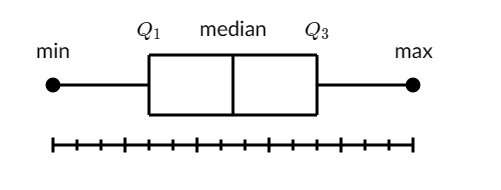
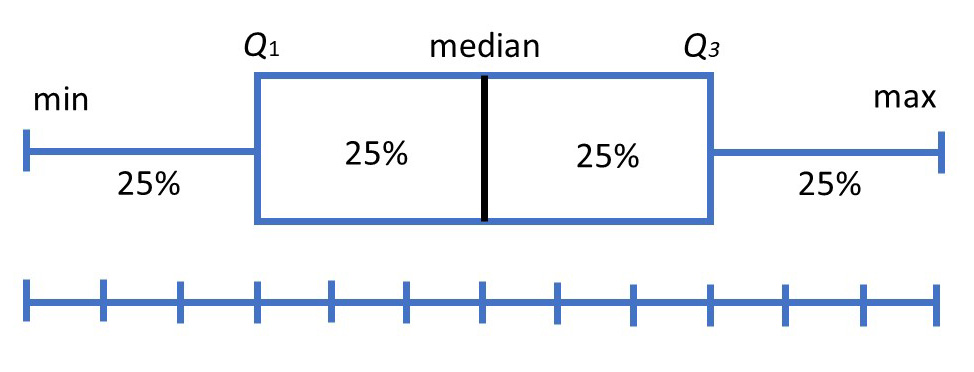
**What is a box and whisker plot?**

A box and whisker plot—also called a box plot—displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum.

In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum.





**Applications**

It is used to know:

* The outliers and their values
* Symmetry of Data
* Tight grouping of data
* Data skewness – if, in which direction and how

**Example: Finding the five-number summary**

A sample of 10 boxes of raisins has these weights (in grams):

25, 28, 29, 29, 30, 34, 35, 35, 37, 38

**Make a box plot of the data.**

**Step 1:** Order the data from smallest to largest.

Our data is already in order

25, 28, 29, 29, 30, 34, 35, 35, 37, 38

**Step 2:** Find the median.

The median is the mean of the middle two numbers:

25, 28, 29, 29, 30, 34, 35, 35, 37, 38

(30 + 34) /2 = 32

The median is 32.

**Step 3:** Find the quartiles.

The first quartile is the median of the data points to the *left* of the median.

25, 28, **29**, 29, 30

*Q*1​=29

The third quartile is the median of the data points to the right of the median

34, 35, **35**, 37, 38

*Q*3​=35

**Step 4:** Complete the five-number summary by finding the min and the max.

The min is the smallest data point, which is 25.

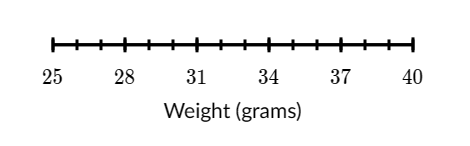
The max is the largest data point, which is 38.

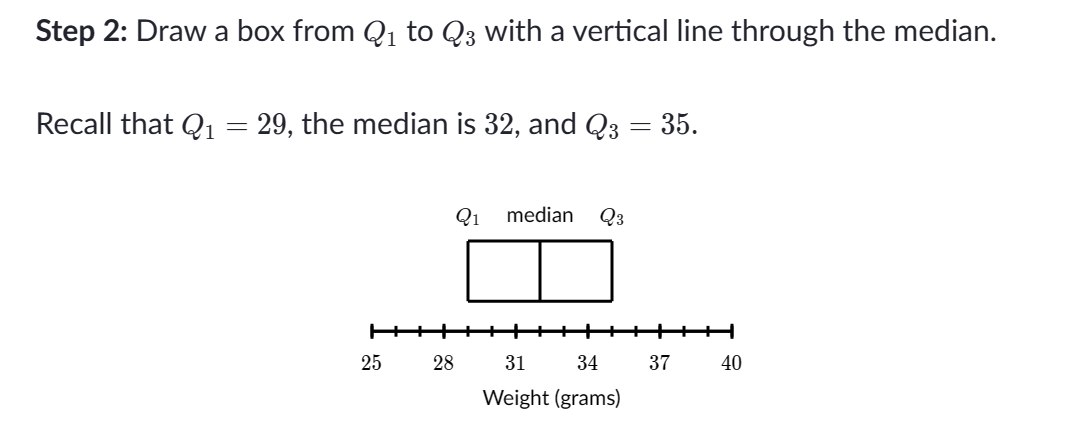
The five-number summary is 25 29 32 35 38

