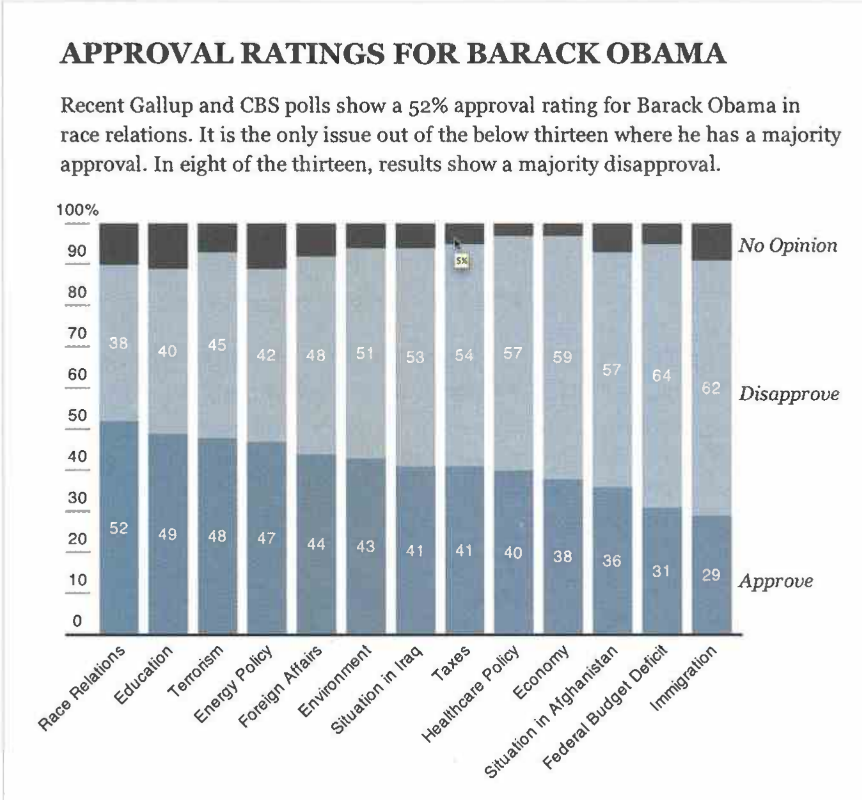


FIGURE 5-13 Interactive stacked bar chart in Protovis

To start, set up the HTML page and load the necessary Protovis JavaScript file.

```
<html>
<head>
    <title>Stacked Bar Chart</title>
    <script type="text/javascript" src="protovis-r3.2.js"></script>
</head>
<body>
    <div id="figure-wrapper">
        <div id="figure">

    </div><!-- @end figure -->
    </div><!-- @end figure-wrapper -->
</body>
</html>
```

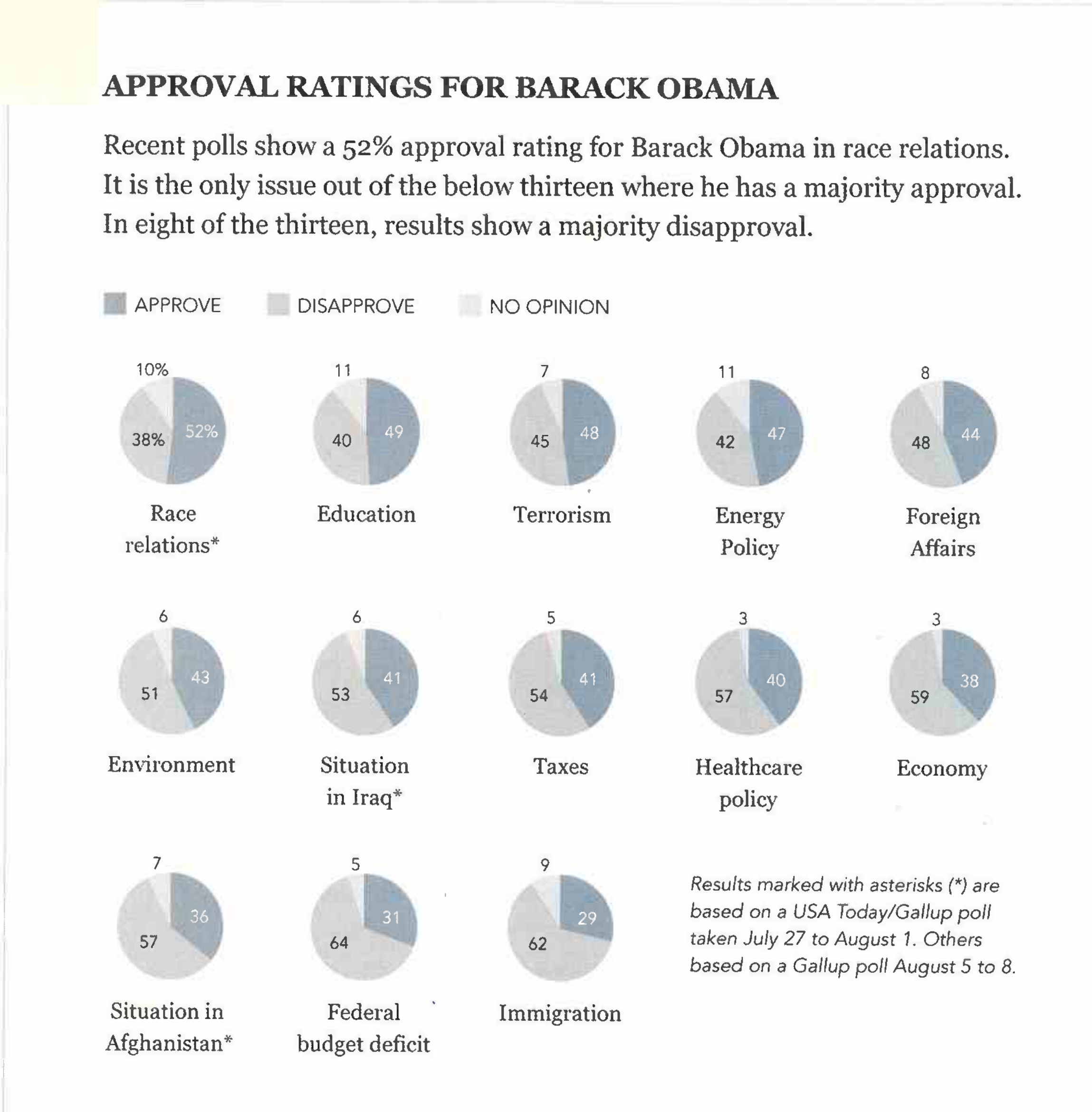
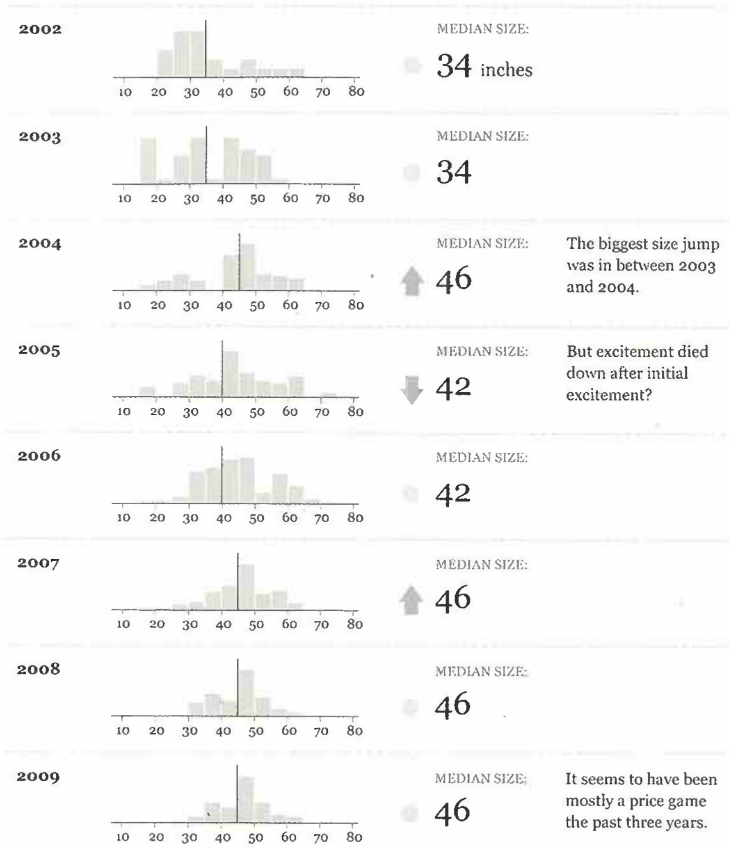


