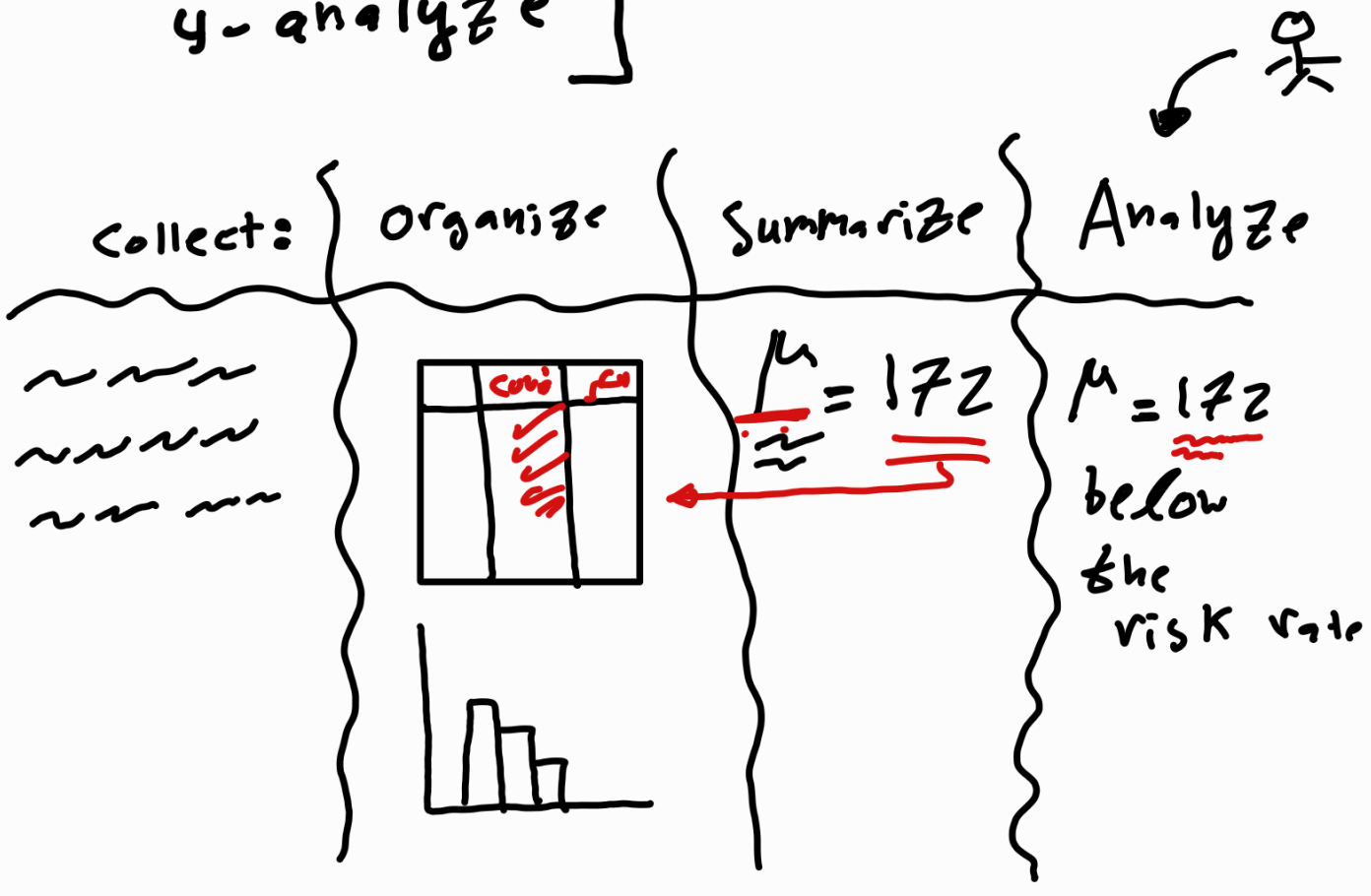
Statistics

→ Is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data

1- collect
2- organize
3- summarize
4- analyze
→ Conclusions



Conclusion:

*** Variable :**

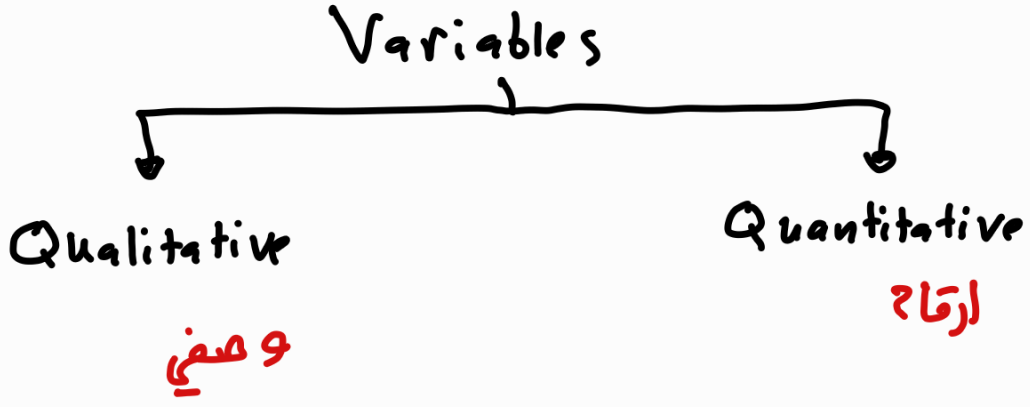
age / Name, Gender

Number

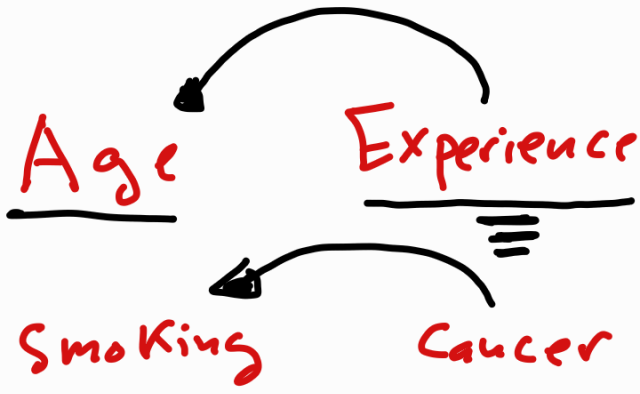
Data :

