

STATISTICS AND GRAPHS

Unit Overview

This unit is about statistics and graphs that help organize data. Graphs are important because they can be used to present large amounts of information in ways that are easily understood. Graphs are used to summarize, analyze, and interpret data in many different types of settings and work places.

Measures of Central Tendency

Measures of central tendency are used in statistics to describe data. The mean, median, mode, and range are the four measures of central tendency we will be discussing.

Mean: the average of a set of data (found by adding all the data together and dividing by the total number of elements).

Median: the middle number in the set when the elements are placed in order from least to greatest or greatest to least (if two numbers appear in the middle, those two numbers are added together and averaged).

Mode: the number that appears most often (there may be no mode, one mode, or more than one mode).

Range: the difference between the greatest and least values in the set.

Example #1: The Rockets scored the following points in their first 10 basketball games. Find the mean, median, mode and range.

32, 27, 38, 43, 29, 51, 32, 34, 49, 41

$$\text{Mean: } \frac{32 + 27 + 38 + 43 + 29 + 51 + 32 + 34 + 49 + 41}{10} = \frac{376}{10} = 37.6$$

Median: Arrange the numbers from least to greatest.

27, 29, 32, 32, 34, 38, 41, 43, 49, 51

Determine the middle number (since there are 2 middle numbers, add them together and divide by 2).

$$\frac{34 + 38}{2} = \frac{72}{2} = 36$$

Mode: Use the arrangement of the numbers from the median to determine the mode.

27, 29, 32, 32, 34, 38, 41, 43, 49, 51

The number that occurs most often is 32.

Range: Subtract the smallest value from the largest value.

$$51 - 27 = 24$$

$$\text{range} = 24$$

Example #2: On the first test, students scored 45, 65, 70, 72, 75, 80, 90, 94, and 100. Find the mean, median, mode, and range.

$$\text{Mean: } \frac{45 + 65 + 70 + 72 + 75 + 80 + 90 + 94 + 100}{9} = \frac{694}{9} = 76.8$$

This value does come out to 76.7 repeating, but we will round it to the tenths place. Just follow directions for rounding if you get an answer like this.

Median: Arrange the numbers from least to greatest.

45, 65, 70, 72, 75, 80, 90, 94, 100

Notice that in this set, there is an exact number in the middle. Therefore, we will not average anything. **75 is the median.**

Mode: Use the arrangement of the numbers from the median to determine the mode.

45, 65, 70, 72, 75, 80, 90, 94, 100

Since all numbers appear exactly once, we say there is NO mode.

You can have multiple modes if multiple numbers appear the most number of times. For example, if you have 4, 4, 4, 5, 5, 5, 6, 6, 6, 7, 8, 8, 9, 10, 10, what is the mode? Notice that 4, 5, and 6 all show up 3 times. This is the most number of times that a number shows up. Therefore the modes are 4, 5, and 6. 8 and 10 are NOT modes because they only show up twice.

****You can have no mode, one mode, or multiple modes.****

Range: Subtract the smallest value from the largest value.

$$100 - 45 = 55$$

Stop! Go to Questions #1-8 about this section, then return to continue on to the next section.

Dot Plots

A dot plot shows the frequency of data. In other words, it shows how often data occurs. It helps to first organize the data into a frequency table.

Watch this video explaining frequency tables and dot plots:

Click on the link to watch the video "[Frequency tables and dot plots](#)" or click on the video.



Example #1: Make a dot plot for the following data.

Temperatures for Cincinnati, OH in July 2015.

80, 77, 76, 84, 80, 85, 83, 73, 82, 77, 84, 85, 84, 89, 83, 84, 90, 91, 89, 87, 85, 81, 84, 87, 87, 86, 86, 90, 91, 85, 87

