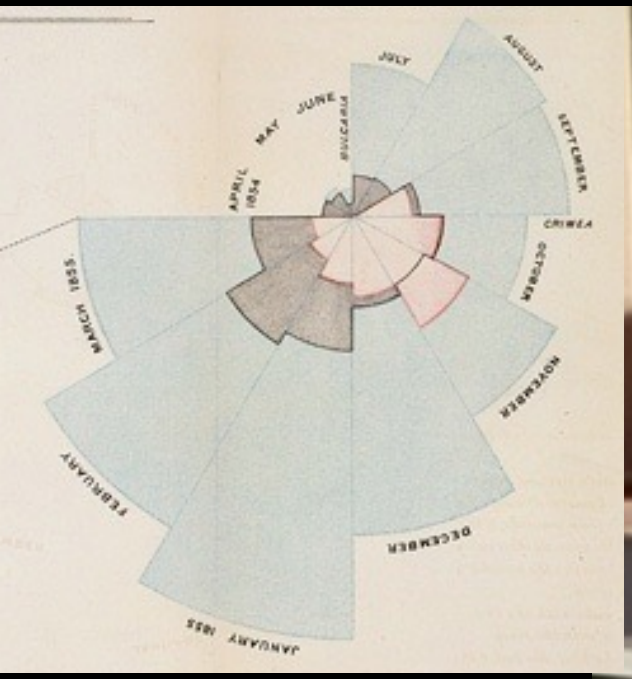
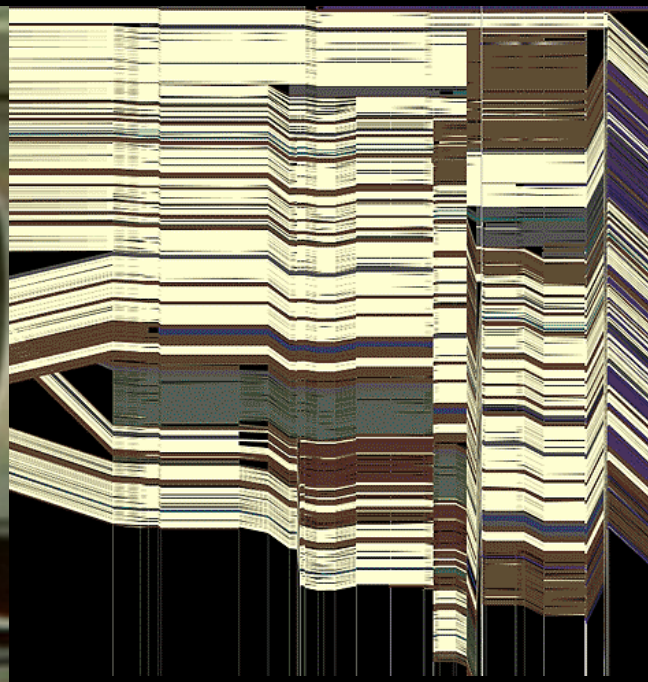
CSE 512 - Data Visualization

Evaluation

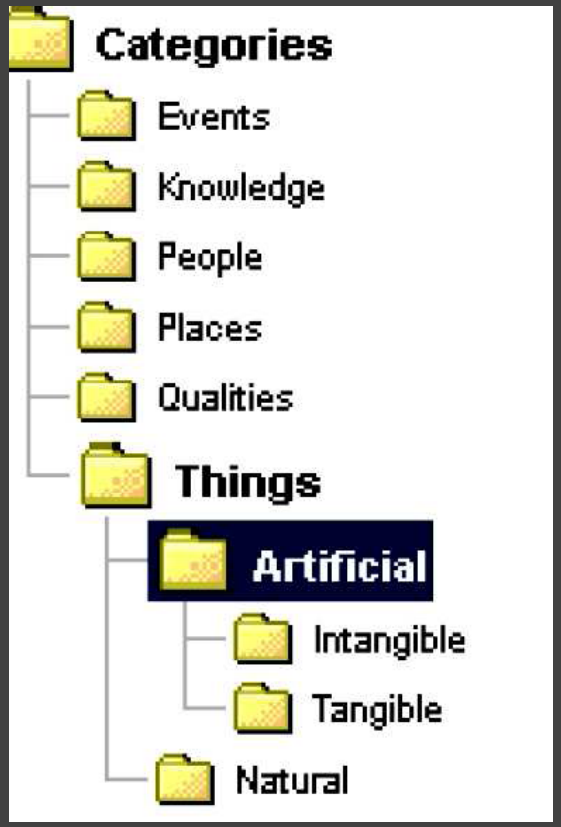




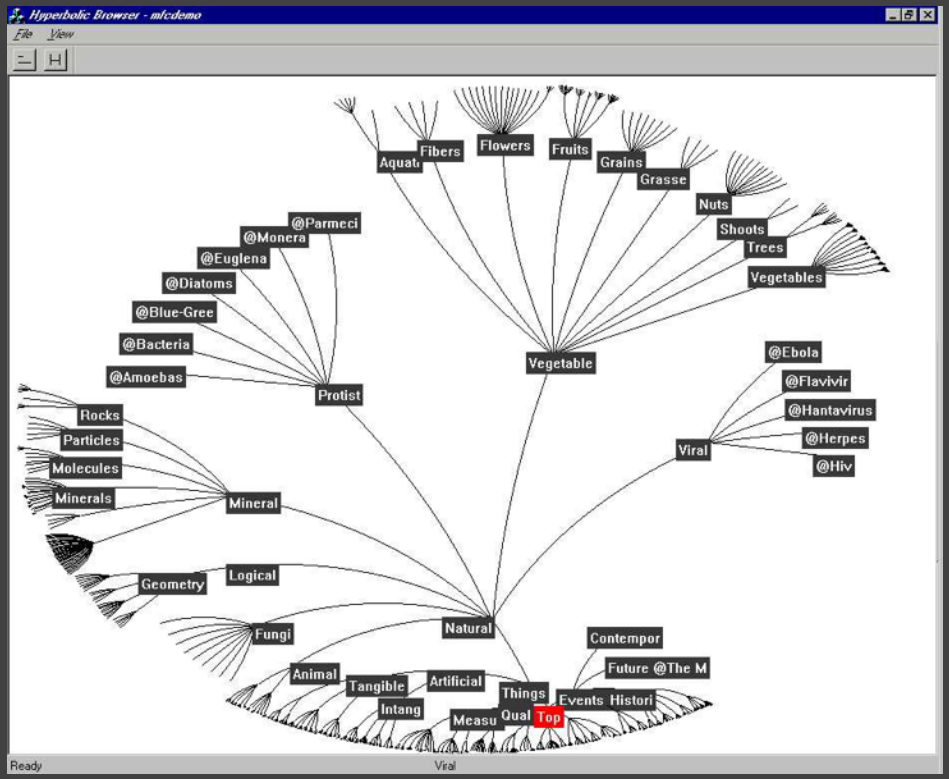
Jeffrey Heer University of Washington

How do we determine if a visualization is *effective*?

Example: Tree Browsers



vs.



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 Methods (Trees, Spatial Navigation)

Data Density of Time Series

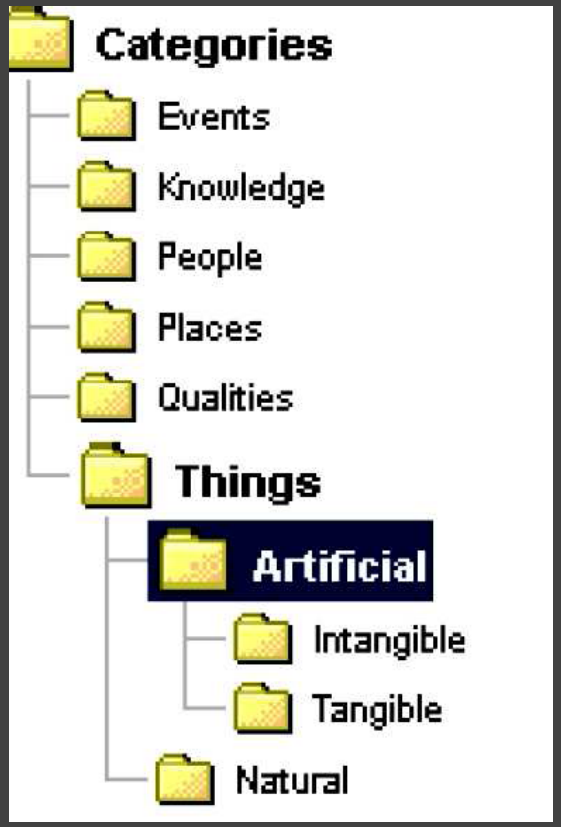
Perceptual Organization of Graphs

Contextual Effects on Visualization Use

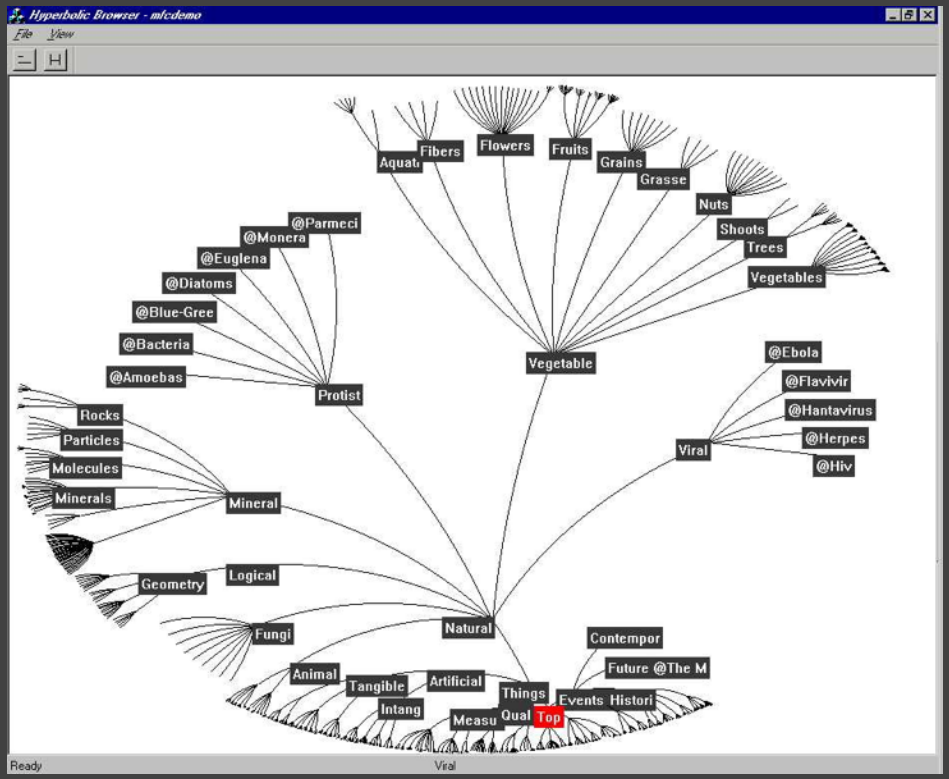
Discussion and Course Evaluation

Trees

The Great Browse-Off! [CHI 97]



vs.



Microsoft File Explorer

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]

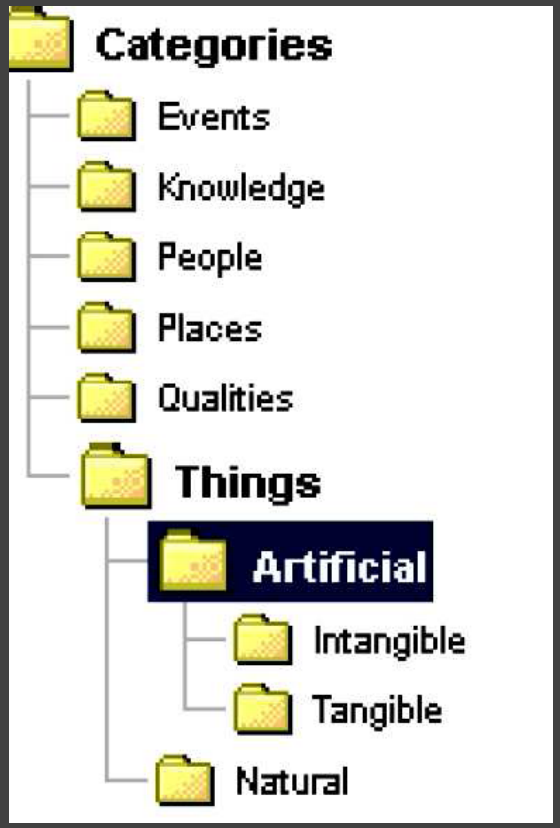
Subjects performed both retrieval and comparison tasks of varying complexity.

