MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Unit 1: Types of Data & Experiments

Mamatha.H.R

Department of Computer Science and Engineering

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Unit 1: Types of Data & Experiments

**Mamatha H R**

Department of Computer Science and Engineering

Topics to be covered

❖ Types of data

❖ Variables or Attributes

❖ Types of studies

Data

- Data refers to **individual facts, statistics, or items of information** that are collected through observation.
- It can also be defined as the **facts and figures** collected, summarized, analyzed and interpreted.
- The **data collected** in a **particular study** are referred to as the **data set**.

Source: twinkl.de

Types of data

Source: lh5.googleusercontent.com
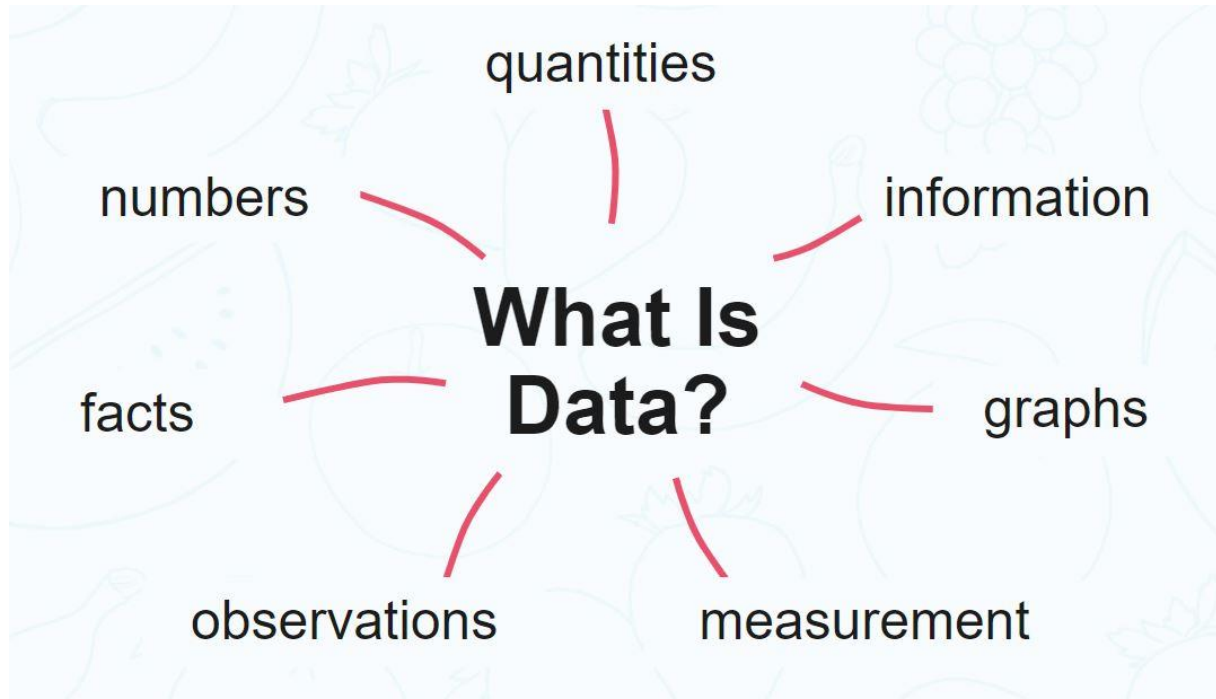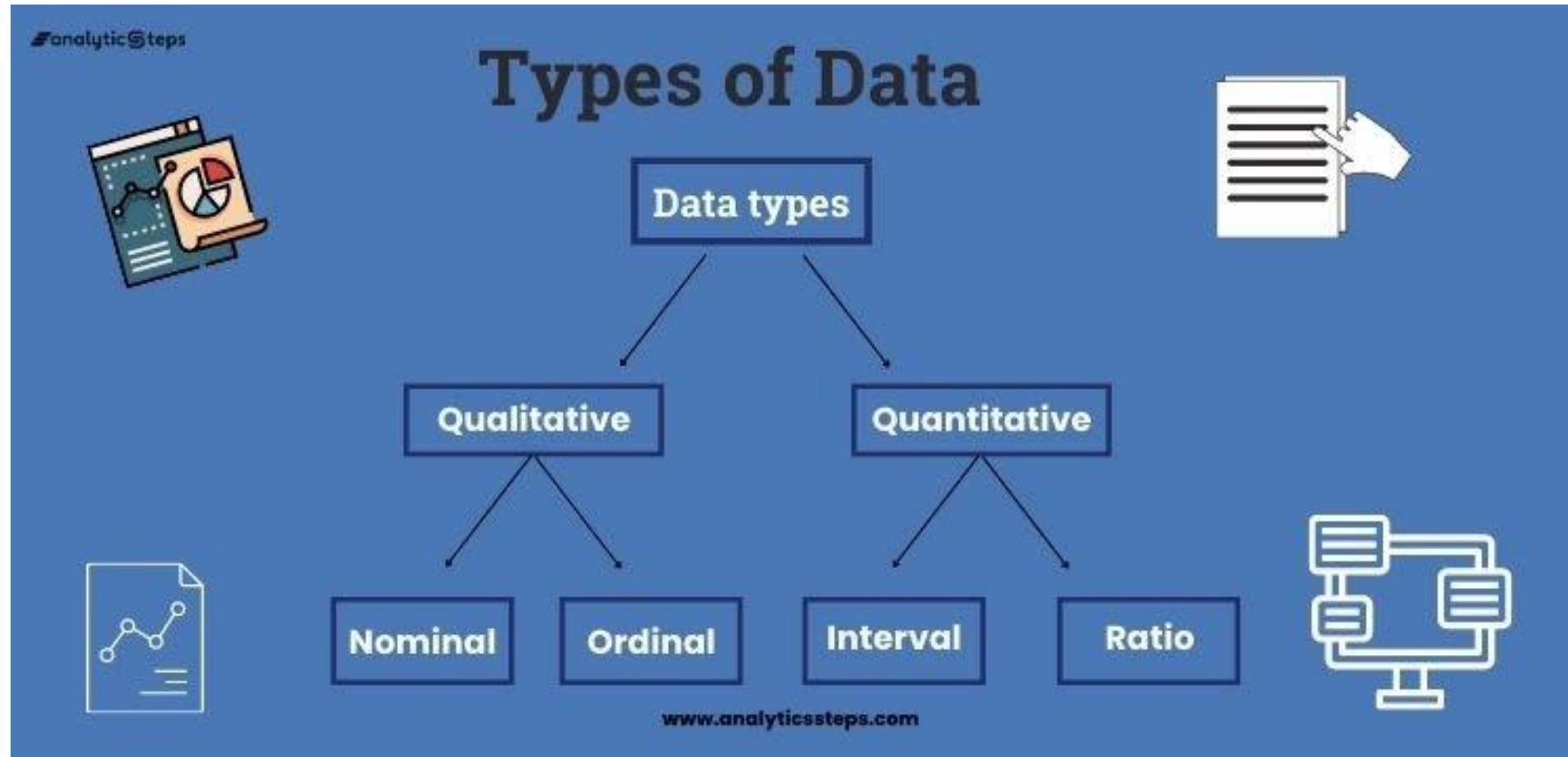
Quantitative data