FIGURE 5-12 Series of pie charts

Television Size Over the Years

A quick search for 'average TV size' quotes Sharp as saying average TV size could be up to 60 inches by 2015. (Although I can't find the actual source.) So by how much has that magical entertainment screen grown during the past 8 years? Probably not as much as you think.



About the Data

The numbers come from the past 745 dated CNet television reviews. While reviews aren't a direct indicator of what televisions people buy, they do show what's on the market.

Data from CNet / *Created by* FlowingData

FIGURE 6-36 Distribution of television size over the years

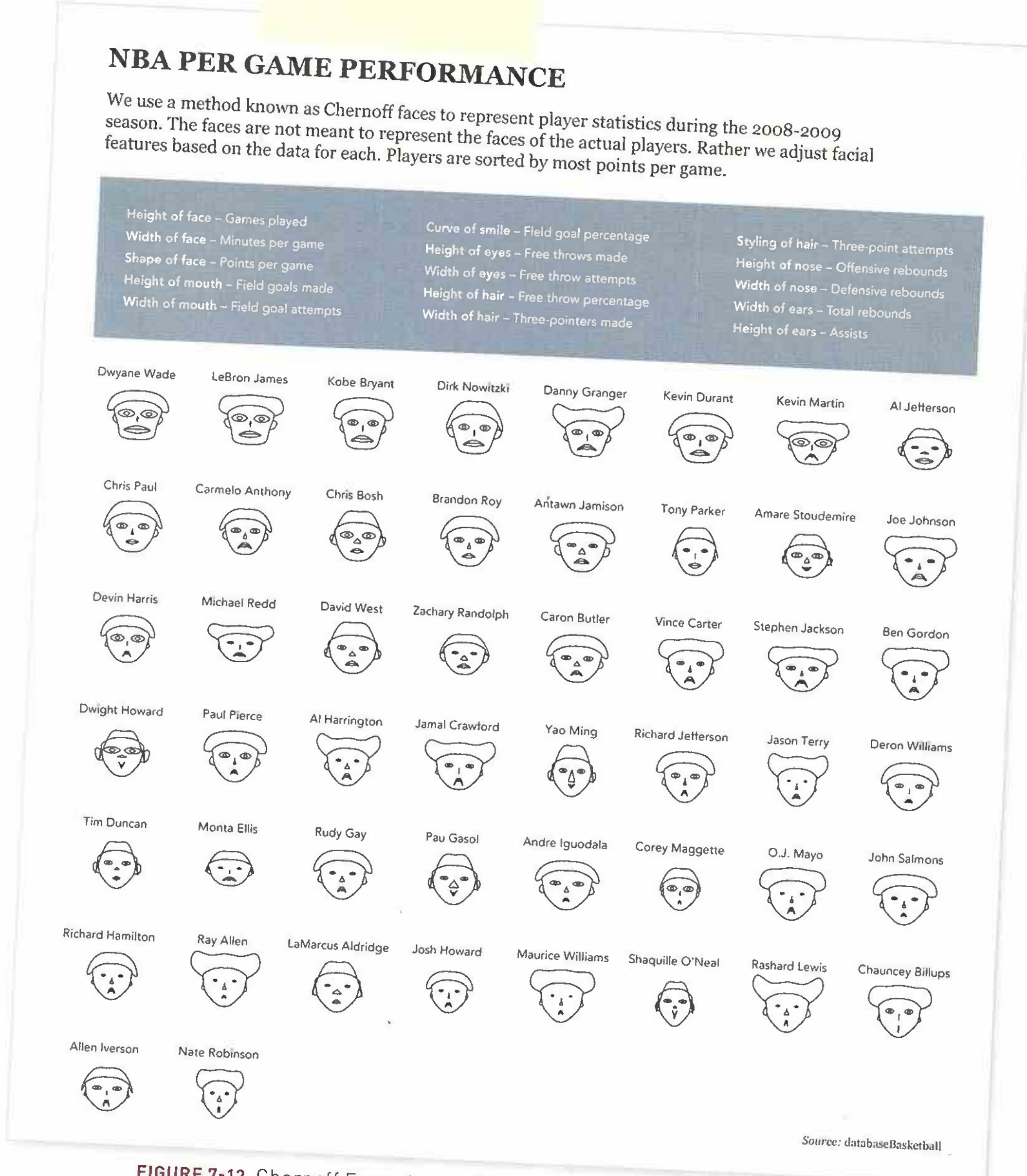


FIGURE 7-12 Chernoff Faces for top NBA scorers during the 2008–2009 season

Along the same lines, I designed a graphic for crime in the United States (Figure 7-13) using Chernoff Faces, and someone actually commented that it was racist because of how the face looked for states with high crime

