1.6.1 Graphing Data

Learning Objective(s)
1 Read and interpret data from tables and pictographs.
2 Read and interpret data from bar graphs
3 Read and interpret data from line graphs

Introduction

A nurse is collecting blood type data from her patients. When a new patient is checked in, the nurse does a simple finger-prick test to see whether the patient's blood is type A, B, AB, or O. (These are the four possible blood types. Each one also carries a + or – to represent the RH factor, but for our purposes, let's just track the type, not the + or – .) She tracks her results by creating a two-column table with the patient's name and blood type.

Name	Blood Type
Dominique	A
Ilya	O
Raul	AB
Madison	O
Philip	AB
Samuel	B
Josefine	O
Brett	O
Paula	B
Leticia	AB

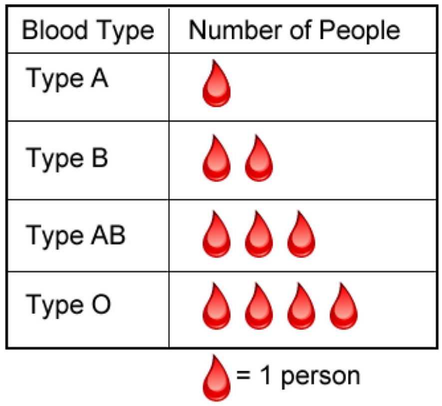
The information in this table is an example of **data**, or information. In this case, the nurse has gathered a fair amount of data about her patients' blood types. By analyzing the data, she can learn more about the range of patients that she serves.

Data helps us make many kinds of decisions. Organizing data into graphs can help us get a clear picture of a situation and can often help us make decisions based on the picture. So how do you take data and make a picture out of it? Let's take a look.

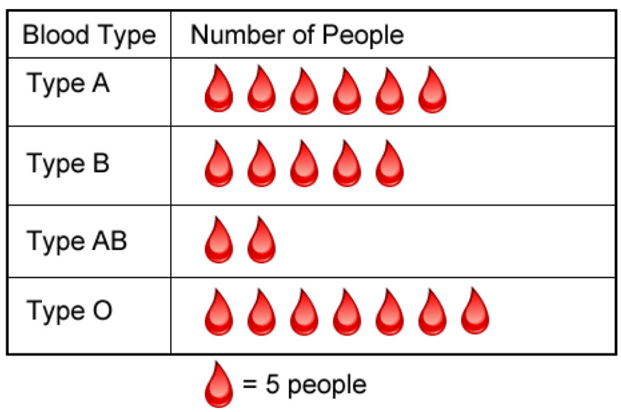
Pictures of Data

Objective 1

Let's return to the data set used previously. If the nurse wanted to represent the data visually, she could use a **pictograph**. Pictographs represent data using images. This visual presentation helps illustrate that for the data in her table, Type O blood is the most common, and Type A blood is the least common.



Interested by the results of this small survey, the nurse continues to document the blood types of her patients until she has surveyed 100 people. She puts all of this data in a table, but she finds that it is hard for her to quickly identify what the data is telling her. She decides to make another pictograph using a different scale.



To read this pictograph, all you need is the scale—the number of people that each blood drop symbol represents. In this graph, each blood drop represents 5 people. There are six drops next to Type A, so 5 • 6 = 30 people had Type A blood. The table below shows the rest of the information.

Blood Type	**Number of People**
Type A	6 drops • 5 people = 30 people
Type B	5 drops • 5 people = 25 people
Type AB	2 drops • 5 people = 10 people
Type O	7 drops • 5 people = 35 people

Example	
Problem	**The pictograph below shows the number of medals earned at an international competition. How many medals did Japan earn?**

Country	Medals
Japan	♦♦♦♦♦
Argentina	♦♦
Germany	♦♦♦♦♦♦
Egypt	♦♦♦♦

♦ = 4 medals

	Look at the scale of the pictograph. Each ♦ represents 4 medals.
$5 \bullet 4 = 20$	Japan has 5 ♦s, so the total number of medals is $5 \bullet 4 = 20$.
Answer	Japan earned 20 medals.

Self Check A

Which table accurately represents the data shown in the pictograph below?

Employee	Hourly wage
Wayne	$ $ $ $ $
Sarah	$ $ $ $ $ $ $
Leigh	$ $ $ $

$ = $4

A)

Employee	Hourly wage
Wayne	$5
Sarah	$7
Leigh	$4

B)

Employee	Hourly wage
Wayne	$20
Sarah	$22
Leigh	$19

C)

Employee	Hourly wage
Wayne	$10
Sarah	$14
Leigh	$8

D)

Employee	Hourly wage
Wayne	$20
Sarah	$28
Leigh	$16

Bar Graphs

Objective **2**

Representing data as pictures doesn't always make sense either. **Bar graphs** are an alternative (and popular) way to represent data sets, especially those with large amounts of data or which do not lend themselves well to individual symbols. In a bar graph, the number of items in a data category is represented by the height or length of bars.

