

Box and Whisker Plot

By [BrushMyQuant](#)



YouTube Video Link to this Post is [Here](#)

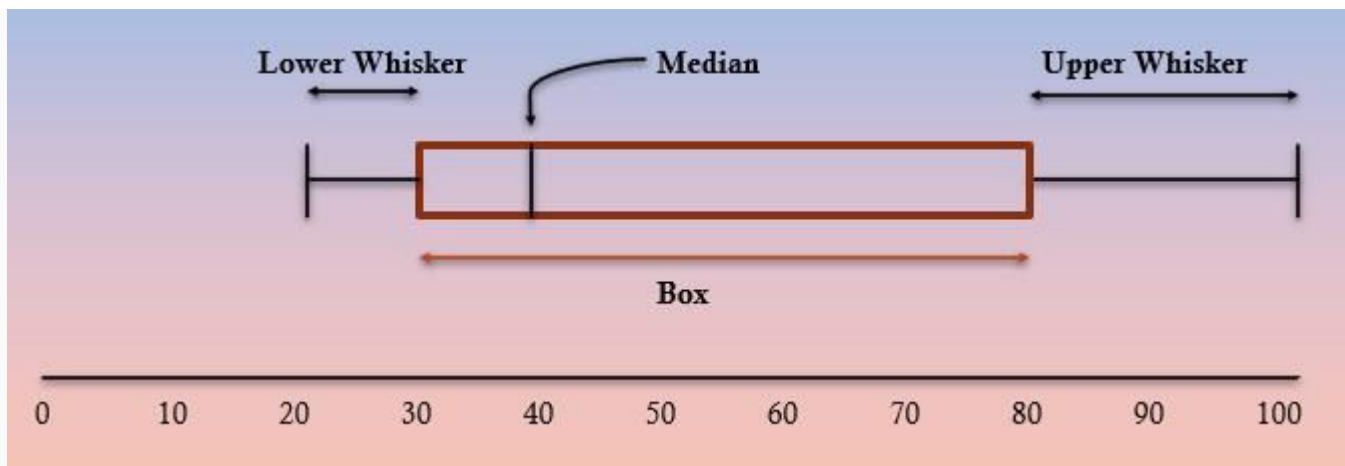
Following is covered in the video

- ▣ What is Box and Whisker Plot?
- ▣ Five Number Summary: Basics of Box and Whisker Plot.
- ▣ Insights from Box and Whisker Plot. (Top and Bottom Values)
- ▣ Inter Quartile Range, Outliers and Modified Box and Whisker Plot
- ▣ Comparing two Box and Whisker Plots
- ▣ Skewness in Box and Whisker Plot

What is Box and Whisker Plot?

A Box and Whisker plot is a graphical method of representing the data in Quartiles. This also helps in easily identifying the locality, spread and skewness in the dataset.

Example: Let's say the marks of the students in a class (from 0 to 100) are given to us in the form of below Box and Whisker Plot.



