



Simple Displays of Complex Quantitative Relationships

Stephen Few

October 10, 2006

This article features the winning solution, submitted by Jock Mackinlay of Tableau Software, for the third scenario in the 2006 Data Visualization Competition.

This article is the third in a five-part series that features the winning solutions to our 2006 Data Visualization Competition. The third scenario in our competition asked participants to create a visual display that would enable real estate agents to monitor what's going on in the housing market.

Here's the scenario as it was described to participants:

As an analyst for a group of real estate agents, you want to create a visualization that will allow them to view several characteristics of house sales in a given month to help them better track and understand what's happening in the housing market. This group of agents deals with properties in five neighborhoods. You believe that they would gain meaningful insights if they could simultaneously examine several sales-related variables at once to make useful connections, so you want to display everything on a single page (but not necessarily in a single graph). You believe that each of the data items that appear below in the data section is significant.

I supplied participants with an Excel spreadsheet that reported individual house sales for the month divided into five neighborhoods, including the actual sales amount, the original asking price, and the number of days each house was on the market. Also, per neighborhood, the spreadsheet provided the median actual sales amount and the median original asking price for the month and for the same month of the previous year.

While reviewing the many solutions that were submitted for this scenario, in addition clarity of communication and ease of use, I was looking for a display that would allow people to see the following:

- Deviation of the actual sales amount from the original asking price.
- How long it took for houses to sell.
- Change from the previous year, both in actual sales amount and deviation from asking price.
- Comparisons of all these characteristics between the neighborhoods.

All of these characteristics displayed within eye span on a single screen or page.

The Winning Solution

Jock Mackinlay of [Tableau Software](#) submitted the winning solution for this scenario, which he created using Tableau 2.0, which has since been updated to the 2.1 release. Tableau Software is uniquely capable of displaying complex quantitative relationships, such as those featured in this solution.

Take a moment to look at Jock's solution (Figure 1) to determine for your self how well he succeeded in communicating this complex picture of the real estate market.

