Do You See What I See? A Qualitative Study Eliciting High-Level Visualization Comprehension

Ghulam Jilani Quadri
jiquad@cs.unc.edu
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

Arran Zeyu Wang
zeyuwang@cs.unc.edu
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

Zhehao Wang
zhehaow@email.unc.edu
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

Jennifer Adorno
jadorno4@usf.edu
University of South Florida
Tampa, FL, USA

Paul Rosen
paul.rosen@utah.edu
University of Utah
Salt Lake City, UT, USA

Danielle Albers Szafir
danielle.szafir@cs.unc.edu
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA

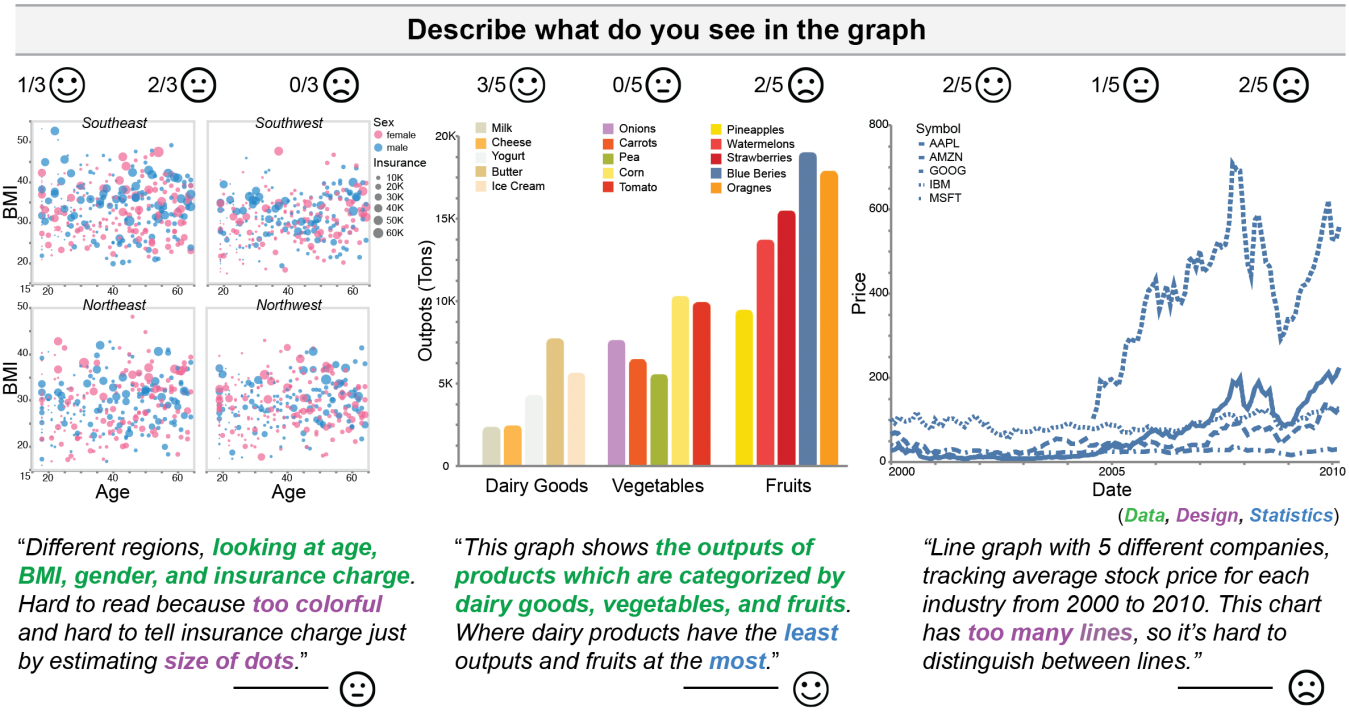


Figure 1: Sample responses participants provided when asked to describe the contents of three graphs: (a) multi-class juxtaposed scatterplot, (b) multi-class juxtaposed bar graph, and (c) multi-class line graph. Happy faces ☺ indicate the number of responses that match the designers' intended communication goals, neutral faces 😐 indicate responses that partially matched, and sad faces ☹ indicate responses that failed to match. The color code in the responses highlight discussions about the salient patterns in *data*, giving *design critiques*, or mentioning *statistical tasks* people observed.

ABSTRACT

Designers often create visualizations to achieve specific high-level analytical or communication goals. These goals require people to naturally extract complex, contextualized, and interconnected

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642813

patterns in data. While limited prior work has studied general high-level interpretation, prevailing perceptual studies of visualization effectiveness primarily focus on isolated, predefined, low-level tasks, such as estimating statistical quantities. This study more holistically explores visualization interpretation to examine the alignment between designers' communicative goals and what their audience sees in a visualization, which we refer to as their *comprehension*. We found that statistics people effectively estimate from visualizations in classical graphical perception studies may differ from the patterns people intuitively comprehend in a visualization. We conducted a qualitative study on three types of visualizations—line graphs, bar graphs, and scatterplots—to investigate the high-level patterns people naturally draw from a visualization. Participants described a series of graphs using natural language and think-aloud protocols. We found that comprehension varies with a range of factors, including graph complexity and data distribution. Specifically, 1) a visualization's stated objective often does not align with people's comprehension, 2) results from traditional experiments may not predict the knowledge people build with a graph, and 3) chart type alone is insufficient to predict the information people extract from a graph. Our study confirms the importance of defining visualization effectiveness from multiple perspectives to assess and inform visualization practices.

CCS CONCEPTS

• **Human-centered computing → Information visualization**; **Visualization theory, concepts and paradigms**; **Empirical studies in visualization**.

KEYWORDS

Visualization, Qualitative evaluation, High-level comprehension, Communicative goals, Insight

ACM Reference Format:
Ghulam Jilani Quadri, Arran Zeyu Wang, Zhehao Wang, Jennifer Adorno, Paul Rosen, and Danielle Albers Szafir. 2024. Do You See What I See? A Qualitative Study Eliciting High-Level Visualization Comprehension. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 26 pages. https://doi.org/10.1145/3613904.3642813

1 INTRODUCTION

Information visualizations help people extract meaningful analytical insights from data. For example, visualizations help people make sense of epidemiological data about COVID-19 [13, 63] or make predictions about natural disasters [25, 78, 95, 108]. Graphical perception experiments measure the effectiveness of visualization designs [101], but the prevailing paradigm for graphical perception studies focuses on low-level tasks [6], typically measuring people's abilities to estimate individual, prespecified statistical quantities. For example, such studies might ask people to *"estimate the correlation between IMDB rating and gross movie sales."* (see Figure 5(b)). The cuing used in these studies (i.e., instructing people to estimate a given statistic) may direct people's attention to statistics or patterns that are otherwise not immediately obvious [139]. *High-level comprehension*, in contrast, describes the overall knowledge a viewer intuitively gains about the data without explicit cuing or guidance. Such comprehension reflects the salient statistics and patterns that emerge organically from a particular combination of data and design. While cued, low-level tasks tell if people *can* perform a given task, understanding the data people comprehend in a visualization informs designers as to whether people *will* perform that task.

We more holistically explore visualization interpretation to examine the alignment between designers' communicative goals and what their audience sees in a visualization, which we refer to as their *comprehension*. Understanding people's comprehension helps designers predict whether their objectives are reflected in how people interpret their visualizations [123]. However, existing design guidelines are typically derived from either expert experience or explicitly-cued experiments looking primarily at low-level tasks like correlation or value estimation [71]. For example, the classical graphical perception paradigm found in experiments like Cleveland & McGill [26] measures visualization effectiveness as the ability to estimate specific statistical quantities. While people may be able to efficiently complete a task when cued, that pattern may not stand out when the same people encounter a similar visualization in the wild. Broadening visualization evaluation beyond low-level tasks can offer valuable insights into their effectiveness, considering factors like knowledge acquisition or key messages [16]. For example, past work has assessed people's understanding of a graph using six levels of knowledge acquisition [16], offered frameworks for formulating communication intents as learning objectives [1], and established that interpreting graphs requires more than just visual properties of visualizations [120]. We draw on this tradition [1, 16] to examine a complementary perspective to classical graphical perception paradigms, focusing on how capturing the patterns visualizations intuitively communicate in data aligns with people's expectations about data communication from design practice and past experiments.

We conducted a qualitative study to investigate the high-level patterns people naturally see when they encounter a visualization without a guiding task. This study is a preliminary investigation that aims to reconcile what experiments and guidelines say graphs are "good" at and what people comprehend in visualized data. This study provides an alternative lens on the classical graphical perception paradigm by approaching the same concept (i.e., what statistical estimates do visualizations afford) from a bottom-up perspective, emphasizing statistical concepts in comprehension (i.e., what estimates occur without any statistical task framing) and using the graph's original intention as a lens to connect to target tasks. People described the patterns they saw and questions they could address in variations of three common graph types: scatterplots, bar charts, and line graphs (see Figure 1). Our stimuli reconstructed visualizations from popular media sources to reflect best practices across a number of design dimensions (see Section 3.2.2). We drew inspiration for our graphs from extensively studied visualization types, commonly encountered graphs in daily life and real-world statistical datasets. Verbal and textual responses from participants were coded using axial coding to extract patterns with respect to their *alignment with stated objectives* (as inferred from each source graph's accompanying text) and to identify salient properties of the *data*, *design*, and *statistical quantities* that shaped people's comprehension.

The patterns people reported seeing in a visualization failed to fully match stated intentions in 59% of tested visualizations. We found that *a visualization's communication goals do not always align with the knowledge people draw from a graph* even when following best practices (Section 4.1). People's interpretations of a given visualization varied with both the features of the visualization itself (e.g., the number of subgraphs, amount of data encoded, labels and units of measurements, visual complexity) and people's individual backgrounds. While chart type was a strong predictor of graph comprehension, *chart type was not sufficient to predict the information people extract from a graph* (Section 4.3). Guidelines built on insights from traditional graphical perception experiments did not fully capture people's high-level comprehension (Section 4.2).

These findings confirm prior theoretical and empirical observations in communicative visualizations [1, 58] and visualization sensemaking [16, 36, 120] calling for more comprehensive approaches to visualization evaluation. We cannot fully understand how people derive insights from graphs exclusively by experiments using cued tasks. This confirmation reinforces the need for a range of diverse methodological paradigms in visualization evaluation. We need to simultaneously understand the precision and salience of different analytical tasks in visualization design. The guidelines generated by combining such top-down (i.e., low-level statistical studies) and bottom-up (i.e., high-level comprehension and other forms of cognitive understanding) aspects of visualization interpretation can help designers optimize visualizations to rapidly and efficiently communicate a range of target patterns.

Contributions. Our primary contributions are: 1) a study eliciting the properties people intuitively see in different visualization designs, 2) a preliminary analysis of how high-level visual comprehension aligns with or contradicts design guidelines from isolated low-level task studies, and 3) insight into how data type, complexity, composition, and design influence the patterns people extract from visualizations. Our results indicate a need for general design guidelines that better consider the knowledge people naturally extract from a visualization

2 BACKGROUND

Understanding what people see in visualizations is critical for helping designers create effective visualizations: they can predict whether their intended goals are actually reflected in the audience's interpretation [123]. Established guidelines for creating effective visualizations originate from experiences or are generalized from studies where individuals are explicitly directed to seek specific patterns or statistics within the visualizations. However, real-world scenarios do not typically provide cues or guidance for interpreting graphs encountered in the wild. To inform our study, we draw on past work in understanding and characterizing insight and measuring graphical perception.

2.1 Visualization Tasks & Insight

A well-established maxim in visualization states that "the purpose of visualization is insight" [19]. In visualization, insight defines the knowledge people obtain from data and serves as a unit of discovery. Visualization evaluation often aims to determine to what degree visualizations help people develop insight into their data [93, 112, 113]. However, characterizing insight is complex and requires more than estimating a single statistic. North describes insight along five separate dimensions: complex, deep, qualitative, unexpected, and relevant [89]. More recent definitions integrate ideas from cognition and neuroscience to clarify that the definition of insight should be broader in visualization and analytics [23] or integrate heuristics to characterize the broader value of visualization [134]. While past works vary in how they measure insight, they all agree that insight is nuanced, complex, and key to effective visualization.

Insight emerges as people use visualizations to conduct a series of *tasks*, binding together patterns and statistics to make sense of data [92, 109]. Definitions of visualization tasks characterize the units by which people analyze data to build insights [6, 14]. For example, Brehmer and Munzner [14] provide a hierarchical typology of tasks, exploring how smaller actions combine to complete larger goals. Conventional visualization guidelines and design processes emphasize smaller, focused tasks (often the *what* in Brehmer and Munzner's typology). However, designs constructed using these guidelines may not uniformly support each constituent task: design guidelines may conflict, or attributes of visualization may cause certain patterns to be more salient than others [69]. Adar & Lee explored the perspective of the fundamental mismatch between designers' intended communication goals and the language they employ, approaching it through the lens of communicative visualization as a learning problem [1]. Our work builds on this idea by investigating how the insights people build through visualizations align with those the visualization intends to communicate.

2.2 Graphical Perception & Comprehension

Many design guidelines are grounded in *graphical perception*, the study of how people perceive specific information in visualizations [26, 123]. These studies typically focus on task effectiveness for specific, readily quantified tasks, such as assessing correlations across a range of designs [51, 64, 102], modeling how precisely people estimate statistics across visual channels [66, 116, 122], and measuring how different visualization techniques support a range of tasks [5, 29, 111]. See Quadri and Rosen [101] for a survey.

While significant research has evaluated how well people can extract specific statistical quantities, these studies explicitly cue people as to what patterns to look for, asking participants to answer direct questions such as, "*identify the highest stock price in last decade*" (see Figure 1). We draw on past work in graphical perception to understand the relationship between design guidelines and the natural patterns people comprehend in visualizations (i.e., the tasks they perform without any cues).

We investigate this relationship using scatterplots, bar graphs, and line graphs, which are the most widely used and highly studied graph types according to recent surveys [101]. Their effectiveness has been explored across a range of tasks, including judging values [41, 55, 126] and exploring clusters [62, 99, 100] using scatterplots, finding extremum [118, 133] and comparison [124, 137] with bar graphs, and estimating trends [5, 9, 31] and value retrieval [52, 110] with line graphs.

While these studies inform us of what common visualizations can do, they suffer from two key limitations: 1) they present simpler visualizations compared to most real-world applications, and

2) they provide people with a specific, well-defined goal, engaging a potentially different set of perceptual mechanisms than when encountering a visualization in the wild under less constrained conditions. Prior works demonstrated how the comparison between visualization could be difficult for people when graphs are complex, have more items, and larger size [39]. Our study aims to include complex graphs that resonate with real-world examples. Furthermore, people's interpretation is influenced by additional factors such as rhetorical techniques from other fields [58] or individual background [15, 49, 91]. Additionally, assessing the communicative goal of visualization requires more than just low-level tasks [1], and relying solely on visual properties may not be sufficient to predict its interpretation [120]. We can compare open-ended responses summarizing people's high-level comprehension against design guidelines reflected by expressed design intents and results from these studies to begin to understand people's graph comprehension and scaffold future investigations.

Guidelines from cued studies may not be sufficient to guide design in part because there are two opposing ways people might attend to information in visualizations: top-down and bottom-up [28, 38, 48]. Top-down attentional processes are correlated with goal-driven attentional control (i.e., people attend to marks based on a given task or objective) [132] whereas bottom-up processes correlate with stimulus-driven attentional control (i.e., people attend to marks based on the visual features of the marks) [8]. While most graphical perception studies reflect goal-driven processes, we argue that, in many cases, visualization comprehension relies on stimulus-driven processes (e.g., when exploring unfamiliar data or encountering a narrative visualization without specific guiding text). Zacks & Tversky [141] found that when asked to simply describe a graph, people attend to different features in different graphs and that these differences lead to different conclusions about data. Their results suggest that people attend to information in visualizations in two different ways due to top-down and bottom-up attentional processes. However, their explorations focused on simple, two-point graphs. We extend these ideas to understand stimulus-driven comprehension in more conventional visualizations.

We can understand specific design guidelines by comparing the patterns people comprehend against a visualization's stated communication objectives. A visualization's abilities to achieve these objectives are likely mediated by a variety of factors such as visual encoding [101] and data distributions [66]. For example, people may attend to different patterns based on past knowledge [115, 139, 141]. People's inferences about a graph may be swayed by social cues [67], rhetorical framing [58], audience background [15, 46, 91], affect [73], additional text [120], or language describing intent [1]. Design guidelines grounded solely in graphical perception studies may fail to account for such factors. For example, Stokes et al. [120] leverage open-ended description tasks on line charts containing varying amounts of text, ranging from no text to a written paragraph with no visuals, and found that charts with text or statistical components led to more accurate takeaways than charts with elemental or encoded texts.

