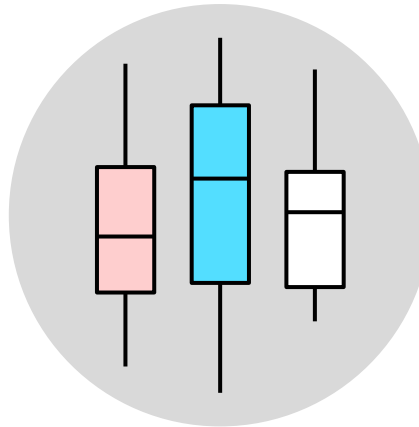


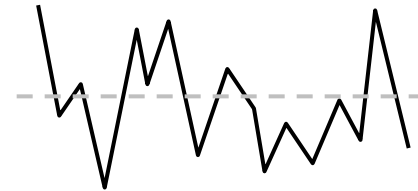
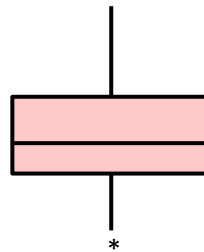
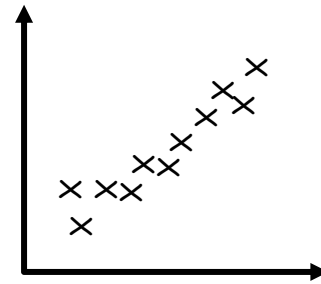
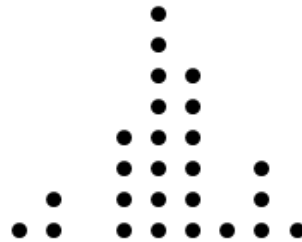
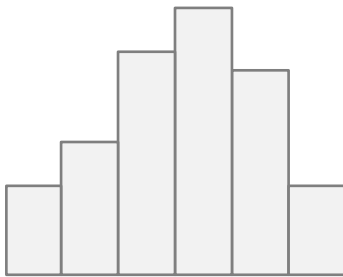
Continuous Improvement Toolkit

Box Plot



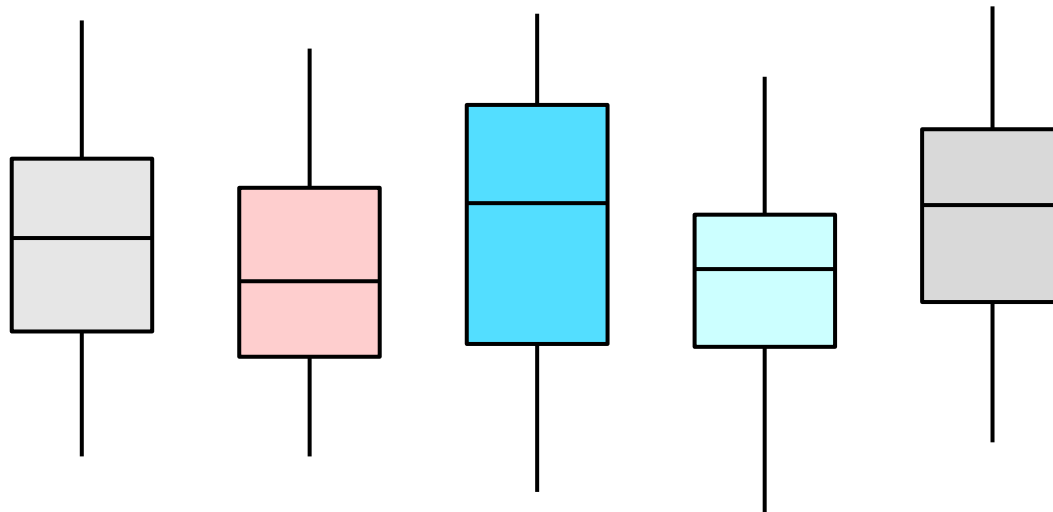
Box Plot

One of the best ways to analyze any process is to **plot the data**



Box Plot

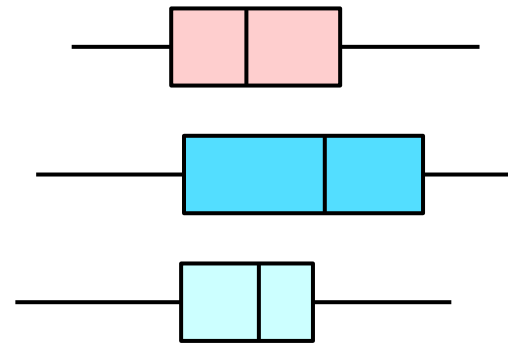
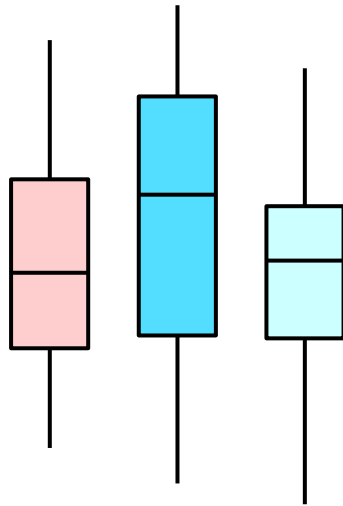
A **box plot** is a graphical way that summarizes the important aspects of the distribution of numeric data



Also referred to as a **Box-and-Whisker Plot** as it displays the data in a box-and-whiskers format