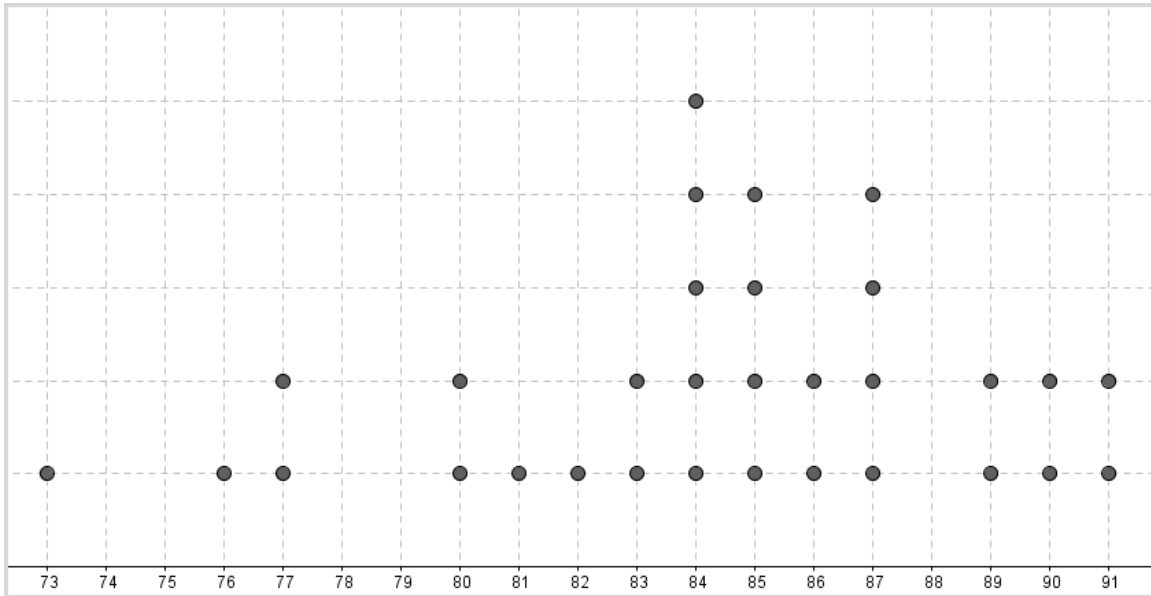
First let's make a frequency table. We will enter all possible values on the left and the number of times they occur on the right. It helps to go ahead and list them from least to greatest. Since 73 is the lowest temperature, we will list it first. We will end with 91.

| Temperature | # of Days with that temperature |
|-------------|---------------------------------|
| 73° | 1 |
| 76 | 1 |
| 77 | 2 |
| 80 | 2 |
| 81 | 1 |
| 82 | 1 |
| 83 | 2 |
| 84 | 5 |
| 85 | 4 |
| 86 | 2 |
| 87 | 4 |
| 89 | 2 |
| 90 | 2 |
| 91 | 2 |

One way to check your frequency table is to add up your right column for # of days and see if the total matches the total number of data. In this case, it should add up to 31 because there are 31 numbers in the data set. By organizing your data this way, it makes it easier to then find measures of central tendency like median and range.

To make the dot plot we now just plot the temperatures on a horizontal number line. We then will make a dot for each piece of data. The number of dots will correspond to the number of days for each temperature. For example, there were 4 days of temperature 87, so at 87, make 4 dots.

The dot plot should look like this:



Stop! Go to Question #9 about this section, then return to continue on to the next section.

Histograms

A histogram is very much like a bar graph except there are **no spaces between the bars**.

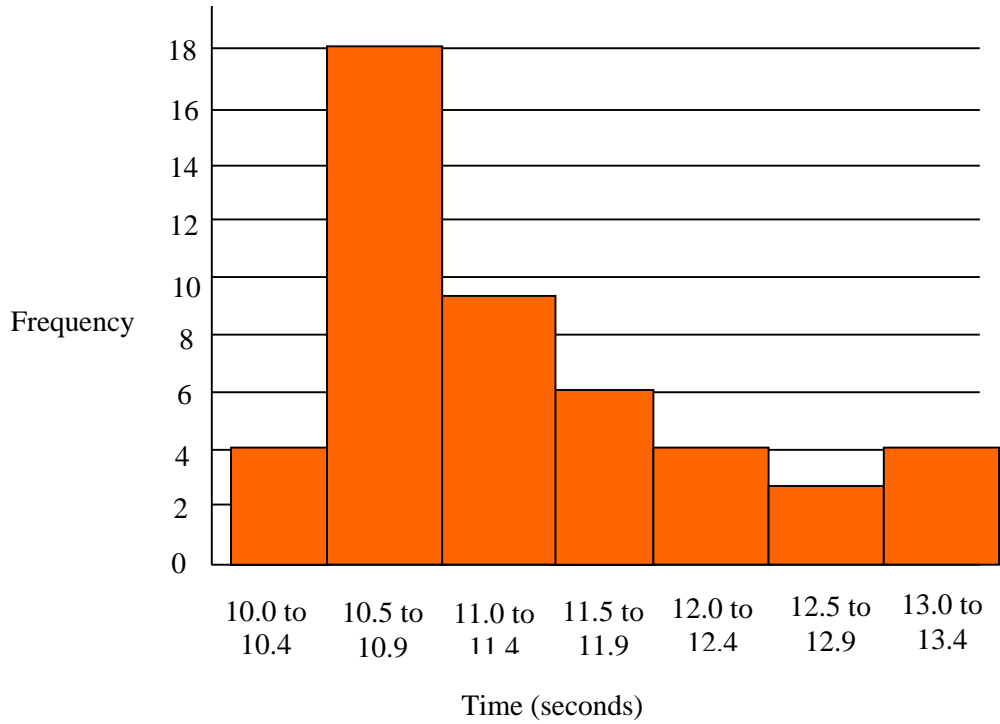
The chart below gives the qualifying times of 48 runners in the 100 meter dash.

