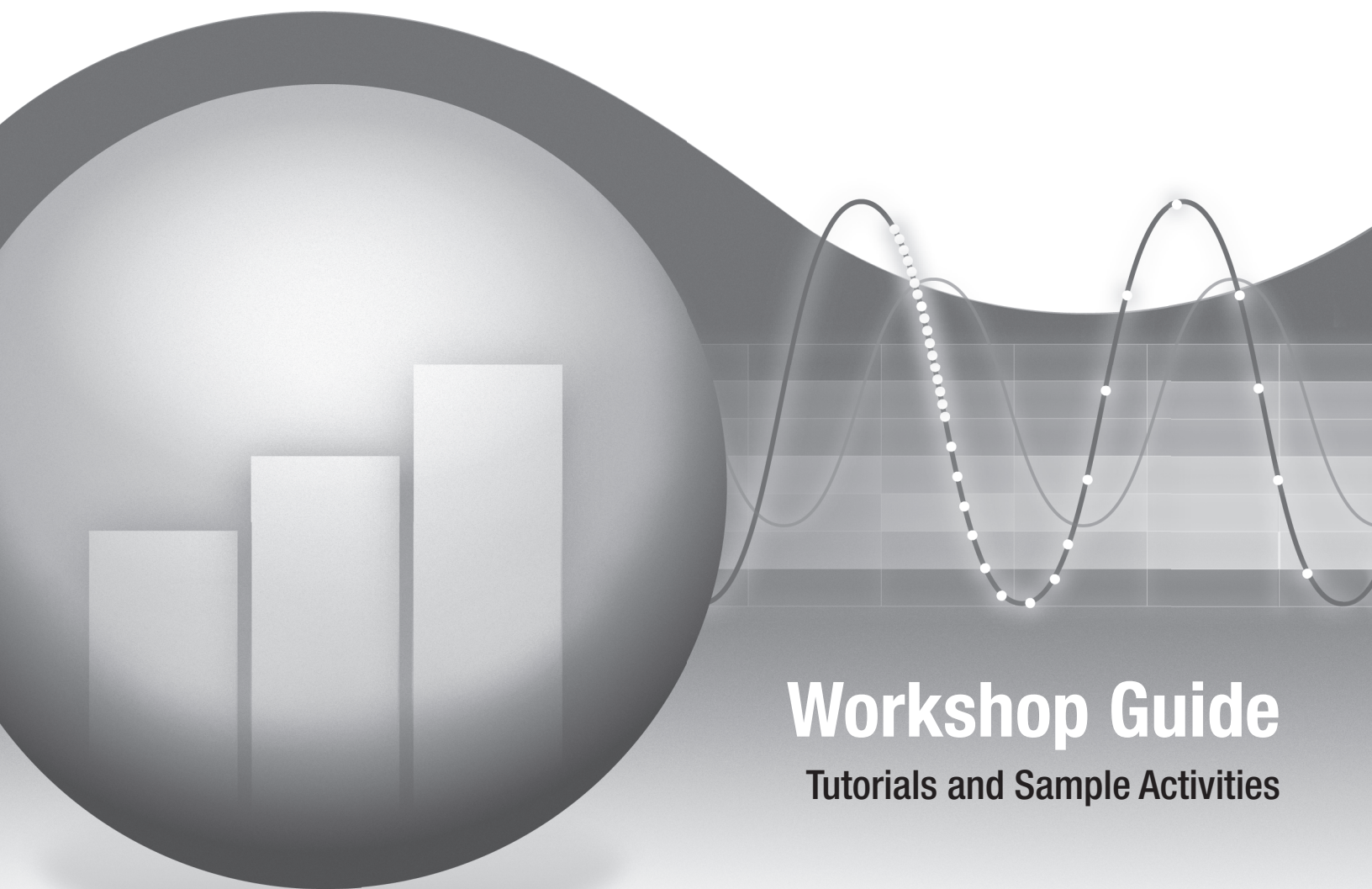


VERSION

2

# Fathom™

Dynamic Data™ Software



## Workshop Guide

Tutorials and Sample Activities

You have permission to make copies of this document for your classroom use only. You may not distribute, copy or otherwise reproduce any part of this document or the lessons contained herein for sale or any other commercial use without permission from the author(s).

**Authors:** William Finzer, Tim Erickson

**Lead Programmer for Fathom:** Kirk Swenson

**Fathom Development Team:** William Finzer, Kirk Swenson, Nick Jackiw, Matthew Litwin, Marina Raskin, Jonas Robledo, Leonid Vainerovsky

**Project Editor:** Heather Dever

**Editorial Consultant:** Jill Binker

**Editorial Assistant:** Aneesa Davenport

**Production Director:** McKinley Williams

**Production Editor:** Kristin Ferraioli

**Copyeditor:** Tara Joffe

**Production Coordinator and Composition:** Ann Rothenbuhler

**Cover Designers:** Keith Nguyen, Marilyn Perry

**Prepress and Printer:** STI Certified Products, Inc.

**Executive Editor:** Casey FitzSimons

**Publisher:** Steven Rasmussen

Portions of this material are based upon work supported by the National Science Foundation under awards numbered III-9400091 and DMI-0131833. Any opinions, findings, and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Fathom © 2014 by William Finzer. All rights reserved.

Fathom's Formula Editor ©1999 by Pacific Tech ([www.PacificT.com](http://www.PacificT.com)). All rights reserved.

™Fathom Dynamic Data and the Fathom logo are trademarks of KCP Technologies. ®Key Curriculum Press is a registered trademark of Key Curriculum Press. All other registered trademarks and trademarks in this book are the property of their respective holders.

The evaluation edition of Fathom and this *Workshop Guide* are for evaluation purposes only. William Finzer grants users the right to use and copy these for the purposes of evaluating the materials or for use in teacher workshops. Copied materials must include this notice and William Finzer copyright. The evaluation software and guide cannot be used for classroom or other extended use. The full Fathom package is available from the Concord Consortium, 25 Love Lane, Concord, MA, at [fathom.concord.org](http://fathom.concord.org).

Printed in China

10 9 8 7 6 5 4 3 2 1 09 08 07 06 05

ISBN 1-55953-683-7

# Contents

<b>Introduction</b> .....	1
To the Workshop Participant .....	1
To the Presenter .....	1
About the Instructor’s Evaluation Edition .....	2
<b>The Structure of Fathom</b> .....	3
The Fathom Document Window .....	3
Cases and Attributes .....	4
Units .....	4
Category Sets .....	5
The Inspector .....	5
Case Tables .....	5
Graphs .....	6
Ubiquitous Formulas .....	6
Using the Formula Editor .....	7
Measures .....	7
Derived Collections .....	7
Sliders .....	8
<b>Activity 1: Printing Paragraphs Exploration</b> .....	9
The Question .....	9
Conjectures .....	9
Measurement .....	9
Comparison .....	9
The Seven Paragraphs .....	10
<b>Presenter Notes for Printing Paragraphs</b> .....	11
<b>Tutorial 1: Working with U.S. Census Microdata</b> .....	12
Accessing Census Microdata .....	12
Choosing Cases and Attributes .....	13
Choosing a Question to Explore .....	13
Making a Case Table .....	14
Recoding an Attribute .....	15
Looking for Trends with a Ribbon Chart .....	16
Making a Summary Table .....	17
Going Further .....	18

<b>Activity 2: Area and Perimeter</b> .....	20
Experiment .....	20
Make a Conjecture .....	21
Investigate .....	22
Explore More .....	23
<b>Tutorial 2: Simulation—Polling Voters</b> .....	24
Defining the Problem .....	24
Making a Simulated Sample of Voters .....	24
Simulating the Votes .....	25
Defining and Collecting a Measure .....	26
Dependence of Spread on Sample Size .....	27
Going Further .....	28
<b>Activity 3: Rolling Dice</b> .....	29
Make a Conjecture .....	29
Experiment .....	29
Explore More .....	30
<b>Tutorial 3: Testing a Hypothesis—Plant Growth</b> .....	31
Looking at Darwin’s Data .....	31
Formulating a Hypothesis .....	31
Deciding on a Test Statistic .....	32
Checking Assumptions .....	33
Testing the Hypothesis .....	34
Looking at the <i>t</i> -Distribution .....	34
Going Further .....	35
<b>Using Fathom as a Demonstration Tool</b> .....	36
Discussion 1—Dice Sums Distribution .....	36
Discussion 2—Guess and Check .....	38
<b>Fathom Demo Documents</b> .....	41
<b>Sample Activities</b> .....	42
Mammals’ Brains .....	43
Activity Notes for Mammals’ Brains .....	47
Sequence of Squares .....	53
Moore’s Law .....	54
Jupiter’s Moons .....	57
Creating Error .....	61

# Introduction

As the title suggests, this guide is meant to be used in a workshop setting, where participants can work together and a workshop leader can give demonstrations, lead discussions, and answer questions. To learn to use Fathom Dynamic Data™ Software outside a workshop setting, try the *Learning Guide* that comes with Fathom or the *Walkthrough Guide* that comes with the Instructor's Evaluation Edition.

## To the Workshop Participant

Welcome to Fathom and the world of Dynamic Data. The tutorials in this guide, which begin on page 12, will introduce you to the most important features of Fathom.

As you work through the tutorials, you will encounter watch icons, such as the one at right. These indicate a good place to pause. Your workshop leader is likely to ask that when you reach these watch icons, you stop working and reflect on what you have learned so far. When most people in your workshop have reached the watch icon, your workshop leader will probably have everyone pause for a group discussion or for the workshop leader to do a brief demo.



At the end of the workshop, take this booklet and the Fathom Instructor's Evaluation CD with you to explore and evaluate Fathom in more detail. But remember that the Instructor's Evaluation CD and the *Workshop Guide* are for evaluation purposes only and are not for classroom or other extended use. As you become comfortable using Fathom, we encourage you to help others by leading workshops yourself.

If you are waiting for the workshop to begin, please read *The Structure of Fathom*, beginning on page 3.

## To the Presenter

We have found that a successful and simple workshop format for Fathom involves little more than turning participants loose to work in pairs with the activities and tutorials in this guide. (If you prefer that participants work individually, have them identify partners they can consult with when they have problems.) It is also extremely useful to have a computer that can project to a screen or large monitor so that everyone can see what you do.

The watch icons interspersed throughout the tutorials indicate reasonable pausing places where you can ask everyone to stop work for a moment so you can answer questions about the preceding section or do a brief demonstration. It's especially useful to have one or more participants come in front of the entire group to show something they have found.



In the later tutorials, as participants have become more independent, you may want to skip watch icons or eliminate their use altogether.

Key Curriculum Press gives copy permission to any presenter who wants to give this Workshop Guide to workshop participants. Key Curriculum Press also gives permission to install multiple copies of the Fathom Instructor's Evaluation Edition or a single network server copy for the purposes of giving a workshop.

You may want to complete the following checklist before the workshop.

- **Is the most recent version of Fathom installed?** This guide assumes that you have Fathom 2, but it's helpful to make sure you have the most recent maintenance update. Look on the Fathom Resource Center to make sure. (Choose **Fathom Resource Center** from the **Help** menu.)
- **Does Fathom Help work properly?** Because Fathom Help opens in your Web browser, this is worth checking. Does Help launch from the **Help** menu? Do the movies play, and is there sound?
- **Are the computers properly connected to the Internet?** In particular, does the census microdata interface work? If not, then you'll need to work around this in Tutorial 1 (page 12). You can have participants use one of the Sample Documents in the folder **Social Science | United States | US Census Files** or download data yourself before the workshop. You can also use the file **Workshop Census.ftm** posted with the *Workshop Guide* on the Fathom Resource Center.

## About the Instructor's Evaluation Edition

The Fathom Instructor's Evaluation Edition has all the features of the full version, with a few important exceptions. You can open documents, import or enter data, do investigations, and use all the features of the full version, but you cannot save any documents you create, export data you generate, copy pictures of Fathom objects to paste into another application, or print your work.

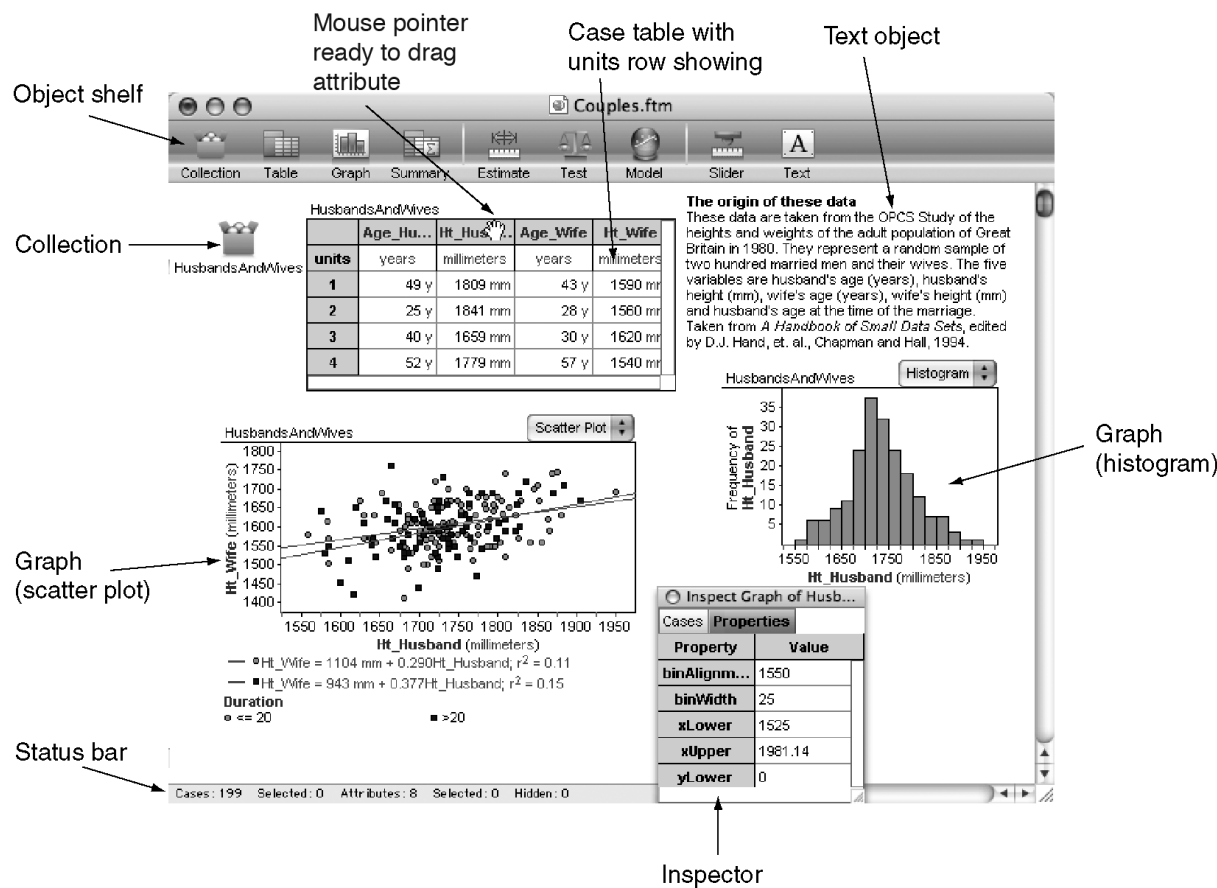
The full version allows you to do all these things and also comes with additional resources:

- The *Learning Guide* contains a much more complete introduction to Fathom's features.
- *Teaching Mathematics with Fathom* contains 25 classroom-ready activities in blackline-master form.
- *Data Are Everywhere: Project Ideas for Fathom* contains more than 50 projects ready to hand to students.

# The Structure of Fathom

Some people want to know what the designers think about the structure of a program. If that's you, this section is what you want. Others like to figure out a program by using it. That's a good way, too. But even if you seldom read manuals, you may want to read this section for background.

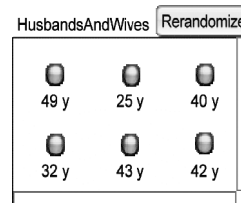
## The Fathom Document Window



A Fathom document lives in a window and contains objects. There are several different kinds of objects, led by the *collection* (which holds the data). The collection looks like a rectangle with gold balls in it or, if you make it small, a box of gold balls.

Graphs, tables, and statistical tests are objects, too. But they do not contain data. They simply provide ways of looking at or altering data. (Deleting a graph or table leaves your data intact; deleting the collection *deletes your data*.) Sliders are objects that control variable parameters, such as the coefficient of a function or a probability in a simulation.

Text resides in objects, too. Text objects allow you to add explanatory notes.



A collection showing its cases



Dice

A collection made small (iconified)

When you first use Fathom, collections are boring and may even appear unnecessary. Why can't the data just live in a table, like in a spreadsheet? The answer is deep. One way to look at it is that Fathom lets you distinguish between the way things *are* and the way they *look*. A table is just one representation of the data, and the designers didn't want to give it the privileged position of being the data's "native" appearance. Another reason for the existence of collections is that it makes some more advanced data manipulation, such as sampling, easier to think about. You sample from among the gold balls in the collection—and the sampled cases appear as balls in another collection.

## Cases and Attributes

Each gold ball is a case. A case has attributes, and attributes have values. If the values are all numeric, then the attribute is a *numeric* (continuous or measurement) attribute; if the values are text, it's a *categorical* attribute. If your cases are people, height and age are numeric; sex and race are categorical.

Whether an attribute is numeric or categorical determines many things, including the kinds of graphs you can make to represent it.

## Units

Numeric attributes may or may not have units. If an attribute has a unit, such as meters, you can enter values using any distance unit, from millimeters to light years, and they will be converted to meters. You can change the units of an attribute to a different commensurate unit, and all its values will display with the new unit. But if you change the unit to something incommensurate—such as meters to years—all the values will become invalid.

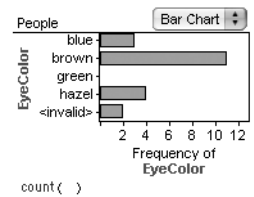
Numeric values without units are just numbers. Numeric values *with* units are *quantities* and are profoundly different from numbers. For example,  $2 + 2 = 4$ , but  $2 \text{ meters} + 2 \text{ feet} = 2.6096 \text{ meters}$ . Using units with your data can really help you understand what the data mean, but the presence of units also makes Fathom a lot pickier. For example, you can't graph two attributes on the same axis if their units are incommensurate.

