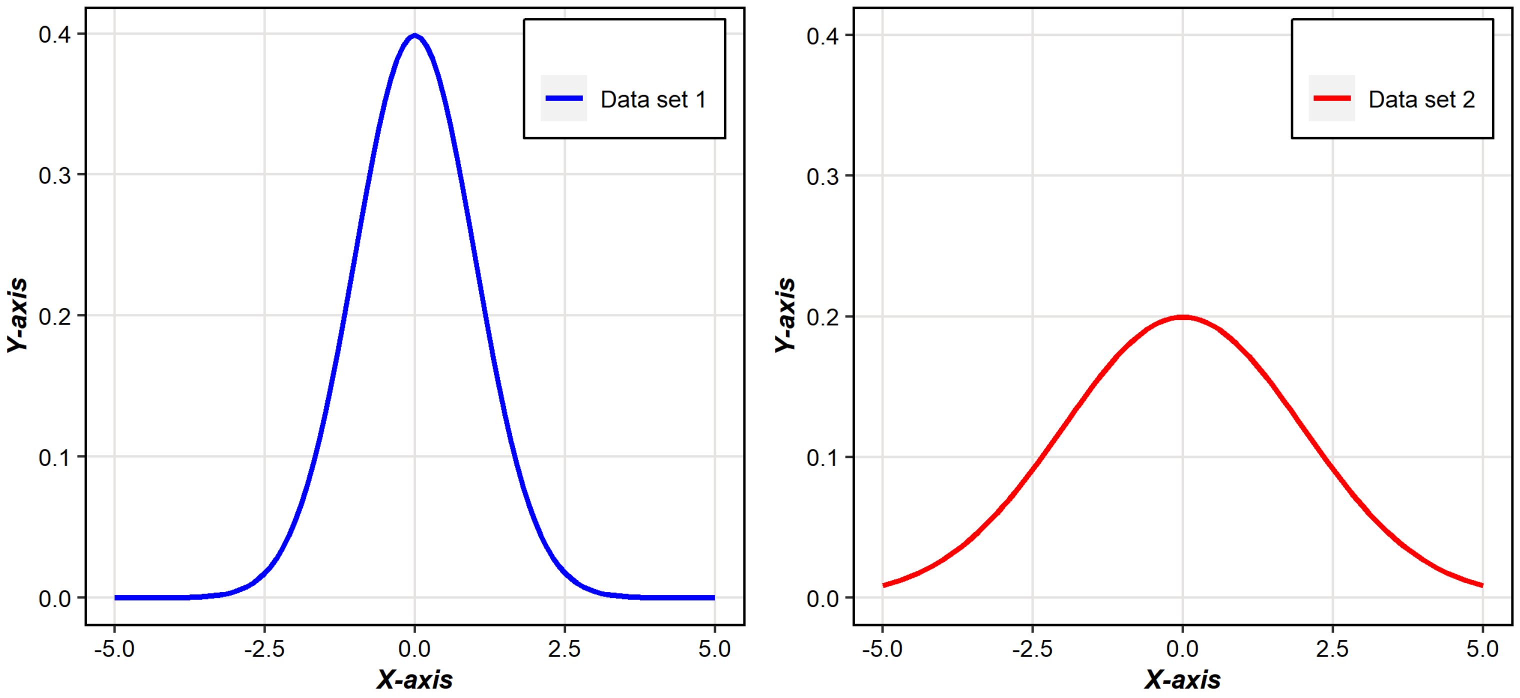
Shape of the distribution

Md. Ismail Hossain Riday

Shape of the distribution

- In previous chapter, we learned about “Location” and “Dispersion” which are two important quantitative concepts.

- Two data sets may have identical means and identical variance, but their graphical shapes may be different



Mean & Variance fail to describe the shape of the data distribution

Moments

- Moments are constant
- Which used to determine some characteristics/properties of frequency distribution

Moments

- Moments are constant
- Which used to determine some characteristics/properties of frequency distribution

Skewness

Moments

- Moments are constant

- Which used to determine some characteristics/properties of frequency distribution

