Insufficient Information to Trust: Visualizations of Averages Rated Less Trustworthy than Visualizations of Individual Data

Hamza Elhamdadi[1] (helhamdadi@stonehill.edu) & Jeremy B. Wilmer[2] (jwilmer@wellesley.edu)
[1] Department of Computer Science, Stonehill College
[2] Department of Psychology, Wellesley College

Abstract

Scientific results are often communicated with bar charts that display category means while omitting the underlying distribution of individual observations. Such visual simplification is commonly assumed to aid accessibility, yet direct evidence about how this design choice affects perceived trustworthiness is limited. We compared initial trustworthiness judgments for average-only bar charts versus sina plots, a data-showing alternative that displays individual observations within each category (Sidiropoulos et al., 2018). In a preregistered online study (N = 162; Prolific), participants viewed two visualizations (bar chart and sina plot) of the same textbook-sourced scientific result (order randomized), completed graph-reading estimates, rated each graph's trustworthiness on a 0–100 scale, rated six antecedents of trustworthiness (e.g., accuracy, completeness, bias), and provided free-response justifications. Sina plots were rated as more trustworthy than bar charts (d = .66), with 70% of participants favoring sina plots versus 17% favoring bar charts; the effect remained substantial when restricted to the first graph viewed (d = .45). Qualitative analysis suggested that bar charts were frequently criticized as providing insufficient information to judge trustworthiness, whereas sina plots more often elicited default trust. Together, these results suggest that showing individual data points can increase perceived trustworthiness even while increasing visual density.

Keywords: trustworthiness; trust; bar graph; sina plot; averages; statistics; data visualization; graph interpretation

Introduction

Although access to scientific information has expanded dramatically in recent decades (Hamilton et al., 2023; NIH, 2025; NSF, 2022), this expansion has not yielded corresponding gains in public understanding, consensus, or trust (Davern et al., 2022; Salmon et al., 2015; Tyson & Kennedy, 2024; Van der Linden et al., 2017; West & Bergstrom, 2021). Bridging this gap requires more than increasing the volume of available data; it also requires translating scientific evidence into forms that support shared and trustworthy knowledge. Achieving this goal depends in part on evidence-based guidance for how scientific information is communicated.

Data visualization plays a central role in this translation by shaping what viewers notice and how they interpret quantitative evidence (Franconeri et al., 2021). Despite its ubiquity in scientific communication, however, the role of visualization in shaping public trust has only recently become the focus of sustained empirical study (Chita-Tegmark et al., 2021; Elhamdadi et al., 2023; McKinley et al., 2025; Padilla et al., 2023; Pandey et al., 2023; Song et al., 2025; Yang, Cai, et al., 2024; Yang, Mortenson, et al., 2024). This growing literature has begun to identify specific design features that influence perceived trustworthiness.

Many of these features reflect a common design tension: how much information to show, and in what form. Higher perceived trust has been linked to the presence of data source citations (Elhamdadi et al., 2023; McKinley et al., 2025; Yang, Cai, et al., 2024), restrained visual presentation (Song et al., 2025), and alignment with viewers' prior beliefs (Yang, Cai, et al., 2024). Yet increasing the amount of visible data can also increase visual complexity, which has been associated with lower trust ratings (Elhamdadi et al., 2023; Padilla et al., 2023).

This tension is especially salient in decisions about whether to present data in aggregated or disaggregated form. Aggregation is often motivated by principled concerns: reducing visual clutter, minimizing cognitive load, and highlighting central tendencies thought to be most relevant for inference or decision-making. Within this framework, abstraction is treated as a virtue, particularly for non-expert audiences.

Disaggregated displays such as text ratios, icon arrays, and quantile dot plots can improve understanding and sustain trust in some contexts (Galesic et al., 2009; Gigerenzer & Hoffrage, 1995). At the same time, aggregate displays, most notably bar charts of means or percentage, are often described as transparent and trustworthy, particularly when contrasted with more embellished infographic styles (McKinley et al., 2025). However, aggregation necessarily removes information about variability and the distribution of individual observations—information that may be relevant for evaluating what the claim is based on and how credible it is. These mixed findings have reinforced a long-standing emphasis on visual simplification in scientific communication, under the assumption that abstraction reduces clutter and improves accessibility for general audiences (Eberhard, 2023; Franzblau & Chung, 2012; Kerns & Wilmer, 2021; Turchioe et al., 2019).

As a result, quantitative data are frequently summarized using average-only bar charts that omit the underlying distribution of individual observations. Despite how common and consequential this design choice is, its implications for perceived trustworthiness remain poorly understood. In particular, prior work has not directly compared trust judgments for average-only bar charts with visualizations that make the underlying data fully visible, such as sina plots (Sidiropoulos et al., 2018). In this work, we directly compare initial trustworthiness judgments of aggregate bar charts and data-showing sina plots.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Methods

We provide our open procedures (a one-page pdf with screenshots of what participants saw per page, with notes on design), open data (raw responses per question, qualitative coding), and preregistered analyses at osf.io/8gpum.