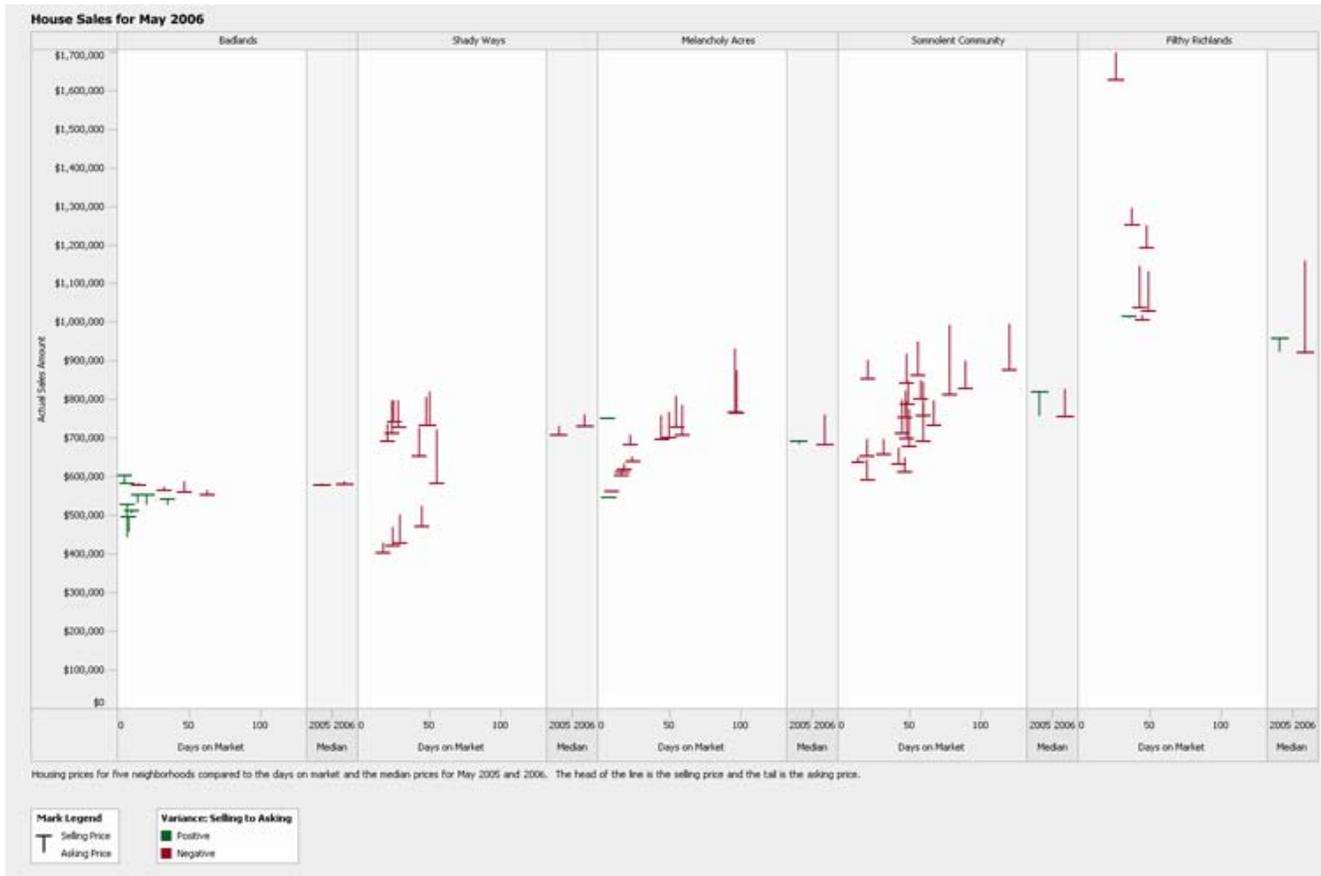


Figure 1: Jock Mackinlay of Tableau Software's winning solution.

Let's begin our examination of Jock's solution by allowing him describe it in his own words:

Real estate agents have a keen interest in the sales price of a property because that determines their commission. They are also interested in how long properties take to sell in a given neighborhood and the relationship between the selling price and asking price. These metrics normally correlate with the satisfaction of their customers. Agents want to get referrals and repeat business.

Our design visually represents the individual property sales for May 2006 as a small-multiple chart for the five neighborhoods. It supports the individual and comparative analysis of these neighborhoods. Each display compares a property's days on market to its sales price with the data shared between the small-multiples. We designed it this way because that is the most important data for real estate

agents. The design also allows us to incorporate the median prices for 2005 and 2006 for each neighborhood.

The marks are fundamentally Gantt bars. One end is the selling price and the other end is the asking price. The length is the variance. We also facilitate the reading of these marks by using color to encode positive and negative variance. The unconventional rendering of these Gantt bars emphasizes the selling price. The rendering also helps to distinguish marks that are adjacent.

Although we could have calculated and visualized aggregate statistics for this data, we have chosen to plot every property sale because real estate agents would find this very accessible. They would be able to find their individual sales in the visualization. They can scan the display to ascertain a general sense of the statistical distributions.

For example:

- *Badlands has a tight cluster of property sales that sell quickly and with a significantly positive variance when visually compared to the other neighborhoods. The median marks show little change from 2005. This is a good neighborhood except that the selling price is low when compared to the other neighborhoods.*
- *Shady Ways has two or three clusters with the lowest cluster below Badlands. This clustering would have been lost with views that used aggregate statistics. The variance is negative and pretty large. After a slight gap, the properties sell pretty quickly. People asked for more this year and got more. This looks like a good neighborhood for a real estate agent, if you can avoid the low cluster.*
- *Melancholy Acres has a reasonably tight selling price but a wide range of days on the market. As you would expect, the longer the property is on the market, the more the selling price drops from the asking price. The median asking price went up from the previous year, but the median selling price is down slightly. This is a good neighborhood for agents, if you avoid the slow selling properties.*
- *Somnolent Community has many property sales, but they vary widely in amount and the days on market. The variance is pretty negative. The median data shows a flip with the previous year. In 2005, people got more than they asked. In 2006, they asked for that amount, but got what was asked in 2005. This looks like a difficult neighborhood. On the positive side, it has many sales.*
- *Filthy Richlands does not have very many sales, but it has a higher commission. The number of days on the market is very predictable. People asked for a lot more in 2006 and got less than in 2005.*

