1. Define a sample with size 30. Construct a frequency distribution table with 6 classes and an appropriate stem-and-leaf display. Sketch the histograms for the frequency distribution table and the stem-and-leaf display. Compare the histograms and discuss the skewness of your dataset.

521, 522, 527, 548, 550, 559, 560, 570, 572, 574, 578, 585, 592, 592, 607, 616, 618, 621, 629, 637, 638, 640, 656, 668, 707, 709, 719, 737, 739, 752

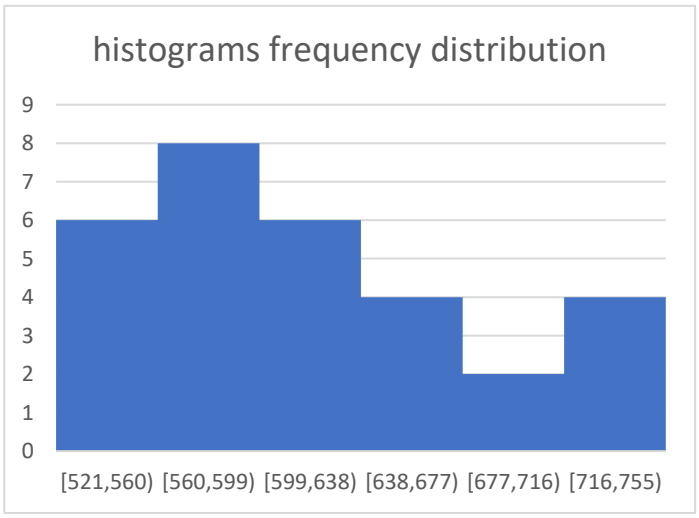
Min: 521, Max: 752, and Range: 231

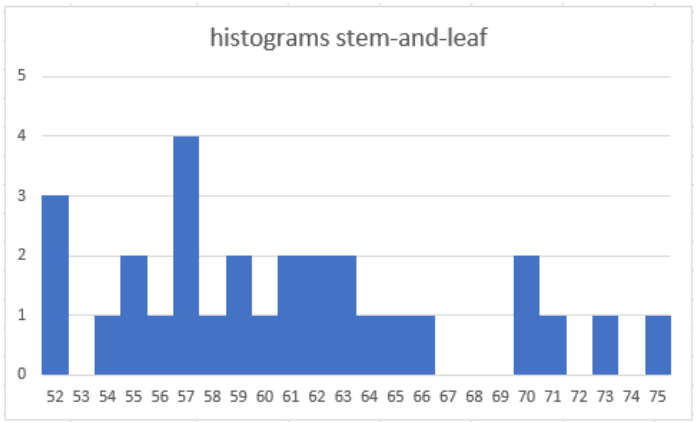
class range: 231/6 = 38.5 ~ 39

Classes	frequency	rel. freq.	percentage	cum. freg.	cum. rel. freq.
[521,560)	6	0.2	20	6	0.2
[560,599)	8	0.267	26.7	14	0.467
[599,638)	6	0.2	20	20	0.667
[638,677)	4	0.133	13.3	24	0.8
[677,716)	2	0.067	6.7	26	0.867
[716,755)	4	0.133	13.3	30	1
Total	30	1	100	-	-

Stem	Leaf
52	1,2,7
53	-
54	8
55	0,9
56	0
57	0,2,4,8
58	5
59	2,2
60	7
61	6,8
62	1,9
63	7,8
64	0
65	6
66	8
67	-
68	-
69	-
70	7,9
71	9
72	-
73	7,9

52|1 = 521

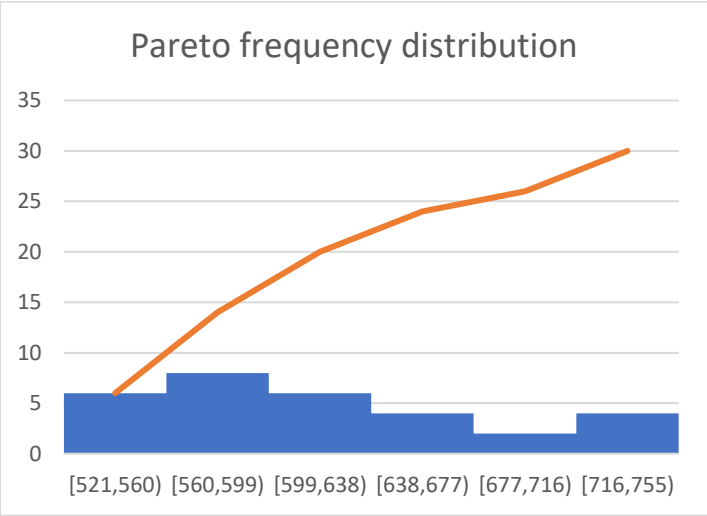




74	-
75	2

According to histograms frequency distribution, this dataset has right-skewed distribution.