The left whisker in above Box and Whisker Plot is called as Lower Whisker

The right whisker in above Box and Whisker Plot is called as Upper Whisker

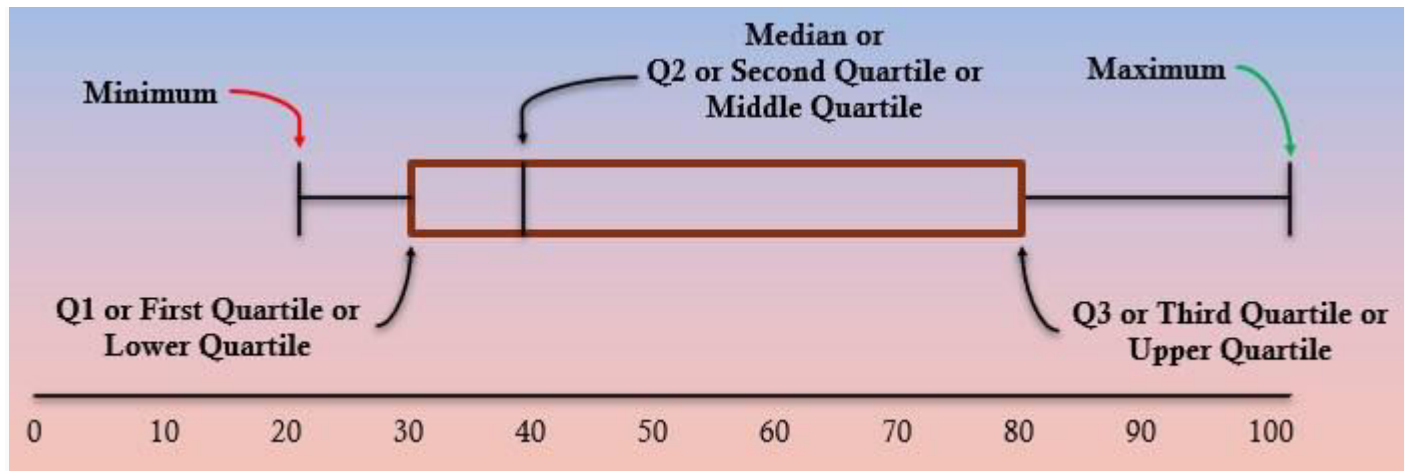
The box in above Box and Whisker Plot is called as Box

The line in between the Box is as the Median of the dataset.

Five Number Summary: Basics of Box and Whisker Plot

Five number summary of a Box and Whisker Plot is given by following values

- ✧ Minimum
- ✧ Q1 or First Quartile or Lower Quartile
- ✧ Median or Q2 or Second Quartile or Middle Quartile
- ✧ Q3 or Third Quartile or Upper Quartile
- ✧ Maximum



In Above Box and Whisker Plot:

- ✧ **Minimum** : Is the starting point from left of lower whisker = 20
- ✧ **Q1** : Is the start of the box from left = 30
- ✧ **Median or Q2** : Is the line in between the box = 40
- ✧ **Q3** : Is the end of the box = 80
- ✧ **Maximum** : Is the ending point of upper whisker = 100

Q 1: Marks of 11 students in a class test are given below: (Min Marks =0, Max Marks = 100)

55, 10, 64, 88, 40, 65, 50, 52, 45, 58, 100

- 1.1 Make a box and Whisker Plot representing this data set.
- 1.2 Find the Highest score.
- 1.3 Find the Lowest score.
- 1.4 Find the Range.
- 1.5 Find the value of Q1
- 1.6 Find the value of Q2
- 1.7 Find the value of Q3

Sol 1: Let's start by arranging the terms in ascending order, we get

10, 40, 45, 50, 52, 55, 58, 64, 65, 88, 100

As there are 11 terms so median will be the $(11+1)/2 = 6^{\text{th}}$ term

=> Median = 55

So, the set will be divided into three parts

First Half = 10, 40, 45, 50, 52 and median of first half = $Q1 = 3^{\text{rd}} = 45$

$Q2 = \text{Median} = 55$

Second Half = 58, 64, 65, 88, 100 and median of second half = $Q3 = 3^{\text{rd}} \text{ term} = 65$

So, the Box and Whisker Plot has

Min = 10 (starting of lower whisker)

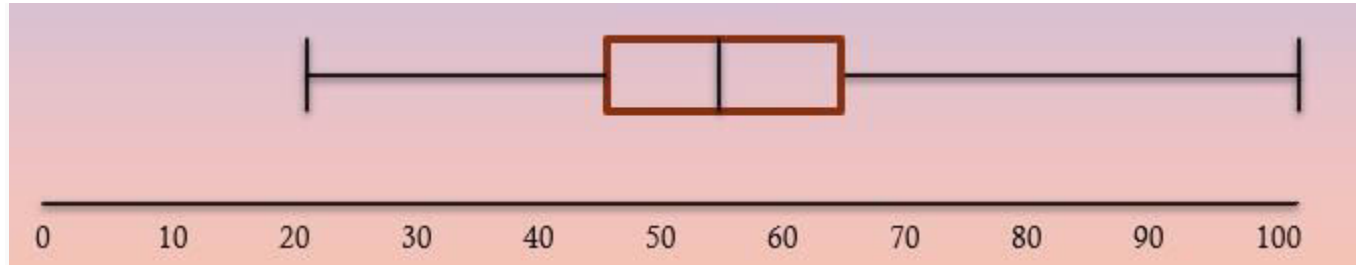
Q1 = 45 (start of the box)

