Article from

CompAct

April 2018
Issue 57

The How of Data Visualization

By Mary Pat Campbell

This is a fifth and final part of a continuing series on data visualization (aka dataviz):

- The Why of Data Visualization—Questions to ask When Visualizing Numerical Information (March 2016)

- The Who of Data Visualization—Major figures and Books in Advocating Data Visualization Best Practices (May 2016)

- The Where of Data Visualization—Websites to Polish Your Data Visualization Game (December 2016)

- The What of Data Visualization—Software to Implement Data Visualization (October 2017)

- The How of Data Visualization—Specific Data Visualization Techniques to Consider in Actuarial Practice

(The when of data visualization being NOW, of course.)

For this article, I'm going to touch on a few data visualization techniques for you to consider using in your own actuarial work, that you may not be as familiar with. I have mentioned a few of these techniques in prior articles, and in this case, I will be supplying a spreadsheet with all these implemented in Excel. These all have analogues in Python and R, as well as other common dataviz systems.

A quick note on software uses: the important thing is to consider the specific technique, why you would want to use it, and what it is best in accomplishing. Some of these methods are easier to implement in speci fic software systems currently. However, software, even technical software, is moving towards perpetual beta (i.e., always in development), you'll find many useful techniques spreading more broadly. The more useful the technique, and the more widely used, the more likely you'll find it easily implementable in multiple places.

In the following examples, I will be using data from the Public Plans Database, specifically looking at funded ratios by state.[1] The spreadsheet with all the finished examples can be found at my dropbox here: *https://www.dropbox.com/s/s9zfib1bvdgjcqq/How%20of%20dataviz%20-funded%20ratio%20viz.xlsx?dl=0*

CONDITIONAL FORMATTING

I demonstrated a use of conditional formatting in the article "The What of Data Visualization," and I saw Bob Crompton showed a use of conditional formatting in his article "Data Visualization for Model Controls" (*Predictive Analytics and Futurism* newsletter, June 2017). Conditional formatting has been available for use in Excel even before Excel 2007, but options have been added over time and certain built-in rules are in Excel itself.

Conditional formatting is unique to spreadsheet and table set-ups, because the concept is that your data—whether numeric or categorical—will be highlighted or colored in some way reflecting the contents of each cell. In Excel, one has been able to do conditional formatting based on a value being greater than or less than certain amounts, within a certain range, etc. One has been able to conditionally format a cell not only based on the value within that cell, but also based on a formula or value in a different cell. It has become quite flexible in its current implementation.[2]

Let me demonstrate some of the more complicated, built-in conditional formatting rules.

In Table 1, I have public pension plans and their funded ratios listed over time by funded ratio. I have selected the largest plans, by assets in 2015. I have data for 2001–2015, but I will show only the last five years in the following table.

Table 1
Our Original Data of Funded Ratios

GASB Funded Ratio by Fiscal Year	2011	2012	2013	2014	2015
California Teachers	69.3%	67.2%	66.9%	68.5%	68.5%
California PERF	82.6%	83.1%	75.2%	76.3%	73.1%
NY State & Local ERS	90.2%	87.2%	88.5%	92.0%	93.8%
Florida RS	86.9%	86.4%	85.4%	86.6%	86.5%
Texas Teachers	82.7%	81.9%	80.8%	80.2%	80.2%
New York State Teachers	96.7%	89.8%	87.5%	92.9%	94.2%
Wisconsin Retirement System	99.9%	99.9%	99.9%	100.0%	100.0%
Ohio PERS	77.4%	80.9%	82.4%	83.8%	85.0%
Ohio Teachers	58.8%	56.0%	66.3%	69.3%	69.3%
North Carolina Teachers and State Employees	94.0%	94.2%	94.8%	95.6%	92.5%

This is difficult to do much in the way of comparison. It's just a bunch of numbers, with a similar number of digits. Nothing stands out visually.

