



Tufte's Design Principles

BAIS 6140 – Information Visualization

L. Miguel Encarnação

Agenda

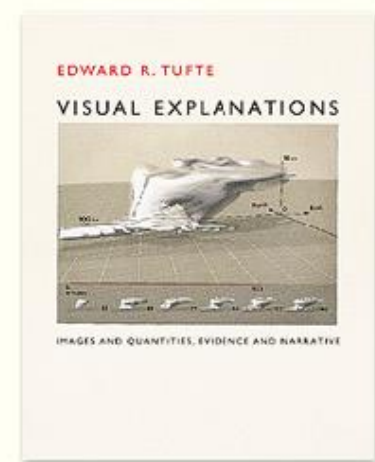
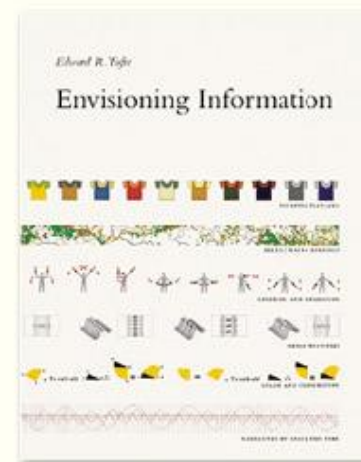
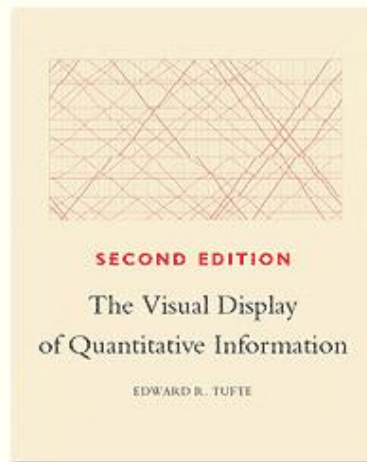
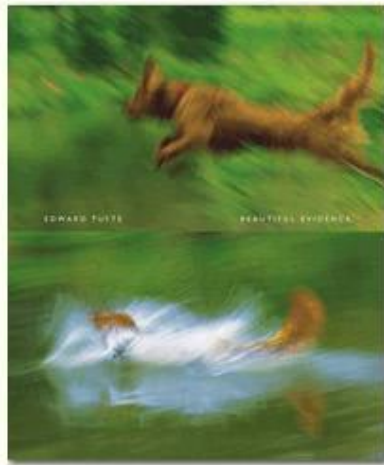


Edward Tufte has written seven books, including *Visual Explanations*, *Envisioning Information*, *The Visual Display of Quantitative Information*, and *Data Analysis for Politics and Policy*. He writes, designs, and self-publishes his books on analytical design, which have received more than 40 awards for content and design. He is Professor Emeritus at Yale University, where he taught courses in statistical evidence, information design, and interface design. His current work includes landscape sculpture, printmaking, video and a new book.

This website describes Edward Tufte's books, one day course, and artwork. For further information, call Graphics Press at 203 272-9187, or fax 203 272-8600, or [email](#).

For a moderated forum on analytical design, go to [ASK E.T.](#)

BOOKS



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
- 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.”

Tufte, Vol.2, page 50

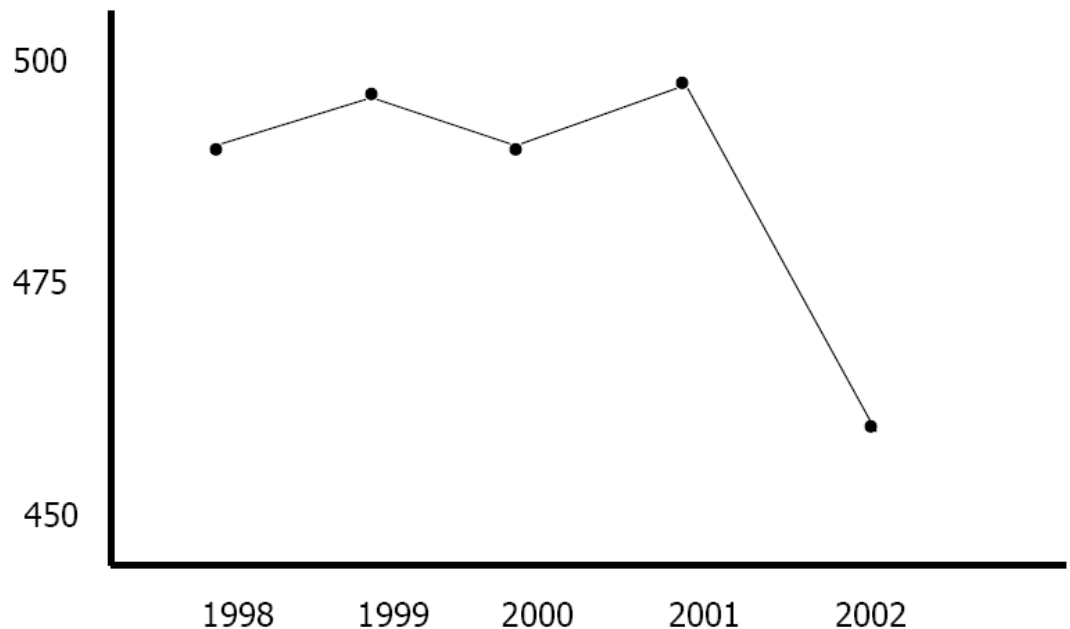


In Summary

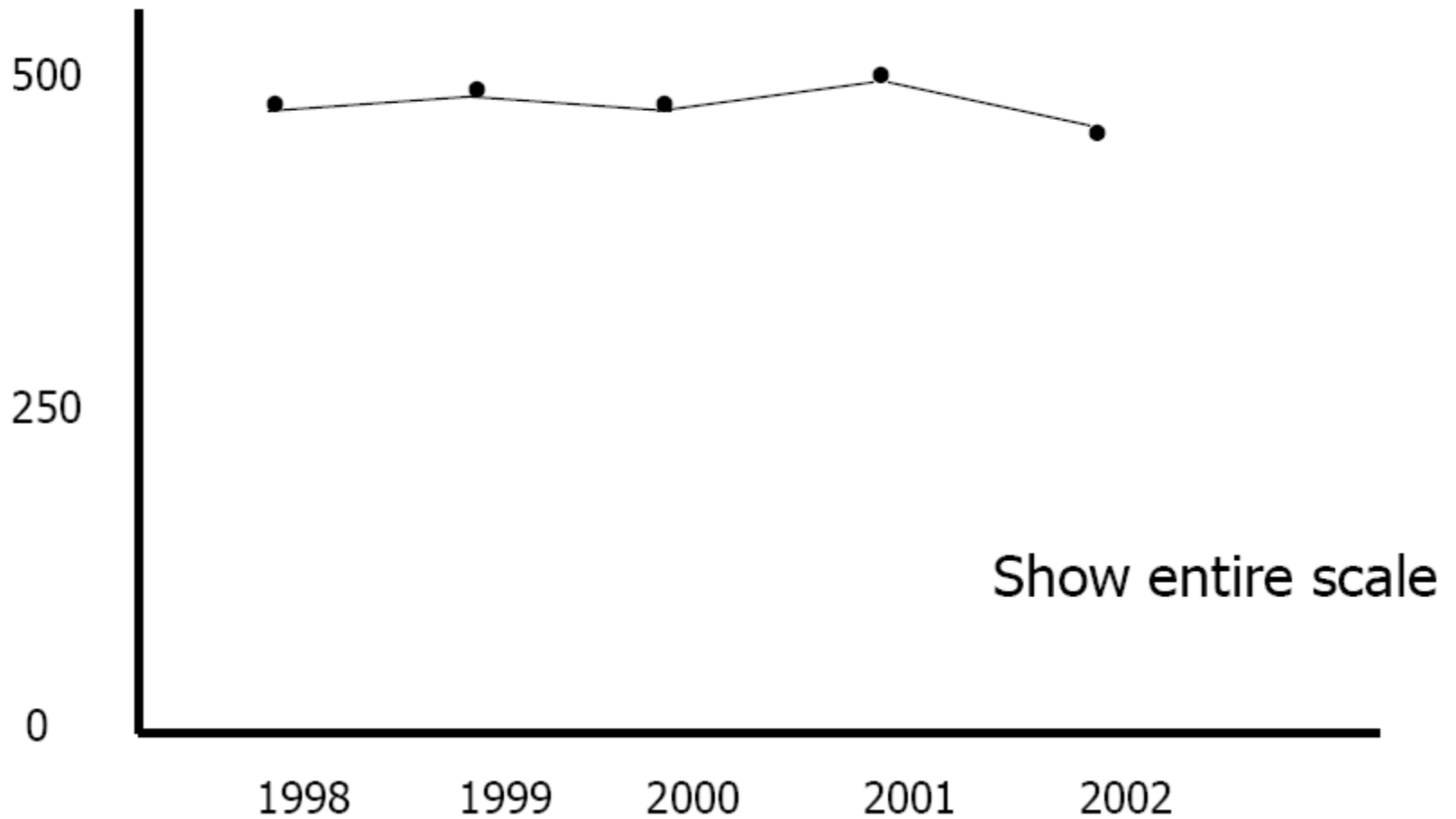
1. Tell the truth
 - Graphical integrity
2. Do it effectively with clarity, precision...
 - Design aesthetics

1. Graphical Integrity

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



Truth being ...



Truth being ...

