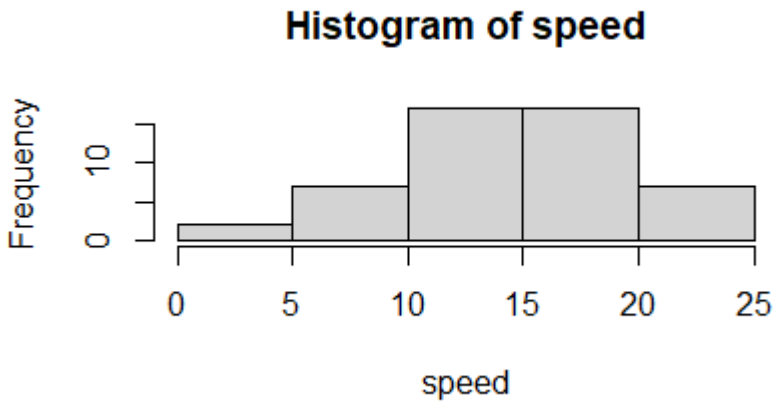
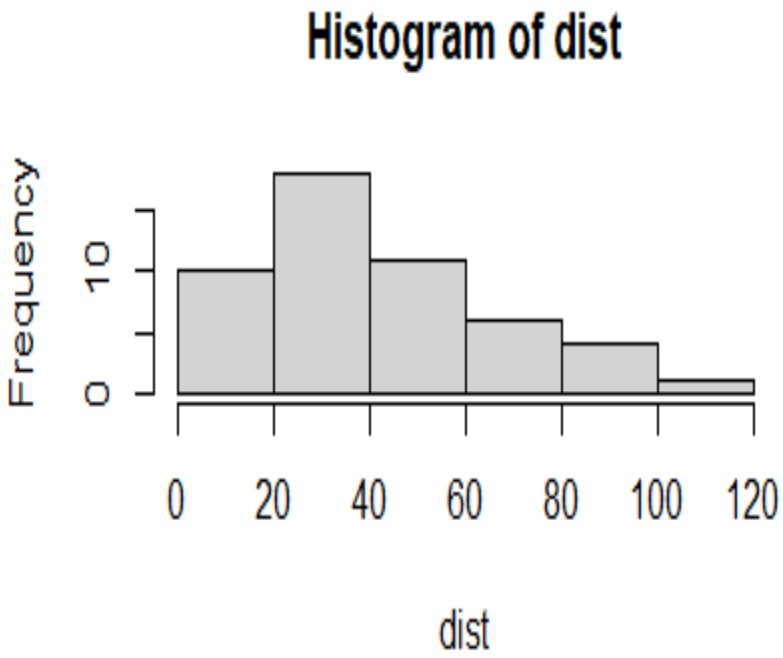
Basic Statistics (Module -3)

Q1) Calculate Skewness, Kurtosis & draw inferences on the following data

a. Cars speed and distance

speed	dist
4	2
4	10
7	4
7	22
8	16
9	10
10	18
10	26
10	34
11	17
11	28
12	14
12	20
12	24
12	28
13	26
13	34
13	34
13	46
14	26
14	36
14	60
14	80
15	20
15	26
15	54
16	32





Ans =

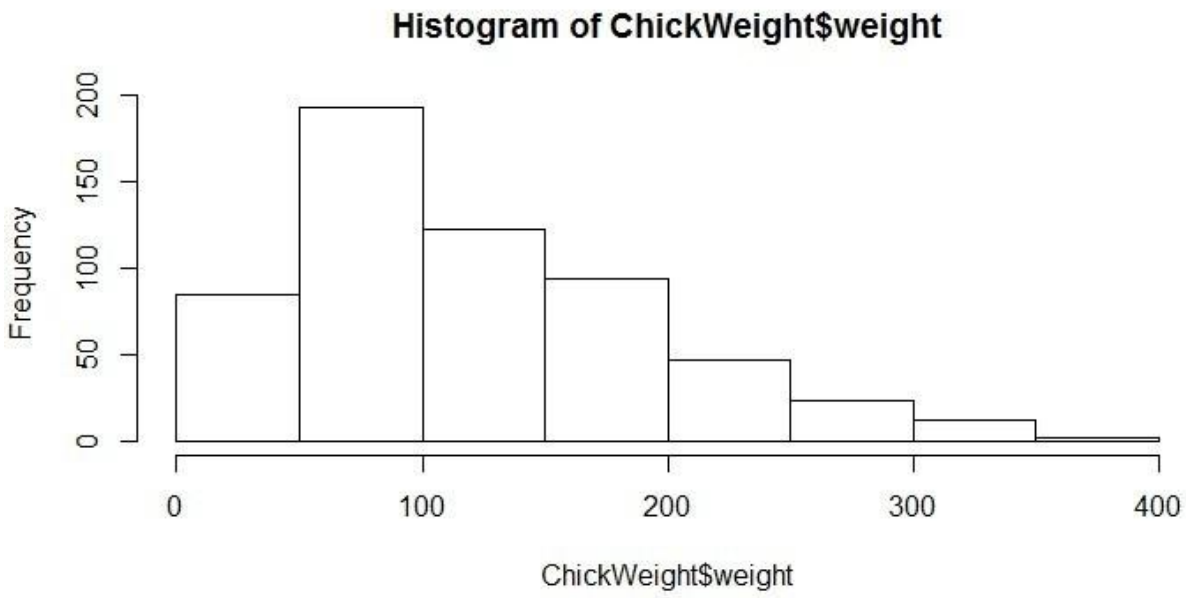
skewness	-0.11751	0.806895
kurtosis	-0.50899	0.405053

