

**Basic 0f Statistics:**

1. Why we need statistics
2. Applications of statistics
3. Types of Data

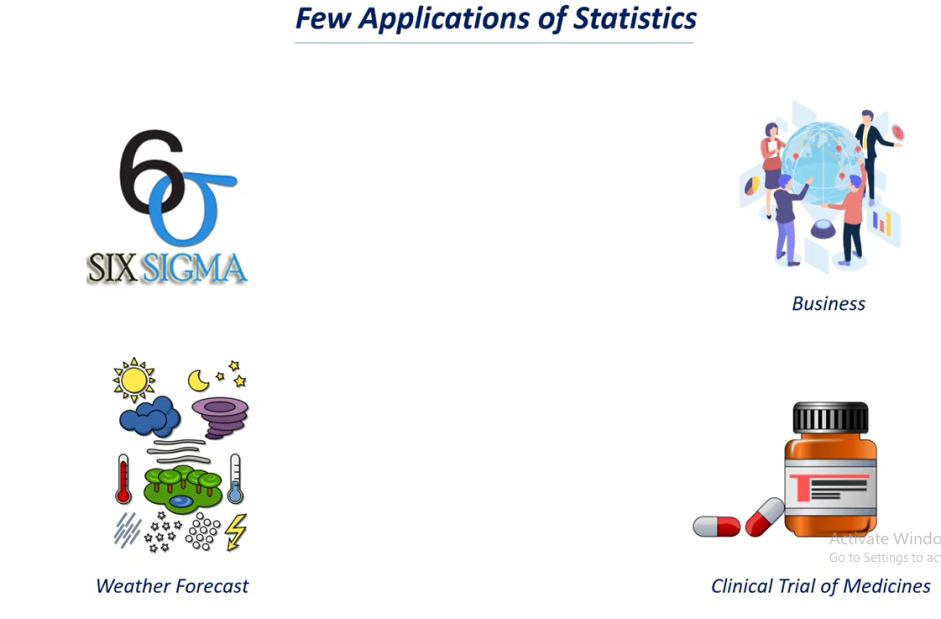
**Why we need statistics or what is the importance of statistics**

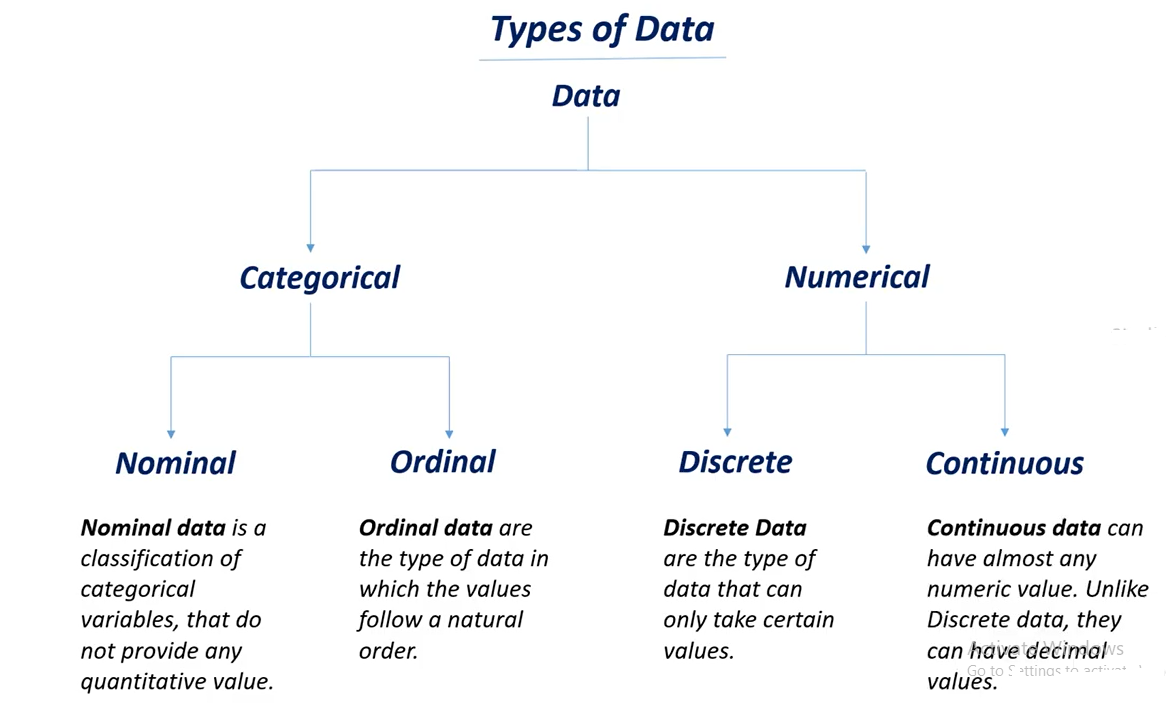
Statistics is a tool that helps us to extract information and knowledge from data

