|  |  |
| --- | --- |
| 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 |

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

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

**Answer**= The probability of getting two heads and one tail when three coins are tossed can be calculated by determining the number of possible outcomes that result in two heads and one tail, and dividing that number of possible outcomes when three coins are tossed.

Since each coin can either come up head or tails, the number of possible outcomes for three coins is 2×2×2=8. out of these 8 possible outcomes, the number of outcomes that result in two heads and one tail is (HHT, HTH, THH).

Therefore, the probability of getting two heads and one tail when three coins are tossed is 3/8 or 0.375

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

1. Equal to 1 = There is no equal number for 1 when two dice are rolled.
2. Less than or equal to 4= The total number of possible outcomes when two dice are rolled is 6 x 6 = 36. Out of these 36 possible outcomes, the number of outcomes that result in a sum less than or equal to 4 is 6/36, since the sums that are less than or equal to 4 are1, 2, 3, and there are (1,1) (1,2)(2,1)(2,2)(3,1)(1,3). Therefore, the probability of getting a sum less than or equal to 4 when two dice are rolled is 6/36 which is 1/6.
3. Sum is divisible by 2 and 3 = The probability of getting a sum that is divisible by 2 and 3 when two dice are rolled is 0, since there is no possible outcome when the sum of two dice is divisible by both 2 and 3.

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?

**Answer** =

Given data ,

2 red, 3 green,2 blue balls = 7 total balls

So, by using permutation and combination concept

=

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

**Answer**=

the formula is

For total,

0.015×1+0.20×4+0.65×3+0.005×5+0.01×6+0.120×2 = 3.036

For child A or child B

0.015×1+ 0.20×4 = 0.815

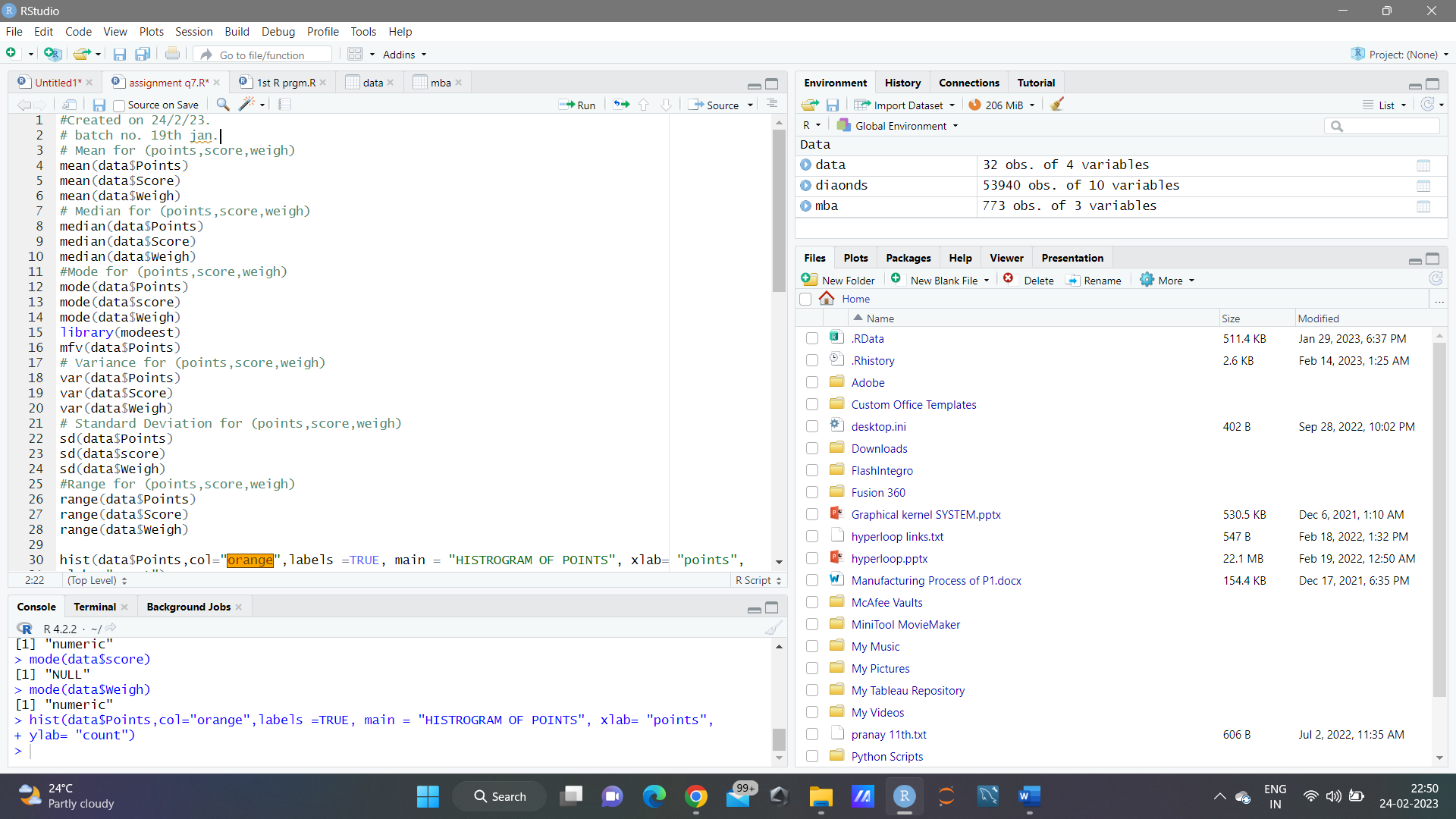
Q7) Calculate Mean, Median, Mode, Variance, Standar Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

