Icons are Best: Ranking Visualizations for Proportion Estimation

Zhengliang Liu* Melanie Bancilhon† Alvitta Ottley‡

Washington University in St. Louis

ABSTRACT

Estimating proportions is a common visualization task. In this paper, we present the findings of a large-scale experiment that compares the effectiveness of helping users in estimating proportions across six common visualization designs. We analyzed 406 participants' decisions over real monetary gains, resulting in 10,150 observations. Participants were most accurate with an *icon array* design and least accurate with an area proportioned *circle* design. They also consistently overestimated low probabilities and underestimated high probabilities across all charts. Using our findings, we rank the representations based on users' ability to accurately estimate the underlying proportions, and provide evidence-based guidelines for visualization designers.

Index Terms: Human-centered computing—Visualization—Empirical studies in visualization techniques;

1 INTRODUCTION AND BACKGROUND

Advances in technology and visualization research enable designers to create a large variety of charts quickly and easily for numerous domains including business, medicine, and government. In each situation, stakeholders commonly use visualizations to estimate proportions and to aid sound decision-making. However, it is often unclear how to choose among the various design options and select the right chart for this task.

Prior work has investigated how the choice of visualization impacts judgments. For example, Cleveland and McGill provided a taxonomy and ranking of necessary judgments that people make when decoding quantitative information from charts: Position along a common scale, position along identical but non-aligned scales, length, angle, slope, and area [2]. They found that people were best at decoding quantitative information when the data was encoded in the position of a visual mark and worst when the encoding used area.

In addition to measuring comparison judgments, Simkin and Hastie tested subjects' ability to make proportion (part-of-the-whole) judgments with different types of encoding [4]. Participants estimated what percentage of the entire chart did an indicated portion represent. The results showed no measurable differences between the angle (pie) and position (bar) estimates, but angle and position were superior to length.

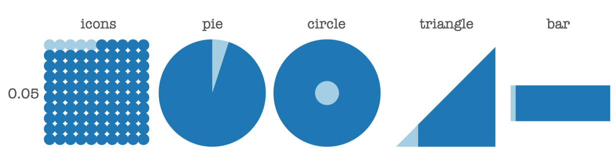

Figure 1: The visualization conditions used in the study.

*e-mail: zhengliang@wustl.edu
†e-mail:mbancilhon@wustl.com
‡e-mail:alvitta@wustl.edu

The previous work in ranking visualizations' effectiveness of estimating proportions, however, did not consider representations with countable, discrete elements, such as the icon array. Icon arrays are effective in communicating and supporting decision-making given their countable and pictorial nature [3]. In this paper, we want to directly investigate what visualization designs help users make the best estimates of proportions with a special attention to icon arrays. We hypothesize that icon arrays are best in helping users estimate proportions and percentages graphically.

We report the findings of a crowd-sourced experiment that tested six different representations of lottery risks, including a text-only condition as our control and five simple visualization designs (icons, pie, circle, triangle, and bar). The lottery scenario is a useful test-bed because it provides a straightforward framework while motivating users to make their best estimates with real monetary incentives. Participants were asked to decode the visualizations and estimate the probability values. They then selected their lottery decisions. We analyzed 406 participants' decisions over real monetary gains which resulted in 10,150 observations.

2 EXPERIMENT

Inspired by Bruhin et al.'s experiment design [1], we presented participants with 25 two-outcome lotteries ($n = 2$) that were choices between risky and certain gains. We used a points system for our compensation where 1 point equalled \$0.01. The probabilities, p_i, were drawn from the set $P = \{.05, .1, .25, .5, .75, .9, .95\}$ and the outcomes x_1 and x_2 ranged from 0 to 150 points. Each lottery sheet contains 20 outcomes that ranged from x_1 to x_2 (see Figure 2). The lottery mechanism motivates the participants to make their best efforts in estimating the probabilities. At the end of the experiment, we randomly selected one row for each of the 25 lottery sheets, and the participant's choice in that row determined the bonus. For example, if we randomly drew the row highlighted in green in figure 2. Suppose that the participant indicated a preference for the guaranteed payment; her bonus for that sheet would be 400 points (\$4.00). Now, suppose instead, the participant opted to enter the lottery, we would simulate the lottery to determine her payment. The bonus was the sum of winnings for the 25 sheets.

2.1 Participants and Procedure

We recruited 406 participants from Amazon Mechanical Turk. For each sheet, participants first entered their best guess for the corresponding probabilities. Then, they selected one of the radio button options and the system would populate the others. To prevent potentially biasing, we used a donut chart for the tutorial which was not a visualization condition in the study. The order of the 25 sheets was counterbalanced to prevent ordering effects.

2.2 Measures

Each participant completed 25 lottery sheets, resulting in 10,150 observations. During the experiment, we recorded the participants' probability estimates and preference selections. The experiment included the following variables:

- **7 Probability values:** $\{.05, .1, .25, .5, .75, .9, .95\}$
- **6 visualizations:** $\{none, icons, pie, circle, triangle, bar\}$

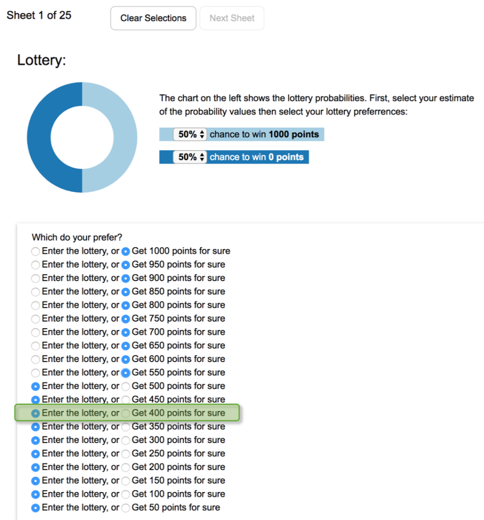

Figure 2: An example of the lottery decision sheet used in our study.

