Taking Truncation to Task

A Task-Based Exploration of Axis Truncation in Bar Charts

Oen G McKinley
Department of Computer Science and Engineering
Washington University in St. Louis
St. Louis, Missouri, USA
m.oen@wustl.edu

Alvitta Ottley
Department of Computer Science and Engineering
Washington University in St. Louis
St. Louis, Missouri, USA
alvitta@wustl.edu

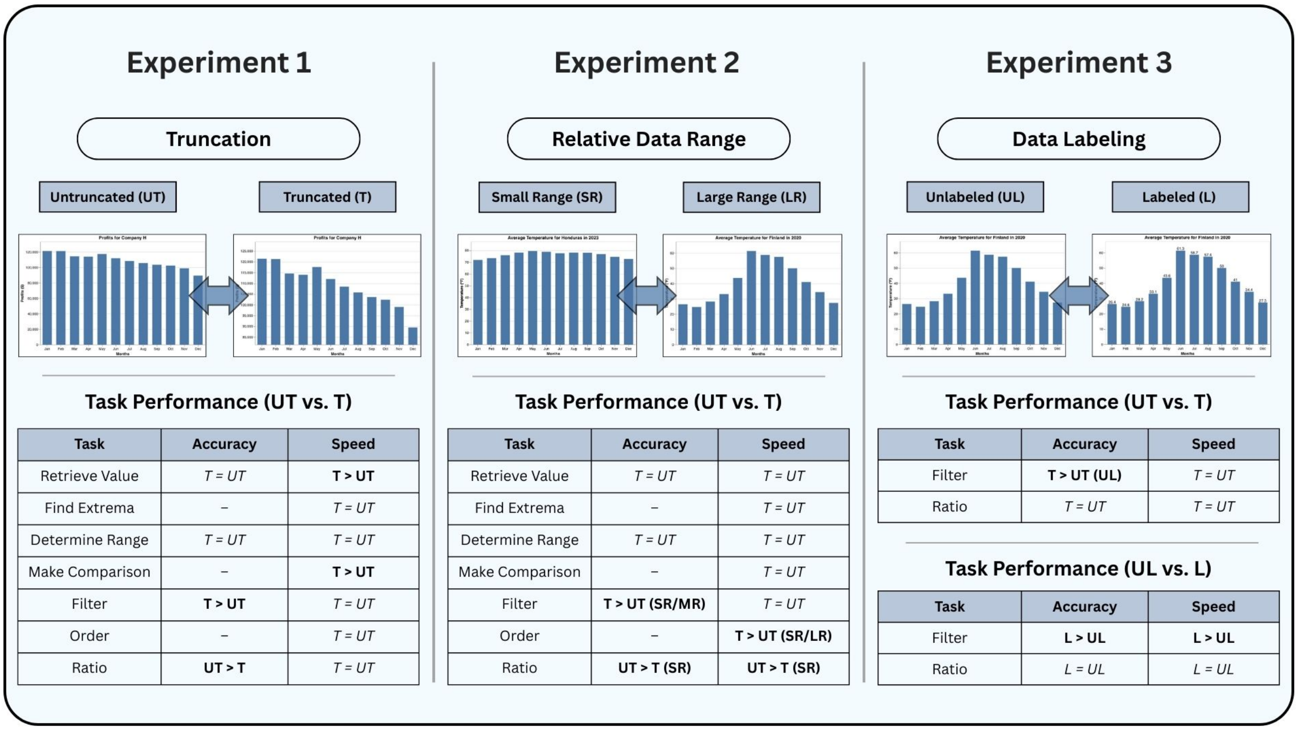


Figure 1: Our 3 within-subjects user experiments presented participants with collections of tasks (see Table 1). Experiment 1 compared performance across truncated vs. untruncated charts, resulting in differences in both accuracy (for Filter and Ratio tasks) and speed (for Retrieve Value and Make Comparison tasks). Experiment 2 compared performance across truncated vs. untruncated charts for small, medium, and large "data ranges". This once again revealed differences in both accuracy (for Filter and Ratio tasks) and speed (for Order and Ratio tasks) that differed by data range. Finally, Experiment 3 investigated data labels as a mitigation strategy, finding that labeled charts performed better in terms of accuracy and speed, while there was no difference in the Ratio task accuracy for both labeled and unlabeled charts, implying that the mitigatory effect of labeling extended to unlabeled charts for the within-subjects design.

Authors' Contact Information: Oen G McKinley, Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, Missouri, USA, m.oen@wustl.edu; Alvitta Ottley, Department of Computer Science and Engineering, Washington University in St. Louis, St. Louis, Missouri, USA, alvitta@wustl.edu.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3790617

Abstract

Axis truncation in bar charts is widely criticized as misleading, often based on ratio judgments or subjective ratings (i.e., Likert scale comparisons). This perspective, however, overly relies on these tasks and lacks nuance. We conducted three experiments to examine the effects of truncation across seven bar chart tasks. Our results show that truncation increases error for ratio calculations but improves accuracy or speed for tasks such as filtering and value retrieval. We further find that the magnitude of these effects depends on the degree of truncation and that direct data labeling substantially

mitigates the negative effects of truncation in our experimental setting. These findings add nuance to bar chart truncation and invite discussion around the inherent "deceptiveness" of design elements.

CCS Concepts

• Human-centered computing → Visualization design and evaluation methods; **Empirical studies in visualization**; **Empirical studies in HCI**.

Keywords

Data Visualization, Truncation, Bar Chart, Task, Axis, Design, Guideline

ACM Reference Format:
Oen G McKinley and Alvitta Ottley. 2026. Taking Truncation to Task: A Task-Based Exploration of Axis Truncation in Bar Charts. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, Article 111, 16 pages. https://doi.org/10.1145/3772318.3790617

1 Introduction

As data visualization has become increasingly ubiquitous, both researchers and practitioners develop guidelines to improve chart clarity [8, 36, 44, 52, 57]. These guidelines help designers communicate clearly, but are often simplified and condensed, leading to a lack of nuance for when the rules should or should not apply. This not only leads to incomplete information for designers who may benefit from understanding the tradeoffs or caveats to different designs, but also stifles designer creativity and leads to less diverse designs.

One prevalent example of this phenomenon is bar chart axis truncation, where the baseline of a continuous axis is above 0. Figure 1 shows how drastically this can change a chart's appearance. Prior work has shown that this truncation, especially in bar charts, can lead to larger perceived disparities between data points and an inflated assessment of comparisons or ratios [5, 12, 41, 46, 60]. Due to this increased scrutiny, we have seen a strong rejection of bar chart axis truncation from both academic research [5, 12, 39, 40, 43, 46, 56, 60] and industry practitioners [14–32].

There are certainly problems with axis truncation in bar charts, and in certain situations, it can lead to worse readability. However, modern guidelines are unnuanced, leading to an opportunity cost in situations where truncating the axis may be beneficial. This can mean that, while trying to follow these guidelines, designers may in fact make their visualizations less clear or readable. In use cases like healthcare or emergency response, this readability and communication can be critical.

Just as prior works have added nuance to the rejection of so-called "chart-junk" [1, 35, 42], and have even shown strong benefits to certain kinds of embellishments or visual flares, this work seeks to add much-needed nuance to bar chart truncation. We do so by taking a task-based approach to measuring the effects of truncation. We conducted a series of three user experiments, which measured the effects of truncation on various tasks, compared different data ranges, and finally leveraged data labeling as a potential mitigating factor. Our findings are three-fold:

- Although our results agree with prior work that truncating the axis is associated with increased error for the Ratio task, we also find evidence of reduced error for the Filter task and potentially increased speed in the Filter, Retrieve Value, Make Comparison, and Order tasks.
- We find that the effects of truncation in our experiments, both positive and negative, vary depending on the amount of truncation. Specifically, for small data ranges, our results show more significant benefits and drawbacks.
- According to our experimental results, data labeling (i.e., directly including the data values in proximity to the bars) seems to mitigate the negative effects of truncation while reducing error and duration overall.

2 Background and Prior Work

2.1 Axis Truncation

There has been much work discussing bar chart axis truncation [4–6, 12, 40, 41, 43, 46, 56, 60]. This work most often results in the same takeaway: in the case of bar charts, truncating the axis can be deceptive, at least in the narrow case of ratio-based or subjective comparisons. While line charts are often recommended to be truncated by both researchers [5, 11, 41, 59] and practitioners [14, 15, 19, 22, 23, 26–30] to highlight the differences between values, bar charts are often highlighted as particularly problematic. This allowance for truncated line charts while still criticizing truncated bar charts is referred to by Correll et al. [5] as "line chart exceptionalism".

This arguably goes all the way back to Tufte, who first coined the "Lie Factor" [57], which is defined as a ratio of the "size of effect shown in graphic" over the "size of effect in data". Essentially, Tufte argued that the further this value was from 1, the more of a "lie" the chart displayed. As we can imagine, bar chart truncation would squarely fall into this category, where a bar that was heavily truncated would have a high Lie Factor.

Earlier empirical works also supported this. Among the first to do so were Pandey et al. [46], who asked participants to subjectively provide a Likert-scale assessment of the perceived difference between two bars in truncated and untruncated bar charts. The findings show that viewers do, indeed, perceive the difference between truncated bars to be subjectively greater than untruncated bars. Similar subjective metrics were later utilized by other works in demonstrating the dangers of truncation [5, 60]. This could potentially deceive viewers by exaggerating small differences between values.

Correll et al. [5] found evidence that this difference in subjective comparisons held true even when using design elements explicitly tailored to draw attention to the truncated axis. For example, even in the case of a "torn paper" or "bent" stimulus (where the bar appears torn or bent), truncation inflated participants' subjective assessment of the differences.

Yet more research then defined the phenomenon mathematically and grounded it in objective differences, asking participants to calculate the ratio of two values in a bar chart [12, 41, 56]. This strategy hearkened back to the Lie Factor by using ratios to measure

[*]See supplemental materials for data, analyses, examples of the stimuli, and survey materials. Pre-registration for Experiment 2 can be found here and Experiment 3 can be found here.

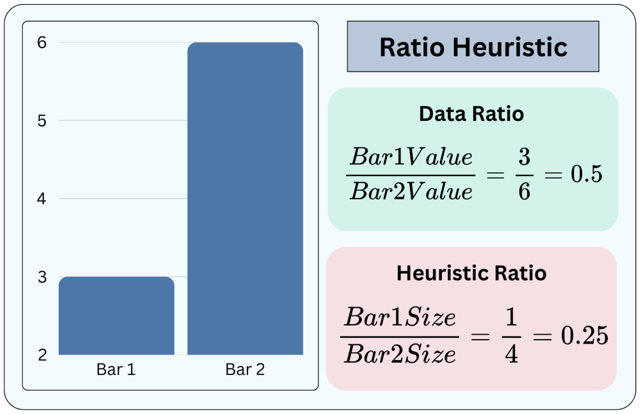


Figure 2: An example of the "ratio heuristic", where viewers assume that the length of the bar is proportional to the value of the bar and therefore estimate ratios based on the length. While this is often useful for untruncated charts, it can lead to drastically incorrect calculations in the case of axis truncation.

the effects of truncation. Such works found that viewers often utilize a heuristic to calculate the ratio based on the lengths of the bars rather than the values of the bars, due to the expectation that lengths are proportional to value [6]. We will refer to this as the "ratio heuristic" for the remainder of this work. An example is provided in Figure 2.

