Introduction to Statistics

Data Science Program

Purwadhika
Startup and Coding School

Outline

- What is Statistics
- Step by step statistics
- Data
- Variable
- Scale of Measurement
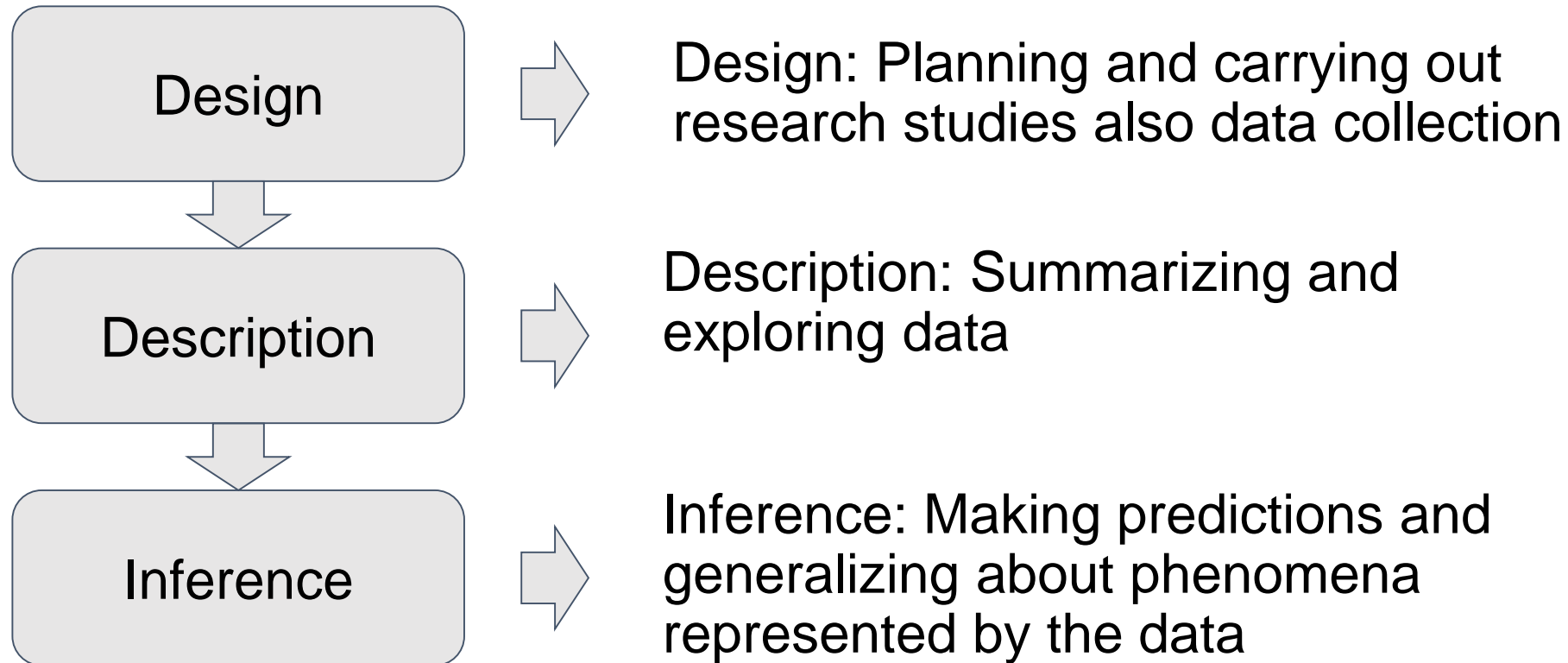

