

To Cut or Not To Cut?

A Systematic Exploration of Y-axis Truncation

Sheng Long

shenglong@u.northwestern.edu
Northwestern University
Evanston, Illinois, USA

Matthew Kay

matthew.kay@northwestern.edu
Northwestern University
Evanston, Illinois, USA

ABSTRACT

Y-axis truncation is a well-known, much-debated visualization practice. Our work complements existing empirical work by providing a systematic analysis of y-axis truncation on grouped bar charts. Drawing upon theoretical frameworks such as Algebraic Visualization Design, we examine how structure-preserving modifications to visualization affect user performance by systematically dividing the space of possible truncations according to their monotonicity and the type of relations in the underlying data. Our results demonstrate that for comparing and estimating the difference between the lengths of two bars, truncating the y-axis does not affect task performance. For comparing or estimating the relative growth between two bars, truncating monotonically has similar performance to no truncation, while truncating non-monotonically is very likely to impair performance. We discuss possible extensions of our work and recommendations for y-axis truncation. All supplementary materials are available at https://osf.io/k4hjd/?view_only=008b087fc3d94be7ba0ce7aea95012a7.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in visualization**; *Visualization theory, concepts and paradigms*.

KEYWORDS

y-axis truncation, deceptive visualization, structure-preserving transformations, algebraic visualization design

ACM Reference Format:

Sheng Long and Matthew Kay. 2024. To Cut or Not To Cut?: A Systematic Exploration of Y-axis Truncation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3613904.3642102>

1 INTRODUCTION

Should a chart of global average temperature start from 0? In a chart of temperature over time created by National Review [6], they extend the y-axis to include 0, but by doing so compress the trend of global warming to obscurity. An alternative design may consider starting the y-axis from a different number, but how do we decide which number? While some guidelines reason that it is

never acceptable to truncate [23, 39], others argue that it might be acceptable for some chart types but not others [4, 8]. Additional guidelines such as Brinton [5] recommend no truncation when the reader's focus is on the relative amount of change, while other guidelines like Skelton [34] recommend no truncation when the reader's focus is on the absolute rate of increase/decrease. Only some of this guidance has been tested by empirical studies, which find that y-axis truncation results in quantifiable differences in how people subjectively interpret the size and significance of effects, even when behavioral and visual interventions (e.g. broken axes) are employed [12, 30, 44, 46].

To understand precisely when it may be appropriate to truncate or not, we systematically analyze how y-axis truncation affects visual properties and identify those that remain invariant after truncation. Specifically, we ask: What happens when we truncate the y-axis in such a way that the underlying relations in data are still preserved in the truncated graph? Motivated by prior work such as Kindlmann and Scheidegger [27] and Demiralp et al. [16], we define y-axis truncations that preserve the underlying relations in data to be **monotonic** and truncations that do not preserve underlying relations in data to be **non-monotonic**.

Our results suggest that for tasks such as comparing or estimating the differences between bars, y-axis truncation (regardless of monotonicity) has little to no effect on user performance. Thus, for these tasks, if the designer wants to elicit a meaningful subjective perception of effect size in the viewers, our results corroborate existing recommendations [12, 31, 44] to truncate the y-axis to show the desired effect size. However, for tasks such as comparing or estimating the percentage change between bars, no truncation, in general, performs similarly to monotonic truncation, and both perform much better than non-monotonic truncation. Thus, we suggest that the designer perform no truncation or monotonic truncation for these tasks.

In real life, designers may find themselves in situations where they face constraints and need to make trade-offs between design choices. Our work provides a rigorous empirical basis that corroborates and expands upon previous recommendations on y-axis truncation by (1) providing a re-evaluation of some existing guidelines about why not to truncate the y-axis, (2) quantifying the degree to which y-axis truncation affects perceptual accuracy (for two data-generating distributions and four tasks), and (3) identifying scenarios where previous research that focused on the subjective perception of effect size can be applied without risk of decreasing performance on judgments of differences or percentage changes between bars.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI '24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0330-0/24/05.

<https://doi.org/10.1145/3613904.3642102>

2 RELATED WORK

2.1 Graphical Perception & Bar Charts

Graphical perception was introduced by Cleveland and McGill [10] as “the visual decoding of information encoded on graphs”. They tested the impact of bar chart designs (simple and stacked) on participants’ ability to estimate the ratio between the lengths of two bars and found that the designs in which the bars are *aligned along a common baseline* scored better than the designs where the comparisons are unaligned. They hypothesized that this difference is due to the use of two different visual estimation strategies – for aligned bars, viewers make a visual comparison of *positions* while for unaligned bars, viewers make a much less accurate visual comparison of *lengths*. They suggest encoding information with position instead of length for better perceptual accuracy. Cleveland and McGill [10] also found that comparisons between adjacent bars are more accurate than between widely separated bars. In our experiment, we had two comparison tasks. The first comparison task asked the participants to compare the difference in heights between two groups of bars, which would require the participants to make an unaligned length comparison. The second comparison task asked the participants to compare the percentage change (i.e., $(\text{tall bar} - \text{short bar}) / \text{short bar}$) between two groups of bars, which would either require the participants to use both unaligned and aligned length comparison or a similar perceptual judgment to that of a perceptual judgment of correlation.

Talbot et al. [37] replicated Cleveland and McGill’s finding that separating bars in space makes the comparison of their height more difficult. They also looked at the effects of distractor bars, but their results were not conclusive. Zhao et al. [50] investigated how neighboring bars affect the perception of order in bar charts, and found that while neighborhood effects do exist in rank estimation tasks, the effect size of neighborhood effects is small compared to other data-inherent effects. From these prior results, we decided to not explicitly model distractor bars because we think the effect size is going to be small. We also did not include the distance between the two groups of bars in our analysis because it is randomized across all conditions.

2.2 Deceptive Visualization & Y-axis Truncation

Pandey et al. [30], the first work to empirically examine deceptive techniques in data visualizations, defined *deceptive visualization* to be “a graphical depiction of information ... that may create a belief about the message and/or its components, which varies from the actual message”. Heer and Correll [13] also presented a definition that emphasizes the encouraged interpretation that deviates from the actual message in the data. Among studied deceptive visualization practices, we focus on **y-axis truncation**, a practice of beginning the vertical axis at a value other than 0. Prior empirical research [12, 30, 44, 45] has studied this topic by focusing on the *subjective* perception of change. These papers find that truncated bar charts persistently increase the perceived magnitude of trends, despite behavioral interventions (e.g., warning and teaching people about y-axis truncation before showing them the stimuli [45]) and visual interventions (e.g., adding broken axes to emphasize the truncated y-axis [14]). As a result, these papers suggest either completely avoiding y-axis truncation for bar charts [45] or truncating it in

a way that aligns with the task at hand or conveys meaningful effect size to the viewers [12, 44]. In our paper, we quantify the degree of this increased perceived magnitude by asking participants to estimate the **absolute change** (i.e., difference) or the **relative change** (i.e., percentage change) between bars. We delay a more detailed discussion of prior recommendations and guidelines to subsection 6.1 to better summarize motivations for these recommendations as well as how our motivations and recommendations fit in with existing ones. We note that truncated axes are not unique to bar charts. Nevertheless, arguments for other chart types tend to focus on how truncation changes the **aspect ratio** of these charts. Existing research has examined how to set the optimal aspect ratio for line charts [9, 21, 36, 42] and for scatter plots [18]. Our work does *not* examine the effects of aspect ratio changes, as we do not manipulate the aspect ratios of the bars or the bar chart itself. We truncate the y-axis, which alters the length of bars within the bar chart. While one could argue that y-axis truncation alters the aspect ratio of the position encodings of bars, our experiment primarily utilizes length judgments, thus the changed position encoding is not relevant to our work.