If I highlight the cells containing the funded ratio amounts, I can select "Conditional Formatting" from the Home ribbon, which gives me a variety of choices, as seen in Figure 1. There are rules

Figure 1
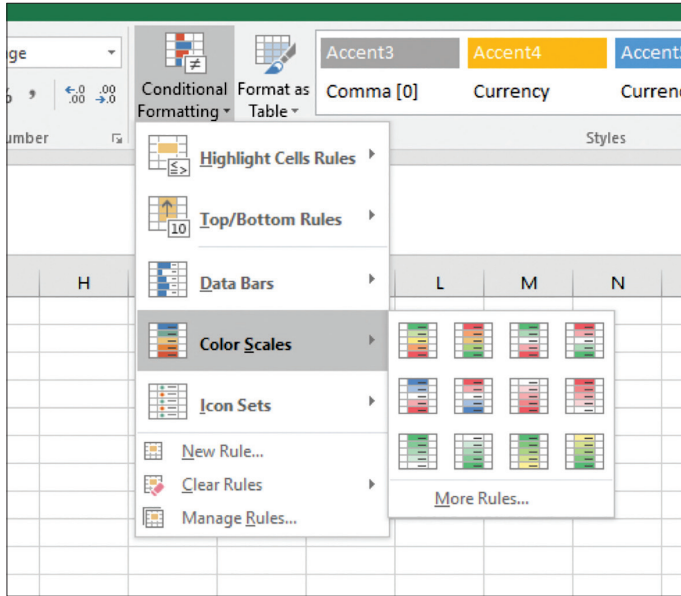Screenshot of the Conditional Formatting Dropdown Menu in Excel

Table 2
Conditionally Formatted Table, Using Excel Default Choices in Three-color Scale

GASB Funded Ratio by Fiscal Year	2011	2012	2013	2014	2015
California Teachers	69.3%	67.2%	66.9%	68.5%	68.5%
California PERF	82.6%	83.1%	75.2%	76.3%	73.1%
NY State & Local ERS	90.2%	87.2%	88.5%	92.0%	93.8%
Florida RS	86.9%	86.4%	85.4%	86.6%	86.5%
Texas Teachers	82.7%	81.9%	80.8%	80.2%	80.2%
New York State Teachers	96.7%	89.8%	87.5%	92.9%	94.2%
Wisconsin Retirement System	99.9%	99.9%	99.9%	100.0%	100.0%
Ohio PERS	77.4%	80.9%	82.4%	83.8%	85.0%
Ohio Teachers	58.8%	56.0%	66.3%	69.3%	69.3%
North Carolina Teachers and State Employees	94.0%	94.2%	94.8%	95.6%	92.5%

Figure 2
Conditional Formatting is Addictive … Beware of too Many Elements!

2011	2012	2013	2014	2015
69.3%	67.2%	66.9%	68.5%	68.5%
82.6%	83.1%	75.2%	76.3%	73.1%
90.2%	87.2%	88.5%	92.0%	93.8%
86.9%	86.4%	85.4%	86.6%	86.5%
82.7%	81.9%	80.8%	80.2%	80.2%
96.7%	89.8%	87.5%	92.9%	94.2%
99.9%	99.9%	99.9%	100.0%	100.0%
77.4%	80.9%	82.4%	83.8%	85.0%
58.8%	56.0%	66.3%	69.3%	69.3%
94.0%	94.2%	94.8%	95.6%	92.5%

to highlight top values (or lowest values), or highlight specific cells, but if I want to do a visualization comparing multiple quantities, I find the color scales the most useful of the choices.

This, after all, is a matter of the states, so I feel like using the Red/White/Blue color scheme, as seen in Table 2. Now we see that the California Teachers plan and Ohio Teachers plan have relatively low funded ratios and Wisconsin and New York State Teachers have relatively high ratios.

Conditional formatting is great for creating dashboards within your Excel spreadsheets in general. It can help you get a high-level view of data you're actively working with, and makes for a relatively simple interface if you want to create reporting dashboards.

Be careful: the conditional formatting rules can "stack." After one rule is implemented, if one highlights the same data … well, check out Figure 2. A little conditional formatting can go a long way.

SPARKLINES

Sparklines were popularized by Edward Tufte, who wrote the following:

> A sparkline is a small, intense, simple, word-sized graphic with typographic resolution.

> Sparklines mean that graphics are no longer cartoonish special occasions with captions and boxes, but rather sparkline graphics can be everywhere a word or number can be: embedded in a sentence, table, headline, map, spreadsheet, graphic. Data graphics should have the resolution of typography.[3]

The concept is to be able to mix text and other elements with data visualization. However, text generally works by having discrete forms from a small set, thus ensuring "readability" even when type size gets small. When graphs are shrunk to the height of text lines, certain small gradations are difficult to distinguish, so most sparkline implementations come with some features to emphasize high or low points. Sparklines tend to be most used with time series, as the natural tracking of quantitative entities over time works well with small lines, especially if the overall trend is of interest and not small differences.

