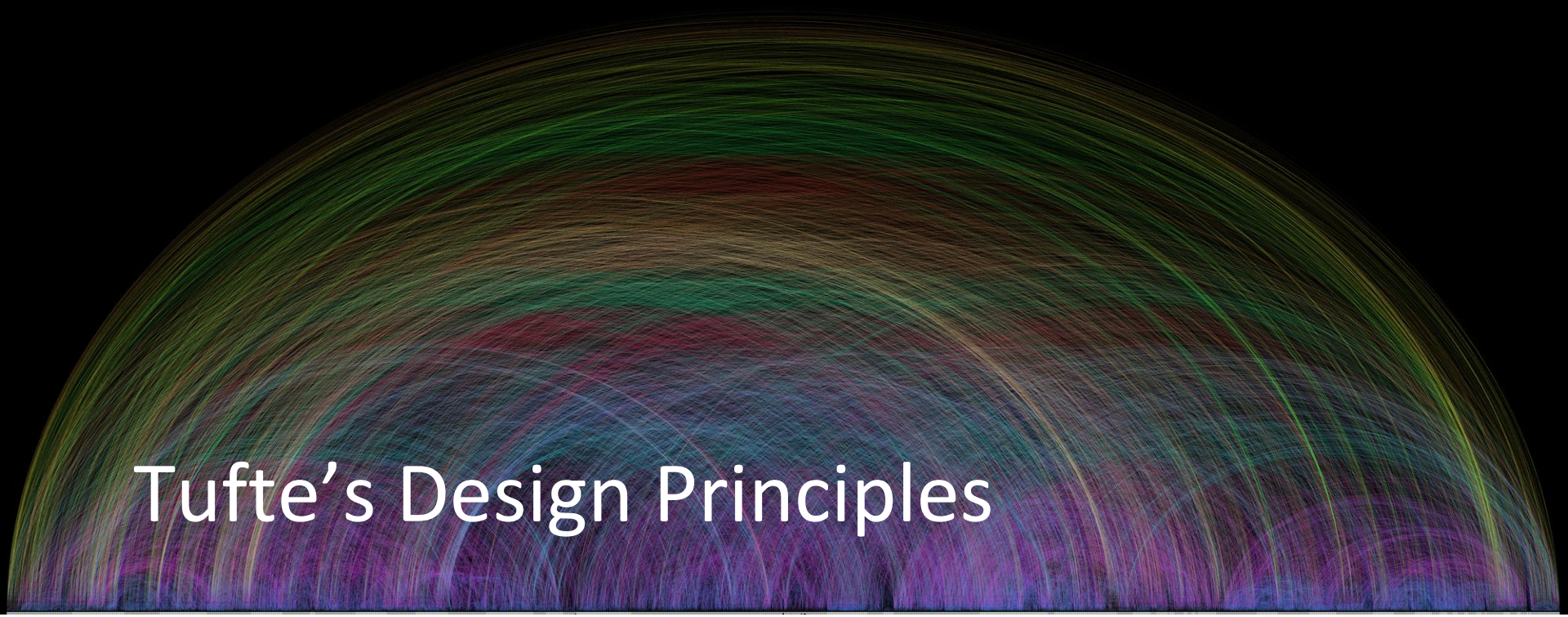
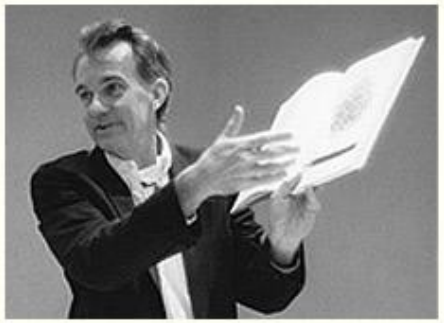


Tufte's Design Principles

BAIS 6140 – Information Visualization

L. Miguel Encarnação

Agenda

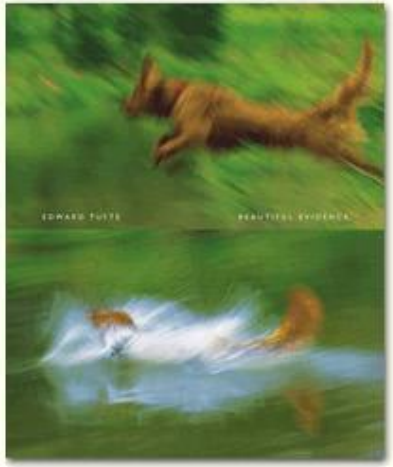


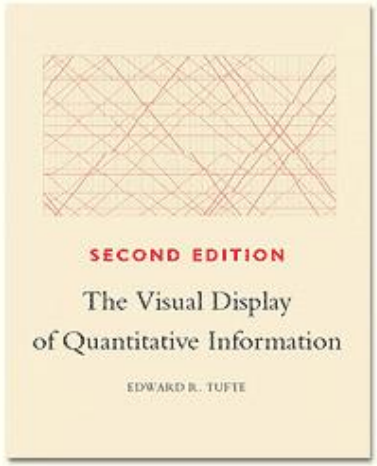
Edward Tufte has written seven books, including *Visual Explanations, Envisioning Information, The Visual Display of Quantitative Information,* and *Data Analysis for Politics and Policy.* He writes, designs, and self-publishes his books on analytical design, which have received more than 40 awards for 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.

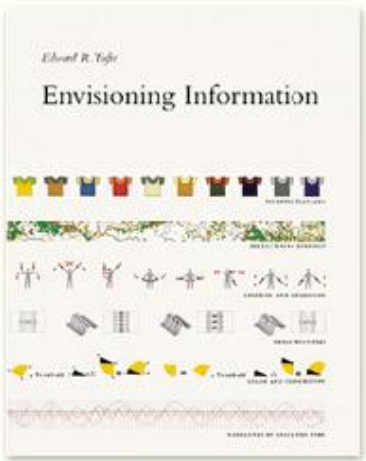
This website describes Edward Tufte's books, one day course, and artwork. For further information, call Graphics Press at 203 272-9187, or fax 203 272-8600, or email.

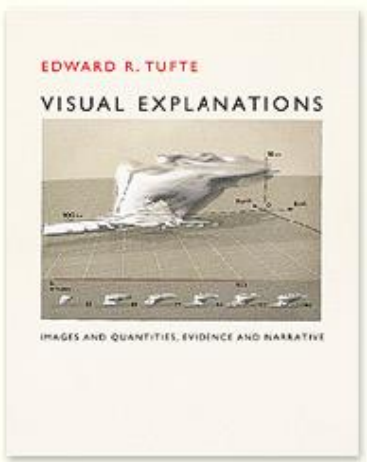
For a moderated forum on analytical design, go to ASK E.T.

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, synopsize, winnow the wheat from the chaff, and separate the sheep from the goats."

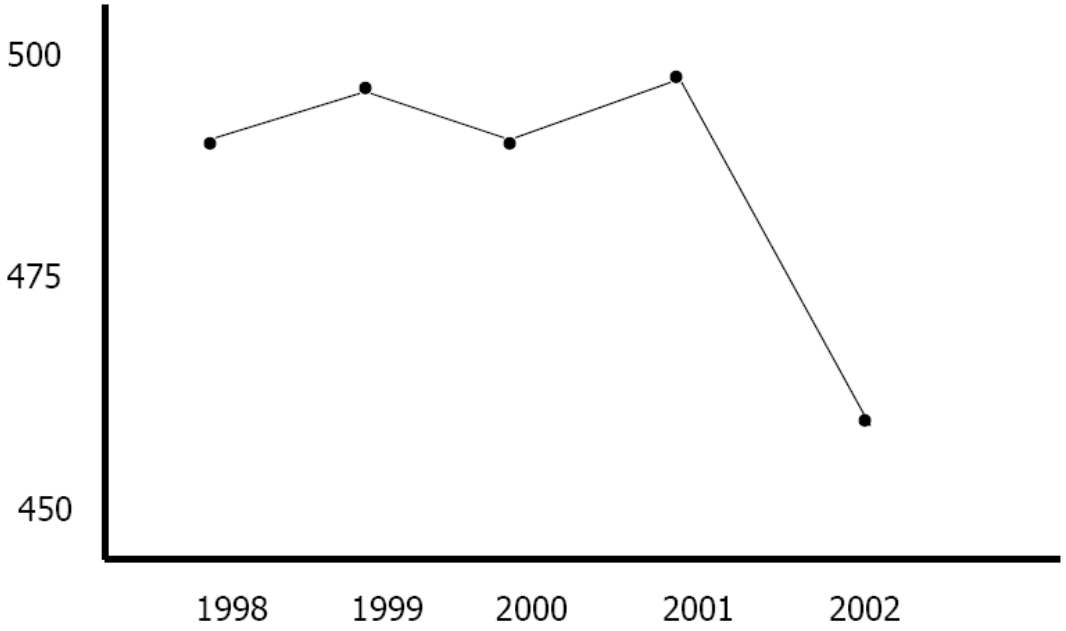
  Tufte, *Vol.2, page 50*

In Summary

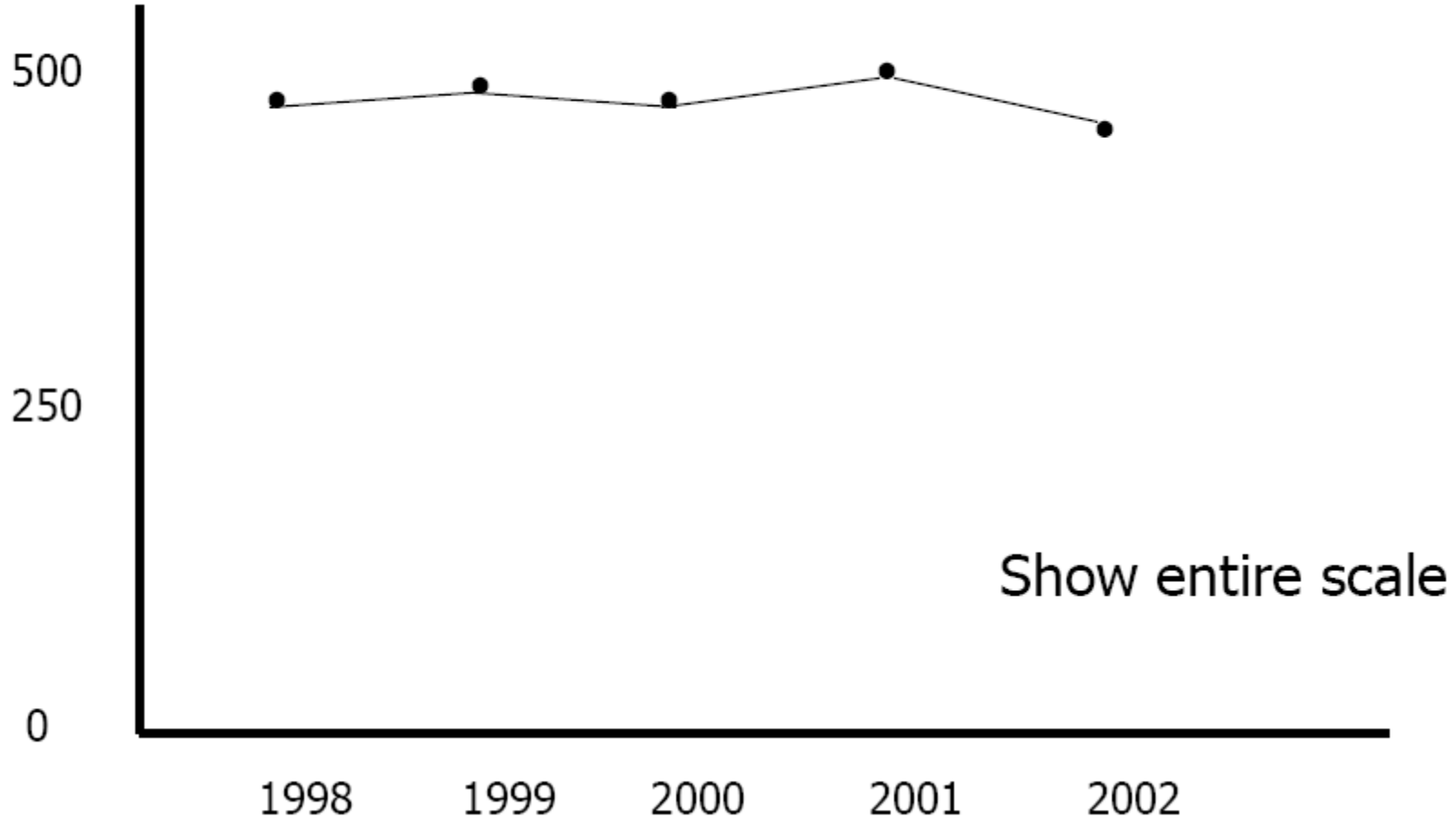
1. Tell the truth
   - Graphical integrity
2. Do it effectively with clarity, precision…
   - Design aesthetics

