Journal of Economic Perspectives—Volume 28, Number 1—Winter 2014—Pages 209–234

An Economist's Guide to Visualizing Data

Jonathan A. Schwabish

Once upon a time, a picture was worth a thousand words. But with online news, blogs, and social media, a good picture can now be worth so much more. Economists who want to disseminate their research, both inside and outside the seminar room, should invest some time in thinking about how to construct compelling and effective graphics.

An effective graph should tap into the brain's "pre-attentive visual processing" (Few 2004; Healey and Enns 2012). Because our eyes detect a limited set of visual characteristics, such as shape or contrast, we easily combine those characteristics and unconsciously perceive them as an image. In contrast to "attentive processing"—the conscious part of perception that allows us to perceive things serially—pre-attentive processing is done in parallel and is much faster. Pre-attentive processing allows the reader to perceive multiple basic visual elements simultaneously. Here is a simple example; count the occurrences of the number 3 in the following set:

1269548523612356987458245
0124036985702069568312781
2439862012478136982173256

Now repeat the task with this set of numbers:

1269548523612356987458245
0124036985702069568312781
2439862012478136982173256

■ *Jonathan A. Schwabish is a Principal Analyst, Congressional Budget Office, Washington, DC. His email address is jonathan.schwabish@cbo.gov.*

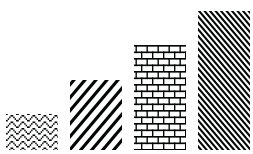
http://dx.doi.org/10.1257/jep.28.1.209 doi=10.1257/jep.28.1.209

The instances in the second set are much easier to find because they are encoded using a different pre-atttentive attribute—in this case, the intensity of boldface type.

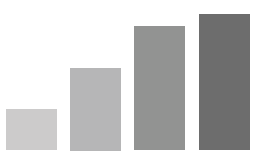
It takes imagination to create high-quality images that illustrate data accurately and effectively and also show some understanding and appreciation of how people acquire information. Indeed, what is known as the "Picture Superiority Effect" refers to our ability to retain more information seen through pictures than through words (for example, Medina 2008; Hockley and Bancroft 2011). There are thousands of approaches to presenting data: for starters, consider the vast information on how to choose fonts, colors, styles, layouts, and chart types. Three basic principles seem especially useful.

First, *show the data*. People read graphs in a research report, article, or blog to understand the story being told. The data are the most important part of the graph and should be presented in the clearest way possible. But that does not mean that *all* of the data must be shown—indeed, many graphs show too much.

Second, *reduce the clutter*. Chart clutter, the use of unnecessary or distracting visual elements, will tend to reduce effectiveness. Clutter comes in dark or heavy gridlines; unnecessary tick marks, labels, or text; unnecessary icons or pictures; ornamental shading and gradients; and unnecessary dimensions. Too often graphs use textured or filled gradients



when simple shades of a color could accomplish the same task.



In some cases, familiar data markers (■♦●×) are used to distinguish between several data series on a graph, but when the markers intersect and overlap they end up cluttering the patterns.

Third, *integrate the text and the graph*. Standard research reports often suffer from the "slideshow effect," in which the writer narrates the text elements that appear in the graph. A better model is one in which visualizations are constructed to complement the text and at the same time to contain enough information to stand alone (Corum 2013). As a simple example, legends that define or explain a line, bar, or point are often placed far from the content of the graph—off to the right or below the graph. Integrated legends—right below the title, directly on the chart, or at the end of a line—are more accessible.

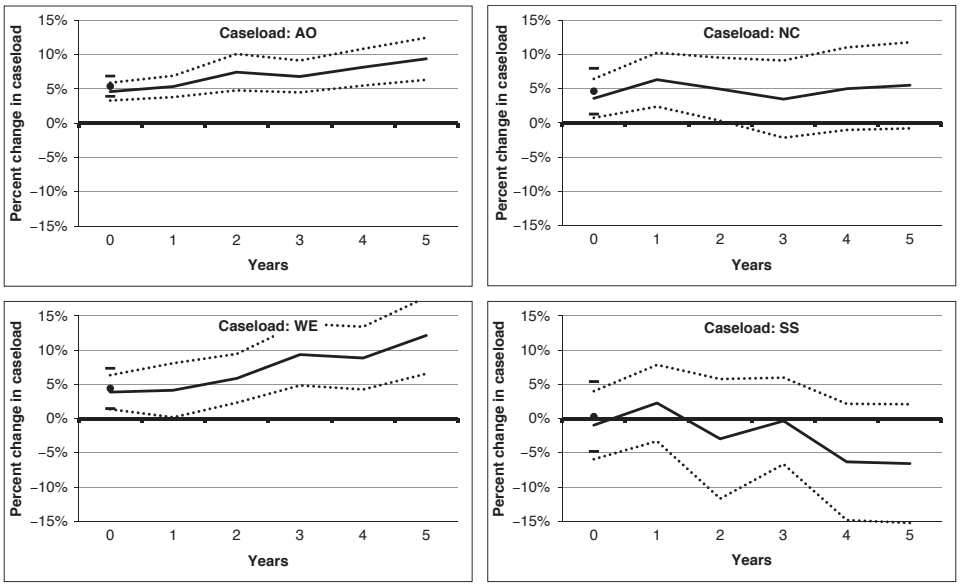
These three principles embody the idea that an author should support the reader's acquisition of information quickly and easily. By stripping out unnecessary clutter, emphasizing the data, and using certain pre-attentive attributes (for example, hue (color), size, orientation, and shape) graphs can more clearly and more effectively communicate information. However, default graph options in many statistical programs tend to add clutter and to separate text and graphs, and so researchers need to consider overriding those defaults and perhaps adding annotation to create graphs that communicate information more effectively to the reader.

This article encourages economists to think more strategically about how to visualize their data and presents some pathways to create better, more effective graphs. The next section demonstrates the principles in eight graphs remade using nothing more complicated than Excel. The discussion then addresses some types and purposes of different data visualizations, and briefly reviews some tools and sources of information that researchers can use to improve their existing graphs or to create new ones. One thing researchers should keep in mind is that graphs in research reports or articles, and even those shown in verbal presentations, are not meant for the author, but for the reader or the seminar audience. The line chart that a researcher uses in the data exploration phase—with default gridlines, tick marks, and colors—may not be the one that will best communicate the researcher's ideas to others. Discussions of data visualization are only now making their way into economics journals and conferences, but perhaps this is just the beginning and our discipline's understanding of the importance of good visualization will expand and grow.

Eight Graphs Transformed

Poor graphs communicate ineffectively, or even worse, provide a distorted impression of the data. This section shows how eight graphs could be redesigned to demonstrate the application of the three principles outlined above. Some decisions are subjective, of course—line thickness, series order, axis label style. Other decisions, I would argue, are just objectively *better* ways to convey meaning. All of the redesigned graphs were constructed in Excel and required slight variations from the program's default settings. Garamond—a classic serif font—is used to slightly distinguish the graphs from the *Journal of Economic Perspectives*'s Baskerville typeface. The electronic version of the *JEP* often uses color, which can be an important tool in data visualization to invoke emotion, emphasize graphical elements, or simply add aesthetic value. The print *JEP* does not use color, but all graphs that use color in the electronic version of the *JEP* are designed to work in greyscale for print readers. The choice of a color palette, like the choice of a font, also can be subjective, but following some basic guidelines can improve communication. In the final section, I discuss some tools that can help with those strategies and selections.

Figure 1A
An Original Line Chart



Source: Klerman and Danielson (2011).

