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

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	Ordinal
Level of Agreement	Ordinal
IQ(Intelligence Scale)	Ratio
Sales Figures	Ratio
Blood Group	Nominal
Time Of Day	Ordinal
Time on a Clock with Hands	Interval
Number of Children	Ordinal
Religious Preference	Nominal
Barometer Pressure	Ratio
SAT Scores	Ratio
Years of Education	Interval

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

Ans) We have different possibilities of trails

H H H

H H T

H T H

T H H

T T T

T H H

T H T

T T H

So, there are total 8 possibilities, in which we have 3 possibilities of two heads and one tail.

Probability = Total of of two heads and one tail.

Total number of possibilities

= 3/8 = 0.375

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

Possibilities

(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)

a) Equal to 1

Ans) 0/36= 0

b) Less than or equal to 4

Ans) possibilities (1,1), (1,2), (1,3), (2,1), (2,2), (3,1) = 6/36 = 1/6

c) Sum is divisible by 2 and 3

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

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 balls = 7, Drawing 2 balls from 7 =
nCr = n!/r!(n-r)!.=7C2= 7*6/2*1 = 42/2 = 21

Total red and green balls are = 5

2 balls needs to drawn from 5 balls(Red and Green)/ = 5C2 = 5*4/2*1 = 20/2 = 10

Probability = 2 balls needs to drawn from 5 balls(Red and Green)/ 2 balls drawn from Total balls

Probability = 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 = E (X) = ∑ x P (x).

Probability of Expected number of candies for a randomly selected child

= each individual Candies count *Probability of each individual =

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

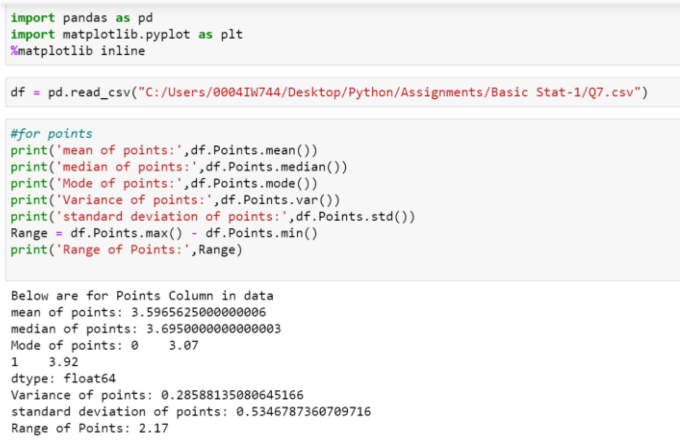
= 3.09

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

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

Use Q7.csv file

Solution:-

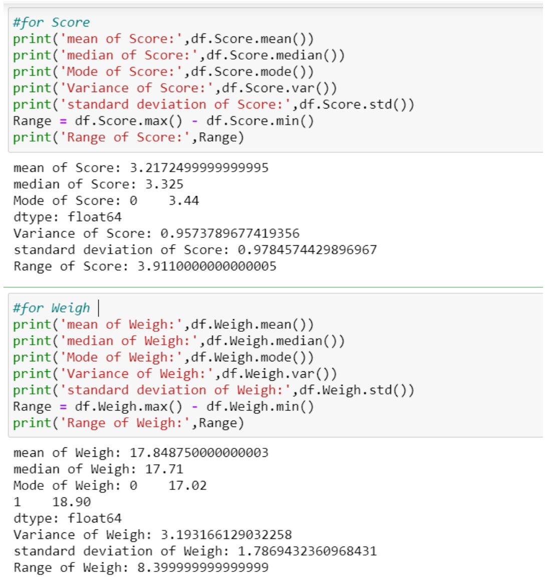


```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

```
df = pd.read_csv("C:/Users/0004IW744/Desktop/Python/Assignments/Basic Stat-1/Q7.csv")
```

```
#for points
print('mean of points:',df.Points.mean())
print('median of points:',df.Points.median())
print('Mode of points:',df.Points.mode())
print('Variance of points:',df.Points.var())
print('standard deviation of points:',df.Points.std())
Range = df.Points.max() - df.Points.min()
print('Range of Points:',Range)
```