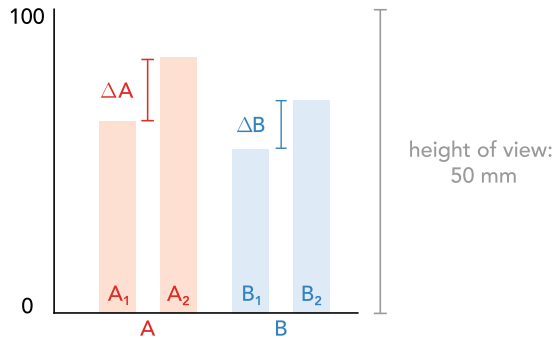
2.3 Structure in visualization

This paper was inspired by two groups of work in formal evaluative theories of visualization. The first group examines how data distributions and user tasks influence the effectiveness of charts [32, 33] and visual encodings [26]. Hu et al. [22] proposed to view every realized visualization as a tuple of (data D , visual form V , task T). McNutt [29] adopted this view and conducted an algebraic analysis of table cartograms by fixing the visual form, varying either data or user task, and observing the remaining element in the triplet. Behrisch et al. [3] provides a detailed survey on research that views visualization design as a multi-objective optimization problem. One such research is Dasgupta et al. [15], in which the authors proposed a model to quantify the different visual structures in parallel coordinate graphs, as well as a way of automatically optimizing the display based on their model.

The second group [16, 27, 48] emphasizes the connection between mathematical structure in the underlying data and mathematical structure in the perception of visualization. Kindlmann and Scheidegger [27] proposed Algebraic Visualization Design (AVD), a framework for reasoning about the design of data visualizations through their intrinsic symmetries. AVD explores the effect of changes in data on resulting images. It is composed of three general principles, including the **principle of visual-data correspondence**, which advocates for “[matching] mathematical structure in data with that in visual perception”. Similarly, Demiralp et al. [16] defined a visualization as a function that maps from a domain of data points to a range of visual primitives. They argued that a visualization is “good” if the embedded visual elements preserve structures present in the data domain.

Albeit not mathematical, Zacks and Tversky [49] shared a similar view, in that there exists *data correspondence* (between graph type and data type) and *message correspondence* between graph type and the intended message. They showed that there was “a strong tendency to portray discrete comparison descriptions as bars and trend assessment descriptions as lines”.

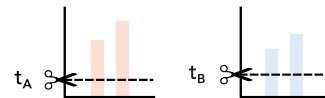
When we have a data set $\begin{cases} A_1: 63 \\ A_2: 84 \end{cases}$, $\begin{cases} B_1: 54 \\ B_2: 70 \end{cases}$ and plot it into a bar chart ...



... the visual relation is consistent with the **concordant data relation**:

$$\begin{aligned} \Delta A &> \Delta B \\ \Delta A / A_1 &> \Delta B / B_1 \end{aligned}$$

If we treat **group A** and **group B** as separate bar plots, and truncate them differently with t_A and t_B ...



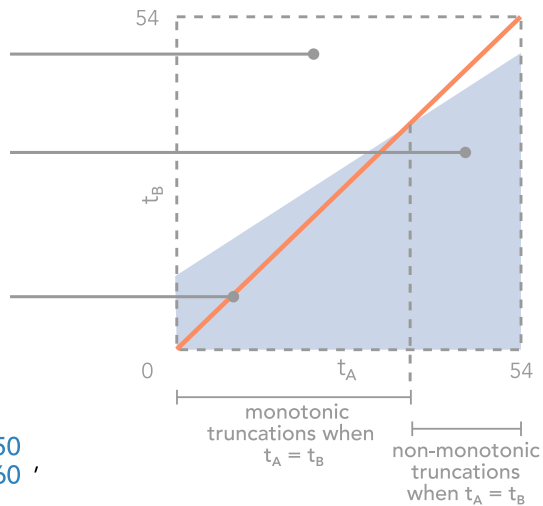
then ...

in the **non-shaded region**, all truncations render a reversed data relation

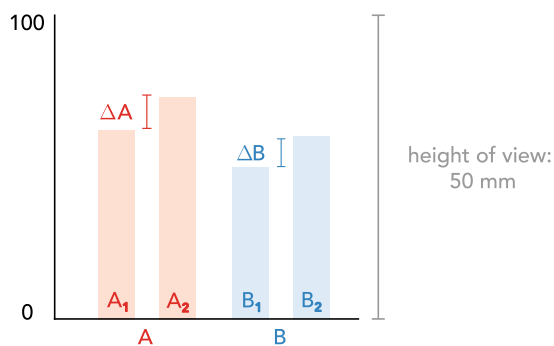
in the **shaded region**, all truncations retain the data relation

truncations on the 45 degree line imply that $t_A = t_B$.

This is the only case in the paper, since **group A** and **group B** are **always** in the same frame.



Likewise, if we have the data set $\begin{cases} A_1: 62 \\ A_2: 73 \end{cases}$, $\begin{cases} B_1: 50 \\ B_2: 60 \end{cases}$, the corresponding bar plot is ...



... which is consistent with the **discordant data relation**:

$$\begin{aligned} \Delta A &> \Delta B \\ \Delta A / A_1 &< \Delta B / B_1 \end{aligned}$$

truncating when $t_A = t_B$ would **always lead to a monotonic truncation**.

In other words, it is impossible to perform non-monotonic truncations on discordant data relations.

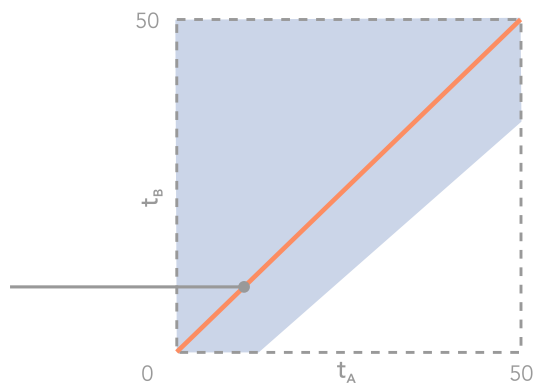


Figure 1: What are monotonic truncations and what are non-monotonic truncations?

3 A MODEL FOR Y-AXIS TRUNCATION

We synthesize the ideas from Demiralp et al., Kindlmann and Scheidegger, and Hu et al. [16, 22, 27] and form the following model for reasoning about visualizations:

$$\text{Data } D \leftrightarrow \text{Visual forms } V \leftrightarrow \text{Tasks } T$$

Given a particular data set d in the space of all possible data sets D , d could be mapped to a set of different visual configurations $V_d \subseteq V$, some more structure-preserving (or having better “data-visual” correspondence) than others. Each feasible visual configuration $v \in V_d$ can be used to answer tasks, be it low-level [2] or high-level [38]. Some visual configurations provide more affordance for certain tasks [32, 49] and may support better perceptual accuracy. Others might not support perceptual accuracy but better decision-making [24] or require a lower level of cognitive effort.

This model allows us to go one step further beyond matching mathematical structure in data with that of visualization. Instead, we want to examine the nature of these correspondences between data, visual forms, and tasks, and be able to characterize these correspondences. Similar to Kindlmann and Scheidegger [27], we start by introducing changes into our model – if we alter the visual configurations in a way that preserves the inherent structures in data, how does it affect the performance of tasks? Formally, if function $f : V \rightarrow V$ is a structure-preserving function applied to visual configurations, would $D \leftrightarrow f(V) \leftrightarrow T$ be much different than $D \leftrightarrow V \leftrightarrow T$? And what can we say about the properties of such a structure-preserving function $f(\cdot)$?

To make our question more concrete, consider the bar charts in Figure 2.

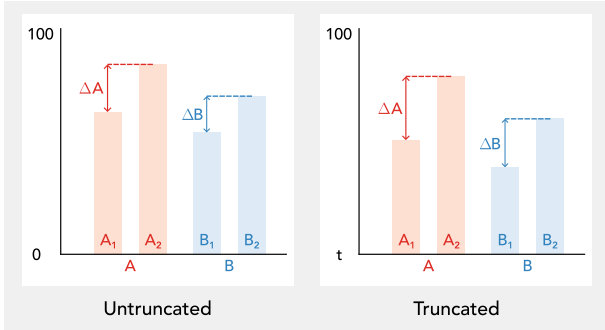


Figure 2: Left: example of untruncated bar chart. Right: example of truncated bar chart. The two charts share the same data set $\{A_1, A_2, B_1, B_2\}$.

First, look at group A in the untruncated bar chart. Group A contains two elements, $\{A_1, A_2\}$, from which we can derive at least three quantities: mean $\frac{1}{2}(A_1 + A_2)$, gap $A_2 - A_1$, and ratio A_1/A_2 . Let L denote the height of the viewport on which the bar chart materializes. Then, the gap ΔA in the data space maps to $\Delta A/100 \times L$ in the visual space. The underlying relation between ΔA and ΔB in data space, i.e., $\Delta A > \Delta B$, still holds in the visual space:

$$\Delta A > \Delta B \implies \Delta A/100 \times L > \Delta B/100 \times L \quad (1)$$

which justifies the usage of length of bars to answer $\max\{\Delta A, \Delta B\}$.