Procedure

Participants were randomly assigned to one of four visualization topics and then viewed two visualizations of the same dataset (a sina plot and a bar chart) presented one at a time in randomized order. For each visualization, participants first estimated group averages and extrema for two independent-variable groups. They then rated the visualization's trustworthiness on a 101-point scale (0 = not at all, 50 = somewhat, 100 = very much) and provided a brief written justification. To reduce noise from the continuous slider in Qualtrics, the selected numerical value was displayed during adjustment.

Participants next rated each visualization on six antecedents of trustworthiness adapted from prior work (Kelton et al., 2008): believability, clarity, bias, accuracy, completeness, and obsolescence, again providing brief written justifications (see **Open Procedure** for item wording). After completing these tasks for both visualizations, participants answered demographic questions (age, gender, education, survey experience, and coursework in psychology and statistics) and were asked to guess the study's hypothesis.

Stimuli

Figure 1 presents the stimuli shown to participants. The bar chart stimuli were drawn from four widely used introductory psychology and statistics textbooks (Gray & Bjorklund, 2018; Grison & Gazzaniga, 2022; Kalat, 2022; Myers et al., 2024), ensuring that the stimuli were not only familiar in form but also ecologically valid examples of data visualizations drawn from real educational materials used in classroom settings. Each bar chart depicted a well-known, published empirical result (DiMascio et al., 1979; Festinger & Carlsmith, 1959; May et al., 1993; Rahman et al., 2004).

For each bar chart, we constructed a corresponding sina plot visualization based on the reported summary statistics from the original publication. Specifically, we generated evenly spaced quantiles from a normal distribution with the reported mean and standard deviation, and then imposed skew as needed to respect the known bounds of each measurement scale. This procedure preserved the reported central tendency and variability while yielding a plausible distribution of individual data points.

From here on, we refer to these visualization pairs using a single keyword corresponding to the independent variable depicted: AGE, CLINICAL, SOCIAL, and GENDER.

Participants

We collected data using the remote survey software Qualtrics and the crowd-sourcing platform Prolific (www.prolific.com). Participants were required to be fluent English speakers. A total of 186 participants started the survey and provided at least partial data. As preregistered, participants were excluded only if they failed to complete the study; 24 participants exited before completion and were therefore excluded, leaving 162 responses, distributed among our four chart topic conditions as such: AGE - 42, CLINICAL - 38, GENDER - 33, SOCIAL - 49. Of these 162 participants, 80 identified as men, 79 as women, two as non-binary, and one preferred not to disclose their gender. Participant ages ranged from 22 to 76 years old (M=44.53, SD=12.66).

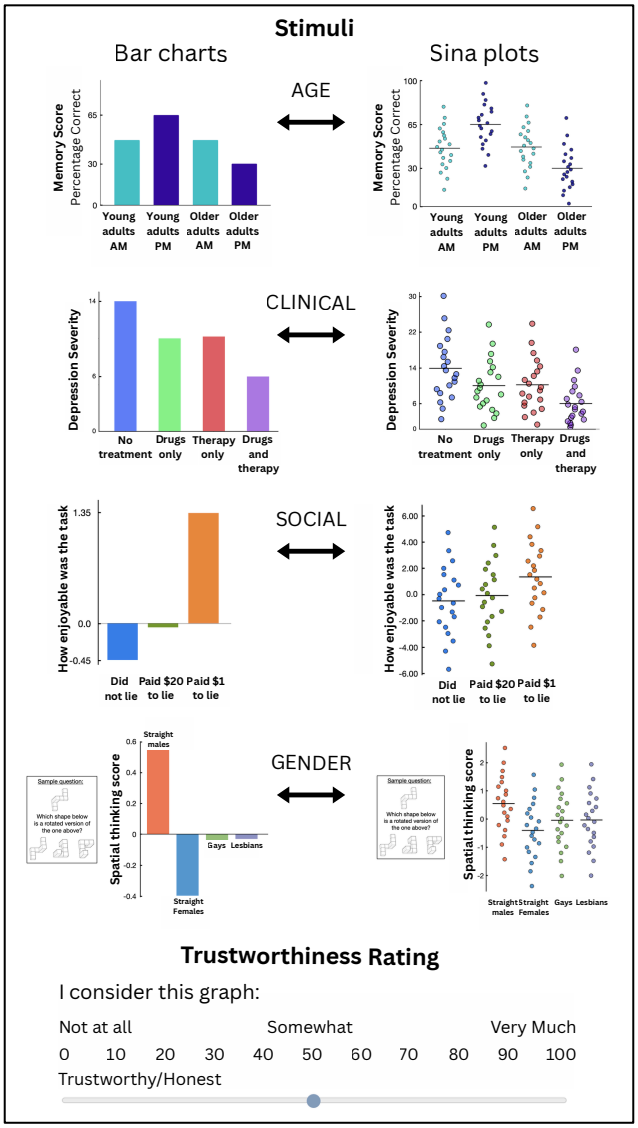

Figure 1: Stimuli and Trustworthiness Rating Bar chart stimuli (left) were adapted from introductory statistics textbooks. Sina plot stimuli (right) were generated from the reported statistics of the original published studies. The sina plots preserved the original axis labels and colors, with y-axis ranges adjusted as needed to display the full data distribution. Font sizes have been enlarged here for readability; original stimuli and the full procedure are available at osf.io/8gpum. Trustworthiness ratings were elicited using the scale shown at bottom. Antecedent ratings were collected using the same slider format, with the prompt text modified to match the specific antecedent being rated.