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

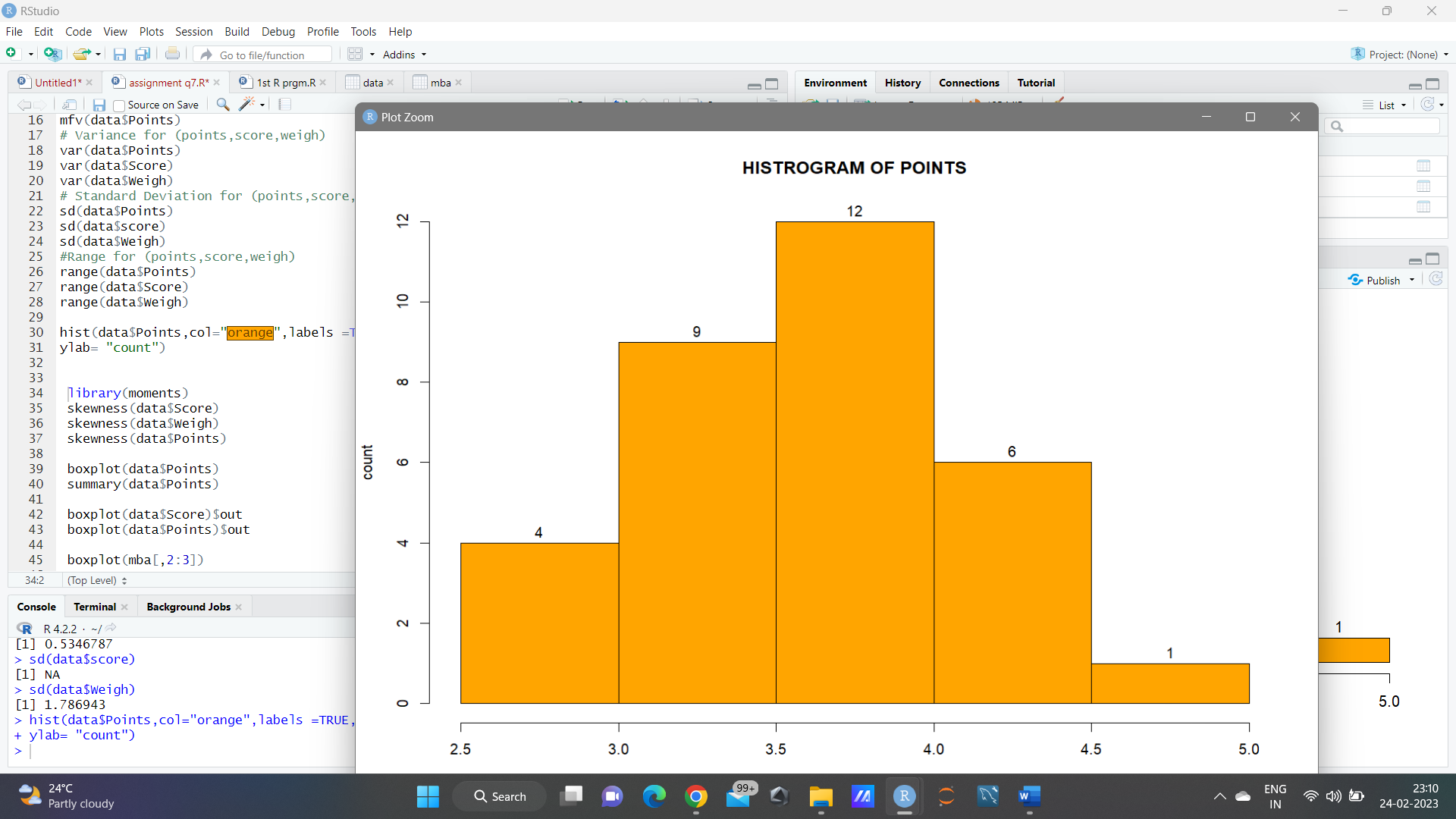
**Use Q7.csv file**

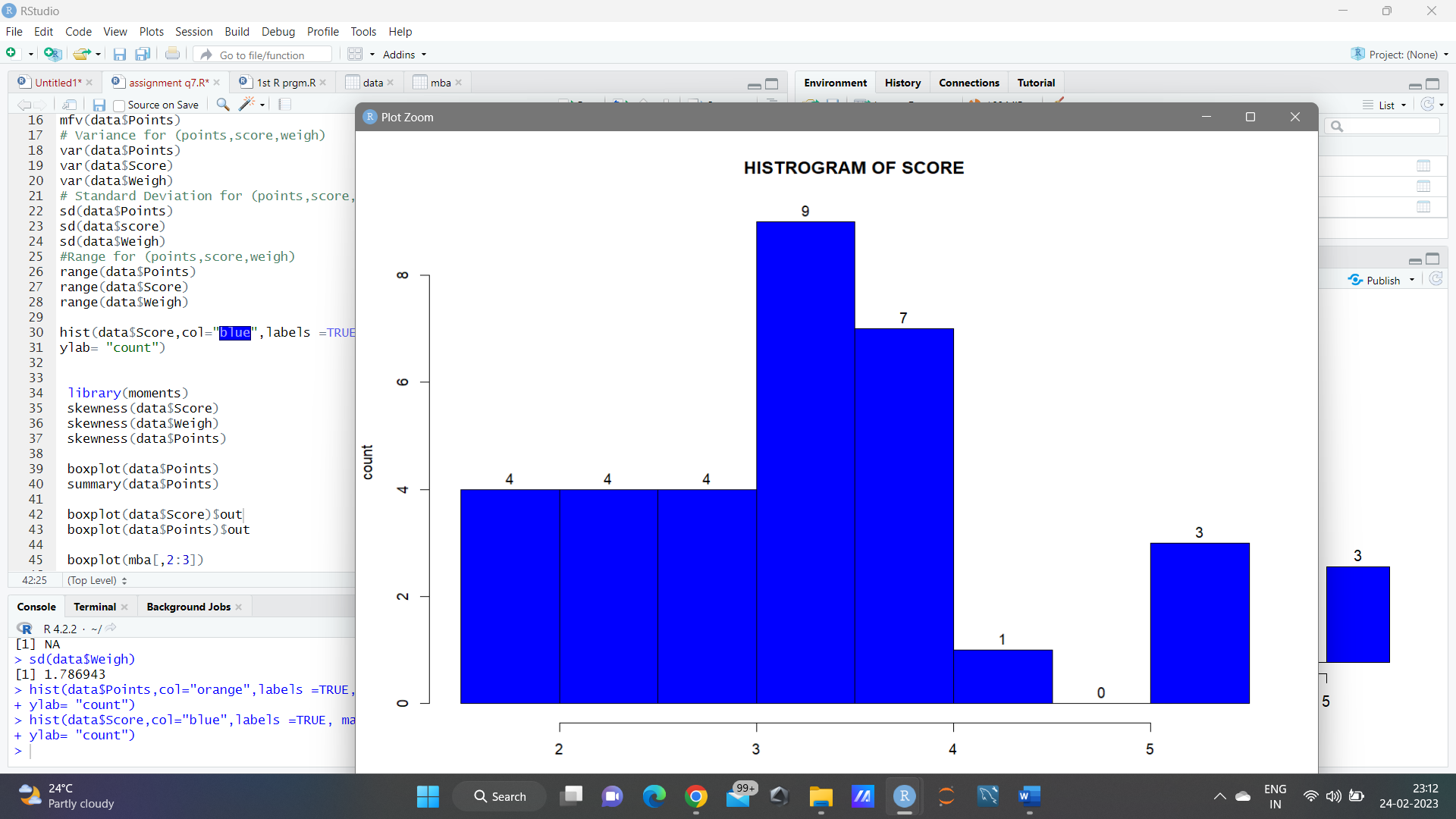
**Ans:** Coding is done in R programming.

