<u>Addendum</u>

Followup Material from
"Best Practices in Graphical Data Presentation" Workshop
2010 Library Assessment Conference Baltimore, MD, October 25 - 27, 2010

COLOR PALETTES

Two slides from the workshop presentation were *color palettes*. The image here is one of those, from Dona Wong's *The Wall Street Journal Guide to Information Graphics*:

When there is a need for a variety of colors in a single graph (e.g., a line graph depicting multiple measures, a scatter plot with a single measure for five or six sub-populations) you read the palette vertically. Select one color for each measure or sub-group being represented. Palette designers chose the hues to make lines and symbols maximally distinguishable from others. In that same vein, you should avoid using too similar hues (such as violet and dark blue and black) together in a line graph or scatter plot. (Do not represent a single measure with multiple colors in line or bar charts!)

To choose colors for bar chars (or pie charts, if you can't avoid using them) read the palette chart horizontally. To represent data on a single measure as they pertain to sub-groups, use different intensities of a single hue or related hues that are adjacent on the color wheel.

The legend in this example shows Tableau Software's default color palette:
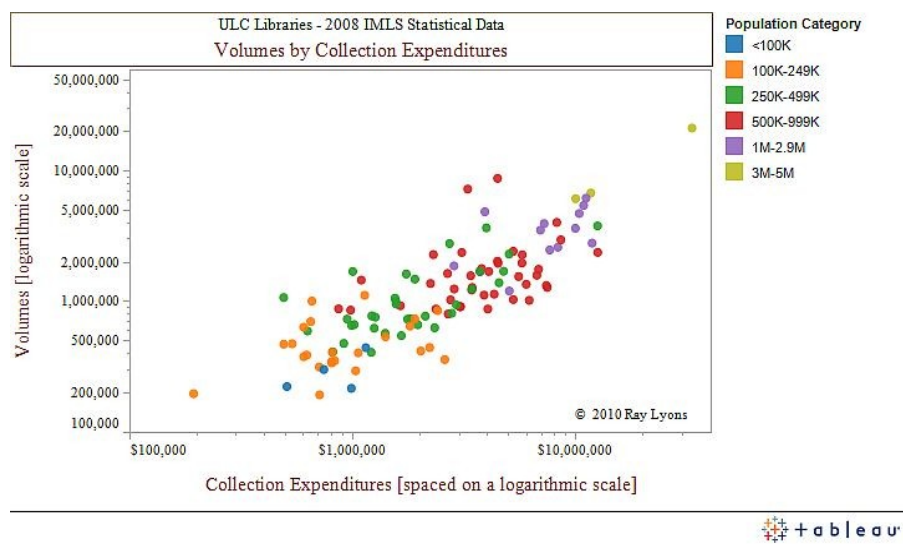
PATTERNED LINES IN BLACK AND WHITE GRAPHS

We didn't talk about this, but there is a movement *away* from dashed and dotted lines in black-and-white graphs these days because they can inadvertently emphasize trends that shouldn't be emphasized. (A solid line implies more importance than a dotted one; thicker or longer dashes can imply the same.) So, be careful before proceeding with dotted or dashed lines in these graphs. The best approach is to *label each line* and forget the patterning, if the lines aren't intertwined. If they are, use identical solid lines with plot symbols (diamonds, triangles, circles, etc.)

