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 | Continues |
| Weight of Gold | Continues |
| Distance between two places | Continues |
| Length of a leaf | Continues |
| Dog's weight | Continues |
| 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 | Ratio |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ratio |
| Height | Interval |
| Type of living accommodation | Ordinal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Nominal |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| 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?

P = {HHH,HHT,HTH,THH,TTH,THT,HTT,TTT}

P(Two heads and one tail) = Total no of outcomes = 3/8 = 0.375

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

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

Number of probability = {(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

1. In first problem as we can see equal to 1 is not possible just because of outcomes comes in format that is eg. (1,1), (1,2) etc.

So probability of getting equal to 1 is zero.

1. In second problem there is probability of getting less then 4 or equal to 4 is P = 6/36 = 1/6 = 0.16
2. Probability of getting sum is divisible by 2 and 3 P = (1,1),(1,2),(1,3),(1,5),(2,1),(2,2),(2,4),(2,6)(3,1),(3,3),(3,5),(3,6),(4,2),(4,4),(4,5),(4,6)

(5,1),(5,3),(5,4),(5,5),(6,2),(6,3),(6,4),(6,6) = 24/36 = 4/6 = 0.66

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?

Ans :

total no of balls=7 balls

n(S)= 2 balls drawn out of 7

7C2=21

n(S1)= 2 balls drawn non of is blue

5C2=10

P(S1)=7C2/5C2

=21/10

P(S1)=2.1

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

E (Expected number of candies for a randomly selected child )

P(E) = 1 \* 0.015 + 4 \* 0.20 + 3 \* 0.65 + 5 \* 0.005 + 6 \* 0.01 + 2 \* 0.120

P(E) = 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

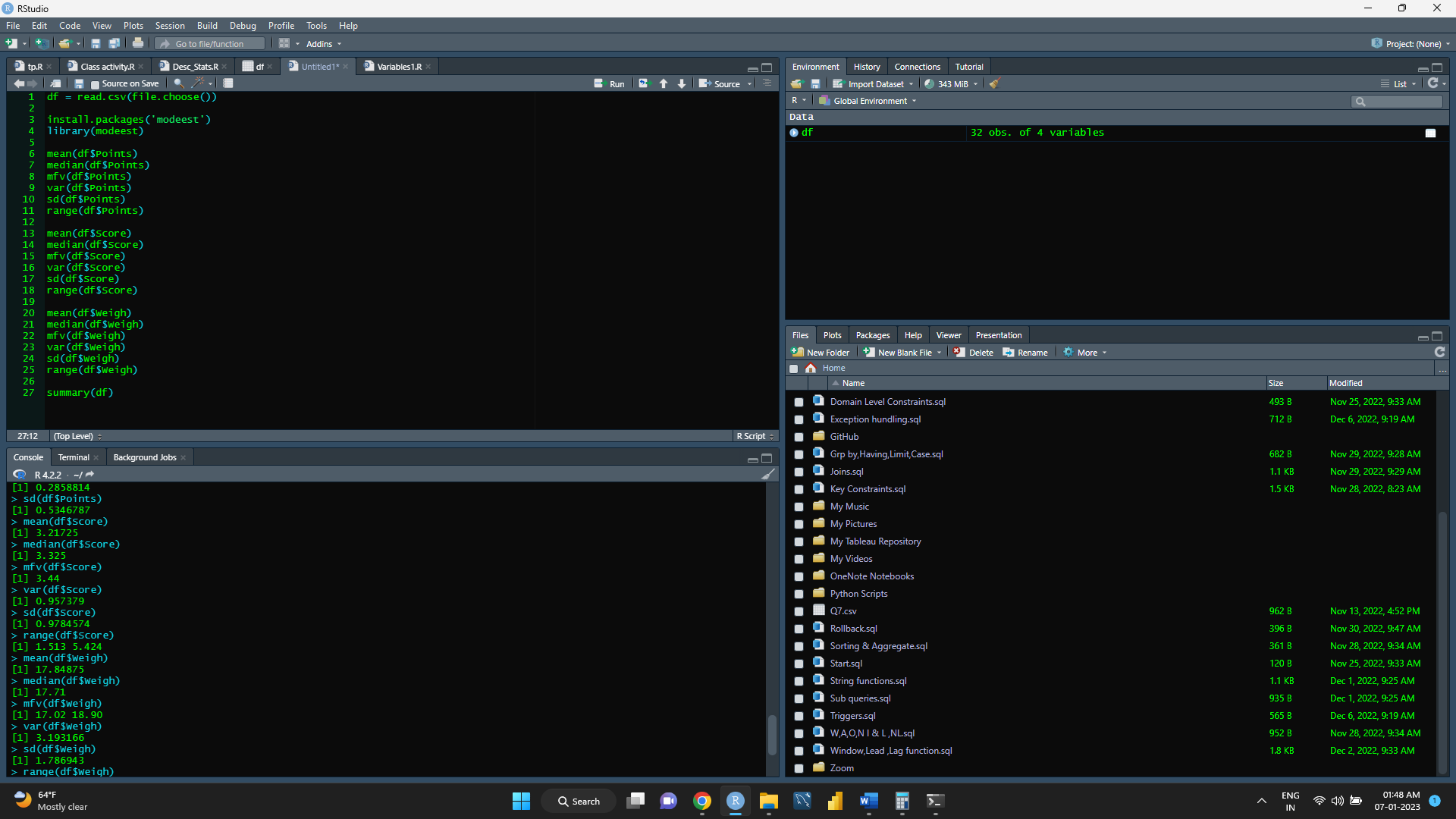
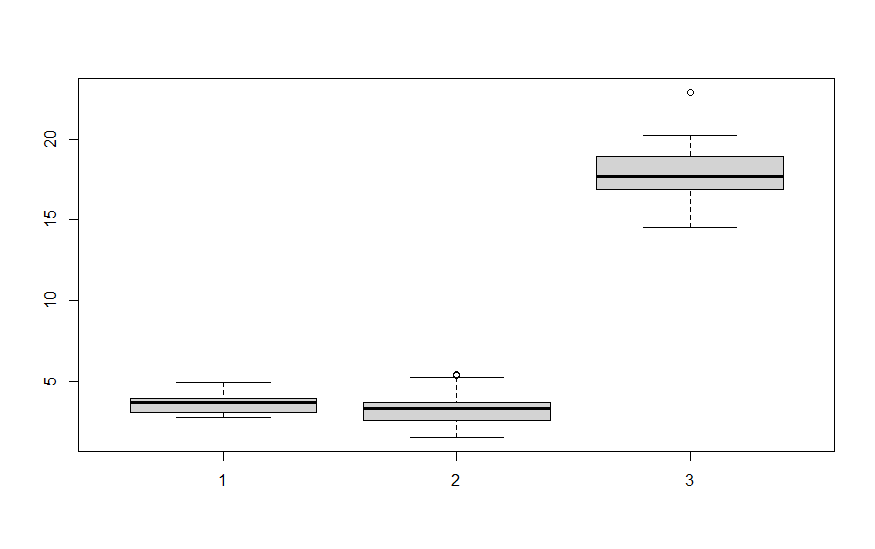
P(E) = 3.09