Measures

Recent data visualization research has emphasized the need for more comprehensive approaches to measuring trust and trustworthiness in data visualizations than single-item questionnaires (Elhamdadi et al., 2023). We therefore included multiple measures to provide a richer and more nuanced picture of trustworthiness judgments.

Subjective Trustworthiness Ratings We elicited subjective judgments of trustworthiness by asking participants to rate their agreement with the statement "I consider this graph: Trustworthy / Honest" using a scale of 0 ("Not at all") to 50 ("Somewhat") to 100 ("Very Much"). Since asking about trustworthiness alone may be insufficient to capture the multidimensional nature of trust (Elhamdadi et al., 2023), we use this score as a baseline measure and supplement it with additional quantitative responses to similar statements about several antecedents of trust. We compare this quantitative trustworthiness measure within and between subjects to assess whether trust is higher in sina plot or bar chart visualizations (i.e., in graphs of individual data vs. aggregates).

Free-Response Justifications of Trustworthiness Ratings To better understand the reasoning underlying participants' quantitative responses, we asked participants to justify their trustworthiness ratings: "In a few sentences, please explain why you rated the chart's trustworthiness/honesty as you did." Below, we conduct a thematic analysis of these responses (see the section entitled "Qualitative analysis").

Trustworthiness Antecedents Exploring the antecedents of trust (factors theorized to occur before and lead to trust) can provide a richer understanding of trustworthiness judgments (Mayer et al., 1995, Elhamdadi et al., 2023). For a more comprehensive yet directed understanding of the factors that influence participants' trustworthiness judgments, we collect responses to an inventory of trustworthiness antecedents adapted from prior work (Kelton et al., 2008). Using the same scale as the general trustworthiness measure, participants were asked to rate their agreements with several statements in the form "I consider this graph <antecedent>".

Estimation Accuracy To ensure that participants exhibited some understanding of the underlying data presented by the visualizations before making trustworthiness judgments, we tasked participants with estimating the average of two of the independent variable groups. Additionally, we explore whether there is a relationship between estimation accuracy and ratings of trustworthiness.

Results

Trustworthiness Ratings

As shown in **Figure 2** (leftmost panel), sina plots were rated as more trustworthy than bar charts, with a medium-to-large effect size (d = .66). More than four times as many participants rated the sina plot as more trustworthy (114; 70%) than rated the bar chart as more trustworthy (28; 17%).

To guard against an influence of outliers, we conducted an equivalent non-parametric analysis on rank-ordered data (Conover & Iman, 1981); the resulting effect size was very similar (d = .63). The effect was statistically significant for all four plotted scientific results (all $p \leq .005$).

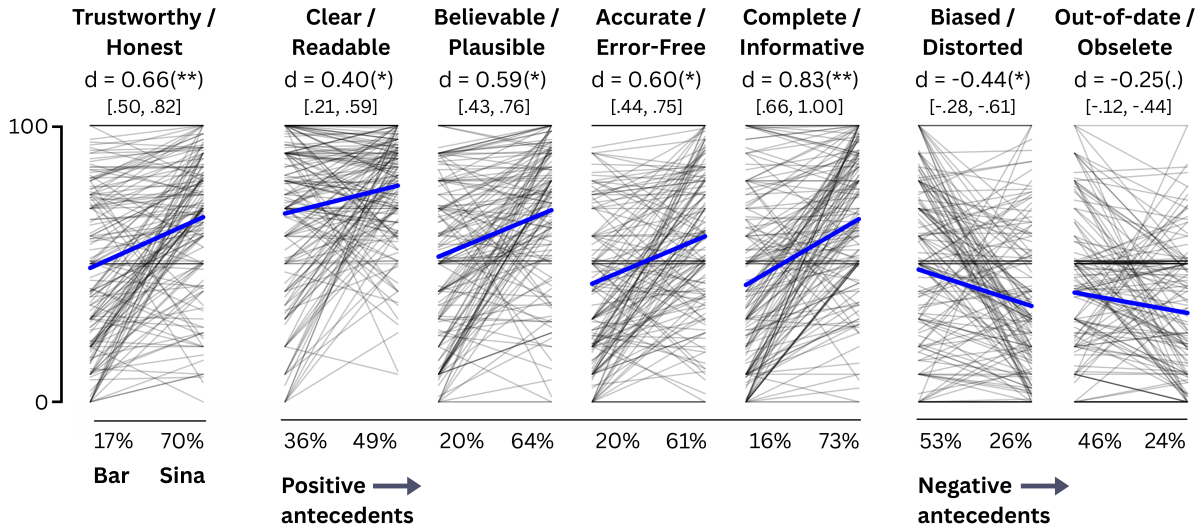

Figure 2: Ratings favored graphs that showed the data. Ratings of trustworthiness (left), and positive (middle) and negative (right) antecedents of trustworthiness are shown. Black sloped lines connect each participant's rating of the bar chart (left) to their rating of the sina plot (right); blue sloped lines show the group mean. Below each subplot, we report the percentage of participants (out of N = 162) who rated the bar chart (left) and the sina plot (right) higher (percentages do not sum to 100% because some participants gave equal ratings). Also shown above each subplot is the standardized mean difference (SD units), categorized by magnitude (after Sawilowsky, 2009; as small (. → .00–.34), medium (* → .35–.64), large (** → .65–.99), or very large (*** → ≥1.00), with 95% confidence intervals in brackets.