Shape of the distribution

Skewness

Kurtosis

Moments

- r^{th} moments can be written as,

- $\mu_r = \frac{\sum(X_i - \bar{X})^r}{N}$; $r = 1,2,3,4, \dots$

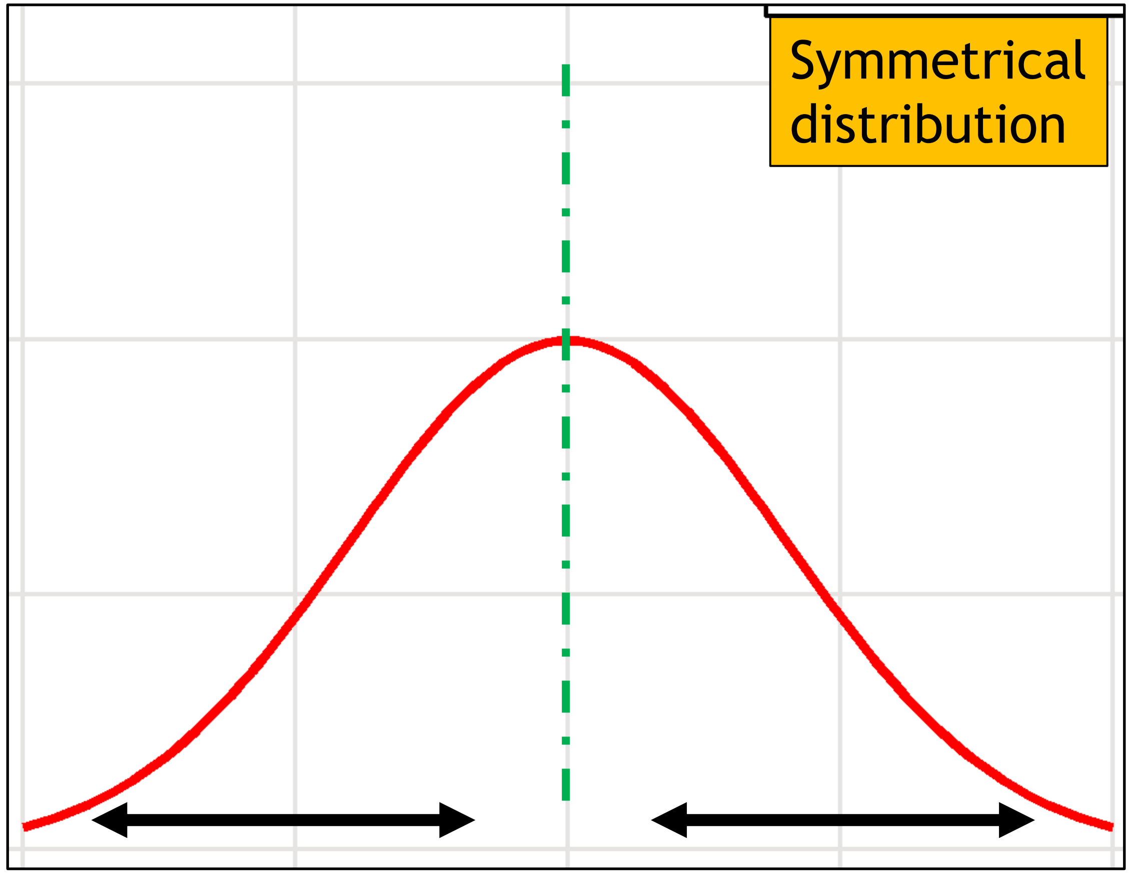
- $If\ r = 1, \mu_1 = First\ moment = \frac{\sum(X_i - \bar{X})^1}{N}$
- $If\ r = 2, \mu_2 = Second\ moment = \frac{\sum(X_i - \bar{X})^2}{N}$
- $If\ r = 3, \mu_3 = Third\ moment = \frac{\sum(X_i - \bar{X})^3}{N}$
- $If\ r = 4, \mu_4 = Fourth\ moment = \frac{\sum(X_i - \bar{X})^4}{N}$

