Truncating the Y-Axis: Threat or Menace?

Michael Correll, Enrico Bertini, Steven Franconeri,

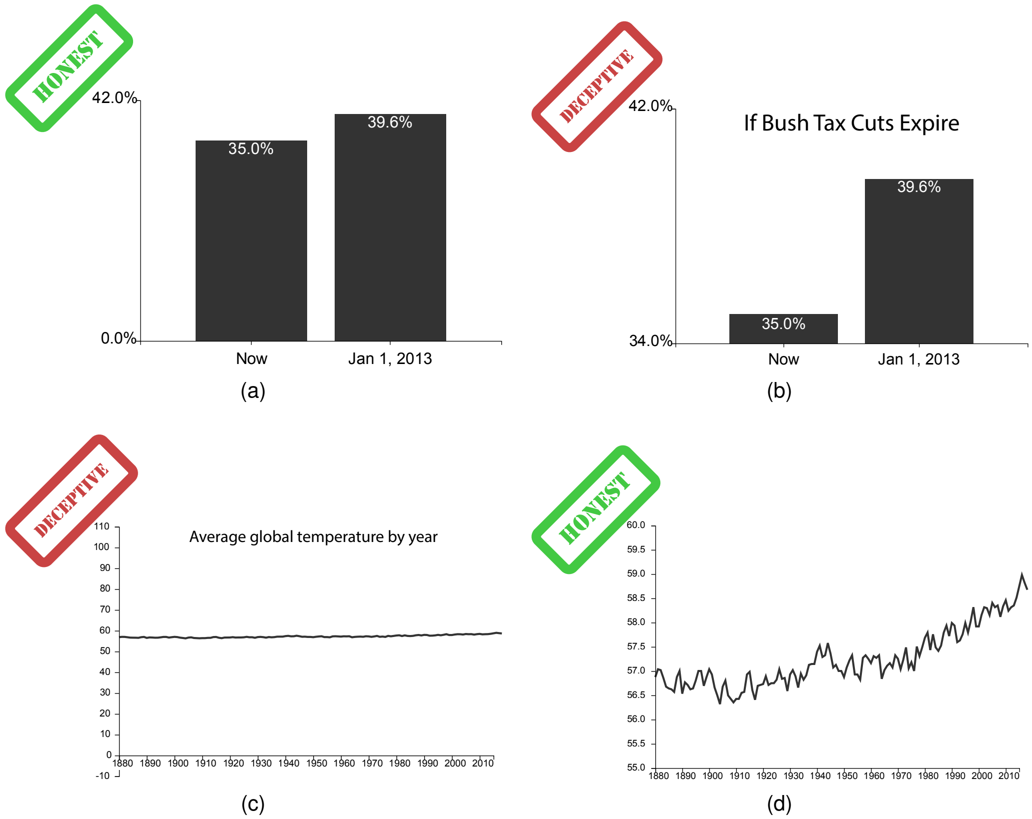


Fig. 1: Two charts that manipulate the y-axis in ways that seem deceptive. In Figure 1b, a reproduction of a chart aired on Fox News [32], the second bar is 6 times taller than the first bar, even though there is only a 4.6% increase in tax rate (ratio of 1.13 to 1). In Figure 1c, a reproduction of a chart tweeted by the *National Review* magazine [26], taken from the Power Line blog, a warming climate is obscured by making the y-axis start at 0 degrees Fahrenheit, compressing the trend into illegibility.

Abstract— Bar charts with y-axes that don't begin at zero can visually exaggerate effect sizes, and in turn lead to unjustified or erroneous judgments. However, advice for whether or not to truncate the y-axis can be equivocal for other visualization types, and there is little existing empirical work on how axis truncation impacts judgments. In this paper we present examples of visualizations where this y-axis truncation can be beneficial as well as harmful, depending on the communicative and analytical intent. We also present the results of a series of crowd-sourced experiments in which we examine how y-axis truncation impacts subjective effect size across visualization types, and we explore alternative designs that more directly alert viewers to this truncation. We find that the subjective impact of axis truncation is persistent across visualizations designs, even for viewers that are aware of the presence of truncated axes. We therefore advise against ironclad rules about when y-axes are appropriate, but ask designers to consider the scale of the meaningful effect sizes and variation they intend to communicate, regardless of the visual encoding.

Index Terms—Information visualization, graphical perception, visualization guidelines

1 Introduction

Starting the quantitative axis of a bar chart from a value other than zero is considered one of the cardinal sins of information visualization. By starting the axis from a value other than zero, the designer *truncates* the range of y-values, over-emphasizing minute differences between values that would otherwise appear very similar in a zero-baseline chart. Bar charts truncated in this manner have been called "biased" [37], "dishonest," [32], "deceptive" [27, 28], "lying with statistics" [19], and "the worst of crimes in data visualisation" [24], with this exaggeration quantified in Tufte's "lie factor" [39]. Prior work has shown that truncation and exaggeration in axes results in quantifiable differences in how people interpret the size and significance of effects [27, 28], affects judgments of correlation [8], and makes trends appear more

- *M. Correll is with Tableau Research. E-mail: mcorrell@tableau.com.*
- *E. Bertini is with New York University. Email: enrico.bertini@nyu.edu.*
- *S. Franconeri is with Northwestern University. E-mail: franconeri@northwestern.edu.*

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

"threatening" [3]. A prescription against non-zero baselines for bar charts is encoded as a hard constraint in automated visualization design tools like Draco [25].

In many guidelines concerning y-axes, bar charts are specifically mentioned as being vulnerable to truncation. By contrast, the injunction against truncating the y-axis is often considered less pressing for line charts (compare the examples in Figure 1). Bar charts use length to encode value, and are often used to afford the quick comparison of individual values. Truncating the y-axis of a bar chart breaks the visual convention that the difference in the height of the bars is proportional to the difference in values, and so is misleading from an encoding standpoint [10]. Line charts do not rely on length but instead use position to encode value, and are often used to afford the quick comparison of trends over time. As such, truncation does not break the convention of the visual encoding in the same way as with bars. Since trends, rather than individual values, are typically important components of the intended messages in line charts, there are some cases where *not* truncating the y-axis is perceived as deceptive (Figure 1c).

However, there is relatively little empirical work on exactly how y-axis truncation inflates judgments across different visual encodings [28, 31]. In this paper, we summarize the current debate over y-axis truncation, drawing on recent work in both visualization and perceptual psychology. We also present the results of a crowd-sourced experiment investigating the impact of y-axis truncation on subjective assessments of data across different visualization designs, with a special focus on designs where the impact of truncation is thought to be lower than for bar charts, or where the presence of truncation is immediately visually detectable. We find that the ability of y-axis truncation to exaggerate effect sizes is not limited to bar charts, but is consistent across multiple chart types, even designs proposed specifically for indicating that truncation has taken place. Our results suggest that the designer therefore has a great deal of control over the perceived effect size in data. There is therefore not a clear binary distinction between "deceptive" versus "truthful" y-axis presentations: designers must take into account the range and magnitude of effect sizes they wish to communicate at a per-data and per-task level.

