Research Article

How People Are Influenced by Deceptive Tactics in Everyday Charts and Graphs

—CLAIRE LAUER AND SHAUN O'BRIEN

***Abstract*—Background:** *Visualizations are used to communicate data about important political, social, environmental, and health topics to a wide range of audiences; however, perceptions of graphs as objective conduits of factual data make them an easy means for spreading misinformation.* **Research questions:** *1. Are people deceived by common deceptive tactics or exaggerated titles used in data visualizations about non-controversial topics? 2. Does a person's previous data visualization coursework mitigate the extent to which they are deceived by deceptive tactics used in data visualizations? 3. What parts of data visualizations (title, shape, data labels) do people use to answer questions about the information being presented in data visualizations?* **Literature review:** *Although scholarship from psychology, human-computer interaction, and computer science has examined how data visualizations are processed by readers, scholars have not adequately researched how susceptible people are to a range of deceptive tactics used in data visualizations, especially when paired with textual content.* **Methodology:** *Participants (n = 329) were randomly assigned to view one of four treatments for four different graph types (bar, line, pie, and bubble) and then asked to answer a question about each graph. Participants were asked to rank the ease with which they read each graph and comment on what they used to respond to the question about each graph.* **Results/Discussion:** *Results show that deceptive tactics caused participants to misinterpret information in the deceptive versus control visualizations across all graph types. Neither graph titles nor previous coursework impacted responses for any of the graphs. Qualitative responses illuminate people's perceptions of graph readability and what information they use to read different types of graphs.* **Conclusions:** *Recommendations are made to improve data visualization instruction, including critically examining software defaults and the ease with which people give agency over to software when preparing data visualizations. Avenues of future research are discussed.*

Index Terms*—Charts, data visualization, deceptive visuals, graphs.*

The era of social media character limits, algorithms that limit post exposure, and the now-popular tl;dr (too long; didn't read) response to content perceived as too long-winded has ushered in new abbreviated modes through which we compose and ingest information. Memes, tweets, news summaries, and clickbait headlines have made us accustomed to expect content in shorter, more compact form, especially online. The influence of these genres has seeped over into technical and professional communication practice so that the packaging of more technical information into succinct visual forms like graphs and other data visualizations is now expected for even complex information. As Edward Tufte famously showed with his analysis of Minard's Napoleonic war graph, no other genre packs as much instant content in a single frame as an effectively designed data visualization [1].

Data visualizations typically use principles of reduction (in which points, lines, and shapes are used to represent relationships between concepts) and space (in which size, position, and scale reveal patterns and capture differences in data) [2]. When data visualizations required hand-drawing and processing, they were almost exclusively developed by technical professionals communicating data in technical contexts (e.g., [3]). However, spreadsheet software like Microsoft Excel and Google Sheets, and more advanced open-source scripting options like R, now allow for the automated processing of data into visual form, enabling even the most novice communicators the ability to utilize a broad array of color and visual effects to construct compelling data-driven arguments. Combine this fact with the vast supply of data that companies collect about their users or that is freely available

Manuscript received May 13, 2020; revised September 1, 2020; accepted September 2, 2020. Date of current version November 24, 2020. *(Corresponding author: Claire Lauer.)*
Claire Lauer is with the Technical Communication Program, Arizona State University, Mesa, AZ 85212 USA (email: claire.lauer@asu.edu).
Shaun O'Brien is with the Human Systems Engineering Program, Arizona State University, Mesa, AZ 85212 USA (email: shaunobrien@asu.edu).
This article has supplementary downloadable material at http://ieeexplore.ieee.org provided by the translators. The files consist of a Chinese translation of the abstract by Z. Xi (94 kB in size) and a Spanish translation of the abstract by E. Cogbill-Seiders (78 kB in size).

IEEE 10.1109/TPC.2020.3032053

Practitioner Takeaway

- Users are deceived by common deceptive data visualization design tactics such as truncated y-axes, 3D beveling, and arbitrary sizing of graph shapes, so designers should take care to avoid such tactics when translating analytics and other technical data into graphs and data visualizations for a wide range of audiences.
- Users report having greater difficulty understanding graph types that they are unfamiliar with, so designers should be sure to properly annotate and label all graphs clearly and not make any assumptions about readability
- Designers should be aware of the potential for misleading defaults in software such as Microsoft Excel and always enhance such defaults in graphics editing programs such as Illustrator to ensure consistent axis ranges and properly labelled graph elements

through receptacles like Data.gov and GapMinder, and anyone with an internet connection, some basic software, and a social media account can develop and share data visualizations on a vast array of topics to a broad range of audiences.

Although graphs and other data visualizations are often perceived as objective conduits of data, scholars who build data visualizations recognize them as highly constructed rhetorical documents that can be framed or designed to suggest certain trends and findings over others. For instance, people tend to be attracted to content that shows clear causations, so political campaigns, news outlets, and amateur developers may choose to deliberately begin and end the time frame shown in a graph to suggest that a single event caused a trend in the data (e.g., the trajectory of the stock market after an election). It also happens that developers substitute photographs or illustrations where numerically accurate graph segments should be, which may visually misrepresent differences in data. And then there is the common practice of truncating the y-axis (or not using a baseline at all) when comparing data points. Scholars and practitioners continually debate whether and when truncating the y-axis is justifiable (e.g., [4]–[7]). Truncating the y-axis may be necessary when working with data that would never start at zero to begin with or when there is a need to more clearly differentiate between small differences in data. Still, it is generally agreed that by default, y-axes should start at zero and that graphs should always include a baseline.

Despite this agreement, during the 2016 US presidential election, the *Washington Post*'s John Muyskens showed how the Trump campaign opted not to show baselines in the barcharts that they often published, depicting leads that they claimed Trump had over his opponent, Hillary Clinton. Muyskens showed how, by fading out baselines or not including them at all in the campaign's barcharts, the Trump campaign was able to arbitrarily position the bars to visually exaggerate the differences between the percentages they wanted to compare. Because barcharts rely on line length and a consistent starting point to illustrate differences between data, without a baseline, it is impossible to compare the lengths of two bars in any meaningful way [8].

