Truncating the Y-Axis: Threat or Menace?

Supplementary Material

1. INTRODUCTION

This document is a supplement to the paper "Truncating the Y-Axis: Threat or Menace?" and includes additional analyses and figures excluded from the main paper for reasons of space. Consult our OSF project at https://osf.io/gz98h/ for raw data tables, analysis scripts, and more reproducibility information.

2. FULL ANOVA TABLES

Our main experimental measure was a rating of the perceived severity of a trend in a chart purporting to show data over time. The main experimental manipulation was the starting location of the y-axis. See the main paper for additional details.

2.1. Experiment One.

Repeated Measures ANOVA Table (Perceived Severity)				
Factor	**df**	**dn**	**F**	**p**
Truncation Level	2	76	89	0.01E-20*
Visualization Type	1	38	0.46	0.5
Framing	1	38	7.4	0.01*
Data Size	1	38	5.5	0.02*
Truncation×Vis	2	76	0.67	0.52
Truncation×Framing	2	76	0.59	0.55
Vis×Framing	1	38	4.9	0.03*
Truncation×Size	2	76	0.24	0.79
Vis×Size	1	38	1.4	0.24
Framing×Size	1	38	2.0	0.16
Trunc×Vis×Fram	2	76	1.3	0.28
Trunc×Vis×Size	2	76	1.1	0.34
Trunc×Fram×Size	2	76	1.6	0.21
Vis×Fram×Size	1	38	2.3	0.14
Trunc×Vis×Fram×Size	2	76	2.0	0.15

2.2. Experiment Two.

Repeated Measures ANOVA Table (Perceived Severity)				
Factor	**df**	**dn**	**F**	**p**
Truncation Level	1	30	39	0.06E-7*
Visualization Type	2	60	3.1	0.05
Data Size	1	30	3.1	0.09
Truncation×Vis	2	60	2.5	0.09
Truncation×Size	1	30	0.57	0.46
Vis×Size	2	60	1.7	0.18
Trunc×Vis×Size	2	60	0.58	0.56

2.3. **Experiment Three.** In Experiment Three, participants had to answer what the difference was between the last and the first data values. We calculated two forms of error: $E_{slope} = |(Q_{last} - Q_{first}) - slope_{actual}|$. and $E_{magnitude} = |(Q_{first} - X_{first}) - (Q_{last} - X_{last})|/2$.

Repeated Measures ANOVA Table (Slope Error)				
Factor	**df**	**dn**	**F**	**p**
Truncation Level	1	20	0.002	0.96
Visualization Type	2	40	0.80	0.45
Data Size	1	20	14	0.001*
Truncation×Vis	2	40	1.4	0.26
Truncation×Size	1	20	0.67	0.42
Vis×Size	2	40	1.6	0.21
Trunc×Vis×Size	2	40	0.95	0.40

Repeated Measures ANOVA Table (Magnitude Error)				
Factor	**df**	**dn**	**F**	**p**
Truncation Level	1	20	8.3	0.009*
Visualization Type	2	40	0.17	0.85
Data Size	1	20	1.7	0.21
Truncation×Vis	2	40	1.4	0.27
Truncation×Size	1	20	0.25	0.63
Vis×Size	2	40	0.05	0.95
Trunc×Vis×Size	2	40	0.29	0.75

We also had the same rating task as the prior two experiments.

Repeated Measures ANOVA Table (Perceived Severity)				
Factor	**df**	**dn**	**F**	**p**
Truncation Level	1	20	11	0.003*
Visualization Type	2	40	0.22	0.80
Data Size	1	20	4.0	0.06
Truncation×Vis	2	40	0.51	0.60
Truncation×Size	1	20	0.11	0.75
Vis×Size	2	40	1.2	0.31
Trunc×Vis×Size	2	40	0.52	0.59