Let’s introduce a change by truncating the y-axis. Let t denote the amount truncated for the bar chart on the right in Figure 2. Then, in the truncated bar chart, the length of ΔA is now $\Delta A/(100-t) \times L$. If we simply compare the length of ΔA in the truncated bar chart to the length of ΔA in the untruncated bar chart, we would conclude that truncation leads to visual exaggeration, since $\Delta A/(100-t) \times L > \Delta A/100 \times L$. However, using the rendered lengths of ΔA and ΔB in the truncated bar chart to answer the question of $\max\{\Delta A, \Delta B\}$ should nevertheless work since the relation that holds in data space still holds in the visual space:

$$\Delta A > \Delta B \implies \Delta A/(100-t) \times L > \Delta B/(100-t) \times L \quad (2)$$

Further, note that the above inequality in Equation 2 implies that regardless of the amount t we truncate, the relation between ΔA and ΔB in data space, $\Delta A > \Delta B$, would still hold in the visual space.

Consider another judgment task: What is the answer to $\max\{A_1/A_2, B_1/B_2\}$? In the example (Figure 2), the underlying relation in data is $A_1/A_2 < B_1/B_2$. In the untruncated bar chart, we could still use the visual length of bar charts and arrive at the same conclusion:

$$\frac{A_1}{A_2} < \frac{B_1}{B_2} \implies \frac{A_1/100 \times L}{A_2/100 \times L} < \frac{B_1/100 \times L}{B_2/100 \times L} \quad (3)$$

However, in the truncated bar chart, if we rely on the visual length of A_1, A_2, B_1, B_2 to judge whether $A_1/A_2 < B_1/B_2$, then for our judgment to be valid, we need

$$\frac{(A_1 - t)/(100 - t) \times L}{(A_2 - t)/(100 - t) \times L} < \frac{(B_1 - t)/(100 - t) \times L}{(B_2 - t)/(100 - t) \times L} \quad (4)$$

$$\implies \frac{A_1 - t}{A_2 - t} < \frac{B_1 - t}{B_2 - t} \quad (5)$$

$$\leftrightarrow \underbrace{(A_2 - A_1)}_{\Delta A} + \underbrace{(B_1 - B_2)}_{-\Delta B} \times t < B_1 A_2 - A_1 B_2 \quad (6)$$

$$\implies t < \frac{B_1 A_2 - A_1 B_2}{\Delta A - \Delta B} \quad (7)$$

In other words, if the truncation t satisfies Equation 7, then using length in the truncated bar chart to judge $\max\{A_1/A_2, B_1/B_2\}$ should not differ from that of using length in the untruncated bar chart, since the relation between A_1/A_2 and B_1/B_2 in data is preserved by the relation in visualization.

3.1 Data relations

Among the different quantities of interest in a grouped bar chart, we choose to focus on two quantities that have been examined by prior work [10]: (1) **gap** $A_2 - A_1$, and (2) **percentage change** $(A_2 - A_1)/A_1$. To ensure completeness, we use Allen’s interval algebra [1], a calculus initially introduced for temporal reasoning but which has also found applications in visualization research [15]. Three out of the thirteen base relations of Allen’s interval algebra apply to the quantities we study: (1) $X > Y$, (2) $X < Y$, and (3) $X = Y$.¹ Ignoring equality, there are two relations in data space for **gaps** between groups A and B: (1) $\Delta A > \Delta B$ and (2) $\Delta A < \Delta B$. There are also two relations in data space for **percentage change** between groups A and B: (1) $\Delta A/A_1 > \Delta B/B_1$ and (2) $\Delta A/A_1 < \Delta B/B_1$. A

¹Because all bars start at the same baseline, most of Allen’s relations do not apply. If we were examining another visual form, e.g., stacked bar charts, more of the relations in Allen’s interval algebra apply.

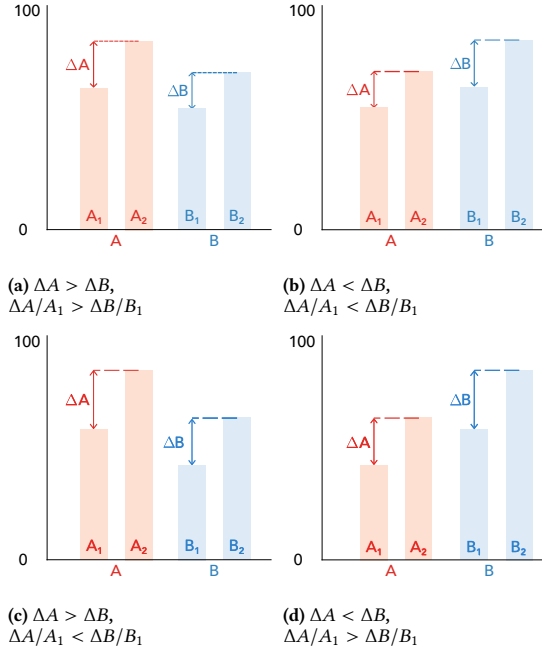


Figure 3: An example of the four kinds of data relations.

combination of these relations in data space gives us a total of four kinds of **data relations** when looking at two groups of bar charts:

- (1) Figure 3a: $\Delta A > \Delta B$, $\Delta A/A_1 > \Delta B/B_1$;
- (2) Figure 3b: $\Delta A < \Delta B$, $\Delta A/A_1 < \Delta B/B_1$;
- (3) Figure 3c: $\Delta A > \Delta B$, $\Delta A/A_1 < \Delta B/B_1$;
- (4) Figure 3d: $\Delta A < \Delta B$, $\Delta A/A_1 > \Delta B/B_1$.

For ease of exposition, we group the first two data relations as **concordant data relations** (since the trend in comparing percentage changes is the same as the trend in comparing gaps) and the later two data relations as **discordant data relations**.

3.2 Monotonicity of truncation

We define a **monotonic** y-axis truncation to be a truncation that does not result in a reversed data relation at hand (Figure 1). Let t denote the amount of truncation performed. Then, for example, a **monotonic truncation** for a concordant data relation $\Delta A > \Delta B$, $\Delta A/A_1 > \Delta B/B_1$ implies that

$$\frac{\Delta A}{A_1 - t} > \frac{\Delta B}{B_1 - t} \quad (8)$$

$$\Leftrightarrow t < \frac{\Delta B A_1 - \Delta A B_1}{\Delta B - \Delta A} \quad (9)$$

while a **non-monotonic truncation** implies that

$$\frac{\Delta A}{A_1 - t} < \frac{\Delta B}{B_1 - t} \quad (10)$$

$$\Leftrightarrow t > \frac{\Delta B A_1 - \Delta A B_1}{\Delta B - \Delta A} \quad (11)$$

Based on our definitions of data relations and monotonic y-axis truncation, it is impossible to perform a *non-monotonic* y-axis truncation for *discordant data relations* (refer to Appendix A). Put

another way, for discordant data relations, it is impossible to visually alter the underlying data relations regardless of how one truncates the y-axis.

Connecting this example back to the $D \leftrightarrow V \leftrightarrow T$ model, a *monotonic truncation* on the y-axis is simply a structure-preserving function $f(\cdot)$. We hypothesize that monotonic y-axis truncations will not affect user performance much (that is, $D \leftrightarrow \text{bar chart} \leftrightarrow T$ should behave similarly to $D \leftrightarrow f(\text{bar chart}) \leftrightarrow T$ when $f(\cdot)$ is monotonic), while non-monotonic truncations would have a more adverse impact on user performance.

4 METHODOLOGY

To test our hypothesis that (1) $D \leftrightarrow f(\text{bar chart}) \leftrightarrow T$ scores similarly to $D \leftrightarrow \text{bar chart} \leftrightarrow T$ when $f(\cdot)$ is monotonic, and (2) $D \leftrightarrow f(\text{bar chart}) \leftrightarrow T$ scores worse compared to $D \leftrightarrow \text{bar chart} \leftrightarrow T$ when $f(\cdot)$ is non-monotonic, we designed our experiment with two different data-generating distributions and four different tasks. Our experiment followed a mixed factorial design. The **between-subjects** factor is the judgment task. The three **within-subjects** factors are (1) data-generating distribution, (2) data relation, and (3) type of truncation. We preregistered our conditions, sample sizes, and analysis on osf.io. The remaining of this section details our choices for each experiment factor.