The Line Chart

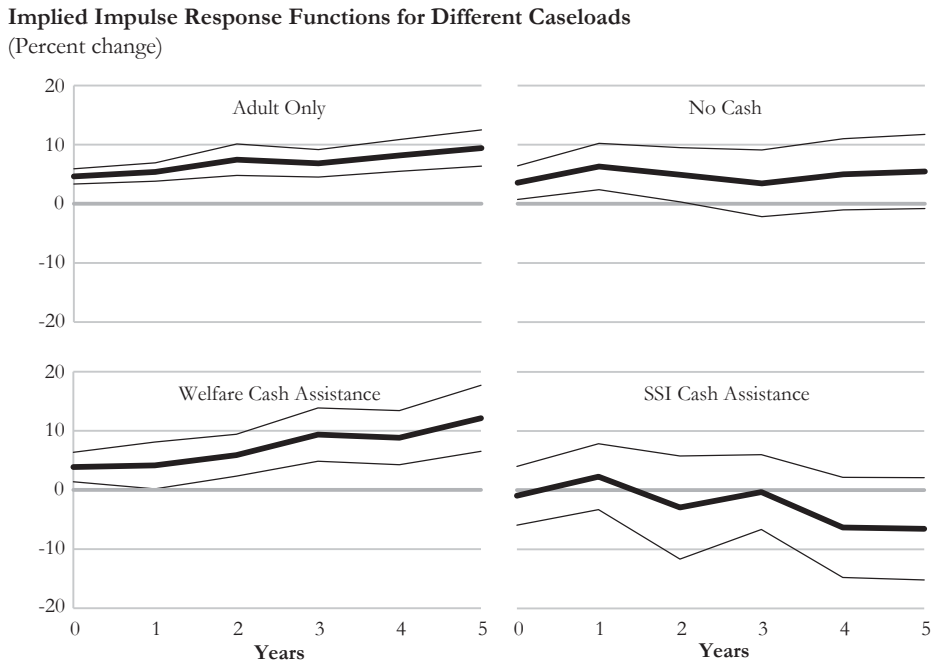
Figure 1A summarizes regression results of the correlation between the long-run unemployment rate in the United States and Supplemental Nutrition Assistance Program caseloads for four groups.[1] Instead of a single line chart with multiple series or four large, separate charts, the authors have used a "small multiples" approach in which four smaller charts are grouped together. Instead of packing as much information as possible into a single graph, researchers should probably use this approach more often. However, in other ways the graphs violate the three principles outlined above.

Perhaps most importantly, a graph should emphasize the data, but the darkest and thickest line on these graphs is the 0 percent gridline. Your eye is immediately drawn to that thick, horizontal gridline rather than to the important parts of the graph, namely the coefficient line and the standard errors. Also notice that the data values in the last point of the WE and SS charts actually exceed the 15 percent data marker—the graphs fail to show all data points.

Some elements add unneeded clutter. For example, the y-axis labels and percentage signs are redundant and add clutter (there are 28 percentage signs in all!) and the tick marks on the y-axes also seem unnecessary.

[1] The authors were kind enough to send me the data behind their figure for reproduction here.

Figure 1B
A Revised Line Chart

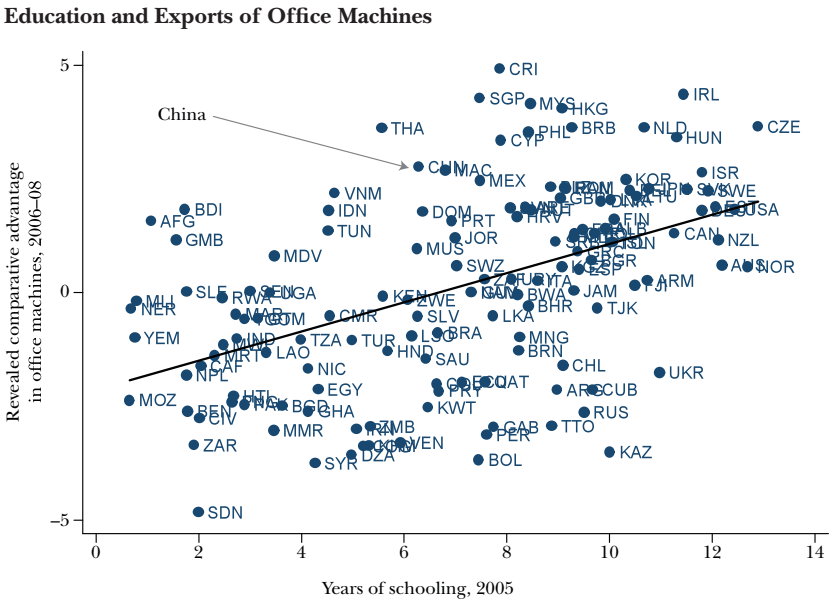


Finally, what do AO, NC, WE, and SS mean in the figure? In this article, these terms are explained on the third and fourth pages, 15 pages before this figure is presented. It seems unfair to ask a reader to search the paper to decode the meaning of those labels. Instead, the text and the graphs can be integrated.

Figure 1B offers an alternative version of this figure. The darkest line now shows the data (in this case, the coefficient estimate). The gridlines have been lightened, but leaving the 0 percent gridline slightly darker so it can provide a baseline for the series that dip below zero. Because the charts are aligned vertically and horizontally, I eliminated two sets of labels to reduce clutter and to help highlight the central message. I eliminated the percent signs and identified the unit in parentheses below the title. This last edit is essentially a style choice—there is nothing inherently wrong with using rotated text on the vertical axis, but it does require readers to turn the page sideways or tilt their heads (Robbins 2013a).

The title is now above the graph—some publishers place titles below graphs even though readers tend to start reading from the top left, move down along the left margin, and then move to the right. Repositioning the word "Caseload" into the title—used four times in the original—leaves room to spell out the abbreviations. One could, of course, argue about many of the choices illustrated here: Perhaps four sets of axis labels are better than two, or dotted standard error lines would be preferable to solid ones. But surely, a number of these changes are clear improvements on the original figure.

Figure 2A
A Clutterplot Example



Source: Hanson (2012).

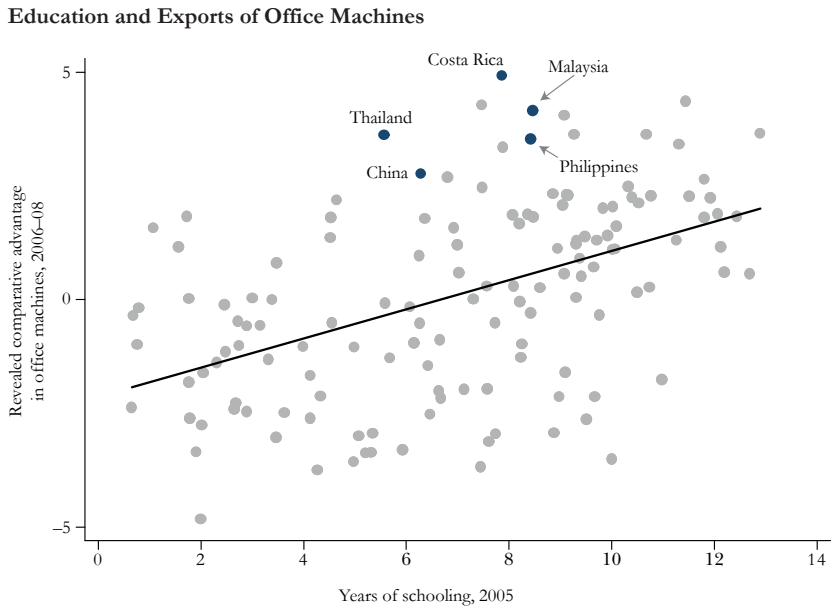
The Clutterplot

The next chart (Figures 2A and 2B) comes from the *JEP* (Hanson 2012, p. 54). The explanation in the original article is as follows (emphasis added):

> [The figure] plots countries' revealed comparative advantage in office machines . . . averaged over 2006 to 2008, against the average years of schooling of the adult population in 2005 . . . *China* is above the regression line, indicating that its specialization in the sector is greater than one would expect given its level of education, but it is hardly an extreme outlier. Other middle-income countries—including *Costa Rica*, the *Philippines, Malaysia*, and *Thailand*—have larger positive residuals.

Congratulations if you already know the three-letter codes for the five countries referenced in the text. Even if you do, try to find them in the haystack of labels and dots. The redesign offers an alternative that shows the data and reduces the clutter. I eliminated all labels other than those for the five countries under discussion, which I spelled out, and I made the five data points darker, thus deemphasizing the other points but still showing the important information.

Figure 2B
Revising the Clutterplot Example

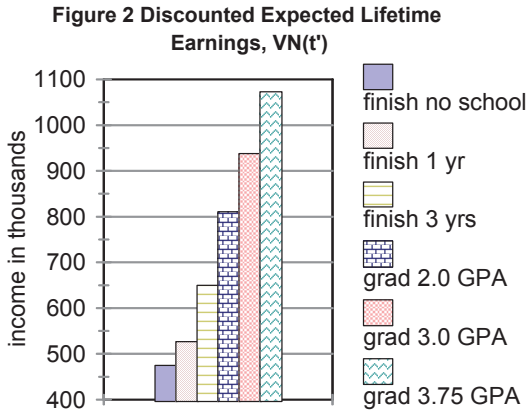


One potential objection to naming only some of the points is the concern that some readers may want to search through the figure for individual countries or identify various outlier points. There will often be some tension between presenting enough data to tell the story and presenting additional data that some readers might want. But most journals and many individual researchers now have their own websites, so posting the complete actual data for interested readers can be relatively simple and inexpensive. If it is important to include the data within the paper itself, a supplemental table in the text or as an appendix could be a more straightforward method of presentation.

