power within race, gender, and geographies (among others) [109, 101]. Critical analyses can range in scope from "macro" studies of society-wide relationships to "micro" scale analyses of the practices within specific organizations [101] and provide an opportunity to examine issues which might have otherwise been ignored.

Critical analyses have been utilized in visualization to explore how visualization techniques may be applied to new domains and to point out existing problems with current practice. For example, researchers have speculated about how visualization could be used for DNA Microarray data analysis [428] and visualizing urban planning data [296]. Other authors have used critical analyses to argue that text visualization is political[24], consider the moral and ethical dilemmas of visualization[87], and critique how visualizations are evaluated for medical use [298].

Scholars have also written about what overarching critical theories for visualization and human-computer interaction might look like. For instance, Dörk et al proposed the idea and need for "Critical InfoVis" which actively considers disclosure, plurality, contingency, and empowerment when building visualizations [109]. Additionally, Ogbonnaya-Ogburu et al. adapted central ideas in Critical Race Theory to the domain of HCI to form a "Critical Race Theory for HCI" and demonstrated how those tenets appear in the lived experiences of the paper's co-authors [266]. In this paper, we aim to examine the underlying assumptions about how authors define audiences. In this process, we aim to apply the ideas from these existing theories to a specific domain rather than to model an overarching critical theory.

Most closely related to this work are critical works about the ways that audiences are defined. One example of this kind of work within the context of visualization focused on the lack of a coherent definition for "domain experts" [403]. While that work aimed to solve the problem of a missing definition directly by proposing one and exploring its implications, this paper focuses on a broader problem of *how* to construct clear definitions. Also related is recent work from HCI questioning the idea of the "user" including work investigating how people are represented in machine learning models [66], arguing that systems are no longer built for a single type of user but a set of diverse stakeholders [129], and suggesting that focusing on only the direct "users" means that other types of human-machine interactions are ignored [23]. Like these works, we too wish to complicate the idea of the audience and encourage readers of this paper to think critically about who they are designing for or studying and how they are communicating that to the readers of their own work.

3.3 Methodology: An Analysis of Audience Definitions

To dissect current practices for defining audiences and their unintentional consequences, we present evidence from an analysis of "core" visualization papers which used the terms novice, non-expert, general public, or layperson[1].

3.3.1 Paper Selection

The focus of our search was papers from "core" visualization research venues during any year of their publication. Namely, we sampled papers published in the IEEE Transactions on Visualization and Computer Graphics (TVCG) journal or published as a part of IEEE VIS, InfoVis, VAST, EuroVis, BELIV, and the ACM Conference on Human Factors in Computing Systems (CHI). We selected these venues based on

[1] We will use "layperson" throughout as a catch-all for "layman" or "layperson" and their plurals.

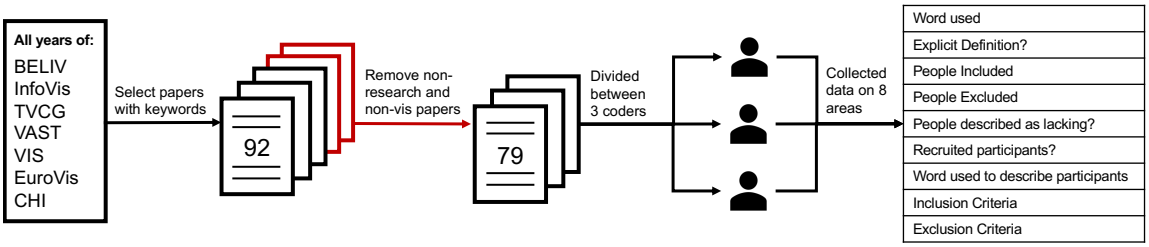


Figure 3.1. We selected all papers published in IEEE, ACM, and Eurographics visualization venues that contained "novice," "non-expert," "general public," or "layperson" in their title or abstract (n = 92). We then excluded 13 papers which were not visualization research papers, resulting in a final set of 79 papers.

past work which similarly surveyed visualization research practices [107]. Our paper selection process is summarized in Figure 3.1.

We conducted a title and abstract keyword search within the publishers' respective digital libraries to retrieve relevant papers. The keywords we used were "novice," "non-expert," "general public," and "lay(man, men, person, people)[2]." We selected these keywords because they represent a set of words used for non-traditional visualization audiences which are sometimes cited as having difficulty understanding visualizations and thus requiring bespoke visualization or visualization creation tools (e.g., [152, 341, 294]).

We used the IEEE, ACM, and Eurographics Digital libraries to obtain papers which matched our keywords. Because the Eurographics Digital Library does not allow users to search within the titles and abstracts directly, we conducted a full-text keyword search of EuroVis papers and then manually filtered the results to find those which satisfied the title and abstract requirement. Additionally, papers from CHI had to also include the word "visualization" in their title or abstract because CHI is not solely a visualization venue. To ensure we did not miss any related results from CHI, we manually skimmed through all CHI results which did not use the word "visualization" but included one of the other terms. Our inclusion criteria resulted

[2]We use the word "layperson" throughout to refer to any matches in this set of keywords.

in a total of 92 papers (89 + 3 from skim). Documentation of the specific queries used to retrieve the papers used in our survey is available on our online repository: `https://osf.io/rdcsf/?view_only=d638db5ca96945ed9765134ffa450d8a`

From our list of 92 results, we excluded 13 results that were not visualization research papers. Namely, we excluded two speeches [91, 258] and eleven papers which focused solely on animation, virtual reality, or augmented reality (all published in TVCG). After applying our exclusion criteria, we ended up with a final set of 79 papers. Note that our set of papers included papers of all lengths (e.g., full-length papers, short papers, extended abstracts). We decided not to apply any exclusion criteria based on the length of the paper because we wanted to get a holistic view of how these terms were being used across paper types. Further, restricting to only papers of a particular length would have been difficult given that the page layout and page limits vary between conferences and may have changed over time.

3.3.2 Data Collection

A team of three coders collected data from each paper with overlap to increase reliability. To begin, the coders created a codebook that described the information they wanted to collect, its purpose, and the process by which they would obtain it. For instance, coders used the keyboard function Control/Command-f to search for the terms novice, non-expert, general public, and lay(man, men, person, people). **The information collected included:**

- the context of each term used (e.g., if "novice" was used as a modifier for "user")
- the groups who were included and excluded as defined directly by the paper
- whether a group was directly described as "lacking" in something (e.g., motivation, literacy, a skill)

- *In cases in which a user study included one of the keyword audiences:* which words the authors used to describe the participants (e.g., students, staff, crowd-workers, volunteers), inclusion criteria, and exclusion criteria

3.3.2.1 Rationale for data collection

Given that our goal was to investigate how audiences were defined in the literature, we chose to collect the context in which words were used to describe audiences, as well as the explicit definitions provided for those terms. After a small, initial pilot suggested that explicit definitions were infrequent, we expanded our data collection to include people described as *included* or *excluded* based on work in social psychology on in- and out-groups (e.g., [62]). To collect these included and excluded groups, we collected the groups named and described within the same paragraph of any of the keywords. We restricted to within one paragraph to be more confident that the words were being used in relation to each other. A group were marked as **included** if they were given as an example (e.g., *"Such comparison tasks are challenging for novice ML practitioners who have primary but not comprehensive ML knowledge background ... For example, a medical school graduate student may want to adopt a CNN for disease detection"* [415]) or if the keyword group was described in terms of what they thought or did (e.g., *"Non-expert users have difficulties to comprehend the coherency of input, parameters, and output of these algorithms"* [45]). We did not count this kind of description as an explicit definition because it is not clear whether it is necessary for a member of the group to have this kind of thought or difficulty or whether it was supplied as an example. Alternately, a group was considered to be **excluded** if they were named in sequence with one of our keywords (e.g., *"Visualization of general relativity illustrates aspects of Einstein's insights into the curved nature of space and time to the expert as well as the layperson."* [153]) or were directly contrasted against them (e.g., *"While most of these systems are geared towards domain experts ...*

CrowdLayout focuses on novice crowds" [352]). We also collected information on participants and how they were selected because participants are supposed to represent the audience. Finally, we coded for descriptions of people as "lacking" because the audiences we surveyed (novices, non-experts, the general public, and laypeople) are non-traditional visualization audiences which may be compared to more traditional audiences and could be described in a negative light in these comparisons.

3.3.2.2 Data collection process

Using the codebook, coders independently coded the same set of 10 randomly selected papers, met to discuss their codes, resolve differences, and make any necessary changes to the code book. The rest of the papers were divided among the coders (19 or 20 per coder). After all coders had coded half of their assigned papers, they all coded a second common set of 10 papers. Again, they met to discuss their codes and resolve differences. The mean pairwise inter-coder reliability for the second set of papers was an average 0.63 (SD=0.02) as calculated using unweighted Cohen's Kappa. Once all differences had been settled, the coders coded the remainder of their papers and adjusted previously coded papers as needed.

After collecting the data, one of the coders categorized the explicit and implicit definitions provided according to their complexity. Each explicit definition provided was tagged as either one-dimensional if inclusion in the audience required only one condition (e.g., "little experience with visualization") or multi-dimensional if inclusion required two or more conditions to be met (e.g., "little experience with visualization and biology concepts). Terms which were not explicitly defined but were defined *implicitly* (through the included or excluded groups) were also categorized according to their complexity. Implicit definitions were tagged as "one-dimensional," "multi-dimensional," or "unclear" when multiple groups were included or excluded but it was difficult to ascertain what criteria united or divided them.

3.4 Analysis and Discussion: Characteristics of Audience Definitions

The purpose of this analysis was to examine how a selection of four audiences are defined in visualization research. In the following section, we will report our findings regarding these definitions and discuss some potential implications. The results of our data collection, per paper, are summarized in Table 3.2.

3.4.1 Audiences were rarely explicitly defined.

Explicit definitions are desirable because they directly and completely describe what is meant by a term, thus fulfilling their explanatory purpose [27]. Among the 79 papers we examined, only 15 contained an *explicit definition* clearly identifying which people were novices, non-experts, members of the general public, or laypeople (see Table 3.1 for counts and example definition). Two explicit definitions included "novice Vega users", which were people who were *"unfamiliar with Vega"* [172], and "novice ML practitioners", which were people with *"primary but not comprehensive ML knowledge background[s]"* [415].

3.4.1.1 Implicit Definitions

In place of explicit definitions, authors provided *implicit definitions* in 56 papers by sharing examples of audience sub-groups or describing characteristics of audience processes or background (see Table 3.1 for counts). For example, novices were said to have difficulty *"effectively utilizing GPU clusters"* [75] or *"lack the knowledge and expertise in data visualization"* [134]. Non-expert users were said to have difficulty *"managing the complexity of visual parameters"* of volumetric functions [327] and think that *"large network visualizations [were] overwhelming, confusing and contain[ed] too much detail"* [384]. Although implicit definitions clarify some of the characteristics of the words being defined, they rarely provide the clarity and precision of explicit definitions [401]. Further, not all examples are equally helpful — in

order for an example to effective, a reader must be able to identify that the example can generalize to a broader set of problems or rules [245]. Without sufficient context around the example, it may not be clear to the reader *what* should be generalized.

In 53 of 79 papers, authors also provided implicit definitions through *counter-examples* by describing people who were *not* in the audience (see Table 3.1 for counts). In many papers, "experts," was used as a contrast to "novices" and "non-experts." For instance, non-expert algorithm users were contrasted against *"experts in data mining techniques (e.g., engineers or biologists)"* [45]. There were also more specific counter-examples provided such as in Chan et al., which contrasted people in the general public against *"music lovers, who have received extensive training in music theory and history"* [65]. We distinguish between "counter-examples" and "negative definitions" as such: "counter-examples" are cases where a group is named or described in a way that makes it ambiguous who else is included or excluded, while "negative definitions" refer to explicitly defining a group by what they are not. For example, *"Non-experts and engineers use visualizations"* contains a counter-example, while *"Non-experts include anyone who is not an engineer"* is a negative definition. Counter-examples can be a helpful tool for learning about the nature of some phenomena (e.g., as observed in [157, 15, 64]), but they must defy the reader's existing expectations in order to be beneficial [420]. In other words, in order to be helpful, a counter-example must (at least) teach the reader that their existing belief is wrong. In this way, specifying that "music lovers" are *not* in the general public may be helpful if the paper reader assumed that "music lovers" would be included, while stating that "experts" are not non-experts or novices may not be helpful if the reader already assumed this was true. Nonetheless, counter-examples that violate a reader's expectations may still be unable to resolve the cognitive dissonance between what a reader previously believed and their new knowledge [420]. In other words, providing

an effective counter-example may teach the reader that they are *wrong* about who belongs in an audience, but does not always help them understand what is *right*.

3.4.1.2 Inclusion and Exclusion Criteria as Implicit Definitions

For papers which reported human-subject studies, a paper-reader could refer to the participant inclusion and exclusion criteria to get an idea of what was meant by the audience (see Table 3.1 for a count of the inclusion and exclusion criteria collected. Inclusion criteria are intended to "specify, in great detail, who qualifies to be included" in a study [222]. Clear inclusion criteria are necessary for replicability and set the scope of who the research question is about [363, 282]; in this way, they too define the audience. Exclusion criteria, on the other hand, describe "features of the potential study participants who meet the inclusion criteria but present with additional characteristics that could interfere with the success of the study" [282]. Together, inclusion and exclusion criteria can also implicitly define the audience by specifying who is included in the audience specified by the research question and who is excluded for the purposes of the study's design.

Of our set of 79 papers, 44 recruited participants which were intended to represent novices, non-experts, the general public, or laypeople. Of the 44 papers which recruited participants, 35 described the inclusion criteria employed. Papers which recruited crowdworkers often reported very specific inclusion criteria such as citing that crowdworkers on Amazon's Mechanical Turk platform were required to have specific task approval metrics (e.g., *"10,000 [human-intelligence tasks] approved"* [218]) and live in particular countries (e.g., *"Canadian or US residents"* [2]). Authors also mentioned sampling participants who were museum visitors ages 6 and older [192] and students enrolled in an *"undergraduate computer science HCI course (taken in person during summer session)"* [169].

On the other hand, most of the papers that we surveyed described no exclusion criteria. Only 5 of 44 papers which recruited participants excluded participants from one of the keyword groups. Of those 5 papers, 2 rejected or re-categorized participants into a different audience based on background knowledge [218, 429], 1 based on color-blindness [425], 1 for poor response quality [385], and 1 on the basis of color-blindness and a non-English native language [226].

Although inclusion and exclusion criteria are supposed to describe what is *required* for an individual to either be included or excluded from a study, it was sometimes ambiguous whether the characteristics of participants described in the papers we surveyed were *required* of participants or if they were disclosed to provide more context to the reader (i.e., were anecdotal). While both can be helpful for placing results in context, it can be difficult to establish what aspects of the participant pool the authors think are relevant or that they intended to control in their study. For instance, among the 44 papers which recruited participants, undergraduate and graduate students were sampled as the *only* participants in 15 papers. The practice of using students as subjects has been hotly debated for decades and can be problematic when students differ from the target population along dimensions which influence the magnitude or direction of the result [112]. However, without a clear description of which dimensions the authors believe matter to the intended audience within the context of the paper, it is difficult to accurately assess how students may differ from it. This, in turn, can make it very difficult to figure out to whom results should apply.

3.4.1.3 Discussion of Implications of Ambiguity

Our analysis suggests that authors may be over-relying on implicit definitions for visualization audiences in place of explicit definitions (e.g., by using examples or counter-examples). This lack of specificity is dangerous because it can result in over-generalizing research outcomes for audiences that were never intended (or evaluated)

in the original research. We can see how this could occur within the context of our case study by examining how the pool of visualization "users" has shifted over time and how implied audiences may have shifted with it. As many authors have pointed out, analysts were once considered the dominant or sole audience for visualizations (e.g., [24, 170]). When authors wrote about "novices," they could have been referring to less experienced analysts; perhaps to statistics- and visualization-savvy people without professional experience in field. Today, this same group of people may be excluded from the audience implied by authors using identical language — as the audience for visualizations has expanded, analysts may not be considered novices at all. Therefore, results about and guidelines for novices 30 years ago might not be applicable to the novices of today. However, it may not be clear to the people who read the work in the future that the audience has shifted at all and that the results do not necessarily generalize, *unless* the authors have defined their audience carefully in the text.

3.4.2 Definitions focused on only one aspect of the individual's identities, interests, or experiences.

In most of the papers that provided explicit or implicit definitions (50/79), authors provided a definition for at least one term that focused on a single aspect of the individual's identities, interests, or experiences. Many of the papers we surveyed focused on an individual's *lack of knowledge* to implicitly or explicitly define novice and non-expert audiences. Most of these papers identified these groups as lacking domain knowledge when defining novices (12/32) or non-experts (6/20). For example, non-experts had *"little technical knowledge in ontological models like [Knowledge Graphs]"* [429] and *"no deep knowledge of flood management"* [86]. Far fewer papers grouped people based on knowledge focused on an individual's knowledge of visualizations (novice: 2/32, non-expert: 2/20, e.g., [215, 388]). Instead, implicit definitions

were based on an individual's job title or relationship to a particular system (novice: 12/32, non-expert: 6/20) such as visualization students [368], crowdworkers [352], and *"users of a distributed file sharing system"* [118].

A few papers that we surveyed defined at least one audience using multiple dimensions (14/79). For instance, South et al. created DebateVis as a tool to explore presidential debate transcripts for "non-experts" who possessed *"limited to no familiarity with the upcoming U.S. Presidential election"* and no *"experience with text analysis or text visualization"* [360]. By defining their intended audience using both domain and visualization knowledge, the authors focus on the unique needs of individuals within that intersection which may have been missed if they had defined the group more generally. For instance, the authors reported that participants had difficulty parsing chord diagrams when evaluating an early prototype [360]. If they had defined "non-experts" more generally (say, only based on domain knowledge), then it is possible the needs of these users may have been masked by users with more experience reading chord diagrams.

Among the 15 papers that provided an explicit definition for at least one of their audiences, 10 defined their audience along a single dimension. The 5 papers which did not defined their audiences using a combination of age and experience using a computer [192], experience with image editing tools and with historical art [289], experience with design and modeling [180], experience and interest in cryptocurrency [431], and experience with presidential elections and text visualization [360].

3.4.2.1 Discussion of Risks of Over-aggregation

Researchers risk ignoring the (potentially unique) needs of different populations them into majority groups when they define audiences based on one aspect of an individual's identities, interests, or experiences. As an example, if an audience was to be categorized based solely on individuals' knowledge of a non-visualization domain

such as Biology, the authors of this paper, who study visualizations, and people who do not need or want to use visualizations would both be grouped together and considered novices despite likely possessing different perspectives on visualization.

The over-aggregation of audiences is likely to disproportionately impact groups that have already been historically marginalized in the visualization literature because they are already minorities or have been minoritized. For example, little visualization literature has focused on the needs and experiences of individuals with disabilities other than blindness and low vision [115]. Without intentionally constructing visualization audiences that consider how disability status interacts with other kinds of knowledge, experiences, interests, or skills, individuals in these groups are aggregated into the majority group.

3.5 Suggestions for Practice & Design

In the previous section, we discussed current practices of defining audiences for visualization work. We demonstrated that in papers which used the terms novice, non-expert, layperson, or general public, explicit definitions of these terms were largely absent. Instead, authors used implicit definitions that may be incomplete or one-dimensional.

In this section, we draw inspiration from research methods and feminist theory to provide three suggestions for defining visualization audiences more completely. This section is not intended to represent all of the actions that researchers could take. Instead, it represents three under-explored areas that mirror our analysis. We encourage readers who connect with the themes that we examine (particularly those on intersectionality and power) consult D'Ignazio and Klein's "Data Feminism" [106] for further discussion and examples of how these themes manifest in data science broadly.

3.5.1 Make the Implicit Explicit – Explicitly define terms; Disclose inclusion and exclusion criteria

Our analysis revealed a lack of clear, explicit definitions for audiences. Instead, authors defined terms by providing examples or counterexamples of people in those groups (e.g., students are considered novices, experts are not). However, explicit definitions are often necessary to ensure that readers understand whom the work and results apply. In this section, we will provide suggestions for how to construct clear definitions of groups and participants.

1. Explicitly define terms used to describe audience(s) by specifying which qualities are required for inclusion.

We propose that papers should clearly define the terms they use for their audience(s). In other contexts, authors are expected to provide definitions when they introduce new terms or utilize a term that has multiple meanings. The terms authors use for their audience(s) should be no different. For instance, in a paper exploring the impact of pictograph arrays on "causal sensemaking" practices of the general public, the authors define the term "causal sensemaking" but do not define "general public" [54]. We suggest introducing a definition of general public such as "people who regularly use static, casual visualizations online or in-print to make decisions."

Definitions should clearly indicate which qualities are considered salient to an individual's inclusion in an audience. We argue that much like the independent and dependent variables ([222]; see [165] for a template of methodology sections in visualization), the intended audience of a study provides critical context to the outcome(s) of the experiment. For instance, many papers in our case study implied that "novices" lacked knowledge about either a particular domain or about visualization, but not both. Defining novices explicitly would make it easier to understand which qualities

novices were required to have and to compare results with other work that focuses on a similar audience. See Table 3.3 for examples of explicit definitions from prior work.

When providing definitions, we encourage authors to think carefully about what dimensions of experience or identity matter. Slater observed that variables such as demographics and geographic location are used most frequently because they are easy to collect and may serve as a starting point for audience differentiation [354]. For example, it may be relatively straightforward to identify an audience for a visualization such as "individuals over the age of 65 who live in Wisconsin." However, "inferred" or psychosocial variables may also provide another option for meaningfully dividing audiences [354]. For instance, Kennedy et al. found that a person's self-efficacy impacts how and why people use visualizations [195]. While it may be possible to divide audiences in this way for experimental purposes (e.g., to employ a pre-test to collect the data and then divide participants based on that), it remains an open question as to how these inferred variables could be employed in real-world visualizations.

It may also be helpful to name people who are *excluded* from an audience definition. Particular groups of people may be excluded because of a design choice made. For example, if a paper focused on designing intricate visualizations for a large screen, mobile phone users may be excluded from the audience as a result of the medium. In this case, we suggest explicitly naming access to particular technologies as necessary criteria (e.g., "our audience is people with access to large high-definition displays, who..."). It also may be advantageous to name who is excluded when defining a particularly broad or diverse audience. For instance, defining the "general public" as "everyone who is not currently employed as a scientist" clearly delineates who is in the audience and why. We suggest that authors who employ this approach note the difference between an explicit "negative" definition (e.g., "The general public is everyone who is not currently employed as a scientist") and providing a counter-example (e.g., "People who are currently employed as scientists are not in the general public").

Notice that the former indicates that the *only* people excluded are people currently employed as scientists, while the latter suggests that scientists are *among* the group of people who are excluded.

2. When papers contain human-subject studies, state inclusion and exclusion criteria.

In papers containing a human-subject study, methods sections should clearly indicate participant inclusion and exclusion criteria. Although this is a practice discussed by many works on research methodology (e.g., [222, 181]), it is still uncommon in the papers we surveyed.

Ideally, there should be a clear connection between the definition of the audience and the inclusion criteria because both define the features relevant to the research question [282, 181]. Further, by definition, exclusion criteria are not just the inverse of inclusion criteria [282] — they describe why a person was excluded who otherwise *would have been included*. Therefore, it is necessary to disclose **both** inclusion and exclusion criteria — the inclusion criteria describe what is required of an individual to be a representative of the audience, while the exclusion criteria describe who is in the audience but not allowed to participate in this study to accomplish study objectives. Regardless if exclusion criteria are applied, authors should state it as to provide clarity for the reader.

There are different ways of disclosing inclusion and exclusion criteria (see examples from existing work in Table 3.4). For example, inclusion criteria can be explicitly stated as such (i.e., the example from [253]) or can be described as requirements (i.e., the examples from [179, 291]). In either case, the descriptions must contain sufficient detail to understand who qualified to be a participant and why [156, 84]. Similarly, exclusion criteria may describe qualities of a person (e.g., *"compromised*

cardiopulmonary function requiring the use of oxygen" [253]) or particular behavior observed once the study has begun (e.g., *"completed a question under three seconds for more than three times"* [226]).

3.5.2 Think Intersectionally – Define audiences multi-dimensionally; Consider how intersections produce unique perspectives

As discussed in Section 3.4.2, the terms that we investigate (e.g., novice, general public) were often (implicitly) defined on the basis of one dimension of a person's identities, knowledge, or experience. In the papers we surveyed, this dimension was commonly the amount of knowledge or experience of an individual. We caution that the practice of defining an audience based on only one dimension may have the unintended consequence of aggregating individuals who exhibit very different behavior or needs. These issues may mimic the well-characterized problems of aggregating non-demographic data, including biased correlation and regression coefficients in statistical analyses (e.g., [77, 31, 11]).

As an alternative, we suggest thinking of visualization audiences through the lens of **intersectionality** — how the intersections of multiple personal dimensions might result in unique experiences with a visualization. The term "Intersectionality" was originally introduced by Kimberlé Crenshaw to describe how thinking of Black women as only Black or women fails to capture their experiences at the intersection of the two identities – as both Black *and* women [60, 94]. Although Crenshaw coined the term in the 21st century, intersectional work has been conducted by predominantly Black, Latina, and Native women in response to systematic injustice stretching back to the 18th century [79, 80, 308]. In recent years, intersectionality as a framework has been applied across many domains including Law, Politics, Psychology, and Social Science Research [60], as well as the closely-related field of Human-Computer Interaction (HCI) where scholars have characterized Critical Race Theory for HCI (including

intersectionality as a central tenet) [266] and applied intersectional lenses to ethnographic analyses [403]. Its roots in gender and race can be thought of as a jumping off point for examining the ways that constellations of an individual's identities may shape their experiences and perspectives [74, 60].

Intersectionality, as a lens, asks researchers to consider how the perspectives of individuals with particular identities or experiences might be unique from the perspectives of individuals who share some, but not all, of the same identities or experiences. In a visualization context, research could apply an "intercategorical method" of intersectionality "to document and compare relationships between social groups along multiple identity category dimensions"[332, 248] and ask questions like:

- How might people who do not read visualizations regularly AND are skeptical of vaccine mandates view this visualization differently than people who do not read visualizations regularly AND are NOT skeptical of vaccine mandates?
- How could we design a visualization to meet the needs of blind users who are also unfamiliar with climate science terminology?
- What does it mean to design a visualization of election results for people with no domain knowledge and years of experience reading visualizations as opposed to people with no domain knowledge and little experience reading visualizations?

Intercategorical methods of intersectionality may be contrasted against "intracategorical methods" which examine the diversity that arises *within* a group [332, 248]. One example of work that engages intracategorically with a visualization audience is Peck et al.'s "Data is Personal" [283]. The authors interviewed 42 adults living in rural Pennsylvania (USA) about their attitudes and perceptions of data visualizations. In their analysis, the authors carefully examine the diversity of opinions which existed within an audience which has been unintentionally excluded from visualization studies in the past. The strength of intracategorical methods is their focus on the nuances that arise within an intersection [22], which might make them particularly

advantageous for projects that build visualizations aimed at a diverse set of readers. By critically engaging with the nuance of the audience, researchers can actively make choices that accommodate for diversity as opposed to building a solution aimed at the (imagined) majority and leaves some readers behind.

Considering intersectional identities could function as a means for uniquely defining terms/groups. Wong et al. provide an example of how this can be done in their work characterizing the specific needs of "Domain Experts" who they defined as *"people who are experts in a specific professional domain where they want to apply visualization and analytics, but who often lack high literacy, training, and motivation in visualization and visual analytics"* [402]. By considering the ways that domain knowledge, visualization literacy and analytical training intersect, the authors can clearly differentiate between the unique needs of this group and other groups such as "Visualization Experts" and "Casual Users." Although a central aim of Wong et al.'s paper was to define the needs of domain experts as a group [402], applying an intersectional lens can also look like defining the intended audience of a visualization through multiple dimensions, but which ones? Past work in technical communication by Albers has argued that authors should consider at least three dimensions when defining an audience: the reader's current knowledge on the topic, the amount of detail the reader wants, and the reader's ability to "comprehend and understand the material" [6]. They also suggest considering "social and cultural factors that affect the reader" if the audience is not expected to be homogeneous [6]. Alternately, research on writing and designing test score reports suggest that authors construct audiences by considering three different areas: Needs (goals and the steps they need to accomplish them), Knowledge (what they already know and will learn), and Attitudes ("feelings or biases" which influence interpretation) [419, 127]. While we suggest that authors define their audiences along more than one dimension, *which* dimensions may

vary according to the project. However, we hope that the dimensions discussed here may serve as jumping-off points for readers of this work.

3.5.3 Beware of Power – Avoid only valuing dominant ways of knowing and doing

In the previous sections, we discussed the importance of clearly identifying one's intended audience and defining that audience with enough depth. For example, we discussed how defining a group of people through multiple dimensions allows for focusing on the specific needs of that group. As Crenshaw writes, *"the process of categorization is itself an exercise of power"* and that the power exerted in categorizing is inherently different from the power to *"cause that categorization to have social and material consequences"* [93]. The ways that researchers categorize the people who read and interact with visualization can have social and material consequences including what information is available to whom. Therefore, in this section we focus on the problem of unintentionally reinforcing dominant systems of power through categorizations to help future researchers use their power responsibly. Our goal is not necessarily to prescribe which audience definitions are "right" or "wrong," but to encourage researchers to think carefully about the ways power moves and is reinforced in visualization work.

People who have power define the knowledge or skills that are deemed valuable and valid [163, 106]. For example, colonial powers have systematically devalued Indigenous ways of knowing and doing in communities across the world in favor of Eurocentric knowledge and practices [343, 399]. Unchecked, this functions as a means to reassert *hierarchical* power structures where those who have the power perpetuate the idea that the things they have are the most important and inherently superior.

Hierarchies are "structure[s] in which groups or people are ranked above or below one another in accordance with authority and status" [217]. While hierarchies

may describe formal relationships among individuals (e.g., many workplaces are organized hierarchically [217, 140]), they can also describe perceived relationships between groups (e.g., racial hierarchies in the United States [359, 61, 375]). While some argue that hierarchies can be helpful organizing principles (e.g., [140]), others emphasize that hierarchies can be harmful because the groups who are not at the top may be seen as the inferior version of the ones who are (rather than as distinct entities) [188]. This, in turn, can result in outcomes including inequality and epistemic injustice (in which minoritized people are not believed, even when speaking from direct experience) [217]. Hierarchies may also be dangerous because they can be self-reinforcing — individuals at the top of hierarchies tend to block individuals who do not conform to behavioral norms from "rising up" [361].

When audiences are defined using only one dimension (as we observed during our analysis), it it may facilitate the creation of a hierarchy between audiences. One sign that a hierarchy may be present is when an audience is defined based on a perceived "lack" of something that another audience has which offers them status or power. For example, defining a group based on a lack of knowledge directly establishes the "lacking" group as having *less* than the group they are being compared to. Further, definitions based *only* on a perceived lack may indicate that a "deficit model" is being applied. In deficit models, people are *"**primarily (or even solely)** [conceptualized] in terms of their perceived deficiencies, dysfunctions, problems, needs, and limitations"* (emphasis added) [108]. Although the papers we surveyed did not specifically use the language of deficits, we argue that discussing the lack of particular knowledge and comparison against experts who *do* possess that knowledge and do not warrant further study or technological intervention satisfies the definition of a deficit posed in past work (i.e., the belief that a person lacks X and ought to have X) [108]. Deficit models have been critiqued for decades in fields like education, disability studies, and philosophy on the bases that they can be de-humanizing, disproportionately

punish minoritized people, and position one group as the "lesser" form of another [108, 168, 188].

One opportunity for authors to value non-dominant ways of knowing and doing (and thus subvert hierarchical power structures) is to be curious about the perspectives of audiences who think about or use visualizations differently from researchers. For example, visualization researchers may believe that visualization is good and worthwhile, so it may be fruitful to consider the perspectives of individuals who do *not* use or do not want to use visualizations, despite the pressures from an increasingly data-focused world [237, 285] to do so. Like the potential for examining who does not engage with particular media [154], considering the non-users of visualizations may be illuminating. One instance of engaging with people who use visualizations differently is present in recent work on visual communication practices in rural Bangladesh [367]. This project offered a look at ways of encoding and communicating information which do not conform to Euro-centric "modern" visualization techniques and values [367]. However, the visualization research community should proceed with caution when looking outside of "traditional" visualization audiences and values in that they do not exoticize and other those who are "out there" [370]. Instead, visualization researchers must actively recognize that they are "in here" – that is, that their knowledge and perspectives are actively shaped by their own experiences and perspectives [163, 370].

We also suggest thinking critically about skills or attributes that might be relevant to an audience member's success and experience with a visualization, but correlate with unequal structures of power (such as wealth and attainment of formal education [125]). Instead, there may be opportunities to anchor analyses to dimensions with weaker ties to power. For example, literature in education and psychology has explored how "grit" (i.e., "perseverance and passion for long-term goals" [113]) impacts success and achievement (e.g., in [114, 92, 113]). Although the evidence is mixed on

the effects of grit as a composite metric, a recent meta-analysis suggested that perseverance component may be positively correlated with academic performance [92]. Further, grit is theorized to differ little between demographic groups [113], though more empirical work needs to be done to clarify the mixed evidence on the relationship between grit scores and specific demographic variables (e.g., age [119], gender [8], ethnicity [67]) [92]. Within a visualization context, one might therefore ask: What strategies do audiences with high grit/perseverance, but little formal visualization training, utilize to accomplish tasks?

Valuing the perspectives of people who use visualizations differently and including skills or attributes which are less tied to institutional power could enable researchers to build visualizations that leverage the knowledge and skills that audience members already have. Connecting visualization practices to the experiences and interests of audiences has been discussed in visualization work for years (e.g., [152, 170, 341, 260, 100]) and could take the shape of alternative visual metaphors (e.g., [260]) or entirely different ways to encode or represent data (e.g., [341]). In Science and Technical Communication, reframing communication messages in ways that connect with the audience's specific values and knowledge has been observed as an effective method to connect with people who were entirely excluded from the conversation before (e.g., [262, 29, 239]). For example, Nisbet and Scheufele observed that framing science as "anti-religion" meant that some religious communities felt that climate change conversations were irrelevant to their lives, but re-emphasizing the moral and ethical dimensions of climate change helped scientists such as E. O. Wilson to convince "religious leaders that environmental sustainability is directly applicable to questions of faith" [262].

