

STATISTICS AND HISTOGRAMS

Data may be analyzed through organizing it into frequency tables, examining the central measures of tendency, and then graphing the results as a histogram. We will look at the statistics of data and graphing a histogram.

Statistics of Data and Graphing a Histogram

Statistics is collecting, organizing, analyzing, and displaying data usually in a graphical manner.

A **frequency table** may be used to organize data. In a frequency table, a tally is made to count the number of times a piece of data occurs.

To analyze data, **measures of central tendency** are often used. Three common measures of central tendency are **mean, median, and mode**.

The **range** of data is the difference between the highest and lowest value.

The following are the heights (in inches) of a sixth grade class. Organize the data into a frequency table, and then make a histogram to display the results in a graph.

54 63 60 57 54 55 60 59 58 56 59 57
57 52 57 59 58 56 58 59 56 56 55 57

Height (inches)	Tally	Frequency
52		1
54		2
55		2
56		4
57		5
58		3
59		4
60		2
63		1

Mode is the item that occurs most often. In this case the mode is a height of **57** inches.

Median is the middle number when the data is arranged in ascending or descending order.

In this case there are two middle numbers, 57; therefore the median is the average of the two numbers.

Median = $(57 + 57) / 2$ which is **57**.

The data can easily be arranged in ascending order.

52
54
54
55
55
56
56
56
56
57
57
57
57
57
57
57
58
58
58
59
59
59
59
60
60
63

The **mean** of the data is the average of the data. The numbers are totaled and then divided by the number of numbers. The *center* is the same thing as mean.

Mean = $1372 / 24 = 57.2$

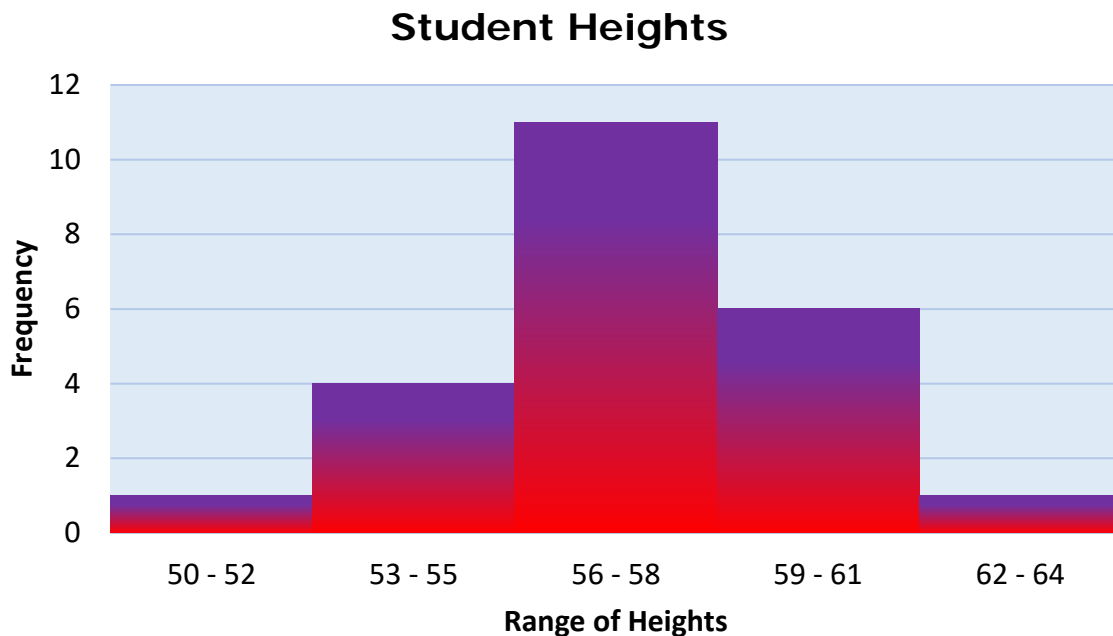
The **range** of the data is the difference between the highest and lowest values. The *spread* is the same thing as range.