The effect sizes varied somewhat across content domains (AGE: d = .67, 95% CI [.33, 1.02]; CLINICAL: d = .76, 95% CI [.46, 1.06]; SOCIAL: d = .38, 95% CI [.12, .65]; GENDER: d = 1.19, 95% CI [.69, 1.70]). An ANCOVA nevertheless revealed only a trend-level interaction of chart type (bar vs. sina) with plotted content (p = .06), indicating that future work will be needed to assess content-specific differences.

Trustworthiness antecedents, too, were all rated as more favorable for sina plots than for bar charts (**Figure 2**). Importantly, sina plots were rated both higher for positive antecedents (clarity, believability, accuracy, completeness), and lower for negative antecedents (bias, obsolescence), demonstrating the thoughtfulness and intentionality of the responses. While all results were highly statistically significant (p < .0001), the largest differences were for completeness, accuracy, and believability, and the smallest were for obsolescence, clarity, and bias (**Figure 2**).

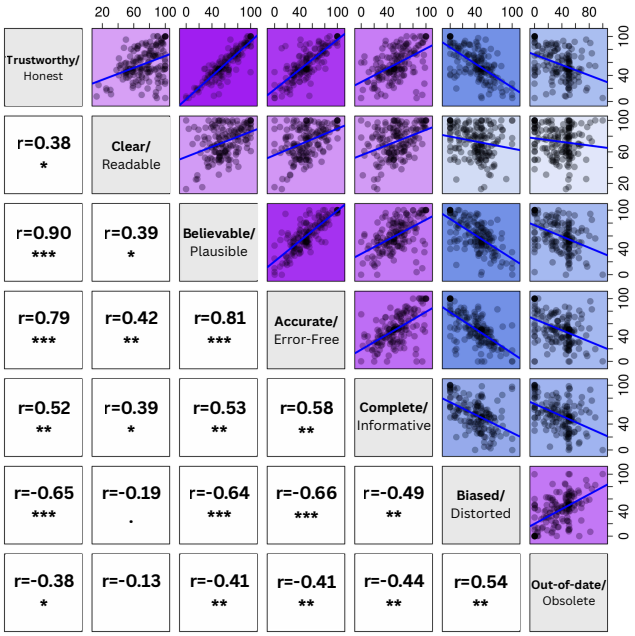

Figure 3. Trustworthiness is strongly correlated with its antecedents. The correlation matrix relates trustworthiness (top-left variable) to six antecedents (remaining diagonal variables). The diagonal contains axis labels. Blue lines show the least-squares fit (physical slope=r). Colors indicate direction and magnitude of the association (purple=positive, blue=negative; saturation∝|r|). Statistically significant correlations (p < .05, two-tailed) are categorized after Cohen (1988): small (· = .00–.19), medium (* = .20–.39), large (** = .40–.59), very large (*** ≥ .60).

Linking trustworthiness to antecedents

To examine the semantic underpinnings of trustworthiness ratings, we correlated them with six other rated characteristics of the same graphs: clarity, believability, accuracy, completeness, bias, and obsolescence. Following prior work, we treat these characteristics as antecedents of trust (Kelton et al., 2008). **Figure 3** shows the pairwise correlations between trustworthiness and these antecedents.

Positively valenced antecedents (clarity, believability, accuracy, completeness) were positively correlated with trustworthiness, whereas negatively valenced antecedents (bias, obsolescence) were negatively correlated. This opposing pattern is consistent with construct-sensitive responding rather than uniform or careless ratings.

Trustworthiness correlated most strongly with believability (r = .90, 95% CI [.86, .92]), followed by accuracy (r = .79, 95% CI [.73, .84]), bias (r = –.65, 95% CI [–.73, –.55]), completeness (r = .52, 95% CI [.39, .62]), clarity (r = .38, 95% CI [.24, .50]), and obsolescence (r = –.38, 95% CI [–.50, –.24]). This graded pattern likely reflects differences in conceptual proximity to trustworthiness, with believability playing a larger role than more distal factors such as obsolescence.

The analyses in **Figure 3** average trust ratings across bar charts and sina plots. To test whether this averaging obscured chart-specific structure, we repeated the analyses separately by chart type, computing all 21 pairwise correlations among the seven trust measures. The resulting correlation vectors were highly similar (r = .97), indicating essentially identical interrelationships across chart types. Raw data for these and all analyses are available in the **Open Data** at osf.io/8gpum.

