Designing with Pictographs: Envision Topics without Sacrificing Understanding

Alyxander Burns, Cindy Xiong, Steven Franconeri, Alberto Cairo, and Narges Mahyar

Abstract—Past studies have shown that when a visualization uses pictographs to encode data, they have a positive effect on memory, engagement, and assessment of risk. However, little is known about how pictographs affect one's ability to understand a visualization, beyond memory for values and trends. We conducted two crowdsourced experiments to compare the effectiveness of using pictographs when showing part-to-whole relationships. In Experiment 1, we compared pictograph arrays to more traditional bar and pie charts. We tested participants' ability to generate high-level insights following Bloom's taxonomy of educational objectives via 6 free-response questions. We found that accuracy for extracting information and generating insights did not differ overall between the two versions. To explore the motivating differences between the designs, we conducted a second experiment where participants compared charts containing pictograph arrays to more traditional charts on 5 metrics and explained their reasoning. We found that some participants preferred the way that pictographs allowed them to envision the topic more easily, while others preferred traditional bar and pie charts because they seem less cluttered and faster to read. These results suggest that, at least in simple visualizations depicting part-to-whole relationships, the choice of using pictographs has little influence on sensemaking and insight extraction. When deciding whether to use pictograph arrays, designers should consider visual appeal, perceived comprehension time, ease of envisioning the topic, and clutteredness.

Index Terms—Infographics, pictographs, design, graph comprehension, understanding, casual sensemaking.

1 Introduction

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

Typically in data visualization, data-driven decision making processes are referred to as **sensemaking**. The most commonly used model of sensemaking, proposed by Pirolli and Card, contains both a foraging loop where data are sought, searched, filtered, and read as well as a sensemaking loop where those data are used to iteratively construct a mental model [50]. This process typically involves time-consuming analysis and re-analysis, with the intention of finding analytical "ah-ha" moments. It is also often conducted by expert analysts who have access to the original data and the resources to independently analyze it [50].

In contrast, the everyday decisions about hurricane paths or COVID-19 might be considered *casual* sensemaking, where the general public understands, reflects, and makes decisions based on information provided by casual visualizations [52] without deep and time-consuming research using raw data. It is therefore more similar to a definition of sensemaking from business and management in which the viewer uses a conceptual model of the world informed by available information "to comprehend, understand, explain, attribute, extrapolate, and predict" [2, 62].

One kind of data visualization often encountered by the general public is the infographic. Although infographics have been used for, by some estimates, as long as bar and pie charts [45], there is no single

- *Alyxander Burns and Narges Mahyar are with the College of Information and Computer Sciences at the University of Massachusetts Amherst. E-mail: [alyxanderbur, nmahyar]@cs.umass.edu*
- *Cindy Xiong and Steven Franconeri are with the Department of Psychology at Northwestern University. E-mail: cxiong@u.northwestern.edu E-mail: franconeri@northwestern.edu*
- *Alberto Cairo is with the School of Communication of the University of Miami. E-mail: a.cairo@miami.edu*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

definition. In the journalism world, the word "infographic" has been historically used to refer to pieces that combine graphical elements with text to convey information [15, 16, 42]. Although infographics in this context might contain data visualizations, they may also contain maps, diagrams, or illustrations combined with text or annotations. Overall, the purpose of an infographic is to present important information in a way that is sensitive to common barriers encountered by the public such as limited time and information overload [45]. Because infographics are used to communicate critical information to the public, they could have a strong impact on the way that everyday people understand science and current events. After encountering an infographic, viewers may not seek out and further analyze the data, even if the data is publicly available, because of a lack of the motivation, time, or skill (for example). In these cases, the viewer must come to conclusions based on the information that they are shown. For example, if a person encounters an infographic about COVID-19 cases in their state while reading, we cannot assume that the person will then seek out the associated data dashboard to do a more in-depth analysis – yet, they might use the graphic as it is presented to make choices about how to safely navigate their community. Therefore, it is important to understand how to design visualizations in a way which communicates information effectively to everyone who encounters it – this is the chief interest of the paper. In other words, if we know that people are using data visualizations to engage in casual sensemaking, then we should be thinking critically about how our designs afford those activities.

To measure the ability of visualizations to afford casual sensemaking or, more traditionally, understanding, we turn to the field of Education for inspiration and operationalize Bloom's taxonomy, following the taxonomy proposed in [13]. Bloom's taxonomy describes 6 aspects of the learning process: knowledge, comprehension, application, analysis, synthesis, and evaluation [7]. Though originally intended as a strict hierarchy, some critiques suggest that while some dependency may exist between the levels, it is not a strict hierarchy [13]. Therefore, we view it instead as 6 complementary skills of differing complexity. We chose to use Bloom's taxonomy because it is commonly used to evaluate learning processes [36] and maps well to the activities necessary for casual sensemaking.

In existing research, infographics show positive empirical findings in terms of memory, engagement, and assessment of risk [9, 24, 27] — particularly when they contain pictographs (simple, iconic pictures that represent a word or topic). Yet, there is little exploration of how pictographs affect and afford different kinds of understanding. Further, there is little work on how the choice to use pictographs affects the personal experience of using infographics across factors such as visual appeal, clutter, and ease of envisioning the topic.

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

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

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

2 Background

2.1 Infographics and relationships with data visualization

For many people, the word "infographic" is a loaded term. The word infographic initially gained traction within the news world to refer to any sort of graphic that was used to display information [16]. As different fields have brought infographics into their practices, the word has changed meaning. A definition used within the communication and design communities similarly requires a combination of text and graphics but stresses the purpose of the chart — they must be created to assist in communication [51]. Alternately, definitions may be very strict, incorporating requirements about the way that the graphic is laid out [67]. In some sources, all data visualizations are considered infographics. This is incorrect for a variety of reasons, but, as it pertains to this paper, infographics are largely intended for effective communication while data visualizations can be used not just to communicate, but for exploration, discovery, and analysis, among other tasks.

Outside of the visualization community, infographics are known for their ability to drive engagement. Many news organizations have teams devoted to developing infographics, where the quality of a graphic is judged by its ability to engage audiences [21]. Within the medical communication community, a study found that when wound care instructions were accompanied by images, patients read the instructions more often, could recall the instructions more accurately, and followed the instructions more often [30]. Another found that infographics that were designed with and for communities of non-native English speakers with low medical literacy were viewed as more likely to influence behavioral change [3].