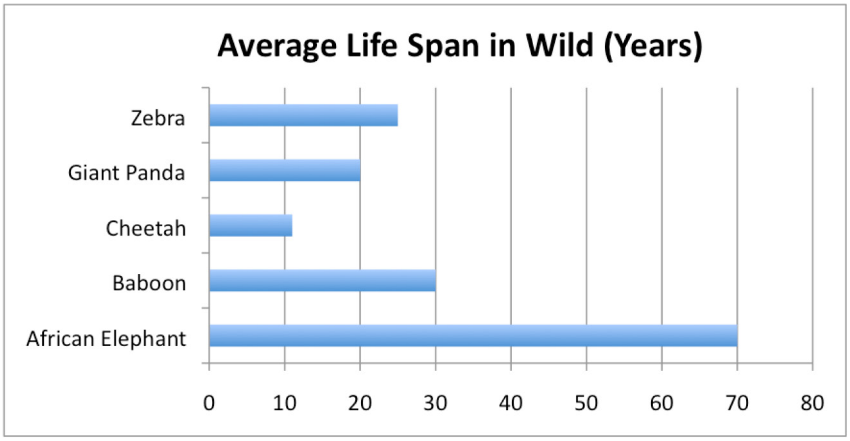
As when reading pictographs, paying attention to the scale is essential—small differences in the height of two bars can sometimes represent thousands of dollars, for example!

Let's look at one example. Here is some information about the average life span of five animals in the wild, presented in a table.

Animal	Average Life Span in the Wild (Years)
Zebra	25
Giant Panda	20
Cheetah	11
Baboon	30
African Elephant	70

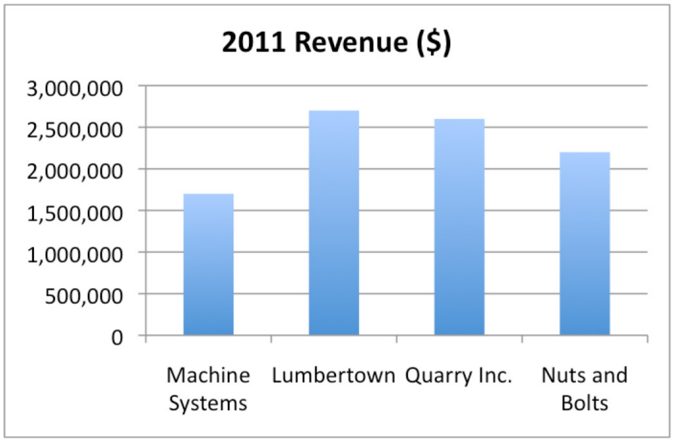
Source: National Geographic, accessed July 2011

This data is fine in a table, but presenting it as a bar graph helps the viewer compare the different life spans more easily. Look at the bar graph below. In this example, the animals are listed on the left side of the graph (also called the **y-axis**), and the life span in years is listed on the bottom (the **x-axis**). The graph shows the information by the length of the bar associated with each animal name.



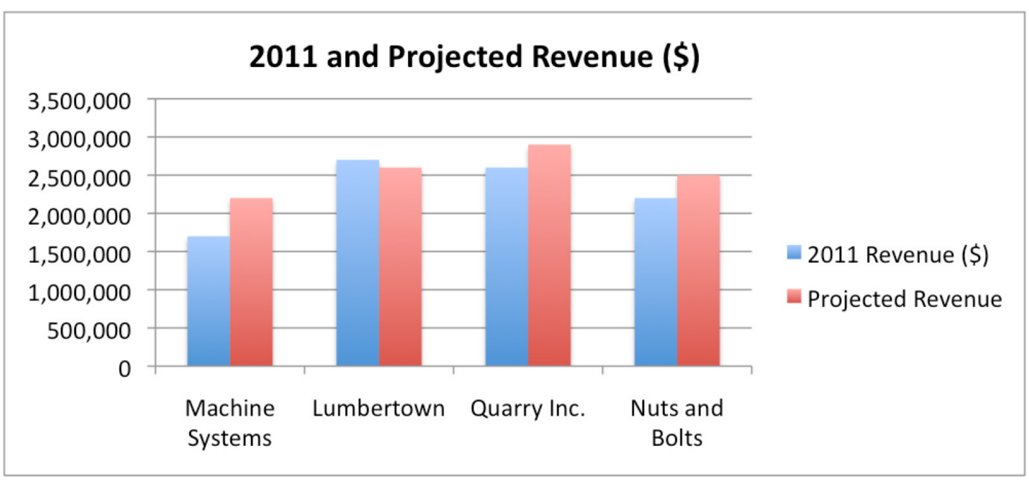
Bar graphs are generally used to compare quantities, not to determine exact quantities, especially when the scale is large, as in the next graph.

