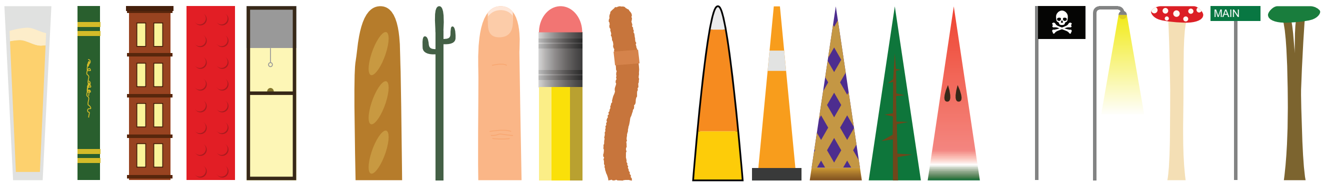
Eurographics Conference on Visualization (EuroVis) 2017
B. Kozlíková, T. Schreck, and T. Wischgoll (Editors)

Short Paper

Readability and Precision in Pictorial Bar Charts

Drew Skau[1] Robert Kosara[2]

[1]Pivotus Ventures
[2]Tableau Research



(a) *Rectangular bars* **(b)** *Rounded-top bars* **(c)** *Triangular bars* **(d)** *T-shaped bars*

Figure 1: *Do elaborate pictorial embellishments of bar charts lead to reduced precision when reading them?*

Abstract

Bar charts embellished with unique artistic styles, or made to look like real objects, are common in information graphics. Embellishments are typically considered detrimental to readability and accuracy, since they add clutter and noise. Previous work has found that some of the shapes used, like rounded tops, triangles, etc., decreased accuracy when judging relative and absolute sizes, while T-shaped bars even showed a slight increase relative to the basic bar chart.

In this paper, we report on a study that adds pictorial elements to bar charts of four different shapes tested previously, thus also including the elements of color and texture.

We find that pictorial bar charts reduce accuracy, but not beyond the effect already observed for their shape. They also do not significantly increase response time. Embellished bar charts may not be as problematic as commonly assumed.

1. Introduction

Information graphics often use highly embellished versions of common chart types like bar charts. They are often added to provide context on the topic of the chart, make it more memorable, and undoubtedly also to make the chart more visually attractive. How much common chart embellishments decrease a chart's readability has not been systematically studied.

Embellishments are stylistic modifications made to a chart that do not add to the representation of data. While there is no general definition of what should be considered an embellished chart, we consider any deviation from a rectangular bar an embellishment.

Changing the shape of bars in ways that the information visualization community generally recommends against is quite common: non-rectangular bars (in particular triangles), rounded tops, etc. The use of color and images in bars is also common and tends to be discouraged by the visualization community.

Research recently found that monochrome pictorial embellishments in ISOTYPE charts did not have a negative impact on accuracy or reading time [HKF15]. Our own previous work showed that changing the shape of bars in a bar chart impacts people's precision when judging their heights [SHK15]. Both studies were limited to just shapes though, with no use of color, texture, or images.

Color and imagery introduce additional visual complexity and potential for confusion and distraction. We therefore tested their impact in a study, which we report on below.

2. Related Work

Any discussion of work on bar charts has to include the seminal work by Cleveland and McGill, who explored differences between different bar chart configurations for different tasks [CM84]. Heer and Bostock [HB10] replicated this study on Amazon's Mechanical Turk platform. Talbot et al. further examined the impact of different bar chart configurations on accuracy [TSA14].

The interplay between bars has also been shown to have an impact on value reading accuracy, as shown by a study by Zacks et al. [ZLTS98]. They examined bars with varying degrees of representation from a single line to a projection of a 3D bar, finding that while the perspective cues reduced accuracy, interplay from the neighboring elements had a larger impact.

More complex tasks have also been evaluated. Estimated averages of all bars in a bar chart have been proven to be low by Newman et al. [NS12]. Work by Elzer et al. [EGCH06] has outlined the need for an understanding of the perceptual effects occurring with bar charts in their model of perceptual task effort.