Within the sphere of medical communication, visual aids are well understood as effective tools to help people understand risk. In that domain, visual aids such as infographics and pictograph arrays are useful for helping individuals with medium to high graph literacy and low numeracy [24]. These infographics have also been effective at encouraging behavioral change and communicating medical information across language barriers [24].

2.2 InfoVis research on infographics

Critics of infographics have argued that embellishments should be avoided, lest they cause distraction from understanding the data presented [64]. However, existing work that compared highly illustrated infographics to informationally equivalent, plain charts found no difference in overall understanding of the data [5]. Further, even when the illustration strictly appeared in place of bars of a bar chart, there was no effect on speed or accuracy [60].

Past research has also revealed memorability to be a strength of infographics, showing that when compared to unembellished charts, participants are able to recall the data, trend, value message, and topic of an infographic more accurately over a long period of time [5]. Further, over a shorter-term, participants were able to remember that they had seen the infographic, could more accurately recall its message, and, when attention was divided, could more accurately recall specific points of data, especially when those infographics were colorful and contained pictographs [5, 8, 9, 27].

Within the data visualization community, infographics have also been shown to be particularly useful for engaging audiences, though an infographic's effectiveness may be linked to its visual appeal [40]. For example, one study found that when news articles were accompanied with visualizations, people who found the infographic visually appealing were more likely to read the accompanying article [19]. Investigations into what makes infographics visually appealing found that the most appealing graphics were those that were very colorful but not visually complex [28]. Some speculate that because infographics are designed to be easy to understand, they are more compelling to share, leading them to reach a wider audience [61].

However, existing work has sometimes also found infographics to be less effective than other techniques. When comparing infographics to data comics, one group found that participants who were shown data comics were able to answer questions more accurately and remember those facts longer than those shown infographics [67]. They also found that, when compared to data comics, infographics were preferred for exploration — a quality seen as effective for building trust that the whole story is being shown [46].

There has also been recent work on infographics-creation tools, particularly aimed at users with limited programming experience. Some of these tools aim to bridge the gap between data analysis and vector graphic tools by allowing the user to assign specific aspects of the illustration to data dimensions via lazy data binding (e.g., [39, 68]). Others are designed to allow users to define custom layouts [58], quickly style unembellished charts [66], and use existing images to copy styles [71] or extend timelines [17]. There are also several tools intended to create infographics specifically for use with personal data, including an end-to-end data collection and visualization generation tool [35] and a tool to modify photographs to fit line charts [48].

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

2.3 InfoVis research on pictographs

Though empirical work on pictographs is fairly limited, the idea of using pictographs for data communication is not new. For example, in the mid 1920s, Otto Neurath, Marie Neurath, and Gerd Arntz created the ISOTYPE system, which used a set of custom pictographs to create data visualizations about social and economic topics [65]. Their designs were guided by 2 simple principles: (1) use pictographs to represent objects and (2) use repetition, not size, to represent quantities, which leads to designs with rows or arrays of pictographs.

Empirical work on bar charts in this style found that when pictographs were used to encode data, they had a positive effect on short-term memorability and engagement [27]. Further, pictographs were found to have no impact on performance in terms of speed or accuracy unless they were used decoratively [27]. More generally, infographics with pictographs have been shown to be more instantly recognizable and to result in more accurate descriptions when recalled from memory [8]. Within medical contexts, pictograph arrays have also been shown to be effective for conveying risk for people with differing levels of graph literacy [23, 25]. Other studies have found pictograph arrays to be less liked and trusted when used to display breast cancer risk when compared to more traditional methods [12].

However, not all pictographs are equally effective. Past work has observed that iconicity can impact both the recall of information and the perception of risk [72]. Additionally, the pictographs might be interpreted in ways the designer did not anticipate (such as when representing categories of items) [3].

The existing work on pictograph arrays is generally positive, but there is much we still do not know. In this paper, we build on existing work in 2 important ways. First, prior work which compares charts with pictographs to those without either uses real charts which are not informationally equivalent (e.g., [9]) or informationally equivalent charts which are not real (e.g., [27]). If we want to know what impact design decisions might have on the audience, then it is important to use charts which are both as realistic as possible and informationally equivalent. Finally, though identifying values in a chart is a critical skill that forms a basis for other more complicated tasks, it does not realistically capture the kinds of understanding-based tasks which readers partake in. Phrased differently, though we have some indication that pictographs do not impact accuracy tasks, we do not yet know what impact pictographs have on understanding beyond these tasks.

2.4 Evaluating sensemaking abilities

A diverse range of casual and professional users are producing and consuming visualizations [59]. While visualizations are increasingly used to communicate important and often sensitive data to the general public, it remains a significant challenge to find appropriate methods to evaluate their effectiveness. We need novel methods to assess what viewers of data visualizations understand. Metrics such as accuracy and response speed have been used as a proxy for knowledge acquisition, but they are unable to capture the full extent of a viewer's insights [44].

North proposed new methods for measuring insights by asking more difficult, open-ended questions [44]. In response to the call for new methods, Mahyar et al. proposed utilizing Bloom's taxonomy to evaluate the depth of users' understanding when engaged with visualization [41]. Bloom's taxonomy is one of the most common learning hierarchies in Education and contains 6-levels of learning objectives: knowledge, comprehension, application, analysis, synthesis, and evaluation [7]. With this system, educators are able to create activities and evaluation questions targeting particular learning outcomes [4, 32].

Casual sensemaking, as defined in this paper, is related to but distinct from the model of sensemaking proposed by Pirolli and Card which is often used in the visualization community [50]. In the Pirolli and Card model, the sensemaking process is made up of the following activities: collecting information, creating a mental model of that information in a way that helps analysis, manipulation of that model to produce an insight, and finally the production of an idea or action based on that insight [50]. This process is not linear and instead involves iteratively looping between the steps until sense is made. We can think of each of the learning objectives in Bloom's taxonomy as a result of a different sensemaking process in the Pirolli and Card model. Viewed this way, lower level objectives in Bloom's taxonomy might require little to no manipulation of a mental model to produce an insight, while higher level objectives may require more iterations of the process.

