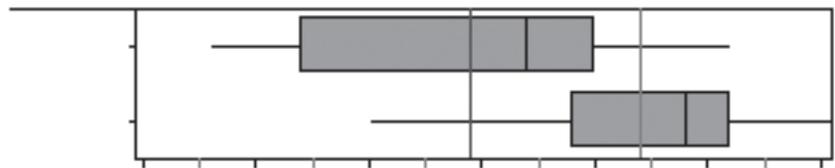


# 1

## Building on Pre-algebra





# Measures of Center—Aircraft Efficiency

The airline business has gone through turmoil recently, with bankruptcies and reorganizations. To keep costs down and remain profitable, it is important for airlines to use efficient aircraft. In this exploration, you'll examine some data and make recommendations about the most efficient aircraft to use.

In describing the airline data, you'll probably want to mention values that are typical, or in the middle of the data. You often hear statements about typical values, such as *Most men have shoe sizes 9 to 11* or *The average university student gains 15 pounds in the freshman year* or even *The median score on the Algebra 1 exam was 72%*. These statements refer to the *central tendency* of data sets.

- Q1** Two measures of central tendency are the mean and median. How do you calculate the *mean* (or *average*) of 36 numbers? How do you calculate the *median* (or middle number) of 36 numbers?
- Q2** Can the mean and median of 36 numbers be different? Can they be the same? Use examples to justify your answers.

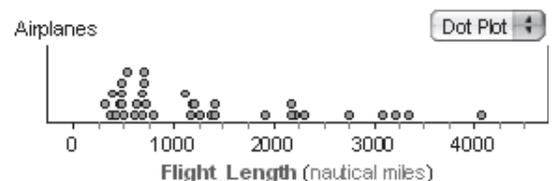
## INVESTIGATE

You have been asked to recommend the best plane for your company to use on a new route. The data you're given about airplanes are contained in a Fathom document.

A shorter version of this set of instructions is Choose **File** | **Open** | **Aircraft.ftm**.

1. Start the Fathom program. Go to the **File** menu and choose **Open**. Browse until you find the document called **Aircraft.ftm**.

On the screen, you will see three objects: a box of balls, called a *collection*; a graph, called a *dot plot*; and a table, called a *case table*. Each ball represents one airplane, or one *case* in the data collection. The planes are also listed as cases in the table. You can see the *attributes*, or the information available for each plane, listed across the top of the case table. Notice that five of the plane attributes have units, which are displayed in green on the screen. The dot plot shows one dot for each plane, representing that plane's flight length.



You can enlarge the collection or the case table by dragging the edge.

2. Make sure you understand what each attribute means and how the collection and the case table are related. Browse through the case table.
- Q3** For how many planes does the case table contain data? Which plane has the longest range (*Flight\_Length*)? Which airplane is the slowest (has the smallest value for *Speed*)?

You can also use the graph to answer Q3.

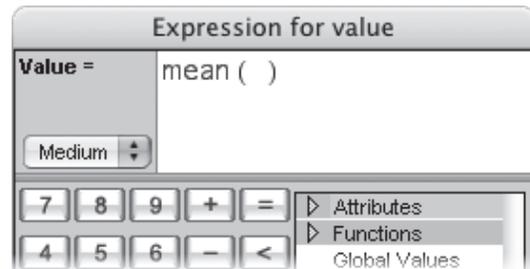
## Measures of Center—Aircraft Efficiency

continued

- Click the rightmost dot on the dot plot. Notice that the plane represented by this dot is now highlighted in the case table. While the cursor is over the dot, Fathom displays information about this airplane in the status bar at the bottom left of the screen. The rightmost dot corresponds to the plane with the longest flight length. The values displayed at the bottom of the screen should agree with your answer to Q3.

**Q4** Look at the 36 dots in the plot and think about their flight lengths. Estimate what you think a typical value for *Flight\_Length* would be.

- You can use Fathom to calculate measures of central tendency. Click on the dot plot to select it. Go to the **Graph** menu and choose **Plot Value**. You'll see a new window, called the *formula editor*.



Type `mean( )` (with the parentheses) and click **OK**. Fathom will calculate the mean of the 36 flight lengths, display the value at the bottom of the graph, and draw a vertical line corresponding to the mean's value on the graph. Again choose **Plot Value**, and plot the value `median( )`.

**Q5** Record the mean and median of these data and compare them with the estimate you made in Q4.

**Q6** Why do you think the mean is greater than the median for this data set?

**Q7** Does the mean or the median give a better indication of the flight length of the typical aircraft? Why?

**Q8** Does your answer to Q7 help narrow down your recommendation for which plane an airline should use? Explain your answer.

- Drag the graph icon from the object shelf to some white space in the Fathom window. Now drag the attribute *Operating\_Cost* from the case table onto this new graph and drop it where it says **Drop Attribute Here**. This will create a dot plot that shows operating cost per hour.

**Q9** Click on the dots on the new graph to determine which planes have the lowest and highest operating costs. Also find the mean and median of the operating costs.

**Q10** Based on your answers to Q9, what recommendations would you make about which plane the airline should use?



## Measures of Center—Aircraft Efficiency

continued

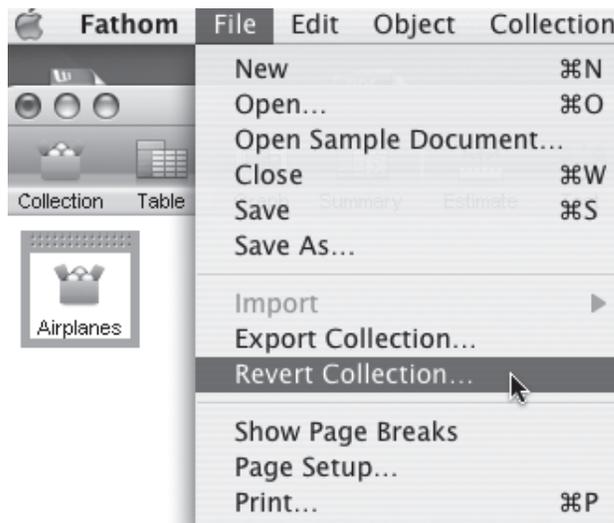
**Q11** Experiment with other attributes. Which attributes are most important when considering aircraft efficiency? What combination of attributes might make a good efficiency rating? You will develop a formula for efficiency involving several attributes when you complete the Explore More.

You may wonder how modifying some of the airplanes changes the mean or median. Could you enlarge the fuel tank to give it a longer flight length? How would that affect the mean and median flight length? With Fathom, you can change the values in the data to see what effect this has on the mean and the median. For any attribute you are considering, drag the rightmost dot to the left. The corresponding value in the table will change.

**Q12** Explore to answer these questions. Use complete sentences to describe how changing values in the data affects the mean and the median.

- How far do you have to move the dot to make the mean change?
- How far do you have to move the dot to make the median change?
- How many dots did you have to move to make the mean equal to the median?

If you want to undo any change in Fathom, choose **Edit | Undo**. You can undo repeatedly until the data set is back to its initial values. Another way to get back to the original data is to select the collection (the box of golden balls) and choose **File | Revert Collection**.



**Q13** Review your answer to Q2. Explain the conditions that make the mean equal to the median.

## Measures of Center—Aircraft Efficiency

continued

---

### EXPLORE MORE

You might use the fact that 1 knot (kt) = 1.152 miles per hour (mph). Or you can change the units from kt to mph in the case table, and Fathom will do the conversion for you.

A December 5, 2005, article in *Newsweek* about the new Boeing 777 compared the plane's efficiency to that of the Airbus A380, saying, "The Boeing burns 24 percent less fuel per seat mile." Create a formula for the aircraft efficiency that will take into account the cost per seat mile. Your formula could take into account other attributes as well.

Scroll to the far right of the table. You will see an attribute called *<new>*; click on it and type *efficiency*. Double-click the gray field below *efficiency*. Enter your formula for *efficiency* and click **OK**. Create the dot plot for your new attribute. What are its units? Are you looking for the plane with the largest or smallest value? Which plane do you think is the most cost efficient to fly? Does your answer agree with the answers of others in your class? What are the mean and median of your measure of efficiency?

**Objective:** Students will investigate the measures mean and median, using the dynamic aspect of Fathom to observe the effect that changing one data value has on the mean and median. Students will consider the ability of the mean and the median to measure central tendency.

**Student Audience:** Pre-algebra, Algebra 1

**Activity Time:** 50–70 minutes

**Setting:** Paired/Individual Activity or Whole-Class Presentation (use **Aircraft.ftm** for either setting)

**Mathematics Prerequisites:** Students can calculate the mean and median of a data set.

**Fathom Prerequisites:** Students can open Fathom sample documents.

**Fathom Skills:** Students will interpret data in case tables and graphs, create a dot plot, use the formula editor to plot a computed value on a graph, change the value of an attribute for one case by dragging a dot on the graph, and use **Undo** or **Revert Collection** to return a collection to its saved state.

**Notes:** This activity requires a long class period (or two periods) for students to complete the activity and define an efficient aircraft in the Explore More. If possible, give students time to explore Q11 and the Explore More. Take time for students to share their answers with the class and to explain the thinking behind their formulas.

**For a Presentation:** To use this activity to introduce Fathom to a class, ask the questions in the investigation as you show the features of Fathom that are introduced. Questions such as Q4, Q6, and Q10 can solicit several answers and begin good discussions. After the demonstration and discussion, you might have student groups work on the Explore More. If groups do not have computer access, you could distribute copies of the table and ask groups to devise formulas, then come up and add their attribute with its formula to the demonstration computer.

**Q1** The mean is the sum of all 36 values divided by 36. The median of a data set with an odd number of cases, ordered from least to greatest or greatest to least, is the middle number. The median of an ordered data set with an even number of cases is the mean of the two middle values. When the data are

sorted from smallest to largest, the median of a data set with 36 values will be the mean of the 18th and 19th value.

**Q2** It is quite likely that the mean and median of a data set are different, because they are calculated differently. It is possible to create a data set for which the mean and median happen to be the same (such as 1, 1, 1, 1, 1 or 1, 2, 3, . . . , 36).

## INVESTIGATE

- Many Fathom sample documents have comments related to their source and composition. Double-click on the collection and click the **Comments** panel tab to view this information.
- Managing screen real estate is an important skill for new Fathom users, as the workspace tends to get cluttered with tables, graphs, and inspector windows. Resizing and positioning these objects will become natural after some practice.
- The table records data for 36 planes. To find the plane with the given characteristics, you can scroll through the data carefully or sort the data. To find the plane with the longest range, select *Flight\_Length* in the case table and choose **Table | Sort Descending**. The B747-400 has the longest range, at 4065 N.M. Sorting on *Speed* in ascending order reveals that the ERJ-135 is the slowest, at 328 kt.
- “B747-400 (4065 N.M.)” is displayed at the bottom left.
- The estimate for the typical value should be between 750 and 2000 N.M.
- The formulas could be entered as  $\text{mean}(\text{Flight\_Length})$  and  $\text{median}(\text{Flight\_Length})$ . (Fathom automatically supplies the closing parenthesis.)
- The mean flight length is 1313.33 N.M., and the median is 949.5 N.M. The comparisons will vary.
- The mean is larger because a couple planes have long flight lengths. This type of data set—with a tail on the right when graphed—is referred to as *skewed right*.
- The median value better reflects where the middle of the majority of the data is located, but the mean takes into account the data spread out on the right. Together, both provide good information.

**Q8** Answers will vary.

**Q9** The operating costs are highest for the B747-100 and lowest for the ERJ-135. The dot plot for *Operating Cost* is similar to the dot plot for *Flight Length*, as they both have a clump of planes at the smaller values and a few planes spread out along the larger values. This spread results in a mean that is greater than the median. Reasonable student predictions should have the median and mean between 2000 and 4000, with the mean being the greater of the two.

**Q10** The ERJ-135 has the lowest operating cost, \$650/h. However, it is the slowest plane, it only seats 37, and it has one of the smallest ranges. This makes it impractical for commercial use.

**Q11** Answers will vary.

**Q12** The mean changes when any value changes. The median changes if a changing value becomes the 18th or 19th largest (the only two values that determine the median) or if values are moved to the other side of the median, which would affect the values in the 18th and 19th places. For these data, because there are 36 cases, the mean changes  $1/36$ th as much as the change in the case's value and in the same direction. Originally, the 18th and 19th values of *Flight Length* are 795 N.M. and 1104 N.M., respectively, and the median is the mean of these two values. Dragging the largest value to the left eventually causes it to become the 19th largest and then the 18th largest (as it joins the smaller half of the data); if it gets smaller than 707 N.M., then 707 N.M. becomes the 18th and 795 N.M. the 19th. The median can take on any value between 751 N.M. (the mean of 707 N.M. and 795 N.M.) and 945.5 N.M. (the mean of 795 N.M. and 1104 N.M.). As the value moves in this range, the median changes half as fast as the value does.

Answers will vary for the number of dots that need to move for the mean to equal the median. For flight length, moving the five longest flight lengths to the left until just before the median makes the mean and median approximately equal.

**Q13** Under certain circumstances, the mean and the median can be the same. The variation on one side of the median needs to be balanced by the variation on the other side; the data are approximately symmetrical, with no outliers on only one side.

**EXPLORE MORE**

Operating cost per seat gives efficiency in dollars per seat hour. To change to dollars per seat mile, divide by speed in miles per hour. According to this formula, the most efficient aircraft is the B737-800, with 149 seats, speed of 454 kt, operating cost of \$1665/h, and efficiency \$0.0185664/seat mile. Note that Fathom originally calculated the formula in dollars per [seat] meter. When you change the *m* to *mi* in the units, Fathom calculates the conversion.

Speed	Efficiency
mi/h	dollars/mi
	$\frac{\text{Operating\_Cost}}{\text{Seats}}$
	$\frac{\text{Speed}}{\text{Speed}}$

Some student equations will not work for planes that have no passenger seats. You might suggest they delete these cases when they look for efficiency formulas for passenger aircraft. Units on some formulas may be confusing; Fathom works in nautical miles and knots rather than miles per hour.