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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596 | 3.21 | 17.848 |
| Median | 3.695 | 3.32 | 17.71 |
| Mode | 3.07, 3.92 | 3.44 | 17.02, 18.90 |
| Variance | 0.285 | 0.957 | 3.193 |
| Standard deviation | 0.534 | 0.978 | 1.786 |
| Range | 2.76 – 4.93 | 1.513 – 5.424 | 14.5 – 22.9 |

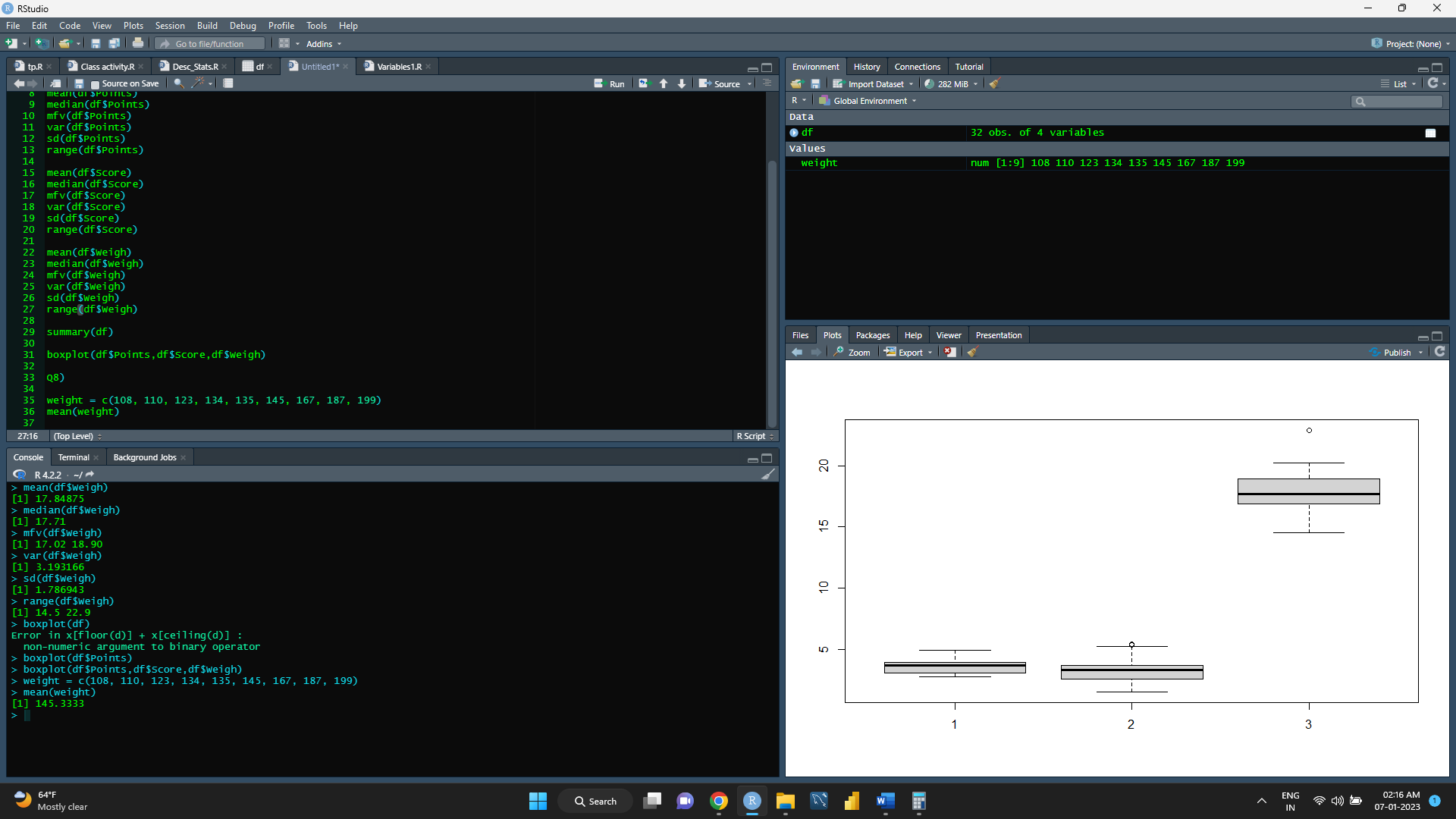
****

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, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