Box Plot

Box plots can be drawn either **vertically** or **horizontally**

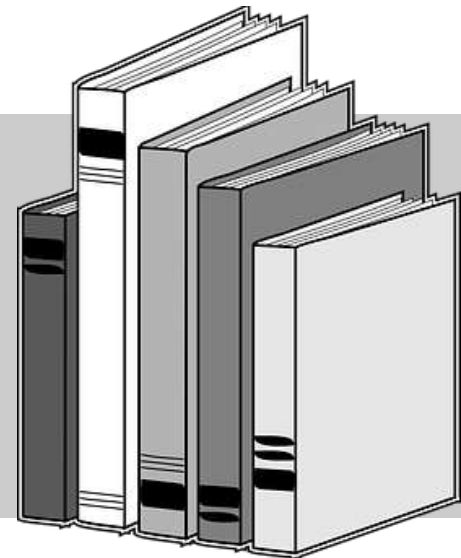


The **length** of the box plot indicates the spread of the data

Box Plot

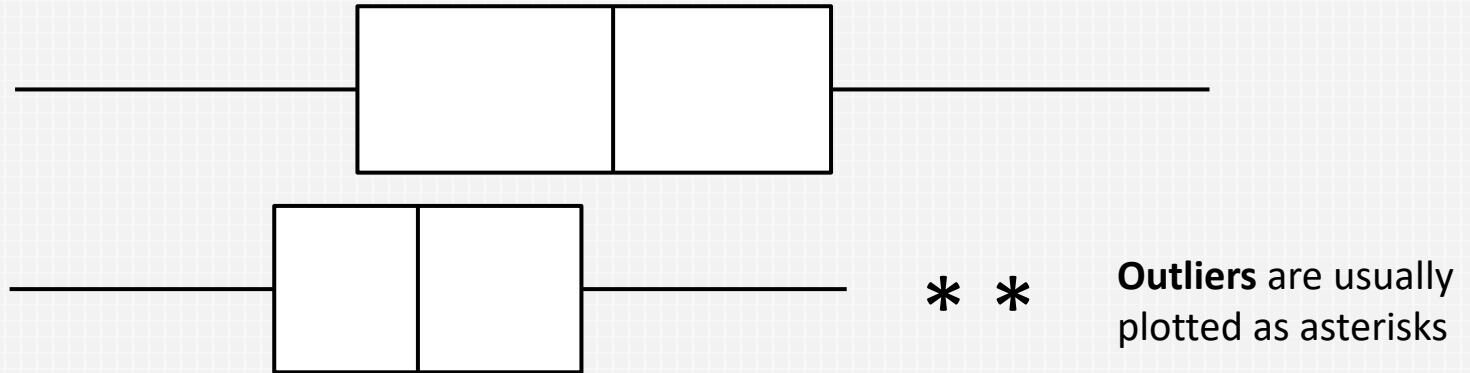
Box plots are widely used in statistics, process improvement, scientific research, economics, and in social and human sciences

Mainly used to **explore** data as well as to **present** the data in an easy and understandable manner.



Box Plot

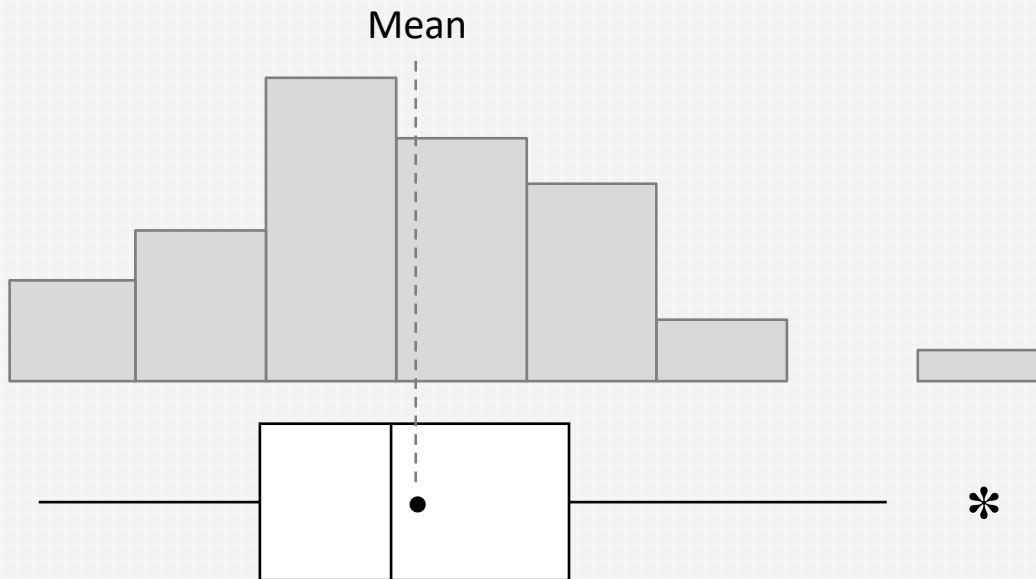
They provide a quick way for examining the **central tendency** and **variation** present in the data