**EXTENSION**

Arthur Hailey wrote *Flight into Danger*, a screenplay that first aired in 1956. It launched his hugely successful career as an author, creating blockbuster movies and even spoofs of those movies. Research the impact of *Flight into Danger* and Hailey's work, and reflect on how society's mixed feelings toward technology still provide plots for today's popular books and movies.

# Spread and Shape—Dice

People have created many games of chance with numbered cubes. To help you analyze the chance of winning these games, Fathom can simulate how dice and their sums behave. Although simulations of random chance will vary in the short term, they will show you what can happen in the long run. You will see that data sets representing the possible outcomes in two games can contain the same elements but have different chances of winning.

- Q1** What are the mean and the median of the data set  $\{1, 2, 3, 4, 5, 6\}$ ?
- Q2** If you roll a fair die 600 times and record the face value (1, 2, 3, 4, 5, or 6) for each roll, roughly how many times would you expect to roll a 1? What would you expect the mean and median of the face values for 600 rolls to be?

## INVESTIGATE

When you start Fathom, a blank document will open. You want to set up a collection that holds 600 simulated rolls of a die.



Attribute names can't have spaces. An underscore can be used in place of a space.

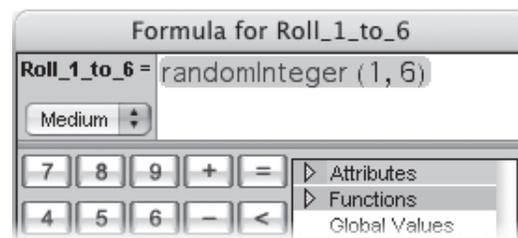
1. From the object shelf, drag a new collection (box of golden balls) to the window. Double-click on the picture of the box to bring up a new window labeled Inspect Collection 1.

If you see a Rename Collection dialog box, then you double-clicked on the collection's name rather than on its picture. You might want to give the collection a new name before you click again on the collection.

2. Data in a Fathom collection have one or more attributes. To define an attribute, click on *<new>* in the inspector's Attribute column. Then enter the name of the attribute. In this case, enter `Roll_1_to_6`.

To get actual numbers into that attribute, you'll use a formula to generate random integers from 1 to 6, representing the face values on a die.

3. Double-click opposite `Roll_1_to_6` in the Formula column. Fathom's formula editor will open. Type `randomInteger(1, 6)` and click **OK**. Move the inspector out of your way, but do not close it.



4. Click on the collection to select it. Choose **Collection | New Cases** and add 600 new cases.

5. Drag a graph from the shelf into your document. Drag the attribute *Roll\_1\_to\_6* from the inspector to the graph's horizontal axis. With the graph selected, go to **Graph | Plot Value** and type the formula `mean( )`. Click **OK**. In the same way, plot the value for the median.
6. It may be difficult to count all those dots. A bar chart with the heights marked might be more useful. Drag down a new graph. Again drag *Roll\_1\_to\_6* to the horizontal axis, but this time hold down the Shift key as you release the mouse button. You have a bar chart!

**Q3** Do these graphs show the values you expected? Explain.

You have done one simulation of 600 cases. You can use Fathom to repeat the simulation many times.

The keyboard shortcut Ctrl+Y (Win) or ⌘+Y (Mac) will come in handy.

7. From the **Collection** menu, choose **Rerandomize**. A new random number will be calculated for each case, and the graphs will be updated. Rerandomize the collection repeatedly.
- Q4** Describe the general shape shared by all the different graphs generated by rerandomizing. What values does the mean have? What values did you get for the median? What three values can the median have?

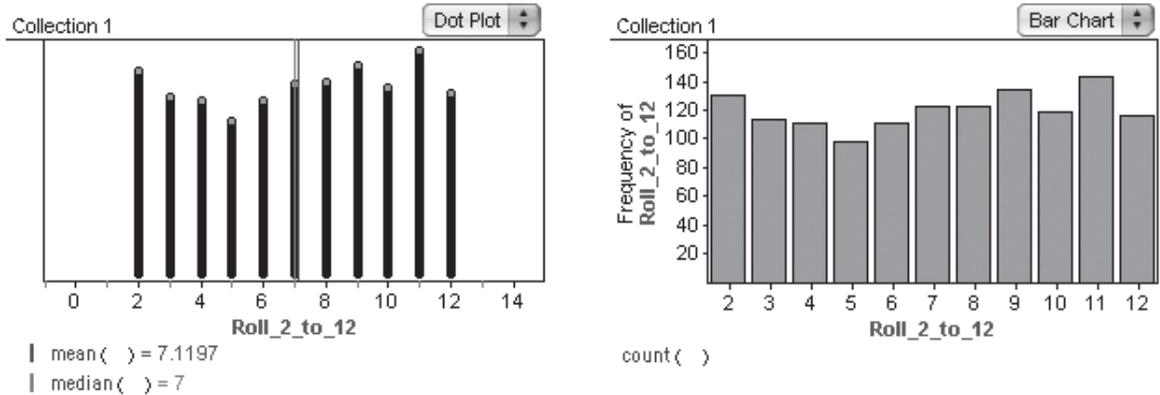
Finding the sum of the rolls is often important in games. A sum of two dice would be one of the values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12.

**Q5** What would you expect the mean and median of the sum to be?

You could simulate the set {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} using random numbers from 2 to 12.

8. Just as you did for rolling 1 to 6, create a new attribute called *Roll\_2\_to\_12*, and enter the formula `randomInteger(2, 12)`. Add several hundred new cases to simulate at least 1000 rolls.

9. Create a dot plot of *Roll\_2\_to\_12* and plot the value(s) of the mean and median. Create its bar chart as well. Rerandomize the values in the collection repeatedly and observe the changes in the dot plot and bar chart.



- Q6** What are the mean and the median face values of these data sets of more than 1000 rolls? Does this agree with your answer to Q5?
- Q7** The *range* of a data set is the difference between the maximum value and the minimum value. As you rerandomize, does the range change?
- Q8** As you rerandomize, the mean changes. What was the smallest and largest value you observed for the mean?
- Q9** How well do you think this formula simulates the sum of two dice?

Another way to simulate the sum would be to add the results for two individual dice.

10. Create a third attribute, *Add\_Rolls*. Give it a formula that simulates two dice and adds their values.
- Q10** What formula did you use?
11. Create both the dot plot (with the mean and median) and the bar chart, as you did for the previous two collections. Rerandomize repeatedly to explore the general features of this collection's dot plot and bar chart.
- Q11** What are the mean, median, minimum value, maximum value, and range of *Add\_Rolls*? Measures of data like these numbers are called *summary statistics*, because they can be used to summarize the behavior of a lot of data (such as more than 1000 rolls of a die). How do these summary statistics compare with those for randomly choosing a number between 2 and 12?

A summary statistic is useful in keeping track of thousands of values, but it can hide or ignore much of the information and thus can be misleading. For example, the

summary statistics we've been looking at don't describe the shape of the data when graphed.

- Q12** How does the graph of *Add\_Rolls* compare with the graph of *Roll\_2\_to\_12*? Which graph would you say is more spread out (less bunched up in the center)? How do the graphs differ in showing the chances to win with a given roll?
- Q13** How well does *Add\_Rolls* simulate finding the sum of two dice rolled randomly?

### EXPLORE MORE

1. If you had three tetrahedral dice (4-sided), then the sum of the dice would always be between 3 (which is  $1 + 1 + 1$ ) and 12 (which is  $4 + 4 + 4$ ). How would the shape and spread of this distribution be different from that of a 6-sided die and that of the sum of two 6-sided dice? Write your prediction and then use Fathom to test it.
2. Plot the value  $s( )$  on your dot plots of *Add\_Rolls* and *Roll\_2\_to\_12*. This is the spread measure called *standard deviation*. How does the value of  $s$  relate to how much the graph is spread out (that is, less bunched up in the center)?

**Objective:** Students will create bar graphs in Fathom to model dice games involving random chance; look at mean, median, maximum, minimum, and range; and interpret the results. Using Fathom to quickly create many simulations, students will see the importance of spread and shape when comparing distributions with the same center and range but with different shapes and spreads.

**Student Audience:** Pre-algebra, Algebra 1

**Activity Time:** 45–60 minutes

**Setting:** Paired/Individual Activity or Whole-Class Presentation (**Dice.ftm** is optional for either setting.)

**Mathematics Prerequisites:** Students understand the terms *mean*, *median*, *range*, *maximum*, and *minimum* for a data set.

**Fathom Prerequisites:** Students can start Fathom.

**Fathom Skills:** Students will learn to create a new collection and add attributes, create a formula involving the `randomInteger` function for an attribute, add cases to a collection, create dot plots and bar charts of numerical data, plot the value of statistical measures on a dot plot, and use **Rerandomize**.

**Notes:** This lesson is an excellent introduction to Fathom and its capacity to simulate probability events. Rather than starting with a table showing the cases in a collection, students start with a blank document and build a collection of dice rolls. Fathom’s dynamic interface will demonstrate the variability that is present in any model involving randomness. The `randomInteger` function is used to simulate an equal-probability event such as rolling a die. Fathom can display the cases as dice; the document **Dice.ftm** includes the dice display and the first attribute, `Roll_1_to_6`. Starting with that document might help students see that a case is one roll of the die. Or you can show this document and rerandomize a couple of times so students see what is happening. Enlarge the collection’s window so students can see the result of many of the 600 rolls. Then let groups work through the activity.

**For a Presentation:** You might use this presentation to introduce Fathom and show how it can be used to simulate a random event. As you facilitate the discussion, encourage several students to describe and explain what they see in

Q3 and Q4. Get several opinions on Q5 and Q9 and several comparisons for Q12. Ask for predictions before you (or a student running the computer) create a graph or add the mean or median (such as in steps 5, 9, and 11).

**Q1** The mean and median are both 3.5.

**Q2** Because each outcome is equally likely, you would expect to roll a 1 about 100 times of the 600. The mean of the face values should be close to 3.5, and the median will likely be 3 or 4.

## INVESTIGATE

- Holding down the Shift key forces Fathom to treat the data as categorical as opposed to numeric, so now 1 is treated as a discrete category rather than as a number. Notice that there is also a way to force Fathom to treat a non-numeric attribute as numeric.
- Being able to model rerolling the dice helps demonstrate the variation experienced. The faces do occur about 100 times each, but the graph is probably more ragged than some students might expect.

**Q3–Q4** In general, the graph has six bars, each with height close to 100 (plus or minus 25 or so). The mean is close to 3.5 (plus or minus 0.2 or so). The median is 3 or 4 with equal likelihood, but every now and again, when the number of rolls from 1 to 3 is exactly equal to the number of rolls from 4 to 6, the median is 3.5 (perhaps 3 times out of 100 tries).

**Q5** Mean and median will be close to 7.

**Q6** The median is almost always 7 for a large number of rolls. The mean is close to 7 (plus or minus 0.3 or so). Actual observations will vary.

**Q7** The range is  $12 - 2$ , or 10. It does not change.

**Q8** The mean is close to 7 (plus or minus 0.2 or so).

**Q9** Students will probably know that 7 is the most likely outcome for the sum of two dice and that 2 and 12 are the least likely, and that choosing a random number between 2 and 12 is not good for simulating the sum.

**Q10** `randomInteger(1, 6) + randomInteger(1, 6)`

**Q11** The mean is 7, the median is 7, the minimum value is 2, the maximum value is 12, and the range is  $12 - 2$ ,

or 10. These summary statistics are the same values as for *Roll\_2\_to\_12*.

In your discussion of this answer, you might ask about the need for other summary statistics that will be different for these two different data sets.

**Q12** The graph of *Roll\_2\_to\_12* has 11 bars of roughly equal height, whereas on the graph of *Add\_Rolls*, the middle bar (at 7) is the tallest, and the bars on either side get progressively shorter until 2 and 12, which have a height usually somewhat less than 30 for 1000 rolls. The graph of *Add\_Rolls* is more bunched up around the center; thus, the graph of *Roll\_2\_to\_12* has a more even distribution.

**Q13** It's a good model.

### EXPLORE MORE

1. This distribution would have a mean and median of 7.5 and would be even more mounded or bunched up in the center. Although the distribution of random integers from 2 to 12 is basically flat and the distribution of the sum of two 6-sided dice is like a tent (two lines meeting), this distribution begins to show the bell shape.

2. The standard deviations observed for *Roll\_2\_to\_12* will likely be between 3.05 and 3.25, whereas the standard deviations for *Add\_Rolls* will likely be between 2.3 and 2.6. This reflects the fact that the graph of *Roll\_2\_to\_12* is significantly more spread out than the graph of *Add\_Rolls*.

### EXTENSIONS

1. Write a formula for the mean of the face value of a die that is numbered consecutively from  $a$  to  $b$  and rolled a large number of times ( $n$ ). Check your formula against your answer to Q2 (when  $a = 1$  and  $b = 6$ ) and your answer to Q5 (when  $a = 2$  and  $b = 12$ ). Your formula will be in terms of  $a$  and  $b$ .

$$\text{Answer: mean} = \frac{a + b}{n}$$

2. Design a single die that would generate the same kind of distribution as the sum of two 6-sided dice. How many faces would it need? Which face values would be used, and how many times would each be repeated?

Sample answer: It would have 36 sides with 2 and 12 each appearing once; 3 and 11, twice; 4 and 10, three times; 5 and 9, four times; 6 and 8, five times each; and 7, six times.

