

Descriptive Statistics II.

Analysing quantitative data:

Histograms:

The most common visual for quantitative data.

Quantitative data Has 4 main aspects: [Center](#), [Spread](#), [Shape](#), [Outliers](#).

Measure of spread:

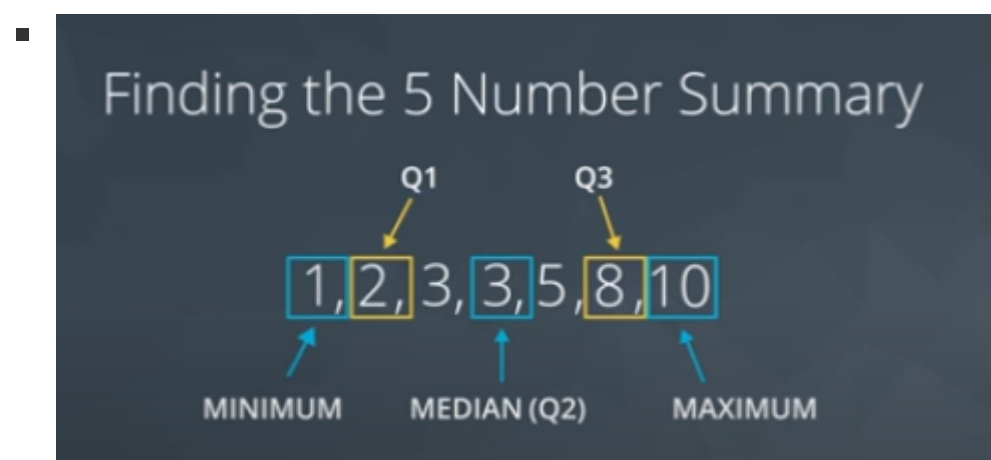
One of the most common ways to measure the spread of data is the [5-Numbers-Summary](#)

[5-Numbers-Summary](#): gives values for calculating the range and interquartile range for a [ordered](#) dataset

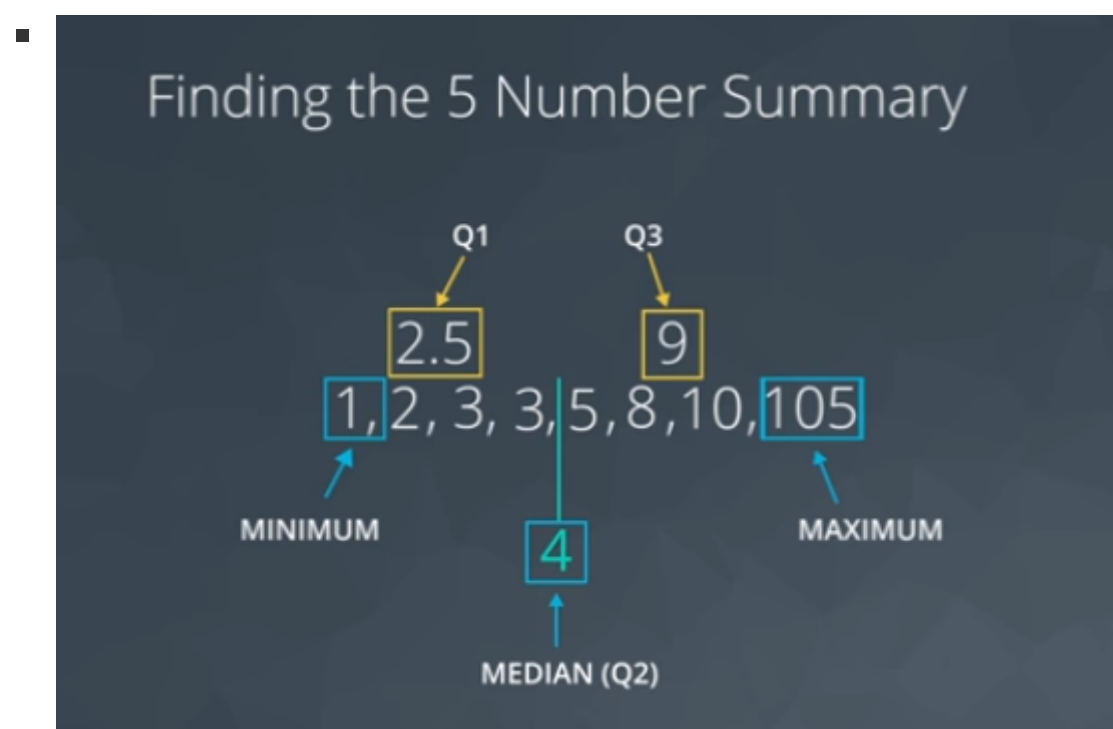
It consists of 5 values:

- [maximum](#) : the biggest value in the data set
- [third quartile](#) : the median of the values between the maximum and second quartile (75% of the data falls below it)
- [second quartile \(median\)](#) : the median of the values
- [first quartile](#) : the median of the values between the minimum and second quartile (25% of the data falls below it)
- [minimum](#) : the smallest value in the data set

◦ odd set of values EX:



◦ Even set of values EX:



The [range](#) is calculated: by subtracting the [maximum](#) from the [minimum](#).