Among the units that Fathom recognizes are some important constants that you can use as units, such as Earth masses ( $1 \text{ Earth mass} = 5.9742 \times 10^2 \text{ kg}$ ) or electron charges ( $1 \text{ electron charge} = 1.602176462 \times 10^{-19} \text{ coulomb}$ ).



## Category Sets

Categorical attributes can't have units, but they can have category sets assigned to them. A *category set* defines the valid values that such an attribute can have. The bar chart at right, for example, shows that the valid values of *EyeColor* are “blue”, “brown”, “green”, and “hazel”. The bar chart leaves a place for “green” even though there are no people with that eye color in the collection. The bar chart also lumps all the values that are not in the category set into a single “<invalid>” group. Using category sets can be very helpful, and sometimes absolutely necessary, but you can go a long way in Fathom without them.



## The Inspector

When you double-click a collection, its inspector appears as a separate window floating above the document. The inspector shows you the names and values of the attributes, one case at a time. You may edit the values in the inspector. You can also use the inspector to add measures (more on this later) and comments to the collection, change how the cases look, and control advanced functions, such as sampling.

Attribute	Value	Formula
face	5	randomPick(1, 2, 3, 4, 5, 6)

Some other objects, particularly graphs and sliders, have inspectors as well. Only one inspector window appears on your screen at a time.

If you're typing a lot of data into Fathom, the inspector is *not* the most convenient place for it; you'll want to use a case table instead.

## Case Tables

The case table presents the familiar tabular view of data. Each case appears as a row; each attribute as a column. In a case table, you can enter and edit data quickly and add new attributes. You can also rename attributes, move columns by dragging them, resize columns by dragging or double-clicking their borders, sort the data, and hide attributes to make the table less crowded. Although you can change data through a case table, it is only one view of the data. If you delete the case table, the data are still present—in the collection.

Elements	Name	Symbol	Number	State
1	Actinium	Ac	89	solid
2	Aluminum	Al	13	solid
3	Americium	Am	95	unstable
4	Antimony	Sb	51	solid
5	Argon	Ar	18	gas
6	Arsenic	As	33	solid
7	Astatine	At	85	solid
8	Barium	Ba	56	solid
9	Berkelium	Bk	97	unstable

## Graphs

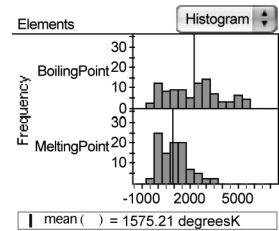
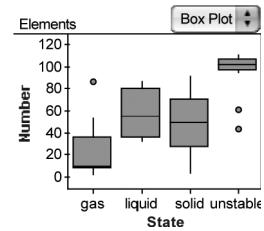
Fathom also lets you view data graphically. The graph object supports various kinds of charts and plots. You tell Fathom what to graph by dragging attribute names to an axis of the graph. You specify the graph type by choosing from the menu in the corner of the graph object.

The configuration of attributes determines what kinds of graphs are available. For example, if you put *age* on an axis and leave the other axis blank, you can get a box plot or a histogram, because *age* is a numeric attribute. But if you put *sex* on that same axis, you can't get a histogram, because *sex* is categorical. You get a bar chart instead.

You can also add things such as functions to graphs. Functions can be anything you can express with a formula (see "Ubiquitous Formulas" below), as well as more familiar lines such as least-squares linear regression lines or median-median lines.

It is frequently possible to place more than one attribute on an axis. In general, an attribute dropped in the middle of a graph produces a legend and changes the color or shape of the plotted points.

In graphs of numeric attributes, you can change the data by dragging (unless the data are locked). For example, you can drag points in a scatter plot to see how the dragging affects a least-squares line.

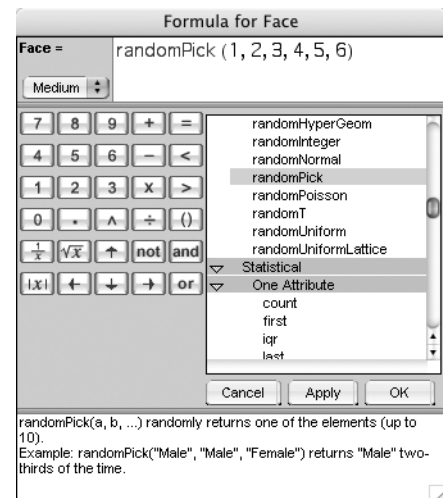


## Ubiquitous Formulas

Formulas are everywhere in Fathom. You can use formulas as filters to tell which cases you want to see. You can also use them to compute values (including random values) for attributes. You can use formulas to devise statistical measures and to plot formulas as functions in graphs.

You write formulas in the formula editor, which appears when you double-click the place where a formula belongs or when you choose certain menu items.

The formula editor supports a wide array of functions, from trig functions to statistical distributions to conditionals such as *if()* and *switch()*. All of the functions are listed in the attribute and function list. Double-click an item in the list to enter it into the formula pane. You can also enter it by typing. The formula editor displays the names of functions, attributes, measures, and



units in special colors. If you type an attribute name or function and it remains black, Fathom does not recognize it (you probably have a typo).

## Using the Formula Editor

The precise nature of mathematical formulas makes editing them different from editing ordinary text. First, certain characters—parentheses, quotes, and absolute-value signs—must appear in pairs. Fathom always makes the second of the pair when you enter the first. Second, grouping and order of operations requires special editing capabilities. New users of Fathom can use their intuition most of the time, but there is one new principle they must learn: The effect of some keys depends on what, if anything, is selected when you press them. For example, if  $a + b$ , without parentheses, is selected, and you press  $($ , you get  $(a + b)$ , with parentheses. If you then press  $*$  (for multiplication), you get  $(a + b) *$ .

Before	After
$a + b $	$a + \frac{b}{4}$
$(a + b $	$\frac{a + b}{4}$
$(a + b) $	$\frac{(a + b)}{4}$

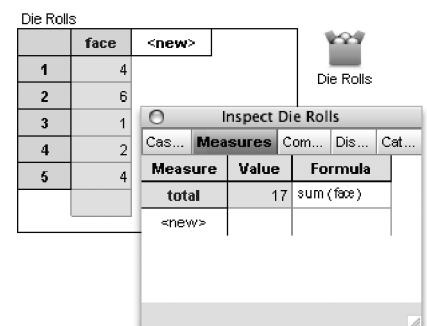
In each of the examples at right, the user types  $/4$ .

Finally, you need to click **OK** to close the formula editor. But sometimes you want to see the result of your formula—to see if it's right—before closing the editor. In that case, click **Apply**; this applies your changes but leaves the editor open. You must close the formula editor before you can go on to do other things.

In most instances, the formula editor is *not* case sensitive; that is, it doesn't care whether you type capital or lowercase letters. But units are an exception.

## Measures

An *attribute* has a separate value for each case in the collection. A collection can also have *measures*, and each measure has one value for the entire collection. For example, a collection of five dice might have an attribute called *face* with five separate values. And the collection could have a measure, named *total*, that applies to the entire collection. Typically, the measure has a formula, such as  $sum(face)$ . You can think of measures as statistics that summarize your collection, though they can serve other functions as well.



You create measures using the inspector. Double-click a collection to show the inspector, and go to the **Measures** panel.

## Derived Collections

Some Fathom collections fill with data automatically, according to rules you specify. These are called *derived collections*, and there are several kinds.

The two most important are *samples* and *measures collections*. A sample is just what it sounds like: a collection that is a sample of some other collection, called its *source*. Select the source and choose **Sample Cases** from the **Collection** menu. The sample collection appears. Use the inspector to control how many cases it samples and whether it samples with or without replacement.

The measures collection is the key to simulation and analysis. It converts measures into case attributes so you can record statistics (measures) about your collections. (It's a bit tricky and is best understood after you've tried it.) Suppose you have a collection with five (randomly generated) dice and a measure, or total, that contains their sum. If you collect measures from that source collection, the measures collection can record the total as a case. If you tell the measures collection to collect 100 measures, it will instruct the dice to reroll 100 times, it will compute each total, and it will record those 100 values.

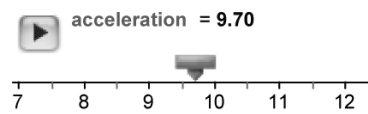
	face
=	randomPick ( 1, 2,
1	2
2	3
3	2
4	1
5	6

	total
1	16
2	20
3	17
4	15
5	20
6	21
7	13

In addition to sample and measures collections, several other derived collections come in handy. A *scrambled collection* reorders the values of one of the source collection's attributes, allowing simulations of situations in which the values of two attributes are guaranteed to be independent. A *cells collection* is generated from a summary table and has one case for each cell in the table and one attribute for each formula. A *results collection* is like a measures collection except that it treats a statistical object as the source of its measures.

## Sliders

A *slider* is a named number whose value can be changed through editing, dragging, or animating. Sliders are commonly used as parameters to models—for example, as the coefficient of a term in an equation being fit to a set of data or as the rate at which customers arrive in a queuing simulation.




Sliders have their own inspector, which gives you control over quite a few properties. Especially powerful is the ability to compute the slider's value with a formula. Unlike other formulas in Fathom, slider formulas can refer to the slider itself so that each time the formula recomputes (for example, when the slider animates), the value changes based on the previous value. By dropping the name of a collection on a slider, you can compute things based on the collection's attributes and measures.

Property	Value	Formula
acceleration	9.7	
Max_updates_per_second		
Lower	-1	
Upper	13	
Restrict_to_multiples_of		
Reverse_scale	false	

# Activity 1: Printing Paragraphs Exploration

---

In this workshop, you'll jump directly into a mathematics activity in which you use Fathom as a tool. You'll see that *using* the software need not consume time for *teaching* the software.

 **Presenter:** To introduce participants to what they need to know for this activity, show them these two 1-minute movies:

- *Entering Data and Making a Graph*
- *Plotting Functions in a Scatter Plot*

## THE QUESTION

You will need

- centimeter ruler



The next page has some paragraphs on it. They're all exactly the same text, but they are set in columns of different widths. How does the height of the column depend on the width?

## CONJECTURES

Look at the paragraphs (but don't measure). Think about how the *height* and *width* of the paragraphs are related. What would a graph of them look like? How would you write that relationship algebraically? If you don't know exactly, what can you say *about* the relationship? Write your conjectures here. *Extension:* These paragraphs are set "ragged right"; that is, the right margin isn't straight. How can you take that into account in your algebra?

## MEASUREMENT

Measure the height and width of the paragraphs to the nearest tenth of a centimeter. You and your partner will have to decide how to measure height and width. Be as precise and consistent as you can. Describe what you decide.

## COMPARISON

Enter your measurements into Fathom and analyze them. What did you discover? Do your measurements support your conjectures? Write your observations here. Be as specific as you can.

## Activity 1: Printing Paragraphs Exploration

(continued)

---

### THE SEVEN PARAGRAPHS

Some rational functions can create very different kinds of graphs from those you have studied previously. The graphs of these functions are often in two or more parts. This is because the denominator, a polynomial function, may be equal to zero at some point, so the function will be undefined at that point. Sometimes it's difficult to see the different parts of the graph because they may be separated by only one missing point, called a hole. At other times, you will see two parts that look very similar—one part may look like a reflection or rotation of the other part. Or you might get multiple parts that look totally different from each other.

Some rational functions can create very different kinds of graphs from those you have studied previously. The graphs of these functions are often in two or more parts. This is because the denominator, a polynomial function, may be equal to zero at some point, so the function will be undefined at that point. Sometimes it's difficult to see the different parts of the graph because they may be separated by only one missing point, called a hole. At other times, you will see two parts that look very similar—one part may look like a reflection or rotation of the other part. Or you might get multiple parts that look totally different from each other.

Some rational functions can create very different kinds of graphs from those you have studied previously. The graphs of these functions are often in two or more parts. This is because the denominator, a polynomial function, may be equal to zero at some point, so the function will be undefined at that point. Sometimes it's difficult to see the different parts of the graph because they may be separated by only one missing point, called a hole. At other times, you will see two parts that look very similar—one part may look like a reflection or rotation of the other part. Or you might get multiple parts that look totally different from each other.

Some rational functions can create very different kinds of graphs from those you have studied previously. The graphs of these functions are often in two or more parts. This is because the denominator, a polynomial function, may be equal to zero at some point, so the function will be undefined at that point. Sometimes it's difficult to see the different parts of the graph because they may be separated by only one missing point, called a hole. At other times, you will see two parts that look very similar—one part may look like a reflection or rotation of the other part. Or you might get multiple parts that look totally different from each other.

Some rational functions can create very different kinds of graphs from those you have studied previously. The graphs of these functions are often in two or more parts. This is because the denominator, a polynomial function, may be equal to zero at some point, so the function will be undefined at that point. Sometimes it's difficult to see the different parts of the graph because they may be separated by only one missing point, called a hole. At other times, you will see two parts that look very similar—one part may look like a reflection or rotation of the other part. Or you might get multiple parts that look totally different from each other.

Some rational functions can create very different kinds of graphs from those you have studied previously. The graphs of these functions are often in two or more parts. This is because the denominator, a polynomial function, may be equal to zero at some point, so the function will be undefined at that point. Sometimes it's difficult to see the different parts of the graph because they may be separated by only one missing point, called a hole. At other times, you will see two parts that look very similar—one part may look like a reflection or rotation of the other part. Or you might get multiple parts that look totally different from each other.

Some rational functions can create very different kinds of graphs from those you have studied previously. The graphs of these functions are often in two or more parts. This is because the denominator, a polynomial function, may be equal to zero at some point, so the function will be undefined at that point. Sometimes it's difficult to see the different parts of the graph because they may be separated by only one missing point, called a hole. At other times, you will see two parts that look very similar—one part may look like a reflection or rotation of the other part. Or you might get multiple parts that look totally different from each other.

from *Discovering Advanced Algebra* by Murdock, Kamischke, and Kamischke

## Presenter Notes for Printing Paragraphs

How you show the movies will depend on the group size and your setup. Remember that 30 people all playing the movies at slightly different times will make quite a lot of noise.

It is *essential* that participants work in pairs so they can help each other past the rough spots and so they have someone to brainstorm with. As you turn participants loose on the investigation, urge them to *write down* their responses in the conjectures and comparison sections so they'll have them for discussion. Ask them to note questions they ask each other and decisions they make.

Inevitably some participants will have completed development of a model long before others have. Here are some prompts for them:

- Using Fathom Help, learn about residual plots, and take a look at residuals for your model.
- Figure out a way to refine your model so that it produces a better fit. (Usually the narrowest paragraphs have the greatest residuals.)

Once everyone has a scatter plot but before everyone has a fitted curve, stop and discuss:

- What conjectures did you make?
- What measurement decisions did you make?
- In what ways did your results conform to your conjectures? How did the results surprise you?

Then demonstrate a movable line model for the data. Demonstrate showing squares of residuals. Demonstrate residual plots. (Or have a participant do the demonstration, but with the straight line not a curve!)

Next, turn the participants loose again, with the goal of finding the “best” model for these data. They should be able to justify their choice of model.

When every group has found a reasonable model, stop and have some participants demonstrate what they did. Here are some discussion prompts:

- In what sense have you developed a model for the printing paragraphs situation?
- What stages were there in the development of your model?
- How much “overhead” was there in learning Fathom in this activity?
- Based on this one experience, what do you think of the statement: “Students can learn how to use Fathom as a part of learning mathematics.”?

# Tutorial 1: Working with U.S. Census Microdata

Fathom provides an interface to data about individuals from 1850 to 2000 who filled out the long forms of the U.S. census. In this tutorial, you'll learn how to access that data and how to make graphs and tables that aid in data exploration.

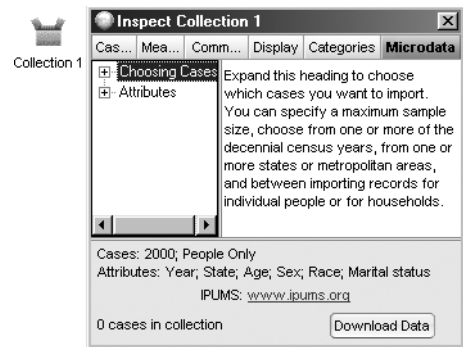
**Presenter:** *If workshop computers are not connected to the Internet and you still wish to do this tutorial, discuss with participants what file or files to use and which sections become irrelevant. Otherwise, allow 10 minutes for the first section.*

## Accessing Census Microdata

1. From the **File** menu choose **Import | U.S. Census Data**.

Two things appear. One is an empty box, a collection that you will fill with cases. The other is an inspector for the collection.

In the inspector you'll choose the people whose census information you want to look at. The bottom pane of the inspector describes the data you've asked for. By default, you'll get people from the year 2000—their state, age, sex, race, and marital status.

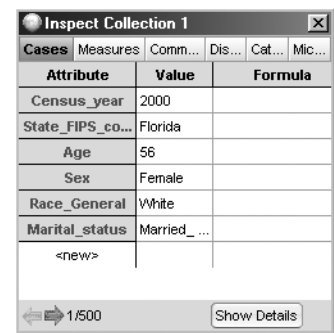


2. Click **Download Data**.

Fathom sends a request to the Minnesota Population Center and, when the data are ready, downloads them. You'll know Fathom has the data when the collection fills with gold balls and the number of cases is listed as 500 (see the bottom-left corner of the inspector).

3. Click the inspector's **Cases** tab.

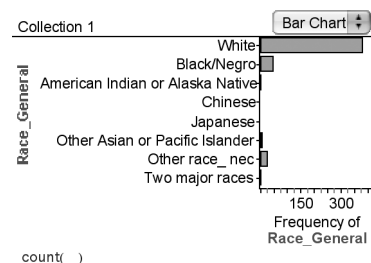
You will see information about one of the 500 people in the collection. You can move from one case to the next using the arrow buttons at the bottom-left corner of the inspector.



