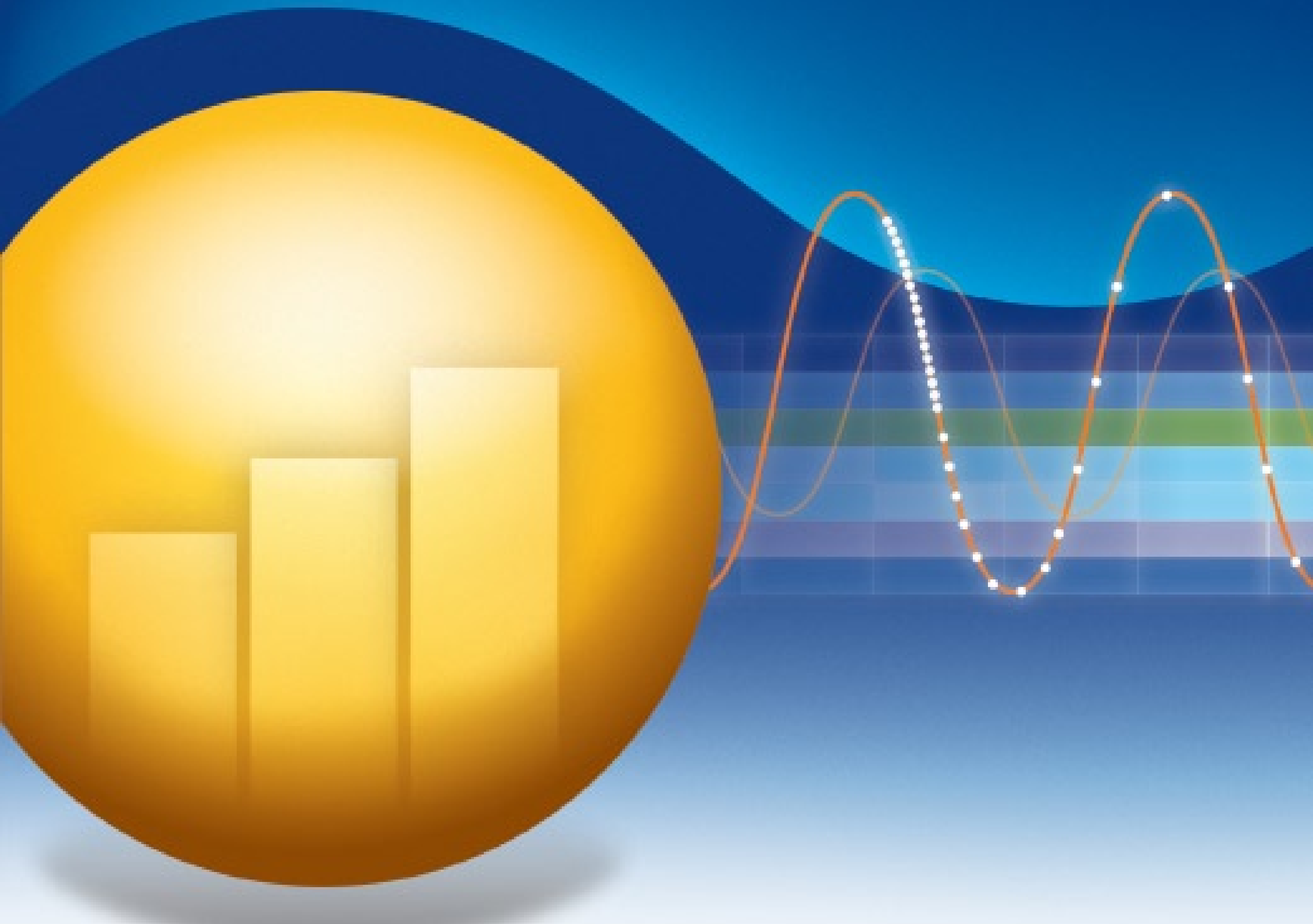
Data Are Everywhere

VERSION 2

PROJECT IDEAS FOR Fathom™

Dynamic Data™ Software



Writers: William Finzer, Stu Schwartz

Project Editor: Heather Dever

Editorial Assistant: Aneesa Davenport

Software Development: KCP Technologies

Production Director: McKinley Williams

Production Editor: Kristin Ferraioli

Copyeditor: Tara Joffe

Production Coordinators: Ann Rothenbuhler, Jennifer Young

Text Designer: Charice Silverman

Compositor: Jennifer Young

Art Editor: Jason Luz

Cover Designers: Keith Nguyen, Marilyn Perry

Photo Credits: p. 7, ©Royalty-Free/CORBIS; p. 22, Satellite image of Hurricane Jeanne courtesy of NASA Earth Observatory http://earthobservatory.nasa.gov; page 16, ©CORBIS SYGMA

Prepress and Printer: Alonzo Printing

Executive Editor: Casey FitzSimons

Publisher: Steven Rasmussen

This material is based upon work supported by the National Science Foundation under award numbers III–9400091 and DMI–0131833. 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.

The Concord Consortium
25 Love Lane
Concord, MA 01742
fathom@concord.org
concord.org

10 9 8 7 6 5 4 3 13 12 11 10 09 08 ISBN 978-1-55953-686-8

Contents

Introduction 2

Guidelines for All Projects 2

Questions to Consider About Data 3

Tips for Collecting Data 3

Projects 4

Everyday Life 4

Games and Sports 9

Social Science 11

United States Census Microdata 14

Mathematics 16

Statistics 20

Science and Nature 21

Language, Art, and Entertainment 26

Introduction

In this book, you'll find projects you can do using Fathom. They come from many different areas, including science, the arts, and everyday life. For some projects you'll collect data yourself, for others you'll do research, and for some you'll build a simulation. Many of the projects only contain questions to get you started. You should follow the project guidelines below, especially if you're doing a project for school. If your project involves real-world data, be sure to address the Questions to Consider About Data. If you're collecting your own data, see the Tips for Collecting Data.

Most of the projects in this book can be done by students from many levels of study, from algebra to statistics and beyond. If you're doing a project for school, your instructor will expect different things depending on your level of study. Before you begin any project, make sure you understand what it involves. It is up to you to decide what project is right for you (when in doubt, ask your instructor).

A few projects in this book are only appropriate if you're taking statistics. These projects are Horoscope Reality Checking, Streaks, Home Field Advantage, and all the Statistics projects. Additionally, some projects, such as Hailstone Numbers, require knowledge of measures. You can read about measures in Fathom Help.

Don't be limited by the ideas in this book! Feel free to make up your own project. Wherever there are data, you can analyze them using Fathom.

GUIDELINES FOR ALL PROJECTS

Part of doing a project is thinking about what you've done and explaining your project to other people. Here are some project guidelines. They may not all apply to the project you're doing.

- Explain your work, using a text object in your Fathom file, or elsewhere. If you collect data yourself, explain your method in as much detail as possible.
- If you found data or other facts somewhere else, be sure to credit your sources. Most data are not copyrightable, but be sure to observe any copyright warnings and get permission if you plan to publish your analysis.
- Be as specific as possible in your analysis. Use any methods you've learned in your mathematics or statistics classes. You should use your data to back up any conclusions you draw. Be especially careful when making inferences based on your data.

