Fathom™

Dynamic Statistics™ Software

Reference Manual

Key Curriculum Press
Key College Publishing

Fathom Dynamic Statistics™ Software February 2001

Authors: William Finzer, Tim Erickson, and Jill Binker

Lead Programmer: Kirk Swenson

Development Team: William Finzer, Kirk Swenson, Nick Jackiw, Jill Binker, Matt Litwin, Eugene Chen, Gavin Peters, Tony Thrall, Tim Erickson, Zach Teitler, Denise Howald, Caroline Wales, and Vadim Keylis

Statistical Consultant: Tony Thrall

Production Editors: Deborah Cogan, Kristin Ferraioli

Copy Editors: Judith Abrahms, Joan D. Saunders, and Erin Milnes

Production and Manufacturing Manager: Diana Jean Parks

Production Coordinator: Ann Rothenbuhler

Art and Design Coordinator: Caroline Ayres

Cover Designer: Jenny Somerville

Prepress and Printer: Data Reproductions

Executive Editor: William Finzer

Publisher: Steven Rasmussen

Portions of this material are based upon work supported by the National Science Foundation under award number III–9400091. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Key Curriculum Press
1150 65th Street
Emeryville, CA 94608

techsupport@keypress.com
http://www.keypress.com

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1 04 03 02 01 ISBN 1–55953–516–4

Acknowledgments

Many people over many years helped us in many ways.

On the original NSF advisory board, Gail Burrill focused us on improving the lot of practicing teachers; Chris Hancock taught us the Tabletop metaphor and how to be concrete; Cliff Konold started us on the fruitful path of resampling; Andee Rubin not only warned us off dice but kept us thinking about learning; Mike Shaughnessy showed us the relationship between offline and online simulation; and Frank Sheehan brought us back time and time again to the statistical and probabilistic heart of the matter.

Of the hundreds of high school teachers who gave us feedback, deserving special thanks are those members of the statistics study group who gave of their Wednesday evenings once a month over two years trying out activities and software: Marianne Nolte for her insights into students' learning; Elisabeth Klein for her appetite for new ideas; Bev Austin for her brave willingness to try new things; Lew Douglas for his ideas on teaching statistics; Carlos Cabana for both his attention to detail and his ability to see the big picture; Judy Quan for her generous hospitality at her school; Don De Pasquale for his reality-tempered enthusiasm; and Carolyn Kirkpatrick for the student work she brought for us all to look at.

Many teachers served as reviewers, consultants, and field testers. We mention Gretchen Davis, who did not fear to tell us the hard truths; Chris Olsen, who told us what he really wanted; Leona Mirza, who had faith in Fathom from the beginning; Howard Amerlan, who funneled both his own and his students' ideas to us; Paul Myers, who helped us see the power of Fathom; Susann Mathews, who helped us think about uses for the network; Dudley Brooks, who gave countless good suggestions; Chris True, who made the first Fathom poster; Sue Gordon, who encouraged us with stories of student understanding; Sharon Shrader, who took Fathom home from the Teaching Statistics with Technology Institute and even surveyed her students about it; Mark Vosskamp, who advocated a balance between discovery and structure; Jim Fensom, who holds the record for bug submissions; David Hoppe, who gave us insightful criticism; and Joyce Smart, who was available for on-the-phone debugging; Len Fellman who has stuck with us through thick and thin, and Robin Lock for invaluable critiques, insights, and suggestions.

Ron Avitzur provided not only an excellent formula engine and editor, but also valuable insights into how his engine should and could link to Fathom.

Many people at Key Curriculum Press contributed to Fathom beyond their normal call. Dan Bennett critiqued just about everything; James Browne found last-minute bugs; Stefanie Liebman tested and told us she enjoyed it; Jennifer Strada tested tours and instructions; and Candice Miller helped us with naming Fathom.

Steve Rasmussen believed in us, supported us, critiqued everything from ad copy to user interface, and did more than anyone else outside the team to make Fathom real.

Nick Jackiw not only contributed in major ways to Fathom's design, he continually brought us back to the fundamental vision of educational software we are pursuing.

No project this size comes to fruition without the help of many family members and friends. Among these are Brigid McCaw, Lauren Finzer, Devin Finzer, Ramona Swenson, Anne Erickson, and Meg Holmberg.

Contents

Fathom Terminology

Case

A case is the fundamental unit of data. For example, in a data set of students, each student is a case. In Fathom, you work with collections of cases. Unless you change their icons, cases appear as gold balls with captions below them.

Collection

Collections hold cases. Any data you have resides in the collection. Graphs and tables allow you to view and change the data, but deleting a graph or table does not delete the data. Delete a collection, though, and your data is gone.

Attribute

Cases have attributes. For example, data about a person might have attributes such as sex, height, and favorite vegetable. The attribute names appear as column headings in the case table. Without attributes, you have no place to enter data.

Value

An attribute has a value. Values for the sex and height and favorite vegetable attributes might be male, 65 inches, and broccoli.

Categorical (or Nominal) Attribute

The values are not numbers, but words (also known as *strings*). Sex, school, favorite vegetable, and marital status are examples.

Continuous (or Numeric or Measurement) Attribute

The values are numbers. Height, age, speed, and price are examples. You can do arithmetic with continuous values, but not with categorical values. So you can't calculate mean or standard deviation for favorite vegetable, but you can for height.

Measure

An attribute belongs to the cases in a collection; a measure belongs to the collection as a whole. Measures have names and, usually, formulas. Examples are a measure named meanHeight that computes the mean of the heights of the cases, and a measure named numFemale that counts the number of females in the collection.

Attribute Formula

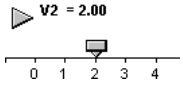
Formulas are mathematical expressions used to compute the values of attributes. For example, you might use a formula to compute the ratio of height to weight, or the total number of minutes a train has traveled, or to put together a full sentence that describes an individual case.

Derived Collection

When you sample from a collection, you get a sample collection. When you collect measures, you get a measures collection. These are both examples of derived collections. The important thing about a derived collection is that its data comes from somewhere else.

Filter

A filter is a mathematical expression that limits which cases in a collection are displayed in graphs or tables connected to that collection.



Slider

Sliders are named numeric values. They appear on the screen as something you can actually change by sliding.

Estimate

An estimate is an object that will calculate confidence intervals and regression lines.

Tests

Test objects calculate statistical results; for most test types you can either drag and drop in attributes or type in summary statistics.

How To's

In this section we've listed many common Fathom tasks and provided directions for each.

How to Do Basic Things

Install Fathom

Windows

- Insert the Fathom CD into your CD drive.
- Double-click the **My Computer** icon on your desktop.

- Double-click the CD icon. (It may be labeled either **D:** or **Fathom**).
- Double-click the **Setup** icon and follow the directions on the screen.

Macintosh

- Insert the Fathom CD into your CD drive.

- Double-click the CD icon.
- Double-click the **Installer** icon and follow the directions on the screen.

Install Fathom on a Network

There are two possible configurations for installing a Fathom multi-user edition: Install on each machine (the type of license you bought tells you how many computers you have permission to install Fathom on) or install on a network server (the type of license you bought tells you how many computers you have permission to run Fathom on). Installing only on a server, and running Fathom from it, is less work to install and upgrade, though it will slow performance (by how much will vary depending on the speed of your computers).

Installation is similar for all (insert the CD, double-click the installer, follow the directions on the screen; at the end, you have a choice of restarting or not restarting the computer).

To Install on Each Computer

Use the Fathom CD to install Fathom onto each computer for which you are licensed.

- Insert the CD into a computer's drive, open it, double-click the installer, follow the directions on the screen, and restart when through.

Or…

Use the CD in the server's CD drive to install onto each machine for which you are licensed (faster than the previous method, because you can install on more than one machine simultaneously).

- Put the CD in the server's CD drive and give sharing privileges for each machine. From each computer, access the server's CD drive, double-click the installer, follow the directions on the screen, and restart when through.

To Install on the Server

- Insert the CD into the server's drive, open it, double-click the installer, and follow the directions on the screen. At the end, you will be asked whether you want to restart the computer (your choice).
- Make sure the Fathom folder is shared to each computer for which you are licensed, by whatever method you usually use to share networked software.
- You may want to put shortcuts or aliases to Fathom on each networked computer.

Note: All installed sample documents are locked so that students cannot easily save over them. When students try to save, they will be prompted with a **Save As...** dialog box. They must rename or move the file in order to complete the save. You may want to restrict students' access to the Data in Depth Solutions folder.

Start Fathom

After you have installed Fathom from the CD, Fathom's icon should appear on your desktop. Double-click it and Fathom will start.

If it is not on your desktop, you can probably find the **Fathom** folder in the **Program Files** folder of your startup disk (Windows platform) or directly on your hard drive (Macintosh platform).

Use Help

When you want to find out something you don't know, you can use Fathom's help system. Choose **Fathom Help** from the **Help** menu—the one on the far right. Fathom launches your Web browser.

Get Technical Support

When things go wrong and you need support, here are some things you can do.

- Check the Fathom Web site at www.keypress.com/fathom. You will find answers to frequently asked questions and the latest patches to the program. The Key Curriculum Press Web site at www.keypress.com has an online tech support form.
- Call our technical support line at Key Curriculum Press, 510-595-7000. We can be most helpful to you if you are sitting in front of a computer with Fathom running.

Save and Back up Your Work

No computer system is infallible. Your time is valuable. If you have done a lot of work to enter data, or you've just done a brilliant analysis, save your work. Save early. Save often. And then make an extra copy just in case.

How to Undo and Redo

Fathom supports (almost) unlimited undo and redo. Choose **Undo** from the **Edit** menu or press **Ctrl-Z** (Win) **⌘-Z** (Mac) to undo the previous action. Choose **Redo** to redo the most recently undone action—**Ctrl-R** (Win) **⌘-R** (Mac).

Collections, Cases, and Attributes

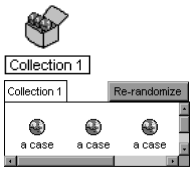
Attributes are also known as variables.

These are the backbone of Fathom. A collection is a data set. Collections contain cases. Cases have attributes. Attributes have values. The cases in a collection all have the same attributes—but these attributes can have different values for different cases.

For example, if you have a collection of people, each person is a case. Each case has attributes: sex, height, age, favorite color, income, and so forth. These attributes take on different values for different cases, because the people have different ages, heights, and so forth.

When creating your own collections, give serious thought to the question, "What is a *case* in this context?" or you may run into problems later. For example, if you are looking at sex and height, don't have men's height in one attribute and women's in another. Instead, make each person a case with sex and height attributes. All the data in a given case should have some real meaning. In this example, case number 3 does not represent any intelligible "thing." Given Fathom's case-centered structure, you won't be able to look at the data the way you want to unless you've correctly identified what the cases are.

In a case table, each row is a different case; each column is an attribute. In a scatter plot, each point represents a case; each axis signifies an attribute.



In Fathom, by default, a case looks like a gold ball (though you can change that). A collection icon looks like a box with gold balls in it. When a collection is open, you can see the gold balls individually.

Get Data into Fathom

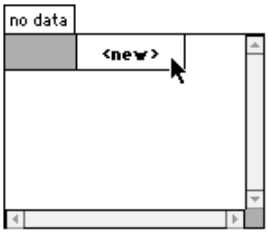
How do you get data to work with? There are five basic strategies:

- Enter the data from scratch.
- Open a Fathom-format file.
- Import data from a text file.
- Paste data from the clipboard.
- Import data from the Internet.

♦ However you get your data, as soon as it's in a collection, *save!*

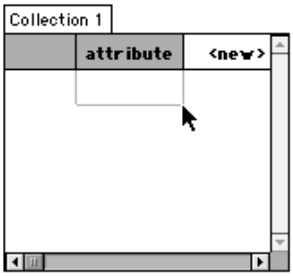
Enter the Data from Scratch

Briefly, make a case table, create new attributes, and type in the data.



- Choose **Case Table** from the **Insert** menu. This creates a spreadsheet-like (but empty) case table.
- Click <new>, type an attribute name, and press **Enter**. (If necessary, repeat this step to add more attributes.)

As soon as you've named your first attribute, Fathom creates an empty collection to hold your data (a little, empty box).



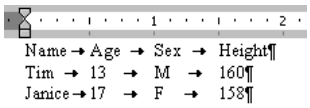
- Click in the blank cell under the attribute name and begin typing values. (Press **Tab** to move from cell to cell.)

(As soon as you begin typing values, the collection icon fills with gold balls.)

Open a Fathom-Format File

- Choose **Open** from the **File** menu.
- Choose the file you want and click **Open**. That's it! Fathom opens the document. Many sample data sets came with Fathom. More can be found on the Fathom Web site at www.keypress.com/fathom.

Import Data from a Text File



Fathom is pretty good about recognizing different text file layouts. The most reliable layout has the names of the attributes in the first line and the values under them, one line (paragraph) per case, as shown here. Fathom likes it if you put tabs between your values, but it will also understand if you just make them all line up using spaces.

Fathom imports text files, so you should first go into the file's creator program, and save the data as a tab-delimited text file.

- Choose **Import From File**... from the **File** menu.
- Select the file you want to import and click **Open**.

Or…

- Drag the file into a Fathom document.

If your data imports, but uses the first case as the attribute names, open the text file in a word processing program and add a line at the top with the attribute names. Then import this edited file, and Fathom will read the first line as the attribute names.

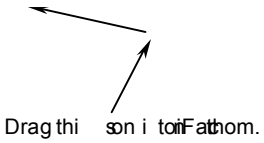
Copy and Paste Data

Copying cells from an Excel spread-sheet works fine. Just be sure attribute names are in the first row.

- Copy the data (tab-delimited, attribute names in the first line).
- Switch to Fathom.
- Activate a collection (click on it; it will get a frame).
- Choose **Paste Text** from the **Edit** menu.

Import Data from the Internet

- Find the data you want on the Internet. Be sure you're looking at the page with the data (not a link to it).
- Drag the URL icon from your browser into an open Fathom document. Fathom imports the data and makes a new collection.



Or…

- Choose **Import From Url...** from the **File** menu. A dialog box appears.
- Type or paste the URL of the data you want into the box and click **OK**. Fathom imports the data and makes a new collection.

After importing, look at the data in a case table to see whether it came in properly. Often you have to do a little fixing before the collection is right.

If Fathom did not put important information into the comments (see next section), you may want to copy and paste comments separately.

Not all Web pages are in a format Fathom can decipher. One site rich in good, comprehensible data sets is the Data and Story Library (DASL) site, currently at

http://lib.stat.cmu.edu/DASL/DataArchive.html. You can find links to other "Fathomable" Web pages by following the "Links to Data" link on the Fathom home page at http://www.keypress.com/fathom.

Add Comments to a Collection

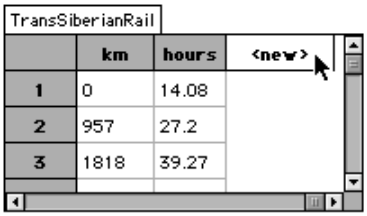
- Double-click the collection (the box of balls) to bring up its inspector.
- Click the **Comments** tab to bring up the **Comments** pane.
- Type the comments into the pane. Fathom will store them when you save the file.

If the comments you want to bring into Fathom exist somewhere else (such as on a Web page), you can copy and paste them into the comments pane (if they weren't imported there already).

Add Attributes

… at the Right Edge of a Case Table

- Make sure you can see the new attribute prompt in the case table. (If not, scroll to the right.)
- Click it once.
- Type the attribute's name and press **Enter**.



- If you need more than one attribute, click <new> again, type the attribute name, and press **Enter**.

Repeat until you have all the attributes you need. (If necessary, enlarge the table or scroll to the right until you see <new>.)

… in the Middle of a Case Table

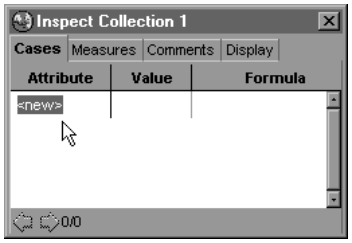
- Select an attribute to the right of where you want the new one to appear.
- Choose **New Attribute** from the **Data** menu. A dialog box appears.

- Type the name of the attribute and click **OK**.

… to a Collection Directly

You can also open an inspector by choosing ***Inspect Collection*** *from the* ***Edit*** *menu, or from the right-click menus for collections, case tables, or graphs. (On Mac,* ***Control-****click to invoke the menu.)*

- Double-click the collection to open its inspector.
- Click once on the word <new>.
- Type the name of the first attribute and press **Enter**.



If you need more than one attribute, click <new> again, type the attribute name, and press **Enter**.

Delete Attributes

- Select an attribute name in a case table.
- Choose **Delete Attribute** from the **Edit** menu.

Add Cases

… in a Case Table

If you are typing data into the new cases, use this method; if your collection is generated by formula, use the menu method below.

- Find the empty row at the bottom of the table. Click in a cell.
- Type in the value for that attribute.
- Press **Tab** to move to the next cell or **Enter** to move down to the next case (row).

This creates a new case. Repeat until you have entered the data you want.

In Fathom, you can always add data, even if you have a bunch of graphs already made. When you add new data, the data will also appear on the graphs.

… Using a Menu

- Activate a graph, case table, or collection.

New Cases *also appears in the right-click (Win)* ***Control-****click (Mac) menus for collections, graphs, and case tables.*

- Choose **New Cases...** from the **Data** menu.
- Type in the number of cases you want and click **OK**.

Delete Cases

- Select the cases you want to delete—in a collection, a case table, or a graph.
- Choose **Delete Cases** from the **Edit** menu.

Note: Deleting cases may not be what you want; it *eliminates* the data. You can also **Cut** the cases (so you can **Paste** them into a different collection), or, if you just want to look at a subset of your data, *filter* out unwanted cases. See "How to Filter Data" on page 45.

Rename an Attribute

- Double-click the attribute name, type a new name, and press **Enter**.

Or…

- Select an attribute (click on it once) and choose **Rename Attribute** from the **Data** menu.

Rename a Collection

- Double-click the collection name, type the new name into the dialog box, and click the **OK** button.

Or…

- Select a collection, graph, or case table (click on it once) and choose **Rename Collection** from the **Data** menu.

Define a Measure

When you're looking at data, there are usually some numbers you would like to compute that apply to the collection as a whole—the sum of the salaries, the median of the weights, the ratio of the second year's yield to the first year's, or the correlation coefficient. These numbers are all statistics, or *measures*. You define measures in the collection's inspector by giving the new measure a name and (usually) a formula.

You can refer to measures by name in all other formulas that apply to that collection.

- Double-click the collection to bring up its inspector.
- Go to the **Measures** pane.
- Click once on <new>.
- Type in the new measure's name and press **Enter**.

Usually you will want to enter a formula for the new measure. See "How to Work with Formulas" starting on page 36.

Prevent Data from Being Changed

- Select the collection.
- Choose **Lock Collection** from the **Data** menu.

Now you cannot change the data: You can't edit it, drag it, or add new attributes. Choose **Unlock Collection** to remove this restriction.

⊗ Is **Lock Collection** unavailable in the file menu? You need to select the collection, not another object (for example, not the case table).

Export Data to Other Programs

Briefly, export it as a text file and then import it to the other program.

This strategy works fine with spreadsheets such as Excel, statistical packages, or calculator-link software.

- Select the collection you want to export.
- Choose **Export File** from the **File** menu.
- Give the file a name and click **OK**.

Fathom will create a tab-delimited text file with the names of the attributes in the first line. Comments *won't* be exported.

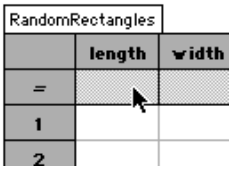
Enter a Formula for an Attribute

For many attributes, their values are computed by formula from other attributes or, in simulations, values generated at random.

Creating a formula deletes any values in the attribute. First create a new attribute unless you want to replace your data with the calculated values.

… in a Case Table

- Select the case table.
- Choose **Show Formulas** from the **Display** menu. A shaded row of formula cells appears.
- Double-click the formula cell for the attribute you want to define by formula.
- Enter the formula in the resulting formula editor and click **OK**.



This method is good if you are creating many formulas or you want your formulas always visible.

Or…

If you don't want the formulas taking up screen space and don't need to see them all the time, you don't have to show the formula row.

- Select the attribute by clicking on its name in a case table.

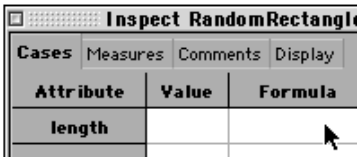
*The shortcut for editing a formula is **Ctrl-E** (Win), **⌘-E** (Mac).*

- Choose **Edit Formula** from the **Edit** menu.
- Enter the formula in the formula editor and click **OK**. (See "How to Work with Formulas" beginning on page 36.)

⊗ ***My formula wiped out my data!*** Immediately **Undo Edit Formula**. You need to add a new attribute to hold the computed values.

… in the Inspector

- Open the collection's inspector. (See page 117.)
- In the **Cases** pane of the inspector, double-click in an attribute's formula cell.
- Type the desired formula and click **OK**.



Delete a Formula for an Attribute

*If you right-click (Win) **Control**-click (Mac) on an attribute in a case table or inspector, you can choose **Clear Formula** from the shortcut menu.*

If you delete a formula, Fathom leaves the computed values in place but treats them as noncomputed (as though you had typed them in). Randomly generated values will no longer rerandomize. You can change the values by dragging.

- Select the attribute whose formula you want to erase.
- Choose **Clear Formula** from the **Edit** menu.

How to Arrange Things in a Document

Move and Resize Objects

- Grab the top bar of an object and drag to move it.
- Grab the sides, bottom, or corners and drag to resize it.

Iconify an Object and Get It Back Again

- When you make an object as small as you can, it will turn into an icon.
- You can drag icons by their top bars or by their centers.
- To restore an iconified object, drag a corner or edge and expand it.

Why Iconify Objects?

- Iconified objects still work—they just take up less space and are easy to move out of the way.
- If updating an object is making a procedure too slow, iconifying it may speed it up. (For example, if Fathom is very slow when you drag data, iconifying the collection will speed things up.)

View an Object in a Separate Window

*The shortcut for **View in Window** is **Alt-V** (Win) **Option-⌘-V** (Mac).*

Sometimes you want to see one object without having to shrink everything else. You can resize the window to fill the screen.

- Select the object by clicking on it.
- Choose **View in Window** from the **Display** menu.

The object now exists in its own window as well as in the document.

- Close it when you're done with it by clicking its close box.

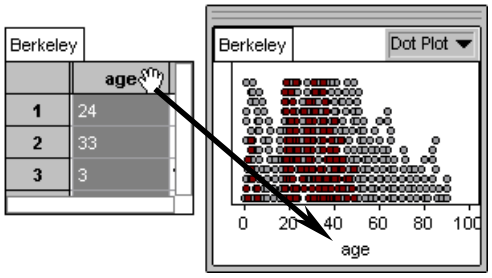
How to Make Graphs

Graphs are visual representations of your data. How do you get the graph you want?

- Make a new graph by choosing **Graph** from the **Insert** menu or dragging one off the shelf. This produces an empty graph.

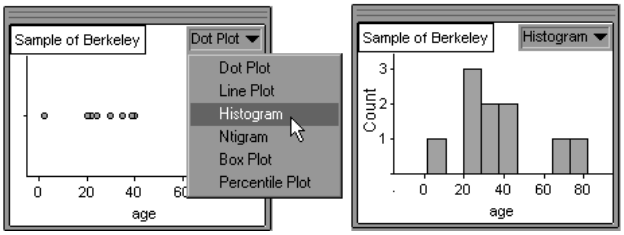
Anywhere you can see an attribute's name, you can drag it from there to another object.

- Drag the name of an attribute you want to study to one of the axes of the graph. This produces a graph of the data, though it may not be the kind you want. Drag another attribute to the other axis if you wish.



How to drag an attribute.

- Choose the kind of graph you want from the popup menu on the graph itself.



- If you want to replace one attribute with another, just drag the new attribute on top of the old one.
- If you want to switch the axes of a graph, just drag an attribute from one axis to the other.

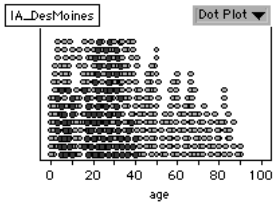
Make, Manipulate, and Understand Each Kind of Graph

What graphs are available to you depends on what kinds of attributes you have on the axes. Fathom's attributes are either *numeric* (values are numbers such as 3.14 or –8) or *categorical* (values are usually text, such as "male" or "orange"). Switch among the available graph types by using the popup menu in the graph. Here is a list of the kinds of graphs you can choose from.

Graphs with One Continuous Attribute

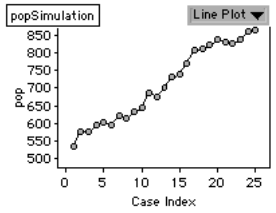
Dot Plot

There is one point for each case. (When you have more data points than can fit on the plot, the overflow is designated with brown dots.) You can unstack the dots by unchecking **Stack Dots** in the **Graph** menu. This puts dots representing the same value directly on top of each other.



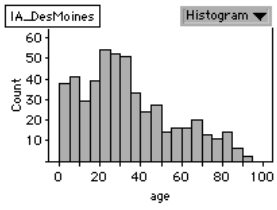
Line Plot

There is one point for each case. They're plotted in order and connected by lines. This kind of plot is good for time series; you typically put the attribute on the vertical axis.



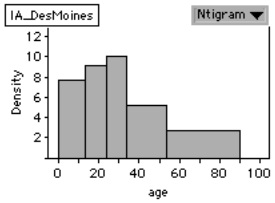
Histogram

Cases are grouped into bins of equal width. A rectangle represents each bin, showing how many cases are in it. Control the width of the bins by dragging a bin edge. You can change the vertical axis to **Relative Frequency** or **Density** scales via the **Scale** submenu in the **Graph** menu.



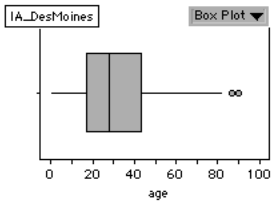
Ntigram

Cases are grouped into bins of equal population. The height of the corresponding rectangle is proportional to the *density* of cases in the bin. You can control the population in each bin by dragging on the bin edge.