No significant performance differences were found across task conditions.

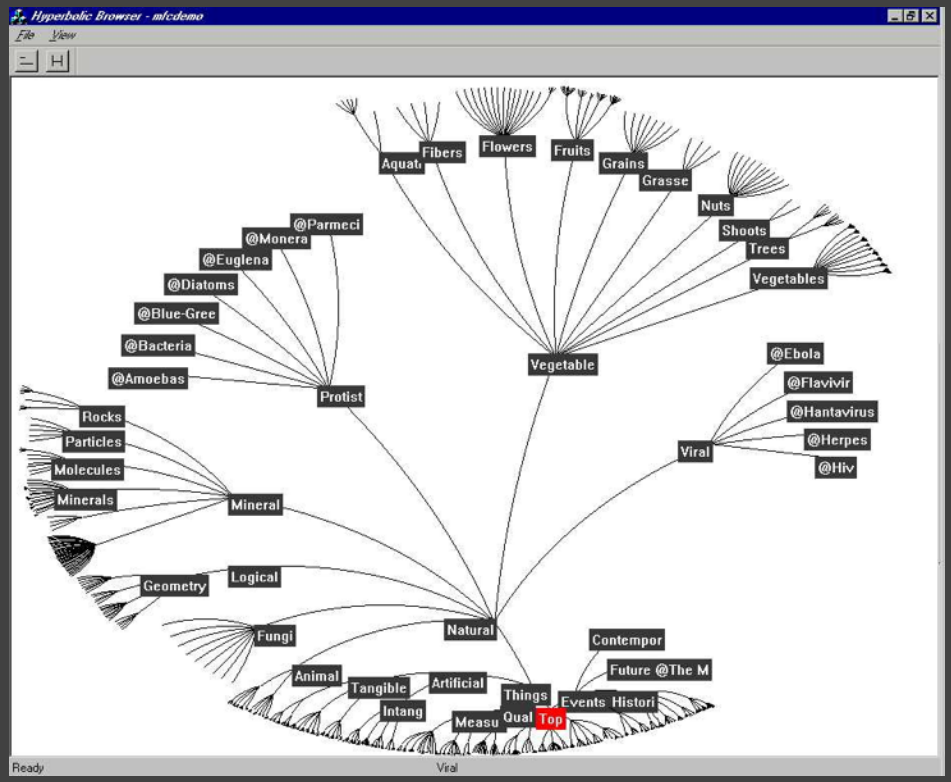
How do users navigate the tree?

How do users navigate the tree?

They read the labels!



vs.



Microsoft File Explorer

Xerox PARC Hyperbolic Tree

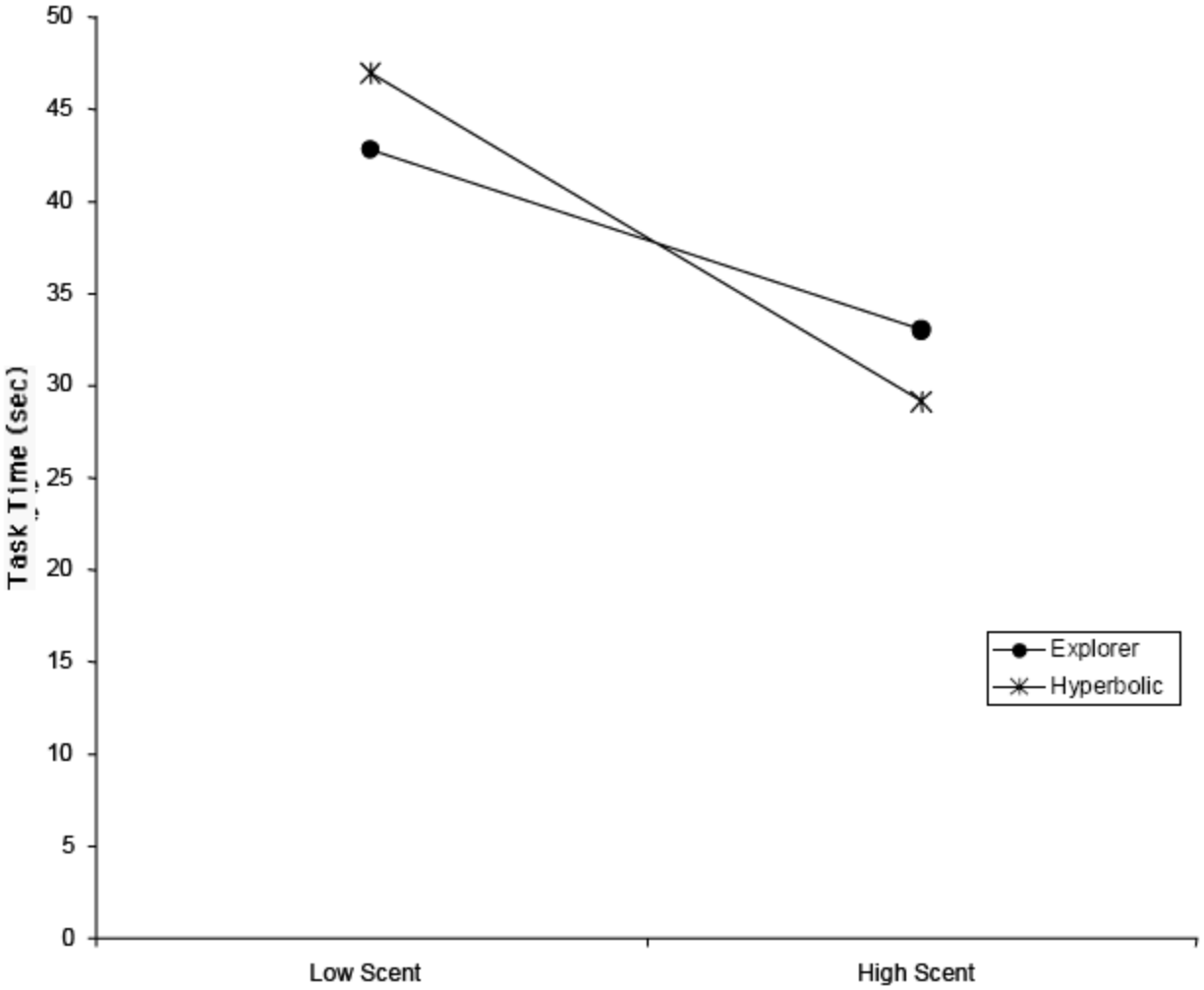
How do users navigate the tree?

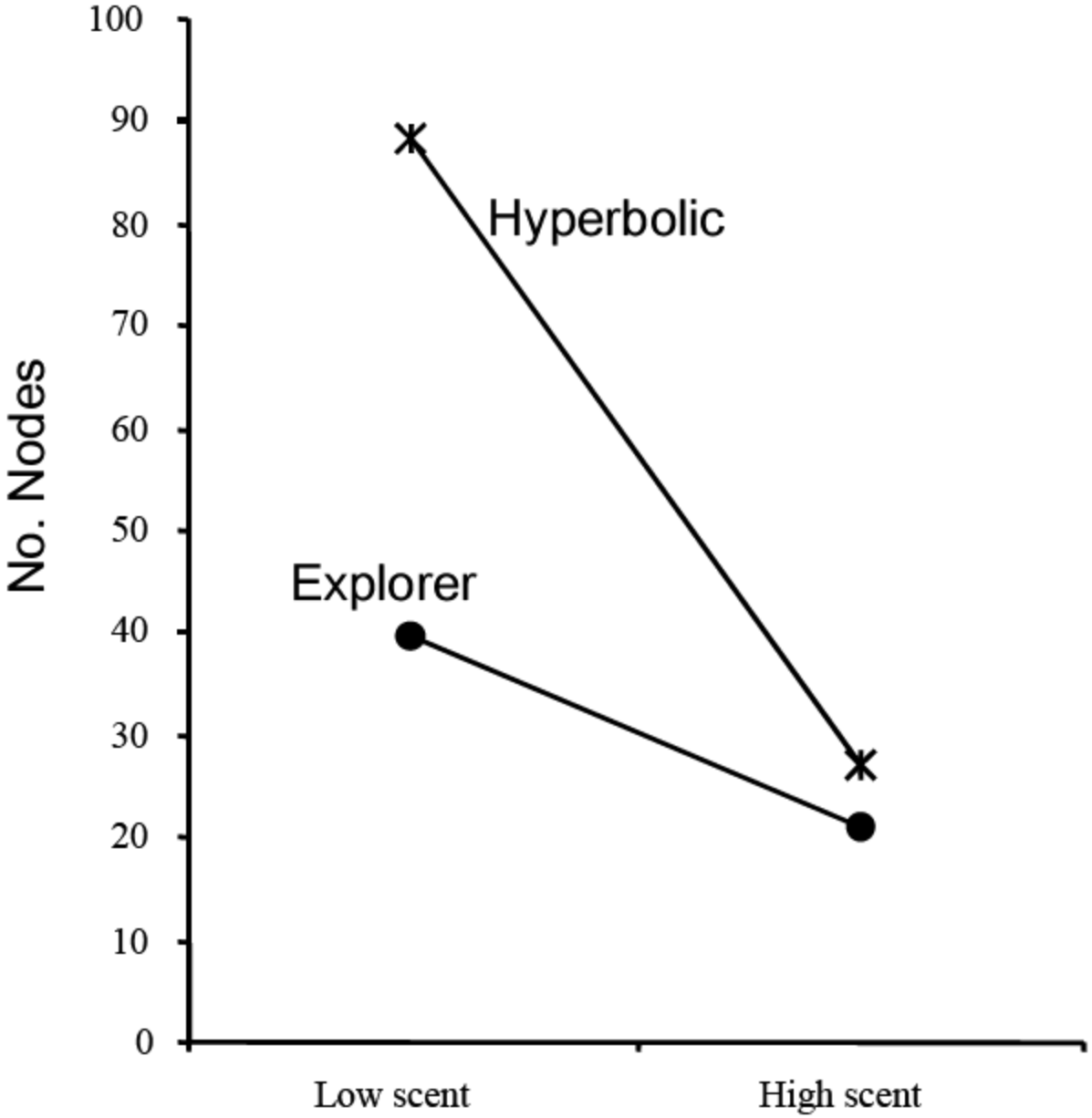
Information Scent: A user's (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues. [Pirolli & Card 99]

How do users navigate the tree?

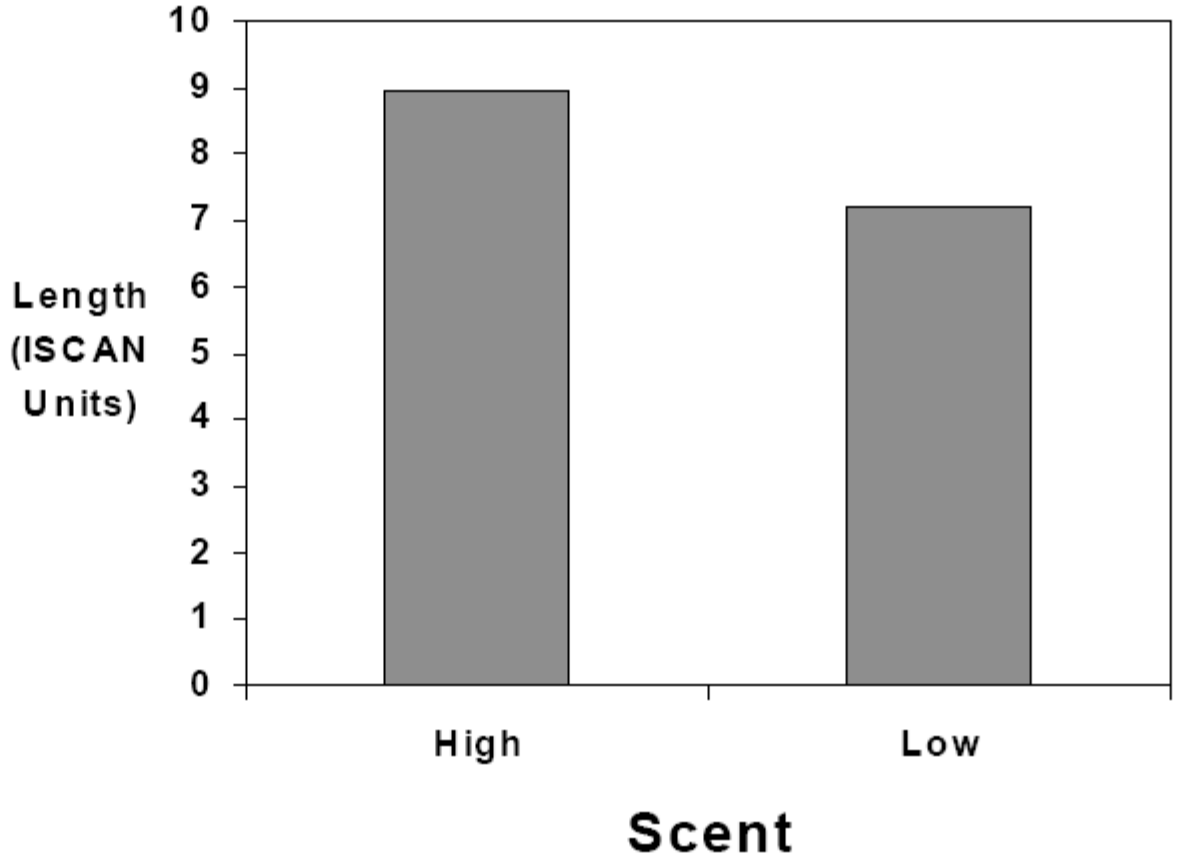
Information Scent: A user's (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues. [Pirolli & Card 99]

Operationalize as: the proportion of participants who correctly identified the location of the task answer from looking at upper branches in the tree.