But does creating visualizations that utilize the unique experiences, attitudes, and perceptions of audiences mean that visualizations which work for "all" cannot or should not be made? We argue that the body of existing work suggests that,

no, "one-sized-fits-all" solutions are not likely to work for "all." For example, there is plenty of evidence supporting that accommodating for the needs of people with extensive visualization knowledge can make tools difficult to use for people with less experience and the inverse – that choices made to fit people without experience can make it harder for those with experience (e.g., [152, 378]). Further, an individual's perceptions are shaped by their existing experiences, knowledge, perception, and values [163, 161]. Therefore, it is necessarily impossible to create a visualization that is uniformly interpreted by all.

3.5.4 Summary

In this work, we proposed four actions that researchers can take to define audiences more inclusively and completely:

1. Treat words for audiences like new or contested terms in papers – define them with precision. Indicate which dimensions are relevant to inclusion in the audience.
2. If participants are recruited, clearly state the inclusion and exclusion criteria used.
3. Think critically about how intersecting dimensions of experience might produce unique experiences; define audiences using multiple dimensions to center those unique experiences.
4. Think broadly about which of skills, knowledge, and experiences might be relevant; look outside of dominant ways of knowing or doing.

3.6 Conclusion

In this paper, we argue that defining the intended audience of a visualization is critically important and that implementing changes will benefit readers and the field as a whole. Through our case study of visualization research papers that used

the terms “novice,” “non-expert,” “general public,” and “layperson,” we discussed the unintended consequences of two problems: explicit definitions are largely absent and the implicit definitions provided are overly simplistic. We proposed that authors ought to improve current practices by: making the implicit explicit and carefully defining terms directly, thinking of audiences intersectionally, and keeping dominant power structures in mind (as to not reinscribe them). Our intention with this work was to identify a potentially dangerous issue in current research practices and provide suggestions for how the visualization research community might avoid those practices in the future. Our suggestions are not intended to be exhaustive, but instead meant to spark a conversation among the visualization research community about how to talk about the audiences of visualizations and the assumptions made about the people inside of them.

Table 3.1. Few of the 79 papers we surveyed provided an explicit definition of a term. Instead, definitions were often provided or audiences were defined implicitly using examples (i.e., a sub-group or description) or counter-examples (i.e., by contrasting). Papers which recruited participants matching one of our surveyed audiences (n=44) provided differing amounts of information on why participants were included or excluded. Counts reported per paper, represent explicit and implicit definitions within 1 paragraph of the term being used, and are not exclusive.

Code	**n**	**Example**	**Source**
Explicit Definition	15/79	*"None of the participants reported expertise in political debates or text visualization. We chose to use this population for our study because they represented the* ***non-expert****, general audience we want to use DebateVis."*	[360]
At least 1 Example	52/79	*"Since* ***novice users*** *of visualization systems lack the knowledge and expertise in data visualization, ..."*	[134]
At least 1 Counter-example	53/79	*"While pop music is intended to be friendly to the* ***general public*** *... [classical music is] usually only understood by music lovers, who have received extensive training in music theory and history.*	[65]
Disclosed Inclusion Information	35/44	*"We conducted our experiment on Amazon Mechanical Turk with the following recruiting specifications ... 10,000 # HITs approved, 99 HIT approval rate (%).*	[218]
Disclosed Exclusion Information	5/44	*"We exclude participants knowledgeable in KG or graph visualization to avoid interference from their prior knowledge."*	[429]

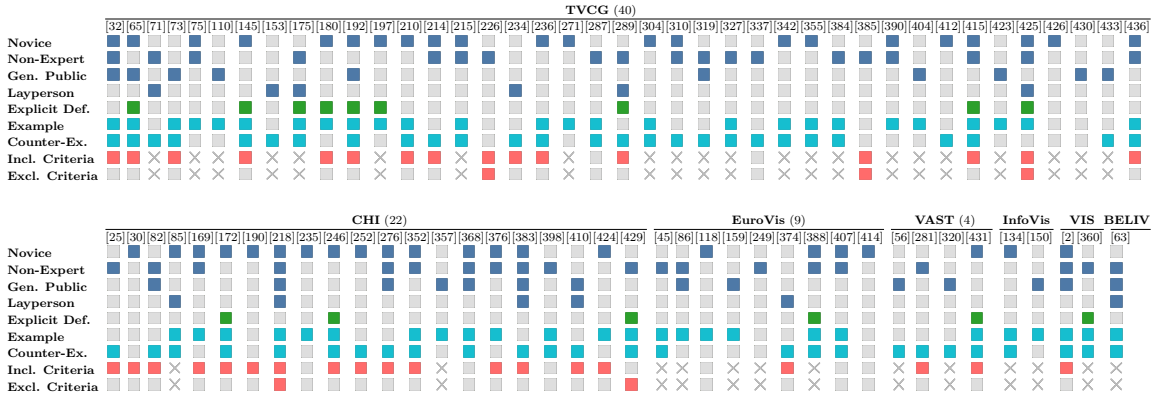


Table 3.2. An overview of the 79 articles we reviewed. The filled boxes indicate that a specific keyword appeared in the paper (■), an explicit definition was provided for at least one keyword (■), an implicit definition was provided via an example or counter-example (■), and inclusion or exclusion criteria were provided (■). We use × to indicate that no participants were recruited to represent any of the keyword audiences we collected.

Table 3.3. Authors should define their audiences carefully by explicitly describing the dimensions they consider. This table contains examples from existing work which provide explicit definitions that describe what constitutes membership, though most are one-dimensional.

Term	**Explicit Definition**	**Source**
Target Users	*"Our **target users** are people who want to understand their personal data with an aesthetically pleasing display, rather than to perform focused tasks in data analysis"*	[280]
General Public	*"The target audience of the science center is the **general public**, ages from six years and up, with no specific prior experience in computer usage."*	[192]
Non-expert	*"None of the participants reported expertise in political debates or text visualization. We chose to use this population for our study because they represented the **non-expert**, general audience we want to use DebateVis."*	[360]
Novice	*"All of our participants were **novice** Vega users (i.e., were unfamiliar with Vega)"*	[172]
Novice	*"There is a need for a novice environment in which first-time (**novice**) users can get acquainted with the basic technology"*	[121]
Novice	*"... we are interested in '**novice users**' that can be defined as users who have seen a particular type of visualization for the first time."*	[225]

Table 3.4. In papers reporting on human-subject studies, authors should carefully disclose inclusion and exclusion criteria. This table contains examples from existing work which do this well.

Inclusion Criteria	**Source**
"Inclusion criteria included community dwelling, 50 years and older, able to read and write in English (no translation services were provided), and able to ambulate without use of assistive devices."	[253]
"Participants were required to be at least 18 years old; have two Instagram accounts, one regarded as their 'real' account (Rinsta) and another one regarded as their 'fake' account (Finsta); and be active on both accounts, making at least one post on each account within the last six months."	[179]
"We had three specific inclusion criteria. First, at least one child had to be experiencing poor sleep, as assessed by a parent-completed Children's Sleep-Habits Questionnaire (CSHQ) ... Second, the child experiencing poor sleep needed to be between the age of 7 and 12 years old ... Finally, all participating family member were required to live in the same home."	[291]

Exclusion Criteria	**Source**
"Exclusion criteria included cognitive impairment and compromised cardiopulmonary function requiring the use of oxygen."	[253]
"We exclude participants knowledgeable in [Knowledge Graphs] or graph visualization to avoid interference from their prior knowledge."	[429]
"We excluded people who have prior knowledge about parallel coordinates (n = 49). To remove random clickers, we filtered out participants who completed a question under three seconds for more than three times."	[226]
"No participants were excluded or withdrawn following randomisation or lost to follow-up."	[78]

CHAPTER 4

DESIGNING WITH PICTOGRAPHS: HOW DOES DESIGN IMPACT CASUAL SENSEMAKING?

In Chapter 3, I discussed how improving how researchers write and think about the audiences of visualizations could result in visualizations that better fit the needs of those audiences. However, improving the design of the visualizations themselves and the methods used to decide which visualizations are understandable could also improve the visualizations which are created. To fill both gaps, I turned my attention to one of the most common types of visualizations encountered by the public – infographics – and used a novel education-inspired method for forming comprehension questions to evaluate how making quantities more explicit through the use of pictograph arrays impacts casual sensemaking. This chapter is a combination of two previously published works [53, 54].

4.1 Introduction

People engage with visualizations to make everyday inferences and decisions. A viewer might consult a hurricane risk map to decide whether to evacuate or inspect a graph of COVID-19 health outcomes for their age group to decide whether to obtain a vaccine. Consider the case of a newspaper reader who encounters a map of COVID-19 positivity rates. Though this static visualization might only contain geographic areas that are categorically coded into risk rates depicted by colors, that reader can still identify trends, compare their county to others, and make choices about whether or not to travel.

Typically in data visualization, data-driven decision making processes are referred to as **sensemaking**. The most commonly used model of sensemaking, proposed by Pirolli and Card, contains both a foraging loop where data are sought, searched, filtered, and read as well as a sensemaking loop where those data are used to iteratively construct a mental model [292]. This process typically involves time-consuming analysis and re-analysis, with the intention of finding analytical "ah-ha" moments. It is also often conducted by expert analysts who have access to the original data and the resources to independently analyze it [292].

In contrast, the everyday decisions about hurricane paths or COVID-19 might be considered *casual* sensemaking, where the general public understands, reflects, and makes decisions based on information provided by casual visualizations [297] without deep and time-consuming research using raw data. It is therefore more similar to a definition of sensemaking from business and management in which the viewer uses a conceptual model of the world informed by available information "to comprehend, understand, explain, attribute, extrapolate, and predict" [12, 362].

One kind of data visualization often encountered by the general public is the infographic. Although infographics have been used for, by some estimates, as long as bar and pie charts [268], there is no single definition. In the journalism world, the word "infographic" has been historically used to refer to pieces that combine graphical elements with text to convey information [58, 57, 243]. Although infographics in this context might contain data visualizations, they may also contain maps, diagrams, or illustrations combined with text or annotations. Overall, the purpose of an infographic is to present important information in a way that is sensitive to common barriers encountered by the public such as limited time and information overload [268]. Because infographics are used to communicate critical information to the public, they could have a strong impact on the way that everyday people understand science and current events. After encountering an infographic, viewers may not seek

out and further analyze the data, even if the data is publicly available, because of a lack of the motivation, time, or skill (for example). In these cases, the viewer must come to conclusions based on the information that they are shown. For example, if a person encounters an infographic about COVID-19 cases in their state while reading, we cannot assume that the person will then seek out the associated data dashboard to do a more in-depth analysis – yet, they might use the graphic as it is presented to make choices about how to safely navigate their community. Therefore, it is important to understand how to design visualizations in a way which communicates information effectively to everyone who encounters it – this is the chief interest of the paper. In other words, if we know that people are using data visualizations to engage in casual sensemaking, then we should be thinking critically about how our designs afford those activities.

To measure the ability of visualizations to afford casual sensemaking or, more traditionally, understanding, we turn to the field of Education for inspiration and operationalize Bloom's taxonomy, following the taxonomy proposed in [53]. Bloom's taxonomy describes 6 aspects of the learning process: knowledge, comprehension, application, analysis, synthesis, and evaluation [33]. Though originally intended as a strict hierarchy, some critiques suggest that while some dependency may exist between the levels, it is not a strict hierarchy [53]. Therefore, we view it instead as 6 complementary skills of differing complexity. We chose to use Bloom's taxonomy because it is commonly used to evaluate learning processes [212] and maps well to the activities necessary for casual sensemaking.

In existing research, infographics show positive empirical findings in terms of memory, engagement, and assessment of risk [166, 36, 138] — particularly when they contain pictographs (simple, iconic pictures that represent a word or topic). Pictographs may also be helpful because they make topics and quantities more explicit than when represented abstractly (as with concrete scales [72]). Yet, there is little exploration

of how pictographs affect and afford different kinds of understanding. Further, there is little work on how the choice to use pictographs affects the personal experience of using infographics across factors such as visual appeal, clutter, and ease of envisioning the topic.

We first demonstrated the applicability of our proposed framework with three case studies comparing two alternative designs of real-world visualizations, selected because they represent several common chart types and were identified by the data visualization or data journalism communities as being confusing or misleading (Section 4.4). Sixty participants from Amazon's Mechanical Turk answered a set of six open-ended comprehension questions (one targeting each of the six levels of Bloom's taxonomy) about three visualizations. Each question was designed to target a specific level of Bloom's taxonomy and was presented in order, beginning with Knowledge and ending with Evaluation. We showed that different versions of each chart afford different interpretations and conclusions across particular levels of the taxonomy, allowing a systematic evaluation of similarities and differences between designs.

Then, we conducted a series of experiments on Amazon's Mechanical Turk where we compared variations of charts that are informationally equivalent but differed in the presence (or absence) of pictographs. In Experiment 1, we examined how well infographics with pictograph arrays afford sensemaking compared to more traditional bar, pie, and donut charts. We created 1 open-response question to target each of the 6 levels of Bloom's taxonomy. We found that participants generated similar quality insights across the two chart versions evaluated. Our data suggest that designers can use pictographs in place of more traditional geometric shapes without impacting user understanding.

While the first experiment helped us understand aspects of casual sensemaking related to learning and understanding, we conducted a second experiment to interrogate the experiential aspects of casual sensemaking. In the second experiment, participants

compared the same charts from Experiment 1 on 5 metrics identified by the visualization community as important to effective infographics: visual appeal, quickness of understanding, ease of envisioning, clutter, and perceived importance. Additionally, we asked participants to explain their reasoning. We found that some participants thought the charts with pictographs helped them envision the topic better, while others thought they required more time to understand and were unnecessarily cluttered. Additionally, we found that charts that were rated as more visually appealing, less cluttered, easier to envision, and faster to understand were also thought to make their topics seem more important.

The main contributions of this chapter are: (1) empirical results that suggest that infographics with pictograph arrays are just as good as more traditional, geometric part-to-whole charts at helping people make sense of data; (2) empirical insights that show some participants view charts containing pictograph arrays as easier to envision, while others view them as unnecessarily cluttered and slower to understand; and (3) qualitative results suggesting that perceptions of visual appeal may be impacted by ease of understanding, making the topic seem real may help readers envision the topic more easily, and feelings of urgency and importance may be influenced by real-world connections.

4.2 Background

4.2.1 Infographics and relationships with data visualization

For many people, the word "infographic" is a loaded term. The word infographic initially gained traction within the news world to refer to any sort of graphic that was used to display information [58]. As different fields have brought infographics into their practices, the word has changed meaning. A definition used within the communication and design communities similarly requires a combination of text and graphics but stresses the purpose of the chart — they must be created to assist

in communication [295]. Alternately, definitions may be very strict, incorporating requirements about the way that the graphic is laid out [394]. In some sources, all data visualizations are considered infographics. This is incorrect for a variety of reasons, but, as it pertains to this chapter, infographics are largely intended for effective communication while data visualizations can be used not just to communicate, but for exploration, discovery, and analysis, among other tasks.

Outside of the visualization community, infographics are known for their ability to drive engagement. Many news organizations have teams devoted to developing infographics, where the quality of a graphic is judged by its ability to engage audiences [104]. Within the medical communication community, a study found that when wound care instructions were accompanied by images, patients read the instructions more often, could recall the instructions more accurately, and followed the instructions more often [177]. Another found that infographics that were designed with and for communities of non-native English speakers with low medical literacy were viewed as more likely to influence behavioral change [16].

Within the sphere of medical communication, visual aids are well understood as effective tools to help people understand risk. In that domain, visual aids such as infographics and pictograph arrays are useful for helping individuals with medium to high graph literacy and low numeracy [138]. These infographics have also been effective at encouraging behavioral change and communicating medical information across language barriers [138].

4.2.2 InfoVis research on infographics

Critics of infographics have argued that embellishments should be avoided, lest they cause distraction from understanding the data presented [379]. However, existing work that compared highly illustrated infographics to informationally equivalent, plain charts found no difference in overall understanding of the data [19]. Further,

even when the illustration strictly appeared in place of bars of a bar chart, there was no effect on speed or accuracy [353].

Past research has also revealed memorability to be a strength of infographics, showing that when compared to unembellished charts, participants are able to recall the data, trend, value message, and topic of an infographic more accurately over a long period of time [19]. Further, over a shorter-term, participants were able to remember that they had seen the infographic, could more accurately recall its message, and, when attention was divided, could more accurately recall specific points of data, especially when those infographics were colorful and contained pictographs [36, 35, 19, 166].

Within the data visualization community, infographics have also been shown to be particularly useful for engaging audiences, though an infographic's effectiveness may be linked to its visual appeal [240]. For example, one study found that when news articles were accompanied with visualizations, people who found the infographic visually appealing were more likely to read the accompanying article [100]. Investigations into what makes infographics visually appealing found that the most appealing graphics were those that were very colorful but not visually complex [167]. Some speculate that because infographics are designed to be easy to understand, they are more compelling to share, leading them to reach a wider audience [356].

However, existing work has sometimes also found infographics to be less effective than other techniques. When comparing infographics to data comics, one group found that participants who were shown data comics were able to answer questions more accurately and remember those facts longer than those shown infographics [394]. They also found that, when compared to data comics, infographics were preferred for exploration — a quality seen as effective for building trust that the whole story is being shown [270].

There has also been recent work on infographics-creation tools, particularly aimed at users with limited programming experience. Some of these tools aim to bridge the gap between data analysis and vector graphic tools by allowing the user to assign specific aspects of the illustration to data dimensions via lazy data binding (e.g., [233, 408]). Others are designed to allow users to define custom layouts [315], quickly style unembellished charts [392], and use existing images to copy styles [427] or extend timelines [71]. There are also several tools intended to create infographics specifically for use with personal data, including an end-to-end data collection and visualization generation tool [200] and a tool to modify photographs to fit line charts [279].

Though interest in infographics within the visualization community is high as evidenced by the bounty of tools available for their creation, the pool of existing research on the design and evaluation of infographics has been limited. However, within the existing research, pictographs appear several times as potentially impactful elements of design.

4.2.3 InfoVis research on pictographs

Though empirical work on pictographs is fairly limited, the idea of using pictographs for data communication is not new. For example, in the mid 1920s, Otto Neurath, Marie Neurath, and Gerd Arntz created the ISOTYPE system, which used a set of custom pictographs to create data visualizations about social and economic topics [380]. Their designs were guided by 2 simple principles: (1) use pictographs to represent objects and (2) use repetition, not size, to represent quantities, which leads to designs with rows or arrays of pictographs.

Empirical work on bar charts in this style found that when pictographs were used to encode data, they had a positive effect on short-term memorability and engagement [166]. Further, pictographs were found to have no impact on performance in terms of speed or accuracy unless they were used decoratively [166]. More generally, info-

graphics with pictographs have been shown to be more instantly recognizable and to result in more accurate descriptions when recalled from memory [35]. Within medical contexts, pictograph arrays have also been shown to be effective for conveying risk for people with differing levels of graph literacy [139, 137]. Other studies have found pictograph arrays to be less liked and trusted when used to display breast cancer risk when compared to more traditional methods [46].

However, not all pictographs are equally effective. Past work has observed that iconicity can impact both the recall of information and the perception of risk [435]. Additionally, the pictographs might be interpreted in ways the designer did not anticipate (such as when representing categories of items) [16].

4.2.4 Realistic evaluation of sensemaking

In this work, we wanted to make sure that we realistically evaluated the ways that pictograph arrays impacted readers' sensemaking processes. Therefore, we employed a novel technique for evaluating visualizations based on Bloom's taxonomy of educational objectives[33].

As discussed in Chapter 2, there are a variety of metrics used to assess particular understanding within visualization research. One of the most popular methods for measuring understanding is by measuring speed and accuracy. Although popular, this method has been critiqued because visualizations are not just intended to help people understand information easier or faster, but also to better afford aspects of understanding such as applying or evaluating the visualized information. Therefore, to truly evaluate the efficacy of a visualization, we need to ask more difficult, open-ended questions [264] which can capture different information than closed-ended methods [265, 314]. Existing methods of forming open-ended questions (e.g., asking participants to describe one thing they found interesting or surprising [387]) lack a system for comprehensively evaluating different levels of understanding.

To fill this gap, we endeavored to create a method which combines open-ended questioning techniques, which have been shown to capture data that would otherwise have been lost [314], based on Bloom's six-level taxonomy of educational objectives. In 2015, Mahyar et al proposed the use of Bloom's taxonomy in information visualization to measure the depth of engagement and knowledge obtained by viewers [242]. However, before our work, there was no concrete description of how the taxonomy could be applied, nor a demonstration of its efficacy. In Section 4.3, we describe Bloom's taxonomy in detail and provide specific examples of how this framework can be used to assess visualizations.

Existing Taxonomies in Visualization

Existing taxonomies within the visualization community can be roughly divided into three categories: taxonomies of visualizations, objectives, and actions. Where taxonomies of visualizations classify types of data visualizations (e.g., [47, 288]), taxonomies of objectives categorize the questions that a user wants the answers to in order to solve a problem (e.g., [44, 256, 432, 382, 69, 316, 299]). Finally, taxonomies of actions classify concrete actions done in pursuit of an objective (e.g., [335, 3, 196, 13, 316]).

Our proposed framework is a taxonomy of objectives and is most similar to the taxonomy of analytic tasks described by Amar and Stasko [10]. In both taxonomies, the categories of objectives correspond to types of tasks which need to be completed in decision-making practices. However, our proposed taxonomy is more comprehensive, expanding the scope of objectives to include both more simplistic tasks and the final conclusions drawn. Some existing work also focuses on participant learning outcomes, but they cannot be used to form evaluation questions – only to evaluate the responses. This work quantifies the completeness and correctness of a response (e.g., in [19, 411])

or the complexity of a reported insight (e.g., in [387]) in order to evaluate open-ended responses in a post-hoc way.

In summary, we build on existing work in 2 important ways. First, prior work which compares charts with pictographs to those without either uses real charts which are not informationally equivalent (e.g., [36]) or informationally equivalent charts which are not real (e.g., [166]). If we want to know what impact design decisions might have on the audience, then it is important to use charts which are both as realistic as possible and informationally equivalent. Finally, though identifying values in a chart is a critical skill that forms a basis for other more complicated tasks, it does not realistically capture the kinds of understanding-based tasks which readers partake in. Phrased differently, though we have some indication that pictographs do not impact accuracy tasks, we do not yet know what impact pictographs have on understanding beyond these tasks.

4.3 Bloom's Taxonomy

In 1948, a group of psychologists and teachers had an informal meeting at the American Psychological Association Convention, brought together by a shared problem. They wanted their students to understand their lessons, but they each had different ideas about what they meant by terms like "understand" or "internalize," leaving them with no way to compare those objectives [33]. Motivated by a desire to find a common, rigorous way to categorize student understanding, the group came up with the Taxonomy of Educational Objectives, described in a handbook published in 1956. The system is now commonly referred to as Bloom's taxonomy, named after Benjamin Bloom, the educational psychologist who edited the handbook.

In the years since its creation, Bloom's taxonomy has been used broadly in Education [212]. Educators can create activities and assessments that target specific levels of the taxonomy [17] or use it to evaluate the assessments they already use to

better understand the objectives being tested and potentially inspire a broadening of those objectives [191]. In addition, Bloom's taxonomy has been applied widely in other fields such as Biology, Business, and Health Sciences, to create and evaluate exam questions [95], teach critical thinking skills [259], guide course creation on substance use disorders [257], and visualize the breadth of goals present while developing curriculum on visual literacy [120].

Drawing inspiration from the taxonomies of Biology, Bloom's taxonomy is intended to be hierarchical, meaning that learning at the higher levels is dependent on demonstrating mastery of lower levels. However, this is not necessarily a realistic representation of the learning process, which is not linear [28]. Therefore, although we will now describe the levels of Bloom's taxonomy in the order originally presented, we reject the assumption that a strict hierarchy exists between levels. Instead, we view them as complementary skills. For a quick reference to the 6 levels and how they can be used for evaluating visualization, refer to Table 4.1.

Level	**Description**	**Example Tasks**
Knowledge	Recall basic facts and definitions.	• Retrieve points • Locate value • Identify axis labels
Comprehension	Understand the information in context.	• Summarize main message/take away • Describe content of visualization • Explain the topic of the visualization
Application	Apply knowledge to a new problem or represent it differently.	• Use a percentage and total population to calculate a number • Calculate the difference between two points • Translate the data in a chart to a table
Analysis	Break down a concept into parts and understand their relationship.	• Describe a trend • Describe the relationship between two variables • Identify what data was used to come to a conclusion
Synthesis	Use knowledge to create something new.	• Predict a future value • Generate a new visual representation
Evaluation	Judge the value of information, backed by evidence.	• Justify a conclusion based on data • Judge which design is more appropriate

Table 4.1. This table presents the 6 levels present in the original Bloom's taxonomy [33], a short description of each, and example tasks specific to the visualization community.

4.3.1 Knowledge

The Knowledge level historically describes the simplest learning objective demonstrable by the learner. Associated with verbs such as retrieve, identify, and recall, at this level a learner is able to accurately **recall or recognize factual information** that they have learned. Note that this does not require the learner to understand any contextual information or the reason behind facts [33].

This level, in its original context, describes a very simple learning task – reporting back something already seen. Therefore, for visualization, we translate this level into tasks which ask participants to locate and report specific pieces of information. In both the original and translation, no transformation is applied to the information by the viewer and no understanding of context is required. In the experiment described later in this paper, we used this question to ask participants to locate specific data points, though other appropriate tasks might include copying text from an annotation layer or identifying what manipulations an interactive visualization offers.

4.3.2 Comprehension

At the Comprehension level, learners begin to **understand the underlying information as a whole** [33]. Traditionally, questions at this level ask learners to write summaries or identify key ideas [98] and use verbs such as describe, explain, and summarize.

When applied to visualizations, we can translate this level's focus on understanding information as a whole into tasks that ask about features present in a dataset as a whole. For example, we suggest asking for a general summary of the data or the key take-away messages. Questions formed at this level can be more open-ended than those at the Knowledge level and, through this, can reveal the different conclusions afforded by the visualization being evaluated.

4.3.3 Application

At the Application level, learners **apply their knowledge** to solve an unfamiliar problem [33]. This level is commonly associated with verbs such as translate, solve, calculate, and apply.

For visualization, this level could be translated to tasks where the participant solves a problem using the data from the visualization, such as identifying a proportion and using it in a simple computation. This approach may be most appropriate when the response modality for the participant is restricted to text. In our experiment, we asked participants to determine the difference between two data values. For situations where there is not an obvious problem to be solved, this level can also be interpreted as translating knowledge from one form to another. For example, one could ask participants to translate the data displayed in a visualization into another visual style, though this approach requires a more complicated response modality.

4.3.4 Analysis

At the Analysis level, the learner is expected to **break down a topic into parts and understand the relationship between each part** [33]. This level is therefore associated with verbs such as classify, break-down, associate, and relate.

Questions targeting this level could ask about trends, as this requires the participant to identify relevant components and then compare their spatial relationship to each other. We rely on this type of question in the experiment presented in this paper. Alternately, questions could also ask participants to identify which pieces of evidence were used to support a specific conclusion drawn from the data. This translation views the conclusion as the "topic" to be broken down and the data points as the components. Acquiring a conclusion from the data points requires understanding the relationship of the points to each other.

4.3.5 Synthesis

Where the focus of Analysis level was to test the learner's ability to decompose a topic into its requisite parts, the Synthesis level focuses on the learner's ability to **put ideas together to create something new** [33]. Among the hierarchy, this is the first level which requires a certain amount of creativity from the learner and is associated with the verbs such as create, invent, predict, and devise [33, 98].

When applied to visualizations, questions targeting this level could ask participants to make predictions about what values will come next in a sequence. In this translation, the participant takes existing trends and values extrapolates on them to form a prediction. Alternately, in the case of interactive visualizations, participants could instead use interactive features to find a view of the data which reveals something new. Because we used static data visualizations in our experiment, we used the prior approach and asked participants to make a prediction.

4.3.6 Evaluation

Finally, we arrive at the sixth and final level of Bloom's taxonomy – Evaluation. This level evaluates a learner's ability to **judge the value of a topic or idea** based on criteria that is either provided or self-derived [33]. Therefore, rather than judgements, tasks at this level may look more like arguments or proofs, as evidenced by the verbs often associated with this level which include judge, justify, argue, and recommend [98].

The most straightforward translation of this level to an evaluation task might be to ask participants to judge the quality of the visualization itself by some provided criteria (e.g., reliability). This method might be appropriate if the experiment aims to evaluate the participant's understanding of the visual encoding of the visualization. Alternately, if the experiment aims to evaluate the participant's understanding of the underlying data, we suggest instead either asking participants to come to a conclusion

and provide a data-based justification for that conclusion or to provide a conclusion and ask participants to only provide the justification. With this tactic, participants judge the value of data features when deciding which are appropriate to justify the conclusion drawn. We use this approach in our experiments.

4.4 Preliminary Demonstration of Evaluation Method

To demonstrate the kind of results an experiment could obtain with our evaluation method, we evaluated three pairs of visualizations as case studies.

4.4.1 Stimuli

For our stimuli, we selected 3 static, real world data visualizations varying in complexity and design that were identified by the internet community as being confusing or misleading [81, 405, 229]. We also collected three corresponding redesigns of the confusing visualization created either by one of the authors who is an expert in visualization design or the original visualization designer at the Economist [229], where the goal of the redesign was to clarify the message conveyed by the visualization.

Markets The first stimulus (as shown on the left in Figure 4.1) was highlighted in an online article from The Economist titled "Mistakes, we've drawn a few" [229]. The original chart depicts the US trade deficit with China and the number of people in the US employed in manufacturing between 1995 and 2016 as two line plots. The cardinal sin committed by this chart is its double y-axes – one positive and one negative, color-coded to match its associated line. In the article, the author presents a redesigned version of the chart that separates the two plotted quantities into separate, side-by-side charts and encodes the trade deficit using bars connected to the 0 baseline to emphasize the directionality of the plot.

Immigration The second stimulus was produced by The Globe and Mail to discuss differences in immigrant populations in Thunder Bay, Ontario, and Canada (as shown

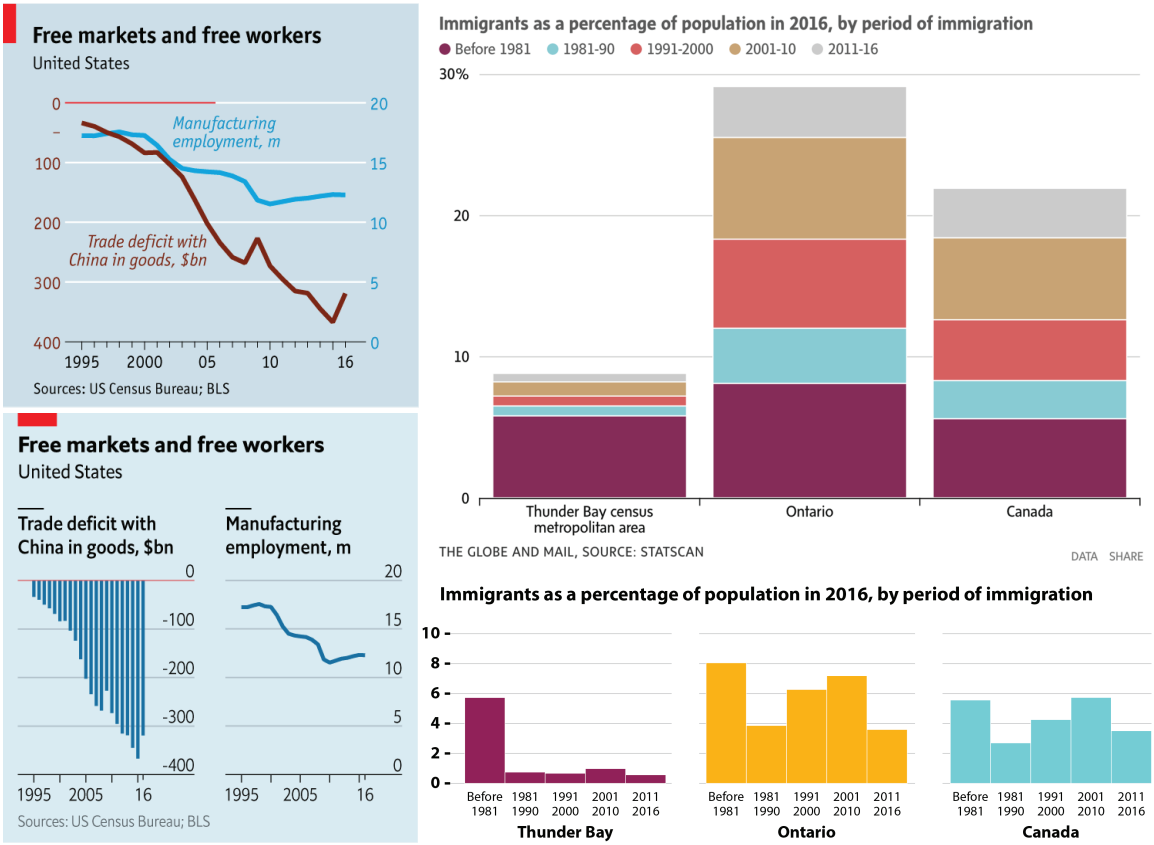


Figure 4.1. The Markets (left) and Immigration (right) charts used in our experiment. The original versions are on top.

on the right in Figure 4.1). The original version of the chart features three stacked bars, one per location. Each stack is divided into five sections, corresponding to the decade when people immigrated to Canada. At a first glance, this chart might be confusing because the sections correspond to time – a variable conventionally reserved for the x-axis.