CRISP-DM Process Diagram

Source: Kenneth Jensen
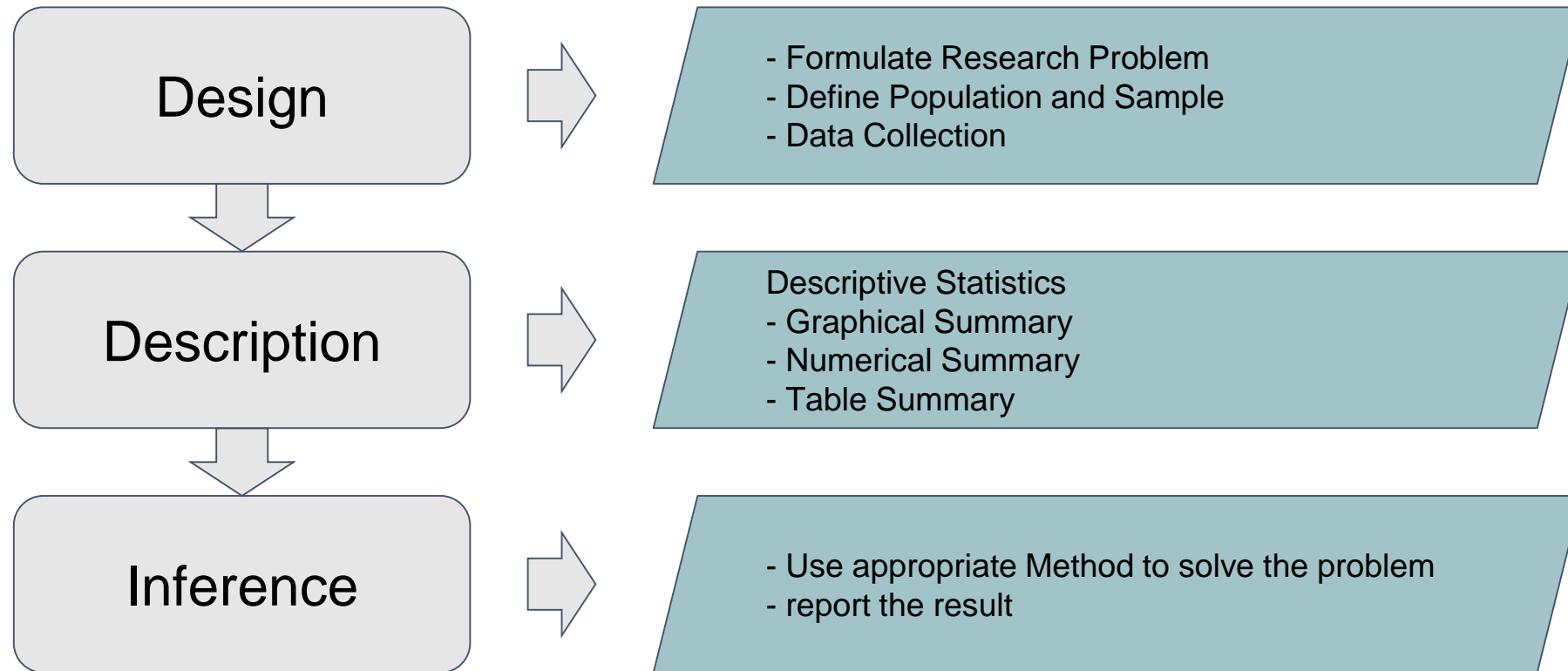
What is Statistic?

- Statistic could be described as methodology for collecting, analyzing, interpreting, and drawing conclusion from data.
- Furthermore, Statistics is the art and science of designing studies and analyzing the data that those studies produce. Its ultimate goal is translating data into knowledge and understanding of the world around us. In short, statistics is the art and science of learning from data.
- It is a very broad subject, with many application on various field:
  - **Experimental Design**, A/B Testing for new web apps design
  - **Survey**, Predicting Election Using an Exit Poll / Quick Count
  - **Research**, Making Conclusion in Medical Research Studies
  - **Quality Control** for any product in a factory, etc

Purwadhika
Startup and Coding School

Step by Step & Types of Statistics

Purwadhika
Startup and Coding School

Step by Step Statistics

Design	→	Design: Planning and carrying out research studies also data collection
↓		
Description	→	Description: Summarizing and exploring data
↓		
Inference	→	Inference: Making predictions and generalizing about phenomena represented by the data

Purwadhika
Startup and Coding School

Step by Step Statistics

Design

- Formulate Research Problem
- Define Population and Sample
- Data Collection

Description

Descriptive Statistics
- Graphical Summary
- Numerical Summary
- Table Summary

Inference

- Use appropriate Method to solve the problem
- report the result

Purwadhika
Startup and Coding School

Type of Statistic

- **Descriptive Statistic**

   Branch of statistic to summarize and describe the data. It consist of methods for organizing and summarizing information.

- **Inferential Statistic**

   Branch of statistic to use the data sample to make an inference about a population. It consist of methods for drawing and measuring the reliability of conclusion based on the sample from the population.

Purwadhika
Startup and Coding School

Data & Variable

Data

Individual units of information

Variables

A characteristic that varies from one person or thing to another is called a variable. Ex: Height, Weight, Eye Color, etc.

