

DT/NT : DT

DATE : 03.07.2024

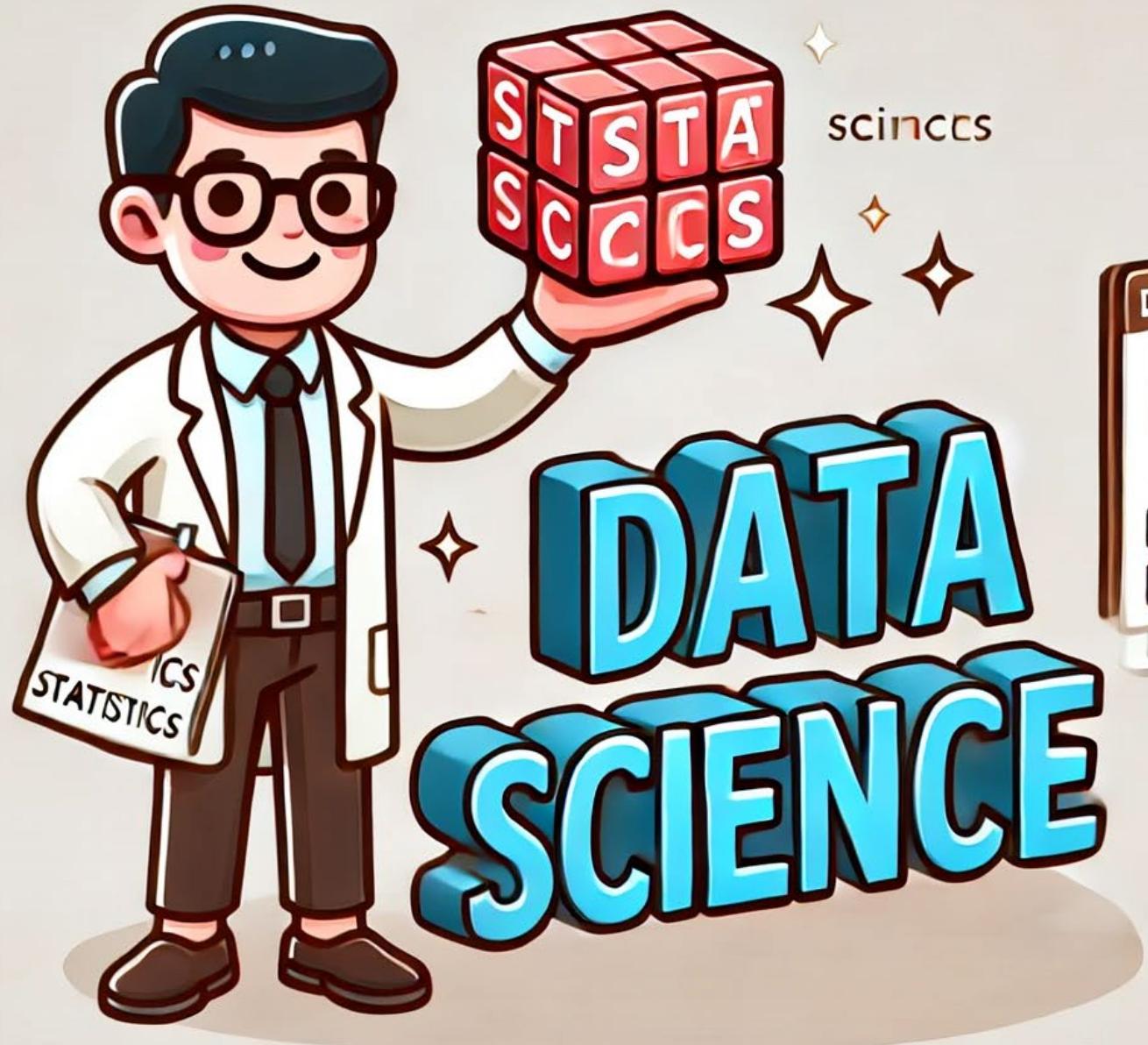
Batch : 250

Lesson: Statistics -1

Subject : Introduction

- Statistics for DS
- Statistics Types
- Data Types





What will we learn in this course?

Part-1

- The meaning of Statistics
 - Eisenhower Matrix
 - Characterization
 - Collection
 - Analyzing
 - Visualization
 - Inference
 - Presentation
 - Why Statistics
 - Importance of Statistics
 - Data Science vs Statistics
 - How much Statistics for us
 - Statistic Types
 - Descriptive
 - Inferential
 - Data Types
 - Parameters and Statistics
 - Probability vs Statistics
 - Level of Measurement
 - Nominal
 - Ordinal
 - Interval
 - Ordinal

Part-2

- Data Visualization - Graphical Represent
- Patterns Center
 - Spread
 - Shape
 - Symmetric
 - Number of peaks
 - Skewness
 - Uniform
 - Unusual Features
 - Gaps
 - Outliers
 - Frequency Table
 - Relative Frequency
 - Cumulative Frequency
 - Bar Chart
 - Pie Chart
 - Histogram
- Populations & Samples
 - Parameters & Statistics
- Central Tendency (Measure of Centre)
 - Mean
 - Median
 - Mode
- Dispersion (Measure of Spread)
 - Range
 - IQR
 - Standard Deviation
 - Empirical Rule
 - Variation

Part-3

- Scatter Plot
 - Linearity
 - Slope
 - Strength
 - Unusual Features
 - Clusters
 - Gaps
 - Outliers
- Box Plot
 - Min & Max Values
 - $1.5 * \text{IQR}$ (John Tukey)
- Covariance
- Correlation
 - Pearson Correlation Coefficient
 - Correlation – Linear Relationship

 Peardeck

 Videos

 Kahoot

 Interview questions

 Notebook hands-on practices

What will we learn in this course?

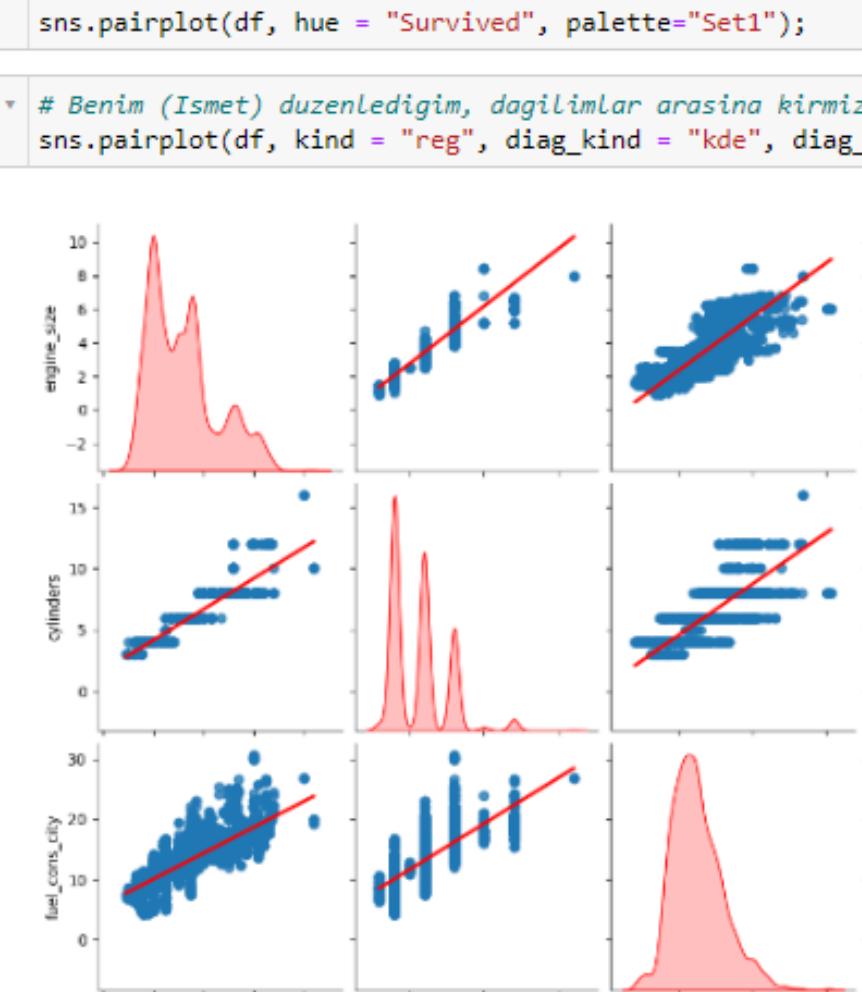


Top 60 Statistics Questions 2024 Interview



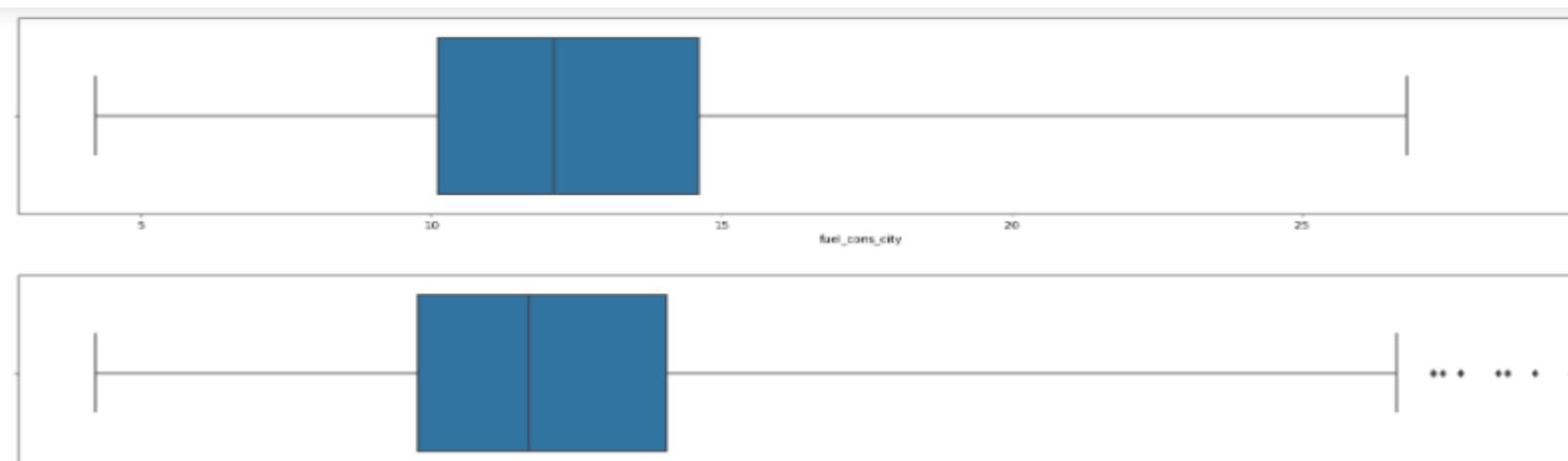
What will we learn in this course?

3.3 Pairplot



3.4 Boxplot

```
sns.boxplot(data=df,y="age",x='income');  
  
index = 0  
plt.figure(figsize=(20,20))  
for feature in df.columns:  
    if feature != "Survived":  
        index += 1  
        plt.subplot(3,3,index)  
        sns.boxplot(x='Survived',y=feature,data=df)  
plt.show()  
  
# Let's draw boxplots and histplots for checking distributions of features;  
index=0  
for feature in df.select_dtypes('number').columns:  
    index+=1  
    plt.figure(figsize=(40,40))  
    plt.subplot((len(df.columns)),2,index)  
    sns.boxplot(x=feature,data=df, whis=3)  
  
    plt.tight_layout()  
    plt.show()
```



Section-1 Content

Content

1. The meaning of Statistics
2. Why Statistics
3. Importance of Statistics
4. How much Statistics for us
5. Statistic Types
6. Data Types
7. Parameters and Statistics
8. Probability vs Statistics
9. Level of Measurements





**Have you practiced
with the pre-class
materials that will
prepare you for
today's lesson?**



Fundamentals Statistics

Online Resources

- YouTube [StatQuest \(Funny\)](#)
- YouTube [CrashCourse](#)
- <https://www.khanacademy.org/>
- [Statistics - A Full University Course on Data Science Basics](#)

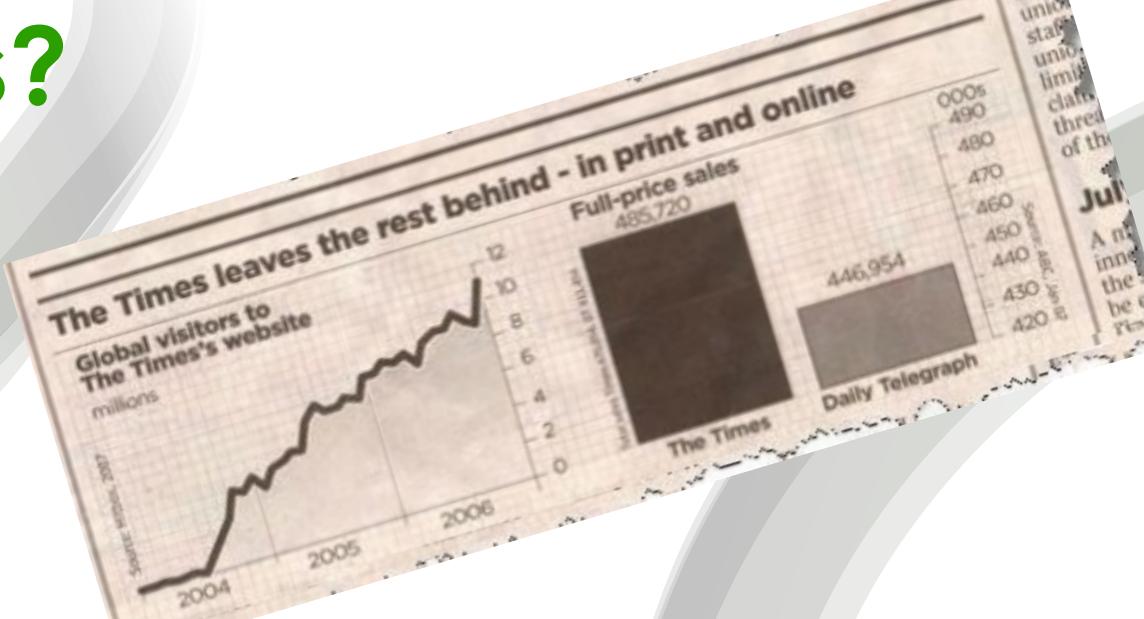
Documents

- LMS page
- PDF Document on Internet

Books:

- Applied Statistics and Probability – Montgomery
- An Introduction to Statistical Learning – James –Witten
- Probability&Statistics - Myers

Why statistics?



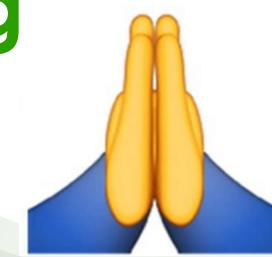
Why statistics?

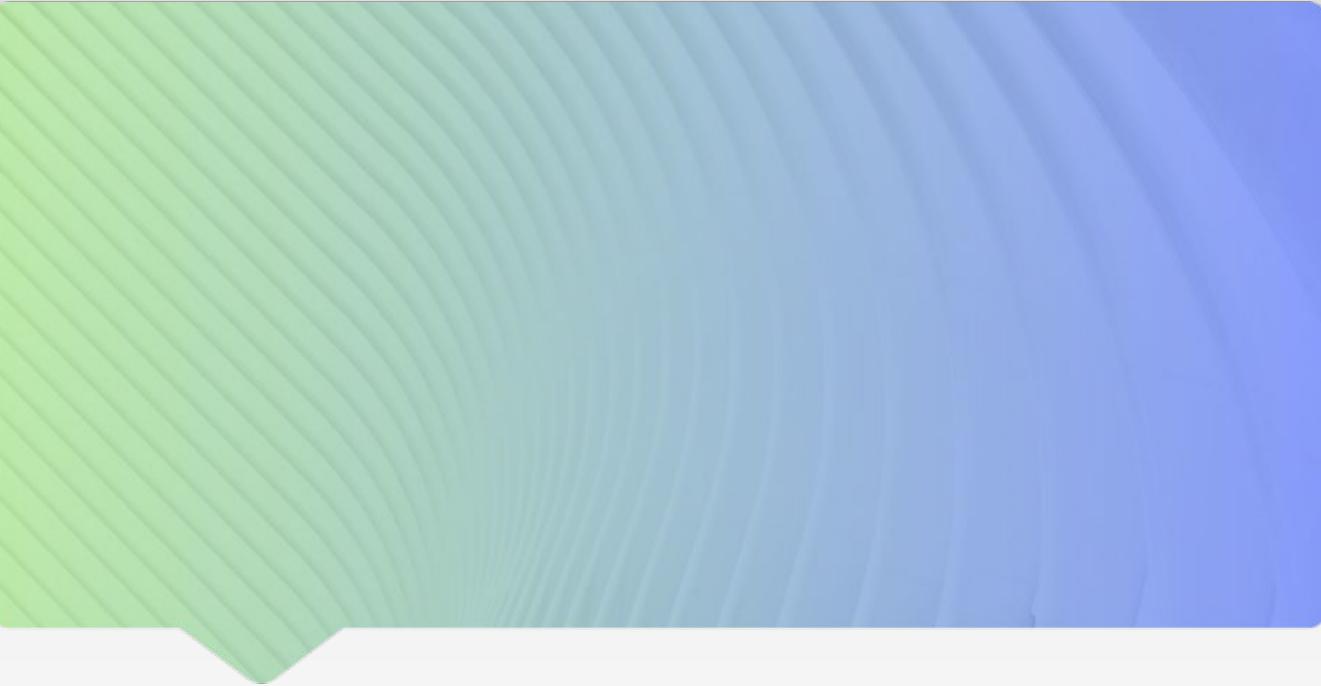


Why statistics?