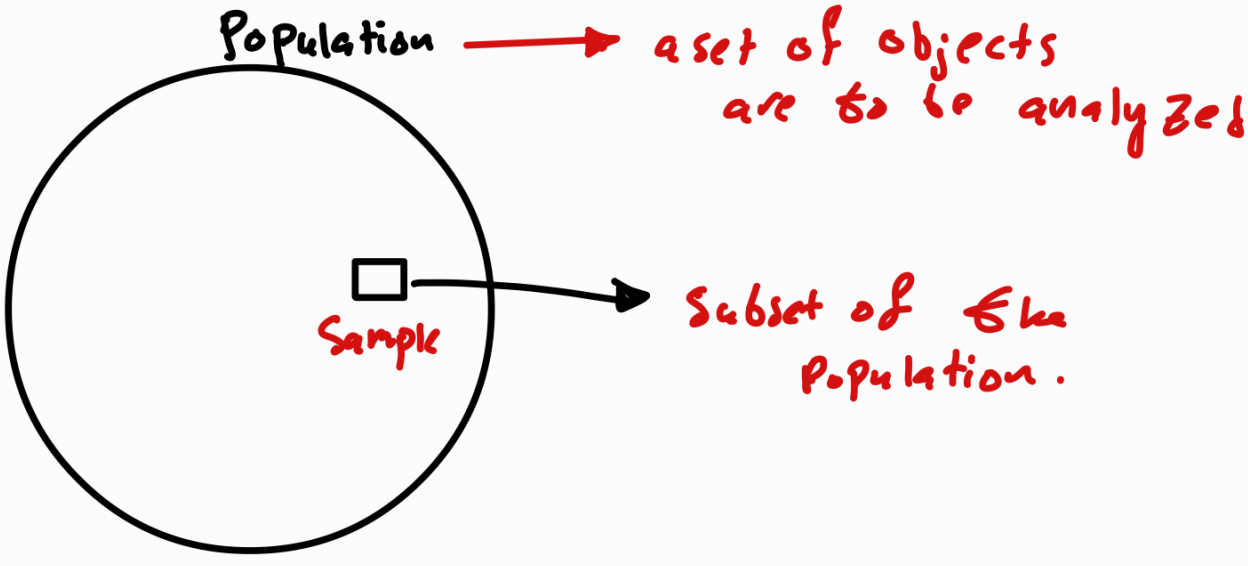
ID	Name	Gender	Age
111	Ali	M	20
222	Alaa	M	30

Data



Independent & Dependent Variables





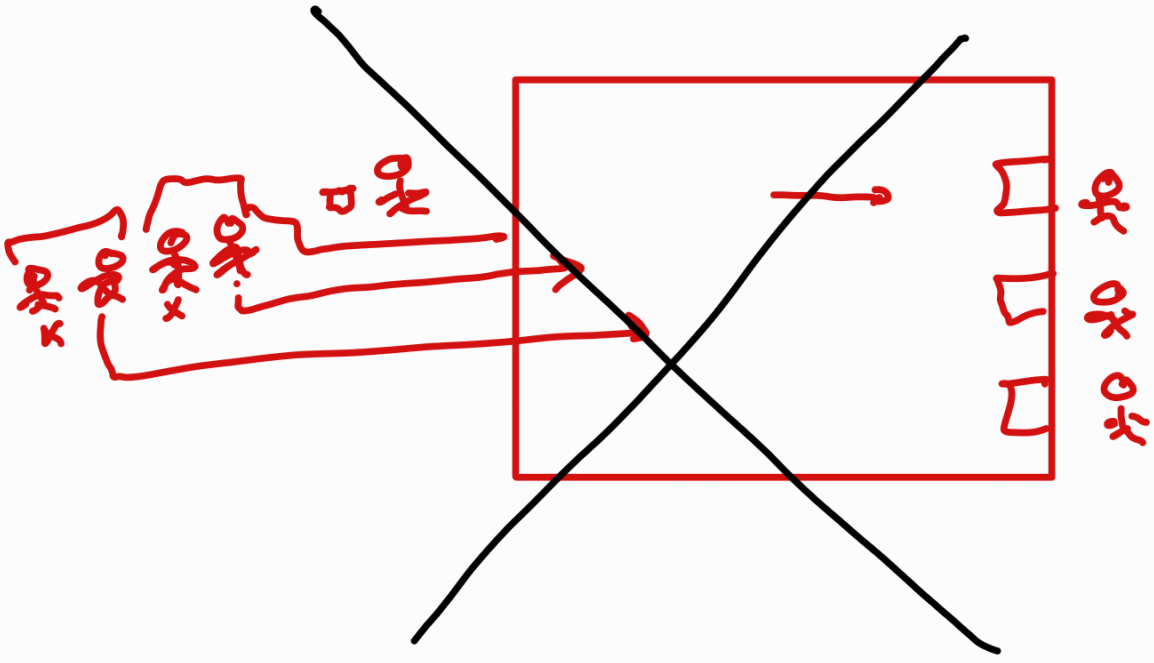
Sampling Techniques

- Random Sampling
- Systematic Sampling
- Stratified Sampling
- Cluster Sampling

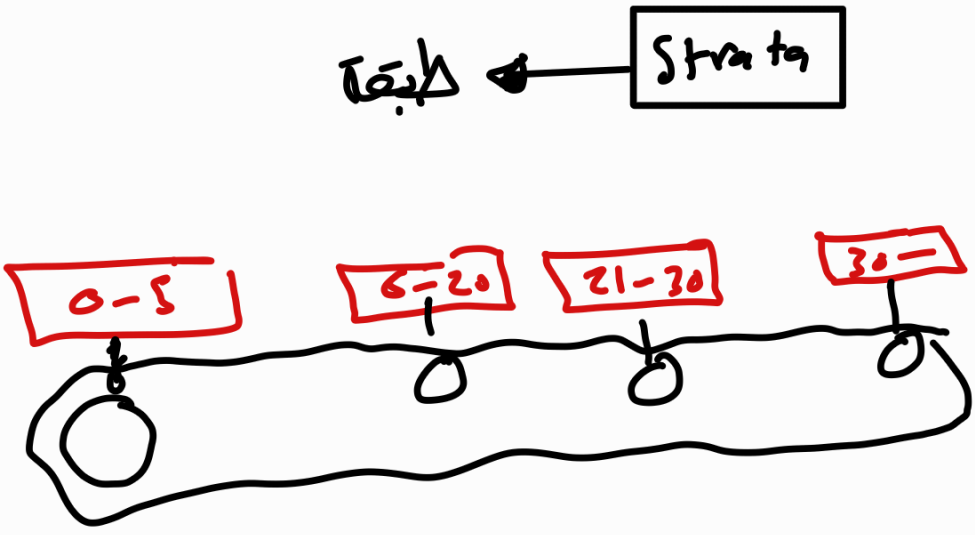
2 - Random Sampling:

based on chance.

Every member of population has an equal chance to be selected

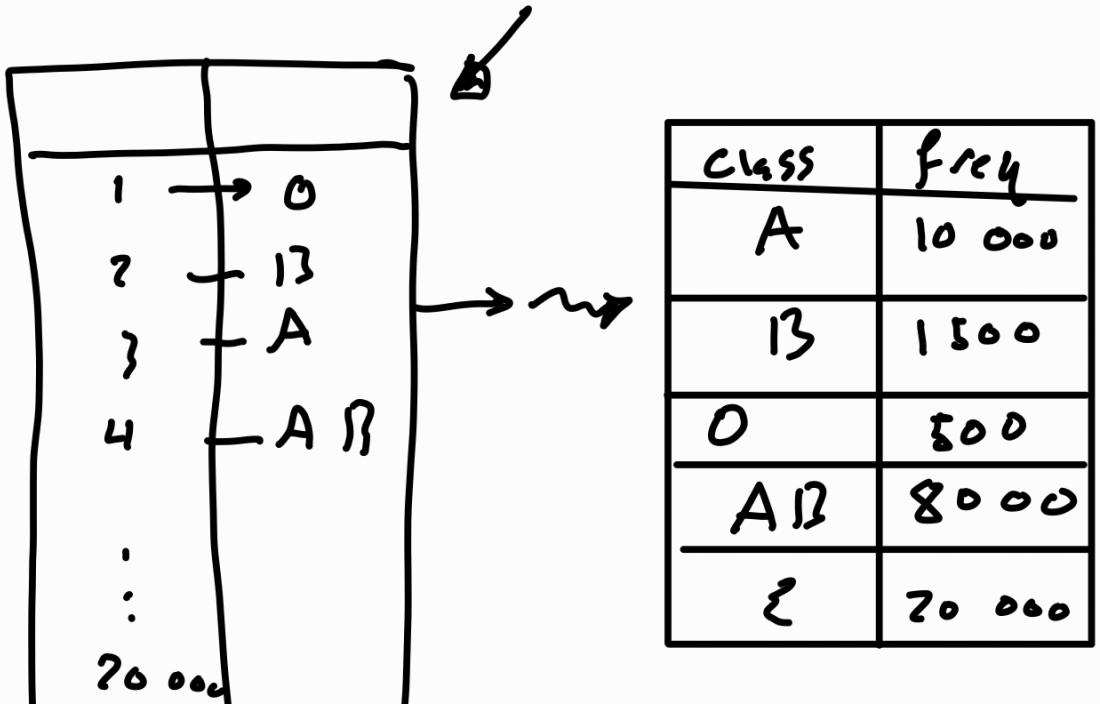


3- Stratified random sample:



