Author Response to Round 1 Reviews

Introduction

The authors thank the reviewers for their thoughtful and constructive feedback on our submission. We have incorporated the vast majority of reviewer feedback in the form of direct edits and additions to the manuscript. We believe these edits strengthen the argument, better situate our work within the relevant literature, clarify several points, and overall improve the paper's contribution.

No analyses or results are changed with our revisions; our changes primarily consist of additional citations and clarifying text. We respond to each reviewers' stated weaknesses point-by-point below. All suggested citations from both reviewers have been incorporated within the text.

Responses to Reviewer 1

The weakest aspect of this paper is some of the representation-based conclusions, which I find difficult to generalize or tie to existing literature. I thought the charts chosen as stimuli contained design choices that may be more common in statistical evaluations, but are less common in public communication of data, and so don't align as strongly with the recruiting (and aren't typically used in visualization literacy work that I'm familiar with). For example, I was surprised to see the use of diverging stacked bar graphs, and line graphs that aren't anchored to time/date information. I wish I had a little clearer justification of these choices.

Author response: *This is a great point. For this smaller-scale study with only three types of charts, we wanted to demonstrate how the same data points could be represented in different ways and study how participant performance differed by only changing the analytic format/mapping of the data. We agree that line graphs that aren't anchored to time/date information are less recommended; however, we do see them in practice and our goal was to understand whether they provided any benefit over a stacked bar presenting the same information. The diverging stacked bar chart format is also used more commonly in communication of survey results, which is a common use case for many of our research collaborators and colleagues. We have added text in the experimental design section of the paper to better clarify these choices for readers.*

While the authors draw connections with Garcia-Retamero's graph literacy work, for a visualization academic audience, I would suggest also drawing connections with visualization literacy work and tests that are more frequently discussed within the visualization community. Right now, the paper seems slightly detached from the primary vis discourse in that space (but no less valuable). I'll put some suggestions in the appropriate place later in this form.

Author response: *Thank you for this constructive feedback. We have incorporated the reviewer's suggested references in the paper's introduction to better connect to existing VIS discourse and explain how our work both complements and differs from prior visualization tests.*

Responses to Reviewer 2

Weakness:

1. Clarify Stimuli Design The study only includes three types of charts—stacked bar, diverging stacked bar, and line charts—which may not fully capture the diversity of data visualization formats commonly encountered in real-world contexts. The paper can justify further why they chose to adopt these three

visualizations from all the charts, and discuss limitations related to the generalizability of the findings to other visualization types. Also, the three charts relies on a single dataset related to living arrangements among older adults, which may not represent the range of contexts where data visualizations are applied. This topic choice could also introduce bias, as it might be more familiar to certain demographic groups. I recommend incorporating multiple datasets to reduce topic-related biases and enhance the robustness of the findings.

Author response: *While we are not able to incorporate multiple topics/datasets within the scope of this paper, we completely agree that this would be beneficial and are in fact in the process of designing a larger study which incorporates varying data topics to control for topic-related biases. Our choice of keeping the dataset consistent was intentional given our limitations on the number of questions we could ask; we intended to see how simply varying the structure impacts understanding. We feel that our existing conclusions section appropriately states the limitations, as we emphasize that it is a limited scope and re-state that this only applies across three chart types. However, we have slightly edited the abstract language to be more clear up-front that this paper only uses three chart formats.*

2. Clarify Data Visualization Skills In this work, the paper aim to evaluate the general public's understanding of data visualization. However, understanding data visualization is a complex cognitive process, and the authors do not clearly define what they mean by this term. It could refer to basic skills, such as identifying chart types or visual elements, or extend to more advanced competencies, such as interpreting data, making inferences, and drawing real-world conclusions. The lack of a clear definition and scope for what constitutes "understanding" makes it difficult for readers to contextualize the study's contributions and assess its relevance within the broader field of data visualization research.

