1. Information Visualization

Introducing Research Need & Concepts

Dr. Thorsten Büring, 18. Oktober 2007, Vorlesung Wintersemester 2007/08

Outline

Literature

- R. Spence: "Information Visualization: Design for Interaction", 2. Auflage, 2006.

- S. Card, J. Mackinlay, B. Shneiderman: "Readings in Information Visualization - Using Vision to Think", 1999.

- B. Bederson, B. Shneiderman: "The Craft of Information Visualization - Readings and Reflections". 2003.

- E. R. Tufte: "The Visual Display of Quantitative Information", 2. Auflage, 2001.
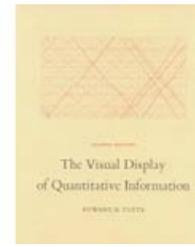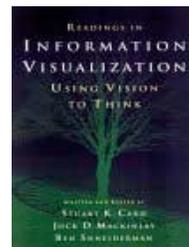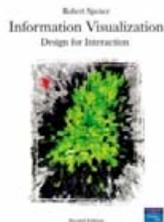
Data Explosion

- Progress in hardware technology allows computers to store an increasing amount of large data

  - "640K ought to be enough for anybody" (Bill Gates, 1981(?))

  - To be fair: "I never said that statement — I said the opposite of that." (Bill Gates, 2001)

- Computers and the Internet let people consume and produce vast amounts of data

- 2002 Study by University of Berkeley (http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/)

  - In 2002, print, film, magnetic, and optical storage media produced about **5 exabytes**

  - 1 exabyte = 1 Million Terabyte = 7500 * the information contained in the Library of Congress (seventeen million books)

  - 92% of new information is stored on magnetic media, primarily hard disks

  - Almost 800 MB of recorded information is produced per person each year

  - New stored information grew about 30% a year between 1999 and 2002

Data Overload

▤ Data, not information!

▤ Main principle of Information Visualization: allow information to be derived from data

▤ How to transfer information to the user?

▤ Use human vision

   ▤ Provides highest bandwidth sense: human retina can transmit data at roughly 10 million bits per second (Koch et al., 2006) (Ethernet connection: 10 to 100 million bits per second)

   ▤ Pattern recognition

   ▤ Pre-attentive

   ▤ Extends memory and cognitive capacity

   ▤ People think visually
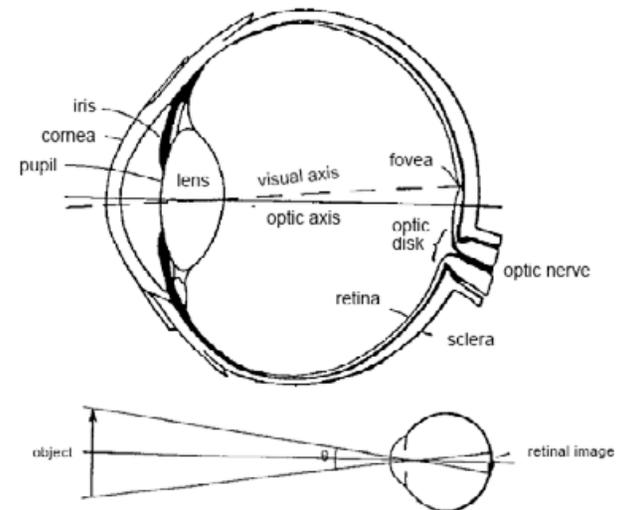
Visualization Example 1
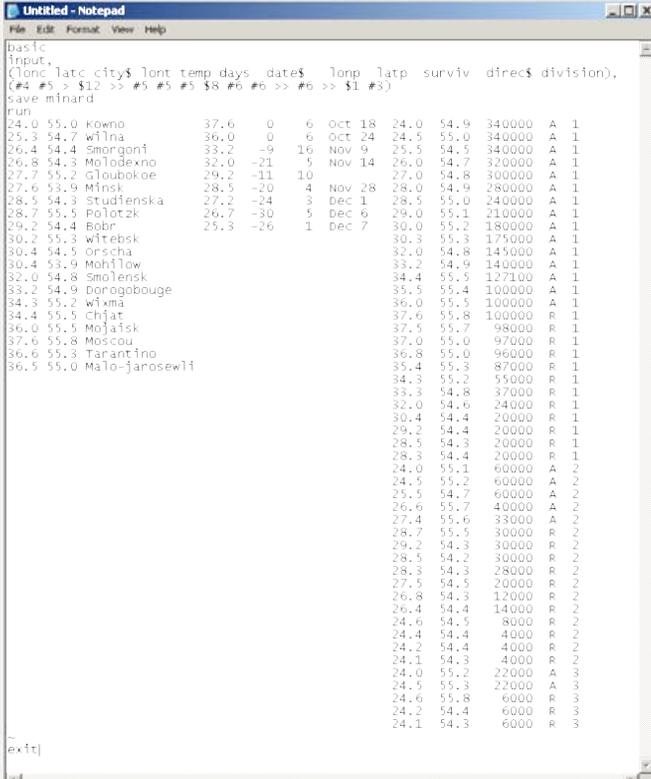
- Napoleon's march to and retreat from Moscow in 1812

- Variables

  - Time

  - Number of soldiers

  - Longitude / Latitude

  - Temperature

  - Direction of movement

- Some main principles of excellence in statistical graphics (Tufte)

