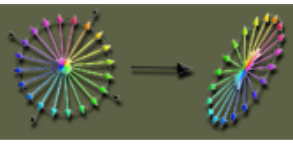
ashleykolodziej@brandeis.edu

TOPIC NOTES  INFORMATION VISUALIZATION

Heat Maps and Scatter Plots

What are Heat Maps and Scatter Plots?

Scatter Plots

From the appearance of these charts they might better be called Scatter Dots. Scatter Plots are also known by other names such as scatter diagram, scatter chart, XY Plot, or dot chart. We have seen these visualizations whether it is from a stats course or in consuming media but maybe didn't know their formal name. So what are they and what do they show?

Scott Berinato in his book, *Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations,* has defined the Scatter Plot chart as "Dots plotted against two variables show the relationship between those two variables for a particular set of data."

Nathan Yau in his book, *Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics*, shows the basic framework of a scatterplot in figure 6-1 below. You should note, in particular, that he explains that the two variables on the x- and y-axes are typically dependent. Then in figure 6-2 Yau shows different types of correlations—positive, negative, and no correlation—illustrated in scatterplots. As you can see when you look at a scatterplot from left to right if the dots move upward then you are looking at a positive correlation between the two variables. Conversely, if the dots get lower as you look from left to right then you are looking at negative correlation. No correlation refers to plots where the dots do not show a distinctive up or down scatter.