I wanted to know how sequels compared to their originals in freshness. It turns out not very well, as shown in Figure 6-38. The median rating of finales was 37 percentage points lower than the median of the originals. In other words, most originals were fresh, and most finales were rotten.

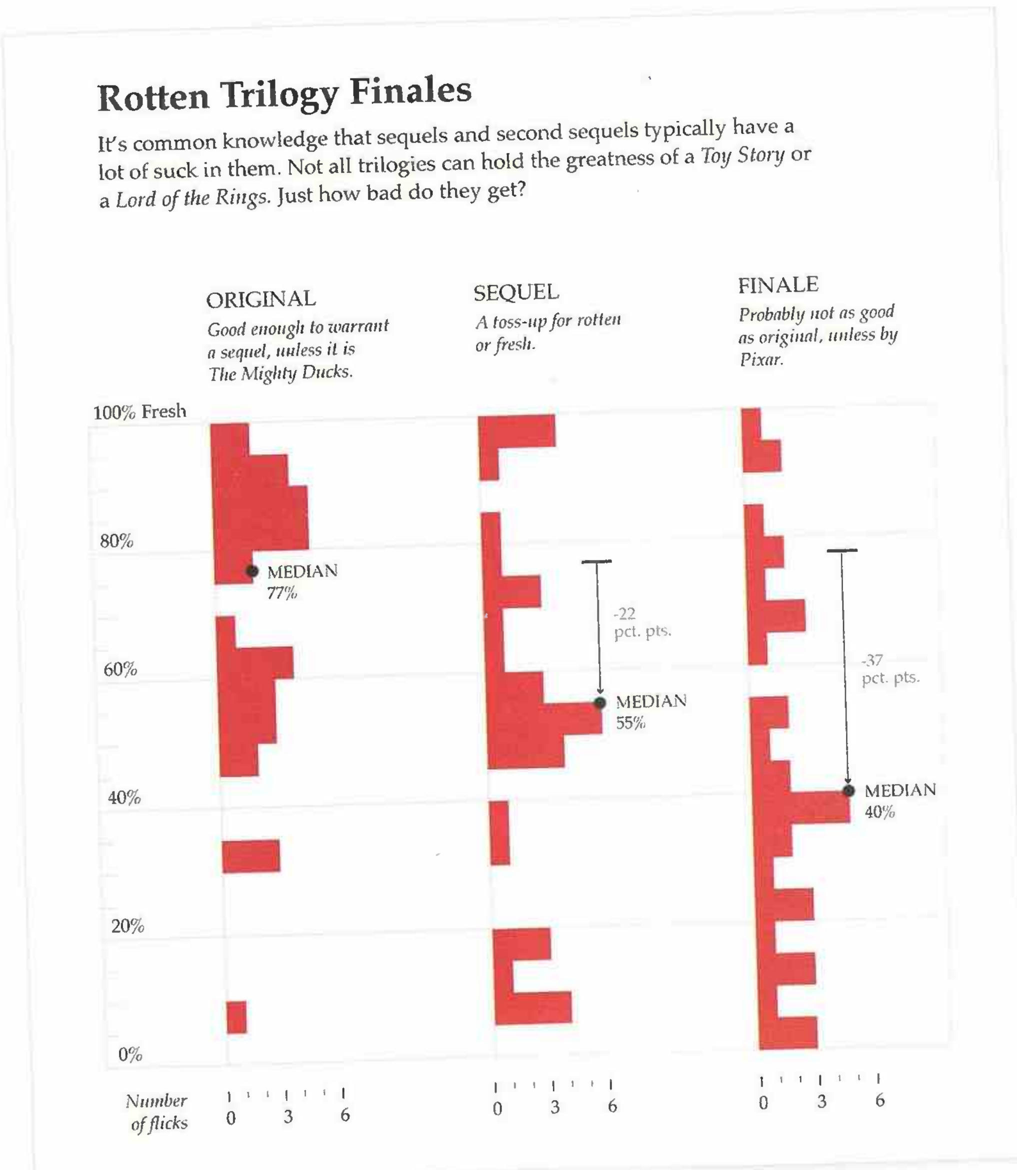


FIGURE 6-38 How trilogies rate from original to finale

In any case, FlowingData readers understood the graphic for the most part—they're a data savvy bunch, naturally. However, the graphic was later linked from IMDB, also known as the Internet Movie Database. IMDB has a much more general audience, and judging by the comments after that link-back, the less data savvy readers had trouble interpreting the distributions.