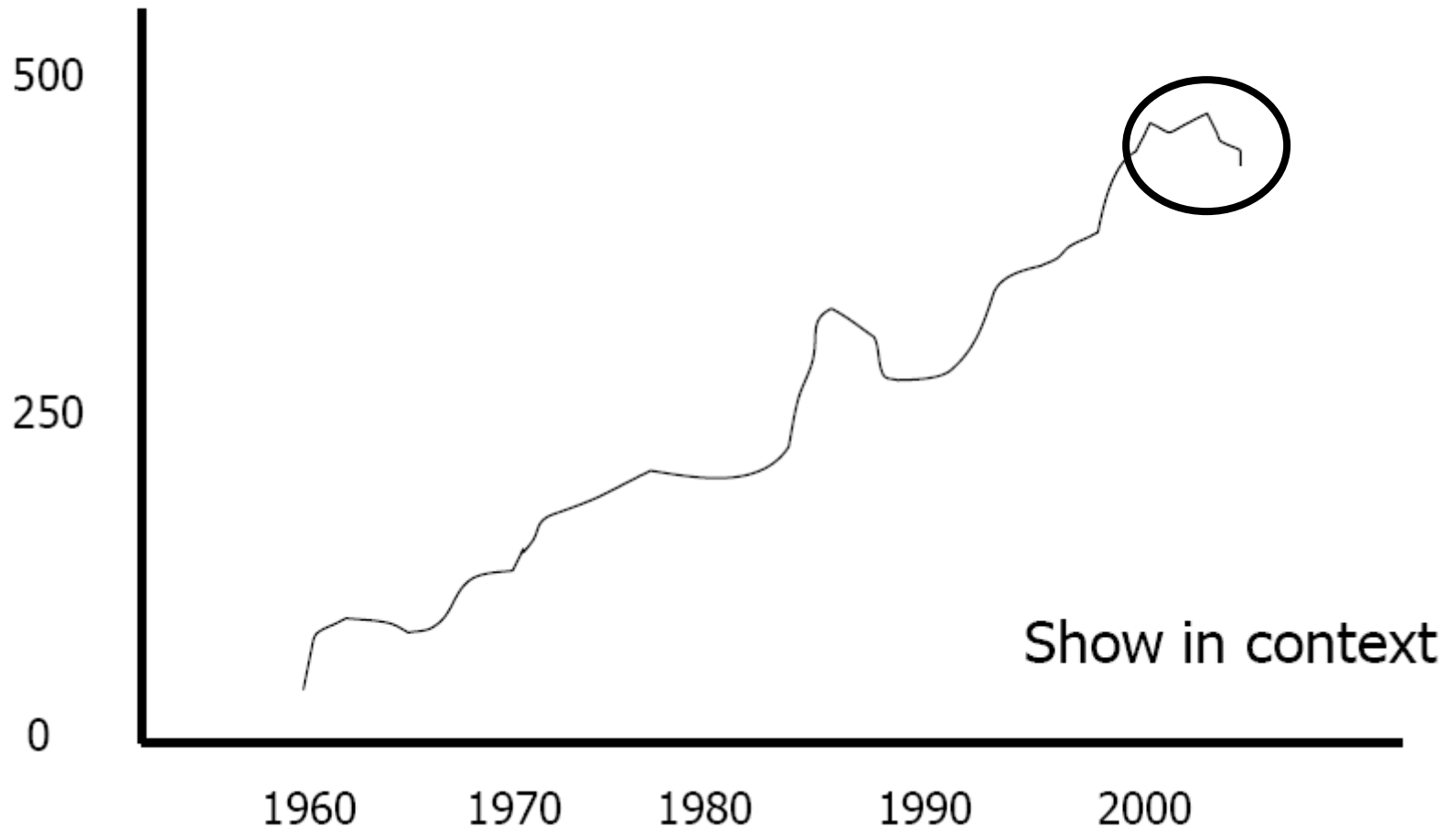
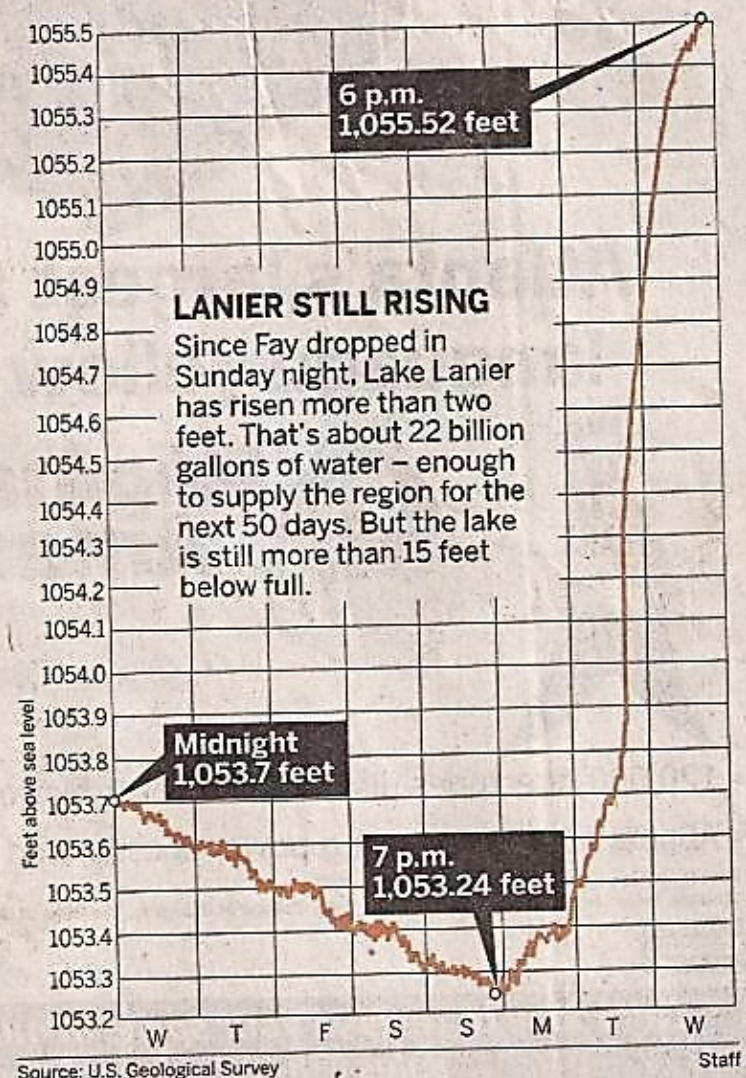


Chart Integrity

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

Example

- A huge rise?



*Atlanta Journal Constitution
Summer '08*

LANIER GETS A FAVOR

Big rains raise big lake, but it's still well shy of full, D3

Example

- More of the data

LANIER STILL ON LIFE SUPPORT

Recent rain helps Lake Lanier, but metro Atlanta's primary water source is a long way from normal.

1 March 2006: Drought begins.

2 April - June 2006: Faulty gauge leads to two-foot drop.

3 June 2006: Drought officially declared, triggering statewide watering restrictions. State officials warn the metro region's water supply is at risk.

4 September 2007: State enacts near-total ban on outdoor watering in North Georgia. Federal officials warn that Lanier is likely to hit a new record low.

5 October 2007: State predicts Lanier could "run dry" in 80 days and mandates 10 percent cut in North Georgia's water use.

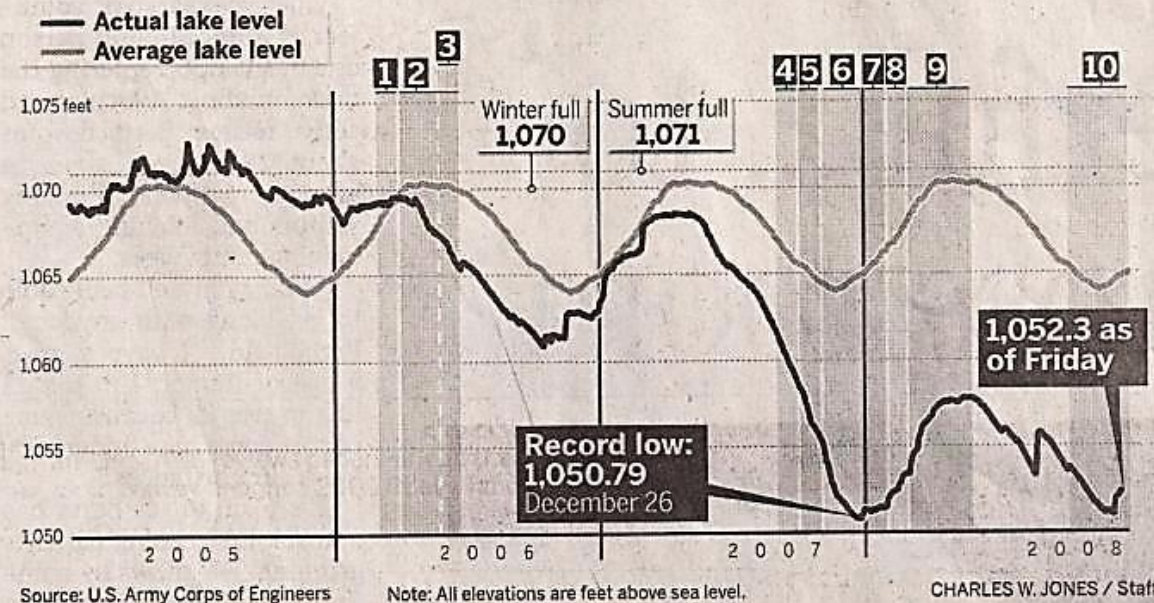
6 November - December 2007: The White House brokers a deal with Georgia, Alabama and Florida to keep more water in Lanier.

7 January 2008: General Assembly passes statewide water plan, but with no new funds.

8 February 2008: Gov. Perdue eases restrictions on outdoor water use to allow some landscape watering and filling of swimming pools.

9 March - May 2008: The U.S. Army Corps of Engineers cuts water released down the Chattahoochee River by 13 percent, holding more in Lanier.

10 November 2008 and continuing through April 2009: The corps minimizes water released from Lanier.



Atlanta Journal Constitution
Dec. '08

Watch size coding

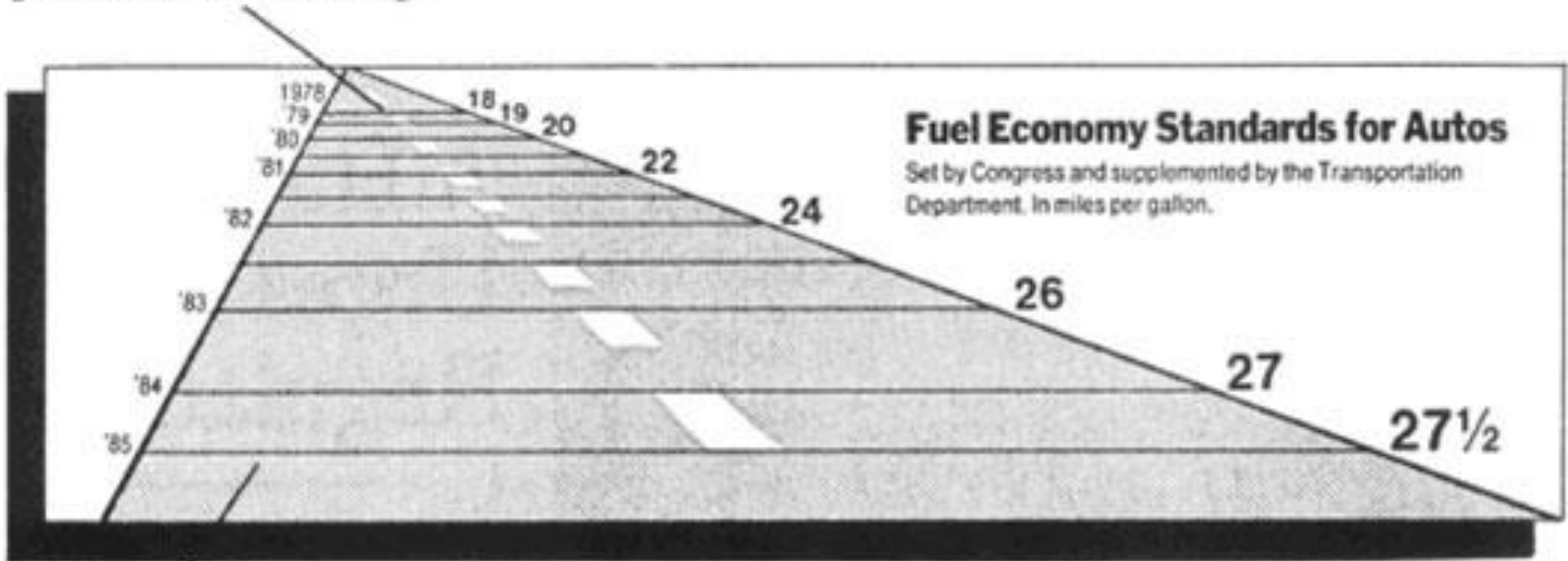
- Height/width vs. area vs. volume
- Measuring misrepresentation
 - Visual attribute value should be directly proportional to data attribute value

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

$$\text{size of effect} = \frac{|\text{second value} - \text{first value}|}{\text{first value}}$$

Lie factor: Example

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

$$\text{Lie Factor} = \frac{\frac{5.3 - 0.6}{0.6}}{\frac{27.5 - 18}{18}} = 14.8$$

Lie factor: Interpretation

- To ensure the Integrity of a graphic, its Lie Factor should have a value between 0.95 and 1.05.
- If the value is less or greater, it indicates a substantial (and often intended) distortion, far beyond minor inaccuracies (e.g. caused by plotting).
- For example, if the lie factor had the value 1.02, then the graphic would not be described as a “lying” one, since a jitter in the graphic or another minor error is more likely the cause for the given distortion.
- On the other hand, if the factor had the value 1.45, we can generally assume that the producer intentionally distorted the graphic.
- Is the Lie Factor greater than 1, the concerned graphic is so called “Overstating”. Is it less than 1, the graphic is “Understating”.