Visualization design guidelines for general audiences should consider the characteristics of the potential audience, including diverse backgrounds [49, 91], graphical literacy [36], and experience levels [15, 46]. Such general communication must be especially cognizant of novices. A 'novice' is a person who is inexperienced or new to *visualization* or *visual analytics*. Past work on novices frequently focuses on more qualitative approaches to understanding visualization use, similar to our approach, to determine how people with limited experience may approach visualization [46, 53]. For example, Grammel et al.[46] investigated how novices create visualizations using commercial visualization software, focusing on the author's perspective rather than the data consumer's to describe barriers novices face in the data exploration process. Burns et al. [15, 16] showed that even when considering novices as a population, visualizations often fail to adequately consider their role in evaluation. While our work does not focus on novices, we draw on a similar body of methods to assess the relationship between visualization design and high-level graph comprehension, focusing on the conclusions people draw from data in practice and contextualizing those against design intentions reflecting conventional design guidelines.

3 METHODOLOGY

We conducted a qualitative experiment to characterize the patterns and statistics people comprehend in common visualizations when they encounter a visualization without a guiding task. Our experiment investigates two primary research questions:

[RQ1] *Do the patterns people see in visualizations match the visualization's stated objective?*

[RQ2] *What types of patterns do people naturally see in common visualization designs?*

We address these research questions in an empirical study using a combination of think-aloud and written free-response methods. The study asked participants to use various methods to describe the content of a sequence of graphs. These graphs were reconstructions of designs from the popular media using available datasets (see Appendix B), where their stated communicative intention was extracted from the text of their accompanying article.

We analyzed responses using axial coding and thematic analysis to identify preliminary patterns in the information that people intuit from visualizations. Our study characterized this intuitive data comprehension across three design dimensions: graph type (scatterplot, bar graph, and line graph), data type (single-class versus multi-class), and graph composition (juxtaposed versus non-juxtaposed).

3.1 Experimental Task

We designed two tasks to investigate whether graphs communicate the information they were designed to communicate (**RQ1**) and to elicit the specific types of patterns or statistics people notice in different visualization types (**RQ2**). To tune our questions to ensure that they elicit the appropriate levels of comprehension, minimize potential bias from question wording, and establish a preliminary codebook, we conducted a pilot study with ten participants, asking them to describe five visualizations from the New York Times (c.f., Appendix A). Our formal study asked participants to use a graph to accomplish the following:

[Description] *Describe what do you see in the graph.*

[Question] *What questions would you be able to answer from this graph?*

Approaching the task from two directions (as a description and a set of questions) reflects best practices in survey design to help overcome potential limitations from question framing [72]. We used the **Description** task to elicit broader, contextualized observations such as insights [89]. The **Question** task implies tasks for which a "correct" answer could be inferred from the graph. This task is intended to prompt people to provide specific analytical tasks (e.g., estimate correlation) that a given visualization enables without cuing participants to individual statistics.

3.2 Study Design

Visualizations in real-world scenarios encompass a vast array of designs, each potentially conveying distinct sets of statistical information. Building on the insights from the pilot study (c.f., Appendix A), our primary aim was to systematically design the stimuli in our study to provide consistency between stimuli while still reflecting the core design approaches in the original visualizations. To gain deeper insights into individuals' high-level comprehension, we need to understand how the interplay of design elements influences the information naturally communicated by diverse graphical representations. The influence of elements such as color, size, or mark shape may not be (and arguably often is not) separable in practice [66, 74, 122]. Our approach instead provides a preliminary investigation of differences across aggregate design parameters in part to inform future studies that systematically vary individual components to understand their influence on comprehension. We chose to conduct the study on multiple visualization types—line graphs, bar charts, and scatterplots—rather than focusing on just one type to gain a broad preliminary understanding of how people interpret graphs, given the complex interactions we anticipate between design factors. These results help identify meaningful future investigations of specific variables in shifting comprehension.

While there exists an abundance of empirical studies on commonly used graphs, there is a lack of guidance on high-level comprehension. Moreover, earlier research has indicated that low-level tasks may not sufficiently address the true communication goals of visualization [1]. In this work, we focused on three graph types designed using best practices in visualization to communicate statistical information effectively. We aimed to draw from a diverse pool of visual designs to ensure a broad examination of the space grounded in real-world practices. We selected a range of frequently encountered and thoroughly studied graphs to provide preliminary comparisons across both heuristic and empirically grounded guidelines. We divided these factors into three categories describing the visualization type, data type (single versus multiclass), and composition (juxtaposed or non-juxtaposed; Figure 2). This approach allowed us to explore the diversity of high-level comprehension across various visualization idioms and approaches. We draw on expert-designed graphs from the popular media to help ensure best practices were enforced and provide context for extracting explicit design intentions.

3.2.1 Independent Variables. Our study investigates visualizations that vary in *visualization type, data type*, and *graph composition* (see Figure 2). These three variables reflect general driving factors in visualization design: the core task people want to accomplish typically dictates the *visualization type* used; the problem space dictates the *data types*; and the combination of tasks and data (i.e., number and type of important attributes) determines *composition* [84].

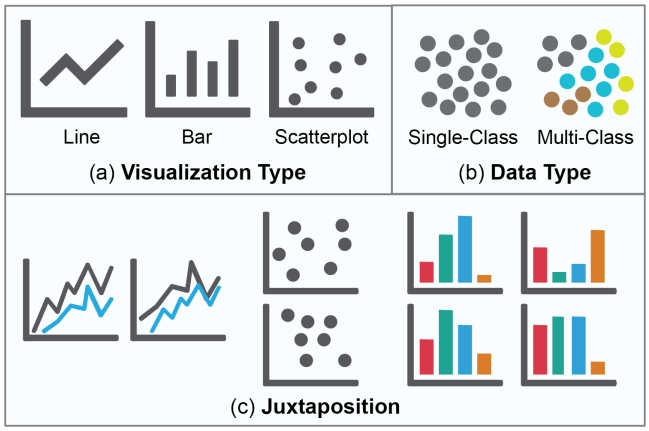


Figure 2: The design dimensions (see section 3.2.1) used in our study: (a) visualization types, (b) data types, and (c) composition type (juxtaposed).

Visualization Type: We selected and tested *scatterplots, line graphs,* and *bar graphs*, as illustrated in Figure 2 and Figure 3. These visualizations commonly appear in the media, and most participants will be familiar with their basic structure [12]. They are also among the most studied in classical graphical perception studies [101], providing a robust baseline for comparison with results from cued experiments.

Data Type: We selected datasets (see Appendix B) to plot graphs with either single-class (no categorical data) or multi-class (data can be decomposed into categories) data (c.f., Figure 2 and Figure 7). Considering data type allowed us to look at graphs across both categorical and continuous encodings as well as to explore responses for more complex visualizations, which often communicate multiple points and have more complex communicative goals, revealing patterns in the ways different design choices interact with one another.

Composition Type: Previous research has indicated that visual tasks often require comparing data across groups or dimensions [39, 40]. These visualizations often require interpreting data across multiple charts and synthesizing those interpretations to understand the collective message. We considered such graph compositions through either juxtaposed (side-by-side and up-down, see Figure 2, or non-juxtaposed [60] designs. Single-class non-juxtaposed graphs consisted of a single graph, while single-class, juxtaposed graphs had multiple graphs, each showing a single class of data. Multi-class juxtaposed graphs had multiple graphs, each showing multiple classes of data, while multi-class non-juxtaposed graphs showed multiple classes of data on the same axes (i.e., superimposed). See Figure 3 for examples.

3.2.2 Stimuli. Visualizations in the wild encompass a wide range of designs; however, different visualizations communicate different

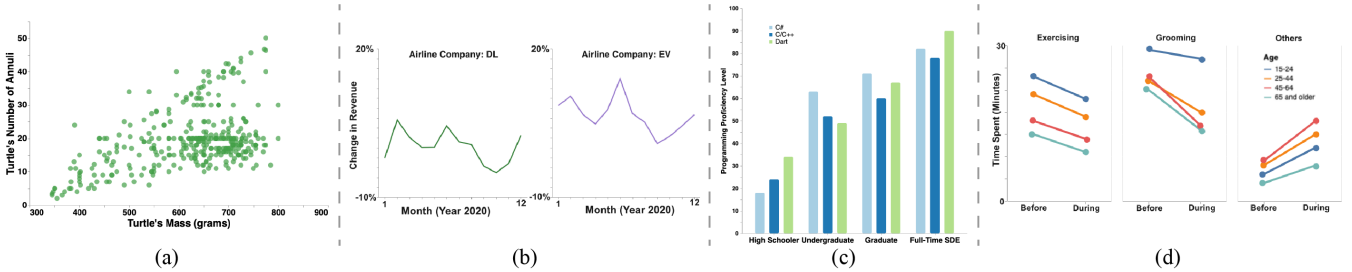


Figure 3: The stimuli samples illustrating design dimensions (see Figure 2 and section 3.2.1) used in our study, where visualizations in (a) use the *Turtle* dataset [128], dimensions: *scatterplot, single-class, non-juxtaposed.* (b) A graph of *Airline* dataset [3], dimensions: *line graph, single-class, juxtaposed.* (c) A graph of *ProgProficiency* dataset [97], dimensions: *bar graph, multi-class, non-juxtaposed.* (d) A graph of *Activity-covid* dataset [77], dimensions: *line graph, multi-class, juxtaposed.*

sets of statistics. As we lack guidance on which visualization design elements shift the information a graph naturally communicates, we needed to draw from a set of designs that represents sufficient diversity to avoid confounds from isolated artifacts (e.g., outliers in a dataset, poor color choices, etc.). We drew the inspiration from graphs in data journalism (i.e., the New York Times or government websites). These graphs (a) allowed us to explicitly understand their intended goals by mining target statistics from the accompanying text and (b) reflect known best practices for communicating this data. We collected a set of 60 graphs from these sources to reproduce (5 per unique setting of our independent variables). Our reproductions used the general design schema from the original graphics, but we replaced the data and associated labels and values with data from 42 stimulus datasets (selected at random) to remove potential interpretive bias associated with data semantics [70, 139]. To maintain ecological validity, we used real-world datasets employed in previous visualization studies. The datasets were drawn from various sources, including past visualization, HCI research, and real-world datasets.

We tried to mirror the design choices from the original graph as faithfully as possible while minimizing potential framing effects when adapting visualizations to the target datasets. Some adaptation was necessary to provide consistency across stimuli to avoid distracting participants (e.g., we used a consistent axis design). We mapped the data using shapes, colors, and data encoding channels (both continuous and categorical) from the original graphics. We maintained the aspect ratio, layout, and legend position. In the reconstruction process, we removed labels, titles, or captions (see Figure 4). As we wanted to understand people's visual comprehension of graphs based on the visualization design and data alone, framing effects from titles [70], captions [24], and labels [70], would have introduced a significant confound by cuing particular interpretations. For example, they may have provided information that guides readers towards a particular conclusion [106].

We selected five graphs for every combination of visualization type (3) × data type (2) × composition (2), resulting in 60 stimuli. Figure 3 and 2 show examples of our designed stimuli where a different dataset is used for the same design intent. See the OSF Supplement for additional details and the full set of tested images and data sources.

3.2.3 Inferring Stated Objectives. The intention behind each visualization in our corpus was inferred from the text accompanying each original source visualization. We first extracted the tasks associated with each source. We then mapped the visualization to the stimulus datasets and manually verified that the tasks remained appropriate for the resulting image (e.g., the intended tasks were still salient given the new data distribution and revised data semantics). For instance, in Figure 5 (a), the scatterplot was designed to visually represent sunny weather conditions. It achieved this by plotting the maximum temperature recorded on each day within a particular calendar year. The visualization's primary goal is to communicate maximum temperature patterns associated with sunny days and/or the distribution of temperature over time. The OSF Supplement contains the stated objectives for all 60 tested stimuli.

3.2.4 Procedure. Participants began by providing informed consent. The experimenter then explained the basic task (to describe each visualization and questions the visualization could address) and directed participants to a web application where they completed the formal trials. We did not include any tutorial examples to avoid potential bias from priming.

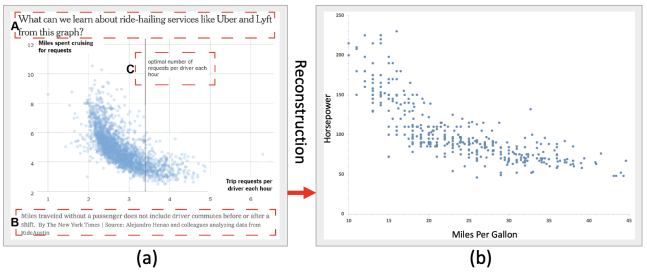


Figure 4: Example figure in (a) illustrating a single-class non-juxtaposed scatterplot from New York Times showing negative correlations between two variables. To reconstruct this graph in (b), we removed additional text and labels (A & B in (a)) and annotation (C in (a)) and plotted using another dataset [7]. The stated objective is extracted from the text accompanied by the article.

Participants completed the two target tasks (section 3.1) for a sequence of twelve visualizations, providing their responses to each question in a textbox. Each participant saw one chart from each combination of visualization type, data type, and composition—12 total visualizations. Visualizations were selected at random from the corresponding rendered stimuli for each combination of independent variables and presented serially in a random order. We encouraged participants to verbalize their reasoning processes to collect additional insight into strategies, points of confusion, and salient information that was not reported in the open-ended response text. To ensure task understanding and broaden our dataset, we instructed participants to be as thorough as possible in their responses, both in the textbox and verbally, and allow participants as much time as they wanted to complete the study. The experimenter was present in the room to address any questions participants had at any point during the study. They finished the study by providing basic demographic information.

Participants completed the study in 30-45 minutes. We recorded all participant study sessions for later transcription and analysis.

3.3 Participant Recruitment

We recruited 24 participants (18-56 years of age; 18 female, 6 male) from a University campus who participated either in-person (21 people) or via Zoom (3 people). Five identified as working professionals, and nineteen as students. We did not intentionally vary or recruit participants with diverse levels of visualization literacy and novice expertise. More details on the participants' demographics are in Appendix B. Potentially identifying information has been redacted from the quotes presented in this paper to protect participant anonymity in accordance with our IRB protocol.

3.4 Data Analysis

We used axial coding to analyze the natural language inputs from both the text responses and transcribed verbal utterances from the think-aloud protocol using the codebook derived in our pilot study (c.f., Appendix B). Two coders each coded responses for 14 of the 24 total participants. They discussed the codes to refine the codebook after four transcripts. A summary of the codes is available in Table 1 and an illustration in Appendix B. They both coded two overlapping sessions to verify interrater reliability. These sessions had a high overall agreement between the coders ($\kappa = 0.81$). In total, we have 288 responses collected from 24 participants. We coded each response for its *comprehension match*, *statistical tasks*, *design critiques*, and *data critiques*.

Comprehension Match quantifies the extent to which a participant's responses align with the stated objectives, as inferred from the accompanying text of the source graph (Section 3.2.3) and the participant's natural language descriptions and verbal responses. This measure is contingent on the degree of granularity in the statistical information discerned by participants from the presented graphs. This metric helps evaluate whether participants accurately perceive and convey the statistical information or insights that the visualization was designed to communicate. The alignment of the reported tasks and stated objectives provides insight into how intuitively the design communicates target information. Codes of the intentions and responses focused on statistical concepts rather than semantic concepts to give consistency across visualizations by translating takeaway messages laden with dataset-specific semantics to objectives that hold across datasets (e.g., "proficiency increases" becomes "increasing trend"). During data coding, we focused on aligning abstract statistical concepts and on word-to-word matching.

Matches aligned at one of four thresholds:

(1) Complete Match (CM): Specific statistics and patterns in the response matched the stated objective. A complete match indicates a strong alignment between participant responses and design objectives. In such cases, participants have effectively comprehended the key statistical information and insights intended by the original graph.
(2) General Match (GM): Participant's responses indicated the general knowledge they took from the graph aligned with the design's stated objective, but not specific statistics or patterns. While there might be some minor differences or details that participants did not capture precisely, the core message and insights align well. For example, a Figure 5(a) intends to *document the pattern in the days where sunny weather occurs and the maximum temperature of each recorded day in a city from January to May and from August to December in 2012.*[P03] response matched with intent– *"Changes in temperature across the year and particular weather characteristics of each day (e.g., sun). The temperature starts low in January, goes up until August, and goes down from there."*.
(3) Partial Match (PM): Participant's responses partially matched the stated objective. They missed main statistical quantities or general knowledge about the data. For example, [P07] responded– *"the graphs describe the temperature of different months in 2012 under the different weather"* on Figure 5(a) and missed statistical quantities on how the temperature follows a pattern- starts low (around 50 F) in January, goes up (above 85 F) until August, and goes down September onwards.
(4) No Match (NM): Participant's responses do not match with the stated objective and missed critical information entirely. Participants have not effectively comprehended or articulated the information the graph was designed to convey. A no-match implies that participants may have misunderstood or missed intended information.

Statistical Tasks encode the patterns and statistics participants provide in their responses. We employed Amar et al.'s low-level task taxonomy [6] to code these tasks as it provides sufficient coverage of the tasks in our stated objectives and participant responses. Codes included *Retrieve Value*, *Filter*, *Compute Derived Value*, *Find Extremum*, *Sort*, *Determine Range*, *Characterize Distribution*, *Find Anomalies*, *Cluster*, *Correlate*, and *Compare*.

Design Critiques identify feedback provided by participants about how a particular design choice assisted or hindered their comprehension of a visualization. We did not ask participants to provide critiques directly, but documented critiques that arose naturally in the responses or think-aloud.

