

# 1. Information Visualization: Introducing Research Need & Concepts

Vorlesung „Informationsvisualisierung“  
Prof. Dr. Andreas Butz, WS 2009/10  
Konzept und Folien: Thorsten Büring

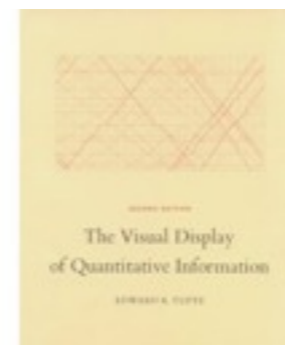
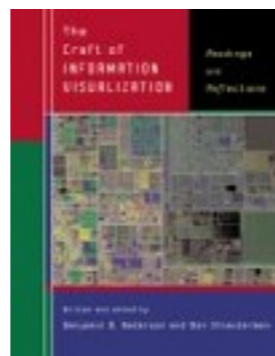
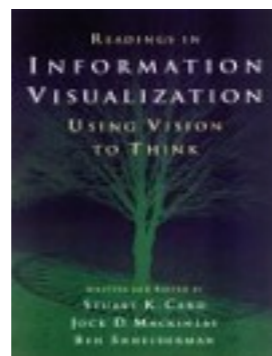


# Outline

- The Problem
- The power of visualization
- Classic examples
- Information Visualization
  - Definition
  - Classification
  - Challenges
  - User tasks
  - Reference Model
- Enron example
- Lecture 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.

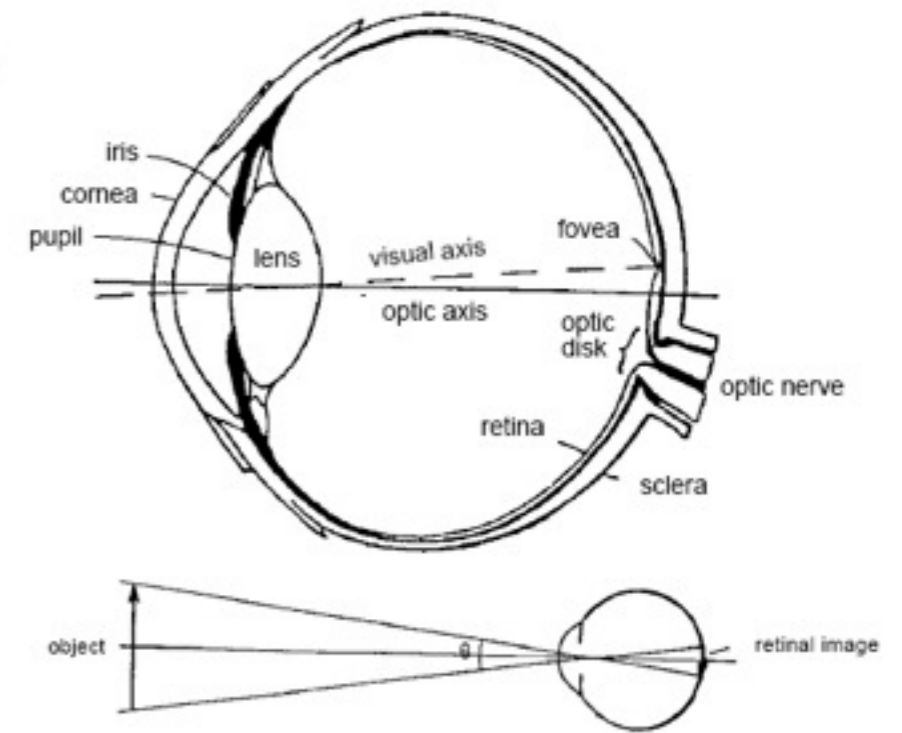


# Data Explosion

- 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 1000 million bits per second)
  - Pattern recognition
  - Pre-attentive perception
  - Extends memory and cognitive capacity
  - People think visually



# Visualization Example 1

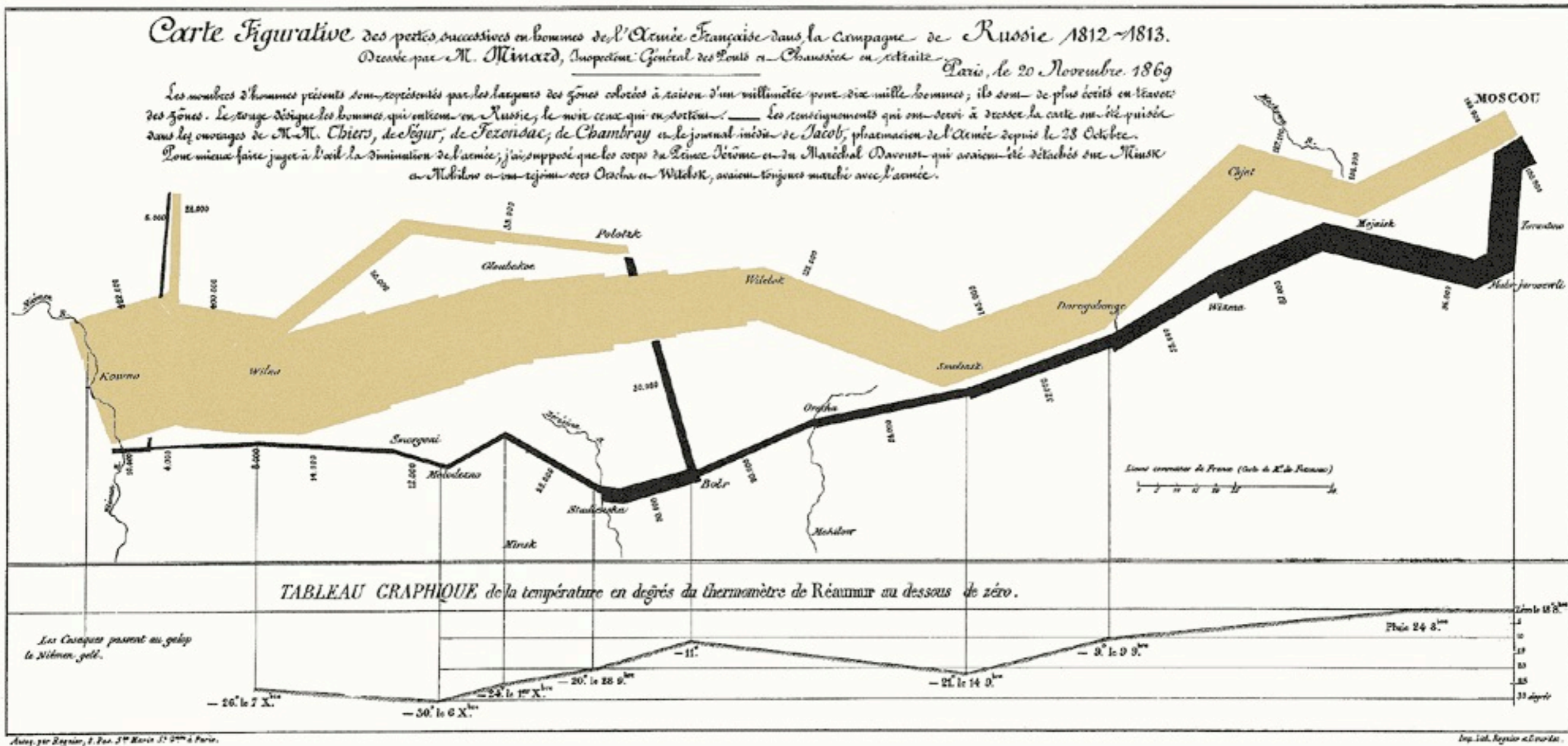
- Napoleon's march to and retreat from Moscow in 1812
- Variables
  - Time
  - Number of soldiers
  - Longitude / Latitude
  - Temperature
  - Direction of movement
- Some main principles of excellence in statistical graphics (Tufte)
  - Show the data
  - Avoid distortion of what the data has to say
  - 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

```
basic
input,
(lonc latc city$ lont temp days date$ lonp latp surviv direc$ division),
(#4 #5 > $12 >> #5 #5 #5 $8 #6 #6 >> #6 >> $1 #3)
save minard
run
24.0 55.0 kowno          37.6  0  6  Oct 18  24.0  54.9  340000  A  1
25.3 54.7 wlna          36.0  0  6  Oct 24  24.5  55.0  340000  A  1
26.4 54.4 Smorgoni     33.2 -9 16  Nov  9  25.5  54.5  340000  A  1
26.8 54.3 Molodexno    32.0 -21 5  Nov 14  26.0  54.7  320000  A  1
27.7 55.2 Gloubokoe    29.2 -11 10         27.0  54.8  300000  A  1
27.6 53.9 Minsk        28.5 -20  4  Nov 28  28.0  54.9  280000  A  1
28.5 54.3 Studfienska  27.2 -24  3  Dec  1  28.5  55.0  240000  A  1
28.7 55.5 Polotzk      26.7 -30  5  Dec  6  29.0  55.1  210000  A  1
29.2 54.4 Bobr         25.3 -26  1  Dec  7  30.0  55.2  180000  A  1
30.2 55.3 witebsk      30.3  55.3  175000  A  1
30.4 54.5 Orscha       32.0  54.8  145000  A  1
30.4 53.9 Mohilow      33.2  54.9  140000  A  1
32.0 54.8 Smolensk     34.4  55.5  127100  A  1
33.2 54.9 Dorogobouge  35.5  55.4  100000  A  1
34.3 55.2 wixma        36.0  55.5  100000  A  1
34.4 55.5 Chjat        37.6  55.8  100000  R  1
36.0 55.5 Mojaisk      37.5  55.7  98000   R  1
37.6 55.8 Moscou       37.0  55.0  97000   R  1
36.6 55.3 Tarantino    36.8  55.0  96000   R  1
36.5 55.0 Malo-jarosew1 35.4  55.3  87000   R  1
34.3 55.2              34.3  55.2  55000   R  1
33.3 54.8              33.3  54.8  37000   R  1
32.0 54.6              32.0  54.6  24000   R  1
30.4 54.4              30.4  54.4  20000   R  1
29.2 54.4              29.2  54.4  20000   R  1
28.5 54.3              28.5  54.3  20000   R  1
28.3 54.4              28.3  54.4  20000   R  1
24.0 55.1              24.0  55.1  60000   A  2
24.5 55.2              24.5  55.2  60000   A  2
25.5 54.7              25.5  54.7  60000   A  2
26.6 55.7              26.6  55.7  40000   A  2
27.4 55.6              27.4  55.6  33000   A  2
28.7 55.5              28.7  55.5  30000   R  2
29.2 54.3              29.2  54.3  30000   R  2
28.5 54.2              28.5  54.2  30000   R  2
28.3 54.3              28.3  54.3  28000   R  2
27.5 54.5              27.5  54.5  20000   R  2
26.8 54.3              26.8  54.3  12000   R  2
26.4 54.4              26.4  54.4  14000   R  2
24.6 54.5              24.6  54.5  8000    R  2
24.4 54.4              24.4  54.4  4000    R  2
24.2 54.4              24.2  54.4  4000    R  2
24.1 54.3              24.1  54.3  4000    R  2
24.0 55.2              24.0  55.2  22000   A  3
24.5 55.3              24.5  55.3  22000   A  3
24.6 55.8              24.6  55.8  6000    R  3
24.2 54.4              24.2  54.4  6000    R  3
24.1 54.3              24.1  54.3  6000    R  3
exitj
```



