Magnitude Judgements Are Influenced by the Relative Positions of Data Points Within Axis Limits

DOI:
10.1109/TVCG.2024.3364069

Document Version
Accepted author manuscript

Citation for published version (APA):
Bradley, D., Strain, G., Jay, C., & Stewart, A. J. (2024). Magnitude Judgements Are Influenced by the Relative Positions of Data Points Within Axis Limits. *IEEE Transactions on Visualization and Computer Graphics*, 1-8. https://doi.org/10.1109/TVCG.2024.3364069

Published in:
IEEE Transactions on Visualization and Computer Graphics

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3364069

Magnitude Judgements Are Influenced by the Relative Positions of Data Points Within Axis Limits

Duncan Bradley, Gabriel Strain, Caroline Jay, and Andrew J. Stewart

Abstract—When visualising data, chart designers have the freedom to choose the upper and lower limits of numerical axes. Axis limits can determine the physical characteristics of plotted values, such as the physical position of data points in dot plots. In two experiments (total N=300), we demonstrate that axis limits affect viewers' interpretations of the magnitudes of plotted values. Participants did not simply associate values presented at higher vertical positions with greater magnitudes. Instead, participants considered the relative positions of data points within the axis limits. Data points were considered to represent larger values when they were closer to the end of the axis associated with greater values, even when they were presented at the *bottom* of a chart. This provides further evidence of framing effects in the display of data, and offers insight into the cognitive mechanisms involved in assessing magnitude in data visualisations.

Index Terms—Magnitude, Axis Manipulation, Cognition, Bias, Framing Effects

1 Introduction

Context is crucial for effectively judging the magnitude of numbers. A 10% probability is twice as great as a 5% probability, but in the absence of context, it is unclear whether this value should be considered large or small. When referring to the chance of losing one's job, a 10% probability may be considered large, but when referring to the chance of losing a sports bet, a 10% probability may be considered small.

Contextual cues may influence interpretation of magnitude in data visualisations. One such cue is a chart's axis limits, which can serve as a frame of reference for assessing whether a data point represents a large or small number. Figure 1 (a reproduction of a similar bar chart from the New York Times), which plots over time the number of Black members of the U.S. senate, provides a striking illustration. Unusually, the y-axis does not terminate just above the highest plotted value. Instead, the y-axis extends all the way to the maximum possible number of senators: 100. As a result, bars representing Black senators are confined to the very bottom, visible just above the x-axis, and a significant expanse of blank space looms above them. This framing situates plotted data points in their numerical context, thus conveying a small magnitude.

It is unclear exactly how a viewer's inferences about magnitude might be influenced by axis limits. Different axis limits present data points at different positions, so one possible explanation is that viewers interpret the magnitude of data points at higher positions as 'high' and those at lower positions as 'low'. Alternatively, axis limits may provide context: magnitudes may be judged as small when the potential for larger values is clearly displayed. The present pair of experiments demonstrates the influence of axis limits on viewers' interpretations and explores which of these two accounts best explains how axis limits contribute to the communication of magnitude.

- *Duncan Bradley is with the Division of Psychology, Communication, and Human Neuroscience, The University of Manchester. E-mail: duncan.bradley@manchester.ac.uk*
- *Gabriel Strain is with the Department of Computer Science, The University of Manchester. E-mail: gabriel.strain@manchester.ac.uk*
- *Caroline Jay is with the Department of Computer Science, The University of Manchester. E-mail: caroline.jay@manchester.ac.uk*
- *Andrew J. Stewart is with the Department of Computer Science, The University of Manchester. E-mail: andrew.stewart@manchester.ac.uk*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

Overview In two experiments, we manipulated the axis limits surrounding plotted data. This affected the physical positions of data points, but also the *relative* positions of data points within axis limits: their proximity to the highest and lowest axis values. Likert scale ratings of the magnitude of data points were higher when data points were positioned close to the end of the axis which was associated with higher numbers. By employing charts with conventional and inverted y-axis orientations to distinguish between possible explanations, we reveal that magnitude judgements are influenced by the *relative* positions of data points within axis limits.

Data, analysis code, experimental scripts, materials, and interactive versions of the experiments are available at https://osf.io/3epm2/. We also provide all necessary resources for running a Docker container, within which the computational environment used for analysis is recreated, meaning a fully-reproducible version of this paper can be generated.

Contributions

1. We provide empirical evidence demonstrating that axis limits influence interpretations of absolute magnitude in dot plot visualisations.
2. We explore the underlying cognitive mechanism by manipulating axis orientations. This reveals that this effect is primarily driven by the *relative position* of data points within axis limits.
3. We present the communication of absolute magnitude as an additional consideration for data visualisation designers.

2 Related Work

2.1 Effects of Axis Limits on Interpretations of Relative Differences

Several studies have explored the role of axis limits in data visualisation. Research has typically focused on how axis limits can alter impressions of the *difference between* presented values. For example, when axis ranges are expanded to create blank space around a cluster of data points, correlation between those points is judged as stronger [7]. Participants also rate the differences between values in bar charts as greater when the vertical gap between bars is larger due to a truncated y-axis [21].

Correll et al.'s [8] experiments found that greater y-axis truncation resulted in higher effect-size judgements in both line charts and bar charts. They found no reduction in effect size judgements when truncation was communicated using graphical techniques (e.g., axis breaks and gradients). Truncation effects also persisted even when participants estimated the values of specific data points. This suggests the bias is

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3364069

driven by initial impressions, rather than by a misinterpretation of the values portrayed by graphical markings. This bias resembles a *framing effect*, wherein the same information is interpreted differently according to the perspective induced by aspects of its presentation [28]. The unavoidable consequence, Correll et al. suggest, is that designers' choices will influence viewers' interpretations whether axes are truncated or not.