Data Critiques identify places where the data itself inhibited high-level visual comprehension. This included comments on missing context, needing additional dimensions, or failure to understand the data directly (e.g., jargon or acronyms). While again we did not

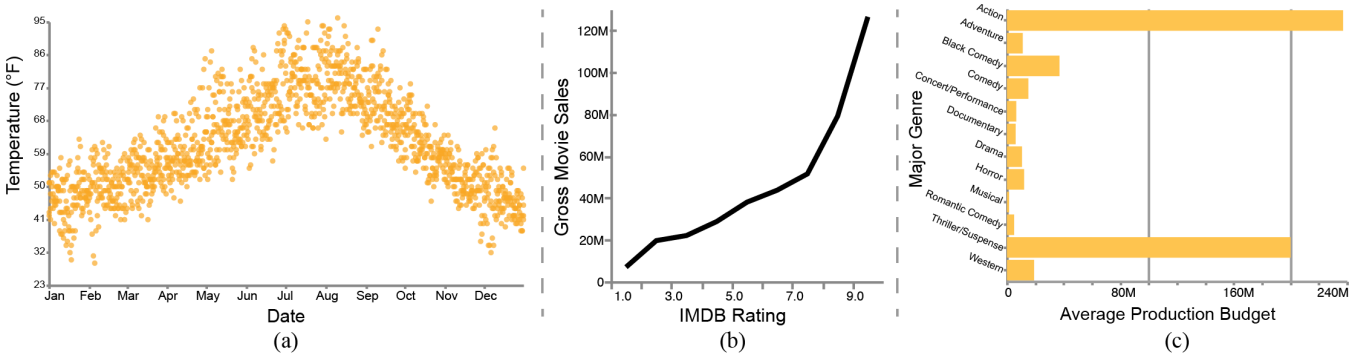


Figure 5: Examples of Single-Class stimuli in our study. (a) is a Single-Class scatterplot with *Sunny day* dataset [42], (b) is a Single-Class line graph with *IMDB* dataset [59], and (c) is a Single-Class bar graph with *Budget* dataset [80].

prompt for these values, they arose in many participants' responses and provided additional insight into gaps in data comprehension and opportunities for improved visualization design.

4 THEMATIC ANALYSIS AND KEY FINDINGS

A key motivation of this work is to characterize visualization design as a function of the patterns people perceive unprompted from a graph. In other words, do people see the information a visualization is designed to communicate? We analyzed participants' response data using thematic analysis and found three overarching themes in order to understand the need for visual comprehension as a metric (see Figure 6 for the roadmap of our thematic analysis):

Theme 1: **Intention does not align with comprehension.** Stated communication goals did not always align with the knowledge people drew from a visualization. Only 41% responses completely matched with the visualization's stated goal even though our study's graphs followed the design schema and matched recommendations from prior work. This result echoes calls made in more comprehension-oriented studies (e.g., [1, 16]) to understand better the connection between design guidelines in theory and their efficacy in practice.

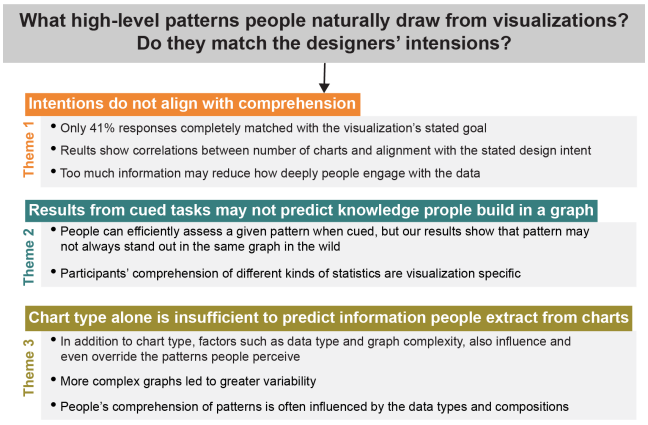


Figure 6: A roadmap of the results from our study. We group our findings into three themes— Theme 1 (see Section 4.1, Theme 2 (see Section 4.2), and Theme 3 (see Section 4.3).

Moreover, the stated designs' intent may not always align with people's comprehension as comprehension is influenced by multiple factors, including visualization design, individuals' background, graphical literacy, and supplemental information.

Theme 2: **Results from cued tasks alone may not predict knowledge people build in a graph.** The prevalent model in visualization research defines visualization effectiveness as the ability of the average user to complete a given task [101]. This assumption has limited ecological validity: people may be able to efficiently assess a given pattern when cued, but that pattern may not stand out in the same graph in the wild. If someone *can* use a given graph to accomplish a directed task, it does not mean that they *will*. Even though prior work showed certain visualization types (e.g., scatterplot, bar, and line graphs) effectively convey various statistics and patterns (e.g., correlation, distribution, cluster, anomalies), less than 50% of participants' responses included such patterns. This observation supports the argument made in past work [1, 71, 89], emphasizing that low-level tasks alone are not sufficient to address the real communication goals of graphs. Our observations suggest guidelines built on insights from cued experiments do not fully capture people's high-level comprehension. Instead, guidelines generated by combining low-level statistical studies and high-level comprehension may help designers optimize visualizations to communicate a range of salient patterns efficiently.

Theme 3: **Chart type alone is not sufficient to predict the information people extract from a visualization** Prior work in visualization effectiveness and corresponding tools like chart choosers often match visualization tasks to chart types [66, 86, 100, 110, 135]. We observed that chart type tended to dictate the patterns people commonly reported (e.g., correlation in a scatterplot, trend in line graphs, and identifying discrete and maxima in bar graphs). However, data type and graph complexity (e.g., composition) also influence, and can even override, the patterns people perceive. We observed many mismatches between participants' responses and the visualizations' stated goals when graphs encoded multi-class data types using juxtaposed compositions (Figure 7). These observations validate past findings that relying solely on chart types or visual properties is insufficient to predict how the communication goal will be interpreted [24, 58, 120].

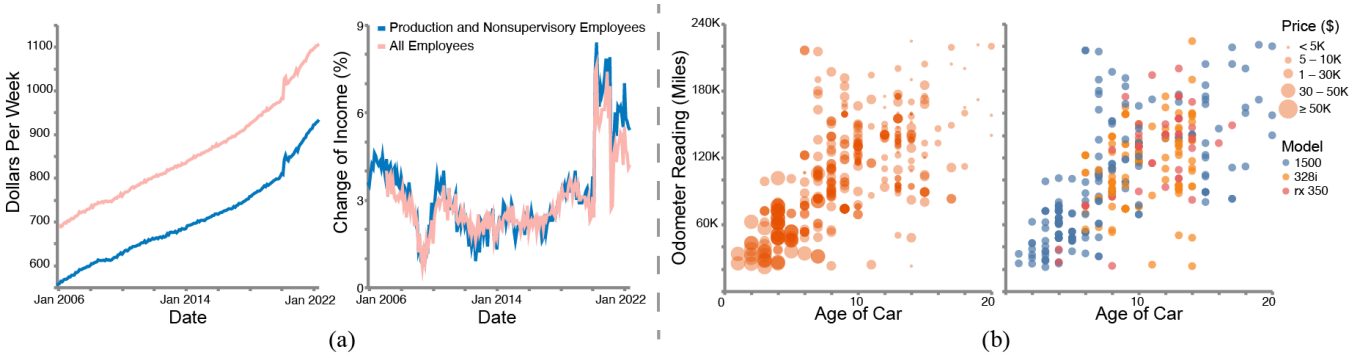


Figure 7: Examples of Multi-Class stimuli in our study. (a) is a Multi-Class Juxtaposed line graph with *Income* dataset [68], and (b) is a Multi-Class Juxtaposed scatterplot with *Car models* dataset [18].

Our thematic analyses also allowed us to investigate various design and data critiques provided by participants (see Section 5). Table 1 summarizes comprehension matches for all 288 collected responses as a function of our independent variables (graph type, data type, and graph composition).

4.1 Theme 1: Intentions Do Not Align with Comprehension

We evaluated the *comprehension match* (see Section 3.4) between the descriptions provided in the participant responses against the stated objectives of the original source graph (see Section 3.2.3). A complete match means the participant's description and questions identify all of the objectives inferred from the original source graph and may include additional information as well. Only 41% (117/288) of responses completely matched the designers' objectives. Additionally, 77% (92/117) of those responses were provided by the 10 participants whose provided complete or general matches for at least 10 of their 12 stimuli (see Appendix B). This discrepancy implies that **an expert-crafted visualization's intended communication goals may not always align with the insights and salient patterns people intuitively extract from that visualization.** The strength of this alignment varied by graph complexity, with multiclass non-juxtaposed line charts (33% of responses were complete matches) and scatterplots (46% complete matches) and multiclass juxtaposed line charts (37%) and scatterplots (33%) seldom matching the target objectives. We also found a correlation between people's comprehension matches and their self-reported familiarity with graphs and background, which we discuss in Section 4.3.4.

The mismatch between reported descriptions and stated intentions signals a need for better guidelines: we assume that professional venues follow best practices for data communication and selected our source visualizations such that they followed this assertion. We can examine participant responses to investigate where discrepancies arise between a visualization's intended message and the knowledge people extract in practice. In the following discussion, we report general patterns but focus specific examples on three graphs—Figure 1, Figure 3, Figure 7, and Figure 5—but additional details are available in Appendix B and OSF Supplement.

The alignment between intention and viewer comprehension is primarily influenced by the visualization design and the complexity of the graph, as summarized in Table 1. People commonly reported specific statistics for certain chart types (e.g., correlation in scatterplots, trend in line graphs, and identifying discrete values and maxima in bar graphs). However, data type and graph composition also influence, and can even override, these patterns (see Section 4.3).

Participants' data interpretation and comprehension varied with the visualization type used. Single-class scatterplot responses aligned with the inferred intention in 57% of cases, while single-class bar graphs (65%) and line graphs (68%) had higher alignment. People tended to quickly identify the peaks and valleys and changes over time in line graphs, maximum values in bar graphs, and associations and correlations in scatterplots.

In other cases, participants' responses significantly deviated from the intended communication goal and focused on what they can read on graphs. For example, for Figure 7(a), which intended to *show positive correlation and compare two types of employees' average weekly incomes and their growth rate over 16 years*, they read the line graph using label and legends, identified 'change' in income, noted that both graphs follow a similar trend, and attended to the highest values. One participant described the visualization as "*an increase in how many dollars are earned per week between earnings of production and non-supervisory employees and earnings of all employees* [P01]". At times, the misalignment resulted from a lack of familiarity: "*I don't know the scatterplot, [I'm] confused with the graph and terminology mentioned in the labels. But I am still understanding the graph about architecture.* [P07]" (see Fig.11 in OSF Supplement).

We found a **correlation between the number of charts and alignment with the stated design intent**. Specifically, fewer charts led to higher alignment. Non-juxtaposed charts tended to support higher alignment than other kinds of composite charts—scatterplot (23/48), line graph (20/48), and bar graph (19/48). Additionally, single-class juxtaposed compositions for both line (12/24) and bar graphs (13/24) achieved their intended goals better than composite graphs. For example, in the single-class, non-juxtaposed

scatterplot in Figure 3(a), participants identified the intended positive correlation in 90% of cases compared to 33% in single-class juxtaposed scatterplots.

People frequently failed to find any relevant patterns in visualizations with multiple categories represented using multiple encodings (e.g., Figure 1(a), 2/3, and Figure 7(b), 2/5) where position, size, and color encode different data aspects, supporting the findings from Gleicher [39] that tasks grow difficult with increases in complexity. One participant mentioned—"*I could understand bar graphs and simple line graphs well but not scatterplots or graphs involving multiple categories.* [P16]" We observed more than 30% of responses failed to identify any alignment with the stated objectives (37% in scatterplots, 32% in line graphs, and 34% in bar graphs).

Several misalignments arose when participants found the graphs difficult to interpret. These instances largely occurred in complex visualizations, most notably juxtaposed graphs and multiclass data (juxtaposed scatterplots, 16/48 responses failed to match; juxtaposed line graphs, 15/48; juxtaposed bar graphs, 14/48; multiclass line graphs, 9/24; and multiclass scatterplots, 4/24). One participant noted that "*It is easier to read bar charts and simple line graphs but not scatterplots or graphs involving multiple categories* [P13]".

In these cases, people tended to describe graphs using surface-level information rather than digging into statistical patterns in the data. For example, participants read and described what a graph represents through legends and axis labels (scatterplots 25/96, line graphs 29/96, bar graphs 24/96). Figure 1(a) is intended to show the distribution between patient's age and their BMI, support comparisons across in four different regions, and build relations between the distributions of four variables—age, BMI, region and insurance charges—across regions. Two of the three participants who interpreted this visualization focused on surface-level data, describing the data in the graph (BMI based on the region, age, and gender) but not mentioning any patterns in the data. One described it as "*looking at patient age, BMI and comparing gender and geographic location.* [P24]" while the other said " *This graph shows client BMI vs. client age in diff regions of the US while differentiating between males and females.* [P05]" Only one out of three participants' responses matched the stated intention described in the source graph—"*The graph seems to be depicting BMI for females and males based on age and geographic location. There does not seem to be a clear trend based on the graphics depicted. BMI, age, and geographic location are not clearly correlated.* [P24]" More complex messages, such as the absence of correlation, can be difficult to communicate. However, our results indicate **a trade-off between complexity and interpretation: too much information may reduce how deeply people engage with the data.**

4.2 Theme 2: Results From Cued Tasks Alone May Not Predict Knowledge People Build in a Graph.

Visualization guidelines suggest what tasks people perform well with a given visualization. For example, scatterplots support clustering, characterizing distributions, and correlation; line graphs convey trends and temporal patterns; and bar charts communicate distributions, discrete value identification, and extremum (see Quadri & Rosen [101] for a survey). However, these assertions about

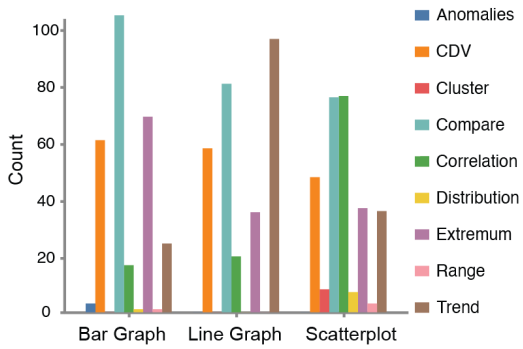


Figure 8: Summary of participants' responses on statistical quantities and patterns in their comprehension. Some statistics and patterns are specific to a particular graph type, for example, extremum on bar graphs, trend on line graphs, and correlation on scatterplots.

task effectiveness correspond to the ability of the average user to complete a given task when prompted. This assumption has limited ecological validity: **people can efficiently assess a given pattern when cued, but our results show that pattern may not always stand out in the same graph in the wild**.

Our analysis showed that just because someone *can* use a given graph to accomplish a directed task does not mean that they *will* without guidance. Measuring performance using abstract statistical quantities may not tell designers whether their visualizations will likely achieve their goals. For example, people did not identify *correlation*, *data distribution*, or *outliers* in the line graph in Figure 7(a) even though line graphs effectively communicate these statistics [5, 50]. This section explores how our results confirm or override previous graphical perception experiments. Figure 8 summarizes our results.

Prior work demonstrated that people are good at estimating distributions [22, 110], clusters [100, 114], and correlations [50, 102] in scatterplots. In contrast to past work, participants did not mention correlations in their scatterplot descriptions for 65% of scatterplots (reported 75 times on 216 tested graphs), clusters for 84% (8 reported of 48), and distributions in 97% (7 reported of 240) of responses. In several cases, participants focused on the meanings of individual points in a scatterplot rather than the relations between them: "*I am seeing a dot graph showing different models of cars with different price points measured by their odometer reading at a specific age. I can tell how many miles a specific car has based on age.* [P15]"; or "*key telling the price of the car and the model of the car. The age of the car and the odometer reading in miles.* [P23]" for Figure 7(b). Further, people did not discuss statistics such as extremum or computing derived values.

People frequently described *trends* in line graphs (96 times); however, they seldom performed other tasks that line graphs are known to be effective for, such as characterizing distributions (no responses) [5, 43], estimating correlation (20 responses)[51, 64], and detecting anomalies (no responses) [5, 110]. Multi-line graphs further failed to communicate distribution (none), clusters (none), and correlation (3 responses). Participants tended to describe each

Table 1: Summary of participant responses alignment with designer's intention per chart type. SC: single-class; MC: multi-class; SC-J: single-class juxtaposed/overlayed; MC-J: multi-class juxtaposed/overlayed. To explore more on comprehension match, see per participant analysis in Appendix B.