However, the second part of the graphic, as shown in Figure 6-40, seemed much easier to understand. It's a use of small multiples where each bar represents the rating for a movie. Bars were colored red for rotten and green for fresh.

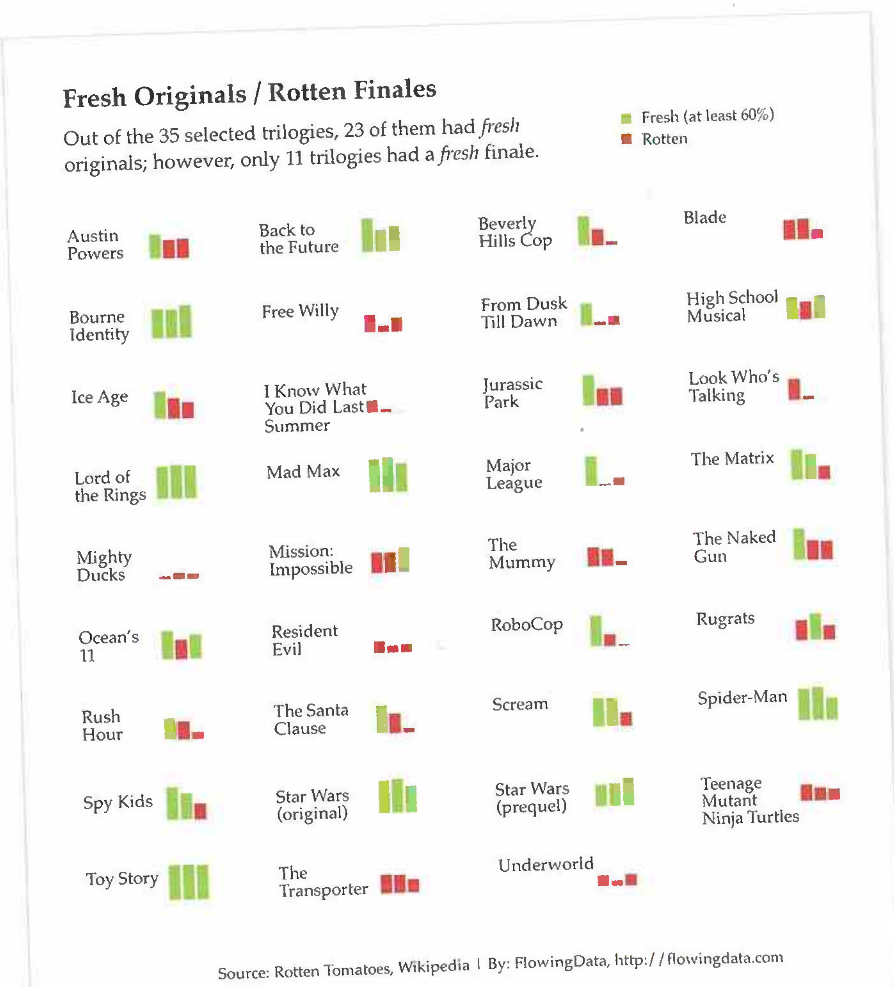


FIGURE 6-40 Small multiples for ratings of trilogies

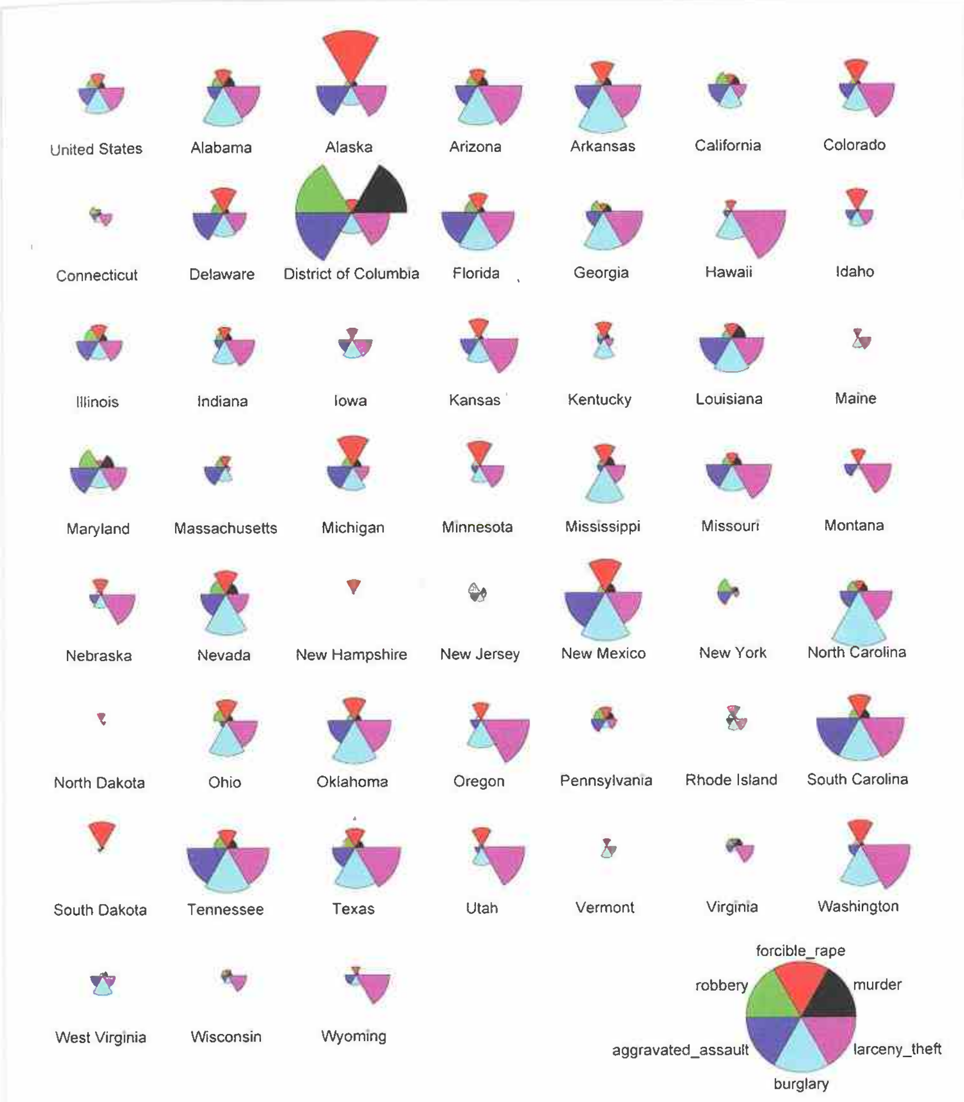


FIGURE 7-18 Crime displayed as Nightingale charts

Like I said though, I'm good with the original format in Figure 7-16, so you can take that into Illustrator to do some cleanup. It doesn't need a whole lot of modification. More white space