

## **Use of Box for Visualization**

John W. Turkey introduced the box plot in his book *Exploratory Data Analysis*. The box plot is also known as a box-and-whisker plot. Tukey was a statistician credited with many contributions to the field of statistics, including the development of the box plot. Similarly, "The box plot is a graphical representation of data that was first introduced by John W. Tukey in his book *Exploratory Data Analysis*." (Morgenthaler, 2009). It shows the distribution of data by dividing the data into four equal parts: the upper quartile, the lower quartile, the interquartile range, and the outliers. The upper quartile is the data point 75% of the way from the lowest to the highest value. The lower quartile is the data point that is 25% of the way from the lowest to the highest value.

The interquartile range is the difference between the upper and lower quartiles. The outliers are the data points that are more than 1.5 times the interquartile range above the upper quartile or below the lower quartile. There are many different ways to make a box plot, but the most common way is to use a software program like Excel or SPSS. To make a box plot in Excel, select the data you want to include in the plot. Then, go to the Insert tab and click on the Box Plot button. In SPSS, go to the Graphs menu and select the Boxplot option. Then, select the variables that you want to include in the plot and click OK. Both Excel and SPSS will generate a box plot based on the data that you selected.

The whiskers extend from the box to show the minimum and maximum values. "Box plots are a standardized way of displaying the distribution of data based on the five number summary: minimum, first quartile, median, third quartile, and maximum." (Komorowski, Marshall, Salciccioli & Crutain, 2016). Any data points that are outside of the whiskers are considered outliers. A box plot is a valuable tool for both exploratory data analysis and for comparing multiple data sets. When looking at a box plot, you can get a quick sense of how the data is distributed and whether there are any outliers. You can also use box plots to compare multiple sets of data to see if they are similar or if there are any significant differences. There are a few things to keep in mind when interpreting box plots.

First, the box only shows the middle 50% of the data. So, if the box is tiny, most data is clustered around the median. If the box is huge, the data is more spread out. Second, the whiskers only show the minimum and maximum values, so they do not give any information about the middle of the data. Finally, outliers can sometimes be hard to spot on a box plot, so it is always good to look at the raw data to ensure you're not missing anything. Overall, the box plot is a helpful tool for understanding data sets. It can be used to quickly identify the distribution of the data and any outliers that may be present. Box plots are easy to construct and understand because all you need is a set of data and a way to divide it into four equal parts. They provide a graphical representation of the data that is easy to interpret.

The boxes represent the data set's 25th, 50th, and 75th percentiles, and the whiskers represent the minimum and maximum values. Box plots can be summarized to compare different data sets to see if they have the same distribution. The box plots can be used to identify outliers since they are not affected by outliers. "A box plot is a graphical way of displaying certain summary statistics from a data set. In particular, the median, first and third quartiles, and extreme values can be plotted." (Cox, 2017). From this, we see the uses of box plots and their advantages.