

Designing Effective Tables and Charts: Theory and Practice

AIR Forum Workshop, June 2012

Rebecca Carr, AAU Data Exchange, rcarr2@unl.edu
Mary Harrington, University of Mississippi, ccmmary@olemiss.edu

During this workshop we will provide an overview of how to use tables and charts as tools for effective *communication*. Our primary goal is to get you to look at charts and graphs more critically and to understand them as tools for communication. When you start to see them differently, you'll have better context as you design your own.

Graphics and design are basic to how we communicate the meaning of data. Edward Tufte – one of the preeminent scholars in the field – said “Graphics reveal data. Indeed graphics can be more precise and revealing than conventional statistical computations” (Tufte, 2001, p 13). But graphics can only reveal data if they are well-designed.

Please notice that we say that you should “design” tables and charts rather than “create” them. Creating tables and charts is easy -- all you need to do is have Microsoft Excel and some data -- but designing them is more complicated. To design a table or chart, you need to actively make decisions about who your audience is, what you want them to know, and how you can present your data to most effectively communicate your message.

Research has shown that of all the sensory receptors in our body, 70% reside in our eyes. So most of what we see and perceive of the world around us is processed through our eyes. That's why “I understand” and “I see” are used interchangeably.

If we want to effectively present information visually, we need to understand the strengths and weaknesses of visual perception. For example, the human visual system has tremendous capacity to recognize patterns – but only if they are presented in certain ways. And it has a tendency to misinterpret or completely miss information if it is presented in a non-intuitive way.

Two simple rules for developing a PowerPoint

1. *Use relevant, striking visual images and only a few words.*
Research has shown that repeating your main message both on the screen and while talking actually inhibits communication. Design slides that capture the audience's attention.
2. *Create a handout (which is not a copy of your slides). Put details in the handout and only present slides that your audience can read.*

Drawing from the research of Stephen Few and Edward Tufte, we have developed three overarching guidelines: (1) organize the data, (2) highlight the data, and (3) show the data.

(1) Organize the data

The two basic possibilities for organizing data are with tables and charts. Neither is better than the other, but each have their own strengths, but before you begin the design process, you need to ask yourself: *What's my point?* and *Why does it matter?*

The same data can be presented in many different ways and the “right” design will depend who is your audience and why you need to present the data. So it's incumbent upon you, as the designer, to ask yourself these two important questions before you begin to ensure the most effective presentation of the data. If you know the point you want to make and the reason your audience should believe it matters will guide your decisions.

Consider your audience and presentation medium

The way you present your data will depend both on who you are presenting to and how it will be presented. Detailed tables work perfectly well in printed documents, but are poor choices for a PowerPoint presentation. The sidebar provides two simple rules for developing PowerPoints.

Use the proper format for your purpose

A good rule of thumb is this: If you want to present your audience with details, then present data in a *table*. But if you want to present patterns or relationships, then you should create a *chart*.

Tables

interact primarily with our verbal system, as we do with written language.

This means that we process information in a sequential fashion, either reading down columns or across rows of numbers, and comparing a handful of numbers at a time (Few, 2004).

The primary benefit of tables is that they make it easy to look up exact values (Few, 2004), so it is best to use tables:

- To look up individual values
- To compare individual values
- When precise values are required

Charts

are perceived primarily with our visual system, which involves the mechanics of sight and principles of visual perception.

It is our visual system that allows us to see and understand patterns and relationships (Few, 2004).

Unlike tables, graphs allow you to illustrate patterns and relations among measurements. The times you want to do this include:

- When the message is in the pattern
- To reveal relationships among multiple values

What type of data are you summarizing?

The *type* of data will have implications on how it would be best presented. In a nutshell, data are either quantitative or categorical. Quantitative data measure things; categorical data group them (Few, 2004).

The basic types of categorical data that have implications for chart selection are as follows. (This section was adapted from Few, 2004, pp 17-19.)