Jock designed an elegant solution. He found a way to use the visualization capabilities of Tableau 2.0 to create an innovative solution that follows the rules of visual perception, resulting in a display that is easy to read and comprehend, despite the many data elements it combines into a single presentation.

Notice how easy it is to see both summary (expressed as medians for each neighborhood) and detail information in the same display. Notice also how easily you can compare neighborhoods. By enabling real estate agents to see all of these critical variables together, connections can be made between them that might not otherwise be discovered.

In this series of articles, I normally make a few suggestions for how the solution can be improved, but I'm at a total loss to find anything significant lacking in Jock's solution. I'm determined, however, to add some value, so I'll get really picky and point out that better colors than red and green could have been used to differentiate positive and negative variances between asking prices and actual sales amounts, because about 10% of males and 1% of females – those who are colorblind – would have difficulty discriminating between these colors. This is not a big deal, however, because the differing shapes of the positive variances (shaped like a "T") and the negative variances (shaped like an upside-down "T") are fairly easy to discriminate, even without the color differences.

Solutions that Fell Short

Let's try to learn something from some of the solutions that failed to work in one way or another. The first appears to have been created with software from Tableau, just like Jock's winning solution, but it doesn't exhibit the same clear representations of the data. In fact, I find it very difficult to interpret. This underscores the fact that having good software doesn't guarantee a good solution. You must still apply your skills to visually encode information in ways that clearly and accurately present it to the eyes of your readers.

