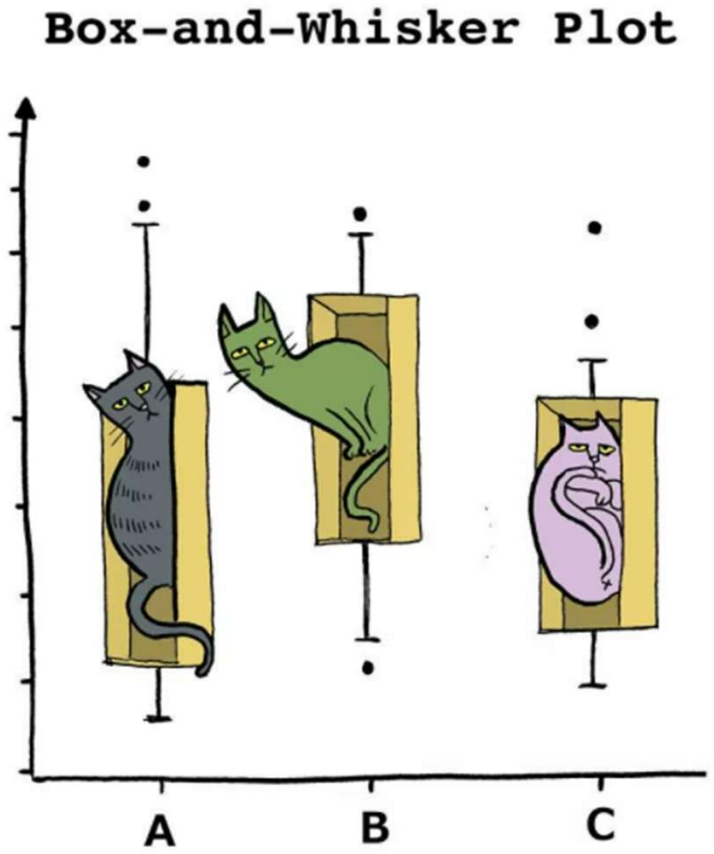
Statistics for Data Analysis

Lecturer: Marina Iantorno

E-mail: miantorno@cct.ie

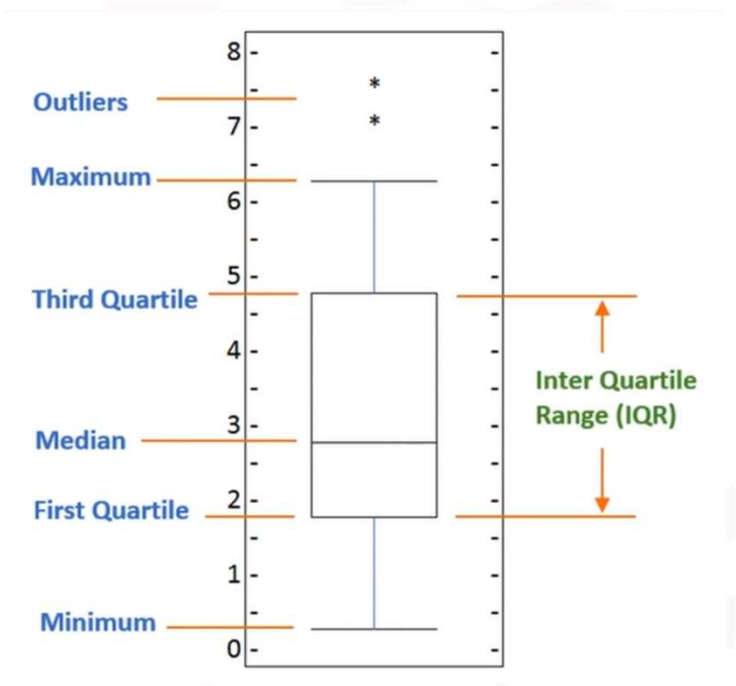
Year 2022



Box Plot

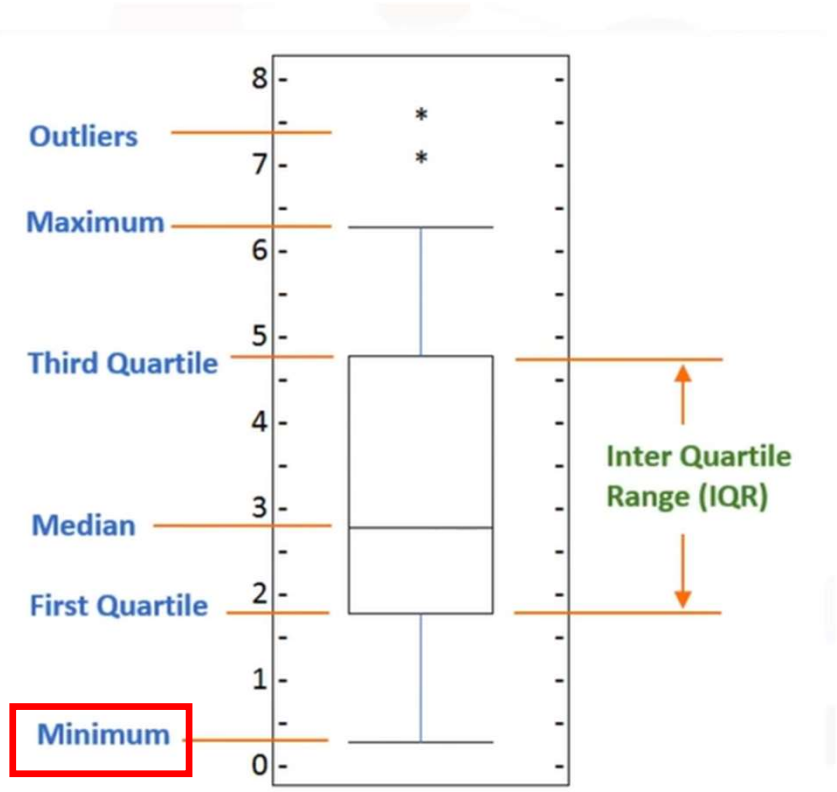
Box Plot

A Box and Whisker plot is commonly known as a Box Plot. It is a way of statistically representing