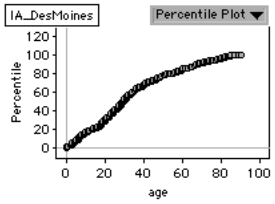
Box Plot

A box extends from the 25th to the 75th percentile and is cut by a line at the median. "Whiskers" extend out from each end of the box, either 1.5 times the interquartile range or to the end of the data, whichever comes first. Points beyond the whiskers are plotted individually as outliers.



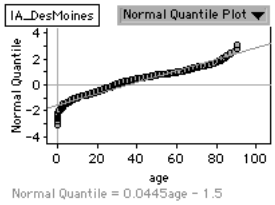
Percentile Plot

Cases are sorted and then plotted in order; the other axis shows the percentile of the points. In the illustration, you can see that there are many people with small incomes and a few with large incomes.



Normal Quantile Plot

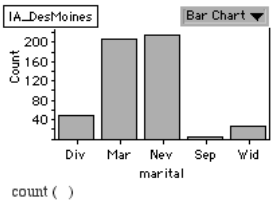
It plots the z-score associated with the percentile of each case if the data were normally distributed. Therefore, if the data are normal, the plot should show a straight line. The line in the plot corresponds to quantiles of a normal distribution whose mean and standard deviation are the same as the mean and standard deviation of the data. The line crosses the *x*-axis at the mean and has a slope equal to the inverse of the standard deviation.



Graphs with One Categorical Attribute

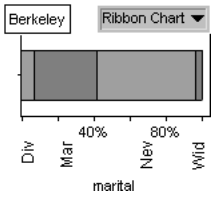
Bar Chart

There is one bar for each category, showing the population in that category. By default, the bar chart shows the count, but you can change this to have the bars represent other values by editing its formula. (For example, by changing the formula to median(age), you would make each bar's height represent the median age within each marital status category.)



Ribbon Chart

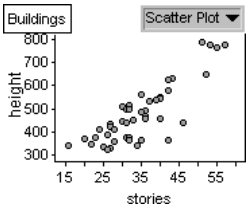
There is a single bar (the ribbon) broken into pieces. Each piece corresponds to one category; its size is proportional to the population. The axis shows percentages.



Graphs with Two Continuous Attributes

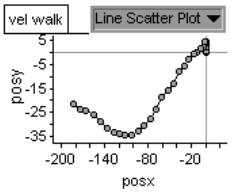
Scatter Plot

There is one point for each case at the position indicated by the values on each axis. The illustration shows the association between the height of a building and the number of stories it has.



Line Scatter Plot

This is just like a scatter plot, except that the points are connected according to the order in the collection. **Note:** This order is the "native" order of the data, *not* the sort order.



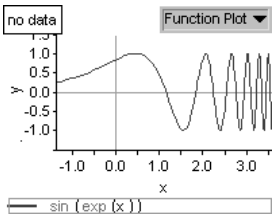
Graphs Without Any Data

The Empty Plot

This is just an empty plot—the kind you get when you first make a graph. Either make it a function plot, or drag one or more attributes to it.

Function Plot

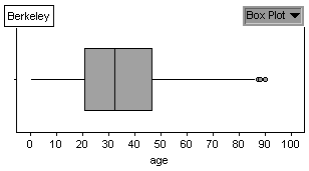
You can use this plot simply to graph a function.



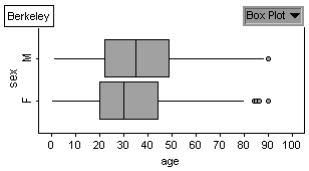
- Make an empty graph.
- Choose **Function Plot** from the graph's popup menu.
- Choose **Plot Function** from the **Graph** menu (along the top of the screen).
- Enter the function in the ensuing formula editor and click **OK**.

Split a Graph

If you have a graph that shows information for an entire population, you can split the graph to get information about its subsets. For example, we changed



to



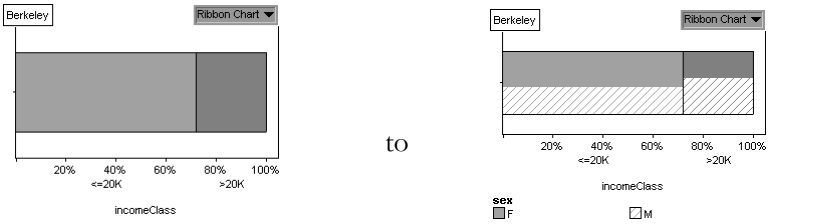
Splitting on an Axis

To split the graph along an axis (as above), drag a categorical attribute to the empty axis. In the above example, we dragged sex to the vertical axis.

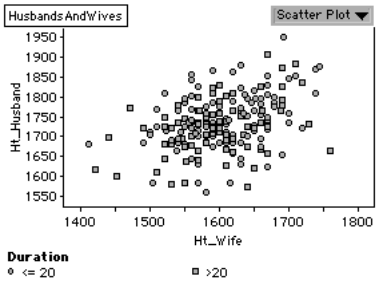
This strategy works on graphs with a single numeric attribute: dot plots, histograms, ntigrams, box plots, percentile plots, and normal quantile plots.

Splitting with a Legend

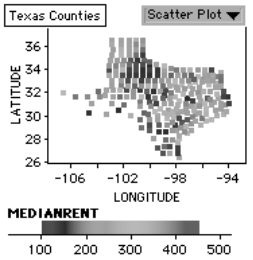
If you're working with a graph that shows a categorical attribute—a bar chart or a ribbon chart—you can split it by dragging the "splitting" attribute to the *middle* of the graph. This splits the individual bars into differently shaded rectangles and creates a *legend* showing what each shading means. For example, dragging sex onto the middle of this graph about income changes



You can also create a legend by dragging a categorical attribute onto the middle of any kind of scatter plot, dot plot, line plot, percentile plot or normal quantile plot. Fathom uses different symbols for the points; these symbols are identified in the legend. You can then select one of the categories by clicking it in the legend.



Finally, you can also drop a numeric attribute into the middle of a scatter plot. The data points become colored, and a continuous legend (a color spectrum) appears below the graph, as at right. You can select a range of values by dragging along the legend. Selected points appear larger in the plot.



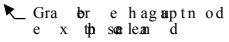
Change Axis Scales

When you first make a graph, Fathom scales the axes so that all the data are included. But you often want to change the scale of the axes to study features of the graph more closely.

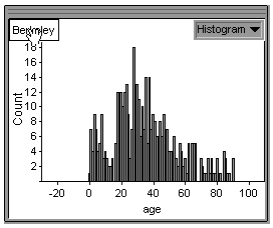
Direct Axis Manipulation

The basic way to work with axes in Fathom is to drag on the numbers of the axis. Dragging in the middle translates the axis, moving the range without changing the scale. Dragging closer to the ends expands or contracts the range, keeping the opposite end of the axis constant. You can think of this action as pulling new numbers onto the axis (zooming out) or pulling them off (zooming in). This action, while hard to describe, is easy to learn—and completely undoable.

Below, we grab the "50" and push it off the top of the graph:



becomes

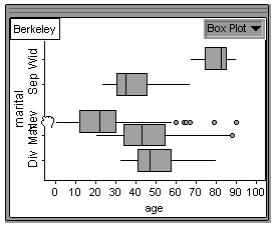


Note: You can always return to the original "all-points" state by choosing **Rescale Graph Axes** in the **Graph** menu or by rechoosing the plot type from the popup menu in the graph.

Note: The number line in a slider is an axis just like an axis on a graph; you can manipulate it the same way.

Manipulating Categorical Axes

If you want to change the order of categories on a categorical axis (they appear alphabetically), you can drag the labels on the axis. In the illustration, the user is dragging the Nev category (showing the ages of never-married people) downward.



This temporarily reorders the categories. The categories go back to the default if you change the plot type or use the **Rescale Axes** command. To permanently force categories into an order you want, you need to rename them (by editing them or creating a switch statement to recode them (for example, **a Nev**, **b Mar**, etc.). (See page 41 for more on switch statements.)

Axis and Graph Controls

A second way to control axes is to double-click an axis. This brings up a control text object in your document describing the axes (see right). This information includes the minimum and maximum values the axis displays. When you manipulate the axis, this text changes; when you edit the text, the axis changes.

Information about this graph:
Histogram: Bin width: **5.0000** starting at: **0**
The age axis is horizontal from **0** to **110.00**
The Count axis is vertical from 0 to **65.000**

For histogram axes, this object also lets you control the binning precisely: You tell Fathom the bin size and any value where a new bin begins.

Zooming

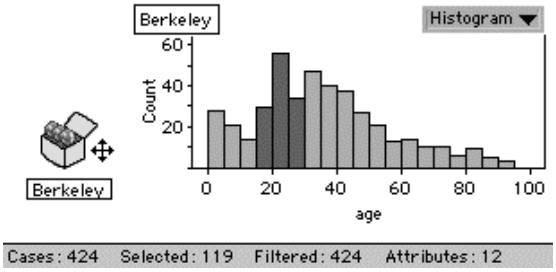
You can also zoom directly. When you hold down the **Ctrl** key (Win) **Option** key (Mac), the cursor changes to a magnifying glass. If you click on an axis, it expands around that point. If you drag a rectangle in a scatter plot (for example, around the points you want to focus on), that rectangle expands. Holding down **Shift** as well reverses the process, making you zoom out.

Select Data in a Graph

Whatever data you select in a graph appear selected in all representations of that data—other graphs, case tables, and the collection itself. Here are the principles of selection:

- Click on a case to select it. If you click on a bar, all of the cases in the bar are selected. This is also true of whiskers in box plots.
- Hold down **Shift** and click to select additional points or bins.
- Drag a rectangle to select all the points within it. If you drag a rectangle in a plot with boxes (for example, a histogram), you select every box the rectangle touches.

♦ You can find out how many cases you have selected by moving the cursor over the collection or a case table and looking in the status bar in the lower left of Fathom's window.



Drag Data in a Graph

If you drag data in a graph, you change the data. (If you don't want this to happen, you can lock the data. Select the collection and choose **Lock Collection** from the **Data** menu.)

Although you can use this feature to adjust your data, its real value is in seeing the effect that changing data has on other representations. For example, if you have a scatter plot with a least-squares line, you can see the effect of an outlier dynamically by dragging a point around.

You can always restore the data you have changed through **Undo**, or you can choose **Revert Collection** from the **File** menu, which restores your data to its last saved state. (**Revert Collection** is only available if the data have been saved and the collection is selected.)

Remove Attributes from a Graph

To remove an attribute from a graph, activate the graph by clicking on it and choosing one of the following options from the **Graph** menu:

*These options also appear in the context menu for the graph: right-**click** (Win) **Control-click** (Mac).*

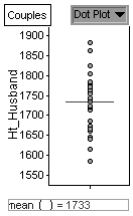
- **Remove X Attribute** removes the attribute from the horizontal axis.
- **Remove Y Attribute** removes the attribute from the vertical axis.
- **Remove Legend Attribute** removes the attribute that has been dropped into the middle of the graph. This is the one that controls different shapes of points, for example, or the splits within the bars of bar or ribbon charts.

Plot Values

You can plot a value on a graph with a continuous axis.

***Plot Value** is also available on the context menu for the graph: right-click (Win) **Control-**click (Mac).*

- Click on the graph to activate it.
- Choose **Plot Value** from the **Graph** menu.
- The formula editor appears. Enter a formula (it can just be a constant) for what you want to plot.
- Click **Apply** or **OK**. The formula result appears as a colored line and as a color-coded value at the bottom of the plot. You can plot as many values as you want.



Refer to "Compute and Plot Statistics in a Graph," page 53.

If the plot is split by some categorical attribute, Fathom will plot a separate line for each category. The calculated quantity at the bottom of the graph is for the entire collection.

Note: If you leave the argument of a function blank (as in the illustration above), Fathom will use the attribute on the axis (in this case, Ht_Husband) as the function's argument.

Plot Lines

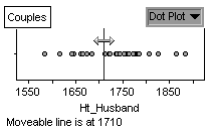
You can plot three different kinds of lines on a scatter plot. Here's how:

*These three options also appear in the graph's context menu—right-click (Win) **Control**-click (Mac).*

- Click on the scatter plot to activate it.
- Choose **Least-Squares Line**, **Median-Median Line**, or **Movable Line** from the **Graph** menu. The corresponding line appears on the graph, and its equation appears at the bottom.
- You can force a movable or least-squares line to go through the origin by choosing **Lock Intercept at Zero** from the **Graph** or context menus.
- You can remove a line by unchecking it on the menu.

Note: The least-squares and median-median lines update dynamically as the data change. The movable line remains where you put it.

Use a Movable Line (One Dimension)

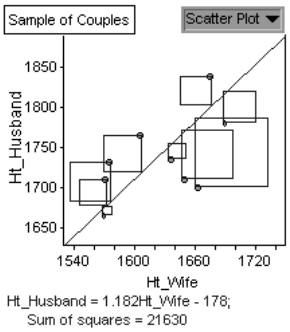


- Click on the graph to activate it.
- Choose **Movable Line** from the **Graph** menu. A line appears, and its value appears at the bottom of the graph.
- Grab the line and drag it.

Use a Movable Line (Two Dimensions)

- To change the movable line, drag its ends to rotate it, or its middle to slide it up and down. Its formula changes dynamically to match the line.

Show Squares



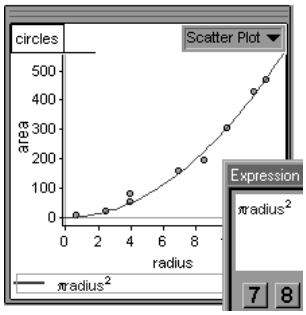
When you choose **Show Squares** from the **Graph** menu, Fathom constructs vertical segments from every point to the line and then makes those segments into sides of squares. Fathom also displays the sum of the areas of those squares.

Show Squares also works with the one-dimensional movable line.

Plot Functions

Fathom will let you plot arbitrary functions on a scatter plot and on plots of distributions. For example, suppose you had measured circles' radii and diameters. Here is how you would plot the πr^2 curve:

- Click on the graph to activate it.
- Choose **Plot Function** from the **Graph** menu.
- Enter the right-hand side of the equation that describes the function. Do not enter the y = part! In this example, you would type pi * radius^2. (Note how ^ made an exponent, and how pi was converted to π.)
- Click **OK** or **Apply** to make the curve show, as in the illustration.



Note: You can use any symbol for the x attribute when you specify a function to plot. Fathom just assumes it's the independent variable. Using the actual attribute name—as we did (radius)—can be useful, however, because then you can copy and paste the formula (to an attribute, say, to make a predicted value) and it will work without modification.

Plot Distributions

Note: This is an advanced topic.

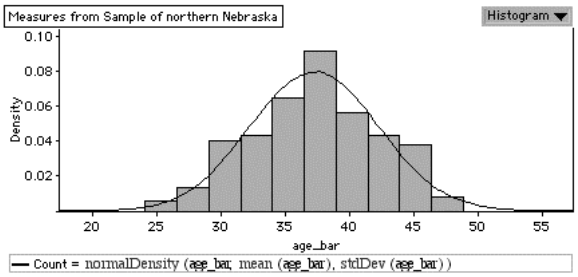
Suppose you had a sampling distribution and you wanted to plot the normal distribution with the same mean and standard deviation. Here's how:

- Make a histogram of the distribution in question.
- Be sure the graph is active.

***Plot Function** is in the graph's context menu: right-click (Win) **Control**-click (Mac).*

- Choose **Use Density Scale** from the **Graph** menu.
- Choose **Plot Function** from the **Graph** menu.
- Enter the formula for the distribution. Your completed graph—with the formula for a normal distribution—will look something like this:

This graph shows the mean ages for 100 samples of 10.



Edit Functions

To edit a function you have entered with the formula editor, double-click it where it appears at the bottom of the graph. The formula editor appears; edit it there.

You can also choose **Edit Formula** from the context menu you get when you right-click (Win) **Control**-click (Mac) on the formula.

Copy Functions from One Graph to Another

You can copy a function from one graph and paste it onto another.

- Invoke the function's context menu—right-click (Win) **Control**-click (Mac) on the function you want to copy.
- Choose **Copy Formula** from the popup menu.

- Invoke the context menu on the receiving graph, and choose **Paste Formula**.

Remove Functions from Graphs

There are two ways to remove a function from a graph:

- Edit the formula and make it blank.

Or…

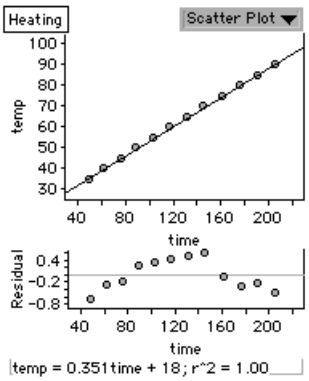
- Right-click (Win) **Control**-click (Mac) on the formula and choose **Clear Formula** from the popup context menu that appears.

Make a Residual Plot

When you have a function plotted on a scatter plot—either one you have specified yourself or a built-in line such as the least-squares line—you can create a *residual plot*. Residual plots help you improve and evaluate your mathematical model. In a residual plot, each point's vertical coordinate is set to its vertical distance from the function on the original graph. So points a little above the curve in the original will be slightly positive. Below the curve, they will be negative. We'll use our circles example:

- Create a scatter plot with a function plotted in it.
- If there is only one function plotted, be sure the graph is active. If there are more than one, click the equation for the function for which you want the residuals computed (so it is highlighted with a gray rectangle).
- Choose **Make Residual Plot** from the **Graph** menu. The residual plot appears.
- To make the plot bigger, stretch the entire graph. The residual plot is always about half the height of the main plot.

Make Residual Plot *also appears in the graph's context menu: right-click (Win)* ***Control****-click (Mac).*



Note: The vertical scale on the residual plot may be very different from the scale in the original. Residual plots can help you see features that are invisible on the original graph.

Filter a Graph

To restrict the cases that appear on a graph (or other object) without deleting them, you can write a filter. Suppose we're looking at the differences in income between men and women, but we realize that we have included all of the collection's children in our analysis. So we decide to look only at people over 20 years old:

- Create a graph of what you want to show, but with all the cases (as in the illustration on the next page on the left).
- Be sure the graph is active.

Add Filter *also appears on the graph's context menu; its shortcut is* ***Ctrl-F*** *(Win)* ***⌘-F*** *(Mac).*

- Choose **Add Filter** from the **Data** menu. The formula editor appears.
- Enter your filter. This should be a Boolean expression; that is, it is either true or false for each case in the collection. In this example, we have used age > 20.
- When you click **OK** or **Apply**, the filter goes into effect, and the graph will show only those cases for which the filter is true, as shown on the next page:

You can remove a filter in several different ways:

- Choose **Remove Filter** from the **Data** menu.
- Activate the graph and press **Ctrl-F** (Win) **⌘-F** (Mac). This toggles the filter on and off without changing its formula.
- Delete it in the formula editor or clear the formula.
- Right-click (Win) **Control**-click (Mac) on the formula and choose **Clear Formula** from the popup menu that appears.

Force a Numeric Attribute to Be Treated Categorically

Sometimes an attribute's values are numeric, but you wish to treat each distinct numeric value as a category.

- Hold down the **Shift** key as you drop the attribute onto the graph.

Force a Categorical Attribute to Be Treated as Numeric

Fathom decides whether an attribute is continuous based on the presence of any non-numeric values for the attribute. But sometimes an attribute has both numbers and other characters, and you want to ignore the string values and treat the attribute as numeric.

- Hold down the **Ctrl** key (Win) **Option** key (Mac) as you drop the attribute onto the graph.

(You could, instead, make a new attribute using the StringToNumber function and plot that attribute.)

How to Work with Case Tables

The case table shows your data in tabular format, with attribute names at the tops of the columns. In a case table, you can see multiple cases at once; add or change data; drag attributes to graphs, summary tables, and tests; or sort your data.

Make a New Case Table Showing Data in a Collection

*The shortcut for inserting a case table is **Ctrl-T** (Win) **⌘-T** (Mac).*

- Select the collection you want to view in case table form by clicking on it.
- Choose **Case Table** from the **Insert** menu.

Or, if you want to control where the case table is placed:

- Drag the case table tool from the shelf to where you want the case table to appear. (A dotted box shows you where the case table will appear.)

⊗ ***Did your case table come up empty?*** If the table is named no data, you forgot to select the collection before making the case table. Drag the *name* of the collection from the collection into the case table, as shown at right. If the

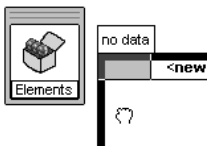


table has the name of the collection but no data, there were probably no data in the collection.

You can have more than one case table for the same collection of data. This can be useful when you want to filter (see page 45) or sort (see page 34) your data in more than one way at the same time.

Make an Empty Case Table

An empty case table gives you a place to start defining your collection of data. You can use it to enter attributes and cases.

*The shortcut for inserting a case table is **Ctrl-T** (Win) **⌘-T** (Mac).*

- Choose **Case Table** from the **Insert** menu.

Or…

- If you want to control where the case table appears, drag the case table icon from the shelf to the area where you want the table to be.

Add Attributes in a Case Table

- Click <new>.
- Type the attribute name and press **Enter**.

⊗ ***Did you get an error message when you defined an attribute?***

Fathom will accept only those attribute names that begin with a letter and contain only letters, digits, or underscores (_). But don't worry; if you type an invalid name, Fathom gives you an error message explaining the rules, then removes the invalid characters, leaving the valid characters in place. For example, if you try to name an attribute 4-hour wait (beginning with a non-letter and containing a hyphen and a space), Fathom will name your attribute hourwait. Simply click **OK** in the error message window to continue working with your data. If you don't want to keep the name Fathom gave you, double-click the name to change it.

Tips for Naming Attributes

- If you want to have numbers in your attribute name, put the number last or in the middle, rather than first. Instead of 1stfavorite, use favorite1, favorite2, and so on.
- If you want multiword attributes, you can use an underscore between words, or capitalize the first letter of each word, as in IncomeAfterTaxes.

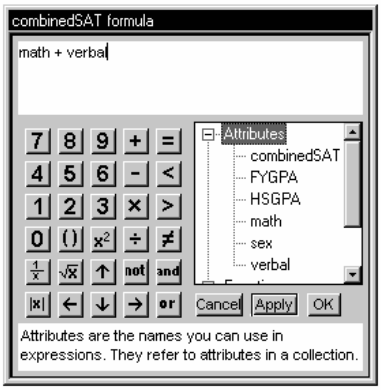
Create a Formula to Define an Attribute

Sometimes you want an attribute to be based on a formula. Perhaps you want Fathom to compute values from other attributes (such as adding math and verbal SAT scores to get a combined SAT score). Or perhaps you want Fathom to generate a collection using one of its random number functions.

- Select the attribute you want to define by a formula by clicking once on its name.

The shortcut for ***Edit Formula*** *is* ***Ctrl-E*** *(Win)* ***⌘-E*** *(Mac).*

- From the **Edit** menu, select **Edit Formula**. This brings up the formula editor for that attribute. (For more about the formula editor, see page 36.)
- Enter the formula (by typing, using the attributes and functions list, clicking a keypad button, or a combination of these).



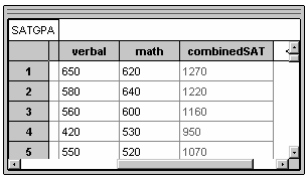
Notice that attribute names appear blue in the formula pane.

- Click **OK** to accept the change and close the formula editor.

Alt-F4 *(Win)* ***⌘-W*** *(Mac) is the shortcut for* ***OK****.*

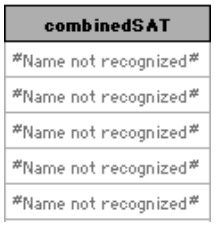
- Or, if you want to see the results of your formula without closing the formula editor, click **Apply** or press **Enter**. When you are through using the formula editor, click **OK**. You cannot do anything else in Fathom until you close the formula editor.

Notice that the results of the formula are displayed in gray to distinguish them from noncomputed values.



If you are going to be defining many attributes by formula, or if you want to see your formulas, you could also use the **Show Formulas** method described in the next section.

⊗ ***Does your attribute show an error, instead of the results you were expecting?*** If the attribute you have tried to define by formula gives you bad results (an error message appears between # symbols) that means there is a problem with your formula.

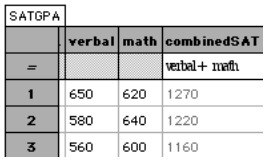


The phrase appearing between the # symbols gives you an idea of what is wrong. In the example to the right, #Name not recognized# means Fathom did not recognize something we typed (for example, we had typed verbl instead of verbal). For more on problems with formulas, see page 133.

View, Resize, and Hide Formulas in a Case Table

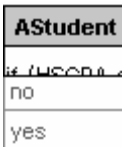
You can make the case table display any formulas you have used to define attributes.

- Click once on the case table to select it, if it isn't already selected.
- From the **Display** menu, choose **Show Formulas**.
- To hide the formula row, choose **Hide Formulas** from the **Display** menu.



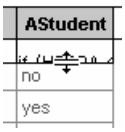
Show Formulas *also appears in the context menu for case tables: right-click (Win)* ***Control****-click (Mac).*

The formula row (labeled with an equals sign) now appears in the case table. The formulas defining attributes now show; the formula cells for noncomputed attributes have cross-hatching to indicate that they have no formula.

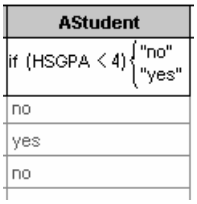


⊗ ***Can't see your formula?*** If you know a formula is there, but you can't see it, or can't see all of it, you probably need to resize the formula row height, the attribute width, or both.

Vertically Resize the Formula Row

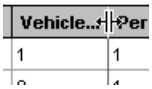


- Move the cursor over the lower border of the formula row until it becomes the resize cursor.
- Drag the border vertically and release.



Change Column Width

A column in a case table may not be wide enough to show its attribute name or its data. Or you may want to make columns narrower so that you can see more columns in a table at once.

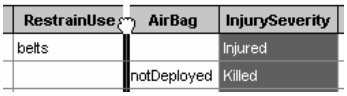


- Move the cursor over the boundary between two attribute names until you get the column resize cursor, as shown at left.
- Drag to resize.

Change the Order of Attributes

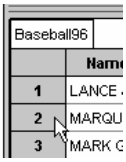
You can change the order in which the attributes appear in the case table—for example, to put all the attributes you are most interested in close to each other.