We can construct a histogram using the data from the table.

| | | | |
|------|------|------|------|
| 10.5 | 12.3 | 11.6 | 13.0 |
| 10.7 | 10.5 | 10.7 | 11.2 |
| 12.3 | 10.9 | 13.1 | 12.2 |
| 10.3 | 10.8 | 10.2 | 11.5 |
| 11.6 | 10.6 | 10.8 | 11.2 |
| 13.4 | 11.4 | 11.0 | 12.9 |
| 10.9 | 12.7 | 11.2 | 11.7 |
| 10.5 | 10.6 | 11.3 | 11.5 |
| 10.8 | 10.6 | 11.2 | 12.2 |
| 10.7 | 10.4 | 12.6 | 10.0 |
| 11.4 | 13.3 | 10.8 | 10.6 |
| 10.5 | 10.8 | 11.2 | 11.5 |

In a histogram, we group data into similar categories and show frequency. In the dot plot, each piece of data was shown, whereas, in a histogram, we lump them together with similar data. For example, with the data above, we will group the data together by 0.5 intervals. For example, we start with the lowest piece of data, 10.0 and create a category of 10.0 – 10.4. Each category will use the same increment of 0.5. Then we will find the frequency for that group of data. See the histogram below.

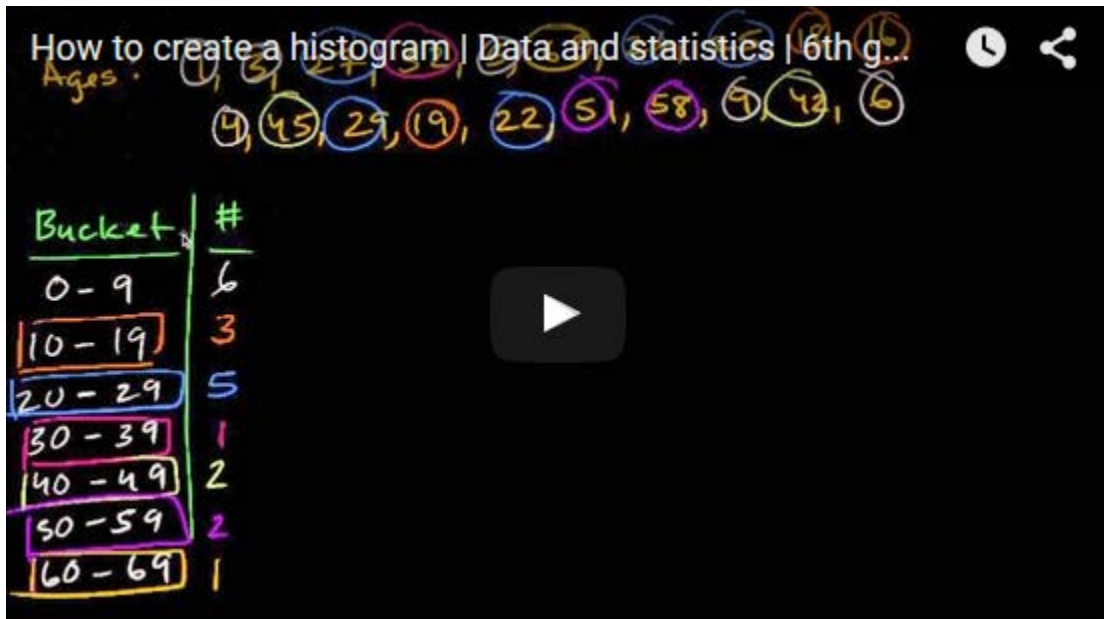
Times for 100 Meter Dash



The times are listed along the horizontal axis and the frequencies are listed on the vertical axis. Notice that the times are listed in equal intervals, 10.5 to 10.9, rather than each individual time.

