

The background features a light blue and white 3D bar chart with several bars of varying heights. In the foreground, there is a 3D pie chart with three segments. The overall aesthetic is clean and professional, typical of a university lecture slide.

STATISTICAL ANALYSIS - LECTURE 01

Dr. Mahmoud Mounir

mahmoud.mounir@cis.asu.edu.eg

STATISTICS

The science that deals with the collection, classification, analysis, interpretation and draw conclusions of numerical facts or data.

INTRODUCTION



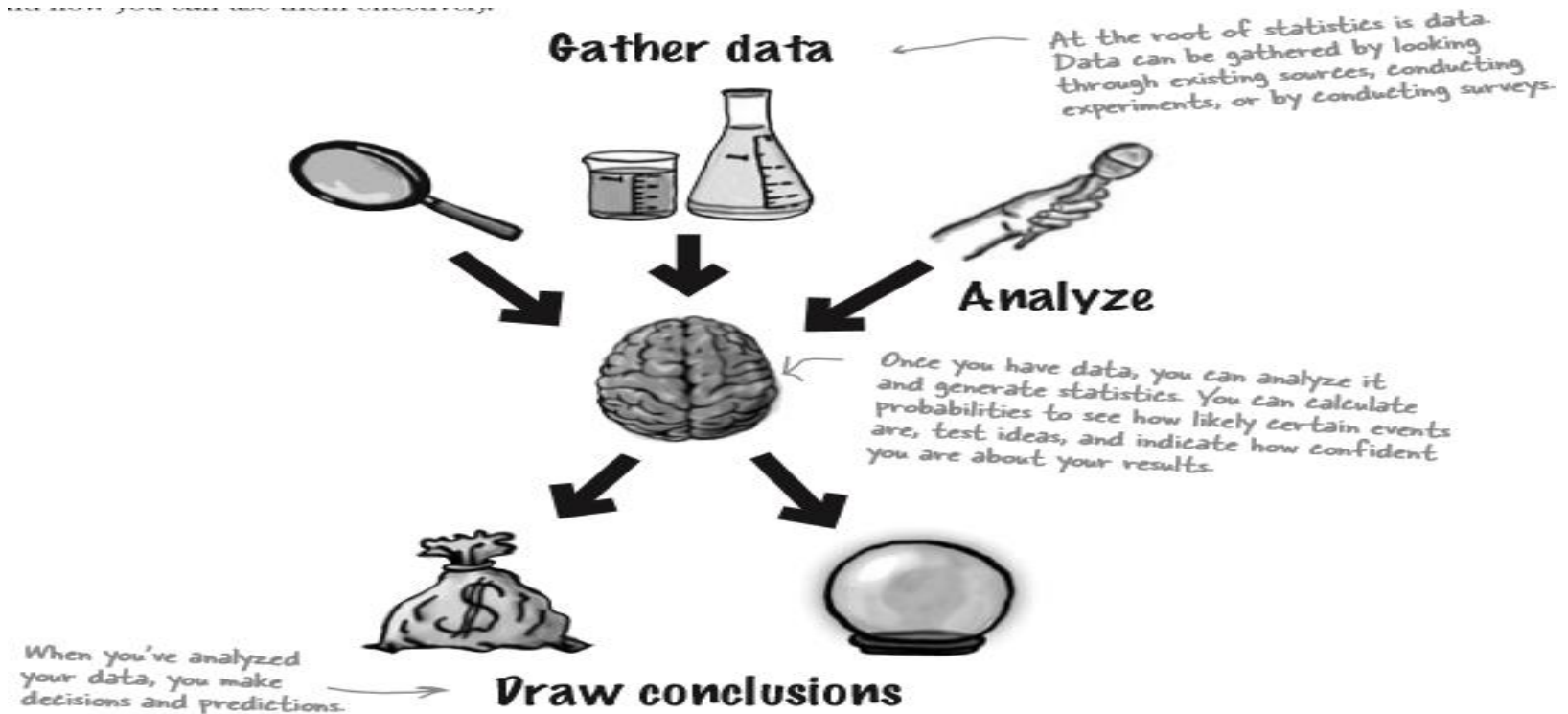
Statistics make the complex simple.

You can discover what's going on, then you need a way of visualizing it and telling everyone else.

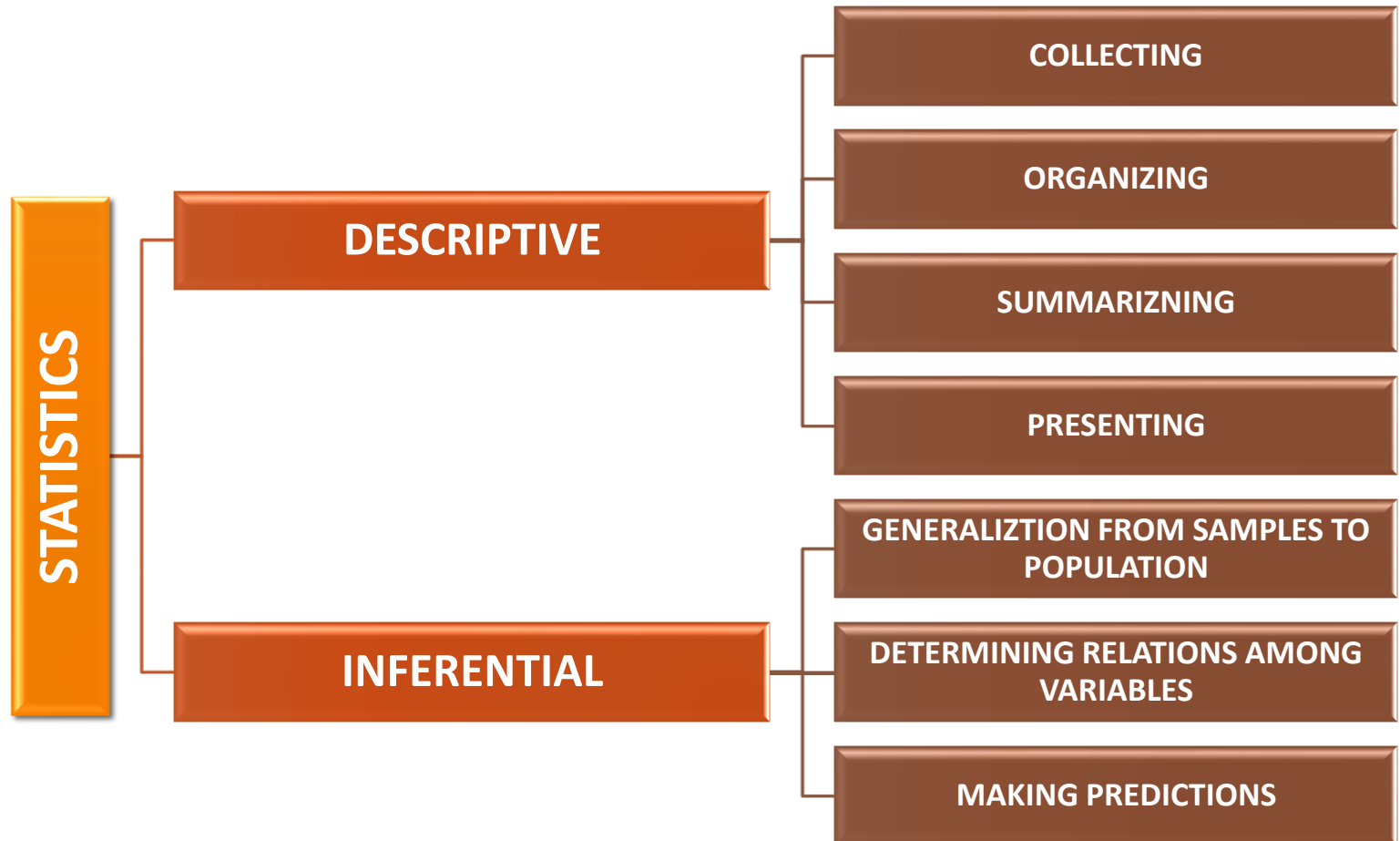
Statistics are numbers that summarize raw facts and figures in some meaningful way.

They present key ideas that may not be immediately apparent by just looking at the raw data, and by data, we mean facts or figures from which we can draw conclusions.

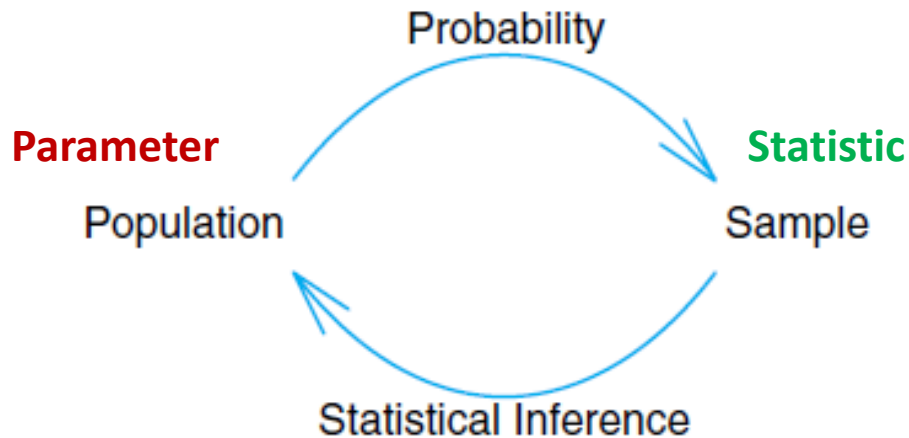
INTRODUCTION



INTRODUCTION

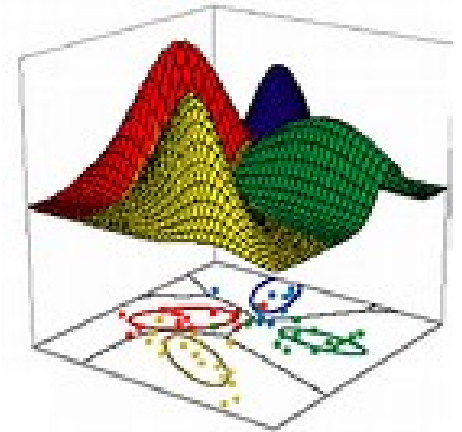


RELATION WITH PROBABILITY



INTRODUCTION

□ Applications:



COURSE OUTLINE

Statistical Analysis

- ❑ Exploring Data, Review on mean and variance, some graphical representations, Quartiles, IQR, Z- scores
- ❑ Review on Some Probability Distributions.
- ❑ Correlation and Regression
- ❑ Data Representation and Description
- ❑ Statistical Inference and Confidence Intervals.
- ❑ Test of Hypotheses

COURSE OUTLINE

A- Methods Used

1. Midterm
2. Quizzes
3. Project
4. Assignments
5. Lecture Tasks
6. Final Exam

B- Assessment Schedule

- | | |
|------------------|---------------|
| 1. Midterm | week 8 |
| 2. Quizzes | week5 ,week10 |
| 3. Project | week9-week11 |
| 4. Assignments | week4, week9 |
| 5. Lecture Tasks | week3 ,week6 |
| 6. Final Exam | week 15 |

C- Weighting of Assessments

- | | |
|------------------|---------------------|
| 1. Midterm | 15% |
| 2. Quizzes | 10% |
| 3. Project | 5% (BONUS) |
| 4. Assignments | 10% |
| 5. Lecture Tasks | 5% |
| 6. Final Exam | 60% |
| Total | 100% |

REFERENCES

- ❑ Robert S. Witte, John S. Witte, *Statistics*, 11th Edition, John Wiley & Sons, Inc., 2017.
- ❑ Ronald E. Walpole et al., *Probability and Statistics for Engineers and Scientists*. 9th Edition, Pearson Education International.
- ❑ Ye and Meyers, *Instructor's Solution Manual for Probability and Statistics for Engineers and Scientists*. 8th Edition.
- ❑ Dawn Griffiths, *Head First Statistics*. O'Reilly
- ❑ Brian Caffo, *Statistical Inference for Data Science*, Leanpub.
- ❑ Gareth James, et al., *An Introduction to Statistical Learning with Applications in R*. Springer.

