Combining Text and Visuals for Effective Data Communication

by

Chase Stokes

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Information Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Marti Hearst, Chair
David Bamman
Aditya Parameswaran
Vidya Setlur

Fall 2025

Combining Text and Visuals for Effective Data Communication

Abstract

Combining Text and Visuals for Effective Data Communication

by

Chase Stokes

Doctor of Philosophy in Information Science

University of California, Berkeley

Professor Marti Hearst, Chair

Although information visualizations are widely used to communicate data-driven insights, the role of text in these visualizations remains understudied. Titles, captions, annotations, and other text components are pervasive and influential in real-world visualization designs. Despite the prevalence of text, the visualization field has limited empirical evidence, theoretical structure, and design guidance for understanding how text and visual elements work together. This dissertation addresses these gaps by investigating the role of text in visualization from both reader and designer perspectives. The first part of the dissertation presents three empirical studies examining how text shapes reader experience and interpretation. These studies show that readers prefer text-rich visualizations and that text influences some interpretations (e.g., takeaways and perceptions of bias) but not others (e.g., predictions). The second part investigates how visualization designers use and reason about text. Through analyses of published visualizations and interviews with practitioners, this work identifies and examines ten functions of text, four recurring design patterns, and six common challenges designers face when adding text, and the value of writing as a steering mechanism to clarify goals and audience needs within the design process. Together, these findings challenge assumptions favoring minimalist visual designs and demonstrate that text is a multifaceted and critical feature of visualizations. These contributions deepen our understanding of visualization as a multimodal communication form and lay the groundwork for future research, empirical guidelines, and design tools that more fully integrate text and visualization.

For my family, in all senses of the word.

Contents

List of Figures

List of Tables

Acknowledgments

This dissertation was made possible by the many individuals who supported and guided me. It's not often that you get a platform to thank everyone who cares about and supports you, so forgive me the verbosity.

I would first and foremost like to thank Simone Laszuk, who has supported me through every step of this journey. Our years in California are filled with memories I'll always cherish: reading together at Hidden Cafe, sharing cheeky pints at Fieldwork, and driving to small towns all around the Bay Area and further. Thank you for showing up in every situation, always with kindness, generosity, and care. Throughout my PhD, you celebrated each win, listened to countless half-formed ideas, and read more drafts than anyone should reasonably be asked to. You piloted surveys, endured every practice talk, and (once) even monitored when MTurk had finished collecting responses. You supported me through the highs and lows of two cross-country moves, long nights of paper revisions, job applications, and all the uncertainty in between. This dissertation may have my name on it, but it reflects your care and support at every stage. I could not have accomplished all of this without you. Finishing this PhD marks one more step toward the future we've been building together: our home in Chicago, our future bookstore, and, now just around the corner, our wedding. I can't wait for all that comes next with you.

I would be remiss if I did not also thank our two cats, Tuesday and Friday. Once I teach you two to read, you're going to love this dissertation.

I am deeply grateful to my parents, Karen and Jeff Stokes, whose steady encouragement and thoughtful planning made graduate school possible. I grew up in a home filled with laughter and music, and the lessons I learned from each of you shape me every day. Mom, your compassion and genuine interest in people taught me to be kind and considerate. Seeing you talk with absolute strangers and instantly make them feel known and cared for has shaped how I try to show up in the world. Because of you, I'll always choose compassion first, I'll always look for ways to give someone that little extra support, and I'll always meet each person I come across with the same grace and understanding that you exemplify every day. Dad, you showed me how to meet life with humor, even when things feel daunting. I think back sometimes to when we watched old SNL and talk-show clips until the last possible minute before I got in the car to drive to California. You have shown me, in that memory and countless others, how to find light and laughter in every moment. I am always so proud to be your son. You both encouraged all of my interests and passions, giving me the drive and curiosity that shape my work today. You demonstrated what hard work looked like, and you continue to embody those ideals. Your support, advice, reassurance, and humor continue to guide me, and I carry those gifts with me in everything I do.

Thanks also to my brothers, Carl and Curtis, who have been role models to me for as long as I can remember. You never told me which direction to take, but you showed me what was possible and trusted me to figure out the rest on my own. You both made the future feel exciting rather than intimidating, and you never wavered in your belief that I could handle whatever came next. Carl, you were always meant to be an older brother. You approach conversations with the right balance of wisdom and curiosity, and you have always listened to my worries and woes with understanding and care. Nora is unbelievably lucky to have you as a dad. Curt, you've always reminded me how to enjoy the life I'm living and take every adventure as they come. When you first moved to Chicago (and again now that I'm back), you welcomed me into your world with enthusiasm. You have always made me feel like I belonged and have continued to be a critical sounding board for all of my life decisions. Through every stage — childhood, high school, college, graduate school, and now the workforce — you both have been a source of perspective, humor, and sanity. I hope, more than anything, to make you proud.

Thanks also to Anna Stokes and Anna Lyons, who have each brought so much warmth and joy into our family. Thank you for welcoming me so genuinely into your lives. You've always treated me as your own brother, and that sense of belonging has meant more to me than I can say. And a preemptive thank you to Nora, who is already growing into such a sweet and brilliant person. It's such a joy to watch you grow up, and I can't wait to see who you become (perhaps a budding visualization researcher, if I have anything to say about it).

My family would also not be complete without my (future) in-laws Michael, Renata, and Charles, who welcomed me with open arms from the very beginning. One of my favorite memories with you all is from the early days of the pandemic, sitting in the backyard together with takeout from Edzo's in Evanston (RIP). We have found ways to enjoy even the most difficult of times, and I could not be more thankful for your support. Thank you in particular to Charles, who has allowed me to fulfill a lifelong dream of being an older brother. Becoming part of your family has been one of the great joys of my life, and I'm so grateful for all the years ahead in Chicago with you three.

There are too many supporting friends to name, but a few deserve special mention for the outsized roles they've played. All three of these people have been my friends since before we even arrived at Northwestern, and they have supported each step of this journey with enthusiasm and encouragement. Erica Washington, you never take a day off from being one of the funniest and kindest people I know. Since returning to Chicago, I have been overjoyed to fall into old and new routines with you, and I look forward to many more years of friendship. Vicente Rudolph, you were sitting in the living room the moment I got the call from UC Berkeley letting me know I'd been accepted. This memory is made even better by having you there to celebrate it, and I cannot imagine anyone else I would have rather shared in those major milestones. Anthony Kolesky, you gave me endless joy and community during the long months of COVID-19 shutdowns. I'm forever grateful for those many evenings on the balcony, finding new inside jokes and sources of laughter. My birthday fell barely a week after starting freshman year at Northwestern, and I, like many young college students, worried I wouldn't yet have anyone to spend it with. Erica, Vicente, and Anthony were ready to show up, celebrate, and make me feel included. You three have done the same ever since, through college, graduate school, and beyond.

I'm also grateful to a myriad of friends: Lucas and Ally, Maggie and Sofia, Jacob, Jack and Teresa, Dani and Alfie, Flori and Ben, Jess, Haritha, Bella, and Reva. Special thanks to Ellie Taft for spending an entire summer indulging whims around the Bay Area, and to

Austin Biehl for many Sunday nights spent commiserating and many well-spent visits to Seattle. Each of you added something irreplaceable to these years.

An additional thank you to all of my fellow PhD students who truly made my experience not only manageable but always enjoyable: Liza Gak, Lauren Chambers, Naitian Zhou, Cedric Whitney, Emma Lurie, Justin Norman, Sarah Barrington, Satej Soman, Susana Constenla, Simón Ramírez Amaya, Jasmine Shih, Gowri Swamy, Connie Hu, Dan Hickey, Paloma Santos, Tonya Nguyen, Ando Shah, Nikita Mehandru, Janiya Peters, and Yangyang Yang.

Thanks also to the many other friends and family who I have not yet mentioned. It is impossible to count the number of people who have made an impact on my life and supported my personal and professional decisions that led me to this moment. Thank you.

Beyond the friends and family who supported me throughout these years, I am also thankful to the many people in my professional community who shaped my development as a researcher and made this work possible.

I am profoundly grateful to my advisor, Marti Hearst, whose guidance shaped every stage of my PhD. Marti, you consistently pushed me to be a sharper researcher and a more thoughtful scholar. Your ability to guide and shape students into the best versions of themselves is something I feel incredibly lucky to have experienced firsthand. Your support, encouragement, and steady enthusiasm for our work made this dissertation not only possible but better in every way. I am grateful for every critique and every accolade you offered over the past four and a half years; each one made me not only a better scientist but a better person. I knew we were a great team from our very first paper, and that quality of collaboration only increased with each project. There were moments when I needed encouragement, when I wasn't sure how to take a project where I wanted it to go, or when I doubted whether I was capable. Each time, you met me not with blanket reassurance but with the questions, frameworks, and ways of thinking that helped me move forward. You taught me how to navigate those moments on my own, giving me skills I will rely on long after this dissertation is finished. Thank you for all of your time and contributions, innumerable as they are.

There are three other key mentors who stand out for me when I think back on my research career. Cindy Xiong Bearfield was the first visualization researcher I met. Cindy, you not only guided me through my first research experiences, but you also introduced me to the concept of a "School of Information" and encouraged me to consider paths I hadn't known existed. Your support shaped both my undergraduate and graduate trajectories, and I'm deeply grateful for the perspective and guidance you've shared over the years. You continue to mentor researchers and inspire success wherever you go, and I thank you for all of the time you spent mentoring and inspiring me.

I also want to thank Steven Franconeri, who played an equally important role in my early forays into visualization research. The Visual Thinking Lab at Northwestern was an ideal place for me to learn and grow, and Steve, your mentorship continues to influence my career. My time at Becise helped me explore new areas and directions, and I remain forever grateful for all the opportunities and support you've provided.

Vidya Setlur has likewise been an important mentor throughout (and beyond) graduate school. Vidya, you were an essential part of how I navigated my first first-author project and have continued to be someone I turn to for candid advice and perspective. My internship with you at Tableau Research was not only influential but incredibly enjoyable, and your guidance has shaped both my research and career decisions. You have demonstrated what it looks like to lead a team with confidence and care, and I look up to the example you have set. I have so deeply appreciated all of the time you have invested in my work and future.

I also want to thank every collaborator not yet mentioned who contributed to the works in this dissertation: Bridget Cogley, Arvind Satyanarayan, Clara Hu, Anajan Arunkumar, and Lace Padilla. I appreciate every brainstorming session, every meeting that went over time, and every time I asked a question or pitched an idea and was met with the most constructive and productive teams I could have asked for. I appreciate each of your contributions to my personal and professional development, and I look forward to following all of your future works and endeavors with a close eye.

Thank you also to the members of my committee not yet mentioned: David Bamman and Aditya Parameswaran. Over the course of graduate school, I have learned much from you both directly as well as the example you each set in your own work. I appreciate your guidance and the attention you've given to the work and perspective in this dissertation. A last (but by no means least) thank you to Inessa Gelfenboym Lee, who is an essential part of the UC Berkeley School of Information. You are always a warm and friendly face to see in the hallway or to grab a conversation with. You are exactly the kind of steady support a PhD student needs, and I can't thank you enough for all your quick responses and your endless fount of institutional knowledge.

The work in this dissertation was supported by several different areas of funding: UC Berkeley's School of Information fellowship, the Effective Programming, Interaction, and Computation (EPIC Lab) with Data Lab at UC Berkeley, gift funding from the Allen Institute for AI, a series of NSF awards (IIS-2237585, IIS-2311575, III-1900991), and by the NSF Graduate Research Fellowship Program under Grant No. DGE 2146752.

AI Disclosure: ChatGPT 5 was used only to provide feedback on the clarity of my writing; all research and writing is my own work.

Part I

Background and Introduction

Chapter 1

Introduction

Information visualizations are a central tool for communicating key data-driven insights to broad audiences. Visualizations are commonly defined as "the use of computer-supported, interactive, visual representations of abstract data in order to amplify cognition" [41]. This definition emphasizes the visual encodings as the primary means by which readers see structure, recognize patterns, and reason about data. Much of the foundational visualization work [18, 41, 46, 135, 211] aligns with this emphasis, developing detailed theories and guidelines for marks and channels, color use, chart types, and evaluation of visual design choices. By comparison, text components, despite their ubiquity in real-world visualizations, have received far less attention, and we still lack a systematic account of how text and visuals jointly shape interpretation.

In practice, most visualizations interweave graphical and text elements. Titles, captions, and annotations all serve to establish narrative structure, connect data to external knowledge or context, and suggest interpretations of complex visual data. Furthermore, legends, axes, and other seemingly structural elements are likewise composed of text. The prevalence of these text elements is illustrated in Fig. 1.1. Prior work suggests that text is often the first part of a chart that readers view [37]. This initial attention on text elements, particularly titles, shapes what readers remember [8, 24] and what patterns or features they notice [105, 114, 115]. Reading the text information is a large part of how someone viewing a visualization makes sense of the information, hence why I refer to them as "readers" rather than "viewers" throughout this dissertation.

When integrating visualization with text, the text components call for special considerations that are often either overlooked or undervalued. Compared to the deep literature on visual encoding and chart type design choices, there is relatively little empirical evidence about how different kinds of text affect interpretation, how much text readers prefer, or how visualization designers think about text throughout the design process.

Understanding the role of text in visualization is important for readers, designers, teachers, and researchers in the field. Text can influence how comfortable a reader might feel interpreting the data or how much they trust the information provided. Design decisions about text elements are not sufficiently supported by current guidelines and tools, often re-

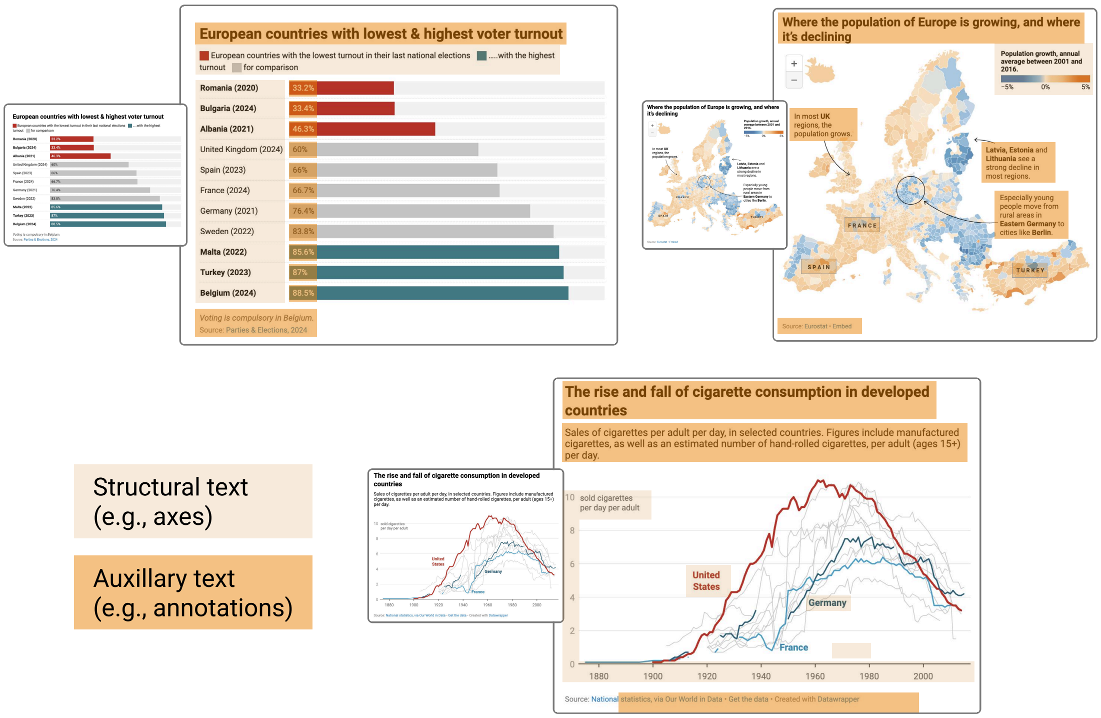


Figure 1.1: Example charts from Datawrapper[1], annotated to illustrate the prevalence of text in visualization design. This includes structural text that is necessary to interpret the data encodings and auxiliary text.

lying solely on a designer's own intuition, preferences, or experience. Teaching visualization design has historically focused on choosing encodings and chart types, with comparatively little attention on creating text for visualizations.

A more systematic account of text in visualization can therefore improve communication, support more reflective design practice, and extend how the field conceptualizes the relationship between language and visual representation. This dissertation attempts to address this issue and provide an empirical and multifaceted depiction of text's role in information visualization.

1.1 Motivating Metaphor

It can be useful to think about the relationship between text and visuals in data communication as a continuum. At one end of this continuum is a purely visual display, where information is conveyed almost entirely through graphical marks and encodings. Aside from

[1] A charting tool currently popular with non-visualization experts; chart images were examples provided at datawrapper.de/charts.

basic axis labels or legends needed for interpretation, the chart contains no additional text. At the other end is a text-only report, in which the same underlying data are described through sentences and paragraphs, without any plotted marks or figures. Between these two extremes lies a wide range of hybrid designs that combine text and charts, as illustrated by the dial in Fig. 1.2.

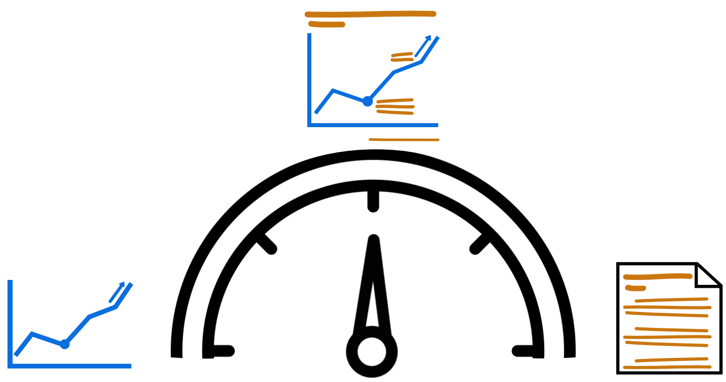


Figure 1.2: Data representations can range from visual-only (left) to text-only (right). This dial illustrates the spectrum of how much text can accompany a visual display.

Most real-world visualizations occupy this middle region. News graphics, dashboards, and scientific figures typically present charts together with text elements that explain what the reader is looking at, highlight notable features, or connect the data to external events. Designers rarely choose either of the extremes. A chart with no explanatory text at all can feel unfinished or underspecified. On the other hand, a long, text-only description can feel disconnected from the visual patterns it describes and require more effort to follow. In practice, designers make many small but consequential decisions about how much text to include, what form it should take, and how it should relate to the data.

When collaborators and I first began exploring the intersection of text and visualization, we used the dial metaphor to expose a basic tension in visualization practice. The visualization field has strong intuitions and norms about how much text is "too much" or "too little," but relatively little empirical or theoretical guidance about where different kinds of designs *should* fall on this continuum. The dial raises foundational questions: How much text do readers actually prefer, and under what conditions? How does the presence, content, and placement of text influence what people take away from a chart? When text expresses a particular stance, how does it affect perceptions of bias or credibility? And on the design side, how do practitioners decide which text elements to include and how to write them?

The studies and analyses that follow are systematic efforts to understand different regions of the dial: how readers respond to more or less text, how specific textual choices (at the *same* point in the dial) influence interpretation and judgments, and how designers select a position on this dial for their designs.

1.2 Specific Research Gaps

There are three primary research gaps that have emerged from this disconnect between visualization research and text elements.

First, we lack **empirical evidence** to guide how different forms of text (i.e., various amounts, content, and placement) affect reader preferences, interpretations, and judgments. Prior work demonstrates that text can be influential but has not examined these influences in a coordinated or comparative way.

The field also lacks a generalizable **theoretical structure** for describing what text does in visualizations. While there are established taxonomies for visual marks and channels [46, 135] or tasks that readers can complete with visualizations [5, 30], we do not have an equivalent vocabulary for the functions of text or for understanding how text and visual elements jointly communicate meaning.

Finally, research provides limited insight into **how designers think about and use text** during the design process. Little is known about how practitioners decide what to write, how they balance text with visuals, or how they manage the constraints and challenges (or even what constraints and challenges *exist*).

Together, these gaps indicate that the field lacks the information needed to guide the use of text components in visualizations. Addressing this area requires studying readers and designers in parallel. The dissertation takes up this challenge by investigating the role of text in visualization from both perspectives.

1.3 Overview of Dissertation Structure

This dissertation is organized into two major parts that reflect the two perspectives needed to understand the role of text in visualization: how readers interpret textual elements, and how designers create and reason about them. Together, these parts build a thorough account of how text functions within visualization design and practice, addressing the three key research gaps.

Part II examines readers' responses to text through a series of empirical studies. Each study was pre-registered, including methods, hypotheses, participant recruitment, and planned analysis. These studies extend our knowledge on how text on a visualization influences reader understandings through a variety of tasks and contexts. In brief summary:

- Chapter 3: Readers prefer text-rich visualizations over minimalist designs.
- Chapter 4: Text influences reader conclusions, with varying effects depending on content and position.
- Chapter 5: Text influences perceptions of designer bias but has only a small impact on predictions.

In more detail, Chapter 3 addresses the central research question: **"What are readers' preferences when viewing information displays with different amounts of text?**

To address this question, collaborators and I conducted an experiment in which participants ranked a series of charts that systematically varied in the amount of accompanying text, ranging from visual-only to text-only displays. Across both studies, readers consistently preferred annotated and text-rich designs, with a substantial minority preferring text-only descriptions over minimalist charts. These findings challenge long-standing assumptions favoring visual minimalism and demonstrate that text plays a central role in how readers experience and evaluate visualizations.

Chapter 4 uses the same experiment to address the research question: **"How do titles and annotations influence the takeaways that readers extract from visualizations?"** We constructed a set of univariate line chart stimuli that varied in text content and placement. Participants were shown a single chart and asked to provide free-response conclusions; these were then coded based on the semantic content of the conclusion [134]. Results indicate that text shapes what readers conclude and that the combination of text content and position informs the degree of influence. Findings from this study provide more precise guidance about when and how specific text elements should be used.

Chapter 5 extends this work to examine additional tasks and contexts, asking two main questions: **"How do titles and annotations influence the predictions that readers make about future data states?"** and **"How does text on a chart influence reader perceptions of author bias?"** Two crowdsourced experiments presented participants with charts that showed two competing groups, with text elements supporting one of the groups. Participants were asked to predict which group would have a greater value at a given point in the chart and to assess the alignment of the visualization designer. The results show that text has little effect on predictions but strongly affects perceptions of author bias, particularly when participants disagreed with the outcome suggested in the text. Implications of these results suggest that text framing for a visualization has a direct and strong impact on how they perceive the neutrality of the data.

Across these chapters, this part of the dissertation provides several important advancements for visualization research and design. We now know that text is part of what readers *want* in a visualization design and that it has varied impacts on their interpretations depending on what task they're completing. This provides concrete evidence to guide long-standing debates about how much text visualizations should include and shows that design choices about text cannot be treated as purely aesthetic or secondary. These studies offer concrete guidelines for designers and researchers alike.

Part III turns the focus to visualization designers specifically and their use of text information. These qualitative works extend the theoretical and practical findings in the field. In brief:

- Chapter 6: An analysis using a framework of ten text functions presented four distinct visualization design patterns.
- Chapter 7: Interviews with visualization designers revealed six key hurdles when designing text elements.

- Chapter 8: An interview study showing that using a lightweight writing step in the design process can focus the process and center the audience in design decisions.

Chapter 6 begins this part of the dissertation by assessing how text supports and influences visualization design practices. This work investigates two research questions: **"What are the *functions* of text in visualization designs?"** and **"What text design patterns emerge across visualizations?"** Through iterative open coding of 120 published visualizations containing more than 800 total text elements, collaborators and I developed a framework of ten functions of text and identified four recurring design patterns that reflect how these functions and other elements of text design (e.g., color) occur in practice. This analysis reveals novel ways that design decisions with text can interact with the chart as a whole. The functions framework also emphasizes the variety of roles that text can play and how combining different text functions can create fundamentally different designs using the same data. This framework provides the conceptual grounding needed to describe text roles systematically, filling a long-standing gap in visualization theory.

Chapter 7 describes the first in a pair of interview studies with visualization designers. This study investigates the research question: **"What challenges arise when visualization designers add text to their designs, and how do they navigate these hurdles?"** Collaborators and I interviewed visualization designers across industries, levels of experience, and design outcomes. We identified six recurring challenges related to choosing how much text to include, what text content to use, and how to implement text elements (e.g., positioning). These challenges highlight critical future directions for research and tool development. Understanding these constraints is essential for developing more supportive workflows and design environments, ultimately paving the way for better visualization designs.

Chapter 8 rounds out this dissertation by examining how language may influence the design process. This chapter examines two research questions: **"How do designers *currently* use writing during the design process?"** and **"What is the perceived *impact* of writing rudders on the design process?"** Through two semi-structured interview studies with visualization designers, we found that writing is not a common step in the visualization design process. However, using writing as a steering mechanism early in the design process can help designers clarify the goals of a design and center their audience's needs or questions.

Across these six chapters, this dissertation provides a strong foundation for current and future study of text and visualization. Synthesis of this work in Part IV Chapter 9 articulates the implications of the findings for visualization researchers, designers, and readers. This synthesis demonstrates that text is not peripheral but rather foundational to visualization, shaping interpretation, credibility, and creative reasoning in ways the field has not previously captured. I outline key future areas of investigation and development in Chapter 10.

1.4 Methodological Approach and Research Rigor

Across the reader-focused studies discussed in this dissertation (Chapter 3, Chapter 4, and Chapter 5), I used a consistent methodological approach grounded in experimental psychology and supported by preregistration, systematic stimulus construction, and, when possible, stepwise statistical modeling. The goal throughout these studies was to isolate the effects of text elements in visualization while ensuring that findings were reproducible, interpretable, and appropriately powered.

The crowdsourced studies followed a classical experimental psychology approach. For each research question, I began by operationalizing the relevant constructs, i.e., identifying how the dependent variables should be measured and which features of the text or visualization needed to be varied. Stimuli were designed to be systematically balanced, with careful control over content, placement, and salience, as detailed in the respective stimuli design sections (Sec. 3.2, Sec. 5.2). Since showing participants multiple versions of the same chart can introduce carryover or comparison effects, these studies primarily used between-participants designs. Each participant viewed a single version of a stimulus, with the exception of ranking tasks which asked for explicit comparisons.

Analyses for Chapter 5 relied on a model comparison framework. When assumptions supported it, I used mixed-effects regression models to account for variation across participants and stimuli and to evaluate whether additional predictors improved model fit. When assumptions (e.g., normality) were not met, other appropriate statistical tests were used based on the distributions of the data. Wherever possible, analysis plans were written in advance and preregistered, and all studies were designed to have adequate statistical power based on formal sample size calculations.

These preregistrations were stored on Open Science Framework (OSF)[2], grouped under the appropriate projects. I wrote the analysis code prior to data collection, and all preregistrations, code, and study materials are publicly archived. This level of transparency was intended to strengthen the interpretability of the findings and to provide a replicable foundation for future work on text in visualization.

To complement the quantitative experiments, several chapters incorporate qualitative methods, including open-ended coding, thematic analysis, and interview-based inquiry. These methods provide insight into how designers interpret, reason about, and use text within their own practice, offering perspectives that cannot be captured through controlled experiments alone.

1.5 Acknowledgment of Collaborative Contributions

All of the work described in this dissertation was completed with invaluable contributions from coauthors, including Marti Hearst, Vidya Setlur, Bridget Cogley, Clara Hu, Arvind Satyanarayan, Cindy Xiong Bearfield, Lace Padilla, and Anjana Arunkumar. Studies and

[2] https://osf.io/, https://osf.io/user/wgm9n?tab=2

analyses are reproduced with their permission. The relevant papers have been edited for clarity and coherence with this dissertation.

Chapter 2

Related Work

This dissertation engages with tensions and gaps present in the field of visualization research. Modern visualization research is often traced back to seminal work from Cleveland and McGill [46]. In this work, authors examined visual "marks" (i.e., the visual shape representing the data item) and "channels" (i.e., the mechanism used to encode data values). By comparing performance on basic perceptual judgments, this study empirically determined a perceptual hierarchy of visual channels. This hierarchy is led by position encodings, supporting the use of bar and dot plots for data communication. These findings laid the groundwork for many future studies examining visual encodings, including replications of the original works [82, 139, 164, 247]. The focus on visual encodings has shaped decades of research, but it leaves open questions about how text elements contribute to interpretation and design. Rankings of visual channels by perceptual accuracy often led to developing new and "improved" guidelines for visualization design [51, 102], though these guidelines are sometimes applied without regard to rhetorical or contextual needs.

There is much more to visualization design than simply using perceptually precise encodings. Bertini, Correll, and Franconeri [19] emphasize that this focus on precision excludes important components of visualization design, such as other low-level tasks that readers may complete [5], affordances of different mark types [64, 157, 206], and the use of rhetoric [90, 189]. Since its origins in perceptual accuracy, data visualization research has extended to many different areas of investigation. This body of work is supported by several of those areas, including the combination of visual and text information, data storytelling, and studies of visualization design practices.

2.1 Visualization and Text

While a great deal of visualization research has focused on perceptions of *visual* information, a growing area of work has focused on the *combination* of text and visuals [81]. This work spans multimodal representations, impacts of specific text elements such as titles, and direct comparisons of text and visual data representations.

Using both written and visual information (multimedia or multimodal communication) can be helpful for learning and comprehension, particularly when text is placed next to explanatory images [84, 141, 259]. This combination may be particularly useful for readers who might typically struggle with data interpretation. In one study examining decision making under uncertainty, participants with lower working memory capacity benefited more from the combination of text and visual information than those with a higher working memory capacity [14]. A dual format might facilitate reasoning by allowing users to offload memory demands onto visual elements while using text for more straightforward processing. This is supported by other work indicating that important information is easier to *identify* with visuals but easier to *comprehend* with text [158].

Text elements of a visualization have a substantial role in directing and attracting attention when viewing a visualization. Titles, paragraphs, legends, labels, and axes are fixated upon within the first three seconds of viewing a chart, typically before the reader examines the data displayed [37]. They also receive the longest cumulative amount of attention across the viewing, indicating that they not only set the stage for the data but also act as a consistent and relevant reference point.

Text elements have a substantial impact on reader interpretations as well. When examining the memorability of real-world visualizations, work by Borkin et al. revealed that content readers recall from visualizations is often based on information they extract from the title [24]. Furthermore, when pairing different titles with the same visualization, participants were more likely to recall information conveyed by slanted framings or emphasis (e.g., emphasizing only part of a chart's message) than the chart's visuals [114, 115]. General narratives describing data patterns can drive people to see those patterns as more visually salient such that they miss other key patterns in the data [250].

Other specific text elements, such as annotations and captions, also affect conclusions drawn from visualizations. When specific features are highlighted in captions, readers were more likely to mention those features in their takeaways [108, 262]. This effect is strongest if the caption refers to salient areas of the chart. Additionally, text information can signal positionally to a reader. Text added to focus the visualization's message by adding emphasis and explanation to certain areas also led to a small number of participants seeing the chart as untrustworthy [1]. Chapter 4 and Chapter 5 follow up on both of these points.

As with the influence of text on reader *actions*, task matters for text's influence on reader *perceptions*. When testing the impact of slanted titles on recalled topics, readers tended to not notice the slant, reporting the visualizations overall as neutral or unbiased even when titles contradicted the messages highlighted in the visualization [115]. Similarly, when answering specific questions about data interpretation, titles (even exaggerated ones) did not have an effect on participant responses but did lead to less attention paid to graphical axes [121, 122].

Text also poses a design challenge, since it takes up significant space in a visualization. Recent studies, including those detailed in this dissertation, reveal that the modern emphasis on minimal design and lack of text may be not representative of participant perspectives. For example, charts that used text to "focus" the view of the data tend to be perceived by readers as clearer and more visually appealing [1, 112]. Simply decluttering a chart's design by removing heavy grid lines or simplifying the color palette improved a reader's experience with the chart. Using text to intentionally describe key parts of the design additionally increased ratings of aesthetics, clarity, and professionalism. Images which have more text are also seen as more informative [7]. Chapter 3 and Chapter 7 follow up on these areas.

However, sometimes *no* visualization is the best representation [81]. Text-only representations of data can help readers recall data trends or specific values better than visualizations [93, 108]. For decision-making, text is sometimes better than visualization representations [95, 140], though the effect is mixed [14].

In several contexts, readers also show a clear preference for text alone over visual *or* mixed representations. For example, 41% of users preferred text-only responses from chatbots compared to text–visual combinations [79], and 10% of participants evaluating scrollytelling designs for data-driven articles chose the text-only condition, which replaced visualizations with written descriptions [144]. Work in this dissertation in Chapter 3 further investigates preferences for text-only representations.

Visualization researchers have demonstrated a growing interest in leveraging text effectively in visualization design [28, 111, 190]. However, the concept of a text's role or function across these studies is often ambiguously defined and context- or task-specific.

2.2 Types of Text for Visualization

Earlier efforts to categorize text for visualization used several different approaches, including drawing on existing task taxonomies for visualization interpretation. Taxonomies in visualization research categorize the elementary tasks that readers complete with charts [5, 123, 196], high-level comprehension tasks [6, 35, 226], and abstract combinations of both [30]. These task taxonomies can help guide designers on how visualizations might support analytical activities.

One of these frameworks in particular is informative to also understanding *text* in visualization design. The abstract task taxonomy proposed by Brehmer and Munzner offers a bridge between low- and high-level task classifications by introducing a multi-level abstract framework structured around the questions of why, how, and what tasks are performed [30]. This work contributes a set of verbs that describe the perspective and goals of users (e.g., discover), search actions (e.g., lookup), and elementary tasks they could perform with the data points (e.g., identify). This encompasses the *why*. The *what* refers to the possible inputs and outputs of the particular task, such as values, structures, or other visualization features. *How* refers to specific actions or interactions from the user (e.g., select, filter).

Rahman et al. [176] built on the verbs used to define the elementary tasks in this abstract task taxonomy (identify, compare, summarize, and present) [30]. Authors conducted an extensive thematic coding of real-world annotations, both visual and text. This work presented a design space of annotation types, including enclosure, connector, text, glyph, color, indicator, and geometric. Using the same "how, why, and what" structure, authors outlined a design space for the analytic purposes of the annotations, the strategies available, and the types of data needed to generate the annotations. The same authors later provided further categorization for text information, comprising *additive* text, which introduces external information, and *observational* text, which describes data features [175].

Similar groupings exist for chart titles. The title of a visualization can be generic (i.e., communicating the basic variables of the chart) or informative (i.e., communicating a message from the chart with explanatory purpose) [131, 241]. For example, a title might simply state the variables shown ("Staff Counts Across Healthcare Systems") or describe a specific trend in the data ("NHS Has Fewer Staff Than Some Counterparts").

Taxonomies of text expand beyond single chart images as well, examining news articles, dashboards, and even spoken explanations of charts. Examination of data-driven news articles [77] found that titles often mirrored headlines by stating issues or emphasizing data features, captions tended to cite sources, and annotations typically labeled data values but sometimes conveyed thresholds or summaries. Prior work on dashboard design [10, 184, 199, 214] also incorporates studies of how text 'blocks' are used to structure and support visualizations. Spoken explanations of data can be broken down into categories including explanations of different chart components, examples for how to interpret charts, descriptions of the chart's construction, and additional external context for the data [253].

A related line of work focuses on the use of text to improve accessibility, particularly through alternative (alt) text for the visually impaired. To support alt text generation, Lundgard and Satyanarayan proposed a conceptual model with four semantic levels:

- **Semantic Level 1 (L1)**: Elemental or encoded aspects of the chart, such as the chart type or variables. Example: *President approval rating over 5 years (2015-2020).*
- **Semantic Level 2 (L2)**: Statistical or relational components, such as a comparison between two points or identification of extrema. Example: *Maximum* or *Approval higher in 2013 than 2009.*
- **Semantic Level 3 (L3)**: Perceptual or cognitive features, such as an overall pattern or changes in trend. Example: *Steep fall slows to a steady decrease.*
- **Semantic Level 4 (L4)**: External context or domain-specific insights, such as real-world events. Example: *President starts popular initiatives against child hunger.*

L1 and L2 are considered *perceiver-independent*, meaning they can be generated directly from the visualization specification or underlying dataset. By contrast, L3 and L4 are *perceiver-dependent*. They require either human or machine interpretation of the rendered chart. For example, "steep fall" requires visual inspection, and domain-specific commentary (L4) requires additional contextual knowledge or expertise.

This model primarily provides guidance for writing alt text but has also been used for evaluating text content in and around visualizations [214, 262], including for the works in this dissertation detailed in Chapter 3, Chapter 4, and Chapter 5. Chapter 6 also incorporates and expands upon the semantic levels and other text classification frameworks.

Across these taxonomies, text is identified as playing many roles related to describing visualization structure, synthesizing data features, and providing external information to further contextualize the data.

2.3 Narrative Visualizations and Data Storytelling

Data stories and narrative visualizations often combine data graphics with textual narration, animation, and other visual devices. Telling a story with data is similar to the process of telling a story with words; both involve understanding the context of the information, focusing attention on important or relevant points, and conveying the story with a structure that engages the audience [112, 116, 258]. Visualizations that draw from narrative structures and rhetoric are useful in many domains, such as education [53] and data journalism [69, 138, 242].

In their influential survey, Segel and Heer [189] identified common genres of narrative visualization (e.g., annotated charts, magazine style, slideshows) and highlighted the continuum between author-driven and reader-driven storytelling approaches. In author-driven visualizations, there is overt messaging, while reader-driven experiences are more exploratory in nature.

Narrative elements do not always result in a higher degree of engagement than non-narrative visualizations; for example, using introductory narratives did not increase interaction in exploratory visualizations [27]. Narrative visualizations do not necessarily improve recall of information [256] or lead to increased attitude changes [128]. The use of storytelling in data communication is not a one-size-fits-all; research in this area continues to explore how different narrative structures affect reader interpretations of data [26, 56, 69].

However, there are many positive effects of using narrative visualization techniques. One study compared data stories to conventional visualizations and reported that data stories improved overall comprehension across multiple tasks [193]. For some simple tasks, narrative visualizations also increased how fast readers were able to come up with the correct answers. In comparison to traditional visualization techniques, storytelling can simplify complex information and improve comprehension by providing additional narrative context and interpretation [61, 193, 221].

Interactivity can also add narrative or storytelling to a visualization. Kim, Reinecke, and Hullman [108] used a trend prediction task to compare how well people recalled data depending on whether they had to first *predict* the trend. Predicting this trend added a reflection component to the visualization that improved recall. Interestingly, making the prediction through text helped people recall the specific data values better than predictions through visualizations. Generally, the insights communicated by narrative visualizations are more memorable and persuasive than statistics alone [116, 124, 189].

A key part of narrative visualizations is the use of text [1, 61, 90, 112]. Text components can provide focus, context, and guidance to the reader as they view the data. In particular, visualization designers can use text to provide structure and framing to the visualization [90] as well as create opportunities for exploration between the data visualization itself and narrative text [120, 147]. In addition to text, narrative visualizations can also use animation to add sequence and structure [40, 44, 240].

The use of storytelling and narrative techniques is increasingly recognized as a key part of how visualization design is taught to novice designers. Teaching narrative infographics helps students learn to communicate insights, not just construct charts according to empirical perceptual principles [80]. The role of narratives or stories in data visualization allows for the development of broader skills alongside visualization techniques [11, 166].

2.4 Visualization Design Tools and Practice

Many systems have been created to help authors coordinate text and visual elements as they structure data stories. Tools such as Kori and related interfaces [119, 120, 147, 213, 260] allow authors to create explicit links between text passages and visual elements, drawing on narrative visualization patterns such as annotation sequences and guided pathways [90]. A few recent systems begin with the designer's core message or written narrative, using natural language processing and generative AI to help authors shape the eventual visualization. For example, Epigraphics [261] takes an author's key message as its starting point and suggests infographic design elements that aligned with this message.

Using natural language as part of a design tool can make visual analysis easier and more straightforward. For example, systems like Articulate [216], DataTone [67], and Eviza [191] allow designers to use their own questions in combination with predefined grammar-based approaches to generate dynamic and interactive data visualizations. InkSight [129] allows visualization creators to augment their iterative sketching practices with generated data facts within a computational notebook environment. Conversational question-answering (QA) pipelines [50, 104] answer users' targeted queries using data underlying the chart. QA interaction mechanisms are increasing in popularity with the prevalence of LLMs [36, 50, 107, 137]. Beyond supporting exploration, QA interfaces also allow visually impaired readers to interrogate visualizations interactively through natural language [107].

A second body of tools focuses on generating text to describe or caption visualizations. Early work on automated descriptions used rule-based systems to produce short summaries of statistical graphics [153]. More recent systems identify data "facts" such as peaks, outliers, or trends and convert these findings into natural language descriptions [43, 174, 198]. Machine-learning-based chart-to-text models such as VisText [219] and Chart-to-Text [98] generate richer text by learning from large corpora of figure captions and summaries. Commercial tools provide similar basic features, such as Tableau's Summary Card [218] and Power BI's insight summaries [150].

AI-supported tools extend text generation into accessibility contexts. Because alt text is often the only way blind or low-vision readers access visualizations [96], recent work explores automatically generating alt text using both rule-based heuristics and neural models [52, 98, 219]. These tools can help authors produce initial drafts more efficiently, though insights from tool use also underscored the importance of human review when text must accurately convey complex data. Other AI-supported design tool features include scaffolding text creation and addition [76, 130], generation or exploration of entire visualizations [33, 223], and detecting misleading design practices in existing charts [132].

Understanding these tools requires grounding them in the realities of visualization design practice. Empirical studies show that visualization design is highly iterative, nonlinear, and shaped by contextual constraints such as audience, data type, and domain considerations [3, 167, 168]. Designers frequently integrate guidelines, heuristics, prior examples, and sketching practices while navigating cycles of critique and revision [13, 20, 169, 237].

To describe and analyze these practices, several frameworks characterize common stages of visualization design. A nested model [154], Design Study Methodology [188], action-oriented models [143], and other frameworks [245] all articulate processes involving understanding the problem, ideating and sketching representations, prototyping, and deploying solutions. In this dissertation, we focus specifically on the Design Activity Framework (DAF), which includes those four overlapping stages [145]: *understand* (i.e., define goals and identify target users), *ideate* (i.e., explore creative approaches to displaying the data), *make* (i.e., use tools to translate ideas into functional charts), and *deploy* (i.e., implement the design in its real-world setting).

Visualization design tools and practices continue to evolve as new technologies introduce different ways of authoring and interpreting charts. Emerging work on large language models and agentic AI systems suggests new possibilities for automatically generating or refining visualizations based on data and intended messages. As these capabilities grow, it becomes increasingly important to ground tool development in accurate representations of designers' real practices [75, 168, 197]. In particular, recommendations for integrating text and visual information must be informed by both empirical evidence on how text shapes interpretation *and* a nuanced understanding of how designers incorporate text elements during their design process.

Part II

Insights from Visualization *Readers*

In the following chapters, I investigate how readers interpret and respond to visualizations and their associated text elements. Visualization research has often prioritized understanding how different components of design (e.g., chart type [64, 248], color associations [185]) affect readers' interpretations of data. However, the role of the language accompanying those visuals is equally important. If visualization research aims to improve data communication, then we must evaluate not only the design of visual elements but also the influence of the text that supports and frames them.

Prior work shows that titles can shape interpretation more strongly than visual information [114], that text is often the most memorable part of an infographic [24], and that many readers even prefer written explanations over visual displays [79]. These findings suggest that text accompanying a visualization has a profound effect on how people understand, remember, and evaluate data.

To more thoroughly explore this area, the following chapters examine how readers respond to text across several dimensions: preferences, interpretations, and perceptions of bias. The first chapter explores readers' preference for text-rich over minimalist designs, revealing a desire for narrative context and explanation. The next investigates how the content and placement of titles and annotations guide interpretation, shaping what information readers take away from a chart. The final chapter examines the limits of textual influence, showing that text can affect perceptions of author bias but has little impact on trend predictions. Together, these studies establish text as a defining factor in how people understand and evaluate data visualizations.

Chapter 3

Readers Prefer Text-Rich Visualizations

This chapter presents two sets of analyses examining reader preferences for different ways of displaying information. Participants were asked to provide preference rankings for a series of charts that varied in the number of text elements. In addition, participants evaluated a text-only representation of the same data, presented as a short paragraph. Alongside rankings, participants provided free-response comments on what they liked and disliked about each display. We found that readers preferred charts with more annotations over those with fewer. Although these richer charts were described as more visually cluttered and somewhat harder to process, participants valued the additional context provided by the annotations. Interestingly, a substantial minority preferred the text-only paragraph. While slower to read, this format offered a more coherent narrative around the data. This chapter combines two previously published studies ([204, 210]) conducted in collaboration with Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti Hearst. For both studies, I served as first author and was responsible for study design, all aspects of the analyses, and the majority of the writing.

3.1 Balancing Text Information in Visualizations

Text reports are often time-consuming to read, whereas charts afford quicker interpretation. However, minimalist charts that rely exclusively on visual encodings may omit critical context, such as changes to a company's leadership, major technological advancements, or global economic events. Visualization designers and researchers often consider visuals to be the better option, but combined visuals and text or text-only displays may have an important role in data communication. In fact, using both written and visual information (multimedia or multimodal communication) can be helpful for learning and comprehension [84, 141, 259].

In several contexts, readers even show a clear preference for text over visual or mixed representations. For example, 41% of users preferred text-only responses from chatbots compared to text–visual combinations [79], and 10% of participants evaluating scrollytelling designs for data-driven articles chose the text-only condition, which replaced visualizations with written descriptions [144]. In decision-making tasks, text alone has been shown to perform similarly to visualizations in terms of decision quality and interpretation accuracy [158]. In that work, while visual displays supported rapid identification of key information, text was more effective for extracting details about such key information. Text also facilitated better recall of specific values than visualization [108].

Work examining bilingual readers' preferences further highlights the role of text-only presentations. When comparing English materials to those in participants' native languages, readers preferred charts with extensive English annotations over other English formats. However, when the same materials were presented in their native languages (Arabic or Tamil), participants overwhelmingly favored text-only representations [9]. Together, these studies suggest that text-only representations of a visualization can be a useful tool for data communication for a broader audience.

Titles, captions and annotations offer designers a means of embedding written context directly into the visualization or highlighting important patterns or data points that warrant attention. Text paragraphs, titles, legends, labels, and axes are among the set of elements fixated upon within the first three seconds of viewing a chart, typically before the reader even examines the data itself [37]. Titles that convey clear information are easier to understand and more visually appealing, and the information presented in titles is often recalled first and foremost [8, 24].

However, long-standing design guidelines typically advise against excessive text at the risk of overwhelming the reader [227, 246]. Many visualization designs aim to stay away from "clutter" introduced by too much text information. From this tension, it follows that text is both an important tool and a potential liability; too little text risks underspecification and ambiguity, while too much text risks being visually displeasing and creating cognitive overload.

In this work, we investigate the core research question: **"What are readers' preferences when viewing information displays with different amounts of text?** These studies set out to understand whether reader preferences indicate an optimal combination of text and visuals. Although such preferences are not necessarily directly related to comprehension or future data use, they provide important insight for improving user experience and for designing more effective integrations of text and visual information.

3.2 Stimuli Design

We selected univariate line charts as the basis for our stimulus set. Line charts are among the most widely used basic chart types, and univariate versions in particular provide ample blank space for adding annotations. Other simple chart types, such as bar or pie charts, offer less open real estate for annotation placement. Because our aim was to test at what point additional text becomes excessive, line charts provided the most suitable foundation.

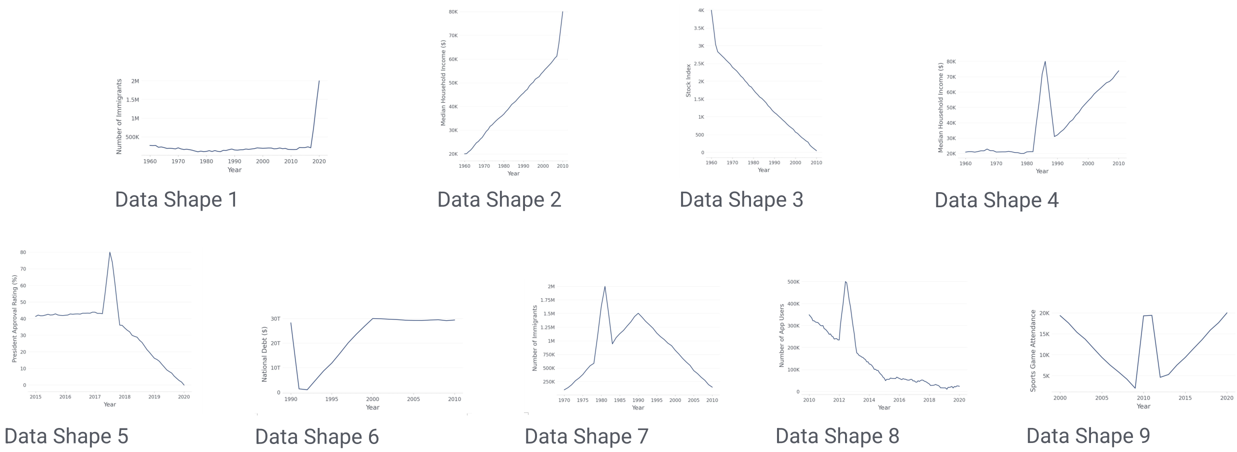


Figure 3.1: Nine univariate line charts generated for this study.

Stimuli for this study build directly on prior work by my collaborator Vidya Setlur [105]. Kim et al. evaluated a set of caption templates designed to highlight different features of a chart, including extrema, trends, inflection points, and individual data points. To test these captions, they generated univariate lines and asked crowdworkers to report feature salience information using bounding boxes. We adopted their generation procedures and their method of predicting salience.

To keep the design space tractable, we created a set of nine charts that contained at most two trends (increasing, decreasing, or flat). Each chart also included a single distinctive feature in the positive direction (e.g., a spike). This approach ensured global shapes with sufficient variation to approximate realistic data trends. We refer to these patterns as data shapes. The complete set of charts, also referred to as *variants*, is shown in Fig. 3.1.

After generating the data shapes, we refined the charts according to recommended visualization design practices [17, 60, 111, 112]. We lightened gridlines and axes ticks while darkening the data line itself, which was rendered in a dark blue consistent with common visualization defaults (e.g., Tableau) and consistent with the Qualtrics survey instrument used for the study. This consistency addressed possible concerns for participants with astigmatism [133].

Chart topics were selected by randomly sampling domains and value ranges from the MassVis dataset [25]. The x-axis displayed time ranges drawn from the overall period of 1900 to 2020. We chose 2020 as the cutoff to avoid manually incorporating the pronounced data fluctuations associated with the COVID-19 pandemic.

3.2.1 Text Content

In this study, we applied the semantic level framework from Lundgard and Satyanarayan [134], which identifies four major categories of text used to describe visualizations: encoded, sta-

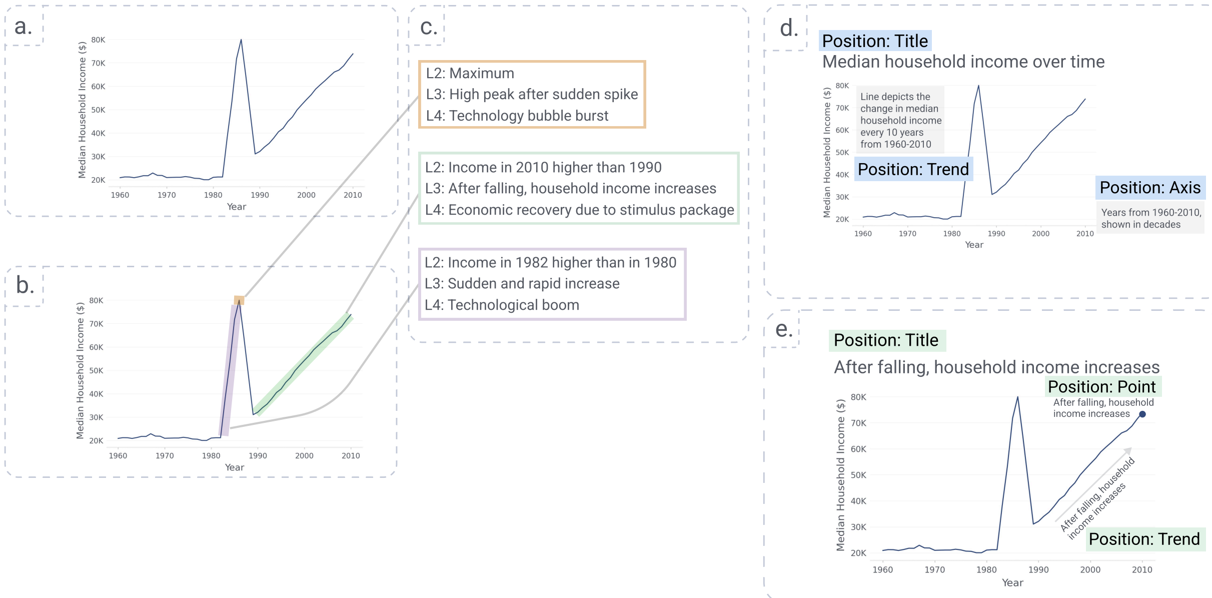


Figure 3.2: Process for adding text to line charts. The initial chart is shown in (a). In (b), red, green, and blue indicate the top three salient features, respectively. In (c), we created potential L2-L4 annotations for each salient region. In (d) and (e), different text position options are shown and labeled for L1-L4.

tistical, perceptual, and contextual. Specific definitions and examples of each level can be found in Sec. 2.2.

