

Tufte's Design Principles



CS 7450 - Information Visualization
February 22, 2011
John Stasko

Please see appropriate books for missing images

Today's Agenda

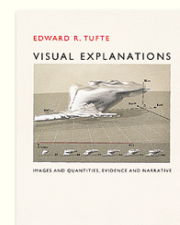
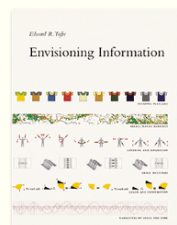
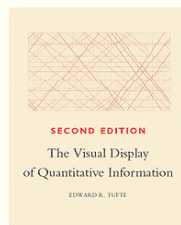


Edward Tufte has written seven books, including *Visual Explanations*, *Envisioning Information*, *The Visual Display of Quantitative Information*, and *Data Analysis for Politics and Policy*. He writes, designs, and self-publishes his books on analytical design, which have received more than 40 awards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

For a moderated forum on analytical design, go to [ASK E.T.](http://www.edwardtufte.com/ask-et/)

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

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, flip 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

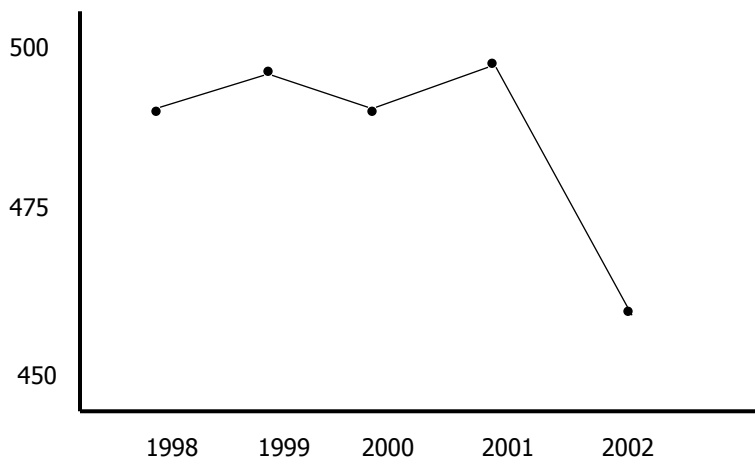
Spring 2011

CS 7450

7

Example

Stock market crash?

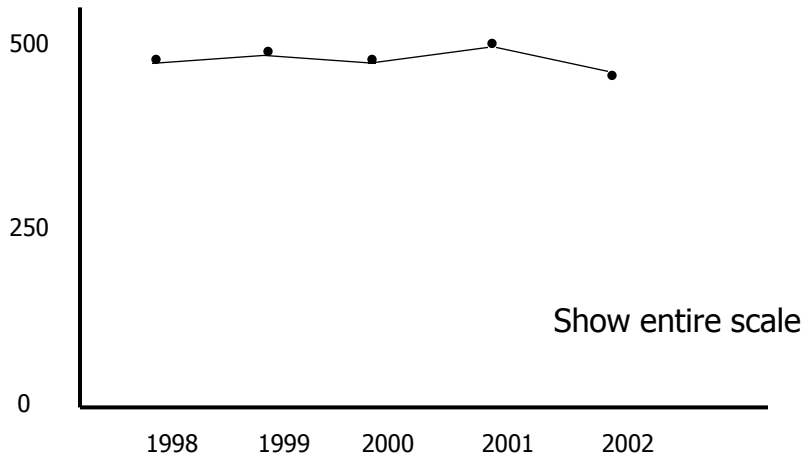


Spring 2011

CS 7450

8

Example

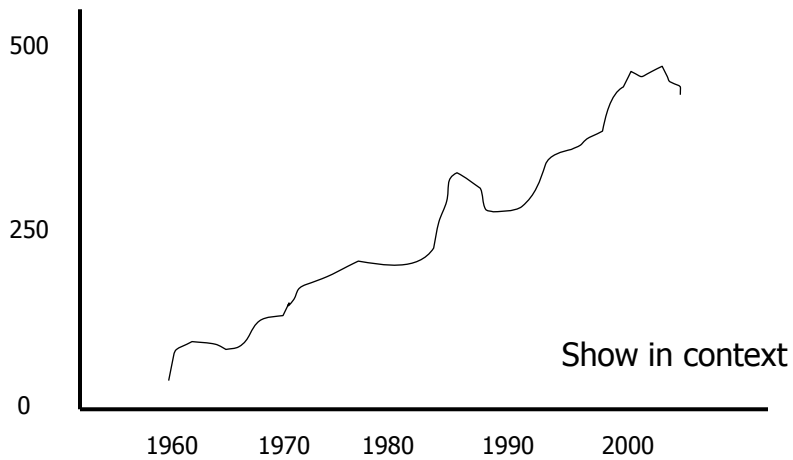


Spring 2011

CS 7450

9

Example



Spring 2011

CS 7450

10

Chart Integrity



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

Vol 1, p. 54 (1)

Where's 0?
Note middle '70



Vol 1, p 54 (2)



What's being compared?