The Basic Column Chart

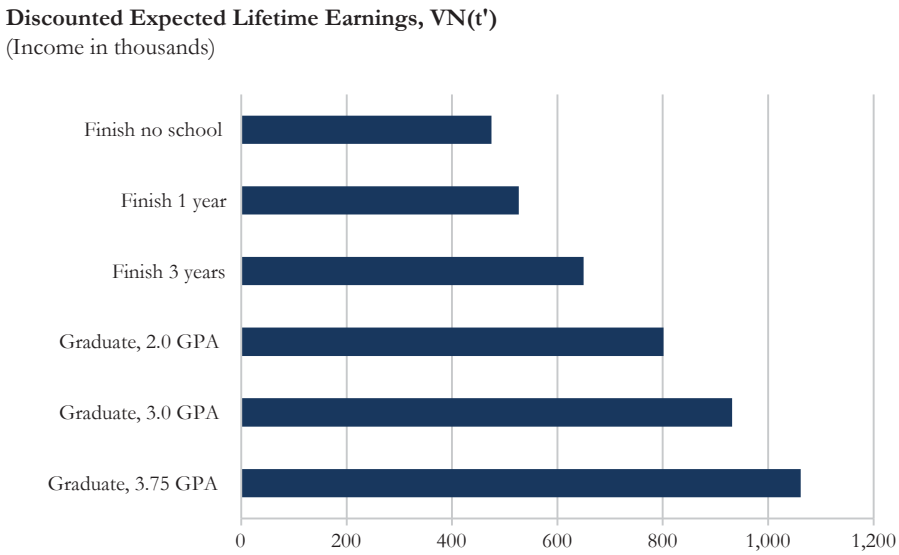
When it comes to bar and column charts, a first rule is to start the chart at zero. Otherwise, the differences between the columns are overemphasized. Figure 3A shows one such example of a column chart that does not start at zero. For example, if you look at the second-shortest bar, with a value just above 500, it certainly does not appear at a glance that it is half the length of the longest bar. This figure had different colors for each bar, which is not necessary but may be useful. The axis in the redesigned version, shown in Figure 3B, starts at zero and, to make room for the labels that are integrated with the chart, the figure is rotated horizontally. In this version, the second-shortest bar now appears at a glance to be about half the length of the longest bar.

Figure 3A
The Basic Column Chart



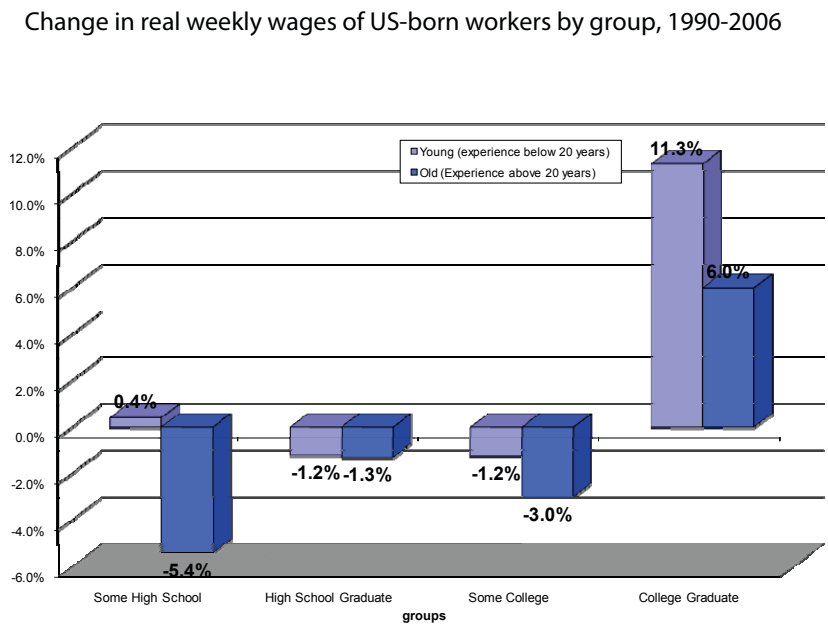
Source: Stinebrickner and Stinebrickner (2013).

Figure 3B
The Revised Column Chart



Source: Author's calculations using numbers inferred from text in Stinebrickner and Stinebrickner (2013).

Figure 4A
A 3D Chart



Source: Ottaviano and Peri (2008).

The 3D Chart

Figure 4A uses the now-familiar 3D effect. In such graphs, the third dimension does not plot data values, but it does add clutter to the chart and, worse, it can distort the information. Look at the far-right-hand bar, labeled 6 percent: No point of the column touches the gridline for that value. This software tool—like many others—uses perspective to give depth to the imaginary plane that runs across the top of the column, intersecting the gridline. But most readers will perceive the actual value of the column as less than 6 percent. Figure 4B shows a redesign: cancel the 3D treatment and integrate the disconnected legend with the graph. Notice that inserting the common baseline—portrayed in the original by a hovering, barely perceptible thin gray line—permits a more effective comparison among groups.

The Unbalanced Chart

The source material for Figure 5A originally appeared in an interactive visualization on the Organisation for Economic Co-operation and Development (OECD) website (http://www.oecd.org/gender/data/proportionofemployedwhoareseniormanagersbysex.htm); a static version was later reproduced in a *New York Times* Economix blog post (http://economix.blogs.nytimes.com/2013/04/02/comparing-the-worlds-glass-ceilings/?_r=2).

Figure 4B
Flattening a 3D Chart

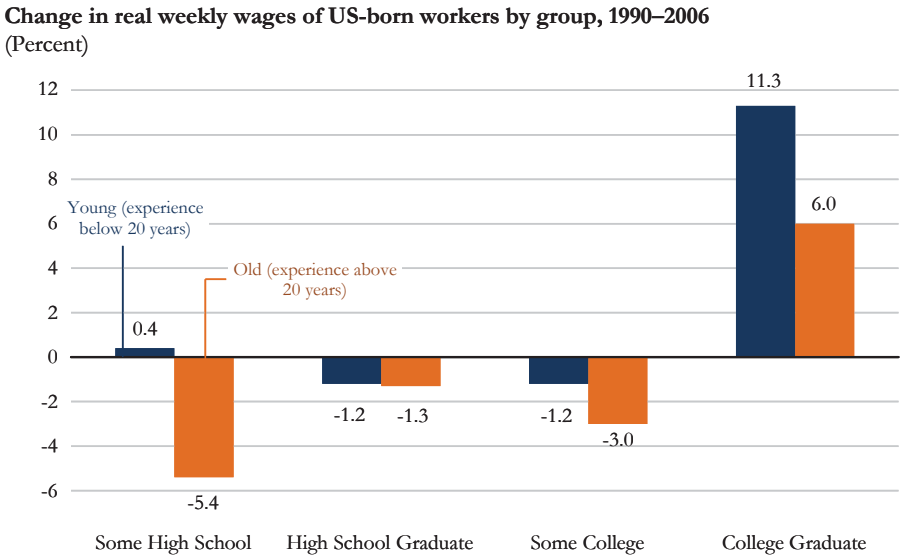


Figure 5A
An Unbalanced Chart

Source: Author, based on OECD (no date) and Rampell (2013).

The chart is an ineffective communication tool in either version for three main reasons. First, the same kinds of data are plotted using different types of encoding so that it is difficult to compare location (diamonds) with length (bars). Further, the relationship between the two values is obscured: The data points for men are too far away from those for women and there is no visual connection between them. The original versions used color—red (for women) and blue (for men) in the *New York Times* and orange and blue in the OECD version—which can serve to draw attention to or from certain elements.

Second, the columns for women take up a much larger proportion of the graph than do the diamonds for men, overemphasizing the data for women. If the intention is to give greater emphasis to the data for women, then a descriptive title—such as "Women's Employment as Senior Managers Averaged 6 Percent in 2008"—would have helped clarify the meaning. Furthermore, the gradient color shading in the columns is darker at the bottom than at the top where the data are truly encoded.

Finally, there are many gridlines, all heavy, and the percent signs on the y-axis labels are redundant. Additionally, the x-axis labels are potentially difficult to scan because they are vertical.

The chart could be redesigned in a number of ways. Paired bars could be used for men and women, or the figure could be turned into a table (Schwabish 2013a shows some other alternatives). Figure 5B shows a less-common type of visual approach. For some readers, these different types of plots may not be perceived as quickly as more commonly used graph types, such as bars or lines, but it's instructive to remember that scatterplots, not so long ago a novelty in mainstream publishing, now appear regularly. Just as our text literacy can expand with experience and exposure, so can our graphic literacy.