For each line chart, we wrote annotations at all four levels. These were authored manually rather than computationally generated, since our goal was to ensure “natural sounding” language, particularly for the perceiver-dependent L3 and L4 categories. Although current state-of-the-art LLMs (e.g., Gemini 2.5 Pro, GPT-5, Claude Opus 4.1) may be able to automatically write text at these semantic levels, doing so manually allows for more control over the content and ensures accuracy to the level definitions.

Annotation content was based on the visually salient features of each line chart, shown in Fig. 3.1. L4 annotations additionally drew on related Wikipedia articles to create plausible domain-specific commentary [243]. Because charts often contained multiple possibly salient features, we relied on crowdsourced salience data from Kim et al. [105] to determine which features should be highlighted. In that study, participants identified the three most visually prominent regions of line charts by drawing bounding boxes. Each bounding box was projected onto the x-axis, weighted by rank order (1st, 2nd, 3rd), and smoothed with a Gaussian kernel centered on the interval. Summing across all participants produced a salience distribution, from which the top three features were extracted.

Although their charts differed from ours, the weighting procedure generalized well to identify salient features in our stimuli. This was possible because the charts shared common visual structures such as spikes, slopes, and plateaus, and the Gaussian-weighting method operates on relative salience rather than absolute values. As a result, the same procedure highlighted features in our charts that aligned with responses from participants in Kim et al. [105].

To balance visual salience across the semantic levels, we ensured that each semantic level was equally likely to describe the three salience rankings across the stimulus set of nine charts. For example, in data shape 4, the L2 text referred to the most salient feature (the spike in 1986), whereas in data shape 8, the L2 text referred to the third most salient feature (the slight decrease from 2015 to 2020). Because L1 annotations simply described the chart type or variables, they did not correspond to any particular visual feature and were not included in the salience counterbalancing procedure. Only L2, L3, and L4 were systematically rotated across salient features.

We implemented the counterbalancing using a 3×3 Latin square. Each row permuted the salience ranks (1, 2, 3); within a chart, each semantic level was assigned a unique rank. Repeating the Latin square three times produced assignments for nine charts, such that each level was associated with each salience rank exactly three times. We also randomized chart order, the assignment of semantic levels to columns, and the ordering of rank values within the Latin square.

To verify that this procedure achieved proper balance, we examined the distribution of salience assignments across all nine charts. Each relevant semantic level (L2–L4) was evenly assigned to the different saliency rankings. This ensured that no semantic level was disproportionately tied to highly salient or less salient features.

Finally, we varied phrasing within semantic levels. This variation occurred naturally for L3, which described different patterns and trend shifts, and for L4, which highlighted distinct contextual events across topics. L1 and L2 required more intentional variation. For L1, we alternated between axis descriptors (3 cases, e.g., "Years span 1960–2020, shown in decades"), topic statements (3 cases, e.g., "Sports game attendance over 20 years"), and encoding specifications (3 cases, e.g., "Line depicts number of immigrants over time").

For L2, we balanced extrema (3 cases, e.g., "Maximum") with pointwise comparisons (6 cases, e.g., "Attendance in 2010 greater than 2009"). The smaller number of extrema reflects the fact that not every annotated feature could be meaningfully described as a maximum or minimum; in such cases, a direct comparison between two values provided a more context-appropriate phrasing. Across L1 and L2, participants encountered a range of naturalistic textual expressions.

The complete set of text elements by data shape, semantic level, and visual salience rank is shown in Tab. 3.1.

For each chart in the stimuli set, we also created a text-only variant, as advocated for by our prior work in this area [203]. Two collaborators and I independently drafted textual descriptions of the chart, which were then discussed and synthesized into a single paragraph with feedback from the entire team. From this paragraph, we produced an additional chart containing the full set of annotations needed to convey the same narrative, ensuring dynamic

Table 3.1: Counterbalanced text content for each chart, semantic level, and visual feature according to estimated salience rank (1 = highest ranked, 3 = lowest ranked).

Data Shape	**Semantic Level**	**Visual Salience**	**Text Content**
1	L1	-	Years span 1960-2020, shown in decades
1	L2	1	Number of immigrants in 2012 is greater than in 2011
1	L3	2	Slight uptick in the overall decreasing trend
1	L4	3	More job opportunities and government policy encouraged immigration
2	L1	-	Years span 1960 to 2010, shown in decades
2	L2	2	Income in 2010 greater than 2008
2	L3	3	After steadily increasing, median household income spiked suddenly in 2008.
2	L4	1	Steady increase from 1960-2008 was caused by an increase in base wages.
3	L1	-	Line depicts stock index over a series of 5 decades
3	L2	1	Stock index at 5 in 2010, less than in 1970
3	L3	3	Pivotal moment for stock index in 1962, decreasing rate changed dramatically
3	L4	2	National debt crisis caused steep decrease.
4	L1	-	Line depicts the change in median household income every 10 years from 1960-2010
4	L2	3	Maximum
4	L3	2	After falling, household income increases.
4	L4	1	Technological boom
5	L1	-	President approval rating over 5 years (2015-2020)
5	L2	3	Maximum
5	L3	1	Steep fall slows to a steady decrease
5	L4	2	President starts popular initiatives against child hunger
6	L1	-	National debt over 20 years
6	L2	3	Minimum
6	L3	2	National debt climbs to previous high
6	L4	1	Government instituted restrictions on international borrowing
7	L1	-	Line depicts number of immigrants over time
7	L2	2	Immigration in 1982 higher than 1980
7	L3	1	Second peak occurs about 10 years after the first
7	L4	3	Changes in administrative policy caused spikes in immigration
8	L1	-	Number of app users ranged from 0 to 500K
8	L2	1	Number of app users in 2020 less than in 2015
8	L3	3	Rapid increase in users from 2012 to 2013
8	L4	2	Update to the app introduced large issues for users
9	L1	-	Sports game attendance (in thousands) over 20 years
9	L2	2	Attendance in 2010 greater than 2009
9	L3	1	Attendance plummets over the span of a year.
9	L4	3	High attendance lasts two years due to consecutive season ticket deals

equivalence between text-only and chart-based representations [148]. In other words, we wanted there to be a match between the information conveyed in the text paragraph and the information available in the annotated chart. These versions combined the semantic-level annotations with additional statements from the text-only paragraph.

3.2.2 Text Position

In addition to counterbalancing text content, we also counterbalanced the positions of titles and annotations across semantic levels. Each semantic level was assigned to the title/subtitle position for three charts, and to annotation positions for the remaining charts. Title positions included subtitles, which allowed more opportunities for each semantic level to exist in the title area. The stimuli creation process, including positioning choices, can be found in Fig. 3.2.

L1 text posed a special design challenge. L1 statements describing axes or encodings are unlikely to appear as chart titles in practice (e.g., "Years span 1960-2020, shown in decades"). To preserve ecological validity, we used the content of the L1 text to inform how we positioned and styled it. Topic statements were always placed in the title or subtitle position. Axis and encoding descriptors were placed as annotations with light gray bounding boxes, positioned near the relevant mark (axis or data line, respectively).

For L2–L4, annotations were distributed across three positions: title/subtitle, trend annotation, and point annotation. Point annotations referred to specific events (e.g., "Technological boom") or data values (e.g., "Maximum"). These annotations were accompanied with a circle mark calling out the relevant point location. Trend annotations described comparisons or changes (e.g., "Income in 2010 greater than 2008") and were accompanied with a light gray arrow parallel to the trend. When possible to preserve readability, this text was rotated in line with the trend angle as well. These positions can be seen in Fig. 3.2.

This positional variation allowed us to compare how readers interpreted information depending on both the semantic content and the position of the text. Placement decisions followed established visualization design principles emphasizing legibility, spatial proximity, and visual clarity [112, 227, 246].

Across the stimuli, each level was used in each of these three positions three times. This achieved positional counterbalancing. Though it was not possible to also perfectly counterbalance the visual salience of the annotation positions, we attempted to make them as similar as possible. As a result, trend annotations tended to coincide with slightly more salient features on average ($Mean = 1.8$), compared to point annotations (2.3) and titles (2.3).

3.2.3 Visual Appearance

To ensure that annotation designs were realistic and aligned with professional practice, all visual design choices were made in close consultation with Bridget Cogley, a professional visualization designer and collaborator on this project. Our stimuli followed visual design conventions that emphasized clarity, focus on the data line, and a consistent annotation style linking text to data features [60, 112, 227, 246]. All text met at least AA accessibility standards under WCAG 2.0 testing [232]. Annotation placement was guided by three key principles:

- **Legibility**: Avoid adding annotations too close to one another or overlapping other text or data features [110, 246]. Avoid text rotation angles that reduce readability [113, 224].
- **Visual language**: Maintain consistent annotation styles across charts. We used shaded bands to represent time ranges, arrows to annotate trends, and dots to highlight point-in-time events.
- **Spatial agreement**: Annotations were placed near their referent, with text angles approximating the slope of the data line when possible, while preserving legibility.

All stimuli underwent several rounds of iteration before the designs were finalized.

3.3 Experiment Design

To evaluate the question **"What are readers' preferences when viewing information displays with different amounts of text?"**, we examined two hypotheses (and their inverses) and conducted an exploratory thematic analysis.

- **Text-Only Hypothesis** (H3.1): Readers who prefer textual information will rank the text-only variant higher than charts with little or no text, whereas readers who prefer visual information will rank the minimalist chart variants over the text paragraph.
- **Annotations Hypothesis** (H3.2): Readers who prefer textual information will rank charts with a greater number of annotations higher than charts with little or no text, whereas readers who prefer visual information will rank the minimalist charts over the heavily annotated charts.

3.3.1 Ranking Sets

We constructed two sets of displays from the stimuli described in this chapter, shown in Fig. 3.3. Data shapes for both ranking sets were randomly assigned for each participant. The first set captured the "extremes" of text use, ranging from a chart with no text (beyond axis labels) to a text-only paragraph. This set contained four variants to encompass the broad range. Throughout this chapter, I refer to this set as the "Broad" set and its variants as:

- **No-Text**: Chart with no text beyond axes labels (i.e., no title or annotations).

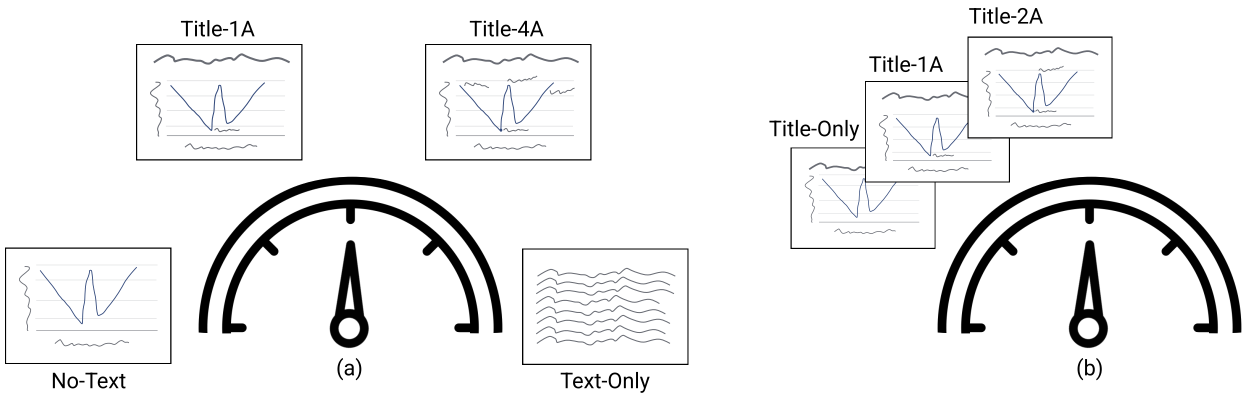


Figure 3.3: Stimuli used for preference ranking and feedback. (a) Broad range of text use, from visual-only to text-only. (b) Stepwise progression in the amount of text.

- **Title-1A**: Chart with a title and a single annotation.
- **Title-4A**: Chart with a high number of annotations to fit the story told by the text-only paragraph. The precise number of annotations varied across these charts, ranging from 3 to 5. On average, these variants had approximately 4 ($Mean = 3.8$) annotations, hence the '4A' designation.
- **Text-Only**: Text paragraph describing the data and related context, highlighting important points ad trends.

The second set presented finer distinctions, ranging from a chart with only a title to one with a title and two annotations. This set had only three variants representing a tighter focus. I refer to this set as the "Focused" set; this set contained:

- **Title-Only**: Chart with only a title.
- **Title-1A**: Chart with a title and a single annotation.
- **Title-2A**: Chart with a title and two annotations.

3.3.2 Survey Procedure

Participants completed a survey consisting of five main sections (Fig. 3.4). One section of this survey is not relevant for the current research question and is instead described in Chapter 4.

Terminology training and comprehension check. To ensure consistent interpretation of text and visual features, participants first viewed a short terminology walkthrough using a sample chart. They clicked through a slide progression which defined a "chart" as a visual representation of data, text as titles and annotations, and visual elements as the data line and other graphical markers (e.g., arrows, dots). Axes were also highlighted and defined. The final slide displayed all terms highlighted together.

Participants then answered 1–2 comprehension questions. Each question showed an illustration of one element (i.e., an axis or a visual feature) and asked participants to identify it. Those who answered the first correctly were not asked the second. If both were answered incorrectly, the survey ended. This check-for-comprehension served to confirm that participants would be familiar with the chart displays used throughout the survey *and* that they were paying attention to the survey instructions and questions.

By calling out and defining text elements of a visualization, this training may have drawn undue attention to text. We acknowledge the possibility that this training may have drawn undue attention to text. To mitigate this, we included axis and visual elements alongside text in the definitions, and used "axis" and "visual" as the answers for comprehension check questions. Although a trade-off, the benefit was consistency in free-response data and participant understanding.

Preference elicitation. After the introduction, participants went on to report their preferences for the randomly selected stimuli. For each ranking set, participants first viewed each variant individually and answered free-response questions about features they liked and disliked. This provided useful qualitative insights into participant preferences and ensured that they had viewed each variant in detail before comparing them to each other in the ranking task.

After reporting their likes and dislikes, participants were instructed to "Rank the images in the order you would prefer to encounter or see them." An optional free-response box allowed them to explain their reasoning. Neither the free-response question nor the ranking task suggested a particular context of use; we preferred to leave this question open-ended to capture general, rather than context-specific, opinions. Following the ranking tasks, participants completed additional questions about their takeaways from the displays; these are described in Chapter 4.

Subjective graphical literacy measure. The fourth section of this survey was a subjective measure of graphical literacy [68], used in prior visualization studies [151, 253]. This measure has been shown to predict performance comparably to objective assessments [65]. Participants rated their ability to work with different chart types (e.g., line, bar, pie) on a 6-point scale (1 = not at all, 6 = extremely good). This measure also contained questions about preferences for visual information (e.g., "When reading books and newspapers, how helpful do you find charts that are part of a story?"; "To what extent do you believe in the saying 'a picture is worth one thousand words'?"). To balance out the focus on visual information, we added a complementary text-focused question: "To what extent do you believe in the saying 'reading expands the mind'?" We calculated the preference scores for each participant by only considering the items relating to the preferences for visual or textual information.

Demographics. Finally, participants reported demographic information, including age range, education level, and prior experience with charts and reading. Chart experience

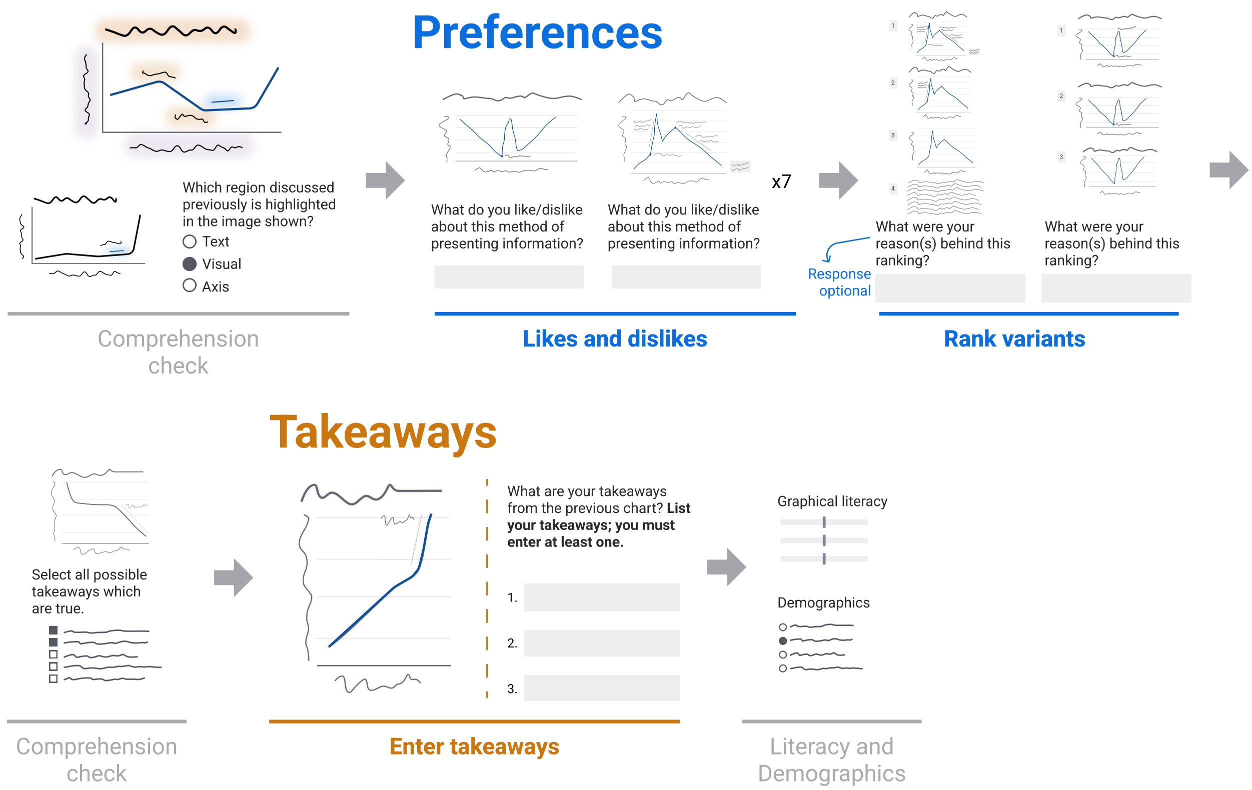


Figure 3.4: Survey flow for assessing preferences (Chapter 3) and takeaways (Chapter 4) for annotated charts. Responses to "likes and dislikes" were the primary dataset for our thematic analysis (reported for each of the 7 variants).

was measured both by frequency of encounter (e.g., "How often do you encounter charts?") and by context (e.g., news articles, government reports). Reading experience was measured for both short-form text (e.g., messages, tweets) and long-form text (e.g., books, magazines). This information can be seen in Tab. 3.2.

3.3.3 Participants

Methods and analyses were pre-registered on OSF[1]. To determine the appropriate sample size for these studies, I conducted a power analysis using G*Power [57]. The choice of statistical test was guided by the type of data collected. Our study design generated two primary data types: (1) ranking data for preference analysis and (2) frequency data for takeaway analysis.

Power analyses were conducted for both statistical tests. For the takeaway task (analyzed via logistic regression), using an odds ratio of 1.5, an alpha level of 0.05, and desired power of 0.8, the required sample size after exclusions was 297 participants. For the preference

[1] https://osf.io/qjhea

Table 3.2: Information about participants' literacy scores and demographics.

(a) Participant demographics.

Participant Information	Count
Age Range	
18-24	3
25-34	80
35-44	105
45-54	67
75-84	1
No answer	3
Education Level	
Less than high school	2
High school graduate	37
Some college	56
2-year degree	25
4-year degree	140
Professional degree	34
Doctorate	5
No answer	3

(b) Participant literacy and preference scores. Higher preference scores indicate a greater preference for visual information; lower scores indicate a preference for text.

Measure	Value
Graphical Literacy	
Mean	46.6
SD	6.71
Min	24
Max (of possible 60)	60
Preference	
Mean	28.9
SD	3.44
Min	16
Max (of possible 48)	40

task (analyzed via Friedman's test and post-hoc Nemenyi testing [170, 194, 263], proxied by a Wilcoxon signed-rank test), using an alpha level of 0.05, desired power of 0.8, and a conservative effect size estimate of 0.2, the required sample size after exclusions was 243 participants. We selected the higher of these estimated sample sizes, at 297.

Anticipating a 30−40% exclusion rate based on prior experience with Amazon Mechanical Turk, we recruited 512 participants [34]. Eligibility criteria required participants to (1) be located in the United States, (2) have a 95% acceptance rate on previous tasks, (3) be fluent in English, and (4) complete the survey on a desktop or laptop computer (i.e., no mobile devices). Participants were compensated at a rate of $15 per hour, consistent with the minimum wage in California at the time of the study, for a total of $4.00 for a 16 minute survey.

After excluding participants who failed any comprehension checks or provided extremely low-quality responses (e.g., unintelligible or off-topic takeaways), the final sample included 302 participants. The majority fell in the 35-44 age range, and most reported completing a 4-year degree. On a scale from 1 (strong preference for text) and 6 (strong preference for visuals) [68], the average preference score was 4.06.

3.3.4 Thematic Analysis

From these 302 participants, we collected a rich set of 2,115 text responses describing perceived likes and dislikes of each representation.

We analyzed these responses using thematic analysis following Braun and Clarke's processes [29, 42]. This procedure began with data familiarization, during which two collaborators read all responses and discussed initial impressions. I then conducted systematic open coding in MAXQDA[2], using an inductive (data-driven) approach so that the codes and themes were derived directly form the data rather than driven by existing theory [45]. Next, I completed axial coding, grouping similar codes into candidate themes. A collaborator reviewed these themes against both the associated codes and the dataset as a whole. Through iterative discussion, these themes were refined, clearly defined, and named.

All counts reported are based on the total sample of 302 participants unless otherwise stated. As shown in Fig. 3.3, one chart variant (chart with a title and one annotation) appeared in both ranking sets, though with different data shapes. Consequently, participants commented on this variant twice: once in the context of the broad visual-text spectrum, and once among more closely related chart variants. Finally, to ensure interpretive robustness, we report only codes that were mentioned by at least 10% of participants ($\geq 31/302$), across all variants. This threshold avoids overemphasizing idiosyncratic responses.

3.4 Ranking Analysis

Overall, participants preferred charts with *more* text elements. For both the Broad and the Focused sets, the highest ranked variant was the one with the most annotations, and the lowest ranked was the one with the fewest.

Notably, Text-Only was ranked above some of the chart variants. Participants with strong baseline preferences for text information ranked Text-Only higher than those with preferences for visuals, supporting previous findings suggesting stable individual differences in preference for text versus visuals.

To analyze the ranking data, we conducted Friedman tests with post-hoc Nemenyi comparisons [170], as outlined in Sec. 3.3.3. We analyzed the two ranking sets (Broad: ranged from No-Text to Title-4A, Focused: ranged from Title-Only to Title-2A).

3.4.1 Visual vs. Textual Preferences

Because our hypotheses included comparisons between participants with stronger preferences for text versus visuals, we derived a composite preference score for each participant. This score combined ratings from the graphical literacy questions and additional questions on reading and chart-viewing frequency from the demographics section. After reverse-coding

[2] https://www.maxqda.com/

where appropriate, higher scores indicated a stronger preference for visual information and lower scores indicated a stronger preference for text.

The maximum possible score was 57 (six preference questions, maximum of 6 points each; three frequency questions, maximum of 7 points each). Participant scores ranged from 16 to 40, with an average of 28.9, as shown in Tab. 3.2.

To create comparison groups, participants in the top quartile of scores (above 31) were classified as the VISUAL group, and those in the bottom quartile (below 27) as the TEXTUAL group. This resulted in 95 participants in the TEXTUAL group, 104 in the VISUAL group, and 103 participants with no strong preference. Analyses focused on the two groups with strong preferences relative to the distribution; participants in the middle quartile were excluded from preference-based comparisons.

3.4.2 Overall Rankings

First, we examined all participant rankings ("Overall" column in Fig. 3.5). There were significant differences across stimuli in both sets (Broad: $\chi^2 = 287.51, df = 3, p < 0.001$; Focused: $\chi^2 = 132.42, df = 2, p < 0.001$). These results indicated that participants ranked the stimuli with sufficient consistency to identify clear preferences among the variants. To compare variants within each set directly, we conducted posthoc Nemenyi tests with single-step p-value adjustment.

In the Broad set, Title-4A was ranked the highest ($Mean\_Rank = 1.60$, $p < 0.001$), followed by Title-1A ($Mean\_Rank = 2.27$, $p < 0.001$). Between the remaining two conditions, participants preferred Text-Only ($Mean\_Rank = 2.84$, $p < 0.001$) over No-Text ($Mean\_Rank = 3.28$, $p < 0.001$), making No-Text the least preferred variant overall. In other words, all chart conditions with text were preferred over Text-Only, but Text-Only was still favored when compared to a purely visual display. Title-4A, the most heavily annotated chart, was ranked the highest, indicating a strong preference for charts with many text elements.

One of our aims with the Broad set was to test for signs of "over-texting," or the possibility that additional text could clutter a chart to the point of lowering its appeal and communicative ability. We did not find evidence of such an effect; in fact, we found the inverse. Even though we had added many text elements with the intent of creating a design with "too much" text, that design was the most preferred by participants. This may be because the designs were carefully constructed in accordance with storytelling and visualization design guidelines. These results suggest that, when text is added thoughtfully, it does not diminish — and may even enhance — participant preferences.

Although most participants ($n = 190; 63\%$) ranked Title-4A highest, data also showed that a substantial minority — 42, or 14% of participants — ranked the Text-Only variant as their first choice. Of these, 25 were in the TEXTUAL preference group and 5 in the VISUAL preference group.

In the Focused set, rankings followed a clear pattern: as more text was added, preference increased. Title-2A was ranked highest ($Mean\_Rank = 1.59, p < 0.001$), followed by Title-

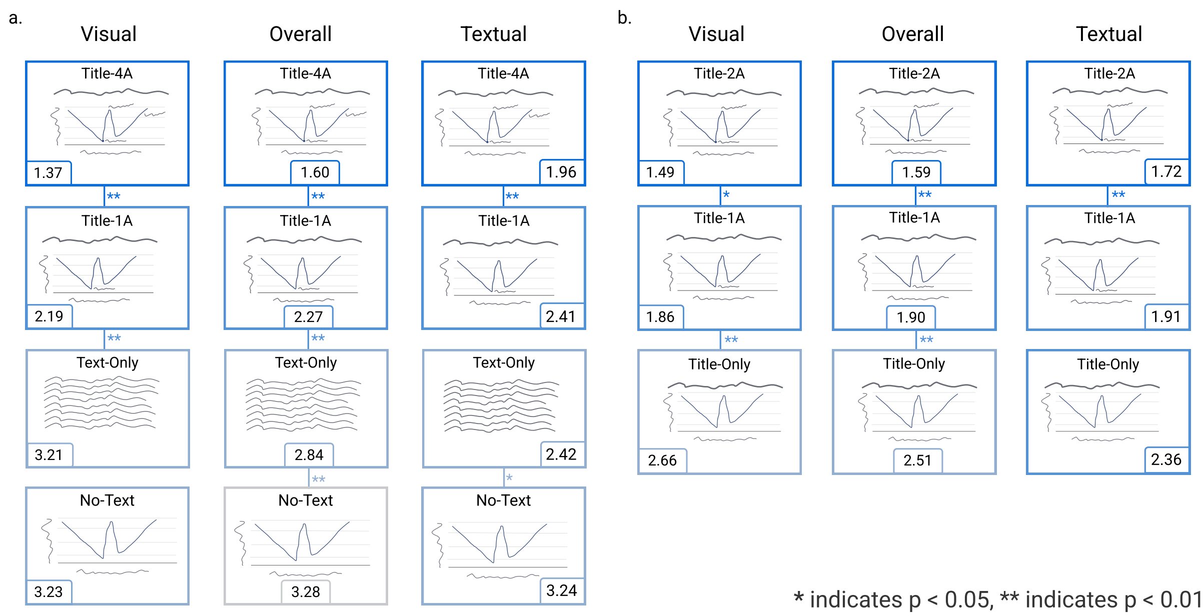


Figure 3.5: Ranking task results, shown for all participants (Overall) as well as the VISUAL and TEXTUAL groups. (a) Broad set. (b) Focused set.

1A ($Mean\_Rank = 1.90, p < 0.001$), and finally Title-Only ($Mean\_Rank = 2.51, p < 0.001$).

3.4.3 Hypothesis Testing

In addition to analyzing the overall average ratings, we tested two pre-registered hypotheses focused on differences between preference groups. One examined preferences for the text-only variant; the other examined preferences for charts with a relatively high number of annotations. While the overall rankings showed that Title-4A was the most preferred chart, the group comparisons revealed how participants made different tradeoffs among the lower-ranked options and how these tradeoffs varied based on underlying preferences.

First, we evaluated the **Text-Only Hypothesis** (H3.1): *Readers who prefer textual information will rank the text-only variant higher than charts with little or no text, whereas readers who prefer visual information will rank the minimalist chart variants over the text paragraph.* This hypothesis refers to the **Broad ranking set**, shown on the left of Fig. 3.3. We found **partial support for both predictions**.

For the TEXTUAL preference group ($\chi^2 = 50.96, df = 3, p < 0.001$), Title-4a remained the highest ranked variant ($Mean\_Rank = 1.93$). Title-1A ($Mean\_Rank = 2.41$) and Text-Only ($Mean\_Rank = 2.42$) were statistically indistinguishable ($p = 1.00$), and both were ranked more closely to Title-4A than in the overall sample ($p = 0.048$; $p = 0.041$). No-Text was consistently ranked lowest ($Mean\_Rank = 3.24, p < 0.001$). Essentially, readers who preferred textual information often ranked the Text-Only variant above the chart with no text.

For the Visual preference group ($\chi^2 = 50.96, df = 3, p < 0.001$), the hypothesis received similar support. Title-4A was again consistently preferred ($Mean\_Rank = 1.37, p < 0.001$). Title-1A ($Mean\_Rank = 2.19$) was preferred significantly more than Text-Only ($Mean\_Rank = 3.21, p < 0.001$). Text-Only and No-Text ($Mean\_Rank = 3.23$) were ranked nearly identically ($p = 1.00$), unlike in the Textual group where No-Text was clearly least preferred. In other words, readers who preferred visual information did rank the minimalist chart variants higher than the text-only condition, particularly in comparison to the Textual group.

Overall, the **Text-Only Hypothesis received partial support**: rankings of the text paragraph varied systematically with the readers' baseline preferences. For text-preferring participants, the paragraph was roughly equivalent to a minimally annotated chart (i.e., ranked higher on average). For visually oriented participants, it was closer in preference to a chart with no annotations at all (i.e., ranked lower on average).

We next evaluated the **Annotations Hypothesis** (H3.2): *Readers who prefer textual information will rank charts with a greater number of annotations higher than charts with little or no text, whereas readers who prefer visual information will rank the minimalist charts over the heavily annotated charts.* This hypothesis centers on the Focused ranking set, the Broad set provides additional context. We found **minimal support overall**: the Visual side of the hypothesis was not supported, and the Textual side received only partial support. This hypothesis focuses more specifically on the **Focused ranking set** shown in Fig. 3.3, allowing us to examine the influence of additive text elements.

For the Visual preference group, the hypothesis was not supported, although rankings did differ significantly ($\chi^2 = 75.25, df = 2, p < 0.001$). Title-2A, which had the greatest number of annotations, was ranked highest ($Mean\_Rank = 1.49, p = 0.028$), followed by Title-1A ($Mean\_Rank = 1.85, p < 0.028$), and finally Title-Only ($Mean\_Rank = 2.66, p < 0.001$). When examining the Broad ranking set, we found a similar pattern: charts with more annotations consistently outranked those with fewer. Thus, even participants who preferred visuals favored more text on their charts.

For the Textual preference group ($\chi^2 = 21.42, df = 2, p < 0.001$), the hypothesis received partial support. Title-Only was reliably ranked lowest ($Mean\_Rank = 2.37, p = 0.005$), but Title-1A ($Mean\_Rank = 1.92$) and Title-2A ($Mean\_Rank = 1.72$) were not significantly different ($p = 0.352$). In other words, readers who preferred text valued annotations over none at all, but they did not show a consistent preference for *more* annotations over fewer. The Broad set for this subgroup also showed a general trend toward favoring more annotations, but this was comparable to the Visual preference group.

The **Annotations Hypothesis received only minimal support**: participants of all preference types tended to favor charts with more annotations, but this pattern reflected a general preference rather than differences based on baseline orientation toward text or visuals.

3.4.4 Summary

Across both ranking sets, participants consistently preferred charts with more annotations. In the Broad set, the most heavily annotated chart (Title-4A) was the clear favorite. The Text-Only variant was also preferred over the No-Text chart, emphasizing that text is a strong component of reader preferences for visual information. The Focused ranking set showed the same pattern, with each additional annotation improving the average ranking.

Preference-group comparisons revealed expected differences only among the lowest-ranked variants (No-Text and Text-Only). Baseline preferences for text or visuals influenced the relative ordering of text-only options but did not alter the broader preference for richer textual support in visualization.

3.5 Thematic Analysis of Preference Comments

In addition to statistical analyses of ranking data, we completed a thematic analysis referencing two types of participant responses: reported likes and dislikes for each variant and the optional elaborations when completing the ranking task itself. These questions and their contexts can be seen in Fig. 3.4.

Analysis of participants' likes and dislikes resulted in three main findings. First, readers commented most frequently on the presence or lack of context and detail. They preferred to be informed, even at the cost of simplicity. Second, readers discussed the story-like component of the text-only variant, making little mention of narrative in relation to the chart variants. Finally, readers showed suspicion around possible misleading elements of the chart or text. Counts of relevant codes for each variant can be found in Fig. 3.6. All participant count information is out of 302 unless otherwise stated. Participants are referred to with an anonymous number (e.g., P1).

3.5.1 Clutter or Context

The most common theme in participant responses was around the balance between *simplicity* and *context*. Participants frequently described charts as simple or cluttered, but the terms were not necessarily positive or negative. "Simple" could mean easy to understand but lacking depth, while "cluttered" could mean visually busy but more informative. The Text-Only variant was rarely framed in these terms, so this theme focuses primarily on the chart variants. Across all chart variants, participants consistently emphasized the importance of context. Ultimately, clutter was less of a concern than whether text served a useful purpose.

Minimal Designs Lack Detail. The minimalist variants (No-Text and Title-Only) were most frequently described as "simple" or "clean" (99 and 77 mentions, respectively). Many appreciated how easy these charts were to interpret (80 and 64 mentions). P3 described them as, *"pretty straightforward and [it] doesn't take long to understand what it's showing."*

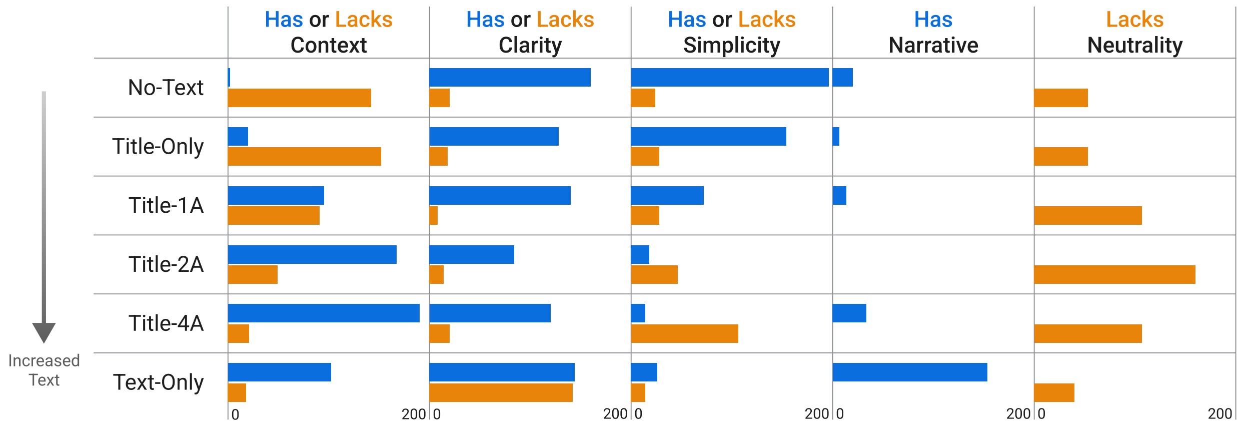


Figure 3.6: Relevant code counts for the themes derived from the thematic analysis. Blue (darker) bars indicate that the variant *has* the respective quality (e.g., context). Orange (lighter) bars indicate that the variant *lacks* that quality. X-axes vary in maximum value according to number of unique participants whose response received the code.

However, these designs also elicited criticism for lacking detail (143 and 152 mentions). As one participant explained, *"I like that I have an image of immigration numbers over time, but I feel like it's not even trying to give me context"* [P64]. 70 participants even specified additional text content that would be helpful for these minimalist designs; *"I don't like that there are no annotations that might explain the spikes and upward/downward trends"* [P212].

Participants also noticed the absence of a title in No-Text, which many found jarring; *"The title is missing. It needs a text title to explain the chart"* [P92]. This reaction suggested that titles serve more than informational purposes. They are part of the schema of a chart and are required to make a chart feel "complete."

Adding Text Adds Context. Adding a single annotation (Title-1A) decreased mentions of simplicity (36 mentions, down from 77 for Title-Only), but the information presented was was still easy to understand (70 mentions, similar to minimalist designs). Many participants saw this variant as balancing visual and text information; P158 responded, *"I like that there are extra facts and text elements to give context to the data visualization. It helps to understand it better and more easily."*

While this context was helpful, it was still too minimal for some participants. Roughly equal numbers described Title-1A as containing sufficient context (96) or lacking it (91). P4 described a need for more specific explanations, *"The title text combined with the added explanation is good, but I'd have questions about why it continued to decrease like that."*

Reactions also depended on the content of the added annotations. Participants particularly disliked redundant information. For example, P106 had an issue with an annotation of extrema, *"Dislike that the maximum is annotated - that is visually obvious, text adds no information."* Usefulness of text, not sheer quantity, seemed to matter most.

Clutter Provides Explanation. As more text was added, perceptions of clutter increased. Title-2A drew 23 mentions of clutter, often linked to the problematic annotations that restated visible features rather than providing new insights. Still, participants valued the added explanatory power when it was present; *"I like this one even more [than Title-1A] because it interprets some of the data and gives a reasoning to the spike"* [P5].

Title-4A contained the most annotations and was the highest-ranked variant overall. Participants praised it for providing rich context (190 mentions) while remaining clear to understand (60 mentions). P20 described this balance, *"It's an easy to read chart with text explanations at the point of interest."* At the same time, clutter concerns persisted (53 mentions). Comments highlighted both the volume of text *and* the accompanying visual markers (19 mentions), which some saw as *"unnecessary visual clutter"* [P233]. On the other hand, several participants welcomed the integration of multiple cues; P45 highlighted this, *"I like that this offers both a visual and written explanation of what is going on."*

Summary. Overall, participants evaluated charts through a tradeoff between clutter and context. Minimalist designs were praised for their clarity but criticized for omitting crucial information. Heavily annotated designs sometimes felt busy but were valued for providing richer explanations. Importantly, perceptions of clutter were less about the *amount* of text and more about whether annotations provided meaningful, non-redundant context.

3.5.2 Text-Only Representation

Participant comments about the Text-Only variant tended to differ meaningfully from those for the chart-based designs. Instead of commenting on aesthetics, they mentioned unique aspects of how text communicates information in the absence of a visual display.

Additional Interpretation Effort. The primary drawback of text-only representations was the lack of a visual trend line (145 mentions). Without a chart, participants had to track the information themselves, constructing and remembering a mental image of the trends. As P37 explained, *"It is hard to keep the years and number of immigrants straight in my mind. This is info that would be better served by being displayed visually."*

This extra *"mental effort"* [P41] was felt by many participants who found the text passage difficult (71) or slow (30) to understand. However, general baseline preferences for methods of information communication made the Text-Only variant uniquely divisive. An almost equal number of participants (72) found the paragraph easy to understand.

Storytelling Capabilities. Participants highlighted a key strength of text communication: the ability to convey a cohesive narrative. While processing the text demanded more cognitive effort, the Text-Only variant provided clear linking of data to events, causes, and effects. The temporal nature of the data afforded storytelling; as P134 put it, the *"narrative clearly describes change over time.. including reasons for the causation."* The emphasis on narrative organization was not mentioned as frequently for the chart variants (11) as it was for Text-Only (23).

Summary. Participants' responses to Text-Only revealed a tradeoff distinct from that observed in chart-based designs. The paragraph demanded more effort and sometimes slowed comprehension, but it also offered narrative coherence and interpretive depth. This theme underscored that, while charts can efficiently convey data trends, text can offer unique affordances for storytelling, context, and explanation.

3.5.3 Misleading and Manipulating

Finally, some participants expressed caution or suspicion toward the stimuli, with 39 comments mentioning that a variant seemed misleading or biased. These concerns appeared most often in charts with annotations (28 mentions) and less frequently for single-modal communication (11 mentions).

Text-Only Limits Bias. For the Text-Only variant, very few (3) participants raised bias concerns. While readers recognized that the hypothetical author shapes the narrative in text, they rarely suggested that the passage was deliberately misleading. Of the chart variants, No-Text and Title-Only also received few comments about bias (4 each). In fact, participants more often emphasized their neutrality. No-Text received 11 mentions of "lack of bias" and Title-Only received 7. The issue posed by the lack of context became a positive attribute in this context; participants felt the minimalist design *"just presents the data"* [P65], and they could *"make [their] own determination as to what is happening"* [P113].

Annotations Narrow Focus. By contrast, charts with annotations led to more reader skepticism and comments about bias or misleading information: Title-1A (8), Title-2A (12), and Title-4A (8). Annotations directed reader attention, possibly giving undue emphasis on specific features, particularly for Title-1A. With only one annotation, the chart seemed to take on too narrow of a focus, with one participant commenting *"hard to believe that that should be the key takeaway given the rest of the chart"* [P123].

With two annotations in Title-2A, some worried the chart was *"purposely trying to lead me into a direction and viewpoint without giving me enough details"* [P65]. Here, the issue was not just what information was included via text but also what was left out. Participants felt the annotations presented a partial story.

Title-4A, which contained the most contextual detail, sometimes provided too much interpretation. While this was the most preferred variant overall, there was a sense of distrust caused by the overt input of the visualization author. P54 commented, *"It makes conclusions based on opinion. It thinks for the reader instead of allowing the reader to do the thinking."*

Summary. In short, participants associated minimal designs with neutrality, while more annotated charts provided valuable context but raised concerns about selective emphasis or over-interpretation. This tension highlights the tradeoff between guiding readers and preserving their sense of autonomy in making meaning from the data.

3.5.4 Analysis of Ranking Justifications

In addition to commenting on what they liked and disliked about each variant, participants were also provided an optional text box to **elaborate on their ranking decisions**, as shown in Fig. 3.4. 200 participants provided at least one response. We analyzed these responses using the same codes developed for the previous analysis on likes and dislikes.

The most common feature that influenced rankings was the presence of context or detail. A majority of participants (159) ranked a variant higher because it offered additional information, and 70 ranked a variant lower because it lacked context compared to alternatives. Ease of understanding also played an important role; 75 participants highlighted clarity and 23 emphasized speed, with 45 linking these qualities directly to the presence of contextual information. P8 explained, *"When the graphs have some text explaining the rise or fall, it gives more context. It makes it easier to understand."*

The visual or textual nature of the variant also played a role in the ranking; 61 participants mentioned preferring a variant specifically because of whether it relied on text or visuals, reflecting their baseline modality preferences. Interestingly, relatively few participants mentioned simplicity (33) or clutter (27) as decisive in their ranking choices.

Participants made relatively few comments regarding simplicity (33) and clutter (27) in comparisons to the responses describing likes and dislikes. While these factors often came up when examining individual variants, they were not significant enough to impact the ranking choices. This further supports the **Clutter or Context** theme; as P55 put it, *"While [the top rank] is too busy, it is also the most informative."* Clutter was a minor issue compared to the benefit of added context.

3.6 Summary

This chapter explored how text information influences reader preferences for different representations of the same data, drawing on both quantitative ranking results and qualitative feedback.

3.6.1 Preferences for Annotation

General guidelines for information sharing and presentation have shied away from adding too much text to visuals at the risk of overwhelming the reader or violating minimalist design principles [227]. Our findings challenge this assumption. More annotation was not penalized; instead, it was often preferred. Readers consistently ranked charts with more annotations above minimalist alternatives. As P31 stated, the combination of text and visuals was *"the best of both worlds,"* and P45 expressed that, *"the chart that combines both visual and written information... helps the widest variety of learning styles to understand the chart."* Although the Text-Only variant was useful for many participants, *"the visuals of the graph + text allow me to more quickly assess the information"* [P10].

This result runs counter to concerns about the potential for such annotations to make the chart appear too busy. On the whole, the presence of clutter does not necessarily make for a poorly designed visualization. However, some participants raised issues with the use of text in visualization designs. Participants valued text that added context or interpretation, but they disliked redundancy in the text annotations. If a feature were visually obvious, they found the additional annotation (e.g., "Maximum") to be annoying and ultimately unnecessary for improving their understanding of the data. Additionally, a minority (11.3%) of readers were aware of possible bias that could be introduced by annotations on a chart and tended to proceed with caution around context shown (or not shown) in a visualization.

3.6.2 Preferences for Text

A non-trivial minority of participants (14%) preferred the Text-Only paragraph over a chart. For these readers, the text was *"easier to understand... than the busy graph"* [P15]. This preference was particularly evident in the TEXTUAL preference group; the Text-Only variant was ranked higher than the No-Text variant. Participants in the VISUAL preference group, on the other hand, ranked Text-Only and No-Text variants similarly. P178, who preferred the text-only variant, explained that they *"like to read the words and use my own imagination for the imagery."* Text offered a story and narrative structure which may be more difficult to convey in a single, static visualization.

This difference in baseline preferences was acknowledged by participants: *"I like this because it's very clear and gives me the information I need to know. I'm content with this. A more visual learner, however, would probably prefer the graph,"* [P139]. Comments like this reinforce findings from prior work and underscore the importance for visualization research to consider the text-only case when comparing visualization options [79, 203].

Beyond the findings from the study, we also found that creating a Text-Only variant was beneficial for the stimuli design as a whole. Creating and structuring a narrative around the data provided a useful starting point for the creation of annotations. We examine this anecdote more empirically in Chapter 8.

Taken together, these results show that most readers preferred annotated charts, and adding more annotations generally increased preferences the displays. Although these findings emphasize the value of annotation for reader experience, they leave open the question of interpretation. The next chapter turns to this issue, asking how annotations actually *affect* readers' conclusions about the data.

Chapter 4

Titles and Annotations Shape Interpretations

This chapter examines how text elements influence what readers take away from visualizations. Participants viewed a chart, randomly selected from a larger stimuli set, with systematically varied text content and positions. They were asked to describe the main message of the chart in their own words. These free-response descriptions were then analyzed to assess the extent to which the content of the text influenced the content of the reader takeaway. In addition, we examined whether different types of content were more effective as a title or as an annotation. We found that the text information had a measurable effect on the takeaways reported by readers. By referring to specific data features, titles and annotations directed attention to those aspects of the chart, and participants tended to mirror the language used in the text. Text that highlighted statistical anomalies or that provided external context were most effective as annotations; text that described the perceptual features of the chart were most effective as titles. This chapter contains the work from a previously published study conducted in collaboration with Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti Hearst [210]. I served as first author and was responsible for study design, all aspects of the analyses, and the majority of the writing.

4.1 Using Text to Guide Takeaways

Findings from the previous chapter suggest that readers prefer visualizations with more annotations than those with fewer. What remains less clear, however, is how these annotations influence the way readers *interpret* the underlying data.

When readers form a conclusion from a visualization, they read the chart in a process akin to reading a paragraph [22, 192, 248]. Prior work demonstrates that chart titles frequently serve as anchors for these takeaways [24, 114, 115], and other textual elements also play an outsized role. Readers often fixate on text within the first few seconds of viewing a visualization [37], and when recalling chart content, they tend to reconstruct narratives by stringing together textual fragments [8].

As in Chapter 3, this study builds again on the earlier work by my collaborator, Vidya Setlur [105]. Results from that work demonstrated that captions shape reader takeaways: when a caption highlighted a specific feature, participants were more likely to include that feature in their interpretations than when viewing charts without captions. These findings have since been replicated [262]; the influence of captions appears to be consistent though somewhat content-dependent.

Building on this foundation, this chapter examines how titles and annotations guide interpretation. We analyzed free-response takeaways from charts that varied in the presence, placement, and content of their text elements. Through this design, we addressed the core research question: **"How do titles and annotations influence the takeaways that readers extract from visualizations?"**

In this chapter, I use the term "conclusion" and "takeaway" interchangeably; both refer to a summary of information that a reader has extracted from a visualization.

4.2 Stimuli Design

The stimuli used in this study were adapted from the full set of line charts described in detail in Sec. 3.2. Briefly, these charts depicted univariate time-series data, generated using methods from prior work [105]. Text for these charts was written based on the semantic level (encoded, statistical, perceptual, and contextual), following the framework introduced by Lundgard and Satyanarayan [134]. For each of the nine data shapes seen in Fig. 3.1, we produced a complete set of single-text variants (one for each semantic level) and paired-text variants representing all pairwise combinations of the four levels. The final set of stimuli tested in this study was evenly balanced across text conditions.

Text elements appeared either as titles or as annotations positioned near relevant data features. Since L1 annotations could not refer to specific data points, we used them to describe the axes or line encodings. L2-L4 text could describe trend or point data features. All visual design choices were developed in collaboration with Bridget Cogley, a professional visualization designer and collaborator, to ensure that the stimuli reflected realistic design practice and adhered to established accessibility and readability standards [187, 232, 233].

4.3 Experiment Design

To evaluate the question **"How do titles and annotations influence the takeaways that readers extract from visualizations?"**, we examined a series of four main hypotheses. The first hypothesis examined the direct impact of text content on reader takeaways, focusing on the semantic levels used to both construct the charts and to analyze participant responses. The remaining hypotheses extended this inquiry to other aspects of how readers considered the information presented, including the participants' takeaways inclusion of information not presented in text or visuals and the participants' self-reported reliance on text versus visual information when constructing their takeaways. The fourth hypothesis examined the effect of text position on participants' takeaways.

- **Semantic Level Hypothesis** (H4.1): Readers' takeaways will be more likely to occur at a given semantic level when the text on the chart also contains that semantic level. This pattern will hold across all semantic levels.
- **Extraneous Information Hypothesis** (H4.2): Readers will be more likely to include information not present in the chart or text (i.e., extraneous information) when L1 content is included in the chart.
- **Reliance Hypotheses** (H4.3): *(H4.3a)* Self-reported reliance on text will be lower when the text is at L1 compared to other semantic levels; *(H4.3b)* Self-reported reliance on text will be higher when the text is at L4 compared to other semantic levels; *(H4.3c)* Self-reported reliance on text will be higher among readers who generally prefer textual information (Textual) compared to those who prefer visual information (Visual).
- **Position Hypothesis** (H4.4): Readers' takeaways will be most likely to be at a given semantic level if the text containing that semantic level is positioned as a title, rather than an annotation.

Together, these hypotheses provide a thorough examination of how text content and position influence the conclusions readers draw from visualizations.

4.3.1 Survey Procedure

Participants completed an online survey consisting of five main sections; the sections relevant for analyses are shown in Fig. 3.4. The initial two sections, comprising comprehension checks and preference-related tasks, are described in detail in Sec. 3.3.2 and are not directly relevant to the current study.

Takeaway training and reporting. Following these sections, participants completed a second comprehension check, which also served as brief training on the study's definition of a "takeaway." Takeaways were defined as "a key fact, point, or idea to be remembered after viewing the chart." In this task, participants viewed a chart with no titles or annotations and were presented with five potential takeaways (e.g., The number of immigrants stayed relatively stagnant from 1965 to 1980.). Three statements were true, and two were false. Because the chart did not contain external content, these takeaways primarily reflected L1–L3 semantic levels. Participants were instructed to select all statements that were true. Those who selected any incorrect options were excluded from the remaining sections of the survey but were compensated for their time spent on the earlier sections.