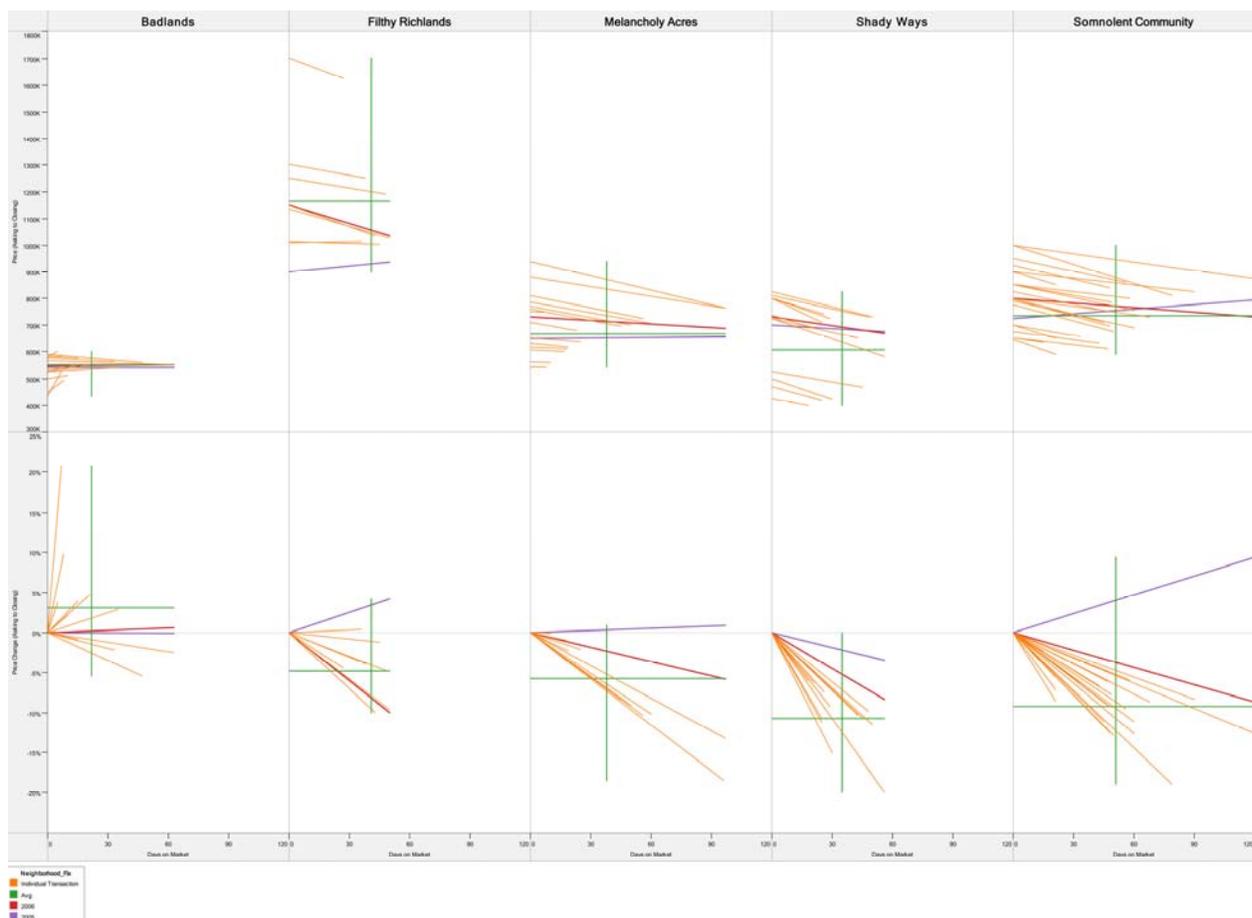


Figure 2: This solution is difficult to interpret.

Let's isolate a single graph (see Figure 3) in this series of 10 graphs and attempt to understand it. Here's how it works:

- Each orange line represents a single house that sold. Each begins at 0% for the asking price and extends to a positive or negative percentage to encode the difference between the asking price and the sales amount. If the line goes up from left to right, the sales amount was greater than the asking price, and a downward sloping line indicates a decrease. The distance that each line extends from left to right represents the number of days the house was on the market.
- The two green lines represent averages for the neighborhood: (1) the average percentage difference between the asking price and sales amount (the vertical line), and (2) the average number of days the that houses were on the market.
- I believe the purple line represents both the average difference in 2005 between the asking price and the sales amount (corresponding to the vertical scale) as well as the average number of days that houses were on the market (corresponding to the horizontal scale), and the red line does the same for 2006.

