

Much Has Changed; Little Has Changed: Revisiting the Role of Technology in Statistics Education 1992-2007

Andee Rubin, TERC

1. INTRODUCTION

In 1992, I wrote a paper for the AAAS called “Technology in Statistics Education: A Double-Edged Sword.” When asked to write a paper for this inaugural issue of TISE, I thought it would be interesting to look back at that paper with the experience I have gained since then. I noted that some of the things I wrote about as exciting innovations are old hat by now. Others have been stuffed in the dead application box. Still others have blossomed into a new line of research and development in statistics education. There are new faces on the scene, as well. In this paper, I will discuss (and even quote) what I offered in 1992 and reflect on the fate of the developments that were worthy of discussion back then.

Since 1992, data analysis and statistics have enjoyed more attention and visibility in educational circles. The National Council of Teachers of Mathematics’ Standards (NCTM 2000) include data analysis and probability as a strand parallel to algebra and geometry, and elementary math curricula have been designed to give students more experience reasoning about data. ETS introduced the AP Statistics test in 1997, and the number of students taking the test has increased every year. Both beginning data analysis and sophisticated statistics courses are more generally available in high school, and statistics is sometimes considered a suitable capstone high school math class for those who are not going on in science or engineering (Rubin 2005).

Even with these hopeful developments, however, statistical literacy in the general public has not changed very much. Few people have any kind thoughts for statistics, and even fewer understand the concepts underlying statistical comparisons and relationships. I still have frequent conversations with taxi drivers, colleagues and friends who believe that no poll can be trusted

because “after all, they didn’t ask me.” So a general caution for readers; as amazing and inspiring as these technologies may seem, none of them can have any educational effect without carefully constructed curriculum and talented teaching.

2. VIDEO AS DATA

Using computers as enhancements and extensions to media has become so common as to be unremarkable these days. A teenager who doesn’t know how to prepare a video for transmission on YouTube is behind the times. Many young people don’t know that video has not always been digital. Yet, with all their familiarity with the flexibility of computer-based media, few students would think that video could be a source of data. In this section, we’ll consider past systems for using video as data and consider how new developments may allow the field to progress.

My 1992 paper explored the mathematical use of video even though, at the time, manipulating video on the computer required a messy and unreliable process of digitizing it from an analog medium. My paper described in some detail my own project, TapeMeasure, which was one of the first pieces of software to test the idea that videotape could be used to explore the mathematics of change over time. The designers of TapeMeasure and its descendants observed that most motion takes place in a time too short to support real-time analysis, so they used the computer to slow down motion so that data could be captured easily. The TapeMeasure system allowed middle and high school students to make measurements on selected frames of a (digitized) videotape of motion and store the data, including the time, for analysis. The pedagogical strength of this approach came from the connections students made among multiple perspectives on a particular motion. Watching the video at standard speed allowed students to see their measurements in context, a reminder of the purpose of the investigation and of the complexity and limitations of taking measurements; watching it at a slower pace revealed other aspects of the motion that were missed otherwise.

In our most extensive investigation, which I discussed at some length in my 1992 paper, eighth-grade students analyzed their running performance. They videotaped themselves running, digitized the video and carried out a variety of measurements on a frame-by-frame basis. Using

TapeMeasure, they could measure position (e.g. where a foot was in each consecutive frame), angle (e.g. of an elbow or knee) or count (e.g. how many times an arm pumped in a second).

TapeMeasure was an awkward system, which required complex technology and considerable human assistance. The next generation of video analysis software could be used more independently, given technological progress in the intervening years. I led a project to develop CamMotion (Camera/Motion), the daughter of TapeMeasure, as part of a project called VIEW (Video for Exploring the World). Although digitizing video was still a hurdle, CamMotion went beyond TapeMeasure in its capabilities and ease of use. The mathematical/data focus of the VIEW project was the representation of motion, i.e. proto-calculus. On the ViEW project, middle and high school students created many of their own videos, and we developed a rich corpus that demonstrated a variety of mathematical aspects of motion, e.g. constant high or low velocity, acceleration, circular motion, and comparisons of simultaneous motions.

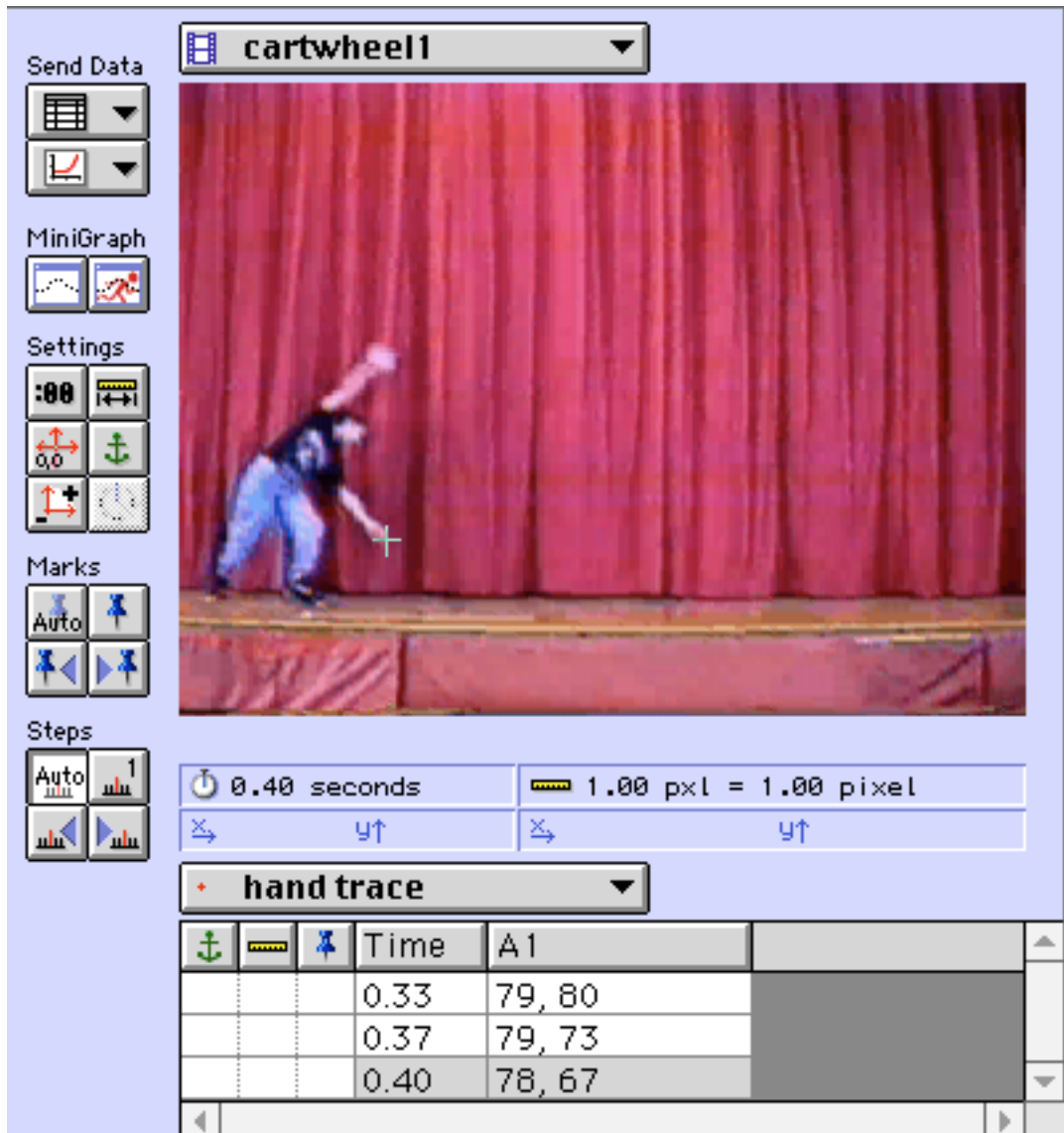


Figure 1. CamMotion: one frame of a girl turning cartwheels

Many of the capabilities of CamMotion can be seen in this example of the analysis of a young girl turning cartwheels. In Figure 1, we see one frame of a video of a girl turning two cartwheels. (Here she is shown in the middle of the first cartwheel.) The question the user is investigating is: what is the path of the girl's left hand as she turns the cartwheels? Note the cross on her bottom hand. This indicates that the user has clicked on this position and corresponding data have been entered into the table, in this case, the time (from the beginning of the video) and the X- and Y- coordinates of the position. The user continues through the entire video in this way, tracking the path of the point. Note that CamMotion does NOT automatically

know where the hand is; it is up to the user to choose a position on the girl and be consistent about clicking on it in each frame.

(Before you read on, try to guess the shape of the path of the girl's hand.)