the distribution of the data, and it allows us to do it through 5 dimensions.



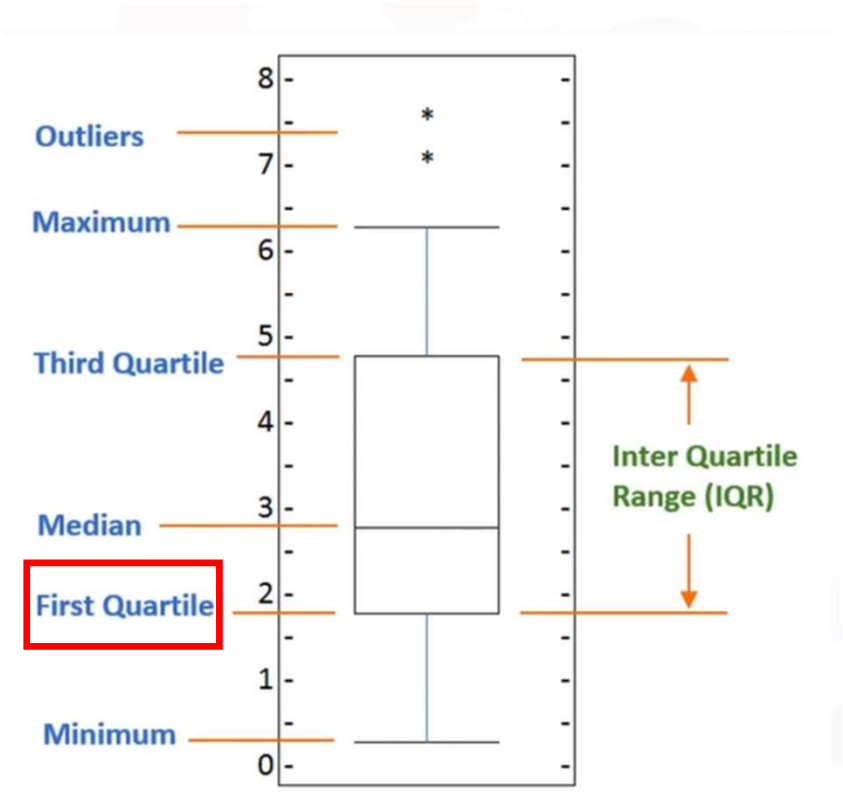
Box Plot

The first dimension is “Minimum”, which is the smallest number in the sorted data. This value can be obtained by subtracting 1.5 times the IQR (interquartile range from the first quartile).