Participants who passed this comprehension check proceeded to the main takeaway task, shown in Fig. 3.4. In this section, each participant viewed a chart and was informed that they would later be asked to report takeaways on the next page, without being able to revisit the chart. They then advanced to a response page where they were prompted to list one to three takeaways. They then rated the extent to which they relied on textual versus visual information when forming their responses, using a five-point scale from 1 (entirely text) to 5 (entirely visual). This rating applied to their overall set of takeaways.

Subjective graphical literacy measure. Following the takeaway task, participants completed a brief subjective measure of graphical literacy [68], adapted from prior visualization studies [151, 253]. Participants rated their ability to interpret common chart types (e.g., line, bar, pie) on a 6-point scale and responded to items assessing their preferences for visual versus textual information. From these items, we calculated a preference score reflecting each participant's orientation toward visual or textual information.

Demographics. Participants also reported demographic information, including age range, education level, and prior experience with charts and reading. Chart experience was assessed by frequency of encounter and by context (e.g., news, work, or government reports). Reading experience was measured separately for short-form (e.g., messages, tweets) and long-form text (e.g., books, magazines).

4.3.2 Participants

This chapter relies on the same pre-registration[1] and the same power analyses described in Sec. 3.3.3. The recommended sample size was 297. Participants were the same as those described in Sec. 3.3.3. In brief, we collected successful (i.e., no failed comprehension checks) responses from 302 participants. The majority were between 35 and 44 years old and held at least a four-year college degree; more details can be found in Tab. 3.2 I refer to participants using an anonymous ID number (e.g., P6).

4.4 Results

4.4.1 Visual vs. Textual Preferences

We used a subset of responses from the graphical literacy measure to classify participants into preference groups. Participants scoring in the lowest 25th percentile on the visual–text preference scale were categorized as the Textual group, while those in the highest 25th percentile were categorized as the Visual group. The middle 50% of participants were excluded from group-level analyses, since their scores indicated no strong preference for either text or visual information.

4.4.2 Coding Takeaways

Before testing the hypotheses, a collaborator and I independently coded participants' takeaways for three attributes: (1) the semantic level of the takeaway, (2) whether the takeaway

[1] https://osf.io/qjhea

directly matched any of the text provided in the chart, and (3) whether it included information external to both the chart and its text (e.g., real-world events not represented in the visualization or annotations).

When coding semantic levels, we applied the hierarchical scheme proposed by Lundgard and Satyanarayan [134], discussed in further detail in Sec. 2.2. Takeaways containing elements of more than one semantic level were coded according to the *highest* level present. For example, P49 wrote, *"There were restrictions on international borrowing that kept the high debt relatively stable since around 2000."* Although *"debt relatively stable"* is a perceptual observation (L3), the mention of *"restrictions on international borrowing"* refers to an external event (L4). Because the takeaway contained any L4 information, it was coded as L4.

Takeaways were coded as "matches" when they repeated or closely paraphrased text that appeared in the chart. For example, after viewing an annotation that read, "Number of immigrants in 2012 is greater than in 2011," P42 concluded, *"There were more immigrants in 2012 than 2011."* This was almost a direct repetition of the text provided in the chart. Matches also included takeaways that referred to the same data feature even if phrased differently. P167 concluded, *"The chart showed a big drop in attendance from 2010 to 2011,"* which was coded as a match to the annotation, "Attendance plummets over the span of a year."

Because the occurrence of matches and extraneous information was relatively low, we use Maxwell's RE instead of κ to calculate inter-rater reliability (IRR) for those codes. Maxwell's RE is more robust than Cohen's κ for binary codes with skewed category distributions [58]. Overall, IRR was high: $\kappa = 0.82$ for semantic level, Maxwell's RE $= 0.73$ for matching, and Maxwell's RE $= 0.91$ for extraneous information. This resulted in an average IRR of 0.82. A third collaborator coded the conflicts and resolved all but four, which were discussed collectively among coders until consensus was reached.

4.4.3 Takeaway Summary Information

Participants provided a total of 736 takeaways, an average of 1.8 takeaways per participant. We did not instruct participants to give takeaways in any particular order and did not treat takeaways differently depending on how many the participant provided. During the coding process, we excluded 11 takeaways (1.5%) that were nonsense responses or unrelated to the chart shown. Tab. 4.1 shows a summary of the results.

L1 takeaways were relatively uncommon, comprising only 6.5% of all responses (e.g., *"The date range was 1960 to 2020,"* P68) L2 takeaways were more common and reflected specific feature-level observations, (e.g., *"Sports games attendance was lower than 5000 around 2009,"* P75). Most takeaways (61%) were classified as L3, reflecting general statements about data features visible in the line charts (e.g., *"The national debt fell steeply in the early 90s before quickly rising to an all-time high,"* P4). L4 takeaways were similarly prevalent as L2 and most strongly influenced by the accompanying text. An example of an L4 takeaway is, *"The tech boom in the early 1980s led to a huge increase in median household incomes,"* (P215); text on the chart viewed for this conclusion read, "Technological boom."

Approximately one-third of all takeaways (245/736) matched the content of the chart's text almost exactly. Only 20 takeaways included extraneous information that was not present in the chart or accompanying text. For example, P94 viewed a chart depicting median household income and wrote, *"Income increased but so did inflation,"* even though the chart contained no reference to inflation.

Table 4.1: Distribution of reader takeaways by semantic level and condition. Rows indicate the total number of takeaways, those produced after viewing text at the same level as the takeaway, and those which matched the accompanying text content.

	L1	**L2**	**L3**	**L4**
Total takeaways	48	132	446	110
After viewing same-level text	23	73	316	100
Takeaways matching text content	17	60	80	88

Participants reported relying primarily on visual information when forming their takeaways, with an average rating of 4 on a six-point scale (1 = entirely text, 6 = entirely visual). Participants with an overall preference for visual information reported slightly higher reliance on visuals ($Mean = 3.9$) compared to those with a preference for text ($Mean = 3.5$). All analyses were pre-registered on OSF[2]. Some analyses were updated for this dissertation to include additional variables (e.g., takeaway rank) not specified in the pre-registration. These updates did not alter the overall published findings [210]; differences were limited to minor changes in specific p-values and confidence interval ranges.

4.4.4 Hypothesis Testing

We used a set of logistic mixed-effects regression models to test the four hypotheses outlined in this chapter. Each model included fixed effects for the semantic level of the text in the chart, the participant's self-reported reliance on text versus visual information, the rank of the takeaway, and whether the takeaway was a match to the text provided. We found that participants' takeaways were influenced by the position and semantic level of the text in the chart. Text at L2 and L4 had the strongest effect on reader takeaways, followed by L3. L1 text had minimal effect. The placement of the text also played a role in these effects.

Semantic Level Hypothesis

We first examined the **Semantic Level** hypothesis (H4.1): *Readers' takeaways will be more likely to occur at a given semantic level when the text on the chart also contains that*

[2] https://osf.io/qjhea

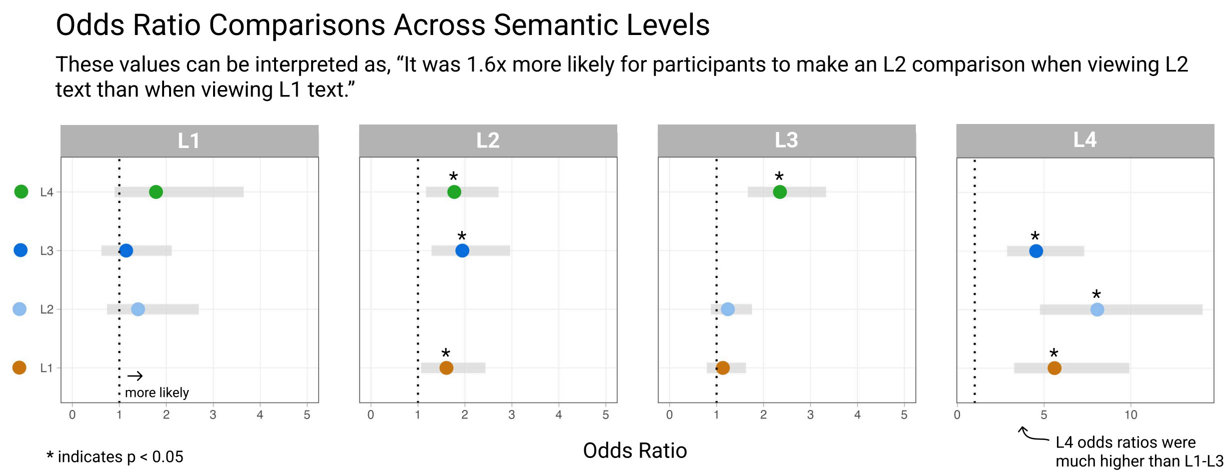


Figure 4.1: Odds ratios across semantic levels. Odds ratios greater than 1 indicate an increased likelihood of making a takeaway at that semantic level; an odds ratio below 1 means they were less likely to do so. Rows indicate the semantic level of participant takeaways when viewing charts containing text at different semantic levels, as indicated by the panel titles. Gray bars represent the 95% confidence interval.

semantic level. This pattern will hold across all semantic levels. To test this hypothesis, we used logistic regression models to analyze the relationship between the semantic level of text on the chart and the semantic level of participants' takeaways. We found **partial support for this hypothesis**, since the influence of text varied across levels. Full sets of comparisons can be found in Fig. 4.1.

We found no evidence that the semantic level of the text predicted the likelihood of an L1 takeaway. The only significant predictor was participants' self-reported reliance on text versus visual information; L1 takeaways were associated with slightly greater reliance on text ($p = 0.033$). As such, we did not find support for this hypothesis for L1.

In contrast, text level had a clear influence on L2 takeaways. Participants were significantly more likely to produce an L2 takeaway when viewing charts with L2 text, compare to any other text level. Specifically, they were 1.6x more likely to produce an L2 takeaway after viewing L2 text than L1 text (95% CI: $[1.1, 2.4]$, $p = .023$), 1.9x more likely than after viewing L3 text (95% CI: $[1.3, 3.0]$, $p = .002$), and 1.7x more likely than after viewing L4 text (95% CI: $[1.2, 2.7]$, $p = .007$). L2 takeaways also tended to appear earlier in participants' lists of conclusions ($p = .030$), suggesting that salient feature-level observations came to mind quickly when interpreting the chart.

Viewing L3 text increased the likelihood that a participant would produce an L3 takeaway but only relative to viewing L4 text (Odds Ratio = 2.2, 95% CI: $[1.7, 3.3]$, $p < 0.001$). L3 takeaways were also less likely to exactly match the information provided in the chart ($p < .001$), likely reflecting their overall frequency in the response dataset.

Finally, L4 text was a strong predictor of L4 takeaways. Participants were significantly more likely to produce L4 takeaways when they were provided with L4 text than when viewing text at any other level: 5.6x more likely than L1 text (95% CI: $[3.3, 9.9]$, $p < 0.001$), 8.1x more likely than L2 text (95% CI: $[4.8, 14.1]$, $p < 0.001$), and 4.5x more likely than L3 text (95% CI: $[2.9, 7.3]$, $p < 0.001$).

L4 information also appeared to be particularly "sticky." Takeaways at L4 were at least 15.2x more likely to match the text provided (95% CI: $[10, 24]$, $p < .001$) compared to takeaways at other levels. Even in charts that combined L4 text with other semantic levels, L4 content made participants significantly more likely to make L4 takeaways than other types of takeaways ($p < 0.001$), suggesting that contextual or explanatory information often overrode perceptual or statistical descriptions when both were available.

Extraneous Information

To better understand how participants incorporated their own external information, we examined the **Extraneous Information Hypothesis** (H4.2): *Readers will be more likely to include information not present in the chart or text (i.e., extraneous information) when L1 content is included in the chart.* We used logistic regression models to analyze the relationship between the semantic level of text in the chart and the presence of extraneous information in the participants' conclusions. **This hypothesis was not supported**.

Contrary to our expectations, we found significant results in the opposite direction. Readers were 9.2x more likely to include extraneous information when L4 text was present in the chart than when L1 text was present (95% CI: $[1.85, 188]$, $p = 0.030$). This wide confidence interval reflects the rarity of extraneous information overall. One possible explanation is that the inclusion of external information within L4 annotations reminded participants of their own related knowledge or experiences, encouraging them to elaborate beyond the data shown.

Reliance on Text

We next examined the **Reliance Hypotheses** (H4.3), which addressed participants' self-reported reliance on text information when making their conclusions. These hypotheses included three subcomponents:

- *(H4.3a)* Self-reported reliance on text will be lower when the text is at L1 compared to other semantic levels.
- *(H4.3b)* Self-reported reliance on text will be higher when the text is at L4 compared to other semantic levels.
- *(H4.3c)* Self-reported reliance on text will be higher among readers who generally prefer textual information (TEXTUAL) compared to those who prefer visual information (VISUAL).

To test these hypotheses, we conducted a one-way ANOVA with Bonferroni-corrected post-hoc comparisons. Participants rated their reliance on text versus visual components of the chart using a five-point Likert scale, where 1 indicated complete reliance on text and 5 indicated complete reliance on visuals. Thus, lower ratings corresponded to greater reliance on text. Overall, we found **minimal support for the Reliance Hypotheses.** H4.3a and H4.3b were not supported, but H4.3c, which compared TEXTUAL and VISUAL preference groups, was supported.

Regarding H4.3a–b, there were significant overall differences in self-reported reliance between semantic levels ($F = 5.81$, $p = .016$, $df = 1$). Participants reported slightly greater reliance on text for charts containing L4 text ($Mean = 3.60, SD = 0.847$) than for charts with L1 text ($Mean = 3.84, SD = 0.841$), although this difference was not statistically significant after correction ($p = 0.170$). None of the other pairwise comparisons between semantic levels reached significance.

In contrast, there was a significant difference between participant preference groups. The TEXTUAL group reported greater reliance on text ($Mean = 3.53, SD = 0.885$) than the VISUAL group ($Mean = 3.87, SD = 0.751$; $p = 0.012$). Participants without a strong preference (BOTH) fell between these groups ($Mean = 3.78, SD = 0.828$), showing a marginal trend towards greater reliance compared to the TEXTUAL group ($p = 0.099$).

Taken together, these findings suggest that self-reported reliance on text is shaped more by *who* the reader is than by *what* the text says. When combined with earlier findings on the influence of semantic level, an interesting discrepancy emerges. Readers' takeaways *were* influenced by L2 and L4 text on the chart, but participants did not report being more influenced by text at those levels. This pattern suggests a disconnect between readers' subjective perceptions and the actual effects of text on their interpretations. As such, behavioral measures of interpretation (e.g., takeaways, decisions) may provide a more reliable indicator of text influence than self-reported information.

Text Position

Finally, we examined the **Position Hypothesis** (H4.4): *Readers' takeaways will be most likely to be at a given semantic level if the text containing that semantic level is positioned as a title, rather than an annotation.* Text on each chart could either appear as a title or as an annotation. Each annotation was assigned a position during the stimuli creation stage, described in Sec. 3.2.2, with examples shown in Fig. 3.2.

For L1, possible positions included 'axis', with the annotation in a gray callout box next to the x- or y-axis, and 'encoding', for which the annotation was positioned in an unanchored callout box near the data line. For L2–L4, annotations could also appear in *trend* or *point* positions. Trend annotations were placed alongside arrows or shaded bands highlighting a range of years, while point annotations were anchored to a specific data point with a small dot matching the line color.

To test the Position Hypothesis, we used logistic regression models to analyze the relationship between the text position and whether the reader's takeaway matched the text in the chart. Unlike the Semantic Level hypothesis, this analysis was restricted to takeaways that explicitly matched the chart text (35% of all takeaways). This approach allowed us to focus on cases in which participants directly incorporated the provided wording, highlighting a significant impact of text information. We found **partial support for the Position**

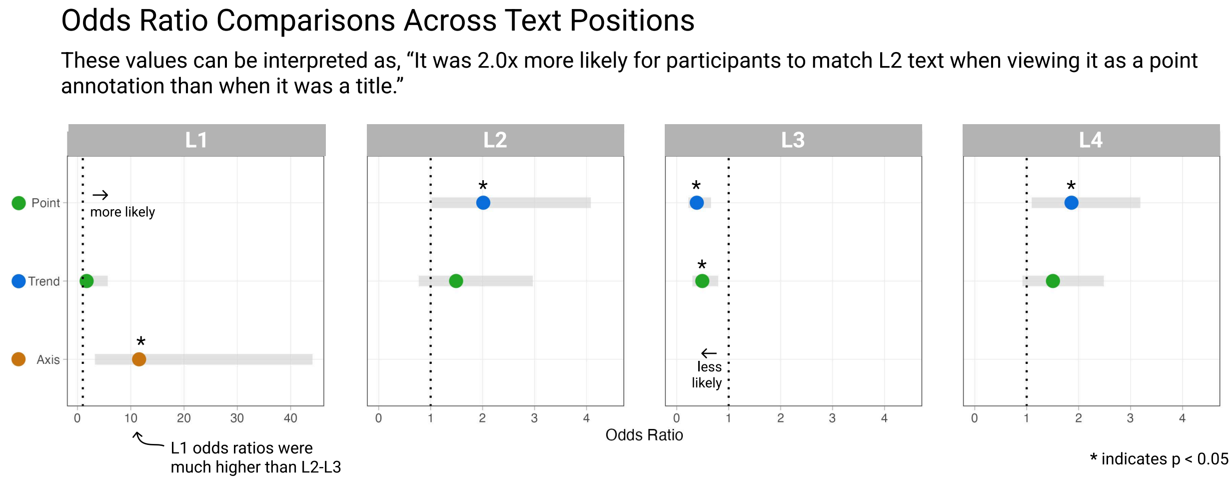


Figure 4.2: Odds ratios across text positions, with "title" as the reference position for all comparisons. Separate analyses were conducted for each semantic level. Missing values for a position indicate that the position was not used for that semantic level (see Sec. 3.2.2 for details). Gray bars indicate the 95% confidence interval.

Hypothesis. Comparisons for each position are shown in Fig. 4.1.

Takeaways matching L1 text were relatively rare, comprising only 30 conclusions (3% of all takeaways). Among these, participants were 11.6x more likely to match their takeaways when the text was positioned by an axis in the chart rather than as a title (95% CI: $[3.27, 44.1]$, $p < 0.001$). This large confidence interval represents the small number of cases but suggests overall that the axis position was more effective than the title position. As such, the hypothesis was not supported for L1 takeaways.

Similarly, we did not find support for this hypothesis for L2 takeaways. Participants were 2.0x more likely to match the annotation as a point than as a title (95% CI: $[1.02, 4.08]$, $p = 0.047$).

L3 text, on the other hand, was most likely to be matched by participant takeaways when it was positioned as a title. They were about half as likely (0.49x) to match L3 text when it appeared as a trend annotation compared to a title (95% CI: $[0.30, 0.80]$, $p < 0.001$) and 0.39x as likely when it appeared as a point annotation (95% CI: $[0.22, 0.66]$, $p = 0.004$). As such, positioning L3 text as a title substantially increased the likelihood of a corresponding L3 takeaway, providing support for the Position Hypothesis.

Continuing the overall pattern, L4 text was most likely to be matched when positioned close to the relevant data feature. Participants were 1.9x more likely to match L4 text when it appeared as a point annotation than as a title (95% CI: $[1.10, 3.19]$, $p = 0.022$).

Overall, with the exception of L3, participants' takeaways were more likely to match text that was positioned closer to the data itself than titles. L2 and L4 annotations often contained specific or data-bound information, such as point comparisons or contextual events, whereas L3 text typically summarized broader patterns or trends. These differences in content may help explain why titles were most effective for L3 but not for the other levels.

4.5 Summary

Authors often include text with charts to provide additional context, and prior work has shown the benefits of aligning text and visuals to emphasize the same message [105]. The findings in this chapter extend this understanding by showing that the semantic level of the text and what kind of information it conveys can also play a crucial role in shaping how readers interpret visualizations.

Readers were most strongly influenced by annotations at the statistical (L2) and contextual (L4) levels. Including external information or context in L4 annotations led readers to incorporate that information into their own takeaways, and L2 annotations prompted participants to comment on specific numerical relationships or comparisons.

In addition to the semantic content of the text, its position within the chart also influenced how readers interpreted information. Takeaways were more likely to match the provided text when annotations were spatially close to the visual elements they described. L1 text placed near chart axes was more frequently recalled than when presented as a title. Similarly, L2 and L4 annotations placed at or near relevant data points were more effective than titles in shaping participant takeaways. Conversely, L3 text, which more often described broader patterns in the chart, was most influential when placed in the title position rather than as an annotation.

These findings highlight an important interplay between spatial and semantic factors: text that refers to *local* features (e.g., a spike or event) is best anchored to those features, while text that conveys *global* patterns (e.g., trends) benefits from a higher-level position in the title area.

4.5.1 Implications for Visualization Design

From a design perspective, these results suggest that the effectiveness of text elements depend on both on what they communicate and where they are placed. Designers aiming to guide reader interpretation or assist readers in interpreting the visualizations can use the following principles:

- **Encoded content (L1)**: Position near relevant axes or encodings.
- **Statistical content (L2)**: Anchor near relevant data point(s) or segments.
- **Perceptual content (L3)**: Present in the title area to describe visual trends.
- **External content (L4)**: Anchor near relevant data point(s) or segments.

These guidelines are most relevant when the designer's goal is for their text to have an influence on how readers interpret the visualization. However if the intent is to support *exploration* rather than *explanation*, a different approach may be preferable. Using only L3 annotations and an L1 title, both of which were shown to have relatively limited effects on takeaways, can preserve flexibility in interpretations while still providing some scaffolding for users. Designer considerations about bias in text information is explored further in Chapter 7 and the impact of this bias is empirically tested in Chapter 5.

4.5.2 Implications for Visualization Research

An important secondary finding highlights the gap between participants' *self-reported* reliance on text and their *actual* behavior. Readers' self-reflections did not always align with the measurable influence of text on their takeaways. This suggests that readers may not be fully aware of how text information can shape their interpretations. This implication underscores the limits of self-report measures for studying visualizations. Future work on text in visualizations should therefore place greater emphasis on observable outputs, such as takeaways, recall, or decisions, rather than relying on introspective measures of influence.

This chapter as a whole examined how text elements shape what readers take away from data visualizations, but interpretation is only one way that text can influence the reading of a chart. In the next chapter, I examine a second type of reader action, *predictions*, and investigate how text influences *perceived bias* in visualization design.

Chapter 5

Text Affects Bias Perceptions but not Predictions

This chapter examines how text affects both the predictions readers make about future data states and their perceptions of author or designer bias. Participants viewed charts containing text that varied in several ways across two studies, including semantic content, position, supported outcome, and degree of bias. The charts were designed to be ambiguous, depicting each outcome as roughly equally likely based on prior trends. Participants first predicted which of two groups they expected to have a greater value at a future point in the chart. They then rated the likelihood that the chart's author favored one group over the other and provided written justifications for their responses. We found that text supporting one group over the other had a minimal effect on how readers perceived the underlying data trends but a substantial effect on how biased they perceived the authors to be. This relationship between the degree of bias in the text and perceived author bias suggests that text in visualizations can strongly influence judgments of credibility, even when the interpretation of the data would be unchanged. Exploratory analyses further revealed an interaction between participants' predictions and their perceptions of bias; participants were less likely to have their prediction influenced by the text if they believed the author to be strongly biased. This chapter contains the work from a previously published study conducted in collaboration with Cindy Xiong Bearfield and Marti Hearst [202]. I served as first author and was responsible for study design, all aspects of the analyses, and the majority of the writing.

5.1 Biases in Visual Data Communication

Building on the previous chapter's findings about how text influences reader takeaways, this chapter examines how text elements in visualizations affect both how readers make predictions from data and how they perceive bias. This combination allows us to also examine how the possible bias presented in a visualization can influence how a reader uses or interprets the data for additional tasks. Within this work, we defined **bias** as "language favoring one side or idea over another without sufficient justification".

Biases can enter a visualization in a number of ways, from the initial perception of the visual data to the final interpretation and decision-making based on the visualization [39]. An emerging body of work has demonstrated that readers can be biased by textual information when making sense of data. For example, readers' recall of key visualization takeaways can be biased by the title to the extent of even contradicting the message present in the visualization [115]. Narratives describing data patterns can drive people to see those patterns as more visually salient such that they miss other key patterns in the data [250]. Some studies have reported that text that signaled the authors' perspective or stance was either not useful or disliked [134]. We found similar results from the thematic analysis detailed in Sec. 3.5.3; a small subset of participants were concerned that the additional text on the visualization was meant to mislead them or present the data in a biased way [204].

However, not all textual information has led readers to assume the visualization is biased. When using pre-generated data facts to explore possible visualizations of a dataset, participants did not report feeling misled by the text presented to them [198]. Additionally, when answering specific questions about data interpretation, titles (even exaggerated ones) did not affect accuracy nor perceptions of chart readability [121].

The studies detailed in this chapter examined the influence of text on reader interpretations of visualized data for two primary tasks: **predictions** about trends in the data and **appraisals** of the bias of the author of the visualization. In doing so, we address two core research questions: **"How do titles and annotations influence the predictions that readers make about future data states?"** and **"How does text on a chart influence reader perceptions of author bias?"**

These two tasks are interesting to study in combination, since prior work indicates that arguments aligned with the participant's own attitudes are interpreted as stronger than arguments that are not aligned [217]. In previous work on crowd-sourced fact-checking, fact-checkers were more likely to correct misinformation from an opposing political party [2]. The degree of perceived bias when evaluating information may influence how that information is evaluated. We expected text to have an effect on predictions, since titles, captions, and annotations are all influential to readers' conclusions from a data visualization [103, 105, 210], as discussed and analyzed in Chapter 4.

5.2 Stimuli Design

The visualizations used in this work depicted a competition between two groups, *Blue* and *Green*, designed to be inherently ambiguous. As shown in Fig. 5.1, the Blue group is consistently in the lead while the Green group steadily increases over time. The slope of this increase was pilot tested to appear likely to reach Blue's value around the prediction point, creating an approximately even split in predicted outcomes [251]. When developing the stimuli, we held the graphical components constant and varied only the text shown on the chart.

Figure 5.1: Study stimuli consisted of line and bar charts that were derived from prior work and designed to have ambiguous prediction outcomes.

5.2.1 Study 1: Semantic Levels

We manipulated two aspects of the text: its *placement* and its *content*. For placement, we tested two common locations: titles and annotations [112, 190, 210]. Figure 5.2 shows where annotations were placed for the Blue and Green conditions. For content, we used different approaches across the two studies. In Study 1, text content was written according to the four semantic levels defined by Lundgard and Satyanarayan [134], also used in Chapter 4 and described in detail in Sec. 2.2: encoded, statistical, perceptual, and contextual.

In total, we constructed 14 distinct text elements for Study 1: two L1 phrases and four sentences each for L2-L4, representing both Blue and Green perspectives for the bar and line charts. These were placed as shown in Fig. 5.2.

5.2.2 Study 2: Levels of Bias

Based on the findings from Study 1 which indicated a possible interaction between predictions and perceived bias, Study 2 focused exclusively on annotations, rather than both titles and annotations. In Study 2, we aimed to control more specifically for the perceived level of bias of the text. To do this, we developed a new set of text stimuli based on annotations written and evaluated by crowdworkers. Figure 5.4 and Fig. 5.5 show the full set of annotations elicited from crowdworkers; Fig. 5.3 shows an example set of the annotations used for Study 2 stimuli conditions.

Crowdworkers were asked to write neutral, low-bias, and high-bias text from both the Blue and Green perspectives. As shown in Fig. 5.4 and Fig. 5.5, this process successfully generated a spectrum of bias levels, confirmed by a second group of crowdworkers who rated each sentence. Higher ratings indicated stronger perceived bias. This process also situated the Study 1 text within this continuum: L2 annotations were generally rated as low-bias and L4 annotations as high-bias.

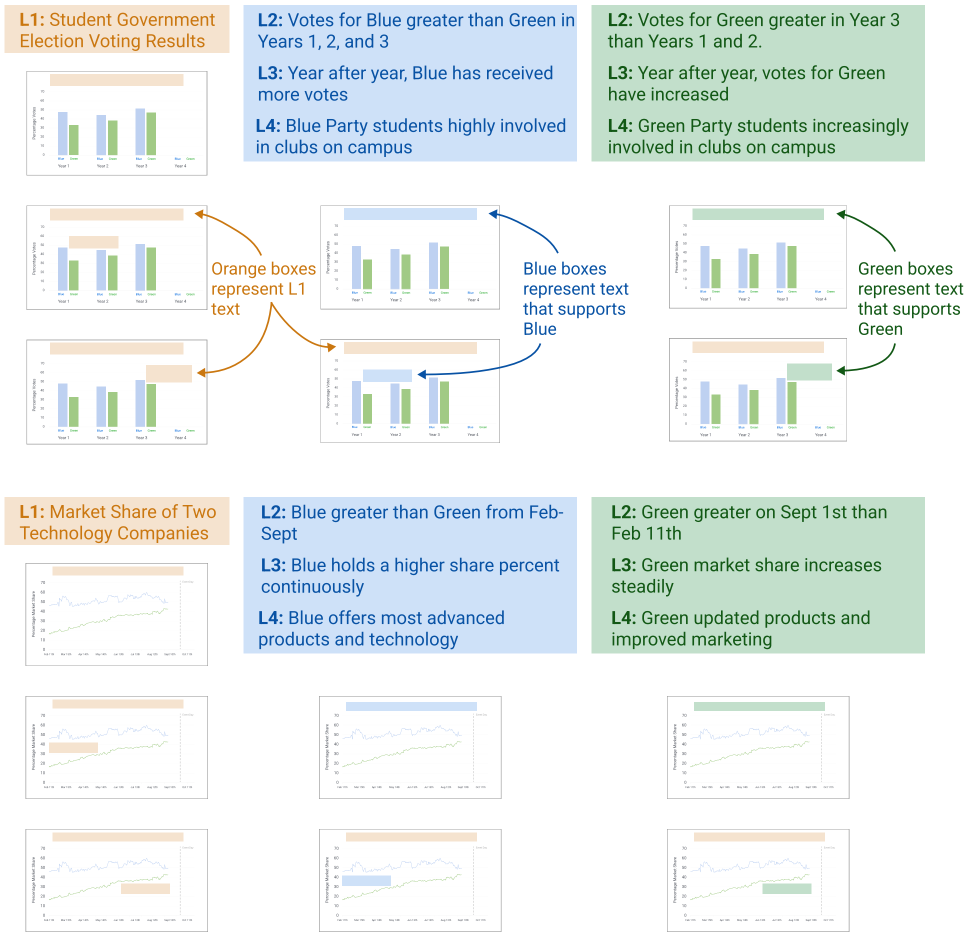


Figure 5.2: Study 1 conditions for bar and line charts. The colored rectangles on the small charts represent text placement: orange for L1 (neutral text), blue for L2–L4 supporting the Blue group, and green for L2–L4 supporting the Green group.

We recruited 20 participants on Prolific with backgrounds in English Language, English Literature, Communications, or Education. These backgrounds were selected based on prior research suggesting that expert writers are better than non-experts at judging the quality of creative work [4, 12].

Participants viewed charts with the L1 title from Study 1 and a single gray box marking where the annotation would appear. They were instructed: “Regardless of your personal opinion, imagine you are a publicist working for the [Blue Group / Green Group]. Help them draft possible annotations to add to this chart that support the [Blue Group / Green Group] winning. The annotation will be placed in the box indicated on the chart and has a character limit.” Neutral annotations were written from the perspective of a “Neutral Organization” but followed the same instructions.

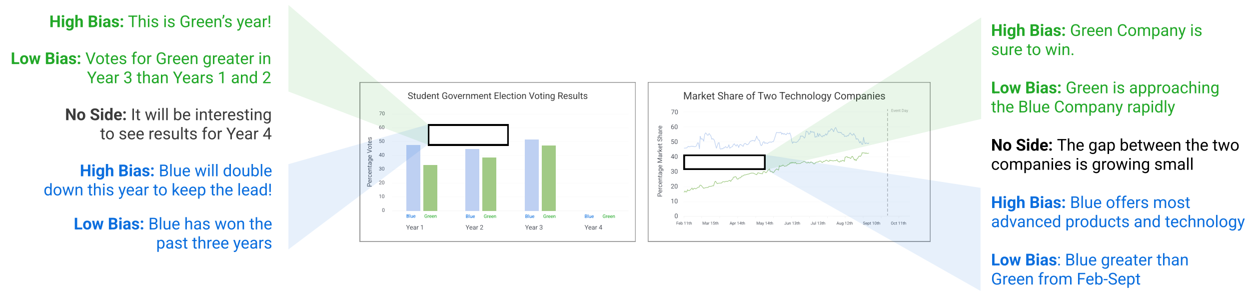


Figure 5.3: Depiction of Study 2 annotation positions and content for both chart types. This figure shows example annotations; four possible annotations were used for each case. Titles were always neutral.

Each participant wrote six annotations (two from each group's perspective). Low-effort responses were excluded if they did not mention the relevant parties or the topic of the chart (e.g., *"ours for the taking"*) or contained incorrect information (e.g., *"Both groups seem to be increasing at the same rate"*). After exclusions, we collected 102 annotations, including the original 12 texts from Study 1 (48 for the bar chart, 54 for the line chart).

To rate bias levels, a separate group of 37 participants on Prolific appraised the 102 annotations on a scale from 0 to 10 in response to the question "To what extent does the annotation favor one side without sufficient justification?" Each participant re-rated four randomly selected annotations at the end of the survey as quality control. Each participant evaluated over 30 annotations (32 for bars, 36 for lines), and each annotation received about ten appraisals on average (bar = 10.42, line = 10.37).

This approach, similar to methods used in prior work [114], offers several advantages. Crowdsourced annotations introduce a broader range of viewpoints than would emerge from a single author or small research team, yielding more diverse and creative interpretations of the data. This diversity can inform more robust experimental stimuli and expose subtleties in how language conveys bias.

As shown in Fig. 5.4 and Fig. 5.5, these appraisals produced a clear distinction between annotations with relatively high and low bias. Text written from a neutral perspective received the lowest bias ratings ($Mean = 2.97$), while text supporting Blue or Green received higher ratings ($Mean = 5.77$). The Study 1 annotations spanned a range of perceived bias ratings, corresponding with their semantic levels ($L2 = 3.07$, $L3 = 4.56$, $L4 = 7.86$). Bar chart annotations tended to use more exclamation marks than annotations written for line charts, but there were no other clear stylistic distinctions between the High-Bias and Low-Bias annotations.

To select the final text stimuli for Study 2, we manually coded annotations that did not mention either group, or mentioned both as having an equivalent chance to win, as "No-Side." The four annotations with the lowest bias ratings formed the No-Side condition. The full set of selected sentences can be found in Tab. 5.1, with examples shown in Fig. 5.3.

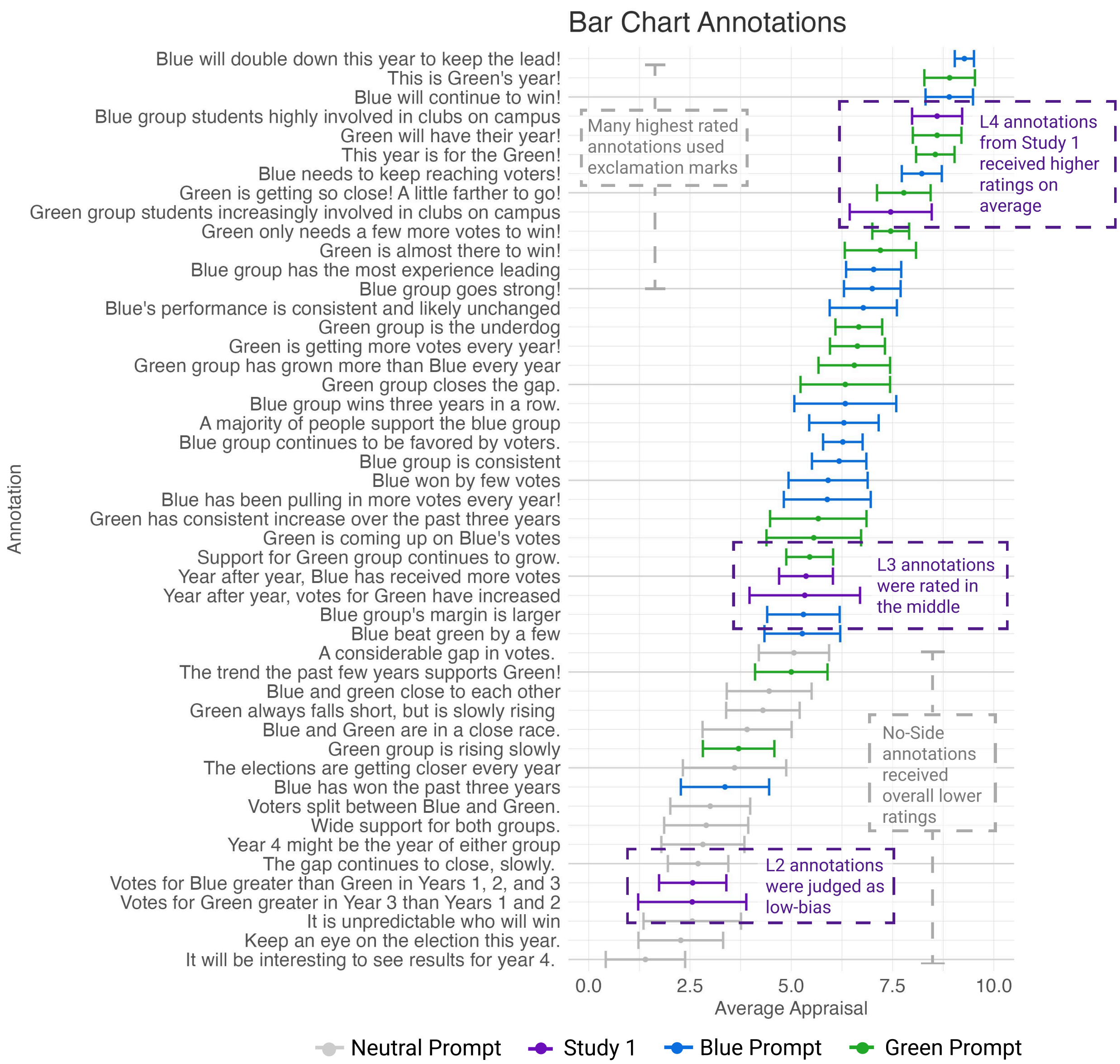


Figure 5.4: Average bias ratings of crowdsourced annotations for bar charts. Error bars show standard error.

From the remaining annotations, the four lowest-rated were selected as the *Low-Bias* condition and the four highest-rated as the *High-Bias* condition. Because these were written from either the Blue or Green perspective, we manually created parallel versions for the opposing group. For example, as shown in Fig. 5.3, a crowdworker-generated High-Bias annotation read, *"Blue will double down this year to keep the lead!"* The corresponding Green version was, *"Green will double down this year to **take** the lead!"* (emphasis added).

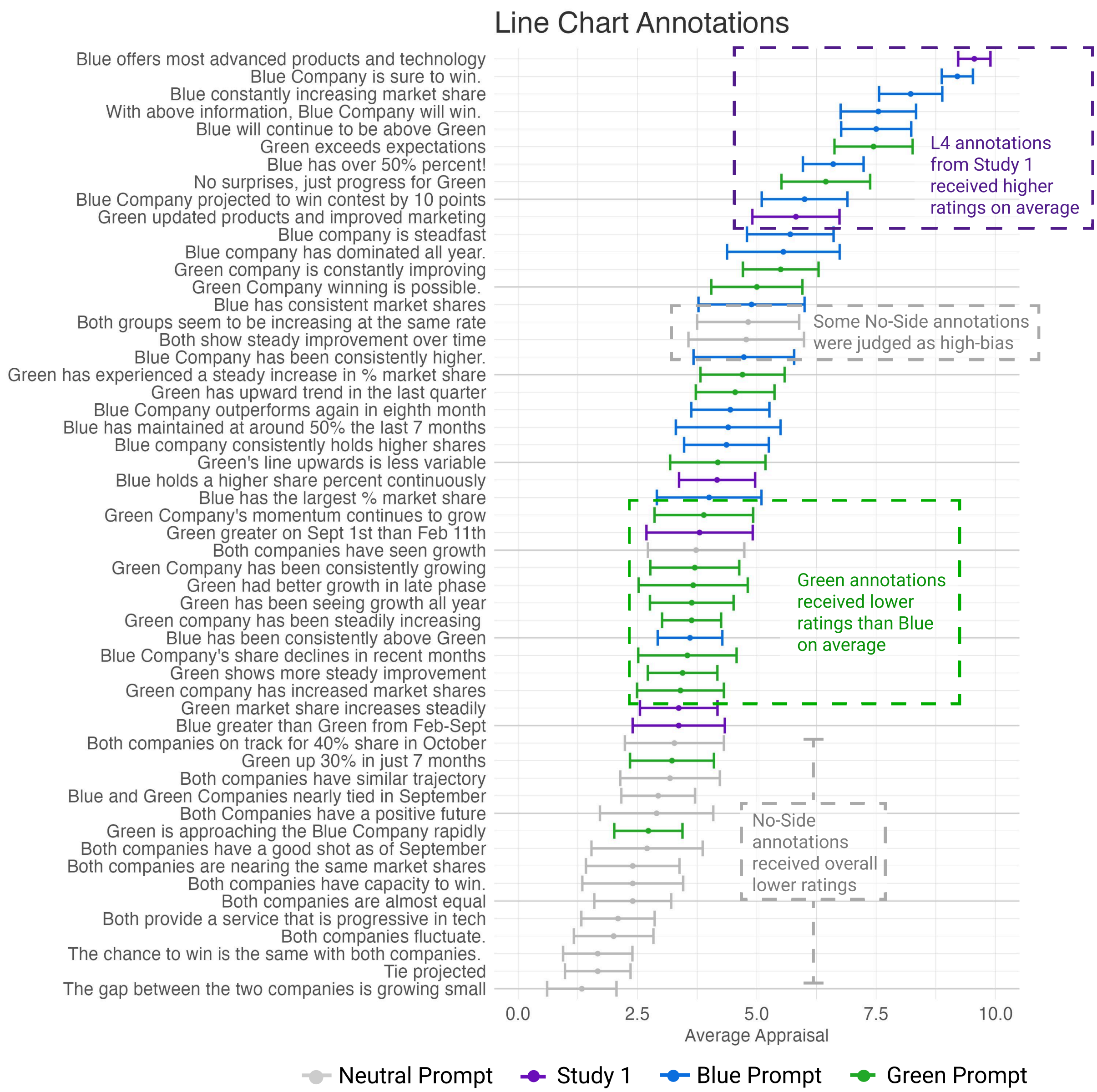


Figure 5.5: Average bias ratings of crowdsourced annotations for line charts. Error bars show standard error.

5.3 Experiment Design

The two studies in this chapter use the ambiguous visualizations shown in Fig. 5.1 to examine how text elements influence data interpretation. Study 1 manipulated both the position and semantic content of text, while Study 2 focused on text content and systematically varied its

Table 5.1: Annotations used for Study 2 stimuli. Blue and Green variants were kept as similar as possible; when changes in wording were necessary, we referenced annotations with similar average bias ratings.

Chart Type	Bias Level	Text
Bar	High	[Blue/Green] will double down this year to [keep/take] the lead!
		This is [Blue/Green]'s year!
		[Blue/Green] will continue to [win/rise]!
		[Blue/Green] group students highly involved in clubs on campus
	Low	Votes for Blue greater than Green in Years, 1, 2, and 3 // Votes for Green greater in Year 3 than Years 1 and 2
		[Blue/Green] has [won/risen] the past three years
		Blue group is winning by a few // Green group is rising slowly
		Blue always wins, but only by a few // Green always falls short, but is slowly rising
	No-Side	It will be interesting to see results for Year 4
		Keep an eye on the election this year.
		It is unpredictable who will win.
		The gap continues to close, slowly.
Line	High	[Blue/Green] offers most advanced products and technology
		[Blue/Green] company is sure to win
		[Blue/Green] constantly increasing market share
		With below information, [Blue/Green] Company will win.
	Low	Blue is consistently above Green // Green is approaching the Blue Company rapidly
		Blue steady at 50% for 7 months // Green up 30% in just 7 months
		Blue greater than Green from Feb-Sept // Green greater on Sept 1st than Feb 11th
		Blue has been consistently above Green // Green market share increases steadily
	No-Side	The gap between two companies is growing small
		Tie projected
		The chance to win is the same with both companies.
		Both companies fluctuate.

degree of bias. In both studies, the text alternated between supporting one outcome (Blue or Green) or remaining neutral, allowing us to measure how textual framing interacted with participants' own interpretations.

We evaluated two core research questions: **"How do titles and annotations influence the predictions that readers make about future data states?"** and **"How does text on a chart influence reader perceptions of author bias?"**. To address these questions, we tested a set of five hypotheses. The phrasing has been adapted from the preregistered materials for clarity and conciseness:

- **Prediction Hypothesis** (H5.1): Readers will more frequently and confidently make predictions consistent with the text's implied outcome than inconsistent with it.
- **Bias Appraisal Hypothesis** (H5.2): Readers will more frequently and confidently appraise author bias in the same direction as the bias conveyed by the text.
- **External Information Hypothesis** (H5.3): Text containing external context will have a stronger influence on both predictions and bias appraisals than other text content.
- **Location Hypothesis** (H5.4): Titles will have a greater effect on predictions and bias appraisals than annotations.
- **Interaction Hypothesis** (H5.5): Participants will rate their predictions as more likely when viewing annotations with a higher degree of bias.

We also conducted exploratory analyses focusing on participants whose interpretations conflicted with the outcome favored by the text. This allowed us to assess, in cases where the text is not effective in guiding a reader's interpretation of the data, how other variables may respond.

5.3.1 Survey Procedures

Each participant completed two main tasks: a **prediction** task and a **bias appraisal** task, shown in Fig. 5.6.

For the **prediction** task, participants identified which of the two groups (Blue or Green) they expected to have a greater value at the future point shown in the chart (P1 in Fig. 5.6). The order of response options ("Blue" or "Green") was randomized and held constant throughout the survey for each participant. After making this binary choice, participants rated their confidence on a sliding scale from –25 to 25, with negative values representing, "likely Blue wins," and positive values representing "likely Green wins." This question is shown as P2 in Fig. 5.6.

Study 2 included an additional prediction question (P3): "What is the percent chance of each possible outcome occurring?" Participants entered a value for the Blue group winning, the Green group winning, and a tie. These values were required to sum to 100. Adding this question allowed us also assess the prediction of a tie occurring.

We then compared participants' predictions with the outcome favored by the text. When a participant's choice agreed with the text's implication, the response was considered *aligned*; when it disagreed, it was *unaligned*. In conditions with neutral text, no alignment could be determined, since no specific outcome was supported by the text.

Participants next completed the **bias appraisal task**, in which they judged whether the visualization's author favored one side. In Study 1, participants were asked to indicate if they thought the author of the visualization favored one side or the other. They rated

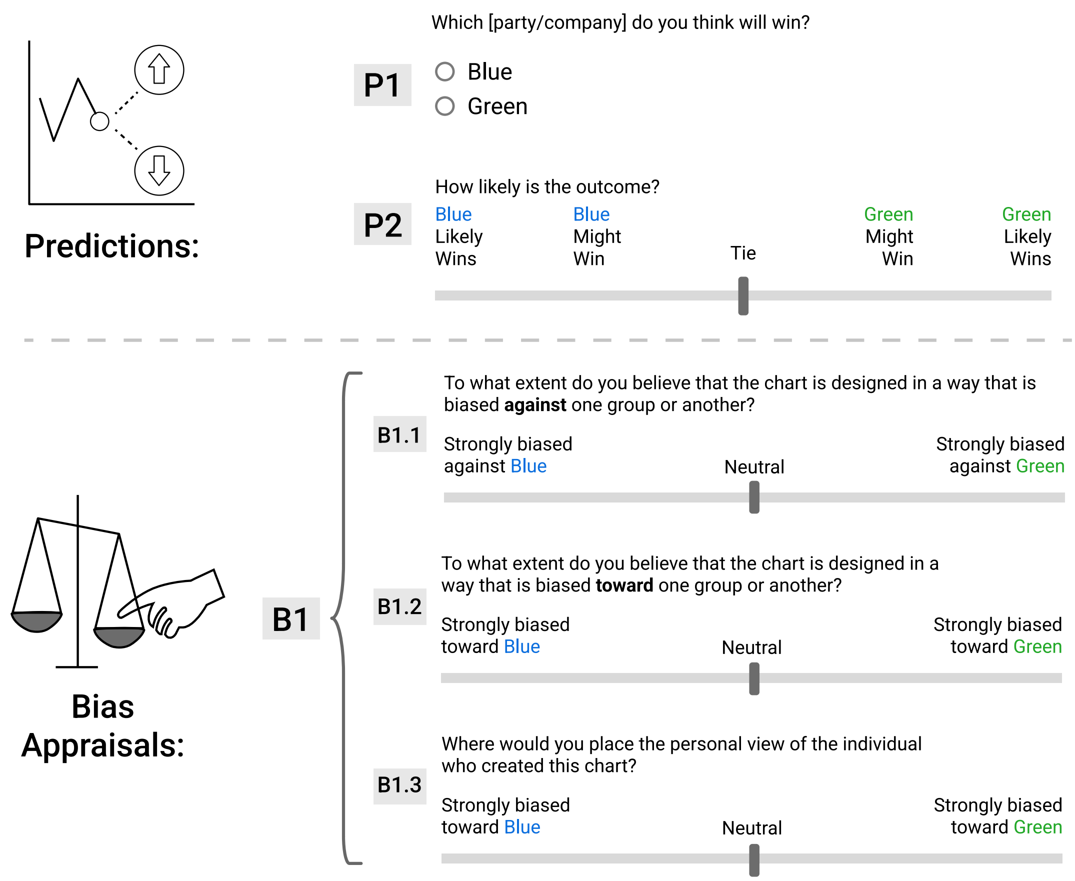


Figure 5.6: Two tasks were studied with crowdsourced participants: prediction of the outcome of the trend and assessment of the bias of the visualization author. P1 provided a binary outcome prediction, and P2 provided a measure of outcome likelihood. B1.1-3 made up a single measure of bias (B1), which was also used to construct a binary bias assessment. Results for these questions can be found in Fig. 5.7 and Fig. 5.8.

this likelihood on a –25 to 25 scale, ranging from "likely Blue author" to "likely Green author." These ratings were also converted to a binary outcome (Blue or Green) for categorical analyses.

Study 2 expanded the bias appraisal task to include three related questions, also shown as B1.1-B1.3 in Fig. 5.6:

- B1.1: "To what extent do you believe that the chart is designed in a way that is biased **against** one group or another?"
- B1.2: "To what extent do you believe that the chart is designed in a way that is biased **toward** one group or another?"

- B1.3: "Where would you place the personal view of the individual who created this chart?"

The second question ("biased against") was reverse-coded, and responses across all three questions were averaged to produce a single bias appraisal rating (B1). This edit to the survey procedure provided a more nuanced assessment of degrees of bias.

To capture neutrality in participants' perceptions, we used a range of five points on either side of zero to indicate a "Neutral" or "No-Side" categorical response. Outside of this range, appraisals that matched the true direction of the text's bias were considered *matched*, while those that favored the opposite side were *not matched*. For example, if the text read "Year after year, Blue has received more votes," the ground truth bias favored Blue. A response indicating that the author favored Blue would therefore be matched; one indicating Green would be not matched. Neutral text, which contained no explicit bias cue, was excluded from this categorization and used to establish baseline perceptions of bias.

Participants also provided written justifications for both tasks. These open-ended responses served as a quality control check, and data were excluded if participants gave inconsistent answers (e.g., predicting a Blue win in one measure but a Green win in another) or submitted nonsensical text. Finally, participants reported demographic information, including age range (e.g., "18-24") and education level (e.g., "Some high school").

5.3.2 Participants

Methods and analyses for Chapter 5 were pre-registered on OSF[1]. Separate power analyses were conducted for each of the two studies. The analytic approach was guided by the data type and research questions. Prediction outcomes (Blue vs. Green) and bias appraisals (Blue, Green, or Neutral) were categorical and analyzed using χ^2 tests of independence. Likelihood ratings were continuous and analyzed mixed-effect regression models and non-parametric tests (Kruskal–Wallis and post-hoc Dunn tests) due to non-normal distributions.