The bar graph below shows total revenue for four fictional companies in 2011. Notice that the scale, on the *y*-axis, is very large—each horizontal line represents an increase of $500,000. For this reason, it is difficult to tell exactly how much money each company made in 2011. However, comparing the bars is straightforward—glancing at the data, you can tell that Lumbertown earned the most (a little over $2,500,000), while Machine Systems earned the least (about $1 million less, at just over $1,500,000).

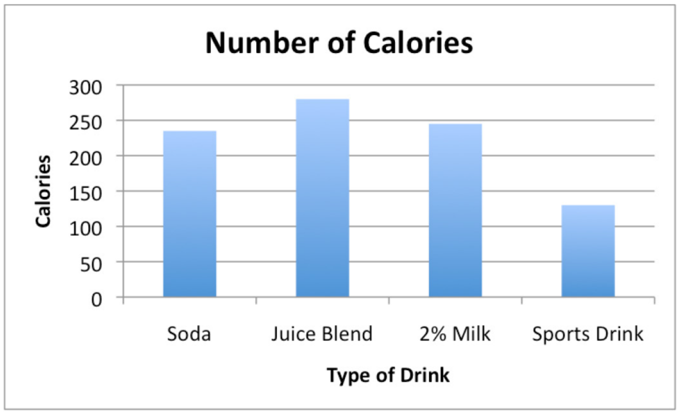


You can also use bar graphs to showcase multiple pieces of information about a specific situation. For example, let's show the next year's projected revenue for each company on the graph that you just looked at. You can leave the existing bars in the graph and just add 4 more.

The blue columns remain, but now they are accompanied by four new red columns that represent the projected revenue for these companies. Again, this data could be expressed in a table—with a bar graph, you gain ease of quick comparison, but lose the detail of the exact values. Looking at this graph tells you that while Lumbertown has the highest revenue for 2011, it is projected to decrease. Conversely, Machine Systems is projected to increase its revenue. Seeing data visually can help you understand the story that the data is telling about a situation.

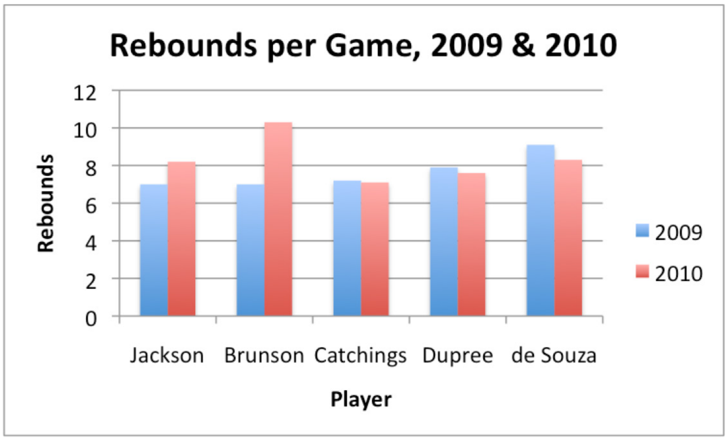


Example		
Problem	**Use the graph to list the drinks from the most number of calories to the least number of calories (serving size: 16 oz).**	



		The *y*-axis shows total calories, and the *x*-axis shows the drink. The taller the bar, the more calories the drink has.
	Juice Blend (≈275)	The juice blend contains over 250 calories, so it has the most calories per serving.
	2% Milk (≈245) Soda (≈230)	Soda and 2% milk are both between 200 and 250 calories, but the bar for 2% milk is taller, so it must contain more calories.
	Sports Drink (≈125)	Sports drink has the shortest bar; it contains about 125 calories.
Answer	From most to least number of calories per serving: Juice Blend, 2% Milk, Soda, Sports Drink	

Example	
Problem	**Based on the graph below, which player's rebounding increased the most from 2009 to 2010?**



Source: WNBA.com, accessed July 2011

	The *y*-axis shows rebounds per game, and the *x*-axis shows the player's name. A taller bar represents more rebounds per game by the player. This graph shows two sets of data—one for 2009, in blue, and one for 2010, in red. To compare the data from one year to the next, compare the heights of the two bars for each player.
Jackson and Brunson	Two players had higher rebound averages in 2010 than they did in 2009. This is indicated by the red bar being taller than the blue bar. The other players' red bars are shorter, so their rebounds decreased.
	Comparing the sizes of the increases, you can tell that Brunson increased her per game rebounding *more* than Jackson did.

Answer The player whose rebounding increased the most from 2009 to 2010 was Brunson.