- *Nominal* categories are discrete and have no intrinsic order. Two examples are gender and race.
- *Ordinal* categories have a prescribed order such as levels of satisfaction from Very Satisfied through Neutral to Very Dissatisfied; Small, Medium, and Large; or tenure status from tenured, tenure-track, to non-tenure-track. To display them in any other order would not make sense, but they are not truly numeric.
- *Interval* data consist of a series of sequential numerical ranges that have a distinct order and can be divided into equal portions.
 - Time categories are the prime example. Twelve months will always add up to one year in ways that (e.g.) three mediums do not add up to a large.
 - In that way, interval measures are meaningful in relation to each other and can be arranged in ascending or descending order.
 - Other examples of interval data in IR would include age, salary and other financial measures.

REVIEW: Before doing any design, you should ask the questions “*What’s my point?*” and “*Why does it matter?*”. These answers, including whether you want to provide exact numbers or patterns and relationships, will guide all your decisions. After identifying the purpose, your presentation medium, and the type of data you have, you’re ready to begin the design process.

Want to provide precise values → **Table**

Sparse data → **Table**

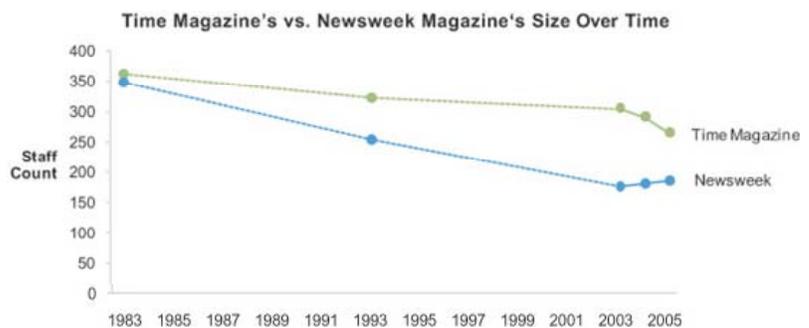
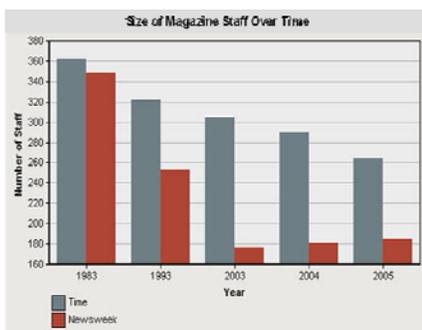
Want to show pattern /relationships of data → **Chart**

Dense data → **Chart**

What type of chart would be best for my data?

Most data can be presented in any chart format, but there are best practices about what you should use. The *purpose* of your presentation and the *type of data* will at least inform if not determine the type of chart to use. The following is a sampling of best practices that was adapted partly from Kosslyn (1994) and Stephen Few.

- **Use a stacked-bar chart for part-to-whole relationships.**
- **Avoid pie charts** and rely on bar graphs instead. Pie charts are terrible ways to present data if you want people to understand the patterns they present. “Save the pies for dessert!”
- **Use a line graph if the x-axis will show interval data.** Remember, interval data are numerical, have a distinct order, and can be divided into equal portions.
 - Intervals should always be equal in size (e.g., time, salary distribution, etc). In this case there are data for every other years from start to finish.
 - Lines should only directly connect values in adjacent intervals. (Leave a gap if have missing data.)
 - The first of the following charts is a **bad** example of how to present interval data with missing data; the second is a good improvement. (These were taken from <http://www.perceptualedge.com>)



- **Use a bar graph to show groups comparison.** Typically this will be nominal or ordinal data.
- **Use a scatterplot to show correlations.**
- **Use stacked bars to present cumulative totals.**

(2) Highlight the data



In designing charts, we need to keep in mind that our brains intuitively recognize differences in color, size, shape, hue, and orientation and attach specific meanings to them. These visual properties are called “pre-attentive variables” because the process of perceiving them is automatic and immediate and does not involve conscious thought. It is up to us as designers to use this knowledge to intentionally highlight the most critical information. One of the best ways to do this is with color.

