DATE : 06.07.2024

DT/NT : DT

LESSON : Statistics 1 - Session 2

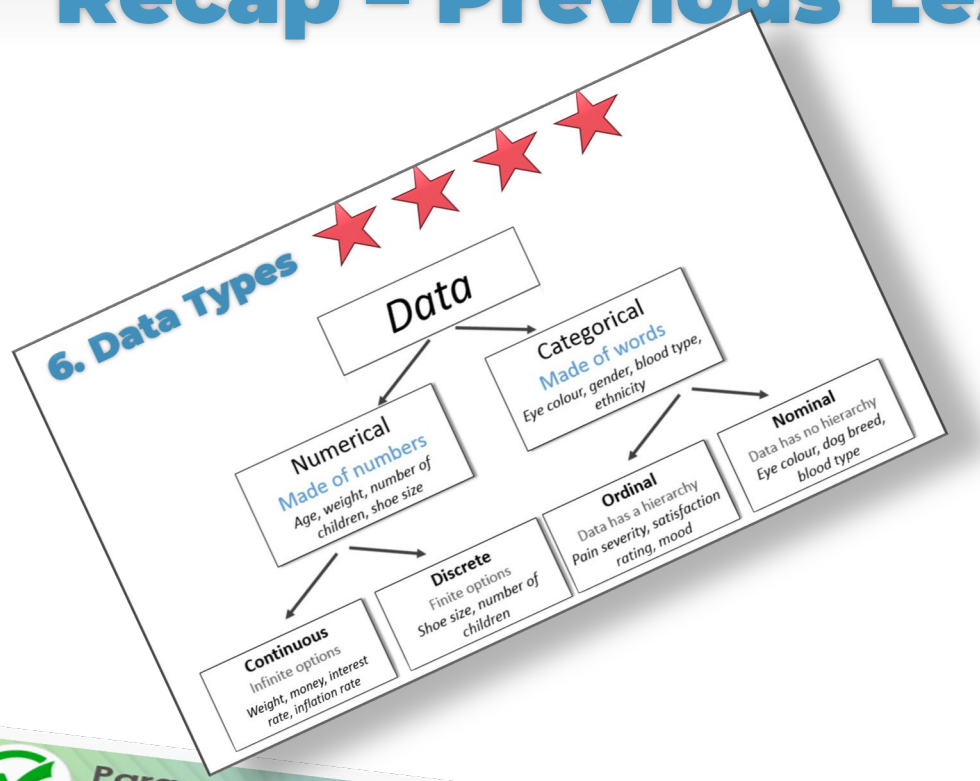
SUBJECT: Graphical Presentations
- Distributions
- Central Tendency

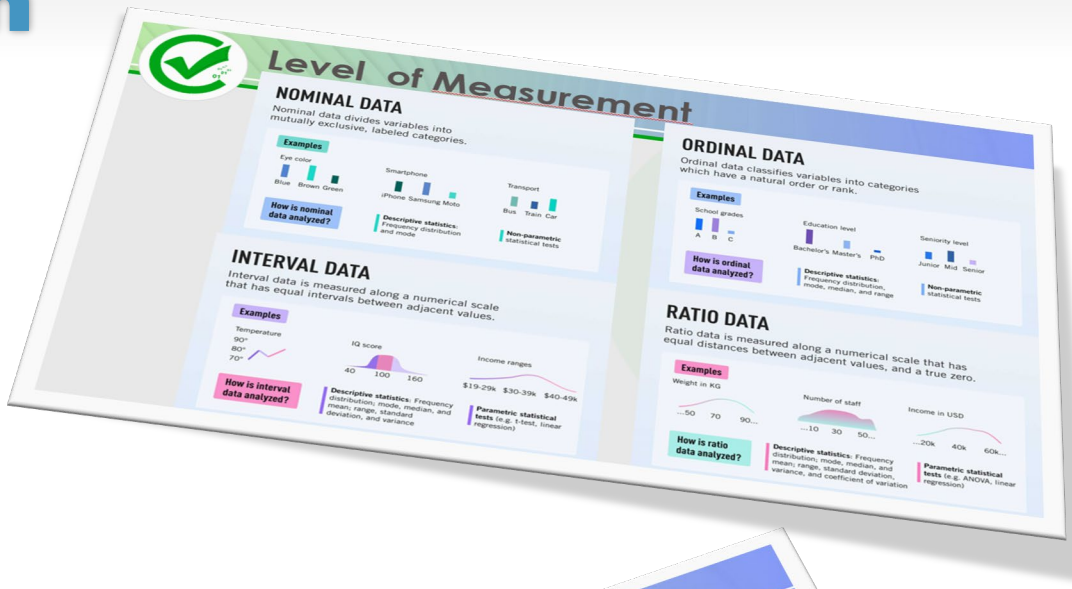
BATCH : 250

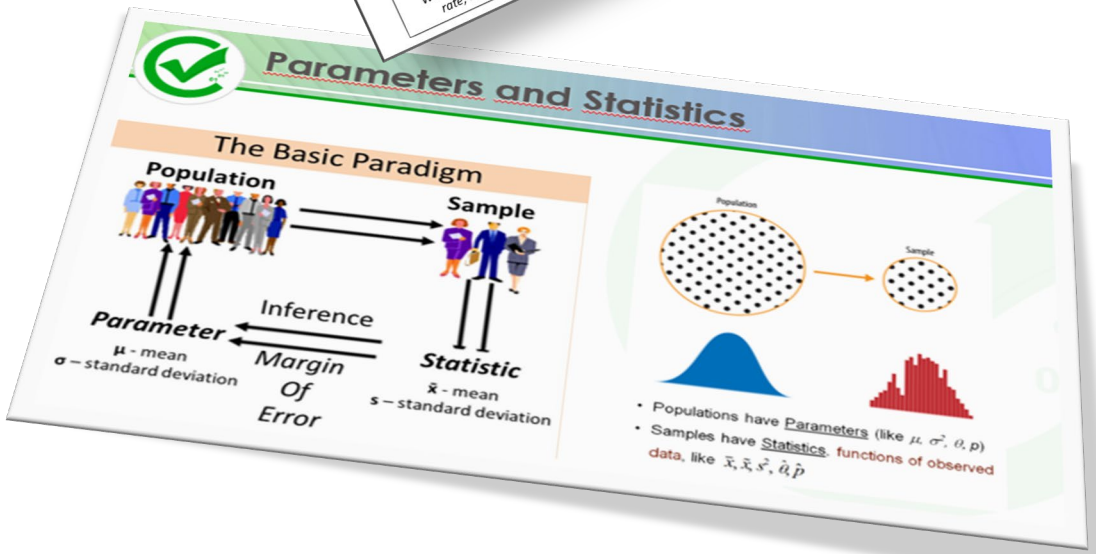
TECHPRO EDUCATION

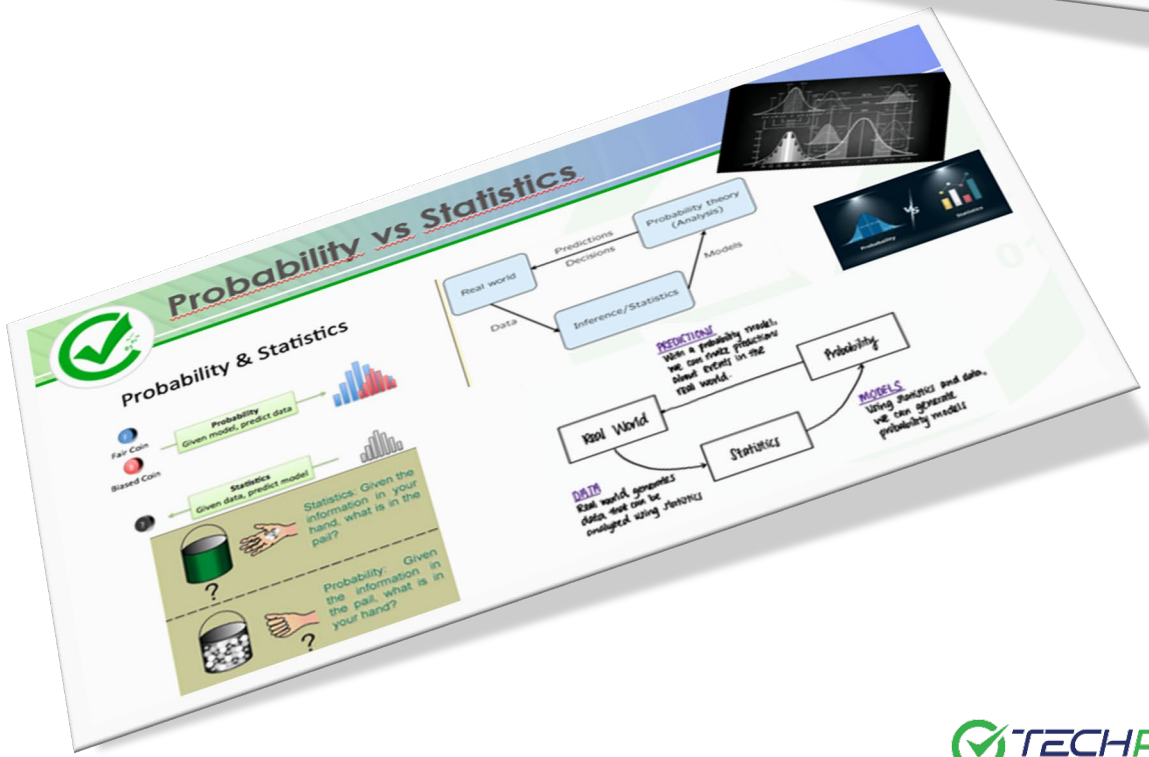
Can you write a sentence that you remember from the previous lesson?

Recap – Previous Lesson







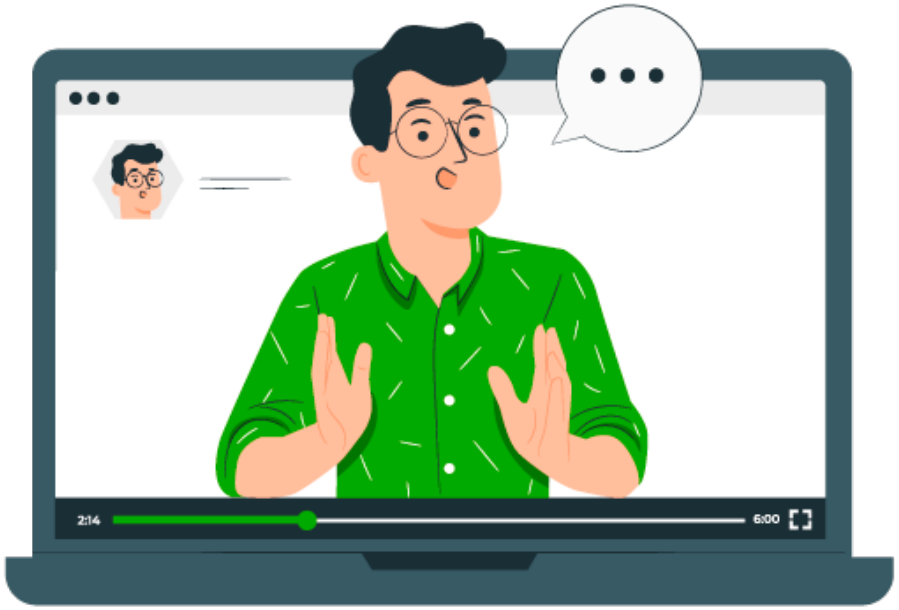


Have you practiced with the pre-class materials that will prepare you for today's lesson?

Content

- **Graphical Representation of Data**
  - Patterns,
  - Frequency Table,
  - Line chart,
  - Pie chart,
  - Bar chart,
  - Box plot
  - Histogram,
  - Distributions
- **Central Tendency**
  - Mod,
  - Median,
  - Mean,
  - Range,
  - IQR
  - Standard Deviation

Session - 2 Content



Graphical Represent

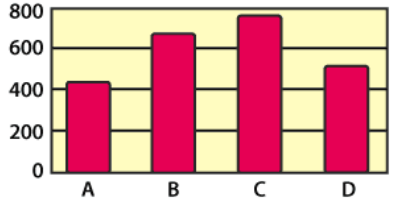
- Patterns
- Distributions
- Frequency Table
- Bar Chart
- Pie Chart
- Line Chart
- Histogram
- Box Plot

What will we learn today?

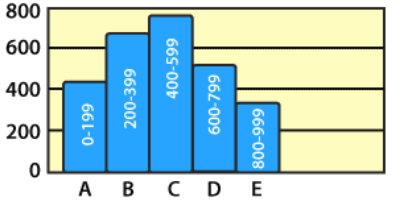
Data Visualization - Graphical Represent

- Graphical Representation of Data
  - Center
  - Spread
  - Shape
  - Unusual Features

Bar Graphs



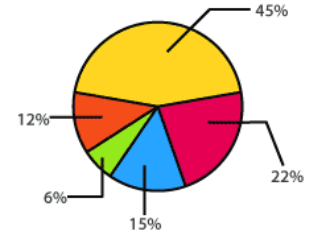
Histograms



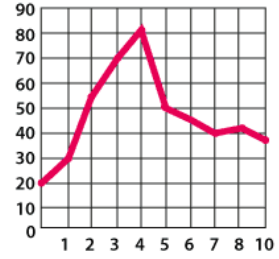
Frequency Table

Rulers of France		
Reign (Years)	Tally	Frequency
1-15	𝍸 𝍸 𝍸 III	18
16-30	𝍸 𝍸 I	11
31-45	𝍸 I	6
46-60	IIII	4
61-75	I	1

Circle Graph



Line Graphs

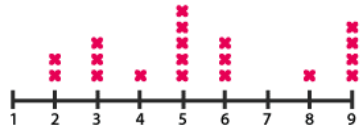


Stem and Leaf Plot

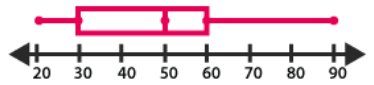
Stem	Leaf
0	1, 1, 2, 2, 3, 4, 4, 4, 4, 5, 8
1	0, 0, 0, 1, 1, 3, 7, 9
2	5, 5, 7, 7, 8, 8, 9, 9
3	0, 1, 1, 1, 2, 2, 2, 4, 5
4	0, 4, 8, 9
5	2, 6, 7, 7, 8
6	3, 6

Key : 6 | 3 = 63 Year

Line Plot

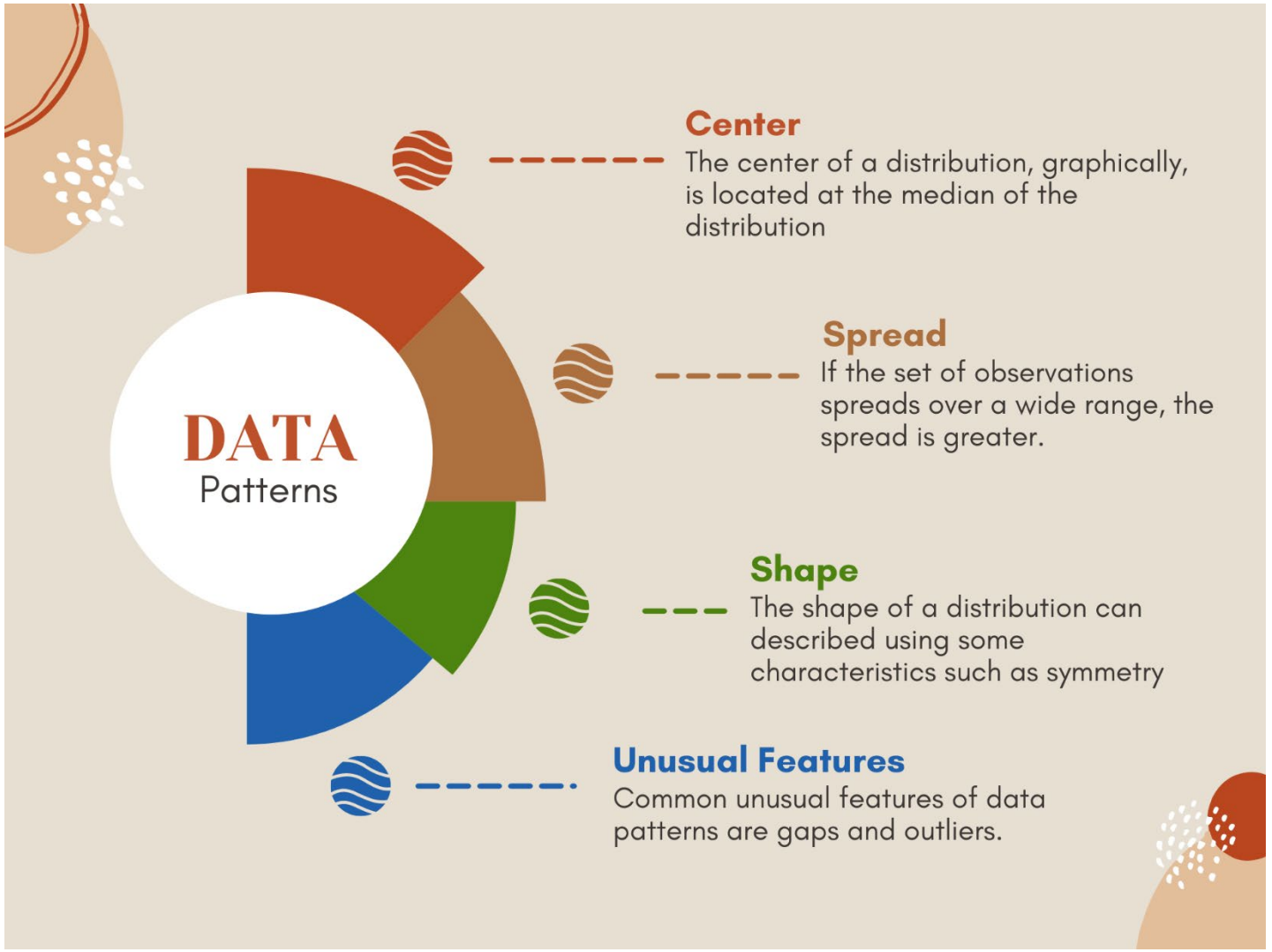


Box and Whisker Plot



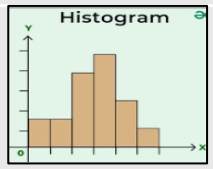
Data Patterns

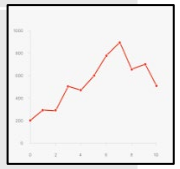
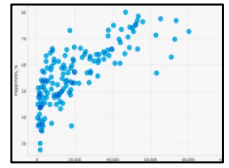
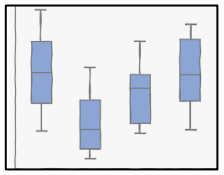
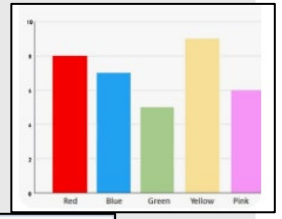
- Data Patterns
  - Center
  - Spread
  - Shape
    - Symmetric
    - Number of peaks
    - Skewness
    - Uniform
  - Unusual Features
    - Gaps
    - Outliers

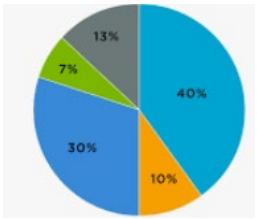


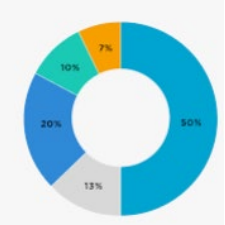
Common data types and corresponding chart types

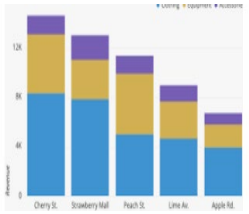
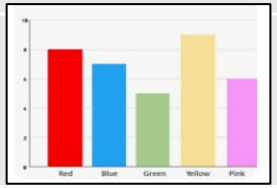
Data Type	Recommended Chart Types
Numeric	Histogram, Line Chart, Scatter Plot, Box Plot
Categorical	Bar Chart, Pie Chart, Donut Chart, Stacked Bar Chart
Ordinal	Bar Chart, Dot Plot, Heatmap
Time Series	Line Chart, Area Chart, Heatmap
Hierarchical	Treemap, Sunburst Chart, Nesting Chart