Range = $63 - 52 = 11$

A frequency table may be arranged into a table that displays **intervals of data** along with the frequency of the intervals. The height of the sixth graders has been organized into intervals of 3, starting with 50 and ending with 64.

Height (inches) (intervals of 3)	Tally	Frequency
50 – 52		1
53 – 55		4
56 – 58	++++ +++++	11
59 – 61	++++	6
62 – 64		1

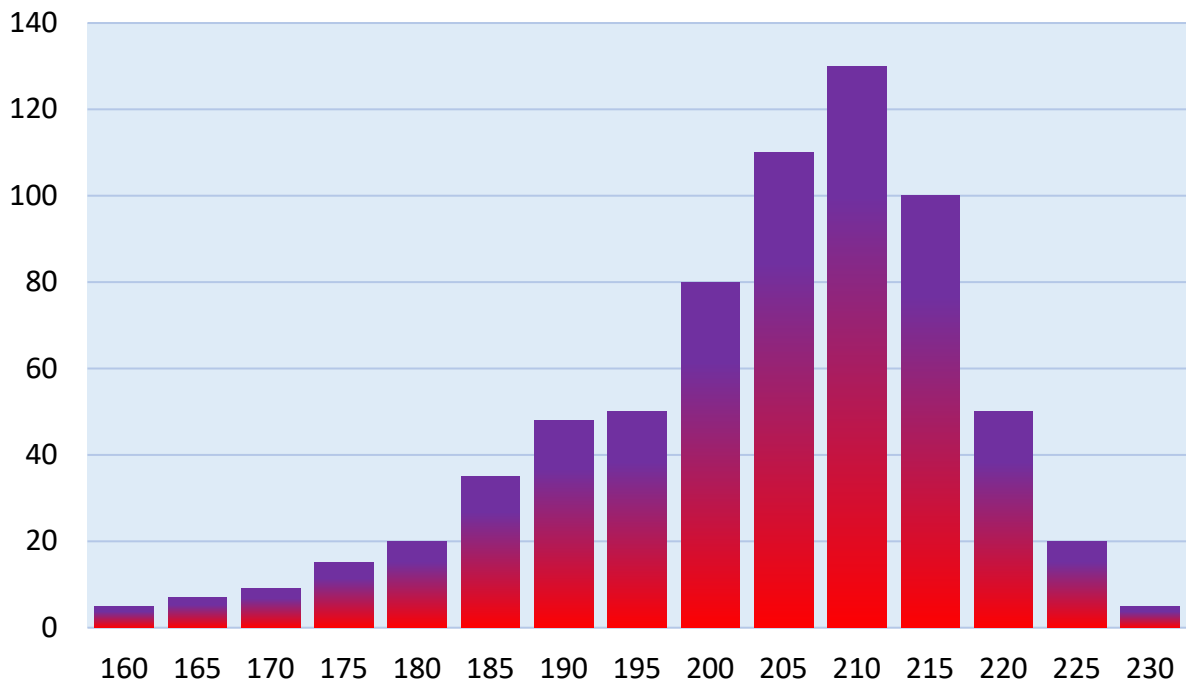
A **histogram** is a graphical representation of the data in the table graphing the intervals of data by the frequency of the heights within the intervals.



The graph shows that most 6th graders' heights fall within a range of 53 inches through 61 inches.

Outliers are data that occur infrequently for the group of data being analyzed. In this case the data falling in the interval of 50 – 52 is considered outlying data, along with data falling in the interval of 62 – 64, where there is only one height each falling in those intervals, a height of 52 inches and a height of 63 inches.

Shape is the term used to describe how the data is distributed. This data is somewhat **symmetrical** with small amounts of data occurring in the outer intervals and a high concentration of the data near the center. Data is symmetrical if it is balanced from left to right. It is considered **skewed** if all of the data occurs near one end or the other.



