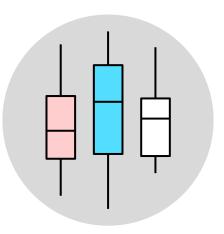
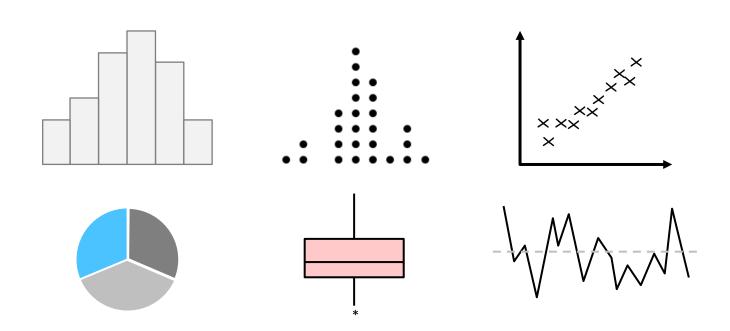
Continuous Improvement Toolkit

BOX PLOT

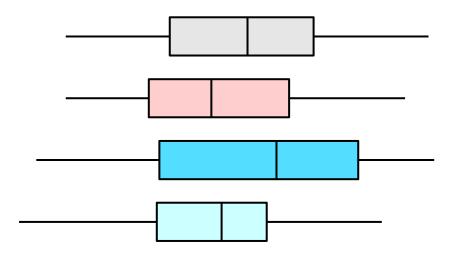


One of the best ways to analyze any process is to plot the data.



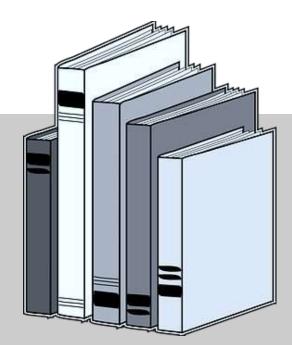
A box plot is a **graph** that shows the frequency of numeric data values.

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



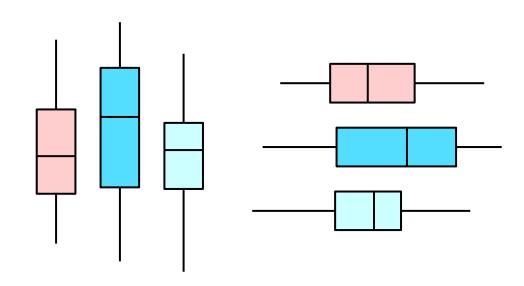
Mainly used to **explore** data as well as to **present** the data in an easy and understandable manner.

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

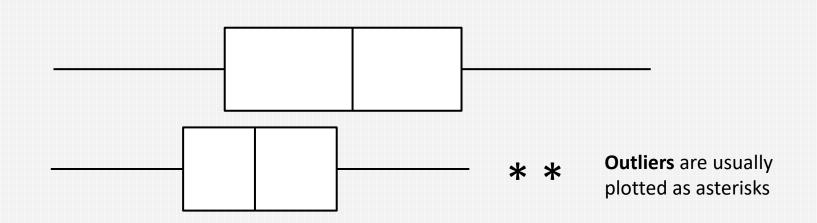


Box plots can be drawn either vertically or horizontally.

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

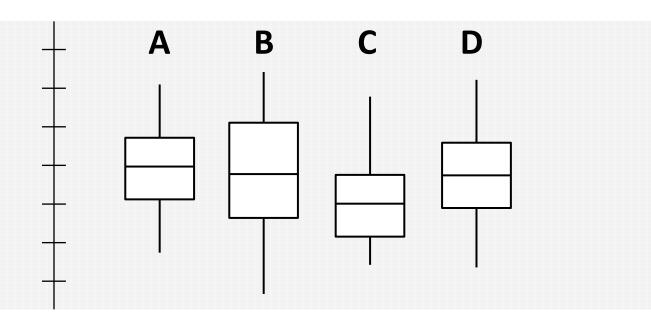


They provide a quick way for examining the central tendency and variation present in the data.



