Welcome everybody

In this chapter, we are going to know about:

- **Data**

- **Types of data**

- **Characteristics of Data**

- **Examples of different types of Data**

- **We should have proper and**

- **clear understanding about the data.**

- **Before making any sort of analysis,**

- **we have to identify them correctly**
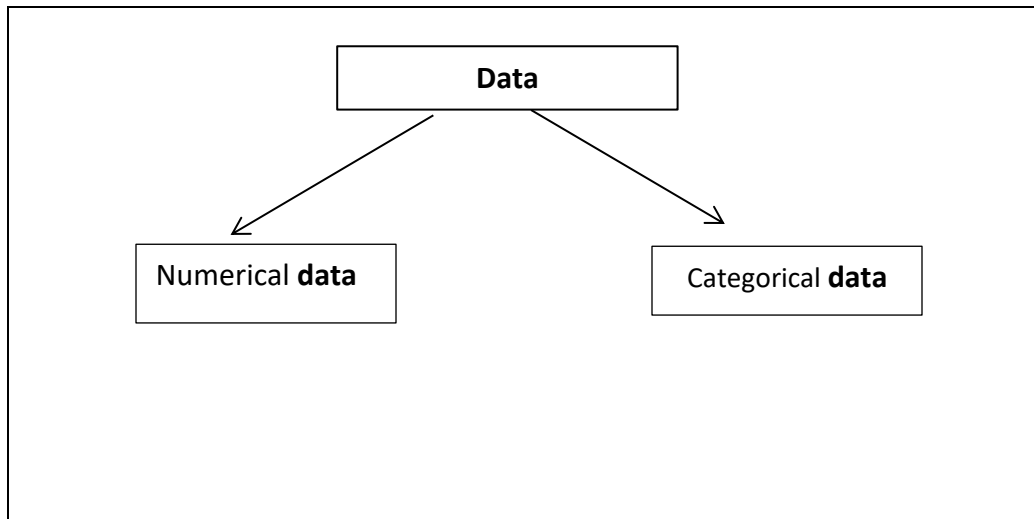
Let us discuss about the Data in details

Data Basics

❖ **In today's world, Data is all around us**

❖ **Data is everywhere**

❖ **It seems the World is made of Data**

❖ **Data, simply means information**

❖ **Your age, your height, your weight, your hobby, the name of your university, your school, college, your department, major, minors**

❖ **Number of friends, your favorite foods, favorite drink, favorite music, films, your birth place, birth date, your CGPA,**

❖ **All are the examples of data**

Types of Data

Data are broadly divides into two parts:

- **Numerical data(Quantitative data)**

- **Categorical data (Qualitative data)**

```
┌─────────────────────────────────────────────────────────────┐
│                      ┌─────────────┐                         │
│                      │    Data     │                         │
│                      └─────────────┘                         │
│                       ╱           ╲                          │
│                      ╱             ╲                         │
│           ┌──────────────┐    ┌──────────────────┐          │
│           │ Numerical data│    │ Categorical data │          │
│           └──────────────┘    └──────────────────┘          │
│                                                              │
└─────────────────────────────────────────────────────────────┘
```
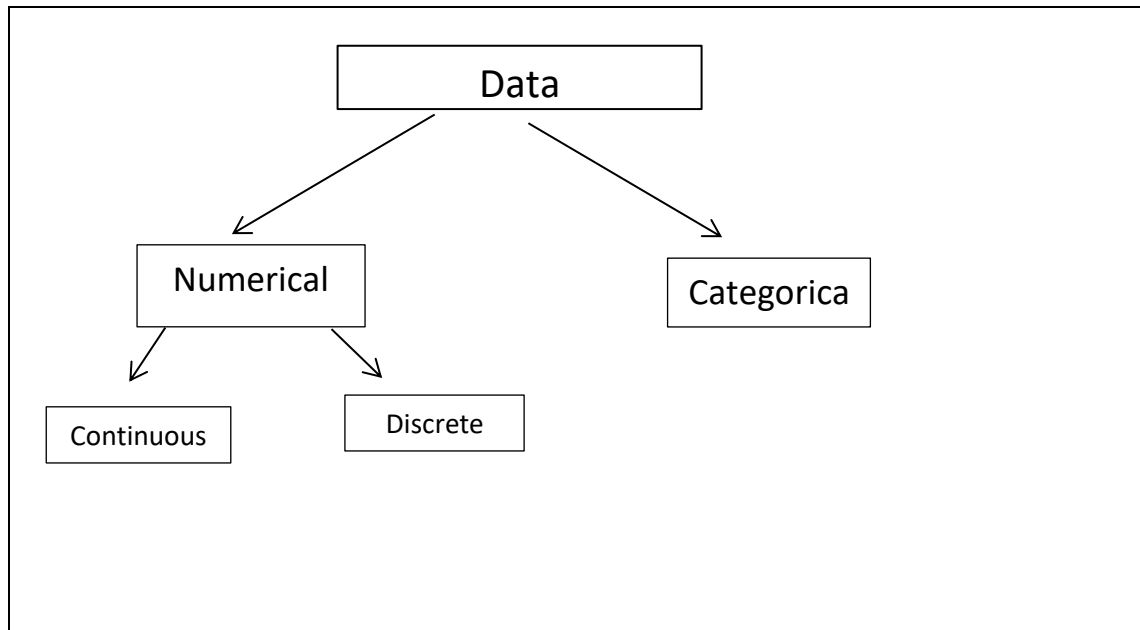
Numerical Data (or Quantitative data)*:*

