what is statistics?

statistics is a science of collecting, organizing, and analyzing of data.

Then what is data?

facts or a piece of information.

ex: **Ages** of students in classroom.
{24,25,32,29,28}

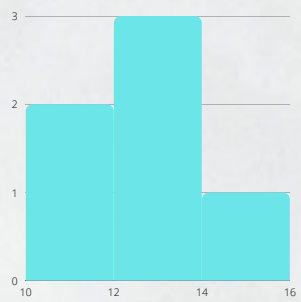
ex: **Weights** of students in classroom.
{85,65,72,73,89}

what are the types statistics?

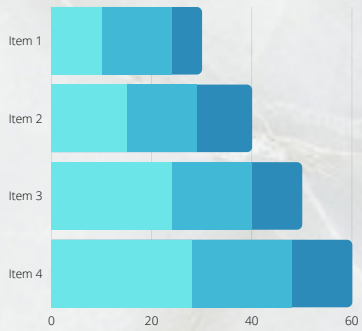
1.Descriptive statistics
2.Inferential statistics

Descriptive Statistics

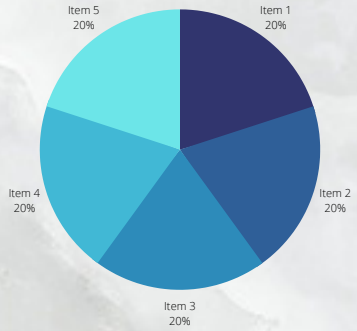
it consists of organizing and summarizing of data using different plots.



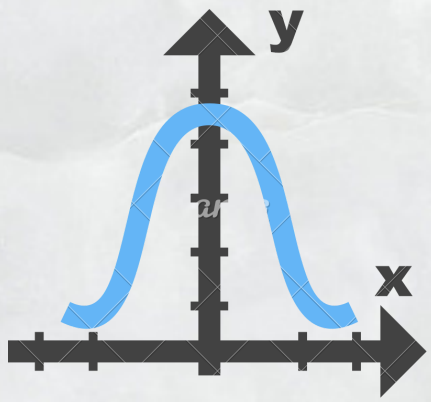
Histogram



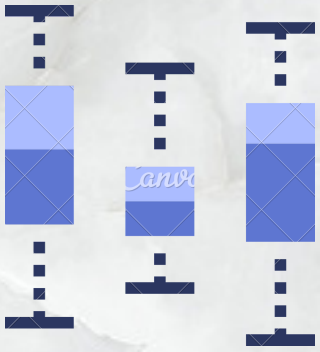
Bar chart



Pie chart



Distribution



Box plot

Candle stick chart

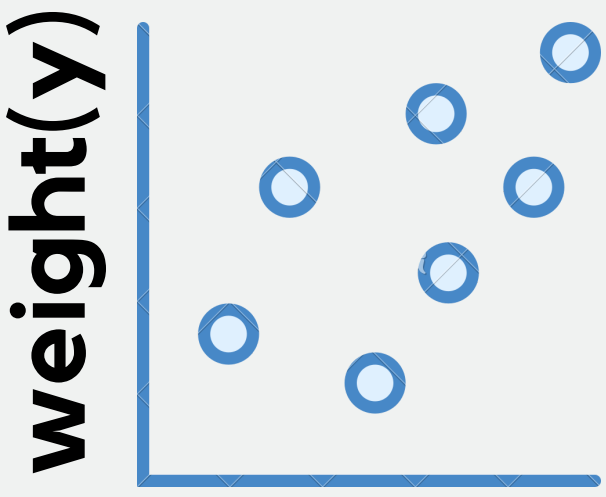
Scatter Plot

Let's how can we apply descriptive statistics on data ?