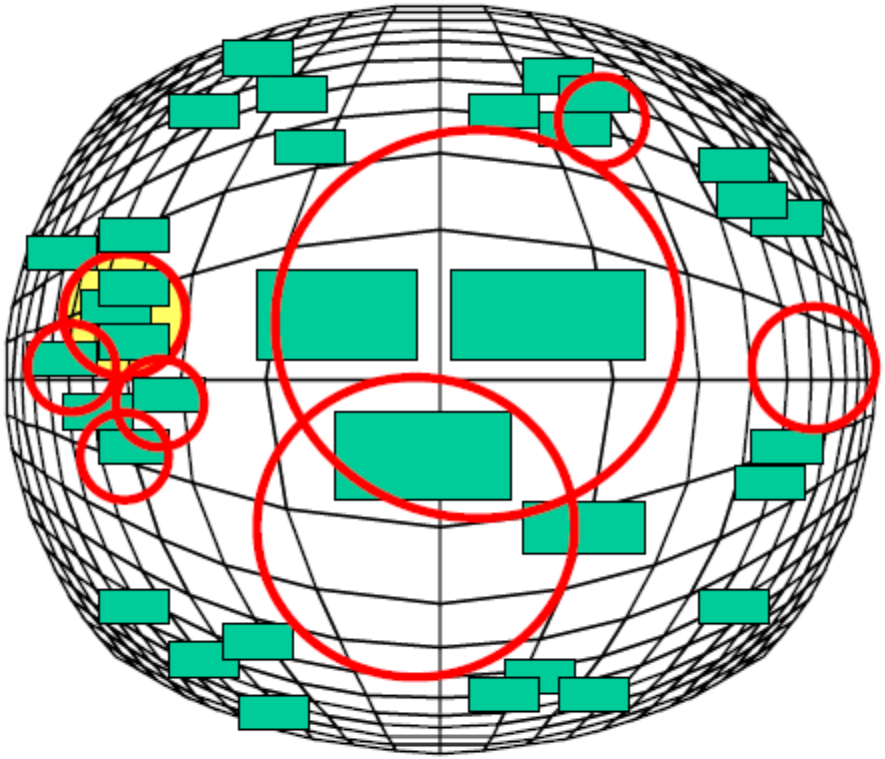

50
45
40
35
30
25
20
15
10
5
0
Task Time (sec)
Low Scent
High Scent
Explorer
Hyperbolic

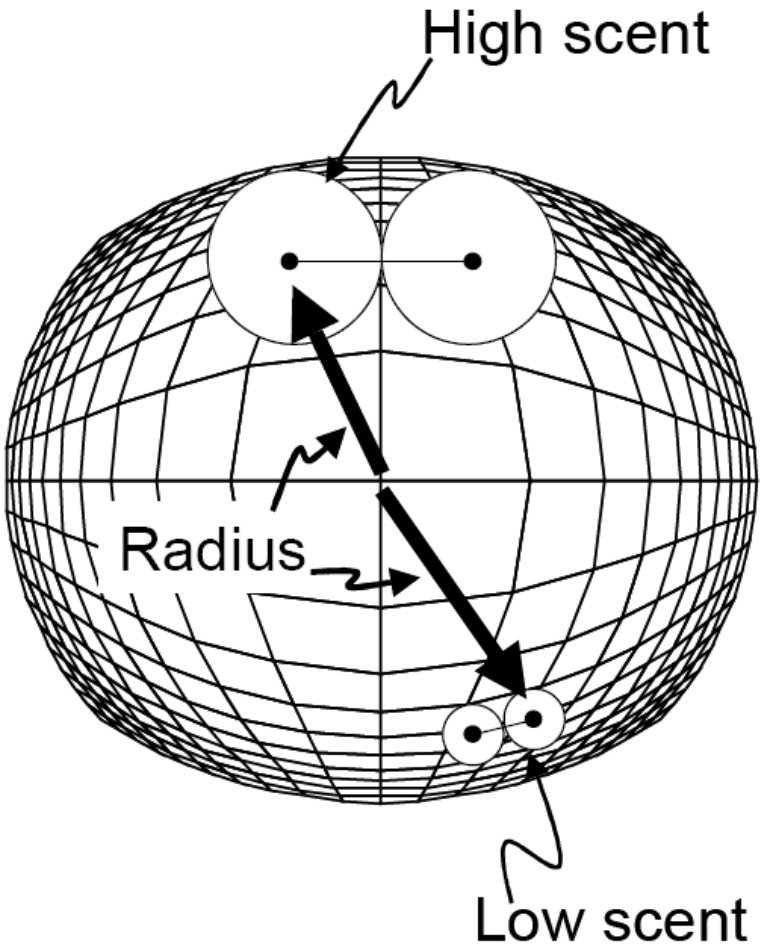

100
90
80
70
60
50
40
30
20
10
0
No. Nodes
Hyperbolic
Explorer
Low scent
High scent

Length of Eye Movements

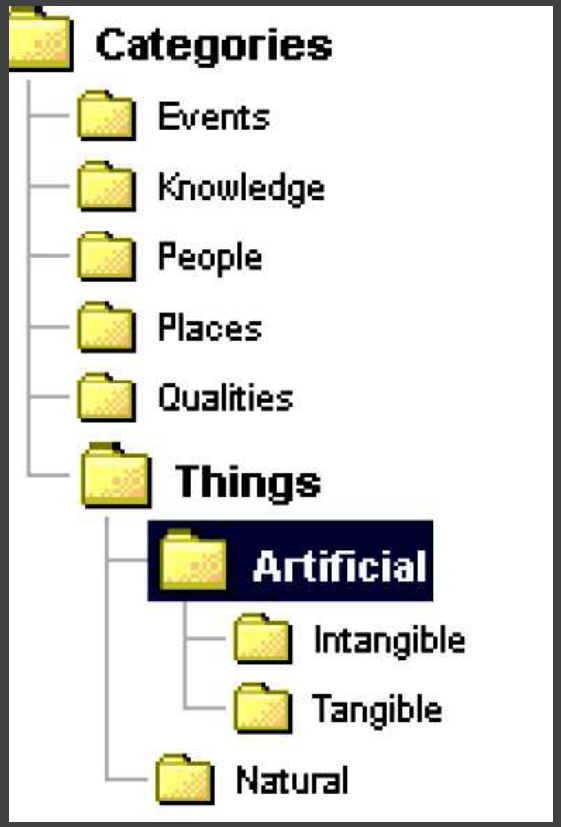


An Adaptive Field of View?

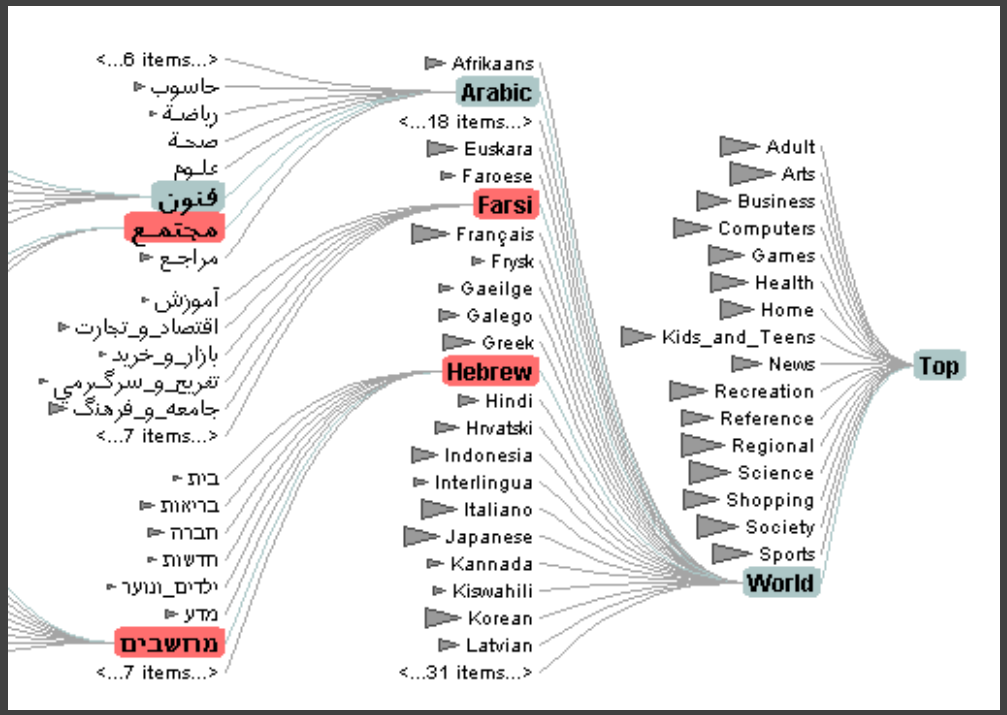




More Evaluations



vs.



Evaluation of DOI Trees

DOITree 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*

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

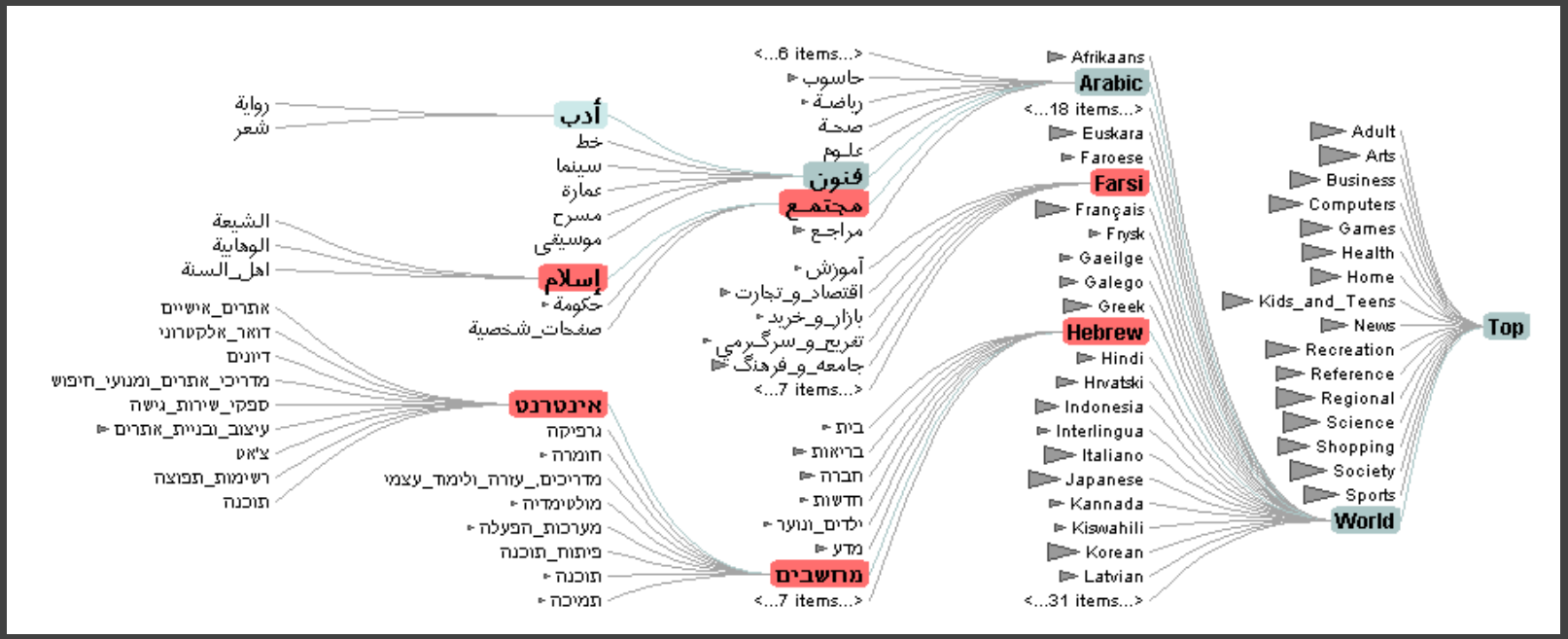
DOITree vs. Google Directory [Pirolli, CHI 06]
DOITree has superior task knowledge transfer

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.