Choosing appropriate axis limits involves a trade-off between participants' bias (reliance on the visual appearance of differences) and their sensitivity (capacity to identify actual differences) [31]. Just as a highly truncated y-axis can exaggerate trivial differences between values, an axis spanning the entire range of possible values can conceal important differences. Based on participants' judgements of effect size, Witt [31] found that bias was reduced and sensitivity increased when using an axis range of approximately 1.5 standard deviations of the plotted data, compared to axes which spanned only the range of the data, or the full range of possible values. This provides further evidence of a powerful association between the appearance of data, when plotted, and subjective interpretations of differences between data points.

Further evidence of truncation effects, provided by Yang et al. [34] improves on the design of previous studies which employed only a few observations per condition [21] or small sample sizes [31]. Participants' ratings of the difference between two bars consistently provided evidence of the exaggerating effects of y-axis truncation. Yang et al. [34] noted that increasing awareness of this bias does not eliminate this effect, which may function like an anchoring effect, in which numerical judgements are influenced by reference points [27]. Another potential explanation discussed by Yang et al. [34] draws upon Grice's co-operative principle [12]. According to this account of effective communication, speakers are assumed to be in cooperation, and so will communicate in a manner that is informative, truthful, relevant, and straightforward. Analogously, a viewer will assume that a numerical difference in a chart must be genuinely large if it appears large, else it would not be presented that way. Effective visualisations should be designed so a viewer's instinctive characterisation of the data corresponds closely to their interpretation following a more detailed inspection [34].

The above research consistently demonstrates that the magnitude of *the difference between* values is interpreted differently depending on the axis limits employed. This helps explain how viewers compare values within a graph. However, little attention has been paid to a different type of judgement: how are interpretations of the *values themselves* (not simply the differences between them) affected by axis limits? The present investigation explores this question by analysing responses to pairs of charts which display the same data using different axis limits. In the following section, we discuss existing work on judgements of *absolute* magnitude in data visualisations, and discuss how our work expands upon these prior investigations.

2.2 Effects of Axis Limits on Interpretations of Absolute Magnitude

Empirical evidence demonstrates that judgement of a value's magnitude can depend on its relationship to a grand total or to surrounding values. This can influence interpretation of verbal approximations and numerical values. For example, participants instructed to take 'a few' marbles picked up more when the total number available was larger [3] and rated satisfaction with the same salary as higher when it appeared in the upper end of a range, compared to the lower end [4]. As well as context, vertical position also plays a role in magnitude judgements. For example, children appear to intuitively understand the relationship between height and value [10]. Both the physical world, and language (e.g., spatial metaphors), provide countless examples where 'higher' is associated with 'more', and 'lower' with 'less', and this principle has been adopted as a convention in data visualisation [29].

In charts, inversions of the typical mapping between magnitude and vertical position charts can lead to misinterpretations [19, 21, 33]. Furthermore, when a company's financial performance was displayed entirely in the bottom fifth of a line chart, the company was perceived as less successful, compared to when the axis did not extend above the maximum value [26]. Sandman et al. [25] investigated assessments of

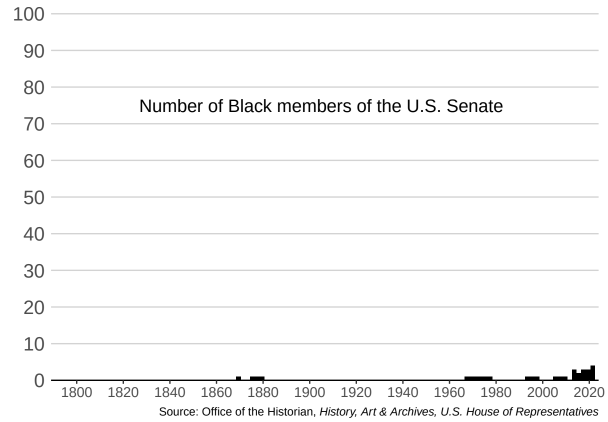

Fig. 1. A reproduction of a bar chart from the New York Times. The y-axis limit is defined by the largest possible value, rather than the largest observed value, thus the magnitude of plotted values appears particularly small.

magnitude in risk ladders, where greater risks are presented at higher positions on a vertical scale. Participants rated asbestos exposure as a greater threat when it was plotted at a higher position, compared to a lower position.

The above findings can be regarded as preliminary evidence that changing axis limits may affect appraisals of the magnitude of data points. However, the evidence is not substantial. Taylor and Anderson [26] did not disclose how judgements were elicited, or provide details of their sample size. Sandman et al. [25] only explored responses to one specific risk (asbestos), and each participant only took part in a single trial. The perceived threat measure was a composite of several separate ratings, preventing diagnosis of whether manipulations affected interpretations of the plotted information in particular, or just related concepts. Further, both studies introduced a confounding variable by adjusting the difference between the minimum and maximum y-axis values across conditions. Stronger evidence is required regarding how axis limits may bias inferences about absolute magnitude, and the cognitive mechanisms involved in generating these inferences.

2.3 Judgements of Event Outcomes

In the present study, participants viewed charts showing fictitious data on the chance of particular events occurring. This provided participants with a purpose; evaluating information about event outcomes is a more meaningful task than assessing how 'large' an abstract value is. Each value was represented using a single dot on a percentage probability scale. Our use of dot plots for conveying percentages was motivated by their simplicity and use of a single encoding channel (position), thus avoiding confounding variables from other encoding channels.

Presenting data about events with negative consequences warranted consideration of the cognitive processing of this information. These events are composed of two core components: 1) chance of occurrence and 2) severity of outcomes. Assessments of chance and severity are not necessarily independent. Events are perceived as more likely when they are described as having more severe consequences [13, 14]. In a similar manner, events are associated with more substantial consequences when they are described as more likely [17].

One account suggests that perceptions of probability and outcome magnitude are related because they are both assumed to reflect the potency of the event's cause. This is known as the probability-outcome correspondence principle [16]. According to this account, probabilities can occasionally provide meaningful indications of severity (e.g., rainfall), but it is inappropriate to apply this perspective to all situations (e.g., volcanic eruptions). Therefore, even though charts in the present study only display the *chance* of events occurring, assessments of the *severity* of events' consequences may also differ between conditions. Collecting separate judgements of chance and severity of consequences

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3364069

provides a clearer picture of how the manipulation affects distinct aspects of participants' representations. Our use of Likert scales forces participants to choose from a limited number of options, thus helping capture their *gist*-level interpretations of information, rather than *verbatim* representations [24].