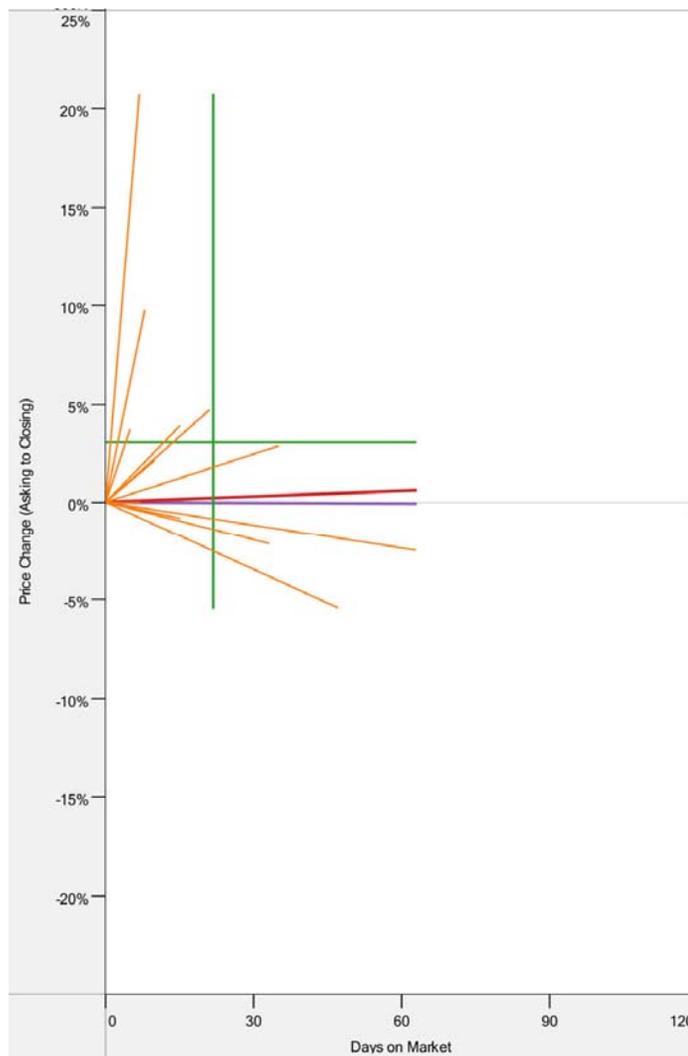


Figure 3: A single graph from the complete solution that appears in Figure 2.

Here are some of the reasons this display doesn't communicate effectively:

- Having this many lines in a graph produces a cluttered effect, making it difficult to read.
- Using lines to represent this many different variables, even though they are distinguished by color, makes it difficult to examine them independently.
- Although the percentage difference between the asking price and sales amount may be encoded as a line, to use the same line to also encode the number of days on the market overcomplicates the display. The slope of each line is determined by a combination of the two variables, which makes it difficult to compare them to one another in a meaningful way. If these lines were used to encode the difference between the asking price and the sales amount alone, with an X-axis that displayed asking price on the left and sales amount on the right, the slopes of the individual lines could be meaningfully compared to determine the differences.

This was a creative attempt to meet the requirements for this display, but it doesn't pass the test of clear and efficient communication.

The next example is not as difficult to understand as the last, but it also fails to communicate effectively, primarily because it breaks several rules of visual perception. Take a couple of minutes to review it on your own and list the problems that you find.