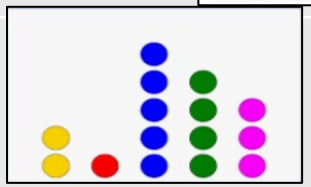
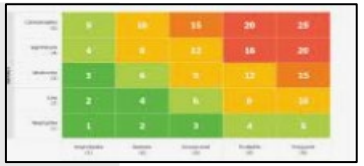
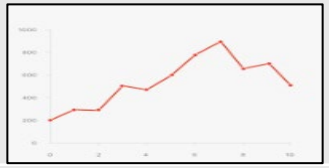
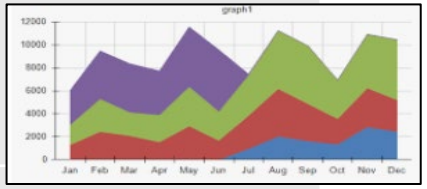


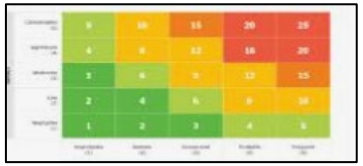







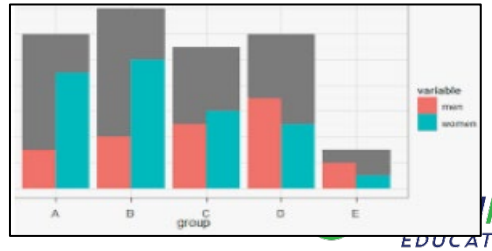




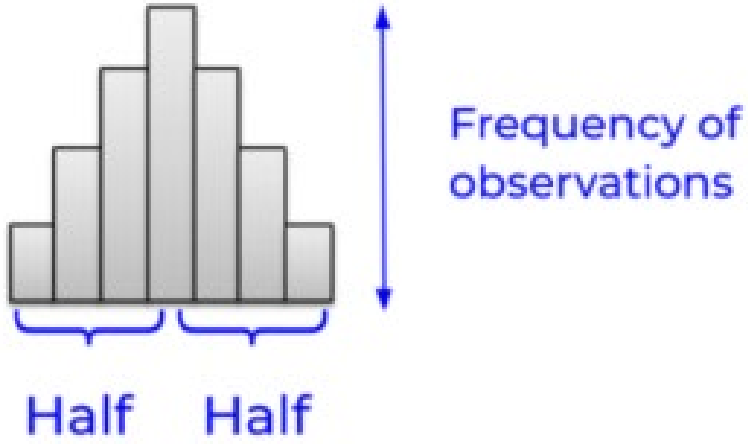




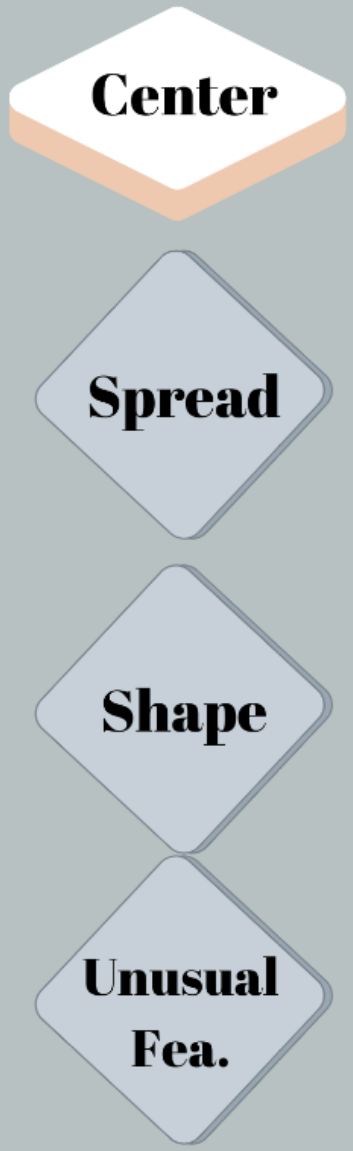
Graphical Representation of Data

Center

- **The center** of the distribution is graphical' at the **median** of the distribution.
- **Half** of the observations are on both sides.
- The **height** of the column indicates the **frequency of observations**.



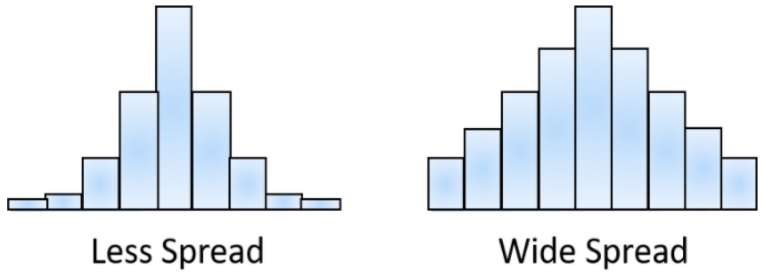
Histogram

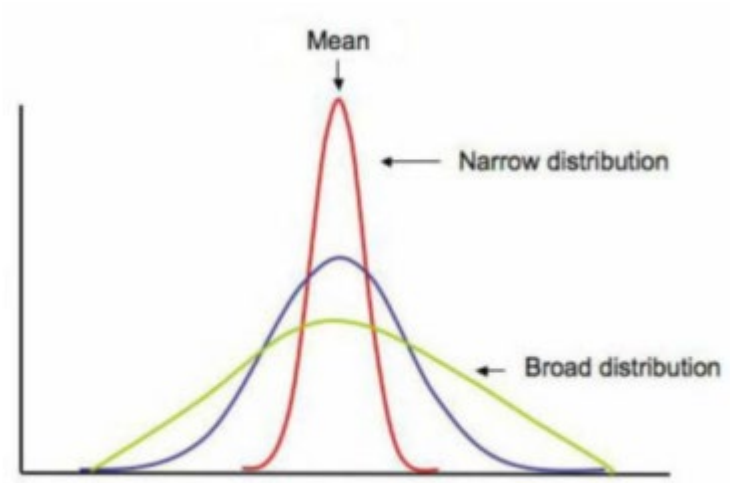


Graphical Representation of Data

Spread

- Variation of data
- If the set of observations spans a wide range .....
- If observations are centered around a single value over a narrower range.....





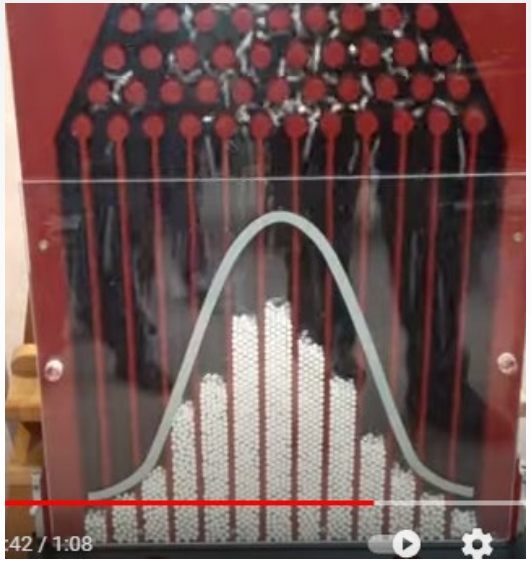
Normally Distribution Videos

Video-1

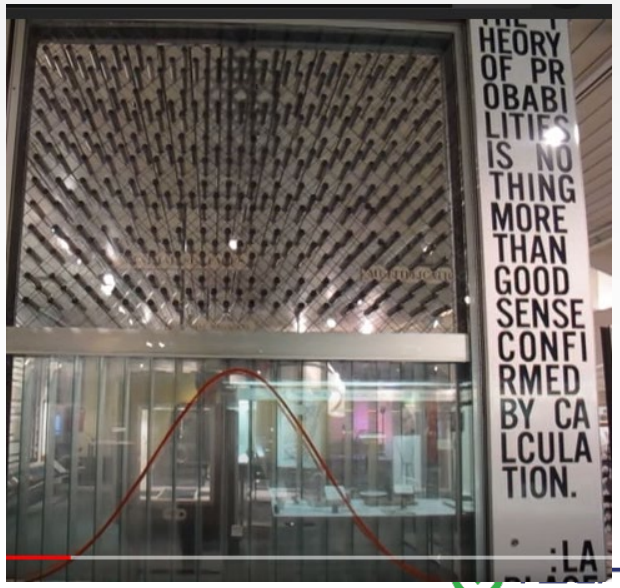
- https://www.youtube.com/watch?v=Bampgm0HKDU

Video-2

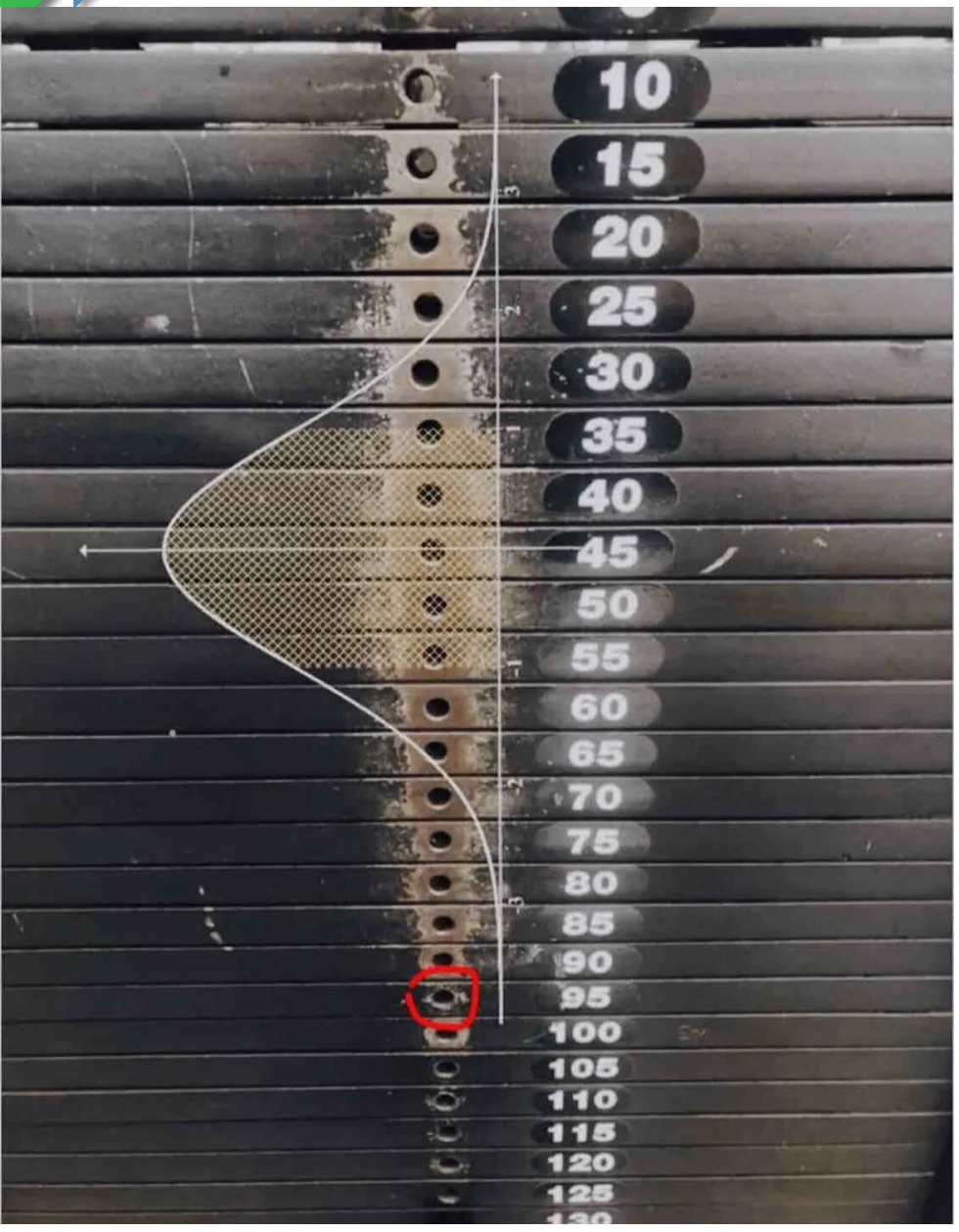
- https://www.youtube.com/watch?v=4HpvBZnHOVI



Video-3



Normal distribution is everywhere..

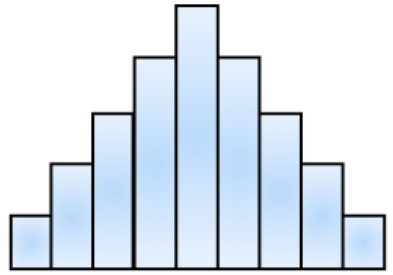


Graphical Representation of Date

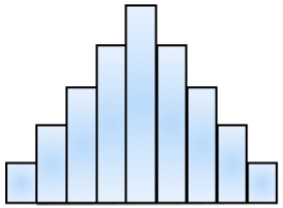
Shape

The shape of a distribution can be defined using the following properties.

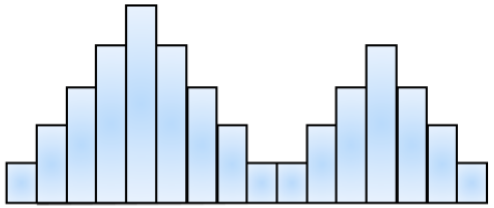
- Symmetric
- Number of Peaks
- Skewness
- Uniform



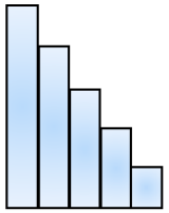
Symmetric



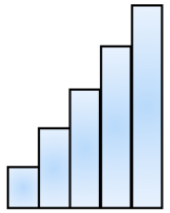
Unimodal,
Bell Shaped



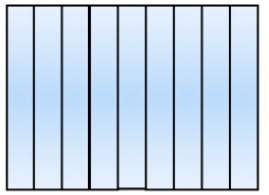
Bimodal



Skewed Right



Skewed Left



Uniform

Center

Spread

Shape

Unusual Fea.

Probability distributions

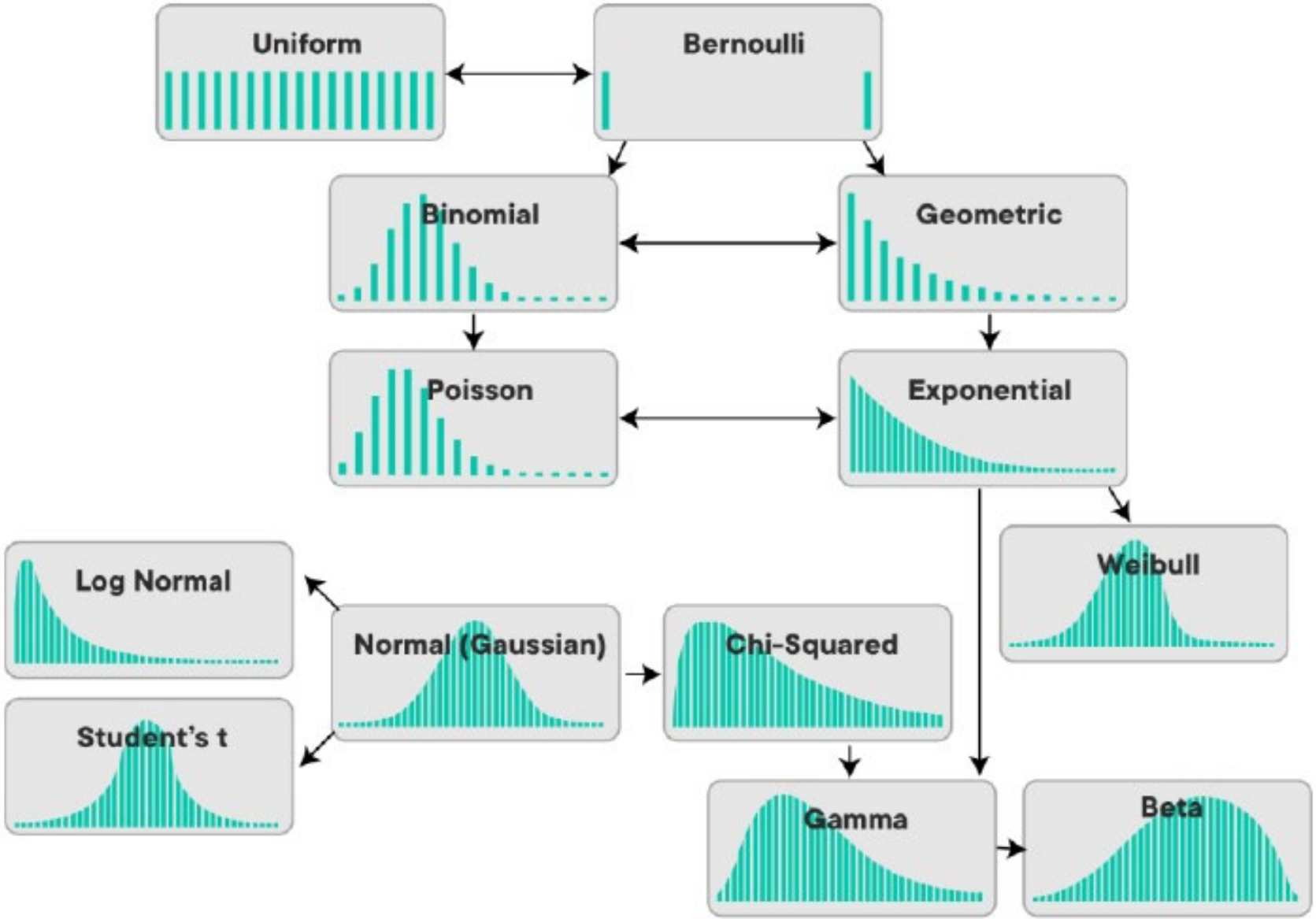
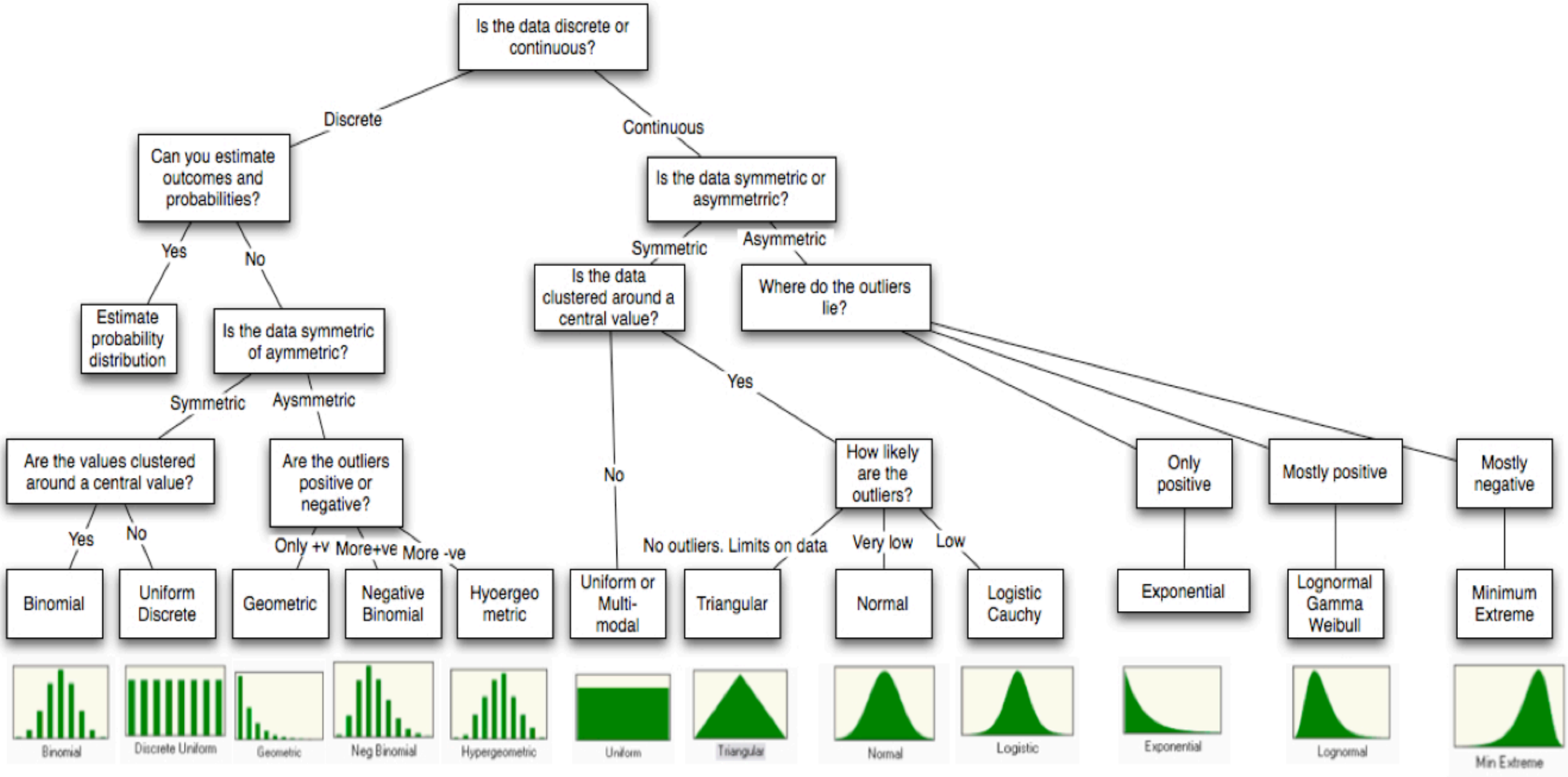