  - Show the data

  - Avoid distortion of what the data has to say

  - Present many numbers in a small space

  - Make large data sets coherent

  - Encourage the eye to compare different pieces of the data
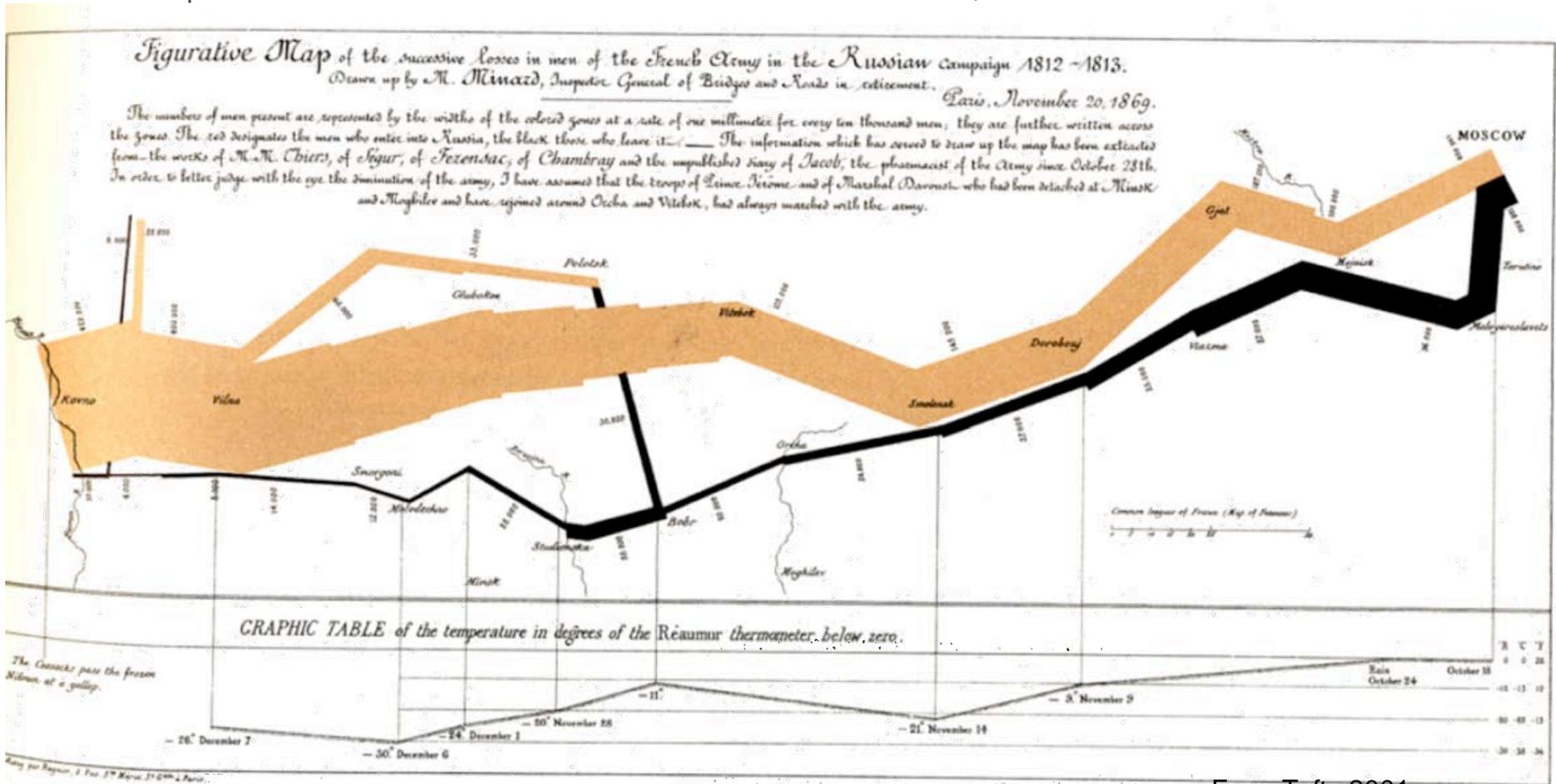
  - Reveal the data at several levels of detail

Visualization Example 1

Napoleon's march to and retreat from Moscow in 1812, Minard 1869 (Translation)

From Tufte 2001

Visualization Example 2

Diagram of the Causes of Mortality, Florence Nightingale (1858)

April 1854 – March 1855

April 1855 – March 1856
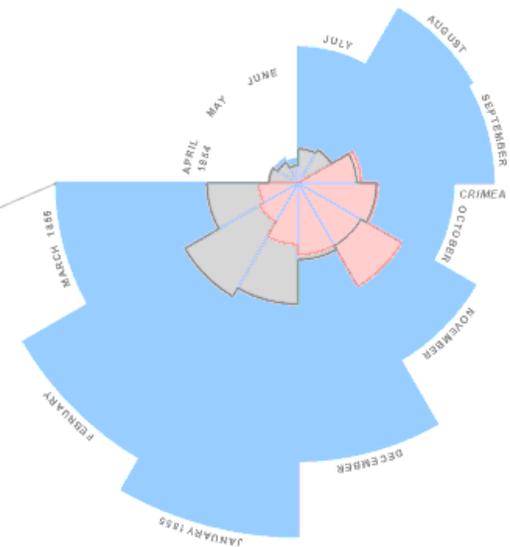
red: wounds, blue: preventable disease, gray: other causes