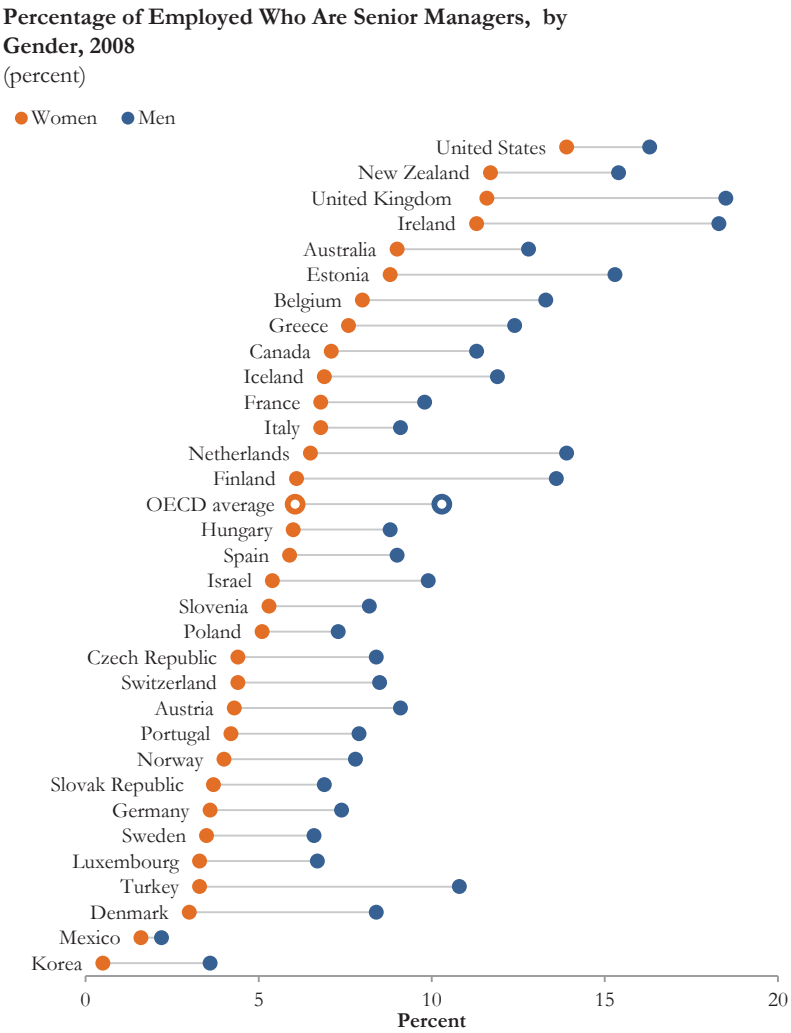
The redesigned chart shown in Figure 5B has several characteristics worth noting. First, the data are encoded similarly for men and women so that the reader can better perceive the connection between the two series and compare them. Second, the title, units, and legend are integrated and placed at the top-left of the chart, which helps the reader "enter" the chart. Third, the country labels are rotated horizontally and incorporated directly onto the chart with the thin gray connecting lines helping to illustrate the comparison between data points for men and women. Finally, the average value for the OECD as a whole is an unfilled circle (in a version where black and white printing is not considered, different shades or colors could be used instead).

One potential shortcoming of the redesign is the lack of vertical gridlines. In the original, the gridlines perhaps helped distinguish more specific values. There is often a tension in this aspect of data visualization—how much of the purpose of the graph is to explain an idea and how much is to provide specific data (Few 2005; Schwabish 2013b)?

The Spaghetti Chart

If a line chart has too many series, any single trend will be obscured. Such charts are sometimes called "spaghetti charts" (Nussbaumer 2013). If too much

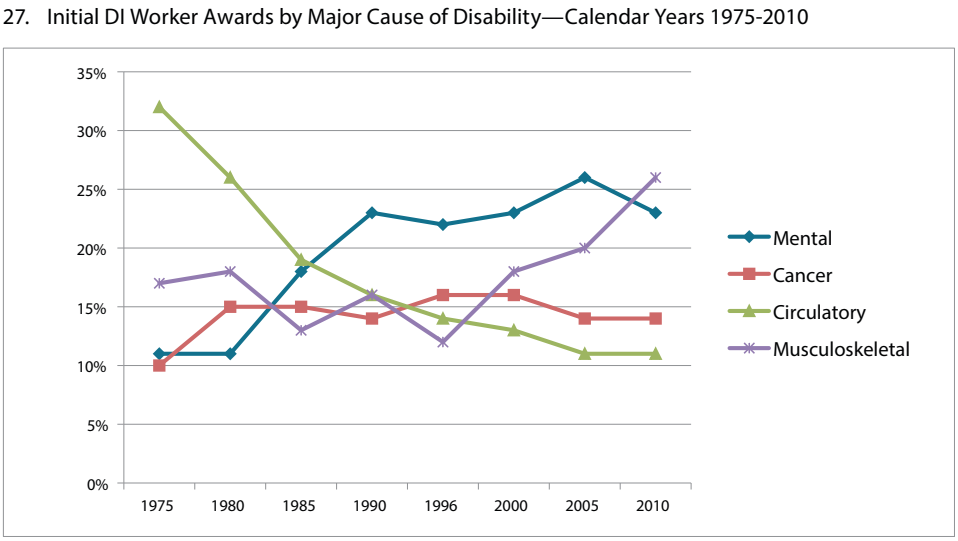
Figure 5B
Revising the Unbalanced Chart



information is plotted, readers can have difficulty pulling out the meaning of a single series or drawing an overall conclusion. Figure 6A is not an extreme example of a spaghetti chart, with only four lines. Nonetheless, consider some of its shortcomings: First, the data markers on every point make it difficult to pre-attentively follow any single series. Second, the legend is located far from the data and the order of the legend does not match the order of the lines.

One alternative to the spaghetti chart is to create smaller charts in series (sometimes called sparklines or small multiples; see Tufte 2006). Instead of a single, dense presentation of data, Figure 6B splits the information into four separate, smaller

Figure 6A
A Spaghetti Chart



Source: Social Security Advisory Board (2012).

Figure 6B
Revising the Spaghetti Chart

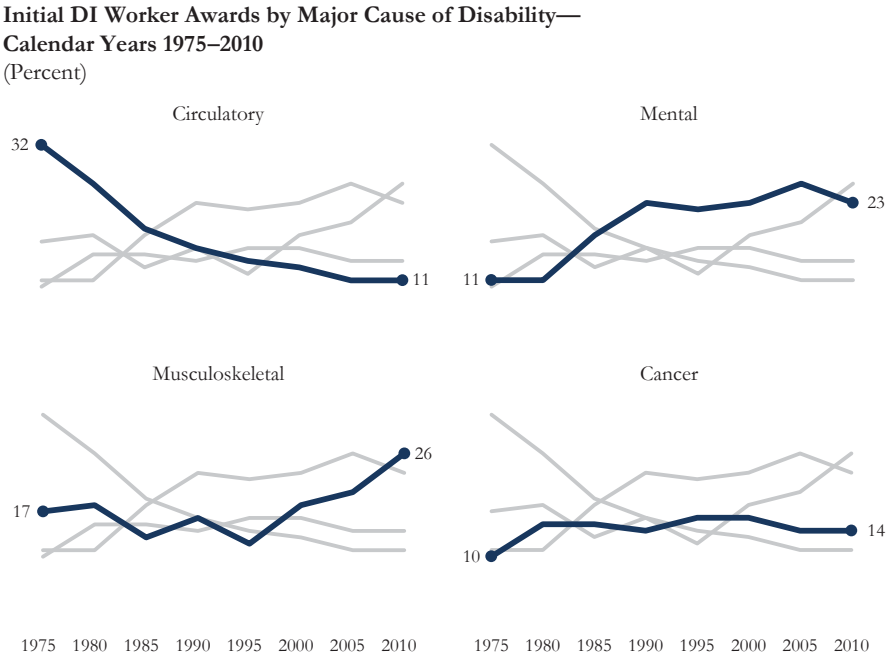


Figure 7A
A Pie Chart

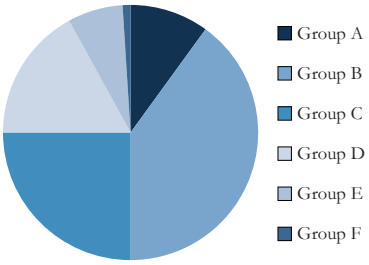
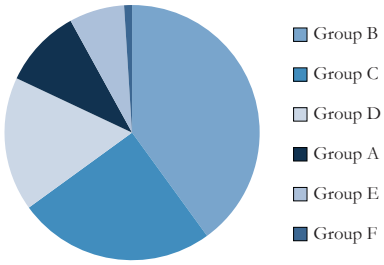


Figure 7B
B: A Pie Chart, Rotated



graphs to highlight the information in each line within the context of all the data. The contrast between light and dark helps to highlight specific trends and reduce clutter, as does the use of a label at either end of the main line in each set. (The y-axes are deleted, but they could be restored.) In this redesign, I have tried to emphasize the trends over time; my approach would differ if I were trying to emphasize specific numbers.

