Visualizing Change

An Innovation in Time-Series Analysis

Stephen Few, Perceptual Edge

Visual Business Intelligence Newsletter, September 2007

(This paper was written as a white paper for SAS Institute)

The paths that our lives take and the speed of passage from one place to the next are difficult to retrace. The story of a life is a story of change as one navigates, in the present, what immediately becomes the past and is mostly forgotten. Wouldn't it be enlightening if your life could be revealed as a picture, a map that displayed your entire journey on a single page, a record of the many intersections, turns, paths, and events that have become lost in the haze of fallible memory? On the pages that follow I examine this problem of tracing change through time, not in terms of our personal journeys, but in terms of the journeys that reside in our data. Those of us who work with information, striving to uncover and understand the stories it has to tell about our businesses or what's happening in the world, struggle to bring these stories to light.

> "Visual representations of data take advantage of the unique ability of visual perception to detect meaningful patterns that might otherwise remain hidden."

I spend my days helping people make sense of data and then present to others what they find using visualization techniques. Visual representations of data take advantage of the unique ability of visual perception to detect meaningful patterns that might otherwise remain hidden. Even highly skilled statisticians recognize when it makes sense to clear their heads of statistics and simply use their eyes to explore data. John Tukey, renowned Princeton statistician and author of *Exploratory Data Analysis*, recognized the important role of our eyes in the analytical process. He wrote:

> *The basic intent of data analysis is simply stated: to seek through a body of data for interesting relationships and information and to exhibit the results in such a way as to make them recognizable to the data analyzer and recordable for posterity.*
>
> *A picture is not merely worth a thousand words, it is much more likely to be scrutinized than words are to be read. Wisely used, graphical representation can be extremely effective in making large amounts of certain kinds of numerical information rapidly available to people.*
>
> *One great virtue of good graphical representation is that it can serve to display clearly and effectively a message carried by quantities whose calculation or observation is far from simple.*
>
> (John W. Tukey, *The Collected Works of John W. Tukey.* Belmont, CA: Wadsworth Inc., 1988)

Since Tukey penned these words, a great deal of attention has been given in academic circles to visual means for exploring and making sense of data. As they have learned more and more about visual perception—how it works and how it can be applied to data analysis—they have developed tools that exploit this potential.

Visual perception and effective thinking are intimately connected. This is why so many of the terms that we use to describe thinking and understanding use visual metaphors, such as "I see." Colin Ware of the University of New Hampshire eloquently describes the significance of visual perception:

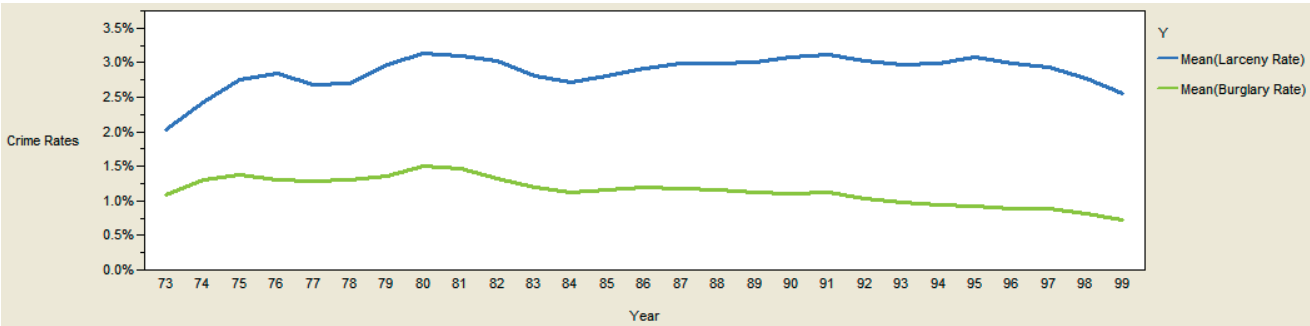
> *Why should we be interested in visualization? Because the human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers. At higher levels of processing, perception and cognition are closely interrelated, which is the reason why the words 'understanding' and 'seeing' are synonymous. However, the visual system has its own rules. We can easily see patterns presented in certain ways, but if they are presented in other ways, they become invisible…The more general point is that when data is presented in certain ways, the patterns can be readily perceived. If we can understand how perception works, our knowledge can be translated into rules for displaying information. Following perception-based rules, we can present our data in such a way that the important and informative patterns stand out. If we disobey the rules, our data will be incomprehensible or misleading.*
>
> (Colin Ware, *Information Visualization,* Second Edition. San Francisco, CA: Morgan Kaufmann Publishers, 2004)

"It is not enough to focus on what's happening today. We must see what's happening in the context of history to understand it fully."

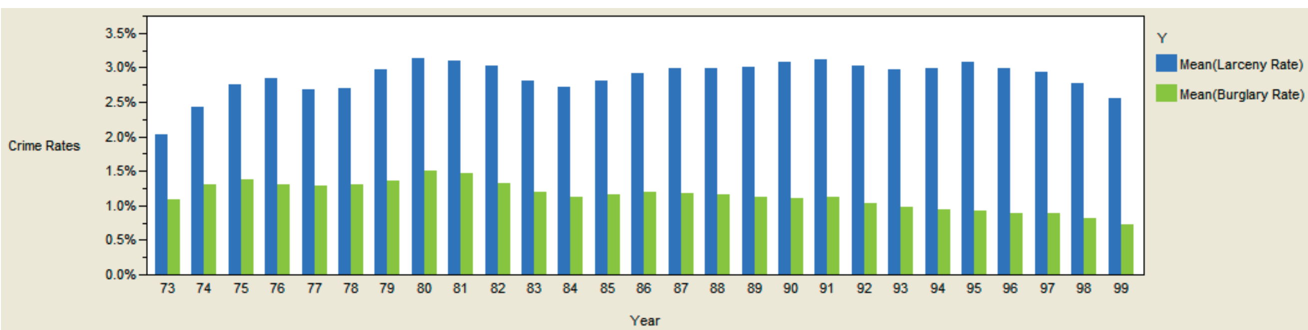
Years before Ware wrote these words, Tukey dreamed of a day when technology would allow us to take full advantage of our eyes, not only by displaying data visually on a computer screen, but by allowing us to interact with those visualizations in ways that help us think more effectively and comprehensively about the data, fully immersed in exploration. Tukey would have taken pleasure in modern software tools for visual exploration, such as JMP®, a statistical analysis product from SAS that makes effective use of analysts' eyes.