Images: http://www.fi.uu.nl/wiskrant/artikelen/hist_grafieken/quetelet/nightingale.html

Visualization Example 3

≡ London's Soho District - Death from cholerea (points) and locations of water pumps (crosses), Dr. John Snow (1854)

Image: http://www.math.yorku.ca/SCS/Gallery/images/tufte/snow.gif
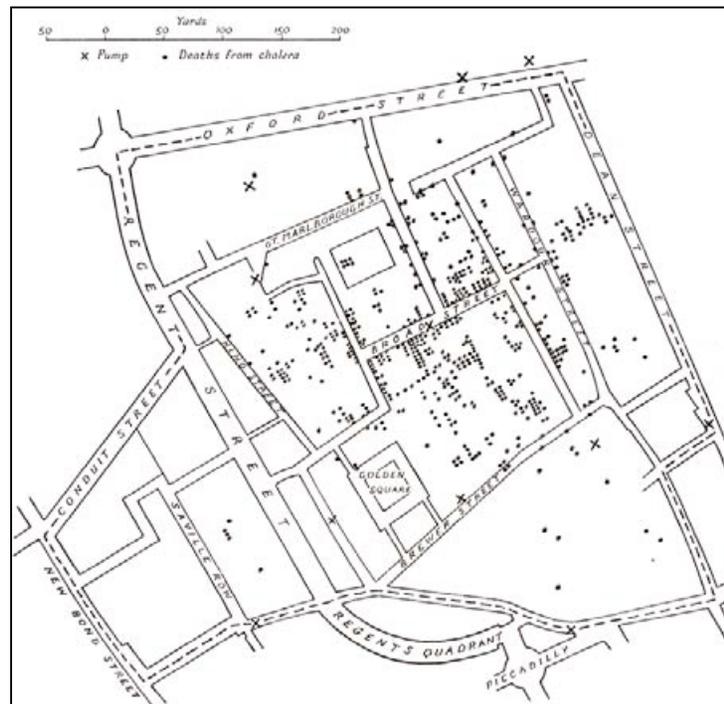
Visualization Example 3

≣ London's Soho Distrcit – Identification of bad water pump causing the cholera epidemic

Images: http://www.math.yorku.ca/SCS/Gallery/images/tufte/snow.gif

Visualization Example 4

≡ London underground map

Souce: http://www.kottke.org/plus/misc/images/tubegeo.gif
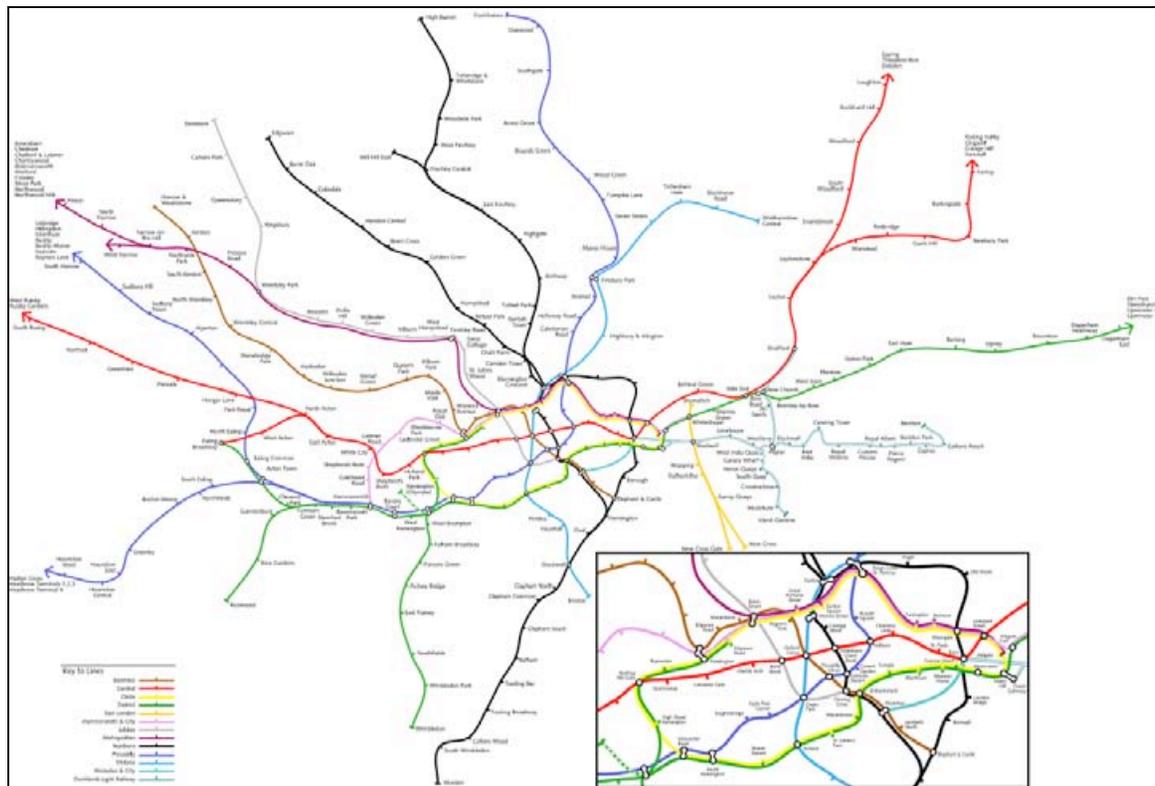
Visualization Example 4

