Linking Performance on Graphical Perception tasks to Visualization Literacy

Hamid Mansoor* Kartik Vasu† Lane Harrison‡

Worcester Polytechnic Institute

Abstract

When it comes to graphical perception tasks, are we all equally skilled? Research in visualization literacy has established methods for objectively measuring a person's literacy level. However, the literacy thread of research has not yet assessed how literacy measures are related to well-established graphical perception tasks. To bridge this gap and contribute to our understanding of how visualization literacy may impact low-level performance with visualizations, we contribute a study replicating a graphical perception study in conjunction with a visualization literacy assessment. Namely, participants with high literacy scores tended to perform not necessarily *accurately* on low-level graphical perception tasks, but *consistently*. These preliminary results suggest a new consideration, consistency, in graphical perception studies, as well as a new dimension relating to the value of visualization literacy.

Index Terms: Literacy, Visualization, Demographics, Perception

1 Introduction

Data visualizations are becoming more commonplace. From consumer goods websites to newscasts, mobile websites, and more, data visualizations provide useful ways for users to factor data into their daily lives.

The mass proliferation of data visualizations, however, means that everyday people encounter visualizations at a higher rate than ever before. This also means that the audience for data visualizations is more diverse than ever. People's abilities to accomplish certain tasks change with factors such as age, various levels of education, and experience with the task itself. This raises several questions for data visualizations. Are we all equally skilled when it comes to basic visualization tasks? Are people's perceptions of skill with data visualizations in line with their actual abilities? These questions suggest a need to examine how individual differences manifest in controlled measures of visualization performance.

The goal of this work is to fill a gap in our understanding of how visualization literacy and other individual differences relate to graphical perception performance. To do so, we replicate the graphical perception study from Cleveland and McGill [2], adding a recent measure of visualization literacy proposed by Lee *et al.* [4], along with other common demographics information. The results of a crowdsourced experiment with $n = 32$ participants indicate that there is little to no effect of the individual differences measured on graphical perception *accuracy*, however, there was a correlation between visualization literacy scores and participants' *consistency* in answers. This result indicates that the value of high visualization literacy may also include a person's consistency of performance with visualizations, a factor that has rarely been discussed compared to more common measures of average accuracy and response times.

*e-mail: hmansoor@wpi.edu

†e-mail:kthooppalvasu@wpi.edu

‡e-mail:ltharrison@wpi.edu

2 Background

Measuring literacy in visualizations can be a challenging task given the range of options available for the design of visualizations. How do we represent a person's ability to read charts i.e. score, percentile or grade etc.? This measure is required to begin the study of individual differences and their relationship with visualization interpreting abilities. Boy *et al.* [1] came up with a principled way of constructing such assessments and implemented 2 online tests as examples. They used Item Response Theory (IRT) to guide this methodology. Another work that attempted to make a valid test was by Lee *et al.* [4] who devised the Visualization Literacy Assessment Test (VLAT). Lee *et al.* [4] used an iterative 6 step method (established in Psychological and Educational Management) to devise this test. Their process involved vetting by experts and they tried to incorporate main tasks associated with reading visualizations into their questions. This test consists of 12 charts and 53 associated multiple choice or True/False.

Researchers are also studying how novices read visualizations, in part due to the massive increase in audience sizes for visualizations. Lee *et al.* [3] propose the "NOVIS" model for the stages a novice encounters when navigating unfamiliar visualizations. These 5 stages are: "encountering visualization, constructing a frame, exploring visualization, questioning the frame, and floundering on visualization". These models provide a solid basis for understanding some of the performance issues that may arise from low visualization literacy.

Certain visual cues are better than others for encoding data in visualizations. People are generally better at assessing position along a common scale (*e.g.* bar charts), than they are at assessing angles or areas (*e.g.* pie charts). This result was established by work from Cleveland and Mcgill, who examined and ranked the use of several visual cues in several experiments [2]. Cleveland and McGill used commonly encountered charts such as pie charts and bar charts and asked participants to estimate what percentage one value in a chart represented of another. They ranked these different cues in terms of the overall log mean error of the responses they received. Given the robustness of Cleveland and McGill's results, which have been replicated by other researchers in the decades since, we take their experiment to be a baseline for studying the impact of visualization literacy and other common demographics factors.

3 Approach

We begin by replicating the comparison experiment from Cleveland and McGill [2], with modifications to include demographics and visualization literacy assessment. All reported studies are conducted on Amazon's Mechanical Turk, with participants receiving at or above the US Minimum Wage, as determined via median experiment time in pilot studies.

Through pilot studies, we found that participant performance degraded over time, possibly due to fatigue. We therefore made changes to the literacy assessment tests and the graphical perception tasks to ensure that the study was appropriately long for crowdsourcing. We found that the original Visualization Literacy Assessment Test (VLAT) [4] test was burdensome on participants due to length (the original questionnaire contains 53 questions). To address this, we reduced the number of visualizations tested from 12 to 5, and

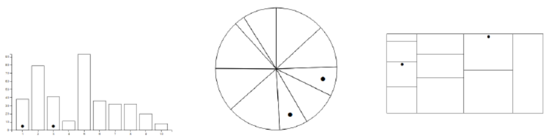

Figure 1: The charts used in our study. Bar charts: easy and familiar. Pie charts: difficult and familiar. Treemaps: Difficult and unfamiliar.