- Drag the attribute by its name to the border between the attributes you want to move it to. In the illustration at right, InjurySeverity is being moved to fall between RestrainUse and AirBag.



Select Cases in a Case Table

When you select a case, it is highlighted in all the graphs and tables in which it appears. This is useful when you want to see how some cases lie in relation to others. Also, to delete cases from a collection, you must first select them.



If you ***Ctrl-click*** *(Win)* ***⌘-click*** *(Mac) on an already selected case, it will deselect.*

- Select a single case by clicking in its row number.
- Select more than one case by dragging along the cases' row numbers.
- Select contiguous cases by clicking on the row number at one end of the desired section. Then hold down **Shift** and click on the row number of the case at the other end of the series.
- To select cases that are not contiguous, hold down **Ctrl** (Win) ⌘ (Mac) and click on each case.

Sort Data

You can change the order in which the cases appear in the case table.

- Click on the name of the attribute you want to sort by.

Choose ***Sort Descending*** *to sort in the other direction.*

- Choose **Sort Ascending** from the **Data** menu to sort from lowest to highest, either alphabetically or numerically.

♦ *How do I sort by more than one attribute?* Suppose you want to sort both by grade level and by name so that all the 11th graders (sorted by name) come before all the 12th graders (sorted by name). First sort by name, then by grade. The sort by grade won't disturb the sort by name. The biggest sort is the one to do last.

Hide Attributes

You may want to make screen space by hiding attributes. Hiding attributes removes them from view in the case table in which they are hidden, but does not delete them or their values. You can bring them back into view by *showing* them. (See next section.)

- Select the attributes you want to hide. (To select more than one, hold down the **Shift** key and click.)
- From the **Display** menu, choose **Hide Attribute**.

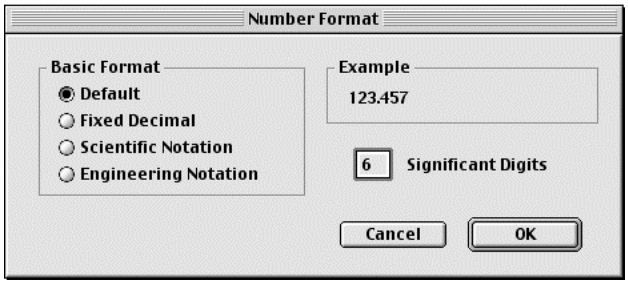
Show Previously Hidden Attributes

- Select the case table by clicking on it once.
- From the **Display** menu, choose **Show Hidden Attributes**.

Change Number Formats

You can control the way numbers are displayed in case tables, determining the number of decimal places shown or formatting numbers in scientific or engineering notation.

- Select the attribute whose number format you wish to change, by clicking on its name in a case table.
- Choose **Number Format** from the **Display** menu.



- Choose the option that you want and click **OK**.

Default

The default numeric format is a good general-purpose format. It keeps things out of scientific format except when doing so would be misleading with respect to significant digits. It doesn't display trailing zeroes after the decimal point.

Fixed Decimal

All numbers will have the same number of digits after the decimal point, the number specified by the **Decimal Places** field.

Scientific Notation

A number expressed in scientific notation has one digit to the left of the decimal point, as many digits as are specified in the **Significant Digits** field, and an exponent that specifies by what power of ten to multiply.

Engineering Notation

In engineering notation, the number of digits to the left of the decimal point is either one, two, or three, adjusted so that the power of ten will always be a multiple of three.

The chart below gives some examples, all with six digits specified in the **Number Format** dialog box.

Entered	Default	FixedDecimal	Scientific	Engineering
"1"	1	1.000000	1.00000e+00	1.00000e+00
"12345"	12345	12345.000000	1.23450e+04	12.3450e+03
".0123"	0.0123	0.012300	1.23000e-02	12.3000e-03
"12345678"	1.23457e+06	1234567.800000	1.23457e+06	1.23457e+06
"12.34567"	12.3457	12.345670	1.23457e+01	12.3457e+00

How to Work with Formulas

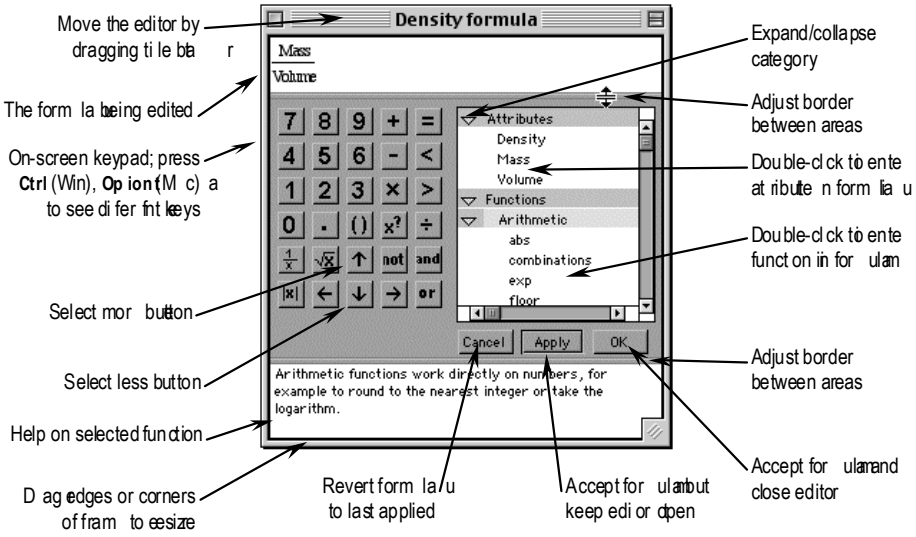
Formulas are used throughout Fathom for a wide variety of purposes. Any time you create or edit a formula, you use the same formula editor interface. You can use formulas to

- define attributes.
- filter cases in collections, tables, graphs, estimates, and tests.
- plot values and functions on graphs.
- determine what information appears in a summary table.
- determine where cases appear in collections and how they look and are captioned.
- define measures.
- control when sampling or measures collections stop.

Use a formula to …	**Example**
Determine the values of a new attribute.	Define an attribute, PopulationDensity, with the formula Population/Area or use a random function.
Define a filter.	Use the formula sex = "M" to look at only the males in a collection of people.
Summarize data.	To summarize student heights, you might use the formulas mean(height), median(height), and (max(height)+min(height))/2.
Define a statistic.	To compare tomato plants grown in sunny locations with those grown in shady locations, you might write this formula: (mean(height, loc = "sunny")–mean(height, loc = "shady")).

Control a simulation.	Suppose you wanted to throw dice until there were two sixes. You would use an "until" command in the **Collect Measures** pane of a measures collection inspector with the formula count(face=6)=2.

With formulas, you don't simply choose from a list of statistics or filters or simple computations. You can compute practically anything you want, as long as you know how to express it. See the next chapter, "Fathom Operators and Functions" for a complete listing of the functions available in Fathom.



This is the formula editor as it appears in the Mac platform.

Type a Formula

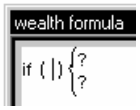
When you call up a formula editor, the cursor is in the formula pane. You can simply start typing the formula. If the formula is straightforward (such as attribute1+attribute2), normal typing will suffice. However, since the formula editor uses standard mathematical notation, there are a few tricks that will come in handy.

- The formula editor is *not* case sensitive. You can type all lowercase, and Fathom will interpret your expression correctly. For example, stddev works as well as stdDev. If your attribute is named DaysSpentAtBeach, typing daysspentatbeach will work.
- Type * (**Shift**-8) for multiplication, / for division, and ^ (**Shift**-6) for exponentiation.
- Exit fractions, exponents, and roots with the right arrow.
- Enter π by typing pi. As soon as you type an operator or a space or press any of the arrow keys, the symbol π appears in place of pi.

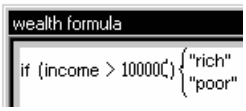
⊗ ***Did π not appear when it should have?*** When you type pi and then immediately type another letter or a number, Fathom doesn't know that you mean pi to be its own word. After typing pi, press one of the arrow keys or the spacebar (or * if you want to multiply). After the π appears, you can type anything you want, and the π will remain π.

- When you have typed an attribute name correctly, it turns blue.
- When you have correctly typed a global value (such as a slider name), it turns magenta.
- When you type the left parenthesis for a correctly spelled function, the function name turns brown.
- When you have correctly typed a constant, it turns red.
- When you type an open parenthesis, Fathom gives you both open and close parentheses, with the cursor between them. You can simply begin typing. To exit the parentheses, type a close parenthesis or use the right arrow.
- When you type an open quote, Fathom gives you both open and close quotes, with the cursor between them. You can simply begin typing. To exit the quotes, type a close quote or use the right

arrow. You must use quotes to enter string values (such as "rich" below right) in formulas.



- When you type if(you get a complete (though empty) if-statement, with three parts for you to fill in (as shown at left). Type the condition (in the example at right, income > 100000), then use **Tab** to move to the results (result-if-true goes on the top, result-if-false on the bottom). If you want the results to be words (also known as *strings*), you must surround them with quotation marks.



- When you have more than two results, instead of nesting if-statements, you can use the switch function. This is useful for recoding attributes. It takes an optional expression inside parentheses and then evaluates each of any number of true/false expressions to determine which value to return. To add an alternative, press **Insert** (Win) or **⌘-Option-Return** (Mac) on your keyboard. (For more detail on using if- and switch statements, see page 41.)

Adjust the Formula Editor Areas

When you want to see more of various portions of the formula editor, for example, when you have a tall formula or you want to see more of the formula list, you can move the borders of the formula and help areas by dragging their borders. (You cannot make the middle portion too small to see the keypad.)

Apply the Formula

You cannot do other things in Fathom with the formula editor open.

*Use **Enter** for **Apply**.*

- The **Apply** button causes the formula to be stored and evaluated, but does not close the formula editor. This can be very helpful when you are not sure how to get the formula you want or when you want to try different options.

*Use **Alt-F4** (Win) **⌘-W** (Mac) for **OK**.*

- The **OK** button causes the formula to be stored and evaluated and closes the formula editor.

*Use **Esc** for **Cancel** or **⌘-**. (Mac).*

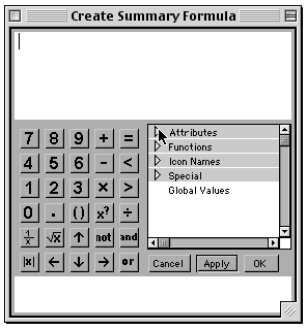
- The **Cancel** button closes the formula editor without changing the formula from what it was the last time you applied changes.

Use the Keypad

Click on the number and symbol buttons in the keypad to enter them into a formula. Everything that can be entered from the keypad can also be entered by typing from your computer's keyboard. When you hold down the **Ctrl** key (Win) **Option** key (Mac), you can see alternative buttons such as ≥.

Use the Attribute and Function List

Enter parts of your expression by double-clicking on items in the attribute and function list. This is handy if you have long attribute names or if you can't remember exactly how to spell a particular function, such as stdError for standard error.



- To see a list of choices in a given category, click the open/close control next to it. In the illustration at right, the user is about to open the attributes list.
- To enter an item from the list into the formula, double-click it.
- To close a list, click the open/close control.
- To see an explanation of an item, click on it once to select it. You can then read the help for that item in the help area at the bottom of the editor. This is especially handy if you can't remember the exact syntax for using a function.

Make Exponents

- Click the x? key on the keypad. When you've typed the exponent, you can exit it using the right arrow key.

Or…

- Type ^ (**Shift**-6) on your computer keyboard. When you've typed the exponent, you can exit it using the right arrow key.

Make Boolean Expressions

```
(sex = "F" ) and  (age > 60)
and  ( (marital = "Mar") or  (marital = "Div") )
```

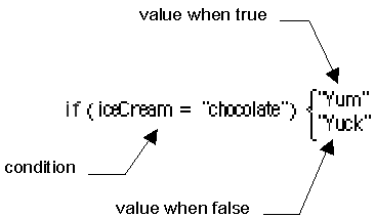
We're talking about true and false here and how to combine them. For example, if you're looking at census data and you want to find women over 60 who are either married or divorced, then the expression shown above will filter out all the other people in the census file.

- The comparison operators =, <, and > can be found on the formula editor's keypad. You can also find ,, ≤, ≥, ≠ by holding down the **Ctrl** key (Win) **Option** key (Mac). You can also enter all of these from the computer keyboard.

In the formula, "not" looks like ¬. This is the logic symbol for not.

- Express the not of an expression (the negative of an expression) by positioning the cursor in front of the expression and pressing the **not** key on the formula editor keypad.
- You can use and and or to string logical expressions together. Use the buttons on the formula editor keypad to enter these. Typing the words from the computer keyboard does *not* work.
- Some keyboard shortcuts: & for and, **Ctrl-Shift-O** for or, and ~ for not.
- When in doubt about which expressions are evaluated first, use parentheses to force the evaluation order you want.
- True and false are allowed values for attributes, so you can write a formula for an attribute that returns a boolean value.

Make If- and Switch Statements



Many times you would like the value of an attribute to depend on whether something is or is not true of another attribute. That's when you use an if-statement, like the one at right. Such a statement has three parts: the condition, the value when the condition is true, and the value when the condition is false.

To create an if-statement,

- Type if(. Fathom fills in the other parenthesis, the brace, and a question mark for each of the values to be filled in.

Or…

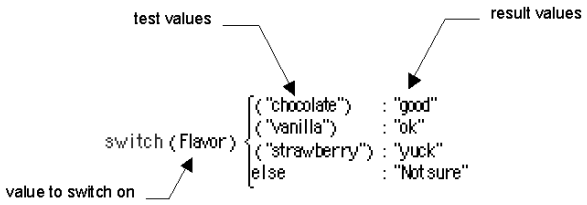
- Open up the function list to the **Conditional** category and double-click if.

You can create nested if-statements by using if-statements as values for if-statements.

```
if (iceCream = "chocolate") { if (sauce = "fudge") { "Wow!"
                                                     "Eat it"
                              if (hour > 8)         { "Eat it anyway"
                                                     "Pass" }
```

Switch statements are incredibly helpful for recoding data. (See "How to Massage Data" on page 46.) Unlike if-statements, switch statements can deal with more than two possibilities.

There are three parts to a switch statement: the value to switch on, the test values, and the result values. Usually you will write switch statements similar to the one shown here. Your value to switch on will be the name of an attribute, the test values will be values that the attribute takes on, and the result values will amount to a recoding of the original attribute. Fathom returns the result value whose test value matches the value being switched on.



A switch statement's tests are not limited to equality. In the statement at right, for example, Fathom first compares the value of N with 0 and returns "zero" if N does equal 0. If not, it tests to see whether N is less than 5 and returns "low" if it is. If not, it goes on to compare N with 10 and returns "medium" if N is less than 10. If no test has succeeded thus far, it returns "high."

```
switch (N) { (0)      : "zero"
             (? < 5)  : "low"
             (? < 10) : "medium"
             else     : "high"
```

Here are the rules to keep in mind when writing switch statements.

- If you do not explicitly specify a comparison operator, Fathom will test for equality.
- Fathom returns the first result value whose test returns true. The fact that later tests might also return true has no effect on the result.

- A switch value can itself be an expression, not just the name of an attribute, for example, $\frac{circum}{diameter}$.

  $$\text{switch } \left(\frac{\text{circum}}{\text{diameter}}\right) \begin{cases} (? < \pi) & : \text{"less than"} \\ (\pi) & : \text{"equal"} \\ \text{else} & : \text{"greater than"} \end{cases}$$

  The result of evaluating the expression is used in the test.
- Result values may be any expression (including if- or switch statements).
- The switch value is substituted for any question marks or missing arguments to functions that appear in the test expression.

Move the Cursor and Select a Portion of an Expression

Moving and positioning the cursor in the formula editor can be tricky. Here are some tips for getting what you want to happen.

- Generally you can click the mouse where you want to position the cursor, but when that fails, try the right and left arrow keys either on the formula editor keypad or on your computer keyboard.
- If you are inside a parenthetical expression, typing a right parenthesis,), will move the cursor to the right and outside of that expression.
- If you are inside a quoted expression (a string), typing a double quote will move the cursor to the right and outside the quoted expression.
- Pressing the **Tab** key will generally select the next selectable chunk of the entire expression. It's particularly helpful when you have if- or switch statements and you want to get to the next section.
- The up-arrow keys on the keypad and the computer keyboard will generally select the next largest chunk of the expression starting from the current selection. (Example: Pressing the up-arrow several times will generally select the entire expression.)
- The down-arrow keys on the keypad and the computer keyboard will generally select the next smallest chunk of the expression. (But this is less useful than the up-arrow, since it is hard to predict *which* smaller expression will be selected.)

The Effect of Selection

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

Selection can play an important role in creating formulas. When an expression is selected, pressing certain keys causes them to operate on the entire selection. For example, if a + b—without parentheses—is selected, and you press * (for multiplication), you get (a + b)*.

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

Copy and Paste Formulas

- To copy a formula in an attribute, click on the attribute name and choose **Copy Formula** from the **Edit** menu.
- To paste an already-copied formula into an attribute, click on the attribute name and choose **Paste Formula** from the **Edit** menu.
- If you can see a formula (in an inspector, in a summary table, in the formula row of a case table, in a filter, or in a graph), right-click (Win) **Control-**click (Mac) it to bring up a popup menu from which you can choose **Edit**, **Copy**, **Cut**, **Paste**, or **Clear**.
- The **Clear Formula** command will remove the formula entirely; for example, if you clear a formula in a list of plotted values in a graph, its entry in the list underneath the graph and the line in the graph will be removed.

Use a Slider Value in a Formula

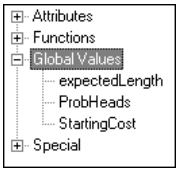
Sliders give you named values that you can use in formulas. (See "How to Use Sliders as Model Parameters" beginning on page 50 for a detailed description of sliders.)

- To use a slider value in a formula, type its name. (When Fathom recognizes it as a slider name, the name turns magenta.)

Or…

In future versions of Fathom there may be other kinds of global values.

- Click the open/close control next to **Global Values** in the function list to see a list of currently defined sliders. Double-click the desired slider name to enter it into the formula.



Other Formula Tips

- Multiplication is sometimes indicated as a dot between terms and sometimes, as in traditional algebra, as nothing between terms (though you need to put the asterisk in).
- Click on an item to get an explanation of it in the help area at the bottom of the formula editor.
- When things don't work the way you want, use parentheses.
- The up-arrow is very useful when you want to understand the structure of an expression.

How to Filter Data

At times you want to see only a subset of the cases in a collection. For example, if you were exploring the incomes of people, you would probably want to filter out the children and retired people in your data. In Fathom, a filter serves this purpose. A filter is an expression with a value of true or false. It can be attached to a collection, table, graph, estimate, or test. Cases for which the value of the expression is true pass *through* the filter; those for which the value is false do not.

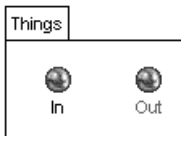
When you add a filter to a collection, anything that you make from the collection will respect the filter. A filter added to any other object applies to only that object.

Add a Filter

The shortcut for adding a filter is ***Ctrl-F*** *(Win)* ***⌘-F*** *(Mac).*

- Select the object to which you want to attach the filter. Choose **Add Filter** from the **Data** menu. The formula editor appears.
- Create the expression you wish to use as a filter and click **OK**.

A filtered collection still shows all its cases, but the captions of the filtered-out cases are gray instead of black. A graph or case table attached to a filtered collection will only show the cases that pass through the filter.



Two graphs attached to the same collection can have different filters. This can be useful when you want to compare the results of a display or analysis under two different conditions.

Remove a Filter

- Select the object whose filter you wish to remove.
- Choose **Remove Filter** from the **Data** menu.

Though the filter is removed, its formula has not been erased. The next time you add the filter back to this object, the original formula will show.

- ♦ Toggle back and forth between the filtered and unfiltered state using the keyboard shortcut **Ctrl-F** (Win) **⌘-F** (Mac).
- ♦ Edit a formula by double-clicking it.

How to Massage Data

Most data that come in from the real world need some work before you can find answers to your questions. In this section we describe some of the more common situations and how to deal with them.

Recode Numeric Codes to Categories

Much data come with categories coded numerically rather than as words. But when you're doing data analysis with Fathom, you would like the categories to be words you can read, not numbers you have to keep translating.

The collection shown in the case table at right has two numerically coded attributes.

People

	Sex	Education
1	1	1
2	2	3
3	2	4
4	2	2
5	1	3
6	1	4

Sex: 1 = Female, 2 = Male

Education: 1 = High School, 2 = College No Degree, 3 = College Degree, 4 = Graduate Work

First we rename the existing attributes to indicate that they are codes, leaving the "nice" names for the new attributes. (These first four steps are optional.)

- Double-click the Sex attribute name and change it to SexCode.
- Double-click the Education attribute name and change it to EdCode.
- Define a new attribute by choosing **New Attribute** from the **Data** menu. Call the new attribute Sex.

- Repeat the previous step for a new attribute named Education.

Now we'll define formulas that do the recoding.

- Click on the Sex column header (the empty attribute you just made), and choose **Edit Formula** from the **Edit** menu.

Be sure to use quotation marks around Female and Male.

- Enter the formula shown at right. You will need to insert the second option in the switch statement: press the **Insert** key on your computer keyboard (**⌘-Option-Enter** also works on a Mac).

```
switch (SexCode) { (1) : "Female"
                   (2) : "Male"
                   else : "Other"
```

We include the "Other" value in the formula so that wrongly entered data doesn't get coded as Male.

- Repeat the steps above for the Education attribute, using the formula at right.

```
switch (EducationCode) { (1) : "High School"
                         (2) : "Some College"
                         (3) : "College Degree"
                         (4) : "Graduate Work"
                         else : "Other"
```

The completed recoding is shown in the illustration below.

People

	SexCode	Sex	EdCode	Education
1	1	Female	1	High School
2	2	Male	3	College Degree
3	2	Male	4	Graduate Work
4	2	Male	2	Some College
5	1	Female	3	College Degree
6	1	Female	4	Graduate Work

You can now delete the recoding formulas and then delete the original attributes, although leaving the formulas and attributes in place allows you to trace things back to the original data.

Recode Ranges of Numeric Values to Categories

You can use the switch statement to recode ranges of numeric values to categories.

Suppose you have heights in inches and you want to group them as short, medium, and tall.

- Define a new attribute called heightGroup.
- Give it the formula shown at right.

```
switch (height) { (? < 60) : "short"
                  (? < 70) : "medium"
                  (? ≥ 70) : "tall"
                  else     : "missing"
```

Tips for entering this formula:

- You do not type the ? characters; Fathom enters them for you when you type the < or > characters.
- To enter ≥, hold down the **Ctrl** key (Win) **Option** key (Mac) and click the ≥ button on the formula editor keypad.
- To get a new line in the switch statement, press the **Insert** key on your computer keyboard (**⌘-Option-Return** also works on a Mac).
- If you get too many lines in your switch statement, you can get rid of them using the **Delete** or **Backspace** keys until there is nothing left in the line.

We include the else line of the switch statement in case there are values that aren't numbers. You could code this to an empty string, that is, nothing in between a pair of quotation marks, and that would allow Fathom to treat the missing values as truly missing.

Recode Missing Values

If your collection is small, you can find cases that need recoding with a filter and change them by hand.

Most real-world data have missing values, but there is no standard coding scheme for missing data. Fathom treats an empty value as missing. So how do you recode something as an empty value? Suppose, as shown in the left column at right, an attribute named thickness has missing values coded as asterisks, *. We want the asterisks recoded to empty values and everything else left intact.

thickness	thicknessRecode
4	4
3	3
2	2
4	4
*	
2	2
1	1

- Make a new attribute named thicknessRecode.
- Give it the formula shown at right.

```
if (isNumber (thickness)) { thickness
                            ""
```

If you want, you can then delete the formula, and then the original thickness attribute.

Restructure Data—Stack Attributes

Fathom expects that cases will be in rows and attributes will be in columns, but that isn't always the format in which people collect data. One common (and natural) format used for collecting data is to make a column for each group and record the values of some attribute under the column header. The example at right shows heights for boys and girls recorded, in no particular order, in two columns.

Heights

	Boys	Girls
1	64	64
2	66	58
3	64	63
4	56	63
5	60	56
6	61	58
7	61	59

The problem is that Fathom treats each row as a case, but, in fact, the two numbers in a row are not logically part of the same case.

- Select the collection (not the case table).
- Choose **Stack Attributes** from the **Analyze** menu.

This creates a new collection called, in this case, Stacked Heights, in which the values for all the columns of attributes are stacked on top of each other and a new attribute, called Group, is created with values corresponding to the names of the original columns. You could—and it would be a good idea to—rename Group to Sex and Value to Height.

Stacked Heights

	Group	Value
1	Boys	57
2	Boys	63
3	Boys	57
4	Boys	61
5	Boys	59
6	Boys	68
7	Boys	64

Compute a Running Sum

Receipts	Running...
49	49
58	107
75	182
55	237
29	266
66	332

Suppose you have a collection of daily sales receipts stored in an attribute named Receipts and you want to keep a running total. Here's how:

- Define a new attribute called RunningTotal.
- Give it the formula shown at right.

```
prev (RunningTotal) + Receipts
```

This works because the prev function for the first case returns 0, unless you specify otherwise. If you wanted the running total to reflect last month's receipts, you would include that number. Assuming you wanted the initial value to be 1,000, you would use this formula:

```
prev(RunningTotal,1000)+Receipts
```

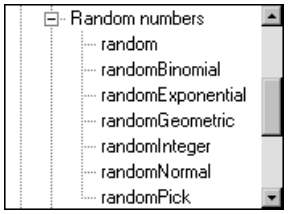
Generate Random Values

Fathom has many built-in functions for generating random values.

- Make a new attribute named RandomValues. (This *can* be the only attribute in a collection.)
- With the RandomValues attribute selected, choose **Edit Formula** from the **Edit** menu, or use the shortcut **Ctrl-E** (Win) **⌘-E** (Mac).

You can see more of the functions list by resizing the formula editor areas.

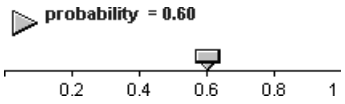
- In the list that appears on the right side of the formula editor, click the open control next to Functions and then open the Random numbers list. You will see a list of random number functions similar to the one at right.