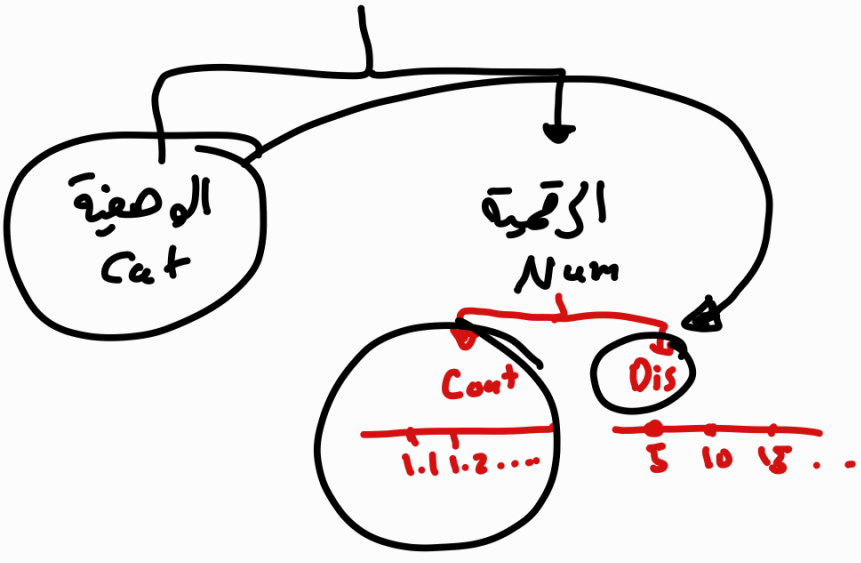
4- Cluster sample:

1- frequency Table :



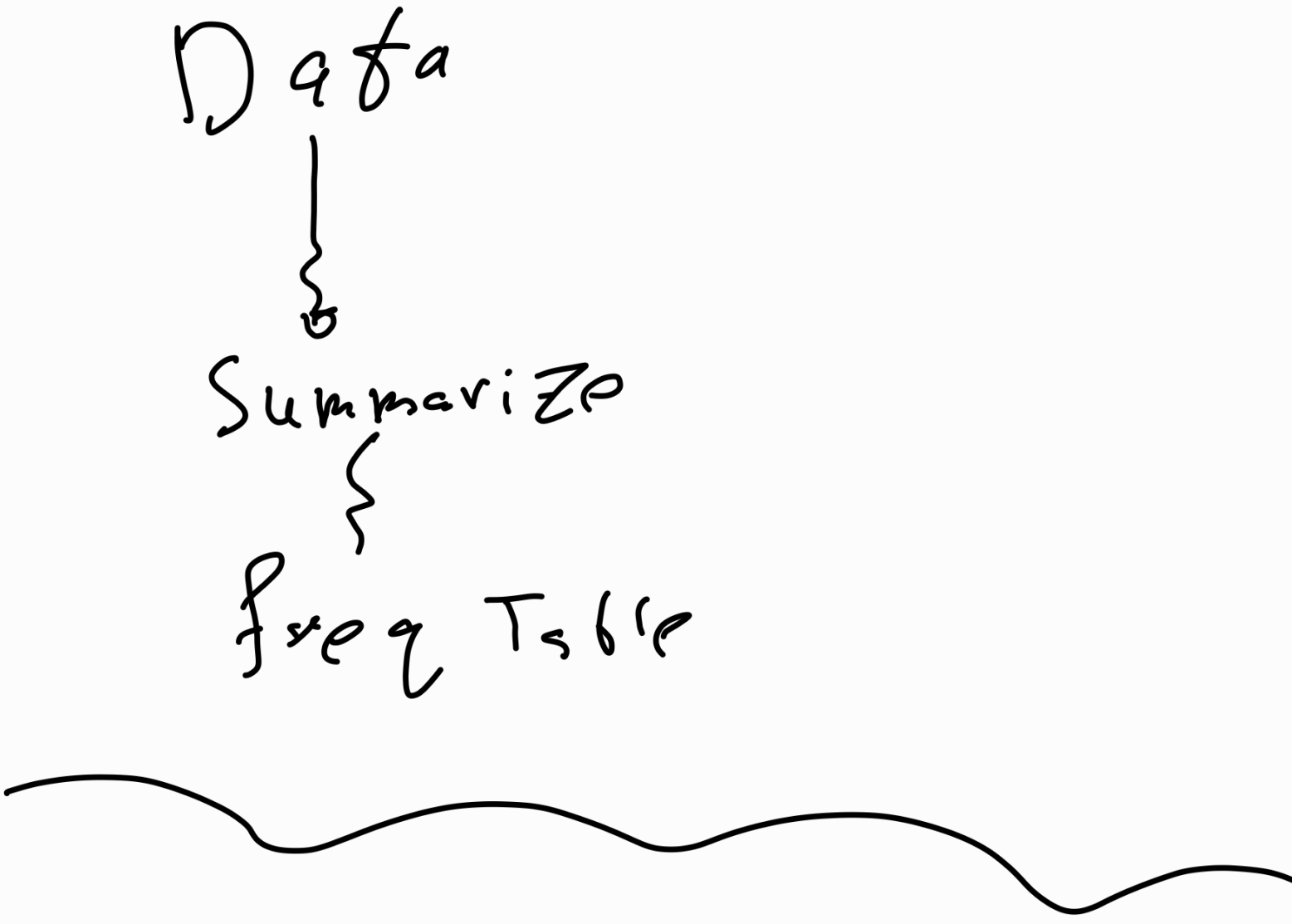
class	freq
A	10 000
B	1500
O	500
AB	8000
∑	20 000

class	freq	relative freq	percent %
A	5	$\frac{freq}{\Sigma} = \frac{5}{25} = 0.2$	r.f * 100% = 20%
B	7	0.28	28%
O	9	0.36	36%
AB	4	0.16	16%
Σ	25	1	100



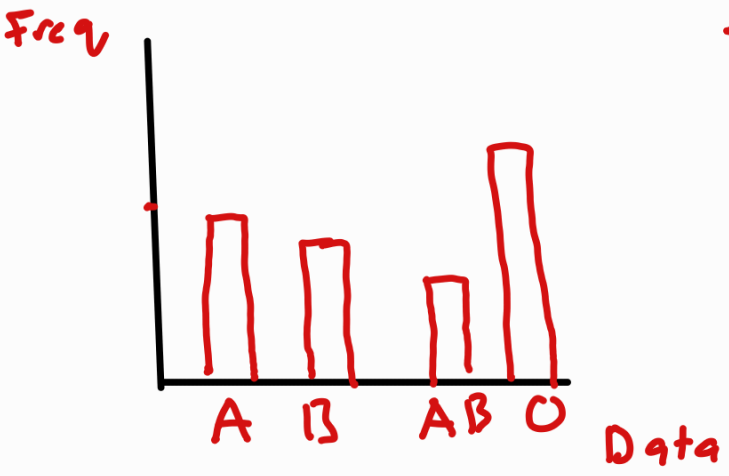
63.1, 63.2, 64, 64.3, 65.4, 65.2

weight	freq
> 60	.
60 - 70	.
71 - 80	.
:	.


