Welcome to Insight Community

Session 01: Basic Statistics

IMI NEW DELHI

INSIGHT
Data Science Society

General Rules:

- Keep yourselves on mute at all the times unless asked otherwise. For any queries, put it in the chat and it'll be taken care of.

- Attendance for the session will be floated randomly during the session. There will be a QR code that you'll need to scan and mark the attendance.

- There will be assignments that need to be submitted before deadline. They shouldn't be copied.

- Regular attendance, regular & quality assignment submissions are an integral part of evaluation for the 2nd round of Insight hiring.

- You can reach out to anyone from the Insight team for any query.

I. Population vs Sample

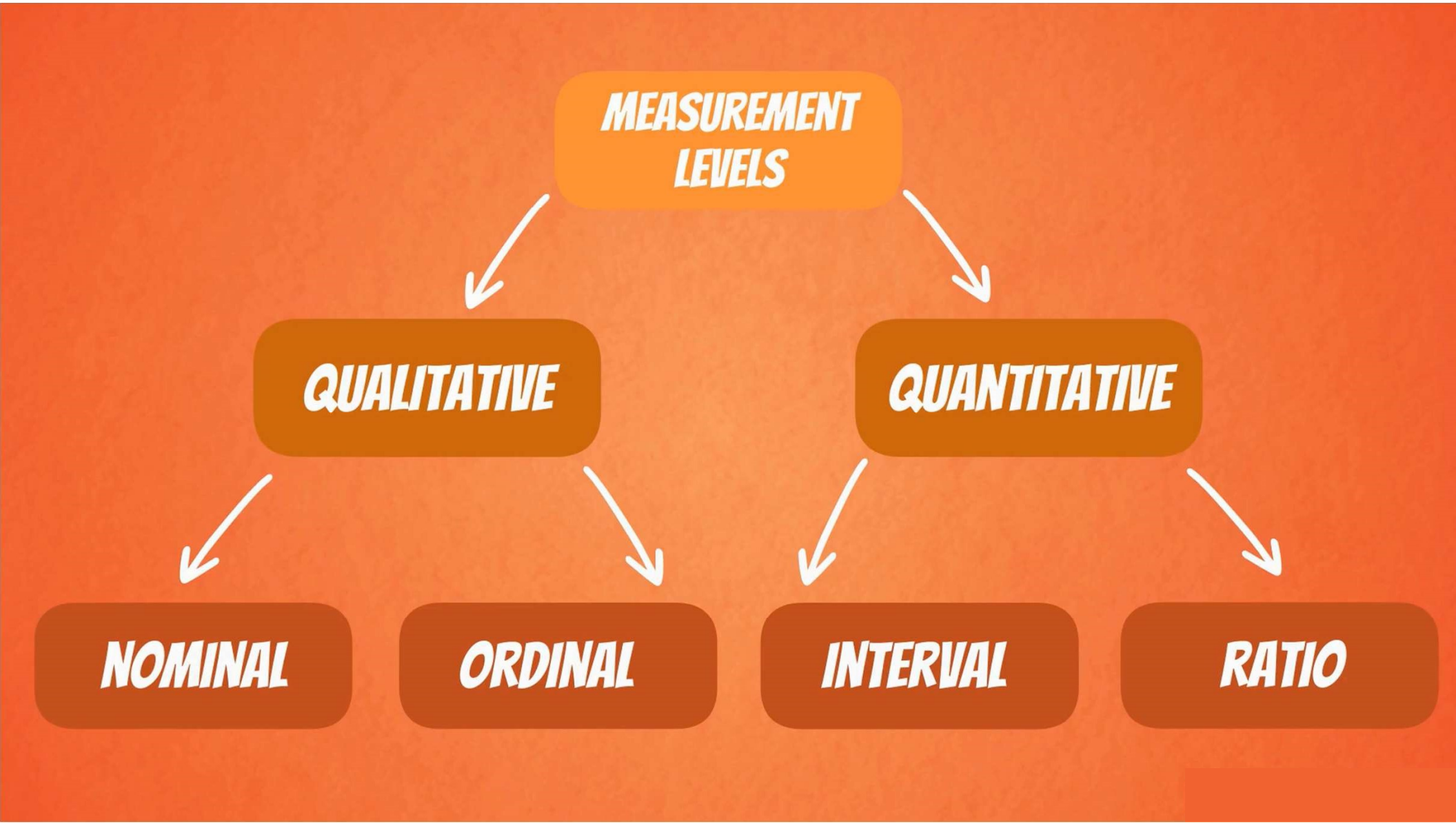
POPULATION	SAMPLE
Collection of all items of interest	A subset of the population
N	n
parameters	statistics