example: we have a data of students of height and their weight

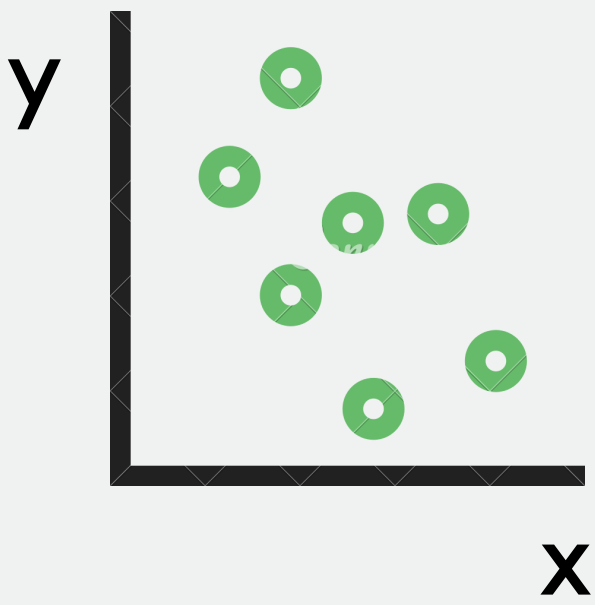
Height(X)	**Weight(Y)**
160 cm	50kg
162 cm	54kg
165 cm	53kg
170cm	60kg

from the above data points we plotted a scatter plot.



from this plot we can observe **x increases** then **y increases.**

we can know that with the help of scatter plot we can find the relation ship between two variables



in this plot we can find that if **x increases** then **y decreases**

X ↑ Y ↓

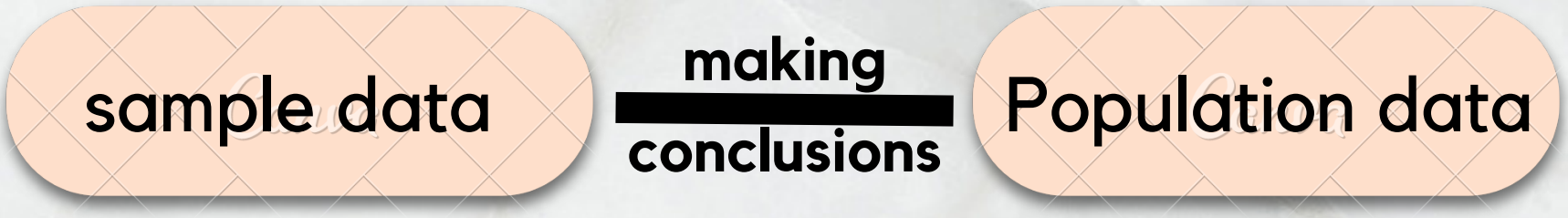
Inferential Statistics

It consists of collecting sample data and making conclusions about population data using some experiments.

Making conclusions can be done by **Hypothesis testing**

ex: In a university there are 500 students.from it we are selecting a class of 60 students.

with the help of 60 students we are finding the average of them and making the conclusions about 500 students.



Sample Data Vs Population data

ex: for example if we have 1000 data points which is population and from it we are considering 100 data points which is known as sample.

- 1000 data points is **population data** and denoted by **N**
- 100 data points is **sample data** and denoted by **n**

What type of questions can be asked and how can you distinguish among descriptive and inferential statistics?

ex: Let's say there are 20 classrooms in a university and you have collected age and weight of students in one classroom?

- age = { 21,20,18,24,28,27,25}
- weight = { 60,65,56,68,69,65}

Descriptive stats:

- what is the average age of students in the class room?
- Relationship between age and weight?