between the rows could make the labels less ambiguous, and you can place the key on top so that readers know what they're getting into (Figure 7-19). Other than that, it's good to go.

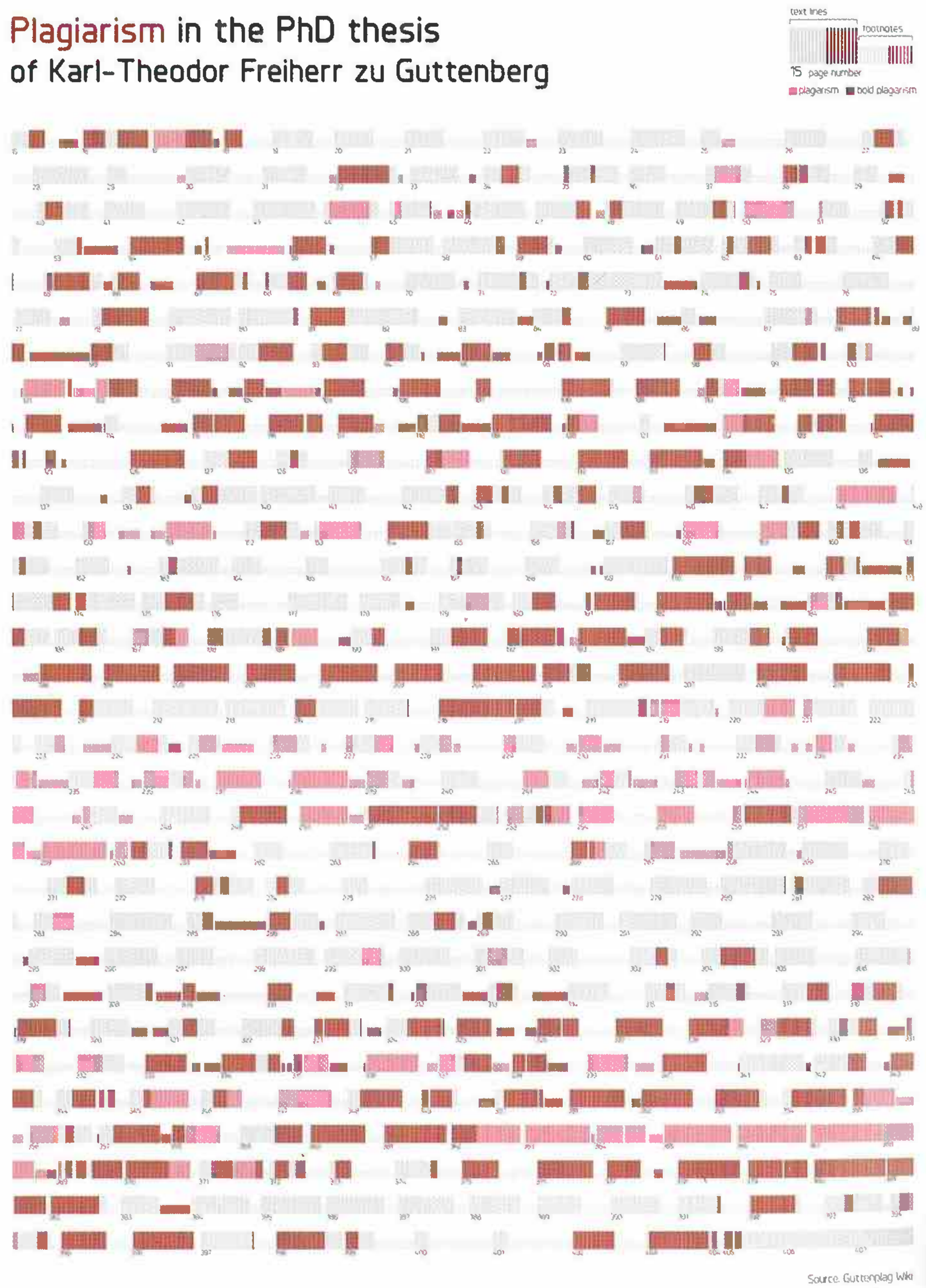


Figure 10.69 Visualizing Plagiarism. http://driven-by-data.net/about/plagiarism/. See the interview for an explanation. By Gregor Aisch (driven-by-data.net)

For this reason, to successfully see meaning in visual displays, we must encode data in ways that allow what's interesting and potentially meaningful to stand out from what's not. What stands out to you as you look at the image below?