A wider range boxplot indicates **more variability**

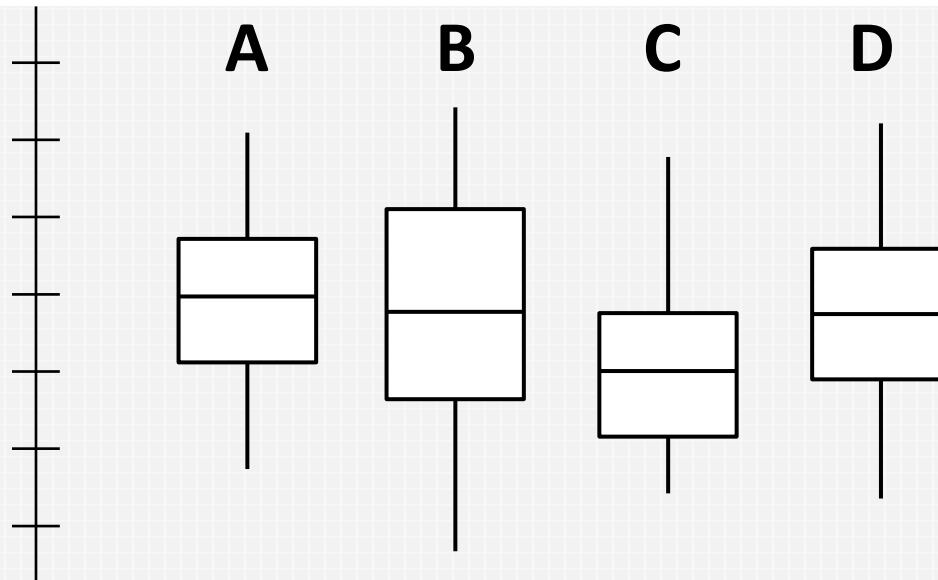
Box Plot

The same continuous data can be presented graphically using histograms and box plots



Box Plot

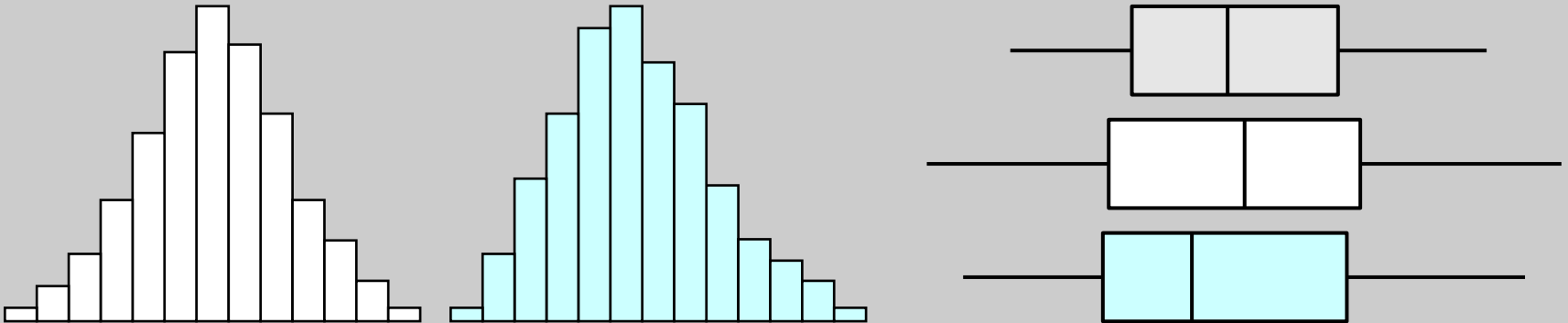
Useful when **comparing** between several data sets



Comparing central tendency and variability

Box Plot

Less detailed than histograms, and take up less space which allows easy comparison of multiple data sets



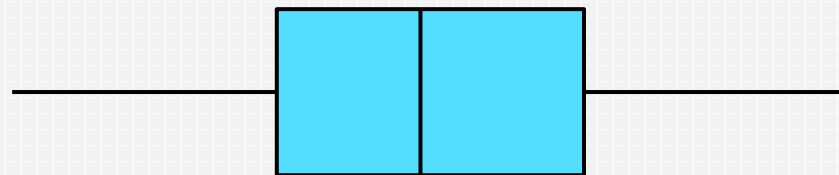
Box Plot

Used to check if there is a significant difference in the process after implementing **process improvement**

BEFORE



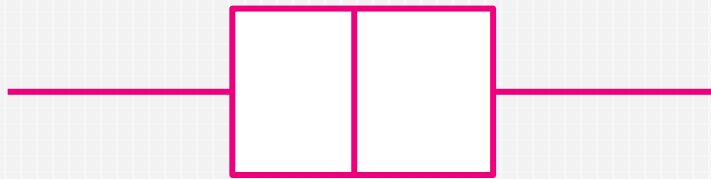
AFTER



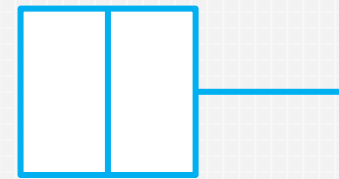
In terms of central tendency and variability

Box Plot

Like histograms, used for moderate to large amount of data



$N = 40$

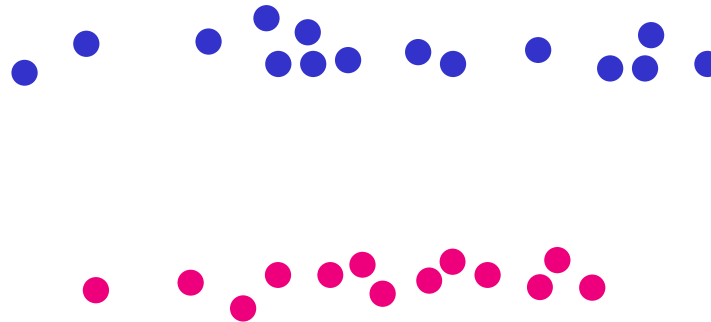


$N = 14$

The **size** of the box plot can vary significantly if the data size is too small

