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	Interval
Time on a Clock with Hands	Interval
Number of Children	Ratio
Religious Preference	Nominal
Barometer Pressure	Interval
SAT Scores	Interval
Years of Education	Ratio

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

Ans: P (2 heads and 1 tail) = N (2 heads and 1 tail) / N (3 coins tossed) = **3/8**

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

a) Equal to 1, **Ans: 0**
b) Less than or equal to 4, **Ans:** P = N (Sum is <= 4) / N (2 dice rolled) = **6/36**
c) Sum is divisible by 2 and 3, **Ans:** P = N (Sum is divisible by 2 and 3) / N(Two dice rolled) = **6/36**

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: Choose 2 from 5(2R3G) / Choose 2 from 7(2R3G2B) = $^5C_2/^7C_2$ = **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: 0.015+0.8+1.95+0.025+0.06+0.2 = **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 Comment about the values/ Draw some inferences.

Use Q7.csv file

Ans:

	Points	Score	Weigh
mean	3.596563	3.21725	17.84875
median	3.695	3.325	17.71
mode	3.07	3.44	17.02
variance	0.285881	0.957379	3.193166
standard deviation	0.534679	0.978457	1.786943
range	(2.76,4.93)	(1.513,5.424)	(14.5,22.9)

points and *score* variables have very close mean, median and mode values but *score* clearly has larger variance and standard deviation. hence, we can see *score* variable has larger range than *points* variable. Looking at *weigh* variable, mode is much smaller than mean and median thus the distribution of *weigh* variable must be little positively-skewed.
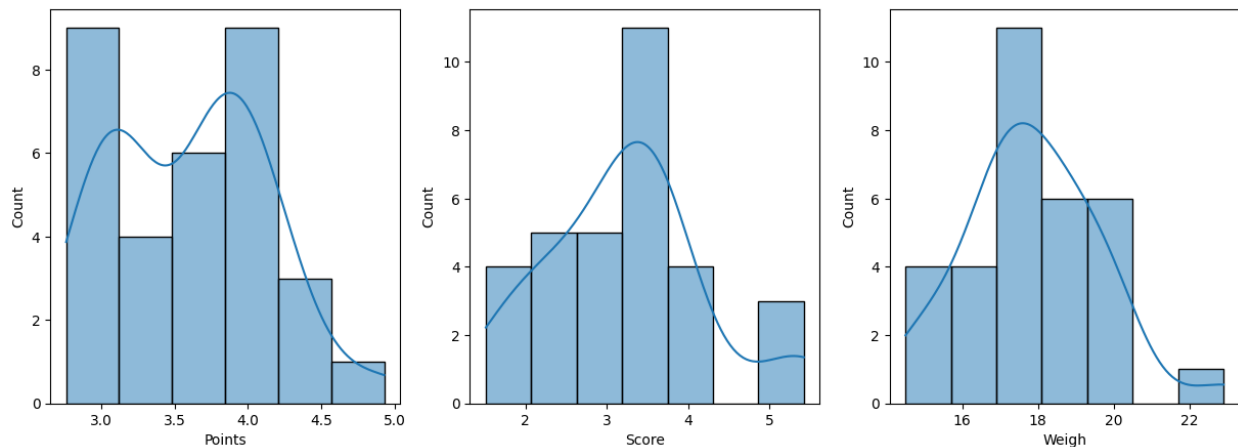
Q8) Calculate Expected Value for the problem below

a) 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:** here expected value will be the mean. i.e., **145.33 pounds**

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

Car's speed and distance: Use Q9_a.csv, SP and Weight (WT): Use Q9_b.csv

	Q9_a		Q9_b	
	speed	**distance**	**SP**	**WT**
Skewness	-0.1175	0.8068	1.6114	-0.6147
Kurtosis	-0.5089	0.4050	2.9773	0.9502

- Q9_a: speed is left skewed & has flatten peak, distance is right skewed has shaper peak.
- Q9_b: SP is right skewed & has sharper peak, WT is left skewed and has sharper peak.

Q10) Draw inferences about the following boxplot & histogram

Histogram of ChickWeight$weight
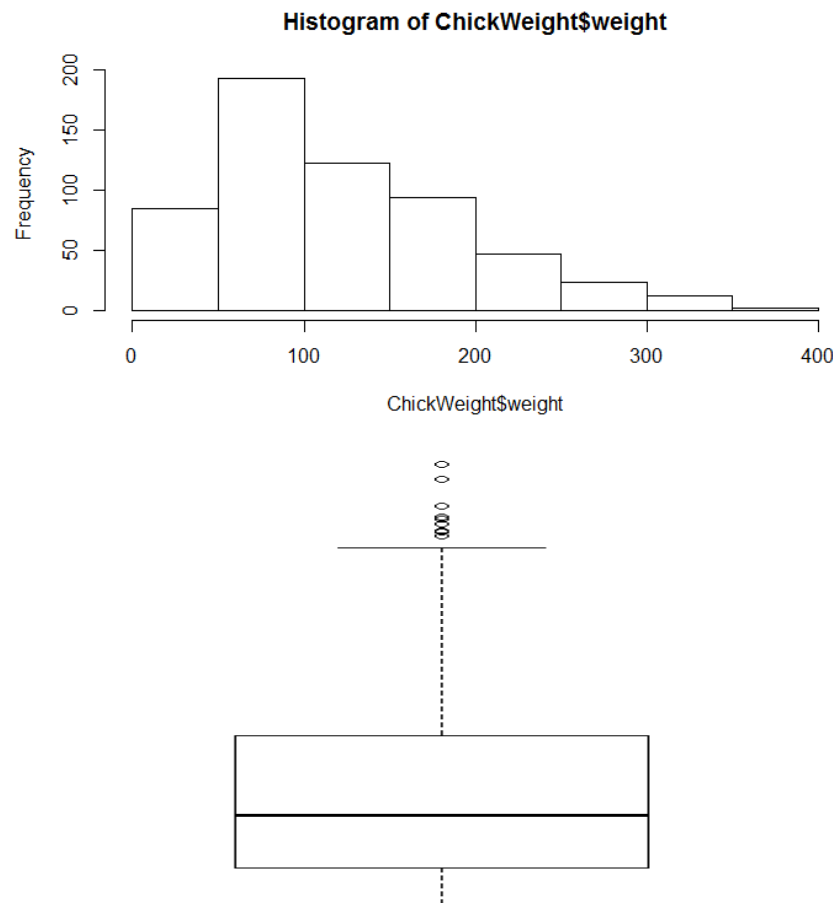
Ans: 1) we can infer that the data is right skewed
2) From boxplot we can see there are outliers present in data
3) mode<median<mean