Line Graphs

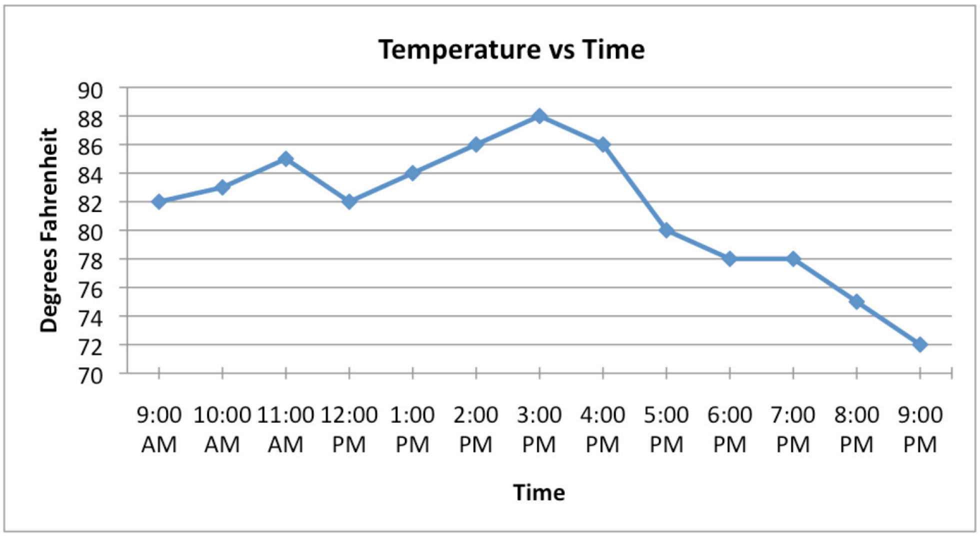
Objective 3

A **line graphs** are often used to relate data over a period of time. In a line graph, the data is shown as individual points on a grid; a trend line connects all data points.

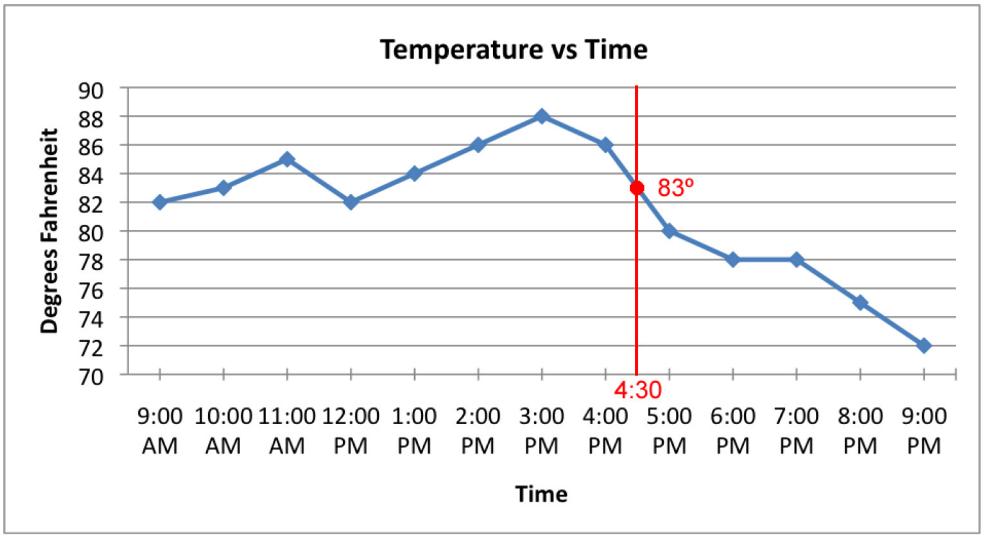
A typical use of a line graph involves the mapping of temperature over time. One example is provided below. Look at how the temperature is mapped on the *y*-axis and the time is mapped on the *x*-axis.

Each point on the grid shows a specific relationship between the temperature and the time. At 9:00 AM, the temperature was 82º. It rose to 83º at 10:00 AM, and then again to 85º at 11:00 AM. It cooled off a bit by noon, as the temperature fell to 82º. What happened the rest of the day?

The data on this graph shows that the temperature peaked at 88º at 3:00 PM. By 9:00 PM that evening, it was down to 72º.