```
Below are for Points Column in data
mean of points: 3.5965625000000006
median of points: 3.6950000000000003
Mode of points: 0    3.07
1    3.92
dtype: float64
Variance of points: 0.28588135080645166
standard deviation of points: 0.5346787360709716
Range of Points: 2.17
```

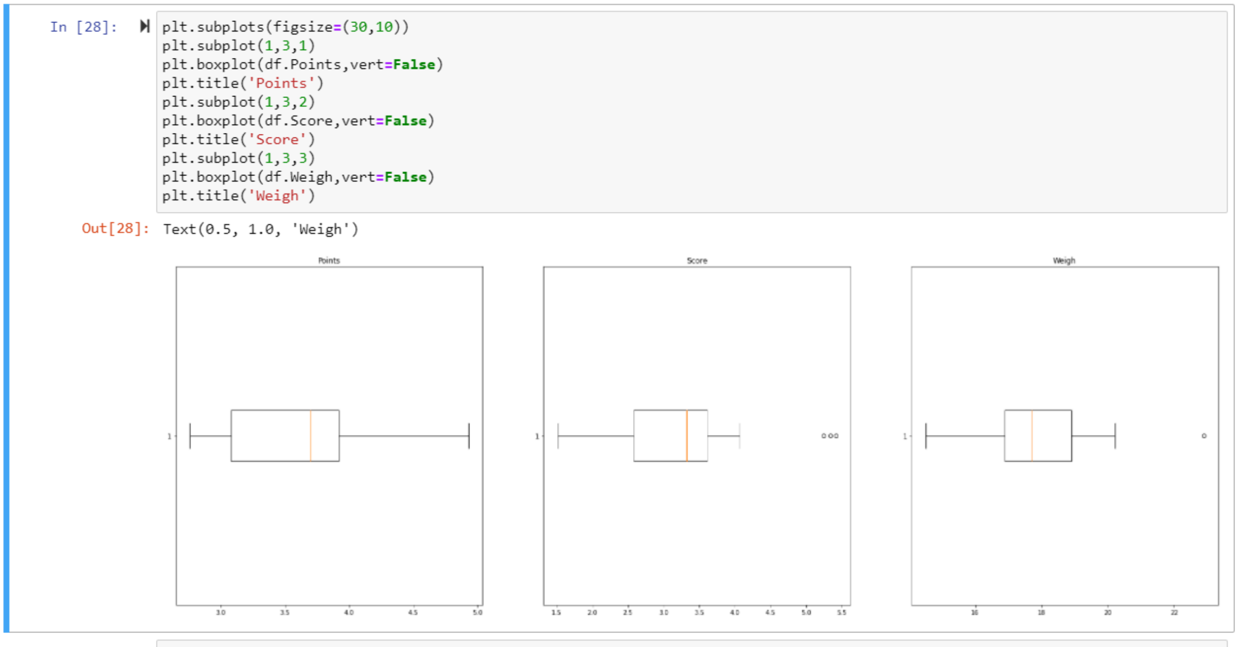


```
#for Score
print('mean of Score:',df.Score.mean())
print('median of Score:',df.Score.median())
print('Mode of Score:',df.Score.mode())
print('Variance of Score:',df.Score.var())
print('standard deviation of Score:',df.Score.std())
Range = df.Score.max() - df.Score.min()
print('Range of Score:',Range)
```

```
mean of Score: 3.2172499999999995
median of Score: 3.325
Mode of Score: 0    3.44
dtype: float64
Variance of Score: 0.9573789677419356
standard deviation of Score: 0.9784574429896967
Range of Score: 3.9110000000000005
```

```
#for Weigh
print('mean of Weigh:',df.Weigh.mean())
print('median of Weigh:',df.Weigh.median())
print('Mode of Weigh:',df.Weigh.mode())
print('Variance of Weigh:',df.Weigh.var())
print('standard deviation of Weigh:',df.Weigh.std())
Range = df.Weigh.max() - df.Weigh.min()
print('Range of Weigh:',Range)
```

```
mean of Weigh: 17.848750000000003
median of Weigh: 17.71
Mode of Weigh: 0    17.02
1    18.90
dtype: float64
Variance of Weigh: 3.193166129032258
standard deviation of Weigh: 1.7869432360968431
Range of Weigh: 8.399999999999999
```



```
In [28]:  plt.subplots(figsize=(30,10))
          plt.subplot(1,3,1)
          plt.boxplot(df.Points,vert=False)
          plt.title('Points')
          plt.subplot(1,3,2)
          plt.boxplot(df.Score,vert=False)
          plt.title('Score')
          plt.subplot(1,3,3)
          plt.boxplot(df.Weigh,vert=False)
          plt.title('Weigh')