Spring 2011

CS 7450

13

Vol 1, 57



Scale?

Spring 2011

CS 7450

14

Vol 1, p. 61



Scale?

Spring 2011

CS 7450

15

Vol 1, p. 74



Great work!

Spring 2011

CS 7450

16

Vol 1, p. 74



Ahhhh

Show the context

Spring 2011

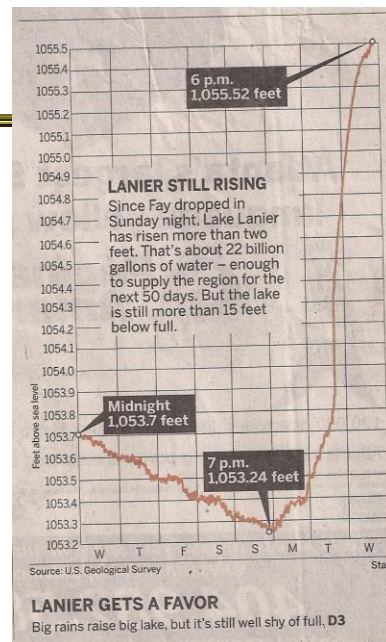
CS 7450

17

Local Example

A huge rise?

Atlanta Journal Constitution
Summer '08



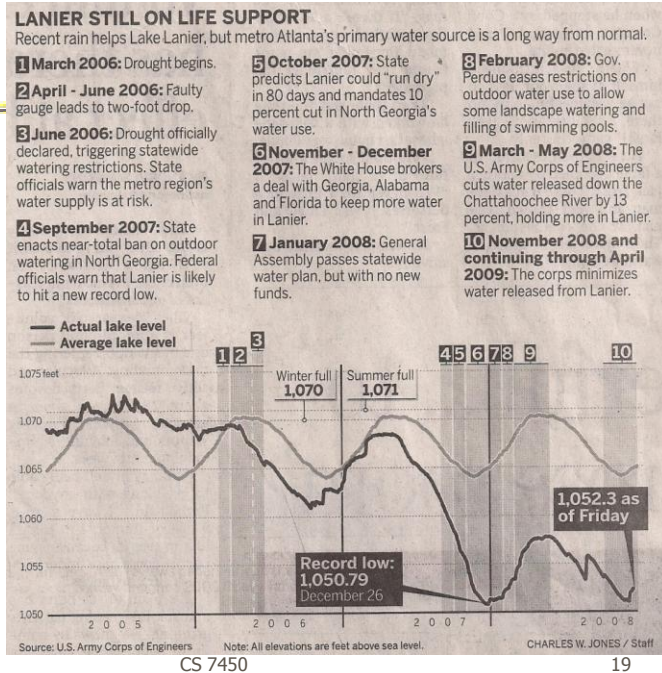
Spring 2011

CS 7450

18

More of the
data

Atlanta Journal
Constitution
Dec. '08



Spring 2011

CS 7450

19

Watch Size Coding



- Height/width vs. area vs. volume

Spring 2011

CS 7450

20

Vol 1, p. 69



area = value?

Spring 2011

CS 7450

21

Vol 1, p. 62



volume = value?

Spring 2011

CS 7450

22

Measuring Misrepresentation



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

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

$$\text{p.62} \quad 9.4 = \frac{4280}{454}$$

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

Vol 1, p. 30



Outstanding

Spring 2011

CS 7450

27

More...



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

Spring 2011

CS 7450

28

More...



- Maximize data density

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

Quote ...

Spring 2011

CS 7450

29

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

Spring 2011

CS 7450

30

Redesign charts



- Bar chart, scatter plot, box plot
(See drawings)

Design Principles



- **Avoid chartjunk**
 - Extraneous visual elements that detract from message

Vol 1, p 108



Spring 2011

CS 7450

33

Vol 2, p.34



A classic

Diamonds Were A
Girl's Best Friend

Spring 2011

CS 7450

34

USA Today

http://www.usatoday.com/news/snapshot.htm



Spring 2011

CS 7450

35

Junk Charts Blog

http://junkcharts.typepad.com/



Spring 2011

CS 7450

36

More Thoughts



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

Spring 2011

CS 7450

37

Rethink That?