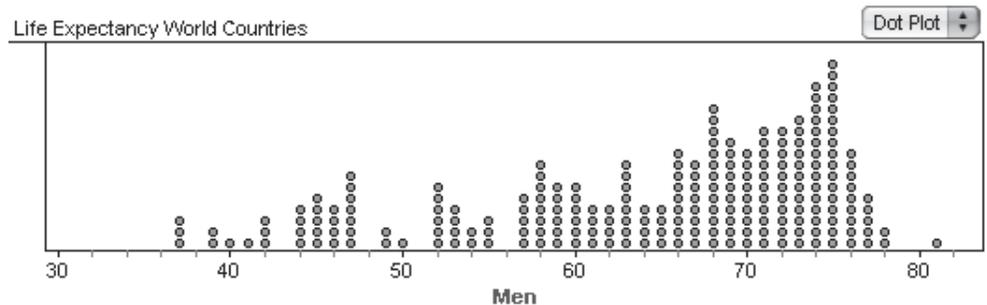
# Box Plots—Insurance

In this activity, you will advise an international insurance company how much to charge for life insurance in various countries. A company selling life insurance must be sure to earn more money in premiums than it pays in death benefits. To accomplish that goal, the company must predict how long customers will live. It can't know when any particular customer will die, but it can gather data about the average (mean) lengths of life for persons in various groups. These data are called *life expectancies*. Groups of people who have a longer life expectancy are charged lower life insurance premiums.

**Q1** Who do you think lives longer on average, men or women? Europeans or Latin Americans? Why?

## INVESTIGATE

1. Open the sample document **Insurance.ftm**. Scroll through the table to find the total number of countries recorded.
2. Drag a graph from the shelf and make a dot plot of all the life expectancies, either for men or for women.



To find the dot that represents the United States, highlight United States in the table. The corresponding dot will be highlighted.

United States (73 )

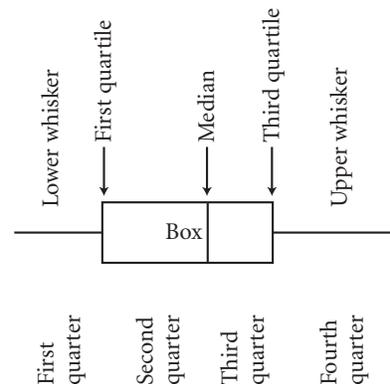
3. Move your mouse over the dots in the dot plot and notice that the country and life expectancy are displayed in the status bar at the bottom left of the window.
- Q2** Determine the maximum, minimum, and range of the life expectancies. Which country has the maximum life expectancy? The minimum?
- Q3** What is the value in the middle of the range? Do you think the mean and median will be close to the middle of the data? Why or why not?
- Q4** Do you expect the mean to be greater or less than the median? From the graph, what values do you predict for the mean and median? Explain your reasoning.
- Data like these are called *skewed left* because instead of being symmetric, the graph has a tail of data to the left.
- Q5** In which countries would you advise that life insurance premiums be the lowest? The highest?

## Box Plots—Insurance

continued

4. Because the data are broken down by sex, you suspect that recommendations for the sexes might differ. Drag a new graph from the shelf and create a dot plot of life expectancies for the other sex.
- Q6 Compare the dot plots for men and women. Identify at least two similarities and two differences between the dot plots.

5. A different kind of graph, called a *box plot*, allows other comparisons. To change a dot plot to a box plot, click on **Dot Plot** in the top right corner of the graph window and choose **Box Plot**. Box plots, sometimes called *box-and-whisker plots*, are based on the five-number summary—minimum, first quartile, median, third quartile, and maximum.



If you move your mouse over these quarters, Fathom displays information about them in the status bar (for example, “Lower whisker from 37 to 58”). The lower whisker is the first quarter, and the upper whisker is the fourth quarter. When you click on a quarter, the countries represented will be selected in the table and in both graphs. Click on each of the four quarters in one box plot and notice which countries are highlighted on the other box plot.

- Q7 Compare the four quarters on the two box plots. Is there evidence that one sex tends to live longer than the other? If so, does that tend to be true within the individual countries as well as overall?
6. Choose **Graph | Plot Value** to plot the mean and the median on each box plot.
- Q8 How good was your prediction in Q3? How do the mean and median compare for data that are skewed left?
7. Comparing two box plots can be difficult because the scales are different. To make this easier, you can put the two plots on the same graph. Drag the *Men* attribute from the table to the box plot for *Women*, aiming for the horizontal axis. A box will appear with a plus sign to the left. Drop the attribute on the plus sign. The mean for *Men* will appear on the graph.



- Q9 What do you notice that you couldn't see before?

8. Choose one region of the world. Scroll through the table to find that region and the total number of countries recorded.
- Q10** Which region did you choose?
  9. To make a graph of data from just the region you chose, select the box plot with life expectancies for both men and women and choose **Object | Duplicate Graph**. Drag the top border of the new graph so that you can easily compare the two graphs. Then choose **Object | Add Filter**. Enter a formula that limits the data to your region; for example, if your region is Latin America, enter the formula  $\text{Region} = \text{"Latin America"}$  and click **OK**. Your new graph will contain data only for the countries in your region.
  - Q11** How does the box plot of your region compare with the box plot of world data? Explain any differences. In particular, consider means, medians, and skew.
  - Q12** Should life insurance premiums for your region be more or less than those for the world as a whole? Defend your recommendation.
10. Examine at least one other region of the world. First, duplicate the graph for your region. Then, double-click the old filter and change the formula to limit data to the new region.
- Q13** Compare the data for this region with the data from your original region. Focus on mean, median, and skew. What advice would you give about life insurance premiums in the new region? Defend your opinions by referring to the graph.

## EXPLORE MORE

1. Investigate how life insurance premiums in the other regions should compare with your original region.
2. Create a new attribute called *Difference* with the formula  $\text{Women} - \text{Men}$ . Create the dot plot for this new attribute. Which country has the largest difference? Which has the smallest difference? How many countries have a negative difference? What does a negative difference mean? Is difference skewed left, skewed right, or symmetrical?

**Objective:** Students will use Fathom’s ability to quickly graph data, then they will compare box plots and characterize the data as skewed left, skewed right, or symmetrical. Fathom’s filters allow students to consider a subset of the data.

**Student Audience:** Pre-algebra, Algebra 1

**Activity Time:** 35–50 minutes

**Setting:** Paired/Individual Activity or Whole-Class Presentation (use **Insurance.ftm** for either setting)

**Mathematics Prerequisites:** Students can calculate the range of a data set.

**Fathom Prerequisites:** Students can create new graphs and plot values on graphs.

**Fathom Skills:** Students will change the graph type, create a box plot with two attributes plotted on the same scale, and add a filter.

**Notes:** Help students make connections as they see different ways of graphing and comparing data. Dot plots show the shape and spread, including skewness. Box plots draw attention to the middle of the data and show the spread but not the shape. How does each graph affect the comparisons of groups?

Ask groups who finish early to think about and present their ideas about the ability of the people from countries in their answer to Q5 to buy insurance. Would the company be likely to sell insurance in countries where the premiums would be highest?

As you listen to students talk about quarters, be aware that some will use the word *larger* to mean “more spread out.” Use the more exact wording, *more spread out*, and discuss how the box plot’s area seems to convey more than it actually represents.

Check that students are aware that axes need to be the same when comparing graphs.

This is an excellent opportunity for several groups to share answers to the same questions. Take time for these group presentations to increase your students’ ability to communicate about the relationships, the mathematics. Their presentations can lead to interesting discussions about what can be seen in different kinds of comparisons.

**For a Presentation:** As you present the activity, ask the student running the computer to create dot plots for both men and women, then skip step 4. Once students have given their opinion on the relationship between the mean and the median in Q4, you might plot the values of the mean and median on each dot plot. As students answer Q6, encourage many different answers. At Q10, to maximize the number of comparisons students will find as they answer Q13, ask for two or three regions and complete step 9 for those regions.

**Q1** Answers will vary. In their predictions, students might consider risky activities, health, standard of living, and other factors.

## INVESTIGATE

1. There are 229 countries recorded in the data set.

**Q2** For men: The maximum life expectancy is 81 years in Andorra; the minimum is 37 years in three African countries; and the range is  $81 - 37$ , or 44 years.

For women: The maximum is 87 years in Andorra; the minimum is 37 years in Zambia and Malawi; and the range is  $87 - 37$ , or 50 years.

**Q3** For men: The value in the middle of the range is  $37 + \frac{44}{2}$ , or 59 years. The mean and the median will be more than 59 years, because most of the data points are more than 59 years.

For women: The value in the middle of the range is  $37 + \frac{50}{2}$ , or 62 years. The mean and the median will be more than 62 years, because most of the data points are more than 62 years.

**Q4** The mean will be less than the median, because there are a few values much less than most of values, but there are no corresponding values much greater than the bulk of values.

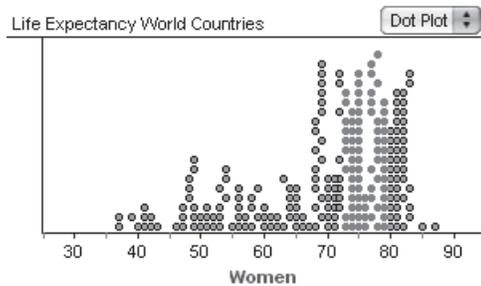
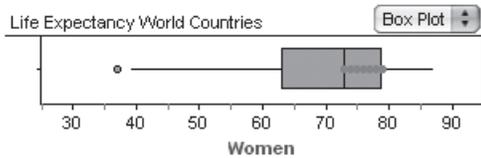
**Q5** Andorra has the highest life expectancy, so the insurance premiums should be lowest; Malawi and Zambia are among the countries with the shortest life expectancy, so their premiums should be highest.

**Q6** The two dot plots are both skewed left, and the minimum values are the same (37). The dot plot for *Women* has many more dots above 80, its maximum

value is larger, and it has fewer dots in the 60s and more in the 70s. (Other comparisons can be made.)

5. If students are having difficulty with the concept of a box plot, create a dot plot and a box plot of the same data. Click on each quarter on the box plot to highlight dots in red and then count the number of corresponding dots in the dot plot (in the diagram, the third quarter was selected on the box plot and the corresponding red dots appear without borders in art below). Each time, one-quarter of the dots will be highlighted on the dot plot, and the range of the highlighted dots will be determined by two numbers from the five-number summary for the data.

(Note: Some textbooks define and calculate the quartiles—the endpoints of quarters—in slightly different ways.)

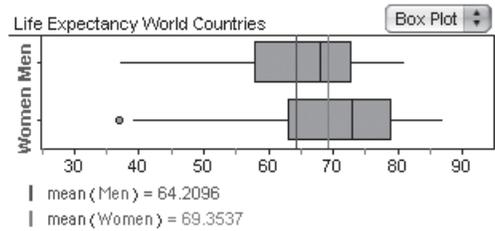


- Q7** The five-number summary for *Men* is 37, 58, 68, 73, and 81, and for *Women*, it is 37, 63, 73, 79, and 87. The median life expectancy for women is five years more than for men. Although the minimum values for the two are identical, the rest of the numbers for women are five or six years greater than the corresponding number for men. There is some evidence that, on average, women live longer than men. Most of the countries in a particular quartile of one box plot are in the corresponding quartile in the other box plot. This indicates that the life expectancy for men and women are ordered similarly.

- Q8** The mean is smaller than the median. The mean for men is 64.2 years, and the median is 68 years. Comparisons with students' predicted values will

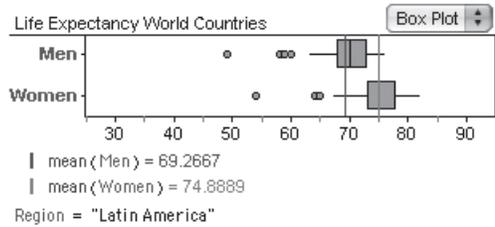
vary. Data sets that are skewed left have means that are less than their medians.

7. When the two attributes are graphed on the same box plot, the results will be similar to this one.



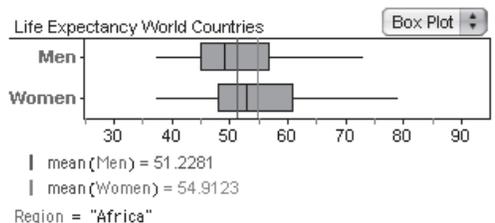
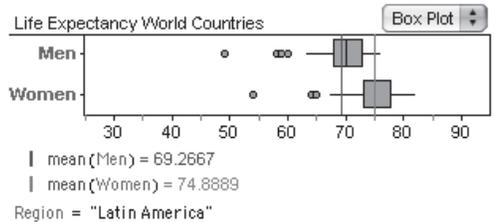
- Q9** Though the women's data are more spread out, the middle half of the data is equally spread out for both groups. The third quartile for *Men* is equal to the median for *Women*.

9. The duplicated graph will be similar to this one.



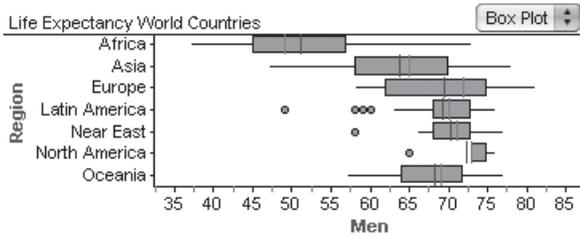
- Q10–Q12** Answers will vary.

- Q13** Sample answer: The middle half of the data for Africa is more spread out than it is for Latin America. Both the means and the medians are higher in Latin America. In Africa, most of the countries have short life expectancies, but a few have much larger values (skewed right); in Latin America, however, most of the countries have long life expectancies, with a few having smaller values (skewed left).



**EXPLORE MORE**

1. Many different graphs and calculations can be made.



Although the median life expectancy in Europe is higher than in Latin America, the means are comparable; otherwise, the most striking difference is the greater spread of the European values.

2. Largest are Belarus, Latvia, and Russia. The attribute *Difference* is smallest for Vanuatu. Nepal, Bangladesh, Zimbabwe, Zambia, and Malawi have the smallest difference in life expectancies. Five countries have a negative difference, meaning the life expectancy for men is higher than for women in those countries. The distribution is approximately symmetric.

**EXTENSION**

Most people can't afford to purchase life insurance. Find data on life expectancies for people of higher incomes in the various countries and regions and adjust the recommendations you'll make to the insurance company.