Out[28]:  Text(0.5, 1.0, 'Weigh')
```

Q8) Calculate Expected Value for the problem below

a) The weights (X) of patients at a 0063linic (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:-

Expected value = Summation of the product of probability choosing each person individually and their weights

Probability of each patients choosing individually is = 1/9

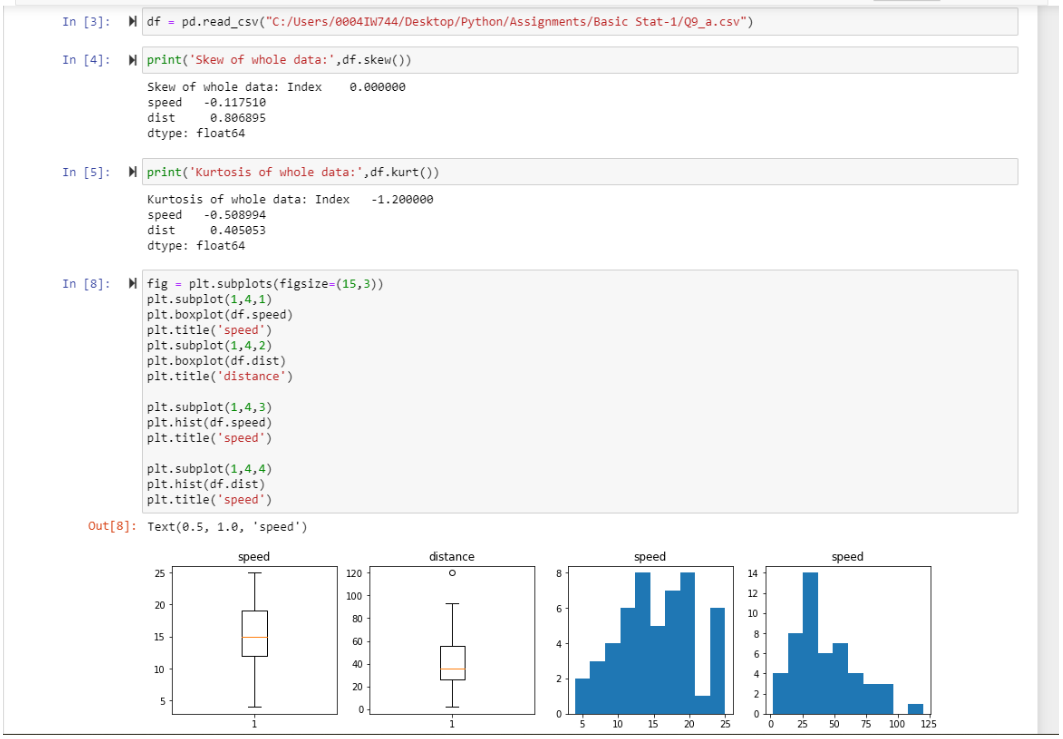
Expected Value = (108+110+ 123+134+ 135+ 145+ 167+ 187+ 199)*1/9

=145.33

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

Cars speed and distance

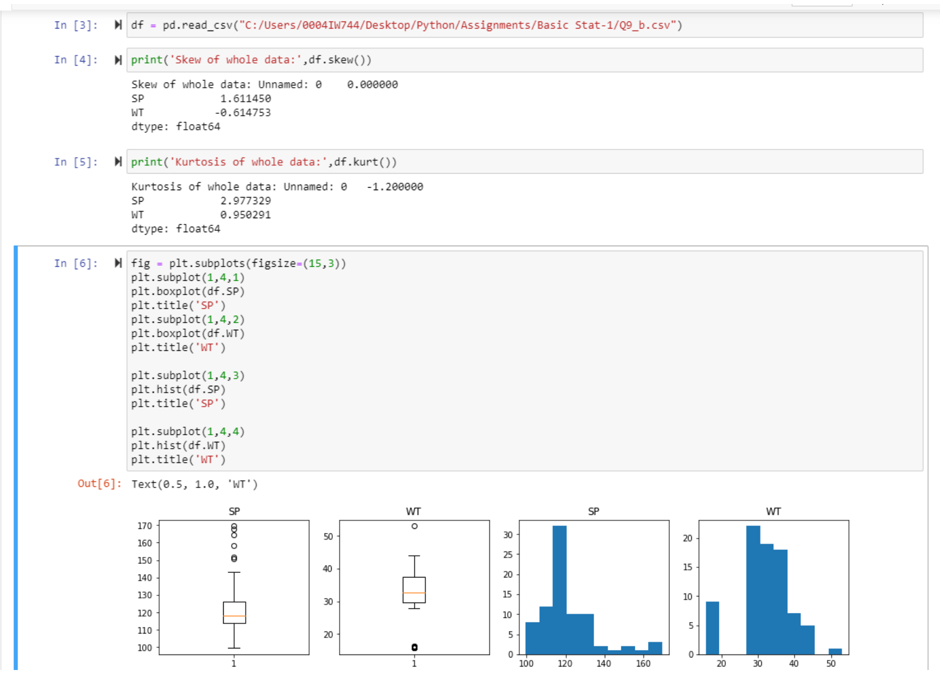
Use Q9_a.csv



```
In [3]:  df = pd.read_csv("C:/Users/0004IW744/Desktop/Python/Assignments/Basic Stat-1/Q9_a.csv")

In [4]:  print('Skew of whole data:',df.skew())

         Skew of whole data: Index    0.000000
         speed   -0.117510
         dist     0.806895
         dtype: float64

In [5]:  print('Kurtosis of whole data:',df.kurt())

         Kurtosis of whole data: Index   -1.200000
         speed   -0.508994
         dist     0.405053
         dtype: float64

In [8]:  fig = plt.subplots(figsize=(15,3))
         plt.subplot(1,4,1)
         plt.boxplot(df.speed)
         plt.title('speed')
         plt.subplot(1,4,2)
         plt.boxplot(df.dist)
         plt.title('distance')

         plt.subplot(1,4,3)
         plt.hist(df.speed)
         plt.title('speed')

         plt.subplot(1,4,4)
         plt.hist(df.dist)
         plt.title('speed')

Out[8]:  Text(0.5, 1.0, 'speed')
```

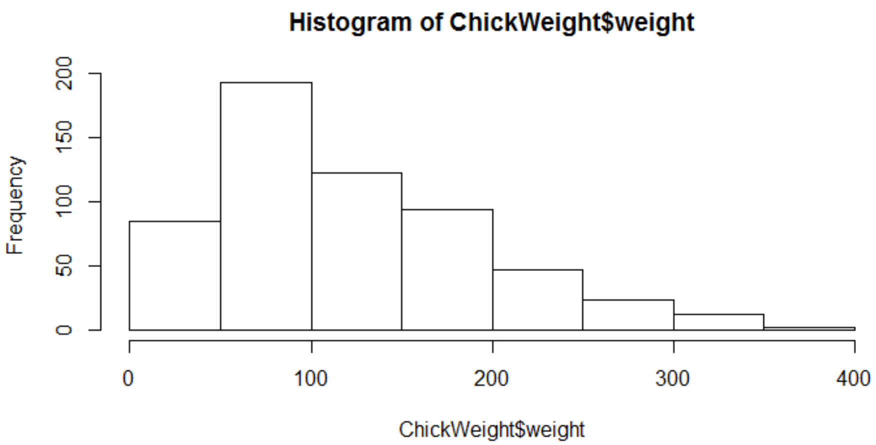
SP and Weight(WT)

Use Q9_b.csv



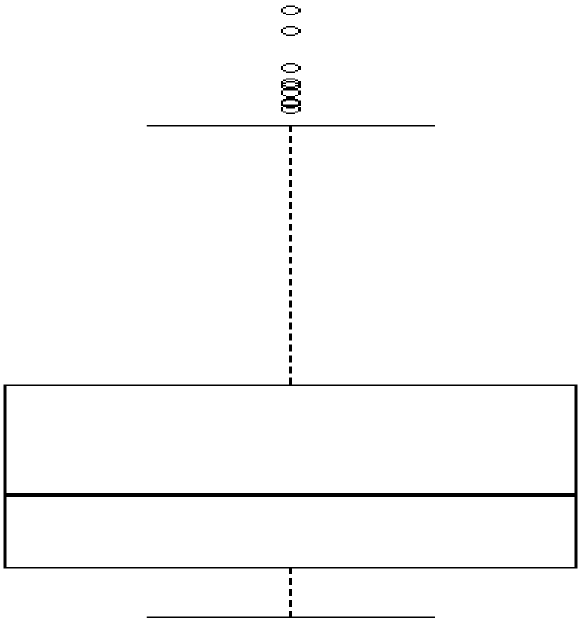
```
In [3]:  df = pd.read_csv("C:/Users/0004IW744/Desktop/Python/Assignments/Basic Stat-1/Q9_b.csv")