b. Top Speed (SP) and Weight (WT)

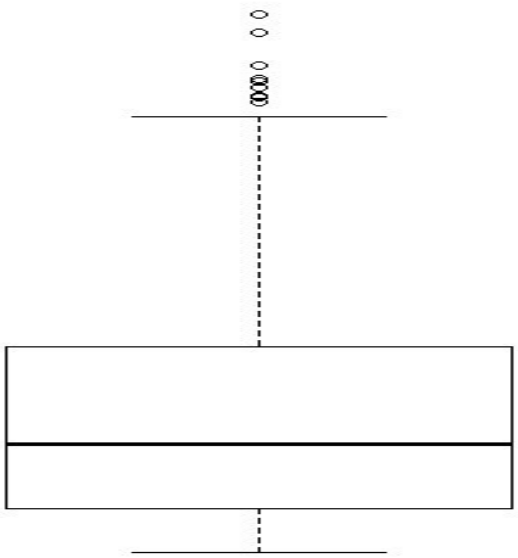
SP	WT
104.1854	28.76206
105.4613	30.46683
105.4613	30.1936
113.4613	30.63211
104.4613	29.88915
113.1854	29.59177
105.4613	30.30848
102.5985	15.84776
102.5985	16.35948
115.6452	30.92015
111.1854	29.36334
117.5985	15.75353
122.1051	32.81359
111.1854	29.37844
108.1854	29.34728
111.1854	29.60453
114.3693	29.53578
117.5985	16.19412
114.3693	29.92939
118.4729	33.51697
119.1051	32.32465
110.8408	34.90821
120.289	32.67583
113.8291	31.83712
119.1854	28.78173
114.5985	16.04317
120.7605	38.06282
119.1051	32.83507
99.56491	34.48321
121.8408	35.54936
113.4846	37.04235
112.289	33.23436
119.9211	31.38004
121.3926	37.57329

skewness	1.61145	-0.61475
kurtosis	2.977329	0.950291

Q2) Draw inferences about the following boxplot & histogram



Ans = Right side skewed or positively skewed



Ans = the interface for this box plot is positively skewed

Q3) 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 = Identify a sample statistic. Since we are trying to estimate the mean weight in the population, we choose the mean weight in our sample (200) as the sample statistic

We are working with a 94 % confidence level

Find standard error. The standard error (SE) of the mean is:

$$SE = \frac{s}{\sqrt{n}} = \frac{30}{\sqrt{2000}} = 0.670$$

= Compute alpha (α): α = 1 - (confidence level / 100) = 0.9933

= Find the critical probability (p*): p* = 1 - α/2 = 1 – 0.9933/2 = 0.4966

= find the degree of freedom (df): df = n-1 = 2000 – 1 = 1999

= The critical value is the t score having 1999 degrees of freedom and a probability equal to 0.4966

= critical value is -0.009

= critical value * standard error = - 0.009 * 0.94 = -0.00846

= - 0.009 * 0.98 = -0.00882

= -0.009 * 0.96 = - 0.00846

Q4) 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 = The scores are in uniformly distribution data in Ascending order

Mean	41
Median	40.5
Variance	25.52
Standard deviation	5.05

Q5) What is the nature of skewness when mean, median of data are equal?

Ans = Normalized Skewness

Q6) What is the nature of skewness when mean > median?

Ans = Right skewed

Q7) What is the nature of skewness when median > mean?