Response Coding	Scatterplot					Line Graph					Bar Graph					All
	SC	MC	SC-J	MC-J	All	SC	MC	SC-J	MC-J	All	SC	MC	SC-J	MC-J	All	
Complete Match	12	11	8	8	39	12	8	11	9	40	11	8	12	7	38	117
General Match	3	3	1	2	9	3	0	1	1	5	3	5	1	3	12	26
Partial Match	4	6	8	5	23	4	7	3	8	22	4	7	6	5	22	67
No Match	5	4	7	9	25	5	9	9	6	29	6	4	5	9	24	78
All	24	24	24	24	96	24	24	24	24	96	24	24	24	24	96	288

individual graph rather than to draw relations between them—"*There are two graphs. From the right one, I see their income change over the years. I am seeing two line graphs showing the average weekly earnings of 2 categories of employees over the period of Jan 2006 to Jan 2022 and their change of income over the previous year.* [P11]" on Figure 7 (a). This lack of synthesis suggests that people may need to be explicitly cued as to how to connect data or findings across visualizations. For example Figure 7(a) aimed to communicate the correlation between earnings of supervisory and non-supervisory employees with the length of employment and their raises over time; however, participants predominantly described the graph according to local trends—"*...change in the employees wage with time*" [P02, P08, P09, P11]—overlooking the comparisons the chart composition afforded.

People's responses generally confirmed prior findings for bar graphs on discrete value identification (in 60 of the responses) [110, 133, 138], extremum (67), and trend (25) [90, 118]. However, participants seldom mentioned distribution (3) and anomalies (5), despite their known effectiveness [110, 118]. Though we saw better correspondence between statistical tasks and known perceptual results in bar graphs, several cases did not fully align with previous findings [110, 118].

Participants' *comprehension of different kinds of statistics are visualization-specific*, see Figure 8. As expected "trend" is heavily associated with line graphs (in 96 of the responses). Data also seemed to play a role: participants focused on extremums at longer bar heights for bar graphs and spikes in line graphs. High correlation (> 0.9) in scatterplots and line graphs and notably valleys and low peaks in line graphs attracted viewers' attention, resulting in higher association with corresponding statistical tasks.

4.3 Theme 3: Chart Type Alone is Not Sufficient to Predict the Information People Extract From a Visualization

Prior work in visualization effectiveness and corresponding tools like chart choosers often match visualization tasks to chart type [101]. As discussed in Section 4.1, chart type dictated the patterns people commonly reported (e.g., correlation in a scatterplot, trend in line graphs, and identifying discrete values and maxima in bar graphs). However, other factors, such as data type and graph complexity (e.g., composition), also influence, and even override, the patterns people perceive.

We observed several effects of design dimensions beyond chart type on visualization interpretation, specifically grouped around chart complexity (as modeled by data type and composition), scaffolding, and individual differences amongst participants. More complex graphs (e.g., multi-class data and juxtaposed sub-graphs) led to greater variability in the people's identified patterns and lower overall alignment with the visualization's stated objective. Added chart scaffolding (e.g., annotations or legends) led to higher alignment, while data anomalies (e.g., outliers) tended to change the patterns people reported from a given graph type.

4.3.1 Data Type & Distribution. People's comprehension of patterns was often influenced by data type—single- and multi-class—and compositions. People more readily identified patterns in single-class graphs as compared to multi-class data. For example, people frequently identified intended statistics in Figure 5(a) (distribution: 4/4), Figure 5(b) (correlation: 2/3), and Figure 5(c) (correlation: 4/5) (see Appendix B for the full list). People's alignment with a visualization's stated goals was significantly lower in multi-line graphs (partial + no-match $> 66\%$ on multi-class non-juxtaposed and 58% on multi-class juxtaposed), scatterplots with multiple data dimensions (partial + no-match $> 41\%$ on multi-class non-juxtaposed and 58% on multi-class juxtaposed), and multi-class bar graphs (partial + no-match $> 46\%$ on multi-class non-juxtaposed and 58% on multi-class juxtaposed).

The observations provided in the descriptions of graphs with multiclass data were limited and tended to focus on correlation and trend while missing other patterns, such as the relations between two sub-graphs. For example, in the case of the multi-class juxtaposed line graph Figure 7(a), only one of eight responses aligned with the stated communication goal of that graph—"*I see changes in dollars (which I assume represents weekly income) of weekly earnings for all employees at an unspecified company versus earnings of employees that work in production or non-supervisory roles... It seems that production and non-supervisory employees tend to make less money per week compared to the employee pool, but salaries have changed at comparable rates.* [P09]".

However, even single-class charts exhibited strong variance between participants. For example, in a single-class scatterplot (see Fig.1 in OSF Supplement) that visualizes the negative correlation between mileage and manufactured year, five participants each identified different statistics—correlation, even distribution, N/A, negative correlation, and trend. We observed a similar case of four participants in a single-class line graph as shown in Figure 5(b), where responses comprised positive correlation (2), compare (1), and trend (1). In Figure 7(a), responses from five participants varied from the increase (1), upward-trend (1), similar trend (1), extremum

(1), and compare-trend (1). Similarly, in Figure 1(b), responses consisted of extremum (3), compare (3), count (2), and correlation (1).

4.3.2 Composition. As discussed in Theme 1 (see Section 4.1), people's comprehension aligned best with designer intent using standard, single-class charts (44/72 complete or general match). With composited charts, alignment was highest with single-class juxtaposed (31/72) followed by multi-class non-juxtaposed (27/72) and finally, multi-class juxtaposed (24/72) graphs. One participant mentioned that these charts were significantly harder to use—"*I could understand bar graphs and simple line graphs well but not scatterplots or graphs involving multiple categories or graphs.* [P16]"

The majority of non-juxtaposed charts with single-class data effectively communicated their target properties (35/72). For example, participants noted a correlation in Figure 5(b) as in "*There is a positive association between gross movie sales and IMDB. The association does not seem to be completely linear, but more dependent on the rating.* [P01]," distribution and extremum in "*scatter plot showing max temp in Edoford during sunny days in a year, definite peak in summer from June to September or October* [P13]" on Figure 5(a), and correlation in "*Easier to read, shows the correlation between the turtle's body mass and flipper length per species in individual graphs* [P02]" on Figure 3(a). These findings suggest that simple, single-class visualizations tended to better align with stated intentions; however, this alignment was still below 50% overall.

More complex data and conclusions drawn on them may require composite chart designs. Juxtaposition is a popular choice for composite charts, but its complexity led to more frequent misalignment and often a full lack of engagement with any specific patterns in the data. Single-class juxtaposed composition for both line (12/24) and bar (13/24) graphs aligned with their intended goals better than other compositions. For example, in Figure 3, participants identified the pattern and performed the comparative analysis of the data—more than 67% of participants' responses aligned with designer's stated intention of comparing trends (5 times), comparing correlations (2 times) and estimating discrete values (1 time) on line and bar graphs respectively.

Viewers found single-class juxtaposed line graphs challenging when the graphs were vertically aligned. Eight responses mentioned that juxtaposed line graphs were "hard" to use, for example, in Figure 7. One participant noted—"*That final graph [single-class juxtaposed line graphs; sub-graphs are stacked on top of each other] was visually very hard for me to parse because all of the lines were stacked on top of one another. It was very hard for me to get a visual frame of reference for how to measure the growth of the stock.* [P07]" This difficulty may have limited people's abilities to synthesize information across pairs of visualizations.

Responses varied more between single- and multi-class versions of the same visualization type on juxtaposed graphs. For example, 45% of responses matched designer intents in single-class juxtaposed line graphs versus 37% in multiclass juxtaposed line graphs. We found an overall inverse correlation between the number of sub-graphs and alignment with the design intent. Specifically, fewer charts led to higher alignment. Non-juxtaposed charts tended to support higher alignment than other kinds of composite charts.

Multi-class juxtaposed graphs more often corresponded to comparisons with a less varied use of other statistics. Despite being data-rich, multiclass juxtaposed designs tended to lead to shallow interpretation—"*In the left, I see plots showing the cost of a car and how old the car is associated with the odometer reading. On the right, I see three different colored graphs. I am assuming they are showing different car models.*" [P06] for Figure 7(b). Multi-class juxtaposed line graph responses focused more on local trends, such as *change* (7/24), *increase-decrease* (8), *upward-downward trend* (9), *spike* (2), *stable trend* (2), *increase* (10), and *peak-valley* (4).

4.3.3 Supplemental Graphical Information. Several charts included basic annotations and added supplemental graphical elements. Although we did not control for the presence or absence of visual annotations (e.g., trend lines or highlights), instead choosing to mimic the widely-used graphs on 42 chosen datsets as closely as possible for ecological validity, we found that annotations assisted people in identifying the intended patterns in the data. Such annotations (e.g., highlighting value on a line graph, a visible baseline or grid line in bar graphs) helped people identify complex intentions, such as quarterly sales in a line graph and 52-week low stock price (see Fig. 29 in OSF Supplement) or the maximum budget beyond a threshold in Figure 5. Statistical baselines helped people identify the trend and correlations in all three types of graphs. However, in a line graph plotting stock data with *52-week low* value as red baseline missed its intended task—highlight days in red when the stock hit below the 52-week low price and assess stock performance over time (correlation, trend, outliers, maxima, discrete value). Only one participant's response matched the intended tasks, reporting correlation and trend. Additionally, grid lines in bar graphs and line graphs helped people identify ranges and extract maximum values.

However, annotations also represent trade-offs. For example, gridlines can lead to attraction and repulsion effects that cause people to incorrectly estimate effect size [138]. Further, calling out key information in a chart may cause people to fail to attend to other salient properties of the data [139]. Future work should investigate the role annotation plays in shaping people's comprehension in the wild.

4.3.4 Individual Differences. Participant responses greatly varied between participants. As shown in Table 1 (and Table 5 in Appendix B), 78% (92/117) of complete match responses were grouped among only ten participants. 59% of participants did not provide responses that reliably aligned with the stated intentions. Only 45% (11/24) of the participants managed to get an average (complete+general > 7/12) to higher (complete+general > 10/12) comprehension match with the stated intention. Three participants (Nursing, Environmental Science, and Computer Science major students) had 11/12 complete match responses, while three other participants had no responses that aligned with the stated intention at all. The fact that only 34% (8/24) participants had an upper high-range comprehension match shows that the designer's intended data communication was not immediately obvious to most readers despite these graphs replicating visualizations intended for mass communication. This internal variance suggests that individual differences may play a role in what information a visualization naturally communicates. For example, people's performance differs with domain expertise [49] or graphical literacy [36]. While we did not systematically sample for specific demographic characteristics (e.g., profession, education, or graphical literacy), a comparative study of high-level

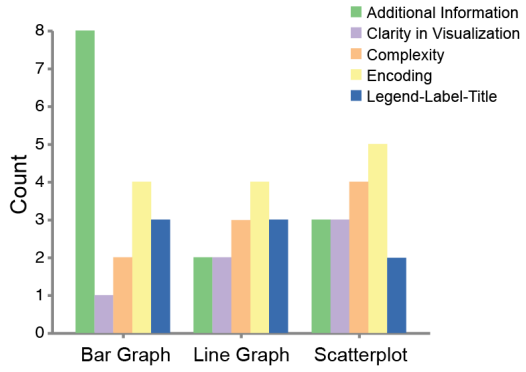


Figure 9: Participants' responses and description of design critiques and need for additional information on the data. We observed that apart from additional information, all graph types received similar critiques.

visual comprehension on diverse populations requires further investigation.

5 ANALYSIS OF DESIGN AND DATA CRITIQUES

Participants noted a range of additional factors that influenced their abilities to read a visualization. We identified and summarized these factors in Figure 9. While not directly related to our core research questions, these observations provide additional insight into visualization design guidelines.

Several observations related to common graphic design principles, such as a visualization was "*easy to read because of [high] color contrast* [P16]" on Figure 1(a); "*hard to read because of dashes* [P01]" or other less common visual metaphors on Figure 1(c). Participants raised four common critiques based on a visualization's content: the need for *additional information*, issues with *legends and labels*, confusion with *visual encodings*, and too much *complexity*.

Additional Information: Participants frequently commented on missing information and context they felt limited their abilities to describe visualizations. 16 participants felt they needed additional information about the data to better comprehend a graph, such as units of measurement (e.g., price in Figure 1(c)), or the demographics of people represented on graphs (Fig. 54 in supplements). Participants commented about acronyms used in the data, even when those acronyms were peripheral to patterns in the data. One participant noted, "*Some graphs are not easy to understand because I don't know what is IMDB* [P01]" in Figure 5(b). People found abbreviated stock names (Figure 1(c)) or unfamiliar acronyms of airlines made charts more challenging to describe. Participants also found jargon off-putting, such as odometer (2/24), CPU architecture & cores (5/24), and annuli (2/24). One participant noted that "*Some of the graphs I had to concentrate on harder as I do not follow stocks or computer programming types.* [P15]"

Legends and Labels: Participants actively attended to legends, labels, and titles, raising concerns when they were unclear (2/24) or missing (7/24). Having clear axis labels, titles, and legends is a known best practice in communicating quantitative information [127]. Our findings support these practices of including labels and legends considering the general audience, enabling them to identify patterns in the data.

Visual Encoding: Participants' comments about data encodings primarily centered on "busy" features of a graph, such as bright colors, dashed lines, and dots in graphs. People generally appreciated the use of color. For example, they noted that a graph may be "easy to read because of color contrast."

Color could also be misused, with one participant noting that visualization had—"*Lots of colorful lines- it's a little jarring at first glance. There's a helpful key at the top though...* [P13]" However, alternative encodings for delineating categories were not always as well-received. For example, all of the participants raised concerns about the dashed-line in Figure 1(c), commenting that it was "*hard to read because of similar dash patterns* [P16]" or "*Difficult to distinguish stroke dash, making the graph confusing.* [P01]" Dashes are more robust across media and more accessible than colors; however, people found them more confusing. This contrast indicates a trade-off in categorical encoding that should be explored in future work.

Colors that followed semantic guidance were also seen as helpful, in line with past recommendations [75, 81]. For example, in Figure 1(b), where colors aligned with specific foods, participants' responses tended to align with the designer intentions: "*Output of specific foods (in tons) for specific food products, grouped by their food group. Almost all of the fruits had higher production than anything else, while dairy had the lowest productions.* [P18]"

Clarity and Complexity: Participants felt that graphs using too many distinct encodings were difficult to read. For example, 1(a) was described as "*Hard to read because of the colorfulness and varying sizes of dots. Hard to estimate the insurance charges certain dot sizes correspond to...* [P13]"

Others noted that in some graphs "*too much information was given, making the graph hard to understand* [P21]" or that there were "*too many elements in the graph being given, making it confusing* [P05]"

Too much data using multiple visual channels increased perceived cognitive load, especially with juxtaposed graphs. The think-aloud responses indicated that people perform graph comprehension in two steps, first processing legends, labels, mark encodings, and subgraph organization and then attending to patterns in the data. Multiple encodings or sub-graphs complicated comprehension by making it harder to complete the first step, in line with our observations in Section 4.1. However, thoughtful organization could increase perceived usability. "*The easier graphs to understand are the more organized and separated due to the clarity provided and lack of over stimulation to the brain; the messier and more colorful, the more my brain shuts down and doesn't want to process it,* [P21]." For example, Figure 1(b) organizes like bars into the same physical group, encouraging people to consider groups as a single unit. A lack of organization led to a similar decrease in perceived usability. Participants found that "*dense clusters make it hard to read the given data points* [P14]" in scatterplots and that it was "*difficult to track the [target] line... due to all the lines overlapping and having similar colors* [P17]" in multi-class line graphs. critiques from the participants on missing information and context about visualized data.

6 DISCUSSION

We investigated the high-level patterns people naturally see when encountering a visualization without a guiding task. This work provides preliminary steps towards characterizing the patterns people perceive unprompted from a graph as a function of its design. Our findings offer preliminary insight into the alignment of design intentions, reader intuitions, and guidelines from graphical perception and expert heuristics. Future research should explore these insights to create more refined heuristics that connect reader intuition with empirical findings.

6.1 Key Themes

Our thematic analysis offers a new lens for understanding visualization effectiveness by modeling the patterns and statistics people intuitively extract from a given visualization. This analysis revealed a misalignment between the guidelines established through experimental research on visualization and what individuals actually perceive when they encounter visualizations in the wild. Our results highlight three key themes:

Intention does not align with comprehension. The patterns that describe a visualization often failed to match its communication goals. The insights and salient patterns people intuitively extract from that visualization may not align with desired communication goals, even for visualizations reflecting best practices. In contrast to previous research, we noted that the degree of alignment is contingent on various factors, which we will discuss in subsequent subsections.

Results from cued tasks may not alone predict the knowledge people build in a graph. Visualization effectiveness is not fully captured by people's abilities to quickly and accurately complete a cued task as in traditional graphical perception paradigms. We need experiments emphasizing both top-down precision (i.e., with cued tasks) and bottom-up high-level visual comprehension (i.e., without cued tasks) to understand the knowledge people build in a graph.

Chart type alone is not sufficient to predict the information people extract from a graph. The patterns people notice in a given visualization depend on several dimensions of visualization design. Mapping tasks to chart type is a powerful paradigm, but fails to account for all of the perceptual variables involved in visualization interpretation.

Our results offer a set of general considerations for visualization designers and researchers, helping reconcile prior findings (section 6.2), providing implications for applying design heuristics (section 6.3) and lending insight into methodological considerations for visualization evaluation (6.4).

