Information Visualization

Session 2: Design Principles

LMU München, Medieninformatik
Prof. Dr. Andreas Butz
WS2018/19
Design Excellence

“Well-designed presentations of interesting data are a matter of substance, of statistics, and of design.”
E. Tufte
Edward Tufte
every time you make a powerpoint


Mark Goetz
InfoVis 2: Design Principles

• Graphical Integrity
• Design Principles
• Design Elements
• Rules and Questions
Missing Scales

- Where's baseline?
- What's scale?
- What's context?

Tufte, VDQI
The Lie Factor

• \( \frac{\text{Size of effect in graphic}}{\text{size of effect in data}} \)

Tufte, VDQI
The Lie Factor

- \((\text{Size of effect in graphic})/(\text{size of effect in data})\)

\[
\frac{5.3 - 0.6}{0.6} / \frac{27.5 - 18}{18} = 14.8
\]
The Lie Factor

THE SHRINKING FAMILY DOCTOR
In California

Percentage of Doctors Devoted Solely to Family Practice

<table>
<thead>
<tr>
<th>Year</th>
<th>1964</th>
<th>1975</th>
<th>1990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>1: 2,247</td>
<td>1: 3,167</td>
<td>1: 4,232</td>
</tr>
<tr>
<td>Ratio to Population</td>
<td>6,694</td>
<td>6,212</td>
<td>1: 6,994</td>
</tr>
<tr>
<td>Ratio to Population</td>
<td>8,023 Doctors</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tufte, VDQI
The Lie Factor

- **1950**: 983 sq. ft.
- **1970**: 1,500 sq. ft.
- **1990**: 2,080 sq. ft.
- **2004**: 2,349 sq. ft.

Source: National Association of Home Builders (Housing Facts, Figures and Trends for March 2006)

http://www.theoildrum.com/node/4645
Design Distortions

• Gratuitous 3D and perspective
• Non-uniform intervals and scales

Principle: Show data variation, not design variation.
Scale Distortions

Based on slide from Stasko
Scale Distortions

Show entire scale

Based on slide from Stasko
Scale Distortions

Show in context

Based on slide from Stasko
Four times the same data

"the latest data shows total consumer credit collapsing at an accelerating rate" (businessinsider.com)

"have to compare credit to something like disposable income." (Rolfe Winkler, reuters blog)

"Per capita is better than total" (chartingtheeconomy.com)

"on-going credit contraction— it’s astonishing" (ritholtz.com – The Big Picture)

Junk Charts
Total Consumer Credit (Annual Rate Change)

Source: Federal Reserve
CHART 1: STEEPEST CONSUMER DELEVERAGING IN OVER 50 YEARS

United States: Consumer Credit Outstanding
(year-over-year percent change)

Shaded region represent periods of U.S. recession

Source: Haver Analytics, Gluskin Sheff
United States: Consumer Credit Outstanding
(year-over-year difference, US$ blns)
InfoVis 2: Design Principles

- Graphical Integrity
- Design Principles
- Design Elements
- Rules and Questions
Use Decomposition

Beer sales

Time series of data with seasonal, trend, and remainder components.
Hierarchical Display
Show Context

100 combat deaths per month

Iraq

M. Ericson, NY Times
2,000 combat deaths per month

Vietnam

Iraq
20,000 combat deaths per month

World War II

Vietnam

Iraq

M. Ericson, NY Times
Readability

Toyota Halts Sales of Eight Models After Recall

By NICK BUNKLEY
Published: January 26, 2010

Toyota Motor, still struggling to resolve a problem with accelerator pedals, said Tuesday it would temporarily stop selling and building eight models in the American market, including the popular Camry and Corolla sedans.

The unusual move has the potential to further damage Toyota, whose reputation for quality has been battered by two recalls of millions of vehicles in the last two months for a problem that the company has described as a “rare” condition in which the gas pedal can stick, and
Gas Pedal Flaw Leads Toyota To Stop Building 8 Models

Toyota Motor, still struggling to resolve a problem with accelerator pedals, said Tuesday it would temporarily stop selling and building eight models in the American market, including the popular Camry and Corolla sedans.

The unusual move has the potential to further damage Toyota, whose reputation for quality has been battered by two recalls of millions of vehicles in the last two months for a problem that the company has described as a “rare” condition in which the gas pedal can stick, and cause a vehicle to speed up unintentionally.