CHI 2010, Graphics April 10-15, 2010, Atlanta, GA, USA

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Gersz, David McDine, Christopher Brooks
Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca, aaron.gersz@usask.ca, dmandr@jmail.usask.ca, c895@jmail.usask.ca

ABSTRACT
Conditions for designing information charts often stress that the presentation should reduce chart junk – visual embellishments that are not essential to understanding the data. In contrast, some graphic chart designers argue the presence of data is detested and obscures memory, making the elements of a chart less likely to be remembered or understood to have been presented, and whether the visual embellishments may have other benefits. To investigate these issues, we conducted an experiment that compared comprehension accuracy and long-term recall. We found that people's accuracy in describing the embellished charts was no worse than the plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, we strongly question some of the premises of the minimalist approach to chart design.

Author Keywords:
Charts, information visualization, imagery, memorability.

ACM Classification Keywords:
H.1.1. Information interfaces and presentation (e.g., HCI); I.2.1. Visualization.

General Terms:
Design, Human Factors

INTRODUCTION

Many experts in the area of chart design, such as Edward Tufte, criticize the inclusion of visual embellishments in charts and graphs, claim guidelines for good chart design often suggest that the addition of chart junk obscures and other kinds of non-essential imagery, in a chart can make comprehension more difficult and can obscure readers from the data [23]. This minimalist perspective advocates plain and simple charts that maximize the proportion of data ink – or the ink to the chart used to represent data.

However, in a study reported in this paper, we found that people who viewed a given chart in a plain style (no chart junk) were no more accurate in describing the data than those who viewed the same chart in an embellished style. In fact, people who viewed the same chart in an embellished style were more accurate in describing the data than those who viewed the same chart in a plain style. This result is surprising because it suggests that the inclusion of chart junk may be beneficial to comprehension and recall of chart data.

CHI 2010, April 10-15, 2010, Atlanta, Georgia, USA.
Copyright 2010 ACM 978-1-60558-335-6/10/04...\$5.00.

Designs from minimalist publications, many designers include a wide variety of visual embellishments in their charts, from small decorative to large images and visual backgrounds. One well-known proponent of visual embellishments in charts is the graphic artist Nigel Holmes, whose work requires comprehension using visual imagery on the basis of the chart [11] (e.g., Figure 1),

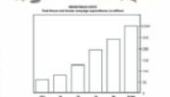


Figure 1. A chart by Holmes [11] (embellished) and a plain version. These kinds of charts appear regularly in many non-media publications, and the widespread use of embellished designs raises questions about whether the minimalist position on chart design is really the better approach. Two issues in particular are raised: first, whether visual embellishments do in fact cause comprehension problems, and second, whether the embellishments may provide additional information that is valuable for the reader. For example, the added visual imagery in a Holmes-style chart could draw

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

Spring 2011

CS 7450

38

Design Principles



- Utilize multifunctioning graphical elements (macro/micro readings)
 - Graphical elements that convey data information and a design function

Vol 1, p 140



Vol 1, p. 141



US Army Divisions
going to France in
WW I

Leonard P. Ayres
The War with Germany
1919

Spring 2011

CS 7450

41

Vol 2, p. 36



Michel E. Turgot
Louis Bretz

Plan de Paris
1739

Spring 2011

CS 7450

42

Vol 2, p. 37

Manhattan 1989
Manhattan Map Company



Spring 2011

CS 7450

43

Vol 2, p. 42



Viet Nam Memorial
in Washington D.C.

Maya Ying Lin

58,000+ dead soldiers

Spring 2011

CS 7450

44

Vol 2, p. 44



Spring 2011

CS 7450

45

Vol 2, p. 43



Names listed
chronologically by death

Spring 2011

CS 7450

46

Design Principles



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

Vol 1, p. 170



23 hours of
LA air pollution

Vol 1, p. 173



Chromosomes of
man, chimpanzee,
gorilla & orangutan

Spring 2011

CS 7450

49

Vol 1, p. 174



Consumer
Reports

Spring 2011

CS 7450

50

Vol 2, p. 68



NY Trains

Spring 2011

CS 7450

51

Vol 2, p. 68



How to draw letters

Spring 2011

CS 7450

52

Vol 2, p. 69



Calligraphy

Spring 2011

CS 7450

53

More Recent Additions



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

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. 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:

Spring 2011

CS 7450

54

Sparkline Examples



Spring 2011

CS 7450

55

Design Principles



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

Space shuttle case we discussed first day

Spring 2011

CS 7450

56

Vol 3, p. 144



Washington Post

Spring 2011

CS 7450

57

Design Principles



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

Spring 2011

CS 7450

58

Vol 2, p. 12



Guide for visitors to Ise Shrine, Japan

Spring 2011

CS 7450

59

Vol 2, p. 24



Timetable for Java railroad line

Spring 2011

CS 7450

60

Vol 3, p. 90

Music history



Steve Chapple and Reebee Garofalo

Spring 2011

CS 7450

61

Design Principles



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

Spring 2011

CS 7450

62

Vol 2, p. 54

IBM Series III Copier



Spring 2011

CS 7450

63

Vol 2, p. 61



Spring 2011

CS 7450

64

Design Principles



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

Spring 2011

CS 7450

65

Vol 1, p.43 & Vol 2, p 110



Life of a beetle

L. Hugh Newman

Spring 2011

CS 7450

66

Vol 2, p. 102



Czech air schedule

Spring 2011

CS 7450

67

Vol 2, p. 103



China railway
timetable

Spring 2011

CS 7450

68

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

Spring 2011

CS 7450

71

Vol 2, p. 63



Before

After

Spring 2011

CS 7450

72

Vol 3, p. 74



Before

After

Spring 2011

CS 7450

73

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.*”