1. Graphical Integrity

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



Truth being …



Truth being …

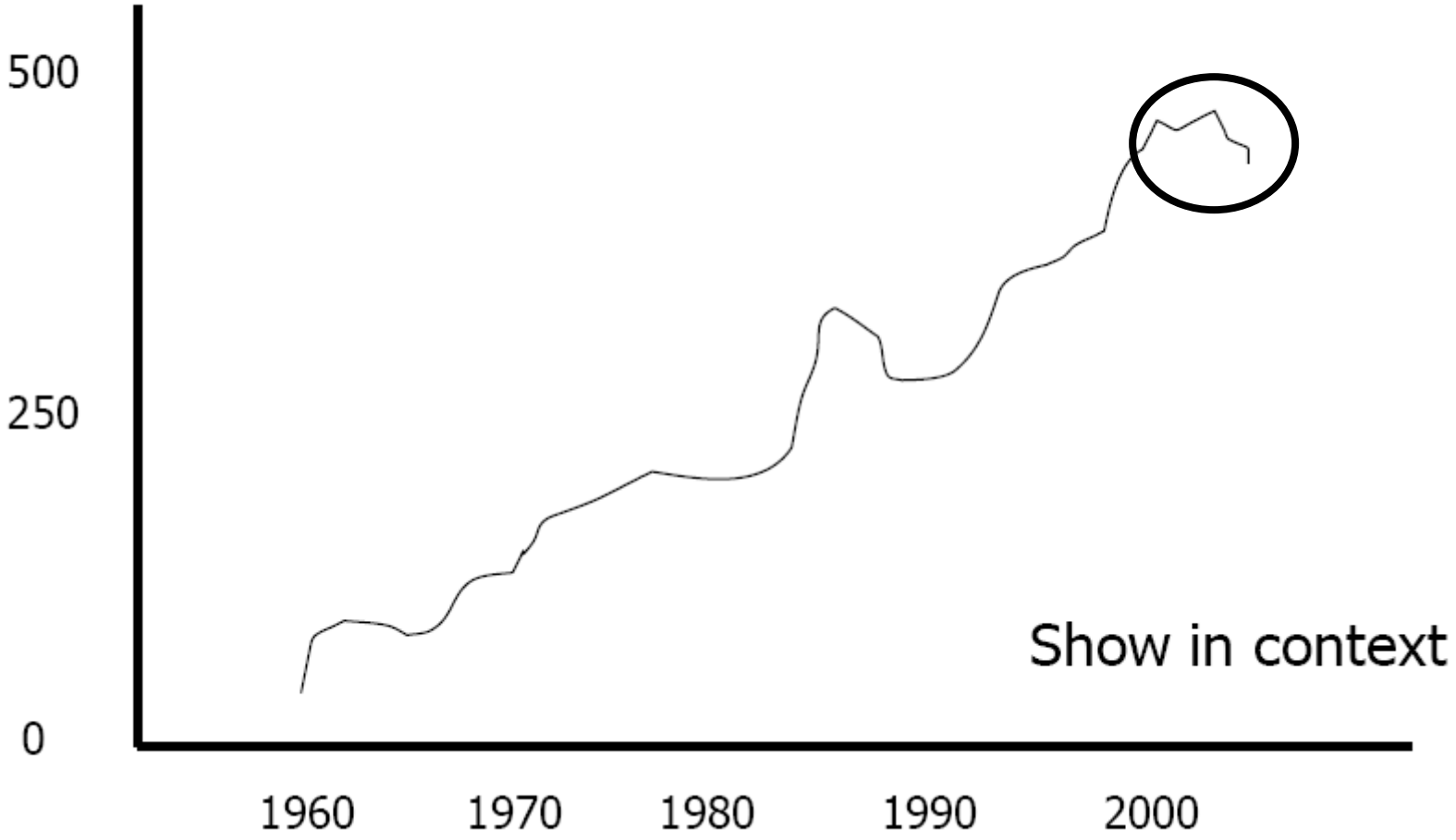
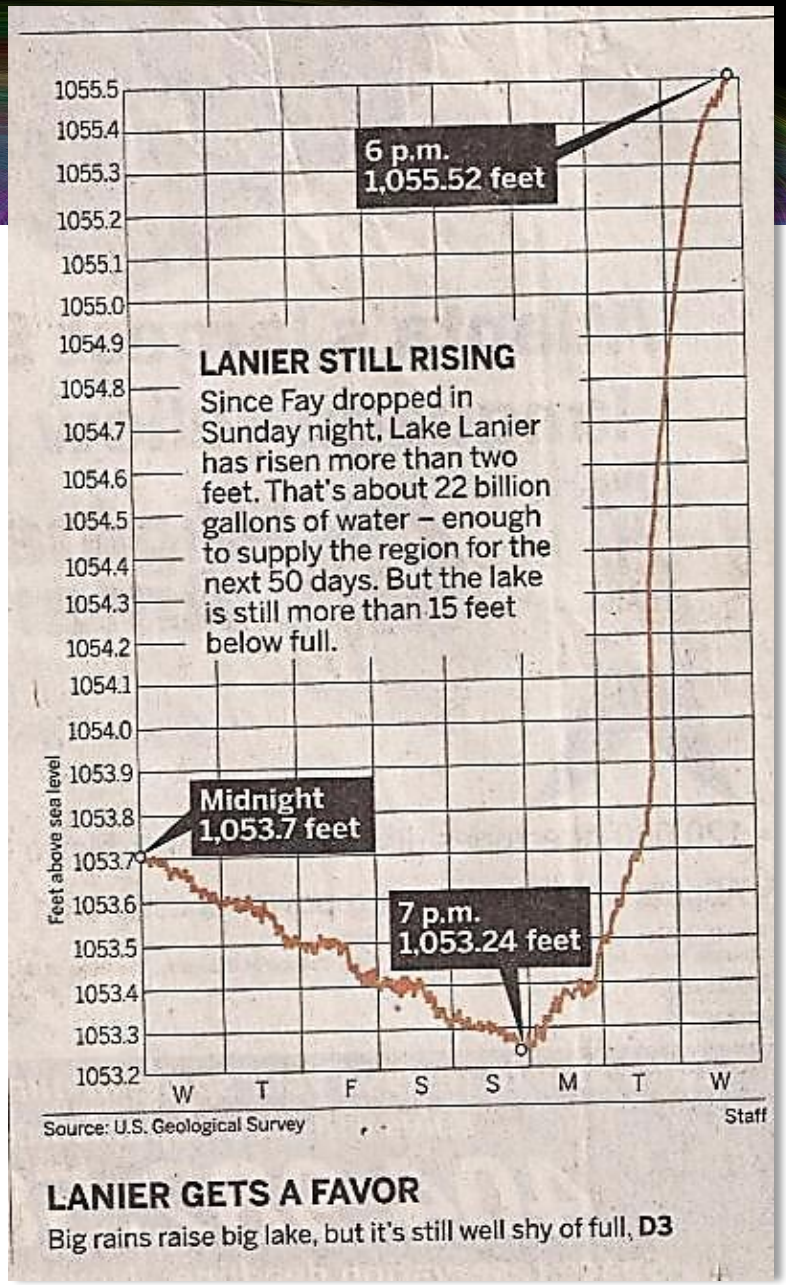


Chart Integrity

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

Example

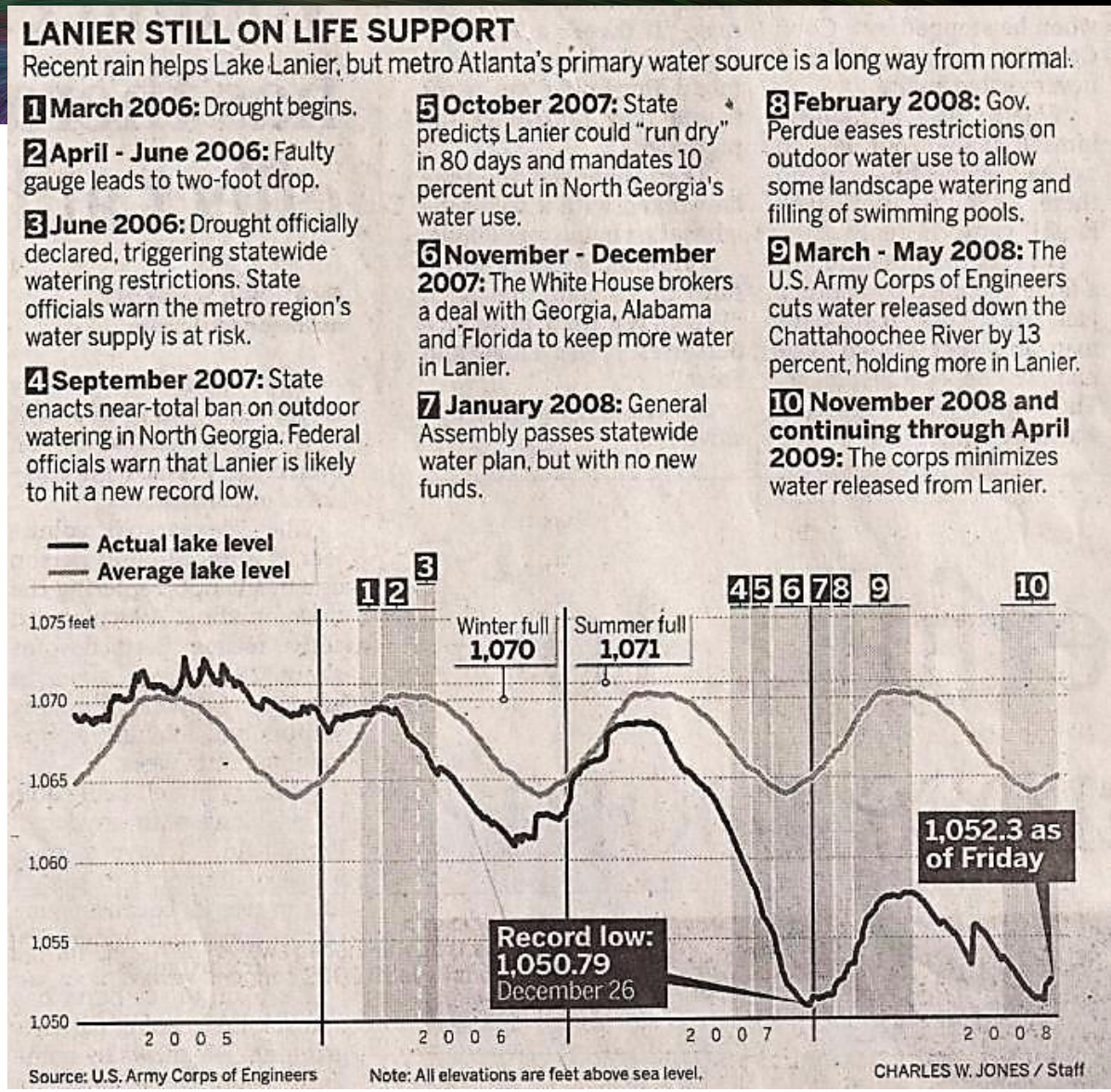
- A huge rise?