Embellished charts have only recently been a topic of research. Research by Bateman et al. [BMG*10] finds that embellished charts can still be read accurately, and are more memorable than plain charts. Borkin et al. [BVB*13] and Borgo et al. [BARM*12] similarly found that embellishments and recognizable pictorial elements can increase chart memorability. ISOTYPE charts were specifically designed to use pictorial elements to make data easier to understand, and recent research has found them to be as accurate as (and more memorable than) plain bar charts [HKF15].

In a previous study [SHK15], we examined the impact of bar chart embellishments on reading accuracy and found that some embellishment shapes can be detrimental. That work only used shapes, however, and did not test images or color.

3. Embellishments and Hypotheses

Our previous study outlines a set of bar chart shape embellishments that occur frequently in the infographic design space. The designs varied shape, but did not use color or texture. We build on this design by adding pictorial content to the basic bars. This impacts their appearance in a number of ways. First, the images may distort the purely geometrical base shape of the bar (Figure 1), and potentially add a third dimension. Some of the designs make shape elements more or less obvious, such as the rounded top in the baguette versus the pencil; the T-shape can be more or less pronounced, e.g., in the Jolly Roger flag versus the street lamp. Different colors also impact the perceived weight of the chart [ZK10].

We picked the four chart shapes with the least negative effect on accuracy from the previous study to base ours on:

- Rectangular bars
- Rounded corner charts
- Triangle charts
- Capped bars

In this study, we hoped to discover if embellishments complete with colors and internal structures increased or decreased the impact of the shape-based embellishments on reading accuracy. We developed the following hypotheses:

1. Any additional embellishment with color and shape will lead to higher error compared to a solid bar of the same shape.
2. The higher complexity of pictorial bars will require more time to read.
3. The previous study had found T-shaped bars to be no different from base bars for absolute judgments. We hypothesize that they will not be impacted by further embellishments.

4. Study

We took the four primary embellishment shape categories and ran a within-subject study on Mechanical Turk, using an experiment design similar to our earlier study [SHK15]. We measured accuracy with the same question structure as our earlier work, with tasks for relative comparisons between bars, and absolute comparisons of one bar to the chart's y-axis.

We compared all embellishments against a baseline chart style with black rectangular bars. All charts used the same axes as in the previous study, with relative questions having no vertical axis, and absolute questions having a vertical axis ranging from 0 to 100.

4.1. Materials

For each shape type we used, we created five different images for a total of twenty (Figure 1). In order to protect the recognizability of each image as much as possible, different images are drawn with different scaling methods.

Some images can be vertically stretched and still look very much like the objects they represent, however, for some images this reduces their recognizability. We therefore decided to scale some of the objects and crop others from the bottom of the bar to achieve the desired bar height.

We created five images for each of the four embellishment shapes, plus one baseline bar chart as a control condition.

The study consisted of two sections, one asking relative questions, the other absolute. In each section, we used each individual image type twice, and also included the baseline bar chart four times. This yielded $4 \cdot 5 \cdot 2 + 4 = 44$ questions for each section, and 88 total. The rectangular charts (Figure 1a) were made up of three scaled objects (pint glass, book, and window) and two cropped objects (building/toy block). The rounded top charts (Figure 1b) were all cropped objects (baguette, cactus, finger, pencil, and worm). The triangular charts (Figure 1c) were all scaled (candy corn, construction cone, blueberry pie, coniferous tree, and watermelon). The T-shaped charts (Figure 1d) were all cropped objects (Jolly Roger flag, lamp post, mushroom, street sign, and deciduous tree).

The maximum value for a bar was 97, the minimum value 3, for a maximum difference of 94 (the smaller bar being ~3% of the larger). Charts were roughly 5-7 cm tall on a typical computer, with variations depending on monitor pixel density, resolution, and browser zoom level. No efforts were made to standardize absolute sizes.

Code and results for this study can be found at `https://github.com/dwskau/embellished-bar-charts`.

4.2. Procedure

Participants began with a page introducing the study and a short demographic form. The next page provided instructions on how to answer the study questions, and briefly discussed the two part structure of the study. The first block of questions ended with an intermission page, giving participants an opportunity for a break before proceeding with the second half of the study.