Please use the graph below to answer the following question:

What is the **ratio** between the profit in 2020 and the profit in 2021 for **April**?

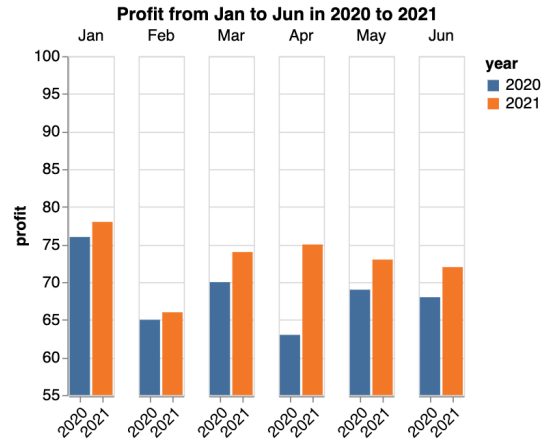


Figure 4: Sample question from the experiment. The participant is asked to estimate the ratio between the profit in 2020 and the profit in 2021 for April. Note that the y-axis is truncated to start at 55.

4.1 Judgment tasks

We used four judgment tasks with varying levels of cognitive and perceptual demand, following recent studies on how tasks affect user performance [26, 32, 33]:

- (1) Compare Gaps – “Which month has a greater difference in profit between 2020 and 2021, [month A] or [month B]?”
- (2) Compare Ratios – “Which month shows a greater percentage change from 2020 to 2021, [month A] or [month B]?”

- (3) Estimate Gap – “What is the difference between the profit in 2020 and the profit in 2021 for [month]?”
- (4) Estimate Ratio – “What is the ratio between the profit in 2020 and the profit in 2021 for [month]?”

For Estimate Gap, we asked for an integer between 1 to 100 (measured in units). For Estimate Ratio, we asked for an integer between 1 to 100 (measured in percentage points). This selection of tasks covers some common tasks used in prior graphical perception research [10], yet is different from prior work on y-axis truncation [12, 30, 31, 44, 45] that elicited responses using Likert items.

4.2 Experiment stimuli and data generation

We generated visual stimuli using Vega-Altair [40] and hosted them on an online survey created via the Qualtrics platform. We framed our questions in the context of comparing profits between 2020 and 2021 for the first six months of the year to simulate a real-world setting where grouped bar charts are used.

4.2.1 Data-generating distribution. Existing research [26, 33] has shown that data impacts task effectiveness, yet prior work on y-axis truncation either did not specify how their data was generated [12, 30] or used only normal distributions [44]. To test if different data-generating distributions have different impacts on y-axis truncations, we selected two distributions with bounded support between 0 and 100:

- (1) We used a **scaled Beta distribution** with parameters $\alpha = 16, \beta = 6$ as an approximation for data that is roughly normally distributed. We thought it necessary to include this data-generating distribution since (approximately) normal distributions are among the most commonly seen distributions in the wild.
- (2) We used a **truncated Pareto distribution** with parameters $b = \log_4(5), c = 5, \text{scale} = 20$ to model real-life scenarios where we have outliers in the data.

We sampled data using the Python packages numpy [20] and scipy [41].

4.2.2 Data relation. As detailed in subsection 3.1, we have four different kinds of data relations grouped into two categories: **concordant** and **discordant** data relations. For any data set, there are four potential judgment tasks that it could answer. The comparison tasks require two months to examine, and these are selected by going through all $\binom{6}{2} = 15$ pairs of months, finding the pairs that satisfy the data relation we want to test, and selecting the pair that has the minimum absolute difference between their gaps, i.e., the pair that has minimum $|\Delta A - \Delta B|$. The estimation tasks require one month to examine, and we randomly selected either the month that has the biggest percentage change or the biggest difference between the two pairs of bars. We did not control for the distance between the two pairs of bars – they could be next to each other or separated by four other pairs of bars. We adjusted the ordering of bars such that for the two months for comparison, within each month, the left bar is always the shorter bar.

4.2.3 Type of truncation. As detailed in subsection 3.2, there are three types of allowed truncation: No truncation, monotonic truncation, and non-monotonic truncation. The actual amount of truncation was *randomly sampled* from the interval of monotonic truncations and the interval of non-monotonic truncations respectively (Figure 1). This differs from prior work where the authors either do not specify how they chose the truncation value [30] or start the y-axis from fixed values [12].

4.3 Experiment Procedure

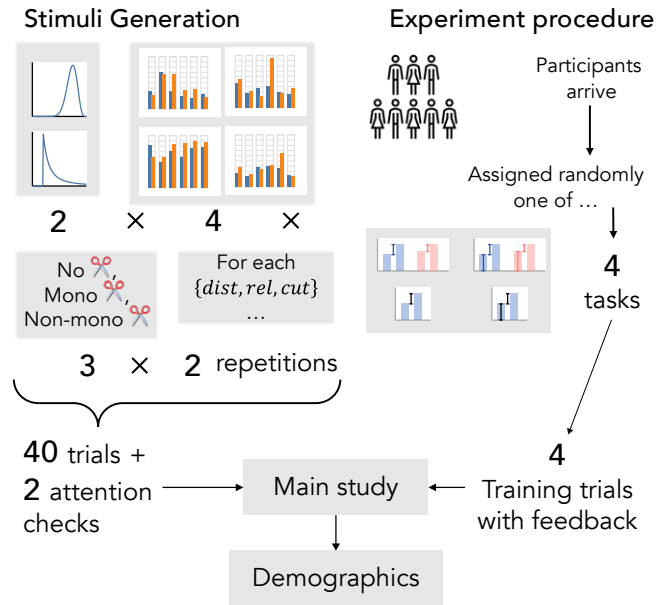


Figure 5: Overview of experiment procedure. Participants were first randomly assigned to one of four tasks, then to the same set of training stimuli in the same order, and then to the same set of 20 conditions, but different testing stimuli. Observe that $2 \times 4 \times 3 \times 2 = 48 \neq 40$ trials – this is because it is impossible to truncate non-monotonically for the two discordant data relations, thus resulting in $2 \times 4 \times 3 \times 2 - 2 \times 2 \times 1 \times 2 = 48 - 8 = 40$ trials.

- **Introduction and Consent.** Participants began by reading about the study goals and compensation. Participants were instructed to not rely on any rulers and to refrain from taking long breaks while answering the questions.
- **Training.** Participants were given 4 practice trials. All participants were assigned the same set of training stimuli, displayed to them in the same order. After answering, participants will see their responses displayed alongside the actual answers for the trial. For example, a participant assigned the task Estimate Ratio will see “You have entered [...] as your answer. The actual ratio between the profit in 2020 and the profit in 2021 for [month] is [...]%.”
- **Testing.** Participants were given 42 testing trials, 2 of which were attention checks. The first attention check was inserted randomly into the first 20 trials, and the second attention

check was inserted randomly into the remaining 20 trials. The attention check stimuli were the same for all participants.

- **Demographics and optional written feedback.** Participants were asked to provide demographic information on their age and gender. Participants were also asked to provide optional feedback on the strategy they used for their assigned perceptual task, as well as whether they followed the instructions in the survey.

4.4 Participants and exclusion criteria

We recruited participants in an IRB-approved study on Prolific.co who are between the ages of 18 to 65, have a normal or corrected-to-normal vision, speak fluent English, and reside in the United States. The study was distributed to a balanced sample of participants.

We excluded participants who failed to correctly answer both of the attention check questions for comparison tasks and participants who provided estimates that were more than 5 units away from the actual answer for estimation tasks. We also excluded one participant who answered "no" to the question in the feedback about whether they followed the instructions or not. Due to a data collection issue, we had two pairs of participants who were assigned to the same task and the same stimuli. We excluded the participants who started the survey later from our analysis.

In total, we collected data from 136 participants, and after excluding 11 participants for failing the attention checks, 1 participant for not following the instructions, and 2 participants for data collection issues, we had a total of 122 participants for analysis. Successful participants were compensated with an average reward of \$17.02 / hour, exceeding the US minimum wage. The median completion time was 17 minutes.

