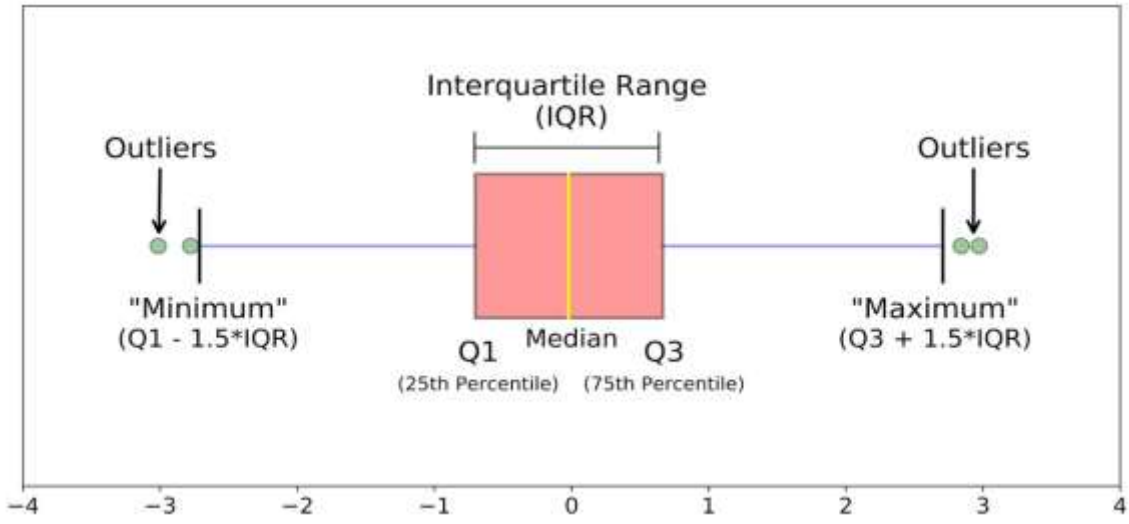
Definition

A box and whisker plot is a graph that exhibits data from a five-number summary, including one of the measures of central tendency. It does not display the distribution as accurately as a stem and leaf plot or histogram does. But, it is principally used to show whether a distribution is skewed or not and if there are potential unusual observations present in the data set, which are also called outliers. Boxplots are also very useful when huge numbers of data collections are involved or compared.

Since the centre, spread and overall range are instantly apparent, using these boxplots the arrangements can be matched easily.

A box and whisker plot is a way of compiling a set of data outlined on an interval scale. It is also used for descriptive data interpretation.

The **box and whisker plot** displays how the data is spread out. In the box and whisker diagram, it has five pieces of information (also called a five-number summary).



Let us list the features of the boxplot:

- The horizontal axis covers all possible data values.
- The box part of a box-and-whisker plot covers the middle 50% of the values in the data set.
- The whiskers each cover 25% of the data values.
    1. The lower whisker covers all the data values from the minimum value up to Q1, that is, the lowest 25% of data values.
    2. The upper whisker covers all the data values between Q3 and the maximum value, that is, the highest 25% of data values.
- The median sits within the box and represents the center of the data. 50% of the data values lie above the median and 50% lie below the median.
- Outliers, or extreme values, in a data set are usually indicated on a box-and-whisker plot by the “star” symbol. If there is one or more outliers in a data set, for the purpose of drawing a box-and-whisker plot, we take the minimum and maximum to be the minimum and maximum values of the data set excluding the outliers.

Why Do We Use Box and Whisker Plot?

Box and Whisker diagrams allow us to read the data very effectively and easily. It summarises the data from multiple sources and displays it in a single graph. It helps us to make an effective decision as it compares the data from different categories.

When to Use Box and Whisker Plot?

The **box and whisker plot** is used if we have multiple datasets from different sources which are related to each other. For example, a test score between classrooms.

How to Draw a Box and Whisker Plot?

The box and whiskers plot can be drawn using five simple steps. To draw a box and whisker diagram, we need to find:

Step 1: The smallest value in the data is called the minimum value.

Step 2: The value below the lower 25% of data contained, called the first quartile.

Step 3: Median value from the given set of data.

Step 4: The value above the lower 25% of data contained, called the third quartile.

Step 5: The largest value in the dataset is called maximum value.

What is Box and Whisker Plot?

Box and whisker plot is one type of graphical representation which shows the five-number summary for the given set of data, such as minimum value, lower quartile, median, upper quartile, maximum value.

Mention the advantages of Box Plot

The advantages of the box and whisker plot is that:

- We can easily identify the data location and data spread.
- It provides the skewness and symmetry of data
- Box and whisker plot show the data outliers.
- A boxplot gives us a very clear visual display of how the data are spread out.

What are the disadvantages of using Box and Whisker Plot?

The disadvantages of a box and whisker plot is that:

- It hides the multimodality and some other characteristics of distributions.
- It confuses the audience sometimes
- Mean cannot be easily located.

What is meant by an outlier in a box plot?

In the box and whisker plot, some data are located outside of the box and the whisker plot, which is numerically different from the rest of the data in the dataset, is called outliers.

How to Draw Box and whisker plot?

- Arrange the data values in the ascending order
- Identify the minimum and maximum values
- Find the median of the data set
- Identify the upper and lower quartile
- Finally, construct the box and whisker plot

Boxplots can be drawn horizontally or vertically.