DON'T FORGET
OUR CLASSROOM
CODE GROUP (1)

dylyahr

DON'T FORGET
OUR CLASSROOM
CODE GROUP (2)

4rhwft6





SAMPLING

SAMPLING

☐ POPULATION SIZE = 30

Street 1



Street 2

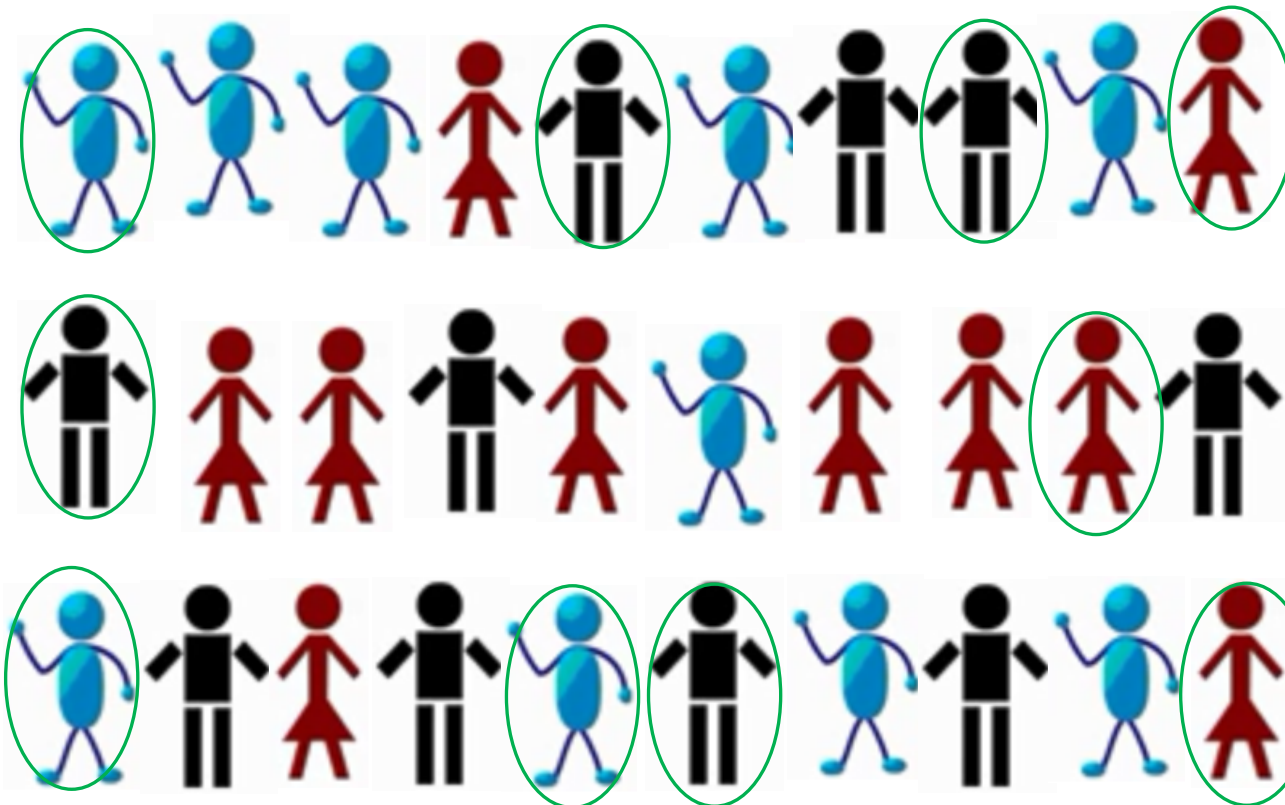


Street 3



SAMPLING

□ SIMPLE RANDOM SAMPLING (SRS)



➤ Population Size = 30

➤ Sample Size = 10

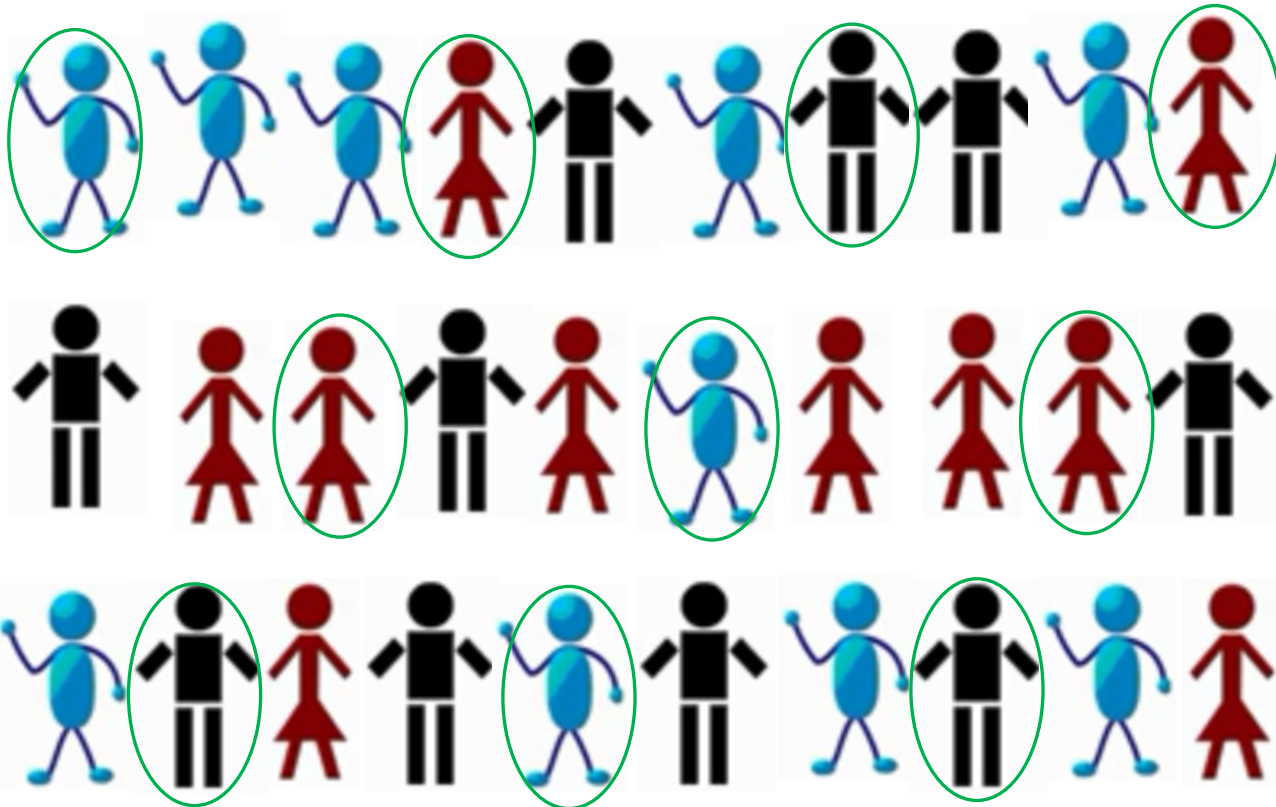
➤ EXAMPLE

RANDOM NUMBER

GENERATOR

SAMPLING

☐ SYSTEMATIC SAMPLING

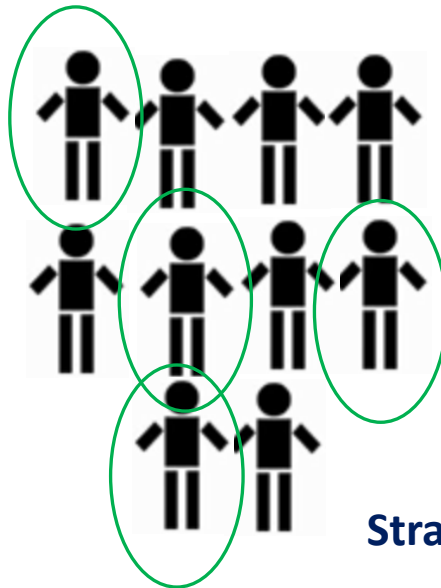
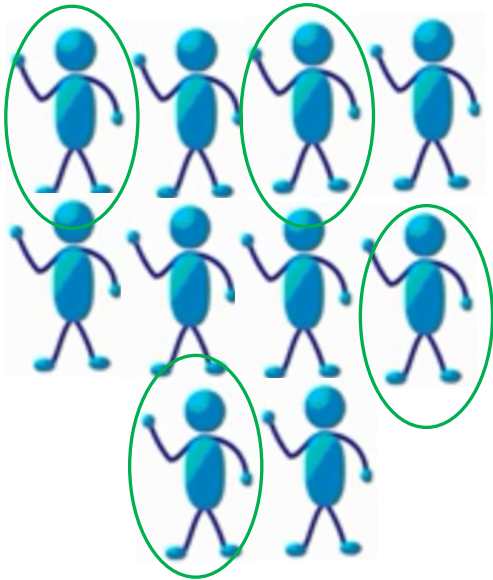