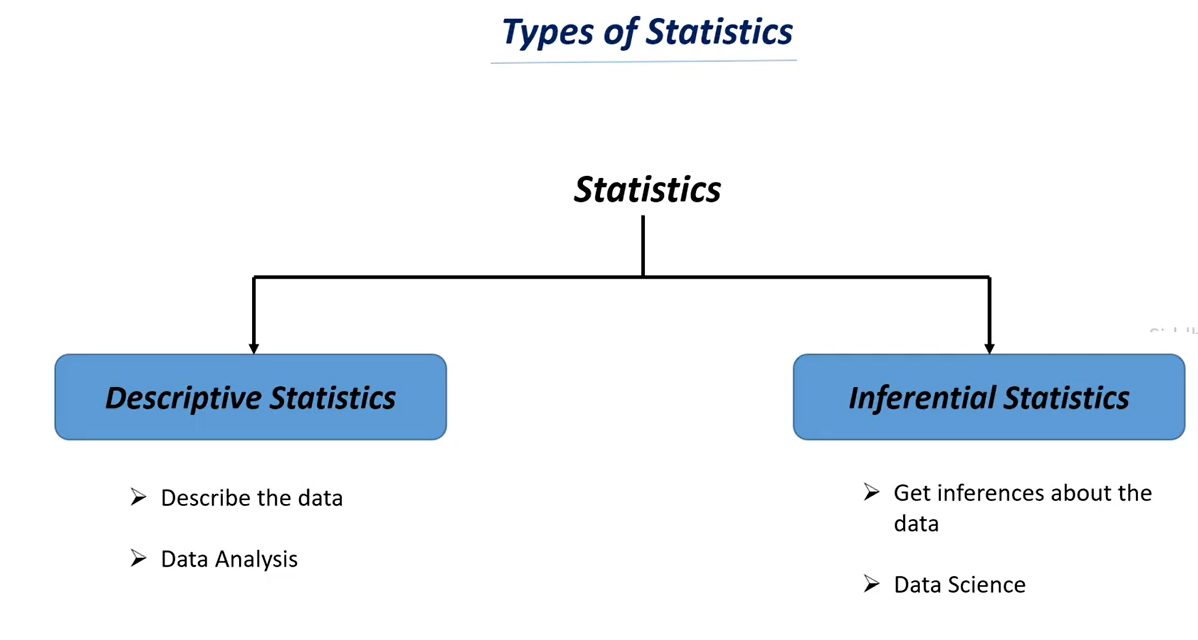
Lets say that we have data and this statistical measure or this statistical technique can help us to find what is the significant of this data.

Avg is nothing but mean

Mean = All the observation / total no. of observation.