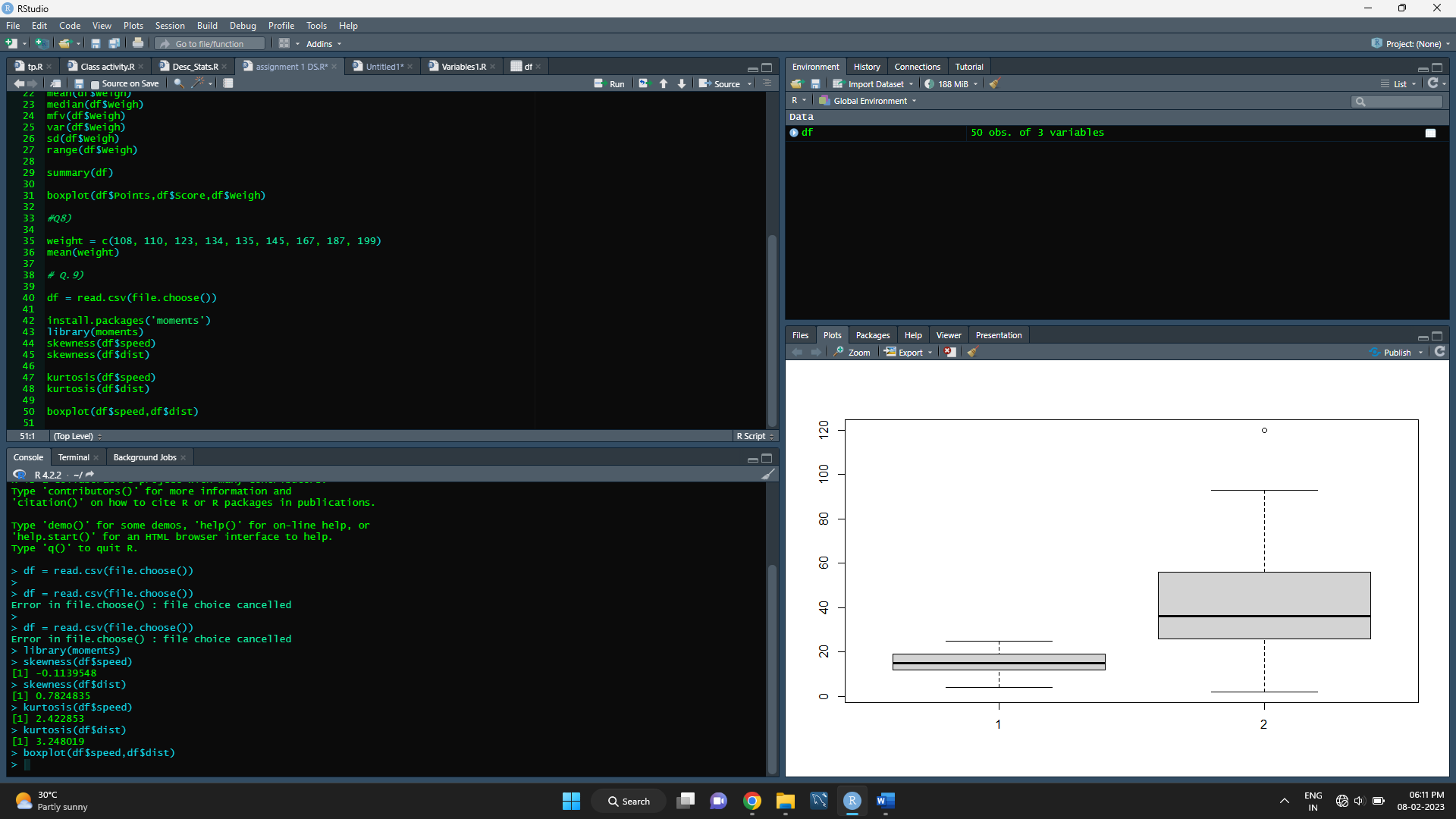
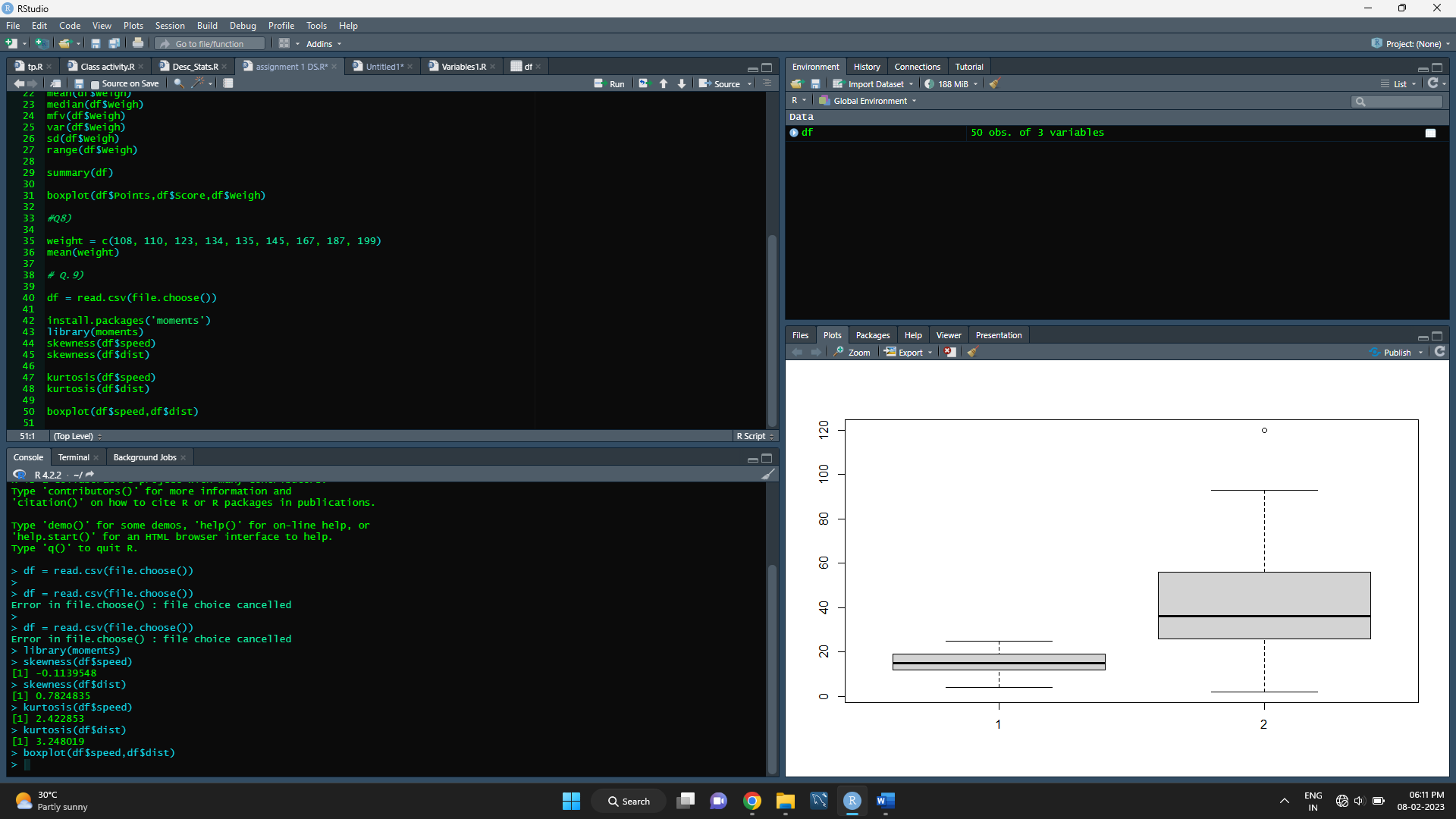
Ans: 

Output : 145.333

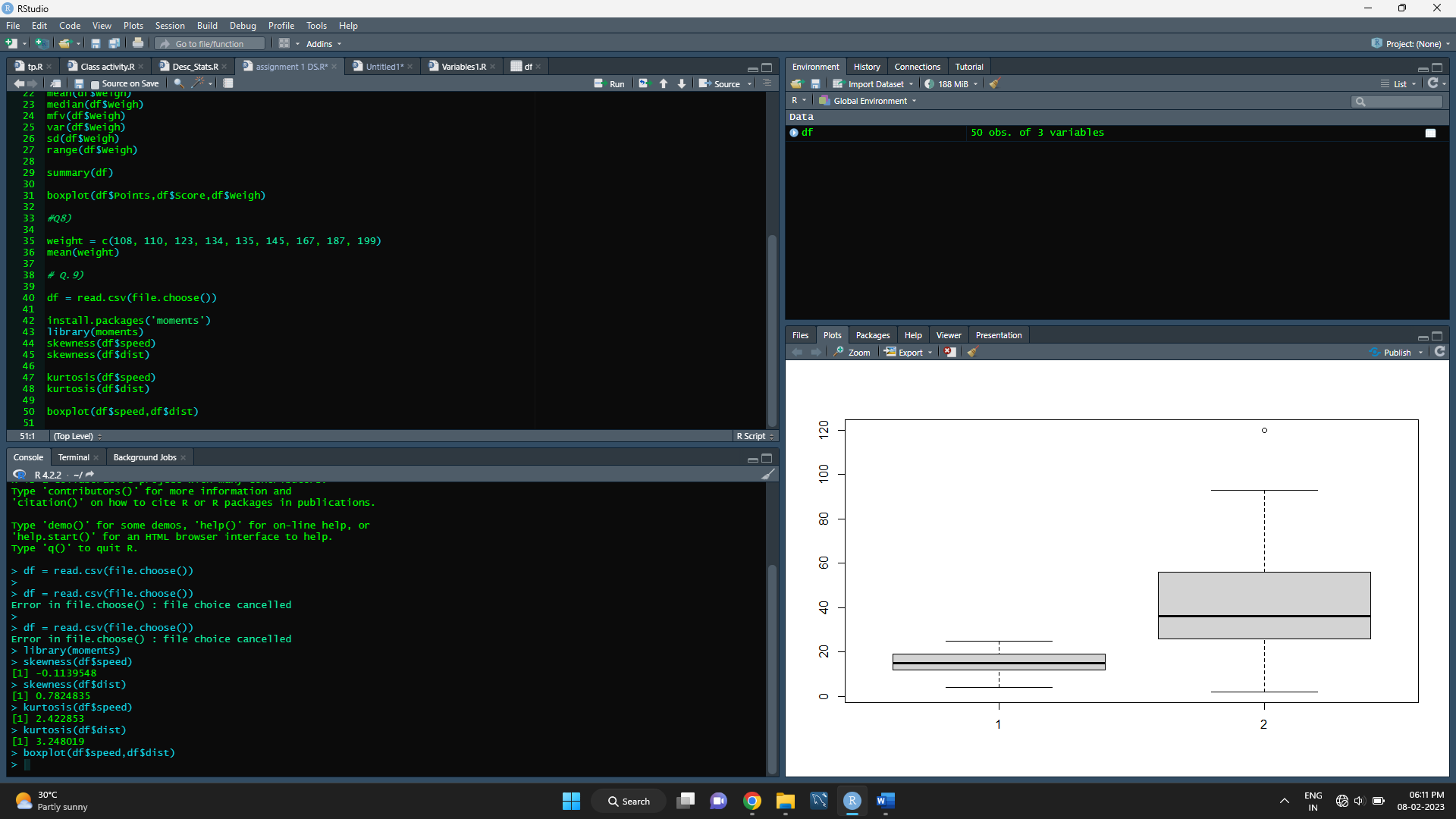
Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