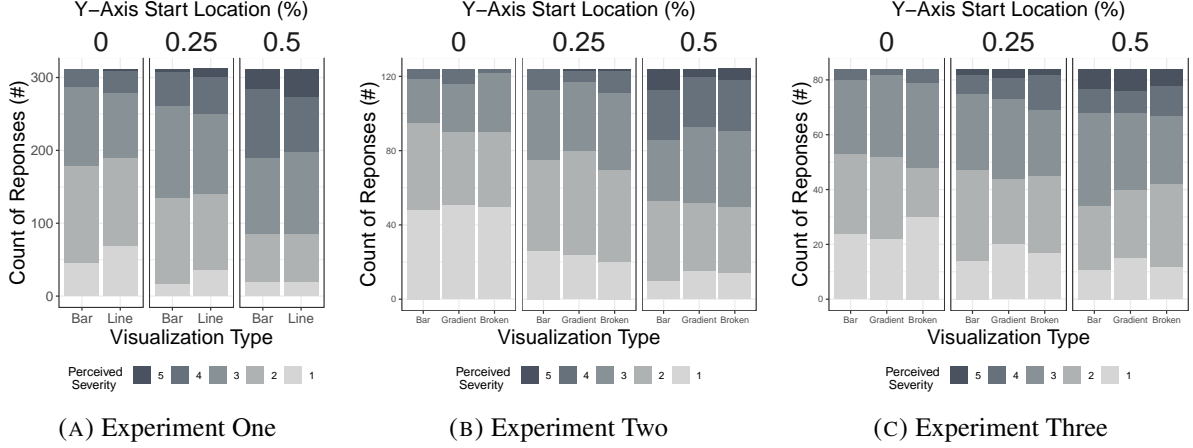(A) Experiment One   (B) Experiment Two   (C) Experiment Three

FIGURE 1. Stacked bar charts of all raw responses from all three experiments. The darker the value, the more severe the effect was judged as being. Note that the actual set effect sizes were numerically identical across all conditions, so if the starting location of the y-axis had no impact, each bar would have a similar proportion of responses.
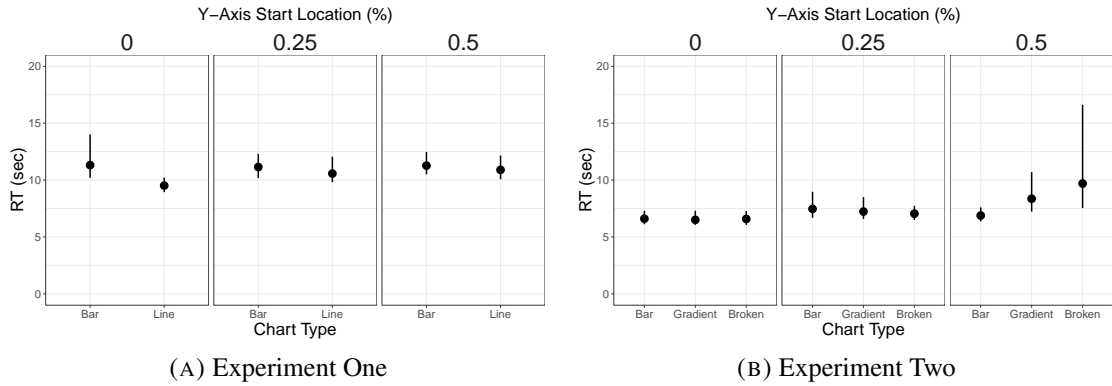

(A) Experiment One   (B) Experiment Two

FIGURE 2. Response time for Experiments One and Two. Error bars are 95% bias-corrected bootstrapped confidence intervals.

3. ADDITIONAL FIGURES & ANALYSES

3.1. **Raw Responses.** The main task in all three experiments was answering a question about the perceived severity of a trend in a data, given the chart used to show the data (and the starting point of the y-axis of that chart). This question was answered as a 5-point rating scale. While we mainly report on the central tendencies of this rating scale in the main paper, we recognize that rating scales are often better conceptualized as ordinal rather than purely numerical data. We include the pattern of raw responses here.