2 Background & Related Work

When it is permissible to truncate the y-axis is a subject of continuous and active debate. Much of this debate occurs in the pages of books, or in informal channels like Twitter and blog posts, rather than in academic articles. In this section we summarize major positions in this debate, with the intent of synthesizing the major rationales behind existing guidelines.

Huff's [19] *How to Lie With Statistics* calls out charts with non-zero axes as "Gee Whiz Graphs." After y-axis truncation:

> The figures are the same and so is the curve. It is the same graph. Nothing has been falsified– except the impression that it gives. But what the hasty reader sees now is a national-income line that has climbed halfway up the paper in twelve months, all because most of the chart isn't there any more... a small rise has become, visually, a big one.

Huff's advice is that therefore all charts of positive values should begin at 0, lest the designer deceive by making a trend appear more "impressive" than it ought to be (see Figure 2).

Brinton's [5] chapter on "Standards for Time Series Charts" in *Graphic Presentation* similarly claims:

> The amount scale should normally include the zero value or other principle point of reference. Departure from this rule should never be made except where there is a special reason for so doing.

Although he includes an exception, and suggests using visual indicators (such as a "torn paper" metaphor [1]) to indicate when an axis has been adjusted:

> When the interest of the reader is in the absolute amount of change rather than in the relative amount of change, it may be safe to omit the principal point of reference and the accompanying horizontal line[... w]hen the zero value or other principal point of reference is omitted the fact should be clearly indicated in a manner that will attract notice.

More recent discussion on the issue has been less dogmatic, and focuses on how different graphs encode data in different ways, and for different purposes. Alberto Cairo, in *How Charts Lie* [6], proposes the following rule:

> I usually advise a baseline of zero when the method of encoding is height or length. If the method of encoding is different, a zero baseline may not always be necessary. The encodings in a line chart are position and angle, and these don't get distorted if we set a baseline that is closer to the first data point.

Similarly, Carl Bergstrom and Jevin West in their critical thinking website "Calling Bullshit" [4] hold that their principal of "proportional ink" (somewhat analogous to Tufte's "lie factor" [39]) does not apply for line charts:

> [...]unlike bar charts, line graphs need not include zero on the dependent variable axis. Why not? The answer is that line charts don't use shaded volumes to indicate quantities; rather, they use positions that indicate quantities. The principle of proportional ink therefore does not apply, because the amount of ink is not used to indicate the magnitude of a variable. Instead, a line chart should be scaled so as to make the position of each point maximally informative, usually by allowing the axis to span the region not much larger than the range of the data values.

Tufte himself, in a posting on his website [39], seems to take a similar view, but narrows his exception to time series data rather than line charts in general:

> In general, in a time-series, use a baseline that shows the *data*, not the zero point. If the zero point reasonably occurs in plotting the data, fine. But don't spend a lot of empty vertical space trying to reach down to the zero point at the cost of hiding what is going on in the data line itself.

However, Chad Skelton believed that line graphs should not be a special category of exemption from truncation guidelines, and proposed the following [35]:

> Most line charts should start at zero.
> BUT not using baseline zero is OK if:
> a) Zero on your scale is completely arbitrary (ie. temperature) OR
> b) A small, but important, change is difficult or impossible to see using baseline zero.

Related to these more permissive guidelines, Ben Jones specifically collected examples of common genres of line charts where non-zero baselines are not only accepted but are integral parts of the message of the chart (see Figure 3). These charts are examples where the analytical goals entail highlighting change from some baseline other than 0, and in fact would mislead or confuse the viewer if they had non-truncated axes. These examples suggest that, for line charts at least, a 0-baseline is not always appropriate.

Moreover, existing guidelines for the design of line graphs and scatterplots focus on making the overall trend as visible (and decodable with the least error) as possible [9, 18, 38, 41]. These optimizations to chart aspect ratio typically assume that the chart covers the range of the data, rather than necessarily beginning at 0.

There are also competing design considerations to consider, across both encodings and visual metaphors. The visual metaphors in charts impact how data are structured and interpreted [45, 46]. Y-axis truncation breaks the visual conventions of bar charts, as the relative ratio of

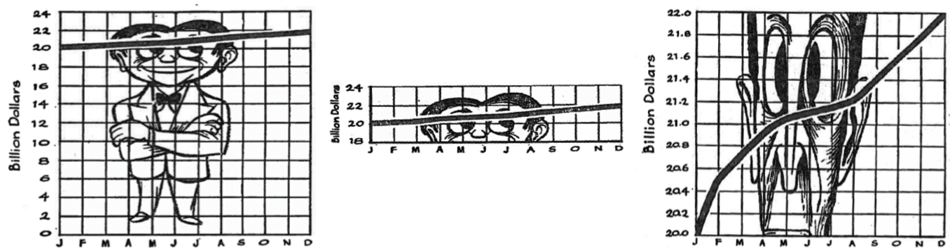
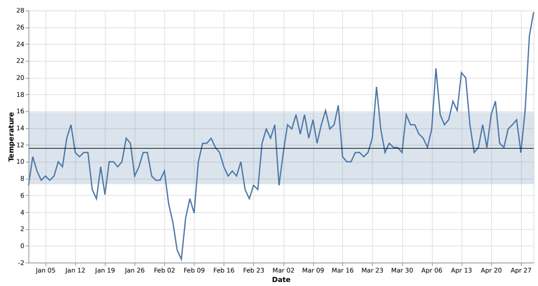
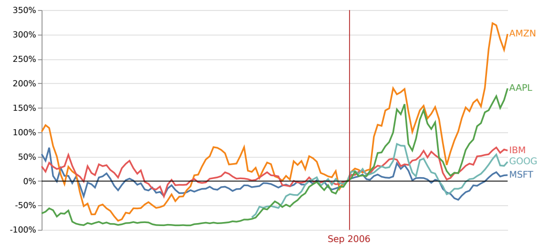


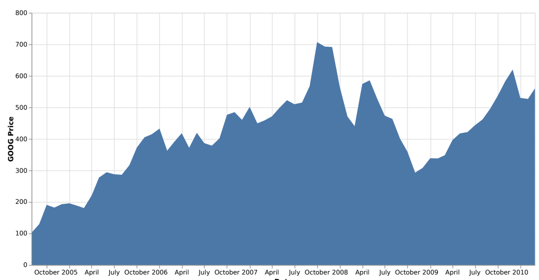
Fig. 2: A visual argument against creating "Gee Whiz Charts," from Huff [19]. For two charts of identical size, starting the y-axis at a point other than 0 creates two sorts of deceptions, for Huff: chopping off most of the context of the chart, and then exaggerating the remaining context. Note that this argument applies broadly to most charts with quantitative y-axes, not just line charts.



