CSE 412 - Intro to Data Visualization

Evaluation

Jane Hoffswell  University of Washington
How do we determine if a visualization is effective?
Example: Tree Browsers
Evaluation Methods

Inspection or Principled Rationale
Apply design heuristics, perceptual principles

Informal User Study
Have people use visualization, observe results

Controlled Experiment
Choose appropriate tasks / users to compare
Choose metrics (time, error, what else?)
Evaluation Methods

Field Deployment or Case Studies
Observation and Interview
Document effects on work practices

Theoretical Analysis
Algorithm time and space complexity

Benchmarks
Performance (e.g., interactive frame rates)
Scalability to larger data sets
Topics

Focus+Context (Trees, Spatial Navigation)
Perceptual Organization of Graphs
Data Density of Time Series
Trees
The Great Browse-Off! [CHI 97]

Microsoft File Explorer

vs.

Xerox PARC Hyperbolic Tree
Which visualization is better?
Which visualization is better?

Xerox PARC researchers ran eye-tracking studies to investigate... [Pirolli et al 00]
Which visualization is better?

Xerox PARC researchers ran eye-tracking studies to investigate... [Pirolli et al 00]

Subjects performed both retrieval and comparison tasks of varying complexity.
Which visualization is better?

Xerox PARC researchers ran eye-tracking studies to investigate... [Pirolli et al 00]

Subjects performed both retrieval and comparison tasks of varying complexity.

No significant performance differences were found across task conditions.
How do users navigate the tree?
How do users navigate the tree?
They read the labels!

Microsoft File Explorer vs. Xerox PARC Hyperbolic Tree
How do users navigate the tree?

**Information Scent:** A user’s (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues. [Pirolli & Card 99]
How do users navigate the tree?

**Information Scent:** A user’s (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues. [Pirolli & Card 99]

**Operationalize as:** the proportion of participants who correctly identified the location of the task answer from looking at upper branches in the tree.
An Adaptive Field of View?
More Evaluations
Evaluation of DOI Trees

DOI Tree vs. Windows Explorer [Budiu, AVI 06]

Nodes visited (avg)  DOI:83  Exp:53  p<.005
Revisitation (avg)    DOI:6.6  Exp:8.2  p<.005
Divergence (avg)     DOI:4.6  Exp:3.9  p<.001

DOI Tree more forgiving to navigation errors
BUT no significant difference in task time

DOI Tree vs. Google Directory [Pirolli, CHI 06]
DOI Tree has superior task knowledge transfer
Design Guidelines
Design Guidelines

Support rapid visual scanning
Most people don’t read in circles!

Degree of Interest Trees  [Heer & Card 04]
Design Guidelines

People don’t read in circles!

Showing more is not always better

Distractors can decrease task performance

Interaction with quality of information scent
Design Guidelines

People don’t read in circles!
Showing more is not always better

Navigation cues critical to search
Informative labels or landmarks needed
Poor *information scent* undermines search
Lessons Learned

Both task and data properties (here, information scent) may interact with the visualization type in unexpected ways.

Equal performance in terms of accuracy or response time is not the whole picture. We often require more detailed study!
Spatial Navigation
An Evaluation of Pan & Zoom and Rubber Sheet Navigation with and without an Overview

Dmitry Nekrasovski, Adam Bodnar, Joanna McGrenere, François Guimbretière, Tamara Munzner
Pan & Zoom vs. Rubber Sheet

(i) PZN

(ii) RSN
Experimental Task

Compare topological distance between nodes in a dendrogram.
Experiment

Compare performance in 4 conditions:
1. Pan & Zoom (no overview)
2. Pan & Zoom (with overview)
3. Rubber Sheet (no overview)
4. Rubber Sheet (with overview)

40 subjects (24F/16M), between 18-39 years old. Right-handed, normal vision. Between-subjects design.
1. Rubber Sheet / No Overview
2. Pan & Zoom / No Overview
3. Rubber Sheet / Overview

Which node is the purple node closer to in terms of topological distance?

- Blue
- Green
- Submit

Drag with LEFT mouse button to ZOOM IN
Drag with RIGHT mouse button to ZOOM OUT
Press R to RESET the visualization
Press ESCAPE to CLEAR the current mouse drag
4. Pan & Zoom / Overview
Hypotheses

1. RSN interfaces perform better than PZN interfaces independently of the presence or absence of an overview.
2. For RSN, the presence of an overview does not result in better performance.
3. For PZN, the presence of an overview results in better performance.
Results: H1 False

Figure 7: Mean completion times per trial for each interface by block in seconds (N=40).
Results: H2 True, H3 False

Figure 9: Block 7 mean per-trial completion times in seconds by navigation technique with and without an overview.
Results

R1. Pan & Zoom had lower completion times, navigation actions, resets, and reported mental demand.

R2. Overview has no significant impact on rubber sheet navigation, though it was reported to reduce physical demand.

R3. Overview has no significant impact on pan & zoom navigation, though it was reported to reduce physical demand.
Thoughts?