Other types of deceptive tactics include cherry picking data points, spacing time segments inconsistently along an x-axis, using 3-D effects to exaggerate the appearance of graph data, suggesting something in the accompanying text that the data in the graph do not support, comparing two different types of data in the same graph, making overly simplistic calculations or comparisons, and others. These deceptions occur for a range of reasons, including that graph developers may lack training in statistics, may value aesthetics over accuracy, may desire to show clear and simple trends, may make mistakes in their calculations or plotting, or may simply desire to mislead their audience.

Scholars in technical and professional communication (TPC) have not yet conducted many of the needed empirical investigations into how people read and understand data visualizations, especially when those visualizations are part of larger information texts [9]. Although scholars can easily identify and recommend best practices in graph design, few have actually studied whether and how people are deceived by deceptive practices when they occur, and how accompanying text plays a role in people's interpretations of data they see in graphs.

Research Questions

This article reports the results of a study that investigated the extent to which people were deceived by common deceptive tactics in data visualizations and the headlines that accompany such visualizations. Specifically, our study asked the following questions.

RQ1. Are people deceived by common deceptive tactics or exaggerated titles used in data visualizations about non-controversial topics.

RQ2. Does a person's previous data visualization coursework mitigate the extent to which they are deceived by deceptive tactics used in data visualizations.

RQ3. What parts of data visualizations (title, shape, data labels) do people use to answer questions about the information being presented in data visualizations.

Our study found that deceptive tactics caused participants to misinterpret information in the deceptive (versus control) visualizations for bar, line, pie, and bubble graphs regardless of previous visualization coursework. However, participants' misinterpretation was not significantly amplified when deceptive graphs were paired with exaggerated titles, and exaggerated titles by themselves did not produce statistically significant differences in responses. Awareness of these findings is important, and technical communicators should continue to study them because their jobs regularly task them with translating analytics and other technical data into graphs and data visualizations for a wide range of audiences. In addition to the inadvertent deceptions that can result from program defaults or plotting mistakes, the temptation to mislead can be strong when the stakes are high for persuading people to adopt certain stances about consequential issues, policies, and practices. In such circumstances, the desire to communicate perspectives as simple, clear, and decisive is strong. Knowing how susceptible people are to being misled through deceptive tactics and suggestive language can help us develop ways to detect and mitigate such deception in technical communication practice and better educate ourselves and the public moving forward.

Literature Review

Scholars have long discussed best practices for creating functional yet ethical graphs [4], [7], [10]–[12]. Compared to TPC, scholarship from psychology, human–computer interaction, and computer science have been more advanced in examining how data visualizations are processed by readers [13]. Researchers from these fields have examined how memorable certain visualization types can be, suggesting that familiar data visualizations are less memorable [14], [15]. Researchers have also discussed the effect of human bias on visual analytics [16] as well as the use and importance of color in data visualizations [17]–[20]. Furthermore, research in such areas as graphical perception [21]; persuasion [22]; user bias [23], [24]; information order [25]; attraction effect [26]; priming [27], [28]; and framing [29], [30] have shown that cognitive processing plays an important role in how readers interpret and understand data visualizations.

Deceptive Tactics Scholars have also theorized about the deceptive potential of data visualizations [1], [28], [31]–[33] and suggested ideas on how to avoid such deception [34]. Although these studies demonstrate a broad interest in ethical data visualization practice, scientists have only recently started to empirically test the extent to which people are actually deceived by deceptive tactics used in data visualizations [31]–[33], [35], [36]. Pandey et al. were the first to empirically show that participants are likely to be misled in their interpretations of data visualizations that employed deceptive tactics such as message reversal (e.g., an inverted axis) and message exaggeration (e.g., a truncated y-axis) [33]. Following them, O'Brien and Lauer found that people's perception of information presented in deceptive data visualizations persisted even when the deceptive graphic was paired with a paragraph of accurate text [32]. Researchers have begun to experiment with strategies to counteract the deceptive potential of data visualizations, such as giving users the option of using tools to zoom in on areas of the data visualization [37], though these options have not been entirely successful and more research is needed to fully understand the extent to which deceptive data visualizations can and do deceive people. Our study extends this line of research by investigating a broader range of visualization types and deceptive tactics than what has been studied previously.

Titles In addition to examining new kinds of deceptive data visualizations, our study also looks at how both control and exaggerated titles interact with visualizations to potentially influence the ways in which people perceive information. Borkin et al. [14] showed that people more accurately recall visualizations that conveyed a primary message in their title, rather than something more generic. Borkin et al. [15] utilized eye-tracking to analyze various chart types to understand what parts of the chart participants fixated on. They demonstrated the extent of user fixation on titles that accompanied visualizations. They also showed that titles were more likely to be fixated on when located above the visual rather than below. They concluded that the titles and text that accompany a data visualization aided people's recall of the main message of the visualization.

Kong et al. [38] studied the effect of data visualizations paired with misaligned or contradictory titles on participants' trust and recall of information. They found that participants recalled the message of the title over the data visualization when the title was presented in a way that misaligned with the data visualization. Another Kong et al. study [30] researched the effect of slanted titles on the perceived main message of data visualizations about controversial topics and found that the slant of the title influenced a participant's perception of the main message; however, slanted titles did not change participants beliefs about the overall topic presented. Kong et al. [30] ended this article by calling for researchers to study the effect of titles on noncontroversial topics.

Our study expands on these studies in several essential ways. Where Pandey et al. quantitatively studied only three kinds of data visualizations (bar, line, and bubble) without titles, and identified only three deceptive tactics, our study also studied pie graphs, line graphs with truncated axes, and the deceptive tactic of spatial exaggeration commonly shown in 3-D beveling. Also, whereas Pandey et al. did not research participant motivation, our study asked participants to reflect in their own words what they used to determine their responses to each study question. These comments provided essential qualitative data that we were able to use to more fully explore what people use to draw conclusions about data they encounter in a visualization. Furthermore, in our deliberate use of noncontroversial data visualizations, our study addresses the call by Kong et al. [30] to research the effects of titles paired with noncontroversial data visualizations. And unlike Kong et al. [38], who studied nondeceptive graphs, our study used deceptive visualizations paired with both control and exaggerated titles. Finally, the results of our study will complicate claims made by Borkin et al. [15] about the use of text to aid in the recall of data shown in visualizations.

METHODS

To test the extent to which people are influenced by deceptive tactics and exaggerated titles in data visualizations, we designed a study to measure people's perceptions of the differences between data points in a data visualization when shown four different graph types across a variety of potential treatments. Those treatments included the following.