Atlanta Journal Constitution
Summer '08

Example

- More of the data



Atlanta Journal Constitution Dec. '08

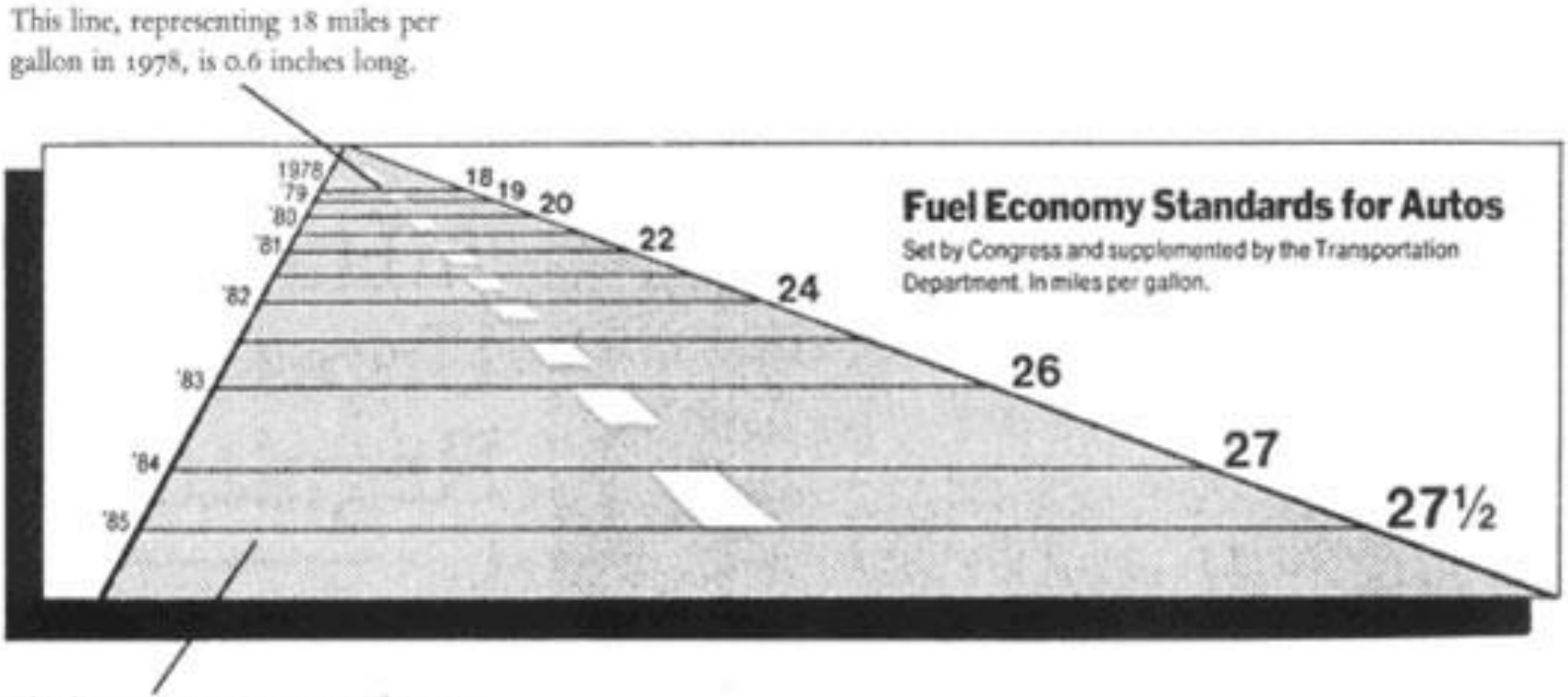
Watch size coding

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

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

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

Lie factor: Example



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

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”.

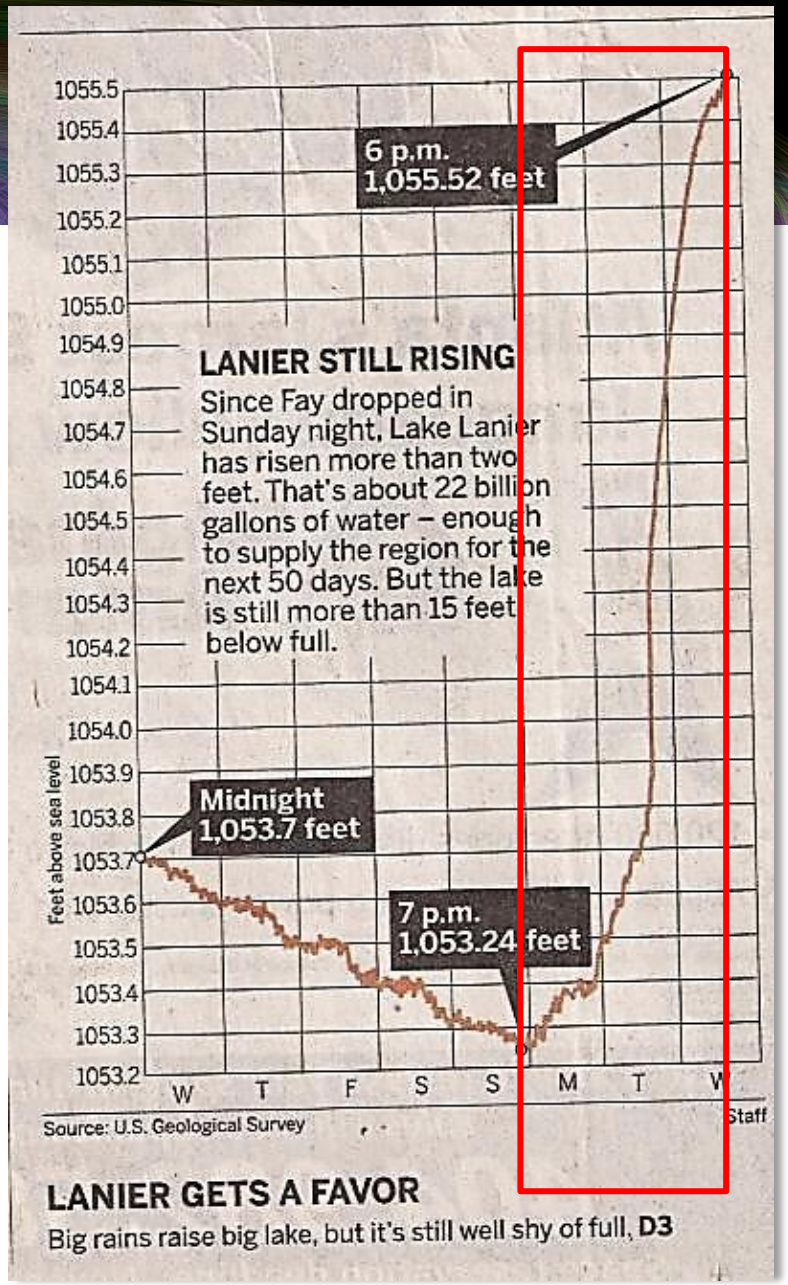
Lie factor: Example

- 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

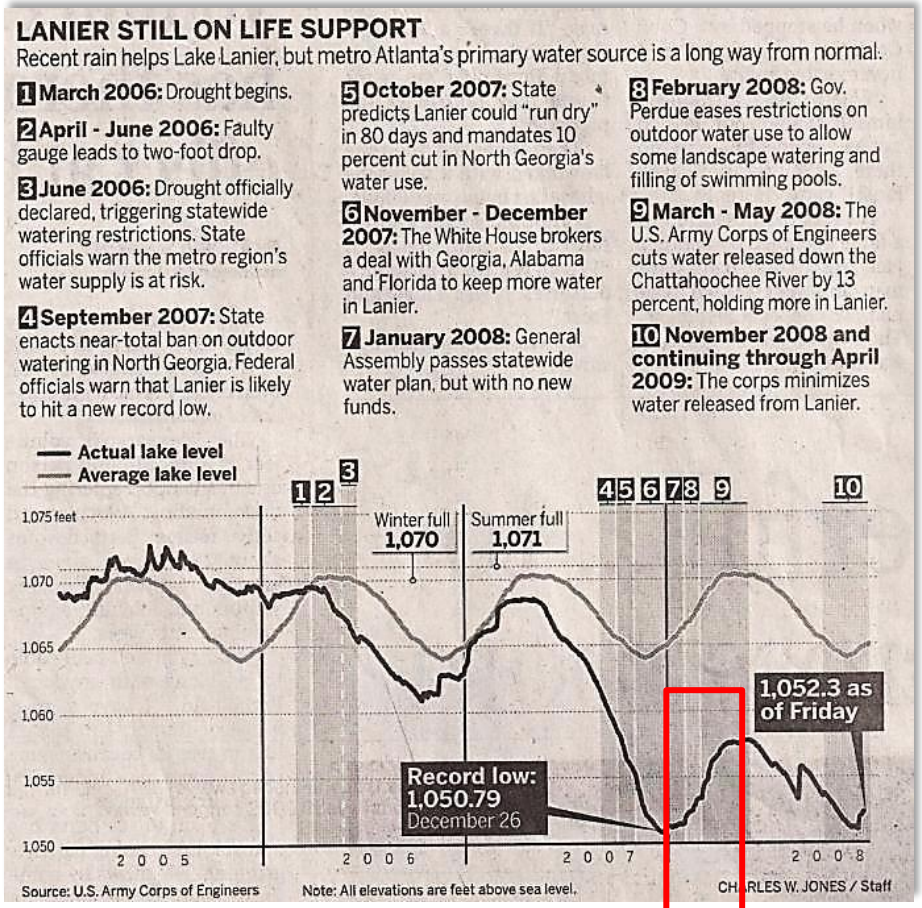
Lie factor: Example

- Lie factor =

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

= 14 / 0.0057088

=2452.3



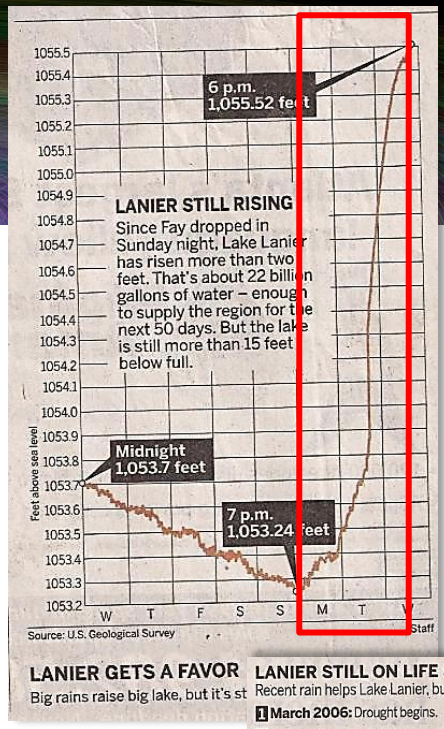
Atlanta Journal Constitution
Summer '08

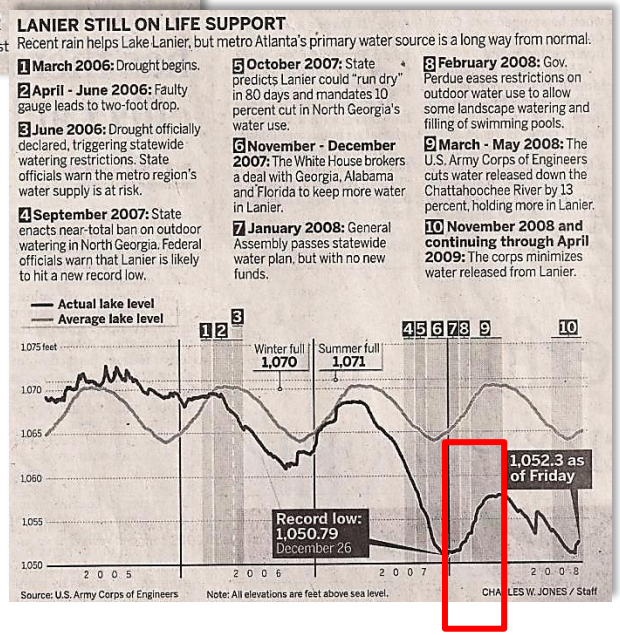
Lie factor: Example

- Lie factor difference

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

= **8.59**





Atlanta Journal Constitution
Summer '08

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

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

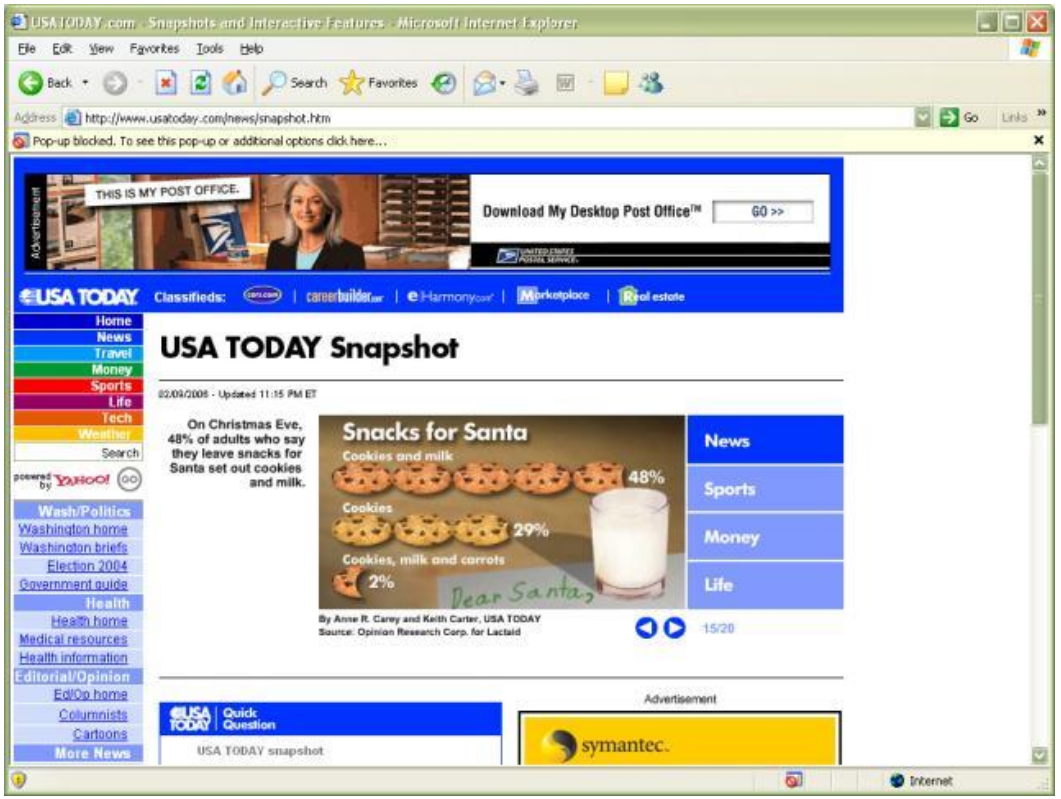
"Data-rich designs give a context and credibility to statistical evidence. Low-information designs are suspect: what is left out, what is hidden, why are we shown so little? High-density graphics help us to compare parts of the data by displaying much information within the view of the eye: we look at 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."