Figure 6A.15: Distributional Choices

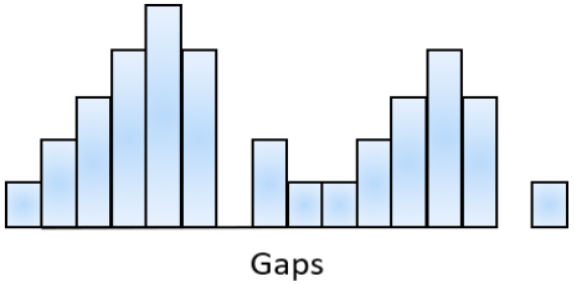


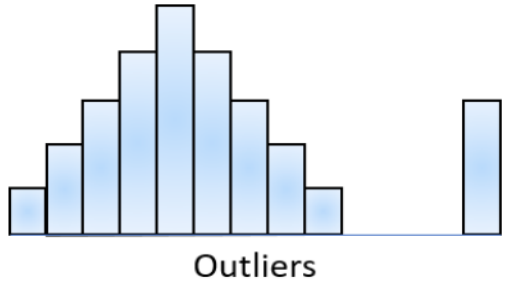
Graphical Representation of Data

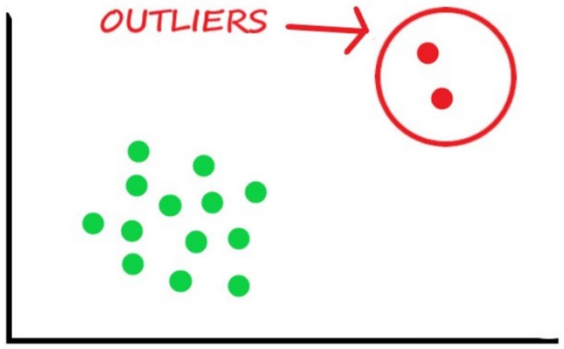
Unusual Features

Common unusual features of data models are:

- Gaps
- Outliers







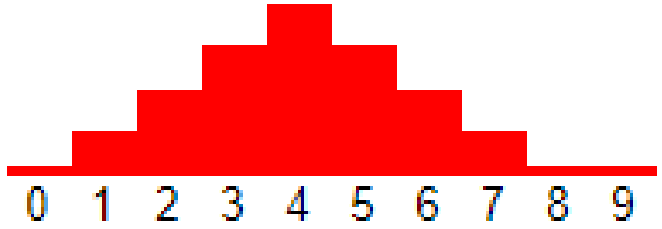
Center

Spread

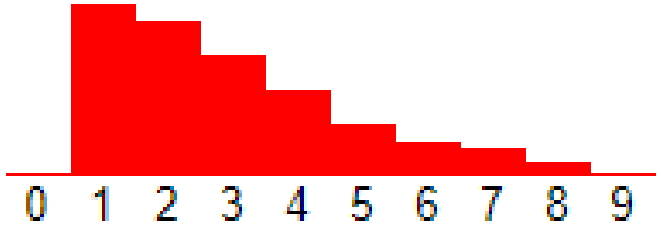
Shape

Unusual Fea.

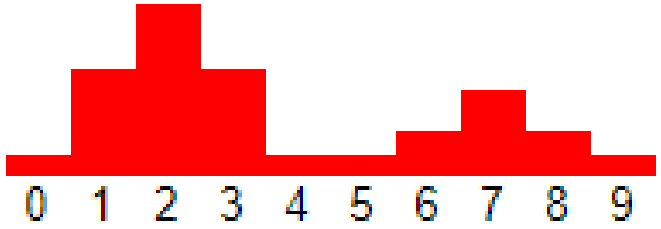
Data Patterns



Symmetric, unimodal, bell-shaped

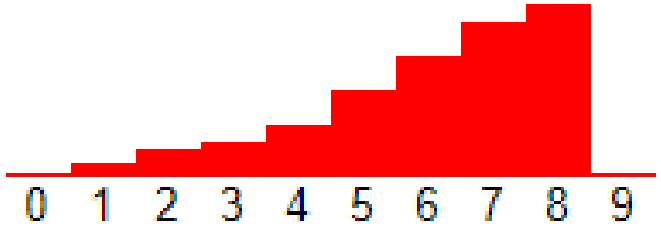


Skewed right

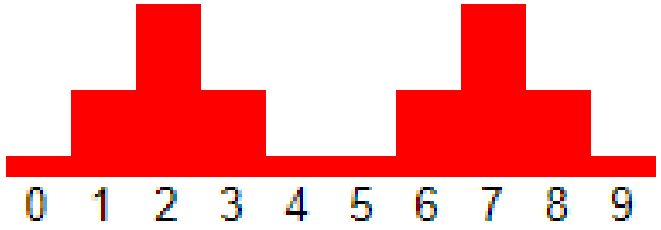


Non-symmetric, bimodal

Uniform



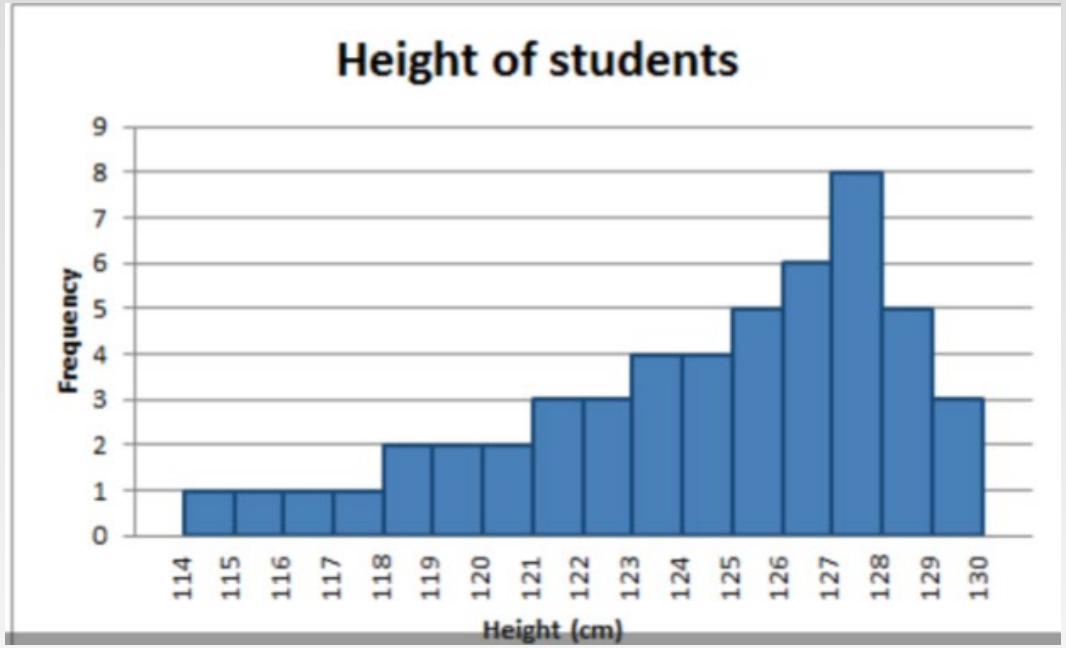
Skewed left



Symmetric, bimodal

Pear Deck

TEST



Which pattern is correct for this figure?

a. Right-skewed with no outliers

b. Right-skewed with one outliers

c. Left-skewed with no outliers

d. Symmetric

Task -2

Task 2: Create a frequency histogram from the data in the table below. What you can conclude about the shape of the distribution?

Income (In thousands of dollars)	Number of families
16-22	2
23-29	3
30-36	5
37-43	8
44-50	8
51-57	10

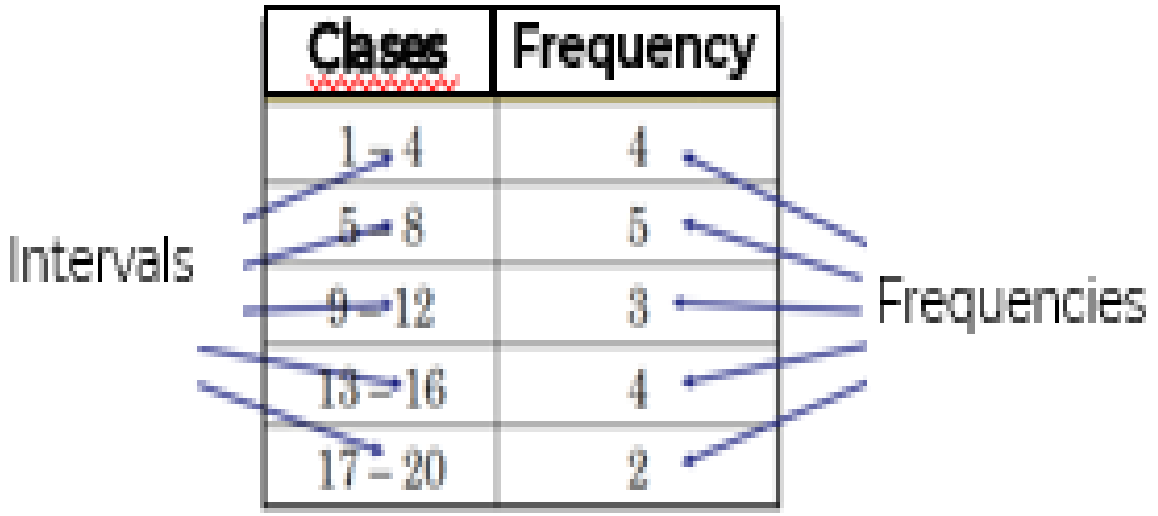
Excel

Frequency

Methods used in descriptive statistics:

- Frequency Tables
- Shapes and Graphics
- Histogram and Frequency Polygons
- Column and Pie Charts

Developer Type	Frequency	Relative Frequency
Front-end Developer	25	0.25
Backend Developer	15	0.15
Full-stack Developer	20	0.20
Data Scientist	40	0.40



Frequency

Frequency

- Number of occurrences of a data value

Relative Frequency

- How often something happens divided by all outcomes

Cumulative Frequency

- Accumulation of previous relative frequencies

DATA VALUE	FREQUENCY
3	5
4	3
5	6
6	2
7	1

DATA VALUE	FREQUENCY	RELATIVE FREQUENCY
2	3	$\frac{3}{20}$ or 0.15
3	5	$\frac{5}{20}$ or 0.25
4	3	$\frac{3}{20}$ or 0.15
5	6	$\frac{6}{20}$ or 0.30
6	2	$\frac{2}{20}$ or 0.10
7	1	$\frac{1}{20}$ or 0.05

DATA VALUE	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
2	3	$\frac{3}{20}$ or 0.15	0.15
3	5	$\frac{5}{20}$ or 0.25	0.15 + 0.25 = 0.40
4	3	$\frac{3}{20}$ or 0.15	0.40 + 0.15 = 0.55
5	6	$\frac{6}{20}$ or 0.30	0.55 + 0.30 = 0.85
6	2	$\frac{2}{20}$ or 0.10	0.85 + 0.10 = 0.95
7	1	$\frac{1}{20}$ or 0.05	0.95 + 0.05 = 1.00

QUESTION

Age	Frequency	Relative Frequency	Cumulative Relative Frequency
0	2	$\frac{2}{19}$	0.1053
2	3	$\frac{3}{19}$	0.2632
4	1	$\frac{1}{19}$	0.3158
5	3	$\frac{3}{19}$	0.4737
7	2	$\frac{2}{19}$	0.5789
10	2	$\frac{2}{19}$	0.6842
12	2	$\frac{2}{19}$	0.7895
15	1	$\frac{1}{19}$	0.8421
20	1	$\frac{1}{19}$	1.0000

What is the proportion of dogs living up to 12 years (at most)?

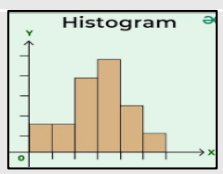
Graphs and Charts

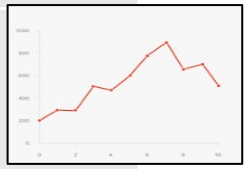
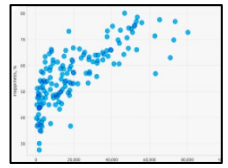
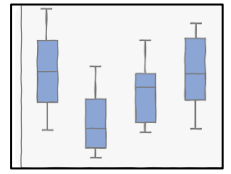
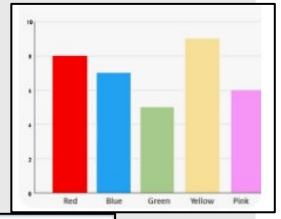
Why Charts ?

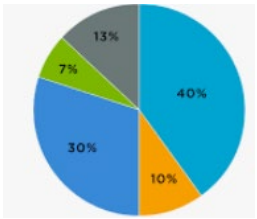
- Makes more understandable/readable.
- Points of interest are indicated.
- Provides information about the shape of the distribution.
- Prediction becomes easier.

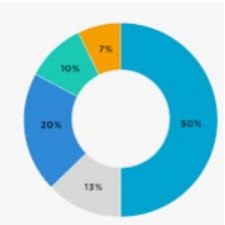
Common data types and corresponding chart types

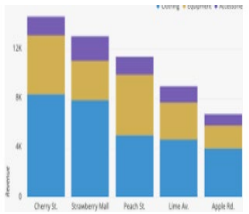
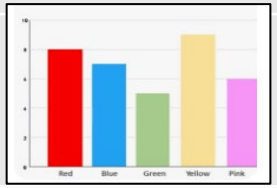
Data Type	Recommended Chart Types
Numeric	Histogram, Line Chart, Scatter Plot, Box Plot
Categorical	Bar Chart, Pie Chart, Donut Chart, Stacked Bar Chart
Ordinal	Bar Chart, Dot Plot, Heatmap
Time Series	Line Chart, Area Chart, Heatmap
Hierarchical	Treemap, Sunburst Chart, Nesting Chart

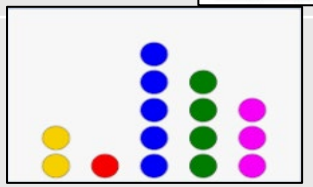
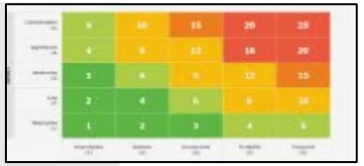
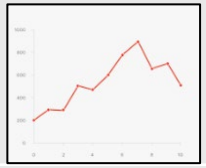
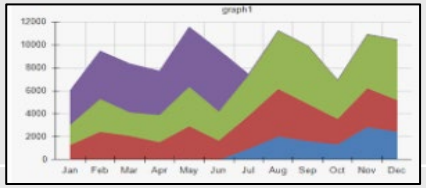


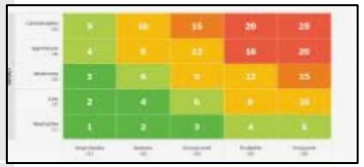







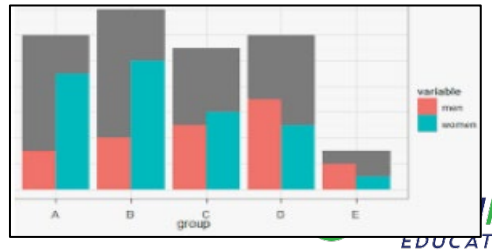


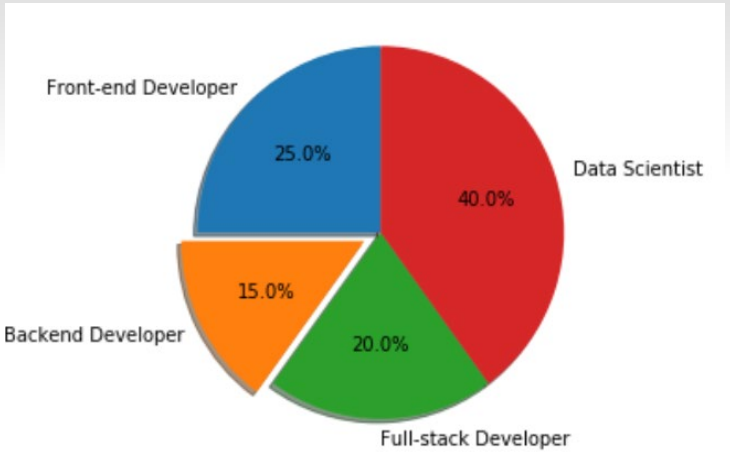











Pie Charts

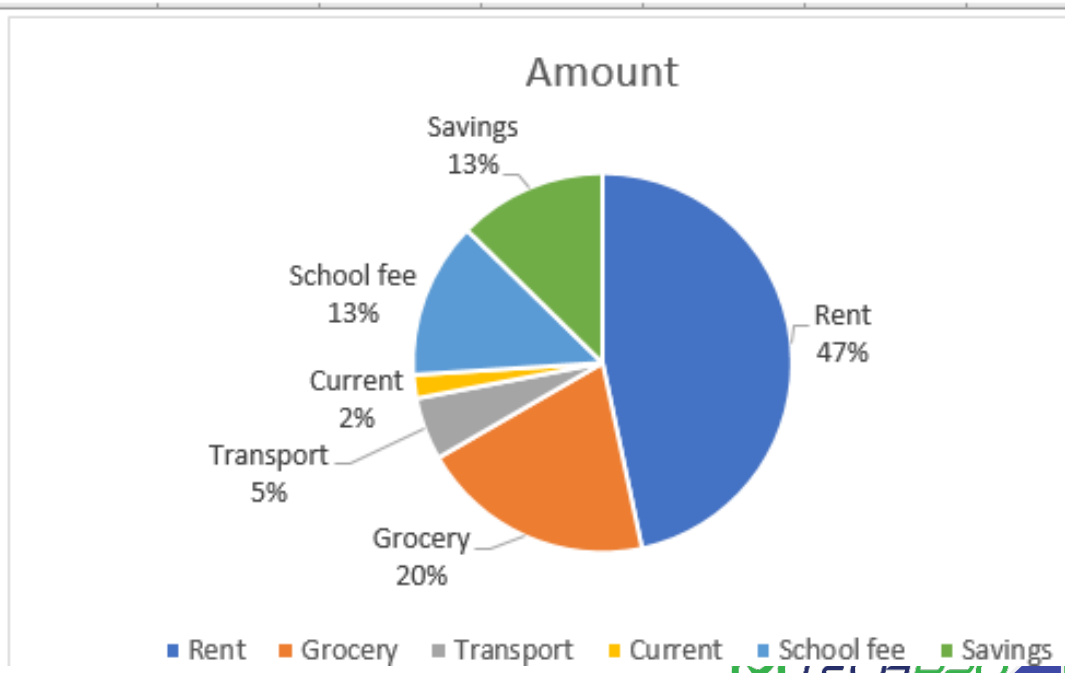


Pie Chart

- For **categorical data!**
- Usually used for **nominal** and **ordinal variables.**
- The circle is sliced into pie slices to complete 100% of the total.
- Each slice presents the attribute of the variable.