2.4 Data Visualisation Literacy

When faced with charts that violate graphical conventions by using atypical scales, individuals with low data visualisation literacy are more likely to draw on the physical positions of data points when making inferences about their magnitudes [19, 20]. We administered Garcia-Retamero et al.'s [9] subjective graph literacy measure to determine whether responses to our manipulation of axis limits were associated with data visualisation literacy.

3 Experiments

We conducted two experiments manipulating y-axis limits in visualisations of fictitious data. This manipulation altered the physical positions of data points in a chart, but crucially the numerical values themselves remained the same.

Experiment 1 sought to establish whether y-axis limits affected magnitude judgements. To provide context for participants, text accompanying the charts outlined (fictitious) scenarios involving a specific negative outcome (e.g., loss on financial investment, delayed flights, etc.). Three plotted data points in each chart represented the chance of the negative outcome occurring (%) for three instances associated with the scenario (e.g., three investment opportunities, three airlines, etc.). Participants rated the chance of events occurring (indicating their interpretation of the magnitude of plotted data points) and also the severity of consequences.

Experiment 2 used the same y-axis manipulation as Experiment 1, but in charts with inverted y-axis orientations, where data points at lower physical positions represented greater values. This allowed us to investigate whether magnitude judgements were driven by the physical positions of data points, or their relative positions within the context of the axis limits. Importantly, the use of inverted charts should not be considered an endorsement (see issues above). However, they serve to distinguish between two possible explanations, since they reverse the typical associations between physical position and magnitude.

Ethical approval was granted by The University of Manchester's Division of Neuroscience & Experimental Psychology Ethics Committee (Experiment 1: Ref 2021-11115-18258; Experiment 2: Ref. 2021-11115-20745). Data, analysis code, experimental scripts, materials, and interactive versions of the experiments are available at https://osf.io/3epm2/.

3.1 Experiment 1

3.1.1 Method

Materials

Datasets For each dataset, we generated three values from a normal distribution. Population means were specified manually in order to represent plausible values for the probability of the event occurring (28% - 72%). All datasets had a population standard deviation of 0.5. The same dataset was employed for both of the experimental conditions associated with a given event scenario.

Charts Datasets were displayed using dot plots. In experimental trials (n = 40), upper and lower axis limits were manipulated such that data points either appeared in the top third of the chart (high physical position: Figure 2, left) or in the bottom third (low physical position: Figure 2, right).

The y-axis range in each chart was 10 percentage points. Horizontal gridlines appeared at one-unit increments. The horizontal gridlines 1.5 units from the extremes were labelled with numerical values.

Filler trials (n = 15) and attention check trials (n = 5, two questions per trial) presented data points in the middle third of the chart. Filler trials employed this additional variation to prevent participants from identifying the purpose of the study.

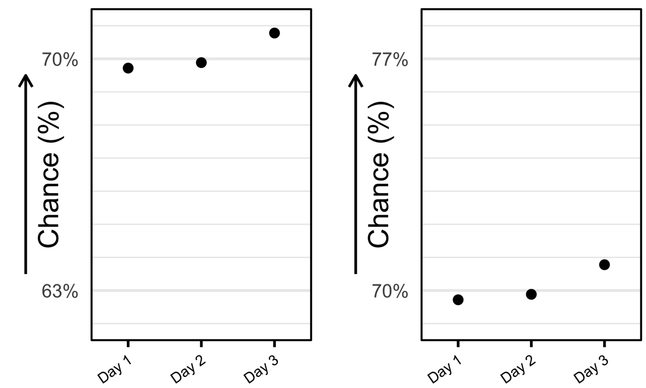

Fig. 2. Example charts, taken from Experiment 1. The *high physical position* condition (left) presents data points near the top of the chart; the *low physical position* condition (right) presents the same data points near the bottom of the chart.

Procedure The experiment was programmed in PsychoPy (version 2021.1.4, [22]) and hosted on pavlovia.org. Participants were instructed to complete the experiment on a desktop computer or laptop, not a tablet or mobile phone. Instructions explained that their task involved assessing the chance and severity of negative outcomes in various scenarios involving risks and noted that some scenarios might appear similar to other scenarios. Participants were asked to complete the task as quickly and accurately as possible. Two practice trials preceded the experiment proper.

An example of a single trial is shown in Figure 3. Participants provided two responses in each trial: a rating of the chance of the negative event occurring; and a rating of the severity of the consequences should that negative event occur. Both 7-point Likert scales had anchors at their extremes: *'Very unlikely'* and *'Very likely'* for the 'chance' scale, and *'Very mild'* to *'Very severe'* for the 'severity' scale. All other response categories were unlabelled. Accompanying text specified that answers should be given in response to the plotted data (e.g., *"If you camp on one of these days..."*). The term 'chance' was used instead of 'probability' to avoid confusion with the standard 0-1 scale for probabilities, and to reflect casual usage.

Participants could change their responses as many times as they wished before proceeding to the next trial, but could not return to previous trials. In attention check trials, participants were instructed not to attend to the chart, and instead to provide specified responses on each of the two Likert scales.

Before exiting the experiment, participants were informed that all data presented were fictitious and guidance was provided in case of distress.

Design We employed a repeated-measures, within-participants design. Participants encountered scenarios from experimental trials twice: once with data presented at a high physical position and once with data presented at a low physical position.

Materials were divided into two lists to minimise the likelihood that different versions of the same scenario appeared in close succession. One list contained half of the high-condition items and half of the low-condition items for the experimental scenarios. The other list contained the alternate versions of each of these experimental scenarios. Fillers and attention check questions were split between the two lists, and did not appear more than once. The order of the two lists was counterbalanced across participants. Within each list, scenarios were presented in a random order.

