Curriculum Innovations Based on Census Microdata: A Meeting of Statistics, Mathematics, and Social Science

William Finzer
KCP Technologies
Emeryville, CA, USA

Tim Erickson
Epistemological Engineering
Oakland, CA, USA

Abstract

Mathematics curriculum materials for early secondary school, based on interactive exploration of the huge database of individuals who filled out the "long" form for the US Census back through 1850, illustrate methods for improving statistical literacy and mathematics conceptual understanding while immersing students in social science content.

Introduction

Census microdata, analyzed with powerful tools, lets us address interesting and thorny curricular issues. Chief among these is how to prepare all students to engage in formal statistical thinking, or at least, for informed citizenship. Determining exactly what we mean by readiness is part of the problem. Our original naïve beliefs about the meaning of this have been challenged in the course of our curriculum-development work and our experiences with students, as well as by available data and tools. In this paper, we begin with an overview of our curriculum-development efforts, then we present some specific examples—interspersed with commentary—that describe the work we have been doing. Finally, we'll reflect back on the main themes discussed in the overview. In the examples, we illustrate ways to promote independent and open-ended investigation, tasks that give many points of access, new ways to make interdisciplinary connections, rich contexts for exploring more advanced topics, and approaches to concepts and skills we might not have realized that students lack. But first, let us set the stage by explaining why we are so excited about census microdata.

Let us begin with microdata. These are data about individual cases rather than predigested, statistics. Others such as Conti and Lombardo (2002), Frey (2003), and the ongoing CensusAtSchool project (see http://censusatschool.ntu.ac.uk/) have used microdata effectively to help students understand the importance of a national census or to give them meaningful data for analysis. To us, the most important thing about microdata is that students can do with them what they want; no one has pre-aggregated the data, possibly losing information that a student could use to explore some interesting question. Even though statistical conclusions are basically about groups and their distributions, the access to individual data both engages students and gives them more options.

The microdata at the heart of this paper are actual national census records about individuals. These are rich and fascinating; they hold student interest and are an inexhaustible source of ideas. We have been using US Census[1] data, although data from other countries are gradually becoming available. The US Census publishes many attributes about each individual ranging from age, race, and sex (e.g., a 23-year-old Black female) to what time she leaves for work in the morning, and whether she drives, bicycles, or takes the bus.[2] Thus the data are about people—not as an aggregate group, but as *individuals*. This contributes more than we had first imagined to students' interest and understanding.

How we obtain and analyze the data is nearly as important as the data themselves. Professional interactive graphical software has been increasingly available since the 1980's (e.g. Data Desk). Such systems makes data accessible and appealing to students. In our materials, students analyze the data using Fathom Dynamic Data™ software (Finzer, 2005), which is described elsewhere (Erickson and Finzer 1998; Finzer 2002), and will be referenced throughout this paper.

The new software capability central to this paper is an interface to census microdata[3] that lets us design activities in which students engage in a process that resembles the scientific method or, perhaps, a learning cycle (as described by Karplus & Their, 1967). The interface lets students specify the data they want: which attributes, which geographical regions, and, interestingly, which years. (The US data extend from 1850 to 2000.) The "cycle" part is important: when students decide that they want to alter their question or method, and that they need different data than they have, they can quickly return—transparently—to the Internet, to get new data. The newly obtained data replace the old, causing all existing tables and graphs to update.

Focus and Philosophy of the Curriculum

We believe that varied experiences with large and rich data sets help students develop data literacy. Learners form conjectures and construct responses to support or rebut them based on data that have many cases and/or many attributes. These responses naturally involve making displays, recoding values, dealing with small subgroups, and thinking about issues that are precursors to inference, such as variability as discussed in Hammerman (2003).

We are interested in exploring the ramifications of some of the curricular innovations made possible by the Fathom software and its seamless interface to census microdata. These ramifications include the ability to:

- Involve students with relatively large ($n \approx 5000$) data sets
- Emphasize exploration and conjecture
- Help learners deepen their understanding of how to summarize and display data through actually using the data to confirm or refute their conjectures
- Lead learners in the direction toward making quantitative statements derived from data and confront them with both the rewards and pitfalls of making such statements.

One question that we confronted when considering the value of census microdata in the curriculum relates to whether the size of these data sets help or hinder learners' ability to focus on trends or differences among groups. One thing we learned relates to the richness of census microdata, which, by presenting learners with a wealth of interesting things to explore, invites conjecture. As do many recent text authors such as Schaeffer (2003), we believe that curriculum materials should strengthen learners' resolve to make both qualitative and quantitative conjectures. As illustrated below, the drive to show that, for example, people are more mobile today than they were 50 years ago can, lead the learner to use and understand new kinds of plots, such as a ribbon chart. The combined use of Fathom and census microdata allows students to explore questions related to how much mobility has changed, or how to quantify the change in income or the difference in diversity between one region and another.

Background

The authors have been involved in the design of the software and the development of accompanying curriculum materials since 1993[4]. A book of classroom activities, *Data in Depth—Exploring Mathematics with Fathom* (Erickson, 2001) comes with the software, and the first chapter of activities in this book, "Come to Your Census," is based on exploration of US Census microdata. A subsequent grant from the National Science Foundation, "Census Microdata in the Mathematics Classroom," has helped us extend the software to include the census interface alluded to above and has supported the development of prototype classroom activities that use these data. The data themselves are served at the Minnesota Population Center (Ruggles, 2004).

Here are five attributes of the software we believe make Fathom especially well-suited for learners who are on the road towards understanding formal, inferential statistics—but who have not yet arrived. A goal of our curriculum development effort is to make appropriate use of these attributes.