For Study 1, a power analysis using G*Power [57] for a difference between two independent means indicated that, with $\alpha = 0.05$, desired power $\beta = 0.8$, and an expected medium effect size of $d = 0.45$, a minimum of 79 participants was required per comparison. To ensure full counterbalancing, we collected 80 responses per distribution. Each chart type (bar and line) included four distributions of interest: Control/L1, L2, L3, and L4 (corresponding to the four semantic levels of text content). This resulted in a target sample of 320 participants per chart type and 640 participants total after exclusions.

For Study 2, we used the average effect size observed in Study 1 ($d = 0.375$) to conduct a similar analysis with the same parameters. This power analysis estimated that 113 participants were required per comparison. Because each chart type included three distributions of interest (No-Side, Low-Bias, and High-Bias), we collected 110 participants per distribution, yielding an ideal post-exclusion sample of 330 participants per chart type and 660 participants total.

[1]https://osf.io/4bysj

Given the relatively large sample sizes, we continuously conducted data quality control checks during data collection to ensure the target numbers were met without oversampling. Recruitment proceeded iteratively until the desired post-exclusion sample size was achieved.

Participants were recruited through Prolific [163] and compensated at a rate of $15 per hour: $0.75 for the three-minute Study 1 survey and $1.00 for the five-minute Study 2 survey. This rate was consistent with minimum wage in California at the time of data collection.

For both studies, participants were required to be fluent in English and to have an approval rate above 95% on Prolific. Data collection proceeded iteratively to reach the target post-exclusion sample size, with quality control checks conducted throughout recruitment. In Study 1, a total of 653 individuals participated; 13 were excluded for failing attention checks or providing inconsistent or nonsensical responses, resulting in the final sample of 640 participants (320 per chart type). The same data quality procedures were followed in Study 2, which achieved the target sample size of 660 participants (330 per chart type) after xxx exclusions.

Across both studies, the sample primarily consisted of young adults with at least some higher education. The most common age range was 25–34, followed by 18–24. The most common education level was a four-year degree, followed by "some college." This distribution is also reflective of the common demographics of the participant pool on Prolific.

5.4 Results

Across both studies, text influenced readers' perceptions in distinct ways. The framing provided by text had a small and inconsistent effect on participants' predictions about data trends but a large and consistent effect on their perceptions of bias. In other words, while text rarely changed what participants predicted would happen, it strongly shaped how biased they perceived the author or designer to be. Readers who disagreed with the text's implication (i.e., made "unaligned" predictions opposite the side supported by the text) tended to rate the visualization's author as being more biased than readers who were aligned with the text.

Before conducting the main analyses, we first tested whether the physical position of annotations (shown in Fig. 5.2 or the chart type affected prediction or bias appraisal ratings in the control conditions. Comparisons between the two annotation positions showed no significant differences in prediction confidence or appraisal likelihood, indicating that position alone did not alter responses. Likewise, overall confidence ratings did not differ significantly between bar and line charts. These preliminary checks confirm that neither annotation placement nor chart type introduced systematic bias into the subsequent analyses.

5.4.1 Overall Influence of Text

We first evaluated the **Prediction Hypothesis** (H5.1): *Readers will more frequently and confidently make predictions consistent with the text's implied outcome than inconsistent with it* and the **Bias Appraisal Hypothesis** (H5.2): *Readers will more frequently and confidently* *appraise author bias in the same direction as the bias conveyed by the text.* Overall, we found **partial support for the Prediction Hypothesis and strong support for the Bias Appraisal Hypothesis**. Because each study used different text content (semantic levels in Study 1 and degrees of bias in Study 2), we present the results separately. This design also allows each study to serve as a conceptual replication of the other. Responses for predictions and bias appraisals for both studies can be found in Fig. 5.7 and Fig. 5.8.

We report two measures of effect size throughout the results. Cohen's h [47] is used to indicate the magnitude of effects in χ^2 tests, while η^2 is used for Kruskal–Wallis rank-sum tests [127]. Because these measures are scaled differently, interpretations of "small," "medium," and "large" vary between them. For clarity, I provide brief interpretations after each set of results.

5.4.1.1 Study 1

In Study 1, results for the prediction task varied across chart types and measurement approaches and can be found on the left side of charts in Fig. 5.7. Categorical analyses (P1, prediction frequencies) showed significant effects for bar charts, whereas continuous analyses (P2, prediction likelihood ratings) revealed significant effects for line charts. In contrast, bias appraisals (left side of charts in Fig. 5.8) were significant across both categorical and continuous analyses, demonstrating a consistent influence of text on perceived author bias.

Outcome Predictions

Participants were somewhat more likely to make predictions consistent with the text's implication when viewing bar charts (57.1% aligned; $\chi^2 = 4.82$, $p = 0.028$, $h = 0.28$) but not when viewing line charts (50.6% aligned; $\chi^2 = 0.04$, $p = 0.846$, $h = 0.03$). Effect sizes for both findings are small, indicating a minimal influence of text on binary predictions. For example, when the text supported the "Blue" group, participants were not necessarily more likely to think the "Blue" group would win.

Prediction likelihood ratings (P2 in Fig. 5.6) were used as a proxy for decision confidence. These ratings showed similarly weak effects. For bar charts, ratings were comparable across aligned, unaligned, and control (L1 text) conditions ($Mean = 12.8$, 11.0, and 12.6, respectively). For line charts, aligned predictions were rated slightly more likely ($Mean = 13.1$) than unaligned predictions ($Mean = 11.2$, $p = 0.030$), though neither differed significantly from the control condition ($Mean = 12.0$). Kruskal–Wallis tests confirmed small effect sizes (bar: K-W $\chi^2 = 5.70$, $p = 0.058$, $\eta^2 = 0.011$; line: K-W $\chi^2 = 6.70$, $p = 0.035$, $\eta^2 = 0.015$).

Bias Appraisals

Participants were far more likely to match their bias appraisals to the chart (i.e., judge the author's bias in the same direction as the text's stance): 67.9% for bar charts and 61.5% for line charts ($\chi^2 = 30.82$, $p < 0.001$, $h = 0.73$; $\chi^2 = 12.66$, $p < 0.001$, $h = 0.46$). These differences can be seen in Fig. 5.8. Values of Cohen's h correspond to a large effect size for the bar results and a moderate effect size for the line results. For example, when the text

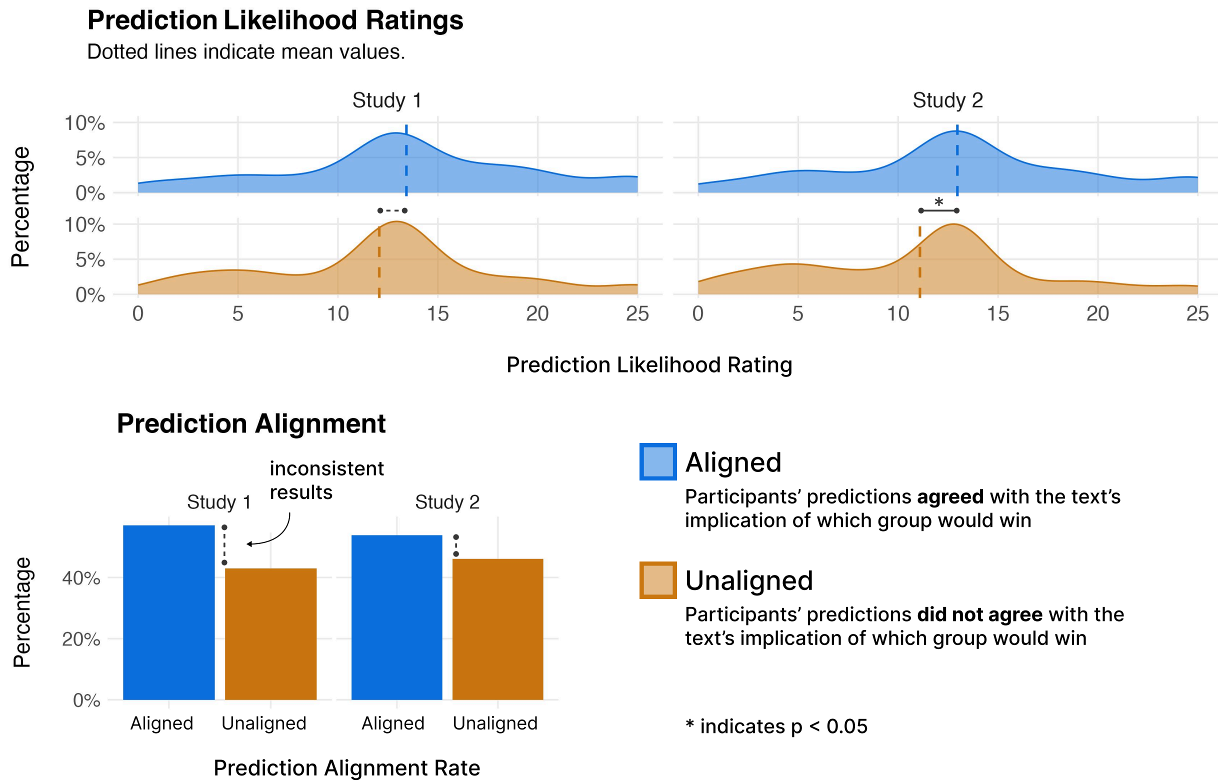


Figure 5.7: Overall results for prediction measures (P1 and P2 in Fig. 5.6) for both studies. Chart types are combined in these visuals. Statistical significance is indicated at the $p < 0.05$ level.

supported the "Blue" group, participants more frequently perceived the visualization author as part of the "Blue" group.

Participants also rated matched appraisals as more likely ($Bar\_Mean = 8.24$; $Line\_Mean = 8.24$) than unmatched ones ($Bar\_Mean = 3.08$, $p < 0.001$; $Line\_Mean = 3.42$, $p < 0.001$). Compared to the neutral control condition ($Bar\_Mean = 3.15$; $Line\_Mean = 4.10$), matched appraisals showed higher likelihood ratings across both chart types. Kruskal–Wallis tests confirmed significant effects (bar: K-W $\chi^2 = 51.14$, $p < 0.001$, $\eta^2 = 0.16$; line: K-W $\chi^2 = 29.34$, $p < 0.001$, $\eta^2 = 0.09$). As in the categorical results, effect size values indicated a large effect size for bar charts and a moderate effect size for line charts.

5.4.1.2 Study 2

In Study 2, we tested the impact of different levels of bias in the text; stimuli annotations can be seen in Tab. 5.1 and Fig. 5.3. Results can be found on the right side of charts in Fig. 5.7 and Fig. 5.8. We again found inconsistent results for the prediction task across chart

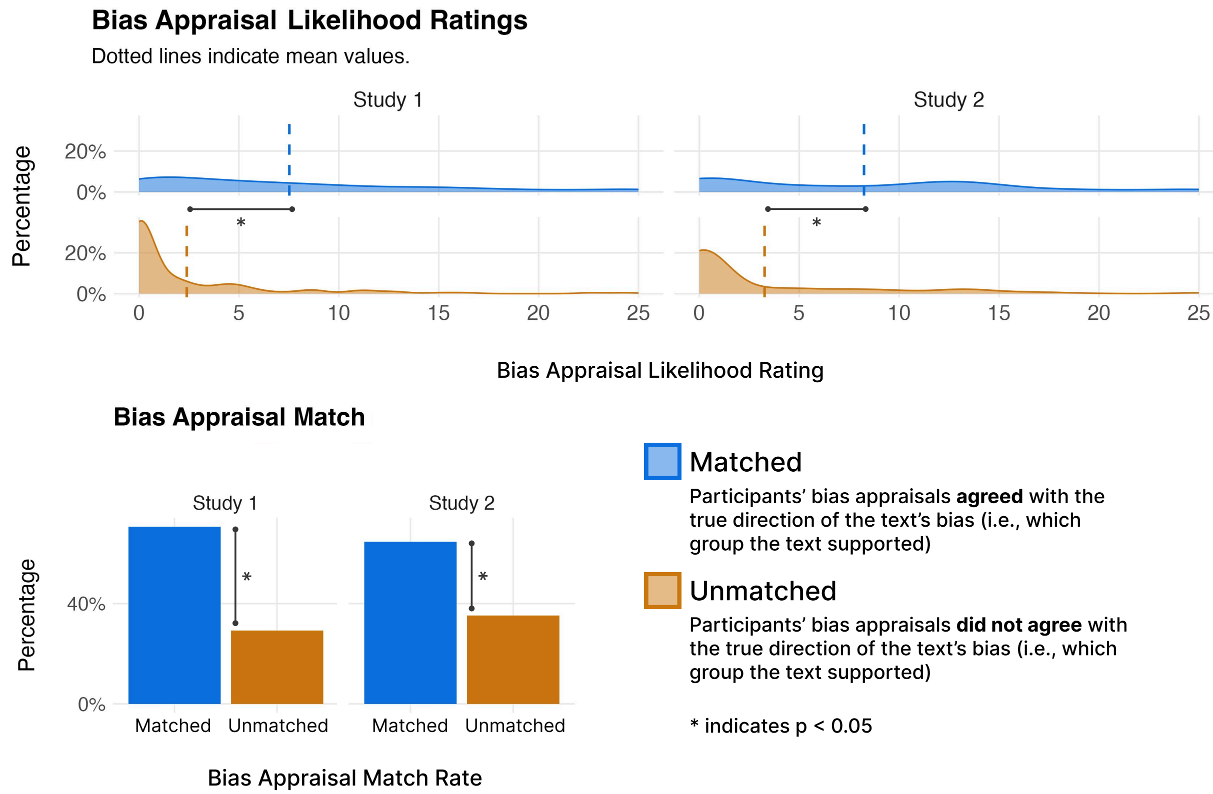


Figure 5.8: Overall results from bias appraisal measures for both studies. B1 in Fig. 5.6 was used both to determine a binary decision and as a continuous measure of bias. Chart types are combined in these visuals. Statistical significance is indicated at the $p < 0.05$ level.

types and measurement approaches. However, the inconsistencies here were reversed from Study 1. Categorical analyses (P1, prediction frequencies) showed significant effects for *line* charts, whereas continuous analyses (P2, prediction likelihood ratings) revealed significant effects for *bar* charts. Consistent with Study 1, bias appraisals were significant across both categorical and continuous measures.

Outcome Predictions

As in Study 1, text had limited and inconsistent effects on prediction outcomes. Participants viewing line charts were slightly more likely to make text-consistent predictions (58.2% aligned; $\chi^2 = 5.89$, $p = 0.015$, $h = 0.33$), whereas this effect was not significant for bar charts (56.0% aligned; $\chi^2 = 3.07$, $p = 0.080$, $h = 0.24$). Although the pattern reversed from Study 1, effect sizes remained small in both cases.

Prediction likelihood ratings were marginally higher for aligned versus unaligned responses, but these effect sizes were again very small (bar: K-W $\chi^2 = 6.25$, $p = 0.044$, $\eta^2 = 0.01$; line: K-W $\chi^2 = 6.28$, $p = 0.043$, $\eta^2 = 0.01$). For line charts, aligned predictions were rated as more likely than the "No-Side" control condition ($Mean$ = 13.7 and 11.7, respectively; $p = 0.039$). For bar charts, responses showed a marginal trend; aligned predictions were also rated slightly more likely than "No-Side" control condition ($Mean$ = 13.1 and 11.3, respectively; $p = 0.099$).

The additional prediction question for this study ("What is the percent chance of each possible outcome occurring?") did show consistent significant effects. Across the three possible outcomes (Blue wins, Green wins, and Tie), only one significant difference emerged. For bar charts, participants whose predictions aligned with the text reported higher likelihood ratings ($Mean$ = 58.9) than those whose predictions were unaligned ($Mean$ = 53.8, $p = 0.029$). This result reflects the same marginal trend observed in the original prediction confidence measure. However, the same comparison was not significant for line charts ($Aligned\_Mean = 59.4$, $Unaligned\_Mean = 56.1$, $p = 0.171$). Again, we found inconsistency in our results between chart types and different measures for prediction responses.

Bias Appraisals

The bias manipulations affected bias appraisals overall. Appraisal match frequency was higher for High-Bias annotations (81.8%) than for Low-Bias annotations (59.5%), and appraisal confidence was greater for High-Bias appraisals ($Mean$ = 8.26) than for Low-Bias ($Mean$ = 3.78) or No-Side ($Mean$ = 2.88) appraisals.

Bias appraisal trends replicated and extended Study 1's findings. Participants overwhelmingly judged the author's perspective to match the stance implied by the text: 70.0% for bar charts and 71.4% for line charts. This effect also increased with greater textual bias. High-Bias annotations produced higher match rates than Low-Bias annotations (bar: $\chi^2 = 35.2$, $p < 0.001$, $h = 0.82$; line: $\chi^2 = 40.16$, $p < 0.001$, $h = 0.88$), demonstrating large effects. Participants were able to more clearly match the appraisal with the perspective when the bias was greater.

Average appraisal ratings followed the same pattern: matched appraisals were rated more likely ($Bar\_Mean = 7.54$; $Line\_Mean = 7.51$) than unmatched ones ($Bar\_Mean = 1.77$, $p < 0.001$; $Line\_Mean = 3.03$, $p < 0.001$). Matched appraisals also exceeded control ratings ($Bar\_Mean = 2.78$, $p < 0.001$; $Line\_Mean = 2.99$, $p < 0.001$). Kruskal–Wallis tests indicated large effect sizes (K-W $\chi^2 = 63.38$, $p < 0.001$, $\eta^2 = 0.19$; line: K-W $\chi^2 = 51.2$, $p < 0.001$, $\eta = 0.15$). Overall, these results were consistent with Study 1 and the finding that participants reliably and confidently correlate the bias in the chart to the perspective displayed in the text.

5.4.2 Text Content and Position

We next evaluated the **External Information Hypothesis** (H5.3): *Text containing external context will have a stronger influence on both predictions and bias appraisals than other text content* and the **Location Hypothesis** (H5.4): *Titles will have a greater effect on predictions and bias appraisals than annotations.* Across analyses, we found **minimal sup-**

port for the External Information Hypothesis and no support for the Location Hypothesis.

These hypotheses applied only to Study 1, which used four semantic levels of text content (L1–L4) and two text positions (title and annotation). The key comparison for the External Information Hypothesis tested whether L4 text, which introduced external context to the data, had stronger effects than the other, more data-focused levels. The Location Hypothesis examined whether text presented as a title, which is typically more visually salient, would have a greater influence than text positioned as an annotation.

For both prediction and bias appraisal tasks, we used stepwise mixed-effects regression models to predict likelihood ratings. These models incorporated both content and position as fixed effects, along with their interactions, and included random effects for response option order and the group supported by the text.

Outcome Predictions

The optimal model for prediction likelihood ratings used only chart type and prediction alignment as fixed effects, with random effects of response option order and supported group. No significant effects were observed in this model, consistent with the overall pattern of weak or inconsistent prediction alignment across conditions. Adding semantic level or text position did not improve model performance ($p = 0.065$), indicating that neither the content nor the placement of text influenced how confidently participants made their predictions.

Consistent with this model, semantic level did not affect the frequency of aligned predictions ($\chi^2 = 1.82$, $p = 0.403$). Alignment rates were similar across semantic levels ($L4 = 54.7\%$, $L3 = 57.1\%$, $L2 = 49.7\%$), suggesting that adding external information did not increase the likelihood that participants would make predictions in the same direction as the text.

Text position also had no significant measurable effect on prediction outcomes. Although titles produced a slightly higher proportion of aligned predictions (56.0%) than annotations (51.7%), this difference was not statistically significant ($\chi^2 = 0.74$, $p = 0.390$). Overall, neither what the text said nor where it appeared substantially altered participants' predictions and their confidence in their prediction.

Bias Appraisals

The optimal model for bias appraisal ratings included chart type, interactions between semantic level and appraisal match, between position and appraisal match, prediction alignment, and random effects. Although text content and position were part of the optimal model, there were no significant differences between conditions. External contextual text (L4) did not increase appraisal likelihood ratings relative to other levels (vs. L3: $p = 0.310$; vs. L2: $p = 0.317$). Titles also did not increase likelihood ratings compared to annotations ($p = 0.812$).

Categorical comparisons revealed that participants were more likely to make matched appraisals when reading external context ($L4 = 73.0\%$) than when reading perceptual or relational text ($L3 = 63.2\%$, $L2 = 58.0\%$; $\chi^2 = 8.03$, $p = 0.018$). Text position did not have this categorical effect; annotations resulted in higher rates of matched appraisals (68.1%)

than titles (61.4%), but this difference was not significant ($\chi^2 = 2.04$, $p = 0.153$). This suggests that **external context may serve as a somewhat stronger signal of author bias** but that whether this text is a title or an annotation matters less.

5.4.3 Interaction Between Predictions and Bias

Finally, we examined the **Interaction Hypothesis** (H5.5): *Participants will rate their predictions as more likely when viewing annotations with a higher degree of bias.* We expected that aligned participants might perceive high-bias annotations as stronger evidence supporting their view, while unaligned participants might respond defensively, discounting the biased text.

Although this hypothesis primarily concerned Study 2, we first conducted exploratory analyses using Study 1 data to assess whether participants' prediction alignment influenced their bias appraisals. Specifically, we tested whether participants were more likely to perceive bias when their own prediction contradicted the text's perspective. In the optimal model for appraisal likelihood ratings, the inclusion of the prediction alignment variable significantly improved model performance. Participants who made predictions unaligned with the text reported higher appraisal likelihood ratings than those who were aligned, by an average of 1.24 points ($SE = 0.60$, $p = 0.039$), or roughly 5% of the total scale. This finding suggests that bias appraisals were influenced in part by disagreement with the perspective conveyed in the chart. These exploratory results motivated a more formal analysis of this interaction in Study 2, where bias was explicitly manipulated.

In Study 2, we found **minimal support for the Interaction Hypothesis**. The level of bias produced only small changes to participants' prediction likelihoods. For aligned participants, who agreed with the direction suggested by the text, the level of bias influenced some likelihood ratings (bar: $\chi^2 = 5.11$, $p = 0.078$, $\eta^2 = 0.01$; line: $\chi^2 = 6.53$, $p = 0.038$, $\eta^2 = 0.02$). However, the significant differences were not driven by the High-Bias condition. Instead, aligned participants showed a small increase in prediction likelihoods when viewing Low-Bias annotations compared to No-Side conditions (bar: $p = 0.078$; line: $p = 0.048$). Although only marginally significant, this pattern suggests that moderate bias may reinforce participants' expectations more effectively than either highly biased or neutral text.

For unaligned participants, prediction likelihoods did not differ across bias conditions (bar: $\chi^2 = 0.64$, $p = 0.726$, $\eta^2 = 0.01$; line: $\chi^2 = 2.42$, $p = 0.298$, $\eta^2 = 0.002$). Increased textual bias did not meaningfully alter prediction confidence among participants whose interpretations opposed the text.

A follow-up exploratory analysis replicated the relationship between prediction alignment and bias appraisals observed in Study 1. Participants again reported lower appraisal likelihood ratings when their prediction aligned with the text than when it was unaligned, by an average of 2.03 points ($SE = 0.56$, $p < 0.001$), or about 8% of the total scale. This consistent effect supports the interpretation that perceptions of bias depend at least partly on whether participants' own conclusions align with the chart's perspective.

5.5 Summary

This chapter investigated how text in visualizations affects readers' predictions about data trends and their perceptions of author bias. Across two studies, we found that text exerted a small and inconsistent influence on readers' predictions but a large and consistent influence on their appraisals of bias. Text containing external or contextual information (Study 1) and text with a higher degree of bias (Study 2) both increased perceptions of author partisanship. Exploratory analyses indicated that participants who agreed with the text's implied outcome tended to view the author as less biased, while those who disagreed rated the author as more biased.

5.5.1 Managing Bias in Visualization Text

While text had little impact on what participants predicted, it consistently shaped how they evaluated the author's intent. Across both studies, participants accurately identified the direction of bias signaled by the text, and strongly worded or contextual information increased perceptions of partisanship. Appraisal confidence and matching rates were highest for High-Bias annotations, indicating that readers can also pick up on the degree of bias present in the text.

These results emphasize the need for careful language choices in visualization design and communication. Text not only clarifies a chart's meaning but can also convey social and rhetorical stances. Designers, journalists, and educators should therefore be aware of this if approaching text as an explanatory or persuasive element. As a result of these findings, we consider it important to equip visualization and visual analytics tools with features that enable users to customize annotations in a manner that aims to reduce bias while effectively conveying intended messages.

Keeping text content descriptive of visible trends and data points may help to decrease the perceptions of bias. However, there are cases where there are important contextual elements that inform the data. For example, many time-series charts that include 2020 also feature data anomalies as a result of the COVID-19 pandemic. Highlighting the impact of the pandemic may, to some readers, introduce bias, but it is also a necessary step to better explain and interpret the data. Constructing charts that feel unbiased to a reader may be difficult if including many text elements, but being aware that the text contributes to possible perceptions of partisanship or influence is a step in the right direction. We further explore designers' perspectives on this difficulty in Chapter 7.

5.5.2 Impact of Visualization Task

The findings highlight how the influence of text depends on the type of cognitive task that is being evaluated. Prediction tasks appeared to rely more heavily on visual reasoning, whereas takeaway and bias appraisal tasks depended more on text information [105, 114, 204]. This difference suggests that some visualization tasks are inherently more susceptible to textual influence than others.

Tasks centered on *explanation* or *interpretation* (e.g., summarizing trends, evaluating intent) appear more affected by language, while tasks focused on *estimation* or *forecasting* (e.g., predicting future outcomes) may depend more on the visual display. These findings extend prior work showing that visuals often inform prediction-oriented tasks, whereas text affects comprehension and recall [108, 158].

The relationship between prediction and appraisal outcomes also underscores the interconnected nature of these processes. Participants whose predictions aligned with the text judged the author as less biased, while those who disagreed perceived the author as more biased. Such cross-task effects suggest that performing one task (e.g., predicting an outcome) can shape how a reader performs another task (e.g., interpreting intent or fairness). Understanding these variations across task type, user goal, and visualization design is crucial for modeling how people interpret visualizations in real-world settings.

The studies presented in this part have examined several different types of visualization interpretation, substantially furthering empirical understandings of the impact of text in visualization design. With the exception of data trend predictions, I have found text to have a significant impact on how readers engage with and interpret data. With this foundation, I now turn to examine the other side of visualization design: the *designers*.

Part III

Insights from Visualization *Designers*

While the previous chapters examined how readers interpret and respond to text in visualizations, this chapter turns to the perspective of the designer. Designers make a range of choices about how to integrate text with visual elements, balancing clarity, framing, and aesthetic goals. Yet unlike visual encodings, for which there are many formal design taxonomies and guidelines [1, 60, 112, 187, 227], text design decisions often rely only on intuition or convention.

Recent scholarship in HCI and visualization has called for stronger connections between research and practice, emphasizing the value of studying how designers actually work [71, 117, 168, 197]. Visualization design studies, for instance, engage practitioners directly, producing insights that inform both design theory and practical outcomes [149, 188]. Yet the role of text within visualization design and design processes remains underexplored. To better understand text within these design practices, the following chapters examine how designers use, organize, and reason about text in visualizations. Drawing on analyses of real-world designs and interviews with visualization designers, I extend the work introduced in Part II.

The first chapter in this section analyzes existing visualizations to propose a framework describing ten distinct functions of text and to identify recurring design patterns that capture how these functions appear in practice. The next chapter builds on this foundation through interviews with professional designers, outlining six core challenges that arise when integrating text and the strategies used to address them. The final chapter extends this focus on design process by investigating how writing activities can help designers reason through and guide their visualization decisions. Together, these chapters examine how text intersects with design to shape visualization practices.

Chapter 6

Common Design Patterns from Text Functions

This chapter examines the diverse roles of text in visualization design and introduces a framework for categorizing text functions in static visualizations. To develop this framework, my collaborators and I reviewed prior taxonomies of visualization tasks, design structures, visualization literacy, and textual communication, and conducted iterative rounds of open coding and group discussion to refine the function categories. Using a dataset of 120 real-world visualizations [7], containing 804 text elements (axes, titles, annotations, and other written elements within each visualization), we identified ten distinct text functions and explored how these functions combine to communicate, synthesize, and frame information. Our analysis revealed cases where text replaces visual elements and highlighted the rhetorical variety of titles and multifunctional text. We further conducted a factor analysis and identified four overarching text-informed design strategies: Attribution and Variables, Annotation-Centric Design, Visual Embellishments, and Narrative Framing. From these factors, we discussed the potential for factor-oriented redesigns to fit different design contexts. This chapter contains work from a previously published study conducted in collaboration with Anjana Arunkumar, Lace Padilla, and Marti Hearst [201]. I served as first author and was responsible for reviews of prior taxonomies, development and coding of the function framework, exploratory analyses of function prevalence, and the majority of the writing. This content has been edited for clarity and coherence with this dissertation.

6.1 Taxonomies of Text Use in Visualizations

Designers make frequent decisions about how to use text in visualizations, but there has been limited effort to characterize the many ways text operates in these designs. Recent research has begun to establish frameworks for text elements such as annotations, titles, and alternative text, but there are still several important gaps in the existing formalizations with respect to text in visualization. This chapter builds on these foundations to propose a thorough framework describing the functions of text in visualization design.

Rahman et al. [176] developed an extensive design space for annotations by applying the "how, why, and what" structure of Brehmer and Munzner's abstract task taxonomy [30]. The original abstract task taxonomy bridges high- and low-level tasks by categorizing what users aim to accomplish (*why*), the data inputs and outputs involved (*what*), and the interactions or operations they perform (*how*).

Rahman and colleagues used this framework to classify annotation types (e.g., text, connector, glyph) and their analytic purposes. The annotation design space also leverages key verbs found in the abstract task taxonomy (identify, compare, summarize, and present). Later work by the same authors [175] distinguished between *additive* text, which introduces external information, and *observational* text, which describes data features. These provide only broad classifications for text function. Other research has provided similar general classifications for visualization titles [131]: generic information (e.g., variables or encodings) and data features (e.g., trends or aggregations).

Lundgard and Satyanarayan [134] proposed a four-level semantic model for alt text (encoded, statistical, perceptual, and contextual) that provides useful structure but is primarily designed for accessibility. This model has since been used to evaluate textual descriptions and chart captions [202, 210, 214, 262] and informed stimuli development in earlier chapters (Chapter 3, Chapter 4, Chapter 5).

Although these frameworks advanced our understanding of how text appears in visualizations, they share several limitations:

1. Text is characterized only at a broad level.
2. One kind of text use is considered (e.g., titles, alt text), rather than classifying across all uses within a visualization.
3. A given text element is assigned only a single function.
4. Interactions between text elements in a design are not considered.

In this chapter, we examine two research questions. First, RQ1 asks **what are the *functions* of text in visualization designs?** This question addresses Gaps 1-3 by clarifying the *specific* roles text can play across a range of elements, including cases where a single element serves multiple functions. Building on existing frameworks and open coding exercises, we propose a detailed set of text functions. We use this set of functions to then investigate RQ2, which asks, **what text design patterns emerge across visualizations?** Addressing Gap 4, this question considers how features of text elements, including their functions, interact and combine within a given design. Together, these questions establish a foundation for a structured and nuanced representation of text in visualization.

6.2 Methodology for Function Framework Creation

To develop a framework describing the functions that text performs in visualization designs, we combined literature review, iterative discussion, and qualitative analysis of existing visualizations. Our process involved three main stages: (1) synthesizing prior frameworks and generating initial categories through discussion, (2) refining and expanding those categories through open and axial coding of real-world designs, and (3) validating the resulting set of functions through systematic coding of a larger corpus. All supplemental materials, including codebooks and visualization examples, are available on OSF[1].

6.2.1 Code Development

Developing the function codes was a multi-stage process that unfolded over three months of weekly meetings and near-continuous discussion among the research team. Each session involved reviewing visualization examples, debating functional distinctions, and refining the language used to describe them. Early classifications were frequently dismantled and rebuilt as we attempted to reconcile overlapping roles and ambiguous cases. The process required the synthesis of perspectives from design, linguistics, and visualization research.

Broad Classification Groups

We began by reviewing prior frameworks for both visualization tasks and textual elements [30, 131, 134, 175, 176]. Through several rounds of collaborative discussion, we proposed an initial classification scheme describing what text can express in relation to data and design. These early conversations emphasized three broad categories:

1. **Data encodings or mappings**, which *must* be conveyed (whether visually or through text) for a reader to understand how visual marks correspond to data values.
2. **Data insights**, which capture relationships or trends visible in the data representation and can be communicated through text or visuals.
3. **Data context**, which encompasses external information (e.g., social, political, methodological) that must be expressed through text rather than visuals.

In these discussions, we also noted several features of text elements that cut across these initial categories: the degree to which a text element attracts attention, whether it is necessary for interpreting the visualization, whether it serves multiple purposes, and how flexibly it can be positioned in the design. We conceptualized these as continuous axes and constructed the diagrams shown in Fig. 6.1 to guide conversation. However, this preliminary framework was too coarse and remained focused on the *type* of text element (e.g., title, caption, axis label) rather than its *function* within the design.

Refinement Through Other Possible Groupings

We were not satisfied with the general grouping we had initially proposed, since it fell victim to several gaps in prior taxonomies as well. To help us move beyond these basic groupings, we used several other possible categorizations schemes to better understand what functions or uses of text might be excluded from possible broad groups.

As our team discussed and coded examples, we drew analogies between visualization text and the microstructures of arguments [31, 101, 181, 200]. In persuasive writing, argument

[1] https://osf.io/swqfc/

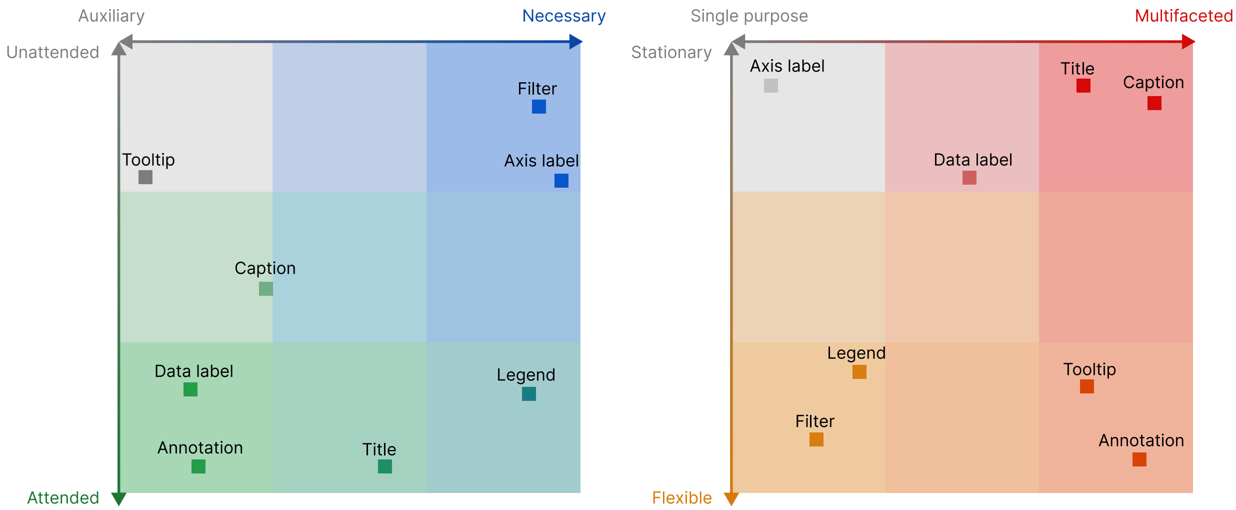


Figure 6.1: A set of diagrams to fuel conversation early in the development of text functions. These representations focused on the type of text, including some text elements present in interactive visualizations.

components (e.g., claims, evidence, warrants) are linked by relations (e.g., support, refute, elaborate) [23, 200]. Similarly, textual components in a visualization often refer to or build upon one another, forming interconnected "argument structures" that communicated and reasoned about data. We began conceptualizing the comparison between text in a visualization and microstructures through a three-level framework inspired by argument analysis for persuasive essays [200]:

1. **Component identification**: isolating text components from visual units and identifying their boundaries.
2. **Component classification**: categorizing the functions of these text components.
3. **Structure identification**: mapping relationships between components (e.g., a subtitle that elaborates on a title, or an annotation that references a legend).

This perspective underscored that text cannot be meaningfully analyzed in isolation; its role and purpose often depend on its relationship to other textual and visual elements. Although we did not end up using microstructure framework as a formal part of our analysis, these discussions informed our later analysis and consideration of interaction and interdependence among text elements.

We also discussed how both text and visual components may work together, recognizing that many communicative functions are distributed across modalities. We organized these into three broad operational categories: **Text Only**, **Text + Visual**, and **Visual Only**. For example, certain communicative purposes, such as providing definitions or source metadata, can only be conveyed through text, whereas others, like showing trends or relationships, can be jointly achieved through text and visual marks. Considering both modalities clarified how textual and visual elements collaborate to achieve communicative goals and informed the development of our later function set.

Finally, we refined our framework through iterative discussions and open coding of diverse visualizations, using our initial broad categories as a foundation but continually revising them to capture more granular and functional distinctions, informed by other possible groupings and analyses. This process was supported by reference to low-level task taxonomies [5, 123, 196], high-level task taxonomies [6, 30, 35, 226], models of reading comprehension [15, 186, 212, 252], and rhetorical frameworks [77, 118, 125, 126, 171, 179]. Together, these perspectives guided the development of a more detailed, functional framework.

Iterative Coding and Function Naming

Using these candidate functions as a foundation, we conducted several rounds of open and axial coding on 18 real-world visualization designs encountered in the real world and drawn from MASSVIS and other chart corpora [25, 134, 176]. Because the open coding was conducted on a small subset of designs in preparation for closed coding on the primary corpus, we drew from diverse sources to avoid missing potential text functions.

We recorded short descriptive phrases for each text segment (e.g., highlighting an outlier, explaining a trend, providing a data source), grouped them into thematic clusters, and refined our definitions through discussion. New text that did not fit existing categories prompted further revision. Through repeated cycles of open and axial coding, we developed a structured set of distinct text functions to capture the range of ways text contributes to visualization design.

A key breakthrough occurred when we adopted verb–noun pairings to name functions, following the structure of prior frameworks for tasks and annotations [30, 176]. There verb-noun pairings also mirrored the kind of argument microstructure concept discussed previously which paired components of an argument with their relations. When renaming our set of candidate functions, we were able to describe functions at the level of specific communicative acts and make the functions more precise and comparable. The resulting set of functions is the one proposed and used for analysis in this chapter.

6.2.2 Corpus Creation

After developing the function set through open coding of a small, diverse group of visualizations, we needed a larger and more consistent collection to assess how the functions captured text use in practice[2]. To do this, we drew on the Image-to-Information corpus [7], which itself was built from MASSVIS and related visualization collections [24, 25]. This corpus was explicitly curated to represent "in the wild" visualizations from multiple public-facing domains, including news media, government reports, and science communication.

The original Image-to-Information corpus contained 500 images. Because our goal was to analyze how text functions within information visualizations, we filtered this corpus to

[2] Three images appeared in both collections.

focus on images that clearly presented data and contained legible text. Following the classifications in the original study, we first removed 273 designs classified as "image" rather than "information" This filtering ensured that our focus remained on visualizations meant to inform readers. We then excluded designs that would complicate or obscure text-function analysis: infographics with highly heterogeneous layouts (54), diagrams (25), tables (10), and interactive designs captured as screenshots (2). Finally, we removed any visualization whose text was unreadable due to resolution or image quality.

These steps resulted in a working corpus of 120 visualizations. This subset leans somewhat more heavily toward news ($n = 46$) and government ($n = 19$) sources than the original 500-image collection, which is consistent with our focus on public-facing, explanatory visualization. At the same time, the subset retained variety in chart type, layout, and text density, providing many different instances of titles, subtitles, annotations, captions, legends, and axes. While our corpus focuses on only a select set of sources, it captures key visualizations aimed at general audiences.

We manually extracted text from each visualization[3], recording either the exact content or summarizing repetitive elements such as categorical headers or date labels. For example, we used "set of categories" to describe repeated labels on a categorical axis. This allowed us to preserve the communicative purpose of the text without inflating the number of components. Multi-line text was treated as a single element to preserve context. This process resulted in 804 text elements or groups. A second collaborator reviewed all extracted text to ensure accuracy and proper grouping of multi-line elements.

6.2.3 Function Coding

We coded the corpus of 120 selected visualizations and 804 text components according to the set of text functions.

To support later analyses, we also recorded a small set of metadata for each text component (see Tab. 6.1). This metadata captured (1) the **text type** (e.g., title, subtitle), (2) any associated **non-data visual element** (e.g., arrows, circles, icons), and (3) the use of **color** within the text (e.g., encoding, highlighting). These metadata categories, particularly the text type, were drawn from prior work on text in dashboards and interactive visualizations [214] and were refined through discussion to match the needs of this study. An additional collaborator independently coded this metadata as well, with discrepancies resolved through discussion.

We then coded each component for the presence of one or more text functions from the finalized framework, found in Tab. 6.2. Because one of the key claims of this work is that text in visualizations is often multifunctional, we allowed multiple function labels to be assigned to the same component.

[3] In initial tests, large-language models (LLMs) extracted text fairly well with some errors (e.g., capturing axes). At the time of analysis, LLMs did not support automatic coding of text with our labels, so we did this work manually.

Table 6.1: Coding types other than text functions. Not all text elements contained visual elements or color, but all text elements were assigned a type.

Metadata	**Description**	**Codes**
Text Type	Category or position of text element used in visualization design	Titles, Subtitles, Annotations, Captions, Axes, Legends, Paragraphs
Non-Data Visual Elements	Visual elements associated with the text, if any	Arrows, Circles, Logos, Icons, Lines, Rectangles
Color Use	Role of color in the text element, if any	Encoding, Highlight, Style

To maintain rigor, the coding process was carried out iteratively, with a focus on achieving high reliability across all function codes. Two coders independently coded a random subset of 20 visualizations from the 120-image corpus. We then calculated interrater reliability using Cohen's κ for each function code [66]. Our target was that at least 75% of the codes would reach $\kappa > 0.8$, following threshold recommendations for qualitative coding studies [87, 155]. When a code failed to meet this threshold, the team revisited the definition in the codebook, clarified edge cases, and recoded a new randomly selected subset of images. The coders discussed discrepancies and refined the codebook.

After two rounds of this iterative process, the coding scheme reached the desired level of agreement. All function codes demonstrated very strong agreement ($Mean\_\kappa = 0.97$). The only function that required additional discussion was the one related to affective or valenced language, which was somewhat more subjective ($\kappa = 0.62$) Each coder then coded the entire corpus independently, with discrepancies resolved through in-depth discussions among all four authors. The final codebook reflects these deliberations, providing a reliable framework for future analysis.

Table 6.2: Text functions in visualizations, constructed with a verb to describe the text's action and a noun to capture what it acts upon. "Identify," "Present," "Compare," and "Summarize" are verbs borrowed from prior work [30, 176].

Verb	**Noun**	**Definition**
IDENTIFY	MAPPINGS	Communicates conventions for reading the charts, i.e., the relationship between data point(s) and corresponding visual elements or structures, such as position, color, shape, or other visual channels.
IDENTIFY	6.2 VALUES	Directly identifies and labels all data points with the relevant value, category, or other point-specific information. Serves as a reference for data values, rather than selective emphasis or comparison.
PRESENT	METADATA	Provides information about 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. Includes definitions of a data category, variable, or other terms used.
REPLACE	MAPPINGS	Provides an unconventional representation for a conventionally presented element used to establish the basic structure of data encoding when such an element (e.g., axis, legend) is omitted from the display.
COMPARE	MAPPINGS	Translates a data mapping into more understandable or contextually relevant terms by rephrasing or interpreting technical, complex, or abstract data mappings into more relational language.
COMPARE	6.2 VALUES	Describes relationships among and between data points through direct comparison of points or groups. Text may also highlight one or multiple data points in comparison to the overall dataset.
SUMMARIZE	6.2 VALUES	Describes relationships among and between data points through aggregation (e.g., average) or mathematical function (e.g., addition). Text may also group or filter points along a given dimension.
SUMMARIZE	CONCEPTS	Provides a high-level summary of some or all aspects of the chart; text can either provide information synthesis (subcode: SYNTHESIS) or describe the variables displayed (subcode: VARIABLES). Augmented by an additional taxonomy of rhetorical strategies [77, 126].
PRESENT	CONTEXT	Integrates contextual information, including background knowledge about the world (such as geographic, cultural, and political relationships), knowledge about current events, and domain-specific information stemming from expertise in a particular field of research or scholarship.
PRESENT	VALENCED SUBTEXT	Promotes an emotional response, appealing to emotions, values, and personal experience. Conveys an emotional tone that would be diminished or lost in a more neutral phrasing.

6.3 Text Functions

To clarify how text and visual elements interact in visualization design, we developed a framework comprising ten distinct text functions. These functions describe the specific communicative actions text performs within a design, including what it does and what it acts upon. Tab. 6.2 summarizes the final set of functions, and Fig. 6.2 illustrates how they appear in practice. Frequency information for each function is provided in Tab. 6.3.

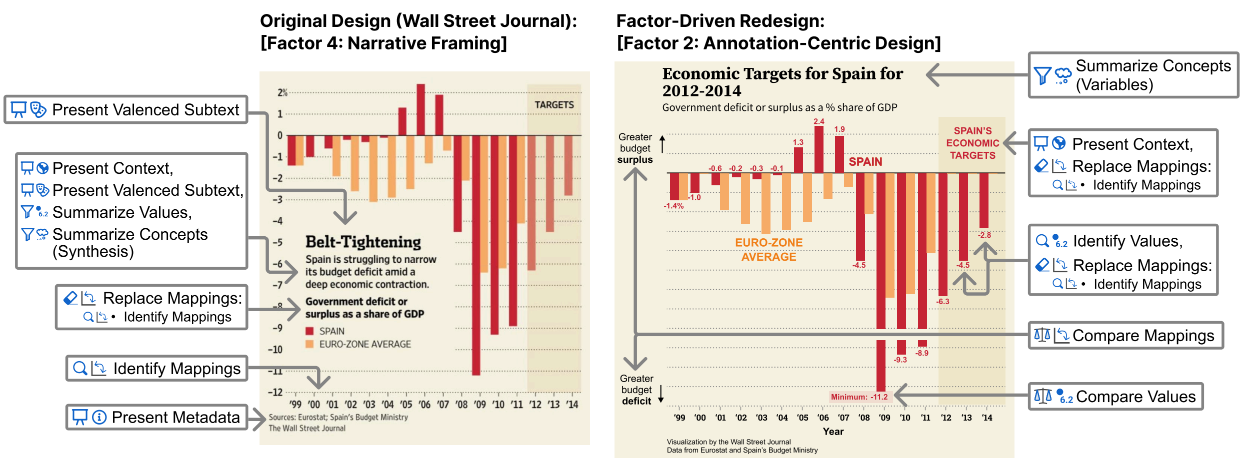


Figure 6.2: Two versions of a Wall Street Journal (WSJ) visualization demonstrate how text functions shape the information provided by the design. **Left**: Original WSJ design [86]. **Right**: A redesign created for this example, illustrating an annotation-centric approach. Redundant functions are omitted from labels for clarity.

6.3.1 Text Function Framework

Text elements that IDENTIFY MAPPINGS communicate the conventions for interpreting the chart, such as how data values correspond to visual channels like position or color. *Axes* and *Legends* frequently serve this function, though it can be handled by other text elements. In the redesigned visualization shown in Fig. 6.2, no formal y-axis exists. Instead, data labels (i.e., the number *Annotations* on the bars) take on the IDENTIFY MAPPINGS function and show how bar height relates to numeric value.

These same *Annotations* also perform IDENTIFY VALUES. By providing a complete set of numeric values for each bar in the 'Spain' category, they allow readers to retrieve exact values without estimating. Crucially, IDENTIFY VALUES only applies when all relevant values are labeled; selective emphasis, such as tagging only an outlier, would fall under COMPARE VALUES, emphasizing relational information.

Text can also provide context about data provenance or production. The PRESENT METADATA function captures instances where text includes information such as data sources, collection methods, or authorship. This function often appears in *Captions* or other footnote positions, such as the note in Fig. 6.2 that reads, "Data from Eurostat and Spain's Budget Ministry," Its goal is often to support transparency and credibility.

During open coding, we identified a function that did not appear in any previous taxonomy: REPLACE MAPPINGS. Unlike text that complements conventional encodings, this function applies to text that fulfills that role entirely, communicating what would otherwise be shown through *Axes* or *Legends*. In the right of Fig. 6.2, for example, numeric labels replace the omitted y-axis. Similarly, *Annotations* such as "Spain" and "Euro-Zone Average" take the place of *Legends*. Even in the original design in Fig. 6.2, the legend title "Government deficit or surplus as a share of GDP" doubles as a replacement for a y-axis label. Without the legend title, it would be unclear what variable was shown on this *Axis*. These examples illustrate how text can step in when conventional elements are minimized or removed. Additional examples of REPLACE MAPPINGS and associated redesigns can be found in Fig. 6.3.

Other text elements interpret or rephrase chart encodings for the viewer. The COMPARE MAPPINGS function translates visual encodings into more accessible or conceptual language. For example, the phrase, "Greater budget surplus," on the right-hand side of Fig. 6.2 qualitatively interprets vertical position, explaining what an upward direction means in the context of the data. Whereas IDENTIFY MAPPINGS indicates the rule(s) for interpreting the display, COMPARE MAPPINGS conveys the meaning of that rule within the context of the data.

The COMPARE VALUES and SUMMARIZE VALUES functions capture relational reasoning between data points. COMPARE VALUES highlights contrasts (e.g., identifying a minimum or maximum), including cases where other points are left unlabeled, since the comparison is implicit between a single point and the rest of the dataset. By contrast, SUMMARIZE VALUES aggregates across multiple points, describing general trends or computed values. The *Subtitle* "Spain is struggling to narrow its budget deficit amid a deep economic contraction" receives the SUMMARIZE VALUES code, since it examines the delta or change between bars over time, rather than referring to any single point. Both functions have a set of more precise subfunctions (e.g., "sum," "group") which parallel low-level task taxonomies [5].

Beyond quantitative summaries, text can also synthesize ideas or provide conceptual framing. SUMMARIZE CONCEPTS captures broader conceptual framing or synthesis, often seen in *Titles* or *Subtitles*. Through close review of title-specific studies and taxonomies [114, 115, 131], we refined this category and identified two subtypes. SUMMARIZE CONCEPTS: SYNTHESIS involves interpretation or synthesis, distilling the chart into a takeaway (e.g., left of Fig. 6.2: "Spain is struggling..."). SUMMARIZE CONCEPTS: VARIABLES, by contrast, lists the variables or chart contents *without* added interpretation (e.g., right of Fig. 6.2: "Economic Targets..."). We also examined rhetorical strategies used in SUMMARIZE CONCEPTS, such as puns, associations, exaggerations, repetitions, etc. [77, 126].

The PRESENT CONTEXT function introduces background or domain knowledge not visible in the chart itself, such as historical events, social conditions, or policy relevance. In Fig. 6.2, the phrase "deep economic contraction" in the *Subtitle* situates the chart within a broader economic narrative, helping readers interpret the stakes of the data. Even brief

Table 6.3: Percent of designs ($n = 120$) containing each function, faceted by text type. "Annot." stands for *Annotation*. "Para." stands for *Paragraph*. Designs could contain multiple types of text serving the same function, so percentages will likely not add to 100%.

Function	**% of Corpus**	***Title***	***Subtitle***	***Annot.***	***Caption***	***Axis***	***Legend***	***Para.***
Identify Mappings	100 %	11 %	9 %	58 %	3 %	75 %	44 %	2 %
Identify Values	56 %	0 %	2 %	52 %	0 %	4 %	1 %	1 %
Present Metadata	68 %	6 %	12 %	7 %	53 %	0 %	4 %	2 %
Replace Mappings	56 %	8 %	8 %	38 %	1 %	4 %	2 %	2 %
Compare Mappings	19 %	0 %	1 %	12 %	0 %	2 %	2 %	2 %
Compare Values	34 %	5 %	6 %	24 %	1 %	1 %	1 %	3 %
Summarize Values	35 %	8 %	12 %	13 %	0 %	0 %	1 %	4 %
Summarize Concepts	88 %	78 %	30 %	13 %	7 %	1 %	2 %	3 %
Present Context	32 %	3 %	12 %	18 %	5 %	2 %	2 %	8 %
Present Valenced Subtext	25 %	21 %	8 %	2 %	0 %	0 %	0 %	2 %

text, like the "Targets" *Annotation*, can signal contextual knowledge that these are not just values, but policy goals.

Finally, text could carry tone and emotion through the Present Valenced Subtext function which promotes emotionally charged or value-laden language. Describing Spain as "struggling," for example, adds emotional charge absent from the display otherwise. To inform our definition and use of this function, we reviewed studies on affect in visualization design [118, 125]. Since we did not have access to designer intents, we focused instead on the text's *potential* to elicit an emotional or affective response in viewers. A practical heuristic is whether the text could be rewritten neutrally without changing the informational content; if so, the original wording likely carries valence.