FILTERING

Just a quick mention of a feature of interactive graphical software that Stephen Few advocates for: *Filtering* is the ability to selectively view sub-groups, or single measures plotted in a line graph of multiple measures. The image below is the same graph as on the prior page (created in Tableau). But this rendition of the graph has libraries serving community populations of 100K to 249K highlighted. Other circles are grayed out, similar to the scheme we discussed for bar charts where a single color can highlight one value and the rest are rendered in a single, blander color.
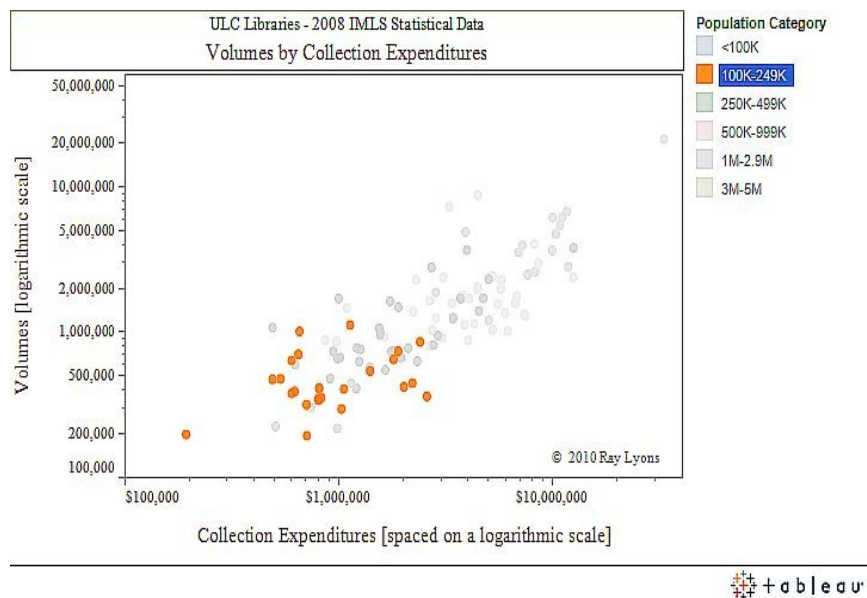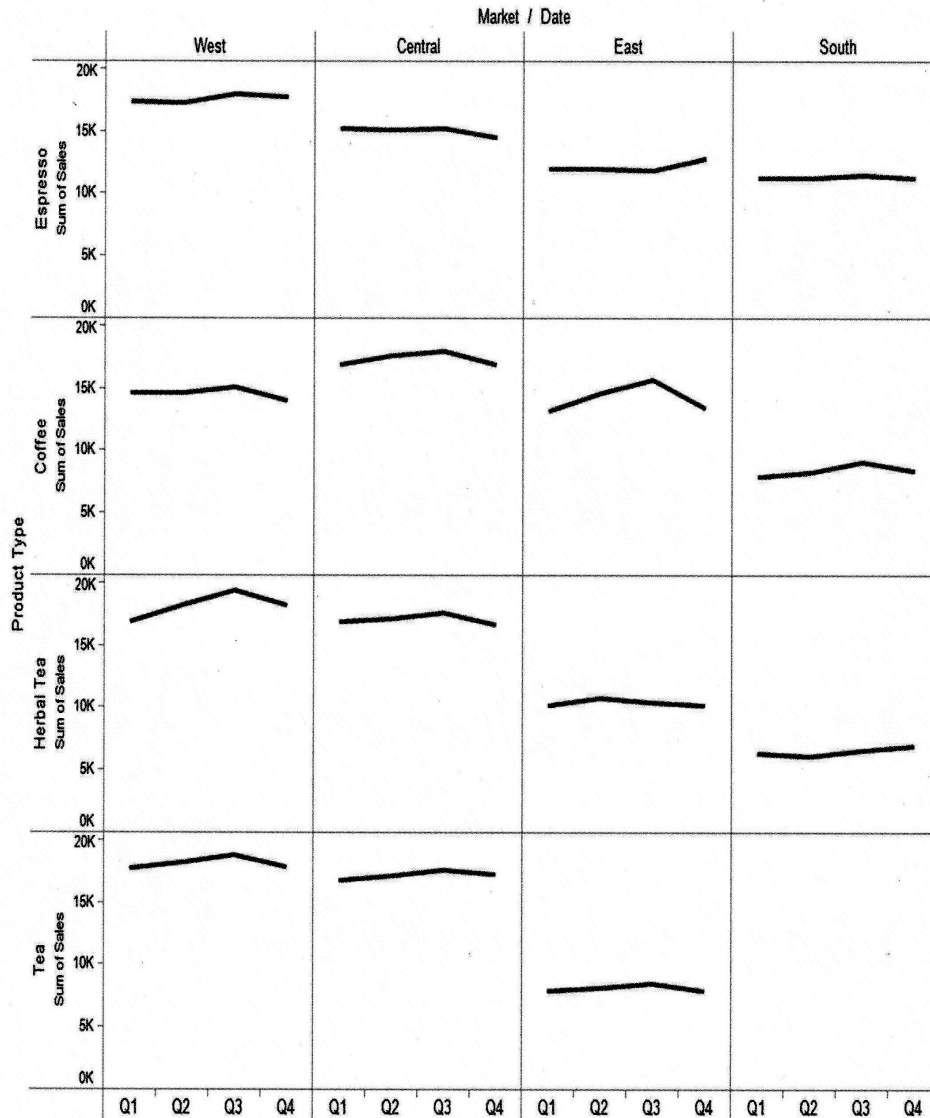