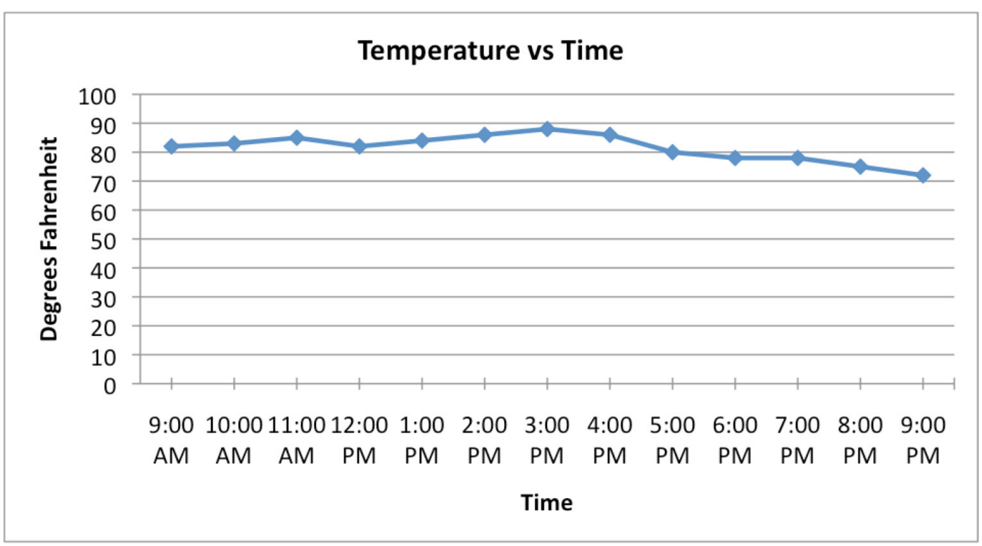
The line segments connecting each data point are important to consider, too. While this graph only provides data points for each hour, you could track the temperature each minute (or second!) if you wanted. The line segments connecting the data points indicate that the temperature vs. time relationship is continuous—it can be read at any point. The line segments also provide an estimate for what the temperature would be if the temperature were measured at any point between two existing readings. For example, if you wanted to estimate the temperature at 4:30, you could find 4:30 on the *x*-axis and draw a vertical line that passes through the trend line; the place where it intersects the graph will be the temperature estimate at that time.



Note that this is just an estimate based on the data—there are many different possible temperature fluctuations between 4:00 PM and 5:00 PM. For example, the temperature could have held steady at 86º for most of the hour, and then dropped sharply to 80º just before 5:00 PM. Alternatively, the temperature could have dropped to 76º due to a sudden storm, and then climbed back up to 80º once the storm passed. In either of these cases, our estimate of 83º would be incorrect! Based on the data, though, 83º seems like a reasonable prediction for 4:30 PM.

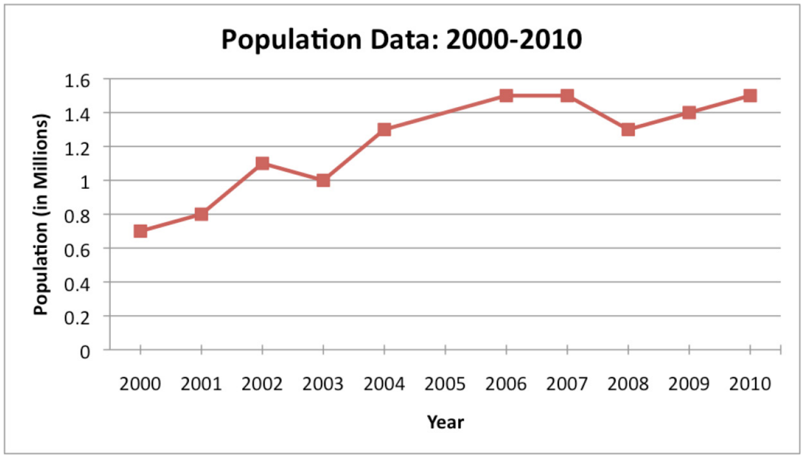
Finally, a quick word about the scale in this graph. Look at the *y*-axis—the vertical line where the Degrees Fahrenheit are listed. Notice that it starts at 70º, and then increases in increments of 2º each time. Since the scale is small and the graph begins at 70º, the temperature data looks pretty volatile—like the temperature went from being warm to hot to very cold! Look at the same data set when plotted on a line graph that begins at 0º and has a scale of 10º. The peaks and valleys are not as apparent!

As you can see, changing the scale of the graph can affect how a viewer perceives the data within the graph.

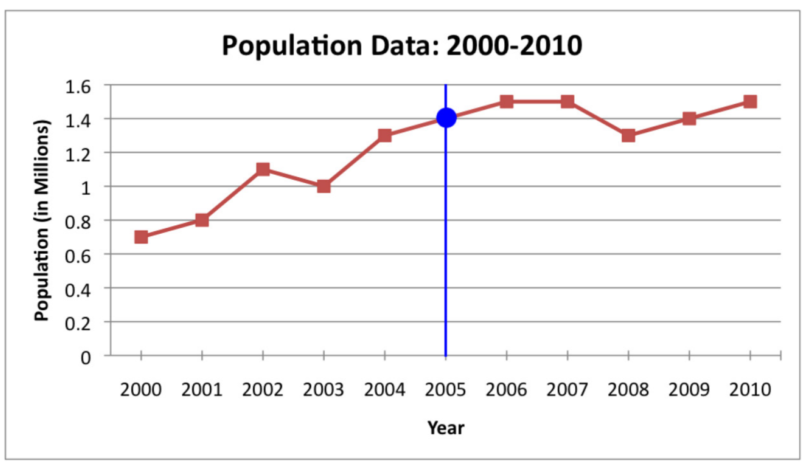


Example

Problem **Population data for a fictional city is given below. Estimate the city's population in 2005.**



