Information Visualization: Introducing Research Need & Concepts

Vorlesung „Informationsvisualisierung”
Prof. Dr. Florian Alt, WS 2016/17

Konzept und Folien (7th revised edition)

- Thorsten Büring (LMU München, LFE Medieninformatik)
- Andreas Butz (LMU München, LFE Medieninformatik)
- Michael Burch (Universität Stuttgart, Institut für Visualisierung und Interaktive Systeme)
Outline of today‘s lecture

• Motivation
• Visualization History (19th century — today)
  – Napoleon‘s March to Moscow
  – Diagram of the Causes of Mortality
  – Death from Cholera
  – London Underground Map
• Definitions, InfoVis Goals, and Challenges
• Bad vs. Good Design
  – The Lie Factor
  – Chart Junk
• Visualization Process
Visualization History
Visualization History

• 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

steven.kean@enron.com

1979-12-31 16:00:00.0 Further information regarding this email:
1979-12-31 16:00:00.0 Fle. Trains - Light
1979-12-31 16:00:00.0 Re:
1993-06-20 08:20:00.0 Ross Perot's EMS
1993-10-06 06:44:00.0 Madera Ranch Prior
1999-01-19 01:47:00.0 Translation of artic

To: kenneth.ley@enron.com

Subject: Further information regarding this email:

steven.kean@enron.com

Date: 1979-12-31 15:00:00.0

Steve Landbeck in government affairs spoke today to you about Amtrak. Specifically, D.C. had quite some time on an electricity
Visualization History

• 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 bring the bill, H.R. 1647 next week (many on and off the Hill) and we are gathering political intelligence on the outcome of the coming week's House vote. We are also gathering political intelligence on the coming week's House vote. If you have any questions or additional comments, please feel free to reach out.

There was considerable emphasis at both hearings on the need for higher natural gas prices into California. While the Senate hearing was dominated by the testimony of the Senate hearing, the House hearing was a more civil and informative discussion of the issues.

The Senate hearing featured the testimony of House Energy and Commerce Committee Chairman Barton, who said that the higher natural gas prices into California are due to the California Public Utilities Commission's decision to increase fees to fund the state's main natural gas pipeline, the California Natural Gas Transmission System. He said that the higher fees would result in higher prices for California consumers.

The House hearing featured testimony from the California Public Utilities Commission, which said that the higher fees were necessary to fund the pipeline and that the higher prices would be offset by lower costs for the state's main natural gas pipeline.

Both hearings also touched on the topic of the price cap at the House hearing, one of those occasions, the witnesses could answer excellent questions. The witnesses just said "cost plus a little bit" and then leave the details to the FERC. At the Senate hearing, the staff bring in 15 boxes from one FERC case, that take too long to bring any relief to California. The week's soft price cap is much better. Also on the price cap, Rep. Walden (R-OR) got the FERC admit that if the price caps had been in place recently, they would have taken the commission's new generation energy effective.

The interplay among the FERC commissioners was covered by a House hearing on Tuesday, although it could have been covered by the Senators' hearings. Senator Chairman Burwell (R-RI) said that "the FERC" of the nominees for the two vacancies, Sen. DeWine, suggested that the confirmation process will not be easy. The Senate confirmed the nominations by a margin of 52-48, with Senator B簇n (D-MT) voting against the nominations. The nominees were then seated on the FERC.

The Enron Corporation Viewer is a tool for visualizing email data, allowing users to explore relationships and patterns in large datasets. It highlights key figures and communications, making it easier to understand complex data sets.

Please advise if you have any questions or need further information.

---

Example: Exploring Enron

Chairman Barton said it is still his intention to bring the bill, H.R. 1647 next week (many on and off the Hill) and we are gathering political intelligence on the outcome of the coming week's House vote. We are also gathering political intelligence on the coming week's House vote. If you have any questions or additional comments, please feel free to reach out.

There was considerable emphasis at both hearings on the need for higher natural gas prices into California. While the Senate hearing was dominated by the testimony of the Senate hearing, the House hearing was a more civil and informative discussion of the issues.

The Senate hearing featured the testimony of House Energy and Commerce Committee Chairman Barton, who said that the higher natural gas prices into California are due to the California Public Utilities Commission's decision to increase fees to fund the state's main natural gas pipeline, the California Natural Gas Transmission System. He said that the higher fees would result in higher prices for California consumers.

The House hearing featured testimony from the California Public Utilities Commission, which said that the higher fees were necessary to fund the pipeline and that the higher prices would be offset by lower costs for the state's main natural gas pipeline.

