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	Ratio
Religious Preference	Nominal
Barometer Pressure	Ordinal

Q1) Identify the Data type for the Following:

SAT Scores	Interval
Years of Education	Ratio

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

P (Two heads and one tail) = N (Event (Two heads and one tail)) / N (Event (Three coins tossed)) = 3/8 = 0.375 = 37.5%

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

- Equal to 1
- Less than or equal to 4
- Sum is divisible by 2 and 3

Number of possible outcomes for the above event is N (Event (Two dice rolled)) = 6^2 = 36

a.) P (sum is Equal to 1) = '0' zero null none.

b.) P (Sum is less than or equal to 4) =

N (Event (Sum is less than or equal to 4)) / N (Event (Two dice rolled))

= 6 / 36 = 1/6 = 0.166 = 16.66%

c.) P (Sum is divisible by 2 and 3) =

N (Event (Sum is divisible by 2 and 3)) / N(Event (Two dice rolled)) = 6 / 36 = 1/6 = 0.16 = 16.66%

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?

Q1) Identify the Data type for the Following:

Total number of balls =7 balls

N (Event (2 balls are drawn randomly from bag) =

7! / 2! * 5! = (7654321) / (21) * (54321)

N (Event (2 balls are drawn randomly from bag) =

(76)/ (21) = 21

If none of them drawn 2 balls are blue = 7 – 2 = 5 N

(Event (None of the balls drawn is blue) = 5! / 2! * 3! = (54) / (2*1) = 10 P (None of the balls drawn is blue) = N (Event (None of the balls drawn is blue) / N (Event (2 balls are drawn randomly from bag) = 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

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: 0.015+0.8+1.95+0.025+0.06+0.24 = 3.09

Q1) Identify the Data type for the Following:

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

```
View(data)

head(data)

tail(data)

# MEAN

mean(data$Points)  #3.596563

mean(data$Score)   #3.21725

mean(data$Weigh)   #17.84875

# MEDIAN

median(data$Points)

3.695

median(data$Score)

3.325

median(data$Weigh)

17.71

# MODE

library('modeest')

mfv(data$Points)

 3.07 3.92
```

Q1) Identify the Data type for the Following:

mfv(data$Score)

3.44

mfv(data$Weigh)

17.02 18.90

 # VARIANCE

var(data$Points) # 0.2858814

var(data$Score) # 0.957379

var(data$Weigh) # 3.193166

 # STANDARD DEVIATION

sd(data$Points) # 0.5346787

sd(data$Score) # 0.9784574

sd(data$Weigh) # 1.786943

 # RANGE

range[min - max]

range(data$Points) # 2.76 4.93

range(data$Score) # 1.513 5.424

range(data$Weigh) # 14.5 22.9

Q8) Calculate Expected Value for the problem below

- 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?

Solution:----

```
x <- data.frame(x=1:9, weights = c(108, 110, 123, 134, 135, 145, 167, 187, 199))
```

Q1) Identify the Data type for the Following:

x

mean(x$weights) # 145.3333

(OR)

solution :--

P(x) 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9

E(x) 108, 110, 123, 134, 135, 145, 167, 187, 199

Expected Value = ∑ (probability * Value)

∑ P(x).E(x)

Expected Value = (1/9)(108) + (1/9)110 + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) (1308) = > 145.33

Expected Value of the Weight of that patient = 145.33

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

Cars speed and distance

Use Q9_a.csv

View(Q9_a)

head(Q9_a)

tail(Q9_a)

mean(Q9_a$speed) # 15.4

mean(Q9_a$dist) # 42.98

Q1) Identify the Data type for the Following:

```
median(Q9_a$speed)  # 15
median(Q9_a$dist)   # 36

mfv(Q9_a$speed)     # 20
mfv(Q9_a$dist)      #26

library(moments)
skewness(Q9_a$speed) # -0.1139548
skewness(Q9_a$dist)  # 0.7824835

kurtosis(Q9_a$speed) # 2.422853
kurtosis(Q9_a$dist)  #  3.248019
```

SP and Weight(WT)

Use Q9_b.csv

```
View(Q9_b)
head(Q9_b)
tail(Q9_b)

mean(Q9_b$ SP)  #  121.5403
mean(Q9_b$ WT)   #  32.41258
```