# Visualization Example 1

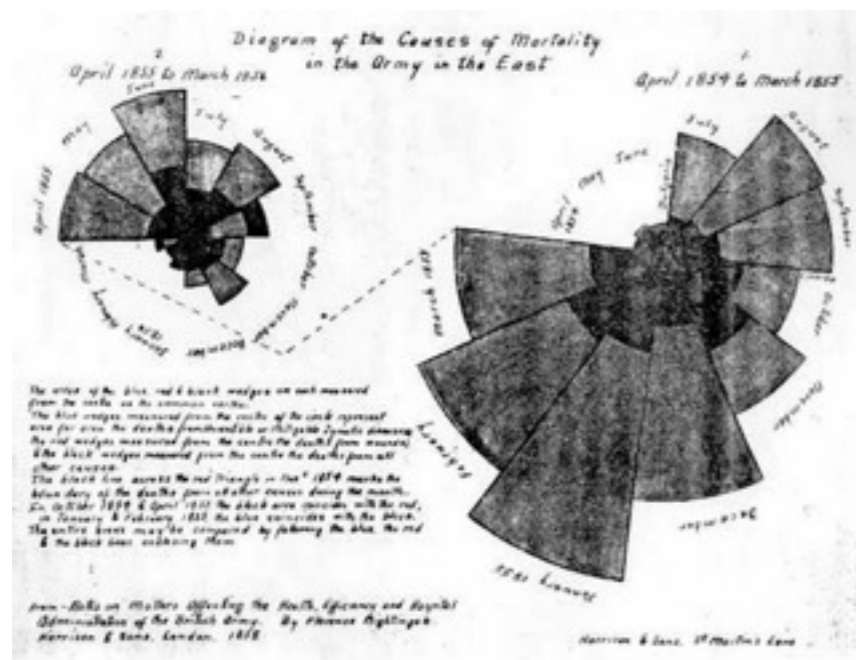
- Napoleon's march to and retreat from Moscow in 1812, Minard 1869 (Translation)



