Westfield
February 1st
Word count: 583

Statistical graphics

Data sets are prevalent, from TV news to academic papers. A data set often contains at least hundreds of data points, and it is important to quickly extract meanings from the data. Group statistics are often the critical information and need to be visualized.

A common message to be extrapolated from a data set is whether the groups examined are different. The average value of each group is an intuitive and widely used group statistic. Therefore, data graphs often represent the average values. By directly comparing the average values, the viewers can quickly visualize how similar or different the groups are. In academia, bar graphs are normally used to represent and compare the group averages. There are advantages in using bar graphs to represent data. For example, higher values are represented by taller bars, consistent with the intuitive representation of intensity. There are also disadvantages in using bar graphs. For example, height differences in spatially separate bars are less salient (Nothelfer & Franconeri, 2017). The reason is that the spatial separation of the bars discourages the viewers to group the bars and directly compare them. The choice of the graph type ultimately depends on the most important message. If the message is the differences between the groups, then the bars need to be at least physically close, so that they can be visually grouped and easily compared. If the message is the specific value of each group, the bars need to be separated. That way, the average value of each group can be quickly perceived against the y-axis.

The representation of average values does miss a few important statistics of the data set. While variance of each group can be represented by adding error bars, it is difficult to simultaneously present other missing statistics such as mode, distribution, and range. An easy fix would be to represent the individual data points on the graph. This is especially effective as people can quickly extract the average value of an object set through ensemble encoding (Chong & Treisman, 2005). However, this method is rather uncommon in academia because: 1) Larger data sets cannot be clearly represented in a size-limited figure. 2) While people can quickly estimate the average, it is hard for them to extract the group variance and infer whether two groups are statistically different (Pak, Hutchinson, & Turk-Browne, 2014). Therefore, representing individual data points should be considered in cases when statistical inferences are unimportant.

In addition to representing the group statistics, the groups also need to be easily distinguishable. A simple group-wise comparison is clear using bar graphs, as each group is represented by a separate bar. However, studies examining multiple interacting factors (e.g., gender, before/after treatment, health) need to apply multiple feature dimensions to the figures. Color is an adequate tool to represent different categories (Wong, 2010) as it has three dimensions: hue, brightness, and saturation (Hurvich & Jameson, 1955). However, when there are more than three factors, features such as texture, shape must also be considered in visualizing the categories.

This essay examined how group statistics can be effectively visualized. The specific method of visualization depends on the most important message to be conveyed. When factors such as statistical significance and individual performance are both important, their compatibility should be considered. If they cannot be satisfactorily represented together, separate figures can be the solution. Please also note that this essay has not examined statistics such as linear relationships. The methods discussed in the current essay may not apply for other types of visualizations.

References

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision research, 45,* 891-900.

Hurvich, L. M., & Jameson, D. (1955). Some quantitative aspects of an opponent-colors theory. II. Brightness, saturation, and hue in normal and dichromatic vision. *JOSA, 45,* 602-616.

Nothelfer, Christie & Franconeri, Steven. (2017). Visual search through displays of data. *Journal of Vision*. 17. 75. 10.1167/17.10.75.

Pak, S. S., Hutchinson, J. B., & Turk-Browne, N. B. (2014). Intuitive statistics from graphical representations of data. *Journal of Vision, 14*, 1361-1361.

Wong, B. (2010). Points of view: Color coding. *nature methods, 7,* 573-573.