Pie Chart Examples

	Expenses	Amount
2	Rent	7000
3	Grocery	3000
4	Transport	800
5	Current	300
6	School fee	2000
7	Savings	1900

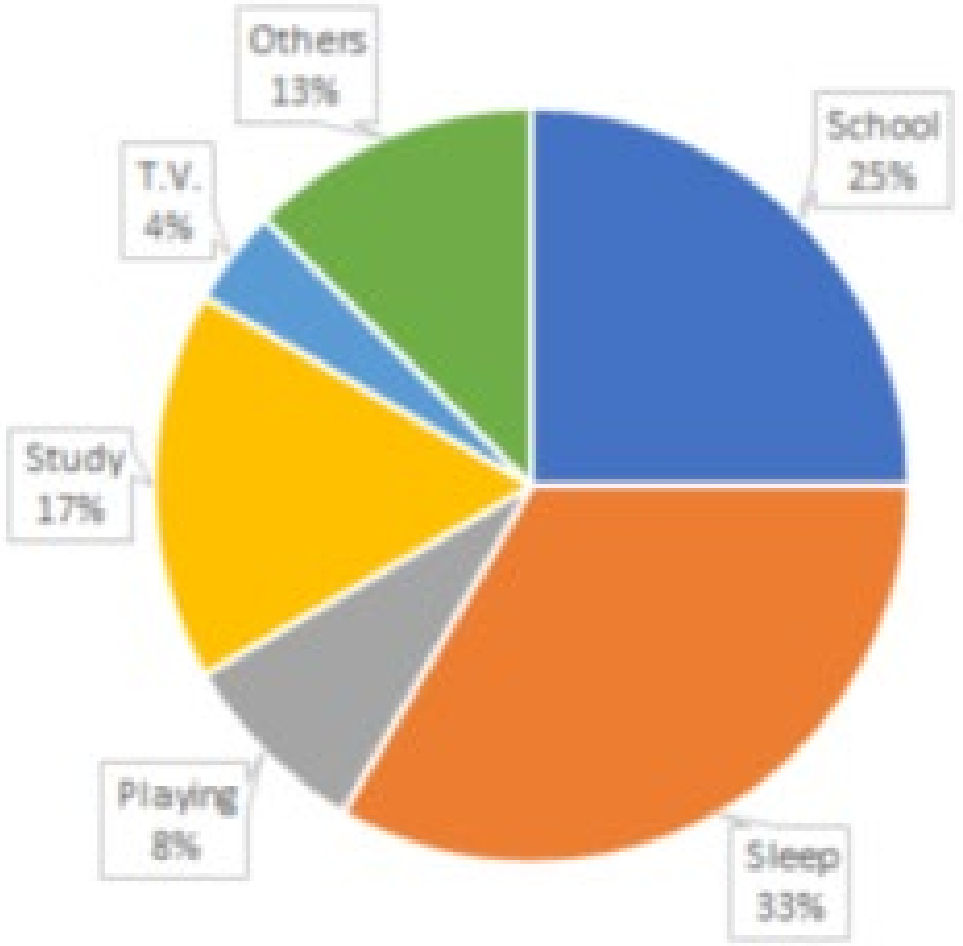


Pie Chart

Example

- Calculating Pie Percentage

Activity	No. of Hours	Measure of central angle
School	6	$(^{6}/_{24} \times 360)° = 90°$
Sleep	8	$(^{8}/_{24} \times 360)° = 120°$
Playing	2	$(^{2}/_{24} \times 360)° = 30°$
Study	4	$(^{4}/_{24} \times 360)° = 60°$
T. V.	1	$(^{1}/_{24} \times 360)° = 15°$
Others	3	$(^{3}/_{24} \times 360)° = 45°$

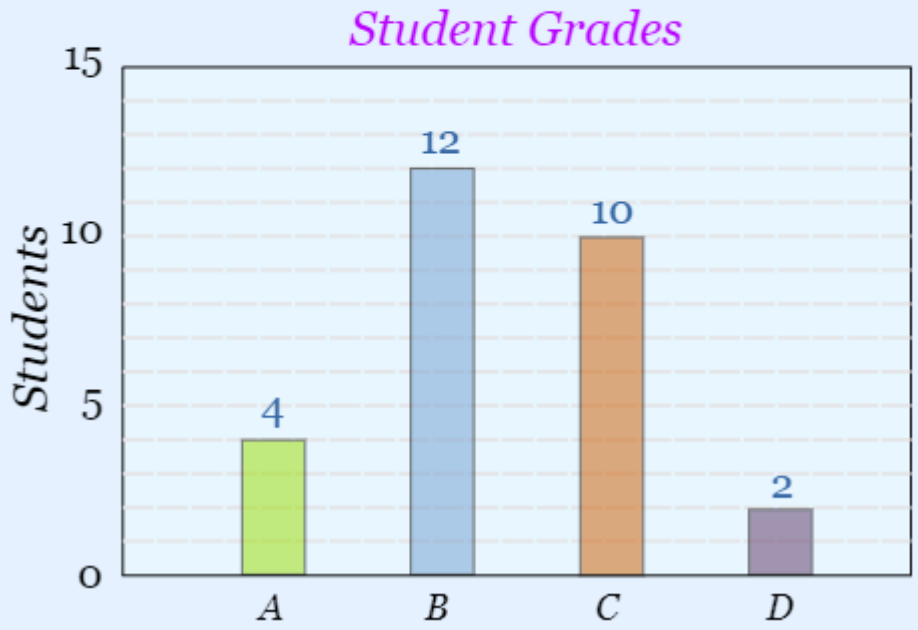


Bar Charts

- For **categorical** data!
- Usually used with **nominal** and **ordinal** variables.
- Each of the bars (columns) represents different values of a variable.
- Each bar's height indicates the frequency of each attribute.

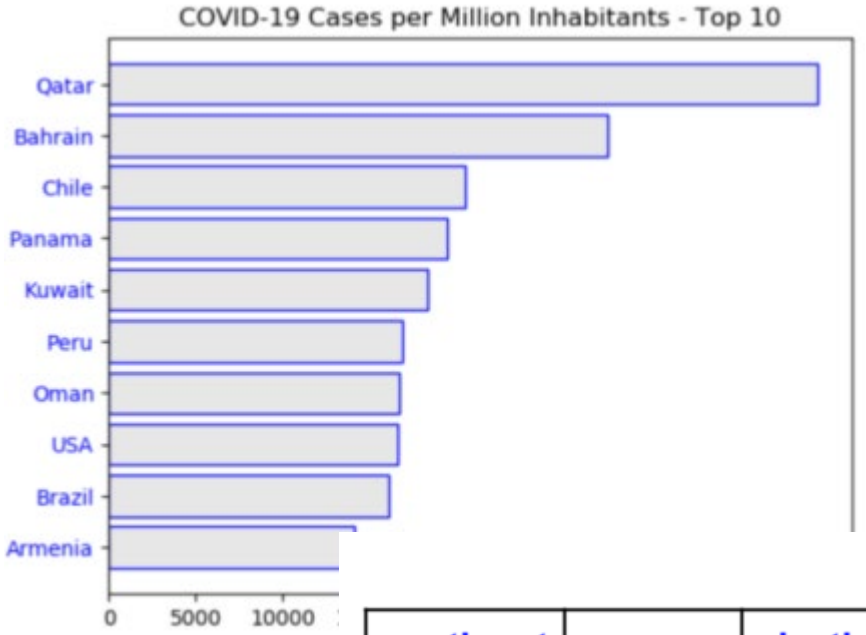
Grade:	A	B	C	D
Students:	4	12	10	2

bar graph:

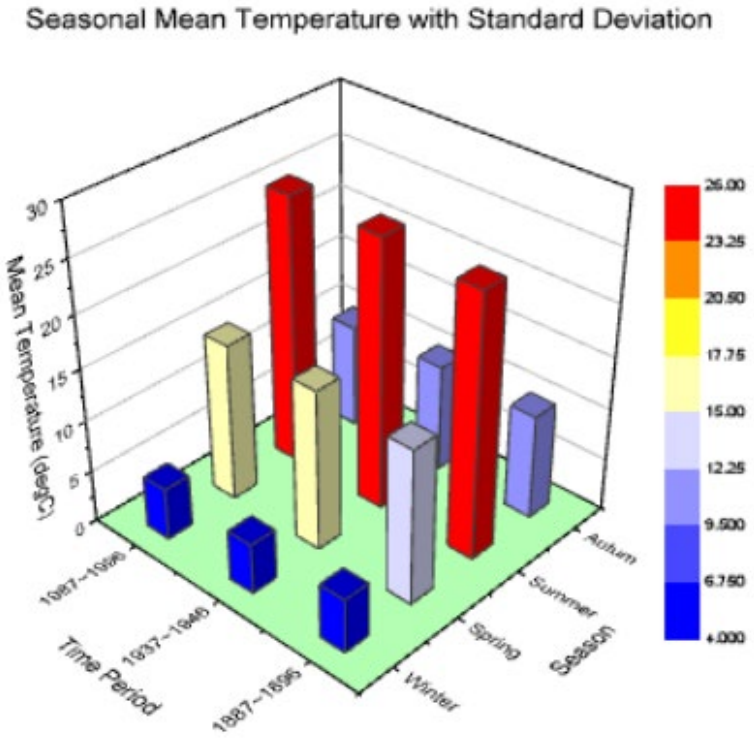


Bar Charts

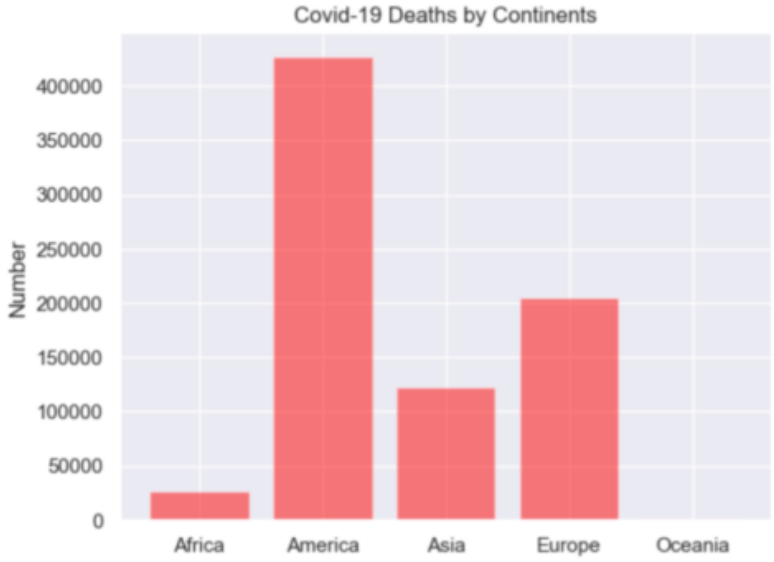
countriesAndTerritories	cases	deaths	popData2019	casesPer1M
Qatar	115661	193	2832071.0	40839.724710
Bahrain	47185	175	1641164.0	28750.935312
Chile	388855	10546	18952035.0	20517.849402
Panama	82790	1809	4246440.0	19496.331044
Kuwait	77470	505	4207077.0	18414.210151
Peru	549321	26658	32510462.0	16896.745423
Oman	83418	597	4974992.0	16767.464149
USA	5482416	171821	329064917.0	16660.591016
Brazil	3407354	109888	211049519.0	16144.808177
Armenia	41846	832	2957728.0	14148.021725



3D



continent	cases	deaths
Africa	1119579	26260
America	11698368	427207
Asia	5606210	122034
Europe	3239237	205144
Oceania	25742	471



Task -3

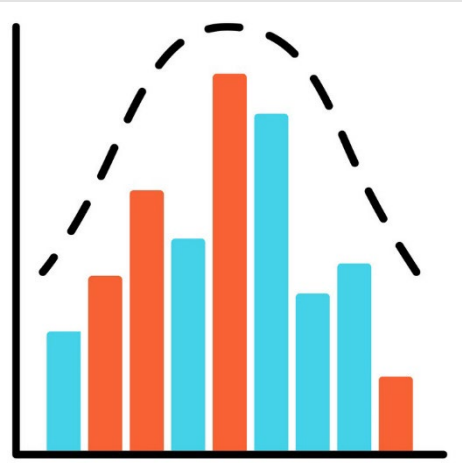
Task 3: The number of passengers of an airline company by years is given in the table below. Create a bar chart based on these data.

The number of passengers per year

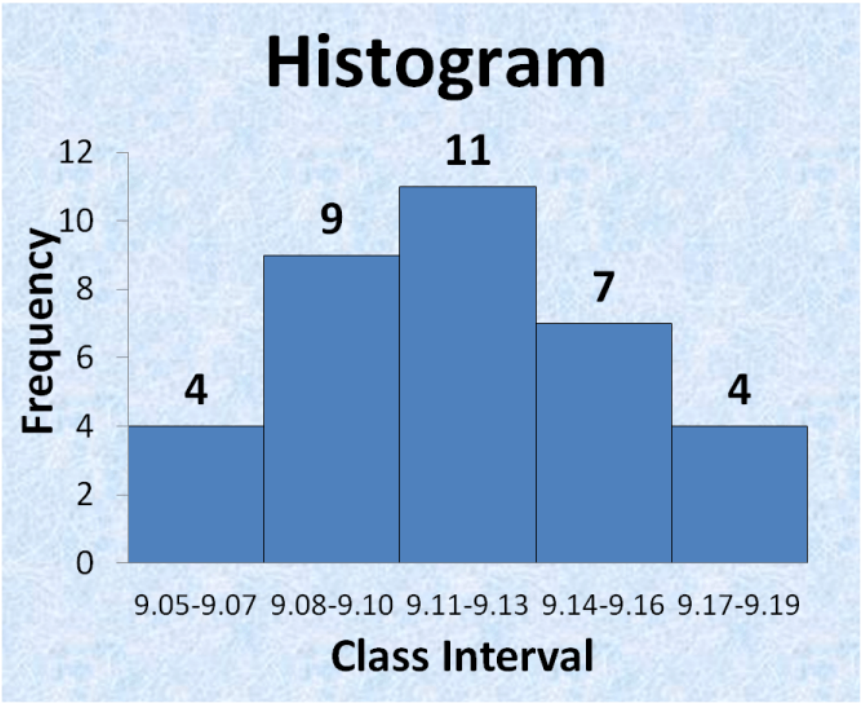
Years	2010	2012	2013	2014	2015	2016	2017	2018	2019
Number of passengers (x1000)	5	7	13	10	20	22	17	16.5	27

Excel

Histogram



- Used with **Interval** or **Ratio** variables.
- Represents the frequency of each attribute for a variable.
- Gives you a good bird's eye view of the distribution of your data.



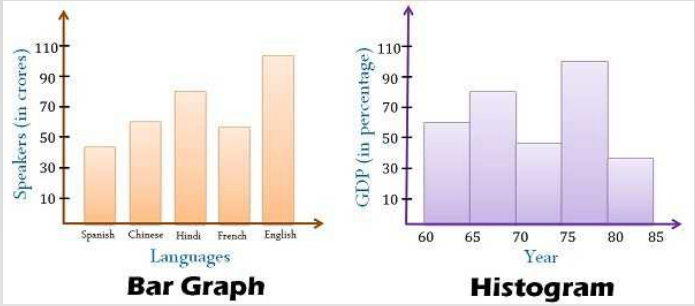
Bar Chart vs. Histogram

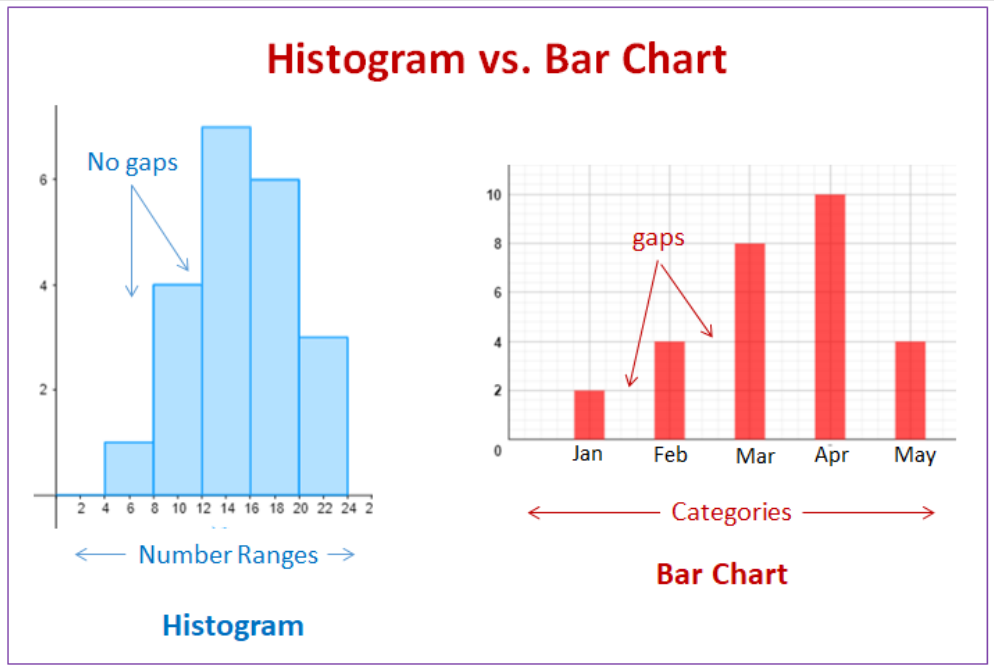
Bar Chart

- Categories are present.
- A schematic comparison of discrete variables.
- Provides categorical data.
- Gaps between bars.