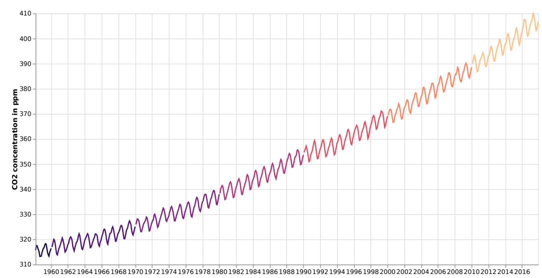
(a) Statistical Process Chart



(b) Index Chart



(c) Stock Chart



(d) Climate Anomaly Chart

Fig. 3: Reproductions of examples, as collected by Ben Jones [21], of line charts where having a non-zero baseline for the y-axis is beneficial. Statistical process charts (3a) rely on comparison to an expected value, and so deviations from that value, not from zero, are important. Index charts (3b) compare to an indexed value rather than zero. In stock charts (3c) small differences in stock value can translate to enormous gains or losses. Similarly, climate anomaly charts (3d) rely on both highlighting deviation from a non-zero expected value but also emphasize the potentially disastrous impact of even minute changes in climate.

heights between two bars is no longer proportional to their difference in value (a bar that is twice as high may not represent a value that is twice as large). However, simply representing the same data as a line chart may not resolve this broken convention. Line charts have a *continuous* encoding of position on the x-axis, and so employ a metaphor of *continuity*. For data with discrete categories on the x-axis, a line chart may therefore be inappropriate. Yet, in many cases data divided into discrete categories have (practical) significant effect sizes that occur in a narrow dynamic range, or important differences in the data that occur in narrow dynamic ranges that are numerically distant from one another. Highlighting changes in narrow ranges of a bar chart when the y-axis starts at 0 is challenging, and existing solutions (such as those in Figure 6) have not been empirically vetted. Designers may not have an ideal solution to the problem of depicting differences in small dynamic ranges of categorical data without violating some expectation or encoding practice in their visualization.

In summary, truncating a y-axis may or may not always be dishonest, and so to be avoided. This anathema may or may not extend to line charts or time series data, or it could just be the case for bar charts. Even if line charts are an exception, whether it is permissible to truncate may depend on task (relative versus absolute change, trends versus values) and on data (meaningful versus non-meaningful baselines).

2.1 Related Work in Perceptual Psychology

The guidelines from the visualization community that we highlight often focus on visual conventions or visual clarity related to issues of truncation. However, it is not clear *a priori* whether these concerns or suggested interventions connect to *graphical perception*. In particular, we were interested if research from perceptual psychology could suggest different patterns of behavior or legibility for different forms of truncated charts.

The research literature on perceptual psychology suggests that bar and line encodings may lead to different effects of y-axis truncation, because the two data encodings may be treated differently by the visual system. If the y-axis truncation manipulation distorts a perceived effect size because it distorts the length encoding of the object that represents the data value, this distortion might be stronger for bars, which explicitly encode length, compared to lines, which do not. The explicit length encoding for bars should not only be more salient, but actually tough to resist for the perceptual system. Decades of work in perceptual psychology suggests that visual attention—the process of filtering visual information to focus processing on subsets at a time

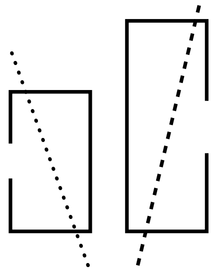


Fig. 4: Demonstration of object-based attention from Duncan [13]. Bars could be either tall or short, with a gap on the left or right. Lines could be tilted left or right, and dotted or dashed. Making two rapid decisions about the same object (the tilt and pattern of the line) was more accurate than two decisions about separate objects (the tilt of the line and the height of the box), suggesting that visual attention has a "preference" for selecting whole objects at once.

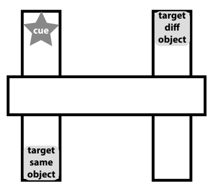


Fig. 5: Demonstration of object based attentional spreading from Egly et al. [14]. When one end of a rectangle was cued with an attention-demanding flash, quick decisions about the identities of targets at the other end of the same object are faster and/or more accurate compared to targets on other objects, despite being equidistant, again suggesting that visual attention tends to focus on entire objects.

(see [15] for review)—can be focused not just on regions of space, but also tends to focus involuntarily on entire visual objects [7]. Figure 4 and Figure 5 and depict two classic examples of this spread of attention (though note that there is debate over whether the spread is due to a true simultaneous spread of attention, or a preference for shifting among different parts of the same object [33]. This mandatory spread of attention should render the length encoding of bars harder to ignore, in contrast to lines where the length encoding is more implicit, requiring the tougher perceptual operation of attending to the negative space between the line and the baseline [2, 40]. On the other hand, lines explicitly encode local deltas between points as orientations of the line segments, while these deltas require the visual system to superimpose orientations that straddle the tops of individual bars, which might be a tougher operation because it requires joint attention to multiple objects at once.

In summary, if the perceptual distortion due to y-axis truncation is primarily driven by the distorted length encoding, we might expect it to be stronger for bars, while if it is primarily driven by the orientation encoding between data values, we might expect it to be stronger for lines.

Finally, work on object-based attention also suggests that any design that attempts to mitigate the truncation distortion by "cutting" bars into multiple parts will fail, because the viewer's attention will tend to be isolated only to one "half" of the bars—the more diagnostic top half—and ignore or at least discount the otherwise important contextual objects underneath them.

2.2 Research Questions

Synthesizing these threads suggests that there is ambiguity in existing guidelines for how to truncate the y-axis, and suggests the following questions for designers seeking to decide whether or not to truncate the y-axis in their chart:

- Do line charts and bar charts have different perceptual or metaphorical properties that suggest different impacts of truncation, or **is it always deceptive to truncate the y-axis, regardless of visual encoding?**
- If not, **is y-axis truncation at least less harmful for line charts than bar charts?**
- Are any of the proposed methods for indicating y-axis truncation, or otherwise highlighting differences in data with narrow dynamic range (as in Figure 6) effective? That is, **can visual designs alleviate the exaggeration caused by truncation while still accurately conveying the values?**

These questions motivate our exploration of graphical perception work from psychology and elsewhere, as well as motivating a series of crowdsourced experiments on how people interpret the *perceived severity* of effect sizes across visual designs, data metaphors, and indications of truncation.

3 Experiment

In order to assess how different visual presentations and analytical frames affect the inflation of perceived effect size introduced by truncating the y-axis, we conducted a series of two crowd-sourced experiments using the `prolific.ac` platform, approved by the IRB of Tableau Software. Experimental data from Prolific is comparable in quality to those from Amazon's Mechanical Turk platform [29]. The Prolific crowd-working platform is focused on studies rather than more general micro-tasks, and enforces minimum compensation rates of at least $6.50/hour compared to the extremely low average earnings of workers on MTurk [17].

The design of our experiments was influenced by Pandey et al. [28], who use a rating scale to assess the effect of various sorts of "deceptive" visualization practices, including y-axis truncation. Our central measure of interest was therefore the response to the 5-point rating item that related to how severe or important the differences in the data series were (the exact question text depended on the intended framing of the question). Higher ratings indicate a higher *Perceived Severity* of the effect size. Our experimental design extends Pandey et al.'s work to a wider range of visual designs and task framings, and includes repeated within-subject trials on an array of different graphs (rather than just single exemplar pairs of deceptive and non-deceptive charts).

