Activity	Data Type
Number of beatings from Wife	Countable
Results of rolling a dice	Nominal / Categorical
Weight of a person	Continuous
Weight of Gold	Continuous
Distance between two places	Continuous
Length of a leaf	Continuous
Dog's weight	Continuous
Blue Color	Discrete
Number of kids	Continuous
Number of tickets in Indian railways	Continuous
Number of times married	Continuous
Gender (Male or Female)	Nominal / Categorical

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

Data	Data Type
Gender	Nominal
High School Class Ranking	Ordinal
Celsius Temperature	Ratio
Weight	Interval
Hair Color	Nominal
Socioeconomic Status	Ordinal
Fahrenheit Temperature	Ratio
Height	Interval

Type of living accommodation	Ordinal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Interval
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Interval
Time on a Clock with Hands	Interval
Number of Children	Ratio
Religious Preference	Nominal
Barometer Pressure	Interval
SAT Scores	Interval
Years of Education	Nominal

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

ANS : 3/8

Q4) Two Dice are rolled, find the probability that sum is

a) Equal to 1

b) Less than or equal to 4

c) Sum is divisible by 2 and 3

ANS :

a) 0 (The sum always exceeds more than 1)

b) 1/6

c) 1/6

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

ANS : 10/21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

CHILD	Candies count	Probability
A	1	0.015
B	4	0.20
C	3	0.65
D	5	0.005
E	6	0.01
F	2	0.120

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

ANS : 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

- For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Use Q7.csv file

ANS :

	Mean	Median	Mode	Variance	Std deviation	Range
Points	3,596	3,69	3,07	0,285	0,534	2,17
Score	3,217	3,32	3,44	0,957	0,978	3,91
Weigh	17,848	17,71	17,02	3,193	1,786	8,40

Mean, median and mode are not the same in any of the cases.
We have outliers in the score and weigh

Q8) Calculate Expected Value for the problem below

a) The weights (X) of patients at a clinic (in pounds), are
108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

ANS : 145.33

Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

   Cars speed and distance

Use Q9_a.csv

SP and Weight(WT)

Use Q9_b.csv

ANS :

	Skewness	Kurtosis
Speed	-0,11	-0,50
Dist	0,80	0,40

Speed: It is negatively skewed

Dist: It is positively skewed

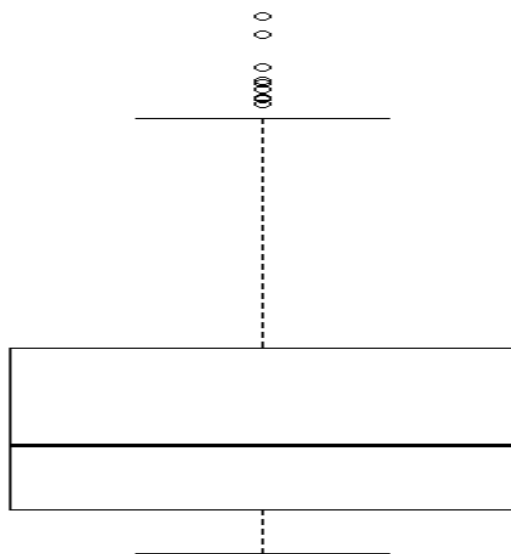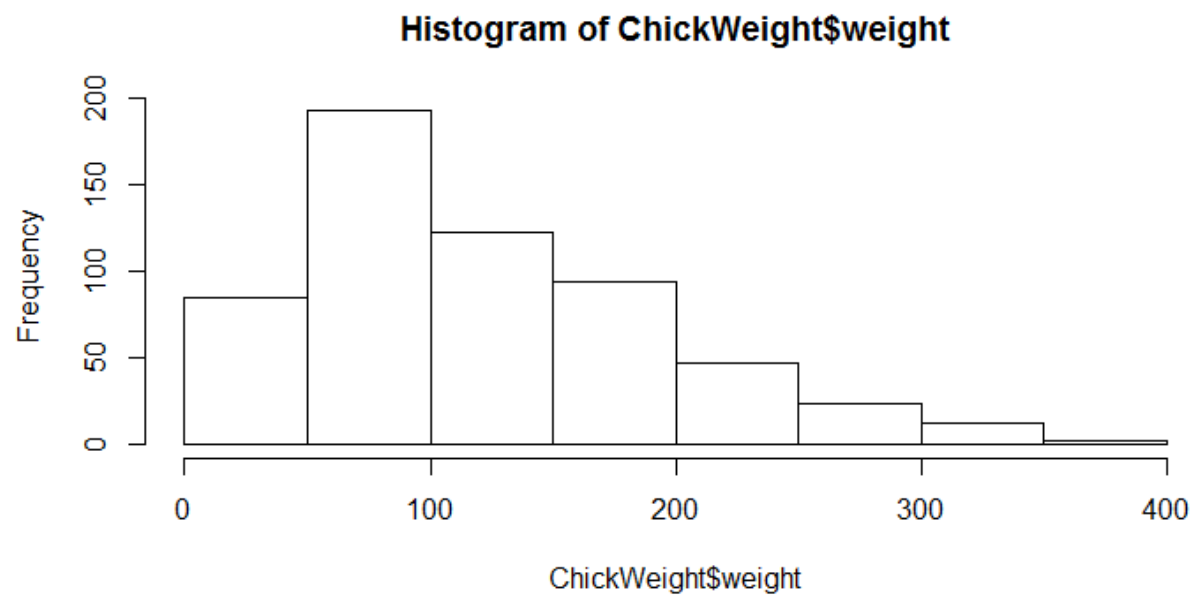Speed has negative kurtosis and dist have positive kurtosis

