UMBER NOREEN

Data Analytics
Week 2

UMBER NOREEN

- Assistant Professor at Faculty of Engineering. Minhaj University Lahore, Pakistan
- Chairperson School of Electrical Engineering
- Director Office of Research, Innovation, & Commercialization.
- Looking after a Business Incubation Center named Minhaj Centre for Startups.
- PhD in Computer Engineering from UBO, & University of Lille, France.
- Taught at University of Applied Sciences Alytus, Lithuania under Erasmus+ Program

WEEK 2

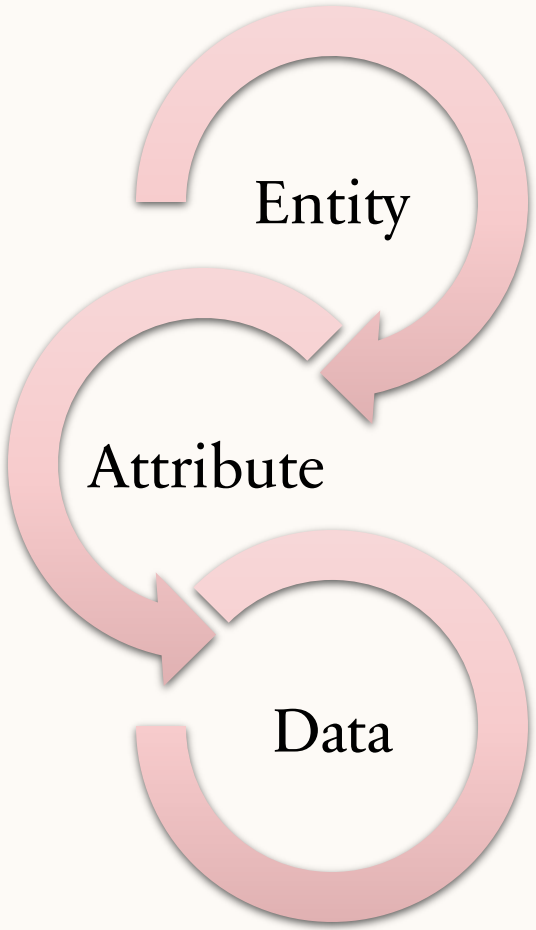
- Data Analytics Process Steps
- 5 V's of Big Data
- Levels of Data Analytics – NOIR
  - Nominal Scale/Measurement
  - Ordinal Scale/Measurement
  - Interval Scale/Measurement
  - Ration Scale/Measurement
- Data Visualization
  - Quantitative Data Graphs
  - Qualitative Data Graphs

BUSINESS ANALYTICS

- What is Happening & Why did Something Happened?
- Will it happen Again?
- What will happen if we make Changes to Some of the Inputs?
- What the Data is telling us that we are not able to see before?