Sparklines are also implemented in Excel, starting with Excel 2010.[4] In their Excel implementation, a single sparkline "lives" within a regular Excel cell, allowing one to put a mini-graph next to data, or making it easy to construct dashboards native to Excel. There are three types of Excel sparklines: line, column and win/loss. I will show examples of each, and why one may wish to use them.

Given that sparklines are intended to live side-by-side with text, the other major package/language they're implemented in is LaTeX. A pdfLaTeX implementation can be seen here: *https://ctan.org/pkg/sparklines*.[5] The sparklines package allows the mini-graphics to be embedded in a final pdf document created via LaTeX.

Let's go back to our table, and put back all the years 2001–2015. Say we want to compare the general trend for the funded ratios for all the plans, but don't want to have 10 lines on a single graph.

Sparklines can be found on the Insert ribbon in Excel, in its own grouping. Place the cursor in the cell in which you want to place the sparkline, choose which type you want, and select the data you want graphed. This can be seen in Figure 3.

The real magic comes when you copy one sparkline over a range. Just as with normal Excel formulas, this is treated in a relative manner, and the copied set are all linked to the original as a grouping. See Figure 4 to see the sparkline design menu, where you can edit features.

One can make several choices, such as forcing all the sparklines into the same vertical axis, highlighting the high or low point, and a few other choices. The point is to make small graphs with few features that are easily readable in a small space. One cannot provide many options at that point.

In my final version in Figure 5, I have hidden my data, marked the high and low points, and let the max and min for each graph differ by plan. The resulting visualization gives general trends but doesn't allow for comparison of actual amounts between plans. As this is a spreadsheet-native graphic, this is also useful for Excel-based dashboards.

SLOPE GRAPHS

Slope graphs are a bit odd, in that they are a type of graph you already know: line graphs. Except there are only two points on each line: the begin and the end. The standard style for a slope graph is that the name of the data series is at the begin and/or end of each line, instead of a legend separate from the data itself.

The point of slope graphs is to give a very high-level feel for changes in different quantities, but it's also a way to compare rankings of many items in a before-and-after sense. I've seen slope graphs used to show how survival rates for different types of cancer have changed for decades ago versus recently. Because each graph element is a straight line segment, it makes it relatively easy for people to have multiple lines on a single graph in order to compare them.

This is not a built-in graph type for Excel, but one can easily create these via the built-in line graphs, adding data labels, and coloring lines differently by slope (if one wants to make that distinction.) A few ways of implementing these in Excel can be seen in the endnotes.[6]

For the funded ratios from 2001 to 2015 for our top 10 pensions, see the trend in Figure 6 on the next page.

TILE GRID MAPS

Here's my favorite new dataviz technique I've been using: tile grid maps. I had originally thought about writing about
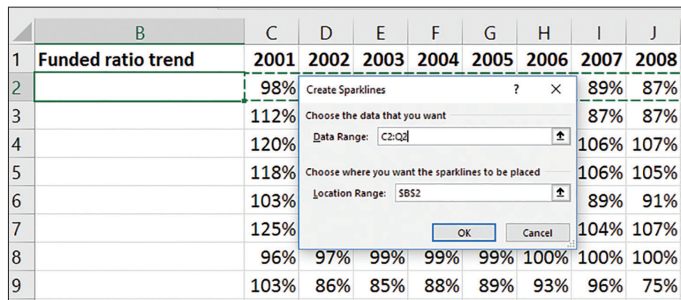
Figure 3
Setting the Sparkline for the First Cell