- ✓ If you say “I know statistics” in an interview, it will increase your offer chance.
- ✓ If you use statistical evidence, figures, and numbers in your presentations, you will get the attention of the audience easily.
- ✓ If you interpret your analysis results by using statistical criteria, your teammates respect you.





Statistics in daily life examples?



Statistics in Daily life examples

▶ Statistics Role In Real Life

- Government
- Emergency Preparedness
- Political Campaigns
- Sports
- Research
- Education
- Prediction
- Predicting Disease
- Insurance
- Financial Market
- Business Statistics

https://medium.com/@john_marshall/statistics-role-in-real-life-a6ba727e0ad8

▶ Statistics In Our Day-to-Day Life

- Medical Study. Statistics are used behind all the medical study. ...
- Weather Forecasts.
- Quality Testing. A company makes thousands of products every day and make sure that they sold the best quality items.
- Stock Market.
- Consumer Goods.

<https://statanalytica.com/blog/uses-of-statistics/>

1. What is the statistics?

- Statistics is the study of how to collect, organize, analyze, and interpret numerical information and data.

= Data Engineering
= Data Scientist

- Statistics is both the science of uncertainty and the technology of extracting information from data.

= Data Analytics, BI

- Statistics is used to help us make decisions. This is especially important in health care and public health.

Statistician
Royal Papworth Hospital NHS Foundation Trust

1. What is the statistics?



Statistics is the grammar of science.

(Karl Pearson) (1857 – 1936)

Statistics as a method:

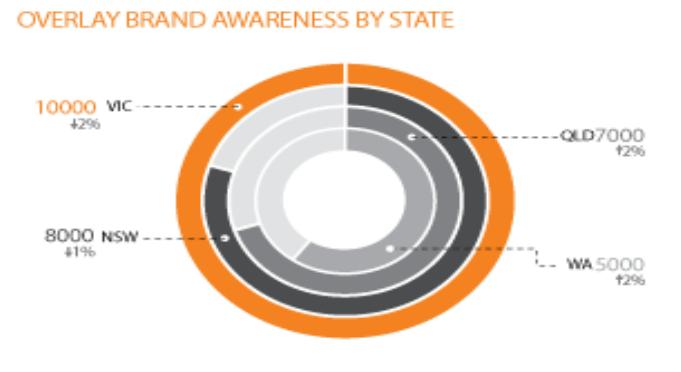
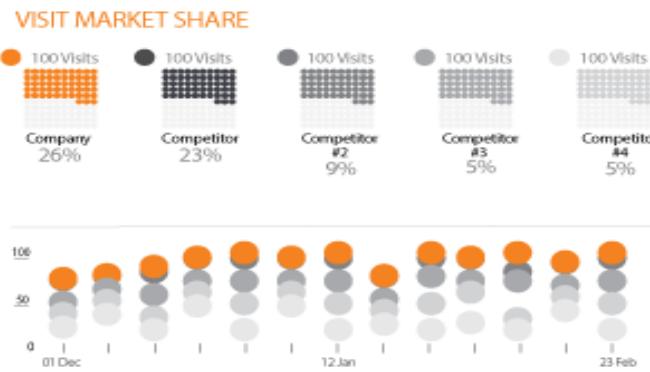
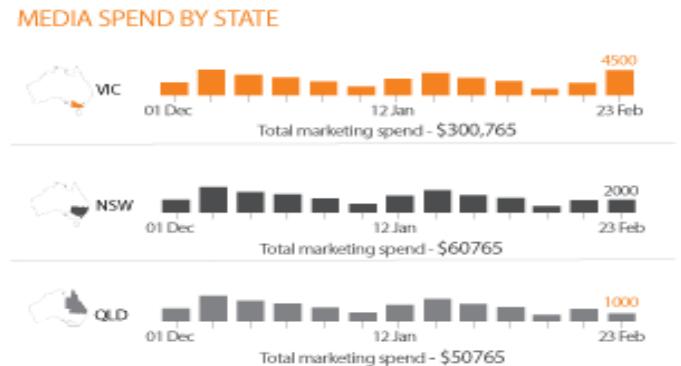
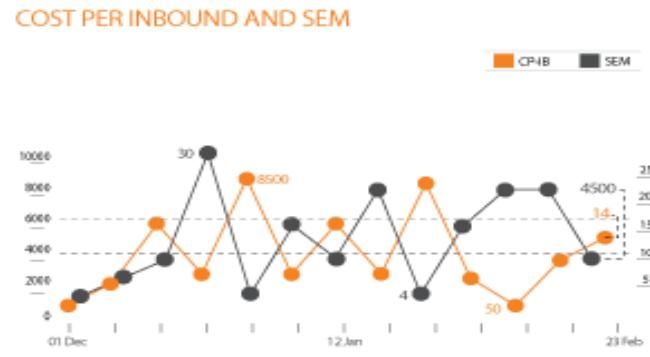
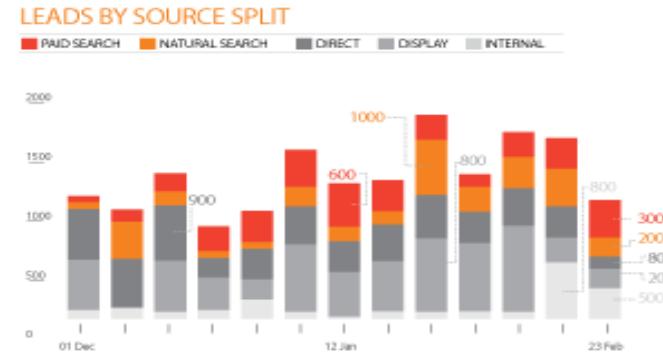
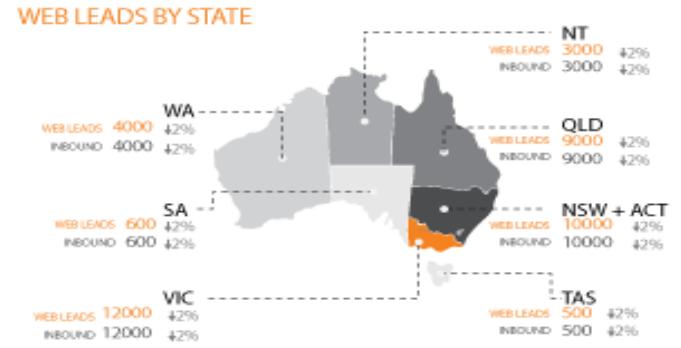
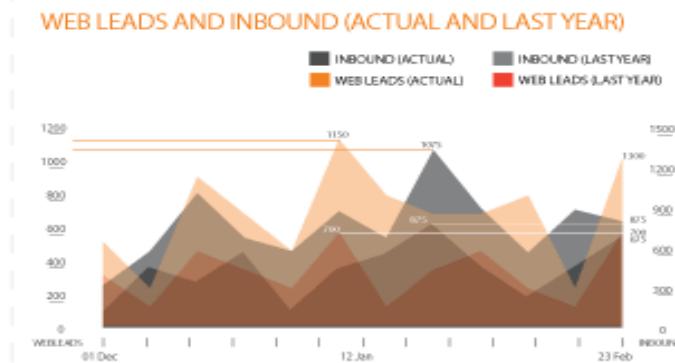
- It is a set of techniques used in **collecting, processing, analysing** and **interpreting quantitative data** on events that can be subject to statistics.

1. What is the statistics?



1. What is the statistics?

Example of Statistics as Dashboard



1. What is Statistics

Statistics is all about data



Collection

Characterization

Analysing

Visualisation

Inference

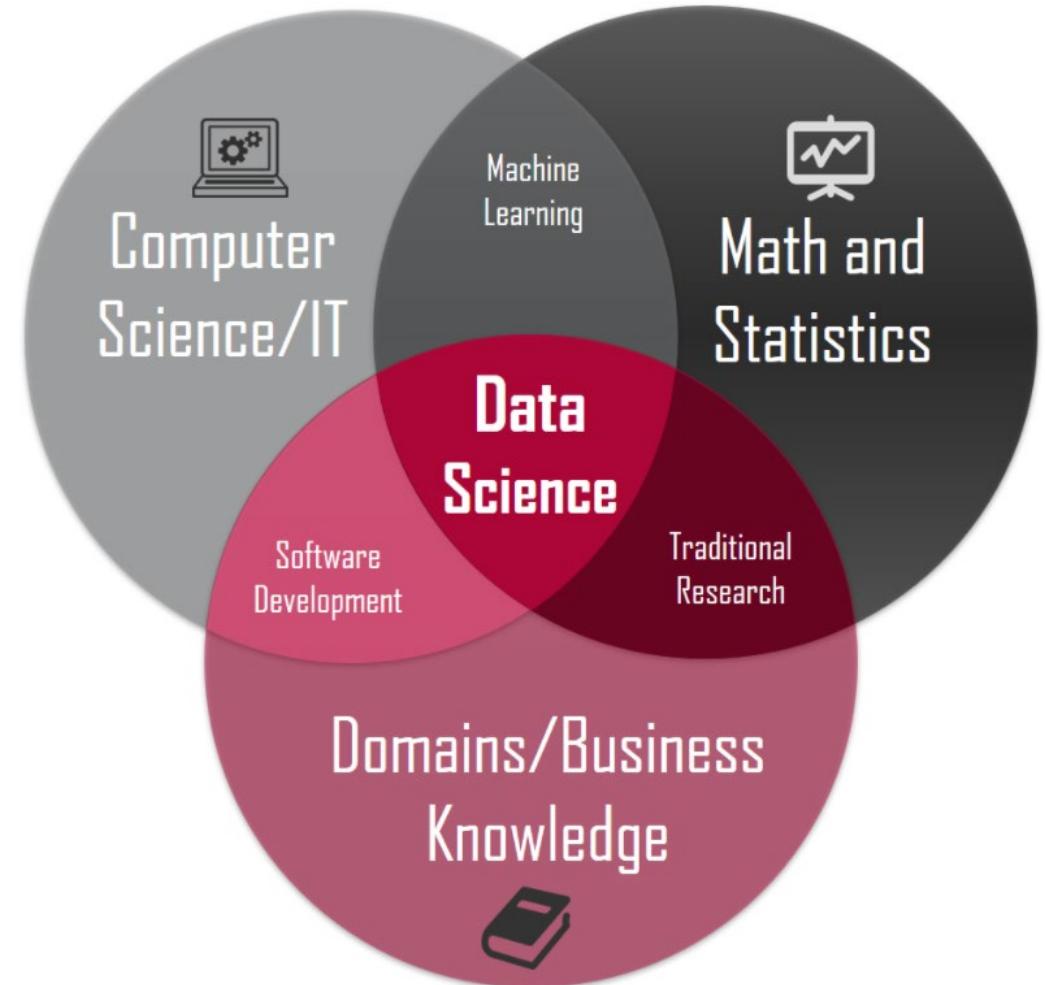
Presentation

Where is the Statistics in Data Science?



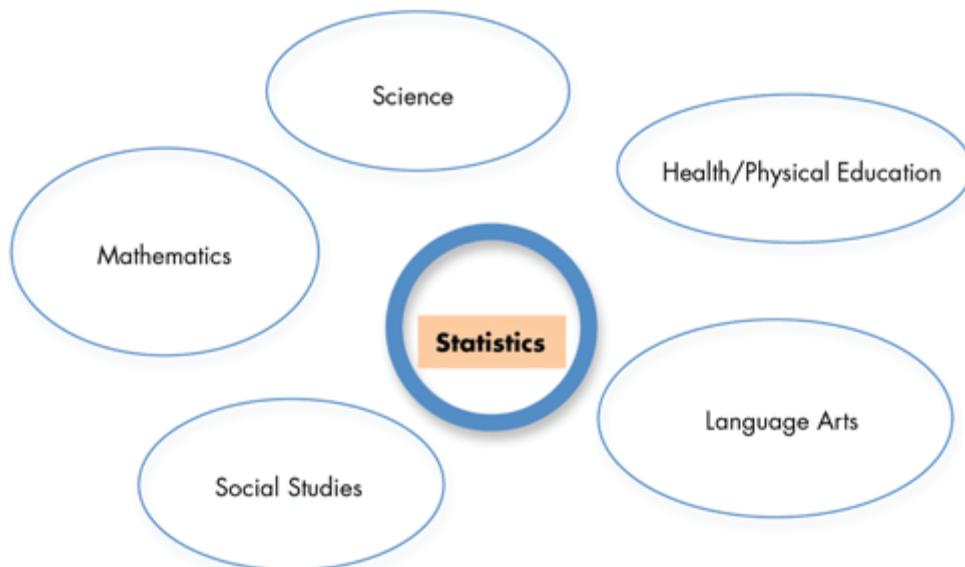
2. Why Statistics

- It helps researchers to explain their findings.
- It helps researchers go beyond the data to more general conclusions.
- Definition of Data Scientist; "**a person who knows more statistics than a programmer and more programming than a statistician**"
- How much Statistics will we learn?



3. Importance of Statistics in Science

Teach Statistics as an Independent Subject



<https://www.emathzone.com/tutorials/basic-statistics/importance-of-statistics-in-different-fields.html>



3. Why statistical knowledge are important in Data Science

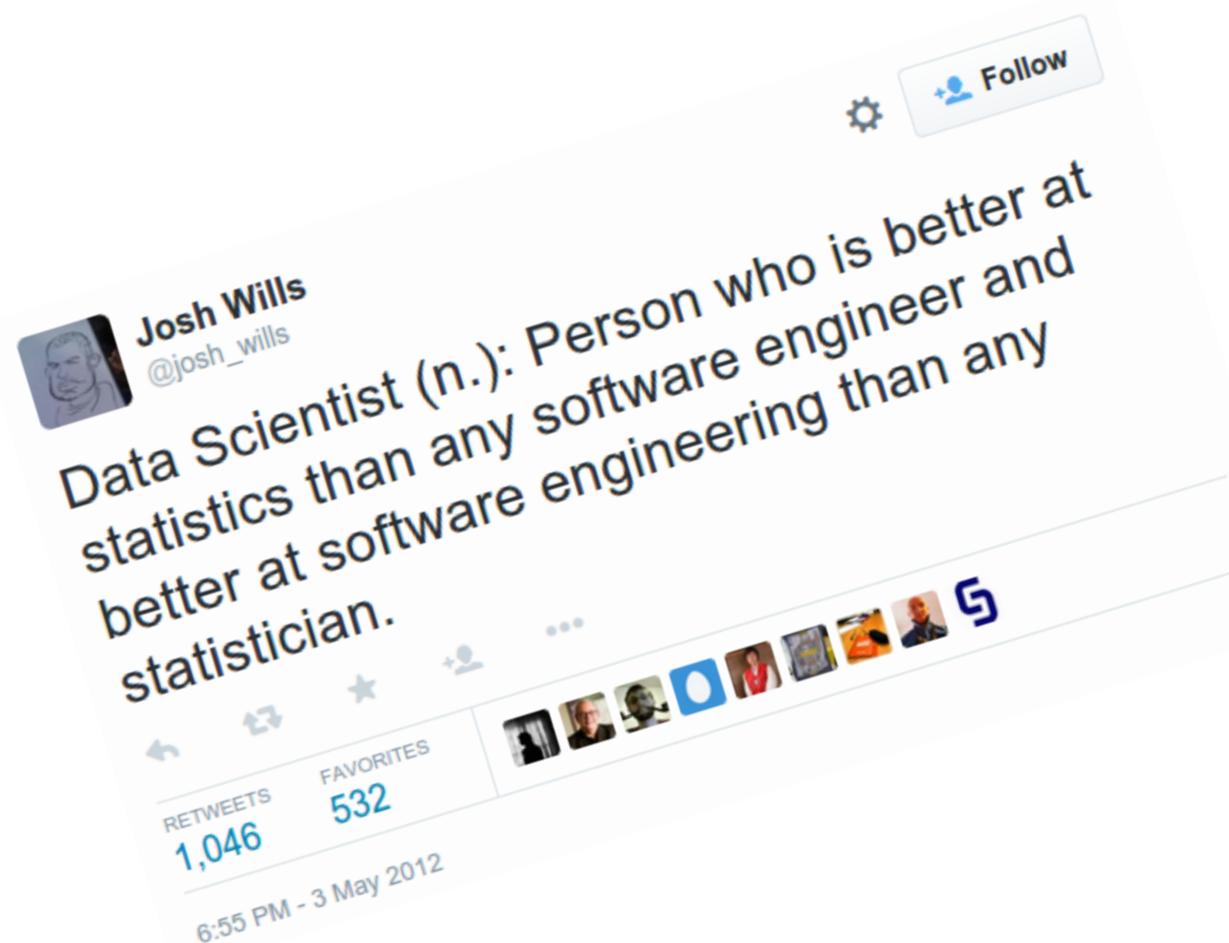
Should I run for the bus?

Which stock should I buy?