- Think about what additional questions you could explore. (You don't need to explore them in this project, but of course you can.)
- If your project involves building a model, think critically about your model. Could it have been better? What assumptions are you making? Is there a specific range of values for which the model is valid? Are you simplifying anything to make the model possible? It's okay to do that, as long as you explain what you did.
- If you publish your project in some way, such as put it on your Web site, be aware of legislation regarding the privacy of people involved in your project. In particular, the Children's Online Privacy Protection Act of 1998 governs how information about people under the age of 13 is used or disseminated.

QUESTIONS TO CONSIDER ABOUT DATA

If your project involves real-world data (which it probably does), think about these questions. Try to answer as many of them as apply.

- Is the source reliable?
- Is your sample representative?
- If there is a strong correlation, does this necessarily imply causation? What other factors might contribute to the correlation?

TIPS FOR COLLECTING DATA

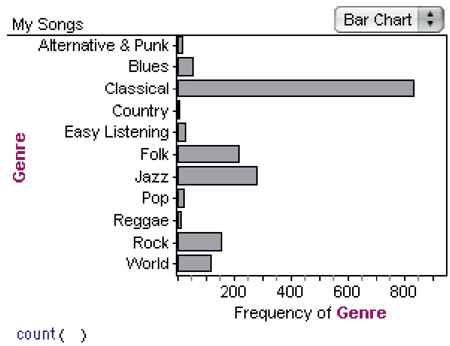
If your project involves collecting your own data, here are some things to keep in mind.

- Think about what attributes you'll need before you start: If you're collecting data over time, remember to record the time (in days or whatever time unit is appropriate).
- Don't wait until you're through collecting data to begin your analysis—you'll probably discover other attributes you want to look at.
- Keep detailed and accurate records.
- Try to make sure your data aren't biased in some way. If bias is unavoidable, be sure to note that in your analysis.

Projects

EVERYDAY LIFE

If you look around, you'll see your life is full of data. How can you use these data to find out more about yourself and the world around you? Here are some ideas.



Your Song Catalog What music do you listen to? What music do you own? What is your taste in music? What kinds of music have you heard little but would like to hear more? How does your taste in music change over time? Of course, you can answer these questions subjectively, but it can be extremely interesting to add data into the mix. There are many ways to gather the information, and, of course, what information you gather will depend on what you want to know. If you have your music stored electronically, find out whether you can export your song list as a text file, then import the text file into Fathom. (The bar chart shown here came from iTunes data.) Write a description of who you are as portrayed by your music data. Get someone else to do this project, too, and compare your musical portraits.

Money Flow Fathom wasn't designed to manage your budget—there are many other computer software packages for that. But Fathom does do a good job of keeping a simple record of money flowing in and out. Where *does* your money go? What does your balance look like over time: steady, cyclic, exponentially increasing or decreasing? Can you predict what you'll have at the end of the month or the end of the year? You'll probably want each case in this collection to be an expense or a deposit. One attribute should record the item's category, for example, Food, Entertainment, or Allowance. Figure out how best to record the date. You'll probably want to have a *DayOfTheYear* attribute or something similar so you can make time-series graphs. What do you learn from looking at the flow of your money? What kind of a saver and spender are you? Are you happy with what you see, or are there things you would like to change?

Contacts What do you know about your family and friends? Make a collection of the people you know, with the goal of finding out something about them as a group and about yourself as the person they all have in common. In addition to "address book" attributes, such as *PhoneNumber* (the area code should be a separate attribute) and *State,* include things like *Relationship* (friend, teacher, parent's friend, relative, and so on), *BirthYear, YearFirstMet, FrequencyOfContact, Ethnicity, PoliticalPersuasion,* and so on. Some attributes will definitely require a subjective response from you, and that's okay. You might also have a set of yes-no attributes, such as *LikesSports, TalksALot,* and *Introspective.* Decide how you'll code things you don't know about a person versus things that don't apply. Include yourself—how do you fit in or stick out?

Don't wait until you're done entering data before beginning the analysis, because the analysis will suggest other attributes you might like to include. What do the data say about you? Were there surprises? Are there kinds of people you don't know that you would like to know? Get someone you know to do this project, too, and compare results. Keep your collection growing over the years and track the changes.

One Day at a Time Are there things about yourself or what you do that you would like to track over time to see if and how they change? Possibilities: How far you walk, time on the telephone, health problems, mood, wake-up time, quarrels, and hugs. Pick at least two or three things (picking only one probably won't be very interesting), and code them as attributes in a collection. Make each case a day. Begin your analysis after you have a week or so of data, so you can tell whether you're going to learn anything interesting. Look for trends or cycles. If there are things you are hoping to change, do you see signs of success in the data? Are there any attributes that correlate with each other?

Exercise Trends and Cycles If you do some kind of regular exercise, keep track of whatever you can about it. Possibilities include running time and distance, swimming laps, weight lifted, and number of sit-ups or push-ups. Also, record things that interest you about your body's response to the exercise—blood pressure, weight, amount of time you can hold your breath, or whatever else you can measure. It might be appropriate to record something about the environment, too, like the temperature or humidity. Include a comment attribute so you can record special occurrences, like "stayed up very late last night" or "feeling especially good!" In addition to recording the date, also record the day number, counting from the first case, so you can make time-series graphs. You may have some personal goals, such as a certain time in which to sprint 100 meters. Plot these goals on the time-series graphs so you can mark your progress toward them (or away from them). Describe what the data tell you and how that makes you feel about yourself.

Mood Swings Do your moods change from day to day (or hour to hour), or are you as steady as a rock? No matter where you fall on the spectrum, you can learn things about yourself by making a collection in which cases represent instants in time and the attributes are various "feeling dimensions"—happiness, confidence, anger, love, irritation, and so on. It's a bit like keeping a diary, but more quantitative.

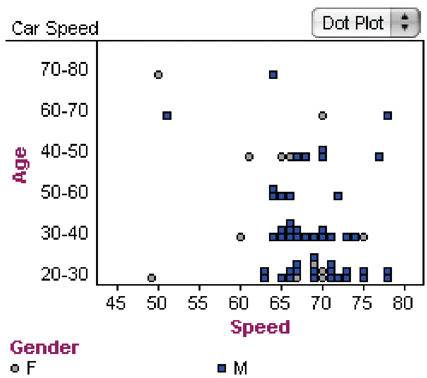
Choose a scale to use. An even number of possible values is recommended because it will force you to commit to one side or the other of the middle. An example scale is "the complete opposite, hardly any, quite a bit, completely." Although you are free to invent different scales for each emotional dimension, there is some advantage to using the same scale for all dimensions. Fathom's category sets work well for this purpose, and the *indexOfCategory* function can come in handy for the analysis. Of course, you'll want to record the date and day number from the start of recording. You may also want to record the time so that you can record your feelings more than once per day. You can use a comment attribute to say something about events that accompanied your feelings.

Make graphs of mood versus time. (Using the *indexOfCategory* function to get numeric values will make this easier than trying to use the categorical values themselves.) Do certain emotions vary together? In opposition? Are there things that surprise you from looking at the data? What generalizations can you make about yourself?

Nutrition It's fairly easy (given sufficient determination) to record the *things* you eat each day. But how do you record what you eat in a way that you can analyze? What would you like to know about what you eat—calories, nutritional information, food groups, percentage of Recommended Dietary Allowances (RDA)? What questions do you want to be able to answer? For example, am I eating a healthy diet? How does my caloric intake look over the long haul, and does it correlate with my mood? My weight? Experiment with different methods of setting up attributes for this one; your first try won't be as satisfying as your fifth.