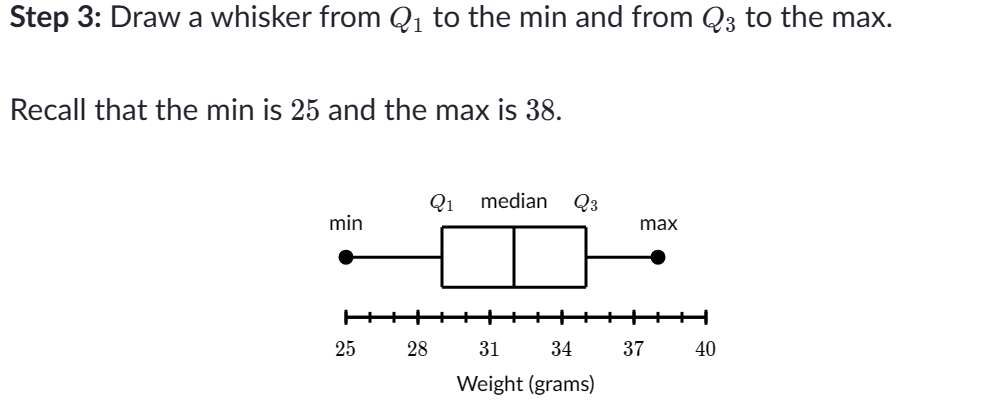
**Making a box plot**

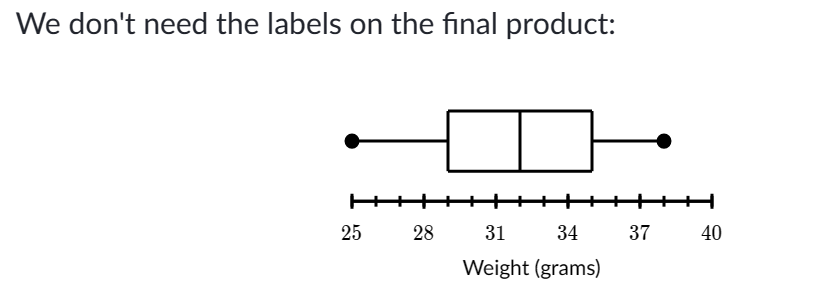
Let's make a box plot for the same dataset from above.

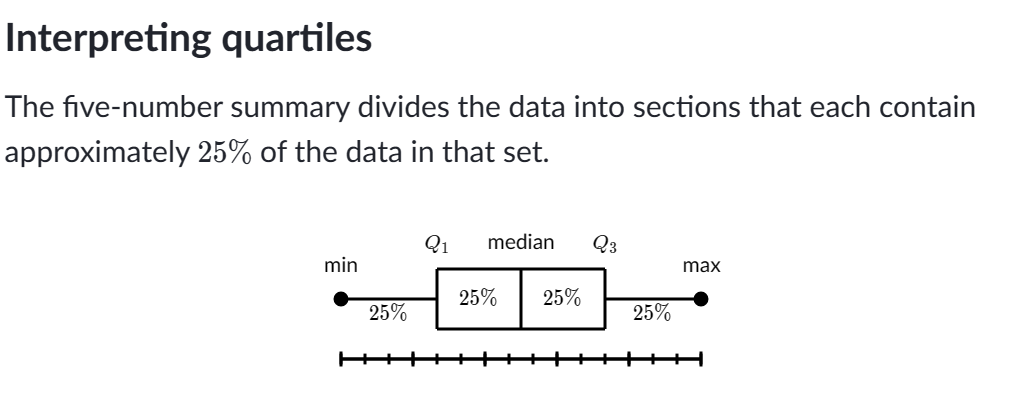
**Step 1:** Scale and label an axis that fits the five-number summary.

