

Recipe 2: Five Number Summary

The five number summary of a set of data consists of the maximum value, the minimum value, the median, and the *lower quartile* (Q_1) and the *upper quartile* (Q_3). The lower quartile is the value of the variable chosen so that 25% of your results are less than the lower quartile, and 75% of the results are more than the lower quartile. The upper quartile reverses these percentages. A couple of examples will make things clearer.

The five number summary contains most of the information that the original data contained: a typical value of the variable, the spread of the data and the extremes of the data range, and first order information about the *shape* of the distribution of the variable (see box and whisker plot).

Example (even number of data values)

Suppose you found the heights of fourteen children in centimetres;

168, 156, 153, 155, 155, 159, 163, 157, 172, 144, 163, 153, 152, 164

The **maximum height** is 172 cm and the **minimum height** is 144 cm by inspection. To find the **median**, we put the data values in order of size (see recipe 1).

144, 152, 153, 153, 155, 155, 156, 157, 159, 163, 163, 164, 168, 172

$N = 14$ so $\frac{N+1}{2} = \frac{15}{2} = 7.5$ so we need to take the mean of the 7th and 8th data items, which gives
 $\frac{156+157}{2} = \frac{313}{2} = 156.5$ for the median.

The median is the value that divides the data set into two halves. You can think of the lower quartile as the median of the lower half, and the upper quartile as the median of the upper half. There are seven numbers in each half of the data set, so $\frac{N+1}{2} = \frac{8}{2} = 4$ so we pick the 4th number in each case;

144, 152, 153, [**153**] 155, 155, 156 | 157, 159, 163, [**163**] 164, 168, 172

So the lower quartile Q_1 is 153 cm and the upper quartile Q_3 is 163 cm. You can organise the five number summary into a table

Number	Value
Min	144
Q_1	153
Median	156.5
Q_3	163
Max	172

Another example (odd number of data values)

Suppose you have an odd number of data items and wish to find the five number summary. If N is the (odd) number of data values, then $\frac{N+1}{2}$ is a whole number, and you can pick a median without having to add up and divide by two.

You can find the lower and upper quartiles by finding the median of the lower half and the median of the upper half of the values. Some authors include the median in both halves of the data, and some authors exclude the median from both halves of the data! In this course, we will adapt the convention of *including* the median in both halves of the data. A concrete example might help.

The supine systolic blood pressure (SSBP) of 19 diabetic patients was measured to provide a baseline for a drug trial. The results in mmHg are shown below.

124 120 115 134 131 119 124 127 113 132 129 124 112 124
144 134 119 123 122

See if you can follow the steps below and reproduce my answers...

Step 1: Maximum SSBP is 144 mmHg and minimum SSBP is 112 mmHg.

Step 2: place the data in order of size. $N = 19$, so median is $\frac{N+1}{2} = 10$ the 10th number in the list, which is 124 mmHg.

Step 3: Using our convention, we divide the data set into a lower half from 112 to 124 mmHg with 10 results including the median, and an upper half from 124 mmHg to 144 mmHg again including the median;

Lower half 112, 113, 115, 119, 119, 120, 122, 123, 124, 124
Upper half 124, 124, 124, 127, 129, 131, 132, 134, 134, 144

For each half, $N=10$ and $\frac{N+1}{2} = 5.5$, so the median value is the mean of the 5th and 6th number in each of the lists. For the lower half, I find 119.5 mmHg and for the upper half I find 130 mmHg.

Step 4: We can now complete the five number summary for this data;

Number	Value / mmHg
Min	112
Q1	119.5
Median	124
Q3	130
Max	144

Your turn

Find the five number summary of the following set of 19 supine diastolic blood pressure (SDBP) results for the same patients as above, again in mmHg...

78, 70, 68, 91, 93, 71, 76, 82, 72, 88, 73, 76, 70, 80
88, 80, 77, 58, 79