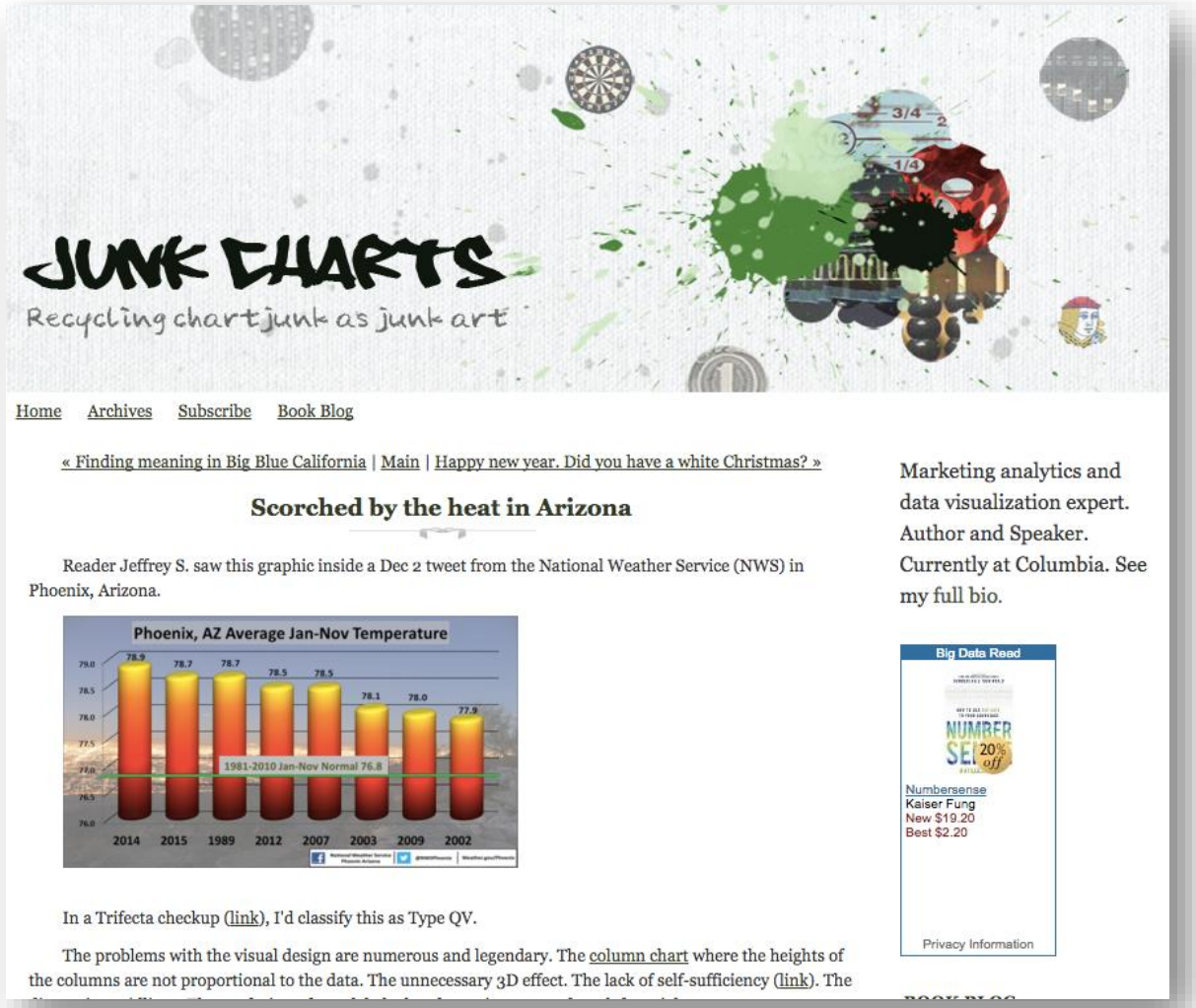
Vol 1, p 168

Design Principles

3. Avoid chart junk
   - Extraneous visual elements that detract from message



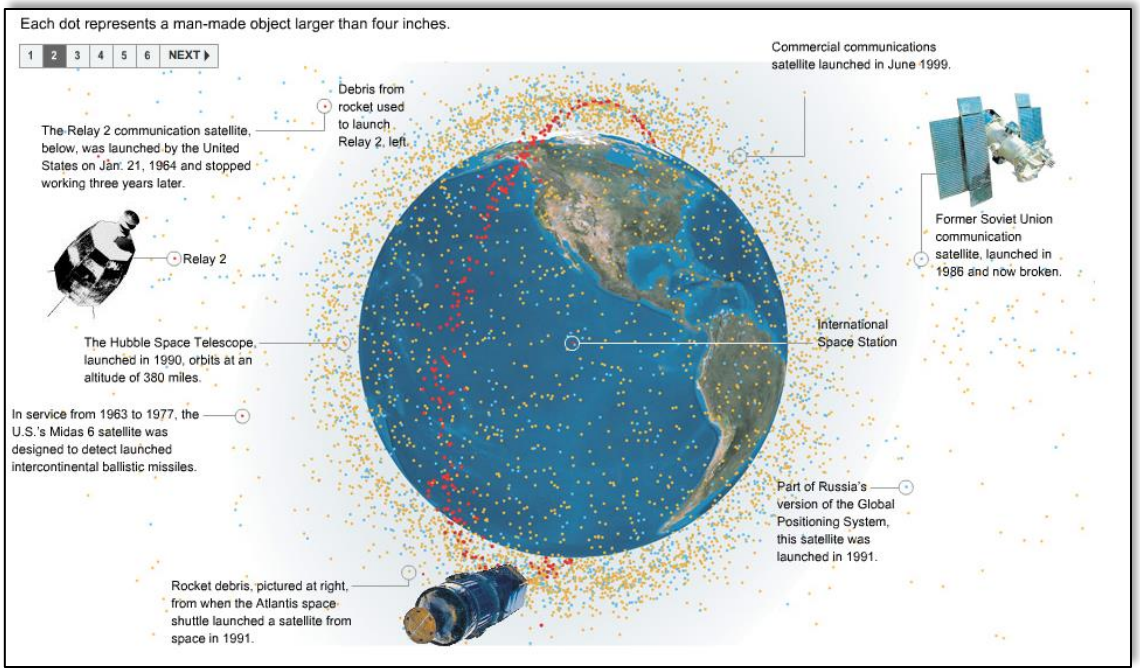
Junk charts blog



http://junkcharts.typepad.com

Design Principles

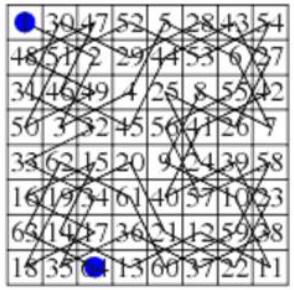
4. Utilize multifunctioning graphical elements (macro/micro readings)
   - Graphical elements that convey data information and a design function
   - Overview with high levels of detais



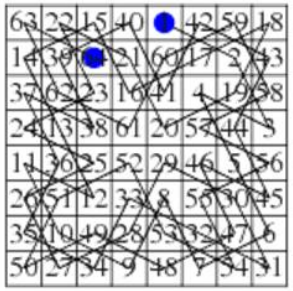
Design Principles

5. Use small multiples
   - Repeat visually similar graphical elements nearby rather than spreading far apart

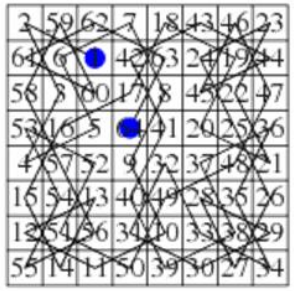
Beverly 1848, Σ = 282, 210



Jaenisch 1862, Σ = 348, 168



Francony 1882, Σ = 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

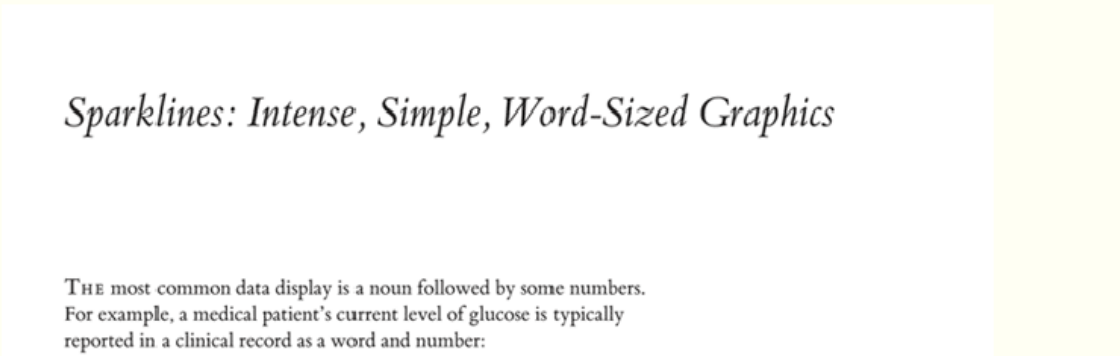
Theory of sparklines (small, intense, simple datawords) along with many practical examples of recent sparkline developments. From Edward Tufte's book *Beautiful Evidence*.

-- 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:

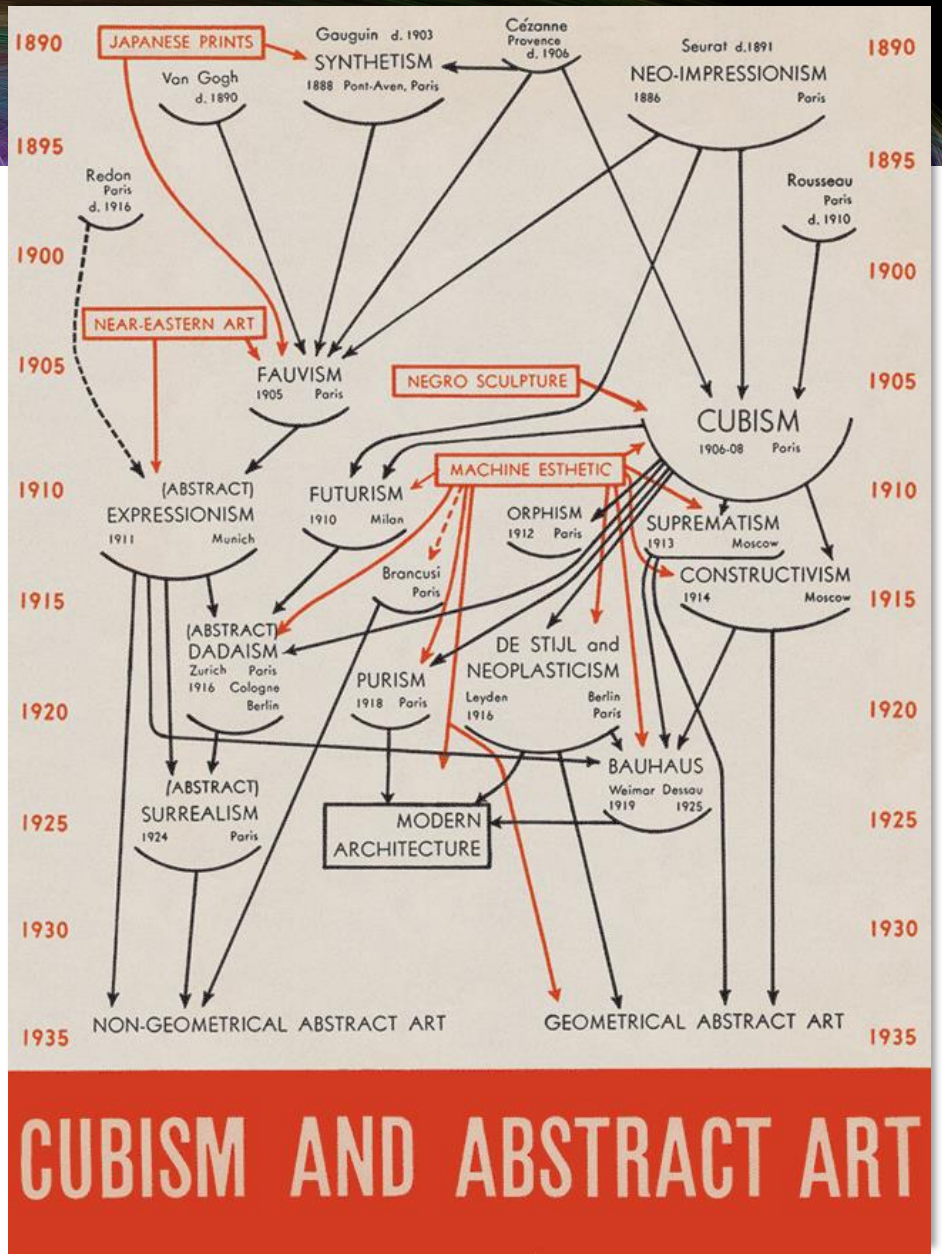
Sparklines Example

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:

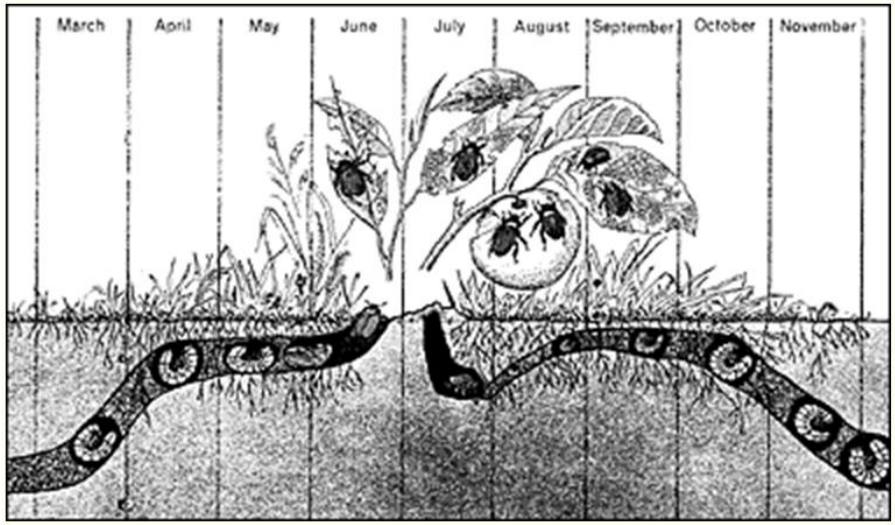
Design Principles

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

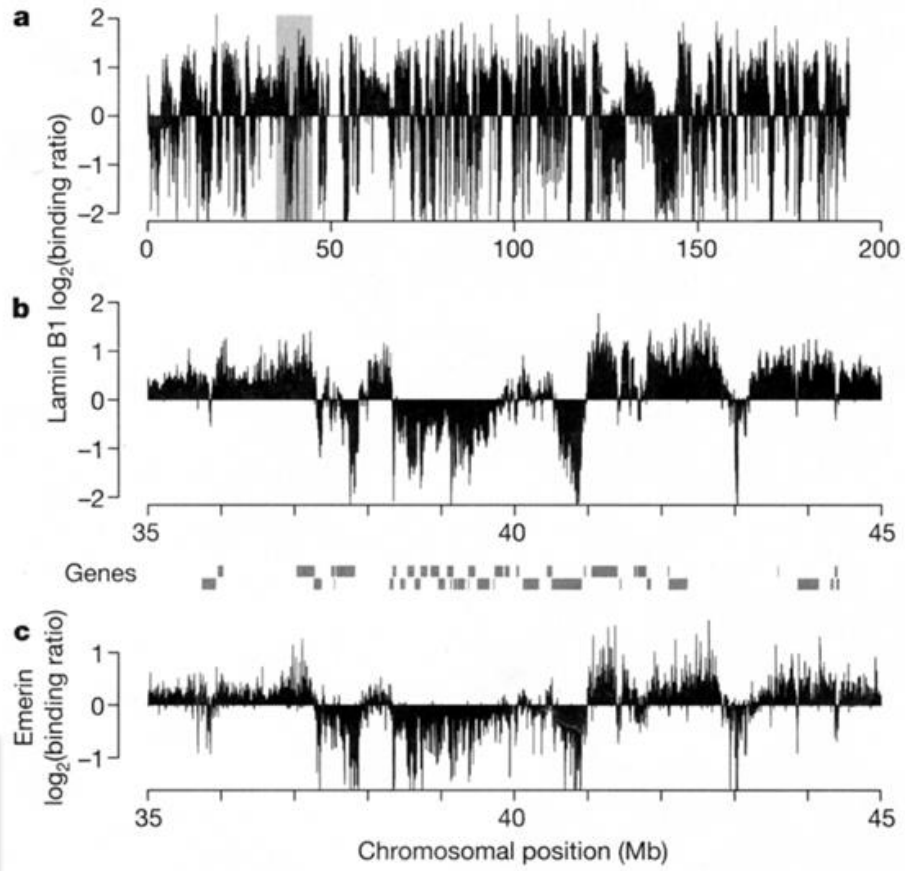


Figure 1 | Mapping of genome–NL interactions in human fibroblasts. **a**, DamID map of lamin B1 interactions for the whole of chromosome 4. **b**, Detailed view of a 10-Mb region (grey box in **a**). **c**, Map of emerin interactions in the same region as in **b**. Genes are indicated by grey rectangles between **b** and **c**.

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

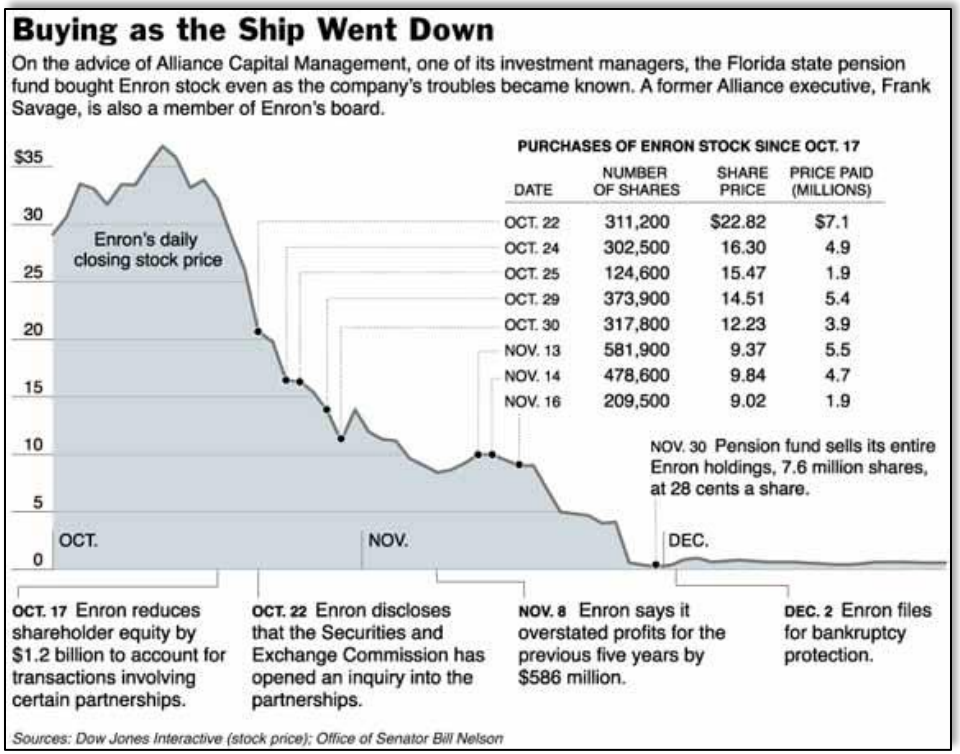
LETTERS

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

Lars Guelen[1], Ludo Pagie[1], Emilie Brasset[2], Wouter Meuleman[1,4], Marius B. Faza[1], Wendy Talhout[1], Bert H. Eussen[3], Annelies de Klein[3], Lodewyk Wessels[1,4], Wouter de Laat[2] & Bas van Steensel[1]