3.2. **Response Time.** We measured response time from when the participant clicked the "Ready" button, to when they finalized their choice with the "Confirm" button. Error bars are 95% bias-corrected bootstrapped confidence intervals. Note that we did not instruct participants to answer as quickly as possible (although on crowdworking platforms there is an inherent pressure to complete tasks as quickly as possible), nor did we alert them that we were recording this timing information, or forbid participants from pausing the task to complete other activities. As such, the values are highly variable.
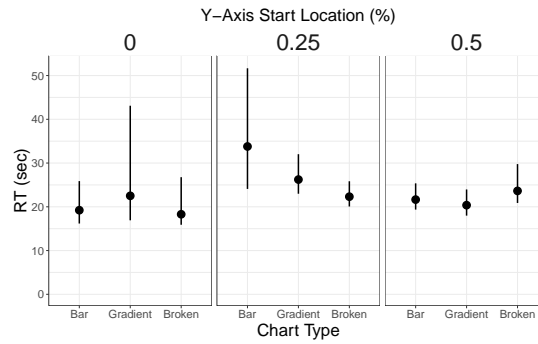
FIGURE 3. Response time for Experiment Three. Note the differing y-axis between this chart and Figure 2: in Experiment Three there was a third, much more involved numerical task, so RTs are not comparable here to the other experiments.

Repeated Measures ANOVA Table				
Factor	**df**	**dn**	**F**	**p**
Truncation Level	2	76	1.4	0.25
Visualization Type	1	38	4.0	0.05
Framing	1	38	17	0.0002*
Data Size	1	38	3.5	0.07
Truncation×Vis	2	76	1.15	0.32
Truncation×Framing	2	76	0.46	0.62
Vis×Framing	1	38	0.003	0.96
Truncation×Size	2	76	0.02	0.97
Vis×Size	1	38	0.60	0.44
Framing×Size	1	38	1.1	0.30
Trunc×Vis×Fram	2	76	0.53	0.59
Trunc×Vis×Size	2	76	0.76	0.47
Trunc×Fram×Size	2	76	1.2	0.32
Vis×Fram×Size	1	38	0.18	0.67
Trunc×Vis×Fram×Size	2	76	1.1	0.33

3.3. **Strategy Information.** After the main task, participants were asked to report on their strategy, and if they noticed anything about the graphs they were looking at. We were specifically looking for if the participants reported noticing that the y-axes of the charts had different start points. For each experiment, two coders independently gave a binary code to the free text responses if they judged that the participants' responses gave such an indication. The coders then met to assess and rectify mismatches. In this section we report on the impact of these codes on the size of the impact of y-axis truncation on subjective impressions of trend.
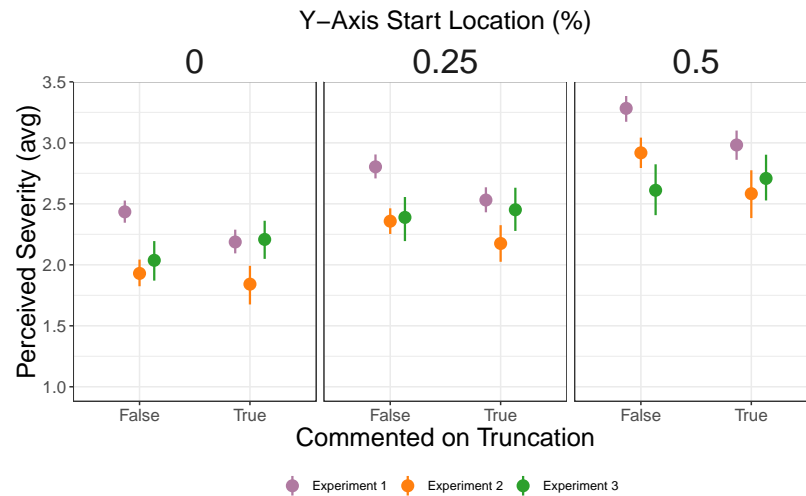
FIGURE 4. Impact of explicit mention of the y-axis manipulation on perceived severity of trend. In all cases, across all experiments there was a consistent impact of the y-axis start location, even for "saavy" participants. In Experiments One and Two there was a slight decrease in perceived severity for those who commented on y-axis truncation. But there was a slight increase in Experiment Three. (Although note that in Experiment Three participants had to answer numerical questions about the data, so it's possible that the truncation was so "obvious," as it was central to the task, that it was not diagnostic to mine responses for its presence or absence. Error bars are 95% bias-corrected bootstrapped confidence intervals.