- Double-click the function you want to use in your formula. (Or single-click it to read the help text at the bottom of the formula editor.)
- Insert parameters between the parentheses to determine what random numbers you get. You can use functions and existing attributes as arguments. For example, if you want random numbers from a distribution with a mean and standard deviation of some attribute already in the collection, you can use mean(attributeName), StdDev(attributeName) as arguments in the random function.
- Be sure your collection has cases. If it doesn't, choose **New Cases** from the **Data** menu.

How to Use Sliders as Model Parameters

Briefly, make a slider, change its name, and use that name in a formula.



A slider is a named value that you can use in any formula in your document. You might use a slider named probability in a formula to determine the proportion of yes votes in a simulation of polling; changing the slider value will change the proportion of yes votes in the collection of voters. Or use a slider named CoEfficient in a function you plot on a graph.

Make a Slider

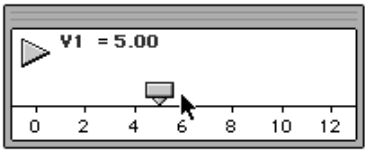
- Drag the slider from the shelf into the document. A dotted outline will show you where the slider is going to appear when you release the mouse button (or click again).

Or…

The shortcut to insert a slider is ***Ctrl-Shift-D*** *(Win)* ***⌘-Shift-D*** *(Mac).*

- Choose **Slider** from the **Insert** menu. The new slider will appear in an empty place in your document.

By default, sliders are named V1, V2, etc.; have a value set at 5; and have an axis that goes roughly from 0 to 12. On this slider, the cursor is near the slider's "thumb" which you can drag to change the slider's value.



Rename a Slider

- Double-click the name and type a new one. When finished, press **Enter**.

♦ A slider's name is important because you use it in formulas. Use meaningful slider names to make it clear what purpose each slider serves. The rules for naming sliders are the same as the rules for naming attributes: Each must begin with a letter and contain only letters, numbers, or underscores (_). Any invalid characters you type are removed when you press **Enter**.

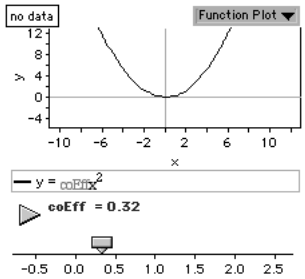
Change a Slider's Value

- Drag the slider thumb to a new position on the number scale.

Or…

- Double-click the numeric value and type in a new one. If necessary, the axis will change to include the value.

Use a Slider in a Formula



You can refer to a slider's value in any formula. When you create the formula, either type the slider's name, or double-click its name from the list of global values.

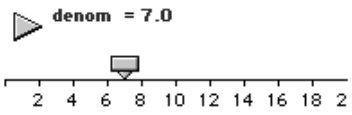
The example at right shows a slider being used to set the coefficient of a quadratic. As you drag the slider, the shape of the parabola changes.

Change a Slider's Scale

The numeric scale for a slider behaves just like a numeric axis of a graph. When the cursor is over the axis, you can drag to rescale or translate. (Also see "Change Axis Scales" on page 20.)

Or…

- Double-click the axis to bring up a Control Text object, as at right.
- Double-click any of the blue text to select and retype.
- Use the **Tab** key to move to the next field.



denom = 7.00000 Slider axis info:
The **denom** axis is horizontal from **0** to **20.000**

Animate a Slider

- Click the triangular button on the slider (to the left of its name). The button changes to an **x** button.

The slider thumb will move back and forth between the ends of the slider axis, and the value corresponding to the slider will change accordingly. Any formula containing the slider's name will be recalculated and all views of it will update.

- To stop the animation, click the **x** button.

While an animation is taking place, you can go ahead and do other things in Fathom (but the response might be a bit slow). You can slow the slider down by making it wider, or speed it up by making it narrower.

How to Compute Statistics

In Fathom, most statistics are computed by formulas that you write. (The exception is statistics that are computed by tests and parameter estimates.) Sometimes you will want to place these formulas and their results in a graph, other times it will be more useful to put them in a summary table, still other times you will want to store them as measures in the collection itself.

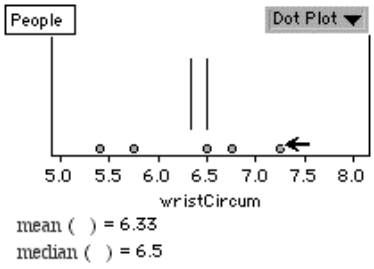
You can compute the value of any expression you can write in the formula editor. You are not limited to Fathom's built-in functions, but you will find the statistics functions (described beginning on page 101) to be very useful.

Compute and Plot Statistics in a Graph

When you compute statistics in a graph, they usually show up as plotted values along with your data, a more useful result than an isolated number. We describe the mechanics of plotting values in "Plot Values" beginning on page 23. Here we give two examples of what you can do with this capability.

Compare the Mean and the Median

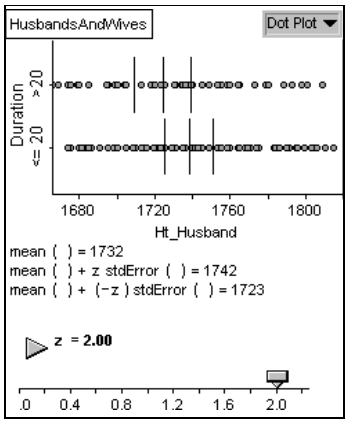
The graph at right shows the mean and median of the wrist circumferences of five people. It is quite instructive to change one of the data points by dragging it with the mouse. The mean changes continuously with changes in the data point, whereas the median changes only when the data point moves past the middle of the five points.



- You can compute anything you want in a plotted value, for example, the count of cases or the difference between the mean and the median. Be aware, though, that if the value falls within the bounds of the graph axis, it will also appear as a line on the graph. This can be misleading. It doesn't really make *sense* to show that there are five people on the same axis as wristCircum.

Show Standard Error in a Graph

The graph at right shows the heights of a random sample of 199 couples in England in the 1980s. The dot plot is split by couples married more than 20 years and those married 20 years or less. Below the graph are three formulas and a slider labeled z. The three formulas plot the mean ± z standard errors.



While the lines in the graphs show the values for each group of couples separately, the values reported below the graph are for the collection as a whole.

Make an Empty Summary Table

When you just want to compute some numbers, a summary table is the thing to use. There are two ways to make one. It starts out empty, and then you add attributes and formulas to it.

- Drag the summary table icon (the table with the big S in the middle) from the shelf into your document. Release it when you have the table positioned where you want it.

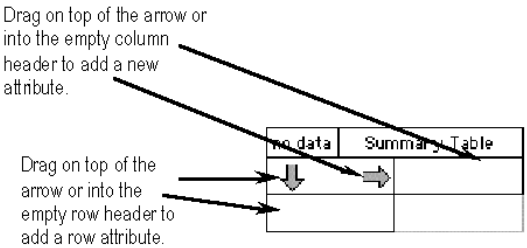
Or…

The shortcut for creating a summary table is ***Ctrl-U*** *(Win)* ***⌘-U*** *(Mac).*

- Choose **Summary Table** from the **Insert** menu. The new table will appear in some empty space in your document.

Add Attributes to an Empty Summary Table

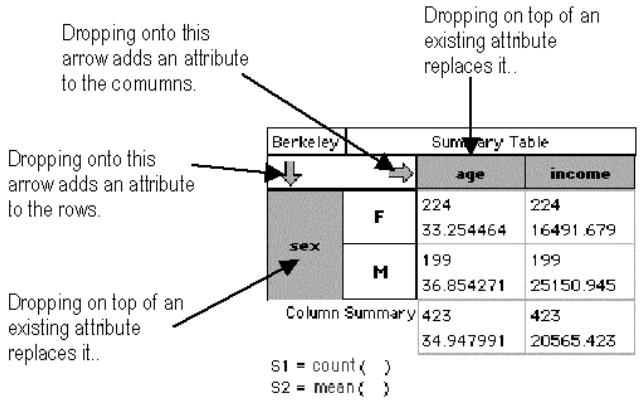
Add attributes by dragging them from somewhere else in your document, such as from the column headers of a case table or from the **Cases** pane in an inspector. As shown at right, you



can drag attributes either to the top of the summary table or to its side. (Although, putting categorical attributes on the side, rather than the top, conserves screen space.)

Add Attributes to a Summary Table

When a summary table already has attributes, there are four distinct places to which you can drag additional attributes, shown in the illustration below.

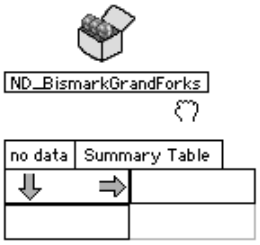


The summary table follows two rules that determine the results you get:

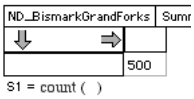
1. You cannot have both categorical attributes and numeric attributes along the same dimension (row or column) of the table. Dropping a categorical attribute onto rows (or columns) when there are numeric attributes present, eliminates the numeric attributes and replaces them with the categorical attribute. Similarly, dropping a numeric attribute onto rows (or columns) where there are categorical attributes present replaces the categorical attributes with the numeric attribute.
2. You cannot have numeric attributes in both the row and column dimensions of the table. Dropping a numeric attribute on one dimension when there are numeric attributes on the other dimension eliminates the attributes in the other dimension.

Connect a Collection to an Empty Summary Table

You can use a summary table without its having attributes on it. An empty summary table, unconnected to a collection, can't recognize any attributes in any collection. But once you connect a collection to a summary table, it will recognize attribute names in formulas you add to it. Here's how:



- Create a summary table.
- Drag the *name* of a collection to one of its drop places (the arrows or cells adjacent to them).



By default, the summary table shows the count—the number of cases in the collection.

- Double-click the formula to change it.
- Add more formulas via the **Summary** menu.

Add Formulas to a Summary Table

A summary table can have any number of formulas. The results of each formula computation are placed in corresponding rows of each of the cells. (That is, S1 is the first value in each cell, S1 the second, etc.) The steps are as follows:

- Select the summary table.

Add Formula *is on the summary tables' right-click menu (**Control-**click on Mac).*

- Choose **Add Formula** from the **Summary** menu.
- Enter the desired formula in the formula editor and click **OK**.
- You can also use the **Add Basic Statistics** command in the **Summary** menu to add count, mean, median, standard deviation, and inter-quartile range all at once.

Edit an Existing Summary Table Formula

- Double-click the formula you want to edit.

Or…

- Click the formula you want to edit and choose **Edit Formula** from the **Edit** menu.

Remove an Attribute from a Summary Table

- Select the attribute you want to remove by clicking once on its name.
- From the **Summary** menu, select **Remove Attribute**.

Remove a Formula from a Summary Table

- Click on the formula you want to remove and choose **Clear Formula** from the **Edit** menu.

Or…

- Double-click the formula to edit it. Delete the formula in the editor and click **OK**.

Using a Summary Table Without Attributes

You don't need to add attributes to use a summary table. An empty summary table can function as a calculator, for example, if you want to perform a calculation without having to leave Fathom. Also, you can connect a collection to an empty summary table and add formulas without having the attributes present.

- Drag a collection's *name* to an empty summary table.

By default, the table will give the number of cases in the collection. You can then add formulas that use attribute names, and the summary table will recognize the attributes and show the formulas' results.

Force a Continuous Attribute to Be Treated Categorically

Sometimes an attribute's values are numeric, but you want to treat each distinct numeric value as a category. For example, you might want to find the frequency of different years.

- Hold down the **Shift** key as you drag the attribute onto the summary table.

♦ This trick also works when dragging attributes to graphs.

Force a Categorical Attribute to Be Treated as Continuous

Fathom decides whether an attribute is continuous based on the presence of any non-numeric values for the attribute. But sometimes you have an attribute that has a mixture of numbers and strings and you want to treat the attribute as numeric, ignoring the string values.

- Hold down the **Ctrl** key (Win) **Option** key (Mac) as you drag the attribute onto the summary table.

♦ This trick also works when dragging attributes to graphs.

How to Sample from a Collection

Sampling in Fathom means to randomly choose some number of cases from a collection and put them into a new collection. The original collection is not changed.

There are two major reasons for sampling in Fathom: You are performing a simulation in which you want to see how the results of repeated sampling vary, or you would like to work with a smaller number of cases than there are in the original collection, perhaps because there are so many cases that computation is too slow.

Collect a Simple Random Sample with Replacement

You must select the collection, not a case table or graph.

- Select the collection from which you want to sample.
- Choose **Collect Sample** from the **Analyze** menu.

Or…

- Drag the collection's name to an empty collection.

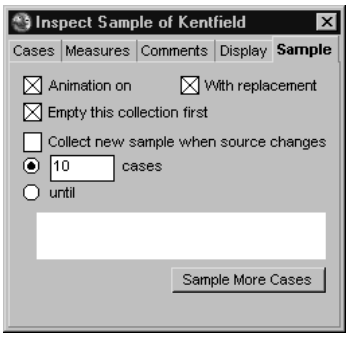
With replacement *means that any given case from the original collection can appear more than once in the sample.*

The result is a new collection labeled Sample of OriginalCollection. This sample collection contains ten cases selected at random with replacement from the original.

You can go on to view these sampled cases in a case table or a graph or do any analysis of them you like.

Change the Number of Cases Sampled

You can change the number of cases in your next sample.



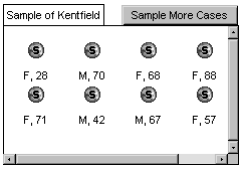
- Open the sample collection's inspector by double-clicking it or by selecting it and then choosing **Inspect Collection** from the **Edit** menu.
- Choose the **Sample** tab in the inspector.

⊗ ***Can't see the* Sample *pane?*** You probably selected the original, or *source,* collection. Close the inspector and make sure you call up the inspector for the *sample* collection (the box of gold balls with S on the lid).

- Replace the number 10 with the number of cases you want to have in your sample.
- Click the **Sample More Cases** button and the sample will be collected again with the specified number of cases.

Collect Another Sample

There are several ways to collect another sample:



- Resize the sample collection frame so that you can see its cases. You will also see a button in the upper right corner labeled **Sample More Cases**. Click this button to collect a new sample.

Or…

With the sample collection selected, ***Ctrl-Y*** *(Win)* ***⌘-Y*** *(Mac) is the shortcut for collecting another sample.*

- Select the sample collection and choose **Sample More Cases** from the **Analyze** menu.

Or…

- Open the sample collection's inspector, go to the **Sample** pane, and click the **Sample More Cases** button.

Change Other Sampling Controls

The last pane of the sample collection's inspector has sampling controls. Here we describe what they do and how to change them. To

try them out, you should first open the sample collection's inspector by double-clicking it or using the **Inspect Collection** command in the **Edit** menu. Then click the **Sample** tab to go to the correct pane.

Animation on

When animation is on during sampling, a case icon moves from the original collection to the sample collection approximately once for every 15 sampled cases. This can be helpful when you are repeatedly sampling and collecting measures from the samples (see page 61), because it makes clear the distinction between sampling cases and collecting measures.

The default is to have animation on, but you will want to turn it off when it slows things down unacceptably.

With Replacement

By default, sampling will happen with replacement. Conceptually, a case is drawn at random from the original collection and then put back so that it can be drawn again. If you turn this option off, each case can appear at most one time in the sample collection.

If the With replacement option is turned off, the greatest number of cases that can be sampled is the number of cases in the original collection.

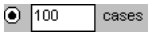
Empty This Collection First

By default, each time you sample, the cases from the previous sample are tossed out. If you uncheck this box, newly sampled cases are added to the old ones. You might use this, for example, to study how a sample proportion converges to the population proportion as the size of the sample increases.

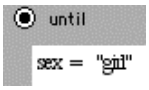
Collect New Sample When Source Changes

By default, when values in the source collection change, it has no effect on the sample. But if you turn this option on, you will get a new sample every time something changes in the original. If you have a fast enough computer, you can use this capability to show how the sample changes dynamically as you drag cases in the original collection.

Number of Cases to Sample



Use this option to set the number of cases in a sample.



Sample Until

Choosing this option allows you to specify a condition under which sampling should stop. First select the radio button next to the word **until**. Then double-click in the empty white rectangle to bring up a formula editor in which you can create the stop condition formula.

How to Collect Measures

In Fathom, when you want to gather results from something that happens repeatedly (for example sampling or rerandomizing), you use the **Collect Measures** command. Here's how.

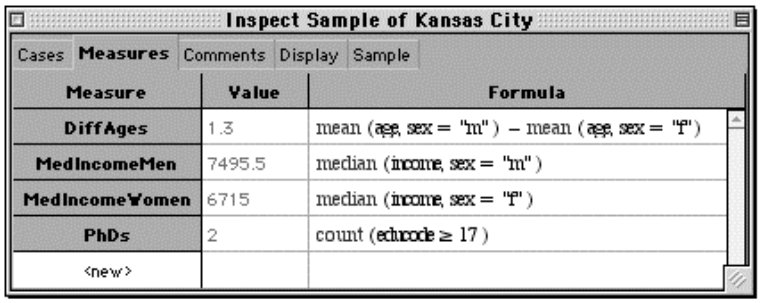
Collect Measures from a Collection

First, you need at least one measure to collect; a measure is a value that applies to the collection as a whole (such as the number of cases, the maximum run length, or the mean of heights). You define measures in the source collection's inspector.

- Choose the collection from which you want to collect measures.

This will usually (but not always) be a sample collection or a collection with randomly defined values for attributes. Each time the measures collection collects, it tells its source collection to update—to rerandomize or sample more cases.

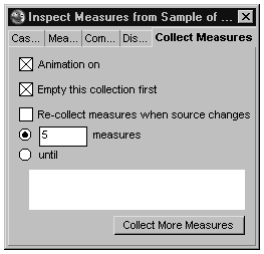
- Using the **Measures** pane of the chosen collection's inspector, define a new measure for each statistic you want to collect. (After naming a measure, double-click in its formula cell to enter a formula.) Some possibilities are shown below.



Another way to collect measures is to drag a measure from the inspector onto an empty collection.

- With the *collection* selected, choose **Collect Measures** from the **Analyze** menu.

A new collection will be created with five cases (the default number). Each case will have an attribute value for each of the measures you defined in the original collection. (You can drag these to graphs or summary tables.)



- Double-click the measures collection to bring up its inspector.
- Click on the **Col**... tab to bring the **Collect Measures** pane to the front.

You can change the defaults for collecting measures. Here we give an explanation for each option.

Animation

☒ Animation on

With animation turned on (the default), you will see a case move from the original collection to the measures collection once for each collected measure. Any graphs you have made of the measures will update case by case. Turn off animation to speed things up.

⊗ ***Do you still get animation, even though you turned it off in the measures collection inspector?*** If you are collecting measures from a sample collection, you need to turn off animation in it, too.

Starting with an Empty Collection

☒ Empty this collection first

The default setting guarantees that each time you collect measures, you will start fresh. If you want to add measures to those already collected, uncheck this option.

Automatic Re-Collection

☐ Re-collect measures when source changes

By default, changes in the source collection do not have any effect on the measures collection. If you want the measures collection to gather more measures whenever anything changes in its source, check this option. For example, if you have a collection driven by a slider you can collect measures each time the slider changes under animation.

⊗ ***Were you unable to stop the collecting process?*** If you can't stop the process by clicking **Cancel** or pressing **Esc**, stop the slider animation instead.

Number of Measures to Collect

Use this option to change the number of measures you want to collect.

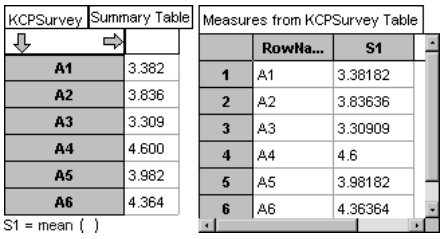
Collect Until Some Condition

Sometimes you don't know how many measures you want to collect; you only know the condition under which you would like to stop collecting measures. Select the **until** radio button, double-click in the white formula area, and type the stop condition into the formula editor that comes up.

♦ You can also collect measures from a measures collection.

Collect Statistics from a Summary Table into a New Collection

At times it is useful to convert the values computed in a summary table into a new collection.



The tables at right show one such situation. The mean scores from a series of survey questions have been computed in a summary table. These mean scores have been collected into a new collection with one case per question. Now you can compute statistics on these means or sort them.

The steps are simple.

- Compute the desired statistics in a summary table.
- With the summary table selected, choose **Collect Measures** from the **Analyze** menu.
- Make a case table for the new collection (see page 29).

Notice that each case corresponds to one cell in the summary table.

Collect Measures from a Test or Estimate

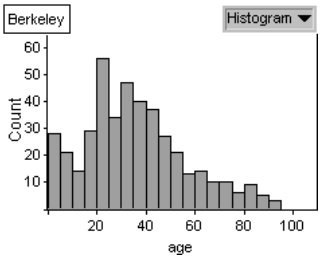
Select the test or estimate object, and choose **Collect Measures** from the **Analyze** menu. For more detail, see "How to Collect Measures from a Test or Estimate" on page 93.

How to Create Sampling Distributions

The concept of a sampling distribution is important in statistics. The idea is that you have defined some statistic for a random sample. Now you sample again and again, keeping the sample size the same and keeping track of this statistic for each sample. The values of the statistic vary, of course, and fall into a distribution. A histogram of these values represents the sampling distribution of the statistic. (More precisely, the histogram approximates the sampling distribution, which is the distribution we would obtain from all possible samples.)

Some statistics, especially those commonly used in statistical practice, are known to have sampling distributions that can be approximated, under certain conditions, by a normal or other probability distribution. For many statistics, however, the only way to come up with the sampling distribution is to use a computer to actually carry out the repeated sampling procedure.

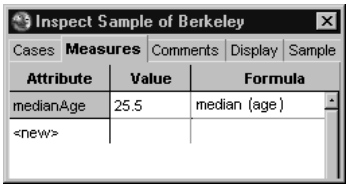
Here is an example in which we create the sampling distribution of the median age of people from a certain collection of individuals from the 1990 census. We consider the collection of 424 ages as the population. We draw simple random samples of 10 ages and compute the median age for each. The distribution of ages in the population is shown at right. Here's how we did it:



- With the original collection selected, choose **Sample Cases** from the **Analyze** menu.

This creates a new collection with 10 cases chosen randomly with replacement. If you want to change the number of cases per sample or to sample without replacement, change the sampling controls in the **Sample** pane of the sample collection's inspector (see page 59).

- Double-click the sample collection to bring up its inspector, then choose the **Measures** tab.
- Define a new measure, as shown at right, that computes the median age of the people in the sample. (You can define as many measures as you want.)

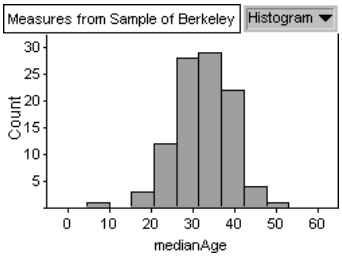


- With the *sample collection* selected, choose **Collect Measures** from the **Analyze** menu.

You now have a measures collection containing five cases, each case with one value for the median age of people in a sample. (If you defined more than one measure, each measure becomes an attribute in the measures collection.)

- Double-click the measures collection to open its inspector.
- If necessary, click the **Col**... (the last) tab to bring the **Collect Measures** pane forward.
- Change the number of measures collected from the default of 5 to 100.
- Click the **Collect More Measures** button.

The collection you get contains 100 values of the median age.



- Make a histogram of the collected measures. (Go to the **Cases** pane and drag medianAge to a graph.)

The histogram is an approximation of the sampling distribution for the statistic.

How to Make a Simulation

For many people, constructing simulations is one of the most interesting and rewarding things they do with Fathom. The possibilities

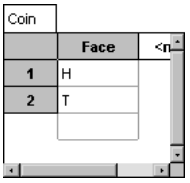
are endless. In this section, we offer two fairly simple examples to give you the idea. (Tour 6 in the *Learning Guide* gives another.)

Make a Coin Flip Simulation

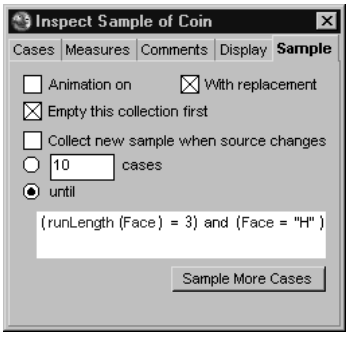
You may need to refer to other sections of this manual to accomplish some of these steps.

Suppose we are interested in the question, "On the average, how many times will we have to flip a coin before we get three heads in a row?"

- Make a collection with one attribute, Face, and two cases, one with Face = H and the other with Face = T. (It's easiest to use a case table, as at right.)

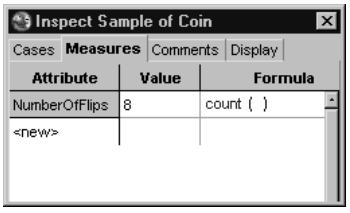


- Name this collection Coin. It represents the coin and its two faces.
- With the Coin collection selected, choose **Sample Cases** from the **Analyze** menu.
- Open the inspector for the sample collection and bring the **Sample** pane to the front.
- Configure the **Sample** pane as shown at right. (Double-click the formula area to create an **until** formula.)



The **until** condition tells Fathom to keep sampling until the three most recent values for the attribute Face are all the same and the most recent value of Face is H.

- Bring the **Measures** pane to the front.
- Create a new measure named NumberOfFlips.
- Double-click in the formula cell for the NumberOfFlips measure and give it the formula count( ).



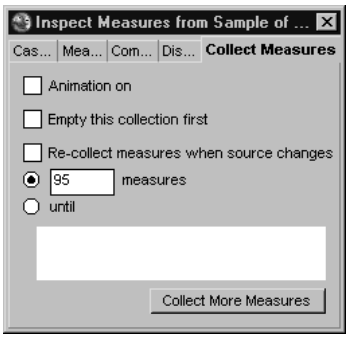
The pane should look similar to that shown here. The formula count(), when you don't specify count of what, gives

the number of cases in the collection.

- Close the inspector for the Sample of Coin collection.
- With the sample collection selected, choose **Collect Measures** from the **Analyze** menu.