Researchers in the psychological or perceptual side of the data visualization research community have also contributed to this growing corpus of critical literature. For example, in the seminal work by Franconeri et al. titled "The Science of Visual Data Communication: What Works" [10], the authors refer to a real-world example of a truncated bar chart that suffers from this ratio heuristic and cite the work by Correll et al. The authors do not conclude with a hard recommendation not to truncate bar charts, but certainly caution about the dangers of bar chart truncation and this ratio heuristic. Other perceptual work, such as Ge et al. [13], refer to "inappropriate y-axis truncation" as being misleading in bar charts due to exaggerated differences, but do not opine on what "appropriate" truncation might entail.

Truncating the axis of bar charts has even been included in several compilations or taxonomies of deceptive design principles [12, 39, 40, 43, 56], and is often a guideline pushed by industry visualization designers. In fact, at least 20 of the 23 organizations that include Bar Chart-specific guidance tracked by the Data Visualization Style Guidelines project [52] advise to never truncate the axis of a bar chart, without exception [14–32]. This appears to be the most widely shared guideline among these organizations according to the Guidelines Explorer [34], which cataloged and categorized the guidelines present in these style guides. Indeed, many designers and researchers alike claim that truncating the axis of a bar chart will inherently lead to a less readable and less accurate chart for viewers.

All in all, the existing literature provides a wealth of evidence for the negative potential for bar chart truncation. This should not be discounted or ignored. However, it also leaves an opportunity to explore the nuances of when such a design may be useful. There has been a recent push within the visualization research community to question how holistically this guideline is to be applied. This includes adding nuance and showing cases where axis truncation leads to no significant difference in the viewer's ability to read the chart, even in a ratio-based task [41, 59]. Correll et al. [5] also came down on the side of nuance, in which they rejected the "line chart exceptionalist" framing or the holistic rejection of bar chart truncation. Long et al. [41] found evidence that smaller ("monotonic") bar chart truncation did not lead to negative effects in grouped bar charts. Schlieder et al. [51] explored the psychological role of metacognition in the effects of bar chart truncation, finding evidence that truncation may be mitigated by meta-cognitive learning, even when correctness of an answer is not explicitly shown. Ritchie et al. [48] further showed that, when paired with untruncated charts in an interactive setting, there were perceptual benefits to displaying truncated bar charts. But many of these studies are limited in scope, focusing on particular chart subtypes (e.g., grouped or interactive bar charts) or omitting broader considerations of when truncation may be justified. Given these gaps, we explore how a broadened space of tasks might clarify when bar chart axis truncation may hinder interpretation and readability or when it might be a defensible design choice.

2.2 Task-Based Visualization Analysis

We began by noting that specific tasks seemed to be used repeatedly within prior work: (1) subjective Likert-scale descriptions of relative differences between bars and (2) mathematical ratio calculations. This commonality led us to consider whether there are tasks with improved accuracy or speed after axis truncation. In order to determine which tasks to include, we surveyed prior works related to task taxonomy.

One of the first papers that looked at data visualizations through the lens of tasks was Shneiderman [53], who specifically created a taxonomy of seven tasks for interactive visualization systems, building on his "Visual Information Seeking Mantra". Although this work explicitly dealt with interactive systems, it nonetheless inspired works to taxonomize general visualization-related tasks. Amar et al. [2] used an affinity diagram approach to analyze student questions regarding datasets and generate a list of tasks that represent how the students seemed to explore the datasets, resulting in ten low-level visualization tasks. Yi et al. [62] similarly considered task in the case of interactive visual systems, building upon both of these works, among others. Chen et al. [61] then created a taxonomy of 12 tasks, sourced from prior works and a scrape of comments from a now-decommissioned IBM visualization-centric web forum. Another touchstone of task-based categorization came a few years later when Brehmer and Munzner [3] created an abstract taxonomy focused on the "why" and "how" of visualization communication, both for exploring and designing visualizations. Though not a comprehensive list, these works are all foundational to categorizing the ways and reasons people use visualizations.

Several works then built upon or leveraged these task-based models. Lee et al. [38], when developing their Visualization Literacy Assessment Test (VLAT), maintained a chart-type-first perspective

Table 1: Seven Canonical Bar Chart Tasks from Prior Works

Task Name	Sources	Combined Definition
Retrieve Value	[2, 38, 49]	Retrieve the y-value associated with a given x-value.
Find Extrema	[2, 38, 49]	Determine which x-values correspond to the highest or lowest y-values.
Determine Range	[2, 38, 49]	Determine the minimum and maximum y-values.
Make Comparison	[2, 38]	Determine if a y-value is higher or lower than another.
Filter	[2, 49]	Determine which x-values correspond to y-values within a specified range.
Order	[2, 49]	Sort the x-values by the corresponding y-value.
Ratio	[6, 12, 41]	Calculate the ratio of one y-value to another.

on the tasks included in the test. From prior works [2, 3, 61], they extracted 8 tasks for users to complete when looking at charts and determined (with expert input) which tasks applied to which chart types, resulting in several tasks being considered relevant for bar charts.

Saket et al. [49] also considered task, but empirically tested different chart types for each task to determine which charts performed well for different tasks. In their analysis, the authors also found that bar charts were useful in many of the tasks.

Notably, despite its prevalence in works from subsection 2.1, none of these taxonomies of visualization tasks included the Ratio task. While this could be a limitation of these taxonomies, it is also emblematic of how uncommon this task seems to be, as multiple of these taxonomies were generated by analyzing real-world users, such as students [2] or visualization enthusiasts [61], who did not seem to emphasize this task. The closest corollary is the "Compute Derived Value" task included by Saket et al. and Amar et al., though these works primarily focused on sums or differences rather than ratios [2, 49]. However, for the purpose of comparing this work to the prior findings regarding bar chart axis truncation presented in subsection 2.1, we included the Ratio task in our experimental designs.

Consolidating these works, we arrived at the following tasks: Retrieve Value, Find Extrema, Determine Range, Make Comparison, Order, Filter, and Ratio. The list of all 7 tasks, along with definitions, is presented in Table 1. The subjective differences, however, were omitted from these experiments for a key reason, as discussed by Correll et al., who note that the subjective nature of this metric means that "There is no a priori, domain-agnostic ground truth for how severe, important, or meaningful an effect size ought to be" [5]. Therefore, we agree with those authors that this decision should be ultimately left with the designer, and is not inherently "deceptive".

Table 2: Examples of Task Instructions for Temperature Charts. These wordings were used for all tasks in all experiments, with variations depending on topic (rainfall, profit, or temperature). All brackets are replaced with the relevant variable, such as the relevant month or country.

Task Name	Task Phrasing
Retrieve Value	What was the average temperature for [Country] in [Month]?
Find Extrema	What was the month with the [highest/lowest] temperature for [Country]?
Determine Range	What is the range of average temperatures for [Country]?
Make Comparison	Was the average temperature in [Month 1] higher or lower than in [Month 2]?
Filter	Which month(s) had average temperature between [Y_1] and [Y_2] (inclusive)? Select all that apply.
Order	Select the option that orders the months from [highest/lowest] to [highest/lowest].
Ratio	Fill in the blank. The average temperature in [Month 1] is _% of [Month 2].

3 Overview, Questions, and Hypotheses

From this prior work and our own judgment, we generated the following research questions regarding axis truncation in bar charts.

RQ1: Is there a difference between various visualization tasks and the effect of truncation? Are there tasks for which truncation improves speed or accuracy? We expect that, as discussed in prior work [6, 12, 41], the Ratio task will see increased error when the axis is truncated due to the use of the ratio heuristic. However, for the other tasks sourced from VLAT [38], Amar et al. [2], and Saket et al. [49], we hypothesize that there may be improvement in speed or accuracy due to the increased granularity of the axis tick marks and the clearer differences between bar heights.

RQ2: Are these effects related to the range of the underlying data? Similar to Long et al. [41], we expect that the amount of truncation (i.e., the size of the new baseline relative to the data size) will affect the results. In their work, results show that smaller ("monotonic") changes to the baseline did not lead to significant changes while larger ("non-monotonic") changes led to poorer performance on ratio-related questions. We expand this theory to hypothesize that every task with a significant difference (both positive and negative) between truncated and untruncated performance will have stronger effects as the truncation becomes more severe.

RQ3: Are there strategies to mitigate the effects of axis truncation? We hypothesize that certain strategies will help to mitigate the negative effects and lower error overall. Correll et al. [5] showed that highlighting the axis truncation did not necessarily lessen the subjective impact of truncation, but we believe that strategies which highlight data values will lead to less usage of the ratio heuristic and higher readability of the chart, leading to lower error and reduced differences between truncation.

4 Experiment 1: Task-Based Performance

We designed three experiments, one for each of these research questions. Each of the studies leveraged the reVISit framework for data collection [7] and Vega-Altair for chart generation [50, 58].

We began by conducting an experiment with the intent of answering RQ1 by measuring the task-based performance of participants when presented with truncated or untruncated bar charts as stimulus.

4.1 Stimulus Design

4.1.1 Tasks. For the first study, our goal was to answer RQ1, which concerns task-based performance. As discussed in subsection 2.2, we started by collecting the tasks from prior works [2, 12, 38, 41, 49, 56] that we believe represent common bar chart tasks, listed in Table 1. For this experiment, all seven tasks were used.

4.1.2 Visualization Designs. All charts were kept simple, with minimal variation or embellishment to avoid confounding design elements. For example, the color palette used for all charts is the default blue from Vega, and textual elements (such as titles and axis labels) were kept consistent. All text was kept minimal, with default axis ticks and descriptive, neutral titles for charts and axes (including units for measurements).

The chart designs only meaningfully varied in the data they presented (which was randomly generated for each chart) and the y-axis truncation level.

4.1.3 Truncation. For this experiment, we wanted to test whether or not truncation led to differences in task performance. We therefore wanted to have varying stimuli for truncation. We settled on a binary truncated vs. untruncated design, for multiple reasons. Firstly, this simplifies the design and allows for fewer categories in our analysis. Secondly, we will test various levels of truncation in Experiment 2 by adjusting the data range of the visualizations, and therefore can focus on a binary variable here.

