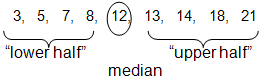
**First and Third quartiles of the data set**

**Definitions:**

* The **lower half** of a data set is the set of all values that are to the left of the median value when the data has been put into increasing order.
* The **upper half** of a data set is the set of all values that are to the right of the median value when the data has been put into increasing order.
* The **first quartile,** denoted by ***Q*1** , is the median of the *lower half* of the data set. This means that about 25% of the numbers in the data set lie below *Q*1 and about 75% lie above *Q*1 .
* The **third quartile,** denoted by ***Q*3** , is the median of the *upper half* of the data set. This means that about 75% of the numbers in the data set lie below *Q*3 and about 25% lie above *Q*3

**Example 1:** Find the first and third quartiles of the data set {3, 7, 8, 5, 12, 14, 21, 13, 18}.

First, we write data in increasing order: 3, 5, 7, 8, 12, 13, 14, 18, 21.



As on the previous page, the median is 12.

Therefore, the lower half of the data is: {3, 5, 7, 8}.

The first quartile, *Q*1, is the median of {3, 5, 7, 8}.

Since there is an even number of values, we need the mean of the middle two values to find the first quartile:

http://web.mnstate.edu/peil/MDEV102/U4/S36/ada-equation.gifhttp://web.mnstate.edu/peil/MDEV102/U4/S36/lessonimages/equation_image4.gif .

Similarly, the upper half of the data is: {13, 14, 18, 21}, so

* http://web.mnstate.edu/peil/MDEV102/U4/S36/lessonimages/equation_image5.gif .

**How to Find a Five-Number Summary: Steps**

* **Step 1:***Put your numbers in ascending order* (from smallest to largest). For this particular data set, the order is:  
  Example: 1, 2, 5, 6, 7, 9, 12, 15, 18, 19, 27.
* **Step 2:***Find the minimum and maximum* for your data set. Now that your numbers are in order, this should be easy to spot.  
  In the example in step 1, the minimum (the smallest number) is 1 and the maximum (the largest number) is 27.
* **Step 3:***Find the median*. The median is the middle number. If you aren’t sure how to find the median, see: How to find the mean mode and median.
* **Step 4:***Place parentheses around the numbers****above and below****the median.*  
  (This is not *technically* necessary, but it makes Q1 and Q3 easier to find).  
  (1, 2, 5, 6, 7), 9, (12, 15, 18, 19, 27).
* **Step 5:***Find Q1 and Q3*. Q1 can be thought of as a median in the lower half of the data, and Q3 can be thought of as a median for the upper half of data.  
  (1, 2, **5**, 6, 7)**, 9**, ( 12, 15,**18**,19,27).
* **Step 6:***Write down your summary found in the above steps*.  
  **minimum = 1, Q1 = 5, median = 9, Q3 = 18, and maximum = 27.**

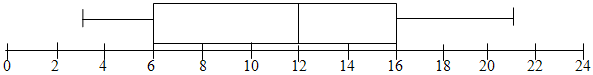
**Box-and-Whisker plot**

**Example 1:** Draw a box-and-whisker plot for the data set {3, 7, 8, 5, 12, 14, 21, 13, 18}.

From our Example 1 on the previous page, we had the five-number summary:

{3,5,7,8,**12**,13,14,18,21}

Minimum: 3, *Q*1 : 6, Median: 12, *Q*3 : 16, and Maximum: 21.



Notice that in any box-and-whisker plot, the left-side whisker represents where we find approximately the lowest 25% of the data and the right-side whisker represents where we find approximately the highest 25% of the data. The box part represents the interquartile range and represents approximately the middle 50% of all the data. The data is divided into four regions, which each represent approximately 25% of the data. This gives us a nice visual representation of how the data is spread out across the range.

**Example 2:**

Find Q1, Q2 , and Q3 for the following data set, and draw a box-and-whisker plot.

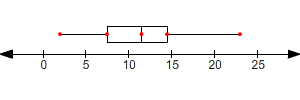
{2,6,7,8,8,11,12,13,14,15,22,23}

There are 12 data points. The middle two are 11 and 12. So the median, Q2, is 11.5.

The "lower half" of the data set is the set {2,6,7,8,8,11}. The median here is 7.5. So Q1=7.5.

The "upper half" of the data set is the set {12,13,14,15,22,23} . The median here is 14.5. So Q3=14.5.

A box-and-whisker plot displays the values Q1, Q2, and Q3, along with the extreme values of the data set (2 and 23, in this case):



A box & whisker plot shows a "box" with left edge at Q1, right edge at Q3 , the "middle" of the box at Q2 (the median) and the maximum and minimum as "whiskers".

Note that the plot divides the data into 4 equal parts. The left whisker represents the bottom 25% of the data, the left half of the box represents the second 25% , the right half of the box represents the third 25% , and the right whisker represents the top 25% .

Example 3

## Outliers

If a data value is very far away from the quartiles (either much less than Q1 or much greater than Q3), it is sometimes designated an outlier. Instead of being shown using the whiskers of the box-and-whisker plot, outliers are usually shown as separately plotted points.

The standard definition for an outlier is a number which is less than Q1 or greater than Q3 by more than 1.5 times the interquartile range (IQR=Q3−Q1).

That is, an outlier is any number less than Q1−(1.5×IQR) or greater than Q3+(1.5×IQR).

**Example 3:**

Find Q1, Q2, and Q3 for the following data set. Identify any outliers, and draw a box-and-whisker plot.

{5,40,42,46,48,49,50,50,52,53,55,56,58,75,102}

{5,40,42,46,48,49,50,**50**,52,53,55,56,58,75,102}

{5,40,42,**46**,48,49,50}**50**{52,53,55,**56**,58,75,102}

There are 15 values, arranged in increasing order. So, Q2 is the 8th data point, 50.

Q1 is the 4th data point, 46, and Q3 is the 12th data point, 56.

The interquartile range **IQR is Q3−Q1 or 56−46=10.**

Now we need to find whether there are values less than Q1−(1.5×IQR)) or greater than Q3+(1.5×IQR).

Q1−(1.5×IQR) =46−15=31

Q3+(1.5×IQR) =56+15=71

**Since 5 is less than 31 and 75 and 102 are greater than 71, there are 3 outliers.**