- Quantitative Data are measurements that are recorded on a naturally occurring numerical scale.
- These are easily open for statistical manipulation and can be represented by a wide variety of statistical types of graphs and charts like line charts, bar graphs, scatter plots, etc.
- These types of data tries to find the answers to questions such as
  - "how many,
  - "how much" and
  - "how often"
- Example: Age, GPA, Salary, Cost of books this semester, Scores of tests and exams, weight of a person, temperature in a room etc.
- There are 2 general types of quantitative data:
  - Discrete data
  - Continuous data

Qualitative data

- Qualitative Data are measurements that cannot be recorded on a natural numerical scale, but are recorded in categories.
- It is also known as Categorical Data as the information can be sorted by category, not by number.
- Example: Year in school, Live on/off campus, Major, Gender, colors etc.
- These can answer the questions like:
  - "how this has happened", or
  - "why this has happened".
- In general, there are 2 types of qualitative data:
  - Nominal data
  - Ordinal data

Qualitative data: Nominal data

- This data type is **used just for labeling variables**, without having any quantitative value.
- Here, the term 'nominal' comes from the Latin word "nomen" which means 'name'.
- They are categories **without any particular order or direction**.
- The nominal data sometimes is referred to as "labels".
- Their **use is restricted** to keeping track of people, objects and events.
- They are **least powerful** in measurement with no arithmetic origin or order.
- Hence, nominal data is of restricted or limited use.
- Examples: Gender (Women, Men), Hair color (Blonde, Brown, Brunette, Red, etc.), Marital status (Married, Single, Widowed) etc.

Slide Courtesy:Dr.Uma

Qualitative data: Nominal data

- It can't be manipulated using mathematical operators.
- But, it can be visualized using pie chart.
- Nominal data can be both quantitative and qualitative.
- Quantitative labels lack a relationship.

NOMINAL DATA

Nominal data divides variables into mutually exclusive, labeled categories.
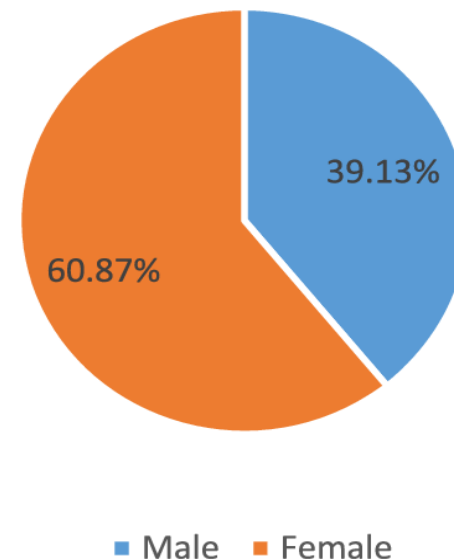
Examples

Eye color
Blue Brown Green

Smartphone
iPhone Samsung Moto

Transport
Bus Train Car

How is nominal data analyzed?

Descriptive statistics: Frequency distribution and mode

Non-parametric statistical tests

Pie chart: Male 39.13%, Female 60.87%

Source: dpbnri2zg3lc2.cloudfront.net, researchgate.net

Slide Courtesy:Dr.Uma

Qualitative data: Nominal data

→ **How to analyze Nominal Data?**

- Using grouping method.

- Group them into categories.

- For each category, frequency or percentage can be calculated.

- Hypothesis testing is carried out using nonparametric tests such as Chi-Square test.

- To determine whether there is a significant difference between the expected frequency and the observed frequency.

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Qualitative data: Nominal data examples

Gender, marital status or any alphabetic / numeric code without intrinsic order or ranking.

Sl. No.	Subject	Code
1	Physics	P
2	Chemistry	C
3	Mathematics	M
4	Biology	B

Source:
www.slideshare.net

Slide Courtesy:Dr.Uma

Qualitative data: Ordinal data

- In simple words, we can understand the ordinal data as qualitative data for which the values are ordered.
- Ordinal data may indicate superiority.
- But, we cannot do arithmetic operations with ordinal data because they only show the sequence.
- Based on the relative position, we can also assign numbers to ordinal data. For example, "first, second, third…etc."
- Ordinal data allows for setting up inequalities, but it has no absolute value.
- More precise comparisons are not possible.
- Examples:
  - Ranking of users in a competition: The first, second, and third, etc.
  - Rating of a product taken by the company on a scale of 1-10.
  - Economic status: low, medium, and high.

Slide Courtesy:Dr.Uma
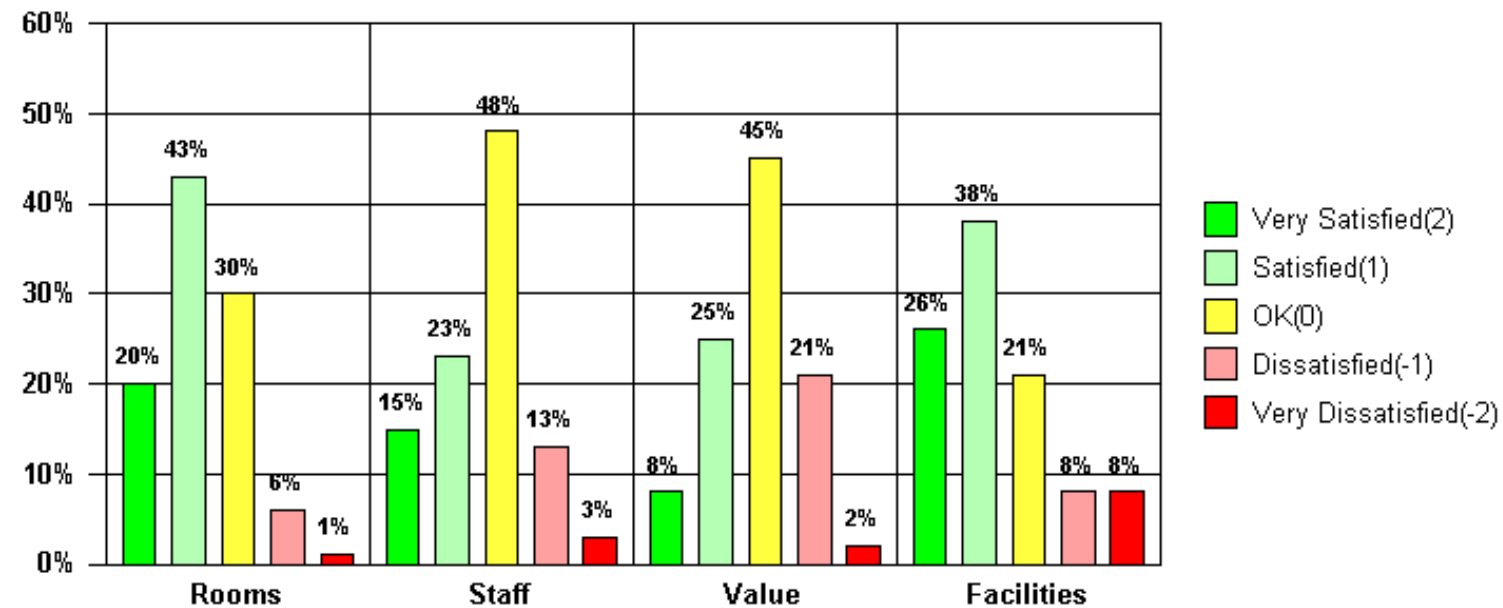
Qualitative data: Ordinal data