Histogram

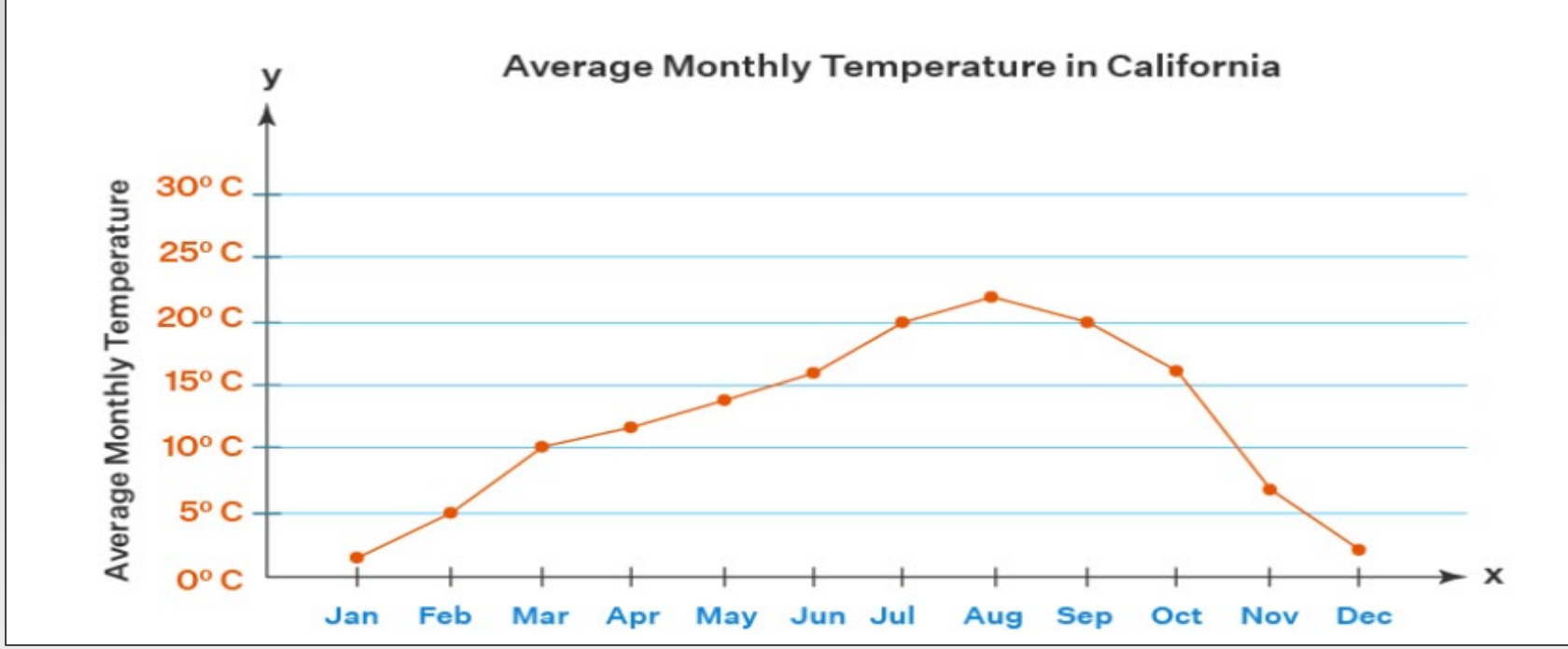
- Refers to graphic representation.
- Frequency distribution of continuous variables
- Provides numerical data.
- No Gaps between bars.

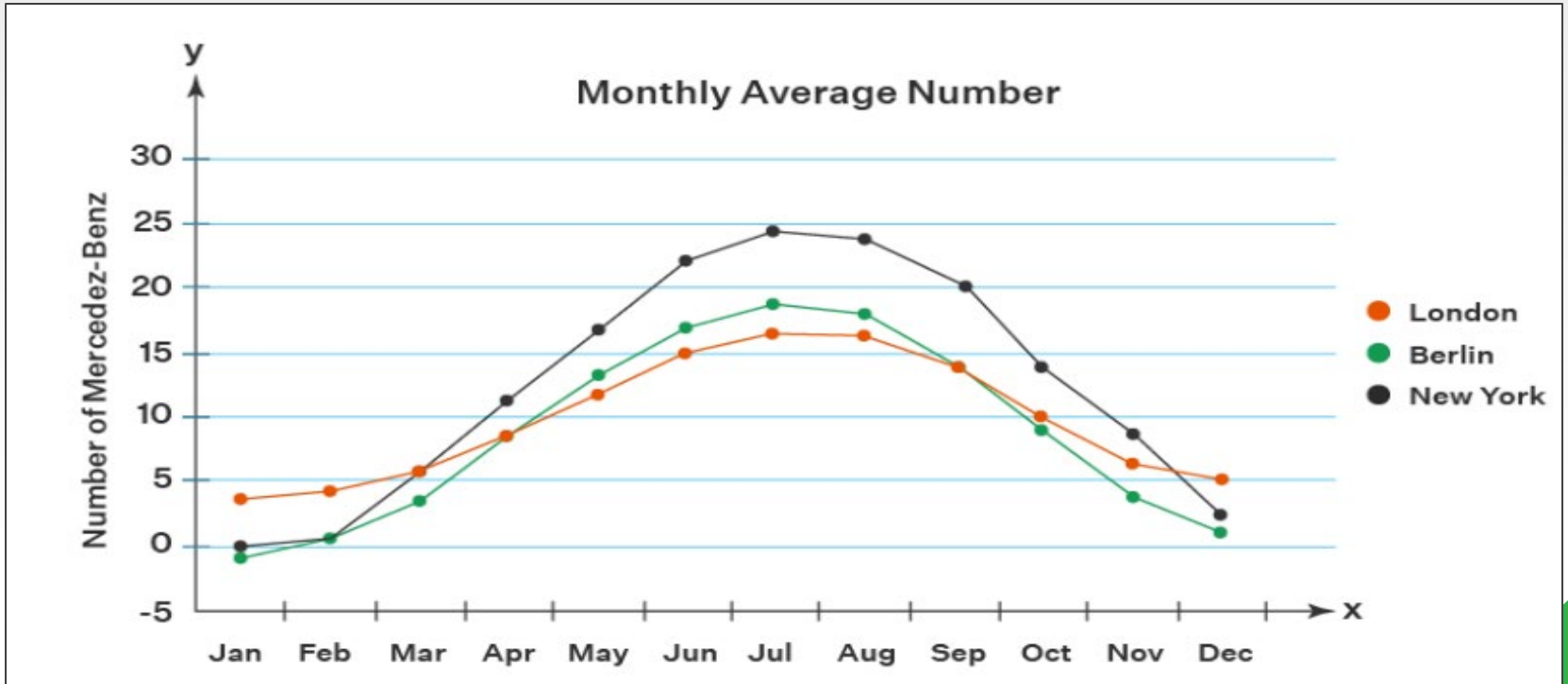




Line Chart

- Useful for **continuous numerical** data!
- Especially for time series.
- We put time to x-axis.
- Values locate at y-axis.





https://www.youtube.com/shorts/wx5mccl3c1k

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

TECHPRO EDUCATION

Tea break...

10:00

Start | Stop | Reset | mins: 10 | secs: 0 | type: Tea

Breaktime for PowerPoint by Flow Simulation Ltd.

Pin controls when stopped

Central Tendency (Measure of Centre)

- **Mod**
- **Median**
- **Mean**
- **Range**
- **IQR**
- **Variance, Standard Deviation**

Content

Central Tendency (Measure of Centre

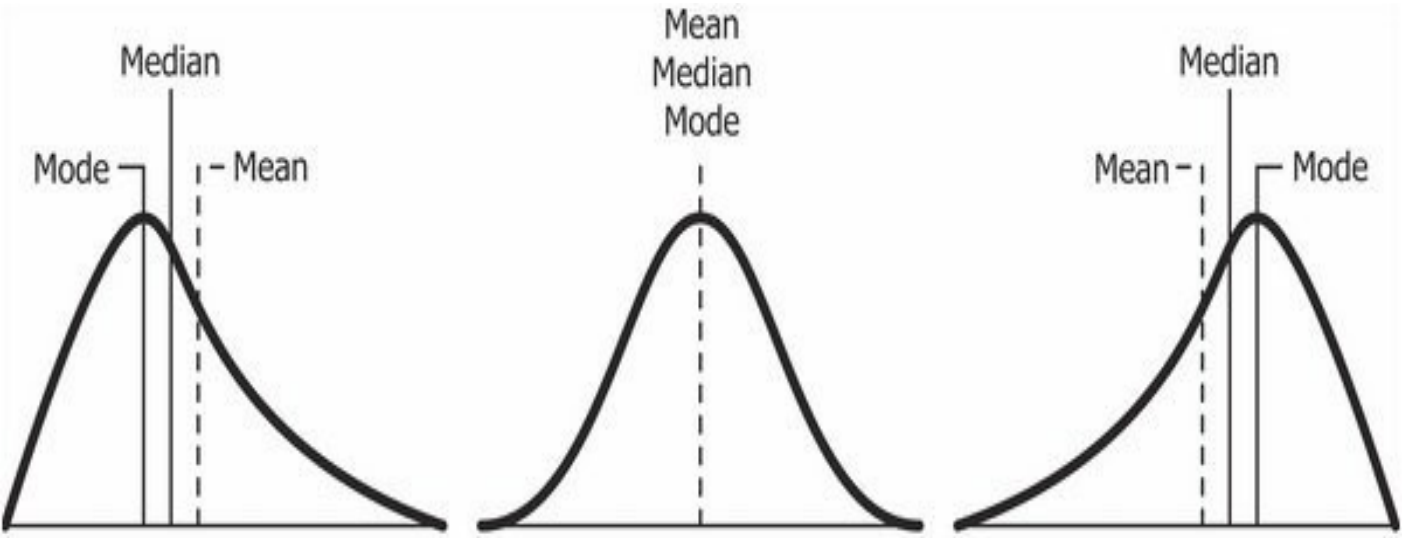
- Mean
- Median
- Mode

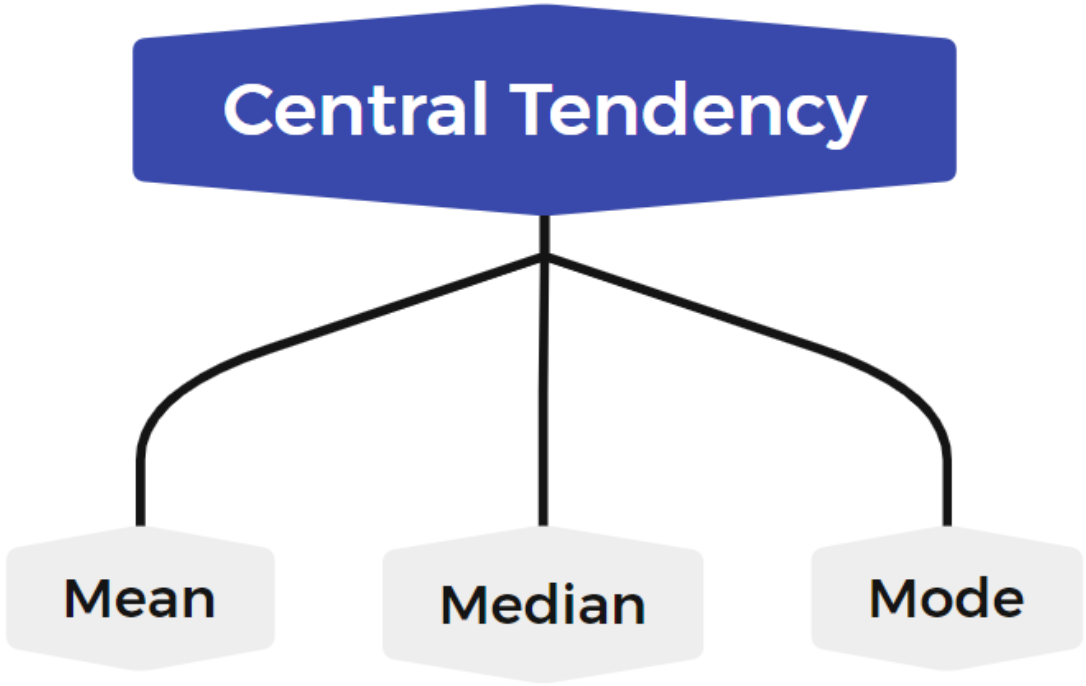
Dispersion (Measure of Spread)

- Range
- IQR
- Variance
- Standard Deviation

Central Tendency (Dispertion/ Measure of Spread)

- Best description of data with a single value;
  - **Mean:** Average, arithmetic mean
  - **Median:** Central point
  - **Mode:** Highest frequency





Mean (Average)

- Divide the sum of the data by the total number of observations
- Used to determine the location of the distribution

Population Mean	Sample Mean
$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$
N = number of items in the population	n = number of items in the sample

Mean Example

Example - 1

73	73	76	77	81	100

Mean (average) = $\frac{\text{Sum}}{\text{Count}}$

$$= \frac{73 + 73 + 76 + 77 + 81 + 100}{6}$$

$$= \frac{480}{6}$$

$$= 80$$

Example - 2

If we have frequencies, how can we calculate mean?

x	frequency
10	3
12	5
15	2
17	6
20	1
24	4

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$$

$$\bar{x} = \frac{10 \times 3 + 12 \times 5 + 15 \times 2 + 17 \times 6 + 20 \times 1 + 24 \times 4}{3 + 5 + 2 + 6 + 1 + 4}$$

$$\bar{x} = \frac{338}{21}$$

$$\bar{x} = 16.095$$

We use “mean” for filling null values in “numeric columns”.

Mean (Average)

Staff	Salary (thousand $)
1	102
2	33
3	26
4	27
5	30
6	25
7	33
8	33
9	24

Mean =37

But except the outlier (102), none of them not close to 37. Outlier caused a misinterpretation. Mean cannot represent the data set while there is outlier in the data set.

Staff	Salary (thousand $)
2	33
3	26
4	27
5	30
6	25
7	33
8	33
9	24

Mean =28.9

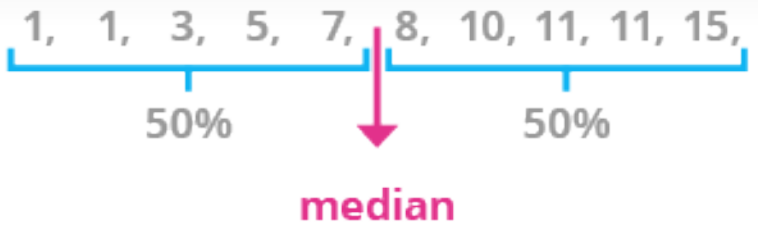
When we eliminate the outlier, mean is 28.9. It is located the centre of data set. It can represent the data set now.

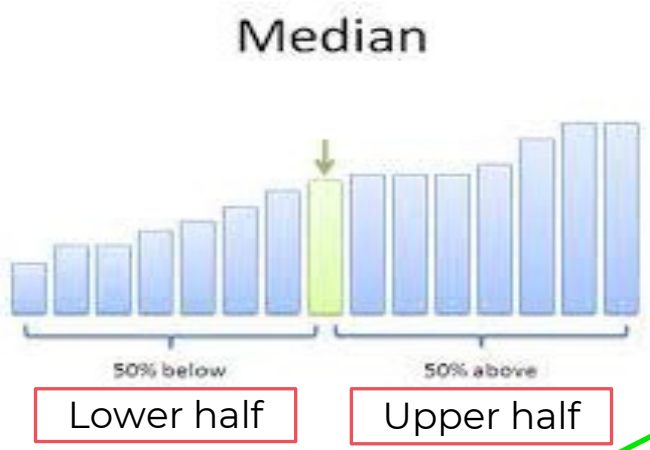
Median

- It is the median (**central**) **score of a data set sorted from smallest to largest**.
- If the number of data is odd, the median is **1** value, but if the number is even, the average of the middle **2** values is taken to find the median.

Median Formula $\left(\frac{n+1}{2}\right)^{th}$

- The median is the middle score. If the sample size is 9, the fifth element is the median





Staff	Salary (thousand $)
1	24
2	25
3	26
4	27
5	30
6	33
7	33
8	33
9	102

Median Example

- Ex-1

Find the median of these ages.

53 32 61 57 39 44 57

Let's sort the numbers as ascending.

32 39 44 **53** 57 57 61

Median is 53.

- Ex-2

Mean:

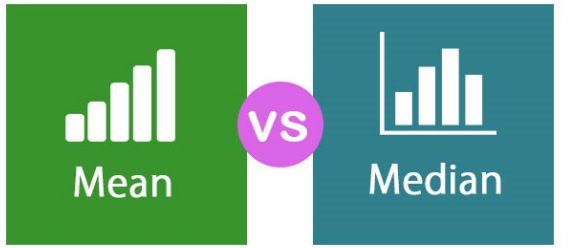
$$\mu = \frac{\sum X}{N}$$

$$\mu = \frac{\$4000 + \$15000 + \$20000 + \$33000 + \$1800000}{5}$$

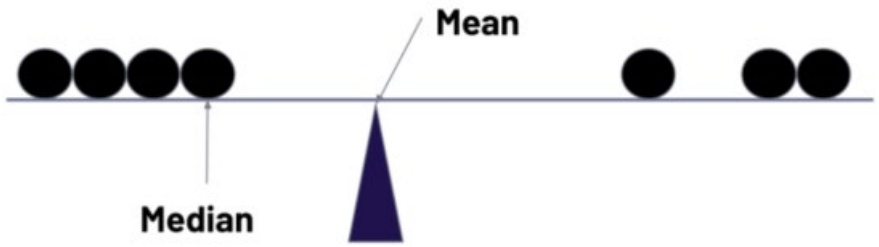
$$\mu = \frac{\$1872000}{5} = \$374400$$

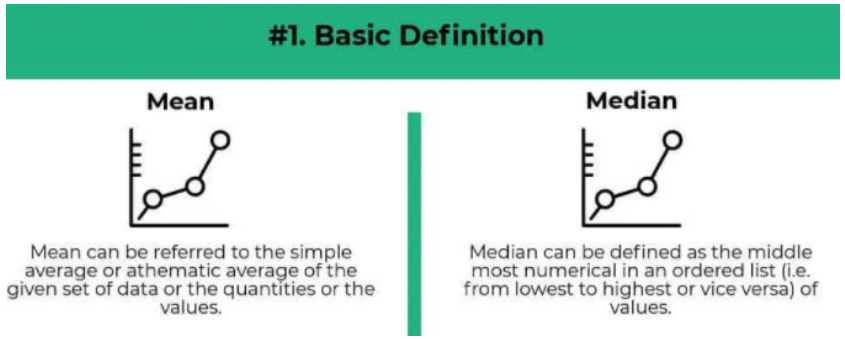
Median: $20000

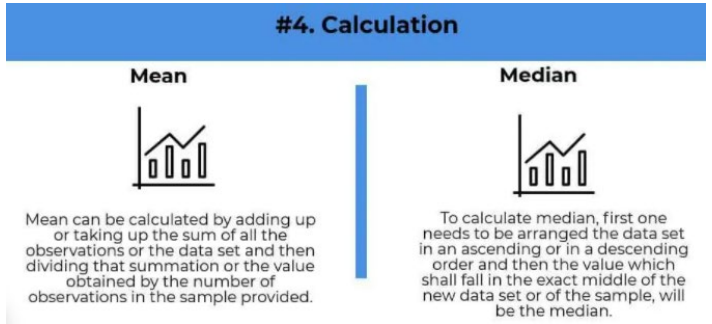
Mean vs. Median

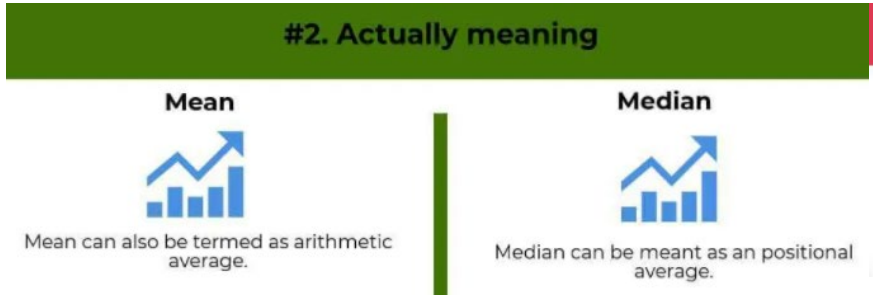


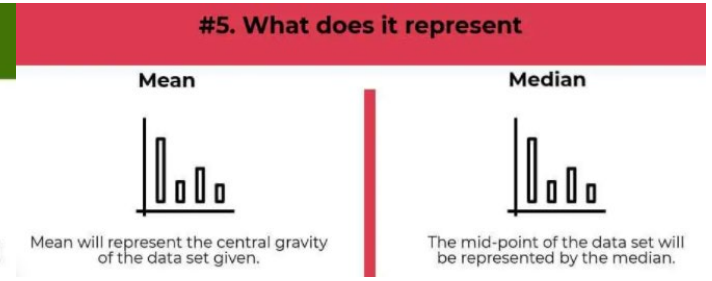
- If there are outliers in a small set of scores, the median is better.
- If there are no outliers in large data sets, the mean is better.
- Median may be better for salary bidding.