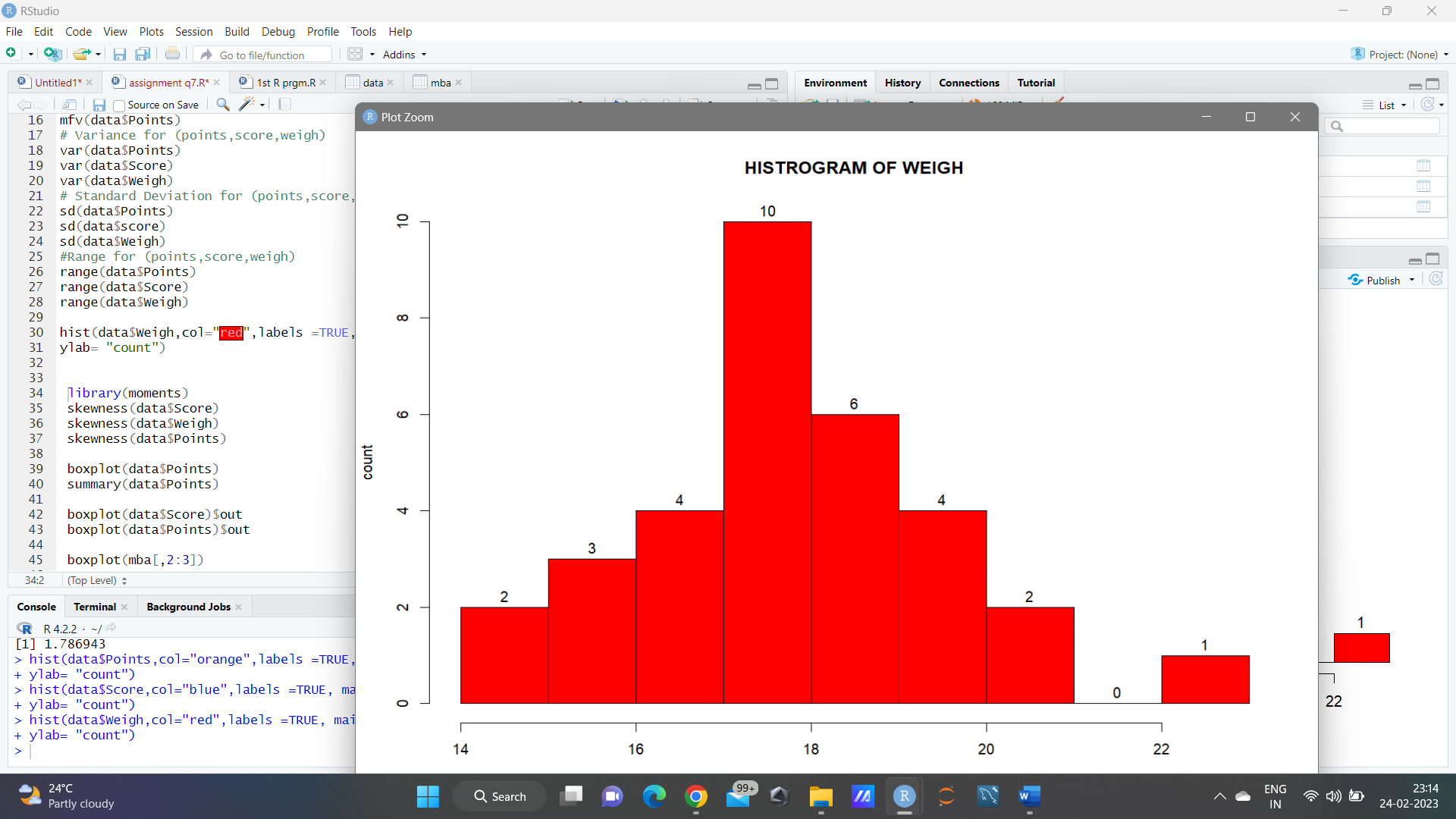


|  |  |  |  |
| --- | --- | --- | --- |
|  | POINTS | SCORE | WEIGH |
| MEAN | 3.69 | 3.21 | 17.84 |
| MEDIAN | 3.69 | 3.325 | 17.71 |
| MODE | 3.9 | 3.44 | 18.90 |
| VARIANCE | 0.285 | 0.95 | 3.19 |
| STANDAR DEVIATION | 0.53 | NA | 1.786 |
| RANGE | 2.76 | 1.513 | 14.5 |

Inferences: Histograms for (points, score, weigh);

 ----- POINTS

 -----SCORE

-----WEIGH

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

**Ans:** There are 9 patients so the probability of selecting each patient P(X)=1/9.

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

Expected value = P(X). E(X)

= (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)