source: <http://www.edwardtufte.com/>

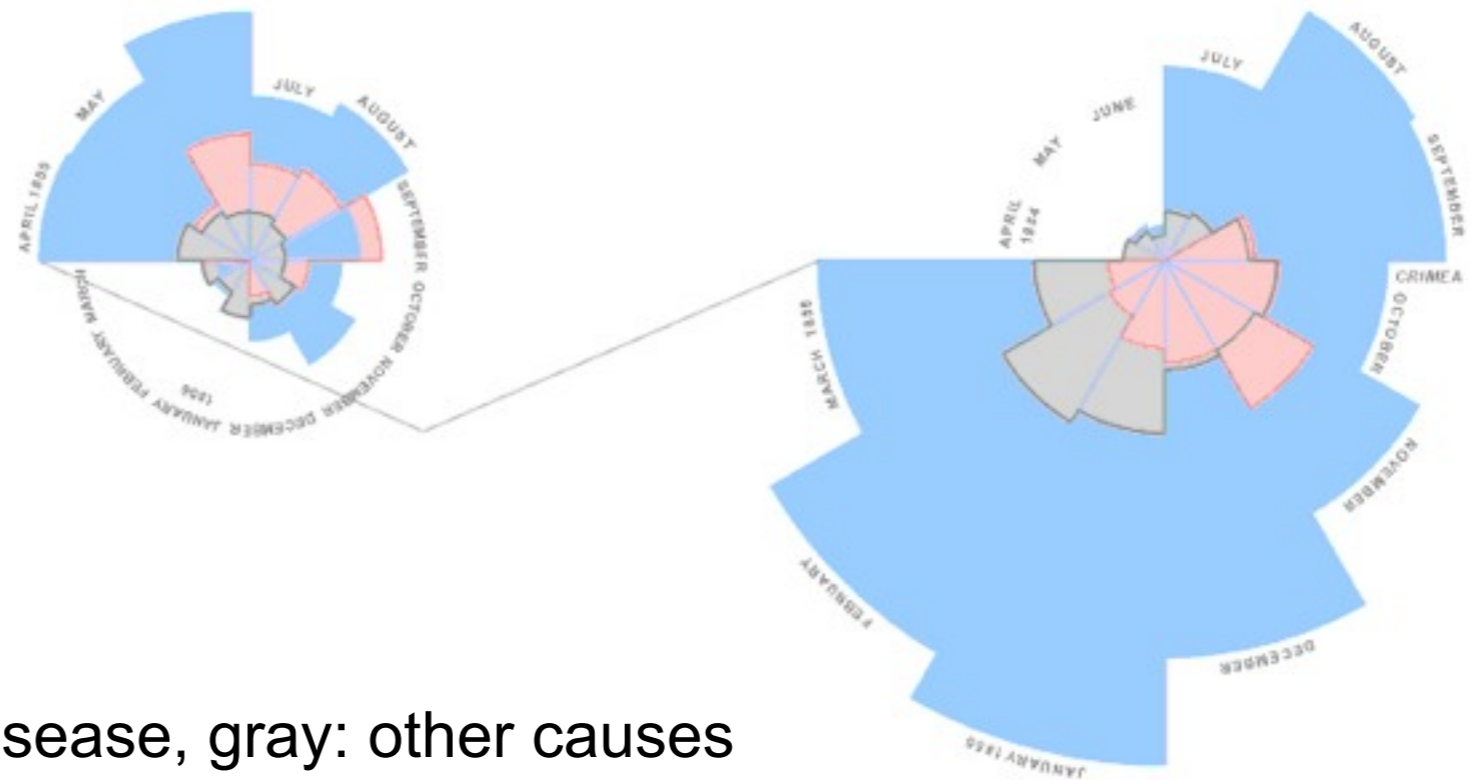
# Visualization Example 2

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



April 1855 – March 1856

April 1854 – March 1855



red: wounds, blue: preventable disease, gray: other causes

Images: [http://www.fi.uu.nl/wiskrant/artikelen/hist\\_grafieken/quetelet/nightingale.html](http://www.fi.uu.nl/wiskrant/artikelen/hist_grafieken/quetelet/nightingale.html)



# Visualization Example 3

- London's Soho District - Death from cholera (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

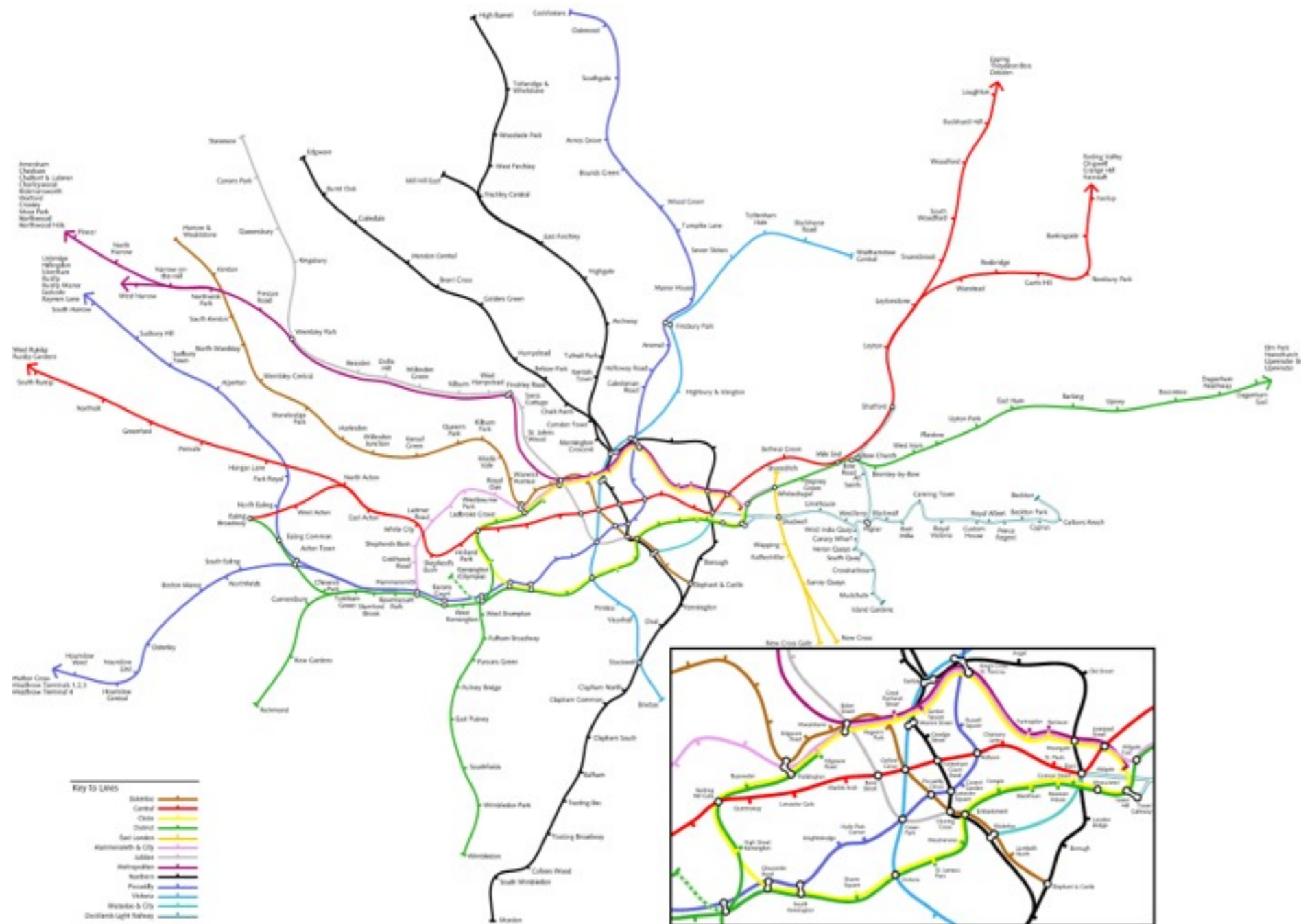
- London's Soho District – Identification of bad water pump causing the cholera epidemic