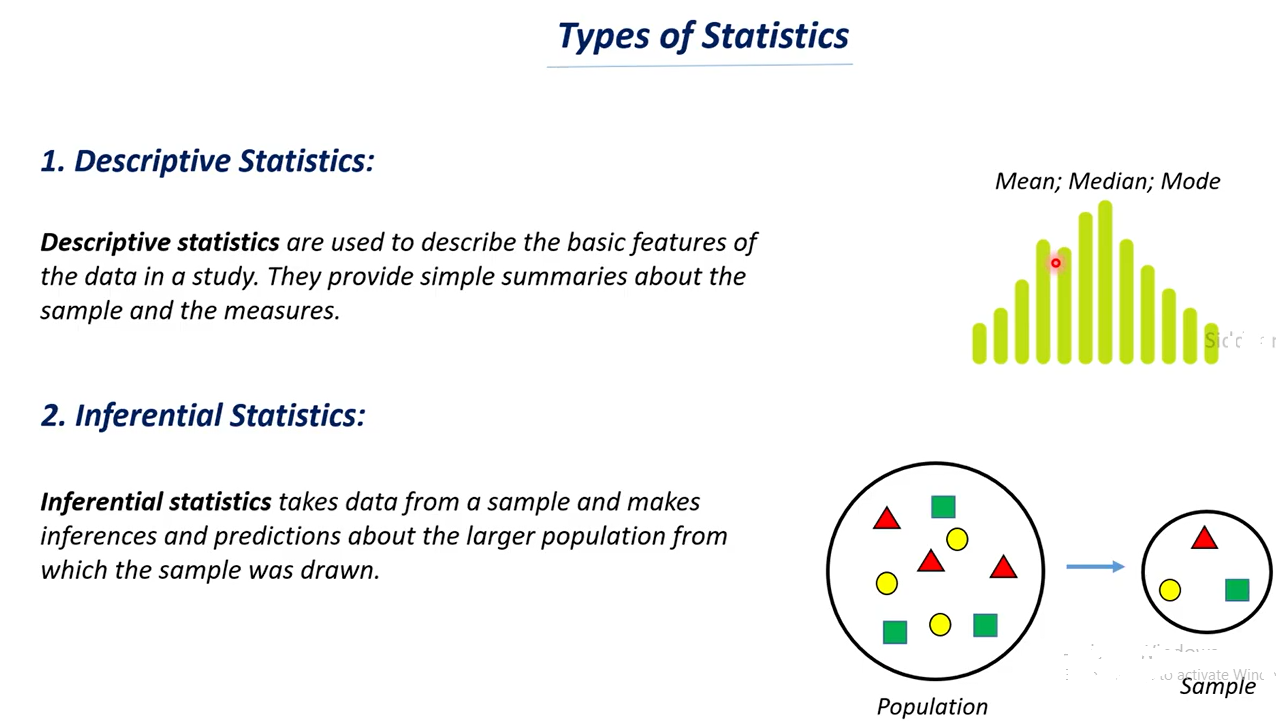


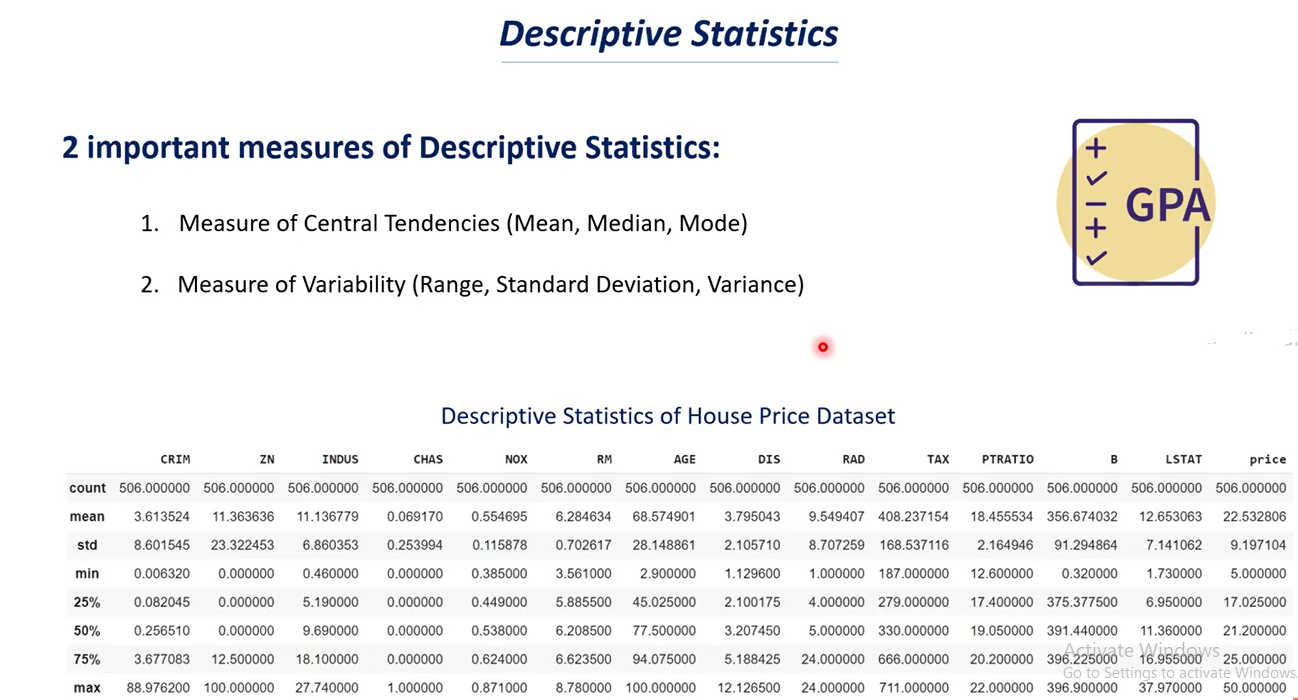
* **Descriptive statistics** means giving some description or giving some summarization to the data.
* **Inferential statistics** means finding some inferences from the data and finding insights from the data and we also try to predict future cases.

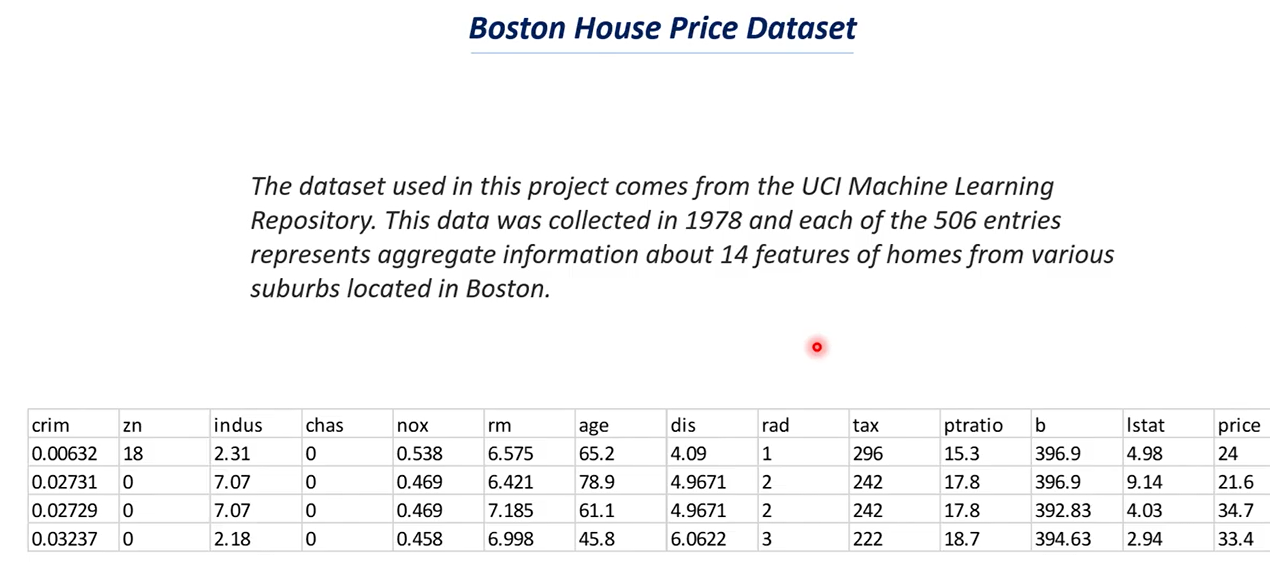
**Descriptive statistics Example: There is data analysis to analysis there previous year’s sales and analysis can go through the sales data and find’s some interesting insights from the data like profit, loss from the data.**