≡ London underground map -- Design initiated by Harry Beck (1931)

Visualization Example 4

≡ Walking lines between stations, which are ca 500 meters apart from each other

Source: http://rodcorp.typepad.com/photos/art_2003/tube_walklines_final_lmfaint.html
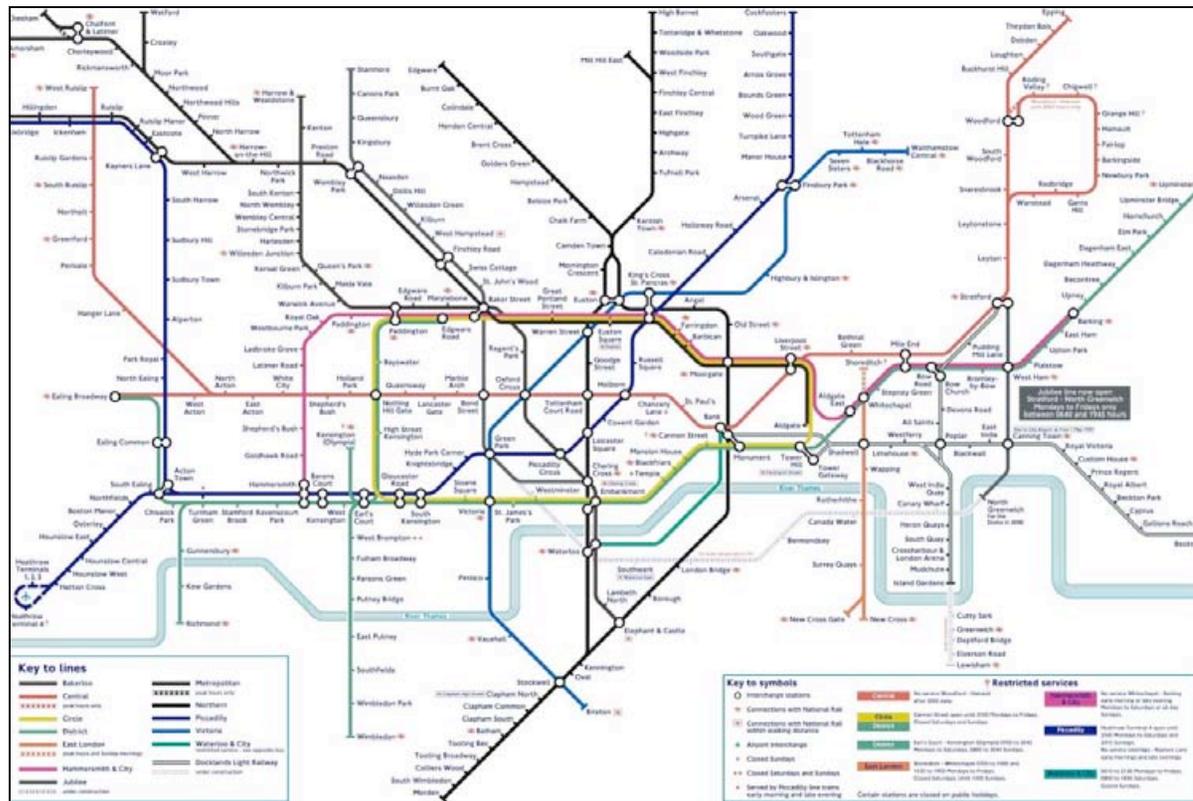
The Lie Factor