Lie factor: Example

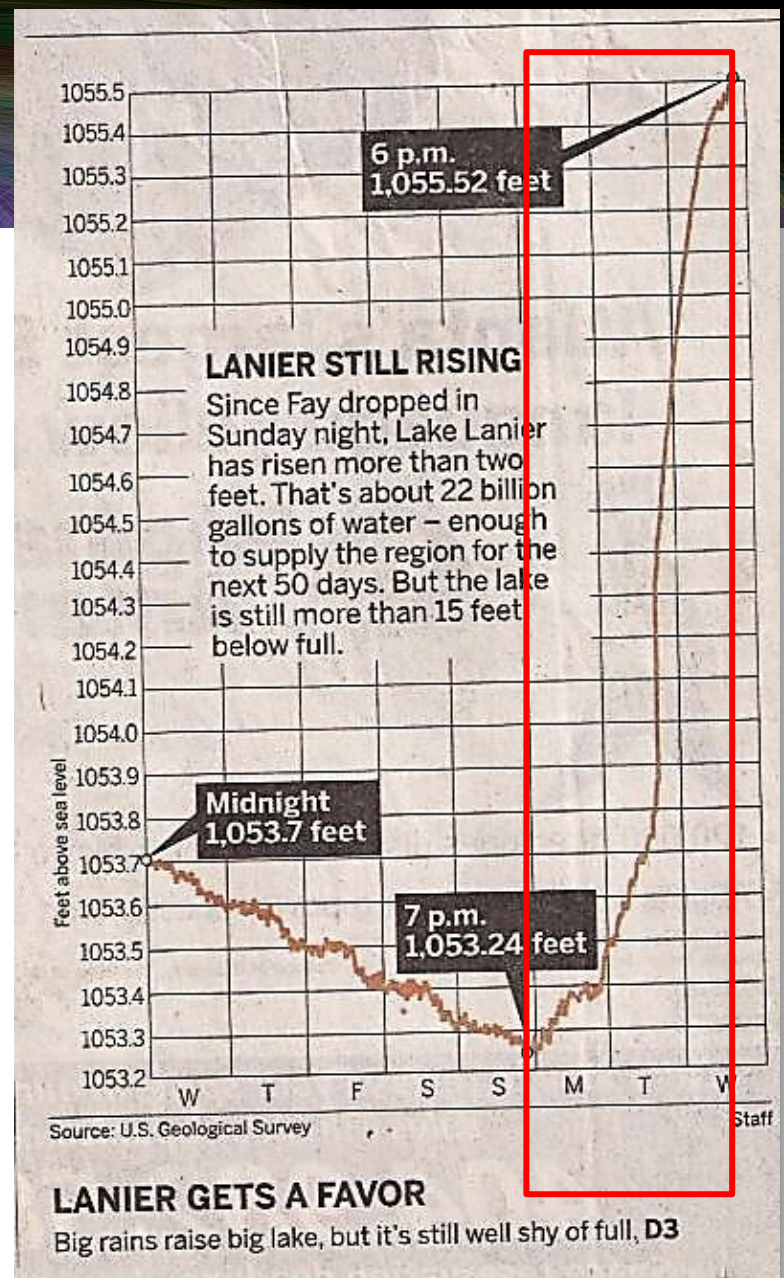
- Lie factor =

$$\frac{(23 - 0.5) / 0.5}{(1055.5 - 1053.25) / 1053.25}$$

$$= 45 / 0.00213624$$

$$= 21,065.0489$$

*Atlanta Journal Constitution
Summer '08*



Lie factor: Example

- Lie factor =

$$\frac{(1.5 - 0.1) / 0.1}{(1057 - 1051) / 1051}$$

$$= 14 / 0.0057088$$

$$= 2452.3$$

LANIER STILL ON LIFE SUPPORT

Recent rain helps Lake Lanier, but metro Atlanta's primary water source is a long way from normal.

1 March 2006: Drought begins.

2 April - June 2006: Faulty gauge leads to two-foot drop.

3 June 2006: Drought officially declared, triggering statewide watering restrictions. State officials warn the metro region's water supply is at risk.

4 September 2007: State enacts near-total ban on outdoor watering in North Georgia. Federal officials warn that Lanier is likely to hit a new record low.

5 October 2007: State predicts Lanier could "run dry" in 80 days and mandates 10 percent cut in North Georgia's water use.

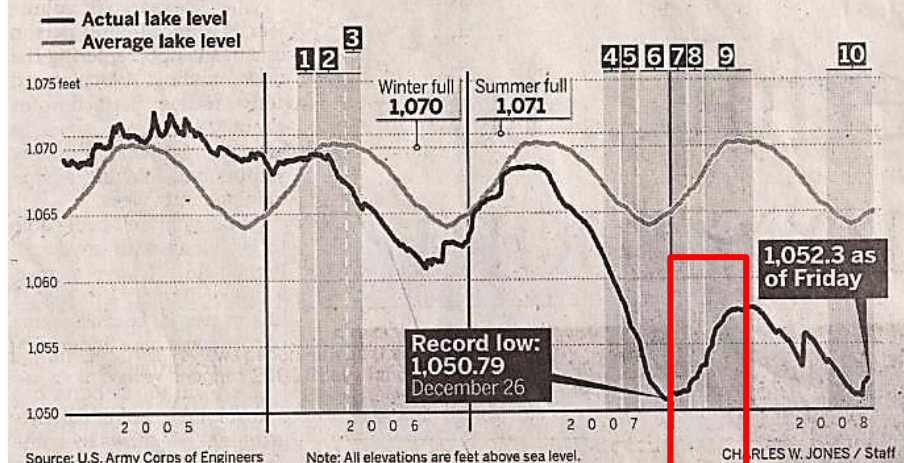
6 November - December 2007: The White House brokers a deal with Georgia, Alabama and Florida to keep more water in Lanier.

7 January 2008: General Assembly passes statewide water plan, but with no new funds.

8 February 2008: Gov. Perdue eases restrictions on outdoor water use to allow some landscape watering and filling of swimming pools.

9 March - May 2008: The U.S. Army Corps of Engineers cuts water released down the Chattahoochee River by 13 percent, holding more in Lanier.

10 November 2008 and continuing through April 2009: The corps minimizes water released from Lanier.



*Atlanta Journal Constitution
Summer '08*

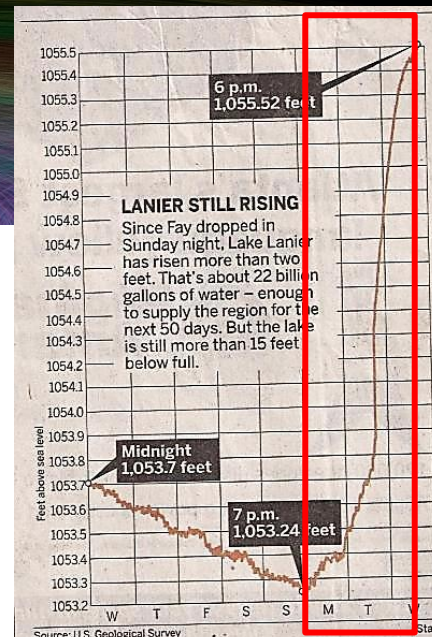
Lie factor: Example

- Lie factor difference

$$\frac{21,065.0489}{2452.3}$$

$$= 8.59$$

Atlanta Journal Constitution
Summer '08



LANIER GETS A FAVOR

Big rains raise big lake, but it's st

LANIER STILL ON LIFE SUPPORT

Recent rain helps Lake Lanier, but metro Atlanta's primary water source is a long way from normal.

1 March 2006: Drought begins.

2 April - June 2006: Faulty gauge leads to two-foot drop.

3 June 2006: Drought officially declared, triggering statewide watering restrictions. State officials warn the metro region's water supply is at risk.

4 September 2007: State enacts near-total ban on outdoor watering in North Georgia. Federal officials warn that Lanier is likely to hit a new record low.

5 October 2007: State predicts Lanier could "run dry" in 80 days and mandates 10 percent cut in North Georgia's water use.

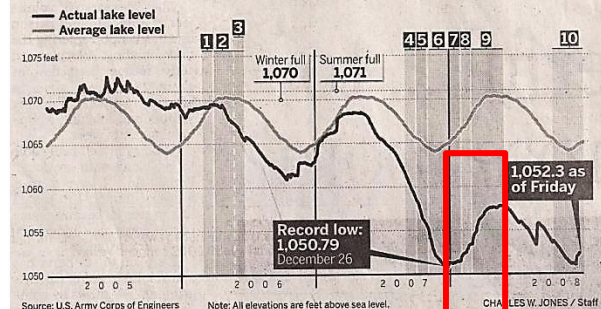
6 November - December 2007: The White House brokers a deal with Georgia, Alabama and Florida to keep more water in Lanier.

7 January 2008: General Assembly passes statewide water plan, but with no new funds.

8 February 2008: Gov. Perdue eases restrictions on outdoor water use to allow some landscape watering and filling of swimming pools.

9 March - May 2008: The U.S. Army Corps of Engineers cuts water released down the Chattahoochee River by 13 percent, holding more in Lanier.

10 November 2008 and continuing through April 2009: The corps minimizes water released from Lanier.



Design Aesthetics

- Guides for enhancing visual quality
- Set of principles to help guide designers

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

Design Principles

1. 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

Design Principles

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

Design Principles

Remove
to improve
(the **data-ink** ratio)

Created by Darkhorse Analytics

www.darkhorseanalytics.com

Design Principles

2. Maximize Data Density

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

Maximize data density

- Tufte quote

“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:

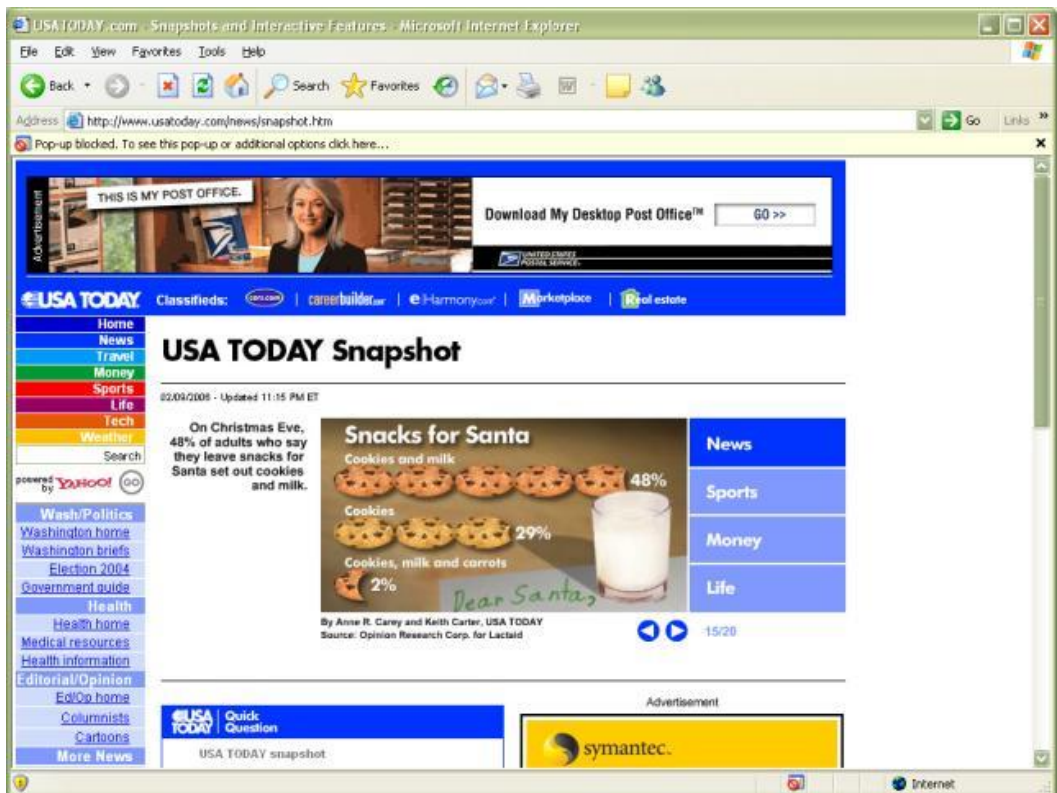
Maximize data density and the size of the data matrix, within reason.”

Vol 1, p 168


Design Principles

3. Avoid chart junk

- Extraneous visual elements that detract from message



Junk charts blog



JUNK CHARTS

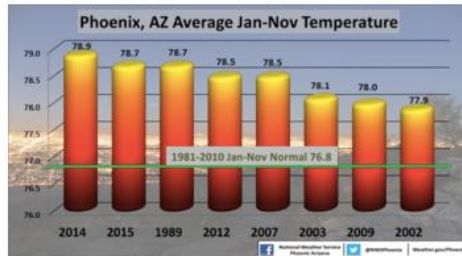
Recycling chartjunk as junk art

[Home](#) [Archives](#) [Subscribe](#) [Book Blog](#)

« [Finding meaning in Big Blue California](#) | [Main](#) | [Happy new year. Did you have a white Christmas?](#) »

Scorched by the heat in Arizona

Reader Jeffrey S. saw this graphic inside a Dec 2 tweet from the National Weather Service (NWS) in Phoenix, Arizona.