Look at the line graph. The population starts at about 0.7 million (or 700,000) in 2000, rises to 0.8 million in 2001, and then again to 1.1 million in 2002. To find the population in 2005, find 2005 on the *x*-axis and draw a vertical line that intersects the trend line.



The lines intersect at 1.4, so 1.4 million (or 1,400,000) would be a good estimate.

Answer The population in 2005 was about 1.4 million.

Summary

Data is mathematical information. Mathematical data is often recorded in tables to organize, or spreadsheets to organize and sort. Graphs can help you see the data visually, which can help you to better understand the data. A pictograph is a graph that uses symbols to represent data. Bar graphs show the frequency of categorical data, using bars instead of symbols. By contrast, line graphs are usually used to relate continuous data over a period of time.

1.6.1 Self Check Solutions

Self Check A

Which table accurately represents the data shown in the pictograph below?

Employee	**Hourly wage**
Wayne	$ $ $ $ $
Sarah	$ $ $ $ $ $ $
Leigh	$ $ $ $

$ = $4

D)

Employee	**Hourly wage**
Wayne	$20
Sarah	$28
Leigh	$16

Each $ symbol represents $4, so if you multiply the number of $s in a row by $4, you will find that Wayne earns $20, Sarah earns $28, and Leigh earns $16.

1.6.2 Measures of Center

Learning Objective(s)
1 Find the mean, median, and mode of a set of numbers.
2 Find the ramge of a set of numbers.
3 Read and interpret data from box-and-whisker plots.
4 Solve application problems that require the calculation of the mean, median, or range.

Introduction

When trying to describe data, **mean**, **median**, and **mode** are important tools. These measures of center all use data points to approximate and understand a "middle value" or "average" of a given data set. Two more measures of interest are the **range** and **midrange**, which use the greatest and least values of the data set to help describe the spread of the data.

So why would you need to find out the middle of a data set? And why do you need three measures instead of just one? Let's look closely at these measures of center and learn how they can help us understand sets of data.

Mean, Median, and Mode

Objective 1

"Mean" is a mathematical term for "average" which you may already know. Also referred to as the "arithmetic mean," it is found by adding together all the data values in a set and dividing that sum by the number of data items.

You can often find the average of two familiar numbers, such as 10 and 16, in your head without much calculation. What number lies half way between them? 13. A mathematical way to solve this, though, is to add 10 and 16 (which gives you 26) and then divide by 2 (since there are 2 numbers in the data set). 26 ÷ 2 = 13

Knowing the process helps when you need to find the mean of more than two numbers. For example, if you are asked to find the mean of the numbers 2, 5, 3, 4, 5, and 5, first find the sum: 2 + 5 + 3 + 4 + 5 + 5 = 24. Then, divide this sum by the number of numbers in the set, which is 6. So the mean of the data is 24 ÷ 6, or 4.

In the previous data set, notice that the mean was 4 and that the set also contained a value of 4. This does not always occur. Look at the example that follows—the mean is 18, although 18 is not in the data set at all.

Example		
Problem	**Find the mean of the set: 4, 7, 28, 33.**	
	$4 + 7 + 28 + 33 = 72$	Add all the values.
	$\frac{72}{4} = 18$	Divide by 4, the number of values.
Answer	The mean is 18.	

Next, let's look at the "median." The median is the middle value when the data is ordered. If there are two middle values, the median is the average of the two middle values.

To calculate the median, you first put your data into numerical order from least to greatest. Then identify the middle value(s).

For example, let's look at the following values: 4, 5, 1, 3, 2, 7, 6. To find the median of this set, you would put it in order from least to greatest.

1 2 3 4 5 6 7

Then identify the middle value. There are three values to the right of four and three values to the left of four. The middle value is 4, so 4 is the median.

If there is an even number of data items, however, the median will be the mean of the two center data items.

Example		
Problem	**Find the median of the set: 2, 6, 4, 3, 6, 7.**	
	2, 3, 4, 6, 6, 7	Arrange the values from least to greatest.
	2, 3, 4, 6, 6, 7	The set has 2 middle values. So take the mean (average) of the two values.
	$\frac{4+6}{2}=\frac{10}{2}=5$	
Answer	The median is 5.	