Typical Time-Series Analysis and Its Limitations

Many of those who dedicate themselves to information visualization research today spend a great deal of time looking for better and better ways to examine quantitative data as they change through time. A recent study reported that 70% of all business graphs display time-series information. It is not enough to focus on what's happening today. We must see what's happening in the context of history to understand it fully. Is crime going up or down? To what degree did violent crime vary in relation to overall crime during the year? Questions like these can usually be answered using line graphs to examine the nature and extent of change through time.



Lines do a wonderful job of displaying the ups and downs of change through time. Notice how much easier it is to see the shape of change in the graph above, compared to the bar graph below. The strength of a bar graph is its ability to feature individual values and support comparison of one value to another, but the overall shape of change gets lost in the forest of bars.



There are other important questions, however, that cannot be examined using line graphs because they involve quantitative relationships other than change through time alone. For example, if you wish to examine the correlation between income and health in America today, a scatterplot could be used to gain insight into this relationship. But what if you also wish to see how this relationship has changed through time? Has the health of the poorest Americans declined over the years, or is it improving? One way to view this involves a graph that is dynamic, in which values actually move to represent change through time.

Displays that use animated graphics in this manner to reveal change through time are not unprecedented, but the new version of JMP does this and a great deal more to enhance what we can see and understand. Even though an animated scatterplot alone could be used to examine the changing relationship of household income and health through time, it could not help you compare values at different points in time because you would be forced to view each point in time as a separate sequential instance. JMP has solved this limitation, however, by adding a feature called "trails." With this feature, you can see the full path that values take from one point in time to another, all at once. This allows you to not only see degree of change from one point in time to the next, but also the shape, velocity, and direction of change. This ability leads to valuable analytical insights.

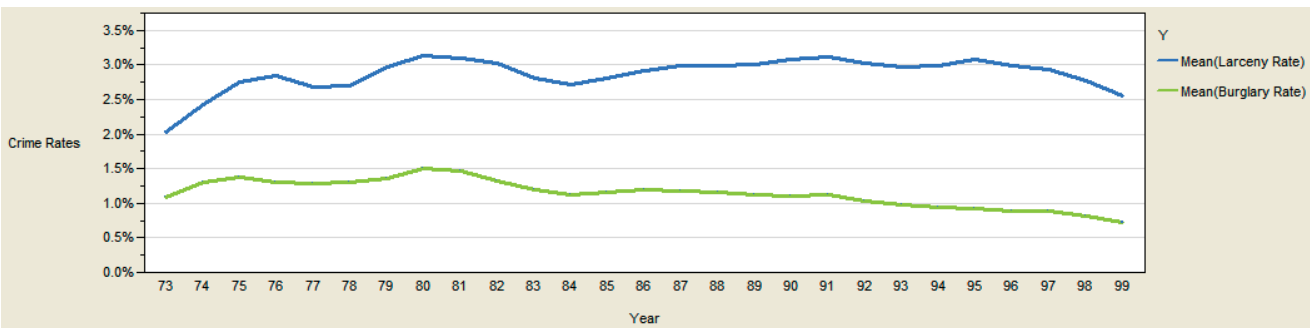
Meaningful Characteristics of Change through Time

Patterns of change through time can take many forms, many of which are meaningful. These patterns are formed primarily by combinations of four characteristics:

- Magnitude of change
- Shape of change
- Velocity of change
- Direction of change

Magnitude of Change

Magnitude of change is the difference between measures of something taken at two points in time. In the following graph, the magnitude of change in the burglary rate from 1973 to 1974 was approximately +0.3%.