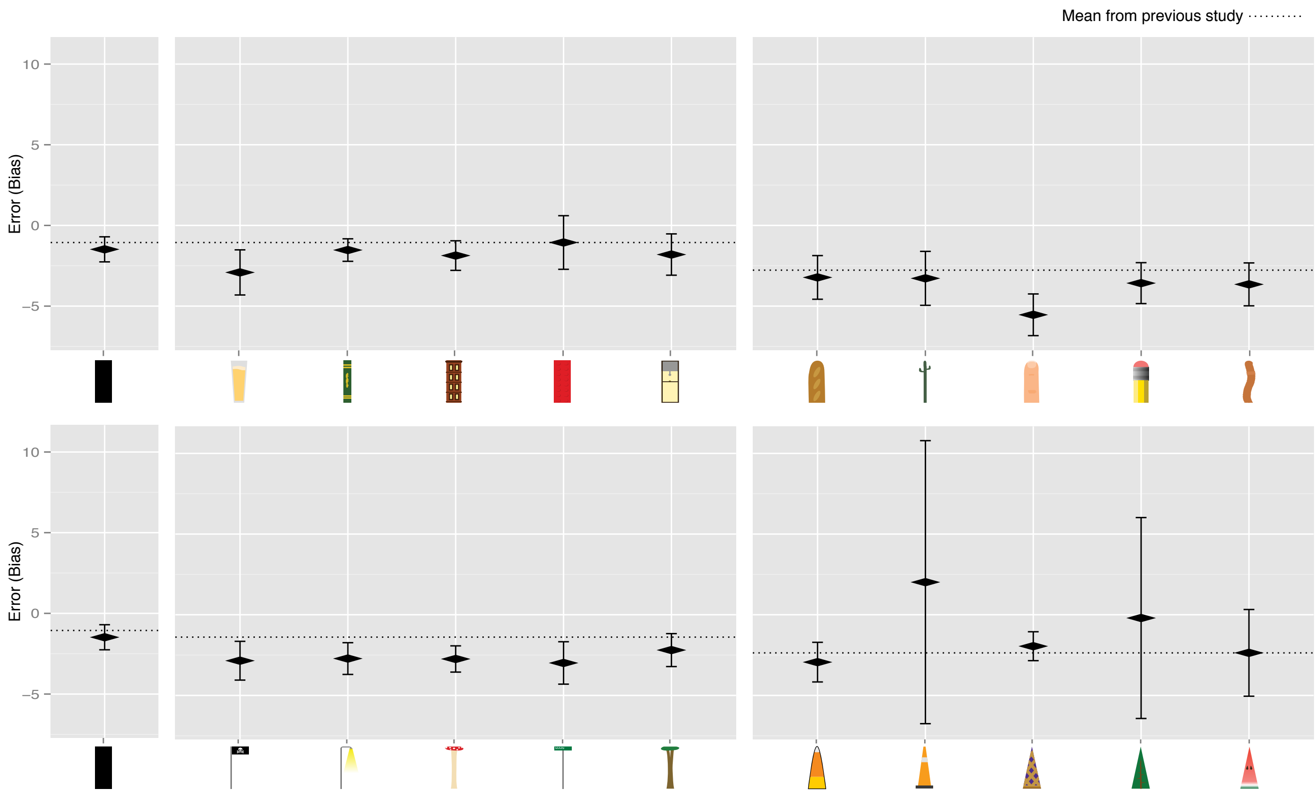


Figure 2: *Signed error for absolute value questions for each embellishment, broken down by class. Error bars show 95% confidence intervals. Dotted lines show the previous study's means for each bar class. The difference between classes is statistically significant (ANOVA:* $F(4,390) = 3.354, p = 0.0102$*), but the differences between the pictorial embellishments within each class are not.*

We randomized the order of the blocks of question types, as well as the order of the embellishment shapes and images within each block to ensure there were no learning effects from one question or embellishment shape or image to another.

We also avoided referring to the charts in the materials as bar charts, instead calling them simply charts, and referring to individual bars just by their labels. This allowed participants to come to their own conclusions on how to interpret the charts. This mimics the experience of most infographics, with no instructions provided to assist with the interpretation of the graphics presented.