Cars speed and distance

Ans :



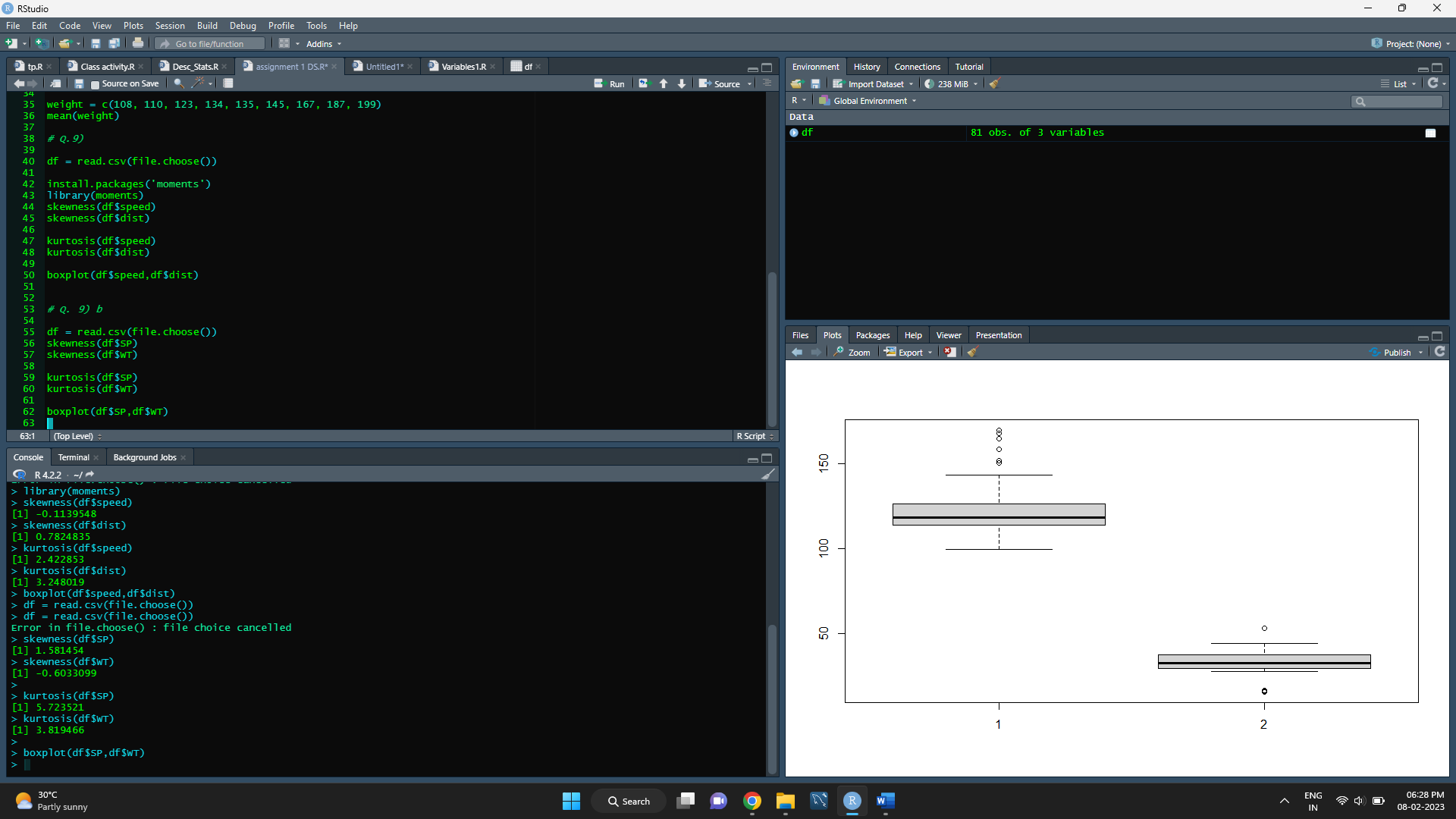
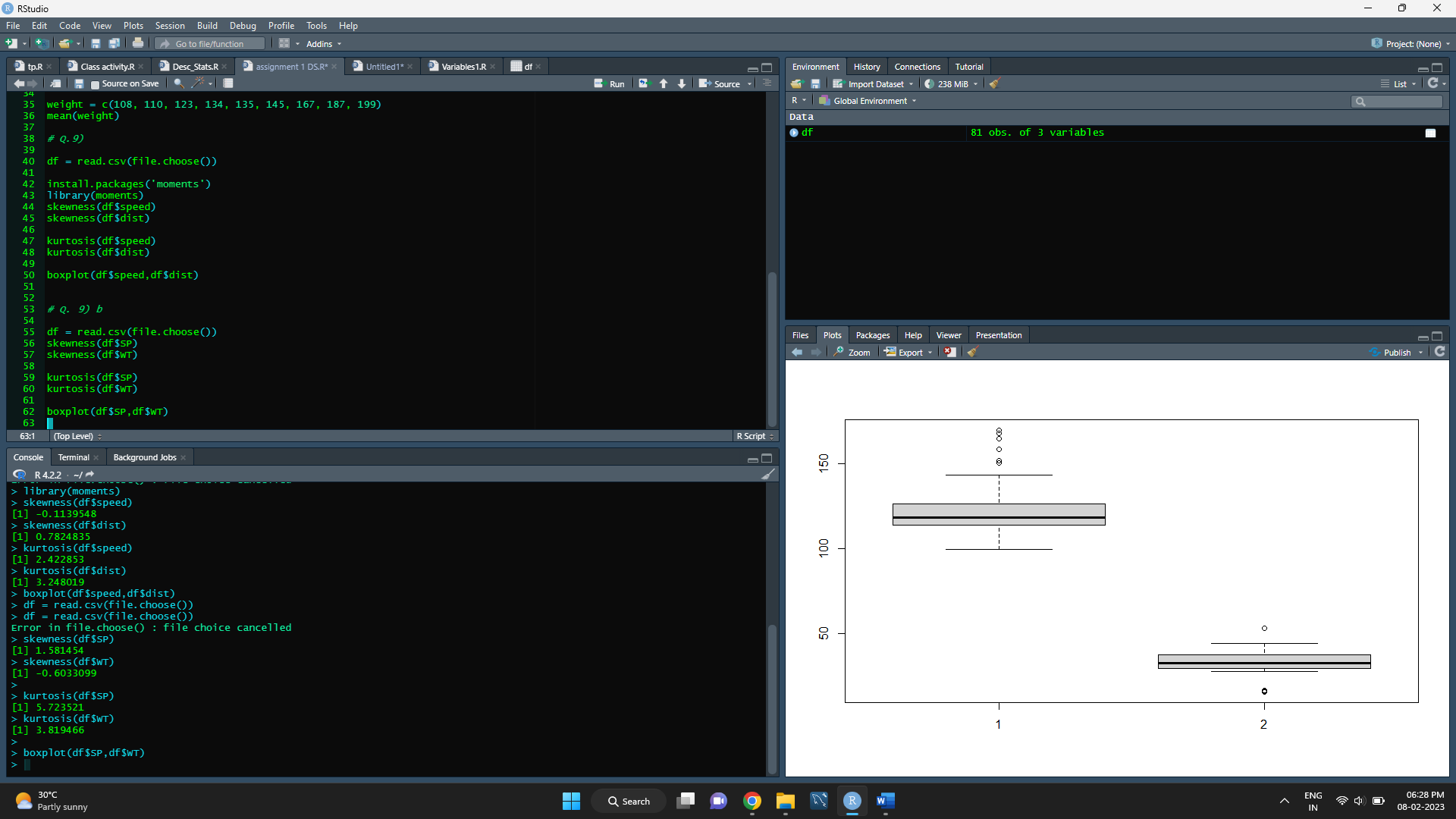
|  |  |  |
| --- | --- | --- |
|  | Speed | Distance |
| Skewness | -0.1139548 | 0.7824835 |
| Kurtosis | 2.422853 | 3.248019 |

