
Effect of Scale, Symbols and Number of Elements in Data Visualization

First Author

Pritam Pebam
IDC, IIT Bombay
Mumbai
pritampebam@iitb.ac.in

Second Author

Girish Dalvi
IDC, IIT Bombay
Mumbai
girish.dalvi@iitb.ac.in

Extended Abstract

In order to identify the optimal visual representation for an specific kind of data and purpose, we need to measure the effectiveness of data visualizations. Several studies have been conducted to understand fundamental questions in this field — How do we define the effectiveness of a data visualization and how to quantitatively measure the effectiveness of data visualizations. McGill (1984) identified the ordering of the visual tasks on the basis of accuracy in terms of elementary perceptual tasks; his paper validates essential criteria over which certain visualization are more efficient than others. Zhu (2007) defines what is an effective data visualization and how to analyse quantitative data to provide a precise measure of effectiveness based on the accuracy, utility or efficiency. However, in all these existing studies, we did not find the effect of a base-scale, symbols or optimal number of elements in a specific data visualisation. In order to fill in this gap, we performed a preliminary experiment to measure the effect of these entities on data visualizations.

In this paper, we present the results of an experiment where we studied i) the effect of having a base measure (scale) in data-visualizations; ii) the effect of using symbols in data-visualizations; iii) the optimal number of elements a visualization should carry. Twenty five participants of both genders between the age range of 20 to 30 were subjected to a total of ten stimuli (2 to study the effect of base-scale, 2 to study the effect of symbols and 6 to find the optimal number of elements to be used in an data visualization). These stimuli were presented in a random order and participants were asked to answer questions based on the visualization, while at the same time, the time taken for each answer was electronically recorded through the system. Participants were asked to answer the questions as fast as they could.

The collected data was analyzed using a method similar to the one suggested by Zhu. On the whole a students T-test was used to measure significant differences between speed and accuracy (error rate) of answers for each of the data visualizations. Our results show that adding a scale has a significant impact on the effectiveness of the visualization. A visualization with a scale performed significantly better ($p < 0.1$) than one without a scale in terms of its speed (12.67% faster), the error rate was also found to be lower ($p < 0.5$) (33.56% less errors). We also found that the usage of symbols also makes an improvement in understanding the visualization and hence making it more effective, the time taken is again less ($p < 0.1$) i.e (8.9% faster) and less errors (19.72% less errors) were recorded ($p < 0.5$) when compared to a visualization without the use of symbols. However, our results to identify the optimal number of elements in a visualization did not yield any significant results. We hope that the findings of this study would help designers create more effective data visualizations.