Bloom's taxonomy has been used in visualization research across design and evaluation processes (e.g., [4, 14, 22]). Adar and Lee proposed a modified version of Bloom's taxonomy and demonstrated that it can be used by designers to frame the information they intended to communicate as a set of learning objectives [1]. Independently, Burns et al. adapted the same taxonomy into an evaluation method where researchers form a set of 6 questions about the visualization(s) of interest [13]. Where the former adaptation might be best used to evaluate learning objectives at levels specifically related to designer intent, the latter evaluates aspects of sensemaking at all 6 levels and may reveal impacts the design has but that the designer did not intend. Because we are not concerned with authorial intent in this work, we operationalize the method proposed by Burns et al. in [13].

3 Experiment 1: Understanding

In infographics, pictographic arrays are often used to depict part-to-whole relationships. We explore how this choice affects the insights that viewers draw from the image, compared to using traditional charts that encode the same information with solid areas. In this experiment, we explore the 3 hypotheses described in Section 3.3.

3.1 Experiment 1: Stimuli

For this experiment, we used 6 pairs of infographics that display part-to-whole relationships. In each pair, one encoded that relationship with solid, geometric areas and the other used pictograph arrays (see Figure and 1 for examples; all 6 pairs can be found in the supplementary materials). We will refer to these two types as "versions" and, for brevity, call the two chart versions "Area" and "Count," respectively.

Our criteria for selecting charts included: diversity of chart type, diversity of topic, and a comparable number of variables. Past research suggests that familiarity with a chart type can influence perceptions of attractiveness and ease of use [54]. Therefore, we wanted our stimuli set to include chart types which were common in media so that participants were unlikely to be encountering a chart type for the first time, but that would vary in familiarity. Our final stimuli set includes 1 bar chart, 1 pie chart, 2 stacked bar charts, 1 donut chart, and 1 treemap (listed in decreasing familiarity for general audiences according to [54]).

We also aimed to select charts with a diverse set of topics that would vary in familiarity and interest among participants. Existing work has shown that background knowledge on and interest in the subject matter of a visualization can impact engagement [34], so we hoped a diverse set of topics would include something for every participant. The topics of the charts in the final set were: Alphabet's earnings, the number of COVID-19 cases and deaths in a US state, the severity of COVID-19 symptoms, the threat of extinction for different kinds of animals, the guns used in mass shootings in the US, and the number of times different diseases were mentioned on Twitter. Of this set, we hypothesized that the charts about COVID-19 would be the most familiar topic and the breakdown of Alphabet's earnings would be the least familiar.

While all 6 pairs take inspiration from real-world infographics, 5 of the pairs contained a modified version of real, published charts gathered from sources including the New York Times, Washington Post, and Visual Capitalist. The sixth pair was created by one of the authors in the style of existing area charts of COVID-19 cases, but using a smaller data set that more closely matched the range of variables present in the other stimuli. Each chart contained between 3 and 6 variables. The specific range of variables was not pre-selected, but emerged through our search for real-world charts with limited visual complexity.

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

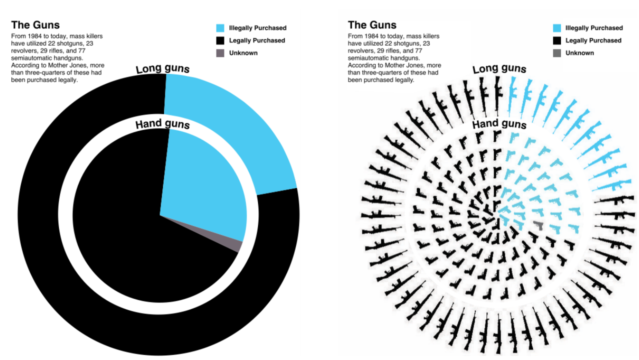

Fig. 1. One of the pairs of infographics used in this study. This chart is about the guns used in mass-shootings in the United States.

3.2 Experiment 1: Method

We conducted this experiment on Amazon's Mechanical Turk. We recruited 60 Workers ($Mean_{age} = 39.44$, $SD_{age} = 11.28$, 18 women, 30 men, 12 others) who had completed at least 100 tasks with an approval rate of at least 95%. We selected this combination of tasks completed and approval rate in line with existing research on quality data without attention-check questions [49]. We decided to survey 60 participants after calculating a power analysis based on the effect size of a pilot study. This experiment was conducted in 2 phases: Comprehension and Comparison. All participants completed both phases. In total, the experiment took about 30 minutes and participants were paid $5.00.

Phase 1: Comprehension. The first phase of the experiment asked participants to answer a series of 6 comprehension questions based on the levels of Bloom's taxonomy [7, 13]. This method, which was proposed in [13], allowed us to comprehensively assess different aspects of understanding which differ in complexity. The possible limitations of this approach (and the newness of the technique) are discussed in Section 5.

Because each chart has a different topic, we created a set of 6 different questions per chart — see Table 1 for a description of each level in the taxonomy and an example question for each level. The entire list of comprehension questions can be found in supplementary materials. This experiment has a 2 by 6 Graeco Latin Square design with repetition, wherein each participant sees one version of all 6 infographics (3 Count, 3 Area).

Phase 2: Comparison. In the second phase of the experiment, participants were shown both versions of each infographic, side by side, and asked to compare them on a series of metrics (see Table 2). Participants viewed the charts in the same order as in the comprehension portion. We used an additional Latin Squares design to randomize whether the Area condition appeared on the left or right. The metrics were selected after compiling a list of qualities remarked in existing literature as being important for effective infographics [37, 69]. From this list, we selected those which we thought could be affected by the presence or absence of pictographs and designed a question based on each. For each metric, participants indicated on a 5-point Likert scale which of the two versions of the chart best satisfied the prompt. The extremes of these scales indicated a strong preference for one chart over the other. Finally, participants scored their familiarity with the topic of each chart on a 5-point Likert scale.