Together, these ten functions describe the myriad of ways that text elements contribute to visualization design. They emphasize text not as a single-purpose supplement to a visual display but as a flexible communicative component that can inform, interpret, substitute, or frame visual information.

6.3.2 Comparison to Other Frameworks

Our framework extends and refines prior taxonomies of text in visualization. Several of our functions, such as IDENTIFY MAPPINGS, SUMMARIZE VALUES, and COMPARE VALUES, align with the action-oriented verbs in Rahman et al.'s annotation taxonomy [176]. Similarly, Lundgard and Satyanarayan's four-level semantic model [134] captures broader categories like contextual or encoded information, which correspond to our PRESENT CONTEXT and IDENTIFY MAPPINGS functions. However, these existing frameworks primarily address text at broad levels of abstraction or within specific contexts (e.g., accessibility), whereas our framework identifies finer-grained, context-sensitive functions grounded in how text operates within the structure of a chart. This added specificity enables more detailed analysis of design choices and more precisely reflects how text operates as an integrated part of visualization structure, rather than as isolated content.

Our approach also introduces novel functions that capture roles largely absent from prior work. REPLACE MAPPINGS identifies instances where text replaces conventional chart components – an interaction between language and design rarely discussed in visualization taxonomies. While the semantic levels address text that describes visual encodings, they do not account for these cases of mutual design between text elements. Additionally, functions such as COMPARE MAPPINGS bridge between semantic levels of description [134], linking encodings (L1) with higher-level relational (L2) interpretations. The potential for text to evoke emotion, represented by PRESENT VALENCED SUBTEXT, also goes beyond the scope of most taxonomies.

Our framework treats multifunctionality as a defining property of visualization text. Rather than assigning each element a single role, it recognizes that text often serves several communicative purposes simultaneously This multidimensional treatment of text functions moves beyond prior one-to-one categorizations, providing a more realistic and flexible account of text in visualization designs.

6.3.3 Exploratory Analysis Findings

This function framework addressed our first research question, **"What are the *functions* of text in visualization designs?"** We conducted additional exploratory analyses examining the prevalence, distribution, and co-occurrence of text functions across our corpus. We examined how function frequencies and usage patterns align with established conventions in text design and to what extent the function framework reveals additional or previously undercharacterized aspects of how text is used in visualizations.

6.3.3.1 Characterizing Titles and Subtitles

Prior research has established that *Titles* strongly influence how readers interpret visualizations [24, 114, 115], yet few studies have analyzed how *Titles* and *Subtitles* work together or how their linguistic construction shapes meaning. Consider the two *Titles* in Fig. 6.2:

T1: "Belt-Tightening: Spain is struggling to narrow its budget deficit amid a deep economic contraction."

T2: "Economic Targets for Spain for 2012-2014: Government deficit or surplus as a % share of GDP"

These examples illustrate two recurrent patterns in our corpus. Type T1 was SUMMARIZE CONCEPTS: SYNTHESIS, which interprets or synthesizes data into a takeaway. These titles often contained verbs (68%), reflecting their interpretive character. Type T2 corresponded to SUMMARIZE CONCEPTS: VARIABLES, where the text indicates which variables are shown in the chart without explicit interpretation. Nearly all such *Titles* (94%) consisted entirely of noun phrases without verbs.

T1 demonstrates a common rhetorical pattern in news visualizations: a short, attention-grabbing *Title* that uses metaphor or wordplay ("Belt-Tightening"), followed by a longer explanatory *Subtitle*. These shorter *Titles* may act to capture the readers' attention, while the longer *Subtitles* would provide them with a description of the important data features. Drawing on rhetorical taxonomies from advertising and data journalism [77, 126], we coded rhetorical devices such as puns, associations, exaggerations, word omissions, and repetition, alongside broader distinctions between formal and informal registers [92].

Short *Titles* (three or fewer words) frequently used rhetorical devices (62%), compared to only 3% of longer *Titles* or *Subtitles*. For example, the *Title* "Falling Rains", was used in one of the images in our corpus,. This *Title* functions as a pun, encapsulating both a data variable (rainfall) and its decline. Additionally, by using the double meaning in a word, *Titles* can simultaneously perform SUMMARIZE CONCEPTS: SYNTHESIS and SUMMARIZE CONCEPTS: VARIABLES. *Titles* that used rhetorical strategies often also PRESENT VALENCED SUBTEXT through associations or wordplay. In contrast, *Subtitles* rarely used distinct headline rhetorical strategies, suggesting that *Subtitles* serve a different role, often providing additional context rather than relying on wordplay or rhetorical techniques.

Although *Titles* are a common element in visualizations, 26 images in our corpus omitted them entirely, instead relying on *Annotations* to convey key information ($Mean$ = 3.65 annotations per image, SD = 2.56). Surprisingly, these *Annotations* were most commonly used to IDENTIFY VALUES, though some also synthesized or compared data. *Annotations* were also the predominant location for COMPARE VALUES. The relative positional flexibility of *Annotations* compared to other text types was mirrored by its flexibility in content as well; *Annotations* could both serve as substitutes for conventional chart elements and contribute to storytelling by emphasizing comparisons.

6.3.3.2 Usage of Replace Mappings

Across our corpus, 56% of designs omitted at least one conventional mapping element, instead using REPLACE MAPPINGS. This function was most common in *Annotations* but also occurred in *Titles* and *Subtitles*, where text assumed the informational roles typically held

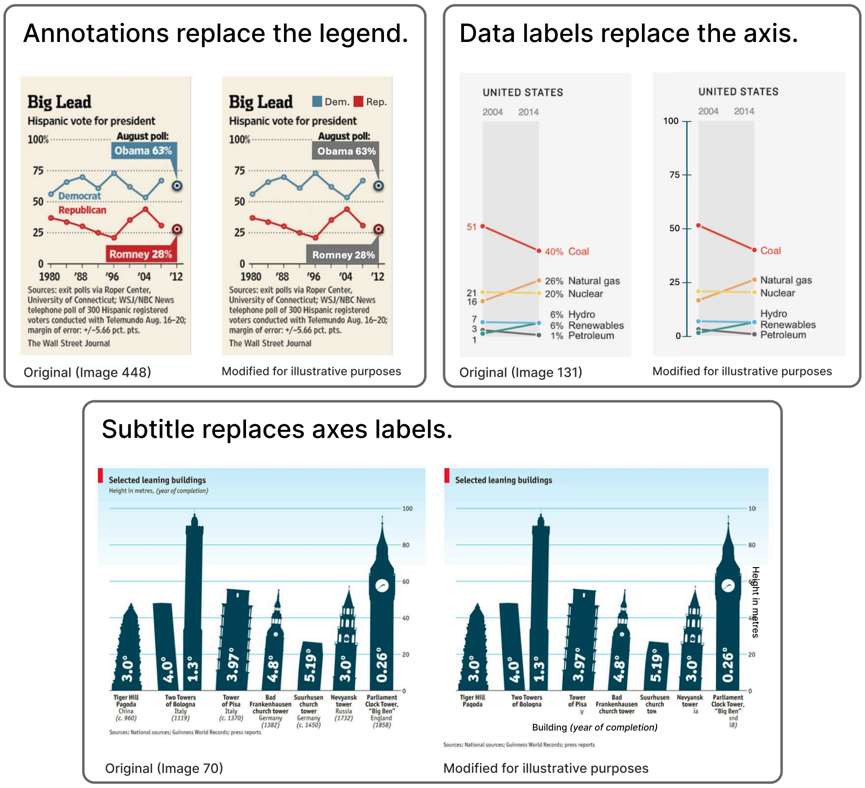


Figure 6.3: These designs demonstrate Replace Mappings: the use of text elements to replace conventional mapping components in a chart, such as *Axes* and *Legends*. Charts on the left of each panel are the original designs[4]; charts on the right are redesigns *without* Replace Mappings.

by *Axes*. *Axes* were replaced slightly more often ($n = 41$) than *Legends* ($n = 31$), with six visualizations replacing both.

We identified three recurring forms of Replace Mappings: (i) *Legends* replaced by *Annotations*, (ii) *Titles* or *Subtitles* conveying *Axes* information, and (iii) *Axes* replaced with direct data labels; Figure 6.3 displays representative examples as well as the non-Replace Mappings version of the designs. These redesigns demonstrate that the use of Replace Mappings does not change the *meaning* of the data, as may be the case for other functions. Instead, it directly affects the aesthetics of the chart and the number or type of other text elements present.

The most common form of Replace Mappings involved *Legends* replaced by *Annotations* ($n = 28$), sometimes incorporating the use of color directly into the text ($n = 14$). For

[4]Figure 6.3 image credits: Img 131: Alyson Hurt / NPR [original [91]] and Flowing Data [reprint [255]], September 2015; Img 70: Graphic Detail / Economist, October 2011 [49]; Img 448: Neil King Jr. / Wall Street Journal, September 2012 [109]

example, the upper left image in Fig. 6.3 uses color in the "Democrat" and "Republican" *Annotations*, allowing the designer to avoid finding suitable space for a *Legend*. This direct-labeling approach [1, 63, 112] streamlines interpretation and reduces spatial complexity.

At times, color encodings were simply described in the text rather than using direct labels, as in the case of Image 403 in Fig. 6.6. An *Annotation* explicitly states that red shading indicates uncertainty, eliminating the need for a separate legend box. This method may be particularly useful when space is limited, or the color encodings require more explanation than a *Legend* affords.

A second form used *Titles* or *Subtitles* to convey *Axis* information ($n = 17$). For example, the bottom image in Fig. 6.3 uses the *Subtitle* "Height in metres, *(year of completion)*," which provides the same information as an axis label but in a more prominent location. This approach could reduce redundancy and align with minimalist design practices, though it requires readers to synthesize information across different text elements rather than relying on a localized label.

Finally, *Axes* could also be replaced with direct data labels in the form of *Annotations* ($n = 14$). These labels eliminated the need for scale estimation, instead embedding precise values within or beside data points. This strategy, as seen in the upper right panel of Fig. 6.3, emphasizes accuracy but can increase text density, potentially overwhelming readers in data-rich displays. For example, in Fig. 6.2, the decision to label Spain's data directly required widening the design in order to make the *Annotations* readable.

This function underscores the importance of examining a visualization as an integrated whole rather than as a collection of isolated components. Without considering how *Axes* and *Legends* interact with *Annotations*, we would not have identified the substitution behavior captured by Replace Mappings. This design pattern has not been described in detail in prior taxonomies of text or tasks, nor in other visualization studies addressing design trade-offs.

6.3.3.3 Single Text, Multiple Functions

One contribution of this framework is its recognition that text in visualizations often serves multiple functions simultaneously. Co-occurrences of functions within text elements can be seen in Fig. 6.4. Within our corpus, 39% of text components performed more than one function. Multifunctionality was most prevalent in *Annotations* (66%), followed by *Subtitles* (58%) and *Titles* (51%). These text elements tended to advance the visualization's narrative or interpretive goals rather than simply describe encodings.

Several patterns of co-occurrence emerged. Summarize Values and Compare Values frequently co-occurred (58% of Summarize Values instances). The combination of these functions suggests that data synthesis often involved both aggregation *and* direct comparisons. Similarly, Present Valenced Subtext almost always appeared in combination with Summarize Concepts (90%) and occasionally with Present Context (28%). This pattern highlights the role of some *Titles* and *Subtitles* in shaping emotional framing and narrative context rather than presenting data in a purely neutral manner.

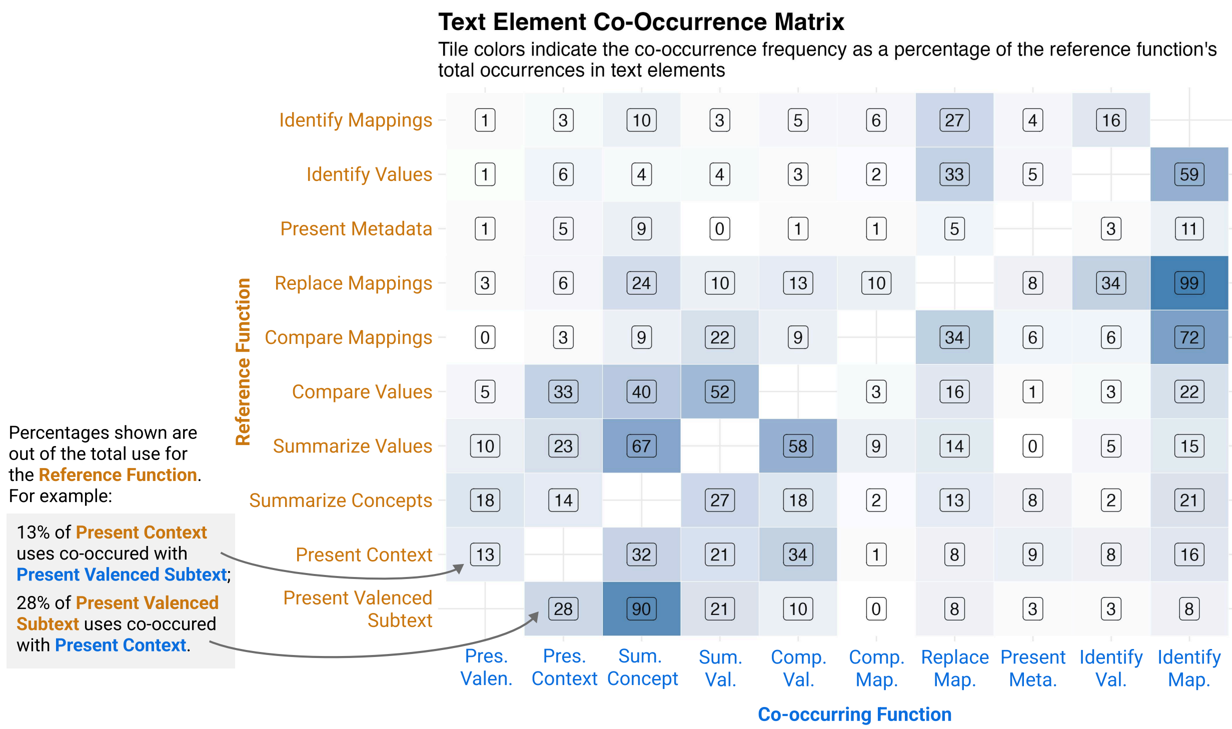


Figure 6.4: Co-occurrence of text functions within individual text elements. Each cell represents the percentage of instances of the reference function (rows) that also co-occurred with the paired function (columns). In other words, rows can be interpreted, " *[value]% of [reference function] co-occurred with [co-occurring function].*" Higher values indicate functions that frequently appear together within the same text component.

Text that performed the Compare Mappings functions frequently co-occurred with Summarize Values (22%), It was also common for Compare Mappings to co-occur (34%) with Replace Mappings; in these cases, the design would typically omit one or both *Axes* in favor of *Annotations*.

By contrast, Present Metadata was largely independent of other functions, with only weak co-occurrence (11%) with Identify Mappings and negligible overlap elsewhere. These findings suggest that although some text functions may combine to form layered interpretative structures, others remain more specialized or independent, serving distinct roles within the visualization.

6.4 Distilling Common Design Patterns

Building on the function framework developed in the previous section, we next examined how these text functions co-occur within real-world visualization designs. Specifically, we ask: **What text design patterns emerge across visualizations?** Because we expected these relationships to be multidimensional rather than categorical, we applied exploratory factor analysis (EFA) [72, 222] to detect latent constructs – underlying patterns that emerge from correlated text and design elements in visualizations. Factor analysis has been used in prior visualization work to reveal structural relationships among design features and user behaviors [10, 184, 199]. Here, it allows us to determine design patterns that reflect the ways text functions operate within visualization layouts.

We conducted the EFA using the `psych` package in R Studio [180, 220], employing a varimax rotation to maximize factor independence and interpretability. Binary features represented the presence or absence of specific properties within each visualization. These included our proposed text **functions**, text **type** (e.g., *Title*), image **domain** (e.g., News), **color** use (e.g., encoding), **visual elements** associated with the text (e.g., arrows), and normalized **word count** measures (e.g,. total word count). See Tab. 6.1 for more information on the text metadata categories. Because IDENTIFY MAPPINGS appeared in every design, it provided no discriminative value and was excluded from the analysis.

6.4.1 Factor Selection

We examined the feature correlation matrix to identify relationships between text properties. As expected, there were strong correlations among certain variables; PRESENT METADATA and *Captions* had the highest correlation ($r = 0.74$). At the same time, many features were only weakly correlated, reinforcing the need for a multifactor solution. To determine the number of factors, we used Bayesian information criterion (BIC), model complexity, and eigenvalues [78]. These measures collectively supported a 4-factor model, which best captured variation in text use while preserving factor independence and interpretability.

6.4.2 Factor Analysis Evaluation

To assess the reliability and robustness of the factor analysis, we implemented multiple validation procedures: (1) masked expert categorization, (2) resampling with intraclass correlation (ICC) assessment, and (3) outlier exclusion. Together, these analyses ensured that the factors reflected stable, interpretable patterns of text use rather than statistical artifacts.

Masked categorization. Three collaborators independently categorized a subset of visualizations from the corpus; they were not aware of which factor these visualization had been assigned (i.e., masked). Using the factor loadings of the identified factors as a reference, each coder predicted which factor best characterized five selected designs. The masked classifications showed strong alignment with the factors derived from the analysis, providing initial support for factor interpretability. The collective loadings of each factor were usable to distinguish between different design patterns.

Resampling and reliability testing. To evaluate the stability of the factor structure, we performed 50 iterations of resampling-based validation. In each iteration, a randomly selected

75% subset of the corpus was analyzed using the same four-factor model. The 75% subset size balanced sample representativeness and computational feasibility. We repeated this process for 50 iterations, determined empirically through stabilization of intraclass correlation (ICC) values. Intraclass correlation coefficients (ICCs) were computed across iterations to assess the consistency of factor loadings. ICC values for all four factors exceeded 0.93 after 50 iterations, indicating convergence and high reliability of the factor solution.

Outlier analysis. We further tested robustness by removing the top 10% of visualizations with the highest factor loadings to ensure that highly weighted designs were not disproportionately influencing the results. The resulting factor structure remained highly similar to the original, confirming that the extracted factors reflected systematic relationships in text design rather than outlier effects.

Together, these validation steps demonstrate that the identified factors represent meaningful and reproducible text-based constructs within the visualization corpus, reinforcing the robustness of our findings.

6.5 Design Patterns Based on Text Functions

Our exploratory factor analysis identified four distinct approaches to text design. We assigned names to these factors to make them easier to reference and understand [184]. While these names serve as interpretive aids rather than strict labels, they summarize the dominant characteristics of each factor. Loadings above 0.6 were considered strongly associated, loadings around 0.5 were moderately associated, and those below 0.3 were considered weakly associated. We interpreted negative values using the same magnitudes (e.g., loadings less than -0.6 were considered strongly inversely associated).

Figure 6.2 illustrates the use of these factors. The original version on the left aligns with `Narrative Framing`. We converted this design to align with the variables that load highly on `Annotation-Centric Design` by shifting the emphasis toward Identify Values and *Annotation*.

6.5.1 Factor 1: Attribution and Variables

We refer to `F1` as `Attribution and Variables`, since this factor showed strong loadings from Present Metadata and Summarize Concepts: Variables, as can be seen in Fig. 6.5. Overall, loadings indicate that text in this factor primarily served a neutral, factual role.

The highest-loading item was *Captions*, which overwhelmingly served the Present Metadata function (93%), following standard journalistic conventions for citing sources and providing attribution [77]. *Subtitles* and *Titles* also loaded on this factor, with strong and moderate loadings, respectively. *Titles* for this factor most often performed Summarize Concepts: Variables, providing no additional framing to the data (87%). Overall,

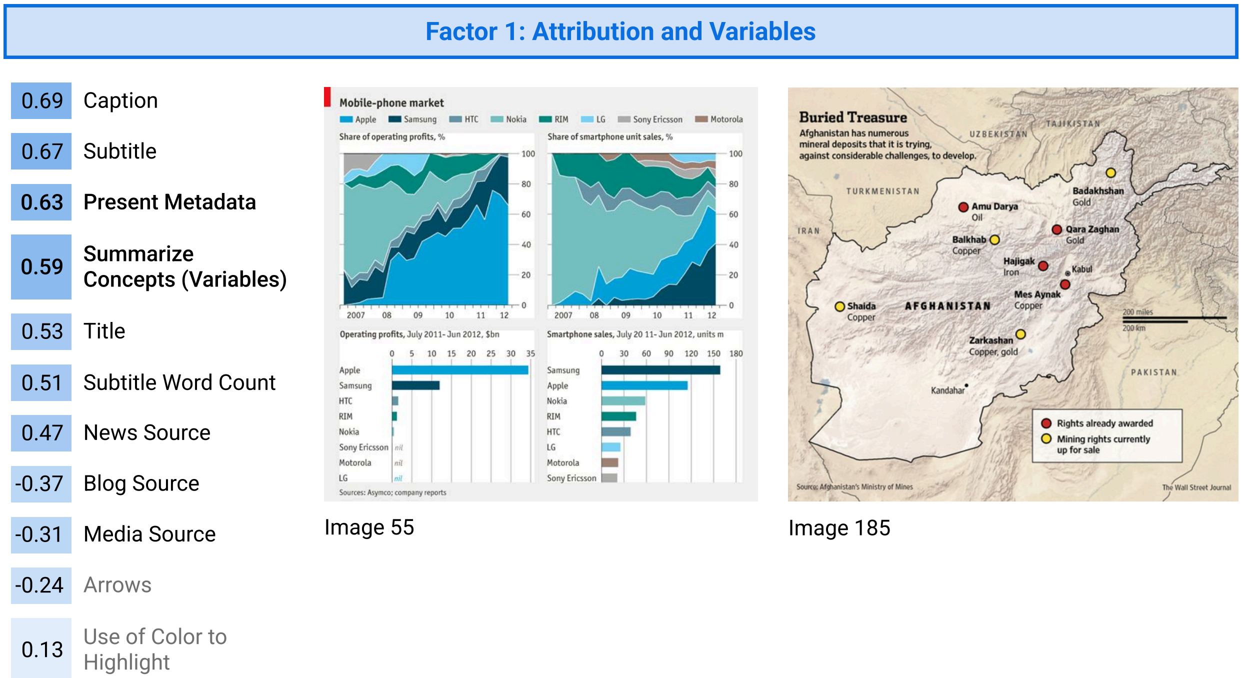


Figure 6.5: F1 (`Attribution and Variables`) variable loadings and representative examples[5]. Loading scores indicate the strength of each variable's association (or inverse association) with the factors. `Attribution and Variables` contained mostly neutral text functions.

`Attribution and Variables` appeared to capture common text design practices in journalism visualizations, with a moderate loading from news sources. This moderate loading likely reflects that metadata-based attribution also appeared in other source categories. Additionally, news visualizations were prevalent in our corpus; they are likely present in multiple factors. Conversely, blog and media-based visualizations were moderately negatively associated with `F1`.

A recurring design theme within this factor involved primarily neutral text; most text served to Present Metadata and Summarize Concepts: Variables. For example, in Image 55 in Fig. 6.5, nearly all non-axis text is dedicated to these neutral functions, producing a design where text organizes and attributes information rather than guiding interpretation.

A second pattern within this factor involved the use of longer *Subtitles* to provide supplementary context, with *Subtitle* word length showing a moderate loading on `F1`. For instance, in Image 239 in Fig. 6.7, the *Title* is the elliptical phrase "Buried Treasure," and the *Subtitle*

[5]Fig. 6.5 image credits: Img 55: Graphic Detail / Economist, September 2012 [48]; Img 185: Dion Nissenbaum / Wall Street Journal, June 2012 [156]

states, “Afghanistan has numerous mineral deposits that it is trying, against considerable challenges, to develop.” The *Title* offers only minimal insight into the data, but the *Subtitle* extends the description by supplying factual context and background.

Although such *Subtitles* introduced contextual information, most examples within this factor remained primarily neutral rather than interpretive. Collectively, these loadings suggested that `F1` represents a broader “fact-based” approach to text design.

6.5.2 Factor 2: Annotation-Centric Design

We refer to `F2` as `Annotation-Centric Design`, since it encompassed a design approach characterized by a high density of *Annotations*, frequent use of text to Identify Values, and substitutions in the form of Replace Mappings. Designs in this factor typically embedded explanatory or quantitative details directly within the chart rather than relying on longer *Titles* to describe data features. Examples can be seen in Fig. 6.6.

Factor 2: Annotation-Centric Design

0.65	Annotation
0.65	Annotation Word Count
-0.63	Title Word Count
-0.55	Government Source
0.53	Identify Values
0.33	Replace Mappings
0.26	Compare Mappings
-0.19	Legend

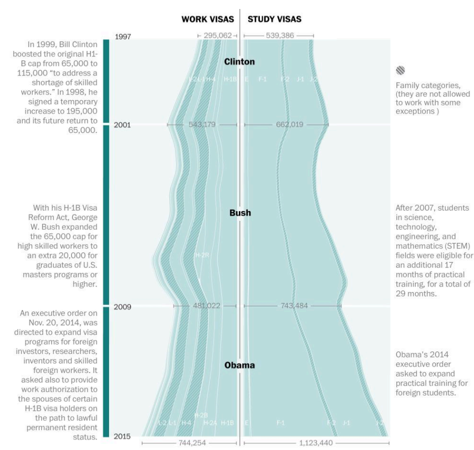


Image 440

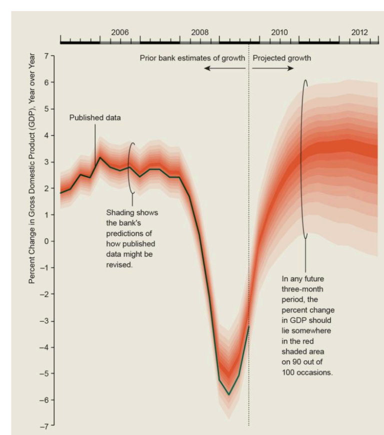


Image 403

Figure 6.6: `F2` (`Annotation-Centric Design`) variable loadings and representative examples[6]. Loading scores indicate the strength of each variable’s association (or inverse association) with the factors. Designs in `Annotation-Centric Design` used a high degree of annotations, often to indicate or explain how to read the design.

This pattern revealed a design strategy that prioritizes in situ explanation, with strong loadings for both *Annotations* and *Annotation* word count, alongside a strong negative loading for *Title* word count. These loadings suggested that information typically conveyed

[6] Fig. 6.6 image credits: Img 440: Samuel Granados / Washington Post, February 2017 [74]; Img 403: Jen Christiansen / Flowing Data [reprint [254]], September 2019.

through a lengthy *Title* is instead distributed across multiple *Annotations* within the visual field, maintaining comparable text density while situating explanations closer to the data they describe. Based on the moderate loading of IDENTIFY VALUES also within F2, these annotations likely also provided additional data precision in many cases.

REPLACE MAPPINGS had a moderate loading in this factor (see 6.3.3.2 for more detail on this function), likely corresponding in part with the use of IDENTIFY VALUES and the weak negative loading of *Legends*. Government-produced visualizations also had a negative loading, which suggests that government designs may rely more on conventional data presentation methods, with fewer *Annotations* replacing traditional mapping elements.

Although not a major component of F2, COMPARE MAPPINGS was weakly associated with the factor. *Annotations* tended to be the primary location for the COMPARE MAPPINGS function (Tab. 6.3). In contrast, COMPARE MAPPINGS had a relatively equal but negative loading on `Attribution and Variables`, indicating that F1 designs tend to rely on more conventional visual mappings rather than creative uses of text, such as COMPARE MAPPINGS or REPLACE MAPPINGS.

6.5.3 Factor 3: Visual Embellishments

We refer to F3 as `Visual Embellishments`, since it captured how text interacts with stylistic and graphical elements within a design, highlighting the use of color, icons, and other graphical elements to add emphasis. Unlike the other factors, which were defined primarily by text functions, variables in this factor reflected aesthetic choices that shape the visual presentation of text rather than its communicative purpose. In addition to using color to style the text, designs in this factor incorporated embellishments, often connecting the text to specific data points (with a circle or line) or higher-level concepts (using logos or icons). These variables and example designs can be seen in Fig. 6.7.

The strongest-loading item in this factor was the use of color in text for stylistic purposes. This included cases where text color serves aesthetic or branding goals rather than data encoding, such as matching a designer's branding. In Image 19 in Fig. 6.7, a section of the *Caption* is rendered in green to align with the logo of the design group (Visual Capitalist), reinforcing the source's visual identity.

Word count also strongly loaded onto this factor, indicating that these designs tended to feature more text components overall. On average, designs in F3 contained 1.5 more text elements per visualization ($Mean = 9.4, SD = 3.1$) than those in other factors ($Mean = 7.9, SD = 3.6$).

Beyond color and word count, the presence of additional visual elements (e.g., logos, icons, circles) was another key feature of `Visual Embellishments`. Most visual elements (87%) appeared alongside *Annotations*, with only a few appearing near *Captions*. None were observed with *Titles* or *Subtitles*, reinforcing that these stylistic additions were primarily applied to supporting text elements rather than to high-level framing components.

While text functions describe what textual content communicates, the combination of variables in `Visual Embellishments` highlights how designers can also use text as an aes-

Factor 3: Visual Embellishments

0.56	Use of Color as Style
0.54	Design Total Word Count
0.52	Logos
0.41	Icons
0.31	Circles
0.30	Lines
0.26	Paragraph
0.20	Use of Color as Encoding

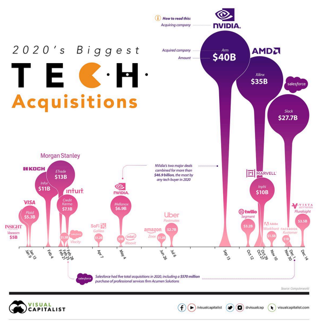


Image 19

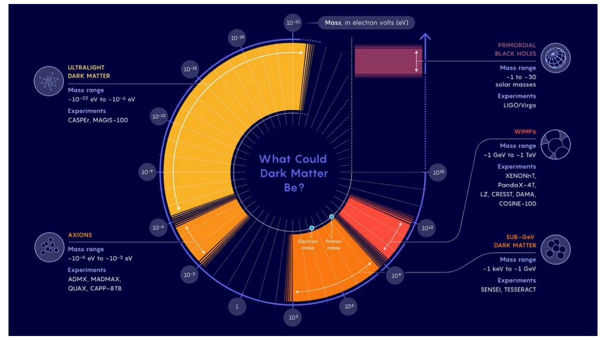


Image 367

Figure 6.7: F3 (`Visual Embellishments`) variable loadings and representative examples[7]. Loading scores indicate the strength of each variable's association (or inverse association) with the factors. Key variables for `Visual Embellishments` primarily considered the display of text elements, rather than the content.

thetic component. The presence of F3 underscores the need for future research into how visual styling may interact with or operate separately from text function in visualization design.

6.5.4 Factor 4: Narrative Framing

We refer to F4 as `Narrative Framing`, due to text usage patterns that aligned with data storytelling practices [178, 189]. This factor was defined by strong loadings for Summarize Concepts: Synthesis and Present Valenced Subtext, followed by Summarize Values and Compare Values, with a moderate loading for Present Context. Together, these loadings described designs that use text to frame data within broader interpretive or emotional narratives; examples can be seen in Fig. 6.8.

Visualizations associated with this factor often emphasized specific takeaways by relating the data to real-world events or human-centered stories. The strong loading for Summarize Concepts: Synthesis indicated that text in these designs distills complex information into concise guiding statements, while the high loading for Present Valenced Subtext suggested that such summaries may incorporate evaluative or emotive language, including

[7] Fig. 6.7 image credits: Img 19: Omri Wallach / Visual Capitalist, December 2020 [236]; Img 367: Samuel Velasco / Quanta Magazine, November 2020 [229]

Factor 4: Narrative Framing

0.71	Summarize Concepts (Synthesis)
0.70	Present Valenced Subtext
0.65	Summarize Values
0.58	Compare Values
0.34	Present Context
-0.34	Social Media Source
0.31	Axis/Axes
0.18	Rectangles
0.16	Flags
0.10	Scientific Source

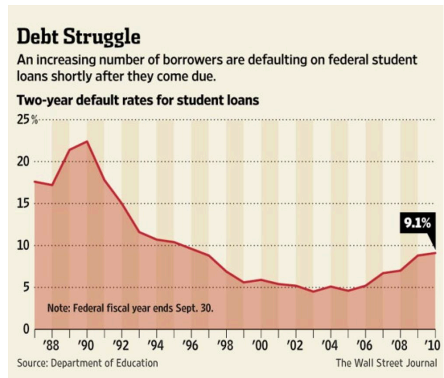


Image 235

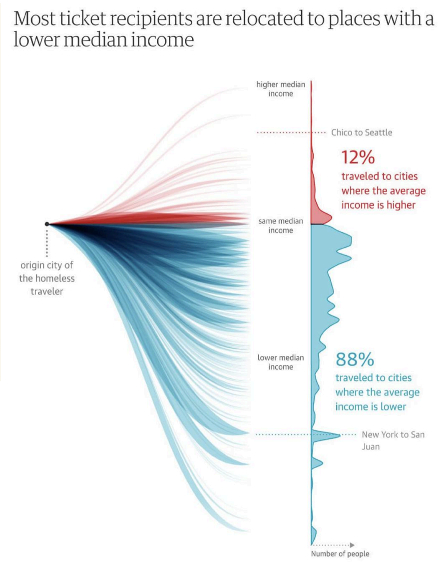


Image 138

Figure 6.8: `F4` (`Narrative Framing`) variable loadings and representative examples[8]. Loading scores indicate the strength of each variable's association (or inverse association) with the factors. `Narrative Framing` designs used text for data storytelling and synthesis.

metaphors or wordplay.

Moderately strong loadings for Summarize Values and Compare Values further characterized this factor through text that highlighted specific data points or relationships. These designs also sometimes included Present Context, which had a moderately weak loading on `F4`. Some of these associations may have stemmed from individual text elements that combined multiple functions within a single phrase or sentence (see Sec. 6.3.3.3 and Fig. 6.4).

Social media visualizations showed a moderate negative association with this factor; charts circulated online may rely less on in-chart synthesis or narrative framing. This absence of embedded framing may reflect the use of accompanying captions or post text on social platforms, which were not captured in the scraped chart images. Although our corpus did not have a high percentage of visualizations that were collected from social media, this suggests an interesting future direction of research to better understand the prevalence of these factors across contexts.

[8] Fig. 6.8 image credits: Img 235: Josh Mitchell and Rachel Louise Ensign / Wall Street Journal, September 2012 [152]; Img 138: Nadieh Bremer / Visual Cinnamon [reprint [32]], December 2017

6.6 Summary

Through this qualitative analysis, we addressed two major research questions: **What are the functions of text in visualization designs, and what text design patterns emerge across visualizations?** By developing and applying a functional framework grounded in real-world examples, we aimed to establish a baseline for how textual elements operate in practice: what they do, where they appear, and how they co-occur. This framework extends previous taxonomies by offering more specific characterizations of textual function, examining multiple text types, considering interactions between text and visual elements, and capturing multifunctionality within a single text component.

A key finding is the prevalence of *multifunctional* text elements. These elements challenge the assumption that each component of a visualization serves a single, fixed role, prompting new design questions: when is multifunctionality effective, and when might it obscure meaning? The framework presented here offers a means to explore such questions systematically, supporting structured audits of existing visualizations through functional analyses, reader assessments, or design reviews. For example, designers could identify where additional text might clarify ambiguous encodings or where overloaded annotations might be better divided into smaller, more focused components.

Beyond its research contributions, this framework supports visualization practitioners who seek greater structure and intentionality in how they use text. By outlining a clear set of functional roles, it provides a practical foundation for planning, revising, and evaluating textual elements. Designers can more effectively align their use of text with specific communicative goals, audiences, or contexts.

The factor-based analysis also illustrates how text functions cluster into distinct design patterns, as exemplified in Fig. 6.2. These factors provide a structured lens for adapting text design based on context. For example, visualizations characterized by `Visual Embellishments` may be well suited for public-facing reports or marketing materials, where color and stylistic embellishment enhance engagement. When presenting the same data to stakeholders, visualizations associated with `Narrative Framing` may be more appropriate, where emphasizing key takeaways and persuasive framing supports interpretive clarity.

We also briefly how these functions and factors might guide AI-generated designs (Fig. 6.9). Using the WSJ data from Fig. 6.2 as input, we gave ChatGPT 5 mini an extensive prompt describing the functions, the factors, the variable loadings of the factors and the goal of producing a chart aligned with `Annotation-Centric Design`. Although generating chart text remains challenging for out-of-the-box models, the result showed a basic grasp of how the functions shape a design. Setting aside the overlapping text, the model's use of functions resembled our own redesign. However, Replace Mappings was not present in the AI design despite their high loadings on the factor. More sophisticated design techniques using these functions may require more iteration or examples to enforce.

By providing a framework of text functions and their co-occurrence patterns, we lay a strong foundation for evaluating the effectiveness, appropriateness, and communicative balance of textual elements. The accompanying codes contribute to a reusable corpus that

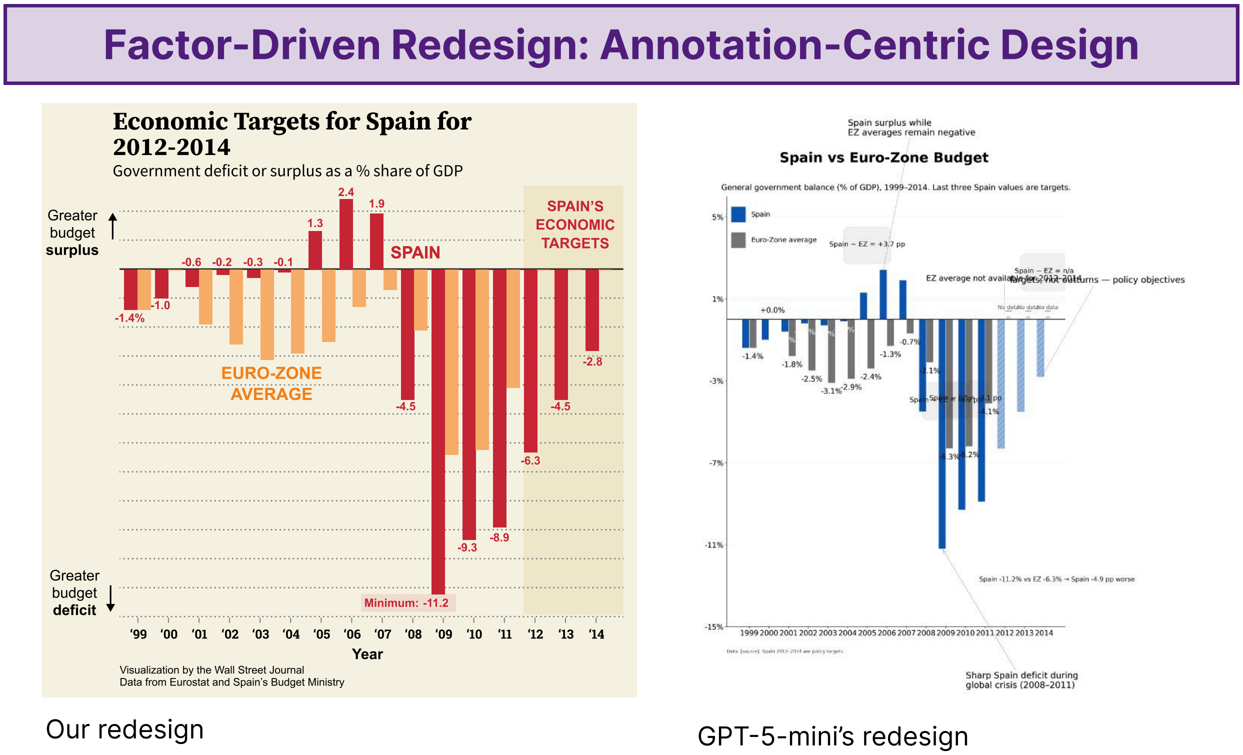


Figure 6.9: Example of applying the functions and factors to an AI-generated redesign (right) compared to our redesign (right), also seen in Fig. 6.2. The original data and design is from the Wall Street Journal [86].

can support future studies of visualization practice and perception.

Ultimately, this work contributes to a deeper theoretical understanding of text in visualization. Researchers and designers can move beyond broad categorizations of text content toward a functional perspective that captures the diverse roles text plays in structuring and shaping visual information. The ability to differentiate designs based solely on textual practices underscores the importance of text as an *active design element*. By systematically defining and analyzing text functions, this framework advances a growing body of research that positions text not as an accessory to visualization, but as a central component.

Chapter 7

Designers Face Persistent Hurdles Using Text

This chapter continues the focus on visualization designers with particular attention to the challenges designers face when incorporating text elements such as titles, annotations, and alternative text into their work. Through interviews with 24 visualization designers across seven industries, we identified six major hurdles: reducing visual clutter, managing dynamic text, understanding and implementing alternative text, avoiding bias, formatting text, and balancing designs across different presentation contexts. These challenges occur throughout the visualization design process but are particularly impactful when the designer is making and deploying the visualization. Designer practice with regards to text is shaped by tool limitations, audience needs, and contextual constraints. While these hurdles were consistently found across design practices, designers have also constructed a variety of strategies to address these challenges, such as dual-encodings, using concise wording, and combining multiple tools to fine-tune the design. Throughout this work, we emphasize a need for improvements for text design in modern visualization tools to better support text integration in visualization design. This chapter contains work from a study conducted in collaboration with Clara Hu and Marti Hearst. I led this project and was responsible for developing the interview protocol, conducting interviews, review of transcripts, coordinating and conducting the coding process, distillation of major hurdles, and the majority of the writing.

7.1 Visualization Design Practice

Designing effective visualizations goes beyond selecting an appropriate chart type or encoding variables. Visualization design requires addressing challenges related to legibility, usability, and accessibility. Although visualization research has traditionally emphasized visual and spatial encoding, recent work highlights the critical role of written language in data communication [203]. Text elements enhance clarity, draw attention to key messages, and help readers interpret complex data patterns [1, 24, 37, 105, 210, 262]. Titles [1, 24, 210], captions [105, 262], and annotations [1, 176, 210] can guide readers through the data and add external context. Earlier chapters of this dissertation (Chapter 4 and Chapter 5) examine how text influences readers' interpretations. Despite these advances, there remains limited understanding of how designers incorporate and manage text within real-world visualization workflows. This chapter addresses that gap.

Prior studies in visualization design practice have emphasized the contextual and iterative nature of design work [167, 168, 188]. Design methods often include sketching, wireframing, interviewing, and usability testing [168]. Frameworks like the Design Activity Framework (DAF) conceptualize this process as four overlapping stages: understanding goals, ideating on design features, making the visualizations, and deploying final designs to product [145].

Recent studies have begun to examine the intersection of text and design. Emerging tools such as InkSight [129] and Epigraphics [261] integrate written key messages directly into the design process. Work in dashboard design highlights persistent challenges around text formatting, live data updates, and determining appropriate levels of detail [214, 225]. We extend this line of inquiry to examine how these challenges manifest across industries and design contexts (e.g., dashboards, reports, or live presentations).

This chapter investigates the questions: **What challenges arise when visualization designers add text to their designs, and how do they navigate these hurdles?** Through a semi-structured interview study with professional visualization designers, we identify key challenges, document strategies for addressing them, and outline opportunities for future research and tool development to better support text integration in visualization design. Throughout this chapter, we use the following definitions:

- *Visualization designer*: A professional who creates visual representations of data as part of a paid role, typically for a specific task or objective (also referred to as a *designer* or *practitioner*).
- *Text elements*: Text content within visual representations of data (e.g., captions, annotations, titles); several examples are visible in Figure 7.1.

7.2 Interview Study Methods

We conducted 60-minute, semi-structured interviews with 24 professional visualization designers to examine how they approach visualization design and how they integrate text elements into their work. These interviews are also explored further in Chapter 8. That analysis does not overlap with these results, as shown in Fig. 7.2. Although text-based visualizations (e.g., word clouds or phrase trees) can include text as a visual element, our scope was limited to standard chart types, including line, bar, and area charts, scatterplots, maps, and dashboards.

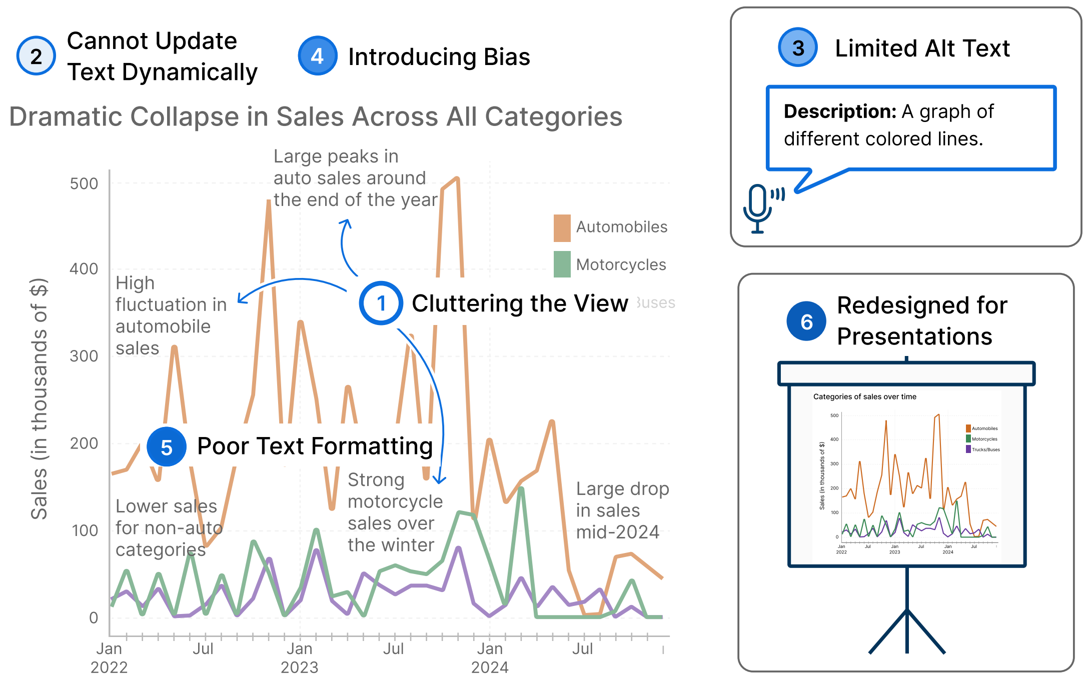


Figure 7.1: Six key challenges designers face when adding text to visualizations: (1) navigating visual clutter, (2) updating text in dynamic designs, (3) providing alternative text, (4) avoiding bias, (5) ensuring proper text positioning and formatting, and (6) varying text use across contexts. Data in these examples is modified from an automobile dataset[1]. The alt text in 3 was generated for this chart by using Microsoft PowerPoint's automatic description generator.

7.2.1 Participants

Participants were recruited through the Data Visualization Society (DVS), News Nerdery, and public posts to X (formerly Twitter) and LinkedIn. Eligibility criteria included being based in the United States, designing or creating visualizations as part of a paid role, and fluency in English. The recruitment materials targeted both data journalists and traditional visualization designers to capture a range of perspectives on the integration of text and visualization. All participants, regardless of their primary role, had experience in visualization design work. Participants in this study are the same as those from Study 1 in Chapter 8.

In total, 24 designers participated in the study, representing a wide range of industries and professional contexts. Most participants (22 of 24) were employed full time. Fourteen participants identified as women, eight as men, and two as trans or non-binary. The majority worked in medium-sized (7) or enterprise-level (11) companies. Further information on

[1] https://www.kaggle.com/datasets/ddosad/auto-sales-data

participants can be found in Tables 7.1 and 7.2.

Table 7.1: Information about (a) participants and (b) work contexts. Overall, participants had a wide range of experience and practice, comprising many different sectors of work. These are also the Study 1 participants in Chapter 8.

(a) Information about participants and their experience in visualization design.

Participant Information	Count
Gender	
Woman	14
Man	8
Trans or Non-binary	2
Years of Experience	
1-3 years	3
4-6 years	9
7-9 years	7
10+ years	5
Time Spent Designing (per week)	
Less than 5 hours	2
5-10 hours	5
11-20 hours	7
21-30 hours	5
30+ hours	5

(b) Information about participants' companies and industries.

Work Information	Count
Company Size	
Micro	3
Small	1
Medium	7
Large	2
Enterprise	11
Industry Sector	
Broadcasting/Journalism	5
Medical/healthcare	4
Scientific or Technical Services	4
Manufacturing	3
Non-profit/Government	3
Software	3
Research	2

7.2.2 Interview Protocol

Interviews were conducted in February and March 2024 over Zoom and lasted approximately 60 minutes. Participants are referenced using assigned ID numbers (P#). All sessions were recorded with participant consent and later transcribed for accuracy and anonymization. Each participant received a $30 Amazon gift card as compensation. An overview of the interview procedure can be found in Figure 7.2; materials can be found on OSF[2]. The study detailed in this chapter also doubles as Study 1 in Chapter 8.

Participants completed a 5-minute pre-interview survey to provide context for their work with visualizations. In the pre-interview survey, participants reported features of their work

[2] https://osf.io/yjsnh/overview

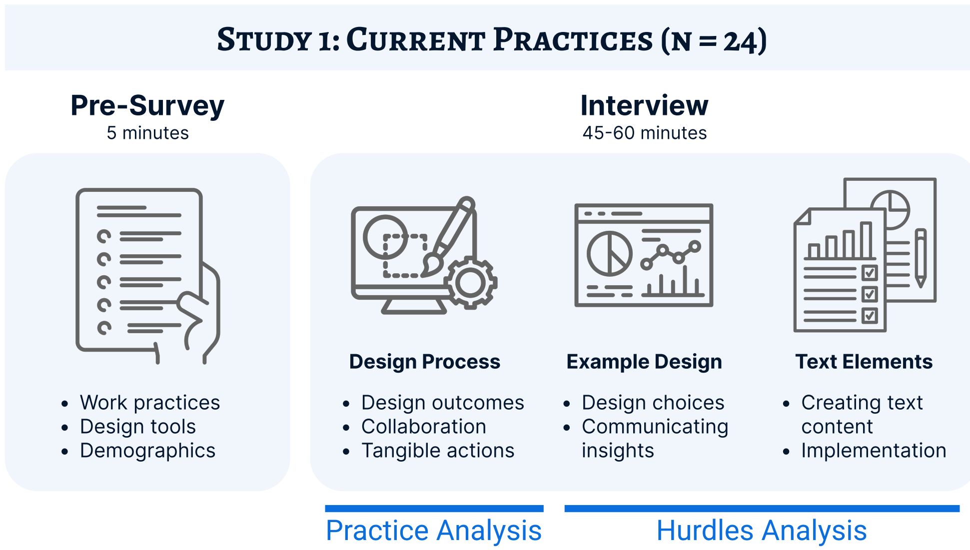


Figure 7.2: Overview of the interview methodology. We conducted two sets of qualitative analyses on the transcripts for this interview. This chapter focuses specifically on the **Hurdles Analysis**. The **Practice Analysis** informs Study 1 in Chapter 8. Data analysis consisted of several rounds of open coding which identified hurdles and relevant solution strategies.

environments (e.g., time per week spent creating or working with data visualizations, company size, the number of collaborators they work with). Participants also listed the types of visualizations they create and the tools used throughout their design process. Finally, the survey collected demographic information such as the participants' gender, employment status, and level of experience (e.g., 1-3 years).

At the beginning of each interview, the interviewer explained the purpose of the study: to understand how designers use language during the visualization design process and how they create and implement the text accompanying their designs. Participants were informed that their data would remain confidential, their participation was voluntary, and they could stop the recording or withdraw at any time.

The interview itself consisted of three major sections, designed to capture both general design practices and text-specific experiences.

Visualization design practices. First, the participant discussed their role and responsibilities in data visualization design, including the specific actions they took to create visualizations (e.g., sketching, writing down questions). They also discussed how they determined what from their datasets to communicate or emphasize and the strategies they used to do so. Example questions from this section included: "What are your specific roles and responsibilities in data visualization design?" and "Do you have a standard design process or use

Design Area	n = 24
Design Outcome	
Dashboards	14
Live Presentations	10
Stand-Alone Charts	10
Text and Visual Reports	10
Enterprise Applications	4
Design Tool	
D3.js	6
Figma	9
Adobe Illustrator	10
Power BI	10
PowerPoint	10
R	10
Tableau	13
Excel	18
Other	15

Table 7.2: Information about participants' design outcomes and tools used to create visualizations, out of a total of 24. These participants are also the Study 1 participants in Chapter 8. "Other" design tools included Python, Flourish, Plot.ly, Datawrapper, Highcharts, Everviz, and web-based languages (e.g., Svelte).

a specific design paradigm?" Follow-up questions typically included queries about different tools, team structures, and design contexts.

