Information Visualization Evaluation and User Study

CSE591 Visual Analytics

Lujin Wang

Evaluation for Information Visualization

Enrico Bertini *http://www.dis.uniroma1.it/~beliv06/infovis-eval.html*

- Component/system level evaluation
- Low level components/perceptual studies
- Longitudinal studies, case studies
- Metrics, benchmarks, model-based evaluation, frameworks, taxonomies
- Novel evaluation methodologies, non-conventional methods/ parameters
- Reviews

Evaluation for Information Visualization

- Component/system level evaluation
- Low level components/perceptual studies
- Longitudinal studies, case studies
  - Martin Wattenberg. **Baby names, visualization, and social data analysis.** *INFOVIS '05*
- Metrics, benchmarks, model-based evaluation, frameworks, taxonomies
  - Robert Amar, James Eagan, and John Stasko. **Low-level components of analytic activity in information visualization.** *INFOVIS '05*
- Novel evaluation methodologies, non-conventional methods/ parameters
- Reviews
  - Catherine Plaisant. **The challenge of information visualization evaluation.** *Advanced Visual Interfaces '04*

The Challenge of Information Visualization Evaluation

Catherine Plaisant

Human-Computer Interaction Laboratory
University of Maryland at College Park

Advanced Visual Interfaces '04

Introduction

- The tools and ideas in information visualization research publications are reaching users
  - Commercial products
    - Spotfire, Inxight, and HumanIT
  - Additions to statistical packages
    - SPAA/SigmaPlot, SAS/GRAPH, DataDesk
  - Commercial development environments
    - ILOG JViews
  - Others
    - SmartMoney financial maps, Peet's Coffee Selector, health information maps, highway traffic information

Information Visualization Evaluation

- The reports of usability studies and controlled experiments are helpful to understand the potential and limitations of tools
- However, we need
  - Consider other evaluation approaches that take into account
    - the long exploratory nature of users tasks
    - the value of potential discoveries
    - or the benefits of overall awareness
  - Better metrics and benchmark repositories to compare tools
  - Seek reports of successful adoption and demonstrated utility
- We need to understand how to improve evaluation methods in order to present actionable evidence of measurable benefits that will ***encourage more widespread adoption***

Current Evaluation Practices

Four thematic areas of evaluation: (fifty user studies of InfoVis system)

- *Controlled experiments* comparing *design elements*
  - Comparing specific widgets, mappings of information to graphical display
- *Usability evaluation* of a tool
  - Provide feedback on problems of a tool, show how to refine the design
- *Controlled experiments* comparing *two or more tools*
  - Compare a novel technique with the state of the art
- *Case studies* of tools in realistic settings
  - Do real tasks, demonstrate feasibility and in-context usefulness
  - Time consuming, results may not be replicable and generalizable

Challenges

- Matching tools with users, tasks and real problems
  - Using real datasets and demonstrating realistic tasks is important
- Improving user testing
  - Looking at the same data from different perspectives , over a long time
  - Answering questions you didn't know you had
  - Factoring in the chances of discovery and benefits of awareness
- Addressing universal usability
  - General public, diverse users

Possible Next Steps

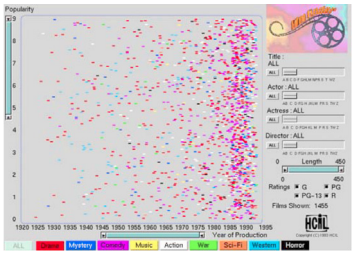
- Development of repositories of data and tasks
  - Create benchmark datasets and tasks
  - InfoVis 2003, 2004, 2005, 2006 contest, VAST 2006 contest
  - Information Visualization Benchmark Repository
    http://www.cs.umd.edu/hcil/InfovisRepository/index.shtml
- Gathering of case studies and success stories
- Strengthening the role of toolkits and development tools
  - Making a technique a part of a toolkit greatly improves its chance of success

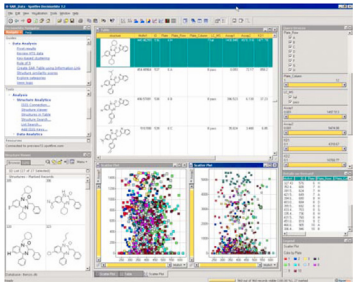
Examples of Technology Transfer

- A common evaluation measure for any technology is adoption by others, and the move into commercial products
- Commercial success depends on
  - The quality of a product
  - Financial alliances
  - Marketing strategies
  - Personal networking
  - Luck
- Gain lessons from examples of transformations from prototypes to products and applications

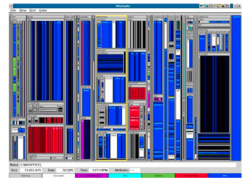
Example 1: From the Film Finder to Spotfire