Figure 4
Sparkline Design Menu can be Seen in the Ribbon
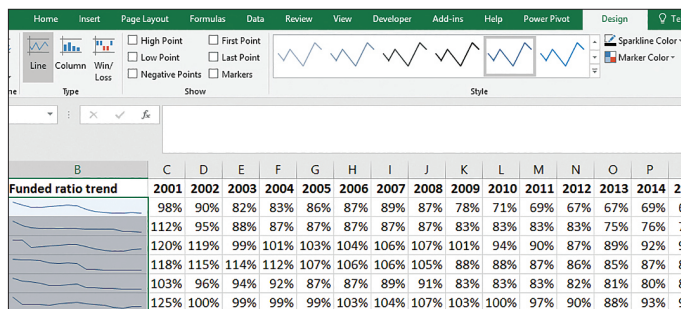
Figure 5
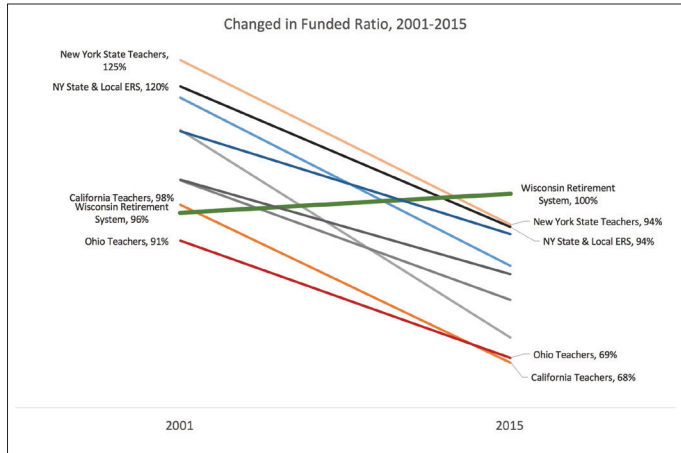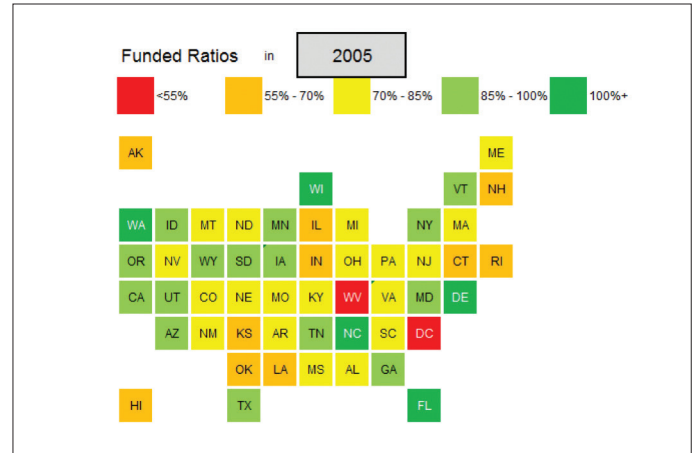A Final Table With Sparklines Within the Spreadsheet

Figure 6
On, Wisconsin!

Figure 7
Public Pension Funded Ratios Before …

choropleths, which is a way of coloring geographically-accurate maps based on underlying data—and even these are built into Excel now (I put one in the spreadsheet accompanying this article).

But do we really need to look at all the squiggly-lined items? Especially if we're looking at data that isn't necessarily geographically-determined (such as hurricane incidence) but more generic? Some people have done tables, column charts, and line graphs with the items marked by state abbreviation, but that's difficult to look at.

The tile grid map is a great solution: one uses the familiar general geographic placement of the states (or countries, or whatever geographic units) in a grid pattern (usually square, but can be hexagonal or other). Then one colors the regions based on value, similar to conditional formatting. Indeed, for an in-cell implementation in Excel, one uses conditional formatting![7] Figures 7, 8 and 9 show the tile grid map for the states for different years. (total pension assets over total pension liabilities). I used a drop-down menu to select the fiscal year, conditional formatting colors the squares based on a lookup formula, and the state abbreviations on each square? Well, check out the resources on tile grid maps to see how it was done.[8]

Tile grid maps are being used more in data journalism, but different publications prefer different configurations for the U.S. states. Where exactly should one put Rhode Island? To the east of Connecticut or east of Massachusetts?[9] (Once National Public Radio put Rhode Island to the west of Massachusetts in a post on its website, which was odd.)