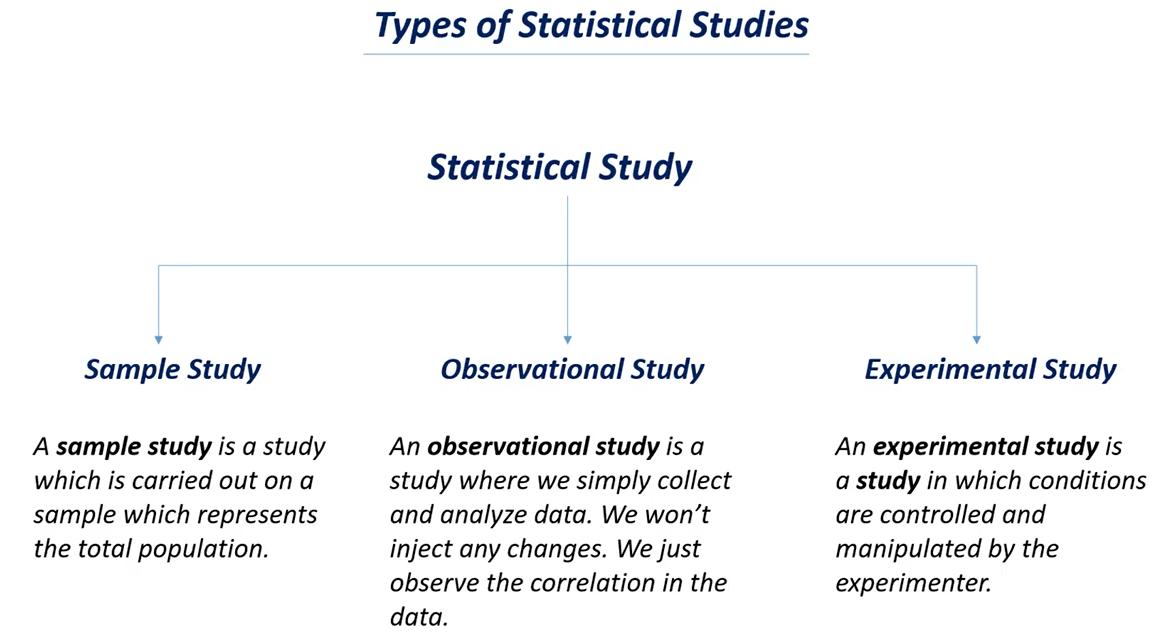
**In case of descriptive statistics we use some numerical measures from the data.**

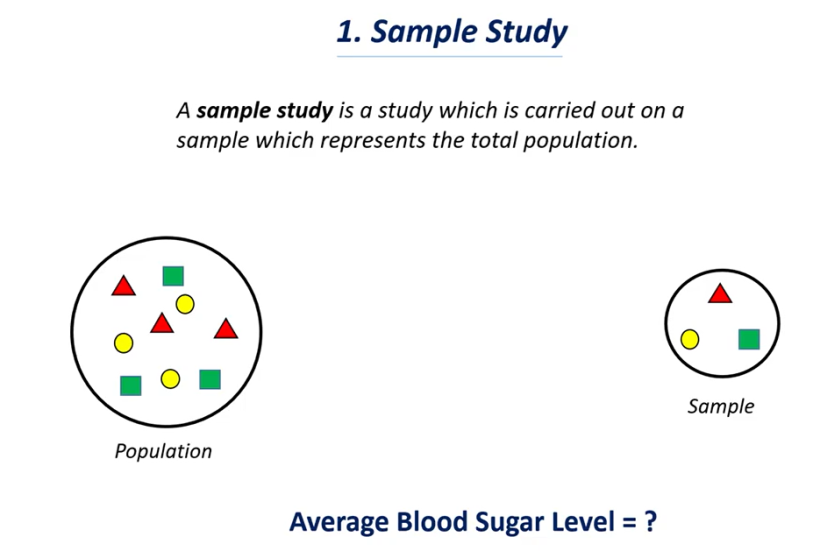
**Whereas data scientist do is he can use the data and try’s to find the insights from the data or find the inferences from the data based on these inferences he can suggest the management of the company to obtain various strategies to increase profit percentage.**