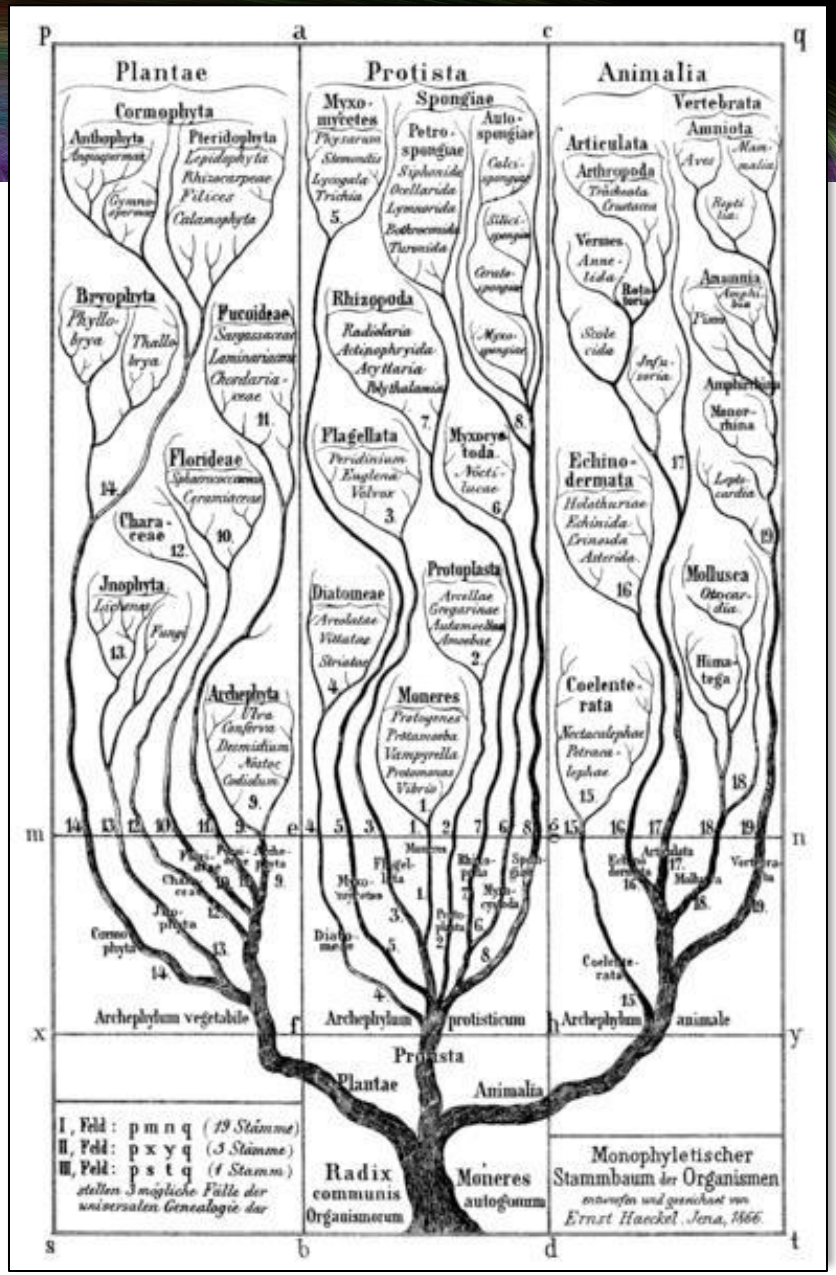
Design Principles

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

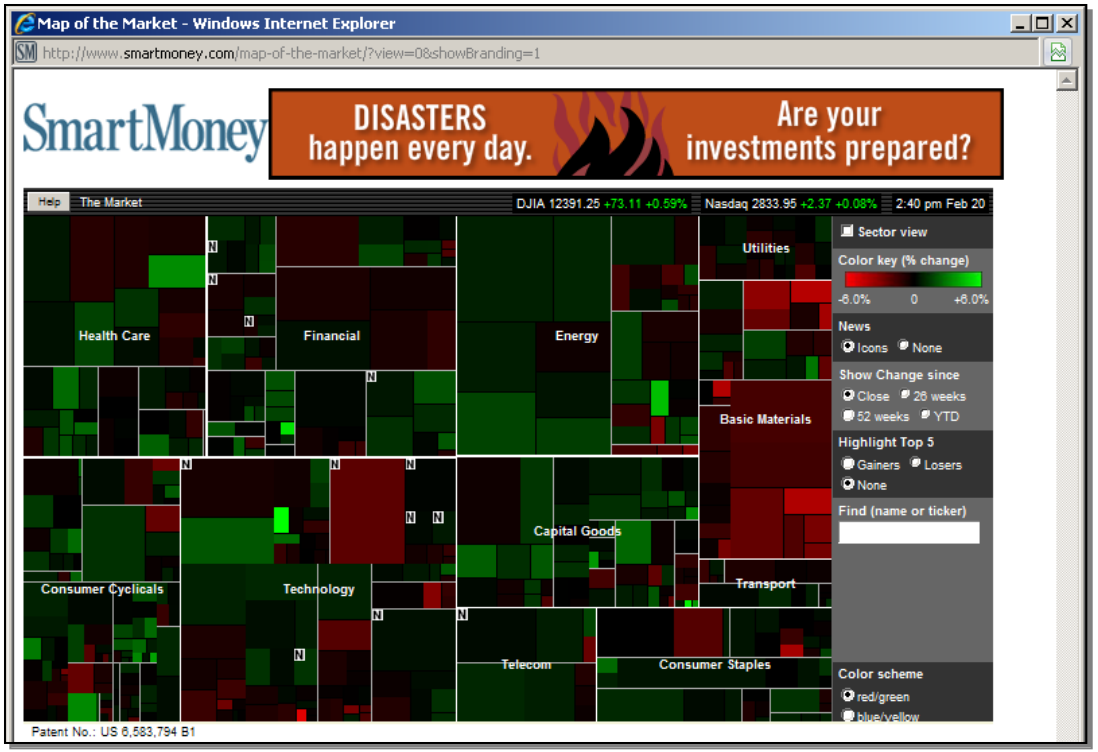
1. Tell the truth
   - Graphical integrity
2. Do it effectively with clarity, precision…
   - Design aesthetics

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

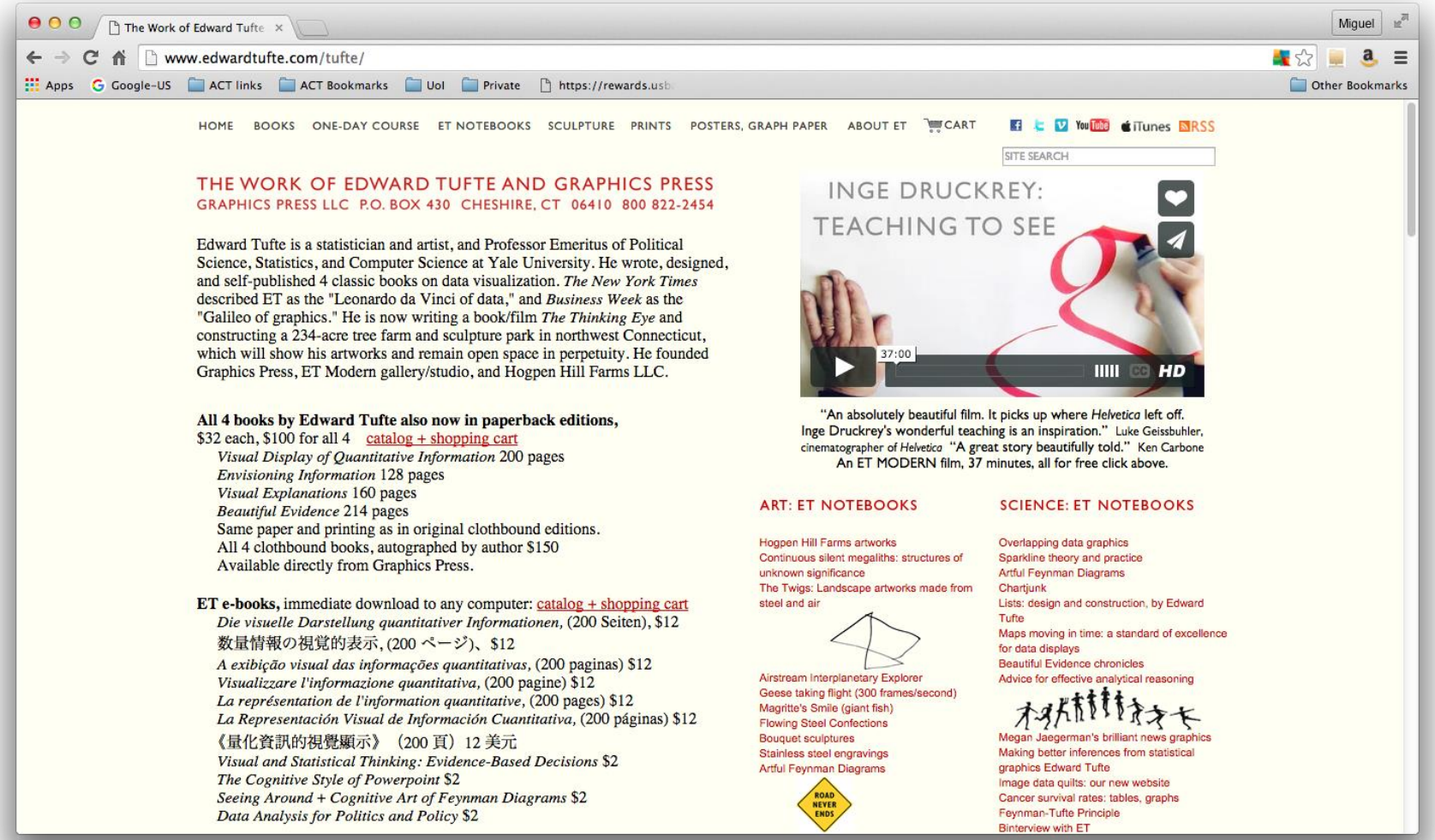
- to Label
- to Measure
- to Represent or imitate reality
- to Enliven or decorate



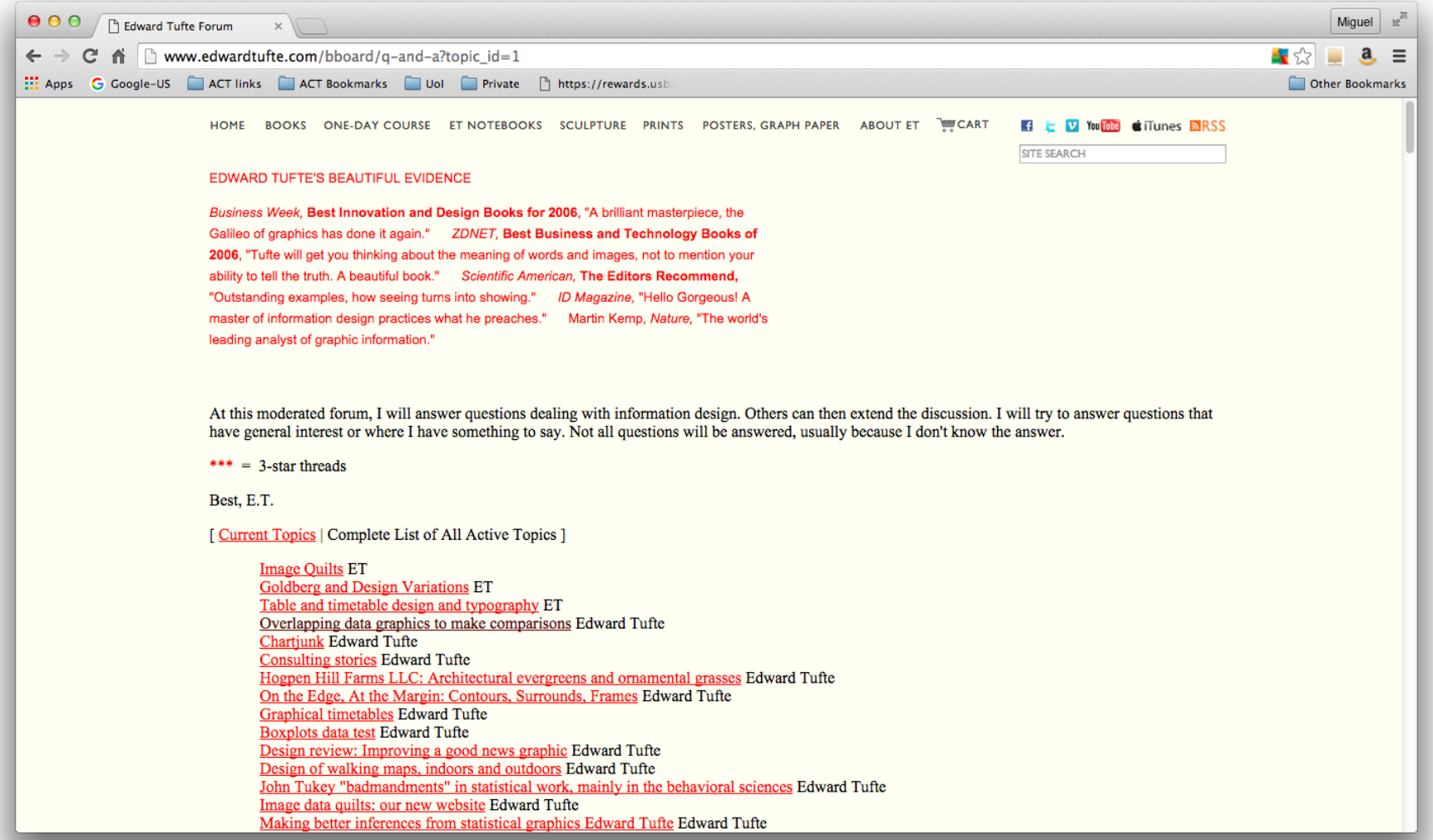
Graphical Displays should …

- 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

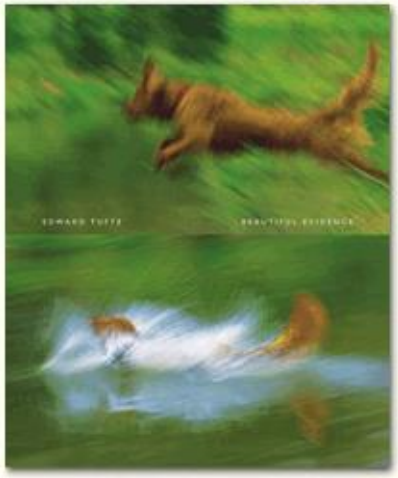


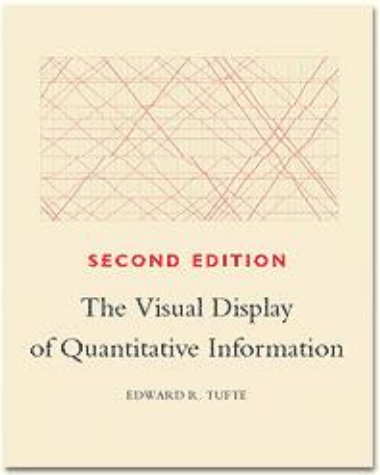
Tufte's discussion forum (ET Notebooks)

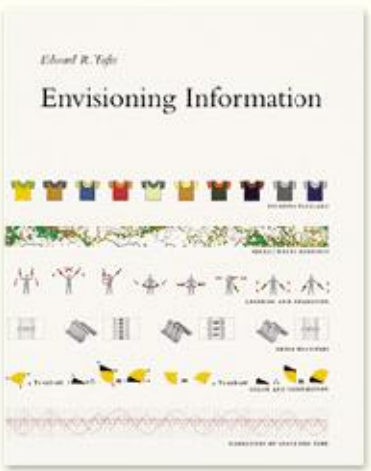


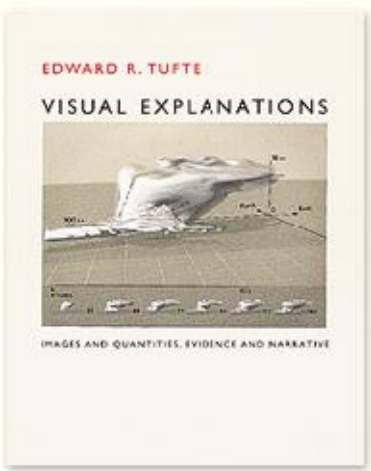
Additional Readings

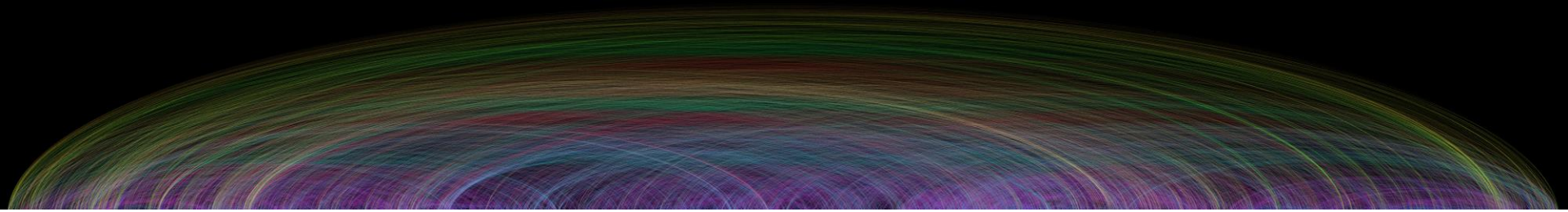
- Tufte, *Envisioning Information*.