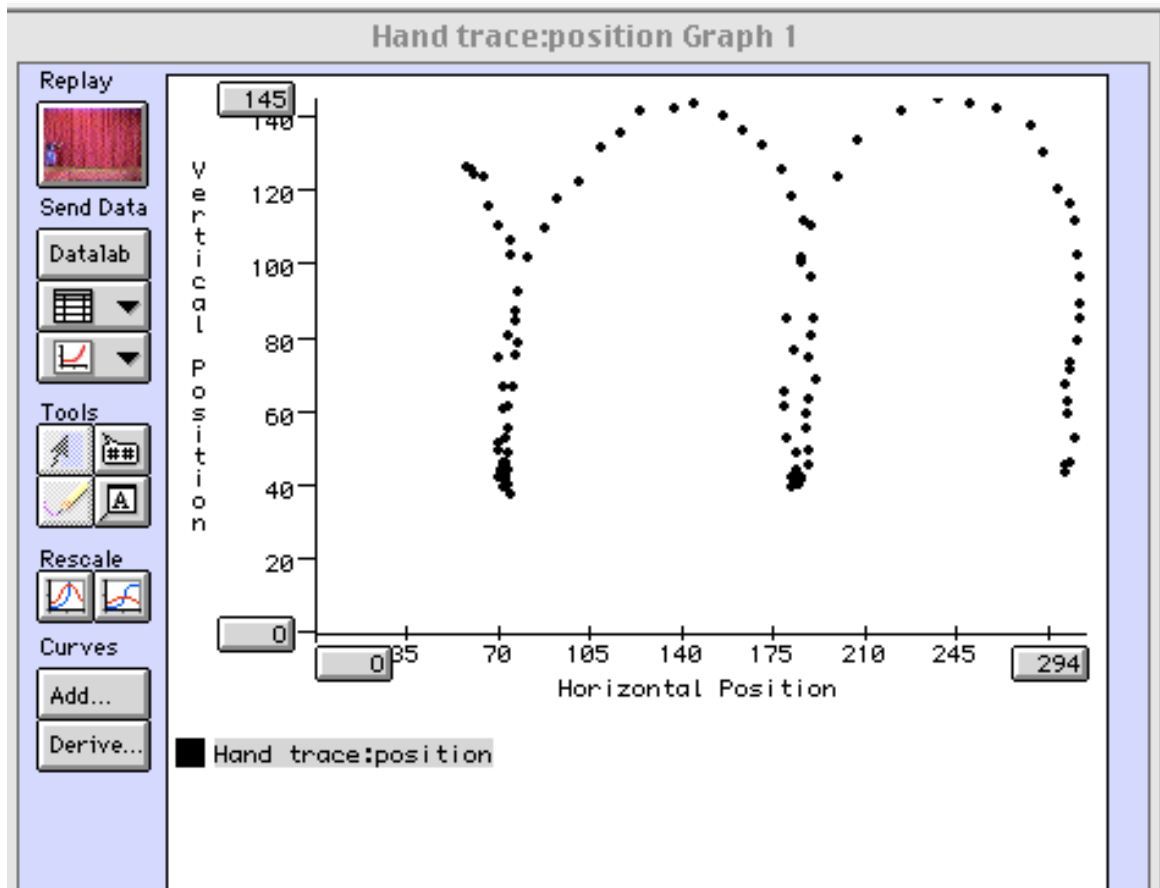


Figure 2. The path of the girl's left hand through two cartwheels

Figure 2 shows the entire path of the girl's left hand through the two cartwheels. How did you do? In addition to the general shape of her hand – going up and down in a generally symmetrical arc – we can see that her hand moved more quickly at the top of the arc where the points are further apart, and much more slowly at the bottom, where the points are almost on top of one another. A more sophisticated analysis of this video might include examining the X vs. time and Y vs. time graphs and attempting to fit a function to each curve. The Y vs. time graph is shown in Figure 3.

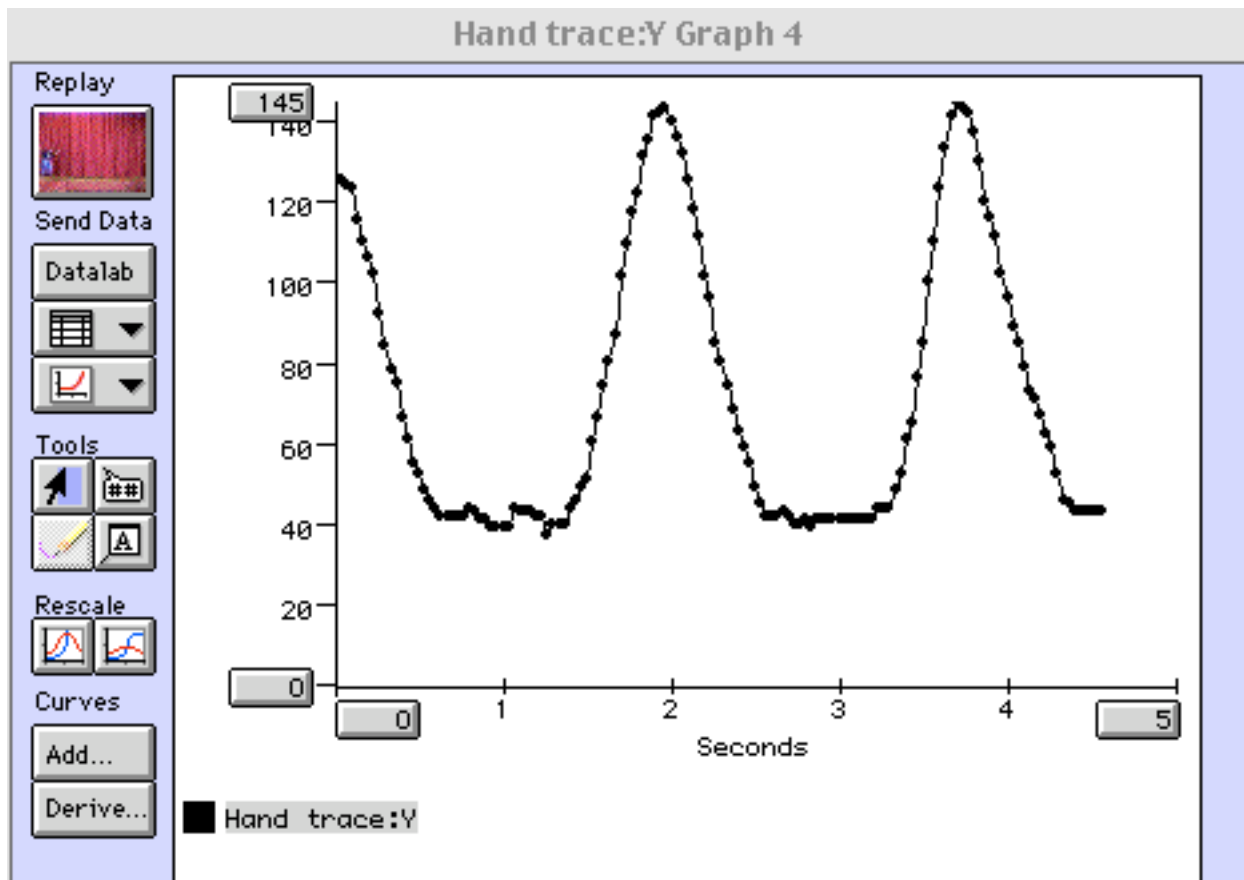


Figure 3 Y-position with respect to time.

Although neither TapeMeasure nor CamMotion was ever commercially published, a number of current products have incorporated and elaborated on their features and attempt to take advantage of advances in video technology, of which the most important is the fact that it's now digital. The digital revolution in video technology has made possible the explosion of YouTube, and videos on cell phones, among others, and many students are facile with editing videos to create their own video products. Now just about anyone can create a video setup, and programmers will likely find it easier to create high-quality video analysis packages to incorporate into data analysis environments. So, what have been the implications of this sea change from analog to digital for education?

Despite the change in video technology, many of the video analysis tools that are currently available differ only in minor ways from those I wrote about in the mid-90's. Measurement in Motion, for example, provides most of the same tools as CamMotion did a decade earlier. In

addition, it has additional tools for creating and evaluating quantitative models e.g. more flexible manipulation of formulas and additional graph types. The Measurement in Motion website describes a variety of high school projects that incorporate video analysis, including investigations of athletes' performance and dynamic forces in nature (e.g. the Coriolis force).

One of the changes we see in the past decade involves the connection between video data and other real-time information about a situation gathered using a variety of sensors. *Logger Pro*, for example, is primarily an interface for collecting and modeling data from many types of sensors (e.g. temperature sensors, motion sensors, accelerometers), but it also can coordinate sensor data with video analysis capabilities similar to those of CamMotion or Measurement in Motion. To provide the visual data collection capability, the developers of *Logger Pro* licensed VideoPoint video capture software, which had originally been developed as a high school and college physics teaching tool. With *Logger Pro/VideoPoint*, for example, one could attach an accelerometer to a gymnast's thigh, take the video of a cartwheel, then see the collected and computed values of position, speed and acceleration update as the video plays.

We can see a truly new direction by following this road into the relatively near future, where we see imaginative tools such as Alberti's Window. Alberti's Window pushes the idea of video as data to literally another dimension. Given a specific configuration of two video cameras and a visually distinguished (by color or contrast) object of interest, the software can automatically find the object's 3-D coordinates in each video frame and produce a variety of complex 3- and 2-D graphical representations of its motion. Digital video makes these feats possible, since the complex calculations involved require accurate and immediate pixel-level data.

While these innovations add to the accessibility and usefulness of video analysis, there are some potential downsides to the use of more and more powerful software. Paradoxically, some of the strengths of early systems such as CamMotion relied on their lack of sophisticated features. One of the advantages of CamMotion and Measurement in Motion was students' ability to use video from any source, be it supplied with the software, taken by students themselves or taped from a TV show; these options were important connections for students with the video's content and context. While the 3-D capabilities of Alberti's Window are remarkable and an important step

forward in motion analysis, using the software requires that the motion take place in a special setting. This makes it more difficult for students to analyze videos they create themselves outside of the classroom or extract from other media.

Similarly, *actively* tracking a selected object across video frames provides students with another intimate connection with the video and an awareness of the subtleties of the motion. Students often find the positions that are revealed in a single frame of their own or their classmates' motion amusing and engaging and can lead to animated discussions of "how that could be possible?" So, while there are many advantages to automating some of these labor-intensive procedures, we need to be careful not to compromise the pedagogical intent in the process.

3. GEOGRAPHICAL INFORMATION SYSTEMS

In 1992, Geographical Information Systems (GIS) were used primarily by scientists; their potential as a source of data in the classroom had not yet been recognized. Although the United States Geological Service (USGS) had provided mapping data for decades, it was only recently that many educational software and curriculum developers recognized the pedagogical value and inherent motivation of such visualization tools. Although their use is currently limited by the small number of geography and earth science courses in middle and high schools, GIS have found their way into a growing number of classes and from this vantage point, they seem poised to become a major technological presence.