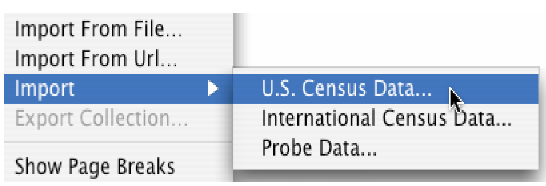
- *Multiple Representations*: It is easy and natural for learners to make multiple graphs, summary tables, data tables, and tests to represent the same data in different ways—and to have all of them visible and usable simultaneously. Thus students who are learning about these representations can "triangulate" the meaning using other representations, and can better learn to relate them. The challenge in the curriculum materials is to encourage students to use more than one representation without dictating which ones and when.
- *Always Current*: Any change in the data results in changes to all representations in which it is displayed. There is never a need to "update a graph." This extends to selection: cases selected in one representation are selected in all. The latter is surprisingly useful in exploratory data analysis, as a later example shows. The curriculum materials model the process of using this "always current" attribute of the software to make discoveries about the data, with the hope that students will add it to their repertoire of data analysis techniques.
- *A Case-centric View of Data:* Fathom helps the learner think about data as a collection of cases, for example, by supporting "drilling down" from representations of a data set to reveal the details of any particular case. Activities we are developing must not neglect the case-centric view of data and often need to return to the meaning of individual cases and possible problems attendant to data collection.
- *Ubiquity of Formulas*: To communicate more sophisticated requests to the program—such as filtering, plotting functions, or specifying new variables—learners often have to write formulas using symbolic expressions. This gives students needed practice making their ideas clear and unambiguous—and in understanding formulas. In writing activities, we need to be sure to allow plenty of room for learners to come up with their own algebraic solutions to a problem.
- *Easy Import*—In addition to the microdata interface, Fathom lets learners import data from a wide variety of sources, especially arbitrarily-formatted tables from the Internet. The value of this capability for census microdata curriculum materials is that we can allow for students to find and import supplemental data whose analysis can aid the overall investigation.

Description of the Curriculum

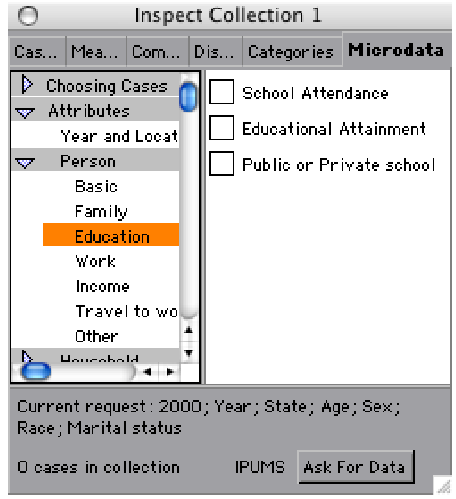
To design an effective curriculum, we need to know what knowledge and skills beginners, especially, need to develop to be successful later. In order to begin to determine this, let us look at examples of activities that use census microdata, and see how the software, the curriculum, the rich data, and learners interact.

Example: Income and Education

Here is an activity from Erickson and Holmberg (2005). Imagine that Anna and Bert are ninth-graders, aged 14 years, and have heard that they should study hard because they can get better jobs if they go to university. Is that really true? Is there evidence for this common assertion? With census data, they can find out.



Anna and Bert start up Fathom, and choose **Import>U.S. Census Data** from a menu. A Fathom *collection* (a data set) appears, and an inspector window opens. In this window, they specify what cases they want to collect and, for each case, what attributes. By default, they will get a sample of the whole United States, and a small set of attributes: the year of the data, state, age, sex, race, and marital status.



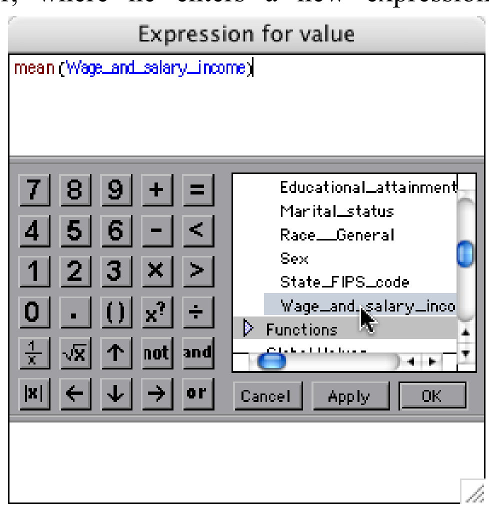
These attributes do not let them study income or education, so they navigate to the obvious heading in the inspector: **Education**. Clicking there, they see three possibilities. They choose "Educational Attainment." In a similar manner, they click on **Income**, and then on "Wage and Salary Income." The list in the box at the bottom of the Inspector shows what data they will be asking for.

They press the **Ask for data** button, and in a minute, their collection fills with data from 5000 individuals from across the United States. Now they can start to analyze the data. But where to start? They begin by making a graph, which first appears empty. We want students to learn to associate *attributes* with *axes*, so to make the graph, students must drag the name of one of the attributes (variables) to an axis of the graph. Anna decides to look at education first, so she grabs the name from the inspector and drops it on an axis of the graph (she picks the vertical axis).