Data
Summarize
Freq Table

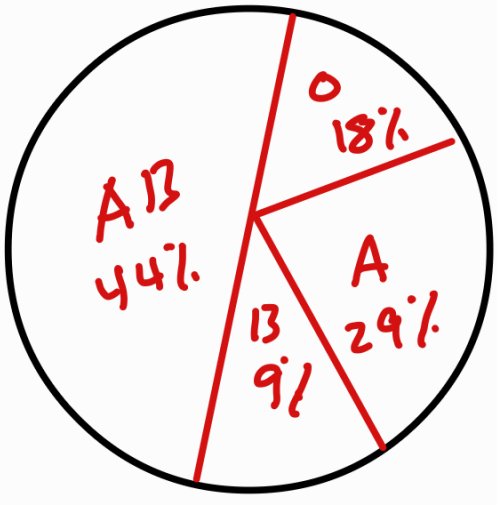
Data Visualization

1- Bar chart



- Categorical
- Discrete

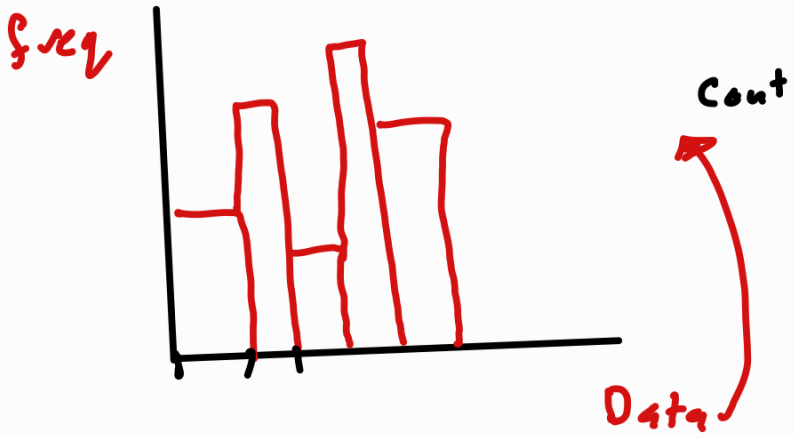
3- Pie chart :



Data & percentage

- Cat
- Discrete

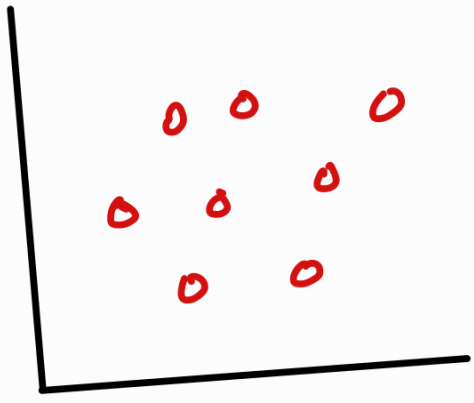
* Histogram:

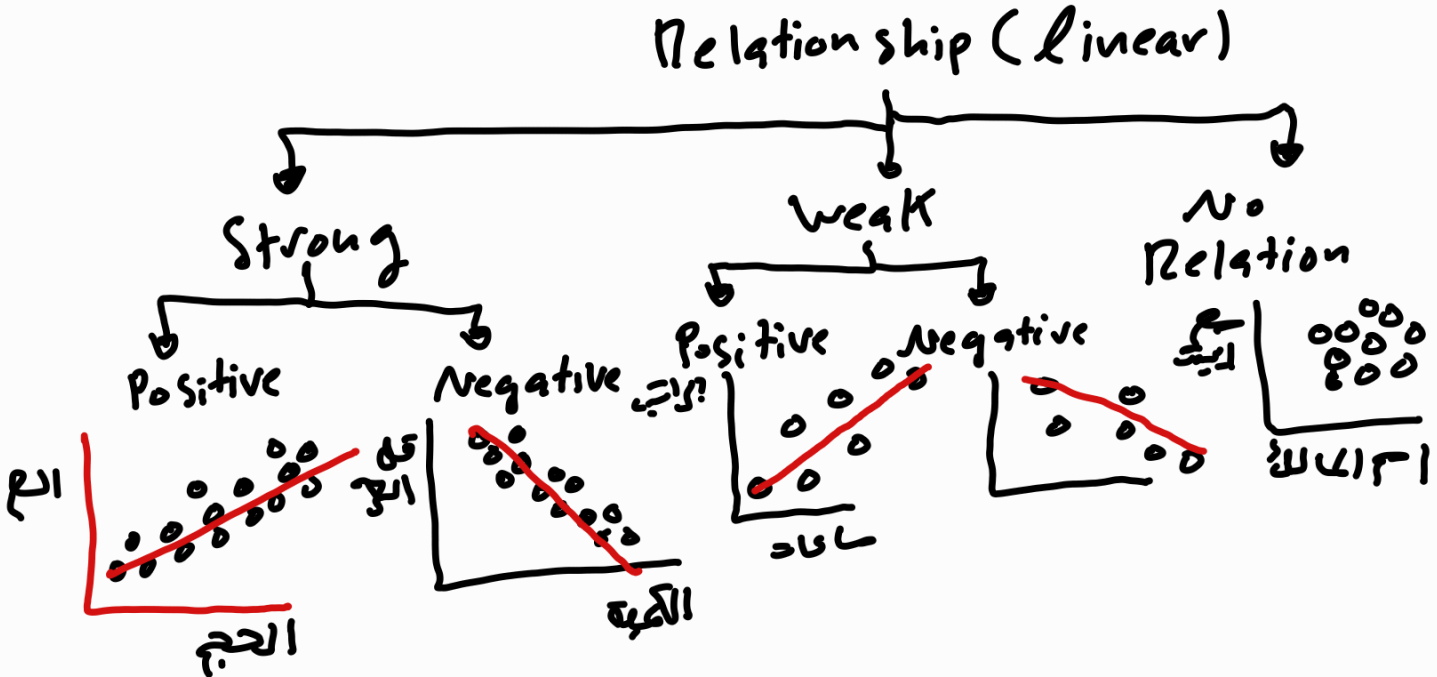


Scatter plot

→ relation between two Variable

Quantitative / Nums





Descriptive Statistical

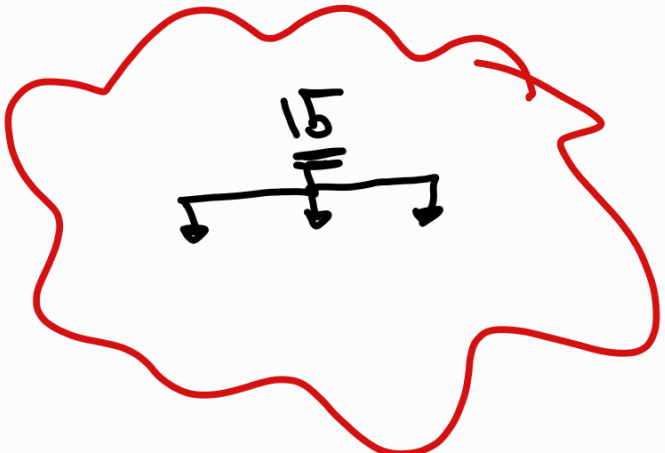
- Measures Central tendency
- Measures of Variation
- Measures of position
- Measures Skewness & Kurtosis