4. Drag a graph from the object shelf, and place it in the document.
5. Drag one of the attribute names from the inspector to either the vertical or horizontal axis of the graph.

For graphs of categorical attributes with long category names, you'll find the graphs are more readable with the attribute on the vertical axis.

6. Take a look at several other attributes. You can reuse the first graph or make new ones.





## Choosing Cases and Attributes

Fathom’s census microdata interface gives you access to a vast amount of data. It is helpful to approach these data with a question in mind. But to ask reasonable questions, you need to know more about what cases you can get and what attributes are available.

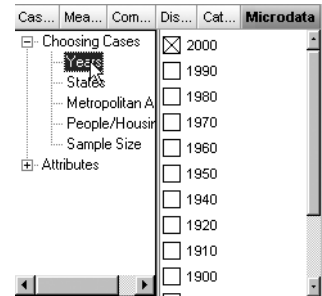
7. In the inspector, click the **Microdata** tab—the rightmost one.
8. Expand the Choosing Cases list.

The subheads under Choosing Cases show you the various limitations you can place on the cases you bring into the collection.



9. Select Years in the left pane.

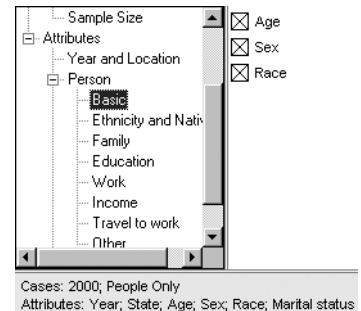
You can import microdata for each year in the right pane. Selecting more than one year from this list allows you to compare people from these years.



10. Select each of the other headings in turn, and look at the choices available.

11. Open the Attributes heading and, under it, the Person heading.

The subheads under Person contain attributes that deal with aspects of an individual person. For example, as shown at right, *Age*, *Sex*, and *Race* are Basic attributes.



12. Select each of the subheads in turn, and note the attributes available.

Notice that if you allow the cursor to linger over an attribute, information about that attribute appears.



## Choosing a Question to Explore

Exploring census microdata is most interesting when you have a question you wish to answer or a conjecture you would like to confirm. We’ll work from a particular example, but feel free to investigate a question of your own instead.

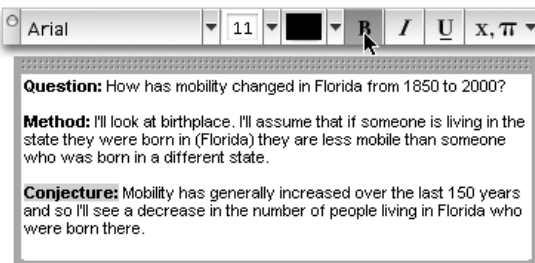
Suppose we’re interested in how “mobile” people are, that is, how much they move around. We want to know how that “mobility” has changed over time. There is no attribute that will answer this question directly. (This is typical in working with

census microdata.) We have to use attributes that give us *indirect* information about the question. We'll use *Birth Place* as an “indicator” of mobility, assuming that the greater the proportion of people who were not born in the state they currently live in, the greater the mobility.

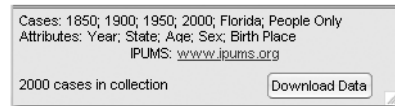
As we begin an investigation, it is extremely helpful to document what we're trying to do and how we intend to do it.

13. Drag a text box from the shelf, and drop it into the document.
14. In the text box, type a description of the investigation.

Below is an example of what you might write. Some formatting has been done using the *text palette*. (Choose **Show Text Palette** from the **Edit** menu.)



15. Use the inspector's **Microdata** panel to request cases from a single state over a range of years. Include *Birth Place* as an attribute.



(Here the sample size has been increased to 2000 cases. This is not necessary.)

16. Download the data.

The new data *replace* the old.

**Participant:** Stop working. **Presenter:** Answer questions. Follow the link in the **Microdata** panel to [www.ipums.org](http://www.ipums.org), and discuss it as documentation for the census microdata. Allow 15 minutes for the next section.

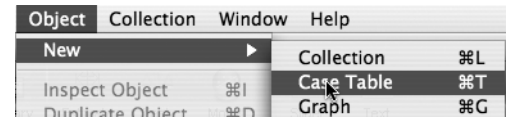
## Making a Case Table

So far we've been using graphs and inspectors to look at the data. To verify that it's possible—and to help you avoid a common problem—let's make a table.

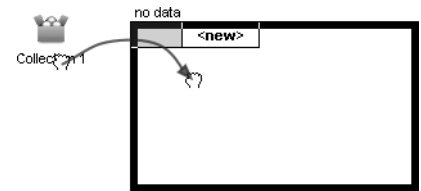
17. Click once on the collection to select it. (When you have it selected, it has a border around it.)

18. Choose **New | Case Table** from the **Object** menu.

The table should have the same name as the collection and should show the attribute names in the top row.



If the table is empty, you probably did not have the collection selected. To connect the collection to the table, drag the *name* of the collection onto the table.



You can resize a case table just as you would resize any other object—by dragging its edges or corners. Many people resize their case table to be very large in an attempt to see all the cases. *This is usually not a good idea.* Even with a large table, you can rarely see all the data; even if you could, you would have trouble making sense of them. Worse, a large case table leaves little room for truly useful objects, especially graphs.

Beginning Fathom users often lose their graphs underneath a large case table. Expert users, if they have a case table in their document at all, make it wide but short, so it can serve as a source of attribute names.

Collection 1

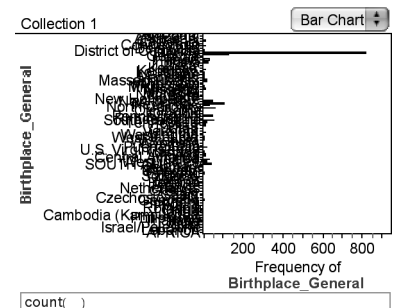
	Census...	State_Fl...	Age	Sex	Birthpla...
1	1850	Florida	26	Male	Georgia
2	1850	Florida	3	Male	Florida

## Recoding an Attribute

It's time to look at *Birth Place*. First, notice that the attribute actually got named *Birthplace\_General* rather than *Birth Place*. Fathom defers to the names used by the census microdata server at the Minnesota Population Center, the actual source of the data.

19. Make a bar chart with *Birthplace\_General* on the vertical axis.

Your graph probably looks similar to this one—it is *unreadable* because there are so many values. For our investigation, we only need to know whether a person was born “in state” or “out of state.”



20. In the inspector's **Cases** panel, click on **<new>** and type `where_born` to define a new attribute.

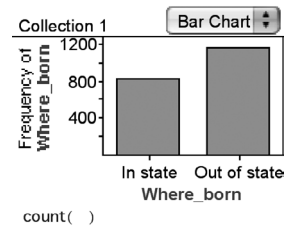
21. Double-click the Formula cell for the new attribute, and enter the formula shown at right. Use the following formula editor hints:

```
if (Birthplace_General = State_FIPS_code) {"In state"} {"Out of state"}
```

- Enter the structured if-statement by typing `if (`. When you type the left parenthesis, the rest of the structure appears.
- Enter long attribute names by expanding the Attributes list and double-clicking the name.
- Move from place to place in the formula expression by pressing the Tab key.
- Typing `"` always produces a pair of double quotes.

22. Click **OK** to close the formula editor and accept the formula.

23. Use the new attribute to make the bar chart shown at right.



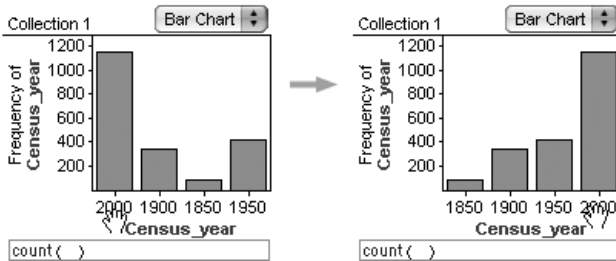
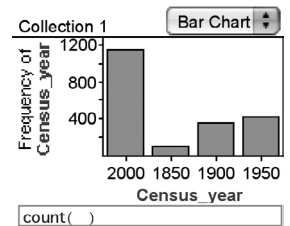
## Looking for Trends with a Ribbon Chart

Now we can get back to the investigation of how the proportion of people born in the state has changed over time.

24. Make a bar chart of *Census\_year*.

Notice that the years are not in chronological order.

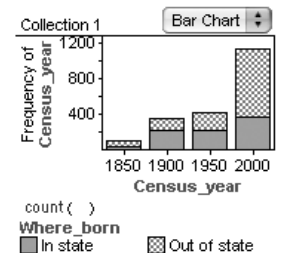
25. Change the order of the years by dragging the label (not the bar) for the year 2000 to the right.



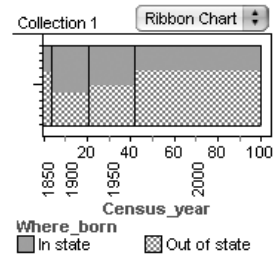
26. Drag *Where\_born* into the middle of the bar chart.

The bar chart shows the counts of people in each year born in state and out of state, but we're more interested in proportions.

27. Change the bar chart to a ribbon chart using the pop-up menu in the upper-right corner of the graph.



The ribbon chart allows you to directly compare the proportions from year to year. The ribbon chart for Florida, for example, shows a more complicated pattern than stated in the initial conjecture.



28. Summarize what you have found in your text object.

**Participant:** Stop working. **Presenter:** Answer questions. Probe a bit on reading the ribbon chart because typically some people will not have appreciated its usefulness. Demonstrate that `Census_year` must be special somehow because it is treated as categorical even though it is purely numeric. Explain category sets and how without them you can still force an attribute to behave the way you want it to. Allow 10 minutes for the next section.

## Making a Summary Table

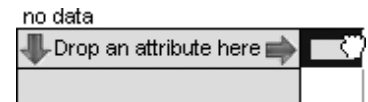
As useful as graphical displays are, sometimes we need the numbers.

29. Drag a summary table from the shelf.

The summary table is empty because Fathom doesn't know which attributes you want to work with.

We'll set up the summary table to mirror the layout of the ribbon chart.

30. Drag `Census_year` to the summary table and drop it on the column header cell, as shown at right (or on the right-pointing arrow). (A black outline appears when you are over a drop area.)



By default, for a categorical attribute, the summary table displays the number of cases in each group. The formula(s) below the table determine what it shows.

Collection 1

	Census_year				Row Summary
	1850	1900	1950	2000	
	90	344	418	1148	2000

S1 = count( )

31. Drag `Where_born` to the summary table and drop it on the row header (or the down-pointing arrow).

Collection 1

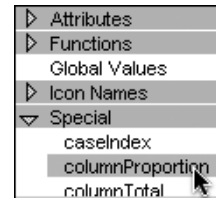
	Census_year				Row Summary	
	1850	1900	1950	2000		
Where_born	In state	31	210	215	370	826
	Out of state	59	134	203	778	1174
	Column Summary	90	344	418	1148	2000

S1 = count( )

Now the table displays counts and summaries for the rows and columns, but, again, we're more interested in proportions.

32. Double-click the formula for *S1* to show the formula editor.

In the Special list is a function called *columnProportion*, which gives the proportion of total cases in a column that are in a given cell. (It is available only when you are writing a formula for a summary table.)



33. Create the expression shown at right and click **OK**. You get a table that mirrors the ribbon chart.

`round(100columnProportion)`

Collection 1

		Census_year				Row Summary
		1850	1900	1950	2000	
Where_born	In state	34	61	51	32	41
	Out of state	66	39	49	68	59
Column Summary		100	100	100	100	100

*S1* = `round(100columnProportion)`


Here are a few useful things to know about summary tables. Try out the ones that interest you.

- Click in a table cell to select all the cases that correspond to that cell. Shift-click extends or inverts the selection.
- To add additional formulas to the table, select the table and choose **Add Formula** from the **Summary** menu. With more than one formula in the table, it is helpful to click on one formula and see all the corresponding values highlighted in the table.
- Change which attributes are displayed in the table by dragging attributes to the top or side column or to the corresponding arrows.
- Replace one attribute with another by dragging an attribute on top of the attribute you wish to replace.
- To remove an attribute from the table, click once on the attribute and choose **Remove Attribute** from the **Summary** menu.

## Going Further

- Choose another state, and compare the patterns of *Where\_born* for the two states.
- Each time you download data, you get a new random sample of the available cases (unless there are fewer cases available than you have set in the sample size). Of course, the proportions you get will be somewhat different each time you download. Explore this variation. Should you question your original conclusion?

- Make a percentile plot of income and figure out how to read it. Use the plot to determine the income corresponding to the 80th percentile.
- Use other attributes to look at mobility. Do they give similar or conflicting results?
- Think about the influence of having children in the sample. Use Fathom Help to learn how to *filter* out the children, and determine whether that changes your conclusion.

 **Participant:** *Stop working. Presenter:* *Answer questions. Demonstrate the various kinds of summary tables you can generate, depending on the attribute configuration. Demonstrate adding a filter to a graph or summary table. Allow participants who have completed one of the explorations to show the group what they found.*

The next activity is taken from *Teaching Mathematics with Fathom*, the activity book that comes with Fathom (2005, Key Curriculum Press). This activity is from Chapter 2: Geometry, but could also be used with algebra, precalculus, or statistics students. The steps are detailed so that it could be a first activity that students undertake with Fathom.

## Activity 2: Area and Perimeter

---

How are the area and perimeter of a rectangle related? You probably know the formulas by heart:

$$\text{Area} = \text{Length} \cdot \text{Width}$$

$$\text{Perimeter} = 2(\text{Length} + \text{Width})$$

But if you look at data for many different rectangles, would you notice any patterns? Are there relationships between area and perimeter? You'll use Fathom to explore these questions.

### EXPERIMENT

The first step is to experiment with many different rectangles and gather data. You could get data by actually drawing or cutting out rectangles and measuring their lengths and widths, but that would get tiring. Instead, you'll use Fathom to generate random numbers for the lengths and widths. Using technology to randomly generate data for an experiment is one method of doing a *simulation*.

1. Open a new Fathom document and create a new case table.
2. Make two new attributes, *Length* and *Width*.
3. With the case table selected, choose **New Cases** from the **Collection** menu. Type 200 in the dialog box and click **OK**. The cases appear in your table, although they have no values.

For this activity, you'll assume that the length or width of a rectangle can be between 0 and 10 units.

4. Select the attribute *Length* and choose **Edit Formula** from the **Edit** menu. Type `random(10)` in the formula editor and click **OK**. The *Length* column fills with randomly generated numbers. If you want, choose **Rerandomize** from the **Collection** menu to see a different set of random numbers.
- Q1 Look at the values of *Length*. What does the formula `random(10)` do? What would `random(100)` do? How about `random()`? (*Hint: If you aren't sure, read about the `random` function in Fathom Help.*)
5. Use the same formula to define *Width*.

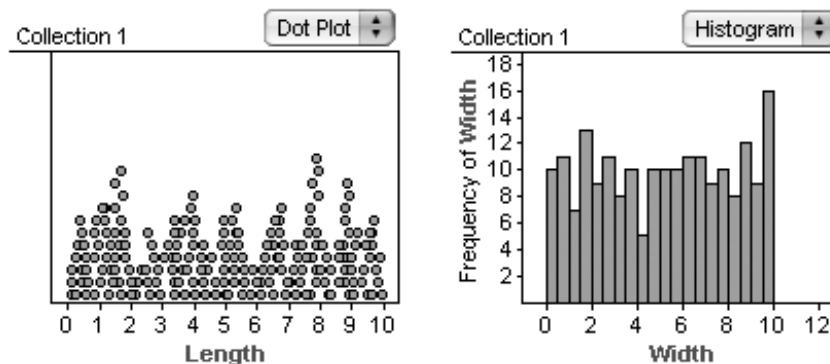


## Activity 2: Area and Perimeter

(continued)



6. Drag a new empty graph from the object shelf. Drag *Length* from the case table to the horizontal axis. You'll see a dot plot; you can change the graph to a histogram if you like. Make a dot plot or histogram of *Width*, too.



- Q2** Describe the general “shape” of the graphs of *Length* and *Width*. If you want, rerandomize the data several times to get a better feel for the general shape. Why do these graphs have this shape?
7. Make another new graph. This time make a scatter plot with *Length* on one axis and *Width* on the other axis. Rerandomize the data a few times.
- Q3** Describe any patterns in the scatter plot. Why does the scatter plot look the way it does?

Now that you have data that represent random rectangles, you need to find the perimeter and area of each. After all, you want to look for relationships between area and perimeter.

8. Make two new attributes, *Area* and *Perimeter*. Define each with a formula that uses *Length* and *Width*.
- Q4** Copy on paper all of the values for the first two cases in your case table. Use by-hand calculations to verify that your formulas for *Area* and *Perimeter* are correct.

## MAKE A CONJECTURE

So far, you’ve simulated many rectangles by randomly generating length and width, and you’ve calculated the area and perimeter. You’ve looked at graphs that help you see how length and width are related. Now, let’s see how area and perimeter are related. But first, predict what you expect to find.

## Activity 2: Area and Perimeter

(continued)

- Q5** If you made a scatter plot of *Area* versus *Perimeter*, what do you think it would look like? Make a sketch of what you would expect to see or describe it thoroughly. Explain why you think it would look that way.

### INVESTIGATE

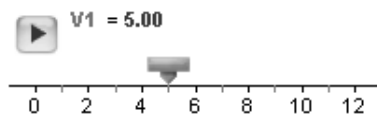
- 9.** In Fathom, actually make the scatter plot of *Area* versus *Perimeter*. Put *Perimeter* on the horizontal axis and *Area* on the vertical axis. You might want to rerandomize a few times to make sure you understand the general relationship.
- Q6** Make a sketch of the actual graph or describe it thoroughly. Did the actual graph match your prediction? Is there anything about the graph that surprises you? Explain.
- Q7** If everything went well, there should be no points in the upper-left part of the graph—a region bounded by a curve. Explain why there are no points in that region.