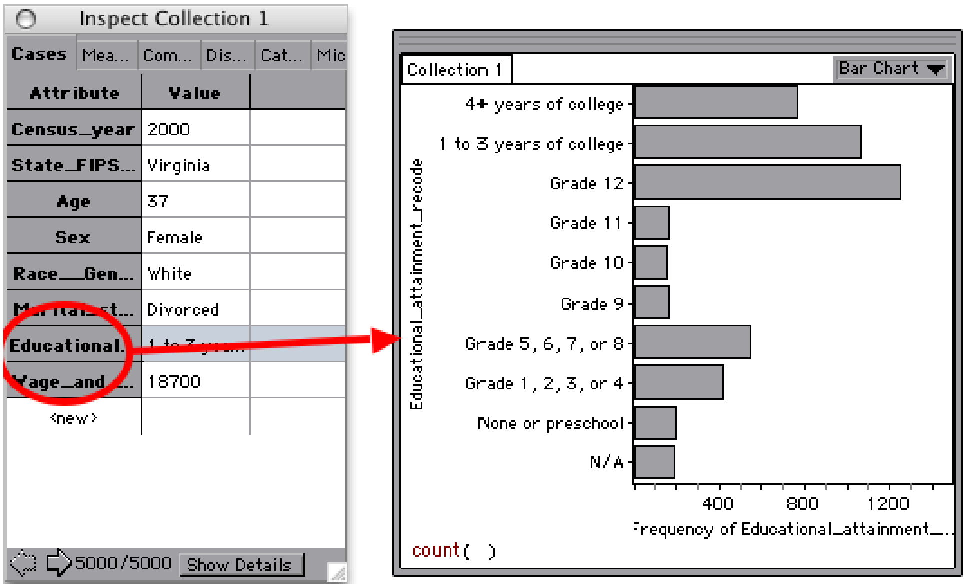
A bar chart appears. It shows the distribution of educational attainment among the 5000 people in the collection. At this point, we will make several observations:

- While the distinction between census and sample, and the details of what kind of sample you have, are crucial for inferential statistics, Anna and Bert do not—and should not—care. At this stage, they see the interface simply as a source of data.[5]
- Anna has great flexibility in how she proceeds. She could have put education on the horizontal axis—to get vertical bars—or even begun with income.
- Anna did not have to choose what kind of graph to make. Fathom noticed that the attribute was categorical, so made an appropriate display. That is, she got something useful in as few steps as possible.
- The categories of educational attainment are already recoded *and ordered*. This may seem simple, but it is not; and dealing with the original data without this help is hard.

Now Bert takes over the mouse to look at income. He notes that Anna's graph shows how many people are in each category—but it does not have to be that way. He double-clicks the expression **count( )** at the bottom of the graph to invoke the formula editor, where he enters a new expression, **mean(wage_and_salary_income)**. He does not have to type in that long attribute name; he gets it from a "browser" in the editor itself.

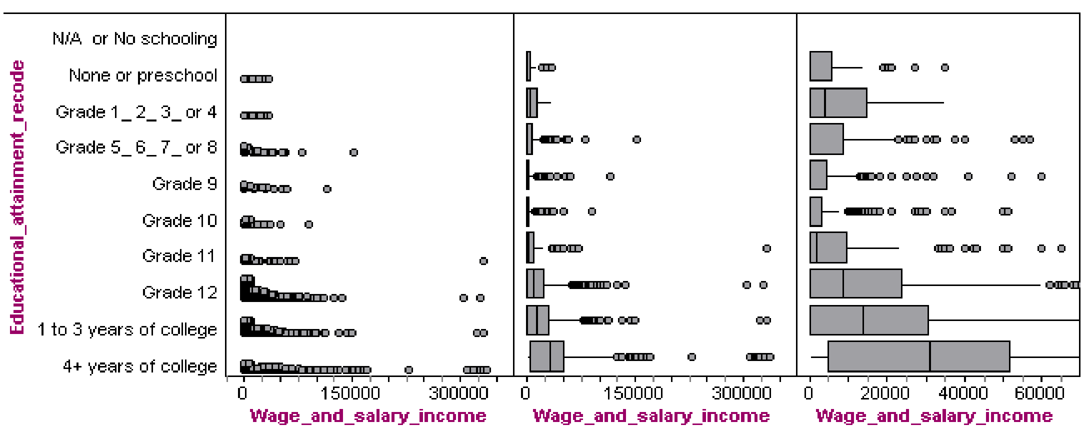


This changes the graph to show these means as bar lengths. Bert is satisfied; it looks as if there is a clear relationship between education and income. But here the teacher (or the worksheet) can help Bert and Anna look more deeply. The teacher might ask, "Does this mean that if you go to college, you're *sure* of making more money than if you don't?"



"Well, no," says Anna. "The bar is the average. Some people make more, some make less. Maybe there are some college people who make less than some high-school people." So the teacher challenges them to find some way to show that aspect of the situation.

So the students drag the **income** attribute to the horizontal axis of the graph. This produces a dense and busy split dot plot (below left), so they use the menu in the corner of the graph to change the dot plots to box plots (center), simplifying the display. Then, because the whiskers extend so far, and there are so many outliers, they rescale the axes to better see the progression of medians (right). This means that they forgo a view of the whole distributions, for a clearer view of the medians.



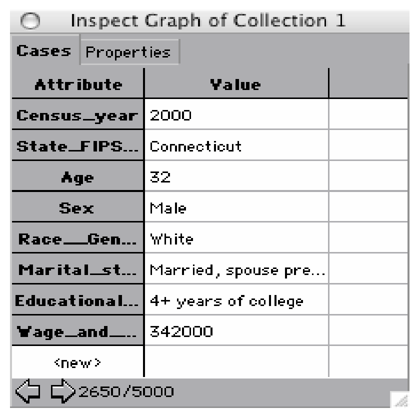
Now their report on the relationship between education and income can be much more nuanced. To be sure, on the average, the more education you have, the more money you will make. But there are plenty of college-educated people who are not earning very much; and there are a few who have never attended college who earn quite a bit. The students see that a fuller description of the situation involves looking at the *distributions*, and that the distributions overlap. Also—and important for their education as citizens—the students see that there are, overall, a small number of people earning a great deal of money, and a large number earning more modest—even paltry—amounts. Now some more comments:

- The process we have described is idealized. In a real classroom, Anna and Bert typically mess about with the data more, try things that do not work, and have more trouble making sense of things that do. Because of the software, however, they can do the messing about quickly. And because of the context—and the prior knowledge they bring to the data—they are better at wrestling with the confusing parts.