Type of variable

Variable

Quantitative

Take On Numerical Value. Sensible to add and subtract

Continuous

Percentage, height, weight, age. etc

Discrete

Number of family member, Number of Child, Number of pets in a household, etc

Qualitative

Gender, Satisfaction Level, Education Level, etc

Observation belong to one set of distinct category
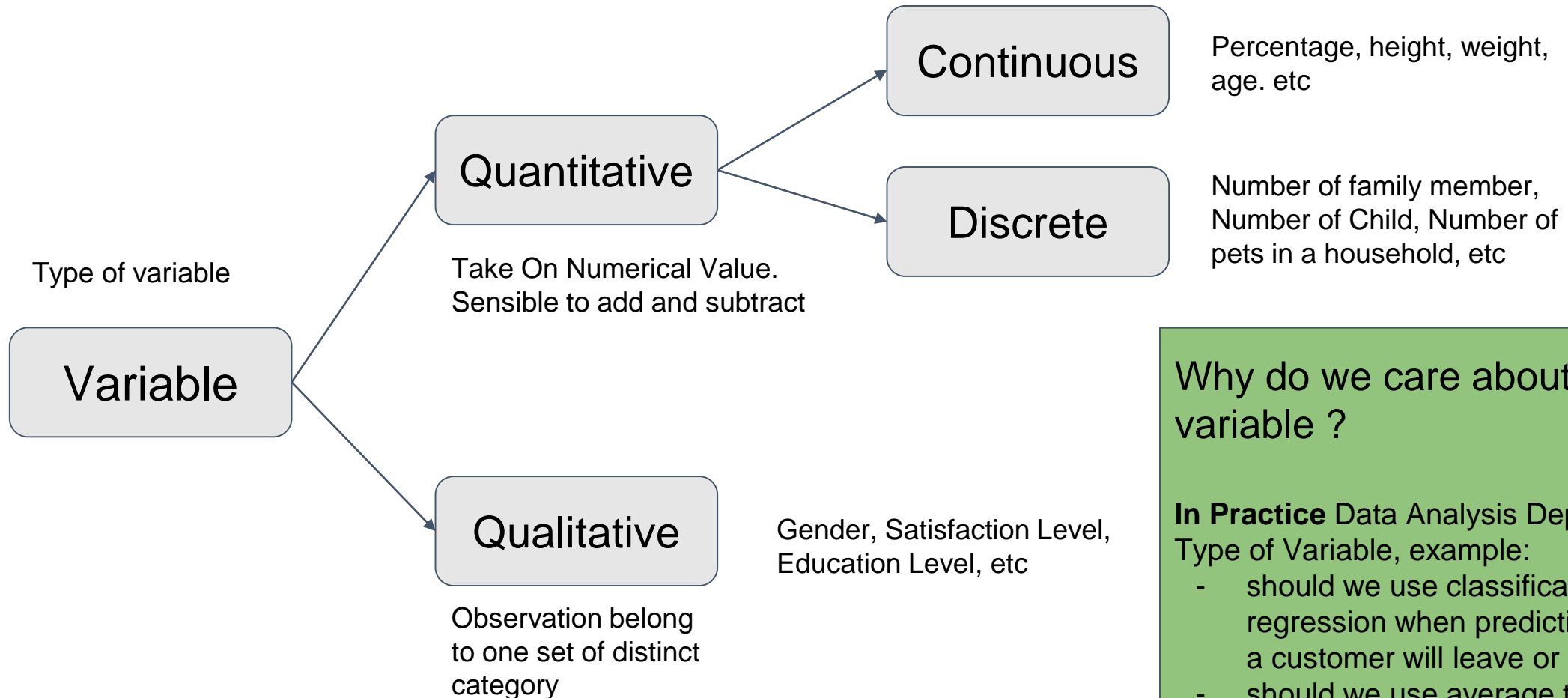
Why do we care about type of variable ?

In Practice Data Analysis Depends on Type of Variable, example:
- should we use classification or regression when predicting whether a customer will leave or not
- should we use average to describe Number of pets in a household, etc

Scale of Measurement

Scale of Measurement

- **Nominal**: Qualitative variables that have two or more categories, but did not have intrinsic order. Ex: Type of fruit (Apple, Banana, Grape), Type of properties(Commercial or Residential), Own a house (Yes or No). Special type of **Nominal** scale when only has two category is **Dichotomous or Binary**.

