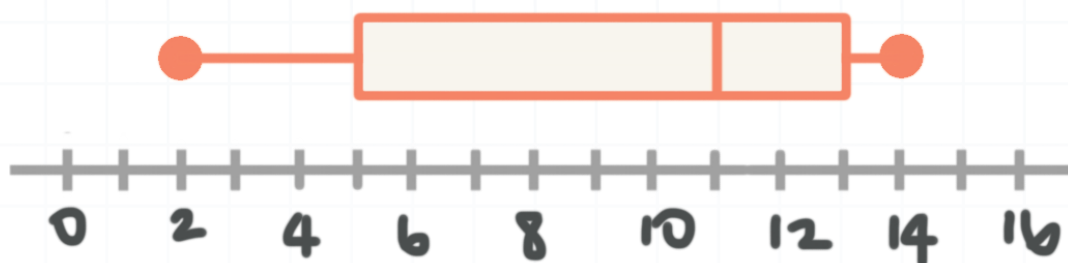


Box-and-whisker plots

Box-and-whisker plots (or **box plots**) are a great way to represent a data set when we 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 rectangle 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 inside the box is the median of the data set, so the median of the data set represented in the plot above is 11.

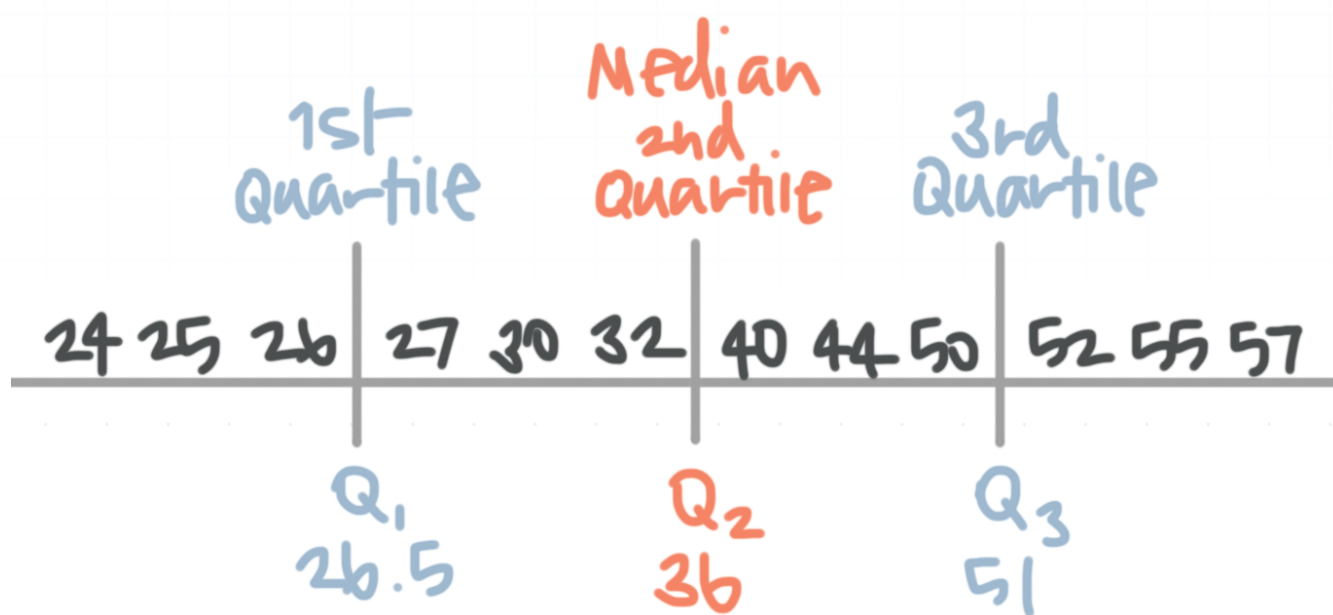
The dot at the end of the left whisker is the minimum of the data set, and the dot at the end of the right whisker is the maximum of the data set. So in this plot, we can say that the minimum is 2, that the maximum is 14, and so we know right away that the range of the data is $14 - 2 = 12$.

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

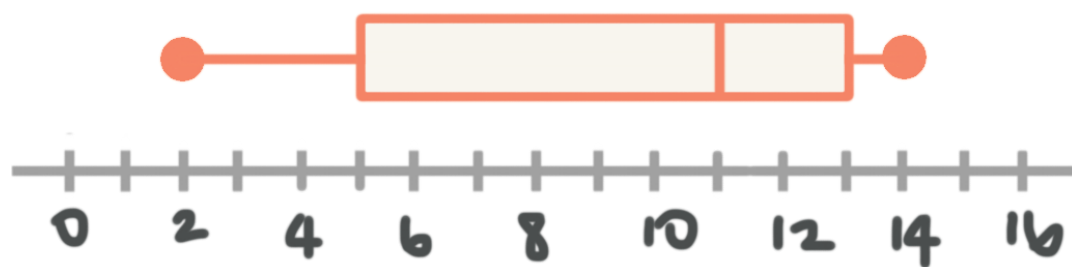


The box-and-whisker plot also shows us where each quartile of the data is located. A **quartile** is a number that divides the data set into quarters. The first quartile, Q_1 , separates the lowest 25 % of data points from the second 25 %. The second quartile, Q_2 , is the median, and it separates the data set into halves. The third quartile, Q_3 , separates the third 25 % of data points from the upper 25 % of data points.

In the ordered data set below, which has 12 values between 24 and 57, there must be $12 \text{ values} / 4 \text{ quartiles} = 3 \text{ values per quartile}$, so we use lines to split the data set into four parts, each with three values. Then we find the mean of the two values on either side of each line to get $Q_1 = (26 + 27)/2 = 26.5$, $Q_2 = (32 + 40)/2 = 36$, and $Q_3 = (50 + 52)/2 = 51$.



In a box-and-whisker plot like the one we've been working with,



the left end of the box represents Q_1 , the median represents Q_2 , and the right end of the box represents Q_3 . So from the box plot we know

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

And therefore that

- 25 % of the data points lie between 2 and 5
- 25 % of the data points lie between 5 and 11
- 25 % of the data points lie between 11 and 13
- 25 % of the data points lie between 13 and 14

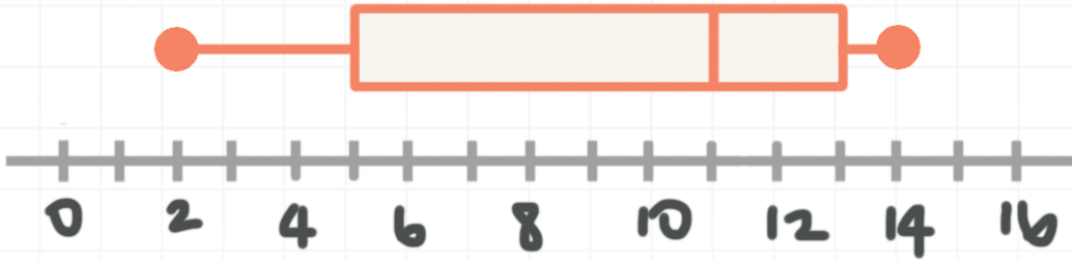
In a box-and-whisker plot, the middle 50 % of the data is represented inside the box, the lowest 25 % in the whisker on the left, and the highest 25 % in the whisker on the right.

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	Q_1	Median	Q_3	Max
2	5	11	13	14