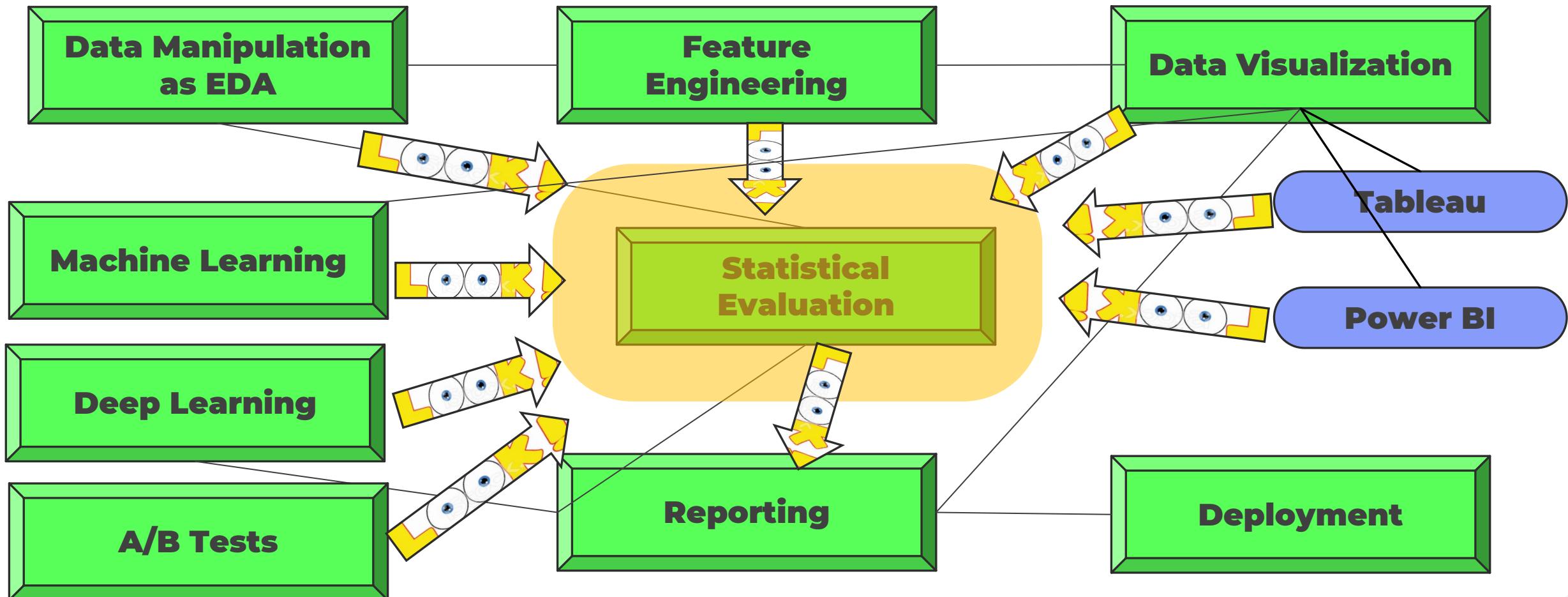
Which products should I list in Website?

Should I take this medication?

Should I have my children vaccinated?



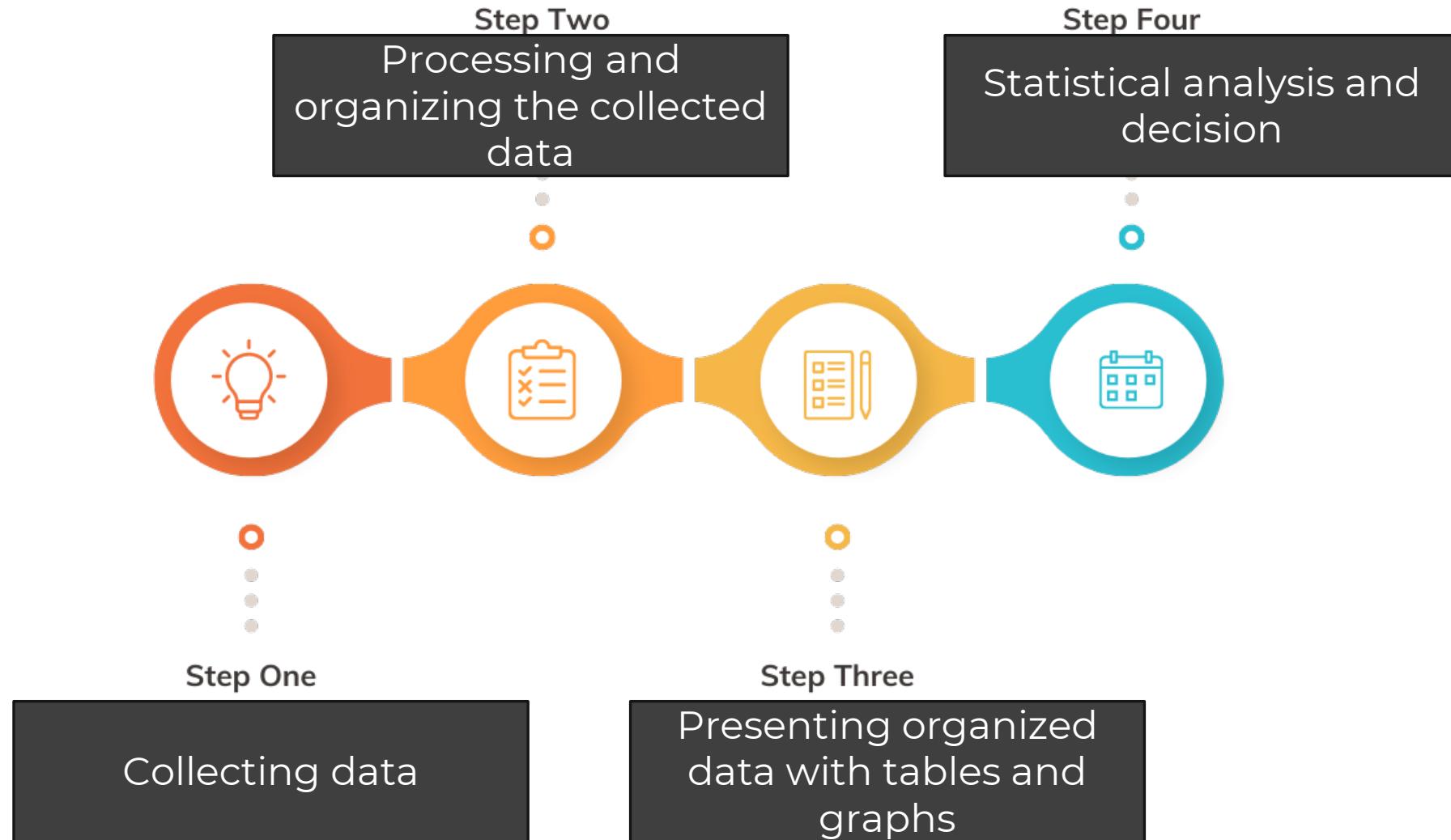
4. Where is the Statistics in Data Science Roadmap



Data Flowing



Statistical Order



Data Science vs Statistics

Differences

Similarities

- Understanding of mathematics
- investigating problems
- exploratory data analysis
- analyzing **trends, patterns**
- creating **forecasts**
- visualizations
- reporting findings to **non-technical users**

Statistics

- one-off reports
- use of SAS programming
- focus on diagnostic plots
- focus on significance testing
- focus on t-tests, ANOVA, and MANOVA, etc.
- more manual data collection (sometimes from surveys)
- usually, you will find Statisticians in the healthcare and economics field
- or more academic settings

Data Science

- automation
- use of SQL querying for data collection
- machine-learning libraries like sklearn and TensorFlow
- use of Python and R programming languages
- deployment of automated models (into an app)
- focus on software engineering practices



DS vs Statistics

DS

A data scientist makes hundreds of decisions every day.

Many of these decisions require a strong foundation in math and statistics.

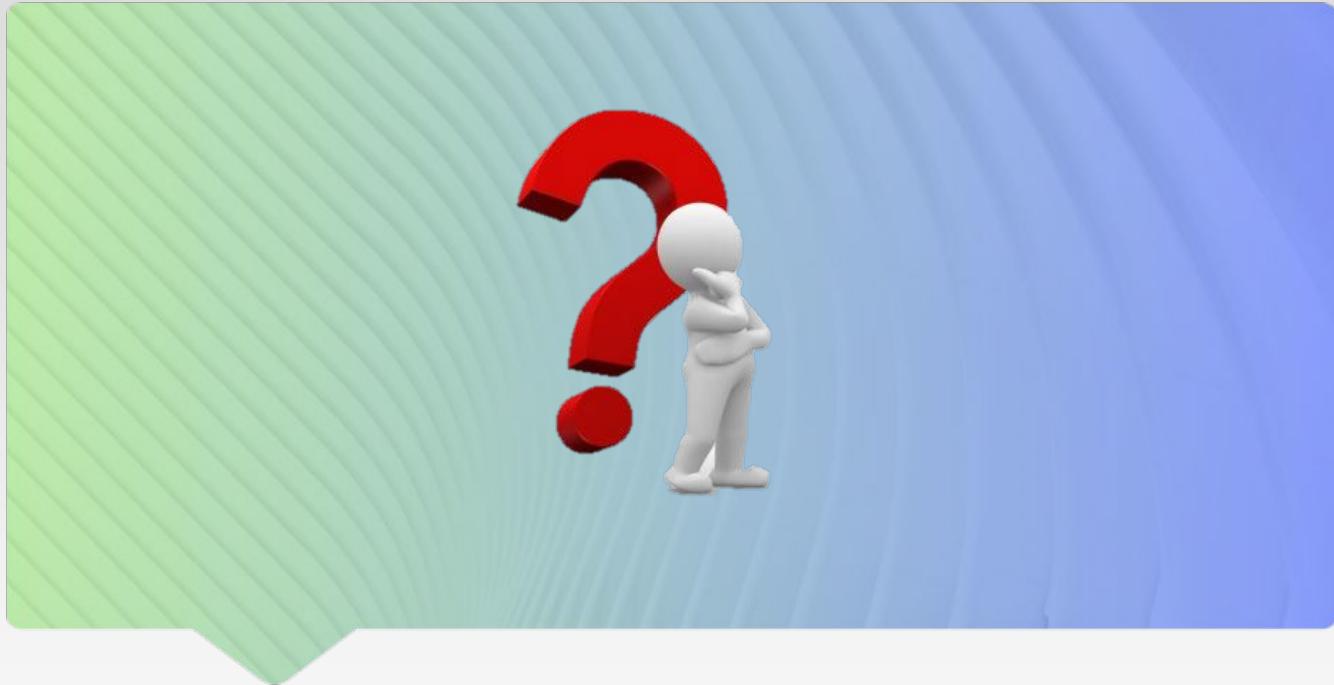
Data science requires descriptive statistics and probability theory, at a minimum.

DS

- ▶ Data science is a multidisciplinary field which uses scientific methods, processes, and systems to extract knowledge from data.
- ▶ Data science problems often relate to making predictions and optimizing search of large databases.
- ▶ Data scientists tend to come from engineering backgrounds.

Statistician

- ▶ Statistics is a mathematically-based field which seeks to collect and interpret quantitative data.
- ▶ In contrast, the problems studied by statistics are more often focused on drawing conclusions about the world at large.
- ▶ Statisticians are usually trained by math departments.



Let's answer the questions in 2 sentences

What is Data Science?

Why are we learning Statistics ?



<https://www.youtube.com/watch?v=bBRKYdkitbU>

<https://www.online-stopwatch.com/timer/10minutes/>

<https://www.youtube.com/watch?v=bBRKYdkitbU>

Tea break...

10:00



Start Stop Reset mins: secs: type:

Second Session

5. Statistics Types

Statistics

Descriptive Statistics

Summarize the data set.

Measures of Central Tendency:

1. Mean: The average of all values.

2. Median: The middle value when data is ordered.

3. Mode: The most frequently occurring value.

Measures of Variability:

1. Range: The difference between the maximum and minimum values.

2. Variance: The average of the squared differences from the mean.

3. Standard Deviation: The square root of the variance.

4. Interquartile Range (IQR): The difference between the 75th and 25th percentiles.

Inferential Statistics

Work on sample, generalize for population.

A branch of statistics that deals with using data from a sample to draw conclusions about a larger population. It allows you to make predictions and generalisations that go beyond the information contained in the sample itself.
For instance; surveys

1. Hypothesis testing:
2. Confidence intervals
3. Regression analysis

5. Statistics Types



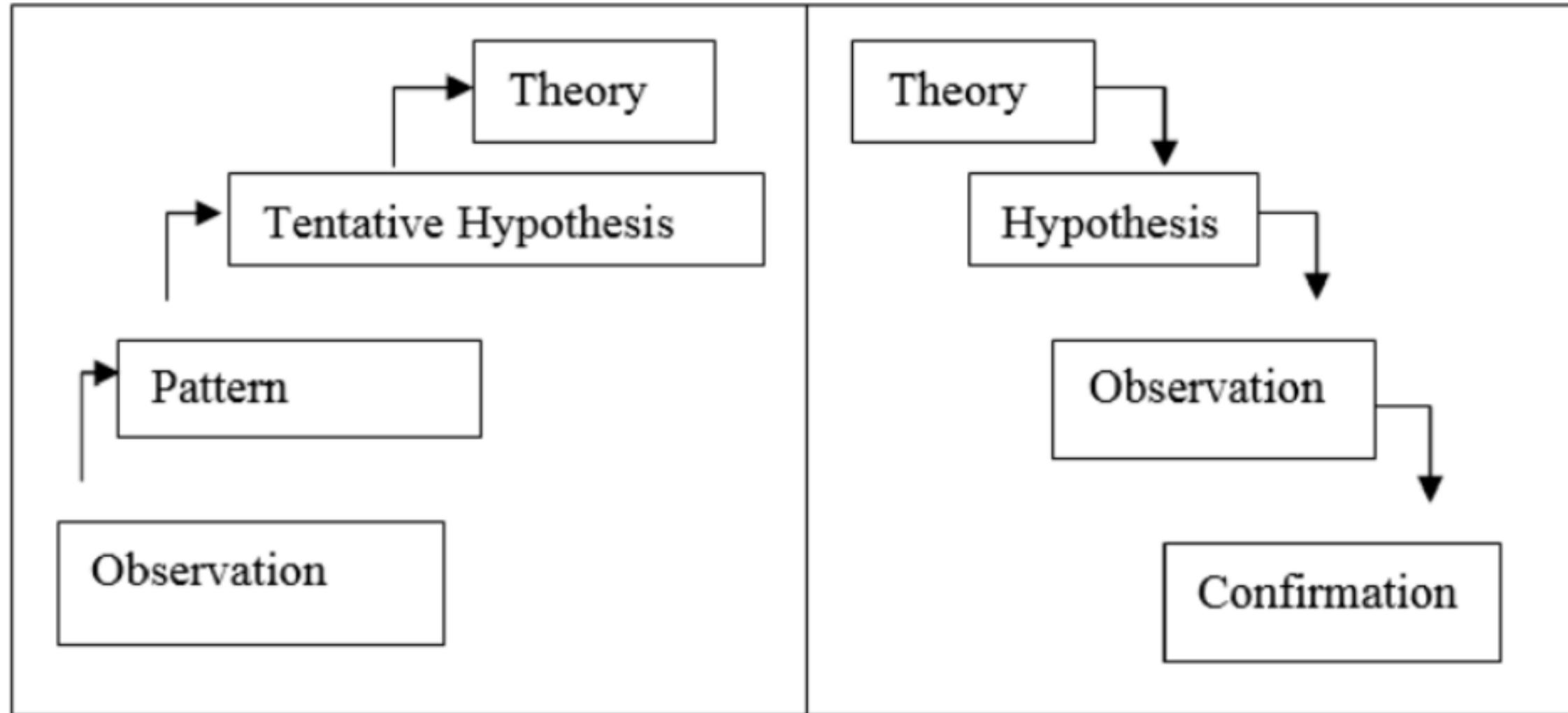
► Descriptive Statistics

- **Collecting data** on collective events,
- **Process and organize the collected data,**
- **Present organized data in the form of tables or graphs,**
- And ultimately revealing the **trend of collective** events.

► Inferential Statistics

- Based on the concepts of sampling and sampling fractionation, With the help of a **sample selected randomly from a population.**
- **Estimate the parameters of the population,**
- **Testing hypotheses about population parameters,**
- **And making predictions for the future....**
- Since these inferences cannot be absolutely true, they are expressed with the word "**probability**".
- In inferential statistics, the **inductive method** is usually used.