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

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

3.3 Experiment 1: Hypotheses, Metrics and Analysis

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

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

3.4 Experiment 1, Phase 1: Results

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

3.4.1 Hypothesis 1.1: Higher-level insights

Insight. Question 2, 5, and 6 tested the ability of participants to extract key messages and gain insight. These questions had no objectively

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

Table 1. We used Bloom's taxonomy [7] to create comprehension questions following the method from [13]. The levels of the original taxonomy are shown here, along with a sample question indicative of those created for the experiment.

correct answers, so we designed a 7 point scale (0–6) judging the kind of insight present. On this scale, a score of 0 indicated no information gathered from the chart. Low insight responses (assigned scores of 1 or 2) referenced the information directly from the chart and demonstrated no further attempt at casual sensemaking. For example, in response to Q2 about the main takeaway of the chart about COVID-19 cases, one participant responded *"Everything is increasing every day."* Medium insight responses focused on a single dimension (score: 3) or offered opinions without coming to conclusions based on the data (score: 4). For example, to the same chart, another participant responded *"Death rate goes down with more testing."* Finally, high insight responses came to conclusions not included in the chart (for a score of 5) and offered justification for those responses (for a score of 6), such as here, where a participant speculates about what is to come: *"This shows that the number of cases increase rapidly and could double in a week."*

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

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

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

Overall, accuracy was high — 92.9% of the participants answered Question 1 correctly, 70.4% of the participants answered Q3 correctly, and 92% of the participants answered Q4 correctly. Participants had the highest accuracy when answering questions about the extinction chart (90.2%) and the lowest accuracy when answering about the chart about guns (80.6%). Post-hoc analysis with Tukey adjustments showed no significant difference in accuracy between topics. Using a mixed-effect linear model to predict accuracy with question type, chart version, and chart topic, we found no main effect of chart version ($\chi^2 = 0.0024$, $p =$ 0.96), but an overall main effect of question ($\chi^2 = 106.56$, $p < 0.001$) and chart topic ($\chi^2 = 12.62$, $p = 0.027$). Post-hoc analysis with Tukey adjustments [38] showed that participants performed significantly worse on Question 3 compared to Question 1 ($Est = 0.23$, $SE = 0.025$, $p <$ 0.001) and 4 ($Est = -0.22$, $SE = 0.025$, $p < 0.001$). Because Question 1 and 4 were tasks related to identifying information in the chart directly and Question 3 asked participants to compute a number based on information from the chart, we hypothesize that the dip in accuracy for Question 3 may be driven by higher graphical literacy than numeracy in our participant population.

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

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

3.4.2 Hypothesis 1.2: Higher Engagement

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

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

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

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

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

3.4.3 Hypothesis 1.3: Stronger Emotional Response

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

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

Surprise. Because past literature identified "surprise" as a component of the insight generation process (e.g., [44]), we also coded for this dimension (binary, contains/does not contain). Chart version did not have a significant effect on eliciting expressions of surprise (χ^2 = 0.092, $p = 0.76$). Very few viewers mentioned surprise in their response (1.69%), with about half of them in response to a Area chart and the other half in response to a Count chart. Responses containing surprise varied across topics, but often related to the size of particular dimensions. One example of this was in response to the chart about confirmed cases of COVID-19, *"They would be surprised that there were so few cases of death and hospitalizations as the number of confirmed cases grows."* They also commented on the overall size of the units depicted such as this response to the Alphabet chart *"People in my community would be very surprised at how much money that these companies make from Google and I think they would be very surprised at these numbers."*

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

3.4.4 Other Observations

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

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

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

We found a similar pattern of effects regarding whether participants described their community or mentioned taking action. For descriptions, there was a main effect of chart topic ($\chi^2 = 19.80$, $p = 0.0014$), but no main effect of chart version ($\chi^2 = 2.42$, $p = 0.12$). Similarly, chart version did not have a significant effect on mentions of taking action ($\chi^2 = 0.82$, $p = 0.37$), but chart topic did ($\chi^2 = 26.91$, $p < 0.001$). Participants were more likely to describe their community and mention taking actions in responses to the charts about extinction and guns. For example, one participant described their geographic location as a means to justify the relevance of the extinction chart *"It would effect it to a small extent although the threat of extinction to Amphibians could pose an issue as we are located on the ocean."* Another mentioned the action their community would take in response to the guns chart, writing *"Smart people would back legislation to make all guns more difficult to obtain. People would be more willing to accept stricter background checks and mandatory registration."*

3.5 Experiment 1, Phase 2: Results

Readability. Our results suggested that there was an overall difference in perceived readability ($\chi^2 = 61.71$, $p < 0.001$). Post-hoc analysis with Bonferroni adjustments [29] suggested that viewers were evenly divided on this issue, such that there were equal amounts of participants who rated the Area (45%) and Count chart (46%) as more readable ($p = 1.00$) (1 5). Details of pair-wise post-hoc comparisons can be found in the supplementary materials.

Visual Appeal. Ratings of visual appeal were similar to those of readability, such that there was an overall difference ($\chi^2 = 39.03$, $p <$

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

Table 3. Correlation table and VIF for metrics from Experiment 1, Phase 2. Note that larger values are darker, regardless of sign. Several metrics were highly correlated, such as readability and understandability, or perceived clutteredness and complexity. VIFs were relatively high. For example, both readability and understandability had VIFs greater than 4, rendering them less-optimal measures of viewer attitudes as they can be highly accounted for by other dimensions. Experiment 2 modified these metrics to reduce the multicollinearity between them.

0.001) driven by polarized ratings (1 5). Viewers were equally likely to rate Area (46%) and Count chart (47%) as more visually appealing ($p = 1.00$), with very few giving neutral ratings (6.5%).

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

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

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

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

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

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

Predicting Comprehension with Comparison Metrics. Because viewers spent a considerable amount of time deciphering the infographics in Phase 1, by the time the participants saw both versions of the infographic in Phase 2, they may have processed the chart seen before more 'fluently' – requiring less time to visually dissect and comprehend it [56]. This processing fluency may impact judgment, such as preference or trustworthiness ratings [55, 57]. Therefore, we investigated whether the chart seen in Phase 1 impacted ratings on the given metrics in Phase 2.

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

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

3.6 Experiment 1: Discussion

In summation, we observed that Count and Area charts did not differ in their ability to elicit higher levels of insight. Participants were not more likely to answer questions correctly across the two chart versions, nor did they show more signs of engagement with the content on metrics of mental effort or response length. We observed that participants were more likely to misinterpret Count charts than Area charts, but because of the small number of misinterpretations overall, this effect may be overstated and explainable by one commonly misinterpreted chart. Additionally, we observed several interesting trends among the results of Phase 2, including polarized views on visual appeal and no observed effect of chart version on relatability.