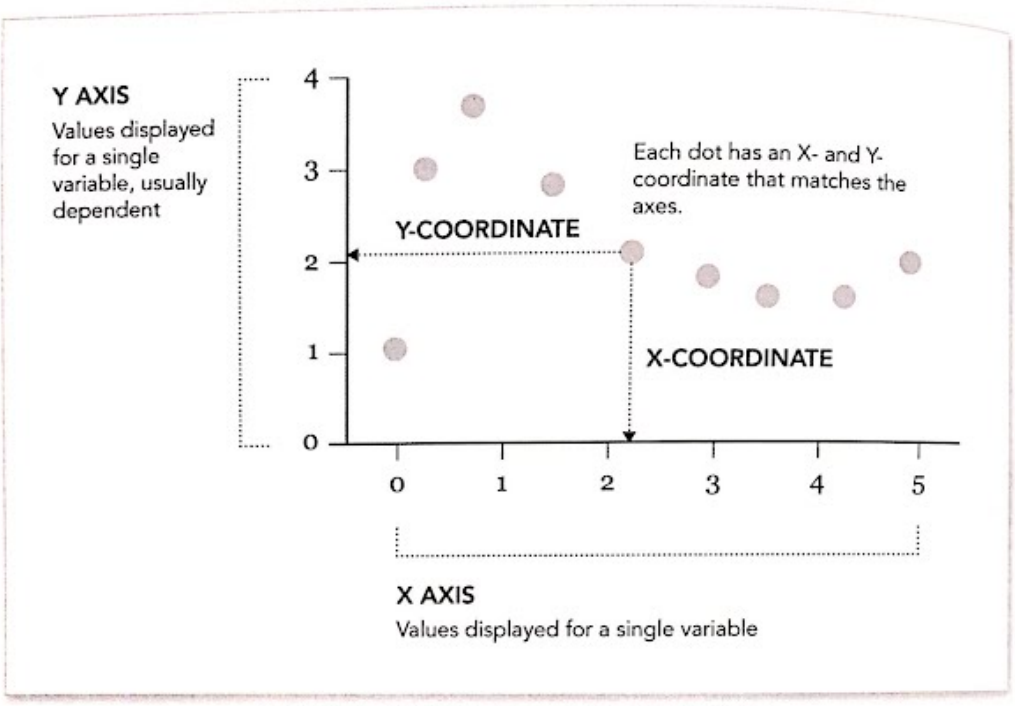


FIGURE 6-1 Scatterplot framework, comparing two variables

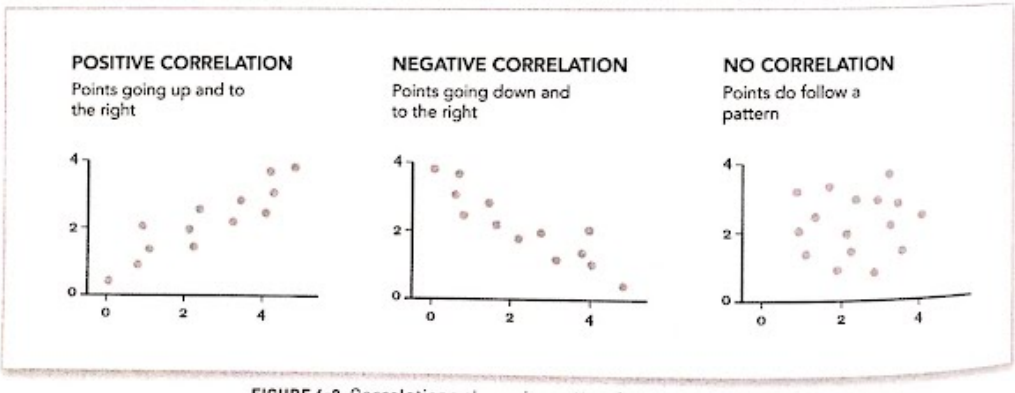
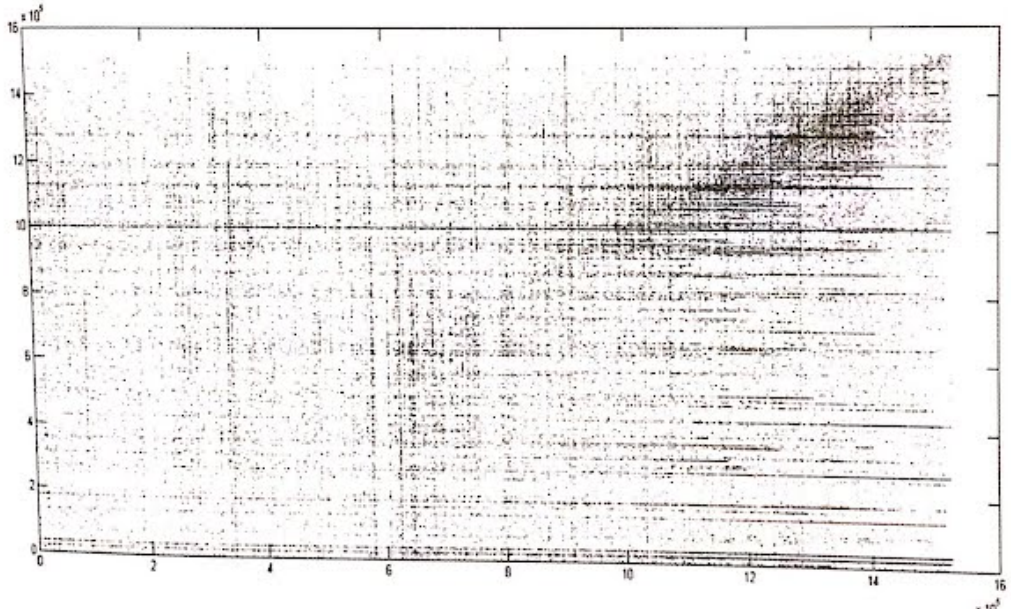


FIGURE 6-2 Correlations shown in scatterplots

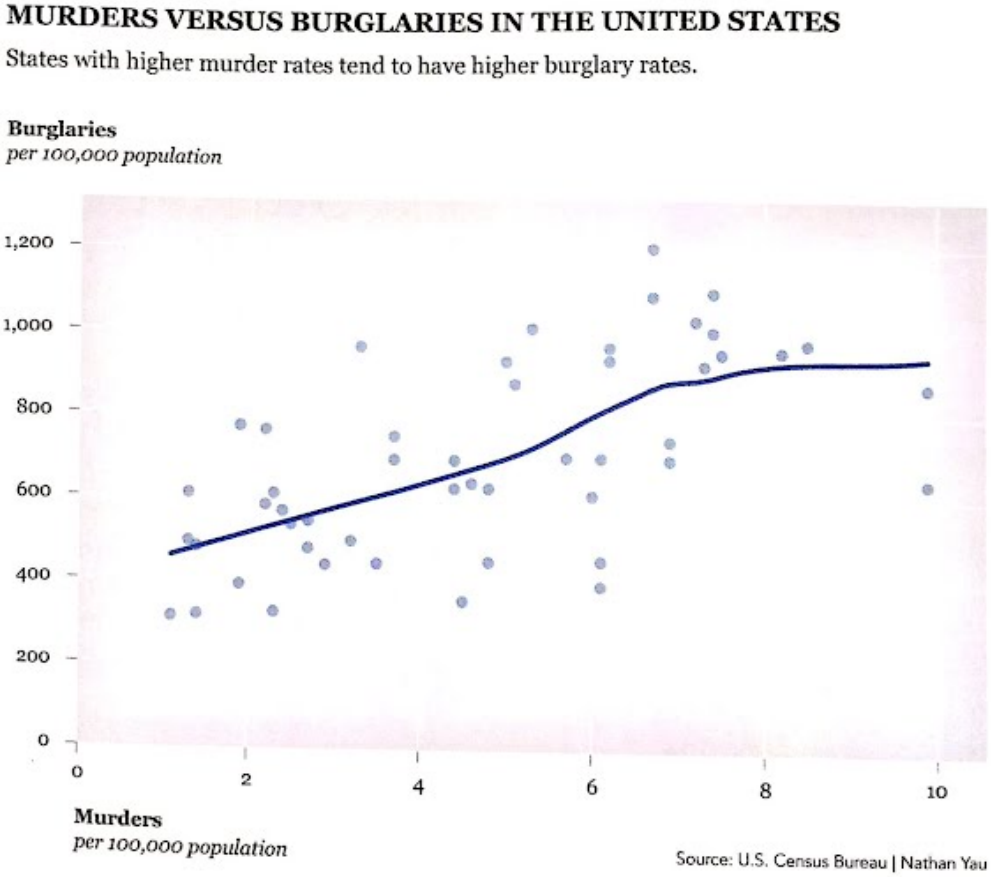
Because of the scatter plots spatial approach, it makes a great candidate to detect and show correlations between two variables. Our eye easily notices clusters of points and outliers. As an example, you might want to see if there is a correlation with your users' ages and their error rates in using your product or website.

