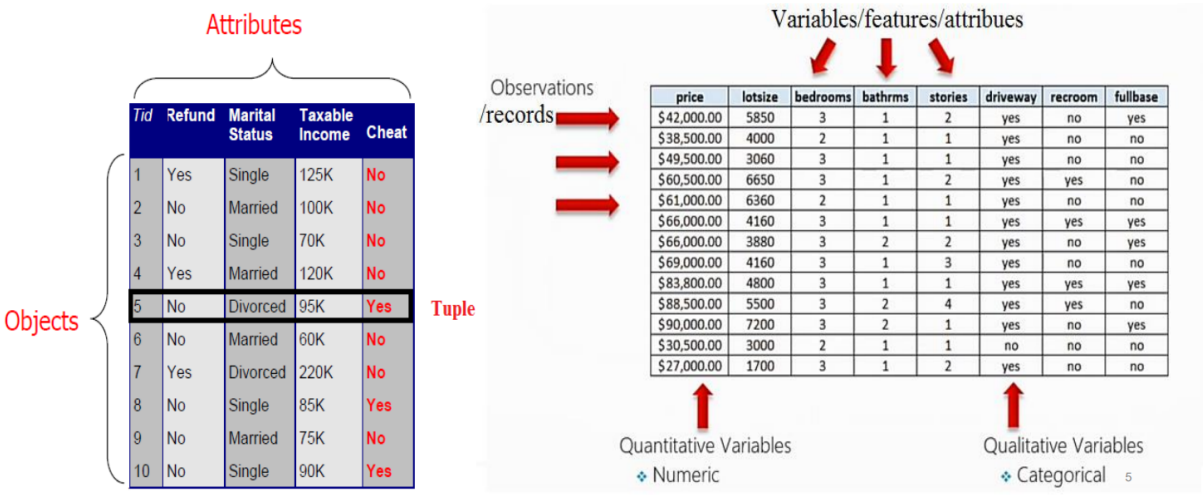
Data Types and their Characteristics

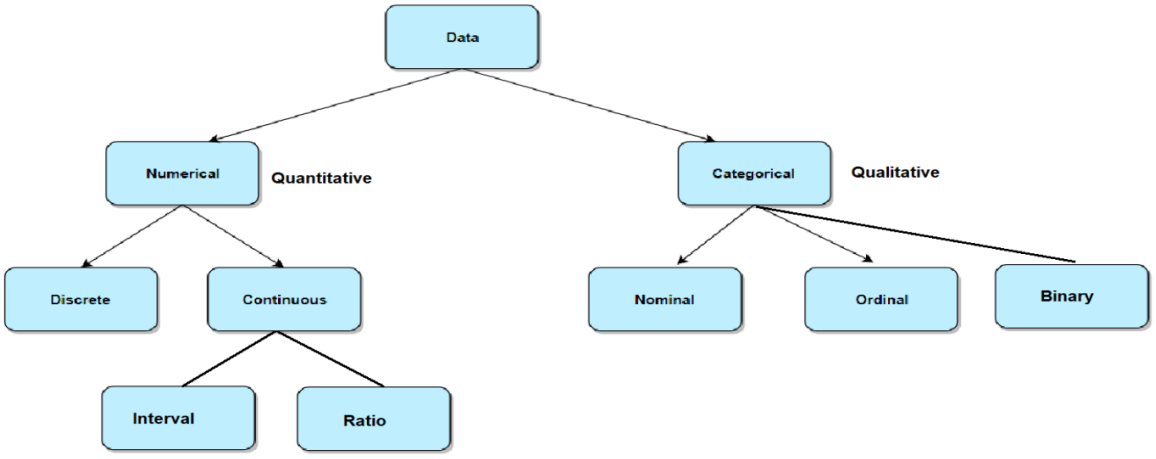
Notes by Mannan Ul Haq (BDS-3C)

Data

In Data Science, data means collection of **Objects** and **Attributes**.



Data Types



Categorical Data:

Categorical data represents categories or characteristics like gender, language, or movie genre. It's also called qualitative data. You can use numbers for them, but those numbers have no real math value (like 0/1 for male/female).

Types:

1. **Nominal Data**:
   - No order.
   - Examples: gender, language, eye color.
   - Analyze with frequencies, pie charts, etc.
2. **Ordinal Data**:
   - Has order.
   - Examples: happiness level, education level, movie ratings.
   - Summarize with median, mean, visualize with bar charts.
3. **Binary Data**:
   - Just two values: yes or no.
   - Represented as "True" and "False" or 1 and 0.

Numerical Data

Numerical data is expressed as numbers, allowing quantification. It represents values like integers or real numbers. It is also called quantitative data. Examples include a person's height, product prices, IQ scores, the number of lessons in a course, etc.

Types:

1. Discrete Data

- Has a finite or countably infinite set of values.
- Values are distinct and separate.
- Examples: zip codes, words in a document collection, number of coin toss heads, students in a classroom, cars in a showroom.

- Often represented as integer variables.
- Analyzed using mean, median, quartiles, box plots, and histograms.

2. Continuous Data

- Cannot be counted but can be measured.
- Represents measurements.
- Examples: market share price, height/weight of a person, amount of rainfall, car speed, Wi-Fi frequency.
- Can be divided into meaningful parts.
- Has real numbers as attribute values.

Types of Continuous Data:

a. Interval Data

- Categorized, ranked, and evenly spaced.
- Values have order and can be positive, zero, or negative.
- Allows comparison and quantification of differences.
- Examples: temperatures in Celsius or Fahrenheit, calendar dates.

b. Ratio Data

- Numeric attribute with an inherent zero-point.
- Values can be multiples or ratios of one another.
- Ordered values with computed differences.
- Examples: Kelvin temperature scale, years of experience, number of words.