However, there was a correlation between some of the metrics assessed in Phase 2 and the chart version participants saw previously in Phase 1. We suspect that either perceptual fluency was a factor or that the questions were not clear to the participants. Further, analysis of multicollinearity using variance inflation factors (VIFs) (see Table 3) suggested that many of our metrics were not independent, as evidenced by several high VIFs of 2.5 and above. Stated differently, the set of metrics interacted with each other and, therefore, unsatisfactorily measured viewers' perceptions. When moving to explore the metrics from Phase 2 more deliberately, we made several changes. Namely, we removed redundancy within the set by removing questions about Readability and Complexity. Additionally, we rephrased several of the questions to increase clarity (see Table 4 for the updated questions).

4 Experiment 2: Personal Experience

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

4.1 Experiment 2: Stimuli and Methods

In this experiment, we utilized the same stimuli as the previous experiment but iterated upon the Comparison phase to see what other insights could be revealed. We collected data from 60 participants via Amazon's Mechanical Turk ($mean_{age} = 39.74$, $SD_{age} = 12.98$, 23 women, 27 men, 10 other). As in Phase 2 of Experiment 1, participants were shown two versions of the same chart and asked to indicate which of the two charts best satisfied the prompt (see Table 4 for prompts). After each rating, with the exception of the questions about Familiarity and Interest, participants were asked to explain their reasoning. Chart order was determined with the Graeco Latin Square order from the previous experiment, but, critically, participants had not seen either chart previously. Participants rated every metric on a separate page and all metrics were presented in the same order throughout.

4.2 Experiment 2: Analysis and Results

To analyze the responses, we utilized the Thematic Analysis method [11]. Two of the authors familiarized themselves with the responses and then independently generated a set of preliminary codes for separate questions based on common topics mentioned. Additional codes were generated and applied iteratively until each author could not come up with more codes, at which point the codes were grouped into broader themes. The 3 to 5 most frequently applied themes for each question are included in Table 4. We now present the numeric and qualitative results for each dimension judged by participants.

Visual Appeal. Participant opinion on visual appeal was polarized ($\chi^2 = 54.43$, $p < 0.001$). A Chi-squared test revealed that there was an equal number of participants that rated Count and Area as more visually appealing ($p_{area/count} = 0.82$). A post-hoc analysis with Bonferroni corrections [29] showed significantly fewer people gave a neutral rating (1 5). Detailed pair-wise comparison p-values can be found in the supplementary materials. We observed a positive correlation between perceived visual appeal and importance ($Est = 0.23$, $SE = 0.051$, $p < 0.001$), such that the chart that was perceived to be more visually appealing was also perceived to make its topic seem more important.

As shown in Table 4, the most common themes identified for this question were "easy to understand," "easy to see," "icon selection," and "simplicity." This is particularly interesting because although this question was asking about visual appeal, the most common themes we found are actually about ease of use. This suggests that for many participants, visual appeal and understanding cannot be separated – a chart which is easy to use is also visually appealing.

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

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

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

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

The most common themes that emerged were "difficulty understanding value," "more to look at," "difficulty focusing," and "difficulty identifying topic" (see Table 4). These themes may give us a window into what kind of information the participants were using as a proxy for the amount of time it would take for them to understand the chart or the kinds of things they imagine would be a hindrance to understanding.

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

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

Envisioning. Overall, when asked about which chart best made it easier to envision what was happening and to relate the data to real world objects, situations, or entities, viewers perceived Count and Area charts differently ($\chi^2 = 29.03$, $p < 0.001$). In particular, significantly more participants rated the Count chart as easier to envision ($p < 0.01$) (1 5). There was also a positive correlation between perceived easiness to envision and perceived importance ($Est = 0.39$, $SE = 0.052$, $p < 0.001$), such that the chart that was perceived to be easier to envision also made its topic seem more important.

The top themes identified in responses to this question were "makes the topic real," "straightforward," and "obvious chart topic" (see Table 4). Though it's unsurprising that participants talked about things that made a topic seem real, the other two popular themes suggest a more surprising potential relationship between relating and ease or speed of understanding. Proponents of Count charts expressed that the pictographs made the topic seem more "real" and encouraged them to relate it to their own experience. As one participant wrote, *"The human icons emphasizes [sic] that COVID-19 is affecting people just like me all over the world."* In addition, participants contended that the pictographs made the topic and message of the chart obvious, especially at a glance. For example, about the chart about social media mentions, one participant wrote *"When I looked at [the count chart] I knew without even looking at the top that it was about Twitter as I saw the Twitter logo. At first glance I don't know what [the area chart] is referring to."* While the bird pictographs were effective for signaling the chart topic to some participants, others found it confusing. As one participant described it *"The birds are somewhat misleading as I think of actual birds instead of Twitter."* Another participant said that they found the Area chart easier to envision *"Because I don't know what birds have to do with the coronavirus."*

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

Unnecessary Clutter. Count charts were significantly more often rated as unnecessarily cluttered, compared to Area charts ($\chi^2 = 227.4$,

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

Table 4. In Exp. 2, participants compared 2 versions of a chart along 7 metrics and then explained their reasoning. These metrics are a subset of those used in Exp. 1 Phase 2, but most are worded differently. The top 3-5 themes for each question are listed with the number of responses they were assigned to. There was no justification requested for Familiarity or Interest, because these questions pertained to the chart topic only and did not compare chart versions.

$p < 0.001$). However, the number of people who gave neutral ratings is significantly higher than all the other categories (1 5). Detailed pair-wise comparisons can be found in the supplementary materials.

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

As Table 4 shows, the main themes for this question were "no added meaning," "organization," and "space." Within these themes, we see both a possible response to the word "needlessly" that was used in the prompt, as well as an idea of what characteristics participants were using to decide what was or was not cluttered.

For some participants, the Count charts felt needlessly cluttered because the pictographs contributed no additional meaning. Exemplifying this idea was this response from one participant *"It doesn't seem like there's much point in using the moneybags when [the area chart] works just as well. I'd rather just have [the area chart] if I were reading this data on a website or document."* Additionally, participants were critical of the number of pictographs and their organization, describing the page as "busy" and "complicated." As one participant put it, *"I can get what the person who made this chart was trying to express that all different types of people get COVID-19 and it effects them in different ways but the presentation is jumbled and distracting."* Some participants who answered neutrally didn't consider either of the the charts cluttered. One reason for this was the layout, such as expressed here: *"Both charts use the space provided well enough and do not feel overwhelming."* Others did find the Count charts cluttered, but not "needlessly" so. For example, one participant wrote *"I wouldn't say either is NEEDLESSLY cluttered. [The Count chart] seems a little more cluttered since it's representing each person but I don't think it's needless. It definitely serves a purpose."*

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

The most common themes present in responses to this question were "real world connection," "color," and "straightforward" (see Table 4). These themes suggest both an impact of realism and ease of understanding on perceptions of importance and urgency.