Year	Temperature (°F)
2014	78.9
2015	78.7
1989	78.7
2012	78.5
2007	78.5
2003	78.1
2009	78.0
2002	77.9
1981-2010 Jan-Nov Normal	76.8

In a Trifecta checkup ([link](#)), I'd classify this as Type QV.

The problems with the visual design are numerous and legendary. The [column chart](#) where the heights of the columns are not proportional to the data. The unnecessary 3D effect. The lack of self-sufficiency ([link](#)). The

Marketing analytics and data visualization expert. Author and Speaker. Currently at Columbia. See my full bio.

Big Data Read



Numbersense
Kaiser Fung
New \$19.20
Best \$2.20

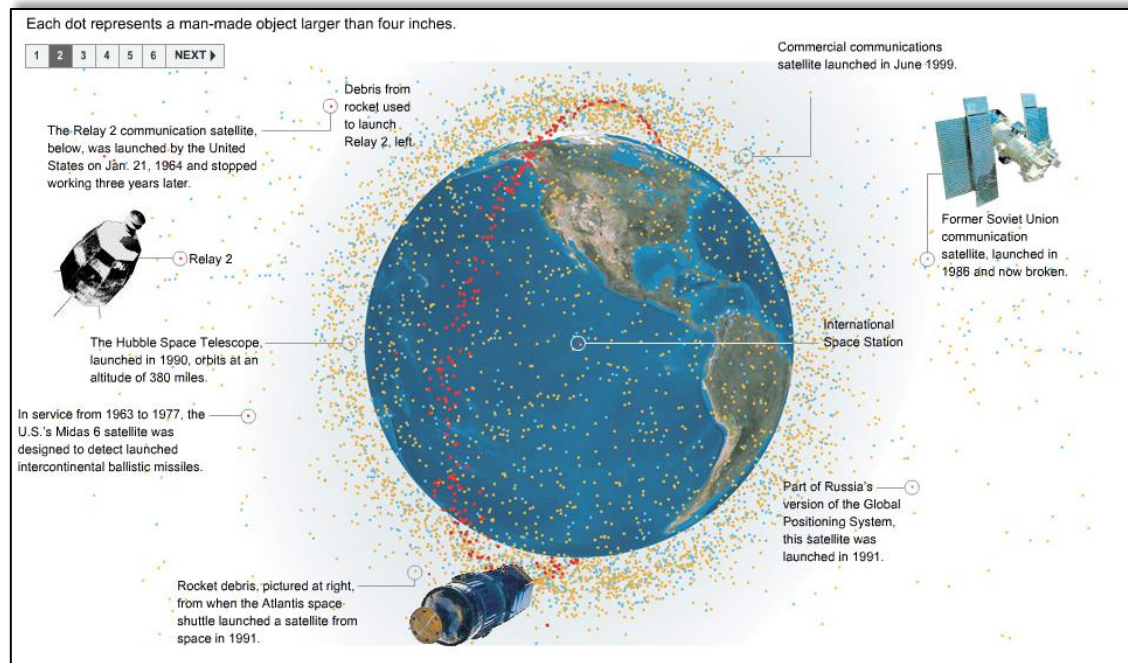
[Privacy Information](#)

<http://junkcharts.typepad.com>

Design Principles

4. Utilize multifunctioning graphical elements (macro/micro readings)

- Graphical elements that convey data information and a design function
- Overview with high levels of details

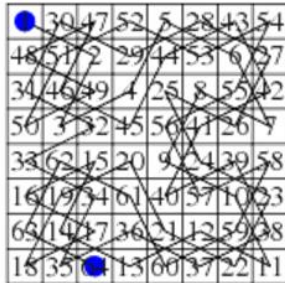


Design Principles

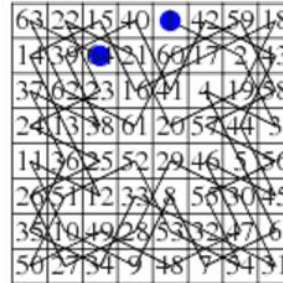
5. Use small multiples

- Repeat visually similar graphical elements nearby rather than spreading far apart

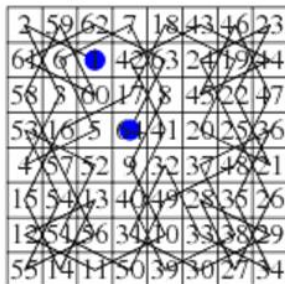
Beverly 1848, $\Sigma = 282, 210$



Jaenisch 1862, $\Sigma = 348, 168$



Francony 1882, $\Sigma = 264, 256$



Magic Night's Tours

<http://mathworld.wolfram.com/MagicTour.html>

Recent additions

- Sparkline
- Small, repeated graphics (frequently line graphs)

Sparklines: theory and practice

Theory of sparklines (small, intense, simple datawords) along with many practical examples of recent sparkline developments. From Edward Tufte's book *Beautiful Evidence*.

-- Edward Tufte, May 27, 2004

Sparklines: theory and practice

Theory of sparklines (small, intense, simple datawords) along with many practical examples of recent sparkline developments. Excerpts from Edward Tufte, *Beautiful Evidence*. New examples or helpful comments much appreciated.

ET

Sparklines: Intense, Simple, Word-Sized Graphics

THE most common data display is a noun followed by some numbers. For example, a medical patient's current level of glucose is typically reported in a clinical record as a word and number:

Sparklines Example

Sparklines have obvious applications for financial and economic data, by tracking changes over time, showing overall trend as well as local detail. Part of a data table, this sparkline depicts the euro exchange rate (dollar cost of one euro) for every day in the previous year:

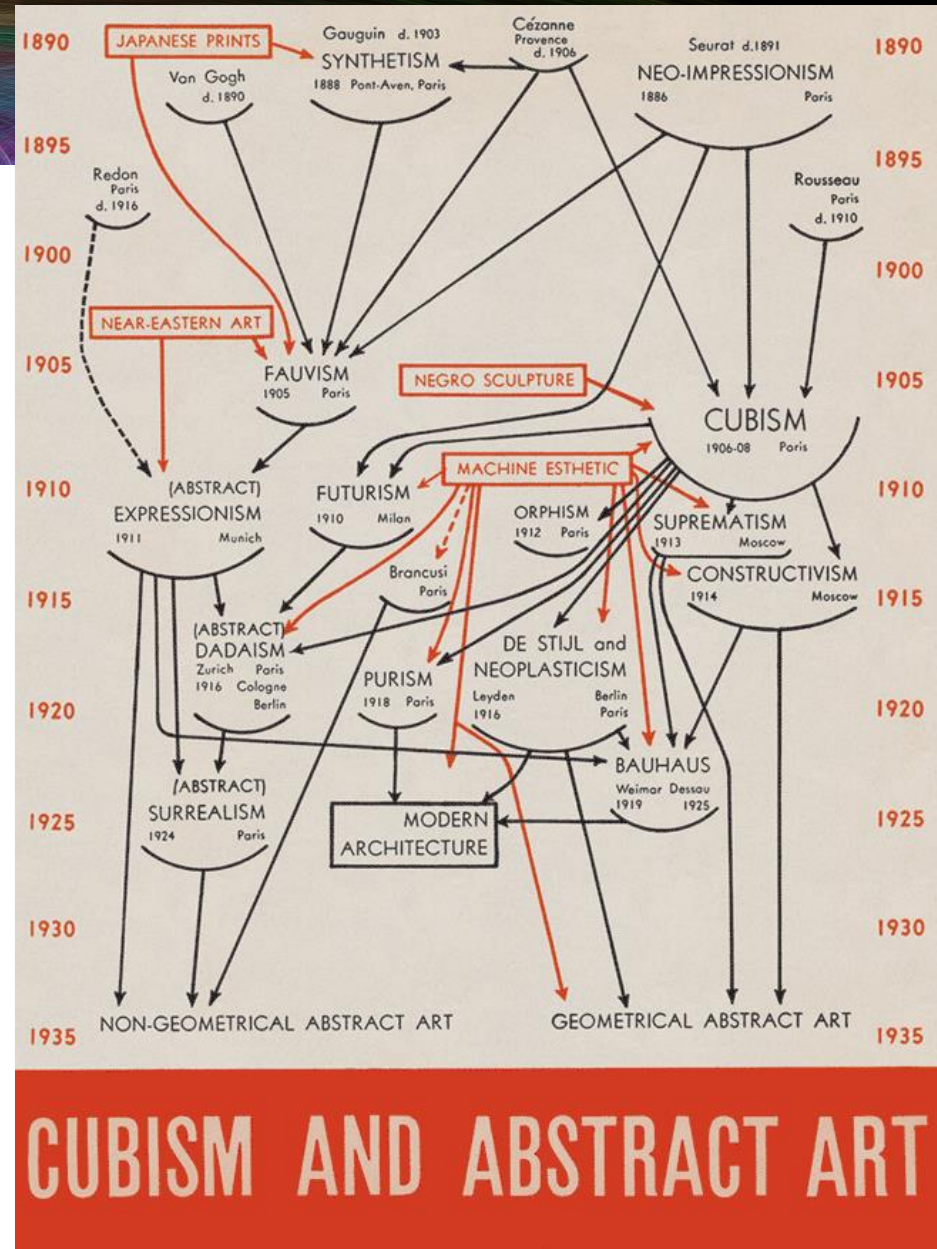


Colors serve to link the sparkline and the numbers: red = the oldest and newest rates in the series; blue = yearly low and high for daily exchange rates. Extending this graphic table is straightforward; here, the price of the euro in terms of 3 other currencies for 65 months and for 12 months:



Design Principles

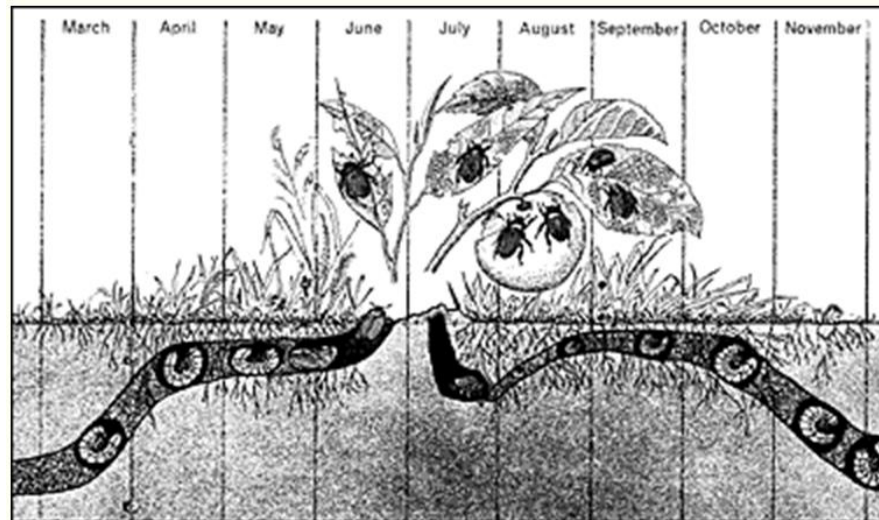
6. Show mechanism, process, dynamics, and causality
- Cause and effect are key
 - Make graphic exhibit causality (if available)



