PM R 8 (2016) 268-272
www.pmrjournal.org

Statistically Speaking

The Box Plots Alternative for Visualizing Quantitative Data

Regina L. Nuzzo, PhD

Introduction

Although bar charts are popular among researchers and are ubiquitous in quantitative software packages, they do not always provide the best visualization for a dataset. This column discusses a simple, alternative graphical method that is often underappreciated: the box plot, also known as the box-and-whisker plot. The basic elements of the box plot are presented, along with how to correctly interpret box plots, variations that are available to provide more information, and free online software that researchers can use to create box plots for publication.

Weaknesses of Using Bar Charts

A bar chart is one of the simplest of all data visualizations and is included in every quantitative software package. The bar chart is a good method for summarizing counts or proportions with categorical data, yet it is not always the best option for summarizing or comparing numeric responses in samples. For example, suppose that a clinical team wanted to summarize the Berg Balance Scale scores of 3 groups of patients before and after a therapy, and again after 1 year (Figure 1). Using a bar chart for this display results in several potential problems or weaknesses:

1. The value of interest is the position of the mean, which is represented in the bar chart only with the top line of the bar. The bar itself is unnecessary and conveys no information, which can be considered a waste of space and ink [1,2].
2. The bar leads the viewer's eye to believe that length is an important dimension. Most bar charts start the vertical axis at 0, yet this choice is arbitrary and often misleading, because the sample data may not necessarily include 0. In fact, the minimum value in the sample may be negative, or it may be orders of magnitude greater than 0, and thus the length of the bar is purely arbitrary.
3. Bar charts display the sample mean for a set of data, yet the mean is not a robust summary of the "average" central tendency of a population. The mean is highly sensitive to extreme values, and so for skewed data or in the presence of outliers, the mean may not lie near the center bulk of the data. When working with small samples, samples with outliers, or populations that are not known to be symmetric, a better measure of average is the median, which is not displayed in bar charts.
4. Error bars are often added to bar charts, but it has been shown that viewers have a difficult time judging and interpreting overlaps in error bars [3]. Plots are not always clearly labeled to convey whether the error bar represents one standard deviation, one standard error of the mean, or one half-width of a 95% confidence interval. Furthermore, confidence intervals require assumptions about the data and can be misleading for small samples or skewed distributions [4].

Box Plots: Styles, Strengths, and Value

Princeton statistician John Tukey designed the box plot as an easy-to-draw data visualization as part of exploratory data analysis [5]. The box plot has persisted into the computer age as an information-rich graphic that conveys key features of a numeric dataset at a glance. Unlike the bar chart, it uses statistical summaries (median and interquartile range) that are robust in the presence of skewness and outliers and require no assumptions about the population. It shows the full range of the sample data, provides information about the tails, and indicates the shape of the data. It can be used for samples as small as $n = 5$ and allows for quick side-by-side comparisons between groups.

The components of a box plot are as follows (Figure 2):

1. The box plot divides the sample data into fourths, or quartiles: 2 box panels and 2 whiskers (plus any outlying values beyond the whiskers).

http://dx.doi.org/10.1016/j.pmrj.2016.02.001

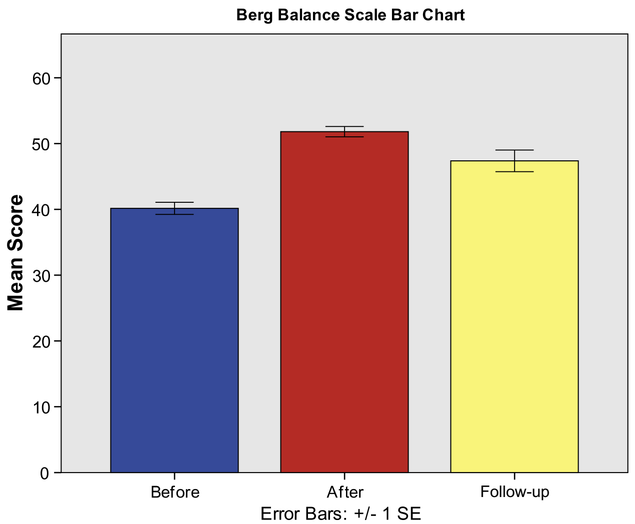

Figure 1. Typical bar chart for the balance scores of a group of patients before therapy, after therapy, and at 1-year follow-up. Mean scores for each group are shown with the top line of the bar only. Error bars show 1 standard error (SE) of the mean. N = 80, 40, and 10 as a result of patients lost to follow-up.