To standardize the level of truncation for our stimuli, we began by once again looking at prior work, which found that the level of truncation was related to its effects on the Ratio task [41]. In other words, the more extreme the truncation, the more error was introduced in the Ratio task. We then used our judgment to determine a level of truncation that is extreme enough to elicit meaningful effects but that does not completely obscure the smallest bar. This led us to generate the following formula to determine the truncated baseline: $b = y_{min} - \frac{y_{max}-y_{min}}{4}$. This means that the data range (between the tallest and shortest bars) will take up the middle two-thirds of the axis, satisfying the requirement that truncation is extreme and also providing a simple way to standardize across datasets with different scales or ranges. Figure 3 shows two examples of charts used in this experiment, each truncated with this paradigm.

This truncation strategy also echoes prior work, most notably Witt. While our truncation makes the axis range 1.5 times the data range, Witt argued that, especially for scientific papers, the y-axis should be ranged at 1.5 standard deviations of the data [59]. There

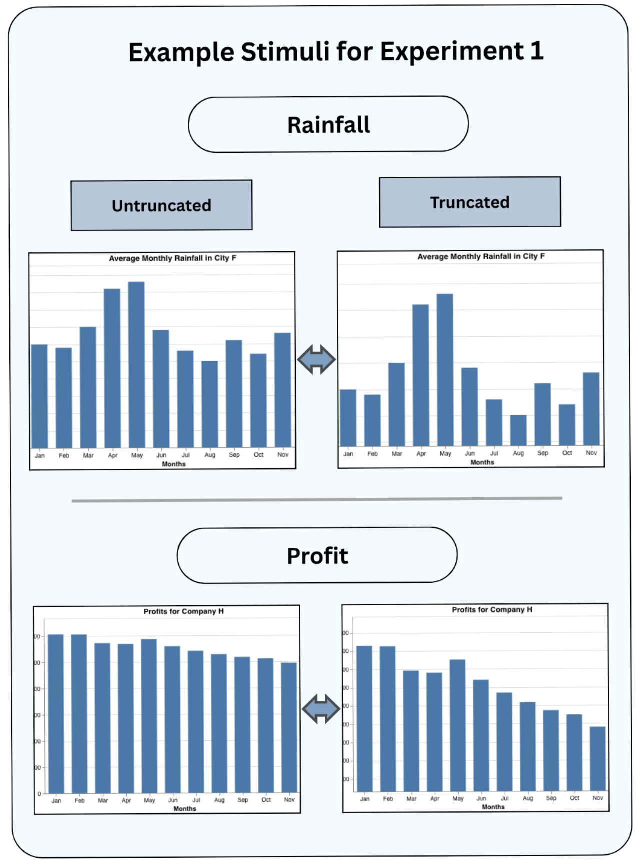


Figure 3: Two examples of charts used in Experiment 1, with both the untruncated (left) and truncated (right) versions. All data and entity IDs were randomly generated.

are also multiple industry guidelines (such as the UNHCR [31] and the Baltimore City Data Fellows [14]) who recommend, in the context of line charts, that the data take up two-thirds of the axis. From this, we refer to this truncation as the "two-thirds rule" in this work. Importantly, we do not believe this to be a hard rule, but rather a useful way to standardize our truncation.

4.2 Procedure

In Experiment 1, charts were randomly generated using the design guidelines in subsection 4.1 according to two topics, which were designed to align with common use cases for bar charts: company performance (specifically monthly profits) and weather events (specifically average monthly rainfall). All data were fabricated and the titles referred to entities named "Company X" or "City X", where X is some randomized letter from the English alphabet. Each chart had both a truncated and an untruncated version. Examples of the charts used in this experiment can be found in Figure 3.

Utilizing a within-subjects design, we presented each participant with seven pairs of such bar charts, one for each task. Each of these seven pairs were randomly chosen to either topic (rainfall and profit). The pairs of charts that were shown for each task were

Table 3: Demographic Information by Study

	Study 1 (N=62)	Study 2 (N=57)	Study 3 (N=60)
Sex Identity			
Male	32	28	31
Female	28	28	28
Non-binary or third gender	1	1	1
Prefer not to answer	1	0	0
Education Level			
Less than High School	2	0	1
High School or GED	10	10	12
Some College	16	13	8
Associate's Degree	8	8	6
Bachelor's Degree	19	21	21
Master's Degree	5	4	6
Professional Degree	0	0	1
Doctorate	2	1	5
Age			
Minimum	19	21	23
Maximum	72	62	73
Mean	42.0	43.8	45.0
Standard Deviation	12.5	10.6	12.9
Mini-VLAT [47]			
Raw Scores			
Minimum	4	2	1
Maximum	11	12	12
Mean	7.7	7.6	7.9
Standard Deviation	1.8	1.9	2.2
Corrected Scores			
Minimum	1.7	-0.3	-1.3
Maximum	11	12	12
Mean	6.8	6.1	6.5
Standard Deviation	2.3	2.5	2.8
8-item SNS [9]			
Minimum	-	15	9
Maximum	-	47	47
Mean	-	34.1	35.0
Standard Deviation	-	6.0	8.2

identical in every way except that one was truncated and the other was not. The order of all 14 charts was randomized using Latin-Square to mitigate learning effects.

Participants were instructed to answer the questions to the best of their ability and informed that their response time would be measured but that there was no time limit (aside from the Mini-VLAT, see subsubsection 4.2.1). The participants did not receive task-specific training to avoid priming them with a particular method of performing each task, which may influence their usage of the ratio heuristic.

4.2.1 Mini-VLAT and Demographic Survey. After completing the above 14 tasks, the survey design then included a Mini-VLAT test for visualization literacy [47] and a brief demographic survey.

Table 4: Results of the ANOVA tests performed for Experiment 1.

Error ANOVA

Effect	F	η_p^2	p
trunc	2.28	.036	.137
task	**44.72**	**.423**	**<.001**
trunc:task	**39.84**	**.395**	**<.001**

Duration ANOVA

Effect	F	η_p^2	p
trunc	**10.97**	**.152**	**.002**
task	**77.66**	**.560**	**<.001**
trunc:task	**5.15**	**.078**	**<.001**

Note: Box-Cox transformed; Greenhouse–Geisser corrected; Type III tests; $\alpha = 0.05$.

The Mini-VLAT, published in 2023, is a popular tool for determining the "visualization literacy" of a given user. Based on the Visualization Literacy Assessment Test (VLAT) [38], it includes 12 visualization types with one targeted question per visualization. Each question is timed and must be completed in 25 seconds, or else the user cannot input a response. Users are also allowed to skip questions if they are unsure of the answer.

The test can be used to get a raw or corrected literacy score. The raw score is simply the number of correct answers, ranging from 0 to 12. The corrected score uses a formula to correct for guessing, where users are penalized for incorrect answers, and therefore can instead range from -6 (if all answers are incorrect) to 12 (if all answers are correct).

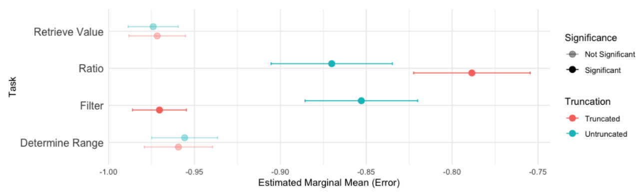
Our demographic survey then asked participants for their age, their sex, and their highest level of education. The age response was a numeric field and the sex and education responses were multiple choice, using the categories in Table 3.

4.3 Data and Analysis

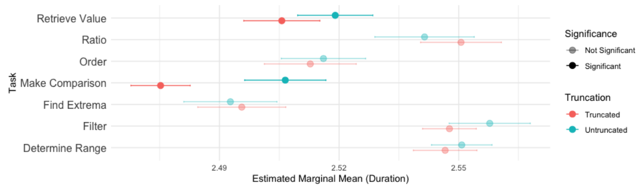
We performed power analyses for repeated-measures within-subjects 2×7 ANOVA with 14 measurements per participant, $\alpha = 0.05$, $\beta = 0.9$, and conservative estimates of nonsphericity correction $\epsilon = 0.5$ and effect size $F = 0.15$. From this calculation, our necessary sample size was 58. Using a non-response rate of 33%, we arrived at a sample size of 77, which we rounded to 80 participants.

We recruited these 80 participants from Prolific to take our survey, all of whom were US-based, English-speaking adults over the age of 18. Participants were paid $3.75 for their participation, based on an expected 15-minute survey duration and an hourly rate of $15. The actual median duration was 13 minutes, for an actual hourly rate of $17.13 per hour.

After data collection, the data consisted of responses for each question and the duration taken by each participant for each question. We began by using the responses to calculate an error score for each task. For the continuous responses (Retrieve Value, Determine Range, and Ratio), this error was calculated as the percent difference from the correct answer to control for value size. For multiple-choice responses (Find Extrema, Make Comparison, and Order), the error was a binary 1 or 0, where 1 represented incorrect



(a) Error EMM for Experiment 1



(b) Duration EMM for Experiment 1

Figure 4: The differences in Estimated Marginal Means (EMMs) for Experiment 1. The opacity of the points corresponds to the significance of the difference between truncation levels. Note that, due to the nature of the EMM calculation and the Box-Cox transformation, the x-axis values have been transformed and shifted, sometimes to negative values. Importantly, for both charts, the right direction of the x-axis means higher error or longer duration.

answers. Lastly, for the Filter task, error was calculated by how many incorrect selections were made (i.e., how many months were selected or deselected differently than the correct answer).

Participants were then removed if they had error or duration outside of 3 standard deviations from the mean to avoid any skew from extreme outliers. The result was the removal of 18 participants for a total sample size of $N = 62$. For the remaining participants' demographic breakdowns and scores for the Mini-VLAT, see Table 3.

We then conducted two 2×7 ANOVA tests using truncation level and task as a predictor of first error and then duration. As our raw data did not meet the ANOVA assumptions of normality or homoscedasticity, even after log transformations, we performed Box-Cox transformations. We also performed Greenhouse-Geisser (GG) corrections to meet assumptions of sphericity and control for the repeated measures in our design. For the error ANOVA, the binary error tasks were removed and only the continuous errors were included, though the binary error tasks were included in the duration ANOVA. Results from both ANOVA tests are presented in Table 4.

4.4 Results

As can be seen in Table 4, there were several meaningful results. Firstly, truncation alone was not a strong predictor of error ($F = 2.28$, $\eta^2 = .036$, $p = .137$), though it was a significant predictor of duration ($F = 10.97$, $\eta^2 = .152$, $p = .002$). There was also a significant interaction effect between task and truncation for both error ($F = 39.84$, $\eta^2 = .395$, $p < .001$) and duration ($F = 5.15$, $\eta^2 = .078$, $p < .001$).

We then ran Estimated Marginal Means (EMMs) on truncation by task, using a Box-Cox transformation of the data and a Holm p-value correction, to determine direction of effect (refer to Figure 4 for the results). From this test, we can see that truncation was associated with lower error for the Filter task ($t = -7.113$, $p < .001$, $d = -0.91$) and higher error for the Ratio task ($t = 4.652$, $p < .001$, $d = 0.60$).