The Pie Chart

The debate concerning the effectiveness of pie charts is among the most contentious in the field of data visualization (much of the discussion in this section is based on Few 2007). Many people love pie charts—they are familiar, easily understood, and present "part-to-whole" relationships in an obvious way. But others argue that because pie charts force readers to make comparisons using the areas of the slices or the angles formed by the slices—something that our visual perception does not accurately support—they are not an effective way to communicate information. Donut charts—in which the center of the pie is punched out—just exacerbate the problem: The empty center makes the reader *estimate* the angle and arrive at other qualitative part-to-whole judgments without being able to see the center where the edges meet.

Pie chart slices that form 90-degree right angles—that is, slices that form one-quarter increments—are the most familiar to our eyes. Other amounts can be far more difficult to discern. The six segments in Figure 7A, for example, are presented clockwise in alphabetical order. Group C is easily identified as being about 25 percent of the whole. If, as in Figure 7B, however, the order of the segments is positioned so that the largest starts at the 12 o'clock position, the value of Group C is not so easily apprehended. One small change obscures the information.

To test your accuracy, try to guess the values of the slices: Figure 8A shows the answers. Was your guess about Group A correct? What about Group E? When I ask this question, guesses ordinarily range between 5% and 17% for each. Figure 8B offers a different approach: Add labels that integrate the data (instead of listing percentages in the notes or attaching them to the legend).

Figure 8A
A Pie Chart, Labeled

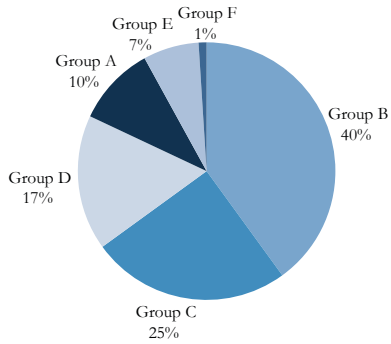
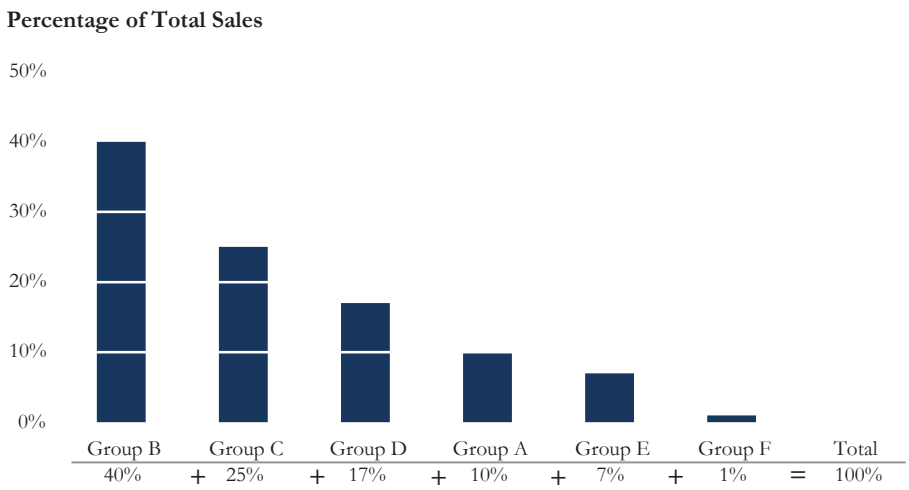
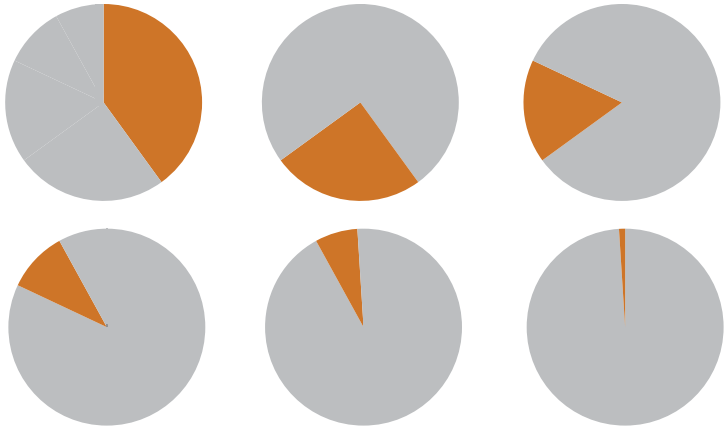


Figure 8B
Pie Chart Alternative: A Bar or Column Chart



This approach results in what amounts to two sets of information: the labels and the values for the slices (which readers might add up to arrive at 100 percent). This defeats the very purpose of the chart, which is to provide a visual representation of the data. It might be more effective just to present this information in the form of a short table. A bar or column chart could be another effective alternative: Figure 8B takes the guesswork out of identifying the value of each group and gives a clear picture of the relationships between the various groups, both in absolute amounts and in relative differences, as well as a ranking of the group from largest to smallest. The bar or column chart representation is best suited for comparing different segments; however, it is less efficient at helping us make part-to-whole

Figure 8C
Part-to-Whole Mini-Pie Charts



comparisons. In an attempt to add "part-to-whole" context—something that is often not that interesting in itself—the descriptive title at the top and the values and plus signs at the bottom reinforce the fact that the sum of the columns is 100 percent. (A completed graph could also delete the percent signs.)

Another point: Although pie charts typically ask us to compare all the data within a single space, they are by definition designed to facilitate part-to-whole comparisons. Thus, Figure 8C shows the true purpose of the pie chart, which is to individually compare each part to the whole (Camões 2013).

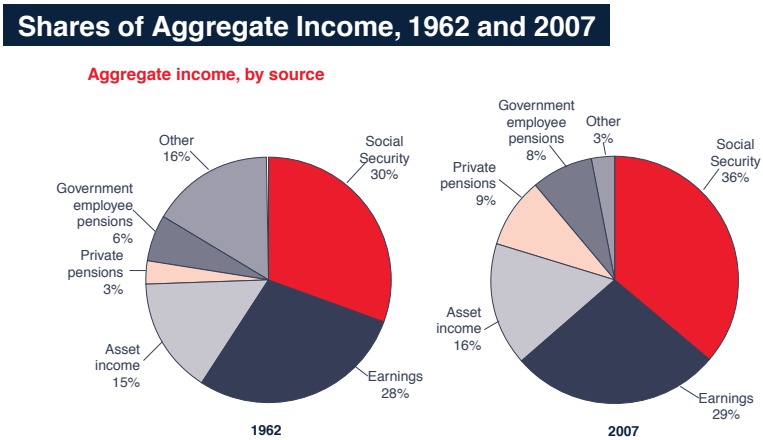
Another common—and less-than-optimal—way to present data is in a 3D pie chart. Like the 3D bar chart, such a treatment often presents the additional drawback of actually distorting the data. Whatever slice of a 3D pie chart that is toward the "front" tends to look bigger, because you can also see the 3D thickness for that slice. However, slices toward the "back" of a 3D pie chart tend to look smaller, because their "thickness" is only partially visible or not visible at all (Skau 2012).

