Tufte’s Design Principles

CS 7450 - Information Visualization
October 17, 2016
John Stasko

Learning Objectives

- Understand and be able to apply Tufte's principles:
  - Graphical integrity (baselines, size coding)
  - Maximize data-ink ratio
  - Avoid chartjunk
  - Macro/micro-readings
  - Small multiples
  - Minimize/unite grids, labeling, legends
  - Appropriate applications of color
Today's Agenda

Graphical Excellence

- Principles
  - Graphical excellence is the well-designed presentation of interesting data—a matter of *substance*, of *statistics*, and of *design*.
  - Graphical excellence consists of complex ideas communicated with clarity, precision and efficiency.

According to Tufte
Graphical Excellence

- Principles
  - 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
  - Graphical excellence is nearly always multivariate
  - And graphical excellence requires telling the truth about the data

Leveraging Human Capabilities

- Data graphics should complement what humans do well

“We thrive in information-thick worlds because of our marvelous and everyday capacities to select, edit, single out, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, list, abstract, scan, look over, sort, integrate, blend, inspect, filter, lump, skip, smooth, chunk, average, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flop through, browse, glance into, leaf through, skim, refine, enumerate, glean, synopsise, winnow the wheat from the chaff, and separate the sheep from the goats.” Vol.2, page 50
Summary

- 1. Tell the truth
  - Graphical integrity

- 2. Do it effectively with clarity, precision...
  - Design aesthetics

Let’s look at each of these

1. Graphical Integrity

- Your graphic should tell the truth about your data
Stock market crash?

Show entire scale
Example

Chart Integrity

- Where’s baseline?
- What’s scale?
- What’s context?
Huge Difference?

Compare area of right bar to that of the left bar.
Huge Difference?

Compare area of right bar to that of the left bar

Baseline?

http://themendozaline.org/post/118146423986/zero-tolerance-why-all-bar-charts-should-have-a
### Suggested Redo

**How Old Are The People Running For President In 2016?**
The chart shows how old each candidate would be at his or her inauguration. The age of each president spans from the year of his inauguration to the last year of his presidency.

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Age Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theodore Roosevelt</td>
<td>62-70</td>
</tr>
<tr>
<td>Bobby Jindal</td>
<td>45-45</td>
</tr>
<tr>
<td>Marco Rubio</td>
<td>45</td>
</tr>
<tr>
<td>Ted Cruz</td>
<td>46</td>
</tr>
<tr>
<td>Barack Obama</td>
<td>47-53</td>
</tr>
<tr>
<td>Scott Walker</td>
<td>49</td>
</tr>
<tr>
<td>Martin O’Malley</td>
<td>54-54</td>
</tr>
<tr>
<td>Rand Paul</td>
<td>54</td>
</tr>
<tr>
<td>Chris Christie</td>
<td>58</td>
</tr>
<tr>
<td>Rick Santorum</td>
<td>58</td>
</tr>
<tr>
<td>Bob Ehrlich</td>
<td>61</td>
</tr>
<tr>
<td>Mike Huckabee</td>
<td>61</td>
</tr>
<tr>
<td>Lindsey Graham</td>
<td>62</td>
</tr>
<tr>
<td>Carly Fiorina</td>
<td>62</td>
</tr>
<tr>
<td>Lincoln Chafee</td>
<td>63</td>
</tr>
<tr>
<td>Jeb Bush</td>
<td>63</td>
</tr>
<tr>
<td>John Kasich</td>
<td>64</td>
</tr>
<tr>
<td>Ben Carson</td>
<td>66</td>
</tr>
<tr>
<td>Rick Perry</td>
<td>67</td>
</tr>
<tr>
<td>Jim Gilmore</td>
<td>67</td>
</tr>
<tr>
<td>Elizabeth Warren</td>
<td>68</td>
</tr>
<tr>
<td>John Bolton</td>
<td>69</td>
</tr>
<tr>
<td>Hillary Clinton</td>
<td>70</td>
</tr>
<tr>
<td>Ronald Reagan</td>
<td>70</td>
</tr>
<tr>
<td>Jim Webb</td>
<td>71</td>
</tr>
<tr>
<td>George Pataki</td>
<td>72</td>
</tr>
<tr>
<td>Peter King</td>
<td>72</td>
</tr>
<tr>
<td>Joe Biden</td>
<td>76</td>
</tr>
<tr>
<td>Bernie Sanders</td>
<td>76</td>
</tr>
</tbody>
</table>

Suggests using dot plots when there is a big range down to zero.