Skewness

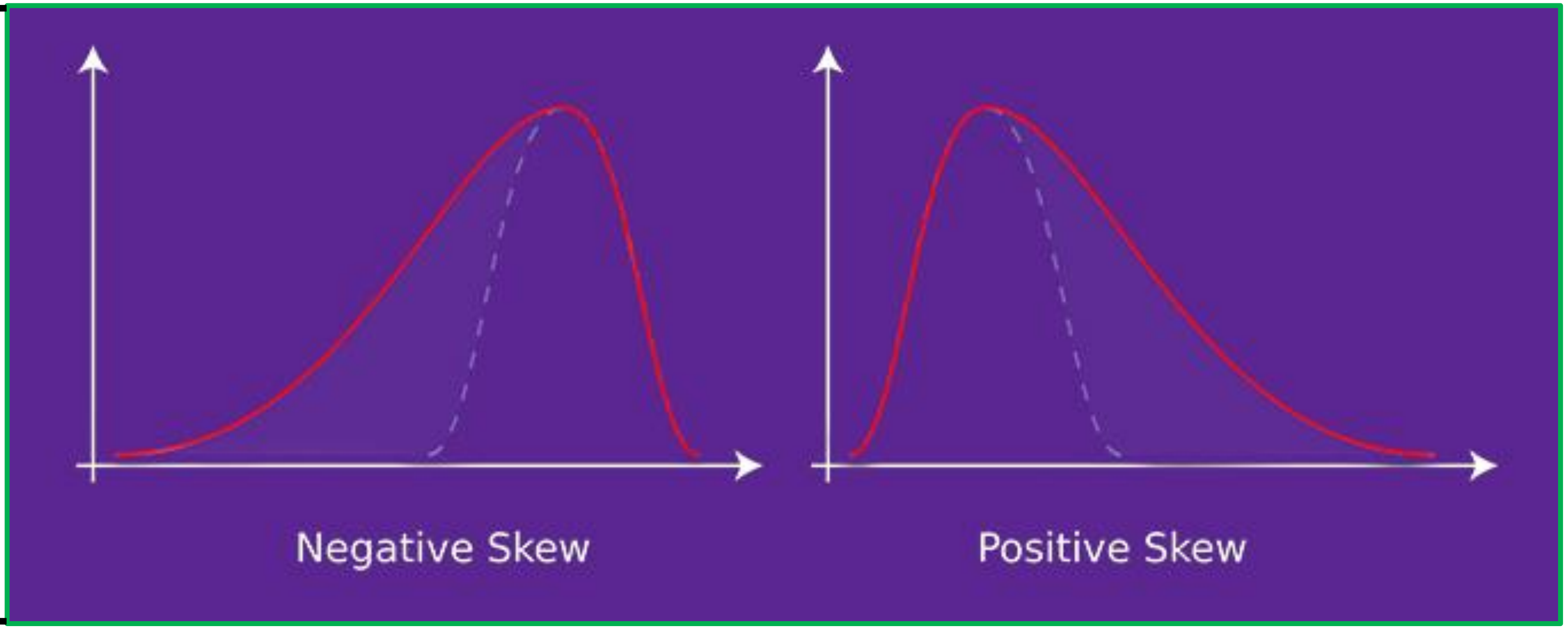
- Lack of symmetry of a distribution

Skewness

- Lack of symmetry of a distribution

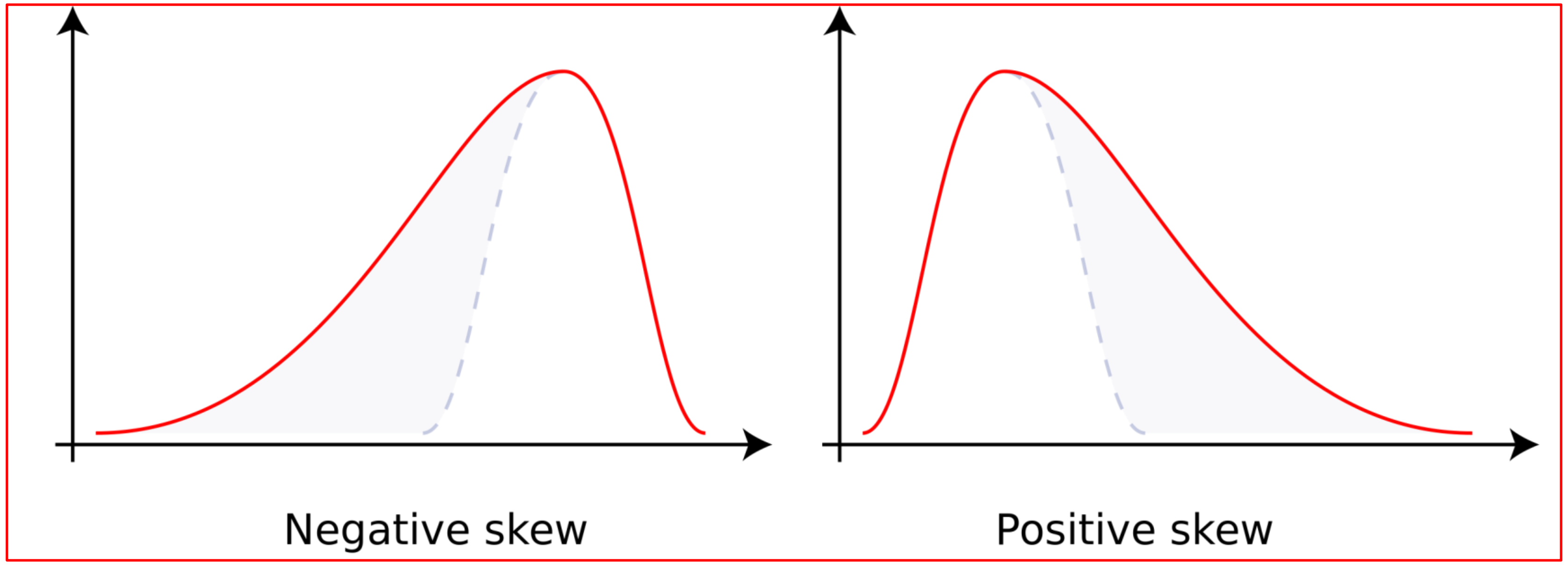


Lack of symmetry



Types of skewness

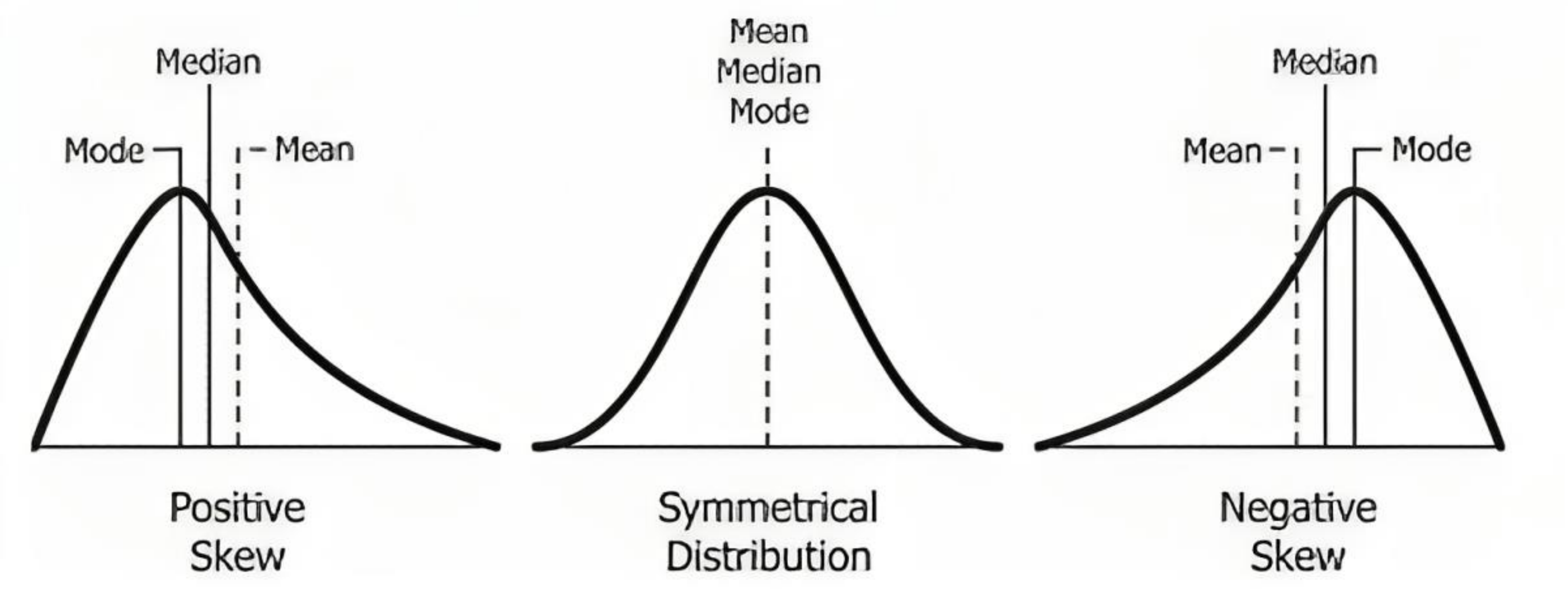
- There are two types of skewness or lack of symmetry occurs:



Positive skewness

Negative skewness

Graph



$Mean > Median > Mode$

$Mean < Median < Mode$

Coefficient of skewness

- $S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$

For example: Calculate the coefficient of skewness: 15,18,2,6,4

- $\boldsymbol{S_k > 0: Positively\ skewed}$
- $\boldsymbol{S_k < 0: Negatively\ skewed}$
- $\boldsymbol{S_k = 0: Symmetric}$

Coefficient of skewness

- $S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$

- $S_k > 0$: $\boldsymbol{Positively\ skewed}$
- $S_k < 0$: $\boldsymbol{Negatively\ skewed}$
- $S_k = 0$: $\boldsymbol{Symmetric}$

For example: Calculate the coefficient of skewness: 15,18,2,6,4

Mean = 9

Coefficient of skewness

- $S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$

- $\boldsymbol{S_k > 0: Positively\ skewed}$
- $\boldsymbol{S_k < 0: Negatively\ skewed}$
- $\boldsymbol{S_k = 0: Symmetric}$

For example: Calculate the coefficient of skewness: 15,18,2,6,4

Mean = 9
Median = 6

Coefficient of skewness

- $S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$