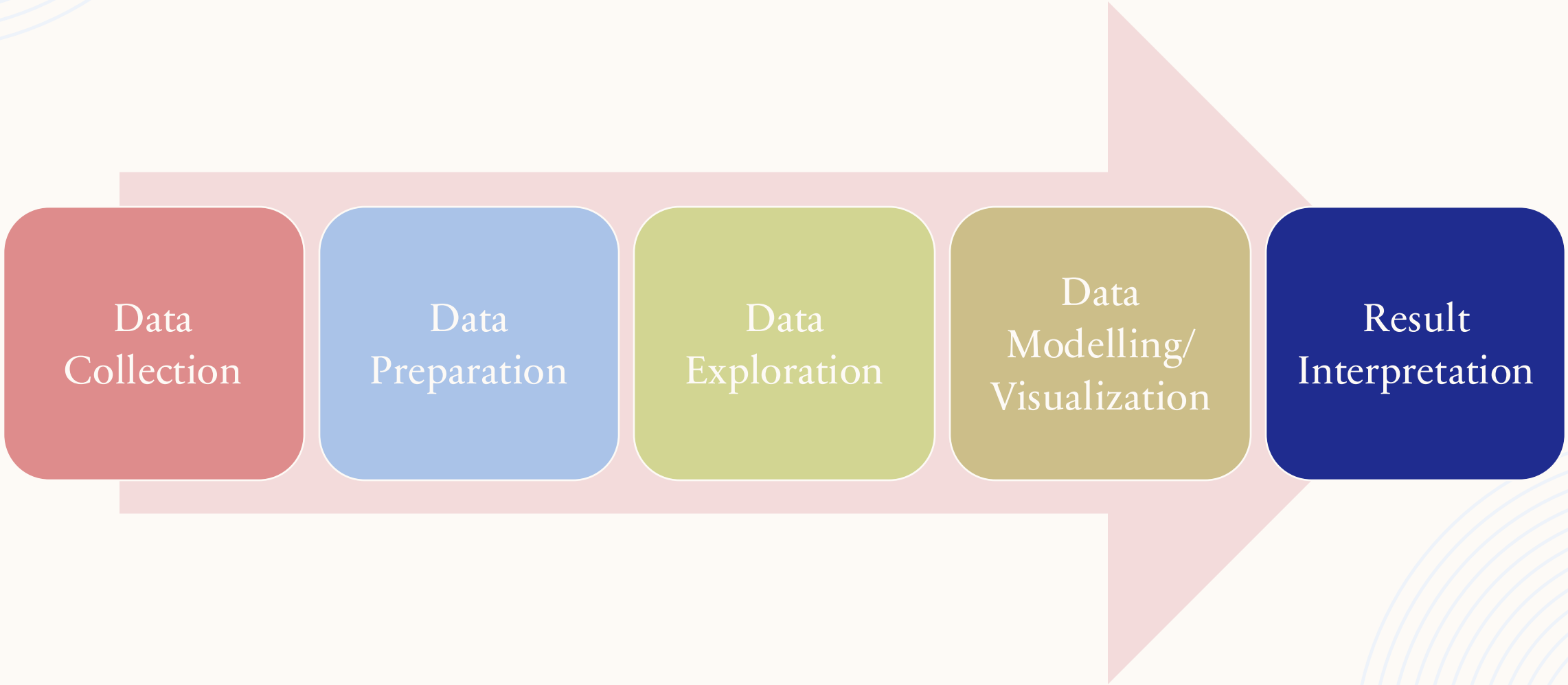
DATA IN DATA ANALYTICS

- **Entity:** A particular is called entity or object.
- **Attribute:** measureable or observable property of an entity.
- **Data:** Measurement of an attribute is called data
- E.g: Information of Student



Data Defines an Entity

DATA ANALYTICS PROCESS STEPS



5V'S OF BIGDATA

Volume

How much Data?
Example: Facebook collects terabytes of data daily.

Variety

Different Formats of data from various Sources
Structured, Unstructured, Semi-structured

Value

Usefulness of data - Extract Useful Data
Data has no real worth unless we can use it to make better decisions.
A company uses customer data to improve services → that's value.

Velocity

How fast data comes? High Speed of Accumulation of Data
Example; Stock market updates every second, or messages on WhatsApp

Veracity

How True & Reliable data is?
Quality and accuracy of the data.