Both hearings also touched on the topic of the price cap at the House hearing, one of those occasions, the witnesses could answer excellent questions. The witnesses just said "cost plus a little bit" and then leave the details to the FERC. At the Senate hearing, the staff bring in 15 boxes from one FERC case, that take too long to bring any relief to California. The week's soft price cap is much better. Also on the price cap, Rep. Walden (R-OR) got the FERC admit that if the price caps had been in place recently, they would have taken the commission's new generation energy effective.

The interplay among the FERC commissioners was covered by a House hearing on Tuesday, although it could have been covered by the Senators' hearings. Senator Chairman Burwell (R-RI) said that "the FERC" of the nominees for the two vacancies, Sen. DeWine, suggested that the confirmation process will not be easy. The nominees were then seated on the FERC.

Please advise if you have any questions or need further information.

---

Example: Exploring Enron

Chairman Barton said it is still his intention to bring the bill, H.R. 1647 next week (many on and off the Hill) and we are gathering political intelligence on the outcome of the coming week's House vote. We are also gathering political intelligence on the coming week's House vote. If you have any questions or additional comments, please feel free to reach out.

There was considerable emphasis at both hearings on the need for higher natural gas prices into California. While the Senate hearing was dominated by the testimony of the Senate hearing, the House hearing was a more civil and informative discussion of the issues.

The Senate hearing featured the testimony of House Energy and Commerce Committee Chairman Barton, who said that the higher natural gas prices into California are due to the California Public Utilities Commission's decision to increase fees to fund the state's main natural gas pipeline, the California Natural Gas Transmission System. He said that the higher fees would result in higher prices for California consumers.

The House hearing featured testimony from the California Public Utilities Commission, which said that the higher fees were necessary to fund the pipeline and that the higher prices would be offset by lower costs for the state's main natural gas pipeline.

Both hearings also touched on the topic of the price cap at the House hearing, one of those occasions, the witnesses could answer excellent questions. The witnesses just said "cost plus a little bit" and then leave the details to the FERC. At the Senate hearing, the staff bring in 15 boxes from one FERC case, that take too long to bring any relief to California. The week's soft price cap is much better. Also on the price cap, Rep. Walden (R-OR) got the FERC admit that if the price caps had been in place recently, they would have taken the commission's new generation energy effective.

The interplay among the FERC commissioners was covered by a House hearing on Tuesday, although it could have been covered by the Senators' hearings. Senator Chairman Burwell (R-RI) said that "the FERC" of the nominees for the two vacancies, Sen. DeWine, suggested that the confirmation process will not be easy. The nominees were then seated on the FERC.

Please advise if you have any questions or need further information.

---

Example: Exploring Enron

Chairman Barton said it is still his intention to bring the bill, H.R. 1647 next week (many on and off the Hill) and we are gathering political intelligence on the outcome of the coming week's House vote. We are also gathering political intelligence on the coming week's House vote. If you have any questions or additional comments, please feel free to reach out.

There was considerable emphasis at both hearings on the need for higher natural gas prices into California. While the Senate hearing was dominated by the testimony of the Senate hearing, the House hearing was a more civil and informative discussion of the issues.

The Senate hearing featured the testimony of House Energy and Commerce Committee Chairman Barton, who said that the higher natural gas prices into California are due to the California Public Utilities Commission's decision to increase fees to fund the state's main natural gas pipeline, the California Natural Gas Transmission System. He said that the higher fees would result in higher prices for California consumers.

The House hearing featured testimony from the California Public Utilities Commission, which said that the higher fees were necessary to fund the pipeline and that the higher prices would be offset by lower costs for the state's main natural gas pipeline.

Both hearings also touched on the topic of the price cap at the House hearing, one of those occasions, the witnesses could answer excellent questions. The witnesses just said "cost plus a little bit" and then leave the details to the FERC. At the Senate hearing, the staff bring in 15 boxes from one FERC case, that take too long to bring any relief to California. The week's soft price cap is much better. Also on the price cap, Rep. Walden (R-OR) got the FERC admit that if the price caps had been in place recently, they would have taken the commission's new generation energy effective.

The interplay among the FERC commissioners was covered by a House hearing on Tuesday, although it could have been covered by the Senators' hearings. Senator Chairman Burwell (R-RI) said that "the FERC" of the nominees for the two vacancies, Sen. DeWine, suggested that the confirmation process will not be easy. The nominees were then seated on the FERC.