Keep in mind though as you interpret or convey this type of chart to your audience that correlation does not imply causation, as you may recall from statistics.

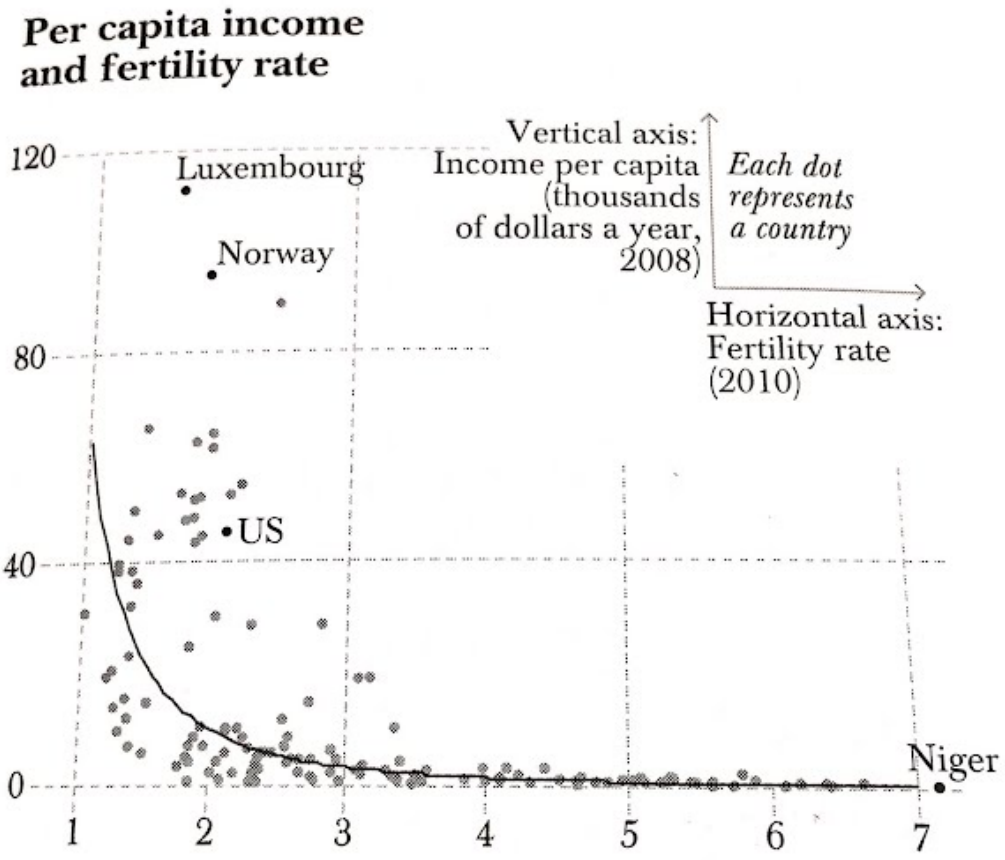
Scatterplots can also be useful in looking at extremely large data sets. For example, look at the scatterplot below from "Good Charts." This scatterplot created by Alex "Sandy" Pentland of MIT shows 10 million data points that chart the social connections between stock traders on a social trading platform. Even though this is a ton of data, our eyes and brains are able to discern a couple of things to focus on. We notice the dense black spot and an upward right increase in density and some striated patterns especially to the right. The creator also had a version where you could zoom into subsets to see all the data points if desired.