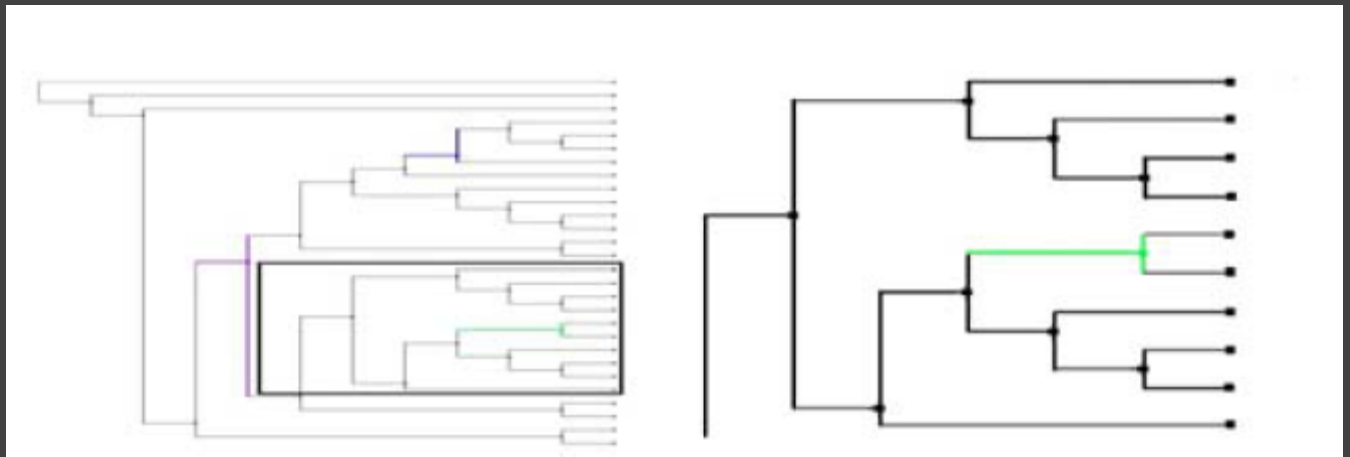
Equal **performance** in terms of accuracy or response time is **not the whole picture**. We often require more detailed study!

Spatial Navigation

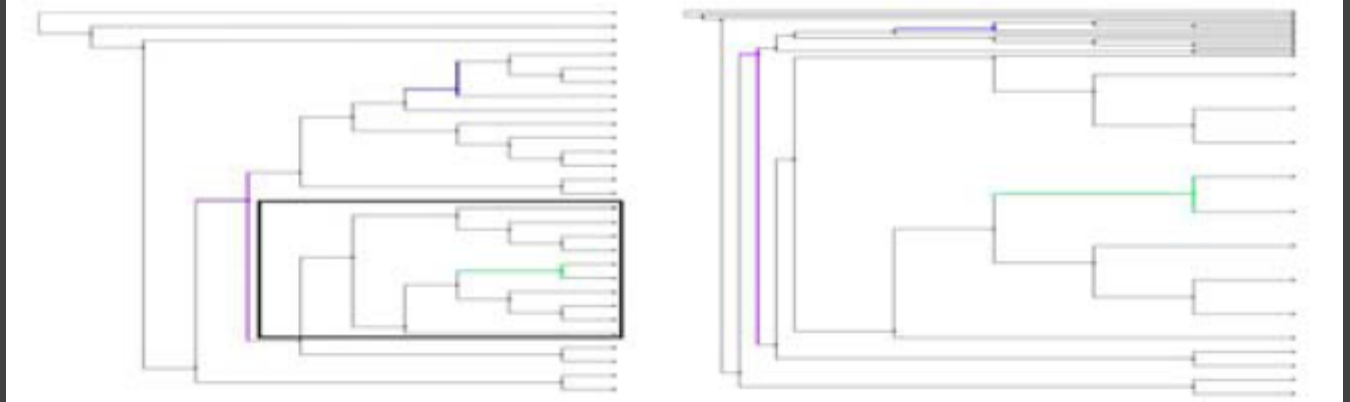
An Evaluation of Pan & Zoom and Rubber Sheet Navigation with and without an Overview

Dmitry Nekrasovski, Adam Bodnar, Joanna McGrenere,
François Guimbretière, Tamara Munzner

Pan & Zoom vs. Rubber Sheet



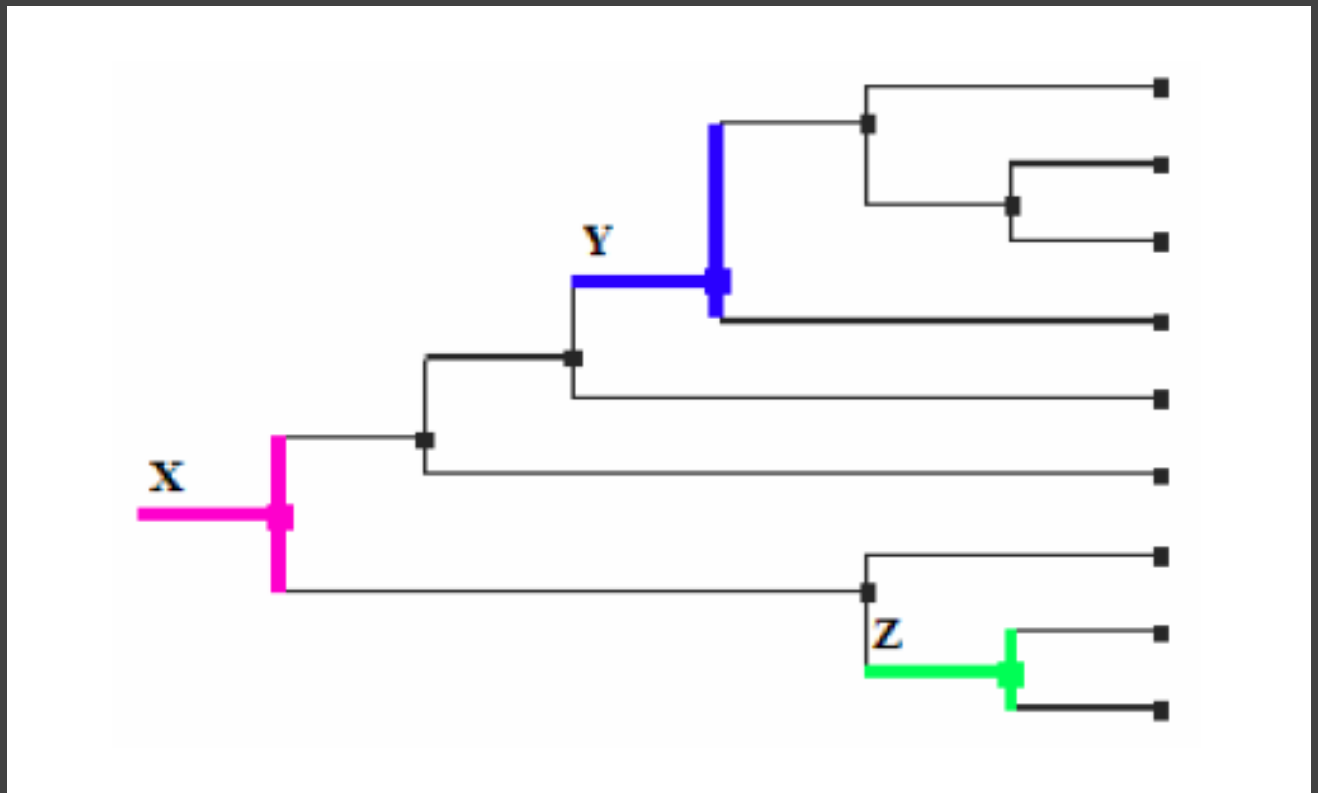
(i) PZN



(ii) RSN

Experimental Task

Compare topological distance between nodes in a dendrogram.



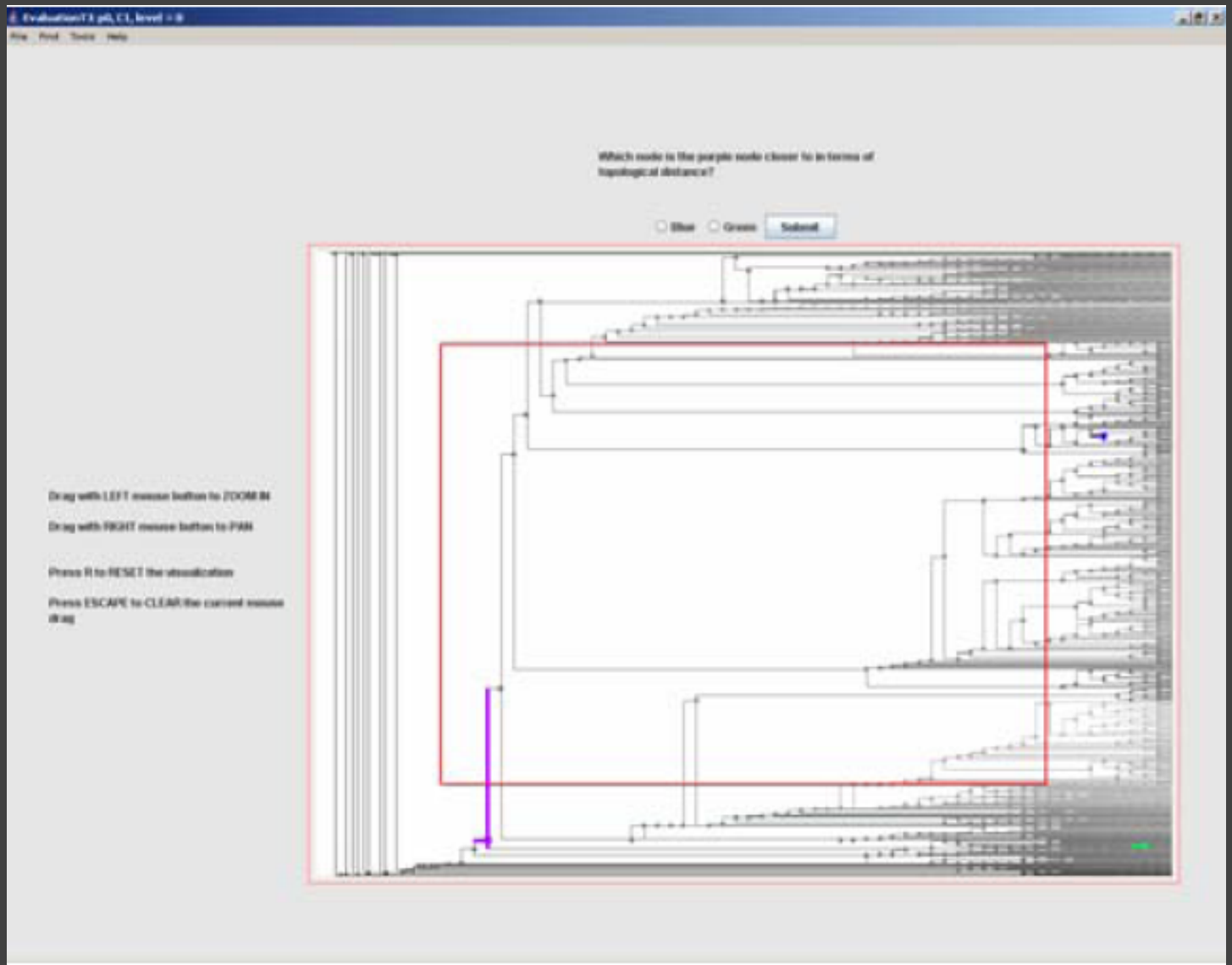
Experiment

Compare performance in 4 conditions:

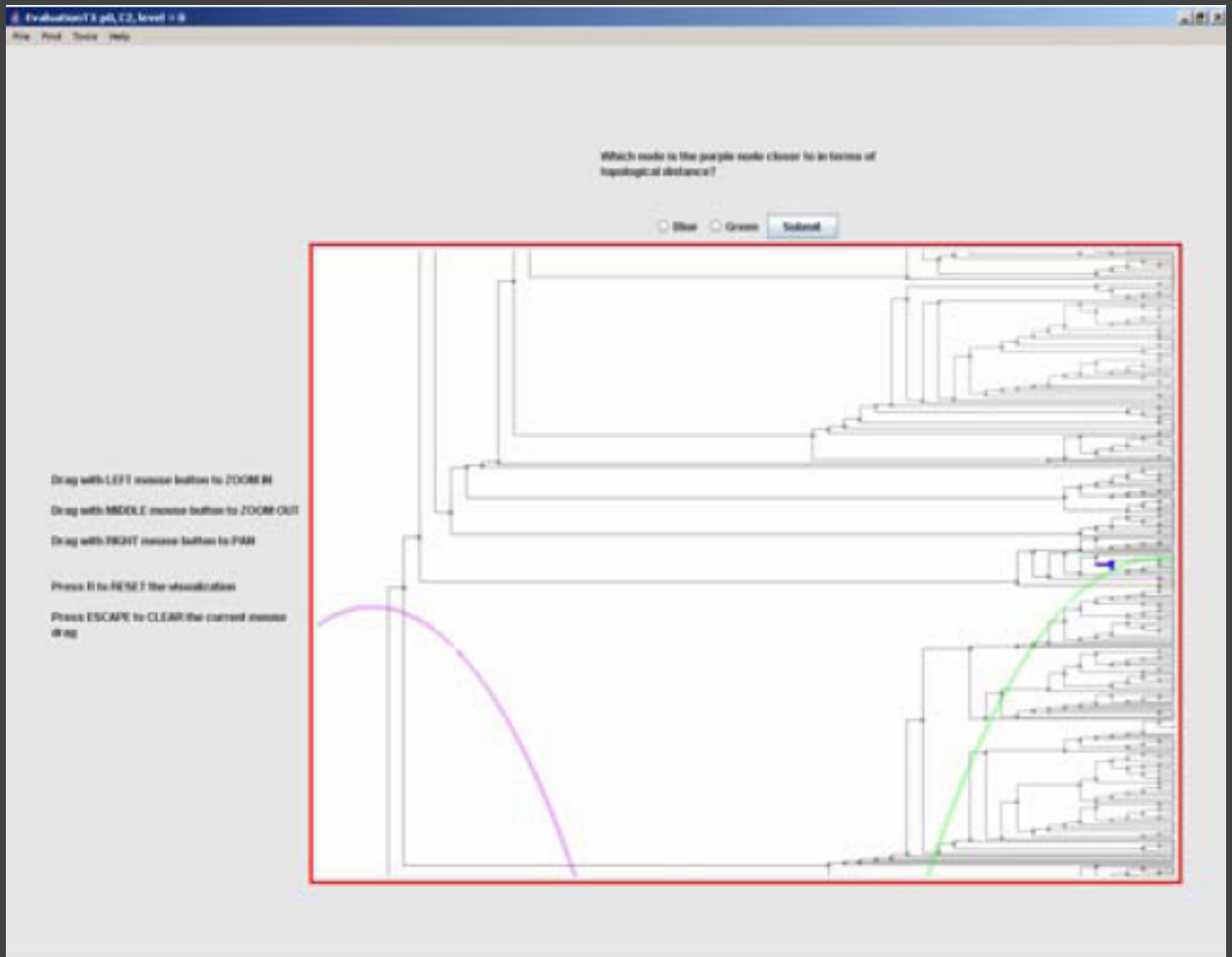
1. Pan & Zoom (no overview)
2. Pan & Zoom (with overview)
3. Rubber Sheet (no overview)
4. Rubber Sheet (with overview)

40 subjects (24F/16M), between 18-39 years old.
Right-handed, normal vision.
Between-subjects design.