Now your goal is to figure out the function for that curved boundary. Because calculating area involves squaring, you might reason that the function is quadratic:  $Area = Perimeter^2$ .

- 10.** Select the graph and choose **Plot Function** from the **Graph** menu. In the formula editor, type  $Perimeter^2$  and click **OK**.
- 11.** Unfortunately, the curve is too steep. Double-click the function equation at the bottom of the graph. Change the formula to divide  $Perimeter^2$  by some number, such as  $Perimeter^2/3$ .
- Q8** What equation did you try? What happened to the curve—did it get steeper or shallower? Do you think the divisor needs to be bigger or smaller than the number you used?

You can use a slider to try lots of divisors quickly.

- 12.** Drag a new slider from the object shelf. The slider that appears looks like a number line with a pointer on it. It's probably called *V1* and has a value of 5.



## Activity 2: Area and Perimeter

(continued)

---

- 13.** Double-click the function equation at the bottom of the graph and change it to use  $VI$  as the divisor. Now you can control the value of the divisor and change the function by dragging the slider. You will probably need to change the scale of the slider. The easiest way is to drag the numbers on the number line.
- Q9** What value of  $VI$  makes the curve fit the boundary the best?
- 14.** Double-click the collection (the box of gold balls) to show the collection inspector. Then go to the scatter plot of *Area* versus *Perimeter* and click any of the points nearest to the curved boundary. When you do, that case appears in the inspector. Do this for several of the points and notice the *Length* and *Width* of each point.
- Q10** What do the cases near the boundary have in common? Why do all those rectangles have that relationship? What type of rectangles are they?
- Q11** Use your observation from Q10 and the formulas for area and perimeter to find the boundary curve's function algebraically. Show that it is the same formula you got by using the slider to fit the curve.

### EXPLORE MORE

- 1.** There is another empty zone in the lower-right part of the scatter plot of *Area* versus *Perimeter*. Why are there no points in that region? What is the equation of that boundary? Why? Use this and your answer to Q11 to completely describe the relationships between area and perimeter.
- 2.** Look at scatter plots of *Area* versus *Length* and *Perimeter* versus *Length*. Are there maximum and minimum relationships between these attributes? If so, find functions for the boundaries and explain why they exist.

## Tutorial 2: Simulation—Polling Voters

In this tutorial, we simulate a population of voters, a certain proportion of whom will vote in favor of a particular proposition. We'll investigate the question of how accurately a random sample of voters can predict the outcome of the election.

In the process, you'll learn how to use a slider to define a population parameter, how to use the *random* function, how to define a measure for a collection, and how to repeatedly collect measures from a collection.

 **Presenter:** Allow 10 minutes for the first section.

### Defining the Problem

The city of Freeport has a rent control initiative, Proposition A, on the ballot. The local newspaper is going to conduct a poll three weeks before the election to gauge public sentiment. Staff members need to know how big a sample to poll. Your job is to set up a simulation they can use to determine, for any given sample size, the accuracy they should expect.

### Making a Simulated Sample of Voters

First, we model the population, the people who will vote in the election either for or against the proposition. This model will consist of a single number—the proportion of voters who will vote yes.

1. Start with a new, empty Fathom document.
2. From the shelf, drag a new slider into the document.

By default, a slider is named *V1*. We want the slider to have a name that fits with the model we're constructing.

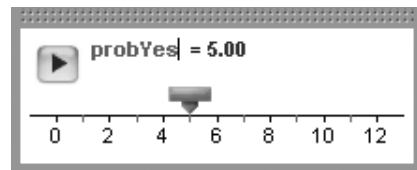
3. Double-click the slider's name, and type `probYes`.

A proportion of voters can only go between 0 and 1, so we need to adjust the slider scale.

4. Drag the numbers on the slider scale until the scale goes approximately from 0 to 1.
5. Set the value of the slider (by dragging its thumb) to around 0.5, simulating a close election.

The slider models the population. We'll use a collection to model a sample of voters.

6. Drag a collection from the shelf into the document.  
The collection will be empty and named Collection 1.
7. Double-click the name below the box. Name the collection Sample of Voters.



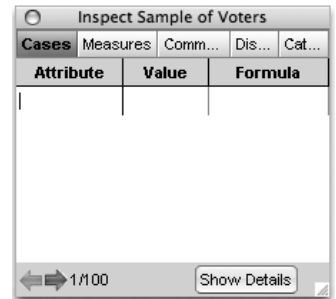
We'll start this collection with 100 cases. Later we can investigate other sample sizes.

8. With the collection selected, choose **New Cases** from the **Collection** menu.
9. Type 100 in the dialog box, and click **OK**.

## Simulating the Votes

So far, the cases in the collection have no attributes. We want one attribute, *Vote*, whose value is either “yes” or “no.”

10. Double-click the collection. The inspector appears.  
Here you will define an attribute for a vote.
11. Click <new>, type *Vote*, and press Enter.
12. Double-click the Formula cell in the right column.



The formula editor appears.

You now need to enter a formula that will choose “yes” the percentage of the time specified in your slider, *probYes*.

13. Enter the formula shown at right.

$$\text{if ( random ( ) < probYes) } \begin{cases} \text{"yes"} \\ \text{"no"} \end{cases}$$

The *random()* function, if you don't give it a minimum and maximum, generates random numbers from 0 to 1.

14. Close the formula editor by clicking **OK**.

You should see the value of the *Vote* attribute in the inspector become either “yes” or “no.” Scroll through a few more cases by clicking the right arrow in the lower-left corner of the inspector. Some of the values should be “yes” and some should be “no.”

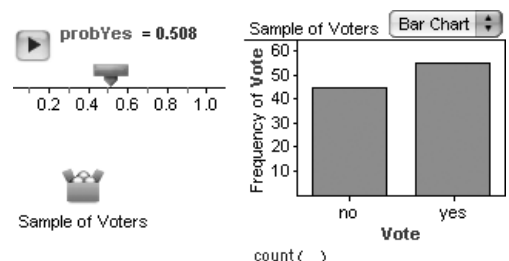
Next, make a graph of the votes. A bar chart will show at a glance whether the poll predicts the proposition will pass or fail.

15. Drag a graph from the shelf.
16. Drag *Vote* from the inspector to the graph's horizontal axis.

You should now have the three objects shown at right.

17. Choose **Rerandomize** from the **Collection** menu. Do this several times.

Each time you rerandomize, the bars in the bar chart will change, reflecting the results of a new sample.



Another way to rerandomize is to drag the slider thumb. When the slider value is near 1, most of the votes will be “yes,” and when the slider value is near 0, most of the votes will be “no.”

**Participant:** Stop working. **Presenter:** Answer questions. Demonstrate use of the slider’s inspector to set its lower and upper bounds. Allow 15 minutes for the remaining section of this tutorial.

## Defining and Collecting a Measure

18. Adjust the slider to about 0.5 to model a close election.

We’ve been hired to predict the accuracy of a poll conducted with a given sample size. To figure that out, we want to record the prediction of each poll taken. You do that with a *measure*.

19. In the inspector, click the **Measures** tab.

As yet, no measures have been defined.

20. Click <new> and type `proportionOfYes` for the measure’s name.

21. Double-click in the Formula cell for *proportionOfYes*.

22. Enter the formula `proportion(Vote="yes")` as shown at right.

23. Rerandomize several times, and observe the values for *proportionOfYes*.

Measure	Value	Formula
proportionOfYes	0.55	proportion(Vote = "yes")
<new>		

Now we’re ready to perform the poll of 100 voters many times, each time recording the proportion of voters who say they will vote in favor of the proposition.

24. With the collection selected, choose **Collect Measures** from the **Collection** menu.



Measures from Sample of Voters

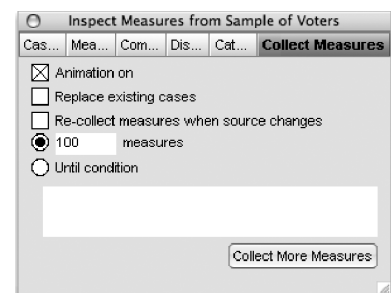
You should see a new collection, as shown at right.

This collection will contain the results of repeatedly conducting the poll.

25. Double-click the measures collection to show its inspector.

26. Change the number of measures to collect from 5 to 100.

27. Click **Collect More Measures**.



Fathom rerandomizes the Sample of Voters collection 100 times, each time computing a value for *proportionOfYes* and storing that value in a new case in the measures collection. The green ball crossing the screen reinforces the idea of a measure being collected.

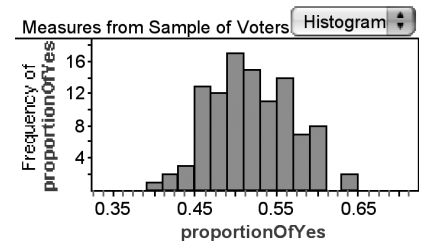
(To interrupt collecting measures, press the Esc key. Turning animation off speeds things up considerably.)

Let's take a look at the results.

28. Make a new graph.
29. Go to the **Cases** panel in the inspector.
30. Drag *proportionOfYes* to the horizontal axis of the graph.
31. Choose **Histogram** from the graph's pop-up menu.

Here are some observations you might make about the histogram of *proportionOfYes*:

- The distribution is mound-shaped and reasonably symmetrical.
- It looks as if the center of the distribution of sample proportions is near 0.5.



The spread of proportions that we get with a sample size of only 100 is quite large. We certainly couldn't accurately predict the results of a close election! Just how large is this spread?

32. Drag a summary table from the object shelf.
33. Drag *proportionOfYes* into the top row of the summary table.
34. Double-click the formula for *S1* and change it to `stdDev( )`.

This will compute the standard deviation of the proportions gathered in the measures collection.

Measures from Sample of Voters	
	<i>proportionOfYes</i>
	0.049165751

*S1* = stdDev( )

## Dependence of Spread on Sample Size

How does the spread of the proportions depend on the sample size? Let's try a larger sample size.

35. With the Sample of Voters collection selected, choose **New Cases** from the **Collection** menu. In the dialog box, type 300 and click **OK**.

The collection now has 400 cases.

36. Go to the **Collect Measures** panel in the measures collection inspector.
37. Uncheck Animation on.
38. Check Replace existing cases. (We don't want the new samples of size 400 to mix with the old samples of size 100.)
39. Click **Collect More Measures** to collect 100 measures of 400 cases each.

We expect that the larger the sample, the smaller the variation in the observed population proportion in favor of Prop A. But *how much* smaller? What do you find for the standard deviation of sample proportions with a sample size of 400? What do you think would be the standard deviation if the sample size were 1600?

You now have a tool to use with the newspaper staff members. You could sit down with them and try different sample sizes until you find one for which the variability is less than some desired number of percentage points. (If you need to make a sample collection with *fewer* cases, select all the cases, delete them, and then add the desired number.)

## Going Further

- Determine a sample size such that twice the standard deviation of sample proportions is about 0.03, or 3%, when *probYes* is 0.5.
- When we actually conduct a poll, we do not know the true proportion of the population. (After all, that’s why we’re doing the poll.) Probability theory predicts that we can expect to capture the true proportion 95% of the time in the range

$$\pm 1.96 \sqrt{\frac{p(1 - p)}{n}}$$

where  $p$  is the proportion found in the sample and  $n$  is the number in the sample. Try modeling this in Fathom to see how close your simulation comes to theory.

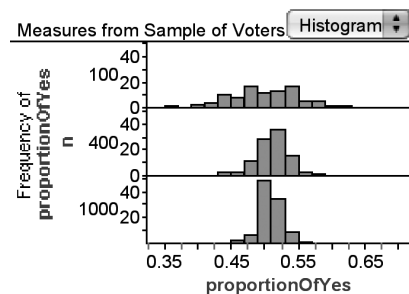
- We set up this simulation so it is easy to change the population proportion. With small sample sizes, population proportions near the extremes, that is, near 0 or 1, are problematic. Investigate why this might be so.



**Participant:** Stop working. **Presenter:** Answer questions. Demonstrate how to gather and display the data shown at right. Do this by defining a new measure in the Sample of Voters collection. Call it  $n$ , and give it the formula `count()`. In the measures collection, uncheck *Replace existing cases*. Collect measures with different sample sizes. Drag  $n$  to the vertical axis of the graph while holding down the Shift key so that  $n$  will be treated as categorical.

Have participants who have done further explorations show what they have found.

The next activity is taken from Chapter 5: Statistics, in *Teaching Mathematics with Fathom*. The steps are clear enough that it could be a first activity that students undertake with Fathom. It builds from a Fathom document that has already been set up to collect measures.





## Activity 3: Rolling Dice

---

Sometimes when you sample or collect measures, you don't know exactly how many things you want to collect. Instead, you want to sample (or collect) until some condition is met. In this activity, you will explore the answer to a traditional question: On the average, how many times do you have to roll a die to get a six?

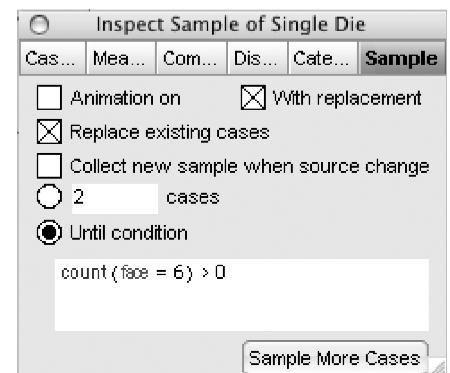
### MAKE A CONJECTURE

Before doing the simulation, think about what you expect to see.

- Q1** On average, how many times do you think you have to roll a six-sided die to get a six?

### EXPERIMENT

1. Open the document **RollingDice.ftm**. It should have two collections: Single Die and Sample of Single Die. Make sure you can see a case table for Sample of Single Die.
2. Double-click the Sample of Single Die collection to show the inspector. Click the **Sample** tab. The **Sample** panel should be set to sample two dice, with replacement.
3. You want to set up the simulation to roll the die until it shows a six. Click the **Until condition** button. The formula editor appears.
4. Enter an expression that will tell Fathom when to *stop* collecting samples: `count ( face=6 ) > 0` . Click **OK** to close the formula editor. Your inspector should look like the one shown here.
- Q2** Click **Sample More Cases** repeatedly. Describe what happens to the case table.
5. Now you will collect the numbers of die rolls that it takes to get a six. Click the **Measures** tab of the inspector.
6. Make a new measure, *numberOfRolls*, and give it the formula `count()`. This will be the number of cases in the Sample of Single Die collection. Close the inspector.
7. Select the Sample of Single Die collection and choose **Collect Measures** from the **Collection** menu. A new measures collection appears, called Measures from Sample of Single Die.



## Activity 3: Rolling Dice

(continued)

---

8. Make a case table for the new collection. You will see *numberOfRolls* for five runs of the simulation, because Fathom collects five measures by default. Show the measures collection's inspector and click the **Collect Measures** tab. Change the number of measures from 5 to 10 and check Replace existing cases. Click **Collect More Measures**. The simulation runs 10 times, collecting the numbers of rolls in the Measures from Sample of Single Die collection's case table.
- Q3 Make a histogram of *numberOfRolls*. Describe the graph. Which bar is the tallest? What is the shape of the graph overall?
9. Select the graph and choose **Plot Value** from the **Graph** menu. Give it the formula  $mean(numberOfRolls)$ . Close the formula editor.
- Q4 Sketch your graph. Explain in words what the number being plotted means.
10. Collect measures again, watching the graph. Continue to collect measures to answer the original question: What is the average number of rolls you need in order to get a six? Alter the parameters on the **Collect Measures** panel of the Measures from Sample of Single Die inspector as you see fit. (*Hint*: If you decide to collect a large number of measures, it will speed things up to uncheck the Animation on box.)
- Q5 What is your answer for the original question: On the average, how many times do you have to roll a die to get a six?
- Q6 Based on the simulation, what do you think is the *most likely* number of rolls you need to get a six (that is, the number that happens most frequently)? Explain. Sketch or refer to a graph if you need to.
- Q7 Explain how the most likely number of rolls can be different from the average number of rolls.


## EXPLORE MORE

1. The most likely number of rolls is the *mode*. You know the mean. What's the *median* number of rolls you need to get a six?
2. Suppose you changed the "middle" collection—the sample collection—to sample *without* replacement. What would happen to the mean number of rolls it would take to get a six?
3. On the average, how many rolls does it take to get a five *and* a six? A five *or* a six? Make predictions, then simulate both situations. Which is more likely, a statement with *and* in it or the same statement with *or* in it? Explain your answer and give examples.

## Tutorial 3: Testing a Hypothesis—Plant Growth

Charles Darwin believed that there were genetic advantages in having two sexes in both the plant and animal kingdoms. Some time after he wrote *Origin of Species*, he performed an experiment in his garden at Down House in Kent. He raised two large beds of snapdragons, one from cross-pollinated seeds, the other from self-pollinated seeds. He observed, “To my surprise, the crossed plants when fully grown were plainly taller and more vigorous than the self-fertilized ones.” This led him to another, more time-consuming experiment in which he raised pairs of plants, one of each type in the same pot, and measured the differences in their heights. He had a rather small sample and was not sure that he could safely conclude that the mean of the differences was greater than zero. His data for these plants were used by statistical pioneer R. A. Fisher to illustrate the use of a  $t$ -test.

In this tutorial, you’ll learn how to perform a  $t$ -test for the mean, generate simulated data from a normal distribution, and get a distribution for the  $t$ -statistic.

 **Presenter:** Allow 10 minutes for the first section.

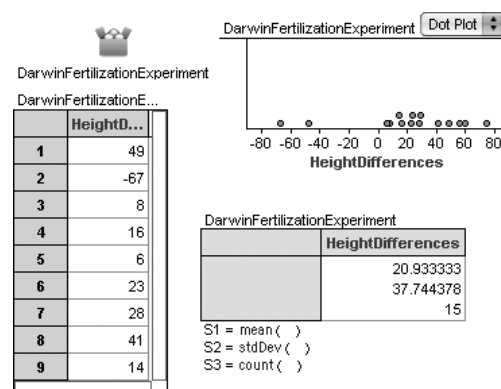
### Looking at Darwin’s Data

1. Choose **Open Sample Document** from the **File** menu.  
Open **Darwin.ftm** in the **Learning Guide Starters** folder.