Unlike the first chart, the structural problem with this chart is harder to identify. If the purpose of the chart is to highlight the differences in distribution or visualize the aspect of time, the original chart does a poor job. However, if the primary purpose of this chart is to highlight differences between the total percentage of immigrants in each area, the original chart might accomplish that goal. For the purpose of this paper, we assumed that the designer of this chart intended to show the differences between the distributions of immigrant year of arrival and emphasized this message in our redesigned chart by unstacking the bars and recoloring the chart such that each location uses a different color (instead of each decade).

COVID-19 The final chart we selected was a now retracted chart created by the Georgia Health Department[144] depicting the number of COVID-19 cases reported over two weeks in the five counties with the largest number of cases (see Figure 4.2). The original version of this chart at first looks uncomplicated. It's just a simple bar chart, but a close reading of the labels along the x-axis reveals the problem – the dates are not in chronological order. Instead, both the days and the counties within each day are sorted by severity.

As made obvious by its retraction, if this chart was intended to show how the number of cases of COVID-19 have changed over time, this chart is plainly misleading. However, like the chart on Canadian Immigration, this chart would be appropriate for answering a different question. Namely: which days saw the largest or smallest number of COVID-19 cases reported, and which counties had the highest or lowest number of cases within each of those days? Unfortunately, this message is not sup-

ported by the annotation layer of the chart, which uses the phrase "the number of cases over time." For comparison, one of the authors created a second version of this chart with its dates in chronological order and with the bars divided by county to make it easier to compare how the number of cases changed over time within each region.

4.4.2 Methods

We conducted our experiment on Amazon's Mechanical Turk and recruited a total of 60 Workers ($\text{Mean}_{\text{age}} = 36.4$, $\text{SD}_{\text{age}} = 10.7$, 18 women, 41 men, 1 other). In line with past research on acquiring quality results without attention check questions, we required that workers had completed at least 100 tasks with an approval rate of at least 95% [284]. The experiment took approximately 30 minutes and participants were compensated with $5.00 for their time.

In this experiment, participants were shown 3 charts and asked to answer 6 questions based on each one. Each question was designed to target a specific level of Bloom's taxonomy and was presented in order, beginning with Knowledge and ending with Evaluation. The order of charts was determined via a 2 by 3 Latin Square design (yielding 6 unique chart orderings and thus 6 conditions). Each participant saw 1 version of each chart in accordance with their assigned condition. In 3 of the conditions, participants saw 2 of the original charts and 1 redesigned chart, while participants assigned to the other 3 conditions saw 1 original and 2 redesigned charts.

We want to emphasize that the purpose of using the alternative designs in this study was to show the range of reader interpretations and visualization affordances our method could evaluate, rather than to generate concrete design guidelines from these case studies. Because of this, we note that our analysis utilizes an exploratory approach – not formal hypothesis testing.

4.4.3 Results

Figure 4.2 provides the questions used for the COVID-19 chart and a summary of results for all 6 levels.

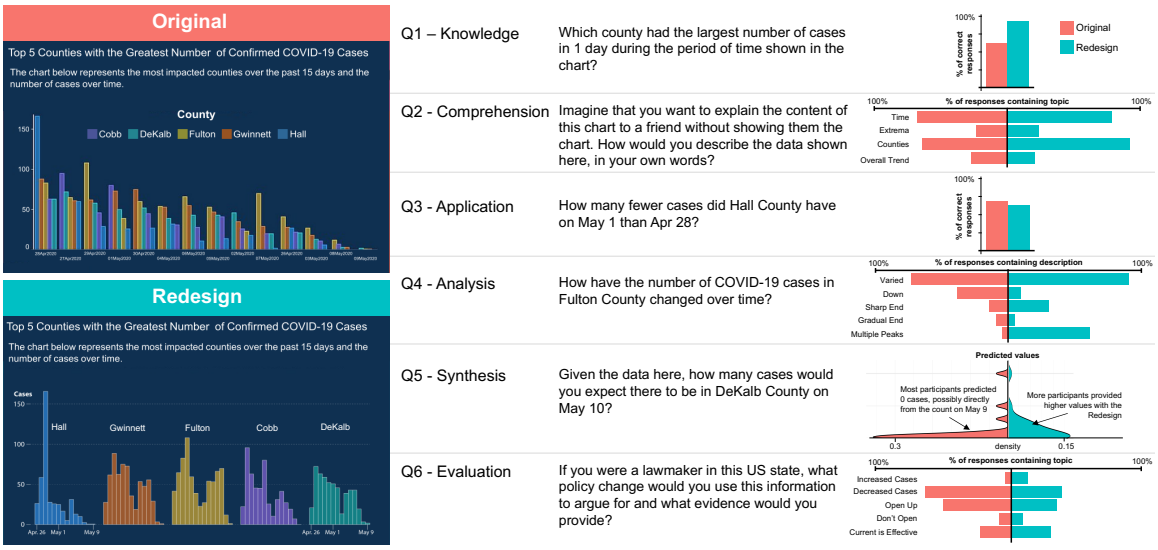


Figure 4.2. A summary of the stimuli and questions used in the experiment for the COVID charts. Charts on the right show results for each level examined. Participants answered Question 1 correctly significantly more often when using the redesigned chart than the original. This does not hold for Question 3, however. When describing the contents of the chart (Q2), participants viewing the redesigned chart were more likely to talk about the chart on a county level, but otherwise answered similarly. Participants more often incorrectly classified the trend of the chart as going down when viewing the original version and were less likely to comment on the bi-modal shape (Q4). We observed no effect of the version on the prediction made about the number of cases (Q5) and, unexpectedly, participants argued for "opening up" and cited decreasing cases as frequently in both groups (Q6).

4.4.4 Question 1: Knowledge

We begin our analyses with the first question, which asked participants to locate particular values in the chart. We constructed a logistic regression predicting response accuracy with visualization design (original vs. redesigned) across the three chart topics. We found that participants were more likely to answer the question correctly when they saw the redesigned visualization compared to the original ($\chi^2(1) = 21.45$,

$p < 0.001$). There was no main effect of the topic on this relationship ($\chi^2(2) = 4.54$, $p = 0.10$) (see Supplemental materials for pair-wise ratios in our online repository: `https://osf.io/85rcf/?view_only=bb1995101d04400bbb70f0b7a8931fc6`). This suggests that participants were retrieving values from the redesigned visualizations more accurately compared to the original version. Further, as the visualization was redesigned following common guidelines to increase clarity and afford more accurate value retrieval, this result suggests that the Knowledge level questions were successful in capturing the improvement brought by the redesign.

4.4.5 Question 2: Comprehension

Questions targeting the Comprehension level were open-ended without defined correct or incorrect answers. Therefore, to analyze these responses, we read through the responses blind to the version seen, and identified varying categories of conclusions. Each response was tagged as either containing or not containing each conclusion. This approach was selected in order to compare differences in patterns identified between the two versions of each chart. We note that while this approach was chosen for demonstration, this kind of analysis is not specific to our proposed method; many existing techniques for analyzing qualitative results could be appropriate here, such as content-based analyses or interpretive analyses (see [26] for some common practices). We conducted exploratory analysis using Pearson's chi-squared tests of independence to examine the relation between visualization design and user-identified salient patterns, with Yates' continuity correction and Bonferroni adjustments. See the Supplemental materials for details.

As shown in Figure 4.2, participants who saw the original and redesigned versions of the COVID chart describe salient features with varying frequencies. For example, they were trendingly more likely to identify and compare the counties included in the redesigned chart, ($\chi^2(1,\ N = 54) = 5.33$, $p = 0.083$). The participants also mentioned

different dates depending on the chart they viewed. Namely, all seven participants who saw the original chart incorrectly identified the earliest date present as April 28, when it is April 26. This was a fact correctly identified by all participants who viewed the redesigned chart.

In the Immigration chart, participants pointed out similar salient patterns in both visualizations. In other words, there is no difference in the distributions of conclusions drawn across both the original and the redesigned visualizations ($\chi^2(1, N = 51) = 0.51, p = 1$). We found this to be somewhat unexpected. We expected participants to more often comment on the total number of immigrants when the data was visualized as a stacked bar chart (as in the original) because they are viewed as the best choice for communicating an overall quantity among bar chart variations [365], but this was not the case. Similar to the immigration chart, we found that participants pointed out similar patterns in both versions of the Market chart (see Supplemental materials for more details).

These three case studies illustrate different outcomes to expect when analyzing responses to this kind of question. We see that the Market and Immigration redesigns did not differ from the original, suggesting that they afford the same set of conclusions. This is different from the redesign of the COVID-19 visualization, which made a different set of patterns more salient to the viewers. We recognize that there could be more subtle differences in affordances between the original and the redesign that this question does not capture. We speculate why this might be in the Discussion section.

4.4.6 Question 3: Application

Questions at this level challenged participants to determine the numerical difference between two specific points. As with Question 1, we constructed a logistic regression predicting response accuracy with the original and redesigned visualization across the three charts. Unlike with Question 1, participants who viewed the

redesigned version of a chart were not statistically more likely to answer this question correctly ($\chi^2(1) = 2.51$, $p = 0.11$). In addition, we observed no statistical difference between the accuracy between charts ($\chi^2(2) = 2.96$, $p = 0.23$) and no significant interaction ($\chi^2(2) = 2.97$, $p = 0.23$).

Although our questions at this level failed to find a significant difference between the two versions of the charts with respect to answer correctness, there may be other factors at play as well, which we discuss in the Discussion section.

4.4.7 Question 4: Analysis

As with Question 2, questions for level 4 were open-ended. Participants were asked to describe a specific trend present in each chart. We identified a series of descriptions for each chart that either were mentioned by the participants or were determined by the authors as reasonable conclusions to mention. Their responses ranged from describing the directions of the trend (e.g., up or down) to commenting on modality. We coded each response blind to the chart version, tagging each description as present or not (e.g., was the trend described as positive or not). We then compared the frequency of descriptions between the two chart versions. We note that, as before, this is not the only way to complete this analysis, but was selected for demonstrative purposes. We conducted chi-square analysis to examine the relation between chart design and chart descriptions, similar to that in Section 4.4.5.

In response to the COVID-19 chart, we can see two distinct points of difference in the distributions (as shown in Figure 4.2). First, although participants were equally likely to describe the trend direction as containing both positive and negative sections, ($\chi^2(1, N = 43) = 1.06$, $p = 1$), more participants incorrectly categorized the trend as decreasing when they viewed the original chart compared to the redesigned chart, although not significantly ($\chi^2(1, N = 43) = 4.25$, $p = 0.20$). In addition, participants more frequently noticed the bi-modal shape in the data when viewing the redesigned

chart ($\chi^2(1,\ N = 43) = 12.34$, $p = 0.0022$), such that only one participant who viewed the original version identified the bi-modal shape in the data whereas over 60% of the participants with the redesigned chart noticed this bi-modal pattern (see Figure 4.2). This suggests that the redesigned version more readily affords viewers the ability to see the bi-modality of the case count. In reality, answering this question correctly with the original version of the COVID-19 chart is extremely difficult. Because the labels on the x-axis are not in chronological order, characterizing the distribution over time would require mentally reorganizing the bars. However, it is important to note that none of the participants who saw the original version of this chart gave any indication on any question that they noticed that the dates were out of order, suggesting that either no participants noticed this feature or found it pertinent.

With the Immigration chart, we observed that most of the topics commented upon by participants are similar across versions, but with one very distinct difference. Namely, participants who viewed the original chart were significantly more likely than those who viewed the redesign to **correctly** describe the trend of immigration as increasing ($\chi^2(1,\ N = 51) = 10.23$, $p = 0.007$). We suspect this unexpected disparity may be driven by the increasing size of the stacked bars over time. Additionally, because of the relatively small change each decade, it is possible that participants viewing the redesigned chart considered the change not remarkable enough to mention. Regardless of the cause, it represents a distinct difference in the affordances between these charts that this question was able to capture. The distribution of topics discussed in response to the Markets chart was highly similar across visualization versions with no significant difference with respect to the version viewed (see Supplemental materials for details).

The difference in distributions here suggests that the different versions of the chart lead to participants to come to describe the underlying data differently. This

difference was not captured by the previous three levels, supporting the importance of a multi-leveled approach to evaluate visualization understanding.

4.4.8 Question 5: Synthesis

At this level, participants were asked to make predictions about future data values and trends for each chart. By looking at the distributions of the predicted values, we begin to unpack the decision making process afforded by each chart and version. In particular, we can glimpse both where the design had an effect on the average prediction made, as well as on the variance of those predictions. To identify statistical differences, we utilized Welch's two-sided t-tests with Bonferroni corrections to compare mean predictions and two-sided F-tests with Bonferroni corrections to compare variances. For this analysis, we only included responses that contained a single, numeric answer, excluding those with ranges or that described a trend. This excluded 17% of responses from the COVID-19 chart, 0% from Immigration, and 7% from the Markets charts.

For the COVID chart, participants were asked to predict the number of COVID-19 cases for one day beyond the dates shown in the chart. Results showed no significant difference between the mean prediction (($M = [5.13,\ 5.13]$, $SD = [14.84,\ 11.77]$, $t(41.83) = 0$, $p = 1$) nor the variance of the predictions made ($F(22, 22) = 1.59$, $p = 0.28$). This suggests that though we observed differences between the versions in previous questions, both charts afforded similar predictions to participants.

For the immigration chart, participants were asked to predict the percentage of the population made up of immigrants in Thunder Bay, Ontario, and Canada. First, we observed that the mean prediction for all 3 locations were not significantly different across the two versions (see Supplemental materials for details). While the variance of predictions between chart versions was not significantly different for Thunder Bay ($F(9, 7) = 0.73$, $p = 1$) or Ontario ($F(9, 7) = 4.75$, $p = 0.31$), it was significantly

different for the predictions about Canada ($F(9, 7) = 15.40$, $p = 0.009$), such that the predictions made about Canada's population were more clustered with the redesigned version than the original version. This suggests that while both chart versions afford similar numeric predictions, the redesigned chart affords less variation in predictions.

When viewing the Market chart, participants were asked to predict the trade deficit and the manufacturing employment. Participants predicted trade deficits to be significantly higher when they viewed the redesigned version ($M = [272.19,\ 355.00]$, $SD = [100.47,\ 49.61]$), $t(21.053) = -3.0032$, $p = 0.027$). We additionally see that the shapes of the distributions were different: the predictions made with the original chart were highly clustered while those made with the redesigned chart were far more variable. Our analysis suggests that the difference between these distributions is significant ($F(15, 18) = 4.10$, $p = 0.021$). There was no significant difference between the predictions made by participants about the number of people employed in manufacturing with respect to chart version ($M = [38.18,\ 11.56]$, $SD = [57.60,\ 1.59]$, $t(16.026) = 1.9044$, $p = 0.30$). However, we observed that the variance of the predictions is, again, significantly different between the two versions ($F(16, 15) = 1311.7$, $p < 0.001$). These results suggests that while both charts afford similar predictions about manufacturing employment, the original chart affords significantly higher predictions about trade deficit than the redesigned chart.

Although all of our questions about the Synthesis level did not reveal differences between versions of every chart, we were still able to reveal some different affordances that were not captured by the previous questions. Evaluating these charts systematically across levels of understanding allows the identification of distinctions between affordances that might otherwise be missed.

4.4.9 Question 6: Evaluation

In the sixth and final question, participants were asked to apply their learning to a real-world situation by describing the argument that they would make. Similar to Section 4.4.5 and 4.4.7, we identified common conclusions and evidence that was cited, tagged each response as containing or not containing mentions of each of these topics, and used the same chi-square analysis approach.

In response to COVID-19 charts, the most common conclusion drawn by participants was for "opening up" (see Figure 4.2). This conclusion was mentioned by participants regardless of which version of the COVID-19 chart they were presented with ($\chi^2(1, N = 44) = 0.76$, $p = 1$). Additionally, the most common evidence cited by participants in defense of their claim was a decrease in COVID-19 cases. Surprisingly, participants were as likely to use this argument when viewing the original as the redesigned chart ($\chi^2(1, N = 44) = 2.32$, $p = 0.76$). This is unexpected, as the original COVID-19 chart shows a strong downward trend in cases (as long as the x-axis ordering is ignored), but the redesigned version of the chart shows increasing numbers in several of the counties. This result suggests that despite the differences in design, both charts afforded participants the ability to come to the same conclusion. The unexpected nature of this result further emphasizes the reason why asking difficult questions in evaluation is important; sometimes conclusions arise because (or in spite) of otherwise careful encoding.

For the Immigration chart, participants were asked to argue why the population of Thunder Bay was not representative of the population of Canada at large. The two main arguments were "there were fewer immigrants in total" and "there were fewer immigrants over time," but participants were equally likely to report these justifications regardless of the chart they saw ($\chi^2(1, N = 47) = 0$, $p = 1$).

As for the Markets chart, we see that participants were very consistent across versions with respect to the conclusions drawn and evidence provided. One repeated

theme present in the responses was a suggestion of a causal relationship between the number of people employed in manufacturing and the trade deficit with China. This relationship was suggested in both directions (i.e. changes in the deficit causes changes in employment and changes in employment cause changes in the deficit) and did not appear statistically more often in response to either chart version ($\chi^2(1, N = 46) = 0.014$, $p = 1$; $\chi^2(1, N = 46) = 0.51$, $p = 1$). This suggests that the two charts did not afford different conclusions nor justifications.

4.4.10 Is this taxonomy really hierarchical?

The original Bloom's taxonomy argues that there is a strict hierarchical relationship between the levels [33]. To shed some light on this idea, we looked for evidence that would suggest that performance on earlier questions is correlated with performance on later ones.

Because Questions 1 and 3 had correct answers, we first examined if success on Question 1 was correlated with success on Question 3. We constructed a logistic regression and observed that participants who answered Question 1 correctly were significantly more likely to answer Question 3 correctly compared to those who answered Question 1 wrong ($\chi^2(1) = 22.3411$, $p < 0.001$) irrespective of chart topic ($\chi^2(2) = 1.02$, $p = 0.60$), version ($\chi^2(1) = 0.0099$, $p = 0.92$), and with no significant interaction effect ($\chi^2(2) = 3.46$, $p = 0.18$). The effect size is quite pronounced for all three charts on both original and redesigned versions. For example, participants who answered Question 1 correctly using the redesigned COVID-19 chart were 1.8 times more likely to answer Question 3 correctly. More dramatically, participants who viewed the original COVID-19 chart were over 10 times more likely to answer Question 3 correctly if they answered Question 1 correctly. This is unsurprising – in order to answer Question 1 correctly, one must estimate one value and Question 3

builds on this skill by requiring participants to estimate two values and then subtract them.

To further explore the relationship between performance on questions, we asked whether performance on Questions 1 or 3 predicted a participant's performance on Question 4. That is, is there a relationship between how well a person locates or determines the difference between points and how well they classify a trends? For this, we reviewed the responses to Question 4 and marked those which were obviously incorrect. We then used another logistic regression that predicted whether a response was obviously incorrect or not based on the chart topic, version, and the participants' performance on Question 1 and 3. Our model suggests that there is no significant relationship between performance on Question 1 or Question 3 on the whether their response to Question 4 was reasonable or not (see Supplemental materials). This suggests that though there may be some overlapping skills required between levels, this taxonomy is not hierarchical in nature.

4.5 Experiment 1: Understanding

In infographics, pictographic arrays are often used to depict part-to-whole relationships. We explore how this choice affects the insights that viewers draw from the image, compared to using traditional charts that encode the same information with solid areas by using the methodology for constructing comprehension questions discussed in the previous section. In this experiment, we explore the 3 hypotheses described in Section 4.5.3.

4.5.1 Experiment 1: Stimuli

For this experiment, we used 6 pairs of infographics that display part-to-whole relationships. In each pair, one encoded that relationship with solid, geometric areas and the other used pictograph arrays (see Figure 4.3). We will refer to these two

types as "versions" and, for brevity, call the two chart versions "Area" and "Count," respectively.

Our criteria for selecting charts included: diversity of chart type, diversity of topic, and a comparable number of variables. Past research suggests that familiarity with a chart type can influence perceptions of attractiveness and ease of use [306]. Therefore, we wanted our stimuli set to include chart types which were common in media so that participants were unlikely to be encountering a chart type for the first time, but that would vary in familiarity. Our final stimuli set includes 1 bar chart, 1 pie chart, 2 stacked bar charts, 1 donut chart, and 1 treemap (listed in decreasing familiarity for general audiences according to [306]).

We also aimed to select charts with a diverse set of topics that would vary in familiarity and interest among participants. Existing work has shown that background knowledge on and interest in the subject matter of a visualization can impact engagement [195], so we hoped a diverse set of topics would include something for every participant. The topics of the charts in the final set were: Alphabet's earnings, the number of COVID-19 cases and deaths in a US state, the severity of COVID-19 symptoms, the threat of extinction for different kinds of animals, the guns used in mass shootings in the US, and the number of times different diseases were mentioned on Twitter. Of this set, we hypothesized that the charts about COVID-19 would be the most familiar topic and the breakdown of Alphabet's earnings would be the least familiar.

While all 6 pairs take inspiration from real-world infographics, 5 of the pairs contained a modified version of real, published charts gathered from sources including the New York Times, Washington Post, and Visual Capitalist. The sixth pair was created by one of the authors in the style of existing area charts of COVID-19 cases, but using a smaller data set that more closely matched the range of variables present in the other stimuli. Each chart contained between 3 and 6 variables. The specific

range of variables was not pre-selected, but emerged through our search for real-world charts with limited visual complexity.

To isolate the effects of using pictographs versus solid areas, when creating the informationally equivalent designs, we maintained other properties, including colors, positions, labels, legends, and shapes. Only 2 of the charts originally contained pictographs, so when creating the alternate version, the research team selected publicly available pictographs that we thought would be reflective of the topics.

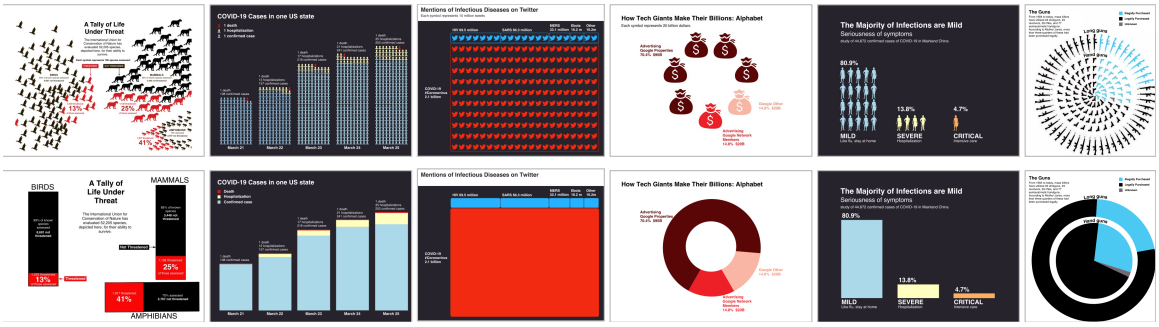


Figure 4.3. During Experiment 1 and 2, participants saw six different visualizations. Each visualization was drawn from a pair of visualizations (columns) which were informationally identical but which either encoded values as pictograph arrays (top row) or solid areas (bottom row).

4.5.2 Experiment 1: Method

We conducted this experiment on Amazon's Mechanical Turk. We recruited 60 Workers ($Mean_{age} = 39.44$, $SD_{age} = 11.28$, 18 women, 30 men, 12 others) who had completed at least 100 tasks with an approval rate of at least 95%. We selected this combination of tasks completed and approval rate in line with existing research on quality data without attention-check questions [284]. We decided to survey 60 participants after calculating a power analysis based on the effect size of a pilot study. This experiment was conducted in 2 phases: Comprehension and Comparison. All participants completed both phases. In total, the experiment took about 30 minutes and participants were paid $5.00.

Level	Description	Example Question
1. Knowledge	Learner can recall facts without an understanding of the meaning behind them.	What percentage of birds are threatened?
2. Comprehension	Learner is able to understand the information in context.	What is the take-away message from this infographic?
3. Application	Learner can apply knowledge to a new problem or represent it in a different way.	If your town has 200 birds, how many of them are threatened?
4. Analysis	Learner can break down a concept into component parts and understand the relationship between them.	Are birds or mammals in more danger of extinction?
5. Synthesis	Learner is able to generalize and use their knowledge to make a prediction or create something new.	How do you think your community would react to this information?
6. Evaluation	Learner is able to assess the value of ideas and methods and make choices based on reasoned arguments.	What do you think public officials should do to combat this problem?

Table 4.2. We used Bloom's taxonomy [33] to create comprehension questions following the method from [53]. The levels of the original taxonomy are shown here, along with a sample question indicative of those created for the experiment.

Phase 1: Comprehension. The first phase of the experiment asked participants to answer a series of 6 comprehension questions based on the levels of Bloom's taxonomy [33, 53]. This method, which was proposed in [53], allowed us to comprehensively assess different aspects of understanding which differ in complexity. The possible limitations of this approach (and the newness of the technique) are discussed in Section 4.7.

Because each chart has a different topic, we created a set of 6 different questions per chart — see Table 4.2 for a description of each level in the taxonomy and an example question for each level. The entire list of comprehension questions can be found in supplementary materials. This experiment has a 2 by 6 Graeco Latin Square design with repetition, wherein each participant sees one version of all 6 infographics (3 Count, 3 Area).

Phase 2: Comparison. In the second phase of the experiment, participants were shown both versions of each infographic, side by side, and asked to compare them on a series of metrics (see Table 4.3). Participants viewed the charts in the same order as in the comprehension portion. We used an additional Latin Squares design to randomize whether the Area condition appeared on the left or right. The metrics

were selected after compiling a list of qualities remarked in existing literature as being important for effective infographics [216, 417]. From this list, we selected those which we thought could be affected by the presence or absence of pictographs and designed a question based on each. For each metric, participants indicated on a 5-point Likert scale which of the two versions of the chart best satisfied the prompt. The extremes of these scales indicated a strong preference for one chart over the other. Finally, participants scored their familiarity with the topic of each chart on a 5-point Likert scale.

Metric	**Question**
Readability	Which of these charts is quicker and easier to read?
Visual Appeal	Which of these charts is more visually appealing?
Understandability	Which of these charts is clearer and easier to understand?
Envisioning	Which of these charts makes the data easier to imagine?
Clutter	Which of these charts is more visually cluttered?
Complexity	Which of these charts is more complex?
Importance	Which of these charts makes its topic seem more important?
Familiarity	How familiar are you with the topic of this chart?

Table 4.3. In Experiment 1, Phase 2 participants were asked to compare 2 versions of the same chart in response to these metrics and questions.

4.5.3 Experiment 1: Hypotheses, Metrics and Analysis

We developed a coding scheme to analyze the open responses from Phase 1 (Comprehension) of the experiment. Of the 6 comprehension questions we used per chart, 3 questions (Questions 1, 3 and 4) have a correct answer, so responses to these were marked as correct or incorrect. For the other 3 questions that have subjective responses (Questions 2, 5, and 6), we developed codes relating to the 3 hypotheses described below. All responses (except for those deemed to be extremely low quality and disqualified) were coded on every dimension by one of the authors. Of the 60 participants sampled, 13 (about 19%) were disqualified. For two of the metrics (mental effort and insight) which are both of high importance to our research question and are based on a subjective scale, we additionally utilized a second coder outside of the research team and computed inter-rater reliability. Data from Phase 2 of the experiment was recorded numerically. As the location of the two chart versions dif-

fered across conditions, we applied a transformation to the scores such that a score of 1 corresponds to strong preference for the Area version and a score of 5 corresponds to a strong preference for the Count version. The following 3 hypotheses guided our analysis:

- **Hypothesis 1.1**: Participants will draw higher-level insights from Count infographics than Area infographics.
- **Hypothesis 1.2**: Participants will have a higher level of engagement with Count infographics than Area infographics.
- **Hypothesis 1.3**: Participants will show a stronger emotional response when they look at Count infographics than Area infographics.

4.5.4 Experiment 1, Phase 1: Results

In the following sections, we describe the results with respect to each hypothesis, then conclude whether the results support the hypothesis.

4.5.4.1 Hypothesis 1.1: Higher-level insights

Insight. Question 2, 5, and 6 tested the ability of participants to extract key messages and gain insight. These questions had no objectively correct answers, so we designed a 7 point scale (0–6) judging the kind of insight present. On this scale, a score of 0 indicated no information gathered from the chart. Low insight responses (assigned scores of 1 or 2) referenced the information directly from the chart and demonstrated no further attempt at casual sensemaking. For example, in response to Q2 about the main takeaway of the chart about COVID-19 cases, one participant responded *"Everything is increasing every day."* Medium insight responses focused on a single dimension (score: 3) or offered opinions without coming to conclusions based on the data (score: 4). For example, to the same chart, another participant responded *"Death rate goes down with more testing."* Finally, high insight responses came to conclusions not included in the chart (for a score of 5) and offered justification for

those responses (for a score of 6), such as here, where a participant speculates about what is to come: *"This shows that the number of cases increase rapidly and could double in a week."*

These insight scores were independently assigned by 2 coders (1 author, 1 outside) using a common scale, then compared for differences and discussed (see supplementary materials for the exact scale used by the coders). Inter-rater reliability was calculated using a weighted Kappa with squared weights for Question 2, 5, and 6 and was 0.959, 0.979, and 0.89, respectively.

Overall, our results indicated that chart version had no significant effect on insight extraction. To obtain this result, we predicted level of insight with a mixed-effect linear model from the lme4 package in R [20]. The model suggested that chart version had no significant effect on insight extraction ($\chi^2 = 0.47$, p = 0.49). However, the model suggested that there was a main effect of question number ($\chi^2 = 902.26$, $p < 0.001$), such that the highest level of insight was displayed in responses to Question 6 (Mean = 4.58, SE = 0.91), followed by Question 5 (Mean = 2.62, SE = 0.91) and Question 2 (Mean = 1.88, SE = 0.91). This is not surprising considering that Question 2, 5, and 6 were designed to probe for increasingly complex aspects of understanding, which yield higher levels of insight. We also observed a main effect of chart topic ($\chi^2 = 21.53$, $p < 0.001$). Viewers reported the highest level insights in response to the chart of confirmed cases of COVID-19 (Mean = 3.33, SE = 0.112). This was significantly higher than the chart about Alphabet's earnings ($p < 0.001$), extinction of animals ($p = 0.016$), and severity of COVID-19 symptoms ($p = 0.022$).

Accuracy. In Phase 1, Questions 1, 3, and 4 had an objectively correct answer. Question 1 asked participants to pull a single datapoint out of the chart, Question 3 asked for a short calculation based on the data, and Question 4 asked for a comparison of two datapoints. Our results revealed that while the question and chart topic had an effect on whether a question was answered correctly, the version did not.

Overall, accuracy was high — 92.9% of the participants answered Question 1 correctly, 70.4% of the participants answered Q3 correctly, and 92% of the participants answered Q4 correctly. Participants had the highest accuracy when answering questions about the extinction chart (90.2%) and the lowest accuracy when answering about the chart about guns (80.6%). Post-hoc analysis with Tukey adjustments showed no significant difference in accuracy between topics. Using a mixed-effect linear model to predict accuracy with question type, chart version, and chart topic, we found no main effect of chart version ($\chi^2 = 0.0024$, $p = 0.96$), but an overall main effect of question ($\chi^2 = 106.56$, $p < 0.001$) and chart topic ($\chi^2 = 12.62$, $p = 0.027$). Post-hoc analysis with Tukey adjustments [228] showed that participants performed significantly worse on Question 3 compared to Question 1 ($Est = 0.23$, $SE = 0.025$, $p < 0.001$) and 4 ($Est = -0.22$, $SE = 0.025$, $p < 0.001$). Because Question 1 and 4 were tasks related to identifying information in the chart directly and Question 3 asked participants to compute a number based on information from the chart, we hypothesize that the dip in accuracy for Question 3 may be driven by higher graphical literacy than numeracy in our participant population.

Misinterpretation. Although few responses contained misinterpretations overall (4.69%), we find that chart version had an effect. A logistic general linear model predicting misinterpretations with chart version and chart topic suggested a trending effect of chart version ($\chi^2 = 3.22$, $p = 0.056$), such that participants were 1.82 times more likely to misinterpret the Count charts. There was also an effect of chart topic ($\chi^2 = 45.74$, $p < 0.001$). A closer look revealed that a majority of these misinterpretations come from responses to the chart about Twitter mentions. Instead of being about Twitter, the chart was misinterpreted to be comparing the severity of infectious diseases, such as in this response from a participant: *"Coronavirus is more dangerous than any other pandemmic [sic]."*

Conclusion: Evidence does not support Hypothesis 1.1 After observing that chart version had no effect on insight and accuracy, as well as a mixed effect on likelihood of misinterpretations, we therefore conclude that the data do not support Hypothesis 1.1: Participants will draw higher-level insights from Count infographics than Area infographics.

4.5.4.2 Hypothesis 1.2: Higher Engagement

Mental Effort. As with insight, we created a 7-point scale (0–6) for rating the mental effort displayed in each response. A comment with a score of 0 mentioned no information from the infographic and a comment with score of 6 considered all information, drew a conclusion, and provided evidence to support that conclusion (the scale used by coders is included in the supplementary materials). Low-mental effort answers copied text from the title (score: 1) or restated text directly from the title, legend, or annotations (score: 2), such as in this response to Question 2 about the main take-away for the chart about extinction: *"To show the percentage of animals that are endangered species."* In contrast, medium effort responses focused on a single dimension (score: 3) or observed a comparison (score: 4), as seen here: *"Birds are less likely to be threatened than mammals or amphibians."* High effort responses made conclusions based on the information (score: 5) and justified their responses (score: 6). For example, this response received a score of 5, because they drew a conservation message which was not present in the chart: *"That we need to be more careful with the earth or we will cause many new extinctions."*

Mental effort scores were independently assigned by 2 coders (1 author) using the mental effort scale paraphrased above (exact scale included in the supplementary materials). After scores were assigned, they were compared and differences were discussed. Inter-rater reliability was calculated using a weighted Kappa with squared weights and was 0.972, 0.986, and 0.936 for Questions 2, 5, and 6, respectively.