- Control (nondeceptive) graph paired with control (non-exaggerated) title
- Control graph paired with exaggerated title
- Deceptive graph paired with control title
- Deceptive graph paired with exaggerated title

Participants were shown a total of four graphs, including one bar graph, one line graph, one pie graph, and one bubble graph. The order of the graphs and the type of treatment assigned to each graph were randomly assigned. Aside from the differences in the y-axis starting point, shape, or scale of the various treatments, the numerical data shown in each graph was identical.

We administered all study components using the online survey platform Qualtrics. Our study was approved by our University's Institutional Review Board. No participants under the age of 18 years old were recruited or allowed to participate in the study, and no payment was provided to participants. We recruited study participants through our first-year psychology student research pool, our first-year writing classes, and instructors whom we knew at other universities or who subscribe to the WPA-L who might be interested in distributing our study to their colleagues and students. Psychology students who participated as part of the research pool received course credit for participating. Other participants were given the opportunity to provide their email address at the conclusion of the survey to be entered in a drawing for a $50 Amazon gift card. Although our study participants were largely convenience sampled from university students, because our university includes students from all 50 US states and 136 countries, it allowed us to sample from a large, diverse group of individuals with varying cultures, perspectives, and experiences.

Survey Design We included a basic demographic questionnaire as part of the study and asked participants about their age, level of education, previous coursework in data visualization, and comfort level reading bar, line, bubble, and pie graphs. Questions about previous coursework and comfort level were included to help us determine whether those factors might play a role in a participant's study responses.

Graph Development To control for variables within and across each graph type, all graphs used the same blue and black color schemes, font styles, and chart elements, such as title and axis labels.

Fig. 1. Control and deceptive graph treatments and corresponding questions for bar, line, pie, and bubble graphs.

To avoid the influence of confirmation bias—or the likelihood that pre-existing opinions about graph topics would factor into the responses participants provided after reading the graphs—we built visualizations using fictional data that we considered apolitical and noncontroversial. For instance, the line graph showed six years of movie ticket sales, the bar graph showed two years of home sales, the pie graph showed concession sales at baseball games, and the bubble chart showed enrollment totals between a fictional university and the national average (see Fig. 1).

The control treatments for both the bar and line graphs set the y-axis to zero. This setting allowed the chart to show the accurate spatial difference between the different bar and line values. The bar and line graph deceptive treatments truncated the y-axis to start at values higher than zero, thus exaggerating the difference between the data points (for the bar) or the trendline (for the line). The pie graph deceptive treatment used a 3-D formatting option available in Excel that has the effect of exaggerating the data point in front of the pie by providing it with visual thickness and placing it at a distorted angle that simultaneously makes the back data segments appear smaller. The bubble chart deceptive treatment arbitrarily altered the bubble sizes for both data points, exaggerating the difference between them.

Title Development After building each graph, we then wrote control and exaggerated titles for each pair of graphs that we positioned at the top of each graph (e.g., where "Control Bar Graph" is in Fig. 1). The titles would serve to introduce the data in each graph. The control titles state accurately what trend the data are showing. The exaggerated titles suggest a much steeper trend or greater difference between data points shown in each graph (see Table I).

Between the control and deceptive versions of each graph, and their control or exaggerated titles, each graph type had four possible iterations that a participant might be shown. The survey instrument randomly distributed one version (e.g., 1. control title-control graph; 2. exaggerated title-control

TABLE I
CONTROL AND EXAGGERATED TITLES FOR BAR, LINE, PIE, AND BUBBLE GRAPH TYPES

	Control Titles	Exaggerated Titles
Bar	Home Sales Show Increase From 2015 - 2016	Huge Increase in Home Sales From 2015 – 2016!
Line	Box Office Continues to Climb In 2016	Box Office Smashes Sales Record!
Pie	Hotdogs Preferred by Small Margin Over Other Concessions	Hotdogs are the Clear Favorite Among Baseball Fans!
Bubble	ABC University Graduates More Students than the National Average	ABC University Boasts Much Higher Graduation Rate than the National Average!

graph; 3. control title-deceptive graph; 4. exaggerated title-deceptive graph) of each of the four graph types (bar, line, pie, bubble), so participants viewed a total of four graphs and answered a total of four questions. The order of treatments and graph types was randomized, but participants were shown only one version of each graph type. After seeing the first graph, participants were asked to assess the difference between two data points. For instance, for the pie graph they were asked how much hotdogs outsold the next highest concession, and for the bar they were asked how much home sales had increased from 2015 to 2016. Participants could answer along a six-option continuum that included “a little” on one end and “a lot” on the other (see Fig. 1). We chose this relative (rather than correct/incorrect) approach to our questions because by focusing only on a participant’s perception of difference between two data points (e.g., housing sales from 2015 to 2016), we could control for variables in computational ability among subjects. Keeping responses relative would allow us to compare the mean response level for each treatment to see if there were differences across treatments.

Qualitative Study Questions After responding to the four graph-related questions, participants were asked to rank the four graphs from easiest to understand to most difficult, and to provide an explanation why. Participants were provided an open text box to allow them to elaborate in their own words. Participants were then asked to reflect in their own words why they chose the responses they did when answering the questions about each graph. For instance, participants were asked:

> When answering the question about housing prices represented in the bar graph, please elaborate on what you used to determine the answer to the question.

Participants were again provided an open text box to allow them to share their responses in their own words. We asked participants these questions to help us determine the extent to which the graph elements—including the graph data, axes, titles, or a combination of these—influenced participant choices. Having more insight into these choices would help us determine what elements were more memorable to participants.

Coding of Qualitative Responses To determine whether there were patterns in qualitative responses, we worked separately and used a process of open coding in which we read through responses and developed codes to apply to what we were seeing referred to in participant responses. We then came together to agree on which codes to use and how to apply them. At several points throughout the process of coding the entire dataset, we discussed codes and formalized a short codebook that included codes for chart data (CD), chart shape (CS), chart axis (CA), and chart title (T). Anytime a participant cited more than one of these sources, several codes were logged. We later combined these into a code that indicated more than one source (M). Responses were coded as unclear (UC) either when the participant did not clearly reference one source or another or when they specified that they were not sure what they had used. There was very high initial agreement among codes assigned by each author; in the few instances where we had coded differently from each other, we discussed our perspectives and came to 100% agreement.