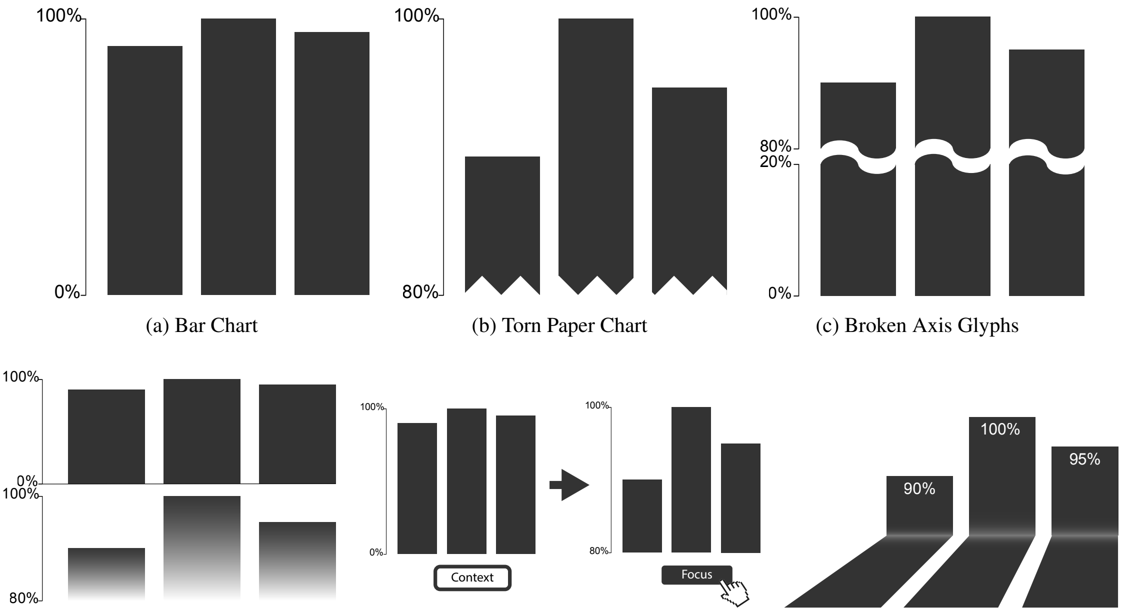
We also asked the participants if they thought the values were *increasing* or *decreasing* (in the trend framing), or if the first value was *smaller* or *larger* than the last value (in the values framing). We used this binary response as an engagement check and to test for comprehension of the chart data. Participants with unacceptably low accuracy at the engagement questions (more than three standard deviations lower than the mean performance) were excluded from analysis, but were compensated for their participation.

Since the main experimental measure was a subjective rating, we presented an initial set of 8 stimuli with every combination of *slope* and *truncation level* in order to present participants with the full range of visual effect sizes and so help to calibrate their rating scale. These initial calibration stimuli were discarded from analysis.

After the main rating task, we gave the participants a 13-item graphical literacy scale developed by Galesic and Garcia-Retamero [16]. Galesic and Garcia-Retamero reported a Cronbach's α of 0.85 for this scale, with some evidence of its utility as a cross-cultural measure of facility in interpreting charts and graphs. The scale does include one item that specifically tests for whether the participant noticed truncation of the y-axis. We collected the answer to this question separately, as well as collecting the overall scale value.

Lastly, we collected demographics data. In addition to standard items such as age and gender, we asked for three free-text responses, the first two of which were required to be non-empty:

- What strategy or procedure did you use to complete the tasks?
- Did you notice anything odd or unusual about the charts you saw during the task?
- Any additional comments of feedback?



(a) Bar Chart (b) Torn Paper Chart (c) Broken Axis Glyphs

(d) Panel Chart (e) Interactive Focus+Context (f) Bent Bar Chart

Fig. 6: Proposed solutions for showing dynamic range in bar charts while indicating truncation or breakage of the y-axis. Figure 6a is a bar chart with three large values in a narrow range. Brinton [5] recommends using a "torn" baseline to indicate an omitted 0 as in Figure 6b. Pelier [30] believes that breaking the axis as in Figure 6c is "a bad idea" and recommends "Panel Charts" as in Figure 6d, which use a gradient to indicate values out of the current scale, with a separate inset on top showing the full range of values without truncation. Ritchie et al. [31], with similar reasoning, uses interaction to animate a transition from focus to context in bar charts as in Figure 6e. Finally, Kosara [23] suggests that "bent" 3D bar charts, as in Figure 6f, make accurate decoding difficult enough that the y-axis truncation is not particularly harmful– the relative rankings of the categories is preserved even in 3D, and to accurately compare values viewers will likely need to consult the labels in any case.

Using these free text responses, a paper author and third party qualitatively coded whether or not the participants' free text responses specifically indicated that they noticed that the y-axes of some of the charts in the experiment were truncated. The coders then discussed mismatches ($\frac{2}{72}$ of codes), which were rectified into a final binary value.

Study materials and data tables are available at `https://osf.io/gz98h/`.

3.1 Experiment One: Framing Interventions

In unconstrained settings, line charts and bar charts can produce different sorts of judgments about the same backing data [44]. Speculatively, one reason for this difference is that bar charts encourage the comparison of individual bars (and thus individual *values*), whereas line charts encourage an assessment of the entire shape (and thus overall *trend*). That is, the bar chart and line chart induce different *framings* of the same data, based on their visual design and resulting *visual metaphors* [45].

While truncation of the y-axes of equally sized line and bar charts both result in magnification of trend (and other measures such as correlation [8]), we were interested in whether or not the differing framings associated with bar charts and line charts would result in differing impacts on judgments when the y-axis is truncated. If so, we were also interested if other strategies to promote different framings (such as text) could have this same impact, without the potential cost or inflexibility of switching visual designs.

3.1.1 Methods

We used a within-subjects study design with the following factors:

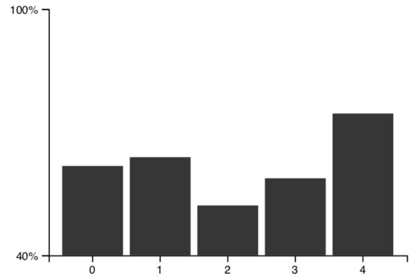
- **Visualization type** (2 levels): whether the data were visualized in a *bar chart* or *line graph*. See Fig. 7 for examples of these designs.
- **Question framing** (2 levels): either a *value-based* or *trend-based* task frame. For the *value-based* framing, participants were asked "Which value is larger, the first value or the last value?" for the engagement question and "Subjectively, how different is the first value compared to the last value?" for the effect size severity question, with the labels "Almost the Same," "Somewhat Different," and "Extremely Different" for the first, third, and fifth items on the rating scale. For the *trend-based* framing, the engagement question was "Are the values increasing or decreasing?", the effect size severity question was "Subjectively, how quickly are the values changing?", and the rating labels were "Barely," "Somewhat," and "Extremely Quickly."
- **Truncation Level** (3 levels): where the y-axis of the visualization began: either at *0*, *25*, or *50%*.
- **Slope** (2 levels) : How much increase (or decrease) there was between the first and last values in the data, either 12.5% or 25%.
- **Data Size** (2 levels): whether there were *two* or *three* data values in the visualization. If there were three data values, the center value was at the midpoint of the first and last items, with a uniform random jitter between $[0, \frac{slope}{4}]$.