DATA PREPARATION

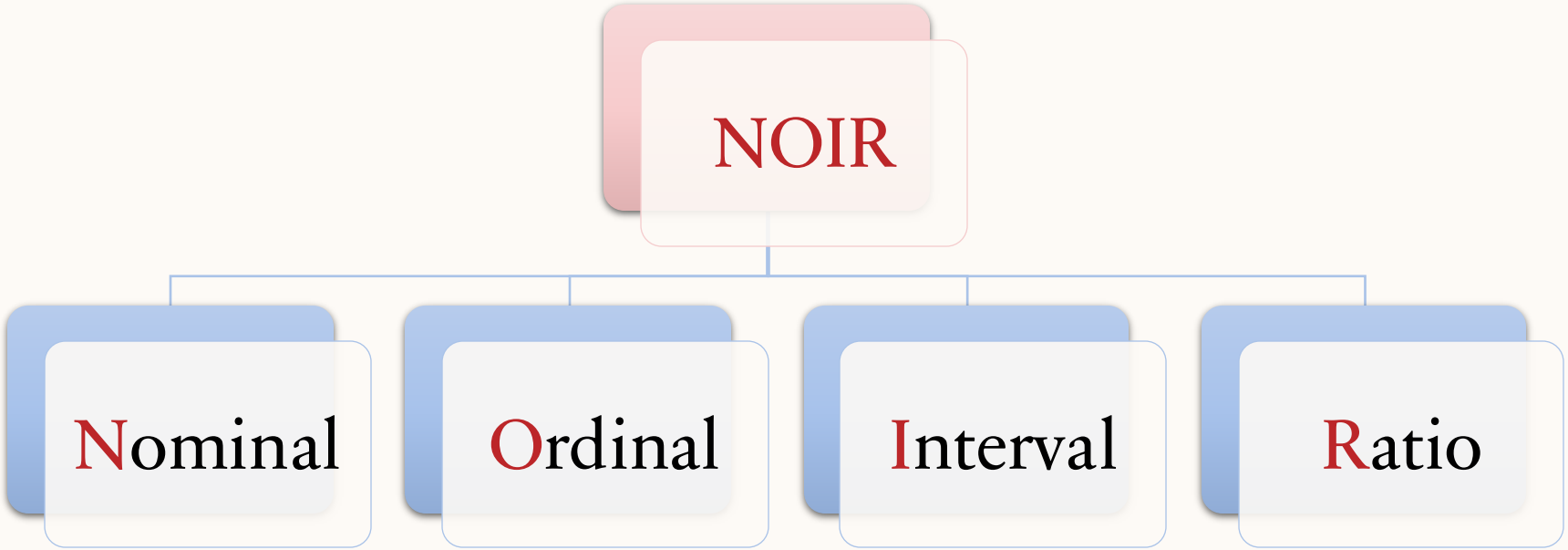
- **Data Cleaning or Data Cleansing:** Data cleansing or data cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and to identify incomplete, incorrect, inaccurate or irrelevant data and then replacing, modifying, or deleting the corrupt data.

- **Data Transformation:** The process of converting data from one format or structure into another format or structure.

