Second response to reviewers of JCGS-22-045

Eye Fitting Straight Lines in the Modern Era

Emily A. Robinson, Reka Howard, and Susan VanderPlas

2022

Reviewer(s)' Comments to Author

All responses to reviewers' comments are written in blue.

Reviewer: 1

I sincerely appreciate the revision made by the authors. They addressed my main concerns and they better clarified the added value of their work in the context of previous and very similar studies. Given the scope of the journal, I think that the article is now suitable for publication. I just have two final suggestions for the authors:

1) I invite them to carefully re-read the entire manuscript, since a few typos can still be found.

- We have carefully reviewed the manuscript for typos.

2) My previous concern about the lack of clear recognition of previous similar work on the topic (which has been very well addressed by the authors, concerning the introduction, the methods and the results sections) seems somehow still applicable to the two last sections. For example, the possible role of training in the task is only discussed in terms of "movements". What about the training itself in performing the task? (i.e., becoming better and better at performing mental regressions? For example, Cui et al., 2018 precisely studied the question of training in correlation estimations in scatterplots). In the "future work" section, some of the suggestions made by the authors have indeed already been tested: concerning the resistance to outliers, quite a few studies already investigated human ability to perform regressions over scatterplots with outliers (as an example, but other studies can be found: Bobko & Karren, 1979 and Correll & Heer, 2017). Lastly, authors suggest to conduct studies on human ability to extrapolate data from trends; again, studies exist on the topic (as an example, Ciccione et al., 2022, who asked participants precisely to extrapolate new datapoints from exponential and linear

trends plotted on log and linear scales). I am aware that the studies I mention come from the literature in cognitive psychology and/or psychophysics, rather than from the literature in data visualization, statistics or interactive visualization. However, I think that the authors, given the research work they conducted and presented, are ideally placed to fill the gap between the fields of cognitive psychology and data visualization. Therefore, they could consider adding those and/or other references to their last sections in order to better acknowledge the work conducted in fields other than data visualization.

- We included and acknowledged the work suggested from cognitive psychology. Thank you for bringing close attention to this field of work!

Good job and good luck with your future work !

Reviewer: 2

The authors present a study where human participants had to fit straight lines through points in a scatterplot. On average, the participants' fitted lines closely matched a principal axis line and not the least squares regression line.

The authors addressed my previous comments very well. Particular thanks for the latexdiff version that made it easy to review the revision!

Only a few very minor corrections remain to be done.

- p.7, l.23-24: use past tense here: "participants from Twitter and Reddit pages related to data visualization voluntarily completed the study and likely HAD an interest in fields related to statistics and WANTED to help advance research in graphics."
  - Changed from present to past tense: have -> had and want -> wanted.
- p.7, l.37: use the full URL (not every pdf viewer / web browser correctly translates an incomplete URL): https://emily-robinson.shinyapps.io/you-draw-it-validation-applet/ [same for the URLs on p.17]
  - Double checked and provided full URLs. We also provided the publishers with a .zip file of supplementary material.
- p.8, l.46: after eq (1), continue with "…, where b_1 represents … and y_x-bar …", i.e., verbally explain earlier in the text what these mean [your explanation comes too late later on]
  - This was a very helpful suggestion, we clarified under equation 1 with "where β_1 represents the slope of the trend and $y_{\bar{x}}$ represents the y value at the mean of the generated x values." and later stated "The point $(x, y_{\bar{x}})$ was used in the point-slope equation of a line."

- p.8, l.27 & p. 11, l.54 & p.14, l.5: speak of "R" package as many authors these days make joint use of R and Python (and others) for their work (so always clarify which "package" from which software environment is used) [check elsewhere]}
  - We clarified the use of R packages after each use of a package from R statistical software.
- p.14, l.11: "REML is used" is not clear - is this another R function from the mgcv package or the method in general? - be more specific}
  - We clarified "...; the 'bam' function implements restricted maximum likelihood (REML) to estimate parameters and smoothing splines."
- p.14, l.19: place e_{ijk,fit} into ...
  - Added math typeset around e_ijk,fit for $e_{ijk,fit}$.
- p.14, l.30: "eijk,fit is the same as equation" likely should be "eijk,fit is the same as IN equation **(2)**"
  - Added 'in' and '(2)' to defining $e_{ijk,fit}$ after equation (3).
- p.19, l.54: remove the extra author from "Kosslyn, S. M. & Kosslyn, S. M. (2006)" - this is just Kosslyn
  - Removed extra author in references.
- p.20, l.45: for Rousset et al., add "ACM, New York, NY" as publisher for these proceedings - see the citation at https://dl.acm.org/doi/10.1145/2820619.2820633
  - Exported bibtex citation from the reccommended link.

Thank you for your valuable feedback, suggestions, and review of our manuscript.

Associate Editor's Comments to Author

(There are no comments.)