

CS500-Data Science Tools and Technique

Introduction to Data Types

Bahar Ali
PhD Scholar,
National University Of Computer and Emerging Sciences,
Peshawar.

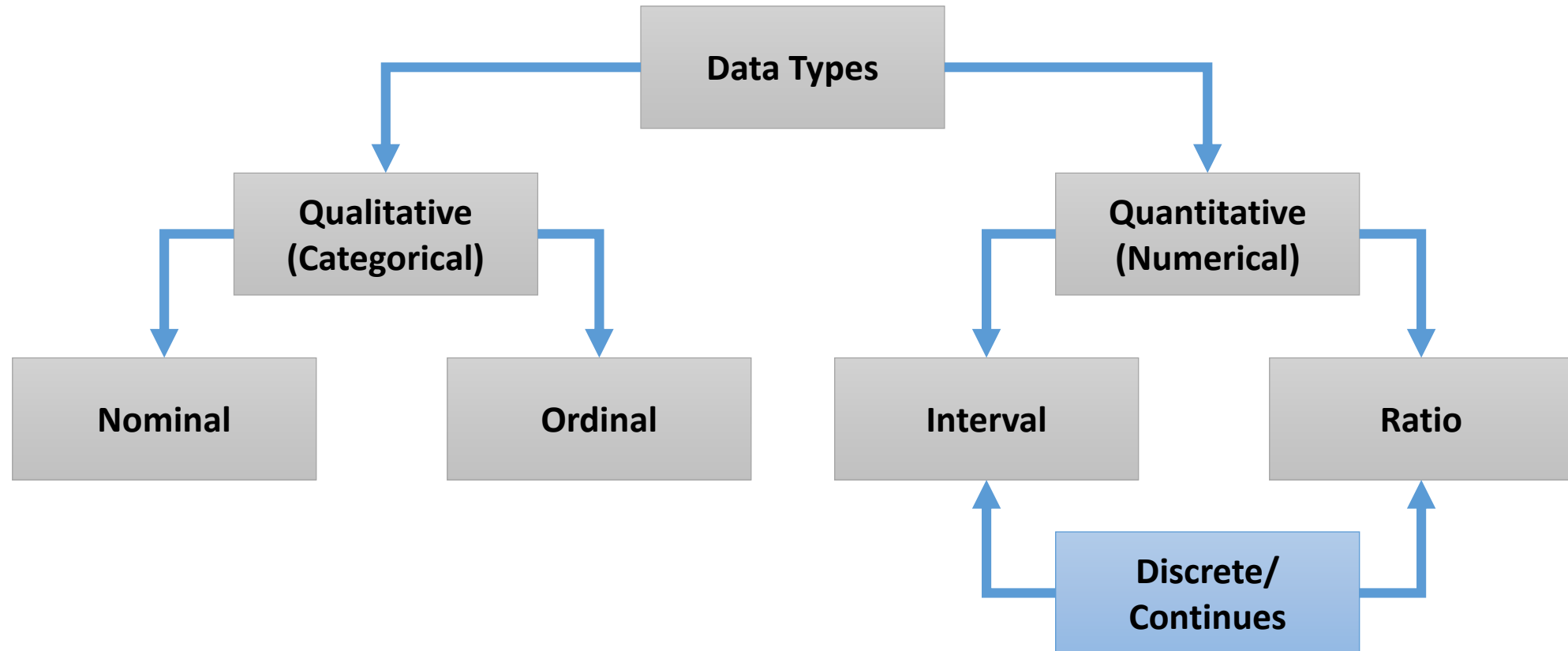
What is Data?

- Data are characteristics or information, that are collected through observation. In a more technical sense, data are a set of values of qualitative or quantitative variables about an object.
- Raw data ("unprocessed data") is a collection of numbers or characters before it has been "cleaned" and corrected.

Data Types

- Data Types are an important concept of statistics
- To do proper exploratory data analysis (EDA), it is very important;
 - ❑ To know about data types, since you can use certain statistical measurements for specific data types.
 - ❑ To correctly apply statistical measurements to data and therefore help to choose the right analytics tools and visualization techniques.

Overview of Data Types



Quantitative vs Qualitative data

Qualitative/ Categorical

- Qualitative means you can not count it
- Categorical

Quantitative/ Numerical

- Quantitative means you can count it
- Numerical
- Two distinctions:
 - ❑ Discrete (whole numbers)
 - ❑ Continuous (real numbers)

Qualitative Data Types

Nominal data

- Nominal data are related to names
- Used to label variables without any quantitative value.
- Represent category
- Also called categorical attributes
- Examples:
 - ❑ Nationality
 - ❑ Brands Name

Qualitative Data Types

- **Binary** is a special type of nominal data, having only two states (0 and 1).
 - ❑ **Symmetric binary:** Both outcomes equally important
Example: Gender (Male/Female)
 - ❑ **Asymmetric binary:** Outcomes not equally important.
Example: Medical Test (positive vs. negative)
Convention: assign 1 to most important outcome (e.g., HIV positive)

Qualitative Data Types

Ordinal data

- Represent a meaningful order and ranking
- Categorical values where the order of the variables matters.
- Ordinal scales are often used for measures of satisfaction, happiness etc.
- Examples:
 - ☐ Strongly Agree/ Agree/ Neutral/ Disagree/ Strongly Disagree
 - ☐ High/ Moderate/ Low
 - ☐ Outstanding=3/ Excellent=2/ Good=1

Quantitative Data Types

Interval Data

- Interval data is concerned with both the order and difference between variables.
- Allows to compare and quantify the difference between values.
- There is no “True zero“, therefore, you can't calculate ratios.
- Example:
 - ❑ Temperatures in degrees Celsius. The difference between 20 degrees and 10 degrees is 10 degrees. The difference between 10 and 0 is also 10 degrees. Moreover, a true zero means none of that thing exists, but 0 degree definitely has a value.

Quantitative Data Types

Ratio data

- Ratio data tells us about the order of variables.
- It measures differences between values
- Have “True zero”, therefore you can calculate ratios.
- Examples:
 - ☐ Weight of body
 - ☐ Height of a person
 - ☐ Length of road

Data Types and Scale Measures

Operations	Nominal	Ordinal	Interval	Ratio
Equality	✓	✓	✓	✓
Order		✓	✓	✓
Add/ Subtract			✓	✓
Multiply/ Divide				✓
Mode	✓	✓	✓	✓
Median		✓	✓	✓
Arithmetic Mean			✓	✓
Geometric Mean				✓
Count/ Frequency	✓	✓	✓	✓
Quantify difference b/w values			✓	✓
Has “True Zero”				✓