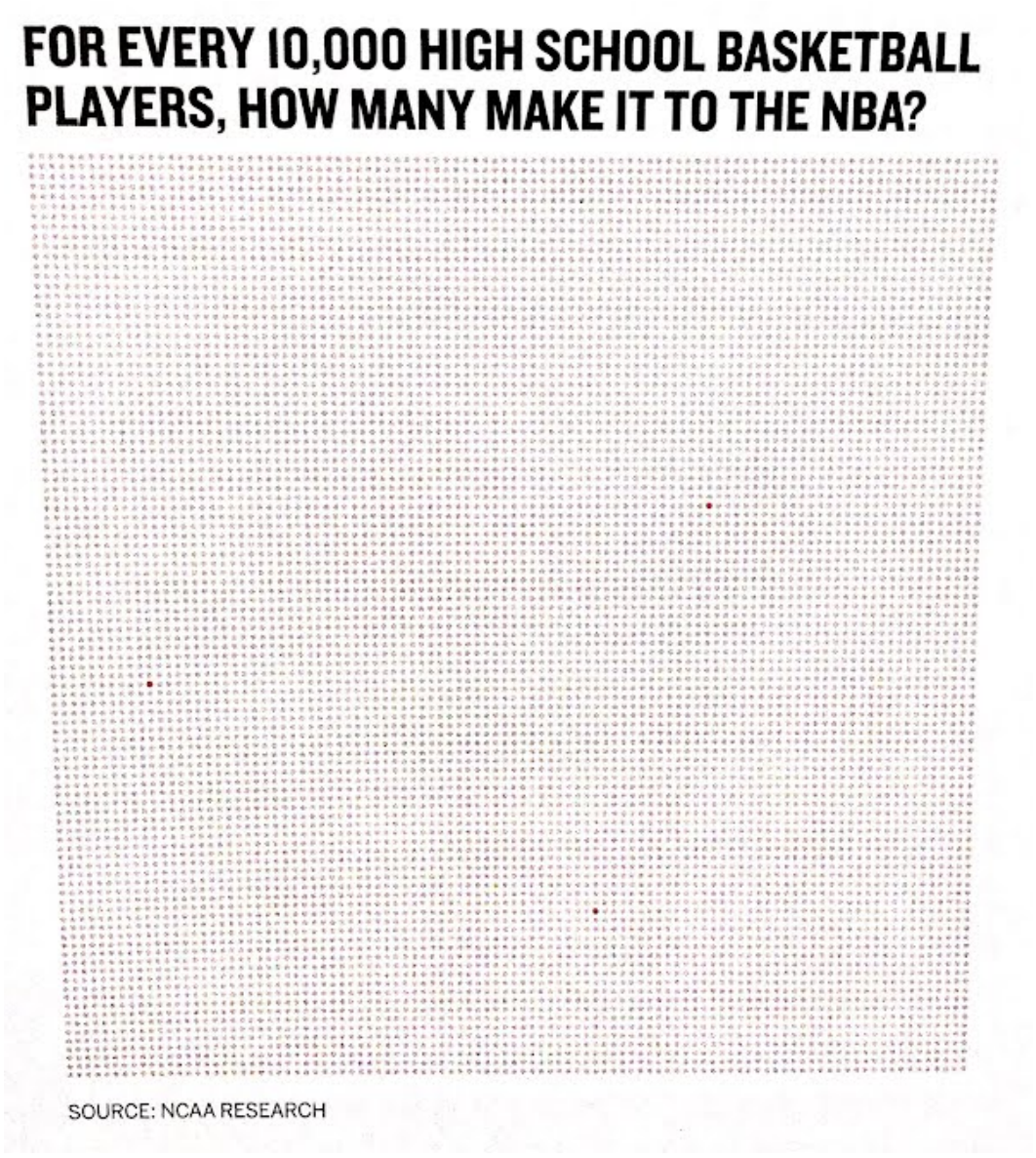
Often times you will see scatterplots displayed with a line of best fit, also called a trendline, to aid the viewer in being able to study the relationship between the variables. The closer the dots are to the line, the more the variables correlate. You can see a trendline drawn in the scatterplot below showing murders versus burglaries in the United States. Also note some design choices in the first visualization that allow the data to speak. Distractions are minimized by not using a border or drawing the axes with lines. The chart also includes a title describing the relationship and the source in the lower right.