As an example of the difficulty in discerning quantities from pie charts, Figure 9A makes the reader undertake a difficult comparison of a large number of segments—not just internally, but with another similar set right beside the first.[2]

The process of information transfer can be simplified by use of other graphic forms, which also offer the benefit of emphasizing different aspects of the analysis. For example, the *paired column chart* in Figure 9B fosters within-category comparisons. The vertical orientation requires that the labels either stretch across several lines or, worse, are rotated 90 degrees. A horizontal orientation would also work

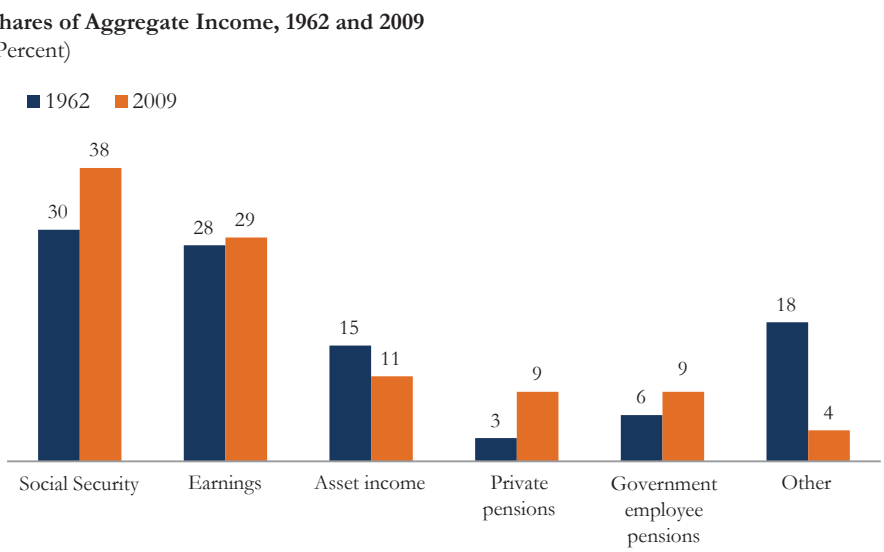
[2] Note that the pie chart for 1962 does not sum to 100 percent and there is a small, unlabeled gap left at the 12 o'clock position. For the redesigned graphs in Figures 9B–9D, I added the missing 2 percent from the 1962 series to the "Other" category. Brock (2013) shows other possible alternatives to paired pie charts.

Figure 9A
Two Pie Charts for Comparison



Source: Social Security Administration (2009).

Figure 9B
Alternative to a Pie Chart: A Paired Column Chart



in this case and is a useful approach when labels are difficult to fit in the vertical column chart layout (recall Figure 3; also see Schwabish 2013c). Also notice the (subjective) decision to omit the y-axis; the usefulness of the y-axis is doubtful with data labels placed on top of each column.

Figure 9C
Alternative to a Pie Chart: A Stacked Bar Chart

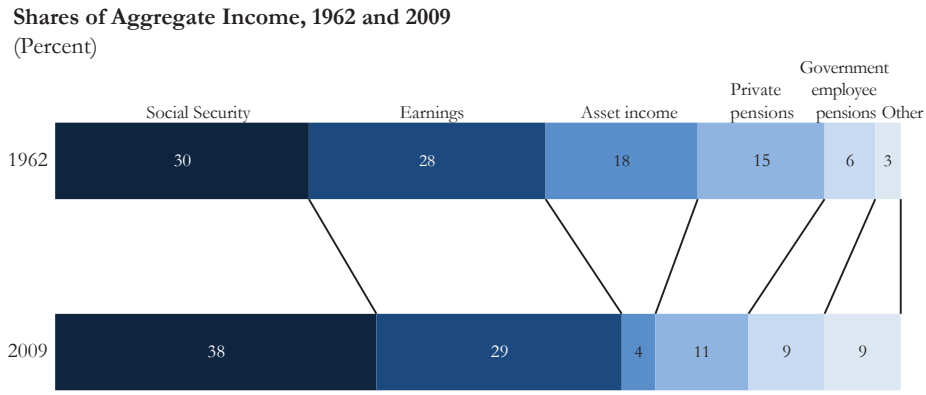
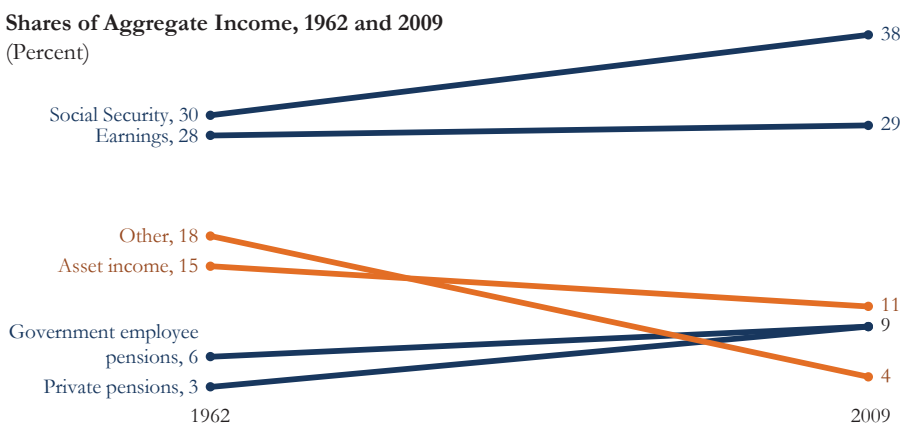
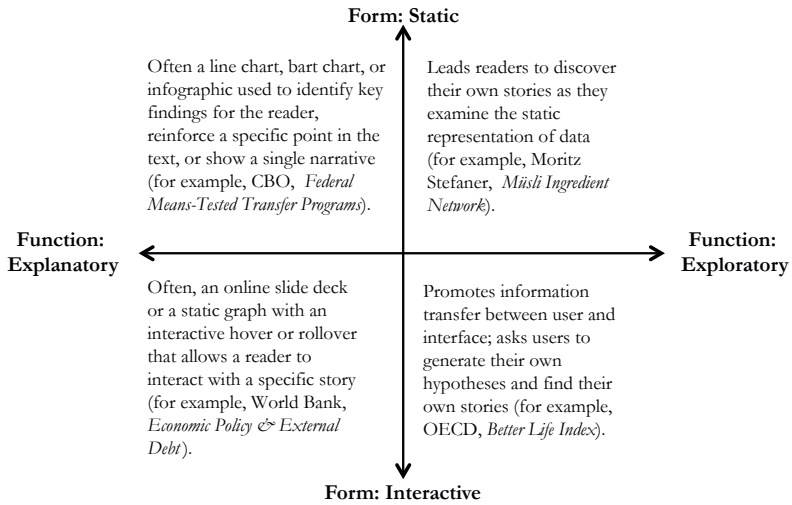


Figure 9D
Alternative to a Pie Chart: The Slope Chart



Alternatively, the *stacked bar chart* in Figure 9C shows the distribution of the various groups and that the groups sum to 100 percent, while also highlighting differences from one year to the other. Finally, the *slope chart* in Figure 9D also shows the difference in each category from the first year to the last by pairing points on two vertical axes. Slope charts can be used for a variety of purposes including showing correlations; for example, the relationship between a state's obesity rate and the share of people with at least a bachelor's degree (Cairo 2013). In this example, the color contrast (or what appears as different shades of grey in the black-and-white printed version) identifies which categories increased over time (blue; darker) and those that declined (orange; lighter).

Figure 10
Data Visualization: Form and Function



Source: CBO (Congressional Budget Office), *Means-Tested Transfer Program* (http://www.cbo.gov/publication/43935); Moritz Stefaner *Müsli Ingredient Network* (http://stefaner.eu/projects/musli-ingredient-network); World Bank, Economic Policy & External Debt (http://data.worldbank.org/topic/economic-policy-and-external-debt); and OECD, *Better Life Index* (http://www.oecdbetterlifeindex.org).

Form and Function

Once we move past the various strategies for presentation, it is useful to examine the various forms and functions of data visualization. First, consider the vertical axis in Figure 10, which illustrates the connection between the two general *forms* of visualization. *Static visualizations* provide all of the information at once and are not active or moving: for example, visualizations that appear on paper are static. *Interactive visualizations* allow a transfer of information between the figure and the user. Animated visualizations, which move but do not necessarily permit manipulation of data points to create alternative results—like movies or online slideshows where the user can control the pace of the story—can be thought of as falling between a static and an interactive visualization. Second, consider the horizontal axis in Figure 10, which considers the *function* of the visualization. *Explanatory visualizations* bring the main results to the forefront—they "surface key findings"—to some extent helping to reveal the story (for further discussion, see Segel and Heer 2010; Kosara and Mackinlay 2013). In comparison, *exploratory visualizations* help users interact with a dataset or subject matter to uncover the findings themselves. Such visualizations do not generally propose a single narrative or draw out specific insights.[3]

[3] Other mappings are of course possible and just as feasible (Bertin 1983; Harris 1996; Heer, Bostock, and Ogievetsky 2010; Kirk 2013; Kosara 2013a).

Economists typically live in a world of explanatory, static graphs. The graphs tend to be static and are used to reinforce a point made in accompanying text. Infographics (a longer form that tends to combine text, graphics, pictures, and icons) also typically are in that quadrant.

An example of a static explanatory figure would be one that portrays a single narrative, like the Congressional Budget Office's *Federal Means-Tested Transfer Programs* infographic (http://www.cbo.gov/publication/43935), which combines images, graphs, and text to tell a specific story. An example of a static exploratory figure is Moritz Stefaner's *Müsli Ingredient Network* (http://stefaner.eu/projects/musli-ingredient-network), which shows combinations of custom-ordered muesli ingredients. This approach leads readers to discover their own stories as they examine the static representation of data.