This document contains the data for the experiment described above: 15 cases with 1 attribute.

2. Make a case table, a dot plot, and a summary table similar to those shown here.

We see that most of the measurements are greater than 0, meaning that the cross-pollinated plants grew bigger. But two of the differences are less than 0. Darwin did not feel justified in tossing out these two values and was faced with a very real statistical question.



### Formulating a Hypothesis

Darwin’s theory—that cross-pollination produced bigger plants than self-pollination—predicts that, on average, the difference between the two heights should be greater than 0. On the other hand, it might be that his 15 pairs of plants have a mean difference as great as they do (21-eighths of an inch) merely by chance. You can write out these two hypotheses in a text object to be stored with your document.

3. From the shelf, drag a text object into the document.

- Write the null hypothesis and the alternative hypothesis.  
At right you can see one way to phrase the hypotheses.

**Null Hypothesis:** The true mean difference of heights of plants is zero. The observed positive mean difference is due to chance alone.

**Alternative Hypothesis:** The true mean difference is greater than zero.

## Deciding on a Test Statistic

At the time of Darwin’s experiment, there was no good theory to deal with a small sample from a population whose standard deviation is not known. It was not until some years later that William Gosset, a student of Karl Pearson, developed a statistic and determined its distribution. Gosset published his result under the pseudonym Student and the statistic became known as Student’s *t*. When the null hypothesis is that the mean is zero, the *t*-statistic is simply

$$\frac{\bar{x}}{s/\sqrt{n}}$$

where  $\bar{x}$  is the observed mean, *s* is the standard deviation, and *n* is the number of observations.

Let’s compute this statistic for Darwin’s data using one of Fathom’s built-in statistics objects.

- Drag a test object from the shelf. An empty test appears.
- From the pop-up menu in the upper-right corner, choose **Test Mean**.

As shown at right, the test object allows us to type summary statistics into the blue text. This is very useful when you don’t have raw data.

- Try editing the blue text. You can, for example, enter the summary statistics for Darwin’s data.

Here are some things to notice.

- Changing something in one part of the test may affect other parts. For example, editing the AttributeName field in the first line also changes it in the hypothesis line and in the last paragraph.
- In the hypothesis, clicking on the “is not equal to” phrase brings up a pop-up menu from which you can choose one of three options. For Darwin’s experiment, we want the third option because his hypothesis is that the true mean difference is greater than 0. (In statistics terminology, we want this to be a one-tailed, rather than a two-tailed, test.) Notice that making this change alters the phrasing of the last line of the test as well.

From Summary Statistics Test Mean ▾

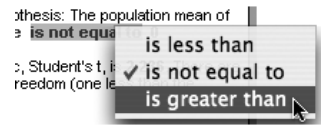
Attribute (numeric): unassigned

Attribute: **AttributeName**  
 Sample count: 20  
 Sample mean: 0.5  
 Standard deviation: 1  
 Standard error: **0.223607**  
 Alternative hypothesis: The population mean of **AttributeName** is not equal to 0

The test statistic, Student's *t*, is **2.236**. There are **19** degrees of freedom (one less than the sample size).

If it were true that the population mean of **AttributeName** were equal to 0 (the null hypothesis), and the sampling process were performed repeatedly, the probability of getting a value for Student's *t* with an absolute value this great or greater would be **0.038**.

- In addition to simple editing of numbers, we can also determine their value with a formula. For example, we might want to tie the sample count to a slider named  $n$  so that we could investigate the effect of different sample sizes. To show the formula editor, choose **Edit | Edit Formula** with the text cursor in the number whose value you wish to determine. These computed values display in gray instead of blue. Editing the value itself deletes the formula.



**Participant:** Stop working. **Presenter:** Answer questions. Make sure everyone understands how to edit the test hypothesis object. Demonstrate using a slider for a parameter such as the alternative hypothesis mean. Allow 10 minutes for the next section.

## Checking Assumptions

Gosset’s work with the  $t$ -statistic relied on an assumption about the population from which measurements would be drawn, namely, that the values in the population are normally distributed. Is this a reasonable assumption for Darwin’s data?

Height measurements of living things, both plants and animals, are usually normally distributed, and so are differences between heights. But we might worry, because the two negative values give a decidedly skewed appearance to the distribution.

Fathom can help us determine qualitatively whether this amount of skew is unusual. We’ll generate measurements randomly from a normal distribution and compare the results with the original data.

8. Make a new attribute in the collection. Call it *SimDiff* for simulated difference.

9. Give *SimDiff* the formula

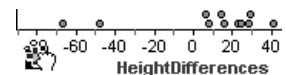
```
randomNormal ( mean (HeightDifferences), s (HeightDifferences) )
```

This formula tells Fathom to generate random values from a normal distribution whose mean and standard deviation are the same as for Darwin’s data.

We want to compare the distribution of these simulated heights with the distribution of the original data. We can do that directly in the dot plot that already shows *HeightDifferences*.

10. Drop *SimDiff* on the plus sign on the horizontal axis.

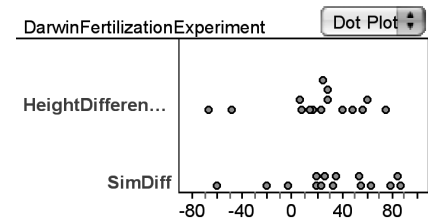
The graph now shows both attributes.



One set of simulated data doesn’t tell the story. We need to look at a bunch.

- With the graph selected, choose **Rerandomize** from the **Collection** menu.

Each time you rerandomize, you get a new set of 15 values from a population with the same mean and standard deviation as the original 15 measurements.



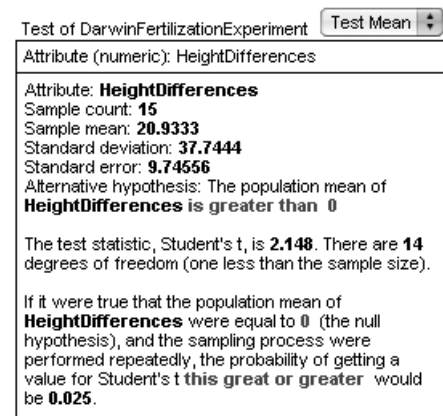
A bit of subjectivity is called for here. Does it appear that the original distribution is very unusual or does it fit in with the simulated distributions?

## Testing the Hypothesis

Once we have decided that the assumption of normality is met, we can go on to determine whether the  $t$ -statistic for Darwin’s data is large enough to allow us to reject the null hypothesis.

In step 7, we typed the summary values into the test as though we didn’t have the raw data. But we are in the fortunate position of having the raw data, so we can ask Fathom to figure out all the statistics using those data.

- Drag *HeightDifferences* from the case table to the top pane of the test, where it says “Attribute (numeric): unassigned.”
- If the hypothesis line does not already say “is greater than,” then choose it from the pop-up menu.



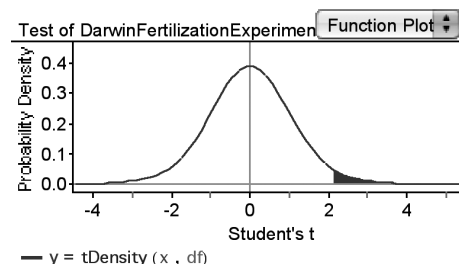
The last paragraph of the test describes the results. If the null hypothesis were true and the experiment were performed repeatedly, the probability of getting a value for Student’s  $t$  this great or greater would be 0.025. This is a low  $P$ -value, so we can safely reject the null hypothesis and, with Darwin, pursue the theory that cross-pollination increases a plant’s height compared to self-pollination.

**Participant:** Stop working. **Presenter:** Answer questions. Demonstrate the non-verbose form of the hypothesis test and how to set the verbose/terse preference. Demonstrate how to hide or iconify objects to free up screen space. Allow 10 minutes for the next section.

## Looking at the $t$ -Distribution

It is helpful to be able to visualize the  $P$ -value as an area under a distribution.


- With the test selected, choose **Show Test Statistic Distribution** from the **Test** menu.



The curve shows the probability density for the  $t$ -statistic with 14 degrees of freedom. The shaded area shows the portion of the area under the curve to the right of the test statistic for Darwin's data. The total area under the curve is 1, so the area of the shaded portion corresponds to the  $P$ -value for Darwin's data.

## Going Further

- What effect do the negative values have on the results? Drag one of the negative points to the left to make it even more negative. How far can you go before the  $P$ -value increases to 0.05?
- Darwin had 15 observations to work with. Could he have gotten away with fewer? Set up a test mean in which you type values into all the parameters of the test except the number in the sample to which you attach a slider,  $n$ , using the formula  $\text{round}(n)$ . What is the smallest sample size (with the same mean and standard deviation) that results in a  $P$ -value of 0.05 or less?

 **Participant:** Stop working. **Presenter:** Answer questions. Demonstrate how to connect a slider to a value in the test object. A good candidate is the value against which the observed mean of height differences is tested. Have participants show what they found for further explorations.

## Using Fathom as a Demonstration Tool

Sometimes you don't have time for students to do a full activity in the computer lab. You can use Fathom as a demonstration tool in the classroom to dynamically illustrate mathematical concepts and easily test thought experiments. Here are two examples of discussions you might have with a class. Work your way through the instructions for each discussion as though you were preparing to use it tomorrow.

**Presenter:** *Participants should work in pairs for this activity. There should be enough time for everyone to work through at least one of the interactive discussions. Instruct some people to start with the first discussion and others to start with the second. Have a volunteer pair go through each discussion as if they were teachers preparing to use the material in a lesson the next day.*

### Discussion 1—Dice Sums Distributions

You are teaching a probability unit. Your objective in this discussion is for students to be able to describe the frequency distributions for the rolls of a single die and for the sum of a pair of dice. You also want students to be able to explain why the distribution for the sum of two dice is “triangular.” Before this discussion students have (possibly in a math class) rolled dice by hand and gathered data.

#### Distribution for One Die

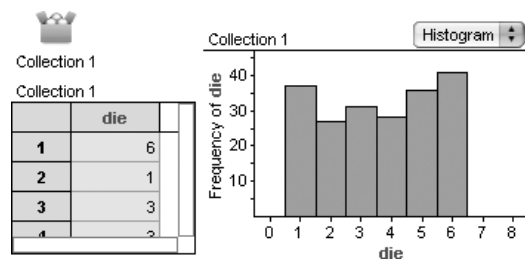
**Teacher:** “Suppose you roll a die a couple hundred times and you record each roll. Working with a partner, sketch a graph that shows a summary of what you would expect to get.”

1. Have a few students show and explain their sketch of a graph.

The accuracy of the graphs will strongly depend on students' prior experience with graphing data.

2. Simulate the situation in Fathom, as shown at right.
3. Ask students to talk with their partners about the ways in which their graph is similar to and different from the Fathom graph. Rerandomize a couple of times to show several different graphs.

**Teacher:** “Suppose I want to plot a function that shows where the tops of the histogram bars should be. What function should I plot?”

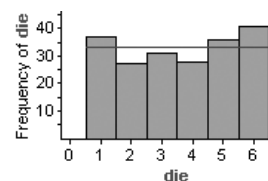




- Take student suggestions, and plot them.

Students may not quite get to the correct answer, which is

$$\text{Frequency of die} = \text{if } \{ \langle \text{die} > 0.5 \rangle \text{ and } \langle \text{die} < 6.5 \rangle \} \left\{ \begin{array}{l} \text{count} ( \quad ) \\ \dots \\ 6 \end{array} \right.$$



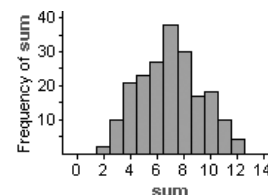
### Discussion Questions

- What is the total area of the gray bars? What is the area of the rectangle formed by the blue line, the outsides of the first and last bar, and the horizontal axis?
- How would this graph change if there were 1000 dice instead of 200? (*Teacher:* Demonstrate by adding new cases.)
- Explain why the graph looks the way it does. Why aren't some of the bars much higher than others?

### Distribution of Sums for Two Dice

*Teacher:* “We’re going to do this simulation again, but this time we’ll throw two dice instead of one, and we’ll record the sum of the two dice. Again, work with a partner to sketch a graph that shows what you expect to get.”

- Have various students show and explain their sketches.
- In Fathom, add an attribute for a second die and the sum of the two dice. (It helps to rename the first attribute.) Plot the sum as shown at right.



*Teacher:* “Discuss with each other the ways in which your graph is different from and similar to the graph that Fathom produced.” Again, rerandomize to show different graphs.

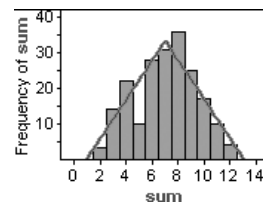
- After some whole-group discussion, make the breakdown plot shown at right and ask, “How does this plot help explain the distribution for the sum?” (To make the breakdown plot, hold down the Shift key when you release the attribute on the axis.)
- As students discuss the question, click on one bar after another in the histogram. They should see that one diagonal lights up in the breakdown plot for each bar in the histogram. Each cell in the breakdown plot represents an equally likely roll of the two dice. The height of each bar in the histogram is proportional to the number of cells in the breakdown plot that correspond to it.



**Teacher:** “What function can I plot on top of the histogram that will best show its shape?”

- Students may need time to work on this before you try plotting some of their ideas. One solution is

$$\text{Frequency of sum} = \text{count} \left( \frac{6 - |7 - \text{sum}|}{36} \right)$$



### Discussion Questions

These questions could be homework. You could also model these situations for the class.

- What will the distribution look like if we have 1000 rolls of the dice? How will it differ from the one with 200 rolls? Will the plotted function still match? Why or why not?
- Suppose we used 10-sided dice instead of 6-sided dice. What would the distribution of sums look like then?
- Suppose we used three dice instead of just two. What would the distribution of sums look like?

### Discussion 2—Guess and Check

Imagine that you have been using the “guess and check” method in teaching a beginning algebra class. In this discussion, you’ll demonstrate how Fathom can work as a spreadsheet to ease the burden of calculation and to get from guessing to an equation to solve.

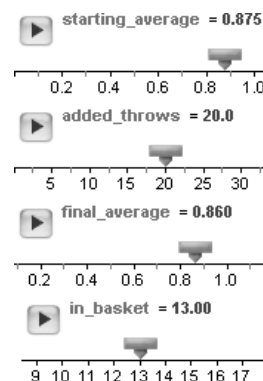
There is a Fathom sample document, **Guess and Check.ftm**, that works through this same problem. Here is the problem:

Melissa’s free throw average is 0.875. If she were given 20 more free throws and got 13 of them in the basket, her new free throw rate would be 0.860. How many free throw opportunities has she had so far?

**Teacher:** “There are four numbers in this problem. I’ll make a slider for each number. Give me a good name for each slider.”

- Make four sliders using names students suggest. Adjust the slider bounds to something appropriate to each.

**Teacher:** “Make a guess at the answer, and work your way through the calculations to check whether her free throw average is the specified new free throw rate.”



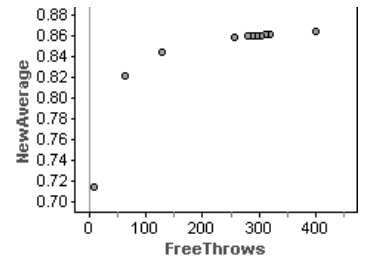
2. After giving students some time to do work on their guesses, take someone's work and convert it to attributes in a Fathom collection. (Be sure to use slider names rather than numbers.) Here is one possibility.

FreeThrows	NumGottenSoFar	NewFreeThrows	NewNumGotten	NewAverage
	FreeThrows-starting_average	FreeThrows + added_throws	NumGottenSoFar + in_basket	$\frac{\text{NewNumGotten}}{\text{NewFreeThrows}}$
64	56	84	69	0.821429
128	112	148	125	0.844595

3. Ask for a second guess and fill it in as the value of *FreeThrows* for the second case. **Teacher:** "Is this guess correct? If not, is it closer to or farther from the correct guess than the first guess?"
4. Continue to ask for guesses. Even if you get the "correct" guess, keep asking for more so that you have six cases or more.
5. Plot the checks versus the guesses, as shown at right.
6. Plot the function

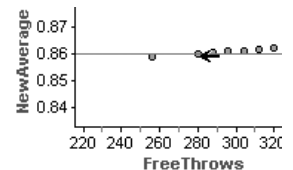
$$\text{NewAverage} = \text{final\_average}$$

This will appear as a horizontal line.



**Teacher:** "Just looking at this graph, how can we determine the answer to the question?"

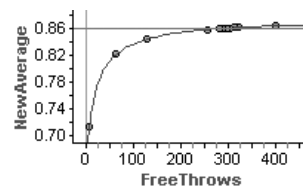
Students should conclude that the point that is on the horizontal line is the one that solves the problem. It helps to zoom in a bit.



**Teacher:** "What function could we plot that will go through all the points?"

7. Through discussion, arrive at the conclusion that the formula for *NewAverage* with the prior expressions substituted so that everything is in terms of *FreeThrows* is exactly what is needed.

$$\text{NewAverage} = \frac{\text{FreeThrows-starting\_average} + \text{in\_basket}}{\text{FreeThrows} + \text{added\_throws}}$$



8. Drag or animate the *final\_average* slider.

**Teacher:** "When I drag this slider, what am I changing about the problem? (Melissa's new free throw rate) What happens to the 'answer' as I decrease Melissa's final average? (Melissa has started with fewer free throws.) Where is that answer on the graph? (at the intersection of the curve and the horizontal line)"

9. We want *NewAverage* to be *final\_average*, so we have

$$\text{final\_average} = \frac{\text{FreeThrows} \cdot \text{starting\_average} + \text{in\_basket}}{\text{FreeThrows} + \text{added\_throws}}$$

**Teacher:** “Solve this equation for *FreeThrows*. Plug in the numbers, and see what you get for an answer. Does it fit what we got by zooming into the graph?”