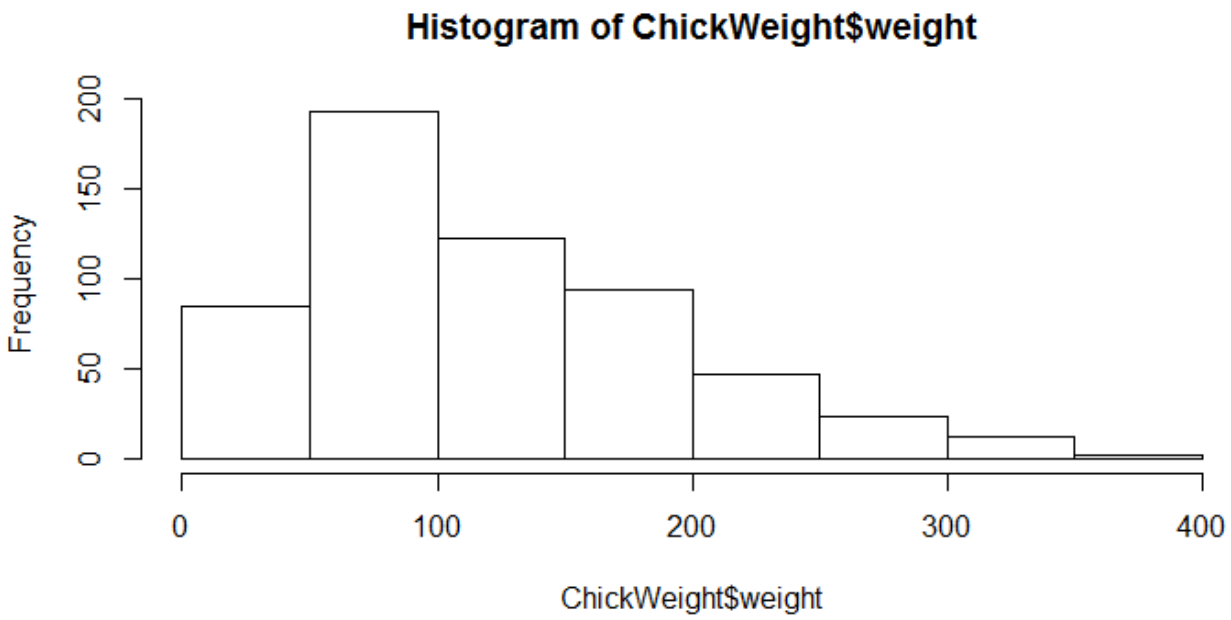
Q1) Identify the Data type for the Following:

```
median(Q9_b$SP)  #  118.2087
median(Q9_b$WT)   #   32.73452

mfv(Q9_b$SP)     #  118.289
```

```
library(moments)
skewness(Q9_b$SP)   # 1.581454
skewness(Q9_b$WT)   # -0.6033099

kurtosis(Q9_b$SP)   # 5.723521
kurtosis(Q9_b$WT)   # 3.819466
```

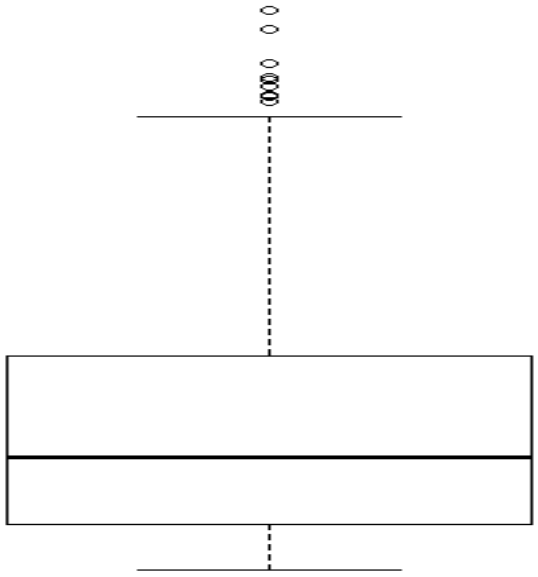
Q10) Draw inferences about the following boxplot & histogram

Q1) Identify the Data type for the Following:



The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

.



The boxplot has outliers on the maximum side.

Q1) Identify the Data type for the Following:

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?

```
import numpy as np
import pandas as pd
from scipy import stats
from scipy.stats import norm
# Avg. weight of Adult in Mexico with 94% CI
stats.norm.interval(0.94,200,30/(2000**0.5))
```

(198.738325292158, 201.261674707842)

```
# Avg. weight of Adult in Mexico with 98% CI
stats.norm.interval(0.98,200,30/(2000**0.5))
```

(198.43943840429978, 201.56056159570022)

```
# Avg. weight of Adult in Mexico with 96% CI
stats.norm.interval(0.96,200,30/(2000**0.5))
```

(198.62230334813333, 201.37769665186667)

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

- Find mean, median, variance, standard deviation.
- What can we say about the student marks?

Q1) Identify the Data type for the Following:

```
x=data.frame(x=1:18, scores =
c(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56))

x

mean(x$scores)    # 41

median(x$scores)  # 40.5

mfv(x$scores)   #  41

var(x$scores)  #  25.52941

sd(x$scores)  #   5.052664

boxplot(x$scores)

hist(x$scores)
```

we don't have outliers and the data is slightly skewed towards right because mean is greater than median.