- Here, the order matters but not the difference between values.
- Example: Pain Scales
  - Patients are asked to express the amount of pain they are feeling on a scale of 1 to 10.
  - A score of 7 means more pain than a score of 5, and that is more pain than a score of 3.
  - But the difference between the 7 and the 5 may not be the same as that between 5 and 3.
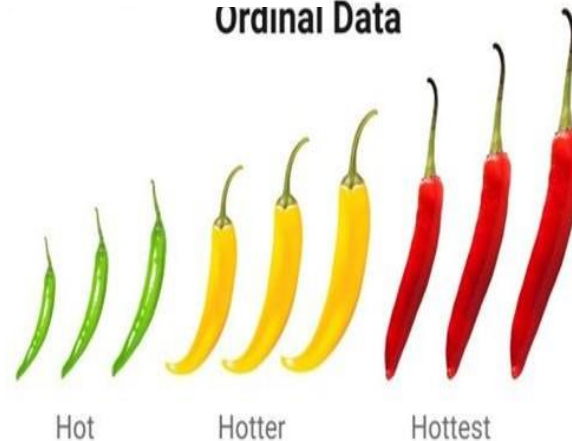 - The values simply express an order.
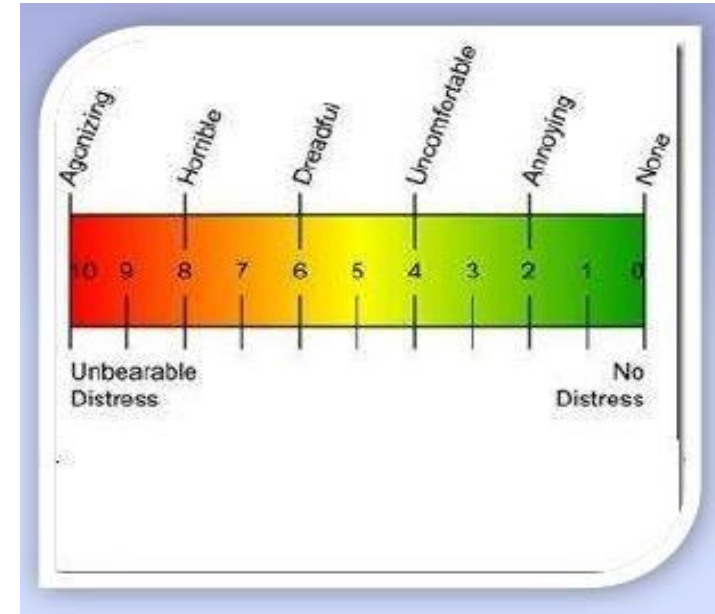
Source: Questionpro, slideshare.net

Slide Courtesy:Dr.Uma

Qualitative data: Ordinal data examples

Chart of Ordinal data

Slide Courtesy:Dr.Uma

Quantitative data: Discrete data

- A set of data is said to be discrete if the values belonging to the set are discrete and separate.
- The data values cannot be divided into smaller parts. For Example, the number of students in a class is an example of discrete data since we can count whole individuals but can't count in fractions like 2.5, 3.75, kids.
- It has a limited number of possible values e.g. days of the month.
- Discrete data can take only certain values by a finite 'jumps' i.e. It 'jumps' from one value to another but does not take any intermediate value between them.
- Examples of discrete data:
  - The number of students in a class.
  - The number of workers in a company.
  - The number of test questions you answered correctly.
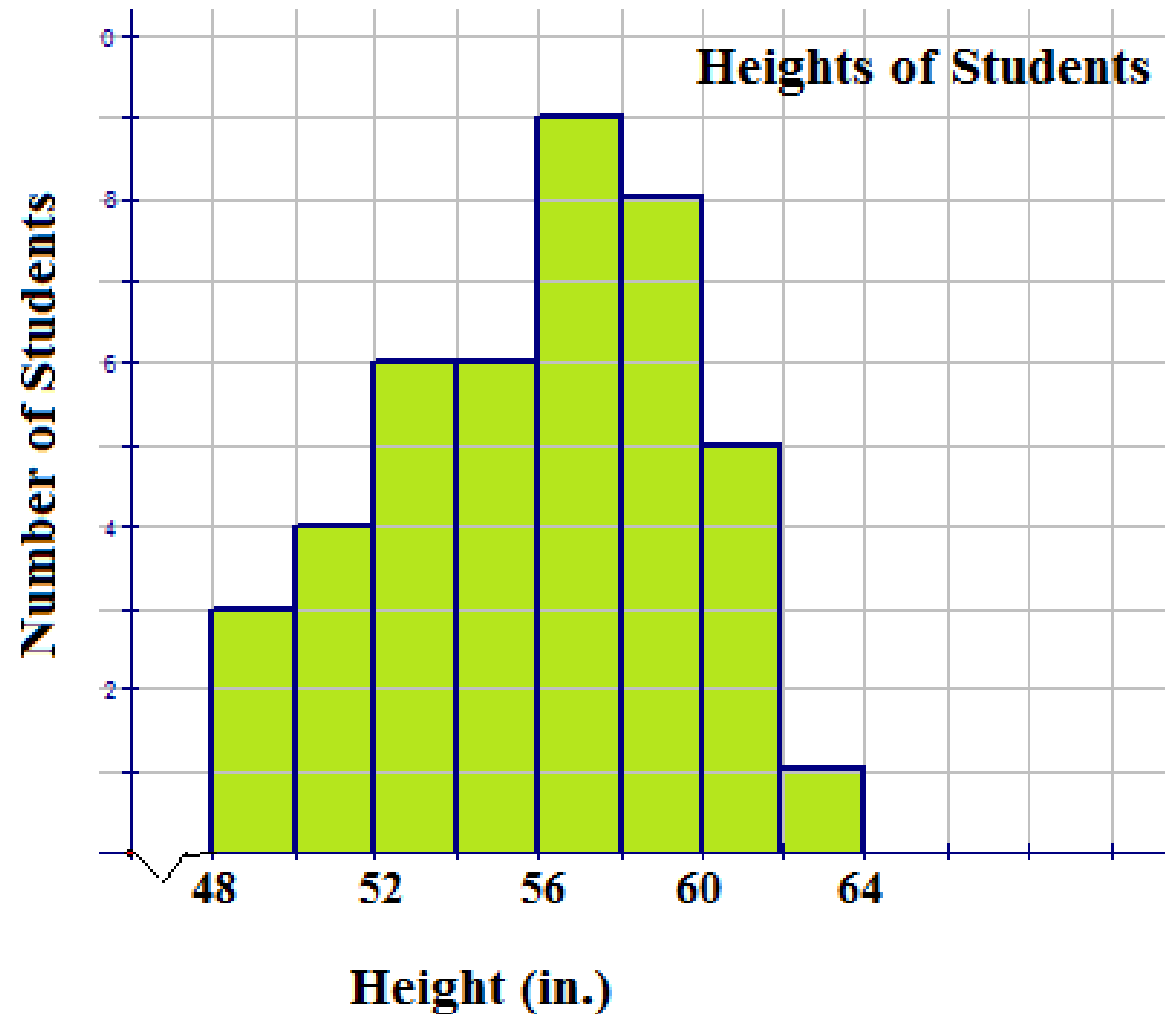
Quantitative data: Discrete data