Make Good Color Decisions

The color wheel is a simple way to introduce concepts and terms related to the use of color. We all remember from grade school that the primary colors are Red, Blue and Yellow, which are organized in a triangle on the wheel. Nestled between the primary colors are the secondary colors of Purple, Green, and Orange.

Complementary colors



Colors¹ that appear on opposite sides of the color wheel are *complementary colors*. The primary combinations are Red/Green, Orange/Blue, and Purple/Yellow. When placed next to each other, complementary colors appear brighter and their intensity is heightened. Color schemes based on complementary color *can* result in “vibrations” that are distracting and difficult to look at. As you can see here, the middle set is more difficult to read than either of the other two. This complementary combination does not work.

Not complementary colors

Complementary colors

Not complementary colors

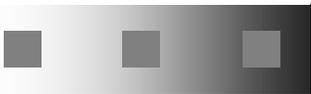
Analogous colors



Analogous colors are next to each other on the color wheel. Generally speaking analogous colors tend to create harmony or unity, but you may need more contrast.

Color has three basic properties. *Hue* is the name of the color. *Value* refers to how light or dark the hue is and *intensity* refers to how bright or dull the hue is. These qualities are independent of each other. Any hue can be light, midtone, or dark; any hue can be intense or neutral (Dunn, 1995).

Color is contextual



One important thing to remember is that color is contextual and we understand it in relation to its environment (Bowers, 1999). In other words, an individual color will change as the colors surrounding it change. These grey boxes are all the same color, but when they are surrounded by different colors they seem different. Specifically, the one on the left looks a lot darker than the one on the right. So a color may be right in one context but not in another.

Temperature of colors

All colors have a “temperature”, which refers to its relative warmth or coolness. The temperature of colors can add an emotional message to the graphic and affects how they appear on the page.

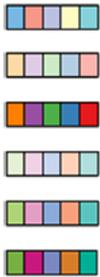
- *Cool colors* like purple, blue and green tend to have a calming effect. They appear smaller than warm colors and visually recede on the page.
- *Warm colors* energize rather than calm and appear more prominent than cool counterparts. So, a red can visually overpower a blue, even if used in equal amounts.
- *Neutral colors* of black, white, silver, gray, and brown put the focus on other colors or tone down colors that might otherwise be overpowering on their own.

¹ Some of these images were taken from <http://www.worqx.com> (which is a terrific resource on color) and used through their Creative Commons License.

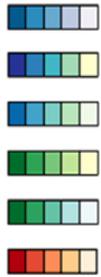
Color palettes have meaning

One resource for selecting colors for your palette is <http://www.colorbrewer.org>. The site -- created by Cynthia Brewer, a faculty member at Penn State -- provides a variety of tested palettes along with their CMYK and RGB values so you can recreate the values. She describes three types of palettes.

Qualitative



Sequential



Diverging



Qualitative palettes include colors that are distinctly different from each other. They are most successfully used to represent categorical data.

Sequential palettes will typically be a single hue that ranges either from neutral to bright or -- as in this case -- from light to dark. They are used successfully with ordinal or interval data.

Diverging palettes are basically two sequential palettes that meet in the middle with a neutral color. They allow you to have two anchors, such as agree and disagree, and easily distinguish by color. They are used successfully with ordinal or interval data.

Principles in practice

Now I'd like to show examples of these principles. Rather than using examples from IR, we'll look at some paintings from well-known artists. My intention in doing this is to focus your attention on how color is used rather than on the data.



In this first painting -- by Wilhelm Hammershoi -- I want to point out the contrast in value and complementary colors. The contrast in value between the white of the chair and the black of her dress clearly defines the shape of the chair. The black dress and the brown piano are very close in value; thus there is lower contrast and less ability to distinguish between the two shapes. The same is true of the white tablecloth and bowls.