10. Make a summary table, and use it as a calculator, as shown at right.

no data	
Drop an attribute here	
	280
S1 = $\frac{\text{added\_throws} \cdot \text{final\_average} - \text{in\_basket}}{\text{starting\_average} - \text{final\_average}}$	

**Teacher:** “Now we have a ‘complete’ solution to the original problem that will work for any values of the four numbers (often called *parameters*). Do you think there is *always* a solution?”

11. Choose one of the sliders and vary its value to see if there is always a solution. Discuss situations in which a number of throws or baskets is a noninteger. Discuss situations in which the answer is negative or infinite.

### Discussion Questions

- Often, the hard part of solving an algebra word problem is writing the equation that describes the situation. How does the guess-and-check method attempt to help with this?
- What use is there in defining sliders for the numbers in the original problem?
- Do you think there might be problems that can only be solved by guess and check; that is, problems that do not have an equation that can be solved? (Yes, there are!)

**Presenter:** *Have participants discuss things to consider when using Fathom as a demonstration tool. What are some advantages? Obstacles? What other tools might you use for this, and how is Fathom different? What are some other topics that would benefit from dynamic demonstration?*

## Fathom Demo Documents

The activities and tutorials in this guide introduce you to a small portion of what you can do with Fathom. Fathom comes with more than 300 sample documents. About 100 of these use techniques in Fathom; in other words, they are more than data. A tour through these sample documents is a fairly quick way to get an idea of the capabilities and possibilities for using Fathom in teaching math and statistics.

Here are just a few of the documents you might want to look at. To open a sample document, choose **Open Sample Document** from the **File** menu. To browse descriptions of all the sample documents, open Fathom Help and look in the Sample Documents section.

**ScreenSaver.ftm:** This document has a kind of parametric plot. Animate the sliders to make interesting Lissajous figures. This could be an inspiration for student projects. Look in **Mathematics | Algebra**.

**TexasCounties.ftm:** This document shows how to use an open collection to display meaningful information—in this case, median values of homes in the counties of Texas. Look in **Social Science | United States**.

**Normal.ftm:** The collection in this document has 100 numbers from a normal distribution whose mean and standard deviation are determined by sliders. A histogram with a normal distribution curve overlay and a normal quantile plot allow you to gain some insight into randomness and normal distributions. Look in **Statistics | Distributions**.

**DerivativeDef.ftm:** This document graphically illustrates the definition of a derivative. Look in **Mathematics | Calculus**.

**CompoundInterest.ftm:** This is a slider-parameterized model of compound interest based on initial investment, interest rate, periods per year, and payment per period. It comes with questions for students to investigate, but it also can be used (without the text object) as a demonstration document. Look in **Mathematics | Algebra**.

**Random Triangle.ftm:** This is a simulation of breaking a stick in two places and determining if they can make a triangle. The plot of segment lengths shows an interesting pattern. Look in **Mathematics | Geometry**.

**CentralLimitTheorem.ftm:** This document gives you a population that is nonnormal and the ability to sample from it and observe the sampling distribution's approach to normal as the sample size increases. Look in **Statistics | Sampling**.

## Sample Activities

*Teaching Mathematics with Fathom*, which comes with Fathom, contains 25 activities for using Fathom in teaching mathematics and statistics. We've included a few of these activities here to help show how Fathom can illuminate concepts and pose worthwhile challenges. The activity notes for one activity are also included.

You can open any files required for the activities by navigating to the **Teaching Math with Fathom** folder within the **Sample Documents** folder.

**Mammals' Brains:** In this Algebra 1 activity, students use ratios and one-variable graphs to answer the question, "Which mammal has the largest brain?" Activity notes are included.

**Sequence of Squares:** This open-ended activity could be used in Algebra 1 or Geometry. Students look for a pattern in the differences of squares and then verify their conjectures.

**Moore's Law:** In this Algebra 2 activity, students fit a curve to data about the power of computers. They use residuals to judge their fit.

**Jupiter's Moons:** In this Precalculus-level activity, students use sliders to adjust the phase of a periodic function and determine its equation.

**Creating Error:** In this Statistics activity, students study how error and scale affect making a model for data and the accuracy of that model.

# Mammals' Brains

In this activity, you'll study the brain sizes of 62 mammals. You'll use Fathom to find the mammals with the largest and smallest brains. You'll see that your answers depend on how you define "largest" and "smallest."

## MAKE A CONJECTURE

Before looking at the data, think about what you expect to see.

- Q1** Assume that you could weigh the brains of the African elephant, mountain beaver, human, and rat. Predict their order from largest to smallest brain, based on brain weight. Explain your reasoning.
- Q2** Assume that you could weigh these mammals' bodies, too. Would the four body weights go in the same order as the brain weights?
- Q3** Assume that you calculate the ratio of brain weight to body weight for these four mammals. Predict their order from largest to smallest brain, based on brain-to-body ratio. Explain your reasoning.

## INVESTIGATE

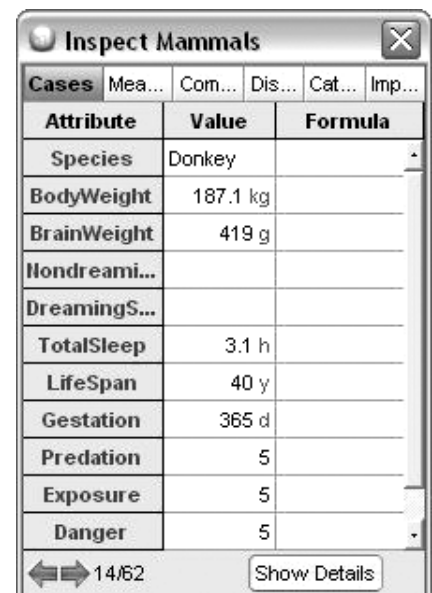
Now you'll use Fathom to look at the actual data.



- 1.** Open the document **Mammals.ftm**. A box of gold balls appears. This is a *collection*. It stores the data.
- 2.** Double-click the collection. The collection *inspector* opens. Click the **Cases** tab to show the data about one mammal. Each mammal is a *case*. Each piece of data, such as *Species*, *BodyWeight*, or *BrainWeight*, is an *attribute*. Use the arrows at the bottom to browse through the cases.
- Q4** Find the brain weight (with units) of a human.
- Q5** Find the body weight (with units) of a rat.

A *case table* is a good way to look at all of the data at once.

- 3.** Select the collection by clicking it. Choose **New | Case Table** from the **Object** menu. A table appears that shows all of the data.



Attribute	Value	Formula
Species	Donkey	
BodyWeight	187.1 kg	
BrainWeight	419 g	
Hondreami...		
DreamingS...		
TotalSleep	3.1 h	
LifeSpan	40 y	
Gestation	365 d	
Predation	5	
Exposure	5	
Danger	5	

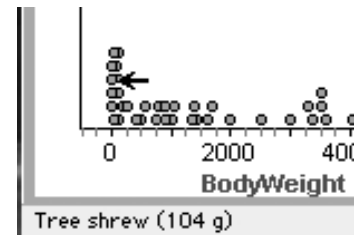
14/62 Show Details

- Q6** Which mammal has the brain that weighs the most? The least? (*Hint: You can order the data by selecting an attribute's column and choosing **Sort Ascending** or **Sort Descending** from the **Table** menu.*)
- Q7** List the African elephant, mountain beaver, human, and rat in order from largest brain weight to smallest brain weight.
- Q8** List these four mammals in order from largest body weight to smallest body weight. Is this the same order as for their brain weights?



Graphs are another convenient way to look at the data.

- 4.** Drag a new graph from the object shelf into your document. An empty graph appears.
- 5.** Drag the attribute name *BodyWeight* from the case table to the graph. Drop the attribute below the horizontal axis to create a dot plot. Point the cursor at one dot to see that body weight in the lower-left corner of the Fathom window. Click the dot to highlight the case everywhere and show that case in the inspector.



- 6.** Make another dot plot of *BrainWeight*.
- Q9** Using a dot plot, find the mammal whose brain weighs the most.
- Q10** Are you able to use a dot plot to find the mammal whose brain weighs the least? Explain how you found it, or explain why you couldn't.
- Q11** Do the brain weights of all 62 mammals go in the same order as their body weights? Explain how you found your answer, and compare this answer to your answer for Q8. What do the results mean?
- 7.** Select the dot plot of *BrainWeight*. Use the pop-up menu in the upper-right corner to change to a histogram. A *histogram* shows the number (or *frequency*) of values that fall in certain intervals (or *bins*). Put your cursor over one of the bars to see the frequency of cases in that bin. Click a bar to highlight all of those cases everywhere.
- Q12** Which mammals have a brain weight between 500 g and 1000 g?

You've found the mammals that have the largest and smallest brains, based on brain weight. Does that mean they have the largest and smallest brains for their overall size?

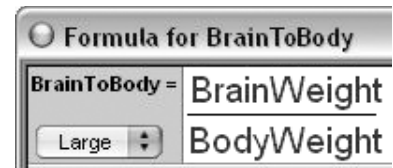


- Q13** By hand or with a calculator, find the ratio of brain weight to body weight for the mountain beaver. What are the units of your ratio? What are some equivalent ways to write your ratio?

Next you'll create an attribute that finds each mammal's brain-to-body ratio.

- 8.** In the case table, scroll to the far right. Click **<new>** in the last column, type `BrainToBody`, and press **Enter**. An empty column appears.

- 9.** Select the *BrainToBody* column and choose **Edit Formula** from the **Edit** menu. The *formula editor* opens. Type `BrainWeight/BodyWeight` and click **OK**. The ratios appear.



- Q14** In the case table, find the brain-to-body ratio (with units) for the mountain beaver. Is this one of the answers you got for Q13?
- Q15** Which mammal has the largest brain for its overall size? The smallest? What are the ratios? Are these the mammals with the brains that weigh the most and the least?
- Q16** List the African elephant, mountain beaver, human, and rat in order from largest brain-to-body ratio to smallest brain-to-body ratio. Is this the same order that you predicted in Q3?
- Q17** The order in Q16 should not be the same as the order in Q7. What does it mean about the sizes of these mammals' brains compared to the sizes of their bodies?
- 10.** Make a graph of *BrainToBody* and change it to a box plot. A *box plot* shows how the data values divide into fourths (or *quartiles*). The upper and lower segments (or *whiskers*) each represent one-fourth of the values, and each half of the box represents one-fourth of the values. The vertical segment in the middle of the box is the *median*, and the distance between the ends of the box is the *interquartile range*.
- Q18** Do more mammals have brain-to-body ratios above 0.01 or below 0.01?
- Q19** Approximate the median and interquartile range for a mammal's brain-to-body ratio.

A *summary table* is a more precise way to find numerical statistics.



- 11.** Drag a new summary table from the object shelf into your document. Drop the attribute *BrainToBody* in the row or column of the summary table. The mean appears automatically.
  - 12.** Select the summary table and choose **Add Five-Number Summary** from the **Summary** menu. The five number summary is added to the summary table.
  - 13.** Choose **Add Formula** from the **Summary** menu. Type `iqr()` and press **Enter**. The interquartile range is added to the summary table. You may have to resize the table to see all the formulas. If you click a formula, its value will be highlighted in the table.
- Q20** What are the median and interquartile range? Round your answers to four decimal places.
- Q21** What is a typical value for a mammal's brain-to-body ratio? Explain how you chose your answer.

## EXPLORE MORE

- 1.** Create a new attribute that gives each mammal's brain weight as a *percentage* of its body weight. What formula did you use? How do the brain-to-body percentages compare to the brain-to-body ratios? How does a graph of the percentages compare to a graph of the ratios? Does this make sense mathematically?
- 2.** From the data that you have, discuss whether brain weight would be a good way to judge intelligence. Would brain-to-body ratio be better?

### MAMMALS' BRAINS (PAGE 43)

**Activity Time:** 30–50 minutes

**Required Document:** *Mammals.ftm*

**Fathom Prerequisites:** Students should be able to

- Start Fathom
- Open a document

**Fathom Skills:** Students will learn how to

- Work with case tables
- Define an attribute with a formula
- Create graphs (dot plots, histograms, box plots)
- Select cases in a case table or graph
- Create a summary table of basic statistics

**Mathematics Prerequisites:** Students should be able to calculate the ratio of two quantities and find the median of a data set.

**Mathematics Skills:** Students will learn how to use one-variable graphs (dot plots, histograms, and box plots) to find particular values (such as greatest, least, and median) and use ratios to compare quantities. *Optional:* Students will explore bivariate data (see Extensions 3 and 5).

**General Notes:** At first, the mathematics in this activity may seem relatively simple—to identify particular values in a data set and to calculate ratios. However, you may be surprised that many students don't fully understand the need for ratios when comparing quantities of unequal size. For example, when a student looks only at the brain weights, she might say, "The African elephant has the largest brain." However, it's unfair to judge a mammal's brain size without considering the overall size of the mammal. When the same student looks at the ratios of brain weights to body weights, she may revise the conclusion to "The African elephant has the smallest brain."

The idea that the same data set can give two entirely different results is a subtle, yet important, aspect of this activity. The catch is to fully understand the definitions of "largest" and "smallest." Absolute brain weight and the ratio of brain weight to body weight are both acceptable ways to define brain size, albeit the latter is arguably more

meaningful. Seeing firsthand that the same data can be "manipulated" to give different conclusions will help students be careful and critical when they run across data in the media.

While comparing attributes and calculating ratios, Algebra 1 students will also get valuable practice in dimensional analysis. In the activity proper, students need to convert between grams and kilograms and understand what it means for a ratio to have no units. If you use Extension 2, students can also get exposure to derived units.

This activity is designed to be an easy first-time introduction to Fathom. The step-by-step worksheet leads students through many Fathom fundamentals, such as what an attribute is, how to create a case table, and how to make a graph. In order to save room, the activity does not give specific instructions for starting Fathom or opening a document; most students will have enough computer experience to accomplish these tasks without explicit instructions.

Although written primarily for Algebra 1, this activity can be successfully used for Algebra 2, Precalculus, or Statistics. First, if you have never used Fathom in these upper-level classes, you may want to use this activity as a gentle introduction to the software. Second, the concept of "largest" brain naturally leads to the question of intelligence, and the scientific study of animal intelligence provides an excellent opportunity for curve fitting and data transformation (see Extension 5).

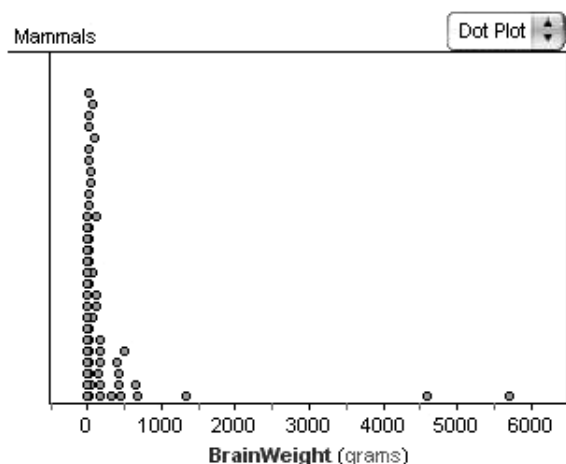
### MAKE A CONJECTURE

- Q1 Possible answer: elephant, human, beaver, rat.  
Students might pick this order based on relative size of the mammals.
- Q2 Possible answer: yes
- Q3 Possible answer: human, elephant, rat, beaver.  
Students might pick this order by assuming that more intelligent mammals have larger brains compared to their overall body size.

# Activity Notes

## INVESTIGATE

- Q4 1320 g
- Q5 0.28 kg. *Note:* Students will notice that the collection inspector and/or case table has some blank cells. This means that data were not available. Encourage students to leave these cells blank and to avoid the temptation to fill in meaningless data, such as zeros.
- Q6 Most: African elephant, 5712 g; least: lesser short-tailed shrew, 0.14 g
- Q7 African elephant, human, mountain beaver, rat
- Q8 African elephant, human, mountain beaver, rat. Yes, this is the same order as for brain weights.
- Q9 Students should use the dot farthest right in the graph of *BrainWeight*: African elephant, 5172 g.



- Q10 Because the dots are so densely clustered at 0 g, some students may have trouble picking out the least value. Other students may discover ways to adjust the axes and be able to zoom in enough to find the least value. The least value is still 0.14 g, the lesser short-tailed shrew.
- Note:* Even with the densely clustered dots, some students may be lucky enough to click at the bottom of the first stack of dots and find the lesser short-tailed shrew. This may lead them to erroneously believe that the stacks of dots always go least to greatest from bottom to top. This is not always the case. Actually, the dots are stacked relative to the cases' order in the case table. Students who find the lesser short-tailed shrew at the bottom of the stack

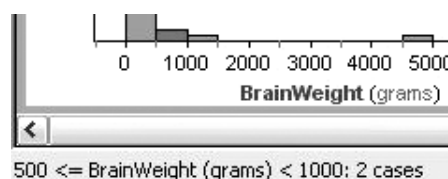
probably used **Sort Ascending** when answering Q6. As a counterexample, if students had last used **Sort Descending**, the lesser short-tailed shrew would have been at the top of the first stack of dots.

- Q11 No, not all of the brain weights go in the same order as the body weights. For example, the third largest body weight belongs to the giraffe, which has the fourth largest brain weight, while the third largest brain weight belongs to the human, which has the thirteenth largest body weight.

Students could answer this by clicking dots in reverse order in the dot plot of *BodyWeight* and noticing that the dots in the dot plot of *BrainWeight* are not in the same order. Students could also sort the case table by *BrainWeight*, and then by *BodyWeight*, and notice that the species do not stay in the same order.

This means that some mammals have proportionately larger or smaller brains for their overall size. The brain-to-body ratios (coming up in steps 8–11) will account for brain size relative to the overall size of the mammal.

- Q12 Giraffe (680 g) and horse (655 g). Students will probably click the bar between 500 and 1000, and see which cases are highlighted in the case table.