Participants saw one of each combination of factors, for a total of $2 \times 2 \times 3 \times 2 \times 2 = 48$ stimuli, presented in a randomized order. Whether the values were *increasing* or *decreasing* was an additional random factor.

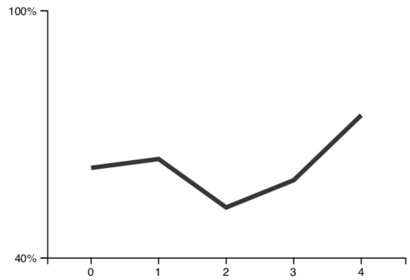
3.1.2 Hypotheses

As with Pandey et al. [28] for bar charts, and Berger [3] for line charts, we expected that **charts with more axis truncation would be perceived as having more severe effect sizes than those with less truncation**.

While there has been some initial work on the impact of framing in visualization [20, 22, 43], and how different visual metaphors can impact how we make use of information in charts [45, 46], there is



(a) Bar Chart



(b) Line Chart

Fig. 7: The two visualization designs in Experiment One. In this example both have truncated axes.

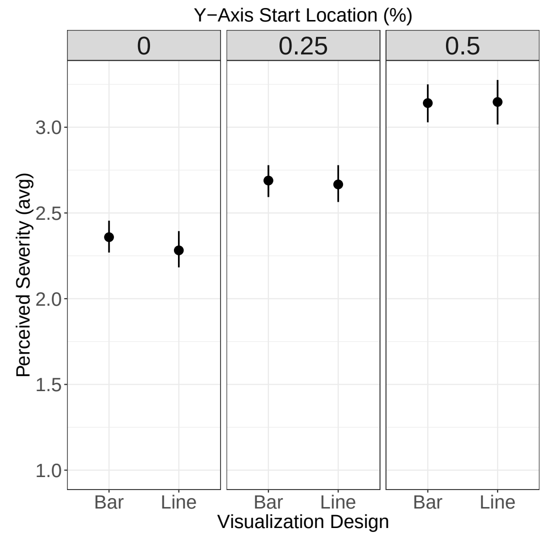


Fig. 8: Results from Experiment One. Increasing the starting point of the y-axis results in larger perceived severity in effect size. This exaggeration appears consistent across bar and line charts. Error bars represent 95% bootstrapped confidence intervals of the mean.

relatively little empirical work on how different visual designs impact framing. Therefore our hypotheses about framing were weakly held. In particular, based on the misleading visual metaphor introduced by truncation in bar charts (where the relative size of the bars is not a proxy for the relative difference in values) we expected that **Bar charts would be more greatly impacted by truncation than line charts** in terms of amplifying perceived severity of effect sizes. Similarly, we believed that **the values framing would be more greatly impacted by truncation than the trend framing**, as comparison of individual values in a truncated graph is fraught and potentially misleading (especially if the axis legend is ignored).

3.1.3 Results

We recruited 40 participants for this task (21 male, 18 female, 1 with a non-binary gender identity, $M_{age} = 27.7$, $SD_{age} = 8.4$). We paid participants \$4 for this task, for an empirical effective hourly rate of \$12/hour. On average, participants scored well on the 13-item Galesic and Garcia-Retamero graphical literacy scale ($M_{correct} = 10$, $SD_{correct} = 2$). In particular, 31 (78%) participants correctly answered the scale item associated with y-axis truncation. Participants also scored highly on our engagement question, correctly labeling the direction of the effect ($M_{correct} = 0.98$, $SD_{correct} = 0.07$), with the exception of one participant, whose performance was more than three standard deviations from the mean (62.5%). The data from this participant was excluded from our analysis.

We conducted a repeated measures ANOVA testing the effect of truncation level, visualization type, question framing, data size, and their interactions on perceived severity.

Our results support our first hypothesis: **increased y-axis truncation results in increased perceived severity** ($F(2,76) = 89$, $p < 0.0001$). A post-hoc pairwise t-test with a Bonferroni correction confirmed that the perceived severity of all three levels of truncation were significantly different from each other. Figure 8 illustrates this result, broken out by visualization type.

Our results fail to support our second hypothesis. **There was no significant effect of visualization design on perceived effect size** ($F(1,38) = 0.5$, $p = 0.50$). Figure 8 shows similar responses to different visualizations across all levels of truncation.

Our results only weakly support our last hypothesis. While there *was* a significant effect of framing on perceived effect size (($F(1,38) = 7.4$, $p = 0.01$), a post-hoc pairwise t-test with a Bonferroni correction did not find a significant difference between the value and trend question framings. Additionally, this effect was quite small: an average decrease in perceived severity of 0.07 for responses using the *trend* framing (based on a $1-5$ rating scale), compared to an increase of 0.36 for starting the y-axis at 25% rather than 0%. Figure 9 shows this result, broken out by visualization design.

In addition to our stated hypotheses, we assessed whether or not our participants, in their free-text responses, mentioned the y-axis manipulation. 45% of participants mentioned this manipulation. The participants who did so reported smaller perceived severity in effect sizes, but this effect occurred at all levels of axis truncation. Figure 10 shows this result.

3.2 Experiment Two: Visual Design Interventions

The results of our first experiment show no robust difference in the impact of truncation on bar charts and line charts: truncation serves to exaggerate effect sizes in both types of graphs. This result suggests that designers may have to employ other methods to indicate that a y-axis has been truncated. A common solution to this problem is to employ the visual metaphor of the "broken" or "continued" axis. Wikipedia recommends indicating truncated axes with glyphs [42] that convey a break from 0 to the start of the truncated axis. To our knowledge, there is no empirical work on whether or not these indications of breaks alter judgments about values. As such, we performed an experiment with similar methodology to our first experiment in order to assess the impact of visual design elements in bar charts that indicate truncation or continuation on perceived effect size.

3.2.1 Methods

Our results from Experiment One were initial evidence that subtle framing effects were not sufficient to reliably impact estimations. As such, we excluded that factor, sticking with the trend framing from the first experiment. Instead, we focused on a narrower set of designs that have been proposed to ameliorate the impact of y-axis truncation in bar charts.