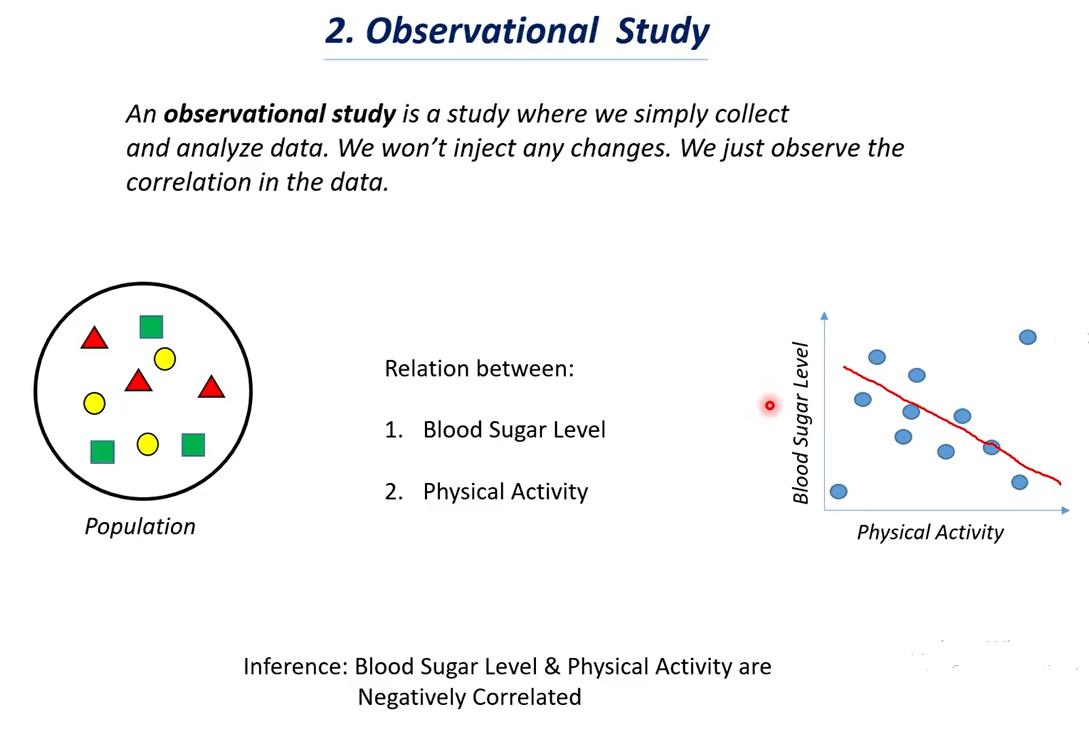


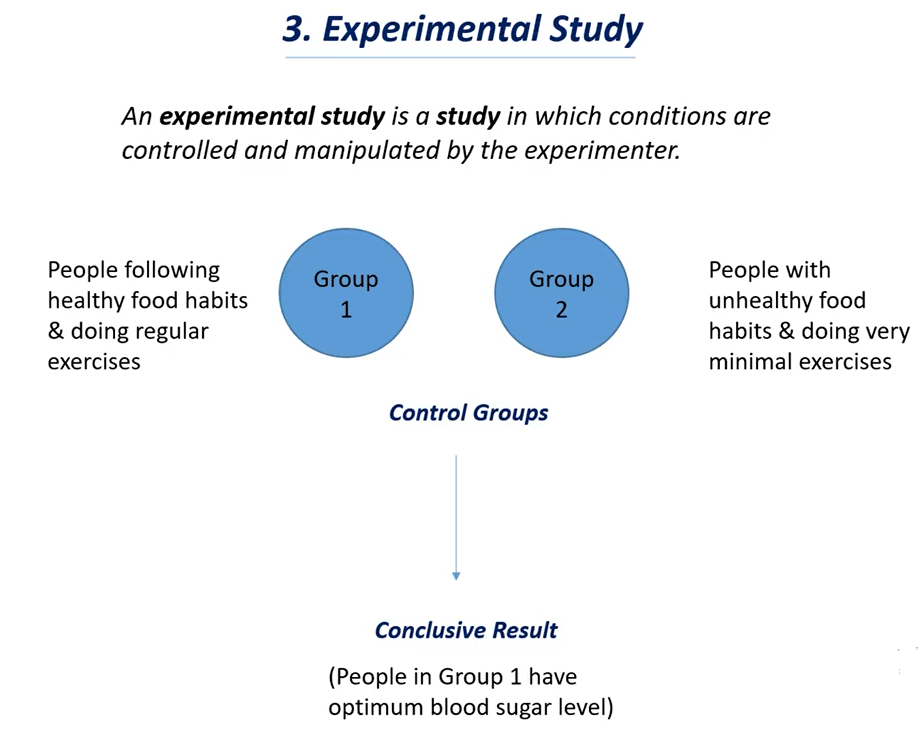


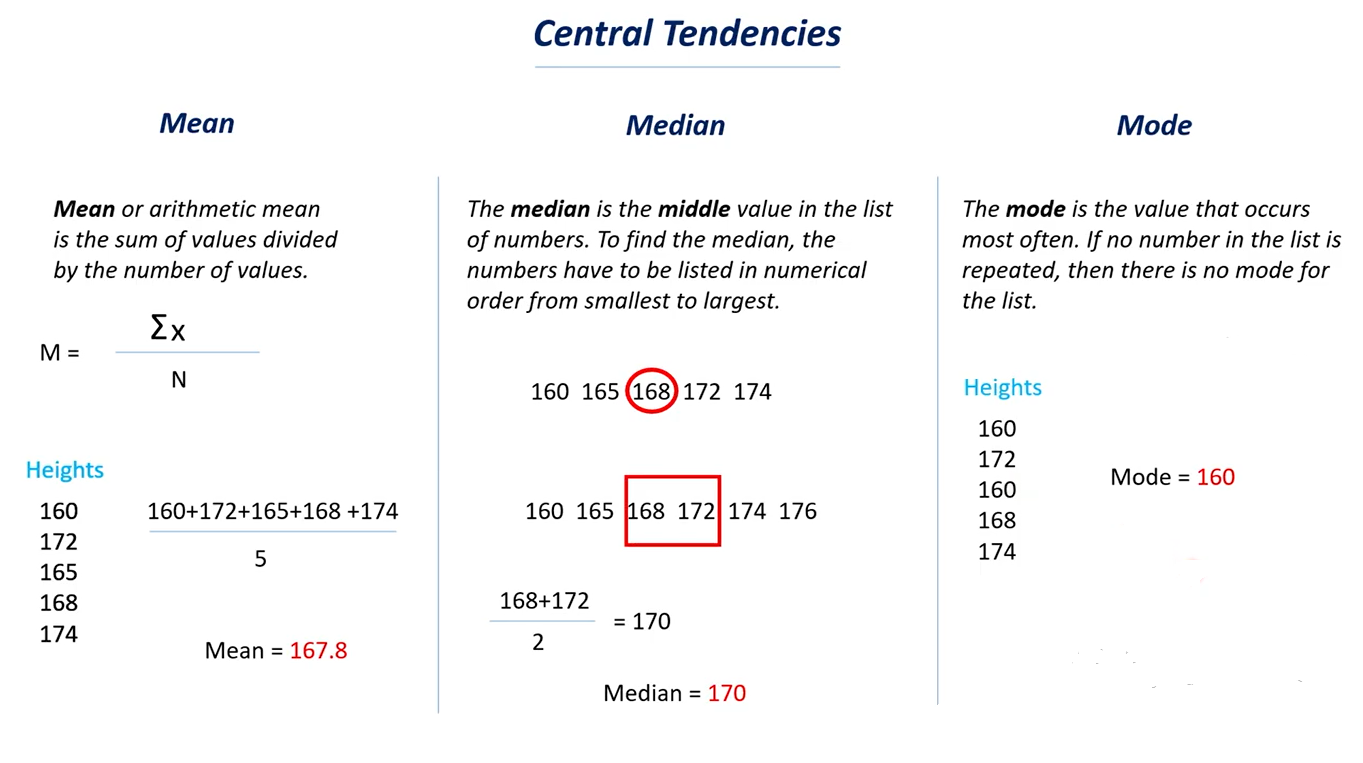












**You should have a question in mind that when we have a mean why we need median**

* The importance of median is mean can’t be a good measure in all cases.

i.e mean is a good measure but not for all cases.