You should see a new collection named Measures from Sample of Coin. Fathom collects five measures immediately. Depending on the speed of your computer and the luck of the draw, this can take a while.

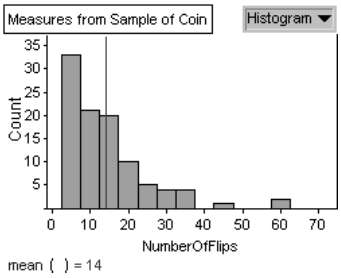
- Double-click the measures collection to open its inspector.
- If necessary, click the tab labeled **Col**... to bring the **Collect Measures** pane to the front.
- Uncheck the animation and empty-this-collection-first controls, and change the 5 to 95.



The pane should look like that shown at right.

- Click the **Collect More Measures** button.

It may take a few minutes to collect 95 more measures. If it is taking longer than you want to wait, click the **Cancel** button or press **Esc** on your computer keyboard. **⌘-.** (period) on a Mac also works.



- Create an empty graph.
- From the **Cases** pane of the inspector, drag the NumberOfFlips attribute to the *x*-axis of the graph.
- Change the plot to a histogram, as shown above.
- With the graph selected, choose **Plot Value** from the **Graph** menu.
- In the resulting formula editor, type the formula mean() and click **OK**.

The plotted mean value is the average number of flips, or "wait time," it took to get three heads in a row.

Make a Brownian Motion Simulation

Suppose we have a molecule bouncing around randomly and we want to know how far from some starting location it ends up after a certain number of bounces. A simulation can tell us.

We're going to assume that the molecule lives in a two-dimensional space and that on any given bounce it goes some distance between –1 and +1 in the *x*-direction and between –1 and +1 in the *y*-direction measured in some units, such as micrometers. We'll simulate 100 bounces and see how far the molecule gets.

- Make a case table and choose **Show Formulas** from the **Display** menu.
- Make two attributes, xPos and yPos.
- Give the xPos attribute the formula prev(xPos) + random(–1, 1) and the yPos attribute the formula prev(yPos) + random(–1, 1).

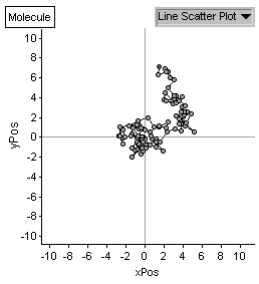
To get the formula editor, double-click in the formula cells.

Molecule

	xPos	yPos
=	prev (xPos) +	prev (yPos) +
1	0.608722	0.843619
2	0.0420229	0.153233
3	0.61864	0.937734
4	1.4863	0.744314
5	2.45099	1.66803
6	2.42396	1.24776

Since we didn't give the prev function an initial value, the molecule will start at point 0, 0. It hasn't gone anywhere, yet, because there are no steps, or cases.

- With the case table selected, choose **New Cases** from the **Data** menu.
- Specify that you want 100 new cases and click the **OK** button.
- Name the collection Molecule, because its cases represent the position of the molecule over 100 time intervals.
- Make a line scatter plot, as shown at right.



This graph represents the motion of the molecule. Each time you choose **Rerandomize** from the **Analyze** menu, you see a new possible path. If you adjust the axes to go roughly from –10 to 10 in both dimensions, you'll usually be able to see the whole "walk" after each rerandomize (if you leave the axes at their

default, you have to keep rescaling the axes to see each random walk).

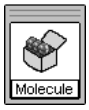
- Bring up the inspector for the Molecule collection.
- In the **Measures** pane, define a new measure, called distance.

We want this measure to compute how far away the molecule is from the origin in the last case.

- Double-click the formula cell for the distance measure and define the following formula:

$$\sqrt{(last(xPos))^2 + (last(yPos))^2}$$

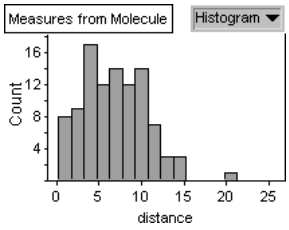
By squaring each final position, then taking the square root of their sum, you get the distance from the origin, regardless of the direction.



- Close the inspector.
- With the Molecule collection selected, choose **Collect Measures** from the **Analyze** menu.

The molecule collection rerandomizes five times and a new collection, Measures from Molecule, appears. It contains the final distance from the origin for each walk.

- Double-click the new collection to bring up its inspector.
- Set the inspector's **Collect Measures** pane to collect 100 measures, and click the **Collect More Measures** button.
- Make a histogram of the values for distance in the measures collection.



It should look similar to the graph at right. If your results were similar, you would have to say that zero distance is not the most likely value!

How to Use Scrambling to Test for Independence

Suppose you have two attributes that you think might be related to each other, but you have to acknowledge that it could be just chance that has produced the observed relationship. So is the relationship real or spurious? You can investigate this question by using a technique called scrambling. First you define some statistic—call it relatedness—

that quantifies the strength of the relationship between the two attributes. Then you scramble the values of one of the attributes, thereby breaking any relationship that does exist. You record the resulting value of the relatedness statistic. By repeating the scrambling and recording process, you see how the relatedness statistic varies when the relationship is guaranteed to be broken. Comparing the value of the relatedness statistic from the original data to the distribution of the relatedness statistic under pure chance variation allows you to estimate the probability that relatedness as extreme as the original value would have been produced purely by chance.

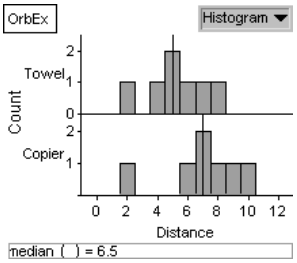
What scrambling does is very simple. It takes all the values of an attribute of your choice and rearranges them randomly. The other attributes are left untouched. Whatever relationship you observe between the scrambled attribute and other attributes after scrambling has to be due to chance.

Test for a Difference in Medians Between Two Groups

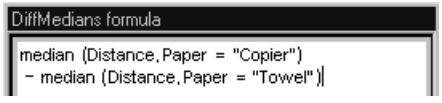
Briefly, define a measure in your collection, use ***Scramble Attribute Values*** *in the* ***Analyze*** *menu, then collect measures from the scrambled collection.*

Here we work through an example to show how scrambling works. Suppose we have conducted an experiment in which we drop crumpled-up paper from a certain height above a target.* After each drop, we measure the distance of the paper from the target. There are two kinds of paper, copier paper and paper towels, and we are interested in whether there is a difference in the distances they end up from the target.

The data are shown at right in a split histogram. The median for each group is plotted. We'll use the difference of the medians as our statistic. In this case, the difference is 7–5, or 2.



- Define a measure in the OrbEx collection as shown below.



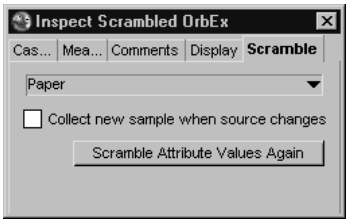
* See Erickson, Tim, *Data in Depth: Exploring Mathematics with Fathom*, "Orbital Express," Key Curriculum Press, 2000.

- With the OrbEx collection selected, choose **Scramble Attribute Values** from the **Analyze** menu.

A new collection, Scrambled OrbEx, results.

- Make a histogram of Distance split by Paper for the scrambled collection.

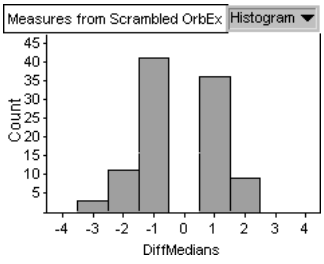
In the scrambled collection's inspector, notice that the DiffMedians measure is defined and that it computes the difference of medians for the two groups.



- Look at the **Scramble** pane in the scrambled collection's inspector.

You could have used the popup menu to scramble Distance instead of Paper. (Would it make a difference?)

- With the scrambled collection selected, choose **Collect Measures** from the **Analyze** menu.



- Use the techniques described in "How to Collect Measures" on page 61, to collect 100 values for DiffMedians and make a histogram for them.

Sample results are shown at right. The evidence is not very convincing. About 20% of the time, scrambling resulted in a difference of medians with an absolute value greater than or equal to 2, the original difference of medians. So we can't rule out the null hypothesis.

How to Estimate Population Parameters

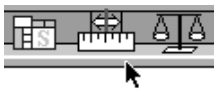
In the practice of statistics, we often have a randomly chosen sample and wish to make statements about the population as a whole. For example, we may have polled from the population of likely voters in an election and found the proportion in favor of a particular candidate or proposition. In this situation, and this is usually the case, what we find for the sample constitutes our best estimate of what is true for the population as a whole. Of course, our estimate based on the sample is

likely to be off a bit from the true value in the population. We construct a confidence interval to show the range of values for the population statistic that would most likely produce the estimation we have observed.

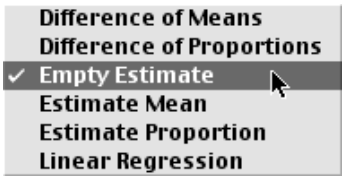
Fathom includes a basic set of estimation procedures that it can carry out for you, and these are described in this section. Alternatively, using Fathom's **Collect Measures** command, you can use powerful repeated sampling techniques to estimate parameters. (The *Learning Guide* that comes with Fathom contains a detailed example of parameter estimation.)

Fathom will perform the following estimates: **Difference of Means**, **Difference of Proportions**, **Estimate Mean**, **Estimate Proportion**, and **Linear Regression**. For all but **Linear Regression**, you can either drag attributes to the object to compute from raw data or compute from summary statistics. (**Linear Regression** requires raw data and cannot be computed easily from summary statistics.)

Here are basic principles for using Fathom's estimates.



- Choose **Estimate Parameters** from the **Analyze** menu or drag an estimate object (think "confidence interval") from the shelf.
- Choose the kind of estimate from the object's popup menu.



You can then either drag attributes to the top of the estimate, or (for most estimate types) input summary statistics. The estimate tells you how many attributes, and of what kind—categorical or continuous (numeric)—to drag. As with other Fathom objects, when an attribute is dragged over a drop area, that area becomes highlighted with a thickened border.


edit text cursor


popup menu crsror


edit formula cursor

The blue text in an estimate object is editable. There are three kinds of editable fields, distinguishable by the kind of cursor you see when over an editable area: The edit text cursor (I-beam), the popup menu cursor (triangle), the edit formula cursor f(x). The edit text cursor tells you that you can double-click, type, and press **Enter** to change the text (such as an attribute name). The popup menu cursor means that you can choose one of two or more given options. The edit formula cursor means that

you can bring up the formula editor, into which you can type a number or a slider name.

Each estimate describes the results in detail, explaining their meaning. You can toggle this feature by choosing **Verbose** from the **Estimate** menu. Unchecking **Verbose** gives the results in the form of more standard computer statistical output.

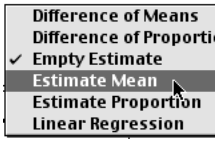
Estimate the Population Mean

Fathom computes a confidence interval for a population mean from the sample mean and the sample standard deviation. The assumption is that the distribution of sample means can be adequately approximated by the normal distribution. The assumption can be justified for a large sample by appealing to the central limit theorem from probability theory, or it may be justified if the original data have an approximately normal distribution.

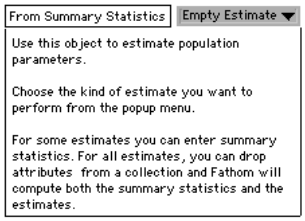
Estimate the Population Mean from Summary Statistics

- From the **Analyze** menu, choose **Estimate Parameters** or drag an estimation object from the shelf.

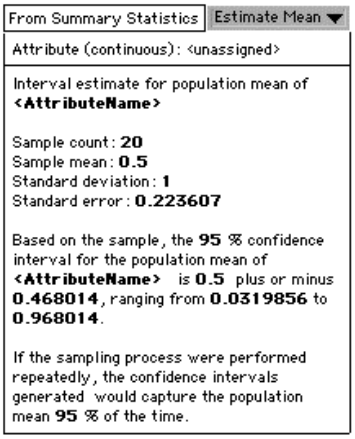
The object that comes up is empty because you haven't told Fathom what attributes you want to look at or what kind of estimate you want.



- From the popup menu in the upper right corner, choose **Estimate Mean**.



The estimation object should look similar to that shown at right. The text that is colored blue is editable. You can either fill in an attribute name and summary statistics or drag an attribute from a collection into the top portion of the object and have Fathom fill in all the statistics.



When you hold the cursor over <AttributeName> you get the edit (I-beam) cursor.

- Double-click <AttributeName>.
- Type an appropriate identifier and press **Enter**.

For all the other fields, you get an edit function cursor, as shown at left. Clicking the mouse brings up a menu with one option: **Change formula for value**.

- Hold down the mouse button and select **Change formula for value**....

This brings up the formula editor.

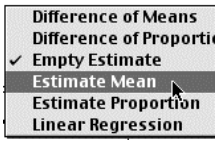
- Enter a value or a slider name and press **OK**.

As you enter values, the text changes. You might end up with something similar to the estimate shown at right.

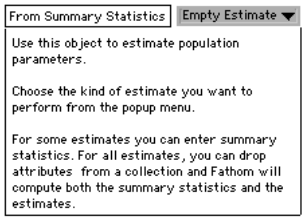
Estimate the Population Mean from Raw Data

- From the **Analyze** menu, choose **Estimate Parameters** to create an estimation object or drag one from the shelf.

The object is empty because we haven't told it the kind of estimate you want or given it any data to work from.



- From the popup menu in the upper right corner, choose **Estimate Mean**.
- Drag a continuous attribute from another object (such as a case table or graph) into the top pane of the estimate object, where it says that the attribute is unassigned.

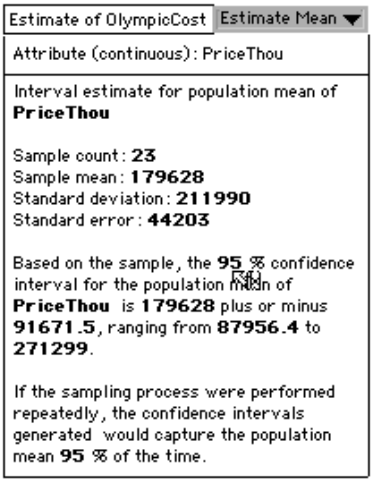


The relevant quantities in the estimate object recompute.

You can change the confidence level, if you want (95% by default).

- Click on the confidence level number and choose **Change formula for value**... from the popup menu.
- Enter either a number or an expression (such as the name of a slider) for the confidence level and click **OK**.

If you used a slider, you can see the effect that changing the confidence level has on the interval.



Make the Description Shorter

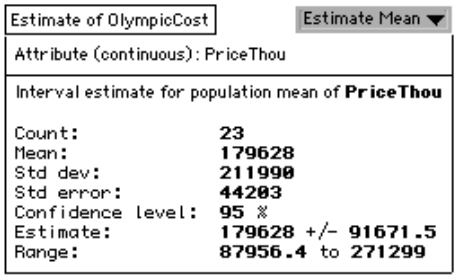
By default Fathom gives a verbose description of the estimation. You can make the description shorter.

- With the estimate object selected, click on the **Estimate** menu.

Notice that the **Verbose** command is checked, indicating that verbose descriptions are in place.

- Choose **Verbose**.

  This will uncheck the **Verbose** command and shorten the description of the estimation.



Estimate the Population Proportion

Estimating a population proportion is very similar to estimating a population mean. You can either enter known summary statistics or drag an attribute into the estimation object and have Fathom compute the summary statistics and the confidence interval.

Fathom uses the normal approximation to the binomial distribution when $n \cdot p \geq 5$ and $n \cdot (1 - p) \geq 5$. When these assumptions do not hold, Fathom uses the binomial distribution to compute confidence intervals.

Estimate the Population Proportion from Summary Statistics

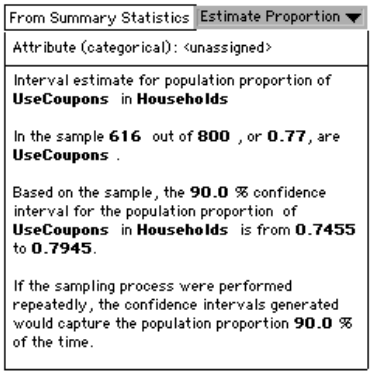
- Choose **Estimate Proportion** from the estimation object's popup menu.

Tab *moves you between attribute and category.*

- Enter the name of the attribute and the category whose proportion you are measuring.
- Click on the sample size and choose **Change formula for value...** from the popup menu.

This opens the formula editor.

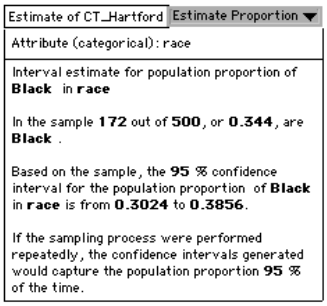
- Enter the sample size and click **OK**.
- Do the same for the number of cases in the category and the confidence level.



Shown above is the computation of the confidence interval for a population of sample households that were asked if they use coupons. The confidence level is tied to the slider labeled alpha.

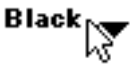
Estimate the Population Proportion from Raw Data

If you have the actual values for the attribute, Fathom will compute the summary statistics for you.



- With the Estimate Proportion object showing, drag an attribute into the top portion of the object where it says Attribute (categorical): <unassigned>.

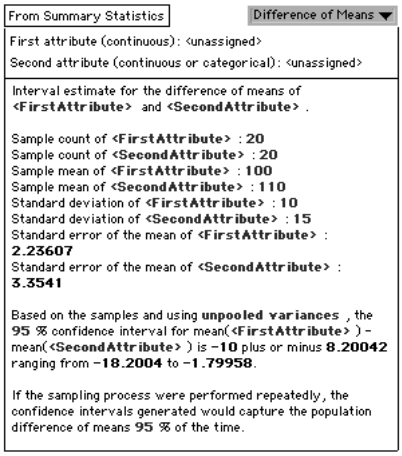
Typically, the attribute is categorical. If the attribute has numeric values, each value is treated as a separate category.



- Click on the category value to bring up a popup menu of possible categories. (The cursor changes to indicate you are over a popup menu item, as shown at left.
- Click on the confidence level and then on the popup menu.
- In the formula editor, enter the desired confidence level or a formula (such as the name of a slider) that will compute one.

Estimate the Difference of Means

Estimating the difference of means is much like the other estimates covered so far, but requires two attributes and takes two different forms depending on the configuration of the data you want to test. You can either "split" a numeric attribute by a categorical attribute (if the categorical attribute has only two values), or you can compare the difference of means between two numeric attributes. The former requires raw data; the latter can be computed either from raw data or from summary statistics.



- Make an estimation object by dragging one from the shelf or by choosing **Estimate Parameters** from the **Analyze** menu.
- Choose **Difference of Means** from the object's popup menu.

Estimate Difference of Means from Summary Statistics

- If you are comparing the means of two numeric attributes from summary statistics, double-click <FirstAttribute>, type an attribute name, **Tab** to <SecondAttribute>, and type its name.

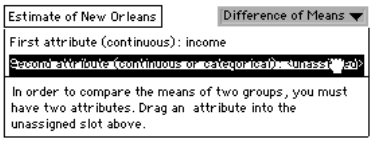
As soon as you type each attribute, its name is filled in throughout the object.

- As with the other estimates, you input the summary statistics by clicking on each value, choosing **Change formula for value**... from the popup menu, and entering either a value or a slider name in the ensuing formula editor.

By default, Fathom uses the unpooled variances for its calculations. This can be changed by clicking the phrase unpooled variances and choosing **the pooled variance** from the popup menu.

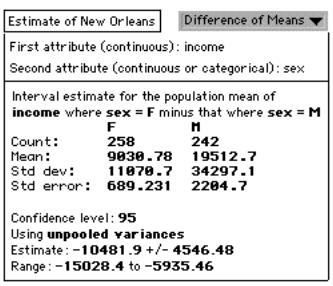
Estimate Difference of Means from Raw Data

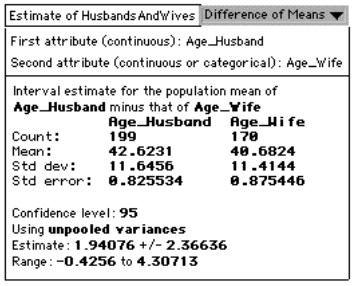
- Drag the attributes you want estimates for into the two drop lines in the top portion of the object.



In the first example below, the attribute **sex** was dropped on the **Second attribute** prompt, and then the **Verbose** option in the **Estimate** menu was unchecked.

The second example below illustrates the comparison of means of two numeric attributes (also with **Verbose** unchecked).



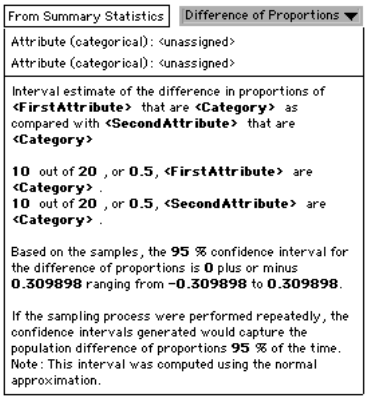


As with the other estimates, you can switch from unpooled variances to the pooled variance by clicking the phrase **unpooled variances**, and choosing the pooled option from the popup menu. You can change the confidence level by clicking it and changing its formula.

Estimate the Difference of Proportions

- Make an estimation object by dragging one from the shelf or choosing **Estimate Parameters** from the **Analyze** menu.
- Choose **Difference of Proportions** from the estimation popup menu.

Estimate Difference of Proportion from Summary Statistics



- Double-click the First attribute prompt, and type the first attribute name.
- **Tab** to the first Category prompt and type the category.
- Continue, entering the second attribute and category.
- When you have finished with the attributes and categories, press **Enter** to accept the final edit.

As in the other estimates, clicking on a value brings up a popup menu with the sole option **Change formula for value**..., allowing you to type the desired number or input a slider name.

Estimate Difference of Means from Raw Data

- Drag two categorical attributes (one at a time) to the top portion of the estimate.

By default, the first category (alphabetically) is used for each attribute.

- If necessary, change the category whose proportion you want to compare, by clicking on the category and selecting what you want from the ensuing popup menu.

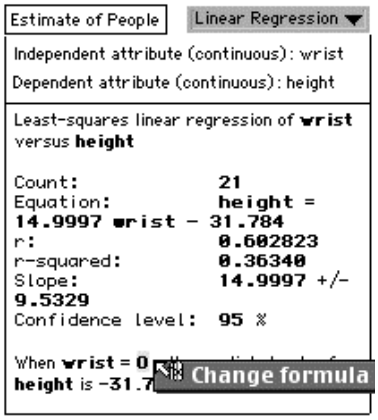
As always, you can change the confidence level by clicking it and changing it in the formula editor, and turn off **Verbose** by unchecking it in the **Estimate** menu.

Estimate the Slope and Correlation Coefficient

Given two numeric attributes, you can estimate the slope and intercept of the least-squares linear regression line and get confidence intervals for predictions based on that line.

- From the popup menu of an estimation object, choose **Estimate Linear Regression**.

- Drag two continuous attributes from some other object, such as a case table or inspector, into the top pane of the estimation object.
- To change the value for prediction or the confidence level, click the current value and choose **Change formula for value...** from the popup menu (illustrated at right, with the **Verbose** command unchecked).
- Enter a value or slider name, and close the formula editor.



How to Test Hypotheses

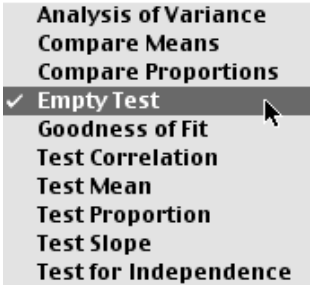
Though hypothesis testing in statistics can take many forms, the basic structure is always the same. We have a hypothesis, usually called the alternative hypothesis, that something is true—that the treatment reduces the risk of lung cancer, that the incidence of domestic violence is greater than it was five years ago, or that the kind of wood used in a wine cask has a discernable effect on the wine's taste. In classical inferential statistics, we want to know the probability that the observed results could be accounted for by chance, that is, that the so-called null hypothesis could plausibly account for what we see. To calculate this probability, we have to settle on a statistic that measures the amount of the effect and then determine a distribution for this statistic when the null hypothesis is true.

Fathom provides a small, basic set of hypothesis tests, described in this section. In addition to these built-in tests, you can use simulation techniques, especially scrambling (see page 69) and collecting measures (see page 61) to handle a wide variety of other situations.

Here are basic principles for using Fathom's tests.

- Choose **Test Hypothesis** from the **Analyze** menu or drag a test object (the balance icon) from the shelf.

- Choose the test type from the test object's popup menu.



You can then either drag attributes to the top of the test, or (for most tests) input summary statistics. The test tells you how many attributes, and of what kind—categorical or continuous (numeric)—to drag. As with other Fathom objects, when an attribute is dragged over a drop area, that area becomes highlighted with a border.


edit text cursor


popup menu cursor


edit formula cursor

The blue text in a hypothesis test object is editable. There are three kinds of editable fields, distinguishable by the kind of cursor you see when over an editable area: the edit text cursor (I-beam), the popup menu cursor (triangle), the edit formula cursor f(x). The edit text cursor tells you that you can double-click, type, and press **Enter** to change the test (attribute name). The popup menu cursor means that you can choose one of two or more given options. The edit formula cursor means that you can bring up the formula editor, into which you can type a number or a slider name.

By default, all test results are given in "verbose" form; you can see a briefer (and more standard-looking) set of results by choosing (unchecking) **Verbose** in the **Test** menu.

Each test offers you the option of seeing the test statistic distribution (the last item in the **Test** menu).

Test the Mean Against a Value

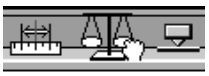
A simple random sample has been taken from some population, and the values of a certain attribute have been measured. The mean of these values has been computed. The question is whether, in the population as a whole, the mean for the attribute is different from some hypothesized value.

The alternative hypothesis is that the population mean is *different from* the hypothesized mean. The null hypothesis is that the population mean is *equal to* the hypothesized mean. There are three possible meanings of *different from*. These are *not equal to*, *greater than*, and *less than*.

Fathom uses the Student's *t*-statistic to carry out a test of a mean. The assumption for this test is that the distribution of sample means is

normal; this can be satisfied either by the population itself being normal or by the sample size being large enough that the central limit theorem is satisfied.