Author response: *This is an excellent point of clarification. While we do describe the instrument and items and how they test a user's abilities in the experimental design, we agree a better clarification should be made throughout. We have added the following details within the introduction to better define our definition of understanding: "Our experiment asks users to estimate presented values within the presented context, testing their basic understanding of labels, legends, and visual mapping to a value, as well as asks participants to assess whether provided real-world conclusions are supported by the visualized data. With this structure, we test both a user's basic understanding of the chart elements and their understanding of how the mapped data relate to one another and correspond to their real-world meaning."*

3. Justify the Evaluation Method Building on the previously mentioned points, participants are asked to answer two types of questions: identifying and estimating values. However, the paper does not provide a rationale for why these tasks are sufficient to represent how people understand data visualization. Additionally, there are established data visualization literacy assessment tools in the field, such as VLAT. The paper does not explain why they chose to design a new assessment instead of utilizing existing frameworks, nor do they clarify how their approach differs. These comparisons and justifications are crucial for validating the research design. While it seems the paper intended to create a setup that reflects real-world contexts—an emphasis throughout the paper—the tasks appear to focus more on basic chart-reading skills rather than evaluating participants' ability to draw "real-world conclusions." Clarifying how these questions capture "real-world conclusion generation" would strengthen the study's contribution.

Author response: *Our work is not intended to be an overall measure of visualization literacy but rather measure how understanding of the same quantitative information differs based on design choices, and further measure how that understanding differs across groups. We agree a more comprehensive test framework such as VLAT is beneficial. Our items do also cover several levels of task difficulty covered in the VLAT including retrieving values, assessing trends, and making comparisons. We have better incorporated the VLAT framework in our introduction to situate how and why our experimental design differs from frameworks such as VLAT. We have also, in response to the prior point, enhanced our definition of "understanding" in the introduction. We hope both these edits together address this concern.*

4. Missing Important Related Work The paper misses several important related works that could strengthen its argument. For instance, while the paper claims that most prior studies are "performed with small groups of undergraduate students," research by Börner et al. demonstrated the use of data visualizations with general museum visitors, and work by Peck et al. evaluated the general public's understanding of data visualizations in a local market setting. Both studies are highly relevant as they assess data visualization comprehension among diverse, real-world audiences outside of controlled lab environments. Including these works would provide valuable context and help position this study within the broader landscape of data visualization research conducted in the wild.

Author response: *There is not evidence to support that the Börner et al. paper represents a diverse audience as the study only reports perceived sex of participants and whether participants were youth or adults. That said, we agree that both mentioned studies provide an important view of oft-understudied groups in the literature and have added a sentence in the paper's introduction to call out these prior contributions and better situate our work among them.*

Minor:

1. Consider accounting for rior data visualization experience: The study does not account for participants' prior exposure to or familiarity with data visualizations, which could significantly influence their performance and interpretation skills.

Author response: *This is a good point. While we are unable to incorporate this into this paper post-data collection, we believe that some of our findings suggest a strong correlation between higher levels of education and exposure to or familiarity with data visualizations given the higher performance among higher education groups. This is also a reason to study these differences further in the future – if prior data visualization experience influences performance, that tells us something about how to improve visualization literacy among adults as a whole.*

2. Consider supplementing the results with Qualitative Data: The study relies solely on quantitative measures without incorporating qualitative data (e.g., think-aloud protocols or post-task interviews) that could provide more information on participants' reasoning processes and potential misconceptions.

Author response: *This is out of scope for this paper as this was not part of our data collection for these tasks. However, we also believe this is an important angle to study – in ongoing related work we have asked participants to provide open-ended text descriptions of what a chart tells them. A research manuscript on those findings is currently in preparation.*

3. Consider showing multiple visualizations per study: Participants were shown to only one chart type in the experiment. This between-subjects design limits the ability to compare individual differences in performance across different visualization formats, which could have provided richer within-subject insights.

Author response: *We agree this would be beneficial for larger validation of performance comparison. We were constrained by being able to ask a limited number of questions in a given round of the survey, but are currently preparing to field a larger survey in which we show participants multiple types of charts and can account for subject-level performance across items.*