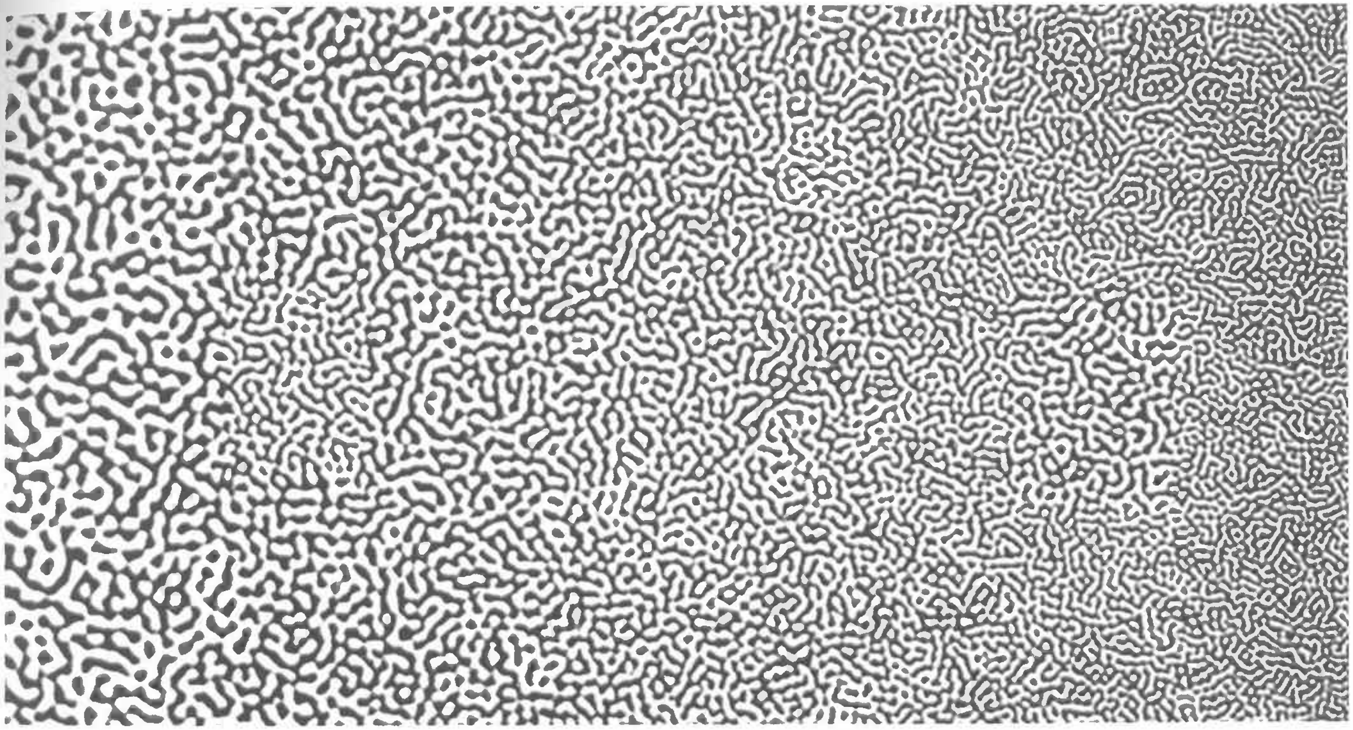


Figure 3.5. This image appears in *Information Visualization: Perception for Design*, Second Edition, Colin Ware, Morgan Kaufmann Publishers, San Francisco CA, 2004, p. 171.

Among the things you notice are probably two roughly oval-shaped areas of texture embedded within the pattern that stand out from the rest: one is in the left half and one in the right half of the image. They stand out because they differ from what surrounds them. What's not obvious is that these two regions that catch our attention are exactly the same. The area that stands out on the left is made up of lines that are less thick than those that surround them. By contrast, the area on the right is surrounded by lines that are thicker than the surrounding lines. Because the two areas that stand out are embedded in contexts that differ, our perception of them is affected so that it is difficult to see that they are made up of lines of the same thickness.

We learn from this fact that information visualizations should cause what's potentially meaningful to stand out in contrast to what's not worth our attention.

Fact #2: Our eyes are drawn to familiar patterns. We see what we know and expect.

When we look at the image below, our eyes see the familiar shape of the rose and our minds quickly categorize it as fitting a recognizable pattern that we know: a rose. However, another distinct image has been worked into the familiar image of the rose, which isn't noticeable unless we know to look for it. Take a few seconds right now to see if you can spot the image that's embedded in the rose.

Figure 3.6. This image was found at www.coolbubble.com.

• • • • • • •

Did you spot the dolphin? Once we have been primed with the image of the dolphin (turn to page 36 to see it), we can easily spot it in the rose. This second fact teaches us that visualizations work best when they display information as patterns that are both familiar and easy to spot.

In addition to visual perception, information visualization must also be rooted in an understanding of how people think. Only then can visualizations support the cognitive operations that make sense of information.

Fact #3: Memory plays an important role in human cognition, but working memory is extremely limited.

The two photographs on the next page illustrate one of the limitations of working memory. We only remember the elements to which we attend. Imagine

Making Abstract Data Visible

We'll begin this section with an illustration. Something is wrong with the following graph. Take a moment to see if you can find a problem.