Participants The experiments were advertised on Prolific.co, a platform for recruiting participants for online studies. Normal or

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3364069

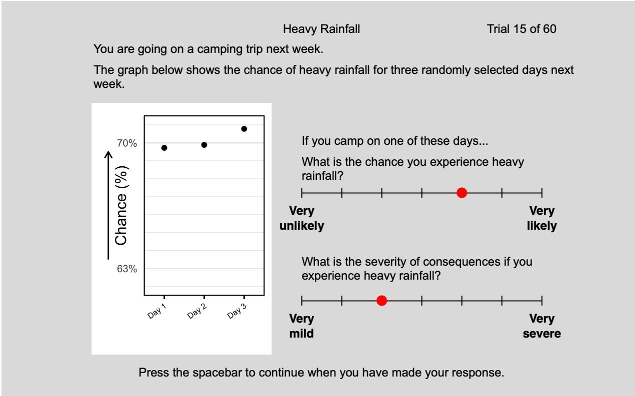

Fig. 3. An example trial, taken from Experiment 1. Participants provided two ratings in each trial: the chance of an event occurring (magnitude rating), and the severity of consequences should the event occur (severity rating).

corrected-to-normal vision and English fluency were required for participation.

We planned to recruit 150 participants. Ten additional participants were excluded for answering more than two of 10 attention check questions incorrectly. The remaining 150 submissions were used in the analysis: (52.00% male, 45.33% female, 2.67% non-binary). Mean age was 31.49 (*SD* = 12.47)[1]. The mean data visualisation literacy score was 21.28 (*SD* = 4.58), out of a maximum of 30. Participants whose submissions were approved were paid £3.55. Average completion time was 25 minutes [2].

3.2 Analysis Technique

Analyses were conducted using R [version 4.2.1, 23].

Likert scales express granularity at the level of ordinal data. They record whether one rating is higher or lower than another, but not the extent of this difference. Therefore, Likert scales do not necessarily capture values from latent distributions (mental representations) in a linear manner. The distance between one pair of points and another pair may appear equal, but may represent different distances on the latent distribution. Therefore, it is inappropriate to analyse Likert scale data with metric models, such as linear regression [18]. Throughout this paper, we construct cumulative link mixed-effects models, using the *ordinal* package [version 2019.12-10, 6] to analyse Likert scale ratings. Odds ratio effect sizes were converted to Cohen's d values using the *effectsize* package [version 0.8.3, 2].

Selection of model random effects structures was automated using the *buildmer* package in R [version 2.8, 30]. The maximal random effects structure included random intercepts for participants and scenarios, plus corresponding slopes for the position variable [1]. The *buildmer* package initially identified the most complex model which could successfully converge. It subsequently removed terms which did not contribute substantially to explaining variance in ratings.

3.3 Results

3.3.1 Magnitude Ratings

Figure 4 plots the distribution of magnitude ratings. Values presented at high physical positions elicited a greater proportion of responses at the higher end of the rating scale than values presented at low physical positions.

A likelihood ratio test reveals that a model including physical position as a fixed effect explains significantly more variability in ratings than a model which does not include physical position as a fixed effect: $\chi^2(1) = 74.21$, $p < .001$. Data points' magnitudes were rated as greater when those data points were presented at high physical positions, compared to when the same data points were presented at low physical positions ($z = 8.57$, $p < .001$).

The odds ratio for the difference between conditions is 1.61, 95% CI [1.44, 1.79]. Participants were 1.61 times more likely to respond with a higher magnitude rating to data points presented at high positions than data points presented at low positions. This is equivalent to a Cohen's d value of 0.26, a small effect size.

This model included random intercepts for each participant and each scenario.

3.3.2 Severity Ratings

For ratings of the severity of consequences, a likelihood ratio test reveals that a model including physical position as a fixed effect explains significantly more variability in ratings than a model which does not include condition as a fixed effect: $\chi^2(1) = 6.16$, $p = .013$. The severity of consequences was rated as greater when data points representing the chance of an event occurring were presented at high physical positions, compared to when the same data points were presented at low physical positions ($z = 2.50$, $p = .012$).

The odds ratio for the difference between conditions is 1.21, 95% CI [1.04, 1.41]. Participants were 1.21 times more likely to respond with a higher severity rating to data points presented at high positions than data points presented at low positions. This is equivalent to a Cohen's d value of 0.11, a very small effect size.

This model employed random intercepts for each scenario, plus random intercepts and slopes for each participant. The slopes modelled, for each participant, the average difference between responses to data presented at different positions.

3.3.3 Data Visualisation Literacy

Adjusting for participants' data visualisation literacy scores did not eliminate the effect of the position of data points on ratings of magnitude ($z = 8.57$, $p < .001$, odds ratio = 1.61, 95% CI [1.44, 1.79]) or the severity of consequences ($z = 2.51$, $p = .012$, odds ratio = 1.21, 95% CI [1.04, 1.41]). These models were identical to the above models except for the inclusion of subjective data visualisation literacy scores as an additional fixed effect.

However, further analysis found that, for magnitude ratings, there was an interaction between physical position and data visualisation literacy: $z = 4.46$, $p < .001$, odds ratio = 0.95, 95% CI [0.92, 0.97]. The difference between the literacy trends for data points presented at high and low positions was -0.05, indicating that this difference *decreased* as data visualisation literacy increased. This interaction is equivalent to a Cohen's d value of 0.03, a very small effect size. This model employed random intercepts for scenarios with random slopes for literacy, plus random intercepts for participants.

For ratings of the severity of consequences, there was no significant interaction between physical position and data visualisation literacy: $z = 1.88$, $p = .061$, odds ratio = 0.97, 95% CI [0.94, 1.00]. This is equivalent to a Cohen's d value of 0.02, a very small effect size. This model employed random intercepts for participants with random slopes for position, plus random intercepts for scenarios with random slopes for literacy.

3.4 Discussion

This experiment demonstrates that axis limits, which determine the position of plotted values, influence inferences about the magnitude of data points. Participants rated *the same values* as greater when these values were plotted at high positions, compared to low positions. This difference is associated with a small effect size. Higher data visualisation literacy levels were associated with a reduced effect, but accounting for data visualisation did not eliminate this effect.