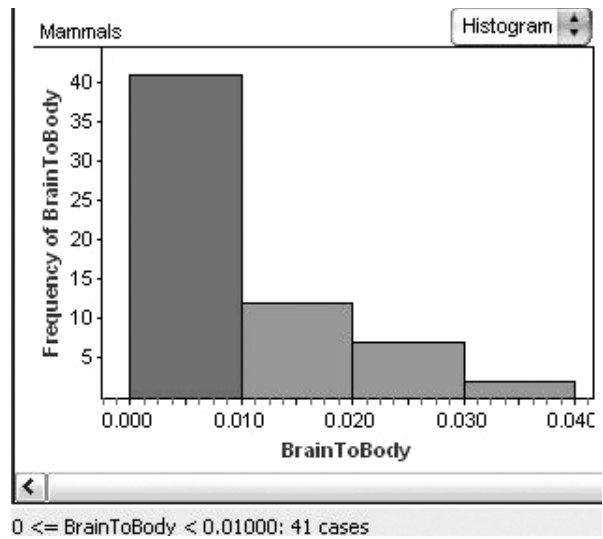
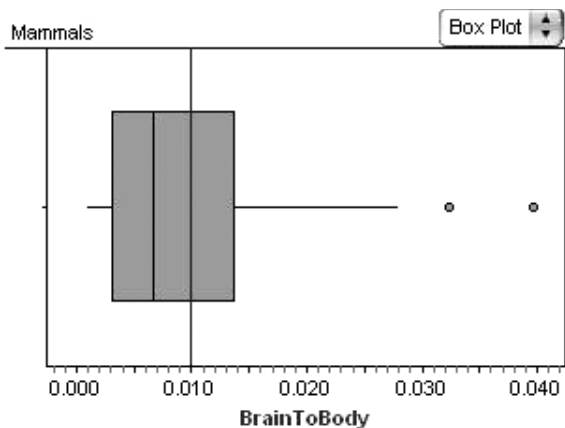
*Note:* If your students have formally learned about histograms, you should point out that the bin is officially  $500 \leq x < 1000$ , which does not include the upper bound.

- Q13 Possible answers: 6 g/kg, 0.006 g/g, 0.006 kg/kg, or 0.006 (no units). All four of these ratios are equivalent. Because both measurements are weights, it is best to convert to the same units (either grams or kilograms) before finding the ratio, and then write the ratio 0.006 without units.

## Activity Notes

- Q14 0.006; technically there are no units. *Note:* For students who could find only the ratio 6 g/kg for Q13, you should explain how Fathom performs unit conversions on compatible units, resulting in a value that has no units.
- Q15 Largest: ground squirrel, 0.0396; smallest: African elephant, 0.00086. No, these are not the mammals that had the brains that weigh the most and least; in fact, the African elephant is now on the opposite end of the spectrum.
- Q16 Human, rat, mountain beaver, African elephant. This may not be the same order predicted in Q3.
- Q17 It means, for example, that the human's brain is proportionately larger than the brains of the other mammals.
- Q18 More mammals have brain-to-body ratios below 0.01. In the box plot, the median value falls below 0.01, so more than 50% of the mammals have brain-to-body ratios below 0.01.

Students could also create a histogram and adjust the bin widths to show that 41 cases have brain-to-body ratios below 0.01.



- Q19 Possible answers: median: 0.0065; interquartile range: 0.0105
- Q20 Median: 0.0066; interquartile range: 0.0108

Mammals	
	0.0096242136
	0.00085843102
	0.0030857143
<b>BrainToBody</b>	0.0066108564
	0.013842482
	0.03960396
	0.010756768

S1 = mean( )  
 S2 = min( )  
 S3 = Q1( )  
 S4 = median( )  
 S5 = Q3( )  
 S6 = max( )  
 S7 = iqr( )

- Q21 Any answer with a reasonable explanation is acceptable. Some students might use the median or mean exactly as calculated in the summary table; they might try combining a measure of center with a measure of spread to give a range of values; they might use the distribution in a dot plot to make a subjective answer based on where the dots cluster; or they might use a histogram and pick out bins that have more values than others (essentially the mode). Some students may not give a value at all, preferring to argue that the data have so much spread that it is meaningless to state a "typical" value.

## Activity Notes

### EXPLORE MORE

1. The formula for brain-to-body percentage is  $100(\text{BrainWeight}/\text{BodyWeight})$ . Students should notice that the only difference between these values and the brain-to-body ratios is that the decimal point is moved two places to the right. Even the graphs will look the same except the axes will differ by a factor of 100. Students should realize that this makes mathematical sense because they've only multiplied the brain-to-body ratios by 100.
2. Students might say that neither brain weight nor brain-to-body ratio is a good way to judge intelligence. To support their answers, students might point to the position of “intelligent mammals” (such as human and chimpanzee) relative to less intelligent mammals. For example, if we use brain weight as a judge of intelligence, both species of elephants are more intelligent than humans, and the chimpanzee is on about the same level as a donkey. Using brain-to-body ratio as a judge of intelligence makes the ground squirrel the most intelligent mammal and there are seven other mammals more intelligent than humans.

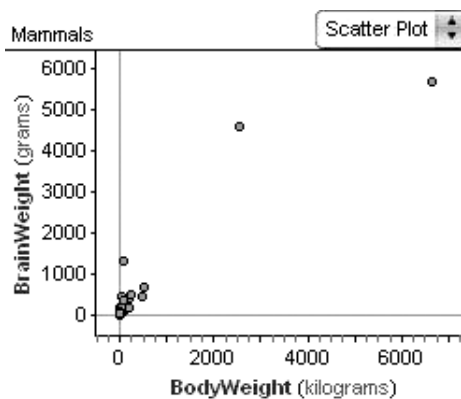
### EXTENSIONS

1. When making the box plot of *BrainToBody*, students will notice that two points appear at the end of the right whisker. Discuss why these two points may be excluded from the rest of the whisker. You may want to introduce the term *outlier*.  
  
For students who have already learned about outliers and different methods of calculating outliers, ask them to verify that these two points surpass the criterion  $\text{median} \pm 1.5 \text{ IQR}$ .
2. The data used in this activity come from a classic study of the sleep habits of 62 mammals (originally published in the article “Sleep in Mammals: Ecological and Constitutional Correlates” by Allison Truett and Domenic V. Cicchetti (1976), *Science*, November 12, vol. 194, pp. 732–734). Although the activity proper examines only brain weight and body weight, there are rich connections between the other attributes as well.  
  
Have students explore ratios of other attributes for these mammals, such as dreaming sleep to total

sleep, length of gestation to life span, or brain weight to length of gestation. Use the ratios to phrase and answer questions such as “Which mammal dreams the most?” or “Which mammal spends relatively the most time developing in the womb?”

*Note:* Depending on the attributes chosen, students may encounter situations when Fathom will create a “derived unit.” For example, when comparing brain weight to length of gestation, the units (grams and days) cannot be converted into compatible units. Instead, the attribute *BrainToGestation* has the unit grams per day.

3. To extend this activity beyond one-variable statistics, have students make scatter plots for pairs of attributes, such as *BrainWeight* versus *BodyWeight*.



Some bivariate graphs show clear relationships; this one, in particular, implies that as body weight increases, brain weight also tends to increase. Students can try fitting a line or any other function to the graph. (See Extension 5 for an exponential function that models this relationship.) *LifeSpan* versus *Gestation*, *TotalSleep* versus *Gestation*, and any combination of the three sleep attributes also provide graphs that have recognizable patterns.

As a further extension, students could try comparing some of the categorical attributes (*Predation*, *Exposure*, and *Danger*) with the numerical attributes or with each other. They can find an explanation of these attributes by double-clicking the collection to show the inspector and then clicking the **Comments** tab.

# Activity Notes

4. Some students may be interested in researching the study of intelligence in different species and writing a short paper on the mathematics used in this field of science. Two key words that students should research are *allometry* (the study of the growth of one part of an organism in relation to the growth of the whole organism) and *encephalization* (the amount of brain mass relative to body size).

Students will learn that some scientists use exponential functions to estimate brain weight as a function of body weight. Then the *encephalization quotient*—the ratio of actual brain weight to estimated brain weight for an animal of that size—is used to judge relative intelligence. Animals that have an actual brain weight greater than the estimated brain weight (an encephalization quotient greater than 1) are considered more intelligent.

Students may also run across competing theories on intelligence, such as judging intelligence based on the number of neurons in the brain rather than on brain size.

5. Students in Algebra 2, Precalculus, or Statistics can perform log-log transformations to find an exponential model for predicting brain weight as a function of body weight. As mentioned in Extension 4, some scientists use exponential functions to calculate the ratio of actual brain weight to expected brain weight (the encephalization quotient), which is an indicator of relative intelligence.

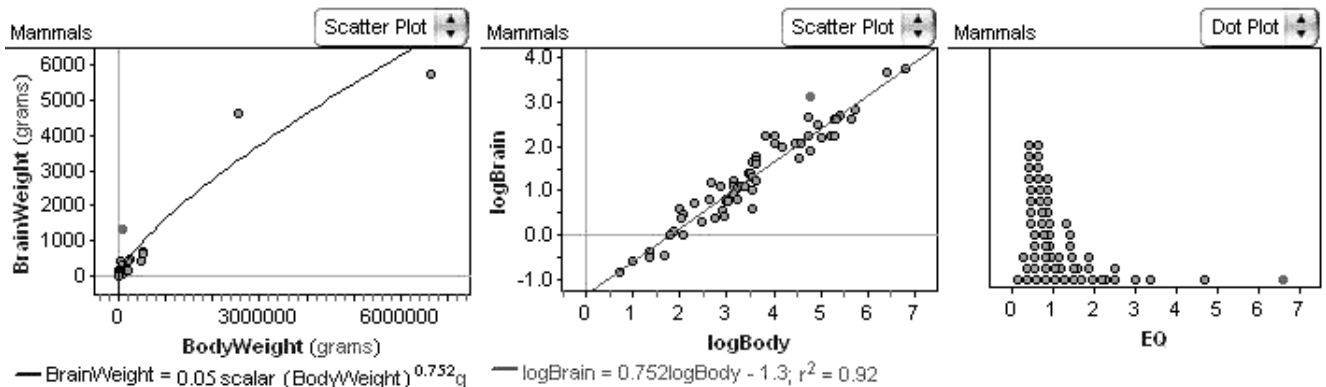
First, create the attributes *logBody* and *logBrain* and create a scatter plot. (*Note:* Before taking logarithms, change the units for both *BodyWeight* and *BrainWeight* to grams or kilograms.) Use a least-squares line on the transformed data and work backward to derive an exponential function that fits the original data. Below, *logBrain* versus *logBody* is modeled by the least-squares line  $\logBrain = 0.752 \logBody - 1.3$ . This converts to the exponential function  $BrainWeight = 0.05 \text{ BodyWeight}^{0.752}$ . (*Note:* In order to get appropriate units, enter the function as

$$BrainWeight = 0.05 \text{ scalar}(\text{BodyWeight})^{0.752} \text{ g}$$

## Mammals' Brains, Extension 5

Mammals

	Species	BodyWe...	logBody	BrainWe...	logBrain	EstBrain	EQ
<b>units</b>		grams		grams		grams	
<b>1</b>	Human	62000 g	4.79239	1320 g	3.12057	200.839 g	6.57242
<b>2</b>	Rhesus monkey	6800 g	3.83251	179 g	2.25285	38.108 g	4.69717
<b>3</b>	Baboon	10550 g	4.02325	179.5 g	2.25406	53.022 g	3.38539
<b>4</b>	Owl monkey	480 g	2.68124	15.5 g	1.19033	5.19115 g	2.98585
<b>5</b>	Chimpanzee	52160 g	4.71734	440 g	2.64345	176.363 g	2.49485
<b>6</b>	Ground squirrel	101 g	2.00432	4 g	0.60206	1.60775 g	2.48794
<b>7</b>	Patas monkey	10000 g	4	115 g	2.0607	50.9296 g	2.25802



## Activity Notes

---

The function *scalar* removes the units from *BodyWeight* during the calculation, and then “g” reapplies grams. Otherwise, the units would be  $\text{g}^{0.752}$ .)

The same exponential function is used to define an attribute (*EstBrain*) that calculates the estimated brain weights, and then another attribute (*EQ*) calculates the encephalization quotient. Notice that *EQ* puts the mammals in an order that most people (and many scientists) find reasonable, with humans as the most intelligent, followed by many of the primates. (See *bottom of previous page*.)



# Sequence of Squares

---

In this activity you'll explore the square numbers:

1    4    9    16    25    ...

You'll use mental math to recognize a relationship between these numbers and then use Fathom to confirm or refute your conjecture.

## MAKE A CONJECTURE

Without using Fathom or a calculator, look for a relationship between each square number and the one that follows it. Think about ways to use words, numbers, or algebra to make the conjecture as general as possible.

**Q1** What is the relationship between any pair of square numbers? Be as specific as you can.

## INVESTIGATE

Now use Fathom to test your conjecture. Make a collection in which one attribute contains the square numbers and another computes the relationship you predicted. Use any attributes or formulas that you want. In case you are stuck, here are some functions that you can research in Fathom Help: *caseIndex*, *prev*, *next*. You can also choose **New Cases** from the **Collection** menu to add cases.

**Q2** List the formulas that you used and describe what happened.

Look at your results in Fathom and compare them to your original conjecture. Use the next questions to help summarize your findings.

**Q3** Did your calculations match your conjecture? If not, explain why and write a revised conjecture.

**Q4** Did your calculations reveal any other patterns in the sequence? If so, explain.

## EXPLORE MORE

One of the problems with exploring data is that you don't know if your conjecture is *always* true. Even if your Fathom document has 5000 cases that show your conjecture to be true, you don't know for sure about case 5001. Make a convincing explanation (verbal, algebraic, or geometric) telling why your conjecture is true for all square numbers.

# Moore's Law

---

Gordon Moore (b. 1929) is cofounder and past chief executive officer (CEO) of Intel Corporation, a company known for making computer central processing units (CPUs). In 1965, three years before founding Intel, Moore made a famous observation that computer processors would get more powerful exponentially. The press called his observation “Moore’s Law,” and different variations of it have been stated ever since.

One way to measure a CPU’s power is to count the number of transistors on it. In this activity, you’ll look at data about Intel’s processors. You’ll test Moore’s Law by using an exponential function to model the data, determine how frequently the number of transistors doubles, and predict the future of CPUs.

## INVESTIGATE

1. Open the document **MooresLaw2004.ftm**. You’ll see a collection and case table. There is one case for each major Intel CPU from 1970 to 2004. The attributes tell you the name of the CPU, the year it was introduced, the number of transistors, and the number of years since 1970. (*Note: The year 1970 was chosen as the nearest multiple of 10 before the year of the first Intel processor.*)
2. Make a scatter plot of *Transistors* versus *Since1970*.
- Q1** Sketch the graph and describe its shape in one or two sentences. What mathematical curve does the graph resemble? What does that mean about the number of transistors over time?
3. Moore’s original observation was that “complexity . . . has increased at a rate of roughly a factor of two per year.” If you assume that “complexity” can be measured by number of transistors, you get the function  $Transistors = 2^{Since1970}$ . Plot this function on your graph.
- Q2** Describe how well the function fits the scatter plot. How could you transform the function to make it fit the data better?
4. Make sliders and use them as parameters to transform the function  $Transistors = 2^{Since1970}$ . Try any transformations that you think are necessary—translations, dilations, or reflections. Adjust your sliders to get the best fit possible. (*Hint: You may want to zoom in to check your fit. Select the graph. Hold down Ctrl (Win) Option (Mac) and click in the graph. Hold down Shift as well to zoom out.*)

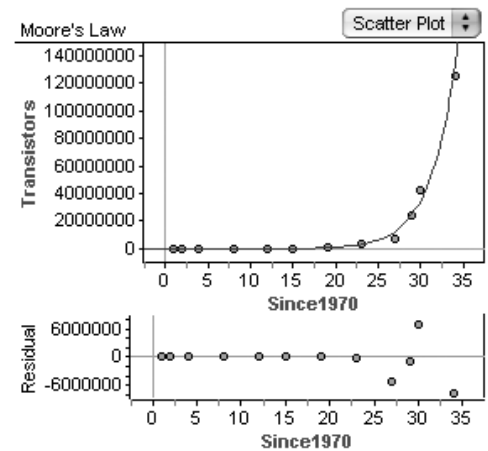
- Q3** List the sliders and transformations that you tried. Which ones helped you get the best fit? Are they the same transformations that you predicted in question Q2?
- Q4** Add the curve to your sketch in Q1. Record the final function, inserting the values of the sliders where necessary.
- Q5** According to your function, how many years does it take for the number of transistors to double? That is, what is the *doubling time*?
- 5.** Move the cursor over the graph of your function. The arrow cursor changes to a red point. Click and drag to trace the graph with the red point. Notice that the coordinates appear beside the point and in the bottom-left corner of the Fathom window. This gives you a way to approximate how many transistors a CPU would have in future years, according to your function model.
- Q6** Approximately how many transistors would a CPU have in 2005? In 2010?

A *residual* is the vertical distance between your function and an actual data point. A *residual plot* shows all of the residuals and helps you judge how well the function fits.

- 6.** Select the graph and choose **Make Residual Plot** from the **Graph** menu. A residual plot appears below your original graph.

- 7.** Adjust your sliders and watch how the residual plot changes as the function's fit changes.

- Q7** What happens to a residual when the function passes through the original data point? What happens when the function doesn't pass through the data point? What happens to the residual when the fit gets better? Gets worse?
- Q8** Which data points in your original graph have the most problematic residuals? Why did you pick these points?
- Q9** Use your sliders to find the best residual plot. Record this new function, inserting the values of the sliders where necessary. How does it compare to your function in Q4?



- Q10** According to your function, was Gordon Moore correct to observe that computer processors get more “complex” exponentially? Was he correct that the complexity doubles each year? How would you state Moore's Law based on your function?
- Q11** What do you think about using your Moore's Law function to make predictions for the future? Do you think transistors will continue to increase at the same exponential rate? Why or why not?

