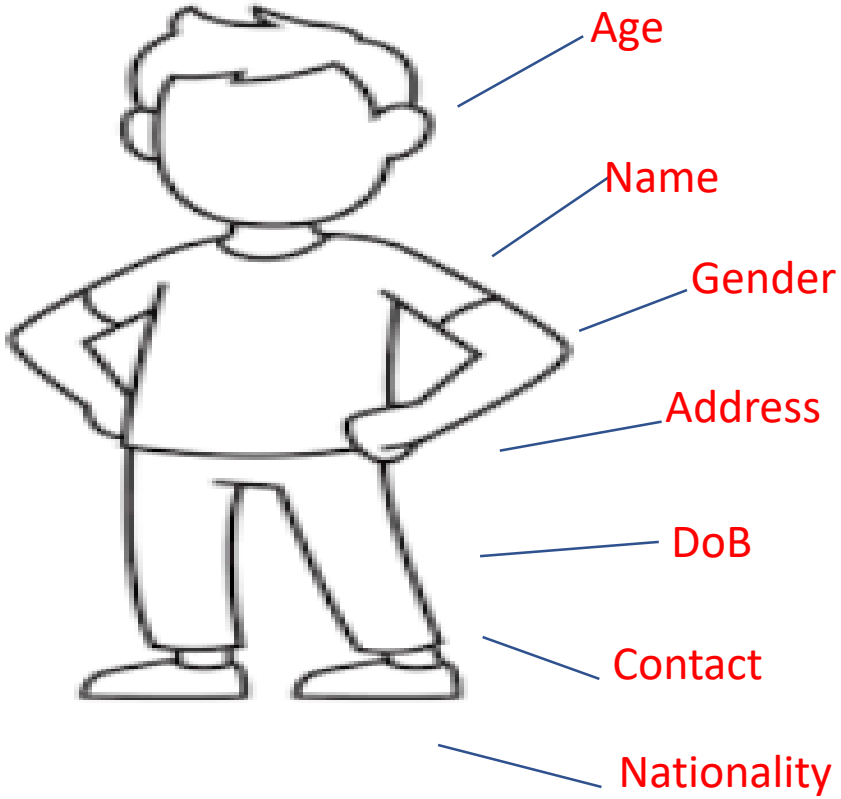
DATA TYPES

Example

Person



Name	Age	Gender	DoB	Address	Nationality	Contact

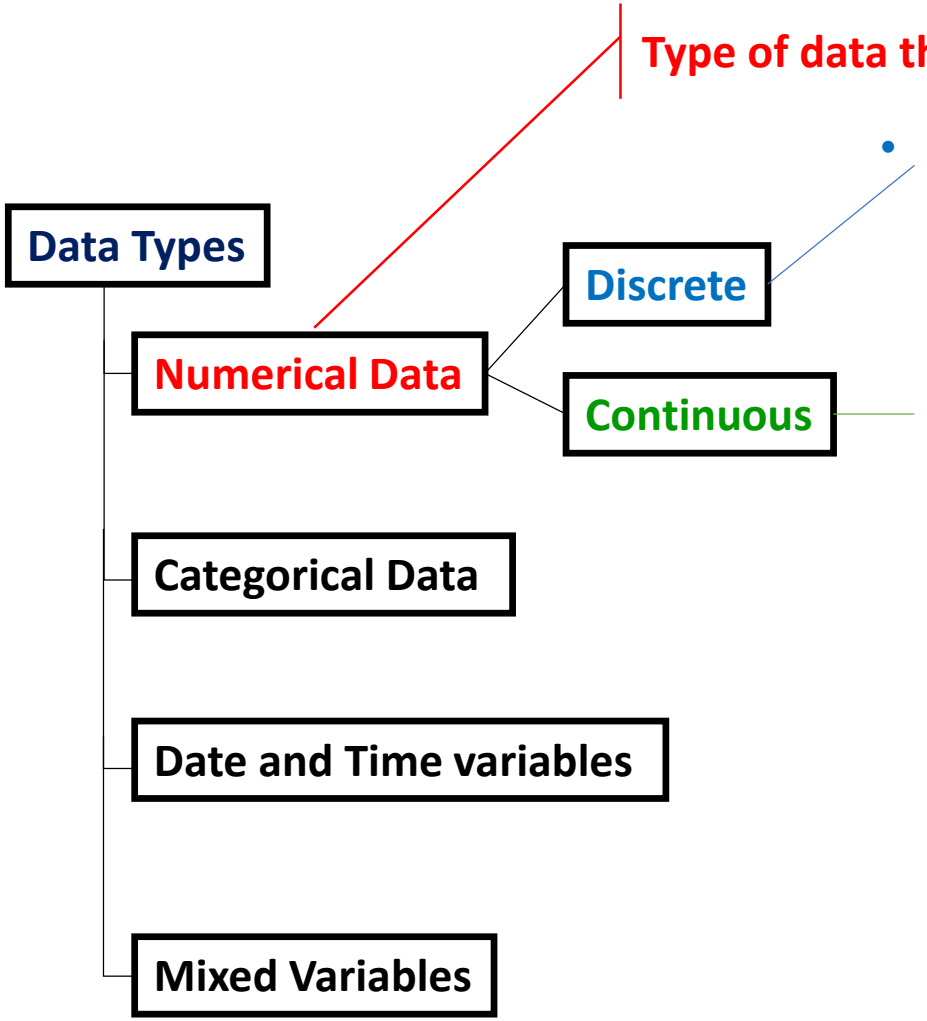
INTRODUCTION

- A **dataset** can contain variables of **different types** depending upon **the data**, **the request**, and **the storage**.
- It is important to know the different types of data that a variable can store since different techniques are required to handle data of various types.

