1. Information Visualization: Introducing Research Need & Concepts

Vorlesung „Informationsvisualisierung”
Prof. Dr. Andreas Butz, WS 2012/13
Konzept und Folien: Thorsten Büring, 3rd, revised edition
Outline of today’s lecture

• The Problem
• The power of visualization
• Classic examples
• Information Visualization
  – Definition
  – Classification
  – Challenges
  – User tasks
  – Reference Model
• Enron example
• Lecture Outline
Literature


Data Explosion

• Progress in hardware technology allows computers to store an increasing amount of large data
• 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
• 2012, Oct 25th: Archive.org „ten petabyte Party“ (http://blog.archive.org/2012/10/10/the-ten-petabyte-party/)
  – 10,000,000,000,000,000 bytes
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)
  - Can not be used at this theoretical limit for raw data!!!
  - 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
Visualization Example 1

• Napoleon’s march to and retreat from Moscow in 1812, Minard 1869 (Translation)
Visualization Example 2

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


red: wounds, blue: preventable disease, gray: other causes
Visualization Example 3

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

[Story: http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak]
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

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

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)

Lie factor: 2.8

Volume of barrels: 59.4

Lie factor: area of barrels: 9.4
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
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
  - 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

From Johnson et al., Univ. of Utah
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

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

**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 Heer

• http://jheer.org/enron
Example: Exploring Enron

Subject: Further information regarding...

From: <steven.kean@enron.com>
Date: 1979-12-31 18:00:00.0
To: <kenneth.ley@enron.com>
Cc: Rosales Fleming <rosales.fleming@enron.com>, <susan.landsehr@enron.com>, <kevin.halder@enron.com>, <don.black@enron.com>

Due Landsehr in government affairs spoke with the administration who indicated that they

are talking to you about Amerak. Specifically, they are looking 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.“
Chairman Barton said it is still his intention to hold H.R. 1647 next week (many on and off this bill, we are gathering political intelligence to be report back with more information and conclude). Chairman Barton also said he may go to California.

There was considerable emphasis at both hearings about higher natural gas prices into California play a major part in the increases; several witnesses focused on different issues. One example was that BPA would be forced to purchase power at a higher price than FERC would have it at market rates and capture the difference. BPA would make it difficult if not impossible to manage prices. The Senate hearing said that today the commission under the Senate conference for later this month on all aspects of the issue related to California; a House member from California has mentioned that BPA has petitioned California for using only natural gas in the state of California. The House witnesses, primarily BPA but also the Senate, attacked the natural gas provision in sec. 102 of the act. It is clear that BPA would be forced to purchase power at a higher price than FERC has it. In summary, 102 would make it difficult if not impossible for BPA to avoid as much as a 200 percent rate increase. The Senate testimony will make it tough for Members of Congress to support sec. 102.

Both hearings also touched on how difficult it is to increase prices in the price cap, at the House hearing, none of those present for the committee could answer the question of how it would be done; the witnesses just said "cost plus", but not the details to FERC. An afternoon Senate hearing, the staff bring in 15 boxes from one FERC case, too long to bring any relief to California. The Senate’s soft price cap is much better.

Also on the price cap, Rep. Walden (R-OR) got to admit that if the price caps had been in place, they have taken the conservation and new generation seriously.

The interplay among the FERC commissioners was a House hearing on Tuesday, although it comes with Senate Chairman Murray (R-AK) said that "The relationship of the nominees for the two vacancies, Sen. Dole, had suggested that the confirmation process will not proceed, and the Senate of the nominees, the nominees are content at times, but:

Please advise if you have any questions or need further assistance.
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
  - Applet: http://download.macrofocus.com/infoscope/
- Advizor Analyst: http://www.advizorsolutions.com/desktop.htm
Lecture Outline

1 Introduction
2 Perception
3 Multidimensional I
4 Multidimensional II
5 Interaction
6 Graphs&Networks
7 Hierarchies&Trees
8 Time-Based Data
  w Weihnachtsvorlesung (nicht prüfungsrel.)
9 Text & Document
10 Presentation I
11 Presentation II
  g Gastvortrag??
12 AmbientInfoVis