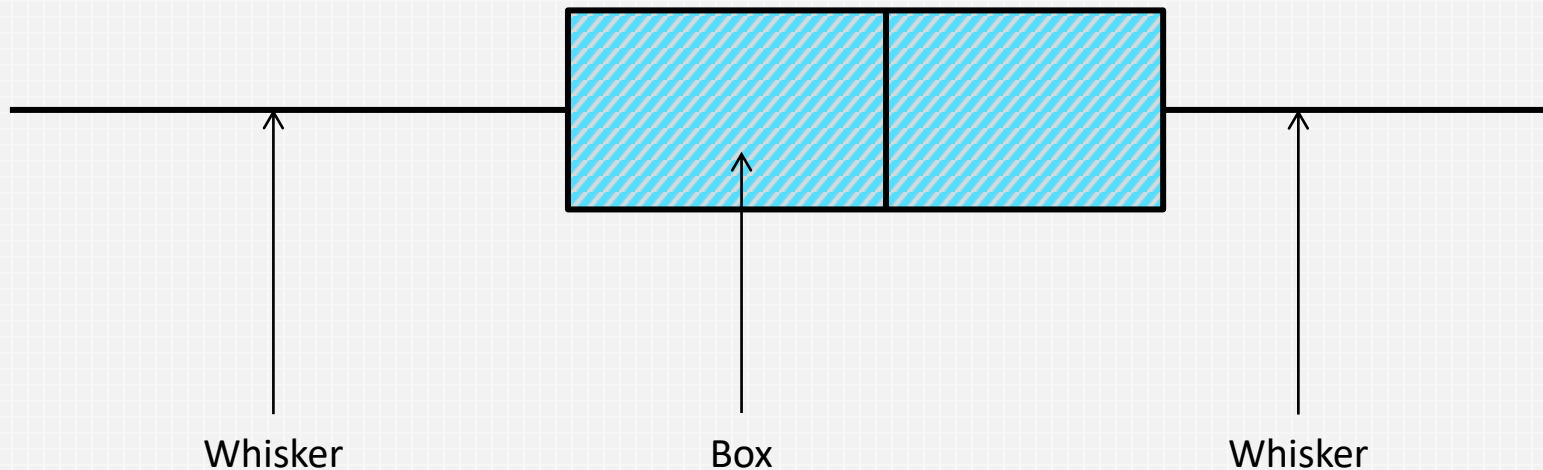
Box Plot

Individual Value Plots are preferred over boxplots when representing small amount of data



Box Plot

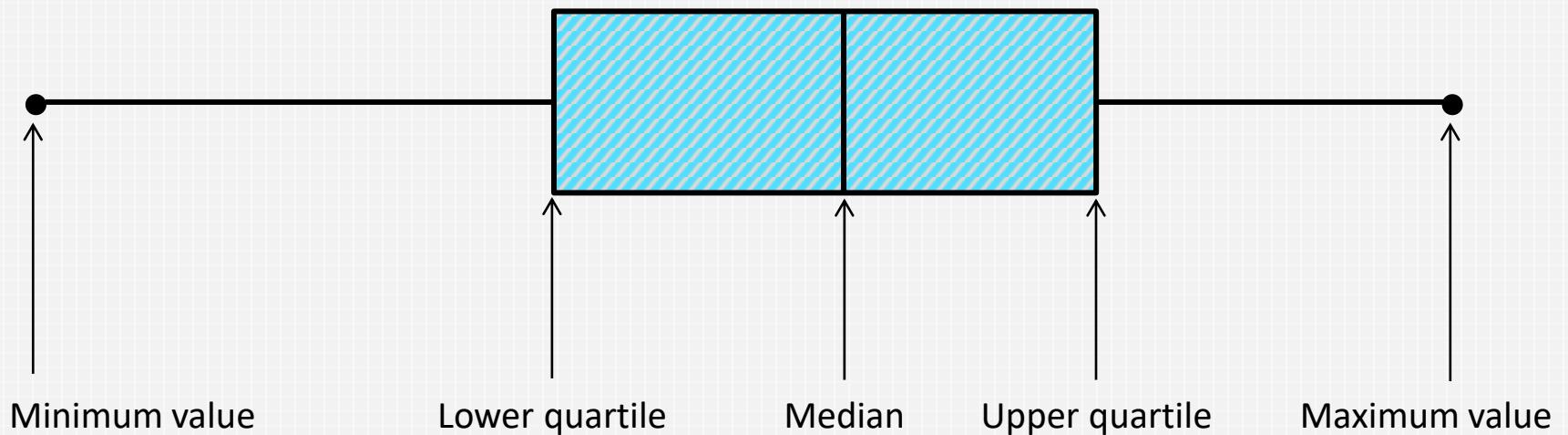
A box plot is made up of a **box** and two **'whiskers'**



The whiskers represent all data values, and the maximum length of a whisker is 1.5