One driving force in this development has been the Environmental Sciences Research Institute, Inc. (ESRI, www.esri.com), who have created a comprehensive set of tools (the most commonly used in schools is ArcSchool), curriculum, data, case studies and seeded a user community. Users of ESRI tools focus on geographic problems; one well-publicized example is the group of Native American students at the Santa Fe Indian School who use GIS skills to participate in environmental and water management programs in their communities.

The GLOBE project takes a different perspective, using the capabilities of a GIS to create "a worldwide community of students, teachers, scientists, and citizens working together to better

understand, sustain, and improve Earth's environment at local, regional, and global scales.” (www.globe.gov). GLOBE projects engage students from many countries in investigations that rely on integrating data from many locations. A representative project is Season and Biomes, which is gathering data on seasonal variability and its effects on local communities; the project will use both satellite data and student field data (which will be used partly to validate the satellite data).

Another group who has customized a GIS for education is the Geographic Data in Education Initiative (GEODE) at Northwestern University, whose mission is to improve Earth and environmental science education by using data visualization and analysis. Their particular GIS is MyWorld, designed for use in courses in middle school through college that involve investigations of geographic data. In a particularly timely example, MyWorld has been used by middle-school students in Chicago to study the effects of road salt on the soil in their neighborhoods. Using Google aerial maps, they chose a set of data collection locations where, in addition to collecting soil samples, students recorded information about the road near their sampling locations (number of lanes, speed limit, curb height), nearby vegetation, weather conditions at time of sampling and the slope of the sampling site. They analyzed the soil samples to determine their salinity, then entered all of the data into the GIS at the appropriate locations, creating a data set that had a true spatial organization. Figure GIS-1 shows a map of the Chicago area with the sampling sites marked by small squares colored according to the degree of salinity found at that site, based on the scale at the bottom of the map. The “Salt of the Earth” project’s results to date indicate that salinity decreases as the distance from the road increases and the salinity of soil in the winter is higher than that in the spring or fall, but there is no significant difference between the spring and fall. These data are consistent with the hypothesis that salt is leaching into the soil from the road, primarily from the use of salt on the roads in winter (but there are likely other hypotheses that fit the data, as well).

The most striking and fast-growing use of GIS is the Google Earth “phenomenon.” (earth.google.com) A cross among a GIS, a collection of data and projects and a community wiki, Google Earth has captured the attention of adults and students alike who find exploring maps and geographic phenomena fun. Using Google Earth, one can find such far-flung

resources as a project on reports of Bigfoot sighting and one on geographical aspects of the Mongol Empire.

The buzz around Google Earth suggests that we may have just scratched the surface of the potential of GIS to engage students. Earth science is already becoming more available as a topic in high school and as it grows in popularity, so will the pedagogical uses of GIS. Aside from that general prediction, the future is not too clear. However, from the perspective of the data analysis community, these developments provide an interesting opportunity. An effective GIS requires both powerful map manipulation capabilities and quantitative analysis tools; current GISs excel in the first of these, but may need more support from the data analysis community to achieve the same strength in the second. Here is an opportunity for statistics educators to collaborate with science educators in a project to increase the pedagogical impact of both communities.

4. GRAPH CONSTRUCTION SOFTWARE

Of course, collecting data from video, graphs or any other source must be followed by analyzing them in order to make meaning of the patterns they contain. In 1992, most pieces of statistical analysis software produced mammoth tables and included a limited menu of “professional” graphs. But in the early 90’s a new kind of data analytic software was beginning to be designed. In my previous paper, I described the TableTop, one of the earliest pieces of software that gave primarily elementary students considerable freedom in creating a series of graphs with which they could describe and make meaning of patterns in the data. Each item in the dataset was represented by an icon on the screen that “carried with it” all of the values of its attributes, which could be accessed simply by double-clicking on the icon. When a user defined a graph (e.g. a Venn diagram, with logical expressions describing each set), the icons would literally move into their new organization with a simple animation. In this way, students had a sense of the unchanging identity of each icon, even though it appeared in different places in different graphs. TableTop was easy enough to use that students could carry out investigations with their data after only a short introduction to the software.

The use of data analysis “construction” software has proved fruitful and, in some statistics education circles, taken for granted. The next generation of these data analysis tools reaches far beyond the original designs that TableTop introduced – in ways that were not even on the agenda or perhaps even imagined in 1992. For example, the designers of TableTop continued to develop the software, culminating in InspireData, just published in 2007. Fathom and TinkerPlots, both published by KCP Technologies, take this idea much further, providing many “primitive” operations that can be combined into an extraordinary number of graphs designed to carry out different statistical investigations. Below, I describe TinkerPlots, which is designed primarily for middle school students (but which has been successfully used with high school students and adults). In the section on Playing Around With Data, I will discuss Fathom, designed primarily for high school and college students.

The series of graphs below illustrates an analysis of a data set about Australian teenagers, including their gender, the year they were born, and the amount of money they earned per week. (downloaded 2003). (These data formed the basis of a set of interviews we carried out with middle and high school teachers. Results from these studies are reported in Rubin and Hammerman 2006; Hammerman and Rubin 2005.)

The distribution is quite skewed (as are most income distributions); while many students earned very little, there were a few who earned a considerable amount, including one who made almost 450 dollars a week.

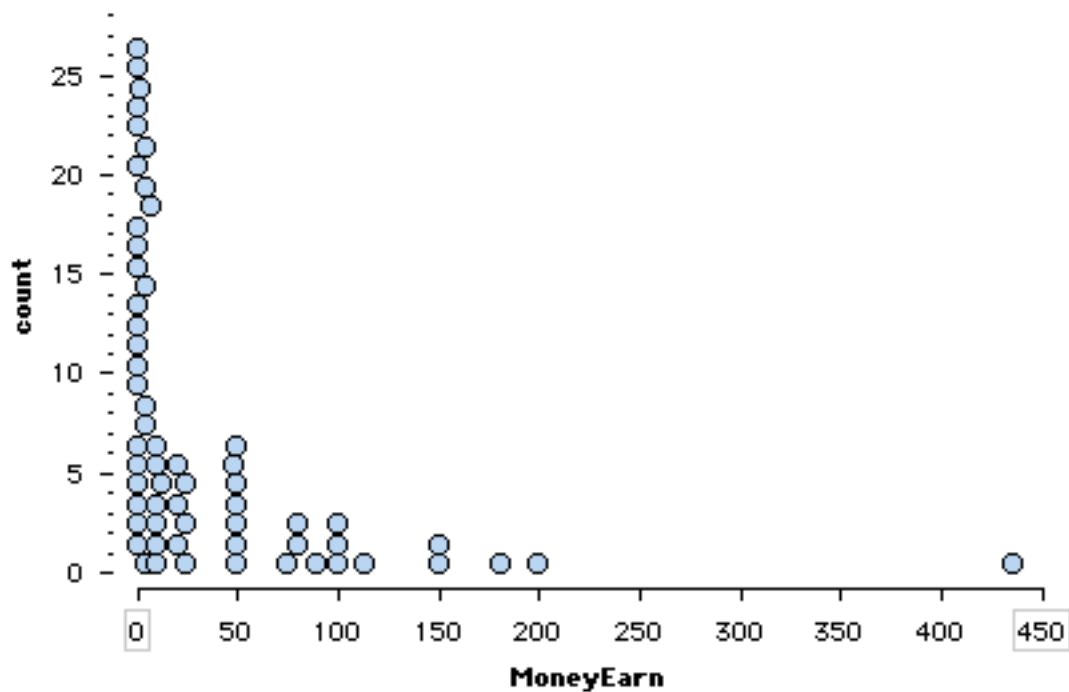


Figure 4. The distribution of the amount of money earned per week in a sample of 60 Australian teenagers.

One might wonder if gender affects the amount of money a teenager earned. The next graph, in which we can see separate distributions for males and females, helps to answer that question. Creating this graph in TinkerPlots requires just one action: adding Gender on the Y-axis to the distribution of MoneyEarn on the X-axis.

The top distribution in Figure 5 represents females (the purple icons) while the bottom represents males (the yellow icons). It's pretty clear even at a first glance that males in this sample generally make more than females. To illustrate this relationship, the user has chosen to display the data with the central 50% of each distribution enclosed in a gray rectangle. From this graph, we can see that 75% of the boys made \$90 or less, while 75% of the girls earned less than \$25. The medians of the distributions, marked by the upside-down T are quite different as well (\$5 vs. \$25); this sad statistic tells us that, at least in this sample, half of the girls make \$5 a week or less. In fact, with a few more manipulations of the graph, we can discover that 15 of the girls earn no money at all.