- Population Size = 30
- Sample Size = 10
- $K = 30 / 10 = 3$

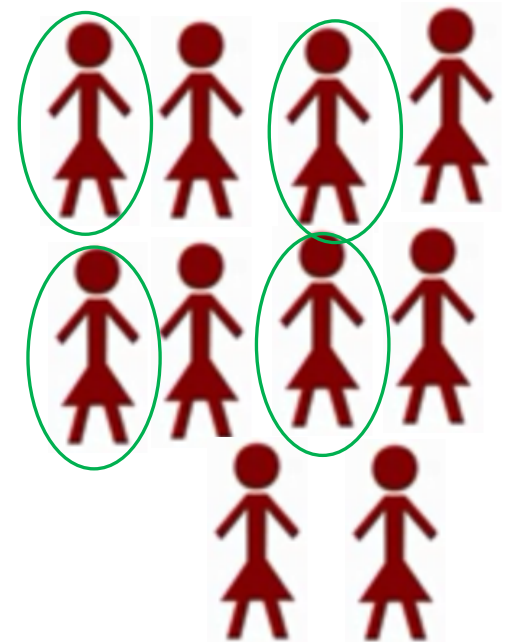
SAMPLING

☐ STRATIFIED SAMPLING

- Population Size = 30
- Sample Size = 12

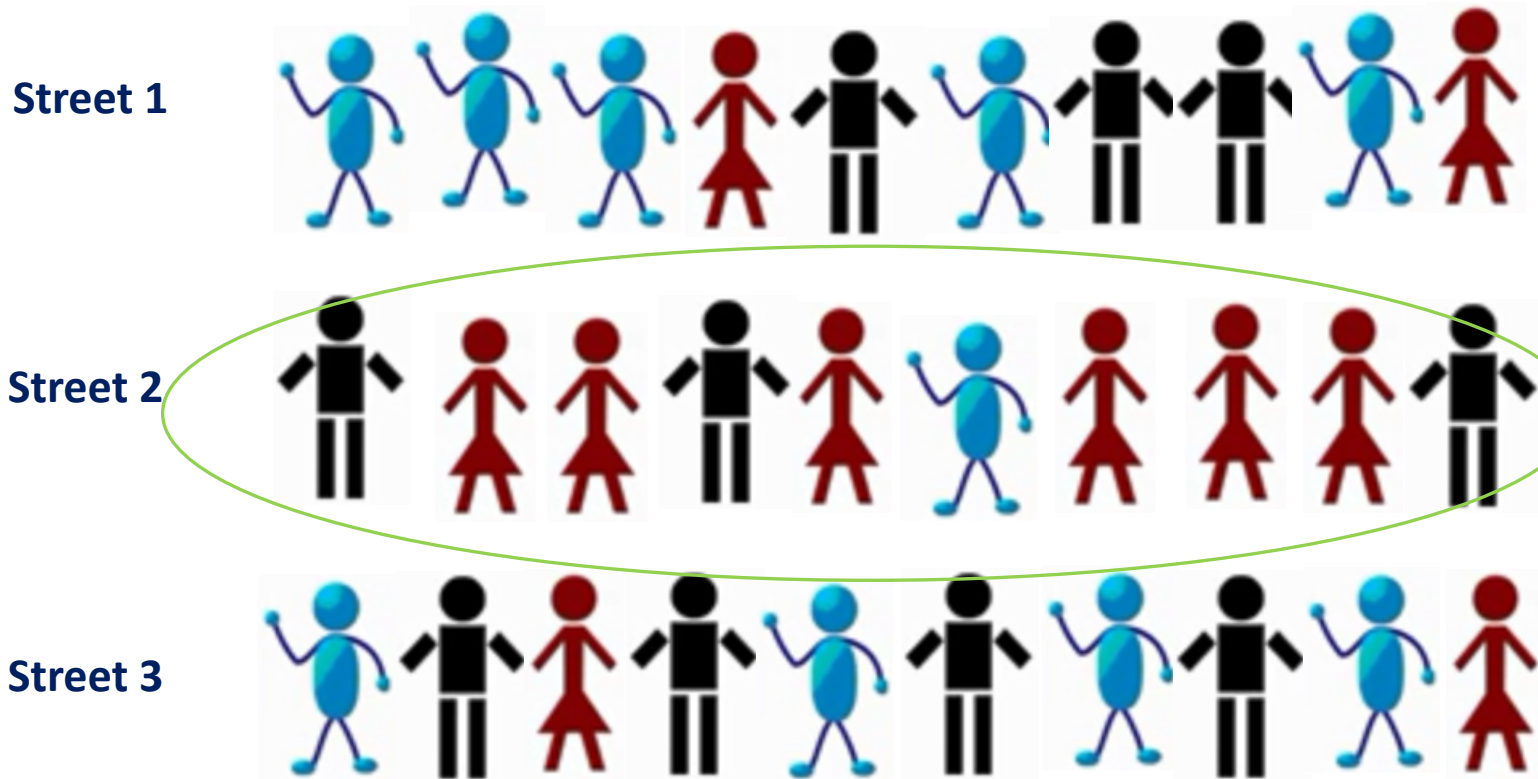


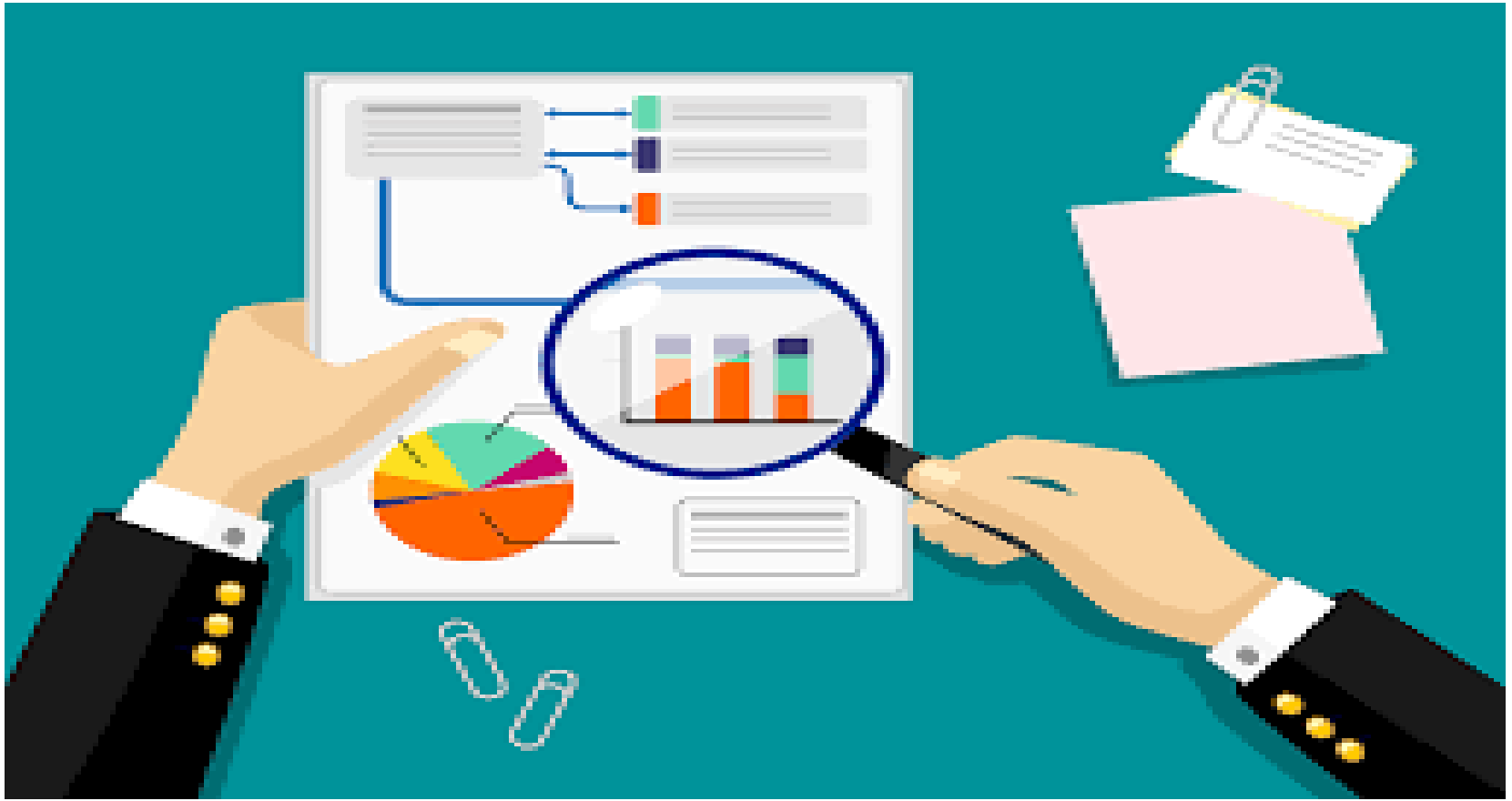
Stratum



SAMPLING

□ CLUSTER SAMPLING





EXPLORING DATA

EXPLORING DATA

☐ Cases, Variables and Levels of Measurements.

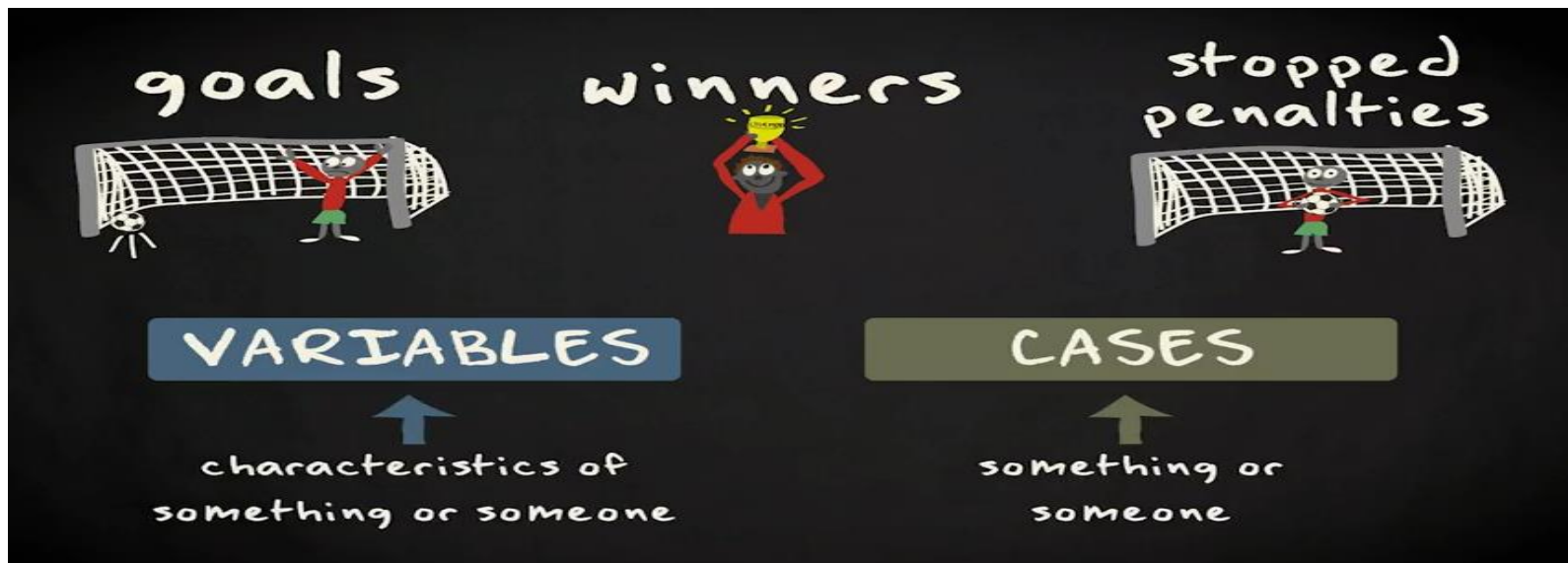
EXPLORING DATA

- Cases, Variables and Levels of Measurements.