Housing Market Analysis May 2006

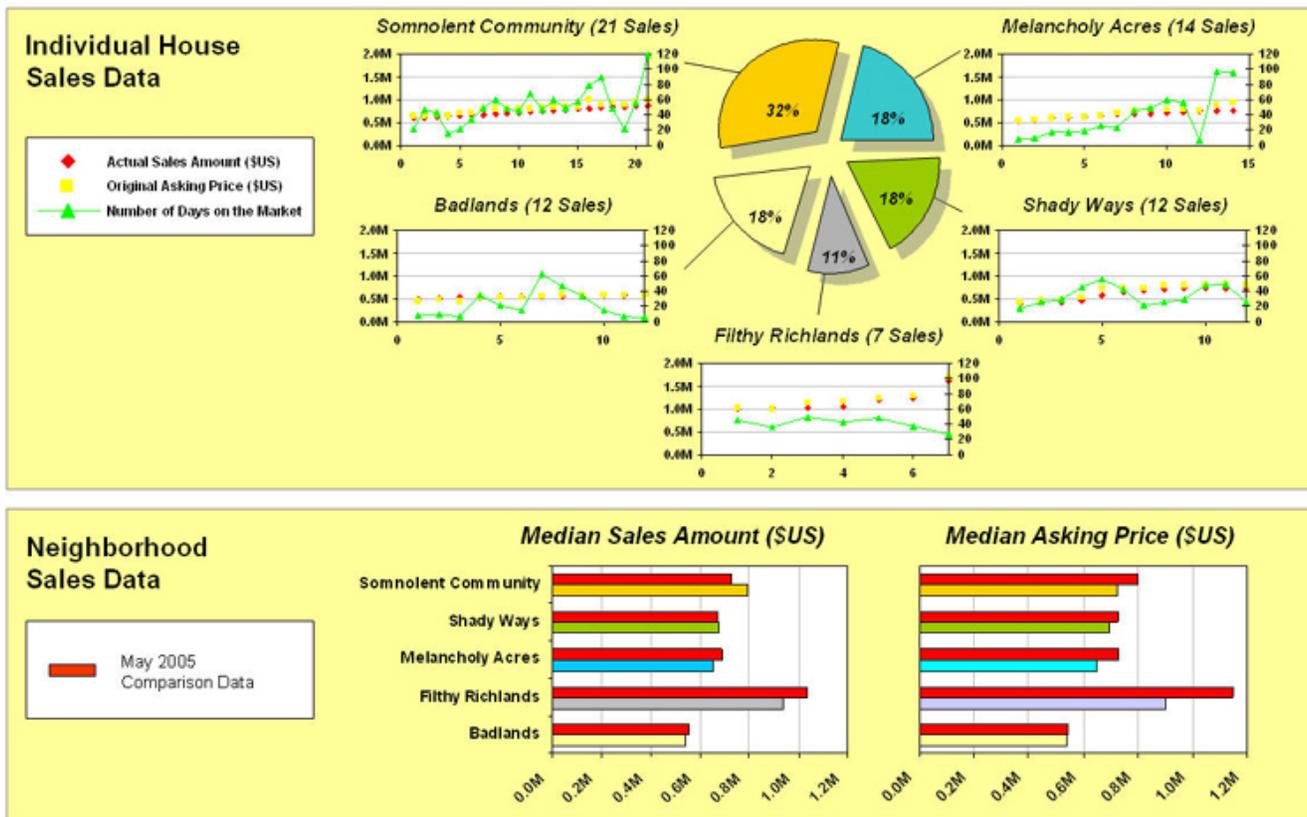


Figure 4: This solution breaks several rules of visual perception.

Here are a few of the problems that I found:

- Pie charts do not display portions of a whole as clearly as bar graphs. It is much easier to compare the lengths of bars than the two-dimensional areas of pie slices. Exploding the pie and adding drop shadows to the slices does not help matters.
- The graphs surrounding the pie chart were properly designed to support comparisons by giving them common quantitative scales (\$0–2,000,000 for house prices and 0–120 days for the amount of time on the market), but they failed in two other ways: (1) it is not appropriate to connect the number of days on the market for each house with a line (encoded as the green line) because these values are discrete (one for each house) and the up and down patterns formed by the lines are not meaningful, and (2) the yellow squares, which represent the asking price for each house, are difficult to see, because there is insufficient visual contrast between the yellow squares and the white background.
- Assigning a different color to each bar in the horizontal bar graphs is meaningless. The bars are labeled along the Y-axis, so there is no need for various colors to distinguish or identify them, yet people reading them will be led astray by these visual differences and waste time searching for meaning that doesn't exist.
- The two legends on the left only play a supporting role, so they should not be so visually dominant. Readers' eyes should be drawn primarily to the data in the graphs.
- Because the summary information (that is, the medians for each neighborhood) is displayed separately and differently from the individual home sales information, summaries and corresponding details are difficult to view in relation to one another.

The final example is typical of many displays today in that it focuses on flashy visual effects, rather than clear and efficient communication. Here are some of its specific problems:

- The background images are distracting and certainly unnecessary. Why would real estate agents who are trying to understand what's going on in the housing market need to see these vague photos of houses in the background?
- By including values as text on each of the bars and along the lines, the patterns represented by the bars and lines have been partially obscured in clutter. When precise values are necessary, which is not the case here, they could have been provided in tables rather than in the graphs themselves.
- The visual effects, such as drop shadows and gradients of color in the bars, are flashy, but add no real value. Any visual content that doesn't add value simply distracts from the message.