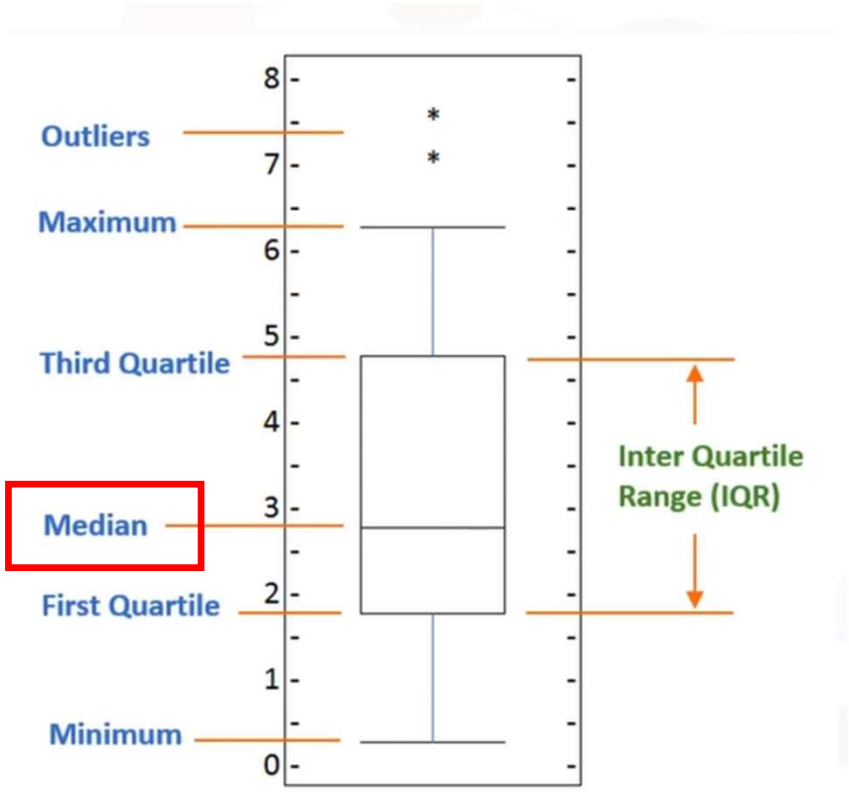
Box Plot

The second dimension is “First Quartile”, which represents the 25% of the way through the sorted data. In simple words, ¼ of the datapoints are below this value.