TRELLIS GRAPHS

While we discussed the idea, I never put the official name on a very useful alternative graphing format: *trellis graphs*. Multiple measures can be displayed in a series of small graphs, with their axes appropriately scaled and aligned. This is a good solution for graphs containing too many measures, like the multiply-colored bar charts and "spidery" graphs we used in the workshop exercises (also shown below).
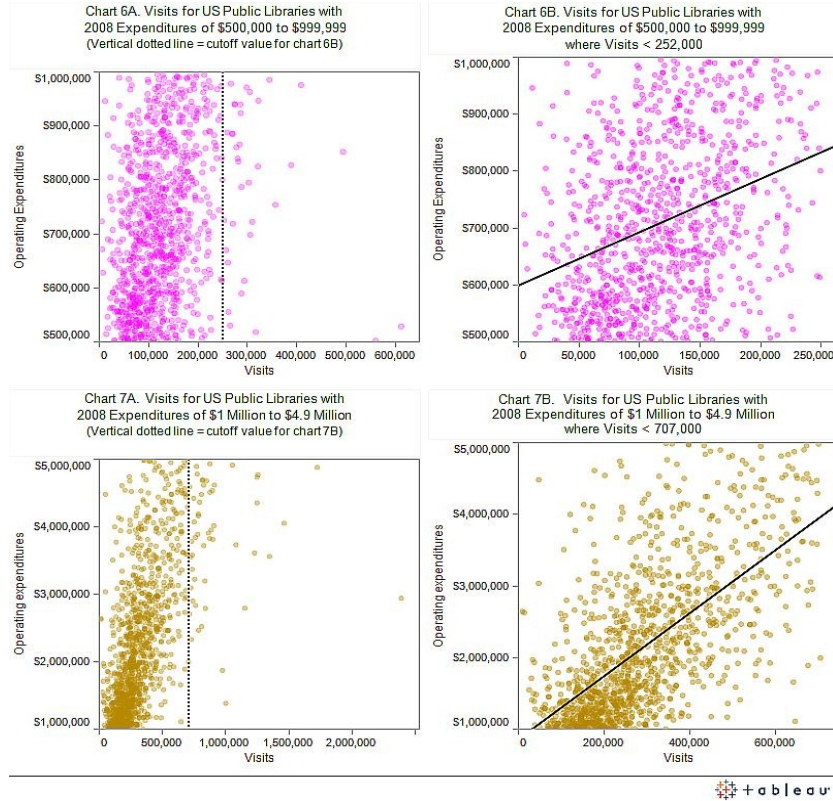
raylyons@gmail.com

Vertical and horizontal axes need to have equivalent scaling on one or both axes, depending on the data and your purpose(s). This is important for making visual comparisons accurate and fair.

A minimalist version of this basic trellis layout is shown below. A second example follows, showing scatter plots comparing two different subsets of public libraries on a single measure with an equivalent vertical scale.

Careful, because trellis graphs can be *misused* to compare apples and oranges. So, the responsibility is yours to ensure that comparisons are justifiable.