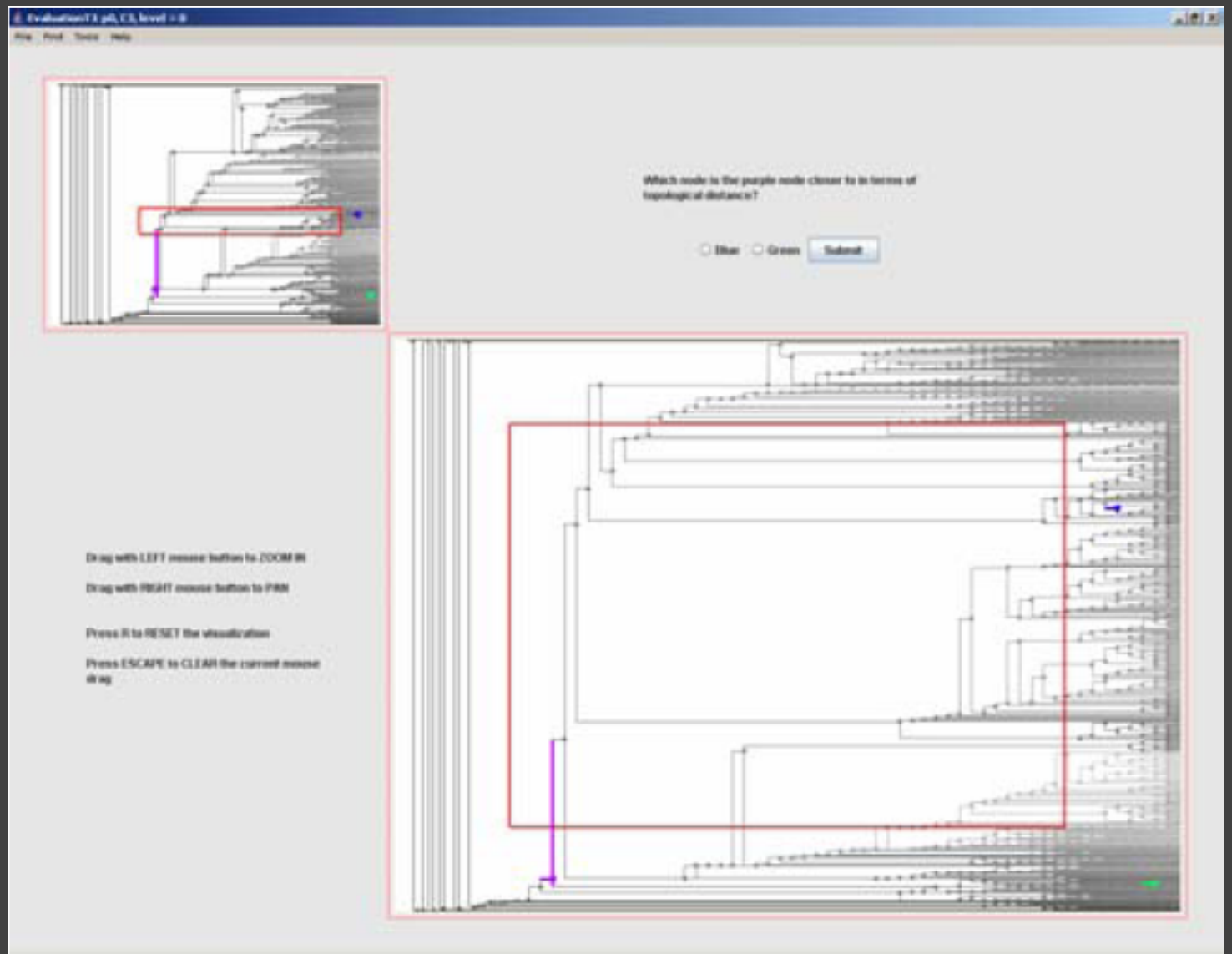
1. Rubber Sheet / No Overview



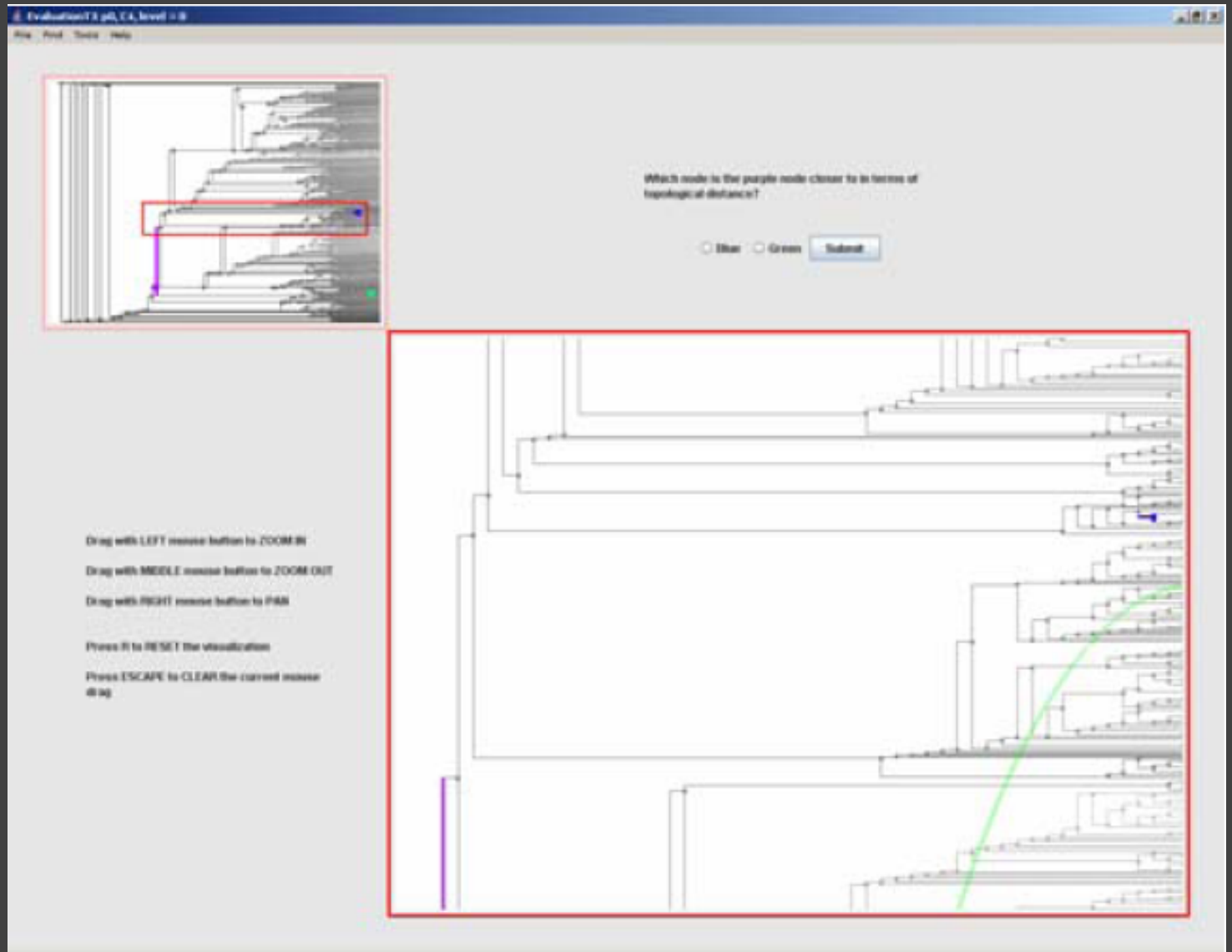
2. Pan & Zoom / No Overview



3. Rubber Sheet / Overview



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

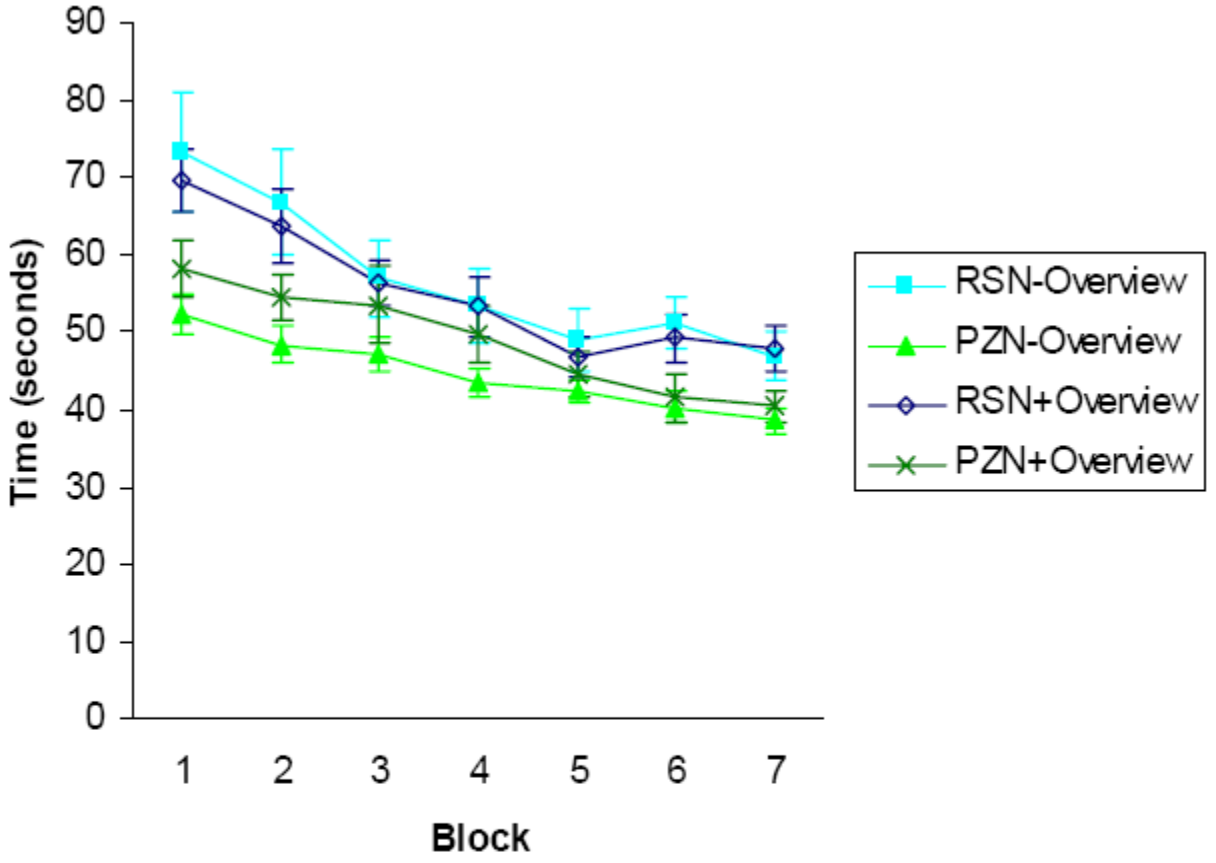


Figure 7: Mean completion times per trial for each interface by block in seconds (N=40).

Results: H2 True, H3 False

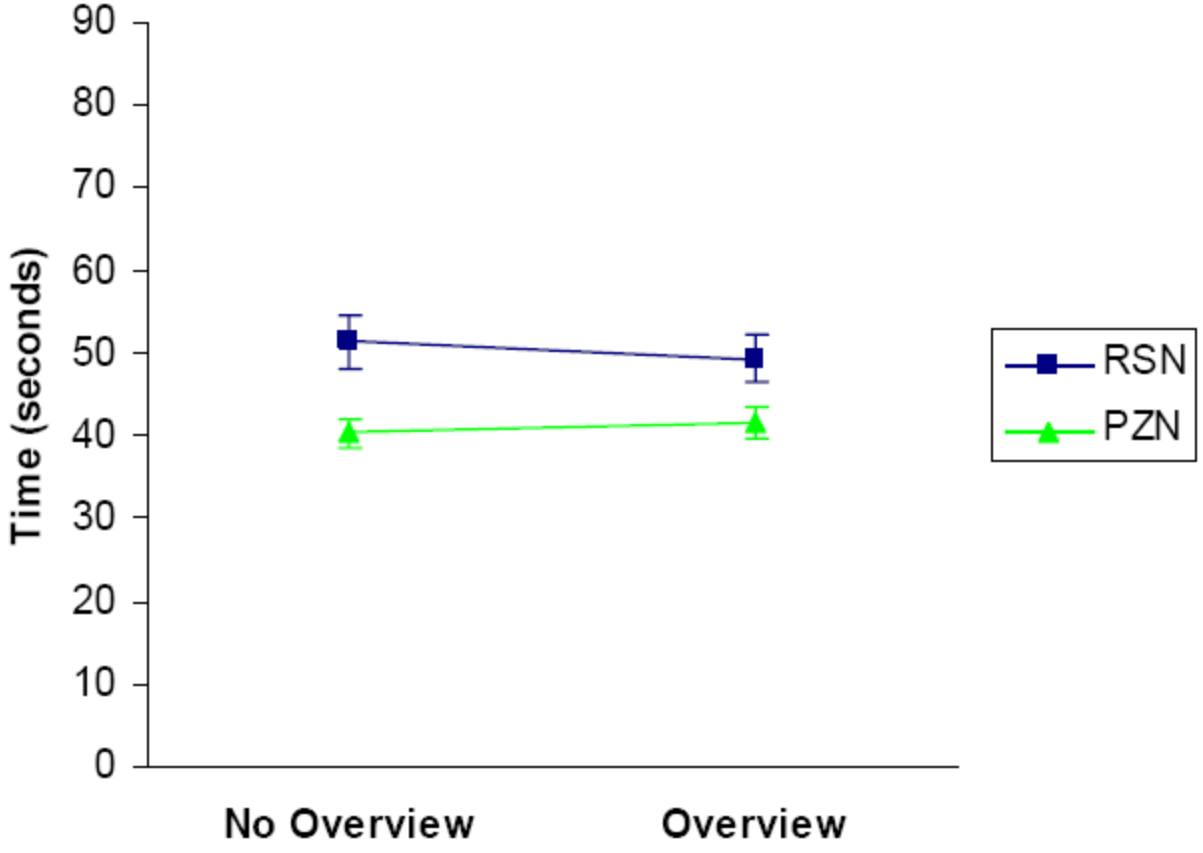


Figure 9: Block 7 mean per-trial completion times in seconds by navigation technique with and without an overview.

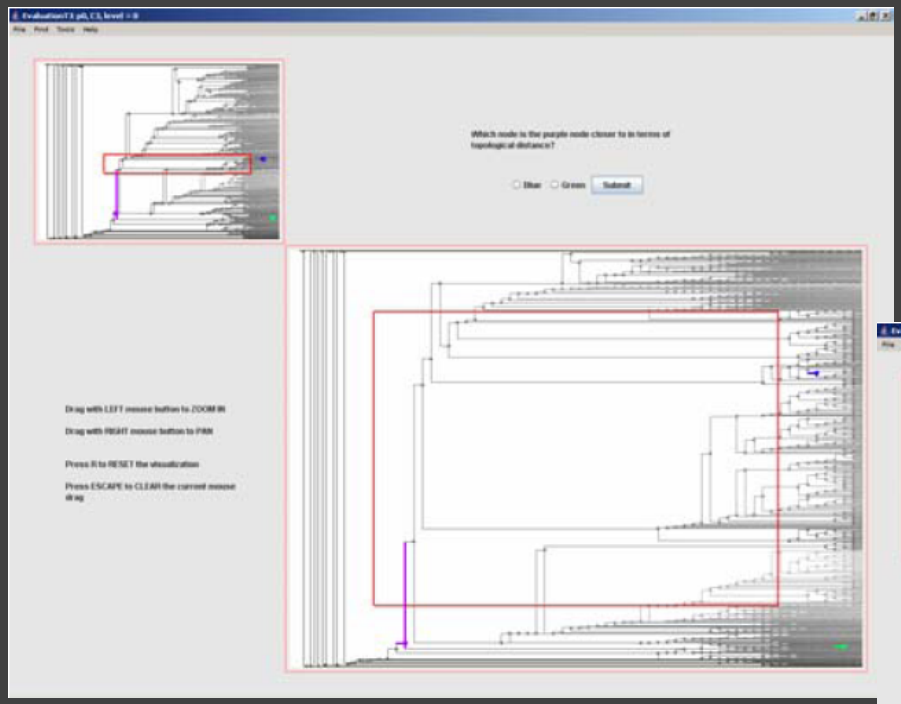
Results

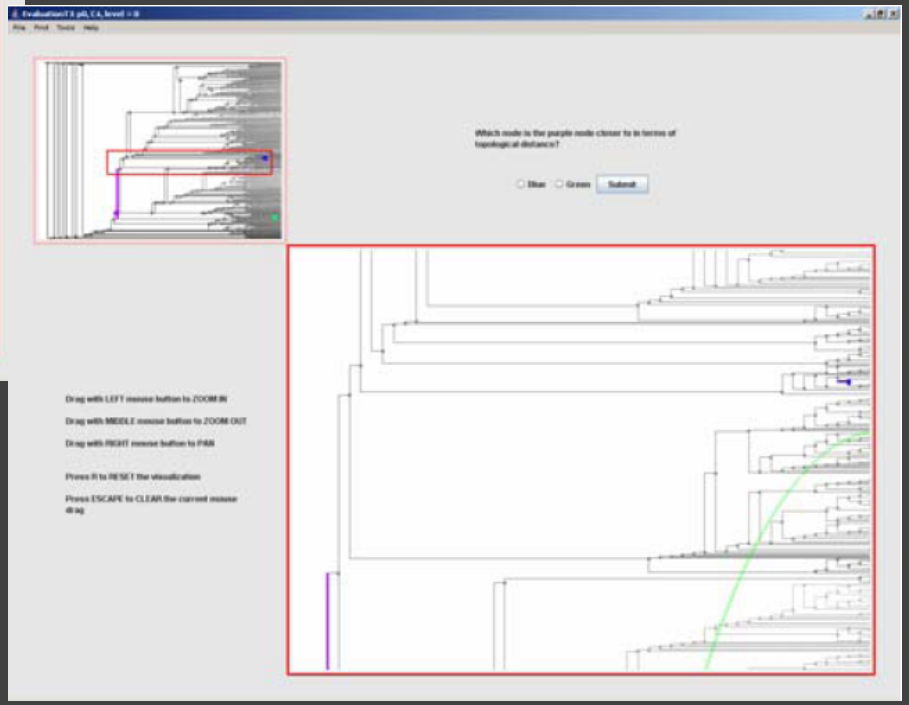
R1. Pan & Zoom had lower completion times, navigation actions, resets, and reported mental demand.

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.