“This action is necessary until a remedy is finalized,” Robert S. Carter, a Toyota group vice president, said in a statement. “We’re making every effort to address this situation for our customers as quickly as possible.”

Toyota said it would immediately stop selling the Camry, Corolla and Avalon sedans, Matrix wagon, RAV4 crossover, Tundra pickup, and Highlander and Sequoia sport utility vehicles.

It will also stop building those models the week of Feb. 1. All of the vehicles are assembled in the United States or Canada, at a total of five plants.

The models affected accounted for more than a million sales in 2009, 57 percent of Toyota’s American total for the year.

The company said the move was intended to restore confidence in the automaker, and the safety of its products. One analyst said many consumers may have a different reaction.

“The problem seems to be getting larger than anyone was led to believe at first,” said Erich Merkle, an analyst at Robert W. Baird & Co., which has a “market perform” rating on Toyota.
Maximize Data-Ink Ratio

- Data-ink = the ink used to show data
- Data-ink ratio = data-ink / total ink used
Maximize Data-Ink Ratio

- Data-ink = the ink used to show data
- Data-ink ratio = data-ink / total ink used
2.2 High Data Density

Another theme of Tufte's is that good graphics have high "data density". To calculate this quantity, note that the numbers to be plotted in a graphic can be written as an array of one or more dimensions.

Definition 3 (Data Density)

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

Figure 2.12: High data density graphic: illustration of many experimental measurements of the thermal conductivity of copper. Each string of dots, connected by a thin curve, is from a single publication which is identified by a number enclosed in a circle. Each of these 200-plus articles is listed with its identification number in the bibliography of the source of this figure, C. Y. Ho, R. W. Powell, P. E. Liley, "Thermal conductivity of the elements: A comprehensive review", supplement no. 1, J. Phys. Chem. Reference Data, 3, pg. 1-244 (1974). The thick solid line is the recommended curve.

Notwithstanding this formal definition, the concept of a "high data density" graphic is perhaps best illustrated through some examples. Fig. 2.12 compares a large number of simultaneous measurements of thermal conductivity as a function of temperature. Because more than 200 different curves, each labelled, are combined into a single graph, Tufte has canonized this as a Good Example. Is it?

Well, yes, but only up to a point. First, there is no really good way to present 200 data sets. A single figure is certainly much easier to grasp than a whole of lot of figures.

### Sparklines

<table>
<thead>
<tr>
<th>Currency</th>
<th>1999.1.1</th>
<th>65 months</th>
<th>2004.4.28</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euro foreign exchange $</td>
<td>1.1608</td>
<td></td>
<td>1.1907</td>
<td>.8252</td>
<td>1.2858</td>
</tr>
<tr>
<td>Euro foreign exchange ¥</td>
<td>121.32</td>
<td></td>
<td>130.17</td>
<td>89.30</td>
<td>140.31</td>
</tr>
<tr>
<td>Euro foreign exchange £</td>
<td>0.7111</td>
<td></td>
<td>0.6665</td>
<td>.5711</td>
<td>0.7235</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Currency</th>
<th>2003.4.28</th>
<th>12 months</th>
<th>2004.4.28</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>1.1025</td>
<td></td>
<td>1.1907</td>
<td>1.0783</td>
<td>1.2858</td>
</tr>
<tr>
<td>¥</td>
<td>132.54</td>
<td></td>
<td>130.17</td>
<td>124.80</td>
<td>140.31</td>
</tr>
<tr>
<td>£</td>
<td>0.6914</td>
<td></td>
<td>0.6665</td>
<td>0.6556</td>
<td>0.7235</td>
</tr>
</tbody>
</table>

Tufte, 1990
Avoid Chartjunk

- Extraneous visual elements that distract from the message
Avoid Chartjunk

- Extraneous visual elements that distract from the message
Avoid Chartjunk

• Extraneous visual elements that distract from the message
Avoid Chartjunk

- Extraneous visual elements that distract from the message
Avoid Chartjunk

- Extraneous visual elements that distract from the message

ongoing, Tim Brey
Avoid Chartjunk

• Extraneous visual elements that distract from the message
Before

Number of bikes sold (2002-2007)

After

Over 5,000 bikes sold in 2007

G. Reynolds, Presentation Zen
Before

After

G. Reynolds, Presentation Zen
Bring in the Clowns…

World Population in 2008

PTS Blog
Bring in the Clowns...