= 145.3

- The Expected Value of the Weight of that patient is 145.3

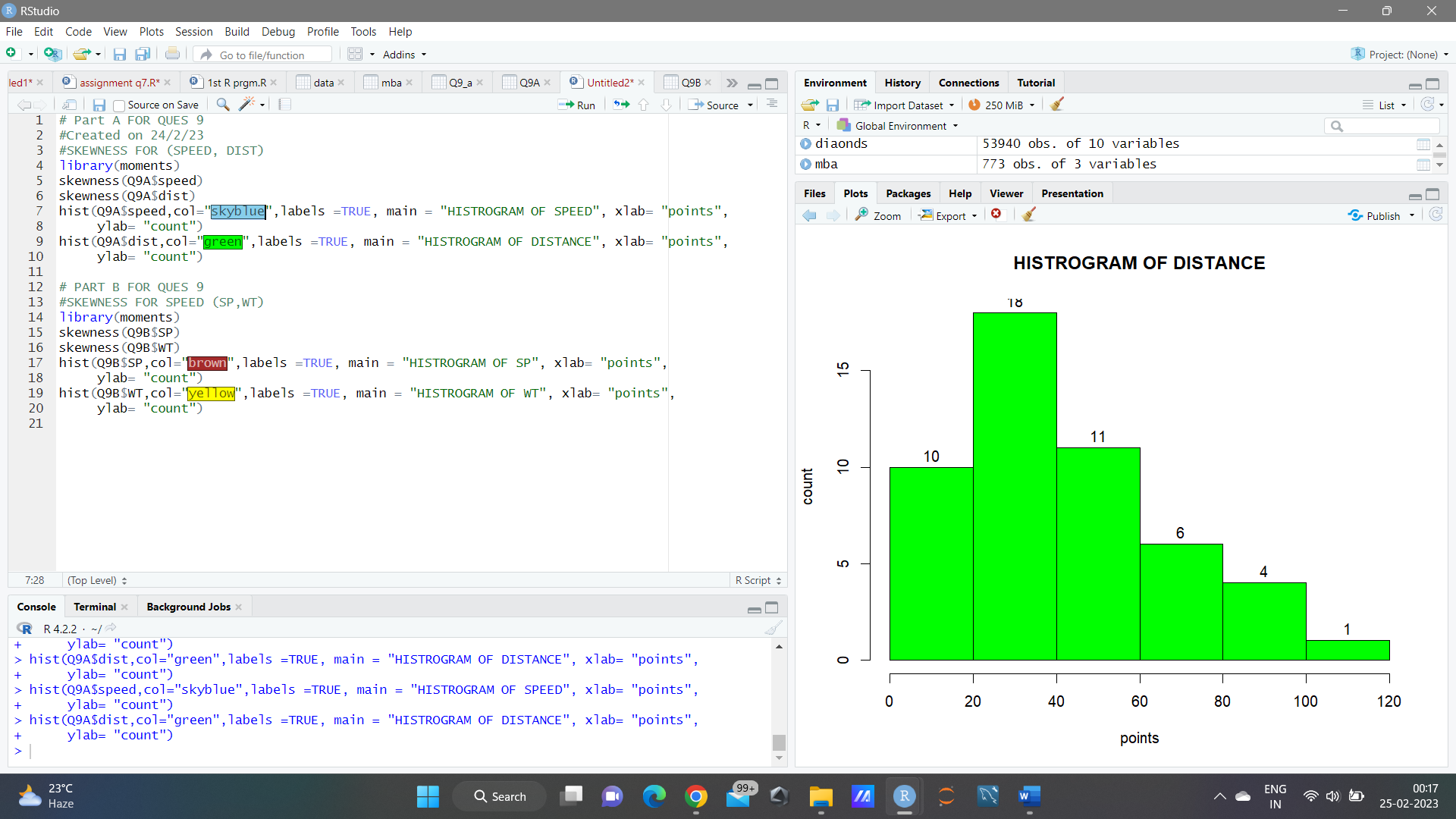
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans : part 1**

Coding is done in R programming.



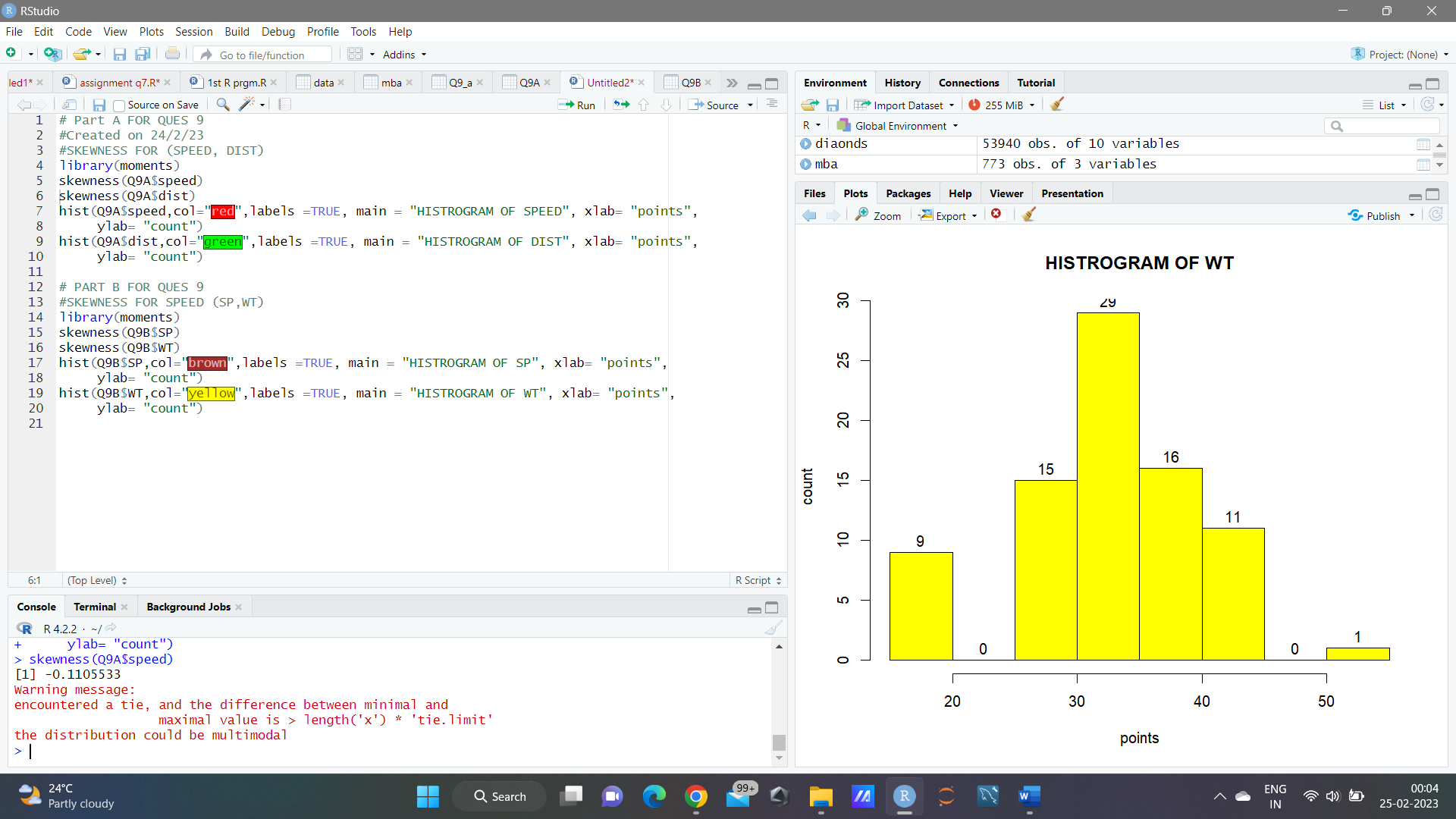
-kurtosis(Q9A$speed)