What is a Variable?

- "A variable is an **entity** which stores a value that corresponds to a characteristic, quantity, or a number that can be counted or measured."
  - Ex: Name, Age, Address, Marks, etc.

Data Types

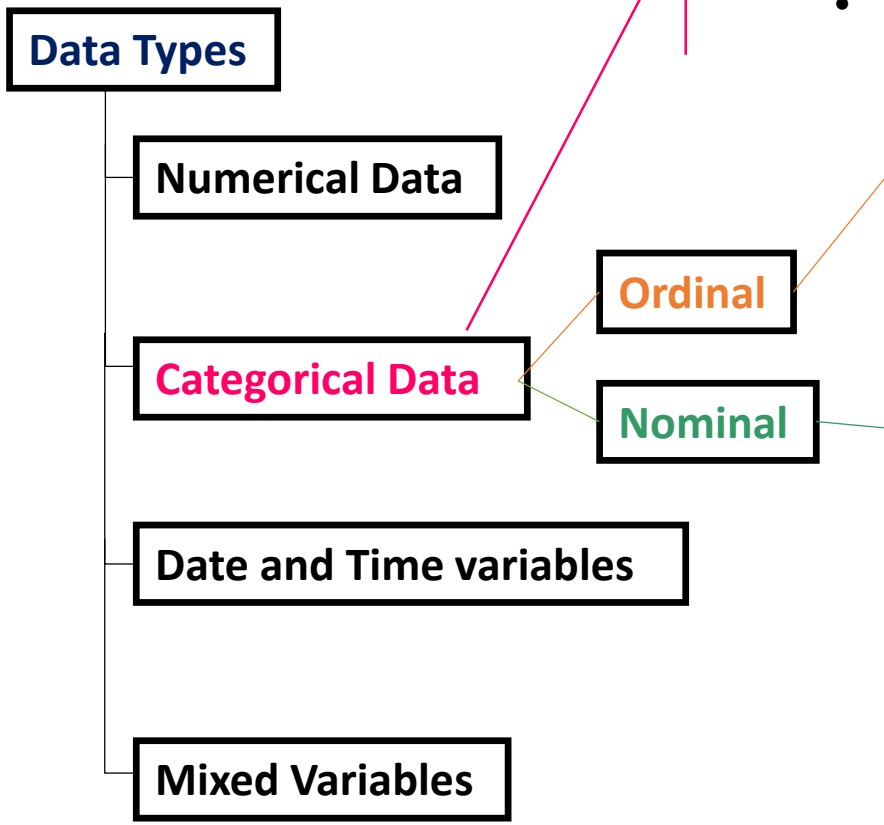


Type of data that consists of numbers

- **Stores a whole number**
  - Ex: Number of children of a person can be 1,2,3,4
    Number of laptops, Number of siblings, etc.

- **Holds a continuous values**
  - Ex: The amount in a person's bank account ($1500.55)
    height of a person (5.6"), weight (55.5kg), price of a car, etc.

Data Types

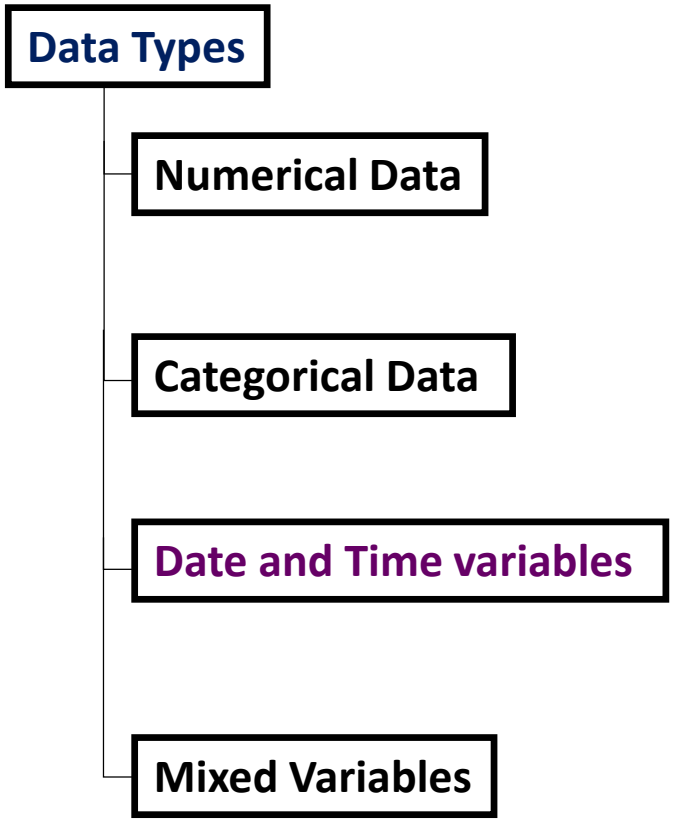


- Categorical data is a type of data where values are selected from a predefined set.
  - Ex: Gender, Nationality