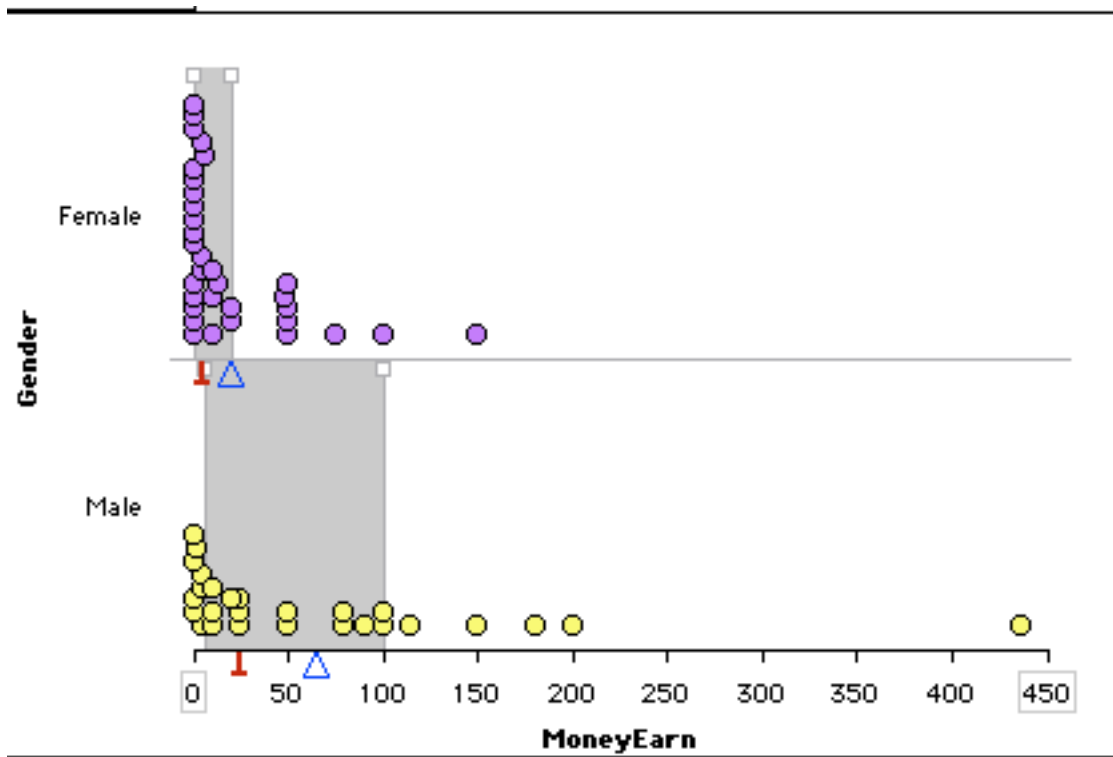


Figure 5. Money Earned for the sample of 60 students. The top half of the graph represents the females, the bottom half represents the males.

If we had chosen to compare males and females by mean, marked by the small triangle, rather than median, we would have found an equally dramatic situation; the mean of the male distribution is over three times that of the female distribution (\$66 vs. \$20). At this point, we may begin to wonder about the effect of the outlier, the male who makes nearly \$450, on our comparison. In fact, we have no way of judging the legitimacy of this data point; perhaps it was \$45 mistyped. But even if it weren't an error, this data point is having a significant influence on the comparison. We might wonder what our conclusions would have been if this data point had not been in our sample.

With TinkerPlots, it's simple to explore the effects of this possibility. We can temporarily hide the outlier so that it does not appear on the graph or figure into the computation of the mean or median. Figure 6 shows the graph with the male outlier hidden and the mean for the males changed accordingly. The male mean decreases rather drastically from \$66 to \$51; still a sad state of affairs for the girls, but not quite as drastic. The median amount of money earned for the

boys, however, barely changes, going from \$25 to \$24.50. For a student, it's an interesting lesson on the different effect an outlier can have on the mean and median.

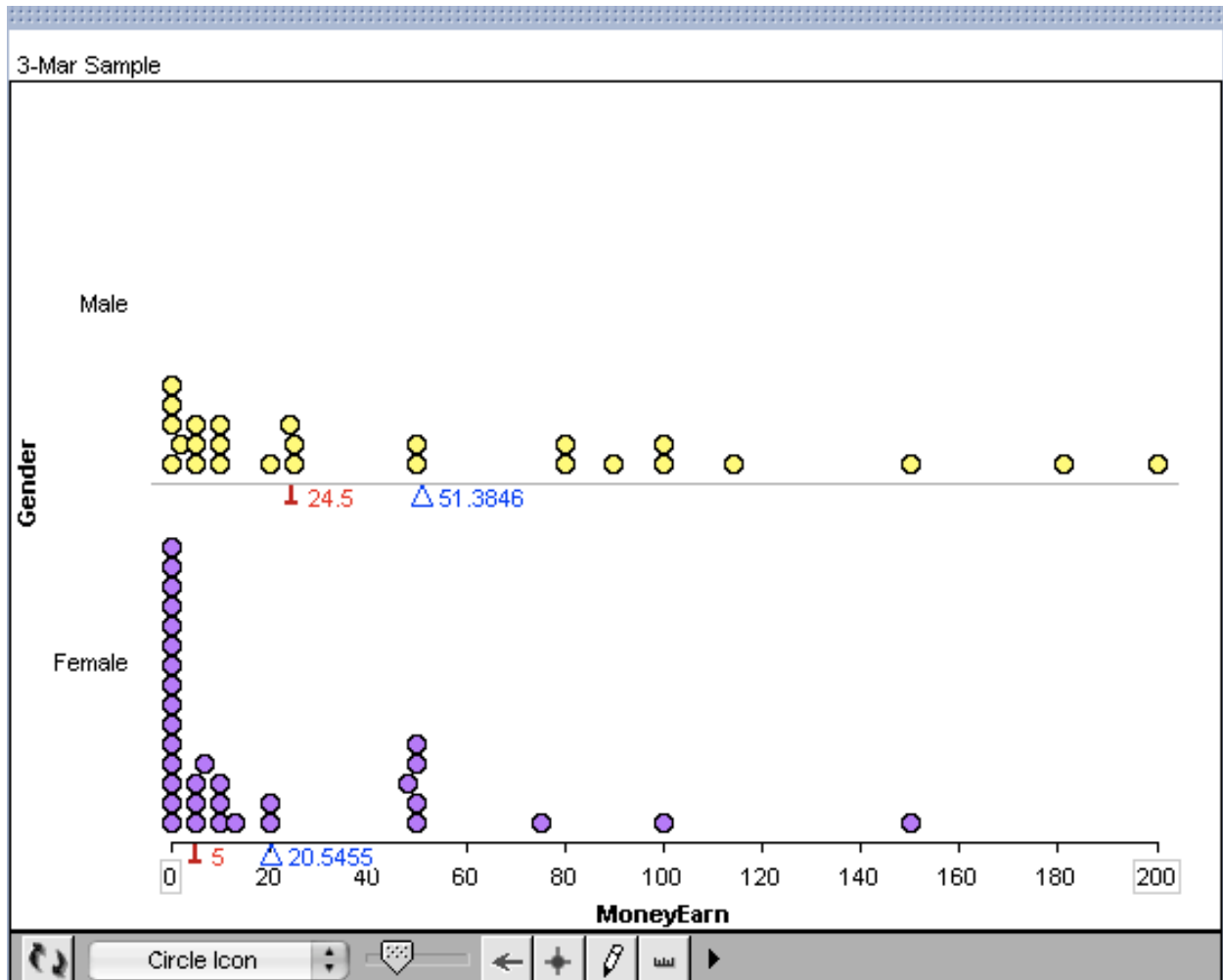


Figure 6. A comparison of money earned between males and females using medians (upside-down T's) and means (triangles).

The teachers whom we interviewed had many possible explanations for the discrepancy between boys and girls. It's possible that girls "work" more at home, helping with household chores, than boys do. Although they may be putting in long hours, housework doesn't pay when it's in your own house. Boys are more likely to work outside the house and thus are more likely to get paid at least something. An additional hypothesis was that when girls weren't working at home, they were likely to be earning money by babysitting, also not a high-paying occupation. Finally, everyone had some story to tell about the boy who earned over \$400 per week. For many, his earnings could be explained only by some illicit activity such as dealing drugs; others believed

he must be a rich kid, perhaps the son of a CEO, who got large amounts of money from his family for doing relatively little.

As this example demonstrates, analyses in TinkerPlots can take many paths, as is appropriate in authentic data analysis. Because graphs are constructed using a sequence of small steps, the number and variety of graphs is beyond that any one individual will create. Has the situation changed in the last 15 years? Here is what I noted in 1992 about the data analysis programs that existed at the time. “While <it creates> a clever and communicative representation, the software limits the set of available representations. Computers can lead to a valuing of convention over invention, stealing from students their opportunity to innovate.” I contend that data analysis software has expanded so much and in so many directions that this limitation is much less of a concern than it was in the past. While it is always important for students to have experience with creating their own graphs from scratch on paper, going on to use a program like TinkerPlots can expand their creative potential rather than constraining their choices.

An advantage of software like TinkerPlots (and others described below) is that students can engage in more fruitful conversations with the teacher and other students with a collection of TinkerPlots graphs to use for discussion and demonstration. In fact, TinkerPlots has the tools for students to create an entire document reporting on an analysis, including their alternate analyses, the hypotheses they entertained, how they interpreted the evidence, what confidence they have in their methods and conclusions, and what additional data they would need to collect in order to learn more about the phenomenon. This is an advantage I mentioned in 1992, but since then the number of new tools that embody an inquiry-based approach to data analysis has expanded and, hopefully, their use will support more productive classroom conversations.

5. PLAYING AROUND WITH DATA

One notable aspect of our analysis of Australian teenagers’ money-earning behavior is that we temporarily deleted a point in order to explore alternate hypotheses. The idea of changing data may seem at odds with the process of analysis, but there *is* a place for working with and manipulating hypothetical data. Statistics Workshop (previously ELASTIC (Rosebery and