Q2 = 55 (place of line inside the box)

Q3 = 65 (end of the box)

Max = 100 (end of upper whisker)

1.1 So, the Box and Whisker Plot will be as below:



1.2 Find the Highest score - 100

1.3 Find the Lowest score - 10

1.4 Find the Range = $100 - 10 = 90$

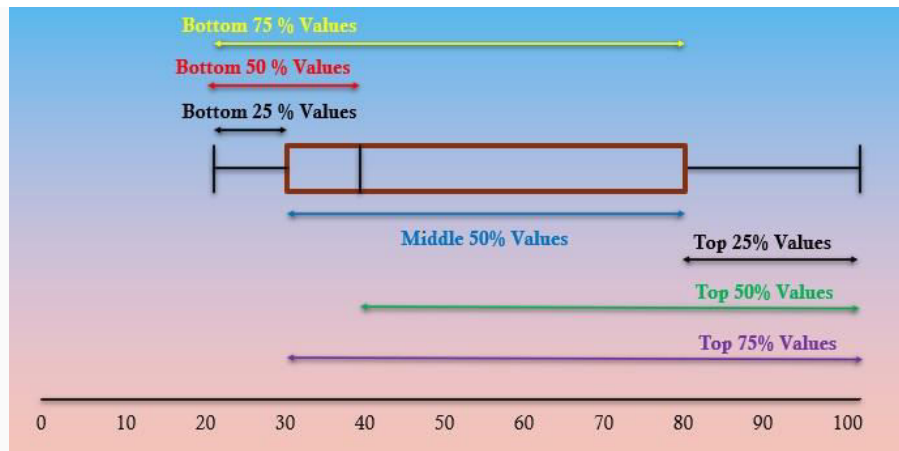
1.5 Find the value of $Q1 = 45$

1.6 Find the value of $Q2 = 55$

1.7 Find the value of $Q3 = 65$

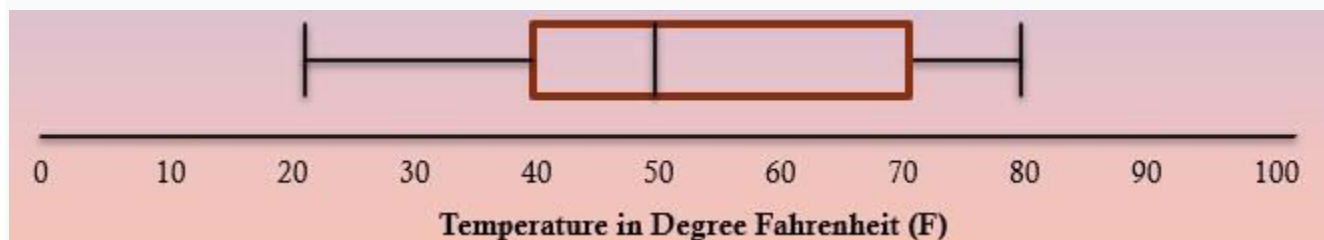
Insights from Box and Whisker Plot. (Top and Bottom Values)

Bottom 25% of Values are between Min to Q1
Bottom 50% of Values are between Min to Q2
Bottom 75% of Values are between Min to Q3
Middle 50% values are between Q1 to Q3
Top 25% of Values are between Q3 to Max
Top 50% of Values are between Q2 to Max
Top 75% of Values are between Q1 to Max



Q2. Answer the following questions based on the information given in the Box and Whisker Plot below:

- 2.1 What was the lowest temperature across the days?
- 2.2 What was the Highest temperature across the days?
- 2.3 What was the Median temperature across the days?
- 2.4 More than 50% of the days had a temperature of more than 42F? True/False?
- 2.5 Colder 25% of the days had a temperature between?
- 2.6 Hotter 25% of the days had a temperature between?
- 2.7 Middle 50% of the days had a temperature between?
- 2.8 Majority of the days had temperature of more than 45F? True/False?
- 2.9 At least 75% of the days had temperature less than 75F? True/False?