Finally, let's consider the "mode." The mode is found by looking for the data value that appears most often. If there is a two-way tie for most often, the data is bimodal and you use both data values as the modes. Sometimes there is no mode. This happens when there is no data value that occurs most often. In our example data set (2, 3, 4, 6, 6, 7), the number 5 appears 2 times and all other numbers appear once, so the mode is 6.

Example		
Problem	**Find the mode of the set: 12, 4, 12, 5, 5, 8, 12, 0, 1, 12.**	
	0, 1, 4, 5, 5, 8, 12, 12, 12, 12	Arrange the values from least to greatest

		(although this is not a necessary step, it sometimes helps to find the mode if the numbers are arranged in ascending order).
	0, 1, 4, 5, 5, 8, 12, 12, 12, 12	Find the value that occurs most often.
Answer	The mode is 12.	

Let's look at an example with some relevant data.

Example		
Problem	**Carlos received the following scores on his mathematics exams: 84, 92, 74, 98, and 82. Find the mean, median, and mode of his scores.**	
	$\frac{84+92+74+98+82}{5}$	To find the mean, add all the tests scores together and divide by the number of tests.
	$\frac{430}{5}=86$	The mean is 86.
	74, 82, 84, 92, 98	To find the median, order the test scores from least to greatest.
	84	There are five scores, so the middle test score is the third in the ordered list. This is the median.
	74, 82, 84, 92, 98	Since each number appears exactly one time, there is no mode.
Answer	The mean is 86. The median is 84. There is no mode.	

What can be learned from the mean, median, and mode of Carlos' test scores? Notice that these values are not the same.

Both the mean and the median give us a picture of how Carlos is doing. Looking at these measures, you notice that the middle of the data set is in the mid-80s: the mean value is 86, and the median value is 84. That's all you are really after when using median and mean—finding the center, or middle, of the data. Notice, also, that there is no mode, since Carlos did not score the same on two tests. In the case of test taking, the mode is often meaningless—unless there are a lot of 0s, which could mean that the student didn't do his homework, or really doesn't know what's going on!

Self Check A

During a seven-day period in July, a meteorologist recorded that the median daily high temperature was 91º.

Which of the following are true statements? There may be more than one answer.

i) The high temperature was exactly 91º on each of the seven days.
ii) The high temperature was never lower than 92º.
iii) Half the high temperatures were above 91º and half were below 91º.

Range

Objective **2**

There are other useful measures other than mean, median, and mode to help you analyze a data set. When looking at data, you often want to understand the spread of the data: the gap between the greatest number and the least number. This is the range of the data. To find the range, subtract the least value of the data set from the greatest value. For example, in the data of 2, 5, 3, 4, 5, and 5, the least value is 2 and the greatest value is 5, so the range is 5 – 2, or 3.

Let's look at a couple of examples.

Example		
Problem	**Find the range for the following set of numbers: 2, 4, 7, 10, 14, 35.**	
	range: 35 – 2 = 33	Subtract the least value from the greatest value to find the range.
Answer	The range is 33.	

Example		
Problem	**Find the range and midrange for the following set of numbers: 62, 88, 20, 145, 37, 105, 93, 22.**	
	least number: 20 greatest number: 145	Since the data set isn't ordered from least to greatest, identify the least and greatest numbers.
	range: 145 – 20 = 125	Subtract the least value from the greatest value to find the range.
Answer	The range is 125.	

Box-and Whisker-Plots

Objective 3

Another type of graph that you might see is called a **box-and-whisker plot**. These graphs provide a visual way of understanding both the range and the middle of a data set.

Here is a sample set of 15 numbers to get us started.

12, 5, 18, 20, 11, 9, 3, 5, 7, 18, 12, 15, 6, 10, 11

Creating a box-and-whisker plot from this data requires finding the median of the set. To do this, order the data.

3, 5, 5, 6, 7, 9, 10, 11, 11, 12, 12, 15, 18, 18, 20

This data set has 15 numbers, so the median will be the 8th number in the set: 11.

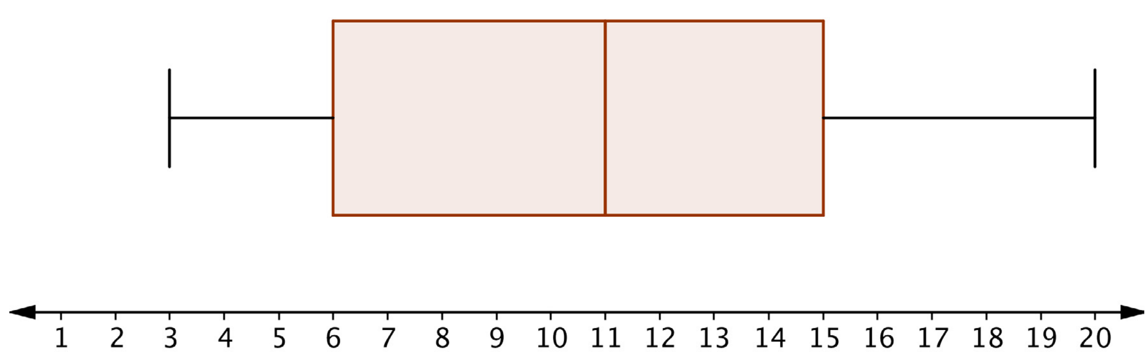
Finding the median of the data set essentially divides it into two—a set of numbers below the median, and a set of numbers above the median. A box-and-whisker plot requires you to find the median of these numbers as well!