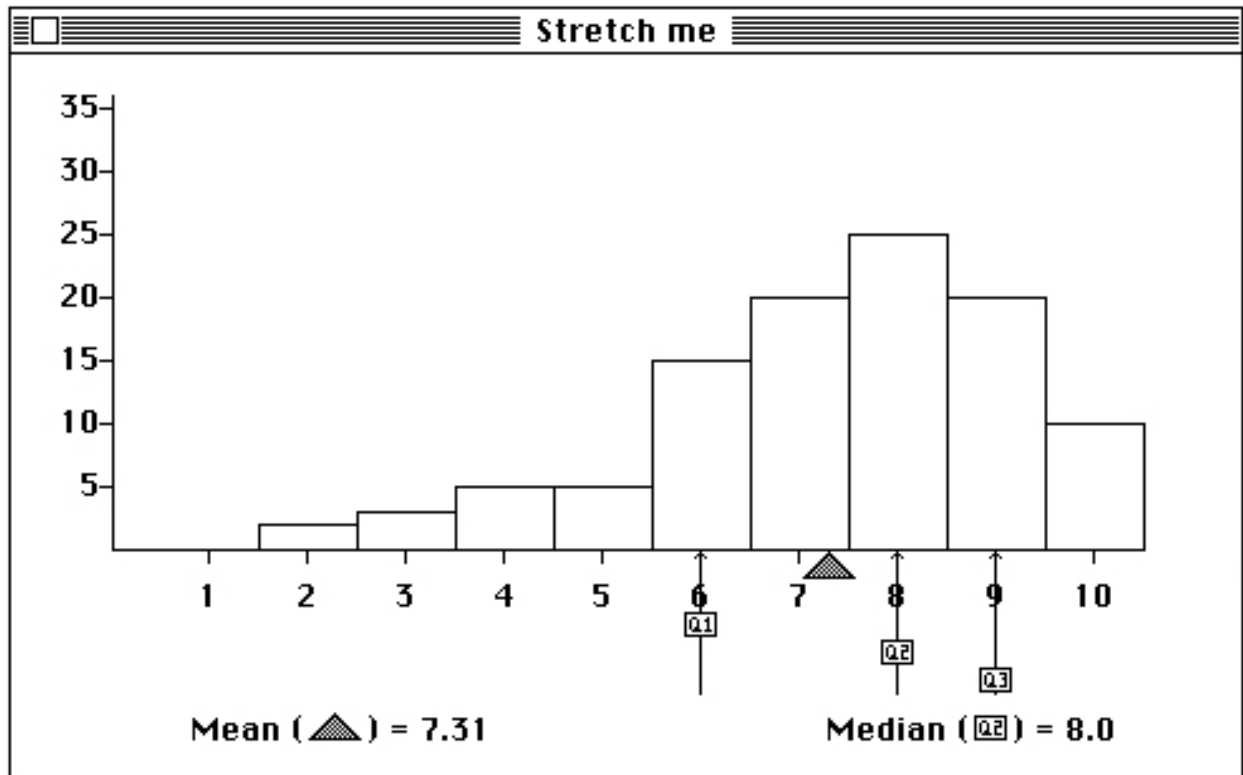
Rubin 1989)) was one of the first pieces of software that included special tools that allowed students to manipulate data in order to learn how to think about numerical distributions and their measures of location (mean and median) and variability (quartiles). “Stretchy Histograms,” as it was called, allowed students to manipulate the shape of a hypothetical distribution as represented in a histogram. Using the mouse, they could stretch or shrink the frequencies of the classes in a distribution and watch as graphical representations of mean, median, and quartiles were dynamically updated to reflect the changes. They took on challenges such as: “Create a distribution in which the median is much lower than the mean.” Or “Create two distributions that have the same mean and median – but whose shapes are different. In this way, students were able to explore the relationships among a distribution's shape, its measures of center, and its measures of variability.

Here’s the example I described in the earlier paper.

An incident in a high school statistics class shows how Stretchy Histogram might be used to explore how mean and median provide different measures of location in a distribution. Students were asked to graph the distribution of a sample of 105 people who had been asked how many hours they slept each night. Figure 7 is a copy of the Stretchy Histograms screen for these data. The second problem asked them to add additional data to the right end of the distribution. When they raised the 10-hour bar of the histogram so that 35 people, rather than 10, slept 10 hours (which they could do instantly with the mouse), the median value was, surprisingly to them, still 8, as shown on Figure 8. Thinking they had done something wrong, they asked their teacher. She suggested they skew the distribution even more and raise the 9-hour bar from 20 to 35 people, as on Figure 9.

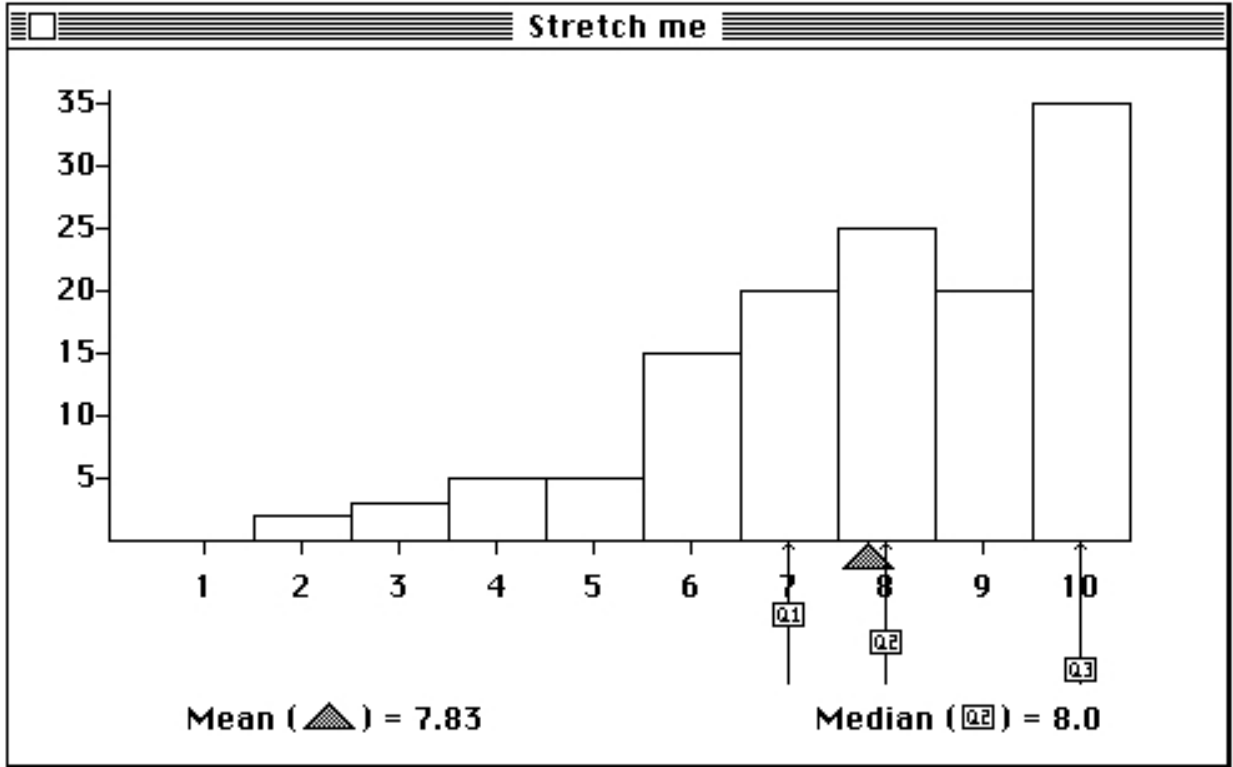
Imagine their surprise when the median continued to be 8! The teacher was surprised - and confused - as well. It took a long class discussion in which students and teacher participated as equals in understanding how stable a measure of central tendency the median really is. By stretching their histogram even more, they discovered an interesting property of a distribution with many repeats of the same value (e.g. many people who

slept 9 hours, in this example). They saw that adding relatively large numbers of new data points to the distribution might not change the value of the median. The explanation of this seeming paradox was that the median, in fact, did change, but it changed to another data point with the same value. Thus, the particular data point that was the median changed - but its value remained the same.



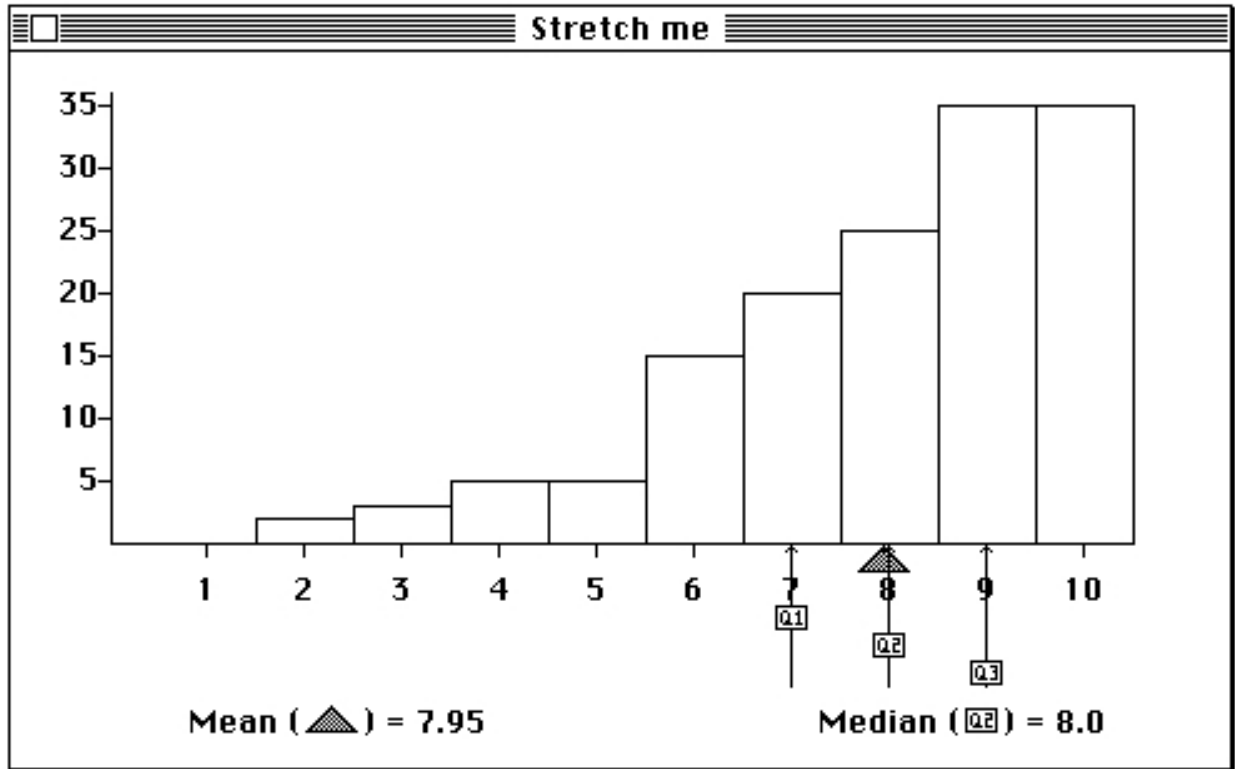
Panel 1

Figure 7. Stretchy Histograms: Original distribution.