When running the EMMs for duration, we can see significant reductions in duration for both Retrieve Value ($t = -2.869$, $p = .006$, $d = -0.37$) and Make Comparison ($t = -5.442$, $p < .001$, $d = -0.70$) when the axis was truncated. This is coupled with the lack of a significant difference in error for Retrieve Value ($t = .490$, $p = .626$, $d = 0.06$), meaning the increased speed did not come with any trade-offs. Importantly, the Make Comparison task had high accuracy (with 155/160 instances giving the correct answer) and all incorrect answers were given in untruncated charts, thus it appears to similarly not include any drawbacks to this increased efficiency.

To summarize, our participants showed significantly lower error for the Filter task, higher error for the Ratio task, and faster durations for the Retrieve Value and Make Comparison tasks. Neither of the reductions in duration were associated with any tradeoffs in accuracy.

4.5 Exploratory Demographic Analyses

4.5.1 Demographic Predictors. We performed exploratory analyses on the relationship between the error and duration for each task and the demographic variables we collected for each experiment: Age, Sex, and Education.

For Age, we analyzed both the correlation between age and error and the correlation between age and duration. None of the correlations had a Pearson's r coefficient with a magnitude greater than 0.1, and therefore we conclude that in our experiments, there were no strong relations between Age and our target variables.

For Sex and Education, we ran exploratory ANOVA tests to determine if there were any notable patterns. For both variables, there were no meaningful relationships with error, duration, or task.

Overall, we find no reason to believe that, in our experiments, Age, Sex, or Education are significant predictors of performance on these tasks or on the effect of truncation.

4.5.2 Visualization Literacy. As for Visualization Literacy (measured by the Mini-VLAT), we once again used a correlational analysis to see if there was any relationship between the literacy scores and the effects of truncation. We did find a slightly stronger correlation between the Mini-VLAT Raw Scores and duration ($r = -0.19$, $p < 0.001$), implying that participants with higher literacy scores seemed to spend slightly less time on the tasks. However, this was most pronounced in the Retrieve Value task ($r = -0.32$, $p < 0.001$), Determine Range task ($r = -0.28$, $p = 0.002$), and Ratio task ($r = -0.25$, $p = 0.004$), with other tasks seeing smaller correlational coefficients. In terms of error, we see a similarly strong correlation with the Ratio task and Mini-VLAT scores ($r = -0.22$, $p < 0.014$). This implies that there is at least some moderate correlation between literacy and performance on these tasks.

4.6 Discussion

In this experiment, we replicated prior findings that participants had significantly higher error when calculating the ratio between two bars. However, we also found that truncating the bar chart decreased error for the Filter task and lowered the duration it took participants to complete the Retrieve Value and Make Comparison tasks. This provides us with evidence that, at least in our controlled experimental setting, participants saw a mixture of results from truncation, with substantial improvement in accuracy for the Filter task and varying levels of improvement in speed for the Retrieve Value and Make Comparison tasks.

Depending on the perspective, one could interpret this as higher error on the Ratio task when the axis was truncated or as higher error on the Filter task and somewhat slower durations for the Retrieve Value and Make Comparison tasks when the baseline was restricted to 0. This indicates that a holistic rejection of bar chart axis truncation may be inadvertently missing certain benefits, at least in specific cases.

In Experiment 2, we then focused on RQ2 to test our hypothesis that the effects of truncation are related to the data range.

5 Experiment 2: Relative Data Range

We approached RQ2 by designing an experiment that presented participants with truncated or untruncated bar charts that varied in the range of the data. In other words, the longest and shortest bars were more similar (a "smaller" range) or more different (a "larger" range). In this way, we investigated the interaction between the variance of the data and the effects of truncation, hypothesizing that smaller data ranges, which result in more extreme truncation, will then cause the effects of truncation to be more pronounced.

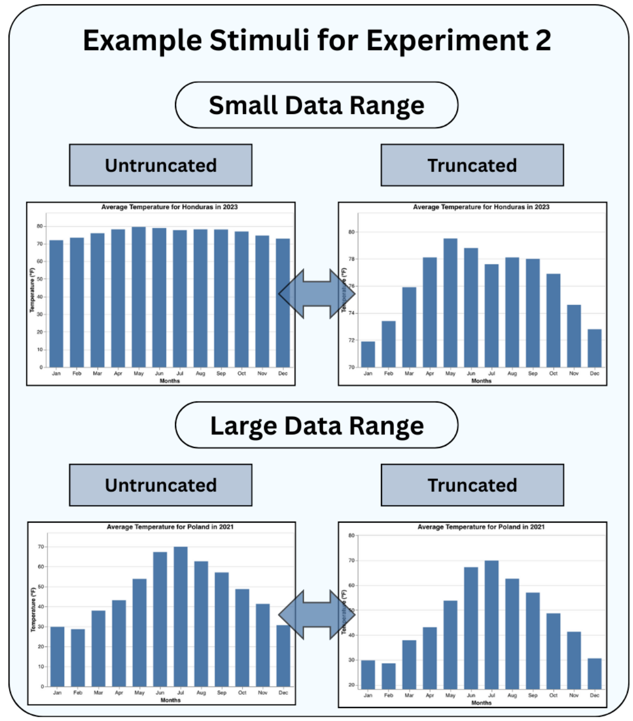


Figure 5: Examples of charts used in Experiment 2. These charts have varying truncation levels and relative data ranges. For Experiment 2, The Small (0.1) and Large (0.6) data ranges pictured here were used in both Experiment 2 and 3, and Experiment 2 also included a Medium (0.35) range, not pictured.

5.1 Stimulus Design

5.1.1 Tasks and Visualization Design. Once again, for the reasons discussed in subsection 4.1, we utilized all seven of the tasks in Table 1 to measure the effect of truncation broadly. Similarly, we utilized the same design conventions for visualizations as in Experiment 1.

5.1.2 Relative Data Range. Our second study was aimed at answering RQ2, which gets at the interactions between data range and truncation. To do this, we needed a way to standardize the data range relative to the scale of the data. We settled on calculating a percentage difference between the minimum and maximum value in the dataset, given by $\frac{y_{max}-y_{min}}{y_{max}}$. For this work, we will refer to this value as the "relative data range" for a dataset or visualization. This is also, incidentally, equivalent to one minus the ratio of the smallest to the largest bar.

For example, the minimum and maximum data values from the Small Data Range charts in Figure 5 are 71.87 and 79.52, leading to a relative data range of $RDR = \frac{79.52-71.87}{79.52} = 0.096$. Similarly, the ratio of the smallest to the largest bar is given as $Ratio = \frac{71.87}{79.52} = 0.904$.

We also utilized real-world temperature data sourced from Copernicus Climate Change Service information (2025) with major processing by Our World in Data [33], which had 5 years of monthly average temperature data from a majority of the world's countries. We chose to use temperature data due to the wide array of data ranges available and the wide use of data visualization in temperature and weather-related phenomena. All data values were converted to Fahrenheit to make the values more familiar to a US-based audience, but no other transformations were performed on the datasets.

Countries and years were then selected based on their relative data ranges, and were filtered to include only countries in the Northern hemisphere to avoid any effects of hesitation or suspicion from participants who are not familiar with the shape of temperature data in the Southern hemisphere (where winter and summer months are inverted from the US perspective). Specifically, the 15 datasets that had relative data ranges closest to 0.1 (which we deemed "small"), 0.6 (which we deemed "large"), and 0.35 (which we deemed "medium") were selected. For the list of country-year pairs that were used, see Table 5.

5.2 Procedure

After agreeing to participate, each participant was shown 6 charts, one from each data range and truncation level, each randomly chosen from the 15 charts generated for each data range. All seven tasks were presented before moving on to the next chart. The order of both the charts and the tasks were independently randomized using Latin Square. As before, instructions were given to the participants but no trainings or practice rounds were conducted.

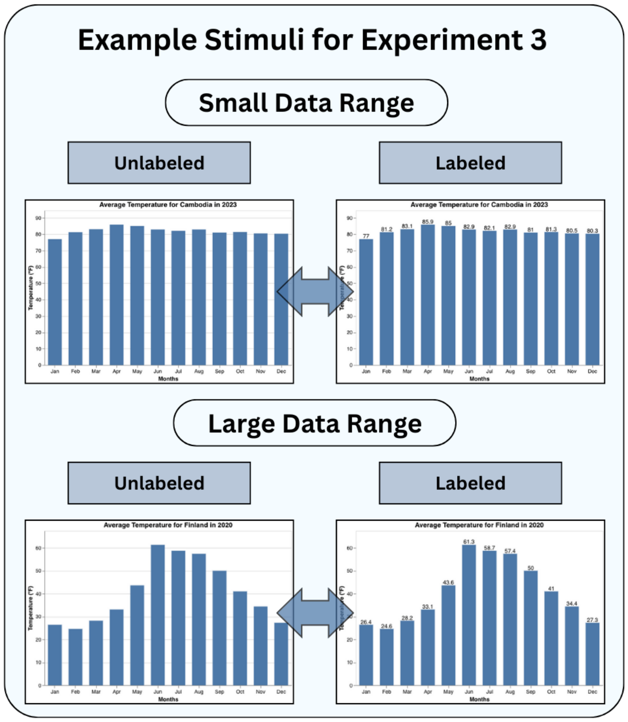


Figure 6: Examples of charts used in Experiment 3. All charts in Figure 5 and Figure 6 are examples used in Experiment 3, in which all charts were either small or large (0.1 or 0.6) data range, labeled or unlabeled, and truncated or untruncated.

5.2.1 Literacy, Numeracy, and Demographic Survey. After the visualization tasks were completed, the survey once again included a Mini-VLAT test for visualization literacy and a brief demographic survey. We also included a numeracy test, namely the 8-item Subjective Numeracy Scale (SNS-8) [9].

The SNS-8 was chosen for its compactness, similar to the Mini-VLAT, and lower cognitive demand than more objective numeracy tests. In prior work, the SNS-8 and related tests have been shown to reliably predict numeracy through the course of 8 subjective questions about their numeracy abilities, such as calculating a 15% tip or the sale price of an item when shopping, and their numeracy preferences, such as how useful they find mathematical skills and whether they prefer weather forecasts to include percentages or words. The responses are Likert scales from 1 to 6. For all but one question, 1 is considered "low" and 6 is considered "high" (for instance, low or high confidence); the exception is question 7, which is purposefully flipped to mitigate the effects of attention-less clicking (e.g., clicking all "6" answers). The score is then given by adding all responses together (after flipping question 7), ranging from 6 (being the lowest subjective numeracy) to 48 (being the highest subjective numeracy).