4.5 Analysis approach

We used multi-level Bayesian regression to analyze our results. Specifically, we used **logistic regression** for the comparison tasks and **hurdle log-normal regression** for estimation tasks. The hurdle log-normal distribution is a modified log-normal distribution that allows zeros by modeling the probability of zero as a separate process. For computing and presenting the results, we used the following R packages: `brms` [7], `tidybayes` [25], `rstan` [35] and `tidyverse` [43].

All of our models are random effects models. Due to the nature of the small sample size we collected for each task, we believe it is best to be conservative in our modeling approach, and thus adopted random effects to maximize partial pooling of information with the amount of data we had [19].

Let $PID[i]$ denote the participant ID associated with the i -th observation. Let $cut:dist:rel[i]$ denote the combination of the type of truncation, data-generating distribution, and data relation for the i -th observation. For the two **comparison tasks**, let y denote the binary dependent variable and π denote the probability of success. Then, we have the following model:

$$y \sim \text{Bernoulli}(\pi)$$

$$\text{logit}(\pi) = \bar{\beta} + \beta_{PID[i]} + \beta_{cut:dist:rel[i]}$$

The corresponding brms formula is `correct ~ 1 + (1|pid)+ (1|rel:dist:cut)`, `family = bernoulli`.

For the two **estimation tasks**, let $y_i := |\text{response}_i - \text{answer}_i|$ denote the numerical response variable, i.e., y_i is the absolute error for the i -th observation. Let π denote the probability of zero. Then, we have the following model:

$$y \sim \text{Hurdle Log-Normal}(\mu, \sigma, \pi)$$

$$\mu = \bar{\beta} + \beta_{PID[i]} + \beta_{cut:dist:rel[i]}$$

$$\text{logit}(\pi) = \bar{\beta} + \beta_{PID[i]} + \beta_{cut:dist:rel[i]}$$

The corresponding brms formula is `bf(diff ~ 1 + (1|pid)+ (1|cut:rel:dist))`, `hu ~ 1 + (1|pid)+ (1|cut:rel:dist)`, `family = hurdle_lognormal`.

4.5.1 Priors. We used the following priors for all of our models:

- $\bar{\beta} \sim \text{Normal}(0, 1.5)$
- $\beta_{PID} \sim \text{Normal}(0, \sigma_{PID})$
- $\beta_{cut:dist:rel} \sim \text{Normal}(0, \sigma_{cut:dist:rel})$
- $\sigma_{PID}, \sigma_{cut:dist:rel}, \sigma \sim \text{Exponential}(1.5)$

5 RESULTS

5.1 Data preliminaries

Among the 122 participants, there were 58 female participants (47.5%), 59 male participants (48.4%), and 5 (4%) non-binary participants. The self-reported age of participants ranges from 20 to 60, with a median age of 32 and a mean age of 35.31. The average time (in seconds) for a participant to answer a question for each task, ordered by magnitude, is Compare Gaps (13.76) < Estimate Gap (13.86) < Compare Ratios (17.99) < Estimate Ratio (20.55).

5.2 Gap

Task: Compare Gap

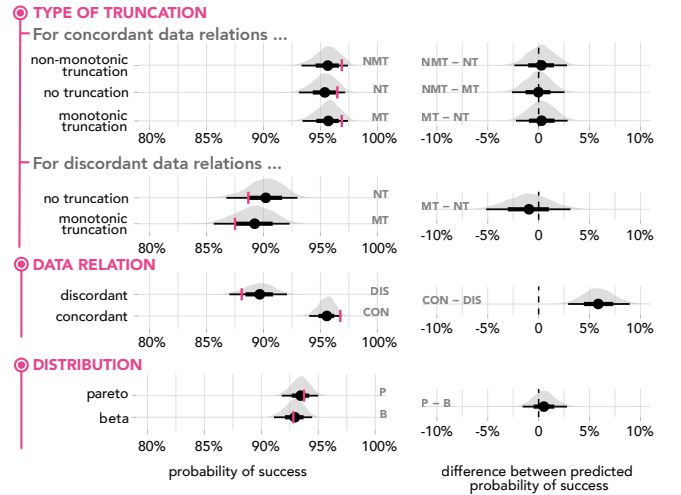


Figure 6: *Left:* Posterior density, median, 66% and 95% quantile interval of the probability of correctly answering. *Right:* Expected difference in the means between different values of a variable. **Pink ticks** represent the percentage of correct response answers.

The 32 participants performed well for Compare Gap (Figure 6).

We found no significant difference in performance between the different types of truncation: for concordant data relations, the difference between non-monotonic truncation and no truncation has a mean of 0.26% with 95% CI $[-1.89\%, 2.39\%]$, and the difference between monotonic truncation and no truncation has a mean of 0.29% with CI $[-1.83\%, 2.41\%]$. For discordant data relations, the difference between monotonic truncation and no truncation has a mean of -0.99% with CI $[-4.53\%, 2.52\%]$.

We also found no significant difference in performance between a Beta distribution and a truncated Pareto distribution, given that the difference in the means between the two is 0.53% (CI $[-1.25\%, 2.38\%]$).

We did find a difference between performance for discordant data relations and concordant data relations: on average, concordant data relations lead to a higher probability of correctness compared to discordant relations, with a difference of mean 5.88% (CI $[8.46\%, 3.41\%]$).

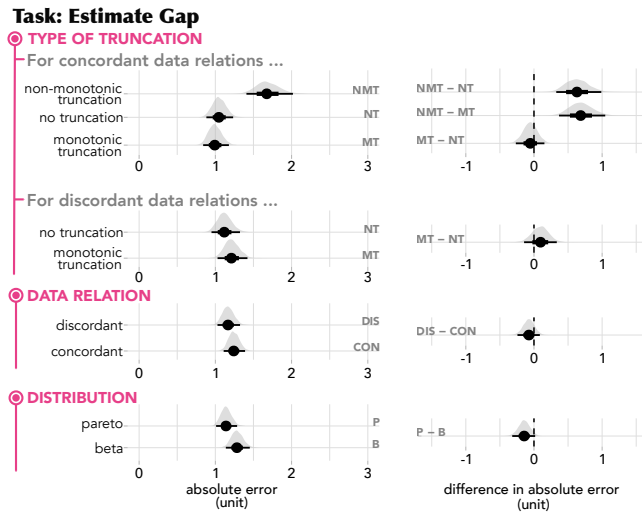


Figure 7: *Left*: Posterior density, median, 65% and 95% quantile interval of absolute error. *Right*: Expected difference in the means between different values of a variable.

The 31 participants performed fairly accurately for the task of Estimate Gap (Figure 7): 52% of responses had an absolute error of 0, and the posterior distributions show a high concentration of absolute error around 1.

For both concordant and discordant data relations, we found no significant difference in performance between monotonic truncation and no truncation: the difference in absolute error between no truncation and monotonic truncation has a mean of -0.05 (CI $[-0.23, 0.12]$) for concordant relations and a mean of 0.09 (CI $[-0.11, 0.29]$) for discordant relations.

Similarly, there is no significant difference between the two different kinds of data relations or the two different types of data-generating distributions. The difference between discordant and concordant data relations has a mean of -0.08 (CI $[-0.22, 0.06]$), and the difference between Pareto and Beta distributions has a mean of -0.14 (CI $[-0.29, -0.01]$).

However, where non-monotonic truncation is allowed (in concordant data relations), the difference in absolute error between non-monotonic truncation and no truncation has a mean of 0.635 (CI $[0.371, 0.926]$), suggesting that non-monotonic truncation causes a small exaggeration in people's estimation of gaps between groups of bars.

5.3 Ratio

Task: Compare Ratio

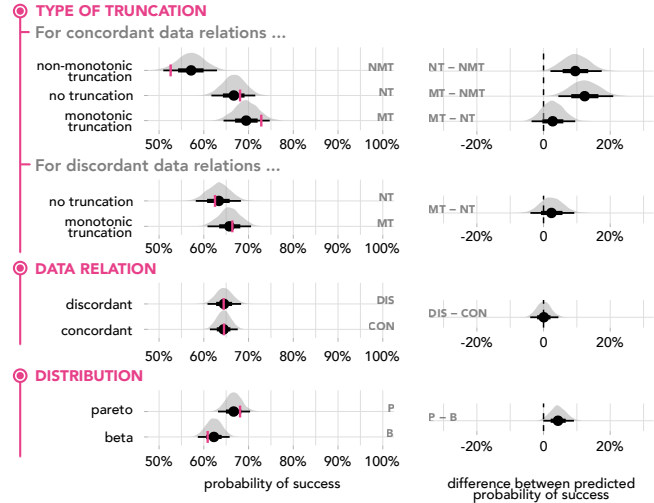


Figure 8: *Left*: Posterior density, median, 66% and 95% quantile interval of the probability of correctly answering. *Pink ticks* represent the percentage of correct response answers. *Right*: Expected difference in the means between different values of a variable. Similar to our results for Compare Gap in Figure 6, the median of posterior predictions (represented by black dots) are closer together than the raw data (represented by pink ticks), which demonstrates *model shrinkage*. This will make estimates of differences more conservative and is thus suitable for the sample size we've collected.