- A 11-year-long voyage
- Dynamic scattergrams
- Other visualizations such as parallel coordinates, table views or standard business graphics were combined





Example 2: Treemap

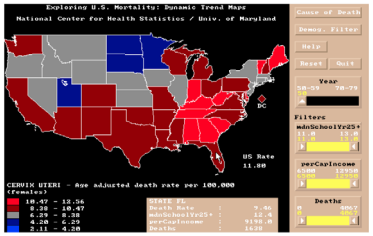


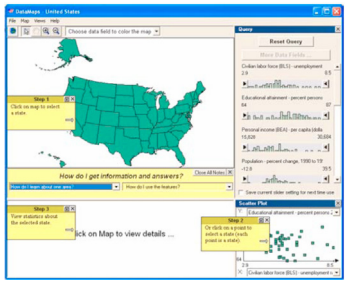
- SmartMoney Map of Market
  - The first widely known commercial application of treemaps

- Hive Group developed an application used by US marines for inventory management
  - The availability of tools for data preparation and publishing, provides automatically updated views
  - A simplified interface allows end-users to view the data and perform limited filtering and grouping operations

Example 3: DataMap

- 1993 Dynamap
  - For the National Center for Health Statistics
- DataMap
  - Might be used by the Census Bureau to release data on CDs





Summary

- Research prototypes are finding ways into commercial products
- We must carefully integrate the visualization tools into solutions for real problems
  - Facilitating the importation of data
  - Coping with large volumes of incomplete data
  - Enabling users to integrate with other tools and collaborate with others
  - Reporting on long term use in natural settings
- We should promote field studies, investigate new evaluation procedures, and celebrate successes

Low-Level Components of Analytic Activity in Information Visualization

Robert Amar, James Eagan, and John Stasko

College of Computing/GVU Center
Georgia Institute of Technology

INFOVIS '05

Goal

- Present a set of primitive analysis task types
  - **Representative of the kinds of specific questions** that a person may ask when working with a data set
  - Serves as **a form of common language or vocabulary** when discussing the capabilities, advantages, and weakness of different information visualization systems
  - Fosters **an emphasis on the importance of analytic measures** of information visualization systems
  - Aids designers in creating novel presentations that amplify users' analytic abilities
  - Serves as **an informal checklist** to assess and evaluate new systems and techniques

The Nature of Analytic Activity

- User analyzes questions and tasks as part of analytic activity typically range from broader "high-level" goals to much more specific "low-level" inquiries
  - For example, a person studying the history of motion picture films may have "high-level", uncertainty-tinged knowledge goals such as
    - *understanding trends in popularity over time* or
    - *determining how to predict which movies will win Academy Awards.*

    In the process of acquiring this knowledge, the person may generate more specific, low-level queries such as
    - *identifying the Academy Award-winning pictures of the past ten years* and
    - *determining whether or not movie length correlates to the film's popularity.*

Concrete Tasks for Analytic Activity

- Data Collection
  - Students are asked to generate data analysis questions for provided data sets, and then evaluate how well the questions could be answered using particular commercial visualization tools

Data Set	Dimensionality	Cardinality	Questions Generated
Cereals	15	78	43
Mutual funds	14	987	14
Cars	10	407	53
Films	10	1742	47
Grocery surveys	8	5164	39

- Analysis Approach
  - Affinity diagramming approach
  - Grouping similar questions and iteratively refining the groups

Analytic Task Taxonomy

Ten primitive analysis task types:

- Retrieve Value
- Filter
- Compute Derived Value
- Find Extremum
- Sort
- Determine Range
- Characterize Distribution
- Find Anomalies
- Cluster
- Correlate

Analytic Task Taxonomy

- Retrieve Value
  - *What are the values of attributes { X, Y, Z, …} in data cases {A, B, C, …}?*
- Filter
  - *Which data cases satisfy conditions {A, B, C…}?*
- Compute Derived Value
  - *What is the value of aggregation function F on a given set S of data cases?*
- Find Extremum
  - *What are the top/bottom N data cases with respect to attribute A?*
- Sort

Analytic Task Taxonomy

- Determine Range
  - *What is the range of values of attribute A in a set S of data cases?*
- Characterize Distribution
  - *What is the distribution of values of attribute A in a set S of data cases?*
- Find Anomalies
  - *Which data cases have unexpected/exceptional values?*
- Cluster
  - *Which data cases are similar in value for attributes {X, Y, Z, …}?*
- Correlate
  - *What is the correlation between attributes X and Y over a given set S of data cases?*

Discussion

- Compound Tasks
  - Compositions of primitive tasks
- Omissions from the Taxonomy
  - Low-level mathematical and cognitive actions
  - High-level questions
  - Uncertain criteria – involves a more subjective evaluation of a data case
- Concerns
  - Similarity to Existing Taxonomies
  - Relationships to high-level tasks
  - Methodological concerns