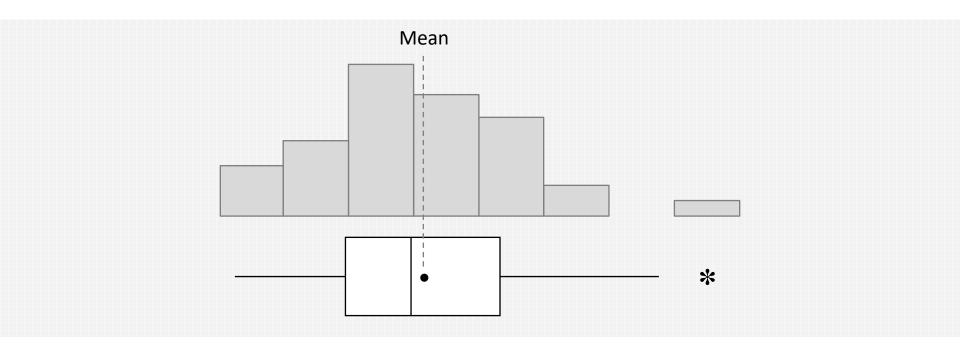
A wider range boxplot indicates more variability

Box plots are useful when **comparing** between several data sets.

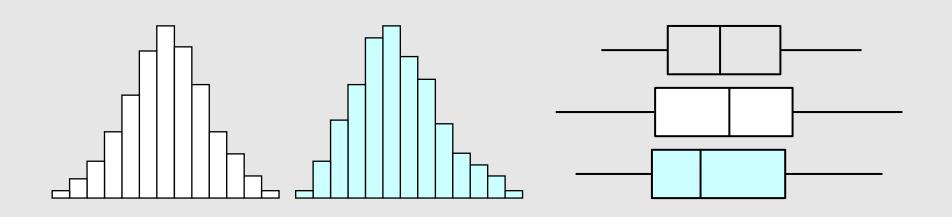


In terms of central tendency and variability

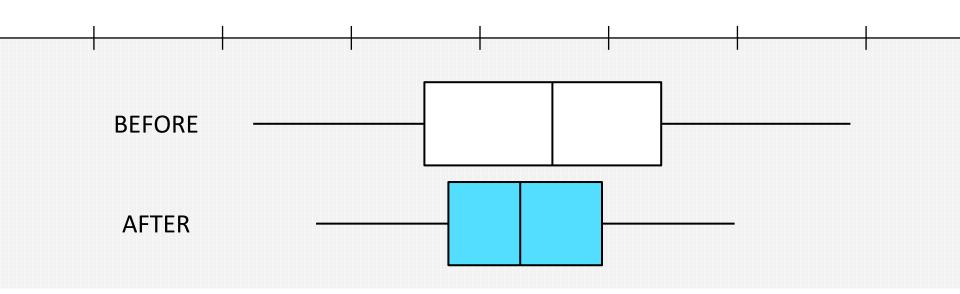
The same **continuous data** can be presented graphically using histograms and box plots.



Less detailed than histograms and take up less space which make them more practical when comparing multiple data sets.

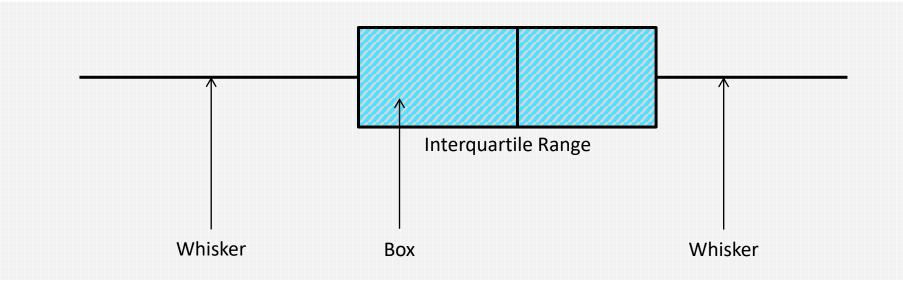


Used to check if there is a significant difference in the process after implementing **process improvement**.



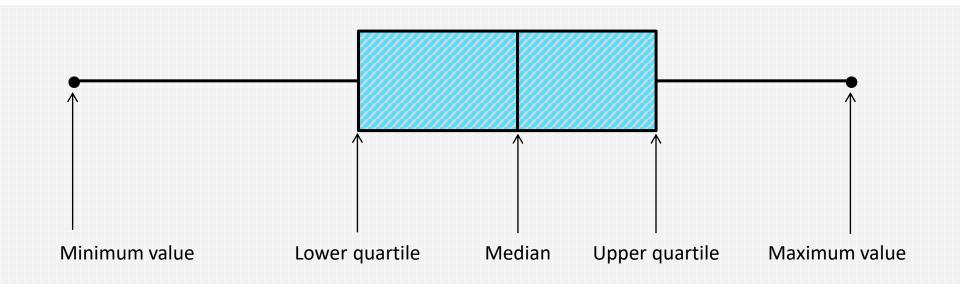
In terms of central tendency and variability