EXPLORING DATA

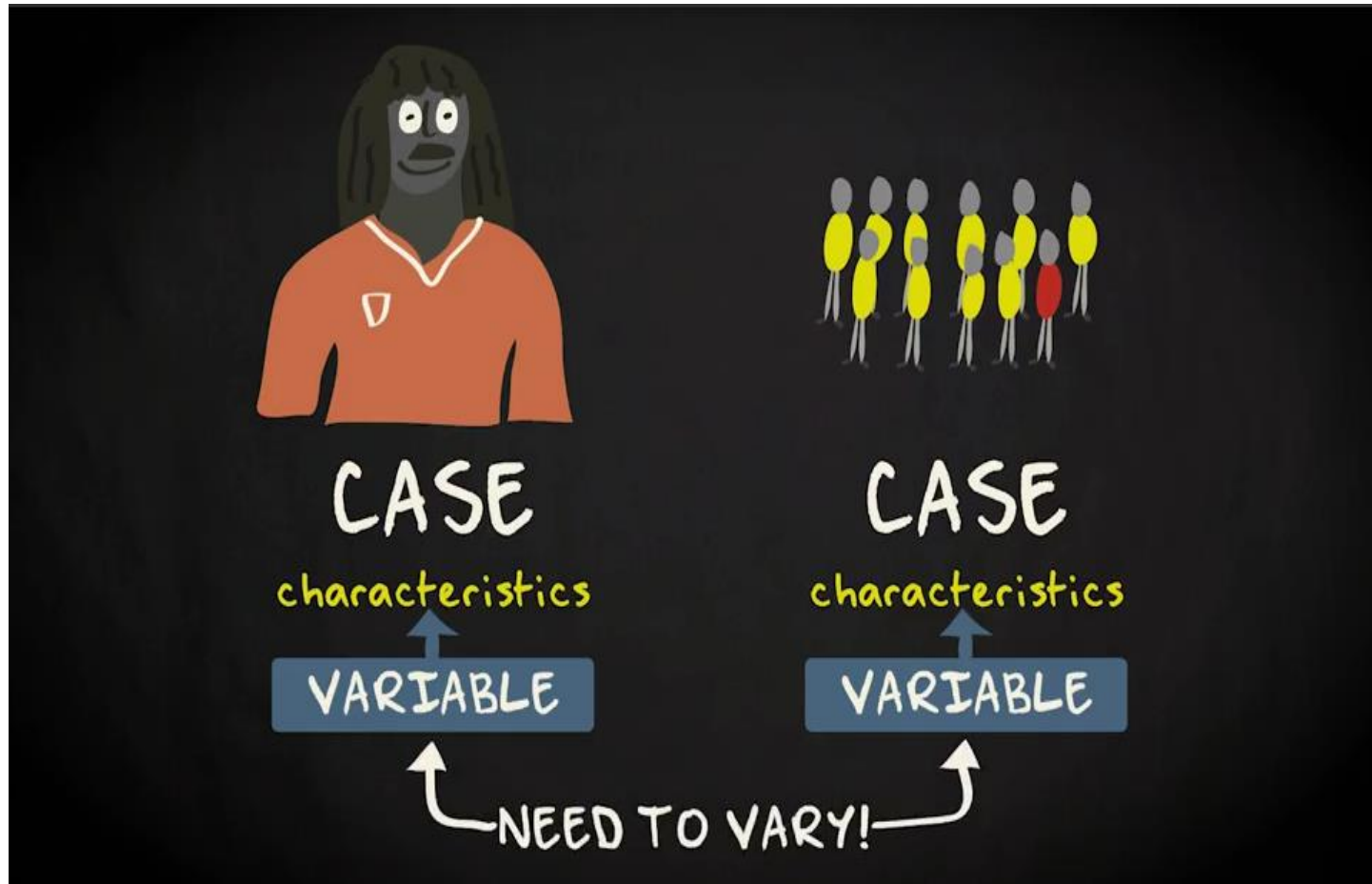
- Cases, Variables and Levels of Measurements.



EXPLORING DATA



EXPLORING DATA



EXPLORING DATA



EXPLORING DATA

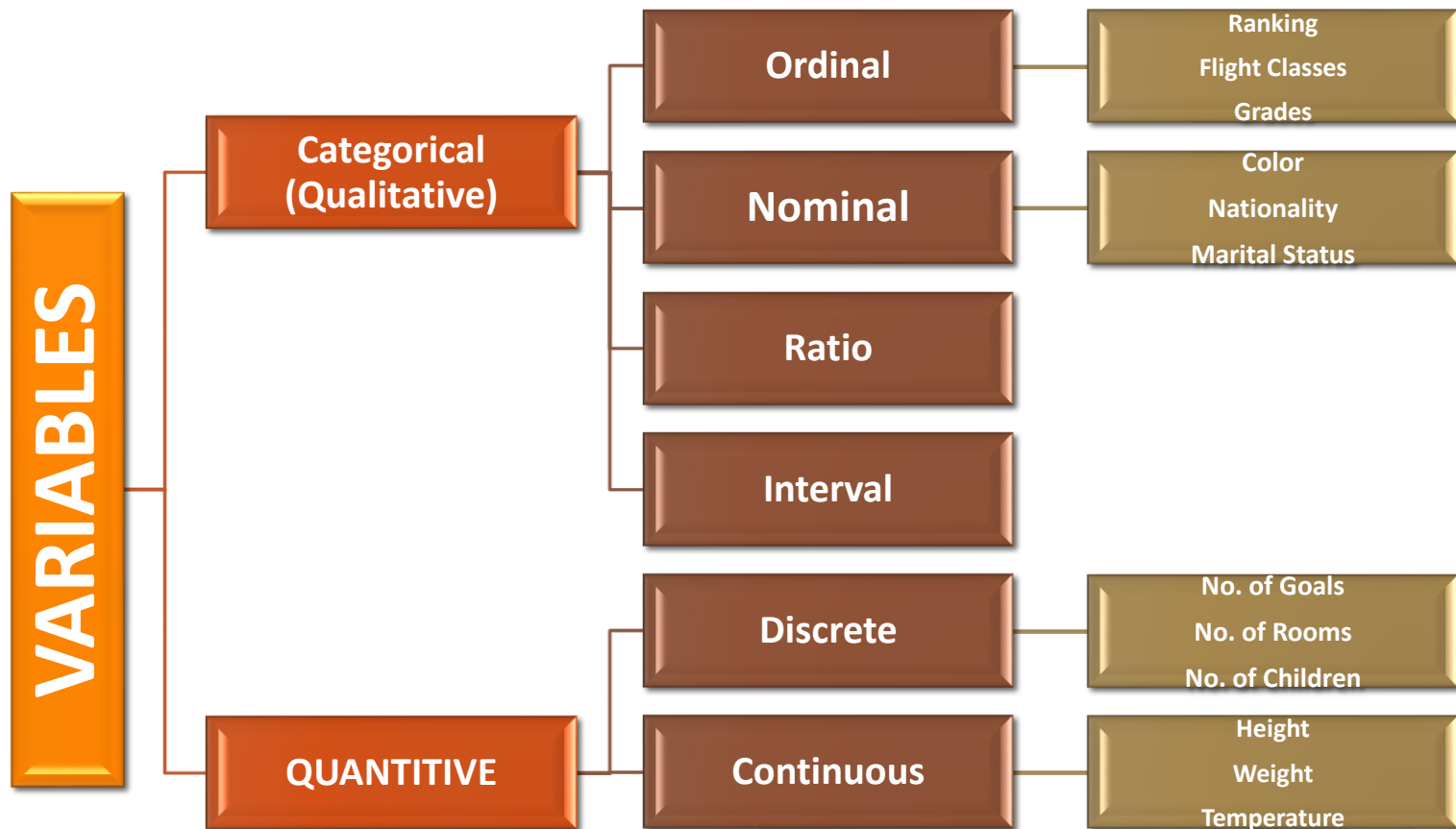


MEASUREMENT LEVELS



LEVELS OF MEASUREMENTS

LEVELS OF MEASUREMENTS



MEASUREMENT LEVELS



LEVELS OF MEASUREMENTS USING
STATISTICAL METHODS

DATA MATRIX AND FREQUENCY TABLE



The diagram features a central logo for 'Primera División' at the top, which is a soccer ball with a colorful ring around it. Below the logo, the text 'Primera División' is written in a white, handwritten-style font. The diagram is divided into two main sections: 'CASES' and 'VARIABLES'. The 'CASES' section is on the left, with a grey box containing the word 'CASES' in white. Below this box are three stylized soccer players: one with long black hair wearing a white jersey, one with short brown hair wearing a yellow and red striped jersey, and one with short yellow hair wearing a purple and red striped jersey. The 'VARIABLES' section is on the right, with a blue box containing the word 'VARIABLES' in white. Below this box, a list of variables is written in orange: 'Age', 'Weight', 'Goals scored', 'Team membership', and 'Hair color'.

Primera División

CASES

VARIABLES

- Age
- Weight
- Goals scored
- Team membership
- Hair color

DATA MATRIX AND FREQUENCY TABLE

□ **DATA MATRIX:** It is an overview of the cases and variables

DATA MATRIX		VARIABLES				
		Age	Weight	Goals scored	Team membership	Hair color
CASES	Player 1	18	72.6	0	Real Zaragoza	Blond
	Player 2	21	71.4	0	Real Betis	Black
	Player 3	26	74.8	8	Sevilla	Black
	Player 4	22	76.8	12	Barcelona	Black
	Player 5	22	74.1	17	Valencia	Other
	Player 6	27	78.9	3	Real Sociedad	Other
	Player 7	30	80.3	2	Real Madrid	Blond
	Player 8	24	73.3	1	Athletic Bilbao	Brown
	Player 9	23	76.9	5	Valencia	Brown
	...					
Player 400	26	77.2	0	Athletic Madrid	Other	

DATA MATRIX AND FREQUENCY TABLE

□ **DATA MATRIX:** It is an overview of the cases and variables

DATA MATRIX	OBSERVATIONS		VARIABLES			
	Age	Weight	Goals scored	Team membership	Hair color	
CASES	Player 1	18	72.6	0	Real Zaragoza	Blond
	Player 2	21	71.4	0	Real Betis	Black
	Player 3	26	74.8	8	Sevilla	Black
	Player 4	22	76.8	12	Barcelona	Black
	Player 5	22	74.1	17	Valencia	Other
	Player 6	27	75.9	3	Real Sociedad	Other
	Player 7	30	80.3	2	Real Madrid	Blond
	Player 8	24	75.3	1	Athletic Bilbao	Brown
	Player 9	23	76.9	5	Valencia	Brown
	...					
Player 400	26	77.2	0	Athletic Madrid	Other	

DATA MATRIX AND FREQUENCY TABLE

DATA MATRIX		VARIABLES				
		Age	Weight	Goals scored	Team membership	Hair color
CASES	Player 22	23	75.1	3	Valencia	Other
	Player 23	22	76.3	5	Barcelona	Black
	Player 24	26	■	0	Málaga	Black
	Player 25	■	80.1	0	Villareal	Blond
	Player 26	27	77.1	7	Getafe	Brown

you might have to remove them

DATA MATRIX AND FREQUENCY TABLE

- Data Matrices can be summarized using tables and graphs.



FREQUENCY TABLE

shows how the values are distributed over the cases

Hair color	Frequency		
Blond	76		
Brown	134		
Black	160		
other	30		
Total	400		

DATA MATRIX AND FREQUENCY TABLE



FREQUENCY TABLE


shows how the values are distributed over the cases

Hair color	Frequency	Percentage
Blond	76	19
Brown	134	33.5
Black	160	40
other	30	7.5
Total	400	100

Calculation for Blond percentage: $76/400 * 100 = 19$

DATA MATRIX AND FREQUENCY TABLE

☐ CUMMULATIVE FREQUENCY



FREQUENCY TABLE
shows how the values are distributed over the cases

Hair color	Frequency	Percentage	Cumulative percentage
Blond	76	19	19
Brown	134	33.5	52.5
Black	160	40	92.5
other	30	7.5	100
Total	400	100	

The table illustrates the calculation of cumulative percentages. The percentage for Blond (19) is added to the percentage for Brown (33.5) to get the cumulative percentage for Brown (52.5). This is shown by a yellow box containing '19+33.5' with an arrow pointing to the '52.5' in the cumulative percentage column.