The 29 participants performed on average much worse in Compare Ratios than Compare Gaps (Figure 8). For both concordant and discordant relations, we found that monotonic truncations slightly improve the probability of success compared to no truncation – for concordant data relations, the difference in probability of success between monotonic truncation and no truncation for concordant data relations has a mean of 2.89% (CI $[-2.43\%, 8.38\%]$) and for discordant data relations has a mean of 2.44% (CI $[-3.18\%, 8.19\%]$).

We also found that non-monotonic truncation decreases the probability of success compared to no truncation, with the difference between the two having a mean of -12.5% (CI $[-19.5\%, -5.47\%]$).

We did not find a significant difference in performance between different kinds of data relations – between concordant and discordant relations, the difference in probability of success has a mean of 0.08% (CI $[-3.51\%, 3.65\%]$). We found a slightly higher probability of success in Pareto distributions compared to Beta distributions, with the difference between the two having a mean of 4.47% (CI $[0.83\%, 8.32\%]$).

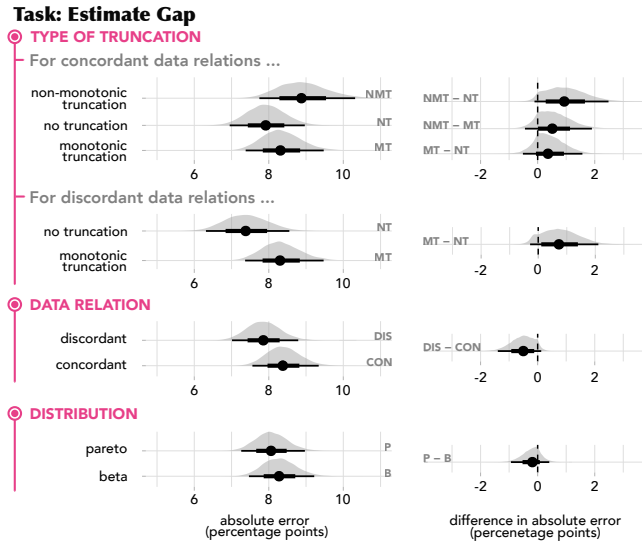


Figure 9: Left: Posterior density, median, 66% and 95% quantile interval of the absolute error for task Estimate Ratio. Right: Expected difference in the means between different values of a variable.

Previously, we saw a 13% decrease in the predicted probability of success between non-monotonic truncation and no truncation and a 3% increase between monotonic truncation and no truncation. Could this be due to how y-axis truncation alters the precision of the estimation of ratios (Figure 9)?

We found that no truncation has a smaller absolute error compared to monotonic truncation for both concordant and discordant data relations – the difference in absolute error between the two is 0.42 (CI [−0.37, 1.35]) for concordant relations and 0.94 (CI [−0.02, 2.04]) for discordant relations. We also found that no truncation has a smaller absolute error compared to non-monotonic truncation – the difference in absolute error between the two is 0.99 (CI [−0.01, 2.22]). This implies that truncation, regardless of its monotonicity, is likely to bias the estimate of percentage change upwards, but non-monotonic truncations have larger absolute errors compared to monotonic truncations.

For different data relations, we found that discordant data relations have a smaller absolute error compared to concordant data relations – the difference between the two has a mean of −0.53 (CI [−1.25, 0.05]).

We did not find much difference between the two kinds of data-generating distributions. The difference between Pareto and Beta in predicted absolute error has a mean of −0.22 (CI [−0.81, 0.29]).

5.4 Summary

At the end of subsection 3.2, we proposed two hypothesis:

- (1) Monotonic truncation would *not* impact $D \leftrightarrow \text{Bar} \leftrightarrow T$.
- (2) Non-monotonic truncation would *adversely* impact $D \leftrightarrow \text{Bar} \leftrightarrow T$.

We found that for tasks that examine a quantity whose data relation is *unaffected by truncation* (i.e., gap), **truncation has little to no effect on user performance, regardless of whether it is monotonic or not**. Specifically, for Compare Gaps, there is

essentially no difference in user performance between no truncation and monotonic truncation, as well as no truncation and non-monotonic truncation. For Estimate Gap, non-monotonic truncation leads to a larger estimated magnitude (as also seen in prior work [12, 30, 44, 45]), but the degree of increase is small.

On the other hand, for tasks that examine a quantity whose data relation is *affected by truncation* (i.e., ratio), **monotonic truncation has similar user performance as no truncation**, but both are much better than non-monotonic truncation. Specifically, for Compare Ratios, monotonic truncation slightly improves the probability of success compared to no truncation, while non-monotonic truncation decreases the probability of success. For Estimate Ratio, truncations (regardless of monotonicity) have larger absolute errors compared to no truncation, but non-monotonic truncation has a larger absolute error compared to monotonic truncation.

6 DISCUSSION

6.1 Recommendations for y-axis truncation

In a blog post [11] accompanying their paper [12], Correll categorized y-axis truncation recommendations into four groups: the Anathemists, the Line Chart Exceptionists, the Signalers, and the Libertines. We take a different approach and employ the $D \leftrightarrow V \leftrightarrow T$ model to categorize recommendations into three broad groups: those emphasizing the $D \leftrightarrow V$ correspondence, those emphasizing the $V \leftrightarrow T$ correspondence, and those that focus on either D or T .

6.1.1 $D \leftrightarrow V$ correspondence. Arguments emphasizing the $D \leftrightarrow V$ correspondence are prominently featured in the bar vs line debate [4, 8, 17, 28, 34]. People have long disagreed on whether the no truncation rule only applies to bar charts or includes line charts as well. Advocates of the “only bars” position argue that the correspondence between data and the **visual encoding** of bar charts is distorted by truncation. Since bar charts use either height/length [8, 17, 34, 47] or size [4] to encode data, starting the y-axis from a non-zero value disrupts the correspondence between bar height/length/size and the original data value. On the other hand, since line charts and dot charts use position/angle to encode information [4, 8, 34, 45], it is not necessary to hold them to the zero-baseline rule because these encodings do not get distorted [8].

6.1.2 $V \leftrightarrow T$ correspondence. While previous arguments focus on how truncation affects the $D \leftrightarrow V$ correspondence differently for different visual encodings, other arguments focus on how the **decoding** of information in line charts and bar charts could be practically the same. For example, Skelton [34] and Kosara [28] argue that people decode line charts and bar charts both by the distance of the mark from the baseline and thus a non-zero line chart poses a similar risk as a non-zero bar chart. This decoding argument is also empirically supported by Correll et al.’s finding that there is no significant difference between bars and lines when participants are decoding charts into a measure of subjective trend strength [11, 12].

Other people whose arguments emphasize the $V \leftrightarrow T$ correspondence care about **visual impression**, or **graphical perception**, which is operationalized by tasks T in $D \leftrightarrow V \leftrightarrow T$. For example, Huff [23] argues that in a truncated graph, “nothing has been falsified – except the impression that it gives” and thus suggests that all

charts should begin their y-axis from zero. Witt [44] reasons that “when the [perceived] visual size of the effect aligns with the actual size of the effect”, the reader exerts less mental effort decoding information. Witt suggests setting the y-axis range to 1.5 SDs because she found “improved sensitivity in effect size” when doing so. However, Witt did not explicitly state whether her recommendation applies to bar charts because her experiment task “did not permit measuring the spontaneous impression given by the graphs”, and this impression could be either the *differences* or the *ratios* between bars.