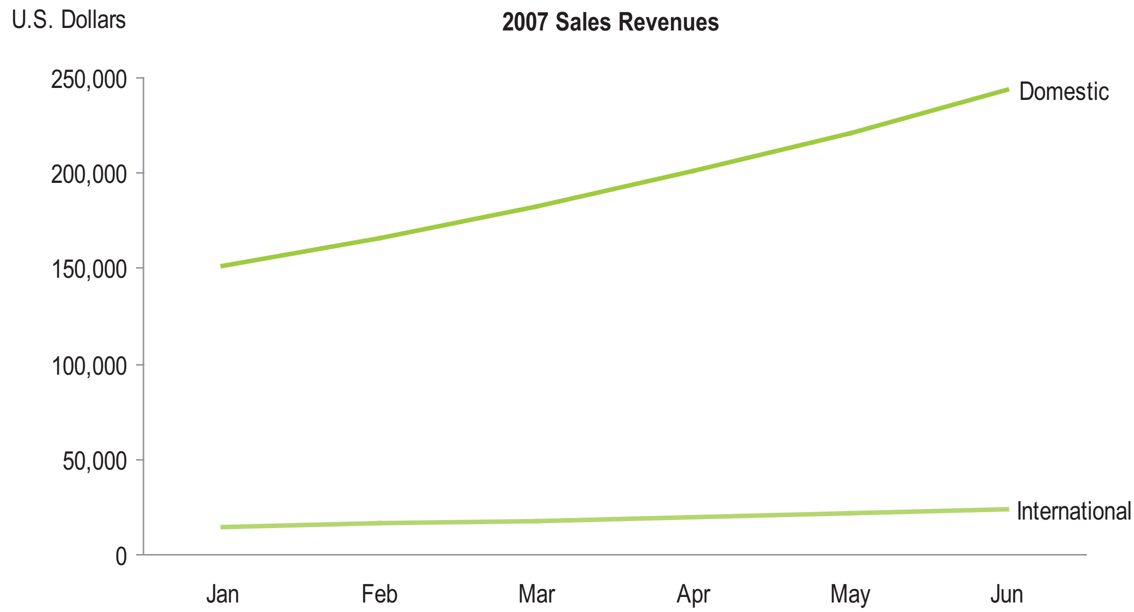


Shape of Change

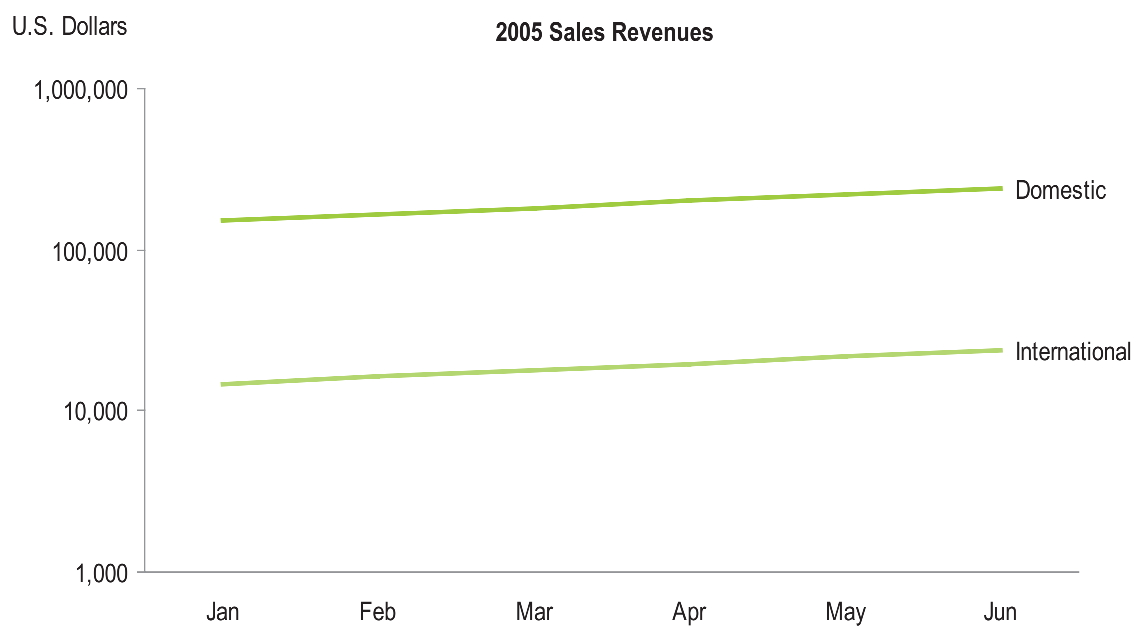
The path that is taken by the value of something as it varies through time, moving up and down, right and left, or both, is what I mean by the shape of change. In the graph above, because time moves from left to right at a constant rate, the shape of the burglary and larceny rates are formed only by the up and down slopes of the lines from one year to the next or across longer periods of time. The burglary rate (green line) exhibits less variation in shape than the larceny rate (blue line). Both exhibit a rise and a fall between the years 1978 and 1984. Their overall trends, however, are different: the burglary rate has decreased, while the larceny rate has increased from 1973 to 1999.

Velocity of Change

In this context, velocity refers to the speed or rate at which changes occur. As time passes, change can crawl slowly like a baby, it can sprint like an Olympic runner, or pace itself moderately. Notice in the example above how quickly the larceny rate increased from 1973 to 1975, compared to what appears to be a slower rate of change in the burglary rate. I said “what appears to be a slower rate” because when a line graph is use to compare rates of change between two or more data sets, equal slopes in the lines do not necessarily represent equal rates of change. For example, the graph on the following page displays two lines: one for domestic revenues (dark green) and one for international revenues (light green). Although it appears that domestic revenues increased throughout the year at a faster rate than international revenues, in fact both increased at exactly the same rate of 10%.

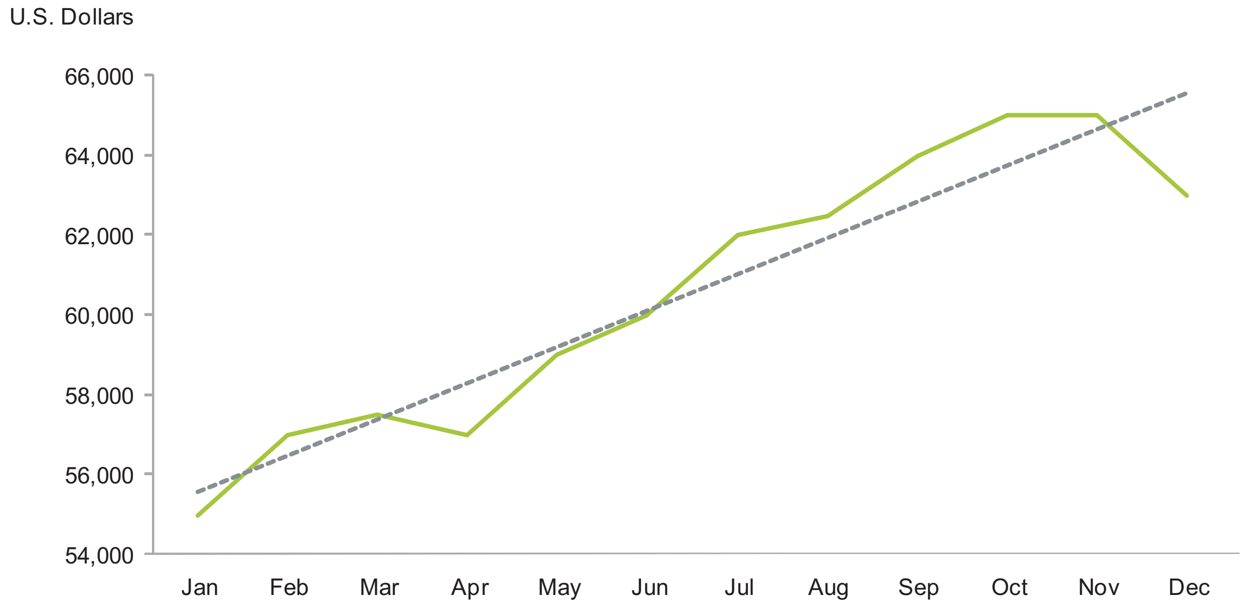


The slope of the domestic line is much steeper than the other because a 10% increase of large values such as $150,000, which is $15,000 per month, is greater than a 10% increase of much smaller values such as $15,000, which is only $1,500 per month. If you wish to compare rates of change easily and accurately, you can overcome the problem illustrated above simply by changing the quantitative scale from a normal scale to a logarithmic scale. With a logarithmic scale, two data sets that exhibit the same rate of change will also exhibit the same slope, even if the actual values of one set are much greater than the other. Below is the same sales information that appears above, this time displayed along a logarithmic scale.



Direction of Change

The overall or general direction of change in a series of time-series values is called the trend, which is often displayed in a graph as a trend line. During a specified period of time, have values mostly gone up, gone down, remained relatively flat, or gone both up and down to a nearly equal extent? Sales could be trending upward (shown as the dashed gray line below), even though last month's sales were down (the green line).



