Designing Effective Tables and Charts: Theory and Practice

ly, youll have better context as you design your ow.

communicate the meaning of data. Edward holars in the field -said **G**aphics reveal data. precise and revealing than conventional statistical

) Bit graphics can only reveal data if they are ell-

u should design tables and charts rather than and charts is easy -- all you need to do is have a -- but designing them is more complicated. To design ctively ma ke decisions about **h**o your audience is, nowou ca n present your data to most effectively

the sensory receptors in our body, 70% eside in our perceive of the wrld around us is processed nderstand" and f see" are used interchangeably.

nformation visually, evneed to understand the visual perception. For example, the human visual city to recognize patterns -but only if they are it has a tendency to misinterpret or completely miss a non-intuitive by.

he research of Sephen Fewand Edward Tifte, whave ree overarching guidelines: 1) organize the data, 2)



Put details in the handout and only present slides that your audience can read.

(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:

- x To look up individual values
- x To compare individual values
- x 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:

- x When the message is in the pattern
- x 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.)

- x Nominal categories are discrete and have no intrinsic order. Two examples are gender and race.
- x 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.
- x Interval data consist of a series of sequential numerical ranges that have a distinct order and can be divided into equal portions.
 - o 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.
 - o In that way, interval measures are meaningful in relation to each other and can be arranged in ascending or descending order.
 - o Other examples of interval data in IR would include age, salary and other financial measures.

REVIEW Beforedoing 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 guideall 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 î Table		Sparsedata Î Table
Want to showpattern /relationshipsof data Î	Chart	Densedata Î 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.

- x Use a stacked-bar chart for part-to-whole relationships
- x 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!"
- x 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.
 - o 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.
 - o Lines should only directly connect valu es in adjacent intervals. (Leave a gap if have missi ng data.)
 - o 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)

х	Use a bar graph to show groups comparison	. Typically this will be nominal
	or ordinal data.	

- x Use a scatterplot to show correlations
- x Use stacked bars to present cumulative totals

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 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

(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.

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

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 pretest/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-dataink (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 <u>show</u> 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

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

- x Grid and axes: A grid can be muted or absent, so its presence is only implicit; otherwise, it could compete with important data.
- x Similarly, excessive **ticks** along the axis should be avoided.
- x Legends and legend borders: Legends can also be extraneous and unnecessary to convey the meaning of a graphic. Effective chart titles and axis



References and resources

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

Bowers, J. (1999). Introduction to Two-Dimensional