- $\boldsymbol{S_k > 0: Positively\ skewed}$
- $\boldsymbol{S_k < 0: Negatively\ skewed}$
- $\boldsymbol{S_k = 0: Symmetric}$

For example: Calculate the coefficient of skewness: 15,18,2,6,4

Mean = 9
Median = 6
SD = 7.07

Coefficient of skewness

- $S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$

- $\boldsymbol{S_k > 0: Positively\ skewed}$
- $\boldsymbol{S_k < 0: Negatively\ skewed}$
- $\boldsymbol{S_k = 0: Symmetric}$

For example: Calculate the coefficient of skewness: 15,18,2,6,4

Mean = 9
Median = 6
SD = 7.07

$$S_k = \frac{3(Mean - Median)}{SD} = \frac{3(9-6)}{7.07} = 1.27$$

Coefficient of skewness

- $S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$

- $\boldsymbol{S_k > 0}$: ***Positively skewed***
- $\boldsymbol{S_k < 0}$: ***Negatively skewed***
- $\boldsymbol{S_k = 0}$: ***Symmetric***

For example: Calculate the coefficient of skewness: 15,18,2,6,4

Mean = 9
Median = 6
SD = 7.07

$$S_k = \frac{3(Mean - Median)}{SD} = \frac{3(9-6)}{7.07} = 1.27$$

Since $S_k > 0$, thus the distribution is positively skewed distribution.

Kurtosis

- $If\ r = 1, \mu_1 = First\ moment = \frac{\sum(X_i - \bar{X})^1}{N}$
- $If\ r = 2, \mu_2 = Second\ moment = \frac{\sum(X_i - \bar{X})^2}{N}$
- $If\ r = 3, \mu_3 = Third\ moment = \frac{\sum(X_i - \bar{X})^3}{N}$
- $If\ r = 4, \mu_4 = Fourth\ moment = \frac{\sum(X_i - \bar{X})^4}{N}$