Panel 2

Figure 8. Stretchy Histogram with the number of people at 10 hours increased to 35.



Panel 3

Figure 9. Stretchy Histogram with the 9-hour bar raised from 20 to 35 people.

While manipulating data in this way clearly has pedagogical value, it could be seen as the top of a slippery slope; when is it OK to change data? And how does one keep track of the changes? Stretchy Histograms approached this problem by making stretchable histograms separate from the rest of the software. The user could create them from real data, but they were always labeled with the title “Stretch me,” to be clear that the distribution they represented had been created manually and weren’t “real.”

The concept behind Stretchy Histograms – that much can be learned by changing data with an experimental mindset – has grown dramatically in the past 15 years, made possible by advances in both hardware and software that support more complex experiments with data. New pieces of software provide tools to temporarily modify many values in an analysis other than the location of a point, e. g., the slope of a line, the location of the origin on a graph, or the probability of an event occurring. Fathom (the predecessor and “mother” of TinkerPlots, designed for high school through college students), for example, allows users to change a variety of values in analyzing a

graph, including individual points. Here's an example of an exploration of the meaning of R^2 within a Fathom document.

Each distribution shows seven points, the Least Squares Regression Line for them and an R^2 for that set of values. Figure 10 is the original distribution included with the sample document. The legend reads: $y=0.229x+4.7$; $r^2=0.041$

The instructions to the student are: "Move the points into various different configurations and notice the effect on r^2 ." The first posed challenge is: "Under what circumstances is $r^2 = 0$?" There is at least one relatively simple way to decrease r^2 significantly: move the (red) dot down and to the left, as shown in Figure 11. Moving this point also changes the Least Squares Regression Line, of course, to one that is close to horizontal, another indication that there is little correlation between the two variables. In fact, r^2 in this graph is 0.00059. But there are many other ways to decrease the correlation between x and y , and part of the learning experience for students is investigating patterns in those strategies.

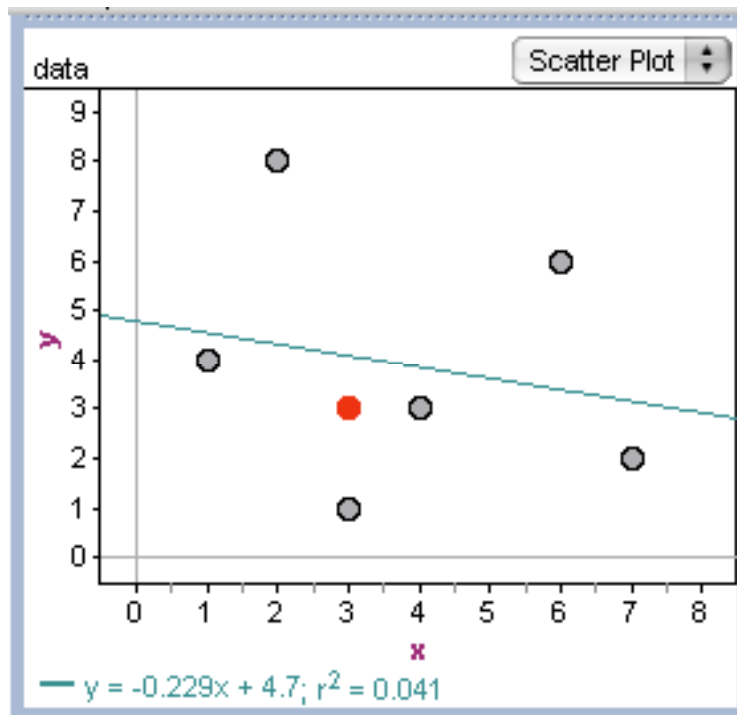


Figure 10. A Fathom scatterplot with a selected point in red.

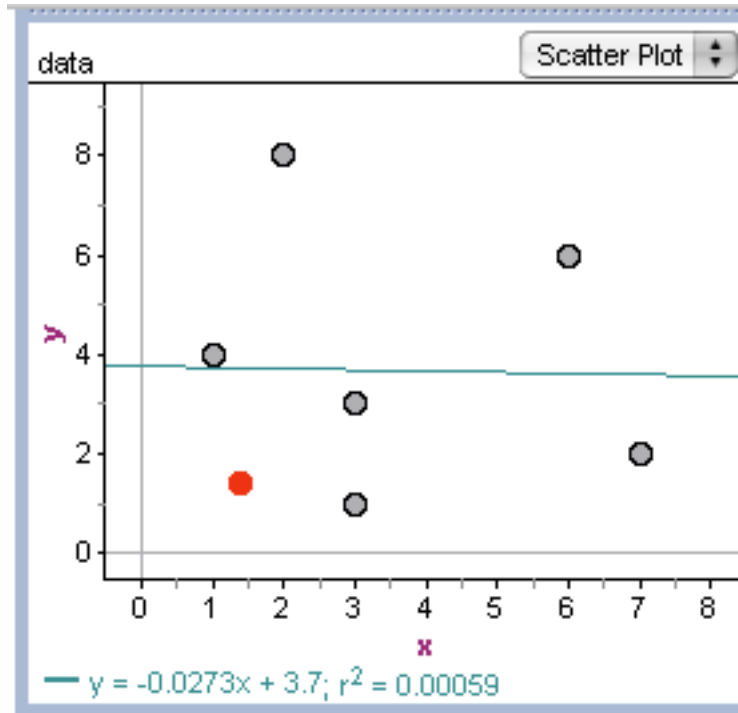


Figure 11. A Fathom scatterplot with several points moved to create a small r-squared.

Another challenge is posed to students: Under what circumstances is $r^2 = 1$? In this case, the manipulations are more complex, as there is no way to create a graph with very high correlation without moving several points. Figure 12 is one possible solution, a distribution with $r^2 = 0.92$; most of the points on the graph have been moved quite a distance from where they began.

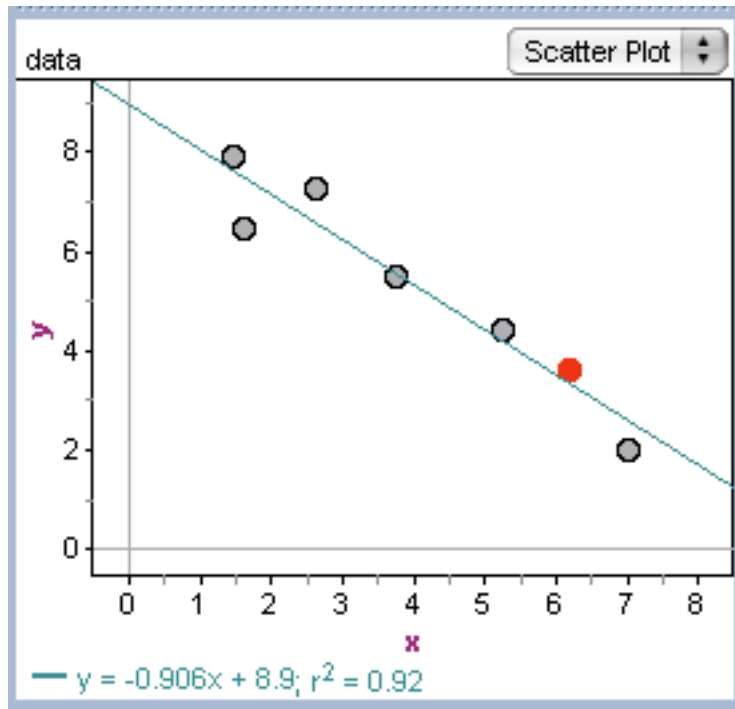


Figure 12. Fathom scatterplot with several points moved to create a large r-squared.

While all six of these figures (from both Stretchy Histograms and Fathom) capture the end result of students' manipulations of the data, they can't convey the kinesthetic aspect of interacting with data in this way. Seeing a line of best fit change as points are moved feels a bit like driving a car, although one where it's hard to predict where it's going based on how you turn the wheel. Moving a point back and forth several times to see the effect of its position is much more compelling than just seeing the beginning and ending states.

What of the concern about the status of data: when are data "real," and when are they the result of a user's modifications? We saw that Stretchy Histograms dealt with this concern by creating a different data structure that had its own rules, but couldn't be brought back together with unmodified data. Fathom has another take on this dilemma. There is no distinction in Fathom between original data and modified data – but there is a "Revert Collection" option that will restore the original values. Other software, such as Logger Pro, plots hidden data as grayed-out points, a reminder to the user of their existence and their current status. Of course, the user still has to decide to take advantage of these features, but even their existence are reminders that there is an original set of values that are being kept safe while the user is fiddling with the data.

6. PROBABILITY AND SAMPLING

Robust understandings of probability and sampling have always been difficult for students of statistics to develop, but many people have claimed that computers have the potential to make these ideas more accessible by being a “probability workhorse.” In 1992, I wrote

Teachers have often approached the task of explaining the behavior of samples by repeating random processes, such as rolling dice, many times. But after a few hundred rolls, students (and teacher) are bored with the process. And understanding how samples behave in different populations requires even more sampling: more tedious dice rolling. Computers are an obvious answer here because they can take many samples of arbitrarily weighted populations and keep track of them in a variety of ways. This power and flexibility give students the tools to do many sampling experiments and compare the results, giving them the data from which they might construct an understanding of sampling.

I described in that paper several approaches to using computers to support students in learning about sampling and probability. Since then, it has become clear that having a piece of software that creates sampling distributions quickly may be necessary for students to understand sampling, but it is far from sufficient. Some of the difficulties are matters of curriculum development, but there have also been some breakthroughs in software development that make effective curriculum easier to develop.