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

# Visualization Example 4

- London underground map



Source: <http://www.kottke.org/plus/misc/images/tubegeo.gif>



# Visualization Example 4

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

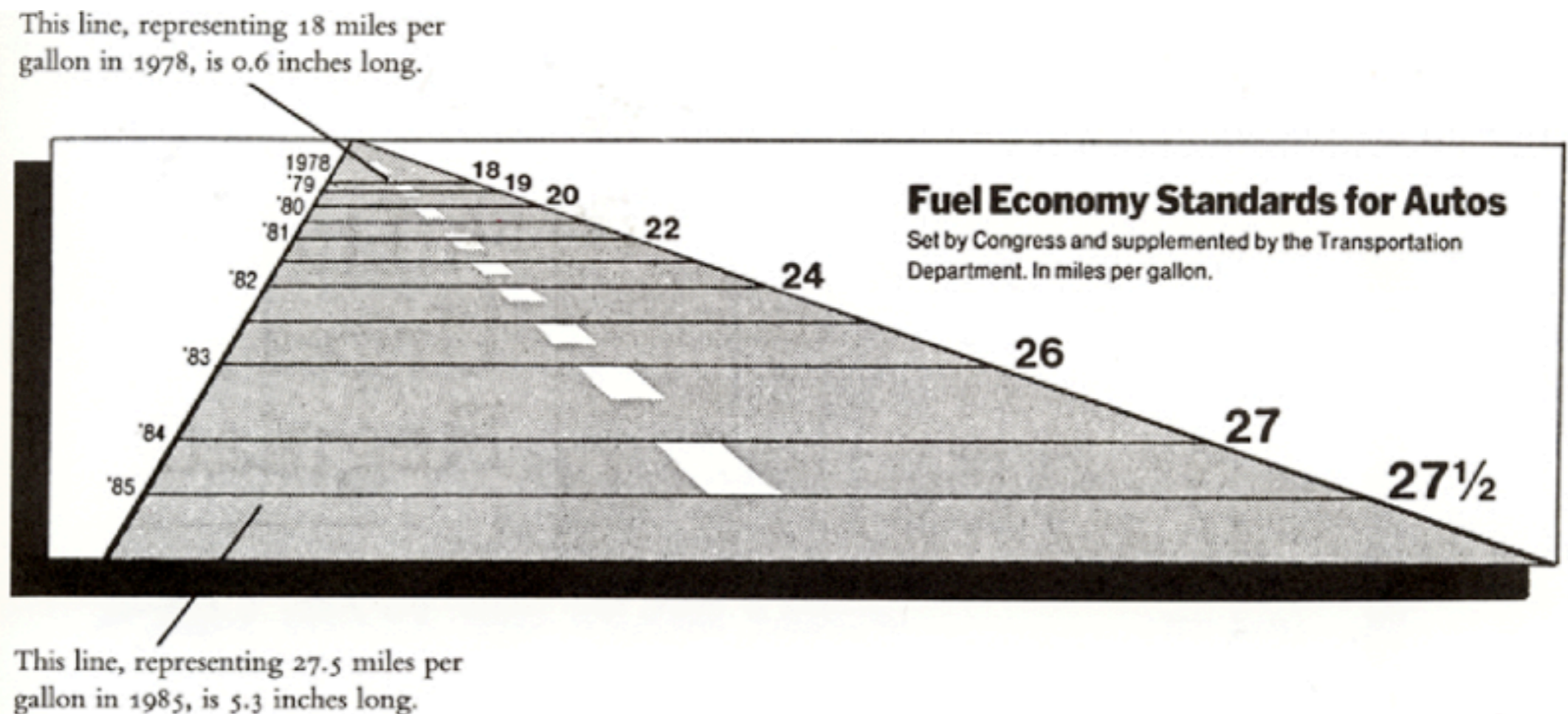






# The Lie Factor

- 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



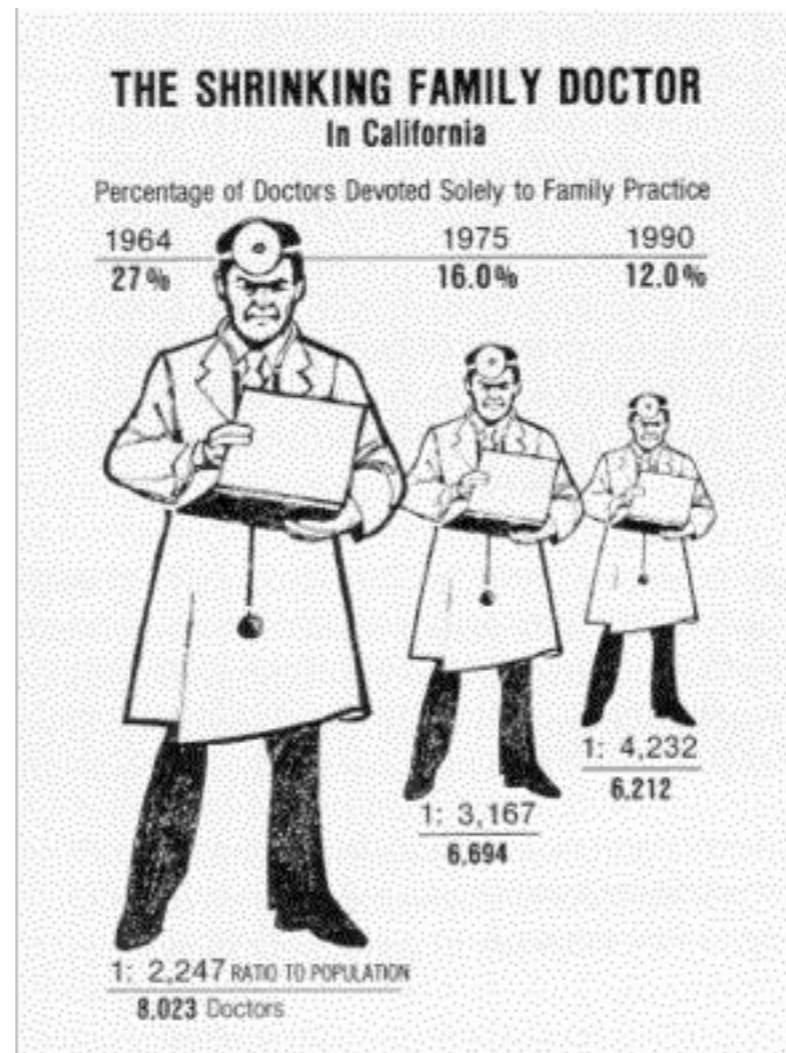


# The Lie Factor

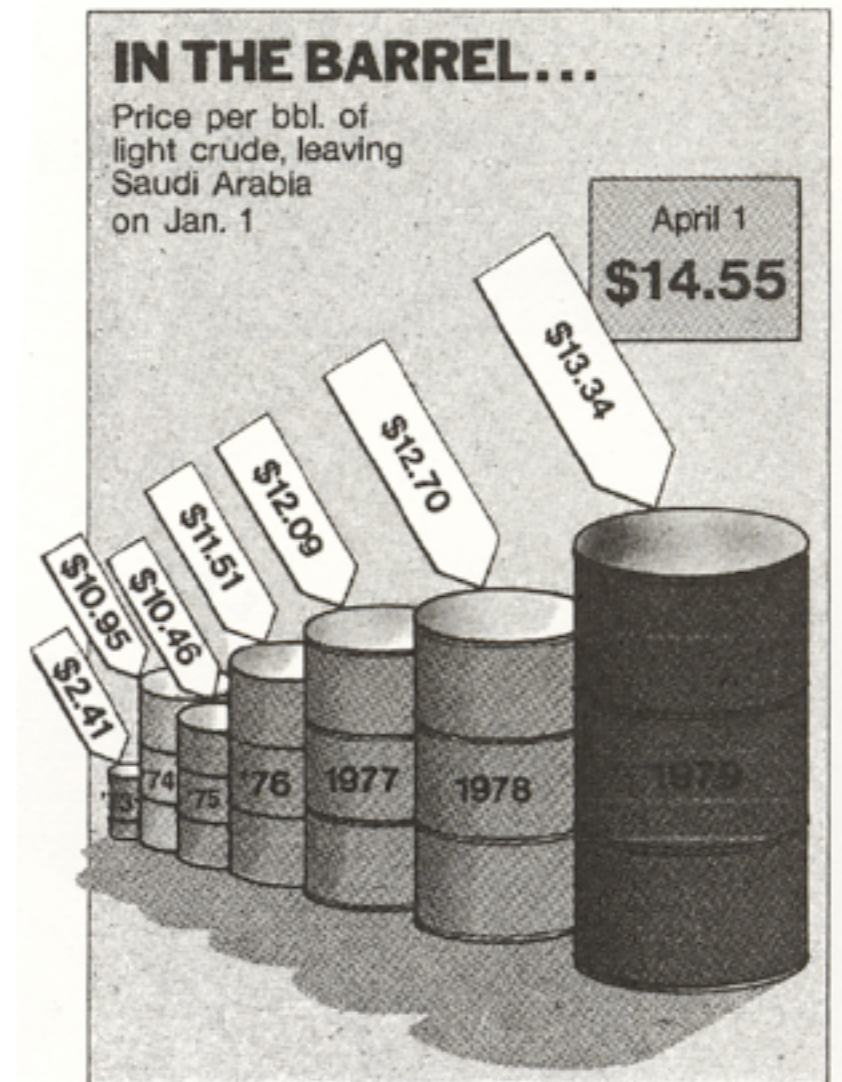
- Lie by area: varying both dimensions simultaneously for change in 1D data (Tufte 2001)



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: 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

$$\begin{array}{r} 17 \\ + 129 \\ \hline = 146 \end{array}$$





# Definition: Visualization

- 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.**”