6.2 Relation to Prior Findings

While specifically focused on statistical patterns perceived in the data, our findings complement observations from past work focusing on related elements of visualization use, such as preference and higher-level cognition and sensemaking [16, 120, 123]. While these past findings address either specific audiences or complementary dependent measures, juxtaposing them against our results highlights two major considerations for visualization design that our findings confirm: the importance of multiple perspectives on "effectiveness" and the need to consider diverse audiences.

6.2.1 Defining Effectiveness. The patterns people perceive in data are influenced by a range of factors, including the visual channels used [26, 66], the settings of those channels [66, 122], the distribution and semantics of the data [45, 117], past knowledge about the data [139], and even supplementary information like annotations and titles [69]. Our results confirm that relying on cued statistical measures to assess performance provides important feedback but may not be sufficient to capture what a visualization actually communicates given this complex space.

Our findings resonate with previous work, highlighting the value of considering comprehension as a goal for interpreting communication intention [16, 89] as cued tasks, particularly low-level tasks, alone may not suffice to capture the real communication goals and interpretations of visualizations [1, 141]. These observations further confirm that encoding channels and graph types are insufficient to predict effectiveness [24, 120]. Connecting findings across these different perspectives can help shape more holistic and robust guidance for effective visualization design, especially if paired with theoretical frameworks that offer grounded approaches to considering complex interactions between aspects of visualization design and use (e.g., rhetorical approaches [58] or communication frameworks [1]). However, bringing these disparate sources of data and theory together under a unifying framework remains an open research challenge.

6.2.2 Who is the Consumer? Our study aligns with Burns et al.'s call [16] for evaluations that consider different levels of visualization understanding. The observed misalignment between visualization intentions and participant interpretation highlights the need for multi-tiered evaluation frameworks that assess both low-level task performance and high-level visual comprehension, which we discuss further in Section 6.4. While our studies offer only a limited preliminary lens on comprehension, as our understanding of visualization comprehension grows, having these frameworks in place will allow the visualization community to act more readily upon the resulting findings.

A key component of this analysis will be accounting for demographic factors that play a crucial role in how individuals interpret visualizations [20, 46, 91, 136]. Our findings confirm that comprehension varies significantly between individuals: to some degree, insight is in the eye of the beholder. Even the concept of a 'novice' could be a multi-faceted label [15], not only pertaining to one's unfamiliarity with visualization tools or techniques but also their ability to extract and interpret complex data patterns or even work with the same representations across data from different domains, as seen in hesitancy introduced by unfamiliar concepts or acronyms. This variability may also inform improved literacy assessments and other methods for characterizing visualization comprehension. While not within the scope of this study, the effects of personal demographics and literacy on visual comprehension are critical to future work.

6.3 Implications for Design

Effective designs make key patterns salient. Design practices use target tasks [6, 83] and problem domain information [83] to drive a particular design choice. Our results indicate that this relationship may be powerful but insufficient: aspects of chart composition, data, and other design elements might alter the "right" visualization choice for a given task. For example, design guidelines suggest that Figure 1 and Figure 7 should communicate the target tasks, but participants did not reliably use those tasks to describe the content of the graphs.

Visualization designers should be aware that what they intend to communicate through a graph may not always align with what viewers naturally perceive. Relying solely on cued statistical measures to assess performance may not accurately determine whether their visualizations will successfully convey their objectives. This underscores the need for additional guidelines derived from future high-level comprehension studies to enhance the effectiveness of visualizations for a given communication goal.

Our study highlights that these design guidelines should go beyond simply mapping chart types to data types or statistics. Data complexity, composition, and supplemental information such as captions [24], titles [69], and additional text [120] also influence how viewers interpret visualizations and what statistics they extract from those graphs. Designers should consider these factors when selecting visualization design. Our results offer insights into trade-offs associated with common design choices, including:

Juxtaposition. Past work provides conflicting perspectives on juxtaposition versus superposition [61, 90]. In our study, people found juxtaposed graphs more difficult to use, provided less synthesis of the data when describing the graphs, and were less likely to describe the graph's intended message. However, superimposed graphs (in our case, multiclass, non-juxtaposed) can also lead to increased visual complexity when communicating differences in classes with different visual channels, albeit less often than juxtaposed graphs in our results. These conflicts suggest a need to understand better different types of complexity that may arise in more complex analysis scenarios.

Expressiveness. More complex graphs can be more expressive and communicate a broader range of information in a single chart [27] but were less likely to achieve their intended goals. People's descriptions of more complex graphs tended to focus on surface features and described fewer patterns and statistics, even though more questions could be answered with those charts. Designers should think about ways of encouraging engagement with more complex charts and help people more rapidly and effectively orient themselves within complex data. This finding also supports more minimalist design practice: Simplicity and clarity remain essential principles in visualization design.

Chart Scaffolding. Misinterpretations drawn from shallow, at-a-glance readings of misleading charts may suggest that people do not pay close attention to labels and legends [30]. However, past work in visual attention suggests that people prioritize axes and titles [65]. We found that participants spent significant time understanding the visualization environment (e.g., graph itself, encodings, and text) and actively processed legends and labels. Such labels should carefully consider their content as well: abbreviations, jargon, or missing units of measure can be distracting, alienate readers, and cause people to second guess their interpretations.

Supplemental Graphical Information. Additional visual cues to critical data or patterns, such as annotations, lead to more consistent interpretation. Designers can use these cues to direct attention to key aspects of the data and reduce ambiguity in data interpretation. Visualizations presenting data without statistical or written annotations, explicitly showing a target pattern, and missing contextual information (e.g., unit, topic, title, acronyms) led to greater variance in reported patterns. However, using explicit visual cues to highlight some patterns may cause people to miss other key takeaways [10, 139].

6.4 Implications for Methods

Conventional design guidelines in visualization often suggest which tasks a given visualization type is effective for, such as scatterplots for clustering, distribution characterization, and correlation. However, these guidelines are typically based on experiments where people are explicitly cued to identify specific statistics. Task effectiveness in controlled settings may not always translate to real-world scenarios, where people may approach visualizations without predefined goals in mind. Accompanying text and other chart contexts (e.g., labels and titles) may provide relevant cues in some cases [24, 69, 120]. However, our results indicate a need for a broader range of methodological considerations in generalizing from empirical studies to design guidelines.

6.4.1 Rethinking Graphical Perception. Visualization effectiveness is typically measured through performance (e.g., accuracy, completion time, and error rate) and subjective experience (e.g., confidence, familiarity, and subjective preferences) in performing a specific task, such as estimating a statistic or finding an object [35, 101]. **Such cued experimental tasks may not adequately capture how people use visualizations in the wild** or determine how well they achieve their true communication goal. These results align with previous studies that suggest visual properties [120] or low-level tasks [1] alone are not sufficient to capture the designer's intention or determine the reader's interpretation. Visualizations are often contextualized in text with a descriptive title, caption, or accompanying prose that cues people to focus on the specific information. However, charts may also appear without this scaffolding, such as in presentations, exploratory analysis tools, or as "teaser" figures that people see before reading the accompanying article. Measuring performance using abstract statistical quantities may not tell designers whether visualizations in these contexts will achieve their goals. Further, failure to see a described pattern may lead to mistrust in information even when cued.

A statistics-focused and directed research task may invoke a cognitive process known as the *cognitivisation of perception* [106]. This cognitivization may change how people engage with a visualization, causing them to note different patterns as being more salient than actually are [139]. In practice, this implies that just because people can estimate a statistic from a given graph does not mean that they *will* estimate that statistic, at least unprompted. Our findings offer preliminary steps towards designing for graphs in the wild without the risk of cognitivization.

Future studies adding visual comprehension evaluations should complement traditional graphical perception approaches to explore methods for understanding how visualization design, data, and visual encodings can more universally predict whether a visualization will achieve its goals or helped readers in interpreting the insights. Doing so requires understanding both goal-directed (i.e., traditional graphical perception to understand accuracy and speed) and feature-driven (i.e., uncued comprehension to understand interpretation) visualization use.

6.4.2 The Importance of Task Framing. Our pilot study only asked people to describe a visualization. Statistical tasks such as estimate correlation [51], identify outliers [5], and characterize distribution [66] did not arise in these descriptions. As a result, we amended our study design to ask participants to identify questions each graph addressed. This new experimental task, in turn, led to significantly more statistical tasks and specific patterns emerging in people's responses. Best practices in survey design encourage asking about a target topic from different perspectives to avoid framing bias [82]. We found systematic differences in the ways people described the patterns in graphs depending on which experimental task (question or description) they were addressing.

Question: Asking people what questions one could answer with a graph led to more statistically-oriented responses, such as "*How do the sales of item X change throughout the year and the months?*" or "*How much, on average, does each genre need for production budget?*" People frequently formed analytical questions with statistical quantities, and these responses tended to better align with the original goals of the graph: Question-oriented responses aligned in 182 of 288 responses, compared to 117 of 288 description-oriented responses (Table 1).

Description: Asking people to describe a graph led to responses that were more deeply grounded in the graph's semantics, closer to traditional insights [89] and level of understanding [16]. We found from the think-aloud that people initially spent time on understanding the graph environment (e.g., labels and legends) when describing the graph, leading to responses reflecting contextualized knowledge rather than raw statistics. However, for more complex graphs, descriptions are more often correlated with reading what one can physically see in the graph. When describing a graph, people primarily browsed at a high level, and if they could not see salient patterns or statistics, they moved on. For example, one person verbally described a complex graph with large pauses between observations–"*this graph has age of car and odometer reading on both graphs....[paused]...there difference price and age and model information....[paused]* [P05]" (see Figure 7(b)).

Prompting people to provide different kinds of information about a graph (i.e., descriptions versus questions) may allow experiments to tease apart the different levels of reasoning people engage within a graph. These tensions should be explored in future work to understand trade-offs for experimental evaluation.

6.5 Limitations & Future Directions

Our study emphasizes the need for further research to build a deeper understanding of high-level comprehension in data visualizations. This includes investigating new chart types, individual differences, data complexity, and the impact of design choices on data communication.

Design Factors: Our study evaluated the kinds of information people draw from scatterplots, line graphs, and bar graphs with different data types and compositions. While these reflect commonly studied visualization types [101], they are only a small portion of visualization types. Future work should explore other visualization types, such as maps, pie charts, and bubble charts, as well as alternative designs for the current graphs. We synthesize our results across higher-level properties of a graph (e.g., its composition and visualization technique) to retain a manageable scope for our study; however, our results also indicate lower-level design factors like individual visual channels and their interactions also drive data interpretation. Future work should provide a more systematic investigation of how specific variables in visualization design affect high-level interpretation. Our visualizations also draw from narrative visualizations to provide a ground truth for a graph's intended purpose, reflecting best practices. For future work, we plan to extend our method to exploratory scenarios, where knowledge evolves over time and, with use and visualizations, often aim to support a wider range of tasks.

Limitation Due to Variability in Results: We found a person's background, such as their education or profession, may affect their comprehension as in prior work [49]. Understanding the nature and extent of these differences can be instrumental in refining visualization design strategies to cater to diverse audiences effectively, as demonstrated in past studies [15, 46]. However, we did not systematically control for these factors across participants. Moreover, our study did not aim to assess novices' performance in visual comprehension of widely-used visualizations. Hence, participant recruitment did not specifically target novices, contrary to suggestions in prior works [15, 16, 46, 91]. While we believe our study reflects a typical audience for most applications, future work should be conducted on a larger population using systematic sampling methods to better understand how demographic factors influence data interpretation.

Study Procedure: We focused on verbal and textual descriptions of graphs, but such descriptions omit additional data about what patterns are salient, such as what features in a graph people actively attend to. Future work can provide additional response data, such as eye-tracking, to further investigate the most salient elements of a graph and their influence on data interpretation. We studied peoples' high-level comprehension of visualization via natural language and think-aloud protocols. However, previous studies [11, 70] showed visual recognition and people's recall of the message in visualization also play a vital role in visual memory. Future work should focus on viewers' visual recall and examine what high-level comprehension people can recall from visualization and which designs may help people recall the messages. We believe these directions would help us further understand how to retain the key messages of a visualization.

7 CONCLUSION

Designers create visualizations to achieve specific high-level analytical or communication goals. These goals often require people to naturally extract complex, interconnected patterns in data. However, perceptual studies of visualization effectiveness focus on isolated, predefined, low-level tasks, such as estimating statistical quantities. People may efficiently assess that pattern when cued, but that pattern may not stand out when the same people encounter a similar visualization in the wild. We studied three visualization types—scatterplots, line graphs, and bar graphs—to investigate the high-level patterns people naturally see when they encounter a visualization. While graph type predominantly affected the statistics people noted, we found that interpretation varies with the data type, graph composition, and specific design elements. These findings enable us to look at visualization design effectiveness and their communication goals with a new lens of high-level visual comprehension. Through our results, we highlight the significance of incorporating both high-level comprehension and low-level tasks in assessing visualization effectiveness.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grant No. 2127309 to the Computing Research Association for the CIFellows project, NSF IIS-2046725, NSF IIS-1764089, and NSF IIS-2316496.

REFERENCES

[1] Eytan Adar and Elsie Lee. 2020. Communicative visualizations as a learning problem. *IEEE Trans Vis Comput Graph* 27, 2 (2020), 946–956.
[2] Afghanistan population dataset 2023. Afghanistan population dataset. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=AF.
[3] Airlines 2022. Airlines Dataset. https://www.bts.gov/newsroom/2022-annual-and-4th-quarter-us-airline-financial-data.
[4] Oluwatobi Noah Akande, Taofeeq Alabi Badmus, Akinyinka Tosin Akindele, and Oladiran Tayo Arulogun. 2020. Dataset to support the adoption of social media and emerging technologies for students' continuous engagement. *Data in Brief* 31 (2020), 105926.
[5] Danielle Albers, Michael Correll, and Michael Gleicher. 2014. Task-driven evaluation of aggregation in time series visualization. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*. 551–560.
[6] Robert Amar, James Eagan, and John Stasko. 2005. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization (InfoVis)*. 111–117.
[7] Auto MPG 2017. Auto MPG Dataset. https://www.kaggle.com/datasets/uciml/autompg-dataset.
[8] Edward Awh, Artem V Belopolsky, and Jan Theeuwes. 2012. Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends Cogn. Sci.* 16, 8 (2012), 437–443.
[9] Lisa Best, Aren Hunter, and Brandie Stewart. 2006. Perceiving relationships: A physiological examination of the perception of scatterplots. In *Proc. 4th Conference on Diagrammatic Representation and Inference*.
[10] Tal Boger, Steven B Most, and Steven L Franconeri. 2021. Jurassic mark: Inattentional blindness for a datasaurus reveals that visualizations are explored, not seen. In *2021 IEEE Visualization Conference (VIS)*. IEEE, 71–75.
[11] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2015. Beyond memorability: Visualization recognition and recall. *IEEE Trans Vis Comput Graph* 22, 1 (2015), 519–528.
[12] Michelle A Borkin, Azalea Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What makes a visualization memorable? *IEEE Trans Vis Comput Graph* 19, 12 (2013), 2306–2315.
[13] David Borland, Irena Brain, Karamarie Fecho, Emily Pfaff, Hao Xu, James Champion, Chris Bizon, and David Gotz. 2021. Enabling Longitudinal Exploratory Analysis of Clinical COVID Data. In *IEEE Workshop on Visual Analytics in Healthcare (VAHC)*. 19–24.
[14] Matthew Brehmer and Tamara Munzner. 2013. A multi-level typology of abstract visualization tasks. *IEEE Trans Vis Comput Graph* 19, 12 (2013), 2376–2385.
[15] Alyxander Burns, Christiana Lee, Ria Chawla, Evan Peck, and Narges Mahyar. 2023. Who Do We Mean When We Talk About Visualization Novices?. In *Proc. 2023 ACM SIGCHI Hum Factor Comput Syst*. 1–16.
[16] Alyxander Burns, Cindy Xiong, Steven Franconeri, Alberto Cairo, and Narges Mahyar. 2020. How to evaluate data visualizations across different levels of understanding. In *IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*. IEEE, 19–28.
[17] Calories Burned 2020. Calories Burned During Exercise and Activities. https://www.kaggle.com/datasets/aadhavvignesh/calories-burned-during-exercise-and-activities.
[18] Car Model Dataset 2023. Car Model Dataset. https://www.kaggle.com/datasets/peshimaammuzammil/2023-car-model-dataset-all-data-you-need.
[19] Mackinlay Card. 1999. *Readings in information visualization: using vision to think*. Morgan Kaufmann.
[20] Sheelagh Carpendale, Alice Thudt, Charles Perin, and Wesley Willett. 2017. Subjectivity in personal storytelling with visualization. *Inf. Des. J.* 23, 1 (2017), 48–64.
[21] Certificated Air Carrier 2021. Certificated Air Carrier Fuel Consumption and Travel. https://www.bts.gov/content/certificated-air-carrier-fuel-consumption-and-travel.
[22] Yu-Hsuan Chan, Carlos Correa, and Kwan-Liu Ma. 2013. The generalized sensitivity scatterplot. *IEEE Trans Vis Comput Graph* 19, 10 (2013), 1768–1781.
[23] Remco Chang, Caroline Ziemkiewicz, Tera Marie Green, and William Ribarsky. 2009. Defining insight for visual analytics. *IEEE Comput Graph Appl* 29, 2 (2009), 14–17.
[24] Shelly Cheng, Hazel Zhu, and Eugene Wu. 2022. How Do Captions Affect Visualization Reading? *arXiv preprint* (2022).
[25] Lisa Cheong, Susanne Bleisch, Allison Kealy, Kevin Tolhurst, Tom Wilkening, and Matt Duckham. 2016. Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *Int J Geogr Inf Sci* 30, 7 (2016), 1377–1404.
[26] William Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554.
[27] William S Cleveland and Robert McGill. 1986. An experiment in graphical perception. *International Journal of Man-Machine Studies* 25, 5 (1986), 491–500.
[28] Charles E Connor, Howard E Egeth, and Steven Yantis. 2004. Visual attention: bottom-up versus top-down. *Current Biology* 14, 19 (2004), R850–R852.
[29] Michael Correll, Danielle Albers, Steven Franconeri, and Michael Gleicher. 2012. Comparing averages in time series data. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*. 1095–1104.
[30] Michael Correll and Jeffrey Heer. 2017. Black hat visualization. In *IEEE Workshop on Dealing with Cognitive Biases in Visualisations (DECISIVe)*, Vol. 1. 10.
[31] Michael Correll and Jeffrey Heer. 2017. Regression by eye: Estimating trends in bivariate visualizations. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*. 1387–1396.
[32] CPU Specifications 2022. CPU Specifications Dataset. https://www.kaggle.com/datasets/lincolnzh/cpu-specifications-dataset.
[33] Will Cukierski. 2012. Titanic - Machine Learning from Disaster. https://kaggle.com/competitions/titanic
[34] S Dixon. 2022. Average daily time spent on social media worldwide 2012-2022. *Erişim Tarihi* 22 (2022), 2022.
[35] Madison A Elliott, Christine Nothelfer, Cindy Xiong, and Danielle Albers Szafir. 2020. A Design Space of Vision Science Methods for Visualization Research. *IEEE Trans Vis Comput Graph* (2020).
[36] Steven L Franconeri, Lace M Padilla, Priti Shah, Jeffrey Zacks, and Jessica Hullman. 2021. The science of visual data communication: What works. *Psychol. Sci. Public Interest* 22, 3 (2021), 110–161. https://doi.org/10.1177/1529100621105195 6
[37] Taher M Ghazal, Sajid Hussain, Muhammad Farhan Khan, Muhammad Adnan Khan, Raed AT Said, Munir Ahmad, et al. 2022. Detection of benign and malignant tumors in skin empowered with transfer learning. *Comput. Intell. Neurosci.* 2022 (2022).
[38] JJ Gibson. 1972. A theory of direct visual perception. In the Psychology of Knowing, ed. JR Royce, WW Roze-boom, 215-27. *New York: Gordon & Breach* 63 (1972), 396–97.
[39] Michael Gleicher. 2017. Considerations for visualizing comparison. *IEEE Trans Vis Comput Graph* 24, 1 (2017), 413–423.
[40] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. 2011. Visual comparison for information visualization. *Inf Vis* 10, 4 (2011), 289–309.
[41] Michael Gleicher, Michael Correll, Christine Nothelfer, and Steven Franconeri. 2013. Perception of average value in multiclass scatterplots. *IEEE Trans Vis Comput Graph* 19 (2013).
[42] Global Summary 2023. Global Summary of the Day station observations. http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.GLOBAL/.STATION.cuf/.
[43] Anna Gogolou, Theophanis Tsandilas, Themis Palpanas, and Anastasia Bezerianos. 2018. Comparing similarity perception in time series visualizations. *IEEE*