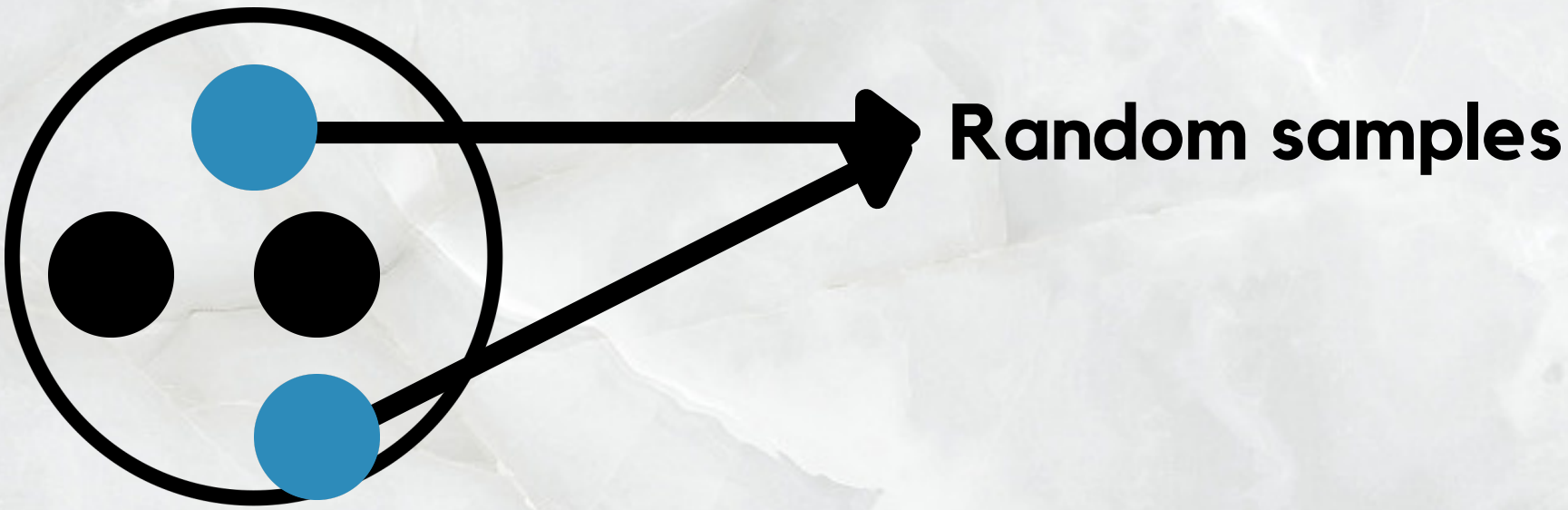
Inferential stats:

- Are the average age of the students in class room less than the average age of the students in the country?

Sampling Techniques

Simple random sampling

Every member of population has an equal chance of being selected for your sample.



we can perform random sampling for exit polls and general survey.

Stratified sampling

group the population and then partitioned into subpopulations

Gender → **Male**, **Female**

Education → **High school**, **UG**, **Masters**, **Phd**

Exit poll survey can be done with people only has 18+ age

- > 18 + can vote
- < 18 can't vote

Systematic sampling

selecting every n'th individual out of population.

- selecting every 3rd individual out of population.

Convenience sampling

Only those who are interested in the survey can participate.

if it is a data science related survey then people who are intrested init can participate.

Questions for you?

1. what type of sampling if it is a survey regarding new technology?

2. what type of sampling can be done by credit card sellers?

write your answears in comments.

what is a variable?

- A variable is a property that can take any values.

Ex: Age = 25, weight=65

Ex: age = [25,26,18,32,27] is a collection

Two types of Variables

1. Quantitative Variable
2. Qualitative Variable

1.Quantitative Variable

- Measured Numarically (mathematical operations).

Ex: Age, weight, Rainfall, Temperature, Distance

2.Qualitative Variable

- Categorical variables (based on some characters they are grouped together.).

Ex: Gender,Type of flowers, Type of movies.

- **Quantitative Variable again divided into two types**

    1. Discrete variable
    2. Continuous variable

1.Discrete Variable

- it is a whole number
- ex: no of bank accounts, no of children

2. Contineous Variable

- it is a decimal values
- ex: Height,weight,Rainfall,Speed

Questions for you?

1. what kind of variable is a marital status?
2. what kind of variable is a ganga river length?
3. what kind of variable is a movie duration?
4. what kind of variable is pin code?

write your answears in comments.