



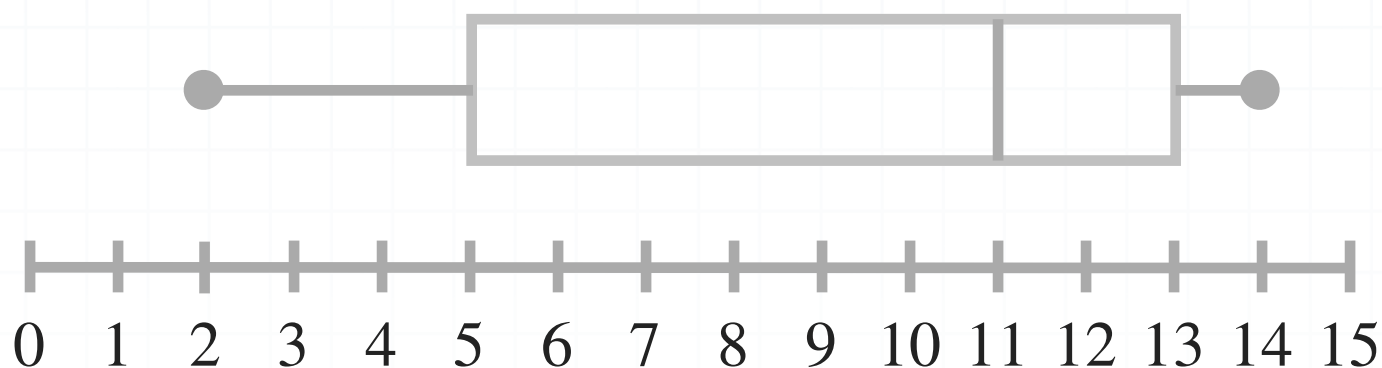
# Probability & Statistics Notes

---

# Box-and-whisker plots

**Box-and-whisker plots** (also called box plots) are a great way to represent a data set when you want to show the median and spread of the data at the same time.

In general, a box-and-whisker plot might look like this:



The big box in the center is the **box**, and the little lines extending out from the sides are the **whiskers**.

The great thing about a box plot is that we know the median, range, upper and lower bounds just by looking at it. The interquartile range is also just a simple calculation.

The vertical line in the center of the box is the median of the data set, so the median of the data set represented in the plot above is 11.

The dot at the end of the left whisker is the minimum of the data set, and the dot at the end of the right whisker is the maximum of the data set. So in this plot, we can say that the minimum is 2, that the maximum is 14, and so we know right away that the range of the data is  $14 - 2 = 12$ .



The IQR is given by the ends of the box. Since the box above extends from 5 to 13, the IQR is  $13 - 5 = 8$ .

The box-and-whisker plot also shows us where each quartile of the data is located. A **quartile** is one quarter of the data points, which means that the bottom 25 % of data points in the set are found in the **first quartile**, the next 25 % of data points in the set are found in the **second quartile**, the next 25 % of data points in the set are found in the **third quartile**, and the top 25 % of the data points are in the **fourth quartile**. We also call these quartiles  $Q_1$ ,  $Q_2$ ,  $Q_3$  and  $Q_4$ .

In a box-and-whisker plot, the left whisker represents  $Q_1$ , the box to the left of the median represents  $Q_2$ , the box to the right of the median represents  $Q_3$ , and the right whisker represents  $Q_4$ . Based on the box plot above,

- 25 % of the data points have a value between 2 and 5
- 25 % of the data points have a value between 5 and 11
- 25 % of the data points have a value between 11 and 13
- 25 % of the data points have a value between 13 and 14

It's also correct to describe the quartiles this way:

- 5 is the first quartile
- 11 is the second quartile
- 13 is the third quartile



- 14 is the fourth quartile

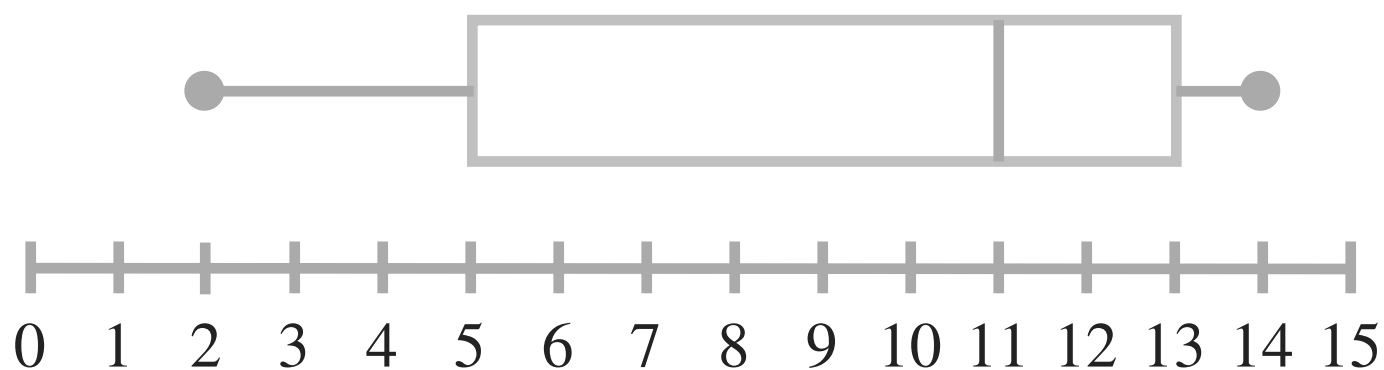
In a box-and-whisker plot, the middle 50 % of the data is represented inside the box, the lowest 25 % in the whisker on the left, and the highest 25 % in the whisker on the right.

As far as our quartiles, this means that the first quartile is represented by the whisker on the left, the second quartile is represented by the part of the box to the left of the median, the third quartile is represented by the part of the box to the right of the median, and the fourth quartile is represented by the whisker on the right.

## Five-number summary

The **five-number summary**, also called the five-figure summary, for any set of data will include the minimum and maximum values, the median, and  $Q_1$  and  $Q_3$  for the data set. We usually give the five-number summary in a table, and we can easily gather all of this information from a box-plot.

The five-number summary for the box plot



is



Min	Q1	Median	Q3	Max
2	5	11	13	14