Interactive visualizations often are popular because they open the possibility of new and independent conclusions. They also enable the user to take those discoveries and produce something more explanatory. In general, the most effective interactive visualizations follow a three-step mantra: "Overview first, zoom and filter, then details-on-demand" (Shneiderman 1996). Such visualizations give a broad look at the graphic space, then allow readers to further define the space of interest, and finally permit them to capture specific details.

Perhaps the easiest explanatory interactive graph type to consider is a static graph that has an interactive hover or rollover layered on top (for example, the line graphs of economic indicators produced by the World Bank, http://data.worldbank.org/topic/economic-policy-and-external-debt). Exploratory interactive visualizations (such as OECD's *Better Life Index*, http://www.oecdbetterlifeindex.org) graphically present a complete dataset and ask users to find interesting stories. Subsequent links and reports then allow readers to find the details on demand.

Tools and Resources

Standard statistical software packages can generate basic static graphs, but people who want to improve their visualizations can take a few steps to reduce clutter and move away from default layouts, gridlines, colors, and fonts. The discussion below is not comprehensive and mentions of specific products are not intended as an endorsement but may serve as a starting point. A longer list can be found at the "Resources" page of my website (http://policyviz.com/resources/). Several tools—often free—can help users to make better use of their data in analysis or as they prepare for presentation or publication.

Color

Color could be the most misused and misunderstood aspect of design (Kosara 2013b). One simple strategy for improving the finished product is simply to avoid using default palettes. The red, green, and blue default in Excel is so pervasive that just changing to a different set of colors can make a graph more appealing.

Free online tools like Adobe Kuler (kuler.adobe.com/create/color-wheel), ColorBrewer2.0 (colorbrewer2.org), ColorSchemeDesigner (colorschemedesigner.com), and Colrd (colrd.com), allow users to create, modify, and export color palettes. The Instant Eyedropper tool (instant-eyedropper.com) allows users to choose colors from any image on a computer screen. Much has been written on color, but Ware (2012) offers perhaps the most comprehensive discussion. For situations such as publication in this journal, where one's graph may appear in both color and greyscale, there are tools—especially the ColorBrewer tool—that allow one to test grayscale-consistent color palettes.

Another important note about color is that about 10 percent of the population has some form of a color vision deficiency, and many of those people have difficulties distinguishing between greens and reds (see, for example, Coady 2013). Color Oracle (colororacle.org) and Vischeck (www.vischeck.com/) are two free tools that can simulate varieties of color blindness.

Fonts

With so many type choices on most people's computers or available online for free, it seems a shame to use the boring and overused Arial, Calibri, and Times New Roman typefaces. New fonts are designed all the time, and many old ones can give an image a fresh look. Good starting places are Font Squirrel (www.fontsquirrel.com) and Google Fonts (www.google.com/fonts). Spiekermann and Ginger (2003) and Bringhurst (2013) explain some of the science of typography and offer guides to choosing type for different purposes.

Visualization Tools

It is relatively simple to venture beyond the basic graphs available through the default settings in Stata, SAS, Excel, and other commonly available programs. The default graph in Stata, for example, has a blue background, and the first set of tick marks are not always located where the x- and y-axes intersect. Both of those conditions are easily changed. As alternatives, the open-source language R (www.r-project.org) offers more graphing capabilities. And although Excel is often dismissed as unimaginative, a variety of blogs, books, and websites provide tips and strategies to extend its capabilities (www.peltiertech.com is one).

It used to be that knowledge of HTML, JavaScript, or some other web-based programming language was a prerequisite for creating interactive visualizations. But an expanding set of available graphics tools require no more than the ability to drag and drop. Several different file types can be imported into Tableau (www.tableausoftware.com), for example, to create a variety of interactive graphics. Custom visualizations can be built using different programming languages such as HTML, JavaScript, or Processing (http://processing.org/). The R programming language also has a series of additions that enable users to create interactive visualizations: for examples, see, rCharts (http://ramnathv.github.io/rCharts/) and Shiny (http://www.rstudio.com/shiny/). The *New York Times*, for example, makes extensive use of the JavaScript library D3. Creator Mike Bostock (http://bost.ocks.org/mike/ and

bl.ocks.org/mbostock) provides a library of D3 visualizations, and Murray (2013) offers an introduction to the language.

Layout

The workhorse of most graphic designers is the Adobe Creative Suite, which includes programs such as InDesign, Illustrator, and PhotoShop. The free, open-source software Inkscape (http://inkscape.org/) is an alternative. Many books have been published on the topic of layout and design: Golombisky and Hagen (2010) offer a starting point to better understanding of design techniques, and Tondreau (2009) offers a good introduction to layout.

Mapping

Formal mapping software can be quite expensive and free online versions—from ArcGIS and ESRI, for example—are often relatively inflexible. Stata offers the "spmap" add-in (www.stata.com/support/faqs/graphics/spmap-and-maps), but the quality of the images is disappointing. StatPlanet (www.statsilk.com) is a free Flash-based program that imports data from Excel to create interactive visualizations. Interactive maps can also be constructed in Tableau, mentioned above. Another free tool, TileMill (http://www.mapbox.com/tilemill/) is HTML-based and thus may require a bit more time to learn and use, but it is slightly more flexible.

Infographic Tools

Rising interest in infographics has led to the creation of services that guide users through the design process. Like those for interactive visualizations, the new infographic packages are more user friendly than older tools. Examples include Datawrapper (http://datawrapper.de/), Infogr.am (http://infogr.am/), and Lemon.ly (http://lemon.ly/).

Resources

The amount of writing on data visualization has exploded over the past few years. The data visualization field moves fast—new work is constantly emerging, new products are constantly being released, and discussion and debates about best practices are continuous. Many books and blogs offer in-depth discussions on data visualization techniques and strategies or offer tutorials for the tools listed above.

Early and fundamental books are those by Tukey (1977), Bertin (1983), Cleveland (1993), and Tufte (2001[1983]). Wong (2010) and Robbins (2013b) are excellent sources for classifying specific charts for specific data types. In his books, Few (2009, 2012) dedicates a bit more time to examine the cognitive theory of data visualization and effective data visualization techniques. Recent books by Cairo (2013) and Yau (2011, 2013) are also excellent newer contributions to the field.

The number of blogs dedicated to the data visualization field is constantly growing. What follows is a very short list: Eager Eyes (eagereyes.org) is produced by Robert Kosara, a visual analysis researcher at Tableau Software and former computer science professor at UNC-Charlotte. Kosara often writes about the research side of

data and information visualization and offers constructive criticism and explorations. Flowing Data (flowingdata.com) is produced by author and statistician Nathan Yau, who provides a daily showcase of visualizations from around the web. He also posts visualization tutorials, primarily in the R programming language. Perceptual Edge (perceptualedge.com) is produced by author and consultant Stephen Few, who writes about good and bad trends in data visualization and promotes best practices with a focus on the visual aspects of human cognition. Junk Charts (junkcharts.typepad.com) is where Kaiser Fung collects and offers criticism of charts that do a poor job of presenting data. Visualising Data (visualisingdata.com) author Andy Kirk details the design process and trends in the field. Storytelling with Data (storytellingwithdata.com) is Cole Nussbaumer's blog where she offers practical examples for creating more effective visualizations. Finally, I offer practical data visualization examples on my website (policyviz.com/), as well as issues pertaining to creating more effective verbal presentations. On my companion site, HelpMeViz.com, readers can submit works in progress to seek advice and feedback from the data visualization community.

Conclusion

For economists who want readers to apprehend results quickly and accurately, presentation matters. Effective visualizations show the data to tell the story, reduce clutter to keep the focus on the important points, and integrate the text with the graphs to transfer information efficiently. With the increased flexibility of even fairly basic software programs (like Excel), it is now more cost-effective in terms of time and energy for researchers to invest some time learning and thinking about the details of graphical presentation.

To create great, effective visualizations, carefully consider the needs of your audience—the numbers, facts, or stories that will help them understand your ideas and your arguments. Consider the interfaces—static versus interactive—they will use. And pair the depth and clarity of your data, models, and writing with visualizations that are just as clear and compelling.