RESULTS/DISCUSSION

A total of 329 people provided informed consent and participated in the study between February 23 and November 30, 2018. Table II shows the participant distribution for each of the treatment and graph types.

Fig. 2 shows participants’ mean responses to the graph questions for each treatment of each graph type.

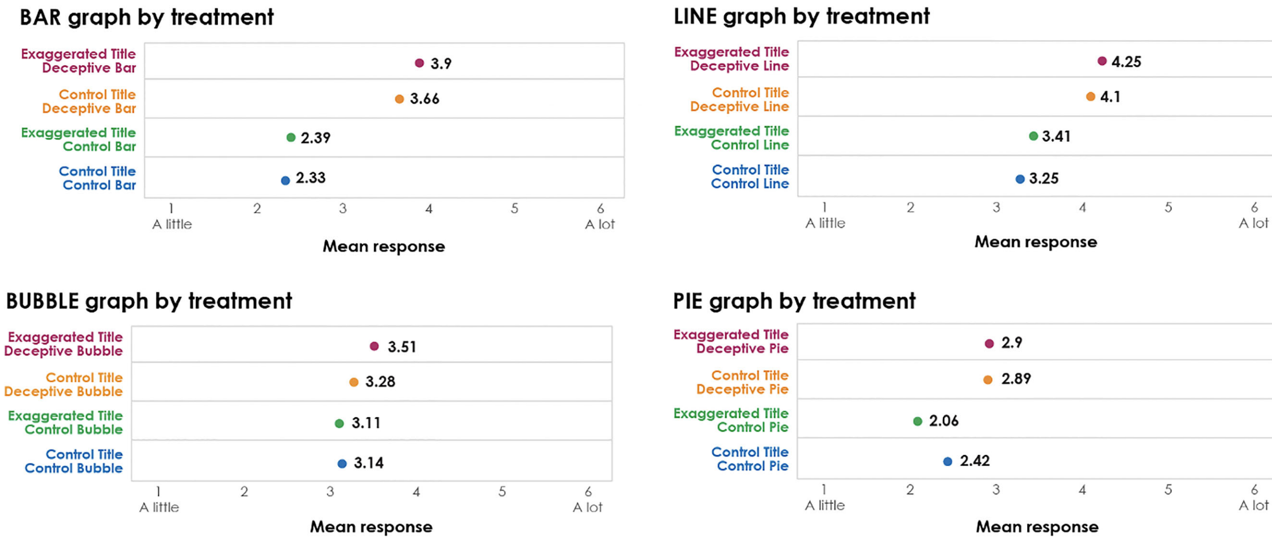

Fig. 2. Mean responses to graphs by treatment.

TABLE II
DISTRIBUTION OF PARTICIPANTS AMONG TREATMENTS

	Control Graph + Control Title	Control Graph + Exaggerated Title	Deceptive Graph + Control Title	Deceptive Graph + Exaggerated Title
Bar	81	85	82	81
Line	81	80	86	81
Bubble	86	81	83	78
Pie	79	84	85	79

To evaluate whether the manipulation of either graph title or graph content impacted participant responses after viewing each graph, a 2 (title) x 2 (content) between-groups ANOVA was calculated for each of the four graph types. The reporting of previous coursework was used as a covariate for each of these analyses in an attempt to control for prior knowledge or experience with such graphics. As can be seen in Table III, the pattern of effects was consistent across all four graph types. Graph title did not reliably impact responses for any of the graphs; however, graph content did significantly impact responses for all four graphs ($p < 0.05$). There was no interaction between these factors, and previous coursework was not a reliable covariate for any of the four graphs. In other words, all participants, whether or not they had reported previously taking a data visualization course, were equally susceptible to deception in graph content.

As confirmed by the exaggerated responses to all deceptive graph types, the visual effect created by a truncated axis (for the bar and line), arbitrary sizing (for the bubble), and 3-D effects (for the pie) all caused participants to perceive a greater difference between data points than they perceived in the control graphs where deceptive tactics were not employed. This finding holds true across varying amounts of data shown in the graphs, as there were only two data points in the bar and bubble graphs, 7 in the pie, and 11 in the line graph. These results confirm the findings of Pandey et al. [33] and O'Brien and Lauer [32], while extending those studies to show deceptiveness in the new genres of a line graph with a truncated axis and a pie graph with a 3-D bevel effect.

Use of Titles Examining the differences in mean scores, it appears that deceptive titles by themselves were influential, though not significantly so, across all graph types (see Fig. 2).

Because the role of exaggeration in titles did not prove to significantly affect how people responded to the graph question, it follows that the potential persuasive effect of titles was largely unnoticed by participants in the study. In our coding of the qualitative response data that participants provided explaining what they used to determine their answers to the graph questions, we were stuck by the extent to which participants *did not cite* using the chart titles. Only 8 responses out of 1064 total qualitative responses provided (0.75%) even referenced using the title, and *none* of those cited using the title exclusively.

Unlike previous studies of (nondeceptive) graphs, we deliberately pursued graphs about noncontroversial themes to test whether the titles of such graphs had the same effect as other studies had found. In our case, they did not. We suggest that when

TABLE III
ANOVA RESULTS (F-VALUE)

	Title	Graph	Course	Title * Graph	Title * Course	Graph * Course	Title * Graph * Course
Bar	1.63	***80.69**	2.27	0.42	2.75	0.08	0.03
Line	0.34	***26.95**	0.38	0.27	1.72	1.57	0.25
Bubble	0.78	***4.11**	0.15	1.1	0.34	1.26	0.03
Pie	1.01	***12.55**	2.68	0.9	0.02	1.08	0.07

Note: $*p < 0.05$. Degrees of freedom for bar was (1, 318); for line (1, 320); for bubble (1, 317); and for pie (1, 316).