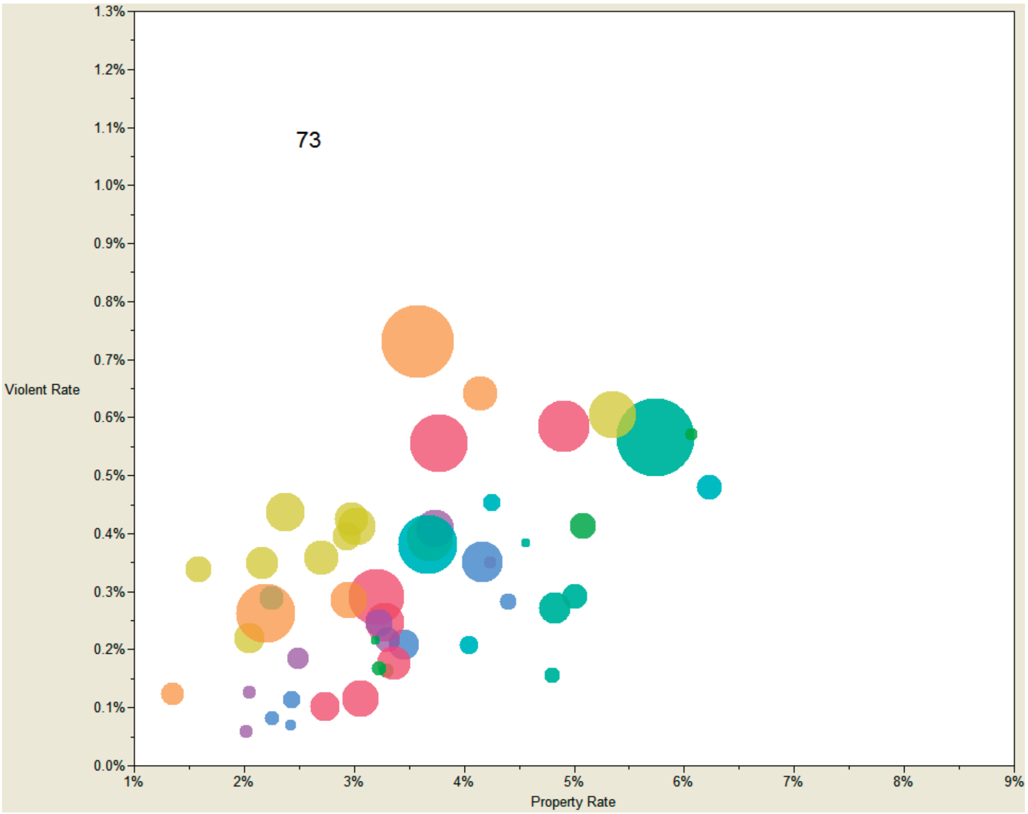
Every example so far has been displayed effectively in a line graph. Change through time was directly encoded as the ups and downs of position along a line. This works fine if you are only examining change through time, but what if you primarily want to examine another relationship in the data and only secondarily see how it changes through time? For example, what if you want to see if there is a correlation between property crime and violent crime rates, but also see if the nature of this correlation, assuming there is one, has changed

> "How can we add the variable of time to a scatterplot? The answer can be found in animation..."

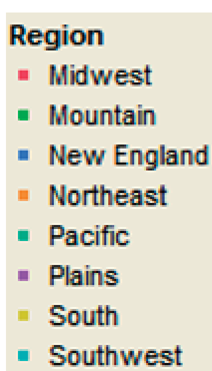
through time? A scatterplot provides the best means to display a correlation between two quantitative variables, but a scatterplot doesn't show change through time. How can we add the variable of time to a scatterplot? The answer can be found in animation; causing the data points in the scatterplot to move from position to position to trace the passage of time and the changes that accompanied it.

Motion as a Means of Displaying Change

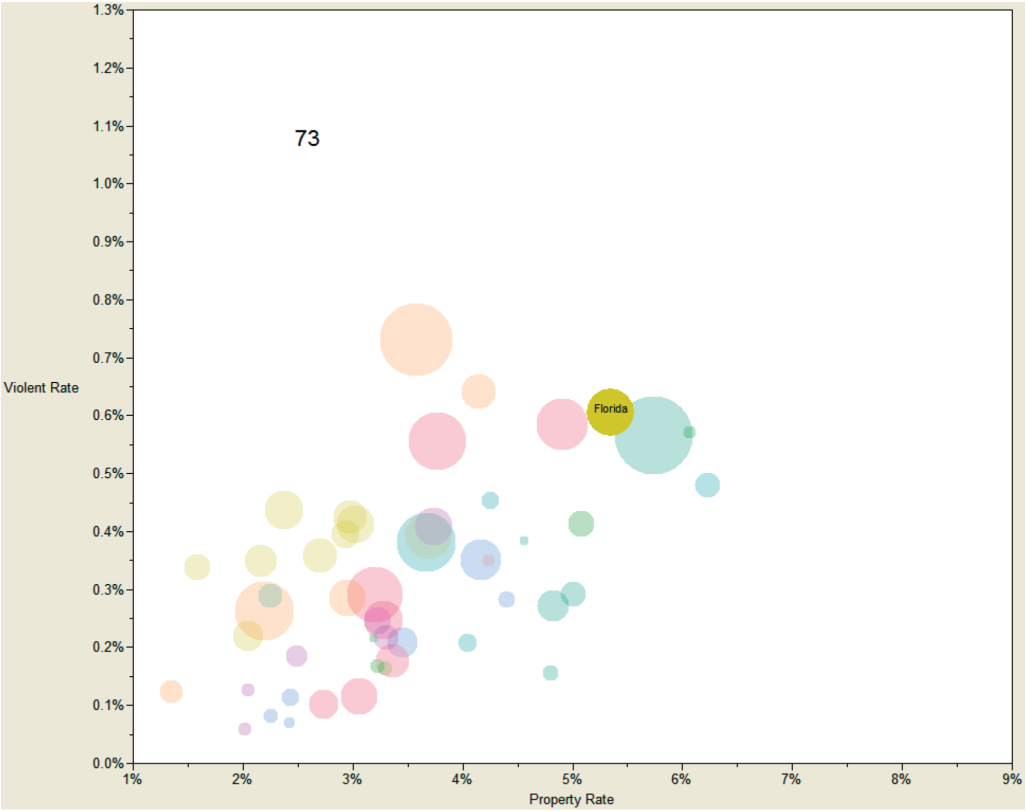
Consider the example that I proposed above regarding crime rates. Rather than a scatterplot, let's display the relationship between property crime and violent crime by state in the form of a bubble plot, using the size of each mark (circles in this case) to add a third variable—population—to the picture. To add even more meaning to the picture, color has also been used to group the states into geographical regions.