Horizontal boxplot

Vertical boxplot

Example:

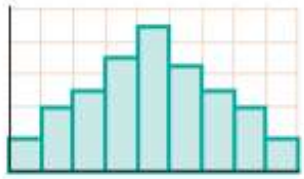
The boxplot below shows the distribution of the part-time weekly earnings of a group of Year-11 students. Write down the range, the median and the interquartile range for these data.

Solution:

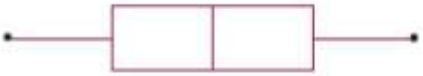
1 Range = Maximum value − Minimum value. The minimum value is 20 and the maximum value is 90.	Range = 90 − 20 = 70
2 The median is located at the bar inside the box.	Median = 50
3 The ends of the box are at 40 and 80. IQR = $Q_3 - Q_1$	$Q_1 = 40$ and $Q_3 = 80$ IQR = 80 − 40 = 40

Skewness:

In the figure below a symmetric distribution is represented in the histogram and in the boxplot. The characteristics of this boxplot are that the whiskers are about the same length and the median is located about halfway along the box.

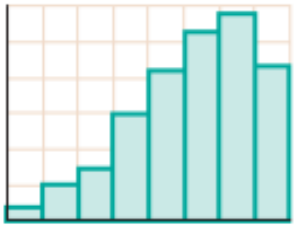


Symmetric histogram

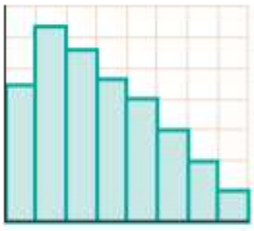


Symmetric boxplot

The figure below shows a negatively skewed distribution. In such a distribution, the data peak to the right on the histogram and trail off to the left. In corresponding fashion on the boxplot, the bunching of the data to the right means that the left-hand whisker is longer and the right-hand whisker is shorter; that is, the lower 25% of data are sparse and spread out whereas the top 25% of data are bunched up. The median occurs further towards the right end of the box.



In the figure below we have a positively skewed distribution. In such a distribution, the data peak to the left on the histogram and trail off to the right. In corresponding fashion on the boxplot, the bunching of the data that the left-hand whisker is shorter and the right-hand whisker is longer; that is, the upper 25% of data are sparse and spread out whereas the lower 25% of data are bunched up. The median occurs further towards the left end of the box.

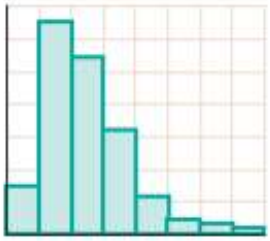


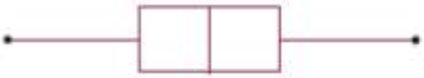
Positively skewed histogram

Positively skewed boxplot

Example:

Explain whether or not the histogram and the boxplot shown below could represent the same data.





THINK

The histogram shows a distribution which is positively skewed.
The boxplot shows a distribution which is approximately symmetric.

WRITE

The histogram and the boxplot could not represent the same data since the histogram shows a distribution that is positively skewed and the boxplot shows a distribution that is approximately symmetric.

Example:

The results out of 20 of oral tests in a Year-12 Indonesian class are:

15 12 17 8 13 18 14 16 17 13 11 12

Display these data using a boxplot.

THINK

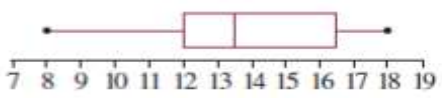
1. Find the lowest and highest scores, Q_1, the median (Q_2) and Q_3 by first ordering the data.

WRITE

8 11 12 12 13 13 14 15 16 17 17 18
The median score is 13.5.
The lower half of the scores are
8 11 12 12 13 13.
So, $Q_1 = 12$
The upper half of the scores are
14 15 16 17 17 18.
So, $Q_3 = 16.5$

The lowest score is 8.
The highest score is 18.

2. Using these 5 key scores, draw the boxplot.



Parallel Boxplot

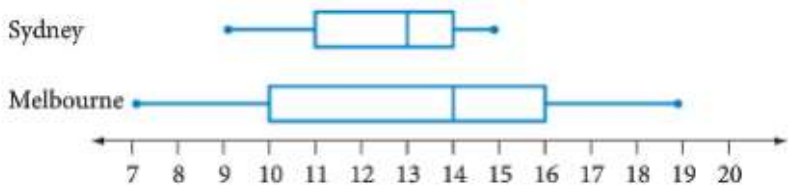
Example:

The following two five number summaries for Sydney and Melbourne describe the number of rainy days per month over two years.

Sydney: 9, 11, 13, 14, 15

Melbourne: 7, 10, 14, 16, 19

The boxplots are placed on a common scale.



Median for Sydney = 13 Median for Melbourne = 14

Interquartile range for Sydney = 14 – 11 = 3 Interquartile range for Melbourne = 16 – 10 = 6

Comparison: Melbourne has more rainy days per month. Its median is higher and half its scores are above 14 compared to one quarter of the scores for Sydney.
Sydney has a more consistent pattern of rainy days because its range and interquartile range are smaller than Melbourne's.

Practice the Outlier Problems