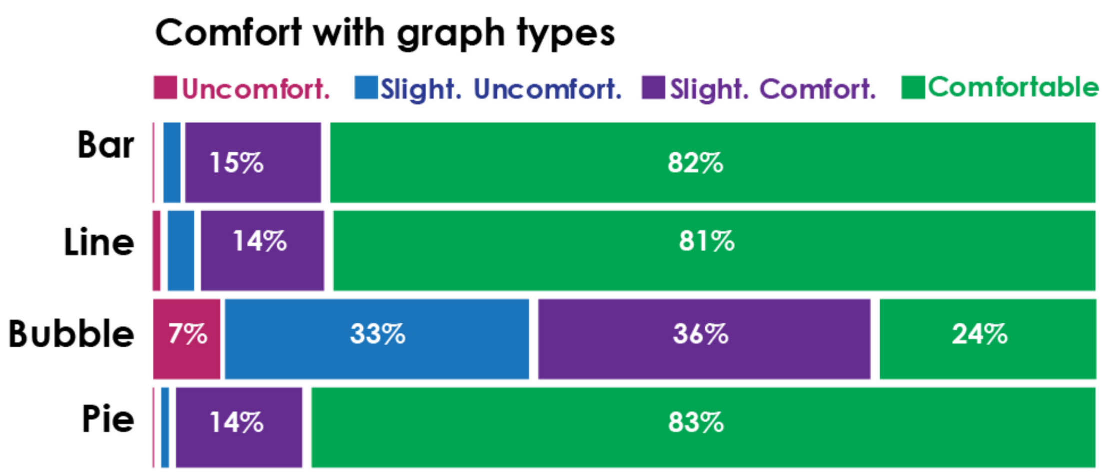

Fig. 3. Comfort level by graph type.

not provoked by particular topics, participants seemed to barely register graph titles and did not cite using titles to determine their responses. This finding is significant because noncontroversial topics remove the potential for personal bias in the reading process. Without such potential for bias, participants appear to rely almost exclusively on the data visualizations themselves in their responses rather than being additionally influenced by the content or slant of the title that accompanies the visualization. These results counter claims made by Borkin et al. [15] about the use of text to aid in the recall of data shown in visualizations and the claims made by Kong et al. [38] about the role of slanted titles. In fact, the complete lack of reference to titles in participants' qualitative responses suggests that data visualizations dominate other textual elements when subjects perceive that the visualization is not attempting to persuade them in some way.

Previous Coursework The results show that participants who cited previously taking a data visualization course were equally susceptible to deceptive tactics as those who had not. This finding may result from data visualization coursework typically being more focused on teaching students how to construct data visualizations than on analyzing ethical considerations or developing critical reading practices. Because data visualizations are often depicted as arhetorical and objective, visualization best practices are commonly communicated in the form of straightforward rules to follow; it may thus not even enter into the minds of participants that deception is possible, especially regarding noncontroversial topics like those used in our study.

Chart Comfort Level A total of 270 participants responded to the "comfort level" questions, and Fig. 3 shows the percentage breakdown of those responses. The results of the graph comfort-level questions show that participants were most comfortable with bar, line, and pie graphs by an overwhelming margin, while fewer reported being as comfortable with the bubble graph. While the exceedingly high percentage of participants who

TABLE IV
EASE WITH WHICH PEOPLE RANKED THE GRAPHS AFTER COMPLETING THE STUDY

	Easiest	Second Easiest	Third Easiest	Fourth Easiest	Total
Bar	133 * 1 = 133	106 * 2 =	48 * 3 =	5 * 4 =	**509**
Pie	109 * 1 = 5	71 * 2 =	91 * 3 =	21 * 4 =	**545**
Line	37 * 1 = 106	100 * 2 =	107 * 3 =	48 * 4 =	**750**
Bubble	13 * 1 = 48	15 * 2 =	46 * 3 =	218 * 4 =	**1053**

cited being "comfortable" with these graph types did not allow for further analysis within subgroupings, this high percentage, combined with all four deceptive graphs proving to be significantly deceptive, shows that the perceived comfort level with a graph type does not appear to mitigate how susceptible people are to deception.

Perception of Graph Readability After viewing and answering a question about each graph, participants were asked to rank, in order from easiest to most difficult, how easy they found it was to read each graph. We calculated ease of readability by multiplying the rank that participants assigned to each graph type (on a scale of 1–4, with 1 being easiest and 4 being most difficult) by the number of participants who assigned that particular rank. So if 133 people ranked the bar graph as the easiest to read, and 106 ranked it as second easiest, we would multiply 106 x 1 (106) and add it to 100 x 2, and so on to achieve a total score. The lower the score, the easier the graph was perceived to read. Table IV shows the scores for each graph type.

These results suggest that participants found the bar graph easiest to read, followed by the pie, the line, and, in a distant fourth place, the bubble. Participants' rankings mirrored, to some extent, the comfort level with each graph type that they reported before viewing the graphs, especially with regard to the bubble graph. Participants reported both the least comfort level with and the most difficulty reading the bubble graph, suggesting that people's perceptions of their lack of familiarity with a graph type likely influences their ease with reading such graph types. The bubble, as we presented it in this study, was an intentionally simplified version of a bubble graph in which only the bubble size (not its position on the *x*-axis) conveyed data, and there were only two data points presented, each clearly labeled. Despite this fact, the bubble graph was overwhelmingly chosen as the least readable graph. We found further evidence about why participants thought this in the qualitative feedback they provided about their rankings. Although those who ranked the bubble as more readable confirmed that the bubble was easy to read because there were fewer data points and "the numbers were clear," the overwhelming sentiment was that people were not familiar with bubble charts or had never seen them before, facts that made reading them more difficult. For instance, participants made the following comments.

> I am not very familiar with bubble charts, which is why I ranked it last.
>
> I have rarely seen a bubble chart in constant use, as compared to other visual data representations, so I was less familiar with how to garner information from it.
>
> I feel like I understand line charts best simply because I have had the most exposure to them. Bubble charts, on the other hand, I have seen very little of.
>
> I have never worked with a bubble chart in the past, and am not comfortable reading them.
>
> I don't know why the bubble charts don't work so well for me.
>
> I have never seen the bubble chart before compared to the other three.
>
> I am not sure how a bubble chart works. The other charts I can understand well.
>
> I am just not a fan of bubble charts
>
> The explanations are similar for all 218 participants out of 270 who ranked bubble graphs the most difficult to read.