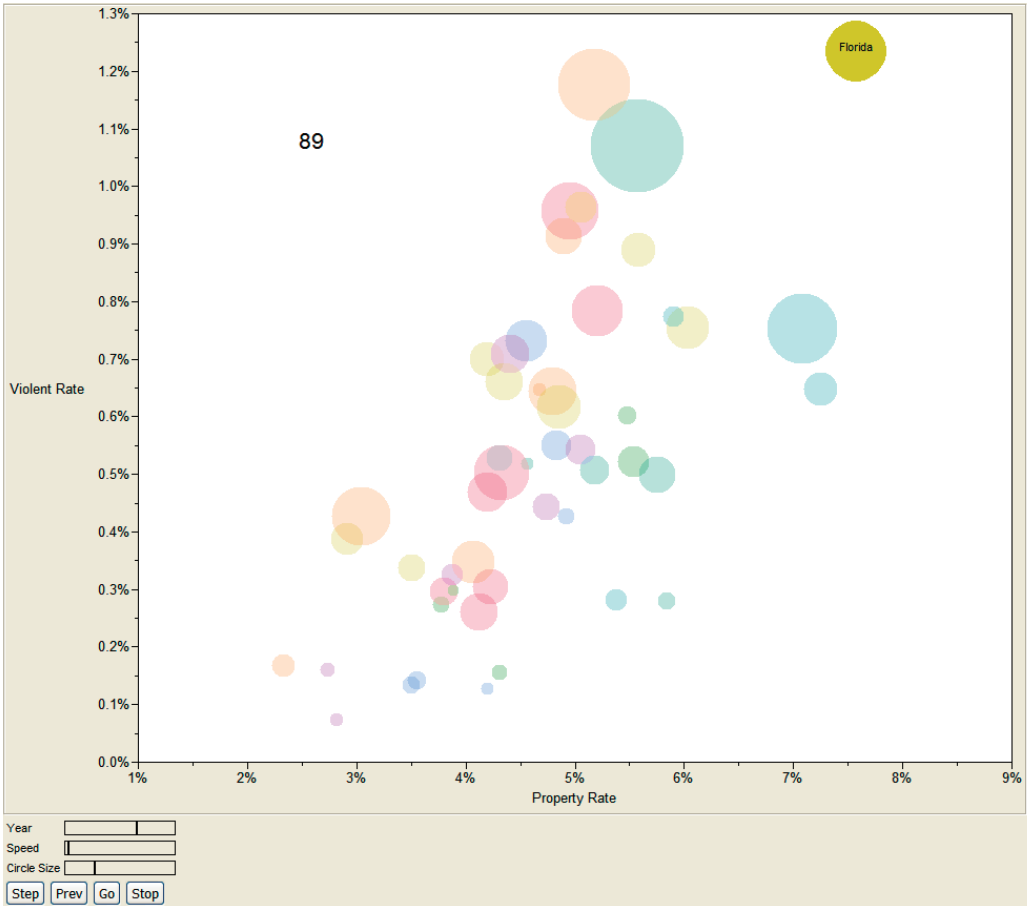
The number "73" that appears in the plot area indicates the year 1973. Here's a legend for decoding the colors that have been assigned to the regions:



We can see that in 1973, most states in the South had relatively low rates of property crime, below 3.5%, and violent crime rates below 0.5%. The one exception is the state of Florida, which I've made apparent below by highlighting and labeling the circle that represents it.



Noticing that Florida stands out among the southern states with higher rates of both property and violent crime led me to wonder: "During the years since 1973, does Florida continue to have exceptionally high crime rates?" With the graph on the following page, I used the slider control labeled "Year" that appears below the graph to march through time, keeping my eye on Florida to see if its crime rates ever ceased to be high. I discovered that they remained high throughout the years, reaching their apex in 1989, when they surpassed all other states.



Notice that below the sliders appears a series of command buttons labeled Step, Prev, Go, and Stop. These controls can be used to animate time, causing the data in the bubble chart to change year by year to display how these values changed through time. Pressing the Go button causes time to play across the screen, displaying change through time. This makes it possible for me to watch time unfold like a movie on the screen.

Utility of Sequential Motion for Visualizing Change

The use of motion in this way to represent changes that occur as time passes is very natural and effective. I can stop the passage of time at any moment by pressing the Stop button, I can go backwards in time by pressing the Prev button, and I can control the passage of time, pacing it one year to the next, by pressing the Step button. No means of displaying time is more natural than this, for it reproduces the chronological sequence of events that time itself previously produced. Much like the actual passage of time, however, viewing change in this sequential manner suffers from the same limitations that make the passage of time itself so difficult to remember.

Limitations of Motion for Visualizing Change

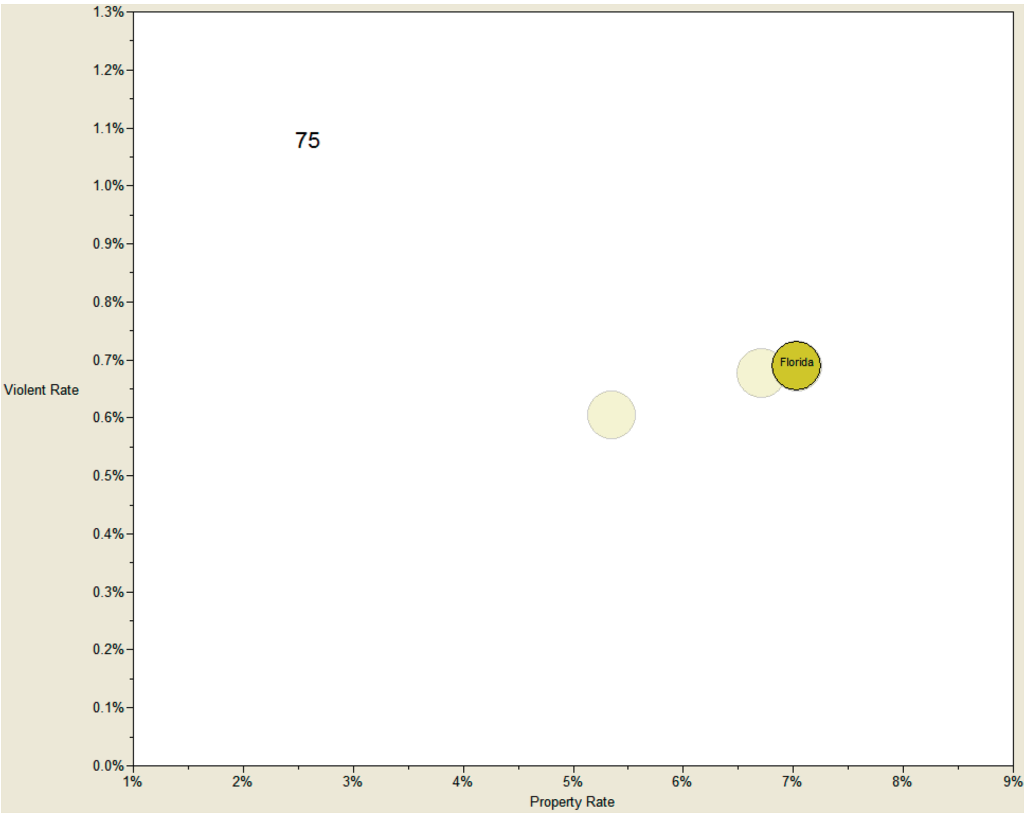
Despite the tremendous power of our brains, memory is severely limited. It is difficult to reconstruct the past from memory, to recapture the sequence of events and the nuance of each moment. As I watch the state of Florida bounce around the bubble plot as time passes, I can observe what's happening at each point in time, but afterwards I can only vaguely retrace the steps that unfolded before my eyes. The details were not stored in memory.