- Degree of peaked or flatness of a distribution

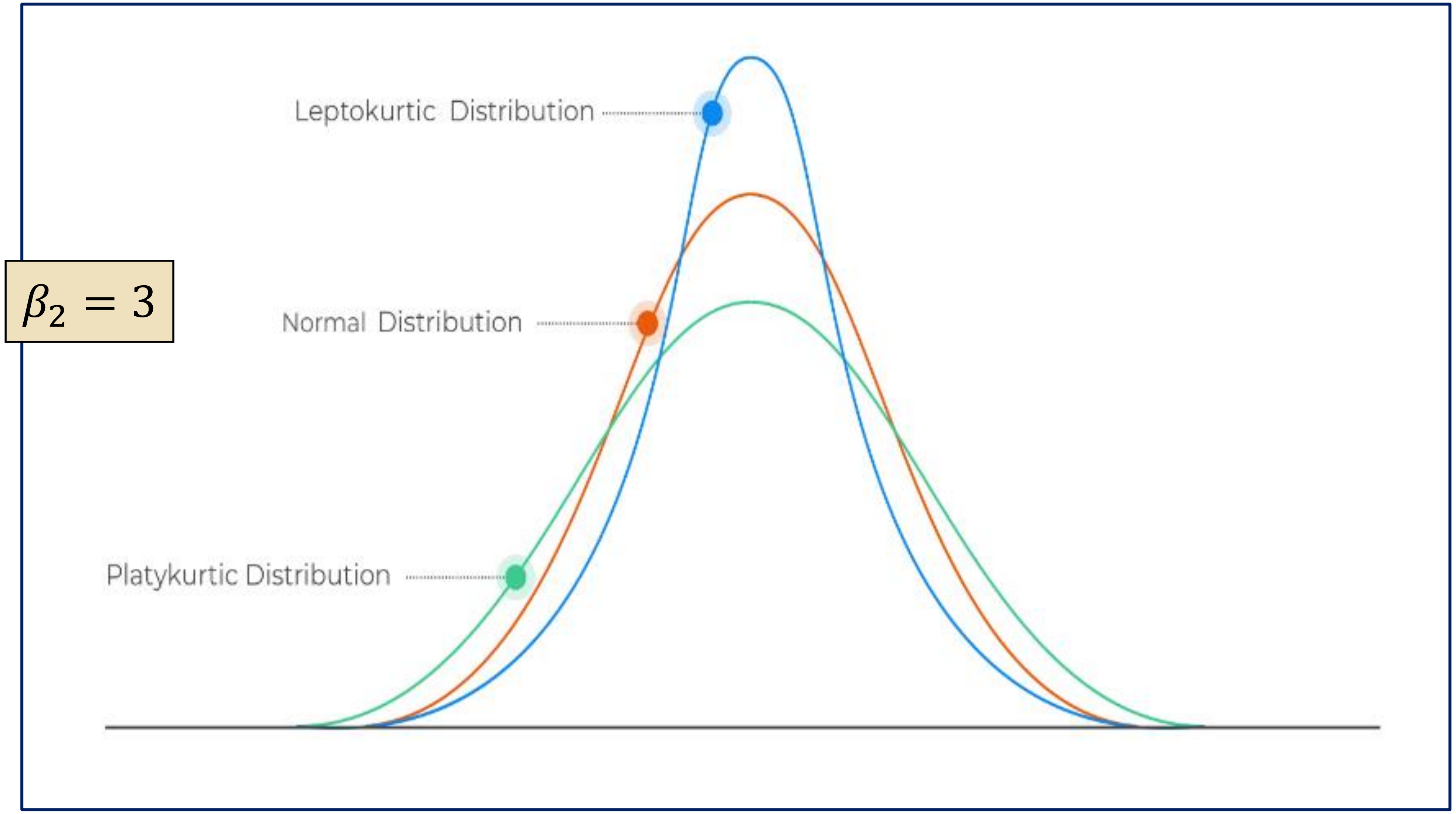
- Three types-

a) Leptokurtic $\beta_2 > 3$

b) Mesokurtic/Normal $\beta_2 = 3$

c) Platykurtic $\beta_2 < 3$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$