We selected two alternative designs to bar charts as exemplars of designs where the truncation of the y-axis is not only visible in reading the y-axis labels, but is an integral component of the visual metaphor of the chart (see Figure 11). *Bar charts with broken axes* are a common choice to indicate a truncated axis. We used a broken axis design with both a break on the y-axis as well as a break in the bars themselves to

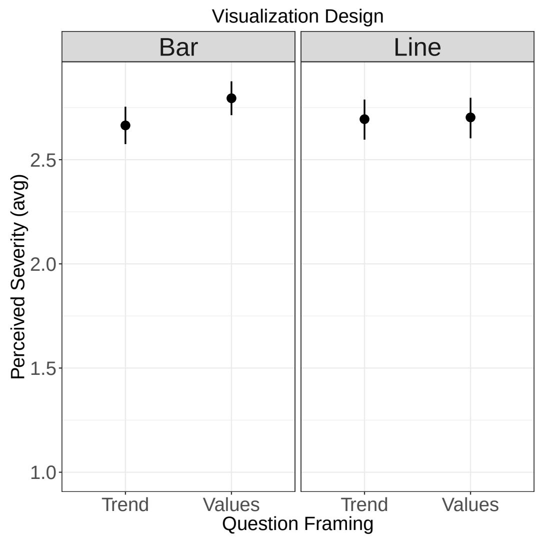


Fig. 9: Results from Experiment One. Using different ways to solicit perceived severity (focusing on either individual *values* or overall *trend*) did not produce robustly different perceptions of effect size. Error bars represent 95% bootstrapped confidence intervals of the mean.

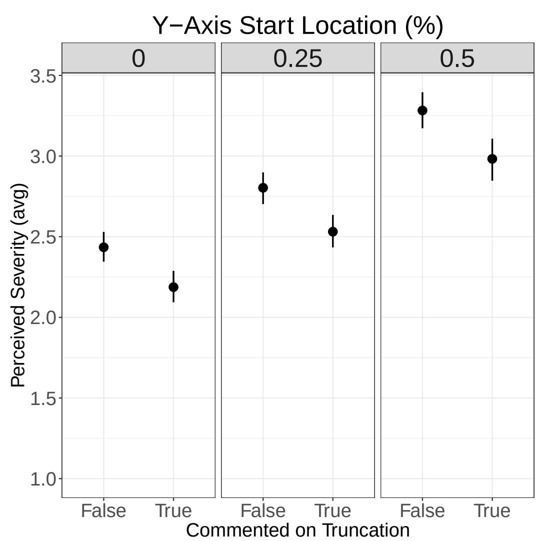


Fig. 10: Results from Experiment One. People who explicitly mentioned the truncated or potentially misleading y-axes reported less severity in effect sizes. Results were similar for Experiment Two. Error bars represent 95% bootstrapped confidence intervals of the mean.

reinforce the metaphor across the entire chart. Bar charts with irregular shapes on the bottom have been reported as complicating decoding [34], so we use rectangular glyphs to indicate breakage rather than the "wavy" or "jagged" glyphs commonly uses to indicate breaks [42](as in Figure 6c). Our second design, a *bar chart with a gradient bottom*, has been empirically considered for other scenarios by both Skau et al. [34] and Diaz et al. [11]. Skau et al. in particular were investigating the case where the gradient is an artistic embellishment (mean to convey e.g., sitting on a reflective surface) rather than conveying continuation. They found that the gradient caused overestimation of value in their task setting, which is potentially advantageous for our task (where exaggeration of value might counteract the effect of truncation). The heights of the bars above the break (for the broken axis chart) and above the y-axis (for the gradient bottom chart) were equal to the height of the bars in the standard truncated bar chart. The pre-break axes and under-bar gradients were therefore constrained to a narrow area underneath the chart.

For this experiment, we had the following within-person factors:

- **Visualization type** (3 levels): whether the data was visualized in a *bar chart*, *broken axis bar chart*, or *gradient bar chart*. See Fig. 11 for examples of these designs.
- **Truncation Level** (3 levels): where the y-axis of the visualization began: either at *0*, *25*, or *50%*. Note that, in this experiment, all three visualizations were visually identical when the truncation level was 0%.
- **Slope** (2 levels) : How much increase (or decrease) there was between the first and last values in the data, either 12.5% or 25%.
- **Data Size** (2 levels): whether there were *two* or *three* data values in the visualization. If there were three data values, the center value was at the midpoint of the first and last items, with a uniform random jitter between $[0, \frac{slope}{4}]$.

Participants saw one of each combination of factors, for a total of $3 \times 3 \times 2 \times 2 = 36$ stimuli, in a random order. Whether the value was increasing or decreasing was a random factor. As with the previous experiment, we included an initial set of 8 stimuli illustrating the full range of effect sizes in order to assist in calibrating the participants' subjective judgments, for a total of 44 total stimuli, but these calibration stimuli were excluded from analysis.

3.2.2 Hypotheses

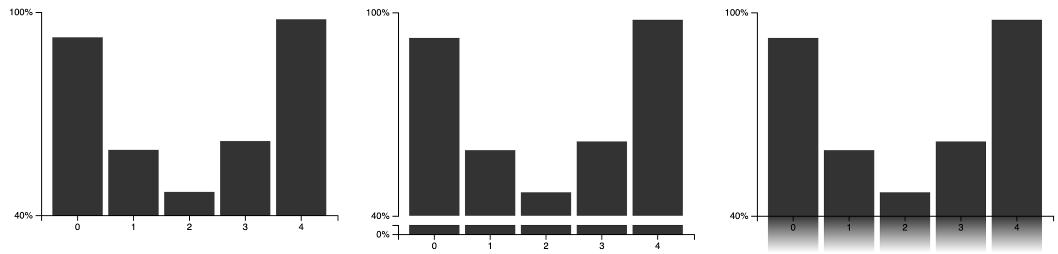
We had only a single hypothesis for this experiment: **Visual designs that indicate y-axis breaks or continuations would be perceived as having smaller effect sizes than standard bar charts**. We believed that these visual indications would make the truncation harder to ignore or overlook, and promote caution or reflection in judgments.

3.2.3 Results

We recruited 32 participants for this task (20 female, 12 male, $M_{age} = 29.0$, $SD_{age} = 11.7$). We paid participants $4 for this task, for an empirical effective hourly rate of $16/hour. Participants scored well on the Galesic and Garcia-Retamero graphical literacy scale ($M_{correct} = 10.8$, $SD_{correct} = 2$). Similar to the previous study, 25 (78%) participants correctly answered the scale item connected with y-axis truncation. Participants also scored highly on our engagement question, correctly labeling the direction of the effect ($M_{correct} = 0.99$, $SD_{correct} = 0.02$). One participant had performance more than three standard deviations from the mean (93.2%). The data from this participant was excluded from our analysis.

We conducted a repeated measures ANOVA testing the effect of truncation level, visualization type, data size, and their interactions on perceived severity. We built our model on the subset of trials where the truncation level was > 0, as those were the trials with visual differences between designs.

Our results fail to support our first hypothesis: **there was no significant difference between perceived severity among visualization**



(a) Bar Chart (b) Bar Chart with Broken Axis (c) Bar Chart with Gradient

Fig. 11: The three visualization designs in Experiment Two. A broken bar and axis legend indicates truncation in Figure 11b, whereas the continuation of the axis beyond the chart is indicated by a gradient in Figure 11c. When there is no truncation, breaking the axis and indicating that the bars continue is inappropriate. In those cases, the alternative designs devolve into traditional bar charts, as in Figure 11a.