Source: Stephen Few, 2009, *Now You See It*

Chart 6A. Visits for US Public Libraries with
2008 Expenditures of $500,000 to $999,999
(Vertical dotted line = cutoff value for chart 6B)

Chart 6B. Visits for US Public Libraries with
2008 Expenditures of $500,000 to $999,999
where Visits < 252,000

Chart 7A. Visits for US Public Libraries with
2008 Expenditures of $1 Million to $4.9 Million
(Vertical dotted line = cutoff value for chart 7B)

Chart 7B. Visits for US Public Libraries with
2008 Expenditures of $1 Million to $4.9 Million
where Visits < 707,000

Source: Ray Lyons, *Lib(rary) Performance Blog,* Aug. 24, 2010. http://libperform.wordpress.com/2010/08/24/e-pluribus-unum/

Note: One difficulty with Tableau Public Software is getting it to align the graphs exactly.

Data from two graphs shown below, from the group exercise in the workshop, would be much better presented in a trellis graph format:
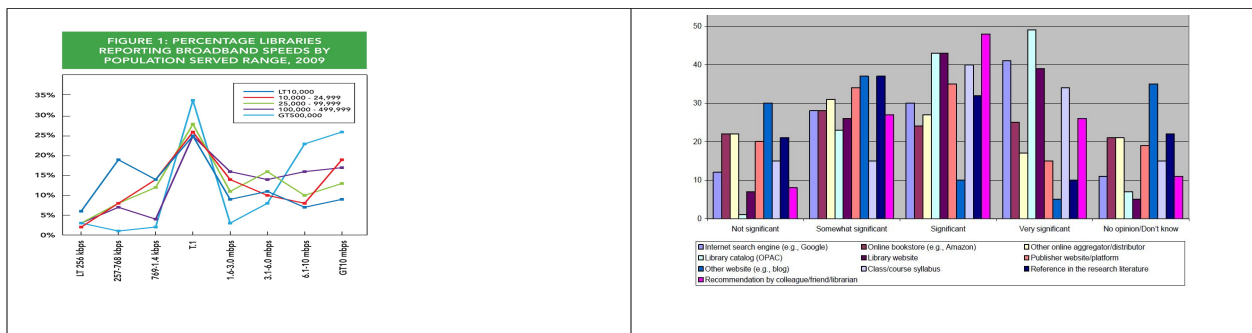
DECIDING GRAPH PROPORTIONS

As emphasized by Tufte, Wainer, Cleveland, Wong, and Few, relative sizes of a graph's vertical and horizontal scales determine how emphatic, or understated, comparisons between data on a graph are. I have not found a definitive answer to this puzzle, which makes me suspect there is no final and objective answer.   Displaying the same data replotted with different scales is, perhaps, an honest way to address this difficulty.

Also, I mentioned in the class that there is a method developed by statistician William S. Cleveland called "*banking to 45 degrees*." The idea, which applies only to line graphs, is to choose a ratio between the units of the vertical and horizontal axes that makes the angles of lines on the plot, in a sense, "average out" to 45 degrees. The approach is more complicated than this and is based on studies of human perception of the angles on lines—not on any absolute mathematical answer about what is the fair or right method. Statistical software like SPSS and SAS do provide this option for line graphs. Doing it in MS Excel is possible, but a hassle.

For more details, see William S. Cleveland, *The Elements of Graphing Data.*

GRAPHING SOFTWARE

Speaking of graphing software, I don't know a good answer to the question of what is reasonable software libraries can use to create well-designed graphs. Tableau Software offers Tableau Public along with their regular Tableau Desktop (about $1000). That package has several advanced interactive features, supports clean presentation and newer features like the 'trellis graphs' I described above. But it has some cumbersome aspects (e.g., lack of templates to standardize title and caption fonts).

The open source statistical program *R* is freely available (http://www.r-project.org/). But it apparently has a fairly steep learning curve. Like SPSS and SAS, it has a graphical interface, but still requires learning program commands and syntax. Of course SPSS and SAS have graphical capabilities, but can be inflexible.

R. Lyons   11/1/2010   raylyons@gmail.com