Lower set: 3, 5, 5, 6, 7, 9, 10. Median: 6
Upper set: 11, 12, 12, 15, 18, 18, 20. Median: 15

So, the median of the set is 11, the median of the lower half is 6, and the median of the upper half is 15.

3, 5, 5, **6**, 7, 9, 10, **11**, 11, 12, 12, **15**, 18, 18, 20

A box-and-whisker plot for this data set is shown here. Do you see any similarities between the numbers above and the location of the box?



Notice that one “box” (rectangle section) begins at 6 (the median of the lower set) and goes to 11 (the median of the full set), and the other box goes from 11 to 15 (the median of the upper set).

The “whiskers” are the line segments on either end. One stretches from 3 (the least value in the set) to 6, and the other goes from 15 to 20 (the greatest value in the set).

The box-and-whisker plot essentially divides the data set into four sections (or **quartiles**): whisker, box, box, whisker. The size of the quartiles may be different, but *the number of data points* in each quartile is the same.

You can use a box-and-whisker plot to analyze how data in a set are distributed. You can also the box-and-whisker plots to compare two sets of data.

Using Measures of Center to Solve Problems

Objective **4**

Using mean, median, and mode, as well as range and midrange can help you to analyze situations and make decisions about things like which is the best, whether it is more reliable to walk or take the bus to school, or even whether to buy or sell a particular stock on the stock market.

Let’s look at an example of how analyzing data using measures of center can help you to make choices (and even get to school on time!).

FIX THIS

Example

Problem **Below, is a table listing the amount of time it took Marta to get to school by either riding the bus or by walking, on 12 separate days. The times are door to door, meaning the clock starts when she leaves her front door and ends when she enters school.**

Bus	Walking
16 min	23 min
14 min	19 min
15 min	21 min
14 min	21 min
28 min	22 min
15 min	20 min

- **Which method of travel is faster?**
- **If she leaves her house 25 minutes before school starts, should she walk or take the bus to be assured of arriving at school on time?**

$$bus: \frac{16+14+15+14+28+15}{6} = 17$$

$$walking: \frac{23+19+21+21+22+20}{6} = 21$$

Determine the mean of each travel method.

bus: 28 –14 = 14
walking: 23 – 19 = 4

Determine the range of each travel method.

bus: 14, 14, 15, 15, 16, 28

$$\frac{15+15}{2} = 15$$

walking: 19, 20, 21, 21, 22, 23

$$\frac{21+21}{2} = 21$$

Determine the median for each travel method.

bus: 14, 15
walking: 21

Determine the mode for each travel method.

	Bus	**Walking**
Mean	17	21
Median	15	21
Mode	14, 15	21
Range	14	4

Answer Looking at the mean, median, and the mode, the faster way to school is riding the bus. The data also shows that the bus is the most variable, with a range of 14, so if Marta wants to be *sure* that she gets to school on time, she should walk.

In the previous example, riding the bus is, *on average*, a faster way to school than walking. This is revealed in the mean of each method, which shows that the bus is 3 minutes faster. The mode and median show an even greater time advantage to riding the bus, and this is due to the one time high value of 28 minutes that isn't really accounted for in these measures. Notice the difference in the mean (17) and the median (15) for riding the bus, which lets you know there is some variance in the data.

As far as getting to school on time is concerned, while not being the fastest method, walking is the most reliable, with consistent values for mean, median, and mode, and a low value for the range, meaning that the spread of the data is very small.

Summary

Measures of center help you to analyze numerical data. The mean (or arithmetic mean) is often called the "average", and is found by dividing the sum of the data items by the number of items. The median is the number that is in the middle when the data is ordered from least to greatest, and the mode is the number that appears most often. The range is the difference between the least number and the greatest number. Box-and-whisker plots use the median and range to help you to interpret the data visually.

8.2.1 Self Check Solutions

Self Check A

During a seven-day period in July, a meteorologist recorded that the median daily high temperature was 91º.

Which of the following are true statements?

i) The high temperature was exactly 91º on each of the seven days.
ii) The high temperature was never lower than 92º.
iii) Half the high temperatures were above 91º and half were below 91º.

iii only
Half the high temperatures were above 91º and half were below 91º since the median will always represent the value where half the data is higher and half the data is lower.