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:

Confidence level	Confidence Interval
94%	(198.73, 201.26)
98%	(198.43, 201.56)
96%	(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.
    2) What can we say about the student marks?

Ans: 1) mean=41, median=40.5, variance= 25.529, std=5.052, mode=41
      2) The data is slightly right skewed

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

Ans: it will be symmetrical data, no skewness

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

Ans: Positive or right skewed data

Q15) What is the nature of skewness when median > mean?
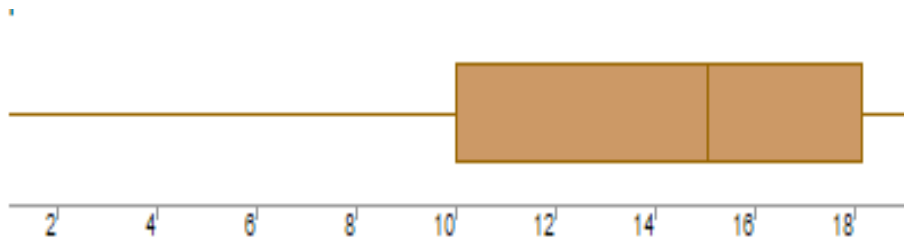
Ans: Negative or left skewed data

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

Ans: Data is more peaked than normal distribution

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

Ans: Data is less peaked (flat) than normal distribution

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

What can we say about the distribution of the data?

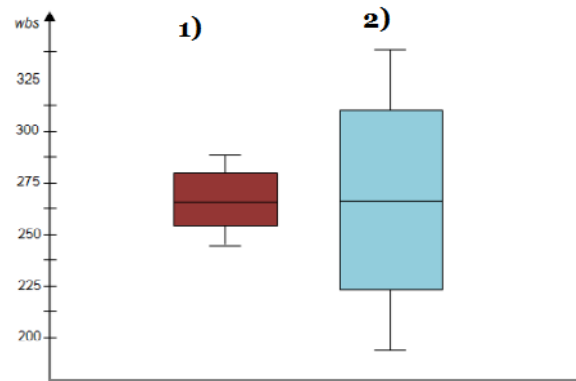**Ans:** The distribution is left skewed, mode>median>mean

What is nature of skewness of the data?

Ans: Negatively skewed

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

Ans: 18-10 = 8 units

Q19) Comment on the below Boxplot visualizations?

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

Ans: 1) The median of both boxplots is equal
2) The data 2 has more spread than data 1
3) The IQR of 2 is larger than 1, 4) There are no outliers, 5) Both has almost normal distribution

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), **Ans:** 0.347
  b. P(MPG<40), **Ans:** 0.729
 c.   P (20<MPG<50), **Ans:** 0.898

Q 21) Check whether the data follows normal distribution
  a) Check whether the MPG of Cars follows Normal Distribution, Dataset: Cars.csv
  Ans: No, MPG is not normal distribution

  b) Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution, Dataset: wc-at.csv
  Ans: No, both AT and Waist are not normal distribution
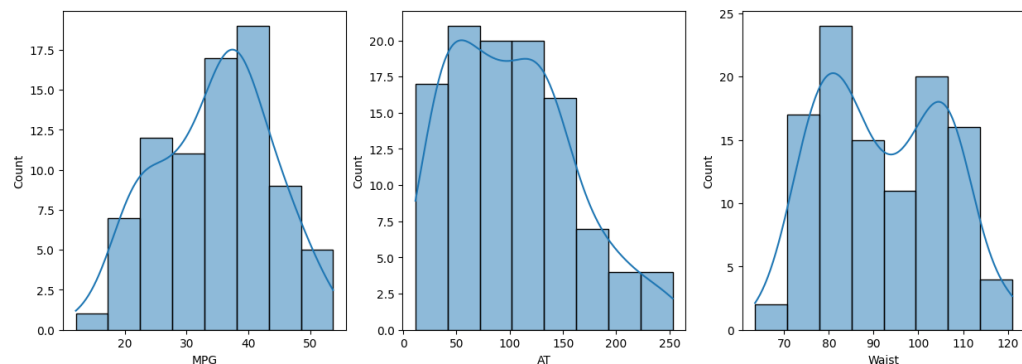
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

90%	-1.6449
94%	-1.8808
60%	-0.8416

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

95%	-2.064
96%	-2.172
99%	-2.797

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

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{\{n\}}}} = \frac{260 - 270}{90/\sqrt{18}} = -0.471$$

```
from scipy import stats

stats.t.cdf(-0.471,df=17)
```
= 0.321

=32 % probability