DATA IN DATA ANALYTICS

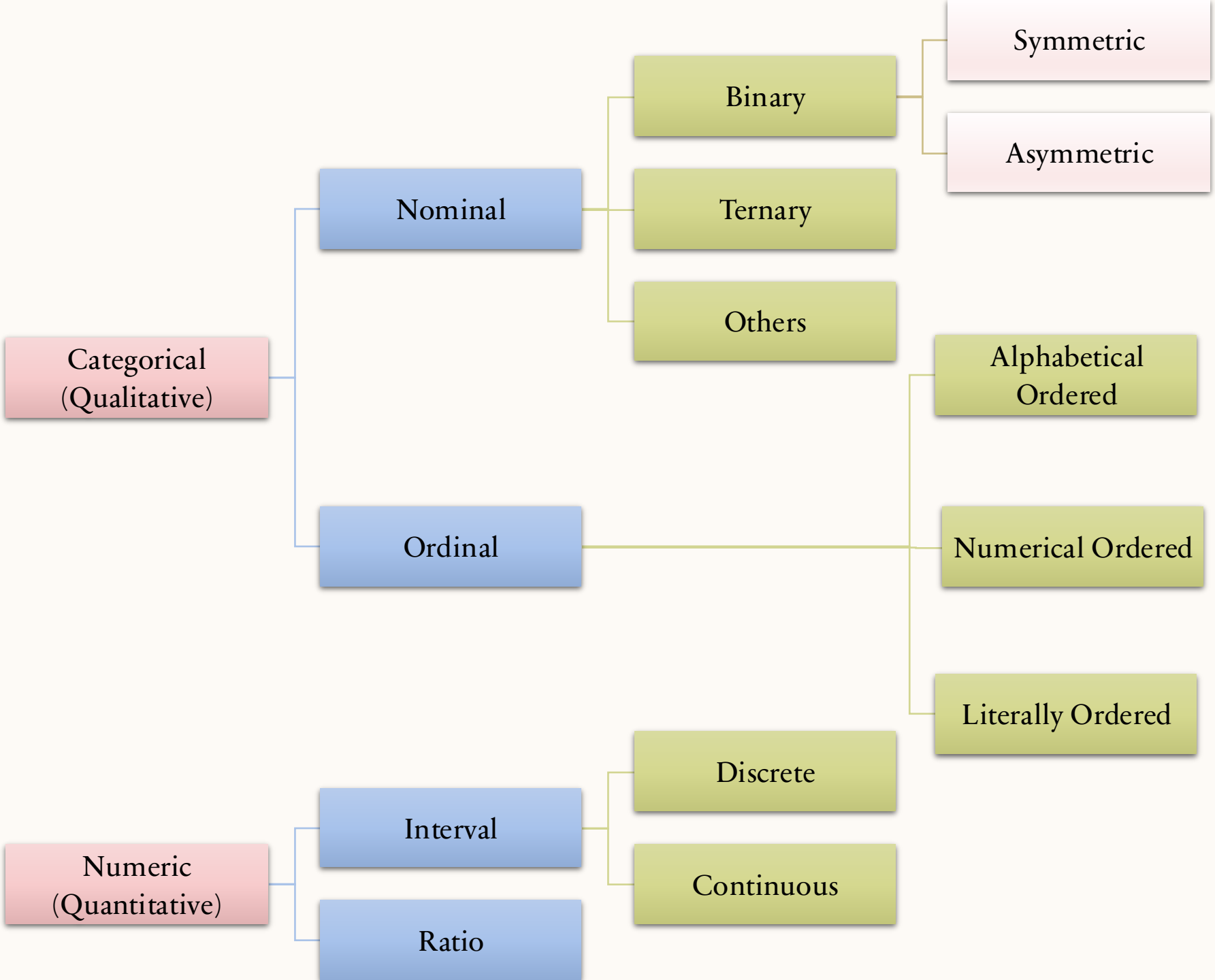
- There are many types of data that can be use to measure the properties of an entity.
- Scales of Measurement are important.
- Depending on the scales of measurement, different techniques are followed to derive the hitherto unknown knowledge in the form of
  - Pattern, association, anomolies, or similatrities, from the volume of data.
- Makes it easy to Write Algorithms