POPULATION
SAMPLE
Hard to observe
Hard to contact
Easy to observe
Easy to contact

Features of a sample

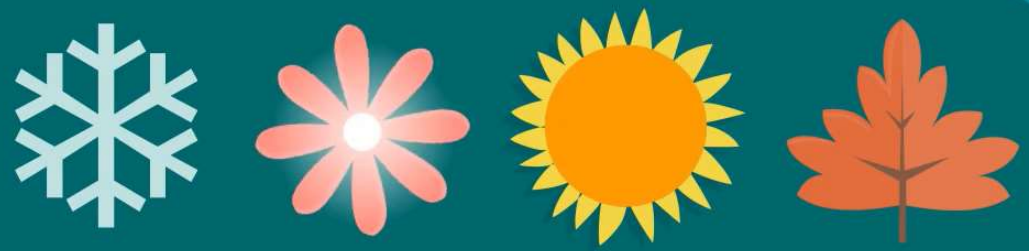
Randomness	Representativeness
When each member of the sample is chosen from the population **strictly by chance.**	A subset of the population that accurately reflects the members of the entire population.

II. Type of Variables


MEASUREMENT LEVELS
QUALITATIVE
QUANTITATIVE
NOMINAL
ORDINAL
INTERVAL
RATIO

NOMINAL

Mercedez BMW Audi



Winter Spring Summer Autumn

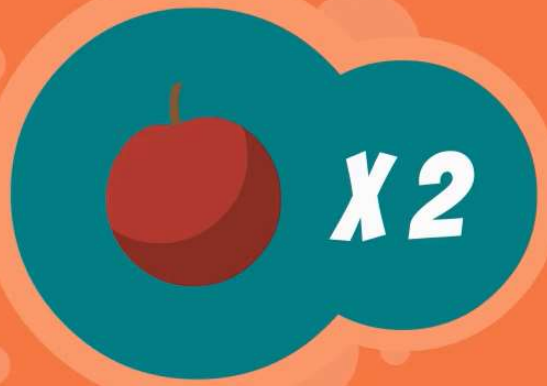
(CANNOT BE ORDERED)

ORDINAL

(HAS A SPECIFIED ORDER)

(Ratio has a true zero)

RATIO



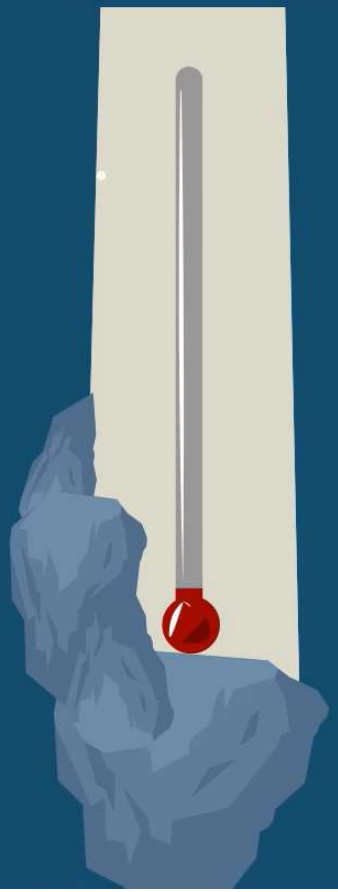
You have 3 times as many as I do.

RATIO OF 6/2 IS 3

INTERVAL
(Interval doesn't have a true zero)
TODAY
5°C or 41°F
YESTERDAY
10°C or 50°F