2. The box spans the middle 50% of the data. The outer edges of the box, often called the "hinges," indicate the first quartile (the 25th percentile, or the value at which 25% of the data fall below) and the third quartile (the 75th percentile, or the value at which 25% of the data fall above).

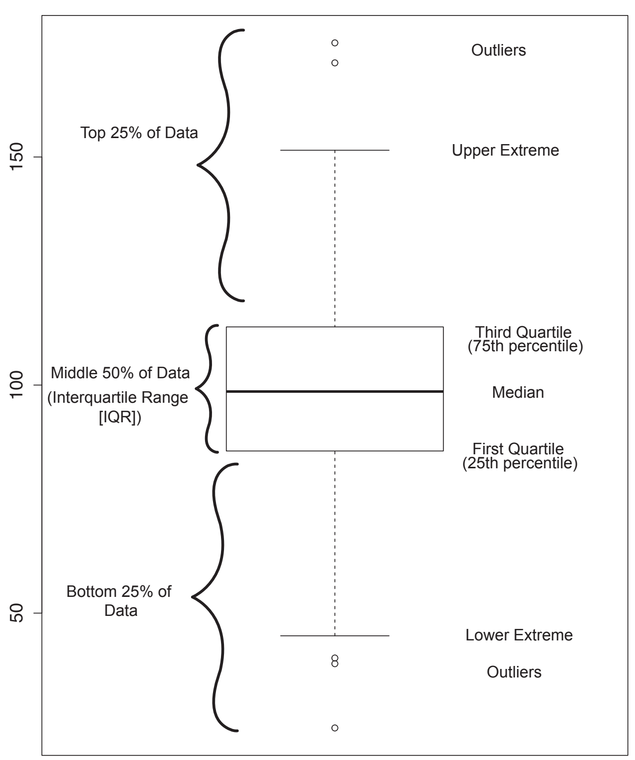

Figure 2. Annotated box plot of 1000 points from a normal distribution with a mean of 100 and a standard deviation of 20.

3. The middle line of the box indicates the median (the second quartile, or the 50th percentile).
4. The length of the box is the interquartile range (IQR), which is a measure of spread similar to the standard deviation.
5. The whiskers show the extent of the data range for the other 50% of the data. The whiskers start at the edges of the center bulk (the first and third quartiles) and extend to what is considered "extreme" values in the data, typically taken to be a distance of 1.5 IQR beyond the first and third quartiles, although other variations are possible (described in a subsequent section).
6. Data values beyond the extremes are considered outliers or potential outliers and are marked as individual points.

Box Plot Demonstration Using Dataset From Figure 1

The data in Figure 1 yield more information when expressed as box plots (Figure 3). Here the viewer can compare the medians among the groups, and other information is also apparent. Patients before therapy had a median balance score of about 39, which rose to about 53 after therapy and dropped to about 48 a year later. The positive skew of the patients' scores before therapy can be seen from the long right (upper) whisker, revealing that whereas most patients initially tended to score fairly low, a few (including 2 high outliers) scored much higher. Immediately after therapy, the patients' scores were much higher, showing almost no overlap with scores before therapy. The data had a negative skew, seen in the longer left (lower) whisker and the one moderately low outlier. The maximum possible score of the Berg Balance Scale is 56, would explain the ceiling effect in this group of data. One year after