CLASSIFICATION OF SCALES OF MEASUREMENTS



Fundamental Building Block on which the extended data types are Built

LEVELS OF MEASUREMENTS



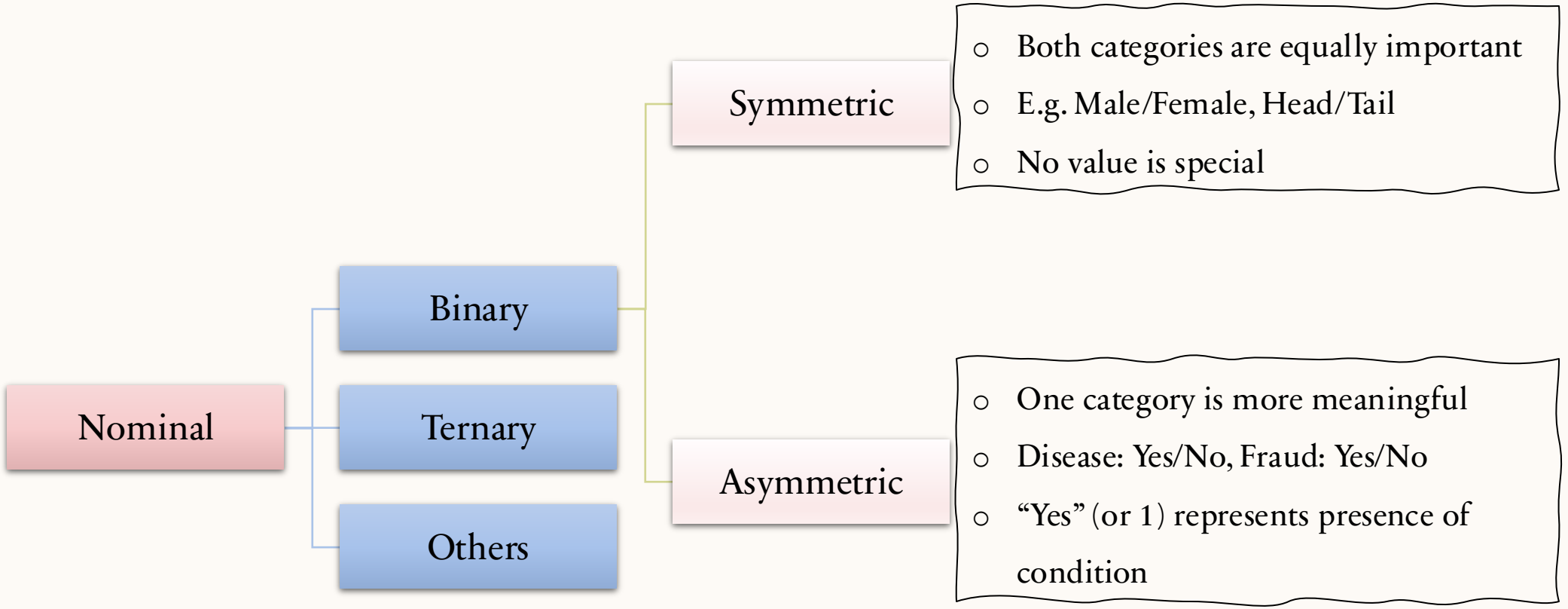
NOMINAL SCALE(NAMES ONLY)

- If the observed data are classified into various distinct categories in which no ordering is implied, a nominal level of measurement is achieved.
- Data are just labels or categories, with no order.

- Example:
  - Gender: Male, Female
  - Colors: Red, Blue, Green
  - City names: Lahore, Karachi, Islamabad
  - Type of Exercise: Yoga, Running, Swimming, Cycling
  - You can't rank or measure them — they're just names.

- *Nominal scale is the weakest form of measurement.*

LEVELS OF MEASUREMENTS



OPERATIONS ON NOMINAL VARIABLES

- Arithmetic (+,-, ×, ÷), and Logical Operations (>, < etc), Mean, Median, Sort, Rank are NOT APPLICABLE.

- Count Frequency (e.g. How many Male?), Mode (e.g. Most common color), percentage or Proportion (40% red), Check Equality (e.g. City A = City B), Compare (Gender vs Favorite Sport, or, City vs Purchase Decision)are VALID OPERATIONS.

- Nominal data can be visualized using Line Charts, Bar Charts, or Pie Charts etc.

ORDINAL SCALE (ORDER MATTERS)

- If the observed data are classified into distinct categories in which ordering is implied.
- Data can be **ranked or ordered**, but **differences between values aren't equal**.
- Ordinal scale is also a **WEAK** form of measurement because no meaningful numerical statements can be made about the different categories.
- For example, the ordinal scale only tells which category is greater but does not tell how much greater.
- *You can say which is higher or lower, but not by how much.*
- Example:
  - Satisfaction: Poor, Fair, Good, Excellent
  - Education level: High school, Bachelor, Master, PhD

LEVELS OF MEASUREMENTS

Ordinal

Alphabetical Ordered

- The order is based on the alphabet, not on real meaning or value.
- Grades in Exam
  - A, B, C, D, F