In [4]:  print('Skew of whole data:',df.skew())

         Skew of whole data: Unnamed: 0    0.000000
         SP             1.611450
         WT            -0.614753
         dtype: float64

In [5]:  print('Kurtosis of whole data:',df.kurt())

         Kurtosis of whole data: Unnamed: 0   -1.200000
         SP             2.977329
         WT             0.950291
         dtype: float64

In [6]:  fig = plt.subplots(figsize=(15,3))
         plt.subplot(1,4,1)
         plt.boxplot(df.SP)
         plt.title('SP')
         plt.subplot(1,4,2)
         plt.boxplot(df.WT)
         plt.title('WT')

         plt.subplot(1,4,3)
         plt.hist(df.SP)
         plt.title('SP')

         plt.subplot(1,4,4)
         plt.hist(df.WT)
         plt.title('WT')

Out[6]:  Text(0.5, 1.0, 'WT')
```

Q10) Draw inferences about the following boxplot & histogram



1) The data is skewed on the right side. So, data is positively skewed.

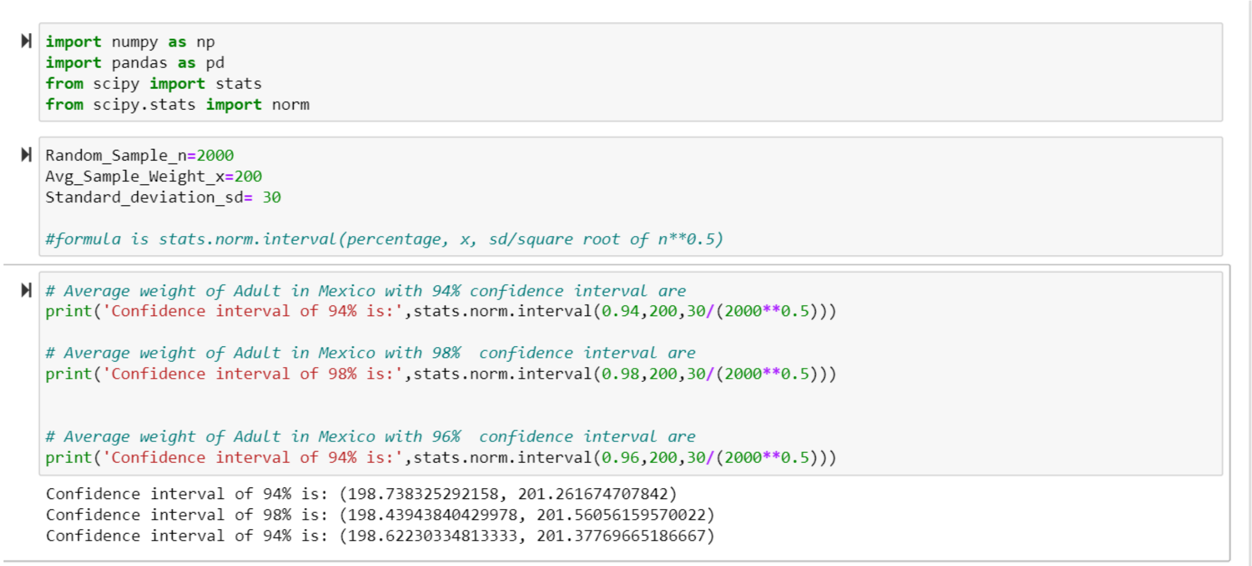
2) There are no outliers for the given data



1) Outliers exists
2) Data is distributed on the right
3) Positive skew

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?

Sol:-



```python
import numpy as np
import pandas as pd
from scipy import stats
from scipy.stats import norm
```

```python
Random_Sample_n=2000
Avg_Sample_Weight_x=200
Standard_deviation_sd= 30

#formula is stats.norm.interval(percentage, x, sd/square root of n**0.5)
```

```python
# Average weight of Adult in Mexico with 94% confidence interval are
print('Confidence interval of 94% is:',stats.norm.interval(0.94,200,30/(2000**0.5)))

# Average weight of Adult in Mexico with 98%  confidence interval are
print('Confidence interval of 98% is:',stats.norm.interval(0.98,200,30/(2000**0.5)))


# Average weight of Adult in Mexico with 96%  confidence interval are
print('Confidence interval of 94% is:',stats.norm.interval(0.96,200,30/(2000**0.5)))
```

```
Confidence interval of 94% is: (198.738325292158, 201.261674707842)
Confidence interval of 98% is: (198.43943840429978, 201.56056159570022)
Confidence interval of 94% is: (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

1) What can we say about the student marks?