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

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

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

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

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

Familiarity. Familiarity is a different metric than the others because it measures viewer attitude towards the topic rather than the chart. We conducted a one-way ANOVA comparing familiarity for each topic which suggested differing familiarity across the topics represented ($F = 19.63$, $p < 0.001$). Post-hoc analysis with Tukey adjustment [38] suggests that participants were significantly more familiar with COVID-19 related topics than other topics ($MD = 1.28$, $p < 0.001$).

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

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

Table 5. Correlation table and VIF for metrics from Experiment 2. Note that larger values are darker, regardless of sign. No two metrics appear to be strongly correlated. Overall VIFs decreased compared to those in Experiment 1, suggesting that the metrics used in Experiment 2 more orthogonally capture different participant attitudes and thus were better metrics to use in this type of work.

as more interesting ($Est = 0.69$, $SE = 0.041$, $p < 0.001$).

Interest. Like familiarity, interest is another metric that measures viewer attitude towards the topic rather than the chart itself. One-way ANOVA comparing interest for each topic suggested that viewers were more interested in some topics than others ($F = 14.9$, $p < 0.001$). Post-hoc analysis with Tukey adjustment suggested that participants were significantly more interested in COVID-19 related topics than other topics ($MD = 1.03$, $p < 0.001$). Controlling for other metrics, interest in the topic did **not** significantly predict perceived topic importance/urgency ($Est = 0.021$, $SE = 0.062$, $p = 0.74$), easiness to understand ($Est = 0.096$, $SE = 0.66$, $p = 0.14$), or perceived unnecesary clutter ($Est = -0.026$, $SE = 0.046$, $p = 0.58$). However, it **did** significantly predict familiarity, such that more familiar topics were rated as more interesting ($Est = 0.71$, $SE = 0.042$, $p < 0.001$). It also significantly predicted visual appeal – charts depicting more interesting topics were also rated as more visually appealing ($Est = 0.20$, $SE = 0.068$, $p = 0.004$). Finally, it trendingly predicted easiness to envision such that charts depicting more interesting topics were rated as harder to envision ($Est = -0.11$, $SE = 0.063$, $p = 0.07$).

Correlation and VIF of Metrics Used. Analysis of multicollinearity using variance inflation factors (VIF) (see Table 5) suggested that although our rating metrics were not fully independent, there was only a small amount of variance inflation (most VIFs were around 2 or below). In other words, these metrics were satisfactory measures of viewer perceptions of visualization designs. In addition, both VIFs and the number of highly-correlated metrics decreased from Experiment 1, which demonstrates that the changes made to the metrics were effective.

5 Discussion and Future Work

This study revealed several important findings. First, we found no difference in the aspects of understanding demonstrated by viewers of infographics containing pictograph arrays and those containing solid areas. However, using pictograph arrays significantly influenced the participants' experience. In particular, we found a visual complexity-relatability trade-off. Infographics with pictograph arrays were thought to require more time to understand and were perceived as more often visually cluttered, but they enabled viewers to more easily relate the chart topic to the real world. Additionally, two distinct perspectives on the use of pictographs emerged in our experiment. For some, pictographs are preferable to abstract areas because they are visually appealing and make the topic more relatable. For others, charts with pictographs are not preferred because they are seen as cluttered, complex, and less serious. We will now speculate as to why these results occurred and then discuss their potential impacts on the design of visualizations.

Bloom's taxonomy as an evaluation method

In this paper, we used Bloom's taxonomy as a formal evaluation method for visualizations. Through this method, we were able to comprehensively probe 6 different aspects of understanding. Although we found no difference between the two chart versions we evaluated, the breadth of aspects covered and the real-world applicability of these aspects suggests that our inquiry was still valuable. Since conducting these experiments, the authors have written a more in-depth description of the method of using Bloom's taxonomy for evaluation [13].

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

We can imagine other translations of these levels to the causal sensemaking context and we *did* translate them differently in other related work [13]. In that work, participants predicted a value beyond the chart (Level 5) and generated and justified a conclusion (Level 6). This is in contrast to this paper, where we asked about the participant's community (Level 5) and what they thought a public official should do (Level 6). Future work that uses this method may find yet new ways of translating these upper levels across contexts.

Why was there no effect on understanding?

In our experiments, replacing the geometric shapes of the plain charts with pictograph arrays changed the way that the information was encoded. It would be reasonable to think that a change like this would have some effect on the understanding obtained by viewers, but we observed no such effect. There are several reasons why this might be.

First, it is possible that although one version uses solid shapes and the other individual icons, participants may process the images similarly – as masses of color [27]. Future work with more complex charts or with fewer pictographs may produce different results. Second, it is possible that while there was no effect for participants in aggregate, there were subsets of people who did experience some effect. For example, particular combinations of graph literacy and numeracy may have led to larger effects such as those observed in the medical risk literature (e.g., [24]). Similarly, some of the factors which have been observed to affect engagement such as subject matter, source, or self-efficacy [34] may also affect understanding. Future work could examine the possible effect of these factors and could help the community better understand what about an audience matters when trying to communicate facts and ideas. Third, it could be a feature of the infographics that we used in our experiments. The stimuli we chose to use were relatively simplistic as they only contained a few categories and under a few hundred points. It is possible that more complex charts might lead to larger effects. Future work with more complex infographics could help determine the veracity of this possibility.

It is also possible that this manipulation affected casual sensemaking, but not understanding, in ways that could not be detected by our measures, e.g., emotional aspects of sensemaking or sensemaking as an ongoing personal experience (e.g., as in the notion of sensemaking in [20]). Past work has shown that emotions have a strong impact on the way that data and visualizations are perceived and understood – it is not just the data themselves that are important, but how they feel [33]. Given the range of factors identified in Experiment 2, it may be true that though the comprehension aspects of casual sensemaking (measured in Experiment 1) were not effected, the **meaning** made was.

Why did participants find pictographs easier to relate to?