4.2.1. Question Types

We adopted the question types used in the previous embellishment study and tested the accuracy of comparisons between bars and reading single bar values.

a) *In the chart below, what is the value of A?*
b) *In the chart below, what percentage is B of A?*

We chose to not address the subject matter of each chart's images with the questions. Our goals were to test accuracy, not other effects like comprehension or memorability.

Each question type was asked twice for each bar image, and four times for the baseline chart for a total of 44 questions in each section and a grand total of 88 questions in the study. We did not display a y-axis in the relative questions (question type *b*) to discourage participants from mathematically computing the percentage using the absolute heights of each bar.

4.3. Results

We used Amazon's Mechanical Turk to recruit 81 participants for our study. Each participant was paid US $2.50, with an average completion time of 16 minutes for an hourly rate of $9.07. Of those participants, 37 identified themselves as female and 44 as male. There were seven in the 18-24 age range, 18 between 25-29, 33 between 30-39, 16 between 40-49, five between 50-59, and two older than 60. Education levels were also fairly wide ranging, with 30 who finished high school, 42 with a bachelor's degree, six with a master's, and three reporting as other.

We eliminated the results of two participants who had error rates almost twice as high as the average. Both were males with high school degrees. This resulted in a total of 79 participants.

To report results, we use signed error (*answer-correct*) and absolute error (absolute value of signed error). Signed error is useful to gauge over- and underestimation, with positive values for overestimation and negative ones for underestimation. Absolute error is

	Absolute Questions		Relative Questions	
Embellishment	Mean	95% CI	Mean	95% CI
Baseline	4.278	±0.472	5.539	±1.875
Rectangular	4.678	±0.451	5.277	±0.608
Triangular	6.863	±2.185	7.062	±1.364
T-shaped	4.858	±0.361	6.506	±1.227
Rounded tops	5.815	±0.473	6.341	±0.519

Table 1: *Absolute error, means and 95% confidence intervals by question and embellishment type. Differences statistically significant for absolute (ANOVA: $F(4,390) = 3.254, p = 0.0121 < 0.05$), but not relative questions (ANOVA: $F(4,390) = 0.668, p = 0.615$).*

a better measure for precision, since it measures the distance from the correct value without the averaging-out effect between over- and underestimates the signed error suffers from.

4.3.1. Absolute Judgements

We divide our analysis into absolute and relative questions.

The answers for the absolute judgement questions, where participants were asked to estimate the value of one bar using the y-axis as a reference, generally led to underestimations. (Figure 2). Participants underestimated the values for all of the embellishment images, although their estimations were in line with the results from the shape-based embellishments from the previous study.

There were significant differences observed in both the mean signed error ($p = 0.0102$) and the mean absolute error ($p = 0.0121$) between different shapes (Table 1), however the different pictorial embellishments within each shape group did not differ in a statistically significant way. There was no significant difference in mean absolute error between any of the images of a given embellishment type, and the means of each image were largely the same as the means for their embellishment shape.

We also found no significant difference in response times for any of the embellishment images, consistent with the previous study (ANOVA: $F(4,390) = 0.11, p = 0.967$).

4.3.2. Relative Judgements

The answers for the relative judgement questions, where participants were asked to estimate the percentage one bar was of another, were not consistently under- or overestimated. Mean absolute error, however, was higher than mean absolute error of the absolute judgements. This confirms that this is a difficult task for people.

As with the absolute judgements, we observed significant differences in the mean signed error ($p = 0.0154$) between different shapes, but not the mean absolute error ($p = 0.615$, Table 1). Similarly, there were no significant differences in mean absolute error between embellishment images within each embellishment shape.

We did not find a significant effect of shape on response time, though since the p-value is just over 0.05, we feel that this data is worth reporting (Table 2). The difference in time is rather small, however: half a second, or about 7%, between the fastest and slowest charts.

All of the above results are consistent with the previous study.