Example design. The interviewer and participant would then turn to an example design, selected by the participant prior to the interview. Participants were informed that these example designs would be kept private and not shared as part of any research output, in order to limit concerns about privacy. As such, we do not show any of the example designs discussed by participants. In a few cases, the participant's company did not allow for the sharing of previous designs due to confidential data; in these situations, participants could request that this portion of the interview not be recorded, fill in an existing design with "dummy data" (i.e., simulated information to mimic the structure and characteristics of real data), or use a less recent design example that did not have confidentiality concerns. Example questions from this section included: "Walk me through your design process." and "What were the specific design decisions you made in this example?" This section of the interview allowed discussion that grounded the participants' previous answers about their design process in a representative example of their work. The discussion revisited previous answers and verified subsequent insights.

Integrating text elements. In the final portion of the interview, the participant discussed their experiences with text elements of charts (e.g., titles, annotations) specifically. This text element portion of the interview is the most relevant for this chapter, though we examined the entire interview in the analysis. Example questions included: “Who typically creates the text content for visualizations?”, “What kind of text do you typically consider for your designs?”, and “How do you add text elements to your design?” Questions also addressed accessibility and multi-modality, including the creation of alternative text and the use of other communication modes such as animation or narration. At the end of this section, we explicitly probed for challenges encountered when adding text elements to the visualization to ensure we had not missed any major difficulties.

7.2.3 Analytic Approach

We conducted an inductive thematic analysis to explore how visualization designers incorporate text elements in their designs. This approach derived themes directly from participant data, ensuring that our analysis remained grounded in their language and perspectives.

Another collaborator and I independently reviewed all transcripts multiple times to gain familiarity with the data and to identify potential patterns in participants' descriptions of challenges and strategies. Following established procedures for thematic analysis [29, 42, 45], we each conducted open coding to label relevant excerpts related to the integration of text, including pain points, tool limitations, and collaborative practices. We then met to compare and refine our initial codes, merge overlapping themes, and resolve discrepancies through discussion and review of the raw data.

Following this, the transcripts were re-examined to verify our interpretations of participant comments. The refined codes were then grouped into broader categories that represented recurring themes representing recurring patterns across participants' accounts. Throughout the analysis, we maintained detailed notes to document interpretive decisions in code development. Themes were finalized through consensus, with regular discussion to ensure that each theme accurately represented multiple participants' experiences. Representative quotes were selected to illustrate each challenge and strategy, emphasizing clarity and groundedness. These six key hurdles captured both the barriers designers encountered and the strategies they used when combining text with visualization.

7.3 Design Hurdles for Text in Visualization

We identified six key hurdles that visualization designers encounter across different stages of the design process, summarized in Fig. 7.1 and Tab. 7.3. Each hurdle reflects a distinct but interconnected challenge related to how designers integrate text into visualizations. To contextualize these findings within the design process, we describe which stage of the Design Activity Framework (DAF) each challenge most directly affects, as well as which designer

Design Challenge	Challenge Summary	Current Solutions	Future Development	Context(s) Affected
Reducing Visual Clutter	Minimizing visual clutter while maintaining necessary context.	Dual-encode text with color	Intelligent recommendations to balance design density	Significant space constraints and minimalist style
Managing Dynamic Text	Challenges with updating text content or position in dynamic designs	Neutral, non-descriptive text	AI-driven text generation for data updates	Real-time or interactive data
Understanding Alternative Text	Lack of standardized practices for including alternative text.	Focus on other accessibility features (e.g., text size)	Clearer standardizations and improved alt text automation	Designers with a distributed workflow or specific client requests
Avoiding Bias	Concerns about text biasing readers in data exploration	Neutral labels and concise text	Text guidelines for data exploration	Exploratory data analysis
Formatting and Positioning Text	Difficulty positioning and sizing text accurately	Additional tools (e.g., Figma) for precise formatting	Improve precise text control in current tools	Complete control over text elements
Balancing Different Presentation Contexts	Balancing design requirement trade-offs for different contexts	Tailored text strategies for various contexts	Automatic adaption of text styles and formats	Multiple output formats or screenshots of designs

Table 7.3: Summary of six design challenges identified throughout the design process. The table includes a brief description of each challenge, current strategies employed by designers, opportunities for further development, and the designer groups primarily affected.

groups were most impacted, based on patterns observed in our participant demographics. We also draw connections to prior research to identify broader opportunities for improving text integration in visualization design.

The DAF framework considers visualization design as four overlapping stages:

- **Understand**: Designers define the goals of the visualization, identify the target users, and assess any design constraints.
- **Ideate**: Designers generate and explore creative approaches to communicating the data, often using techniques like sketching, prototyping, and iterative concept development.
- **Make**: Designers translate ideas to a fully functional visualization, using tools and technologies to refine both visual and textual elements.

- **Deploy**: Designers integrate the design into its real-world setting, emphasizing decisions related to implementation, such as compatibility with existing software or databases.

7.3.1 Reducing Visual Clutter

Many current visualization design practices favor a minimalist aesthetic [16, 227]. A majority of participants (20/24) described difficulty balancing textual and visual elements, with two-thirds (16/24) citing visual clutter as a persistent challenge. While nearly all participants recognized the value of text for providing context (22/24), communicating key insights (15/24), or improving reader understanding (14/24), they struggled with the tension between informativeness and simplicity. As P11 noted, adding several text elements to an already complex chart could make it feel, *"way too busy way too fast, and then you have too many things going on."* Designers were therefore tasked with finding a balance between enhancing comprehension and avoiding cognitive overload.

This challenge was particularly pronounced because of limited space within visualization layouts. As P9 explained,

> "It's just a real estate problem, you know, trying to put [text] where there's enough white space around it that people actually see it and read it. So it's often just a trade off with balancing [visuals] against other text."

Because the visual elements were typically tied to underlying data, adjusting them could compromise accuracy or meaning. As a result, designers often focused on modifying text elements (e.g., annotations, captions) to manage clutter while preserving key insights.

7.3.1.1 Designer Context: Customizing Designs with Space Constraints

This hurdle arose predominantly during the *make* stage of the design process, when participants began assembling final layouts and the physical constraints of the design became apparent. At this stage, the space constraints of the final product became tangible, forcing designers to make compromises.

The issue was most acute for stand-alone charts and enterprise applications (11/11 participants), where designers had more flexibility to customize textual elements and thus faced greater risk of overcrowding. On the other hand, participants working in nonprofit and government contexts reported fewer issues with clutter (1/3 participants). These organizations were more likely to use templated report formats that restricted text customization.

7.3.1.2 Strategies for Addressing Hurdle: Dual Encoding and Interactive Text

To mitigate visual clutter, many participants chose to reduce the use of text in their designs, believing that text might overwhelm the visualization. However, research suggests that readers often prefer context-rich visualizations and are less hindered by clutter than designers may anticipate [204] (see Chapter 3 for details). In order to strike the right balance between content and clutter, participants (17/24) also shortened text elements, using concise phrasing to convey essential information without occupying excessive space. Shortening text took several rounds of iteration, often (11/24) incorporating feedback from other designers or members of the participant's team. P15 mentioned frequent conversations with their design team centering around the key question, *"How do you rephrase that [content] in a way that it keeps its original meaning?"*

Another common strategy involved combining color and text through dual encoding (8/24), thereby eliminating separate legends and allowing text to serve multiple communicative roles [63, 172] (also discussed in Chapter 6). Emphasizing a clear visual hierarchy could also support the effort to reduce visual clutter. If designers de-emphasized less critical elements by using smaller fonts or lighter colors (while maintaining readability), the chart itself may appear less visually complex.

In interactive settings, designers leveraged details on demand, shifting some annotations into tooltips or secondary views. This approach allowed readers to access the same information while keeping the initial view clean and focused. While not yet supported by major tools, layout optimization could help designers easily adjust text placement and spacing, enabling cleaner and more effective layouts.

7.3.2 Managing Dynamic Text

Designers frequently struggled with the challenges of dynamic or data-driven text, particularly when text needed to update in response to reader interactions or live data changes. Participants (9/24) described situations in which text content or position had to shift based on filters, user-specific data, or continuously updated datasets (e.g., regularly updated COVID-19 dashboards). In these contexts, static annotations or fixed textual descriptions became impractical. As P12 explained, aligning text with varying data displays was a persistent obstacle:

> "I could [annotate] if I have a single static version of this. But if I have this dynamic view that's going to look different for all the people that are going to look at this, it's not going to be in the right place for all of them depending on the display."

Beyond the placement of text, the same participants also faced barriers in updating content dynamically, particularly for user-specific details such as personalized annotations or account-based summaries. Designs that updated with new data or users required manual adjustments for annotations, an approach that was time-intensive and unsustainable at scale. These findings reinforce prior research on dashboard design, which highlights similar barriers related to dynamic content generation and textual integration [214].

7.3.2.1 Designer Context: Providing User-Specific or Live Data

Difficulties with dynamic text primarily arose during the *deploy* stage of the design process, when designers implemented visualizations in their final software environment. At this stage, tool limitations often prevented text from updating automatically to reflect changing data or user inputs. The hurdle was most pronounced in designs that incorporated live or user-specific data, such as enterprise dashboards, technical monitoring systems, and interactive news applications. These types of projects were especially common among participants working in scientific or technical services (e.g., consulting).

7.3.2.2 Strategies for Addressing Hurdle: Language Models and Adaptive Templates

There were few effective workarounds for the challenges posed by dynamic displays; most participants who faced this issue resorted to removing detailed annotations altogether or replacing them with generic placeholders to avoid misalignment or factual errors. P23 shared one instance of using AI to dynamically update text, but this application was limited to numerical values, not annotations or contextual insights. While manual updates for text elements were possible in some specific cases, this approach was impractical for most dynamic design situations. Overall, dynamic text was a persistent pain point in current visualization tools and workflows.

Advances in NLP and AI provide an avenue for addressing these challenges. Systems such as Contextifier [89] and DataTales [215] have demonstrated how saliency analysis and LLMs can be leveraged to identify interesting data features and assist in generating meaningful annotations. Recent work on displays of financial data demonstrates the feasibility of using LLMs for adhoc annotation generation [76]. Such systems focus on supporting data storytelling using the combination of text and visual information and represent a promising solution for dynamic visualization challenges.

A practical, less computationally intensive solution could involve the use of adaptive templates [83, 94] for positioning dynamic text elements. Designers could create templates tailored to specific data structures and visualization types, embedding conditional rules to control the placement and formatting of text based on the data appearance or changes. Conditional display rules could specify when certain text elements, such as annotations or tooltips, appear or adjust based on user interactions or filtered views. This approach would use modular design principles to streamline dynamic text management without relying on computationally intensive resources.

7.3.3 Understanding Alternative Text

Text plays a crucial role in visualization accessibility, particularly for ensuring inclusive design for blind or low-vision (BLV) users. Alternative text descriptions (alt text) are a central component of accessibility, providing textual representations of visual elements so that data can be interpreted without visual perception [52, 96, 134]. Although not visible within the visualization itself, alt text highlights the critical intersection between text and accessibility in visualization design and demonstrates how design decisions extend beyond aesthetics to usability..

Despite this importance, participants often overlooked alt text in their workflows. Many (10/24) reported not including alt text in their designs at all, with client-facing designers noting that clients *"haven't had that asked for yet,"* [P10]. Another group of participants (9/24) created alt text inconsistently, often citing the absence of organizational policies or established workflows. As P22 explained, '*It's something that we are working on. We don't have any policies or or protocols in place to do it for everything, which I think is a little unfortunate."*

Only a small minority of participants (5/24) viewed alt text as an essential part of the design process. As P4 summarized:

> "It's not really built into our structure to always write [alternative] text... The graphics I write are code and are not necessarily one image. So there isn't as strict of an alt text field necessarily. But frankly - we need to be better with accessibility."

Instead, participants tended to focus on other aspects of accessibility, particularly color contrast and text size. About one-third (7/24) mentioned designing for colorblind users, with P16 emphasizing, *"your chart might look beautiful, but it might be awful for somebody who's color blind."* Additionally, nine participants described attention to font size and adherence to accessibility standards such as WCAG AA [232]. A few participants (3/24) discussed internationalization or making visualizations accessible to non-English speakers. While these efforts broaden access, they do not address the full range of accessibility needs. Alt text and related practices remain a critical but underutilized area of visualization design, where improved standards and integration could enhance inclusivity across audiences.

7.3.3.1 Designer Context: Managing Distributed Workflows and Responsibility

Alt text creation typically occurred during the **deploy** stage of the design process, often as an afterthought once the visualization was already finalized. This approach contributed to its inconsistent implementation and reinforced the perception that accessibility was separate from, rather than integral to, design.

The challenge was most pronounced in manufacturing and healthcare industries, where none of the seven participants reported including alt text. These fields often involve highly interactive or complex visual systems and distributed workflows in which design, development, and accessibility are handled by different teams. This fragmentation made it difficult to assign responsibility for creating alt text or verifying its accuracy. Designers in journalism and scientific or technical services faced this challenge less frequently (4/9 participants).

Journalists in particular often interacted with publishing platforms and content management systems that provided built-in fields or accessibility checks.

7.3.3.2 Strategies for Addressing Hurdle: Standardized Practices and Text Generation

Participants emphasized that standardized accessibility practices would make the inclusion of alt text more consistent and achievable. Clear organizational guidelines and templates could help designers determine what aspects of a visualization to prioritize (e.g., data trends or specific values) while reducing reliance on ad-hoc approaches. Standardization should also assign responsibility for accessibility, ensuring accountability across teams. Integrating these standards into the design process, rather than adding them retroactively, would make accessibility efforts more transparent and sustainable.

AI-assisted tools also present promising opportunities to support alt text creation. Using Vision-Language Models (VLMs) or combining traditional computer vision with LLMs can automate the generation of descriptive text by analyzing both the visual structure and underlying data of a chart, improving the accuracy of descriptions [195, 257]. Designers can then review and refine these automatically generated descriptions to ensure alignment with audience needs and communication goals.

While commercial platforms such as Power BI offer basic capabilities, the generated text is often too simplistic. For example, in Fig. 7.1, the example design presents three data series in a multivariate line chart showing trends in sales across transportation types. There are many data features that may be of interest to a reader, but the alt text generated by Microsoft PowerPoint in late 2024 simply reads, “A graph of different colored lines.” In late 2025, the output is word-for-word identical, with the added caveat: “AI-generated content may be incorrect.” Alt text generation may be more successful for charts created in Excel that are inserted into PowerPoint, but the chart-as-image processing capabilities of the current pipeline are significantly limited.

Finally, participatory design practices offer an essential path toward improving accessibility. Collaborating directly with BLV users to test and refine alt text can ensure that descriptions meet user needs and reflect real accessibility challenges [134, 238, 239]. Incorporating usability testing and feedback loops focused specifically on accessibility would help organizations identify gaps in their workflows and prioritize meaningful improvements. Such practices would not only expand access to visualization but also strengthen the inclusivity of design practice as a whole.

7.3.4 Avoiding Bias

When text highlights or interprets specific aspects of a chart, it can influence what readers take away from a visualization [108, 210]. Nearly half of participants (11/24) expressed concern that their use of text might unintentionally introduce bias, shaping how audiences interpret data or emphasizing particular trends over others. In this context, bias referred to the potential for text to guide readers toward a specific interpretation, potentially obscuring alternative perspectives or the raw data itself. P16 emphasized the importance of neutrality in textual elements:

> "We have what we call the chart title, which is non analytical, very explicit, objective text describing the chart. This is not the place to, you know, make your commentary. This is just a place to describe what the chart is doing."

Adhering to this ideal of neutrality, however, was often difficult, especially in situations where the data revealed an important message or trend that needed emphasis. As P20 reflected, "It has to sound very objective sort of, even though there's always a point of view." This tension between neutrality and narrative framing mirrors findings from prior research presented in this dissertation (Chapter 5) showing that readers perceive visualizations as more biased when designers explicitly support one interpretation [202]. Yet, even participants most wary of bias acknowledged the communicative importance of text. Titles, captions, and annotations were viewed as essential tools for highlighting context and key insights, though as designers grappled with how to use them responsibly.

7.3.4.1 Designer Context: Prioritizing Data Exploration

Concerns about bias most often arose during the **ideate** stage of the design process, when participants explored alternative ways of framing and presenting their data. At this stage, participants faced internal questions about how to communicate data clearly without leading the audience toward a predefined conclusion.

These concerns were particularly pronounced among participants working in research, consulting, and science or technology services (3/4). In these contexts, visualizations were often created for external stakeholders or clients who needed to form their own interpretations or for situations where users should be encouraged to explore data independently. Maintaining objectivity and enabling independent exploration were seen as markers of credibility and professionalism. Designers described being careful not to frame visualizations in ways that might be perceived as persuasive, even when underlying data trends were unambiguous. This challenge was also relatively common in the creation of stand-alone charts, which tended to traditionally lack the supporting context or verbal explanation that might accompany a dashboard or live presentation.

7.3.4.2 Strategies for Addressing Hurdle: Conciseness, Transparency, and Progressive Disclosure

To mitigate potential bias, participants often opted to use minimal or purely descriptive text. Some chose to label features without interpreting them, allowing readers to draw their own conclusions. As P5 described, *"Here's the dashed line, and I'll label that... You can interpret for yourself if the metrics changed before or after that point."* Designers frequently used short phrases or single keywords rather than full sentences to avoid implying causality or intent.

As P21 explained, they preferred *"very basic, simple, small sentences or even key words"*. By keeping text brief and factual, participants aimed to support comprehension while minimizing interpretive influence. Other uses of text (10/24) were instructional, describing the kinds of data being shown or how to interact with or interpret the chart rather than interpreting the presented information.

Still, many recognized that complete neutrality was neither possible nor desirable. Most participants (17/24) spoke about the importance of text for helping readers quickly grasp overarching insights. As P9 put it, text elements made it *"easy for [readers] to extract a high level understanding"* of the data. Thus, these two goals are in tension with each other. Results from Chapter 5 indicate that using more factual or statistical descriptions of the data can help to reduce overall perceptions of bias.

While participants rarely mentioned them directly, additional strategies from prior research could help manage this balance. Progressive disclosure [142, 228] allows viewers to form initial interpretations before revealing explanatory text, helping preserve autonomy while still offering context. Visualizations as a whole could also promote a higher degree of trust through more transparency and description of data provenance (e.g., notes explaining the source of the data, methodologies used, and any assumptions made). In collaborative or public-facing projects, community or crowdsourced feedback could serve as an additional check, helping designers identify language that unintentionally signals bias.

7.3.5 Formatting and Positioning Text

Formatting text elements in visualizations, from basic data labeling to the placement of annotations and titles, posed significant hurdle for most designers (16/24). Achieving the intended visual and typographic appearance was often hindered by overcrowding, tool limitations, and inconsistencies across design platforms. P11 highlighted a common challenge: *"a ton of labels that are all on top of each other and impossible to read."* Direct data labels were among the most common forms of annotation (as also discussed in Chapter 6), but they frequently overlapped, particularly in dense visualizations with multiple data points.

Beyond label placement, participants described recurring frustrations with formatting titles and ad-hoc annotations. Tools such as Excel or Google Sheets imposed constraints on title length or text boundaries, truncating content that exceeded predefined layout areas. For annotations, controlling spacing, font size, and alignment to ensure readability required extensive manual adjustment. Even when designers had a clear vision for their layout, implementing it across tools demanded substantial effort. As P10 explained:

> "In Figma, we've got the correct sizing... We've got it all laid out.. We want to make it exactly like this [in Tableau]. It's tedious. It's time consuming to put in spacing and padding, to make sure that my fonts are the correct size."

Transferring designs between tools added another layer of complexity. Participants frequently recreated text formatting multiple times to adapt to different environments, a process that introduced inconsistencies and errors. This issue was also particularly difficult to navigate for the few participants who designed visualizations for mobile contexts (3/24), where limited screen space amplified layout constraints.

7.3.5.1 Designer Context: Implementing Design with Tool Constraints

These formatting challenges primarily occurred during the *make* stage of the design process, when designers transitioned from low-fidelity prototypes to finalized visualizations. Participants in research, software, and technological industries (7/9) encountered these issues most acutely. Participants in the journalism industry encountered formatting issues less often (3/5 participants), possibly since their organizations relied on established style guides and templates that standardized text formatting.

Participants designing enterprise applications also faced this issue frequently (4/4 participants), since they typically created prototypes in one tool and implemented the final products in another. These projects often required strict adherence to platform-specific constraints, making it difficult to replicate the precision achieved during the prototyping phase. Replicating the precision of a mock-up during implementation was thus a time-consuming and error-prone process.

7.3.5.2 Strategies for Addressing Hurdle: Combination of Tools and Increased Flexibility

In some cases, participants opted to omit select text elements from their designs to preserve visual clarity. While this strategy effectively avoided formatting challenges, it came at the cost of sacrificing some of the visualization's communicative potential.

To navigate this hurdle more effectively, participants often treated text addition as a separate step in the design process. They created the base visualization in tools like Excel (18/24) or Tableau (13/24), then exported the visualization as an image to refine text elements in more flexible design platforms such as PowerPoint (7/24) or Figma (6/24). As P20 explained, "we've got those visual encodings [from Excel or Tableau], and [I] add the text in PowerPoint, Keynote or Illustrator." While this multi-tool approach allowed for precise adjustments to text placement, font sizing, and spacing, it also added added complexity and increased production time.

At a broader level, increased standardization of text-related properties across visualization tools could help minimize the need for manual formatting and tool-switching. Integrating responsive design principles for text elements could ensure legibility and proper alignment across devices and screen sizes [83, 106]. Likewise, batch-editing capabilities for annotations or labels could streamline formatting efforts. Improved text formatting is critical for supporting well-designed, highly communicative visualizations without overburdening the designer.

7.3.6 Balancing Different Presentation Contexts

Participants frequently (19/24) had to consider how text and other design elements would translate across design outputs, such as interactive dashboards, static screenshots, and live presentations. Most participants (17/24) designed visualizations for more than one presentation context (Table 7.2). These shifting contexts required designers to balance trade-offs between minimalism and detail. Live presentations typically favored minimalist layouts to avoid clutter and allow presenters to provide explanations verbally. In contrast, static dashboards or reports demanded more comprehensive annotations to ensure clarity in the absence of a narrator. This tension was especially acute when a single visualization had to serve both purposes, such as being displayed during a talk and later circulated in a report.

For many participants, the need for multi-purpose designs was a source of ongoing trade-offs. While dashboards often served as the primary design product, screenshots of those same dashboards were widely reused in slides or documents. As P11 noted, sharing a screenshot was *"a primary use case"* P23 described the balancing act this required:

> "With this dashboard, I had to balance screenshots versus interactivity. So [users] would take a screenshot of this [chart], maybe of those [charts]. But then they could get more information by looking at the tooltips or the other pages. So you see, there's a design challenge here."

Some participants framed this issue less as a technical constraint and more as a question of audience accessibility. P7, a designer working on enterprise applications, described the challenge of addressing distinct audiences, explaining that they share the same design with *"the buyers of our product and... the users of our product."* In this case, the design had to communicate broader business value clearly to the buyers while also including the technical detail and language necessary for end users to extract meaningful insights from the data.

7.3.6.1 Designer Context: Understanding Diverse User Perspectives

This hurdle typically arose during the early understand stage of the design process, when designers identified their target users and anticipated the contexts in which the visualization would appear. Participants across several industries (10/24) described grappling with competing audience needs, but for different reasons. In journalism, the issue stemmed from the need to make visualizations accessible to both general audiences and expert readers, balancing readability with analytical depth. In research and manufacturing contexts, the challenge centered on the dual use of the same design (e.g., dashboards for internal monitoring and reports for stakeholders). These varied demands often forced designers to create compromises that prioritized some audiences or uses over others.

7.3.6.2 Strategies for Addressing Hurdle: Interactivity and Minimal Design

Designing for multiple contexts often led participants to minimize visible text, particularly in live presentations where spoken commentary could provide additional explanation. As P5 explained, *"I will leave a lot of text out [for a slide deck]. Sometimes I'll put it in the speaker notes if I'm distributing it, and I want someone to see it."* Participants also described using progressive disclosure, revealing textual information incrementally to prevent clutter and maintain focus. In such cases, presenters (not designers) often added their own visual or text annotations.

When text was omitted from the main visualization, designers sought other channels for providing contextual information. In interactive dashboards, many relied on tooltips or layered interactions to present additional details on demand. This approach preserved a clean visual layout while allowing users to access deeper information as needed. However, a few participants (3/24) restricted the use of tooltips in dashboards due to low user engagement with these interactive features.

Context-aware design [54, 230] could also help address the challenge of adapting visualizations for multiple contexts by dynamically tailoring text and other elements based on the intended use or audience. For example, a visualization might automatically reduce annotations and increase font sizes for a live presentation mode, while expanding text context for a static report. This approach would reduce the need for manual adaptation and ensure that visualizations remain effective across different use contexts.

7.4 Summary

This study identified six recurring hurdles that visualization designers face when incorporating text into their work: reducing visual clutter, managing dynamic text, understanding alternative text, avoiding bias, formatting and positioning text, and balancing different presentation contexts. Each challenge reflects a different tension in the design process, revealing how visualization designers think about text as both an opportunity and a constraint within their workflows. Together, these hurdles reveal that text integration is not a discrete task but a continuous negotiation across stages of visualization design. Many challenges share common stages or focus areas, highlighting the interconnected nature of visualization design, particularly during the *make* and *deploy* stages of design.

Hurdles such as **formatting and positioning text** and **reducing visual clutter** both primarily occurred during the *make* stage, where designers refine layouts and optimize aesthetics. These challenges are closely linked: difficulties in formatting text elements can exacerbate issues with clutter, as poorly positioned or oversized text disrupts the balance and visual hierarchy of a design. Similarly, **managing dynamic text** and **understanding alternative text** both occurred during the deploy stage, highlighting how existing software makes it difficult to maintain accuracy and accessibility once designs move into production. Designers want to use text in flexible, context-aware ways but are constrained by the systems through which they work.

Other challenges, such as **avoiding bias** and **balancing different presentation contexts**, pointed to the social and rhetorical dimensions of text in visualization. Avoiding bias is critical for neutral data exploration, but providing clear explanations is essential for complex visualizations or for audiences unfamiliar with the data. Knowing the audience and context of use for a given design is essential to support informed decision making with text. Thinking of multiple presentations contexts at once requires designers to explicitly consider their audiences and the ways in which they may view or interact with the data provided. **Understanding alternative text** adds another layer of audience consideration. While there may be cases where designers know that their audience is entirely sighted (e.g., designing for a specific client team), creating text descriptions can be beneficial to audience members who may prefer text over visuals [79, 81, 210] (see Chapter 3).

The visualization design process is iterative and multifaceted, involving multiple tools, diverse requirements, and needs of various audiences. Text is a crucial component of visualization design, influencing how information is communicated, interpreted, and understood. Addressing challenges of text design offers an opportunity for researchers and tool developers to better support designers in creating effective data visualizations. With these understandings of how visualization designers grapple with text design, we turn to another study of the visualization design process, this time examining how *writing* while designing may help to focus the design and document decision rationale.

Chapter 8

Writing Can Guide Visualization Design

This chapter examines how written language can support the early stages of visualization design through the use of "writing rudders," short written artifacts that help designers articulate goals, questions, or possible takeaways. We conducted two interview studies with professional visualization designers to document current writing practices and to evaluate four rudder variants: key questions, possible conclusions, narratives, and potential titles. Overall, writing was rarely used deliberately in designers' workflows, and when it appeared, it tended to surface early in the process to clarify goals or context. Designers responded most positively to writing key questions and possible conclusions, which they felt helped focus the design and maintain alignment with user needs. Narratives and titles were viewed as more appropriate for later stages and raised concerns about biasing data exploration. Together, these studies suggest that lightweight writing rudders can guide early design decisions and provide useful artifacts for evaluating emerging or finished designs. This work represents one of the first examinations of how writing may support visualization design outside of its role in crafting final text elements. Study 1 in this chapter is the same interview protocol and participant set from Chapter 7, but we apply a new research question and set of codes, conducting the Practice Analysis shown in Fig. 7.2. This chapter also contains work from a previously published study conducted in collaboration with Clara Hu and Marti Hearst [205]. I served as first author and was responsible for developing the interview protocol, conducting and coding interviews, synthesizing themes, and a majority of the writing. This content has been edited for clarity and coherence with this dissertation.

8.1 Writing in Design Processes

The study and practice of visualization design puts great attention on the creation of visual elements. Activities that produce visual artifacts, such as sketching and wireframing, are central steps in visualization design [167]. However, taxonomies and studies of design practice have not focused on the use of *writing* in the design of visualizations. Insights from other fields indicate that writing preliminary notes, questions, or outlines is a useful step for clarifying goals and shaping direction. In library science, for example, researchers have long advised writing down the information need as a prelude to effective search [182, 244]; guidance for effective web search similarly recommends articulating questions in advance [183].

For more creative fields, research on essay writing instruction also finds benefits in prewriting steps, including writing outlines, lists, notes, or concept webs [73]. In visual endeavors such as animation and film, scripts and screenplays provide a narrative foundation for eventual visual output. Visualization design is both creative and analytical, yet the potential role of writing in the design process remains underexplored.

Our interest in this topic also stemmed from our own experience designing the chart stimuli for the preference studies in Chapter 3. When creating charts with many annotations, we first struggled to determine what the narrative of the visualization would be and how we should write the accompanying text. We took a step back from the visualization itself to write a short paragraph describing the data and its possible story. After writing this narrative, we found the process of designing the text for the chart stimuli to be easier and better scaffolded.

In this work, we examine how language can support early design framing through what we call **writing rudders**. A **rudder** refers to a mechanism to steer a boat, as well as more metaphorically, "a guiding force or strategy" [146]. Writing rudders serve this purpose by providing direction in the design process and maintaining focus on the message and goals of the project. These messages may shift over the design process, just as a boat's rudder may be pivoted to move the boat in a new direction. The purpose of a rudder is to *guide* the design, similar to how a sketch acts as a starting point for determining visual representations.

Researchers have tangentially examined forms of writing in visualization design workflows, but in ways distinct from rudders. Writing commonly appears in user research summaries, design documentation, or formal design requirements [145, 154, 237]. Design requirements are the most similar to a writing rudder but typically specify technical constraints, user needs, data inputs, and project limitations, rather than guiding the design process or story of the design.

This chapter investigates two research questions. First, **how do designers *currently* use writing during the design process?** Prior research has not explicitly examined this question, even as writing may be present informally or implicitly. Second, **what is the perceived *impact* of writing rudders on the design process?** We introduce and test several rudder variants to evaluate how designers respond to these interventions in early-stage design work.

Throughout this chapter, we use the following definitions:

- **Visualization designer**: a professional who creates visual representations of data as part of a paid role, typically for a specific task or objective. Also referred to as **designer** or **practitioner**.

- **Design process**: the dynamic, iterative set of activities undertaken while creating visual representations from raw data. Also referred to as design **practice**.
- **Writing rudder**: hand-written or typed language created and/or used during the design process, describing the message, story, or key goals of the design itself. Also referred to as a **written rudder** or **rudder**. A **rudder variant** or **variant** is a specific form of this language.
- **Text elements**: written content within visual representations of data (e.g., captions, annotations, etc.).

8.2 Interview Study Methods

Study 1 was the same study examined in Chapter 7 with the same set of participants. A new set of participants was recruited for Study 2. These two semi-structured interview studies with professional visualization designers investigate how writing appeared in current design practice and how designers responded to writing rudders. Because the structure, recruitment approach, and analysis procedures were similar across both studies, I describe them together here. An overview of the full study process is shown in Fig. 8.1. More detail on Study 1 procedures and participants can be found in Sec. 7.2.

8.2.1 Interview Protocols

Across both studies, participants completed a short pre-interview survey followed by a 60-minute semi-structured interview conducted over Zoom. All interviews were recorded and automatically transcribed. Participants were compensated with a $30 Amazon gift card. Transcriptions were manually reviewed for accuracy. The first section of these interviews contained questions about the participant's role and responsibilities in visualization design. Details on the pre-interview survey and the visualization design process interview questions can be found in Sec. 7.2, Fig. 7.2, and Fig. 8.1.

After discussing the participants' visualization design process, the two studies diverged in methodologies. Study 1 participants went on to discuss an example visualization they had selected prior to the interview, ending the interview with a segment on participants' experiences with text elements in visualizations, such as titles, annotations, and captions. This final portion of the protocol is examined in detail in Chapter 7; only questions related to design process and writing are analyzed for this chapter. More detail on specific Study 1 procedures can be found in Sec. 7.2. Study 2 went on to discuss and test writing rudder variants.

Introducing rudders. The interviewer introduced four writing rudder variants that could be used at the beginning of the design process, before bringing data into a visualization tool:

- **Key questions**: Write down the key questions that a user/reader may use the visualization(s) to address.

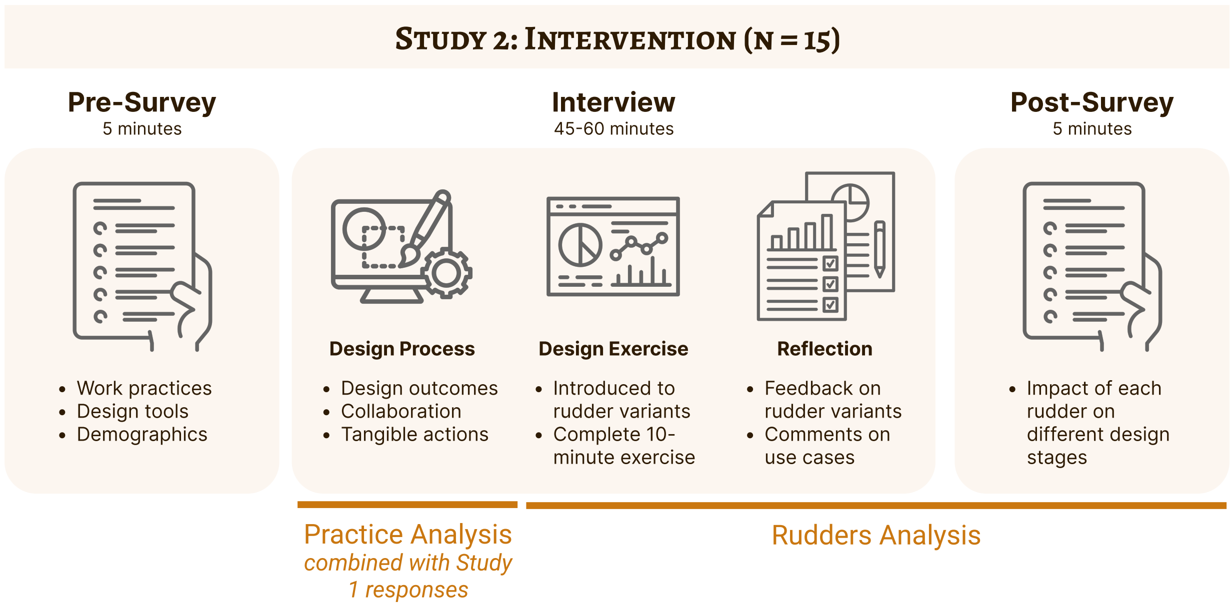


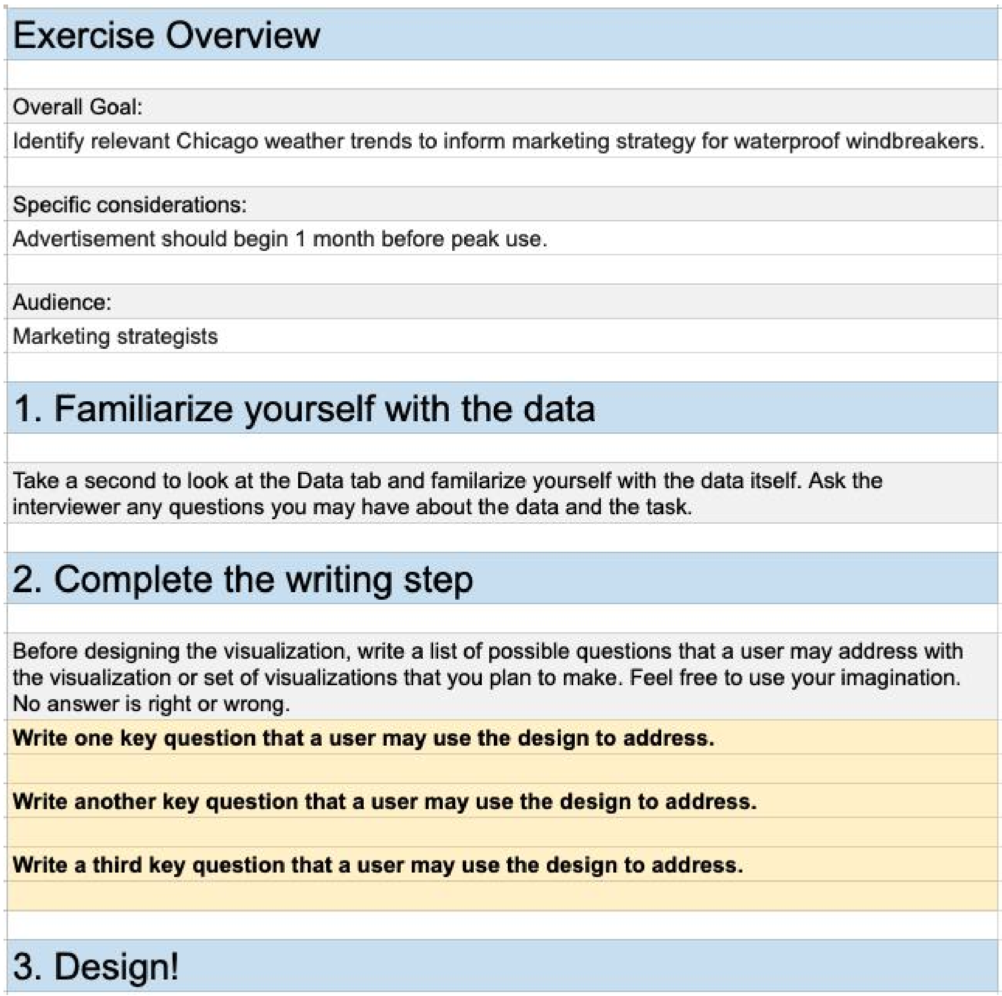
Figure 8.1: Overview of Study 2 design and methodology. Study 1 was the same as the study reported in Chapter 7 and shown in Fig. 7.2.

- **Possible conclusions/takeaways**: Write down ideas for possible conclusions readers might make when viewing the visualization(s).
- **Narrative/story**: Write down a brief story that conveys the main points the visualization(s) might express.
- **Possible titles**: Write down ideas for possible titles for the visualization(s).

These variants were directly based on discussions with the participants in Study 1, allowing us to remain within the scope of real-world experiences and avoid extrapolating to novel cases. As the nature of this study is exploratory, this approach helped maintain relevance and applicability to participants' existing workflows.

The interviewer introduced the variants in a random order, then walked through an example of what each rudder variant might look like in a sample design prompt. The interviewer then asked if the participant had any questions on each variant or how it might be applied in the design process. Participants could ask clarifying questions about how the variants could be applied. They then selected the rudder variant that they found most useful or interesting. Allowing participants to choose supported ecological validity but introduced expected trade-offs, such as limiting comparisons across variants.

Design exercise. Participants then used their selected writing rudder in a short design exercise. Instructions were delivered through a pre-recorded video to ensure consistent presentation. Participants received one year's worth of Chicago weather data (temperature, precipitation, wind speed) and a design prompt: create a visualization (or set of visualizations) to help a marketing agency determine when to begin marketing a waterproof windbreaker. A template, shown in Fig. 8.2, provided the project goal, client considerations, and audience description, mirroring a typical requirements-gathering scenario. If participants had other questions about the client's goals or interests (e.g., is there a particular temperature at which customers begin wearing windbreakers?), they were told to use their best judgment.



Exercise Overview

Overall Goal:

Identify relevant Chicago weather trends to inform marketing strategy for waterproof windbreakers.

Specific considerations:

Advertisement should begin 1 month before peak use.

Audience:

Marketing strategists

1. Familiarize yourself with the data

Take a second to look at the Data tab and familarize yourself with the data itself. Ask the interviewer any questions you may have about the data and the task.

2. Complete the writing step

Before designing the visualization, write a list of possible questions that a user may address with the visualization or set of visualizations that you plan to make. Feel free to use your imagination. No answer is right or wrong.

Write one key question that a user may use the design to address.

Write another key question that a user may use the design to address.

Write a third key question that a user may use the design to address.

3. Design!

Figure 8.2: Template for the design exercise in Study 2. The "writing step" was filled in with the rudder variant selected by the participant. An example is shown here for "User's key questions."

The "writing step" in the template was replaced with the participant's selected rudder variant. Depending on the variant, participants produced either three key questions, three possible conclusions, a 3–4 sentence narrative, or three potential titles.

Based on pilot testing, we chose not to observe the design process directly. Pilot participants reported feeling pressured when watched, especially with a timed task, so interview participants were not asked to share their screen or narrate their workflow. They were reminded that the exercise focused only on getting started, not producing a final design. Participants could use any tools they preferred.

After 10 minutes, the exercise ended, and the interviewer asked participants to reflect on how the writing step shaped their ability to start the design. Participants also considered how the other three rudder variants might have influenced their process. Example prompts for this section included "Describe your experience during the design process of this visualization." and "What, if anything, was different about this exercise compared to the way you typically start designing?"

Post-interview survey. Following the completion of the interview, participants completed a brief (5–10 minute) post-interview survey evaluating the perceived impact of each rudder variant. Using a 5-point scale, they rated how positively (5) or negatively (1) each variant would influence different stages of the design process selected from the Design Activity Framework [145] (Understand, Ideate, and Make).

They also rated the impact of the writing step on "getting started on the design." This additional rating was added to the set of pre-defined stages to provide a more holistic evaluation of the rudders. By including "getting started," we captured insights into the impact of writing rudders on how designers initiate the workflow, which could include a combination or non-linear progression of the stages from the Design Activity Framework.

All ratings were made in comparison to the participant's current design process. Participants also reported if they currently use a similar step (which did not have to be written) and if they would consider using a step like this in the future. At the end of the survey, they reported their overall industry of work and their typical design outcomes, also shown in Tab. 8.1 and Tab. 8.2.

8.2.2 Participants

Participants for both studies were recruited from the Data Visualization Society (DVS), News Nerdery, and public posts to X (Twitter) and LinkedIn. Recruitment materials had an emphasis on recruiting data journalists in addition to more traditional visualization designers to account for the possibility of their unique perspectives on the integration of text and visualization. All participants were actively involved in designing or creating data visualizations as part of their professional roles.

Eligibility criteria were consistent across both studies. Participants were required to be based in the United States, fluent in English, and spend at least part of their work time creating or designing visualizations. These criteria ensured a shared professional context and study-relevant experience.

Study 1 included 24 visualization designers. These were the same participants from Chapter 7 and are described in more detail in Sec. 7.2.1. Study 2 included 15 visualization designers, distinct from those in Study 1. Twelve participants were employed full time, two

Demographic	**Study 1** (n = 24)	**Study 2** (n = 15)
Gender		
Woman	14	13
Man	8	2
Trans or Non-binary	2	0
Years of Experience		
1-3 years	3	4
4-6 years	9	4
7-9 years	7	2
10+ years	5	5
Time Spent Designing (per week)		
Less than 5 hours	2	2
5-10 hours	5	5
11-20 hours	7	2
21-30 hours	5	2
30+ hours	5	4
Company Size		
Micro	3	2
Small	1	1
Medium	7	2
Large	2	2
Enterprise	11	8
Industry Sector		
Broadcasting/Journalism	5	3
Manufacturing	3	0
Medical/healthcare	4	1
Non-profit/Government	3	4
Research	2	5
Scientific or Technical Services	4	1
Software	3	1

Table 8.1: Information about participants' experience and work context. Study 1 participants are the same participants from Chapter 7.

were students, and one was on leave. Thirteen participants identified as women and two as men. Additional detail on both sets of participants can be found in Tab. 8.1 and Tab. 8.2.

Participants are referred to using ID numbers in the format [P#]. To distinguish between studies, Study 1 participants were assigned IDs beginning at 1, and Study 2 participants were assigned IDs beginning at 101.

Design Area	**Study 1** n = 24	**Study 2** n = 15
Design Outcome		
Dashboards	14	6
Live Presentations	10	4
Stand-Alone Charts	10	10
Text and Visual Reports	10	10
Enterprise Applications	4	1
Design Tool		
D3.js	6	2
Figma	9	3
Adobe Illustrator	10	9
Power BI	10	1
PowerPoint	10	4
R	10	3
Tableau	13	6
Excel	18	9
Other	15	8

Table 8.2: Information about participants' design outcomes and tools used to create visualizations. Study 1 participants are the same participants from Chapter 7 (Tab. 7.2).

8.2.3 Analytic Approach

We conducted a multi-stage qualitative analysis across both studies, using a combination of structured coding based on predefined dimensions and open coding to capture emerging themes. Study 1 was conducted primarily to assess current practices, while Study 2 allowed us to examine the impact of writing rudders on designer processes. Although the Study 1 participants and protocols were the same as Chapter 7, the analysis framework used in this chapter provides insights into design processes, rather than challenges with text elements.

All coding procedures were carried out by two independent coders, with a third serving as a tiebreaker when needed. Across both studies, interviews were coded along three primary dimensions: design outcomes produced in participants' typical work, the use of writing within the design process, and the stage of the design process in which writing occurred. This comprised the **Design Analysis** shown in Fig. 7.2 and Fig. 8.1.

The design **outcome** was defined as: "the outcome and context of participant work. This code refers to the specific output of the participant's work (i.e., their deliverable) and/or how it may be applied (i.e., the use context)." Codes included dashboards, stand-alone charts, text and visual reports, enterprise applications, and live presentations. These codes were intended to contextualize design practices rather than serve as a central analytic focus. In Study 2, we replaced this coding process with a direct question in the post-interview survey to minimize any ambiguity.

We also coded for the use of **writing** in the design process. While we knew we would code a dimension for writing prior to conducting the interviews, the precise codes and definitions were developed through the coding process, with repeated discussion between coders to group practices and their frequencies. This code "evaluated the extent to which designers use written language to describe the message/story/key parts of the actual design itself to support their visualization design process. It did not include steps in user research, documentation of design decisions, or the actual creation of the text for the visualization design." The final set of possible codes comprised: No use/none, Incidental, and Deliberate.

The differentiation between incidental and deliberate use of writing was based on the regularity and impact of the written elements on the design process. Deliberate use was identified when participants systematically created and referred to written notes as a main step in their design process. Incidental use, on the other hand, was identified when such notes were created sporadically and had minimal impact on the design's development. It was challenging to make this subjective judgment from interview transcripts; we mitigated this by using detailed coding guidelines, shown in Tab. 8.3, and seeking consensus among coders when assigning these labels.

Finally, we accounted for the **stage** of the design process that the rudders impacted. Stage codes (Understand, Ideate, Make, Deploy) were drawn from the Design Activity Framework (DAF) [145]. Prior to conducting the interviews, we considered different design frameworks. The DAF was chosen for its compatibility with other frameworks [143, 154] and its separate but intersecting stages of design.

The coding process was consistent across both studies. A collaborator and I jointly coded two initial transcripts using a first draft of the codes and their definitions. We then refined the codebook for improved clarity and specificity. Following this, the two coders independently coded the entire set of 24 transcripts. They met to discuss discrepancies and reach consensus on disagreeing codes. For the codes where consensus was not found, the third coder was brought in as a tiebreaker to review the relevant transcripts, without prior knowledge of any previous labels. After considering responses from all three coders, consensus was reached on all coding categories. Interrater reliability (Cohen's κ) for the two original coders was calculated for all codes and can be found in supplemental materials. There was moderate agreement between coders for Study 1 ($Mean\_\kappa = 0.696$) and strong agreement between coders in Study 2 ($Mean\_\kappa = 0.775$).

Study 2 required an additional layer of analysis to capture designers' reflections on the rudder variants. For this, we conducted open coding on all comments related to the rudders, including how participants interpreted them, their perceived benefits or drawbacks, and the contexts in which they might be useful. The first two authors independently generated initial codes, then met to group these into axial codes and, finally, broader themes. This process of open coding allowed us to uncover themes and insights into how these steps influenced the design approach. After individual analysis, both authors met to discuss the emergent

Table 8.3: Definitions of codes used to evaluate the extent to which designers used written language to support their visualization design process.

Use of Writing	Definition of Code
No use/none (0)	The designer does not use written language as a tool to articulate or plan the message, story, or key components of the visualization design. They may use written language as part of user testing, documenting design decisions for collaborators or colleagues, taking notes during conversations/meetings with the client/design team, compiling research on a topic, or creating text elements (e.g., titles) of the visualization, but this is not the same as individually-focused language articulating or planning specific components of the visualization design.
Incidental (1)	The designer sometimes creates and/or uses written language to note thoughts on the message, story, or key parts of the design, but this occurs more incidentally than intentionally. The notes may not be systematically created or referred to throughout the design process, and their impact on the design's development is somewhat minimal.
Deliberate (2)	The designer often or always creates and/or uses written language to note thoughts on the message, story, or key parts of the design. This is an intentional part of the design process, and the designer does this in a systematic way. This written step has a substantial impact on the design's development.

themes across the sets of codes.

Together, these analyses allowed us to characterize both current writing practices in visualization design and designers' perceptions of writing rudders as potential design interventions.

8.3 Practice Analysis: Current Use of Writing

We first addressed our research question, **how do designers *currently* use writing during the design process?** We considered the use of writing on two dimensions: its frequency and the stage of design in which it appears. This analysis combined Study 1 and Study 2 responses, as shown in Fig. 7.2 and Fig. 8.1. Quotes are primarily drawn from Study 1, with Study 2 findings reinforcing and extending these observations. Counts are out of the 39 total participant across both studies.

8.3.1 Writing is Relatively Uncommon

Overall, participants overwhelmingly relied on visual methods when beginning or developing a visualization. About one-third of participants (14/39) reported starting their design process with sketching, and a large majority (27/39) used sketching at some point while designing. For example, P19 said that while on an initial call with a client, they would *"just sketch out the chart... just quickly take my pen and sketch out. Other times, if it's more complex, I'll draw a more complex chart in my notes."* These sketches ranged from quick pencil drawings to digital mockups in tools like Figma. P11 noted that the specific tool changed frequently within their team, *"So we started out with Axure and Illustrator... now we are on Figma. God knows what we will be on next month."*

A slightly smaller number of participants began by loading the data directly into a visualization tool (11/39). This approach was often motivated by efficiency and ease of experimentation. P18 described, *"I'll try to sketch it out or just mock something up... to try to get an idea of what something is gonna look like. Or I'll take Tableau, Power BI and just throw the data in there, see what happens. And then start refining if it's less complicated."* Putting the data directly into the tool facilitated speed and ease in the design process: *"it's just easier to test different chart types that we're looking at"* [P14]. The ability to rapidly switch encodings or variables in software made this approach faster than sketching and well suited for early exploratory data analysis.

For most participants (27/39), writing was either not used at all (14/39) or not used as a distinct part of the design process (13/39). In the latter case, participants would mention taking notes or having written documents, but these were not integral to their design process. Around a third of participants (12/39) used writing in a deliberate way. Study-specific information can be found in Fig. 8.3. Only three participants *started* their design with some form of writing similar to a written rudder.

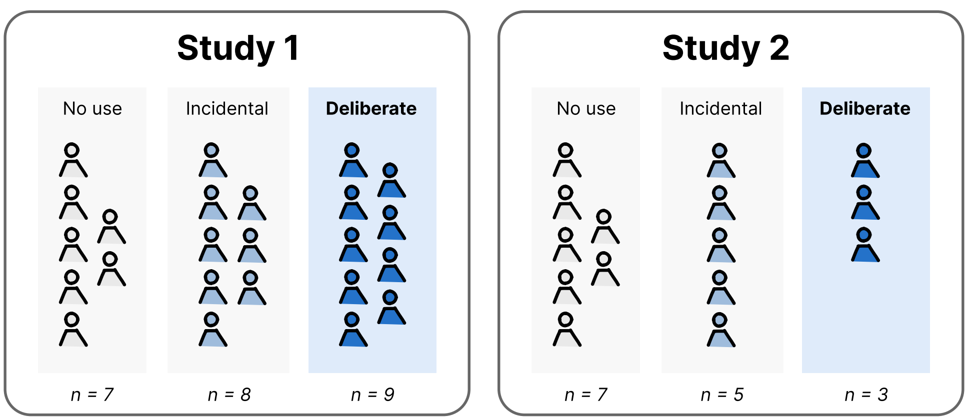


Figure 8.3: Frequencies for different levels of writing in the design process for both Study 1 and 2.

Participants (14/39) who did not incorporate writing *at all* during their design process tended to think of the process as more internal. P1 stated, *"I think it's happening internally.*

I don't list out key takeaways ... I guess it comes up in the process." In some cases, despite explicit probing from the interviewer, writing was never mentioned.