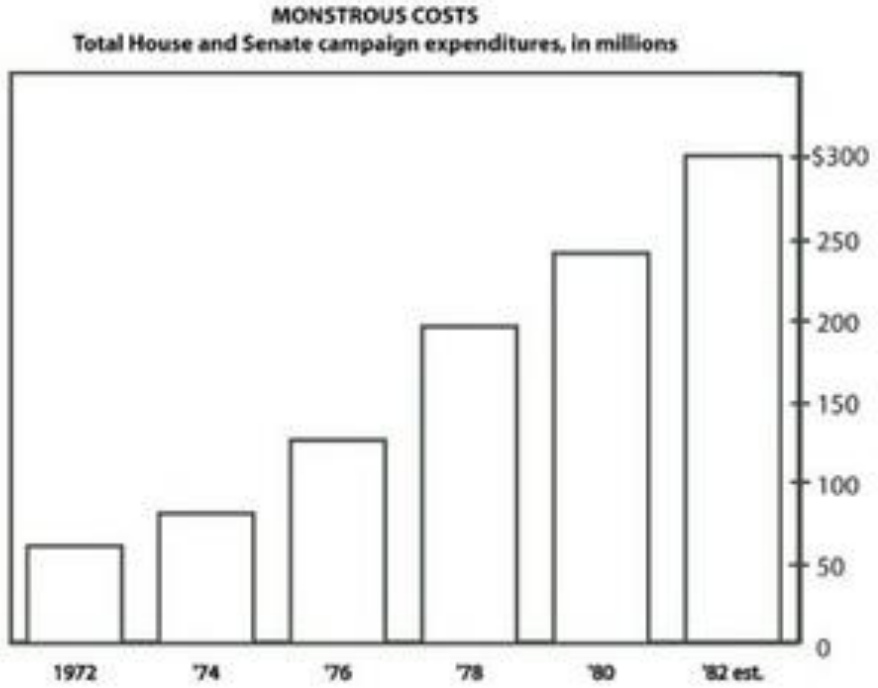


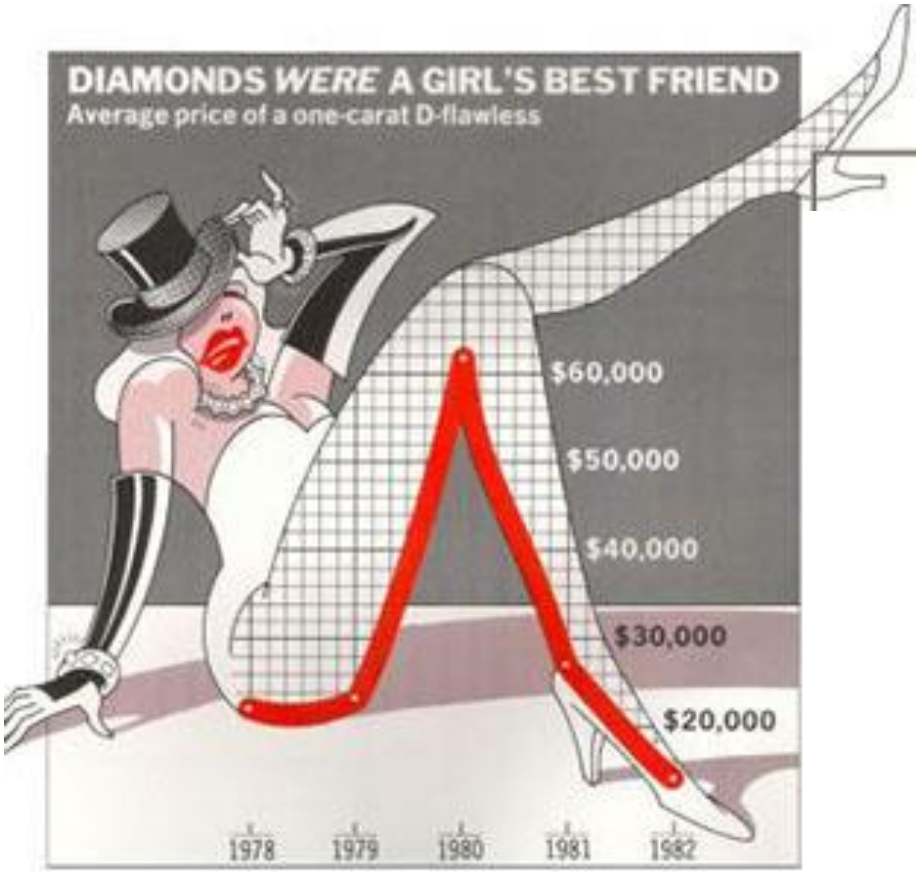
When More Is Better

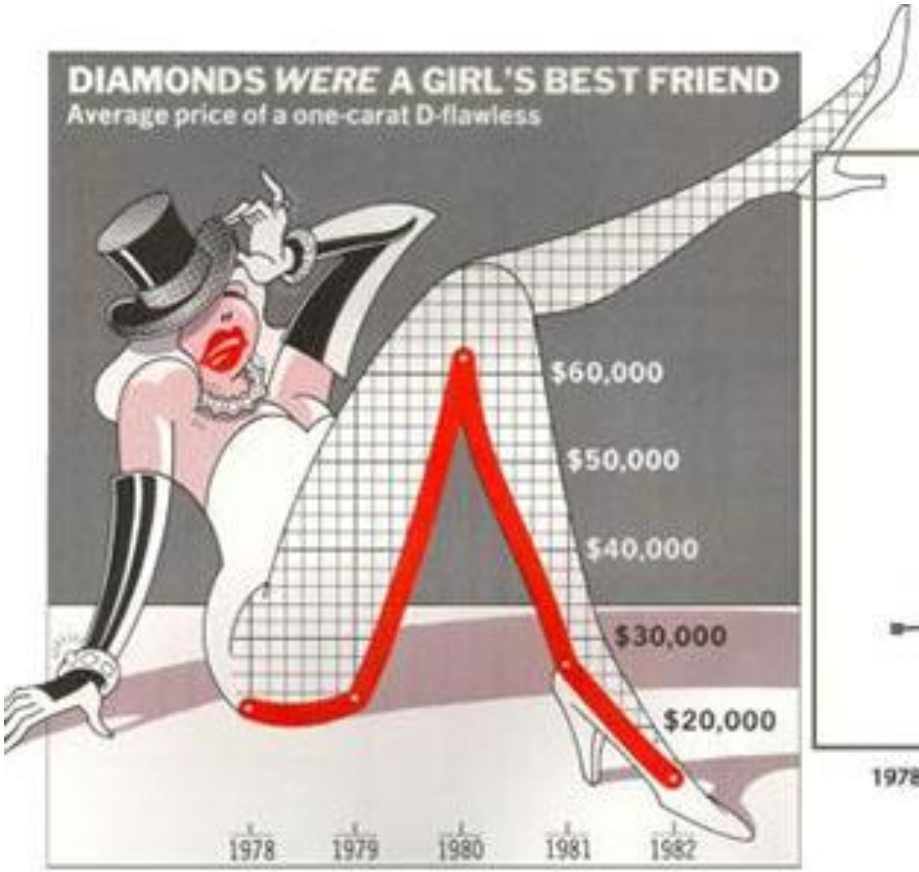
THE FLIPSIDE

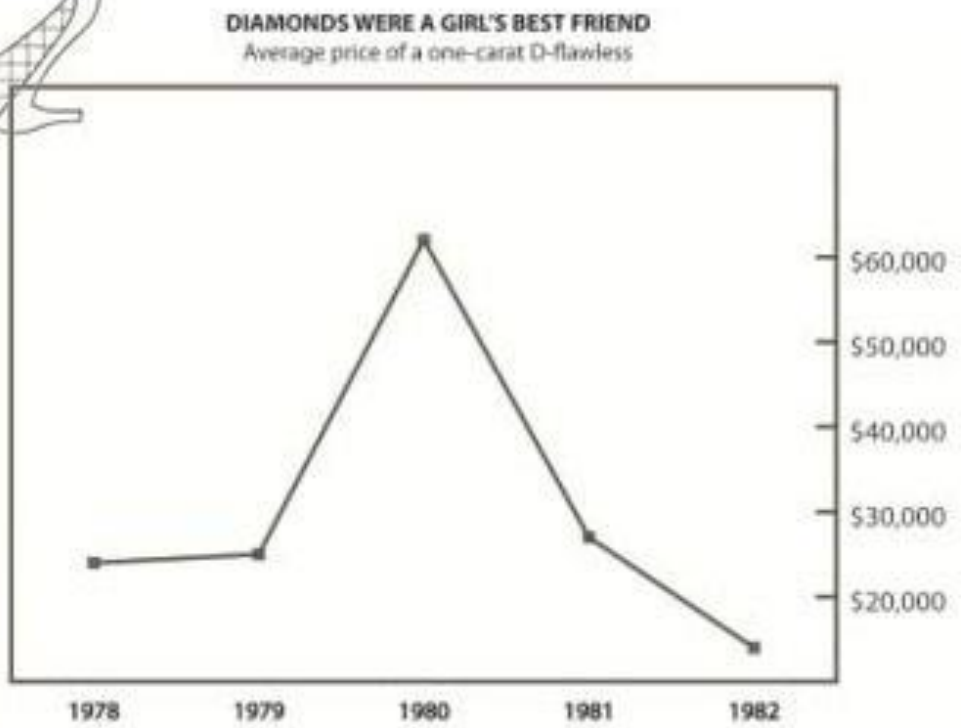
www.nigelholmes.com


MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions
$300
250
200
150
100
50
1972
'74
'76
'78
'80
'82 est.
VOTE