Design Principles

7. Escape flatland

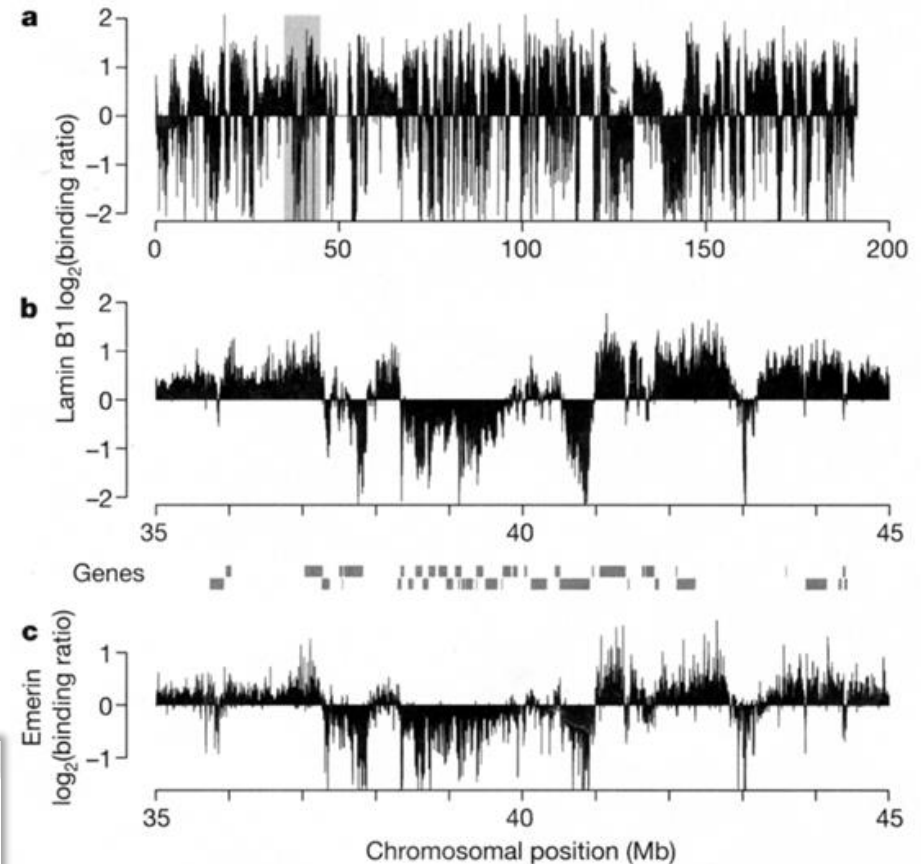
- Data is multivariate
- Doesn't necessarily mean 3D projection



Nasty, Brutish and Short: The life cycle of the Japanese beetle (detail), a great graphic, and "a smooth escape from flatland," says Tufte.

Design Principles

8. Utilize layering and separation
- Independent elements become more organized and discernible.
 - Relationships between data points are more clearly defined.
 - 1+1 = 3 or more
 - Good or bad



nature

Vol 453 | 12 June 2008 | doi:10.1038/nature06947

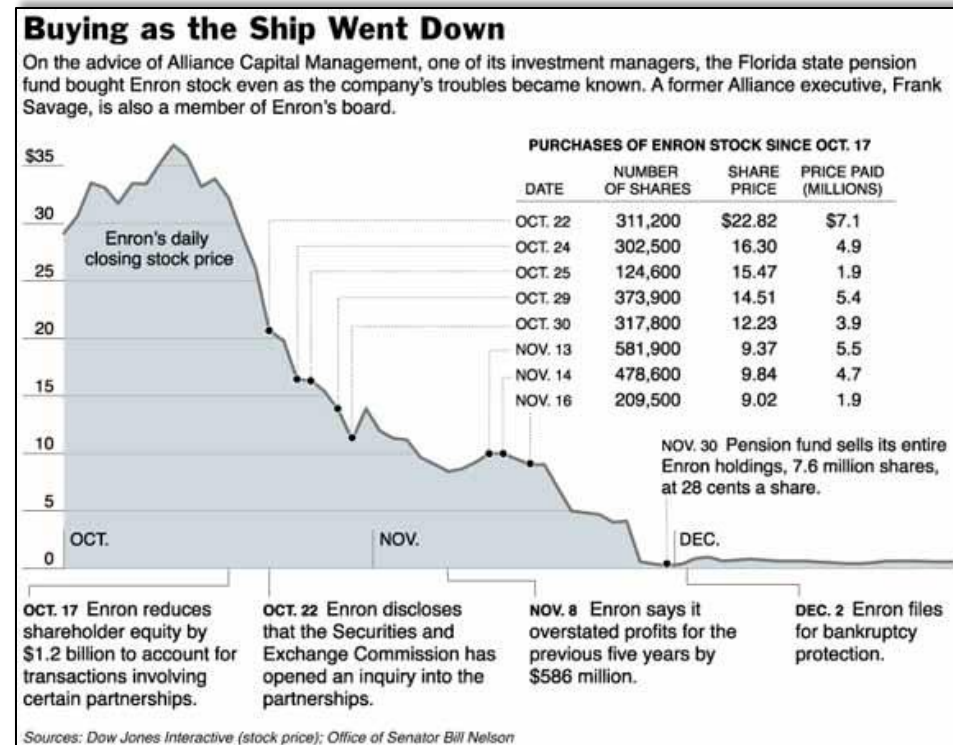
LETTERS

Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions

Lars Guelen¹, Ludo Pagie¹, Emilie Brasset², Wouter Meuleman^{1,4}, Marius B. Faza¹, Wendy Talhout¹, Bert H. Eussen³, Annelies de Klein³, Lodewyk Wessels^{1,4}, Wouter de Laat² & Bas van Steensel¹

Design Principles

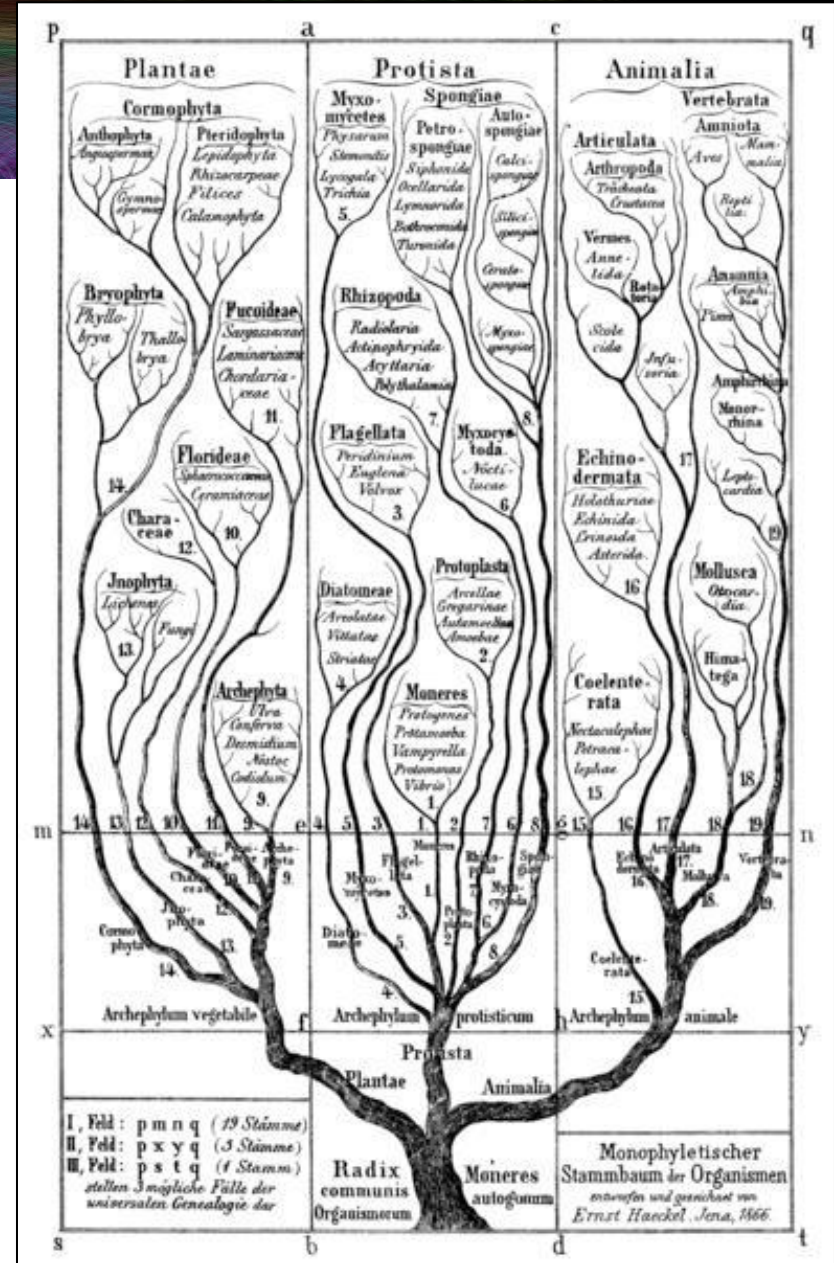
9. Utilize narratives of space and time
- Tell a story of position and chronology through visual elements



Design Principles

10. Content is king

- Quality, relevance and integrity of the content is fundamental
- What the data structure? Find the right visual metaphor
- What's the analysis task? Make the visual design reflect that
- Integrate text, chart, graphic, map into a coherent narrative



10 Design Principles

1. Maximize Data-Ink Ratio
2. Maximize Data Density
3. Avoid Chart Junk
4. Utilize multifunctioning graphical elements (macro/micro readings)
5. Use Small Multiples
6. Show mechanism, process, dynamics, and causality
7. Escape Flatland
8. Utilize Layering and Separation
9. Utilize Narratives of Space and Time
10. Content Is King



In Summary

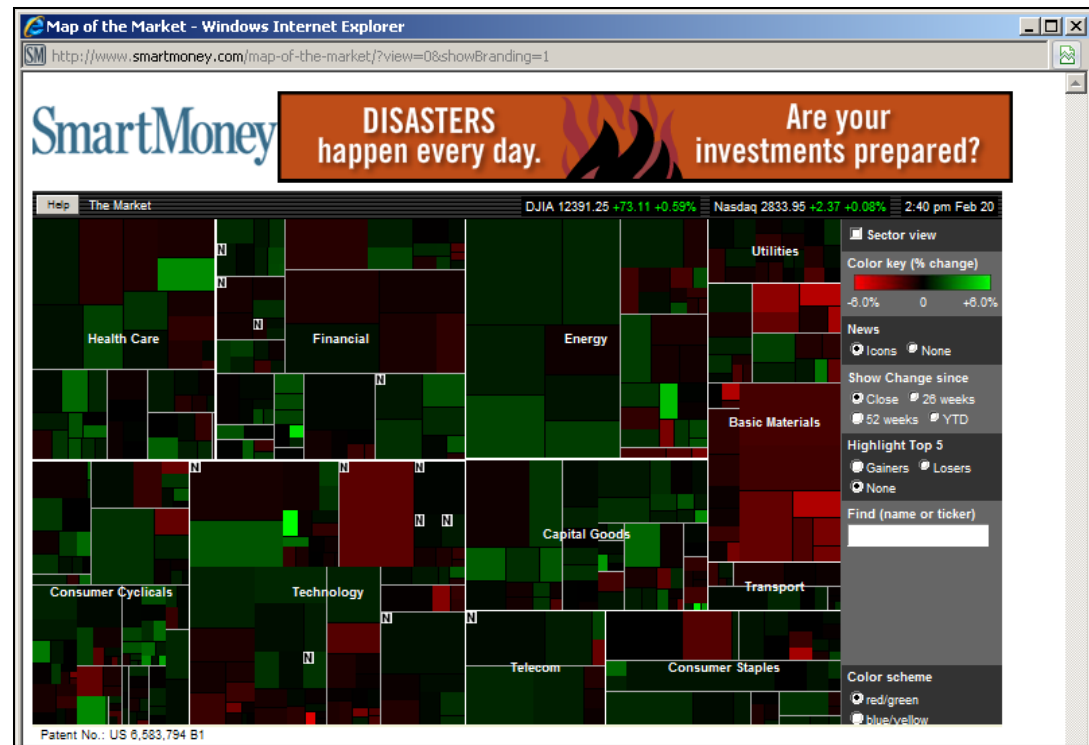
1. Tell the truth
 - Graphical integrity
2. Do it effectively with clarity, precision...
 - Design aesthetics