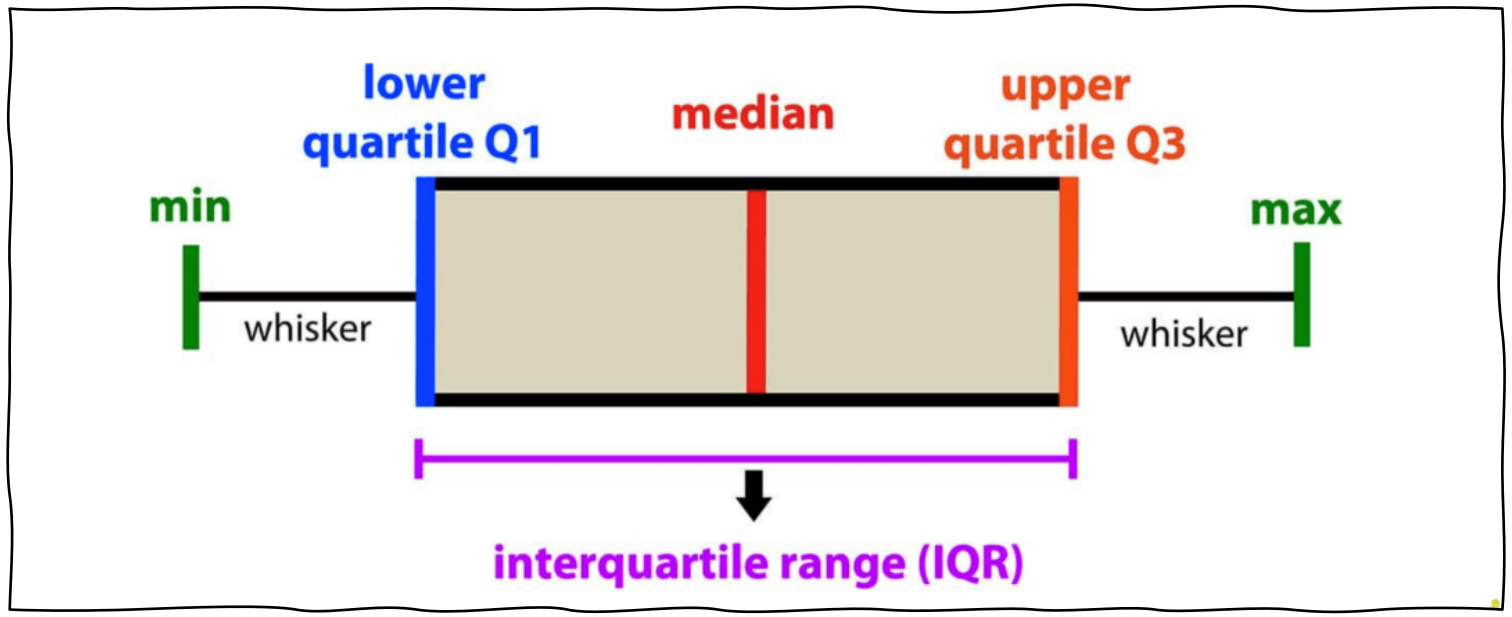
Box & Whisker plot

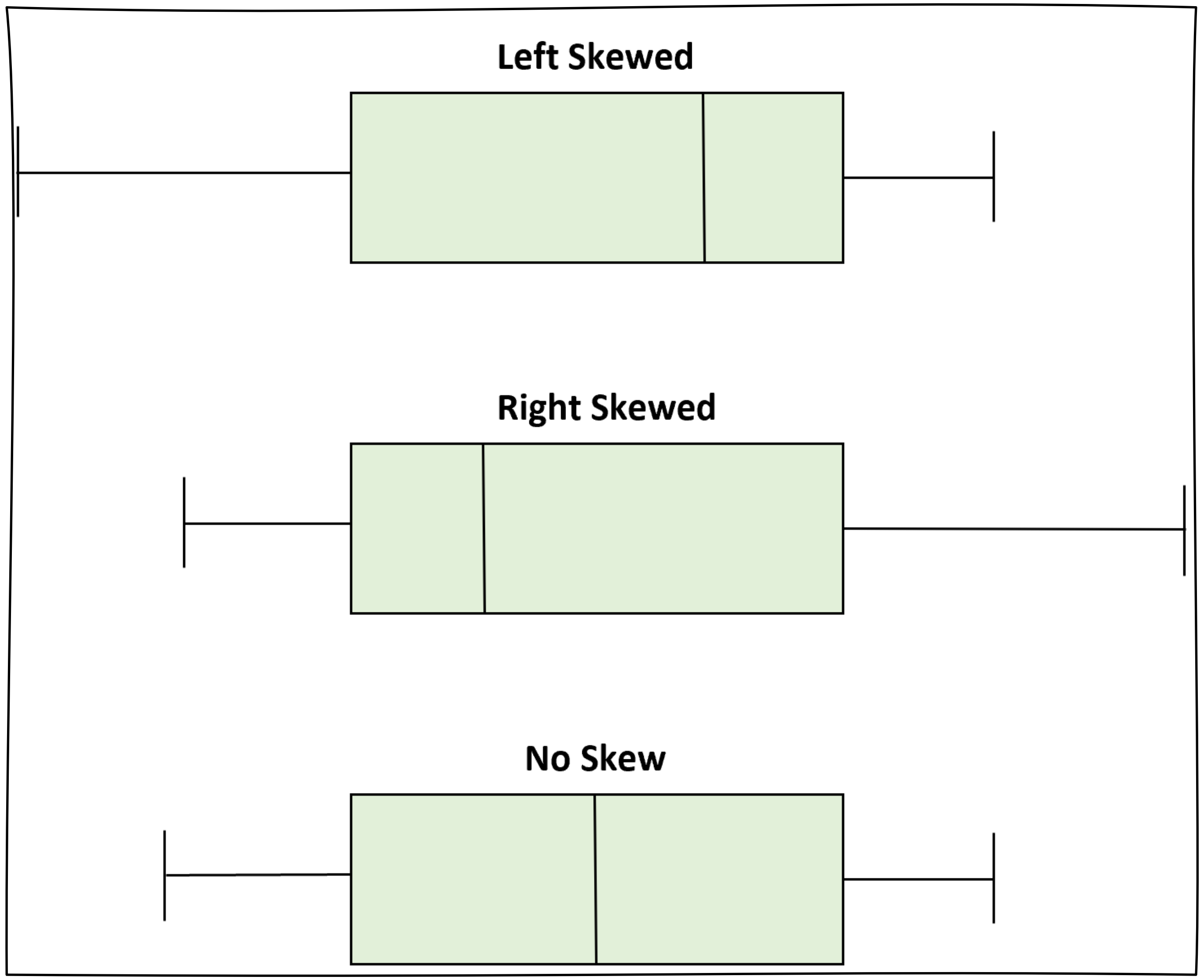
- A box-whisker plot, also known as a box plot, is a graphical representation of the distribution of a dataset.
- It displays the five-number summary of the data:

1. Minimum value
2. First Quartile (Q1)
3. Second Quartile/Median (Q2)
4. Third Quartile (Q3)
5. Maximum value

Importance of Box plot

- To get idea of the shape of the distribution
- To detect outliers from the data.