- Tufte 2001

- Lie factor = size of effect shown in graphic / size of effect in data

- Magnitude of change mpg: 53%

- Magnitude of the change of line size: 783%

- Lie factor = 14.8
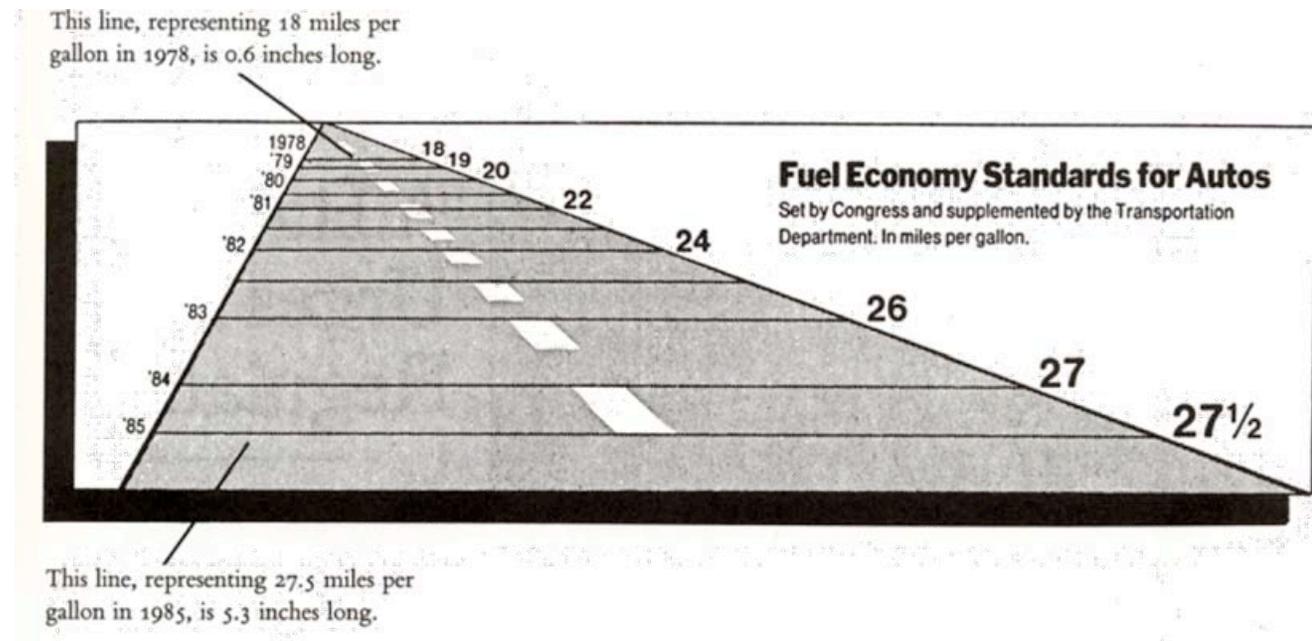
The Lie Factor

Lie by area: varying both dimensions simultaneously for change in 1D data (Tufte 2001)

If the area of the dollar is accurately to reflect its purchasing power, then the 1978 dollar should be about twice as big as that shown.
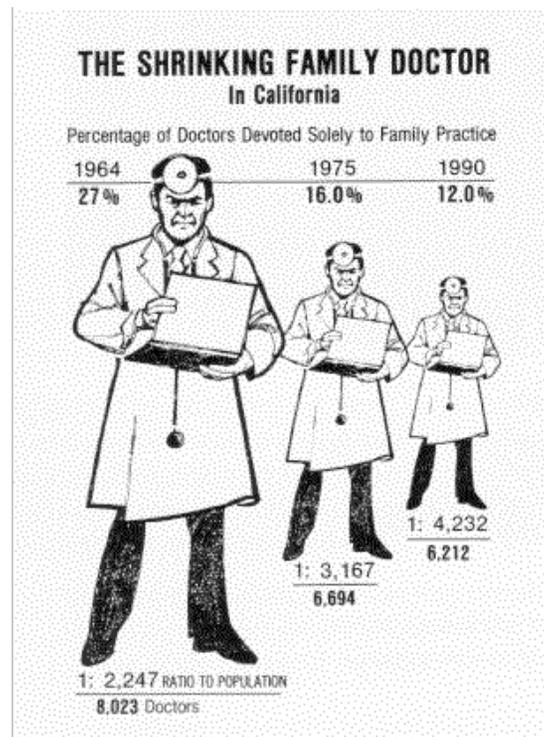
Lie factor: 2.8
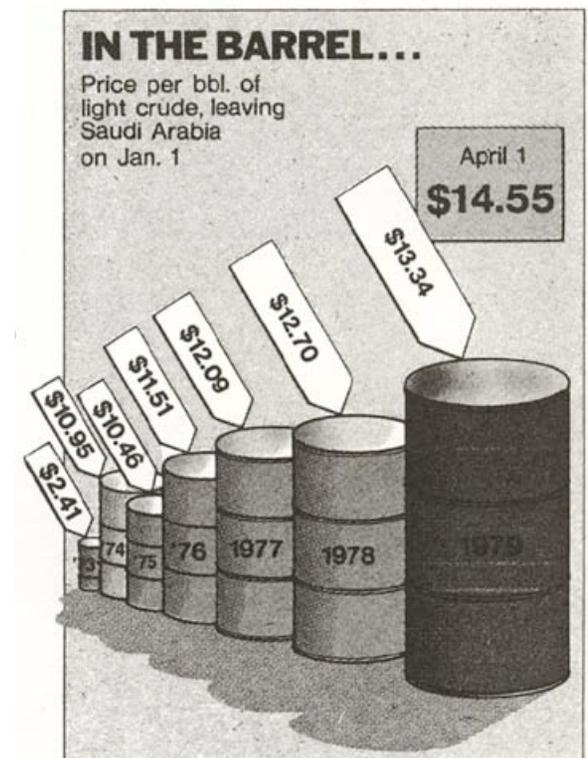
Lie factor: draw area of barrels: 9.4
Volume of barrels: 59.4

Definition Visualization

- Visualization
  - Dictionary definition: *"... to form a mental model or mental image of something."*
  - Edward R. Tufte: *"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."*
  - Tai-Hsi Fan: *"The purpose of visualization is insight, not pictures."*
    - Discovery
    - Decision making
    - Explanation
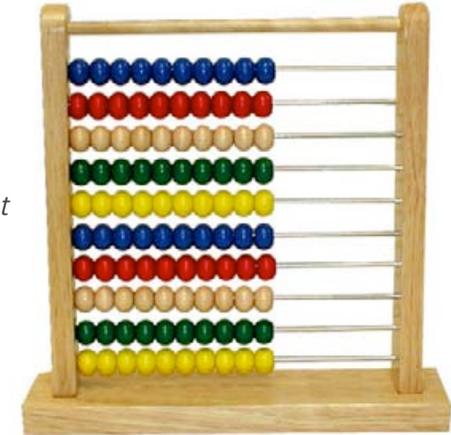
- Visuals provide a frame of reference, a temporary storage area to help us think