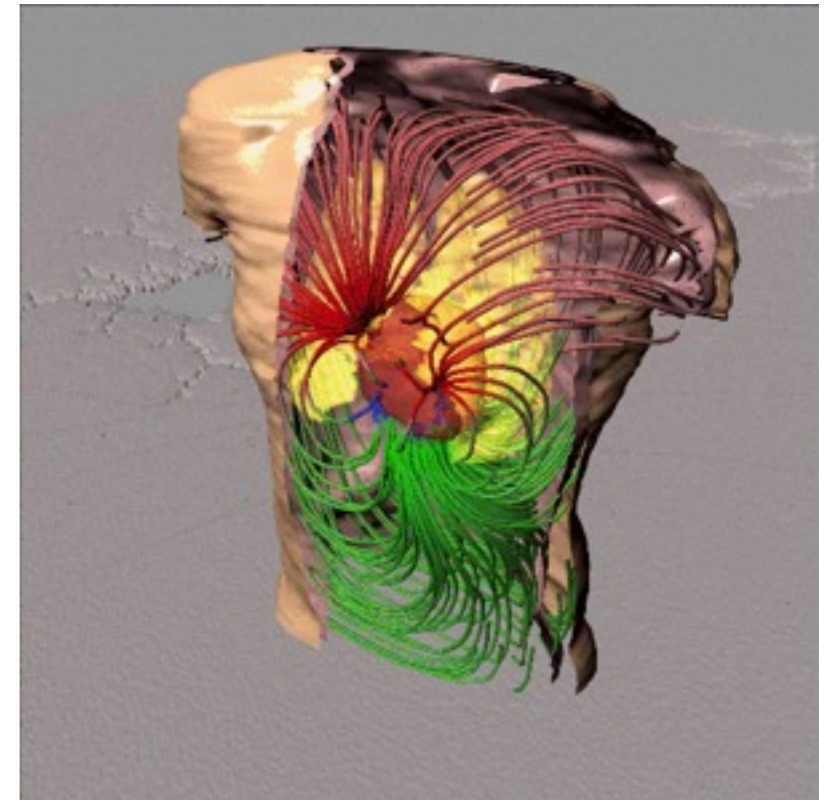
# Information Visualization

- 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

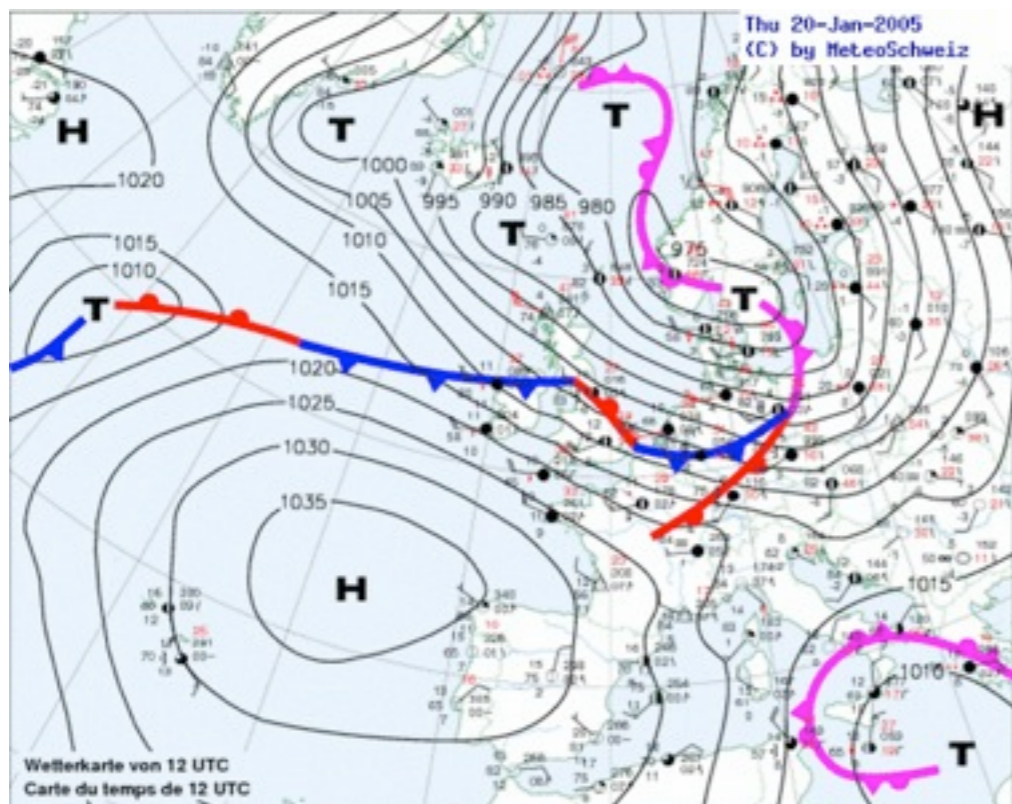
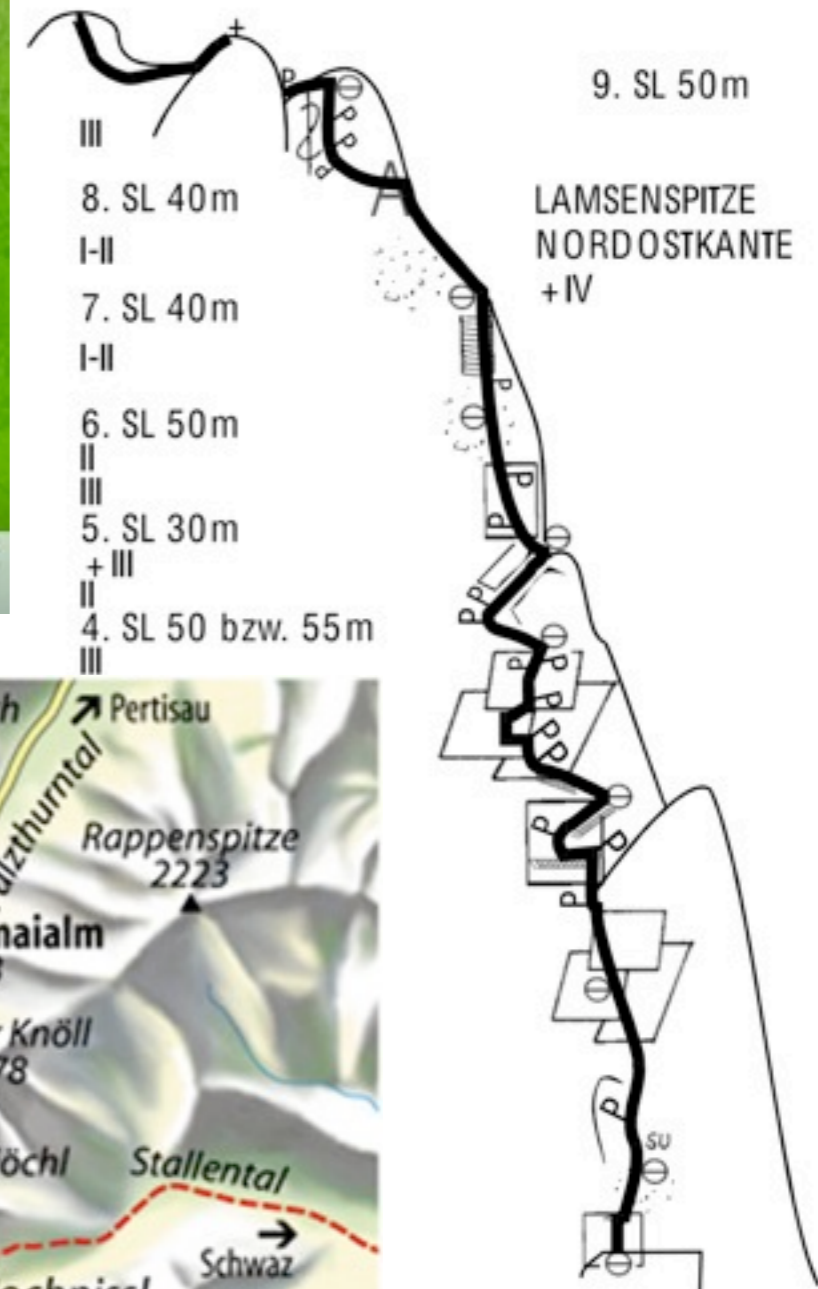
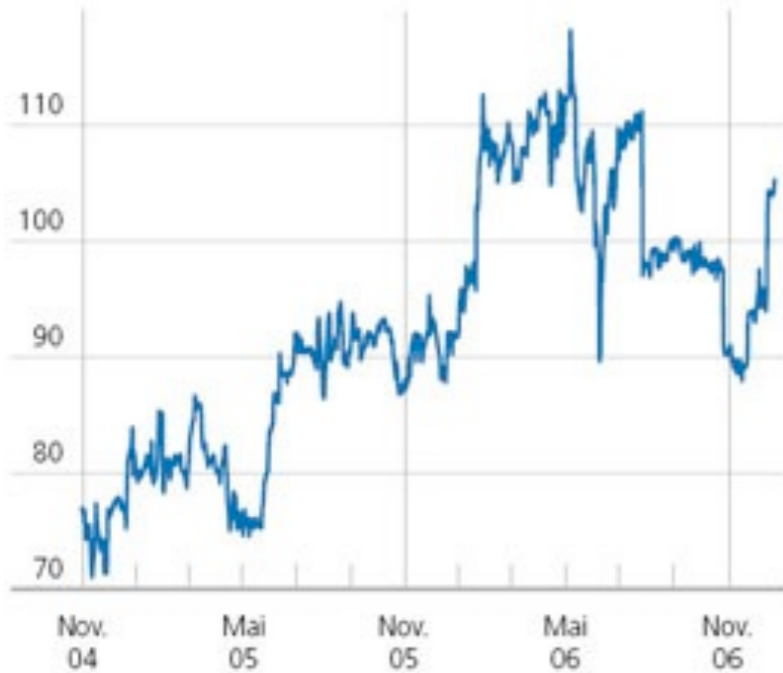