Dissociating trustworthiness from training and understanding

Expertise could plausibly influence trustworthiness judgments. For example, sina plots may require prior statistical training to read and interpret, and thus to trust, whereas bar charts may be more readily trusted by viewers with less statistical training. To examine this possibility, we tested whether trust judgments (separately for each chart type) were correlated with chart-reading performance and with prior coursework in statistics or psychology.

Trust ratings were largely independent of both performance and coursework: all such correlations were non-significant, despite the large sample size (N = 162). Although null or near-null correlations could in principle reflect measurement noise, this explanation is unlikely given the presence of substantial correlations elsewhere in the data. As shown in **Figure 3**, trust was strongly correlated with its antecedents; and as shown in **Figure 4**, trust judgments were strongly correlated across chart types (r = .48, 95% CI [.35, .59]), training was strongly correlated across domains (r = .51, 95% CI [.39, .62]), and performance showed a moderate cross-chart correlation (r = .20, 95% CI [.05, .34]).

One might expect individuals with greater training in statistics or psychology to make fewer errors when reading these graphs, which depict statistical results adapted from introductory psychology textbooks. Yet prior work has found little evidence for such relationships: even well-educated viewers commonly misread average-only graphs, whereas lay viewers often accurately interpret graphs that display individual data (Cui et al., 2024; Kerns & Wilmer, 2021; Reimann, 2026; Wang et al., 2025). Consistent with this prior work, we found no reliable relationship between participants' statistical or psychological training and graph-reading performance; all associations were non-significant (**Figure 4**). These results suggest that misinterpretation of

average-only graphs reflects a broadly shared perceptual or inferential challenge, rather than one readily corrected by formal training.

Trustworthiness by Order of Exposure

Conceivably, the advantage of data-showing sina plots over average-only bar charts could arise primarily for the second stimulus, when participants can make an explicit sequential comparison. Such an effect would still be informative, suggesting that when shown both formats in close succession, viewers come to trust sina plots and mistrust bar charts. However, it would carry less practical significance, because outside the laboratory people rarely encounter two charts of the same scientific result in immediate succession.

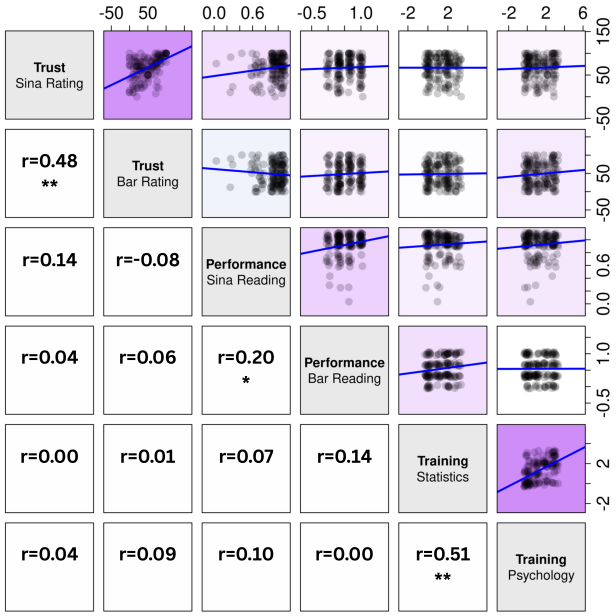

Figure 4. Trustworthiness is largely independent of present performance and prior training. Correlation matrix relating trustworthiness ratings to estimation accuracy (by chart type) and participant statistics and psychology training, formatted as in **Figure 3**. Performance is the percentage of plausible average and extrema values read from the graph, and training responses are treated as an ordinal variable (see **Methods** and **Open Procedure**).

With this possibility in mind, we analyzed trustworthiness ratings separately for the first and second graphs viewed (**Figure 5**). When restricted to the first viewed graph, the sina plot advantage remained substantial: a 12.5-point difference on the 100-point scale (95% CI [3.9, 21.1], $p < .005$), corresponding to a medium effect size ($d = .45$, 95% CI [.14, .77]). The second ratings showed a numerically larger effect, with a 24.5-point sina plot advantage (95% CI [15.9, 33.2]; $d = .88$, 95% CI [.56, 1.21]; $p < 1 \times 10^{-7}$). However, the interaction between rating order and chart type fell short of statistical significance ($p = .11$).

In sum, although there is suggestive evidence that explicit comparison may amplify the effect, this difference did not reach statistical significance and thus remains for future work to establish conclusively. To guard against the influence of outliers, we conducted parallel non-parametric analyses on rank-ordered data (Conover & Iman, 1981); the resulting effect sizes were very similar ($d = .45$ and $d = .82$, respectively).

Qualitative Analysis

To identify common themes in participants' justifications of their trustworthiness ratings, and to better understand why sina plots were rated more trustworthy than bar charts, we conducted a preliminary thematic analysis (Naeem et al., 2023) of free-response justifications provided for the first chart viewed by each participant. Responses were deidentified, assigned anonymous participant IDs (P1-162), and randomized to avoid coding by chart type or topic.

