Q1) Identify the Data type for the Following:

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	Discrete
Number of kids	Discrete
Number of tickets in Indian railways	Discrete
Number of times married	Discrete
Gender (Male or Female)	Discrete

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	Interval
Weight	Ratio
Hair Color	Nominal
Socioeconomic Status	Ordinal
Fahrenheit Temperature	Interval
Height	Ratio
Type of living accommodation	Nominal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Interval
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Ordinal
Time on a Clock with Hands	Interval
Number of Children	Nominal
Religious Preference	Nominal
Barometer Pressure	Ratio
SAT Scores	Interval

Years of Education	Interval

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

Total Possible outcomes = 8

= (H,H,H), (H,H,T), (H,T,H), (H,T,T), (T,H,H), (T,H,T), (T,T,H), (T,T,T)

Number of outcomes that gives two heads and one tail = 3

=(H,H,T), (H,T,H), (T,H,H)

Probability of getting two heads and one tail = Number of favorable outcomes / Total number of outcome = 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 : Total possible number of outcomes = 6 x 6 = 3

a) Sum is equal to 1= 0 /36

b) Less than equal to 4= {(1,1)(2,1),(3,1),(1,2).(2,2),(1,3)}
= 6/36= 0.17

c) Sum is divisible by 2 and 3= {(1,5)(2,4),(3,3),(4,2).(5,2),(6,6)}
= 6/36=0.17

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: Total possible Outcome = {R,R,G,G,G,B,B}

= Probability of first ball being not blue is 5/7

= Probability of second ball being not blue is 4/6

= 5/7 and 4/6 = 5/7*4/6

= 20/42

=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: Expected number of candies = Sum of (X* p(x))

= Sum of (Candies count * Probability)

= 1*0.015 + 4*0.20 + 3*0.65 + 5*0.005 + 6*0.01 + 2*0.120

= 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:
Points : 3.596563
Score : 3.21725
Weigh : 17.84875

median :
Points : 3.695
Score : 3.325
Weigh : 17.71

mode :
Points: 3.92
Score : 3.44
Weigh:17.02

variance :
Points : 0.2858814
Score : 0.957379
Weigh: 3.193166

Standard deviation:

Points : 0.5346787
Score : 0.9784574
Weigh : 1.786943

Range:

	Min	Max
Points:	2.76	4.93
Score:	1.513	5.424
Weigh:	14.5	22.9

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: The mean is: Expected Value=(108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9 = 145.333.

Expected value =Sum(x.p(x)) = 150.976

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

Cars speed and distance

Use Q9_a.csv

skewness for speed= -0.352894, skewness value is negative so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed
kurtosis for speed= -0.035315 this kurtosis is negative is a flatter than normal curve

And for distance= 0.764757, right skewed (Positive) slight magnitude to right
Kurtosis for distance = 0.360105 . Positive values of kurtosis indicate that a distribution is peaked and possess thick tails

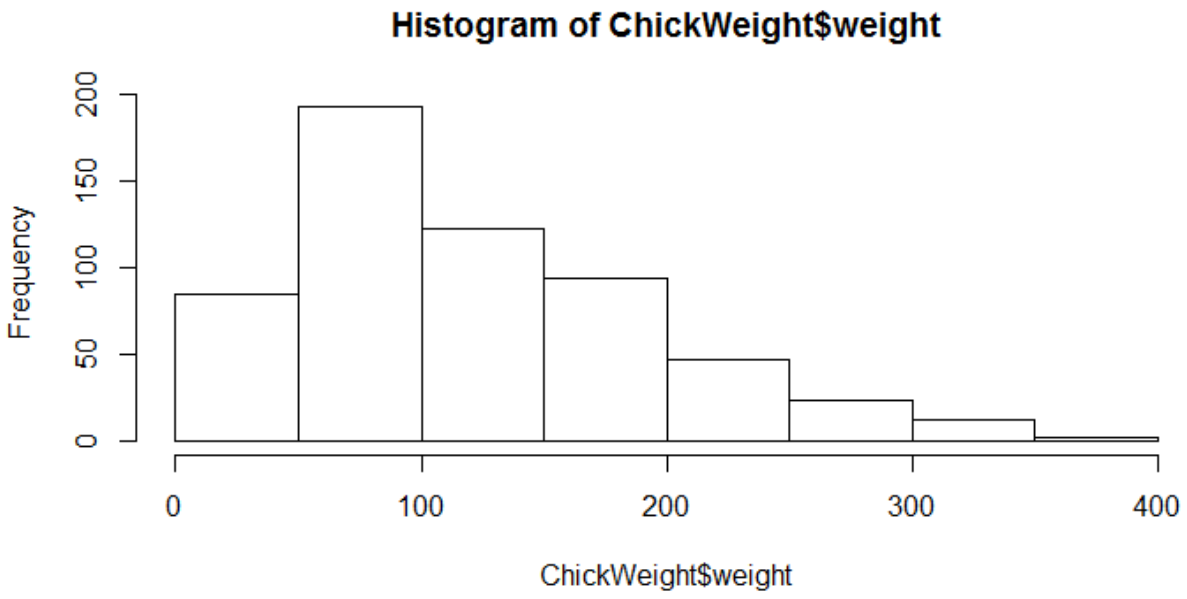
SP and Weight(WT)