■ *The views in this article are mine and should not be interpreted as those of the Congressional Budget Office. For more information on data visualization and presentation techniques, see my site www.PolicyViz.com. I am grateful to the editors of this journal—David Autor, Chang-Tai Hseih, Ulrike Malmendier, and Timothy Taylor—for recognizing the importance of data visualization in the economics profession, and to Ann Norman and Annette Blanar for their painstaking proofreading and typesetting work. I also thank the authors of the graphs used in the paper and Alberto Cairo, Stephen Few, Molly Dahl, Kate Kelly, and Robert Kosara.*

References

Bertin, Jacques. 1983. *Semiology of Graphics: Diagrams, Networks, Maps*. New York, NY: Esri Press.

Bringhurst, Robert. 2013. *Elements of Typographic Style: Version 4.0, 20th Anniversary Edition*. Point Roberts, WA: Hartley and Marks.

Brock, Tim. 2013. "Improving on a Pair of Pie Charts." http://datatodisplay.com/blog/chart-design/improving-pair-pie-charts/.

Cairo, Alberto. 2013. *The Functional Art: An Introduction to Information Graphics and Visualization*. Berkeley, CA: New Riders.

Camões, Jorge. 2013. "Finally Revealed: The Optimal Number of Categories in a Pie Chart." http://www.excelcharts.com/blog/optimal-number-categories-pie-chart/.

Cleveland, William S. 1993. *Visualizing Data*. Summit, NJ: Hobart Press.

Coady, Geri. 2013. *Colour Accessibility*. (The Pocket Guide series: Collection 3.) Five Simple Steps. http://www.fivesimplesteps.com/products/the-pocket-guide-series-collection-three.

Corum, Jonathan. 2013. "Storytelling with Data." Opening keynote address at the Tapestry Conference, held in Nashville, TN, February 23. http://style.org/tapestry/.

Few, Stephen. 2004. "Tapping the Power of Visual Perception." *Visual Business Intelligence Newsletter,* September 4. http://www.perceptualedge.com/articles/ie/visual_perception.pdf.

Few, Stephen. 2005. "Grid Lines in Graphs are Rarely Useful." *Visual Business Intelligence Newsletter,* February. http://www.perceptualedge.com/articles/dmreview/grid_lines.pdf.

Few, Stephen. 2007. "Save the Pies for Dessert." *Visual Business Intelligence Newsletter* (August).

Few, Stephen. 2009. *Now You See It*. Burlingame, CA: Analytics Press.

Few, Stephen. 2012. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. 2nd ed. Burlingame, CA: Analytics Press.

Golombisky, Kim, and Rebecca Hagen. 2010. *White Space is Not Your Enemy: A Beginner's Guide to Communicating Visually through Graphic, Web & Multimedia Design*. Burlington, MA: Elsevier, Inc.

Hanson, Gordon H. 2012. "The Rise of Middle Kingdoms: Emerging Economies in Global Trade." *Journal of Economic Perspectives* 26(2): 41–64.

Harris, Robert L. 1996. *Information Graphics: A Comprehensive Illustrated Reference*. New York, NY: Oxford University Press.

Healey, Christopher G., and James T. Enns. 2012. "Attention and Visual Memory in Visualization and Computer Graphics." *IEEE Transactions on Visualization and Computer Graphics* 18(7): 1170–88.

Heer, Jeffrey, Michael Bostock, and Vadim Ogievetsky. 2010. "A Tour through the Visualization Zoo." ACMQueue, May 13. http://queue.acm.org/detail.cfm?id=1805128.

Hockley, William E., and Tyler Bancroft. 2011. "Extensions of the Picture Superiority Effect in Associative Recognition." *Canadian Journal of Experimental Psychology* 65(4): 236–56.

Kirk, Andy. 2013. "Discussion: Storytelling and Success Stories." April. http://www.visualisingdata.com/index.php/2013/04/discussion-storytelling-and-success-stories/.

Klerman, Jacob Alex, and Caroline Danielson. 2011. "The Transformation of the Supplemental Nutrition Assistance Program." *Journal of Policy Analysis and Management* 30(4): 863–88.

Kosara, Robert. 2013a. "Visual Storytelling in the Age of Data." Keynote address at the Tapestry Conference, held February in Nashville, TN. http://www.youtube.com/watch?v=qSYEjhR2AwQ&feature=youtu.be.

Kosara, Robert. 2013b. "How the Rainbow Color Map Misleads." July 7. http://eagereyes.org/basics/rainbow-color-map.

Kosara, Robert, and Jock Mackinlay. 2013. "Storytelling: The Next Step for Visualization." *Computer (Special Issue on Cutting-Edge Research in Visualization)* 46(5): 44–50.

Medina, John. 2008. *Brain Rules: 12 Principles for Surviving and Thriving at Work, Home, and School*. Pear Press.

Murray, Scott. 2013. *Interactive Data Visualization for the Web: An Introduction to Designing with D3*. Sebastopol, CA: O'Reilly Media, Inc.

Nussbaumer, Cole. 2013. "Strategies for Avoiding the Spaghetti Graph." March 14. http://www.storytellingwithdata.com/2013/03/avoiding-spaghetti-graph.html.

Organisation for Economic Co-operation and Development. Accessed, August 2013. "Percentage of Employed Who are Senior Managers, by Sex." http://www.oecd.org/gender/data/proportionofemployedwhoareseniormanagersbysex.htm.

Ottaviano, Gianmarco I. P., and Giovanni Peri. 2008. "Immigration and National Wages: Clarifying the Theory and the Empirics." NBER Working Paper 14188 (July).

Rampell, Catherine. 2013. "Comparing the World's Glass Ceilings." *Economix*, April 2. http://economix.blogs.nytimes.com/2013/04/02/comparing-the-worlds-glass-ceilings/?_r=0.

Robbins, Naomi. 2013a. "How to Position Y-Axis Labels in Graphs." February 12. http://www.forbes.com/sites/naomirobbins/2013/02/12/how-to-position-y-axis-labels-in-graphs/.

Robbins, Naomi. 2013b. *Creating More Effective Graphs*. 2nd ed. Chart House.

Schwabish, Jonathan. 2013a. "Mind the Gap—An Economic Remake." The Why Axis, April. http://thewhyaxis.info/gap-remake/.

Schwabish, Jonathan. 2013b. "To Label or Not to Label? That *Is* the Question." June 10. http://www.allanalytics.com/author.asp?section_id=3072&doc_id=264322.

Schwabish, Jonathan. 2013c. "Visualizing Data: Bad Labels, Easy Fix." May 2. http://www.allanalytics.com/author.asp?section_id=3072&doc_id=262539.

Segel, Edward, and Jeffrey Heer. 2010. "Narrative Visualization: Telling Stories with Data." *IEEE Transactions on Visualization and Computer Graphics* 16(6): 1139–48.

Shneiderman, Ben. 1996. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." In *Proceedings of the IEEE Symposium on Visual Languages*, IEEE Computer Society Press, 336–43.

Skau, Drew. 2012. "2D's Company, 3D's a Crowd." http://blog.visual.ly/2ds-company-3ds-a-crowd/.

Social Security Administration. 2009. *Fast Facts & Figures About Social Security, 2009*. Social Security Administration (July). http://www.ssa.gov/policy/docs/chartbooks/fast_facts/2009/fast_facts09.pdf.

Social Security Advisory Board. 2012. *Aspects of Disability Decision Making: Data and Materials*. Social Security Administration (February), http://www.ssab.gov/Publications/Disability/GPO_Chartbook_FINAL_06122012.pdf.

Spiekermann, Erik, and E. M. Ginger. 2003. *Stop Stealing Sheep & Find Out How Type Works,* 2nd ed. Berkeley, CA: Adobe Press.

Stinebrickner, Ralph, and Todd Stinebrickner. 2013. "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model." *Western University CIBC Working Paper 2013-5* (July), http://economics.uwo.ca/cibc/workingpapers_docs/wp2013/Stinebrickner_Stinebrickner05.pdf.

Tondreau, Beth. 2009. *Layout Essentials: 100 Design Principles for Using Grids*. Beverly, MA: Rockport Publishers.

Tukey, John W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison Wesley.

Tufte, Edward. 2001 [1983]. *The Visual Display of Quantitative Information*. 2nd ed. (First edition 1983). Cheshire, CT: Graphics Press.

Tufte, Edward. 2006. *Beautiful Evidence*. Cheshire, CT: Graphics Press.

Ware, Colin. 2012. *Information Visualization: Perception for Design*. 3rd ed. Waltham, MA: Morgan Kaufman.

Wong, Dona M. 2010. *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures*. New York: W. W. Norton and Company.

Yau, Nathan. 2011. *Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics*. Indianapolis, IN: Wiley Publishing, Inc.

Yau, Nathan. 2013. *Data Points: Visualization that Means Something*. Indianapolis, IN: Wiley Publishing, Inc.