As with insight and accuracy, we found that the question and chart topic had an effect on mental effort, but the chart version did not. A mixed-effect linear model predicting mental effort suggested a main effect of question ($\chi^2 = 385.95$, $p < 0.001$) and a main effect of chart topic ($\chi^2 = 27.31$, p = 0.49), but no effect of chart version ($\chi^2 = 0.05$, p = 0.82). Post-hoc analysis with Tukey's adjustment suggested that, like with insight, Question 6 elicited significantly higher mental effort compared to that of Question 5 ($p < 0.001$) and Question 2 ($p < 0.001$). Viewers seemed to exert the most mental effort when responding to the chart about Alphabet (M = 2.84, SE = 0.14), which is significantly higher than the chart about extinction (M = 2.14, p = 0.0058), and COVID-19 cases (M = 2.66, p = 0.017). See supplementary for full pair-wise comparisons.

Response length. We also examined engagement via response length. Our mixed-effect linear model predicting the number of tokens suggested a trending main effect of chart version ($\chi^2 = 3.59$, p = 0.058), such that viewers wrote on average one more word when viewing an Area chart (M=14.4, SE = 0.93) than a Count chart (M=13.4, SE = 0.93). Additionally, there was a main effect of question ($\chi^2 = 75.94$, $p < 0.001$) such that participants wrote the longest responses for Question 6 (M = 16.6 words, SE = 0.97) and a main overall effect of chart topic ($\chi^2 = 11.17$, p = 0.048), although post-hoc analysis reveals no particular chart elicited a longer response than others.

Conclusion: Evidence does not support Hypothesis 1.2 Considering that chart version had no effect on mental effort and participants wrote more words when viewing an Area chart, we therefore conclude that our results do not support Hypothesis 1.2: Participants will have a higher level of engagement with Count infographics than Area infographics.

4.5.4.3 Hypothesis 1.3: Stronger Emotional Response

Emotional Response. While we hypothesized that viewers would show stronger emotional response when looking at Count than Area charts, Pearson's Chi-squared test showed that viewers were significantly more likely to respond in a neutral fashion ($p < 0.001$), and overall there was no significant difference in Area and Count chart in eliciting emotional responses ($\chi^2 = 0.43$, p = 0.81).

Negative responses reflected a variety of emotions including shock, worry, or frustration. An example of this was a response to the Alphabet chart in which a participant wrote *"if i was a decision making[sic] at google i would stop caring about growth and instead focus on using that wealth to help struggling people."* Neutral responses reflected facts or preferences, such as this response to the same prompt *"I would invest more in Google Properties."* Finally, positive responses also varied in emotion including relief, interest, and hope, as is expressed here by one participant about using the chart about Symptoms of COVID-19, *"to calm them and provide data and let them know that it's not that bad."*

Surprise. Because past literature identified "surprise" as a component of the insight generation process (e.g., [264]), we also coded for this dimension (binary, contains/-does not contain). Chart version did not have a significant effect on eliciting expressions of surprise ($\chi^2 = 0.092$, $p = 0.76$). Very few viewers mentioned surprise in their response (1.69%), with about half of them in response to a Area chart and the other half in response to a Count chart. Responses containing surprise varied across topics, but often related to the size of particular dimensions. One example of this was in response to the chart about confirmed cases of COVID-19, *"They would be surprised that there were so few cases of death and hospitalizations as the number of confirmed cases grows."* They also commented on the overall size of the units depicted such as this response to the Alphabet chart *"People in my community would be very surprised*

at how much money that these companies make from Google and I think they would be very surprised at these numbers."

Conclusion: Evidence does not support Hypothesis 1.3 We observed that chart version had neither an effect on emotional response nor surprise. We can therefore conclude that our evidence does not support Hypothesis 1.3: Participants will show a stronger emotional response when they look at Count infographics than Area infographics.

	Readability	Visual Appeal	Understandability	Relatability	Clutter	Complexity	Importance	Familiarity	VIF
Readability	1.00								4.52
Visual Appeal	0.38	1.00							1.85
Understandability	0.86	0.40	1.00						4.68
Relatability	0.34	0.59	0.46	1.00					1.97
Clutter	-0.66	0.00014	-0.61	0.0058	1.00				3.29
Complexity	-0.58	0.059	-0.54	0.0029	0.77	1.00			2.65
Importance	0.10	0.46	0.17	0.49	0.25	0.25	1.00		1.59
Familiarity	-0.017	0.061	-0.04	-0.0087	0.12	0.069	0.13	1.00	1.04

Table 4.4. Correlation table and VIF for metrics from Experiment 1, Phase 2. Note that larger values are darker, regardless of sign. Several metrics were highly correlated, such as readability and understandability, or perceived clutteredness and complexity. VIFs were relatively high. For example, both readability and understandability had VIFs greater than 4, rendering them less-optimal measures of viewer attitudes as they can be highly accounted for by other dimensions. Experiment 2 modified these metrics to reduce the multicollinearity between them.

4.5.4.4 Other Observations

Numerical Thinking. Though numerical thinking was not directly related to any hypothesis, we noticed evidence of it while analysing the results. We identified two categories of numerical thinking: references to frequency (e.g. 8 in 10) and probability (e.g. 80%). Overall, participants were significantly more likely to mention probability than frequency when viewing Area chart ($p < 0.001$), but were equally likely to mention frequency and probability when viewing Count chart ($p < 0.001$). Note, however, that very few participants mentioned frequency or probabilities overall ($\sim$ 3.9%), meaning that the small percentage of frequency/probabilities mentions likely skewed the Chi-square approximation to exaggerate effects ($\chi^2 = 12.82$, $p = 0.0016$).

Relating to Community. Though it was not connected to any particular hypothesis, we also examined the responses to Question 5 more closely because it asked viewers to speculate and discuss how their own community would react to the information shown. Specifically, we coded for (1) whether participants deemed the information to be relevant enough to affect their own community, (2) whether they described their community in their response, and (3) whether they explicitly mentioned their community would take action. We found that chart version had no effect on the prevalence of any of these connections.

With chart topic and chart version as fixed effects and participant as random effects, mixed-model regression analysis showed that chart topic significantly predicted whether a viewer would judge the information as relevant to their community ($\chi^2 = 44.12$, $p < 0.001$), with the chart about COVID-19 cases as the most relevant (detailed pair-wise comparisons can be found in the supplementary). There was, again, no significant effect of chart version ($\chi^2 = 1.01$, $p = 0.31$). Participants that referenced their community often did so while reflecting on the relevance of the information to them. For example, a participant expressed why the chart about COVID-19 symptoms was very relevant: *"I think that people in my community would be relieved that most cases are mild but also wary that the percentages are still significant that infection will require hospitalization and possibly intensive care. They might be more prone to stay home and isolate."* Another reflected on why a chart about extinction wouldn't have any effect on their community: *"It wouldn't as my community is already a bird sanctuary area and has been for decades."*

We found a similar pattern of effects regarding whether participants described their community or mentioned taking action. For descriptions, there was a main effect of chart topic ($\chi^2 = 19.80$, $p = 0.0014$), but no main effect of chart version ($\chi^2 = 2.42$, $p = 0.12$). Similarly, chart version did not have a significant effect on mentions of taking action ($\chi^2 = 0.82$, $p = 0.37$), but chart topic did ($\chi^2 = 26.91$,

$p < 0.001$). Participants were more likely to describe their community and mention taking actions in responses to the charts about extinction and guns. For example, one participant described their geographic location as a means to justify the relevance of the extinction chart *"It would effect it to a small extent although the threat of extinction to Amphibians could pose an issue as we are located on the ocean."* Another mentioned the action their community would take in response to the guns chart, writing *"Smart people would back legislation to make all guns more difficult to obtain. People would be more willing to accept stricter background checks and mandatory registration."*

4.5.5 Experiment 1, Phase 2: Results

Readability. Our results suggested that there was an overall difference in perceived readability ($\chi^2 = 61.71$, $p < 0.001$). Post-hoc analysis with Bonferroni adjustments [171] suggested that viewers were evenly divided on this issue, such that there were equal amounts of participants who rated the Area (45%) and Count chart (46%) as more readable ($p = 1.00$) (1 ▮▪▪▪▮ 5). Details of pair-wise post-hoc comparisons can be found in the supplementary materials.

Visual Appeal. Ratings of visual appeal were similar to those of readability, such that there was an overall difference ($\chi^2 = 39.03$, $p < 0.001$) driven by polarized ratings (1 ▮▮▪▮▮ 5). Viewers were equally likely to rate Area (46%) and Count chart (47%) as more visually appealing ($p = 1.00$), with very few giving neutral ratings (6.5%).

Understandability. Similarly, viewers were evenly divided on understandability ratings ($\chi^2 = 25.13$, $p < 0.001$) (1 ▮▪▪▪▮ 5). They were equally likely to rate Area (43%) and Count charts (44%) as easier to understand ($p = 1.00$), with very few giving neutral ratings (13%).

Relatability. Although there was an overall difference in relatability ratings ($\chi^2 = 23.49$, $p < 0.001$) (1 ▅▃▂▆▇ 5), post-hoc analysis revealed that viewers were not more likely to rate one chart version as more relatable ($p = [0.25, 1.00]$).

Clutter. There was no overall difference in ratings of perceived clutter ($\chi^2 = 2.21$, $p = 0.70$) (1 ▇▇▇▇▇ 5), such that participants rated Area and Count as equally cluttered.

Complexity. Participants gave differing complexity ratings ($\chi^2 = 27.07$, $p < 0.001$), but an equal number of participants rated Area and Count charts as more complex ($p = 1.00$). The majority of participants gave neutral ratings, suggesting that they perceived Area and Count charts to have similarly complexity ($p_{area/neutral} = 0.0054$, $p_{count/neutral} = 0.0023$) (1 ▂▇▇▇▂ 5).

Perceived Importance. Participant ratings of importance followed a similar trend to complexity: there was an overall difference between scores ($\chi^2 = 24.80$, $p < 0.001$), but participants were equally likely to rate the two chart versions as seeming more important ($p = 0.15$). Significantly more participants gave neutral ratings than chose Area charts as more important ($p = 0.00026$), but there was no significant difference between the number of neutral responses and participants who rated Count charts as more important ($p = 0.69$) (1 ▂▃▇▃▅ 5).

Familiarity. One-way ANOVA suggested that chart topic and perceived importance significantly predicted familiarity ratings. Perceived importance was positively correlated with familiarity, such that more familiar topics were rated as more important (or vice versa, as we cannot determine the direction of their relationship) ($Est = 0.13$, $SE = 0.062$, $p = 0.043$). Post-hoc analysis with Tukey's adjustment revealed that charts depicting COVID-19 related information was rated as more familiar (Mean Difference = 1.02).

Predicting Comprehension with Comparison Metrics. Because viewers spent a considerable amount of time deciphering the infographics in Phase 1, by the time

the participants saw both versions of the infographic in Phase 2, they may have processed the chart seen before more 'fluently' – requiring less time to visually dissect and comprehend it [312]. This processing fluency may impact judgment, such as preference or trustworthiness ratings [311, 313]. Therefore, we investigated whether the chart seen in Phase 1 impacted ratings on the given metrics in Phase 2.

We constructed 8 mixed-effect linear models, where each predicted 1 of the 8 metrics evaluated in Phase 2 and controlled for the 7 other metrics. With these models, we aimed to establish if chart topic or the chart version seen in Phase 1 had any effect on the metrics from Phase 2. The chart version seen before did not significantly predict visual appeal ($\chi^2 = 0.44$, $p = 0.50$), readability ($\chi^2 = 0.75$, $p = 0.39$), understandability ($\chi^2 = 0.70$, $p = 0.40$), relatability ($\chi^2 = 0.98$, $p = 0.32$), importance ($\chi^2 = 1.86$, $p = 0.17$), or topic familiarity ($\chi^2 = 0.46$, $p = 0.50$). However viewers were trendingly more likely to rate the chart version they saw in Phase 1 as more cluttered ($\chi^2 = 3.45$, $p = 0.06$) and to rate the chart version they saw in Phase 1 as less complex ($\chi^2 = 6.74$, $p = 0.0094$).

We also examined the inverse, looking at whether the metrics measured in Phase 2 had predictive power on accuracy, insight, or mental effort in Phase 1. We conducted similar mixed-effect linear model as before, but with the 8 metrics measured in Phase 2 as additional fixed effects. Overall, none of the metrics in Phase 2 predicted Phase 1 accuracy (measured by performance in Question 1, 3, 4), insight, mental effort, or response length (measured by coder ratings for Question 2, 5, 6). Statistical details can be found in the supplementary materials.

4.5.6 Experiment 1: Discussion

In summation, we observed that Count and Area charts did not differ in their ability to elicit higher levels of insight. Participants were not more likely to answer questions correctly across the two chart versions, nor did they show more signs of

engagement with the content on metrics of mental effort or response length. We observed that participants were more likely to misinterpret Count charts than Area charts, but because of the small number of misinterpretations overall, this effect may be overstated and explainable by one commonly misinterpreted chart. Additionally, we observed several interesting trends among the results of Phase 2, including polarized views on visual appeal and no observed effect of chart version on relatability.

However, there was a correlation between some of the metrics assessed in Phase 2 and the chart version participants saw previously in Phase 1. We suspect that either perceptual fluency was a factor or that the questions were not clear to the participants. Further, analysis of multicollinearity using variance inflation factors (VIFs) (see Table 4.4) suggested that many of our metrics were not independent, as evidenced by several high VIFs of 2.5 and above. Stated differently, the set of metrics interacted with each other and, therefore, unsatisfactorily measured viewers' perceptions. When moving to explore the metrics from Phase 2 more deliberately, we made several changes. Namely, we removed redundancy within the set by removing questions about Readability and Complexity. Additionally, we rephrased several of the questions to increase clarity (see Table 4.5 for the updated questions).

4.6 Experiment 2: Personal Experience

Experiment 2 aimed to better understand how participants experienced the infographics and tease apart the factors that made a difference.

4.6.1 Experiment 2: Stimuli and Methods

In this experiment, we utilized the same stimuli as the previous experiment but iterated upon the Comparison phase to see what other insights could be revealed. We collected data from 60 participants via Amazon's Mechanical Turk ($mean_{age} = 39.74$, $SD_{age} = 12.98$, 23 women, 27 men, 10 other). As in Phase 2 of Experiment 1,

participants were shown two versions of the same chart and asked to indicate which of the two charts best satisfied the prompt (see Table 4.5 for prompts). After each rating, with the exception of the questions about Familiarity and Interest, participants were asked to explain their reasoning. Chart order was determined with the Graeco Latin Square order from the previous experiment, but, critically, participants had not seen either chart previously. Participants rated every metric on a separate page and all metrics were presented in the same order throughout.

Metric	**Question**	**Themes (Number of associated responses)**
Visual Appeal	Which of these charts is more visually appealing to you?	Easy to understand (54), Easy to see (34), Icon selection (29), Simplicity (24)
Understandability	Which of these charts takes more time to understand?	Difficulty understanding value (54), More to look at (49), Difficulty focusing (33), Difficulty identifying topic (27)
Envisioning	Which of these charts makes it easier for you to envision what is happening, by relating the data to real world objects, situations, or entities?	Makes topic real (83), Straightforward (47), Obvious topic (23)
Unnecessary Clutter	Do either of these charts feel needlessly cluttered?	No added meaning (29), Organization (30), Space (27)
Importance and Urgency	Which of these charts makes the topic seem more urgent or important to you?	Real-world connection (77), Color (40), Straightforward (28), Size (19), Effort/Professionalism (14)
Familiarity	How familiar are you with the topic of this chart?	None (Likert scale only)
Interest	How interested are you in the topic of this chart?	None (Likert scale only)

Table 4.5. In Exp. 2, participants compared 2 versions of a chart along 7 metrics and then explained their reasoning. These metrics are a subset of those used in Exp. 1 Phase 2, but most are worded differently. The top 3-5 themes for each question are listed with the number of responses they were assigned to. There was no justification requested for Familiarity or Interest, because these questions pertained to the chart topic only and did not compare chart versions.

4.6.2 Experiment 2: Analysis and Results

To analyze the responses, we utilized the Thematic Analysis method [41]. Two of the authors familiarized themselves with the responses and then independently generated a set of preliminary codes for separate questions based on common topics mentioned. Additional codes were generated and applied iteratively until each author could not come up with more codes, at which point the codes were grouped into

broader themes. The 3 to 5 most frequently applied themes for each question are included in Table 4.5. We now present the numeric and qualitative results for each dimension judged by participants.

Visual Appeal. Participant opinion on visual appeal was polarized ($\chi^2 = 54.43$, $p < 0.001$). A Chi-squared test revealed that there was an equal number of participants that rated Count and Area as more visually appealing ($p_{area/count} = 0.82$). A post-hoc analysis with Bonferroni corrections [171] showed significantly fewer people gave a neutral rating (1 ▮▂▁▂▮ 5). Detailed pair-wise comparison p-values can be found in the supplementary materials. We observed a positive correlation between perceived visual appeal and importance ($Est = 0.23$, $SE = 0.051$, $p < 0.001$), such that the chart that was perceived to be more visually appealing was also perceived to make its topic seem more important.

As shown in Table 4.5, the most common themes identified for this question were "easy to understand," "easy to see," "icon selection," and "simplicity." This is particularly interesting because although this question was asking about visual appeal, the most common themes we found are actually about ease of use. This suggests that for many participants, visual appeal and understanding cannot be separated – a chart which is easy to use is also visually appealing.

Participants who preferred Count charts appreciated how quickly they could grasp the topic of the chart and mentioned how the individual pictographs allowed them to see and compare quantities more easily. As one participant put it, *"Understanding the [Count] chart is immediate[.] I see the drawings of the different kind of animals and I understand immediately the quantities. I don't need to read words [to] understand the issue."* Other participants struggled to explain why they liked the Count charts better, but attributed it to the pictographs, for example saying: *"Neither is very easy to read but the animal shapes are nicer to look at."*

Proponents of Area charts argued that the simplicity or clarity of this style made the charts easier to understand. One participant expressed this opinion as *"[The Area chart] has a much simpler design and in turn offers a clearer and more understandable message."* Here, the participant is very clearly describing the implicit relationship between understanding and visual appeal for them. Beyond simplicity, some other participants (11%) cited precision as factor as in: *"Although the [pictographs] are visually appealing I like the precision of the donut chart since the percentages aren't exactly equal."*

Understandability. Overall, participants rated the Count charts as taking more time to understand ($\chi^2 = 50.23$, $p < 0.001$). Post-hoc analysis with Bonferroni corrections revealed that the majority of viewers gave neutral ratings or rated Count charts as taking longer to understand (1 ▂▃▅▆▇ 5). We also observed a negative correlation between perceived time to understand and importance ($Est = -0.11$, $SE = 0.054$, $p = 0.049$), such that the chart that seemed to take longer to understand was perceived as seeming less important. Detailed pair-wise comparisons can be found in the supplementary materials.

Although participants rated the Count charts as requiring more time to understand, an analysis of the amount of time it took for participants to answer each question in Experiment 1 does not support this belief. A mixed-effect model predicting response time for all 6 questions in Experiment 1 revealed that question ($\chi^2 = 1.63$, $p = 0.44$), chart topic ($\chi^2 = 2.35$, $p = 0.80$), nor chart version ($\chi^2 = 3.59$, $p = 0.058$), seem to be significant predictors of how long a participant spent on the chart. We speculate as to why this mismatch occurred in the Discussion.

The most common themes that emerged were "difficulty understanding value," "more to look at," "difficulty focusing," and "difficulty identifying topic" (see Table 4.5). These themes may give us a window into what kind of information the participants were using as a proxy for the amount of time it would take for them to

understand the chart or the kinds of things they imagine would be a hindrance to understanding.

Within responses that described Count charts as requiring more time to understand, many participants (55%) focused on the number of items in the image. Some offered that Count charts required more time purely because there was more to look at. Others viewed the number of items more negatively, describing it as "distracting" and a hindrance to finding critical information. Another point of disagreement was whether the possibility of counting the pictographs was helpful or harmful. Some participants thought seeing the individual pictographs was helpful for understanding ratios quickly, while others preferred the solid blocks of the Area charts. For some, this preference came down to familiarity. As one participant wrote *"[The count chart] is a new form of chart for me so I really need to put some effort in order to read it properly."*

Participants who thought that the Area chart required more time to understand cited the need to read more text to figure out the topic of the chart. As one participant put it *"The bars are more abstract and therefore [require] more time to understand."* Ironically, a similar sentiment was expressed by other participants as a reason that Count charts required more time to read, arguing that the viewer had to figure out what each of the symbols represented.

Envisioning. Overall, when asked about which chart best made it easier to envision what was happening and to relate the data to real world objects, situations, or entities, viewers perceived Count and Area charts differently ($\chi^2 = 29.03$, $p < 0.001$). In particular, significantly more participants rated the Count chart as easier to envision ($p < 0.01$)

(1 ▪▪▪▪▮ 5). There was also a positive correlation between perceived easiness to envision and perceived importance ($Est = 0.39$, $SE = 0.052$, $p < 0.001$), such that

the chart that was perceived to be easier to envision also made its topic seem more important.

The top themes identified in responses to this question were "makes the topic real," "straightforward," and "obvious chart topic" (see Table 4.5). Though it's unsurprising that participants talked about things that made a topic seem real, the other two popular themes suggest a more surprising potential relationship between relating and ease or speed of understanding. Proponents of Count charts expressed that the pictographs made the topic seem more "real" and encouraged them to relate it to their own experience. As one participant wrote, *"The human icons emphasizes [sic] that COVID-19 is affecting people just like me all over the world."* In addition, participants contended that the pictographs made the topic and message of the chart obvious, especially at a glance. For example, about the chart about social media mentions, one participant wrote *"When I looked at [the count chart] I knew without even looking at the top that it was about Twitter as I saw the Twitter logo. At first glance I don't know what [the area chart] is referring to."* While the bird pictographs were effective for signaling the chart topic to some participants, others found it confusing. As one participant described it *"The birds are somewhat misleading as I think of actual birds instead of Twitter."* Another participant said that they found the Area chart easier to envision *"Because I don't know what birds have to do with the coronavirus."*

Participants who favored Area charts found the straightforwardness of the charts to assist in envisioning the topic. As one participant wrote *"[The Area chart] is using a good old traditional bar chart which is easy to read at a glance and therefore it is easy to see what is actually happening without over thinking it."* Those who preferred this style also found the presentation of the numerical information to be easier to grasp. For example, writing *"[The area chart] doesn't need me to do my math meaning that I can instantly feel on my skin what the chart wants to tell me."* In contrast, others found interpreting the proportions easier with the individual pictographs. As one

wrote *"I can easily visualize the statistics because the objects represent them at a glance."*

Unnecessary Clutter. Count charts were significantly more often rated as unnecessarily cluttered, compared to Area charts ($\chi^2 = 227.4$, $p < 0.001$). However, the number of people who gave neutral ratings is significantly higher than all the other categories (1 ▁▁▆▃▃ 5). Detailed pair-wise comparisons can be found in the supplementary materials.

There is a negative, but not statistically significant, correlation between perceived visual clutter and importance ($Est = 0.035$, $SE = 0.078$, $p = 0.75$) such that the chart that is perceived to be less cluttered is perceived to make its topic seem more important. The reason this correlation was not statistically significant was due to that fact that very few people rated the Area chart as cluttered.

As Table 4.5 shows, the main themes for this question were "no added meaning," "organization," and "space." Within these themes, we see both a possible response to the word "needlessly" that was used in the prompt, as well as an idea of what characteristics participants were using to decide what was or was not cluttered.

For some participants, the Count charts felt needlessly cluttered because the pictographs contributed no additional meaning. Exemplifying this idea was this response from one participant *"It doesn't seem like there's much point in using the moneybags when [the area chart] works just as well. I'd rather just have [the area chart] if I were reading this data on a website or document."* Additionally, participants were critical of the number of pictographs and their organization, describing the page as "busy" and "complicated." As one participant put it, *"I can get what the person who made this chart was trying to express that all different types of people get COVID-19 and it effects them in different ways but the presentation is jumbled and distracting."* Some participants who answered neutrally didn't consider either of the the charts cluttered. One reason for this was the layout, such as expressed here: *"Both charts*

use the space provided well enough and do not feel overwhelming." Others did find the Count charts cluttered, but not "needlessly" so. For example, one participant wrote *"I wouldn't say either is NEEDLESSLY cluttered. [The Count chart] seems a little more cluttered since it's representing each person but I don't think it's needless. It definitely serves a purpose."*

Importance and Urgency. There was an equal number of participants that rated Area and Count charts as more urgent/important ($p = 1.00$). A Chi-squared test for given probabilities revealed an overall difference in importance/urgency ratings ($\chi^2 = 30.578$, $p < 0.001$). However, as established with Post-hoc analysis with Bonferroni adjustments, significantly more participants gave neutral ratings (1 ▪▪▮▪▪ 5).

The most common themes present in responses to this question were "real world connection," "color," and "straightforward" (see Table 4.5). These themes suggest both an impact of realism and ease of understanding on perceptions of importance and urgency.

Participants who thought Area charts made their topic seem more urgent and important reasoned that the straightforward style of the chart enabled them to immediately get the gist of the data. Further, they criticized icon charts as too playful and cartoonish for important topics. For example, one participant said *"it looks like [the Area charts] are trying to give you straight facts not make it fun with little pictures."*

In contrast, viewers who preferred the Count chart expressed that the relatability of the symbols connected it to real-world objects and made it seem more important, for example, saying *"Seeing the guns makes it less about numbers and more about actual guns."* Some explained that the number of individual items helped the graphic seem more impactful. For example, one participant summarized, *"The sheer amount of red animals should raise a red flag in the reader's mind."*

Among those that felt that the chart versions made the topic seem equally important, there were several different factors cited. Some participants stressed that the design didn't have much to do with their perception of urgency or importance; instead mentioning that it was the topic that mattered. This sentiment appeared both in reference to topics perceived as unimportant, as in: *"Number of tweets doesn't seem like a particular[ly] urgent or important topic so neither stands out that way to me."* and in reference to topics perceived as critical: *"Both talk about a serious topic. It is serious and urgent regardless of how the charts are displayed in this case."*

Just under half of all participants (47%) referenced the influence of color and size on their perceptions of importance. Red and black were cited as contributing the feeling of importance, especially when paired with large areas, as in: *"The giant red block screams 'urgency' when I see it."* Additionally, the accessibility of the numeric data was also cited as a factor of perceived importance. For some, the ease of seeing the proportions in the data, assisted by the individual pictographs, made the message seem more important. Others disagreed, arguing that the precision of the Area charts was better.

Finally, important-seeming designs were also praised for perceived professionalism and effort. Participants that preferred Area charts argued that this style looked more professional and therefore more important. This opinion was not shared by those who preferred Count charts, who expressed that Count charts looked like more effort had been put into them and therefore looked more important. Of the Extinction chart, one participant wrote: *"[The Area chart] just feels so lazy and underdeveloped that it is hard [for] to me read it, much less care. [The Count chart] at least feels like there is some passion there."*

Familiarity. Familiarity is a different metric than the others because it measures viewer attitude towards the topic rather than the chart. We conducted a one-way ANOVA comparing familiarity for each topic which suggested differing familiarity

	Visual Appeal	Understandability	Envision	Clutter	Importance	Familiarity	Interest	VIF
Visual Appeal	1.00							2.18
Understandability	-0.48	1.00						1.45
Envision	0.59	-0.46	1.00					2.07
Clutter	-0.59	0.42	-0.50	1.00				1.67
Importance	0.57	-0.42	0.63	-0.40	1.00			1.85
Familiarity	0.076	-0.055	-0.008	-0.054	0.029	1.00		2.00
Interest	0.16	-0.019	0.17	-0.016	-0.094	-0.051	1.00	2.07

Table 4.6. Correlation table and VIF for metrics from Experiment 2. Note that larger values are darker, regardless of sign. No two metrics appear to be strongly correlated. Overall VIFs decreased compared to those in Experiment 1, suggesting that the metrics used in Experiment 2 more orthogonally capture different participant attitudes and thus were better metrics to use in this type of work.

across the topics represented ($F = 19.63$, $p < 0.001$). Post-hoc analysis with Tukey adjustment [228] suggests that participants were significantly more familiar with COVID-19 related topics than other topics ($MD = 1.28$, $p < 0.001$).

We further examined if topic familiarity significantly influenced participants' rating on the above metrics via linear regression models. Controlling for other metrics, familiarity does **not** significantly predict visual appeal ($Est = -0.082$, $SE = 0.068$, $p = 0.023$), perceived importance/urgency ($Est = -0.008$, $SE = 0.061$, $p = 0.90$), easiness to understand ($Est = -0.094$, $SE = 0.065$, $p = 0.15$), easiness to envision ($Est = 0.023$, $SE = 0.063$, $p = 0.71$), or unnecessary clutter ($Est = 0.0065$, $SE = 0.046$, $p = 0.89$). However, it **does** significantly predict interest in the topic, such that more familiar topics were rated as more interesting ($Est = 0.69$, $SE = 0.041$, $p < 0.001$).

Interest. Like familiarity, interest is another metric that measures viewer attitude towards the topic rather than the chart itself. One-way ANOVA comparing interest for each topic suggested that viewers were more interested in some topics than others ($F = 14.9$, $p < 0.001$). Post-hoc analysis with Tukey adjustment suggested that participants were significantly more interested in COVID-19 related topics than other topics ($MD = 1.03$, $p < 0.001$). Controlling for other metrics, interest in the topic did **not** significantly predict perceived topic importance/urgency ($Est = 0.021$, $SE =$

0.062, $p = 0.74$), easiness to understand ($Est = 0.096$, $SE = 0.66$, $p = 0.14$), or perceived unnecesary clutter ($Est = -0.026$, $SE = 0.046$, $p = 0.58$). However, it **did** significantly predict familiarity, such that more familiar topics were rated as more interesting ($Est = 0.71$, $SE = 0.042$, $p < 0.001$). It also significantly predicted visual appeal – charts depicting more interesting topics were also rated as more visually appealing ($Est = 0.20$, $SE = 0.068$, $p = 0.004$). Finally, it trendingly predicted easiness to envision such that charts depicting more interesting topics were rated as harder to envision ($Est = -0.11$, $SE = 0.063$, $p = 0.07$).

Correlation and VIF of Metrics Used. Analysis of multicollinearity using variance inflation factors (VIF) (see Table 4.6) suggested that although our rating metrics were not fully independent, there was only a small amount of variance inflation (most VIFs were around 2 or below). In other words, these metrics were satisfactory measures of viewer perceptions of visualization designs. In addition, both VIFs and the number of highly-correlated metrics decreased from Experiment 1, which demonstrates that the changes made to the metrics were effective.

4.7 Discussion and Future Work

This study revealed several important findings. First, we found no difference in the aspects of understanding demonstrated by viewers of infographics containing pictograph arrays and those containing solid areas. However, using pictograph arrays significantly influenced the participants' experience. In particular, we found a visual complexity-relatability trade-off. Infographics with pictograph arrays were thought to require more time to understand and were perceived as more often visually cluttered, but they enabled viewers to more easily relate the chart topic to the real world. Additionally, two distinct perspectives on the use of pictographs emerged in our experiment. For some, pictographs are preferable to abstract areas because they are visually appealing and make the topic more relatable. For others, charts with

pictographs are not preferred because they are seen as cluttered, complex, and less serious. We will now speculate as to why these results occurred and then discuss their potential impacts on the design of visualizations.

Bloom's taxonomy as an evaluation method

In this chapter, we used Bloom's taxonomy as a formal evaluation method for visualizations. Through this method, we were able to comprehensively probe 6 different aspects of understanding. Although we found no difference between the two chart versions we evaluated, the breadth of aspects covered and the real-world applicability of these aspects suggests that our inquiry was still valuable. Since conducting these experiments, the authors have written a more in-depth description of the method of using Bloom's taxonomy for evaluation [53].

The questions used in the experiments we conducted were designed to reflect the spirit of the levels of Bloom's taxonomy in the context of *casual* sensemaking. It is relatively straightforward to translate the aspects of understanding measured by Levels 1 to 4 to casual sensemaking, as they are largely similar to simple tasks completed in the original, educational context. But for Levels 5 and 6, the type of learning evaluated in an educational context differs from that of casual sensemaking. Specifically, Levels 5 and 6 are intended to evaluate understanding in cases where students could gather more information, ask questions, and take lots of time to produce responses. In contrast, in the casual sensemaking context, such as when someone sees a visualization in a news article or a tweet, that process typically does not involve seeking more information or asking questions, and typically unfolds across a few minutes at maximum. We translated Levels 5 and 6 in ways that we saw as most similar to the intent of the original taxonomy, but our translation between these mismatched contexts is only one of many possible translations and represents one limitation of this work.

We can imagine other translations of these levels to the causal sensemaking context and we *did* translate them differently in other related work [53]. In that work, participants predicted a value beyond the chart (Level 5) and generated and justified a conclusion (Level 6). This is in contrast to this chapter, where we asked about the participant's community (Level 5) and what they thought a public official should do (Level 6). Future work that uses this method may find yet new ways of translating these upper levels across contexts.

Future iteration of this line of work could also extend the list of example tasks in this paradigm to evoke specific, detailed responses that help designers and researchers gain insights regarding how a visualization reader is reacting to a visualization. We see strong potential for this method to evolve into a useful technique supplementing in-person interviews in remote environments where in-depth interviews may be difficult or impossible. Future researchers could also combine this evaluation method with other measures of graphical, linguistic, or numerical literacy to generate a more comprehensive evaluation method, or use this method to identify concrete design guidelines. Alternatively, researchers could diversify the data analysis approaches to extend our method beyond just evaluating affordances to cover graphical literacy or numeracy. For example, one could compare the participants' trend predictions in the synthesis (prediction) task to a ground truth to determine the accuracy of their prediction, which could inform researchers about their numeracy skills (e.g., how well the participant could interpolate/extrapolate trends in data). Additionally, although in our paradigm, we only compared two alternative designs of the same chart, this framework is flexible enough to allow for single or multi-chart comparisons.