- Bar charts can be used to display discrete numerical data.
- For example, the bar chart below shows the number of CDs bought by a group of children in a given month.

Number of CDs bought in a month

Source: slideplayer.com
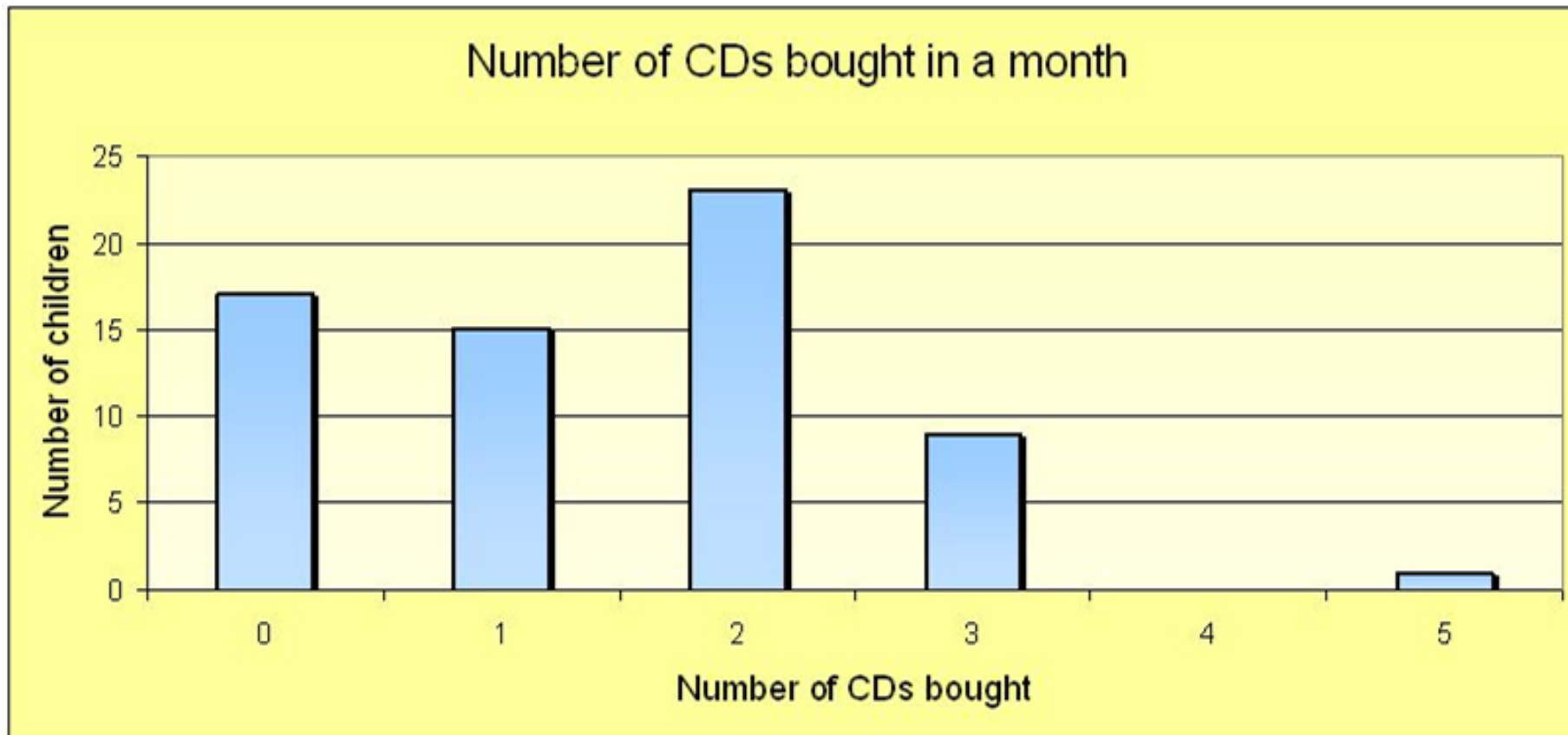
Quantitative data: Continuous data

- A set of data is said to be continuous if the values belonging to the set can take on any value within a finite or infinite interval.
- It represents the information that could be meaningfully divided into its finer levels.
- It can be measured on a scale or continuum and can have almost any numeric value. For Example, we can measure our height at very precise scales in different units such as meters, centimeters, millimeters, etc.
- Examples of continuous data:
  - The amount of time required to complete a project.
  - The height of children.
  - The speed of cars.

Quantitative data: Continuous data

Quantitative data: Interval data

- It is a data type which is measured along a scale, in which each point is placed at equal distance from one another.
- These data types are measurable and ordered with the nearest items but have no meaningful zero.
- Interval scales not only educate us about the order of the items but in addition, give information about the value between every item.
- There are some descriptive statistics that we can calculate for interval data such as :
  - Central measures of tendency (mean, median, mode)
  - Range (minimum, maximum)
  - Spread (percentiles, interquartile range, and standard deviation).
- Examples: Temperature (°C or F, but not Kelvin),

  Dates (1055, 1297, 1976 etc), Time Gap on a 12-hour clock (6 am, 6pm)

Quantitative data: Interval data

Slide Courtesy:Dr.Uma
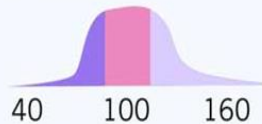
Quantitative data: Ratio data

- Ratio data **classifies and ranks data**, and uses measured, continuous intervals, just like interval data.
- However, unlike interval data, ratio data **has a true zero**.
- This basically means that zero is an absolute, below which there are no meaningful values.
- Speed, age, or weight are all excellent examples since none can have a negative value (you cannot be -10 years old or weigh -160 pounds)
- These data are also in the ordered units that have the same difference.
- Ratio data allow for forming quotients inaddition to setting up inequalities and forming differences.
- All mathematical operations(manipulations   with real numbers) are possible on ratio data.
- It is the most precise data and allow for application of all statistical techniques.