5.3 Data and Analysis

We performed a second power analyses for repeated-measures $2 \times 3 \times 7$ ANOVA with 42 measurements per participant, $\alpha = 0.05$, $\beta = 0.9$, and conservative estimates of nonsphericity correction

Table 5: Relative Data Ranges for Temperature Datasets

Small Range	
Dataset	*RDR*
Belize (2021)	.095
Belize (2022)	.097
Benin (2023)	.101
Cambodia (2023)	.104
Cameroon (2024)	.095
Cape Verde (2024)	.101
Cuba (2021)	.103
Dominican Republic (2023)	.104
Guinea-Bissau (2020)	.095
Guinea-Bissau (2024)	.100
Honduras (2023)	.096
Nigeria (2021)	.102
Nigeria (2023)	.105
Suriname (2023)	.105
Thailand (2022)	.100
Medium Range	
Cyprus (2024)	.346
Egypt (2021)	.353
France (2021)	.357
France (2023)	.350
France (2024)	.352
Ireland (2021)	.355
Ireland (2022)	.340
Israel (2023)	.357
Palestine (2024)	.343
Portugal (2021)	.361
Qatar (2020)	.352
Qatar (2022)	.343
Qatar (2023)	.346
UAE (2020)	.345
UAE (2021)	.342
Large Range	
Afghanistan (2022)	.605
Afghanistan (2024)	.596
Andorra (2021)	.592
Belarus (2022)	.609
Estonia (2022)	.607
Finland (2020)	.599
Lithuania (2022)	.603
Moldova (2024)	.604
Poland (2021)	.591
South Korea (2020)	.593
South Korea (2024)	.596
Turkey (2022)	.601
Turkmenistan (2021)	.607
Ukraine (2022)	.598
USA (2020)	.606

Table 6: Results of the ANOVA tests performed for Experiment 2.

Error ANOVA

Effect	F	η_p^2	p
range	2.18	.038	.119
trunc	1.63	.028	.207
task	**28.83**	**.340**	**<.001**
range:trunc	2.96	.050	.056
range:task	**9.21**	**.141**	**<.001**
trunc:task	**11.56**	**.171**	**<.001**
range:trunc:task	**2.84**	**.048**	**.024**

Duration ANOVA

Effect	F	η_p^2	p
range	3.05	.052	.052
trunc	1.57	.027	.216
task	**158.30**	**.739**	**<.001**
range:trunc	2.58	.044	.088
range:task	1.78	.031	.068
trunc:task	**5.47**	**.089**	**<.001**
range:trunc:task	1.46	.025	.155

Note: Box-Cox transformed; Greenhouse–Geisser corrected; Type III tests; $\alpha = 0.05$.

$\epsilon = 0.5$ and effect size $F = 0.11$. From this calculation, our necessary sample size was 53. Using a non-response rate of 33%, we arrived at a sample size of 71, which we once again rounded to 80 participants.

80 US-based, English-speaking adults were once again recruited from Prolific. Participants were paid $5 for their participation, based on an expected 20-minute survey duration and an hourly rate of $15. The actual median duration was 25 minutes, for an actual hourly rate of $11.63 per hour.

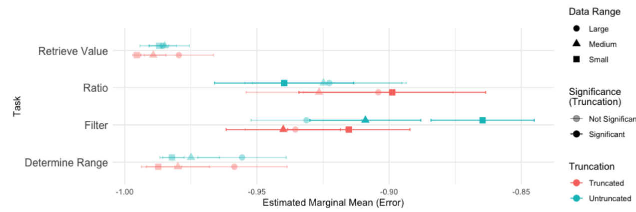
Data was collected in the same format, with responses and duration, from which we calculated error in the same manner as Experiment 1. Participants were also removed in the same way, based on 3 standard deviations in error or duration, leading to the removal of 23 participants and a final sample of $N = 57$. For demographic breakdowns and scores for the Mini-VLAT and SNS-8, see Table 3.

We again ran two ANOVA tests, but this time they were 2×3×7 ANOVA tests with the two truncation levels, three data ranges, and seven tasks. Once again, we included both Box-Cox transformations and GG corrections.

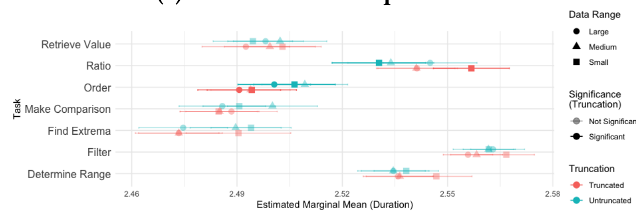
5.4 Results

As can be seen in Table 6, task ($F = 28.83$, $\eta^2 = .340$, $p < .001$), range x task ($F = 9.21$, $\eta^2 = .141$, $p < .001$), truncation x task ($F = 11.56$, $\eta^2 = .171$, $p < .001$), and range x truncation x task ($F = 2.84$, $\eta^2 = .048$, $p = .024$) were all significant predictors of error, while task ($F = 158.30$, $\eta^2 = .739$, $p < .001$) and truncation x task ($F = 5.47$, $\eta^2 = .089$, $p < .001$) were significant predictors of duration.

As range and truncation both interacted with task, we then ran post-hoc EMMs consistent with Experiment 1, using Box-Cox transformations for the data and Holm corrections to the p-value, including all three variables. The results are displayed in Figure 7.



(a) Error EMM for Experiment 2



(b) Duration EMM for Experiment 2

Figure 7: The differences in Estimated Marginal Means (EMMs) for Experiment 2. The opacity of the points corresponds to the significance of the difference between truncation levels. Note that, due to the nature of the EMM calculation and the Box-Cox transformation, the x-axis values have been transformed and shifted, sometimes to negative values. Importantly, for both charts, the right direction of the x-axis means higher error or longer duration.

For error, some clear patterns emerge. Notably, when the range was small, we see a similar result as Experiment 1, where truncated charts had significantly lower error on the Filter task ($t = . - 4.664$, $p < .001$, $d = -0.62$) and significantly higher error on the Ratio task ($t = 2.839$, $p = .006$, $d = 0.38$). However, when the range was medium, we see Ratio decrease drastically ($t = -0.125$, $p = .901$, $d = -0.02$) while Filter decreases only slightly ($t = -2.483$, $p = .016$, $d = -0.33$), to the point that Ratio no longer has a significant change between truncated and untruncated charts. Finally, in the large data ranges, we were unable to reject the null hypothesis that truncated and untruncated charts did not differ in error. This shows a clear trend to the findings in Experiment 1, in which all effects of truncation are inflated at smaller ranges.

As for duration, when the data range was small, we see slower times for the Ratio task when the axis was truncated ($t = 4.632$, $p < .001$, $d = 0.62$), which contradicts Experiment 1. We suspect that this is due to the larger data ranges overpowering the smaller ranges in that analysis. We also did not replicate the increased speed for the Retrieve Value and Make Comparison tasks, which may be due to various factors in the changed experimental design (such as the change from two different topics/units to one or the more standardized shape of temperature data leading to more consistent overall speed).

To summarize, our participants again showed significantly lower error for the Filter task when the data range was small and medium, but the higher error for the Ratio task was only apparent in small *data ranges. In these same small data ranges, the duration was also significantly longer for the Ratio task.*

5.5 Exploratory Demographic Analyses

5.5.1 Demographic Predictors. We again performed exploratory analyses on the demographic variables we collected for each experiment: Age, Sex, and Education.

For Age, we again ran correlational analyses for error and duration. Once again, the correlations were below $r = 0.1$, and are therefore small enough to not be meaningful. For Sex and Education, as with Experiment 1, we do not find any reason to believe that there is a meaningful relationship with the target variables or the effects of truncation. We once again find no reason to believe that Age, Sex, or Education are significant predictors of performance in these tasks.

5.5.2 Visualization Literacy and Subjective Numeracy. For Mini-VLAT scores, we replicated the trend in Experiment 1 that literacy was negatively correlated with duration, this time for the Retrieve Value task ($r = -0.26$, $p < 0.001$), Determine Range task ($r = -0.26$, $p < 0.001$), Filter task ($r = -0.29$, $p < 0.001$), and Ratio task ($r = -0.24$, $p < 0.001$). However, we did not replicate the moderate correlation between literacy and error for the Ratio task.

As for subjective numeracy (measured by the SNS-8 scores), none of the tasks showed a correlation stronger than $r = 0.1$.

This implies that, while higher visualization literacy may be correlated with faster performance on these tasks, numeracy does not seem to have a significant relationship with error or duration in our experiment.

5.6 Discussion

Once again, in addition to higher error for the Ratio task, we also saw that truncation led to improved accuracy for the Filter task. However, there were additional takeaways. Firstly, the Ratio task only saw significant differences in error and duration when the data range was specifically small, rather than large or medium. On the other hand, we see that the Filter task had decreased error for both the small and medium data ranges, implying that it may be more sensitive to changes in the data range than the Ratio task. At least in our controlled experimental conditions, this appears to indicate that there are situations where we see no significant difference in error or duration for the Ratio task but do see significant improvements to accuracy for the Filter task.

In Experiment 3, we then address our final research question and test the hypothesis that data labeling will significantly mitigate the effects of truncation.

6 Experiment 3: Data Labeling

RQ3 specifically asks whether there are mitigatory strategies to reduce the effects of truncation. To this end, we conducted a third and final experiment, which introduced data labeling as a candidate for reducing the impact of axis truncation.

6.1 Stimulus Design

6.1.1 Tasks and Visualization Design. The third study sought to answer RQ3, which referred to strategies to mitigate the effects of truncation. To this end, we looked specifically at the tasks that showed consistent, strong effects in the prior two experiments: the Filter and Ratio tasks. With these two tasks, we had one that was consistently associated with improved accuracy and one that was consistently associated with increased error.

We also once again leveraged the same design standards as Experiments 1 and 2, with the exception of including data labels as a binary variable in our stimuli, as discussed below.

6.1.2 Data Labeling. For this experiment, we began with the expectation that the increased error in the Ratio task was primarily due to the use of the ratio heuristic rather than direct calculation, as discussed by prior works. For example, when there is a bar that is half the size of another, users may often just assume the value is also half, even when truncation has changed this underlying proportionality.

From this expectation, we decided that a potential mitigation would be to directly show the data on the chart. Data labeling and chart annotation have a long history in data visualization. Prior works have found evidence that viewers often prefer charts with more text and labels, claiming them to be easier to read [55] and taxonomies of helpful chart overlays (such as gridlines) have highlighted numerical data labels as a useful alternative [37]. Similarly, practitioners often use direct labeling of some or all data points for either readability or emphasis [15, 16, 19, 20, 22, 23, 28–31].

Versions of data labeling as a mitigation strategy have been tested in prior works, such as Pandey et al. [46], who found that subjective differences were not meaningfully impacted by data labels for bubble charts. However, it is unclear whether this effect will translate to the Ratio task in bar charts . We hypothesize that labeling the bars will prime participants to calculate the ratio rather than rely on the ratio heuristic. For this experiment, we therefore added a new binary independent variable: labeled vs. unlabeled charts. Otherwise, the chart stimuli were identical to those used in Experiment 2, but with only the small and large data ranges to avoid unnecessary complexity in the experimental design.