****

SP and Weight(WT)

Use Q9\_b.csv

Ans :

****

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.581454 | -0.6033099 |
| Kurtosis | 5.723521 | 3.819466 |

Q10) Draw inferences about the following boxplot & histogram



Ans : we can in histogram it is not normally distributed as well as we can see skewness in this plot, that is right skewed it is also called as positively skewed.



Ans : in this boxplot we can conclude there us a outliers in dataset .

**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 : = stats .t. interval (alpha= 0.94, df= 1999, loc=200, scale=30/np . sqrt (2000))

for 94% confidence interval range is [198.73 – 201.26]

for 98% confidence interval range is [198.43 – 201.56]

For 96% confidence interval range is [198.62 – 201.37]

**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 = 34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18 = 41

Median = 40+41/2 = 40.5

Variance = (34 – 41)^2+(36 – 41)^2+(36 – 41)^2+(38 – 41)^2+(38 – 41)^2+(39 – 41)^2+(39 – 41)^2+(40 – 41)^2+(40 – 41)^2+(41 – 41)^2+(41 – 41)^2+(41 – 41)^2+(41 – 41)^2+(42 – 41)^2+(42 – 41)^2+(45 – 41)^2+(49 – 41)^2+(56 – 41)^2/18-1 = 25.5294118

Standard deviation = ✓25.5294118 = 5.05266383

1. What can we say about the student marks?

Ans: - normal distribution

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

Ans: - zero skewness no skewness it is a normal distribution

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

Ans: - positive skewed

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

Ans: -negative skewed

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

Ans: -positive kurtosis telling us about peaked ness of distribution

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

Ans: - negative kurtosis is telling us about how flatten peak of your distribution

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



What can we say about the distribution of the data?

Ans: - this is a box plot and the main purpose of boxplot is detection of outliers so 1st interpretation about data is data didn’t have outlier. also, data telling us about approximate values of 25th, 50th, 75th percentile and upper extreme, lower extreme. There is a skewness in a data

What is nature of skewness of the data?

Ans: -the nature of skewness in this boxplot is left skewed. (Negative skewed)

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

Ans: - 1st qtrl 25th = 10 or maybe 10.2

2nd qtrl 50th = 15 or maybe 15.2

3rd qtrl 75th = 18 or maybe 18.2

Q19) Comment on the below Boxplot visualizations?



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

Ans : In this boxplot we can conclude about there range in first boxplot there is very low range and second boxplot there is very high range as well as both boxplot are normally distributed also there is no outliers in datasets .

May be there is a kurtosis between both boxplot.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

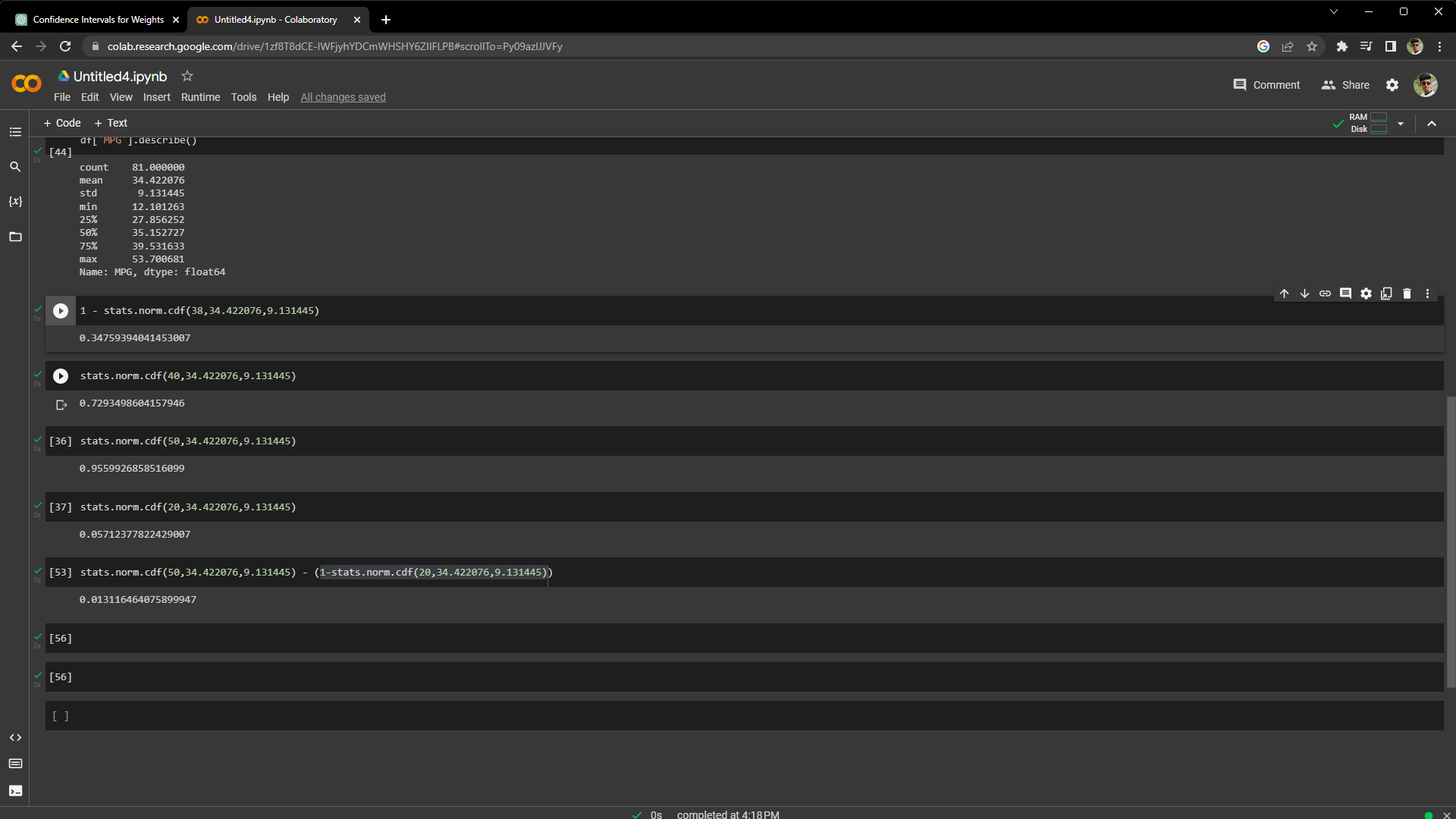
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