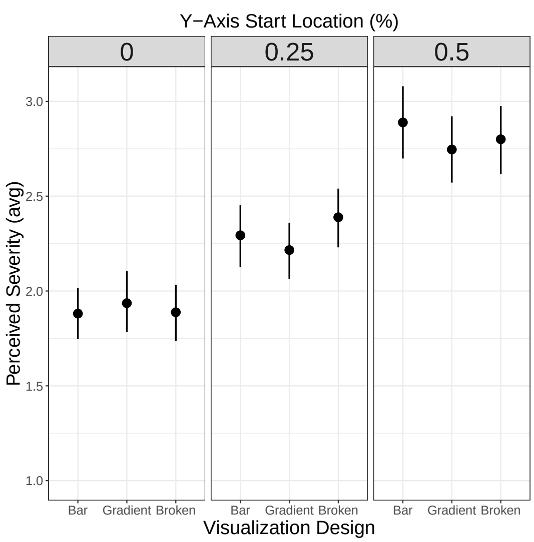


Fig. 12: Results from Experiment Two. While broken axes may *indicate* that a y-axis is truncated, and a gradient fill may *connote* that the bars extend beyond the visualized chart area, neither intervention had a consistent impact on perceived severity; increased axis truncation resulted in similar increases in perceived severity. Error bars represent 95% bootstrapped confidence intervals of the mean. Note that when there is no truncation (the y-axis begins at 0%), all three designs were visually identical.

designs ($F(2,60) = 3.1$, $p = 0.05$). A post-hoc pairwise t-test with a Bonferroni correction failed to find any significant difference between visualization designs. Figure 12 illustrates the performance of all three designs across different truncation levels.

Additionally, we only coded 31% of participants as having specifically mentioned y-axis truncation in their post-task free text responses, compared to 47% in the first experiment. It is possible that the alternative conditions made the truncation of the axis so "obvious" that it was not felt necessary to comment upon, but we had no specific hypothesis to this effect. As in the first experiment, participants who mentioned the y-axis reported less severe effect sizes than those who did not across all truncation levels.

4 Discussion

Our experimental results suggest that truncating the y-axis has a consistent and significant impact on the perceived importance of effect sizes. This exaggeration occurs in both line charts and bar charts, as well as in bar charts that visually indicate either broken axes or the continuation of bars beyond the bounds of the chart.

These results suggest that, regardless of differences in the visual metaphors or encodings of line charts, there does not appear to be a significant *practical* difference in the impact of exaggeration due to y-axis truncation across different visualization: the visual design alone is not sufficient to shape guidelines around how to define charts. For the same data, the narrower the range of values in the y-axis, the larger the visual effect size and so the larger the subjective effect size. Different designs might provide more visual *indications* that this exaggeration is occurring, but did not substantially alter the reported impact of the exaggeration.

Moreover, we cannot *rely* on visual indicators of broken or truncated axes to counteract the exaggeration caused by y-axis truncation. Much less visual real estate is dedicated to axis labels than the actual encoded data. Viewers can and do overlook or misread these labels. Even if they do explicitly notice manipulations in the y-axis (as did many of our participants), the sheer visual effect of exaggerating the height of bars or the steepness of lines has an impact nonetheless. Judgments about effect size (at least for our experimental context) may be *visual* rather than *mathematical or statistical* judgments. In which case merely indicating that truncation has occurred, even in a prominent and unambiguous way, may not be sufficient to "de-bias" viewers of truncated charts.

However, we resist the interpretation of our experimental results to mean that, as Huff suggests [19], all charts with quantitative axes should include 0. An important consideration of our experimental measure is that there is no *a priori*, domain-agnostic *ground truth* for how severe, important, or meaningful an effect size ought to be. Rather, we interpret these results as meaning that there is no obvious way for designers to *relinquish the responsibility* of considering effect size in their charts. There is not an unequivocal dichotomy of "honest" and "dishonest" charts (for instance, as presented in Figure 1). Rather, designers have a great deal of power over the message that viewers take away from a chart, and ought to employ that power responsibly.

In particular, designers ought to consider:

1. **What is the full range of the quantitative data, in *principle*?** Temperature cannot dip below 0K, stock prices can't drop below 0, percentages of a whole cannot exceed 100%, etc. Axes should not depict nonsensical or undefined values.

2. **What is the full range of the quantitative data, in *practice*?** Observed data has limited extent. At the very least axes should not be *smaller* than the extent of the data one wishes to visualize.

3. **What range of differences are important for the viewer to detect?** A difference in global temperature of 1 degree can be tremendously important in climate change contexts. A difference in a few millimeters in a factory process control context is likewise potentially very important. If a significant difference is invisible or visually insignificant in the chart, then the chart fails at its task of communicating the important features of the data.

4. **Is there a "natural" or task-relevant baseline?** Revenue data is often compared to revenue in the previous quarter. Climate data or medical data are often compared to benchmark values or expected ranges. In many cases, there exists a baseline value of expectation that can be used to provide a more meaningful "zero" than a literal 0 y-axis.

5. **Are there more direct visual encodings that can highlight task-relevant differences?** In the presence of a task-relevant baseline, designers can choose to overlay or explicitly encode difference from this baseline rather than raw values. Srinivasan et al. [36] had success using these "difference overlays" to communicate both value and difference in bar charts. In general, if the viewers are interested in a rate of change, then it is often possible to plot this rate of change directly, rather than rely on variability in bar heights to act as a proxy.

The answers to these questions entail knowledge of the domain, the audience, and the analytical objective of the chart. In short, they are questions about intent and communication, rather than questions about graphical perception.

4.1 Limitations & Future Work

Our experiments focus on a limited set of designs and employ a subjective rating metric to assess the impact of truncation on perceived effect size. We also focus on detecting the relative difference in subjective effect size across a few different levels of truncation, rather than attempting to fully model the complex relationship between slope, axis truncation, and perceived severity. It remains to future work to develop a full response curve for the interplay of these variables.

Similarly, we tested only two potential designs for indicating axis truncation in bar charts as representatives of common classes of design interventions. Even of the designs we considered, we focused only on methods for static charts. Other methods using animation or interaction (such as in Ritchie et al. [31]) could result in different patterns of subjective judgments by allowing the viewer to switch between truncated and non-truncated axes.

Wishing to avoid the complications involved in narrative or domain-focused crowdsourced studies (as discussed in Dimara et al. [12]) our designs were presented in a relatively context-free manner. We believe that analysts in different domains have different internal models of effect size severity that would therefore not be captured in our results. We anticipate that different data domains and analytical contexts can impact the perceived importance or severity of effect sizes.

Connected with the issue of domain relevance is that of authority: visualizations from different sources or presented with different levels of perceived expertise or authority could potentially produce differing patterns of judgment in different audiences. While we are actively working in understanding how visualizations persuade, and the rhetorical strategies that designers use to increase the persuasive power of visualization, a quantitative study of the persuasive power of y-axis truncation (especially for decision-making tasks) remains future work.

4.2 Conclusion

Experts in the field of information visualization and statistical graphics have produced conflicting advice on how harmful it is to start the y-axis of a chart from values other than 0. This conflict has often centered on the distinction between line graphs and bar charts, or on best practices for depicting axis breaks. We find that truncating the y-axis produces consistent patterns of exaggeration across bar charts and line charts, and across different strategies for indicating axis breaks. Exaggeration is not only the purview of bar charts.