Horoscope Reality Checking Do you have any faith in horoscopes—daily predictions based on your astrological sun sign? See if you can come up with an experimental method for determining the accuracy of horoscopes. Here's a simple example. Each day, at the end of the day, get a friend or family member to read your horoscope and the horoscope for a different sign to you without telling you which is which. (He or she should probably use coin flips to determine which other sign to pick and which horoscope to read first.) You decide which horoscope more accurately describes your day and then record that sign. After a fair number of days of this, you can use Fathom's Test Proportion statistical tests to determine whether chance alone can explain the results. This is just one example—you can think of many other methods and get others to join you in collecting data.

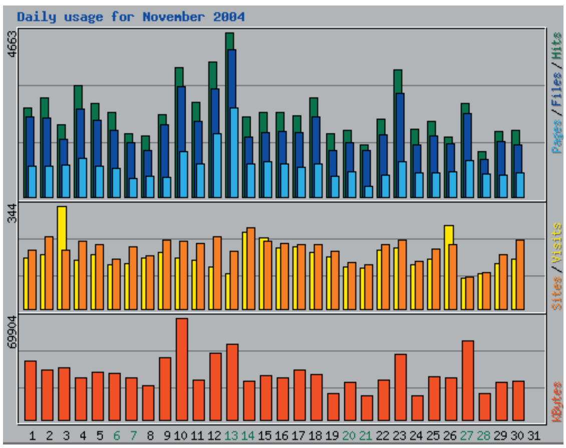
Counting Cars There are many possible projects you can do by counting cars or, more generally, measuring something about cars. For example, answer questions like, What are the relative frequencies of colors, makes, models, or styles of cars on the road near my home or school? At a busy intersection? On the freeway between here and somewhere else? How fast do cars drive on the street at some specific location? If I drive at the speed limit on the freeway, how many cars per minute pass me, and how many cars per minute do I pass? How do the cars parked on one street compare with the cars parked on another? You get the idea.



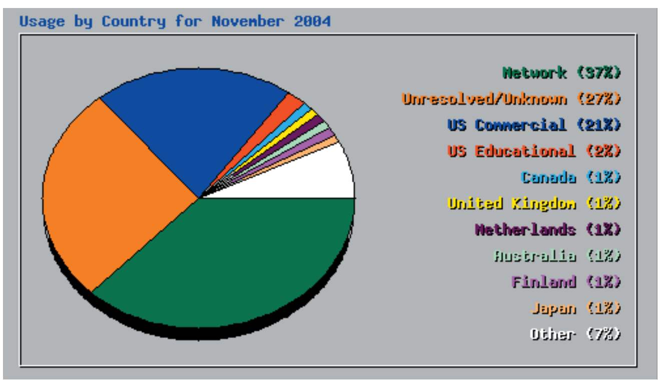
Start with a question that really interests you, and then decide how you're going to gather the data to answer it. Then do a pilot of your method by gathering a relatively small amount of data and doing an analysis with that. *Before* you actually invest the time to gather a lot of data, revise and pilot again until you're pretty sure your method is going to work. If you come up with something interesting, get someone else to attempt to replicate your result. Then write it up as an article for the local or school newspaper.

Collection Mania If you collect something, analyze your collection using Fathom. For example, if you collect coins, your attributes might be *Year, Country, Condition*, and *Value*. What patterns do you find in your collection? Are any of the items in your collection outliers? If so, why?

Who Visits Your Web Site? If you have a Web site, your Internet Service Provider probably provides statistics about how many people visit your Web site each month, how they got there, and where they're from. Choose several statistics, and keep track of them over time. How do your visitor profiles change? When you change your site, does that appear to affect the number of visitors? How quickly is a change in your site picked up by search engines? What search terms lead visitors to your site? What are you best known for, and who knows about it?



Number of pages and files used and hits for one Web site



Percentage of use by country for one Web site

GAMES AND SPORTS

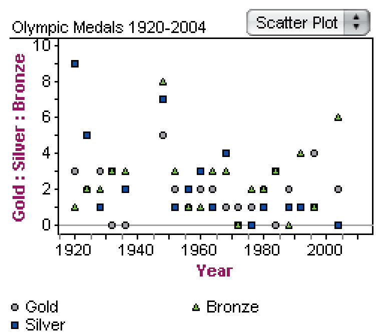
People have been collecting data about sports for at least a hundred years. In fact, the study of probability started with the study of games. Return to the roots of data analysis.

Streaks There has long been controversy about whether athletes have periods of time when they are *on,* and other periods of time when they are *off.* Does the baseball batter have hitting streaks; the basketball player, free throw streaks; the soccer player, shooting streaks? Some people (especially fans) swear that there are streaks, and others (especially statisticians) claim that such occurrences are merely random fluctuations. What do you think? Choose a sport and some aspect of that sport you think might exhibit "streakiness." Gather as much past data as possible. Use Fathom to simulate the same number of events (times at bat, kicks at the goal, and so on) as you have data for. In your simulation, the chance of success (getting a hit, making a goal, and so on) should be the overall success rate exhibited by the data you have gathered.

Now decide on a method for *measuring* the data's streakiness. It has to be a quantitative measure that Fathom can calculate for the simulated data. Get Fathom to collect this measure each time you run the simulation. The result will be a distribution of streakiness for the simulation. You can use this distribution to decide how likely it is that the streakiness in the actual data would come about by chance. To learn more, see Fathom Help, especially the section on collecting measures.

Predicting the Gold, Silver, and Bronze

Can you predict how many Olympic medals a country will win in any given year? Could you predict this based on characteristics of the countries, like population, gross national product (GNP), and state sponsorship of athletes? What other attributes do you think might be relevant? You can find the medal data from past Olympics in **Summer Olympic Medals.ftm,** a sample document that comes with Fathom. Gather the relevant information about each participating country for the year or years you have chosen. Build your model for making the prediction, and then apply it to a different year to test how well it works.



Denmark's Olympic medal counts

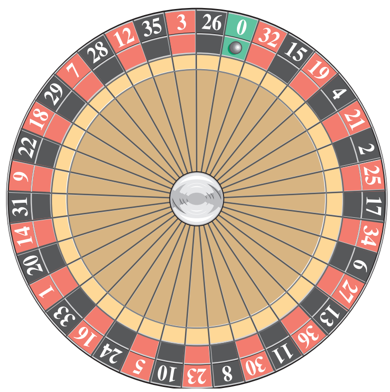
The Game of Hog Hog is played with up to 15 dice. You pick as many dice as you want and roll them. If none of the dice are 1, you score the sum of the dice for that turn. If any one of the dice is 1, you get 0 points for that turn. You win when you have 100 points. Simulate the game.

Start by assuming that you always choose to roll 3 dice. How many rolls will you need, on average, to get 100 points? Then change the number of dice that you choose to 10. How many rolls will you now need to accumulate 100 points? Create a slider that represents the number of dice you choose. How many rolls will you need, on average, to get 100 points? Which strategy seems to be the best?

Home Field Advantage Is there really a home field advantage in sports? Choose a professional or college-level sport that has large crowds at games. Using the Internet, find the number of times, *n*, the home team won in a recent season. Then use Fathom to simulate 100 repetitions of the number of games in your sport's season where the home team wins 50% of the time (there is no home field advantage). Finally, find the percentage of your trials where the home team won at least *n* times. Based on this result, would you conclude that there is or is not a home field advantage in that sport? Explain.