Inductive and Deductive Analyses Algorithms



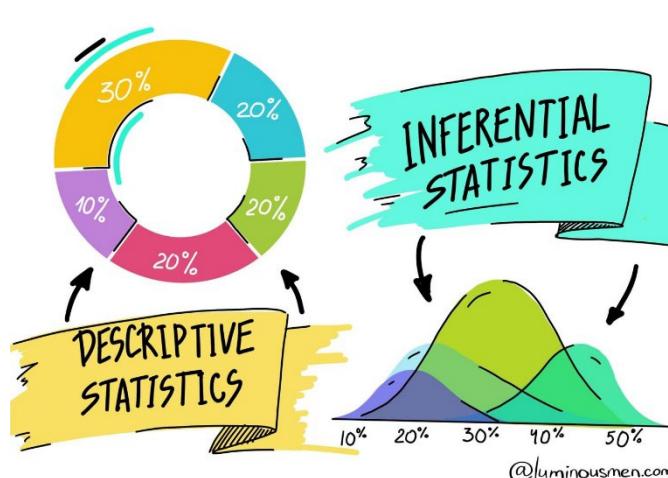
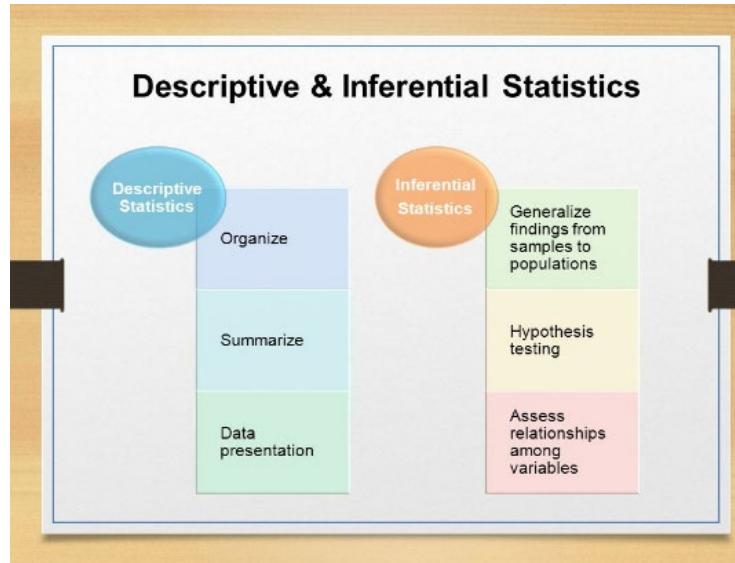
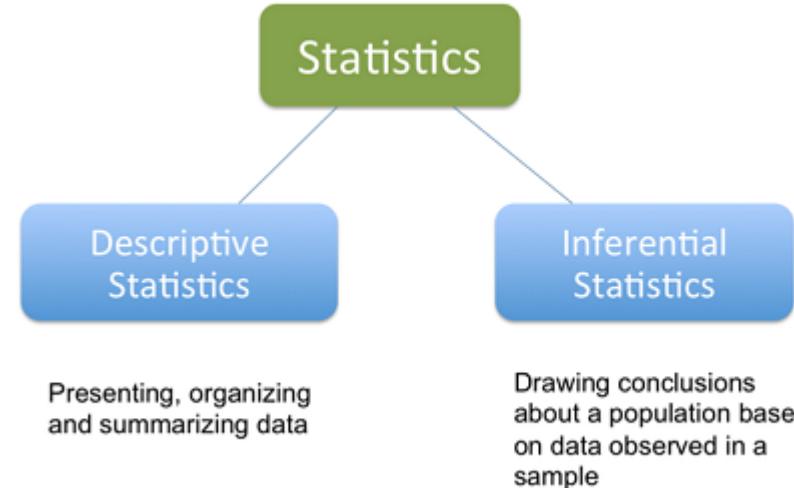
Inductive Reasoning

Deductive Reasoning

5. Statistics Types

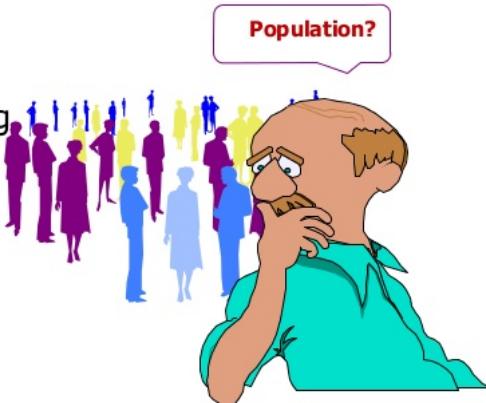


5. Descriptive vs Inferential Statistics



INFERENTIAL STATISTICS

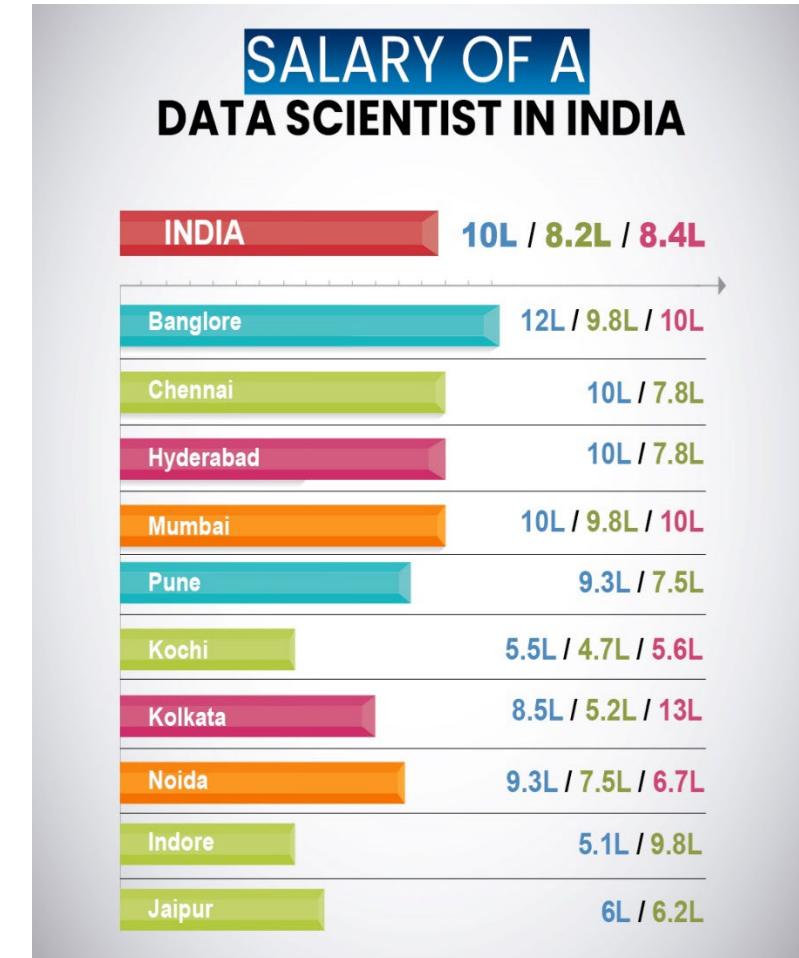
1. Involves
 - Estimation
 - Hypothesis Testing
2. Purpose
 - Make Decisions About Population Characteristics



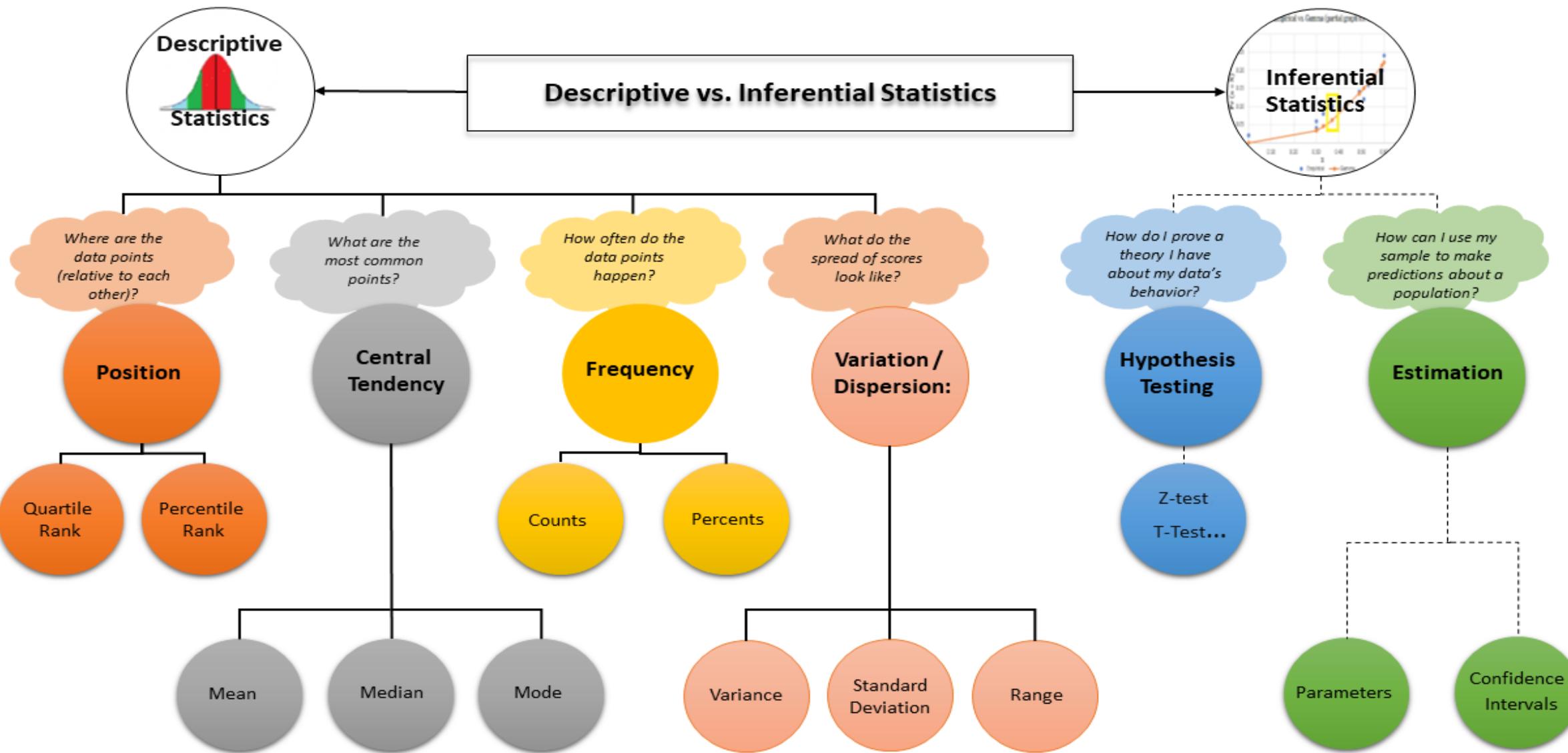
<https://www.kaggle.com/code/yashvi/practical-statistics-1-descriptive-statistics>

Why do we need Inferential Statistics?

- What is the average salary of a Data Scientist in India?
- We can use inferential statistics since collecting all data is hard/expensive.**



Descriptive vs Inferential Statictics



Descriptive vs Inferential Statictics

Examples of Descriptive and Inferential Statistics

Descriptive Statistics

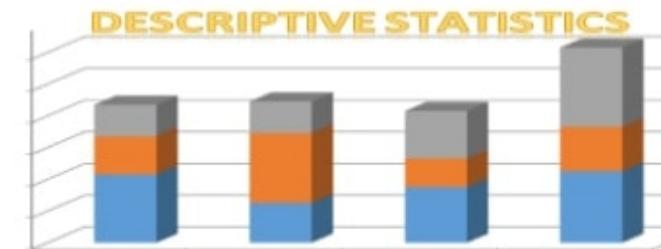
- Graphical
 - Arrange data in tables
 - Bar graphs and pie charts
- Numerical
 - Percentages
 - Averages
 - Range
- Relationships
 - Correlation coefficient
 - Regression analysis

Inferential Statistics

- Confidence interval
- Margin of error
- Compare means of two samples
 - Pre/post scores
 - t Test
- Compare means from three samples
 - Pre/post and follow-up
 - ANOVA = analysis of variance

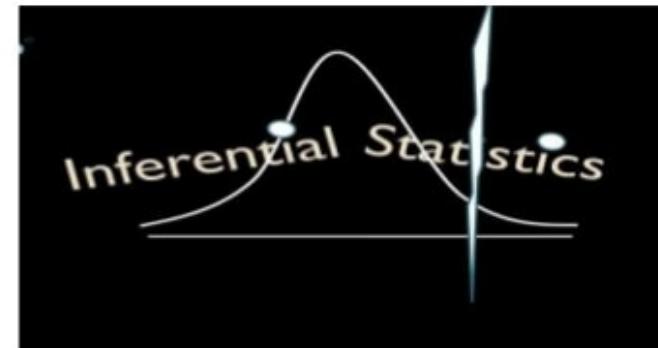
Descriptive Statistics

Descriptive statistics are numbers that are used to summarize and describe data. The word "data" refers to the information that has been collected from an experiment, a survey, a historical record, etc.



Inferential Statistics

The sample is a set of data taken from the population to represent the population. Probability distributions, hypothesis testing, correlation testing and regression analysis all fall under the category of **inferential statistics**.



<https://www.youtube.com/watch?v=MUyUaouisZE>

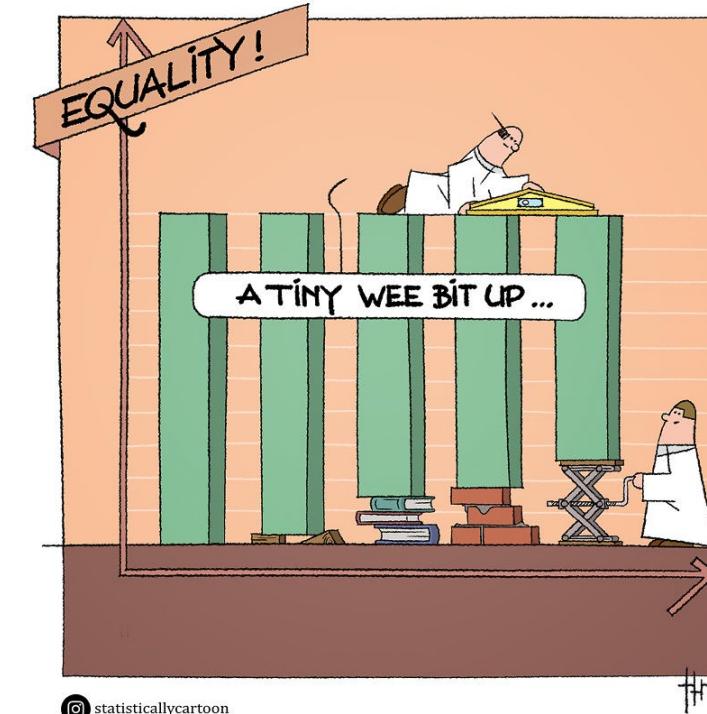
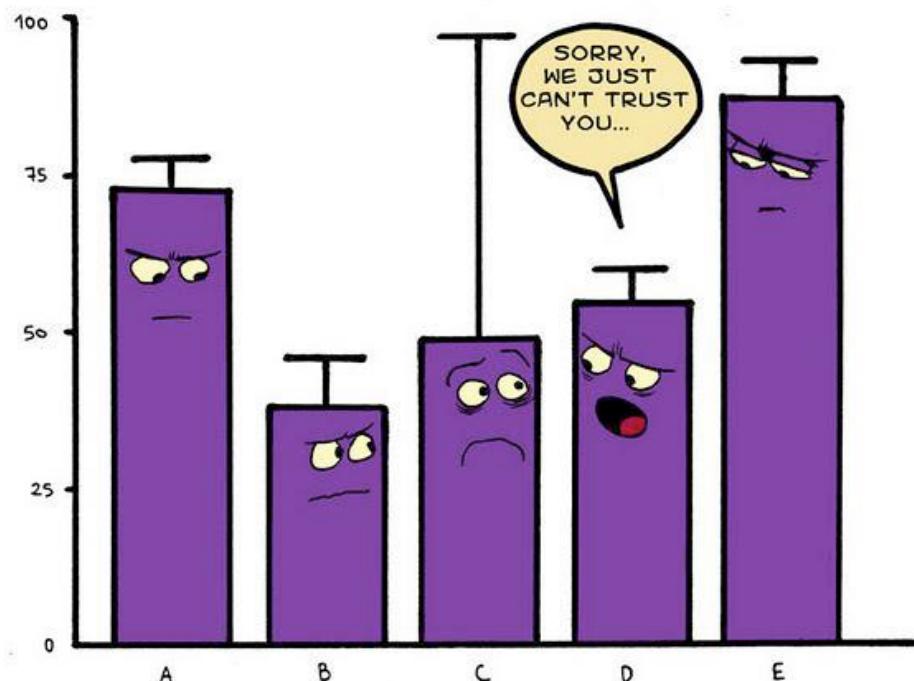
QUESTION

Which is descriptive statistics?

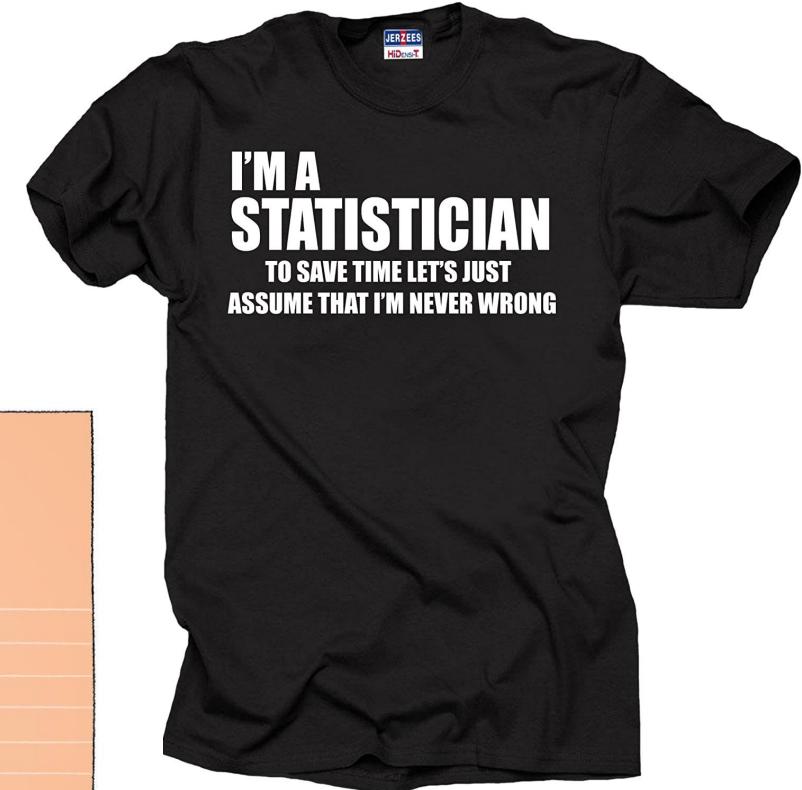
A survey for
next
election?

Average age
of DS-Batch
250 students
at TechProEd

Sometimes funny content



@ statisticallycartoon



6. Data Types

■ What is Data?

Data refers to **information**, often facts or **figures**, **amounts**, **labels** (groups, kinds), **colours**, shapes, **images**, **videos**, **voices**, and **text**.