Ans = Left Skewed

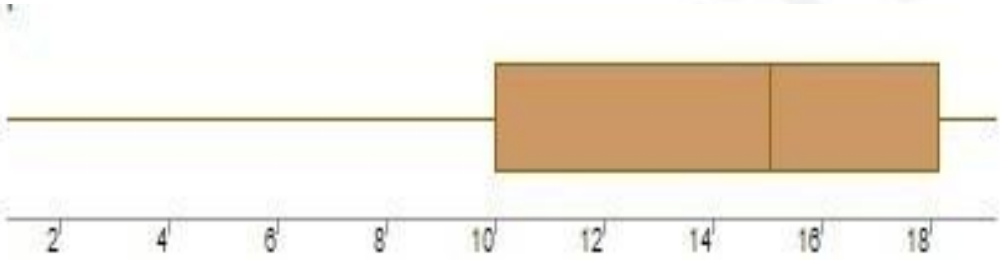
Q8) What does positive kurtosis value indicates for a data?

Ans = Sharp peak in the plot. less gap between tails to x-axis

Q9) What does negative kurtosis value indicates for a data?

Ans = Border peak under the curve and more gap between tails and x-axis

Q10) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

Ans = The data is distributed in De-assigned format

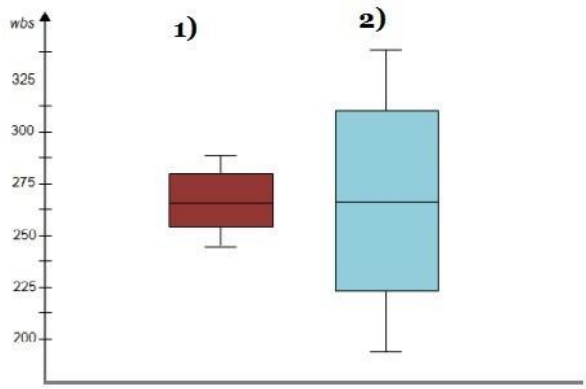
What is nature of skewness of the data?

Ans = Left side skewed

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

Ans = Q3-Q1
= 18-10
= 8 is IQR

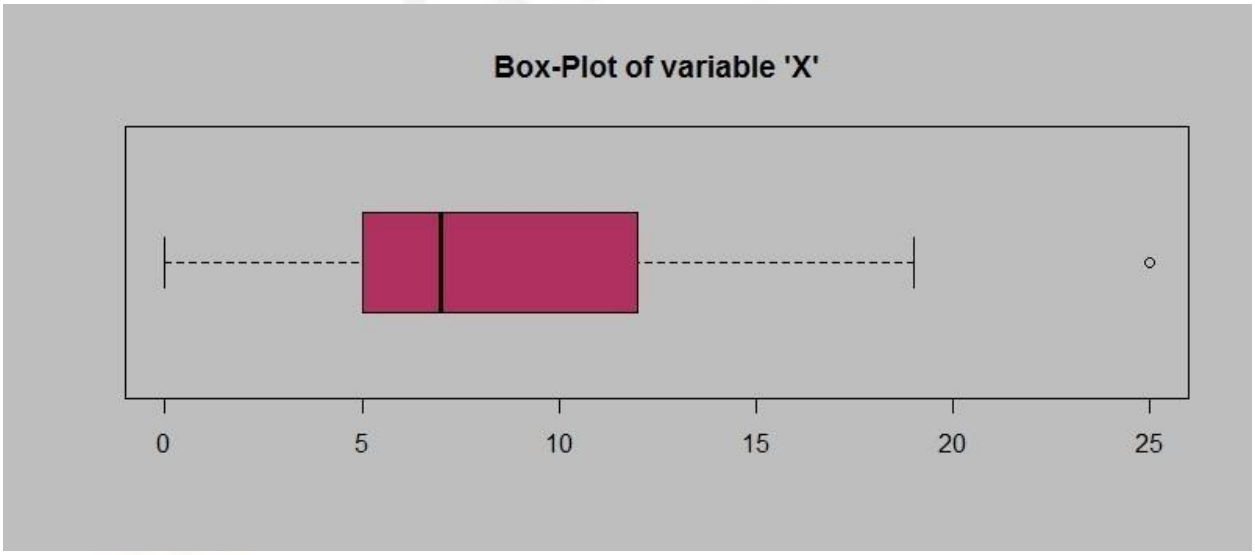
Q11) Comment on the below Boxplot visualizations?



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

Ans = The box plot 1 designed with range = 3 , The second one range is = 1.5

Q12)



Answer the following three questions based on the boxplot above.

(i) What is inter-quartile range of thisdataset? (please approximate the numbers) In one line, explain what this value implies.

Ans = IQR is Q3-Q1

= 12 – 5

= 7 (The value implies that is Mean > median)

(ii) What can we say about the skewness of this dataset?