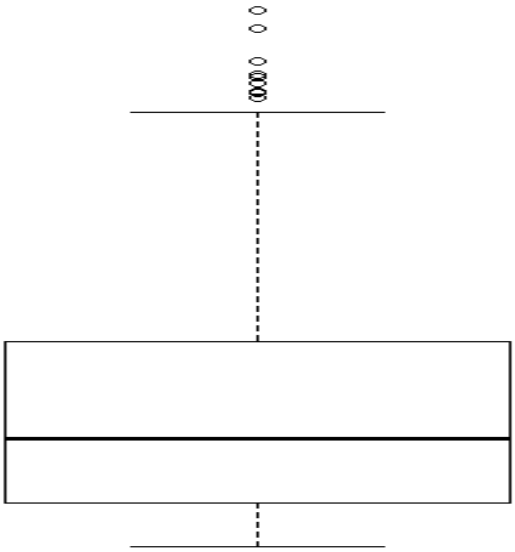
Use Q9_b.csv

skewness for sp= 1.611450, skewness value is positive so it is right skewed. Since magnitude is slightly greater than 0 it is slightly right skewed
kurtosis=2.977329

And for weight= -0.614753, left skewed (Negative) slight magnitude toright
Kurtosis = 0.950291

Q10) Draw inferences about the following boxplot & histogram





Ans: The maximum of the data points is present in the range 50-100 with frequency 200.

minimum range of weight is 400 and its frequency around 0-10.

Skewness- we can notice a long tail towards right so it is heavily right skewed.

Boxplot : Median is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between Q1 and bottom point.

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: formula = xbar+-z(sigma/sqrt(n))

xbar= 200 ,sigma = 30 ,n = 2000

1) Confidence interval= 94%

(1- σ/2)= 1-0.03) =0.97

error1<-qnorm(0.97)*sigma/sqrt(n)

left <- xbar- error1
right <- xbar+error1
left 198.7383
right 201.2617

2) Confidence interval= 96%

(1- σ/2)= 1-0.02) =0.98

error1<-qnorm(0.98)*sigma/sqrt(n)

left <- xbar- error1
right <- xbar+error1
left 198.6223
right 201.3777

3) Confidence interval= 98%

(1- σ/2)= 1-0.01) =0.99

error1<-qnorm(0.99)*sigma/sqrt(n)

left <- xbar- error1
right <- xbar+error1
left 198.4394
right 201.5606

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= 25.5294, Standard deviation= 5.052664

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

Ans: distribution is symmetric,

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

Ans: If the mean is greater than the median, the distribution is positively skewed.(right skewed)

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

Ans: If the median is greater than the mean, the distribution is negatively skewed.(left skewed)

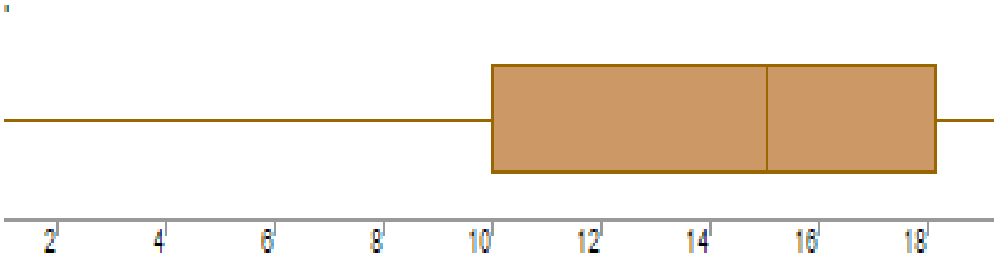
Q16) What does positive kurtosis value indicates for a data ?

Ans: The data is notmally distributed and kurtosis value is 0. that a distribution is fatter and has heavy tails.

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

Ans: Negative values of kurtosis indicate that a distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

Ans: Let's assume above box plot is about ages of the students in a school.

50% of the people are above 10 yrs. old and remaining are less.