Even though the charts only displayed data on the chance of negative outcomes occurring, ratings of severity of consequences were also greater when data points were presented at high positions. The effect on severity ratings was not associated with differences in data visualisation literacy.

[1]Age data were unavailable for one participant.

[2]Timing data were unavailable for two participants.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3364069

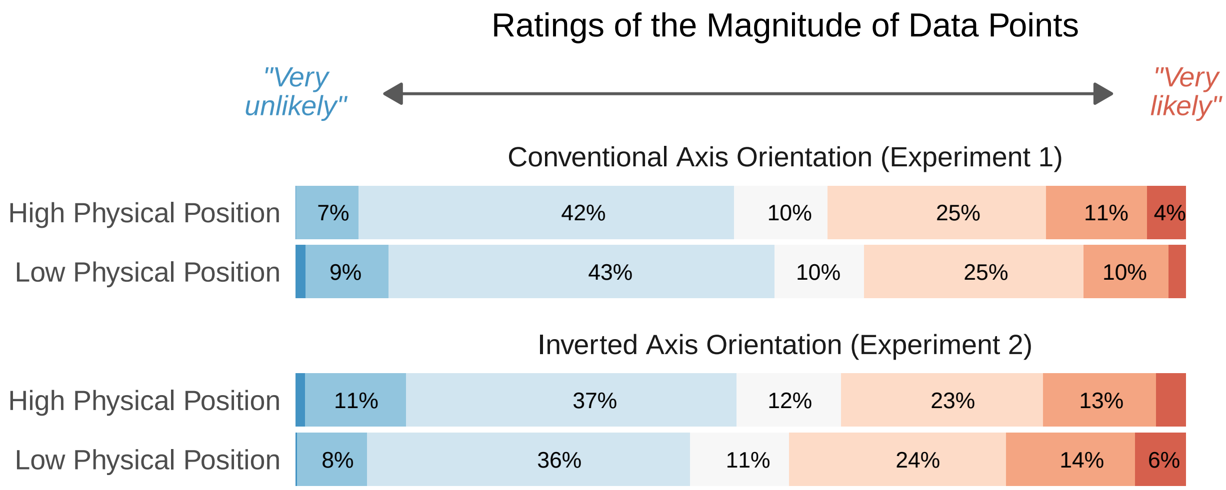

Fig. 4. The distribution of Likert scale ratings of the magnitude of data points, shown separately for charts with conventional axes (Experiment 1) and charts with inverted axes (Experiment 2). The width of each response option represents the proportion of ratings recorded for that option. In Experiment 1, data points presented at high physical positions (top) elicited a larger proportion of ratings on the right-hand side (representing greater magnitudes), compared to data points at low physical positions (bottom), which elicited a larger proportion of ratings on the left-hand side (representing smaller magnitudes). In Experiment 2, this pattern was reversed, with data points at low physical positions (bottom), which elicited a larger proportion of ratings on the right-hand side (representing greater magnitudes).

4 Experiment 2

4.1 Introduction

Experiment 1 (E1) found that participants associated data points with greater magnitudes when those data points were positioned near the *top* of a chart, compared to when the same data points were positioned near the *bottom* of a chart.

One possible explanation for this finding is that participants made simple associations between physical position and magnitude, equating physically higher data points with larger magnitudes and physically lower data points with smaller magnitudes. This is congruent with well-established conceptual metaphors for magnitude, where greater vertical positions denote greater magnitudes [29].

An alternative explanation is that participants used the y-axis as a frame of reference for assessing the magnitude of plotted values. For example, when considering data points near the bottom of the axis, participants may have recognised the potential for values larger than those observed, consequently associating plotted values with smaller magnitudes.

Experiment 1 does not provide a means of differentiating these competing explanations. Drawing inferences from the physical positions of data points would bias magnitude judgements in the same direction as drawing inferences from their relative position within axis limits. A high magnitude is implied by a data point's high physical position *and* its superior position in the context other of presented values. Therefore, an additional experiment is required in order to distinguish between the two competing explanations.

Plotting numerical values along the x-axis would not assist in answering this question, since values that are large in the context of the x-axis limits would be positioned on the right-hand side, which is also typically associated with larger magnitudes [32]. However, inverting a vertical axis changes the typical relationship between physical position and numerical value: increasingly *lower* positions represent increasingly *higher* numerical values. This means data points presented near the *bottom* of a chart are numerically *larger* than the accompanying y-axis values. Therefore, inferences invoking relative position would bias magnitude judgements in the opposite direction compared to inferences invoking the physical positions of data points. This is illustrated in Figure 5. In Experiment 2, we manipulate the physical positions of data points by changing axis limits (as in Experiment 1), but employ inverted y-axes.

Previous research suggests that charts with inverted axes can be prone to misinterpretation when viewers are not informed about the inversion [21, 33]. Therefore, we provided explicit instruction to ensure participants were aware that inverted charts were presented.

4.2 Method

4.2.1 Materials

Materials were identical to E1, except for the inversion of the y-axis in all charts, including practice trials.

4.2.2 Procedure

The experiment used PsychoPy version 2021.2.3. One slide in the instructions explained to participants how charts with inverted axes function: *"In all graphs in this experiment, the arrow on the 'Chance' axis points downwards, meaning the numbers get bigger as the axis goes down."*. Otherwise, the procedure was identical to E1.

4.2.3 Design

As in E1, we employed a repeated-measures, within-participants design.

4.2.4 Participants

A viral social media post on 24th July 2021 endorsing the Prolific.co platform attracted many new users from a narrow demographic, heavily skewing the distribution of participants [5]. Therefore, the experiment was not advertised to users who signed-up to Prolific.co after 24th July 2021, or to those who had participated in E1.

We planned to recruit 150 participants. Eleven additional participants were excluded because they answered more than two of 10 attention check questions incorrectly (10 participants) or because they exceeded the maximum completion time (87 minutes; one participant).