- **Ordinal data is a type of categorical data where values are meaningfully ordered and have some relationship between them.**
  - Ex: Height of a person can be Short, Medium, Tall
    Marks can be A, B, C, etc

- **Nominal data is a type of data where values have no meaningful ordering.**
  - Ex: Nationality of a person.

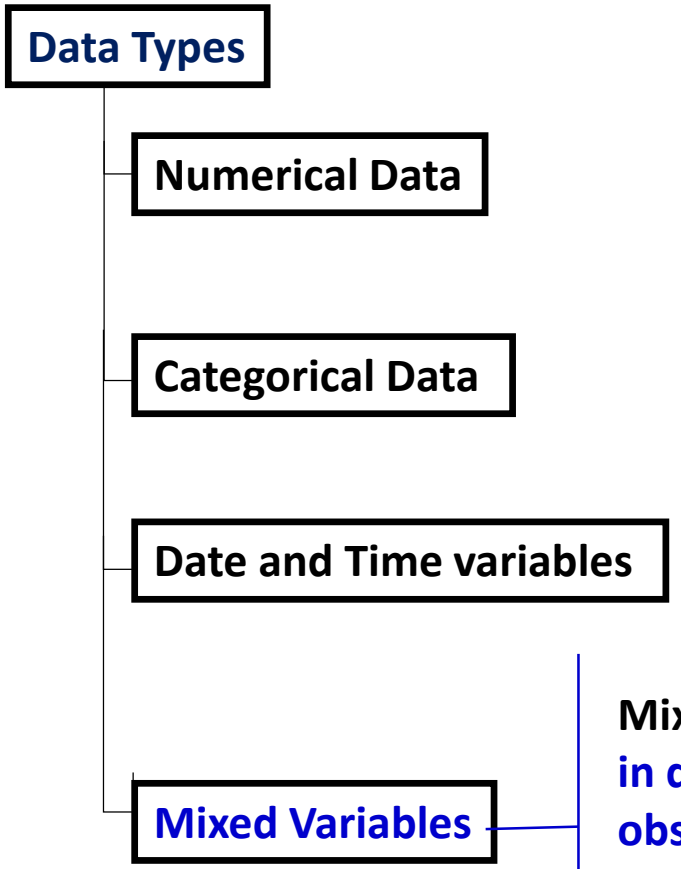
Data Types



The date and time variable can consist of date and time, date only, and time only.
Ex:

- **The date of birth of a person**
  Ex: 10-12-1995, December 10, 1995, 10th December 1995.
- **The date and time of a football match**
  Ex: 11:30 am, December 10, 2021.
- **The time at which the office closes**
  Ex: 5:00 pm.

Data Types



Mixed data type variables are those variables that have either numerical and categorical data in different observations exclusively or contain numerical and categorical data in single observations of a data column.

Ex:

- Experience (1,2,3, 5.5, 6.8, Fresher, etc.)
- The quantity of the product bought by the customer (5 Kgs, 3Litres, 2 dozens, etc.)

Interval Data

- Interval data builds on top of ordinal data. In addition to ordering the values, it also specifies that the intervals between subsequent values are the same.
  - Ex: the temperature measured in degrees Celsius: the difference between 1 degree and 5 degrees is the same as between 20 and 24: it's 4 degrees.

- Note that this was not the case for ordinal data: we cannot say that the difference between graduating from a high school versus from an elementary school only is the same as the one between a university and a high school.

Ratio Data

- Ratio data builds on top of interval data.
- The difference is that ratio type variables have a meaningful zero value.
  - Ex: price, length, weight, amount of something, or temperature measured in Kelvin.
- The meaningful zero allows us to calculate ratios between two data points: we can say that 4 apples are twice as much as 2, or that $5 is half as expensive as $10. This was not the case for interval data: in the case of temperature measured in degrees Celsius, we cannot say 10 degrees is twice as warm as 5 degrees. Ratios make no sense for scales without a meaningful zero.

Summary

Data Measurement Type	Property	Central Tendency Measure	Allowed math	Example
NOMINAL	grouping	mode	-	color
ORDINAL	order	mode, median	monotone transformations	education level
INTERVAL	equal intervals	mode, median, mean	linear transformations	temperature in °C
RATIO	meaningful zero	mode, median, mean	scaling transformations	price