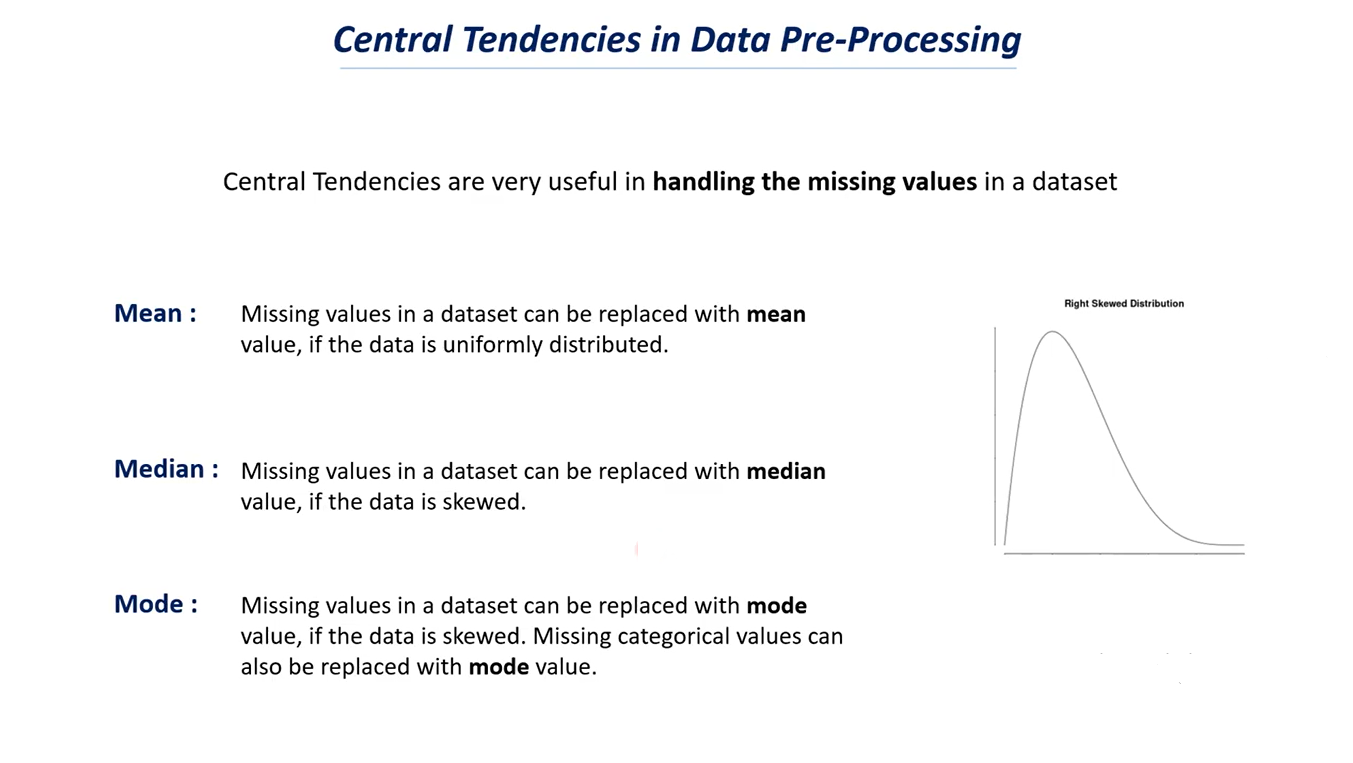
Consider the same example of heights we have one more value as 185cm he is a very tall person, so we

Have total 6 observation or value in that case here if we calculate mean the mean value will be very huge and if you compare all the data points with the mean value there will be a very huge variation .so this is an outlier and if we see other values they are in similar range and one value is in a huge range.

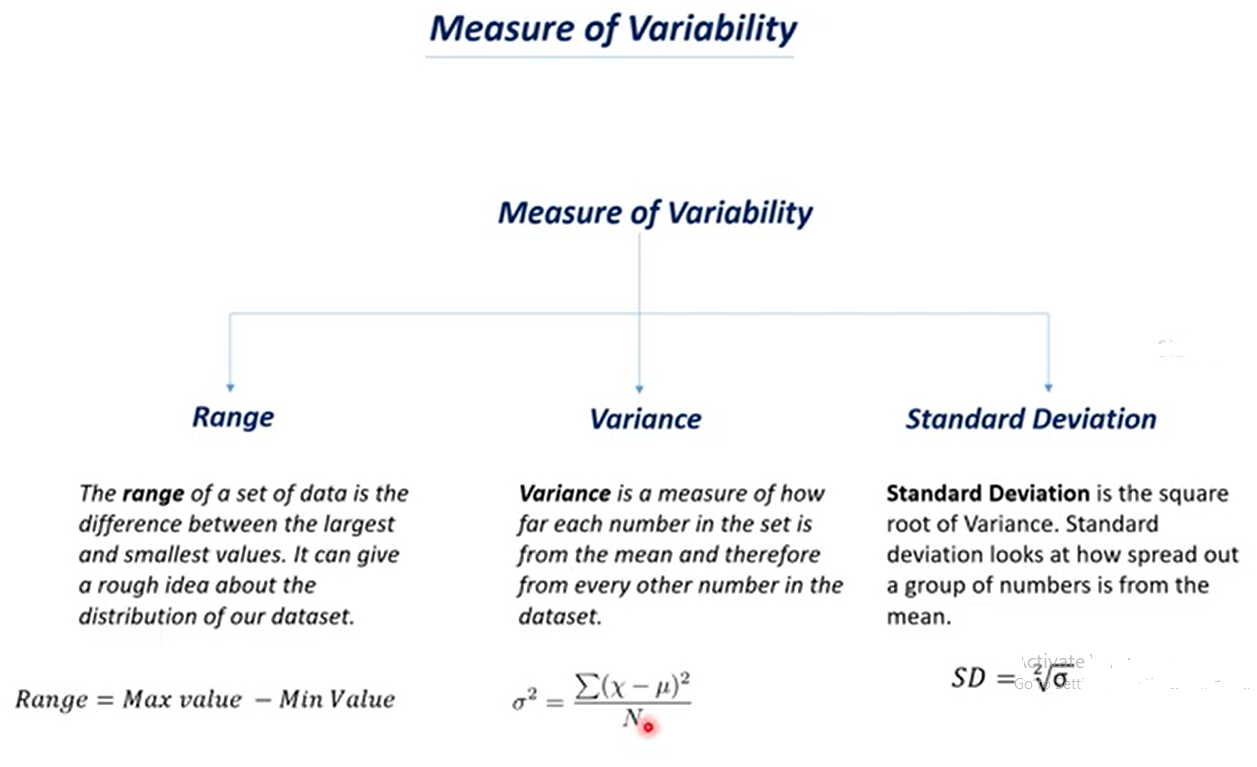
In that cases when you have outliers in your data mean value will be affected drastically because of that outliers.