Now look at the use of complementary colors. The gold of the picture frames and the white wall trim are roughly the same shape, but the gold appears richer and stands out more from the blue wall. This is because blue and gold are complementary colors. The yellow of the butter on the table is a brighter color by itself, but because it's surrounded by white, it does not stand out nearly as much.



There is a lot to point out about the color and color interactions in this painting by Kandinsky. In the third square on the bottom row, the colors used are the warm and analogous colors of yellow, orange and red. That square feels quite unified and doesn't have a large amount of contrast. Now, compare those colors to those in the first square of the middle row. That square also uses yellow and red, but exchanges the orange for purple. The yellow and purple are complementary colors and provide more drama, more contrast, and a different level of interest. The same is true of the third square in the middle row. It too uses red, orange, and yellow but also includes a blue-turquoise, which is the complement of the orange. This combination of colors results in a different feel than the square of colors just below it.



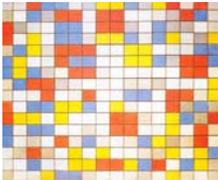
These two paintings -- both by Matisse -- demonstrate how temperature of colors can influence your reaction to the paintings. This first one uses primarily cool, analogous colors. It is calm and relaxing.

In contrast, the second painting almost throbs with excitement and energy. Color-wise it is dominated by the warm colors of red-orange and yellow. The use of the complementary color of blue-violet in the wallpaper and tablecloth adds to

the energy of the painting. Also, in the second Matisse painting, please notice that the red-orange is very demanding of your attention. It takes a while, even a conscious effort, to notice the blues and greens outside the window in the corner. The warm colors come forward and the cool colors recede.



The next painting, by Picasso, makes heavy use of the neutral colors of brown, black, grey and white. They serve as a background or frame for the brighter blue, yellow and oranges in the painting. The neutral colors are important colors in the painting, but they do not draw your attention.



This final painting, by Mondrian, looks more like the graphs you might create during your work than any of the other paintings included in this brief slide show. It is included to point out how the warm colors of the red and yellow come forward. You definitely know the other colors are present, but the red and yellow demand your attention before the cool blue and neutral grey and white.

The color decisions you make in IR will be much simpler than those made by visual artists, but having a basic understanding of the properties of colors and how they interact will better inform your decisions.

Practical tips about highlighting data: See Few (Feb 2008) for more detail and examples.

Use color meaningfully and with restraint.

- Color is not always necessary; shades of grey will be adequate for many needs. It is also most is cheap to print and easy to duplicate.
- Too many or too intense colors may overwhelm your data. You want people's focus to be on the data not its decoration.
- Use different colors to correspond to differences of meaning in the data. Select a color palette that matches your purpose and your data.
- Don't always rely on your school colors; they may not be the best choice.

Use the three properties of color to highlight data.

- *Intensity*: Highlight with a bright color.
- *Value*: Highlight with a dark value on a white background or light color on a dark background.
- *Hue*: Highlight with warm hues.
- Be aware of the context of your colors.
 - Complimentary color schemes provide the most contrast, but can be overpowering.
 - Use a background color with sufficient contrast so objects in a table or graph can be seen.
 - If you want different objects of the same color in a table or graph to look the same, make sure the background is consistent.

Additional tips.

- Utilize the color palettes available at <http://www.colorbrewer.org>.
- To guarantee that most people who are colorblind can distinguish groups of data, avoid using a combination of red and green.
- Use colors that are already associated with the data content. For example, use black/red for financial measures; and red/blue for the political parties or ideology.

(3) Show the data

Don't decorate your data

Research clearly shows that features like pictures and 3-D rendering are visually distracting and hinder the audience from understanding what is important and what is irrelevant. As designers, we need to display only meaningful content and abandon extraneous items.