- Donald Norman: *"The power of the unaided mind is highly overrated. Without external aids, memory, thought, and reasoning are all constrained. But human intelligence is highly flexible and adaptive, superb at inventing procedures and objects that overcome its own limits. The real powers come from devising external aids that enhance cognitive abilities. How have we increased memory, thought, and reasoning? By the inventions of external aids: It is things that make us smart."*

$$\begin{array}{r} 1\ 7 \\ +\ 1\ 2_1\ 9 \\ \hline 1\ 4\ 6 \end{array}$$

Information Visualization

- Limitation of static visualization: only communicates information, which has been previously extracted and compiled by a designer

- **Information Visualization:** *"The use of computer-supported, interactive visual representations of abstract data to amplify cognition"* (Card et al. 1998)

- Abstract data

  - Items, entities, things which do not have a direct physical correspondence

  - Examples: football statistics, currency fluctuations, co-citation between scientists

- Amplify cognition by

  - Increasing memory and processing resources available

  - Reducing the amount of time to search

  - Enhancing the detections of patterns and enabling perceptual inference operations

  - Aiding perceptual monitoring

  - Encoding information in a manipulable medium

Draws from Several Domains

- Human-Computer Interaction

- Information Science

- Computer Graphics

- Cognitive Psychology

- Related / overlapping disciplines:

- Scientific Visualization

  - Visualize aspects of the 'natural world',

  - Data has physical representation, e.g. air flow over a wing, ozone concentration

  - **Example image** shows electric current within a thorax

- Visual Analytics

  - Science of analytical reasoning facilitated by interactive visual interfaces

  - An integrated approach combining visualization, human factors and data analysis

  - Research Agenda (U.S. Department of Homeland Security): http://nvac.pnl.gov/agenda.stm
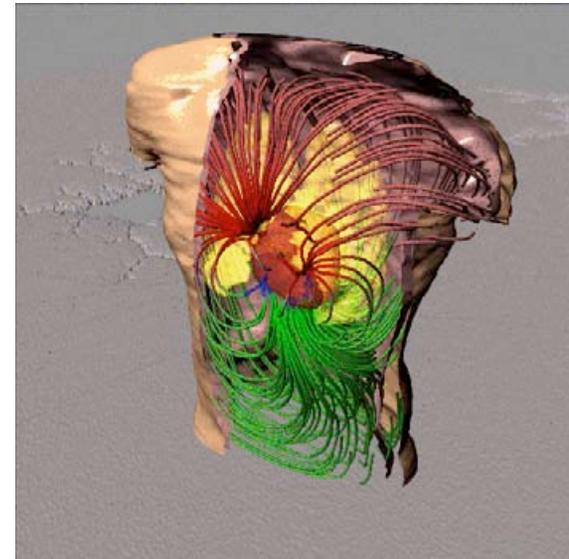
From Johnson et al., Univ. of Utah

Goals of Information Visualization

- Make large datasets coherent

- Compress data to a visual quintessence

- Present information from various viewpoints

- Present information at several levels of detail

- Support visual comparisons

- Tell stories about the data

InfoVis Challenges

≡ High data dimensionality

≡ Scale

≡ Advanced filtering mechanisms – which variables produce a potentially interesting
visualization?

≡ Usability

≡ Evaluation of usability

≡ Catherine Plaisant: "The Challenge of Information Visualization Evaluation" (2004):
http://hcil.cs.umd.edu/trs/2004-19/2004-19.pdf

User Tasks in InfoVis

☰ John Stasko:

☰ Search

   ☰ Finding a specific piece of information in a data set

   ☰ How many games did the Braves win in 1995?

   ☰ What novels did Ian Fleming author?

☰ Browsing

   ☰ Look over or inspect something in a more casual manner, seek interesting information

   ☰ Learn about crystallography

   ☰ What has Jane been up to lately?