INTERVAL
TODAY
seems colder in C°
in terms of F°... not really

INTERVAL



$0^{\circ}C$ AND $0^{\circ}F$ ARE NOT TRUE ZEROS

$0^{\circ}K = -273.15^{\circ}C = -459.67^{\circ}F$

The Temperature Dilemma

0 degree Celsius and 0 degree Fahrenheit are not true zeroes

These scales were artificially created by humans for convenience

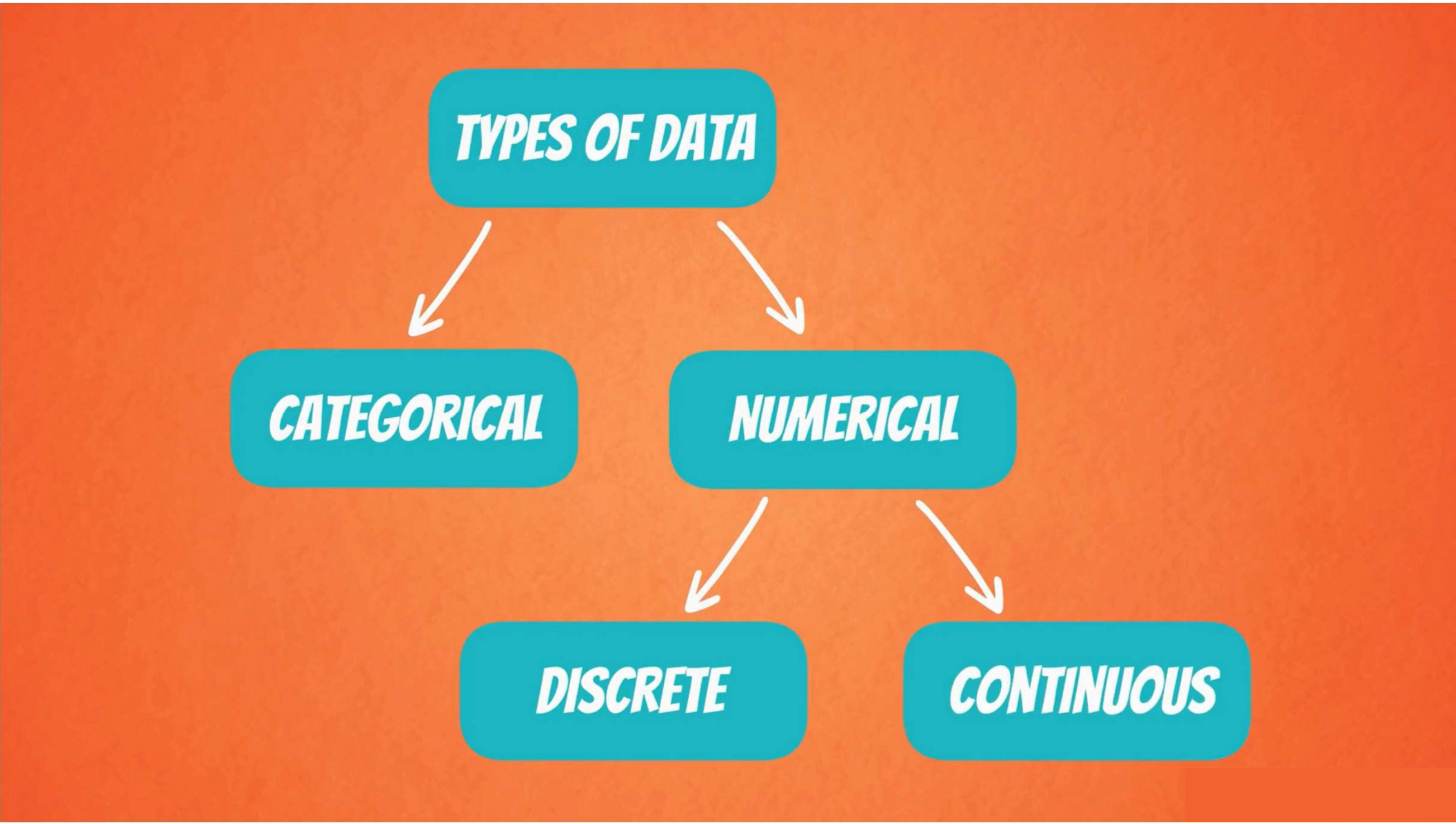
0 degree Kelvin is a true zero. That is the coldest anything can get.

Variables in Kelvin are **Ratios** while Variables in Celsius & Fahrenheit are **Intervals.**

All numbers like 1, 2, 3, 10, 12.42342, e, pi, etc. can be **both** Interval and Ratio.

Just be careful of the context you are operating in!!

III. Distributions


TYPES OF DATA
CATEGORICAL
NUMERICAL
DISCRETE
CONTINUOUS

Thank you…