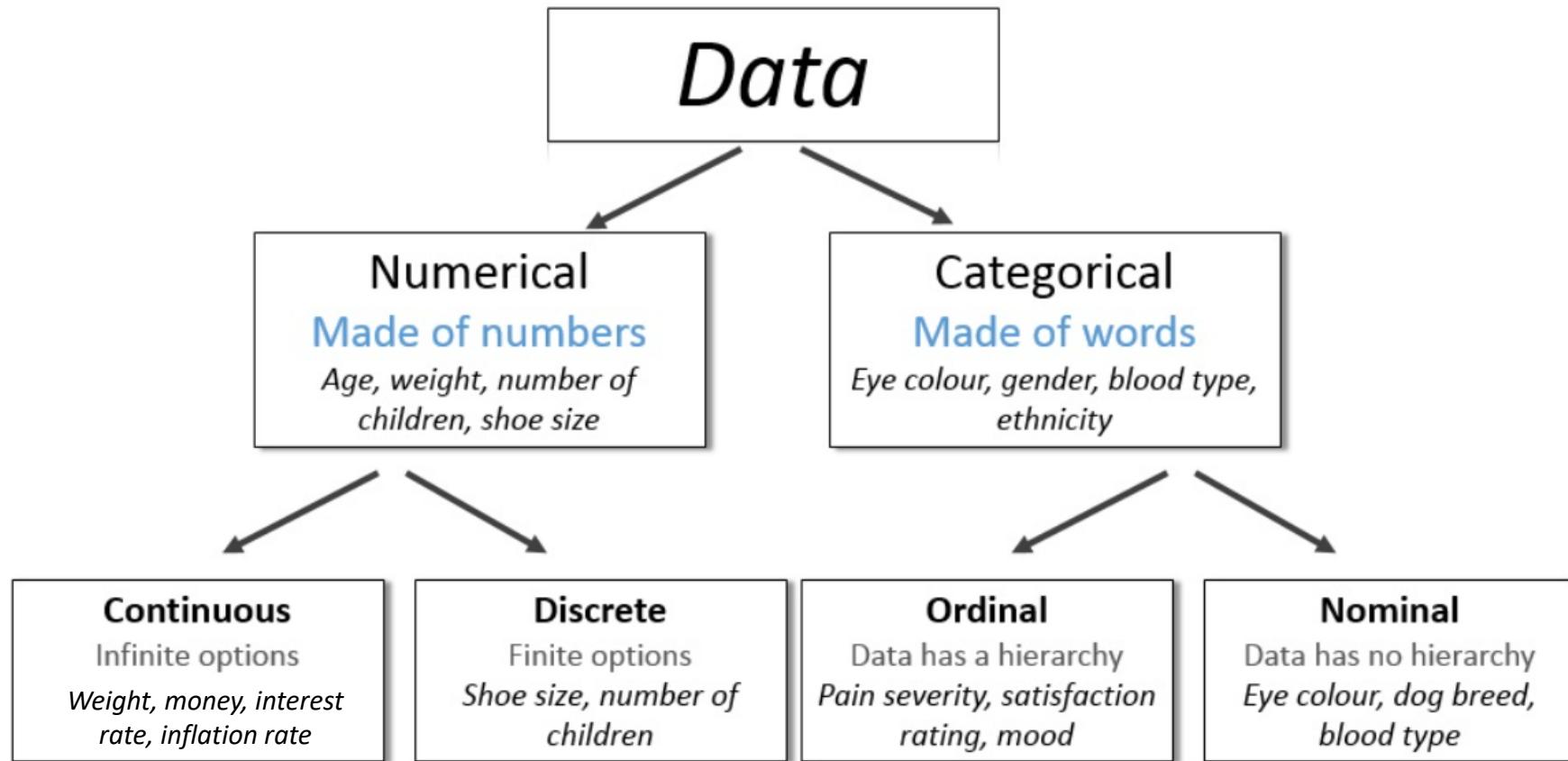
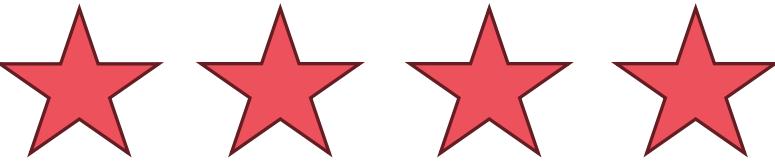
Data is everything.

Posts, messages, emails, shopping information, temperature, vote, prices, inflation, interest rate, GDP, population, ...

Fred's Classic Movie Events					
A	B	C	D	E	F
1					
2	Venue	Fred's Movie Emporium		Today	13/12/2012
3	Capacity	150			
4					
5					
Date	Movie	Tickets Sold	% of Capacity Sold	Tickets Remaining	% of Tickets to Sell
Wednesday 11 May 2011	Grease	50	33%	100	67%
Sunday 15 May 2011	Jaws	150	100%	0	0%
Monday 23 May 2011	Citizen Kane	105	70%	45	30%
Wednesday 01 Jun 2011	The Wizard of Oz	150	100%	0	0%
Friday 10 Jun 2011	Singin' In The Rain	85	57%	65	43%
11					
12					

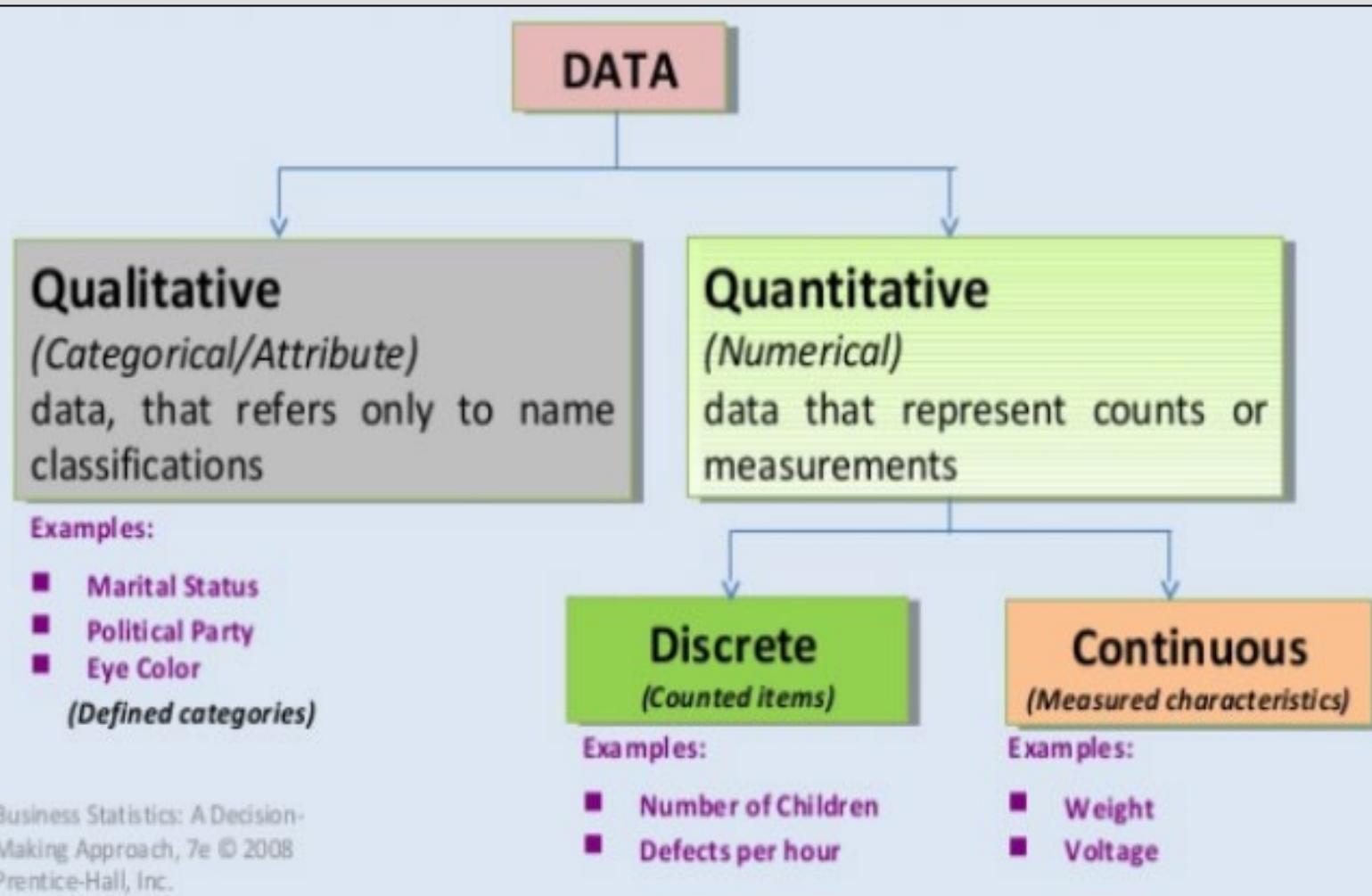


6. Data Types



<https://www.youtube.com/watch?v=pgI2U1BAnoA>

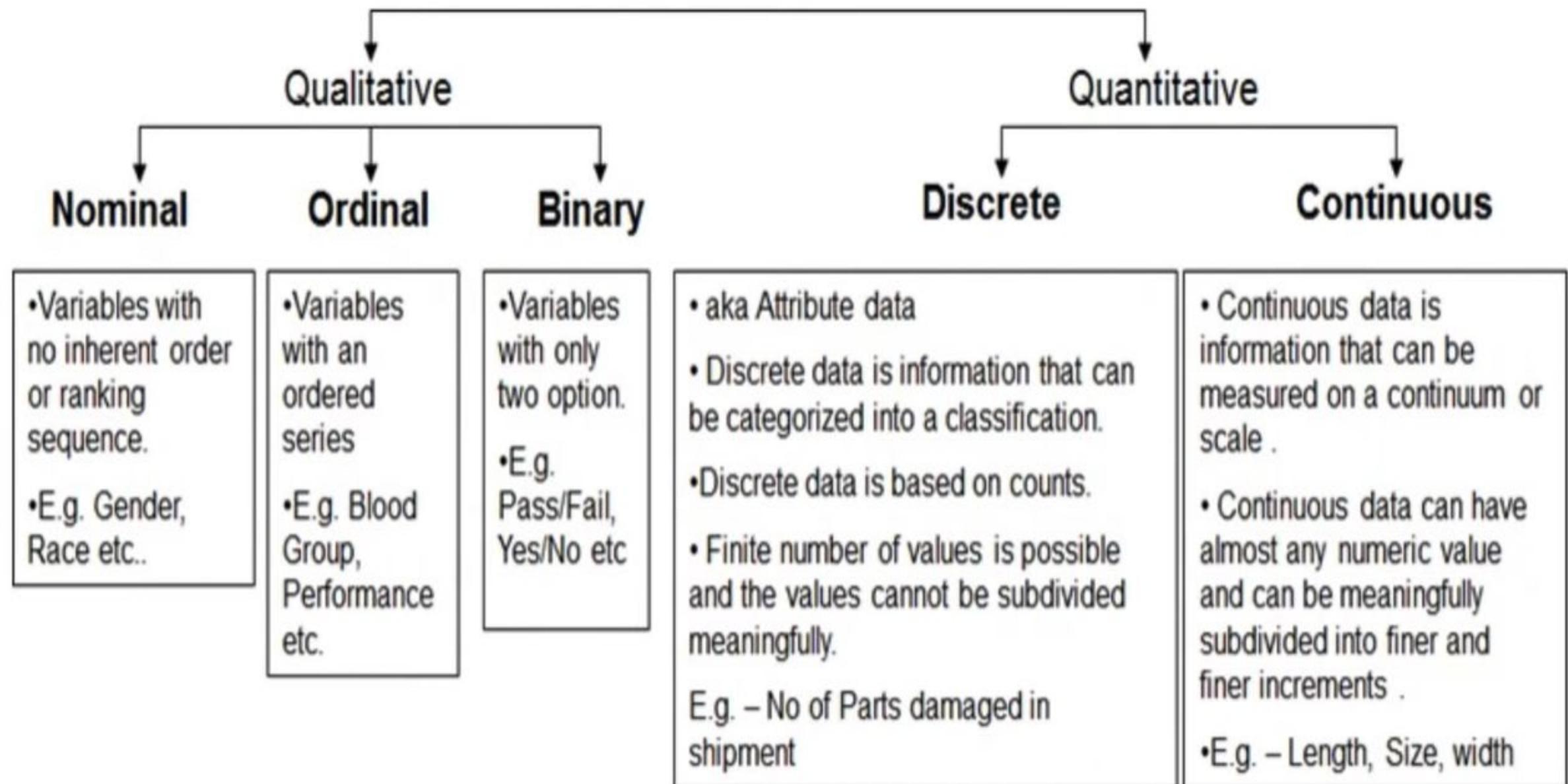
Data Types



If you see a job position named
“Quantitative Analysis/Modelling”,
jump on it. That person is you.

Jobs

- Past 3 days Full-time Work from home Data analyst
- 
Specialist, Data Management & Quantitative Analysis
The Bank Of New York Mellon Corporation
Manchester
via Sercanto
🕒 3 days ago Full-time
- 
Head Of Quantitative Modelling
Endeavor
London
via ShowbizJobs
🕒 21 days ago Full-time
- 
6 Head of Quantitative Modelling
682 OpenBet Limited
Anywhere
via Workday
🕒 21 days ago Work from home Full-time



Top 60 Statistics Interview Questions 2024



Question 1: What are quantitative data and qualitative data?

Answer: Qualitative data is used to describe the characteristics of data and is also known as Categorical data. For example, how many types.

Quantitative data is a measure of numerical values or counts. For example, how much or how often. It is also known as Numeric data.

QUESTION

Which is continuous data ?

Race

Height

QUESTION

Which is discrete data ?

Population

**Exchange
Rate**

QUESTION

Which is ordinal data ?

**Satisfaction
Level (e.g.
Low, High)**

Bill Amount

QUESTION

Which is nominal data ?

Days of week

**Corporate
Hierarchy**

ML Models Content as Categorical Data

Label Encoder

Raw Data

- We encode the country before exerting it to a ML model

Ready for ML/DL Model

Encoding

- Prepared for ML model

ID	Country	Population
1	Japan	127185332
2	U.S	326766748
3	India	1354051854
4	China	1415045928
5	U.S	326766748
6	India	1354051854

ID	Country	Population
1	0	127185332
2	1	326766748
3	2	1354051854
4	3	1415045928
5	1	326766748
6	2	1354051854

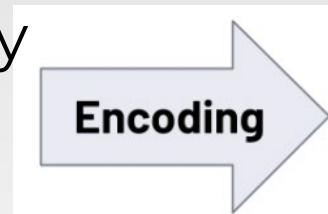
ML Models Content as Categorical Data

One Hot Encoder

Raw Data

- Performing one hot encoding on the country column

ID	Country
1	Japan
2	U.S
3	India
4	China
5	U.S
6	India



Ready for Model (ML Model)

- We have **n new** dummy variables (columns) in OneHot Encoder.

ID	Country_Japan	Country_U.S	Country_India	Country_China
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
5	0	1	0	0
6	0	0	1	0

ML Models Content as Categorical Data

Dummy Encoder

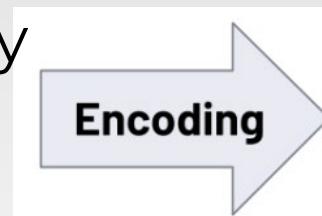
Raw Data

- Performing dummy encoding on the country column

ID	Country
1	Japan
2	U.S
3	India
4	China
5	U.S
6	India

Ready for Model (ML Model)

- We have **n-1** new dummy variables at Dummy Encoder.



ID	Country_Japan	Country_U.S	Country_India
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0
5	0	1	0
6	0	0	1

ML Models Content as Categorical Data

Ordinal Encoding



Ordinal Encoding	
Grades	
A	
B	
C	
D	
Fail	

→

Grades	Encoded
A	4
B	3
C	2
D	1
Fail	0



Original Encoding	Ordinal Encoding
Poor	1
Good	2
Very Good	3
Excellent	4

Student	Grade
John	A
Mike	B
Elena	A
Adam	D

→ Encoding →

Student	Grade
John	4
Mike	3
Elena	4
Adam	1

<https://www.youtube.com/watch?v=15uClAVV-rl>

Interview Question from me



Question: What is the difference of Label Encoder and Ordinal Encoder?



Answer: While Label Encoder converts categorical data into numeric and gives numbers from smallest to largest, starting from the alphabetically preceding category, Ordinal Encoder assigns numbers according to the order of importance given by the user.

Interview Question from me



Question: What is the difference of One Hot Encoder and Dummy Encoder?

Answer: One Hot Encoding creates binary columns for each category, representing the presence or absence of that category. Dummy Encoding is a specific case of One Hot Encoding where one category is implicitly represented (by dropping its corresponding column) to avoid redundancy in model fitting.



Interview Question from me

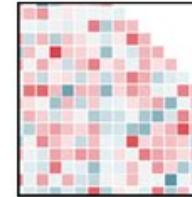
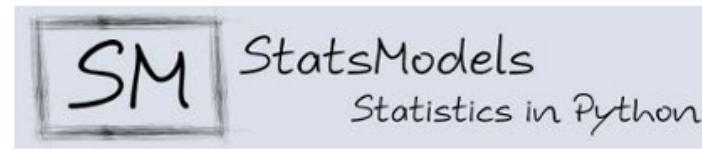


Question: Which method(s) is suitable for “Gender”?