Spring 2011

CS 7450

74

Proper Color Use



- To label
- To measure
- To represent or imitate reality
- To enliven or decorate

Examples



- The bad...

Vol 1, p. 153



Spring 2011

CS 7450

77

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..."

Spring 2011

CS 7450

78

Vol 2, p. 82



Spring 2011

CS 7450

79

Vol 2, p. 88



“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.”

Spring 2011

CS 7450

80

Vol 3, p. 77



Spring 2011

CS 7450

81

Examples



- The good...

Spring 2011

CS 7450

82

Vol 2, p. 91 & Vol 3, p. 76



Spring 2011

CS 7450

83

Vol 2, p. 80



Swiss Mountain Map

Spring 2011

CS 7450

84

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
direction

latitude
longitude

temperature
date

Spring 2011

CS 7450

87

Graphical Displays Should



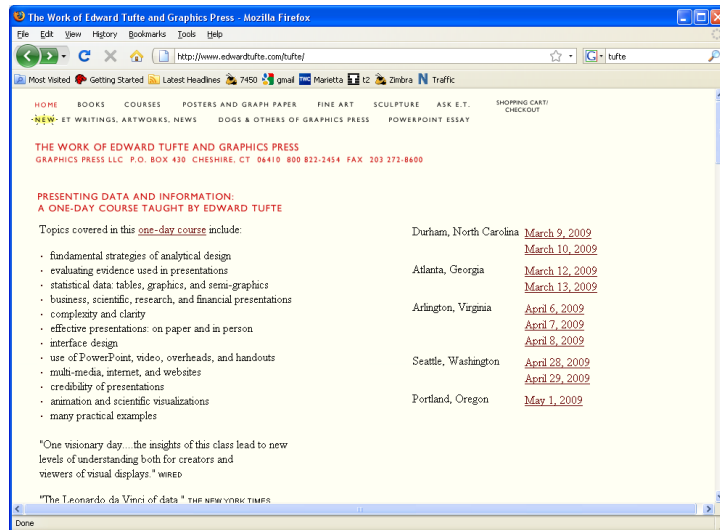
- Show the data
- Induce the viewer to think about substance rather than about methodology, graphic design the technology of graphic production, or something else
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets coherent
- Encourage the eye to 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

Spring 2011

CS 7450

88

Website & Seminar

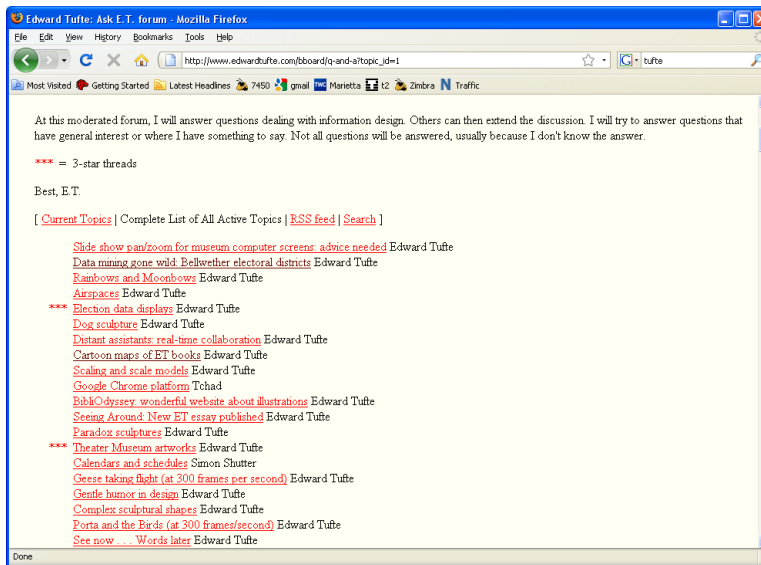


Spring 2011

CS 7450

89

Discussion Forum



Spring 2011

CS 7450

90

Upcoming



- Few's Design Principles
 - Reading
 - Few chapters 5-12

- Hierarchical Data (Node-link reps)
 - Reading
 - Card & Nation

Sources Used



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