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

Zero skew and Perfectly symmetrical

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

Positively skewed

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

Negatively skewed

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

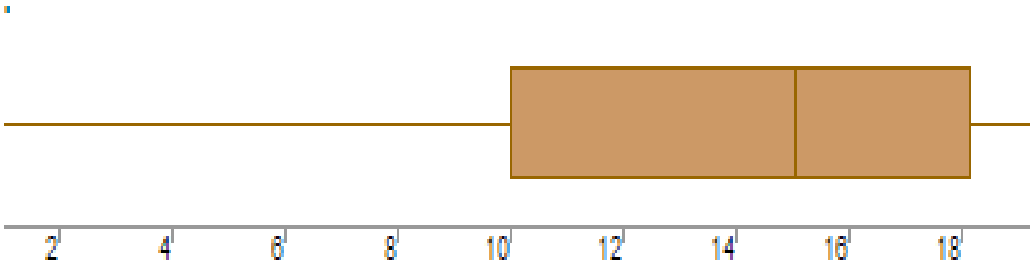
A positive value tells you that you have heavy-tails

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

A negative value means that you have light-tails

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

Q1) Identify the Data type for the Following:



What can we say about the distribution of the data?

It is not normally distributed the median is towards the higher value

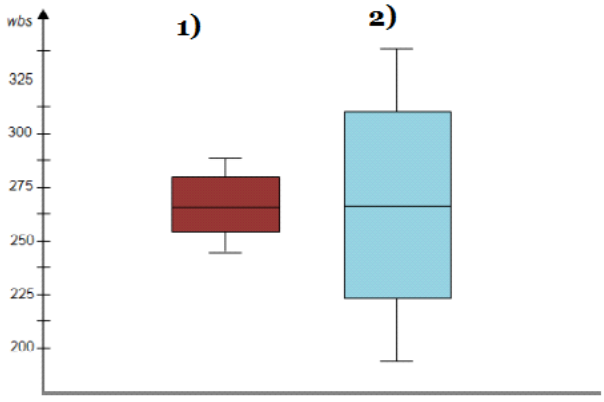
What is nature of skewness of the data?

It is a skewed towards left. The whisker range of minimum value is greater than maximum

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

The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 18 – 10 =8

Q19) Comment on the below Boxplot visualizations?



Q1) Identify the Data type for the Following:

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

Here firstly there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker 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

- P(MPG>38)

```
Prob_MPG_greater_than_38 = np.round(1 - stats.norm.cdf(38, loc=
q20.MPG.mean(), scale= q20.MPG.std()),3)
print('P(MPG>38)=',Prob_MPG_greater_than_38)
```

P(MPG>38)= 0.348

- P(MPG<40)

```
prob_MPG_less_than_40 = np.round(stats.norm.cdf(40, loc = q20.MPG.mean(),
scale = q20.MPG.std()),3) print('P(MPG<40)=',prob_MPG_less_than_40)
```

P(MPG<40)= 0.729

c. P (20<MPG<50)

```
prob_MPG_greater_than_20 = np.round(1-stats.norm.cdf(20, loc =
q20.MPG.mean(), scale = q20.MPG.std()),3)
print('p(MPG>20)=',(prob_MPG_greater_than_20)) p(MPG>20)= 0.943
```

```
prob_MPG_less_than_50 = np.round(stats.norm.cdf(50, loc = q20.MPG.mean(),
scale = q20.MPG.std()),3) print('P(MPG<50)=',(prob_MPG_less_than_50))
```

P(MPG<50)= 0.956

Q1) Identify the Data type for the Following:

prob_MPG_greaterthan20_and_lessthan50= (prob_MPG_less_than_50) - (prob_MPG_greater_than_20)
print('P(20<MPG<50)=',(prob_MPG_greaterthan20_and_lessthan50))
P(20<MPG<50)= 0.0130000000000000012

Q 21) Check whether the data follows normal distribution

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

MPG of cars follows normal distribution

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

Adipose Tissue (AT) and Waist does not follow Normal Distribution

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

z value for 90% confidence interval

print('Z score for 60% Conifidence Intervla ='',np.round(stats.norm.ppf(.05),4)) Z score for 60% Conifidence Intervla = -1.6449

z value for 94% confidence interval

print('Z score for 60% Conifidence Intervla ='',np.round(stats.norm.ppf(.03),4)) Z score for 60% Conifidence Intervla = -1.8808

z value for 60% confidence interval

Q1) Identify the Data type for the Following:

```
print('Z score for 60% Conifidence Intervla
=',np.round(stats.norm.ppf(.2),4)) Z score for 60% Conifidence Intervla = -
0.8416
```

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

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

t score for 95% confidence interval

```
print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4))
T score for 95% Confidence Interval = -2.0639
```

t value for 94% confidence interval

```
print('T score for 94% Confidence Inteval =',np.round(stats.t.ppf(0.03,df=24),4)) T
score for 94% Confidence Inteval = -1.974
```

t value for 99% Confidence Interval

```
print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4))
T score for 95% Confidence Interval = -2.7969
```

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

Q1) Identify the Data type for the Following:

import numpy as np

Import scipy as stats t_score

= (x - pop mean) / (sample standard daviation / square root of sample size) (260-270)/90/np.sqrt(18)) t_score = -0.471 stats.t.cdf(t_score, df = 17) 0.32 = 32%