- A. Label Encoding
- B. Ordinal Encoding
- C. One Hot Encoding
- D. Dummy Encoding

LABEL ENCODING
VS
ORDINAL ENCODING

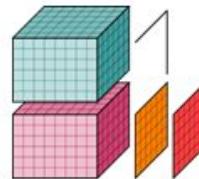
Hands-on Practise Time



Seaborn

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



xarray



**scikit
learn**



scikit-image
image processing in python



matplotlib



IP[y]:
IPython





<https://www.youtube.com/watch?v=bBRKYdkitbU>

<https://www.online-stopwatch.com/timer/10minutes/>

<https://www.youtube.com/watch?v=bBRKYdkitbU>

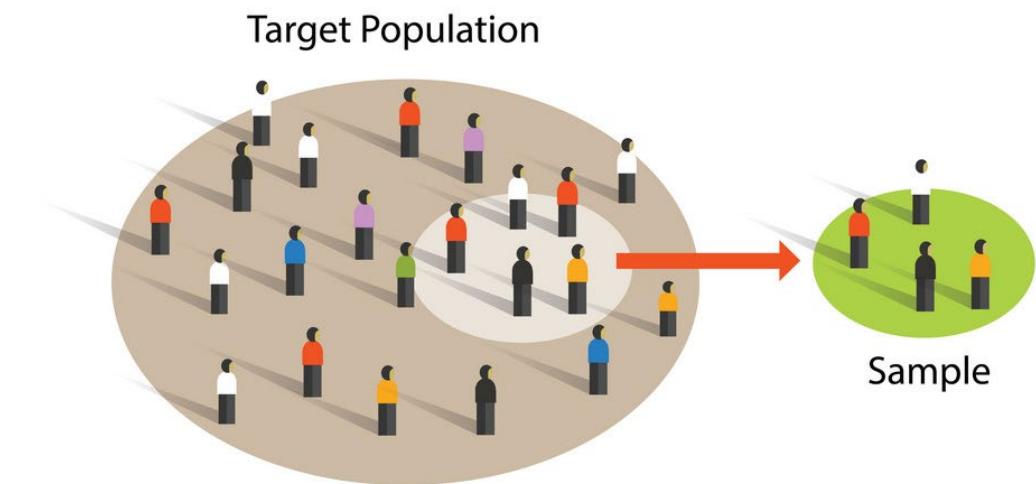
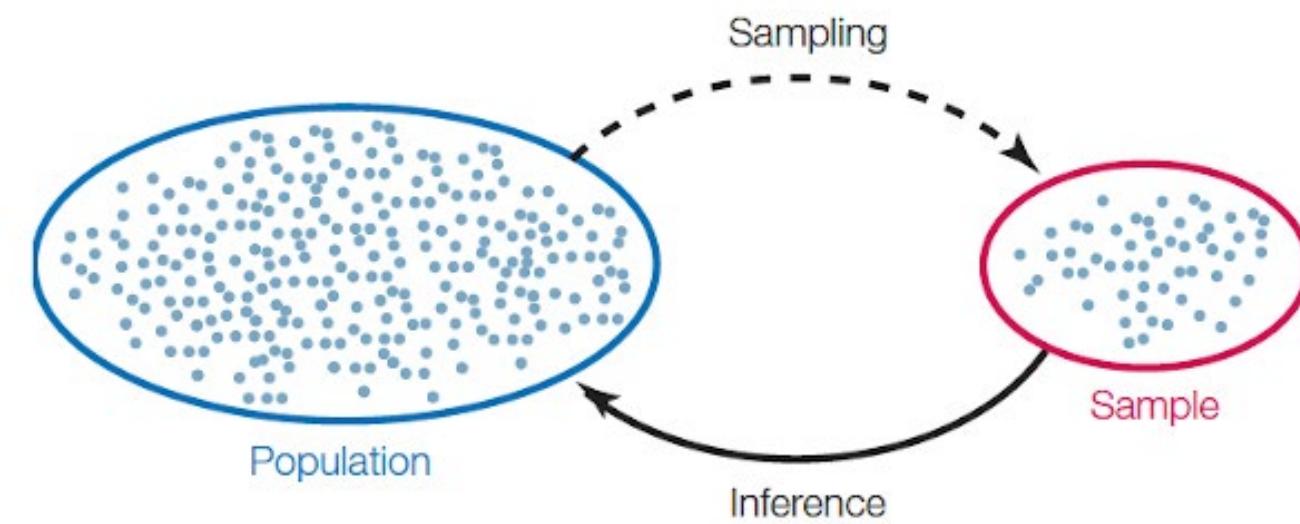
Tea break...

10:00



Start Stop Reset mins: secs: type:

7. Parameters and Statistics

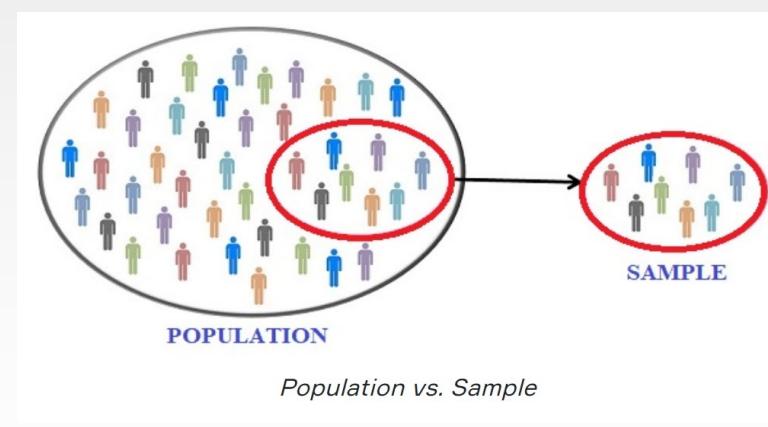


7. Populations & Samples

► **Population**



► **Sample**



!!! We observe samples but we are interested in populations.

7. Parameters & Statistics

Population
Attributes



Parameters

Sample Attributes

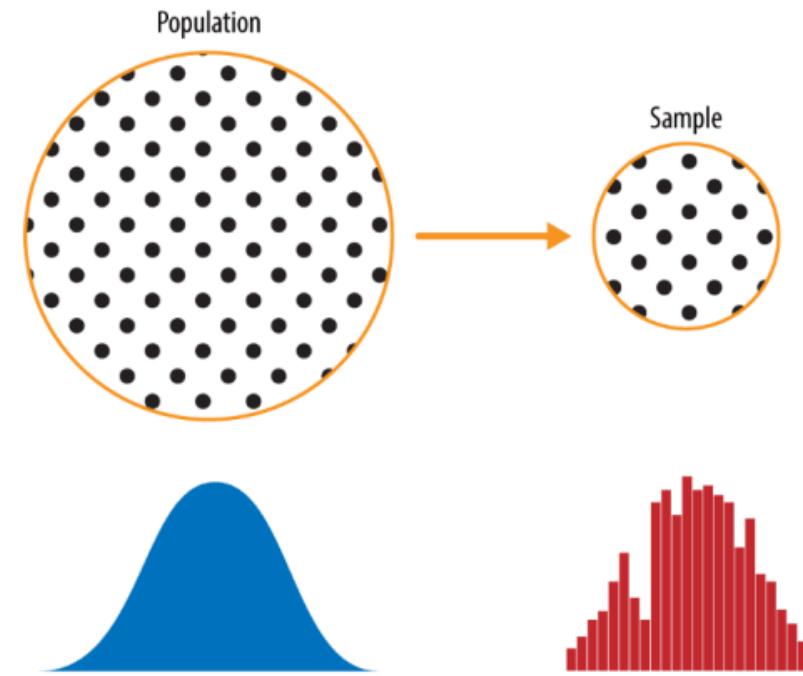
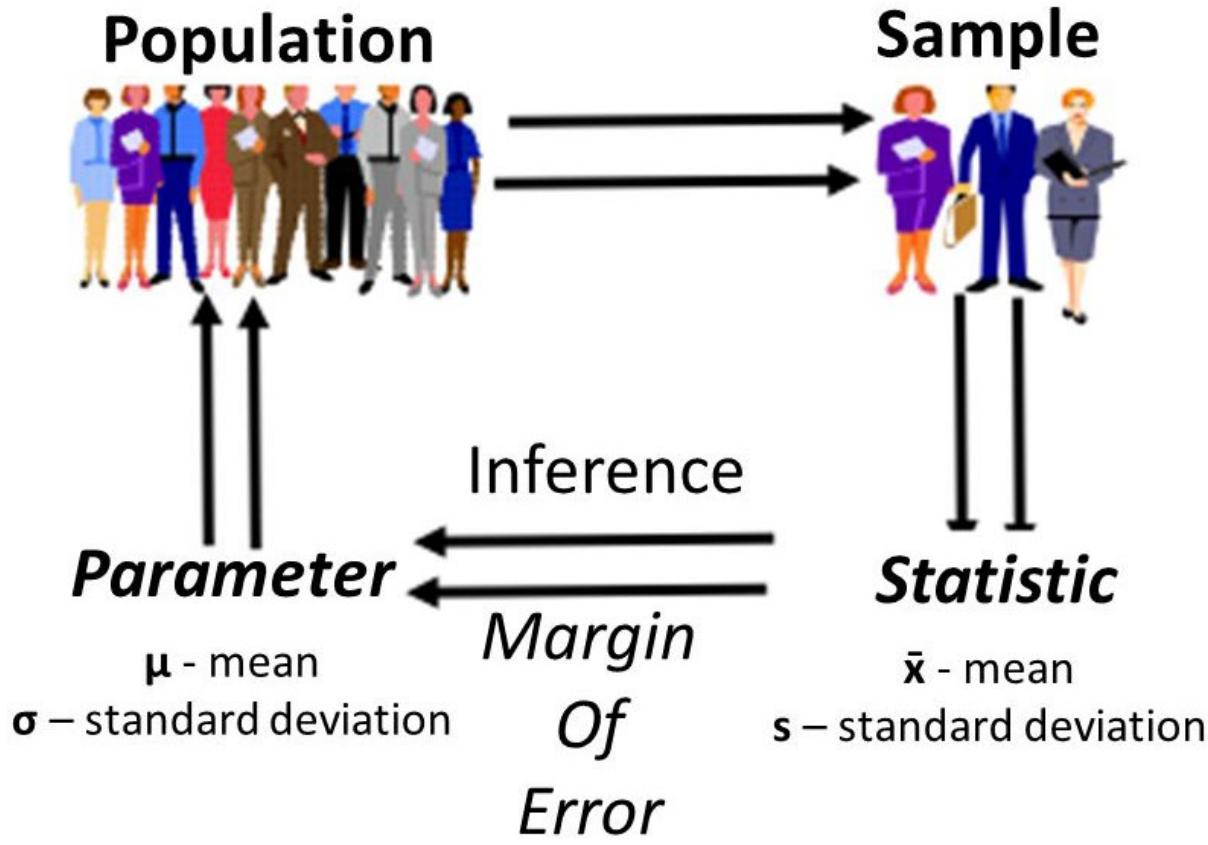


Statistics

A parameter is a numerical summary of the population and a statistic is a numerical summary of a sample.

7. Parameters and Statistics

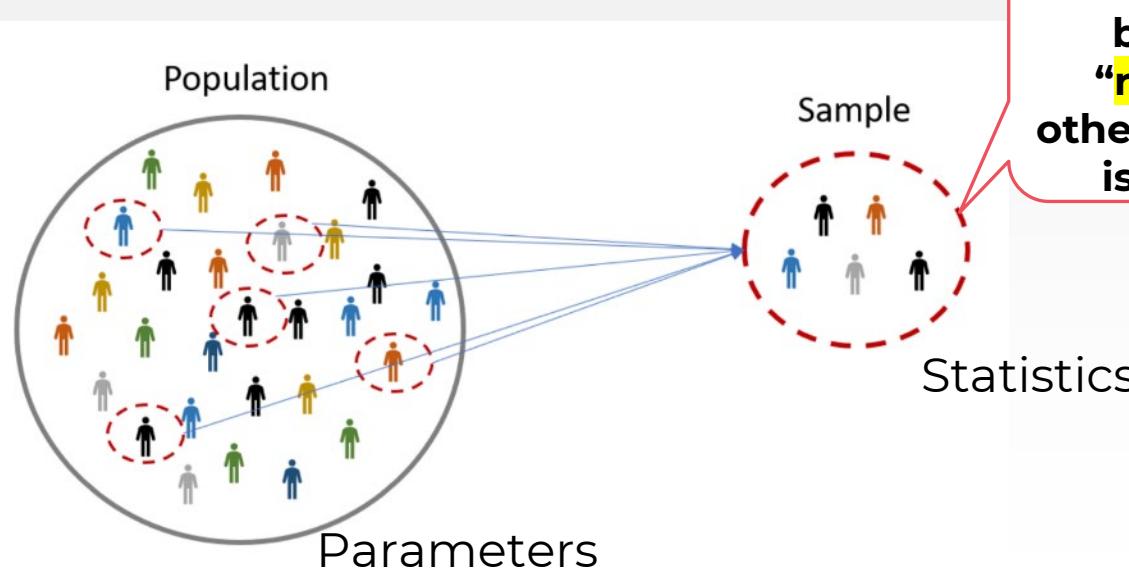
The Basic Paradigm



- Populations have Parameters (like μ , σ^2 , θ , p)
- Samples have Statistics, functions of observed data, like \bar{x} , \tilde{x} , s^2 , $\hat{\theta}$, \hat{p}

7. Parameters and Statistics

- A parameter is a numerical description of the characteristics of the population.
- A statistic is a numerical description of the characteristics of the sample.



The sample should be chosen “randomly”; otherwise “, bias” is revealed.

Sample (statistic):

- Let us decide whether the given value describes a population parameter or a sample statistic.
 - a.) A survey of a sample of 450 college students reports that the average weekly income for students is \$325.
 - Since the average of \$325 is based on a sample, this is a sample statistic.
 - b.) The average weekly income for all students is 405 dollars.
 - Since the average of \$405 is based on a population, this is a population parameter.

Top 60 Statistics Interview Questions 2024



Question 2: What is Random Sampling? Give some examples of some random sampling techniques.

Answer: Random sampling is a sampling method in which each sample has an equal probability of being chosen as a sample. It is also known as probability sampling.

Top 60 Statistics Interview Questions 2024



Question 3: What is the difference between population and sample in inferential statistics?

Answer: A **population** in inferential statistics refers to the entire group we take samples from and are used to draw conclusions. A **sample**, on the other hand, is a specific group we take data from population. This data is used to calculate the statistics. **Sample size is always less than that of the population.**

8. Probability vs Statistics

▶ Probability

- In **probability**, the **process** of randomization is assumed or **known**, and used to make an inference about the result.

Example: Given a proper or error-free coin (**PROCESS is known**), **what is the probability that 7 out of 10-coin tosses will land heads** (We are trying to estimate the **RESULT**)?