There have been tile grid maps created for the countries of Europe, for the provinces of Canada, for the states of Mexico
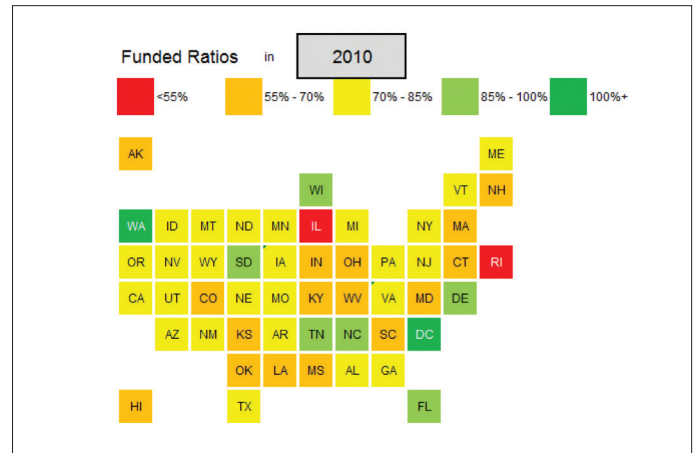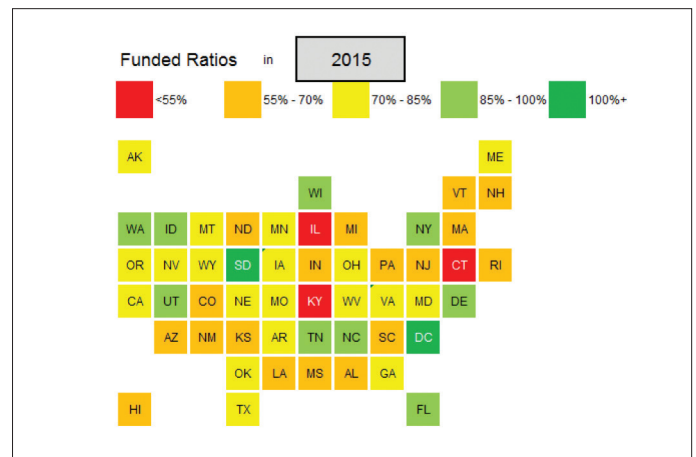
Figure 8
… and After the 2008 Market Crash

Figure 9
It's Still Getting Worse?

… for pretty much anything one wants, as they are easy to build compared to some other visualizations.

EXPLORE ON YOUR OWN!

In putting together this article, I had thought of techniques I like using that I don't think are necessarily commonly used, but I also used some online dataviz catalogs I found over the past year. I found the tile grid map through one of these galleries and realized it was a great solution for a problem I was having with particular data sets.

Why not check out some of these yourselves and try something new today!

Financial Times' visual vocabulary: *https://github.com/ft-iwnteractive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf*

Chartmaker Directory: *http://chartmaker.visualisingdata.com/*

Python Graph Gallery: *https://python-graph-gallery.com/*

R Graph Gallery: *www.r-graph-gallery.com*

Visualization Universe: *http://visualizationuniverse.com/charts/*

Graphic Continuum: *https://www.informationisbeautifulawards.com/showcase/611-the-graphic-continuum*

Data Viz Project: *http://datavizproject.com/*—2017 Kantar Information is Beautiful award winner

Let me know what your favorite new dataviz technique is! ■

Mary Pat Campbell, FSA, MAAA, is a vice president, Insurance Research at Conning in Hartford, Conn. She can be reached at *marypat.campbell@gmail.com.*

ENDNOTES

1 Public Plans Database: *http://publicplansdata.org/, downloadable data: http://publicplansdata.org/public-plans-database/download-full-data-set/* Accessed 11 December 2017.

2 Check out the Microsoft support on Conditional Formatting in Excel for some tips:*https://support.office.com/en-us/article/Enter-and-format-data-fef13169-0a84-4b92-a5ab-d856b0d7c1f7#ID0EAABAAA=Conditional_formatting*

3 Edward Tufte Forum: "Sparkline theory and practice". *https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001OR*, accessed 12 Jan 2018.

4 Office Support: "Use sparklines to show data trends". *https://support.office.com/en-us/article/use-sparklines-to-show-data-trends-1474e169-008c-4783-926b-5c60e620f5ca*, accessed 12 Jan 2018.

5 Sparklines package, version 1.7 as of 12 Jan 2018 (version date: 27 Dec 2016)

6 Make a slope graph in Excel, *http://stephanieevergreen.com/slopegraph/*, accessed 30 Jan 2018. How to Make Slope Graphs in Excel, *https://peltiertech.com/slope-graphs-in-excel/*, accessed 30 Jan 2018.

7 Tile grid maps in Excel, *https://policyviz.com/2016/04/13/tile-grid-maps-in-excel/*, accessed 30 Jan 2018

8 Other resources: *https://www.gislounge.com/how-to-make-a-tile-grid-map-using-excel/*, *https://policyviz.com/2016/05/05/hexagon-tile-map-excel/*, *https://www.linkedin.com/pulse/excel-map-hack-john-nelson/*, *https://policyviz.com/2017/05/04/european-tile-grid-map/*

9 *http://blog.yanofsky.info/post/117635988235/there-appears-to-be-some-disagreement-on-the*, accessed 30 Jan 2018.