Box-and-Whisker Plots

A box-and-whisker plot displays data in four parts, the lower extreme, the lower quartile, the upper quartile, and upper extreme.

Michael has been keeping track of the high temperatures of the day for the past two weeks. Construct a box-and-whisker plot to display the data collect.

74, 82, 90, 91, 91, 85, 87, 77, 82, 83, 85, 79, 90, 92

First order the data in numerical order.

74, 77, 79, 82, 82, 83, 85, 85, 87, 90, 90, 91, 91, 92

Next find the median of the data.

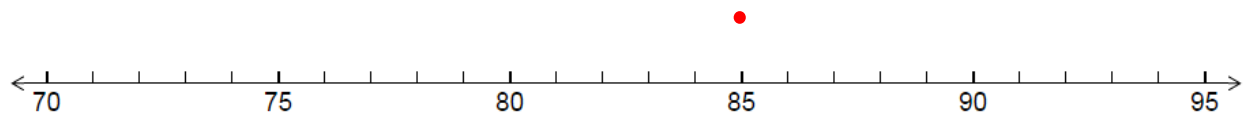
74, 77, 79, 82, 82, 83, 85, 85, 87, 90, 90, 91, 91, 92

$$85 + 85 = 170$$

$$170/2 = 85$$

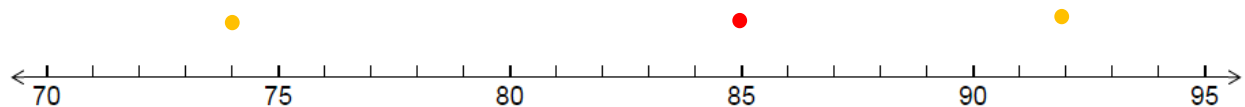
85 is the **median**.

On a number line put a dot by 85



Next plot the lower extreme and the upper extreme. The **lower extreme** is the smallest number in the data set. The **upper extreme** is the biggest number.

74, 77, 79, 82, 82, 83, 85, 85, 87, 90, 90, 91, 91, 92



Now we need to find the **lower quartile** and **upper quartile** medians.

74, 77, 79, 82, 82, 83, 85, 85, 87, 90, 90, 91, 91, 92

To find the **lower quartile** we find the median of the numbers less than the median of the whole set.

74, 77, 79, 82, 83

$$79 + 82 = 161$$

$$161/2 = 80.5$$

The **lower quartile** is 80.5



To find the **upper quartile** we find the median of the numbers greater than the median of the whole set.

87, 90, 90, 91, 92

$$90 + 91 = 181$$

$$181/2 = 90.5$$

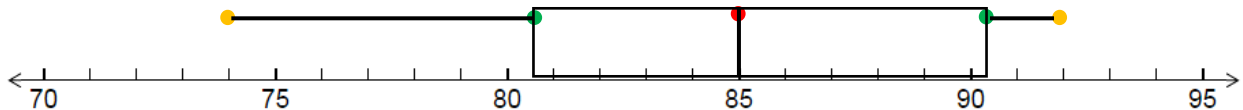
The **upper quartile** is 90.5



Now its time to draw the box and the whiskers to the plot.

Draw a box around the lower and upper quartile.

Draw a line to show the median. Now connect the lower and upper extremes to the box with lines.



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

Constructing a box and whisker plot | Probability and Statis

The owner of a restaurant wants to find out more about where his patrons are coming from. One day he decided to gather data about the distance (in miles) that people commuted to get to his restaurant. People reported the following distances traveled:

14, 6, 3, 2, 4, 15, 11, 8, 1, 7, 2, 1, 3, 4, 10, 22, 20

He wants to create a graph that helps him understand the spread of distances (and the median distance) that people travel. What kind of a graph should he create?

1, 1, 2, 2, 3, 3, 4, 4, 6, 7, 8, 10, 11, 14, 15, 20, 22

median