6.2 Procedure

As before, after agreeing to participate, our participants were presented with the stimulus charts. Our final experimental design consisted of eight visualizations per participant, which were chosen based on the two truncation levels, two data ranges, and two labeling paradigms. Each chart was randomly ordered and presented with the Filter and Ratio tasks, also in a random order. Once again, aside from general instructions shown to each participant, no trainings or practice rounds were conducted.

6.2.1 Literacy, Numeracy, and Demographic Survey. The participants were then shown the same Mini-VLAT, SNS-8, and demographic questionnaire as the prior experiments to determine their visualization literacy, subjective numeracy, and demographic identifiers.

6.3 Data and Analysis

Using the same power analysis, this time for a within-subjects repeated measures $2 \times 2 \times 2 \times 2$ ANOVA with 14 measurements per participant, $\alpha = 0.05$, $\beta = 0.9$, and conservative estimates of

Table 7: Results of the ANOVA tests performed for Experiment 3.

Error ANOVA

Effect	F	η_p^2	p
range	2.40	.039	.127
label	**103.08**	**.636**	**<.001**
trunc	**14.19**	**.194**	**<.001**
task	**29.96**	**.337**	**<.001**
range:label	0.27	.004	.608
range:trunc	3.07	.050	.085
label:trunc	**8.10**	**.121**	**.006**
range:task	**20.59**	**.259**	**<.001**
label:task	**78.50**	**.571**	**<.001**
trunc:task	**19.18**	**.245**	**<.001**
range:label:trunc	1.44	.024	.235
range:label:task	0.11	.002	.742
range:trunc:task	2.60	.042	.112
label:trunc:task	**6.83**	**.104**	**.011**
range:label:trunc:task	0.64	.011	.427

Duration ANOVA

Effect	F	η_p^2	p
range	0.47	.020	.494
label	**11.76**	**.127**	**.001**
trunc	0.08	.006	.781
task	**20.43**	**.257**	**<.001**
range:label	1.40	.012	.242
range:trunc	0.40	.021	.532
label:trunc	1.66	.012	.203
range:task	**7.85**	**.031**	**.007**
label:task	**28.99**	**.156**	**<.001**
trunc:task	1.79	.005	.186
range:label:trunc	0.23	.002	.631
range:label:task	0.69	.011	.410
range:trunc:task	2.39	.048	.127
label:trunc:task	0.30	.008	.588
range:label:trunc:task	0.04	.008	.525

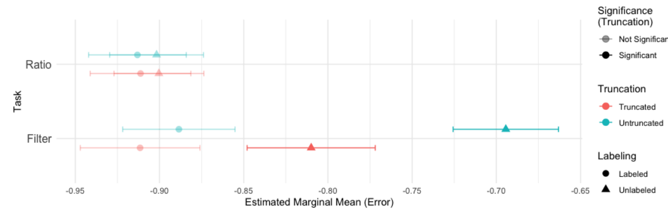
Note: Box-Cox transformed; Greenhouse–Geisser corrected; Type III tests; $\alpha = 0.05$.

nonsphericity correction $\epsilon = 0.5$ and effect size $F = 0.15$, we arrived at a necessary sample size of 53. Using a non-response rate of 33%, we arrived at a sample size of 71, which we once again rounded to 80 participants.

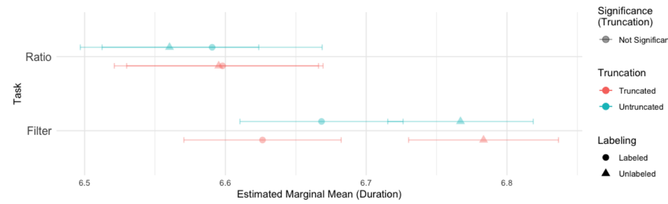
From Prolific, we once again recruited 80 participants, who were all US-based, English-speaking adults, all of whom were paid a base pay of $3.75, based on an expected survey duration of 15 minutes and a base pay of $15 per hour. The actual median duration was roughly 17 minutes, for an actual hourly rate of $13.04 per hour.

As before, we removed 20 participants with error or duration higher than 3 standard deviations from the mean, leaving $N = 60$ as our sample. For demographic breakdowns and scores for the Mini-VLAT and SNS-8, see Table 3.

Once the data was collected, we again ran two 2×2×2×2 ANOVA tests, one for error (which was calculated similar to Experiments



(a) Error EMM for Experiment 3



(b) Duration EMM for Experiment 3

Figure 8: The differences in Estimated Marginal Means (EMMs) for Experiment 3. The opacity of the points corresponds to the significance of the difference between truncation levels. Note that, due to the nature of the EMM calculation and the Box-Cox transformation, the x-axis values have been transformed and shifted, sometimes to negative values. Importantly, for both charts, the right direction of the x-axis means higher error or longer duration.

1 and 2) and one for duration, transforming and correcting in the same way. This time, the independent variables were all binary, with truncated vs. untruncated axes, small vs. large ranges, Filter and Ratio tasks, and labeled or unlabeled data. The results are shown in Table 7.

6.4 Results

For the error ANOVA, we saw significant effects for several variable combinations (refer to Table 7 for a complete list), with labeling alone actually being the largest η^2 effect size in the whole model ($F = 103.08$, $\eta^2 = .636$, $p < .001$). This was followed by the interaction effect between labeling and task ($F = 78.50$, $\eta^2 = .571$, $p < .001$). We also saw significant effects from the interaction of task with both range ($F = 20.59$, $\eta^2 = .259$, $p < .001$) and truncation ($F = 19.18$, $\eta^2 = .245$, $p < .001$). As for the duration, there were far fewer significant effects, as seen with the Filter and Ratio tasks in the prior experiments. However, the significant effects were label x task ($F = 28.99$, $\eta^2 = .156$, $p < .001$), task ($F = 20.43$, $\eta^2 = .257$, $p < .001$), label ($F = 11.76$, $\eta^2 = .127$, $p = .001$), and range x task ($F = 5.85$, $\eta^2 = .031$, $p = .007$).

Due to these results, for the error, we ran post-hoc EMMs (using a Box-Cox transformation and a Holm p-value correction as with prior analyses) for the truncation by the labels and task (refer to Figure 8) and found that when looking at the interaction between task and truncation, there was no longer an effect from truncation on the error for the Ratio task for either the labeled ($t = 0.306$, $p = .760$, $d = 0.04$) or unlabeled charts ($t = 0.147$, $p = .883$,

$d = 0.02$). We also see that, while the same significant improvement occurred in the Filter task when the unlabeled chart was truncated ($t = -4.826$, $p < .001$, $d = -0.63$), the presence of labels was associated with a smaller difference ($t = -1.07$, $p = .289$, $d = -0.14$), such that the null hypothesis could no longer be rejected between truncated and untruncated charts.

We also ran post-hoc EMMs on duration for the labeling by the task (as truncation was not significant in our model). Note that, for consistency with the other figures in this work, the significance in Figure 8 is reported using truncation rather than labeling (to avoid confusion). We found, as expected, that labeling the charts improved the completion time for the Filter task significantly, for both truncated ($t = -5.725$, $p < .001$, $d = -0.75$) and untruncated charts ($t = -3.810$, $p < .001$, $d = -0.50$), and did not meaningfully affect the speed of the Ratio task for either truncated ($t = 0.086$, $p = .932$, $d = 0.01$) or untruncated charts ($t = 0.856$, $p = .396$, $d = 0.11$).

In summary, our participants yet again showed significantly lower error for the Filter task when the charts were unlabeled, but this difference was no longer significant when the charts were labeled. Overall, we saw a reduction in error and duration for the Filter task when the charts were labeled. We also saw no evidence that the Ratio task had higher error or duration, regardless of truncation level, for either the labeled or unlabeled charts, in contradiction to the prior experimental results.

6.5 Exploratory Demographic Analyses

6.5.1 Demographic Predictors. Once more, the Age, Sex, and Education of participants was explored for potential demographic or personal differences in the results of Experiment 3. Age, for the first time in these experiments, did seem to have some relationship with specifically duration. Specifically, age had a positive correlation with duration for both the Filter task ($r = .23$, $p < 0.001$) and Ratio task ($r = .19$, $p < 0.001$). As for Sex and Education, none of the groups showed any significant differences or interactions with task or labeling for error or duration. Once more, these results showed a lack of meaningful significant differences between demographic groups, with the only significant correlations being between Age and duration.

6.5.2 Visualization Literacy and Subjective Numeracy. For visualization literacy, we found evidence that it correlated with error for the Ratio task ($r = -.18$, $p < 0.001$) and with duration for the Filter task ($r = -.26$, $p < 0.001$), implying that higher visualization literacy was associated with faster performance on the Filter task and more accuracy on the Ratio task. Regarding numeracy, we found the same correlations with error for the Ratio task ($r = -.22$, $p < 0.001$) and with duration for the Filter task ($r = -.18$, $p < 0.001$). Both of these correlations are notably higher than in Experiment 2, with the Ratio task having a small correlation with error ($r = -0.12$, $p = 0.02$) and the Filter task having an even smaller correlation ($r = -0.004$, $p = 0.94$).

6.6 Discussion

In this experiment, we once again found that, when the axis was truncated and labels were not included, the Filter task was performed with significantly lower error. However, when the charts were labeled, the Filter task was performed with holistically lower error, regardless of truncation, and the difference between the truncated and untruncated charts is no longer significant. We also see the same trend in duration where, while the difference between truncation levels was not significant for either labeled or unlabeled charts, the labeled charts had significantly lower duration regardless of truncation.

However, we also found that the error for the Ratio task, which had been one of the largest and most consistent differences in both of the prior experiments, was no longer significant. In terms of error, none of the four cases (truncated vs. untruncated and labeled vs. unlabeled) were significantly different, rather being remarkably close in error. This was also true for duration. Our interpretation, based on the consistency and strength of the difference in prior experiments, is that the presence of the data labels not only lowered error for the charts that had the data labels, but also primed participants to notice the data values, therefore highlighting the truncated axis. This echoes the finding from Schlieder et al. [51] that participants can learn to detect and mitigate truncated charts, and shows that direct data labels may potentially be a catalyst for this effect. At least in our controlled experimental conditions, this implies that data labeling has a strong priming effect that may mitigate the effects of truncation and holistically lower error.

7 General Discussion

7.1 Takeaways

In this work, we have provided a closer look at some of the potential pros and cons of axis truncation in bar charts. Through the three experiments discussed throughout this paper, we have found evidence for three key takeaways: truncation does not seem to be entirely harmful when considering tasks such as Filter and Retrieve Value, the effects of truncation appear to be related to the amount of truncation, and data labeling may mitigate the effects of axis truncation by reducing error for both the Ratio and Filter tasks and lessening the difference in error between truncated and untruncated charts.