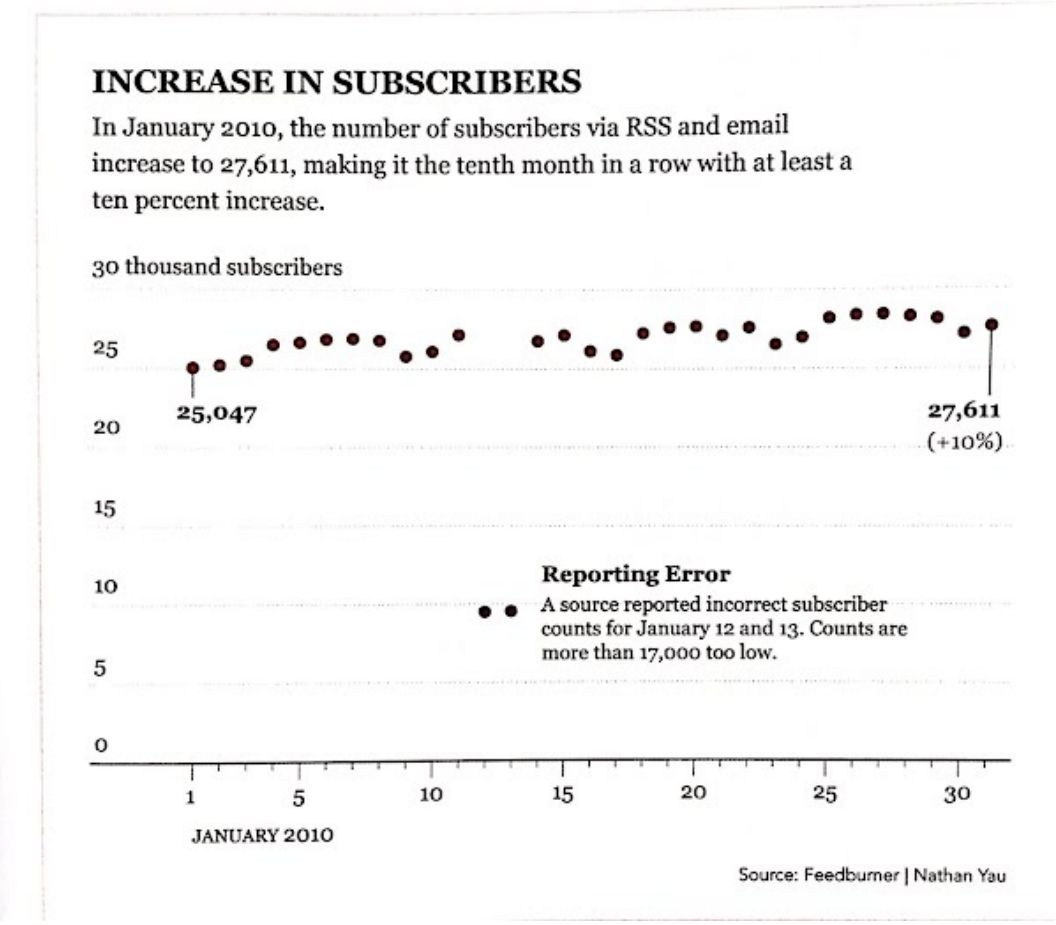
Below is another example of a scatter plot with a trendline when plotting per capita income (y axis) and fertility rates (x axis) in a number of countries.



Sometimes seeing data as dots rather than conveyed in percentages can be more powerful or persuasive. For example, rather than showing a pie chart that reveals that only .03% of high school basketball players make it to the NBA, the chart below (from Scott Berinato's book *Good Charts*) 10,000 high school players with each player represented as a dot, reveals that same percentage in each red dot (only 3) as representative of those that make it. Seeing this is more relatable than reading statistics. It really has that memorable wow effect.



Sometimes you may also see a scatterplot with temporal data, illustrating one variable over time. In the example below from Nathan Yau on Feedburner, the chart shows the number of subscribers through email and RSS over time throughout the month of January. Please take note of the design choices to show the increments without the axis lines and the use of light horizontal lines in the background that do not detract from the bold dots showing the data story. He also provides a title and brief description of the highlight that he is illustrating with the visualization. Notice he also directly labels the data at the beginning and end of the month that shows an increase of 10%.



Heat Maps

One way to visualize a table of data all at once is to use colors to represent numerical values or ranges. These are called heat maps and were invented by Cormak Kinney in the mid 90's to help investors. (For more info on Kinney check out this Forbes article at http://www.forbes.com/global/1999/0517/0210064a.html)

Heat maps makes it easy to find relatively high and low values based on the color. Most of the times when displaying heat maps, you would use darker colors to represent greater values and lighter colors to represent lower values. Where tables work well for finding specific values, heat maps enable you to scan tabular values for trends that are easily visible through the color patterns and changes. Nathan Yao presents the basic framework in the graphic below from his book *Visualize This.* Units or observations are calculated on the vertical or row while unit variables are presented in columns or along the horizontal axis. Each cell is colored according to its value.

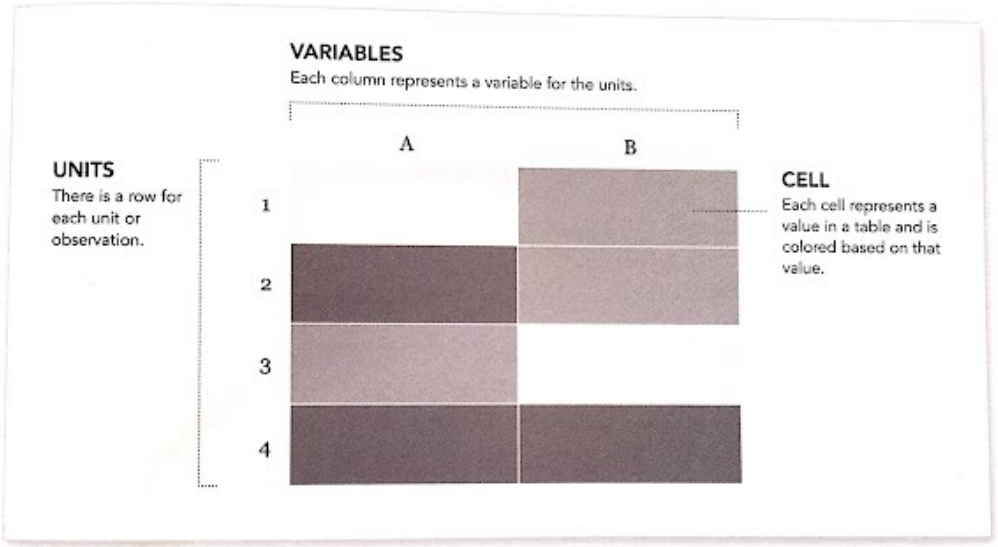


FIGURE 7-1 Heatmap framework

An example of a heat map or heat matrix in action from *Visualize This* shows a comparison for various stats for some NBA players. The colors range from yellow to shades of orange to dark red.