Identify outliers from boxplot

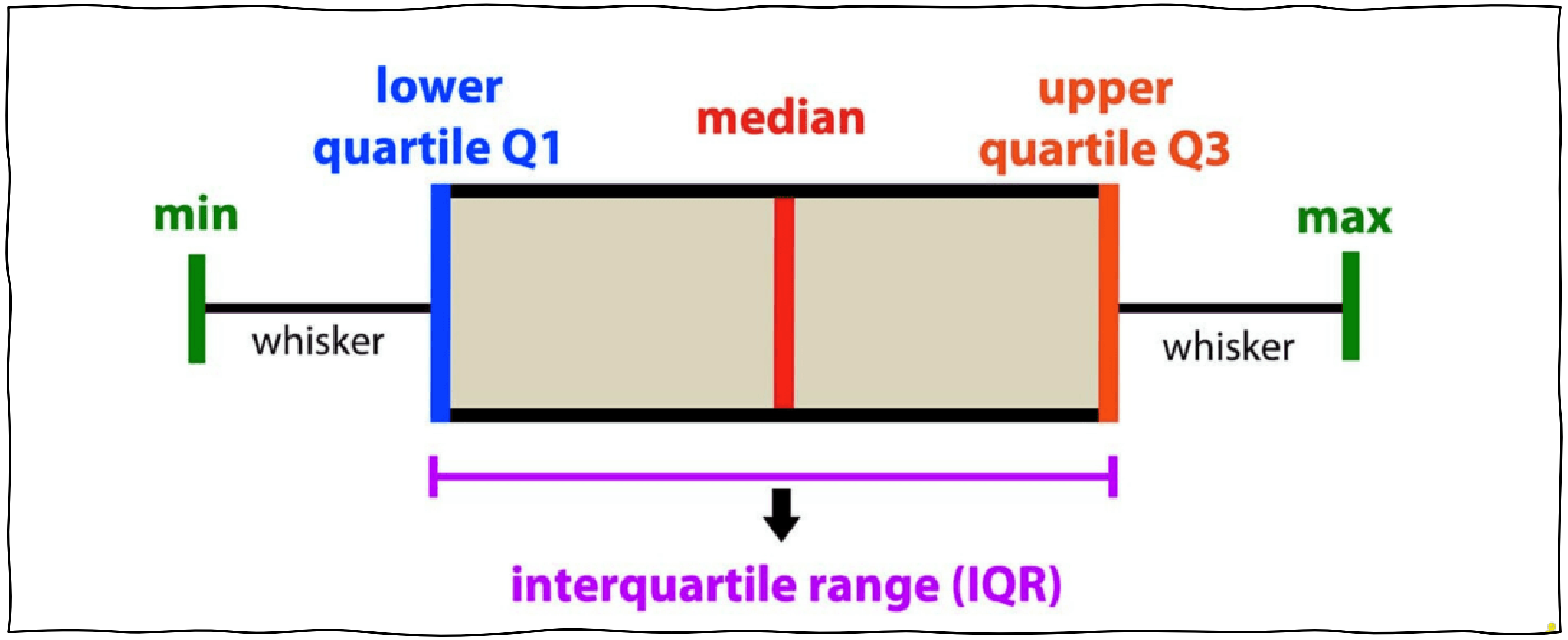
$$IQR = Inter\ Quartile\ Range$$
$$IQR = Q_3 - Q_1$$

Lower Fences:

$$\{Q_1 - 1.5 \times IQR\}$$

Upper Fences:

$$\{Q_3 + 1.5 \times IQR\}$$



Identify outliers from boxplot

2900, 2765, 2960, 2890, 2880, 2720, 2930, 2950, 2860, 3060, 3260, 3525

Draw the box-whisker plot and detect the outlier(s)