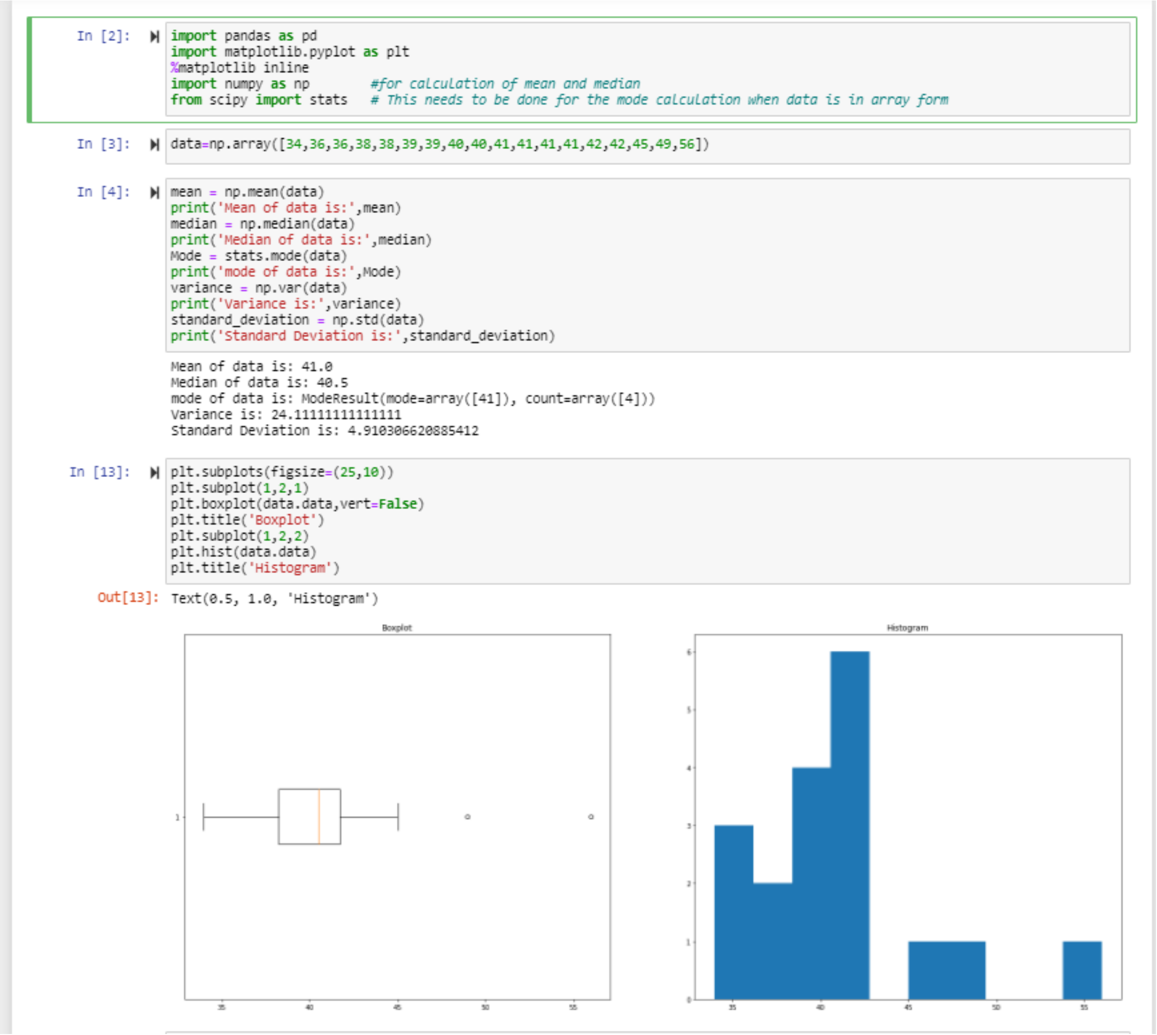
Sol:- --

- Two outliers in the Student's marks: 49 and 56
- It is not following the Normal Distribution

2) Find mean, median, variance, standard deviation.

- Mean of data is: 41.0
- Median of data is: 40.5
- mode of data is: ModeResult(mode=array([41]), count=array([4]))
- Variance is: 24.11111111111111
- Standard Deviation is: 4.910306620885412

Ans)



In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import numpy as np        #for calculation of mean and median
from scipy import stats   # This needs to be done for the mode calculation when data is in array form
```

In [3]:
```
data=np.array([34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56])
```

In [4]:
```
mean = np.mean(data)
print('Mean of data is:',mean)
median = np.median(data)
print('Median of data is:',median)
Mode = stats.mode(data)
print('mode of data is:',Mode)
variance = np.var(data)
print('Variance is:',variance)
standard_deviation = np.std(data)
print('Standard Deviation is:',standard_deviation)
```