Roulette and the Problem of the Bank Account

Suppose you play roulette by betting on either red or black. Your first bet is $1. Any time you lose, you double your bet for the next round. Any time you win, you start over again with a $1 bet. When your total winnings reach $100, you stop. Game over. It sounds foolproof, doesn't it? Construct a Fathom simulation to play the game using this strategy. For each game, keep track of the number of rounds it took to get to the $100 goal and how much money you had to have in your bank account to keep placing bets.



Consider the effect of the places where the house always wins. How does this affect the size of the bank account you need? Vary the simulation in other ways by, for example, changing the betting strategy. What recommendations would you make to gamblers based on your study?

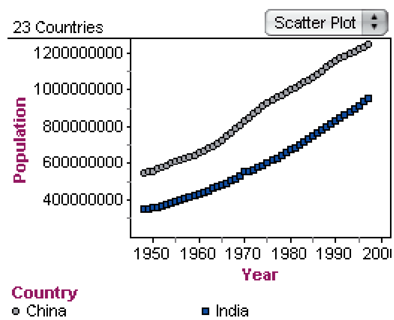
SOCIAL SCIENCE

Many data exist about the societies of the world and the people who live in them. You can use these data to find out more about the world.

Magazine Ads Do some types of magazines have more ads than others? Are the various ads in one type of magazine usually for similar products? Analyze the number of ad pages versus the number of content pages in at least 20 different magazines. What trends do you notice? How strong are the trends? Why might one type of magazine have more ads than another type?

Spam Nearly everyone who has an e-mail address receives anywhere from a trickle to a torrent of noxious spam. Make it the object of your study. Devise a classification scheme for the content. Record where it's from. Get your e-mail program to show you the "long header," which includes lots of information about the journey of e-mail to your inbox. After you get a feel for what there is in the spam, decide what you want to know. Do you want to know what spam says about our culture? Are you interested in the change in spam over time? Or are you looking for clues about what should be done about spam?

Country Populations Compare the countries of the world. Included with Fathom's sample documents are several that provide data about the countries of the world, including population data. You'll find them in **Sample Documents | Social Science | World**. These data came from the United Nations Statistics Division. (You'll probably encounter the names of a *lot* of countries you've never heard of before. Get out your atlas!) Here are some questions to get you started: How much do the ratios of men to women vary from one country to another? Which countries are growing fastest? Which regions of the world are growing fastest? Which are growing slowest? Which countries contribute most to world population growth? What explains population growth differences? Do family-planning programs influence population growth? How do levels of education for men and women in a country correlate with population growth? Once you start looking at the data, you'll probably think of many other questions you can explore. Do additional research to find the answers if necessary.

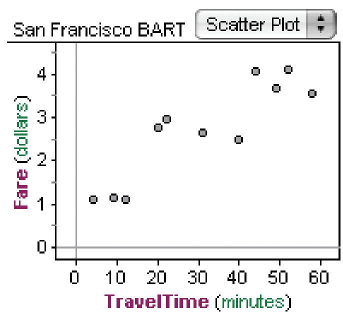


Populations of China and India

The Apportionment Problem In some national government bodies that have a representative form of government, each state is assigned a number of delegates proportional to the population of that state, or, the number of parliament seats are assigned to a particular party proportional to the number of votes the party received. In both cases, the total number of delegates stays constant. Research apportionment methods, such as Hamilton, Jefferson, Adam, Webster, and Huntingdon-Hill. Investigate past or current data for things like the population of each state in the United States. Then use Fathom to simulate each method. Analyze the methods. Which methods seem more or less fair? Which give citizens more or less power?

The Stock Market Is there a relationship between the Dow Jones Industrial Average and the NASDAQ (National Association of Securities Dealers Automated Quotation System)? Are there relationships among American markets, the European market, and the Japanese market? Investigate these indicators over the latter half of the twentieth century, import the data into Fathom, and see if there are any relationships.

Riding on the Metro Many people take some form of mass transit every day. But how much do they know about the train they ride? Pick a transit system you'd like to learn more about. It could be air travel, your local bus system, or the trains in your country. You could also pick a transit system in some place where you'd like to travel. Then think of what you'd like to know about this transit system. For instance, how have airline ticket prices changed over the past few years? What factors determine bus prices? Does our suburban rail adequately cover the population in the various towns it serves? Research the transit system, import the data into Fathom, and report on your findings. Do you have any recommendations to the body that governs the transit system or to the traveling public?

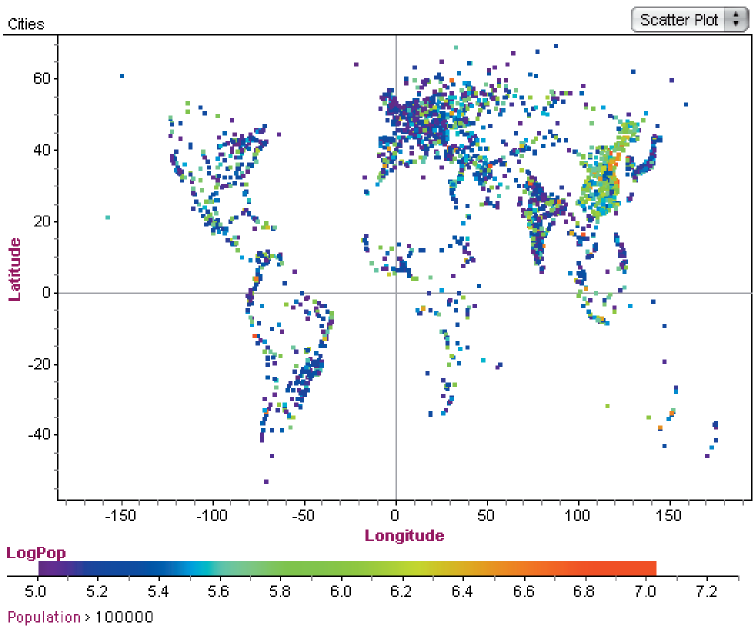


What's in a Name? In **Sample Documents | Social Science | United States**, you'll find a file called **Names.ftm.** This file contains a collection of more than 6,000 first and last names drawn from the 1990 U.S. Census and ranked from most to least frequent. (You may wish to research names from another country or another time. You could, for example, use Fathom's census microdata import capability to get names of people from 1850 to 1920.) Analyze the collection. Can you find your first and last names? What patterns do you find? Why are there more female first names than male first names? What do you find if you attempt to classify the names in some way, say by country of origin?

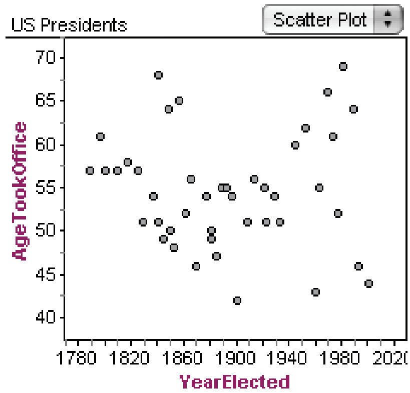
On the Internet, find lists of popular baby names, and compare these names with the names drawn from the general population in 1990. Using the *leftString* function, it's easy (and surprising) to get the distribution of first letters for the names. Can you figure out how to use the frequency of occurrence to get the distribution of first letters in the actual population?

Finally, create a simulation to help you estimate the probability that two randomly chosen people have the same name.