Numerical Ordered

- The order is based on numbers, from smallest to largest (or vice versa).
- Satisfaction levels coded as:
  - 1 = Poor, 2 = Fair, 3 = Good, 4 = Excellent
- If sorted numerically, it's naturally: $1 < 2 < 3 < 4$

Literally Ordered

- The order is based on the logical or natural meaning of the words, not by alphabet or numeric codes.
  - Education level: Primary → Secondary → Bachelor → Master → PhD
- If we go by literal order, this is correct (based on real-world meaning).
- If we sort alphabetically, "Bachelor" would come first — which is wrong.

OPERATIONS ON ORDINAL DATA

- Arithmetic (+,-, ×, ÷), Mean, are NOT APPLICABLE.
- Logical Operations (>, <, etc) *(e.g. Customer 1's rating (Poor) < Customer 3's rating (Good))*, Sort, Count Frequency *(e.g. How many time each category appers?)*, Mode (e.g. Most common rating), percentage or Proportion (80% rating excellent), are VALID OPERATIONS.
- Calculation based on order (e.g. count, min, max etc)
- All operations on Nominal data defined are also valid.

INTERVAL SCALE (ORDER & EQUAL DISTANCE)

- Values have meaningful order and equal gaps, but no true zero.
- Ordered scale: the difference between any two measurements is a meaningful quantity

- Example:
  - a person who is 70 inches tall is 2 inches taller than someone who is 68 inches tall.
  - Temperature in °C or °F (difference between 20°C and 30°C is same as between 30°C and 40°C)
  - But 0°C doesn't mean "no temperature."
  - So, you can add and subtract, but ratios don't make sense.

LEVELS OF MEASUREMENTS

Interval

Discrete

- Data can take only certain fixed or separate values (whole numbers).
- You can count the possible values.
- No fractions or decimals within the scale.
- *Example: IQ scores (e.g., 85, 90, 95, 100…) OR Calendar years (e.g., 1990, 1991, 1992, no 1995.3)*
- The difference between values is equal, but you can't have in-between values.

Continuous

- Data can take any value within a range, including decimals or fractions.
- You cannot count all possible values because they are infinite within a range.
- *Example: Temperature in °C or °F (e.g., 36.5°C, 36.6°C…) OR Time of day (e.g., 2:30 PM, 2:31 PM, 2:31:45 PM)*
- equal intervals,
- values can be infinitely precise.

OPERATIONS ON INTERVAL DATA

- Arithmetic (×, ÷), Mean, are NOT APPLICABLE. Only Multiplication by a constant is permissible.
- Arithmetic (+,-) are VALID OPERATIONS.
- All operations on Ordinal data defined are also valid.
- Interval Data can be Graphed using Histogram & Frequency polygon etc

RATIO SCALE

- Has everything: order, equal gaps, and a true zero.
- If in addition to the difference being meaningful and equal at all points on a scale, there is also a “true zero” point in which the ratio of measurements is sensible to consider, then the scale is a ratio scale.
- 0 means nothing of that quantity exists

- Example:
  - Height, Weight, Age, Income
  - 0 kg means no weight; 0 rupees means no money
  - You can add, subtract, multiply, and divide these values.