# Box Plots—Think Box

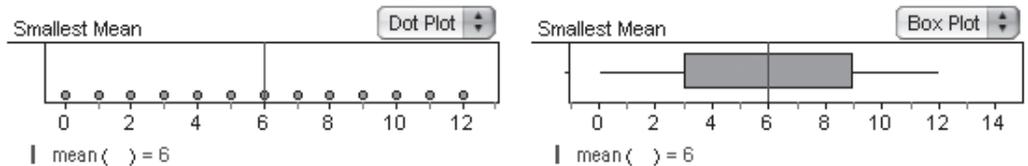
It can be fun to see how far you can go without anyone noticing. In this exploration, you'll play some games to see how much you can push around data points without changing their box plot.

- Q1** Draw a box plot on your paper for the data set  $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ . How much would the box plot change if the 12 became a 13?

## INVESTIGATE

Neither method is wrong; they just follow slightly different rules.

1. Open the document **Think Box.ftm**. You will see a collection, Original Data, together with a dot plot and a box plot of the data.
- Q2** Is the box plot that Fathom created the same as the one you completed by hand in Q1? It might not be. Some books outline methods of graphing box plots that are different from those used by Fathom. If your plot is different, explain how Fathom calculates the five numbers determining the box plot.
2. You will also see an exact copy of the original data in the Smallest Mean collection, together with its dot plot and box plot for the first game. Notice that the mean is computed.



- Q3** What is the mean of the data?

Here are two games to play with the data in the Smallest Mean collection. The object of both games is to move dots in the dot plot to get the smallest and largest values of the mean. Here are the rules.

### The Rules of Game 1

- You can drag only one dot at a time in the dot plot.
  - You cannot drag a dot past another one in the dot plot.
  - You cannot change any of the five-number summary values (minimum, first quartile, median, third quartile, or maximum). That is, the box plot should always look exactly like the box plot for the original data. The mean is not part of the five-number summary and is free to change.
3. Play Game 1 to find the smallest mean.
- Q4** What strategy did you use in playing the game?

## Box Plots—Think Box

continued

- Q5** What is the smallest value of the mean? Check your answer against those of others in the class. What numbers give the smallest mean?
- Q6** As another check, calculate the mean of your Smallest Mean data by hand and write it as a fraction in lowest terms.

$$\frac{0 + 0 + 0 + 3 + 3 + \cdots + 9 + 9 + 12}{13}$$

Highlight the original collection and choose **Object | Duplicate Collection**.



4. To find the largest value of the mean, make another copy of the data collection. Double-click the name of this new collection and call it Largest Mean.
5. Drag a new graph from the shelf. Double-click the Largest Mean collection to show its inspector. Drag the attribute *Value* to the graph to create a dot plot.

Attribute	Value	Formula
value	3	
<new>		

Use **Graph | Plot Value** to plot `mean( )`. Now drag down a new graph and make the corresponding box plot. You have duplicated the original situation.

- Q7** Find the largest possible value of the mean, following the rules of Game 1. What values will make the largest possible mean? Write the mean of these values as a fraction in lowest terms.
- Q8** What is the difference between the largest and smallest possible values of the mean for Game 1? Express this as a fraction in lowest terms. Devise a shortcut for finding this difference.
- Q9** Compare the completed dot plots for Smallest Mean and Largest Mean. Which of the two graphs has the bigger spread? Justify your answer.

The object of Game 2, like Game 1, is to find the largest and smallest values of the mean, but the rules are a little different.

## The Rules of Game 2

Follow all the rules of Game 1, except you may change the value of the median. None of the other four values of the five-number summary can change.

6. Click on the dot plot for the Largest Mean collection and use **Graph | Plot Value** to plot `median( )`.
- Q10** Use this dot plot to play Game 2, finding the largest possible value for the mean under the relaxed rules. What strategy did you use? What is the largest possible

## Box Plots—Think Box

continued

---

value for the mean now? What is the median value when the mean is the largest possible? Is this graph skewed left or skewed right?

- Q11** Without moving any dots, determine the smallest possible mean and the corresponding value of the median.

### EXPLORE MORE

1. Find examples of data sets that can be reconstructed from seeing just the box plot and knowing how many data points are in the data set. Make some conjectures about the kinds of sets that can be reconstructed from their box plots. Find examples of data sets that can be reconstructed from knowing the number of data points, the box plot, and the mean.
2. Create a data set and a set of rules different from those in this activity, then challenge other students to play.

**Objective:** Using the ability to drag data points in Fathom, students will explore summary data and realize that it does not reflect all the information in a data set. Because Fathom allows students to change data by dragging dots on a dot plot, students can visualize changes that affect the box plot.

**Student Audience:** Pre-algebra, Algebra 1

**Activity Time:** 40–60 minutes

**Setting:** Paired/Individual Activity (use **Think Box.ftm**) or Whole-Class Presentation (use **Think Box Present.ftm**)

**Mathematics Prerequisites:** Students can calculate the five-number summary for a data set and create a box plot.

**Fathom Prerequisites:** Students can create new dot plots and box plots, duplicate objects, and plot the value of statistical measures on graphs.

**Fathom Skills:** Students will rename a collection and use the inspector as a source of data attributes.

**Notes:** Box plots are useful for describing the center and spread of data, but the data can change without affecting the box plot. To set values precisely, students might find that double-clicking on a dot and typing the exact value in the inspector works best.

As you watch pairs of students work through the activity, choose groups that have slightly different answers to present their methods and results. Move down to the lowest mean and up to the highest in the order of presentations. Talking about methods is as important as talking about results.

**For a Presentation:** To orchestrate a whole-class discussion, use **Think Box Present.ftm**, which already has the Largest Mean collection. Take suggestions from the class on which dots the student running the computer should move. Build your discussion around the questions, asking students to make predictions and then using Fathom to check them out. You might assign Explore More 1 as homework after the whole-class discussion.

- Q1** Student box plots may look like the one in **Think Box.ftm**. Fathom calculates the five-number summary as 0, 3, 6, 9, and 12; the median is considered part of both the lower half of the data and

the upper half of the data. If 12 became 13, the right whisker would become 1 unit longer.

## INVESTIGATE

- Q2** Answers will vary. For example, Fathom includes the middle value, 6, in the lower half of the data when calculating the first quartile, but some conventions identify the first quartile by finding the middle number of those numbers less than 6.
- Q3** The mean of the data is 6.
- Q4** As soon as 0, 3, 6, 9, or 12 is moved, the box plot changes shape and the rules are broken. Move the two dots just above 0 down to 0, those just above 3 down to 3, and so on.
- Q5** The smallest mean is about 5.07692.
- Q6** Using the numbers 0, 0, 0, 3, 3, 3, 6, 6, 6, 9, 9, 9, 13, the smallest value for the mean in lowest terms is  $\frac{0+0+0+3+3+3+6+6+6+9+9+9+12}{13} = \frac{66}{13}$ .
- Q7** Using the values  $0 + 3(3) + 3(6) + 3(9) + 3(12)$ , the largest possible value for the mean is  $\frac{0 + 3(3) + 3(6) + 3(9) + 3(12)}{13} = \frac{90}{13}$ .
- Q8** The difference between the largest and smallest possible values is  $\frac{90}{13} - \frac{66}{13} = \frac{24}{13}$ . Both means have three each of 3, 6, and 9, so the difference will depend only on the 12; that is,  $3(12) - 12$ , or 24, will be in the numerator and 13 will be in the denominator.
- Q9** The two dot plots are mirror images of one another with the value 6 as the axis of reflection. Their spreads will be identical.
- Q10** The largest possible value for the mean is  $\frac{99}{13} \approx 7.61538$ , and it occurs when the median is 9. Move the data points that are not 0, 3, 9, or 12 as far to the right as possible without moving any dot past its neighbor. Six dots will be at 9, making the median 9.
- Q11** Starting with the Smallest Mean collection, move the three dots at 6 to the left by 3 units. This will make the total go down by 9, from 66 to 57. The smallest mean for Game 2 is  $\frac{55}{13} \approx 4.38462$  and will occur when the median is 3.

**EXPLORE MORE**

1. Data sets whose cases can be determined from a box plot include those with just five cases, those with six or seven cases and the median equal to one of the quartiles or with outliers, and those with any number of cases and all five-summary values the same. If you also know the mean, you can identify any six-member data set from its five-number summary.
2. Answers will vary.

**EXTENSIONS**

1. Compare the spread of the two graphs created for Game 1 and the graph created for Game 2. Plot the standard deviation,  $s( )$ , on the two graphs. The graph with the greater spread should have the greater standard deviation.

Answers:

Game 1 smallest mean:  $s \approx 3.95$

largest mean:  $s \approx 3.92$

Game 2  $s \approx 4.0$

2. Create a measure for spread that produces the same result for the data  $\{0, 0, 0, 3, 3, 3, 6, 6, 6, 9, 9, 12\}$  as it does for  $\{0, 3, 3, 3, 6, 6, 6, 9, 9, 9, 12, 12, 12\}$ .

Answers will vary. Sample answer: Measure the spread by taking the sum of the distances from the median and divide by the number of elements:

$$\frac{1}{n} \sum_{i=1}^n | \text{median} - x_i |, \text{ or } \frac{1}{13} (6 + 3 + 3 + 3 + 0 + 0 + 0 + 3 + 3 + 3 + 6 + 6 + 6) \approx 3.23$$



# Histograms—Ontario Communities

Governments collect census data to answer interesting questions about their populations. For example, Statistics Canada reports the median 2001 income in each community in the province of Ontario. What do the data say about communities in which the average annual income is above the median?

- Q1** How much money do you think the average person (15 years old or older) in your own community makes in a year?
- Q2** Do you think annual income is higher in cities than it is in towns? Explain.

## INVESTIGATE

1. Start Fathom and open **Ontario.ftm**.

The scatter plot shows 445 communities in Ontario, colored by median income. For example, in each community colored red, the median income is greater than about \$30,000.

- Q3** How does the scatter plot relate to the map of Ontario?



- Q4** On the scatter plot, about how many of the 445 communities have incomes greater than \$30,000?

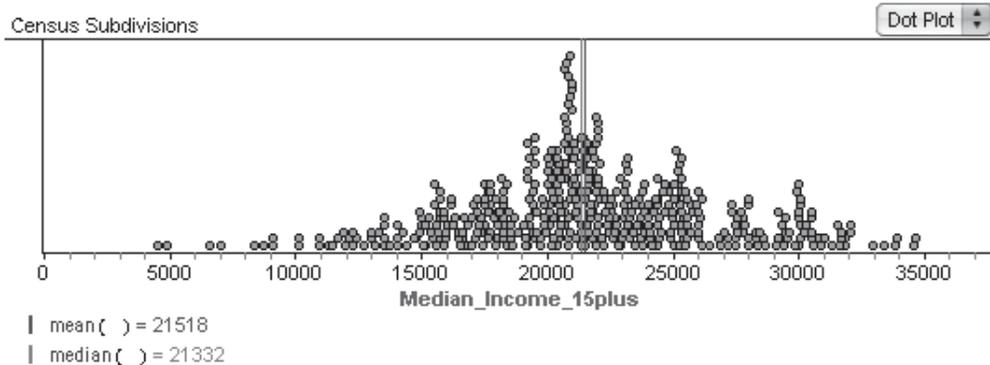
## Histograms—Ontario Communities

continued

Double-click the graph to show the inspector, or add a case table for the Census Subdivisions data and drag the attribute from the table.

- It's difficult to count the communities. Create a new dot plot of the attribute *Median\_Income\_15plus*. You can drag the attribute name from the graph's inspector.

- Q5** Is the dot plot skewed left, symmetrical, or skewed right? What does the shape indicate about whether the mean or the median will be larger?



With the graph selected, choose **Graph | Plot Value** and type `mean()`; repeat for `median()`.

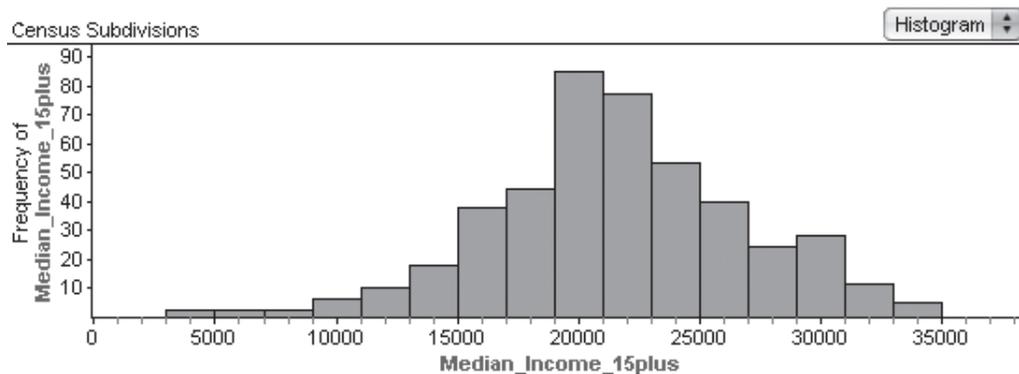
- Plot the value of the mean and the median on the dot plot.

- Q6** Are the values related in the way you indicated in Q5? You may need to drag to adjust the horizontal scale until you can see both lines.

- Q7** Is the median you plotted the median income for the entire province? Explain.

- Q8** From the dot plot, try to determine the number of communities that appear to have a median income greater than \$30,000.

- It may be easier to count these communities in a histogram. Choose **Histogram** from the pop-up menu at the top right of the graph.



The general shape of the histogram is similar to the shape of the dot plot, but now the data have been collected into groups, often called *bins*. The frequency, or number of communities in each bin, is represented by the height of the bin. You can see this frequency by looking at the scale on the vertical axis or by holding the

cursor over the bin and looking at the status bar. You can click on the bin to see those communities highlighted on the scatter plot.

**Q9** Move your mouse over the tallest bin and observe the information displayed at the bottom left of the Fathom window. In what interval of incomes does the greatest number of communities lie? How many communities have incomes in this interval?



5. Move your mouse between two bins until a double-tipped horizontal arrow is displayed.

Drag to change the bin widths. Notice that the vertical scale does not change as you do this. To get a better vertical scale, you can drag the scale details or simply reselect **Histogram** as the graph type.