Box Plot

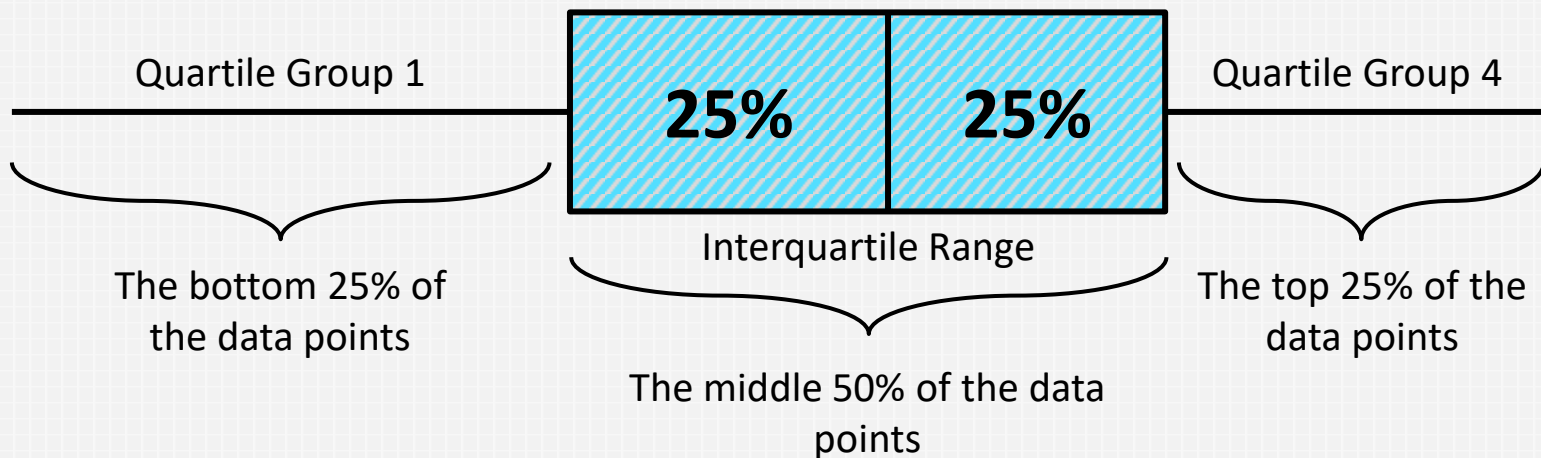
Box plots summarize **key statistics** from the data



Including the median, maximum and minimum values, as well as the lower and upper quartiles (Q1 and Q3)

Box Plot

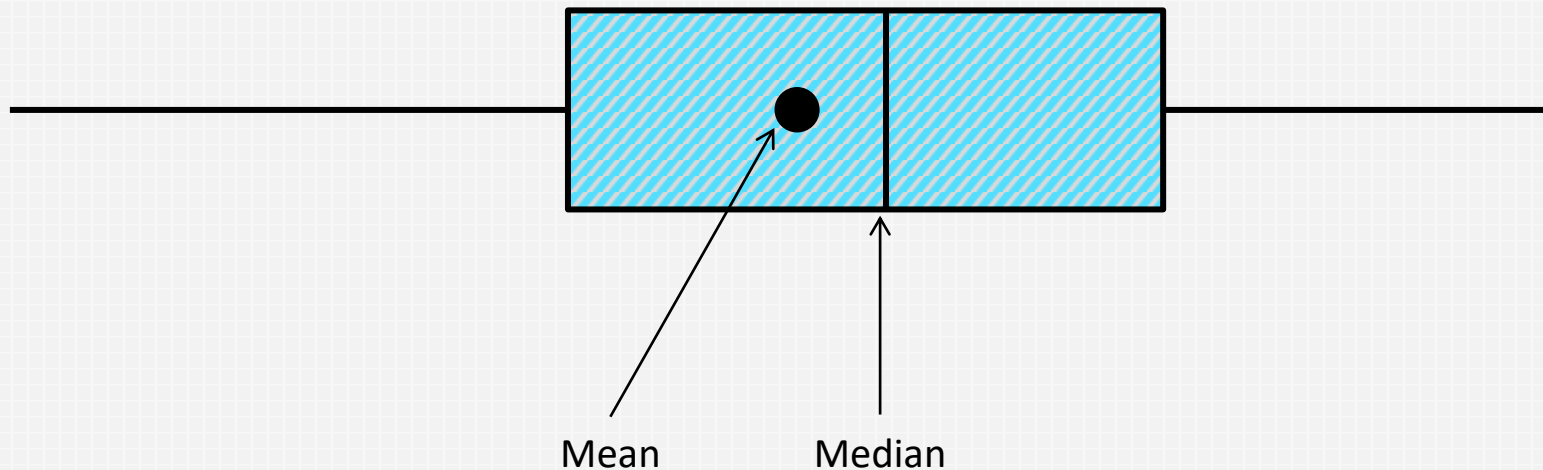
The data is plotted such as . . .



When the median line is not present in the box plot, it suggests that it coincides with one of the quartiles