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



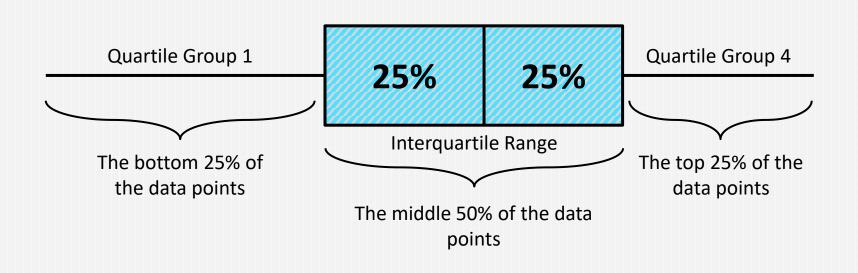
The maximum length of a whisker is limited to 1.5 times the interquartile range (IQR)

Box plots summarize key statistics from the data.



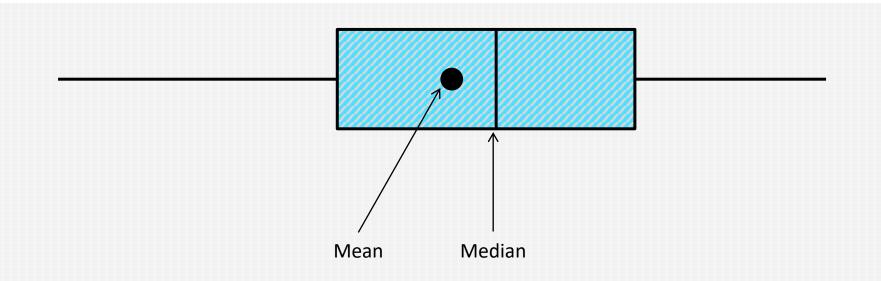
Including the median, maximum and minimum values, as well as the lower and upper quartiles (Q1 and Q3)

The data is plotted such as . . .



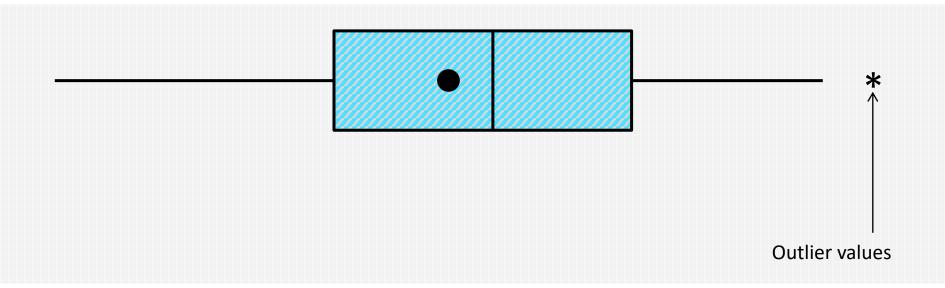
Note: When the median line is not present in the box plot, it suggests that it coincides with one of the quartiles

Sometimes the **mean** is displayed with a special character.

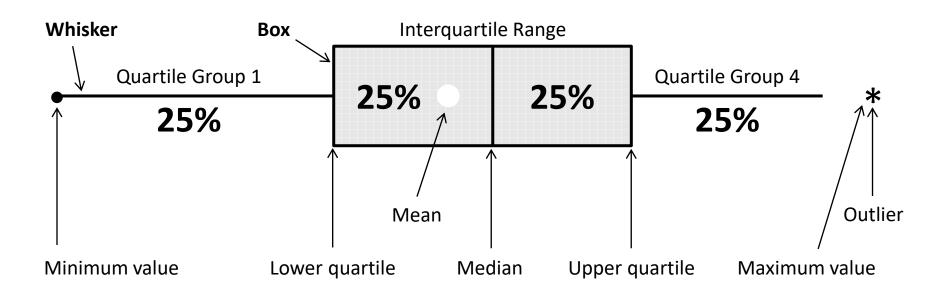


Other character can indicate the mean such as a diamond, a plus, etc.

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.

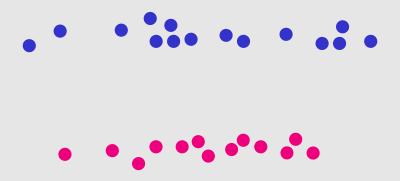


Like histograms, used for moderate to large amount of data.