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3364069

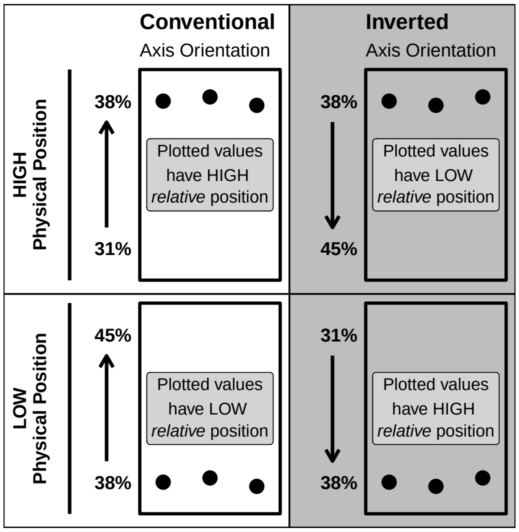

Fig. 5. Rationale for Experiment 2: distinguishing the roles of physical position and relative position. In charts with conventional axis orientations (left column), there is congruence between data points' physical positions and their relative positions within axis limits. In charts with inverted axis orientations (right column), there is incongruence between data points' physical positions and their relative positions within axis limits. This allows us to examine which cue informs magnitude judgements.

The remaining 150 submissions were used in the analysis: (60.00% male, 40.00% female). Mean age was 29.64 (*SD* = 9.56)[3]. 100% had completed at least secondary education. The mean data visualisation literacy score was 21.87 (*SD* = 4.28). Participants whose submissions were approved were paid £3.45, and average completion time was 24 minutes.

4.3 Analysis

We used the same analysis methods as in Experiment 1.

4.4 Results

4.4.1 Magnitude Ratings

Figure 4 plots the distribution of magnitude ratings. Values presented at *low* physical positions elicited a greater proportion of responses at the higher end of the rating scale than values presented at *high* physical positions.

A likelihood ratio test reveals that a model including physical position as a fixed effect explains significantly more variability in ratings than a model which does not include physical position as a fixed effect ($\chi^2(1) = 46.45$, $p < .001$). Data points' magnitudes were rated as larger when those data points were presented at *low* physical positions, compared to when the same data points were presented at high physical positions, in contrast to the findings in Experiment 1 (z = 6.80, $p < .001$).

The odds ratio for the difference between conditions is 1.39, 95% CI [1.27, 1.53]). Participants were 1.39 times more likely to respond with a higher magnitude rating to data points presented at *low* positions than data points presented at high positions. This is equivalent to a Cohen's d value of 0.18, a very small effect size.

This model employed random intercepts for each scenario.

4.4.2 Severity Ratings

For ratings of the severity of consequences, a likelihood ratio test reveals that a model including physical position as a fixed effect did not explain significantly more variability in ratings than a model without physical condition as a fixed effect: ($\chi^2(1) = 3.40$, p = .065). The odds ratio for the difference between conditions is 1.13, 95% CI [0.99, 1.28]). This is equivalent to a Cohen's d value of 0.07, a very small effect size.

This model employed random intercepts for each scenario, plus random intercepts for each participant with random slopes for position.

4.4.3 Data Visualisation Literacy

Adjusting for participants' data visualisation literacy scores did not change the pattern of results regarding ratings of the magnitude of data points themselves (z = 7.51, $p < .001$, odds ratio = 0.68, 95% CI [0.61, 0.75]) or the severity of consequences (z = 1.85, p = .064, odds ratio = 1.13, 95% CI [0.99, 1.28]).

For magnitude ratings, there was an interaction between data visualisation literacy and physical position: z = 2.12, p = .034, odds ratio = 1.02, 95% CI [1.00, 1.05]. This is equivalent to a Cohen's d value of 0.01, a very small effect size. The difference between the literacy trends for data points presented at low and high positions was -0.02, indicating that this difference *decreased* as data visualisation literacy increased. This interaction is equivalent to a Cohen's d value of 0.01, a very small effect size. This model employed random intercepts for scenarios with slopes for literacy.

For ratings of the severity of consequences, there was no significant interaction between physical position and data visualisation literacy: z = 0.70, p = .485, odds ratio = 1.01, 95% CI [0.98, 1.04]. This is equivalent to a Cohen's d value of 0.01, a very small effect size. This model employed random intercepts for scenarios with slopes for literacy, plus random intercepts for participants with slopes for position.

4.5 Discussion

This experiment demonstrates that inferences about the magnitude of data points are informed primarily by their *relative* positions within axis limits, as opposed to their physical positions. Viewing data in charts with inverted y-axes, participants rated the same values as greater when these values were plotted at *low* physical positions, compared to *high* physical positions. Ratings reflecting higher magnitudes were awarded more frequently to charts where data points were positioned near the bottom, where the axis limit was associated with a *higher* numerical value. Therefore, axis limits inform judgements primarily through inferences about the numerical context surrounding data points, rather than the connotations of their physical positions.

Ratings of the severity of consequences were not significantly affected by the position of data points representing the chance of negative events occurring. Overall, accounting for differences in data visualisation literacy did not alter the pattern of results, but higher data visualisation literacy scores were associated with a diminished effect of physical position, for magnitude ratings.

5 General Discussion

Given the use of data visualisation for the communication of numerical information, understanding how design choices affect interpretations is an important matter. In a pair of experiments, we demonstrate that judgements of the magnitudes of data points are influenced by a chart's axis limits. These experiments provide insight into the cognitive processes involved in assessing magnitudes in data visualisations.

We manipulated the axis limits accompanying plotted data, which affected the context in which data appeared *and* the physical positions of data points. However, regardless of their physical positions, data points were consistently associated with greater magnitudes when they appeared close to the end of the axis associated with higher values. Interpretation of the same numerical value is biased by its relative position within axis limits. This highlights viewers' sensitivity to surrounding information when assessing data. We illustrate that this framing effect occurs even when no contrasting data points are present to provide context: axis limits are sufficient for informing magnitude judgements. Therefore, charts which aim to communicate absolute magnitude, like the New York Times example discussed above Figure 1, may do so by employing axes which contextualise the magnitude of data points through axis limits.