To begin the testing process,



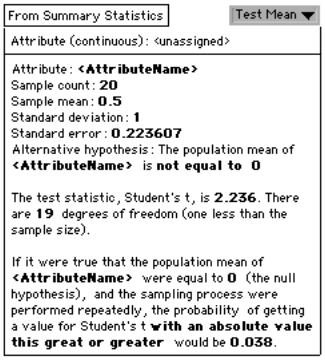
- Choose **Test Hypothesis** from the **Analyze** menu or drag a test object (the balance icon) from the shelf.
- From the popup menu in the upper right corner of the test object, choose **Test Mean**.

There are two possibilities: Either you have summary statistics for the attribute and no raw data, or you have raw data.

From Summary Statistics

When you choose Test Mean as the test type, you get a test already filled out with arbitrarily chosen values, which you can change.

- Double-click <AttributeName>, type a meaningful name, and press **Enter**.
- To change the sample count, click on the 20 and choose **Change formula for value**....

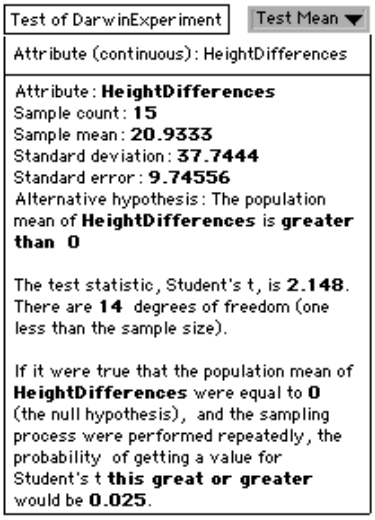


This brings up a formula editor in which you can input either a value or a slider name.

- When you have the value you want, click **OK**.
- Do the same two steps for the sample mean and standard deviation.
- To change to a one-tailed test, click on the phrase is not equal to and choose the desired form of the test from the popup menu.
- Click on the hypothesized mean and choose **Change formula for value**....

The values in the test are recomputed any time you accept a new value. If you used any sliders, changing their values will also force the test to recompute.

From Raw Data



- Drag a continuous attribute into the top pane of the test object where it says Attribute (continuous): <unassigned>.

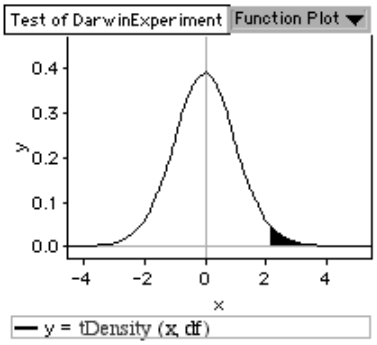
Fathom determines the number in the sample, the sample mean, and the standard deviation. Then it reports the results of the test just as it does when you enter summary statistics by hand.

- Choose the form of the alternative hypothesis in the popup menu that comes up when you click is not equal to.
- Change the hypothesized mean by clicking its value and choosing **Change formula for value...** from the popup menu, entering a formula in the formula editor, and clicking **OK**.

Show the Distribution for the Student's t-Statistic

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

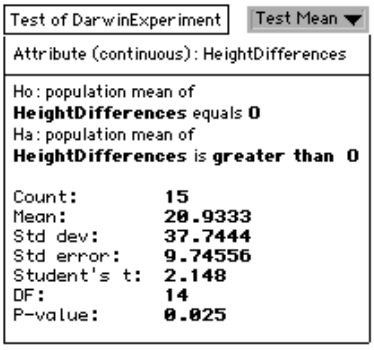
A function plot appears showing the distribution of Student's *t*. The shaded portion corresponds to the currently chosen alternative hypothesis. In the illustration at right, for example, the alternative hypothesis would be that the population mean is *greater than* the hypothesized mean.



It is especially instructive to change data by dragging in a graph while this distribution is showing or to link the hypothesized mean to a slider and drag the slider and watch the effect on the distribution's shading.

Make the Description Shorter

By default Fathom gives a verbose description of the test. You can make the description shorter.



- Click on the **Test** menu. Notice that the **Verbose** command is checked, indicating that verbose descriptions are in place.
- Choose **Verbose**. This will uncheck the **Verbose** command and shorten the description of the test.

Test the Mean Difference of Two Attributes

Fathom has no explicit test for paired means, but such a test is very simple to perform. Define a new attribute that is the difference between the two given attributes. The case table at right gives the idea.* Suppose we conduct an experiment by having each member of a football team kick two footballs: one filled with air, and the other filled with helium. Does the helium-filled football go farther? To answer the question, we define an attribute, Difference, using the formula Air – Helium, and then test the mean of this attribute against zero. Such a test of the mean difference between two attributes is called a paired *t*-test.

Heliumfootball

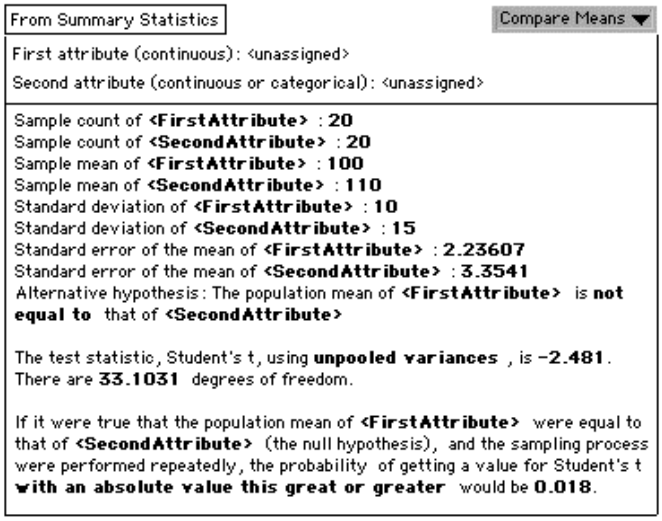
	Air	Helium	Difference
=			Air - Helium
1	25	25	0
2	23	16	7
3	18	25	-7
4	16	14	2
5	35	23	12

* From the Data and Story Library, http://lib.stat.cmu.edu/DASL/Datafiles/Heliumfootball.html

Compare the Means Between Two Groups

In the example above, suppose one group of people kicked the air-filled football and another group kicked the helium-filled football. Now there is no way of matching a given air-filled kick with any particular helium-filled kick. In the first example, each case was a person, for whom two kicks were recorded. In the example below, your data could be in two forms: either a case is a kick with two values (which kind of air, how far the ball went) or each case has two numeric attributes for distance. A set of individual differences, an Air – Helium attribute, would not make sense. Instead, we calculate the mean of the air-filled group and the mean of the helium-filled group, and then compare the two means. The comparison can be achieved by a using a two-sample *t*-statistic. In Fathom this is a **Compare Means** test.

- Choose **Hypothesis Test** from the **Analyze** menu or drag a test from the shelf.
- Choose **Compare Means** from the popup menu in the test object.



What you do next depends on whether you have only summary statistics or raw data and on how your data is structured.

From Summary Statistics

If you have only summary statistics, or you want to change values with one or more sliders, you can enter them into the test object.

- Move from the <FirstAttribute> to the <SecondAttribute> prompt using **Tab**.

As soon as Fathom has accepted an attribute name, that name appears throughout the test object.

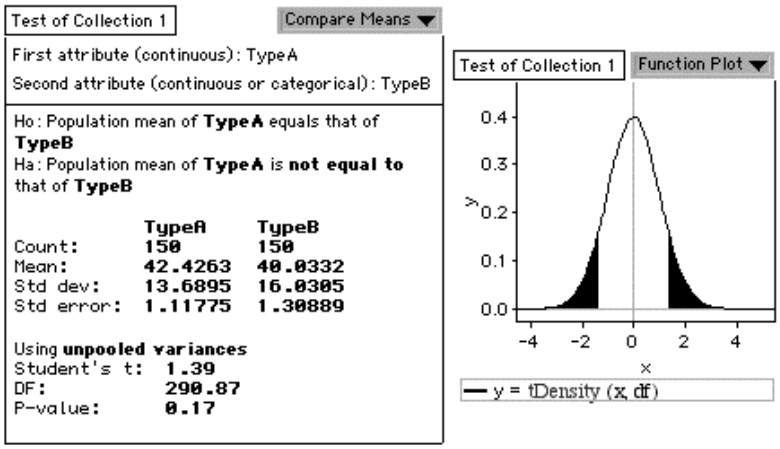
- Input the values or slider names via the formula editor you get when you click a value and select **Change formula for value**....
- If you want a one-tailed test, choose one of the options from the menu you get by clicking either **not equal to** or **with an absolute value this great or greater**.

As with the other tests, you can uncheck **Verbose** in the **Test** menu to get briefer output, and you can show the test statistic distribution, via the **Test** menu.

From Two Numeric Attributes

- If you have two numeric attributes representing the two groups, drag each one into the top portion of the test object.

In the example below, shown with the **Verbose** option turned off, the mean size of washers made by machine A are compared with the mean size of washers made by machine B.



By default, Fathom uses the unpooled variances. Clicking the phrase **unpooled variances** gives you a popup menu which allows you to select **the pooled variance**.

From One Numeric and One Categorical Attribute

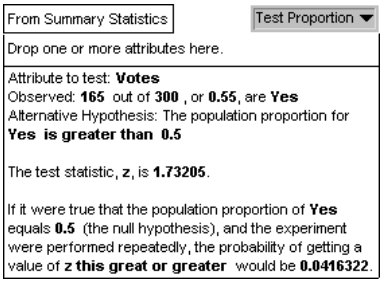
You may have two groups defined by a categorical attribute and want to compare their means for some continuous attribute.

- Drag the two attributes into the top area of the **Compare Means** test.

Fathom determines that one attribute is continuous and the other is categorical and checks to make sure that there are exactly two values for the categorical attribute. If so, it proceeds to compute the *t*-statistic and *P*-value.

Test a Proportion Against a Value

Suppose you have conducted a poll of a random sample of 300 voters before an election and that 55% of those polled said they would vote for your candidate. Can you safely predict victory?

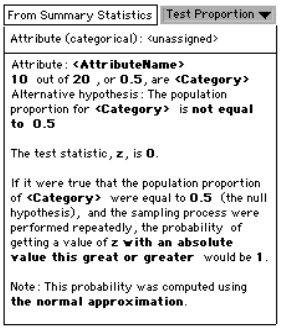


Your hypothesis is that the proportion of voters in the population that are in favor of your candidate is greater than 50%. The null hypothesis is that it equals 50%.

- Make a test object and choose **Test Proportion** from its menu.
- Type in the values from summary statistics.

Or…

- Drag the relevant attribute to the top part of the test object.



The hypothesized proportion, when clicked, gives you a popup menu. Choose **Change formula for value...** to type in a test proportion or a formula that yields a test proportion. It is especially interesting to type the name of a

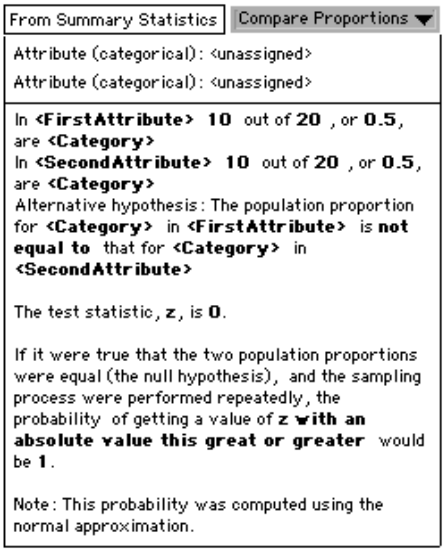
slider here.

For this test, Fathom uses the normal approximation to the binomial distribution when $n \cdot p \geq 5$ and $n \cdot (1 - p) \geq 5$. Under these conditions, it will compute a z-statistic. Otherwise, it will show a binomial statistic. (The test itself tells you which it used.)

Choosing **Show p_hat Distribution** from the **Test** menu, gives you a graph.

Compare Proportions

- Make a test object and choose **Test Proportion** from its menu.



If you have the raw data, drag the attributes to the top of the test.

- Change the categories by clicking on them and choosing the category you want from the popup menu.

If you have only summary statistics, enter the information. As usual, the cursors you get over various fields tell you how to change them.

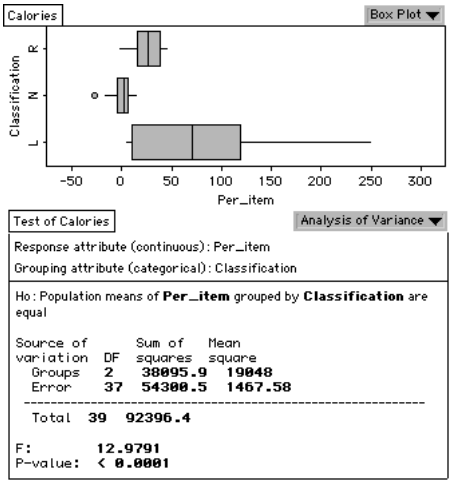
- **Tab** takes you to the next field among the attribute and category fields; the numbers are done via the formula editor.

As usual, you can get the non-verbose version and see the test statistic distribution via the **Test** menu.

Perform an Analysis of Variance (One-Way ANOVA)

Methods for comparing the means of two groups are described starting on page 86, but what do you do when you have *more* then two groups? The null hypothesis is that there is *no* difference in the means of the groups and the alternative hypothesis is that there is *some* difference.

The example at right shows a typical situation.* In this case the attribute is the difference between the measured number of calories per item and the advertised number of calories. The groups are L for locally advertised, R for regionally advertised, and N for nationally advertised. You can see from the box plot that there does appear to be some difference.



The basic question is whether the variation *between* the groups is sufficiently greater than the variation *within* the groups that we can conclude that something other than chance is at work. The statistic most often chosen to measure this is the *F*-statistic. The results of doing the analysis for the calories example are shown above.

Here are the steps:

- Create a test object and, from the popup menu in its upper-right corner, choose **Analysis of Variance**.

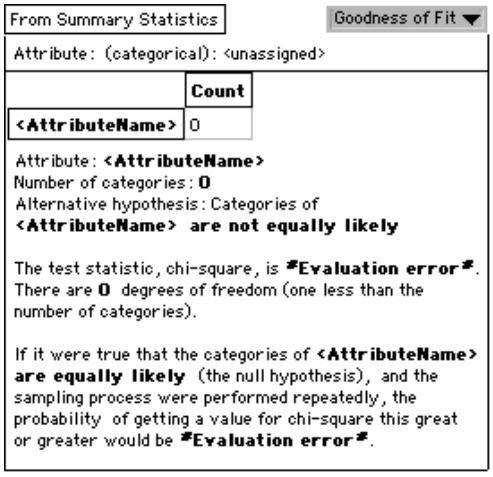
By studying the results of the analysis, you can make a judgment as to whether you are justified in rejecting the null hypothesis or not.

Test Frequencies of Groups (Chi-Square Goodness of Fit)

Consider rolling a die and discovering that the number 3 was coming up more often than any of the other numbers. To test the hypothesis that the die might be loaded, you would want to determine the probability that such a discrepancy (or one even worse) would occur if the die were *not* loaded. This is the purpose of the **Goodness of Fit** test.

* From the Data and Story Library, http://lib.stat.cmu.edu/DASL/Datafiles/Calories.html

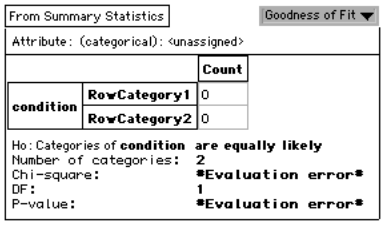
- Make a test object and choose **Goodness of Fit** from its popup menu.



Fathom allows you to specify summary statistics for the test by typing in an attribute name and a number of categories.

- Double-click one of the blue <AttributeName> prompts in the body of the test.
- Type a meaningful name.
- Press **Tab** to move to the Number of categories field.
- Type the number of categories and press **Enter**.

As soon as you do so, Fathom gives you a grid at the top of the test to input the categories and their counts, as shown (in non-verbose form) at right.



- Click in the RowCategory1 cell, and type the first category.
- Use **Tab** or **Enter** to accept the name and move to the next.
- When the category names are filled in, click in the first count cell to begin typing the counts.

To compute the goodness of fit from raw data, drag an attribute to the drop area in the top of the test.

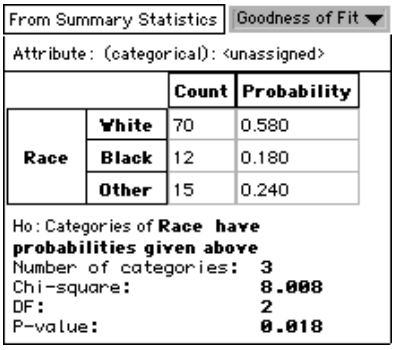
- Choose **Show Test Statistic Distribution** from the **Test** menu to bring up a graph of the chi-square distribution in which the shaded area corresponds to the probability of getting a chi-square as great or greater than that observed if the null hypothesis were true.

When not all categories are equally likely

Your null hypothesis may involve something other than equal probabilities for each category; for example, you might be wondering whether observed frequencies of people from different racial groups is significantly different from the proportions of those groups in the population as a whole. You can specify the expected probabilities by clicking on the phrase are not equally likely and choosing have probabilities different from those given above from the popup menu.

You get a Probability column in the table. Enter into that column the probability for each category you wish to use for your null hypothesis.

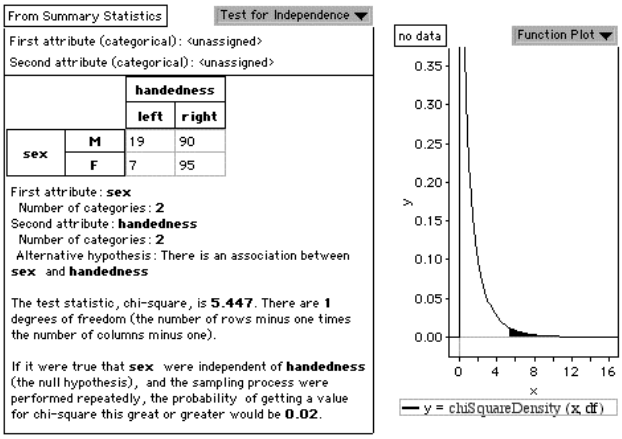
Fathom will adjust the last probability so that the sum of the probabilities is one. (Or, if you change the last probability, it will adjust the one above.)



The example at right shows a test in which counts for three racial groups are being compared against proportions found in the population.

Test for Independence of Two Categorical Attributes

A random sample of 200 students from a college campus in which, among other things, students were asked about handedness, reveals what might be a slightly increased prevalence of left-handedness among men as compared to women. The data, and a chi-square test for independence are shown on the following page.



In this situation, the null hypothesis is that there is no relationship between the two attributes, sex and handedness, that is, that the two are independent. The alternative hypothesis is that there is a relationship.

- Make a test object and choose **Test for Independence** from its popup menu.
- Type the names of the attributes.
- Specify the number of categories for each attribute and type the names of the categories in the table.
- Type the values for each cell in the table.

If you have the raw data, you can get results by dragging the attributes to the two places at the top of the test.

Fathom computes the chi-square statistic and *P*-value for the test.

- If you wish to see the chi-square distribution graphed, choose **Show Test Statistic Distribution** from the **Test** menu.

How to Collect Measures from a Test or Estimate

The values computed in a test or estimation object (see the sections starting with "How to Estimate Population Parameters" on page 71) are also available for collecting as measures. This can be useful when the estimate or test is of a simulated situation or when you are manually changing the parameters of the estimate or test and want to record each set of results.

- Select the estimation or test object.
- Choose **Collect Measures** from the **Analyze** menu.

The resulting collection will contain five identical cases (the default number of cases to collect). You can delete four of them. Some of the attributes in the measures collection for comparing a proportion are shown below. Each test or estimate has its own set of computed values that result in attributes in the measures collection.

Measures from Test of coins

	pValue	statValue	countInC...	countInC...	sample1...	sample2...	category1
1	0.687013	0.402911	29	27	0.58	0.54	heads

As usual with a measures collection, you can collect more measures and control the number and conditions for collecting measures in the **Collect Measures** pane of the collection's inspector. (See "How to Collect Measures" beginning on page 61.) If the attributes in the test or estimate are randomly generated, Fathom rerandomizes before collecting each set of measures.

How to Write a Report and Print It

Fathom has the basic tools necessary to write and print a report that you might do as part of a project. But Fathom is *not* a word processor. As your report gets longer and more complex, you will do better to copy your results from Fathom and paste them into another application that is designed specifically for report writing.

Type Text in Fathom

- Drag the text tool from the shelf into your document or use the shortcut **Ctrl-Shift-T** (Win) or **⌘-Option-T** (Mac).

As you move the mouse in the document, you will see the outline of a rectangle telling you where the text object will appear when you release the mouse (or click again).

Once you have positioned the text object, you can type whatever you like. Resize by dragging a corner or edge.

Here are some things to know about text objects:

- Fathom has no built-in formatting for text. If you want fancy formatting, type the text in a word processor and paste it into Fathom.
- You can have as many text objects as you like in your document.
- If a text object is empty, when it isn't selected, it will have a gray border around it to remind you where it is.

Paste Formatted Text into Fathom

Though you cannot format text within Fathom, you can create the text in another application and then paste it into Fathom.

- Create a text object.
- Choose **Paste** from the **Edit** menu.

The pasted text is not only formatted, but also editable.

⊗ ***The text isn't editable.*** You must create a text object and have it selected before you paste.

⊗ ***The text didn't appear.*** Perhaps you had some other object selected. If you had a collection selected, Fathom attempted to combine the text you pasted with the cases already in the collection. You should **Undo** and try again.

Paste Pictures

You cannot make drawings in Fathom, but you can paste in pictures. If you are writing your report in Fathom, this is a good way to illustrate parts of your report in which the data do not speak for themselves.

- In some other application, copy a picture onto your clipboard.
- Make sure that nothing is selected in your document. (Click in empty space to deselect all objects.)
- Choose **Paste Picture** from the **Edit** menu.

You can resize the picture just as you would any other object in Fathom by dragging on the edges of its frame.

You can also copy and paste pictures of Fathom objects in a Fathom document (for example, if you want several, static graphs of several randomizations).

- Select the object.
- Choose **Copy Picture** from the **Edit** menu.
- Click in an empty area of the Fathom document (to de-select all objects).
- Choose **Paste Picture** from the **Edit** menu.

Copy from Fathom to Another Application

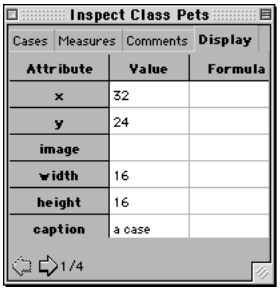
You can copy objects from your Fathom document to other applications. They will be pasted into the other application as pictures.

- Select the object you wish to copy.
- Choose **Copy Picture** from the **Edit** menu.
- Switch to the application into which you wish to paste the picture and choose **Paste** from that application's **Edit** menu.

♦ The picture you paste is *not* a live Fathom object. If you change something in the original Fathom document that you want reflected in the other document, you need to paste again.

How to Change the Appearance of Cases in a Collection

You can control the appearance of cases: the icons that represent them, their width and height, their positions, and the captions that appear under them. Here's how you do it.



- Double-click a collection to bring up its inspector.
- Click on the **Display** tab.
- For any of the attributes shown in this pane, you can enter values or create a formula. The x and y attributes determine the horizontal and vertical positions of the

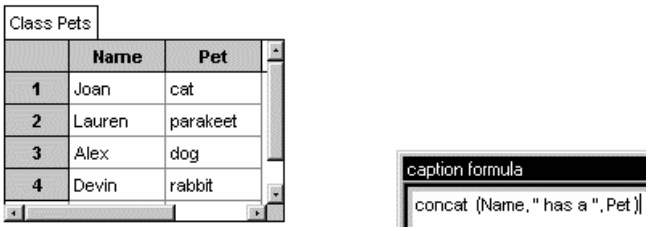
icons (by default, a gold ball) in the collection. (The modulo function is useful for the x attribute.) The caption attribute determines what is written underneath the case icon, and what appears in the status bar when you hold the cursor over a data point in a graph.

- To quickly change the caption, in a case table, select the attribute you want to use as a caption, and choose **Use as Caption** from the **Display** menu. If you select more than one attribute, Fathom concatenates them.

Example

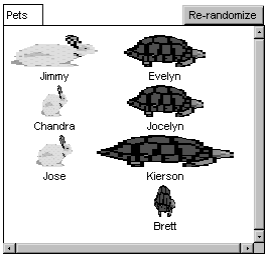
Suppose you have a collection with two attributes, Name and Pet. You could write a formula that produces captions such as "Lauren has a parakeet."

Caption: concat(Name, " has a ", Pet)

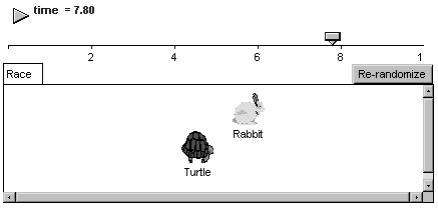


When you are editing the formula for Image, there will be a list of Fathom's built-in icons. You can look these over in the **Icon List** sample document (in the **Fathom Demos** folder under **Data Display and Exploration**). By using if- or switch statements in the formula, you can have the image depend on data values (see the example above, and the files in the **Data Display and Exploration** folder).

Since you can paste pictures onto cases in a collection, and you can control their length and width formulaically, you can position cases in a collection so that they convey information that would be difficult to convey in a more ordinary graph. The example on the right shows a Pets collection with pictures for the pets, the pets positioned according to kind, and the width of the pet proportional to the number of months the owner has had it.



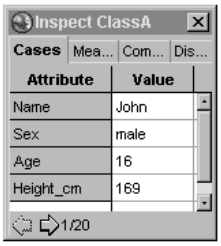
At right, we have a race between a rabbit and a turtle, where the time in the race is controlled by a slider. When you animate the slider the animals move. You can find this file in the **Algebra** folder in the **Sample Documents** that come with Fathom.