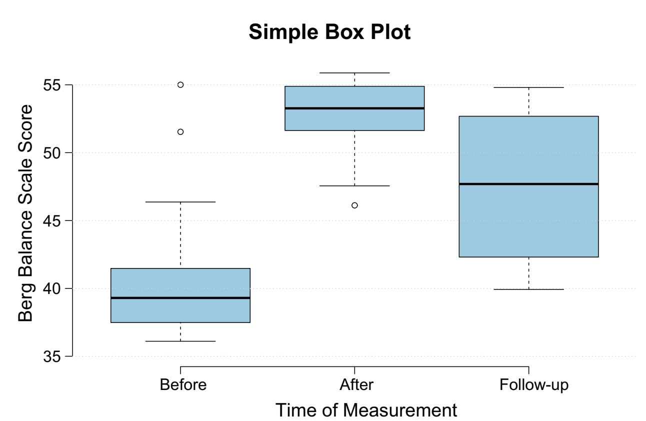

Figure 3. Simple side-by-side box plots for the data in Figure 1. Center lines show the medians, box limits indicate the 25th and 75th percentiles as determined by R software, whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, and outliers are represented by dots. n = 80, 40, and 10 sample points.

therapy, patients' scores dropped somewhat, but the data were fairly symmetric and the range containing the middle 50% of the patients was quite wide, which indicates the scores were sparsely spread out.

To aid in visualization, the original data can be seen overlaid the box plots in Figure 4.

Variations in Box Plot Presentations

The basic box plot can be enhanced to convey even more information. Here are common options:

1. Whiskers: As noted earlier, box plot whiskers show the range of the data set. All reasonably "extreme" data are contained between the 2 ends of the whiskers. Whiskers are typically extended to be 1.5 times the interquartile range beyond the first and third quartiles [5]. For a normal distribution, this results in about 99% of the data ($\pm$2.7 standard deviations) being contained between the 2 whisker ends. Data values beyond this are flagged as possible outliers and plotted as individual points. Variants use different definitions of "extreme" to cover different amounts of data. The Spear variant extends the whiskers to the minimum and maximum data values, so that 100% of the data are contained between the whisker ends [6] (Figure 5). This approach can make skewness more apparent, but it also does not allow for the quick identification of outliers. Other common variants extend the whiskers to specified percentiles; for example, extending the whiskers to the 5th and 95th percentiles of the data will capture 90% of the data within the whisker ends [7] (Figure 6).
2. Sample sizes with unequal distributions: For unequal group sizes, box widths can be scaled proportional to the square root of the sample size, which encourages the viewer's eye to place more weight on a wider box than a narrower one, reflecting the increased precision of larger samples. This variant is recommended for data with unequal sample sizes. Figure 7 shows the greater visual weight placed on the data collected during the first time point, and decreasing thereafter.
3. Samples with a large size or from a normal distribution:
   A. 95% confidence intervals for the population median: Confidence intervals for the median can be presented as "notches" around the median [8]. When the notches do not overlap, the medians can be assumed to be significantly different at the 95% confidence level. Overlapping notches do not rule out a significant difference. Figure 8 shows the 95% confidence interval for the median patient score before therapy to be about 39 to 40, which does not overlap with the confidence interval after therapy, so the medians of the 2 groups can be assumed to be significantly different. Although the

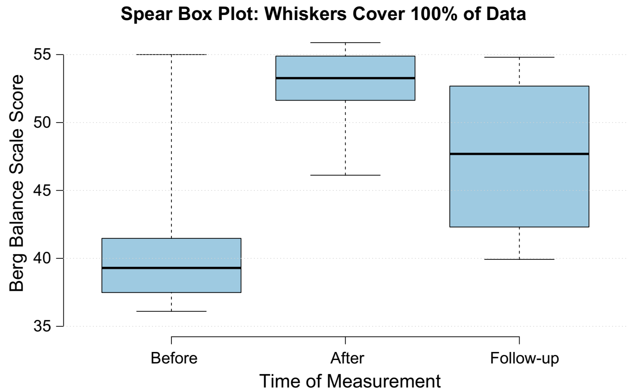

Figure 5. Whiskers are defined to cover 100% of the data (thus no outliers can be identified). Center lines show the medians, box limits indicate the 25th and 75th percentiles as determined by R software, and whiskers extend to minimum and maximum values. n = 80, 40, and 10 sample points.