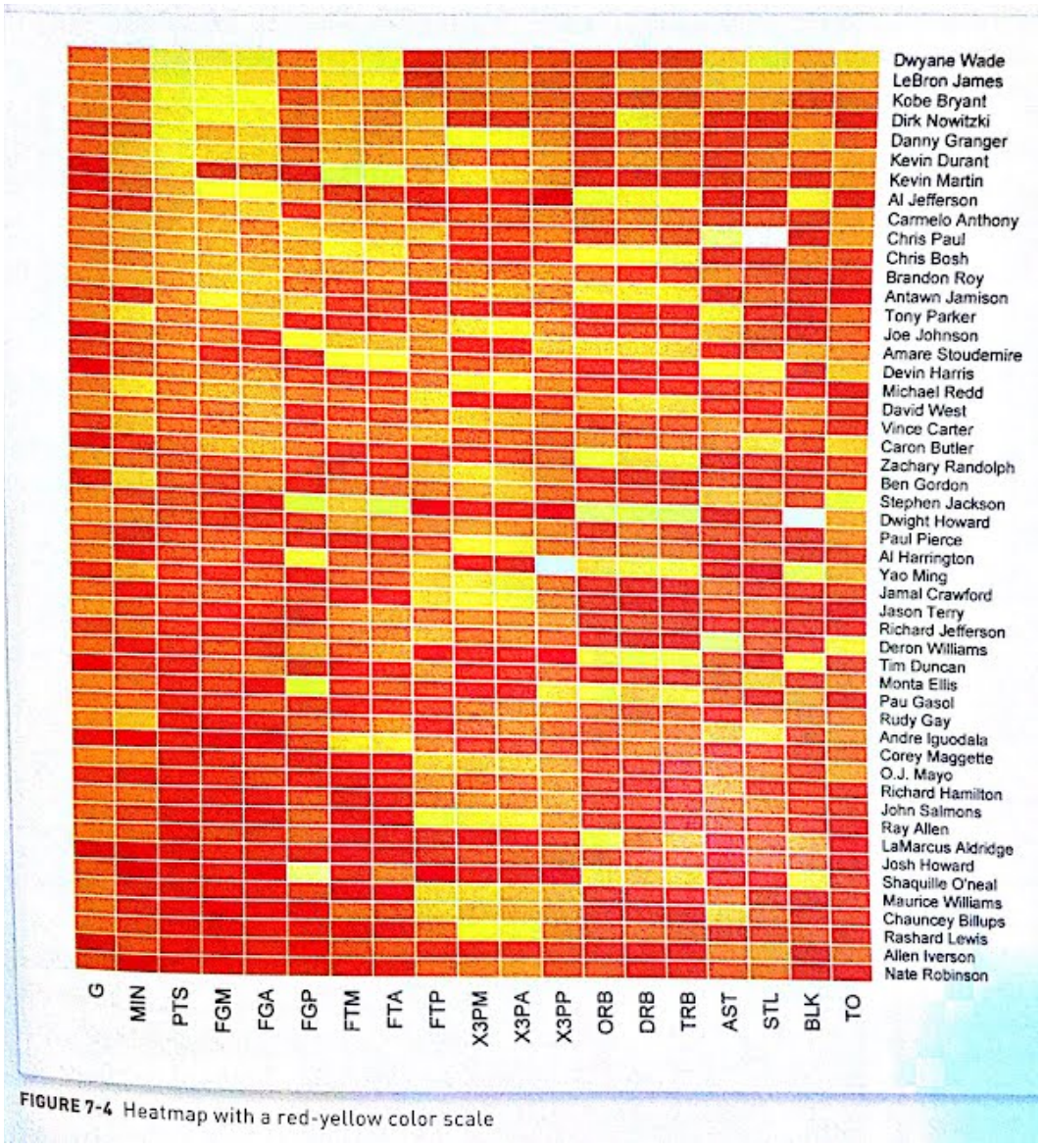


FIGURE 7-4 Heatmap with a red-yellow color scale

The next chart shows the same data with a title and better formatting. Notice the use of slanted labels to disambiguate the abbreviations and putting the players on the left rather than the right. They also use a blue color gradient that is easier on the eyes and border each cell with a thin white line. Consider that cell borders can add more definition to each cell which makes it easier to scan left to right and top to bottom. In addition, the contrast has been toned down using transparency.