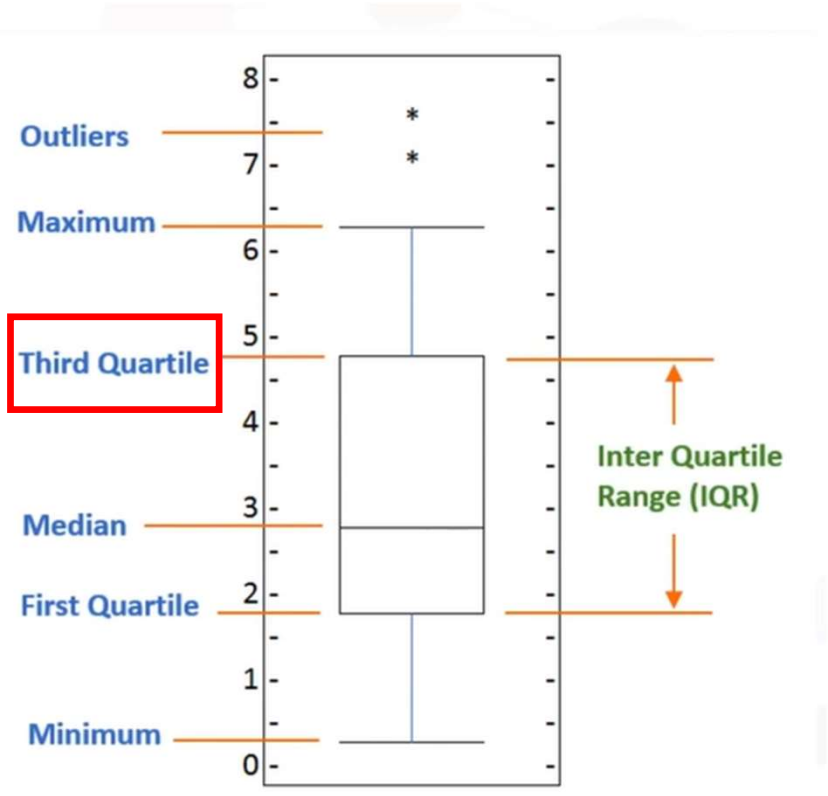
Box Plot

The third dimension is “Median”, which represents exactly the 50% of the datapoints. We can look at it saying that ½ of the data is below this value and ½ of the data is above this value.



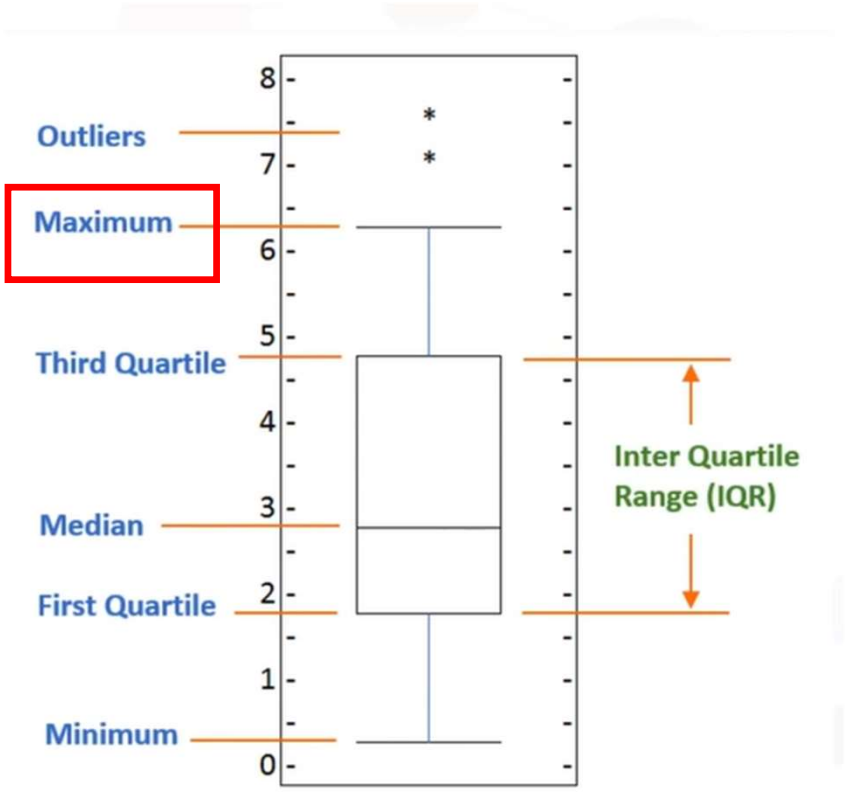
Box Plot

The fourth dimension is "3rd quartile", which is 75% of the way through the sorted data. ¾ of the datapoints are below this value, while ¼ of the greater datapoints are above this value.



Box Plot

The fifth dimension is “Maximum”, which is the highest number in the sorted data. We can say that this value equals the third quartile summed with 1.5 multiplied by IQR.



Box Plot

Finally, this plot displays the outliers as individual dots that occur outside the upper and/or lower extremes.