Organize the data into ascending order,

2720, 2765, 2860, 2880, 2890, 2900, 2930, 2950, 2960, 3060, 3260, 3525

Now,

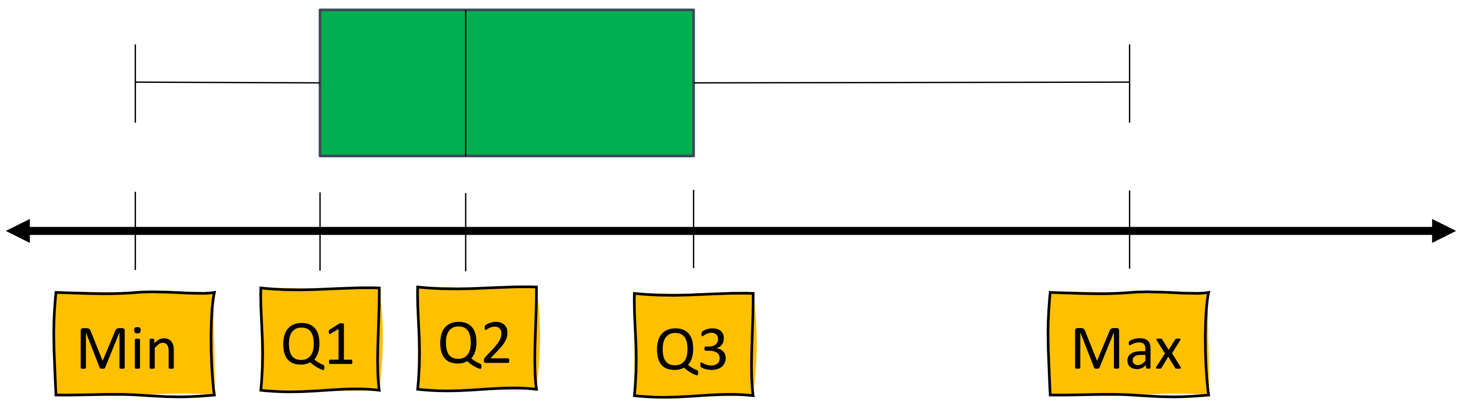
Min=2720

Q1=2870

Q2=2915

Q3=3010

Max=3525



Identify outliers from boxplot

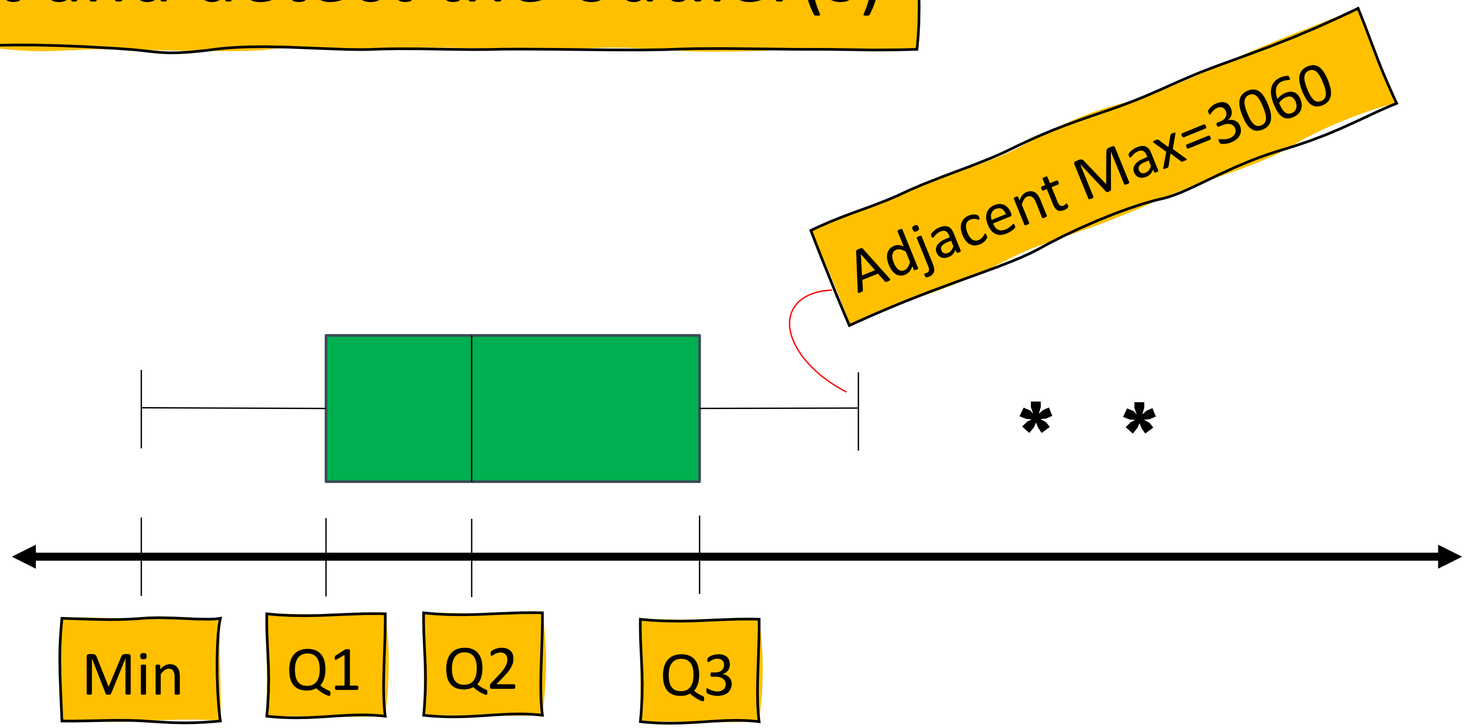
2720, 2765, 2860, 2880, 2890, 2900, 2930, 2950, 2960, 3060, 3260, 3525

Draw the box-whisker plot and detect the outlier(s)

$$\therefore IQR = Q_3 - Q_1 = 140$$

$$Lower = \{Q_1 - 1.5 \times IQR\} = 2660$$

$$Upper = \{Q_3 + 1.5 \times IQR\} = 3220$$



Thank You