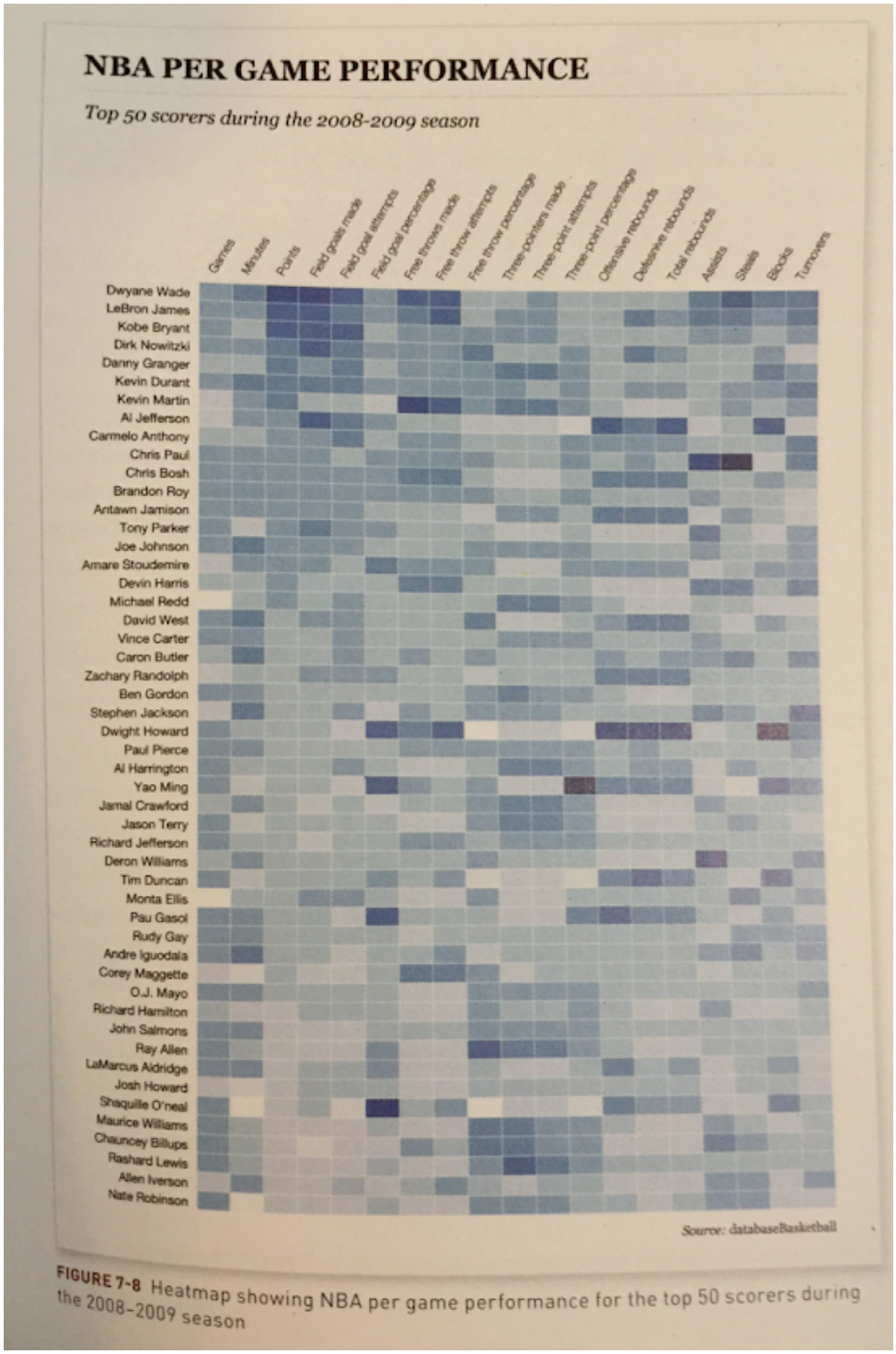
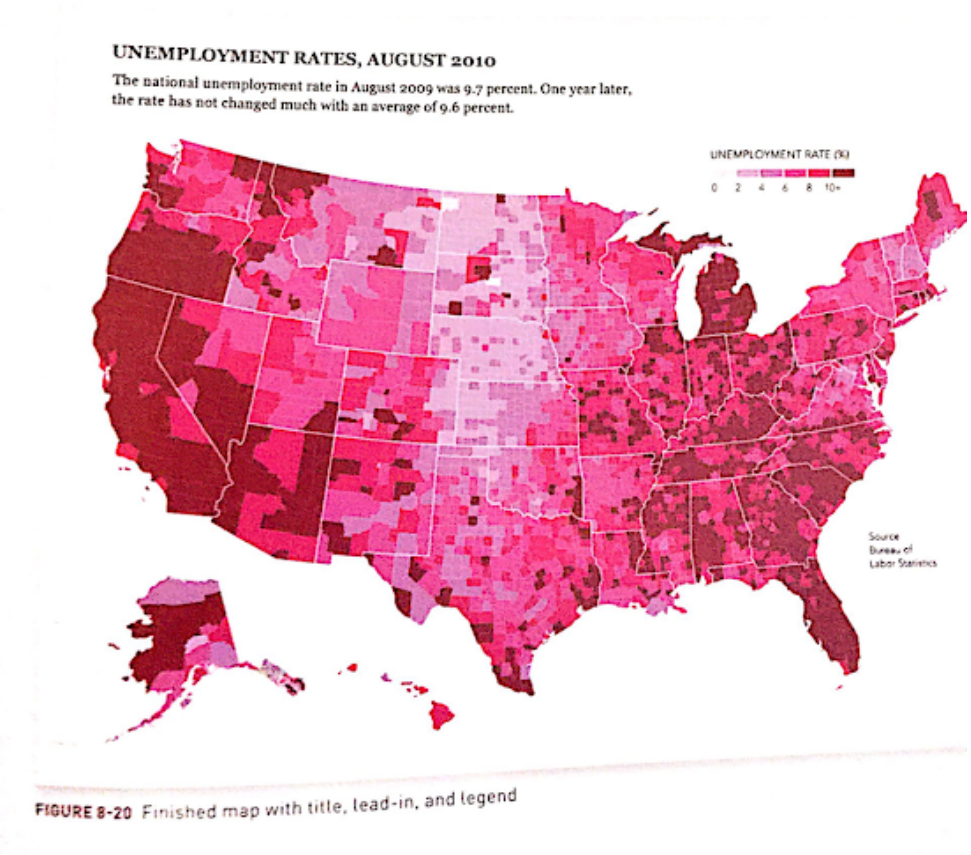