[3]Age data were unavailable for two participants.

Our findings suggest that axis limits influence, but do not wholly *dictate*, impressions of magnitude. The distribution of magnitude judgements across scenarios approximately followed the distribution of plotted numerical values, suggesting that numerical values also contributed to magnitude judgements. In addition, the effect size associated with the physical positions of data points was larger for charts with conventional axis orientations, compared to charts with inverted y-axes. This suggests that the physical position of data points partially contributed to participants' assessments. Despite this, it is evident from the pattern of results for inverted charts that relative position exerts a greater influence on magnitude judgements. Whilst we cannot conclude that viewers interpret the axis range as the complete context for assessing plotted data, it is clear that axes primarily inform magnitude judgements by defining the relative positions of data points.

5.1 Relationship to Prior Work

The present data complement findings from research on y-axis truncation, which has observed that axis limits accompanying plotted values can influence viewers' impressions of those values. While previous investigations have shown that y-axis limits affect judgements of the relative difference between plotted values [8, 21, 31, 34], the present findings show that they also influence judgements of the absolute magnitude of plotted values. This finding supports the notion that viewers are sensitive to visualisation rhetoric [15] and framing effects [28], wherein particular presentations of numerical information provokes specific interpretations.

A previous study addressing a similar question also concluded that a data point's location within a range of values affects interpretation of its magnitude [25]. The present study builds upon this research by identifying the mechanism behind this effect and removing the confound of variable axes ranges. It also extends the finding beyond a single scenario (asbestos) to a wider range of situations. By analysing different types of judgement separately, rather than using a combined measure, we verify that axis limits affect interpretations of the specific variable displayed in a chart, but not related variables (e.g., severity).

In addition to the conceptual metaphor for magnitude, physical positions are also linked to *emotional valence*, with high positions typically associated with positive valence. Woodin et al. [33] found that physical arrangements of data which are consistent with the conceptual metaphor for valence somewhat facilitate comprehension, but that associations between position and magnitude affect interpretations more strongly. Visualisations in the present experiments displayed data on negative events, so data were aligned with the conceptual metaphor for valence in inverted charts, and misaligned in conventional charts. Participants evidently did not use valence metaphors to interpret values in conventional charts; this would have produced the opposite pattern of results to those observed. For inverted charts, we cannot differentiate between responses based on relative position and responses based on a conceptual metaphor for valence. However, the simplest explanation for our results suggests that participants relied on the same relative position cue when interpreting both conventional and inverted charts.

5.2 Additional Findings

Prior research has observed positive correlations between perceptions of event probability and outcome magnitude [13, 14, 17]. We did not find robust evidence that assessments of the severity of consequences were affected by our manipulation of data points representing the chance of events occurring. However, whereas prior work substantially manipulated underlying scenarios, our subtler manipulation retained the same probability values, changing only the surrounding context. In addition, participants evaluated the severity of an event's *consequences*, which is one step removed from the property explored in prior research: the severity of the event itself. The effects of axis limits on interpretation of data about the incidence of events do not reliably extend to judgements about their consequences.

Adjusting for subjective data visualisation literacy did not change the overall pattern of results in either of the experiments. This indicates that any biases observed cannot be fully explained by differences in data visualisation literacy. Despite observing that bias *reduced* as data visualisation literacy increased, the effect sizes associated with these interactions were very small. Thus, the data visualisation literacy measure provided low explanatory power. Yang et al. [34] also observed that data visualisation literacy could not sufficiently explain variance in the degree of bias caused by y-axis truncation. This measure reflects comprehension of the conventions of data visualisation, indicating receipt of elementary instruction [20]. Therefore, it is perhaps better suited for capturing viewers' application of basic knowledge in interpretation [34], whereas Ge et al.'s CALVI test [11] may be more appropriate for predicting susceptibility to differences in presentation format.

5.3 Limitations and Future Directions

We employed inverted y-axes solely for the purpose of distinguishing competing explanations; their use should not be considered an endorsement. To avoid misinterpretations, participants were given instruction on how to read inverted charts. With this explicit instruction, our data provide evidence *contrary* to the typical finding of misinterpretation resulting from associating higher positions with higher values [21, 33]. However, this instruction may have suppressed a spontaneous interpretation of magnitude, based on physical position, in favour of a learned interpretation. Our investigation therefore only explores the cognitive processing associated with assessing magnitude in charts which viewers know how to read.

This study was designed to explore one factor involved in assessing magnitude. The influence of axis limits on interpretations was relatively small, raising the question of how much this factor influences real-world decision-making and behaviour. A forced choice measure, or response scale with concrete values, would be suitable for capturing these outcomes in future work. Addressing this question will help to quantify how much a designer's choice of axis limits affects a viewer's choices and actions. An appreciable effect would have implications for visualisation design, suggesting use of axis limits which convey magnitude appropriately to avoid misleading users. Suitable axis limits cannot be objectively determined, but must be informed by the designer, based on their assessment of the data [8]. The effects of axis limits on discrimination ability would also warrant consideration, taking account of the intended application. In addition, future work could systematically manipulate the range of the axes (e.g., expanding beyond 10 percentage points) to investigate how this influences magnitude judgements. Pre-registration of experiments and analysis protocol would also increase the credibility of future work on this topic.

6 Conclusion

We conducted two experiments investigating how axis limits inform interpretations of the magnitude of plotted values. Subjective judgements were affected by the positions of data points in relation to accompanying axis limits. The association between the positions of data points and magnitude judgements critically depends on whether plotted data appear closer to the axis limit associated with higher or lower values. The cognitive processes associated with assessing magnitude in data visualisations involve taking into account the context in which the data appear.

References

[1] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278, doi:10.1016/j.jml.2012.11.001.

[2] M. Ben-Shachar, D. Lüdecke, and D. Makowski. effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56):2815, doi:10.21105/joss.02815.

[3] M. A. Borges and B. K. Sawyers. Common verbal quantifiers: Usage and interpretation. *Journal of Experimental Psychology*, 102(2):335–338, doi:10.1037/h0036023.