```
Mean of data is: 41.0
Median of data is: 40.5
mode of data is: ModeResult(mode=array([41]), count=array([4]))
Variance is: 24.11111111111111
Standard Deviation is: 4.910306620885412
```

In [13]:
```
plt.subplots(figsize=(25,10))
plt.subplot(1,2,1)
plt.boxplot(data.data,vert=False)
plt.title('Boxplot')
plt.subplot(1,2,2)
plt.hist(data.data)
plt.title('Histogram')
```

Out[13]:
```
Text(0.5, 1.0, 'Histogram')
```

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

Sol) Symmetrical distribution

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

Sol) Right-skewed

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

Sol) Left-skewed

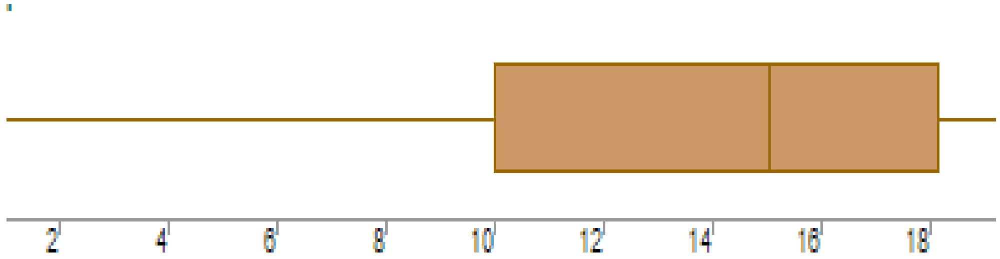
Q16) What does positive kurtosis value indicates for a data ?

Sol) Distribution is peaked and possess thick tails for given data

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

Sol) Distribution is not peaked and don’t have thick tails for given data

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



What can we say about the distribution of the data?

Sol) No Outliers

q1= 18

Median= 15.2

Q3 = 10

What is nature of skewness of the data?

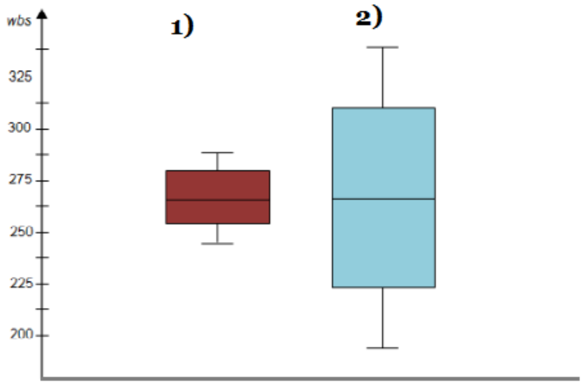
Sol)Negative Skewness

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

Sol) IQR = Q3-Q1

= 8

Q19) Comment on the below Boxplot visualizations?



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

For the above data, we can say that

- Skewness is '0'
- Normal Distribution Exists for both
- There are no Outliers for both plots

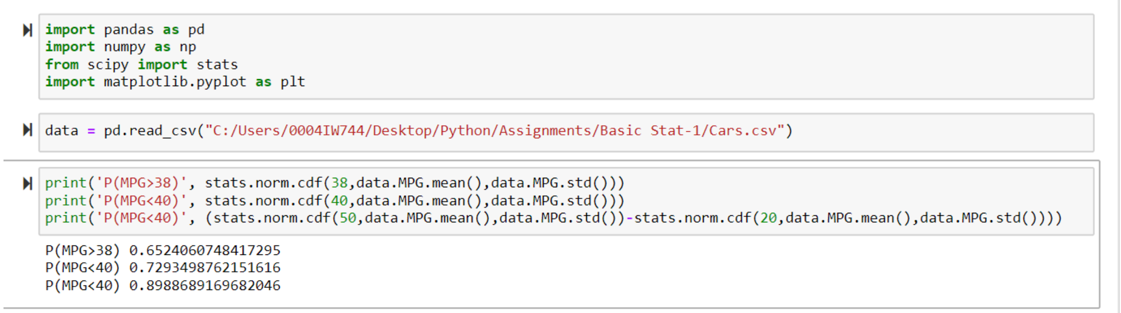
Q 20) Calculate probability from the given dataset for the below cases
Data _set: Cars.csv

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)

Sol:-



```
import pandas as pd
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
```

```
data = pd.read_csv("C:/Users/0004IW744/Desktop/Python/Assignments/Basic Stat-1/Cars.csv")
```

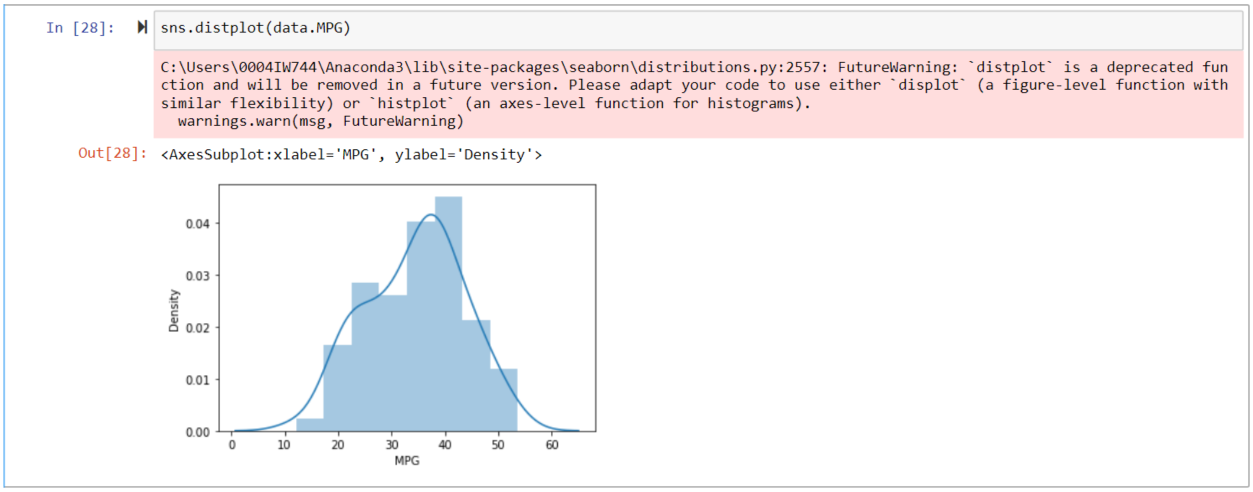
```
print('P(MPG>38)', stats.norm.cdf(38,data.MPG.mean(),data.MPG.std()))
print('P(MPG<40)', stats.norm.cdf(40,data.MPG.mean(),data.MPG.std()))
print('P(MPG<40)', (stats.norm.cdf(50,data.MPG.mean(),data.MPG.std())-stats.norm.cdf(20,data.MPG.mean(),data.MPG.std())))
```