The [interquartile range](#) is calculated: by subtracting the values of the 3rd & 1st quartiles.

The spread of data is measured most commonly using a single value is with [Standard deviation](#) or with [Variance](#).

[Standard Deviation](#): How much each point on average varies from the mean of the points (EX: how much on [average](#) the distance of [each of the employees](#) of a company differs from the [average distance all employees are](#) from work).

(IT IS THE SQRT OF VARIANCE)

[Variance](#): The average squared difference of each observation of data from the mean

Calculating the standard deviation:

- get the [mean](#) (\bar{x})
- [square the difference](#) between each value of the data set and the mean ($x_i - \bar{x}$)
- [get the average squared distance](#) of each observation of the mean ([variance](#))
- [square root the ending value](#) and we get the [standard deviation](#)

◦ EX:

▪

DATASET
10, 14, 10, 6

$$(x_i - \bar{x})^2 =$$

$$(10 - 10)^2 = 0^2 = 0$$

$$(14 - 10)^2 = 4^2 = 16$$

$$(10 - 10)^2 = 0^2 = 0$$

$$(6 - 10)^2 = -4^2 = 16$$

VARIANCE $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{4} (0 + 16 + 0 + 16) = \frac{32}{4} = 8$

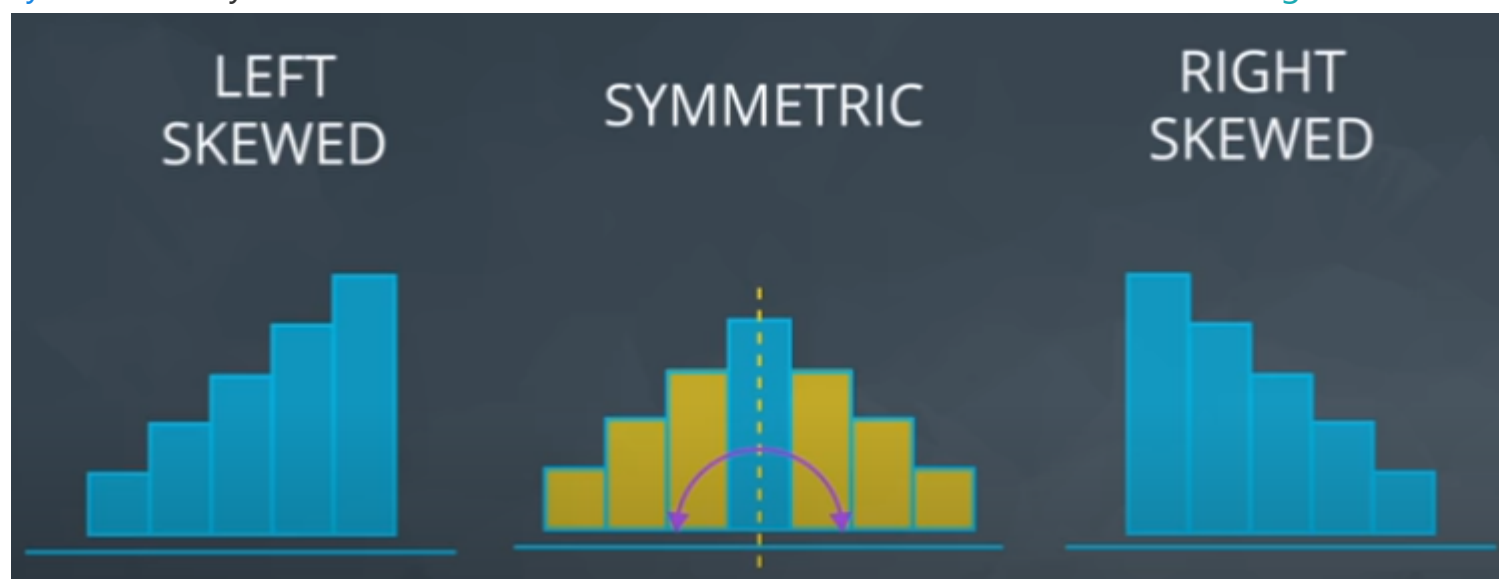
- The standard deviation is the sqrt of the variance
- The higher the mean value is the lower the standard deviation and variance are

Measures Of Shape:

Shape : is how to use histograms to determine the shape associated with data.

here are 3 examples of histogram shapes:

- **Left skewed**: the left most bin is smaller than the right most bin
- **Right Skewed** : the right most bin is smaller than the left most bin
- **Symmetrical** : you can draw a line down the middle and have both sides mirroring



Outliers:

Data points that falls very far from the rest of the data values in out dataset.

with outliers you should:

- Note the impact they have on summary
- Remove / Fix them if they're typos
- Understand why they exist and their impact on questions we're trying to answer
- be careful when reporting and ask the right questions

(When outliers are present it's better to use the 5-number-system instead of the mean or median)