Activity	Data Type
Number of beatings from Wife	Discrete
Results of rolling a dice	Discrete
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	Categorical
Number of kids	Discrete
Number of tickets in Indian railways	Discrete
Number of times married	Discrete
Gender (Male or Female)	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	Ratio
Hair Color	Nominal
Socioeconomic Status	Ordinal
Fahrenheit Temperature	Ratio
Height	Ratio

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

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

Ans- P - 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

       b) 0.167

       c) 0.139

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 = 0.476

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

	Mean	Median	Mode	Variance	Std. Dev	Range
Points	3.59	3.70	3.92	0.28	0.53	
Score	3.21	3.32	3.44	0.95	0.98	
Weight	17.85	17.71	17.02	3.19	1.78	

a) No one has the variable Mean = Median = Mode

b) Thus dataset score and weight has outliers.

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

Ans =

	Skewness	Kurtosis
Speed	-0.11	-0.50
Dist	0.80	0.40

A) dist is positively skewed where as speed is negatively skewed

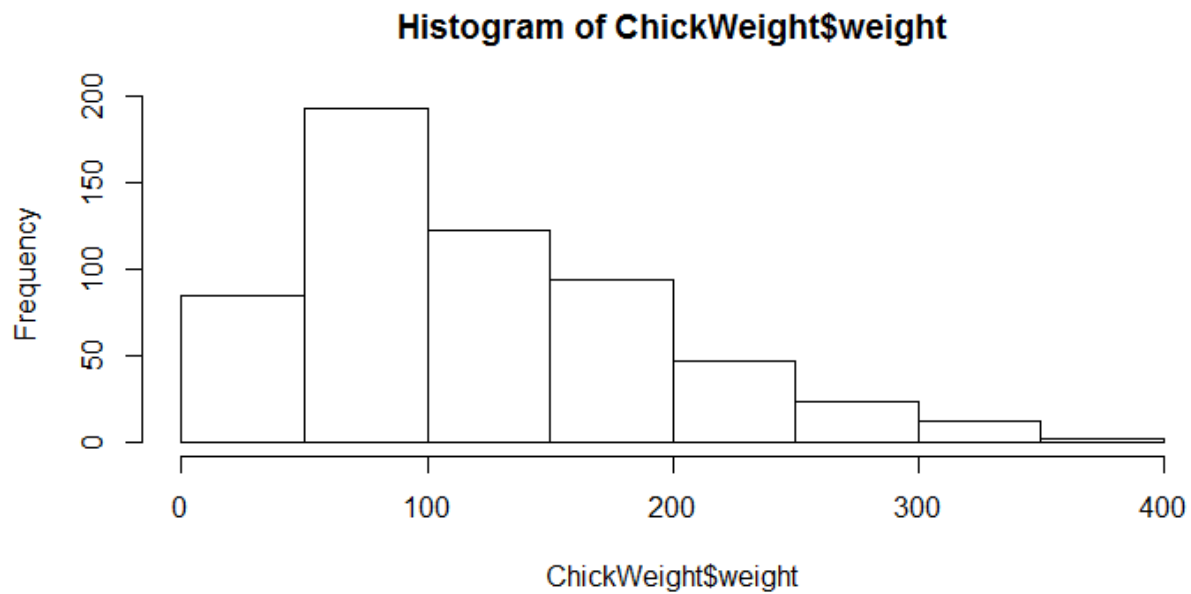B) Both dist and speed has positive kurtosis

Use Q9_b.csv

Ans)

	Skewness	Kurtosis
SP	1.55	5.72
WT	-0.59	3.82

A) "SP" is positively skewed where as "WT" is negatively skewed.

B) Both WT and SP has positive Kurtosis.