```
P(MPG>38) 0.6524060748417295
P(MPG<40) 0.7293498762151616
P(MPG<40) 0.8988689169682046
```

Q 21) Check whether the data follows normal distribution

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

Sol:-



```
In [28]:  sns.distplot(data.MPG)
```

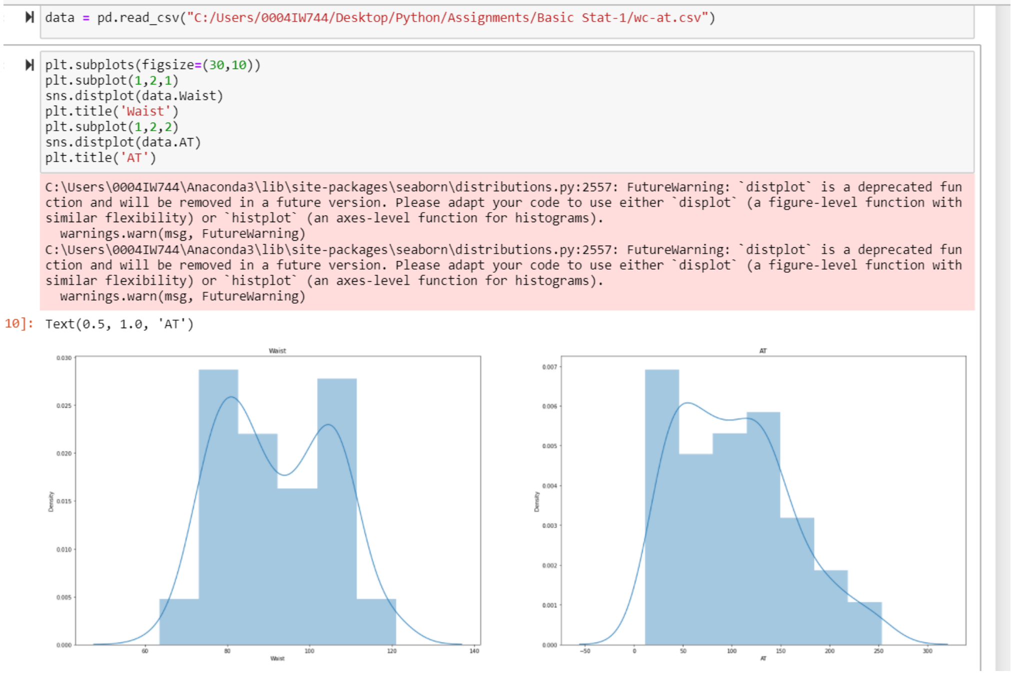
```
C:\Users\0004IW744\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated fun
ction and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

```
Out[28]: <AxesSubplot:xlabel='MPG', ylabel='Density'>
```

BellShape – Yes(Approximately), Skewness – Negative

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

Sol:-



```
data = pd.read_csv("C:/Users/0004IW744/Desktop/Python/Assignments/Basic Stat-1/wc-at.csv")
```

```
plt.subplots(figsize=(30,10))
plt.subplot(1,2,1)
sns.distplot(data.Waist)
plt.title('Waist')
plt.subplot(1,2,2)
sns.distplot(data.AT)
plt.title('AT')
```

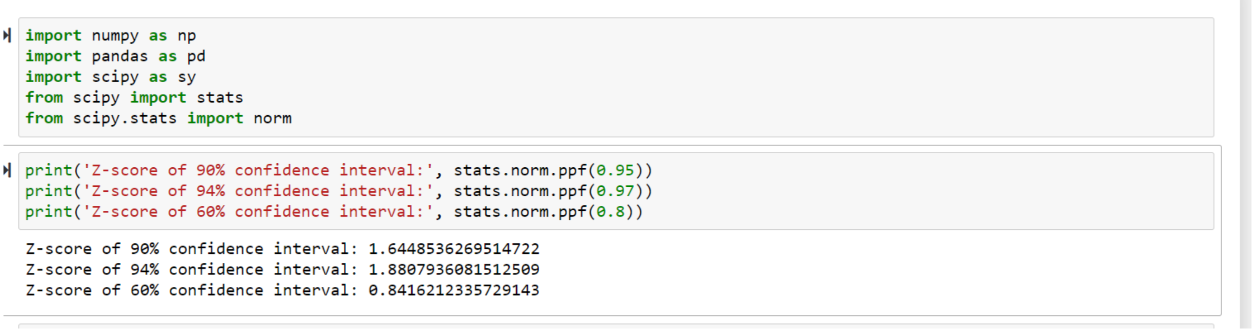
```
C:\Users\0004IW744\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated fun
ction and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\Users\0004IW744\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated fun
ction and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