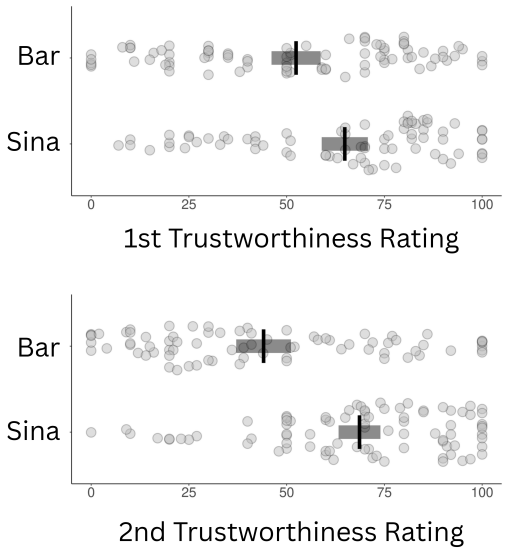

Figure 5. Data-showing graphs are rated as more trustworthy for both first and second judgments. The figure shows trustworthiness ratings for bar charts and sina plots when each was the first graph viewed (top) and the second graph viewed (bottom). Each dot represents a single participant's rating; the vertical black line indicates the mean, and the horizontal gray bar shows the 95% CI.

Both authors jointly generated a list of keywords from the first ten responses. These keywords were then combined into conceptually-related codes (e.g., data seems real and fabrication of data were subsumed under a single code with potential values of yes, no, somewhat, or left blank). Each author then independently coded twenty responses, introducing new codes as needed. These codes were used to compute inter-rater reliability (κ=.76; "substantial" according to Landis & Koch, 1977). The resulting codebook guided focused coding (Charmaz, 2006) of the remaining responses. Qualitative codes were linked to quantitative trustworthiness ratings after coding was complete.

To facilitate interpretation, responses were grouped by trustworthiness rating into three bins: untrustworthy (0-40), undecided (40-60), and trustworthy (60-100). This grouping distinguishes participants who clearly distrusted or trusted the chart from those with ratings nearer to the midpoint.

Examining themes by rating group suggested a preliminary conceptual model of how participants reasoned about visualization trustworthiness in this experimental context. Participants in the undecided group were more likely to mention missing provenance data (e.g., visualization creator, data source, experimental design) as limiting their ability to judge trustworthiness. For example, P98 explained that they "don't know anything about who produced the graph and how they obtained the information" and P148 noted that "it's impossible to see how the study was conducted."

Conversely, participants with ratings farther from the midpoint (both higher and lower trustworthiness) often grounded their judgment in the chart content or data itself, explicitly evaluating whether the depicted results seemed plausible or aligned with their expectations. For example, P134 has "a hard time believing that straight female score would be that much lower than straight males", whereas P94 considered the data credible because "the numbers do make sense for the age groups: older people having lower scores and the younger generation having higher ones."

We next examined these patterns separately by chart type. The broad distinctions by trustworthiness rating persisted, but two additional chart-specific patterns emerged. First, participants who viewed the bar charts were more likely to explicitly state that the visualization provided insufficient information to judge trustworthiness. For example, P5 wrote "It's hard for me to tell just through the bar graph whether this is something trustworthy or not" and P88 noted that the bar graph "shows numbers [but] doesn't show much more of anything. [H]ow can I estimate right if it only shows us only a couple of things." Second, participants who viewed sina plots tended to treat them as trustworthy by default, expressing an absence of reasons to doubt rather than pointing to specific evidence, as P81 explained, "I will just assume positive intent and say it's fairly trustworthy," and P159 stated "I have no reason not to believe this chart."

Discussion

In this work, we used quantitative and qualitative self-report measures to assess perceived trustworthiness of charts showing individual data (sina plots) versus aggregated data (bar charts). We further validated participants' semantic understanding of trustworthiness using established antecedents from prior work (Kelton et al., 2008). Across measures, charts displaying individual data were judged more trustworthy than charts of aggregated averages ($d = .66$): 70% of participants rated the individual-data charts as more trustworthy, compared to 17% for bar charts. This effect was not explained by prior coursework in statistics or psychology, nor by chart-interpretation performance.

These findings complicate claims that visual simplicity reliably increases perceived trustworthiness. Prior work has suggested that increased visual complexity can reduce trust (Elhamdadi et al., 2023) and that bar charts are more transparent than more elaborate visualizations (McKinley et al., 2025). In contrast, we find that aggregating data into summary statistics can reduce perceived transparency and trust relative to charts that reveal the full distribution.

Our thematic analysis suggests a conceptual account of this effect. When viewing aggregate-only charts, participants often reported insufficient information to assess trustworthiness, frequently noting missing details such as data sources or provenance. By contrast, visualizations that displayed individual data points were perceived as providing sufficient information about the underlying distribution, leading participants to infer positive intent on the part of the chart designer. In some cases, this inferred intent appeared to outweigh other potential cues of untrustworthiness, such as the absence of explicit source information.