- **Ordinal**: Qualitative variables that have two or more categories, but the categories could be ranked or ordered. Ex: Satisfaction level (Satisfied, Normal, Not Satisfied), Education Level (SD, SMP, SMA).
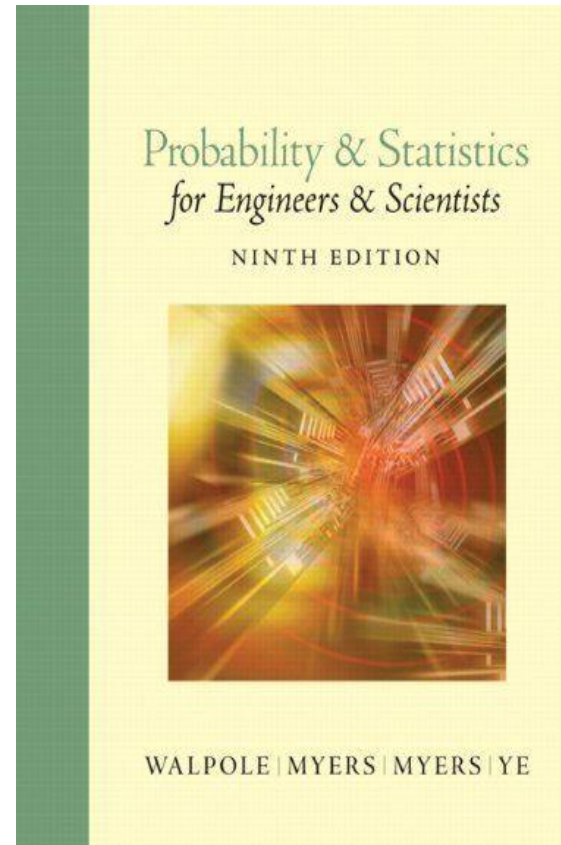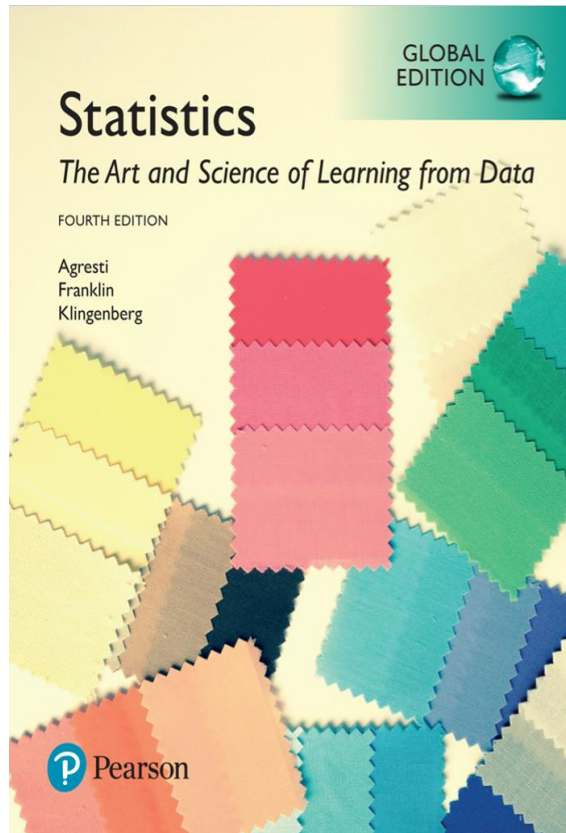
Purwadhika
Startup and Coding School

Scale of Measurement

- **Interval**: Quantitative variable which their central characteristic could be measured along continuum value and have numerical value. Multiplication or division are not sensible. Ex: Temperature, difference between 20C and 30C is the same as difference between 30C and 40C and 40C is not as hot as 2 x 20C.

- **Ratio**: Interval variables, but with the added condition that 0 (zero) of the measurement indicates that there is none of that variable. Multiplication or division are sensible. Values from ratio variable usually is greater than 0. So, temperature measured in degrees Celsius or Fahrenheit is not a ratio variable because 0C does not mean there is no temperature. Ex: Weight 0 kg meaning the unit doesn't exist, 2 x 4 kg is equal to 8 kg. However, temperature measured in Kelvin is a ratio variable as 0 Kelvin (often called absolute zero) indicates that there is no temperature whatsoever.

Scale of Measurement Summary

Scale	Compare	Distance	Order	Classify	Zero	Multiplication or division
Nominal	-	-	-	v	-	-
Ordinal	-	-	v	v	-	-
Interval	-	v	v	v	Non-absolute and can be negative	Not Sensible
Ratio	v	v	v	v	Absolute and usually > 0	Sensible

Purwadhika
Startup and Coding School

Reference

Reference

https://towardsdatascience.com/data-science-you-need-to-know-a-b-testing-f2f12aff619a

https://towardsdatascience.com/data-science-fundamentals-a-b-testing-cb371ceecc27

https://www.niagahoster.co.id/blog/ab-testing-adalah/

https://vwo.com/blog/ab-testing-examples/

https://www.scribbr.com/methodology/sampling-methods/

Purwadhika
Startup and Coding School