Why was there no effect on understanding?

In our experiments, replacing the geometric shapes of the plain charts with pictograph arrays changed the way that the information was encoded. It would be

reasonable to think that a change like this would have some effect on the understanding obtained by viewers, but we observed no such effect. There are several reasons why this might be.

First, it is possible that although one version uses solid shapes and the other individual icons, participants may process the images similarly – as masses of color [166]. Future work with more complex charts or with fewer pictographs may produce different results. Second, it is possible that while there was no effect for participants in aggregate, there were subsets of people who did experience some effect. For example, particular combinations of graph literacy and numeracy may have led to larger effects such as those observed in the medical risk literature (e.g., [138]). Similarly, some of the factors which have been observed to affect engagement such as subject matter, source, or self-efficacy [195] may also affect understanding. Future work could examine the possible effect of these factors and could help the community better understand what about an audience matters when trying to communicate facts and ideas. Third, it could be a feature of the infographics that we used in our experiments. The stimuli we chose to use were relatively simplistic as they only contained a few categories and under a few hundred points. It is possible that more complex charts might lead to larger effects. Future work with more complex infographics could help determine the veracity of this possibility.

It is also possible that this manipulation affected casual sensemaking, but not understanding, in ways that could not be detected by our measures, e.g., emotional aspects of sensemaking or sensemaking as an ongoing personal experience (e.g., as in the notion of sensemaking in [103]). Past work has shown that emotions have a strong impact on the way that data and visualizations are perceived and understood – it is not just the data themselves that are important, but how they feel [194]. Given the range of factors identified in Experiment 2, it may be true that though the

comprehension aspects of casual sensemaking (measured in Experiment 1) were not effected, the **meaning** made was.

Why did participants find pictographs easier to relate to?

Our results for Experiment 2 indicate that participants thought that charts with pictographs helped them better envision the chart topic. One possible explanation is that including pictographs related to the topic reduces the mental burden of relating abstract textual content to a visual depiction. This could be related to the factors thought to make concrete scales effective – by relating abstract depictions of data to something more familiar, concrete scales reduce the cognitive load of comprehending the underlying numerical values [72]. This explanation could still be consistent with our results from Experiment 1 that observed no difference in response accuracy between the pictograph and area conditions, as reducing cognitive load does not necessarily increase performance [178]. Instead, it could indicate that even the most complex questions did not overtax the participant's cognitive resources. Nonetheless, to investigate this conjecture, future work could incorporate subjective or psychological measures of mental effort and cognitive load (such as those reviewed in [273]).

Alternately, strong positive emotions could influence participant perceptions. Previous work on emotional design has shown that tools which are perceived as beautiful or attractive are also considered to work better [263]. Following this model, when a chart containing pictographs evoked positive emotions for participants, it may have helped them feel more willing to engage and ultimately produced a feeling of ease and reduced effort. It is worth noting, however, that a viewer may experience positive emotions in response to charts as a whole, even when the topic or pictographs are not happy or joyful. For example, in our study, sometimes the pictographs were described by participants as "scary," even in a chart that was otherwise noted to be aesthetically beautiful (e.g., in response to the guns chart in Figure 4.3).

It is also possible that the "realism" of a pictograph contributes to how effective they are at making the topic seem real. There is conflicting evidence supporting the effectiveness of anthropomorphized images of people on inducing empathy (such as in [40, 143]), but little is known about their effect in other realms such as those explored in this chapter. Future research may investigate the use of more realistic images in infographics.

Why did participants think that the pictograph arrays took more time to understand? (and why were they wrong?)

On the other hand, though participants reported that charts with pictographs were easier to envision, they also perceived them to require more time to understand. Though this was believed to be true by participants, our results from Experiment 1 do not support this – we found no significant difference in response time on any question with respect to the version of chart viewed.

Estimating the time it takes to complete a task is not straightforward. Therefore, instead of asking why participants thought the charts with pictograph arrays would take more time to understand, a better place to start may be to ask: What were participants using as a proxy for estimating the time? It is possible that participants were using something like visual complexity or the total number of items on the screen. Investigating exactly what factor is being used to estimate time could be a future direction in itself, but our results imply that the actual effect of this factor is smaller than people think and may not actually have an effect at all. Further, this could suggest that there is a difference between the features that people think contribute to their understanding and those that actually do.

Design considerations for the use of pictographs

Our results suggest three design considerations. First, if a designer is looking to make their topic easy to envision, they should consider using pictograph arrays in

place of geometric areas. Second, if a design contains pictographs and the designer is concerned that viewers could think the chart will take too long to understand, they may wish to consider the number of items present in the design and how easy it is to understand the value of each component. Third, to mitigate potential perceptions of unnecessary clutter in a design containing pictographs, the designer may wish to consider what additional meaning the pictographs contribute over a more traditional representation.

Further, our results suggest that when producing graphics which are not very data dense, pictographs are best when used in two ways: to help identify what the chart is about or when organized spatially into clusters that should be interpreted collectively. Critically, this means that pictographs should not be used to force the viewer to count [377]. Instead, designers should combine pictographs with textual labels containing the absolute values where they are important to avoid making viewers feel like they need to count in order to understand.

Though our results indicate that pictographs can be effective for helping the viewer envision the topic, finding representative pictographs for abstract concepts is not a trivial task. For example, while there was little confusion about the symbols we used to represent people, some participants found the Twitter bird to be helpful, while others found it confusing. Existing studies have shown that even when icons are specifically designed to cross language and cultural divides, they are often not understood as intended [422]. However, having a set of pictographs which is as inclusive as possible in terms of subject and multi-cultural clues can help. One set which we utilized in our redesigned chart about COVID-19 symptoms (see top row, second column of Figure 4.3) was WeePeople [302] which contains silhouettes of individuals of different genders and races, with and without mobility aids. While there are some excellent inclusive pictograph sets, such as the WeePeople set, there is still work to

be done so that designers can choose resonant pictographs based on the background of their anticipated audiences.

4.8 Conclusion

When designing a data visualization, infographic, map, or diagram for the general public, designers need to weigh trade-offs in visual complexity, relatability, and clarity of design. One decision to make is whether to show the data through bare geometric objects, such as familiar bars and lines found in conventional charts, or icon arrays, such as human silhouettes. We explored the effect of encoding information with pictograph arrays and more traditional solid areas in part-to-whole relationships as a case study, referencing Bloom's taxonomy to design comprehension tasks. We found that using familiar geometric objects and icon arrays have no significant impact on sensemaking activities, at least in the context of our study participants and set of visualizations that we tested. We found individual differences in design preference, but, overall, viewers considered infographics with pictograph arrays to require more time to read, but easier for envisioning the topic and associating it with real-world entities.

CHAPTER 5

FROM INVISIBLE TO VISIBLE: IMPACTS OF METADATA IN COMMUNICATIVE DATA VISUALIZATION

In Chapter 4, I explored how altering the design of a visualization may impact casual sensemaking. One reason that a visualization may be difficult to understand could be because of what is *not* present. For instance, communicative visualizations are often disseminated without "metadata:" information that is not directly represented in a visualization and 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. Although metadata are typically absent, their content nonetheless could impact the way that people understand or experience a visualization. To investigate the effects of metadata further, I first proposed a taxonomy of types of metadata and then conducted two experiments to investigate what kinds of metadata visualizations might want access to and then evaluate what impact having some of that information had on their understanding and experience of a set of four real-world visualizations. This chapter represents the work from two papers: [52, 51].

5.1 Introduction

Although designers and practitioners aim to create communicative visualizations which are easily understandable, relevant, and trustworthy, 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 [225] (e.g., logarithmic scales [318], truncated y-axes [88]). The same of lack of transparency may also impact how much readers trust the visualization (i.e., believe that the information communicated by visualizations is truthful and worth sharing with other people). 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 all together [187].

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, perceived transparency, and ultimately trust (e.g., [109, 183]). 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 [389]). The practice of disclosing metadata may also increase the transparency of the processes underlying a visualization, helping the public build trust [116, 176], and encouraging people to use visualizations [247]. 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*[52]. This definition of metadata is broad and intentionally encompasses

types of data that are typically considered metadata (e.g., about data collection or methods [90, 272]) 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 [173]) [52].

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 trust). We presented 64 participants from Prolific[301] with set of four random visualizations from five award-winning data journalism projects. For each visualization, participants were given four goals and selected the type of metadata they most wanted to have from a set of 18 types of metadata to accomplish each of them. 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 confidence that you understand the chart well) and metadata about the source of the data (Data Source) for assessing trust (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 transparency, trust, information relevance, and understanding. 144 participants from Prolific were shown pairs of visualizations, both with and without metadata. We collected open-response feedback related to the visualization's message and Likert-scales corresponding to dimensions of transparency (accuracy, clarity, completeness, and thoroughness) and 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. Our results suggest that people view visualizations with metadata as more thorough than those without, but found no

evidence that it contributes to perceptions of accuracy, clarity, completeness, or relevance. 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 trust, understand perspectives); 2) quantitative results suggesting that metadata can increase perceptions of thoroughness, which may positively impact the perceived transparency of a visualization; 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 the possible implications of our results on trust, understanding, and the tension between disclosing information and providing more textual information.

5.2 Background

5.2.1 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 [413, 48, 351, 307, 164, 21, 49], which can help establish the objectivity and context of research claims [34, 219]. This practice has also been argued to be beneficial for prospective users by helping them interpret data [351], judge whether it is appropriate to apply a dataset to a problem [141], and determine what kinds of conclusions can be responsibly inferred from data [135].

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 world view and the ways that they think about the social and political context of their research are commonplace made in social science [173, 128, 331]. 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 [163, 106, 224, 322]. 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 is highly related to what they will perceive as useful [283], believe to be too difficult to understand [195], and think others will see in the visualization [412]. Finally, feminist scholars have emphasized the importance of understanding the social and historical contexts of data to challenge hegemonic power structures (e.g., [106, 163, 141]).

5.2.2 Increasing Transparency & Trust

One of our primary motivations for this work is to examine the potential for metadata to influence and increase transparency and trust. Past research in visualization concluded that providing metadata like data provenance can be a tool for signalling transparency and trustworthiness to end-users [183, 109]. While increased transparency is often associated with increased trust (e.g., as in [254]), this relationship does not always hold. Instead, transparency increases trust to a critical point, at which increasing transparency further can counter-intuitively decrease trust — especially when expectations are broken [176, 204]. This erosion of trust is theorized to occur because too much explanation confuses readers and directs their attention toward unexpected outcomes [204].

Additionally, past work on transparency and trust in map-based visualizations found that elements of transparency from existing work (accuracy, clarity, disclosure, and thoroughness) did not predict participants' perceived level of trust [409, 333]. 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 [409]. 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 [293]. 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 [247, 372]. 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 [42, 400]. 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 [247, 193]. 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 [193, 89, 206, 230]. It is conceivable that the presence of metadata might influence trust for readers engaging in comprehensive processes (who want or need to look through all of the related information) and readers making quicker judgements 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 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 [416] and educational resources [132]. Notably, the disclosure of metadata was specifically named as a sign of trustworthiness for many participants [416]. Recent work also found that metadata which indicated a software was reputable and would perform well increased perceptions of trustworthiness and willingness to reuse the software [5].

The influence of metadata on credibility has also been explored in non-visualization domains. Credibility refers to the extent to which something is thought to be believable, trustworthy, accurate, and valid [396]. Past work in this space has observed, for example, that the presence of sources believed to be experts increased the perceived credibility of memes [396], tweets [231], and social media posts [358]. 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 more transparent).

5.2.3 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 [183]. Therefore, metadata

which 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 [393] or slide shows [389] 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 [160].

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 [162]. 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., [283, 183, 195, 202, 201]). 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 theses 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.

5.3 Types of Metadata

In this section, we discuss the potential pros and cons of two broad groups of metadata: metadata about the visualization pipeline from "raw" data to final visualization and metadata about the people behind the pipeline. We do not wish to suggest that there is a distinct separation between pipeline and people – there is, in fact, quite the opposite because data are necessarily the result of human decisions and processes. However, we decided to divide these two groups in order to focus on what kinds of information we could provide about each and because they offer different benefits and challenges.

We identified the two groups based on existing research including Gebru et al.'s "Datasheets for Datasets" [141] and Krause's notion of "Data Biographies" [213]. Gebru et al. [141] divided metadata into seven thematic categories based on important components of a dataset's lifecycle, where five categories are encompassed within our section about the visualization pipeline (Collection Process, Preprocessing/Cleaning/Labeling, Maintenance, Composition, and Uses) and the remaining two fit within our category about people (Motivation and Distribution). Krause [213] approached the categorization differently by creating five groupings based on the central questions of When, Where, Why, How, and Who. Of these categories, When, Where and How are covered by our section about the visualization pipeline and Who and Why fall within our section on people.

Previously, we discussed that providing any kind of truthful information increases transparency and trust up to a threshold, after which disclosing more may decrease trust. Therefore, in the remainder of this section, we will only mention this relationship when there are novel considerations.

5.3.1 Metadata about the Visualization Pipeline

We consider what metadata could be disclosed to the readers regarding three stages of the data visualization pipeline: data sourcing, data transformation, and data visualization. We primarily focus on the transformation and visualization components because existing work has extensively researched what should be collected and disseminated about the data sourcing step [141].

5.3.1.1 Data Source

Documenting the data collection process itself is a critical component of documenting a dataset and is covered in depth in [141]. However, describing which dataset was used in a visualization may still be an important piece of metadata to include. In fact, this practice is already an important part of the visualization style guides for some organizations (e.g., the Urban Institute [184]).

Pros: Providing the name of (or, ideally, a reference link to) the original dataset, may allow readers to replicate the visualization or conduct their own analyses on the data. While not all readers will have the time or expertise to do further analysis, they still might use this information to judge the usefulness of the visualization if they are able to discern its collectors. Evidence of using the source to judge usefulness was observed in past work, though the source was of the visualization, not of the underlying dataset [283].

Cons: However, it may be difficult to clearly describe the source of data without also disclosing who collected it, which may have negative impacts on privacy and trust if, for example, a dataset is associated with a particular individual or organization that readers already distrust.

5.3.1.2 Cleaning and Processing

However, disclosing *only* the dataset without also describing which transformations were applied to the data may lead to a loss of trust from data analysis-savvy readers who conduct their own analyses. This is because the readers could make different data analysis choices and come to different (or even contradictory) results [350]. Data processing is the middle-stage between the data obtained from the dataset and the visualization. It is important to consider what one could disclose about this stage because the methods chosen can introduce uncertainty into the data in ways which are not immediately apparent in the visualization [277, 183]. For example, the choice to keep or filter out outliers may change the mean and variation of the data. This, in turn, will change the visualization in ways which may not be apparent to the reader. Within this stage, one could also consider disclosing whether the data will be cleaned or processed again in the advent of updates to the underlying dataset. This kind of information might be particularly important in domains like public health where the recency of data is critical because there can be consequences for making a decision with out-of-date data.

Pros: Disclosing one's data cleaning and processing methods make the visualization more credible [244] and replicable, especially if the methods are sufficiently detailed such that another person could follow the steps to recreate the values represented in the final visualization. Further, this information may also allow readers to independently extract insights that can improve their own data cleaning methods in the future, as these insights "cannot be ensured by merely observing the results from data analytics" [76].

Cons: On the other hand, being transparent about one's methods of data cleaning and processing might open the creators up to undue or pedantic critique, especially if the creators belong to minoritized groups (e.g., because minoritized individuals are not always believed even when they are experts [220, 133]) . Furthermore, while

pressure to include how the data were processed might push some creators to select their methods more carefully if they know it will be public information, any diversion from the methodological norm might incur the wrath of the crowd and result in creator harassment or a loss of trust in the visualization – even if the methods are sound. Finally, it may be deceptively difficult to explain methodology in a way that is understandable for the general public, especially when analysis techniques are complicated or counter-intuitive [321]. Science communication techniques might provide guidance on this front, but explanation remains a challenge.

5.3.1.3 Visual Encoding: Explaining Perceptual Challenges

The final stage in the visualization pipeline is visualizing the data. Because this space is less explored in existing literature, we will discuss two different options for this stage.

One possible piece of information one could provide is an explanation of problems that the reader might face while trying to decode the marks, channels, or other design elements used in the visualization. For example, log scales have been shown to be difficult for the general public to understand [318]. Offering information that explains the log scale or simply points out its existence might help readers better understand the scale and the visualization as a whole.

Pros: Describing the potential difficulties of a visualization's encoding may help readers understand what problems exist so that they might try to avoid them. Existing work has shown that when flaws in visualizations are pointed out to readers, they are able to identify similar errors in other charts more often in the future [174]. This suggests that explaining what may go wrong with a visualization may also help readers learn over time what to look out for. Finally, it is possible that visualization creators will also learn how to create better designs over time by thinking critically about the negative impacts their design choices. Reflection is already an integral

part of generating insights for visualization design studies [250] and it stands to reason that design insights could result from asking designers to reflect on their designs more generally.

Cons: Nevertheless, when a reader is made aware of potential pitfalls in a visualization, it is possible that it becomes irredeemable in their opinion. The relationship between trust and use after a system is known to be flawed seems to not be very well understood in data visualization, but is an ongoing research direction in the related field of human-robot interaction where existing work suggests that the impact of known flaws on the user's behavior changes with the kind of task they are conducting (e.g., in [329]). Additionally, telling the reader that there are potential issues with a design begs the question: Why wasn't a different design choice made? Every design choice involves trade-offs that may be clear to the visualization creator but not to the reader. Finding ways to communicate what choice was made and why could bring the readers and creators onto the same page regarding the visualization design, but justification is particularly difficult when readers do not posess sufficient knowledge of the alternatives. Finally, explaining to the reader that they may make a mistake understanding an aspect of a visualization takes up space in the metadata that could be used to include other useful information and does not necessarily prevent the reader from making that mistake. For example, past work on truncated y-axes found that participants using visualizations that explicitly pointed out truncated axes made similar errors to participants using designs that did not point out the potential problem [88].

5.3.1.4 Visual Encoding: How to read the visualization

Finally, one could also consider providing information that helps the reader make sense of a visualization. Two possible ways of doing this are (1) directly telling the reader what the intended message of the visualization is and (2) providing a tutorial

describing how to "decode" the encoding. For example, when explaining a scatterplot, one could directly report that sugar intake and the number of cavities have a positive correlation or break down the encoding by describing what each point represents and what the axes mean without providing a specific conclusion that could or should be drawn. These two techniques have overlapping benefits but different challenges.

Pros: Directly providing information about what is being shown through either of the methods described may primarily benefit readers who are less experienced in decoding visualizations. Past work has shown that generating slideshows and "cheat sheets," which explained the encodings of complicated data visualizations, helped readers without a background in visualization successfully extract information from the charts [389, 393]. It is also possible that this type of information may motivate people to engage with and try to understand complex or unfamiliar looking visualizations by boosting self-efficacy, a factor known to impact engagement [195]. Finally, one possible benefit of providing a tutorial (which is not necessarily afforded by describing the intended message) is that readers may learn how to decode similar visualizations in the future.

Cons: However, there are also drawbacks to providing either the intended message or a tutorial. For instance, being explicit about the intended message might bias attention and interpretation, as well as change the reader's behavior. Pointing out a feature or conclusion to a visualization reader can change what they pay attention to [412]. For example, the content of captions can bias reader's attention toward specific points instead of other salient ones [199] and the content of titles have a significant impact on what the reader understands and remembers about a visualization [35]. Additionally, one may also consider that providing the intended message of a visualization may also change the way that the reader engages with the visualization. For example, readers might be less inclined to spend time exploring the visualization if they are explicitly told the key insight. At worst, this might result in a reader skip-

ping the visualization entirely, as has been experimentally observed with students [324].

Utilizing a tutorial about how to understand the visualization suffers from different challenges than describing the intended message because it tells the reader *how* to see something instead of *what* to see. While a lack of guidance might be desirable when exploration of the data is important, it is much less helpful for readers who are less experienced in drawing conclusions from what they are seeing or lack the time to sit with the visualization long enough to parse it.

5.3.2 Metadata about People

Metadata does not only refer to information about the data, but also includes information about the people involved in the process of visualization creation. We have identified two groups of people which have an impact on a visualization: the creators who make choices about what and how to visualize and the intended audience whose wants and needs inform what choices the creators make.

5.3.2.1 Creators

One could consider including information about the identities and lived experiences of the visualization creators. At its simplest, one could imagine that this could include pieces of information like an individual's name, job title, gender, or country where they live. It could also include information about the creator's motivation to create the visualization or anything that the authors deem relevant to convey. For example, depending on the topic of the visualization, one might consider including information about their political affiliation or their relationship to wealth, sexuality, or privilege. It may also be pertinent to think broadly about who qualifies as a "creator" of a visualization. For example, if a visualization was produced on behalf of an organization, one could also imagine disclosing the identities of the organization's leaders or its management structure, as they can reflect the priorities of that organization

and how much they value the perspectives of minoritized groups [106]. Additionally, information about the funding organizations may also help answer questions about who owns the visualization or how it can be used.

Pros: Sharing aspects of the authors' identities may increase transparency and lend credibility to both the individuals who created the visualization and the visualization itself [254]. This is consistent with the feedback-loops present in data provenance and trustworthiness theories where the trustworthiness of data is reflective of the trustworthiness of the provider and the trustworthiness of the provider is reflective of the trustworthiness of the data (e.g., [97]). Further, acknowledging that a visualization is ultimately constructed by people recognizes their unseen labor [105, 87] and breaks down what Donna Haraway calls the "god trick" — where a visualization appears to "[see] everything from nowhere" [163] and thus appears to represent objective, global truth rather than a situated viewpoint. Data products like data visualizations only "seem objective... because the perspectives of those who produce them... pass for the default" when really they represent one of many perspectives [106]. This also means that providing information about the identities of the creator(s) allows the reader to more clearly see who is involved in the creation of data visualizations and to uplift the voices and work of authors in minoritized groups.

Cons: On the other hand, if visualization readers feel that they have some reason to distrust the individual or organization who produced the visualization, they may distrust the visualization too. While in some situations this loss of trust might be considered warranted, it is also important to recognize that trust is socially constructed and that what/who is considered trustworthy is a reflection of existing power structures. This means that individuals in minoritized groups may not be trusted or believed even when they are experts – a phenomenon commonly referred to as testimonial injustice [220, 133]. It is not yet clear how disbelief of an individual or institution transfers to the visualizations that they create, though recent work on

visualization recommendations found that reader's trust was directly impacted by how capable they thought the creator was of making accurate visualizations [421]. Unfortunately, this suggests that individuals from minoritized groups may be disproportionately hurt by the disclosure of information about identity. However, there is evidence that readers make assumptions and judgements about the identities held by authors even when it is not disclosed (e.g., using the subject or content of a paper [203]), which suggests that witholding identity information is not an effective way to prevent the effect of harmful biases such as racism, xenophobia, or misogyny.

5.3.2.2 Intended Audience

In addition to the creators of a visualization, it may also be fruitful to consider *for whom* the visualization was made. One could imagine identifying the intended audience through demographic features (e.g., Georgia residents) or by specifying a group who use visualizations in a particular way (e.g., casual visualization readers).

Pros: One benefit of disclosing the intended audience of a visualization is formally recognizing that every design does not work equally well for every reader. By making the intended audience of a visualization explicit, readers might get a better understanding of why particular design choices were made. For example, explaining that a map of sea temperature was intended for use by museum visitors might help a scientist understand why a red-blue scale, which maps to customary ideas of hot and cold, was used instead of the rainbow scale commonly used by experts in the field (example from [290]).

Cons: Nonetheless, making the intended audience of a visualization explicit may also alienate some readers. For example, if a reader belongs to the intended group and does not understand what is being shown or otherwise does not find the visualization helpful, they may disengage or feel disillusioned. Alternately, if a reader does not belong to the intended group, they may feel as if they cannot or should not try to

understand what is being shown. It is possible that this could lead to disengaging entirely with a visualization. For example, people who felt that particular visualizations were too hard for them, but not too hard for others, were resistant to spend time with the visualization [195].

5.4 Experiment 1: Desirable Metadata

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 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 takeaways 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 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 5.1. We used six categories of metadata defined by the authors in prior work [52] (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, highlighted). Note: no examples for "Intended Audience" were collected during the survey and were instead generated by the researchers.

5.4.1 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 [344]. Sigma Awards are 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 [348], *Mapping Makoko* [346], *Land-Grab Universities* [345], *Rough Justice* [347], and *Who Gets to Breathe Clean Air in New Delhi?* [349]. 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

In Section 5.3, we defined six categories of metadata that could be provided to visualization readers (summaried in Column 1 of Table 5.1). 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 5.1.

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 5.1). 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.

[1]Three visualizations contained two types of graphs within the same image.

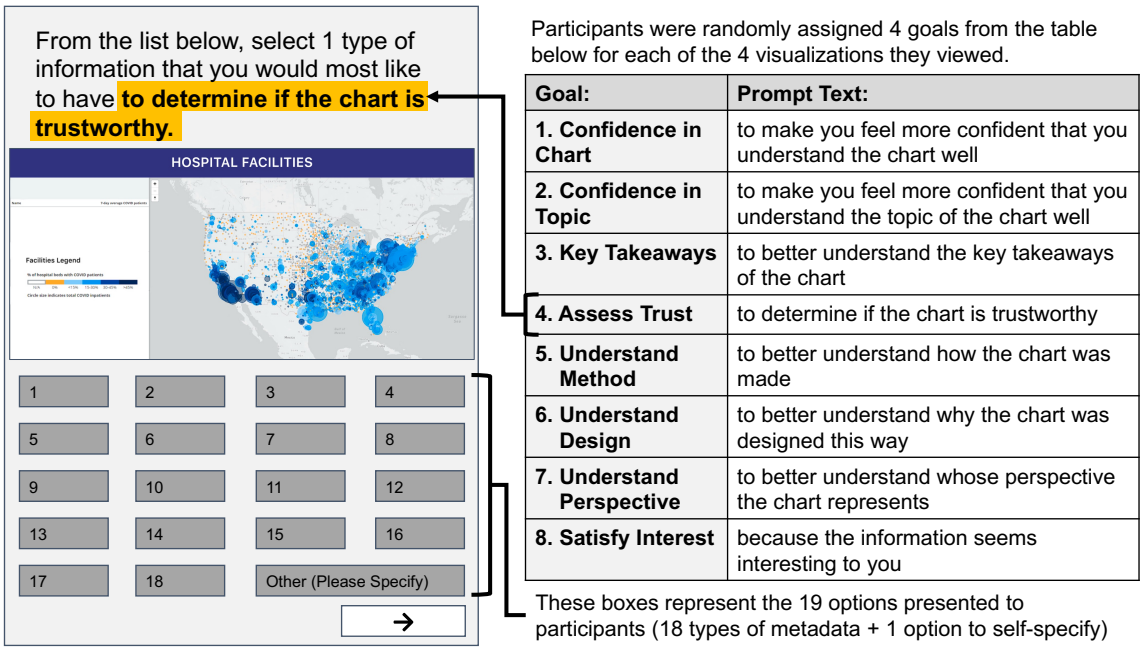
At least three examples of metadata were collected for all of the metadata categories except for "Intended Audience," of which we found no explicit examples. In lieu of "Intended Audience" metadata, the research team generated three examples that represent ways in which intended audiences are described 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 [261]).

Goals

There were a total of eight goals that participants could encounter during the study (listed in Figure 5.1). The set of eight goals was generated to represent different types of outcomes which have been theorized to result from metadata access (e.g., in [52]) or were shown in existing literature to be a reason that people sought additional information (e.g., as in [186]). 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."

Participants

We recruited 64 participants who self-identified as fluent in English from Prolific [301]. 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.



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 Trust	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

Figure 5.1. In Experiment 1, participants indicated which type of metadata they were most interested in having to accomplish a specified goal. Participants were randomly assigned 4 goals for each of the 4 visualizations they were shown.

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[305] 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 5.1 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.

		Categories of Metadata						
		Data Sources	Cleaning & Processing	Perceptual Challenges	Encoding Explanation	Creators	Intended Audience	Other
Goals	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 Trust	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)

Table 5.2. 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.

5.4.2 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 5.2 contains an overview of the distribution of participant responses for each goal. Therefore, in the following sub-sections, we explore which kinds of metadata were selected when participants were presented with each goals.

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 [171] to further 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 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 Trust

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 Trust**. 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 were 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 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 were 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.

5.4.3 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 Trust. 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.

5.5 Experiment 2: Impacts of Metadata

The previous experiment identified metadata categories that participants were interested in accessing. This experiment investigates the impacts of metadata on understanding and perceptions of relevance, transparency, trust, and persuasion.

5.5.1 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* [345] 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 5.2).

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 5.2 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 the 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 trust. There were no goals from Experiment 1 which were explicitly related to relevance.

Transparency

The disclosure of metadata is often cited as a means to communicate transparency (e.g., in [109]), but do visualization readers see it that way? Previous work concluded that accuracy, clarity, and completeness[2] were essential components of transparency [333] and that, in a visualization context, thoroughness captures the extent to which the data visualized represents the possible design space [409]. To examine perceived transparency, we asked participants to rate their agreement with four statements about the visualizations: accuracy, clarity, completeness, and thoroughness (see Ta-

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

ble 5.3 for the statements provided to participants and our definitions for each of the dimensions). We compared responses to these questions across participants to compare whether the presence of metadata had an effect on the perceived data transparency.

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 5.3. Participants rated each visualization by responding to seven 5-point Likert scales (1 - Strongly disagree to 5 - Strongly agree). Four scales assessed Subjective Transparency (Accuracy, Clarity, Completeness, Thoroughness) and three scales assessed Relevance (Meaningfulness, Relevance to Self, Relevance to Others).

Trust & Persuasion

In addition to measuring perceived transparency, we also wanted to capture trust more directly. Based on our goal to explore how disclosing metadata might influence readers' conceptions of trust, 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 policy makers. 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" trust and persuasion to select the visualization to ensure that participants used similar metrics to make judgements, rendering the predictions comparable between participants. Further, it mentioned both trust and persuasiveness because the two have been shown in past work to be closely linked (e.g., in [328]) and a visualization which is selected because it is trusted *and* persuasive may 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 [412]. 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. Existing literature defined the goal of journalism as making information more meaningful and relevant to readers [211], and thus we also measure perceived relevance of the visualizations. We wanted to know if disclosing metadata supported journalists' goal of explaining data in a way that is meaningful and relevant to readers [43]. We asked participants about three dimensions: Meaningfulness, Relevance to Self, and Relevance to Others (see Table 5.3 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 process (e.g., [54]). 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 is described in depth in Section 5.5.6.

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 5.5.7.

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 a 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.

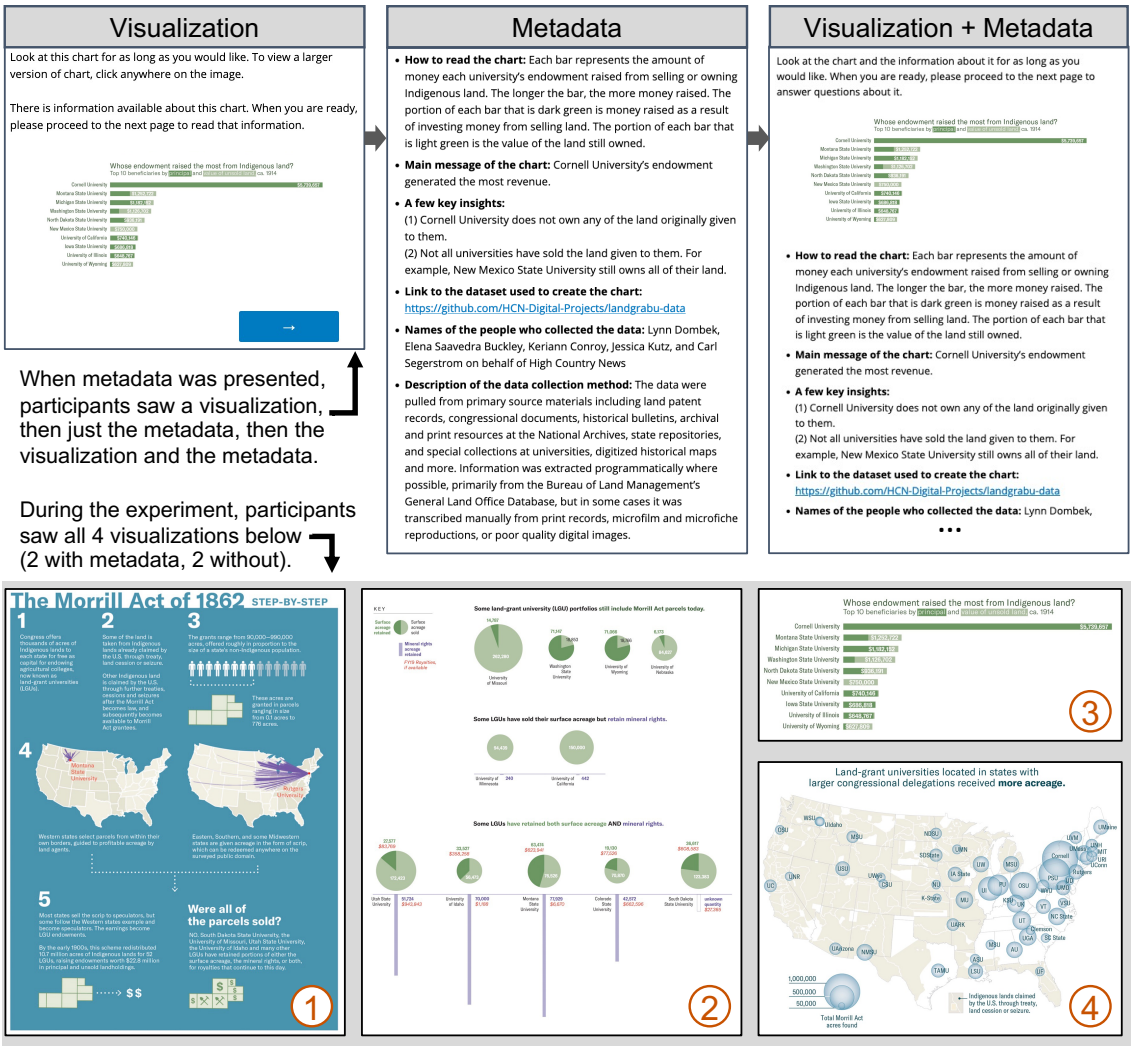


Figure 5.2. When participants were shown metadata, they were presented with a screen containing just the visualization, a screen with only metadata, and a screen with both the visualization and metadata. Participants saw four visualizations from "Land-Grab Universities" [345] including an infographic 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).