- In this situation, some areas of statistics education that give students trouble (such as sampling error) do not matter. Students are to make a display that produces an "interocular" effect—it hits you right between the eyes (Edwards, et. al. 1963). This is *pre-inferential* statistics; we claim that many students have trouble with inferential statistics later because they do not have enough experience with making these more obvious quantitative arguments and displays. The large number of cases means that students can see and compare distributions even when they split the data along some other variable.
- Finally—and perhaps most important—this activity invites students to be skeptical and to use data to study an assertion they generally believe to be true. Such activities are not only engaging to students (they present a challenge and they always hold the possibility of surprise), they are also natural bridges to social-science topics such as bias and stereotyping. What's more, such activities are impossible unless the learner has any reason to believe the original assertion, that is, they must be in areas where the learner has prior knowledge and experience.

We have omitted one of the most effective elements that the confluence of data and technology makes possible: looking at the individuals. We often begin with outliers. Who makes a lot of money? What do they do? We can ask such questions in written materials, or encourage the teacher to ask them, either as part of the main lesson, or during a debriefing.

Students have a number of strategies for exploring these outliers (as well as for getting more information about people "in the pack"). One of the simplest is to double-click one of the dots in a dot plot or box plot. The inspector changes to show information about that particular case. In our data set, this is a 32-year-old married white male from Connecticut.

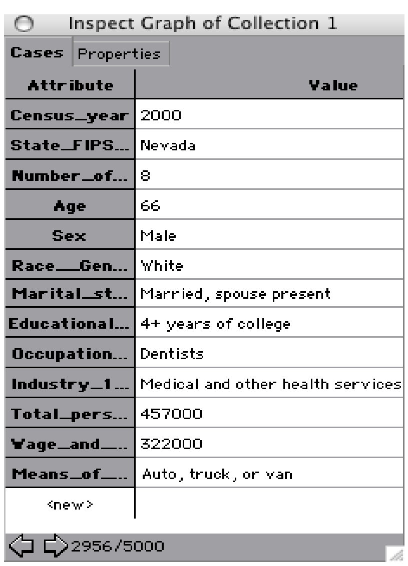

Inspect Graph of Collection 1

Cases | Properties

Attribute	Value
Census_year	2000
State_FIPS...	Connecticut
Age	32
Sex	Male
Race__Gen...	White
Marital_st...	Married, spouse pre...
Educational...	4+ years of college
Wage_and_...	342000
<new>	

2650/5000

But suppose Anna and Bert want to know more. For that, they need more data. They return to where they specified the original data to be downloaded. Now they choose, for example, **Occupation** and **Industry**, and then add other things of interest (such as the number of rooms in the house or apartment). When they click **Ask for data**, Fathom returns to the Minnesota Census server, and gets a new sample of 5000 people, replacing the old ones in the analysis. Now the biggest earner is a dentist in Nevada, married, living in a house with eight rooms, and commuting to work in an "Auto, truck, or van."


Inspect Graph of Collection 1

Cases | Properties

Attribute	Value
Census_year	2000
State_FIPS...	Nevada
Number_of...	8
Age	66
Sex	Male
Race__Gen...	White
Marital_st...	Married, spouse present
Educational...	4+ years of college
Occupation...	Dentists
Industry_1...	Medical and other health services
Total_pers...	457000
Wage_and_...	322000
Means_of_...	Auto, truck, or van
<new>	

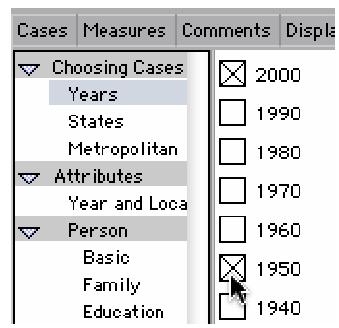
2956/5000

The ability to get details about individuals serves at least two purposes. First, it lends a human face to the data, which makes it intrinsically more interesting. Second, it gives us a chance to address a gap in students' understanding. Some students lose track of where data come from; they do not see that every aggregate graph or statistic actually arises from data about individuals. [6] When Anna and Bert look at **income**, we want them to realize how this particular dentist contributes to that distribution. To the student, a distribution should not be an abstract curve or histogram—easily detached from its meaning—but rather a summary of data about many actual people.

At this point, the astute reader may have ideas for extensions, or additional questions. Let us mention just two to give ideas of what is possible:

- We can use the data to create traditional statistics problem-solving opportunities. Suppose Anna and Bert remove the dentist from the data set. How does that affect the overall appearance of the distribution? How does it affect the median? The mean? More experienced students can calculate the change in mean before trying it out; but even less-experienced students can make qualitative predictions: it will decrease, but not by much.
- Studying those who have only a high-school education, our students notice that many have no job or only part-time employment. But this group includes many who are quite young—even children. Perhaps we should look only at older people, or people not currently in school, or perhaps only *employed* people when comparing incomes. In Fathom, Bert and Anna create a *filter*—a Boolean expression they enter using the formula editor—to specify which cases appear in the graph. For this, they could use an attribute they have already (**age**) or go back to the Internet to get a new one (e.g., **labor force status**, which tells whether the person is in the labor force, and if so, if he or she has a job).