The Stem-and-Leaf graph shows more distribution in detail and we can do analyze and make a better decision

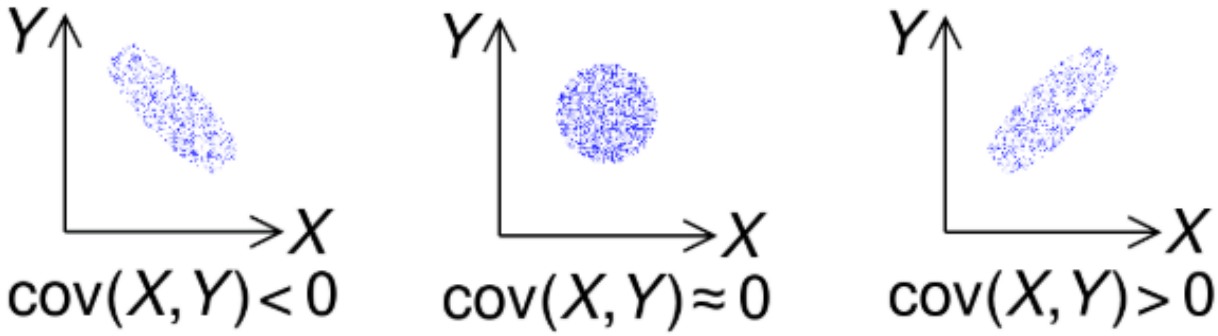


2. What are covariance and correlation coefficient? Write a question for which you are asked to obtain the values of those. Then construct a scatter plot for your variables. Finally, discuss the relationship between the variables using the plot and the values of covariance and correlation coefficient.

Covariance and Correlation are very helpful in understanding the relationship between two continuous variables.

Covariance signifies the direction of the linear relationship between the two variables. By direction we mean if the variables are directly proportional or inversely proportional to each other. (Increasing the value of one variable might have a positive or a negative impact on the value of the other variable).

$$\operatorname{cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))(y_i - E(Y)).$$

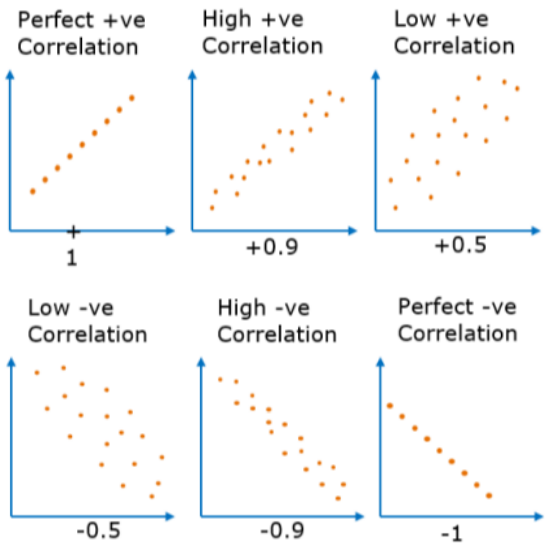


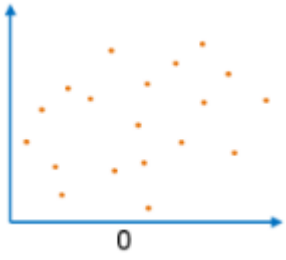
Correlation analysis is a method of statistical evaluation used to study the strength of a relationship between two, numerically measured, continuous variables.

$$Correlation = \frac{Cov\ (x,y)}{\sigma x * \sigma y}$$

where:

- cov is the covariance
- σ_X is the standard deviation of X
- σ_Y is the standard deviation of Y





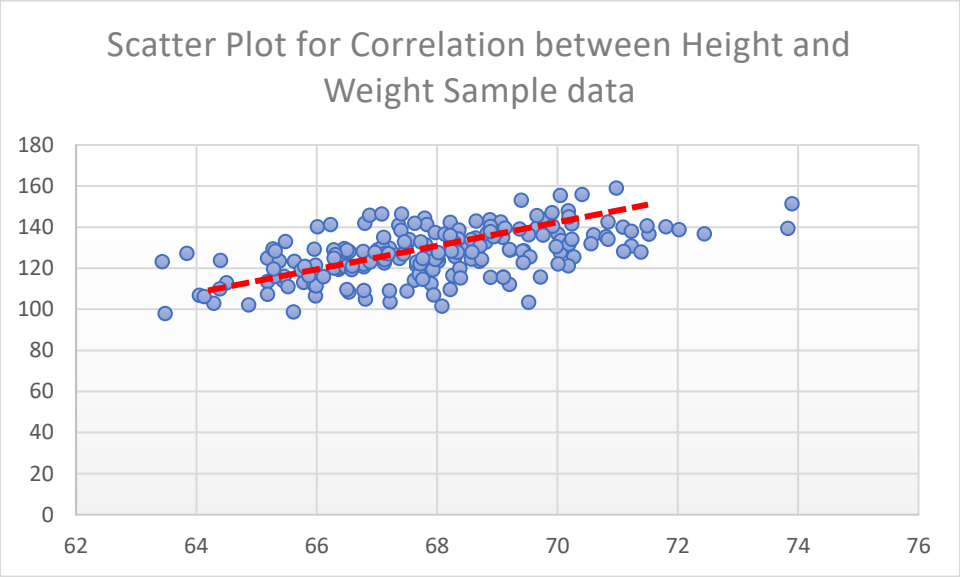
However, if it is 0 then we can only say that there is no linear relationship. There could exist other functional relationships between the variables.

For the second part, the sample data is extracted from the below site that related to the relation between Height and Weight of random people,

http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_Dinov_020108_HeightsWeights

The output is,

Covariance: 12.92405
Correlation: 0.556865



According to the amount of Covariance and correlation, these two values have a low positive correlation that you can see in the scatter graph as well.

Reference,

https://towardsdatascience.com/covariance-and-correlation-321fdacab168

https://www.mygreatlearning.com/blog/covariance-vs-correlation/