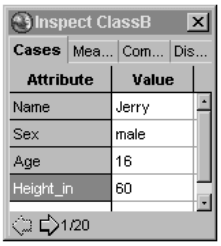


How to Merge Two Collections

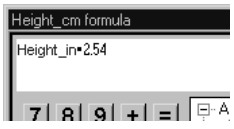
It frequently happens that you have two collections of data that have nearly the same attributes and you want to merge them into one collection.

The illustrations at right show the particular problem in which you have two collections of people and they differ in their use of the Height attribute. ClassA has chosen to measure height in centimeters, while ClassB has chosen to measure height in inches. How do you combine them into one collection in which height is measured in centimeters?





- In ClassB, define a new attribute, Height_cm. Give it the formula shown at right.



You now have the values you want in an attribute whose name matches the attribute in the ClassA

collection.

- Clear the formula for Height_cm so that when you merge, Fathom will not attempt to apply the formula to the attribute in the ClassA collection.
- Click once on the ClassB collection and choose **Select All Cases** from the **Edit** menu.
- Choose **Copy Cases** from the **Edit** menu.
- Click on the ClassA collection. Make sure no cases are selected in this collection.
- Choose **Paste** from the **Edit** menu.

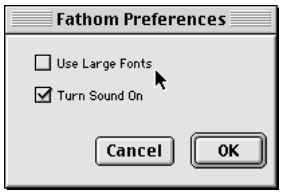
ClassA

	Name	Sex	Age	Height_cm	Height_in
17	George	male	16	157	
18	Laura	female	17	144	
19	Matt	male	15	156	
20	Nancy	female	18	148	
21	Jerry	male	16	152.4	60
22	Francine	female	17	139.7	55
23	Sarah	female	14	139.7	55

- Examine the cases in a case table to make sure the merge happened properly.

How to Display with Large Fonts

You can force Fathom to use large fonts globally. This is especially useful if you are displaying to a large group of people.



- Choose **Preferences** from the **Edit** menu.
- Check the **Use Large Fonts** box.
- Click **OK**.

How to Turn off Sound

- Choose **Preferences** from the **Edit** menu.
- Uncheck **Turn Sound on**.
- Click **OK**.

Fathom Operators and Functions

Here are the operations and functions currently supported in Fathom's formula editor.

Operators

- **Arithmetic operators**: +, –, *, /, ^ (**Shift**-6), !
  These obey normal algebraic precedence, but use parentheses when in doubt. The ^ operator raises to a power, and the ! operator computes the factorial of a number.
- **Logical operators**: and, or, not
  Use the keypad buttons on the screen to enter these. (You can type these operators if you are careful about using parentheses.)
- **Comparison operators**: <, >, =, ≠
  These are available on the keypad. You can also type <, >, or = from the computer keyboard. To type ≠, hold down **Ctrl** (Win) **Option** (Mac) and type =.
- **Comparison operators**: ≥, ≤
  Enter these by holding down the **Ctrl** (Win) **Option** (Mac) and **Shift** keys while typing > or <, or by holding down **Ctrl** (Win) **Option** (Mac) while clicking the buttons on the on-screen keypad.

Arithmetic Functions

abs	Use either the function, as in abs(first – second), or use the vertical bar button, | on the keypad, or use the | key on your computer keyboard.
combinations	Returns the number of combinations of *n* things taken *k* at a time. For example, combinations(5,2) is 10.
exp	Exponential function ("*e*-to-the…").
floor	Truncates to the next integer toward zero.
ln	Natural logarithm.
log	Common (base 10) logarithm.
modulo	Returns the remainder after one number is divided by another; modulo(11, 4) returns 3.
round	Rounds to the nearest integer.
square root	Use **Ctrl-Shift-R** or use the button on the keypad. You can also use the sqrt function from the functions list.

sgn	Signum function (+1 if it's > 0; –1 if it's < 0; 0 if it's 0).
trunc	Truncates, that is, lops off the decimals.

Trigonometric Functions

All of the trig functions you could want are here, plus hyperbolic functions and their inverses. They all use radians.

sin, cos, ...	All six trig functions and their inverses, including hyperbolic functions. sin(pi / 2) This gives you 1.00. You also get cos, tan, sec, csc, and cot.
asin(1.00)	Gives you 1.57, which is pi / 2. You also get acos, atan, asec, acsc, and acot.
sinh(1)	Gives you 1.18. The other hyperbolic functions and their inverses are constructed the same way: by adding an *h* to the end of the name. These hyperbolic functions and their inverses are not in the list of trigonometric functions, so you must type them from the keyboard.

Statistics Functions

One Attribute

For an attribute whose values are true *and* false, count *will return the number of cases for which the value is* true.

count	Returns the number of cases for which the expression is true. For example, count(NumberOfPets > 0) will return the number of cases for which NumberOfPets is greater than zero. Similarly, count(exists(Gender)) will return the number of cases for which the attribute Gender is defined. count() returns the number of cases in the collection.
first	Returns the first value in the collection for the given attribute; for example, first(height) would be 61 for a collection of people in which the first person's height is 61 inches.
iqr	Interquartile range, for example, iqr(blood_pressure). This function returns the value at the 75th percentile minus the value at the 25th percentile.

last	Returns the last value in the collection for the given attribute; for example, last(name) would be Zelda for a collection of ducks in which the last duck's name is Zelda.
max	Maximum value; for example, max(age).
mean	The arithmetic mean; for example, mean(height).
median	The median; for example, median(speed).
min	Minimum value; for example, min(salary).
percentile	Returns the value with a given percentile. For example, percentile(50, speed) is another way to compute the median. Or percentile(95, score) will return the score corresponding to the 95th percentile.
popStdDev	The standard deviation of the attribute you give it. This is the "population standard deviation."
popVariance	The variance of the values. This is also popStdDev squared.
proportion	Gives the proportion of cases for which the argument is true. For example, if 12 out of 24 people are over 12 years old, proportion(age > 12) will yield 0.5.
Q1	The value that lies at the 25th percentile; for example, the first quartile.
Q3	The value that lies at the 75th percentile; for example, the third quartile.
s sampleStdDev	Computes the sample standard deviation according to the formula $\sqrt{\sum \frac{(x-\bar{x})^2}{n-1}}$. It is an unbiased estimate of the population standard deviation. For example, s(pressure) computes the sample standard deviation of the attribute pressure.
sampleVariance	Computes the square of the sample standard deviation according to the formula $\sum \frac{(x-\bar{x})^2}{n-1}$. For example, sampleVariance(voltage) would compute the sample variance of the attribute voltage.

stdDev — Standard deviation; for example, stdDev(score). Computes the standard deviation of the cases in the collection using the formula $\sqrt{\sum \frac{(x-\bar{x})^2}{n}}$.

stdError — Returns the standard error; for example, stdError(score). The formula used is $\frac{s}{\sqrt{n}}$ where s is the sample standard deviation and n is the number of cases.

sum — Returns the sum of the values over all the cases. For example, sum(time)/count(isNumber(Time)) is another way to compute the mean of the attribute Time.

uniqueValues — The number of unique values that an attribute has in the collection. For example, uniqueValues(sex) will be 2 if there are only two values (male and female) for sex. (Missing values are ignored.)

variance — Computes the variance of an attribute, that is, the square of the standard deviation, according to the formula $\sum \frac{(x-\bar{x})^2}{n}$. For example, variance(before – after) computes the variance of the difference of the two attributes before and after.

Transformations

bin — Takes the form bin(a, bin, min, max) where a = attribute, bin = bin width, min = start of bin 1, and max = end. bin gives you a string (category value) for a—its "bin" as defined by the other arguments. For example; bin(3.14, 2, 0, 10) gives "b02" because the value (pi) is in bin #2 in [0, 10] with bins of width 2. (The last two arguments are optional.)

next — The value for the next case. If this is the last case, next returns 0. For example, next(year) returns, for each case, the value of the next year. As with prev, next takes an optional second argument that specifies the value to be returned for the last case.

popZScore — Returns the number of population standard deviations a value is from the mean. For example, popZScore(finalExam)

computes a standard score for each value of the attribute finalExam.

prev — The value for the previous case. If this is the first case, prev returns 0. For example, prev(year) returns, for each case, the value of the previous year. A second, optional, argument allows you to specify the value prev should take if there is no previous case. For example, prev(Factor, 1) will return the previous value of Factor for all cases except the first, for which it returns 1.

rank — Returns the position of the value when cases are ordered from lowest to highest. For example, rank(Population) used as an attribute in a collection of states assigns to each state its rank according to population. Note that if there are duplicate values, the rank will be fractional and the same for all the values. See also uniqueRank.

runLength(flip) — This one's wild! It gives the number of identical values immediately prior to and including the current value. For example, if flip contained {H, H, H, T, H, T, T}, this example would return {1, 2, 3, 1, 1, 1, 2}. You can use max(runLength(flip)) to compute the longest streak of heads or tails in a coin-flipping simulation.

sampleZScore — Returns the number of sample standard deviations a value is from the mean. For example, popZScore(height) computes a standard score for each value of the attribute height.

uniqueRank — Returns the unique position of a value in a list of values sorted from smallest to largest. Each value in the list gets assigned a different rank, even if there are duplicate values. For example, if attribute N contains the values {1, 2, 3, 2}, an attribute using the expression uniqueRank(N) will have values {1, 2, 4, 3}. See also rank.

zScore — Same as sampleZScore.

Two Attributes

correlation — Returns the correlation coefficient for two continuous attributes. For example, correlation(stories, height) will return the correlation coefficient for stories and height. This value will be between –1 and +1 and is a measure of how closely the values of

	one attribute follow those of the other.
covariance	Returns the average of the products of the deviations of each of two attributes from the mean. For example, covariance(hp, mpg) / variance(hp) would give the slope of the least square regression line for hp versus mpg.
linRegrIntercept	Returns the intercept of the least-squares regression line with x as the independent attribute and y as the dependent attribute.
linRegrSlope	Returns the slope of the least-squares regression line with x as the independent attribute and y as the dependent attribute.
rSquared	The square of the correlation coefficient for two attributes. covariance(x, y) represents the proportion of the variation of y that is accounted for by the variation in x. It takes on values between 0 and 1.

Expressing Qualifiers in Statistics Functions

Let's say you want to know the mean age of females in the collection. You need a qualifier. Your formula should be mean(age, gender = "female"). The last parameter in any statistics function is an optional true/false expression, that is, a qualifier.

Random Functions

random( )	A random number between 0 and 1.
random(max)	A random (real) number between 0 and max.
random(min, max)	A random (real) number between min and max.
randomInteger(min, max)	A random integer between its two arguments, *inclusive*. For example, randomInteger(1, 6) gives 1, 2, 3, 4, 5, or 6, chosen at random.
randomPick(a1, a2, ...)	This function gives you an element randomly chosen from a list of any number of arguments. For example: randomPick(1, 2, 3, 4, 5, 6) makes a die. randomPick("heads", "tails") makes a coin. randomPick("Male", "Male", "Female") gives you a

	population that is two-thirds "Male".
randomBinomial(n, p) n = number of trials p = probability of success	This function gives you a random integer from a binomial distribution. For example, randomBinomial(20, 0.5) gives the number of heads in 20 tosses of a fair coin.
randomNormal(mu, sd) mu = the mean sd = standard deviation	A random real number pulled from a normal distribution. For example, randomNormal(0, 1) gives you a number from a distribution with a mean of 0 and a standard deviation of 1.
randomGeometric(p) p = the probability of a "catch." Must be between 0 and 1.	A random nonnegative integer from a geometric distribution. You can think of the result as the number of turns it takes before you "catch" the case if the probability of catching it is the argument. Catching on the first try yields 0.
randomExponential(mu) mu = the mean. Must be positive.	A random real number greater than zero, pulled from a distribution that declines exponentially (so there are more near zero, just as in the geometric distribution). The larger the argument, the shallower the curve. Some people use the reciprocal of mu as the argument of their exponential distributions.

Other Random Number Functions

Fathom has a number of less commonly encountered random number functions, each of which returns numbers from a different distribution. These include randomBeta, randomCauchy, randomChiSquare, randomF, randomGamma, randomLattice, randomPoisson, and randomT. The help text that appears at the bottom of the formula editor will provide some guidance for using these functions.

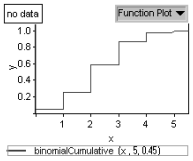
Distribution Functions

These functions give you access to statistical distributions. You can use distribution functions as part of case functions or part of aggregate functions.

Each distribution has three functions associated with it:

1. A cumulative function that returns the integral of the probability distribution from its lower bound up to a given value, x. This function is typically used to determine the likelihood that a value as extreme as, or more extreme than, some observed value would occur.
2. A density function that returns the probability density of the distribution at a given value, x. You can plot this function to see the shape of the distribution.
3. A quantile function that returns the value whose cumulative probability is equal to a given value, c. You can use this function to, for example, determine the critical value for a test statistic. You can think about it with reference to a standardized test where you are interested in the score on the test corresponding to some percentile. A quantile function returns the score.

Binomial Distribution



A plot of the cumulative distribution function when p = 0.45

Think of this distribution in terms of coin flips. You have a certain number of coins, n, and a certain probability of getting heads, p. This distribution deals with the probability of getting x heads. The possibilities are 0 heads, 1 head, 2 heads, …, n heads.

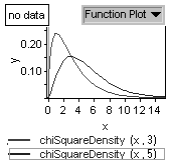
Instead of going from 0 to n, you can use the min and max parameters to go from min to max in intervals of (max – min) / n.

binomialCumulative(x, n, p, min, max) The binomial cumulative distribution function computes the cumulative probability, $Pr(X \leq x)$, where X is a random variable having a binomial distribution of n choices and the probability of success equal to p.

binomialProbability(x, n, p, min, max) This probability function computes the probability that $X = x$, where X is a random variable chosen from the set of possible values.

binomialQuantile(c, n, p, min, max) The binomial quantile function computes the value x such that $Pr(X \leq x) = c$.

Chi-Square Distribution



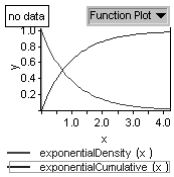
A plot of two chi-square probability densities, one with three and the other with five degrees of freedom

The chi-square distribution is used in Fathom when you are testing for goodness of fit or independence of two categorical attributes. The distribution has a degrees of freedom parameter, df, that determines its shape and must be a positive integer. In a goodness of fit test, for example, the number of degrees of freedom is one less than the number of categories.

These functions take optional parameters that determine a minimum value and a scale so that $Pr(X \leq x)$, where X = min + scale· C and C is a random variable from a chi-square distribution with df degrees of freedom. scale defaults to 1. min defaults to 0.

chiSquareCumulative(x, df, scale, min)	The cumulative chi-square distribution function calculates the cumulative probability, $Pr(X \leq x)$.
chiSquareDensity(x, df, scale, min)	The chi-square density function calculates the probability density, $(d/dx)Pr(X \leq x)$.
chiSquareQuantile(c, df, scale, min)	The chi-square quantile function computes the value x, such that $Pr(X \leq x)$ = c.

Exponential Distribution



A plot of both an exponential density and its cumulative probability

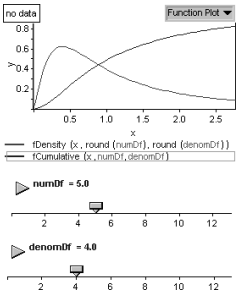
For the default values of scale and min (see below), the density function for an exponential distribution is just $\frac{1}{e^x}$. This means that random values that come from such an exponential distribution are concentrated in the smaller values.

The exponential functions described below take two optional parameters, scale, which defaults to 1, and min, which defaults to 0. The standard deviation of exponentially distributed numbers equals the scale, and the mean equals the sum of the minimum and the scale. We let E represent a random variable having an exponential distribution.

exponentialCumulative(x, scale, min)	This function computes the cumulative probability, $Pr(X \leq x)$, where X = min + scale· E.

exponentialDensity(x, scale, min) The exponential probability density function is given by $\frac{e^{\frac{x-\text{min}}{scale}}}{scale}$.

exponentialQuantile(c, scale, min) The exponential quantile function computes the value x, such that $Pr(X \leq x) = c$. It is the inverse of exponentialCumulative.



F-Distribution

See "Perform an Analysis of Variance (One-Way ANOVA)," on page 89.

The *F*-distribution is used in Fathom to compute *P*-values in a one-way analysis of variance. Under the null hypothesis, the test statistic in this analysis (called, appropriately enough, the *F*-ratio) is a random variable from an *F*-distribution.

The ratio referred to is the ratio of the mean square for treatments and the mean square for error. Each of these has associated with it a number of degrees of freedom. There is a different *F*-distribution for each distinct pair of degrees of freedom. These are labeled numDf and denomDf in the list that follows, corresponding to numerator and denominator. The function plot above right shows a plot of both the density and cumulative probability for the *F*-distribution, where sliders have been used to make it easy to explore the effect of the two parameters on the shape of the distribution.

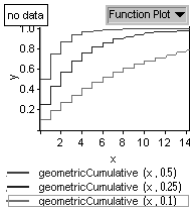
F-distribution functions also take two optional parameters: scale, which defaults to 1, and min, which defaults to 0.

fCumulative(x, numDf, denomDf, scale, min) This function computes the cumulative probability, $Pr(X \leq x)$, where X = min + scale· F where F is a random variable drawn from an *F*-distribution with the give pair of degrees of freedom.

fDensity(x, numDf, denomDf, scale, min) This function calculates the probability density, $(d/dx)Pr(X \leq x)$.

fQuantile(c, numDf, denomDf, scale, min) This function computes the value x, such that $Pr(X \leq x) = c$.

Geometric Distribution

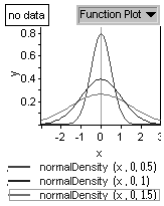


Shows three different probabilities comparing how long, on average, you have to wait for success

The geometric distribution can be used to explore how long you have to wait for a given event to occur, given the probability of that event. The probability, p, defaults to 0.5. We speak of X being randomly chosen with geometric probability from the set of values {min, min + scale, min + 2 scale, ..., } where min and scale default to 0 and 1.

geometricCumulative(x, p, scale, min)	This function computes the cumulative probability, $Pr(X \leq x)$, where X is randomly chosen as described above and where p is the probability of success on a given trial.
geometricProbability(x, p, scale, min)	The geometric probability function calculates the probability, $P(X = x)$.
geometricQuantile(c, p, scale, min)	The geometric quantile function computes the value x, such that $Pr(X \leq x) =$ c.

Normal Distribution



Shows how the standard deviation parameter of a normal density plot increases as the spread increases

The normal distribution has the familiar bell-shaped curve as its density function, coming from $\frac{1}{\sqrt{2\pi}} e^{-\frac{((x-\mu)/\sigma)^2}{2}}$, where μ is the mean and σ is the standard deviation of the distribution.

normalCumulative(x, mu, sigma)	This function computes the cumulative probability, $Pr(X \leq x)$, where X is a normally distributed random variable whose mean is mu and standard deviation is sigma.
normalDensity(x, mu, sigma)	This function calculates the derivative of $Pr(X \leq x)$. It produces the normal curve with a mean mu and standard deviation sigma.
normalQuantile(c, mu, sigma)	The normal quantile function computes the value x, such that $Pr(X \leq x) =$ c.

Other Distributions

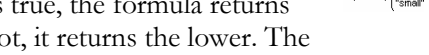
Fathom provides functions to compute a number of additional distributions, including beta, Cauchy, gamma, Poisson, *t*-, uniform, and uniform lattice. You can learn about these distributions using Fathom's online help.

Logical Functions

even	Returns true if the argument is even. even(10) returns true. even(11) returns false. If the argument is not an integer, you will get a #DOMAIN# error.
exists	True if the value exists for the indicated attribute. For example, exists(pressure) will be true for each case for which there is a value for pressure.
includes	Takes two arguments and returns true if the second argument is a substring of the first (also treated as a string). includes("the", "he") returns true. includes("dancing", "joy") returns false. includes(1234, 23) returns true.
inRange	True when a number is in a given range. For example, inRange(height, 60, 66) will be true for people whose height is greater than or equal to 60 and less than 66. If the third argument is less than or equal to the second argument, you will get an error.
isNumber	True if the value is numeric. For example, isNumber(date) will be true for the value 27 but false for the value June.
odd	Returns true if the argument is odd. odd(15) returns true. odd(20) returns false. If the argument is not an integer, you will get an error.

Conditionals

if	Creates an "if" block. If the expression in parentheses after if is true, the formula returns the upper value; if not, it returns the lower. The example at right returns "big" if the value for xx is greater than 2 and "small" otherwise.
switch	Acts like a sophisticated if, very useful for recoding attributes. It takes an optional expression inside parentheses and then evaluates each of any number of



true/false expressions to determine which value to return. This is best done with a couple of examples.

```
switch (age)
(? < 2)  : "infant"
(? < 13) : "child"
(? < 20) : "teen"
(? < 30) : "young adult"
(? < 50) : "mature"
else     : "wise"
```

Fathom interprets this by substituting age for ? in each of the expressions, starting at the top, until it finds one that is true. It returns the value to the right of the colon for that expression. So, if age is 45, the value of the entire expression is "mature."

```
switch (eduCode)
(10) : "High school"
(14) : "Bachelor's degree"
(15) : "Master's degree"
else : "Something else"
```

Here's another example, one that might be used to recode numeric data from the census into meaningful phrases. Fathom compares the value of eduCode with each of the numbers in parentheses and returns the expression to the right of the first one that matches. If none match, it returns the value of the expression to the right of else.

Text Functions

beginsWith Takes two arguments and returns true if the first begins with the second. For example, beginsWith(LastName,"Mc") will return true for McBribe, false for Binker.

concat Takes up to 10 arguments and returns a string. For example, if a case has the value "Denise" for the attribute Name, and the value "likes dogs" for the attribute Hobby, then concat(Name, " ", Hobby) gives "Denise likes dogs." The arguments can be numeric or strings. You have to include an argument for the space, to get a space in the caption.

endsWith Takes two arguments and returns true if the first ends with the second. For example, endsWith(LastName,"er") returns true for Binker, false for McBride.

includes Takes two arguments and returns true if the second argument is a substring of the first (also treated as a string). includes("the", "he") returns true. includes("dancing", "joy") returns false. includes(1234, 23) returns true.

stringLength Returns the number of characters for each case in the specified attribute.

stringToNumber Returns the first number in a value. For example, suppose you have dinosaur weights, but the attribute includes the units, such as 14kg. stringToNumber(DinoWt) will return 14 for that case. Fathom will then treat the recoded attribute as numeric (very handy for cleaning up imported data).

Special Values

binWidth The width of bins in a histogram. This special value is only available when you are plotting a function or value on a histogram.

caseIndex A value equivalent to the "row number." No parentheses are needed. This is the order in the *collection,* not the order you see in a sorted case table.

columnProportion The proportion of cases in a column of a summary table that belong in a cell. This special value is available only for formulas in a summary table, and it does not appear in the function list.

columnTotal The number of cases in a column of a summary table. This special value is available only for formulas in a summary table, and it does not appear in the function list.

expected The expected number of cases for a cell in a two-way table. This special value is available only for formulas in a summary table.

false A constant whose value is always false.

grandTotal The total number of cases that appear in a summary table. Note that this may not be the same as the total number of cases in the collection because some cases in the collection may not have the requisite values to show themselves in the summary table. This special value is available only for writing formulas in a summary table, and it does not appear in the function list.

π You can enter the value for π by typing pi or by double-clicking the entry for **pi** under **Special** in the function list.

rowProportion The proportion of cases in a row of a summary table that belong in a cell. This special value is available only for

	formulas in a summary table, and it does not appear in the function list.
rowTotal	The number of cases in a row of a summary table. This special value is available only for formulas in a summary table, and it does not appear in the function list.
true	A constant whose value is always true.

Fathom Objects

In this section we describe each of the kinds of objects in Fathom and provide a summary of how to manipulate that object.

General Principles

You can create a new object by dragging its icon off the shelf in the top of the Fathom window or by choosing a command (in the **Insert** or **Analyze** menus). Dragging from the shelf has the advantage of allowing you to place the object where you want it.

For most kinds of Fathom objects, you then drag one or more attributes to that object. When you drag an attribute over a drop area, you can see the drop area highlighted on your screen. You can also add formulas to most objects.

Some objects have their own menu in the menu bar at the top of the window. These menus are visible only when the object is selected.

Most objects have a shortcut menu that can be evoked by right-clicking it (Win) or by holding down **Control** and clicking it (Mac).

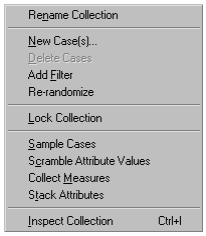
You can turn most objects into icons that can be easily moved out of the way by dragging from the lower right corner all the way up and left. Get the object back by enlarging it by dragging a corner, or see an enlargement of the object with the **View in Window** command in the **Display** menu.

You can delete an object by selecting it and then choosing **Delete [object name]** from the **Edit** menu or by using the keyboard shortcut **Del** (Win) **⌘-D** (Mac).

Collections

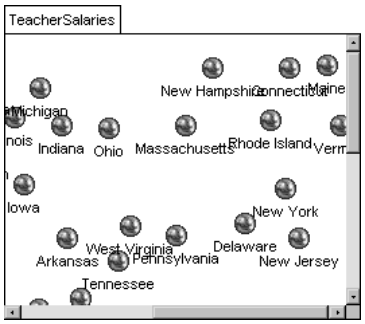
In Fathom, collections hold the data. Tables and graphs help you *view* the data. Delete a table or graph, and the data is still there. Delete a collection and the data is gone.

The collection at left has been iconified. When a collection is iconified, you can see its name, move it around in your document, sample from it, or inspect it.



The collection's context menu: right-click (Win) ***Control-****click (Mac).*

The collection at right is open. You can see each case as an icon, double-click a case to inspect it, change the case icons, move the cases around, and select individual cases.



- You can copy cases from one collection and paste them into another. Select the cases you want to copy, choose **Copy Cases** from the **Edit** menu, select the collection into which you want to paste the cases (make sure no cases are selected, or you will replace, rather than add to them), and choose **Paste** from the **Edit** menu.
- You can position cases wherever you like in the open collection. Click on a case and drag it.
- To display a collection in a case table, either choose **Case Table** from the **Insert** menu while the collection is selected or drag the name of the collection into a case table.
- You can change the icon appearance (by default, a gold ball) of a case by copying the picture you want (usually from some other application), selecting the case or cases you want to be displayed with that picture, and choosing **Paste as Case Icon** from the **Display** menu.
- You can also change the icon formulaically using the icons in the formula editor list.
- Use formulas to control the size and position of cases within an open collection by writing formulas for the *x*-position, *y*-position, width, and height from within the **Display** pane of the collection's inspector.
- A shortcut for sampling from a collection is to drag the collection name into an empty collection.