OPERATIONS ON RATIO DATA

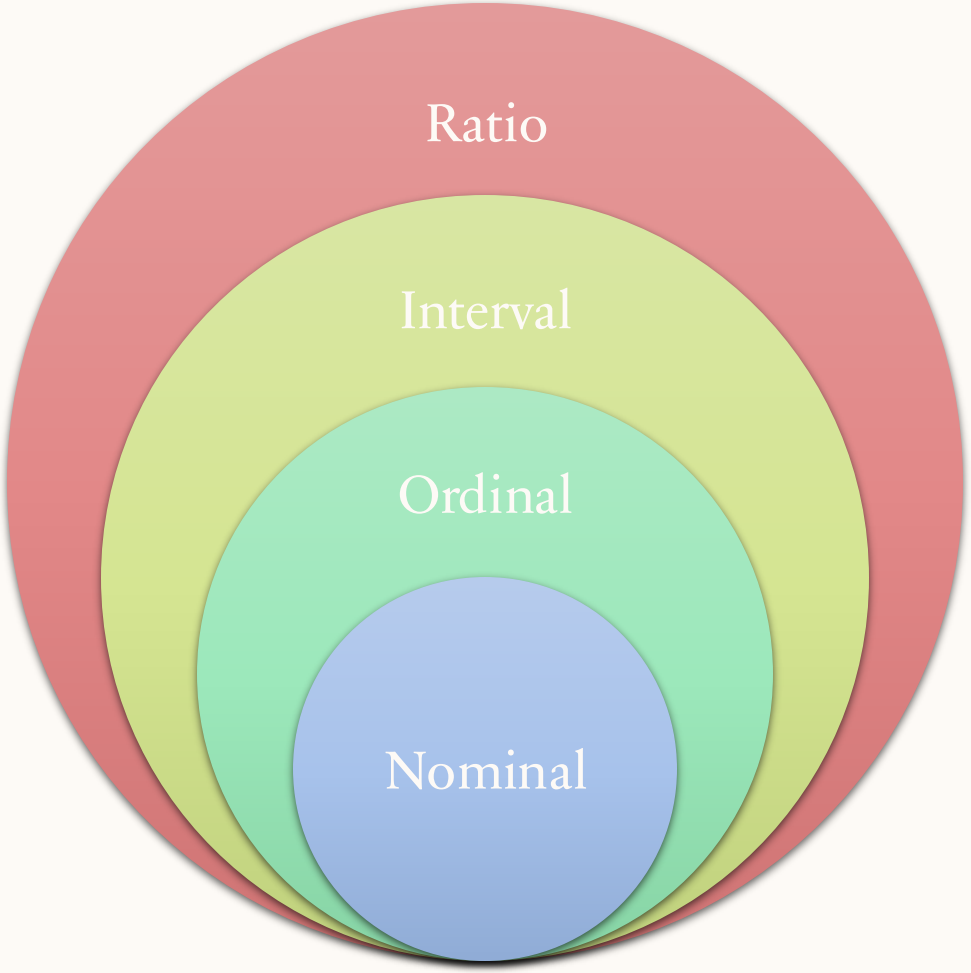
- Arithmetic ($\times, \div$), Mean, are VALID OPERATIONS.
- All operations on Interval data defined are also valid.
- All Ratio data is Interval Data but reverse is not true.
- Both Interval, and Ratio data can be stored in the same data type (e.g. integer, float, etc)