Cities of the World **Cities of the World.ftm** in **Sample Documents | Social Science | World** is a collection of more than 3,000 cities—country capitals and cities with a population over 100,000. Because the collection has attributes for latitude and longitude, it is easy to make a map in Fathom and color it by population. What clustering of cities do you see, and how do you explain it? Examine the distribution of populations and explain its shape. What patterns do you see in the distribution of cities by latitude? What questions do you have as you look at these data, and how can you go about answering them?



Leaders of the Country What can you predict about the next leader of your country? Collect data on the past leaders. Try to find anything you think might be helpful. Some possible predictors are age when taking office, gender, profession, birthplace, and religious affiliation. Analyze the data using Fathom, and write a profile of the next leader of your country. Be as quantitative as possible, using the data to support your claims. You might consider researching a different country to see if the same predictors apply.



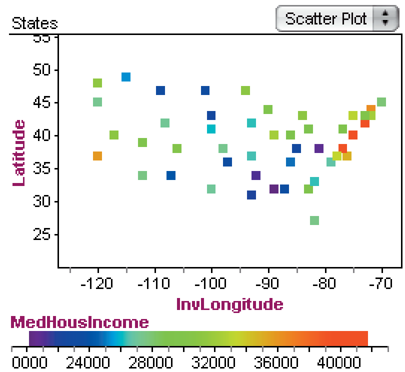
UNITED STATES CENSUS MICRODATA

Fathom can easily provide you with data about individual Americans from 1850 to 2000 who filled out the long form of the U.S. census. These data provide an endless source of data-sleuthing projects. Discoveries you make with these data might well be brand new—until fairly recently, it wasn't possible for many people to get access to the data or to use tools, such as Fathom, that make the work manageable.

The Demographic History of a Place You can get census microdata information from as early as 1850 for a state or metropolitan area (although the data for many metropolitan areas don't go back that far). Choose a place, a time period, and a question you want to investigate. The question will almost certainly evolve as you discover what it is possible to learn from the data. Limiting yourself to only two census years will give you room to compare those years in many ways. However, choosing a long range of years will let you look for trends and correlations. As you generate conclusions, you'll be settling on a methodology for analyzing the data. Chances are, this same methodology will work for a different place or time. Try it! Was it much easier to reach your conclusions for the second place or time than it was for the first? Write about *how* your methodology works and see if you can get others to try it. If they choose a place and time you have also worked on, do they draw the same conclusions?

The Difference a Place Makes How is one place different from another? How does the place where you live compare with a place you've thought about living? Use Fathom's ability to import census microdata to do a comparison. What places do you want to compare, and will you compare two states or two metropolitan areas? Will you attempt to compare them along as many dimensions as possible, or will you concentrate on particular kinds of differences? Some things you might compare are costs of living, travel time to work, and how long people have lived there. Before you start, be sure to write down your conjectures about what you will find. Develop a methodology for comparing places, as described in the project The Demographic History of a Place. Try your methodology on a different pair of places. Does it work as well?

Order the States It's easy enough to find an ordering of the states in the United States from biggest to smallest area or from greatest to least population. But what about their order for median income, percentage of immigrants, or ethnic diversity? Which states have the largest and smallest family sizes, or the youngest and oldest people? Choose your own set of questions. Start with something that you can get explicit microdata about, but know that soon you'll want to order by less direct measures. Start with the default sample size of 500, and when you have figured out *how* to do things, increase the sample size to 5,000 or even 10,000. If you live in the United States, find a ranking or two that puts your home state at one extreme or the other. How "stable" are your rankings—if you get a new sample, are the states in the same order? Come up with a quality of life ranking. (*Hint:* You can use **StatesMap.ftm** in **Sample Documents | United States | States** to color maps with your rankings.)



Diversity Fathom's census microdata include several attributes that provide insight into the diversity of the people who live in a specific place: birthplace, citizenship status, ancestry, and race are obvious ones. But income and property value attributes also allow you to investigate economic diversity. How do you quantify diversity using census microdata? Once you have a reasonable measure, compare some places. Where are the most diverse and least diverse places in the country? Study the change in diversity over time for a single place. Write a piece for your school or community newspaper about the diversity of your chosen area.

Mobility How likely are Americans to stay put throughout their lives? This information is not directly encoded in the census microdata, but certain attributes, particularly birthplace and years in current residence, can help measure mobility. Pick a place, and study mobility over time. Has it steadily increased, or was there a period of increased mobility followed by stability? By looking at multiple places, can you detect westward migration? Is it still happening? Where are people moving to now? What places have the *least* mobility?

Steam Fitters and Computer Techs How has people's work changed over time? The attributes *Occupation* and *Industry,* combined with some judicious filtering, allow you to study the changes. (Both attributes have been recoded to 1950 standards to lessen the effect of changing job titles.) What jobs do people no longer do? When did certain jobs first appear? Has the value that society puts on certain jobs changed over the years? How about the percentages of women doing a job?

History Detective Pick a fact you've heard about American citizens (perhaps in history class or in the news). Use the census microdata to research the accuracy of this fact. Some examples are: Many African American people moved from the South to the North in the mid-twentieth century; families have gotten smaller over the years; or women earn less than men earn. Explore your fact as thoroughly as possible. Explain what the microdata do or do not show and how you know.

MATHEMATICS

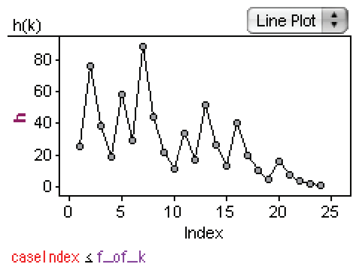
In mathematics, it can be very difficult to figure out an abstract solution to a problem. Using Fathom, you can experiment with many possible cases of the problem to see what the solution might look like. You can also use Fathom to simulate abstract or other problems and to analyze the data you generate.

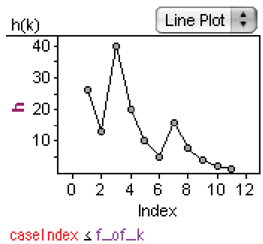
Symmetry Many things in the world, including people, have bilateral symmetry. (In other words, the left side of your body is a mirror image of the right side.) But how perfectly symmetrical are these things? Investigate bilateral symmetry. You can either compare many people (How much higher is your left eye than your right eye? How much bigger is your right foot than your left foot?), or you can compare many different objects. Either way, look at how perfectly symmetrical they are or are not. Develop a measure of symmetry. Does the symmetry depend on the size of the object? The type of the object?

Hailstone Numbers Some functions behave nicely and simply: You graph them, and you get a smooth curve. The $3k + 1$ function is different because its value at any point depends on the number of steps in a recursive sequence: If k is any positive integer and

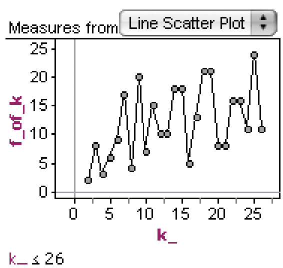
$$h(k) = \begin{cases} 1 & \text{if } k = 1 \\ h\left(\frac{k}{2}\right) & \text{if } k \text{ is even} \\ h(3k+1) & \text{if } k \text{ is odd} \end{cases}$$

then $f(k) = n$, where n is the number of integers to get to a final answer in $h(k)$. That answer is always 1 (unproven!). For example, $f(5) = 6$ because $h(5) = h(16) = h(8) = h(4) = h(2) = h(1) = 1$. Experiment with some numbers and chart their path as they move toward 1. Why should $k = 27$ take so many steps and $k = 29$ comparatively few? Model this function in Fathom, and analyze it in whatever way makes sense to you. You could graph lines of fit for patterns you see in the numbers and try to explain those patterns. You might keep track of the maximum value of $h(k)$ and analyze those patterns too. No one knows why this function acts the way it does. If you figure something out, let us know!