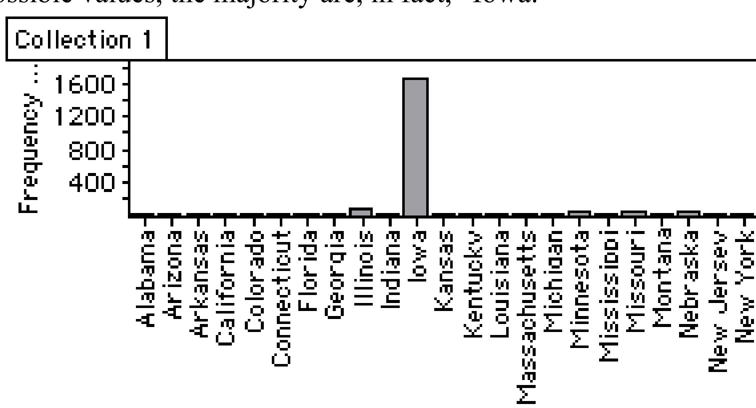
Example—Proportion and Ribbon Charts

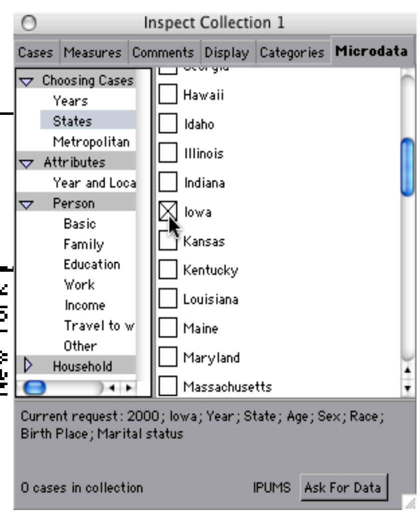


Another activity asks students to study whether people move more often today than they did fifty years ago. Our conjecture might be that they do: families used to be more stable, the country is becoming more mobile with time. Our data do not include the length of time since the last move, so we have to sift through the available attributes to find a suitable stand-in (bearing in mind from then on in what way our data are—and are not—related to the original question). There are many possibilities; one of them is to compare the states individuals were born in to the ones in which they now live.

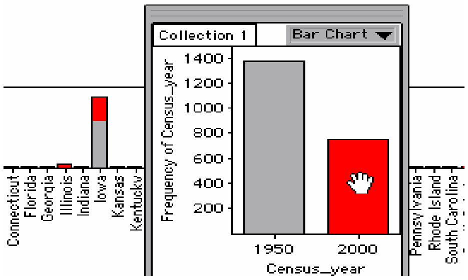
Anna and Bert decide to study their home state, Iowa. In the inspector where they specify the data, they choose **Birth Place** as an attribute, and also choose Iowa from a list of states. In another panel (**Years**) they choose 2000 and 1950.

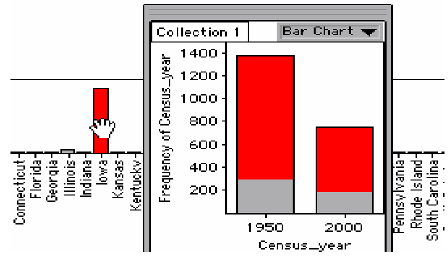
When they get the data (this time there are only 2126 cases since they're getting data only from Iowa), they make a graph of **Birthplace** to see what sorts of values it has. That creates a new graph, a portion of which is shown below. We can see that while there are many possible values, the majority are, in fact, "Iowa."





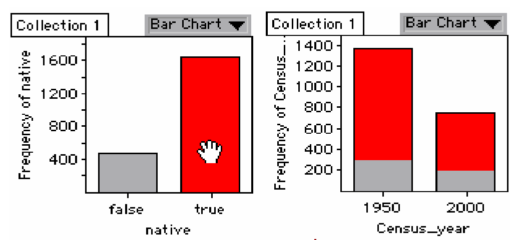
But how can they compare the mobilities from the different census years? First, Bert makes a graph of the census years themselves. He sees that there are more cases from 1950 than from 2000.[7] Clicking on the 2000 bar, he sees that part of the Iowa bar lights up. The students puzzle over what that means; eventually Anna clicks on the Iowa bar (in the other graph) and they realize this is what they are after: the proportion of those not selected—the "migrants"—looks smaller in 1950 than in 2000 (though the *number* of migrants in the sample is larger).





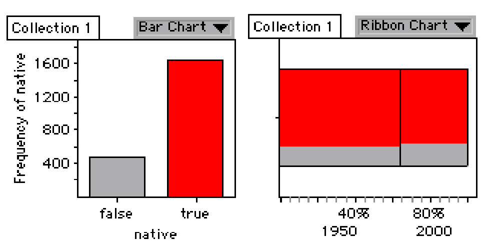
This situation holds in it central problems in statistics learning: the importance of both *proportion* and *conditions*. The latter arises as we consider the selected portions of the bars. Bert found the proportion of all natives who were in the 2000 sample; Anna found the proportion of those in the 2000 sample who are natives. Even we teachers, even in such a simple situation, must take a moment to make sure we're thinking the right way around, and using the right words in the right order. Is it any wonder our students get confused? Bert and Anna are confused, too, but they can click, and look, and click elsewhere, and talk about what they see. By addressing the issue themselves, using a powerful, easy-to-use tool, Anna and Bert are developing the experiences they will need to be comfortable (or at least not miserable) later on when they have to parse two-by-two tables and use conditional probability.

With regard to proportion, what the students see in Anna's graph is not completely convincing. Does the proportion of migrants really increase? They decide to go deeper, first *recoding* the **Birthplace** data to make its graph easier to deal with.



Anna creates a new attribute, called **native**, and gives it a formula, **Birthplace = "Iowa"**. Now the pair of graphs is simpler; she has lumped all those whose birthplace is not Iowa into a single category: **false**.

But Bert is not satisfied with the graph. The different sample sizes obscure the matter. So Bert changes the **Year** graph to a *ribbon chart*. There they see that the proportion of natives does decrease with time—but not by much. Other states give very different results—in both directions.