6.1.3 D or T. Other guidelines have proposed different exceptions for the no-truncation rule, focusing on either the nature of data *D* or the reader’s task/quantity of interest *T*. On the data side, a non-zero baseline is encouraged when it is meaningful or shows small but meaningful changes in the data [8, 34]. On the task side, it depends on whether the visualization is communicating absolute change or relative change [5, 34]. On one hand, Brinton [5] suggests that time-series line charts do not need to include the zero baselines when the reader is interested in “the absolute amount of change rather than ... the relative amount of change”. On the other hand, Skelton [34] argues that for comparisons between the relative rate of increase/decrease on the line chart, baseline zero is irrelevant, and for “communicating the actual rate of increase/decrease ..., baseline zero can be very important (and its absence potentially misleading).”

6.1.4 Our approach and recommendation. In our paper, we proposed the $D \leftrightarrow V \leftrightarrow T$ model, defined and examined properties of truncations that are structure-preserving transformations of the visual component of a $D \leftrightarrow V \leftrightarrow T$ correspondence, proposed hypotheses of the effects of such truncations, and designed experiments to test our hypotheses. As such, our recommendations stem from a rigorous theoretical and empirical basis. Combined with findings from prior studies [12, 44, 46], we propose the following suggestions that we believe may extend well beyond grouped bar charts:

When the designer has control over user tasks or knows with high certainty the tasks a visualization will be used for, and when the task examines quantities that are unaffected by truncation (e.g., Compare Gaps and Estimate Gap), the designer can truncate the axis to elicit the desired subjective perception of effect size or make the visuals more task-aligned, as recommended by previous work [12, 44]. However, if the tasks examine quantities that are affected by truncation (e.g., Compare Ratios and Estimate Ratio), then we suggest performing monotonic truncations or no truncations to avoid creating visual relations that are contrary to data relations.

In other scenarios **when the designer does not have control over user tasks or is uncertain about the tasks that visualizations will be used for**, then fixed truncation in a static visualization, such as tested in this paper, is likely inappropriate. However, other approaches that use interactivity to gain the potential benefits of truncation without introducing deception [31] may be effective.

Ultimately, it is the designer who faces constraints [8] and shoulders the responsibility to determine the effect size they want to show [12], but we provide a clear, empirically-verified rule on the conditions under which a designer can justify their choice for the

amount they truncate. Our recommendations highlight the importance of tasks, data relations (as revealed by the examination of structure-preserving transformations in our work) and subjective perception (as revealed by previous work [12, 31, 44, 45]) to making an informed decision about when and how to truncate the y-axis.

6.2 Limitations

Our research did not center on the cognitive aspect of graphical perception. We did not investigate the causes of deception, nor did we examine the relationship between inattention and deceptive visualizations. Additionally, we did not analyze how *framing*, *priming*, or *anchoring* impacts the graphical perception of bar charts.

There are many interesting directions as potential follow-ups to our work. A natural extension is to examine other tasks and see if we can offer structure-preserving transformations for those tasks. Additionally, we could investigate the feasibility of providing these transformations to tasks that involve subjective perception of bar charts rather than numerical perception. A third extension is to impose a time limit on tasks and see if our results still hold while participants are more time-constrained. With more data, we could also characterize the Pareto-frontier of tradeoffs for different tasks. Finally, we could empirically investigate the just noticeable difference (JND) for the perception of percentage change and investigate its alignment with mathematical concepts of monotonicity.

7 CONCLUSION

In this paper, we systematically examined the controversial visualization technique, y-axis truncation, applied to grouped bar charts. We find that for simpler tasks that examine visual quantities whose data relations are unaffected by truncation, such as comparing or estimating differences between bars, y-axis truncation has minimal impact on user performance. However, for tasks that examine visual quantities whose data relations are affected by truncation, such as comparing or estimating percentage changes between bars, truncating the y-axis, specifically non-monotonically truncating the y-axis, worsens user performance. We provide suggestions for when designers can leverage previous recommendations on y-axis truncation regarding the subjective perception of effect size without compromising judgment accuracy.

ACKNOWLEDGMENTS

The authors would like to thank the HCI group and the theory group at Northwestern University for providing thoughtful feedback during every stage of this project.

REFERENCES

- [1] James F. Allen. 1983. Maintaining Knowledge about Temporal Intervals. 26, 11 (1983), 832–843. <https://doi.org/10.1145/182.358434>
- [2] Robert Amar, James Eagan, and John Stasko. 2005. Low-Level Components of Analytic Activity in Information Visualization. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization (USA, 2005-10-23) (INFOVIS '05)*. IEEE Computer Society, 15. <https://doi.org/10.1109/INFOVIS.2005.24>
- [3] Michael Behrisch, Michael Blumenschein, Nam Wook Kim, Lin Shao, Mennatallah El-Assady, Johannes Fuchs, Daniel Seebacher, Alexandra Diehl, Ulrik Brandes, Hanspeter Pfister, et al. 2018. Quality metrics for information visualization. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 625–662.
- [4] Carl Bergstrom and Jevin West. 2017. Calling bullshit: Proportional ink. https://callingbullshit.org/tools/tools_proportional_ink.html
- [5] Willard Cope Brinton. 1939. *Graphic presentation*.