☰ Analysis

   ☰ Comparison-Difference, find outliers and extremes, spot patterns

   ☰ Categorize, associate

   ☰ Locate, rank

   ☰ Identify, reveal

   ☰ Monitor, maintain awareness

InfoVis Reference Model

- Raw table to data table: filtering, data cleaning

- Data table to visual structures: pick mappings

- Visual structures to views: viewpoints, distortion etc.
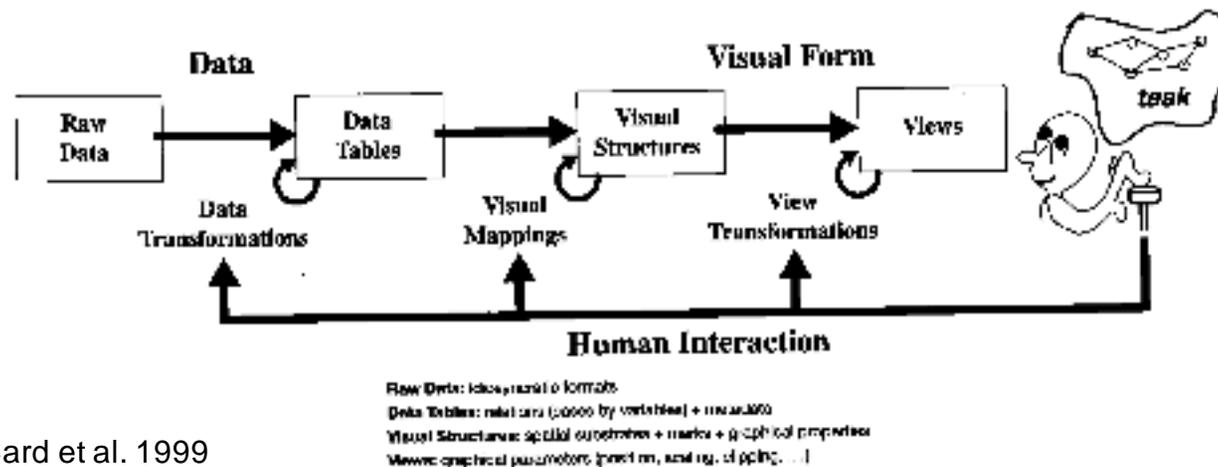
Card et al. 1999

Example: Exploring Enron

≡ Enron

≡ Leading energy company in the US

≡ Large-scale accounting fraud

≡ Bankruptcy in late 2001

≡ Played an important role in the California energy crisis

≡ Archive of 500.000 emails of 150 persons

≡ Social network visualization by Jeffrey Hear

≡ http://jheer.org/enron

Example: Exploring Enron

Example: Exploring Enron

≡ Identification of hubs and authorities in the network via visual filtering – who sent / received many emails involving the California Energy Crisis