Sol 2:

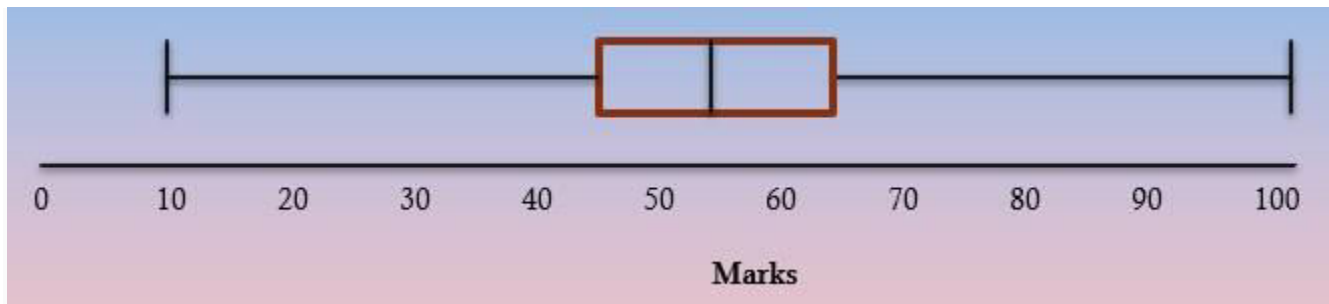
- 2.1 What was the lowest temperature across the days? : **20F as Min value is 20**
2.2 What was the Highest temperature across the days? : **80F as Max value is 80**
2.3 What was the Median temperature across the days? : **80F as Median value is 50**
2.4 More than 50% of the days had a temperature of more than 42F? True/False? : **True as top 50% values lies between Median and Max and Median is 50F**
2.5 Colder 25% of the days had a temperature between? : **20(Min) and 40(Q1)**
2.6 Hotter 25% of the days had a temperature between? : **70(Q3) and 80(Max)**
2.7 Middle 50% of the days had a temperature between? : **40(Q1) and 70(Q3)**
2.8 Majority of the days had temperature of more than 45F? True/False? : **True as top 50% values lies between Median and Max and Median is 50F and 45F is to the left of Median**
2.9 At least 75% of the days had temperature less than 75F? True/False? : **True as bottom 75% values lies between Min and Q3 and Q3 is 70F and 75F is to the right of Q3**

Inter Quartile Range, Outliers and Modified Box and Whisker Plot

Inter Quartile Range = $Q3 - Q1$

Q3. Marks of 11 students in a class test are given below. (Min marks = 0, Max Marks = 100). Find the IQR of the set.

55, 10, 64, 88, 40, 65, 50, 52, 45, 58, 100



Sol 3: $IQR = Q3 - Q1 = 65 - 45 = 20$

Outliers: Any extremely lower or higher value in the set as compared to other values in the set

Any value which is $< Q1 - 1.5 \cdot IQR$ and $> Q3 + 1.5 \cdot IQR$ is an outlier