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 which 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 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 were 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.

5.5.2 Approach to Quantitative Analysis

In Experiment 2, participants rated each visualization on four dimensions of Transparency and three dimensions of Relevance. For both Transparency and Relevance,

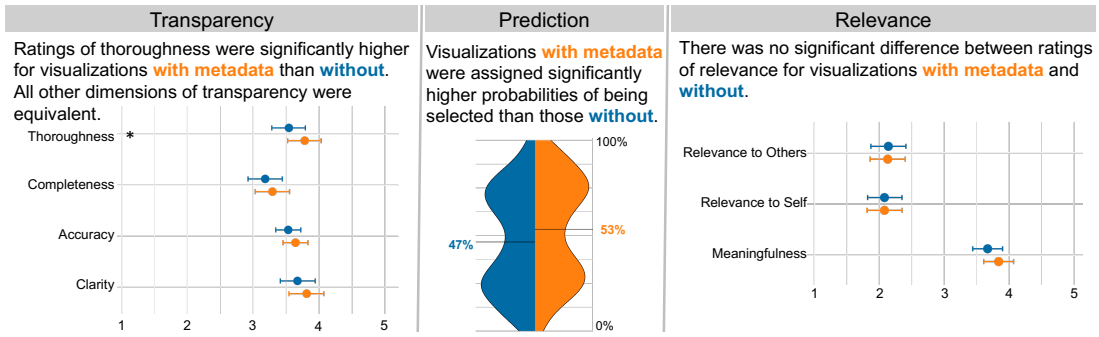


Figure 5.3. Our results from Exp. 2 suggest that visualizations with metadata were perceived as more thorough, but similarly clear, accurate, complete, and relevant. Participants also assigned higher probabilities that visualizations with metadata were chosen for a presentation to policy makers. Error bars of ratings depict 95% Confidence Intervals.

we analyzed our results using a multivariate analysis of variance (MANOVA), which is an analysis of variance (ANOVA) with two or more dependent variables [395]. 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 [395]. In each MANOVA, we considered the effects of the presence of metadata, chart type, order in which the charts were shown, 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 [126, 395]. 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 [395] (higher values suggest that the null hypothesis should be rejected, moderated by the degrees of freedom). 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 [20] to correct p-values for multiple comparisons.

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 strong negative correlation between two variables, close to 0 suggest no correlation, and close to 1 suggest strong positive correlation [371]. 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 between values [373].

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.

5.5.3 Results: Subjective Transparency Scales

In our second experiment, we measured transparency subjectively by asking participants to rate each visualization on a set of four dimensions: accuracy, clarity, completeness, and thoroughness. Our results indicated that the dimensions of transparency were impacted differently by the presence of metadata, chart type, importance of the topic, and how frequently participants used charts. Our MANOVA analysis revealed a trending effect of the presence of metadata (Pillai's value = 0.02) and a significant effect of the chart type (Pillai's value = 0.04), importance of the

topic (Pillai's value = 0.06), and frequency that participants used charts (Pillai's value = 0.06). 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 of transparency which was significantly impacted by the presence of metadata was thoroughness ($F(1) = 7.29, p < 0.01$). Post-hoc analysis with Tukey's adjustment suggested that visualizations with metadata were considered to be significantly more Thorough than those without ($Est = 0.24, p = 0.002$).

Effect of Chart Type: Clarity was the only dimension of transparency significantly impacted by the chart type ($F(1) = 15.93, p < 0.001$) such that the pie chart visualization was perceived as significantly *less* clear than the other three visualizations.

Effect of Topic Importance: Both thoroughness ($F(5) = 2.44, p = 0.03$) and accuracy scores ($F(5) = 2.73, p = 0.02$) 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: Accuracy scores were the only dimension of transparency significantly impacted by the frequency participants used charts ($F(5) = 2.56, p = 0.03$).

Collinearity: The four dimensions of transparency that we collected were distinct, but not entirely independent (see Table 5.4). 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 5.4).

5.5.4 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 policy makers.

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

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 ($\chi^2 = 8.06$, $p = 0.004$). 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 Transparency: Of the four dimensions of transparency, our model suggested a significant effect of Clarity ($\chi^2 = 13.53$, $p < 0.001$), but not of Accuracy ($\chi^2 = 0.12$, $p = 0.72$), Completeness ($\chi^2 = 0.26$, $p = 0.61$), or Thoroughness ($\chi^2 = 1.48$, $p = 0.22$). These results suggest that the probabilities that were assigned by participants may have been informed, in part, by participants' perceptions of Clarity.

We observed a significant effect on the predictions from the chart ($\chi^2 = 40.70$, $p < 0.001$). 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 also assigned lower Clarity scores.

5.5.5 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, 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. 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 5.5). 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 5.5. The three measures of relevance were not entirely independent. Relevance to self and relevance to others were highly correlated.

5.5.6 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 correct 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 [19]). 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 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 ($\chi^2 = 2.69$, $p = 0.44$).

5.5.7 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 5.6) using an iterative method similar to the one used in [241]. 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 Supplemental 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 5.6. 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 are not surprising, and instead reflect 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%, $n = 12$). A Chi-squared test of independence comparing the number of times that a specific university was mentioned (all Cornell) confirmed that the difference was significant ($\chi^2 = 4.00$, $p = 0.04$). 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 5.2 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, a one participant concluded that: "*...The chart showed that the amount of land was not equal and Eastern universities got more on average.*" A Chi-squared test of the frequencies that comparisons were mentioned for this chart suggests that participants mentioned comparisons significantly more often ($\chi^2 = 5.12$, $p = 0.02$).

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).

5.6 Discussion & Future Work

Our two experiments showed the following main results:

- Among four dimensions of transparency, 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.
- 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.

5.6.1 How does metadata impact trust?

Our motivation for this study was to investigate whether the disclosure of metadata alongside visualizations would impact the way that visualization readers perceive transparency and therefore trust. Our results suggested that participants rated visualizations with metadata higher on one of the dimensions of transparency (thoroughness), but we observed no evidence of an effect of metadata on the other three dimensions (accuracy, clarity, completeness). Further, participants predicted that it was significantly more likely that an organization would select the visualization with metadata for a presentation. These predictions were, in turn, impacted by participants' perceptions of clarity. These results suggest that metadata may impact perceptions of thoroughness and clarity, which, in turn, positively impact transparency. But how well does this transparency translate to trust? Past work that investigated the relationship between trust and same dimensions of transparency that we employed

found that neither thoroughness nor clarity were a good predictor of participants' reported levels of trust [409]. Additionally, while prediction scores were higher for visualizations with metadata, our experimental design did not disambiguate whether participants were more motivated to select the visualizations because of trust or persuasion. Therefore, we must conclude that our results show weak support for the idea that metadata improve readers' trust. However, studying trust and the mechanisms behind building trust are difficult [247]. We hope that the visualization community will see this as a call to investigate the effects of metadata further and to explore additional ways of measuring transparency and trust. 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 [274]. 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 [325]. While uncertainty visualizations may use visual encoding to communicate the quantifiable uncertainty within the data [274], the practice of disclosing metadata may be able to elucidate the *source* of uncertainty or aspects of uncertainty which cannot be quantified easily. On the other hand, it is yet unclear to whether visualization creators face the same apprehensions to disclosing metadata that they face to communicating uncertainty, such as a lack of resources and worry that disclosing uncertainty will impair understanding [182]. The visualizations that we used in our experiments did not directly visualize uncertainty, but future researchers may explore the ways that communicating metadata and visualizing uncertainty can support each other.

5.6.2 How does 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 which 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 [54]. Outside of visualization, researchers have measured perceived relevance by measuring whether participants choose to engage in particular behaviors (e.g., signing a petition [238]). Future work may therefore draw inspiration from other fields and explore the space of perceived relevance further.

5.6.3 How does 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 [35, 206]. 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 [206]. The ability of textual elements to bias what participants remember

as the main message is consistent with our results regarding the topics mentioned in responses by participants. It is therefore possible that participants utilized the 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. One possible direction might be to explore the impact of perceived misalignment between the information in the visualization and metadata, as has been explored with other elements of the textual layer (e.g., [206]).

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 [35]). 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 [306, 37]. 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 [19, 206]), 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 [53, 1].

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 regards 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 past work on the impact of pictographs [54]).

5.6.4 Tension between providing text information 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 [364]. By definition, visualizations are intended to provide a *visual* representation of data to help readers complete tasks [255]. However, despite this intention, textual information has proven to be a successful component of narrative visualizations [339], better for extracting critical information [269], and can be added to existing visualizations in order to add context, heighten empathy, and introduce temporal references [124]. 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 [59]. However, visualizing textual data is also well understood to be very difficult (see [7] for extensive discussion of the challenges including high dimensionality and irregularity) and some categories of metadata resist quan-

tification (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 [18], or cheat sheets [393] 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 [198]. 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.

5.6.5 Investigating the Impact of Animation & Interactivity

While experiment 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 [303] or help readers better understand unfamiliar visualizations [323]. 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., [340]). 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 which, 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 an 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.

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 which "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 between 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 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.

5.7 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 and ultimately trust. 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 is relevant to specific goals (e.g., assess trust, 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 perceptions of transparency,

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.

CHAPTER 6

CONTRIBUTIONS AND FUTURE WORK

In this chapter, I will summarize the contributions of each of the chapters of my thesis. Then, I will discuss future directions which may build upon my thesis work.

6.1 Review of Thesis Contributions

6.1.1 The Audience Matters! A Critical Analysis of Audiences in Visualization

In Chapter 3, I conducted an analysis of how the words "novice," "non-expert," "layperson," and "general public" were defined in existing visualization papers as a means to better understand how audiences are defined. The key contributions were:

- **An analysis of how "novice," "non-expert," "layperson," and "general public" are defined in visualization research papers revealing that these audiences are rarely defined explicitly and definitions which are present rely on only one aspect of an individual's attitudes, experiences, or skills.**
- **A discussion of the unintended consequences of defining audiences implicitly and one-dimensionally.**
- **Suggestions for improving current practice in three areas which draw inspiration from fields such as critical race theory and feminist theory.** Namely, I suggest that: (1) make the audiences clearer for researchers and paper readers by carefully defining who belongs in an audience and the criteria used to select participants; (2) define audiences multi-dimensionally and consider

how intersections of identity/experience produce unique experiences; and (3) be careful to avoid recreating hierarchical power structures which value some kinds of knowledge over others.

6.1.2 Designing with Pictographs: How does Design Impact Casual Sensemaking?

In Chapter 4, I investigated what impact using pictograph arrays had on casual sensemaking when used in place of solid, abstract shapes. The key contributions were:

- **A novel framework that provides a systematic way to evaluate levels of understanding in a visualization based on Bloom's Taxonomy.** This framework can be used to generate a set of six comprehension questions which each assess a different level of understanding.
- **Empirical results that suggest that infographics with pictograph arrays are just as good as more traditional, geometric part-to-whole charts at helping people make sense of data.**
- **Empirical insights that show some participants view charts containing pictograph arrays as easier to envision, while others view them as unnecessarily cluttered and slower to understand.** Although participants thought that visualizations with pictographs would be slower, we observed no significant difference between the amount of time it took for participants with pictographs or solid areas to answer the questions in our study.
- **Qualitative results suggesting that perceptions of visual appeal may be impacted by ease of understanding, making the topic seem real may help readers envision the topic more easily, and feelings of urgency and importance may be influenced by real-world connections.**

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

Finally, in Chapter 5, I explored what kinds of metadata people want access to and how having it impacts their understanding and experience of the visualization. The key contributions were:

- **A comprehensive taxonomy of six types of metadata and discussion of the potential benefits and risks of disclosing each.** This taxonomy includes metadata about the people who influenced the visualization's creation (Creators, Intended Audience) as well as several stages of the visualization creation pipeline (Data Source, Cleaning & Processing, Perceptual Challenges, Visual Encoding).
- **Results from an exploratory study suggesting that what kinds of metadata people may perceive as most important varies according to their goals (e.g., assess trust, understand perspectives).** For instance, participants were most interested in metadata about the Data Source in order to assess the trustworthiness of the visualization.
- **Quantitative results suggesting that metadata can increase perceptions of thoroughness, which may positively impact the perceived transparency of a visualization.**
- **Quantitative results suggesting that the information in visualizations with metadata are not considered more relevant than in visualizations without.** Participants rated visualizations with and without metadata similarly on three measures of relevance (Relevance to Self, Relevance to Others, Meaningfulness).
- **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.**

- **A discussion the possible implications of the results on trust, understanding, and the tension between disclosing information and providing more textual information.**

6.2 Future Work

Ultimately, neither disclosing metadata nor using pictographs arrays noticeably altered the ways that participants understood the visualizations they saw. However, both interventions impacted readers' experience of the visualizations. This suggests that altering the ways that visualizations are designed and communicated could impact readers' casual sensemaking processes, though the complexity of this relationship remains to be seen. Future work may build upon the work of this thesis in several different directions, which I will now discuss.

6.2.1 Complex, Unusual, or Interactive Visualizations

In the experiments in my thesis, I used real-world visualizations as stimuli. However, all of the visualizations were static, used common visual encodings, and presented data with low dimensionality. These qualities may have meant that the visualizations were relatively easy for most participants to make sense of the visualizations with or without the interventions and, therefore, hid effects that could be present with complicated, interactive, animated, or unfamiliar charts. Future work may therefore wish to replicate this work with different visualizations to establish how our results generalize to different encodings.

6.2.2 Alternative Methods of Presentation

Alternatively, future work could explore different methods for providing information to visualization readers. In my thesis, I presented information to readers by adding text alongside the visualization (Chapter 5) or by using pictographs to alter the visualization's design (Chapter 4). Yet, there may be other ways to present in-

formation that might have alternative effects. For instance, future work may explore how pictographs or graphics could be used to communicate metadata such as data cleaning processes or to signal that particular types of information are available for view. Another potential direction may be to use interactivity to allow readers to show or hide the information that is available or to ask for particular information (e.g., using a chatbot or natural-language interface).

6.2.3 Specialized Audiences

In this work, I focused on visualizations for people in the general public which was defined quite broadly. As I discussed in Chapter 3, audiences that are defined *too* broadly may mask the unique needs of sub-groups. It is possible that by focusing on the broad group, I was unable to detect an effect experienced by some, but not all of the participants. Future work may investigate the impact of disclosing information on narrower audiences such as people who have never seen a particular visualization before or people who might face a social or economic consequence if a particular visualization is unexpectedly deceptive.

6.2.4 Attitudes & Barriers Toward Disclosure

One question that I did not explore in this thesis is how stakeholders (e.g., designers, researchers) feel about disclosing different types of information such as metadata or details about their audiences. This direction may build upon past work that investigated why visualization creators do not disclose uncertainty [182]. Similarly, it may be relevant to explore what tools or structures would encourage visualization designers and researchers to disclose information (or otherwise reduce the difficulty of doing so). For instance, future researchers might ask: what effect could a data visualization "code of ethics" that prioritizes transparency and accountability have on practice? What tools might make it easier to disclose information in a way that is accessible and useful to readers?

6.3 Concluding Thoughts

In my thesis, I explored three ways that changing current practice may result in communicative data visualizations that are more understandable by members of the general public. I began with the possibility of changing research methods and investigated current practices for defining the audiences of data visualizations and how those current practices may have unintended consequences. Next, I explored how changing the design of visualizations might change how they are understood and experienced by replacing abstract areas with pictograph arrays. Finally, I investigated how re-contextualizing a visualization with additional text altered how it was understood and perceived. There is still much to be learned about how to make visualizations that "work" for the people who use them and I hope that my work in this thesis will inspire future inquiry into this important topic.

BIBLIOGRAPHY

[1] Adar, Eytan, and Lee, Elsie. Communicative visualizations as a learning problem. *IEEE Trans. Vis. Comput. Graph. 27*, 2 (2020), 946–956.

[2] Ahmad, Jarryullah, Huynh, Elaine, and Chevalier, Fanny. When red means good, bad, or canada: Exploring people's reasoning for choosing color palettes. In *2021 IEEE Visualization Conference (VIS)* (CA, USA, 2021), IEEE, pp. 56–60.

[3] Ahn, J., Plaisant, C., and Shneiderman, B. A task taxonomy for network evolution analysis. *IEEE Trans. Vis. Comput. Graph. 20*, 3 (2014), 365–376.

[4] Ajani, Kiran, Lee, Elsie, Xiong, Cindy, Knaflic, Cole Nussbaumer, Kemper, William, and Franconeri, Steven. Declutter and focus: Empirically evaluating design guidelines for effective data communication. *IEEE Trans. Vis. Comput. Graph.* (2021).

[5] Alarcon, Gene M, Gibson, Anthony M, Walter, Charles, Gamble, Rose F, Ryan, Tyler J, Jessup, Sarah A, Boyd, Brian E, and Capiola, August. Trust perceptions of metadata in open-source software: The role of performance and reputation. *Systems 8*, 3 (2020), 28.

[6] Albers, Michael J. Multidimensional audience analysis for dynamic information. *Journal of Technical Writing and Communication 33*, 3 (2003), 263–279.

[7] Alharbi, Mohammad, and Laramee, Robert S. Sos textvis: An extended survey of surveys on text visualization. *Computers 8*, 1 (2019), 17.

[8] Allen, Benita R. *An analysis of student success predictors for technical college persistence*. PhD thesis, Edgewood College, 2014.

[9] Allor, Martin. Relocating the site of the audience. *Critical Studies in Mass Communication 5*, 3 (1988), 217–233.

[10] Amar, R. A., and Stasko, J. T. Knowledge precepts for design and evaluation of information visualizations. *IEEE Trans. Vis. Comput. Graph. 11*, 4 (2005), 432–442.

[11] Amrhein, C G. Searching for the elusive aggregation effect: Evidence from statistical simulations. *Environment and Planning A: Economy and Space 27*, 1 (1995), 105–119.

[12] Ancona, Deborah. Framing and acting in the unknown. *S. Snook, N. Nohria, & R. Khurana, The Handbook for Teaching Leadership* (2012), 3–19.

[13] Andrienko, Natalia, and Andrienko, Gennady. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach.* Springer Science & Business Media, 2006.

[14] Andry, Tiffany, Hurter, Christophe, Lambotte, François, Fastrez, Pierre, and Telea, Alexandru. Interpreting the effect of embellishment on chart visualizations. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2021), pp. 1–15.

[15] Angluin, Dana. Learning regular sets from queries and counterexamples. *Information and computation 75*, 2 (1987), 87–106.

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

[17] Arneson, Jessie B, and Offerdahl, Erika G. Visual literacy in Bloom: Using Bloom's taxonomy to support visual learning skills. *CBE—Life Sciences Education 17*, 1 (2018), ar7.

[18] Bach, Benjamin, Riche, Nathalie Henry, Carpendale, Sheelagh, and Pfister, Hanspeter. The emerging genre of data comics. *IEEE Computer Graphics and Applications 37*, 3 (2017), 6–13.

[19] Bateman, Scott, Mandryk, Regan L., Gutwin, Carl, Genest, Aaron, McDine, David, and Brooks, Christopher. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proc. of the SIGCHI Conf. on Human Factors in Comput. Systems* (New York, NY, USA, 2010), Association for Computing Machinery, p. 2573–2582.

[20] Bates, Douglas, Maechler, Martin, Bolker, Ben, Walker, Steven, Christensen, Rune Haubo Bojesen, Singmann, Henrik, Dai, Bin, Grothendieck, Gabor, Green, Peter, and Bolker, Maintainer Ben. Package 'lme4'. *Convergence 12*, 1 (2015), 2.

[21] Battle, Leilani, Angelini, Marco, Binnig, Carsten, Catarci, Tiziana, Eichmann, Philipp, Fekete, Jean-Daniel, Santucci, Giuseppe, Sedlmair, Michael, and Willett, Wesley. 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] Bauer, Greta R., and Scheim, Ayden I. Advancing quantitative intersectionality research methods: Intracategorical and intercategorical approaches to shared and differential constructs. *Social Science & Medicine 226* (2019), 260–262.

[23] Baumer, Eric P. S., and Brubaker, Jed R. Post-userism. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2017), ACM, p. 6291–6303.

[24] Baumer, Eric P.S., Jasim, Mahmood, Sarvghad, Ali, and Mahyar, Narges. Of course it's political! a critical inquiry into underemphasized dimensions in civic text visualization. To appear in the Proceedings of EuroVis 2022, 2022.

[25] Baumgärtner, Sebastian, Ebert, Achim, Deller, Matthias, and Agne, Stefan. 2d meets 3d: a human-centered interface for visual data exploration. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems* (NY, USA, 2007), ACM, pp. 2273–2278.

[26] Bazeley, Patricia. *Qualitative data analysis: Practical strategies.* Sage, 2013.

[27] Belnap, Nuel. On rigorous definitions. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 72*, 2/3 (1993), 115–146.

[28] Berger, Ron. Here's what's wrong with Bloom's taxonomy: A deeper learning perspective, 2018. `https://blogs.edweek.org/edweek/learning_deeply/2018/03/heres_whats_wrong_with_blooms_taxonomy_a_deeper_learning_perspective.html`.

[29] Bickford, David, Posa, Mary Rose C, Qie, Lan, Campos-Arceiz, Ahimsa, and Kudavidanage, Enoka P. Science communication for biodiversity conservation. *Biological Conservation 151*, 1 (2012), 74–76.

[30] Björk, Staffan, Holmquist, Lars Erik, Ljungstrand, Peter, and Redström, Johan. Powerview: structured access to integrated information on small screens. In *CHI'00 extended abstracts on Human factors in computing systems* (NY, USA, 2000), ACM, pp. 265–266.

[31] Blalock Jr, Hubert M. *Causal inferences in nonexperimental research.* UNC Press Books, USA, 2018.

[32] Blascheck, Tanja, Vermeulen, Lindsay MacDonald, Vermeulen, Jo, Perin, Charles, Willett, Wesley, Ertl, Thomas, and Carpendale, Sheelagh. Exploration strategies for discovery of interactivity in visualizations. *IEEE Trans. Vis. Comput. Graph. 25*, 2 (2018), 1407–1420.

[33] Bloom, Benjamin Samuel. *Taxonomy of educational objectives: The classification of educational goals.* Longman, 1956.

[34] Borenstein, Michael, Hedges, Larry V, Higgins, Julian PT, and Rothstein, Hannah R. *Introduction to meta-analysis.* John Wiley & Sons, 2021.

[35] Borkin, Michelle A, Bylinskii, Zoya, Kim, Nam Wook, Bainbridge, Constance May, Yeh, Chelsea S, Borkin, Daniel, Pfister, Hanspeter, and Oliva, Aude. Beyond memorability: Visualization recognition and recall. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2015), 519–528.

[36] Borkin, Michelle A, Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., and Pfister, H. What makes a visualization memorable? *IEEE Trans. Vis. Comput. Graph. 19*, 12 (2013), 2306–2315.

[37] Börner, Katy, Maltese, Adam, Balliet, Russell Nelson, and Heimlich, Joe. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Vis. 15*, 3 (2016), 198–213.

[38] Bostock, Michael, Ogievetsky, Vadim, and Heer, Jeffrey. D^3 data-driven documents. *IEEE Trans. Vis. Comput. Graph. 17*, 12 (2011), 2301–2309.

[39] Böttinger, Michael, Kostis, Helen-Nicole, Velez-Rojas, Maria, Rheingans, Penny, and Ynnerman, Anders. Reflections on visualization for broad audiences. In *Foundations of Data Visualization*. Springer, USA, 2020, pp. 297–305.

[40] Boy, Jeremy, Pandey, Anshul Vikram, Emerson, John, Satterthwaite, Margaret, Nov, Oded, and Bertini, Enrico. Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data? In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2017), Association for Computing Machinery, p. 5462–5474.

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

[42] Boyd Davis, Stephen, Vane, Olivia, and Kräutli, Florian. Can I believe what I see? data visualization and trust in the humanities. *Interdisciplinary Science Reviews 46*, 4 (2021), 522–546.

[43] Bradshaw, Paul. Data journalism. In *The Online Journalism Handbook*. Routledge, 2017, pp. 250–280.

[44] Brehmer, M., and Munzner, T. A multi-level typology of abstract visualization tasks. *IEEE Trans. Vis. Comput. Graph. 19*, 12 (2013), 2376–2385.

[45] Bremm, Sebastian, von Landesberger, Tatiana, Bernard, Jürgen, and Schreck, Tobias. Assisted descriptor selection based on visual comparative data analysis. *Computer Graphics Forum 30*, 3 (2011), 891–900.

[46] Brewer, Noel T, Richman, Alice R, DeFrank, Jessica T, Reyna, Valerie F, and Carey, Lisa A. Improving communication of breast cancer recurrence risk. *Breast cancer research and treatment 133*, 2 (2012), 553–561.

[47] Buja, Andreas, Cook, Dianne, and Swayne, Deborah F. Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics 5*, 1 (1996), 78–99.

[48] Buneman, Peter, Khanna, Sanjeev, and Wang-Chiew, Tan. Why and where: A characterization of data provenance. In *Int. Conf. on Database Theory* (2001), Springer, pp. 316–330.

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

[50] Burns, Alyxander, Lee, Christiana, Chawla, Ria, Peck, Evan, and Mahyar, Narges. The audience matters! A critical analysis of audiences in visualization, 2022. Under review.

[51] Burns, Alyxander, Lee, Christiana, On, Thai, Xiong, Cindy, Peck, Evan, and Mahyar, Narges. From invisible to visible: Impacts of metadata in communicative data visualization, 2022. Under review.

[52] Burns, Alyxander, On, Thai, Lee, Christiana, Shapiro, Rachel, Xiong, Cindy, and Mahyar, Narges. 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.

[53] Burns, Alyxander, Xiong, Cindy, Franconeri, Steven, Cairo, Alberto, and Mahyar, Narges. How to evaluate data visualizations across different levels of understanding. In *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Vis. (BELIV)* (2020), IEEE, pp. 19–28.

[54] Burns, Alyxander, Xiong, Cindy, Franconeri, Steven, Cairo, Alberto, and Mahyar, Narges. Designing with pictographs: Envision topics without sacrificing understanding. *IEEE Trans. Vis. Comput. Graph.* (2021).

[55] Butsch, Richard. *The citizen audience: Crowds, publics, and individuals*. Routledge, England, UK, 2008.

[56] Cabrera, Ángel Alexander, Epperson, Will, Hohman, Fred, Kahng, Minsuk, Morgenstern, Jamie, and Chau, Duen Horng. Fairvis: Visual analytics for discovering intersectional bias in machine learning. In *2019 IEEE Conf. on Visual Analytics Science and Technology (VAST)* (CA, USA, 2019), IEEE, pp. 46–56.

[57] Cairo, Alberto. *The Functional Art: An introduction to information graphics and visualization*. New Riders, 2012.

[58] Cairo, Alberto. *The Truthful Art: Data, Charts, and Maps for Communication*. New Riders, 2016.

[59] Cao, Nan, and Cui, Weiwei. *Overview of Text Visualization Techniques*. Atlantis Press, Paris, 2016, pp. 11–40.

[60] Carbado, Devon W, Crenshaw, Kimberlé Williams, Mays, Vickie M, and Tomlinson, Barbara. Intersectionality: Mapping the movements of a theory1. *Du Bois review: social science research on race 10*, 2 (2013), 303–312.

[61] Carter, Niambi M, and Pérez, Efrén O. Race and nation: How racial hierarchy shapes national attachments. *Political Psychology 37*, 4 (2016), 497–513.

[62] Castano, Emanuele, Yzerbyt, Vincent, Bourguignon, David, and Seron, Eléonore. Who may enter? the impact of in-group identification on in-group/out-group categorization. *Journal of Experimental Social Psychology 38*, 3 (2002), 315–322.

[63] Cay, Damla, Nagel, Till, and Yantaç, Asım Evren. Understanding user experience of COVID-19 maps through remote elicitation interviews. In *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)* (CA, USA, 2020), IEEE, pp. 65–73.

[64] Cebron, Nicolas, Richter, Fabian, and Lienhart, Rainer. "i can tell you what it's not": active learning from counterexamples. *Progress in artificial intelligence 1*, 4 (2012), 291–301.

[65] Chan, Wing-Yi, Qu, Huamin, and Mak, Wai-Ho. Visualizing the semantic structure in classical music works. *IEEE Trans. Vis. Comput. Graph. 16*, 1 (2009), 161–173.

[66] Chancellor, Stevie, Baumer, Eric P. S., and De Choudhury, Munmun. Who is the "human" in human-centered machine learning: The case of predicting mental health from social media. *CSCW 3* (2019).

[67] Chang, Winnie. *Grit and academic performance: Is being grittier better?* PhD thesis, University of Miami, 2014.

[68] Chen, Dong, Bellamy, Rachel KE, Malkin, Peter K, and Erickson, Thomas. Diagnostic visualization for non-expert machine learning practitioners: A design study. In *2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (CA, USA, 2016), IEEE, pp. 87–95.

[69] Chen, M., and Golan, A. What may visualization processes optimize? *IEEE Trans. Vis. Comput. Graph. 22*, 12 (2016), 2619–2632.

[70] Chen, Min, Floridi, Luciano, and Borgo, Rita. *What Is Visualization Really For?* Springer International Publishing, Cham, 2014, pp. 75–93.

[71] Chen, Zhutian, Wang, Yun, Wang, Qianwen, Wang, Yong, and Qu, Huamin. Towards automated infographic design: Deep learning-based auto-extraction of extensible timeline. *IEEE Trans. Vis. Comput. Graph. 26*, 1 (2019), 917–926.

[72] Chevalier, F., Vuillemot, R., and Gali, G. Using concrete scales: A practical framework for effective visual depiction of complex measures. *IEEE Trans. Vis. Comput. Graph. 19*, 12 (2013), 2426–2435.

[73] Chittaro, Luca, and Buttussi, Fabio. Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety. *IEEE Trans. Vis. Comput. Graph. 21*, 4 (2015), 529–538.

[74] Cho, Sumi. Post-intersectionality: The curious reception of intersectionality in legal scholarship1. *Du Bois Review: Social Science Research on Race 10*, 2 (2013), 385–404.

[75] Choi, Hyungsuk, Choi, Woohyuk, Quan, Tran Minh, Hildebrand, David GC, Pfister, Hanspeter, and Jeong, Won-Ki. Vivaldi: A domain-specific language for volume processing and visualization on distributed heterogeneous systems. *IEEE Trans. Vis. Comput. Graph. 20*, 12 (2014), 2407–2416.

[76] Choi, Jin-Wook. Meaning and significance of data analytics for the public sector. `https://www.apo-tokyo.org/resources/articles/meaning-and-significance-of-data-analytics-for-the-public-sector/`.

[77] Clark, William AV, and Avery, Karen L. The effects of data aggregation in statistical analysis. *Geographical Analysis 8*, 4 (1976), 428–438.

[78] Cohen, Jordana B, Hanff, Thomas C, William, Preethi, Sweitzer, Nancy, Rosado-Santander, Nelson R, Medina, Carola, Rodriguez-Mori, Juan E, Renna, Nicolás, Chang, Tara I, Corrales-Medina, Vicente, et al. Continuation versus discontinuation of renin–angiotensin system inhibitors in patients admitted to hospital with COVID-19: a prospective, randomised, open-label trial. *The lancet respiratory medicine 9*, 3 (2021), 275–284.

[79] Collins, Patricia Hill. *Black feminist thought: Knowledge, consciousness, and the politics of empowerment*. Routledge, England, UK, 2002.

[80] Collins, Patricia Hill, and Bilge, Sirma. *Intersectionality*. John Wiley & Sons, USA, 2020.

[81] Collins, Sean. Georgia's COVID-19 cases aren't declining as quickly as initial data suggested they were, May 2020. `https://www.vox.com/covid-19-coronavirus-us-response-trump/2020/5/18/21262265/georgia-covid-19-cases-declining-reopening`.

[82] Concannon, Shauna, Rajan, Natasha, Shah, Parthiv, Smith, Davy, Ursu, Marian, and Hook, Jonathan. Brooke leave home: Designing a personalized film to support public engagement with open data. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2020), ACM, pp. 1–14.

[83] Conklin, Jeff. *Dialogue mapping: Building shared understanding of wicked problems*. John Wiley & Sons, Inc., 2005.

[84] Connelly, Lynne M. Inclusion and exclusion criteria. *Medsurg Nursing 29*, 2 (2020).

[85] Cooney, Sarah. Toward co-creative tools for tactical urban revitalization. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (NY, USA, 2021), ACM, pp. 1–5.

[86] Cornel, Daniel, Konev, Artem, Sadransky, Bernhard, Horvath, Zsolt, Gröller, Eduard, and Waser, Jürgen. Visualization of object-centered vulnerability to possible flood hazards. *Computer Graphics Forum 34*, 3 (2015), 331–340.

[87] Correll, Michael. Ethical dimensions of visualization research. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2019), pp. 1–13.

[88] Correll, Michael, Bertini, Enrico, and Franconeri, Steven. Truncating the y-axis: Threat or menace? In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2020), pp. 1–12.

[89] Costante, Elisa, den Hartog, Jerry, and Petkovic, Milan. On-line trust perception: What really matters. In *2011 1st Workshop on Socio-Technical Aspects in Security and Trust (STAST)* (2011), pp. 52–59.