Box Plot

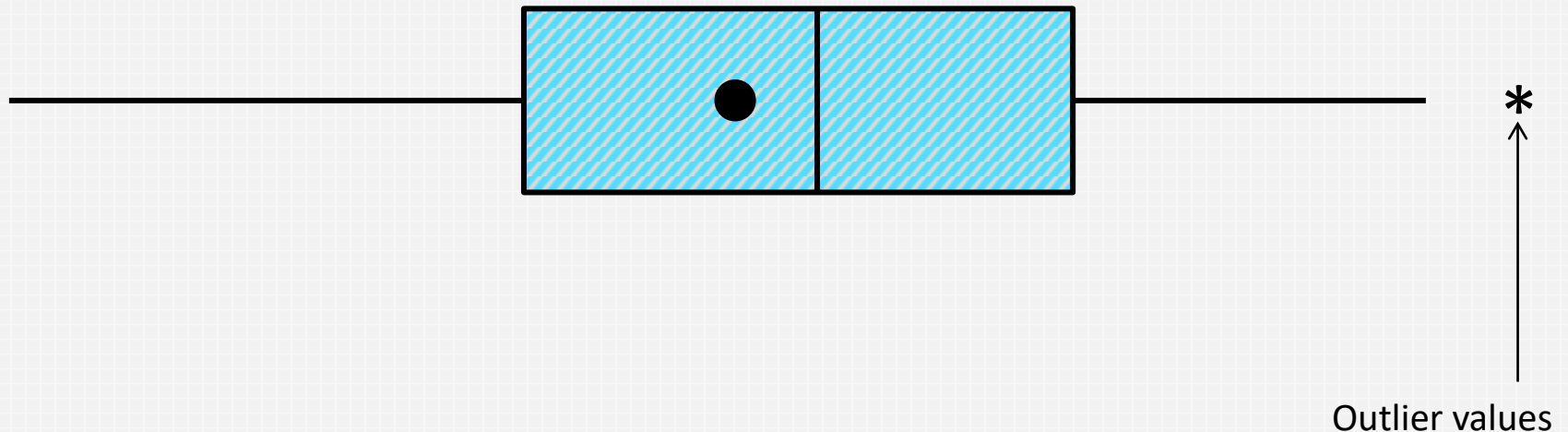
Sometimes they display the **mean** with a special character



Other character can indicate the mean such as a diamond, a plus, etc.

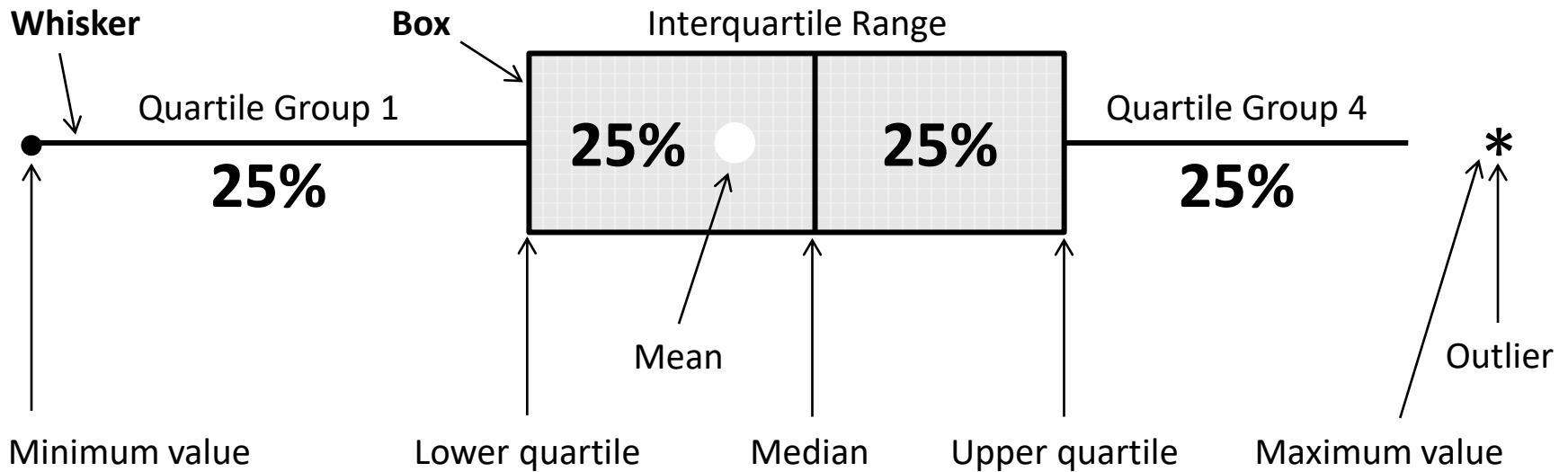
Box Plot

Any data beyond the whiskers are considered **outliers**



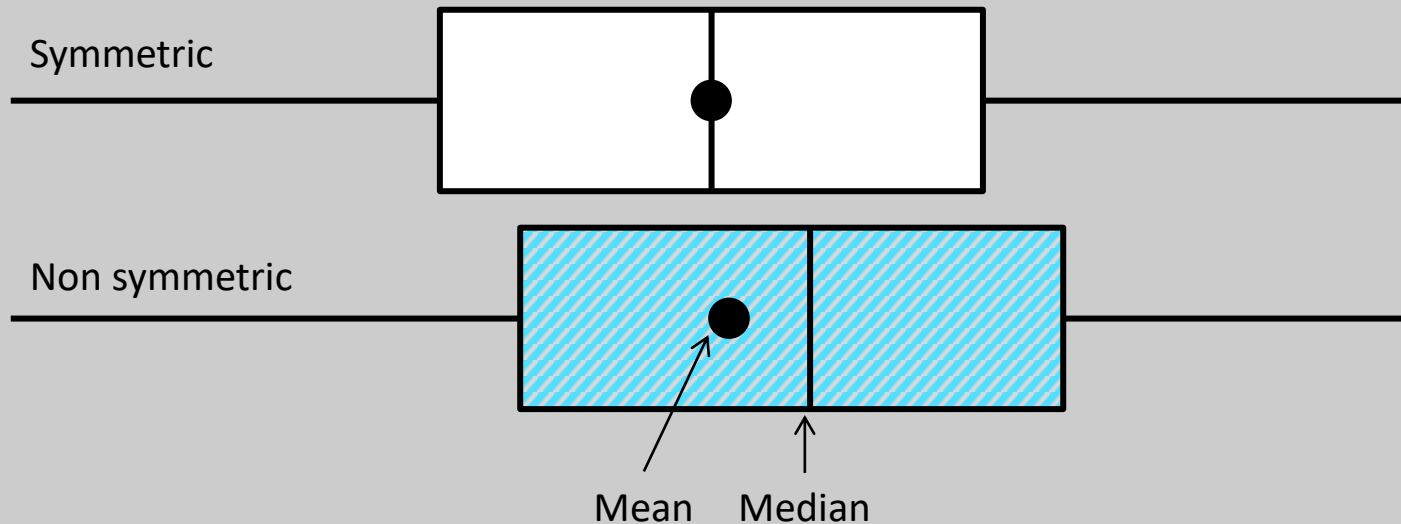
Outliers often reflect errors in data recording or data entry. If the values are real, you should investigate what was going on in the process at that time