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight

ChickWeight$weight

ANS :

a) Majority of the Chicks has weight in range 50 – 100, followed by 100 -150 and 150 – 200

b) The data is positively Skewed.

c) The distribution is right skewed.

d) Mean value is greater than median
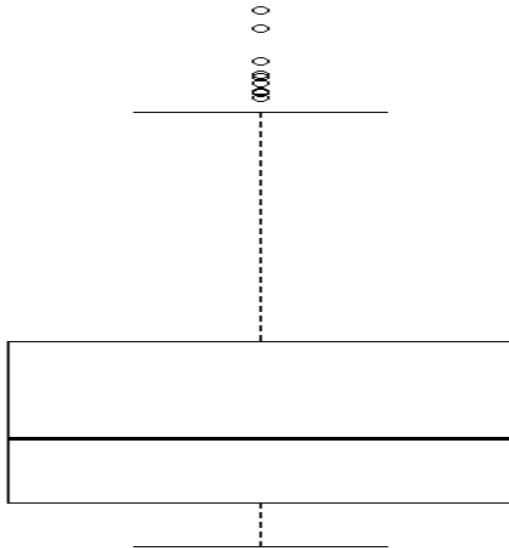
ANS:

a) Data has outliers

b) Data is positively skewed

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?

	94%	98%	96%
Upper	201.04	201.38	201.17
Lower	198.96	198.62	198.83

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.

Ans)

Mean	41
Median	40.50
Variance	25.53
std. Dev	5.05

2) What can we say about the student marks?

ANS:

a) Not normally distributed

b) Data has outlier

c) Majority of the students scored between 35 – 45 Marks

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

ANS: Skewness = 0. Perfectly symmetric bell shaped curve

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

ANS: Skewness = Positive. Data is distributed more on left

Q15) What is the nature of skewness when median > mean?

ANS: Skewness = Negative. Data is distributed more on right
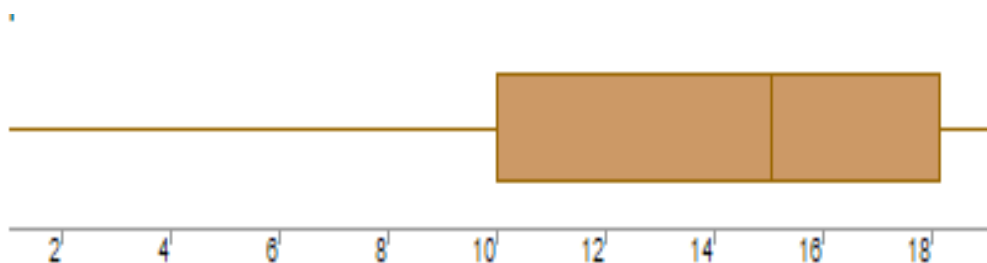
Q16) What does positive kurtosis value indicates for a data ?

ANS: High and narrow peak on central part of the data

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

ANS: wider peak on central part of the data

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

What can we say about the distribution of the data?

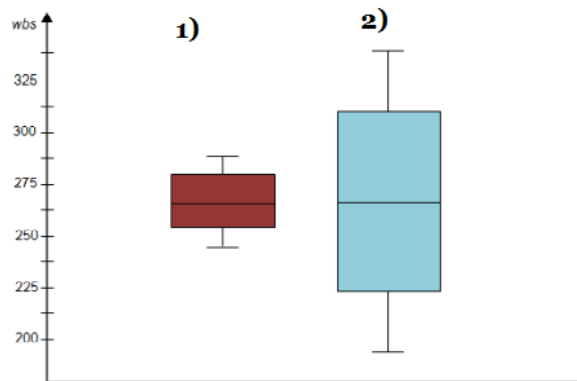ANS : The data is not symmetric. Data is more concentrated towards right side

What is nature of skewness of the data?

ANS : Skewness = Negative

What will be the IQR of the data (approximately)?
ANS: IQR data is 8 (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:

1) Data is Normally Distributed. No Outliers. Center around 262.5. Comparatively, first graph has less range

2) Data is Normally Distributed. No Outliers. Center around 262.5

   Comparatively, second graph has more range

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)
   ANS: 0.3476

b. P(MPG<40)
   ANS: 0.7293

c. P (20<MPG<50)

ANS: 0.0131

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

ANS: MPG is 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 AT and Waist doesn't follow Normal Distribution

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

Ans)

95%	1.880
96%	0.841
99%	1.644

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

Ans)

95%	2.063
96%	2.171
60%	2.796

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:

Given:

$\mu = 270$, n= 18, x bar = 260, s = 90

t score = (x bar - μ)/(s/sqrt(n))

=(260 − 270)/(90/sqrt(18))

=-10/21.23=-0.47

=0.322163

   So, Required Probability = 0.32