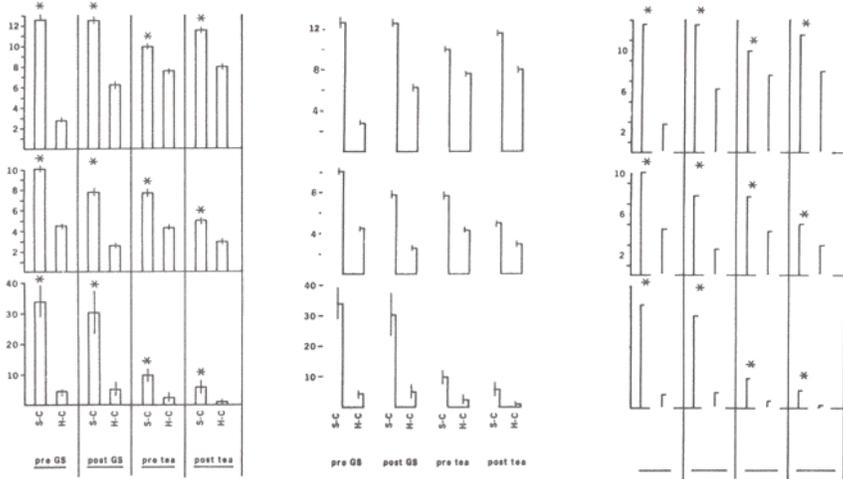
“Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space”. –Edward Tufte

One of Tufte’s rules is to maximize the “data-ink ratio”, which he describes with this formula.

$$\text{Data-ink ratio} = \frac{\text{“Data-ink”}}{\text{Total ink used to print the graphic}}$$

In other words, you want to look at the total amount of ink that is used in the display and minimize the amount that is not necessary to provide or describe the data. Increasing the data-ink ratio places more emphasis on the data than on the features of design. Tufte argues that – within reason – each pixel of a graphic should add meaning or context.

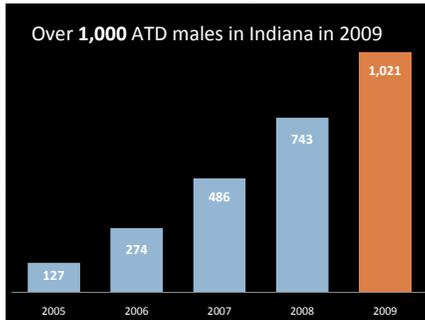
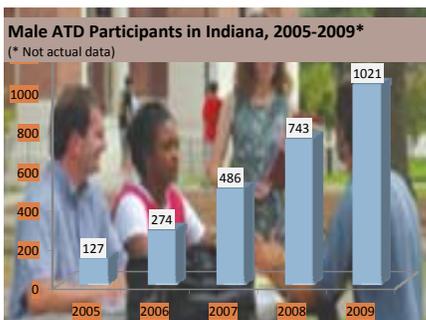
These three images come from Tufte (2001, pp 100-102) and show the results of a pre-test/post-test designed experiment. The first of the images is the original produced by the study’s authors.



The second shows Tufte’s revision of the graph. As you can see, he has removed the axes, reduced the bars to lines, and made the baseline implicit only. Yet by removing all this non-data-ink (the third image), the data is easier to understand.

Many charts are graphs are decorated in ways that complicate the display with unnecessary and distracting graphics. Ultimately, decoration is a product of poor design that deemphasizes the data being communicated.

The following are two versions of the same data. In the one on the top, your eyes are inevitably drawn to the background picture rather than the data. The picture



decorates, but does not enhance the chart and should be eliminated.

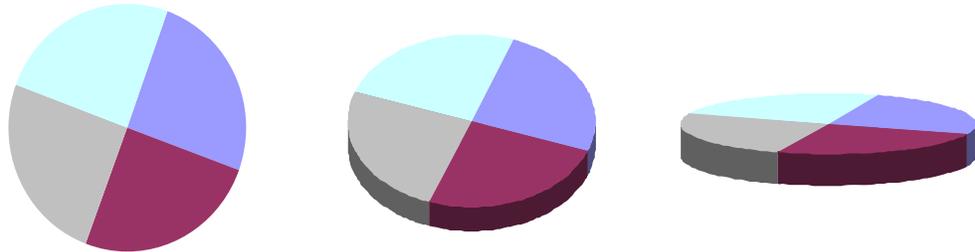
When we eliminate all decoration, as shown in the graph on the bottom, we focus entirely on the data and actually show the data. The original version might look pretty, but it *decorates* rather than *communicates*.