DATA MATRIX AND FREQUENCY TABLE

FREQUENCY TABLE

shows how the values are distributed over the cases

QUANTITATIVE



Weight (in kgs)	Frequency	Percentage
65.3	2	0.5
65.4	1	0.25
65.5	3	0.75
65.6	1	0.25
65.7	0	0
65.8	0	0
65.9	1	0.25
...		



DATA MATRIX AND FREQUENCY TABLE

FREQUENCY TABLE

shows how the values are distributed over the cases

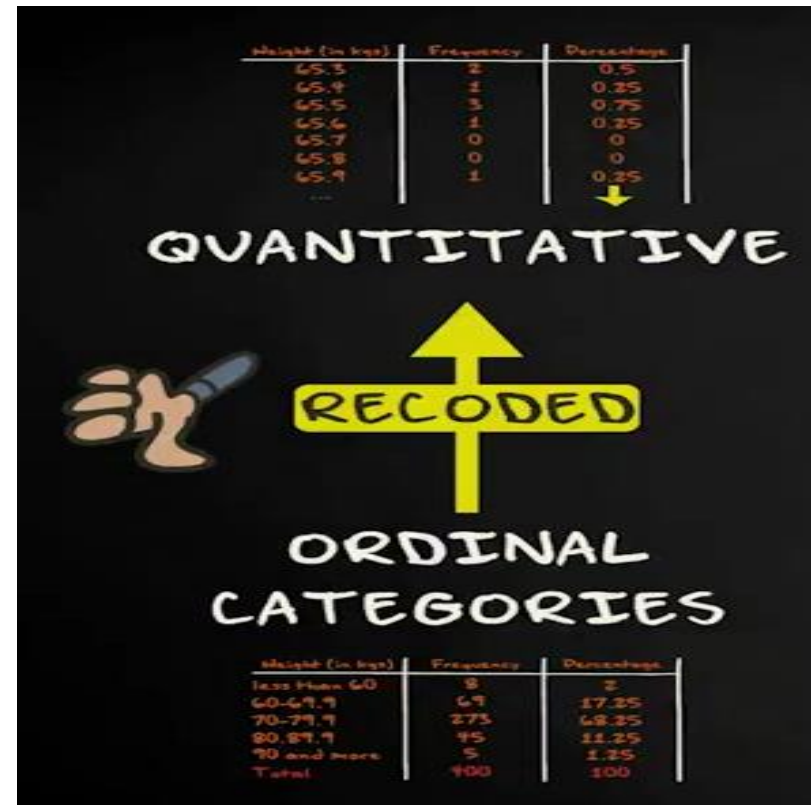
QUANTITATIVE

ORDINAL CATEGORIES

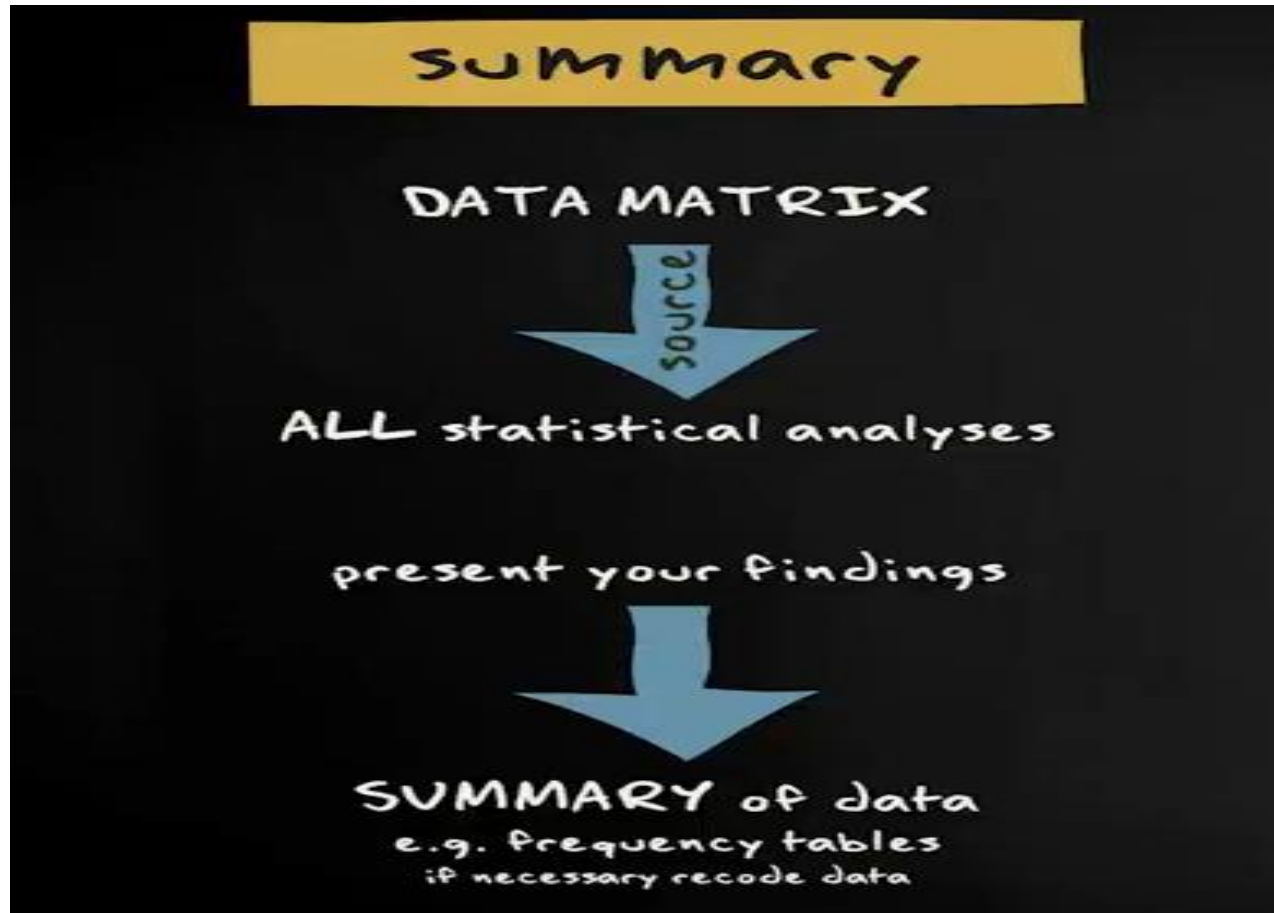
- you lose information
+ much better overview

Weight (in kgs)	Frequency	Percentage
less than 60	8	2
60-69.9	69	17.25
70-79.9	273	68.25
80.89.9	45	11.25
90 and more	5	1.25
Total	400	100

DATA MATRIX AND FREQUENCY TABLE



DATA MATRIX AND FREQUENCY TABLE



Categorical Frequency Distribution

☐ NOMINAL OR ORDINAL DATA

- ☐ Blood types for different persons
- ☐ Sample size (n) = 28

A	B	B	AB	O	O	A	O	O	B	A	B	O	AB
O	AB	B	B	A	A	O	B	B	O	O	O	A	O

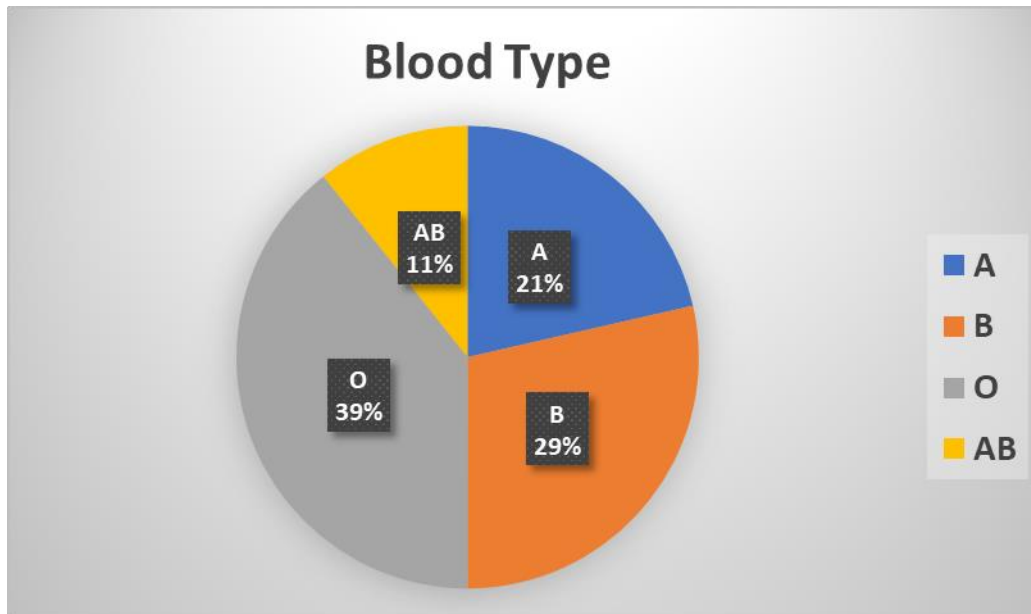
Class / Category	Frequency	Relative Frequency / Percent	Cumulative Frequency
A	6	21.43 %	21.43 %
B	8	28.57 %	50.00 %
O	11	39.29 %	89.29 %
AB	3	10.71 %	100.00 %
Total	28	100%	

NOMINAL DATA

$$\text{Percent} = \frac{f}{n} \times 100$$

DATA MATRIX AND FREQUENCY TABLE

❑ **Pie Chart** : A circle that is divided into sections according to the percentage of frequencies in each category.



**CATEGORICAL
NOMINAL DATA**

$$\text{Angle} = \frac{f}{n} \times 360^\circ$$

DATA MATRIX AND FREQUENCY TABLE

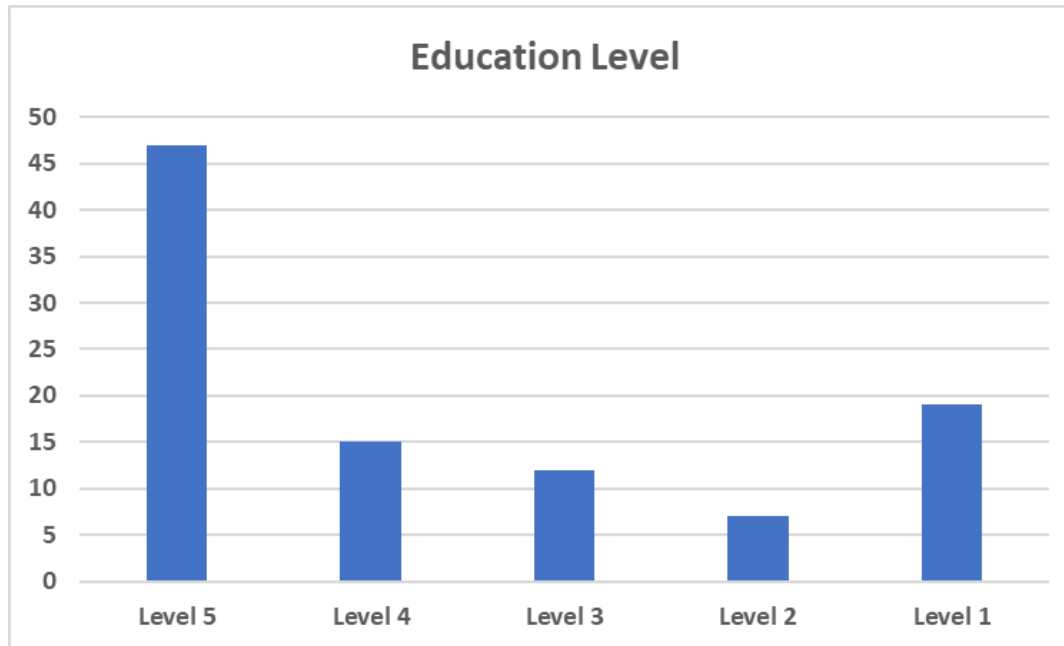
□ Education level for different students.