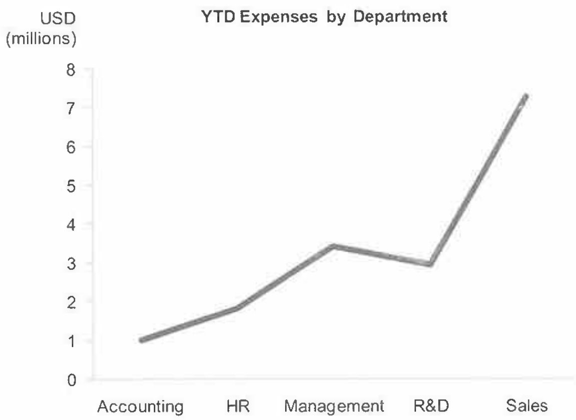
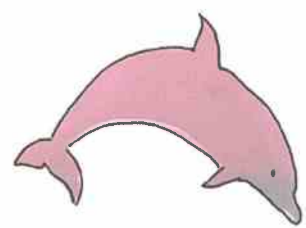


Figure 3.8

Does anything about this graph bother you? Does any aspect of its design undermine its ability to represent data appropriately? This is a case where it doesn't make sense to encode the values as a line. The line connects values for a series of categorical items—departments in this case—that are completely independent from one another. These are discrete items in the category called "departments," which have no particular order and no close connection to one another. To connect these values with a line visually suggests a relationship that doesn't exist in the data. We are used to interpreting a line like this as indicating an increase or decrease in some variable, in this case, expenses on the vertical axis, in relation to some variable on the horizontal axis that might reasonably be expected to affect or have a comprehensible relationship to expenses, such as time. The up and down slopes of the line and the pattern formed by them are meaningless.



This picture of a dolphin can be found embedded in the rose in Figure 3.6.

Lines work well for connecting values through time, such as months in a year, but are inappropriate for connecting categorical items such as departments. In the following graph, separate bars accurately encode and visually reinforce the independent nature of these departments and their expenses.

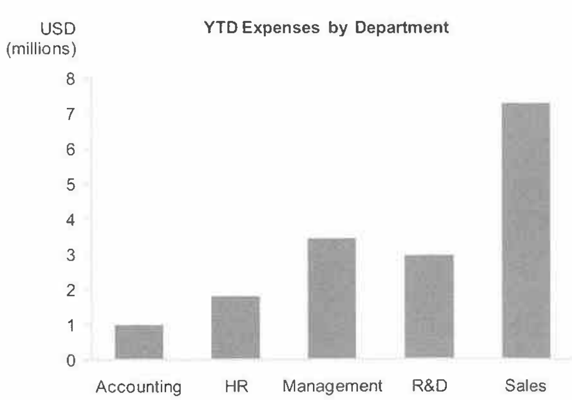


Figure 3.9

divided into equal intervals. For example, if we want to count and compare sales orders of various dollar sizes, and the smallest order is 50¢ and the largest is $100, we could break order sizes into intervals of $10 each: $0.00-9.99, $10.00-19.99, and so on. This would be an interval scale. A range of time beginning at one point and ending at some later point is a quantitative range of values. Breaking time into intervals, such as years, also results in an interval scale. Because interval scales consist of ordered and intimately connected items, such as the years 2004, 2005, 2006, and 2007, it is appropriate to display values across those years using a line to connect them. It is natural and effective. Our eyes can easily trace how a set of values change through time when these values are displayed as a line, and our minds are able to easily understand the nature of this change.

Although it is true that years vary in length because of leap years and months have different numbers of days, for most analytical purposes we consider these intervals equal in size.

The point that I'm trying to make is that there are ways to visually display data that are effective because they correspond naturally to the workings of vision and cognition, and there are ways that break the rules and consequently don't work. If we wish to display information in a way that will enable us and others to make sense of it, we must understand and follow the rules.

I'll show another example to drive this truth home. If we wish to rank and compare the sales performance of the 10 products that are displayed in each of the two graphs below, which supports this task most effectively?

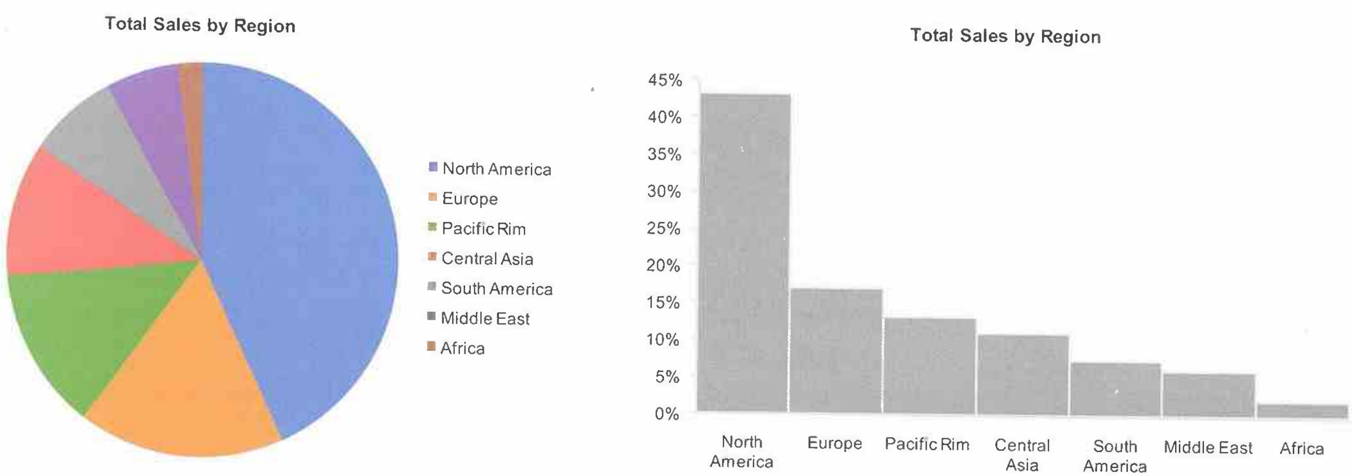


Figure 3.10

The pie chart doesn't work nearly as well as the bar graph because, to decode it, we must compare the 2-D areas or the angles formed by the slices, but visual

aps don't have to be associated with a geographical map. Another com-
eatmap is composed of cells (square or rectangular areas) arranged as a
r matrix with each cell color-coded to display a quantitative value, as in
lowing example, which shows variations in gas mileage (miles per gallon,
horsepower, and weight for several cars in the year 1982.