From Johnson et al., Univ. of Utah

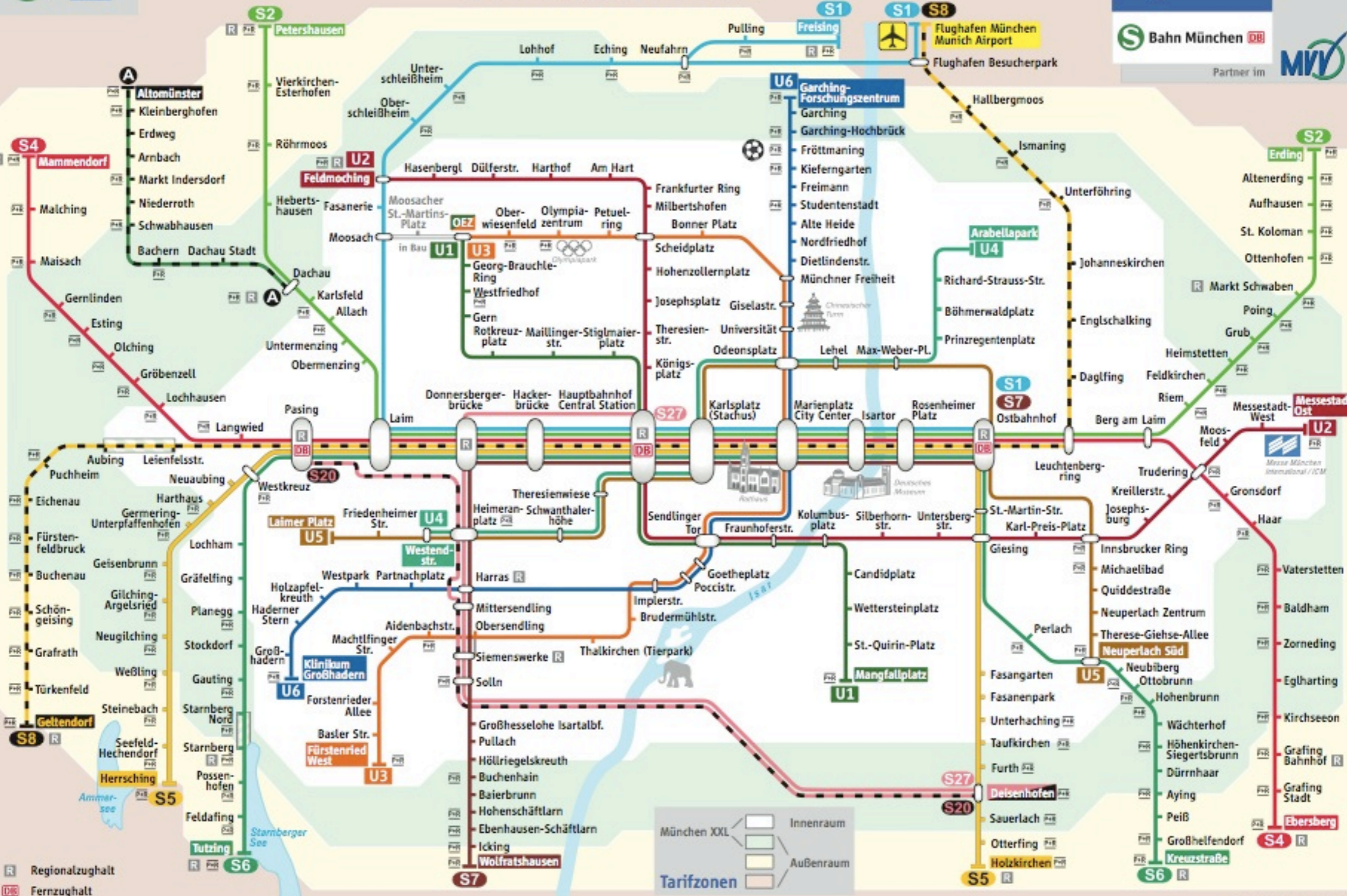
- Visualize aspects of the ‘natural world’,
- Data has physical representation, e.g. air flow over a wing, ozone concentration
- Example image shows electric current within a thorax
  
- Visual Analytics
  - Science of analytical reasoning facilitated by interactive visual interfaces
  - An integrated approach combining visualization, human factors and data analysis
  - Research Agenda (U.S. Department of Homeland Security): <http://nvac.pnl.gov/agenda.stm>



# Some more daily life examples of InfoVis









# 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>

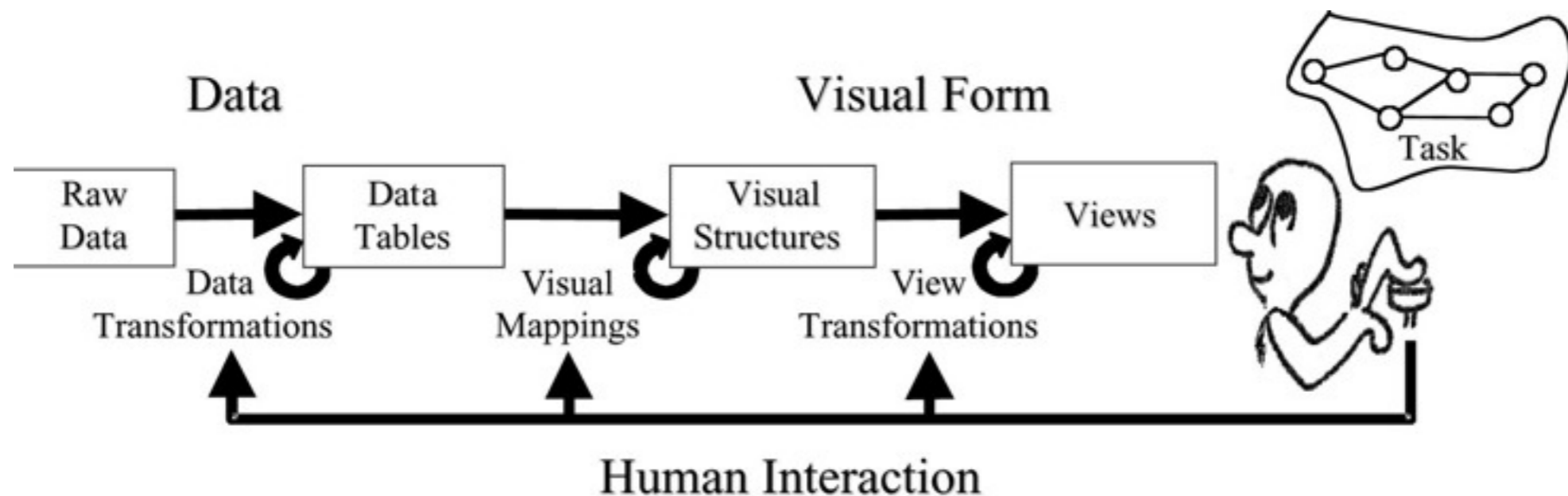


# User Tasks in InfoVis

- 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.