Note that the students could also make a table, and compute the change in percentage. But it is easy to look at the numbers and say, "78 to 75, it's decreasing"—and fail to grasp that the decrease is not very much. The graphic helps us assess whether the difference is meaningful, without resorting to a statistical test.

Collection 1	Summary Table		
	Census_year		Row Summary
	1950	2000	
	1372 78.571429	754 75.464191	2126 77.469426

S1 = count( )

S2 = count(native) / count( ) 100

Again, some comments:

- This is a place to make sure that students can connect the representation to the original question or conjecture. What does the graph *mean* in terms of the proportion of people moving to Iowa in 1950 and 2000? Are people more mobile now? Students need to stop and re-connect their work to the original questions or risk being lost in abstraction.
- Bert selected the bar in one graph to see a relationship in another. Notice how this lets students build up multivariate representations from univariate ones. This helps get us around a curricular dilemma: it makes sense to teach univariate plots (such as the bar chart or histogram) first, and

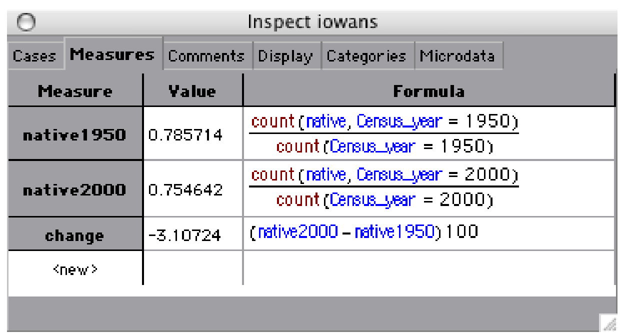
then move on to multivariate. But multivariate graphs are inherently more interesting because they show *relationships*. This selection feature essentially lets the student dynamically lay another variable's worth of information over the well-understood univariate plot.

- We find the ribbon chart to be a particularly effective and interesting graph, helping students focus on important issues in proportion. For example, in this configuration of the ribbon chart, if the two variables are independent, then the horizontal lines *meet at a point*. The size of the "break" is the difference of proportions, which is a measure of how much the two variables are associated.

Measures and Simulation

More experienced students could perform a statistical test. (In this case, the chi-square test for independence yields an unconvincing $p = 0.1$.) But what about less-experienced students? How can we help them explore whether the observed difference in proportions could be due to chance? Here we can construct one of the simulations we alluded to earlier. We will define a measure of association, and then simulate the null hypothesis (using a randomization test, though we would probably not use that term with students). That simulation will yield a sampling distribution of our measure, to which we compare the test statistic, in this case, the measure applied to the original data.

The advantage, pedagogically, is that students get to use their own measures; they do not have to wrestle with the underlying meaning of chi-square (or *t* or whatever statistic is appropriate). In this case, many students use the change in proportions as a measure, that is, the "natives" proportion declined by about 3 percent from 1950 to 2000.



When we first ask students to define a measure, they will say something woefully incomplete: their measure is "the percentage," or "the change," or even "the percent change." So it is especially useful for them to have to define the measure in a way the computer understands.

In Fathom, they construct this measure in the "measures" panel, a place to define attributes that apply to the whole collection rather than to individual cases. In the illustration, we show one way to compute this measure—called **change**—breaking up the larger calculation into manageable parts. In our experience, just writing the expression for a measure like this can be challenging in any software environment. You have to specify so carefully which numbers to use, which cases to count. To describe it precisely in natural language is even more tortured—which may be a clue to why conditional and proportional thinking are so difficult.

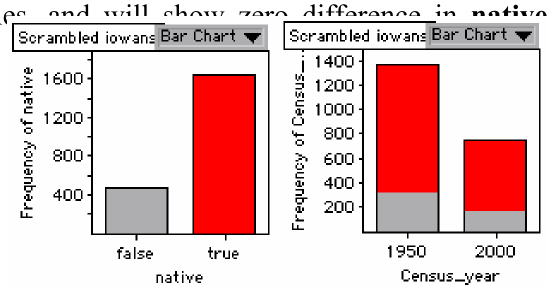
With these problems, is there benefit to students defining measures even if they do not go on to run a simulation and look at inference? We think there is.

Graphs such as the ribbon chart present a phenomenon visually and convincingly. Such a graph usually shows that one group (people from 1950) is different from another (those from 2000) in some aspect (the proportion that were born in Iowa). But this is a qualitative difference. The moment we ask, *how different* the groups are, we need a quantitative measure. It is that step—taking something you're looking at and quantifying it—that a lesson like this makes real for students.

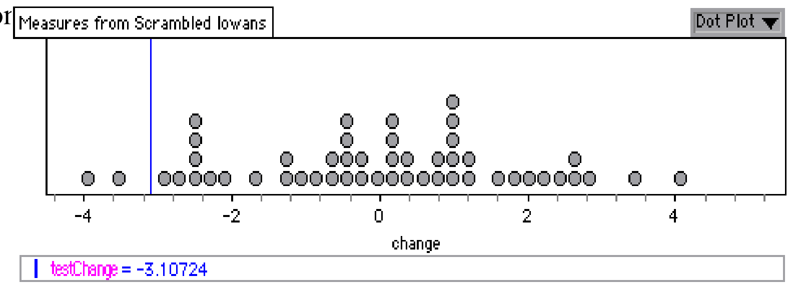
Students could do this step without computer assistance. But besides getting help with laborious calculations, there is learning value in the interaction: they must communicate with the computer unambiguously using symbolic mathematics. This does at least two good things: it gives students practice with a nonjudgmental partner as they struggle to get the formula right; and when they're done, they can (we hope, if we draw attention to it) see the value and efficacy of symbolic statements as a means of communicating complicated ideas—in this case, a difference of two carefully-worded proportions.