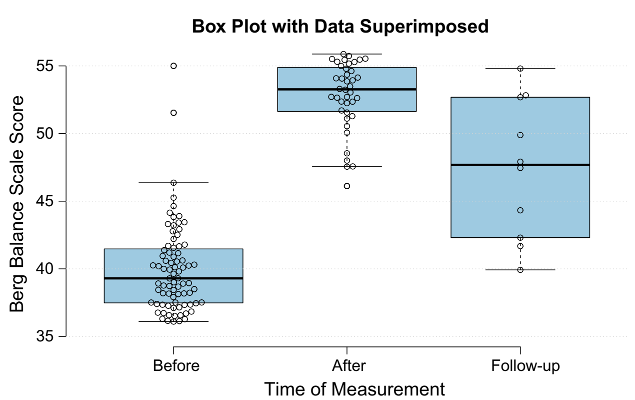

Figure 4. Data values superimposed on the box plots to aid in visualization. Center lines show the medians, box limits indicate the 25th and 75th percentiles as determined by R software, whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots, and data points are plotted as open circles. n = 80, 40, and 10 sample points.

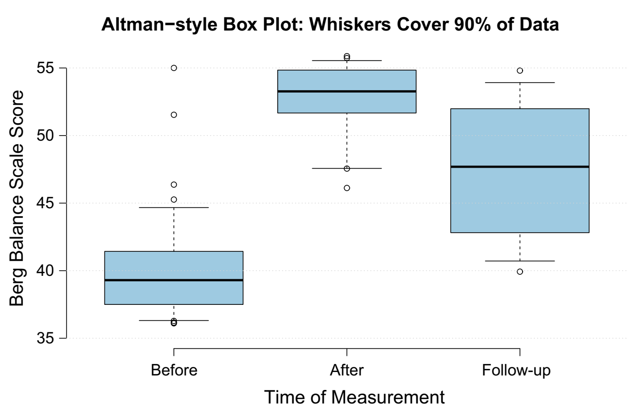

Figure 6. Whiskers are defined to cover 90% of the data. Center lines show the medians, box limits indicate the 25th and 75th percentiles as determined by R software, whiskers extend to 5th and 95th percentiles, and outliers are represented by dots. n = 80, 40, 10 sample points.

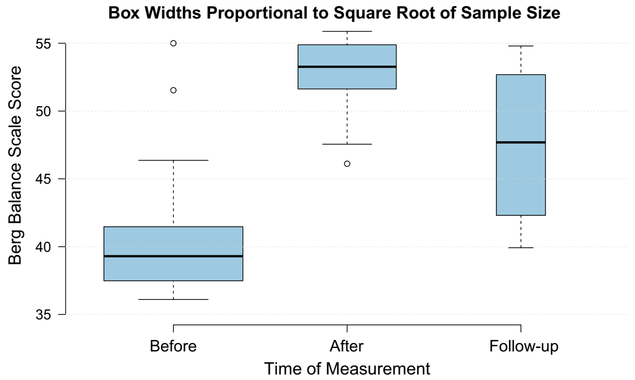

Figure 7. The width of boxes is proportional to the square root of the sample size for each group, and thus wider boxes reflect increased precision of estimates. Center lines show the medians, box limits indicate the 25th and 75th percentiles as determined by R software, whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, and outliers are represented by dots. n = 80, 40, and 10 sample points.

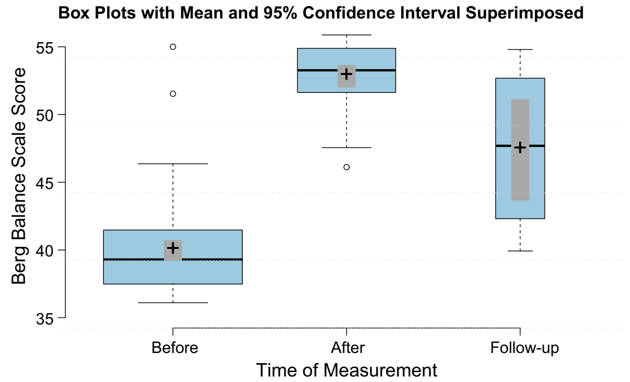

Figure 9. Sample means (indicated by crosses) and 95% confidence intervals of the means (indicated by shaded bars) superimposed on the box plots show the same information traditionally seen on bar charts. Note that confidence intervals rely are not recommended for skewed data or small sample sizes, both of which are present here. Center lines show the medians, box limits indicate the 25th and 75th percentiles as determined by R software, whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots, and the width of the boxes is proportional to the square root of the sample size. n = 80, 40, and 10 sample points.

median does not rely on assumptions about the data, the validity of the confidence interval depends on approximations of a normal distribution, and thus care should be taken when interpreting the intervals for small samples, such as those in the last group.