7.1.1 Truncation Pros and Cons. Throughout the three experiments, certain themes held true about the benefits and detriments to truncating bar charts. While the detriments were largely similar to prior work on the subject, some of these benefits have been underexplored in the literature and should be investigated with more nuance.

Beginning with the dangers, the clearest takeaway is the ratio heuristic. Truncated charts consistently led to strong, significant increases in error for the Ratio Task. If the Ratio task is important for a given use case, this would be a significant detriment to the readability of the chart.

The benefits, however, were also significant. Especially with the Filter task, there were significant improvements in the accuracy of users when the axis was truncated. This follows expectations, as the finer detail of the axis would allow for more precision. We also saw that the Retrieve Value, Make Comparison, and Order tasks saw somewhat increased efficiency when the axis was truncated, though these findings were notably inconsistent between differing experimental designs.

Altogether, this paints a picture of one important drawback in the increased error of the Ratio task and several potential benefits in other tasks. These benefits should be included in a nuanced evaluation of axis truncation in bar charts.

7.1.2 How Much to Cut? We also see that there is an interaction between these effects and the amount of truncation, as expected. It stands to reason that, when truncation is minimal (and the baseline is close to 0), the effects of that truncation will be smaller than when the baseline is much higher, especially relative to the data values.

We found exactly this in our analysis, in that the positive effects on the Filter task and negative effects on the Ratio task both became more severe as the relative data range decreased. For the Ratio task, the results began to differ significantly between the medium and small ranges, which would be between the 0.35 and 0.1 data ranges, while the Filter task began to see benefits at the medium range (between 0.6 and 0.35). This implies that the Filter task benefits are more sensitive to the data range than the Ratio task drawbacks, and implies that there may be a Goldilocks zone of truncation that allows for the benefits of the Filter task to possibly eclipse the detriments of the Ratio task. Future work should validate these results and investigate this possibility.

7.1.3 Mitigating with Labels. Finally, we see some notable takeaways from data labels. As discussed in Experiment 3, we found evidence that the presence of data labels significantly decreased the effect of axis truncation for the Ratio task, which was the primary negative effect. In fact, in Experiment 3, our results showed no significant difference in Ratio task performance for unlabeled charts, despite the Ratio task being one of the most consistent effects of truncation in prior experiments.

We hypothesize that this is due to the presence of data labels in the experiment overall. We believe that the data labels drew the users' attention to the truncation, as it showed that the values were not directly proportional to the height of the bars, and subsequently primed them to forego the use of the ratio heuristic. Therefore, even when the charts were not labeled, the users opted to complete the Ratio task as intended (through value-based calculation) rather than through trivial bar length comparisons.

This shows that data labeling may have not only reduced the negative effects of axis truncation, but that labeling some truncated charts seemed to improve the performance for unlabeled truncated charts, suggesting a strong neutralizing effect. This is also in line with our exploratory analysis of SNS-8 numeracy scores, which showed that even in unlabeled charts, numeracy was a significant predictor of the Ratio task in Experiment 3, implying that participants were more likely to actually use numeracy skills to attempt to solve the Ratio task, unlike in Experiment 2. Therefore, we see data labeling as a potential counter to the negative effects of truncation.

7.1.4 Nuance. In this work, we have provided empirical evidence that axis truncation should not be treated as a universal faux pas, and should not be universally criticized as "misleading". That said, this work should in no way be used to imply that bar chart truncation is a universally positive practice. We simply provide evidence that the story is complex and warrants nuanced decision making on a case-by-case basis. We argue that truncation is more of a tradeoff with contextual drawbacks or benefits that a given designer should weigh to determine the best design to accomplish their goals. In other words, there is not and should not be a universally "optimized" level of truncation for bar charts. Until more work is done to tease apart these subtle nuances, we recommend that truncation can be beneficial, but advise caution, especially where ratios are important.

7.2 A Critique of the Ratio Task

We also want to address the elephant in the room: the Ratio task. While this task has its basis in the literature and its practical uses, we believe this task to be uncommon and unintuitive, confusing participants and relying more on numeracy skills than on chart comprehension.

In terms of prevalence, we struggled to find any prior works (outside of those explicitly discussing the problems with bar chart truncation) that list Ratio as a bar chart task. It could easily be argued that, under the definitions presented by Saket et al. [49] and Amar et al. [2], it could fall under "Compute Derived Value", but that is a broad umbrella that would also include any mathematical operations. It could also be argued that it can be included in "Make Comparison" as described in the VLAT [38], but this again is broad enough to include many operations that are far more common that computing ratios.

We also note that the taxonomies of tasks generated by Amar et al. [2] and Chen et al. [61] both at least partially sourced their tasks from questions or uses for charts identified by chart audiences, namely students for Amar et al. and practitioners in online forums for Chen et al. This certainly does not invalidate the Ratio task, but rather asks why so few of these students or visualization practitioners thought to calculate the ratio of two bars, but rather referred to comparing values in the abstract or finding differences. We expect the answer is due to the low prevalence of this task in practice.

It is true that certain use cases may focus heavily on the proportion between two bars. For example, the margin of victory for political candidates or the relative survival rate of medical interventions may affect popular support or life-changing decision making, respectively. But, just as there are cases where the Ratio task may be somewhat valuable, there are also cases where the Ratio task is irrelevant, such as when the 0 baseline is not mathematically meaningful. For temperatures (unless using Kelvin), for example, it does not make sense to say that 20 degrees is "twice as hot" as 10 degrees. Therefore, we believe that although the Ratio task illustrates the dangers of unfettered truncation, it is not universally applicable or useful and should not be used as an unnuanced bludgeon to label all truncation as problematic.

7.3 Limitations and Future Work

In this work, we showed experimental evidence for potential benefits to bar chart truncation, how the task performance interacts with the data range and amount of truncation, and a potential strategy to improve the readability of truncated bar charts, even for the Ratio task. We hope for future work to build on this foundation to validate our findings, gain an even more nuanced view of truncated bar charts, and give designers the tools they need to maximize the effectiveness of their data communication.

An important limitation of this work is generalizability. Specifically, all of these results were found in controlled experiments in which participants had consented to participating in a research study. This will clearly impact the attention and care they give to their answers. Therefore, future work should explore if these findings, especially the mitigation offered by data labeling, is as effective under time pressure or lowered attention, as these conditions may be more accurate to real-world visualization scenarios (i.e., scrolling idly through social media).

This work also focuses on task-based effectiveness, as this was a strong avenue to show the different perspectives on the effects of truncation. However, this is also only one potential aspect of truncation. Future work would do well to investigate other potential metrics, such as impacts to decision making, aesthetic appeal, or trust.

Similarly, this work only utilizes a binary view of truncation, either a 0 baseline or the two-thirds rule. This was intended to simplify the experimental design with an extreme binary, but future work should investigate more subtle truncations to determine if there are specific levels of truncation that are more or less effective. We also treated the data range as a categorical variable for these analyses, but there would also be benefit in determining (via regression or other means) if there is a specific data range or level of truncation where the truncation becomes impactful.

This work also focuses on popular guidelines regarding bar chart truncation and therefore does not consider alternatives to truncation. Such alternatives could include manipulating the aspect ratio of the chart, using alternative tick intervals (such as logarithmic scales or bar compression), or reporting differences in charts rather than the values themselves. In some situations, these techniques may offer better readability or more consistent performance on the Ratio task. Future work should explore these alternative designs and potential tradeoffs with bar chart axis truncation.

Finally, although data labeling seems to be a promising way of mitigating the negative impacts of truncation, there are potential issues with text and labels, such as interpretations of the chart message being biased [54] or viewers relying too heavily on either text or visualization without synthesizing the two [45]. Our stimuli for data labeling also utilized both direct labels and axes, rather than replacing the axes with direct labels, which some practitioners claim may lead to unnecessary clutter [16, 19, 20, 22, 29, 31]. Future work can build on these findings to compare labeling with other mitigation techniques and test if our findings are replicated when only direct labeling is used.

8 Conclusion

This work is an attempt to bring a more nuanced view of bar chart axis truncation to light. As seen throughout this work, there are good-faith reasons that designers may want to truncate the axis, and it is important to give them the tools to do so effectively and without incident. Even those who are staunchly against truncating bar charts may still see the value in learning more about its effects or developing strategies to mitigate its impacts.

We have provided further validation that truncating bar charts leads to worse accuracy for the Ratio task, but have also found evidence that other tasks (most notably Filter) see improved performance under certain experimental conditions when the axis is truncated. We have also shown at least one variable, namely the data range, that interacts to increase or reduce the effects of truncation. Finally, we supplied data labeling as a potential mitigation strategy to reduce the impacts of axis truncation, showing that the labeling led to a significant reduction in the overall error for the Ratio task, as well as the Filter task.

Overall, we find reason to believe that bar chart axis truncation is not nearly as black-and-white as it is often portrayed. Although there are certainly associated risks, there are also situations where axis truncation may be beneficial to the readability of the chart. These pros and cons should be carefully weighed by designers and thoroughly investigated in future research, rather than summarily dismissed out of hand.

Acknowledgments

This work was partially supported by the National Science Foundation (NSF) under Award Numbers 2142977 and 2330245.

References

[1] Derya Akbaba, Jack Wilburn, Main T. Nance, and Miriah Meyer. 2021. Manifesto for Putting 'Chartjunk' in the Trash 2021! https://altvis.github.io/papers/2021/akbaba.pdf

[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 (INFOVIS '05)*. IEEE Computer Society, USA, 15. https://doi.org/10.1109/INFOVIS.2005.24

[3] Matthew Brehmer and Tamara Munzner. 2013. A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (Dec. 2013), 2376–2385. doi:10.1109/tvcg.2013.124 Publisher: Institute of Electrical and Electronics Engineers (IEEE).

[4] Alberto Cairo. 2015. *Graphics Lies, Misleading Visuals.* Springer-Verlag, London, 103–116. doi:10.1007/978-1-4471-6596-5_5

[5] 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*. ACM, Honolulu HI USA, 1–12. doi:10.1145/3313831.3376222

[6] Michael Correll and Jeffrey Heer. 2017. Black Hat Visualization. In *Workshop on Dealing with Cognitive Biases in Visualisations (DECISIVe)*. IEEE VIS, Phoenix, AZ, 1–4. https://idl.uw.edu/papers/blackhatvis

[7] Yiren Ding, Jack Wilburn, Hilson Shrestha, Akim Ndlovu, Kiran Gadhave, Carolina Nobre, Alexander Lex, and Lane Harrison. 2023. reVISit: Supporting Scalable Evaluation of Interactive Visualizations. In *2023 IEEE Visualization and Visual Analytics (VIS)*. IEEE VIS, Melbourne, 31–35. doi:10.1109/VIS54172.2023.00015

[8] Stephanie Evergreen and Chris Metzner. 2013. Design Principles for Data Visualization in Evaluation. *New Directions for Evaluation* 2013, 140 (2013), 5–20. doi:10.1002/ev.20071

[9] Angela Fagerlin, Brian J. Zikmund-Fisher, Peter A. Ubel, Aleksandra Jankovic, Holly A. Derry, and Dylan M. Smith. 2007. Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. *Medical Decision Making* 27, 5 (2007), 672–680. arXiv:https://doi.org/10.1177/0272989X07304449 doi:10.1177/0272989X07304449 PMID: 17641137.

