Box Plot

Idea In Short

- In descriptive statistics, a box plot is a method for graphically representing numerical data using quartiles
- A box plot is a graph that gives you a good indication of how the values in the data are spread out
- Box plots take up less space, which is useful when comparing distributions between many groups / datasets

For some distributions/datasets, you will find that you need more information than the measures of central tendency (median, mean, and mode). The mathematician John W. Tukey introduced this type of visual data display in 19691. Since then, several variations on the traditional box plot have been described2. Two of the most common are variable width box plots and notched box plots. According to Wikipedia:

> In descriptive statistics, a box plot or boxplot is a method for graphically depicting groups of numerical data through their quartiles. Box plots may also have lines extending vertically from the boxes [whiskers] indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot and box-and-whisker diagram.

Box and whisker plots are very effective and easy to read. They summarize data from multiple sources and display the results in a single graph. Box and whisker plots allow for comparison of data from different categories for easier, more effective decision-making.

How to interpret Box & Whisker plots?

The box and whisker plots show the spread of your data using five pieces of information. Correspondingly, box plots show the five number summary:

- The minimum (the smallest number in the data set). The left whisker shows the minimum
- First quartile, Q1, is the far left of the box (or the far right of the left whisker)
- The median is shown as a line in the centre of the box
- Third quartile, Q3, shown at the far right of the box (at the far left of the right whisker)
- The maximum (the largest number in the data set), shown at the far right of the box

When to use box & whisker plots?

Use box and whisker plots when you have multiple data sets from independent sources that are related to each other in some way. Examples include test scores between schools or classrooms, data from before and after a process change, data from different machines producing the same products, etc. Box plot takes up less space, which is useful when comparing distributions between many groups or datasets.

Summary

- Box plot is a convenient way of visually displaying the data distribution through their quartiles
- They are a standardized way of displaying the distribution of data based on a five number summary