**Raw Data:** idiosyncratic formats

**Data Tables:** relations (cases by variables) + meta-data

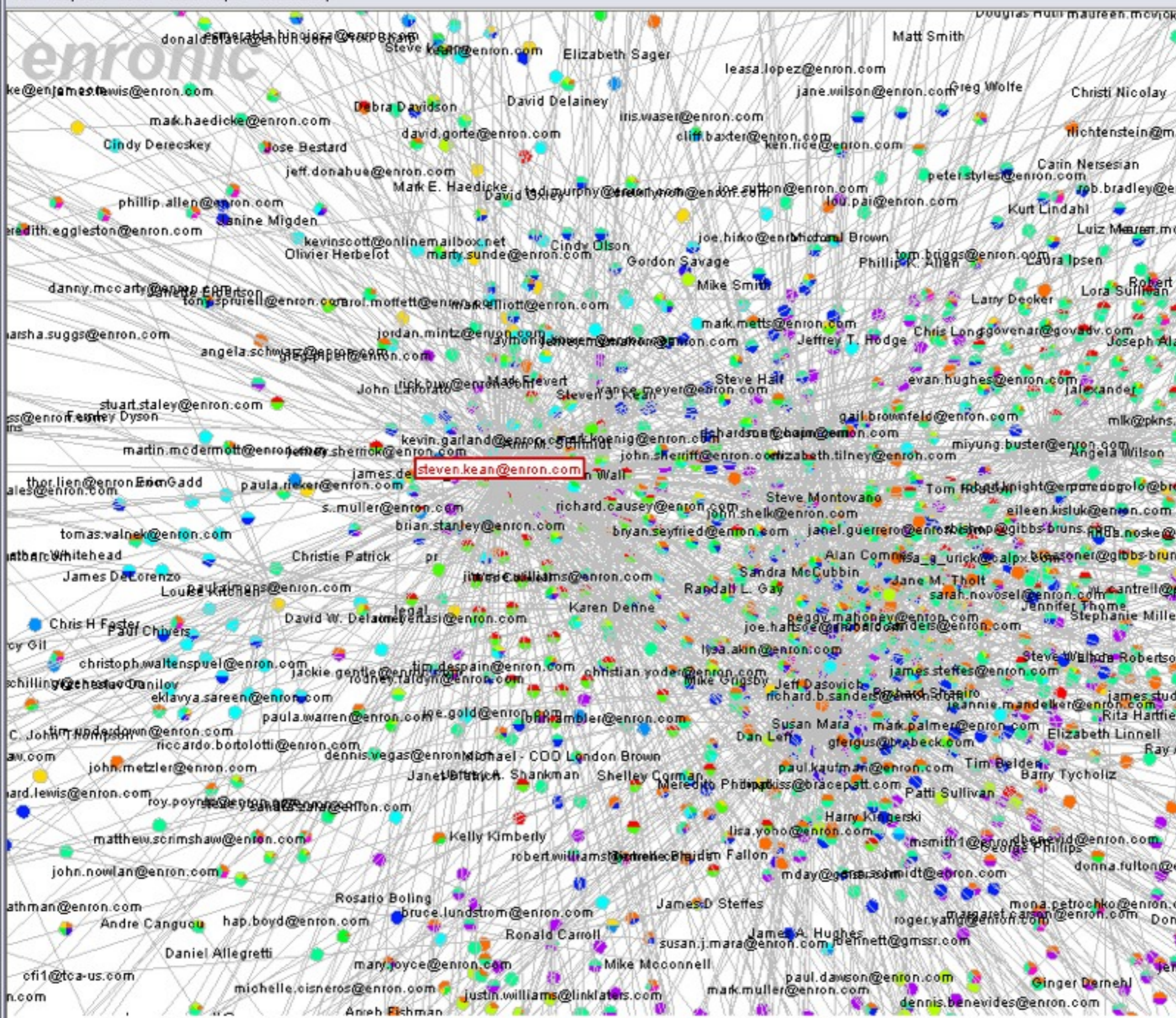
**Visual Structures:** spatial substrates + marks + graphical properties

**Views:** graphical parameters (position, scaling, clipping, ...)

# 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>





- steven.kean@enron.com**
- 1979-12-31 16:00:00.0 Further information regarding...
  - 1979-12-31 16:00:00.0 Re: Trains - Light r...
  - 1979-12-31 16:00:00.0 Re:
  - 1979-12-31 16:00:00.0 Re: Job Titles and...
  - 1999-09-06 07:06:00.0 Monthly Billing - D...
  - 1999-09-10 06:06:00.0 Commission Meeti...
  - 1999-09-10 06:22:00.0 Speech to the Briti...
  - 1999-09-21 08:20:00.0 Ross Perot's EMS...
  - 1999-10-06 06:44:00.0 Madera Ranch Pr...
  - 1999-10-18 01:47:00.0 Translation of artic...

0 0 0 0 0 0 2 0

ID: 172241  
 Subject: Further information regarding...  
 From: <steven.kean@enron.com>  
 Date: 1979-12-31 16:00:00.0  
 To: <kenneth.lay@enron.com>  
 Cc: Rosalee Fleming <rosalee.fleming@enron.com>, <susan.landwehr@enron.com>, <kevin.ha...@enron.com>, <david.delainey@enron.com>, Don Black...

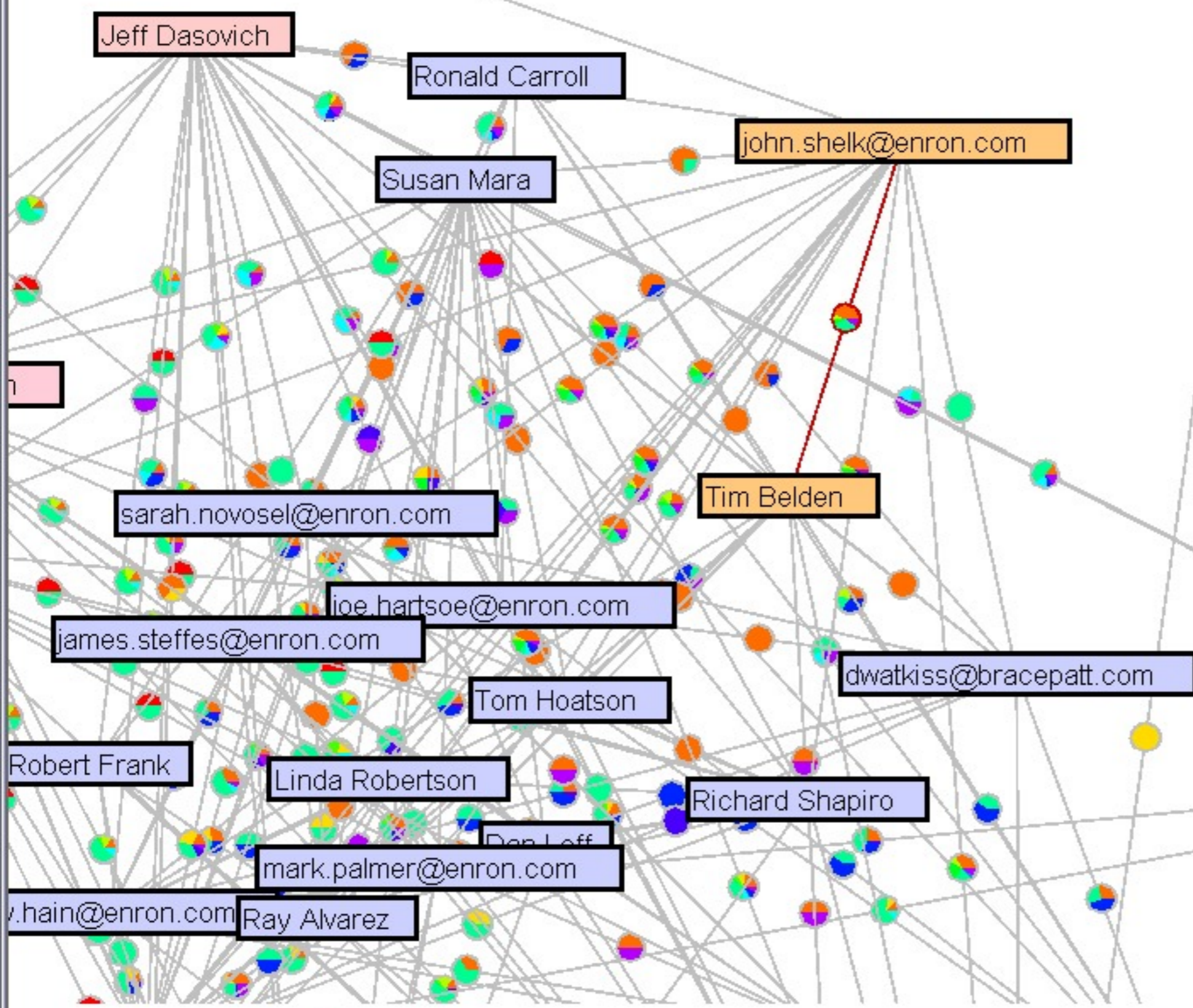
Sue Landwehr in government affairs sp...  
 administration who indicated that the...  
 talking...  
 to you about Antrak. Specifically, E...  
 for...  
 quite some time on an electricity