Onward to Inference

That number, –3.1 percent, is our test statistic. How plausible is it that that difference of proportions, that **change**, could be due to chance? To answer this question, we construct a situation in which **year** and **native** are truly independent. To do that, we ask Fathom to *scramble* one of the attributes (we'll pick **native**), essentially picking up all of the values, shuffling them, and reassigning them to the cases in the collection regardless of which **year** they came from. On the average, they will fall into the two years in proportion to the sizes of those two samples, and will show zero difference in **native** proportion, that is, **change** = **0**.



But in practice, they will not fall evenly. The illustration shows the same graph as before, but for our "scrambled Iowans." It looks about the same, but in this case, the difference on proportions is +0.79 percent. Does that mean that –3.1 percent is unusual? Not necessarily; we have to do this many times. Fathom makes it relatively easy; we collect measures as many times as we like, and display the distribution of the result. The illustration next shows such a distribution.



Here we can address important issues in learning inferential statistics. There are only two points out of fifty "beyond" our test statistic (that is, $p = 0.04$). So do we believe the change we observed to be implausible? If we look at the other end of the graph (ah! a two-tailed test), two more are beyond +3.1 percent ($p = 0.08$). Does that change our assessment? What is the meaning of an individual point in this graph? (It's the difference of proportions for a single random scrambling of the **native** attribute.) Does our formula actually measure what we think it does? Once students understand questions like these, and have compared sampling distributions using different measures, they will be better prepared to look at a more esoteric measure such as chi-square: they can create its formula, build up its sampling distribution, and see how the formula measures what it is supposed to. Students also learn that their measure may be every bit as good for their purposes.

Discussion

It is time to draw some of the threads from these examples together. Let's begin with the data at the center of this project. Why are census microdata so interesting? Observing students, teachers, and ourselves interacting with the data, the first reason is clear: the data are about us—people in everyday life. But also, because the data are about people, we all bring knowledge to the table. So our first analyses show us something we already know. Our graphs make sense right away, and we gain confidence in the data and in our ability to study them. And then, because our prior knowledge makes us masters of the overall shape of the data, we can immediately look more deeply, and discover things that may surprise us. Even then, we can often offer explanations for these surprises, and we have the background knowledge to study them. Put another way, we often think of data analysis as a way to probe the unknown. But it also helps us see the known more clearly.

Thus students working with these data are both more interested in what the data have to say and better equipped to engage the data than with the traditional data we may give them. As a consequence, we believe that activities like those we have described, and data like the census microdata, may be useful for preparing students for more formal statistical training.

We also think that there are a number of topics and skills that textbook writers often wrongly assume students have previously mastered. These are both essential for good understanding of statistics and accessible with interesting data, good tools, and suitable curriculum materials. Here are three—consider it a list in progress—drawn from our work.

- *Connecting summary displays and calculations to the individual data cases from which they arise*. Students need to know that a bar in a bar chart summarizes (by counting) many individual pieces of data; that each point in a scatter plot comes from one case; that not every case contributes equally to the IQR. This is especially important as students come to grips with distributions; if students cannot imagine disaggregating a histogram or box plot, they do not understand it. We believe that an understanding of the relationship between cases and summaries helps combat the "tyranny of the center," in which students mistakenly impute all members of a group with some summary statistic: while a mean or median summarizes many cases, it may not resemble any particular one.

- *Finding displays that show what you intend*. Some students have a hard time producing a suitable display. Sometimes the display is wrong (for example, if Anna and Bert had been satisfied with their first Iowa/native graph, which essentially cut the conditions the wrong way) or just not as interesting as possible (as when they considered a bar graph instead of the box plots for comparing income to education). Students need practice displaying data they understand well, especially when the conclusion is obvious—and then practice being critical of these displays and of the displays they see every day in the media. We want students to say, for example, "this graph only shows the mean—what about the distribution?"

- *Making statements clear and unambiguous by "mathematicizing" them*. We cannot study an association without describing it. For example, if we're studying whether women live longer than men and by how much, and we ask students how they will figure it out, they say, "we'll compare their ages." That's reasonable, but it does not describe precisely what they have in mind. They probably imagine computing the *mean* age (and not the median), but if they look to see what is really different about the age distributions, they might decide to use the 95th percentile, or look at the proportion of people over 80 who are women. Then students have to render those ideas into mathematical language so that the relevant measures can be computed. Becoming comfortable with this process is a prerequisite for independent investigation.

So let us now talk about independent investigations. Ideally, curriculum materials prepare students to develop their own questions and fully study them, all the way through producing the definitive report on the results. Yet most statistics textbooks concern themselves—understandably—with the statistics *content* students need to learn, especially conceptually difficult topics such as the normal distribution, statistical tests, confidence intervals, the importance of randomness, and the Central Limit Theorem. Conducting an independent investigation is often relegated to an end of term project. Alas, many student products we have seen seem ill-conceived or executed, not in their use of the statistical armamentarium but in more fundamental areas such as asking good questions and looking critically at data.

Once topics and skills such as those listed above are acknowledged as critical to successful execution of an independent investigation and generally lacking, we can devise ways to strengthen them. Prior coursework, particularly at the early secondary level, and the beginning weeks of an introductory statistics course are attractive targets. Our hope is that the census microdata materials discussed here can serve as prototypes of how to accomplish this goal.

But there are "independent-investigation" skills that may be useful for general quantitative literacy and for being a critical thinker. Here are four:

- *Finding suitable data to address your question*. Measurement is indirect. We hardly ever actually measure what we really want: we measure blood chemistry to assess pregnancy, income as a stand-in for wealth, we take an opinion poll to predict a future vote. Problems in a stats textbook almost always tell us what data to use, but real-life investigations require imagination. We can give students situations in which they choose what data to use to get at a particular idea—and then compare the consequences of different choices.