Please advise if you have any questions or need further information.
Visualization History

• “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/
Some more daily life examples of InfoVis

[Graphs and maps illustrating various data visualizations]
Definitions and Challenges
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

• What is Information Visualization?
  – The use of computer-supported, interactive, visual representations of abstract data to amplify cognition. [Card et al., 1999]
    • computer-supported: visualization is displayed by a computer, usually on a screen
    • interactive: visualization can be manipulated simply and directly (includes filtering data, drilling down to focus on information)
    • visual representation: information is display in visual form using attributes such as location, length, shape, color, and size of objects to form a picture of the data and thereby allow us to see patterns, trends and exceptions otherwise not visible
    • abstract data: information such as quantitative data, processes or relationships (in contrast to visual representations of physical objects)
    • amplify cognition: ability to think about information by assisting memory and representing the data in ways that our brains can easily comprehend
Definition: Visualization

• What is Information Visualization?
  – The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.
    [Card et al., 1999]
  – Information visualizations attempt to efficiently map data variables onto visual dimensions in order to create graphic representations.
    [Gee et al., 2005]
  – Information Visualization (InfoVis) is the communication of abstract data through the use of interactive visual interfaces.
    [Keim et al., 2006]
  – Information Visualization utilizes computer graphics and interaction to assist humans in solving problems.
    [Purchase et al., 2008]
  – Information Visualization is a set of technologies that use visual computing to amplify human cognition with abstract information.
    [Card, 2008, p. 542]
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

• When is Visualization required?
  – When the goal is to augment human capabilities in situations where the problem is not sufficiently well defined for a computer to handle algorithmically

• When is Visualization not required?
  – If totally automatic solution can completely replace human judgement

(Tamara Munzner)
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
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 according to 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
Good vs. Bad Design
Good vs. Bad Design — The Lie Factor

• Tufte 2001
  – “The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the quantities represented." [Tufte, 1983]
  – ...defined as the ratio of the size of an effect shown in the graphic to the size of the effect in the data

\[
\text{Lie Factor} = \frac{\text{Size of effect in represented graphic}}{\text{Size of actual effect in data}}
\]
Good vs. Bad Design — The Lie Factor

- Magnitude of change mpg: 53%
- Magnitude of the change of line size: 783%
- Lie factor = 14.8
Good vs. Bad Design — The Lie Factor

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

Lie factor: 2.8

Lie factor: area of barrels: 9.4
Volume of barrels: 59.4
Good vs. Bad Design — Chart Junk

• Chart junk
  – “Chart junk refers to all visual elements in charts and graphs that are not necessary to comprehend the information represented on the graph, or that distract the viewer from this information.” [Tufte, 1983]
  – “The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new. The purpose of decoration varies — to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistic skills. Regardless of its cause, it is all non-data-ink or redundant data-ink, and it is often chart junk.” [Tufte, 1983]
Visualization Process
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, ...)
Visualization Pipeline

1. **Data Analysis**: Data is prepared for visualization (e.g., by applying a smoothing filter, interpolating missing values, or correcting erroneous measurements). This is usually computer-centered, little or no user interaction.

2. **Filtering**: Selection of data portions to be visualized. This is usually user-centered.

3. **Mapping**: Focus data is mapped to geometric primitives (e.g., points, lines) and their attributes (e.g., color, position, size). This is the most critical step for achieving expressiveness and effectiveness.

4. **Rendering**: Geometric data are transformed to image data.
Visual Encoding

The process of mapping data to visual variables is called **visual mapping**. Choosing different visual variables for representing different aspects of the same information can greatly influence the perception and understanding of the presented information. It is therefore important to know and appropriately use the characteristics of visual variables when creating any visual data representation.

(used by Mackinlay, 1986, for quantitative data)

*Figure 14: Accuracy Ranking of Quantitative Perceptual Tasks. Higher tasks are accomplished more accurately than lower tasks. Cleveland and McGill empirically verified the basic properties of this ranking.*
Visual Encoding

• Visual Variables are a specified set of symbols that can be applied to data in order to translate information.

• Bertin defined seven different Visual Variables.

![Bertin’s Original Visual Variables](http://www.infovis-wiki.net/images/8/89/VisualVariables.png)
Visualization Resource Limitations

• Technological and algorithmic limitations
  – Computational capacities
    • Is there an efficient algorithm?
  – Display capacities
    • How many objects can be rendered?

• Human limitations
  – Perceptual capacities
    • How many colors can be distinguished?
  – Cognitive capacities
    • How much can be remembered?
Commercial InfoVis Systems

- Tableau Software http://www.tableausoftware.com/
- Spotfire (Tibco): http://spotfire.tibco.com/products/compare-spotfire?capabilities
- InfoZoom (Siemens): http://www.infozoom.com/deu/infozoom/video.htm
  - Applet: http://download.macrofocus.com/infoscope/
- Advizor Analyst: http://www.advizorsolutions.com/desktop.htm
- D3.js: http://d3js.org/
Visualcomplexity.com

- Webpage Visualcomplexity collects data visualizations