	Skewness	Kurtosis
SP	1,58	2,97
WT	-0,61	0,95

Sp: It is positively skewed

WT: It is negatively skewed

Both have positive kurtosis

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight

ANS :

Histogram is positively skewed.The majority of the data lies in the range of 50-100

From the boxplot, we can conclude that the dataset contains outliers

Q11) Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

ANS :
Intervals

94 %	143	256
96 %	138	261
98 %	130	269

Q12) Below are the scores obtained by a student in tests

34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56

1) Find mean, median, variance, standard deviation.

2) What can we say about the student marks?

ANS :

Mean	41
Median	40,5
Variance	24,11
Std deviataion	4,91

Data has outliers.Maximum values lies between 35 - 40

Q13) What is the nature of skewness when mean, median of data are equal?

ANS: Symmetrical distribution

Q14) What is the nature of skewness when mean > median ?

ANS: Positively skewed

Q15) What is the nature of skewness when median > mean?
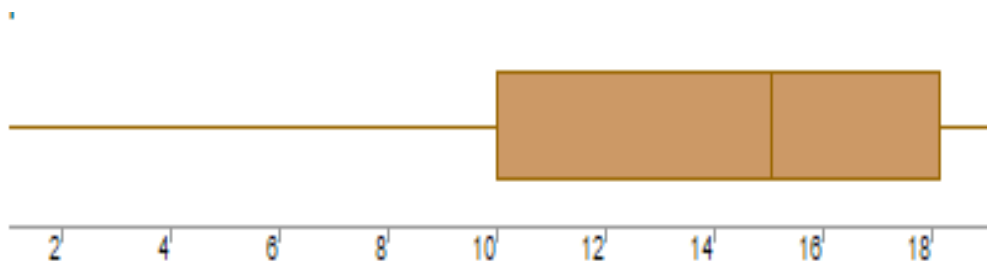
ANS: negatively skewed

Q16) What does positive kurtosis value indicates for a data ?

ANS: It indicates that the data is peaked and the graph is high.

Q17) What does negative kurtosis value indicates for a data?

ANS: It indicates that wider peak of data.

Q18) Answer the below questions using the below boxplot visualization.

What can we say about the distribution of the data?

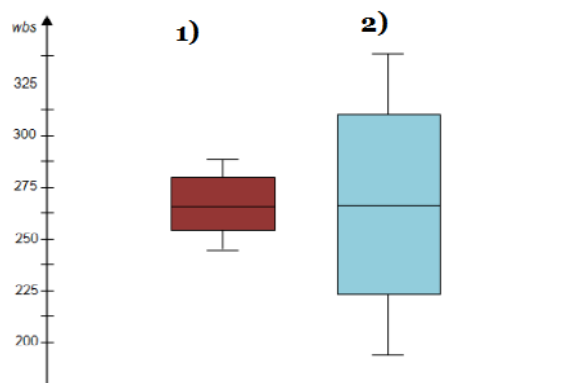ANS : The data does not have outliers and the data is more towards the right side.

What is nature of skewness of the data?

ANS : Negatively skewed

What will be the IQR of the data (approximately)?

ANS : 18-10=8

Q19) Comment on the below Boxplot visualizations?

Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

ANS.: Data has no outliers and boxplot1 has less range compared to boxplot2.

Q 20) Calculate probability from the given dataset for the below cases

Data _set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

a. P(MPG>38)
b. P(MPG<40)
c. P (20<MPG<50)

ANS : a) 0.3466

b) 0.7306

c) 0.900

Q 21) Check whether the data follows normal distribution
a) Check whether the MPG of Cars follows Normal Distribution
Dataset: Cars.csv

ANS: It is following a normal distribution

b) Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution
Dataset: wc-at.csv

ANS: Both are not following the normal distribution

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

ANS :

60 %	0,25
90 %	1,28
94 %	1,55

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

ANS :

95 %	1,710
96 %	1,828
99 %	2,492

Q 24) A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode → pt(tscore,df)

df → degrees of freedom.

ANS: The probability that 18 randomly selected bulbs would have an average life of no more than 260 days if the CEO's claim were true is: 0.678