Numerical data take on numerical values. Numerical variables measure the actual magnitude of some characteristic for each of the individuals or units under consideration. This type of data is also called quantitative data. It is sensible to add, subtract, take averages, etc. with the numerical values.

Example:

- height of a person,

- price of a commodity,

- weight of students,

- CGPA of a student,

- Time

- Temperature

- Number of students in a class,

All are the examples of numerical data

Again, **Numerical data** can also be Discrete or Continuous:

```
                        ┌──────────────┐
                        │     Data     │
                        └──────────────┘
                         ╱            ╲
                        ╱              ╲
              ┌──────────────┐    ┌──────────────┐
              │  Numerical   │    │  Categorica  │
              └──────────────┘    └──────────────┘
                ╱          ╲
               ╱            ╲
        ┌─────────────┐  ┌──────────┐
        │ Continuous  │  │ Discrete │
        └─────────────┘  └──────────┘
```

Discrete Data:

Discrete numerical data are counted, and can take on only whole non-negative numbers can only take certain values (like whole numbers).

Example:

- **number of students,**
- **number of chairs in a room,**
- **number of children in a family,**
- **number of days in a month,**
- **number of courses**
- **runs**
- **wicket number**
- **goals score**
- **points in a game.**

Continuous Data:

Continuous numerical data can be measured, and can take any numerical value (within a range). Thus, the numerical data can be whole number or be fraction

Example:

- **Age of the students**
- **height of a person,**
- **price of a commodity,**
- **weight of students,**
- **CGPA of a student,**
- **Time**
- **temperature**
- **Sales of a shop,**

  All are the examples of numerical data

Note: It is important to note that

- **NID number**

- **Telephone number**

- **PIN/TIN number**

- **Social security number**

Although all these are discrete number but are in the class of categorical data. As the addition or subtraction of these numbers do not carry any meaning result.

Categorical Data (or Qualitative Data)*:*

Categorical data is just sort of descriptive. They are simply names. Categorical data take on a limited number of distinct categories. It is concerned only with the presence or absence of some characteristics in a set of objects or individuals. This type of data is called qualitative or enumeration data and the characteristic is used to classify an individual into different categories is called an attribute.

Example:

- **Name**

- **Your department,**

- **Hobby**

- **Gender,**

- **Colour (red or green),**

- **Passed or failed,**

- **Religion,**

- **rich or poor.**

Categorical data can be divided into two parts Nominal or Ordinal

Nominal Data:

Nominal data is classified by quality (attribute) rather than numerical scale. The levels of the data do not have ordering. A good way to remember all of this is that "nominal" sounds a lot like "name" and nominal data are kind of like "names" or labels.The labels or names used to identify an attribute of the element.

Examples of Nominal Scales are:

- **Gender: Male, Female**

- **Marital status: Married, unmarried, divorced**

- **Eye color: brown, green, or blue**

- **Department: CSE, EEE, MNS**

- **Hobby: Reading. Programming, gardening**

Note that we can only summarize the nominal data by frequency table and cannot compare them.

Ordinal Data:

Categorical variables that have the order or rank or have a rating scale of values are meaningful are called ordinal. For the ordinal data have relative differences and consist of ordering or ranking the differences. Thus the ordinal data can be compared. One can count and order, but not measure, ordinal data. Ordinal data are typically measures of non-numeric concepts like satisfaction, happiness, discomfort, etc.
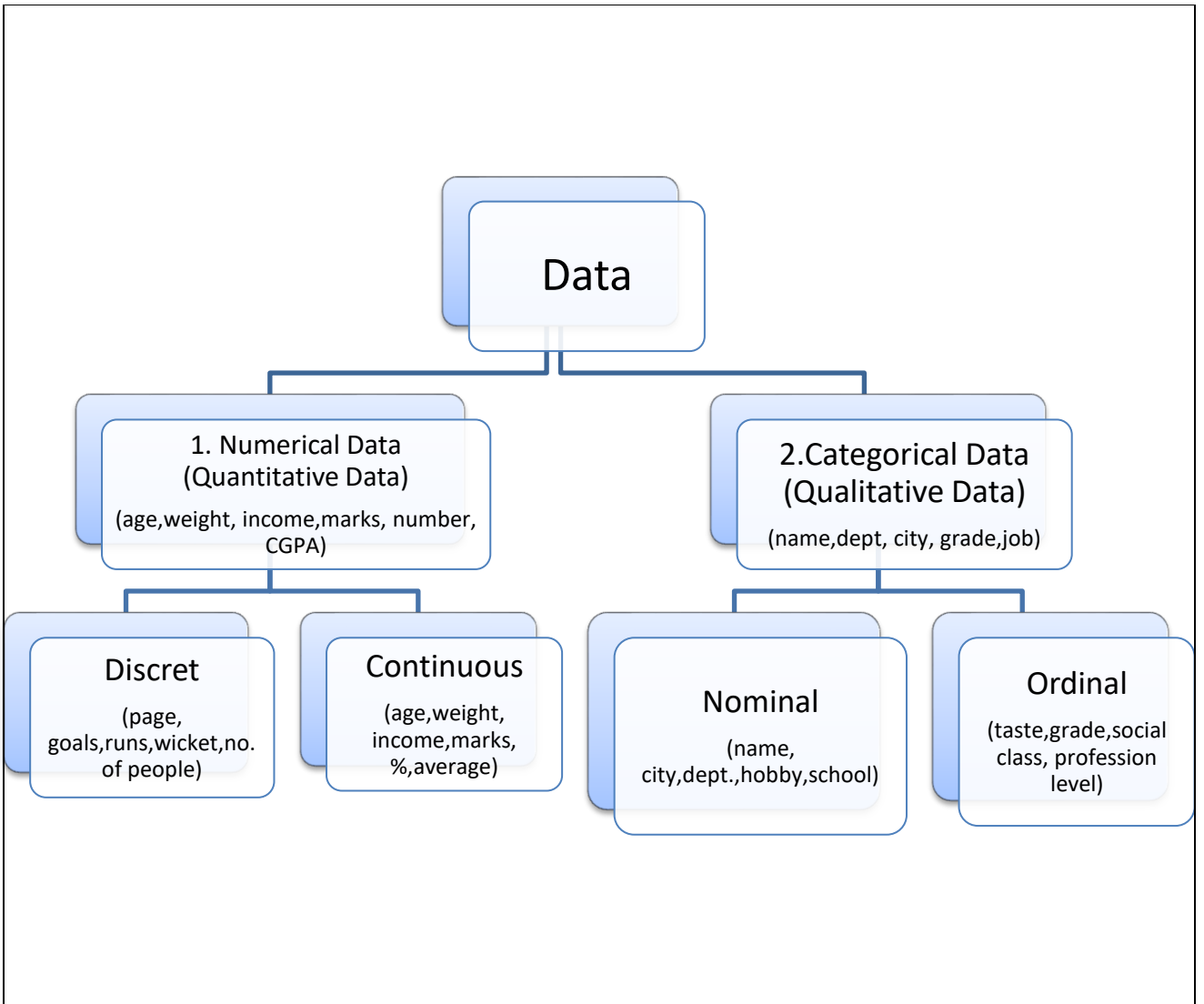
Examples:

- Grades

- Shoe size 4, 5, 6, 7,etc.

- Rich or poor,

- Social Class,

- Level of satisfaction

- Professional level

- Level of education, etc.

Data recorded as excellent indicate the best service, followed by good and then poor; and we can assign 3 for best 2 for good and 1 for poor service.The categories for an ordinal set of data have a natural order, for example, suppose a group of people were asked to taste varieties of biscuit and classify each biscuit on a rating scale of 1 to 5, representing strongly dislike, dislike, neutral, like, strongly like. A rating of 5 indicates more enjoyment than a rating of 4, for example, so such data are ordinal.Thus, the scale of measurement is ordinal.