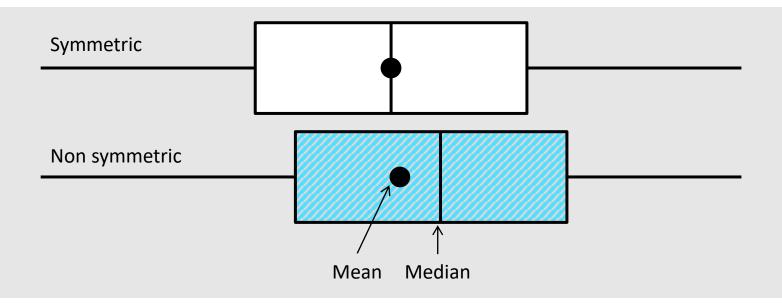


The **size** of the box plot can vary significantly if the data size is too small

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



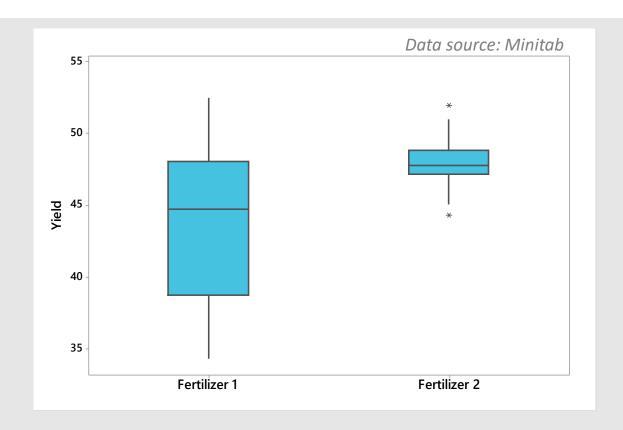
Can tell whether the **distribution** is symmetrical or skewed.



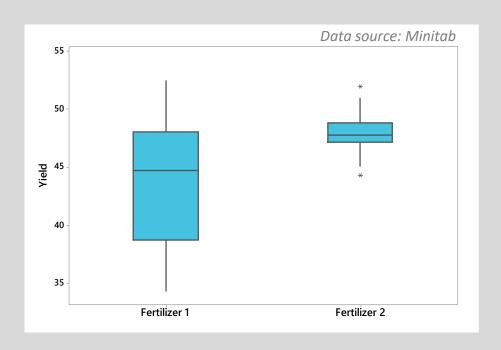
In a **symmetric distribution**, the mean and median are nearly the same, and the two whiskers has almost the same length

Example - Crop Yield

Fertilizer 2 seems to have a higher yield value than Fertilizer 1.



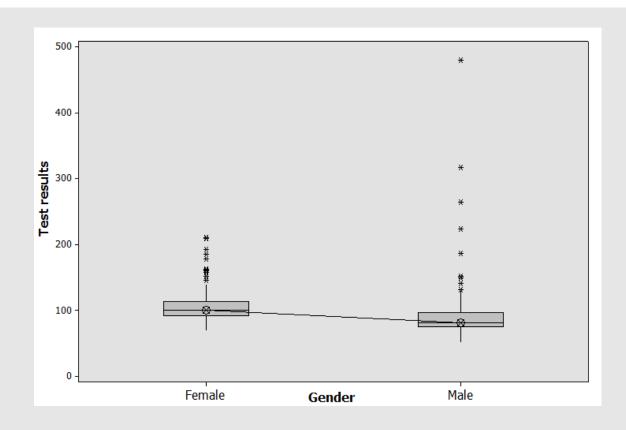
Example – Crop Yield



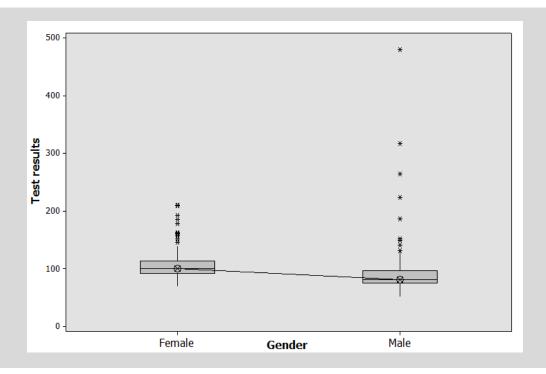
- What other comments would you make about the above boxplots?
- Think about the variation as well as the presence of any unusual values.

Example – Diabetes Test

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



Example – Diabetes Test



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

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).



Made with **y** by

CITOOLKIT

The Continuous Improvement Toolkit www.citoolkit.com

© Copyright Citoolkit.com. All Rights Reserved.