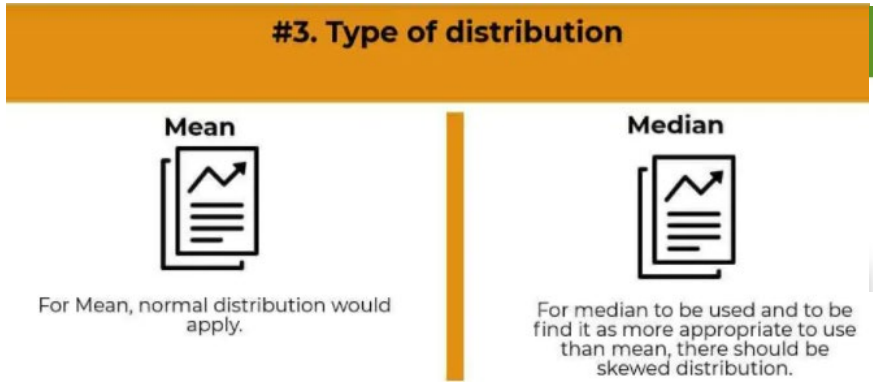


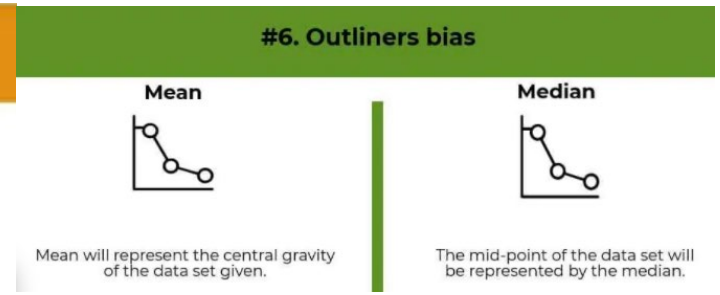








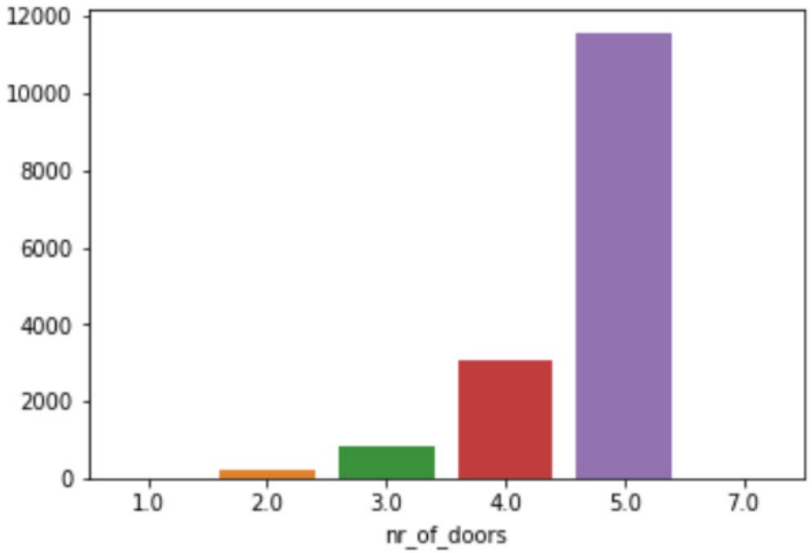


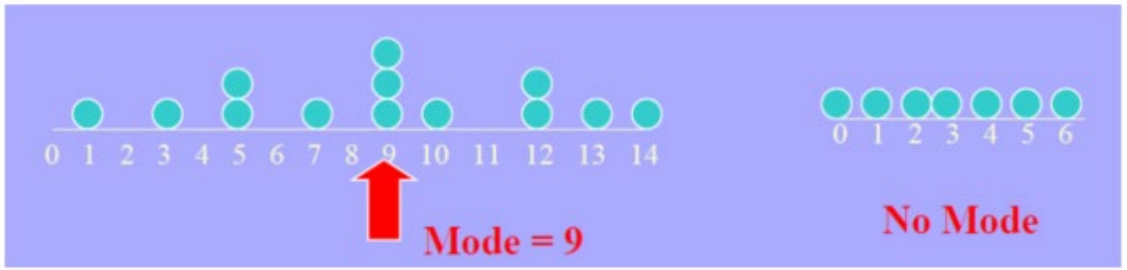


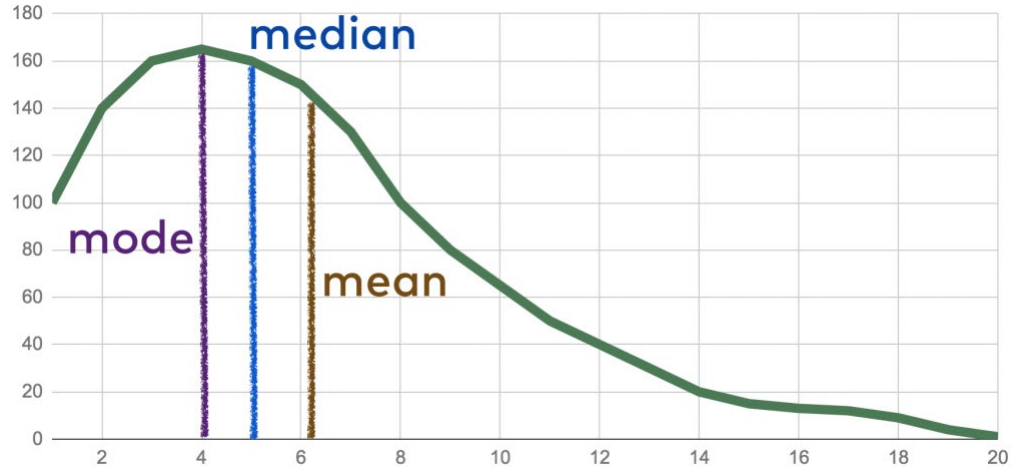
TECHPRO EDUCATION

Mode

- Mode is called peak value
- Mode: The most popular, most frequently encountered value in the data set.
- It can be used for both numeric and categorical variables
- Advantages and Disadvantages







TECHPRO EDUCATION

Mode Example

- Ex-1

What is the mode of these numbers?

53 32 61 **57** 39 44 **57**

Mode is 57 since there are two 57.

- Ex-2

What is the mode of these numbers?

120 100 130 100 160 130 86 100 94 90

Let's sort the numbers as ascending;

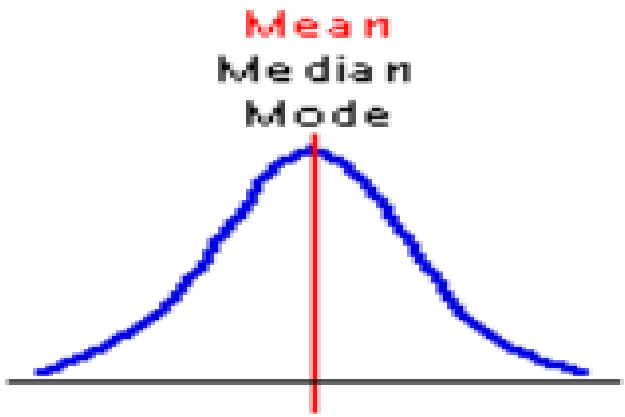
86	90	94	100	100	100	120	130	130	160

Mode = 100

Mode is 100, because it has most frequency.

Mean, Median and Mode

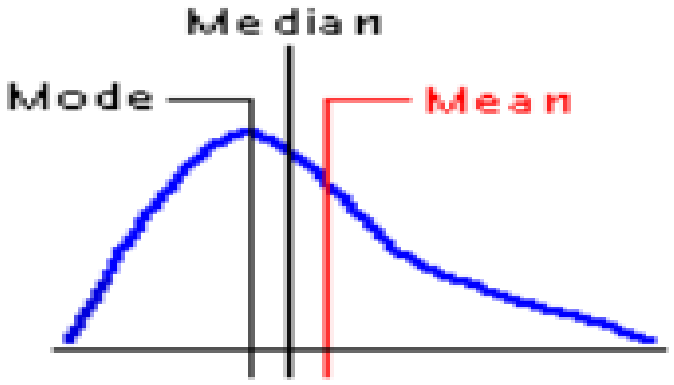
Normal Symmetric Distribution



Symmetrical Distribution

Mean=Median=Mode

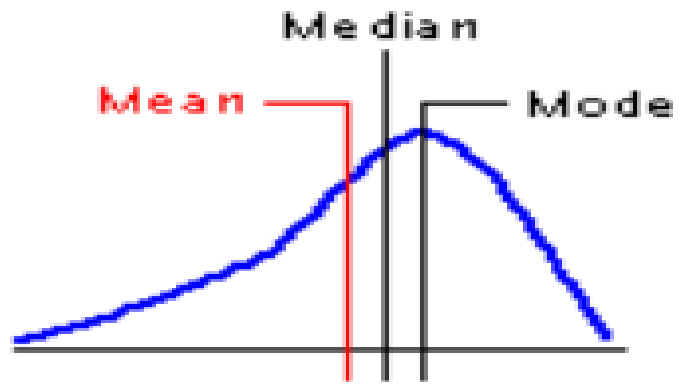
Right-Skewed Distribution



Positive Skew

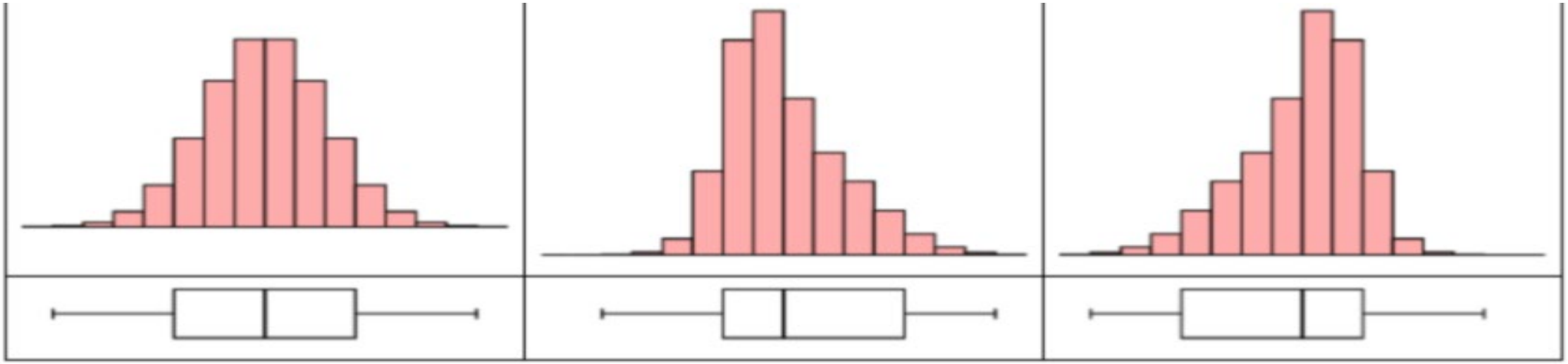
Mode < Median < Mean

Left-Skewed Distribution



Negative Skew

Mean<Median<Mode



Top 60 Statistics Interview Questions 2024

Question 4: What is the relationship between mean and median in normal distribution?

Answer: In a normal distribution, the mean and the median are equal.

https://www.mygreatlearning.com/blog/statistics-interview-questions/

Top 60 Statistics Interview Questions 2024

Question 5: What is the left-skewed distribution and the right-skewed distribution?

Answer: In the left-skewed distribution, the left tail is longer than the right side. It is also known as **negative-skew** distribution.

Mean < median < mode

In the right-skewed distribution, the right tail is longer. It is also known as **positive-skew** distribution.

Mode < median < mean

https://www.mygreatlearning.com/blog/statistics-interview-questions/

Mode – Median-Mean Example

Pear Deck time

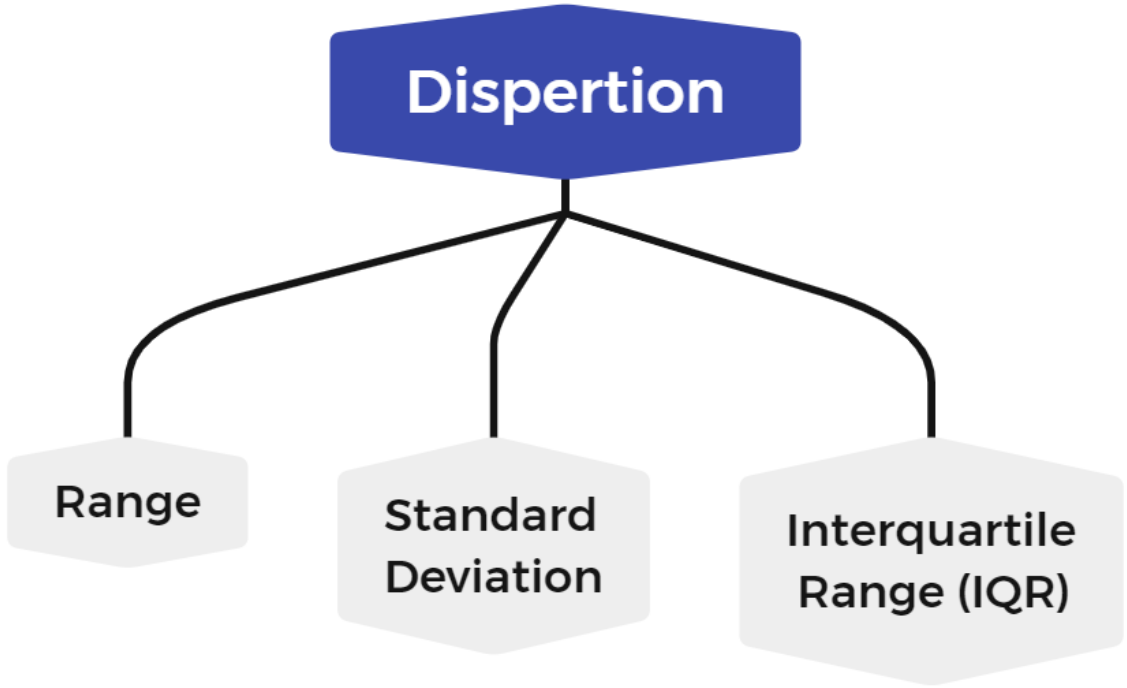
27-33

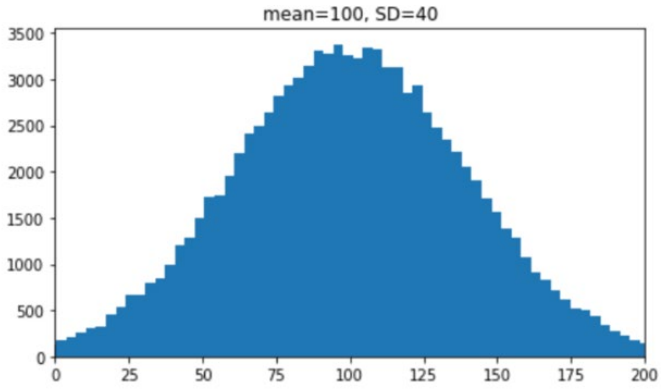
We use “mode” for filling null values in “categorical columns”.

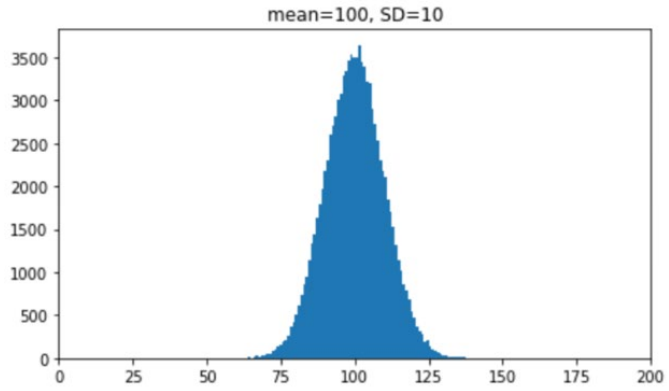
Dispertion (Measure of Spread)

Dispersion

- Central tendency measures alone do not characterize the distribution.
- The fact that **the mean of two data groups is equal** does not **require that their distributions are the same**.
- **A distribution explains more than the central tendency** does.





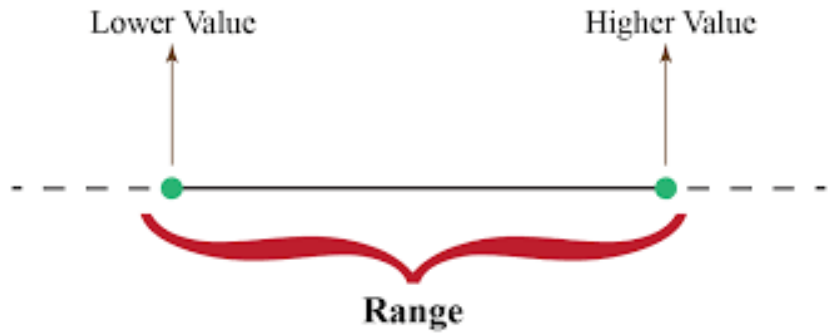


TECHPRO EDUCATION

Range

Range - Span - Width of Variation

- The **range** of a data set is the **difference** between the **maximum** and **minimum** number of entries in the set.
- It is the simplest measure of variability.



2 4 9 5 7 3

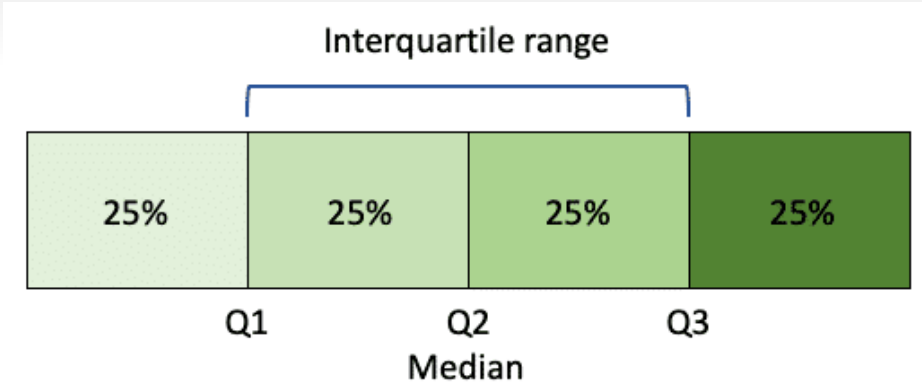
Range = Largest - Smallest = 9 - 2 = 7

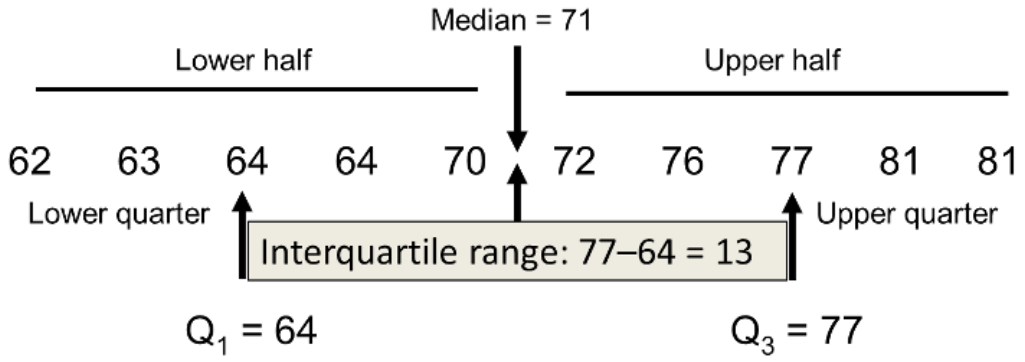
Recap Mean-Median-Mode-Range

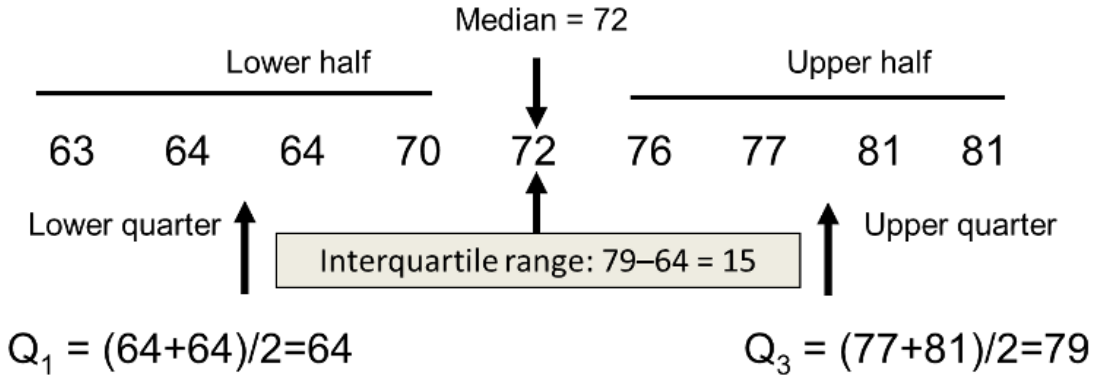
Inter Quartile Range (IQR)

- Quartiles are the values that divide a group of numbers by four parts.
- Q2 is the median of the whole dataset
- Q1 is the median of the portion below the median
- Q3 is the median of the portion above the median.