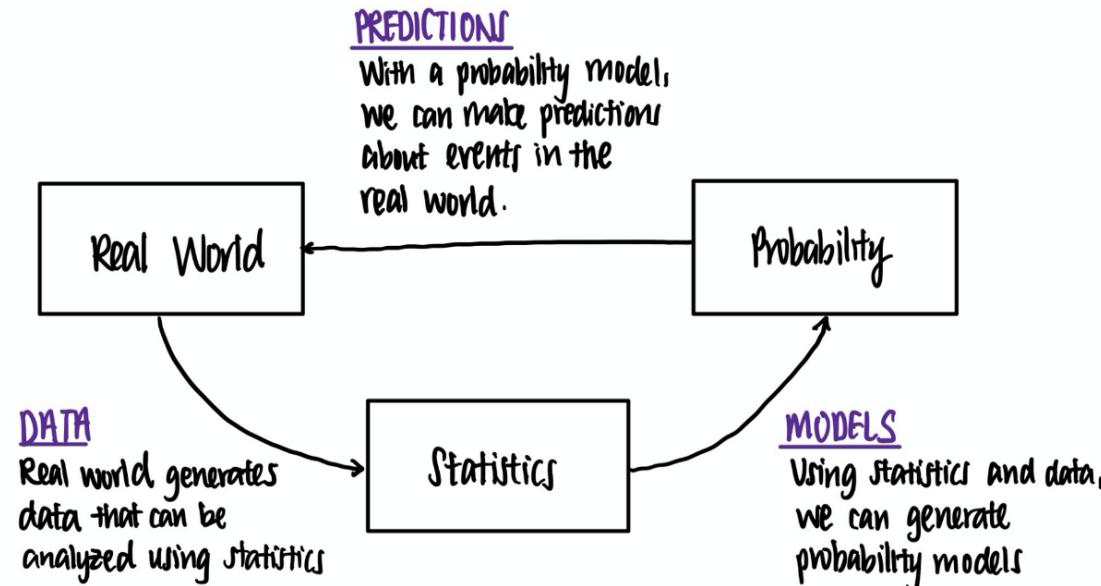
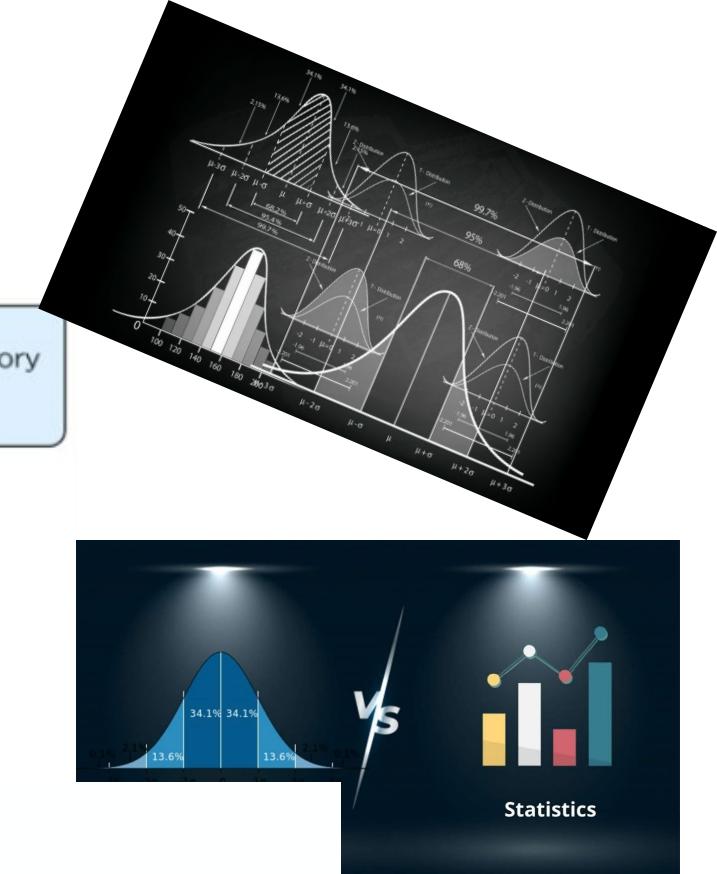
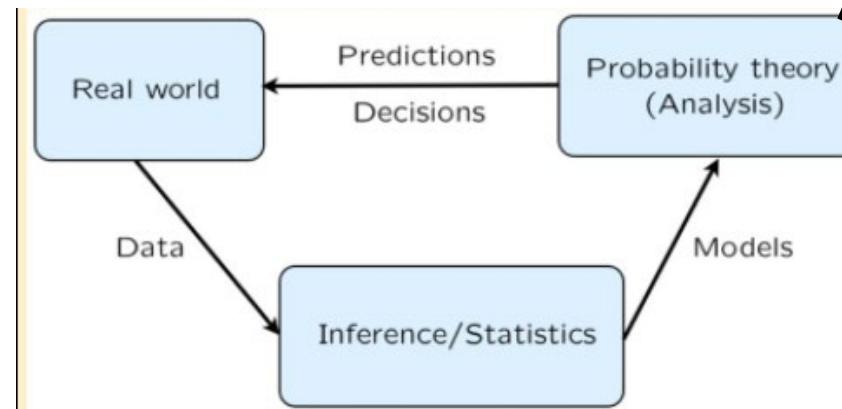
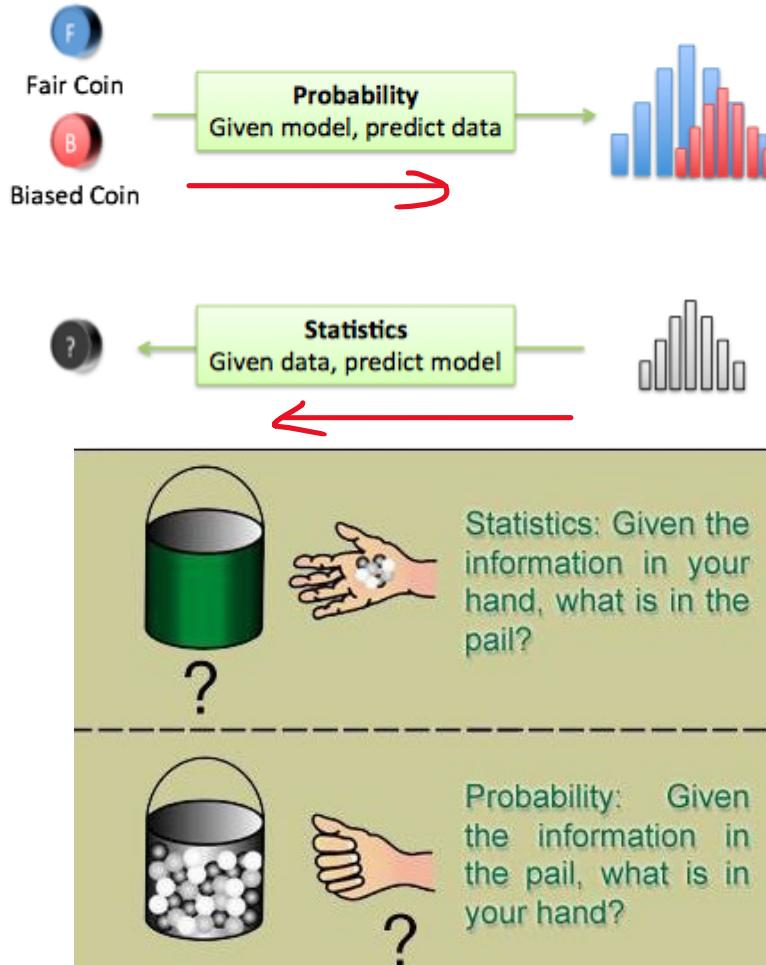
▶ Statistics

- In **statistics**, the **result** is **known** and used to make an inference about the process.

Example: Since I just observed 7 out of 10-coin flips (**RESULT is known**), **are the coin flip results reasonable** (We are trying to understand process)?

8. Probability vs. Statistics

Probability & Statistics



9. Level of Measurement

- In the 1940s, Stanley Smith Stevens introduced four levels of measurement
 - scales: nominal, ordinal, interval and ratio.
- These are still widely used today as a way of describing the characteristics of a variable. Knowing the level-scale of measurement of a variable is an important consideration in choosing the right statistical analysis.

LEVEL OF MEASUREMENT

TYPE	MEASURE PROPERTY
NOMINAL	Classification, membership = , ≠
ORDINAL	Comparison Level > , <
INTERVAL	Difference , closeness + , -
RATIO	Magnitude, amount *, /

9. Level of Measurement

NOMINAL DATA

Nominal data divides variables into mutually exclusive, labeled categories.

Examples



How is nominal data analyzed?

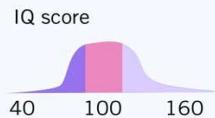
Descriptive statistics:
Frequency distribution and mode

Non-parametric statistical tests

INTERVAL DATA

Interval data is measured along a numerical scale that has equal intervals between adjacent values.

Examples



How is interval data analyzed?

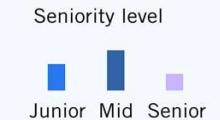
Descriptive statistics: Frequency distribution; mode, median, and mean; range, standard deviation, and variance

Parametric statistical tests (e.g. t-test, linear regression)

ORDINAL DATA

Ordinal data classifies variables into categories which have a natural order or rank.

Examples



How is ordinal data analyzed?

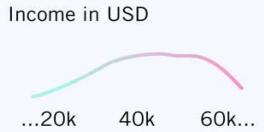
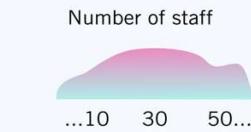
Descriptive statistics:
Frequency distribution, mode, median, and range

Non-parametric statistical tests

RATIO DATA

Ratio data is measured along a numerical scale that has equal distances between adjacent values, and a true zero.

Examples

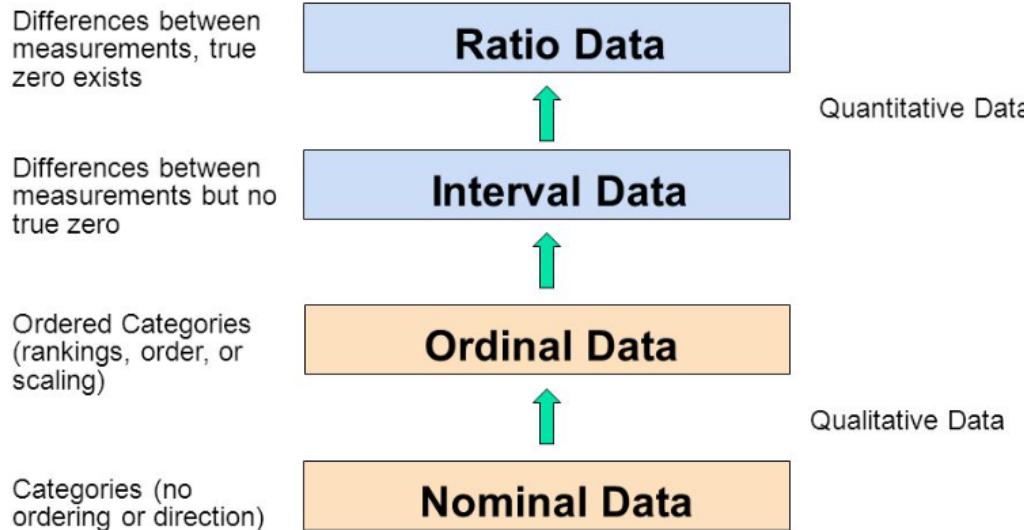
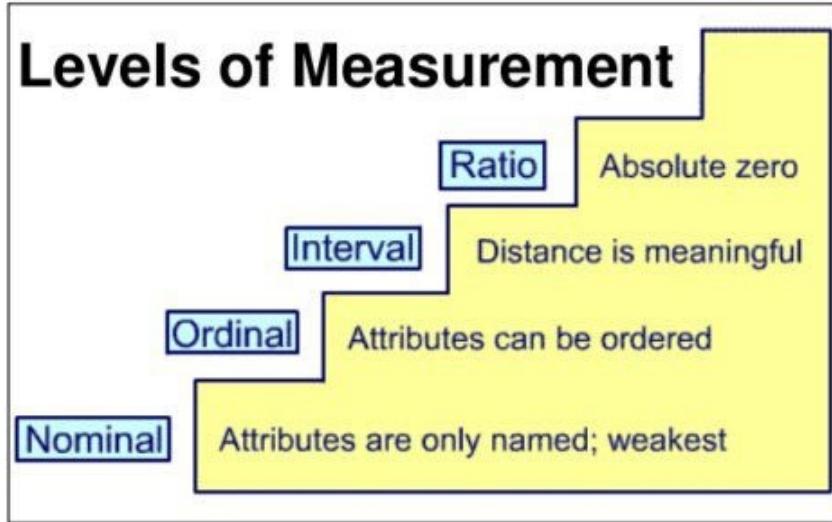


How is ratio data analyzed?

Descriptive statistics: Frequency distribution; mode, median, and mean; range, standard deviation, variance, and coefficient of variation

Parametric statistical tests (e.g. ANOVA, linear regression)

9. Level of Measurement



THE FOUR LEVELS OF MEASUREMENT:

Nominal	Ordinal	Interval	Ratio
Categorizes and labels variables	✓	✓	✓
Ranks categories in order	✓	✓	✓
Has known, equal intervals		✓	✓
Has a true or meaningful zero			✓

Levels of Measurement

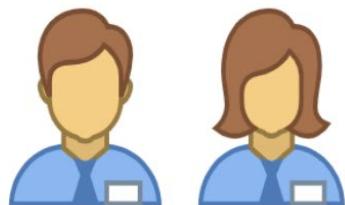
Nominal	Ordinal	Interval	Ratio
"Eye color"	"Level of satisfaction"	"Temperature"	"Height"
Named	Named	Named	Named
	Natural order	Natural order	Natural order
	Equal interval between variables	Equal interval between variables	Has a "true zero" value, thus ratio between values can be calculated

9. Level of Measurement

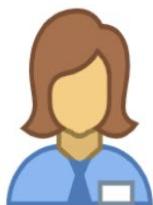
▶ Nominal Level o Measurement

Also called classificatory level of measurement

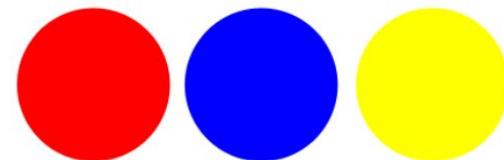
- No ranking
- There is categorization
- There is belonging



MALE



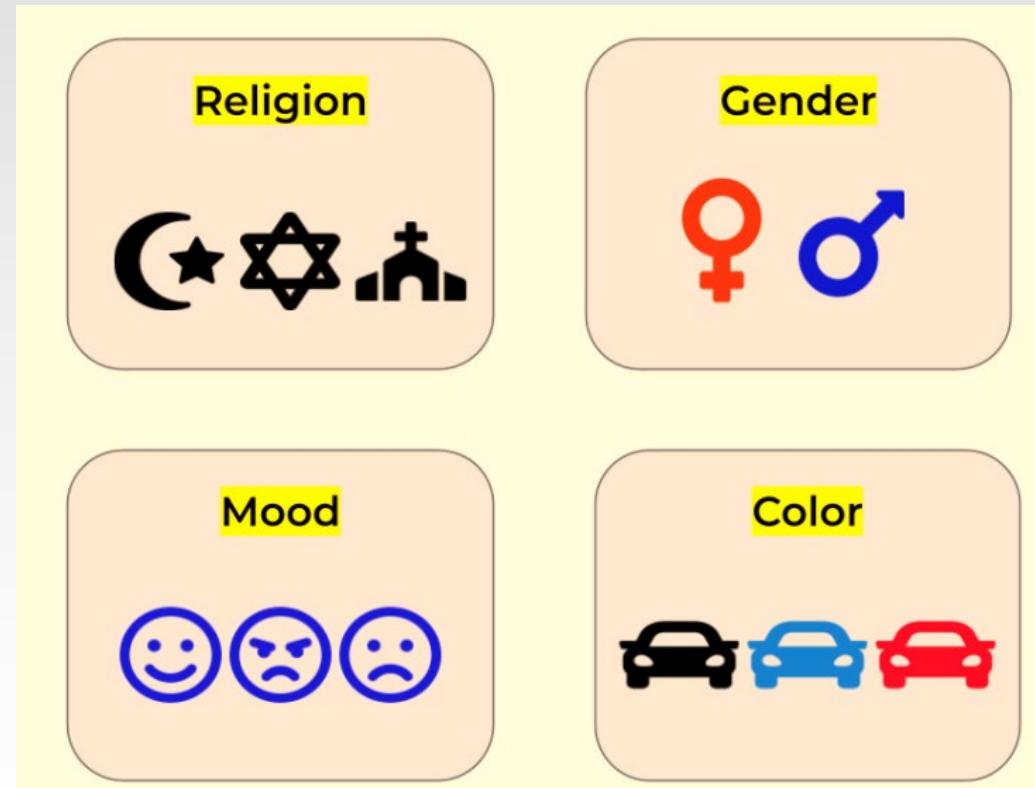
FEMALE



RED

BLUE

YELLOW



9. Level of Measurement

NOMINAL DATA

Nominal data divides variables into mutually exclusive, labeled categories.

Examples

Eye color



Smartphone



Transport



How is nominal data analyzed?

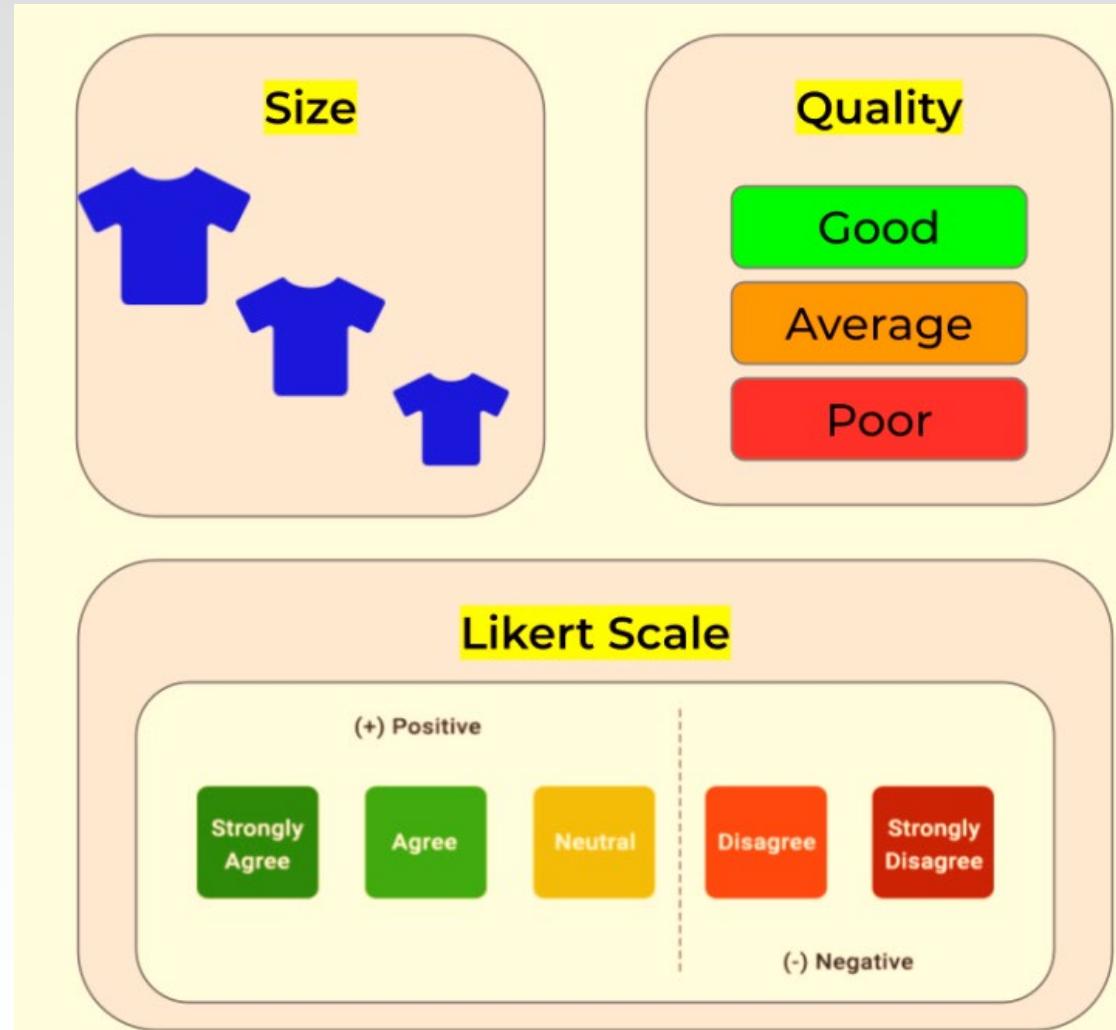
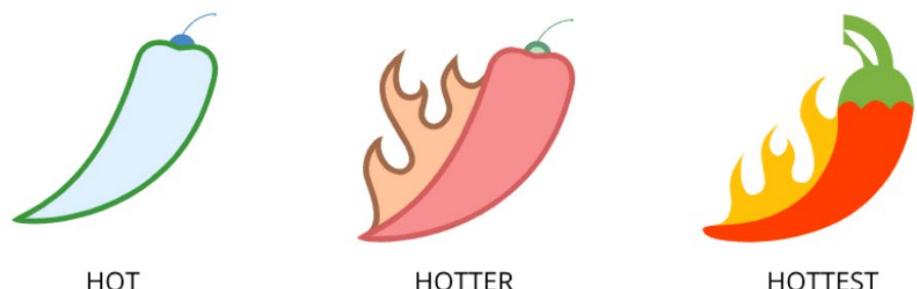
Descriptive statistics:
Frequency distribution
and mode

Non-parametric statistical tests

9. Level of Measurement

▶ Ordinal Level of Measurement

- There is a ranking
- No real value
- Comparison available



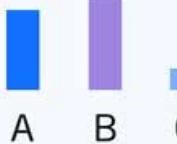
9. Level of Measurement

ORDINAL DATA

Ordinal data classifies variables into categories which have a natural order or rank.

