computational because data science typical involves some sort of algorithm, methods which are written inside code.

statistical because inferences based on statistics helps us to build predictions that we make.

Raw pata

Also caued automic data, source data, unprocessed data.

This information may be stored in the file and it can contains collection of numbers. characters, etc.

This data can be entered in computer or it can be generated by computer.

It is hard to passe or analyze.

- Dota analysis includes processing or cleaning of data.

Data scientists job is to do processing of such data.

Processed Oata

is perform on raw data to get i processed data, which can be analyzed and visualized.

Example: Voltage signal of microphone is raw data and when voltage is filtered and noise is removed, it is processed data.

Processing can include merging, subsetting. transforming, etc.

sources of row data:

- Binary file generated by measurement machine
- Unformatted excel file
- Ison from twitter API
- Hand entered numbers (readings) you collected.

8. multivariate Data

- multivariate is more complex form of statistical analysis technique
- Used when there are more than two variable in the data set.

Example: An advertiser want to compare the popularity of four advertisements on a website, then their click rates could be measured for both men and women and relationship between these variables can be examined.

Data Processing:

It occurs when data is collected and translated into useful information.

At initial, the data is unprocessed or raw data so to use or make that data clean we perform data processing on it to get processed data.

six stages of data processing:

Data collection, Data preparation, Data input, Processing, Data output, Data strage

pata collection

Is the first step of data processing is the process of measuring, collecting and analyzing different type of data using set standard validation technique.

Data Preparation

once data is collected, it then enters

the data preparation stage.

Data preparation, often referred as

" Pre-processing" is the stage at which

raw data is cleaned.

Purpose of this step is to eliminate row data &

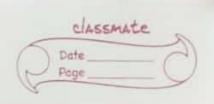
begin to create high quality data.

Data visualization and its techniques

note visualization is graphical representation of information and data by using visual elements like graphs, charts and maps.

Data visulitation convert large and small data sets into visuals, which is easy to understand and process for humans.

Tools & techniques are required to analyze vast amount of information.



Data visualization are common in every day life, but they always appear in the form of graphs & chasts.

The term inference and prediction both discribes tasks where we learn from data in a supervised manner in order to find a model that describes the relationship between independent variables and the outcome

Scatter Plot

Scatter plot uses dots to represent values for two different numeric variables.

The position of each dot on horizontal & vertical axis indicates values for an individual data point.

uariables.

scatter matrix

Each scatter plot in grid is created between different combinations of variables.

matrix represents bi-variate or pairwise relationship between different combination of variables while lying them in and form.

Histogram

Histogram provides visual representation of the distribution of data set:

location, spread and skewness of the data.

- It also helps to visualize whether the distribution is symmetric or skewed left or right.

Box plot

Also known as five Number summary,
mininum
upper
tower quartile or
lower quartile or
median
maximum

It allows you to see important characteristics of the data visually.
Box plot is a graph that gives you a good indication of how the values in the data spread out.