A complementary explanation of these findings draws on a broader preference for concrete over abstract information. Research on risk and statistical reasoning shows that people reason more effectively with formats that preserve natural frequencies rather than abstract probabilities (Gigerenzer & Hoffrage, 1995), and that visually concrete displays such as icon arrays improve risk interpretation (Garcia-Retamero & Cokely, 2013). From this perspective, charts displaying individual data may better align with how people naturally reason about quantities, whereas aggregation obscures frequencies, reducing interpretability and trustworthiness. Future work can probe this distinction to examine whether trustworthiness judgments differ for charts that display the underlying distribution without concrete data (e.g., violin plots, bar charts with error bars).

Before concluding, three potential limitations are worth addressing. First, a common concern with data-showing visualizations is that displaying many individual observations could introduce perceptual issues such as overplotting, potentially reducing clarity (Hullman, 2020). However, this concern is mitigated by visualization techniques designed to preserve individual data while managing density. For example, the sina plots used here normalize local density to reduce overplotting (Sidiropoulos et al., 2018) and thus scale effectively to datasets with tens of thousands of observations (Kim et al., 2024). Second, our sina plot stimuli were generated from reported summary statistics under the general assumption of normality, yielding comparatively "well-behaved" data visualizations. Future work should explore how perceptions of a sina plot's trustworthiness differ when presenting "messy," real-world data. Third, our conclusions concern initial impressions of trustworthiness. Trust is a dynamic process (Kelton et al., 2008), and judgments may evolve with repeated exposure or deeper engagement with the data. Future work should examine how displaying underlying distributions shapes the development of trust in visualizations over time.

Conclusion We find that visualizations displaying individual data are perceived as more trustworthy than conventional charts of aggregated averages. These findings challenge the assumption that simplification necessarily promotes trust, suggesting instead that omitting underlying distributions may signal insufficient information. When trust is critical, showing the data, rather than abstracting it away, may be a simple yet powerful design choice.

References

Charmaz, K. (2006). Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. SAGE.

Chita-Tegmark, M., Law, T., Rabb, N., & Scheutz, M. (2021). Can you trust your trust measure? Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, 92–100.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. The American Statistician, 35(3), 124–129.

Cui, L., Wang, C.-Y., Wang, Y., Li, P., Kini, M., & Liu, Z. (2024). Bar tip limit error and characteristics of drawn data distributions on bar graphs. Proceedings of the Annual Meeting of the Cognitive Science Society, 46.

Davern, M., Bautista, R., Freese, J., Morgan, S., & Smith, T. (2022). Major declines in the public's confidence in science in the wake of the pandemic. AP NORC at the University of Chicago.

DiMascio, A., Weissman, M. M., Prusoff, B. A., Neu, C., Zwilling, M., & Klerman, G. L. (1979). Differential symptom reduction by drugs and psychotherapy in acute depression. Archives of General Psychiatry, 36(13), 1450–1456.

Eberhard, K. (2023). The effects of visualization on judgment and decision-making: A systematic literature review. Management Review Quarterly, 73(1), 167–214.

Elhamdadi, H., Stefkovics, A., Beyer, J., Moerth, E., Pfister, H., Bearfield, C. X., & Nobre, C. (2023). Vistrust: A Multidimensional Framework and Empirical Study of Trust in Data Visualizations. IEEE Transactions on Visualization and Computer Graphics, 30(1), 348–358. https://doi.org/10.1109/TVCG.2023.3326579

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. The Journal of Abnormal and Social Psychology, 58(2), 203.

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. Psychological Science in the Public Interest, 22(3), 110–161.

Franzblau, L. E., & Chung, K. C. (2012). Graphs, tables, and figures in scientific publications: The good, the bad, and how not to be the latter. The Journal of Hand Surgery, 37(3), 591–596.

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association, 28(2), 210–216. https://doi.org/10.1037/a0014474.

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

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. Psychological Review, 102(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Gray, P. O., & Bjorklund, D. F. (2018). Psychology (8th ed.). Worth Publishers.

Grison, S., & Gazzaniga, M. S. (2022). Psychology in Your Life. W. W. Norton & Company.

Hamilton, D. G., Hong, K., Fraser, H., Rowhani-Farid, A., Fidler, F., & Page, M. J. (2023). Prevalence and predictors of data and code sharing in the medical and health sciences: Systematic review with meta-analysis of individual participant data. BMJ (Clinical Research Ed.), 382.

Hullman, J. (2020). Why Authors Don't Visualize Uncertainty. IEEE Transactions on Visualization and Computer Graphics, 26(1), 130–139. https://doi.org/10.1109/TVCG.2019.2934287

Kalat, J. (2022). Introduction to Psychology (12th ed.). Cengage Learning.

Kelton, K., Fleischmann, K. R., & Wallace, W. A. (2008). Trust in digital information. Journal of the American Society for Information Science and Technology, 59(3), 363–374. https://doi.org/10.1002/asi.20722

Kerns, S. H., & Wilmer, J. B. (2021). Two graphs walk into a bar: Readout-based measurement reveals the Bar-Tip Limit error, a common, categorical misinterpretation of mean bar graphs. Journal of Vision, 21(12), 1–36.