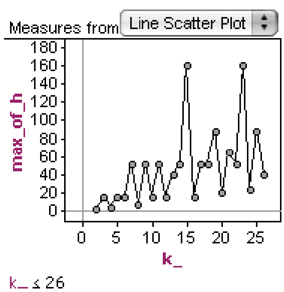




The hailstone sequence, $h(k)$, for $k = 25$ (left) and $k = 26$ (right)



The $3k + 1$ function for $k \leq 26$



Maximum values of $h(k)$ for $k \leq 26$

Johnson Solids The *Johnson solids* are all the convex polyhedra that have regular faces and equal edge lengths. (*Convex* means having no indentations.) They do not include the completely regular Platonic solids, the "semiregular" Archimedean solids, or the two infinite families of prisms and antiprisms. There are exactly 92 of them. (At the Math World Web site, you can see images and nets of the solids.) In **Sample Documents | Mathematics | Geometry**, you will find a file called **JohnsonSolids.ftm.** Use this file to investigate the Johnson solids. Here are some questions you might consider: How are the numbers of polygons related to the numbers of faces, edges, or vertices? How are the surface areas of the polyhedra related to the numbers of faces, edges, or vertices? What do the different parts of the names mean? Try filtering for different prefixes and suffixes and looking for patterns. Do research to find out if you are right. Add data for the Platonic and Archimedean solids. Do they conform to the patterns you found for the Johnson solids?

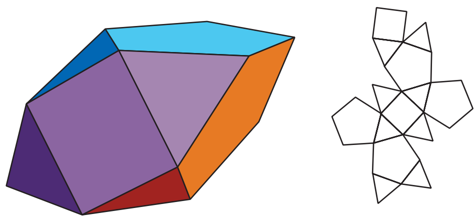


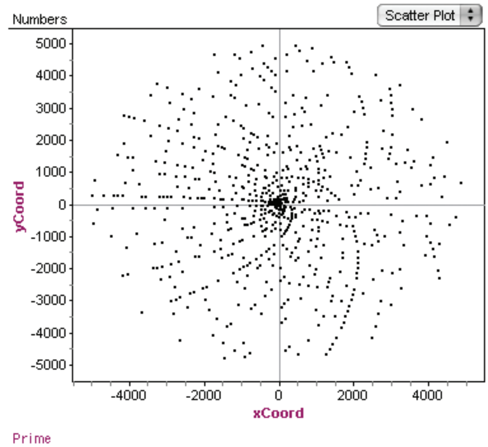
Image and net of Johnson solid 91, the bilunabirotunda

Family Planning Construct a Fathom simulation to model this situation:

> Suppose every family continues having children until they have a girl. They then stop having children. What is the effect of this practice on gender balance—in the long run, would there be more boys, more girls, or the same number of each?

Use your simulation to determine the answer, but don't stop there! Modify the problem parameters by, for example, changing the implicit assumption that the probability of giving birth to a girl is the same as that of giving birth to a boy or by limiting the family size. How do your modifications affect the outcome? Come up with a completely different family-planning scenario, and attempt to model it. Think about what kinds of simulations are easy to do in Fathom and what kinds are difficult.

Prime Patterns Many famous mathematicians have spent years pondering the prime numbers and the patterns they contain. Fathom has a function, *isPrime*, that tells you whether an integer is prime or not. Use this function to explore the prime numbers. For example, you probably know that the number of primes is infinite. (You might even know how to prove this.) But what do you know about how spread out prime numbers get? Are there about the same number of primes around one million as there are around one billion? Or you might want to research *prime spirals*. These patterns are made by arranging the positive integers in a spiral and then removing all but the primes. The example above uses polar coordinates, but you can also use a grid. There are many patterns in the prime numbers. Perhaps you can find more!



The Prisoner's Dilemma This famous problem is one that mathematicians and psychologists have used to study enlightened self-interest—when is it in your best interests to cooperate with someone else, and when is it not? Here is the situation.

> Two people are picked up by the police for committing a crime. They are interrogated in separate cells without the opportunity to communicate with each other. For the purpose of this problem, it makes no difference whether or not either of them actually committed the crime. They are both told the same thing:
>
> - If they both confess, they each get eight years in prison.
> - If neither of them confesses, the police will be able to pin part of the crime on them, and they'll each get four years.
> - If only one of the people confesses, the confessor will make a deal with the police and will go free, but the other person gets ten years in prison.

Simulate the dilemma in Fathom with each person deciding randomly what to do. Then create sliders to represent the proportion of times both of you will adopt a strategy to confess. Determine the average prison times with these strategies. What do you think is the best strategy? Suppose you vary the prison times for each decision. How do the average prison times and strategies change?

STATISTICS

Statistics provides powerful tools for working with data to find patterns, test hypotheses, and make predictions and decisions. Typically, you use statistics to find out something about another content area. So there aren't many projects in this section. If you're a statistics student looking for project ideas, by all means look in this section, but look throughout this book, too.

Correlation Versus Causation Think of a relationship that someone claims involves causation but that you think might only involve correlation. Your claim can be about anything—science, popular culture, sociology—but it must be something that can be tested. First, research data related to the claim, and determine whether or not the data seem to show a relationship between the two variables. Then, think about whether or not one event really causes the other. What other factors might be involved? Might the data you found be misleading in some way? If possible, find data on any other factors to see how these data are related to your claim.

This project is adapted from *Discovering Advanced Algebra* by Jerald Murdock, Ellen Kamischke, and Eric Kamischke.

A Recipe Book As a statistics student, it can be difficult to see that the practice of statistics is much more unified than the chapter titles in your statistics book might indicate. If you have trouble seeing the forest for the trees, this project may be for you. Put together one or more Fathom documents that provide at least one example for each of the statistical "recipes" in a standard introductory statistics course. Find your own data for each example. The links to data listed at the Fathom Resource Center are a good place to start. (Choose **Links to Data** from the **Help** menu.) As you carry out the analyses and write up your results, think about what is the same for each—look for the pattern in the process. Write up what you find, and share it with a statistics teacher.

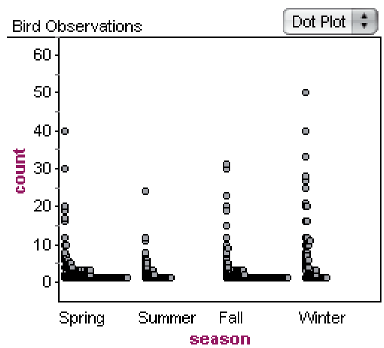
Can You Be Replaced? In any discussion of sampling, there's always a careful distinction made between sampling "with replacement" or "without replacement." Most sampling done in the real world is *without* replacement. (Imagine conducting a poll of voters and calling the same voter back again.) But most of the statistics you study in an introductory course is based on sampling *with* replacement. Use Fathom's ability to sample either with or without replacement to investigate when it really does matter which method you use. As an example, consider sampling 20 voters from a population of 100 to determine the proportion in favor of a proposition to require students to wear uniforms to school. Is your estimate likely to be more or less accurate using sampling *with* or *without* replacement? (A whole section of Fathom Help explains how to create simulations. That section is well worth reading before you start this project.)