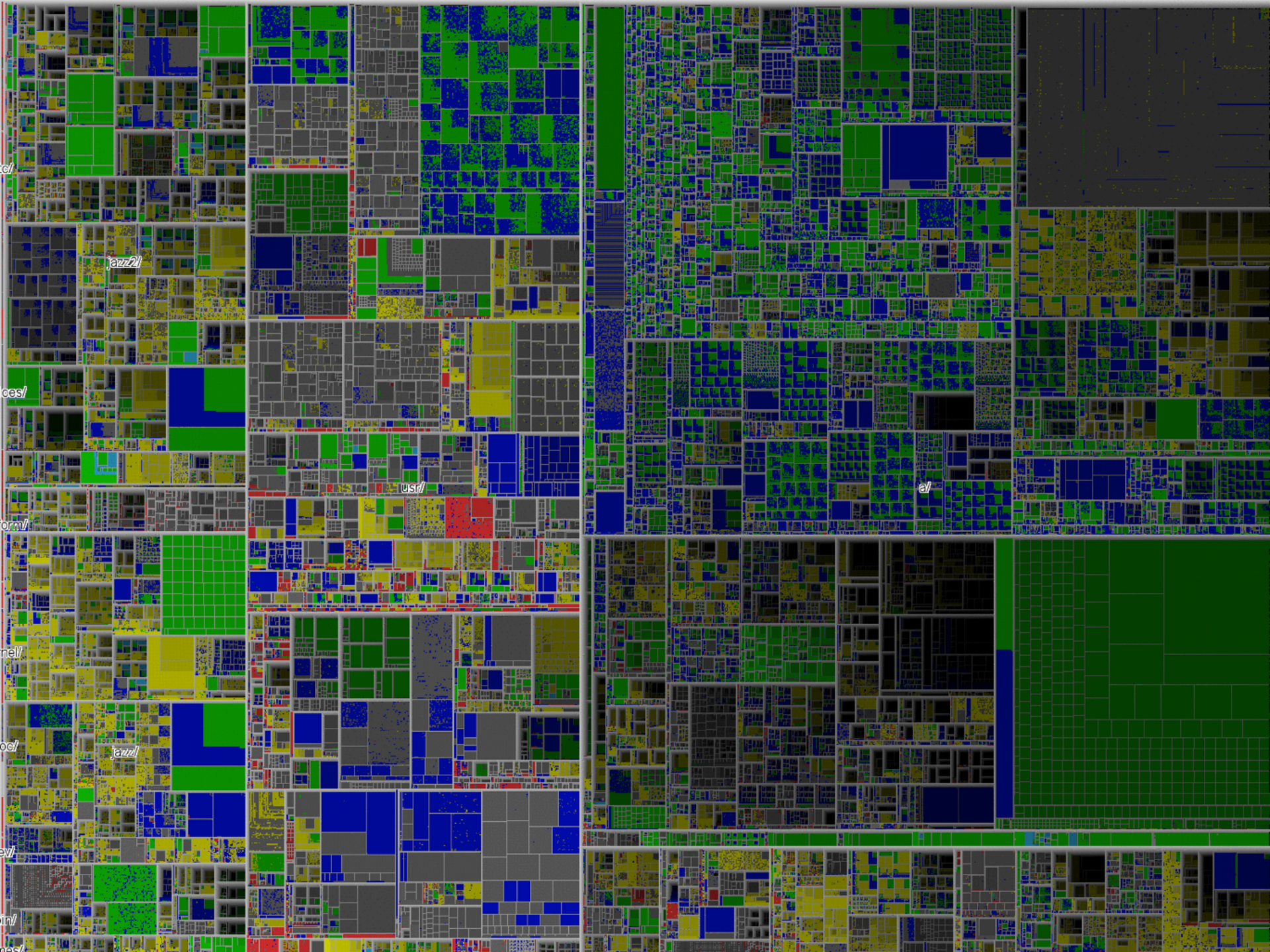
Thoughts?





Does this generalize
for overview displays?

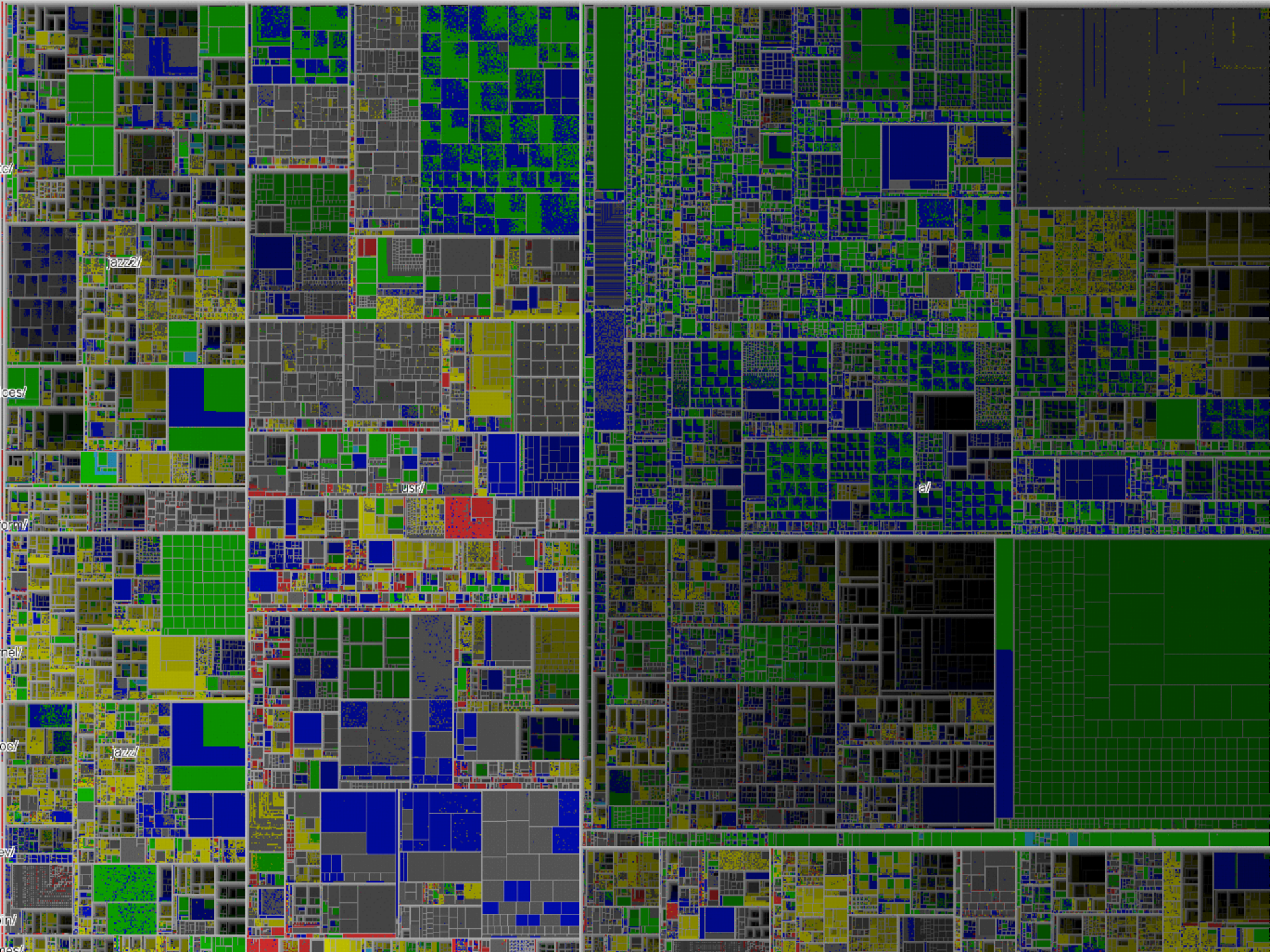
Data Density

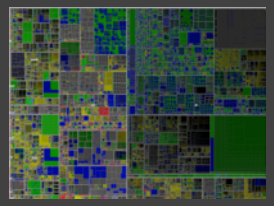

jazz2/
usr/
a/
jazz/

$$\text{Data Density} = \frac{(\#\text{ entries in data})}{(\text{area of graphic})}$$

"Graphical excellence… gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space"

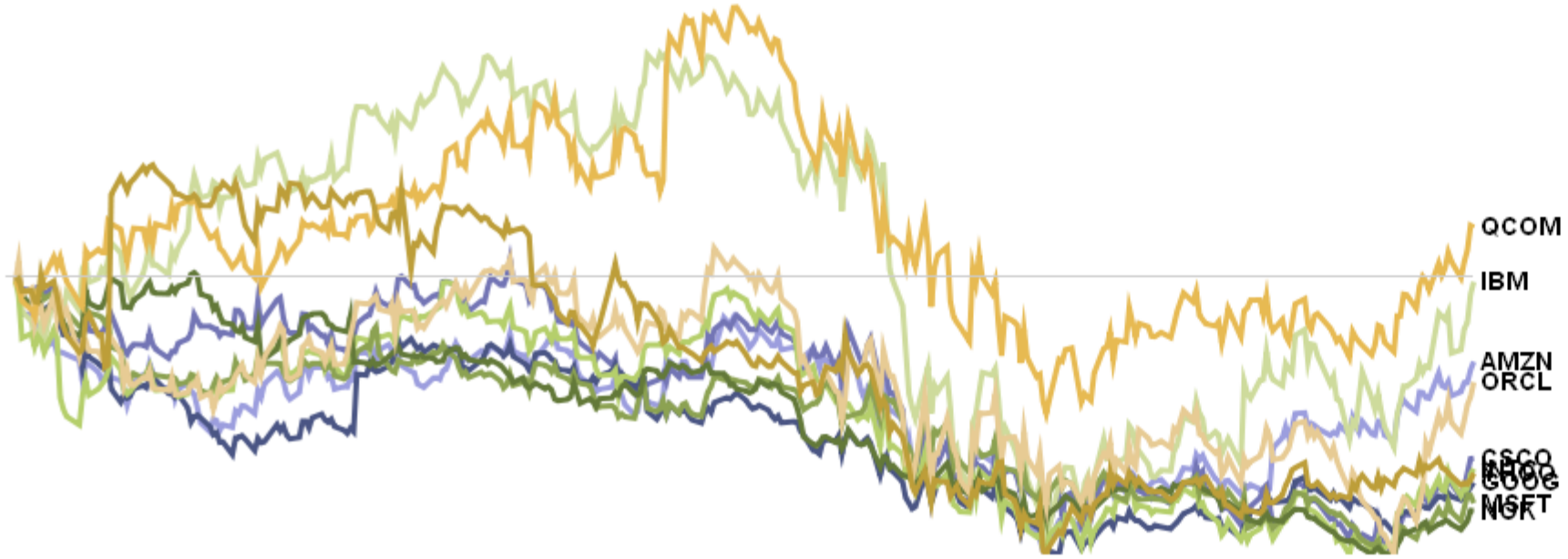
[Tufte 83]