Our results for Experiment 2 indicate that participants thought that charts with pictographs helped them better envision the chart topic. One possible explanation is that including pictographs related to the topic reduces the mental burden of relating abstract textual content to a visual depiction. This could be related to the factors thought to make concrete scales effective – by relating abstract depictions of data to something more familiar, concrete scales reduce the cognitive load of comprehending the underlying numerical values [18]. This explanation could still be consistent with our results from Experiment 1 that observed no difference in response accuracy between the pictograph and area conditions, as reducing cognitive load does not necessarily increase performance [31]. Instead, it could indicate that even the most complex questions did not overtax the participant's cognitive resources. Nonetheless, to investigate this conjecture, future work could incorporate subjective or psychological measures of mental effort and cognitive load (such as those reviewed in [47]).

Alternately, strong positive emotions could influence participant perceptions. Previous work on emotional design has shown that tools which are perceived as beautiful or attractive are also considered to work better [43]. Following this model, when a chart containing pictographs evoked positive emotions for participants, it may have helped them feel more willing to engage and ultimately produced a feeling of ease and reduced effort. It is worth noting, however, that a viewer may experience positive emotions in response to charts as a whole, even when the topic or pictographs are not happy or joyful. For example, in our study, sometimes the pictographs were described by participants as "scary," even in a chart that was otherwise noted to be aesthetically beautiful (e.g., in response to the guns chart in Figure 1).

It is also possible that the "realism" of a pictograph contributes to how effective they are at making the topic seem real. There is conflicting evidence supporting the effectiveness of anthropomorphized images of people on inducing empathy (such as in [10, 26]), but little is known about their effect in other realms such as those explored in this paper. Future research may investigate the use of more realistic images in infographics.

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

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

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

Design considerations for the use of pictographs

Our results suggest three design considerations. First, if a designer is looking to make their topic easy to envision, they should consider using pictograph arrays in place of geometric areas. Second, if a design contains pictographs and the designer is concerned that viewers could think the chart will take too long to understand, they may wish to consider the number of items present in the design and how easy it is to understand the value of each component. Third, to mitigate potential perceptions of unnecessary clutter in a design containing pictographs, the designer may wish to consider what additional meaning the pictographs contribute over a more traditional representation.

Further, our results suggest that when producing graphics which are not very data dense, pictographs are best when used in two ways: to help identify what the chart is about or when organized spatially into clusters that should be interpreted collectively. Critically, this means that pictographs should not be used to force the viewer to count [63]. Instead, designers should combine pictographs with textual labels containing the absolute values where they are important to avoid making viewers feel like they need to count in order to understand.

Though our results indicate that pictographs can be effective for helping the viewer envision the topic, finding representative pictographs for abstract concepts is not a trivial task. For example, while there was little confusion about the symbols we used to represent people, some participants found the Twitter bird to be helpful, while others found it confusing. Existing studies have shown that even when icons are specifically designed to cross language and cultural divides, they are often not understood as intended [70]. However, having a set of pictographs which is as inclusive as possible in terms of subject and multi-cultural clues can help. One set which we utilized in our redesigned chart about COVID-19 symptoms (see top row, second column of Figure) was WeePeople [53] which contains silhouettes of individuals of different genders and races, with and without mobility aids. While there are some excellent inclusive pictograph sets, such as the WeePeople set, there is still work to be done so that designers can choose resonant pictographs based on the background of their anticipated audiences.

6 Conclusion

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

Acknowledgments

The authors wish to thank the anonymous reviewers for their thoughtful feedback, Evan Anderson of the Visual Thinking Lab for assisting with the qualitative coding, Mahmood Jasim of the HCI-VIS lab for his feedback on early versions of the paper, and Mackenzie Matscherz for her help with designing stimuli.

References

[1] E. Adar and E. Lee. Communicative visualizations as a learning problem. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):946–956, 2021.

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

[3] A. Arcia, N. Suero-Tejeda, M. E. Bales, J. A. Merrill, S. Yoon, J. Woollen, and S. Bakken. Sometimes more is more: iterative participatory design of infographics for engagement of community members with varying levels of health literacy. *Journal of the American Medical Informatics Association*, 23(1):174–183, 2016.

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

[5] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proc. of the SIGCHI Conf. on*

Human Factors in Comput. Systems, CHI '10, p. 2573–2582. Association for Computing Machinery, New York, NY, USA, 2010.

[6] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmann, B. Dai, G. Grothendieck, P. Green, and M. B. Bolker. Package 'lme4'. *Convergence*, 12(1):2, 2015.

[7] B. S. Bloom. *Taxonomy of educational objectives: The classification of educational goals*. Longman, 1956.

[8] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics*, 22(1):519–528, 2015.

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

[10] J. Boy, A. V. Pandey, J. Emerson, M. Satterthwaite, O. Nov, and E. Bertini. Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data? In *Proc. of the 2017 CHI Conf. on Human Factors in Computing Systems*, CHI '17, p. 5462–5474. Association for Computing Machinery, 2017.

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

[12] N. T. Brewer, A. R. Richman, J. T. DeFrank, V. F. Reyna, and L. A. Carey. Improving communication of breast cancer recurrence risk. *Breast cancer research and treatment*, 133(2):553–561, 2012.

[13] A. Burns, C. Xiong, S. Franconeri, A. Cairo, and N. Mahyar. How to evaluate data visualizations across different levels of understanding. In *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*, pp. 19–28, 2020.

[14] V. Byrd. Using bloom's taxonomy to support data visualization capacity skills. In S. Carliner, ed., *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2019*, pp. 1039–1053. Association for the Advancement of Computing in Education (AACE), New Orleans, Louisiana, United States, November 2019.

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

[16] A. Cairo. *The truthful art: data, charts, and maps for communication*. New Riders, 2016.

[17] Z. Chen, Y. Wang, Q. Wang, Y. Wang, and H. Qu. Towards automated infographic design: Deep learning-based auto-extraction of extensible timeline. *IEEE Trans. Vis. Comput. Graphics*, 26(1):917–926, 2020.

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

[19] Y. De Haan, S. Kruikemeier, S. Lecheler, G. Smit, and R. van der Nat. When does an infographic say more than a thousand words? audience evaluations of news visualizations. *Journalism Studies*, 19(9):1293–1312, 2018.

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

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

[22] J. Fuchs, P. Isenberg, A. Bezerianos, M. Miller, and D. Keim. Educlust - a visualization application for teaching clustering algorithms. In *Eurographics 2019 - 40th Annual Conf. of the European Association for Computer Graphics*, pp. 1–8. Genova, Italy, 2019. Education paper.