**Q10** Adjust the histogram so that exactly eight bins are displayed and you can see all the data. Now what interval of incomes has the most communities in it? Do the mean and median values lie within this interval? How many communities are in this interval? From this graph, can you determine the exact number of communities with median incomes greater than \$30,000?

6. You may need more precise control over the histogram's display. Double-click on any white space on your histogram to show the graph inspector (or choose **Object | Inspect Graph**). Adjust the *binWidth* and *binAlignmentPosition* fields until the last bin starts at \$30,000. You may need to experiment with a few values to figure out how these fields affect the histogram. The graph will adjust after you change a field's value and press **Enter**. There is no need to hide the inspector. Select the bin or bins representing median incomes of \$30,000 or more, and they will turn red.

**Q11** What values of *binWidth* and *binAlignmentPosition* are you using? How does the exact count compare with your earlier estimates? What percentage of communities have a median income of \$30,000 or more?

7. To discuss locations of the communities with the highest incomes, look at the original scatter plot and notice the communities that are selected.

**Q12** Where on the map are the communities that have the highest median incomes? Where are the communities with the lowest median incomes?

## EXPLORE MORE

1. Duplicate your histogram and add a filter  $\text{Population\_2001} > 50000$ . How many communities have a population greater than 50,000? What percentage of the communities with population greater than 50,000 have a median income of

## Histograms—Ontario Communities

continued

---

\$30,000 or greater? How does this compare with your percentage in Q11?  
Does this provide evidence that people in cities make more money on average?  
Explain.

2. Go back to the scatter plot of Ontario that shows the latitude and the longitude of each community. Drag an attribute other than *Median\_Income\_15plus* to the center of the graph. What does this tell you about the population of Ontario? For example, how is family income or the value of dwellings distributed across the province?

**Objective:** Students will interpret data displayed in scatter plots and histograms. In Fathom, students explore how changing data affects the five-number summary used to create a box plot and how a histogram changes as the bin width changes.

**Student Audience:** Pre-algebra, Algebra 1

**Activity Time:** 20–35 minutes

**Setting:** Paired/Individual Activity (use **Ontario.ftm**)

**Mathematics Prerequisites:** Students can characterize a data set as skewed left, skewed right, or symmetrical.

**Fathom Prerequisites:** Students can open a Fathom document, use inspectors, create dot plots, and plot values on graphs.

**Fathom Skills:** Students will create histograms, adjust the display of a histogram, and select the cases within a bin of a histogram.

**Notes:** As you talk with students, ask them what they see in each type of display. What does the display show best? Which display is best to answer a specific question? Which questions does a particular display answer? As students give answers to Q7, ask why. For questions whose answers vary, choose a variety of displays to be shared. As you ask students to share a particular histogram, make sure they are able to copy that display and put it aside where it won't get changed. As students share, you might bring up the question of which histogram fairly portrays the data and for what purpose. Extensive information on the sources of the data in this activity is given in the Census Subdivision inspector **Comments** panel. A central point in the community was used for the latitude and longitude measurements.

- Q1** Answers will vary. At the time of the census, the Canadian dollar was worth less than \$0.75 American dollars, so CAN\$30,000 would be around US\$22,000.
- Q2** Answers will vary. Often, the suburbs have the highest income, and the answer might depend on how suburbs are treated as communities.

## INVESTIGATE

- Q3** Communities are concentrated in the large urban centers around Toronto, Ottawa, and Windsor, as well as along a line of industrial towns in northwest Ontario.
- Q4** Answers will vary.
- Q5** The dot plot appears approximately symmetrical. The mean and the median will be relatively close to one another.
- Q6** The mean and median are within \$200 of each other, with the mean larger than the median.
- Q7** No; this is the median of the median values in the communities. Because all communities are not the same size, large communities are underrepresented in the median of the medians.
- Q8** It is difficult to count these dots, even if students resize the graph. There appear to be approximately 30 communities. By creating a table and sorting by that attribute in the table, students will see that 29 communities have median incomes of greater than \$30,000.  
  
For example, the greatest number of communities in the histogram lies in the range  $19,000 \leq \text{Median\_Income\_15plus} < 21,000$ , which contains 85 communities.
- Q9** Answers will vary. One possible answer is that the greatest number of communities lies in the range  $19,000 \leq \text{Median\_Income\_15plus} < 21,000$ , which contains 85 communities, or  $20,500 < \text{Median\_Income\_15plus} \leq 25,000$ , which contains 163 communities. Students would be very lucky to create a histogram with \$30,000 as one of the divisions.
- Q10** Answers will vary. Most students will need to adjust the histogram further.
- Q11** One possible set of values is  $\text{binWidth} = 5000$  and  $\text{binAlignmentPosition} = 0$ . There are 29 communities with median income of \$30,000 or more. The percentage of communities in this range is  $\frac{29}{445}$ , or approximately 7%.

**Q12** Most of the higher incomes are found around the Great Lakes. A few of the communities with a low median income are in urban areas, but most of them are more remote.

### EXPLORE MORE

1. Thirty-seven communities have a population greater than 50,000. Eight of these communities have median incomes that are \$30,000 or more. Thus,  $\frac{8}{37}$ , or approximately 22%, of the larger communities have larger incomes, compared with only 8% of all

communities. This means that if you are in a larger community, you are about three times more likely to be in a community with a median income of \$30,000 or more.

2. Answers will vary.

### EXTENSION

Use other attributes in the collection to compare the median incomes of males 15 and older with females 15 and older. Is there evidence of a difference? If so, what factors might contribute to this difference?

# Ratios—Surveys

Your teacher will ask your class for opinions on some questions. Your goal is to use the responses of your classmates to estimate how all the kids in your school would answer those questions.

**Q1** Before you see the actual results, guess how many students in your class answered *Yes* to the first question and how many answered *Yes* to the second question. Now write at least three statements that summarize your guesses in different ways.

One way to compare two things is to use a ratio.

**Q2** For the first question your teacher asked, give two different ratios that include your guesses as to the number of *Yes* votes. Do the same for the second question.

## INVESTIGATE



You might choose different attribute names or omit the underscore between the *Response* and the 1 or 2, but attribute names can't contain spaces.



You can set up a data collection by entering data into a Fathom case table.

1. Open Fathom and drag a new case table from the shelf into the Fathom window.
2. To set up an attribute, click on *<new>* in the table and enter *Response\_1* for the attribute. Enter another attribute, *Response\_2*. If your teacher asked three questions, you would add a third attribute, *Response\_3*. Your teacher will now provide you with the results of the survey. Enter *Yes* or *No* (or *Y* or *N*) for each student response to the first two questions. When you are done, check that you have entered the correct number of responses.
3. Graphs will help you visualize and interpret the data. Drag a new graph from the shelf. Drag the attribute *Response\_1* from the case table to the graph. Drop the attribute under the graph's horizontal axis. You may be surprised to see a bar chart rather than a dot plot.

You get a bar chart because the responses are not numbers, they're *categories*. A bar chart allows you to see the number of data points in each category. By comparing the lengths of the bars, you can see which category holds more responses.

32 cases (80.0%) are No.

4. To see the actual numbers, watch the status bar at the bottom left of the Fathom window while you move the cursor over the bar that represents the *No* responses and then on the bar that represents the *Yes* responses.

**Q3** How many students answered *No*? How many said *Yes*?

The graph must be selected for the option **Object | Duplicate** to appear.

5. To approximate how much each category contributes to the whole, you might make a ribbon chart. Make a copy of the bar chart by choosing **Object | Duplicate Graph**.

Move the new graph to the side. Use the pop-up menu in the upper right corner to change the graph from a bar chart to a ribbon chart.

- Q4** What fraction of the ribbon represents *Yes* responses?
- Q5** Look again at the bar chart. Estimate a fraction that describes how large the *Yes* bin is compared with the *No* bin.
- Q6** Calculate each of the following ratios. Express your answer as both a reduced fraction and a decimal.

a.  $\frac{\text{Number of Yes}}{\text{Total number}}$       b.  $\frac{\text{Number of No}}{\text{Total number}}$       c.  $\frac{\text{Number of Yes}}{\text{Number of No}}$

d.  $\frac{\text{Number of No}}{\text{Number of Yes}}$       e.  $\frac{\text{Total number}}{\text{Number of No}}$       f.  $\frac{\text{Total number}}{\text{Number of Yes}}$

- Q7** Use complete sentences to explain the two ratios that compare the number of *Yes* responses (in the numerator) to another quantity.

- Q8** Which chart was better in helping you visualize each ratio?

6. Analyze the results for the second question. Make a bar chart and a ribbon chart with *Response\_2* on the horizontal axes.

- Q9** Use at least two ratios to summarize the class responses to the second question.

To use the data from your class to estimate how many students in your school would answer *Yes* to these questions, you'll be assuming that the responses from the whole school would have the same ratios as the responses in your class.

- Q10** Estimate how many *Yes* responses to each question would come from your entire school. Be sure to justify your reasoning and be clear about which ratio you chose and how you used it.
- Q11** Estimate how many *No* responses to each question would come from your entire grade. Explain.

## EXPLORE MORE

1. If your teacher asks your class a question that has three answers, use graphs and ratios to estimate how many students in your school would give each answer, assuming that your class is representative of the school.
2. Look for patterns in the responses. For example, did students who answered *Yes* on the first question also answer *Yes* on the second question?
3. Your own class might not be very representative of the entire school. How might you select a small group that would be more representative of all the students?

**Objective:** Students deepen their understanding of ratios as they look at different ratios from the same information. They will look at different ways to represent those ratios in bar charts and ribbon charts, which are quickly created in Fathom, and will use different ratios to solve proportions for an unknown.

**Student Audience:** Pre-algebra, Algebra 1

**Activity Time:** 25–40 minutes

**Setting:** Paired/Individual Activity (based on survey results from the whole class)

**Mathematics Prerequisites:** Students can reduce a fraction and convert a fraction to a decimal.

**Fathom Prerequisites:** Students can duplicate Fathom objects.

**Fathom Skills:** Students will open a case table and create a new attribute, a bar chart, and a ribbon chart.

**Notes:** Start the activity by asking two yes-no questions, such as those listed below. Students might write their answers on slips of paper. If you want students to do Explore More 1, ask a third question with three answers. Any questions will work, as long as not all students will give the same answer. While you gather and tally the responses, students can work on Q1 and Q2.

Report the responses student by student, providing the data for the table case by case. The quickest way might be to ask two students to work quickly to enter the data into a Fathom document, then have the whole class access that document. Alternatively for two questions, you could tally the ballots for the number of YY, YN, NY, and NN so students can quickly enter data and copy responses down rows.

You want questions for which students would like to hear the opinion of their peers, yet that don't reveal personal or sensitive information. Here are some suggestions:

- Should there be a student uniform that all students are required to wear?
- Would it be better to have fewer classes with longer periods? (Or the opposite question if you are on a block system.)
- Should the cafeteria serve more healthful options and fewer junk foods?

The third question for Explore More 1 should have three responses, such as *Earlier*, *Later*, or *Same time*. Here are some examples:

- Should school start and end earlier, later, or at the same time?
- Should (insert school function: dance, football game, band concert) start earlier, later, or at the same time?
- If you were in charge of the TV network, would you move the evening news (or a current show or a student-designated favorite show) to start earlier, later, or at the same time?

Many ratios can come from the same information. Using the right ratio at the right time is an important skill. If students simply take any ratio given and use it, they will often have poor results.

Different graphs of the same information can help students visualize that information in different ways. Both bar charts and ribbon charts can be used to display categorical information, but each allows you to see different ratios. Give students opportunities to explain why they are using a certain ratio or proportion. As you circulate, ask groups what they have done and why. During presentations, student groups can show their charts and give opinions on Q8.

**Q1** Answers will vary.

**Q2** Students could give the ratio of *Yes* to total responses, the ratio *Yes* to *No*, the ratio of total to *Yes*, or the ratio of *No* to *Yes*. The investigation will focus on the first two:  $\frac{\text{Number of Yes}}{\text{Total number}}$  and  $\frac{\text{Number of Yes}}{\text{Number of No}}$ .

## INVESTIGATE

The answers to the questions will depend on the survey results. Here are sample results from a survey with 40 responses, 8 *Yes* and 32 *No*, on the first question.

**Q3** 32 *No* responses; 8 *Yes* responses

**Q4** About  $\frac{1}{5}$  of the ribbon is *Yes*.

**Q5** The *Yes* bin is about one-fourth the size of the *No* bin.

**Q6** a.  $\frac{8}{40} = \frac{1}{5} = 0.2$     b.  $\frac{32}{40} = \frac{4}{5} = 0.8$     c.  $\frac{8}{32} = \frac{1}{4} = 0.25$   
 d.  $\frac{32}{8} = \frac{4}{1} = 4.0$     e.  $\frac{40}{32} = \frac{5}{4} = 1.25$     f.  $\frac{40}{8} = \frac{5}{1} = 5.0$

**Q7** Both  $\frac{1}{5}$  and  $\frac{1}{4}$  are ratios of *Yes* responses, but they compare the *Yes* response to different things. One-fifth of all responders thought that students should wear uniforms. One-fourth as many responders said *Yes* to uniforms as said *No* to uniforms. Some students may write sentences about the ratios that have *Yes* in the denominator.

**Q8** The ribbon chart shows the ratio of *Yes* to total, and the bar chart shows the ratio of *Yes* to *No*.

6. The following answers are for data with 16 *Yes* and 24 *No* responses for the second question.

**Q9** Any two of the ratios for *Response\_2* answers can be included, along with sentences explaining their meaning.

$$\frac{\text{Number of Yes}}{\text{Total number}} = \frac{2}{5}, \quad \frac{\text{Number of No}}{\text{Total number}} = \frac{3}{5},$$

$$\frac{\text{Number of Yes}}{\text{Number of No}} = \frac{2}{3}, \quad \frac{\text{Number of No}}{\text{Number of Yes}} = \frac{3}{2},$$

$$\frac{\text{Total number}}{\text{Number of No}} = \frac{5}{3}, \quad \frac{\text{Total number}}{\text{Number of Yes}} = \frac{5}{2}$$

**Q10** One way is to choose the first ratio and solve the proportion  $\frac{\text{Yes responses in class}}{\text{Students in class}} = \frac{x}{\text{Students in school}}$ . Students might choose different proportions for each question.