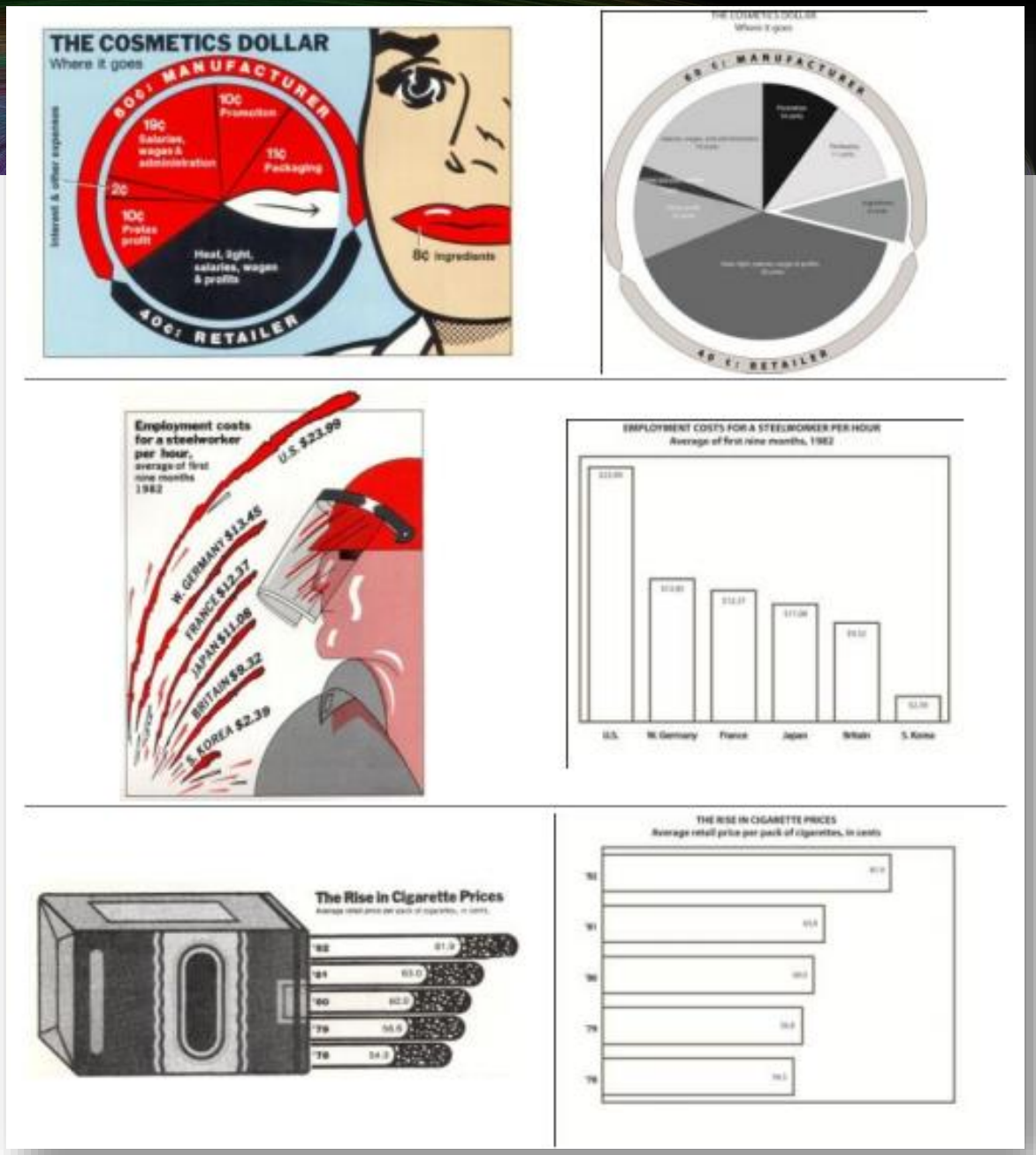

MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions
$300
250
200
150
100
50
0
1972
'74
'76
'78
'80
'82 est.


DIAMONDS WERE A GIRL'S BEST FRIEND
Average price of a one-carat D-flawless
$60,000
$50,000
$40,000
$30,000
$20,000
1978
1979
1980
1981
1982


DIAMONDS *WERE* A GIRL'S BEST FRIEND
Average price of a one-carat D-flawless
$60,000
$50,000
$40,000
$30,000
$20,000
1978
1979
1980
1981
1982


DIAMONDS WERE A GIRL'S BEST FRIEND
Average price of a one-carat D-flawless
$60,000
$50,000
$40,000
$30,000
$20,000
1978
1979
1980
1981
1982

Experiment


THE COSMETICS DOLLAR
Where it goes
60¢: MANUFACTURER
10C Promotion
19C Salaries, wages & administration
11C Packaging
2C
10C Profits profit
Heat, light, salaries, wages & profits
8C Ingredients
40¢: RETAILER
Employment costs for a steelworker per hour, average of first nine months 1982
U.S. $23.99
W. GERMANY $13.45
FRANCE $12.37
JAPAN $11.06
BRITAIN $9.32
S. KOREA $2.38
EMPLOYMENT COSTS FOR A STEELWORKER PER HOUR
Average of first nine months, 1982
U.S.
W. Germany
France
Japan
Britain
S. Korea
The Rise in Cigarette Prices
'82
'81
'80
'79
'78
THE RISE IN CIGARETTE PRICES
Average retail price per pack of cigarettes, in cents

Gaze tracking

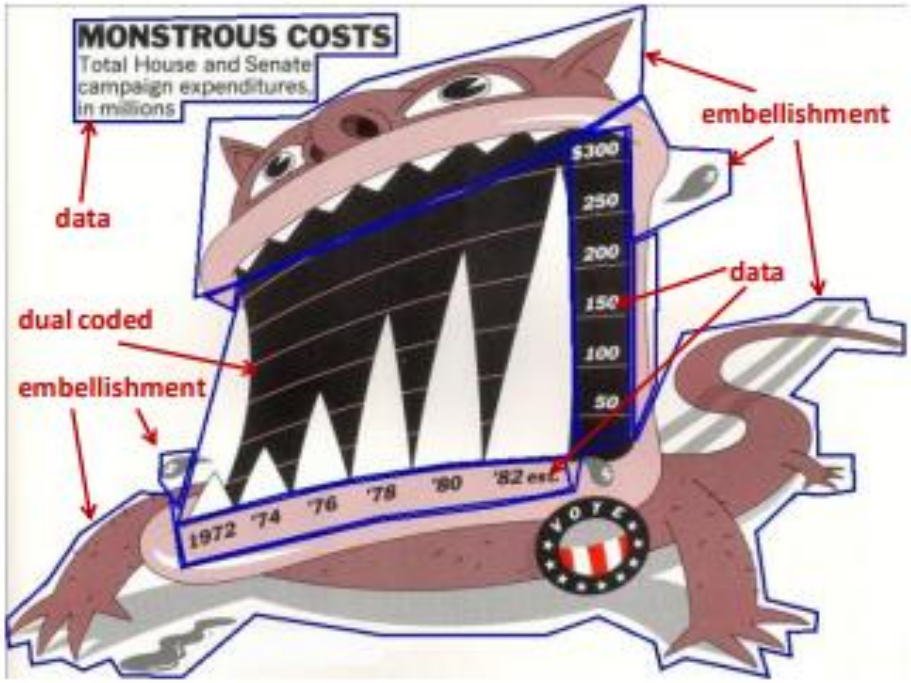


Figure 3. Area of interest analysis: gaze falling in any of the areas defined by the blue borders were labeled as data, embellishment, dual encoded, or other.

Results

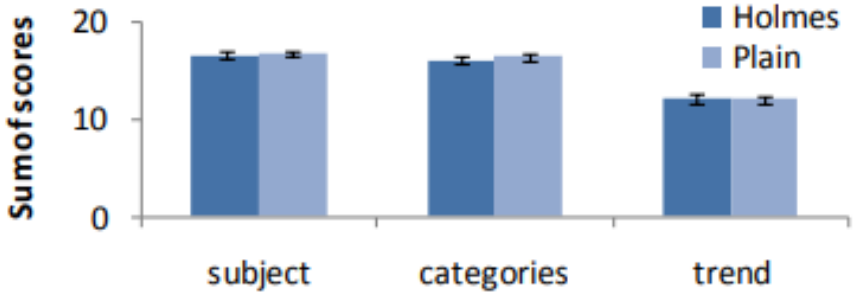


Figure 4. Means ± SE for description scores.

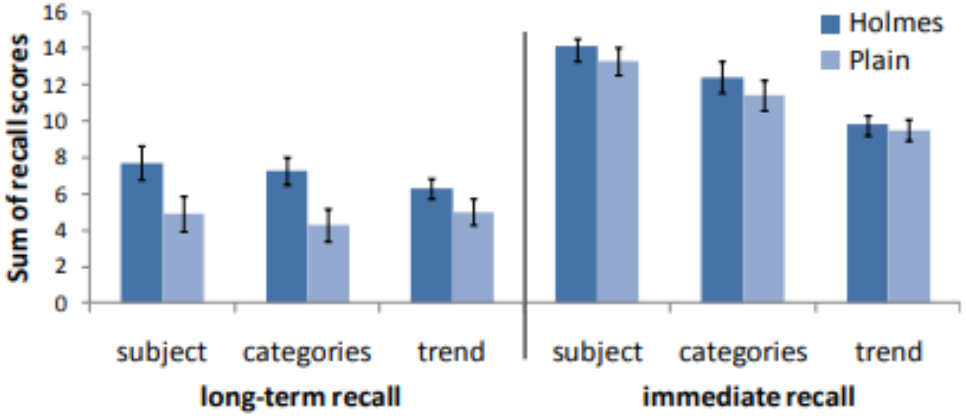


Figure 5. Means ±S E for recall scores for long-term and immediate recall.

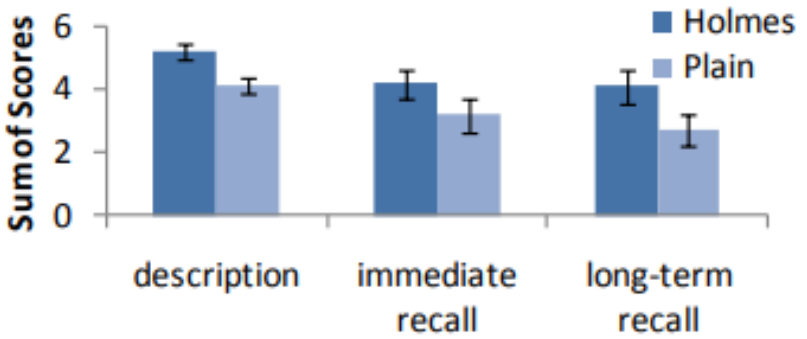
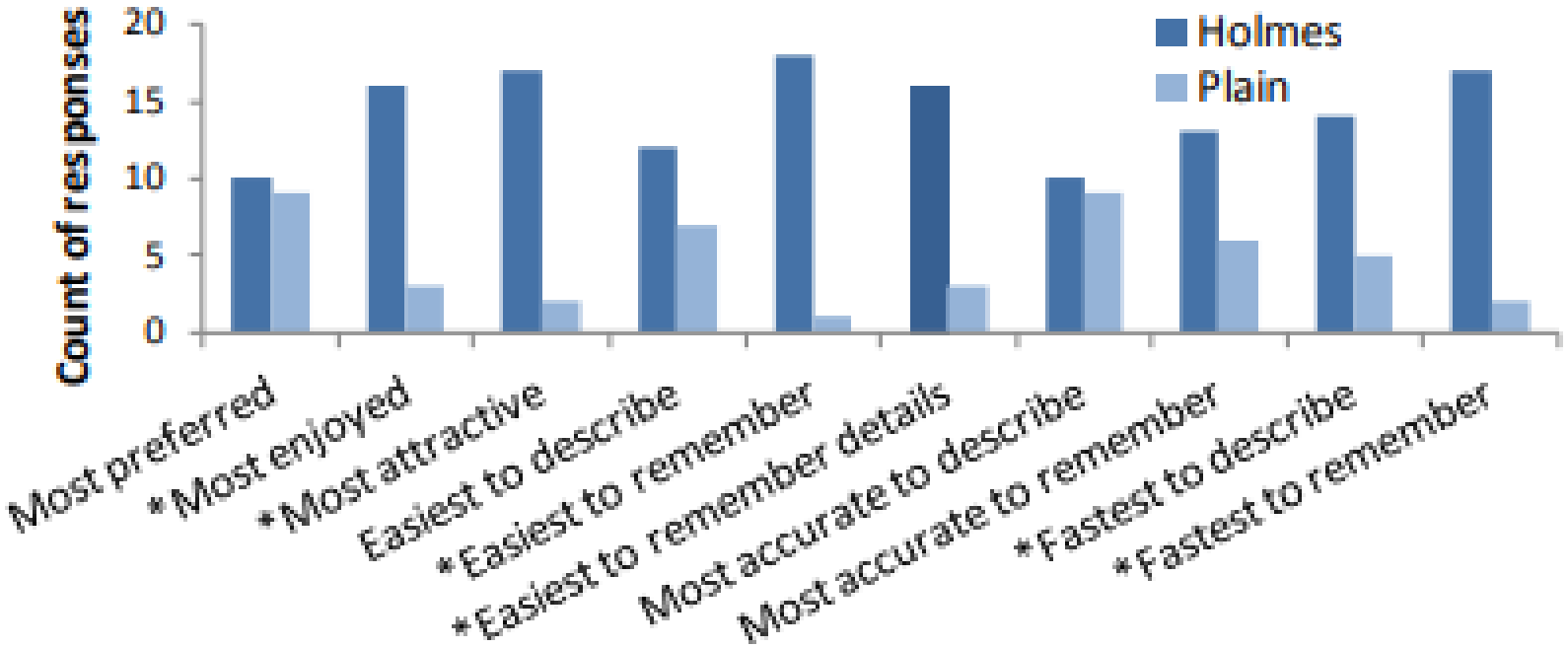


Figure 6. Means ± SE for sum of value message scores.


Count of responses
20
15
10
5
0
Holmes
Plain
Most preferred
*Most enjoyed
*Most attractive
Easiest to describe
*Easiest to remember
*Easiest to remember details
Most accurate to describe
Most accurate to remember
*Fastest to describe
*Fastest to remember

- 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.