IQR = Q3 – Q1

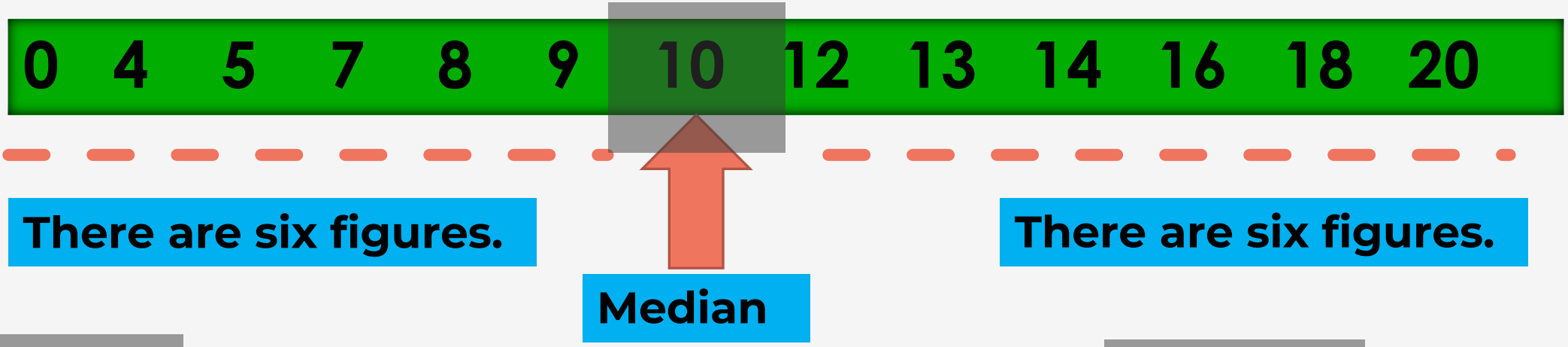


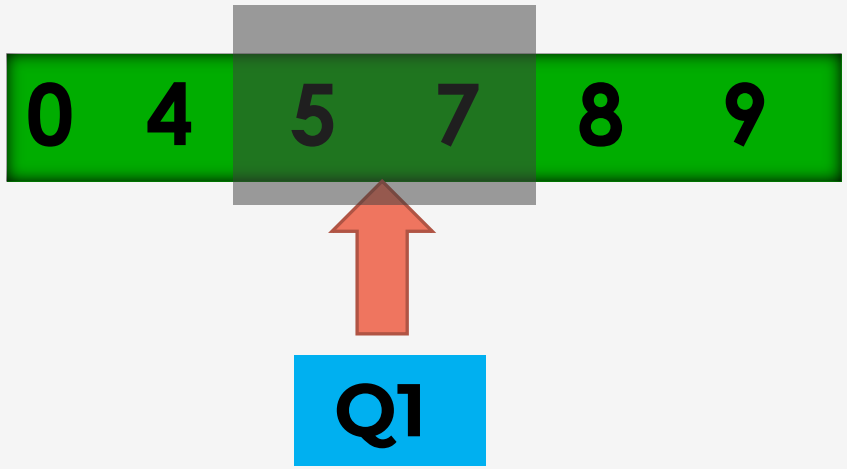


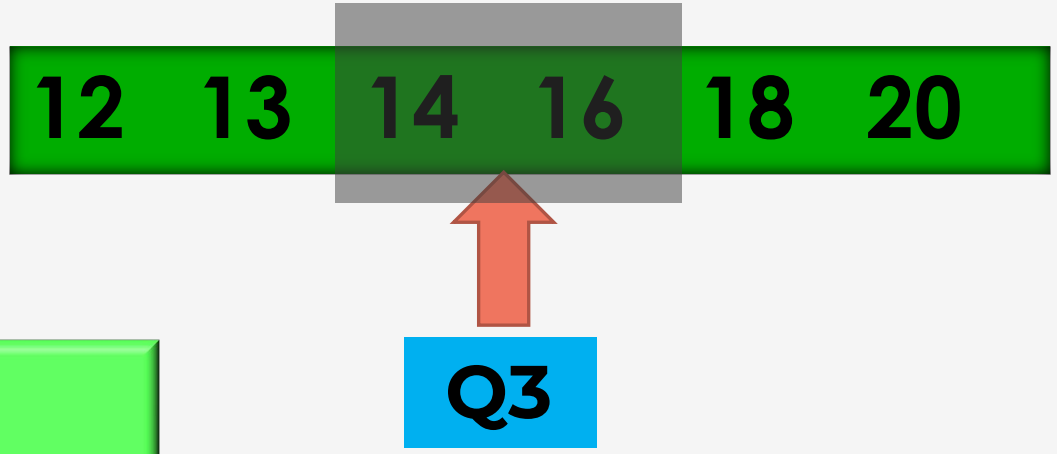


TECHPRO EDUCATION

IQR Example







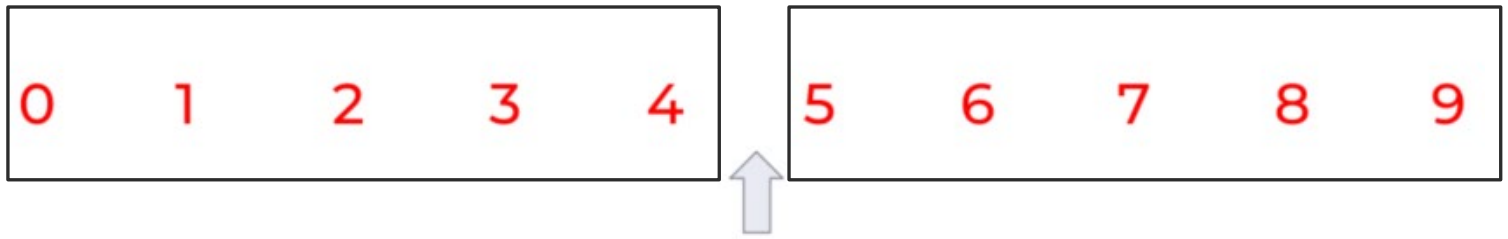
IQR = Q3 - Q1

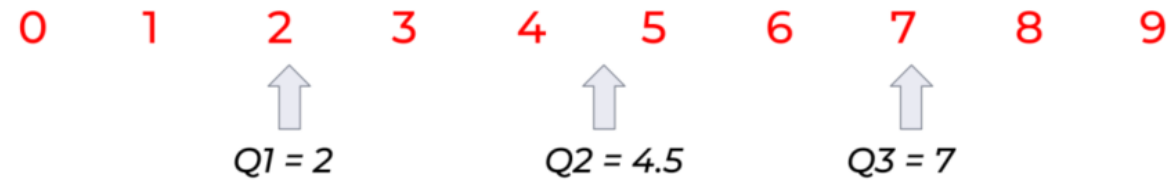
IQR = 15 - 6 = 9

Q1 = (5+7)/2 = 6

Q3 = (14+16)/2 = 15

IQR Example - 2

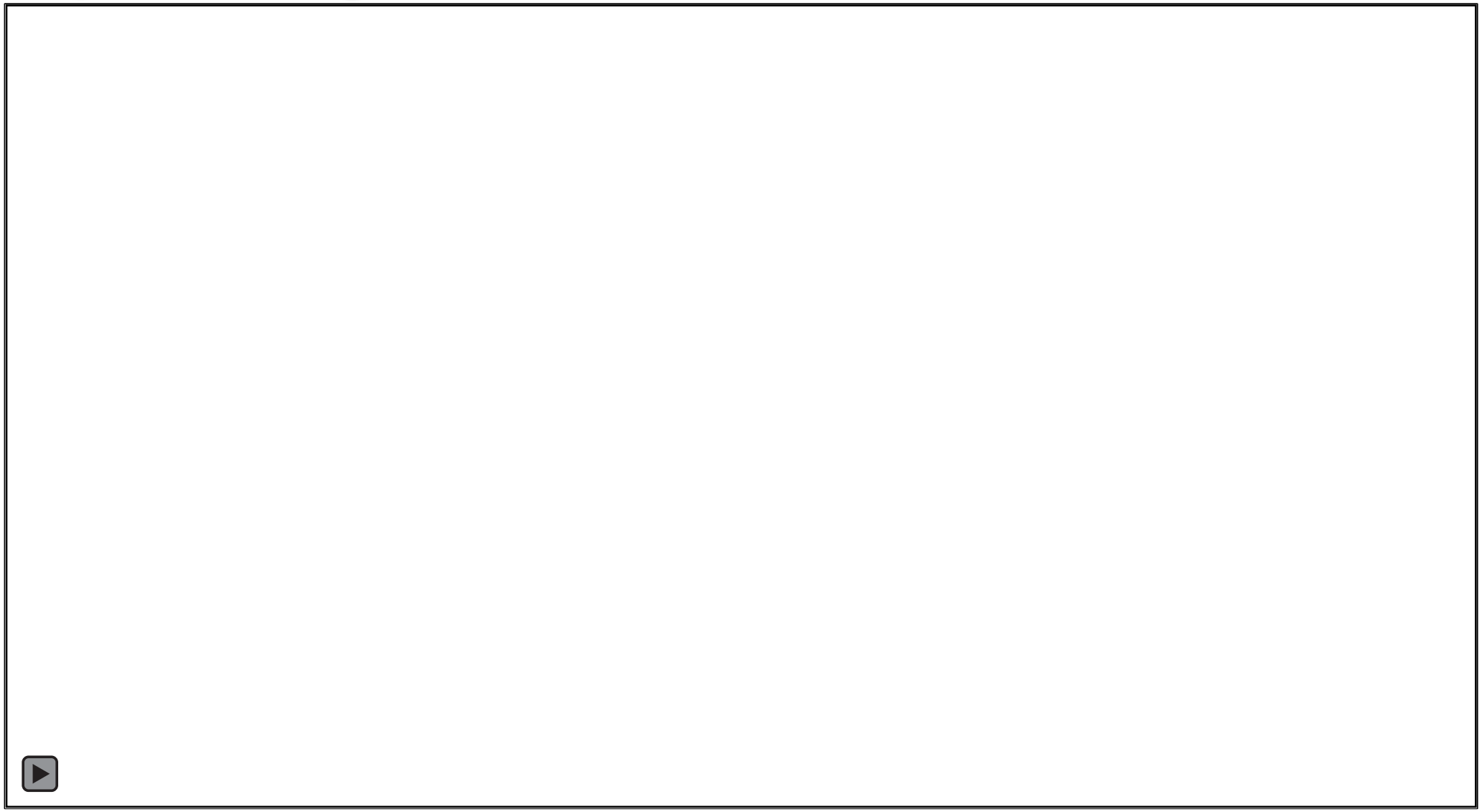




Interquartile Range = 7 - 2
IQR = 5

Interquartile Range = 7 - 2
IQR = 5

IQR Example – 2 - on Video



Top 60 Statistics Interview Questions 2024

Question 6: How to calculate range and interquartile range?

Answer: The range is the difference between the highest and the lowest values whereas the Interquartile range is the difference between upper and lower medians.

```
Range (X) = Max(X) - Min(X)

IQR = Q3 - Q1
```

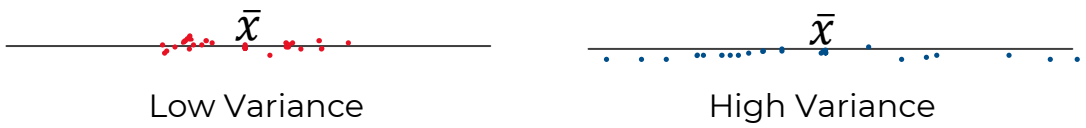
Here, Q3 is the third quartile (75 percentile)

Here, Q1 is the first quartile (25 percentile)

https://www.mygreatlearning.com/blog/statistics-interview-questions/

Variance

- Variance is defined as the mean of the squares of the differences from the mean.
- The amount by each score moves away from the mean.



Low Variance

High Variance

Variance

Sample variance

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

S^2=sample variance

x_i =value of i th element

$\bar{x}$ = sample mean

n=sample size

Population variance

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

σ^2=population variance

x_i =value of i th element

μ=population mean

N=population size

sample variance — $S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$ — observation, mean, number of observations

variance — $\sigma^2 = \frac{\sum(x - \mu)^2}{N}$ — element, mean, number of elements

Variance Example

- Variance for the following 4 numbers

0 1 5 6

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

0 1 5 6

Mean: $\mu = \frac{\sum X}{N} = \frac{0+1+5+6}{4} = \frac{12}{4} = 3$

Dev Sum of Squares: $SS = \sum(X-\mu)^2$

$SS = (0-3)^2 + (1-3)^2 + (5-3)^2 + (6-3)^2$

$SS = 9 + 4 + 4 + 9 = 26$

Variance: $\sigma^2 = \frac{\sum(X-\mu)^2}{N}$

$\sigma^2 = \frac{26}{4} = 6.5$

- Ex- 2

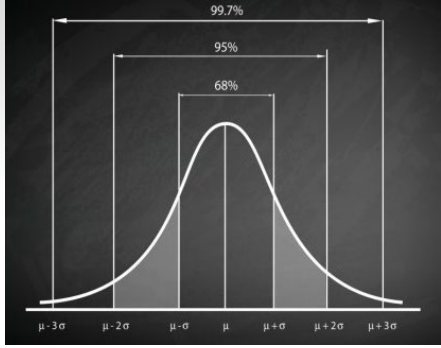
10 12 17 20 25 27 42 45

- Find for both sample and population.

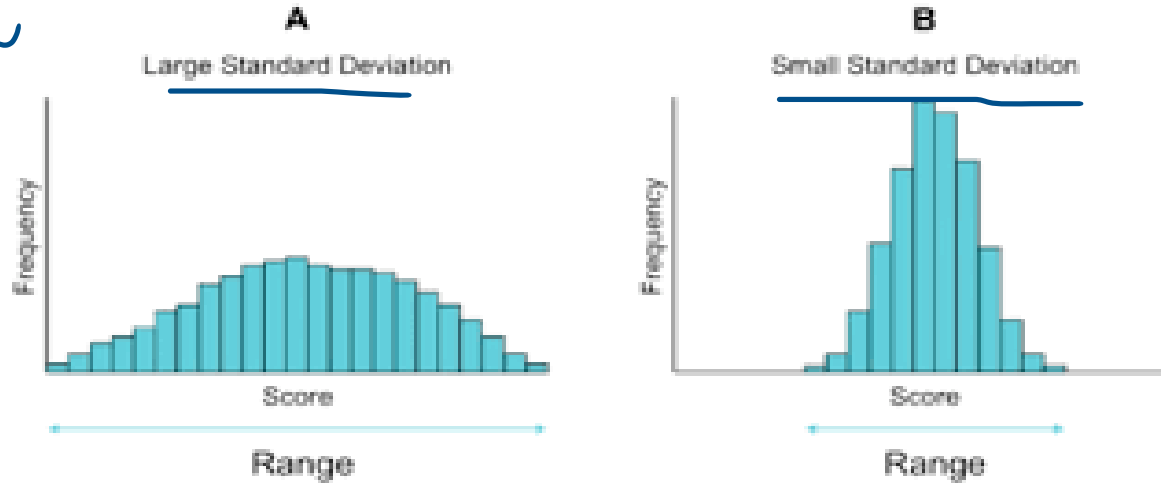
$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

$$S^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$$

Standard Deviation



- It is the square root of the variance.
- The more spread out the data, the larger the standard deviation.
- If std dev is large, data is spread over an extensive range.



Sample

$$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Population

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

element mean

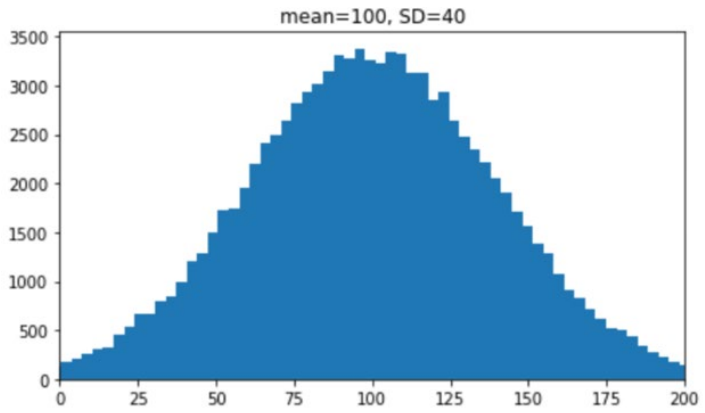
standard deviation $\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$

number of elements

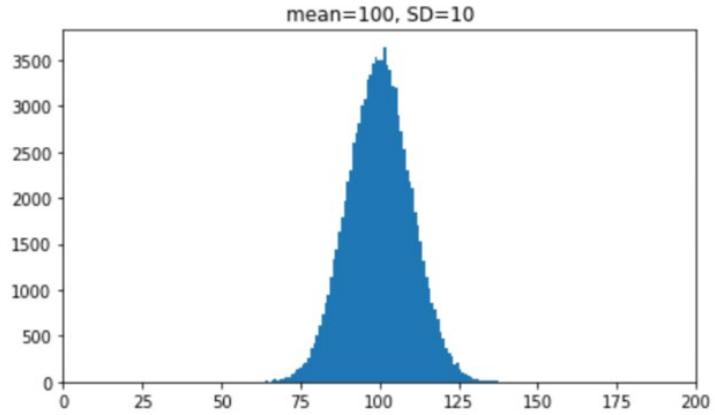
TECHPRO EDUCATION

Standard Deviation

While means equal in both data set, the first one has bigger std dev since it spread in a large range.



The second one has smaller std dev since it spread in a narrow range.