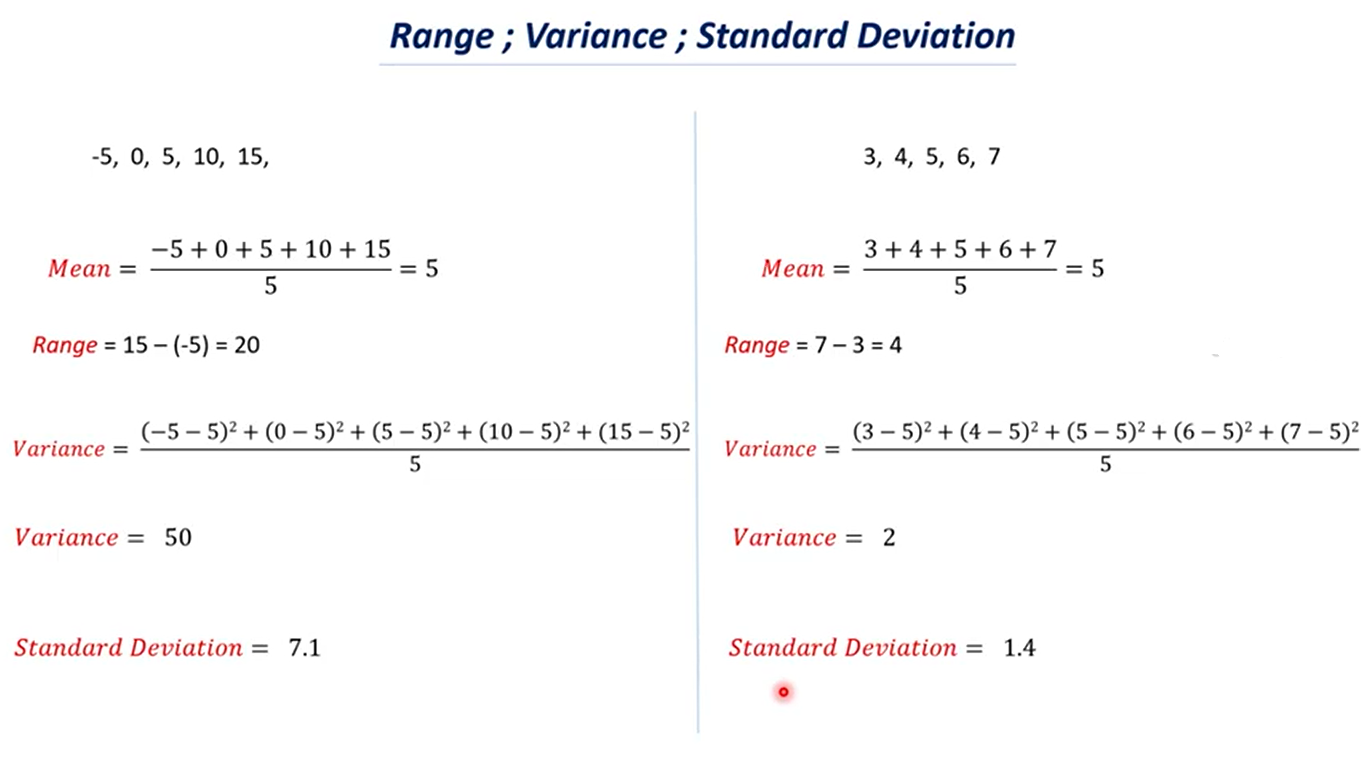
If you have outlier it is best to arrange the data in order and calculate median.

In this cases median will be good measure as central value.



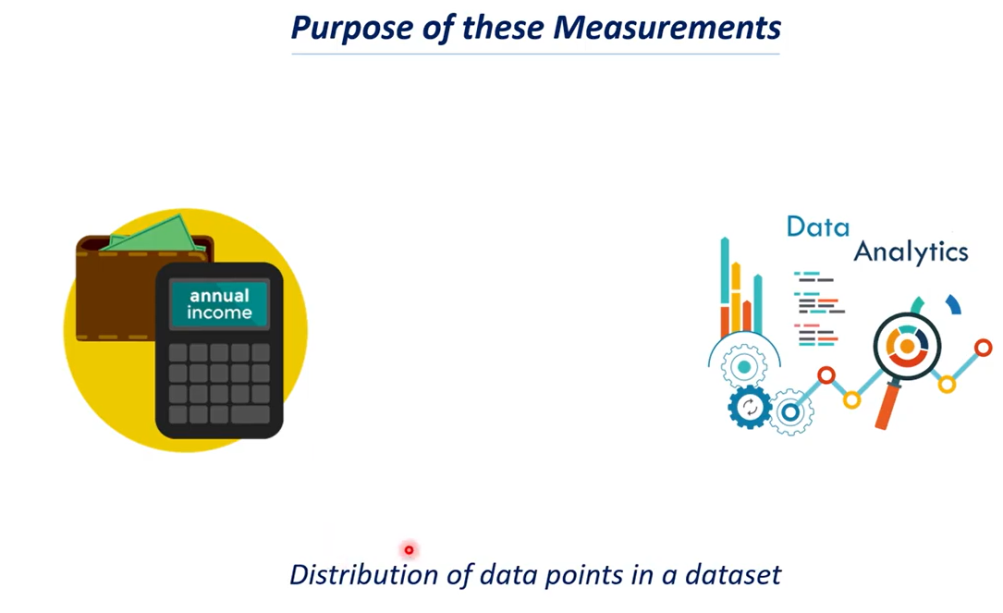
**Measure of Variability - Range, Variance, Standard Deviation**