One of the most interesting new pieces of software is actually a descendant of one that I described in 1992. As part of the early work on TinkerPlots, Konold et al developed a Probability Simulator with which students could build and “run” probability models, which produced distributions based on random number generation. Giving students the tools to be explicit about their models of probability and sampling provided teachers with information about the strengths

and weaknesses of their thinking. Here is an example from the previous paper of the kind of difficulty that students have in building probabilistic models.

Students were given the following problem: “According to a report from the Centers for Disease Control, the chances of getting the flu during the next 4 months is .30. What is the probability that at least 1 of 6 of your teachers will get the flu?” Many students successfully built models but this particular problem caused considerable trouble for others, who modeled it as drawing 6 times without replacement from a box that contained 2 elements (F and W) in the ratio 3 to 7. They apparently were thinking of the elements in the box as teachers, rather than as traits, and drew without replacement because they were concerned about drawing the same teacher twice.

As TinkerPlots has developed, the designers have pursued a vision in which exploring data and exploring probability have not been kept separate, as has been the tendency until now. Since the tools that TinkerPlots provides have been well-received and integrated into several curricula, using the same structures to analyze the output of probabilistic models would appear to have great potential. In the past two years, the Model Chance project has explored the results of joining data analysis and probability generating tools and the early returns are promising. Here’s an example of how the new probability capabilities of TinkerPlots work in a middle school classroom setting.

In the two dice problem, students are asked to create a distribution of the sum of the results of a roll of each die. So, for example, one possibility is “2,” which would result from a roll of 1 and 1; another is “7,” which would result from a roll of 3 and 4 (or several other possibilities.). What shape will the distribution have? (Answer this for yourself before reading on.)

Most pieces of software built for sampling simulations can handle such a problem easily, but there are some interesting characteristics in the TinkerPlots approach that grow from a different vision. Simulating this problem in TinkerPlots starts with building a generator that resembles as

much as possible the physical act of sampling. Figure 13 is one possible way to build a generator for the two dice problem:

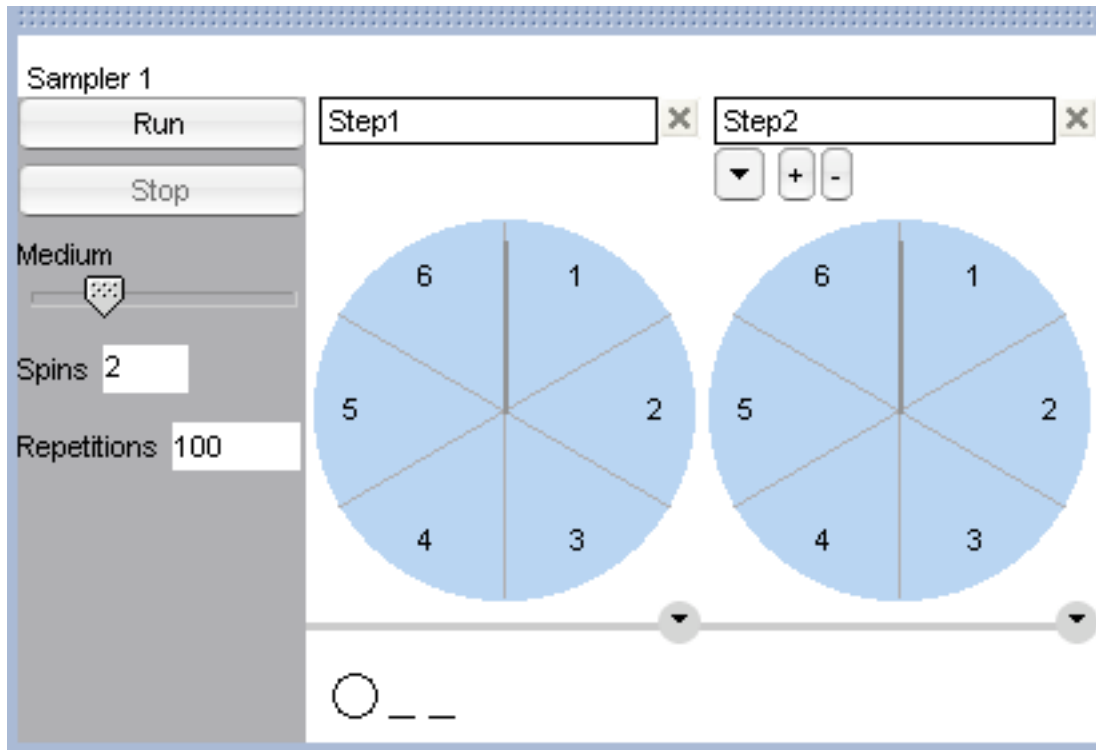
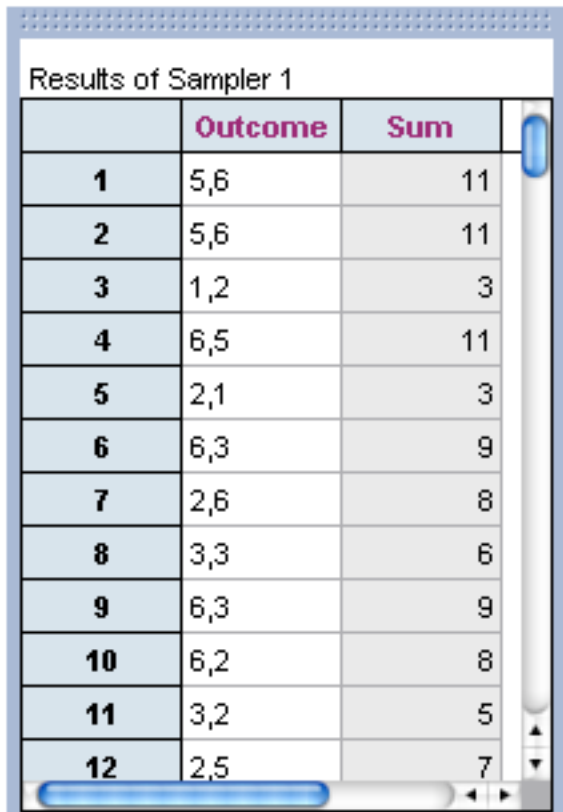


Figure 13. One possible TinkerPlots generator for the two-dice problem .

In the sampler above, there are two spinners, each representing one die. Each spinner has six slices of equal area marked with the numbers 1 – 6 and they are hooked together so that one spins, then the other. The output of this pair of spinners (pair of dice) is a table of pairs, such as (6,1) and (3,4). Up to here, the software mimics many other probability simulations, except that the process of setting up a sampler resembles the act of throwing one die and then the other more than most other simulations.

At this point, the student has a data table of 100 pairs, but nothing else. She has lots of options: she could graph the first roll, the second roll, the sum of the two rolls, even the difference between them. Solving the two dice problem requires using the “Sum” function, just as one would in an ordinary table, as in Figure 14. We are now living in what looks like a “data analysis” world, but we are working with the output of a probability simulation. TinkerPlots graphing functions are now available, so Figure 15, a frequency graph of Sum, is easily created.

(The shading in the graph indicates the value of Sum, with darker being larger; in this case, the color of each dot is redundant with its position.)



The image shows a screenshot of a data table from a software application. The table is titled "Results of Sampler 1" and contains 12 rows of data. The first column is labeled "Outcome" and the second column is labeled "Sum". The "Sum" column is shaded light gray. The table is displayed within a window that has a blue title bar and a scroll bar on the right side.

	Outcome	Sum
1	5,6	11
2	5,6	11
3	1,2	3
4	6,5	11
5	2,1	3
6	6,3	9
7	2,6	8
8	3,3	6
9	6,3	9
10	6,2	8
11	3,2	5
12	2,5	7

Figure 14. "Data" output of a probability generator in TinkerPlots

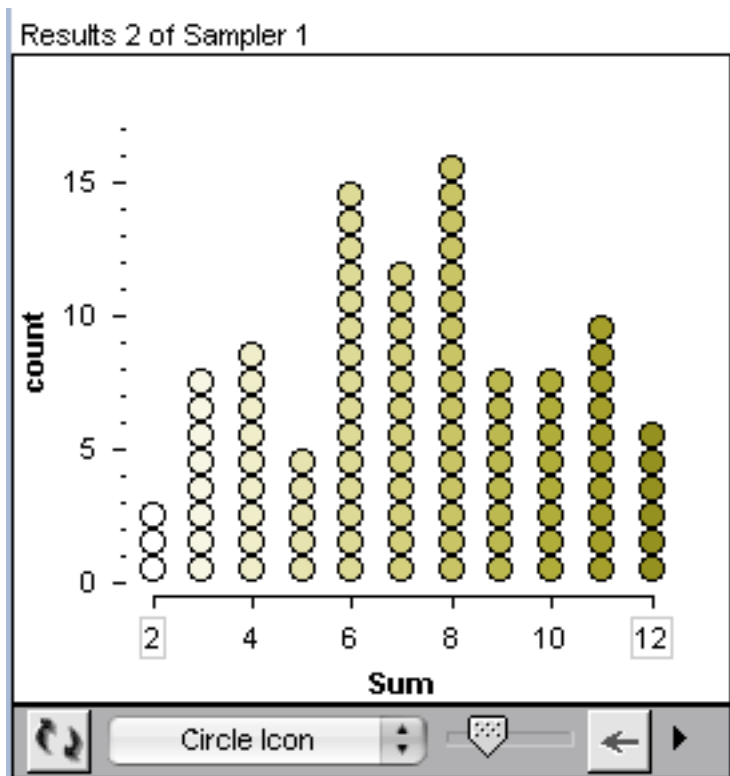


Figure 15. Plot of probability generation results in TinkerPlots.