Does this generalize for overview displays?
Perceptual Organization of Node-Link Diagrams
Perceptual Organization of Graphs

Circular

Force-Directed
Experiment Design

Factors
Circular or Force-Directed Seed Layout
# of Between-Cluster Edges ("masking")
  All graphs had two primary clusters

Measures
# of Edge Crossings
Average Edge Length
Average Node Distance
  within or between clusters
Circular
Force-Directed
Figure 4. Edge Crossings. Human observers produced graph layouts with fewer edge crossings than the force-directed graph algorithm.
Figure 5. Edge Length Distribution. Human observers did not focus on maintaining equal edge length as much as the force directed algorithm.
Figure 7. Cluster Extraction. For all levels of masking, the distance between nodes within a cluster is significantly smaller than the overall inter-node distance, demonstrating perceptual grouping. Error bars show 95% confidence intervals.
Figure 9. Cluster Hulls. Two examples of user-generated layouts where cluster edges formed a hull enclosing the cluster, organizing it into a single perceptual group.
Administrivia
Final Project Deliverables

Demonstration Video (<= 2 min)
Due on YouTube & Canvas by midnight Wed 3/10.

Communicate topics and project goals
Do: Show what viewers can learn from your page
Don't: Enumerate every feature of the page

Video should include: project name, team members' names, link to your website

For other tips, see the video production guide!
Final Project Deliverables

**Demonstration Video** (<= 2 min)
Due on YouTube & Canvas by midnight Wed 3/10.

**Final Project Showcase**
We will show demo videos in class, Fri 3/12.

**Interactive Web Page & GitHub Repo**
All materials online by midnight Mon 3/15.

Read assignment description for more!
Course Evaluation

Official course evaluation, due by 3/14
Your opinion is valued!

https://uw.iasystem.org/survey/236202
Data Density
Data Density = \frac{(# \text{ entries in data})}{(\text{area of graphic})}

“Graphical excellence... gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space”

[Tufte 83]
Relative Technology Stock Performance: Jan 2008 - Present
Relative Technology Stock Performance: Jan 2008 - Present
Relative Technology Stock Performance: Jan 2008 - Present
Relative Technology Stock Performance: Jan 2008 - Present
Relative Technology Stock Performance: Jan 2008 - Present

AMZN
CSGO
GOOG
IBM
INTC
MSFT
NOK
ORCL
QCOM
YHOO
Relative Technology Stock Performance: Jan 2008 - Present

AMZN
CSCO
GOOG
IBM
INTC
MSFT
NOK
ORCL
QCOM
YHOO
Relative Technology Stock Performance: Jan 2008 - Present
Relative Technology Stock Performance: Jan 2008 - Present
Relative Technology Stock Performance: Jan 2008 - Present
Horizon Graphs

Segment Peaks

Layer Segments

Mirror Negative Values
Relative Technology Stock Performance: Jan 2008 - Present

AMZN
CSCO
GOOG
IBM
INTC
MSFT
NOK
ORCL
QCOM
YHOO
Relative Technology Stock Performance: Jan 2008 - Present
Relative Technology Stock Performance: Jan 2008 - Present

AMZN
CSCO
GOOG
IBM
INTC
MSFT
NOK
ORCL
QCOM
YHOO
Relative Technology Stock Performance: Jan 2008 - Present

AMZN
CSCO
GOOG
IBM
INTC
MSFT
NOK
ORCL
QCOM
YHOO
Relative Technology Stock Performance: Jan 2008 - Present
Experiment: Chart Type & Size

Q1: How do mirroring and layering affect estimation time and accuracy compared to line charts?

Q2: How does chart size affect estimation time and accuracy?
Estimate the difference between T and B (0-200) to within 5 values.
3 (chart type) x 4 (size) within-subjects design

- N = 30 (17 male, 13 female), undergrads
- 14.1 inch LCD display, 1024 x 768 resolution
- At scale = 1, chart is 13.9 x 1.35 cm (48 px)
Experiment Design

3 (type) x 4 (size) within-subjects design
N = 30 (17 male, 13 female), undergrads

2 (type) x 3 (size: 1/8, 1/12, 1/24) follow-up
N = 8 (6 male, 2 female), engineering grads
Virtual Resolution (VR)

The un-mirrored, un-layered height of a chart

\[ VR = h \]

\[ VR = 2h' = h \]

\[ VR = 4h'' = h \]
Experiment Results

**Q1**: 2-band horizon graph (but not mirrored graph) has higher baseline estimation time and error.

**Q2**: Estimation error increases as the virtual resolution decreases. Estimation time decreases as the physical height decreases.
Design Guidelines

Mirroring does not hamper perception
Design Guidelines

Mirroring does not hamper perception

Layered bands beneficial for smaller charts

2-band mirror charts more accurate for heights under 6.8mm (24 pixels @ 1024x768)

Predict benefits for 3 bands under 1.7mm (6 px)
Design Guidelines

Mirroring does not hamper perception
Layered bands beneficial for smaller charts

Optimal chart sizing

**Sweet spots** in time/error curves
6.8mm (24 px) for line chart & mirrored chart
3.4mm (12 px) for 2-band horizon graph
FOLLOW-UP QUESTION: What other tasks and performance measures should one test?
Summary

Design and analyze visualization techniques in context of real-world use. Time/error analyses can be insightful, but they don’t provide a complete picture. Performance measures may be more suited to serious analysis than casual use?
Encoding

Data — Task

Users & Domain