Box Plot



Box Plot

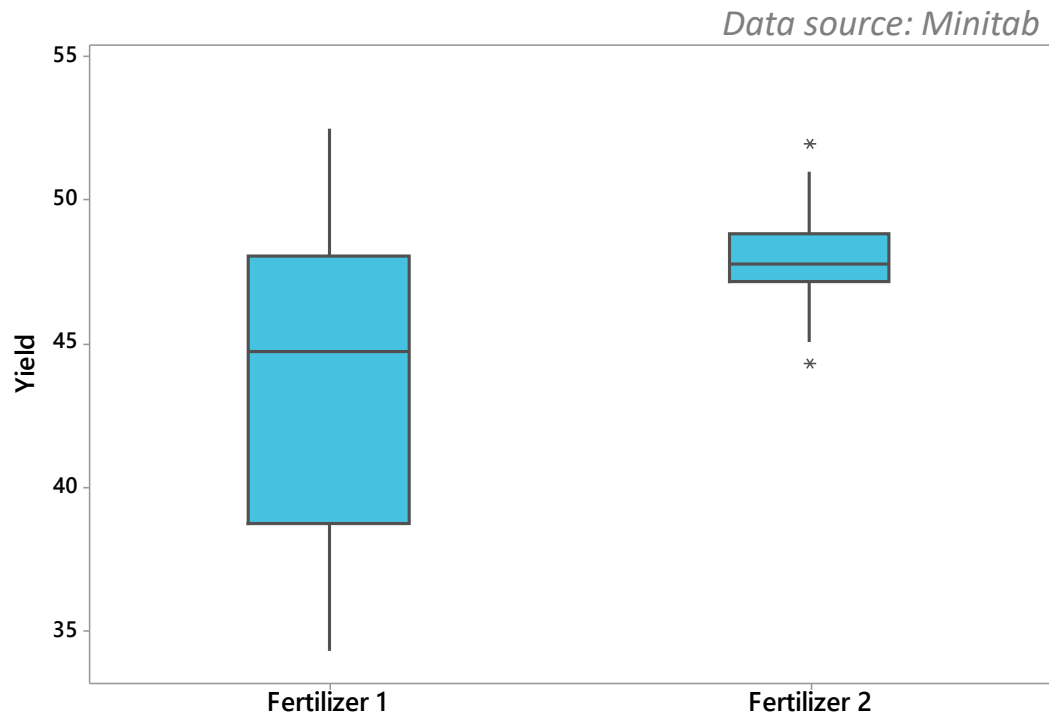
Can tell whether the **distribution** is symmetrical or skewed



In a **symmetric distribution**, the mean and median are nearly the same, and the two whiskers has almost the same length