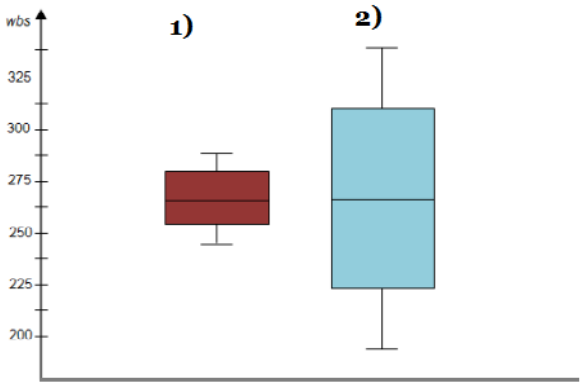
And students whose age is above 15 are approx. 40%.

What is nature of skewness of the data?

Ans: Left skewed, median is greater than mean.

What will be the IQR of the data (approximately)?
Ans: Approximately= -8

Q19) Comment on the below Boxplot visualizations?



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

Ans : By observing both the plots whisker's level is high in boxplot 2, mean and median are equal hence distribution is symmetrical.

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)

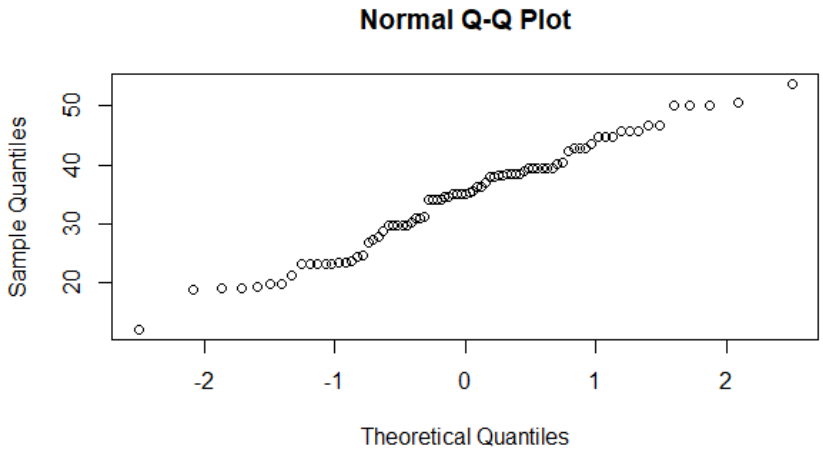
Ans:
```
MPG <- read.csv("cars.csv")

mean(MPG$MPG)
>> 34.42
sd(MPG$MPG)
>>9.13
pnorm(38,34.42,9.131)
>>0.65
1-pnorm(38,34.42,9.131)
>>0.34
pnorm(40,34.42,9.131)
>>0.72
pnorm(50,34.42,9.131)-pnorm(20,34.42,9.131)
>>0.89
pnorm(50,34.42,9.131)
>>0.95
pnorm(20,34.42,9.131)
>>0.057
```

Q 21) Check whether the data follows normal distribution

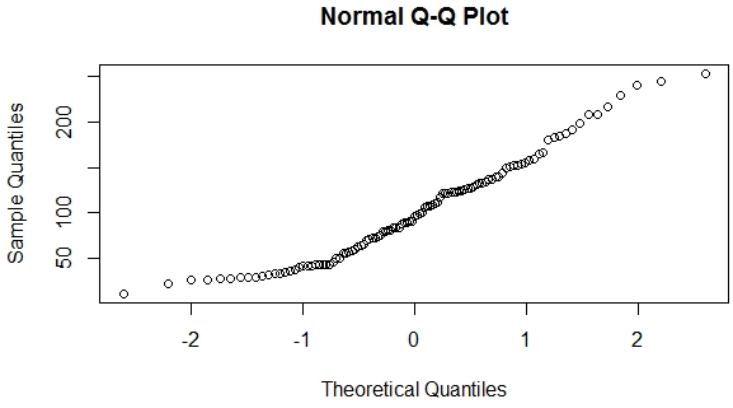
a) Check whether the MPG of Cars follows Normal Distribution
   Dataset: Cars.csv

```
qqnorm(MPG$MPG)
```



No it don't follow normal distribution curve, data is left skewed and have a positive kurtosis.

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



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

(1- σ/2)= 1-0.05) =0.95 confidence interval is 1.644

Z score of 94%

(1- σ/2)= 1-0.03) =0.97 confidence interval is 1.88

Z score of 60%

(1- σ/2)= 1-0.20) =0.80 confidence interval is 0.841

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

Ans: TSCORE CALCULATION

T((1,alpha),(n-1))
Here n = 25
n-1 = 24

Hence t score values will be:
95%= qt(0.975,24)= 2.06389996

96%= qt(0.98,24)=2.171545

99%= qt(0.995,24)= 2.79694

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:

Mu= 270
Xbar= 260

N2=17
S=90
terror<-S/sqrt(N2)
tscore<-(Xbar-Mu)/terror
tscore = -0.4581
pt(tscore,df=17) =0.3263

so the average life of bulb not more than 260 days will have 32.6% probability