**Q11** One way is to choose the second ratio and solve the proportion

$$\frac{\text{No responses in class}}{\text{Students in class}} = \frac{\text{No responses in your grade}}{\text{Students in school}}$$

### EXPLORE MORE

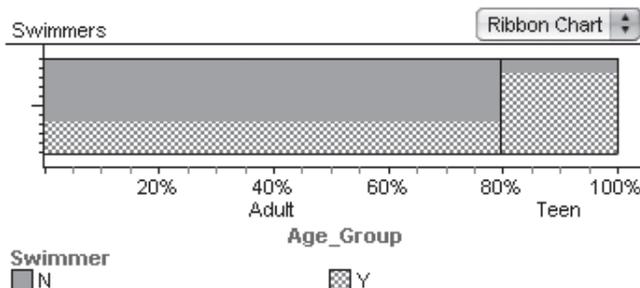
- Answers will depend on the data.
- Answers depend on class data.
- People cannot help thinking in terms of patterns, so any group they deliberately select is likely to have one or more things in common. Statisticians use randomness to reduce this bias. In a random selection, each subject has an equal chance of being selected. The selection is not based on any characteristics the group has in common. Randomness will not definitely eliminate all commonalities, but it is as fair as we can be. Any nonrandom selection process may shift the results of the survey one direction or another.

### EXTENSIONS

- Add the second attribute by dropping it in the middle of the ribbon chart of the first response. What can students say about this graph? What ratios appear here, and how are they interpreted?
- If the two questions you asked are related (such as *Response\_1* indicating opinions about longer classes and *Response\_2* asking whether school should start earlier), students might look at and analyze the numbers that responded *Yes* to both, *No* to the first and *Yes* to the second, and so on.

# Weighted Average—Swimmers

As part of a study on the impact of adding a swimming pool at the middle school, a student gathered these data on a sample of 200 people from her community, grouped by age (adult or teen) and by ability to swim (N = can't swim; Y = can swim).



- Q1** From the graph, estimate the percentage of adults who can't swim and the percentage of teens who can't swim.
- Q2** Now estimate the percentage of the entire group who can't swim.

## INVESTIGATE



The meaning of the numbers appears below the summary table.

You might need to drag a corner of the summary window to see the second number in each box and the description below the table.

- To see the data for these 200 people, open **Swimmers.ftm**.
- To see statistics that summarize a data set, drag a new summary table into the document.
- Drag the *Swimmer* attribute from the case table to the summary table. Drop it over the arrow pointing down. Then drag the *Age\_Group* attribute from the table and drop it over the arrow pointing to the right. Each box of the table now contains a number.
- Three of the numbers appearing in the summary table are 107, 40, and 86. Use a complete sentence to describe what each number means.
- By hand, calculate the percentage of the individuals who are teens.
- You can have Fathom calculate the fraction of each category in the column. Select the summary table and choose **Summary | Add Formula**. Enter the single word `columnProportion`. (The word will change color when it is spelled correctly.)
- Use complete sentences to describe the meaning of the table's numbers 0.825, 0.66875, and 0.57 (or their percentage equivalents).
- What is the average (mean) of the ratio of adults who can't swim to all adults and the ratio of teens who can't swim

		Age_Group		Row Summary
		Adult	Teen	
Swimmer	N	107	7	114
	Y	0.66875	0.175	0.57
Column Summary		53	33	86
		0.33125	0.825	0.43
		160	40	200
		1	1	1

S1 = count ( )  
S2 = columnProportion

## Weighted Average—Swimmers

continued

to all teens? Compare your answer with the information in the table that says 57% of people can't swim. Explain why the numbers are the same or different.

Is there a way to find the ratio for the full group from the ratios of the parts? You might think of the average you just found,  $\frac{0.66875 + 0.175}{2}$ , as  $0.66875 \left(\frac{1}{2}\right) + 0.175 \left(\frac{1}{2}\right)$ . Think of  $\frac{1}{2}$  as being the *weight* given to each part. To get from the ratios of unequal parts to the ratio of the whole, you can give a weight of more than  $\frac{1}{2}$  to the larger part and a weight of less than  $\frac{1}{2}$  to the smaller part. The result is called a *weighted average*. Experiment to find what weights you should use for a weighted average in this study.



When two words have to be made into one word, programmers sometimes use CamelCase, using capital letters to help distinguish the words. You can use CamelCase rather than an underscore for your slider names.

- To experiment with different values for the weights, drag a slider from the shelf to your document. Be sure the name *V1* is selected and enter a name for the slider, such as *Weight\_for\_Adults*. Pull down another slider and give it a name like *Weight\_for\_Teens*.



- You want the slider values to be between 0 and 1, because two weights will add to 1. Double-click on a slider "thumb" to open the slider inspector. Set *Lower\_* to 0 and *Upper\_* to 1. Do the same for the other slider.

Inspect Slider	
Properties	
Property	Value
Weight_for_Adults	0.5
Max_updates_per_second	
Lower_	0
Upper_	1

- You can also make a slider whose value depends on the values of other sliders. Drag down a third slider, and name it something like *Weighted\_Ave*. Go to **Edit | Edit Formula** and enter the formula

$$0.66875 \cdot \text{Weight\_for\_Adults} + 0.175 \cdot \text{Weight\_for\_Teens}$$

- Q7** Explain this formula. Where do the numbers 0.66875 and 0.175 come from?

- Slide *Weight\_for\_Adults* and *Weight\_for\_Teens* until the value *Weighted\_Ave* equals 0.57. There is more than one possible pair of answers.

- Q8** What values did you find for *Weight\_for\_Adults* and *Weight\_for\_Teens*?

- In Q4, you found the ratio of the number of teens to the number of subjects in the whole group. Now find the ratio of the number of adults to the number of subjects in the whole group. The sum of these ratios should be 1. Use the fact that the sum is 1 to write an equation expressing one weight ratio in terms of the other; use the slider names *Weight\_for\_Adults* and *Weight\_for\_Teens*. Enter the formula in the appropriate slider.

- Q9** What formula did you use?

You might use 1 – *Weight\_for\_Teens*.

## Weighted Average—Swimmers

continued

---

10. Adjust the slider that can still be moved independently until the weighted average of ratios of parts equals the ratio for the whole.

**Q10** What weighted average gives you the ratio for the whole?

**Q11** Restate in words, and explain, the equation for the weighted average.

### EXPLORE MORE

1. The table also contains an *Age\_Range* attribute. In the graph and in the summary chart, replace *Age\_Group* with *Age\_Range*. Find a formula for the weighted average of the three swimmer ratios to equal 0.57, the ratio in the summary table.
2. A student has a homework average of 80% and a test average of 95%. Her teacher gives grades based on a weighted average of these two averages. The student's weighted average is 90%. Another student has a homework average of 90% and a test average of 70%. What is his weighted average? Demonstrate this with sliders.

**Objective:** Students will explore the ratios used in calculating a weighted average. Fathom calculates the ratios of two groups and the ratio for the total. Using sliders, students can determine the weights that make the weighted average of ratios of parts equal to the ratio for the total.

**Student Audience:** Pre-algebra, Algebra 1

**Activity Time:** 30–45 minutes

**Setting:** Paired/Individual Activity (use **Swimmers.ftm**)

**Mathematics Prerequisites:** Students can use ratios and average two numbers.

**Fathom Prerequisites:** Students can open a Fathom document.

**Fathom Skills:** Students will create and use formulas in two-way summary tables and use sliders with formulas.

**Notes:** Estimating the two ratios from the graph and recognizing the connection between the ratio and a percentage may be easy for some students but difficult for others. Encourage students to give complete answers to Q5–Q7 and Q11. Opportunities to explain what they are doing and the meaning of the result deepen students' understanding. As students share a variety of answers for Q8, ask how they can all be valid answers.

- Q1** Answers should approximate  $\frac{2}{3}$ , or 0.67, for adults and  $\frac{1}{6}$ , or 0.17, for teens.
- Q2** Some students may get 0.42 by averaging the averages rather than the actual ratio of 0.57. This question might cause some disagreement among students. Use discussion to spark interest in exploring this discrepancy.

## INVESTIGATE

- Q3** The group contains 107 adults who don't swim, 40 teens, and 86 people who do swim.
- Q4** Twenty percent, or  $\frac{1}{5}$ , of the group are teens.
- Q5** The ratio of teens who swim to all teens is 0.825; that is, 82.5% of teens are swimmers. Of the adults, 66.9% don't swim. The ratio of people who don't swim to the whole group is 0.57; equivalently, 57% of the group are not swimmers.

**Q6**  $\frac{0.66875 + 0.175}{2}$ , or approximately 0.42. Students may have different ideas about why this average differs from 57%. Through discussion, elicit the idea that the discrepancy is due to having different numbers of teens (40) and adults (160). If there had been the same number of adults and teens among the 200 people, then the average would have given the ratio for the entire group. Because there are different numbers of adults and teens, the average of the ratios for the parts is different from the ratio for the whole. The true ratio (57%) is closer to the ratio for the larger group (67%) than to the ratio for teens (18%).

**Q7** 0.66875 and 0.175 are the ratios of parts from the summary table. These will be weighted differently to find the ratio for the whole.

**Q8** Several answers are possible.

9. Eighty percent, or  $\frac{4}{5}$ , of the group are adults;  
 $\frac{4}{5} + \frac{1}{5} = 1$ .

$$\begin{aligned} \text{Weight\_for\_Adults} + \text{Weight\_for\_Teens} &= 1 \\ \text{Weight\_for\_Adults} &= 1 - \text{Weight\_for\_Teens} \end{aligned}$$

**Q9** In the *Weight\_for\_Adults* slider, enter  $1 - \text{Weight\_for\_Teens}$ , or in the *Weight\_for\_Teens* slider, enter  $1 - \text{Weight\_for\_Adults}$ .

**Q10**  $0.66875 \cdot 0.8 + 0.175 \cdot 0.2 = 0.57$

**Q11** Because adults make up 80% of the population, their ratio is multiplied by 0.80, and the teens' ratio is multiplied by 0.20. If you multiply the ratio for each group by the importance of that group (that is, the ratio of the number of people in that group compared with the total number), you get the overall ratio.

## EXPLORE MORE

- $0.175 \cdot 0.2 + 0.627 \cdot 0.55 + 0.76 \cdot 0.25 \approx 0.57$
- $(0.80)\left(\frac{1}{3}\right) + (0.95)\left(\frac{2}{3}\right) \approx 0.90$ , or 90%  
 $(0.90)\left(\frac{1}{3}\right) + (0.70)\left(\frac{2}{3}\right) \approx 0.77$ , or 77%

One way is to set the slider formula on the weighted average slider to  $0.8 \cdot \text{hw\_Weight} + 0.95 \cdot (1 - \text{hw\_Weight})$  and move *hw\_Weight* to 0.333 to get the formula to equal 0.90. Then, without moving *hw\_Weight*, set a formula to  $0.9 \cdot \text{hw\_Weight} + 0.7 \cdot (1 - \text{hw\_Weight})$ .

# Proportions—Veterans

Data from a sample of a population are often used to predict something about that population. For example, of 500 Iowa residents in a random sample taken from the 2000 U.S. census, 55 people had served in the military.

Veterans		
	N/A	135
Veteran_status	No Service	310
	Yes	55
Column Summary		500

The population of the state of Iowa in 2000 was 2,926,324. The sample might be used to estimate the number of veterans in Iowa. How good are estimates of the number of veterans based on this sample?

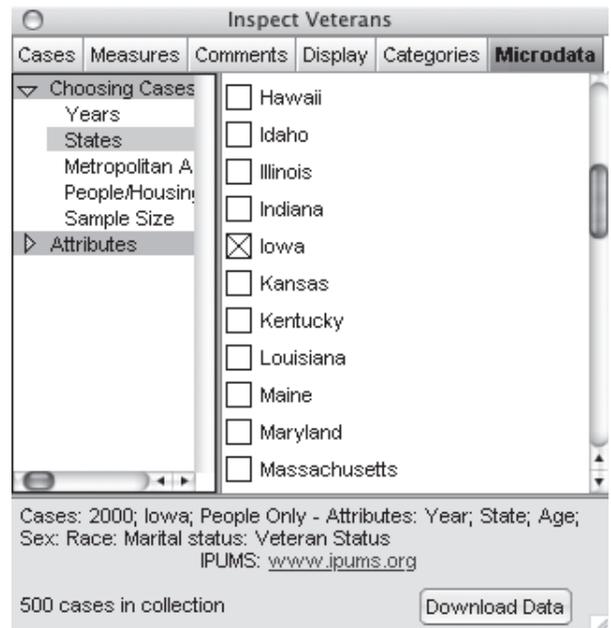
S1 = count ( )

- Q1** Set up a proportion stating that the ratio of veterans to the number of Iowans in the sample equals the ratio of Iowa veterans to the population of Iowa. A variable will represent Iowa veterans in your proportion. Solve this proportion to estimate the number of veterans in Iowa.
- Q2** The U.S. population in 2000 was 281,421,906. Use the sample data to set up and solve a proportion for estimating the number of veterans in the United States. What percentage are veterans?
- Q3** How accurate do you think these estimates might be? Explain.

## INVESTIGATE

You must be connected to the Internet to download data.

1. Open **Veterans.ftm**. You will see a collection (a box of balls) and a table and summary for Iowa data.
2. Look at data from another state. Double-click on the collection to show its inspector and click on the last tab, **Microdata**. Under Choosing Cases, choose States. You will see a list of states, with an X in the box beside Iowa. Click on the X to remove it, then click on the box beside a different state. Click the **Download Data** button.
- Q4** About how many veterans are in the state you chose?