More practically, here are some specific examples of features to avoid include:

- **3-dimensional graphs:** Nearly all data needs in IR can be presented in 2-D; the use of 3-D graphs adds only chart-junk.
- **Pattern or gradient-fills:** At times, you'll need to make distinctions between different data points, you should do so with color (or shades of grey) rather than patterns or gradients.
- **Excess decoration,** whether in background or otherwise.

Because so many people prepare 3-D charts, we would like to spend a little time specifically telling you why *you* shouldn't.

This is a very basic pie chart with all pieces equal to 25% of the total. As we flatten it into a 3-D chart, it becomes increasingly more difficult to tell that the four pieces are the same size. Thus, 3-D style charts are solely decorative and distort the data along with any message you want to communicate. There are virtually no data in IR that require three dimensions. Stay away from 3-D charts!



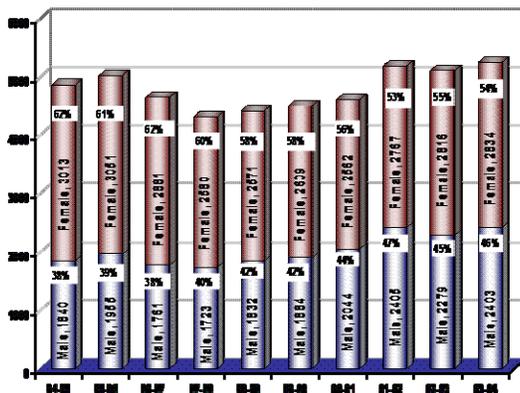
Simplify, simplify, simplify

A similar recommendation is to simplify, simplify, simplify. In a nutshell, we can summarize the *show the data* section with a quote from Einstein:

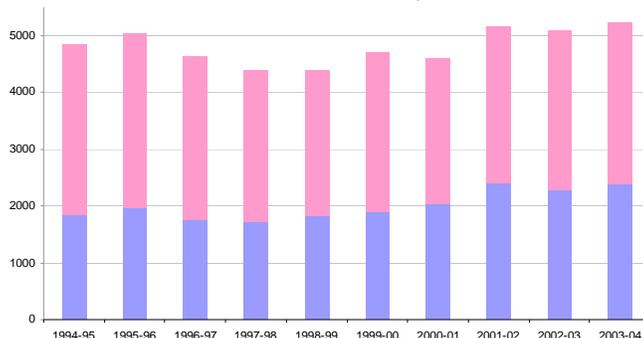
“Everything should be made as simple as possible but not simpler”.

This same thinking applies to charts and tables. Keep what you need and discard the rest. As an example, The chart on the top can be simplified and streamlined by removing the values for count and percentage and adding a simple legend or using the stereotypical colors of pink and blue in place of a legend.

**Curriculum Enrollment by Gender
1994-95 - 2003-04**



Curriculum Enrollment Trend by Gender



As you simplify your graph and maximize the data-ink ratio, some portions of a graphic that you should consider erasing or muting include:

- **Grid and axes:** A grid can be muted or absent, so its presence is only implicit; otherwise, it could compete with important data.
- Similarly, excessive **ticks** along the axis should be avoided.
- **Legends and legend borders:** Legends can also be extraneous and unnecessary to convey the meaning of a graphic. Effective chart titles and axis labeling often reduce the need for legends. When legends are necessary, their borders can be eliminated to facilitate the visual connection between the legend and the data.
- **Redundant data-ink:** Avoid redundant data-ink that depicts the same number over and over, which would include showing numerical values on top of bar graphs.
- **Lines around bars in bar graph:** Generally you can eliminate lines around bar charts since they are redundant to the shape of the bar itself.