Kim, H. A., Kaduthodil, J., Strong, R. W., Germine, L. T., Cohan, S., & Wilmer, J. B. (2024). Multiracial Reading the Mind in the Eyes Test (MRMET): An inclusive version of an influential measure. Behavior Research Methods, 56(6), 5900–5917.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics. Journal of the International Biometric Society, 159–174.

May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal time of day and the magnitude of age differences in memory. Psychological Science, 4(5), 326–330.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

McKinley, O., Pandey, S., & Ottley, A. (2025). Trustworthy by Design: The Viewer's Perspective on Trust in Data Visualization. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, 1–17. https://doi.org/10.1145/3706598.3713824

Myers, D. G., DeWall, C. N., & Gruber, J. (2024). Psychology (14th ed.). Worth Publishers.

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. International Journal of Qualitative Methods, 22, 16094069231205789. https://doi.org/10.1177/16094069231205789

NIH. (2025). NOT-OD-21-013: Final NIH Policy for Data Management and Sharing. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

NSF. (2022, March). Effective Practices for Making Research Data Discoverable and Citable (Data Sharing) | NSF - National Science Foundation. https://www.nsf.gov/funding/information/dcl-effective-practices-making-research-data-discoverable

Padilla, L., Fygenson, R., Castro, S. C., & Bertini, E. (2023). Multiple Forecast Visualizations (MFVs): Trade-offs in Trust and Performance in Multiple COVID-19

Forecast Visualizations. IEEE Transactions on Visualization and Computer Graphics, 29(1), 12–22. https://doi.org/10.1109/TVCG.2022.3209457

Pandey, S., McKinley, O. G., Crouser, R. J., & Ottley, A. (2023). Do You Trust What You See? Toward A Multidimensional Measure of Trust in Visualization (arXiv:2308.04727). arXiv. http://arxiv.org/abs/2308.04727

Rahman, Q., Wilson, G. D., & Abrahams, S. (2004). Biosocial factors, sexual orientation and neurocognitive functioning. Psychoneuroendocrinology, 29(7), 867–881.

Reimann, D. (2026). The bar-tip limit error in bar charts: Exploring its relationship to the within-the-bar bias. IEEE Transactions on Visualization and Computer Graphics, (99), 1–8.

Salmon, D. A., Dudley, M. Z., Glanz, J. M., & Omer, S. B. (2015). Vaccine hesitancy: Causes, consequences, and a call to action. Vaccine, 33, D66–D71.

Sawilowsky, S. S. (2009). New effect size rules of thumb. Journal of Modern Applied Statistical Methods, 8(2), 26.

Sidiropoulos, N., Sohi, S. H., Pedersen, T. L., Porse, B. T., Winther, O., Rapin, N., & Bagger, F. O. (2018). SinaPlot: An Enhanced Chart for Simple and Truthful Representation of Single Observations Over Multiple Classes. Journal of Computational and Graphical Statistics, 27(3), 673–676. https://doi.org/10.1080/10618600.2017.1366914

Song, H., Cho, A., Bearfield, C. X., & Stasko, J. (2025). Visualizing Trust: How Chart Embellishments Influence Perceptions of Credibility. IEEE Transactions on Visualization and Computer Graphics, 1–11. https://doi.org/10.1109/TVCG.2025.3634785

Turchioe, M. R., Myers, A., Isaac, S., Baik, D., Grossman, L. V., Ancker, J. S., & Creber, R. M. (2019). A systematic review of patient-facing visualizations of personal health data. Applied Clinical Informatics, 10(04), 751–770.

Tyson, A., & Kennedy, B. (2024). Public trust in scientists and views on their role in policymaking. Pew Research Center. https://www.pewresearch.org/science/2024/11/14/public-trust-in-scientists-and-views-on-their-role-in-policymaking/

Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. Global Challenges, 1(2), 1600008.

Wang, Y., Kerns, S., Brady, T., & Wilmer, J. (2025). The paradox of certainty: When graphed ensembles convey averages better than graphed averages. Proceedings of the Annual Meeting of the Cognitive Science Society, 47.

West, J. D., & Bergstrom, C. T. (2021). Misinformation in and about science. Proceedings of the National Academy of Sciences, 118(15), e1912444117.

Yang, F., Cai, M., Mortenson, C., Fakhari, H., Lokmanoglu, A. D., Hullman, J., Franconeri, S., Diakopoulos, N., Nisbet, E. C., & Kay, M. (2024). Swaying the Public? Impacts of Election Forecast Visualizations on Emotion, Trust, and Intention in the 2022 U.S. Midterms. IEEE Transactions on Visualization and Computer Graphics, 30(1), 23–33. https://doi.org/10.1109/TVCG.2023.3327356

Yang, F., Mortenson, C. R., Nisbet, E., Diakopoulos, N., & Kay, M. (2024). In Dice We Trust: Uncertainty Displays for Maintaining Trust in Election Forecasts Over Time. Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24, 1–24. https://doi.org/10.1145/3613904.3642371