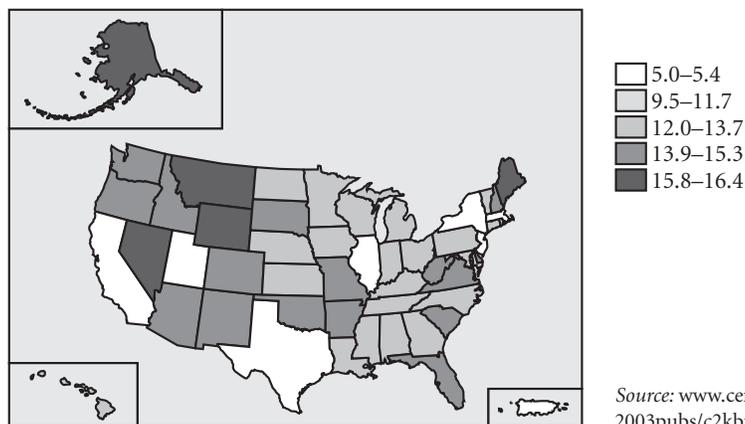
## Proportions—Veterans

continued

The collection State Populations has 2000 census data.

- Q5** According to the summary table, what is the ratio of veterans to residents of this new state?
- Q6** If you used this ratio and the U.S. population (281,421,906), what would be the estimated number of veterans in the United States? What percentage is that? How accurate is this estimate? Explain.
- Q7** Combine the information from Iowa and from your chosen state (giving you a sample of 1000 people) to give another estimate of the number of veterans in the United States. Is this estimate more accurate or less accurate than the estimate in Q5? Explain.
3. Return to the inspector and remove the X from your selected state. Now select a new state and download its data.
- Q8** What is the ratio of veterans to residents in this sample? About how many veterans live in your new state?
- Q9** Combine the information from Q7 with the information for your second state and Iowa (giving you a sample of 1500 people) and give a ratio of veterans and another estimate of veterans in the United States. Is this estimate more accurate or less accurate than your previous estimates? Why?
4. Look at the map, which shows how veterans are distributed across the United States. It shows why selecting some states might give an inaccurate estimate for the country.

Percentage of Population Who Are Veterans



- Q10** After looking at the map, how accurate do you think your latest estimate is?

Statisticians are very careful about drawing inferences, or conclusions, from a sample. Statistical inference is only valid when the sample is chosen randomly. A sample of 50 from Nevada is likely to have more veterans than a sample of 50 from the entire United States.

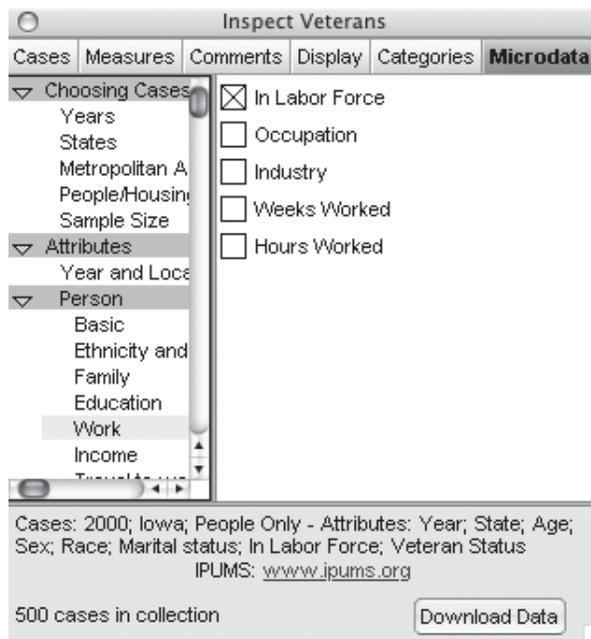
## Proportions—Veterans

continued

- Q11** Choose five states randomly and use this sample to estimate the percentage of veterans in the United States. Describe how you took your random sample and the estimate your sample predicts. How close do you come to the actual percentage of a little more than 9%?

### EXPLORE MORE

1. On the **Microdata** panel of the collection inspector, pick a state or states. Then select Attributes: Person: Work: In Labor Force and collect a set of data so you can estimate the ratio of persons in the labor force. Use that number to estimate the number of U.S. residents in the labor force.



2. Choose another attribute from the **Microdata** panel and explore those data from several states. Describe the patterns you find.

**Objective:** Students will deepen their understanding of setting up and solving proportions. Fathom allows students to gather census microdata about any of the 50 states and make estimates about statistics for the entire country. This activity is also an informal look at how biased sampling can affect predictions and estimates.

**Student Audience:** Pre-algebra, Algebra 1

**Activity Time:** 20–30 minutes

**Setting:** Paired/Individual Activity (use **Veterans.ftm**)

**Mathematics Prerequisites:** Students can solve a proportion and read and understand a graph key.

**Fathom Prerequisites:** Students can use a summary table.

**Fathom Skills:** Students will use a collection inspector to collect microdata from a U.S. census.

**Notes:** Computers used for this activity must have access to the Internet for sampling U.S. census data. Samples of individual cases are drawn from the 2000 census data in steps 2 and 3. Because only one state can be sampled at a time, students need to make new proportions by adding numbers from two, three, then five samples. Students might use technology to help them combine information, or they might use a pencil and paper. If you have students who make a table in Fathom or on their calculators, you might ask them to present it to the class. As you visit groups, ask for oral explanations to Q6 and Q7; encourage different explanations and note those that the class could benefit from hearing. You might ask groups to put their answers to Q9 on a class table. How do these estimates compare? Doing the same for Q11 could lead to some interesting class discussion. You might ask students with good methods of taking a random sample of states to share their methods with the class.

**Q1**  $\frac{55}{500} = \frac{v}{2,926,324}$ ; the estimated number of veterans is 321,896.

**Q2**  $\frac{55}{500} = \frac{v}{281,421,906}$ , so  $v = 30,956,410$ , or 11%

**Q3** The estimate for Iowa might be fairly accurate, because it was a random sample of Iowans. The estimated number of veterans in the U.S. will be accurate only if Iowa represents the whole country.

## INVESTIGATE

**Q4–Q9** Answers will vary depending on the states and the samples. Sample answers below are for adding data from New York and Minnesota. State populations are in a separate collection in the Fathom document.

**Q4** There are about 1,783,800 veterans in New York.

**Q5** New York:  $\frac{47}{500}$ , or 9.4%

**Q6**  $\frac{47}{500} = \frac{v}{281,421,906}$ ;  $v = 26,453,659$

This is somewhat lower than the answer using Iowa. Perhaps the true number is in between.

**Q7**  $\frac{102}{1000} = \frac{v}{281,421,906}$

$v = 28,705,034$ , or about 10.2%

This estimate is probably more accurate, because there is a larger, more diverse sample.

**Q8** Minnesota:  $\frac{65}{500}$ ; Veterans in Minnesota:

$\frac{65}{500} = \frac{v}{28,705,034} = 3,731,654$

**Q9**  $\frac{167}{1500} = \frac{v}{281,421,906}$ ;  $v = 31,331,639$

This is probably no more accurate than just Iowa. Minnesota and New York were on opposite extremes from Iowa.

**Q10** Different samples give different estimates. Because not all states are equally represented in the military, selecting only one or two states will bias the estimate for the country.

**Q11** In general, unbiased samples including more states give estimates closer to the 9.4% for the population, or about 26,454,000 veterans in the United States. Here are two good descriptions of the random sampling.

- Sample answer: We used a random number generator on our calculator to select numbers from 1 to 50 and used the case numbers from the population table to find the corresponding states.
- Sample answer: We used Fathom to take a random sample from the case table of State Populations.

## EXPLORE MORE

1. The actual number in the labor force is about 142,600,000 people.
2. Answers will vary.

# Dimensional Analysis—Fastest Animals

The book *Top 10 of Everything in 2004* lists the ten fastest animals with their speeds in meters per second. Some characteristics of these ten animals are given with measurement units. You will often find it useful to convert to different measurement units. Your goal is to find quick ways to do that conversion.

- Q1** The fastest animal is the cheetah, with a speed of 29.1 meters per second (m/s). How fast do you think that is in miles per hour (mi/h)?
- Q2** Is it possible to convert from m/s to mi/h in a single step? Explain why or why not.

## INVESTIGATE

1. Fathom can convert among many common measurement units. To see how, start by opening the document **FastestAnimals.ftm**.
2. Select the case table. Display the units by choosing **Table | Show Units**.
3. Select the attribute *Speed* by clicking on the attribute name. Choose **Edit | Copy Attribute**, then immediately choose **Edit | Paste Attribute**.

Fastest Animals

	Name	Speed1	Speed
<b>units</b>		mi/h	m/s
<b>1</b>	Cheetah	65.1 mi/h	29.1 m/s
<b>2</b>	Pronghorn antelope	55.0 mi/h	24.6 m/s
<b>3</b>	Mongolian gazelle	50.1 mi/h	22.4 m/s
<b>4</b>	Springbok	50.1 mi/h	22.4 m/s
<b>5</b>	Grant's gazelle	47.0 mi/h	21.0 m/s
<b>6</b>	Thomson's gazelle	47.0 mi/h	21.0 m/s
<b>7</b>	Brown hare	45.0 mi/h	20.1 m/s
<b>8</b>	Horse	42.9 mi/h	19.2 m/s
<b>9</b>	Greyhound	42.1 mi/h	18.8 m/s
<b>10</b>	Red deer	42.1 mi/h	18.8 m/s

Fathom abbreviates miles per hour as mph or mi/h.

4. In the units row of the new attribute, *Speed1* (to the left of *Speed*), change the unit to miles per hour.
- Q3** What is the speed of a cheetah in mi/h?

- Q4** Which of these animals could outrun a car going 50 mi/h?

When converting from one unit to another, you can use the idea that multiplying by 1 does not change a number. One inch (in.) equals 2.54 centimeters (cm); so the fraction  $\frac{2.54 \text{ cm}}{1 \text{ in.}}$  and the fraction  $\frac{1 \text{ in.}}{2.54 \text{ cm}}$  both equal 1; they are unit fractions. To change 30 cm to inches, multiply  $30 \text{ cm} \cdot \frac{1 \text{ in.}}{2.54 \text{ cm}}$ , which is about 12 in. To change 50 in. to centimeters, multiply  $50 \text{ in.} \cdot \frac{2.54 \text{ cm}}{1 \text{ in.}}$ , which is about 127 cm. Products of unit fractions are like those of numeric fractions: the units in the numerator of one cancel the same units in the denominator of another. But how do you know that 1 in. equals 2.54 cm, as claimed above? Fortunately, Fathom can help you find conversion factors like this.

5. In the case table, to the right of all existing attributes, click on *<new>* to create a new attribute called *Change*.

## Dimensional Analysis—Fastest Animals

continued

Fastest Animals

	Change
units	meters
1	

- Enter the units for this attribute as meters.
- For the first data item in this column, enter 1 mi. When you press **Enter**, the value will change to the number of meters that are in 1 mi.

**Q5** Based on this change, write two fractions that equal the number 1. (Remember to include the units.)

**Q6** Which of these fractions would you multiply by  $\frac{32 \text{ m}}{\text{s}}$  to change to  $\frac{\text{mi}}{\text{s}}$ ?

- Change the units of your new attribute to **seconds** and enter the first value of 1 min. When you hit **Enter**, the 1 min changes to the equivalent measure in seconds.

Fastest Animals

	Change
units	seconds
1	1 min

**Q7** How could you multiply  $\frac{29.1 \text{ m}}{\text{s}}$  by *two* fractions to change to  $\frac{\text{mi}}{\text{min}}$ ? Remember that a unit you do not want in the final results must appear in both a numerator and a denominator in order to cancel.

**Q8** How could you multiply  $\frac{29.1 \text{ m}}{\text{s}}$  by *three* fractions to change to  $\frac{\text{mi}}{\text{h}}$ ?

- Select the units on *Change* and delete them. Also delete the units on *Speed* and *Speed1*. For the attribute *Change*, enter the formula to change the units that you developed in Q8 (without the unit labels),  $\text{Speed} \left(\frac{?}{?}\right)\left(\frac{?}{?}\right)\left(\frac{?}{?}\right)$ .

**Q9** Compare the values of *Change* with the values in the other columns.

- Create another new attribute named *Ratio*. Select the column, choose **Edit | Edit Formula**, and enter the formula  $\text{Speed1}/\text{Speed}$ .

Fastest Animals

	Ratio
units	
=	$\frac{\text{Speed1}}{\text{Speed}}$

**Q10** How do the values in the new column relate to the conversion you found in step 9? Explain.

With the table selected, choose **Table | Show Formulas**.

Formula for Change

$$\text{Speed} \left(\frac{1}{1609.3}\right) \left(\frac{60}{1}\right)$$

## EXPLORE MORE

- Reenter the units in *Speed1* (mi/h). Next change the units to **mach** (the speed of sound). Then delete the units. How fast is the speed of sound in m/s? Explain how you found this value. How fast is the speed of sound in mi/h?
- Use Fathom to determine the conversion from the speed of sound (mach) to the speed of light (lightspeed).
- Drag a new graph from the shelf. From the case table, drag *Speed* to the graph's horizontal axis and *Speed1* to its vertical axis. Make sure that *Speed1* does not have units. Now choose **Graph | Add Movable Line** then **Graph | Lock Intercept At Zero**. Adjust the line to fit the data. How does the equation of this line relate to the conversion of *Speed* to *Speed1*?

**Objective:** Students will use ratios equal to one, called *unit ratios*, to convert units. As Fathom calculates the conversion factor, students will see validation of the conversion factor they found from dimensional analysis.

**Student Audience:** Pre-algebra, Algebra 1

**Activity Time:** 30–40 minutes

**Setting:** Paired/Individual Activity (use **Fastest Animals.ftm**)

**Mathematics Prerequisites:** Students understand units and can reduce fractions by dividing the numerator and denominator by the same quantity.

**Fathom Prerequisites:** Students can open Fathom documents and add new attributes and their own formulas.

**Fathom Skills:** Students will duplicate an attribute, change or delete attribute units, and add a line through (0, 0) to a scatter plot (Explore More 3).