-kurtosis(Q9A$dist)

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS ; PART B**

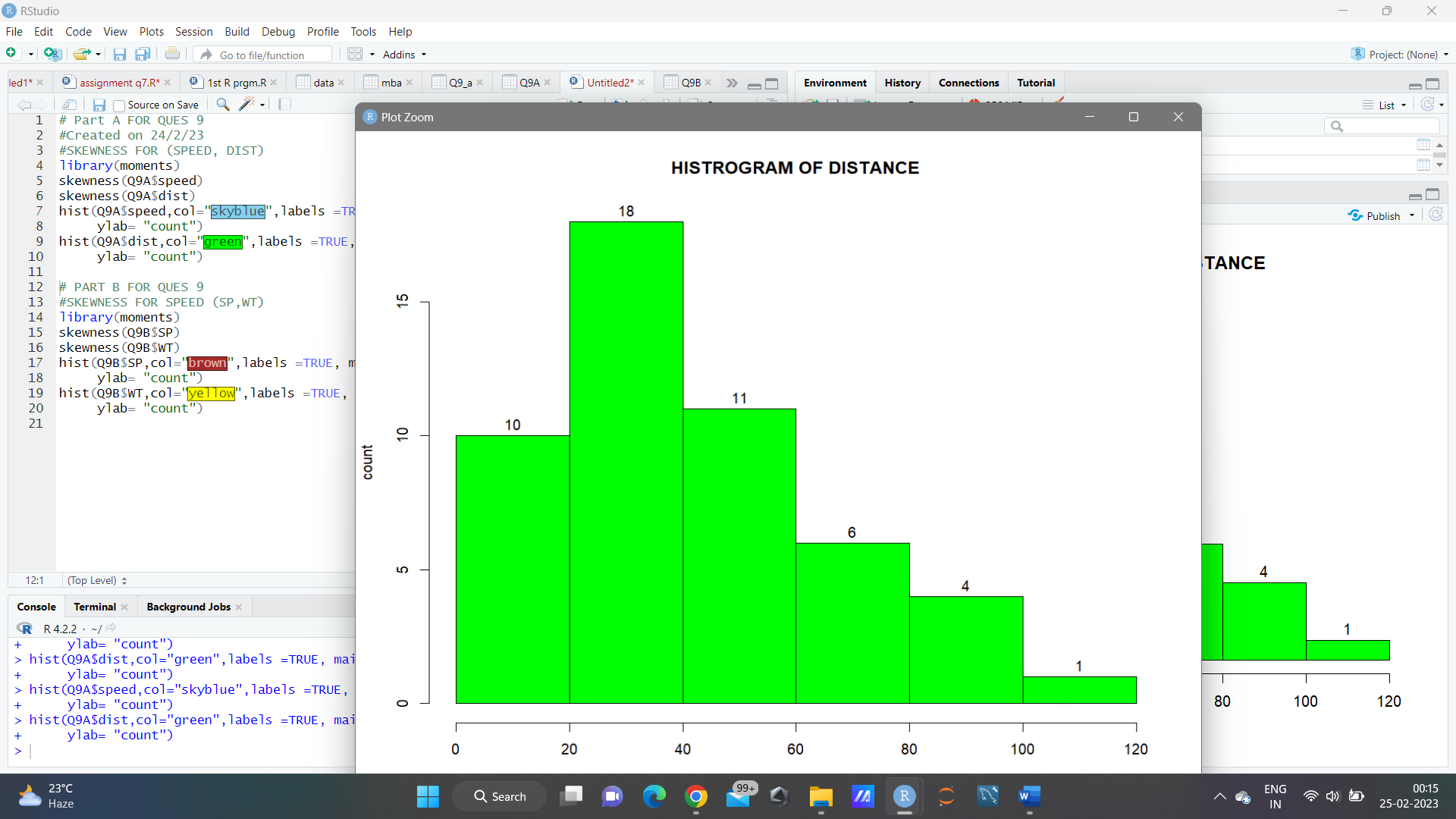
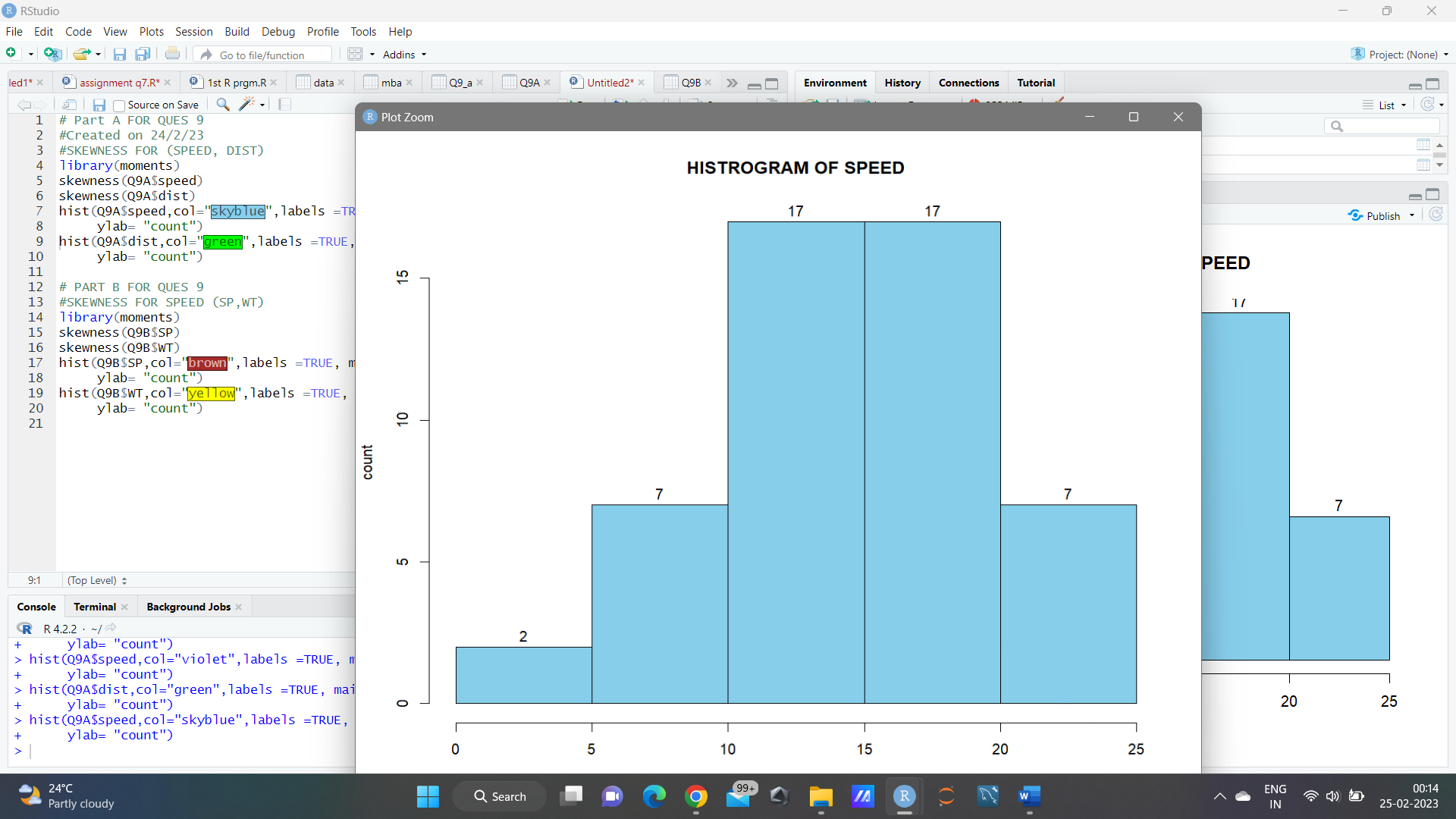


kurtosis(Q9A$speed)

kurtosis(Q9A$dist)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **SPEED** | **DIST** | **ST** | **WT** |
| SKEWNESS | **-**0.110 | 0.75 | 1.55 | -0.59 |
| KURTOSIS | 2.422 | 3.422 | NA | 3.81 |

Inferences Histrogram for car speed and distance ;



**Q10) Draw inferences about the following boxplot & histogram**





Ans :

From the above boxplot and box plot , it show that the distribution has outliers at the end (means in the histogram tail side and in box plot at in upper extreme). the outliers on the maximum side are the upper side. The distribution is positively skewed or right 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?