[23] M. Galesic, R. Garcia-Retamero, and G. Gigerenzer. Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychology*, 28(2):210, 2009.

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

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

[26] A. Genevsky, D. Västfjäll, P. Slovic, and B. Knutson. Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. *Journal of Neuroscience*, 33(43):17188–17196, 2013.

[27] S. Haroz, R. Kosara, and S. L. Franconeri. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proc. of the 33rd Annual ACM Conf. on Human Factors in Comput. Systems*, CHI '15, p. 1191–1200. Association for Computing Machinery, 2015.

[28] L. Harrison, K. Reinecke, and R. Chang. Infographic aesthetics: Designing for the first impression. In *Proc. of the 33rd Annual ACM Conf. on Human Factors in Computing Systems*, CHI '15, p. 1187–1190. Association for Computing Machinery, 2015.

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

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

[31] W. Huang, P. Eades, and S.-H. Hong. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3):139–152, 2009.

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

[33] H. Kennedy and R. L. Hill. The feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology*, 52(4):830–848, 2018.

[34] H. Kennedy, R. L. Hill, W. Allen, and A. Kirk. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday*, 21(11), 2016.

[35] N. W. Kim, H. Im, N. Henry Riche, A. Wang, K. Gajos, and H. Pfister. Dataselfie: Empowering people to design personalized visuals to represent their data. In *Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems*, CHI '19, p. 1–12. Association for Computing Machinery, 2019.

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

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

[38] R. Lenth, H. Singmann, J. Love, et al. *Emmeans: Estimated marginal means, aka least-squares means*, 2018.

[39] Z. Liu, J. Thompson, A. Wilson, M. Dontcheva, J. Delorey, S. Grigg, B. Kerr, and J. Stasko. Data illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. In *Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems*, CHI '18, p. 1–13. Association for Computing Machinery, 2018.

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

[41] N. Mahyar, S.-H. Kim, and B. C. Kwon. Towards a taxonomy for evaluating user engagement in information visualization. In *Workshop on Personal Visualization: Exploring Everyday Life*, vol. 3, p. 2, 2015.

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

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

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

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

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

[47] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. V. Gerven. Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1):63–71, 2003.

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

[49] E. Peer, J. Vosgerau, and A. Acquisti. Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior research methods*, 46(4):1023–1031, 2014.

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

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

[52] Z. Pousman, J. Stasko, and M. Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE transactions on visualization and computer graphics*, 13(6):1145–1152, 2007.

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

[54] A. Quispel, A. Maes, and J. Schilperoord. Graph and chart aesthetics for experts and laymen in design: The role of familiarity and perceived ease of use. *Information Visualization*, 15(3):238–252, 2016.

[55] R. Reber and N. Schwarz. Effects of perceptual fluency on judgments of truth. *Consciousness and cognition*, 8(3):338–342, 1999.

[56] R. Reber, N. Schwarz, and P. Winkielman. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review*, 8(4):364–382, 2004.

[57] R. Reber, P. Winkielman, and N. Schwarz. Effects of perceptual fluency on affective judgments. *Psychological science*, 9(1):45–48, 1998.

[58] D. Ren, B. Lee, and M. Brehmer. Charticulator: Interactive construction of bespoke chart layouts. *IEEE Trans. Vis. Comput. Graphics*, 25(1):789–799, 2018.

[59] M. Sedlmair, P. Isenberg, T. Isenberg, N. Mahyar, and H. Lam. Preface. In *BELIV '16: Proc. of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. Association for Computing Machinery, New York, NY, USA, 2016.

[60] D. Skau and R. Kosara. Readability and precision in pictorial bar charts. In *Proc. of the Eurographics/IEEE VGTC Conf. on Visualization: Short Papers*, EuroVis '17, p. 91–95. Eurographics Association, 2017.

[61] M. Smiciklas. *The power of infographics: Using pictures to communicate and connect with your audiences*. Que Publishing, 2012.

[62] W. H. Starbuck and F. J. Milliken. Executives' perceptual filters: What they notice and how they make sense. In D. Hambrick, ed., *The executive effect: concepts and methods for studying top managers*, pp. 35–65. JAI Press, Greenwich, CT, 1988.

[63] P. Trogu. Counting but losing count: the legacy of Otto Neurath's ISOTYPE charts. *Visible Language*, 52(2), 2018.

[64] E. R. Tufte. *The visual display of quantitative information*, vol. 2. Graphics press, 2001.

[65] M. Twyman. The significance of isotype. *Graphic communication through ISOTYPE*, 7, 1975.

[66] Y. Wang, H. Zhang, H. Huang, X. Chen, Q. Yin, Z. Hou, D. Zhang, Q. Luo, and H. Qu. Infonice: Easy creation of information graphics. In *Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems*, CHI '18, p. 1–12. Association for Computing Machinery, 2018.

[67] Z. Wang, S. Wang, M. Farinella, D. Murray-Rust, N. Henry Riche, and B. Bach. Comparing effectiveness and engagement of data comics and infographics. In *Proc. of the 2019 CHI Conf. on Human Factors in Comput. Systems*, CHI '19. Association for Computing Machinery, 2019.

[68] H. Xia, N. Henry Riche, F. Chevalier, B. De Araujo, and D. Wigdor. Dataink: Direct and creative data-oriented drawing. In *Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems*, CHI '18, p. 1–13. Association for Computing Machinery, 2018.

[69] M. Yin, D. Hollender, L. Condelli, D. Shewitz, A. Duffy, and M. Movit. *The Power of Data Visualization: Advanced Presentations of NRS Data*, January 2014.

[70] M. Zender and G. M. Mejía. Improving icon design: through focus on the role of individual symbols in the construction of meaning. *Visible Language*, 47(1):66–89, 2013.

[71] J. E. Zhang, N. Sultanum, A. Bezerianos, and F. Chevalier. Dataquilt: Extracting visual elements from images to craft pictorial visualizations. In *Proc. of the 2020 CHI Conf. on Human Factors in Computing Systems*, CHI '20, p. 1–13. Association for Computing Machinery, 2020.

[72] B. J. Zikmund-Fisher, H. O. Witteman, M. Dickson, A. Fuhrel-Forbis, V. C. Kahn, N. L. Exe, M. Valerio, L. G. Holtzman, L. D. Scherer, and A. Fagerlin. Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical decision making*, 34(4):443–453, 2014.