Remarkably, people appear to relinquish a great deal of agency to the authority of graphs if they do not have previous experience with a graph type or are unsure about the way that data are being shown in a graph. Despite the fact that the bubble

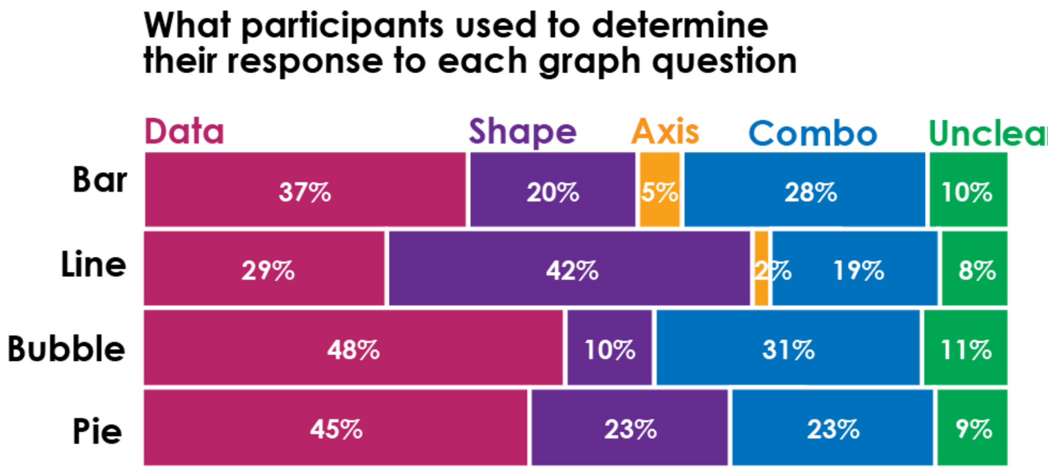

Fig. 4. Coded qualitative response data.

graph showed only two data points, each as a number within a bubble next to another number within a differently sized bubble, participants expressed very little confidence in their ability to figure out what was happening in the graph simply because of its unfamiliar name or style. Their lack of confidence is striking because the graph is actually quite easy to read. In our experience teaching data visualization, we have always taught that a graph is only as good as a reader's ability to understand what it is communicating. But in this case, even the simplest version of an unfamiliar graph type caused people to doubt their own critical reading ability.

Coded Qualitative Data Fig. 4 shows the breakdown for each graph type in what participants reported that they used to determine their responses across all treatments. Some interesting patterns emerge from these data. First, participants cited relying on chart data less in the bar and the line graphs, the two graphs whose deceptive iterations utilized a truncated axis. Those were also the graphs with the largest difference between the mean scores of the control–control graph and the deceptive–deceptive graph. For the bubble and pie graphs, a much higher percentage of participants reported relying on the data to determine their answer. Because neither of these graphs included an axis nor a legend, and the data for both were overlaid directly onto the shape, it is not surprising that people would cite relying on the data exclusively.

The line graph stood out with 42% of participants citing using the shape of the line graph to determine their response. This finding can perhaps be explained by the gestalt law of continuation, which follows that "elements that are arranged on a line or curve are perceived to be more related than elements not on the line or curve" [39]. This simply means that our brain would see the trend line of a line graph as easier to read than the individual data points, so people are likely to notice the shape of the trend line over the data that are labeled at each point. Here are two responses that illustrate the large segment of participants who cited using the shape of the line to determine their response.

> I remembered the pattern of the line chart. I tried to remember the values but I could not. So my answer was based on the pattern of the line chart.

> The connections are more apparent in line, then bar, then pie, then bubble. I think this is so because there is a visual relationship and a relatedness—contiguity—that is established, whereas in the bubble chart the elements are separate.

Perhaps because the line graph included 11 data points, remembering the shape of the line was more manageable than attempting to remember the data points or deciding which of the data to remember. Interestingly, the line appeared to be so persuasive that even when a participant admitted to being aware that the line was perhaps exaggerating the differences between data points (this participant was given a deceptive graph with a truncated axis), he or she was nonetheless persuaded by it.

> The box office sales line chart made it seem as if a large change in revenue was happening. But

> nevertheless, the scaling of the graph and the average slope of the line made the change seem more dramatic and made me want to choose a more dramatic option.

Reflections such as this one show how people can recognize the cognitive dissonance that may occur when they simultaneously identify deceptive tactics used in a graph that they are reading but are nonetheless persuaded by such tactics. It shows how unwilling people may feel to exert agency over the seemingly objective truth of a graph, especially one about a noncontroversial topic.

The qualitative data about the pie and the bubble stood out in how overwhelmingly people cited relying on the graph data or some combination of the data and shape. Unlike the bar or line, the bubble and pie were designed in a way that overlapped the data onto the shape, so a person would not really be able to see the shape without also seeing the data. This approach follows the common design principle of "grouping" data as closely as possible with their corresponding shape to reinforce the relationship between the shape and number. We assume that had we engaged in less effective chart design and separated the data from the shape in a legend positioned off to the side, we might have seen different results. It also appears that participants' unfamiliarity with bubble graphs caused them to look more closely at the data on those graphs, as several examples show.

> I only looked at the numbers on this because I didn't know how a bubble chart worked.

> The values were evident but the visualization of the bubble chart itself is confusing to me.

> I chose the answer based on the percentages inside the bubble, since I have never encountered this type of graph before.

These responses suggest that if a particular graph type being employed is unfamiliar to an audience, labeling all the data clearly is particularly important so that the amounts can be understood by an audience whether or not the audience members are confident about reading the shape or style of the graph.

Conclusion

We live in an era in which it is easier than ever to visualize and circulate technical and scientific data about issues of critical importance to people's lives, their families, their health, and the well-being of their communities. Such visualization practices are vulnerable to both plotting and visualization mistakes—or worse, manipulations by developers actively attempting to misinform. Technical communication researchers must thus continue to study the extent to which data visualizations have the potential to employ deceptive tactics that can work to influence users' perception and understanding. Our study showed that people are deceived by common deceptive tactics used in data visualizations. This finding extended research conducted by Panday et al. [33] and O'Brien and Lauer [32] by studying additional graph types and deceptive tactics, including pie graphs constructed with spatial exaggeration and line graphs constructed with truncated axes. Our study also introduced insightful qualitative data that enabled us to better understand how people were reading graphs and, more important, how they were positioning themselves and their sense of agency as readers in relation to the various graphs that they were studying.

