

# Tufte Design Principles

## Sources:

Edward R. Tufte, The Visual Display of Quantitative Information, Graphics Press, Cheshire, CT 1983.

Yoko Mizuno, Josh Sheldon, and Mun-See Tham, Tufte Design Principle Project, January 26, 1999, <http://ldt.stanford.edu/ldt1999/Students/mizuno/Portfolio/Work/reports/tufte/ed229c-tufte-outline.html>

## Principles of Graphical Excellence

Graphical excellence is the well-designed presentation of interesting data - a matter of **substance**, of **statistics**, and of **design**.

It consists of complex ideas communicated with clarity, precision, and efficiency.

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.

It is nearly always multivariate.

And it requires telling the truth about the data.

## Principles of Graphical Integrity

The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.

$$\text{Lie Factor} = (\text{size of effect shown in graphic}) / (\text{size of effect in data})$$

Clear, detailed, and thorough labeling should be used.

Show data variation, not design variation.

Graphics must not quote data out of context.

The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.

## Data-Ink

This is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented.

$$\text{Data-ink ratio} = (\text{data ink}) / (\text{total ink used to print the graphic})$$

## Principles of data graphics

Above all else show the data.

Maximize the data-ink ratio.

Erase non-data-ink.

Erase redundant data-ink.

Revise and edit.

## Chartjunk

This refers to over-busy grid lines and excess tricks, redundant representations of the simplest data, the debris of computer plotting, and many of the devices generating design variation.

Forgo chartjunk, including moire vibration, the grid, and the duck.

## Data-Ink Maximization

When the principles of maximizing data-ink and erasing are applied to some graphical designs, e.g., the box plot and the scatter plot, new designs emerged.

For the scatter plot, the frame lines can be edited to display the maximum, minimum, two quartiles, and the median for both variables.

## Multifunctioning Graphical Elements

A graphical element may carry data information and also perform a design function usually left to non-data-ink. Or it might show several different pieces of data. It is then called a multifunctioning graphical element.

Principle is: Mobilize every graphical element, perhaps several times over, to show data.

The graphical element that actually locates or plots the data is the data measure.

Building data measures out of data increases the quantitative detail and dimensionality of a graphic, e.g., stem-and-leaf plot.

Varying shades of gray show varying quantities better than color, they have a natural visual hierarchy.

Different visual angles for different aspects of the data also organize graphical information. Each separate line of sight should remain unchanging (preferably horizontal or vertical) as the eye watches for data variation off the flat of the line of sight. For multivariate work, several clear lines can be created.

## Data Density

Maximize data density and the size of the data matrix, within reason.

*Data density = (number of entries in data matrix) / (area of data graphic)*

As the volume of data increases, data measures must shrink.

Graphics can be shrunk way down.

## Small Multiples

Well-designed small multiples are:

- comparative

- multivariate

- shrunk, high-density graphics

- usually based on large data matrix

- drawn almost entirely with data ink

- efficient in interpretation

- often narrative in content, showing shifts in the relationship between variables as the index variable changes.

## Aesthetics and Technique in Design

Graphical elegance is often found in simplicity of design and complexity of data.

Tables are the best way to show numerical values, although the entries can also be arranged in semi-graphical form.

Tables also work well when the data presentation requires many localized comparisons.

## The Friendly Data Graphic

- words are spelt out, elaborate encoding avoided

- words run from left to right

- little messages help explain data

- labels are placed on the graphic itself; no legend is required

- graphics attract viewer, provoke curiosity

- colors, if used, are chosen so that the color-deficient and color-blind can make sense of the graphic

type is clear, precise, modest

type is upper-and-lower case, with serifs

## The Golden Rectangle

If the nature of the data suggests the shape of the graphic, follow that suggestion. Otherwise, move toward horizontal graphics about 50 percent wider than tall.

# The LATCH Principle

Source:

Richard Saul Wurman, Information Anxiety. Doubleday Books, New York, 1989.

**Location** - Location is chosen when the information who you are comparing comes from several different sources or locales. Doctors use different locations of the body to group and study medicine. Concerning an industry you might want to know where on the world goods are distributed.

**Alphabet** - Alphabet is best used when you have enormous amount of data. For example words in a dictionary or names in a telephone. As usually everybody is familiar with the Alphabet, categorizing by Alphabet is recommendable when not all the audience is familiar with different kind of groupings or categories you could use instead.

**Time** - Time is the best form of categorization for events that happen over fixed durations. Meeting schedules or our calendar are examples. The work of important persons might be displayed as timeline as well. Time is an easily framework in which changes can be observed and comparisons made.

**Category** - Category is an organization type often used for goods and industries. Shops and services in the yellow pages are easy to find by category. Retail stores are divided into e.g. men- and woman-clothing. This mode works well to organizing items of similar importance.

**Hierarchy** - Hierarchy organizes by magnitude. From small to large, least expensive to most expensive, by order of importance, etc. Hierarchy is to be used if you want to assign weight or value to the ordered information.