B. Sample mean: Because the sample mean is sensitive to skewness and outliers, it should only be used as a supplement to the median. It can be added to the box in the box plot, usually marked with a cross (Figure 9).
C. Confidence intervals for the population mean: Caution should be used in interpreting confidence intervals for the mean for small samples or if the population is not known to be unimodal and symmetric. The confidence intervals can be overlaid on the box [4]. Figure 9 shows the sample means marked with a cross and the 95% confidence intervals marked with a shaded bar. The mean's sensitivity to skewness and outliers can also be seen in groups 1 and 2, as the mean is pulled toward the upper and lower tails, respectively.

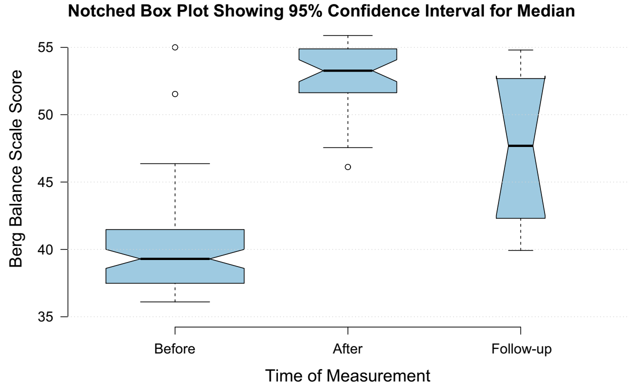

Figure 8. Notches are $\pm$1.58*IQR/sqrt(n) and represent the 95% confidence interval for each median. Non-overlapping notches give roughly 95% confidence that 2 medians differ, such that in 19 of 20 cases, the population medians (estimated based on the samples) are in fact different. Center lines show the medians, box limits indicate the 25th and 75th percentiles as determined by R software, whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots, and the width of the boxes is proportional to the square root of the sample size. n = 80, 40, and 10 sample points.

Pitfalls of Box Plots

Although box plots are flexible and require no assumptions of the underlying population, the following warnings should be noted:

1. Box plots should be used only with sample sizes of at least 5.
2. Box plots are not appropriate for nominal or ordinal data.
3. Box plot whiskers should not be confused with standard error bars or standard deviation bars and cannot be used for inference.
4. Box plots are not well designed to reveal the presence of multimodality. A histogram should always be performed as part of exploratory data analysis, which provides more detail about the shape of the data sample.

Box Plot Software

Box plots are available on most software packages, including SPSS, SAS, Stata, and R. A free online tool, BoxPlotR, is also available at http://boxplot.tyerslab.com/. This program allows users to upload data, create box plots (including the variants discussed here), and download plots in various graphical formats for inclusion in a manuscript [9].

References

1. Tufte ER. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press; 1983.
2. Cleveland WS. The Elements of Graphing Data. Monterey, CA: Wadsworth; 1985.
3. Belia S, Fidler F, Williams J, Cumming G. Researchers misunderstand confidence intervals and standard error bars. Psychol Methods 2005;10:389-396.
4. Sainani KL. A closer look at confidence intervals. PM R 2011;3:1134-1141.
5. Tukey JW. Exploratory Data Analysis. Boston, MA: Addison-Wesley; 1977.
6. Spear ME. Charting Statistics. New York, NY: McGraw-Hill Book Company; 1952.
7. Altman DG. Practical Statistics for Medical Research. London, England: Chapman and Hall; 1991.
8. McGill R, Tukey JW, Larsen WA. Variations of box plots. Am Stat 1978;32:12-16.
9. Spitzer M, Wildenhain J, Rappsilber J, Tyers M. BoxPlotR: A web tool for generation of box plots. Nat Methods 2014;11:121-122.

Disclosure

R.L.N. Department of Science, Technology, and Mathematics, Department of Psychology, HMB S340F, Gallaudet University, 800 Florida Ave NE, Washington, DC 20002-3695. Address correspondence to: R.L.N.; e-mail: regina.nuzzo@gallaudet.edu

Disclosure: nothing to disclose