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 use of color

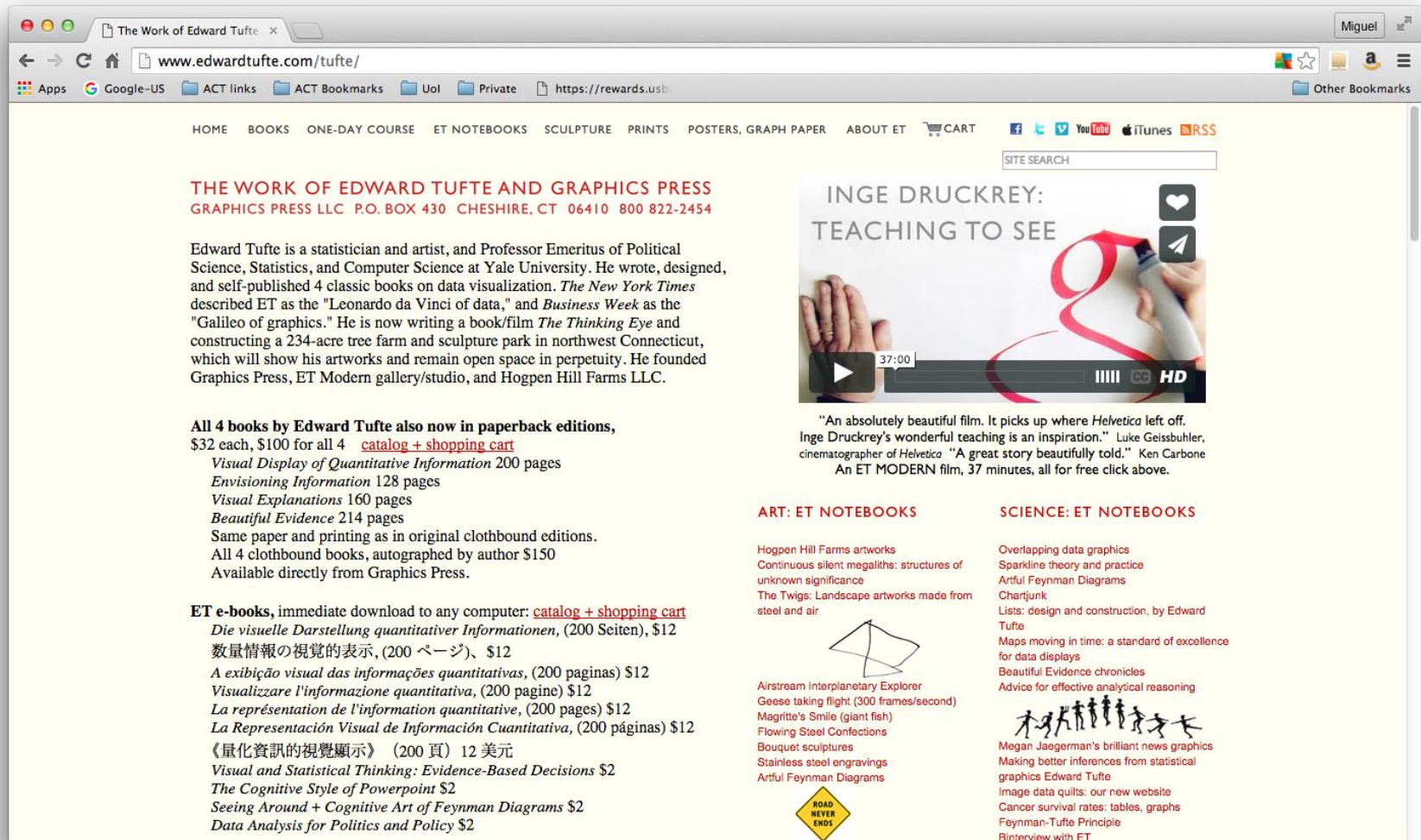
- to Label
- to Measure
- to Represent or imitate reality
- to Enliven or decorate



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

Tufte's website

A screenshot of a web browser displaying the website of Edward Tufte. The browser's address bar shows 'www.edwardtufte.com/tufte/'. The website has a clean, minimalist design with a white background and black text. The navigation bar at the top includes links for HOME, BOOKS, ONE-DAY COURSE, ET NOTEBOOKS, SCULPTURE, PRINTS, POSTERS, GRAPH PAPER, ABOUT ET, and a CART icon. There are also social media icons for Facebook, Twitter, YouTube, and iTunes, along with an RSS feed icon. A search bar is located on the right side of the page. The main content area is divided into two columns. The left column features a section titled 'THE WORK OF EDWARD TUFTE AND GRAPHICS PRESS' with contact information. Below this, there is a paragraph about Edward Tufte's background and his work. Further down, a section titled 'All 4 books by Edward Tufte also now in paperback editions,' lists four books with their page counts and prices. The right column features a video player for a film titled 'INGE DRUCKREY: TEACHING TO SEE'. Below the video player, there is a quote from Luke Geissbuhler and Ken Carbone. At the bottom of the page, there are two columns of text under the headings 'ART: ET NOTEBOOKS' and 'SCIENCE: ET NOTEBOOKS', listing various artworks and publications. The browser's status bar at the bottom shows the user's name 'Miguel' and the page's URL.

Tufte's discussion forum (ET Notebooks)

The screenshot shows a web browser window with the address bar displaying www.edwardtufte.com/bboard/q-and-a?topic_id=1. The page features a navigation bar with links to HOME, BOOKS, ONE-DAY COURSE, ET NOTEBOOKS, SCULPTURE, PRINTS, POSTERS, GRAPH PAPER, ABOUT ET, and a CART icon. Social media icons for Facebook, Twitter, YouTube, iTunes, and RSS are also present. A search bar labeled "SITE SEARCH" is located on the right. The main content area is titled "EDWARD TUFTES BEAUTIFUL EVIDENCE" and contains a list of endorsements from various publications and individuals, including *Business Week*, *ZDNET*, *Scientific American*, *ID Magazine*, and *Nature*. Below this, a paragraph explains the purpose of the moderated forum. A legend indicates that "***" represents 3-star threads. The section "Best, E.T." lists a link to "Current Topics" and a link to a "Complete List of All Active Topics". A list of links follows, each followed by "Edward Tufte": [Image Quilts](#), [Goldberg and Design Variations](#), [Table and timetable design and typography](#), [Overlapping data graphics to make comparisons](#), [Chartjunk](#), [Consulting stories](#), [Hogpen Hill Farms LLC: Architectural evergreens and ornamental grasses](#), [On the Edge, At the Margin: Contours, Surrounds, Frames](#), [Graphical timetables](#), [Boxplots data test](#), [Design review: Improving a good news graphic](#), [Design of walking maps, indoors and outdoors](#), [John Tukey "badmandments" in statistical work, mainly in the behavioral sciences](#), [Image data quilts: our new website](#), and [Making better inferences from statistical graphics](#).

Edward Tufte Forum

www.edwardtufte.com/bboard/q-and-a?topic_id=1

Apps Google-US ACT links ACT Bookmarks Uol Private https://rewards.usb

HOME BOOKS ONE-DAY COURSE ET NOTEBOOKS SCULPTURE PRINTS POSTERS, GRAPH PAPER ABOUT ET CART

Facebook Twitter YouTube iTunes RSS

SITE SEARCH

EDWARD TUFTES BEAUTIFUL EVIDENCE

Business Week, **Best Innovation and Design Books for 2006**, "A brilliant masterpiece, the Galileo of graphics has done it again." *ZDNET*, **Best Business and Technology Books of 2006**, "Tufte will get you thinking about the meaning of words and images, not to mention your ability to tell the truth. A beautiful book." *Scientific American*, **The Editors Recommend**, "Outstanding examples, how seeing turns into showing." *ID Magazine*, "Hello Gorgeous! A master of information design practices what he preaches." Martin Kemp, *Nature*, "The world's leading analyst of graphic information."

At this moderated forum, I will answer questions dealing with information design. Others can then extend the discussion. I will try to answer questions that have general interest or where I have something to say. Not all questions will be answered, usually because I don't know the answer.

*** = 3-star threads

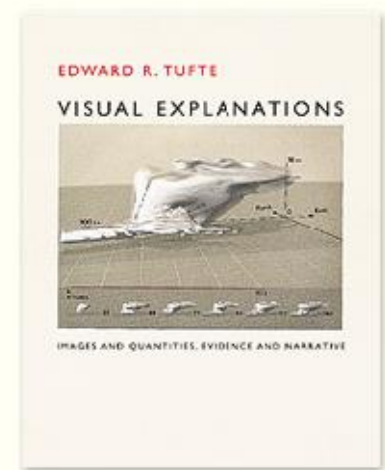
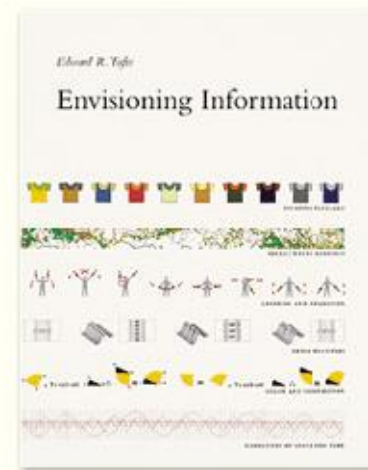
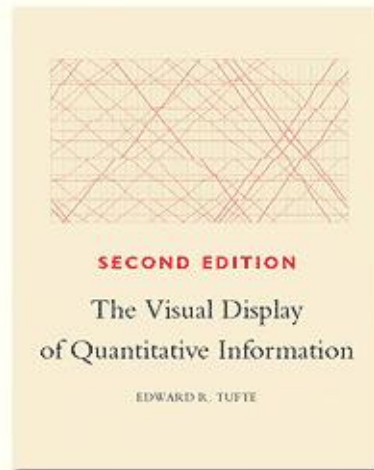
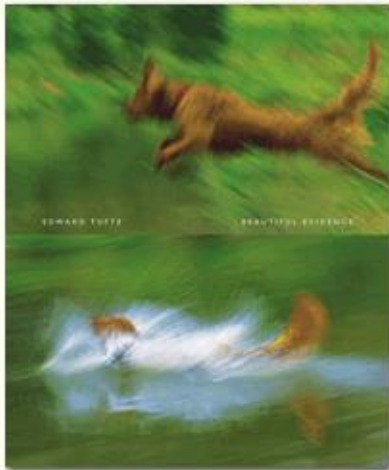
Best, E.T.

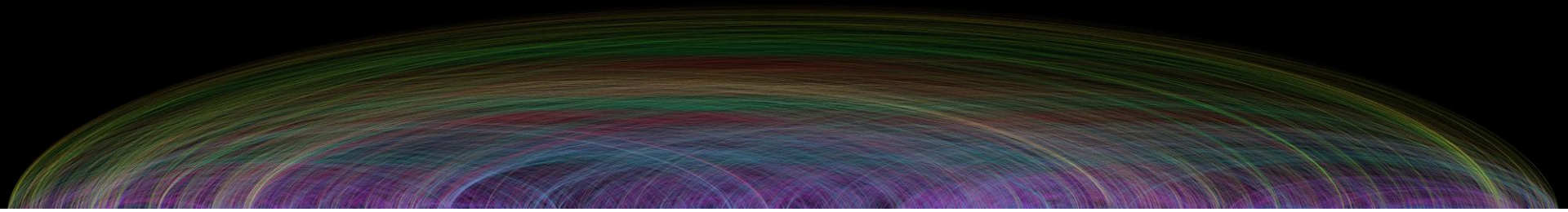
[[Current Topics](#) | [Complete List of All Active Topics](#)]

[Image Quilts](#) ET
[Goldberg and Design Variations](#) ET
[Table and timetable design and typography](#) ET
[Overlapping data graphics to make comparisons](#) Edward Tufte
[Chartjunk](#) Edward Tufte
[Consulting stories](#) Edward Tufte
[Hogpen Hill Farms LLC: Architectural evergreens and ornamental grasses](#) Edward Tufte
[On the Edge, At the Margin: Contours, Surrounds, Frames](#) Edward Tufte
[Graphical timetables](#) Edward Tufte
[Boxplots data test](#) Edward Tufte
[Design review: Improving a good news graphic](#) Edward Tufte
[Design of walking maps, indoors and outdoors](#) Edward Tufte
[John Tukey "badmandments" in statistical work, mainly in the behavioral sciences](#) Edward Tufte
[Image data quilts: our new website](#) Edward Tufte
[Making better inferences from statistical graphics](#) Edward Tufte

Additional Readings

- Tufte, *Envisioning Information*.