Inspectors

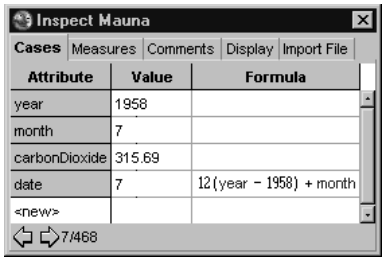
Each collection has an inspector with four or five panes, each of which controls a different aspect of the collection. An inspector, unlike graphs

or tables, does not live *inside* the document. Instead, an inspector is a separate window that always floats above the document window.

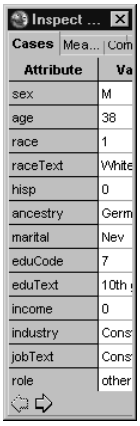
Here we describe each of the four panes that are always present in an inspector. Other panes—**Sample**, **Collect Measures**, **Scramble**, and **Import** are described in their appropriate sections.

Cases Pane

The **Cases** pane, an example of which is shown at right, allows you to view one case at a time and edit attributes, values, and formulas. Use the arrow buttons at the bottom of the pane to move from case to case.

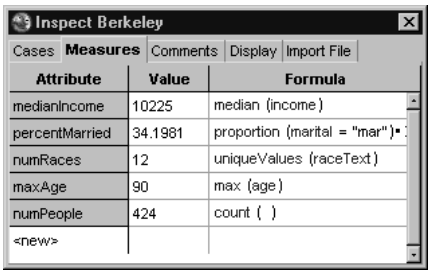


- Change the width of columns by dragging the border between column headers.
- Resize rows by dragging their borders.
- Edit values by clicking the value cell. (This won't work if the collection is locked or if the value is defined by a formula.)
- Drag attributes from the **Cases** pane to Fathom objects, such as graphs.
- Double-click an attribute name to rename it.
- Double-click a formula cell to edit the formula.
- Click <new> and type a new attribute name to make a new attribute.
- Right-click (Win) **Control-**click (Mac) to bring up a menu with which you can add new cases, delete the current case, and delete attributes.
- The **Cases** pane is particularly useful for collections that have lots of attributes. Set the size of the inspector to long and skinny, as shown at left, and use it as a kind of palette from which you can draw the attributes you need.



Measures Pane

Use the **Measures** pane to compute measures (or statistics) for the collection as a whole. A few possibilities are shown at right.

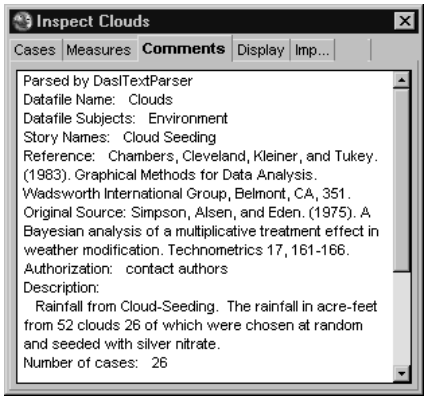


- Click in the cell labeled <new> and type a name of a measure to make a new measure.
- Double-click a name to edit it.
- Double-click in a formula cell to edit the formula.
- You don't have to define a formula for a measure. This can be useful when you want a constant value to use in attribute formulas.
- Use the **Collect Measures** command in the **Analyze** menu to gather these measures into a new collection. This is especially useful when measures change due to randomization, sampling, or scrambling.
- A shortcut for collecting measures is to drag any measure name from the **Measures** pane into an empty collection.
- Use the right-click (Win) **Control**-click (Mac) shortcut menu to delete a selected measure.

Comments Pane

For a collection that you create yourself from scratch, the **Comments** pane is initially empty. You can type documentation about the collection there, and it will be saved with the collection. It is a particularly good idea to explain attributes as comments.

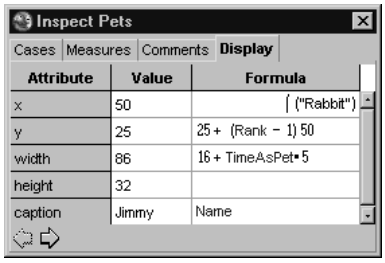
When you import a collection, anything that Fathom does not recognize as data is stored in the **Comments** pane.



- Right-click (Win) **Control**-click (Mac) in the **Comments** pane to bring up a popup menu with which you can cut, copy, and paste text.

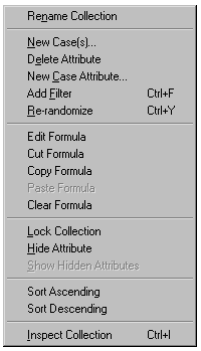
Display Pane

The **Display** pane gives you control over the way a case appears in an open collection. It is possible to type values for display attributes for individual cases, but it is more useful to specify them with formulas. See "How to Change the Appearance of Cases in a Collection" on page 96 for more details.



Also see the **IconNames** demo document to see what case icons are available. (**Sample Documents/Fathom Demos/Data Display and Exploration**)

Case Tables



The case table's shortcut menu: right-click (Win) ***Control****-click (Mac).*

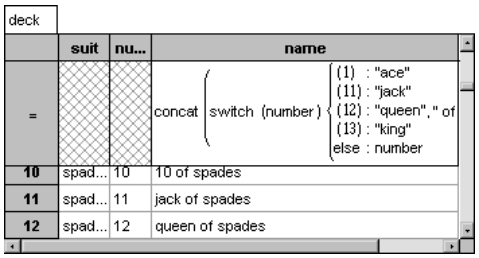
Case tables display data in a collection such that the columns correspond to attributes and the rows correspond to cases.

deck

	suit	number	name	<new>
10	spades	10	10 of spa...	
11	spades	11	jack of s...	
12	spades	12	queen of ...	
13	spades	13	king of s...	
14	hearts	1	ace of he...	
15	hearts	2	2 of hearts	
16	hearts	3	3 of hearts	

- Click on an attribute name to select the attribute.
- Click on a row number to select a case.
- Click in the column labeled <new> to add a new attribute to the collection.
- To insert an attribute in a particular place, select the attribute to the right of that place, and then choose **New Case Attribute** from the **Data** menu.
- Type new values in the bottom row of the collection to enter data for a new case.
- Use **Tab** to move to the next cell; **Enter** to move down one row.
- Drag column headers to new positions within the table.
- Drag between columns to change column width.

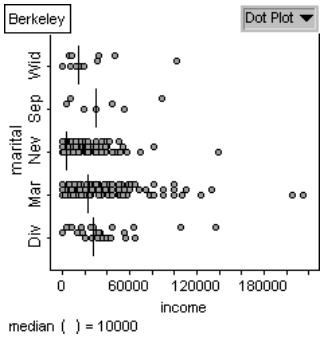
- Drag column headers (attribute names) to add them to other Fathom objects (graphs, summary tables, analyses).
- Double-click an attribute name to rename that attribute.



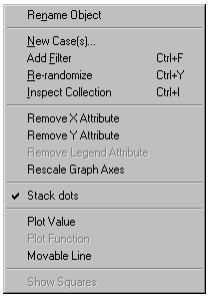
- Show the formula row (as above) by choosing **Show Formulas** from the **Display** menu.
- Resize the formula row by dragging the bottom of its row.
- Edit a formula by double-clicking it or by selecting the attribute and pressing **Ctrl-E** (Win) **⌘-E** (Mac).
- Right-click (Win) **Control**-click (Mac) in a value cell to bring up a simple edit popup menu. The usual keyboard shortcuts for cut, copy, and paste also work while editing values.

Graphs

Graphs display data, well, *graphically*. Use them when you're exploring and looking for patterns. But also use them to display computed quantities, such as the median incomes shown at right.



- To use an attribute in a graph, drag it from somewhere else (an inspector, a case table, another graph) and drop it on either of the axes or in the middle of the plot.



The shortcut menu for graphs depends on what plot you are showing. Shown here is the menu for a dot plot.

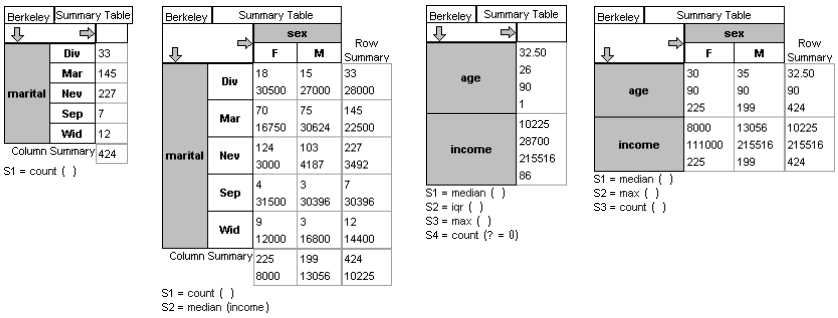
- The plot menu in the upper right corner of the graph displays the kinds of plots Fathom can make given the attributes that are currently part of the graph.
- When a graph is selected, a **Graph** menu appears in the menu bar. Many of the commands in this menu apply specifically to the kind of plot currently displayed, and they will change when the plot changes.
- You can select cases in graphs by clicking on points or bars. **Shift**-click toggles the selection state of the cases.
- Drag, starting in empty space in a graph, to select all cases within the selection marquee. A complete bar will be selected if any part of it is inside the marquee.
- In plots that display values for continuous attributes, you can change those values by dragging the points or bars in which they appear. (You can't drag to change data if the collection is locked or if the values are computed formulaically.)
- To swap the attributes on the axes, drag the attribute on the *x*-axis to the *y*-axis, or vice versa.
- In most graphs, you can zoom in to a point in the graph by holding down the **Ctrl** key (Win) **Option** key (Mac) and clicking; add the **Shift** key to this combination to zoom out.
- In many graphs you can also zoom into (or out from) a region of the graph by holding down the **Ctrl** key (Win) **Option** key (Mac) and dragging in the graph to show the region you want. Adding the **Shift** key zooms out.
- In graphs that contain points, you can double-click on a point to bring up the inspector for that case.
- In most graphs, as you move the mouse over portions of the graph, the status bar will display information relevant to that portion of the graph.

Summary Tables

Like graphs, summary tables help you find out things about your data. Also like graphs, the form a summary table takes depends on what kind of attributes you put in its rows and columns. Categorical attributes

break things down into groups. With continuous attributes, you can quickly compute statistics.

Here are some examples of summary tables in various configurations.



When you add a continuous attribute to a summary table, Fathom adds the formula mean( ). When you add a categorical attribute, Fathom appends the formula count( ). You can edit these formulas, and you can add any number of additional formulas.

- Formulas such as mean( ) that do not specify an attribute have the attribute filled in as the appropriate continuous attribute for the cell. At times you will want to be explicit about the attribute, especially when there are no continuous attributes.
- Using the **Add Basic Statistics** command in the **Summary** menu adds count( ), mean( ), median( ), stdDev( ), and iqr( ) to the table.
- Dragging an attribute on top of an attribute already in the summary table replaces that attribute.
- Dragging an attribute on top of one of the gray arrows adds that attribute to the appropriate dimension of the table.
- See "Add Attributes to a Summary Table" on page 55 for rules about attribute placement in a summary table.
- The context menu for a summary table provides a shortcut for adding formulas and deleting the selected attribute.
- See "Collect Statistics from a Summary Table into a New Collection" on page 63 for information about how to perform computations based on computed summary statistics.

You can also connect a collection to an empty summary table by dragging the collection's name into the summary table. By default, you

get the count of cases in the collection. You can then add other formulas (or edit the default formula).

Finally, you can also use an empty formula to perform calculations on numbers. Add a formula and enter the values and operators you need into it. When you OK the formula editor, the formula appears below, and the result appears in the cell.

Analyses

Analysis objects in Fathom provide standard statistical tests and parameter estimation. They are explained in "How to Estimate Population Parameters" starting on page 71 and in "How to Test Hypotheses" starting on page 81.

Sliders

Sliders are explained in "How to Use Sliders as Model Parameters" starting on page 50.

Text

Text objects are useful for documenting your work. You can make small text objects to serve as captions, or you can make long text objects that contain a complete report.

There is no control over formatting text in Fathom. You can paste formatted text from another application into a Fathom text object. For more information see "How to Write a Report and Print It" on page 94.

Fathom Menus

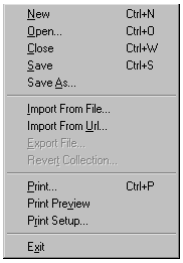
Here is a list of Fathom's menus; each is described in detail on the following pages.

File	Controls opening, saving, and printing files.
Edit	Controls the clipboard, undoing, renaming, and preferences.
Display	Controls how various items appear on the screen.
Insert	Inserts objects into your Fathom document.
Data	Controls new cases, randomization, sorting, and inspecting.
Analyze	Performs analyses including sampling, simulation, and statistical inference.
Window	Chooses among the current windows.
Help	Launches your Web browser to display the help system.
Graph	Controls graphs. (Available only when a graph is selected.)
Summary	Controls summary tables. (Available only when a summary table is selected.)
Estimate	Works with statistical estimate objects. (Available only when an estimate object is selected.)
Test	Enhances statistical tests. (Available only when a test object is selected.)

Contextual (Right-Button) Menus

It's not always easy to tell which menu you need. So Fathom also provides context or shortcut menus. On Windows computers, click the button on the right side of the mouse; on Macintosh computers, hold down the **Control** key and click to invoke these menus. This brings up a menu of items that may be relevant to that object—collected from many of the menu-bar menus.

For example, if you show the context menu for a graph, you'll get many items from the **Graph** menu and also items from the **Data** menu, such as **New Cases...**.



File Menu

The **File** menu controls opening, saving, and printing files.

New — Brings up a new, blank Fathom document.

Open — Opens an existing Fathom document.

Import From File... — Imports data from a text file. You specify the name of the file in the dialog box that appears.

Import From Url... — Imports data from the Internet. You type in or paste the URL, and Fathom imports the data.

Export File... — Exports a collection to a tab-delimited text file. You must select a collection to make this active.

Revert Collection — Restores the data in the collection to the way it was when you last saved. Note that this is different from simply reopening the document: If you have made graphs or other analyses, they are not affected.

You must select a *collection* to make this active. (If you have entered or imported data without saving, this command will not be available.)

Close — Closes the current document. This is the same as clicking the close box on the window. If the document has not been saved, Fathom will ask whether you want to save it.

Save — Saves the document. You give it a name. If it has already been saved, Fathom updates the saved file to the current state.

Save As — Saves the document, but give it a new name (thus preserving the previously saved file).

Print — Prints your document. This item brings up a standard **Print** dialog box where you specify how many copies you want, etc.

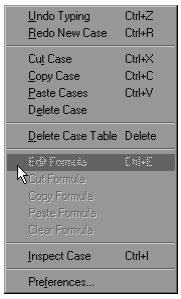
Page Setup — Chooses a printer, page size, etc.

Print Preview — Shows what your document will look like printed. This is especially useful if you're worried about where the edge of the page is. (Available only in Windows.)

Exit/Quit — Quits the program.

Edit Menu

For example, if a case is selected, you might see:



or, if an attribute is selected, you might see:

The **Edit** menu controls the clipboard, undoing, renaming, and preferences. The specific names of the items change to reflect what is selected.

Undo — Undoes the last action you took. The item changes to indicate what kind of action it was (such as **Typing**.) Fathom has unlimited undo.

Redo — Redoes the last action you undid.

Cut or **Cut Case** or **Cut Attribute** — Cuts whatever is selected and puts it on the clipboard. Fathom supports cutting text and cases. (So you can cut cases from one collection and paste them into another.) You can cut cases from collections, case tables, or graphs.

Copy Picture — Puts a picture of whatever you have selected onto the clipboard. This is great for taking snapshots of graphs and putting them into word-processing documents.

Copy — Copies whatever you have selected. This is just like **Cut**, except that the thing you selected remains where it was in addition to being placed on the clipboard.

Paste — Puts whatever is on the clipboard into the selected location. Exactly what happens depends on the situation.

In general, if the clipboard contains text, **Paste** replaces selected text or inserts the text at the insertion point.

If the clipboard contains cases, **Paste** adds cases to a selected collection.

Delete Case or **Delete Attribute** — Deletes the selected case or attribute (but not the object).

For example, if you have selected cases in a graph, this command deletes those cases. If you have selected an attribute in a case table, this deletes the attribute.

Delete (object) — Deletes the selected object. This is how you eliminate unwanted tables and graphs. Select one and choose this item—or press **Delete** (Win) **⌘-D** (Mac)—to get rid of it.

Select All — When a collection is selected, this selects all cases; when a text object is selected, it selects all text.

Edit Formula Invokes the formula editor. If an attribute or a formula (for example, a value plotted in a graph or a calculation in a summary table) is selected, this item lets you edit that formula. This command is equivalent to double-clicking in the formula box.

Cut Formula Cuts a formula from the selected object (attribute, calculation, or plot), if one exists.

Copy Formula Copies a formula from the selected object, if one exists.

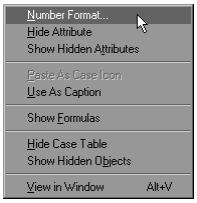
Paste Formula Pastes the formula in the clipboard into the selected object (attribute or plot). This is perfect for avoiding retyping a complex formula: copy it from a function, say, and paste it into an attribute.

Clear Formula Removes the formula from the selected plot or attribute. If it is a plot or calculation, that plot or calculation is removed. If it is an attribute, the formula is removed, leaving editable data.

Inspect Opens an inspector for the selected case or collection. If, say, a graph is selected, this will inspect the collection whose data the graph displays.

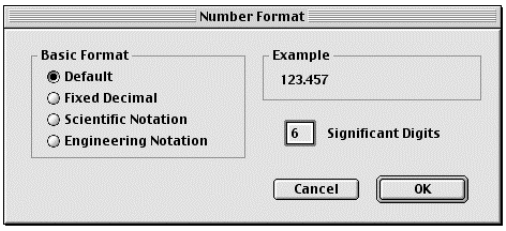
Preferences Invokes Fathom's preferences dialog box, which currently lets you set the program to display in large fonts (or not), and to turn sound off (or on).

Display Menu



The **Display** menu controls how various items appear on the screen.

Number Format... Controls how the selected attributes' numbers are displayed (for example, how many decimal places show). See "Change Number Formats" on page 35, for a detailed explanation and examples.



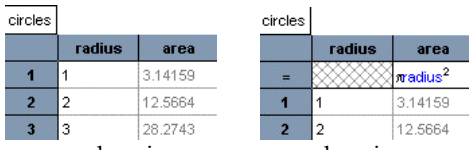
Hide Attribute If an attribute is selected in the case table, hides the attribute without eliminating it (great for reducing clutter).

Show Hidden Attributes Makes any hidden attributes reappear.

Paste As Case Icon Uses the clipboard graphic for the icon of the selected cases in the open collection.

Use As Caption Uses the selected attribute(s) as captions for cases.

Show/Hide Formulas Shows (or hides) the formula row in the selected case table. The formula row displays formulas for attributes. Double-click in a formula to edit it.



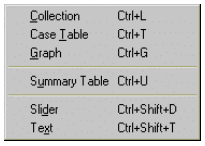
not showing showing

Hide (object) Hides the selected object.

Show Hidden Objects Makes hidden objects reappear.

View in Window Makes a separate window for the currently selected object. This is great for making really big graphs.

Insert Menu



The **Insert** menu puts objects into your Fathom document.

Note: You can invoke these commands by dragging icons off the shelf.

Collection Puts a new, empty collection in your document.

Case Table Puts a new, empty case table in your document. As soon as you add an attribute, Fathom creates a collection as well. If you select a collection before using this command, you get a case table filled with the collection's data.

Graph Puts a new, empty graph in your document.

Summary Table Puts an empty summary table in your document.

Slider Puts a new, empty slider in your document.

Text Puts a new, empty text object in your document.

Type into it to put (printable) text on your page.

Data Menu

The **Data** menu controls new cases, filters, sorting, and renaming.

New Case(s)... Adds new cases to your collection. You specify how many.

New Attribute Adds a new case attribute to your collection. If an attribute is selected in a case table, the new attribute is placed to the left of the one selected.

Add/Remove Filter Adds or removes a filter to the active object.

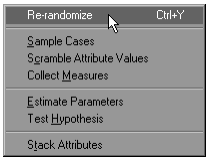
Lock/Unlock Collection Makes it so you can't change the data. You can still analyze it, but you can't edit it or make new attributes. **Unlock** reverses this action.

Sort Ascending/Descending Sorts the selected attribute in the chosen order.

Rename Attribute Changes an attribute's name (same as double-clicking its name).

Rename Collection Changes a collection's name (same as double-clicking its name).

Analyze Menu



Use the **Analyze** menu to perform Fathom analyses including sampling, simulation, and statistical inference.

Rerandomize Rerandomizes any values determined by random numbers in the active object. If no object is active, Fathom rerandomizes for the entire document.

Shortcut: **Ctrl-Y (**Win**) ⌘-Y (**Mac)

Sample More Cases
Scramble Again
Collect More Measures

If the active object is a derived collection, one of these items replaces **Rerandomize**. Choosing this option updates the derived collection (sample, scrambled, measure, or stacked) according to the settings on the corresponding pane in its inspector. For example, if the collection is a

sample, this command collects a new sample.

Shortcut: **Ctrl-Y (**Win**) ⌘-Y (**Mac)

Sample Cases Creates a new collection consisting of a sample from the selected collection.

Scramble Attribute Values Creates a new collection with the same attributes as the selected collection, except that one of the attributes' values is scrambled.

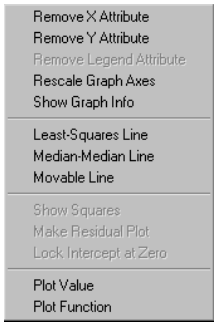
Collect Measures Creates a new collection made from the measures of the selected collection.

Estimate Parameters Creates an analysis object that will calculate confidence intervals.

Test Hypothesis Creates an analysis object that will perform hypothesis tests.

Stack Attributes Creates a new object that will convert multiple attributes into two by "stacking" the cases on top of one another.

Graph Menu



The **Graph** menu only appears when a graph is currently active (that is, it has a border).

Note: Relevant items from this menu also appear in the context menu for the graph.

Remove X Attribute Removes the attribute on the horizontal axis.

Remove Y Attribute Removes the attribute on the vertical axis.

Remove Legend Attribute Removes the attribute that has been dragged to the middle of the graph (it makes different-shaped points or shaded areas in bar charts, for example).

Rescale Graph Axes Returns the graph axes to their defaults (encompassing all of the points). This is equivalent to rechoosing the kind of plot (for example, **Box Plot**) from the popup menu in the graph.

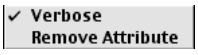
Show Graph Info Brings up a Control Text object that allows you to rescale the graph (and bin widths) by typing.

Stack Dots	In a dot plot, toggles between stack and unstack (dot plots appear as stacked by default).
Least-Squares Line	In a scatter plot, places (or removes) a least-squares linear regression line.
Median-Median Line	In a scatter plot, places (or removes) a median-median line.
Movable Line	In a box plot, histogram, dot plot, ntigram, or scatter plot, places (or removes) a movable line.
Show Squares	Constructs vertical segments from all plotted functions (including lines) to the data points and then build squares on them. Reports the sum of squares (the total area).
Make Residual Plot	Makes a plot of the residuals from the selected function. If more than one function is plotted, you have to select one first by clicking on it once.
Lock Intercept at Zero	Ensures that the movable line or least-squares line has a zero intercept.
Plot Value	Plots a single value on a graph.
Plot Function	Plots an arbitrary function on the graph.

Summary Menu

This menu controls summary tables. It appears only when a summary table is active (that is, it has a border).

Add Formula	Adds a new formula to the list of quantities computed for each cell.
Remove Attribute	Selects an attribute and chooses this item; the attribute then goes away.
Add Basic Statistics	Adds formulas for count, mean, median, standard deviation, and interquartile range to the table.



Estimate Menu

This menu works with statistics estimation objects. It appears only when an estimate is active.

Verbose Toggles the verbose mode off or on to switch between Fathom's default wordy mode, and more standard-looking statistical output.

Remove Attribute Removes the selected attribute from the estimate.

Test Menu

This menu enhances statistical tests. It appears only when a test object is active (that is, it has a border).

Verbose Toggles the verbose mode off or on to switch between Fathom's default wordy mode, and more standard-looking statistical output.

Remove Attribute Removes the selected attribute from the estimate.

Show Test Statistic Distribution/Show p_hat Distribution Makes a graph of the theoretical distribution of the relevant statistic for your test, with a shaded portion showing the relevant region.

Window Menu

Use the **Window** menu to choose among the current windows.

Cascade Lays out the windows so that you can see all their title bars. (Available in Windows only.)

(names) All of the windows are listed here. Chooses the one you want to bring to the top.

Help Menu

Use the **Help** menu to invoke **Help**.

Fathom Help Invokes the help system. This launches your browser software and uses it to display **Fathom Help**.

About Fathom Here you can find out what version you have and who made it. (On the Mac, you find it in the Apple menu.)

Error Messages for Formulas

Despite your best efforts, you will probably commit formula errors. Here is a list of what they are and what they mean.

#Name not recognized#	You have referred to a symbol that Fathom doesn't recognize. For example, you may have misspelled the name of an attribute.
#Type(s) incompatible#	You have given a function an argument of the wrong type. For example, sin("hello") gives this error.
#Argument count error#	You've given a function the wrong number of arguments. For example, even(3, 4, 5) gives this error.
#Domain error#	You've tried to evaluate a function outside its domain. For example, percentile(110, height) returns this error because percentiles must be between 0 and 100.
#Circular reference#	You've used the attribute you're defining in the formula (perhaps indirectly, through another attribute).
#Format error#	Something is wrong with the way the formula is written. For example, x < ? will produce this error.
#Evaluation error#	This is a catchall used when Fathom cannot otherwise explain why the formula did not produce a result.

Index

#

A

B

C

D

E

F

G

H

I

K

L

M

N

S