[90] Couper, Mick. Measuring survey quality in a casic environment. *Proc. of the Survey Research Methods Section of the ASA at JSM1998* (1998), 41–49.

[91] Cox, Donna J. Vis keynote address. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)* (CA, USA, 2015), IEEE, pp. 1–1.

[92] Credé, Marcus, Tynan, Michael C, and Harms, Peter D. Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and social Psychology 113*, 3 (2017), 492.

[93] Crenshaw, Kimberlé. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stan. L. Rev. 43* (1990), 1241.

[94] Crenshaw, Kimberlé. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics [1989]. In *Feminist legal theory*. Routledge, England, UK, 2018, pp. 57–80.

[95] Crowe, Alison, Dirks, Clarissa, and Wenderoth, Mary Pat. Biology in bloom: Implementing bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education 7*, 4 (2008), 368–381.

[96] Cui, Weiwei, Zhang, Xiaoyu, Wang, Yun, Huang, He, Chen, Bei, Fang, Lei, Zhang, Haidong, Lou, Jian-Guan, and Zhang, Dongmei. Text-to-viz: Automatic generation of infographics from proportion-related natural language statements. *IEEE Trans. Vis. Comput. Graph. 26*, 1 (2019), 906–916.

[97] Dai, Chenyun, Lin, Dan, Bertino, Elisa, and Kantarcioglu, Murat. An approach to evaluate data trustworthiness based on data provenance. In *Workshop on Secure Data Management* (2008), Springer, pp. 82–98.

[98] Dalton, Joan, and Smith, David. *Extending Children's Special Abilities: Strategies for primary classrooms*. Office of Schools Administration, Ministry of Education, Victoria, 1989.

[99] Danielson, Robert, Sinatra, Gale, and Kendeou, Panayiota. Augmenting the refutation text effect with analogies and graphics. *Discourse Processes 53* (03 2016).

[100] De Haan, Yael, Kruikemeier, Sanne, Lecheler, Sophie, Smit, Gerard, and van der Nat, Renee. When does an infographic say more than a thousand words? audience evaluations of news visualizations. *Journalism Studies 19*, 9 (2018), 1293–1312.

[101] Deetz, Stanley. Critical theory. In *Engaging organizational communication theory & research: Multiple perspectives*. Sage Publications, CA, USA, 2005, pp. 85–111.

[102] Deitrick, Stephanie. Evaluating implicit visualization of uncertainty for public policy decision support. In *Proc. AutoCarto* (MO, USA, 2012), Cartography and Geographic Information Society.

[103] Dervin, Brenda. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of knowledge management* (1998).

[104] Dick, Murray. Interactive infographics and news values. *Digital Journalism 2*, 4 (2014), 490–506.

[105] D'Ignazio, Catherine, and Klein, Lauren F. Feminist data visualization. In *Proceedings of the Workshop on Vis. for the Digital Humanities (VIS4DH)* (2016), Workshop on Vis. for the Digital Humanities (VIS4DH), IEEE.

[106] D'Ignazio, Catherine, and Klein, Lauren F. *Data feminism*. MIT press, MA, USA, 2020.

[107] Dimara, Evanthia, and Stasko, John. A critical reflection on visualization research: Where do decision making tasks hide? *IEEE Trans. Vis. Comput. Graph. 28*, 1 (2021), 1128–1138.

[108] Dinishak, Janette. The deficit view and its critics. *Disability Studies Quarterly 36*, 4 (2016).

[109] Dörk, Marian, Feng, Patrick, Collins, Christopher, and Carpendale, Sheelagh. Critical infovis: exploring the politics of visualization. In *CHI'13 Extended Abstracts on Hum. Factors in Comput. Sys.* ACM, 2013, pp. 2189–2198.

[110] Douma, Michael, Ligierko, Grzegorz, Ancuta, Ovidiu, Gritsai, Pavel, and Liu, Sean. Spicynodes: Radial layout authoring for the general public. *IEEE Trans. Vis. Comput. Graph. 15*, 6 (2009), 1089–1096.

[111] Drucker, Steven, Huron, Samuel, Kosara, Robert, Schwabish, Jonathan, and Diakopoulos, Nicholas. *Communicating data to an audience*, 1 ed. CRC Press, 2018, ch. 9, p. 21.

[112] Druckman, James N, and Kam, Cindy D. Students as experimental participants. *Cambridge handbook of experimental political science 1* (2011), 41–57.

[113] Duckworth, Angela L, Peterson, Christopher, Matthews, Michael D, and Kelly, Dennis R. Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology 92*, 6 (2007), 1087.

[114] Duckworth, Angela L, and Quinn, Patrick D. Development and validation of the short grit scale (grit–s). *Journal of personality assessment 91*, 2 (2009), 166–174.

[115] Elavsky, Frank, Bennett, Cynthia, and Moritz, Dominik. How accessible is my visualization? evaluating visualization accessibility with chartability. *Computer Graphics Forum 41*, 3 (2022).

[116] Elliott, Kevin C, and Resnik, David B. Science, policy, and the transparency of values. *Environmental health perspectives 122*, 7 (2014), 647–650.

[117] Ellis, Geoffrey, and Dix, Alan. An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization* (2006), pp. 1–7.

[118] Elmqvist, Niklas, and Tsigas, Philippas. TrustNeighborhoods: Visualizing Trust in Distributed File Sharing Systems. In *Eurographics/ IEEE-VGTC Symposium on Visualization* (Switzerland, 2007), K. Museth, T. Moeller, and A. Ynnerman, Eds., The Eurographics Association.

[119] Eskreis-Winkler, Lauren, Shulman, Elizabeth P, Beal, Scott A, and Duckworth, Angela L. The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in psychology 5* (2014), 36.

[120] Faccin-Herman, Vitoria. Visual literacy education: Developing a curriculum for designers and nondesigners. Master's thesis, Iowa State University, Ames, Iowa, 2020.

[121] Fairchild, K.M., Lee, B.H., Loo, J., Ng, H., and Serra, L. The heaven and earth virtual reality: Designing applications for novice users. In *Proceedings of IEEE Virtual Reality Annual International Symposium* (CA, USA, 1993), IEEE, pp. 47–53.

[122] Feenberg, Andrew. *Transforming technology: A critical theory revisited.* Oxford University Press, UK, 2002.

[123] Ferres, Leo, Lindgaard, Gitte, Sumegi, Livia, and Tsuji, Bruce. Evaluating a tool for improving accessibility to charts and graphs. *ACM Trans. Comput.-Hum. Interact. 20*, 5 (2013).

[124] Figueiras, Ana. 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.

[125] Filmer, Deon, and Pritchett, Lant. The effect of household wealth on educational attainment: evidence from 35 countries. *Population and development review 25*, 1 (1999), 85–120.

[126] Finch, Holmes. 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 1*, 1 (2005), 27.

[127] Flower, Linda. *Problem-solving strategies for writing*, 2 ed. Harcourt Brace Jovanovich College Publishers, USA, 1989.

[128] Foote, Mary Q, and Gau Bartell, Tonya. Pathways to equity in mathematics education: How life experiences impact researcher positionality. *Educational Studies in Mathematics 78*, 1 (2011), 45–68.

[129] Forlizzi, Jodi. Moving beyond user-centered design. *Interactions 25*, 5 (2018), 22–23.

[130] Fouad, Nadya, Grus, Catherine, Hatcher, Robert, Kaslow, Nadine, Hutchings, Philinda, Madson, Michael, Jr, Frank, and Crossman, Raymond. Competency benchmarks: A model for understanding and measuring competence in professional psychology across training levels. *Training and Education in Professional Psychology 3* (11 2009).

[131] Franconeri, Steven L, Padilla, Lace M, Shah, Priti, Zacks, Jeffrey M, and Hullman, Jessica. The science of visual data communication: What works. *Psychological Science in the Public Interest 22*, 3 (2021), 110–161.

[132] Frank, Rebecca D, Chen, Zui, Crawford, Erica, Suzuka, Kara, and Yakel, Elizabeth. Trust in qualitative data repositories. *Proc. of the Association for Information Science and Technology 54*, 1 (2017), 102–111.

[133] Fricker, Miranda. *Epistemic injustice: Power and the ethics of knowing.* Oxford University Press, 2007.

[134] Fujishiro, Issei, Ichikawa, Yoshihiko, Furuhata, Rika, and Takeshima, Yuriko. Gadget/iv: a taxonomic approach to semi-automatic design of information visualization applications using modular visualization environment. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings* (CA, USA, 2000), IEEE, pp. 77–83.

[135] Fyfe, Paul. An archaeology of victorian newspapers. *Victorian Periodicals Review 49*, 4 (2016), 546–577.

[136] Galesic, Mirta, and Garcia-Retamero, Rocio. Graph literacy: A cross-cultural comparison. *Medical Decision Making 31*, 3 (2011), 444–457.

[137] Galesic, Mirta, Garcia-Retamero, Rocio, and Gigerenzer, Gerd. Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychology 28*, 2 (2009), 210.

[138] Garcia-Retamero, Rocio, and Cokely, Edward T. Communicating health risks with visual aids. *Current Directions in Psychological Science 22*, 5 (2013), 392–399.

[139] Garcia-Retamero, Rocio, and Galesic, Mirta. Who profits from visual aids: Overcoming challenges in people's understanding of risks. *Social Science & Medicine 70*, 7 (2010), 1019 – 1025.

[140] Garicano, Luis. Hierarchies and the organization of knowledge in production. *Journal of political economy 108*, 5 (2000), 874–904.

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

[142] Gegenfurtner, Andreas, Lehtinen, Erno, and Säljö, Roger. Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational psychology review 23*, 4 (2011), 523–552.

[143] Genevsky, Alexander, Västfjäll, Daniel, Slovic, Paul, and Knutson, Brian. Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. *Journal of Neuroscience 33*, 43 (2013), 17188–17196.

[144] Georgia Department of Public Health. Georgia department of public health daily status report, 2020. `https://dph.georgia.gov/covid-19-daily-status-report`.

[145] Gerry, Lynda Joy. Paint with me: stimulating creativity and empathy while painting with a painter in virtual reality. *IEEE Trans. Vis. Comput. Graph. 23*, 4 (2017), 1418–1426.

[146] Glaas, Erik, Ballantyne, Anne Gammelgaard, Neset, Tina-Simone, Linnér, Björn-Ola, Navarra, Carlo, Johansson, Jimmy, Opach, Tomasz, Rød, Jan Ketil, and Goodsite, Michael E. Facilitating climate change adaptation through communication: Insights from the development of a visualization tool. *Energy Research & Social Science 10* (2015), 57–61.

[147] Goldberg, Lewis R. The structure of phenotypic personality traits. *American psychologist 48*, 1 (1993), 26.

[148] Goo, Sara Kehaulani. The art and science of the scatterplot, 2015.

[149] Google. Insert and edit charts. `https://support.google.com/a/users/answer/9308981?hl=en`. Accessed: 2022-02-06.

[150] Gossett, Nathan, and Chen, Baoquan. Paint inspired color mixing and compositing for visualization. In *IEEE Symposium on Information Visualization* (CA, USA, 2004), IEEE, pp. 113–118.

[151] Gough, Phillip, de Berigny Wall, Caitilin, and Bednarz, Tomasz. Affective and effective visualisation: Communicating science to non-expert users. In *2014 IEEE Pacific Visualization Symposium* (2014), IEEE, pp. 335–339.

[152] Grammel, Lars, Tory, Melanie, and Storey, Margaret-Anne. How information visualization novices construct visualizations. *IEEE Trans. Vis. Comput. Graph. 16*, 6 (2010), 943–952.

[153] Grave, Frank, and Buser, Michael. Visiting the gödel universe. *IEEE Trans. Vis. Comput. Graph. 14*, 6 (2008), 1563–1570.

[154] Gray, Jonathan. Reviving audience studies. *Critical Studies in Media Communication 34*, 1 (2017), 79–83.

[155] Green, Tear Marie, and Fisher, Brian. Towards the personal equation of interaction: The impact of personality factors on visual analytics interface interaction. In *2010 IEEE Symposium on Visual Analytics Science and Technology* (CA, USA, 2010), IEEE, pp. 203–210.

[156] Grove, Susan K, Burns, Nancy, and Gray, Jennifer. *The practice of nursing research: Appraisal, synthesis, and generation of evidence*. Elsevier, Netherlands, 2012.

[157] Gruenwald, Norbert, and Klymchuk, Sergiy. Using counter-examples in teaching calculus. *The New Zealand Mathematics Magazine 40* (2003), 33 – 41.

[158] Grunig, James E. *Publics, audiences and market segments: Segmentation principles for campaigns*, vol. 18. Sage Publications, CA, USA, 1989, pp. 199–228.

[159] Haase, H, Bock, M, Hergenröther, Elke, Knöpfte, C, Koppert, H-J, Schröder, Florian, Trembilski, Andrzej, and Weidenhausen, Jens. Where weather meets the eye—a case study on a wide range of meteorological visualisations for diverse audiences. In *Computer Graphics Forum* (USA, 1999), Springer, pp. 261–266.

[160] Hall, Crystal C, Ariss, Lynn, and Todorov, Alexander. The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior and Hum. Decision Processes 103*, 2 (2007), 277–290.

[161] Hall, Stuart. Encoding/decoding. In *Culture, media, language*. Routledge, England, UK, 2003, pp. 127–137.

[162] Hall, Stuart. 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.

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

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

[165] Haroz, Steve. Experiment methods template, 2022. `http://steveharoz.com/public/experimentmethods/Experiment_Methods_Template.html`.

[166] Haroz, Steve, Kosara, Robert, and Franconeri, Steven L. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2015), Association for Computing Machinery, p. 1191–1200.

[167] Harrison, Lane, Reinecke, Katharina, and Chang, Remco. Infographic aesthetics: Designing for the first impression. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2015), Association for Computing Machinery, p. 1187–1190.

[168] Harry, Beth, and Klingner, Janette. Discarding the deficit model. *Educational leadership 64*, 5 (2007), 16.

[169] Hearst, Marti A, Laskowski, Paul, and Silva, Luis. Evaluating information visualization via the interplay of heuristic evaluation and question-based scoring. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2016), ACM, pp. 5028–5033.

[170] Heer, Jeffrey, Ham, Frank van, Carpendale, Sheelagh, Weaver, Chris, and Isenberg, Petra. Creation and collaboration: Engaging new audiences for information visualization. In *Information visualization*. Springer, USA, 2008, pp. 92–133.

[171] Hervé, Maxime. Package 'RVAideMemoire', 2020. `https://CRAN.R-project.org/package=RVAideMemoire`.

[172] Hoffswell, Jane, Satyanarayan, Arvind, and Heer, Jeffrey. Augmenting code with in situ visualizations to aid program understanding. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2018), ACM, pp. 1–12.

[173] Holmes, Andrew Gary Darwin. Researcher positionality–a consideration of its influence and place in qualitative research–a new researcher guide. *Shanlax Int. J. of Education 8*, 4 (2020), 1–10.

[174] Hopkins, Aspen, Correll, Michael, and Satyanarayan, Arvind. VisuaLint: Sketchy In Situ Annotations of Chart Construction Errors. *Comput. Graphics Forum (Proc. EuroVis)* (2020).

[175] Hoque, Md Naimul, and Mueller, Klaus. Outcome-explorer: A causality guided interactive visual interface for interpretable algorithmic decision making. *IEEE Trans. Vis. Comput. Graph.* (2021).

[176] Horvath, Roman, and Katuscakova, Dominika. Transparency and trust: The case of the european central bank. *Applied Economics 48*, 57 (2016), 5625–5638.

[177] Houts, Peter S, Doak, Cecilia C, Doak, Leonard G, and Loscalzo, Matthew J. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. *Patient education and counseling 61*, 2 (2006), 173–190.

[178] Huang, Weidong, Eades, Peter, and Hong, Seok-Hee. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization 8*, 3 (2009), 139–152.

[179] Huang, Xiaoyun, and Vitak, Jessica. "finsta gets all my bad pictures": Instagram users' self-presentation across finsta and rinsta accounts. *Proc. ACM Hum.-Comput. Interact. 6*, CSCW1 (apr 2022).

[180] Huang, Yi-Jheng, Lin, Wen-Chieh, Yeh, I-Cheng, and Lee, Tong-Yee. Geometric and textural blending for 3d model stylization. *IEEE Trans. Vis. Comput. Graph. 24*, 2 (2017), 1114–1126.

[181] Hulley, Stephen B. *Designing clinical research*. Lippincott Williams & Wilkins, USA, 2007.

[182] Hullman, Jessica. Why authors don't visualize uncertainty. *IEEE Trans. Vis. Comput. Graph. 26*, 1 (2019), 130–139.

[183] Hullman, Jessica, and Diakopoulos, Nick. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Trans. Vis. Comput. Graph. 17*, 12 (2011), 2231–2240.

[184] Insitute, Urban. Urban institute data visualization style guide. `https://urbaninstitute.github.io/graphics-styleguide/`.

[185] Isenberg, Tobias, Isenberg, Petra, Chen, Jian, Sedlmair, Michael, and Möller, Torsten. A systematic review on the practice of evaluating visualization. *IEEE Trans. Vis. Comput. Graph. 19*, 12 (2013), 2818–2827.

[186] Jach, Hayley K, DeYoung, Colin G, and Smillie, Luke D. Why do people seek information? the role of personality traits and situation perception. *J. of Experimental Psychology: General* (2021).

[187] Jaimes, Luis G., Murray, Tylar, and Raij, Andrew. Increasing trust in personal informatics tools. In *Design, User Experience, and Usability. User Experience in Novel Technological Environments* (Berlin, Heidelberg, 2013), Aaron Marcus, Ed., Springer Berlin Heidelberg, pp. 520–529.

[188] Jay, Nancy. Gender and dichotomy. *Feminist studies 7*, 1 (1981), 38–56.

[189] Joffe, Steven, Cook, E. Francis, Cleary, Paul D., Clark, Jeffrey W., and Weeks, Jane C. Quality of Informed Consent: a New Measure of Understanding Among Research Subjects. *JNCI: Journal of the National Cancer Institute 93*, 2 (01 2001), 139–147.

[190] Johansen, Stine S, Merritt, Timothy, Jacobsen, Rune Møberg, Nielsen, Peter Axel, and Kjeldskov, Jesper. Investigating potentials of shape-changing displays for sound zones. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2022), ACM, pp. 1–12.

[191] Jones, K. O., Harland, J., Reid, J. M. V., and Bartlett, R. Relationship between examination questions and bloom's taxonomy. In *2009 39th IEEE Frontiers in Education Conf.* (2009), pp. 1–6.

[192] Jönsson, Daniel, Falk, Martin, and Ynnerman, Anders. Intuitive exploration of volumetric data using dynamic galleries. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2015), 896–905.

[193] Kelton, Kari, Fleischmann, Kenneth R, and Wallace, William A. Trust in digital information. *J. of the American Society for Information Science and Technology 59*, 3 (2008), 363–374.

[194] Kennedy, Helen, and Hill, Rosemary Lucy. The feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology 52*, 4 (2018), 830–848.

[195] Kennedy, Helen, Hill, Rosemary Lucy, Allen, William, and Kirk, Andy. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday 21*, 11 (2016).

[196] Kerracher, N., Kennedy, J., and Chalmers, K. A task taxonomy for temporal graph visualisation. *IEEE Trans. Vis. Comput. Graph. 21*, 10 (2015), 1160–1172.

[197] Kersten-Oertel, Marta, Chen, Sean Jy-Shyang, and Collins, D Louis. An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery. *IEEE Trans. Vis. Comput. Graph. 20*, 3 (2013), 391–403.

[198] Kiernan, Michaela, Oppezzo, Marily A, Resnicow, Kenneth, and Alexander, Gwen L. Effects of a methodological infographic on research participants' knowledge, transparency, and trust. *Health Psychology 37*, 8 (2018), 782.

[199] Kim, Dae Hyun, Setlur, Vidya, and Agrawala, Maneesh. Towards understanding how readers integrate charts and captions: A case study with line charts. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (New York, NY, USA, 2021), ACM.

[200] Kim, Nam Wook, Im, Hyejin, Henry Riche, Nathalie, Wang, Alicia, Gajos, Krzysztof, and Pfister, Hanspeter. Dataselfie: Empowering people to design personalized visuals to represent their data. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2019), Association for Computing Machinery, p. 1–12.

[201] Kim, Yea-Seul, Reinecke, Katharina, and Hullman, Jessica. Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation. *IEEE Trans. Vis. Comput. Graph. 24*, 1 (2017), 760–769.

[202] Kim, Yea-Seul, Walls, Logan A, Krafft, Peter, and Hullman, Jessica. A bayesian cognition approach to improve data visualization. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2019), pp. 1–14.

[203] King, Eden B., Avery, Derek R., Hebl, Mikki R., and Cortina, Jose M. Systematic subjectivity: How subtle biases infect the scholarship review process. *J. of Management 44*, 3 (2018), 843–853.

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

[205] Knaflic, Cole Nussbaumer. *Storytelling with data: A data visualization guide for business professionals.* John Wiley & Sons, 2015.

[206] Kong, Ha-Kyung, Liu, Zhicheng, and Karahalios, Karrie. Trust and recall of information across varying degrees of title-visualization misalignment. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2019), pp. 1–13.

[207] Kosara, Robert. An empire built on sand: Reexamining what we think we know about visualization. In *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization* (2016), pp. 162–168.

[208] Kosara, Robert. The impact of distribution and chart type on part-to-whole comparisons. In *EuroVis (Short Papers)* (2019), pp. 7–11.

[209] Kosara, Robert, Healey, Christopher G, Interrante, Victoria, Laidlaw, David H, and Ware, Colin. Thoughts on user studies: Why, how, and when. *IEEE Computer Graphics and Applications 23*, 4 (2003), 20–25.

[210] Kotranza, Aaron, Lind, D Scott, and Lok, Benjamin. Real-time evaluation and visualization of learner performance in a mixed-reality environment for clinical breast examination. *IEEE Trans. Vis. Comput. Graph. 18*, 7 (2011), 1101–1114.

[211] Kovach, Bill, and Rosenstiel, Tom. *The Elements of Journalism, Revised and Updated 3rd Edition: What Newspeople Should Know and the Public Should Expect*. Crown, 2014.

[212] Krathwohl, David R. A revision of bloom's taxonomy: An overview. *Theory Into Practice 41*, 4 (2002), 212–218.

[213] Krause, Heather. An introduction to the data biography, 2019. `https://weallcount.com/2019/01/21/an-introduction-to-the-data-biography/`.

[214] Kristensson, Per Ola, Dahlback, Nils, Anundi, Daniel, Bjornstad, Marius, Gillberg, Hanna, Haraldsson, Jonas, Martensson, Ingrid, Nordvall, Mathias, and Stahl, Josefine. An evaluation of space time cube representation of spatiotemporal patterns. *IEEE Trans. Vis. Comput. Graph. 15*, 4 (2008), 696–702.

[215] Kristiansen, Yngve S, Garrison, Laura, and Bruckner, Stefan. Semantic snapping for guided multi-view visualization design. *IEEE Trans. Vis. Comput. Graph. 28*, 1 (2021), 43–53.

[216] Krum, Randy. *Cool infographics: Effective communication with data visualization and design*. John Wiley & Sons, 2013.

[217] Kwok, Chi. Epistemic injustice in workplace hierarchies: Power, knowledge and status. *Philosophy & Social Criticism 47*, 9 (2021), 1104–1131.

[218] Kwon, Bum Chul, and Lee, Bongshin. A comparative evaluation on online learning approaches using parallel coordinate visualization. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2016), ACM, pp. 993–997.

[219] Lamal, Peter A. On the importance of replication. *J. of Social Behavior and Personality 5*, 4 (1990), 31.

[220] Lanius, Candice. Fact check: Your demand for statistical proof is racist, 2015. `https://thesocietypages.org/cyborgology/2015/01/12/fact-check-your-demand-for-statistical-proof-is-racist/`.

[221] Law, Chris M, Soo YI, Ji, Choi, Young Sang, and Jacko, Julie A. Unresolved problems in accessibility and universal design guidelines. *Ergonomics in Design 15*, 3 (2007), 7–11.

[222] Lazar, Jonathan, Feng, Jinjuan Heidi, and Hochheiser, Harry. *Research methods in human-computer interaction*. Morgan Kaufmann, MA, USA, 2017.

[223] Lee, Bongshin, Choe, Eun Kyoung, Isenberg, Petra, Marriott, Kim, and Stasko, John. Reaching broader audiences with data visualization. *IEEE Computer Graphics and Applications 40*, 2 (2020), 82–90.

[224] Lee, Crystal, Yang, Tanya, Inchoco, Gabrielle D, Jones, Graham M, and Satyanarayan, Arvind. 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.

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

[226] Lee, Sukwon, Kim, Sung-Hee, and Kwon, Bum Chul. VLAT: Development of a visualization literacy assessment test. *IEEE Trans. Vis. Comput. Graph. 23*, 1 (2016), 551–560.

[227] Lee, Sukwon, Kim, Sung-Hee, and Kwon, Bum Chul. VLAT: Development of a visualization literacy assessment test. *IEEE Trans. Vis. Comput. Graph. 23*, 1 (2017), 551–560.

[228] Lenth, Russell, Singmann, Henrik, Love, Jonathon, et al. *Emmeans: Estimated marginal means, aka least-squares means*, 2018.

[229] Leo, Sarah. Mistakes, we've drawn a few, 2019. `https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368`.

[230] Li, Nan, Brossard, Dominique, Scheufele, Dietram, Wilson, Paul H, and Rose, Kathleen M. Communicating data: interactive infographics, scientific data and credibility. *J. of Science Communication 17*, 2 (2018), A06.

[231] Lin, Xialing, Spence, Patric R, and Lachlan, Kenneth A. Social media and credibility indicators: The effect of influence cues. *Computers in Hum. behavior 63* (2016), 264–271.

[232] Liu, Zhengliang, Crouser, R Jordan, and Ottley, Alvitta. Survey on individual differences in visualization. *Computer Graphics Forum 39*, 3 (2020), 693–712.

[233] Liu, Zhicheng, Thompson, John, Wilson, Alan, Dontcheva, Mira, Delorey, James, Grigg, Sam, Kerr, Bernard, and Stasko, John. Data illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2018), Association for Computing Machinery, p. 1–13.

[234] Lohfink, Anna-Pia, Anton, Simon D Duque, Schotten, Hans Dieter, Leitte, Heike, and Garth, Christoph. Security in process: Visually supported triage analysis in industrial process data. *IEEE Trans. Vis. Comput. Graph. 26*, 4 (2020), 1638–1649.

[235] Lu, Weiquan, Sun, Chenchen, Bleeker, Timo, You, Yingdan, Kitazawa, Shintaro, and Do, Ellen Yi-Luen. Sensorendipity: a real-time web-enabled smartphone sensor platform for idea generation and prototyping. In *Proceedings of the Second International Symposium of Chinese CHI* (NY, USA, 2014), ACM, pp. 11–18.

[236] Lu, Zhicong, Fan, Mingming, Wang, Yun, Zhao, Jian, Annett, Michelle, and Wigdor, Daniel. Inkplanner: Supporting prewriting via intelligent visual diagramming. *IEEE Trans. Vis. Comput. Graph. 25*, 1 (2018), 277–287.

[237] Luca, Michael, and Bazerman, Max H. *The power of experiments: Decision making in a data-driven world*. MIT Press, MA, USA, 2021.

[238] Luo, Yu, and Zhao, Jiaying. Motivated attention in climate change perception and action. *Frontiers in Psychology* (2019), 1541.

[239] Luzón, María José. Public communication of science in blogs: Recontextualizing scientific discourse for a diversified audience. *Written Communication 30*, 4 (2013), 428–457.

[240] Lyra, K. T., Isotani, S., Reis, R. C. D., Marques, L. B., Pedro, L. Z., Jaques, P. A., and Bitencourt, I. I. Infographics or graphics+text: Which material is best for robust learning? In *2016 IEEE 16th Int. Conf. on Adv. Learning Technologies (ICALT)* (2016), pp. 366–370.

[241] Mack, Kelly, McDonnell, Emma, Jain, Dhruv, Lu Wang, Lucy, E. Froehlich, Jon, and Findlater, Leah. 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.

[242] Mahyar, Narges, Kim, Sung-Hee, and Kwon, Bum Chul. Towards a taxonomy for evaluating user engagement in information visualization. In *Workshop on Personal Visualization: Exploring Everyday Life* (2015), vol. 3, p. 2.

[243] Malamed, Connie. *Visual language for designers: Principles for creating graphics that people understand*. Rockport Publishers, 2009.

[244] Marcus, Emilie. Credibility and reproducibility. *Chemistry & biology 22*, 1 (2015), 3–4.

[245] Mason, John. *What makes an example exemplary: Pedagogical and didactical issues in appreciating multiplicative structures*. Routledge, England, UK, 2006, pp. 41–68.

[246] Mauriello, Matthew Louis, Shneiderman, Ben, Du, Fan, Malik, Sana, and Plaisant, Catherine. Simplifying overviews of temporal event sequences. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (NY, USA, 2016), ACM, pp. 2217–2224.

[247] Mayr, Eva, Hynek, Nicole, Salisu, Saminu, and Windhager, Florian. Trust in information visualization. In *EuroVis Workshop on Trustworthy Vis. (TrustVis)* (2019), The Eurographics Association.

[248] McCall, Leslie. The complexity of intersectionality. *Signs: Journal of women in culture and society 30*, 3 (2005), 1771–1800.

[249] McDonnell, Kevin T, and Mueller, Klaus. Illustrative parallel coordinates. *Computer Graphics Forum 27*, 3 (2008), 1031–1038.

[250] Meyer, Miriah, and Dykes, Jason. Reflection on reflection in applied visualization research. *IEEE Comput. Graphics and Applications 38*, 6 (2018), 9–16.

[251] Microsoft Support. Use charts and graphs in your presentation. `https://support.microsoft.com/en-us/office/use-charts-and-graphs-in-your-presentation-c74616f1-a5b2-4a37-8695-fbcc043bf526`. Accessed: 2022-02-06.

[252] Mullet, Kevin, Fry, Christopher, and Schiano, Diane. On your marks, get set, browse! In *CHI'97 Extended Abstracts on Human Factors in Computing Systems*. ACM, NY, USA, 1997, pp. 113–114.

[253] Mumba, Mercy Ngosa, Mugoya, George, Jurczyk, Abigail, and Robb, MacKenzie. Associations among balance self-efficacy, physical function, and pain interference among community-dwelling older adults. *MEDSURG Nursing 29* (2020), 19–26.

[254] Munzel, Andreas. Assisting consumers in detecting fake reviews: The role of identity information disclosure and consensus. *J. of Retailing and Consumer Services 32* (2016), 96–108.

[255] Munzner, Tamara. *Visualization analysis and design*. CRC press, 2014.

[256] Murray, Paul, McGee, Fintan, and Forbes, Angus G. A taxonomy of visualization tasks for the analysis of biological pathway data. *BMC Bioinformatics 18*, 2 (2017), 21.

[257] Muzyk, Andrew J., Tew, Chris, Thomas-Fannin, Allie, Dayal, Sanjai, Maeda, Reina, Schramm-Sapyta, Nicole, Andolsek, Kathryn, and Holmer, Shelley. Utilizing bloom's taxonomy to design a substance use disorders course for health professions students. *Substance Abuse 39*, 3 (2018), 348–353.

[258] Myers, Brad A. Sigchi lifetime research award talk–ruby: Reminiscing about user interfaces by brad over the years. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (NY, USA, 2017), ACM, pp. 10–11.

[259] Nentl, Nancy, and Zietlow, Ruth. Using Bloom's taxonomy to teach critical thinking skills to business students. *College & Undergraduate Libraries 15*, 1-2 (2008), 159–172.

[260] Nieman, Adam. Concrete vs abstract visualisation: the real world as a canvas for data visualisation. In *Proceedings of ADS-VIS2011: Making visible the invisible: Art, design and science in data visualisation* (UK, 2012), University of Huddersfield Press, pp. 49–56.

[261] Nightingale, Virginia, et al. *The handbook of media audiences*. Wiley Online Library, 2011.

[262] Nisbet, Matthew C, and Scheufele, Dietram A. What's next for science communication? promising directions and lingering distractions. *American journal of botany 96*, 10 (2009), 1767–1778.

[263] Norman, Donald A. *Emotional design: Why we love (or hate) everyday things*. Basic Civitas Books, 2004.

[264] North, Chris. Toward measuring visualization insight. *IEEE Computer Graphics and Applications 26*, 3 (2006), 6–9.

[265] North, Chris, Saraiya, Purvi, and Duca, Karen. A comparison of benchmark task and insight evaluation methods for information visualization. *Information Visualization 10*, 3 (2011), 162–181.

[266] Ogbonnaya-Ogburu, Ihudiya Finda, Smith, Angela DR, To, Alexandra, and Toyama, Kentaro. Critical race theory for hci. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2020), ACM, pp. 1–16.

[267] Okan, Yasmina, Janssen, Eva, Galesic, Mirta, and Waters, Erika A. Using the short graph literacy scale to predict precursors of health behavior change. *Medical Decision Making 39*, 3 (2019), 183–195.

[268] Otten, Jennifer J, Cheng, Karen, and Drewnowski, Adam. Infographics and public policy: using data visualization to convey complex information. *Health Affairs 34*, 11 (2015), 1901–1907.

[269] Ottley, Alvitta, Kaszowska, Aleksandra, Crouser, R Jordan, and Peck, Evan M. The curious case of combining text and visualization. In *EuroVis (Short Papers)* (2019), pp. 121–125.

[270] Ovans, Andrea. What makes the best infographics so convincing, April 2014. https://hbr.org/2014/04/what-makes-the-best-infographics-so-convincing.

[271] O'Donovan, Peter, Agarwala, Aseem, and Hertzmann, Aaron. Learning layouts for single-pagegraphic designs. *IEEE Trans. Vis. Comput. Graph. 20*, 8 (2014), 1200–1213.