Doing Away With Deceptive Defaults Scholars and teachers often assume that students producing graphs using widely available programs like Excel will also critically examine how those graphs represent their data when they are finished. And yet, the default settings and ease of use of programs like Excel process the data so quickly that novice users may not feel qualified to question the results of the graphs nor equipped to change them in any way. This article's first author has taught a data visualization class to graduate and undergraduate students for nine years and consistently sees how unwilling students are to modify what a program outputs for them, especially if what they see is not especially understandable. She has witnessed how students are willing to relinquish a great deal of agency to the software for performing the precise calculations necessary to turn a table of data into a visually pleasing display.

The programs' defaults adapt the visual range of a graph's axes to the data source. This fact becomes especially problematic when a user wants to compare several graphs from the same dataset and may not notice that each graph is plotted with a different axis range. As a result, the data shown across the several graphs are visually misleading when they are compared with each other.

These programs also make choosing 3-D options entirely too easy for users, when, as our research here has shown, 3-D beveling in pie graphs significantly affects how users interpret differences in data points. To combat this problem, it would be useful to lobby Microsoft, Google, and other software vendors to design their products more responsibly by demoting or hiding altogether unnecessary "eye candy" effects, such as 3-D beveling, that can so easily result in deception, especially among novice users who may be more attracted to such styles.

It would also be useful for these programs to default to a zero y-axis starting point or consistent range rather than that point being relative to the data being displayed. We have seen many highly professional documents that compare two graphs with different y-axes apparently unbeknownst to the graph developer. Students can be taught to modify these settings in Excel or, ideally, open the Excel graphs in Adobe Illustrator or Microsoft PowerPoint to enable more extensive editing and annotation capabilities. It would also be useful to introduce more advanced students to open source graphing solutions such as R and Python, which are more difficult to learn but ultimately provide greater control over the visualized output.

Graphs Versus Titles and Text Our research emphasizes how much more compelling graphs are in communicating data than paragraphs of text [31] or titles. Engaging with principles of design, we believe that graphs provide a *visual focal point* for readers that directs their eyes to the data visualization as the prominent figure almost immediately and leaves the textual data to recede into the ground area, especially if that text is communicating information about a noncontroversial topic. Graphs often employ color, contrast, and other gestalt and design principles that explain why people may be more compelled by their presence on a page over the surrounding text.

Considering both the susceptibility of people to deceptive tactics and the inability of correct paragraphs of text [31] or exaggerated titles to significantly influence how people discern differences in graph data, we conclude that emphasizing graph literacy across all areas of our pedagogy is more important than ever. Graphs are not objective vehicles, delivering visualized data in an arhetorical and straightforward manner; rather, they are rhetorical constructions, subject to deception, and often developed using software that does not make it easy to modify them once they have been developed. Faculty and programs should continue to pursue this fact with vigilance if we are going to prepare critically conscious professionals for the future.

Future Research Further research currently underway in our lab includes eye-tracking research to analyze the patterns of eye movements when deceptive data visualizations are combined with both control and exaggerated titles. This research seeks to provide some understanding of the interplay between graphical and textual elements on the screen. Combined with participants' free form responses from this study, incorporating eye-tracking data should provide a richer context of how people read and understand information from graphs.

Our future research also involves large-scale analysis of persuasive graph use. For instance, the first author is working with a research group that has collected 2.4 million images from years of Kickstarter campaigns and will be processing the data visualizations to determine how many employ deceptive tactics. For those that do, we will examine whether the data warrant such tactics (e.g., use of a truncated axis) or whether they were used to exaggerate differences between data points.

With the onset of the worldwide pandemic caused by COVID-19, as well as the continued concerns around climate change and other issues of social justice, technical communication researchers need to design more studies to understand the complex nature of data visualizations and the ways that they can be effectively constructed to communicate information. This is an interdisciplinary research area spanning technical communication, media studies, psychology, and computer science. With increasingly shorter attention spans among readers and the desire among developers to quickly and effectively communicate information for user consumption, developers must take the extra step to ensure that the information that they are presenting in the form of data visualizations refrains from engaging in deceptive tactics. With our backgrounds in technical content development, our emphasis on the user, and our familiarity with UX research methods, technical communication researchers are perfectly positioned to take up this call.

ACKNOWLEDGMENT

The authors would like to thank their study participants and the instructors who so graciously distributed the study to their students.