Participants (13/39) who *incidentally* incorporated writing did not consider it an important or consistent part of their design process. For example, P18 mentioned that they *"try to keep notes as I'm going, cause as I'm coming up with a design, or really, as I'm working on it, I'll just have random stuff pop on my head."* In this case, the use of language was sporadic and informal, often for personal memory rather than design guidance.

Only about one-third of participants (12/39) used writing *deliberately* as part of their design process. These designers described writing as an intentional step, often mentioned without interviewer prompting. This could take the form of a formal document, such as a *"written and approved strategic plan"* [P3] or *"paragraphs that are really data heavy"* [P13]. Journalists (6/39) and participants who frequently produced text-and-visual reports (10/39) were more likely to use writing deliberately, though these groups overlapped substantially, making it difficult to attribute differences to role alone.

Overall, the use of writing in the design process was relatively uncommon in our sample, with only about one-third of participants considering it a pivotal or concrete step in their design process. A smaller proportion of Study 2 participants reported deliberate writing (20%) compared to Study 1 (38%), which may reflect that the Study 2 protocol included fewer explicit questions about current writing practices.

8.3.2 Language Used in Early Design Stages

Among participants who used writing in *any* capacity (25/39), most described using it during the early stages of the design process, particularly the **understand** stage. In these cases, these writings set the scene for the content of the visualization and the specific user needs met by the design.

Seventeen participants (17/39) used writing to articulate the core ideas guiding the design. These written notes often served as a structured summary of the problem and the intended message. As P20 explained, *"After I have that initial conversation, I like write it all up. This is the question. This is the context. This is the data we're going to use. This is how we think we're going to communicate it."* Similarly, P17 described drafting a short description of, *"what the graphic is supposed to show, which is usually two sentences."* In other words, the preparation for the design is written out, with key questions and data attributes captured in concrete language prior to beginning the design.

In data journalism contexts, written artifacts frequently originated outside the visualization team. P6 described requesting a story draft from a reporter because, *"I'm less likely to make a mistake if I see the whole story, even if it's just reporter notes."* In cases where journalists were writing their own reports, a similar process took place where the text draft was written first, and data-heavy paragraphs were replaced with preliminary charts.

Writing also appeared during the **ideate** stage (10/39), in which participants were brainstorming different ways to address the needs of the design. For example, after finalizing the goals and intents of the design, P23 has *"a whole notes document going of things that just* *occur to me."* This writing served less as formal structure and more as a generative space for exploring options.

Together, these findings reveal that when designers used writing, they employed it primarily to frame the problem and shape initial ideas rather than to guide later stages of development. We used this information to construct the context for the design exercise used in Study 2.

8.4 Rudders Analysis: Impact of Writing Rudders

Having established that designers currently use writing in fairly limited and early-stage ways, we now turn to our second research question: **What is the perceived impact of writing rudders on the design process?** This analysis focuses specifically on responses from Study 2.

Study 2 introduced designers to four structured rudder variants and asked them to incorporate one into a brief design exercise. This approach allowed us to evaluate not only whether designers saw value in these interventions, but also which type of intervention would be most helpful. Participants' reflections, combined with post-interview ratings of each variant, provide insight into the potential for writing rudders in the design process. Representative comments for each participant can be found in Tab. 8.4. Overall impressions for each rudder can be found in Fig. 8.4.

8.4.1 Written Rudders Add Design Focus

Participants consistently noted that writing rudders added structure and focus to the early stages of the design process. These reactions were most strongly associated with the key questions (9/15) and possible conclusions (9/15) variants, which participants viewed as helping them clarify what the visualization should accomplish (see Fig. 8.5). In contrast, The narratives (3/15) and titles (3/15) were not often seen as providing additional useful direction to the design process.

Relative to how participants typically began their workflows, writing rudders allowed for a more guided process. Participants felt they had a greater degree of focus in the initial stages of the design process (12/15). For example, P108 said that writing out possible conclusions helped to narrow down, *"which of the metrics would be most important to someone."* Written rudders also acted as guardrails to the design process, protecting against, *"getting too excited and diving into the data, potentially losing focus of what the purpose is"* [P103].

Participants also highlighted the way rudders, particularly questions and conclusions, helped them foreground the audience's perspective (8/15). P107 summarized this advantage, *"I like to make sure that I am doing what the audience wants... The data isn't valuable unless you're giving it to the right people in the right format."* The narrative and title variants did not prompt this same user-centered framing and were generally viewed as less effective for aligning design decisions with audience needs.

Table 8.4: Participant responses and quotes, grouped according to the chosen rudder. "Process Start" refers to how the participant usually began their design process.

ID	Industry	Process Start	Representative Quotes from Participants about the Selected Rudder
			Questions
101	Research	Raw data	"You need an objective and a plan, and you need to make sure those questions are open enough."
103	Public Sector	Sketch	"The establishing [of] the questions beforehand makes you sit down and just focus on the client first."
104	Public Sector	Sketch	"Going back, saying, this is my goal... Can people answer this question?... I think that's super helpful."
106	Healthcare	Raw data	"It's a non- event. It's just part of [design]... It's just pretty fundamental."
107	Software	Tool	"Building something that the user wants is the main goal. I think that doing [questions] is the most effective way."
113	Research	Sketch	"I think it's a good way to kind of organize things, cause I feel like a lot of times, it's just kind of in my head."
114	Public Sector	Sketch	"I like this approach and that it does require me to begin more with those [questions]."
115	Journalism	Raw data	"I think it was a good framing to have in mind. But it definitely changed a lot as I like explored the data more."
			Conclusions
102	Journalism	Sketch	"Coming up with what you want people to get out of this data set... helps me figure out what I'm gonna be visualizing"
108	Journalism	Tool	"Having to write out kind of the actual conclusion that someone would see forced me to really be strategic."
110	Research	Tool	"I was surprised at how much it guided me in the process. Hadn't really occurred to me to to do it like that before."
111	Technical	Writing	"You want to make sure that you haven't gone down a rabbit hole too far, and you're straying from the main point."
112	Research	Tool	"I would just have it as [a] starting point, because the takeaway can change."
			Narrative
105	Research	Sketch	"It's a good idea to try to put it into words... It helped to figure out what the point is."
109	Public Sector	Tool	"It really focused me. I used what I wrote to immediately start thinking about what graph type I was going to use."

Though hypothetical, future use of these writing steps was viewed positively. Nearly all participants (14/15) said they would consider using key questions or possible conclusions in at least one area of their own workflows. These variants were viewed as lightweight ways to sharpen intent, set priorities, and establish direction before moving into visual

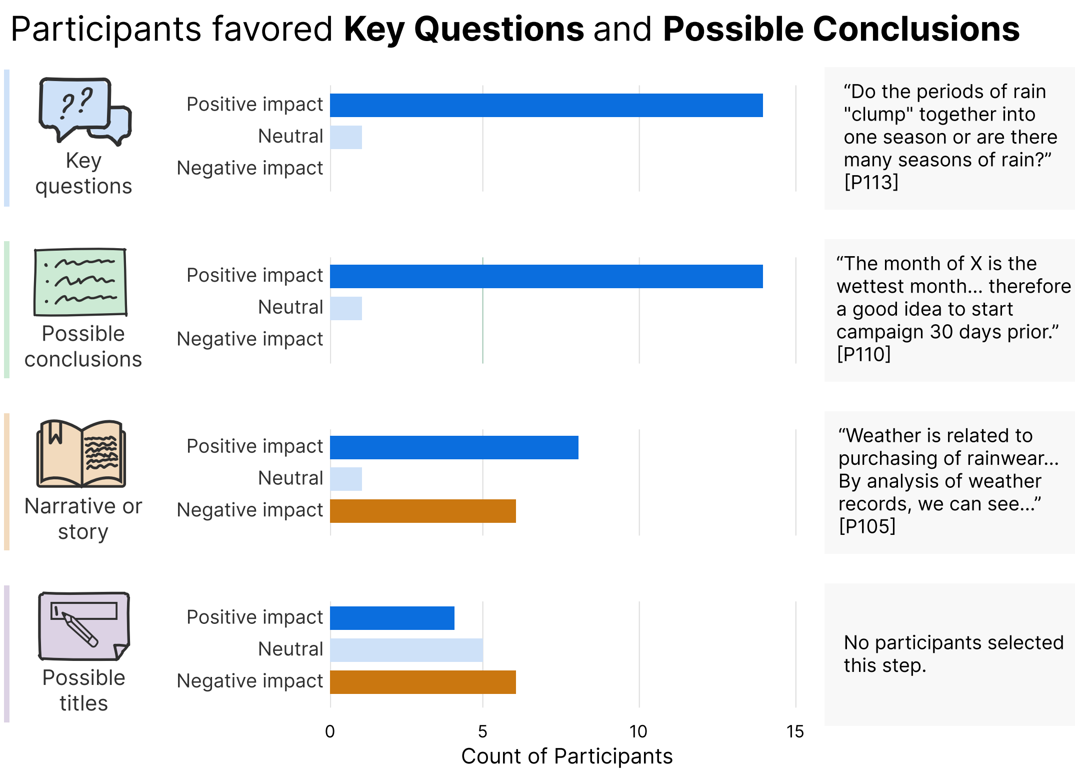


Figure 8.4: Four types of writing rudders tested in Study 2, participants ratings of each type, and examples of participant-written rudders.

exploration. These steps, primarily the **questions**, helped designers articulate the purpose of the visualization and remain anchored to audience needs. Additional quotes can be found in Tab. 8.4.

8.4.2 Using Rudders for Evaluation or Instruction

Participants identified potential uses for written rudders later in the design process (12/15), even though the design exercise itself focused only on the initial stages. One consistent idea was that written rudders (key questions (9/15) or conclusions (3/15)) could serve as evaluation artifacts. Because rudders capture the intended goals or message at the start of a project, participants (11/15) felt they could act as a reference point when assessing whether the final visualization achieved its aims. As P102 explained, rudders could help designers ask, *"Are people actually coming away with what I wanted them to come away with?"*

Shown in Fig. 8.5, the questions rudder was most frequently associated with possible evaluative use. However, a narrative (3/15) could also provide design justification for the client (*"Definitely when you're communicating with a team or with the client... it would be* *nice to have a narrative"* [P105]) or simply to provide a point of engagement with the client about the visualization's goals.

Participants also noted that written rudders could play a role in instructional or educational contexts. Some described the question-writing step as especially helpful for novices; P110 described it as "t*he more intuitive place to start for, particularly a beginner,"* even if unnecessary for their own practice. Others reflected that the rudders themselves felt pedagogical in a way that was unexpectedly beneficial. After completing the design exercise, P113 stated, *"I'm always a little bit skeptical of this sort of thing because I'm like, This feels like school. But I actually really liked it."* These reactions suggest that structured writing may be particularly valuable for early-career designers or students who are still developing their approaches to visualization design.

The comments made by participants for this theme were not based in their direct experience using the rudders for evaluation or instruction; the design exercise was completed solo and only focused on the initial stages of the design process. However, reflecting on the rudder exercise and drawing from prior design experiences, participants indicated that these features could have use cases *beyond* those tested in this study.

8.4.3 Narrative Rudders Suited for Later Stages

Although the narrative and title rudder variants were the least preferred overall (as seen in Fig. 8.4), many participants (9/15) felt these approaches could be more appropriate for later stages of the design process. Participants frequently described narratives as difficult to write before the visual form of the data was known. P115 compared the narrative rudder to, ' *"alt text for graphic,"* stating, *"I can't imagine writing it before [creating the design]."* After constructing a set of possible designs, P110 would use a narrative step: *"I'll write text for each [visualization option]... I'll look at the text by itself to see which one reads better."*

Titles followed a similar pattern. Seven participants viewed writing a title before exploring the data as premature, noting that titles often change once the visual design and key message become clearer. Because the participants typically finalized titles near the end of the design process, completing this step too early felt counterintuitive.

Shifting the more directive rudders (conclusions, narratives, and possible titles) to later in the design process may have additional benefits for data exploration. While written rudders provided a clear direction to the design process, this direction could also introduce possible bias to data exploration (10/15). The importance of neutral data exploration varied across contexts. P106 (researcher) described, *"For me, quality data vis means you don't know the answers. You're exploring the data."* However, P102 (journalist) emphasized that, *"if there's no lead, there's no story... It's a very easy test that we really have to be able to answer,"* making early hypotheses or takeaways a useful guide for assessing newsworthiness.

To balance focus with openness, some participants proposed workarounds. One strategy was to use placeholder language or a "fill in the blank" approach, rather than referring to specifics of the data. An example of this approach is shown for the conclusions rudder in Fig. 8.5, using an "X" instead of a specific month. Keeping an open mind was important

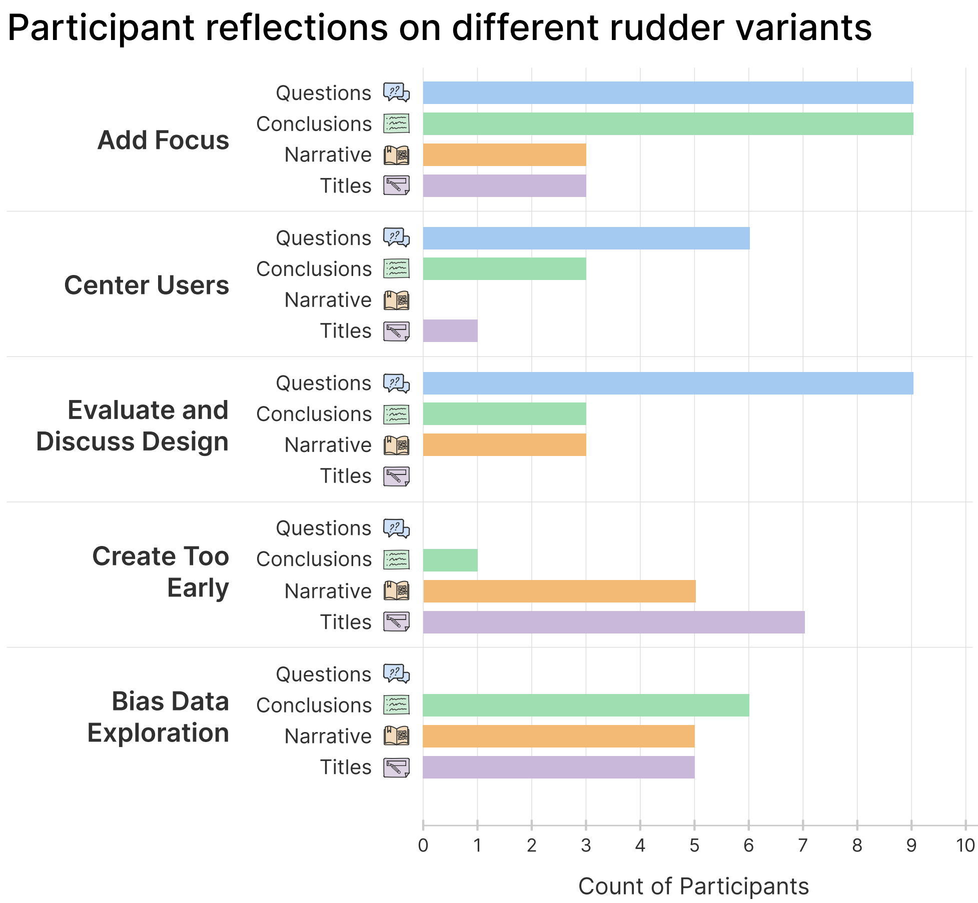


Figure 8.5: Participant reflections on the impact of different rudder variants. Counts shown here represent how many participants ($n = 15$) mentioned the topic for the specific variant.

when using these rudders. As P108 noted, it is helpful to imagine what a reader might think, *"but, on the other hand, I think you also have to be open to the idea of the data not saying what you might want it to say."*

8.5 Summary

In exploring the role of writing in visualization design, these two interview studies offer a nuanced understanding of how written rudders can potentially enrich current practices. Study 1 showed that writing is rarely a formal part of visualization workflows, and when it does appear, it typically helps designers clarify goals or organize early ideas before sketching or exploring data in tools. Study 2 demonstrated that structured writing rudders, especially key questions and possible conclusions, can support these early stages by helping designers articulate intent, maintain focus, and keep audience needs in view. Participants also saw potential value for using rudders later in the process, such as evaluating whether a visualization achieved its goals or facilitating communication with clients and collaborators.

The broader value of writing down goals and intentions is well documented [70, 99, 165, 183]. This principle extends naturally to visualization design: writing helped designers slow down, articulate what they hope to achieve, and engage more deeply with the problem at hand. Written rudders served to externalize a designer's intent in a concrete, discussable form, offering a fast, low-effort complement to existing user-centered practices [157, 231]. Compared to more involved methods like cognitive walkthroughs, rudders required minimal time investment while still prompting reflection on users, goals, and possible interpretations. They also provided a narrative focus, building on work in narrative visualization [90, 189] by making space for early consideration of story and meaning.

At the same time, writing rudders were not universally applicable, and participants raised important considerations for their use. More prescriptive variants, such as narratives or titles, can anchor designers too early, particularly when the task requires open-ended exploration. Using placeholders, focusing on open questions, or adopting hypothesis-testing strategies may help preserve exploratory flexibility [108, 136]. The usefulness of rudders may also vary by designer experience, team structure, and timing within the design process. Early-career designers, students, or teams needing more explicit design alignment may benefit more from these steps than individuals working within established personal workflows.

Overall, the findings support the value of incorporating short written guidance into visualization design. Rudders can act as a scaffold for early design decisions, a shared artifact for aligning collaborators, and a reference for evaluating whether a final visualization communicates its intended message. Notably, all participants in Study 2 expressed willingness to adopt some form of this practice, even though only a small subset (3/15) previously used writing as part of their typical workflow. By explicitly combining language and visualization, rudders help bridge the gap between the data and its interpretation, facilitating a deeper connection with the audience and a better understanding of the design process goals.

More broadly, this work contributes to growing evidence that language plays a central role in both interpreting and constructing visualizations. Visualization is often treated as a primarily visual medium, yet these studies presented in this dissertation emphasize that text, whether written by designers or interpreted by readers, shapes how data is understood, framed, and communicated.. Recognizing visualizations as inherently multimodal underscores the need for design practices, tools, and research frameworks that explicitly account for the interplay between text and visuals in data communication.

Part IV

Conclusions

Chapter 9

Key Findings

Throughout this dissertation, I show that text plays a central and often under-recognized role in how people perceive, interpret, and create data visualizations.

Readers consistently favored visualizations that included richer text content, ranking annotated charts above minimalist designs. This text content also shaped interpretation. Titles and annotations guided what readers noticed and repeated in their takeaways, with statistical and contextual statements exerting the strongest influence. These same elements could shift perceptions of bias as well, leading readers to view the designer as supporting one particular data interpretation over others.

At the same time, text did not uniformly affect all forms of reasoning. We had assumed that, because text affected conclusions, it would have a similar influence on predictions of future data trends. However, the effect was minimal and inconsistent, suggesting that task and context are important for how readers use written or visual information. This contrast highlights the importance of examining text influence across multiple cognitive tasks rather than assuming its effects generalize.

Turning to designers, we identified ten functions of text that appeared across real-world visualizations and revealed four recurring design patterns that practitioners use when communicating data. These functions provide a theoretical foundation for analyzing text across a wide variety of text components, including structural elements that are often overlooked, such as axes. In doing so, we uncovered unique functionality of text and interactions across text elements and visual design choices.

Interviews with designers further illustrated persistent challenges in incorporating text, from justified concerns about clutter and bias to workflow and tooling limitations that complicated text formatting and content updates. Writing also emerged as a potentially valuable but underused design aid: short written rudders helped designers clarify goals and make more intentional user-centered design decisions.

Taken together, these findings emphasize that we cannot understand data comprehension without understanding the impact and use of text in visualization design. They also carry distinct implications for different audiences: for visualization researchers, for designers in practice, and for the many readers who engage with charts as part of their everyday lives.

9.1 For Visualization Researchers

The insights from this dissertation are most directly relevant to other visualization researchers. While only a small subset of the community focuses explicitly on text, the findings here apply broadly to work on visual design, narrative visualization, explainable AI, uncertainty communication, and other areas that rely on charts to convey information. The set of recommendations below highlights key considerations for researchers studying how people interpret or interact with visualizations.

Treat text as part of the experimental manipulation, not as the background. Much empirical work focuses on the variation of visual encodings or design choices. Even in these cases, researchers should treat text as an important part of stimulus design. This includes documenting how text is written or generated and considering whether the text may introduce confounds attributed to visual design choices. If this work were to culminate in one recommendation, it would be this: text should not be treated as a disposable or secondary element of visualization design.

Consider text-only conditions as first-class baselines. Visualization research often only tests variations of *visual* displays. However, this relies on the assumption that the best way to communicate data is visual, rather than through some other medium. By using text-only representations as a baseline condition along with a visualization control, researchers can test this assumption. Doing so not only strengthens empirical claims but also encourages more critical examination of how text contributes to or substitutes for visual communication. These conditions may also be preferred by a subset of readers (see Chapter 3).

Use text function analysis when designing controlled stimuli. When constructing experimental manipulations, the functions of text (e.g., IDENTIFY VALUES) offer a systematic way to create comparable conditions that differ in meaningful ways. These functions can be used alongside broader frameworks (e.g., four-level semantic model [134]) to design text systematically and analyze its effects with more granularity. The text functions outlined in Chapter 6 provide researchers with a variety of ways to consider text design in their stimuli *and* their paper figures.

Build study materials that reflect real-world practices. Researchers should also consider the four design patterns identified in Chapter 6 when creating stimuli. These patterns reflect common ways that text appears in practice and can help ensure greater ecological validity. Using naturalistic combinations of different text elements and functions can make study findings more applicable to real-world visualization settings.

Use multiple types of measures when examining effects of text. In Chapter 4, we observed that participants reported low reliance on text information, even when their takeaways were clearly influenced by the text. This divergence illustrates an important methodological point found in other work [206]: the measurement method shapes the conclusions researchers can draw. Incorporating multiple measures and tasks, such as behavioral responses, scalar ratings, recall measures, or decision tasks, would provide a more complete picture of how text influences interpretation.

Design visualization platforms that allow for easy text editing. Many existing visualization platforms provide rich control over visual encodings but make it difficult to adjust text beyond basic labels. Enhancing these tools with more flexible formatting, reliable handling of dynamic text, and built-in support for producing alternative text would directly address several hurdles designers encounter. These improvements would also give other researchers greater capability to test how different text functions and design patterns influence interpretation in empirical work.

Together, these recommendations encourage researchers to take text seriously as a central component of visualization. Text is not merely decoration around a chart; it is an active part of how viewers process, interpret, and recall information. By designing studies that account for text more systematically, researchers can deepen empirical understandings of how visualizations communicate meaning.

9.2 For Visualization Designers

These findings also carry important implications for visualization designers. Text is often treated as the secondary layer added after a chart is complete, but the studies in this dissertation show that text design choices influence how readers interpret data, what they remember, and how they perceive the designer's intent. Considering text early and deliberately can improve the clarity, inclusiveness, and impact of a design.

Treat text as a core design element, not an afterthought. Similar to the recommendations for researchers, designers should actively consider how they use text throughout the design process. Text shapes how readers understand the data and perceive the message behind the visualization. While this makes text a powerful communicative tool, it can also act as a potential source of unintentional bias. Designers should evaluate not only whether titles or annotations are clear but also whether they may inadvertently overemphasize certain interpretations or guide readers too strongly.

Consider the needs and preferences of different users. Readers vary in how much they prefer and rely on text. On average, people favor text-rich displays, but these preferences differ based on individual tendencies toward visual or text information. Designing with these differences in mind could look like including adjustable text density, details on demand, or optional text explanations. These design approaches can make visualizations more flexible and more inclusive. Our findings also indicate that some readers prefer text-only descriptions over visual displays. Designers should consider when a paragraph might serve the audience better than a chart.

Use writing techniques early in the design process. Short written rudders can help designers clarify what the visualization should communicate before beginning the design process. Drafting a set of user questions, possible takeaways, or a brief narrative can focus the design and improve alignment with collaborator or client expectations. Integrating this step with accessibility considerations (e.g., drafting alternative text early) can also make it easier to produce high-quality descriptions rather than leaving them as an afterthought. Writing alt text during the design encourages more thoughtful choices about what data information is communicated or emphasized.

Use placement intentionally: annotations do different work than titles. Positioning text close to the relevant data feature strengthens the link between the annotation and the visual evidence, while broader statements that comment on more global visual features are better suited for titles. Designers should consider content and placement together rather than treating them as independent decisions; the communicative role of the text should guide where it appears in the visualization.

Taken together, this set of recommendations emphasizes how designers should think about using text elements intentionally. In addition to these comments, designers can use the functions detailed in Chapter 6 to think more specifically about how each of their text elements supports or communicates the key data insights. Effective visualization design depends on thoughtful coordination of text and visuals.

9.3 For Readers of Data Visualizations

Many people encounter visualizations as part of their work, studies, or everyday information consumption – not as visualization specialists. The findings in this dissertation are also relevant for anyone who reads charts to understand evidence, evaluate claims, or make decisions. Below are several practical habits that can improve how readers interpret data visualizations.

Notice how text frames what is seen. Titles, captions, and annotations guide attention to particular aspects of the data, and they can subtly shape what reader conclusions from a chart. This does not mean that text is necessarily designed to manipulate or mislead; in many cases, designers know their data well and highlight details that matter. Still, recognizing how text highlights certain parts of the data can help readers approach visualizations with more intention and awareness.

Compare the text with the visual evidence. When reading charts, particularly those from unknown sources, check how the text and the visual information aligns to tell the story. If the title emphasizes a "dramatic increase," look to see whether the chart supports that claim. If a caption highlights a specific point or trend, locate it in the visual. Looking for agreement or tension between text and visuals can reveal where the data may need more context, and understanding the designers' choices can help construct a more comprehensive view of the data information.

Pay attention to what's *not* said in the text. This follows with the previous recommendation, representing the inverse of evaluating text information. Text is used to highlight certain aspects of the data, but it can also leave things out. If a chart's title describes an overall trend but says nothing about other salient data features, that absence may matter for the interpretation. Simply asking "What else might be important that isn't mentioned here?" can provide a reader with a more holistic interpretation of data in news stories, dashboards, and reports.

Consider your own preferences for visual or text information. Use both text and visuals in ways that align with individual preferences. People differ in how they interact with information. Some prefer concrete text explanations and have no problem reading a set of paragraphs describing data, while others would prefer to see information presented visually. Recognizing one's own preferences for different kinds of information can shape how a reader approaches and engages with data.

These habits can help readers engage more thoughtfully with visualized data. Text influences how people interpret charts, but it works best when considered alongside the visual evidence and with awareness of one's own information preferences. Overall, understanding text as a meaningful part of the visualization can support clearer, more confident understanding of data in everyday contexts.

Chapter 10

Future Directions

The work presented in this dissertation has made significant strides in research at the intersection of text and visualization design. I provide a foundation for understanding how text influences visualization design, interpretation, and the design process. At the same time, the designs and limitations of these studies reveal several important areas for continued investigation. Future work can expand the scope of design settings, empirically evaluate the frameworks introduced here, develop tools and practices that support designers in real-world contexts, and examine further nuance of individual interactions with visualizations.

10.1 Broader Visualization Contexts: Chart Types, Settings, and Interactivity

The work in this dissertation relied on scoping choices that made controlled experimentation possible but necessarily limited the real-world generalization of the findings. We only examined English text and often recruited participants who were native English speakers. Prior work suggests that bilingual participants' preferences and chart comprehension can differ depending on whether the information is presented in English or their native language [9]. Exploring other languages is a key area of future research for visualization research as a field but is particularly relevant for text and visualization studies [177].

Additionally, expanding these studies to additional chart types, such as bar charts, scatterplots, multivariate line charts, and real-world graphics from data journalism, would provide further nuance to the results. Uncertainty visualizations, such as quantile dot plots [59, 100, 161] or forecast visualizations [159, 160, 162], also present a promising direction. Both designers and readers tend to struggle with communicating and understanding uncertainty [63, 88] and so may benefit from additional explanatory text.

User interactions or understandings of text may also differ depending on the situation they are in. For example, we did not provide a specific context for participants to base their preference rankings off of in Chapter 3. If we had instead provided some specific context of use (e.g., sharing this data with a coworker, adding this data to a report, skimming this data in a hurry), the results would likely have differed across these contexts. Interpretations or interactions with data would also look different when viewing visualizations on a mobile (i.e., smartphone) screen than on a desktop [21, 85]. Examining these different contexts and use cases is critical for further investigations into text and visualization.

In addition to text, there are other modes of communication (e.g., speech, haptic feedback) that could communicate data information. Some of my prior work in this area suggests that speech, text, and visualizations all have different trade-offs for communicating uncertainty information for decision making [207, 208, 209]. Other research has begun to explore even more out-of-the-box data communication mediums, including textile data artifacts [173]. Future work in text and visualization should compare to or incorporate other modes of information.

While there has been work in text and visualization for dashboard designs [214], the studies presented in this dissertation are limited to individual, static visualizations. Dashboards provide a series of other types of text information, such as filters or tooltips, and they almost always present multiple visualizations at once. Future work should also examine the impact of this text in interactive environments, where tooltips and hover interactions introduce new forms of engagement with text information. Additionally, domain-specific analyses (e.g., social media, news articles) also provide rich opportunities for studying how text use and influence varies across contexts.

10.2 Further Evaluations: Text Design, Text Functions, and Writing Rudders

Our priority when creating stimuli was to keep the chart's text in line with best practices for visualization design. As a result, we did not investigate a variety of visual design choices with text. For example, exploring the use of color and typeface would allow for a different set of design recommendations. Similar work exists for document readability [38, 234, 235], but the readability for text *within* visualization designs is still an open and underexplored area of investigation.

In Chapter 6, we introduce a framework for understanding text functions in visualization, but additional research is still needed to assess its usability and effectiveness. Studies evaluating how the different factors (i.e., text design patterns) influence data comprehension, trust, and engagement would help translate the framework into further actionable guidance for designers. Additionally, more empirical studies on the impact of different text functions could help to construct a model of text influence on reader interpretations.

Similarly, the work on writing rudders provides an initial examination of how writing shapes early design stages. Future research should include controlled studies comparing rudders to other design interventions and assessing design outcomes (e.g., quality, creativity, workflow efficiency). Evaluating how rudders influence final design products would help build the evidence base informing written interventions. Longitudinal studies that examine the use and impact of these rudders over time could also clarify how rudders fit into real-world workflows and identify organizational or individual barriers to adoption.

10.3 Tool Development, AI, and Accessibility

Several tool-development directions can support more effective use of text in visualization design, many of which are detailed in Chapter 7. Designers often face repetitive tasks when formatting text, and features such as batch editing, responsive text layout for mobile environments, and context-aware text adjustments are underdeveloped in mainstream visualization tools. Additional work could explore tools that convert visualizations across presentation contexts or support collaborative authoring of text, especially given that many visualization teams work in distributed workflows.

Design tools could also support the use of writing in the design process, as explored in Chapter 8. By providing a 'scratchpad' within the design tool, designers could be better supported to use writing as part of their design process. InkSight [129] and Epigraphics [261] represent two recent initiatives that forward the use of writing in design, and future work can continue developing features and tools that support this direction.

AI-supported interactions with data also provide new areas of exploration. Question-answer (Q/A) benchmarks for querying data visualizations [62, 97, 137] not only allow us to understand the capabilities and limits of LLMs for visualization comprehension, but they also emphasize a new kind of user experience to examine. Comparing Q/A systems with interactive or static annotations can help identify the strengths and weaknesses of each approach. This area of research would inform the broader field of generative AI and interactive visualization, providing evidence-based guidelines for integrating AI-driven Q/A systems into data presentations.

AI-based approaches could also assist in generating text that identifies key data points and highlighting them for users. While platforms like Power BI offers basic AI-driven summaries, these typically require significant manual refinement and are not positioned on the charts themselves. More advanced applications, such as models trained on organization-specific data, could provide accurate, context-sensitive text to support dynamic text.

Accessibility remains a critical area of future work. Integrating writing earlier in the design process may streamline the creation of alt text. Developing standardized guidelines for visualization-specific alternative text could promote more inclusive design practices. Alternative text descriptions should also provide any written text visible in the chart, in order to provide parity across data experiences. Research has already begun to develop critical initiates and toolkits along these lines [55].

10.4 Individual Differences and Response Measures

Finally, expanding research to broader participant populations will improve the external validity of findings. The studies in this dissertation primarily reflect responses from young, highly educated adults, limiting generalizability to other groups. Differences in age, education, graphical literacy, and domain familiarity may influence how readers respond to text-rich visualizations. Targeting specific populations, such as older adults, children, people with a lower education level, or individuals with lower data literacy, would provide a more complete understanding of how text shapes interpretation across diverse contexts.

In this work, our experimental stimuli were designed to be relatively neutral; the data was simulated, and we introduced no real external or political information to the visualizations. However, participants still considered their own experiences when interpreting even simulated data. For example, we used a line chart depicting a fictional stock index in Chapter 4. From this chart, one participant concluded, *"Something happened to some stock index in 1962 to make to decline to nothing, when the stock market has increased over the same period."* Their own knowledge of the stock market affected their conclusion from the generated data.

Considering the impact of real-world beliefs, experiences, and expertise is an important next step in visualization research. Some of my related prior work has examined the influence of prior beliefs on correlation estimations from scatterplots [249]; having a strong belief in a relationship between two variables can increase estimations of correlation (and vice versa) by about 19% on average. The influence of belief or experience has not yet been examined in the context of text and visualization, but it represents an important area of consideration.

Finally, variations observed across binary choice questions and likelihood ratings suggest that question format can meaningfully shape participant responses. Understanding how different elicitation formats influence judgments would help researchers design more robust measures of data interpretation. Combining multiple measures of an outcome allows a researcher to draw the most informed conclusion regarding the impact of experimental conditions. Future work could also simulate real-world tasks involving comparison, estimation, or forecasting to better capture the role of text under realistic cognitive demands.

10.5 Closing Remarks

Taken together, the studies in this dissertation demonstrate that text is a central component of how people make sense of data. The findings highlight both the promise of well-designed text elements and the challenges that accompany their use, offering evidence that can guide researchers, practitioners, and the general public alike. As information visualization continues to evolve in increasingly interactive, multimodal, and AI-supported environments, the study of text will remain essential for designing communication that is clear, trustworthy, and engaging. The future directions outlined for this space provide a path forward for deepening this inquiry and extending the insights of this dissertation into the future of visualization research and practice.

Bibliography

[1] Kiran Ajani, Elsie Lee, Cindy Xiong, Cole Nussbaumer Knaflic, William Kemper, and Steven Franconeri. "Declutter and focus: Empirically evaluating design guidelines for effective data communication". In: *IEEE Transactions on Visualization and Computer Graphics* 28.10 (2022), pp. 3351–3364. DOI: 10.1109/TVCG.2021.3068337 (cit. on pp. 11, 12, 14, 72, 87, 98, 99).

[2] Jennifer Allen, Cameron Martel, and David G Rand. "Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program". In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2022. DOI: 10.1145/3491102.3502040 (cit. on p. 53).

[3] Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Jin, and Marti A Hearst. "Futzing and moseying: Interviews with professional data analysts on exploration practices". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 22–31. DOI: 10.1109/TVCG.2018.2865040 (cit. on p. 16).

[4] Teresa M Amabile. "Social psychology of creativity: A consensual assessment technique". In: *Journal of Personality and Social Psychology* 43.5 (Nov. 1982), pp. 997–1013. DOI: 10.1037/0022-3514.43.5.997 (cit. on p. 55).

[5] Robert Amar, James Eagan, and John Stasko. "Low-level components of analytic activity in information visualization". In: *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*. IEEE Computer Society, 2005, pp. 111–117. DOI: 10.1109/INFOVIS.2005.24 (cit. on pp. 5, 10, 12, 77, 82).

[6] Robert Amar and John Stasko. "A knowledge task-based framework for design and evaluation of information visualizations". In: *Proceedings of the IEEE Symposium on Information Visualization*. IEEE Computer Society, 2004, pp. 143–150. DOI: 10.1109/INFVIS.2004.10 (cit. on pp. 12, 77).

[7] Anjana Arunkumar, Lace Padilla, Gi-Yeul Bae, and Chris Bryan. "Image or information? Examining the nature and impact of visualization perceptual classification". In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (2023), pp. 1030–1040. DOI: 10.1109/TVCG.2023.3326919 (cit. on pp. 12, 73, 77).

[8] Anjana Arunkumar, Lace Padilla, and Chris Bryan. "Modeling and measuring the chart communication recall process". In: *Computer Graphics Forum*. Wiley Online Library. 2025, e70099. DOI: 10.1111/cgf.70099 (cit. on pp. 2, 19, 41).

[9] Anjana Arunkumar, Lace M. Padilla, and Chris Bryan. "Lost in translation: How does bilingualism shape reader preferences for annotated charts?" In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2025. DOI: 10.1145/3706598.3713380 (cit. on pp. 19, 141).

[10] Benjamin Bach, Euan Freeman, Alfie Abdul-Rahman, Cagatay Turkay, Saiful Khan, Yulei Fan, and Min Chen. "Dashboard design patterns". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (2023), pp. 342–352. DOI: 10.1109/TVCG.2022.3209448 (cit. on pp. 13, 89).

[11] Benjamin Bach, Mandy Keck, Fateme Rajabiyazdi, Tatiana Losev, Isabel Meirelles, Jason Dykes, Robert S. Laramee, Mashael AlKadi, Christina Stoiber, Samuel Huron, Charles Perin, Luiz Morais, Wolfgang Aigner, Doris Kosminsky, Magdalena Boucher, Søren Knudsen, Areti Manataki, Jan Aerts, Uta Hinrichs, Jonathan C. Roberts, and Sheelagh Carpendale. "Challenges and opportunities in data visualization education: A call to action". In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (2024), pp. 649–660. DOI: 10.1109/TVCG.2023.3327378 (cit. on p. 15).

[12] John Baer, James C. Kaufman, and Claudia A. Gentile. "Extension of the consensual assessment technique to nonparallel creative products". In: *Creativity Research Journal* 16.1 (2004), pp. 113–117. DOI: 10.1207/s15326934crj1601_11 (cit. on p. 55).

[13] Hannah K. Bako, Xinyi Liu, Leilani Battle, and Zhicheng Liu. "Understanding how designers find and use data visualization examples". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (2023), pp. 1048–1058. DOI: 10.1109/TVCG.2022.3209490 (cit. on p. 16).

[14] Melanie Bancilhon, Amanda Wright, Sunwoo Ha, R. Jordan Crouser, and Alvitta Ottley. "Why combining text and visualization could improve bayesian reasoning: A cognitive load perspective". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2023. DOI: 10.1145/3544548.3581218 (cit. on pp. 11, 12).

[15] Deni Basaraba, Paul Yovanoff, Julie Alonzo, and Gerald Tindal. "Examining the structure of reading comprehension: Do literal, inferential, and evaluative comprehension truly exist?" In: *Reading and writing* 26.3 (2013), pp. 349–379. DOI: 10.1007/s11145-012-9372-9 (cit. on p. 77).

[16] Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. "Useful junk? The effects of visual embellishment on comprehension and memorability of charts". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2010, pp. 2573–2582. DOI: 10.1145/1753326.1753716 (cit. on p. 106).

[17] Scott Berinato. *Good charts: The HBR guide to making smarter, more persuasive data visualizations*. Harvard Business Review Press, 2016. ISBN: 1633690709 (cit. on p. 20).

[18] Jacques Bertin. *Sémiologie graphique. Les diagrammes, les réseaux, les cartes*. Gauthier-Villars, 1967. ISBN: 3111282171 (cit. on p. 2).

[19] Enrico Bertini, Michael Correll, and Steven Franconeri. "Why shouldn't all charts be scatter plots? beyond precision-driven visualizations". In: *2020 IEEE Visualization Conference (VIS)*. 2020, pp. 206–210. DOI: 10.1109/VIS47514.2020.00048 (cit. on p. 10).

[20] Alex Bigelow, Steven Drucker, Danyel Fisher, and Miriah Meyer. "Reflections on how designers design with data". In: *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. Association for Computing Machinery, 2014, pp. 17–24. DOI: 10.1145/2598153.2598175 (cit. on p. 16).

[21] Kerstin Blumenstein, Christina Niederer, Markus Wagner, Grischa Schmiedl, Alexander Rind, and Wolfgang Aigner. "Evaluating information visualization on mobile devices: Gaps and challenges in the empirical evaluation design space". In: *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. Association for Computing Machinery, 2016, pp. 125–132. DOI: 10.1145/2993901.2993906 (cit. on p. 142).

[22] Tal Boger and Steven Franconeri. "Reading a graph is like reading a paragraph". In: *Journal of Experimental Psychology: General* 153.7 (2024), pp. 1699–1704. DOI: 10.1037/xge0001604 (cit. on p. 40).

[23] Boltužić, Filip and Šnajder, Jan. "Toward stance classification based on claim microstructures". In: *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Association for Computational Linguistics, 2017, pp. 74–80. DOI: 10.18653/v1/W17-5210 (cit. on p. 76).

[24] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. "Beyond memorability: visualization recognition and recall". In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (2015), pp. 519–528. DOI: 10.1109/TVCG.2015.2467732 (cit. on pp. 2, 11, 17, 19, 40, 77, 84, 98).

[25] Michelle A. Borkin, Azalea A. Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. "What makes a visualization memorable?" In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2306–2315. DOI: 10.1109/TVCG.2013.234 (cit. on pp. 20, 77).

[26] Magdalena Boucher, Benjamin Bach, Christina Stoiber, Zezhong Wang, and Wolfgang Aigner. "Educational data comics: What can comics do for education in visualization?" In: *2023 IEEE VIS Workshop on Visualization Education, Literacy, and Activities (EduVis)*. 2023, pp. 34–40. DOI: 10.1109/EduVis60792.2023.00012 (cit. on p. 14).

[27] Jeremy Boy, Francoise Detienne, and Jean-Daniel Fekete. "Storytelling in information visualizations: Does it engage users to explore data?" In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2015, pp. 1449–1458. DOI: 10.1145/2702123.2702452 (cit. on p. 14).

[28] Richard Brath. *Visualizing with text*. 1st ed. A K Peters/CRC Press, 2020. ISBN: 9780429290565. DOI: 10.1201/9780429290565 (cit. on p. 12).

[29] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology". In: *Qualitative Research in Psychology* 3.2 (2006), pp. 77–101. URL: https://doi.org/10.1191/1478088706qp063oa (cit. on pp. 30, 104).

[30] Matthew Brehmer and Tamara Munzner. "A multi-level typology of abstract visualization tasks". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2376–2385. DOI: 10.1109/TVCG.2013.124 (cit. on pp. 5, 12, 74, 75, 77, 80).

[31] Jens Breivik. "Argumentative patterns in students' online discussions in an introductory philosophy course: Micro-and macrostructures of argumentation as analytic tools". In: *Nordic Journal of Digital Literacy* 15.1 (2020), pp. 8–23. DOI: 10.18261/issn.1891-943x-2020-01-02 (cit. on p. 75).

[32] Nadieh Bremer. *Techniques for data visualization on both mobile & desktop*. https://www.visualcinnamon.com/2019/04/mobile-vs-desktop-dataviz/. (accessed March 26, 2025). 2019 (cit. on p. 95).

[33] Dennis Bromley and Vidya Setlur. "Dash: A bimodal data exploration tool for interactive text and visualizations". In: *2024 IEEE Visualization and Visual Analytics (VIS)*. 2024, pp. 256–260. DOI: 10.1109/VIS55277.2024.00059 (cit. on p. 16).

[34] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. "Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data?" In: *Perspectives on psychological science* 6.1 (2011), pp. 3–5. DOI: 10.1037/14805-009 (cit. on p. 29).

[35] Alyxander Burns, Cindy Xiong, Steven Franconeri, Alberto Cairo, and Narges Mahyar. "How to evaluate data visualizations across different levels of understanding". In: *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*. Piscataway, NJ: IEEE, 2020, pp. 19–28. DOI: 10.1109/BELIV51497.2020.00010 (cit. on pp. 12, 77).

[36] Victor Soares Bursztyn, Jane Hoffswell, Eunyee Koh, and Shunan Guo. "Representing charts as text for language models: An in-depth study of question answering for bar charts". In: *2024 IEEE Visualization and Visual Analytics (VIS)*. IEEE. 2024, pp. 266–270. DOI: 10.1109/VIS55277.2024.00061 (cit. on p. 15).

[37] Zoya Bylinskii, Michelle A Borkin, Nam Wook Kim, Hanspeter Pfister, and Aude Oliva. "Eye fixation metrics for large scale evaluation and comparison of information visualizations". In: *Workshop on Eye Tracking and Visualization*. Springer. 2015, pp. 235–255. DOI: 10.1007/978-3-319-47024-5_14 (cit. on pp. 2, 11, 19, 40, 98).

[38] Tianyuan Cai, Shaun Wallace, Tina Rezvanian, Jonathan Dobres, Bernard Kerr, Samuel Berlow, Jeff Huang, Ben D Sawyer, and Zoya Bylinskii. "Personalized font recommendations: Combining ML and typographic guidelines to optimize readability". In: *Proceedings of the 2022 ACM designing interactive systems conference*. 2022, pp. 1–25. DOI: 10.1145/3532106.3533457 (cit. on p. 142).

[39] André Calero Valdez, Martina Ziefle, and Michael Sedlmair. "Studying biases in visualization research: Framework and methods". In: *Cognitive biases in visualizations*. Springer, 2018, pp. 13–27. DOI: 10.1007/978-3-319-95831-6_2 (cit. on p. 53).

[40] Yining Cao, Jane L E, Chen Zhu-Tian, and Haijun Xia. "DataParticles: Block-based and language-oriented authoring of animated unit visualizations". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2023. DOI: 10.1145/3544548.3581472 (cit. on p. 15).

[41] Stuart K Card, Jock D Mackinlay, and Ben Shneiderman. *Readings in information visualization - using vision to think*. Academic Press, 1999. ISBN: 978-1-55860-533-6 (cit. on p. 2).

[42] Ashley Castleberry and Amanda Nolen. "Thematic analysis of qualitative research data: Is it as easy as it sounds?" In: *Currents in pharmacy teaching and learning* 10.6 (2018), pp. 807–815. DOI: 10.1016/j.cptl.2018.03.019 (cit. on pp. 30, 104).

[43] Charles Chen, Ruiyi Zhang, Eunyee Koh, Sungchul Kim, Scott Cohen, and Ryan Rossi. "Figure captioning with relation maps for reasoning". In: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2020, pp. 1526–1534. DOI: 10.1109/WACV45572.2020.9093592 (cit. on p. 15).

[44] Hao Cheng, Junhong Wang, Yun Wang, Bongshin Lee, Haidong Zhang, and Dongmei Zhang. "Investigating the role and interplay of narrations and animations in data videos". In: *Computer Graphics Forum* 41 (2022). DOI: 10.1111/cgf.14560 (cit. on p. 15).

[45] Victoria Clarke, Virginia Braun, and Nikki Hayfield. "Thematic analysis". In: *Qualitative Psychology: A Practical Guide to Research Methods*. SAGE, 2015. ISBN: 1526422867 (cit. on pp. 30, 104).

[46] William S Cleveland and Robert McGill. "Graphical perception: Theory, experimentation, and application to the development of graphical methods". In: *Journal of the American Statistical Association* 79.387 (1984), pp. 531–554. DOI: 10.2307/2288400 (cit. on pp. 2, 5, 10).

[47] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. 2nd ed. Routledge, 1988. ISBN: 0805802835 (cit. on p. 64).

[48] Graphic Detail. *Mega phone*. https://www.economist.com/graphic-detail/2012/09/12/mega-phone. (accessed March 26, 2025). 2012 (cit. on p. 91).

[49] Graphic Detail. *Tottering towers*. https://www.economist.com/graphic-detail/2011/10/24/tottering-towers. (accessed March 26, 2025). 2011 (cit. on p. 86).

[50] Kedar Dhamdhere, Kevin S McCurley, Ralfi Nahmias, Mukund Sundararajan, and Qiqi Yan. "Analyza: Exploring data with conversation". In: *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. 2017, pp. 493–504. DOI: 10.1145/3025171.302522 (cit. on p. 15).

[51] Alexandra Diehl, Alfie Abdul-Rahman, Mennatallah El-Assady, Benjamin Bach, Daniel Keim, and Min Chen. "VisGuides: A forum for discussing visualization guidelines". In: *EuroVis 2018 - Short Papers*. Ed. by Jimmy Johansson, Filip Sadlo, and Tobias Schreck. The Eurographics Association, 2018. DOI: 10.2312/eurovisshort.20181079 (cit. on p. 10).

[52] Diogo Duarte, Rita Costa, Pedro Bizarro, and Carlos Duarte. "AutoVizuA11y: A tool to automate screen reader accessibility in charts". In: *Computer Graphics Forum*. The Eurographics Association and John Wiley & Sons Ltd, 2024. DOI: 10.1111/cgf.15099 (cit. on pp. 15, 109).

[53] Vanessa Echeverria, Roberto Martinez-Maldonado, and Simon Buckingham Shum. "Towards data storytelling to support teaching and learning". In: *Proceedings of the 29th Australian Conference on Computer-Human Interaction*. Association for Computing Machinery, 2017, pp. 347–351. DOI: 10.1145/3152771.3156134 (cit. on p. 14).

[54] Mike Eissele, Daniel Weiskopf, and Thomas Ertl. "Interactive context-aware visualization for mobile devices". In: *Smart Graphics: 10th International Symposium, SG 2009*. Springer. 2009, pp. 167–178. DOI: 10.1007/978-3-642-02115-2_14 (cit. on p. 115).

[55] Frank Elavsky, Lucas Nadolskis, and Dominik Moritz. "Data navigator: An accessibility-centered data navigation toolkit". In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (2023), pp. 803–813. DOI: 10.1109/TVCG.2023.3327393 (cit. on p. 143).

[56] Nina Errey, Jie Liang, Tuck Wah Leong, and Didar Zowghi. "Evaluating narrative visualization: A survey of practitioners". In: *International Journal of Data Science and Analytics* 18.1 (2024), pp. 19–34. DOI: 10.1007/s41060-023-00394-9 (cit. on p. 14).

[57] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. "G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences". In: *Behavior Research Methods* 39.2 (2007), pp. 175–191. DOI: 10.3758/BF03193146 (cit. on pp. 28, 62).

[58] Guangchao Charles Feng. "Factors affecting intercoder reliability: A Monte Carlo experiment". In: *Quality & Quantity* 47.5 (2013), pp. 2959–2982. DOI: 10.1007/s11135-012-9745-9 (cit. on p. 44).

[59] Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. "Uncertainty displays using quantile dotplots or CDFs improve transit decision-making". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2018, pp. 1–12. DOI: 10.1145/3173574.3173718 (cit. on p. 141).

[60] Stephen Few. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. 2nd ed. Analytics Press, 2012. ISBN: 0970601972 (cit. on pp. 20, 25, 72).

[61] Ana Figueiras. "Narrative visualization: A case study of how to incorporate narrative elements in existing visualizations". In: *2014 18th International Conference on Information Visualisation*. 2014, pp. 46–52. DOI: 10.1109/IV.2014.79 (cit. on p. 14).

[62] James Ford, Xingmeng Zhao, Dan Schumacher, and Anthony Rios. "Charting the future: Using chart question-answering for scalable evaluation of LLM-driven data visualizations". In: *Proceedings of the 31st International Conference on Computational Linguistics*. 2025, pp. 7497–7510. URL: https://aclanthology.org/2025.coling-main.501/ (cit. on p. 143).

[63] Steven L Franconeri, Lace M Padilla, Priti Shah, Jeffrey M Zacks, and Jessica Hullman. "The science of visual data communication: What works". In: *Psychological Science in the Public Interest* 22.3 (2021), pp. 110–161. URL: https://doi.org/10.1177/15291006211051956 (cit. on pp. 87, 107, 141).

[64] Racquel Fygenson, Lace Padilla, and Enrico Bertini. "Cognitive affordances in visualization: Related constructs, design factors, and framework". In: *IEEE Transactions on Visualization and Computer Graphics* (2025). URL: 10.1109/TVCG.2025.3610803 (cit. on pp. 10, 17).

[65] Mirta Galesic and Rocio Garcia-Retamero. "Graph literacy: A cross-cultural comparison". In: *Medical Decision Making* 31.3 (2011), pp. 444–457. DOI: 10.1177/0272989X10373805 (cit. on p. 27).

[66] Matthias Gamer, Jim Lemon, Ian Fellows, and Puspendra Singh. *Various coefficients of interrater reliability and agreement*. 2012. DOI: 10.32614/CRAN.package.irr (cit. on p. 79).

[67] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G Karahalios. "Datatone: Managing ambiguity in natural language interfaces for data visualization". In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*. 2015, pp. 489–500. DOI: 10.1145/2807442.2807478 (cit. on p. 15).