### Vol 1, p 54 (2)

What’s being compared?
Vol 1, 57

Scale?

Vol 1, p. 61

Scale?
Great work!

Ahhh

Show the context
Local Example

A huge rise?

Atlanta Journal Constitution
Summer ‘08

More of the data

Atlanta Journal Constitution
Dec. ‘08
Baselines & heights?

A Redesign

Watch Size Coding

- Height/width vs. area vs. volume

Vol 1, p. 69

area = value?
volume = value?

Circle Width vs. Area

https://www.huffingtonpost.com/sandy-krum/false-visualizations-when_b_5736106.html
Areas? Not Sure

Measuring Misrepresentation

- Visual attribute value should be directly proportional to data attribute value

\[
\text{Lie factor} = \frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}
\]

p.62 9.4 = \frac{4280}{454} \text{ oil barrels}
2. Design Aesthetics

• Set of principles to help guide designers

Design Principles

• Maximize data-ink ratio

\[
\text{Data ink ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}
\]

= proportion of graphic’s ink devoted to the non-redundant display of data-information
Vol 1, p. 94

Good

Bad

Fall 2016
CS 7450
35

Vol 1, p. 30

Outstanding

Fall 2016
CS 7450
36
More...

• Above all else, show the data
• Maximize the data-ink ratio
• Erase non-data-ink
• Erase redundant data-ink
• Revise and edit

More...

• Maximize data density

\[
data \text{ density of graphic} = \frac{\text{number of entries in data matrix}}{\text{area of data graphic}}\]

Quote ...
Maximize Data Density

“Data-rich designs give a context and credibility to statistical evidence. Low-information designs are suspect: what is left out, what is hidden, why are we shown so little? High-density graphics help us to compare parts of the data by displaying much information within the view of the eye: we look at one page at a time and the more on the page, the more effective and comparative our eye can be. The principle, then, is:

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

Vol 1, p 168

Redesign charts

• Bar chart, scatter plot, box plot
Bar chart
Box plot

Fall 2016   CS 7450   47

Box plot

Fall 2016   CS 7450   48
Box plot

Scatter plot
Scatter plot

Scatter plot
Design Principles

- Avoid chartjunk
  - Extraneous visual elements that detract from message
Vol 1, p 108

Vol 2, p.34

A classic

Diamonds Were A Girl’s Best Friend
USA Today


Junk Charts Blog

http://junkcharts.typepad.com/
More Thoughts

Great narrative: Vol.2, bottom page 33-34

Rethink That?

Compared plain charts to “embellished” charts

Found that the embellished charts were just as good on interpretation accuracy and were recalled better weeks later

Participants also preferred the embellished ones

Some caveats:
- Very simple data
- Very plain plain charts
- Each chart/data is different

My take: It’s all about purpose
Memorability

Mechanical Turk study

- Color and human recognizable objects enhance memorability
- Common graphs are less memorable than unique visualization types

But memorability is only one dimension of utility ... and perhaps not one of the most important ones

Design Principles

- Utilize multifunctioning graphical elements (macro/micro readings)
  - Graphical elements that convey data information and a design function
US Army Divisions going to France in WW I

Leonard P. Ayres
_The War with Germany_
1919
Vol 2, p. 36

Michel E. Turgot
Louis Bretz

Plan de Paris
1739

Vol 2, p. 37

Manhattan 1989
Manhattan Map Company
Viet Nam Memorial in Washington D.C.

Maya Ying Lin

58,000+ dead soldiers
Vol 2, p. 43

Names listed chronologically by death

Design Principles

• Use small multiples
  – Repeat visually similar graphical elements nearby rather than spreading far apart
23 hours of LA air pollution

Chromosomes of man, chimpanzee, gorilla & orangutan
Vol 1, p. 174

Consumer Reports

Vol 2, p. 68

NY Trains
How to draw letters

Calligraphy
More Recent Additions

Sparklines
Small, repeated graphics (frequently line graphs)

Sparklines: Intense, Simple, Word-Sized Graphics

Test most common data display is a mean followed by some number. For example, a medical patient’s current level of glucose is typically reported in a clinical record as a word and number.