Ans :

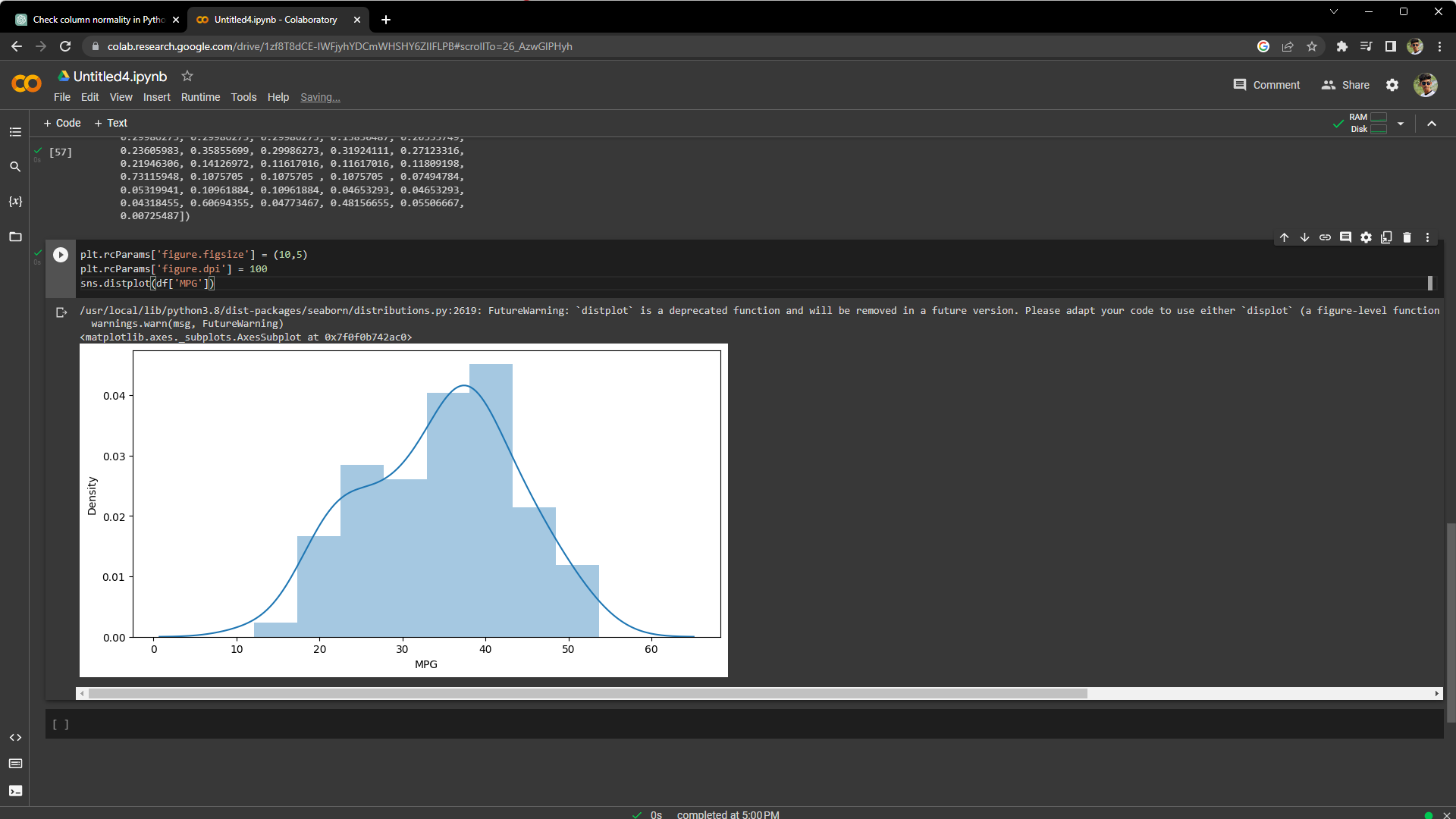


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans :