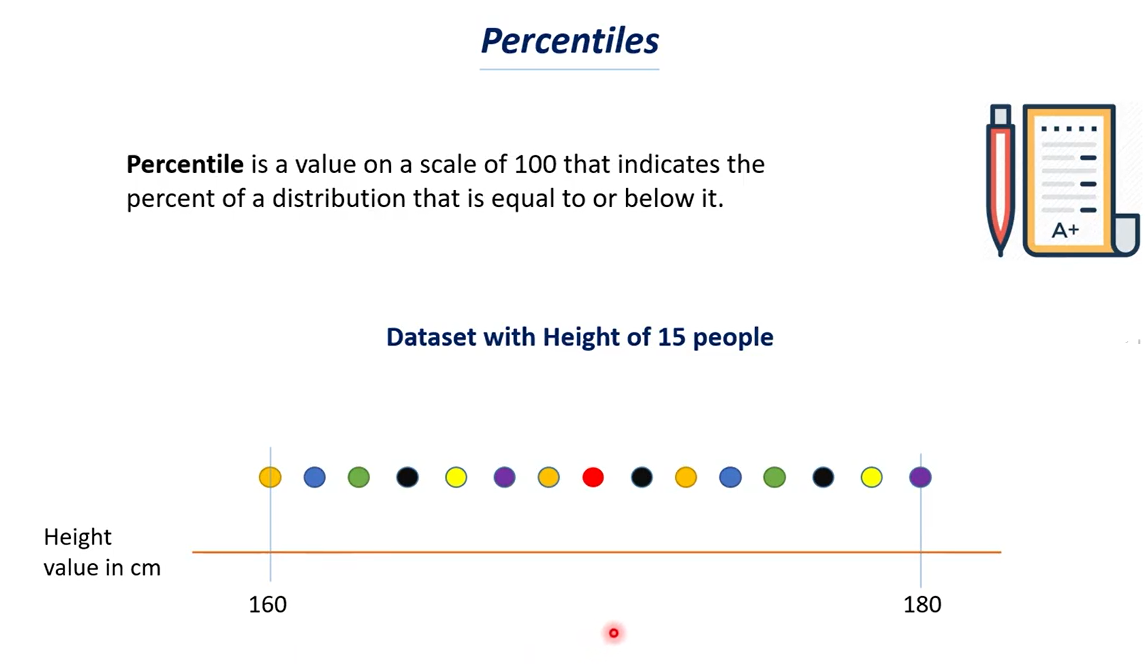


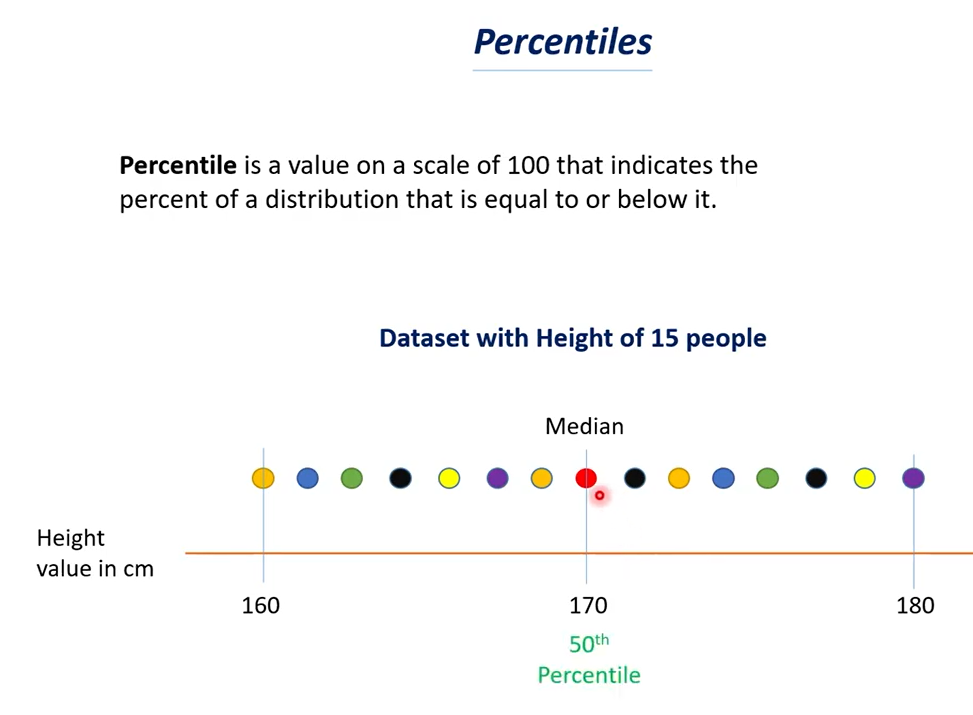


* We observed these two datasets are different but they have the same mean values.
* We know that mean value is representational value or summary value.
* In both these dataset have same mean value now can I say both these two datasets are similar.
* This is where the measure of variability come in grows.
* So sometimes mean value doesn’t give up much information when you think measure of variability it gives us weather these two datasets are similar or different.

# Percentiles and Quantiles







Calculate median for this dataset say its 170 which is 50th percentage.