Moreover, in many cases the "deceptive" practice of axis truncation is one of the most straightforward ways of communicating important and consequential effects and trends in the data. We therefore advise designers of visualizations to consider effect sizes in a domain- and audience-specific way, and choose chart designs that are effective for their communicative goals.

References

[1] R. J. Andrews. Tear up your basline, 2019. `https://medium.com/data-visualization-society/tear-up-your-baseline-b6b68a2a60f1`.

[2] G. C. Baylis and J. Driver. One-sided edge assignment in vision: 1. figure-ground segmentation and attention to objects. *Current Directions in Psychological Science*, 4(5):140–146, 1995.

[3] C. R. Berger. Slippery slopes to apprehension: Rationality and graphical depictions of increasingly threatening trends. *Communication Research*, 32(1):3–28, 2005.

[4] C. Bergstrom and J. West. The principle of proportional ink, 2017. `https://callingbullshit.org/tools/tools_proportional_ink.html`.

[5] W. C. Brinton. *Graphic Presentation*. Brinton Associates, 1939.

[6] A. Cairo. *How Charts Lie*. W.W. Norton, 2019.

[7] Z. Chen. Object-based attention: A tutorial review. *Attention, Perception, & Psychophysics*, 74(5):784–802, 2012.

[8] W. S. Cleveland, P. Diaconis, and R. McGill. Variables on scatterplots look more highly correlated when the scales are increased. *Science*, 216(4550):1138–1141, 1982.

[9] W. S. Cleveland, M. E. McGill, and R. McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402):289–300, 1988.

[10] M. Correll and J. Heer. Black hat visualization. In *Workshop on Dealing with Cognitive Biases in Visualisations (DECISIVe), IEEE VIS*, 2017.

[11] J. Díaz, O. Meruvia-Pastor, and P.-P. Vázquez. Improving perception accuracy in bar charts with internal contrast and framing enhancements. In *2018 22nd International Conference Information Visualisation (IV)*, pp. 159–168. IEEE, 2018.

[12] E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE transactions on visualization and computer graphics*, 2018.

[13] J. Duncan. Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113(4):501, 1984.

[14] R. Egly, J. Driver, and R. D. Rafal. Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123(2):161, 1994.

[15] S. L. Franconeri. The nature and status of visual resources. *Oxford handbook of cognitive psychology*, 8481:147–162, 2013.

[16] M. Galesic and R. Garcia-Retamero. Graph literacy: A cross-cultural comparison. *Medical Decision Making*, 31(3):444–457, 2011.

[17] K. Hara, A. Adams, K. Milland, S. Savage, C. Callison-Burch, and J. P. Bigham. A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 449. ACM, 2018.

[18] J. Heer and M. Agrawala. Multi-scale banking to 45 degrees. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):701–708, 2006.

[19] D. Huff. *How to lie with statistics*. WW Norton & Company, 1993.

[20] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics*, 17(12):2231–2240, 2011.

[21] B. Jones, 2015. `https://twitter.com/DataRemixed/status/671832482279395328`.

[22] H.-K. Kong, Z. Liu, and K. Karahalios. Frames and slants in titles of visualizations on controversial topics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 438. ACM, 2018.
[23] R. Kosara. 3d bar charts considered not that harmful, 2016. https://eagereyes.org/blog/2016/3d-bar-charts-considered-not-that-harmful.
[24] S. Leo. Mistakes, we've drawn a few, 2019. https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368.
[25] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. Formalizing visualization design knowledge as constraints: actionable and extensible models in draco. *IEEE transactions on visualization and computer graphics*, 25(1):438–448, 2019.
[26] National Review. The only #climatechange chart you need to see, 2015. https://twitter.com/nro/status/676516015078039556?lang=en.
[27] S. O'Brien and C. Lauer. Testing the susceptibility of users to deceptive data visualizations when paired with explanatory text. In *Proceedings of the 36th ACM International Conference on the Design of Communication*, p. 7. ACM, 2018.
[28] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini. How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 1469–1478. ACM, 2015.
[29] E. Peer, L. Brandimarte, S. Samat, and A. Acquisti. Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017.
[30] J. Peltier. Broken y axis in an excel chart, 2011. https://peltiertech.com/broken-y-axis-in-excel-chart/.
[31] J. Ritchie, D. Wigdor, and F. Chavalier. A lie reveals the truth: Quasimodes for task-aligned data presentation. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. To appear.
[32] D. Shere and H. Groch-Begley. A history of dishonest fox charts. https://www.mediamatters.org/research/2012/10/01/a-history-of-dishonest-fox-charts/190225.
[33] S. Shomstein. Object-based attention: strategy versus automaticity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(2):163–169, 2012.
[34] D. Skau, L. Harrison, and R. Kosara. An evaluation of the impact of visual embellishments in bar charts. In *Computer Graphics Forum*, vol. 34, pp. 221–230. Wiley Online Library, 2015.
[35] C. Skelton. Bar charts should always start at zero. but what about line charts?, 2018. http://www.chadskelton.com/2018/06/bar-charts-should-always-start-at-zero.html.
[36] A. Srinivasan, M. Brehmer, B. Lee, and S. M. Drucker. What's the difference?: Evaluating variations of multi-series bar charts for visual comparison tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, p. 304. ACM, 2018.
[37] D. A. Szafir. The good, the bad, and the biased: five ways visualizations can mislead (and how to fix them). *interactions*, 25(4):26–33, 2018.
[38] J. Talbot, J. Gerth, and P. Hanrahan. Arc length-based aspect ratio selection. *IEEE transactions on visualization and computer graphics*, 17(12):2276–2282, 2011.
[39] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press; 2nd edition, Cheshire, CT, USA, 2001.
[40] J. Wagemans, J. H. Elder, M. Kubovy, S. E. Palmer, M. A. Peterson, M. Singh, and R. von der Heydt. A century of gestalt psychology in visual perception: I. perceptual grouping and figure–ground organization. *Psychological bulletin*, 138(6):1172, 2012.
[41] Y. Wang, F. Han, L. Zhu, O. Deussen, and B. Chen. Line graph or scatter plot? automatic selection of methods for visualizing trends in time series. *IEEE transactions on visualization and computer graphics*, 24(2):1141–1154, 2018.
[42] Wikipedia contributors. Misleading graphs. https://en.wikipedia.org/wiki/Misleading_graph#Truncated_graph, 2019.
[43] C. Xiong, L. van Weelden, and S. Franconeri. The curse of knowledge in visual data communication. In *DECISIVe: Workshop on Dealing with Cognitive Biases in Visualizations. IEEE VIS*, 2017.
[44] J. Zacks and B. Tversky. Bars and lines: A study of graphic communication. *Memory & Cognition*, 27(6):1073–1079, 1999.
[45] C. Ziemkiewicz and R. Kosara. The shaping of information by visual metaphors. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), 2008.
[46] C. Ziemkiewicz and R. Kosara. Preconceptions and individual differences in understanding visual metaphors. In *Computer Graphics Forum*, vol. 28, pp. 911–918. Wiley Online Library, 2009.