Trans Vis Comput Graph (2018).
[44] Goverment Bond Dataset 2023. Goverment Bond Dataset. https://fred.stlouisfed.org/series/IRLTLT01USM156N.
[45] Connor Gramazio, Karen Schloss, and David Laidlaw. 2014. The relation between visualization size, grouping, and user performance. *IEEE Trans Vis Comput Graph* (2014).
[46] Lars Grammel, Melanie Tory, and Margaret-Anne Storey. 2010. How information visualization novices construct visualizations. *IEEE Trans Vis Comput Graph* 16, 6 (2010), 943–952.
[47] Grand Slam Title Winners 2023. Grand Slam Title Winners. https://www.kaggle.com/datasets/abrafey/mens-womans-grand-slam-title-winners.
[48] Richard Langton Gregory. 1974. *Concepts and mechanisms of perception.* Charles Scribner's Sons.
[49] Kyle Wm Hall, Anthony Kouroupis, Anastasia Bezerianos, Danielle Albers Szafir, and Christopher Collins. 2021. Professional differences: A comparative study of visualization task performance and spatial ability across disciplines. *IEEE Trans Vis Comput Graph* 28, 1 (2021), 654–664.
[50] Lane Harrison, Drew Skau, Steven Franconeri, Aidong Lu, and Remco Chang. 2013. Influencing visual judgment through affective priming. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*.
[51] Lane Harrison, Fumeng Yang, Steven Franconeri, and Remco Chang. 2014. Ranking visualizations of correlation using weber's law. *IEEE Trans Vis Comput Graph* 20, 12 (2014), 1943–1952.
[52] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*.
[53] Jeffrey Heer, Frank Van Ham, Sheelagh Carpendale, Chris Weaver, and Petra Isenberg. 2008. Creation and collaboration: Engaging new audiences for information visualization. *Information Visualization: Human-Centered Issues and Perspectives* (2008), 92–133.
[54] Historical Weather Dataset 2017. Historical Weather Dataset. https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data.
[55] Matt-Heun Hong, Jessica Witt, and Danielle Albers Szafir. 2021. The Weighted Average Illusion: Biases in Perceived Mean Position in Scatterplots. *IEEE Trans Vis Comput Graph* 28, 1 (2021), 987–997.
[56] Allison Marie Horst, Alison Presmanes Hill, and Kristen B Gorman. 2020. palmerpenguins: Palmer Archipelago (Antarctica) penguin data. *R package version 0.1.0* (2020).
[57] Household monthly electricity bill 2020. Household monthly electricity bill. https://www.kaggle.com/datasets/gireeshs/household-monthly-electricity-bill.
[58] Jessica Hullman and Nick Diakopoulos. 2011. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Trans Vis Comput Graph* 17, 12 (2011), 2231–2240.
[59] IMDB 2019. IMDB Dataset. https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews.
[60] Waqas Javed and Niklas Elmqvist. 2012. Exploring the design space of composite visualization. In *IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 1–8.
[61] Waqas Javed, Bryan McDonnel, and Niklas Elmqvist. 2010. Graphical Perception of Multiple Time Series. *IEEE Trans Vis Comput Graph* 16, 6 (2010), 927–934.
[62] Hyeon Jeon, Ghulam Jilani Quadri, Hyunwook Lee, Paul Rosen, Danielle Albers Szafir, and Jinwook Seo. 2023. Clams: a cluster ambiguity measure for estimating perceptual variability in visual clustering. *IEEE Trans Vis Comput Graph* (2023).
[63] Smiti Kaul, Cameron Coleman, and David Gotz. 2020. A rapidly deployed, interactive, online visualization system to support fatality management during the coronavirus disease 2019 (COVID-19) pandemic. *J Am Med Inform Assoc* 27, 12 (2020), 1943–1948.
[64] Matthew Kay and Jeffrey Heer. 2016. Beyond weber's law: A second look at ranking visualizations of correlation. *IEEE Trans Vis Comput Graph* 22 (2016).
[65] Nam Wook Kim, Zoya Bylinskii, Michelle A Borkin, Krzysztof Z Gajos, Aude Oliva, Fredo Durand, and Hanspeter Pfister. 2017. BubbleView: an interface for crowdsourcing image importance maps and tracking visual attention. *ACM Trans Comput Hum Interact* 24, 5 (2017), 1–40.
[66] Younghoon Kim and Jeffrey Heer. 2018. Assessing effects of task and data distribution on the effectiveness of visual encodings. *Comput Graph Forum* 37, 3 (2018), 157–167.
[67] Yea-Seul Kim, Katharina Reinecke, and Jessica Hullman. 2017. Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation. *IEEE Trans Vis Comput Graph* 24, 1 (2017), 760–769.
[68] Ron Kohavi et al. 1996. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid.. In *Proc. ACM SIGKDD Int. (KDD)*, Vol. 96. 202–207.
[69] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and slants in titles of visualizations on controversial topics. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*. 1–12.
[70] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2019. Trust and recall of information across varying degrees of title-visualization misalignment. In *Proc. 2019 ACM SIGCHI Hum Factor Comput Syst.* 1–13.
[71] Robert Kosara. 2016. An empire built on sand: Reexamining what we think we know about visualization. In *IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*.
[72] Jon A Krosnick. 2018. Questionnaire design. *The Palgrave handbook of survey research* (2018), 439–455.
[73] Elsie Lee-Robbins and Eytan Adar. 2022. Affective Learning Objectives for Communicative Visualizations. *IEEE Trans Vis Comput Graph* 29, 1 (2022), 1–11.
[74] Jing Li, Jean-Bernard Martens, and Jarke van Wijk. 2010. A model of symbol size discrimination in scatterplots. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*. 2553–2562. https://doi.org/10.1145/1753326.1753714
[75] Sharon Lin, Julie Fortuna, Chinmay Kulkarni, Maureen Stone, and Jeffrey Heer. 2013. Selecting Semantically-Resonant Colors for Data Visualization. *Comput Graph Forum* 32, 3pt4 (2013), 401–410. https://doi.org/10.1111/cgf.12127
[76] Claudia B Maier, Hilary Barnes, Linda H Aiken, and Reinhard Busse. 2016. Descriptive, cross-country analysis of the nurse practitioner workforce in six countries: size, growth, physician substitution potential. *BMJ Open* 6, 9 (2016), e011901.
[77] Hannah McCarthy, Henry WW Potts, Abigail Fisher, et al. 2021. Physical activity behavior before, during, and after COVID-19 restrictions: longitudinal smartphone-tracking study of adults in the United Kingdom. *J. Medical Internet Res.* 23, 2 (2021), e23701.
[78] Barbara Millet, Andrew P Carter, Kenneth Broad, Alberto Cairo, Scotney D Evans, and Sharanya Majumdar. 2020. Hurricane risk communication: visualization and behavioral science concepts. *Weather Clim Soc* 12, 2 (2020), 193–211.
[79] Minnesota Agricultural Product 2023. Minnesota Agricultural Product. https://www.nass.usda.gov/Statistics_by_State/Minnesota/Publications/County_Estimates.
[80] Movie Dataset 2023. Movie Dataset: Budgets, Genres, Insights. https://www.kaggle.com/datasets/utkarshx27/movies-dataset.
[81] Kushin Mukherjee, Brian Yin, Brianne E Sherman, Laurent Lessard, and Karen B Schloss. 2021. Context matters: A theory of semantic discriminability for perceptual encoding systems. *IEEE Trans Vis Comput Graph* 28, 1 (2021), 697–706.
[82] Hendrik Müller and Aaron Sedley. 2015. Designing surveys for HCI research. In *Extended Abstracts of the 2015 ACM SIGCHI Hum Factor Comput Syst.* 2485–2486.
[83] Tamara Munzner. 2009. A nested model for visualization design and validation. *IEEE Trans Vis Comput Graph* 15, 6 (2009), 921–928.
[84] Tamara Munzner. 2014. *Visualization analysis and design.* CRC press.
[85] Museum visitors dataset 2019. Museum visitors dataset. https://www.opendatanetwork.com/dataset/data.lacity.org/trxm-jn3c.
[86] Pranathi Mylavarapu, Adil Yalcin, Xan Gregg, and Niklas Elmqvist. 2019. Ranked-List Visualization: A Graphical Perception Study. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*. 192.
[87] National Health 2023. National Health Expenditure Data. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata.
[88] NOAA Monthly 2023. NOAA Monthly U.S. Climate Divisional Database. https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00005.
[89] Chris North. 2006. Toward measuring visualization insight. *IEEE Comput Graph Appl* 26, 3 (2006), 6–9.
[90] Brian Ondov, Nicole Jardine, Niklas Elmqvist, and Steven Franconeri. 2018. Face to face: Evaluating visual comparison. *IEEE Trans Vis Comput Graph* 25, 1 (2018), 861–871.
[91] Evan M Peck, Sofia E Ayuso, and Omar El-Etr. 2019. Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. In *Proc. 2019 ACM SIGCHI Hum Factor Comput Syst.* 1–12.
[92] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Intelligence Analysis*, Vol. 5. 2–4.
[93] Catherine Plaisant, Jean-Daniel Fekete, and Georges Grinstein. 2007. Promoting insight-based evaluation of visualizations: From contest to benchmark repository. *IEEE Trans Vis Comput Graph* 14, 1 (2007), 120–134.
[94] PM 2.5 dataset 2016. PM 2.5 dataset. https://data.cdc.gov/browse/select_dataset?tags=pm2.5.
[95] Annie Preston, Maksim Gomov, and Kwan-Liu Ma. 2019. Uncertainty-aware visualization for analyzing heterogeneous wildfire detections. *IEEE Comput Graph Appl* 39, 5 (2019), 72–82.
[96] Product Clustering Dataset 2020. Product Clustering Dataset. https://www.kaggle.com/datasets/lakritidis/product-classification-and-categorization.
[97] Programming 2023. Programming Languages Dataset. https://www.kaggle.com/datasets/muhammadkhalid/most-popular-programming-languages-since-2004.
[98] Protein Products 2022. Protein Products Market Dataset. https://www.statista.com/topics/4232/protein-market.
[99] Ghulam Jilani Quadri, Jennifer Adorno Nieves, Brenton Wiernik, and Paul Rosen. 2022. Automatic Scatterplot Design Optimization for Clustering Identification. *IEEE Trans Vis Comput Graph* (2022).