How can we overcome this limitation of memory? How can we examine the magnitude of change, shape of change, velocity of change, and direction of change that occurred during the passage of time? We need a picture of the entire path the state of Florida took as the years marched forward, laid out before our eyes from beginning to end so we can see the passage of time as a single image.

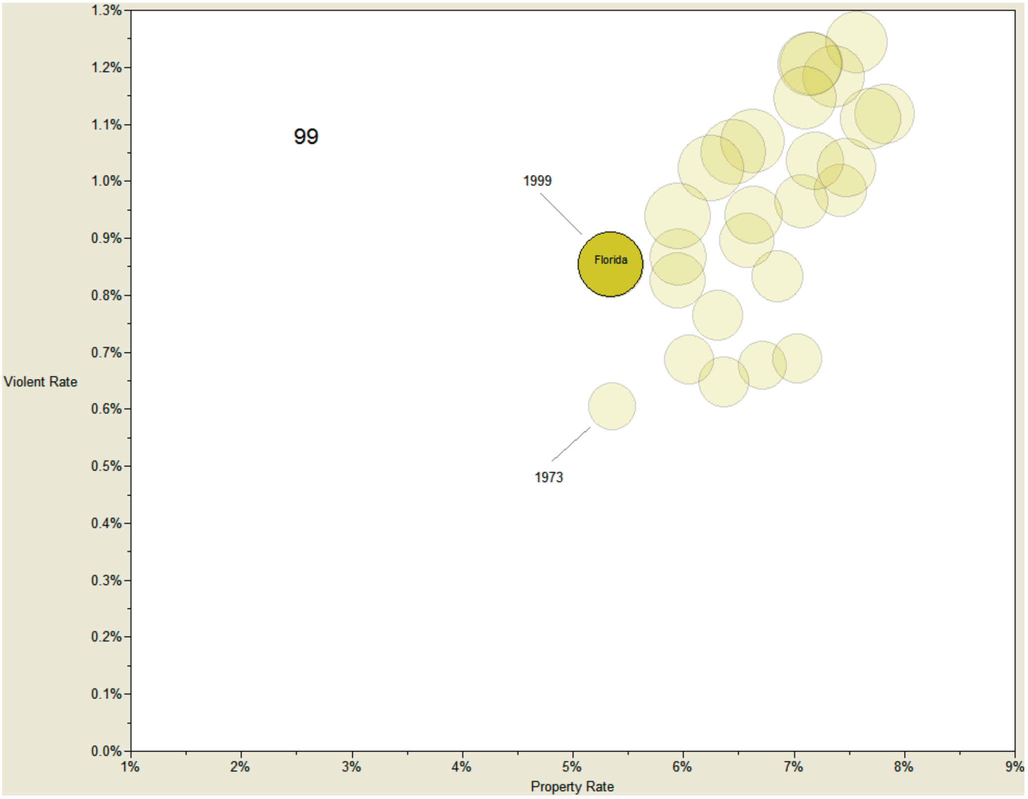
Trails for Overcoming the Limitations of Motion

A nice solution to this problem can be provided in the form of visual trails that capture and retain the state of change at regular intervals of time. JMP supports trails, which do a good job of overcoming the limitations of memory.

Let's return to the previous example. How have the state of Florida's property and violent crime rates changed from 1977 through 1999, the most recent year for which we have data? Let's see how trails can be used to reveal the nature of change. In the next example, I have removed all but the state of Florida from the bubble plot so we can see how its values have changed without distraction from the other states. Let's begin by looking only at the first three years' worth of change.



The darkest circle of the three represents the crime rate value in 1975. The three circles together can be viewed as bread crumbs that were dropped along a trail, one for each year. The leftmost circle represents 1973 and the one in the middle represents 1974. In other words, the path that change took from 1973 to 1975 proceeded from left to right and at an incline. Notice also that the velocity of change from 1973 to 1974 was high, which we can easily see as the distance between the first two circles. From 1974 to 1975, the velocity of change slowed down considerably. Now let's look at the entire period of time from 1973 to 1999.



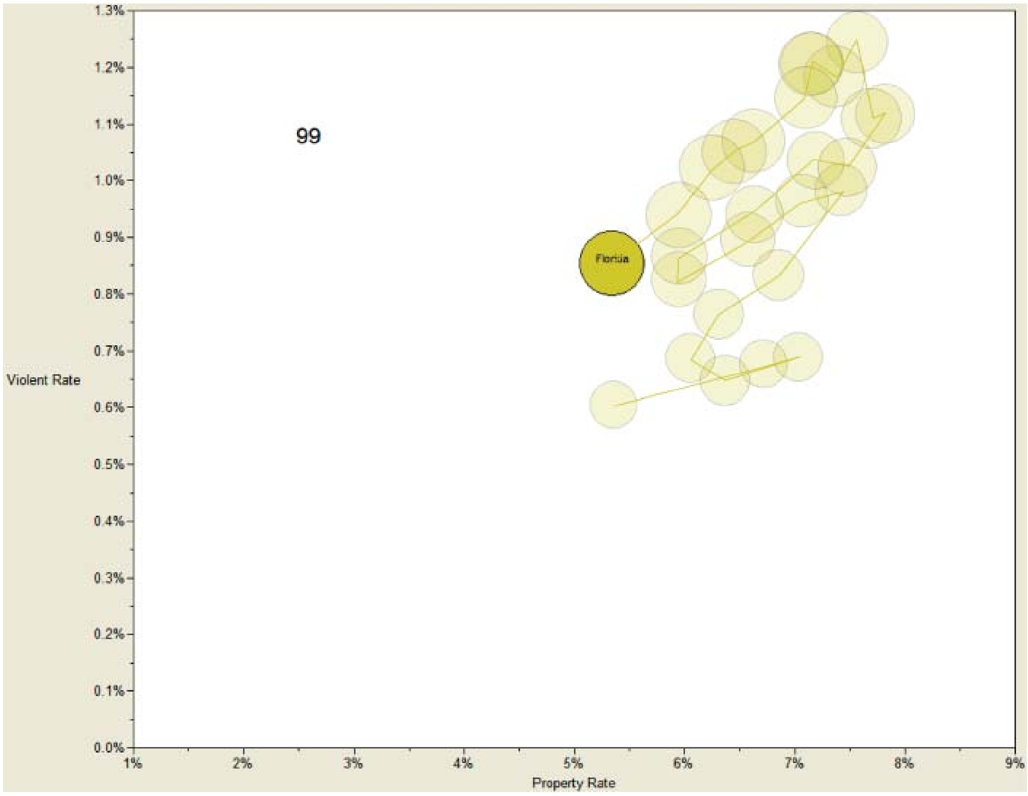
This displays, in a single picture, the story of property and violent crime across 27 years.