[4] G. D. A. Brown, J. Gardner, A. J. Oswald, and J. Qian. Does wage rank affect employees' well-being? *Industrial Relations*, 47 (3):355–389, DOI:10.1111/J.1468-232X.2008.00525.X.

[5] N. Charalambides. We recently went viral on TikTok - here's what we learned, 2021. [Online]. Available: https://blog.prolific.co/we-recently-went-viral-on-tiktok-heres-what-we-learned/.

[6] R. H. B. Christensen. ordinal—Regression models for ordinal data, 2019. [Online]. Available: https://CRAN.R-project.org/package=ordinal.

[7] W. S. Cleveland, P. Diaconis, and R. McGill. Variables on scatterplots look more highly correlated when the scales are increased. *Science*, 216(4550):1138–1141, DOI:10.1126/SCIENCE.216.4550.1138.

[8] M. Correll, E. Bertini, and S. Franconeri. Truncating the y-axis: Threat or menace? In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Honolulu HI USA, 2020. DOI:10.1145/3313831.3376222.

[9] R. Garcia-Retamero, E. T. Cokely, S. Ghazal, and A. Joeris. Measuring graph literacy without a test: A brief subjective assessment. *Medical Decision Making*, 36(7):854–867, DOI:10.1177/0272989X16655334.

[10] M. Gattis. Structure mapping in spatial reasoning. *Cognitive Development*, 17(2):1157–1183, DOI:10.1016/S0885-2014(02)00095-3.

[11] L. W. Ge, Y. Cui, and M. Kay. CALVI: Critical Thinking Assessment for Literacy in Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–18, Hamburg Germany, 2023. DOI:10.1145/3544548.3581406.

[12] P. Grice. Logic and Conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics Vol.3: Speech Acts*, pages 41–58. Academic Press, New York, 1975.

[13] A. J. L. Harris and A. Corner. Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6):1571–1578, DOI:10.1037/A0024195.

[14] A. J. L. Harris, A. Corner, and U. Hahn. Estimating the probability of negative events. *Cognition*, 110(1):51–64, DOI:10.1016/J.COGNITION.2008.10.006.

[15] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2231–2240, DOI:10.1109/TVCG.2011.255, Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[16] G. Keren and K. H. Teigen. The probability—outcome correspondence principle: A dispositional view of the interpretation of probability statements. *Memory and Cognition*, 29(7):1010–1021, DOI:10.3758/BF03195763.

[17] D. Kupor and K. Laurin. Probable cause: The influence of prior probabilities on forecasts and perceptions of magnitude. *Journal of Consumer Research*, 46(5):833–852, DOI:10.1093/JCR/UCZ025.

[18] T. M. Liddell and J. K. Kruschke. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79:328–348, DOI:10.1016/J.JESP.2018.08.009.

[19] Y. Okan, R. Garcia-Retamero, M. Galesic, and E. T. Cokely. When higher bars are not larger quantities: On individual differences in the use of spatial information in graph comprehension. *Spatial Cognition and Computation*, 12(2-3):195–218, DOI:10.1080/13875868.2012.659302.

[20] Y. Okan, M. Galesic, and R. Garcia-Retamero. How people with low and high graph literacy process health graphs: evidence from eye-tracking: graph literacy and health graph processing. *Journal of Behavioral Decision Making*, 29(2-3):271–294, DOI:10.1002/BDM.1891.

[21] 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 *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 1469–1478, Seoul, Republic of Korea, 2015. DOI:10.1145/2702123.2702608.

[22] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1): 195–203, DOI:10.3758/S13428-018-01193-Y.

[23] R Core Team. R: A language and environment for statistical computing, 2022. [Online]. Available: https://www.R-project.org/.

[24] V. F. Reyna. A theory of medical decision making and health: Fuzzy trace theory. *Medical Decision Making*, 28(6):850–865, DOI:10.1177/0272989X08327066.

[25] P. M. Sandman, N. D. Weinstein, and P. Miller. High risk or low: How location on a "risk ladder" affects perceived risk. *Risk Analysis*, 14(1):35–45, DOI:10.1111/J.1539-6924.1994.TB00026.X.

[26] B. G. Taylor and L. K. Anderson. Misleading graphs: Guidelines for the accountant. *Journal of Accountancy*, 162(4):126–135, 1986.

[27] A. Tversky and D. Kahneman. Judgment under uncertainty: heuristics and biases. *Science*, 185(4157):1124–1131, DOI:10.1126/SCIENCE.185.4157.1124, Publisher: American Association for the Advancement of Science Section: Articles.

[28] A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458, DOI:10.1126/SCIENCE.7455683.

[29] B. Tversky. Cognitive principles of graphic displays. In *Proceedings of the AAAI 1997 Fall Symposium on Reasoning with Diagrammatic Representations*, pages 116–124, Menlo Park, CA, 1997. AAAI Press,.

[30] C. C. Voeten. buildmer: Stepwise elimination and term reordering for mixed-effects, 2022. [Online]. Available: https://CRAN.R-project.org/package=buildmer.

[31] J. K. Witt. Graph construction: An empirical investigation on setting the range of the y-axis. *Meta-Psychology*, 2:1–20, DOI:10.15626/MP.2018.895.

[32] G. Woodin and B. Winter. Placing abstract concepts in space: Quantity, time and emotional valence. *Frontiers in Psychology*, 9 (2169):1–14, DOI:10.3389/FPSYG.2018.02169.

[33] G. Woodin, B. Winter, and L. Padilla. Conceptual metaphor and graphical convention influence the interpretation of line graphs. *IEEE Transactions on Visualization and Computer Graphics*, 28 (2):1209–1221, DOI:10.1109/TVCG.2021.3088343.

[34] B. W. Yang, C. Vargas Restrepo, M. L. Stanley, and E. J. Marsh. Truncating bar graphs persistently misleads viewers. *Journal of Applied Research in Memory and Cognition*, 10(2):298–311, DOI:10.1016/J.JARMAC.2020.10.002.