Sparkline Examples
Design Principles

• Show mechanism, process, dynamics, and causality
  – Cause and effect are key
  – Make graphic exhibit causality

Space shuttle case we discussed first day

Vol 3, p. 144

Washington Post
Design Principles

• **Escape flatland**
  – Data is multivariate
  – Doesn’t necessarily mean 3D projection

Vol 2, p. 12

Guide for visitors to Ise Shrine, Japan
Timetable for Java railroad line

Vol 3, p. 90

Music history

Steve Chapple and Reebee Garofalo
Design Principles

• Utilize layering and separation
  – 1+1 = 3 or more
  – Good or bad

Vol 2, p. 54 IBM Series III Copier
Design Principles

• Utilize narratives of space and time
  – Tell a story of position and chronology through visual elements
Vol 1, p.43 & Vol 2, p 110

Life of a beetle

L. Hugh Newman

Vol 2, p. 102

Czech air schedule
Design Principles

• Content is king
  – Quality, relevance and integrity of the content is fundamental
  – What’s the analysis task? Make the visual design reflect that
  – Integrate text, chart, graphic, map into a coherent narrative
Graph and Chart Tips

• Avoid separate legends and keys -- Just have that information in the graphic
• Make grids, labeling, etc., very faint so that they recede into background

Vol 2, p. 54

New Jersey Transit

Before

After
Using Color Effectively

• “The often scant benefits derived from coloring data indicate that even putting a good color in a good place is a complex matter. Indeed, so difficult and subtle that avoiding catastrophe becomes the first principle in bringing color to information: *Above all, do no harm.*”

Proper Color Use

• To label
• To measure
• To represent or imitate reality
• To enliven or decorate
Examples

- The bad...

Vol 1, p. 153
Description

“..despite its clever and multifunctioning data measure, formed by crossing two four-colored grids, this is a puzzle graphic. Deployed here, in a feat of technological virtuosity, are 16 shades of color spread on 3,056 counties, a monument to a sophisticated computer graphics system. But it is surely a graphic experienced verbally not visually. Over and over, the viewers must run little phrases through their minds, trying to maintain the right pattern of words to make sense of the visual montage: “Now let's see, purple represents counties where there are both high levels of male cardiovascular disease mortality and 11.6 to 56.0 percent of the households have more than 1.01 persons per room...”
"Color’s multidimensionality can also enliven and inform what users must face at computer terminals, although some color applied to display screens has made what should be a straight-forward tool into something that looks like a grim parody of a video game."

Vol 2, p. 88

Vol 3, p. 77
Examples

• The good...

Vol 2, p. 91 & Vol 3, p. 76
Swiss Mountain Map

Guides for Enhancing Visual Quality

• Attractive displays of statistical info
  – have a properly chosen format and design
  – use words, numbers and drawing together
  – reflect a balance, a proportion, a sense of relevant scale
  – display an accessible complexity of detail
  – often have a narrative quality, a story to tell about the data
  – are drawn in a professional manner, with the technical details of production done with care
  – avoid content-free decoration, including chartjunk
Information Overload

What about confusing clutter? Information overload? Doesn’t data have to “boiled down” and “simplified”? These common questions miss the point, for the quantity of detail is an issue completely separate from the difficulty of reading. Clutter and confusion are failures of design, not attributes of information. Often the less complex and less subtle the line, the more ambiguous and less interesting is the reading. Stripping the detail out of data is a style based on personal preference and fashion, considerations utterly indifferent to substantive content. Vol. 2, p. 51

Minard graphic

size of army

latitude

temperature

date

direction

longitude
Graphical Displays Should

- Show the data
- Induce the viewer to think about substance rather than about methodology, graphic design the technology of graphic production, or something else
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail, from a broad overview to the fine structure
- Serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- Be closely integrated with statistical and verbal descriptions of a data set

Website & Seminar
Discussion Forum

Interesting Contrast

Nigel Holmes
http://www.nigelholmes.com
More Bad Examples


Learning Objectives

• Understand and be able to apply Tufte's principles:
  - Graphical integrity (baselines, size coding)
  - Maximize data-ink ratio
  - Avoid chartjunk
  - Macro/micro-readings
  - Small multiples
  - Minimize/unite grids, labeling, legends
  - Appropriate applications of color
**Reading**

- Read the quartz.com website just shown

**Upcoming**

- Geospatial visualization

- No class next week
  - Assignment: Watch a video
Sources Used

E. Tufte, *The Visual Display of Quantitative Information*
E. Tufte, *Envisioning Information*
E. Tufte, *Visual Explanations*