***Watch this video for further explanation on making histograms:**

Click on the link to watch the video "[How to create a histogram](#)" or click on the video.



Stop! Go to Questions #10-16 about this section, then return to continue on to the next section.

Box-and-whisker Plot

A box-and-whisker plot is a method of showing how data is distributed using the median, upper and lower quartiles, and the greatest and least values known as extremes.

In this unit, you will be responsible for interpreting and creating box-and-whisker plots. However, to fully understand box-and-whisker plots, we will go through step by step instructions on how they are constructed.

Recall that the **median** of a set of data is the number in the center when the data is ordered from least to greatest. Basically, the median separates the data in half.

The median of the lower half of the set of data is called the **lower quartile**. The median of the upper half of the set of data is called the **upper quartile**. The **inter-quartile range** is the range of the middle half of a set of numbers and is found by finding the difference between the upper quartile and the lower quartile.

***Watch this video for an introduction to box and whisker plots:**

Click on the link to watch the video "[Box and whisker plot](#)" or click on the video.



Follow along with the example:

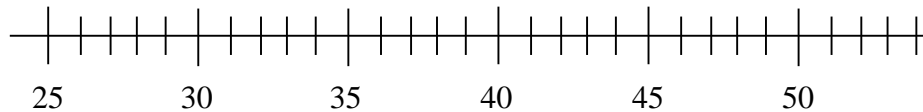
Example #1: Mr. Zehnder's science class wanted to study weather patterns and decided to do this using outside temperatures. Every day at 1:00, the students would check the temperature outside the science class window. Use the data provided to construct a box-and-whisker plot.

38 33 45 42 39 48 49 43 52

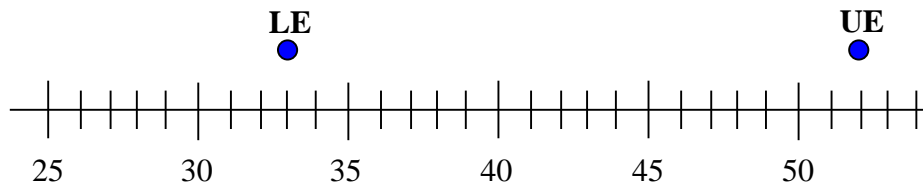
Step #1: Arrange the data in order from least to greatest.

33 38 39 42 43 45 48 49 52

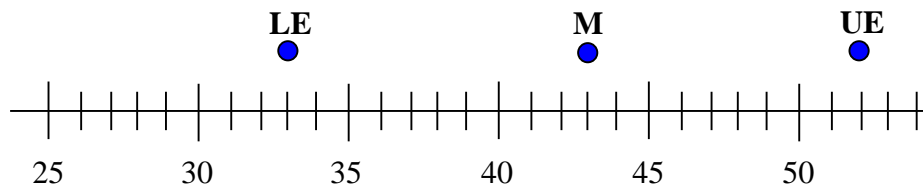
Step #2: Draw a number line to accommodate all the data.



Step #3: Plot the smallest and largest value (above the number line) and label them **LE** (lower extreme) and **UE** (upper extreme).



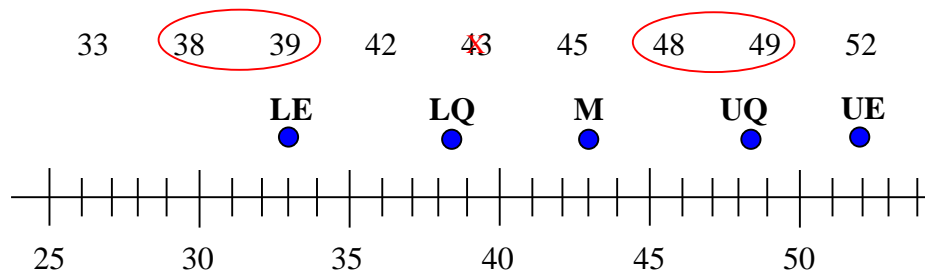
Step #4: Find the **median** of the set of data (plot a point and label it **M**). Remember that the median is the middle data value. Because there is an exact middle value, we do not have to find an average. Notice that 43 is the median. This is also known as Q2.