≡ Applied Natural Language Processing

≡ *"Such an analysis revealed the role of John Shelk, who regularly reported on Congressional meetings, sending all such meeting reports to Tim Belden. In fact, the visualization reveals that their conversation is completely one-sided, with John sending reports to Tim, with no back-traffic occurring. This is a bit suspicious. Clicking on Tim Belden then reveals that according to the database he hasn't sent ANY e-mails, but receives various legal reports from throughout the company."*
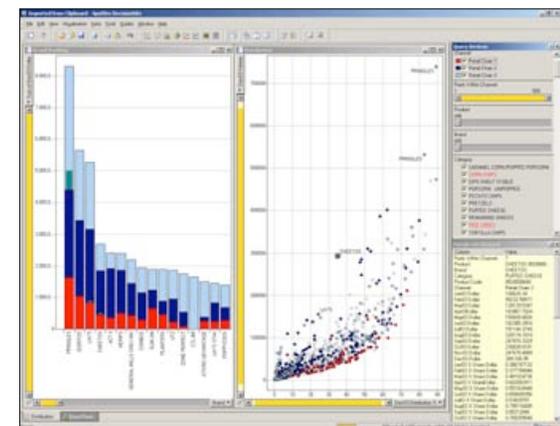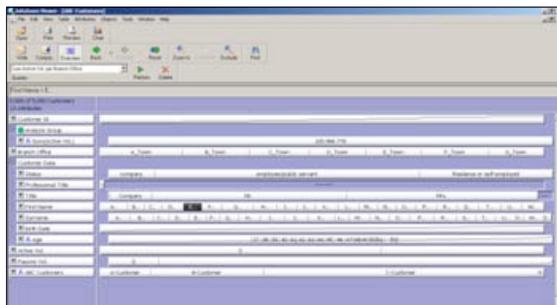
Example: Exploring Enron
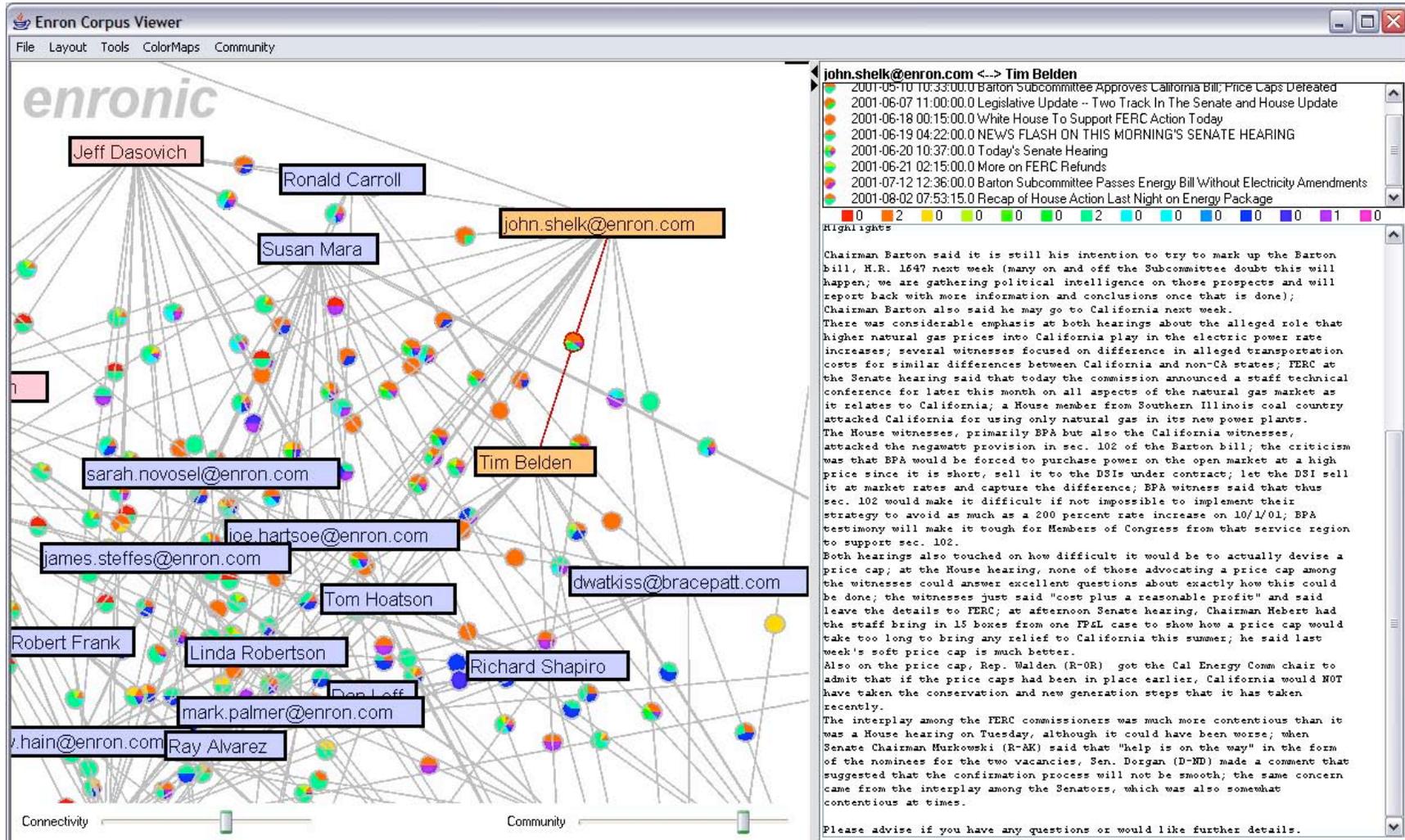
Example: Exploring Enron
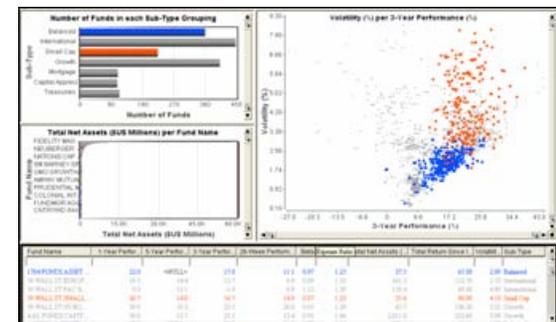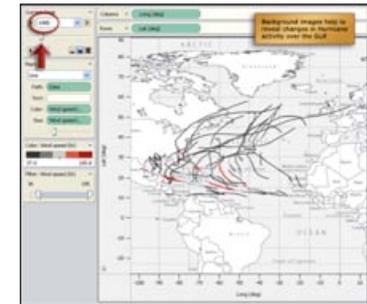
- *"After performing this analysis, I did a web search on Google for "Tim Belden". I had never heard his name before doing this analysis exercise. Little did I know he was the first person charged by prosecutors, considered the "mastermind" of Enron's manipulation of California's markets, and was found guilty on charges of federal conspiracy. "*

- Have a look at prefuse: http://prefuse.org/

Commercial InfoVis Systems

≡ Tableau Software http://www.tableausoftware.com/

≡ Spotfire (Tibco): http://spotfire.tibco.com/index.cfm

≡ InoZoom (Siemens):

http://www.infozoom.com/deu/infozoom/video.htm

≡ InfoScope (Macrofocus):

http://www.macrofocus.com/public/products/infoscope.html

≡ Applet: http://download.macrofocus.com/infoscope/

≡ Advizor Analyst: http://www.advizorsolutions.com/desktop.htm

Lecture Outline

18.10.2007	Introduction InfoVis
25.10.2007	Visual Perception
01.11.2007	Frei - Allerheiligen
08.11.2007	Multidimensional Information Visualization 1
15.11.2007	Multidimensional Information Visualization 2
22.11.2007	Interaction - Dynamic Queries
29.11.2007	Graphs & Networks
06.12.2007	Hierarchies & Trees
13.12.2007	Time-based data
20.12.2007	Text & Documents
10.01.2008	Presentation Approaches 1
17.01.2008	Prof. Dr. Harald Reiterer - Visuelle Suchsysteme
24.01.2008	Presentation Approaches 2
31.01.2008	Michael Sedlmair - DA BMW (InfoVis in der Praxis)
07.02.2008	Ambient InfoVis

Student Tasks

- Summary of **6** research papers á **100 words in English**

  - Papers will be announced during the lecture and provided via the website

  - No copy of the abstract!

  - Send complete set of summaries to sebastian.boring@ifi.lmu.de

- Successful participation in the tutorial

  - 11.02.2008 - 20.02.2008

  - Identify and design an effective solution for a given visualization problem