**Notes:** The first two questions are for thought and need not be discussed beyond the pair or small group working together. Return to question Q2 at the end of the activity to make sure all students agree with the answer. As you visit groups, ask why. Why does  $\frac{1 \text{ mi}}{1609.34 \text{ m}} = 1$ ? Why does  $\frac{60 \text{ s}}{1 \text{ min}} = 1$ ? Why can you use a ratio that equals 1 (a unit ratio) to convert a measurement to new units? Why can you cancel units? How can you be sure you are using the correct ratios?

These data come from the book *The Top 10 of Everything in 2004*; publication information can be found on the

**Comments** panel of the collection's inspector. The book is a good source of small data sets to study and graph.

**Q1** About 65 mi/h

**Q2** Yes; any units that can be converted in multiple steps can be converted in one step by combining all the steps into a single expression, a unit ratio.

## INVESTIGATE

**Q3** 65.1 mi/h

**Q4** The top two (cheetah and pronghorn antelope) would outrun the car, and the next two (Mongolian gazelle and springbok) would keep up with the car.

**Q5**  $\frac{1609.34 \text{ m}}{1 \text{ mi}}$  and  $\frac{1 \text{ mi}}{1609.34 \text{ m}}$

**Q6**  $\frac{1 \text{ mi}}{1609.34 \text{ m}}$

**Q7**  $\left(\frac{29.1 \text{ m}}{\text{s}}\right)\left(\frac{1 \text{ mi}}{1609.34 \text{ m}}\right)\left(\frac{60 \text{ s}}{1 \text{ min}}\right) = \frac{1.085 \text{ mi}}{\text{min}}$

**Q8**  $\left(\frac{29.1 \text{ m}}{\text{s}}\right)\left(\frac{1 \text{ mi}}{1609.34 \text{ m}}\right)\left(\frac{60 \text{ s}}{1 \text{ min}}\right)\left(\frac{60 \text{ min}}{1 \text{ h}}\right) = \frac{65.095 \text{ mi}}{\text{h}}$

**Q9** The values of the *Change* attribute now match those of *Speed1*. With the Change column selected, choose **Table | Format Attribute** and set the attribute to a fixed decimal of one decimal place. Because of rounding, one of the measurements is slightly off.

**Q10** This column is the conversion factor for m/s to mi/h. It is also the product of  $\left(\frac{1 \text{ mi}}{1609.34 \text{ m}}\right)\left(\frac{60 \text{ s}}{1 \text{ min}}\right)\left(\frac{60 \text{ min}}{1 \text{ h}}\right)$ .

## EXPLORE MORE

- 1 mach = 331.46 m/s = 741.455 mi/h
- 1 lightspeed = 904,460 mach
- The slope of the line is the conversion factor that you have found. (The intercept of the line should be 0.)



# Proportions—Squirrel Population

To estimate populations in the wild, naturalists use a technique called *capture-recapture*. They capture and tag some of a population before releasing them. After giving the tagged animals some time to mix in with the others, they make another round of captures. They carefully note how many of those captured in the second round were tagged in the first round. They get data like these.

Squirrels Caught	
Both first and second	15
First only	270
Second only	153
Total	438

- Q1** How many squirrels were captured and tagged in the first round? How many were captured in the second round? How many caught in the second round were already tagged?

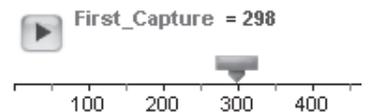
Naturalists calculate the total population using a proportion that equates the ratio of the number of squirrels caught in the second capture to the number of those second-capture squirrels who are tagged with the ratio of the total population to the total number who were tagged on the first round.

- Q2** What proportion that uses the three values from Q1 and a variable for the number in the total population can be used to estimate the population? What is the estimate of the total population?

## INVESTIGATE

In this investigation, you will explore how naturalists decide how many animals to capture in each round.

1. Open the document **Squirrels.ftm**, which contains a collection consisting of the population of squirrels.
2. There is a slider named *First\_Capture*, which allows you to capture up to 500 squirrels in the first round. Set this slider to something near 300.
3. There is a second slider named *Second\_Capture*, which allows you to capture up to 500 squirrels in the second round. Set this slider, also, to something near 300.
4. The first and second capture will happen each time you click on the **Sample More Cases** button. Do that now. The results of the second capture appear in the graph and the summary box.

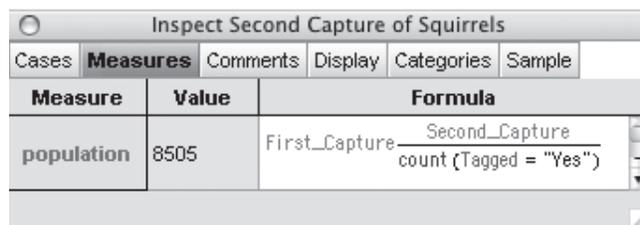


## Proportions—Squirrel Population

continued

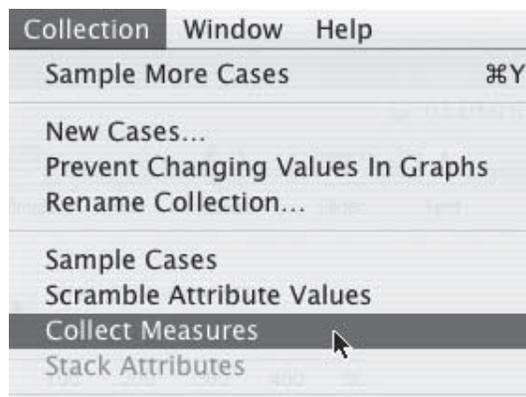
- Q3** Use each of the numbers in the summary box to write a sentence about what happened in the second capture.
- Q4** Write a proportion using this information, the size of the first capture, and the unknown size of the full population. Then solve this proportion to predict the size of the squirrel population.
- Q5** Compare your result with those of others around you. How satisfied are you with your prediction?
5. Set both of the sliders to only about 100 squirrels, then sample more cases.
- Q6** Calculate the population size based on these values. Compare these results with those of others around you or repeat the simulation and compare the results.
6. You can build a population calculator. Open the inspector for the second capture collection and click on the tab for **Measures**. Enter a new measure named **Population**. Rewrite this equation so that *Population* is alone on the left side:  $\frac{\text{Population}}{\text{First\_Capture}} = \frac{\text{Second\_Capture}}{\text{count}(\text{Tagged} = \text{"Yes"})}$ . Enter that formula.

Click on the Second Capture collection to open the inspector.



Inspect Second Capture of Squirrels		
Cases	Measures	Comments
Measure	Value	Formula
population	8505	$\frac{\text{First\_Capture} \cdot \text{Second\_Capture}}{\text{count}(\text{Tagged} = \text{"Yes"})}$

7. Select the Second Capture **Collection** | **Collect Measures**.



This will create another collection of population estimates. Each estimate is based on a new simulation of the capture-recapture procedure.

# Proportions—Squirrel Population

continued

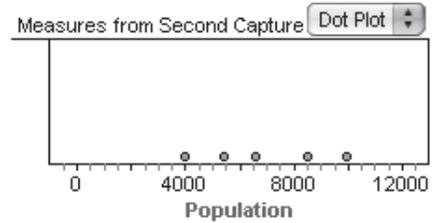
Highlight the Measures collection and drag a table from the shelf.



8. Create a case table of these measures and a dot plot of *Population*.

Measures from Second Capture of Squirrels

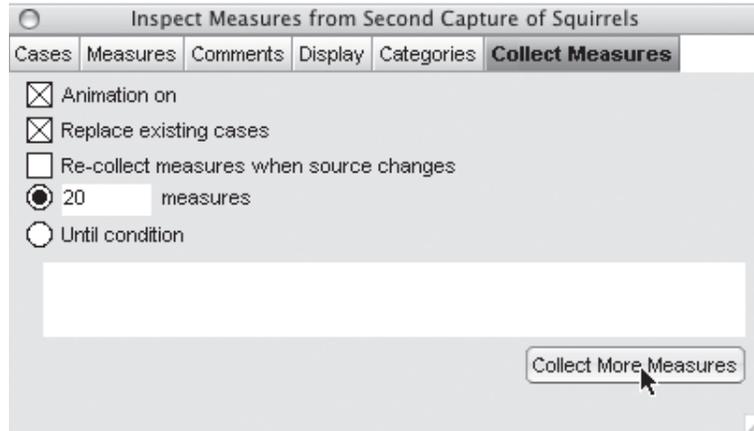
	Population	<new>
1	9922.5	
2	6615	
3	3969	
4	5412.27	
5	8505	



**Q7** Interpret the results of this graph.

Once you understand where the measures are coming from, uncheck the **Animation on** box.

9. Open the inspector for the Measures collection; under the **Collect Measures** tab, set it to replace the existing cases and collect 20 measures. Reset one or both of the sliders, then collect more measures.



**Q8** After exploring different sample sizes, what recommendations would you make about the sample size for an accurate estimate of the population?

Imagine that each capture costs \$10 and the budget for the capture-recapture project is \$6000.

10. Use your sliders to discover if it is better to make a large capture first or second or the same size each time. Remember that you are only going to capture a total of 600 squirrels.

**Q9** What is your recommendation to the capture team?

## EXPLORE MORE

IQR =  $Q3 - Q1$ . Here, Q3 is the third quartile and Q1 is the first quartile.

Calculate the interquartile range (IQR) of the 20 simulations using 600 captures total on each simulation. Then calculate the IQRs for simulations using 200, 400, 800, and 1000 captures. Summarize what you found.

**Objective:** As students simulate several capture-recapture scenarios, they will use ratios and proportions to make estimates. Students will see the effect of different scenarios on the accuracy of the estimate.

**Student Audience:** Algebra 1

**Activity Time:** 30–45 minutes

**Setting:** Paired/Individual Activity or Whole-Class Presentation (use **Squirrels.ftm** for either setting)

**Mathematics Prerequisites:** Students can set up and solve a proportion, and they understand what sampling is.

**Fathom Prerequisites:** Students can open a Fathom document, read a summary table, work with sliders, and create a dot plot.

**Fathom Skills:** Students will take samples from a collection and create and collect measures.

**Notes:** As you talk with students, help them understand two powerful tools of Fathom—the ability to collect samples and to collect measures on samples. Collecting samples involves cases selected at random from the original collection and assembled into a new collection. This sample collection will be a subset of the original. The original collection represents the entire population of squirrels; some squirrels were tagged in the first capture (the number depends on the value in the *First\_Capture* slider). When Fathom collects a sample of Second Capture of Squirrels, it is simulating the second capture, which will have some tagged squirrels and many untagged squirrels. To estimate the population of squirrels, assume that the ratio of tagged to total sample in the second capture is the same as the ratio of total tagged (squirrels in the first capture) to the population.

A measure is a calculation based on the collection. For example, you could create a measure that counted the tagged squirrels in the collection. Any summary statistic, or any formula based on the summary statistic, can be made into a measure. After students have solved the proportion for the population, they are asked to turn this formula into a measure. When Fathom collects measures, it re-collects the sample, calculates the formula created for each measure, and then builds a new collection from the results of the calculation. The attributes in this new collection are

the measures, and the cases each represent a measure taken from a different random sample.

**For a Presentation:** Using this activity as a presentation gives you a chance to make sure students understand what is happening when samples are taken and when measures of the samples are calculated. The animation helps students see that the sample comes from the Squirrels collection into the Second Capture of Squirrels collection and the measures come from that collection to the Measures from Second Capture of Squirrels. To collect large samples, turn off the animation. Take several samples for Q3 and Q4 and ask students to compare the results. Before taking samples or collecting measures, ask students what they think will happen. Building a graph like the one that answers Explore More gives several opportunities to ask the what-do-you-think question, which can keep students engaged.

**Q1** In the first round, 285 ( $270 + 15$ ) squirrels were captured and tagged. In the second round, 168 ( $153 + 15$ ) were captured, and 15 of those had been tagged in the first round.

**Q2**  $\frac{\text{Total caught in second round}}{\text{Caught both first and second}} = \frac{\text{Population}}{\text{Total caught in first round}}$ ,  
 $\frac{168}{15} = \frac{\text{Population}}{285}$ ; the estimated population is 3192 squirrels.

## INVESTIGATE

**Q3** Answers will vary. Sample answer: There were 305 squirrels caught—285 of them were caught for the first time, and 20 of them had already been caught and tagged in the first round of captures.

**Q4** Answers in the form  $\frac{\text{Second_Capture}}{\text{Number tagged in Second Capture}} = \frac{\text{Population}}{\text{First_Capture}}$  will vary. Sample answer:  $\frac{305}{20} = \frac{\text{Population}}{298}$ ,  
 $\text{Population} = 4544$  squirrels.

**Q5** There will be a variety of estimates.

**Q6** These results will likely be more varied than those for Q4. Sample answer:  $\frac{103}{1} = \frac{\text{Population}}{100}$ ,  $\text{Population} = 10,300$  squirrels.

**Q7** Each dot represents one estimate for the population's size based on one simulation of capture-recapture. The graph shows the spread in the estimates.

continued

9. Measures can be collected faster with animation off, but watching the animation gives students a sense of where the numbers are coming from.

**Q8** Most students will find that for large catches, there is less spread in the estimates of the population. Large samples provide better estimates. Some student results may not show this.

**Q9** In general, the least spread occurs when the two samples are about the same size. Some student results may not show this.

This graph was made by creating the collection Measures from Measures from Second Capture of Squirrels. Three measures were collected—*Spread*, with the formula  $\text{iqr}(\text{Population})$ ; *Second\_Capture*, with the formula  $\text{Second\_Capture}$ ; and *Total*, with the formula  $\text{First\_Capture} + \text{Second\_Capture}$ . To collect the measures, clear any data from Measures from Measures from Second Capture. Set the sliders to 100 and 100 and collect five measures. Reset the sliders for 200 and 200 and add five more measures. Repeat for 300 and 300, 400 and 400, and 500 and 500, giving 25 cases.

**EXPLORE MORE**

As the sample size increases, the IQR decreases (accuracy increases). The decrease is not linear. The improved accuracy going from 600 to 800 is much less than the improvement from 400 to 600.