### EXPLORE MORE

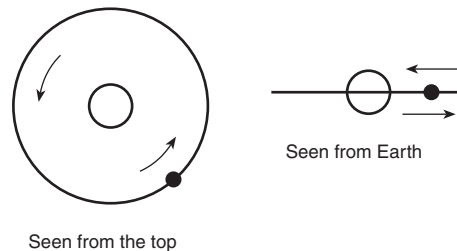
1. Research any new Intel processors that have been released since the Pentium 4. Compare the processor's actual number of transistors to the number of transistors that your function predicts for that year of release. How well did your model work?
2. You may have learned about logarithms. Create a new attribute that takes the logarithm of *Transistors*, and make a scatter plot of  $\log Transistors$  versus *Since1970*. Describe what happens and find a function to fit this new graph. How can you use this new graph to predict values for the future? How can you use the equation of the function of fit to get something like the exponential function you used before?

# Jupiter's Moons

---

Astronomers have been able to model many parts of the universe to accurately predict where planets, stars, comets, and moons will be at any time. In this activity, you'll look at data about one of Jupiter's four largest moons and estimate how long it will take for the moon to complete one full revolution around the planet.

If you look at one of Jupiter's moons from night to night, it seems to be swinging back and forth. It's really going around the planet, of course, but you're seeing the moon's orbit edge-on.



If you record the distance of the moon from Jupiter—positive to the west, negative to the east—as a function of time, you get a wavy curve. You will find the period of this curve.



## INVESTIGATE

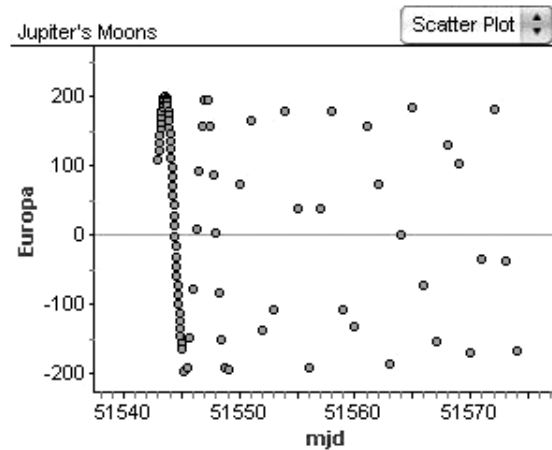
1. Open the document **JupiterMoons.ftm**. You should see a case table with ten attributes.

The attribute *mjd* stands for Modified Julian Date in decimal days. (The Modified Julian Date measures days since November 17, 1858.) The attributes for the four moons, *Io*, *Europa*, *Ganymede*, and *Callisto*, give the distance of the moon from Jupiter as seen from Earth, measured in arc seconds ( $1 \text{ arc second} = \frac{1}{3600} \text{ degree}$ ). Positive distances are to the west; negative distances are to the east.

## Jupiter's Moons

(continued)

- Choose one of the moons to use in this activity. You'll be able to experiment with the other moons later. Create a graph of the distance of the moon you chose as a function of  $mjd$ . Examine the table to understand why the initial points in the graph are tightly clustered together.



- What specific kind of periodic function does this graph appear to be? Estimate the period of the graph.

Now you'll attempt to find a function to model this graph.

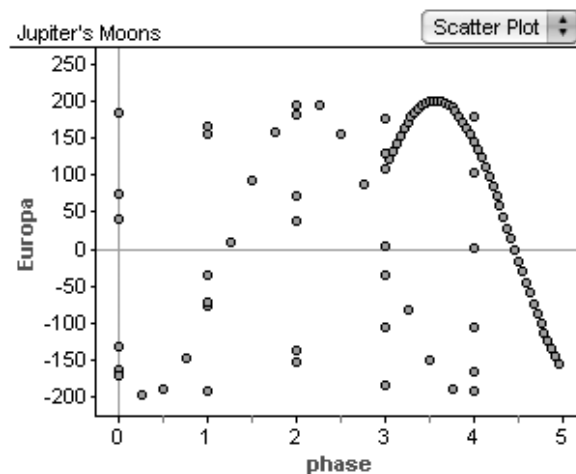
- Create a new slider called *period*.

The *phase* of a periodic function is how far a function is into its current cycle.

- In the case table, create a new attribute called *phase*. Enter a formula for *phase* that calculates the phase of the function using your slider for the period. (*Hint*: You want to find the time since the function has completed its last cycle. Use the *floor* function to help you find the number of whole cycles completed.)

- What formula did you come up with?

- Make a new graph that plots the distance of your moon as a function of *phase*.



- Adjust the *period* slider. Describe what you see in the graph. What period seems to best fit the data for your moon? How did you choose that period?

The *mjd* values are so large that any changes in the estimate for the period lead to a large change in the calculation of the *phase*.

**6.** Add a new attribute, *newjd*. Edit the formula for this attribute to be  $mjd - 51544$ .

**Q4** What date does the number 51544 correspond to in your model?

**7.** Edit the *phase* formula to use *newjd* instead of *mjd*.

**8.** Adjust the *period* slider until the points line up and show just a single cycle. (You may have to rescale the graph axes to see the complete cycle.) Try to be as accurate as possible by shortening the range of values for the slider. Do this by double-clicking on the slider axis and changing the lower and upper values. You will probably want to adjust the slider several times as you get more and more accurate.

**Q5** What value of *period* to three decimal places makes the points in the graph line up? How accurate do you think your estimate is?

You now have an estimate for the period. But your data cover only a few cycles. Next you'll test your estimate by using data from several months later.

**9.** Add these data to your case table.

Date	Modified Julian Date	Io	Europa	Ganymede	Callisto
July 1	51726.0	27.28	-26.23	-239.64	-196.76
July 2	51727.0	14.45	-152.89	-83.11	-34.57

**10.** Select the new points by holding down the Shift key as you click on the case number for each new point in your case table. The points should be highlighted red in the second graph.

**11.** Now adjust the value of *period* to make the two highlighted points in the *phase* graph fit the curve.

**Q6** Does your estimate for *period* change due to adding the new points? If so, what is your new estimate and why do you think it changed? Which estimate do you think is more accurate?

Now you'll create a function that models the data. You already know the period of the function (although the true period requires a bit of manipulation of the period you've already found). You'll need to come up with values for amplitude and phase displacement and decide whether to use the sine or cosine function.

12. Use your graphs to make an estimate for the amplitude of the function. Try using the *max* and *min* functions of Fathom to create a formula for the amplitude.
- Q7 State the formula and the value you found. Explain why this is a good starting point for estimation, but probably isn't the actual amplitude of the function.
13. Before you can find the phase displacement, you need to decide which function to use. Although you can use either sine or cosine, one may be easier than the other. Once you have chosen which function to use, estimate the phase displacement by inspecting the graph. (*Hint*: When you click on a point on the graph, the coordinates of the point are shown in the bottom-left corner of the Fathom window. Think of points that may be helpful in estimating the displacement.)
14. You may need to adjust the amplitude and phase displacement, because they are only estimates. To help with adjustment, create two sliders—*amplitude* and *displacement*.
15. Now you have all the information for the equation. Click on the *phase* graph and then choose **Plot Function** from the **Graph** menu. Enter a function for the graph, using *amplitude*, *displacement*, and *period* as constants in your equation. Adjust the sliders to improve your model.
- Q8 State the function that best matched your data. Could you write a different function that would model the data equally well? Explain.

## EXPLORE MORE

1. Plot your function on the first graph you created, replacing *phase* with *mjd*. Does the graph fit the data? Explain.
2. Predict the position of the moon you chose at midnight on your birthday in 2006. Remember to convert to the New Modified Julian Date.
3. Compare your findings with those of other members of the class. Compare your graphs. Which planet has the longest orbital period? Can you determine which moon is closest to Jupiter? Which moon is farthest away?



# Creating Error

---

In this activity, you will look at how errors affect the way you interpret linear fits to data.

## EXPERIMENT

1. Open a new Fathom document. Drag a new case table from the shelf and add 100 new cases.
  2. Create two attributes—*cause* and *effect*—with these formulas: For *cause*, use  $\text{randomNormal}(100, 10)$ . For *effect*, use *cause*.
  3. Make a scatter plot of *effect* as a function of *cause*, and show the least-squares linear regression line.
- Q1** What is the slope of the least-squares line? What is the value of  $r^2$ ? Is this what you would expect?

You have set up a very simple situation. Now let's make it more complicated. So far, *effect* has been perfectly related to *cause*. Now you'll add error to *effect* and see what happens.

4. Make a slider and name it *SD* (for standard deviation).
  5. Make a new attribute, *error*, and give it the formula  $\text{randomNormal}(0, SD)$ .
  6. Now change the formula for *effect* to  $\text{cause} + \text{error}$ . This adds a random, normally distributed amount to *cause*. That random amount has a mean of 0 and a standard deviation of *SD*, in this case, 5.
- Q2** Adjust the slider. What happens when *SD* is near 0?
- Q3** As *SD* gets larger, what happens to  $r^2$ , the square of the correlation coefficient? As *SD* gets larger, what happens to the slope of the least-squares line? (You can increase the scale of *SD* by dragging on the axis.) Explain why this happens.
7. Sometimes we get cause and effect mixed up. Make a second scatter plot. Put *effect* on the horizontal axis and *cause* on the vertical axis. Add a least-squares line.
- Q4** Move the slider some more. What happens to the slope and the correlation coefficient in the new graph? Explain why the slope and the correlation coefficient behave differently in the other scatter plot.

## INVESTIGATE

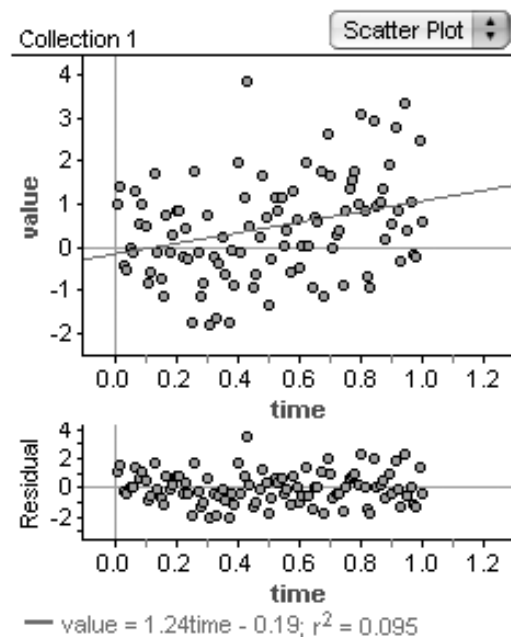
So far you've been looking at basically linear data. Now you'll explore how measurement error can obscure the true shapes of functions. You will also see why residual plots are useful.

8. Make two new attributes, *time* and *value*. Give *time* the formula  $\frac{\text{caseIndex}}{100}$ . Give *value* the formula  $\text{time}^2$ . (Type `time^2`. The ^ will not appear—it just lets you type the exponent.)
9. Make a graph of *value* as a function of *time*. The graph you see should be familiar: It's the  $y = x^2$  parabola between 0 and 1. It looks curved, obviously curved—for the moment. Now you will try to fit a line to it.
10. With the graph selected, choose **Least-Squares Line** from the **Graph** menu. A line appears.
11. Choose **Make Residual Plot** from the **Graph** menu. This shows how far each of the points in the first graph is, vertically, from the line. You may want to stretch the graph vertically to make more room for the residual plot.
- Q5 Describe what is happening in the function and in the residual plot between  $\text{time} = 0$  and  $\text{time} = 0.5$ . Explain why this happens.

Right now, everything is nice and smooth. To shake things up, you'll add randomness as you did with the line.

12. Double-click the *SD* slider near the axis and change its axis to go from 0 to 1.2. Set the slider to about 1.00.
13. Now edit the formula for *value*. Change it to  $\text{time}^2 + \text{randomNormal}(0, SD)$ .
14. The graph will change, and many points will be out of range. Choose **Scatter Plot** again from the pop-up menu to rescale the graph.

If everything has gone according to plan, you probably see that the graph seems to increase somewhat, but that *you can no longer tell that it's curved*. The randomness is so great that the curve is



overwhelmed. You can confirm this by looking at the residual plot. It looks flat and random. That is the sign that the model—the straight line—is a good model for your data.

Yet you know that the “real” data are curved. The question you will explore is: How much randomness can that curved model take before it no longer looks curved?

- 15.** Adjust the *SD* slider to near 0 and rescale the graph.
- Q6** What do you notice about the original graph? What about the residual plot?
- 16.** Gradually increase the slider, rescaling the graph as necessary.
- Q7** What is the smallest value of *SD* that makes the residual plot look flat and random?
- 17.** Now repeatedly choose **Rerandomize** from the **Collection** menu. (This makes Fathom reassign all the error values.)
- Q8** What changes in the graph when you rerandomize? Does the graph generally appear curved—or flat and random?
- 18.** Readjust the slider, repeatedly rerandomizing, to find the largest value for *SD* you can use and *still see the curvature in the residual plot most of the time*.
- Q9** What value did you find for *SD* in step 18?

## EXPLORE MORE

- 1.** Leave *SD* at the critical value you found in step 18. Add 10 to the formula for *time*, so it is  $10 + \left(\frac{\text{caseIndex}}{100}\right)$ . Sketch the new graph (including the residual plot) and explain why it looks so different.
- 2.** Change the formula for *time* so that the positions of the points are random within the interval  $[0, 1]$  instead of evenly spaced. What formula did you use? Does that make any difference in the amount of error you can put in and still detect the curvature?
- 3.** Change the number of cases (by adding and deleting them) without changing the range of *time* values. Does that make any difference in the amount of error you can put in and still detect the curvature?
- 4.** If you’re trying to detect an effect (for example, curvature), you have a better chance if your measurements have less error. This activity helped quantify the amount of error you can have and still detect the effect. Give some suggestions as to how you can improve your chance of detecting an effect if you’re looking for one that’s right on the edge of detectability.

# Fathom™

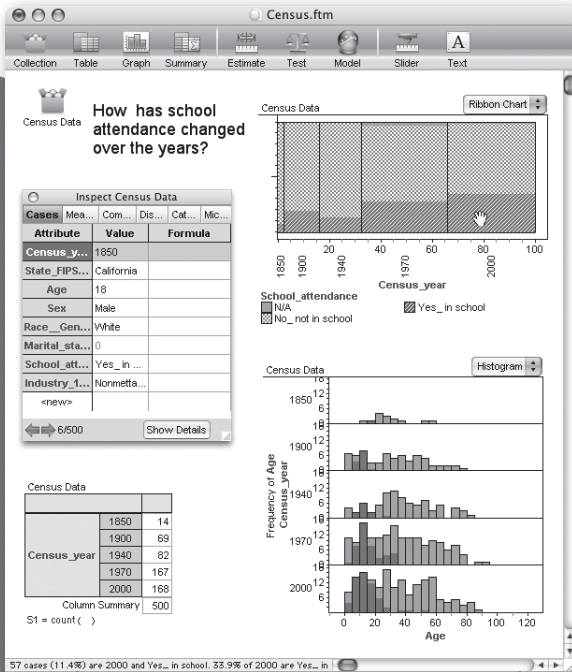
Dynamic Data™ Software

## Data are everywhere!

Forever change your students' view of data analysis, statistics, and mathematics! *Fathom* enables and motivates your students to gather, graph, analyze, and animate data of all kinds. *Fathom 2*'s new full-color design makes it easier than ever for your students to identify patterns and make sense of data.

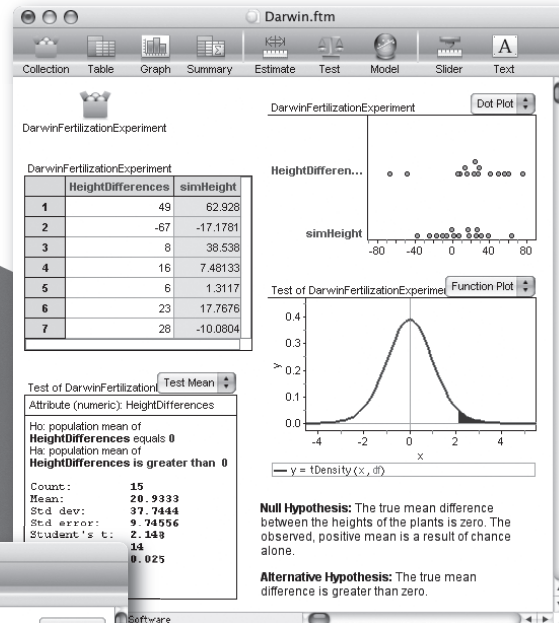
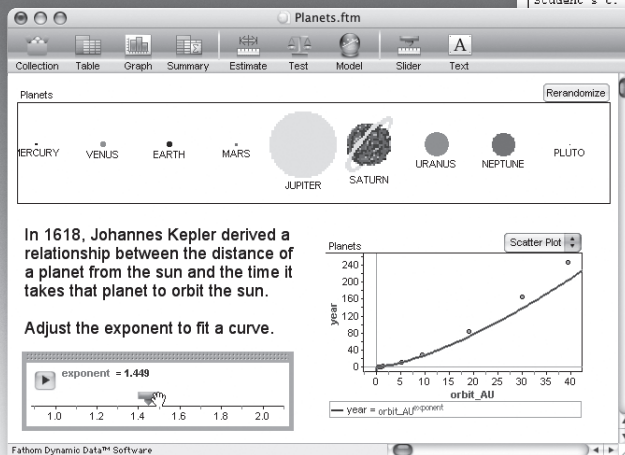
## Workshop Guide

The *Workshop Guide* includes everything you need to take or lead a one-day *Fathom* workshop. Learn how *Fathom* can enhance your students' learning as you explore the range of what *Fathom* can do. Plus, you'll walk away with tested classroom activities and ideas on how to use *Fathom* in your class today.



Learn the basics of exploring data with *Fathom*.

Help your students model problems from mathematics, science, and social science.



Enhance your students' data analysis with *Fathom*'s powerful statistical tools.

## Key Curriculum Press Key College Publishing

Founded in 1971, Key Curriculum Press is the leading publisher of innovative mathematics books and dynamic mathematics software for high school and middle school students. Key College Publishing, founded in 1999 in cooperation with Springer Science and Business Media as Key's higher education division, publishes innovative courseware in undergraduate and pre-service mathematics and statistics.

Fathom Dynamic Data™ and the Fathom logo are trademarks of KCP Technologies. All other trademarks are held by their respective owners.

ISBN 1-55953-683-7



9 781559 536837