Education Level	Frequency	Relative Frequency / Percent	Cumulative Frequency
Level 5	47	47.00 %	47 %
Level 4	15	15.00 %	62.00 %
Level 3	12	12.00 %	74.00 %
Level 2	7	7.00 %	81.00 %
Level 1	19	19.00 %	100.00 %
Total	100	100%	

ORDINAL DATA

DATA MATRIX AND FREQUENCY TABLE

□ **Bar Chart** : Vertical bars of various heights to reflect the frequencies of the categories.



**ORDINAL
OR
DISCRETE**

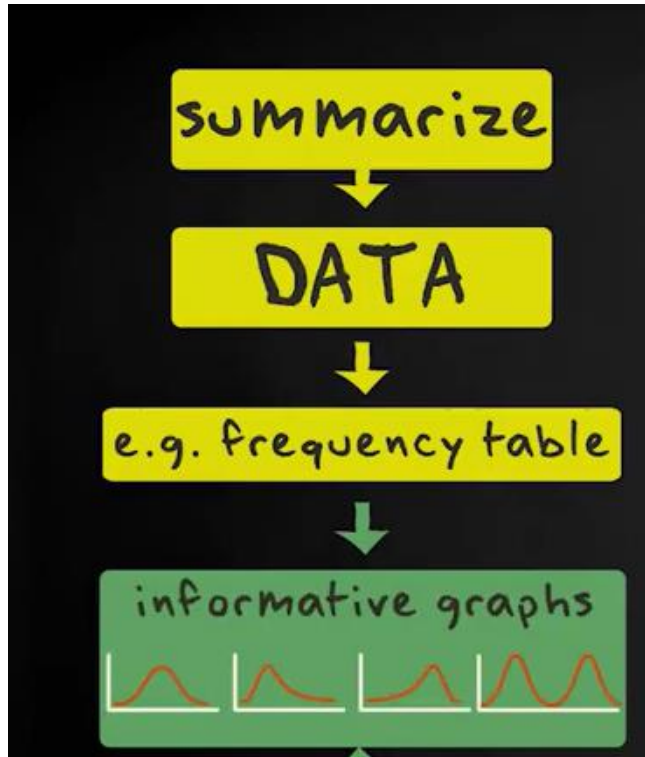
Ungrouped Frequency Distribution

☐ DISCRETE DATA WITH SMALL RANGE

☐ Sample size (n) = 16

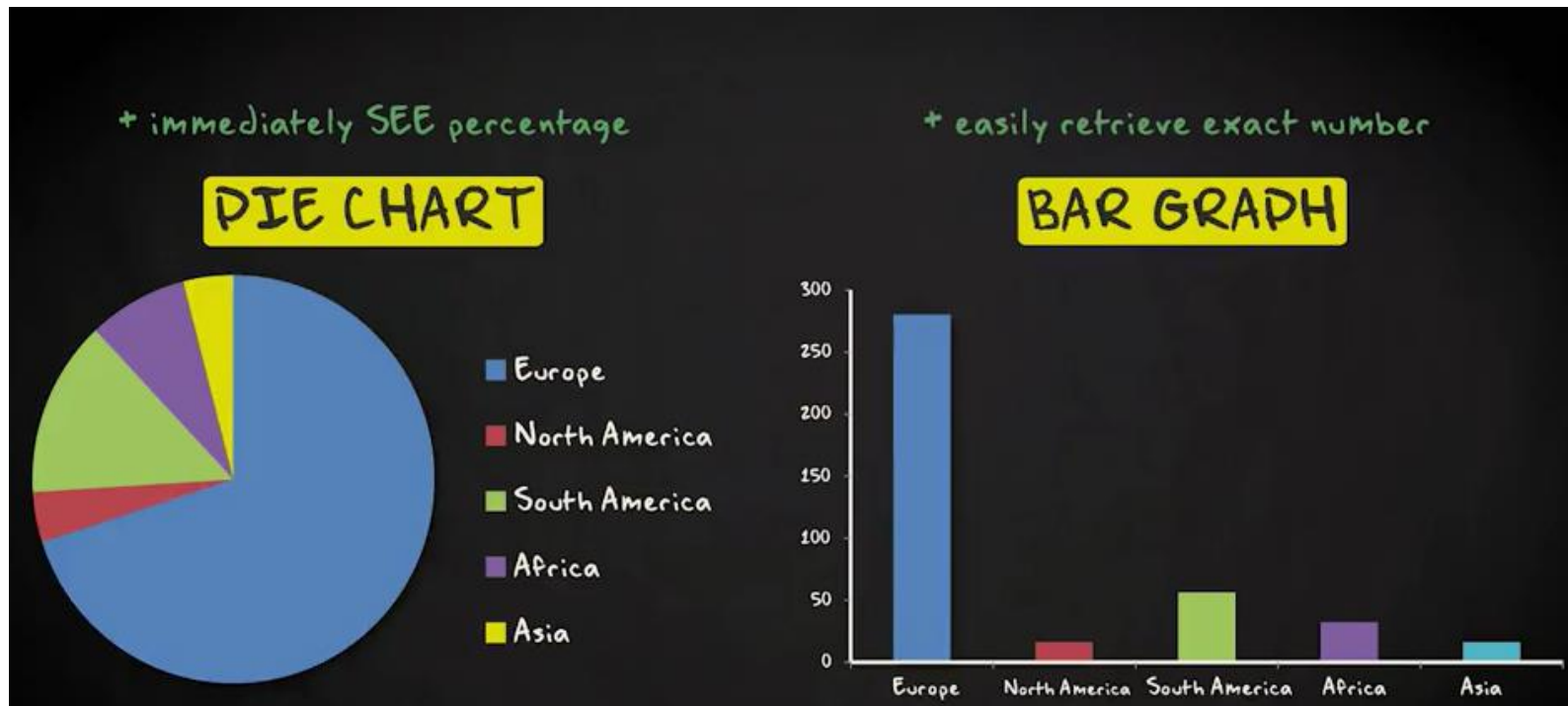
5	4	4	8
8	5	8	4
4	4	8	4
5	8	4	4

Class / Category	Frequency	Relative Frequency / Percent	Cumulative Frequency
4	8	50 %	50 %
5	3	18.75 %	68.75 %
8	5	31.25 %	100 %
Total	16	100%	