Quantitative data: Ratio data

Slide Courtesy:Dr.Uma

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Quantitative data: Ratio data

● Example:

Roll No.	Name	Gender	Rank	Height	Weight In Kgs
1	Amar	M	9	4' 8"	51
2	Asha	F	1	3' 10"	39
3	Bhaskar	M	5	4' 5"	48
4	Chandru	M	3	4' 3"	41

Slide Courtesy:Dr.Uma

Identify the type of data

➜ Number of cartons of milk manufactured each day.
  Quantitative data, Discrete data

➜ Temperatures of airplane interiors at a given airport in Celsius.
  Quantitative data, Continuous data, Interval data.

➜ College major of each student in a class.
  Qualitative data, Nominal data

➜ Method of payment
  Qualitative data, Nominal data

➜ Incomes of college students on work study programs.
  Quantitative data, Discrete data

➜ Weights of newborn calves.
  Quantitative data, Continuous data, Ratio data.

Identify the type of data

➜ Gender of each employee at a company.
  Qualitative data, Nominal data
➜ Number of tomatoes on each plant in a field.
  Quantitative data, Discrete data
➜ Number of defective items in a lot.
  Quantitative data, Discrete data
➜ Salaries of CEOs of oil companies.
  Quantitative data, Discrete data

Attribute or Variable

- **Attribute**(or variable, feature, dimension) is a data field, representing a characteristic or feature of a data object.
- It is a property of a data object which is measured for each observation or record.
- It can vary from one observation to another.
- Example : name, age, Student-ID, address, marks, gender etc.
- There are different types of attributes or variables such as:
  - Nominal
  - Ordinal
  - Discrete
  - Continuous
  - Binary
  - Interval
  - Ratio

Quantitative Variable

- A variable that can be <u>measured numerically</u> is called a quantitative variable.
- The <u>data collected</u> on a quantitative variable are called <u>quantitative data</u>.
- Thus, a quantitative variable represents a measure and is numeric.
  Its values can be recorded on a numeric scale.
- Example: a country's population, a book's price, height, weight, number of items sold to a shopper, time in 100 yard dash etc.

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS
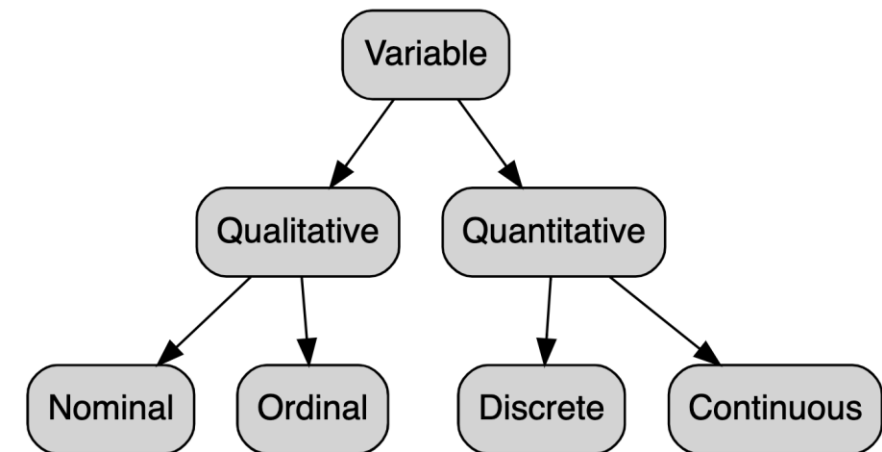
Quantitative Variable: Discrete

- A variable whose <u>values are countable</u> is called a discrete variable.
- In other words, a discrete variable can assume only certain values with no intermediate values.
- Its number of values is finite or limited.
- Example: number of oranges in a bag, number of students in a classroom, shoe size etc.

Days	Number of students present
Monday	24
Tuesday	26
Wednesday	28
Thursday	30
Friday	29
Saturday	22

Sources: previews.123rf.com, teachoo.com

Quantitative Variable: Continuous

- A variable that can assume any numerical value over a certain interval or intervals is called a continuous variable.
- It represents the numerical data as accurately as possible.
- It can take unlimited number of values between the lowest and highest points of measurements.
- Continuous attributes are typically represented as floating-point variables.
- Example: height, weight, temperature

Continuous Variables	Quantile				
	Zeroth	First	Second	Third	Fourth
Apparent Temperature	28.11	67.25	75.09	82.495	107.23
Cloud Cover	0.0	0.03	0.1	0.22	1.0
Dew Point	16.55	58.16	66.0	73.08	82.14
Relative Humidity	0.16	0.62	0.79	0.89	1.0
Temperature	35.61	67.25	75.09	80.37	94.99
Wind Speed	0.0	3.87	5.66	7.82	26.55

Temperature of Heated Water

Dependent: Water temperature (oC)

Independent: Heating time (min)
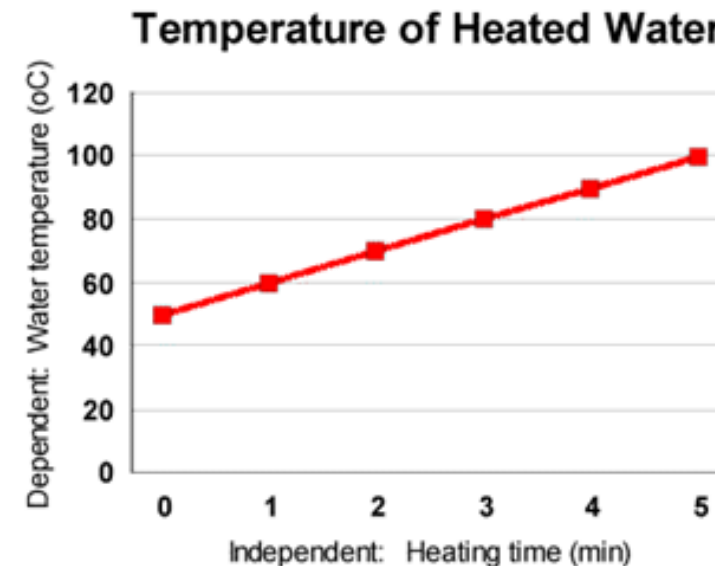
Qualitative Variable

- A variable that <u>can not assume a numerical value</u> but can be <u>classified into two or more non-numeric categories</u> is called a qualitative
- It is also known as a **categorical variable**.
- The **data collected** on such a variable are called **qualitative data.**
- The values of this variable are not numeric as they do not result from counting or measuring.
- Thus, arithmetic operations can't be applied on these variables.
- Example: hair colour, favorite books, religion, political party in power, profession, name etc.