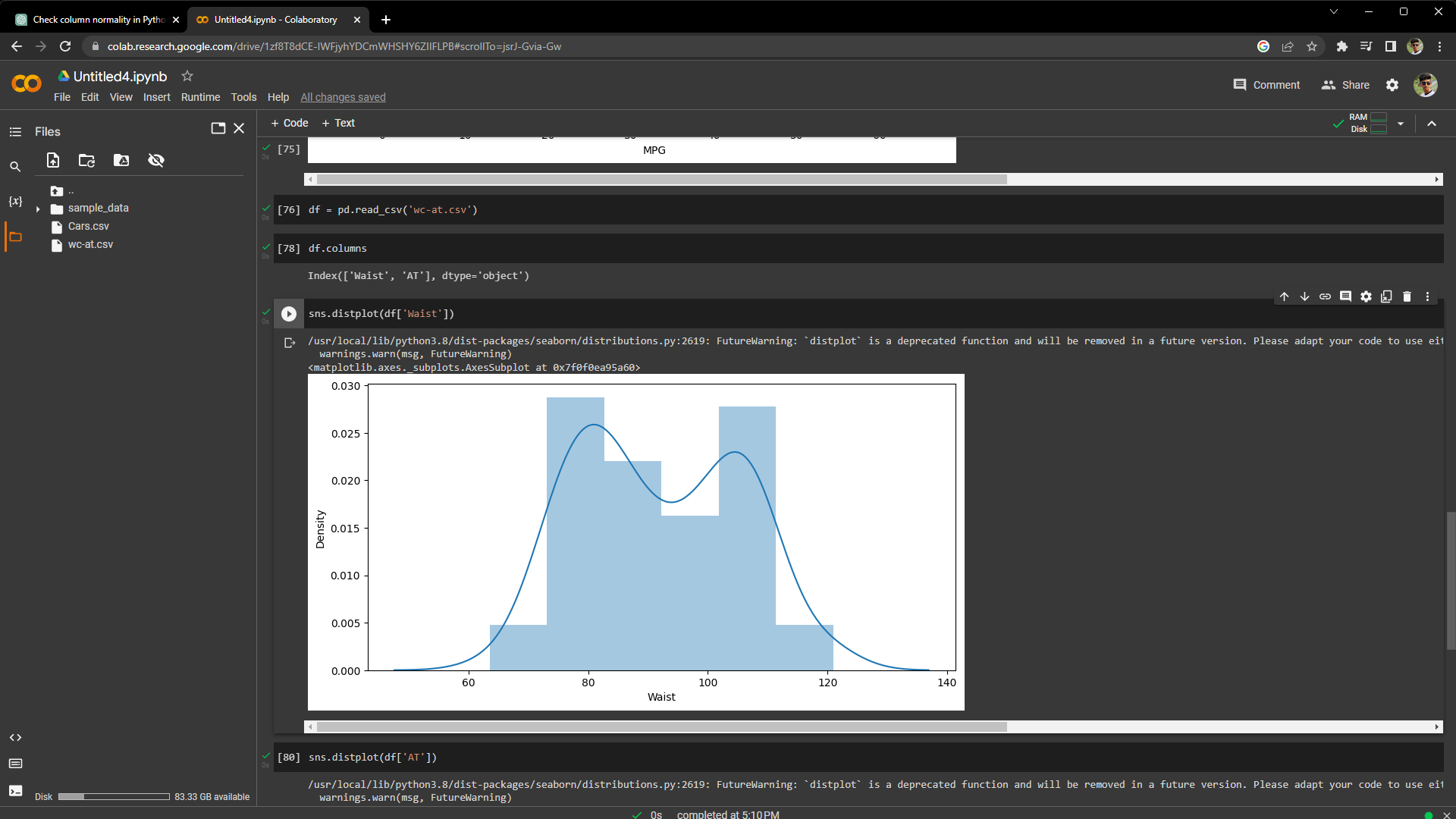


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

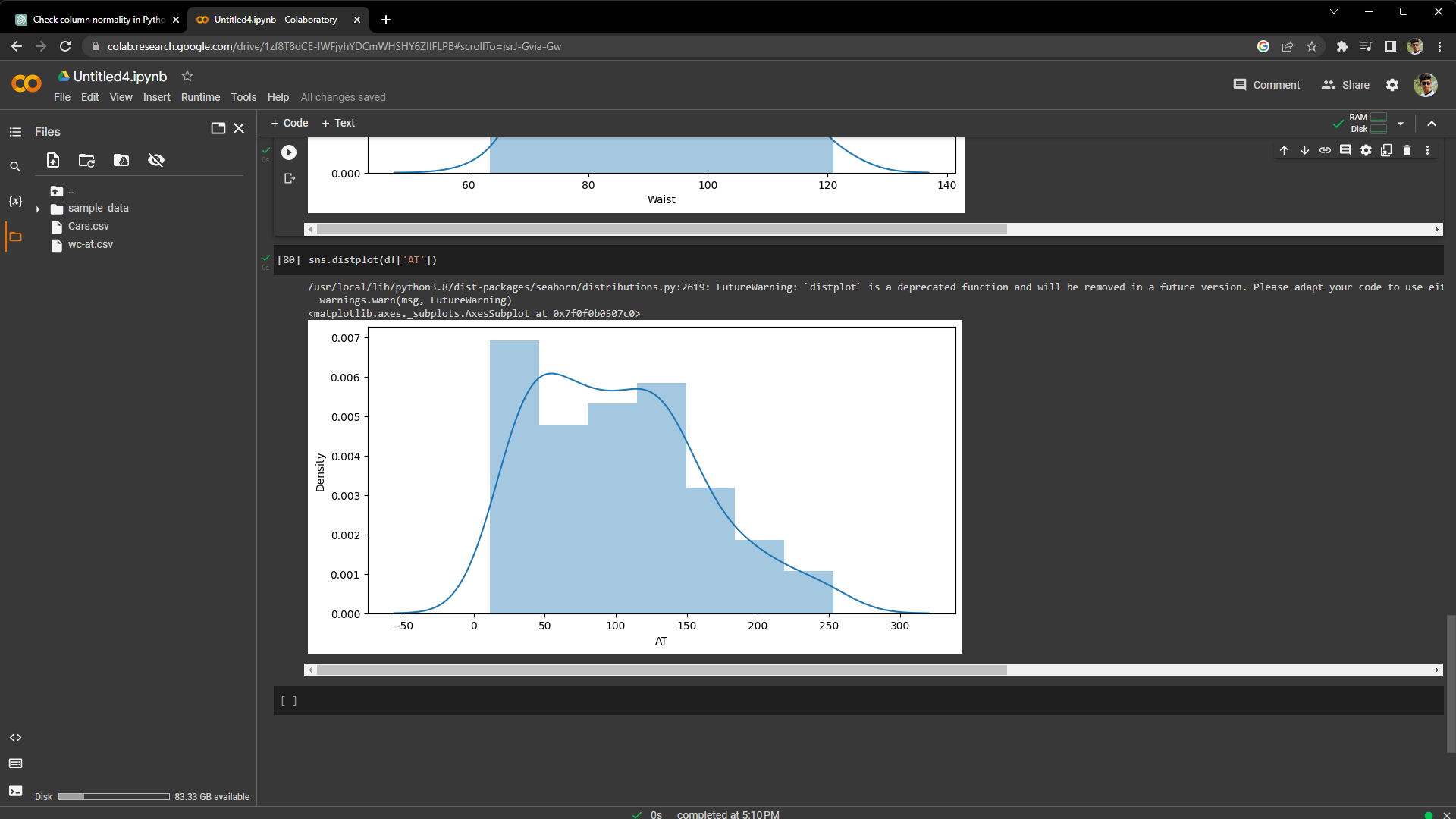
Dataset: wc-at.csv

Ans : In first column of data Adipose Tissue(AT) as we can see it is normally distributed. And in second column of data wc–at there is slightly skewed data.(Positive Skewness/Right Skewness)