In above Box and Whisker plot

$$Q1 - 1.5 \cdot IQR = 45 - 1.5 \cdot 20 = 45 - 30 = 15$$

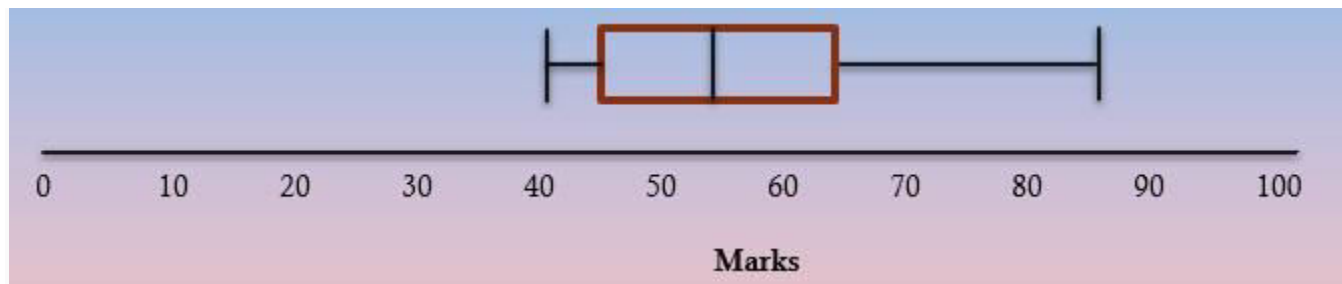
$$Q3 + 1.5 \cdot IQR = 65 + 1.5 \cdot 20 = 65 + 30 = 95$$

So, Any value < 15 and > 95 is an outlier

\Rightarrow 10 and 100 are outliers

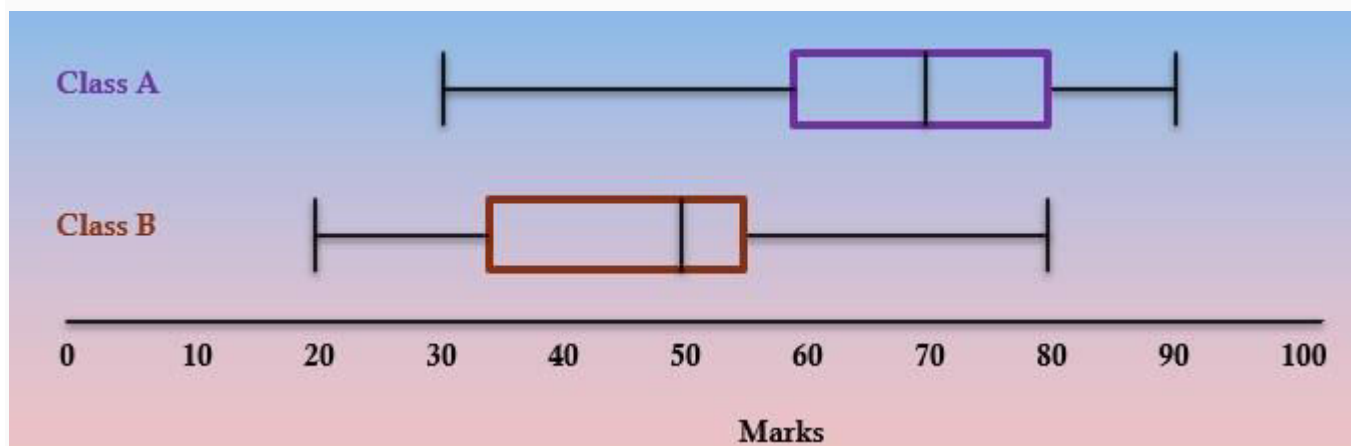
Modified Box and Whisker Plot: If we remove the outlier values from the set and redraw the Box and Whisker plot then the new plot is called as a Modified Box and Whisker Plot.

After removing 10 and 100 following is the Modified Box and Whisker Plot



Comparing two Box and Whisker Plots

Let's say Marks of students from Class A and Class B are given in terms of Box and Whisker Plots as below:



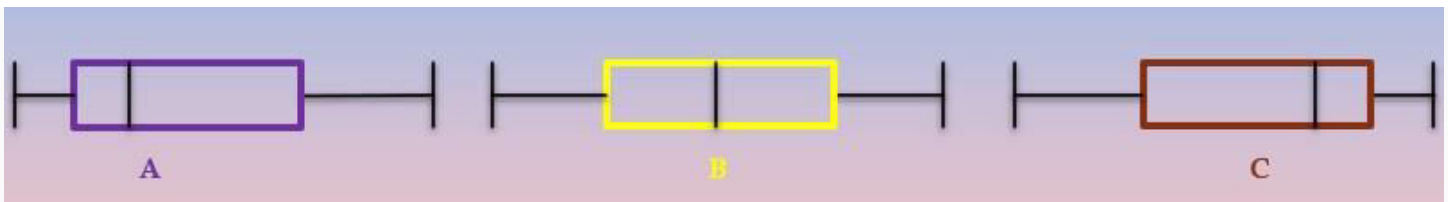
Insights from these Box and Whisker Plots