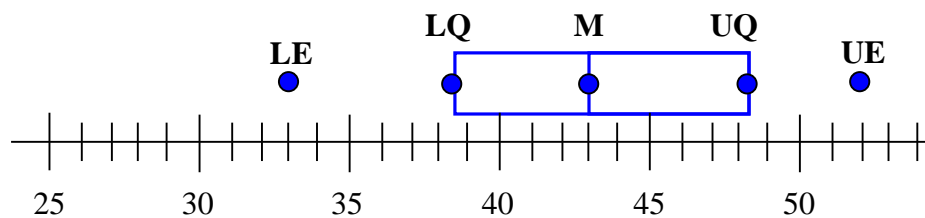
Step #5: Find the lower and upper quartile (the median between LE and M, and M and UE), label these numbers **LQ (lower quartile or Q1)** and **UQ (upper quartile or Q3)**.

To find the lower quartile, you will find the median of the first half of the data. Because there is an odd number of data, the median is not counted as part of the lower or upper quartiles. To find the lower quartile, find the median of 33, 38, 39, 42. However, we end up with 2 numbers in the middle, so take the average of 38 and 39. The LQ (lower quartile) is 38.5.

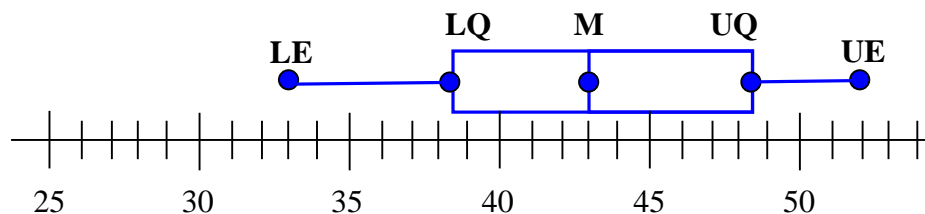
To find the upper quartile, we will do the same thing with the other half of the data. We will not count the median as part of the upper set. Find the median of 45, 48, 49, 52. We have two middle numbers, 48 and 49. Average these to get 48.5. This is our upper quartile.



Step #6: Draw a vertical line through the lower quartile, median, and upper quartile and construct a box.



Step #7: Whiskers are now extended from the ends of the box to the lower and upper extremes.



The box-and-whisker plot has now separated our data into fourths. For example, the whisker from 33 to 39 contains 25% of the data.

We will also find something called the **Interquartile Range**. This is the range of the middle 50% of our data. We find this by subtracting the lower quartile from the upper quartile.

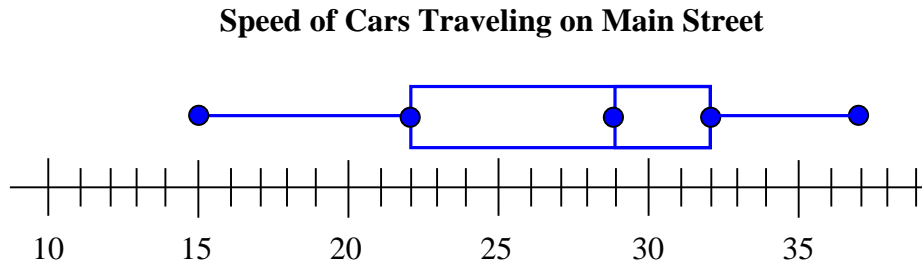
In this case, we subtract $48.5 - 38.5 = 10$. 10 is the interquartile range for this data.

***Watch this video for help constructing a box-and-whisker-plot:**

Click on the link to watch the video "[Constructing a box and whisker plot](#)" or click on the video.



Example #2: Use the box-and-whisker plot to answer the questions below.

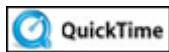


- a.) What is the slowest speed? 15 MPH
- b.) What is the median of the set of data? 29 MPH
- c.) What is the value of the upper quartile? 32 MPH
- d.) What is the interquartile range? 10

*Click [here](#) to practice reading a box-and-whisker plot.

You may also encounter data that may have **outliers**. Outliers are data that do not easily fit with the rest of the data. For example, if you measured all of the students in the class as well as your teacher, you may find that most of the data is similar (all of the students) but you may have an outlier (the teacher's height). When outliers exist, we plot them on the horizontal number line but do not connect them to the box and whisker plot. Outliers are data that are much lower or much higher than the rest of the data. We can calculate outliers mathematically this way: $\text{data} < Q1 - 1.5 \times \text{IQR}$ or $\text{data} > Q3 + 1.5 \times \text{IQR}$. Q1 is the lower quartile, Q3 is the upper quartile, and IQR is the Interquartile Range.