SCIENCE AND NATURE

Scientists have been using data about the natural world for centuries to try to describe and explain the world we live in. You can, too.

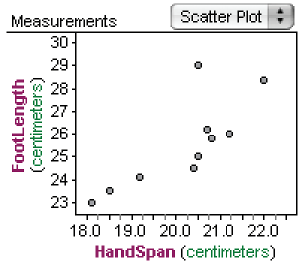
Bacterial Population Growth Consider the lowly bacterium. One way it reproduces is through fission. Where there was one, now there are two. You won't have trouble modeling that, will you? (Perhaps the model will consist of a collection in which each case is the number of bacteria at the end of the reproductive cycle, and a slider says how many bacteria to start with.) But a fundamental fact of nature is that all growth is limited—possibly because of scarcity of food or inability to get rid of waste. Model this fact. (You'll need more sliders.) Try to get a model that looks realistic, and then research what happens in the actual laboratory.

For the Birds If you have an active bird feeder where you live, then you can learn a lot about the natural world in your area. Spend five or ten minutes a day recording the activity on and around your bird feeder. How many different kinds of birds do you see? Do you see the same kinds every day? What about at different times of day—or at different times of the year? What are the most and least common species? What's the most you see of each kind? How often did a given kind appear as a percentage of the number of observation periods? (Use descriptions if you don't know their real names.) If you get hooked, you may want to investigate Project FeederWatch at Cornell University, which organizes bird feeder observation data from all over the United States. If you don't have access to a bird feeder, you can do the same research by taking regular walks in a neighborhood park.



Counts of species seen by one observer

Body Measurements Are any measurements of the human body correlated? For instance, if you knew someone's wrist circumference, could you predict her height? Choose at least eight measurements, such as hand span, foot length, length of little finger, height, arm span, length of lower leg (from knee to floor), and width of thumbnail. Take these measurements for as many people as possible. (If your entire class does this, you can pool the data.) Be as precise and consistent as you can in your measurements. Then analyze your data. Which measurements have the most spread? The least spread? Look for correlations between various measurements. Is any one measurement the best predictor of any other measurement? What pair of measurements appear to be the least correlated?

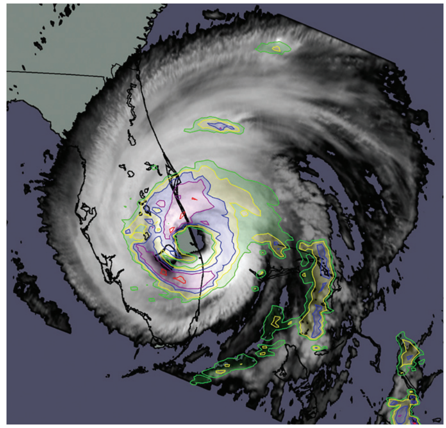


Bay of Fundy The Bay of Fundy in eastern Canada has the highest tidal bore in the world. This means the difference between the high tide and the low tide is greater than in any other place. Research the Bay of Fundy tides. Choose a month this year. Input the times of low and high tides and determine relationships between these times. Once you have done that, choose a coastal area in a different part of the world or facing a different direction, and input the times for the same month. Find relationships for this area. How much do the tides vary between the two places?

The Big One Can you predict where the next earthquake will occur? Research major and minor earthquakes that have occurred around the world in the past 20 years. Note their latitude, longitude, intensity, casualties, and any other attributes that seem interesting. Report any patterns you observe. (*Hint:* You can graph latitude and longitude to make a map of the quakes. You can also use a slider to replay the quakes over time.)

Sun Cycles The times of sunrise and sunset are cyclical—that is, they occur in cycles over a period of time. Choose a location and research the times of sunrise and sunset. Use Fathom to find a pattern. If you know trigonometry, fit a sinusoidal equation to the data, and use it to predict sunrise and sunset times for that area. When will the sun rise there on your birthday in the year 3000? How confident are you about this prediction?

Hurricane Season If you live near the ocean, you know that hurricanes can be a very real danger. Research hurricanes over the past 20 years. Is there any relationship among the location of the storm, the size of the storm, the barometric pressure, the storm surge, maximum wind speed, and the rainfall in inches? Can you predict how bad next season will be? Which coastal places in the world are the most likely to be hit by a hurricane? The least likely? Some scientists think hurricanes are increasing in number or intensity due to global warming. Do you find any evidence of this?



False Positives Some medical screening tests are routinely conducted on people who don't have specific symptoms. These include the skin test for tuberculosis and mammograms for breast cancer. A result is called a *false positive* if the test indicates the presence of the disease but the person doesn't have it. False positives are very common. Because of this, people who test positive are given additional, and more diagnostic, tests. But because there is no way to avoid false-positive readings, understanding what causes them and how likely they are can help reduce the anxiety of people who falsely test positive. In this project, you'll simulate a medical test, varying the rate of people who do have the disease and the test's accuracy rate. Analyze how the number of false positives changes as you vary these factors. Here is one scenario to get you started.

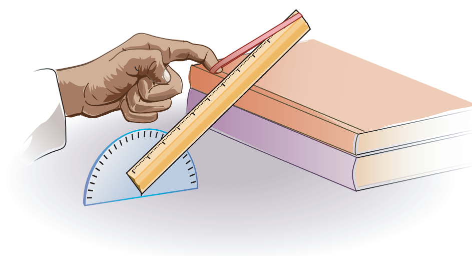
> A medical clinic tests 1,043 people to see whether they have a particular disease. About 5% of people in the general population have the disease. Based on this, the clinic expects that about 50 of the people tested will have the disease. The clinic is using a test that is 80% accurate. This means that about 8 of 10 people who have the disease will test positive, correctly indicating that they have the disease. It also means that about 8 of 10 people who don't have the disease will test negative, correctly indicating that they don't have the disease.

This project is adapted from a demonstration by Clifford Konold and is used with permission.

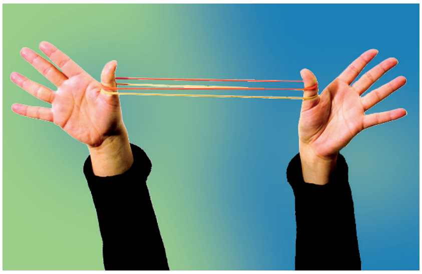
Life-Form Census Biologists believe, based on considerable evidence, that we're in the midst of a massive species extinction and that humankind is responsible. But before they can verify this, they need to know how many species of plants and animals there are. Try to help address this question. Pick a place, any place. (But if it's an indoor place, you'd better have access to a good microscope.) Your place could be a square meter of grass, a portion of a park, or a particular tree. Create a Fathom document with a collection that will contain a census of all the life-forms you are able to find there. Decide what you want to record about the life-forms based on questions you want to answer. Some possibilities include: location, date, time, description, name, size, plant or animal, and abundance. Don't be discouraged if it takes a while before you really start finding a lot of life-forms—they're there. They're everywhere!

The Watched Pot How long does it take a pot of water to boil? Does twice as much water take twice as long? As the water is heating, how does the temperature vary as a function of time? What about when the water is cooling? Does covering the pot while it's heating change the time to reach a boil? Do some pots boil the water faster? Does adding another substance, like salt, change the boiling time or temperature? Decide on the question(s) you want to investigate, and plan your observation. (Your planning should include safety precautions.) If you decide you need a thermometer, make sure it's one that can handle the temperature range you have planned for it (you may need to borrow one from school) and wear safety goggles. Keep careful records not only of what you do but also of your conjectures at various stages. In your analysis, be quantitative. Report findings such as the degrees of heating per second and the minutes per liter needed to reach a boil.