There are some cases when you will want to include some of these “unnecessary items”. If you do so, then do so intentionally for a specific reason rather than by default.

Review: Let’s wrap-up with a summary of what you should do in preparing tables and charts:

- **Know what is your point and why it matters.**
 - If you know what message you want your audience to receive and why it should matter to them, then your design decisions will be much easier and, in the end, you’ll know whether you’ve achieved your goal.
 - Ensure your chart/table emphasizes what you want to highlight.
- **Choose a presentation style to fit your data and its purpose**
 - If you need to present precise numbers, use a table.
 - If you need to present a pattern or set of relationships, use a chart.
 - If you only have a couple of numbers to present, simply use text.
 - “Save your pies for dessert” and do NOT use pie charts.
- **Be intentional about color decisions**
 - Don’t use too much of it.
 - Be sure it helps explain or show the data.
- **Don’t decorate your data and simplify, simplify, simplify**
 - Don’t use 3-D charts!
 - Eliminate or minimize things like axes, grids, legends, and redundant data-ink.
- **When you’re done, look at it again.**
 - ...and ask yourself “what does it look like I’m trying to highlight”.
 - If you don’t have it quite right, then tweak the display and keep trying. Over time this will come much more naturally.

References and resources

Brewer, C. (2009). *Color Brewer*. Retrieved April 1, 2009: <http://www.colorbrewer.org>.

Bowers, J. (1999). *Introduction to Two-Dimensional Design: Understanding Form and Function*. New York: John Wiley & Sons, Inc.

Dunn, C. (1995). *Conversations in Paint: A Notebook of Fundamentals*. New York: Workman Publishing.

Few, S. (2004). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Oakland, CA: Analytics Press.

Few, S. (Aug 2007). *Save the Pies for Dessert*. Retrieved March 1, 2009, from Perceptual Edge Website: http://www.perceptualedge.com/articles/visual_business_intelligence/save_the_pies_for_dessert.pdf.

Few, S. (Feb 2008). *Sometimes Practical Rules for Using Color in Charts*. Retrieved March 1, 2009, from Perceptual Edge Website: http://www.perceptualedge.com/articles/visual_business_intelligence/rules_for_using_color.pdf.

Few, S. (Nov/Dec 2008). *Line Graphs and Irregular Intervals: An Incompatible Partnership*. Retrieved March 1, 2009, from Perceptual Edge Website: http://www.perceptualedge.com/articles/visual_business_intelligence/line_graphs_and_irregular_intervals.pdf.

Few, S. (Feb 2009). *Sometimes We Must Raise Our Voices*. Retrieved March 1, 2009, from Perceptual Edge Website: http://www.perceptualedge.com/articles/visual_business_intelligence/sometimes_we_must_raise_our_voices.pdf.

Ford, J. L. (2009). *Color Theory Overview*. Retrieved March 15, 2009 from <http://www.worqx.com>.

Klass, G. (2002). *Presenting data: Tabular and graphic display of social indicators*. Retrieved April 25, 2006, from Illinois State University Web site: <http://lilt.ilstu.edu/gmklass/pos138/datadisplay>.

Kosslyn, S. M. (1994). *Elements of Graph Design*. New York: W.H. Freeman and Co.

Lauer, D. A. and Pentak, S. (2002). *Design Basics, Fifth Edition*. Philadelphia: Wadsworth Publishing.

Tufte, E. (1990). *Envisioning Information*. Cheshire, CT: Graphics Press.

Five tips on fonts and alignment

1. Use serif fonts, like Garamond or Times, for printed documents.
2. Use sans serif fonts, like Arial or Calibri, for on-screen presentations.
3. Avoid specialty fonts.
4. Right align numbers and column headers.
5. Left align dates and text.

Tufte, E. (1997). *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press.

Tufte, E. (2001). *The Visual Display of Quantitative Information: Second Edition*. Cheshire, CT: Graphics Press.

Williams, R. (2004). *The Non-designer's Design Book, Second Edition*. Berkeley: Peachpit Press.