```
10]: Text(0.5, 1.0, 'AT')
```

In Above:- Waist doesn’t following normal distribution and AT is following approximately

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

Sol:-



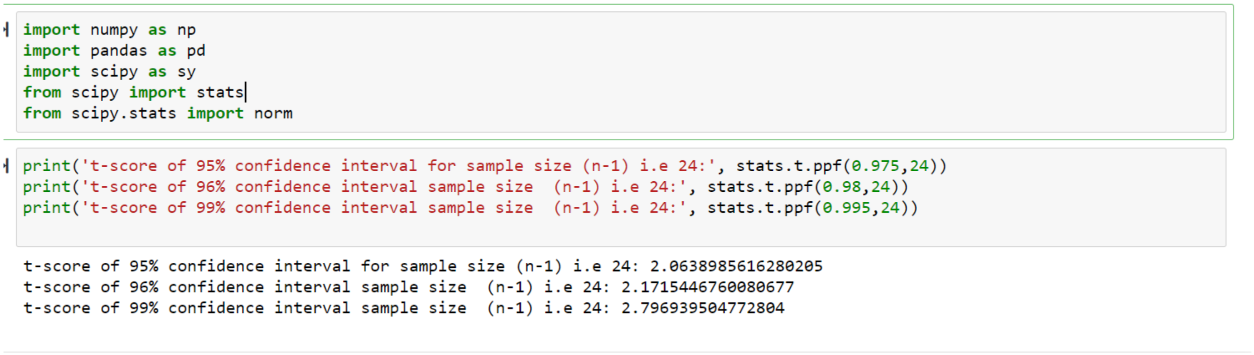
```
import numpy as np
import pandas as pd
import scipy as sy
from scipy import stats
from scipy.stats import norm
```

```
print('Z-score of 90% confidence interval:', stats.norm.ppf(0.95))
print('Z-score of 94% confidence interval:', stats.norm.ppf(0.97))
print('Z-score of 60% confidence interval:', stats.norm.ppf(0.8))
```

```
Z-score of 90% confidence interval: 1.6448536269514722
Z-score of 94% confidence interval: 1.8807936081512509
Z-score of 60% confidence interval: 0.8416212335729143
```

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

Sol:- (df)= sample size – 1=24



```
import numpy as np
import pandas as pd
import scipy as sy
from scipy import stats
from scipy.stats import norm
```

```
print('t-score of 95% confidence interval for sample size (n-1) i.e 24:', stats.t.ppf(0.975,24))
print('t-score of 96% confidence interval sample size  (n-1) i.e 24:', stats.t.ppf(0.98,24))
print('t-score of 99% confidence interval sample size  (n-1) i.e 24:', stats.t.ppf(0.995,24))
```

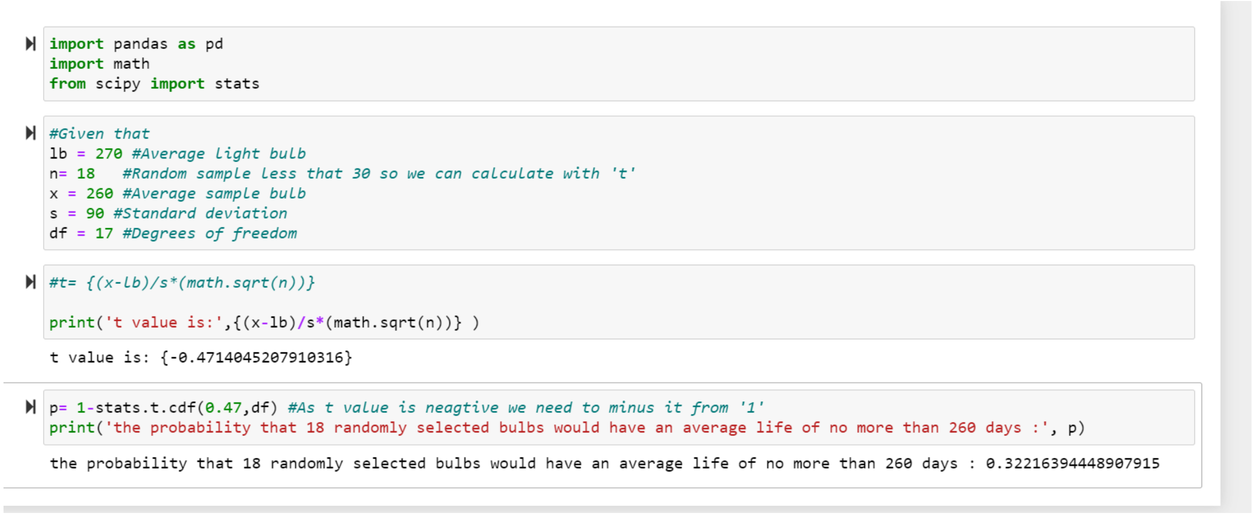
```
t-score of 95% confidence interval for sample size (n-1) i.e 24: 2.0638985616280205
t-score of 96% confidence interval sample size  (n-1) i.e 24: 2.1715446760080677
t-score of 99% confidence interval sample size  (n-1) i.e 24: 2.796939504772804
```

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



```
import pandas as pd
import math
from scipy import stats
```

```
#Given that
lb = 270 #Average light bulb
n= 18   #Random sample less that 30 so we can calculate with 't'
x = 260 #Average sample bulb
s = 90 #Standard deviation
df = 17 #Degrees of freedom
```

```
#t= {(x-lb)/s*(math.sqrt(n))}

print('t value is:',{(x-lb)/s*(math.sqrt(n))} )
```

```
t value is: {-0.4714045207910316}
```

```
p= 1-stats.t.cdf(0.47,df) #As t value is neagtive we need to minus it from '1'
print('the probability that 18 randomly selected bulbs would have an average life of no more than 260 days :', p)
```

```
the probability that 18 randomly selected bulbs would have an average life of no more than 260 days : 0.32216394448907915
```