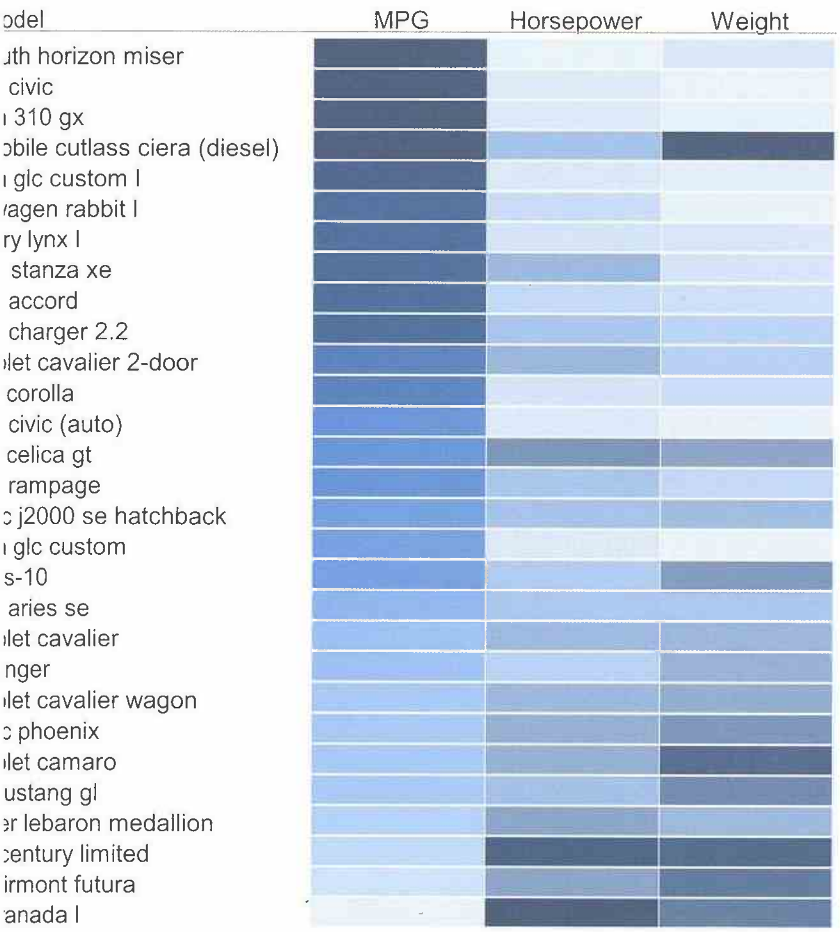


Figure 3.18

special type of information visualization that is similar to a heatmap,
a *treemap*, uses both size and color to encode quantitative values.
aps were originally developed by Ben Shneiderman of the University of
nd as a way to simultaneously display two quantitative variables for a
umber of items, arranged hierarchically.

Another common comparison when analyzing data is making the distinction between values that appear normal—that is, within the quantitative range where most of the values are located—and those that appear abnormal. Values that fall outside the norm are called outliers or exceptions. When values are displayed in a well-designed visualization, it is usually easy to spot outliers and always worthwhile to examine them.

Sorting

Don't underestimate the power of a simple sort. It's amazing how much more meaning surfaces when values are sorted from low to high or high to low. Take a look at the following graph, which displays employee compensation per state:

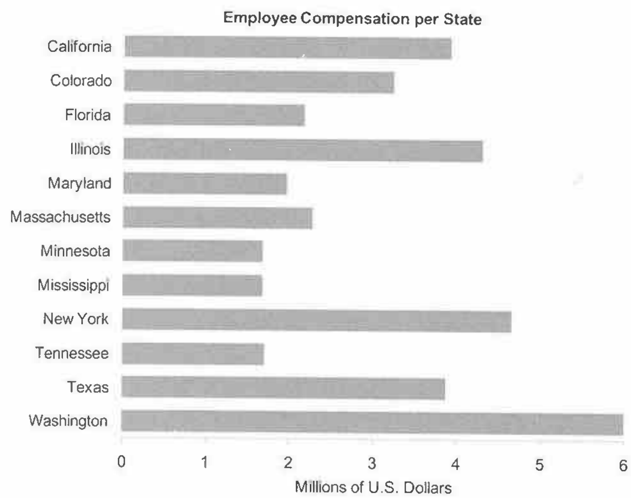


Figure 4.8

With the states in alphabetical order, the only thing we can do with ease is look up employee compensation for a particular state. It is difficult to see any meaningful relationships among the values. Now take a look at the same data, this time sorted from the state with the highest employee compensation to the one with the lowest.

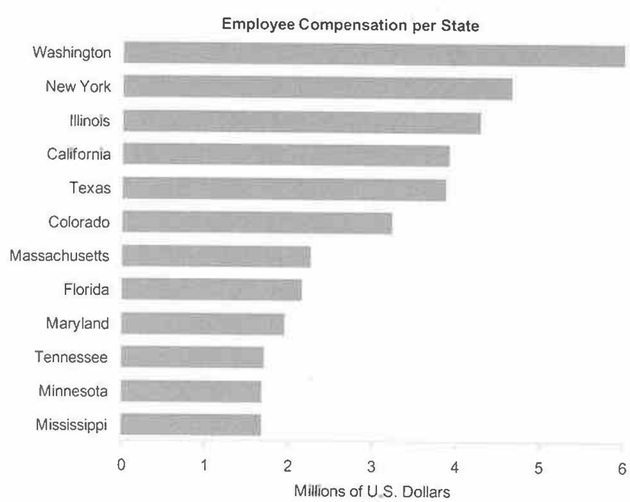


Figure 4.9