- ▣ Highest Marks scored (90) by Class A students is more than the Highest Marks scored (80) by Class B students
- ▣ Lowest Marks scored (20) by Class B students is lower than the Lowest Marks scored (30) by Class A students
- ▣ Median Marks scored (70) by Class A students is more than the Median Marks scored (50) by Class B students
- ▣ 75% of the students in class A have score ≥ 60 marks
- ▣ 75% of the students in class B have score ≥ 35 marks

based on Above information we can conclude that **Students from Class A on an average have scored more than Students from Class B.**

Skewness in Box and Whisker Plot

Consider Following Box and Whisker Plots



Box and Whisker Plot A

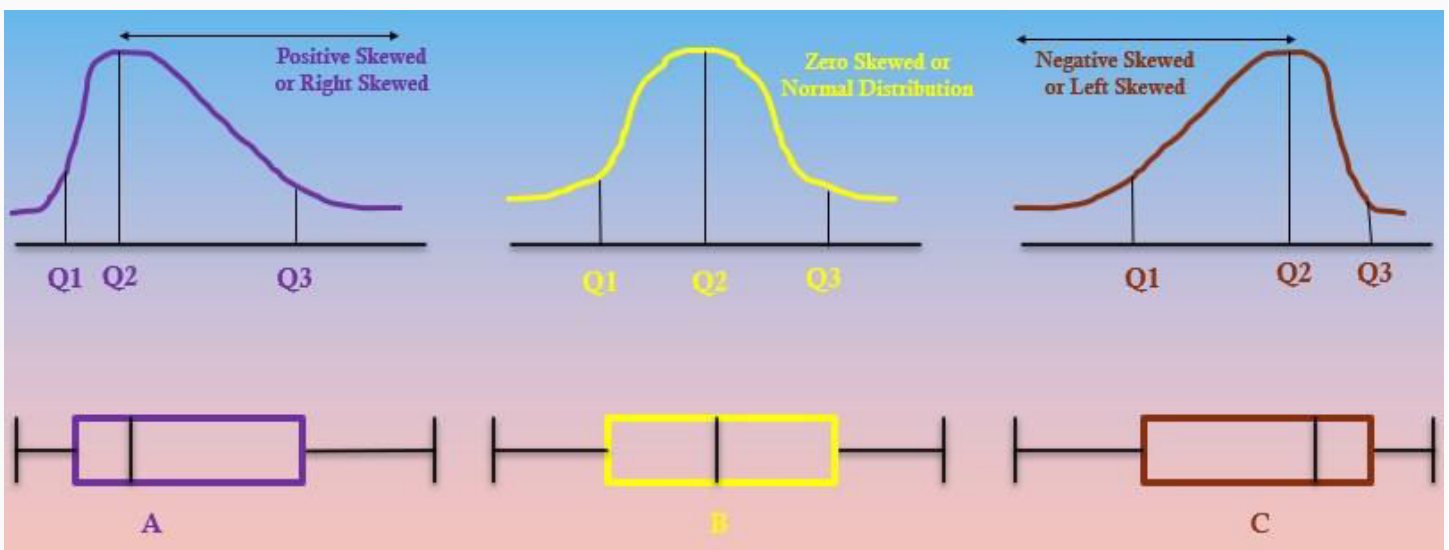
Positive Skewed or Right Skewed: Length of Right Whisker is Larger than the length of Left Whisker, indicating that the graph is skewed towards right.

Box and Whisker Plot B

Zero Skewed or Normal Distribution: Length of Right Whisker is equal to the length of Left Whisker, indicating that the graph is not skewed.

Box and Whisker Plot C

Negative Skewed or Left Skewed: Length of Left Whisker is Larger than the length of Right Whisker, indicating that the graph is skewed towards left.



Hope it helps!
Good Luck!