Ans =

Here total no of sample mean (n) = 2000

The average weight of person in sample (X̅) = 200

Standard deviation of sample () = 30

Confidence Interval =

For 94% of CI value Z score = 1.89

Confidence interval for 94% = 200 ± (1.89 x (30/))

=198.73 to 201.27

For 98% of CI value Z score = 2.33

Confidence interval for 98% = 200 ± (2.33 x (30/))

=198.43 to 201.56

For 96% of CI value Z score = 2.06

Confidence interval for 96% = 200 ± (2.06 x (30/))

=198.62 to 201.38

**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=

import pandas as pd

import statistics as sts

st = [34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56]

sts.mean(st)

sts.median(st)

round(sts.variance(st),4)

round(sts.stdev(st),4)

Mean = 41

Median = 40.5

Variance = 25.5294

Standard Devi ation = 5.0527

1. What can we say about the student marks?

Ans= The student score 41 mark most of the time. He scores an average 41 mark.

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

Ans= There will be no skewness, it will be symmetric.

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

Ans= Positive Skewness.

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

Ans= Negative Skewness.

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

Ans= Positive kurtosis means the curve is more peaked.

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

Ans= Negative Kurtosis means the curve will be flatter.

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



What can we say about the distribution of the data?

Ans= from the above boxplot it is not normally distributed the median is towards the higher value which is on the right side

What is nature of skewness of the data?

Ans=: The data is on left side, whisker range of minimum value is more than the maximum, therefore it is negative skewness.

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

Ans= As we know IQR = Q3 upper quartile – Q1 lower quartile = 18-18= “8”.  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans= From the above boxplots both are having zero errors and both the boxplots median is also in the same range and they are normally distributed.

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

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Ans=

In python

import pandas as pd

cars = pd.read\_csv('D:\\Study\\Assignments\\Cars.csv')

cars.head()

cars[“MPG”].mean()= 34.422

cars[“MPG”].std()= 9.131

A= P(MPG>38)

from scipy import stats

1-stats.norm.cdf(38,34.42,9.13)

= 0.3474

B= P(MPG<40)

stats.norm.cdf(40,34.42,9.13)

=0.7294

C= P(20<MPG<50)

stats.norm.cdf(50,34.42,9.13)-stats.norm.cdf(20,34.42,9.13)

=0.8989

Q 21) Check whether the data follows normal distribution

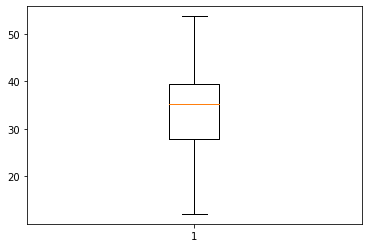
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

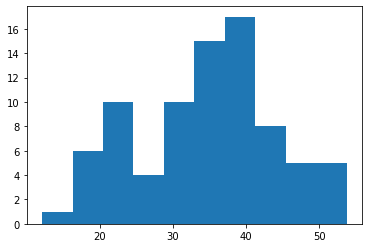
Ans= in python

import matplotlib.pyplot as plt

plt.boxplot(cars['MPG'])



- plt.hist(cars['MPG'])



From this above box plot and histogram we can say the MPG of Cars follows normal distribution

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

Dataset: wc-at.csv

Ans= In python

import pandas as pd

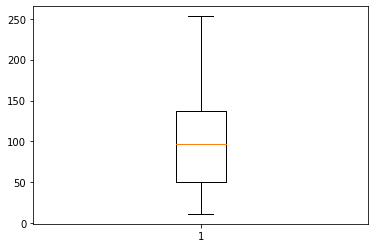
import matplotlib.pyplot as plt

wc\_at = pd.read\_csv('D:wc-at.csv')

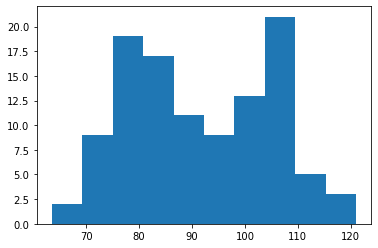
plt.hist(wc\_at["AT"])



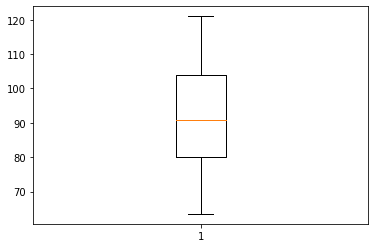
plt.boxplot(wc\_at["AT"])



plt.hist(wc\_at["Waist"])



plt.boxplot(wc\_at["Waist"])



From the above histogram and box plot for both AT & Waist of wc-at data set, it shows that both AT & Waist follows normal distribution.

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

Ans= Z scores of 90% confidence interval is 1.64

Z scores of 94% confidence interval is 1.88

Z scores of 60% confidence interval is 0.84.

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

Ans = t scores of 95% confidence interval is 2.06

t scores of 96% confidence interval is 2.17

t scores of 99% confidence interval is 2.79

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= Population mean,µ = 270

Sample size, n = 18

Sample mean, x̅ = 260

Standard deviation, s =90

t score = (x̅- µ)/(s/sqrt(n))

=(260-270)/(90/sqrt(18))

= -10/21.23

= -0.47

df= degrees of freedom = n-1 = 18-1= 17

Probability:

pt(tscore,df)

pt(-0.47,17)

Answer is = 0.3221639