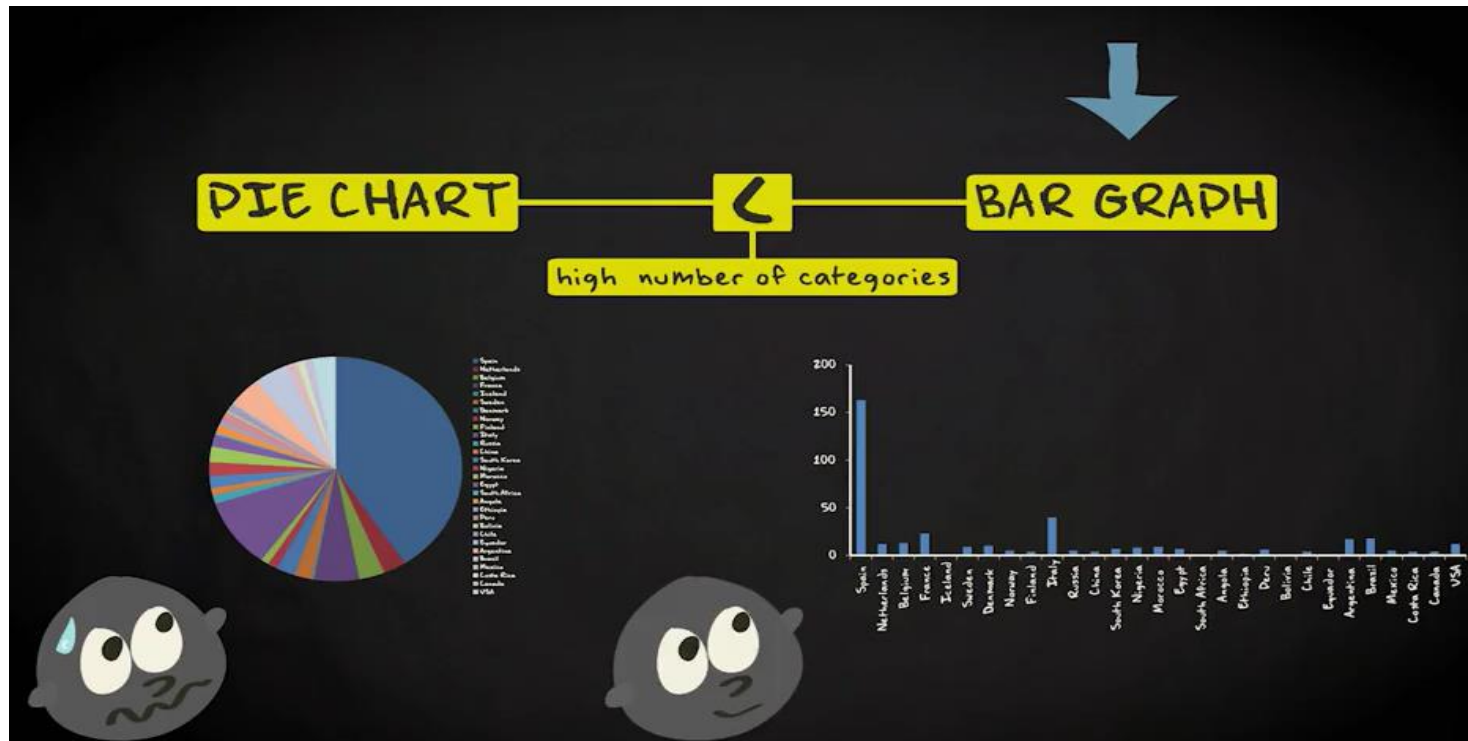


GRAPHS AND SHAPES OF DISTRIBUTIONS

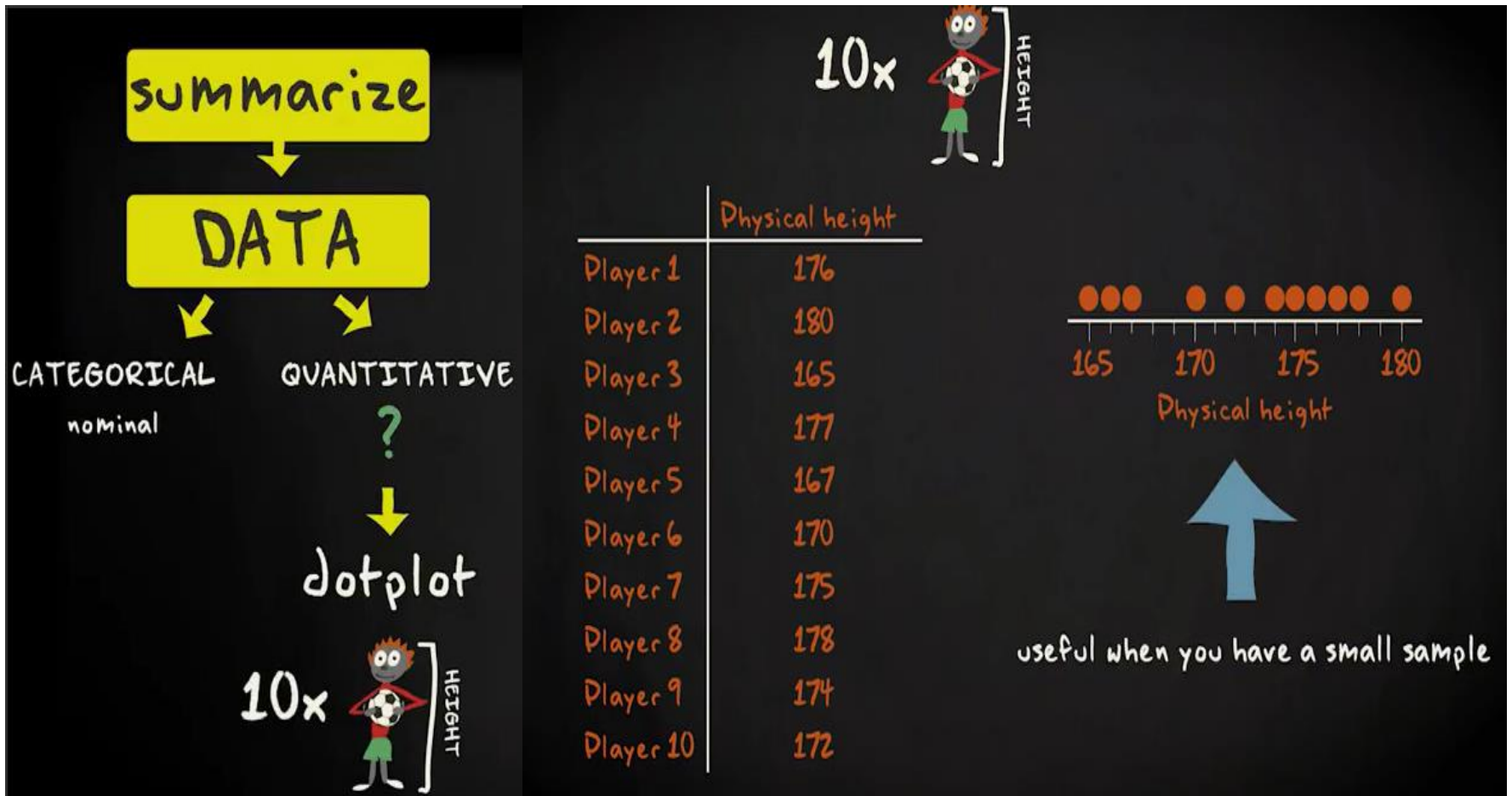
GRAPHS AND SHAPES OF DISTRIBUTIONS



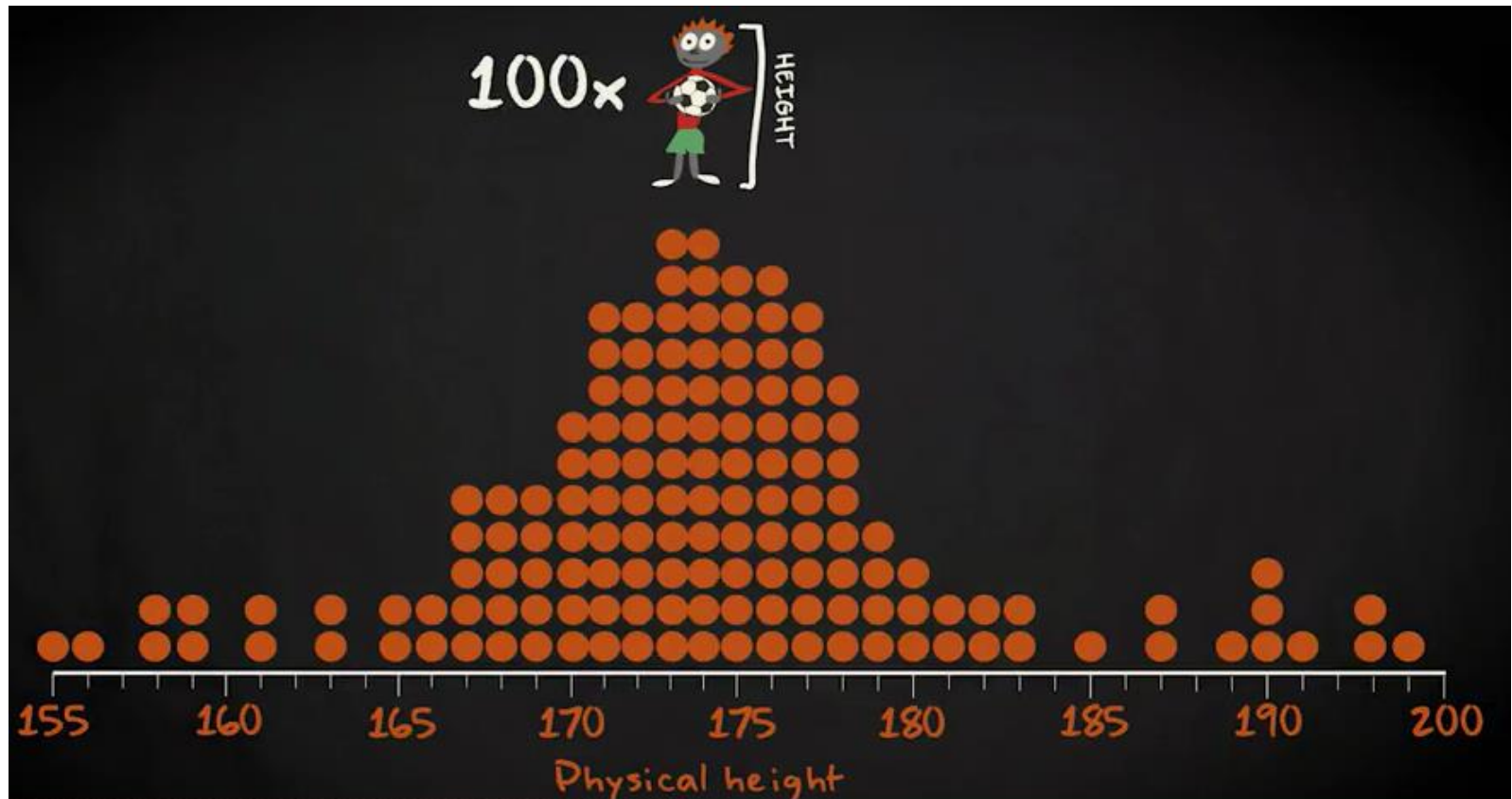
GRAPHS AND SHAPES OF DISTRIBUTIONS



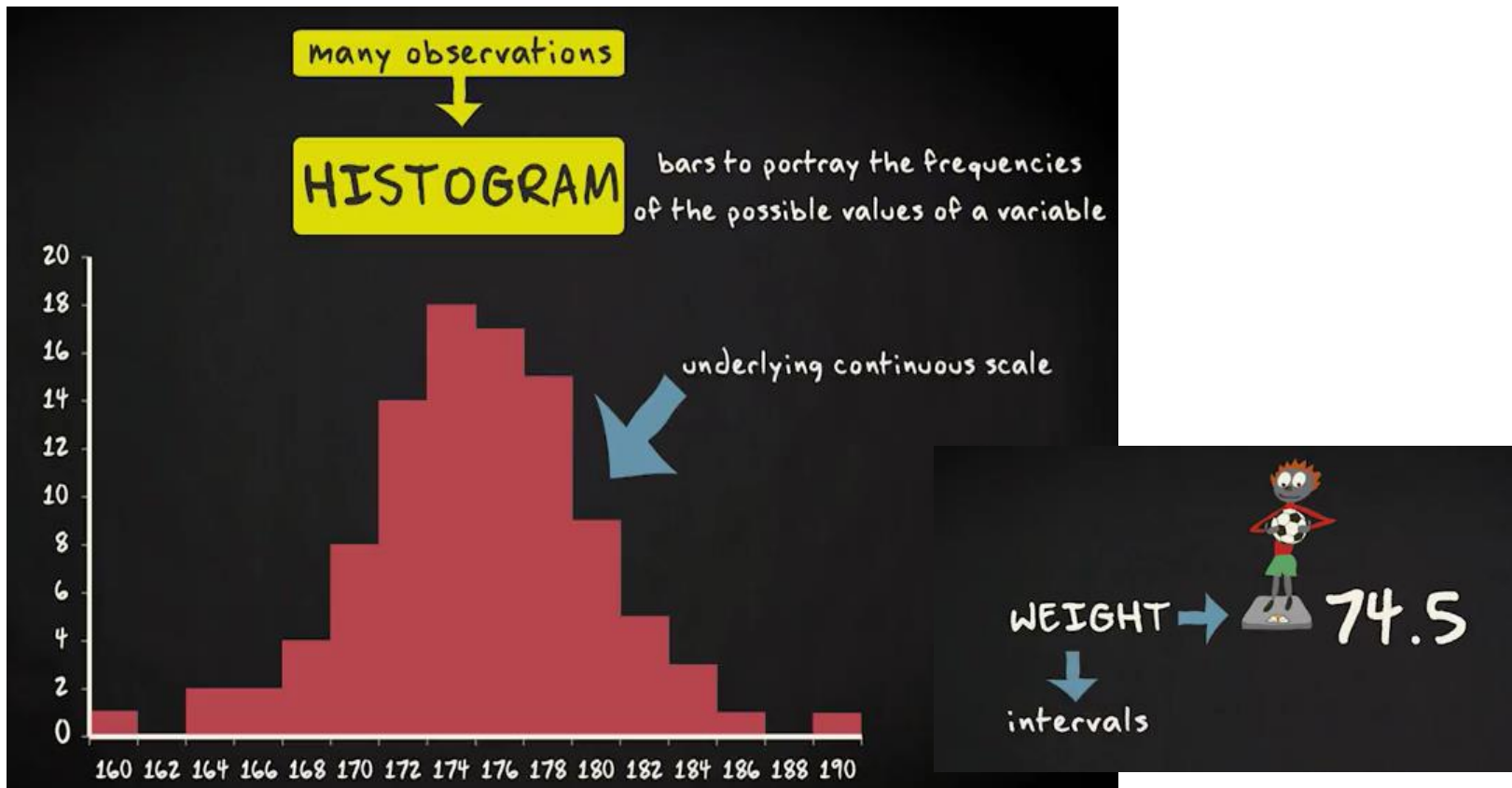
GRAPHS AND SHAPES OF DISTRIBUTIONS



GRAPHS AND SHAPES OF DISTRIBUTIONS



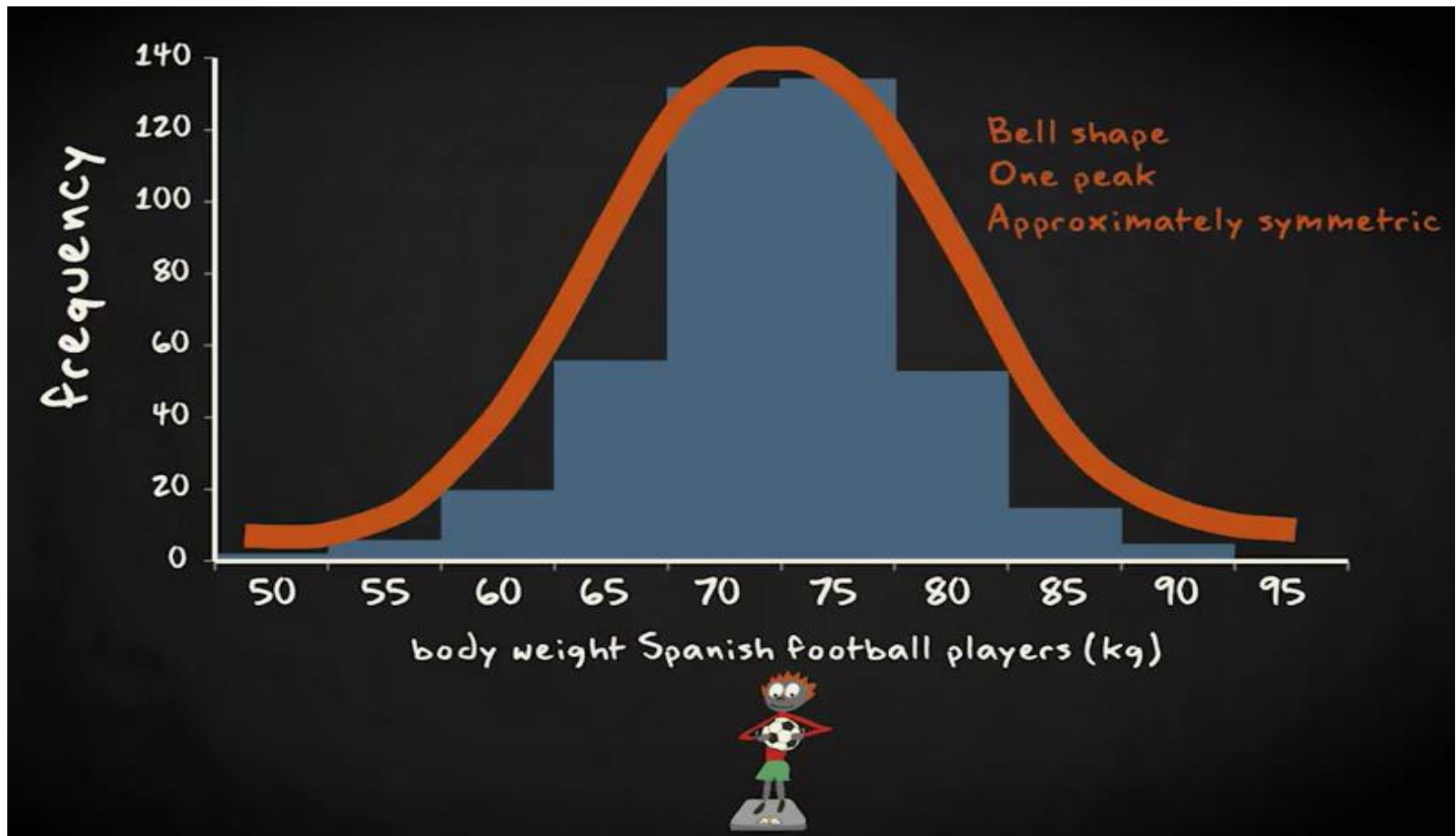