Trails for Viewing Magnitude of Change

We can compare the magnitude of change from one point in time to another simply by focusing on the difference between their locations and the distance between them. For example, although property and violent crime rates meandered significantly during these years, the magnitude difference in the property crime rate (the horizontal scale) in 1973 and 1999 is negligible, but the rate of violent crime increased from about 0.6% to 0.85%, a difference of 0.25%.

Trails for Viewing Shape of Change

We can see the shape of change by following the path formed by the circles sequentially from the first to the last. By playing the sequence of yearly steps one at a time, I was able to follow the course of change more easily than I could by looking at the finished trail above, but if a line were included that connected the center of each circle from one to the next in sequence, this path of change would be crystal clear. The use of a line for this purpose is currently not a standard feature of trails in JMP, but a script can be written to produce this effect with a little effort, illustrated below.



It moved to the right and up slightly in the beginning, ventured to the left a bit, took off at a steep incline for a while, declined in a leftwards direction, then took off again at a steep incline to its peak in 1989, and finally headed back down and to the left just as steeply until its final location in 1999.

Trails for Viewing Velocity of Change

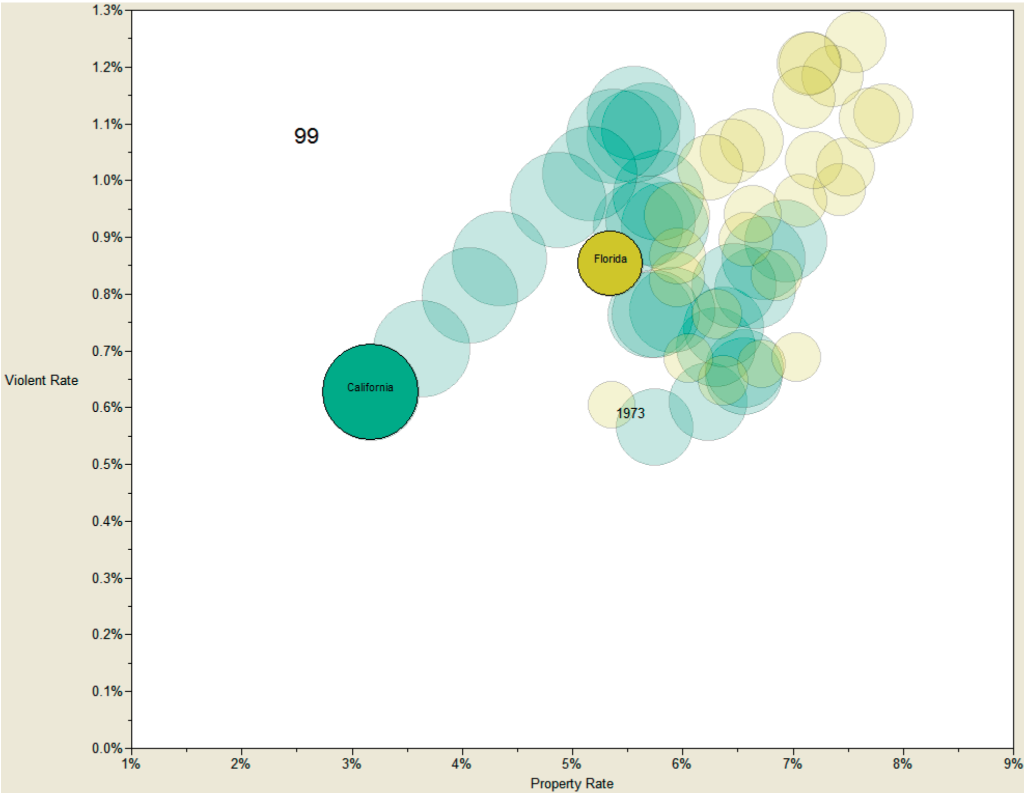
Trails encode the velocity of change as varying distances from one value (circle) to the next. The farther they are apart, the greater the velocity of change. You can see in the example above that change happened at a fast rate from the first to the second instance, but then it slowed down considerably during the next few years. At times velocity was very slow, while at other times it picked up the pace considerably.

Trails for Viewing Direction of Change

By looking at the example above, it is easy to see that at times change was trending upward and to the right (that is, values of property and violent crime rates were both increasing in positive correlation), and at other times the trend was exactly the opposite. During this entire period of time from 1973 to 1999, overall the trend was a slight increase in the rate of property crime and a much greater increase in the rate of violent crime.

Enriching the Story through Multiple Items and Variables

The examples that I've shown have featured a limited set of data. The picture and what we can do with it can be enriched, however, by looking at more data at the same time. For example, rather than focusing on one state only, we could compare the behavior through time of multiple states. In the following example, you can see full trails through time for both Florida and California.