How do you feel today?
- ◉ 1 – Very Unhappy
- ○ 2 – Unhappy
- ○ 3 – OK
- ○ 4 – Happy
- ○ 5 – Very Happy

How satisfied are you with our service?
- ◉ 1 – Very Unsatisfied
- ○ 2 – Somewhat Unsatisfied
- ○ 3 – Neutral
- ○ 4 – Somewhat Satisfied
- ○ 5 – Very Satisfied

1. Which of the is the numerical data
2. An example of qualitative data is
3. PIN is a
4. ID number
5. Age is a
6. Gender is
7. Floor number is a
8. Time is a
9. Level of satisfaction is a
10. Rank of the player is a

Data

1. Numerical Data (Quantitative Data)
(age,weight, income,marks, number, CGPA)

- **Discret**
  (page, goals,runs,wicket,no. of people)

- **Continuous**
  (age,weight, income,marks, %,average)

2.Categorical Data (Qualitative Data)
(name,dept, city, grade,job)

- **Nominal**
  (name, city,dept.,hobby,school)

- **Ordinal**
  (taste,grade,social class, profession level)

Time Series Data

Time Series Data are collected over time. A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. For example, measuring the level of unemployment each month of the year, number of students are admitted into each year, export or import for each month or year are time series

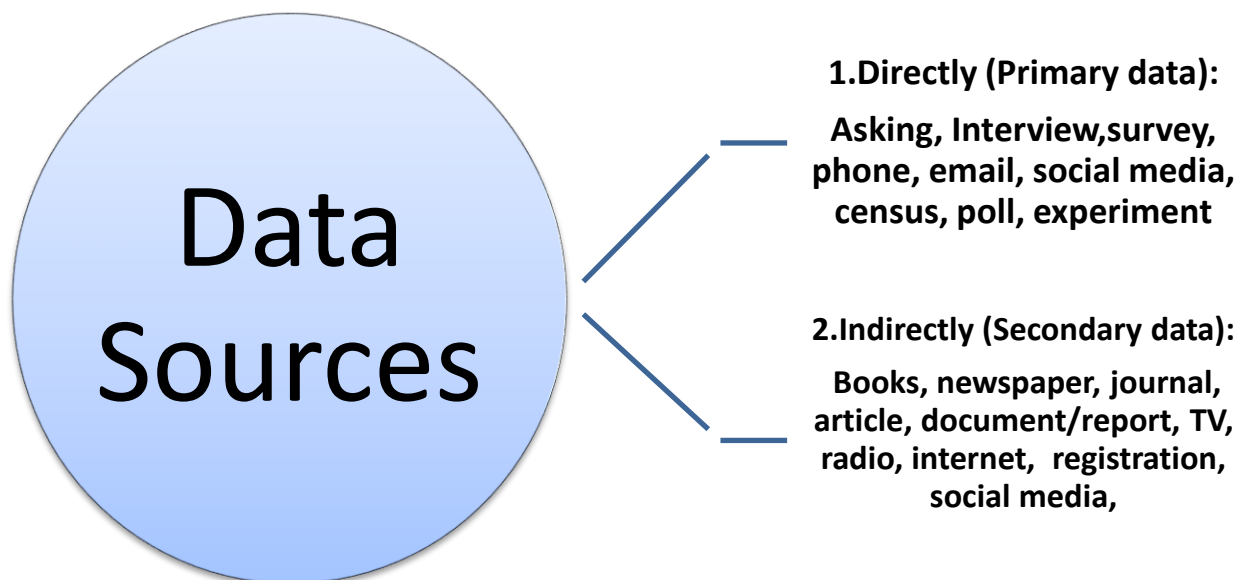**Example GDP of a country for several years is as follows:**

Year	GDP(million$)
2007	300
2008	320
2009	350
2010	455
2011	530

There are many sources to collect dat. Data can be collected

Directly **(Primary data)** or

Indirectly **(Secondary data)**

Data Sources

1.Directly (Primary data):

Asking, Interview,survey, phone, email, social media, census, poll, experiment

2.Indirectly (Secondary data):

Books, newspaper, journal, article, document/report, TV, radio, internet, registration, social media,

Primary data

The data which are collected directly way are known primary data.

Primary data means original data, which were collected specially for a specific study.

Someone collected the data from the original source first hand or directly. Generally, the primary data are collected by an agency or organization.

Methods of direct data collection include:

- Surveys

- Census

- interviews or focus groups discussion

- Observational studies

- Experiments

- Clinical trials

- Social media is a good source of collecting data as a primary source.

Secondary data

The data which are collected indirect way are known secondary data. These data were not originally collected for the purpose of the study. Secondary data is data that is being reused.

Secondary data are usually procured from already published or unpublished documents rather than undertaking first hand field investigation.

Secondary data can be collected by the following ways:

- **Registry,**
- **Website (google, yahoo),**
- **Magazine,**
- **books,**
- **TV,**
- **newspaper,**
- **Radio,**
- **Journals**

Garbage in, Garbage out?

You should have proper data for analysis; otherwise, how well you are making your analysis will go in vain or will be meaning less