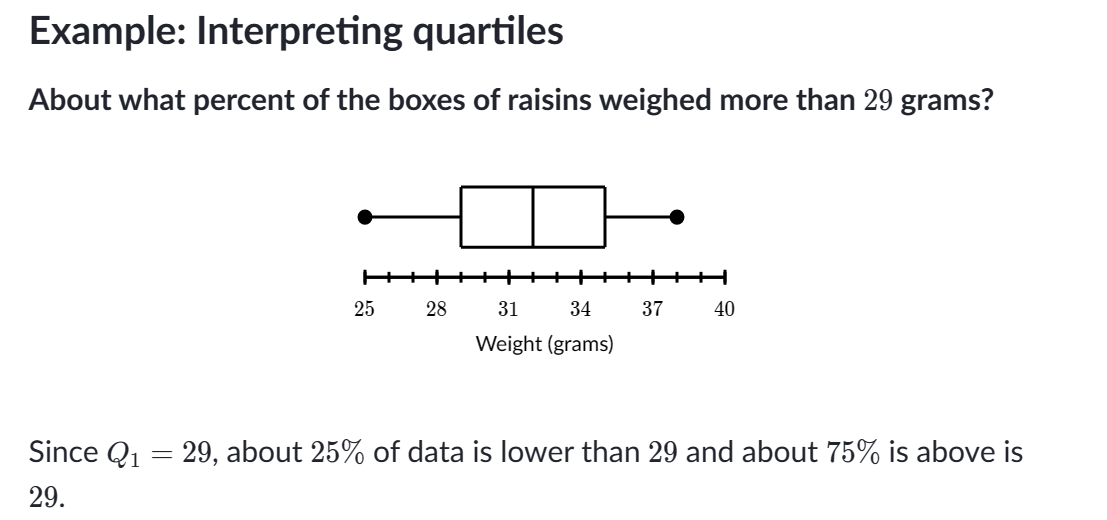


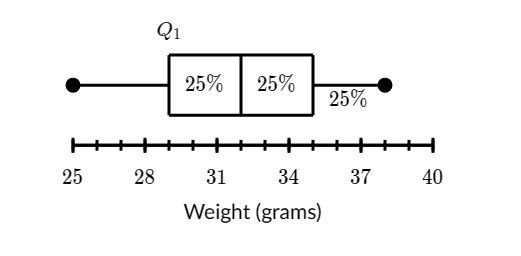


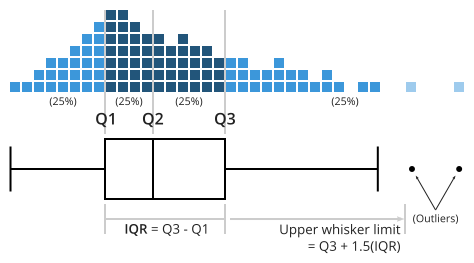


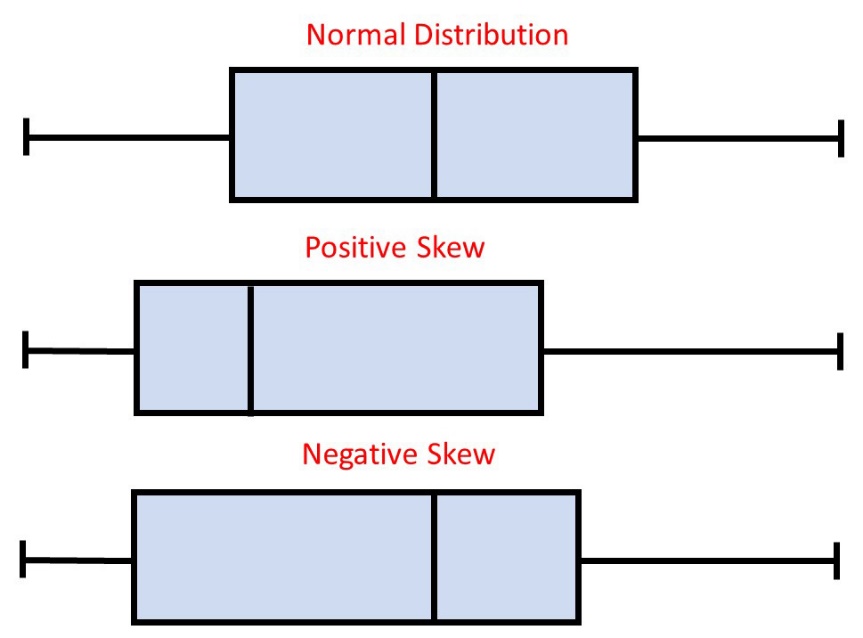


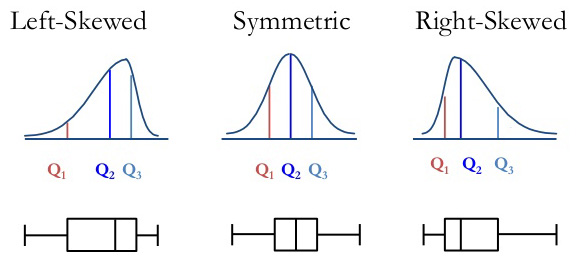












**Example 1:**a simple box and whisker plot

Suppose you have the math test results for a class of 15 students. Here are the results:

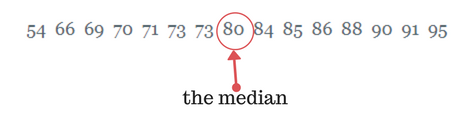
91  95  54  69  80  85  88  73  71  70  66  90  86  84  73

It is hard to say what is the middle point (the median) because the value points are not ordered.