Embellishment	Mean (s)	95% CI
Baseline	6.51	±0.99
Rectangular	6.97	±0.75
Triangular	6.92	±1.03
T-shaped	6.81	±0.61
Rounded tops	7.03	±0.84

Table 2: *Response time for relative questions by embellishment type (ANOVA: $F(4,390) = 2.297, p = 0.0586$).*

5. Discussion

While we found statistically significant differences between different shapes – confirming the effects found in the previous study [SHK15] –, there appears to be no further significant effect from adding pictorial elements to them. Neither accuracy nor response time were impacted by the pictorial embellishment.

As a result, we believe the impact of pictorial elements on bar charts to be overstated. Other shapes are not impacted negatively beyond the effect that the shape alone already has.

5.1. Recommendations

Bar shapes and the pictorial elements of bars usually go hand-in-hand in practice. Designers thus need to carefully look at the shape of the embellished bar charts they create.

Our findings lead us to concrete advice for designers of embellished bar charts in information graphics and other contexts:

- Avoid bars without a strong horizontal mark indicating the top of the bar, but bars shaped like rectangles or Ts are okay.
- Ensure strong boundary contrast so the edges of the bar are clearly visible.
- Within the bounds of the bars, feel free to use any variety of colors, textures, and shapes, as this has little impact people's ability to read the chart accurately.

This also provides some promising news for designers, as they should be able to reap the memory benefits found in recent work [BARM*12, BVB*13] without the negative effects associated with embellished bar charts.

6. Conclusions

In this paper we present a crowdsourced experiment to investigate the impact of pictorial chart embellishments on the accuracy of absolute and relative judgements in bar charts.

The results confirm findings from a previous study that bar chart shape embellishments do indeed have an impact on how well the data within the chart can be accurately read. However, pictorial embellishments within a bar's shape have no discernible impact on how precisely the data within the chart can be read.

The design space of information graphics is certainly much larger than what we were able to test in this study. More work is necessary to test more chart shapes, embellishments, colors, etc. This first step points to bar chart embellishments being less harmful than often assumed, however, at least as long as the shape of the bars is kept.

References

[BARM*12] BORGO R., ABDUL-RAHMAN A., MOHAMED F., GRANT P. W., REPPA I., FLORIDI L., CHEN M.: An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2759–2768.

[BMG*10] BATEMAN S., MANDRYK R. L., GUTWIN C., GENEST A., MCDINE D., BROOKS C.: Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010), ACM, pp. 2573–2582.

[BVB*13] BORKIN M. A., VO A. A., BYLINSKII Z., ISOLA P., SUNKAVALLI S., OLIVA A., PFISTER H.: What makes a visualization memorable. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2306–2315.

[CM84] CLEVELAND W. S., MCGILL R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association 79*, 387 (1984), 531–554.

[EGCH06] ELZER S., GREEN N., CARBERRY S., HOFFMAN J.: A model of perceptual task effort for bar charts and its role in recognizing intention. *User Modelling and User-Adapted Interaction 16*, 1 (2006), 1–30.

[HB10] HEER J., BOSTOCK M.: Crowdsourcing graphical perception: using Mechanical Turk to assess visualisation design. *ACM Human Factors in Computing Systems* (2010), 203–212.

[HKF15] HAROZ S., KOSARA R., FRANCONERI S. L.: ISOTYPE Visualization – Working Memory, Performance, and Engagement with Pictographs. In *Proceedings CHI* (2015), pp. 1191–1200.

[NS12] NEWMAN G. E., SCHOLL B. J.: Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic bulletin & review 19*, 4 (2012), 601–607.

[SHK15] SKAU D., HARRISON L., KOSARA R.: An evaluation of the impact of visual embellishments in bar charts. *Computer Graphics Forum (Proceedings EuroVis) 34*, 3 (2015), 221–230.

[TSA14] TALBOT J., SETLUR V., ANAND A.: Four Experiments on the Perception of Bar Charts. *IEEE Transactions on Visualization and Computer Graphics* (2014).

[ZK10] ZIEMKIEWICZ C., KOSARA R.: Laws of Attraction: From Perceptual Forces to Conceptual Similarity. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1009–1016.

[ZLTS98] ZACKS J., LEVY E., TVERSKY B., SCHIANO D. J.: Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied 4*, 2 (1998), 119.