[100] Ghulam Jilani Quadri and Paul Rosen. 2020. Modeling the influence of visual density on cluster perception in scatterplots using topology. *IEEE Trans Vis Comput Graph* 27, 2 (2020), 1829–1839.
[101] Ghulam Jilani Quadri and Paul Rosen. 2021. A survey of perception-based visualization studies by task. *IEEE Trans Vis Comput Graph* (2021).
[102] Ronald Rensink and Gideon Baldridge. 2010. The perception of correlation in scatterplots. *Comput Graph Forum* 29, 3 (2010), 1203–1210.
[103] Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, and Max Roser. 2020. Coronavirus pandemic (COVID-19). *Our World in Data* (2020).
[104] Hannah Ritchie, Pablo Rosado, and Max Roser. 2023. Agricultural Production. *Our World in Data* (2023). https://ourworldindata.org/agricultural-production.
[105] Hannah Ritchie, Max Roser, and Pablo Rosado. 2020. CO_2 and Greenhouse Gas Emissions. *Our World in Data* (2020). https://ourworldindata.org/co2-and-greenhouse-gas-emissions.
[106] Brian Rogers. 2022. Cues, clues and the cognitivisation of perception: Do words matter? *Perception* 51, 5 (2022), 295–299.
[107] Paul Rosen and Ghulam Jilani Quadri. 2020. Linesmooth: An analytical framework for evaluating the effectiveness of smoothing techniques on line charts. *IEEE Trans Vis Comput Graph* 27, 2 (2020), 1536–1546.
[108] Ian Ruginski, Alexander P Boone, Lace M Padilla, Le Liu, Nahal Heydari, Heidi S Kramer, Mary Hegarty, William B Thompson, Donald H House, and Sarah H Creem-Regehr. 2016. Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation* 16, 2 (2016), 154–172.
[109] Daniel Russell, Mark J Stefik, Peter Pirolli, and Stuart K Card. 1993. The cost structure of sensemaking. In *INTERACT and ACM SIGCHI Hum Factor Comput Syst*. 269–276.
[110] Bahador Saket, Alex Endert, and Çağatay Demiralp. 2018. Task-based effectiveness of basic visualizations. *IEEE Trans Vis Comput Graph* 25, 7 (2018), 2505–2512.
[111] Bahador Saket, Arjun Srinivasan, Eric Ragan, and Alex Endert. 2018. Evaluating interactive graphical encodings for data visualization. *IEEE Trans Vis Comput Graph* 24 (2018).
[112] Purvi Saraiya, Chris North, and Karen Duca. 2005. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Trans Vis Comput Graph* 11, 4 (2005), 443–456.
[113] Purvi Saraiya, Chris North, Vy Lam, and Karen A Duca. 2006. An insight-based longitudinal study of visual analytics. *IEEE Trans Vis Comput Graph* 12, 6 (2006), 1511–1522.
[114] Michael Sedlmair, Andrada Tatu, Tamara Munzner, and Melanie Tory. 2012. A taxonomy of visual cluster separation factors. *Comput Graph Forum* 31, 3pt4 (2012), 1335–1344.
[115] Priti Shah and Eric G Freedman. 2011. Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Top Cogn Sci* 3, 3 (2011), 560–578.
[116] Stephen Smart and Danielle Albers Szafir. 2019. Measuring the Separability of Shape, Size, and Color in Scatterplots. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*. 669. https://doi.org/10.1145/3290605.3300899
[117] Hayeong Song and Danielle Albers Szafir. 2018. Where's My Data? Evaluating Visualizations with Missing Data. *IEEE Trans Vis Comput Graph* (2018). https://doi.org/10.1109/TVCG.2018.2864914
[118] Arjun Srinivasan, Matthew Brehmer, Bongshin Lee, and Steven Drucker. 2018. What's the Difference?: Evaluating Variations of Multi-Series Bar Charts for Visual Comparison Tasks. In *Proc. ACM SIGCHI Hum Factor Comput Syst (CHI)*. 304.
[119] Stock Market Dataset 2020. Stock Market Dataset. https://www.kaggle.com/datasets/jacksoncrow/stock-market-dataset.
[120] Chase Stokes, Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti A Hearst. 2022. Striking a balance: reader takeaways and preferences when integrating text and charts. *IEEE Trans Vis Comput Graph* 29, 1 (2022), 1233–1243.
[121] Supermarket sales 2019. Supermarket sales. https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales.
[122] Danielle Albers Szafir. 2018. Modeling color difference for visualization design. *IEEE Trans Vis Comput Graph* 24, 1 (2018), 392–401.
[123] Danielle Albers Szafir, Rita Borgo, Min Chen, Darren J Edwards, Brian Fisher, and Lace Padilla. 2023. *Visualization Psychology*. Springer Nature.
[124] Justin Talbot, Vidya Setlur, and Anushka Anand. 2014. Four experiments on the perception of bar charts. *IEEE Trans Vis Comput Graph* 20, 12 (2014), 2152–2160.
[125] Top Spotify 2019. Top Spotify Songs. https://www.kaggle.com/datasets/leonardopena/top-spotify-songs-from-20102019-by-year.
[126] Chin Tseng, Ghulam Jilani Quadri, Zeyu Wang, and Danielle Albers Szafir. 2023. Measuring Categorical Perception in Color-Coded Scatterplots. In *Proc. 2023 ACM SIGCHI Hum Factor Comput Syst (CHI)*.
[127] Edward Tufte. 1985. The visual display of quantitative information. *J. Healthc. Qual.* 7, 3 (1985), 15.
[128] Turtles 2013. Turtles Dataset. https://catalog.data.gov/dataset/turtle-reproductive-ecology-data-new-mexico-2012-2013.
[129] Unemployment dataset 2022. Unemployment dataset. https://www.kaggle.com/datasets/pantanjali/unemployment-dataset.
[130] UNICEF and Gallup. 2021. The Changing Childhood Project. https://changingchildhood.unicef.org/.
[131] US Health Insurance Dataset 2019. US Health Insurance Dataset. https://www.kaggle.com/datasets/teertha/ushealthinsurancedataset.
[132] Wieske van Zoest and Mieke Donk. 2004. Bottom-up and top-down control in visual search. *Perception* 33, 8 (2004), 927–937.
[133] Manuela Waldner, Alexandra Diehl, Denis Gracanin, Rainer Splechtna, Claudio Delrieux, and Kresimir Matkovic. 2019. A Comparison of Radial and Linear Charts for Visualizing Daily Patterns. *IEEE Trans Vis Comput Graph* (2019).
[134] Emily Wall, Meeshu Agnihotri, Laura Matzen, Kristin Divis, Michael Haass, Alex Endert, and John Stasko. 2018. A heuristic approach to value-driven evaluation of visualizations. *IEEE Trans Vis Comput Graph* 25, 1 (2018), 491–500.
[135] Arran Zeyu Wang, David Borland, and David Gotz. 2024. An empirical study of counterfactual visualization to support visual causal inference. *Inf Vis* (2024), 14738716241229437.
[136] Keke Wu, Emma Petersen, Tahmina Ahmad, David Burlinson, Shea Tanis, and Danielle Albers Szafir. 2021. Understanding data accessibility for people with intellectual and developmental disabilities. In *Proc. 2021 ACM SIGCHI Hum Factor Comput Syst (CHI)*. 1–16.
[137] Tiffany Wun, Jennifer Payne, Samuel Huron, and Sheelagh Carpendale. 2016. Comparing bar chart authoring with Microsoft Excel and tangible tiles. *Comput Graph Forum* 35, 3 (2016), 111–120.
[138] Cindy Xiong, Cristina R Ceja, Casimir JH Ludwig, and Steven Franconeri. 2019. Biased average position estimates in line and bar graphs: Underestimation, overestimation, and perceptual pull. *IEEE Trans Vis Comput Graph* 26, 1 (2019), 301–310.
[139] Cindy Xiong, Lisanne Van Weelden, and Steven Franconeri. 2019. The curse of knowledge in visual data communication. *IEEE Trans Vis Comput Graph* 26, 10 (2019), 3051–3062.
[140] Youtube dislike dataset 2021. Youtube dislike dataset. https://www.kaggle.com/datasets/dmitrynikolaev/youtube-dislikes-dataset.
[141] Jeff Zacks and Barbara Tversky. 1999. Bars and lines: A study of graphic communication. *Mem Cognit* 27 (1999), 1073–1079.

A APPENDIX A: PILOT STUDY INFORMATION

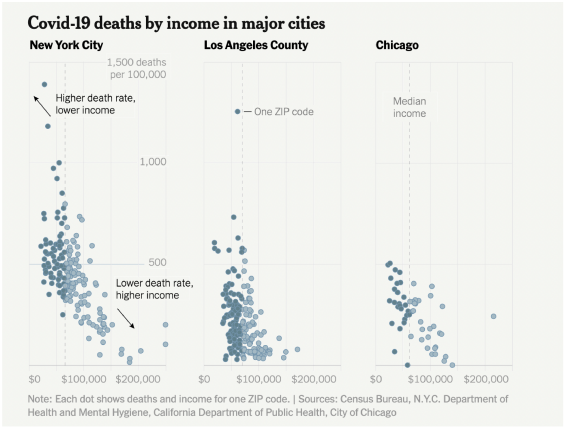
Appendix A introduced more details about our methodology and pilot study.

A.1 Pilot Study

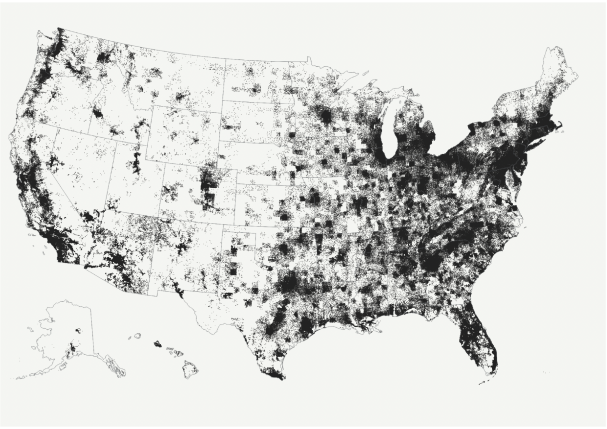
Task: In this pilot study, we conducted an experiment to investigate what people intuitively see in data visualizations. We asked participants to describe what they saw in the given graphs. We recorded their verbal responses to evaluate whether the patterns participants saw matched the designers' goals as stated in the constituent articles.

Study Setup: We selected five different visualizations—one map, one area chart, two scatterplots (one being single class and the other multi-class), and one line graph, as listed in Table 2—from New York Times (see examples in Figure 10). We recruited 10 participants (six males, four females; 21-44 years of age) with varying levels of familiarity with visualizations, two of whom were academic research experts and the other eight more casual visualization users (five from academia and three from industry). We presented the visualizations in random sequential orders and asked the participants to *"describe what you see in the graph"*. The entire study took no more than 10 minutes for each participant. We recorded their verbal responses and later evaluated whether the patterns that participants saw matched the designers' goals as stated in the constituent articles.

Result: We discuss the results of two graphs from Figure 10 categorizing participant responses as 1) understanding the graphs and 2) comments on the design, visualization, or analysis. Here we,



(a) One million Covid-19 deaths.



(b) Covid-19 deaths by income

Figure 10: Examples from New York Times. Graph (a) shows a USA map plotting one million Covid-19 deaths. Graph (b) demonstrates the COVID-19 deaths by income.

Table 2: Information on the five graphs from New York Times used in the experiment. Links to articles and graphs are embedded in the description texts.

ID	Visualization	Description	Visual Encoding
V1	Map	Covid-19 1 million death	Color saturation
V2	Area chart	Covid-19 death per wave	Area
V3	Scatterplot	Covid-19 deaths by income	Position, Color
V4	Scatterplot	Reading time for Popular text	Position
V5	Line graph	Use to tobacco products	Color, stroke or shading

are providing the detailed description only **map** (see Figure 10 and Table 2).

The majority of participants perceived density-based information from the map. However, participants did not identify the primary objective of the map: to convey that the COVID death crisis began in cities and spread to rural areas. Participants generally saw the COVID map as indicating regions of high and low density. Three participants (experts and a Ph.D. student) described the map as pointing to the highly-concentrated region (see Figure 10 (a)), four noted a skew of data towards the east coast, and the remaining 3 participants instead noted that the east and the far west had a greater quantity or higher density of data. Four participants pointed out that the darker region represented higher quantities while the lighter region represented lesser quantities.

The participants who were experts talked in detail about how the design was misleading. For example, they mentioned that the graph Figure 10 (a) lacked a scale to help them further interpret the approximate number of deaths in a given area. Five participants mentioned the lack of scale in the graph, and two participants asked for more information on the context of the graph and data. Six participants felt this graph missed conveying more information regarding quantity and scale.

The designer's objective was to show that income is a predictor of Covid-19 mortality by showing a correlation between income and both death and vaccination rates in major cities. In line with this goal, six people described the graph as showing that Covid-19 deaths are more concentrated in lower-income regions of three cities (see Figure 10 (a)), while one participant focused on correlations between death and income from NYC alone. Two participants compared the overall number of deaths across cities. Only two participants correctly interpreted data points as representing the death and income for a given zip code, with four participants instead seeing dot color as representing incomes as being either above or below the median.

For Figure 10, the responses varied among users when reflecting on the utility of the visualization. People felt the graphs were overloaded (three participants), complicated (two participants), the missing legend on colors (two participants), and were confused by the fact that each graph annotates different information (one participant).

Coding for the Formal Study: Based on the participants' response and axial coding, we decided to code the formal study's response on– *comprehension match, statistical tasks, design critiques, and data critiques.* Additionally, we considered the graph composition of juxtaposition in the formal study.

A.2 More Details of Formal Study

Low-level Task Definition: Only those which are mentioned by participants.

(1) Correlation is defined as "*Given a set of data cases and two attributes, determine useful relationships between the values of those attributes.*" The keywords from response used to code are: *correlation, positive correlation, negative correlation, association, relation, not fully linear, prediction.*

(2) Trend is described as "*Given a set of temporal data cases, determine a pattern.*" The keywords from response used to code are: *increase, decrease, rate of change, change, drastic change, upward trend, downward trend, higher, lower.*
(3) Extremum is defined as "*Find data cases possessing an extreme value of an attribute over its range within the data set.*" The keywords from response used to code are: *maximum, minimum, peak, valley, high, low.*
(4) Compute Derive Value is described as "*Given a set of data cases, compute an aggregate numeric representation of those data cases.*" The keywords from response used to code are: *count, duration, mean.*
(5) Cluster is described as "*Given a set of data cases, find clusters of similar attribute values.*" The keywords from response used in code are: *cluster, grouped.*
(6) Characterize Distribution is described as "*Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute's values over the set.*" The keywords from response used to code are: *even distribution, normal distribution, random, scattered data points, variability.*
(7) Anomalies is described as "*identify any anomalies within a given set of data cases concerning a given relationship or expectation, e.g., statistical outliers*" The keywords from response used to code is: *outlier.*
(8) Determine Range is described as "*Given a set of data cases and an attribute of interest, find the span of values within the set.*" The keywords from response used to code are: *from-to, over.*
(9) Compare is described as "*Given a set of data cases, compare any attributes within and between relations of the given set of data cases for a given relationship condition.*" The keywords from the response used to code is: *compare.*

APPENDIX B: METADATA AND RESULTS

Appendix B introduced the codebook, summarization of study results, the overall metadata, and several instances with different visual encodings.

Table 4 shows details of all the participants' demographics in our study. Table 5 shows the summary results of participants' comprehension per participant's field of education (students) and working professionals. Table 6 shows the summary results of participants' comprehension match per match level in the Description task. Table 7 illustrates the abbreviations used in the rest results of our analysis. Table 8 shows all of the metadata of our study. Figure 11 shows the summary counts of questions participants responded to for the Question Task per graph type and data type.

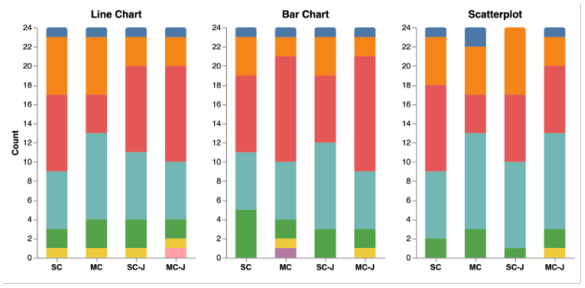


Figure 11: Summary counts of questions participants responded to the Question Task.

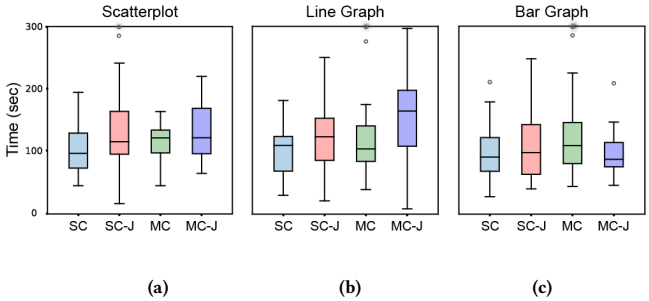


Figure 12: Average response time for four visualization types. SC: single-class; MC: multi-class; SC-J: single-class juxtaposed; MC-J: multi-class juxtaposed. The dark transparent shadows mean this value exceeds the upper bound (300s), and a larger range of shadows means more outliers. As we can see participants took more time to comprehend multi-class graphs in all three graphs and multi-class-juxtaposed in scatterplot and line graphs. The maximum average comprehension time is for multi-class-juxtaposed line graphs.

Table 3: Codebook example using a spreadsheet to code all users' responses.

PID	Name	Graph	Data	Composition	Response-Description	Response-Question	Comprehension	Statistics	Design	Data	Other
X	IMDB	Line graph	Single class	Non-juxtaposed	There is a positive association between gross movie sales and IMDB. The association does not seem to be completely linear, but is more dependent on the rating.	What's the association between gross movie sales and rating? Whether rating impact gross movie sales?	Complete Match	Association ->correlation	NA	NA	NA

Table 4: Summary of 24 participants' backgrounds and visualization experiences in the study. The top rows show participants' backgrounds, where 19 of 24 are students and 5 of 24 are working professions. The bottom rows are their reported visualization experiences.

Fields	**Count**	**Major / Profession**
Humanities	6/24	Social Work, Law, European Studies, Journalism, Economics, Anthropology
Computing	6/24	Computer Science (2), Health Informatics, Mathematics, Biomedical Engineering, Quantitative Biology
Science	3/24	Psychology, Chemistry, Environmental Science
Health	4/24	Nursing (2), Public Health, Pharmacology
Professions	5/24	Health and Physical (2), Communication and Media Relations, Medical Technician, Educational Technology,
Experience	**Count**	**Details**
Extensive	1/24	Almost an expert in visualization
Casual	8/24	Usually use visualizations in projects
Little	13/24	Created some visualizations before
No	2/24	Almost didn't use visualizations before

Table 5: Participant comprehension summary results as per participant's field of education (students) and working professional.

Field of Study	CM	GM	PM	NM
Computing & Related	31	8	15	6
Science	31	2	7	8
Health	17	3	10	18
Humanities	21	9	23	19
Working Professional	17	4	12	27
Count (Total is **288**)	117	26	67	78

Table 6: Participants' comprehension match summary by categories on the Description task (see Sect. 3.1 in the paper). Each participant saw 12 different graphs in the study. Ranked by the number of Complete Match.

Assigned ID	Comprehension				Total
	CM	GM	PM	NM	
05	11	1	0	0	12
19	11	0	1	0	12
22	11	0	1	0	12
09	10	1	1	0	12
10	10	0	2	0	12
20	9	1	2	0	12
08	9	1	1	1	12
17	8	2	1	1	12
12	7	2	1	2	12
01	6	1	4	1	12
13	6	1	2	3	12
18	4	2	3	3	12
16	4	1	3	4	12
07	3	2	3	4	12
11	3	0	2	7	12
06	2	1	4	5	12
21	2	0	4	6	12
14	1	2	6	3	12
24	0	4	3	5	12
02	0	3	9	0	12
03	0	1	7	4	12
04	0	0	4	8	12
23	0	0	2	10	12
15	0	0	1	11	12
Total	117	26	67	78	288

	Class Type				Comprehension Match				Visual Encoding											Statistics								
Abbr.	SC	SC-J	MC	MC-J	CM	GM	PM	NM	C	D	P	S	T	W	BH	BL	CS	SD		Corr	Comp	Tr	Clu	CDV	DR	Xtrm	CD	Anom
Name	Single-Class	Single-Class Juxtapose	Multi-Class	Multi-Class Juxtapose	Complete Match	General Match	Partial Match	No Match	Color	Dots	Position	Size	Texture	Text	Bar Height	Bar Length	Color + Semantics	Stroke Dash		Correlation	Comparison	Trend	Cluster	Compute Derived Value	Determine Range	Extremum	Characterize Distribution	Anomaly

Table 7: Summary of abbreviations used in Table 8.