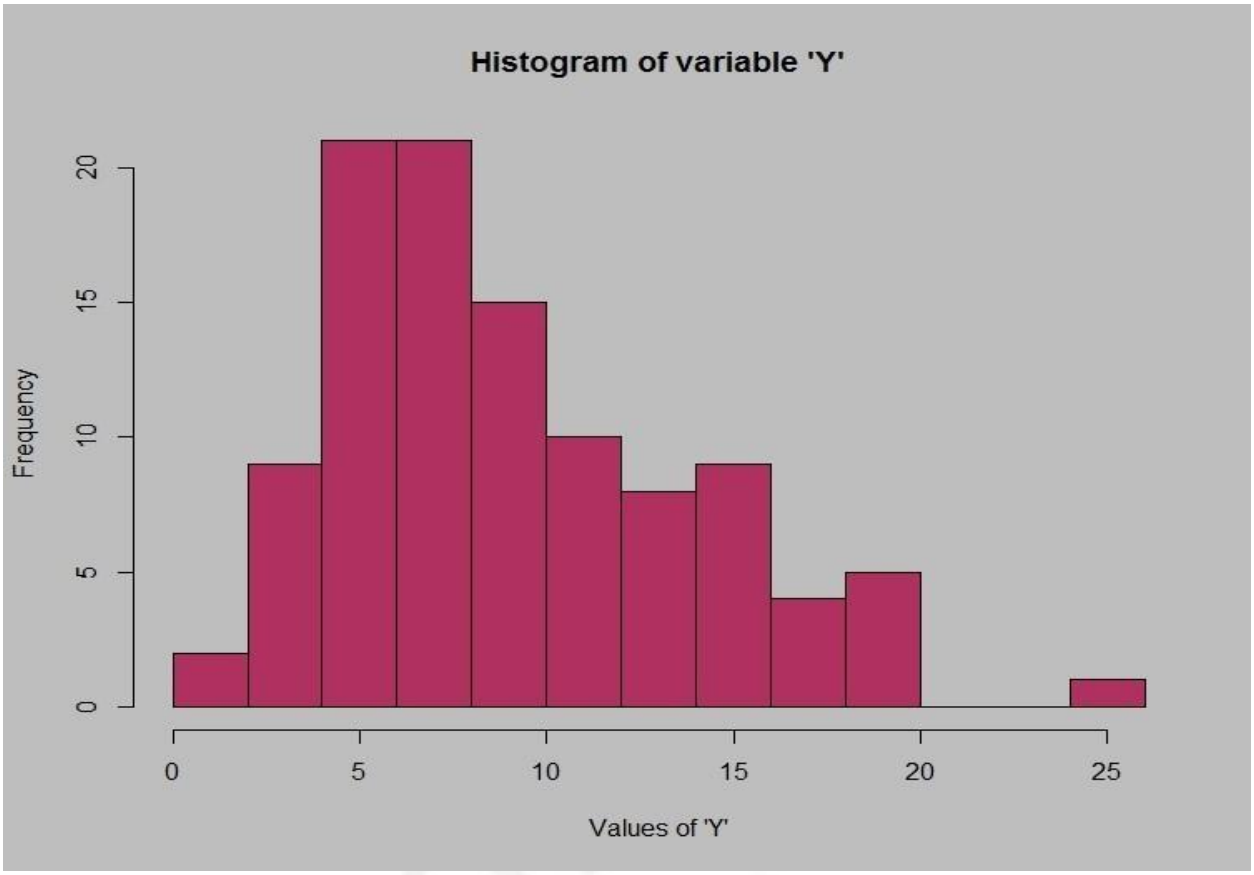
Ans = The data is positively skewed

(ie : The data constitute higher frequency of high value)

(iii) If it was found that the data point with the value 25 is actually 2.5, how would the new boxplot be affected?

Ans = 3

Q13)



Answer the following three questions based on the histogram above.

(i) Where would the mode of this dataset lie?

Ans = The mode lie on the 7 on the X – axis (values of Y)

(ii) Comment on the skewness of the dataset.

Ans = The data is Right side skewed

(iii) Suppose that the above histogram and the boxplot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset.

Hints:

1. Business Problem

    1.1. Objective

    1.2. Constraints (if any)

2. For each assignment the solution should be submitted in the below format

3. Research and Perform all possible steps for obtaining solution

4. For Basic Statistics explanation of the solutions should be documented in black and white along with the codes.

    One must follow these guidelines as well:

    4.1. Be thorough with the concepts of Probability, Central Limit Theorem and Perform the calculation stepwise

    4.2. For True/False Questions, explanation is must.

    4.3. R & Python code for Univariate Analysis (histogram, box plot, bar plots etc.) for data distribution to be attached

5. All the codes (executable programs) should execute without errors