GRAPHS AND SHAPES OF DISTRIBUTIONS



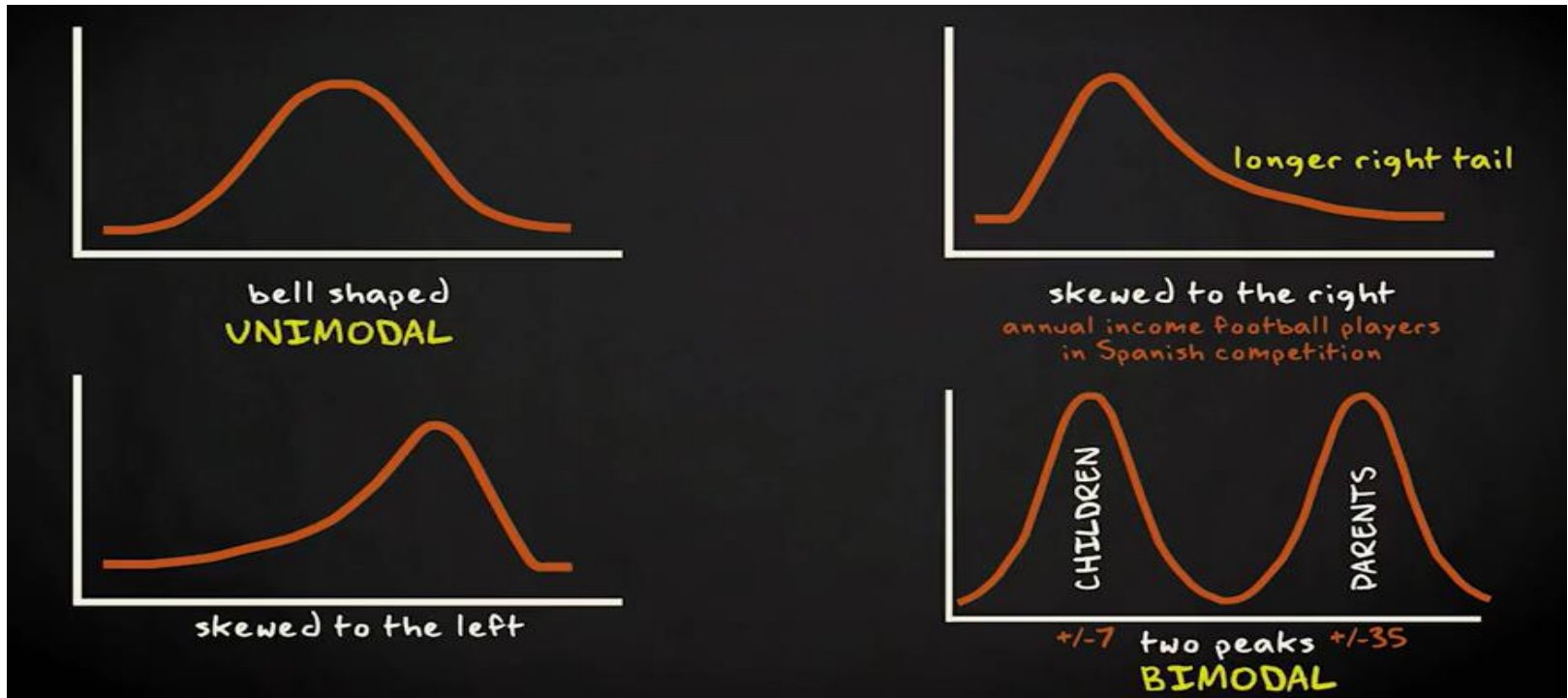
GRAPHS AND SHAPES OF DISTRIBUTIONS



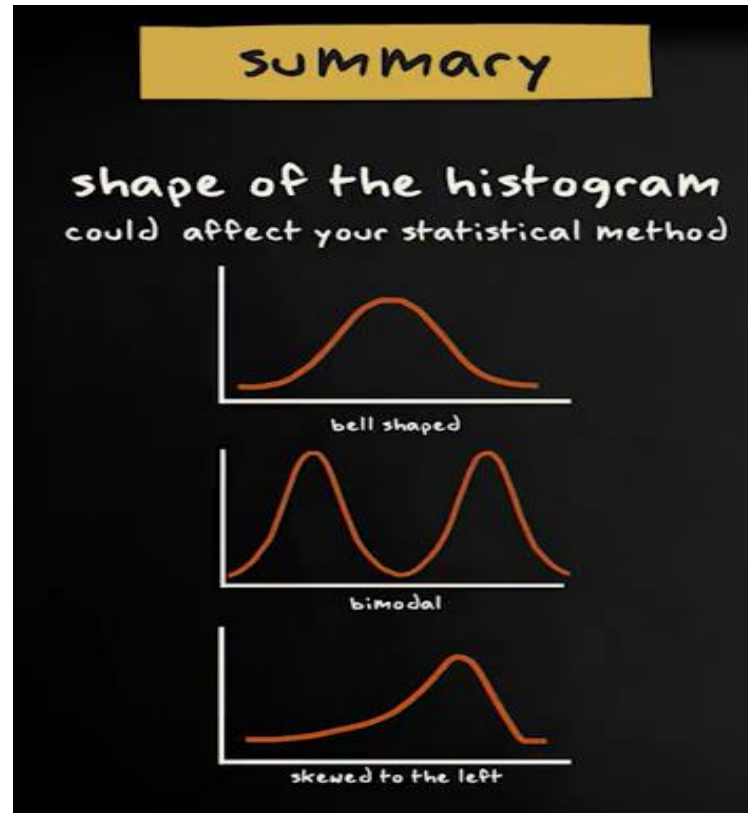
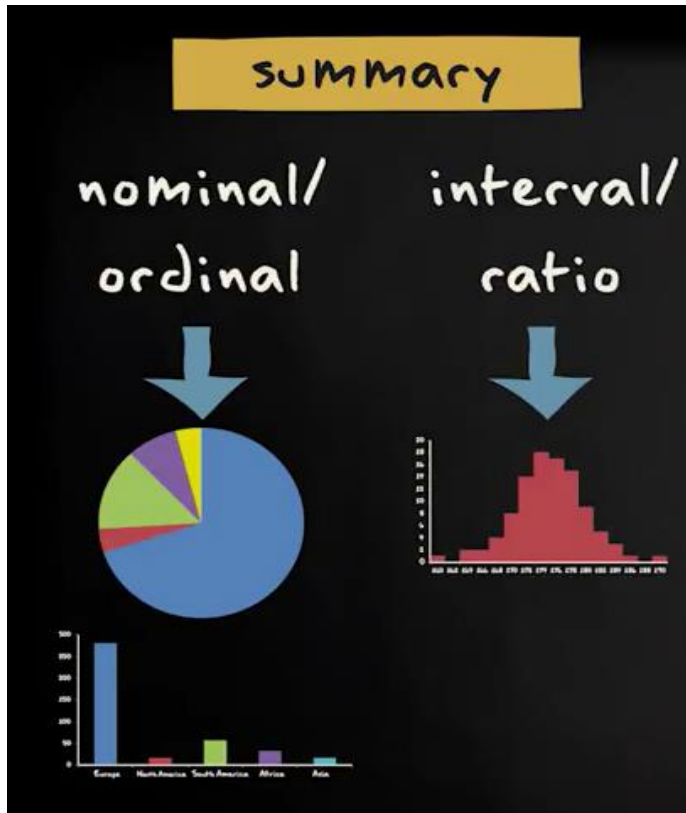
GRAPHS AND SHAPES OF DISTRIBUTIONS



GRAPHS AND SHAPES OF DISTRIBUTIONS



GRAPHS AND SHAPES OF DISTRIBUTIONS



Thank you!

Next Lecture:

- Measures of Central Tendency
- Measures of variability