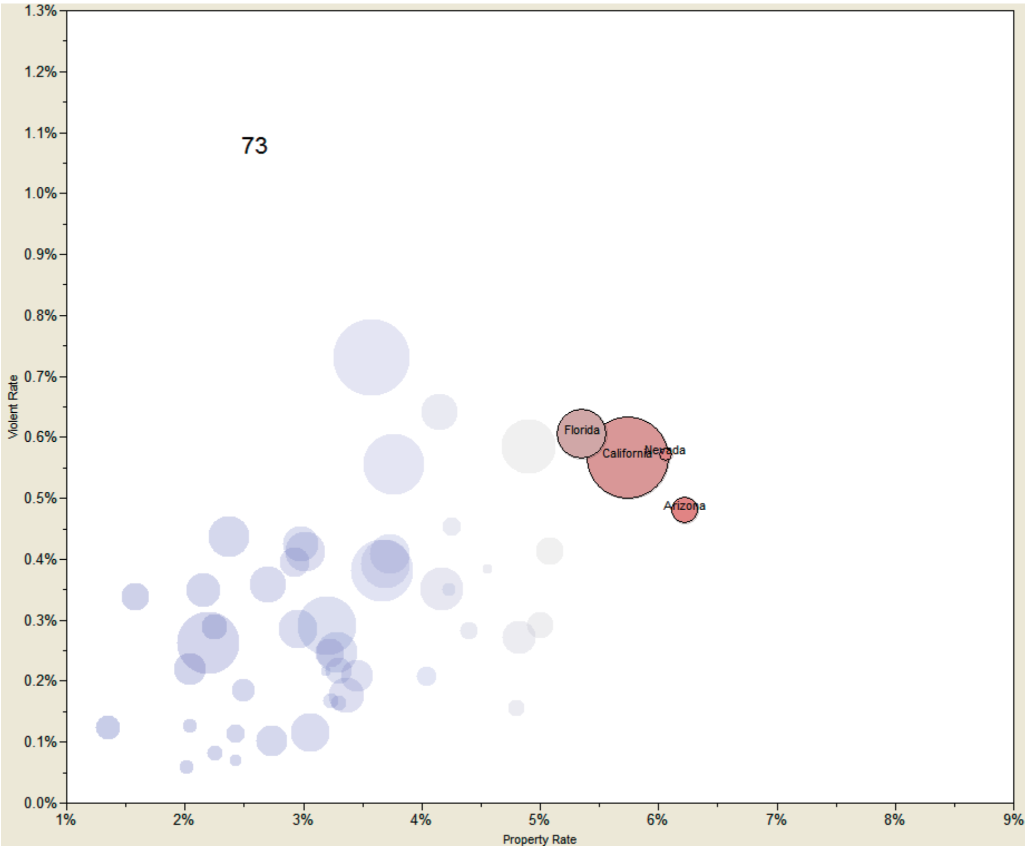


I've labeled the beginning 1973 values for both states so you can easily find the beginning of the trails. Although we might want to step through the sequence of change one year at a time to compare the magnitude, shape, velocity, and direction of change during particular periods, this overview alone allows us to see that the shapes of change for these two states appear to have much in common. It might be interesting to examine several of the states, one or two at a time, to see if any exhibit significantly different behavior.

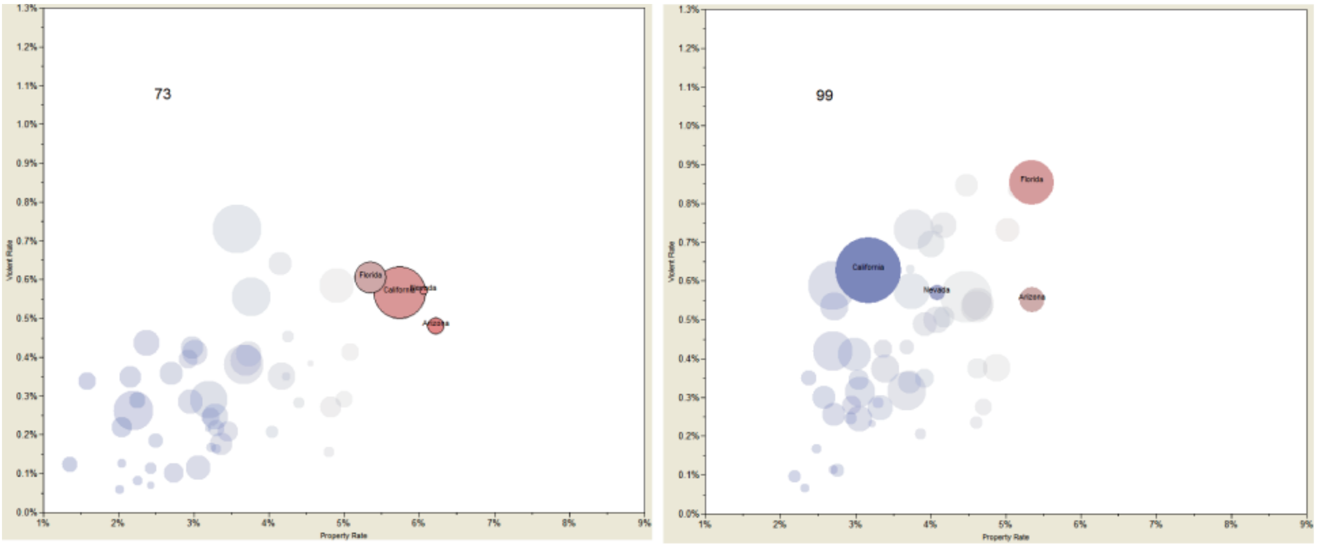
Another way to enrich the story of change through time is to display several variables simultaneously. The examples that we've been examining in the last few pages have perhaps been richer than you realized. They have all included five separate variables:

1. States, each represented by a different circle
2. Regions, each represented by a different color
3. Property crime rates, represented by position along the X axis
4. Violent crimes rates, represented by position along the Y axis
5. Population, represented by the size of the circle

By using several visual attributes that can be seen together, such as the 2-D location, size, and color of the circles, we can view several dimensions of the data at once. If we want to view another quantitative variable in addition to property crime rates, violent crime rates, and populations, we could easily use color to encode another quantitative value, rather than using it to identify regions. In the next example, I've done just that, using a sequential range of color to represent the total crime rate, ranging from dark blue for the lowest values through lighter and lighter blue as the values increase until they cross over into red, with the darkest version of red for the highest values.



What we can tell from looking at this version of the graph is that in 1973 the four states with the highest overall crime rates values (the red circles, which I've labeled) are also the states with the highest property crime rates, but not necessarily those with the highest rates of violent crime. Let's see if these four states with high overall crime rates as well as high property crime rates changed at all from 1973 to 1999. We can do this by comparing 1973 and 1999 versions of the graph, as shown on the following page.



Here we can see that total property, violent, and total crime increased in Florida, but the state of California improved in all but violent crime.

The stories that time-series data have to tell are often rich and important. They are much too important to remain unknown simply because we lack tools that can bring them to light. Tools such as JMP, using visualizations that animate time with trails, can do just that. We cannot retrace the important moments of our lives with precision and clarity, but with tools like this, we can keep the important moments in our data from suffering the same fate.

About the Author

Stephen Few has worked for over 20 years as an IT innovator, consultant, and teacher. Today, as Principal of the consultancy Perceptual Edge, Stephen focuses on data visualization for analyzing and communicating quantitative business information. He provides training and consulting services, writes the monthly *Visual Business Intelligence Newsletter*, speaks frequently at conferences, and teaches in the MBA program at the University of California, Berkeley. He is the author of two books: *Show Me the Numbers: Designing Tables and Graphs to Enlighten* and *Information Dashboard Design: The Effective Visual Communication of Data*. You can learn more about Stephen's work and access an entire library of articles at www.perceptualedge.com. Between articles, you can read Stephen's thoughts on the industry in his blog.