- [6] Philip Bump. 2015. Why This National Review Global Temperature Graph Is so Misleading. (2015). <https://www.washingtonpost.com/news/the-fix/wp/2015/12/14/why-the-national-reviews-global-temperature-graph-is-so-misleading/>
- [7] Paul-Christian Bürkner. 2017. Brms: An R Package for Bayesian Multilevel Models Using Stan. 80, 1 (2017), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- [8] Alberto Cairo. 2019. *How charts lie: Getting smarter about visual information*. WW Norton & Company.
- [9] William S. Cleveland, Marylyn E. McGill, and Robert McGill. 1988. The Shape Parameter of a Two-Variable Graph. 83, 402 (1988), 289–300. <https://doi.org/10.1080/01621459.1988.10478598>
- [10] William S Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. 79, 387 (1984), 531–554. <https://doi.org/10.2307/2288400>
- [11] Michael Correll. 2020. *Truncating the Y-Axis: Threat or Menace?* Medium. <https://engineering.tableau.com/truncating-the-y-axis-threat-or-menace-d0bce66d4d08>
- [12] Michael Correll, Enrico Bertini, and Steven Franconeri. 2020. Truncating the Y-Axis: Threat or Menace? In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 1–12. <https://doi.org/10.1145/3313831.3376222>
- [13] Michael Correll and Jeffrey Heer. 2017. Black Hat Visualization. In *Workshop on Dealing with Cognitive Biases in Visualisations (DECISIVE), IEEE VIS*. <http://idl.cs.washington.edu/papers/blackhatvis>
- [14] Michael Correll, Mingwei Li, Gordon Kindlmann, and Carlos Scheidegger. 2019. Looks Good To Me: Visualizations As Sanity Checks. 25, 1 (01 2019), 830–839. <https://doi.org/10.1109/TVCG.2018.2864907>
- [15] Aritra Dasgupta and Robert Kosara. 2010. Pargnostics: Screen-Space Metrics for Parallel Coordinates. 16, 6 (2010), 1017–1026. <https://doi.org/10.1109/TVCG.2010.184>
- [16] Çağatay Demiralp, Carlos E. Scheidegger, Gordon L. Kindlmann, David H. Laidlaw, and Jeffrey Heer. 2014. Visual Embedding: A Model for Visualization. 34, 1 (2014), 10–15. <https://doi.org/10.1109/MCG.2014.18>
- [17] Stephen Few. 2004. Show me the numbers. *Analytics Pres* (2004).
- [18] Martin Fink, Jan-Henrik Haunert, Joachim Spoerhase, and Alexander Wolff. 2013. Selecting the Aspect Ratio of a Scatter Plot Based on Its Delaunay Triangulation. 19, 12 (2013), 2326–2335. <https://doi.org/10.1109/TVCG.2013.187>
- [19] Andrew Gelman, Jennifer Hill, and Masanao Yajima. 2012. Why We (Usually) Don't Have to Worry About Multiple Comparisons. 5, 2 (04 2012), 189–211. <https://doi.org/10.1080/19345747.2011.618213>
- [20] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. 2020. Array programming with NumPy. *Nature* 585, 7825 (Sept. 2020), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- [21] Jeffrey Heer and Maneesh Agrawala. 2006. Multi-Scale Banking to 45 Degrees. 12, 5 (2006), 701–708. <https://doi.org/10.1109/TVCG.2006.163> PMID:17080790
- [22] Kevin Hu, Snehal Kumar 'neil' S Gaikwad, Madelon Hulsebos, Michiel A Bakker, Emanuel Zraggen, César Hidalgo, Tim Kraska, Zhiqiang Li, Arvind Satyanarayan, and Çağatay Demiralp. 2019. VizNet: Towards A Large-Scale Visualization Learning and Benchmarking Repository. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019-05) (CHI '19). Association for Computing Machinery, 1–12. Issue Paper 662. <https://doi.org/10.1145/3290605.3300892>
- [23] Darrell Huff. 2023. *How to lie with statistics*. Penguin UK.
- [24] Alex Kale, Matthew Kay, and Jessica Hullman. 2021. Visual Reasoning Strategies for Effect Size Judgments and Decisions. 27, 2 (2021), 272–282. <https://doi.org/10.1109/TVCG.2020.3030335> PMID:33048681
- [25] Matthew Kay. 2023. *tidybayes: Tidy Data and Geoms for Bayesian Models*. <https://doi.org/10.5281/zenodo.1308151> R package version 3.0.6.
- [26] Younghoon Kim and Jeffrey Heer. 2018. Assessing Effects of Task and Data Distribution on the Effectiveness of Visual Encodings. 37, 3 (2018), 157–167. <https://doi.org/10.1111/cgf.13409>
- [27] Gordon Kindlmann and Carlos Scheidegger. 2014. An Algebraic Process for Visualization Design. 20, 12 (2014), 2181–2190. <https://doi.org/10.1109/TVCG.2014.2346325>
- [28] Robert Kosara. 2013. *Continuous Values and Baselines*. <https://eagereyes.org/blog/2013/baselines>
- [29] Andrew McNutt. 2021. What Are Table Cartograms Good for Anyway? An Algebraic Analysis. 40, 3 (2021), 61–73. <https://doi.org/10.1111/cgf.14289>
- [30] Anshul Vikram Pandey, Katharina Rall, Margaret L Satterthwaite, Oded Nov, and Enrico Bertini. 2015. 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* (New York, NY, USA, 2015-04) (CHI '15). Association for Computing Machinery, 1469–1478. <https://doi.org/10.1145/2702123.2702608>
- [31] Jacob Ritchie, Daniel Wigdor, and Fanny Chevalier. 2019. A Lie Reveals the Truth: Quasimodes for Task-Aligned Data Presentation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow Scotland Uk, 2019-05-02). ACM, 1–13. <https://doi.org/10.1145/3290605.3300423>
- [32] Bahador Saket, Alex Endert, and Çağatay Demiralp. 2019. Task-Based Effectiveness of Basic Visualizations. 25, 7 (2019), 2505–2512. <https://doi.org/10.1109/TVCG.2018.2829750> PMID:29994001
- [33] Alper Sarikaya and Michael Gleicher. 2018. Scatterplots: Tasks, Data, and Designs. 24, 1 (2018), 402–412. <https://doi.org/10.1109/TVCG.2017.2744184>
- [34] Chad Skelton. 2018. Bar charts should always start at zero. but what about line charts? <http://www.chadskelton.com/2018/06/bar-charts-should-always-start-at-zero.html>
- [35] Stan Development Team. 2023. RStan: the R interface to Stan. <https://mc-stan.org/R-package-version-2.21.8>.
- [36] Justin Talbot, John Gerth, and Pat Hanrahan. 2011. Arc Length-Based Aspect Ratio Selection. 17, 12 (2011), 2276–2282. <https://doi.org/10.1109/TVCG.2011.167>
- [37] Justin Talbot, Vidya Setlur, and Anushka Anand. 2014. Four Experiments on the Perception of Bar Charts. 20, 12 (2014), 2152–2160. <https://doi.org/10.1109/TVCG.2014.2346320> PMID:26356929
- [38] Melanie Tory and Torsten Moller. 2004. Rethinking visualization: A high-level taxonomy. In *IEEE symposium on information visualization*. IEEE, 151–158.
- [39] Edward R Tufte. 2001. *The visual display of quantitative information*. Vol. 2. Graphics press Cheshire, CT.
- [40] Jacob VanderPlas, Brian Granger, Jeffrey Heer, Dominik Moritz, Kanit Wongsuphasawat, Arvind Satyanarayan, Eitan Lees, Ilia Timofeev, Ben Welsh, and Scott Sievert. 2018. Altair: Interactive statistical visualizations for python. *Journal of open source software* 3, 32 (2018), 1057.
- [41] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, Ilhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17 (2020), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- [42] Yunhai Wang, Zeyu Wang, Lifeng Zhu, Jian Zhang, Chi-Wing Fu, Zhonglin Cheng, Changhe Tu, and Baoquan Chen. 2018. Is There a Robust Technique for Selecting Aspect Ratios in Line Charts? 24, 12 (12 2018), 3096–3110. <https://doi.org/10.1109/TVCG.2017.2787113>
- [43] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemond, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4, 43 (2019), 1686. <https://doi.org/10.21105/joss.01686>
- [44] Jessica K Witt. 2019. Graph Construction: An Empirical Investigation on Setting the Range of the y-Axis. (2019).
- [45] Brenda W. Yang, Camila Vargas Restrepo, Matthew L. Stanley, and Elizabeth J. Marsh. 2021. Truncating Bar Graphs Persistently Misleads Viewers. 10, 2 (06 2021), 298–311. <https://doi.org/10.1016/j.jarmac.2020.10.002>
- [46] Fumeng Yang, Lane T. Harrison, Ronald A. Rensink, Steven L. Franconeri, and Remco Chang. 2019. Correlation Judgment and Visualization Features: A Comparative Study. 25, 3 (2019), 1474–1488. <https://doi.org/10.1109/TVCG.2018.2810918>
- [47] Nathan Yau. 2015. *Bar Chart Baselines Start at Zero*. FlowingData. <https://flowingdata.com/2015/08/31/bar-chart-baselines-start-at-zero/>
- [48] Jeff Zacks, Ellen Levy, Barbara Tversky, and Diane J. Schiano. 1998. Reading Bar Graphs: Effects of Extraneous Depth Cues and Graphical Context. 4, 2 (06 1998), 119–138. <https://doi.org/10.1037/1076-898X.4.2.119>
- [49] Jeff Zacks and Barbara Tversky. 1999. Bars and Lines: A Study of Graphic Communication. 27, 6 (1999), 1073–1079. <https://doi.org/10.3758/bf03201236> PMID:10586582
- [50] Mingqian Zhao, Huamin Qu, and Michael Sedlmair. 2019. Neighborhood Perception in Bar Charts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019-05-02) (CHI '19). Association for Computing Machinery, 1–12. <https://doi.org/10.1145/3290605.3300462>

A MATHEMATICAL CALCULATIONS

For discordant data relation $\Delta A > \Delta B, \Delta A/A_1 < \Delta B/B_1$:

PROOF. Suppose for contradiction that truncation $t > 0$ is non-monotonic, i.e., t reverses the data relation. This implies that

$$\begin{aligned} \frac{\Delta A}{A_1 - t} &> \frac{\Delta B}{B_1 - t} \\ \underbrace{(\Delta B - \Delta A)}_{<0} t &> \underbrace{\Delta B A_1 - \Delta A B_1}_{>0} \\ t &< \underbrace{\frac{\Delta B A_1 - \Delta A B_1}{\Delta B - \Delta A}}_{<0} \end{aligned}$$

given by data relation

Since truncation can not be negative, we conclude that it is impossible to have a non-monotonic truncation on data relation $\Delta A > \Delta B, \Delta A/A_1 < \Delta B/B_1$. \square

Similarly, for discordant data relation $\Delta A < \Delta B, \Delta A/A_1 > \Delta B/B_1$:

PROOF. Suppose for contradiction that truncation $t > 0$ is non-monotonic, i.e., t reverses the data relation. This implies that

$$\begin{aligned} \frac{\Delta A}{A_1 - t} &< \frac{\Delta B}{B_1 - t} \\ \underbrace{(\Delta B - \Delta A)}_{>0} t &< \underbrace{\Delta B A_1 - \Delta A B_1}_{<0} \\ t &< \underbrace{\frac{\Delta B A_1 - \Delta A B_1}{\Delta B - \Delta A}}_{<0} \end{aligned}$$

given by data relation

Since truncation can not be negative, we conclude that it is impossible to have a non-monotonic truncation on data relation $\Delta A < \Delta B, \Delta A/A_1 > \Delta B/B_1$. \square