- *Developing appropriate questions or conjectures*. Looking at census microdata almost always generates conjectures. For example, when Anna and Bert see that the high-income outliers are predominantly male, they naturally wonder (or assert) that men generally earn more than women. They can study this, to see if it is true, and make a suitable display. They can ask a more complex question: If you're looking for income, which is better: being a man, or getting a college education? Students also need help turning vague or specific observations into clear and generalized problem statements. In traditional exercises, the question is already well-formed. But we can give students practice with this important step through explicit exercises. We can also help students to ask more useful questions: to move from 'Is there a difference' to 'How big is the difference, and how reliable is our answer'?

- *Developing alternative hypotheses*. We have had interesting results at the end of an activity asking what other reasons could account for the data. For example, we think that you get a better job with a college education and therefore earn more money. So we ask, is there any other reason we could see this association? Perhaps it's because retirees—who earn less because they're on pensions instead of salaries—were educated in an earlier era when fewer people went to college. This exercise is a good one for developing hypothetical thinking skills; it also helps students find lurking variables and guard against biased sampling techniques.

- *Valuing nuance*. Students are often willing to stop investigating when they get a clear, black-and-white answer. But statistics distinguishes itself from most of mathematics by being equivocal. Looking at rich data with powerful tools lets students develop the habit of digging deeper—as Anna and Bert did when they found the overlapping income distributions in the education study. It is not that we want to inculcate in students a habit of indecision (or of moral relativism); rather, we want them to be able to make decisions in the face of uncertainty, recognizing and considering the real complexities that are endemic to data and to life.

Developing skills like these is a long—perhaps life-long—endeavor. And though they tend not to appear in a textbook table of contents, and may be given short shrift in traditional curricula, these skills are not of a wholly new phylum of learning. For example, they fit well within the "process standards" of the NCTM (2000) *Standards* (problem solving, reasoning, communication, connections, and representation). They also form part of the foundation of quantitative literacy that Steen (2004, p. xii) asserts "largely determines an individual's capacity to control his or her quality of life and to participate effectively in social decision-making."

Sources and References

The College Entrance Examination Board. (2003). *Advanced Placement Statistics Course Description* New York: College Entrance Examination Board.

Census at School. This is a multi-year, multi-country, effort to administer surveys at schools and make the microdata public. See www.censusatschool.ntu.ac.uk/.

Conti, C. & Lombardo, E. (2002). "The Italian Census at School®". *Proceedings of ICOTS-6.* Voorburg, the Netherlands: International Statistical Institute.

Edwards, W., Lindman, H. & Savage, L. J. (1963). "Bayesian statistical inference for psychological research." *Psychological Review* 70 : 193-242.

Erickson, T. & Finzer, W. (1998). "Software Design with Learning in Mind." *Proceedings of ICOTS-5.* Voorburg, the Netherlands: International Statistical Institute.

Erickson, T. & Holmberg, M. (2005). Census Microdata in the Mathematics Classroom. Emeryville, CA; Key Curriculum Press. In preparation.

Erickson, T. (2001). *Data in Depth: Exploring Mathematics with Fathom*. Emeryville, CA: Key Curriculum Press.

Finzer, W. (2005). *Fathom Dynamic Statistics Software*, Key Curriculum Press. Version 2, 2005.

Finzer, W. (2002). Is Research-based Development of a Commercial Statistics Learning Environment Possible? - The Fathom Experience. *Proceedings of ICOTS-6*.

Frey, W.H. (2003). *Investigating Change in American Society: Exploring Social Trends With US Census Data (*2nd edition). Wadsworth.

Hammerman, J. & Rubin, A. (2003). Reasoning in the Presence of Variability. Presented at *The Third International Research Forum on Statistical Reasoning, Thinking, and Literacy,* July, 2003.

Karplus, R. & Thier, H. (1967). *A New Look at Elementary School Science*. Chicago: Rand McNally.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

Ruggles, S. (1998). "Microdata in the Classroom: The IPUMS model." Presented at the "2000 and Beyond: Making the Census Accessible" New York: Russell Sage Foundation.

Scheaffer, R., Watkins, A., Wittmer, J., Gnanadesikan, M., & Erickson, T. (2003). *Activity-Based Statistics* (second edition). Emeryville, CA: Key College Publishing.

Steen, L.A., (2004). *Achieving Quantitative Literacy—An Urgent Challenge for Higher Education.* The Mathematical Association of America.

Endnotes

[1] About confidentiality and anonymity: the Census Bureau has taken pains to ensure that we cannot actually identify individuals. The effectiveness of these procedures could be the subject of another paper.

[2] A quick note before we move on: we also know this woman's *income*. Whatever this says about us and our world, some of the most interesting questions we can ask or study are about *money*.

[3] Census microdata is obtained from the IPUMS project at the Minnesota Population Center of the University of Minnesota (www.ipums.umn.edu/). A collaborative agreement makes it possible for Fathom users to send requests directly from Fathom to IPUMS.

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

[5] In fact, this *is* a random sample taken from a 1 percent sample of the U.S. population.

---

[6] Many of the observations we will make here are not supported, to our knowledge, by research. They do, however, seem to be true—they are reasonable conjectures—based on anecdotal classroom observations and informal interviews. Some of our ongoing research is trying to find out more. We would be grateful for appropriate pointers to past literature and for suggestions—and offers—for future research.

[7] This is an artifact of the data source: the number of cases in samples from different places and years is not proportional to their populations. That is, the decrease in sample size doesn't indicate a decrease in population. This is confusing to alert students, of course; we are working to resolve the difficulty.