[68] Rocio Garcia-Retamero, Edward T Cokely, Saima Ghazal, and Alexander Joeris. "Measuring graph literacy without a test: A Brief subjective assessment". In: *Medical Decision Making* 36.7 (2016), pp. 854–867. DOI: 10.1177/0272989X16655334 (cit. on pp. 27, 29, 43).

[69] Manuela Garreton, Francesca Morini, D Paz Moyano, Gianna-Carina Grün, Denis Parra, and Marian Dörk. "Data stories of water: Studying the communicative role of data visualizations within long-form journalism". In: *Computer Graphics Forum*. Vol. 42. 3. 2023. DOI: 10.1111/cgf.14815 (cit. on p. 14).

[70] Peter M Gollwitzer and John A Bargh. *The psychology of action: Linking cognition and motivation to behavior*. Guilford Press, 1996. ISBN: 1572300329 (cit. on p. 134).

[71] Elizabeth Goodman, Erik Stolterman, and Ron Wakkary. "Understanding interaction design practices". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2011, pp. 1061–1070. DOI: 10.1145/1978942.1979100 (cit. on p. 72).

[72] Richard L Gorsuch. "Exploratory factor analysis". In: *Handbook of Multivariate Experimental Psychology*. Springer, 1988, pp. 231–258. ISBN: 978-1-4612-8232-7. DOI: 10.1007/978-1-4613-0893-5 (cit. on p. 89).

[73] Steve Graham and Dolores Perin. "A meta-analysis of writing instruction for adolescent students". In: *Journal of educational psychology* 99.3 (2007), p. 445. DOI: 10.1037/0022-0663.99.3.445 (cit. on p. 118).

[74] Samuel Granados. *How today's visa restrictions might impact tomorrow's America*. https://www.washingtonpost.com/graphics/national/visas-impact/. (accessed March 26, 2025). 2017 (cit. on p. 92).

[75] Colin M. Gray, Erik Stolterman, and Martin A Siegel. "Reprioritizing the relationship between HCI research and practice: bubble-up and trickle-down effects". In: *Proceedings of the 2014 Conference on Designing Interactive Systems*. New York, NY, USA: Association for Computing Machinery, 2014, pp. 725–734. DOI: 10.1145/2598510.2598595 (cit. on p. 16).

[76] Jianing Hao, Manling Yang, Qing Shi, Yuzhe Jiang, Guang Zhang, and Wei Zeng. "FinFlier: Automating graphical overlays for financial visualizations with knowledge-grounding large language model". In: *IEEE Transactions on Visualization and Computer Graphics* (2024). DOI: 10.1109/TVCG.2024.3514138 (cit. on pp. 16, 108).

[77] Shan Hao, Zezhong Wang, Benjamin Bach, and Larissa Pschetz. "Design patterns for data-driven news articles". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2024. DOI: 10.1145/3613904.3641916 (cit. on pp. 13, 77, 80, 82, 85, 90).

[78] James C Hayton, David G Allen, and Vida Scarpello. "Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis". In: *Organizational research methods* 7.2 (2004), pp. 191–205. DOI: 10.1177/1094428104263675 (cit. on p. 89).

[79] Marti Hearst and Melanie Tory. "Would you like a chart with that? Incorporating visualizations into conversational interfaces". In: *2019 IEEE Visualization Conference (VIS)*. IEEE. 2019, pp. 1–5. DOI: 10.1109/VISUAL.2019.8933766 (cit. on pp. 12, 17, 18, 39, 116).

[80] Marti A Hearst. "On teaching narrative infographics in an information visualization course". In: *Pedagogy Data Visualization, IEEE VIS Workshop*. 2017 (cit. on p. 15).

[81] Marti A Hearst. "Show it or tell it? Text, visualization, and their combination". In: *Communications of the ACM* 66.10 (2023), pp. 68–75. DOI: 10.1145/3593580 (cit. on pp. 10, 12, 116).

[82] Jeffrey Heer and Michael Bostock. "Crowdsourcing graphical perception: Using mechanical turk to assess visualization design". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2010, pp. 203–212. DOI: 10.1145/1753326.1753357 (cit. on p. 10).

[83] Jane Hoffswell, Wilmot Li, and Zhicheng Liu. "Techniques for flexible responsive visualization design". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2020. ISBN: 9781450367080. DOI: 10.1145/3313831.3376777 (cit. on pp. 108, 113).

[84] Jana Holsanova, Nils Holmberg, and Kenneth Holmqvist. "Reading information graphics: The role of spatial contiguity and dual attentional guidance". In: *Applied Cognitive Psychology* 23.9 (2009), pp. 1215–1226. DOI: 10.1002/acp.1525 (cit. on pp. 11, 18).

[85] Tom Horak, Wolfgang Aigner, Matthew Brehmer, Alark Joshi, and Christian Tominski. "Responsive visualization design for mobile devices". In: *Mobile Data Visualization*. Chapman and Hall/CRC, 2021, pp. 33–66. ISBN: 9781003090823 (cit. on p. 142).

[86] Jonathan House and David Román. *Spain unveils new set of overhauls*. https://www.wsj.com/articles/SB10000872396390443916104578021692765950384. (accessed March 26, 2025). 2012 (cit. on pp. 81, 97).

[87] Daniel J Hruschka, Deborah Schwartz, Daphne Cobb St. John, Erin Picone-Decaro, Richard A Jenkins, and James W Carey. "Reliability in coding open-ended data: Lessons learned from HIV behavioral research". In: *Field Methods* 16.3 (2004), pp. 307–331. DOI: 10.1177/1525822X04266540 (cit. on p. 79).

[88] Jessica Hullman. “Why authors don’t visualize uncertainty”. In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2019), pp. 130–139. DOI: 10.1109/TVCG.2019.2934287 (cit. on p. 141).

[89] Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. “Contextifier: Automatic generation of annotated stock visualizations”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2013, pp. 2707–2716. DOI: 10.1145/2470654.2481374 (cit. on p. 108).

[90] Jessica Hullman and Nick Diakopoulos. “Visualization Rhetoric: Framing Effects in Narrative Visualization”. In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2231–2240. DOI: 10.1109/TVCG.2011.255 (cit. on pp. 10, 14, 15, 134).

[91] Alyson Hurt. *Coal, gas, nuclear, hydro? How your state generates power*. https://www.npr.org/2015/09/10/319535020/coal-gas-nuclear-hydro-how-your-state-generates-power. (accessed March 26, 2025). 2015 (cit. on p. 86).

[92] ISO/TC 37/SC 1. *A typology of language registers*. Tech. rep. International Organization for Standardization, 2018. URL: https://www.iso.org/standard/68852.html (cit. on p. 85).

[93] Sirkka L Jarvenpaa. “Graphic displays in decision making -— the visual salience effect”. In: *Journal of Behavioral Decision Making* 3.4 (1990), pp. 247–262. DOI: 10.1002/bdm.3960030403 (cit. on p. 12).

[94] Jimmy Johansson, Tomasz Opach, Erik Glaas, Tina-Simone Neset, Carlo Navarra, Björn-Ola Linnér, and Jan Ketil Rød. “VisAdapt: A visualization tool to support climate change adaptation”. In: *IEEE computer graphics and applications* 37.2 (2016), pp. 54–65. DOI: 10.1109/MCG.2016.49 (cit. on p. 108).

[95] Susan L. Joslyn and Jared E LeClerc. “Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error”. In: *Journal of Experimental Psychology: Applied* 18.1 (2012), pp. 126–140. DOI: 10.1037/a0025185 (cit. on p. 12).

[96] Crescentia Jung, Shubham Mehta, Atharva Kulkarni, Yuhang Zhao, and Yea-Seul Kim. “Communicating visualizations without visuals: Investigation of visualization tlternative text for people with visual impairments”. In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2022), pp. 1095–1105. DOI: 10.1109/TVCG.2021.3114846 (cit. on pp. 15, 109).

[97] Kushal Kafle, Brian Price, Scott Cohen, and Christopher Kanan. “DVQA: Understanding data visualizations via question answering”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 5648–5656. DOI: 10.1109/CVPR.2018.00592 (cit. on p. 143).

[98] Shankar Kantharaj, Rixie Tiffany Leong, Xiang Lin, Ahmed Masry, Megh Thakkar, Enamul Hoque, and Shafiq Joty. "Chart-to-text: A large-scale benchmark for chart summarization". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. 2022, pp. 4005–4023. DOI: 10.18653/v1/2022.acl-long.277 (cit. on p. 15).

[99] Harmanpreet Kaur, Alex C. Williams, Anne Loomis Thompson, Walter S. Lasecki, Shamsi T. Iqbal, and Jaime Teevan. "Creating better action plans for writing tasks via vocabulary-based planning". In: *Proceedings of the ACM on Human-Computer Interaction* 2 (2018). DOI: 10.1145/3274355 (cit. on p. 134).

[100] Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. "When (ish) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016, pp. 5092–5103. DOI: 10.1145/2858036.285855 (cit. on p. 141).

[101] Janice M Keenan. "Development of microstructure processes in children's reading comprehension: Effect of number of different arguments". In: *Journal of Experimental Psychology* 12.4 (1986), p. 614. DOI: 10.1037/0278-7393.12.4.614 (cit. on p. 75).

[102] Christa Kelleher and Thorsten Wagener. "Ten guidelines for effective data visualization in scientific publications". In: *Environmental Modelling & Software* 26.6 (2011), pp. 822–827. DOI: 10.1016/j.envsoft.2010.12.006 (cit. on p. 10).

[103] Dae Hyun Kim, Seulgi Choi, Juho Kim, Vidya Setlur, and Maneesh Agrawala. "EmphasisChecker: A tool for guiding chart and caption emphasis". In: *IEEE Transactions on Visualization and Computer Graphics* (2023). DOI: 10.1109/TVCG.2023.3327150 (cit. on p. 53).

[104] Dae Hyun Kim, Enamul Hoque, and Maneesh Agrawala. "Answering questions about charts and generating visual explanations". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–13. DOI: 10.1145/3313831.3376467 (cit. on p. 15).

[105] Dae Hyun Kim, Vidya Setlur, and Maneesh Agrawala. "Towards understanding how readers integrate charts and captions: A case study with line charts". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–11. DOI: 10.1145/3411764.3445443 (cit. on pp. 2, 20–22, 41, 50, 53, 70, 98, 99).

[106] Hyeok Kim, Dominik Moritz, and Jessica Hullman. "Design patterns and trade-offs in responsive visualization for communication". In: *Computer Graphics Forum*. Vol. 40. 3. Wiley Online Library. 2021, pp. 459–470. DOI: 10.1111/cgf.14321 (cit. on p. 113).

[107] Jiho Kim, Arjun Srinivasan, Nam Wook Kim, and Yea-Seul Kim. "Exploring chart question answering for blind and low vision users". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2023. DOI: 10.1145/3544548.3581532 (cit. on p. 15).

[108] Yea-Seul Kim, Katharina Reinecke, and Jessica Hullman. "Explaining the gap: Visualizing one's predictions improves recall and comprehension of data". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 1375–1386. DOI: 10.1145/3025453.3025592 (cit. on pp. 11, 12, 14, 19, 71, 110, 134).

[109] Neil King Jr. *Lost in translation: GOP struggles With hispanics*. https://www.wsj.com/articles/SB10000872396390444082904577605824074935132. (accessed March 26, 2025). 2012 (cit. on p. 86).

[110] Andy Kirk. *Data visualisation: A handbook for data driven design*. Sage, 2016. ISBN: 9781473912144 (cit. on p. 25).

[111] Cole Nussbaumer Knaflic. *So what?* Storytelling with Data Blog. 2017. URL: https://www.storytellingwithdata.com/blog/2017/3/22/so-what (cit. on pp. 12, 20).

[112] Cole Nussbaumer Knaflic. *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, 2015. ISBN: 978-1119002062 (cit. on pp. 12, 14, 20, 24, 25, 54, 72, 87).

[113] Paul A Kolers. "The recognition of geometrically transformed text". In: *Perception & Psychophysics* 3.1 (1968), pp. 57–64. DOI: 10.3758/BF03212713 (cit. on p. 25).

[114] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. "Frames and slants in titles of visualizations on controversial topics". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–12. DOI: 10.1145/3173574.3174012 (cit. on pp. 2, 11, 17, 40, 56, 70, 82, 84).

[115] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. "Trust and recall of information across varying degrees of title-visualization misalignment". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–13. DOI: 10.1145/3290605.3300576 (cit. on pp. 2, 11, 40, 53, 82, 84).

[116] Robert Kosara and Jock Mackinlay. "Storytelling: The next step for visualization". In: *Computer* 46.5 (2013), pp. 44–50. DOI: 10.1109/MC.2013.36 (cit. on p. 14).

[117] Kari Kuutti and Liam J Bannon. "The turn to practice in HCI: towards a research agenda". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014, pp. 3543–3552. DOI: 10.1145/2556288.2557111 (cit. on p. 72).

[118] Xingyu Lan, Yanqiu Wu, and Nan Cao. "Affective visualization design: Leveraging the emotional impact of data". In: *IEEE Transactions on Visualization and Computer Graphics* (2023). DOI: 10.1109/TVCG.2023.3327385 (cit. on pp. 77, 83).

[119] Shahid Latif, Kaidie Su, and Fabian Beck. "Authoring combined textual and visual descriptions of graph data". In: *EuroVis (Short Papers)*. 2019, pp. 115–119 (cit. on p. 15).

[120] Shahid Latif, Zheng Zhou, Yoon Kim, Fabian Beck, and Nam Wook Kim. "Kori: Interactive synthesis of text and charts in data documents". In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2022), pp. 184–194. DOI: 10.1109/TVCG.2021.3114802 (cit. on p. 15).

[121] Claire Lauer and Shaun O'Brien. "How people are influenced by deceptive tactics in everyday charts and graphs". In: *IEEE Transactions on Professional Communication* 63.4 (2020), pp. 327–340. DOI: 10.1109/TPC.2020.3032053 (cit. on pp. 11, 53).

[122] Claire Lauer and Christopher A Sanchez. "What eye tracking can show us about how people are influenced by deceptive tactics in line graphs". In: *IEEE Transactions on Professional Communication* 66.3 (2023), pp. 220–235. DOI: 10.1109/TPC.2023.3290948 (cit. on p. 11).

[123] Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. "Task taxonomy for graph visualization". In: *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors*. New York, NY, USA: Association for Computing Machinery, 2006. DOI: 10.1145/1168149.1168168 (cit. on pp. 12, 77).

[124] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. "More than telling a story: Transforming data into visually shared stories". In: *IEEE Computer Graphics and Applications* 35.5 (2015), pp. 84–90. DOI: 10.1109/MCG.2015.99 (cit. on p. 14).

[125] Elsie Lee-Robbins and Eytan Adar. "Affective learning objectives for communicative visualizations". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (2022), pp. 1–11. DOI: 10.1109/TVCG.2022.3209500 (cit. on pp. 77, 83).

[126] James H Leigh. "The use of figures of speech in print ad headlines". In: *Journal of advertising* 23.2 (1994), pp. 17–33. DOI: 10.1080/00913367.1994.10673439 (cit. on pp. 77, 80, 82, 85).

[127] Timothy R Levine and Craig R Hullett. "Eta squared, partial eta squared, and misreporting of effect size in communication research". In: *Human Communication Research* 28.4 (2002), pp. 612–625. DOI: 10.1111/j.1468-2958.2002.tb00828.x (cit. on p. 64).

[128] Johannes Liem, Charles Perin, and Jo Wood. "Structure and empathy in visual data storytelling: Evaluating their influence on attitude". In: *Computer Graphics Forum*. Vol. 39. 3. Wiley Online Library. 2020, pp. 277–289. DOI: 10.1111/cgf.13980 (cit. on p. 14).

[129] Yanna Lin, Haotian Li, Leni Yang, Aoyu Wu, and Huamin Qu. "Inksight: Leveraging sketch interaction for documenting chart findings in computational notebooks". In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (2024), pp. 944–954. DOI: 10.1109/TVCG.2023.3327170 (cit. on pp. 15, 99, 143).

[130] Maxim Lisnic, Vidya Setlur, and Nicole Sultanum. "Plume: Scaffolding gext composition in dashboards". In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2025. DOI: 10.1145/3706598.3713580 (cit. on p. 16).

[131] Can Liu, Yuhan Guo, and Xiaoru Yuan. "AutoTitle: An interactive title generator for visualizations". In: *IEEE Transactions on Visualization and Computer Graphics* (2023). DOI: 10.1109/TVCG.2023.3290241 (cit. on pp. 13, 74, 75, 82).

[132] Leo Yu-Ho Lo and Huamin Qu. "How good (Or bad) are LLMs at detecting misleading visualizations?" In: *IEEE Transactions on Visualization and Computer Graphics* 31.1 (2024), pp. 1116–1125. DOI: 10.1109/TVCG.2024.3456333 (cit. on p. 16).

[133] H Locke. *Why dark mode causes more accessibility issues than it solves*. Medium blog service. 2021. URL: https://medium.com/@h_locke/why-dark-mode-causes-more-accessibility-issues-than-it-solves-d2f8359bb46a (cit. on p. 20).

[134] Alan Lundgard and Arvind Satyanarayan. "Accessible visualization via natural language descriptions: A four-level model of semantic content". In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021), pp. 1073–1083. DOI: 10.1109/TVCG.2021.3114770 (cit. on pp. 6, 20, 41, 44, 53, 54, 74, 75, 77, 84, 109, 110, 137).

[135] Jock Mackinlay. "Automating the design of graphical presentations of relational information". In: *ACM Transactions On Graphics* 5.2 (1986), pp. 110–141. DOI: 10.1145/22949.22950 (cit. on pp. 2, 5).

[136] Shambhavi Mahajan, Bonnie Chen, Alireza Karduni, Yea-Seul Kim, and Emily Wall. "Vibe: A design space for visual belief elicitation in data journalism". In: *Computer Graphics Forum*. Vol. 41. Wiley Online Library. 2022, pp. 477–488. DOI: 10.1111/cgf.14556 (cit. on p. 134).

[137] Ahmed Masry, Do Long, Jia Qing Tan, Shafiq Joty, and Enamul Hoque. "ChartQA: A benchmark for question answering about charts with visual and logical reasoning". In: *Findings of the Association for Computational Linguistics: ACL 2022*. Association for Computational Linguistics, 2022, pp. 2263–2279. DOI: 10.18653/v1/2022.findings-acl.177 (cit. on pp. 15, 143).

[138] Sorin Adam Matei and Lucas Hunter. "Data storytelling is not storytelling with data: A framework for storytelling in science communication and data journalism". In: *The Information Society* 37.5 (2021), pp. 312–322. DOI: 10.1080/01972243.2021.1951415 (cit. on p. 14).

[139] Bryan J Matlen, Dedre Gentner, and Steven L Franconeri. "Spatial alignment facilitates visual comparison". In: *Journal of Experimental Psychology: Human Perception and Performance* 46.5 (2020), p. 443. DOI: 10.1111/cogs.13182 (cit. on p. 10).

[140] Laura E Matzen, Breannan C Howell, Marie Tuft, and Kristin M Divis. "Transparent risks: The impact of the specificity and visual encoding of uncertainty on decision making". In: *Computer Graphics Forum*. Vol. 43. 3. Wiley Online Library. 2024, e15094. DOI: 10.1111/cgf.15094 (cit. on p. 12).

[141] Richard E Mayer. *Multimedia Learning*. 3rd ed. Cambridge University Press, 2020. DOI: 10.1017/9781316941355 (cit. on pp. 11, 18).

[142] Richard E Mayer and Roxana Moreno. "Nine ways to reduce cognitive load in multimedia learning". In: *Educational psychologist* 38.1 (2003), pp. 43–52. DOI: 10.1207/S15326985EP3801_6 (cit. on p. 112).

[143] Nina McCurdy, Jason Dykes, and Miriah Meyer. "Action design research and visualization design". In: *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. 2016, pp. 10–18. DOI: 10.1145/2993901.2993916 (cit. on pp. 16, 125).

[144] Sean McKenna, Nathalie Henry Riche, Bongshin Lee, Jeremy Boy, and Miriah Meyer. "Visual narrative flow: Exploring factors shaping data visualization story reading experiences". In: *Computer Graphics Forum*. Vol. 36. 3. Wiley Online Library. 2017, pp. 377–387. DOI: 10.1111/cgf.13195 (cit. on pp. 12, 19).

[145] Sean McKenna, Dominika Mazur, James Agutter, and Miriah Meyer. "Design activity framework for visualization design". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 2191–2200. DOI: 10.1109/tvcg.2014.2346331 (cit. on pp. 16, 99, 118, 122, 125).

[146] Merriam-Webster. *Rudder*. URL: https://www.merriam-webster.com/dictionary/rudder (visited on 03/30/2024) (cit. on p. 118).

[147] Ronald Metoyer, Qiyu Zhi, Bart Janczuk, and Walter Scheirer. "Coupling story to visualization: Using textual analysis as a bridge between data and interpretation". In: *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. Association for Computing Machinery, 2018, pp. 503–507. DOI: 10.1145/3172944.3173007 (cit. on p. 15).

[148] Melanie Metzger. "Sign language interpreting: Deconstructing the myth of neutrality". In: Gallaudet University Press, 1999. Chap. Neutrality in Translation and Interpretation. DOI: 10.2307/j.ctv2t5xgrb (cit. on p. 24).

[149] Miriah Meyer and Jason Dykes. "Criteria for rigor in visualization design study". In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2019), pp. 87–97. DOI: 10.1109/TVCG.2019.2934539 (cit. on p. 72).

[150] Microsoft. *Microsoft Power BI*. 2022. URL: https://powerbi.microsoft.com/en-us/ (cit. on p. 15).

[151] Barbara Millet, Sharanya J Majumdar, Alberto Cairo, Carolina Diaz, Qinyu Ding, Scotney D Evans, and Kenneth Broad. "End-user preference for and understanding of hurricane forecast graphs". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 65. SAGE Publications. 2021, pp. 606–610. DOI: 10.1177/1071181321651142 (cit. on pp. 27, 43).

[152] Josh Mitchell and Rachel Louise Ensign. *Student-loan defaults mount again*. https://www.wsj.com/articles/SB10000872396390444712904578024412786219802. (accessed March 26, 2025). 2012 (cit. on p. 95).

[153] Vibhu O Mittal, Johanna D Moore, Giuseppe Carenini, and Steven Roth. "Describing complex charts in natural language: A caption generation system". In: *Computational Linguistics* 24.3 (1998), pp. 431–467. URL: https://aclanthology.org/J98-3004/ (cit. on p. 15).

[154] Tamara Munzner. "A nested model for visualization design and validation". In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (2009), pp. 921–928. DOI: 10.1109/TVCG.2009.111 (cit. on pp. 16, 118, 125).

[155] Alireza Nili, Mary Tate, Alistair Barros, and David Johnstone. "An approach for selecting and using a method of inter-coder reliability in information management research". In: *International Journal of Information Management* 54 (2020), p. 102154. DOI: 10.1016/j.ijinfomgt.2020.102154 (cit. on p. 79).

[156] Dion Nissenbaum. *Delays imperil mining riches Afghans need after pullout*. https://www.wsj.com/articles/SB10001424052702304019404577418363707647028. (accessed March 26, 2025). 2012 (cit. on p. 91).

[157] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic Books, 2013. ISBN: 9780385267748 (cit. on pp. 10, 134).

[158] Alvitta Ottley, Aleksandra Kaszowska, R Jordan Crouser, and Evan M Peck. "The curious case of combining text and visualization". In: *EuroVis (Short Papers)*. 2019, pp. 121–125 (cit. on pp. 11, 19, 71).

[159] Lace Padilla, Racquel Fygenson, Spencer C Castro, and Enrico Bertini. "Multiple forecast visualizations (mfvs): Trade-offs in trust and performance in multiple covid-19 forecast visualizations". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (2022), pp. 12–22. DOI: 10.1109/TVCG.2022.3209457 (cit. on p. 141).

[160] Lace Padilla, Helia Hosseinpour, Racquel Fygenson, Jennifer Howell, Rumi Chunara, and Enrico Bertini. "Impact of COVID-19 forecast visualizations on pandemic risk perceptions". In: *Scientific reports* 12.1 (2022), p. 2014. DOI: 10.1038/s41598-022-07502-y (cit. on p. 141).

[161] Lace M Padilla, Sarah H Creem-Regehr, Mary Hegarty, and Jeanine K Stefanucci. "Decision making with visualizations: A cognitive framework across disciplines". In: *Cognitive research: principles and implications* 3 (2018), pp. 1–25. DOI: 10.1186/s41235-018-0120-9 (cit. on p. 141).

[162] Lace MK Padilla, Maia Powell, Matthew Kay, and Jessica Hullman. “Uncertain about uncertainty: How qualitative expressions of forecaster confidence impact decision-making with uncertainty visualizations”. In: *Frontiers in Psychology* 11 (2021). DOI: 10.3389/fpsyg.2020.579267 (cit. on p. 141).

[163] Stefan Palan and Christian Schitter. “Prolific. ac—A subject pool for online experiments”. In: *Journal of behavioral and experimental finance* 17 (2018), pp. 22–27. DOI: 10.1016/j.jbef.2017.12.004 (cit. on p. 63).

[164] Liudas Panavas, Amy E Worth, Tarik Crnovrsanin, Tejas Sathyamurthi, Sara Cordes, Michelle A Borkin, and Cody Dunne. “Juvenile graphical perception: A comparison between children and adults”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–14. DOI: 10.1145/3491102.3501893 (cit. on p. 10).

[165] Michael R Parke, Justin M Weinhardt, Andrew Brodsky, Subrahmaniam Tangirala, and Sanford E DeVoe. “When daily planning improves employee performance: The importance of planning type, engagement, and interruptions”. In: *Journal of Applied Psychology* 103.3 (2018), p. 300. DOI: 10.1037/apl0000278 (cit. on p. 134).

[166] Paul Parsons. “Preparing future data visualization designers for professional practice”. In: *2023 IEEE VIS Workshop on Visualization Education, Literacy, and Activities (EduVis)*. IEEE. IEEE Computer Society, 2023, pp. 22–27. DOI: 10.1109/EduVis60792.2023.00010 (cit. on p. 15).

[167] Paul Parsons. “Understanding data visualization design practice”. In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2021), pp. 665–675. DOI: 10.1109/TVCG.2021.3114959 (cit. on pp. 16, 99, 117).

[168] Paul Parsons, Ali Baigelenov, Ya-Hsin Hung, and Connor Schrank. “What design methods do dataVis practitioners know and use?” In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–8. DOI: 10.1145/3334480.3383048 (cit. on pp. 16, 72, 99).

[169] Paul Parsons, Prakash Shukla, and Chorong Park. “Fixation and creativity in data visualization design: Experiences and perspectives of practitioners”. In: *2021 IEEE Visualization Conference (VIS)*. IEEE. 2021, pp. 76–80. DOI: 10.1109/VIS49827.2021.9623297 (cit. on p. 16).

[170] Dulce G Pereira, Anabela Afonso, and Fátima Melo Medeiros. “Overview of Friedman’s test and post-hoc analysis”. In: *Communications in Statistics-Simulation and Computation* 44.10 (2015), pp. 2636–2653. DOI: 10.1080/03610918.2014.931971 (cit. on pp. 29, 30).

[171] Jana Pflaeging and Hartmut Stöckl. “The rhetoric of multimodal communication”. In: *Visual Communication* 20.3 (2021), pp. 319–326. DOI: 10.1177/14703572211010200 (cit. on p. 77).

[172] Catherine Plaisant. "Information visualization and the challenge of universal usability". In: *Exploring Geovisualization*. Elsevier, 2005, pp. 53–82. DOI: 10.1016/B978-008044531-1/50421-8 (cit. on p. 107).

[173] Sydney Purdue, Eduardo Puerta, Enrico. Bertini, and Melanie Tory. "Stitching meaning: practices of data textile creators". In: *IEEE Transactions on Visualization and Computer Graphics* (2025). DOI: 10.1109/TVCG.2025.3634255 (cit. on p. 142).

[174] Xin Qian, Eunyee Koh, Fan Du, Sungchul Kim, Joel Chan, Ryan A Rossi, Sana Malik, and Tak Yeon Lee. "Generating accurate caption units for figure captioning". In: *Proceedings of the Web Conference 2021*. 2021, pp. 2792–2804. DOI: 10.1145/3442381.3449923 (cit. on p. 15).

[175] Md Dilshadur Rahman, Bhavana Doppalapudi, Ghulam Jilani Quadri, and Paul Rosen. "A survey on annotations in information visualization: empirical studies, applications and challenges". In: *IEEE Transactions on Visualization and Computer Graphics* 31.12 (2025), pp. 10439–10456. DOI: 10.1109/TVCG.2025.3600957 (cit. on pp. 13, 74, 75).

[176] Md Dilshadur Rahman, Ghulam Jilani Quadri, Bhavana Doppalapudi, Danielle Albers Szafir, and Paul Rosen. "A qualitative analysis of common practices in annotations: A taxonomy and design space". In: *IEEE Transactions on Visualization and Computer Graphics* (2024). DOI: 10.1109/TVCG.2024.3456359 (cit. on pp. 12, 74, 75, 77, 80, 84, 99).

[177] Noëlle Rakotondravony, Priya Dhawka, and Melanie Bancilhon. "Beyond English: Centering multilingualism in data visualization". In: *IEEE VIS Workshop on Visualization for Social Good 2023*. 2023 (cit. on p. 141).

[178] Donghao Ren, Matthew Brehmer, Bongshin Lee, Tobias Höllerer, and Eun Kyoung Choe. "ChartAccent: Annotation for data-driven storytelling". In: *2017 IEEE Pacific Visualization Symposium (PacificVis)*. 2017, pp. 230–239. DOI: 10.1109/PACIFICVIS.2017.8031599 (cit. on p. 94).

[179] Pengkun Ren, Yi Wang, and Fan Zhao. "Re-understanding of data storytelling tools from a narrative perspective". In: *Visual Intelligence* 1.1 (2023). DOI: 10.1007/s44267-023-00011-0 (cit. on p. 77).

[180] William Revelle. *psych: Procedures for psychological, psychometric, and personality research*. 2024. DOI: 10.32614/CRAN.package.psych (cit. on p. 89).

[181] Robert B Ricco. "The macrostructure of informal arguments: A proposed model and analysis". In: *The Quarterly Journal of Experimental Psychology* 56.6 (2003), pp. 1021–1051. DOI: 10.1080/02724980244000701 (cit. on p. 75).

[182] Catherine Sheldrick Ross, Kirsti Nilsen, and Marie L Radford. *Conducting the reference interview*. American Library Association, 2019. ISBN: 978-0838917275 (cit. on p. 118).

[183] Daniel M. Russell. *The Joy of Search: A Google Insider's Guide to Going Beyond the Basics*. MIT Press, 2019. ISBN: 9780262354721. DOI: 10.7551/mitpress/11920.001.0001 (cit. on pp. 118, 134).

[184] Alper Sarikaya, Michael Correll, Lyn Bartram, Melanie Tory, and Danyel Fisher. "What do we talk about when we talk about dashboards?" In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 682–692. DOI: 10.1109/TVCG.2018.2864903 (cit. on pp. 13, 89, 90).

[185] Karen B Schloss, Laurent Lessard, Charlotte S Walmsley, and Kathleen Foley. "Color inference in visual communication: the meaning of colors in recycling". In: *Cognitive research: principles and implications* 3.1 (2018), p. 5. DOI: 10.1186/s41235-018-0090-y (cit. on p. 17).

[186] Jordan T Schugar, Heather Schugar, and Christian Penny. "A nook or a book? Comparing college students' reading comprehension levels, critical reading, and study skills". In: *International Journal of Technology in Teaching & Learning* 7.2 (2011) (cit. on p. 77).

[187] Jonathan Schwabish, Susan J Popkin, and Alice Feng. *Do no harm guide: Centering accessibility in data visualization*. Tech. rep. Urban Institute, 2022. URL: https://www.urban.org/research/publication/do-no-harm-guide-centering-accessibility-data-visualization (cit. on pp. 41, 72).

[188] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. "Design study methodology: Reflections from the trenches and the stacks". In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2431–2440. DOI: 10.1109/TVCG.2012.213 (cit. on pp. 16, 72, 99).

[189] Edward Segel and Jeffrey Heer. "Narrative visualization: Telling stories with data". In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 1139–1148. DOI: 10.1109/TVCG.2010.179 (cit. on pp. 10, 14, 94, 134).

[190] V. Setlur and B Cogley. *Functional aesthetics for data visualization*. John Wiley & Sons, 2022. ISBN: 1119810086 (cit. on pp. 12, 54).

[191] Vidya Setlur, Sarah E Battersby, Melanie Tory, Rich Gossweiler, and Angel X Chang. "Eviza: A natural language interface for visual analysis". In: *Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology*. 2016, pp. 365–377. DOI: 10.1145/2984511.2984588 (cit. on p. 15).

[192] Priti Shah, Richard E Mayer, and Mary Hegarty. "Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension". In: *Journal of Educational Psychology* 91.4 (1999), p. 690. DOI: 10.1037/0022-0663.91.4.690 (cit. on p. 40).

[193] Hongbo Shao, Roberto Martinez-Maldonado, Vanessa Echeverria, Lixiang Yan, and Dragan Gasevic. "Data storytelling in data visualisation: Does it enhance the efficiency and effectiveness of information retrieval and insights comprehension?" In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2024. DOI: 10.1145/3613904.3643022 (cit. on p. 14).

[194] Michael R Sheldon, Michael J Fillyaw, and W Douglas Thompson. "The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs". In: *Physiotherapy Research International* 1.4 (1996), pp. 221–228. DOI: 10.1002/pri.66 (cit. on p. 29).

[195] Leixian Shen, Haotian Li, Yun Wang, and Huamin Qu. "From data to story: Towards automatic animated data video creation with llm-based multi-agent systems". In: *2024 IEEE VIS Workshop on Data Storytelling in an Era of Generative AI (GEN4DS)*. IEEE. 2024, pp. 20–27. DOI: 10.1109/GEN4DS63889.2024.00008 (cit. on p. 110).

[196] Ben Shneiderman. "The eyes have it: A task by data type taxonomy for information visualizations". In: *The craft of information visualization*. Elsevier, 2003, pp. 364–371. DOI: 10.1109/VL.1996.545307 (cit. on pp. 12, 77).

[197] Aletta Smits and Koen van Turnhout. "Towards a practice-led research agenda for user interface design of recommender systems". In: *IFIP Conference on Human-Computer Interaction*. Springer. 2023, pp. 170–190. DOI: 10.1007/978-3-031-42286-7_10 (cit. on pp. 16, 72).

[198] Arjun Srinivasan, Steven M Drucker, Alex Endert, and John Stasko. "Augmenting visualizations with interactive data facts to facilitate interpretation and communication". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2018), pp. 672–681. DOI: 10.1109/TVCG.2018.2865145 (cit. on pp. 15, 53).

[199] Arjun Srinivasan, Joanna Purich, Michael Correll, Leilani Battle, Vidya Setlur, and Anamaria Crisan. "From dashboard zoo to census: A case study with tableau public". In: *IEEE Transactions on Visualization and Computer Graphics* (2024). DOI: 10.1109/TVCG.2024.3490259 (cit. on pp. 13, 89).

[200] Christian Stab and Iryna Gurevych. "Parsing argumentation structures in persuasive essays". In: *Computational Linguistics* 43.3 (2017), pp. 619–659. DOI: 10.1162/COLI_a_00295 (cit. on pp. 75, 76).

[201] Chase Stokes, Anjana Arunkumar, Marti A. Hearst, and Lace Padilla. "An analysis of text functions in information visualization". In: *IEEE Transactions on Visualization and Computer Graphics* (2025). DOI: 10.1109/TVCG.2025.3634632 (cit. on p. 73).

[202] Chase Stokes, Cindy Xiong Bearfield, and Marti A Hearst. "The role of text in visualizations: How annotations shape perceptions of bias and influence predictions". In: *IEEE Transactions on Visualization and Computer Graphics* 30.10 (2023), pp. 6787–6800. DOI: 10.1109/TVCG.2023.3338451 (cit. on pp. 52, 74, 111).

[203] Chase Stokes and Marti A Hearst. "Give text a chance: Advocating for equal consideration for language and visualization". In: *NL VIZ: Workshop on Exploring Opportunities and Challenges for Natural Language Techniques to Support Visual Analysis*. IEEE. 2021 (cit. on pp. 22, 39, 98).

[204] Chase Stokes and Marti A Hearst. "Why more text is (often) better: Themes from reader preferences for integration of charts and text". In: *NL VIZ: Workshop on Exploring Opportunities and Challenges for Natural Language Techniques to Support Visual Analysis*. IEEE. 2022 (cit. on pp. 18, 53, 70, 107).

[205] Chase Stokes, Clara Hu, and Marti A Hearst. ""It's a good idea to put it into words": Writing 'rudders' in the initial stages of visualization design". In: *IEEE Transactions on Visualization and Computer Graphics* 31.1 (2025), pp. 1126–1136. DOI: 10.1109/TVCG.2024.3456324 (cit. on p. 117).

[206] Chase Stokes, Kylie Lin, and Cindy Xiong Bearfield. "Write, rank, or rate: Comparing methods for studying visualization affordances". In: *IEEE Transactions on Visualization and Computer Graphics* (2025). DOI: 10.1109/TVCG.2025.3633872 (cit. on pp. 10, 137).

[207] Chase Stokes, Chelsea Sanker, Bridget Cogley, and Vidya Setlur. "From delays to densities: Exploring data uncertainty through speech, text, and visualization". In: *Computer Graphics Forum*. Vol. 43. Wiley Online Library. 2024, e15100. DOI: 10.1111/cgf.15100 (cit. on p. 142).

[208] Chase Stokes, Chelsea Sanker, Bridget Cogley, and Vidya Setlur. "Mixing modes: Active and passive integration of speech, text, and visualization for communicating data uncertainty". In: *Eurographics Conference on Visualization (EuroVis)*. The Eurographics Association, 2024. DOI: 10.2312/evs.20241072 (cit. on p. 142).

[209] Chase Stokes, Chelsea Sanker, Bridget Cogley, and Vidya Setlur. "Voicing uncertainty: How speech, text, and visualizations influence decisions with data uncertainty". In: *2024 IEEE Workshop on Uncertainty Visualization: Applications, Techniques, Software, and Decision Frameworks*. IEEE. 2024, pp. 17–27. DOI: 10.1109/UncertaintyVisualization63963.2024.00007 (cit. on p. 142).

[210] Chase Stokes, Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti A Hearst. "Striking a balance: Reader takeaways and preferences when integrating text and charts". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (2022), pp. 1233–1243. DOI: 10.1109/TVCG.2022.3209383 (cit. on pp. 18, 40, 45, 53, 54, 74, 98, 99, 110, 116).

[211] Maureen Stone. "Choosing colors for data visualization". In: *Business Intelligence Network* 2.1 (2006) (cit. on p. 2).

[212] Saku Sugawara, Nikita Nangia, Alex Warstadt, and Samuel Bowman. "What makes reading comprehension questions difficult?" In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2022, pp. 6951–6971 (cit. on p. 77).

[213] Nicole Sultanum, Fanny Chevalier, Zoya Bylinskii, and Zhicheng Liu. "Leveraging text-chart links to support authoring of data-driven drticles with VizFlow". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–17. DOI: 0.1145/3411764.3445354 (cit. on p. 15).

[214] Nicole Sultanum and Vidya Setlur. "From instruction to insight: Exploring the functional and semantic roles of text in interactive dashboards". In: *IEEE Transactions on Visualization and Computer Graphics* 31.1 (2025), pp. 382–392. DOI: 10.1109/TVCG.2024.3456601 (cit. on pp. 13, 74, 78, 99, 107, 142).

[215] Nicole Sultanum and Arjun Srinivasan. "DataTales: Investigating the use of large language models for authoring data-driven articles". In: *IEEE Transactions on Visualization and Computer Graphics* (2023). DOI: 10.1109/VIS54172.2023.00055 (cit. on p. 108).

[216] Yiwen Sun, Jason Leigh, Andrew Johnson, and Sangyoon Lee. "Articulate: A semi-automated model for translating natural language queries into meaningful visualizations". In: *International Symposium on Smart Graphics*. Springer. 2010, pp. 184–195. DOI: 10.1007/978-3-642-13544-6_18 (cit. on p. 15).

[217] Charles S Taber and Milton Lodge. "Motivated skepticism in the evaluation of political beliefs". In: *American journal of political science* 50.3 (2006), pp. 755–769. DOI: 10.1111/j.1540-5907.2006.00214.x (cit. on p. 53).

[218] Tableau. *Tableau Summary Card*. Tableau Software help pages. 2022. URL: https://help.tableau.com/current/pro/desktop/en-us/inspectdata_summary.htm (visited on 2022) (cit. on p. 15).

[219] Benny Tang, Angie Boggust, and Arvind Satyanarayan. "VisText: A benchmark for semantically rich chart captioning". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, 2023, pp. 7268–7298. DOI: 10.18653/v1/2023.acl-long.401 (cit. on p. 15).

[220] Posit team. *RStudio: Integrated Development Environment for R*. Posit Software, PBC. Boston, MA, 2025. URL: http://www.posit.co/ (cit. on p. 89).

[221] Marta Terrado, Luz Calvo, and Isadora Christel. "Towards more effective visualisations in climate services: good practices and recommendations". In: *Climatic Change* 172.1 (2022), p. 18. DOI: 10.1007/s10584-022-03365-4 (cit. on p. 14).

[222] Bruce Thompson. *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* American Psychological Association, 2004. DOI: 10.1037/10694-000 (cit. on p. 89).

[223] Yuan Tian, Weiwei Cui, Dazhen Deng, Xinjing Yi, Yurun Yang, Haidong Zhang, and Yingcai Wu. "ChartGPT: Leveraging LLMs to generate charts from abstract natural language". In: *IEEE Transactions on Visualization and Computer Graphics* 31.3 (2024), pp. 1731–1745. DOI: 10.1109/TVCG.2024.3368621 (cit. on p. 16).

[224] Miles A Tinker. "Effects of angular alignment upon readability of print". In: *Journal of Educational Psychology* 47.6 (1956), p. 358. DOI: 10.1037/h0044504 (cit. on p. 25).

[225] Melanie Tory, Lyn Bartram, Brittany Fiore-Gartland, and Anamaria Crisan. "Finding their data voice: Practices and challenges of dashboard users". In: *IEEE Computer Graphics and Applications* 43.1 (2021), pp. 22–36. DOI: 10.1109/MCG.2021.3136545 (cit. on p. 99).

[226] Melanie Tory and Torsten Möller. "Rethinking visualization: A high-level taxonomy". In: *IEEE Symposium on Information Visualization.* 2004, pp. 151–158. DOI: 10.1109/INFVIS.2004.59 (cit. on pp. 12, 77).

[227] Edward R Tufte. *The visual display of quantitative information.* 2nd ed. Graphics Press, 2001. ISBN: 1930824130 (cit. on pp. 19, 24, 25, 38, 72, 106).

[228] Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. "Animation: can it facilitate?" In: *International journal of human-computer studies* 57.4 (2002), pp. 247–262. DOI: 10.1006/ijhc.2002.1017 (cit. on p. 112).

[229] Samuel Velasco. *The search for dark matter Is dramatically expanding.* https://www.quantamagazine.org/physicists-are-expanding-the-search-for-dark-matter-20201123/. (accessed March 26, 2025). 2020 (cit. on p. 94).

[230] Martin Voigt, Stefan Pietschmann, Lars Grammel, and Klaus Meißner. "Context-aware recommendation of visualization components". In: *The Fourth International Conference on Information, Process, and Knowledge Management (eKNOW).* Vol. 1. 2012, pp. 101–109 (cit. on p. 115).

[231] Karel Vredenburg, Ji-Ye Mao, Paul W Smith, and Tom Carey. "A survey of user-centered design practice". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2002, pp. 471–478. DOI: 10.1145/503376.503460 (cit. on p. 134).

[232] W3C. *How to Meet WCAG (Quick Reference).* World Wide Web Consortium website. 2019. URL: https://www.w3.org/WAI/tutorials/images/complex/ (cit. on pp. 25, 41, 109).

[233] W3C. *WAI Web Accessibility Tutorials: Complex Images.* World Wide Web Consortium website. 2019. URL: https://www.w3.org/WAI/tutorials/images/complex/ (visited on 11/03/2020) (cit. on p. 41).

[234] Shaun Wallace, Zoya Bylinskii, Jonathan Dobres, Bernard Kerr, Sam Berlow, Rick Treitman, Nirmal Kumawat, Kathleen Arpin, Dave B Miller, Jeff Huang, et al. "Towards individuated reading experiences: Different fonts increase reading speed for different individuals". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 29.4 (2022), pp. 1–56. DOI: 10.1145/3502222 (cit. on p. 142).

[235] Shaun Wallace, Rick Treitman, Jeff Huang, Ben D Sawyer, and Zoya Bylinskii. "Accelerating adult readers with typeface: A study of individual preferences and effectiveness". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020. DOI: 10.1145/3334480.3382985 (cit. on p. 142).

[236] Omri Wallach. *Visualizing the biggest tech mergers and acquisitions of 2020*. https://www.visualcapitalist.com/visualizing-biggest-tech-mergers-and-acquisitions-of-2020/. (accessed March 26, 2025). 2020 (cit. on p. 94).

[237] Jagoda Walny, Christian Frisson, Mieka West, Doris Kosminsky, Søren Knudsen, Sheelagh Carpendale, and Wesley Willett. "Data changes everything: Challenges and opportunities in data visualization design handoff". In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), pp. 12–22. DOI: 10.1109/TVCG.2019.2934538 (cit. on pp. 16, 118).

[238] Yanan Wang and Yea-Seul Kim. "Making data-driven articles more accessible: An active preference learning approach to data fact personalization". In: *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 2023, pp. 1353–1366. DOI: 10.1145/3563657.3595986 (cit. on p. 110).

[239] Yanan Wang, Arjun Srinivasan, and Yea-Seul Kim. "Enabling tabular data exploration for blind and low-vision users". In: *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 2024, pp. 1218–1233. DOI: 10.1145/3643834.3661609 (cit. on p. 110).

[240] Yun Wang, Leixian Shen, Zhengxin You, Xinhuan Shu, Bongshin Lee, John Thompson, Haidong Zhang, and Dongmei Zhang. "WonderFlow: Narration-centric design of animated data videos". In: *IEEE Transactions on Visualization and Computer Graphics* 31.9 (2025), pp. 4638–4654. DOI: 10.1109/TVCG.2024.3411575 (cit. on p. 15).

[241] Dana Linnell Wanzer, Tarek Azzam, Natalie D Jones, and Darrel Skousen. "The role of titles in enhancing data visualization". In: *Evaluation and Program Planning* 84 (2021). DOI: 10.1016/j.evalprogplan.2020.101896 (cit. on p. 13).

[242] Wibke Weber, Martin Engebretsen, and Helen Kennedy. "Data stories: Rethinking journalistic storytelling in the context of data journalism". In: *Studies in communication sciences* 2018.1 (2018), pp. 191–206. DOI: 10.24434/j.scoms.2018.01.013 (cit. on p. 14).

[243] Wikipedia. *Wikipedia*. The Wikipedia Foundation. 2022. URL: https://www.wikipedia.org/ (cit. on p. 21).

[244] Tom D Wilson. "Talking about the problem: A content analysis of pre-search interviews". In: *Information Research: An International Electronic Journal* 10.1 (2004). ISSN: ISSN:1368-1613 (cit. on p. 118).

[245] Andrew Wodehouse and William Ion. "The integration of information and ideas: Creating linkages through a novel concept design method". In: *Parsons Journal for Information Mapping* 2.2 (2010) (cit. on p. 16).

[246] Dona M Wong. *The Wall Street Journal guide to information graphics: The dos and don'ts of presenting data, facts, and figures*. WW Norton, 2010. ISBN: 978-0393609745 (cit. on pp. 19, 24, 25).

[247] Cindy Xiong, Cristina R Ceja, Casimir JH Ludwig, and Steven Franconeri. "Biased average position estimates in line and bar graphs: Underestimation, overestimation, and perceptual pull". In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2019), pp. 301–310. DOI: 10.1109/TVCG.2019.2934400 (cit. on p. 10).

[248] Cindy Xiong, Elsie Lee-Robbins, Icy Zhang, Aimen Gaba, and Steven Franconeri. "Reasoning affordances with tables and bar charts". In: *IEEE Transactions on Visualization and Computer Graphics* (2022). DOI: 10.1109/TVCG.2022.3232959 (cit. on pp. 17, 40).

[249] Cindy Xiong, Chase Stokes, Yea-Seul Kim, and Steven Franconeri. "Seeing what you believe or believing what you see? Belief biases correlation estimation". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (2023), pp. 493–503. DOI: 10.1109/TVCG.2022.3209405 (cit. on p. 144).

[250] Cindy Xiong, Lisanne Van Weelden, and Steven Franconeri. "The curse of knowledge in visual data communication". In: *IEEE Transactions on Visualization and Computer Graphics* 26.10 (2019), pp. 3051–3062. DOI: 10.1109/TVCG.2019.2917689 (cit. on pp. 11, 53).

[251] Cindy Xiong, Lisanne van Weelden, and Steven Franconeri. "Perceiving graphs as ambiguous figures". In: *Journal of Vision* 18.10 (2018), pp. 1327–1327. DOI: 10.1167/18.10.1327 (cit. on p. 53).

[252] Ying Xu, Dakuo Wang, Mo Yu, Daniel Ritchie, Bingsheng Yao, Tongshuang Wu, Zheng Zhang, Toby Jia-Jun Li, Nora Bradford, Branda Sun, Tran Bao Hoang, Yisi Sang, Yufang Hou, Xiaojuan Ma, Diyi Yang, Nanyun Peng, Zhou Yu, and Mark Warschauer. "Fantastic questions and where to find them: FairytaleQA – an authentic dataset for narrative comprehension". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2022, pp. 447–460. DOI: 10.18653/v1/2022.acl-long.34 (cit. on p. 77).

[253] Leni Yang, Cindy Xiong, Jason K. Wong, Aoyu Wu, and Huamin Qu. "Explaining with examples: Lessons learned from crowdsourced introductory description of information visualizations". In: *IEEE Transactions on Visualization and Computer Graphics* (2021), pp. 1638–1650. DOI: 10.1109/TVCG.2021.3128157 (cit. on pp. 13, 27, 43).

[254] Nathan Yau. *Gallery of uncertainty visualization methods*. https://flowingdata.com/2019/09/03/gallery-of-uncertainty-visualization-methods/. (accessed March 26, 2025). 2019 (cit. on p. 92).

[255] Nathan Yau. *Power sources in each state*. https://flowingdata.com/2015/09/18/power-sources-in-each-state/. (accessed March 26, 2025). 2015 (cit. on p. 86).

[256] Dominyk Zdanovic, Tanja Julie Lembcke, and Toine Bogers. "The influence of data storytelling on the ability to recall information". In: *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. 2022, pp. 67–77. DOI: 10.1145/3498366.3505755 (cit. on p. 14).

[257] Xingchen Zeng, Haichuan Lin, Yilin Ye, and Wei Zeng. "Advancing multimodal large language models in chart question answering with visualization-referenced instruction tuning". In: *IEEE Transactions on Visualization and Computer Graphics* (2024). DOI: 10.1109/TVCG.2024.3456159 (cit. on p. 110).

[258] Yangjinbo Zhang, Mark Reynolds, Artur Lugmayr, Katarina Damjanov, and Ghulam Mubashar Hassan. "A visual data storytelling framework". In: *Informatics*. Vol. 9. 4. 2022, p. 73. DOI: 10.3390/informatics9040073 (cit. on p. 14).

[259] Fang Zhao, Wolfgang Schnotz, Inga Wagner, and Robert Gaschler. "Eye tracking indicators of reading approaches in text-picture comprehension". In: *Frontline Learning Research* 2.5 (2014), pp. 46–66. DOI: 10.14786/flr.v2i4.98 (cit. on pp. 11, 18).

[260] Qiyu Zhi, Alvitta Ottley, and Ronald Metoyer. "Linking and layout: Exploring the integration of text and visualization in storytelling". In: *Computer Graphics Forum*. Vol. 38. 3. Wiley Online Library. 2019, pp. 675–685. DOI: 10.1111/cgf.13719 (cit. on p. 15).

[261] Tongyu Zhou, Jeff Huang, and Gromit Yeuk-Yin Chan. "Epigraphics: Message-driven infographics authoring". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2024. DOI: 10.1145/3613904.3642172 (cit. on pp. 15, 99, 143).

[262] Hanxiu 'Hazel' Zhu, Shelly Shiying Cheng, and Eugene Wu. "How do captions affect visualization reading?" In: *Workshop on Visualization for Communication (VisComm)*. IEEE. 2022 (cit. on pp. 11, 13, 41, 74, 98, 99).

[263] Donald W Zimmerman and Bruno D Zumbo. "Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks". In: *The Journal of Experimental Education* 62.1 (1993), pp. 75–86. DOI: 10.1080/00220973.1993.9943832 (cit. on p. 29).