Given the resources that TinkerPlots makes available, there are other graphs just as easy to create as the simple frequency distribution of Sum. One of the concepts we hope students will understand early in their statistical education is that often results (in this case, the sum) can be obtained in different ways. In this case, the sum of 8 can result from the following throws: (4,4), (3,5), (5,3), (2,6) and (6,2), whereas a sum of 2 can only result from a throw of (1,1). Understanding that difference goes a long way toward understanding the shape of the graph, higher in the center, lowest at the ends. But what if we wanted to see relatively simply how many different ways there are to get each Sum? In TinkerPlots, it is possible to “color” the icons according to a different variable than the one by which they are graphed, as in Figure 16. So it is easy to create the graph above just by coloring the icons in the frequency distribution of Sum according to the different pairs that created the sum.

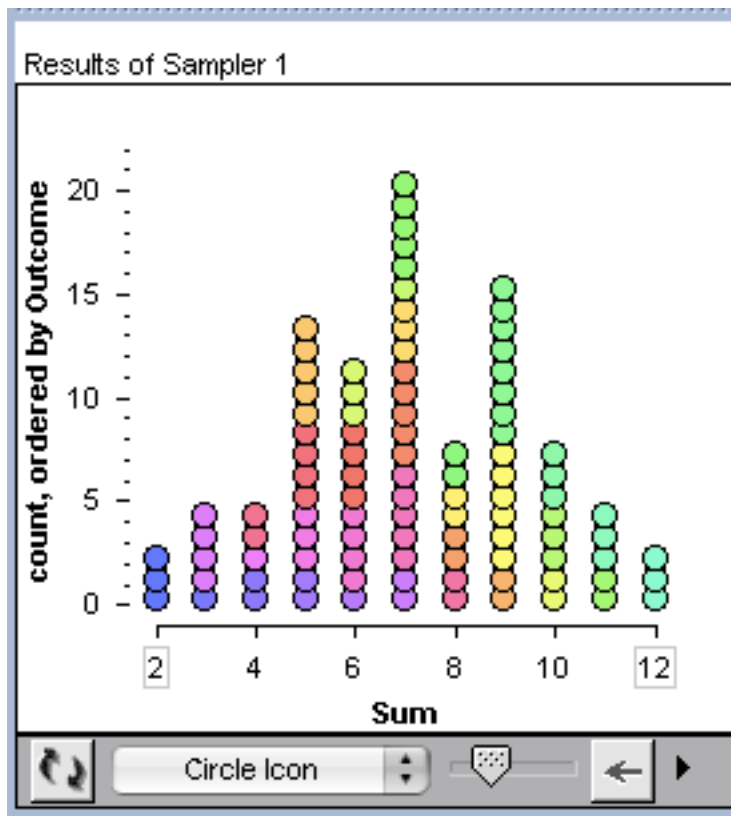


Figure 16. Colored icons in Tinkerplots can help represent the structure of a distribution.

While it's not possible from this graph to know what pair each color represents, we can learn a few interesting things: the icons above 2 and 12 have only 1 color, meaning that just one pair of values that add up to either 2 or 12 occurred (and perhaps there is only one pair, but we can't tell that from this graph.) In contrast, 3 and 11 each have TWO colors of icons in their stack, so there are at least two possible ways to get each. So, even without knowing exactly what the individual pairs are, we can get a quick sense of the number of pairs that can result in each Sum.

What's hidden in the above description is just how simple the process of modeling through simulation is in TinkerPlots. Middle-school students routinely take on problems such as this and many that are more complex. It appears from early results that considering and representing sampling as the process of generating data is more accessible to students than the idea of choosing from a sample space. While integrating probability and data analysis tools is not a completely novel idea, TinkerPlots is pushing the envelope further, using new metaphors and

new technology. In the next decade, the impact of these changes will play themselves out; perhaps most high school students will develop probability literacy.

7. REFLECTIONS AND REACTIONS

In 1992, I described several concerns about the effects of using computers in statistics education. Interestingly, none of these has disappeared and probably never will, given that they arise from deep issues about the ways in which people use computers. Being explicit about them at least raises the probability that software developers will be as careful as possible not to foster these problems, and curriculum developers will be as watchful as possible as they develop materials.

As mentioned above, computers may make graph construction too easy, robbing students of the struggle to determine axes, scales, and other aspects of graph design and limiting their options to a few standard graphs. While software today provides many more design choices, it is still important that students spend time creating their own graphs with lower-tech tools such as paper and markers.

In my previous paper, I noted that students sometimes think that computers have special access to mathematical truth and are unlikely to question anything the computer produces. While students now are more computer-savvy, this problem has not gone away. This attitude can lead to students accepting implausible results, such as probability distributions that violate their expectations. With current technology, though, students can see how changes in graphs highlight different interpretations, how truly random data differ from one run to another, and how seemingly small changes in a probability generator can lead to more significant consequences. With these experiences, students have a better chance of looking critically at their own results and at what they are fed by the media. I've heard several students wonder if they've screwed up when a distribution is not what they expected; it would be better if they considered a larger set of possibilities; yes, they may have screwed up, but they may also have done everything right and be seeing a relatively unlikely result. In the end, there's a delicate balance to be struck here, given the uncertainty that is at the heart of statistical reasoning. We want students to question

what the computer generates, but not to reject results simply because they are not within the expected probability.

Another way in which students may over-believe computers is in interpreting best fit lines and curves. A fit function is only as good as its residuals, but students sometimes argue that if the software puts a line on the graph, it means there is a correlation. For example, in the TapeMeasure experiment described above, several students argued that there was a correlation between weight and running speed, even though the points were almost randomly scattered. Data analysis software is more likely to include R^2 on graphs with fitted curves and to provide an easy way to create a residual plot (Fathom does both of these, for example), but the tendency to see relationships because the computer can calculate them quickly will not disappear.

But even if software developers address these concerns – as they have done to a large extent – no software will ever teach statistics without a curriculum and a teacher. All of the software described here are tools to be used in the context of a well-thought-out idea of what students should be learning, a corresponding set of activities, and enough insight on the part of the teacher to deal with the inevitable sticking points. The good news is that data analysis and statistics are becoming more legitimate and visible parts of the math curriculum, so publishers are willing to invest in curriculum development that makes good use of software. Both Fathom and TinkerPlots have curricula built around them and there are several mainstream statistics textbooks for high school and college that suggest valuable ways to incorporate technology into a course. (Rossman, Chance and Lock 2001; Watkins, Schaeffer and Cobb 2006; DeVeaux and Velleman 2004; Yates, Moore and Starnes 2007).

So where have we come and where should we go? Interestingly, much has changed – but little has changed. The innovative ways of using technology in statistics education today are closely related to those of 15 years ago: direct manipulation graphing software, ways of modifying hypothetical data distributions to explore statistical relationships, understanding of sampling supported by students building probability generation models. Software developers have built on the successes and struggles in these areas to push the field further. We have not yet taken advantage of more “exotic” forms of data, especially those that come from new media or even

from virtual reality, but these developments will surely happen – even if we can't visualize what they will look like at this point.

Where we need new work is in helping a teacher who is faced with thirty students working in parallel at fifteen computers, where she is only able to spend a few seconds at a time with any one student. Tables, samplers, graphs and other data structures come and go so quickly that by the time a teacher responds to a student, she may have done something to change the state of her screen. At the end of a class period in the computer lab, teachers know little about what their students did during class. Of course, that is always the case in a classroom, but the amount of student's thinking that has "disappeared" can be much greater when students have the kind of software described above. I am making a small start on this problem by partnering with artificial intelligence researchers to create a tool that can record students' actions and present them to a teacher in a format that can help her understand what has gone on in the class, on both the aggregate and individual levels. The future of this effort is far less clear than those described above; it is largely uncharted territory. We'll know more for the next iteration of this paper in 2022.

REFERENCES

URLs for software referenced in this article:

Alberti's Window www.albertiswindow.com

Fathom www.keypress.com/fathom

InspireData www.inspiration.com/productinfo/inspiredata

Logger *Pro* www.vernier.com

Measurement in Motion <http://www.motion.com/products/measurement/index.html>

MyWorld <http://www.myworldgis.org/>

TinkerPlots www.keypress.com/tinkerplots

ESRI www.esri.com

GLOBE: www.globe.gov

GOOGLE Earth: www.earth.google.com

LenoxSoftWorks: www.lsw.comn

Australian Teenager Data from <http://www.censusonline.net/sampler2.html>

A survey of high school students in South Australian secondary schools (2003). Downloaded 6/2/03.

DeVeaux, R. D. and Velleman, P. (2004) *Intro Stats*, first edition, Boston: Addison Wesley.

Hammerman, James K., and Rubin, Andee (2004). "Strategies for Managing Statistical Complexity with New Software Tools." *Statistics Education Research Journal*, 3(2) pp. 17–41.

National Council of Teachers of Mathematics (2000). *Principles and Standards for Teaching Mathematics.*, second edition. Reston, VA: NCTM, 2000.

Rosebery, A., and Rubin, A. (1989). "A Computer-Enhanced Approach for Developing Statistical Reasoning." *The Journal of Mathematical Behavior*, Vol. 8, No. 2, August 1989, pp 205-220.

Rossman, A., Chance, B., and Locke, R. (2001). *Workshop Statistics: Discovery with Data and Fathom*, Emeryville, CA: Key College Publishing.

Rubin (2005) "Math That Matters." *Threshold* Vol 3, No.1, Allston, MA: Cable in the Classroom, pp 22 - 25. www.ciconline.org/threshold.

Rubin, Andee, & Hammerman, James K. (2006). "Understanding Data Through New Software Representations." In Gail Burrill (ed.) *Thinking and Reasoning with Data and Chance—2006 NCTM Yearbook*. Reston, VA: National Council of Teachers of Mathematics.

Watkins, A.E., Schaeffer, R. and Cobb, G. (2006). *Statistics in Action: Understanding a World of Data* Emeryville, CA: Key College Publishing. Second Edition. Fathom supplement by Bill Finzer.

Yates, D., Moore, D., and Starnes, D. (2007). *The Practice of Statistics*, Third Edition. New York: W.H. Freeman and Company. Fathom supplement by Tim Erickson.