World Population in 2008

- China: 1300 M
- India: 1150 M
- United States: 315 M
- Indonesia: 250 M
- Brazil: 200 M
- Pakistan: 170 M
- Bangladesh: 140 M
- Nigeria: 130 M
- Russia: 120 M
- Japan: 110 M
Tufte’s Design Principles

- Above all else show the data
- Maximize data-ink ratio
- Erase non-data ink
- Erase redundant data ink
- Revise and edit
Chart Junk: Discussion

http://eagereyes.org/media/2011/memory-monster.jpg
Design Pyramid

- Functionality
  - Usability
    - Aesthetics
      - affective
      - efficient
      - effective
Subjective Dimensions

- **Aesthetics**: Attractive things are perceived as more useful than unattractive ones
- **Style**: Communicates brand, process, who the designer is
- **Playfulness**: Encourages experimentation and exploration
- **Vividness**: Can make a visualization more memorable

Pat Hanrahan, Nov 2007
InfoVis 2: Design Principles

- Graphical Integrity
- Design Principles
- Design Elements
- Rules and Questions
CRAP: *Contrast*, Repetition, Alignment, Proximity

Peter’s cake metaphor ties in nicely with Gall’s Law

A complex system that works is invariably found to have evolved from a simple system that worked. The inverse proposition also appears to be true: A complex system designed from scratch never works and cannot be made to work. You have to start over, beginning with a working simple system.
Before

After

G. Reynolds, Presentation Zen
Before

![Before Diagram]

After

![After Diagram]

G. Reynolds, Presentation Zen
CRAP: Contrast, Repetition, Alignment, Proximity
CRAP: Contrast, Repetition, **Alignment**, Proximity

Before

After

S. Few, Show me the numbers
CRAP: Contrast, Repetition, Alignment, **Proximity**

*S. Few, Show me the numbers*
Small Multiples

- Bookings & Billings (Q)
- Sales Regions (N)
- Sales Channels (N)

S. Few, Show me the numbers
Small Multiples

Cloud water at 0.0125 g/kg

Minutes 20 40 60 80 100 120 140

Tufte, VDQI
Layering and Separation
Layering and Separation

<table>
<thead>
<tr>
<th>Train No.</th>
<th>3701</th>
<th>3301</th>
<th>3801</th>
<th>3542</th>
<th>3765</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>12:10</td>
<td>1:30</td>
<td>3:45</td>
<td>7:30</td>
<td>4:33</td>
</tr>
<tr>
<td>Newark, N. J.</td>
<td>1:43</td>
<td>10:30</td>
<td>5:21</td>
<td>8:50</td>
<td>11:45</td>
</tr>
<tr>
<td>North Elizabeth</td>
<td>....</td>
<td>......</td>
<td>......</td>
<td>......</td>
<td>6:45</td>
</tr>
<tr>
<td>Elizabeth</td>
<td>3:33</td>
<td>2:05</td>
<td>......</td>
<td>......</td>
<td>7:05</td>
</tr>
<tr>
<td>Peekskill</td>
<td>5:34</td>
<td>6:40</td>
<td>......</td>
<td>7:20</td>
<td>8:50</td>
</tr>
<tr>
<td>Ediison, N. J.</td>
<td>4:45</td>
<td>5:20</td>
<td>4:40</td>
<td>2:10</td>
<td>11:05</td>
</tr>
<tr>
<td>Princeton, N. J.</td>
<td>1:30</td>
<td>....</td>
<td>......</td>
<td>3:30</td>
<td>7:30</td>
</tr>
</tbody>
</table>

Figure 2.25: Two versions of a railway timetable with the "bad" version on top. Inspired by an example from pgs. 54-55 of Tufte (1990).
Layering and Separation

<table>
<thead>
<tr>
<th>New York</th>
<th>12:10</th>
<th>1:30</th>
<th>3:45</th>
<th>7:30</th>
<th>4:33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newark, N. J.</td>
<td>1:43</td>
<td>10:30</td>
<td>5:21</td>
<td>8:50</td>
<td>11:45</td>
</tr>
<tr>
<td>North Elizabeth</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>6:45</td>
<td></td>
</tr>
<tr>
<td>Elizabeth</td>
<td>3:33</td>
<td>2:05</td>
<td>...</td>
<td>7:05</td>
<td></td>
</tr>
<tr>
<td>Peekskill</td>
<td>5:34</td>
<td>6:40</td>
<td>7:20</td>
<td>8:50</td>
<td></td>
</tr>
<tr>
<td>Ediison, N. J.</td>
<td>4:45</td>
<td>5:20</td>
<td>2:10</td>
<td>11:05</td>
<td></td>
</tr>
<tr>
<td>Princeton, N. J.</td>
<td>1:30</td>
<td>...</td>
<td>3:30</td>
<td>7:30</td>
<td></td>
</tr>
<tr>
<td>Train No.</td>
<td>3701</td>
<td>3301</td>
<td>3801</td>
<td>3542</td>
<td>3765</td>
</tr>
</tbody>
</table>