1- Measures of central tendency

- is a value that represents atypical, average or central of data set

→ Center of data set

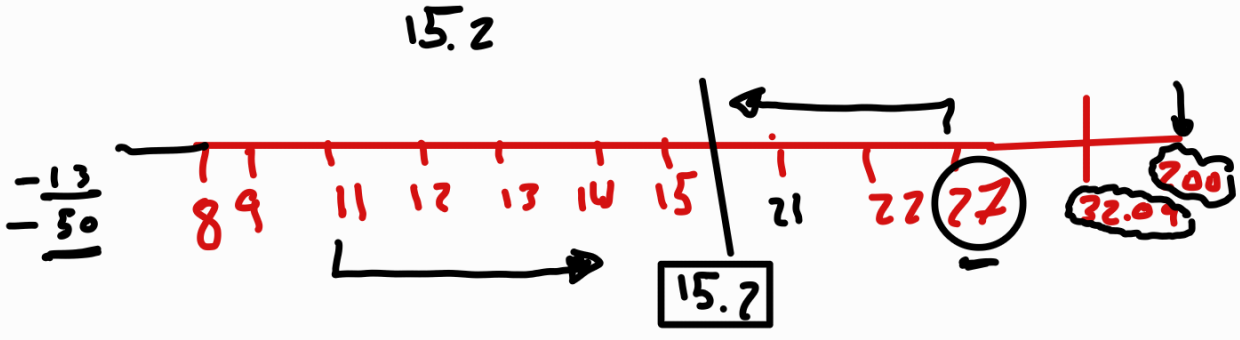
where most of the values in the data are located



1- Mean

μ : Sum of all Values / # of Values

8, 15, 22, 21, 12, 9, 11, 27, 14, 13



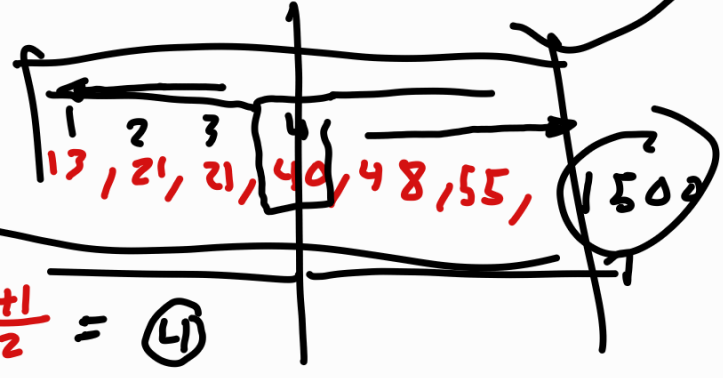
2- Median

- Middle value in the data.

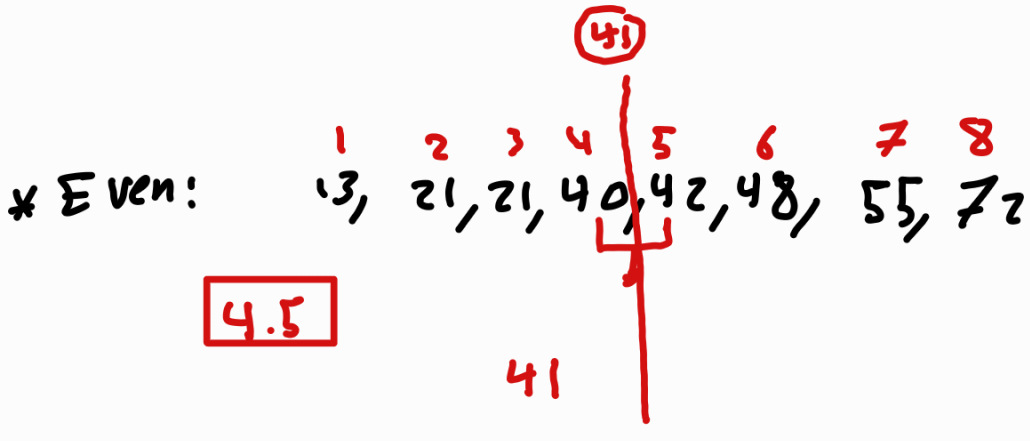
50% < ☐ < 50%

→ odd number!

ترتيب تصاعدي



Median = $\frac{n+1}{2}$ = (4)