COMPARISON OF 4-LEVELS OF DATA

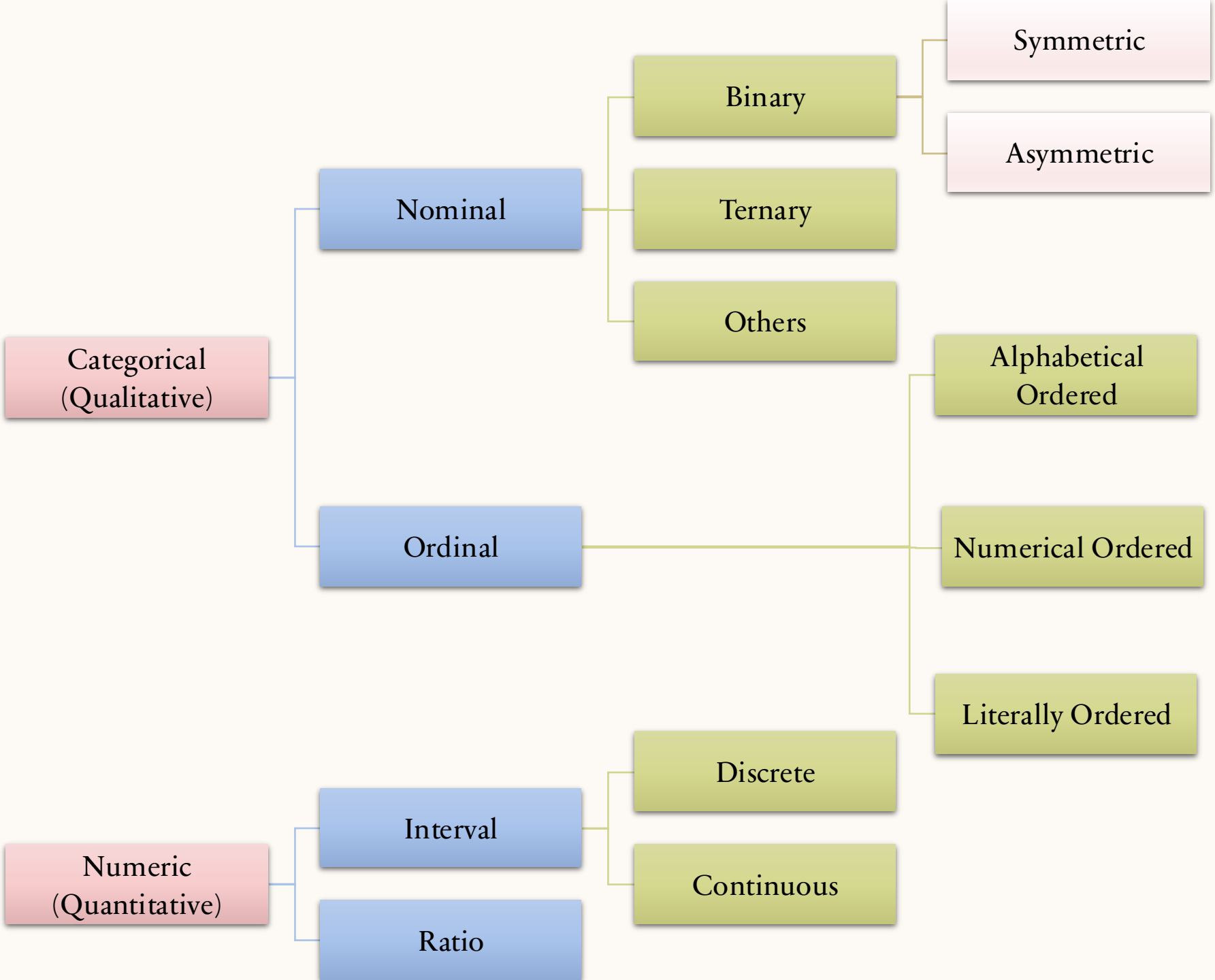
Level	Order	Equal Distance	True Zero	Example
Nominal	❌	❌	❌	Gender, Color,
Ordinal	✅	❌	❌	Rank, Education Level
Interval	✅	✅	❌	Temperature
Ratio	✅	✅	✅	Weight, Age, Slaray

- **Question:** What is your favourite type of exercise? **Options:** Yoga, Running, Swimming, Cycling
- **Question:** How often do you exercise per week? **Options:** Rarely, Sometimes, Often, Very Often
- **Question:** What is the temperature of your gym (in °C)? **Answer:** 25°C, 30°C, etc.
- **Question:** What is your weight (in kilograms)?" **Answer:** 0 kg, 60 kg, 80 kg, etc.

COMPARISON OF 4-LEVELS OF DATA



LEVELS OF MEASUREMENTS



DATA VISUALIZATION

DATA VISUALIZATION

- **Quantitative Data Graphs:**
  - Histograms:
  - Frequency Polygons
  - Ogives
  - Dot Plots
  - Stem and Leaf Plots
- **Qualitative Data Graphs:**
  - Pie Charts
  - Bar Graphs
  - Pareto Charts

Display Data