Examples

School grades



Education level



Seniority level



How is ordinal data analyzed?

Descriptive statistics:
Frequency distribution,
mode, median, and range

Non-parametric
statistical tests

Types of Measurement Scales

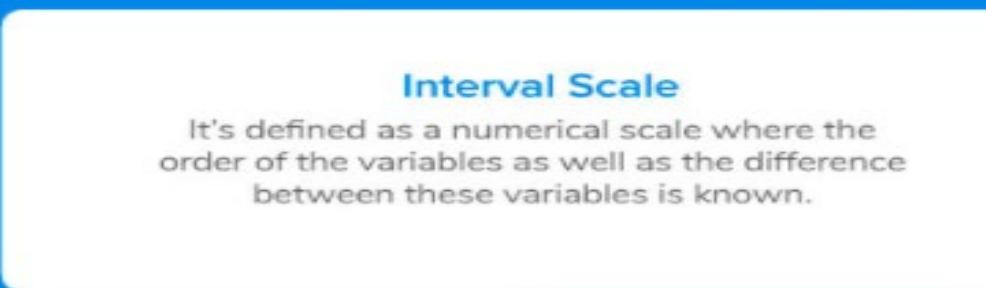
Nominal scale

It's used to label variables in different classifications and does not imply a quantitative value or order.



Ordinal Scale

It's used to represent non-mathematical ideas such as frequency, satisfaction, happiness, a degree of pain, etc.



Ratio Scale

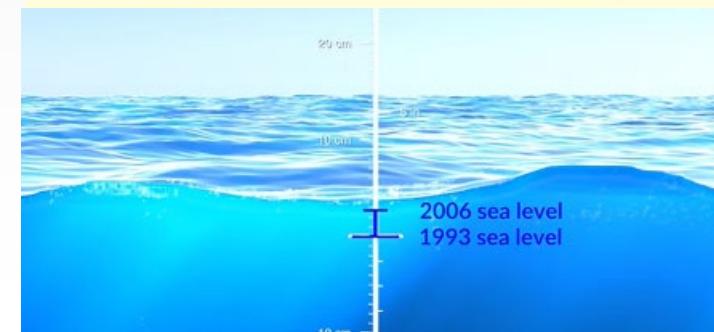
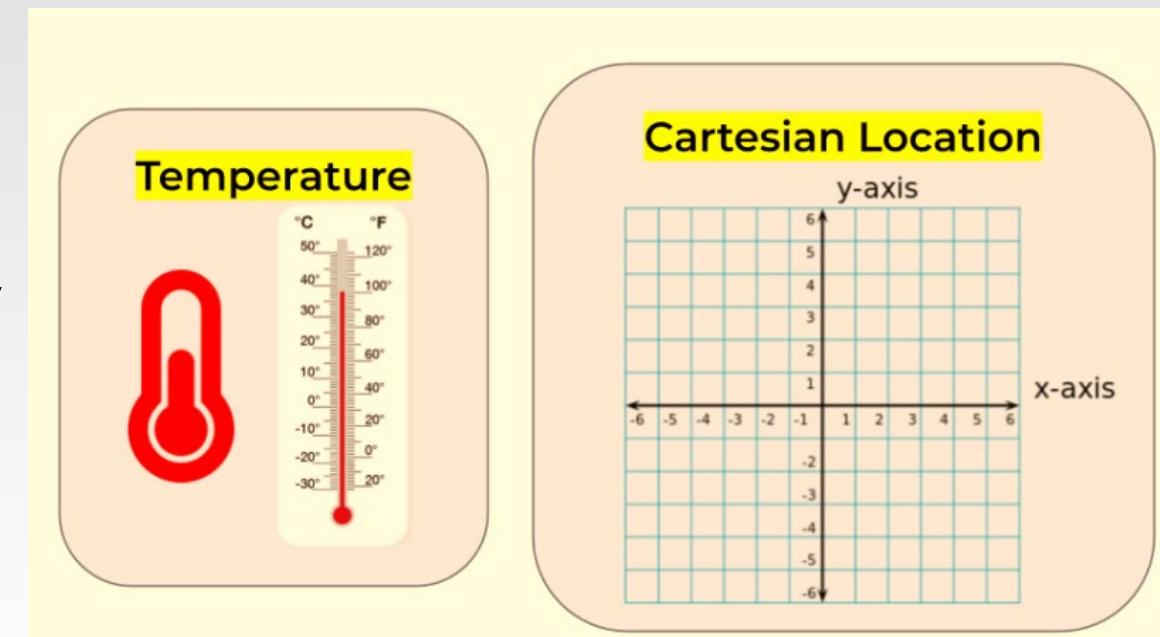
It's a variable measurement scale that not only produces the order of the variables, but also makes the difference between the known variables along with information about the value of the true zero.

9. Level of Measurement

▶ Interval Level o Measurement

Also called **equal interval measurement level**

- Measurable
- There is no true zero, there is **arbitrary zero**
- There are expressions of greatness, but they are not multiples of each other.
- E.g. $20^{\circ}\text{C} \neq 2 * 10^{\circ}\text{C}$
- **Altitude above sea level** $2\text{m} \neq 2 * 1\text{m}$



9. Level of Measurement

INTERVAL DATA

Interval data is measured along a numerical scale that has equal intervals between adjacent values.

Examples

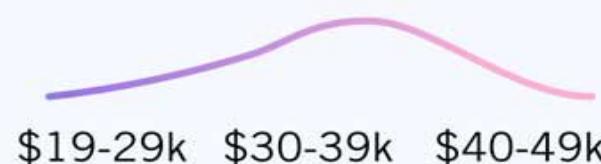
Temperature



IQ score



Income ranges



How is interval data analyzed?

Descriptive statistics: Frequency distribution; mode, median, and mean; range, standard deviation, and variance

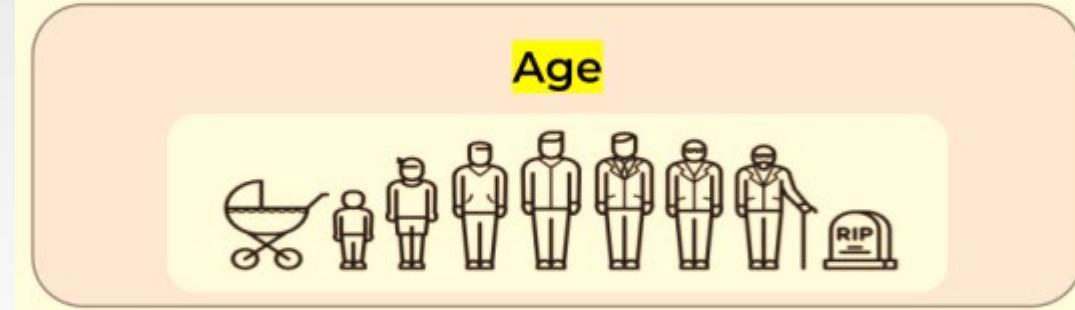
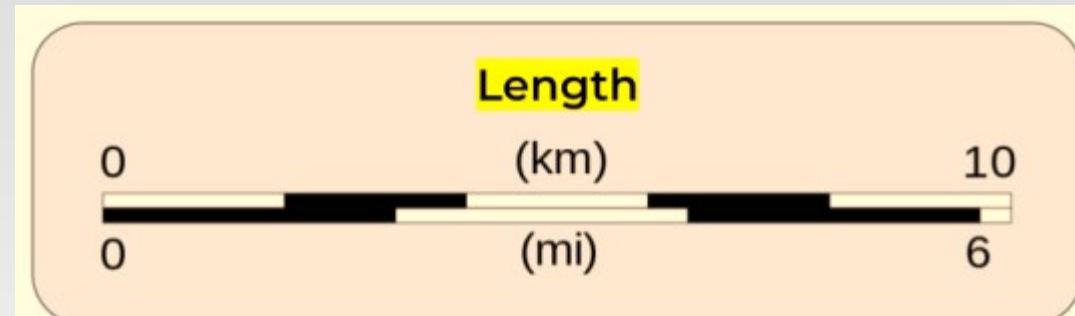
Parametric statistical tests (e.g. t-test, linear regression)

9. Level of Measurement

Ratio Level of Measurement

Also called proportional measurement level

- Measurable
- Has True/Absolute zero
- There is an expression of greatness



9. Level of Measurement

RATIO DATA

Ratio data is measured along a numerical scale that has equal distances between adjacent values, and a true zero.

Examples

Weight in KG



Number of staff



Income in USD



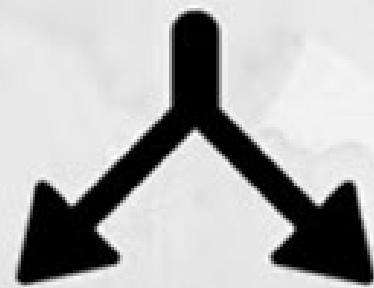
How is ratio data analyzed?

Descriptive statistics: Frequency distribution; mode, median, and mean; range, standard deviation, variance, and coefficient of variation

Parametric statistical tests (e.g. ANOVA, linear regression)

9. Level of Measurement (Operational Level)

Categorical Data



Nominal

No rank or ordering

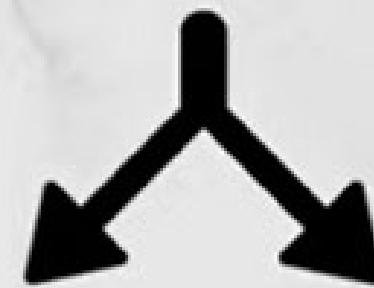
Colours, Days

Ordinal

Natural order or rank

Hot, Hotter, Hottest

Numerical Data



Interval

Measurable, but arbitrary zero

Altitude sea level
IQ level
Magnitude of earthquake

Ratio

Measurable and absolute zero

Income (\$)
Weight, Length

QUESTION

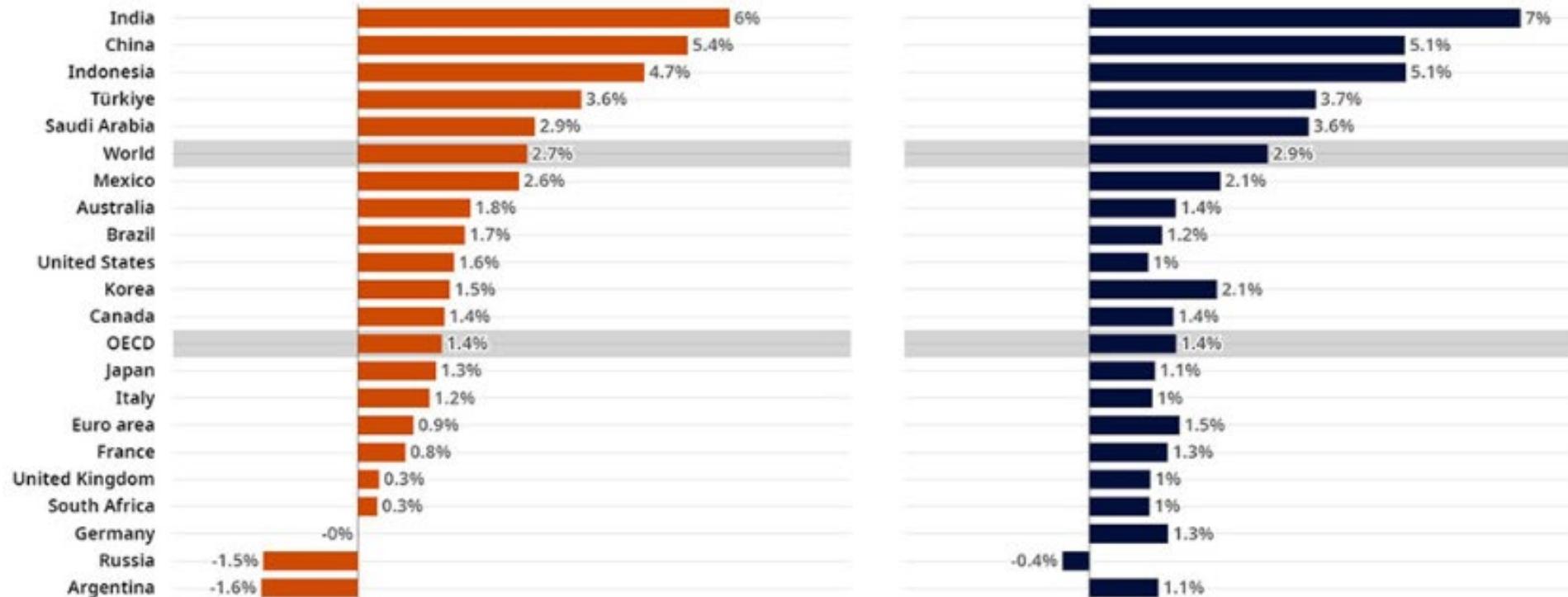
What kind of Level of Measurement is Economic Growth?



Real GDP growth projections for 2023 and 2024

%, year-on-year

Major economies OECD countries



Source: [OECD Economic Outlook, June 2023](#).



Matching Variable vs. Level of Measurement

School grade

Nominal

Temperature

Ordinal

Weight

Interval

Eye Color

Ratio

Length

Task -1

EXERCISE 1. At what level are each of the following variables operationalized?

	Variable Name	Value Labels	Data Type
1	Age	1, 2, 3,	<i>Continuous</i>
2	Sex	male, female, other	
3	Class standing	freshman, sophomore, junior, senior	
4	Marital Status	married, single, divorced, widowed	
5	Median Household Income	annual household income in dollars	
6	How much you like the food on campus?	a lot, some, a little, other	
7	The number of toxic waste sites in your community	0, 1, 2, 3,	
8	The number of toxic waste sites in your community	0, 1–5, 6–10, 11+	
9	Your GPA	below average, average, above average	

Task -1

Variable Name	Value Labels	Data Type	Operational Level
Age	1, 2, 3,	Continuous	Ratio
Sex	male, female, other	Nominal	Nominal
Class standing	freshman, sophomore, junior, senior	Ordinal	Ordinal
Marital Status	married, single, divorced, widowed	Nominal	Nominal
Median Household Income	annual household income in dollars	Continuous	Ratio
How much you like the food on campus?	a lot, some, a little, other	Ordinal	Ordinal
The number of toxic waste sites in your community (exact number)	0, 1, 2, 3,	Discrete	Ratio
The number of toxic waste sites in your community (ranges)	0, 1–5, 6–10, 11+	Ordinal	Ordinal
Your GPA	below average, average, above average	Ordinal	Ordinal

TIMEOUT

How are we doing?

EVERYTHING CLEAR?

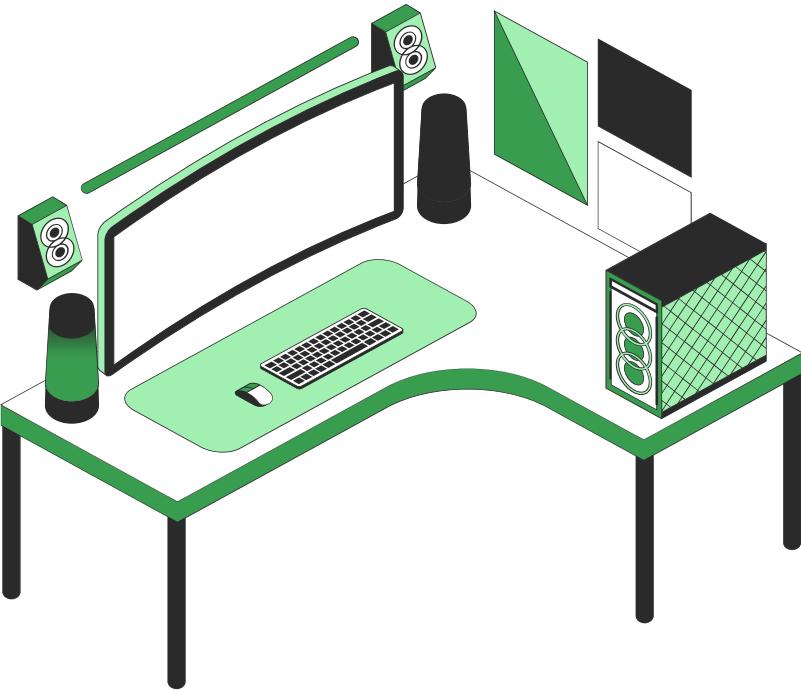
LET'S TEST

memegenerator.net



**Today's session was
productive**





Do you
have any
questions?

Send it to us! We hope you learned
something new.

It is time to Kahoot !!