41
* Even:
1 2 3 4 5 6 7 8
13, 21, 21, 40, 42, 48, 55, 72
4.5
41

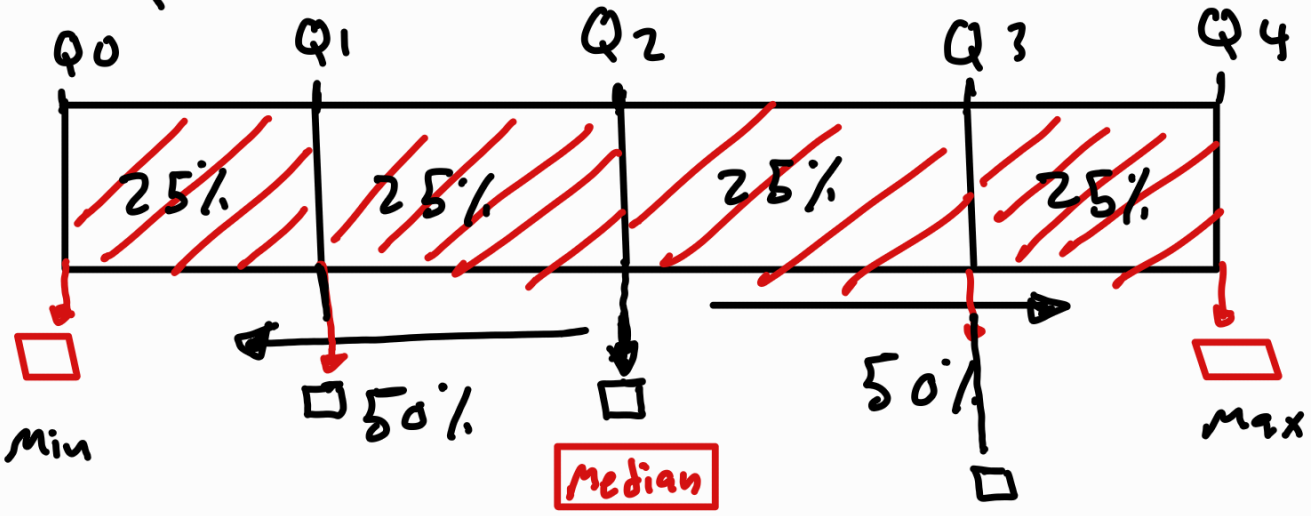
1 2 3 4

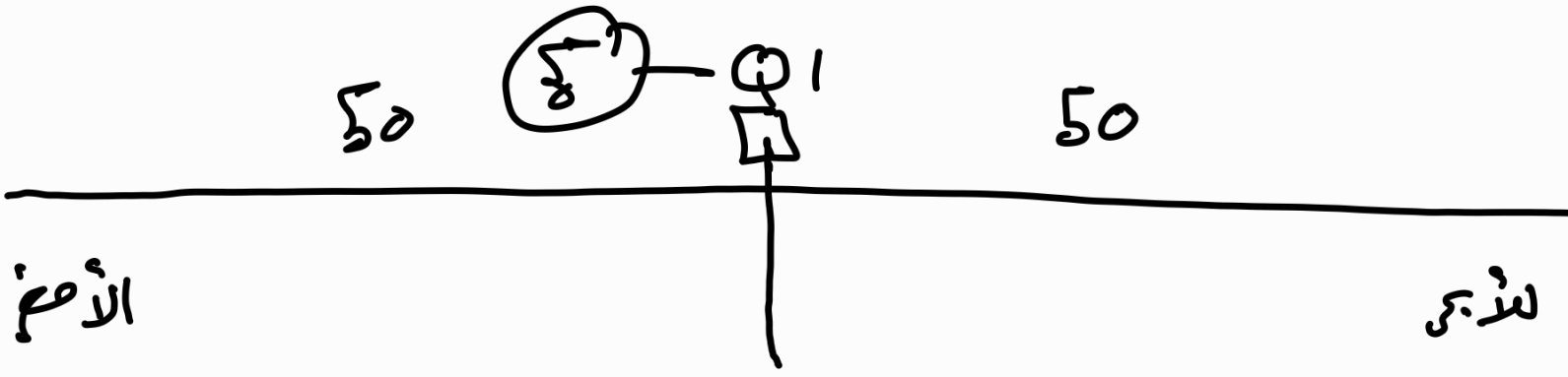
A B AB O A A B O

Median: O

Mode A

* Quartiles:



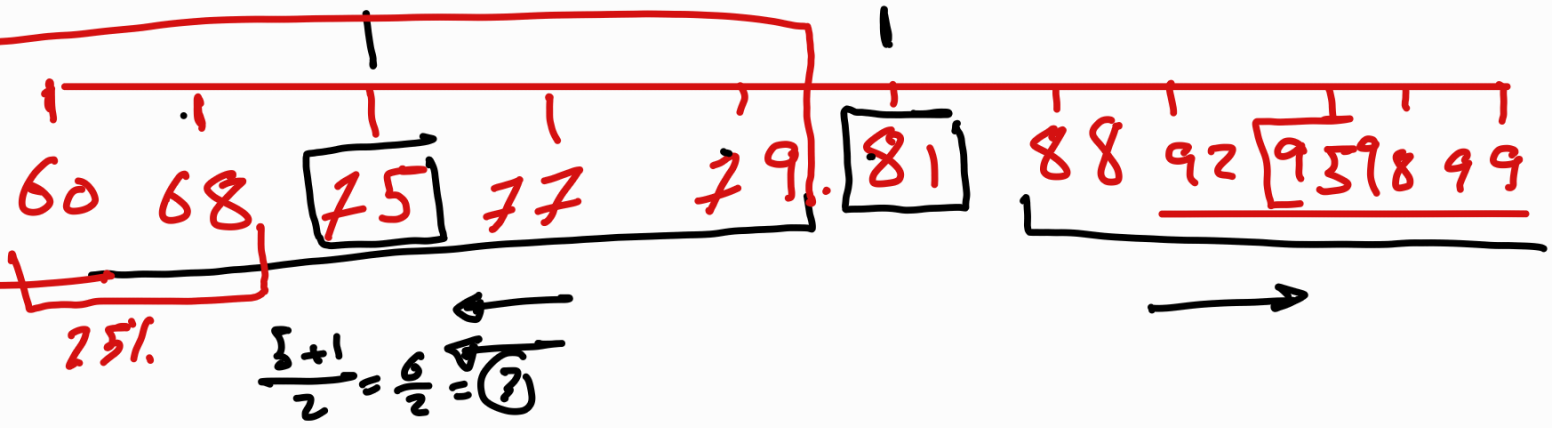


. . . 웃웃 웃웃 !!