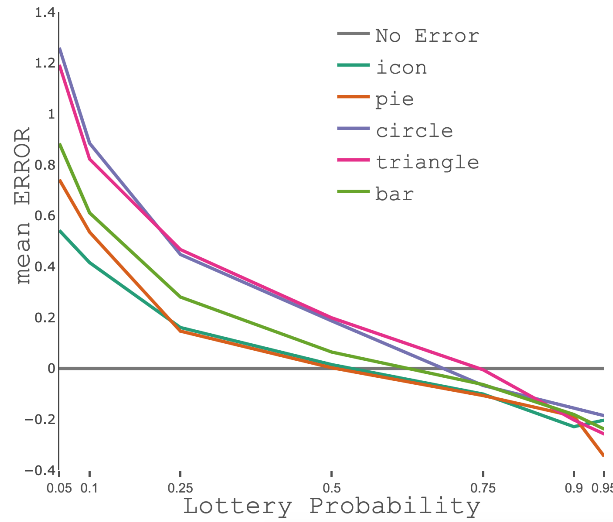

Figure 3: The average ERROR for all charts across the probability values.

- **EXACT:** Binary true or false if the participant's probability estimate matches the true probability.
- **ERROR:** The difference between the participant's probability estimate and the true probability value as a log ratio.

$$ERROR = \log(\hat{p}_i / p_i),$$

where $\hat{p}_i$ is participant's estimate of the probability value and p_i is the true probability.

3 Results and Discussion

We found that participants estimated the correct probability 40% of the time. Subjects in the *icons* condition produced 72.9% accurate estimates, while *pie*, *circle*, *triangle*, and *bar* yielded 59.8%, 10.6%, 13.7%, and 39.4% respectively. A chi-squared test revealed significant differences in the accuracy of participants across the five conditions ($\chi^2(4, N = 8475) < .001$). Pairwise chi-squared tests with a Bonferroni-adjusted alpha ($\alpha = 0.005$) uncovered that accuracy with all but two pairs of charts were statistically different (p-values less than .001). We found no significant differences between ratios of correct estimates within the *circle* and *triangle* groups ($\chi^2(1, N = 3400) = .006$).

The percentage of correct estimates across probability values $\{.05, .1, .25, .5, .75, .9, .95\}$ were 39.3%, 39.9%, 42.7%, 52.8%, 42.8%, 45.4%, 44.7%, and 44.8% respectively. A chi-squared test revealed significant differences in the accuracy across the range of probability values ($\chi^2(6, N = 8475) < .001$). Subsequent pairwise chi-squared tests with a Bonferroni-adjusted alpha ($\alpha = 0.0023$) found that, across all charts, participants were significantly more accurate at estimating a probability value of .5 than any of the other probability values (p-values were all less than .0001).

3.1 The Impact of Visualization on Estimating Proportions

For a more fine-grained analysis, we inspected the errors participants made. Figure 3 summarizes the *ERROR* values for each visualization condition across the different probability values. We saw a systematic tendency for participants to overestimate small probabilities and underestimate large probabilities. Consistent with the *EXACT* finding, we found that participants were least likely to make errors when estimating moderate probabilities.

We averaged the absolute errors across the different probability values. The mean $|ERROR|$ for the *icons*, *pie*, *circle*, *triangle*, and *bar* were .27($\sigma = .66$), .34 ($\sigma = .7$), .6 ($\sigma = .76$), .51 ($\sigma = .72$), and .42 ($\sigma = .74$) respectively. A Kruskal-Wallis H nonparametric test showed that there was a statistically significant difference in the absolute errors made across the visualization groups, $\chi^2(4, N = 8475) = 1,412.61, p < .001$. We observed a mean rank $|ERROR|$ score of 2,975.94, 3,484.03, 5,487.39, 5178.48, and 4218.18 for *icons*, *pie*, *circle*, *triangle*, and *bar* respectively. The pairwise comparisons were all significant, with adjusted p-values less than or equal to .002; this suggests a strict ordering ***icons*>*pie*>*bar*>*triangle*>*circle***.

4 Conclusion

In this paper, we report the findings of a crowd-sourced experiment that investigated the relationship between visualization design and proportion estimates. Our results demonstrate that people tend to overestimate small probabilities and underestimate large probability values. We used these estimation errors to rank the visualizations based on our subjects' ability to decode quantitative information: *icons* > *pie* > *bar* > *triangle* > *circle*. It is important to note that our research employed visualization types that are widely used in various fields for judgement tasks. We believe that these results will have practical impacts on the design and evaluation of data visualizations to assist decision-making with proportion estimates.

Acknowledgments

This project was supported by the National Science Foundation under Grant No. 1755734.

References

[1] A. Bruhin, H. Fehr-Duda, and T. Epper. Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78(4):1375–1412, 2010.

[2] W. S. Cleveland and R. McGill. Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society: Series A (General)*, 150(3):192–210, 1987.

[3] A. Ottley, E. M. Peck, L. T. Harrison, D. Afergan, C. Ziemkiewicz, H. A. Taylor, P. K. Han, and R. Chang. Improving bayesian reasoning: The effects of phrasing, visualization, and spatial ability. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):529–538, 2015.

[4] D. Simkin and R. Hastie. An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82(398):454–465, 1987.