FIGURE 7-8 Heatmap showing NBA per game performance for the top 50 scorers during the 2008–2009 season

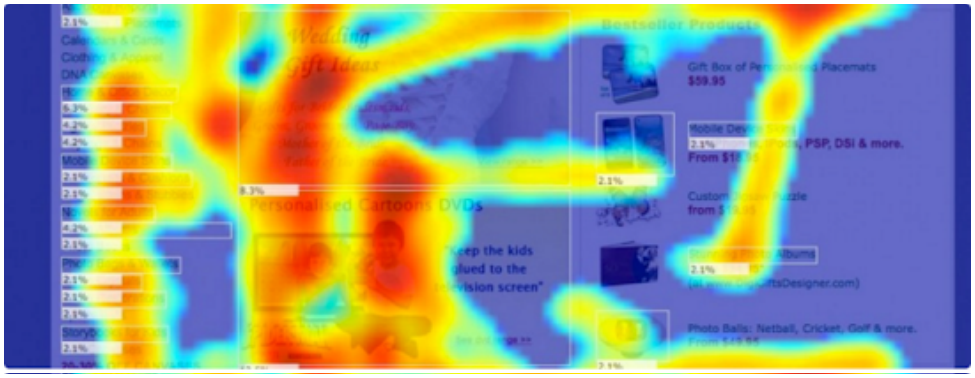
Another heat map example shows unemployment rates in August 2010 superimposed on a map of the United States (Yau, 2011). This is actually called a choropleth map. It uses graded differences in shading or color inside defined areas of the map to indicate the average values of some property or value of that area.

For this visualization, a linear color scale was used because it represents the distribution better and highlights the relatively high unemployment rates across the country. It has a legend, a title, and a brief explanatory paragraph explaining the story the data is telling.



In my own experience as a usability professional I analyze generated heat maps of website behavior. For example, I may want to understand which elements or areas of the page have the most engagement or click rates.

As an example I found online through a google search the heat map below represents the types that I may reveiw. Often times, the percentage of users that clicked on a particular element/area of the page will be overlaid.



Finally, sometimes when we conduct usability testing, we may employ eye tracking software to see where the participants look on the screen during a task. This can aid in knowing where on the screen they expect or tend to look for certain information. The results from eye tracking software are typically displayed in the form of heat maps similar to click based heat maps.

Heat maps are a useful way to communicate problem areas to less analytically savvy folks at a company. However, some studies have shown that there is often little correlation with where folks look and where folks move their mouse so the data needs to be taken with a grain of salt when making decisions about the importance of an element. Despite this, they are a good starting point.

Sources:

Berinato, Scott, (2016). Good Charts: The HBR guide to Making Smarter, More Persuasive Data Visualizations. Harvard Business Review Press. ISBN-10: 1633690709.

Cairo, Alberto, (2016). The Truthful art: data, charts, and maps for communication. New Riders. ISBN-10:0321934075.

Cairo, Alberto, (2013). The Functional art: an introduction to information graphics and visualization. New Riders. ISBN-10:0321834739.

Robbins, Naomi B., (2013). Creating More Effective Graphs: a succinct and highly readable guide to creating effective graphs. Chart House. ISBN: 9780985911126.

Yau, Nathan, (2011). Visualize This: The FlowingData Guide to Design, Visualization, and Statistics. Wiley Publishing Inc. ISBN: 9780470944882.