# Components of Descriptive Statistics

Statistics can be categorized into 2 major parts –

* Descriptive - Organizes and summarizes data using graphs and numbers and also measures the central tendency of the workflow values.
* Inferential – Uses sample data to draw a conclusion or make an inference on given subjects and uses probability to determine the accuracy of our conclusions.

We will be discussing descriptive statistics and its components here –

## Descriptive Statistics

### Measuring Central Tendency –

**Mean** - This is the average of all the numbers present in the dataset. Summation of the values/length of the numbers.

**Median** – In case there is a presence of an outlier in our dataset, the mean value can be misleading to determine central tendency. That’s the reason the median method is used. This process sorts all the data and the central-most number is taken as the median value. In the case of even numbers, 2 central numbers are taken and their average is calculated as the median value.

**Mode** - This method is used to determine the central tendency by selecting the number/object that occurred the most number times in the dataset.

### Measuring Spread –

**Range** – The difference between the largest and the smallest number present in the dataset is called the ‘Range’.

**Variance** **-** A dataset consists of many data points. The difference between the mean of the dataset and each of the data points when squared off and the average of these squared values is called variance.

**Inter-Quartile Range** – In a given list of numbers in our dataset, we first calculate the actual median value. It's called **Q2(50th percentile)**. Once we have **Q2**, we have two halves of the list. One half occurs before the actual median, one half occurs after. The median value of the first half is called **Q1(25th percentile)** and the median value of the second half is called **Q3(75th percentile)**, the difference between the **Q3** and **Q1** is our **interquartile range**(**IQR**).

**Outlier -**  Outlier is the number or the numbers which stays out of the range calculated as below.

**[Q1 – 1.5\*IQR, Q3 + 1.5\*IQR],** yellow marking is the lowest number and the green marking is the largest number in the range specified. Any number that resides outside this range will be called an outlier.

Dependence –

**Co-relation -** co-relation is a statistical measure that describes the size and direction of a relationship between 2 or more variables. However, a co-relation in a variable doesn’t necessarily mean that the change in one variable is the cause of changing the value in the other variable.

**Causation -**  In causation it implies that the change in one variable is the result of the change in other variables present.