1. Adipose Tissue(AT) : -

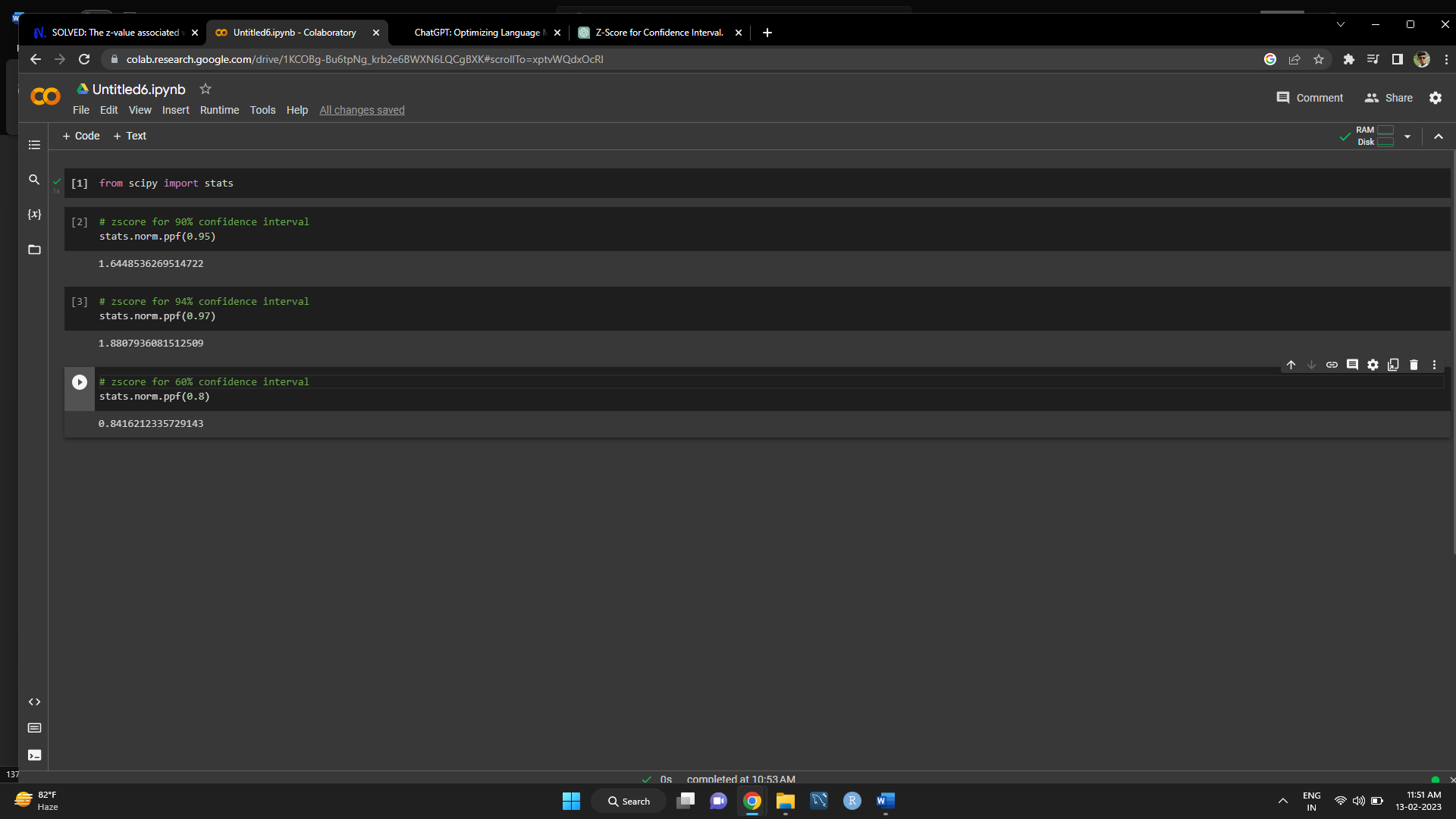


1. wc-at: -



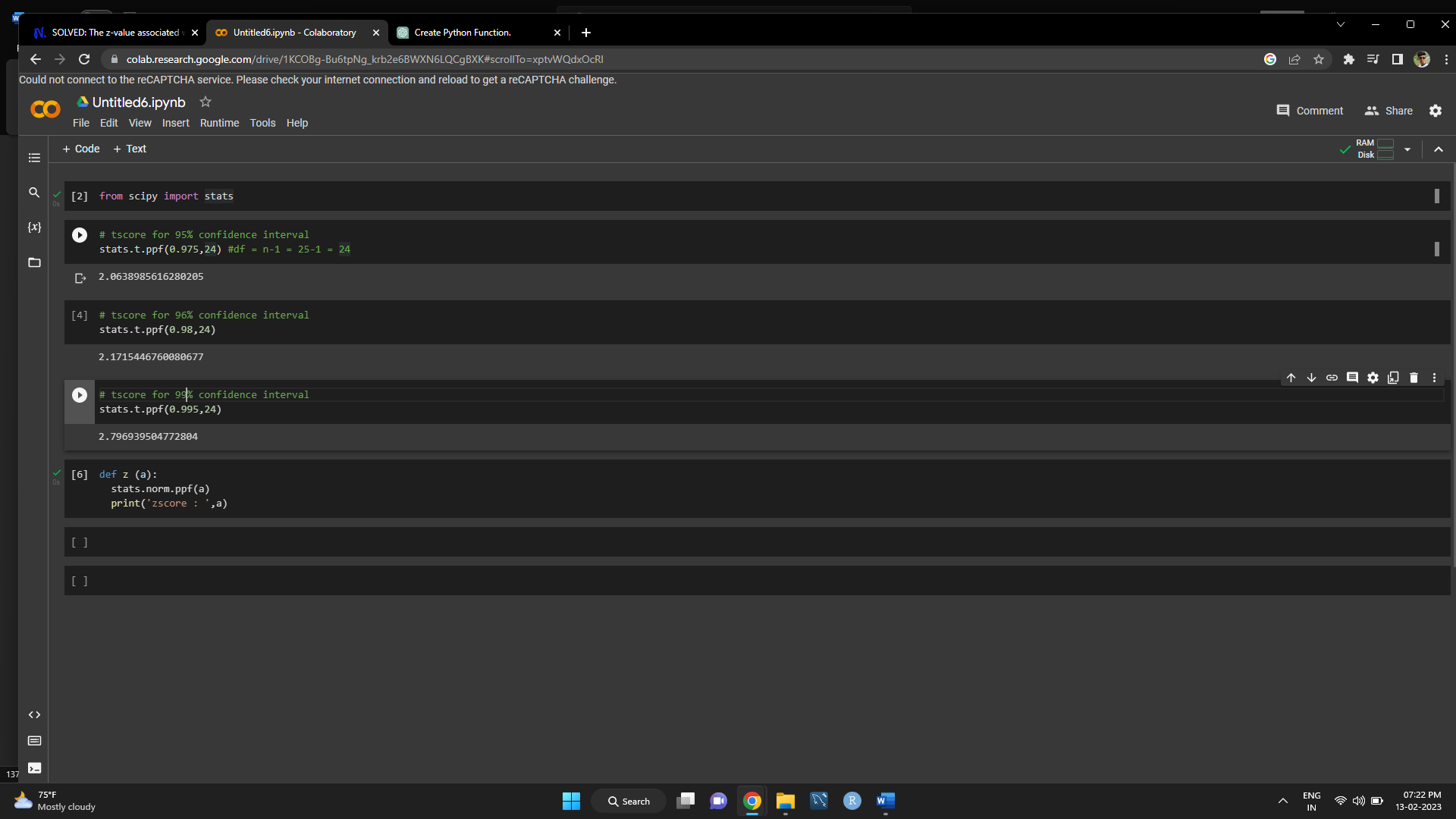
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Ans :



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

Ans :



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 :

t - statistics for the data is given as follows:

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

t =

t = - 0.471

from scipy import stats

stats.t.cdf(- 0.471,17) --> 0.3218140331685075

probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.3218