Figure 2.25: Two versions of a railway timetable with the “bad” version on top. Inspired by an example from pgs. 54-55 of Tufte (1990).
Negative Space

CROSSWIND TAKEOFF
THE SLIP
CROSSWIND LANDINGS
SHORT FIELD TAKE OFF & LANDING
SOFT FIELD TAKE OFF & LANDING
FORCED LANDING
720 POWERTurns

Tufte, EJ (Vol. 2) p. 62
Negative Space Logos
InfoVis 2: Design Principles

- Graphical Integrity
- Design Principles
- Design Elements
- Rules and Questions
The Zen Aesthetic

- Simplicity
- Clarity
- Uncluttered
- Restraint
Before

G. Reynolds, Presentation Zen

After
Garr Reynolds
Tufte’s Graphical Excellence

• Interesting data
  – Complex ideas, multivariate data

• Clear, precise, concise presentation
  – Data-ink ratio

• Accurate communication
  – Lie factor
Graphical displays should...

• Show the data
• Induce the viewer to think about the substance, rather than about methodology, graphic design, [or] the technology of graphic productions...
• 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
• Serve a reasonably clear purpose
• Be closely integrated with the statistical and verbal descriptions

Edward Tufte, The Display of Quantitative Information, page 1
Stephen Few
### Few = Applied Tufte

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Points</th>
<th>Lines</th>
<th>Points &amp; Lines</th>
<th>Bars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal Comparison</td>
<td>When there is a need to narrow the quantitative scale, and in so doing, remove zero from its base</td>
<td>Avoid</td>
<td>Avoid</td>
<td>Either horizontal or vertical bars</td>
</tr>
<tr>
<td>Time Series</td>
<td>Categorical subdivisions on X axis, quantitative values on Y axis; emphasis on overall pattern</td>
<td>Categorical subdivisions on X axis, quantitative values on Y axis; mutual emphasis on overall pattern and individual values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ranking</td>
<td>When there is a need to narrow the quantitative scale, and in so doing, remove zero from its base</td>
<td>Avoid</td>
<td>Avoid</td>
<td>Horizontal bars are preferable, with values sorted in descending order.</td>
</tr>
<tr>
<td>Part-to-Whole</td>
<td>Avoid</td>
<td>Avoid</td>
<td>Avoid</td>
<td>Either horizontal or vertical bars</td>
</tr>
<tr>
<td>Deviation</td>
<td>Especially useful when combined with time series</td>
<td>Useful when combined with time series and when a slight emphasis on individual values is desired</td>
<td>Either horizontal or vertical bars, except when combined with time series, which requires vertical bars</td>
<td></td>
</tr>
<tr>
<td>Distribution</td>
<td>Avoid</td>
<td>Known as a frequency polygon; emphasis on overall pattern</td>
<td>Avoid</td>
<td>Known as a histogram; emphasis primarily on individual values</td>
</tr>
<tr>
<td>Single</td>
<td>Use to mark the median in a box plot</td>
<td>Avoid</td>
<td>Use in the form of range bars in box plots</td>
<td></td>
</tr>
<tr>
<td>Multiple</td>
<td>Known as a scatter plot</td>
<td>Avoid</td>
<td>In this case the line is a trend line, not a line that connects the points.</td>
<td>Either horizontal or vertical bars; can be structured either as a correlation bar graph or a paired bar graph</td>
</tr>
</tbody>
</table>

S. Few, Show Me the Numbers, p. 87
Analysis Questions

• Who is the intended audience?
• What information does this visualization represent?
• How many data dimensions does it encode?
• List several tasks, comparisons or evaluations it enables
• What principles of excellence best describe why it is good / bad?
• Can you suggest any improvements?
• Why do you like / dislike this visualization?
Robin Williams