References

[1] E. R. Tufte, *The Visual Display of Quantitative Information*. Cheshire, CT, USA: Graphics Press, 1983.
[2] L. Manovich, "What is visualisation?" *Vis. Studies*, vol. 26, no. 1, pp. 36–49, 2011, doi: 10.1080/1472586X.2011.548488.
[3] E. R. Tufte, *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT, USA: Graphics Press, 1997.
[4] K. Schriver, *Dynamics in Document Design: Creating Texts for Readers*. New York, NY, USA: Wiley, 1997.
[5] E. R. Tufte. Baseline for amount scale. [Online]. Available: https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00003q
[6] Vox.com, "Shut up about the y-axis. It shouldn't always start at zero," *Video*, 2015. Accessed: Sep. 19, 2019. [Online]. Available: https://www.youtube.com/watch?v=14VYnFhBKcY
[7] N. Yau. (2019). Line chart baselines do not have to start at zero (the process #39). Accessed: Sep. 19, 2019. [Online]. Available: https://flowingdata.com/2019/05/09/process-39/
[8] J. Muyskens, "Most of Trump's charts skew the data and not always in his favor." *Washington Post*. Accessed: Oct. 31, 2016. [Online]. Available: https://www.washingtonpost.com/graphics/politics/2016-election/trump-charts/
[9] L. Meloncon and E. Warner, "Data visualizations: A literature review and opportunities for technical and professional communication," in *Proc. IEEE Int. Prof. Commun. Conf.*, 2017, pp. 1–9.
[10] A. Kirk. *Data Visualization: A Successful Design Process*. Birmingham, AL, USA: Packt Publishing, 2012.
[11] C. Skelton. (2018 Jun.). Bar charts should always start at zero. But what about line charts? [Online]. Available: http://www.chadskelton.com/2018/06/bar-charts-should-always-start-at-zero.html?m=1
[12] N. Yau. (2015). Sometimes the y-axis doesn't start at zero, and it's fine. Accessed: Sep. 19, 2019. [Online]. Available: https://flowingdata.com/2015/11/23/sometimes-the-y-axis-doesnt-start-at-zero-and-its-fine/
[13] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 9, pp. 1520–1536, Sep. 2012.
[14] M. A. Borkin *et al.*, "What makes a visualization memorable?," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2306–2315, Dec. 2013.
[15] M. A. Borkin *et al.*, "Beyond memorability: Visualization recognition and recall," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 1, pp. 519–528, Jan. 2016.
[16] E. Wall, L. M. Blaha, C. L. Paul, K. Cook, and A. Endert, "Four perspectives on human bias in visual analytics," in *Cognitive Biases in Visualizations*, G. Ellis, Ed. Cham, Switzerland: Springer, 2018, pp. 29–42, doi: 10.1007/978-3-319-95831-6_3.
[17] D. Borland and R. M. Taylor Ii, "Rainbow color map (still) considered harmful," *IEEE Comput. Graph. Appl.*, vol. 27, no. 2, pp. 14–17, Mar./Apr. 2007.
[18] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 9, pp. 1520–1536, Sep. 2012.
[19] C. A. Brewer, "Spectral schemes: Controversial color use on maps," *Cartography Geographic Inf. Syst.* vol. 24, no. 4, pp. 203–220, 2013„ doi: 10.1559/152304097782439231.
[20] B. E. Rogowitz, L. A. Treinish, and S. Bryson, "How not to lie with visualization," *Comput. Phys.*, vol. 10, no. 3, pp. 268–273, 1996, doi: 10.1063/1.4822401.
[21] W. S. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *J. Amer. Stat. Assoc.*, vol. 79 no. 387, pp. 531–554, 1984, doi: 10.2307/2288400.
[22] A. Tal and B. Wansink, "Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy," *Public Understanding Sci.*, vol. 25, no. 1, pp. 117–125, 2014, doi: 10.1177/0963662514549688.
[23] E. Kersten van Dijk, W. IJsselsteijn, and J. Westerink, "Deceptive visualizations and user bias: A case for personalization and ambiguity in PI visualizations," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 588–593, doi: 10.1145/2968219.2968326.
[24] H. Wainer, "How to display data badly," *Amer. Stat.*, vol. 38, no. 2, pp. 137–147, 1984, doi: 10.2307/2683253.
[25] I. Cho, R. Wesslen, A. Karduni, S. Santhanam, S. Shaikh, and W. Dou, "The anchoring effect in decision-making with visual analytics," in *Proc. IEEE Conf. Vis. Anal. Sci. Technol.*, 2017, pp. 116–126, doi: 10.1109/VAST.2017.8585665.
[26] E. Dimara, A. Bezerianos, and P. Dragicevic, "The attraction effect in information visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 471–480, Jan. 2017.
[27] L. Harrison, D. Skau, S. Franconeri, A. Lu, and R. Chang, "Influencing visual judgment through affective priming," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2013, pp. 2949–2958, doi: 10.1145/2470654.2481410.
[28] A. C. Valdez, M. Ziefle, and M. Sedlmair, "Priming and anchoring effects in visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 1, pp. 584–594, Jan. 2018.
[29] J. Hullman and N. Diakopoulos, "Visualization rhetoric: Framing effects in narrative visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2231–2240, Dec. 2011.
[30] H. Kong, Z. Liu, and K. Karahalios, "Frames and slants in titles of visualizations on controversial topics," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2018, pp. 1–12, doi: 10.1145/3173574.3174012.
[31] D. Huff and I. Geis, *How to Lie With Statistics*. New York, NY, USA: W. W. Norton Co., 1993.
[32] G. E. Jones, *How to Lie With Charts*. Santa Monica, CA, USA: LaPuerta Books Media, 2011.

[33] S. O'Brien and C. Lauer, "Testing the susceptibility of users to deceptive data visualizations when paired with explanatory text," in *Proc. 36th ACM Int. Conf. Design Commun.*, 2018, pp. 1–9, doi: 10.1145/3233756.3233961.
[34] A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini, "The persuasive power of data visualizations," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2211–2220, Dec. 2014.
[35] M. Monmonier, *How to Lie With Maps*, 3rd ed. Chicago, IL, USA.: Univ. Chicago Press, 1991.
[36] A. Cairo and D. Bihanic, Eds., "Graphics lies, misleading visuals," in *New Challenges for Data Design*. London, UK: Springer, 2015, doi: 10.1007/978-1-4471-6596-5.
[37] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini, "How deceptive are deceptive visualizations? An empirical analysis of common distortion techniques," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 1469–1478, doi: 10.1145/2702123.2702608.
[38] S. Pinker. "A theory of graph comprehension," in *Artificial Intelligence and the Future of Testing*. New York, NY, USA: Psychology Press, 1990.
[39] J. Talbot, V. Setlur, and A. Anand, "Four experiments on the perception of bar charts," *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 12, pp. 2152–2160, Dec. 2014.
[40] J. Ritchie, D. Wigdor, and F. Chevalier, "A lie reveals the truth: Quasimodes for task-aligned data presentation, in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–13, doi: 10.1145/3290605.3300423.
[41] H. Kong, Z. Liu, and K. Karahalios, "Trust and recall of information across varying degrees of title-visualization misalignment," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2019, pp. 1–13, doi: 10.1145/3290605.3300576.
[42] (2019). 7 Gestalt principles of visual perception: Cognitive psychology for UX. Accessed: Feb. 19, 2020. [Online]. Available: https://www.usertesting.com/blog/gestalt-principles#continuity

Claire Lauer received the Ph.D. degree in Rhetoric, Composition, and the Teaching of English from the University of Arizona, in 2006. He is an Associate Professor of Technical Communication and User Experience with Arizona State University, Mesa, AZ, USA. She has published on data-driven versus manual qualitative language analysis research, how technology impacts creative thinking and design, and how the work of technical communicators has evolved within the technological workplace. She researches how people are susceptible to deceptive tactics used in graphs, how to communicate scientific information to public audience, and how to effectively design data-driven interfaces for researchers. Prof. Lauer is the past Chair of ACM's Special Interest Group for the Design of Communication (SIGDOC) and served as the Vice Chair of Operations on the SGB Executive Council of the Association for Computing Machinery.

Shaun O'Brien received the master's degree in Technical Communication from Arizona State University (ASU), Mesa, AZ, USA, in 2017, where he is currently working toward the graduate degree in Human Systems Engineering. He is an Editor of research and special projects for the Edson College of Nursing and Health Innovation, ASU.