Std. Dev. Example

Staff	Salary (thousand $)
1	24
2	25
3	26
4	27
5	30
6	33
7	33
8	33
9	102

$$\mu = \frac{24+25+26+27+30+33+33+33+102}{9}$$

$$\mu = \frac{333}{9} = 37$$

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{N}}$$

$$\sigma = \sqrt{\frac{(24-37)^2+(25-37)^2+(26-37)^2+(27-37)^2+(30-37)^2+(33-37)^2+(33-37)^2+(33-37)^2+(102-37)^2}{9}}$$

$$\sigma = \sqrt{\frac{(-13)^2+(-12)^2+(-11)^2+(-10)^2+(-7)^2+(-4)^2+(-4)^2+(-4)^2+(65)^2}{9}}$$

$$\sigma = \sqrt{\frac{169+144+121+100+49+16+16+16+4225}{9}}$$

$$\sigma = \sqrt{\frac{4856}{9}}$$

$$\sigma = \sqrt{539,55}$$

$$\sigma = 23,22833518$$

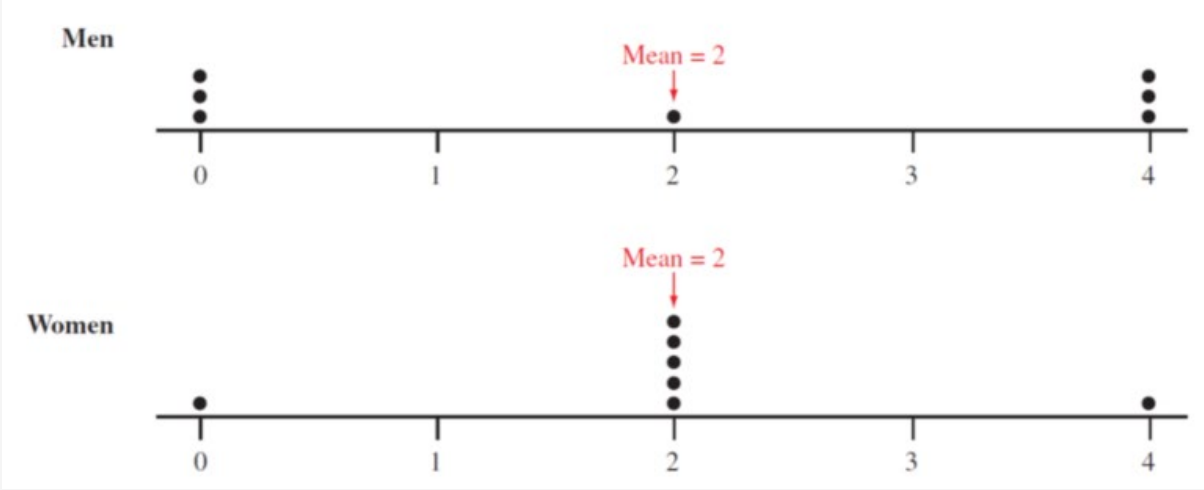
$$\sigma = 3.58$$

Staff	Salary (thousand $)
1	24
2	25
3	26
4	27
5	30
6	33
7	33
8	33

Std. Dev. Example - 2

Men: 0 0 0 2 4 4 4

Women: 0 2 2 2 2 2 4



- For the above 2 groups of respondents who answered the ideal number of children for a family (7 people each).
- What is the variance

Men: $s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{24}{6}} = \sqrt{4} = 2.0$

Women: $s = 1.2$

RANGE STANDARD DEVIATION

Top 60 Statistics Interview Questions 2024

Question 7: What are descriptive statistics?

Answer: Descriptive statistics are used to summarize the basic characteristics of a data set in a study or experiment. It has three main types

- **Distribution:** Refers to the frequencies of responses. ***Example:*** Mode
- **Central Tendency:** Gives a measure or the average of each response. ***Example:*** Mean, median.
- **Variability:** Shows the dispersion of a data set. ***Example:*** Variance, standard deviation, range, IQR

Top 60 Statistics Interview Questions 2024

Question 8: What is the meaning of standard deviation?

Answer: Standard deviation gives the measure of the variation of dispersion of values in a data set. It represents the differences of each observation or data point from the mean.

$$(\sigma) = \sqrt{(\sum (x-\mu)2 \ / \ n)}$$

Where the variance is the square of standard deviation.

https://www.mygreatlearning.com/blog/statistics-interview-questions/

Top 60 Statistics Interview Questions 2024

Question 9: What is Bessel's correction?

Answer: Bessel's correction advocates the use of **n-1** instead of n in the formula of standard deviation at **sample**. It helps to increase the accuracy of results while analyzing a sample of data to derive more general conclusions.

Sample	Population
$S = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$	$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$

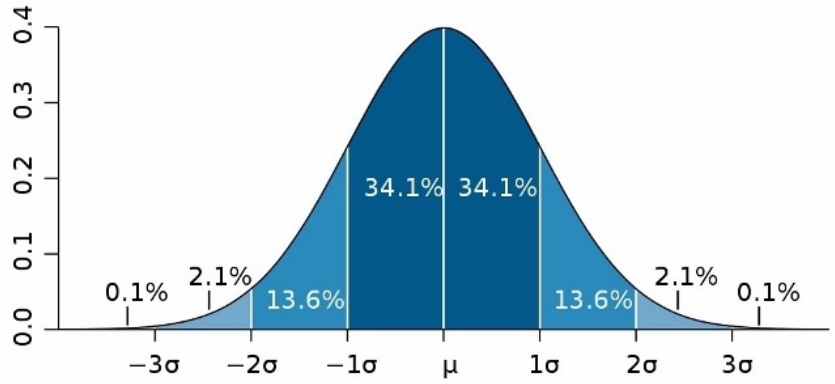
https://www.mygreatlearning.com/blog/statistics-interview-questions/

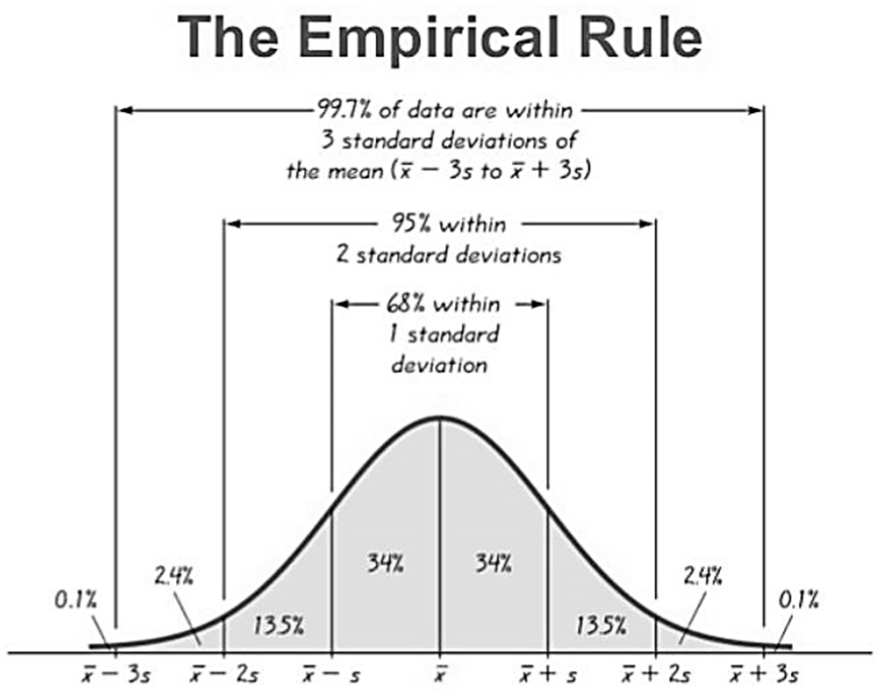
Empirical Rule

3 Sigma Rule

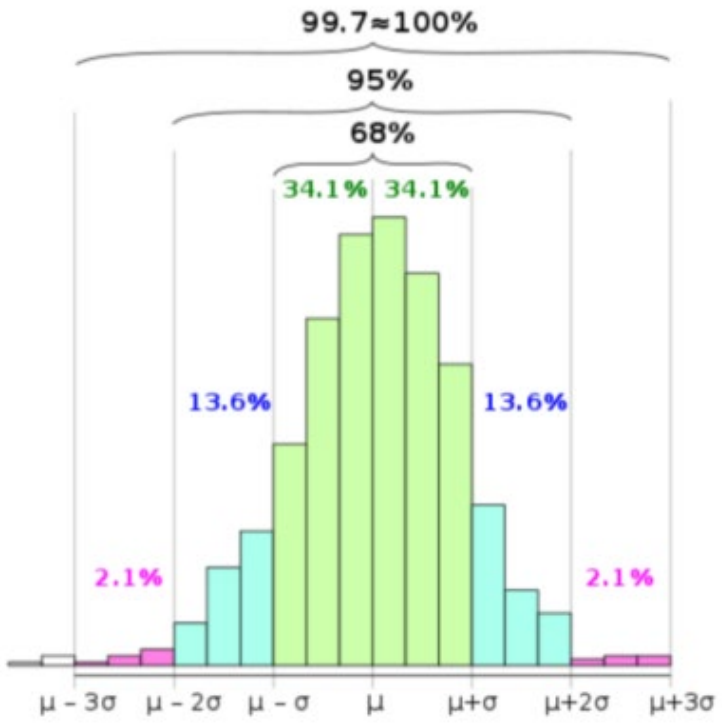
- Three Sigma Rule or also known as 68-95-99.7 Rule.
- **Empirical Rule**

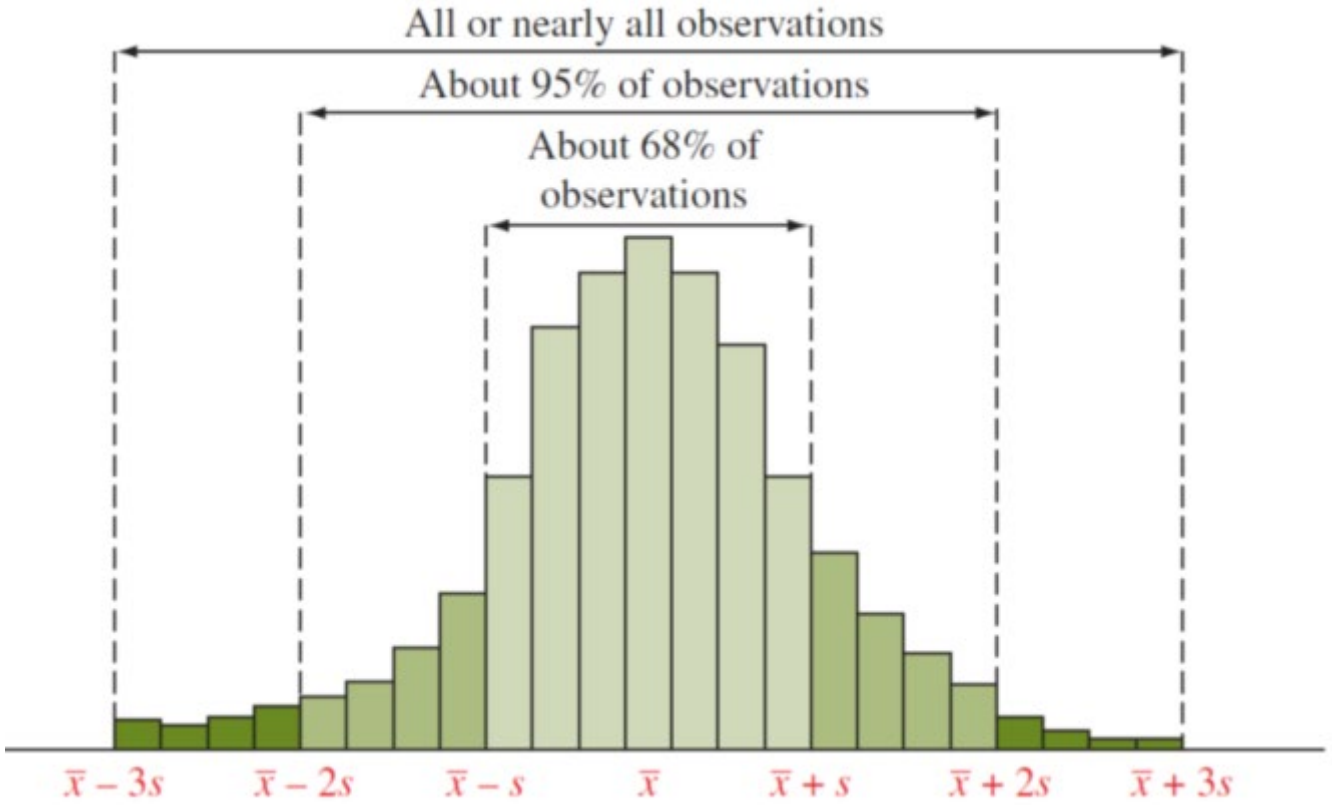
1. rule at 68% = (Mean - standard deviation) and (Mean + standard deviation)
2. rule at 95% = (Mean - 2 × standard deviation) and (Mean + 2 × standard deviation)
3. rule at 99.7% = (Mean - 3 × standard deviation) and (Mean + 3 × standard deviation)

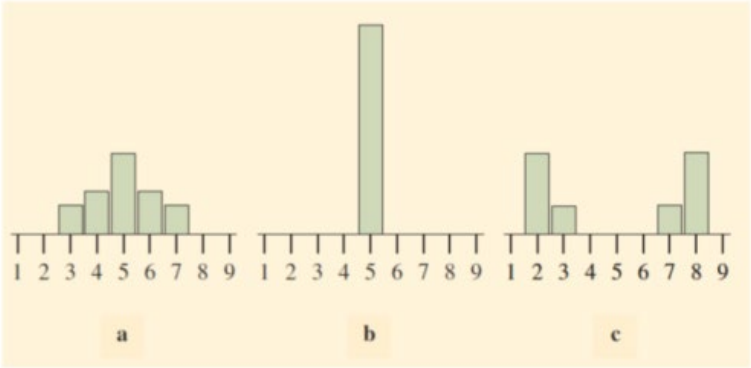




Empirical Rule







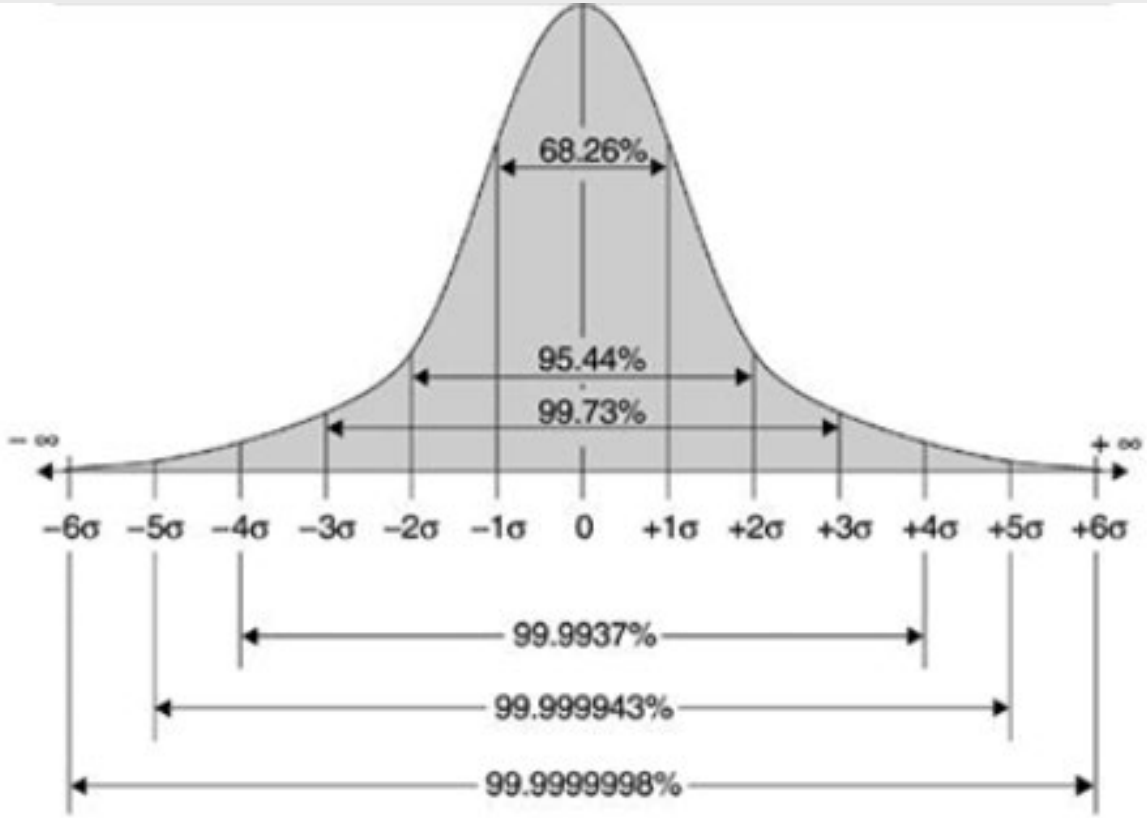
Pear Deck time

- 36-38

Empirical Rule

6 Sigma Rule

- The term "Six Sigma" refers to a statistical measure of how far a process deviates from **perfection.** A process that operates at **six sigma has a failure rate of only 0.00034%, which means it produces virtually no defects.**
- **Six sigma in statistics is a quality control method to produce an error or defect-free data set.** The more the standard deviation, the less likely that process performs with accuracy and causes a defect. **If a process outcome is 99.99966% error-free, it is considered six sigma.** A six-sigma model works better than 1σ, 2σ, 3σ, 4σ, 5σ processes and is reliable enough to produce defect-free work.



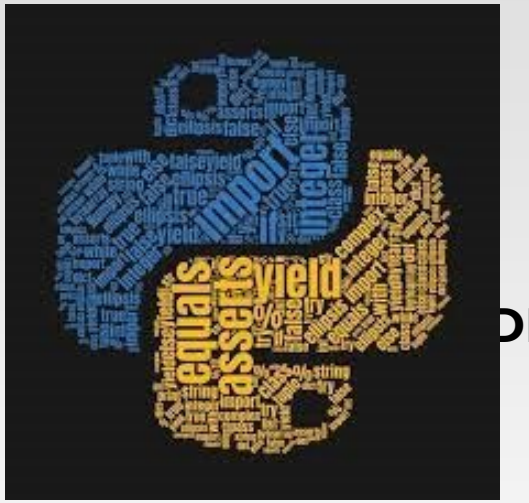
Top 60 Statistics Interview Questions 2024

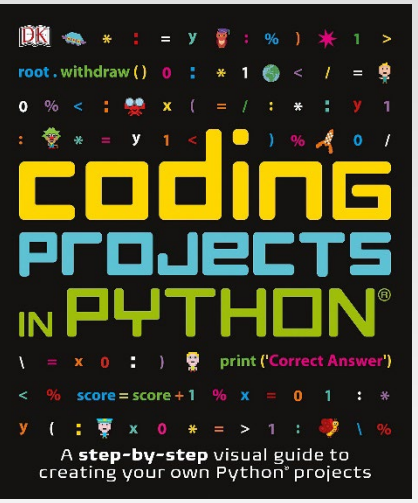
Question 10: What is the meaning of six sigma in statistics?

Answer: **Six sigma in statistics is a quality control method to produce an error** or **defect-free data set.** If a process outcome is 99.9999966% error-free, it is considered six sigma.

https://www.mygreatlearning.com/blog/statistics-interview-questions/

Statistics Practice-1





Python Notebook Time

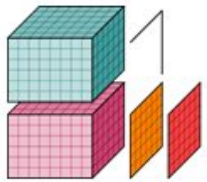
Statistics with Python

SM StatsModels
Statistics in Python

Seaborn

pandas

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



xarray

scikit-image
image processing in python

NumPy

matplotlib

jupyter

python™

IP[y]:
IPython

statsmodels

Statistic with Python

- Input

```
import numpy as np
from scipy import stats

salary = [102, 33, 26, 27, 30, 25, 33, 33, 24]

mean_salary = np.mean(salary)
print("mean:", mean_salary)

median_salary = np.median(salary)
print("median:", median_salary)

mode_salary = stats.mode(salary)
print("mode:", mode_salary)
```

- Output

```
mean: 37.0
median: 30.0
mode: ModeResult(mode=array([33]), count=array([3]))
```

Calculate Mean, Median and Mode with Python

Std. Dev with python

input :

```
import numpy as np

salary = [102, 33, 26, 27, 30, 25, 33, 33, 24]

print("Range: ", (np.max(salary)-np.min(salary)))

print("Variance: ", (np.var(salary)))

print("Std: ", (np.std(salary)))
```

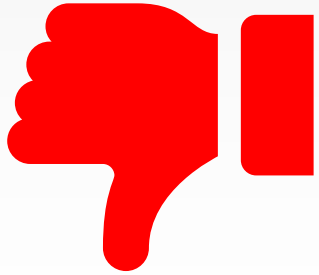
output :

```
Range:  78
Variance:  539.5555555555555
Std:  23.22833518691246
```

BEST PRACTISE

It is time to Kahoot !!

Today's session was effective/fruitful!!



Tea break...

10:00

Start | Stop | Reset | mins: 10 | secs: 0 | type: Tea

Breaktime for PowerPoint by Flow Simulation Ltd.

Pin controls when stopped