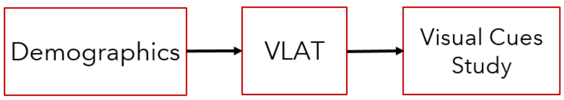

Figure 2: To assess the possible relationships between individual differences and graphical perception tasks, particpants completed a study with the illustrated sequence.

the number of associated questions from 53 to 22. Specifically, we used bar chart, bubble chart, pie chart, line chart, and treemap from the original VLAT. We revisit the possible effects this may have had on the distribution of scores in the results. Further, we abbreviate the number of charts in the graphical perception task to include bar charts, pie charts, and treemaps, to align with the VLAT. We gathered the following demographics: age, educational attainment, perceived experience with statistics (on a scale of 1-7) and perceived experience with data visualizations (on a scale of 1-7).

Participants were given 15 trials each of bar charts, pie charts and treemaps (see Figure 1). We chose these charts not only due to an overlap with the charts tested in the VLAT, but also because they represent differences in ease of perception and familiarity amongst the general population. Namely, bar charts are easy to read and familiar; pie charts are difficult to read but familiar; and treemaps are difficult to read and people are generally unfamiliar with them.

The sequence of the experiment is illustrated in Figure 2.

4 Results

We recruited 32 participants on Amazon's Mechanical Turk, with all tasks being completed within 24 hours of launching the study. We remove data from 3 participants, due to their completion times being well below the average, *e.g.* 5.6 minutes versus the average of 24.58, and large variances in response quality indicating poor attention on the task. Results from the abbreviated VLAT indicate a spread in values, as shown in the y-axis of Figures 3 and 4. These results indicate no relationship between demographics factors or self perceived skills and VLAT or graphical perception accuracy.

As expected, the ranking of overall performance across the 3 charts was bar charts performing the best, followed by pie charts and treemaps. We found no relationship between the VLAT scores and log mean errors across the 3 chart types. Figure 3 shows the scatter plots of the results, with Table 1 showing r and R-squared values for the 3 chart types.

In examining these findings, it was discovered that participants with better VLAT scores tended to have less variance *across* chart types, raising a question of whether their superior performance was not only one of accuracy, but also of consistency. To examine this hypothesis, we transformed the responses from the graphical perception task into measures of variance, that is, when the participants made errors in judging charts, how much variance was in their responses? There was no correlation between literacy and variance in bar charts, a slight relationship with pie charts, but a (relatively) large relationship in treemaps. Figure 4 shows the scatter plots of the results, with Table 2 showing r and R-squared values for the 3 chart types.

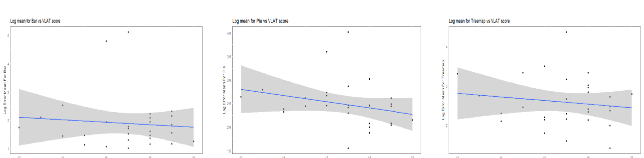

Figure 3: Scatterplots for in log mean error for bar charts, pie charts and treemaps vs VLAT scores.

Chart Type	r	R-squared
Bar	0.037	0.001
Pie	-0.088	0.008
Treemap	-0.032	0.001

Table 1: Values for r and R-squared across the 3 chart types for log mean error.

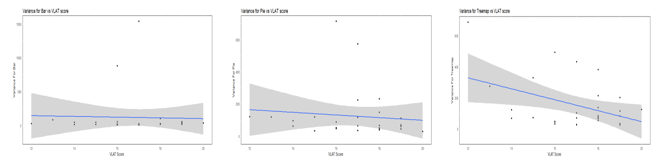

Figure 4: Scatterplots for variance in log mean error for bar charts, pie charts and treemaps vs VLAT scores.

Chart Type	r	R-squared
Bar	-0.03	0.001
Pie	-0.10	0.011
Treemap	-0.42	0.178

Table 2: Values for r and R-squared across the 3 chart types for variance.

5 Conclusion and Future Work

In this work we examine a possible relationship between peoples' performance on graphical perception tasks, namely comparison, with individual differences such as demographics factors and visualization literacy. These preliminary results suggest that participants of low visualization literacy may make comparisons between values in less common chart types more inconsistently than those with higher literacy in visualizations, despite the fact that measures of accuracy are generally more similar. Given differences in visualization literacy, future studies may need to consider explicitly treating *consistency* as a measure of visualization performance, in addition to the often reported *accuracy*. Further, establishing the impact of visualization literacy on low level performance with visualizations may lead to more nuanced models of user abilities, which can inform how we design and evaluate data visualizations for the masses.

References

[1] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics*, 20(12):1963–1972, 2014.

[2] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[3] S. Lee, S.-H. Kim, Y.-H. Hung, H. Lam, Y.-a. Kang, and J. S. Yi. How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE transactions on visualization and computer graphics*, 22(1):499–508, 2016.

[4] S. Lee, S.-H. Kim, and B. C. Kwon. Vlat: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):551–560, 2017.