Qualitative Variable: Nominal

- A nominal variable is a type of variable that is used to name, label or categorize particular attributes that are being measured.
- Nominal means "relating to names" .
- The utilities of a nominal attribute are sign or title of objects .
- Each value represents some kind of category, code or state.
- It takes qualitative values representing different categories.
- There is no intrinsic ordering of these categories.
- Example: Gender- male, female;

       Marital status- married, unmarried;

       Skin colour- dark, white, brown;

Qualitative Variable: Ordinal

- An ordinal variable is a type of categorical variable that **takes values with an order or rank**.
- Qualitative variables have natural, ordered categories and the <u>distances between the categories is not known</u>.
- Example: Likert scale, or the survey question "Is your general health poor, reasonable, good, or excellent?" may have those answers coded respectively as 1, 2, 3, and 4.
- Educational status- undergraduate, postgraduate, matriculate

Likert Scale

5 Point Likert Scale

Numbering	1	2	3	4	5
	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I like Stock Market	○	○	○	○	○
I like Stocks	○	○	○	○	○
I Like Money	○	○	○	○	○
I Like Return	○	○	○	○	○
Total	4	8	12	16	20

Source: cdn.wallstreetmojo.com

Quantitative Variable: Interval

- It is interval measured on a scale of equal-sized units
- Values of inter variables have order
- It has no true zero-point.
- Interval variables allow to rank the items measured in order.
- They also allow to quantify and compare the magnitude of differences between them.
- Example: temperature in ˚ C or ˚ F, calendar dates etc.

Quantitative Variable: Ratio

- Ratio variables represent the highest level of measurement.
- A ratio variable has an inherent or true zero-point.
- The numerical relationship between the values of a ratio variable is meaningful.
- We can speak of values as being an order of magnitude larger than the unit of measurement (10 K is twice as high as 5 K).
- Example: temperature in Kelvin, length, counts, monetary quantities etc.

Properties of Attributes

The type of an attribute depends on which of the following properties it possesses:

- Distinctness: =, ≠
- Order: <, >
- Addition: +, -
- Multiplication: *, /

➔ **Nominal attribute:** distinctness
➔ **Ordinal attribute:** distinctness & order
➔ **Interval attribute:** distinctness, order & addition
➔ **Ratio attribute:** all 4 properties

Properties of Attributes

THE FOUR LEVELS OF MEASUREMENT:

	Nominal	Ordinal	Interval	Ratio
Categorizes and labels variables	✓	✓	✓	✓
Ranks categories in order		✓	✓	✓
Has known, equal intervals			✓	✓
Has a true or meaningful zero				✓

Source: dpbnri2zg3lc2.cloudfront.net

Examples

In the table below identify which columns represent qualitative variables and which columns represent quantitative variables.

Name	Height	River	State	Completed
Oroville	754	Feather	CA	1968
Hoover	725	Colorado	NV	1936
Dworshak	718	N Fork Clearwater	ID	1873
Glen Canyon	708	Colorado	AZ	1966
New Bullards Bar	636	North Yuba	CA	1970
New Melones	626	Stanislaus	CA	1979
Swift	610	Lewis	WA	1958
Mossyrock	607	Cowlitz	WA	1968
Shasta	600	Sacramento	CA	1945
Hungry Horse	564	S Fork Flathead	MT	1953
Grand Coulee	551	Columbia	WA	1942
Ross	541	Skagit	WA	1949

Answer: Qualitative variables: Name, River, State
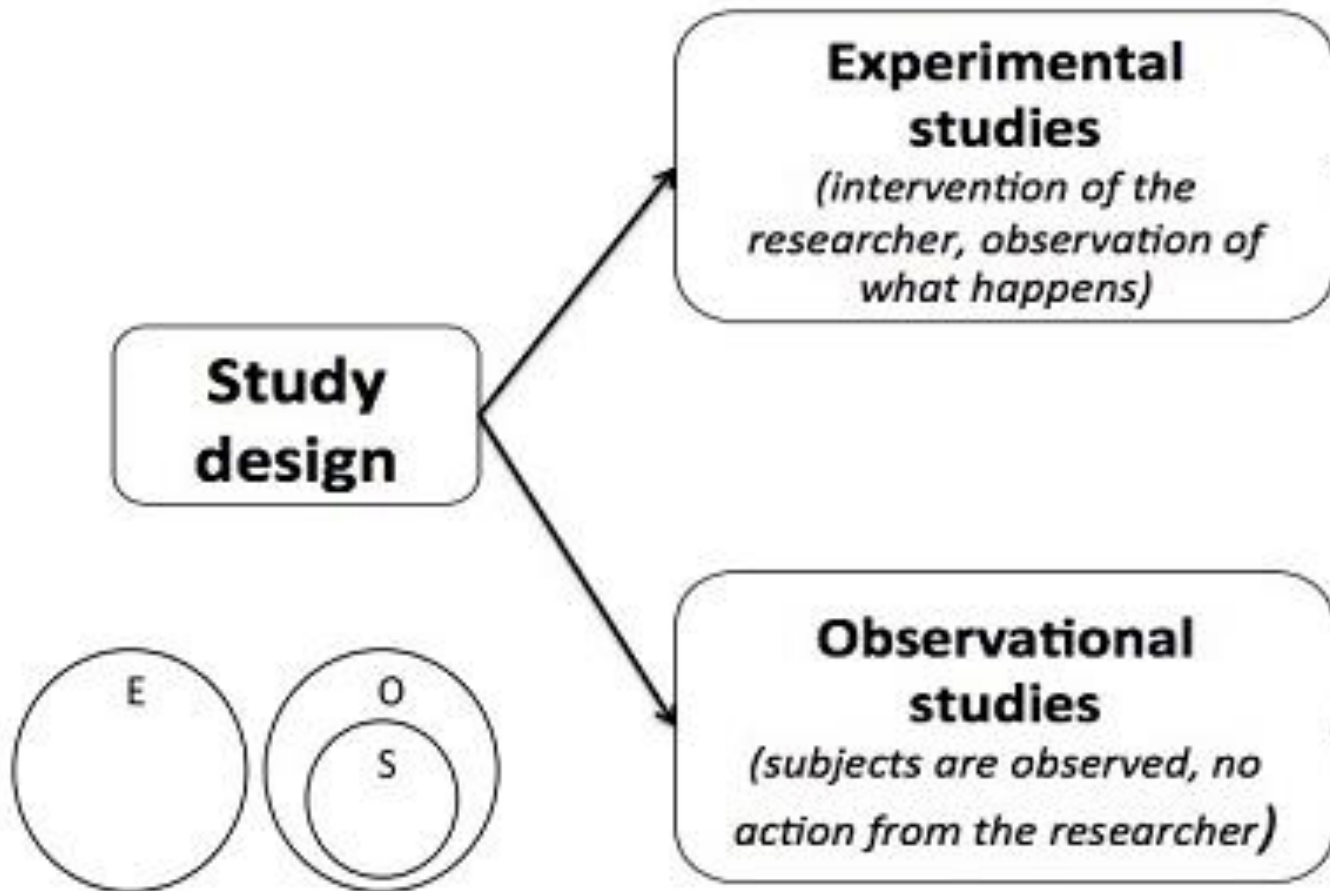
Quantitative variable: height, Completed

Types of studies

We do studies to gather information and draw conclusions. The type of conclusion we draw depends on the study method used:
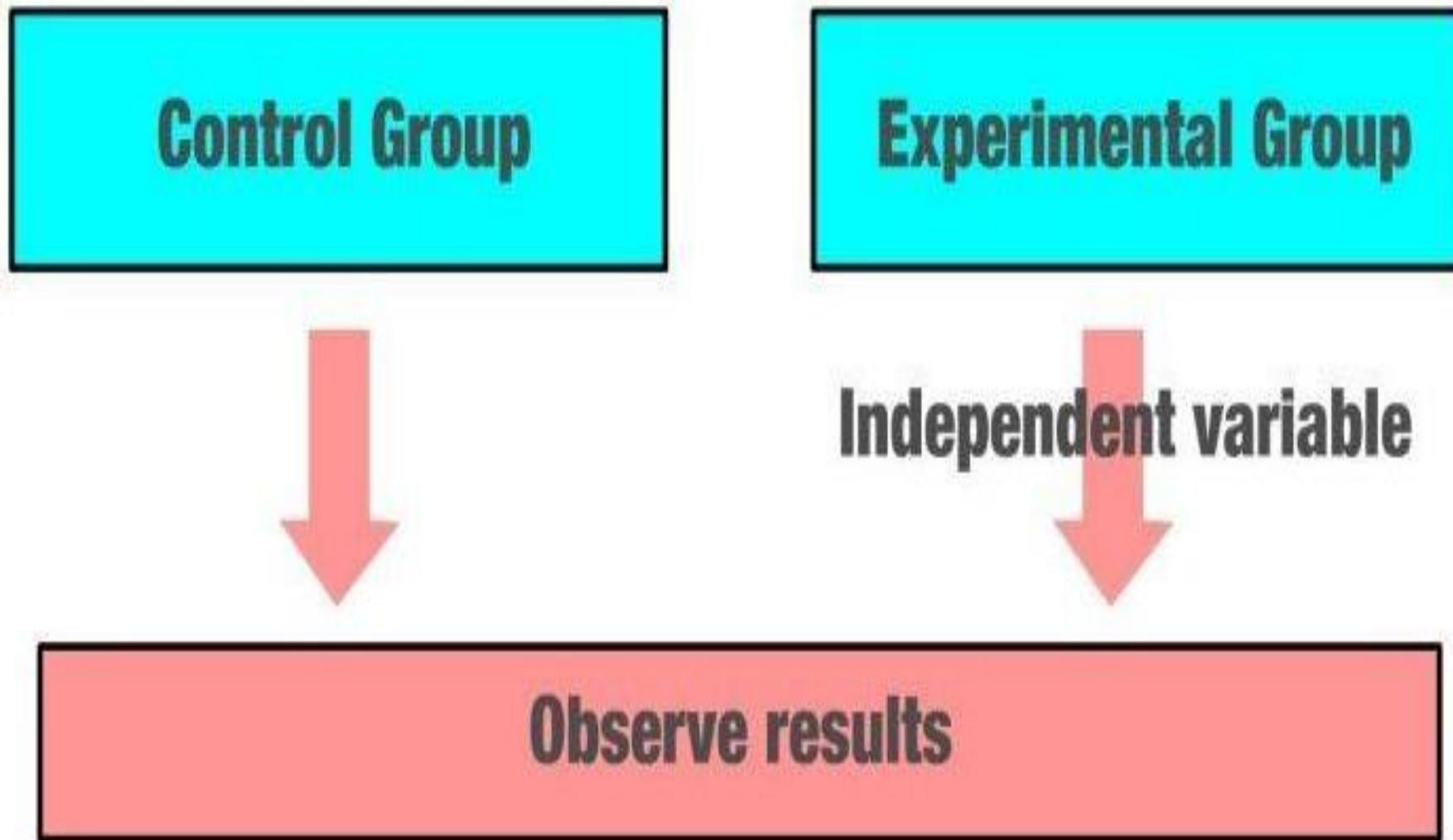
I. **Observational study:** In an observational study, we measure or survey members of a sample without trying to affect them.

II. **Controlled study:** In a controlled experiment, we assign people or things to groups and apply some treatment to one of the groups, while the other group does not receive the treatment.
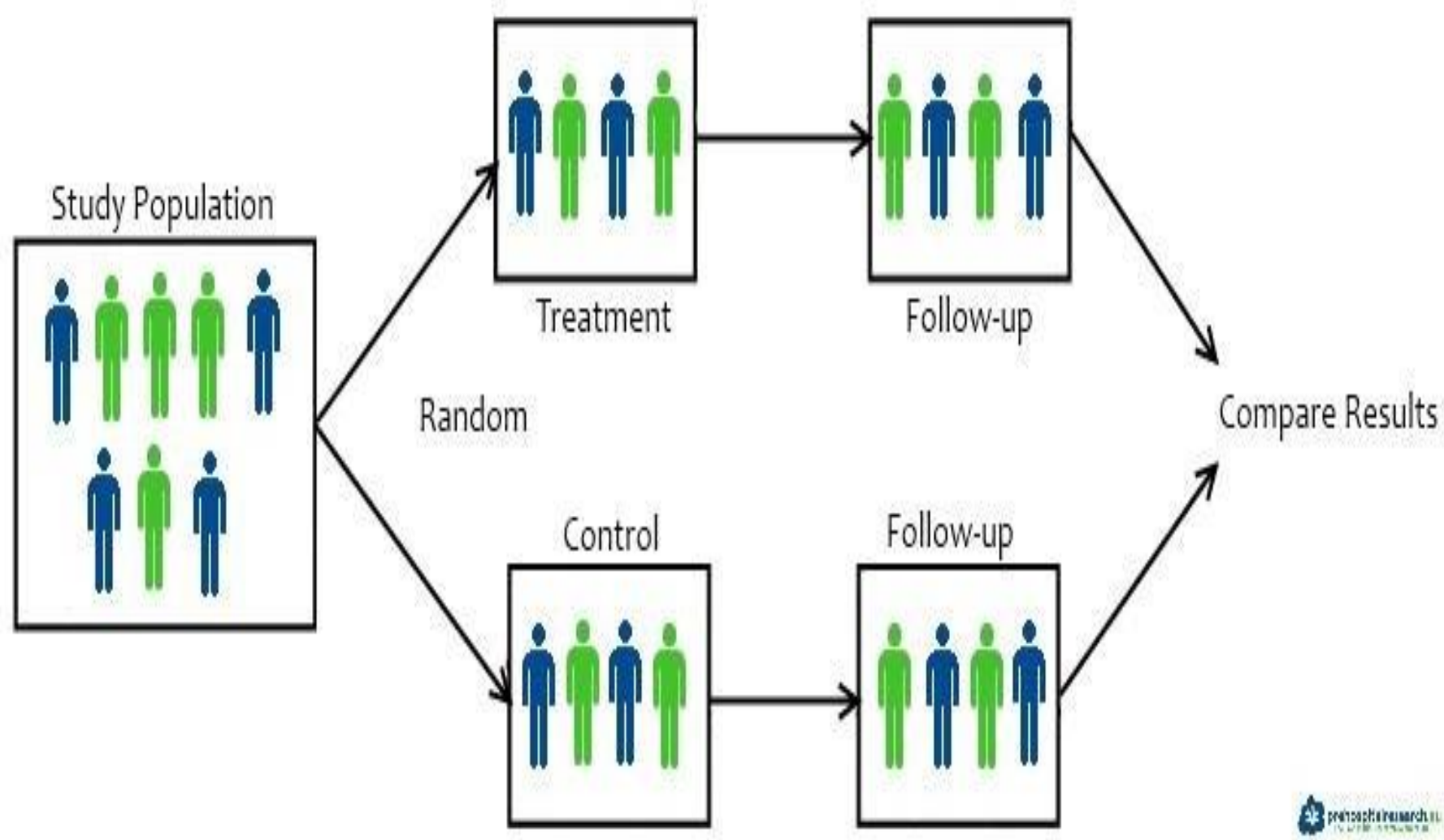
Types of studies