100

99, 75, 60, 98, 88, 79, 81,

92, 77, 68, 95



Q2: Median: $\frac{n+1}{2}$ =

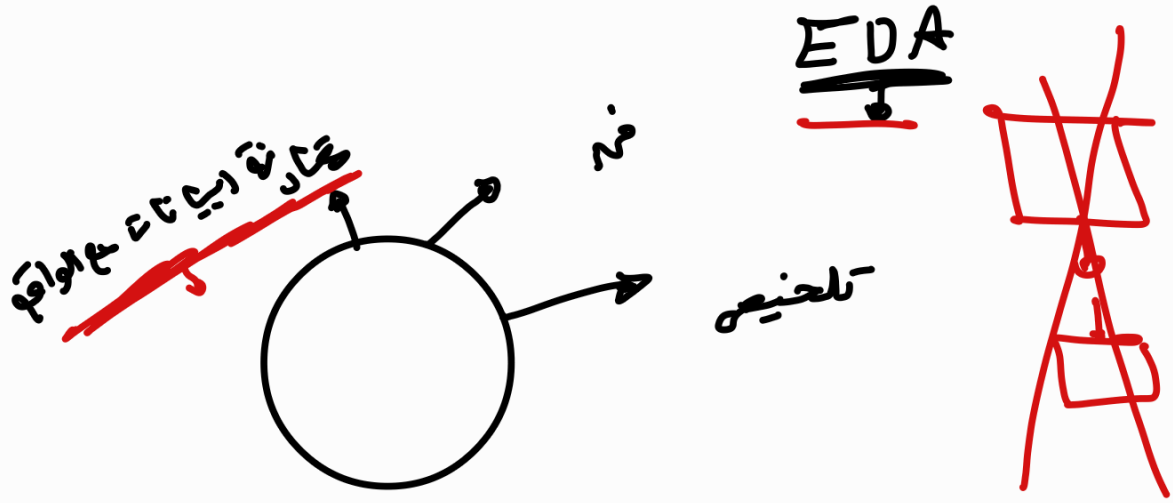
Q0: 60

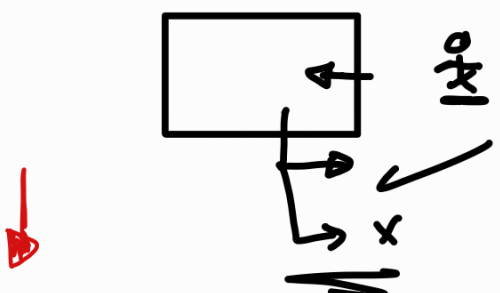
Q5: 99

Q1: 75

Q3: 95

EDA



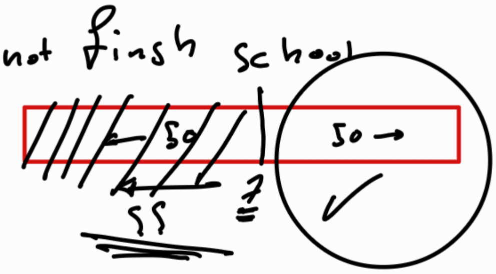


Target

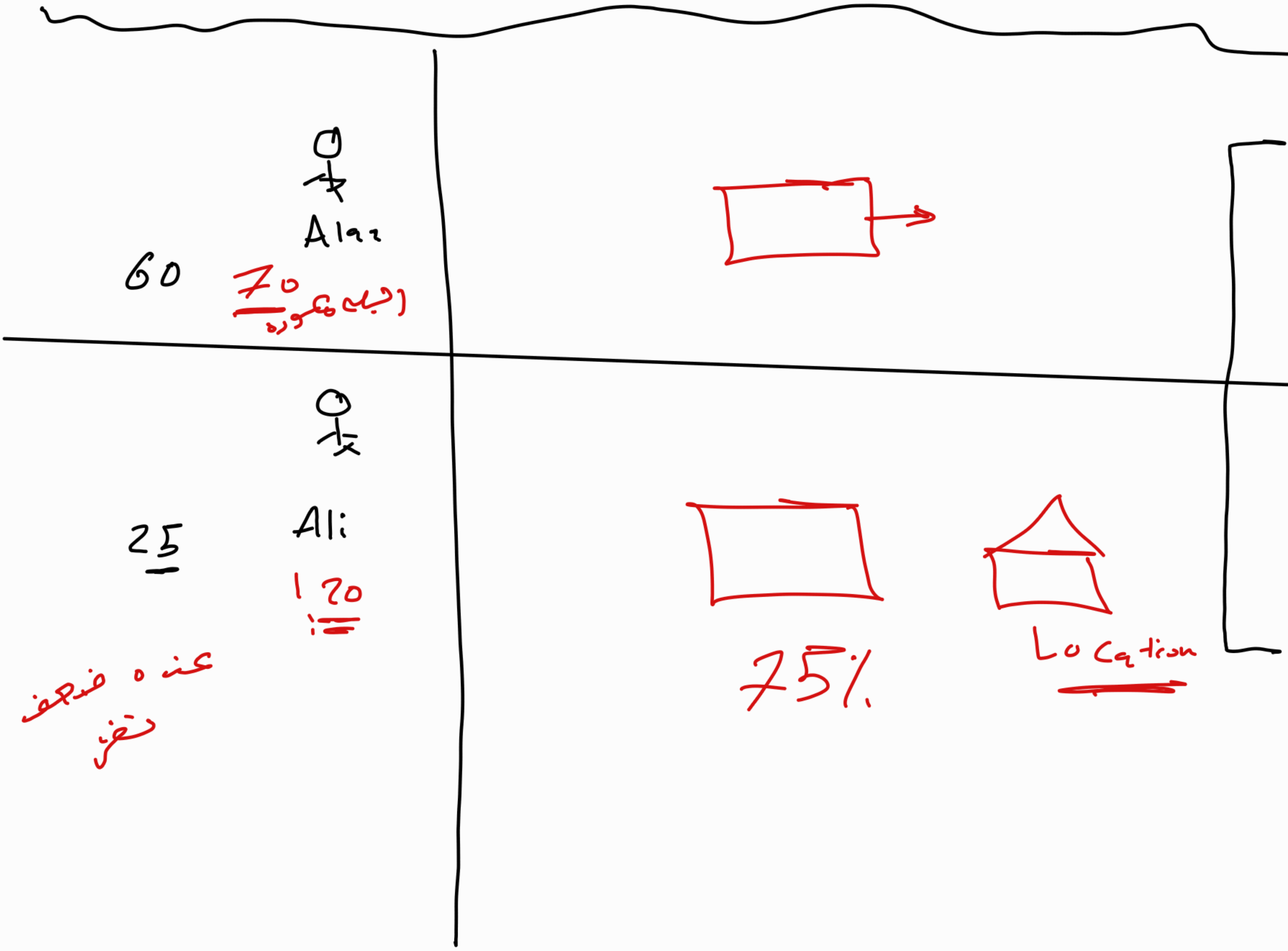
Age	مستوى التعليم	عدد سنوات العمل في المجال	عدد سنوات في المنزل للطفل	في الشهر الماضي	في وقت آخر	هل سبق ان تعرض لـ
41	BA	17	12	167	5.01	Yes
25	Diploma	0	2	32	3.09	No
27	Did not finish School	10	6	31	4	No

10 000

- Age : Mean: 35 Year
- مستوى التعليم : Mode: did not finsh school
- عدد سنوات العمل : Median: 7



- عدد سنوات السكن : Q1 : 3
- دخله السنوي : Q2 : 34
- جنس ونص : Q3 : 3,4
- كحشر ,8 : Mode: NO



mean: 8

5

5

Alaa	Ali
90	100
91	90
88	80

90 90 100

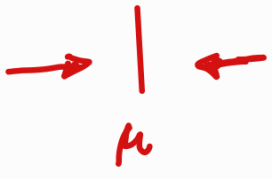
μ: 90

μ: 90

88 88 90 91

Measures of Variation

- Describe if the data close to each other or the Data Spread about each other.

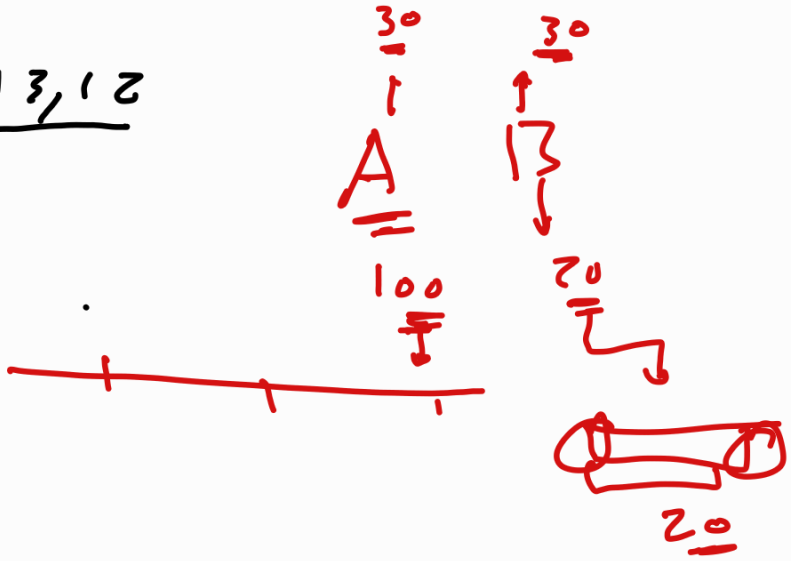


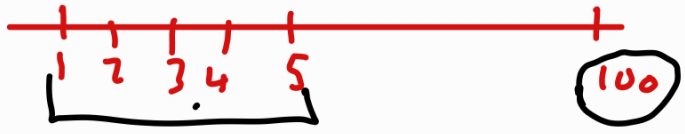
1- Range:

الفرق بين اكبر قيمة وأقل قيمة

5, 6, 10, 15, 21, 13, 12

Range: 21 - 5 = 16



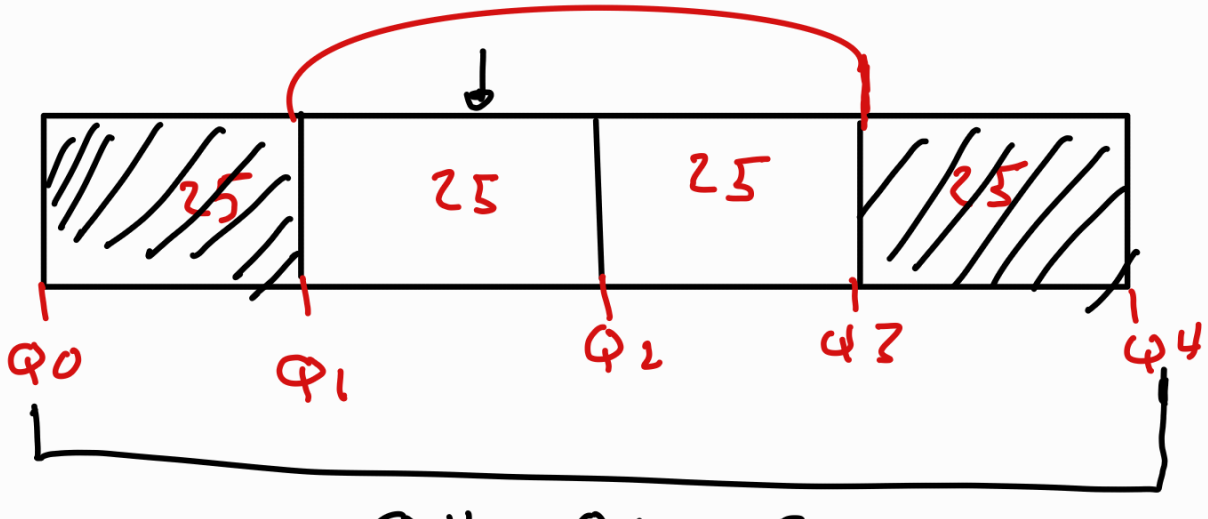


IQR

ان الـ Range يتأثر بـ outliers

IQR والحل

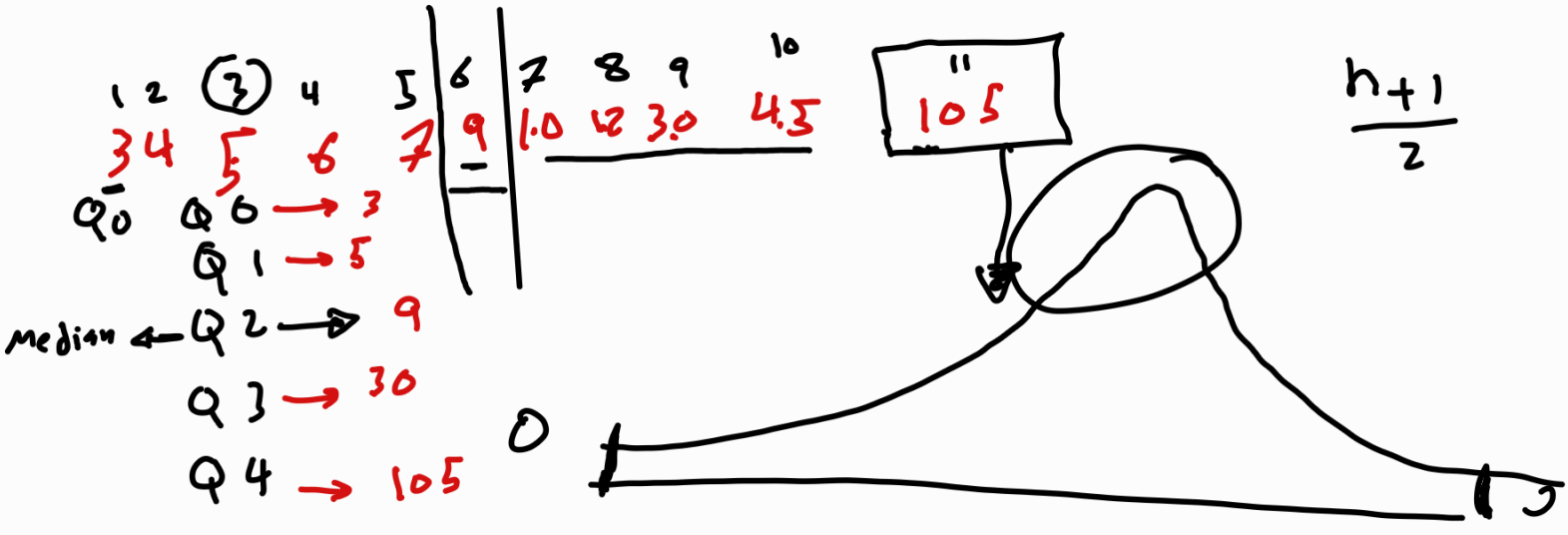
Interquartile Range ⇐ IQR: Q3 - Q1



Q4 - Q0 = Range

$Q1 - 1.5 * IQR =$

$Q3 + 1.5 * IQR$



IQR — $Q3 - Q1 = 30 - 5 = 25$

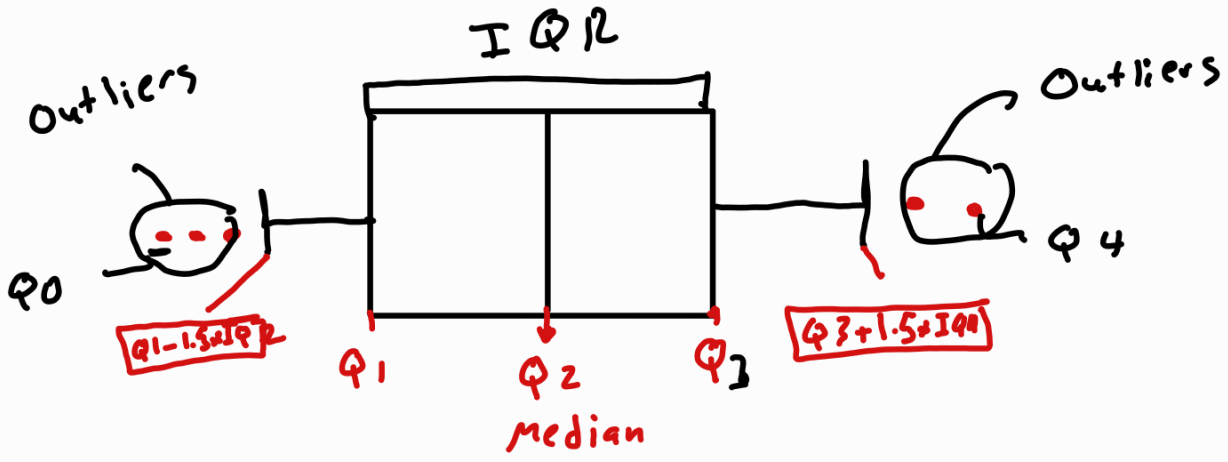
Range → $105 - 3 = 102$

Outliers

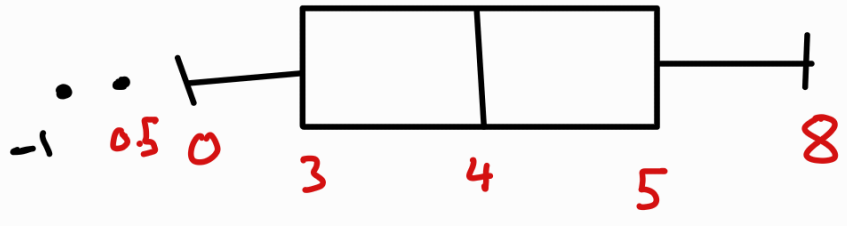
قيم شاذة outliers هي → $[Q1 - 1.5 * IQR,\ Q3 + 1.5 * IQR]$

$[-32.5, 67.5]$

Box plots



Ex: by using boxplot:



- Maximum Value: 8
- minumum Value -1
- median 4
- Q1 3
- Q3 : 5
- Range : 9
- IQR : 2
- how many outliers ? 2
- Non Outliers Interval

[0, 8]

- Outliers Value :

-1 / 0.5