Box Plot

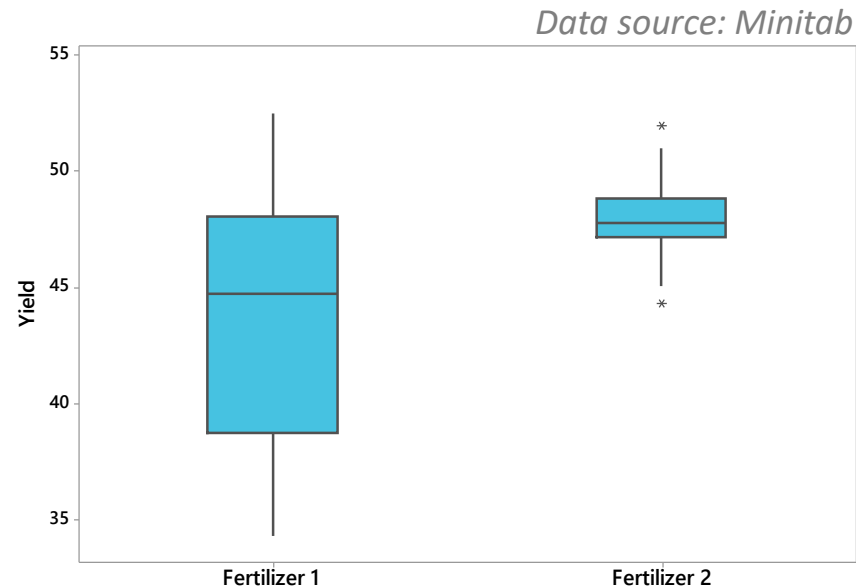
Example – Fertilizers



Fertilizer 2 appears to have a higher yield value than Fertilizer 1

Box Plot

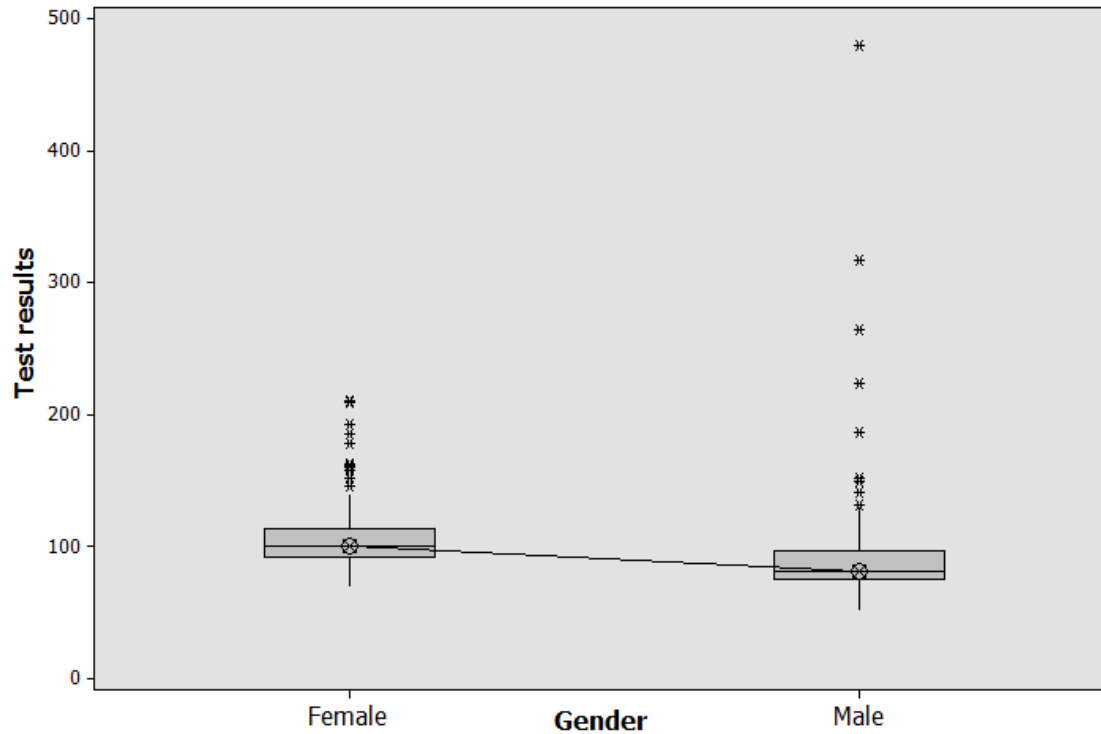
Example – Fertilizers



What other comments would you make about the above boxplots?
Think about the variation as well as the presence of any unusual values

Box Plot

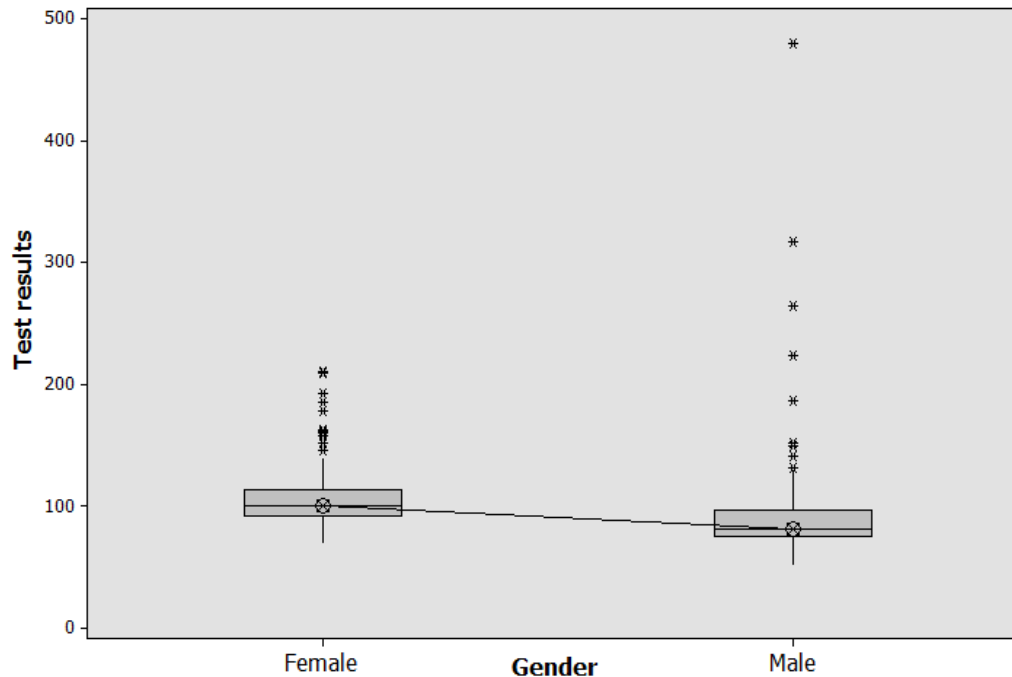
Example – Presence of Diabetes



It is evident that females have in general higher glucose levels than males

Box Plot

Example – Presence of Diabetes



ANOVA can be used here to test the **significance** of the difference between the two means

Box Plot

Further Information

There are many applications and online services that allow the creation of box plots quickly and easily (such as Minitab, JMP, and SPSS)