Baby Names, Visualization, and Social Data Analysis

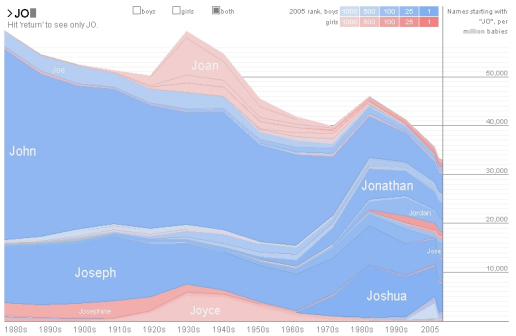
Martin Wattenberg

IBM Research

INFOVIS '05

The NameVoyager

- A web-based visualization of historical trends in baby naming
  http://babynamewizard.com/namevoyager/lnv0105.html

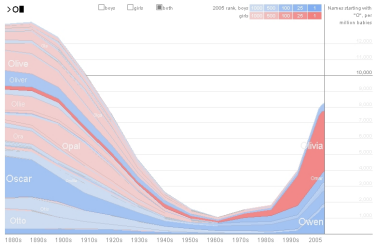


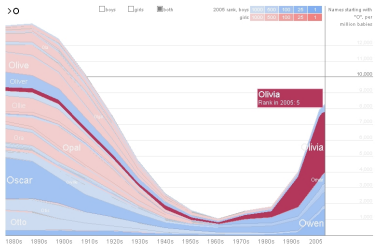
The NameVoyager

- A web-based visualization of historical trends in baby naming
  http://babynamewizard.com/namevoyager/lnv0105.html
- Data *(Book: The Baby Name Wizard)*
  - Public Social Security Administration (SSA) information that tracks baby name trends in US
  - Lists of the most popular 1,000 boys and girls names, for each decade since 1900 and each year since 2001
  - A data set containing popularity time series for 6,000 distinct names
- The Java applet has proven extremely popular
  - More than 500,000 visits in the first two weeks
  - Thousands of comments about the visualization written on the web

Visualization

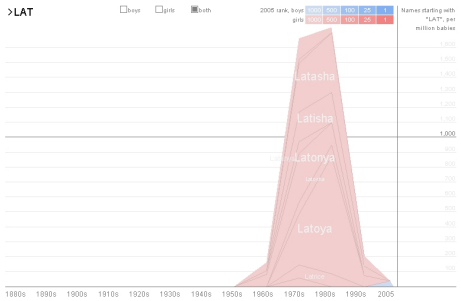
- Given a set of name popularity time series, a set of stacked graphs is produced
- Smooth animated transitions
- Color





Interaction

- "Overview first, zoom and filter, details on demand." (*Shneiderman's mantra)*
- Keyboard-based mechanism for filtering the view



Social Data Analysis

- Web comments show that exploring the data has become a social activity

  One writes, "*For a challenge, try finding a name that was popular at the beginning of the sample (around 1900), went out of style, then came back into vogue recently*"

  Another person responds, "*Take a look at Grace, #18 in the 1900s, #13 in 2003, and down in the 200s and 300s during midcentury*"

  A third writes "*1900's comeback: Porter.*"…..

Social Data Analysis

- Users seem to fall into *Richard Bartle's taxonomy of players* in multi-user online games (MUD)
- Roles in social data analysis
  - Achievers
    - Use the applet to find the good name
  - Socializers
    - Interact with others, use it for conversation and storytelling
  - Explorers
    - Unearth odd names or unusual clusters
  - Killers
    - Take pleasure in singling out names for ridicule

Design Hypotheses for Social Data Analysis

- The NameVoyager encourages a social style of data analysis
- Three hypotheses (the NameVoyager's popularity)
  - Common ground principle
    - Some degree of common knowledge of the underlying data set is necessary
  - Personal perspective principle
    - It is helpful for each person to have a naturally unique perspective on the data
    - Each person is approaching the data in a different way, a group may collectively explore more pieces of the data
  - Deep pointers
    - It allows people to share the state of the visualization at any point in their explorations

Summary

- This paper explores the reaction to the NameVoyager, using web comments as evidence
  - This methodology is unusual
  - However, the online discussion provides a useful source of detailed descriptions from real users, and is fruitful source of hypotheses about how and why the tool is effective
- To design a successful exploratory data analysis tool, one good strategy is to create a system that enables "social" data analysis

Conclusion

- Review
  - The challenge of information visualization evaluation
- Taxonomy
  - Low-level components of analytic activity in information visualization
- Longitudinal study, case study
  - Baby names, visualization, and social data analysis