**Step 1**: Order the data points from least to greatest.

54  66  69  70  71  73  73  80  84  85  86  88  90  91  95

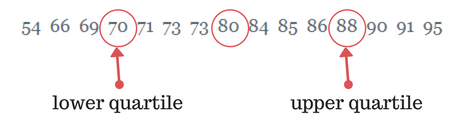
**Step 2**: Find the median of the data:



This is an odd set of data – you have 15 data points. It means the middle point is 80 as there are 7 data points above it and 7 numbers below.

More on how to calculate median, you can see on our post [descriptive statistics examples](http://intellspot.com/descriptive-statistics-examples/).

**Step 3:** Find the middle points of the two halves divided by the median (find the upper and lower quartiles).



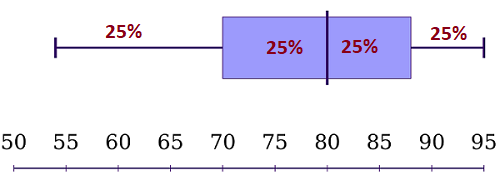
**Step 4:**Find the extreme values.

This is the easiest part. You need to find the largest and smallest data values.

Extreme values = 54 and 95.

So, we can determine that the five-number summary for the class of students is 54, 70, 80, 88, 95.

Now we are absolutely ready to draw our box and whisker plot.



As you see, the plot is divided into four groups: a lower whisker, a lower box half, an upper box half, and an upper whisker. Each of those groups shows 25% of the data because we have an equal amount of data in each group.

**Interpreting the box and whisker plot results:**

The box and whisker plot shows that 50% of the students have scores between 70 and 88 points.

In addition, 75% scored lower than 88 points, and 50% have test results above 80. So, if you have test results somewhere in the lower whisker, you may need to study more.

It was among the simplest box and whisker plot examples just to illustrate what the plot shows. Let’s deep further and see double box and whisker plot examples that help you to compare 2 data sets.

**Example 2**: comparative double box and whisker plot

Suppose an IT company has two stores that sell computers. The company recorded the number of sales each store made each month. In the past 12 months, we have the following numbers of sold computers:

**Store 1**:

350, 460, 20, 160, 580, 250, 210, 120, 200, 510, 290, 380.

**Store 2:**

520, 180, 260, 380, 80, 500, 630, 420, 210, 70, 440, 140.

In order to compare the two stores sales performance, we will make two box and whisker plots, one for Store 1 and one for Store 2.

**Store 1:**

First, we put the data points in ascending order.

20, 120, 160, 200, 210, 250, 290, 350, 380, 460, 510 580.

Now, we need to find the median. However, this is **an even** set of data. There isn’t only one middle point. The middle in our case belongs to sixth + seventh data points e.g. 250 and 290.

And the formula for the median in an even data set is:

(The sum of the two middle numbers) / 2

**The median** is (250 + 290) / 2 = 270

Now let’s see what happens with the lower and upper quartiles in an even data set:

There are six numbers below the median, namely: 20, 120, 160, 200, 210, 250

**Lower quartile** is the median of these six items, so  
= (third + fourth data point) / 2  
= (160 + 200) / 2  
= 180

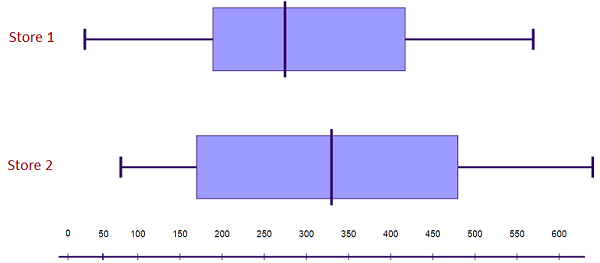
There are also six numbers above the median, namely: 290, 350, 380, 460, 510, 580

**Upper quartile** is the median of these six data points.  
= (third + fourth data points) / 2  
= 420

Finally, the five-number summary for Store 1’s sales is 20, 180, 270, 420, 580

Using the same calculations, we can find that the five-number summary for Store 2 is 70, 160, 320, 470, 630

Now, we are ready to draw our comparative double box and whisker plot example:



**Interpreting the results:**

Store 2’s highest and lowest sales are both higher than Store 1’s relevant sales.

In addition, Store 2’s median sales value is higher than Store 1’s. Also, Store 2’s interquartile range is larger.

These results tell us that Store 2 consistently sells more computers than Store 1.