Graph-type	Data-type	ID	Dataset	Encoding	Count	Comprehension				Critique		Statistics
						CM	GM	PM	NM	Dataset	Design	
Scatterplot	SC	1	*Boeing*	P	5	3	0	1	1	3	0	Corr (3), CD (2), Tr (1)
		2	*CO_2*	P	6	2	1	1	2	0	2	Corr (2), Clu (1), Comp (1)
		3	*Medical*	P	4	1	0	2	1	0	0	Xtrm (2), CDV (1)
		4	*Sunny day*	P	4	2	2	0	0	1	0	Tr (2), Xtrm (1), CD (1)
		5	*Turtle*	P	5	4	0	0	1	4	0	Corr (4), Clu (1), DR (1), CDV (1)
	MC	6	*Activity*	C, P	6	1	2	2	1	2	1	Xtrm (1), CDV (1), Clu (2), Comp (1)
		7	*Car models*	C, S, P	7	4	0	1	2	1	0	Corr (6), Clu (2), DR (1), Xtrm (2)
		8	*Horsepower*	C, P	4	2	1	0	1	1	2	Corr (4), Tr (1)
		9	*TempChange*	C, P	4	1	0	3	0	1	1	Tr (5)
		10	*Penguin*	C, P	3	3	0	0	0	2	0	Corr (2), Comp (1)
	SC-J	11	*CPU*	C, P	6	2	0	2	2	2	0	Corr (4), DR (1), Tr (1)
		12	*Horsepower*	C, P	5	1	1	2	1	2	0	Corr (4), Comp (1), CD (1)
		13	*Penguin*	C, P	5	1	0	2	2	1	1	Corr(3), Comp(1)
		14	*PM 2.5*	C, P	2	1	0	1	0	2	0	Clu (1), CD (1), Comp (1)
		15	*Tumor*	C, P	6	3	0	1	2	5	0	Comp (3), Clu (1), CD (2), Corr (1)
	MC-J	16	*Age-BMI*	C, P	4	1	0	0	3	1	2	CD (1), Corr (1)
		17	*Car models*	C, P, S	5	0	0	1	4	0	0	Comp (1)
		18	*Weather*	CS, P	6	2	1	1	2	4	1	Tr (6), Xtrm (1)
		19	*Insurance*	C, P, S	3	1	0	2	0	2	1	Corr (2), Comp (1)
		20	*Titanic*	C, P	6	4	1	1	0	3	1	Comp (5), Corr (1)
Line graph	SC	21	*Google*	P, D	6	4	1	0	1	3	0	Tr (6)
		22	*Gov bond*	P	5	2	1	2	0	3	0	Tr (5), Xtrm (3)
		23	*IMDB*	P	3	2	0	1	0	1	0	Corr (4), Tr (2)
		24	*Sales*	P	5	3	1	1	0	5	0	Xtrm (1), Tr (6), CDV (1), Comp (1)
		25	*Stock red*	C, P, SD	5	1	0	0	4	1	0	Tr (3)
	MC	26	*5-gov*	C, P	6	4	0	1	1	4	2	Tr (5), Comp (1), Xtrm (1)
		27	*EEG*	C, P	2	0	0	0	2	0	0	Xtrm (2), Tr (1)
		28	*Titles*	C, W, P	7	1	0	4	2	2	1	Tr (2), CDV (2), Xtrm (1), Corr (1)
		29	*Stock red*	C, SD, P	4	1	0	1	2	0	2	Corr (1), Tr (1)
		30	*Stock*	SD, P	5	2	0	1	2	1	1	Tr (4), Comp (1)
	SC-J	31	*Airline*	C, P	6	2	1	1	2	2	0	Tr (4)
		32	*IMDB*	C, P	4	4	0	0	0	2	0	Corr (3), Comp (1), Tr (1)
		33	*Visitors*	P	4	2	0	0	2	2	0	CDV (1), Comp (2)
		34	*Spotify*	C, P	5	2	0	1	2	1	1	CDV (2), Tr (3)
		35	*Stock*	C, P	5	1	0	1	3	1	1	Xtrm (2), Tr (1),
	MC-J	36	*Activity-covid*	C, P	6	3	1	2	0	0	0	Tr (5), Comp (3), CDV (1)
		37	*Unemployed*	C, P	3	2	0	0	1	2	0	Tr (3), Comp (1)
		38	*Covid-3*	C, P	5	1	0	1	3	0	3	Comp (2), CDV (1), Tr (1)
		39	*Income*	C, P	8	1	0	5	2	4	0	Tr (8), Xtrm (2), Comp (1)
		40	*Working*	C, P	2	2	0	0	0	2	0	Tr (1), Comp (1)
Bar graph	SC	41	*12-cat products*	CS, BH, P	6	2	1	1	2	1	1	Xtrm (2), Comp (2), CDV (1)
		42	*Afghan*	BH, P	5	2	0	2	1	0	1	Tr (4)
		43	*TempChange*	C, P	5	3	1	0	1	3	1	Tr (6)
		44	*Budget*	BL, P	5	3	1	1	0	5	1	Xtrm (8), Comp (2)
		45	*Electricity*	BH, P	3	1	0	0	2	1	0	Comp (1)
	MC	46	*Movie*	C, P, W	4	1	2	0	1	2	0	Comp (2)
		47	*ProgProficiency*	C, P, BH	5	3	0	2	0	1	0	Corr (2), CDV (1), Comp (2)
		48	*Textured-Alt*	T, P, BL	5	0	1	2	2	2	0	CDV (1), DR (1), Comp (3)
		49	*Textured*	T, P, BH	5	0	2	3	0	0	2	Xtrm (1), CDV (2), Comp (1)
		50	*Visitors*	C, P, BH	5	4	0	0	1	4	0	Xtrm (6), Anom (3), Comp (1)
	SC-J	51	*Revenue*	C, P, BH	3	2	0	1	0	1	1	Xtrm (2), Corr (3), Comp (1)
		52	*Success*	C, P, BL	5	1	0	1	3	0	2	Xtrm (2), Corr (1), Comp (1)
		53	*Avg sales*	C, P, BH	5	3	0	1	1	2	1	Xtrm (6), Comp (1), Tr (1)
		54	*Media*	C, P, BH	5	3	1	1	0	2	1	Comp (3), CDV (1), Xtrm (3), CD (1)
		55	*Twitter*	C, P, BH	6	3	0	2	1	3	0	Comp (5)
	MC-J	56	*Products*	C, P, BH	5	3	0	0	2	1	0	Xtrm (5), Comp (1)
		57	*Farms*	C, P, BL	7	3	1	1	2	2	1	Xtrm (2), Comp (1), CDV (1),
		58	*ProgPopularity*	C, P, BH	2	0	0	1	1	1	0	Xtrm (2), Comp (1)
		59	*Q&A*	C, P, BL	3	1	2	0	0	1	0	Xtrm (2), Comp (1)
		60	*Stock*	C, P, BH	7	0	0	3	4	2	2	Xtrm (2), Tr (2), Comp (2)

Table 8: Summary of the visualizations shown to users in our user study, grouped by 3 graph types and 4 class types and thus forming 12 categories, each containing 5 visualizations. Check Table 7 for definitions of abbreviations. Lime datasets are instances shown in Figure 1 in the paper, orange datasets are in Figure 5, and magenta datasets are in Figure 7.

Table 9: Summary of the references and stated objectives of each stimulus.

ID	Dataset	Reference	Stated Objective
1	*Boeing*	[21]	To show and demonstrate the correlation between the mileage and year of production of a set of Boeing airplanes.
2	*CO_2*	[105]	To show a set of countries' per capita GDP and their annual CO2 emission and demonstrate the quantity of CO2 emissions concentrated in countries with low per capita income.
3	*Medical*	[87]	To show the distribution of age and medical expenditure of a set of people to understand how one relates to another (correlation).
4	*Sunny day*	[42]	To document the pattern in the days where sunny weather occurs and the maximum temperature of each recorded day in a city from January to May and from August to December in 2012.
5	*Turtle*	[128]	To document turtles' mass and their numbers of annuli and demonstrate the correlation between the two.
6	*Activity*	[17]	To plot the distribution of the number of calories burned vs. time from three workout activities. Also, show the clusters formed by the distribution pattern of burnt calories vs. time in 3 different colors.
7	*Car models*	[18]	To show the correlation between car odometer readings and their ages. Additionally, with two categorical encodings (color and point size) compare the car's odometer reading and age with prices, and models. Demonstrate the correlation between the factors and also see how one predicts the car price for a given age based on the odometer reading. The distribution or relation varies between different models.
8	*Horsepower*	[7]	To show the correlation between miles per gallon and horsepower of cars and also compared the distribution and correlation of these two variables for cars manufactured in 3 different countries.
9	*TempChange*	[88]	To show and compare the changes in US temperature over roughly one and a half centuries using dots whose positions and colors correspond to the change in temperature it indicates. It easily indicates how temperature has risen in one and half centuries.
10	*Penguin*	[56]	To show the correlation between body mass and flipper length of 3 species of penguins. It shows a positive correlation between mass and flipper length. Penguin species can be identified by physical appearance (mass length) and each of their distribution varies.
11	*CPU*	[32]	To show the correlation between time vs no of CPU/core and compare between 6 distinct architectures with six juxtaposed graphs. We showed the completion time for the same task on different numbers of cores, with each graph showing only CPUs of the same architecture.
12	*Horsepower*	[7]	To show and compare the miles per gallon and horsepower of cars manufactured in 3 different countries with three different Scatterplots, with each graph showing cars manufactured in different countries. We want to represent the correlation and distribution between MPG and horsepower for the given car.
13	*Penguin*	[56]	To show the correlation between body mass and flipper length of 3 species of penguins. It shows a positive correlation between mass and flipper length. Penguin species can be identified by physical appearance (mass length) and each of their distribution varies.
14	*PM 2.5*	[94]	To show and compare hundreds of PM 2.5 density readings recorded on two days in 3 different cities, with each separate graph showing readings from the different cities. Understand individual city distribution and compare three.
15	*Tumor*	[37]	To show and compare the radius and concavity of benign and malignant cancer tumors in two separate scatterplots, with each graph showing either benign tumors or other malignant ones.
16	*Age-BMI*	[131]	To show and compare the age and BMI of a set of clients from 4 different regions show the distribution between patient's age and their BMI and compare the same for the information in four different regions in separate scatterplots. We also demonstrated the gender of insurers with different colors. We want to demonstrate the distribution between these factors and compare them among 4 regions.
17	*Car models*	[18]	To show the correlation between car odometer readings, and car ages, in two side-by-side scatterplots. Compare the two graphs on the price of the car, and models of a set of cars, with one graph showing cars' prices and the other showing their models.
18	*Weather*	[54]	To document and demonstrate the trend/pattern in days for 5 types of weather occur and the maximum temperature of each recorded day in a city from 1) January to May (one graph) and from 2) August to December in 2012 (another graph). We want to show the trend in temperature for the two seasons.

19	*Insurance*	[131]	To show the distribution between patients' age and their BMI and compare the same for the information in four different regions in separate scatterplots. We separately encoded the gender of the insurer and their insurance premium amount. We want to demonstrate the distribution between these four factors and compare them among 4 regions.
20	*Titanic*	[33]	To show and compare the distribution of survival/death vs. fare of Titanic passengers separated into 2 sets with one consisting only of males and the other only of females.
21	*Google*	[119]	To document the stock price of Google roughly from 2004 to 2010 in order to show patterns in stock price changes.
22	*Gov bond*	[44]	To document the pattern in the yields of long-term government bonds over roughly 6 decades.
23	*IMDB*	[59]	To document the correlation between movies' IMDB ratings and their revenues made in the US.
24	*Sales*	[98]	To document the pattern of the sales figures of a protein product from 2006 to 2007.
25	*Stock red*	[119]	To document the trend (high and low) in the price of a stock over May 2021 and highlight in red days on which its price hit below a threshold.
26	*5-gov*	[44]	To show the pattern and distribution of annual performance of government bonds issued by 5 countries and highlight years in which the performance hit below 0% return.
27	*EEG*	[107]	To show the pattern of EEG Readings of 500 samples on 3 channels and compare their similarities by superimposition.
28	*Titles*	[47]	To show and compare the number of grand slam titles owned by 5 tennis players at different ages.
29	*Stock red*	[119]	To show the trend in prices of 4 stocks in May 2021 and highlight days on which a certain stock's price hit below a threshold of 52-week-low. The price below the threshold is shown in red color.
30	*Stock*	[119]	To show the pattern of 5 stock prices over roughly 10 years from 2000 to 2010. The objective is to show how each stock performed individually and comparatively with others.
31	*Airline*	[3]	To show and compare the changes in revenue of 4 airline companies over one year and demonstrate when sales for all went down (during COVID).
32	*IMDB*	[59]	To show and compare the pattern in revenues made in the US vs. worldwide of a set of movies with varying IMDB ratings. Both graphs show positive correlations.
33	*Visitors*	[85]	To show and compare patterns in the number of visitors received by 4 museums on a set of days. Show outlier cases.
34	*Spotify*	[125]	To show and compare the streaming pattern of the 3 popular songs on Spotify over April 2017.
35	*Stock*	[119]	To show and compare the prices of 4 different stocks over roughly a decade by the juxtaposition between 2000-2010.
36	*Activity-covid*	[77]	To show and compare the change in time people from 4 age groups spent on different activities before and after the Covid-19 pandemic. The graphs are shown in three categories: grooming, exercise, and mobile.
37	*Unemployed*	[129]	To show and compare patterns in the yearly count of people unemployed in specific industries over roughly 3 decades in three different categories.
38	*Covid-3*	[103]	To show and compare patterns in the monthly count of (Covid-19) cases, deaths, and hospitalizations in 3 counties over one year.
39	*Income*	[68]	To show correlation and compare two types of employees' average weekly incomes and their growth rate over 16 years. There is a positive correlation between dollars per week and years, but the same distribution is not persistent in a change of income from the previous year.
40	*Working*	[76]	To show the pattern and compare the yearly percentage of the population in the workforce of six countries over roughly 3 decades, demonstrated by two graphs: Europe and North America.
41	*12-cat products*	[96]	To show produces output(max-min) belonging to 12 categories using color choices that resemble the colors of their corresponding produce categories.
42	*Afghan*	[2]	To show the census results and trend in Afghanistan population over roughly 7 decades with a design that makes it easy to tell a trend.
43	*TempChange*	[88]	To show the trend in the changes of US temperature over roughly one and a half-century using 2 colors indicating an increase or decrease in temperature to enable users to easily tell a trend.
44	*Budget*	[80]	To show and compare the budget of different genres (min-max) of films by indicating films with very high budgets.
45	*Electricity*	[57]	To show the pattern and compare the prices of monthly electricity bills paid by a household over a year and identify the period where the bill is higher or lower.

46	*Movie*	[80]	To show and compare the average revenues made in the US and worldwide by movies of various types with texts that increase readability.
47	*ProgProficiency*	[97]	To show pattern and compare 4 groups of people's proficiency in various programming languages. The proficiency increases with their education and experience.
48	*Textured-Alt*	[140]	To show and compare (min-max) the count of likes and dislikes received by 4 YouTube videos, categories each belonging to a distinct channel.
49	*Textured*	[140]	To show and compare (min-max)the count of likes and dislikes received by 4 YouTube videos, each belonging to a different channel.
50	*Visitors*	[85]	To show the pattern and compare (min-max) the average monthly count of visitors received by 4 museums in a given year.
51	*Revenue*	[80]	To show patterns and compare the budget and the gross revenue of movies of various types/genres.
52	*Success*	[130]	To show and compare various countries' opinions on how much a specific factor plays a part in contributing to an individual's success.
53	*Avg sales*	[121]	To show patterns and compare the sales figures or average monthly sales and distribution and patterns of 4 locations over a year.
54	*Media*	[34]	To show patterns and compare the time a user spends using social media apps and entertainment apps in a given week.
55	*Snapchat&Ins*	[34]	To show the screen time pattern a user spends on Snapchat and Instagram in a given week and compare patterns between them side by side.
56	*Products*	[104]	To show min-max and compare the output of various foods belonging to 3 categories using colors that resemble their corresponding food's color in real life.
57	*Farms*	[79]	To show and compare (min-max) the output harvest (produce) of 4 types of fruits in 6 locations.
58	*ProgPopularity*	[97]	To show the pattern and compare the monthly market share of 6 programming languages in 2021, demonstrated by quarters.
59	*Q&A*	[4]	To show the pattern and compare the percentage of respondents showing various attitudes towards a set of problems
60	*Stock*	[119]	To show and compare the prices of 5 stocks in the January of 2005-2008. Google stock has the highest price.