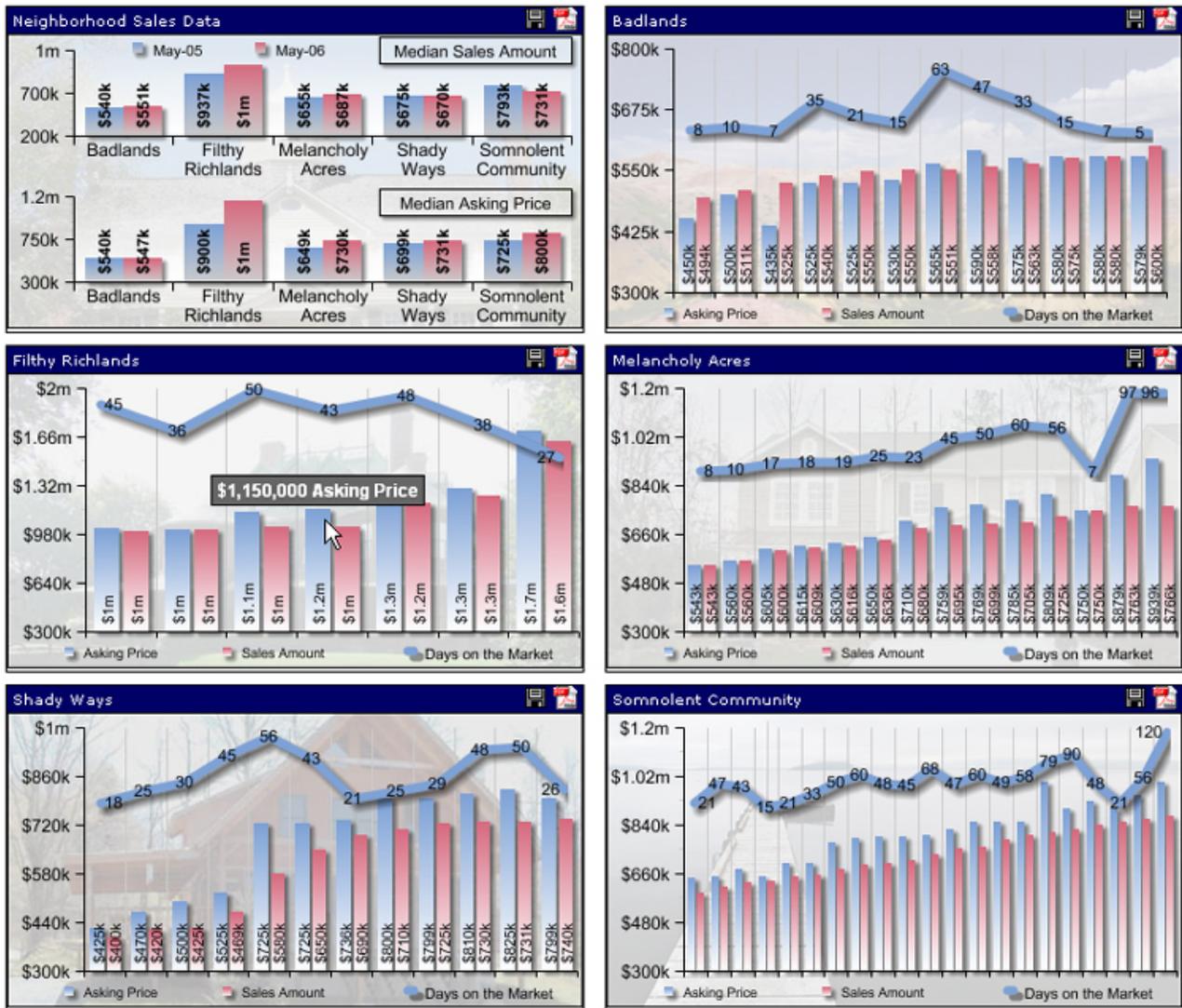


Figure 5: This solution is filled with visual effects, decoration, and textual clutter, which distract from the data.

I hope you've found these solutions and my critiques informative. Next month, I'll feature the winning solution to scenario #4 of our 2006 Data Visualization Competition, which succeeds in designing a dashboard that be used quite effectively by airline executives. If you're interested in dashboard design, I believe you'll find this next article enlightening.

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](http://www.perceptualedge.com) of articles at www.perceptualedge.com. Between articles, you can read Stephen's thoughts on the industry in his [blog](#).