[10] Steven L. Franconeri, Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman. 2021. The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest* 22, 3 (2021), 110–161. arXiv:https://doi.org/10.1177/15291006211051956 doi:10.1177/15291006211051956 PMID: 34907835.

[11] Merideth Gattis and Keith J. Holyoak. 1996. Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 1 (1996), 231–239. doi:10.1037/0278-7393.22.1.231

[12] Lily W. Ge, Yuan Cui, and Matthew Kay. 2023. CALVI: Critical Thinking Assessment for Literacy in Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–18. doi:10.1145/3544548.3581406

[13] Lily W Ge, Matthew Easterday, Matthew Kay, Evanthia Dimara, Peter Cheng, and Steven L Franconeri. 2024. V-FRAMER: Visualization framework for mitigating reasoning errors in public policy. (5 2024). https://sussex.figshare.com/articles/conference_contribution/V-

FRAMER_Visualization_framework_for_mitigating_reasoning_errors_in_public_policy/25452112
[14] Baltimore City Data Fellows Style Guide and Melanie Rogala. 2022. Data Viz Style Guide. https://storymaps.arcgis.com/stories/d19f7d4d2a9b49c7b8f68730e3cda1e6
[15] Carnegie Mellon University Style Guide. [n.d.]. The CMU Brand. https://www.cmu.edu/brand/brand-guidelines/data-viz.html
[16] Delish DS Style Guide. 2024. Data Viz overview. https://delish.supernova-docs.io/latest/data-visualization/data-viz-overview-obPVY4HN
[17] Elastic UI Style Guide. 2021. Data Visualization Overview. https://eui.elastic.co/docs/dataviz/
[18] Gestalt Style Guide. 2025. Data visualization Overview. https://gestalt.pinterest.systems/foundations/data_visualization/overview
[19] HDX Style Guide. 2025. HDX DATA VISUALIZATION GUIDELINES. https://company-200741.frontify.com/d/1GC9otcJfhPe
[20] Justice Innovation Lab Style Guide. 2024. Justice Innovation Lab Data Visualization Guide. https://knowledgehub.justiceinnovationlab.org/reports/JIL_Data_Visualization_Guide.pdf
[21] London Datastore Style Guide and Mike Brondbjerg. 2019. City Intelligence Data Design Guidelines. https://data.london.gov.uk/blog/city-intelligence-data-design-guidelines/
[22] Material Design Style Guide. [n.d.]. Data visualization. https://m2.material.io/design/communication/data-visualization.html#principles
[23] Office For National Statistics Style Guide. [n.d.]. Data visualisation. https://service-manual.ons.gov.uk/data-visualisation
[24] Opattern Style Guide. [n.d.]. How to use charts. https://ux.opower.com/opattern/how-to-charts.html
[25] Sunlight Foundation Style Guide and Amy Cesal. 2014. The Sunlight Foundation's Data Visualization Style Guidelines. https://sunlightfoundation.com/2014/03/12/datavizguide/
[26] Salesforce Lightning Design System Style Guide. 2025. Charts. https://v1.lightningdesignsystem.com/guidelines/data-visualization/charts/#Introduction
[27] Semrush Style Guide. 2025. Charts showcase. https://developer.semrush.com/intergalactic/data-display/chart-showcase/chart-showcase
[28] The Economist Style Guide and Matt McLean. 2017. The Economist visual styleguide. https://design-system.economist.com/documents/CHARTstyleguide_20170505.pdf
[29] TASO Style Guide and Luke Arundel. 2025. Data visualisation style guide. https://taso-he.github.io/technicalguide/data-vis/
[30] Urban Institute Style Guide. 2025. Urban Institute Data Visualization Style Guide. https://urbaninstitute.github.io/graphics-styleguide/
[31] UN Refugee Agency (UNHCR) Style Guide. 2025. UN Refugee Agency Dataviz. https://dataviz.unhcr.org/
[32] VTEX Style Guide. [n.d.]. VTEX Styleguide. https://styleguide.vtex.com/#/Components/%F0%9F%91%BB%20Experimental/Charts
[33] Our World in Data. 2025. Monthly average surface temperatures by year, World. https://ourworldindata.org/grapher/monthly-average-surface-temperatures-by-year
[34] Washington University in St. Louis. 2025. Guidelines Explorer. https://washuvis.github.io/styleguides/
[35] Ohad Inbar, Noam Tractinsky, and Joachim Meyer. 2007. Minimalism in information visualization: attitudes towards maximizing the data-ink ratio. In *Proceedings of the 14th European Conference on Cognitive Ergonomics: Invent! Explore!* (London, United Kingdom) *(ECCE '07)*. Association for Computing Machinery, New York, NY, USA, 185–188. doi:10.1145/1362550.1362587
[36] Christa Kelleher and Thorsten Wagener. 2011. Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software* 26, 6 (2011), 822–827. doi:10.1016/j.envsoft.2010.12.006
[37] Nicholas Kong and Maneesh Agrawala. 2012. Graphical Overlays: Using Layered Elements to Aid Chart Reading. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2631–2638. doi:10.1109/TVCG.2012.229
[38] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2017. VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (Jan. 2017), 551–560. doi:10.1109/tvcg.2016.2598920 Publisher: Institute of Electrical and Electronics Engineers (IEEE).
[39] Maxim Lisnic, Cole Polychronis, Alexander Lex, and Marina Kogan. 2023. Misleading Beyond Visual Tricks: How People Actually Lie with Charts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–21. doi:10.1145/3544548.3580910
[40] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. 2022. Misinformed by Visualization: What Do We Learn From Misinformative Visualizations? http://arxiv.org/abs/2204.09548 arXiv:2204.09548 [cs].
[41] Sheng Long and Matthew Kay. 2024. To Cut or Not To Cut? A Systematic Exploration of Y-Axis Truncation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 207, 12 pages. doi:10.1145/3613904.3642102
[42] Kevin McGurgan, Elena Fedoroksaya, Tina M. Sutton, and Andrew M. Herbert. 2013. Design Principles for Data Visualization in Evaluation. *New Directions for Evaluation* 2013, 140 (2013), 5–20. doi:10.1002/ev.20071
[43] Andrew McNutt, Gordon Kindlmann, and Michael Correll. 2020. Surfacing Visualization Mirages. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3313831.3376420
[44] Stephen R. Midway. 2020. Principles of Effective Data Visualization. *Patterns* 1, 9 (2020), 100141. doi:10.1016/j.patter.2020.100141
[45] Alvitta Ottley, Aleksandra Kaszowska, R. Jordan Crouser, and Evan M. Peck. 2019. The Curious Case of Combining Text and Visualization. In *EuroVis 2019 - Short Papers*, Jimmy Johansson, Filip Sadlo, and G. Elisabeta Marai (Eds.). The Eurographics Association. doi:10.2312/evs.20191181
[46] 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*. ACM, Seoul Republic of Korea, 1469–1478. doi:10.1145/2702123.2702608
[47] Saugat Pandey and Alvitta Ottley. 2023. Mini-VLAT: A Short and Effective Measure of Visualization Literacy. *Computer Graphics Forum* 42, 3 (June 2023), 1–11. doi:10.1111/cgf.14809 arXiv:2304.07905 [cs].
[48] 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) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300423
[49] Bahador Saket, Alex Endert, and Cagatay Demiralp. 2019. Task-Based Effectiveness of Basic Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 25, 7 (July 2019), 2505–2512. doi:10.1109/tvcg.2018.2829750 Publisher: Institute of Electrical and Electronics Engineers (IEEE).
[50] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2017. Vega-Lite: A Grammar of Interactive Graphics. *IEEE transactions on visualization and computer graphics* 23, 1 (2017), 341–350.
[51] Antonia Schlieder, Jan Rummel, Peter Albers, and Filip Sadlo. 2024. The Role of Metacognition in Understanding Deceptive Bar Charts. In *2024 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*. IEEE, 51–59. doi:10.1109/BELIV64461.2024.0001
[52] Jon Schwabish, Amy Cesal, Maxene Graze, and Alan Wilson. [n.d.]. Data Visualization Style Guidelines. http://datavizstyleguide.com/
[53] B. Shneiderman. 1996. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*. IEEE, Boulder, Colorado, 336–343. doi:10.1109/VL.1996.545307
[54] Chase Stokes, Cindy Xiong Bearfield, and Marti A. Hearst. 2024. The Role of Text in Visualizations: How Annotations Shape Perceptions of Bias and Influence Predictions. *IEEE Transactions on Visualization and Computer Graphics* 30, 10 (2024), 6787–6800. doi:10.1109/TVCG.2023.3338451
[55] Chase Stokes and Marti Hearst. 2022. Why More Text is (Often) Better: Themes from Reader Preferences for Integration of Charts and Text. arXiv:2209.10789 [cs.HC] https://arxiv.org/abs/2209.10789
[56] Danielle Albers Szafir. 2018. The good, the bad, and the biased: five ways visualizations can mislead (and how to fix them). *Interactions* 25, 4 (June 2018), 26–33. doi:10.1145/3231772
[57] Edward R. Tufte. 2001. *The Visual Display of Quantitative Information, Second Edition*. Graphics Press, Cheshire, Connecticut. https://www.edwardtufte.com/book/the-visual-display-of-quantitative-information/
[58] 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. doi:10.21105/joss.01057
[59] Jessica K Witt. 2019. Graph Construction: An Empirical Investigation on Setting the Range of the Y-Axis. 17 pages.
[60] Brenda W. Yang, Camila Vargas Restrepo, Matthew L. Stanley, and Elizabeth J. Marsh. 2021. Truncating bar graphs persistently misleads viewers. *Journal of Applied Research in Memory and Cognition* 10, 2 (June 2021), 298–311. doi:10.1016/j.jarmac.2020.10.002 Publisher: American Psychological Association (APA).
[61] Yang Chen, Jing Yang, and William Ribarsky. 2009. Toward effective insight management in visual analytics systems. In *2009 IEEE Pacific Visualization Symposium*. IEEE, Beijing, China, 49–56. doi:10.1109/pacificvis.2009.4906837
[62] Ji Soo Yi, Youn ah Kang, John Stasko, and J.A. Jacko. 2007. Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1224–1231. doi:10.1109/TVCG.2007.70515