Types of Statistical Studies

Experimental studies
(intervention of the researcher, observation of what happens)

Study design

Observational studies
(subjects are observed, no action from the researcher)

E   O
     S

Source: statisticsguruonline.com

Controlled Study

Slide Courtesy:Dr.Uma

Controlled Study

Slide Courtesy:Dr.Uma

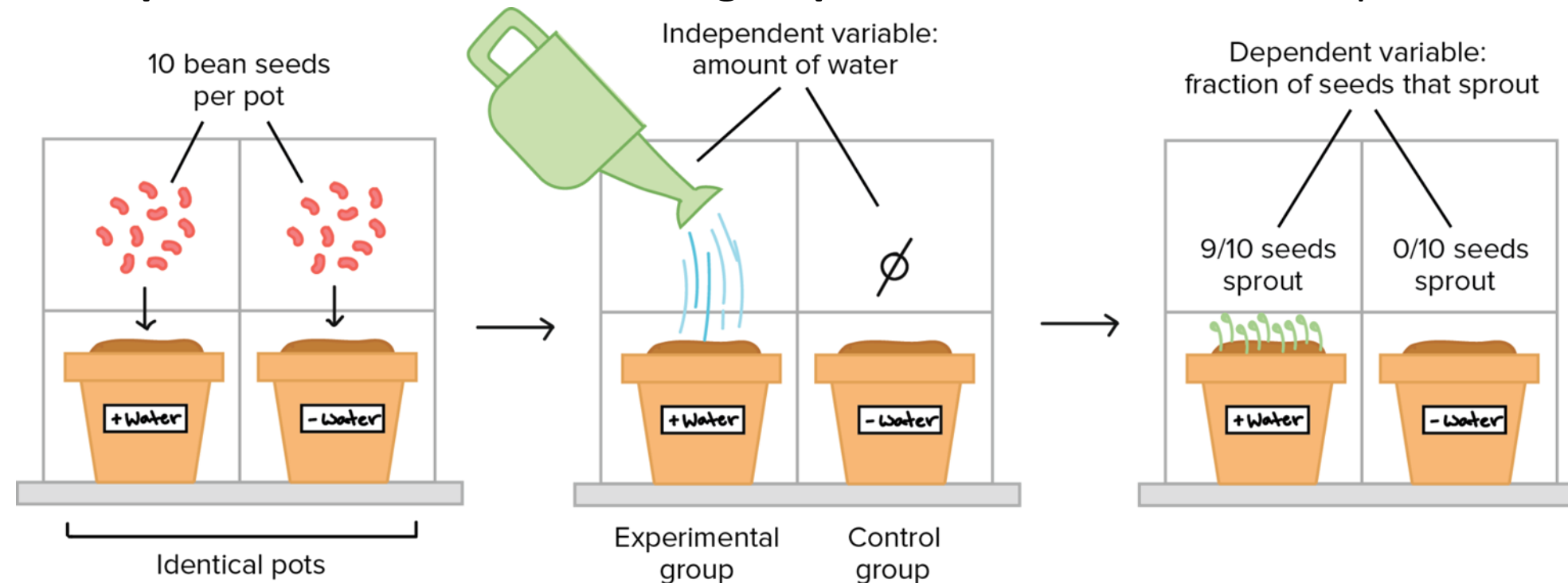## Controlled Experiment

- While all experiments have an experimental group, not all experiments require a control group.
- Controls are extremely useful where the experimental conditions are complex and difficult to isolate.
- **Experiments that use control groups** are called controlled experiments.

Source: cdn.kastatic.org

Observational study vs. Experimental study

Observational Study	Experimental Study
Observe only, no "treatment" assigned.	"Treatment" assigned.
Generally a control group is not needed.	Uses control group for comparison.
Reports an association.	Report a cause and effect.
May (or not) use random sample sets.	Randomization of sample group.
May (or not) generalize to population.	Generalize to population.

Slide Courtesy:Dr.Uma

Identify the types of study

Q1.A study took random sample of adults and asked them about their bedtime habits. The data showed that people who drank a cup of tea before bedtime were more likely to go to sleep earlier than those who didn't drink tea.

Answer : Observation Study

Q2.A study took a group of adults and randomly divided them into two groups. One group was told to drink tea every night for a week, while the other group was told not to drink tea that week. Researchers then compared when each group fell asleep.

Answer : Experimental Study

Identify the types of study

Q3.A study randomly assigned volunteers to one of two groups:
   One group was directed to use social media sites as they usually do.
   One group was blocked from social media sites.

Answer : Experimental Study

Q4.A study took a random sample of people and examined their social media habits. Each person was classified as either a light, moderate, or heavy social media user. The researchers looked at which groups tended to be happier.

Answer : Observation Study

References

https://www.analyticsvidhya.com/blog/2021/06/complete-guide-to-data-types-in-statistics-for-data-science/

Text Book:

Statistics for Engineers and Scientists, William Navidi.

THANK YOU

Dr.Mamatha H R

Professor, Department of Computer Science

mamathahr@pes.edu