Watch this video regarding outliers:



Example 1: Outliers -- Far Out Waves and Surfers (03:20)

Watch this video for further explanation on how to make a box and whisker plot with outliers:

Click on the link to watch the video "[boxplot with outliers](#)" or click on the video.



Example #3: Jara collected data on the ages of people who attended a meeting. It seems it may have been a meeting for parents but a couple kids were brought along. Make a box-and-whisker plot for the data. Find any outliers and adjust the plot accordingly.

Data: 29, 33, 52, 56, 15, 33, 8, 25, 42, 55, 45, 35, 37, 54, 40, 41, 47, 45

First, make sure you arrange your data from least to greatest:

8, 15, 25, 29, 33, 35, 37, 40, 41, 42, 45, 45, 47, 48, 52, 54, 55, 56

Now, find the median. Remember, the median is the middle number. If you have an even amount of data, you will average the two middle numbers. If you have an odd amount of data, you will use the exact middle number. In this data, we have an even amount of data, so we must average the two middle numbers.

8, 15, 25, 29, 33, 35, 37, 40, 41, 42, 45, 45, 47, 48, 52, 54, 55, 56

The average of 41 and 42 is 41.5. Median = 41.5. This is also called Q2.

Now, to find Q1 and Q3 (lower and upper quartiles), we must find the median of the lower half and upper half of the data.

8, 15, 25, 29, 33, 35, 37, 40, 41, 42, 45, 45, 47, 48, 52, 54, 55, 56

Notice, the data is divided in half. Find the median of the lower half and upper half. Since there is an odd amount of data on each side, we will not have to do any averaging.

8, 15, 25, 29, 33, 35, 37, 40, 41, 42, 45, 45, 47, 48, 52, 54, 55, 56

We now have $Q1 = 33$ (lower quartile)

$Q2 = 41.5$ (median)

$Q3 = 48$ (upper quartile)

Now, we will find IQR (interquartile range). $IQR = Q3 - Q1 = 48 - 33 = 15$

Now, we will consider if we have outliers in our data. Outliers are data that are much lower or much higher than the rest of the data. We calculate this mathematically as:

$Q1 - 1.5(IQR)$ and $Q3 + 1.5(IQR)$.

So, for this data, first calculate $1.5 * IQR$. $1.5 * (15) = 22.5$.

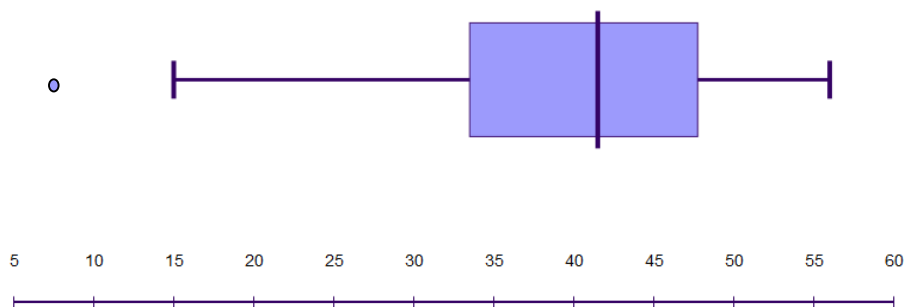
Now, we will see if any data points are more than 22.5 away from $Q1$ or $Q3$.

$$Q1 - 22.5 = 33 - 22.5 = 10.5$$

$$Q3 + 22.5 = 48 + 22.5 = 70.5$$

Does the data have values smaller than 10.5 or larger than 70.5? If so, then we consider these to be outliers. We do have one value, 8, that will be an outlier.

When you make the box and whisker plot, you will plot the 8 as a point, but you will use 15 as then end of your whisker because 8 is an outlier. Your box and whisker plot should look like this:



Notice that 8 is an outlier and plotted separately as an individual point. 15 is marked as the end of the left whisker, while 56 is the end of the right whisker. You get these numbers from the lowest and highest data values (excluding outliers). Your boxes are made from Q1, Q2, and Q3 (lower quartile, median, and upper quartile).

Stop! Go to Questions #17-30 to complete this unit.