# 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.”

# enronic



- john.shelk@enron.com <-> Tim Belden**
- 2001-05-10 10:33:00.0 Barton Subcommittee Approv
  - 2001-06-07 11:00:00.0 Legislative Update -- Two Tra
  - 2001-06-18 00:15:00.0 White House To Support FER
  - 2001-06-19 04:22:00.0 NEWS FLASH ON THIS MO
  - 2001-06-20 10:37:00.0 Today's Senate Hearing
  - 2001-06-21 02:15:00.0 More on FERC Refunds
  - 2001-07-12 12:36:00.0 Barton Subcommittee Passes
  - 2001-08-02 07:53:15.0 Recap of House Action Last

0 2 0 0 0 0 2

Highlights

Chairman Barton said it is still his intention to introduce the bill, H.R. 1547 next week (many on and off the record say it will happen; we are gathering political intelligence on this). Chairman Barton also said he may go to California next week. There was considerable emphasis at both hearings on the impact of higher natural gas prices into California play. The House witnesses, primarily BPA but also the California witnesses, focused on different aspects of the bill. The Senate hearing said that today the commission will report back with more information and conclusions. The House conference for later this month on all aspects of the bill. It relates to California; a House member from California attacked California for using only natural gas. The House witnesses, primarily BPA but also the California witnesses, attacked the negawatt provision in sec. 102 of the bill. The testimony will make it tough for Members of Congress to support sec. 102. Both hearings also touched on how difficult it is to implement a price cap; at the House hearing, none of those questions could be done; the witnesses just said "cost plus a reasonable return" leave the details to FERC; at afternoon Senate hearing, the staff bring in 15 boxes from one FP&L case. The week's soft price cap is much better. Also on the price cap, Rep. Walden (R-OR) got the credit for admitting that if the price caps had been in place, California would have taken the conservation and new generation investments more recently. The interplay among the FERC commissioners was a House hearing on Tuesday, although it came from the interplay among the Senators, which was contentious at times.

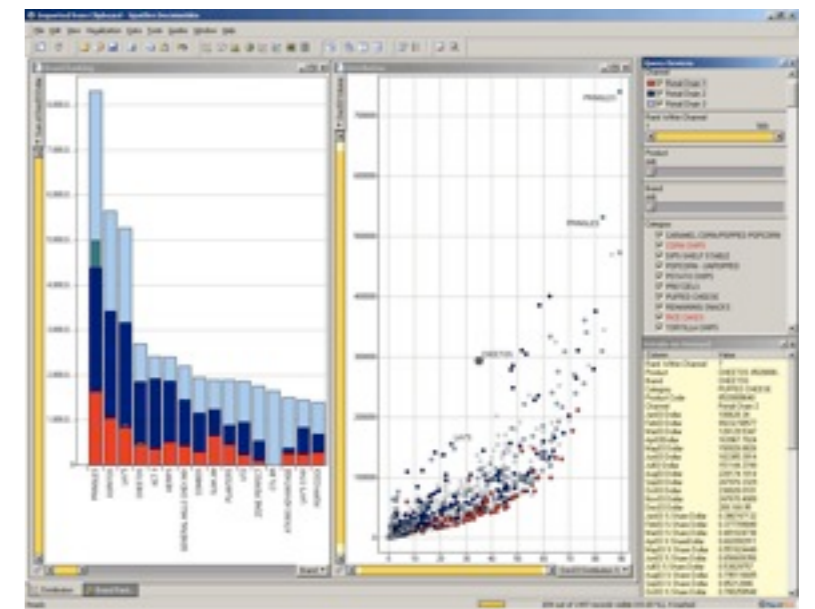
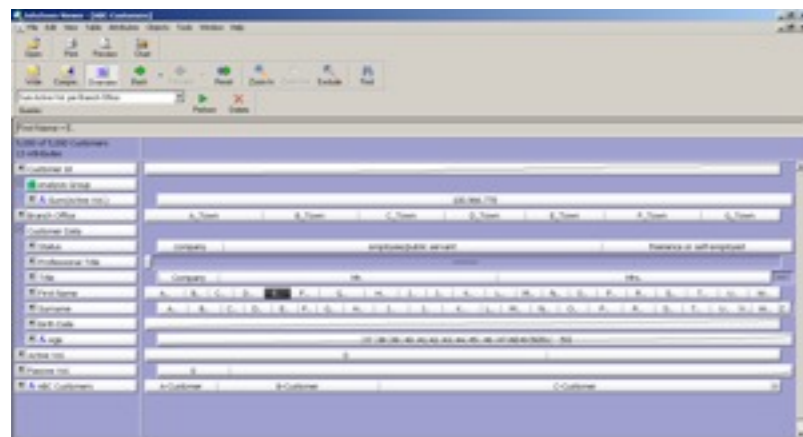
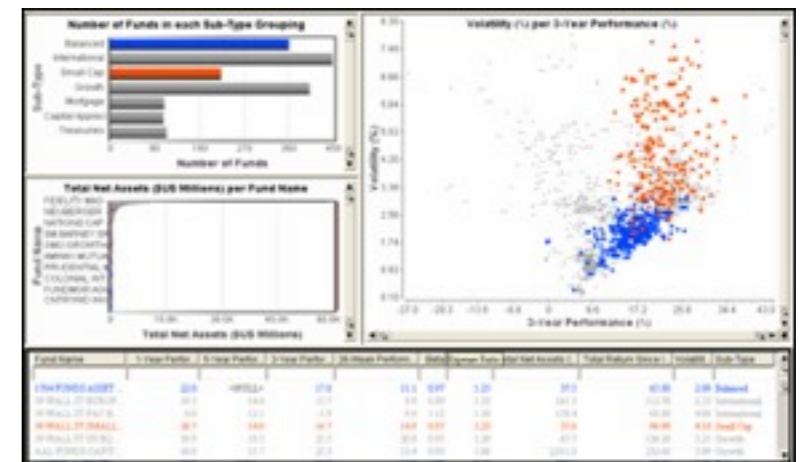
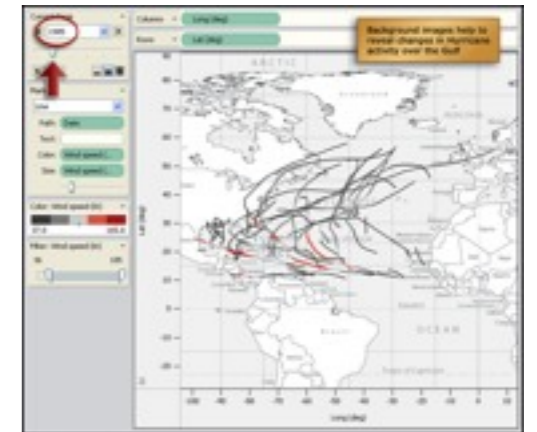


# Example: Exploring Enron

- “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>
- InfoZoom (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

18.10.2011	1 Introduction
25.10.2011	2 Perception
01.11.2011	fällt aus: Allerheiligen
08.11.2011	3 Multidimensional I
15.11.2011	4 Multidimensional II
22.11.2011	5 Interaction
29.11.2011	6 Graphs&Networks
06.12.2011	7 Hierarchies&Trees
13.12.2011	8 Time-Based Data
20.12.2011	w Weihnachtsvorlesung (nicht prüfungsrel.)
27.12.2011	fällt aus: Ferien
03.01.2012	fällt aus: Ferien
10.01.2012	9 Text & Document
17.01.2012	10 Presentation I
24.01.2012	11 Presentation II
31.01.2012	12 AmbientInfoVis
07.02.2012	möglicher Klausurtermin