[272] O'Reilly, Jim. Paradata and blaise: A review of recent applications and research. In *Int. Blaise Users Conf. (IBUC), Latvia* (2009).

[273] Paas, Fred, Tuovinen, Juhani E., Tabbers, Huib, and Gerven, Pascal W. M. Van. Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist 38*, 1 (2003), 63–71.

[274] Padilla, Lace, Kay, Matthew, and Hullman, Jessica. *Uncertainty Visualization.* John Wiley & Sons, Ltd, 2021, pp. 1–18.

[275] Padilla, Lace M, Hansen, Grace, Ruginski, Ian T, Kramer, Heidi S, Thompson, William B, and Creem-Regehr, Sarah H. The influence of different graphical displays on nonexpert decision making under uncertainty. *Journal of Experimental Psychology: Applied 21*, 1 (2015), 37.

[276] Padilla, Stefano, Methven, Thomas S, Robb, David A, and Chantler, Mike J. Understanding concept maps: A closer look at how people organise ideas. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2017), ACM, pp. 815–827.

[277] Pang, Alex, Wittenbrink, Craig, and Lodha, Suresh. Approaches to uncertainty visualization. *The Visual Computer 13* (10 1996), 24.

[278] Park, Douglas B. The meanings of "audience". *College English 44*, 3 (1982), 247–257.

[279] Park, J. H., Kaufman, A., and Mueller, K. Graphoto: Aesthetically pleasing charts for casual information visualization. *IEEE Computer Graphics and Applications 38*, 6 (2018), 67–82.

[280] Park, Ji Hwan, Kaufman, Arie, and Mueller, Klaus. Graphoto: Aesthetically pleasing charts for casual information visualization. *IEEE computer graphics and applications 38*, 6 (2018), 67–82.

[281] Park, Ji Hwan, Nadeem, Saad, Mirhosseini, Seyedkoosha, and Kaufman, Arie. C 2 a: Crowd consensus analytics for virtual colonoscopy. In *2016 IEEE Conf. on Visual Analytics Science and Technology (VAST)* (CA, USA, 2016), IEEE, pp. 21–30.

[282] Patino, Cecilia Maria, and Ferreira, Juliana Carvalho. Inclusion and exclusion criteria in research studies: definitions and why they matter. *Jornal Brasileiro de Pneumologia 44* (2018), 84–84.

[283] Peck, Evan M, Ayuso, Sofia E, and El-Etr, Omar. Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2019), pp. 1–12.

[284] Peer, Eyal, Vosgerau, Joachim, and Acquisti, Alessandro. Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior research methods 46*, 4 (2014), 1023–1031.

[285] Pentland, Alex "Sandy". The data-driven society. *Scientific American 309*, 4 (2013), 78–83.

[286] Perkins, David, and Salomon, Gavriel. Transfer of learning. In *The International Encyclopedia of Education*, 2 ed. Pergamon Press, 1992.

[287] Perovich, Laura J, Wylie, Sara Ann, and Bongiovanni, Roseann. Chemicals in the creek: designing a situated data physicalization of open government data with the community. *IEEE Trans. Vis. Comput. Graph. 27*, 2 (2020), 913–923.

[288] Pfitzner, Darius, Hobbs, Vaughan, and Powers, David. A unified taxonomic framework for information visualization. In *Proceedings of the Asia-Pacific Symposium on Information Visualisation - Volume 24* (AUS, 2003), Australian Computer Society, Inc., p. 57–66.

[289] Pflüger, Hermann, Thom, Dennis, Schütz, Anna, Bohde, Daniela, and Ertl, Thomas. Vechart: visually enhanced comparison of historic art using an automated line-based synchronization technique. *IEEE Trans. Vis. Comput. Graph. 26*, 10 (2019), 3063–3076.

[290] Phipps, Molly, and Rowe, Shawn. Seeing satellite data. *Public Understanding of Science 19*, 3 (2010), 311–321.

[291] Pina, Laura, Sien, Sang-Wha, Song, Clarissa, Ward, Teresa M., Fogarty, James, Munson, Sean A., and Kientz, Julie A. Dreamcatcher: Exploring how parents and school-age children can track and review sleep information together. *Proc. ACM Hum.-Comput. Interact. 4*, CSCW1 (may 2020).

[292] Pirolli, Peter, and Card, Stuart. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (2005), vol. 5, McLean, VA, USA, pp. 2–4.

[293] Pirson, Michael, and Malhotra, Deepak. Foundations of organizational trust: What matters to different stakeholders? *Organization Science 22*, 4 (2011), 1087–1104.

[294] Plaisant, Catherine. Information visualization and the challenge of universal usability. In *Exploring geovisualization*. Elsevier, Netherlands, 2005, pp. 53–82.

[295] Pontis, Sheila. Start with the basics, April 2012. https://sheilapontis.com/2012/04/05/start-with-the-basics/.

[296] Pouliot, Jacynthe, Ellul, Claire, Hubert, Frédéric, Wang, Chen, Rajabifard, Abbas, Kalantari, Mohsen, Shojaei, Davood, Atazadeh, Behnam, Oosterom, Petrus Johannes Maria van, De Vries, Marian, et al. Visualization and new opportunities. In *Best practices 3D cadastres: extended version*. FIG publication, Denmark, 2018, pp. 183–230.

[297] Pousman, Zachary, Stasko, John, and Mateas, Michael. Casual information visualization: Depictions of data in everyday life. *IEEE Trans. Vis. Comput. Graph. 13*, 6 (2007), 1145–1152.

[298] Preim, Bernhard, Ropinski, Timo, and Isenberg, Petra. A critical analysis of the evaluation practice in medical visualization. In *Eurographics 8th EG Workshop on Visual Computing for Biology and Medicine* (NJ, USA, 2018), Wiley Online Library.

[299] Pretorius, AJ, Purchase, HC., and Stasko, JT. Tasks for multivariate network analysis. In *Multivariate Network Visualization* (2014), Kerren A., Purchase H.C., and Ward M.O., Eds., vol. 8380, Springer, Cham.

[300] Prezi Support. Inserting charts in prezi present. `https://support.prezi.com/hc/en-us/articles/360003479074-Inserting-Charts-in-Prezi-Present`. Accessed: 2022-02-06.

[301] Prolific. Prolific, 2022. `https://www.prolific.co/`.

[302] ProPublica. WeePeople font. https://github.com/propublica/weepeople/.

[303] Pu, Xiaoying, Kross, Sean, Hofman, Jake M, and Goldstein, Daniel G. Datamations: Animated explanations of data analysis pipelines. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2021), pp. 1–14.

[304] Qian, Chunyao, Sun, Shizhao, Cui, Weiwei, Lou, Jian-Guang, Zhang, Haidong, and Zhang, Dongmei. Retrieve-then-adapt: Example-based automatic generation for proportion-related infographics. *IEEE Trans. Vis. Comput. Graph. 27*, 2 (2020), 443–452.

[305] Qualtrics. Qualtrics, 2022. `https://www.qualtrics.com`.

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

[307] Ragan, Eric D, Endert, Alex, Sanyal, Jibonananda, and Chen, Jian. Characterizing provenance in visualization and data analysis: an organizational framework of provenance types and purposes. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2015), 31–40.

[308] Rankin, Yolanda A, and Thomas, Jakita O. Straighten up and fly right: Rethinking intersectionality in hci research. *Interactions 26*, 6 (2019), 64–68.

[309] Rao, Goutham. Physician numeracy: Essential skills for practicing evidence-based medicine. *Family medicine 40*, 5 (2008), 354.

[310] Rautek, Peter, Bruckner, Stefan, Gröller, M Eduard, and Hadwiger, Markus. Vislang: A system for interpreted domain-specific languages for scientific visualization. *IEEE Trans. Vis. Comput. Graph. 20*, 12 (2014), 2388–2396.

[311] Reber, Rolf, and Schwarz, Norbert. Effects of perceptual fluency on judgments of truth. *Consciousness and cognition 8*, 3 (1999), 338–342.

[312] Reber, Rolf, Schwarz, Norbert, and Winkielman, Piotr. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review 8*, 4 (2004), 364–382.

[313] Reber, Rolf, Winkielman, Piotr, and Schwarz, Norbert. Effects of perceptual fluency on affective judgments. *Psychological science 9*, 1 (1998), 45–48.

[314] Reja, Urša, Manfreda, Katja Lozar, Hlebec, Valentina, and Vehovar, Vasja. Open-ended vs. close-ended questions in web questionnaires. *Developments in applied statistics 19*, 1 (2003), 159–177.

[315] Ren, Donghao, Lee, Bongshin, and Brehmer, Matthew. Charticulator: Interactive construction of bespoke chart layouts. *IEEE Trans. Vis. Comput. Graph. 25*, 1 (2018), 789–799.

[316] Rind, Alexander, Aigner, Wolfgang, Wagner, Markus, Miksch, Silvia, and Lammarsch, Tim. Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. *Information Visualization 15*, 4 (2016), 288–300.

[317] Rittel, Horst WJ, and Webber, Melvin M. Dilemmas in a general theory of planning. *Policy sciences 4*, 2 (1973), 155–169.

[318] Romano, Alessandro, Sotis, Chiara, Dominioni, Goran, and Guidi, Sebastián. The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences. *Health economics 29*, 11 (2020), 1482–1494.

[319] Rosen, Paul, Popescu, Voicu, Hoffmann, Christoph, and Irfanoglu, Ayhan. A high-quality high-fidelity visualization of the september 11 attack on the world trade center. *IEEE Trans. Vis. Comput. Graph. 14*, 4 (2008), 937–947.

[320] Rothenberger, Scott D, Wenskovitch, John E, and Marai, G Elisabeta. Pexel and heatmap visual analysis of multidimensional gun/homicide data. In *2011 IEEE Conf. on Visual Analytics Science and Technology (VAST)* (CA, USA, 2011), IEEE, pp. 297–298.

[321] Rowan, Katherine E. When simple language fails: Presenting difficult science to the public. *J. of Technical Writing and Commun. 21*, 4 (1991), 369–382.

[322] Rowe, Wendy E. Positionality. In *The SAGE encyclopedia of action research*, David Coghlan and Mary Brydon-Miller, Eds. Sage, 2014.

[323] Ruchikachorn, Puripant, and Mueller, Klaus. Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE Trans. Vis. Comput. Graph. 21*, 9 (2015), 1028–1044.

[324] Ryoo, Kihyun, and Linn, Marcia C. Designing guidance for interpreting dynamic visualizations: Generating versus reading explanations. *J. of Research in Science Teaching 51*, 2 (2014), 147–174.

[325] Sacha, Dominik, Senaratne, Hansi, Kwon, Bum Chul, Ellis, Geoffrey, and Keim, Daniel A. The role of uncertainty, awareness, and trust in visual analytics. *IEEE Trans. Vis. Comput. Graph. 22*, 1 (2015), 240–249.

[326] Saket, Bahador, Endert, Alex, and Stasko, John. Beyond usability and performance: A review of user experience-focused evaluations in visualization. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (2016), pp. 133–142.

[327] Salama, Christof Rezk, Keller, Maik, and Kohlmann, Peter. High-level user interfaces for transfer function design with semantics. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 1021–1028.

[328] Salem, Maha, and Dautenhahn, Kerstin. Evaluating trust and safety in hri: Practical issues and ethical challenges. *Emerging Policy and Ethics of Hum.-Robot Interaction* (2015).

[329] Salem, Maha, Lakatos, Gabriella, Amirabdollahian, Farshid, and Dautenhahn, Kerstin. Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust. In *2015 10th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)* (2015), IEEE, pp. 1–8.

[330] Satyanarayan, Arvind, Moritz, Dominik, Wongsuphasawat, Kanit, and Heer, Jeffrey. Vega-lite: A grammar of interactive graphics. *IEEE Trans. Vis. Comput. Graph. 23*, 1 (2016), 341–350.

[331] Savin-Baden, Maggi, and Howell-Major, Claire. Qualititative research: The essential guide to theory and practice. *Qualitative Research: The Essential Guide to Theory and Practice. Routledge* (2013).

[332] Schlesinger, Ari, Edwards, W Keith, and Grinter, Rebecca E. Intersectional hci: Engaging identity through gender, race, and class. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2017), ACM, pp. 5412–5427.

[333] Schnackenberg, Andrew K, and Tomlinson, Edward C. Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships. *J. of management 42*, 7 (2016), 1784–1810.

[334] Schönborn, K. J., Höst, G. E., and Lundin Palmerius, K. E. Measuring understanding of nanoscience and nanotechnology: development and validation of the nano-knowledge instrument (NanoKI). *Chem. Educ. Res. Pract. 16* (2015), 346–354.

[335] Schulz, H., Nocke, T., Heitzler, M., and Schumann, H. A design space of visualization tasks. *IEEE Trans. Vis. Comput. Graph. 19*, 12 (2013), 2366–2375.

[336] Schwabish, Jonathan. The practice of visual data communication: What works. *Psychological Science in the Public Interest 22*, 3 (2021), 97–109.

[337] Sedlmair, Michael, Frank, Annika, Munzner, Tamara, and Butz, Andreas. Relex: Visualization for actively changing overlay network specifications. *IEEE Trans. Vis. Comput. Graph. 18*, 12 (2012), 2729–2738.

[338] Sedlmair, Michael, Isenberg, Petra, Isenberg, Tobias, Mahyar, Narges, and Lam, Heidi. Preface. In *BELIV '16: Proc. of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization* (New York, NY, USA, 2016), Association for Computing Machinery.

[339] Segel, Edward, and Heer, Jeffrey. Narrative visualization: Telling stories with data. *IEEE Trans. Vis. Comput. Graph. 16*, 6 (2010), 1139–1148.

[340] Setlur, Vidya, and Tory, Melanie. How do you converse with an analytical chatbot? revisiting gricean maxims for designing analytical conversational behavior. *arXiv preprint arXiv:2203.08420* (2022).

[341] Shen, Hong, Jin, Haojian, Cabrera, Ángel Alexander, Perer, Adam, Zhu, Haiyi, and Hong, Jason I. Designing alternative representations of confusion matrices to support non-expert public understanding of algorithm performance. *Proc. of the ACM on Human-Computer Interaction 4*, CSCW2 (2020), 1–22.

[342] Shen, I-Chao, and Chen, Bing-Yu. Clipgen: A deep generative model for clipart vectorization and synthesis. *IEEE Trans. Vis. Comput. Graph.* (2021).

[343] Shizha, Edward. Legitimizing indigenous knowledge in zimbabwe: A theoretical analysis of postcolonial school knowledge and its colonial legacy. *Journal of Contemporary Issues in Education 1* (2006), 20–35.

[344] Sigma Awards. About — the sigma awards. `https://sigmaawards.org/about/`. Accessed: 2022-01-30.

[345] Sigma Awards. 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.

[346] Sigma Awards. Mapping makoko. https://sigmaawards.org/mapping-makoko/. Accessed: 2022-01-30.

[347] Sigma Awards. 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.

[348] Sigma Awards. The COVID Tracking Project at The Atlantic. https://sigmaawards.org/the-covid-tracking-project-at-the-atlantic/. Accessed: 2022-01-30.

[349] Sigma Awards. 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.

[350] Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., et al. Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science 1*, 3 (2018), 337–356.

[351] Simmhan, Yogesh L, Plale, Beth, and Gannon, Dennis. A survey of data provenance in e-science. *ACM Sigmod Record 34*, 3 (2005), 31–36.

[352] Singh, Divit P, Lisle, Lee, Murali, TM, and Luther, Kurt. Crowdlayout: Crowdsourced design and evaluation of biological network visualizations. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2018), ACM, pp. 1–14.

[353] Skau, Drew, and Kosara, Robert. Readability and precision in pictorial bar charts. In *Proc. of the Eurographics/IEEE VGTC Conf. on Visualization: Short Papers* (2017), EuroVis '17, Eurographics Association, p. 91–95.

[354] Slater, Michael D. Theory and method in health audience segmentation. *Journal of health communication 1*, 3 (1996), 267–284.

[355] Smart, Stephen, Wu, Keke, and Szafir, Danielle Albers. Color crafting: Automating the construction of designer quality color ramps. *IEEE Trans. Vis. Comput. Graph. 26*, 1 (2019), 1215–1225.

[356] Smiciklas, Mark. *The power of infographics: Using pictures to communicate and connect with your audiences*. Que Publishing, 2012.

[357] Smid, Henricus, Mast, Patrick, Tromp, Maarten, Winterboer, Andi, and Evers, Vanessa. Canary in a coal mine: monitoring air quality and detecting environmental incidents by harvesting twitter. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, NY, USA, 2011, pp. 1855–1860.

[358] Sohn, Dongyoung, and Choi, Soyoung. Social embeddedness of persuasion: effects of cognitive social structures on information credibility assessment and sharing in social media. *Int. J. of Advertising 38*, 6 (2019), 824–844.

[359] Song, Miri. Introduction: Who's at the bottom? examining claims about racial hierarchy. *Ethnic and Racial Studies 27*, 6 (2004), 859–877.

[360] South, Laura, Schwab, Michail, Beauchamp, Nick, Wang, Lu, Wihbey, John, and Borkin, Michelle A. Debatevis: Visualizing political debates for non-expert users. In *2020 IEEE Visualization Conference (VIS)* (CA, USA, 2020), IEEE, pp. 241–245.

[361] Stamkou, Eftychia, van Kleef, Gerben A, Homan, Astrid C, and Galinsky, Adam D. How norm violations shape social hierarchies: Those who stand on top block norm violators from rising up. *Group Processes & Intergroup Relations 19*, 5 (2016), 608–629.

[362] Starbuck, William H, and Milliken, Frances J. Executives' perceptual filters: What they notice and how they make sense. In *The executive effect: concepts and methods for studying top managers* (Greenwich, CT, 1988), Donald Hambrick, Ed., JAI Press, pp. 35–65.

[363] Stern, Cindy, Jordan, Zoe, and McArthur, Alexa. Developing the review question and inclusion criteria. *AJN The American Journal of Nursing 114*, 4 (2014), 53–56.

[364] Stokes, Chase, and Hearst, Marti A. Give text a chance: Advocating for equal consideration for language and visualization. In *NLViz Workshop* (2021).

[365] Streit, Marc, and Gehlenborg, Nils. Bar charts and box plots. *Nature Methods 11* (2014).

[366] Sullivan, John L. *Media audiences: Effects, users, institutions, and power*. Sage Publications, CA, USA, 2019.

[367] Sultana, Sharifa, Ahmed, Syed Ishtiaque, and Rzeszotarski, Jeffrey M. Seeing in context: Traditional visual communication practices in rural bangladesh. *Proc. ACM Hum.-Comput. Interact. 4*, CSCW3 (jan 2021).

[368] Syeda, Uzma Haque, Murali, Prasanth, Roe, Lisa, Berkey, Becca, and Borkin, Michelle A. Design study" lite" methodology: Expediting design studies and enabling the synergy of visualization pedagogy and social good. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2020), ACM, pp. 1–13.

[369] Tableau Software. Tableau software. `https://www.tableau.com/`. Accessed: 2022-01-30.

[370] Taylor, Alex S. Out there. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (NY, USA, 2011), ACM, pp. 685–694.

[371] Taylor, Richard. Interpretation of the correlation coefficient: a basic review. *J. of diagnostic medical sonography 6*, 1 (1990), 35–39.

[372] Thomas, Jim, and Kielman, Joe. Challenges for visual analytics. *Information Vis. 8*, 4 (2009), 309–314.

[373] Thompson, Christopher Glen, Kim, Rae Seon, Aloe, Ariel M, and Becker, Betsy Jane. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology 39*, 2 (2017), 81–90.

[374] Tietjen, Christian, Isenberg, Tobias, and Preim, Bernhard. Combining silhouettes, surface, and volume rendering for surgery education and planning. In *EuroVis* (Switzerland, 2005), The Eurographics Association, pp. 303–310.

[375] Tomlinson, Barbara. Colonizing intersectionality: Replicating racial hierarchy in feminist academic arguments. *Social Identities 19*, 2 (2013), 254–272.

[376] Torres, Cesar, O'Leary, Jasper, Nicholas, Molly, and Paulos, Eric. Illumination aesthetics: Light as a creative material within computational design. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2017), ACM, pp. 6111–6122.

[377] Trogu, Pino. Counting but losing count: the legacy of Otto Neurath's ISOTYPE charts. *Visible Language 52*, 2 (2018).

[378] Truong, Khai N., Hayes, Gillian R., and Abowd, Gregory D. Storyboarding: An empirical determination of best practices and effective guidelines. In *Proceedings of the 6th Conference on Designing Interactive Systems* (New York, NY, USA, 2006), Association for Computing Machinery, p. 12–21.

[379] Tufte, Edward R. *The visual display of quantitative information*, vol. 2. Graphics press, 2001.

[380] Twyman, Michael. The significance of isotype. *Graphic communication through ISOTYPE 7* (1975).

[381] U.S. Department of Education. Highlights of the 2017 U.S. PIAAC results web report. Tech. rep., Institute of Education Sci., National Center for Education Statistics, 2017. `https://nces.ed.gov/surveys/piaac/current_results.asp`.

[382] Valiati, Eliane R. A., Pimenta, Marcelo S., and Freitas, Carla M. D. S. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization* (2006), BELIV '06, Association for Computing Machinery, p. 1–6.

[383] Van Berkel, Niels, Goncalves, Jorge, Russo, Daniel, Hosio, Simo, and Skov, Mikael B. Effect of information presentation on fairness perceptions of machine learning predictors. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2021), ACM, pp. 1–13.

[384] Van den Elzen, Stef, and Van Wijk, Jarke J. Multivariate network exploration and presentation: From detail to overview via selections and aggregations. *IEEE Trans. Vis. Comput. Graph. 20*, 12 (2014), 2310–2319.

[385] Waldner, Manuela, Diehl, Alexandra, Gračanin, Denis, Splechtna, Rainer, Delrieux, Claudio, and Matković, Krešimir. A comparison of radial and linear charts for visualizing daily patterns. *IEEE Trans. Vis. Comput. Graph. 26*, 1 (2019), 1033–1042.

[386] Wall, Emily, Xiong, Cindy, and Kim, Yea-Seul. Vishikers' guide to evaluation: Competing considerations in study design. *IEEE Computer Graphics and Applications 42*, 3 (2022), 29–38.

[387] Walny, Jagoda, and Carpendale, Sheelagh. Data sketches: An exploratory study. In *InfoVis 2014 Posters Compendium* (2014).

[388] Walny, Jagoda, Huron, Samuel, and Carpendale, Sheelagh. An exploratory study of data sketching for visual representation. *Computer Graphics Forum 34*, 3 (2015), 231–240.

[389] Wang, Qianwen, Li, Zhen, Fu, Siwei, Cui, Weiwei, and Qu, Huamin. Narvis: Authoring narrative slideshows for introducing data visualization designs. *IEEE Trans. Vis. Comput. Graph. 25*, 1 (2019), 779–788.

[390] Wang, Yong, Jin, Zhihua, Wang, Qianwen, Cui, Weiwei, Ma, Tengfei, and Qu, Huamin. Deepdrawing: A deep learning approach to graph drawing. *IEEE Trans. Vis. Comput. Graph. 26*, 1 (2019), 676–686.

[391] Wang, Yun, Sun, Zhida, Zhang, Haidong, Cui, Weiwei, Xu, Ke, Ma, Xiaojuan, and Zhang, Dongmei. Datashot: Automatic generation of fact sheets from tabular data. *IEEE Trans. Vis. Comput. Graph. 26*, 1 (2019), 895–905.

[392] Wang, Yun, Zhang, Haidong, Huang, He, Chen, Xi, Yin, Qiufeng, Hou, Zhitao, Zhang, Dongmei, Luo, Qiong, and Qu, Huamin. Infonice: Easy creation of information graphics. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2018), Association for Computing Machinery, p. 1–12.

[393] Wang, Zezhong, Sundin, Lovisa, Murray-Rust, Dave, and Bach, Benjamin. Cheat sheets for data visualization techniques. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (New York, NY, USA, 2020), ACM, p. 1–13.

[394] Wang, Zezhong, Wang, Shunming, Farinella, Matteo, Murray-Rust, Dave, Henry Riche, Nathalie, and Bach, Benjamin. Comparing effectiveness and engagement of data comics and infographics. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2019), Association for Computing Machinery.

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

[396] Wasike, Ben. Memes, memes, everywhere, nor any meme to trust: Examining the credibility and persuasiveness of COVID-19-related memes. *J. of Computer-Mediated Communication 27*, 2 (2022), zmab024.

[397] Webster, James G. The audience. *Journal of Broadcasting & Electronic Media 42*, 2 (1998), 190–207.

[398] Wiegand, Gesa, Schmidmaier, Matthias, Weber, Thomas, Liu, Yuanting, and Hussmann, Heinrich. I drive-you trust: Explaining driving behavior of autonomous cars. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (NY, USA, 2019), ACM, pp. 1–6.

[399] Wilson, Angela Cavender. *Reclaiming our humanity: Decolonization and the recovery of Indigenous knowledge*. Rowman & Littlefield Lanham, CA, 2005, pp. 255–263.

[400] Windhager, Florian, Federico, Paolo, Schreder, Günther, Glinka, Katrin, Dörk, Marian, Miksch, Silvia, and Mayr, Eva. Visualization of cultural heritage collection data: State of the art and future challenges. *IEEE Trans. Vis. Comput. Graph. 25*, 6 (2018), 2311–2330.

[401] Winnie, John A. The implicit definition of theoretical terms. *The British Journal for the Philosophy of Science 18*, 3 (1967), 223–229.

[402] Wong, Yuet Ling, Madhavan, Krishna, and Elmqvist, Niklas. Towards characterizing domain experts as a user group. In *2018 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)* (CA, USA, 2018), IEEE, pp. 1–10.

[403] Wong-Villacres, Marisol, Kumar, Arkadeep, Vishwanath, Aditya, Karusala, Naveena, DiSalvo, Betsy, and Kumar, Neha. Designing for intersections. In *Proceedings of the 2018 Designing Interactive Systems Conference* (NY, USA, 2018), ACM, pp. 45–58.

[404] Wood, Jo, Beecham, Roger, and Dykes, Jason. Moving beyond sequential design: Reflections on a rich multi-channel approach to data visualization. *IEEE Trans. Vis. Comput. Graph. 20*, 12 (2014), 2171–2180.

[405] WTF visualizations, September 2019. `https://viz.wtf/post/187558414596/the-axis-choices-are-interesting-this-thing-is`.

[406] Wu, Huayi, He, Zhengwei, and Gong, Jianya. A virtual globe-based 3d visualization and interactive framework for public participation in urban planning processes. *Computers, Environment and Urban Systems 34*, 4 (2010), 291–298.

[407] Wun, Tiffany, Payne, Jennifer, Huron, Samuel, and Carpendale, Sheelagh. Comparing bar chart authoring with microsoft excel and tangible tiles. *Computer Graphics Forum 35*, 3 (2016), 111–120.

[408] Xia, Haijun, Henry Riche, Nathalie, Chevalier, Fanny, De Araujo, Bruno, and Wigdor, Daniel. Dataink: Direct and creative data-oriented drawing. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2018), Association for Computing Machinery, p. 1–13.

[409] Xiong, Cindy, Padilla, Lace, Grayson, Kent, and Franconeri, Steven. Examining the components of trust in map-based visualizations. In *EuroVis Workshop on Trustworthy Vis. (TrustVis)* (2019), The Eurographics Association.

[410] Xiong, Cindy, Sarvghad, Ali, Goldstein, Daniel G, Hofman, Jake M, and Demiralp, Çagatay. Investigating perceptual biases in icon arrays. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2022), ACM, pp. 1–12.

[411] Xiong, Cindy, Shapiro, Joel, Hullman, Jessica, and Franconeri, Steven. Illusion of causality in visualized data. *IEEE Trans. Vis. Comput. Graph. 26*, 1 (2019), 853–862.

[412] Xiong, Cindy, Van Weelden, Lisanne, and Franconeri, Steven. The curse of knowledge in visual data communication. *IEEE Trans. Vis. Comput. Graph. 26*, 10 (2019), 3051–3062.

[413] Xu, Kai, Ottley, Alvitta, Walchshofer, Conny, Streit, Marc, Chang, Remco, and Wenskovitch, John. Survey on the analysis of user interactions and visualization provenance. *Comput. Graphics Forum 39*, 3 (2020), 757–783.

[414] Xu, Songhua, Yang, Wenxia, and Lau, Francis CM. A visualization based approach for digital signature authentication. *Computer Graphics Forum 28*, 3 (2009), 935–942.

[415] Xuan, Xiwei, Zhang, Xiaoyu, Kwon, Oh-Hyun, and Ma, Kwan-Liu. Vac-cnn: A visual analytics system for comparative studies of deep convolutional neural networks. *IEEE Trans. Vis. Comput. Graph. 28*, 6 (2022), 2326–2337.

[416] Yakel, Elizabeth, Faniel, Ixchel M, Kriesberg, Adam, and Yoon, Ayoung. Trust in digital repositories. *The Int. J. of Digital Curation 8* (2013).

[417] Yin, Michelle, Hollender, David, Condelli, Larry, Shewitz, Dahlia, Duffy, Amanda, and Movit, Marcela. *The Power of Data Visualization: Advanced Presentations of NRS Data*, January 2014.

[418] Yu, Liangzhi. How poor informationally are the information poor? evidence from an empirical study of daily and regular information practices of individuals. *Journal of Documentation* (2010).

[419] Zapata-Rivera, Juan Diego, and Katz, Irvin R. Keeping your audience in mind: Applying audience analysis to the design of interactive score reports. *Assessment in Education: Principles, Policy & Practice 21*, 4 (2014), 442–463.

[420] Zazkis, Rina, and Chernoff, Egan J. What makes a counterexample exemplary? *Educational Studies in Mathematics 68*, 3 (2008), 195–208.

[421] Zehrung, Rachael, Singhal, Astha, Correll, Michael, and Battle, Leilani. Vis ex machina: An analysis of trust in human versus algorithmically generated visualization recommendations. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2021), pp. 1–12.

[422] Zender, Mike, and Mejía, G Mauricio. Improving icon design: through focus on the role of individual symbols in the construction of meaning. *Visible Language 47*, 1 (2013), 66–89.

[423] Zeng, Wei, Fu, Chi-Wing, Arisona, Stefan Müller, Erath, Alexander, and Qu, Huamin. Visualizing mobility of public transportation system. *IEEE Trans. Vis. Comput. Graph. 20*, 12 (2014), 1833–1842.

[424] Zenner, André, Kriegler, Hannah Maria, and Krüger, Antonio. Hart-the virtual reality hand redirection toolkit. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (NY, USA, 2021), ACM, pp. 1–7.

[425] Zenner, André, Makhsadov, Akhmajon, Klingner, Sören, Liebemann, David, and Krüger, Antonio. Immersive process model exploration in virtual reality. *IEEE Trans. Vis. Comput. Graph. 26*, 5 (2020), 2104–2114.

[426] Zhang, Fang-Lue, Wang, Jue, Shechtman, Eli, Zhou, Zi-Ye, Shi, Jia-Xin, and Hu, Shi-Min. Plenopatch: Patch-based plenoptic image manipulation. *IEEE Trans. Vis. Comput. Graph. 23*, 5 (2016), 1561–1573.

[427] Zhang, Jiayi Eris, Sultanum, Nicole, Bezerianos, Anastasia, and Chevalier, Fanny. Dataquilt: Extracting visual elements from images to craft pictorial visualizations. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (2020), Association for Computing Machinery, p. 1–13.

[428] Zhang, Leishi, Kuljis, Jasna, and Liu, Xiaohui. Information visualization for dna microarray data analysis: A critical review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 38*, 1 (2008), 42–54.

[429] Zhang, Shengchen, Wang, Zixuan, Chen, Chaoran, Dai, Yi, Ye, Lyumanshan, and Sun, Xiaohua. Patterns for representing knowledge graphs to communicate situational knowledge of service robots. In *Proc. CHI Conf. Hum. Factors Comput. Sys.* (NY, USA, 2021), ACM, pp. 1–12.

[430] Zhao, Jian, Xu, Shenyu, Chandrasegaran, Senthil, Bryan, Christopher James, Du, Fan, Mishra, Aditi, Qian, Xin, Li, Yiran, and Ma, Kwan-Liu. Chartstory: Automated partitioning, layout, and captioning of charts into comic-style narratives. *IEEE Trans. Vis. Comput. Graph.* (2021).

[431] Zhong, Zengsheng, Wei, Shuirun, Xu, Yeting, Zhao, Ying, Zhou, Fangfang, Luo, Feng, and Shi, Ronghua. Silkviser: A visual explorer of blockchain-based cryptocurrency transaction data. In *2020 IEEE Conf. on Visual Analytics Science and Technology (VAST)* (CA, USA, 2020), IEEE, pp. 95–106.

[432] Zhou, Michelle X, and Feiner, Steven K. Visual task characterization for automated visual discourse synthesis. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1998), pp. 392–399.

[433] Zhuang, Mengdie, Concannon, David, and Manley, Ed. A framework for evaluating dashboards in healthcare. *IEEE Trans. Vis. Comput. Graph. 28*, 4 (2022), 1715–1731.

[434] Ziemkiewicz, Caroline, Ottley, Alvitta, Crouser, R Jordan, Chauncey, Krysta, Su, Sara L, and Chang, Remco. Understanding visualization by understanding individual users. *IEEE computer graphics and applications 32*, 6 (2012), 88–94.

[435] Zikmund-Fisher, Brian J, Witteman, Holly O, Dickson, Mark, Fuhrel-Forbis, Andrea, Kahn, Valerie C, Exe, Nicole L, Valerio, Melissa, Holtzman, Lisa G, Scherer, Laura D, and Fagerlin, Angela. Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical decision making 34*, 4 (2014), 443–453.

[436] Zytek, Alexandra, Liu, Dongyu, Vaithianathan, Rhema, and Veeramachaneni, Kalyan. Sibyl: Understanding and addressing the usability challenges of machine learning in high-stakes decision making. *IEEE Trans. Vis. Comput. Graph. 28*, 1 (2021), 1161–1171.