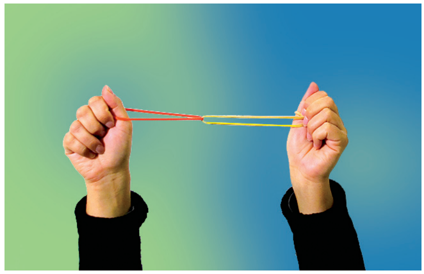
Rubber Band Launch A rubber band is launched into the air. What do you need to know to predict where it will land? If you needed to launch the rubber band so that it landed in a particular place, could you? To find out, you'll need a supply of rubber bands that are as identical as possible, unless you're determined to use the same rubber band over and over. You'll also need safety goggles and a reliable and controllable launching method, such as the one shown here. Don't collect very much data before you attempt some analysis, because you'll probably find that your first guess about how to record the data won't be your last.



Stretch! You pull on a rubber band, and it stretches. The harder you pull, the more it stretches. The "stronger" the rubber band, the less it stretches. But get quantitative. How much does it stretch? Is the relationship between force and distance stretched linear, that is, does pulling twice as hard cause twice the amount of stretch? What happens near the rubber band's breaking point? (This is why you'll need safety goggles!) Come up with a quantitative way to describe the strength of different rubber bands. What happens when you stretch two rubber bands in the two ways shown below? Write something that could be quoted in a catalog, so that when you place an order for rubber bands, you really know what you are getting.



Rubber bands stretched in "parallel"

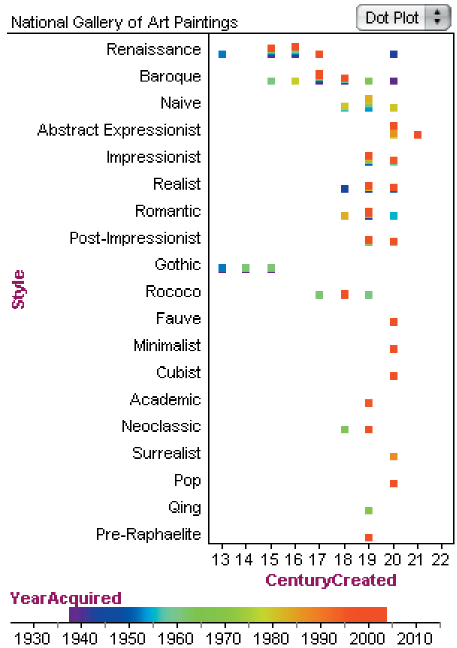


Rubber bands stretched in "series"

The Pendulum Swings What determines how long it takes for a pendulum to swing back and forth? (The time it takes is its *period*.) Don't let the fact that you may have already experimented with a pendulum in a math or science class dissuade you from doing this project. If you learned a rule, it may not apply to *all* pendulums. There's always more to learn about any phenomenon. Make a list of the factors that you think *might* influence a pendulum's period. Then investigate each factor *one at a time*. Think carefully about how you will vary one factor without also varying other factors. After you have developed a basic theory, you may want to test it by going to extremes—a very long, very heavy, or very whatever pendulum. Do you think your methods would work for tiny pendulums 1 millimeter long? What about for huge pendulums 1 kilometer long? Consider writing up your results as an article for a science journal. What aspects of your work should you share with the wider scientific community?

LANGUAGE, ART, AND ENTERTAINMENT

You may not think of data analysis as having much to do with art or entertainment, but you can use data to find out many things in these fields. You can use them to track changes in art styles, to study the lengths of various languages, and much more.



Art Critic The file **National Gallery Collection.ftm** in **Sample Documents | Language and the Arts** contains data from the complete collection of over 3,000 paintings at the National Gallery of Art in Washington, D.C. You could spend hours exploring these data, discovering artists you may never have heard of. (You can view many of the paintings at the National Gallery's Web site.) Analyze the collection in some way. Here are a few ideas. Track the emergence and disappearance over the centuries of the styles of paintings represented in the collection. Similarly, describe the rise and fall of popularity of various media. Based on the data, write a paragraph that gives an overview of the collection. Investigate the gallery's acquisition strategy in the twentieth century. For example, did it acquire work early or late in the life cycle of a style? Devise a scheme for classifying the content of some of the paintings based on their titles, and look at how the content changes over time. Characterize the work of the female artists in the collection as compared with that of the male artists. Compare the work of younger artists with that of older artists.

Gained in Translation Does the English language take more words to express similar thoughts than other languages? Find books that are written in English and then translated into another language. Choose random corresponding paragraphs of both books, and count the words. Enter your data into Fathom, and examine the relationship between paragraph lengths. You could also find several other languages the book has been translated into. Keep track of the lengths of paragraphs in all languages and attempt to generate correlations between paragraph lengths in English and in other languages.

Eloquence In **Sample Documents | Language and the Arts | Text Passages,** you'll find several documents that represent passages from books or speeches, letter by letter. Compare at least two of the documents by examining the word lengths. (You will need to generate several attributes to find the number of letters before and after a space.) What are the distributions of letter frequencies? Who seems to use the longer words? Longer sentences? Does either person use any words that are outliers? What effect might this have on readers of the books or listeners of the speeches? If you know statistics, use a slider to find the percentage of words for each person within *s* standard deviations.

Search and ReSearch Make Internet search engines a topic of research. How many results do particular words or phrases (your name, for example) get? How many results do these words or phrases get the following day or week? Keep track of the search phrase, date, and number of results. You might also try comparing different search engines. How do these numbers change with time? Look at **Googlewhacks.ftm** in **Sample Documents | Language and the Arts.** This collection contains 500 two-word phrases that produced a single result in a Google search at the time they were recorded as Googlewhacks. How many results do they give now?

Bacon Numbers How interconnected is the entertainment industry? You can find the *Bacon numbers* of various actors by connecting them to actor Kevin Bacon through movies or TV shows they've appeared in together. For example, Elvis Presley was in *Live a Little, Love a Little* with John Wheeler, and John Wheeler was in *Apollo 13* with Kevin Bacon. So Elvis Presley has a Bacon number of 2. A few Web sites will allow you to find "Bacon numbers" using another actor in place of Kevin Bacon. At these sites, you can link any actor to another. Choose an actor and find his or her "Bacon number" for as many other actors as you wish. Enter the data into Fathom, including the gender of each actor. What is the average "Bacon number" for your actor? Plot a histogram for the "Bacon numbers" for your actor. Do male or female actors seem to have higher or lower "Bacon numbers" for this actor? If you've ever met a famous actor, find out your own "Bacon number"!

Centuries of Artists In **Sample Documents | Language and the Arts**, you'll find **Artists.ftm**, which contains the names, birth dates, and death dates for over 1,000 artists. Use this document as a starting point for research. Pick a period, and gather more information about the artists in that period. How many works did they produce? Did they become well-known in their own lifetime? Did they belong to a "school" of art or even more than one such school? What media did they work in? Are they well-known today? Search for their names in an online bookstore to find out how many books come up. Investigate their causes of death, especially for those who had a very short life. What museums (especially nearby museums) could you visit to view their work? Write a paper about what you have discovered about the group of artists. Show your work to friends with the plan of getting them to do a similar project on a different period in art history.