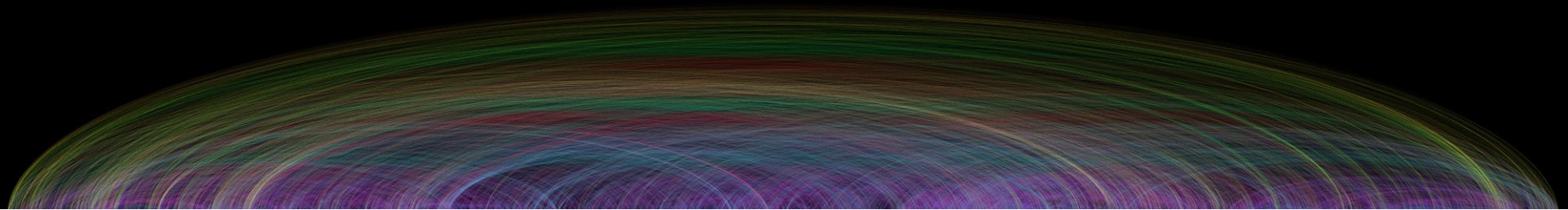




When More Is Better

THE FLIPSIDE



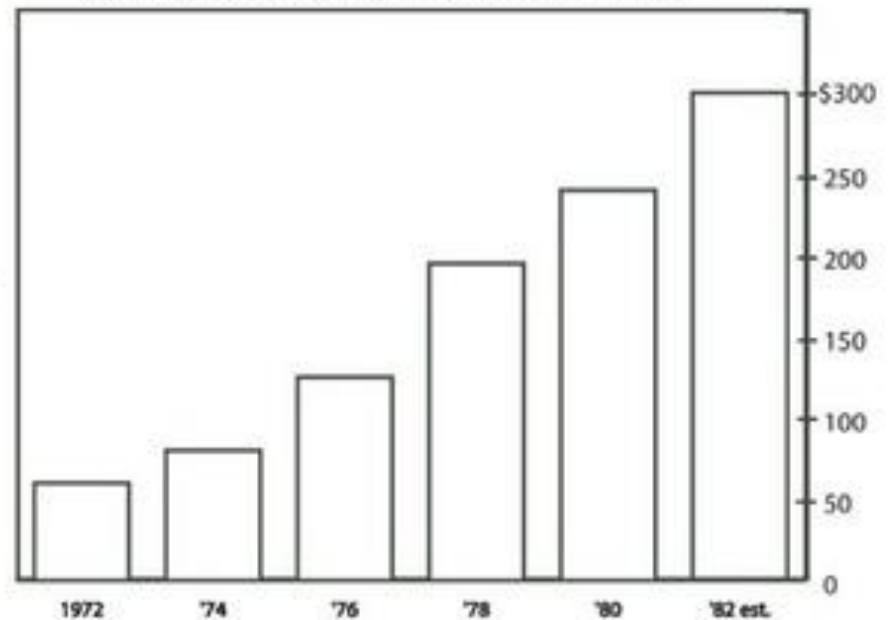


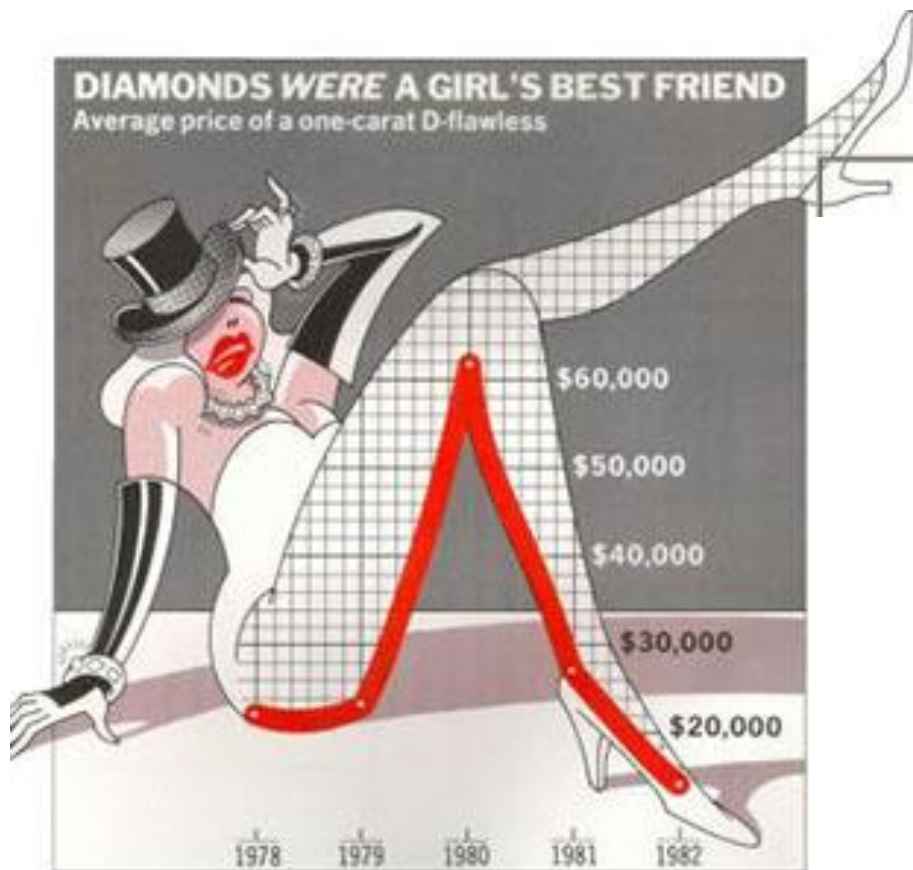
MONSTROUS COSTS

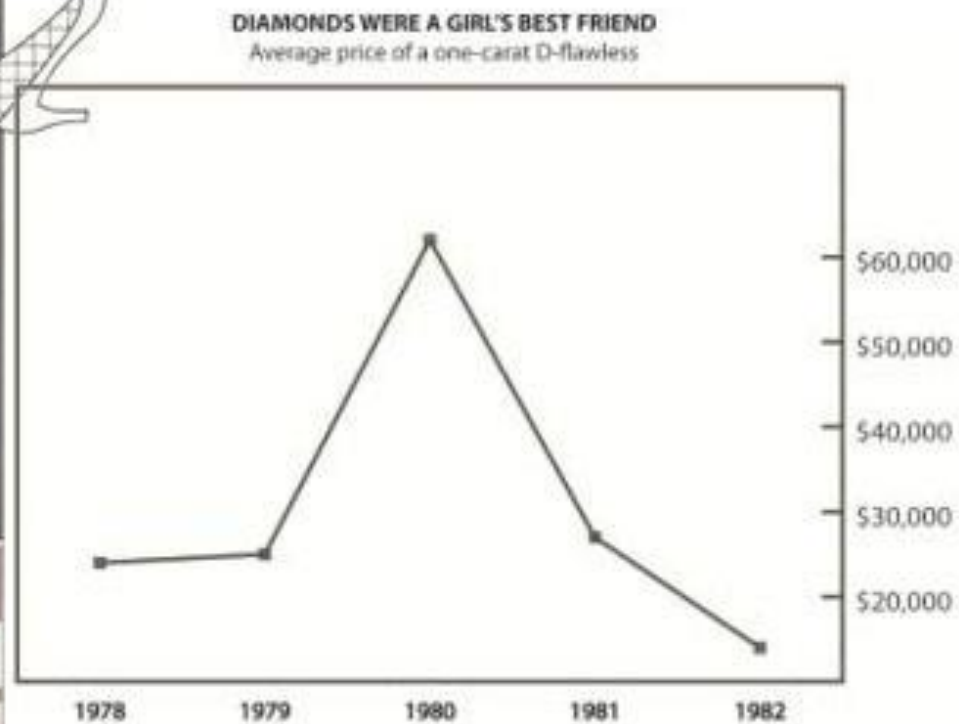
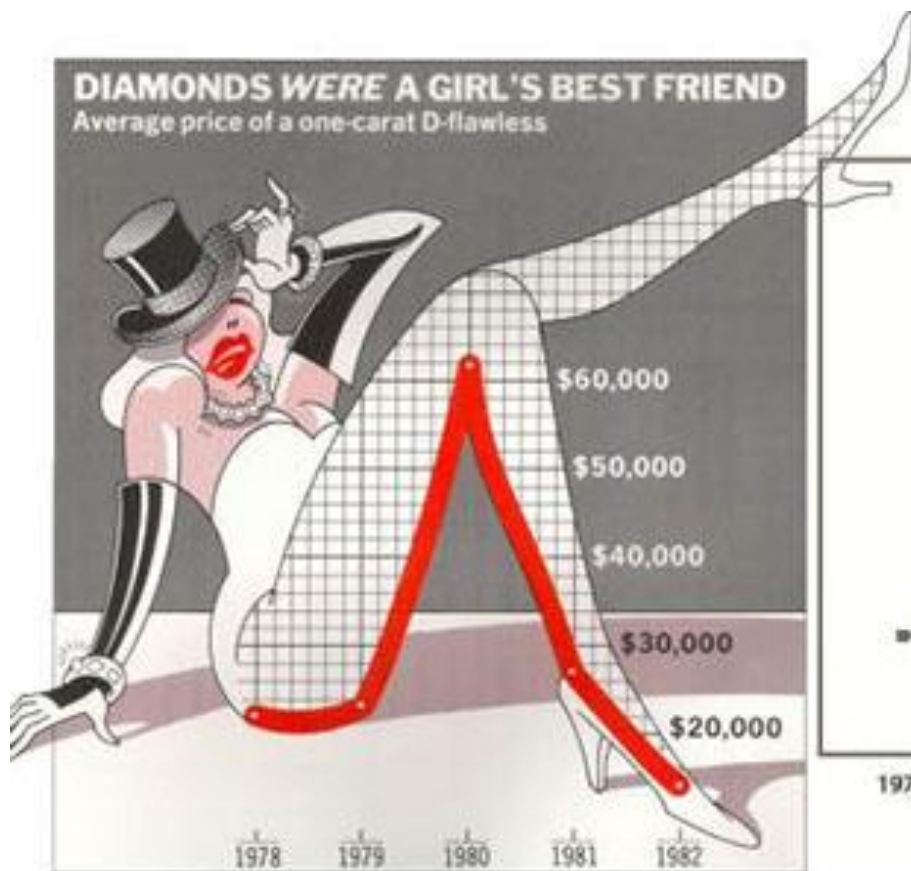
Total House and Senate campaign expenditures, in millions



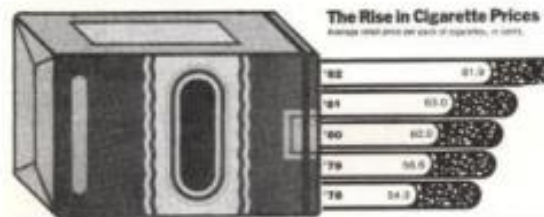
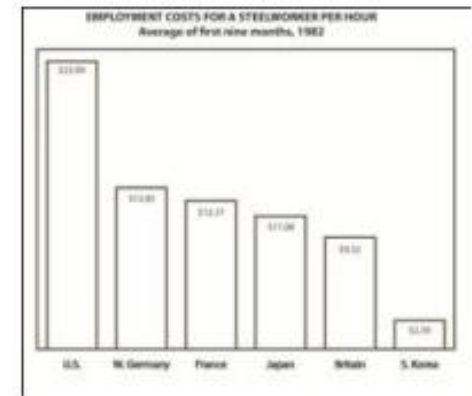
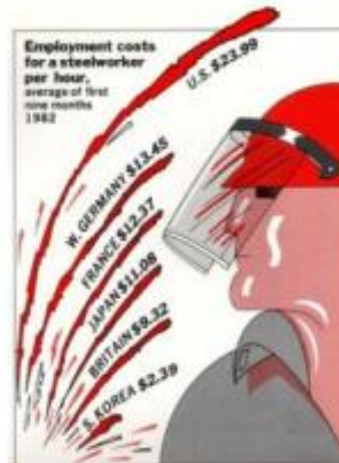
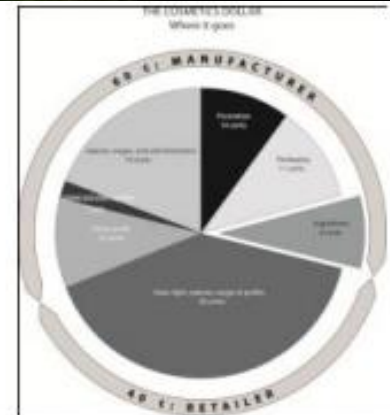
MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions







Experiment



Gaze tracking

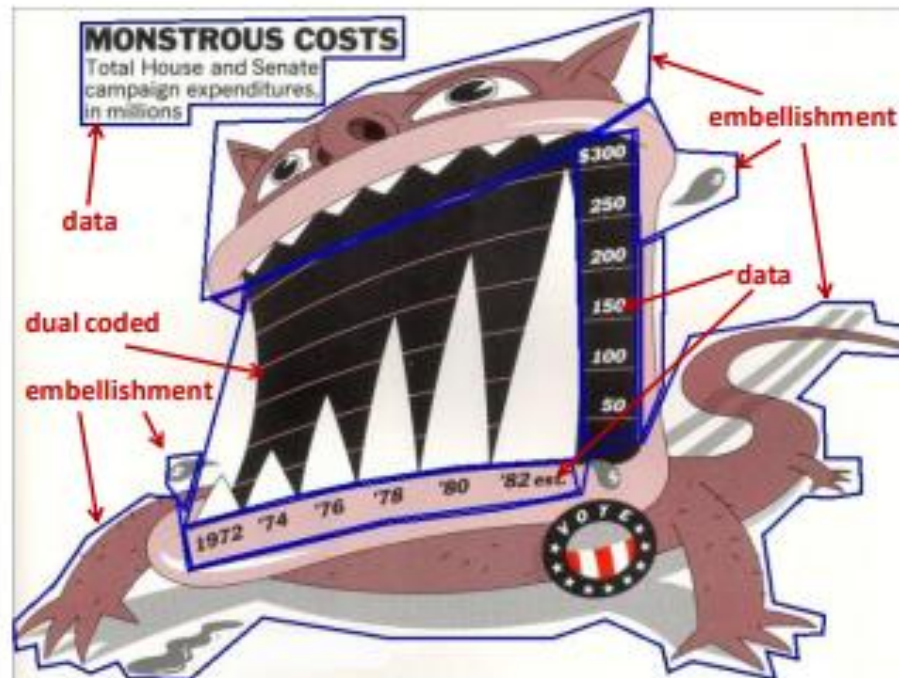


Figure 3. Area of interest analysis: gaze falling in any of the areas defined by the blue borders were labeled as data, embellishment, dual encoded, or other.

Results

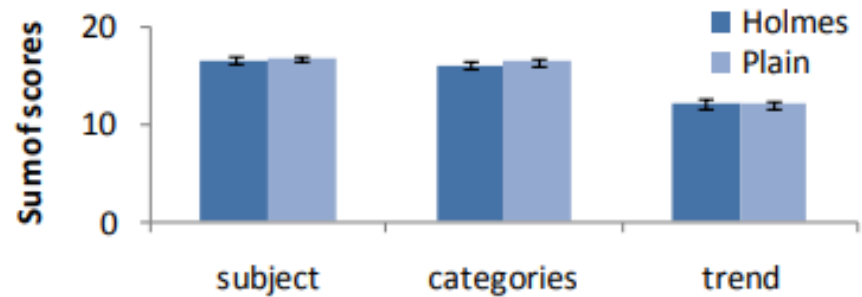


Figure 4. Means \pm SE for description scores.

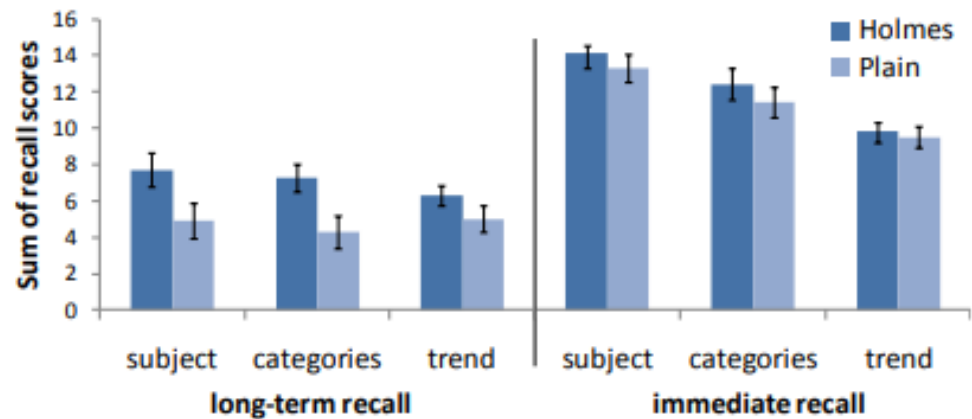


Figure 5. Means \pm SE for recall scores for long-term and immediate recall.

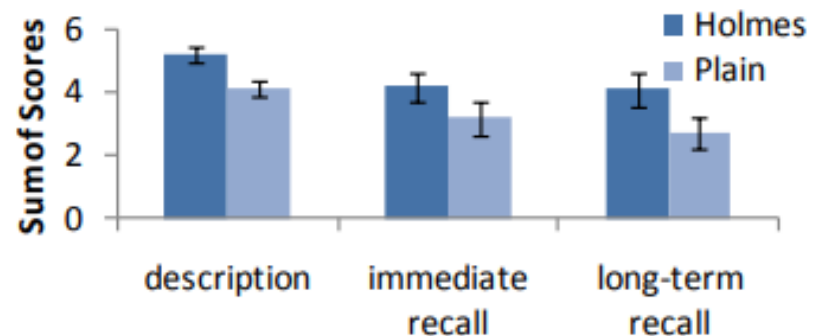
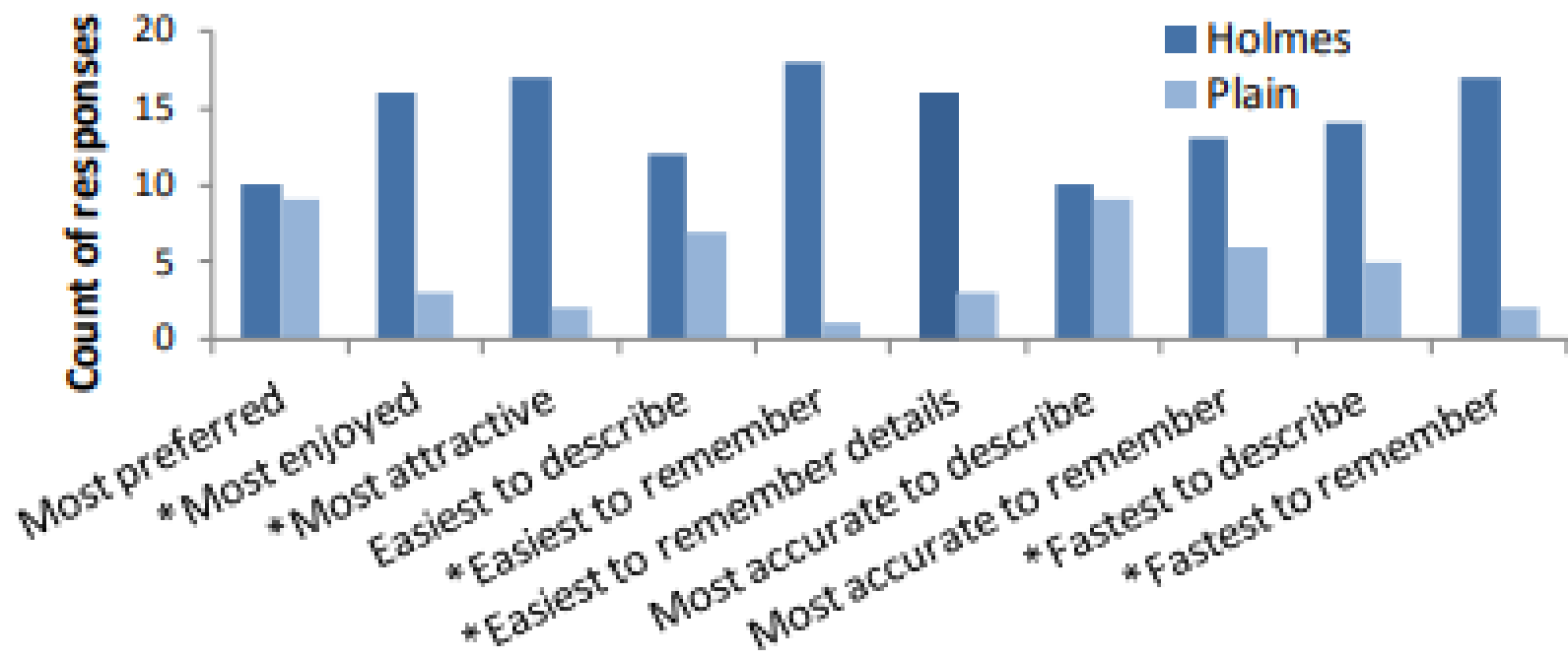
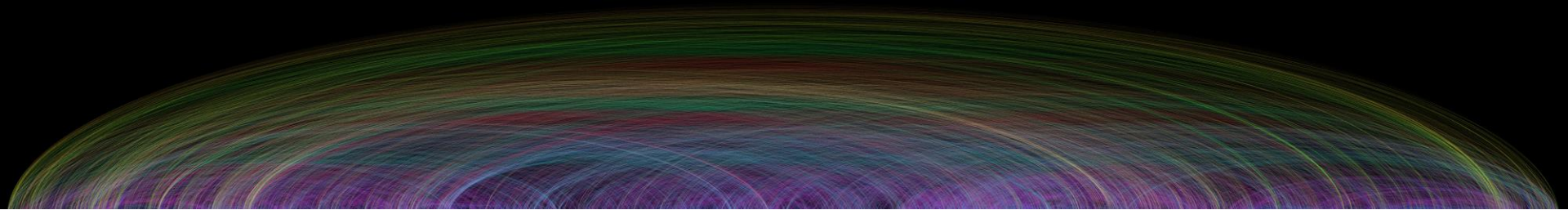
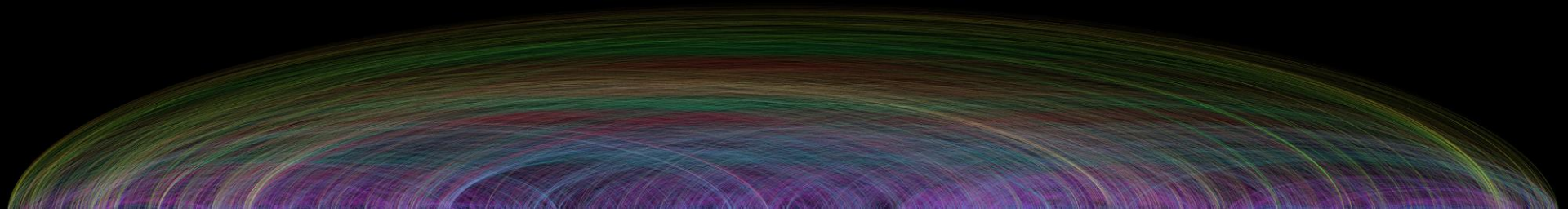


Figure 6. Means \pm SE for sum of value message scores.



- 
- S. Bateman, et al, "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts", *Proceedings of CHI '10*, April 2010, pp. 2573-2582.