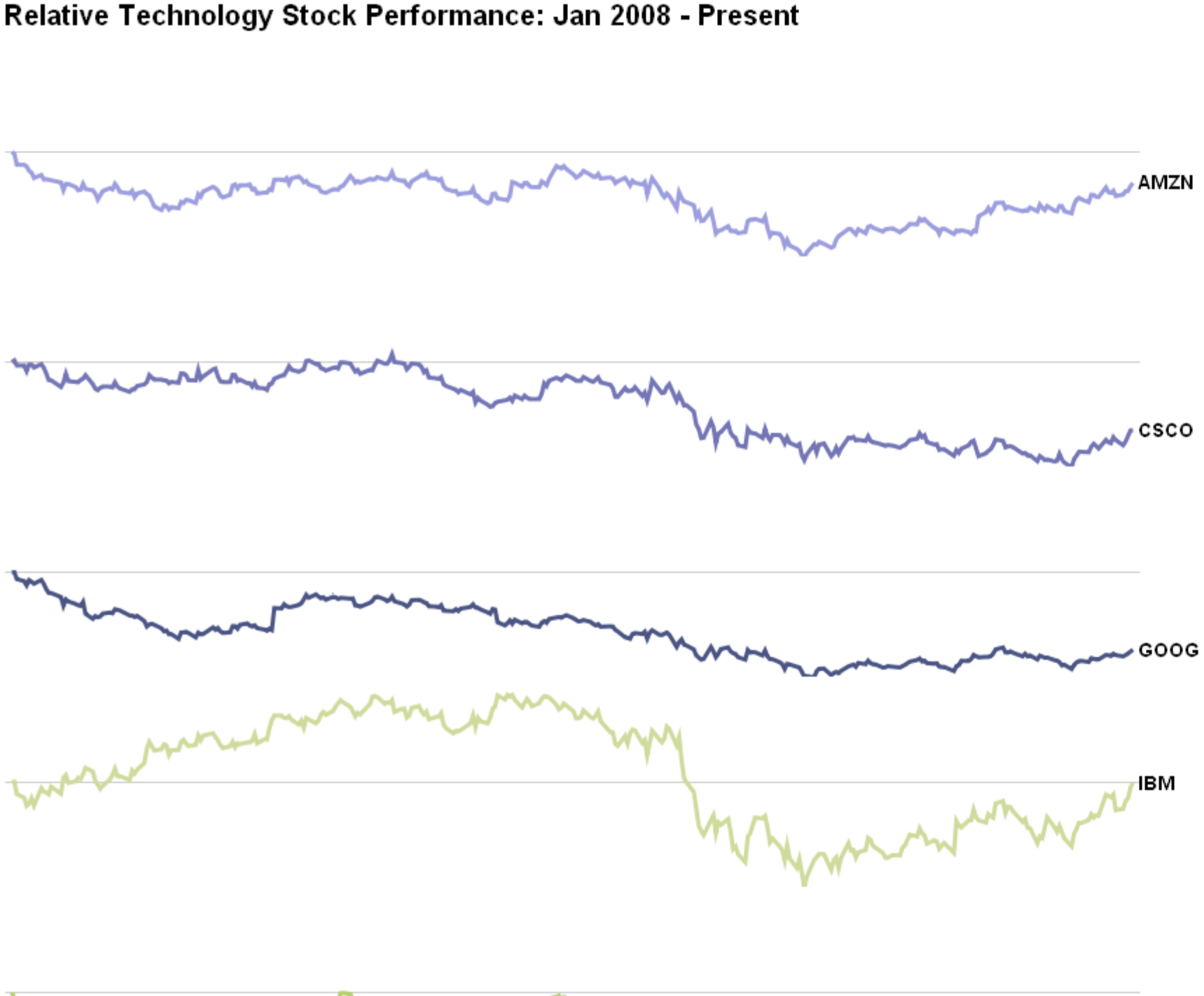

jazz2/
usr/
a/
jazz/

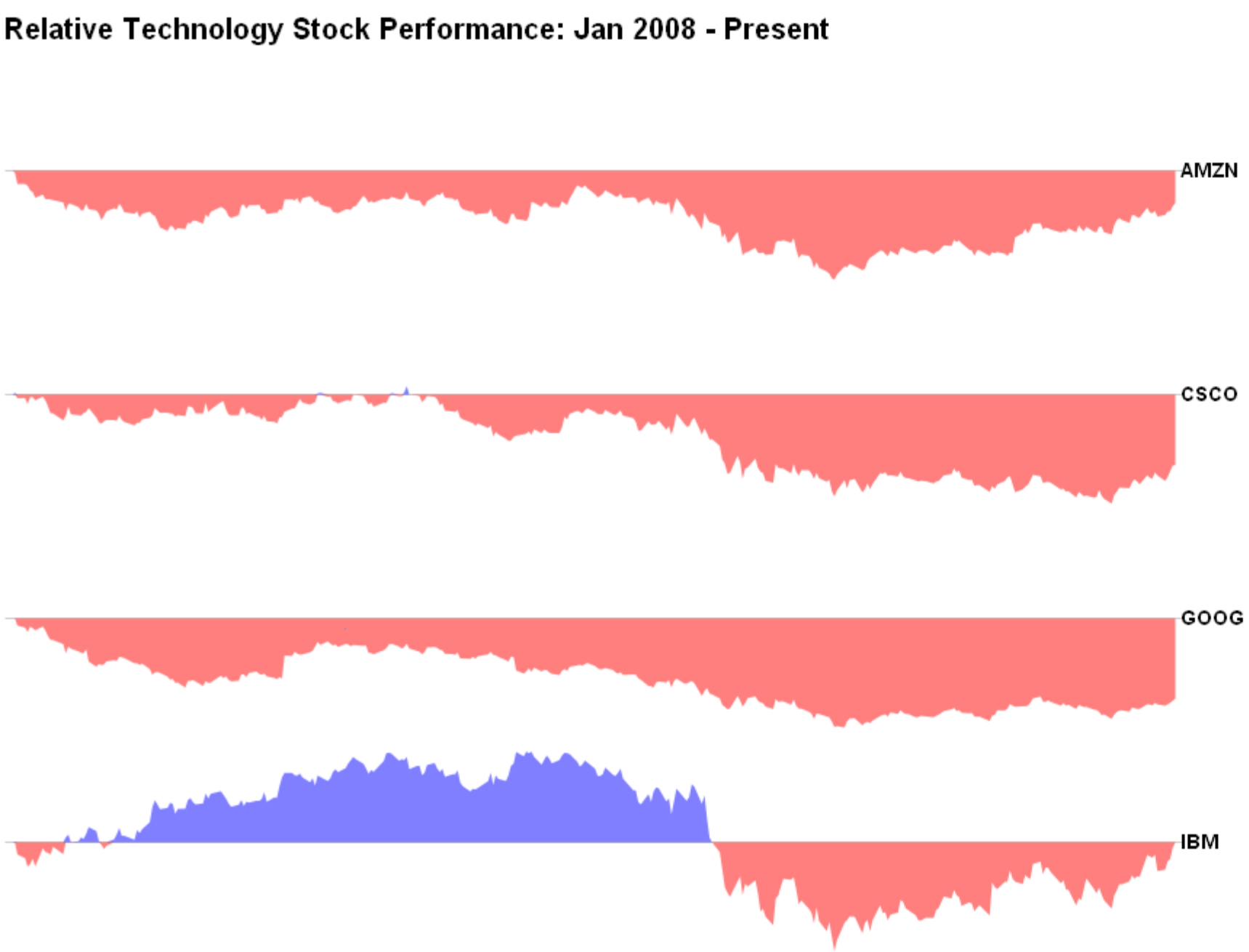


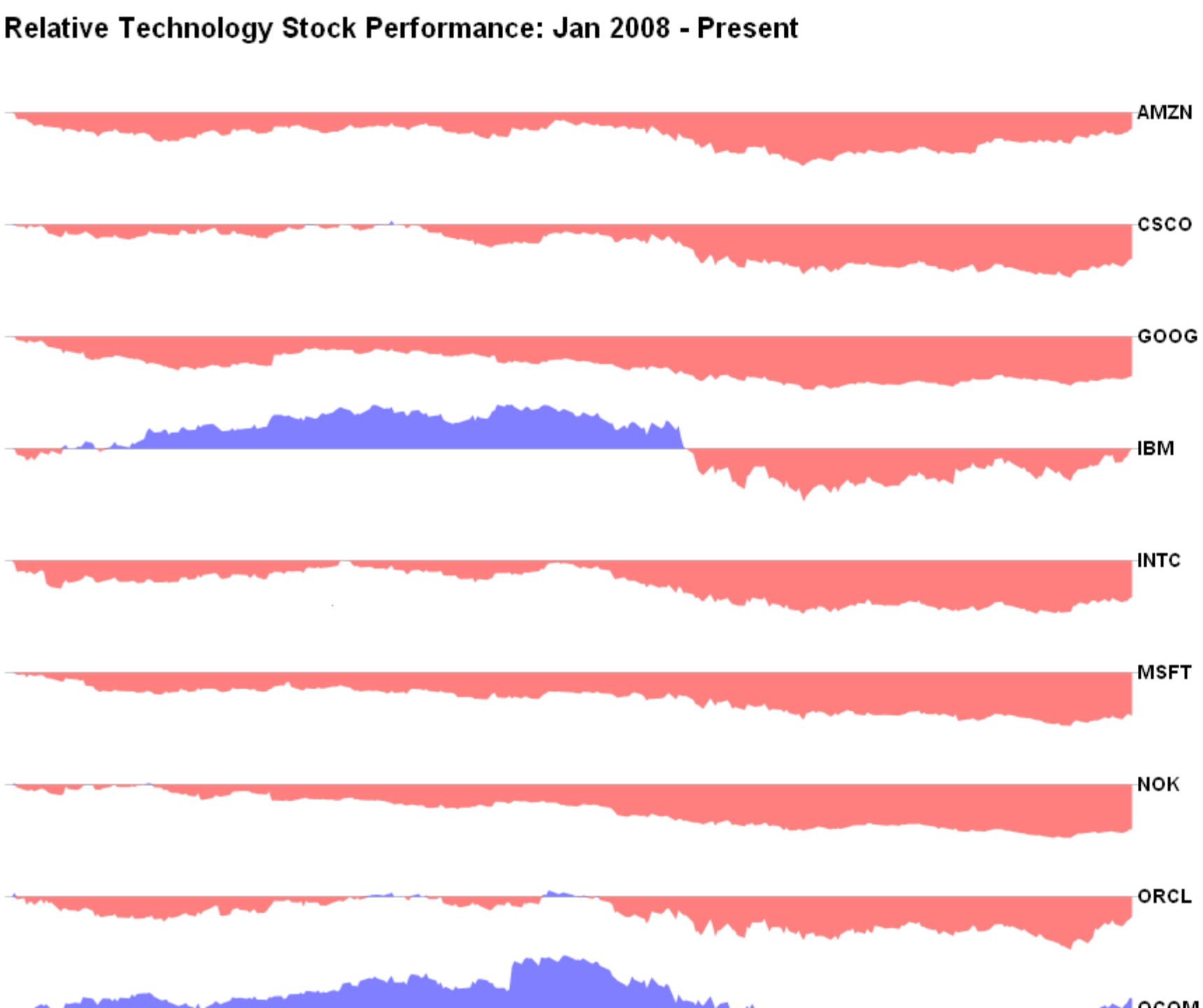
Relative Technology Stock Performance: Jan 2008 - Present

Relative Technology Stock Performance: Jan 2008 - Present

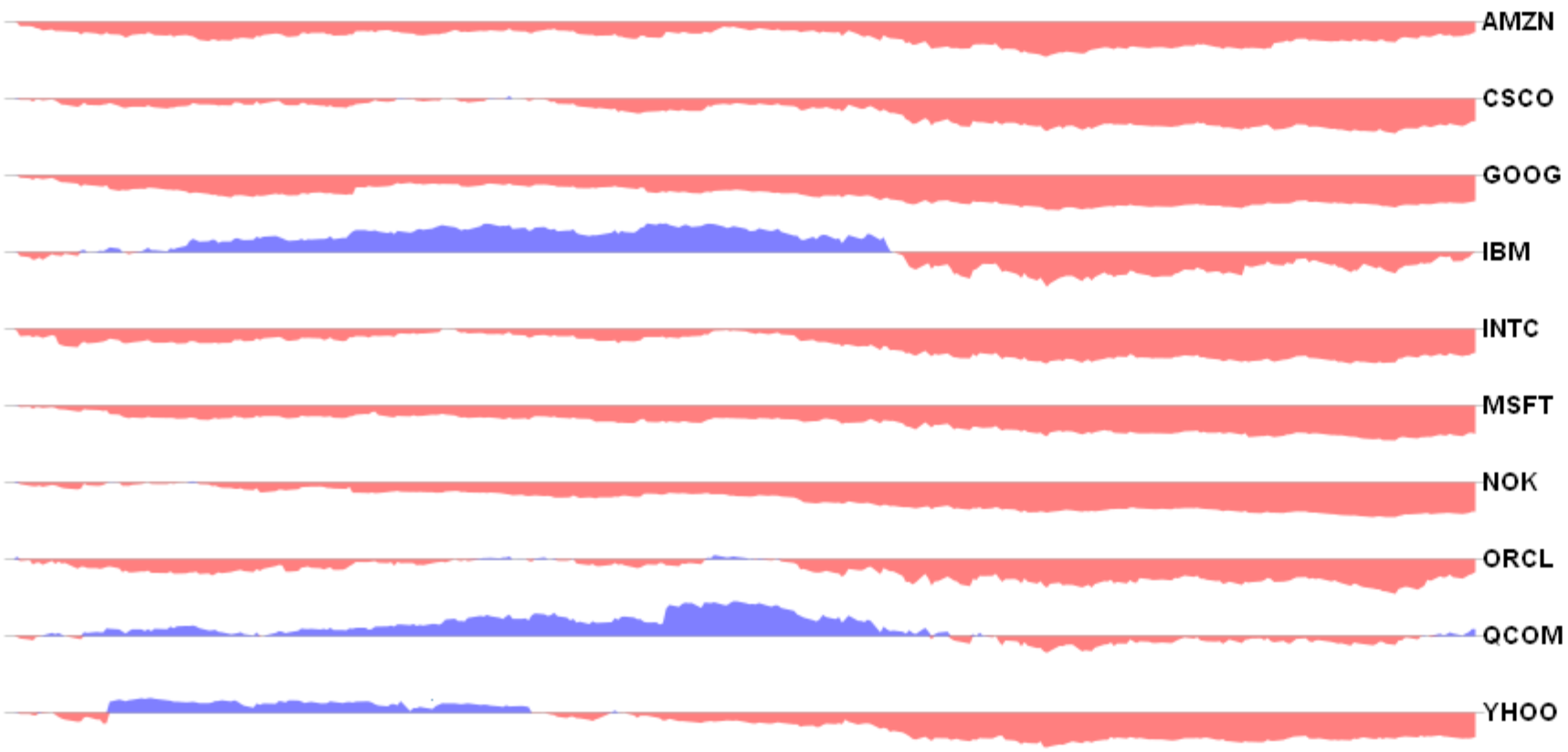



Relative Technology Stock Performance: Jan 2008 - Present
AMZN
CSCO
GOOG
IBM

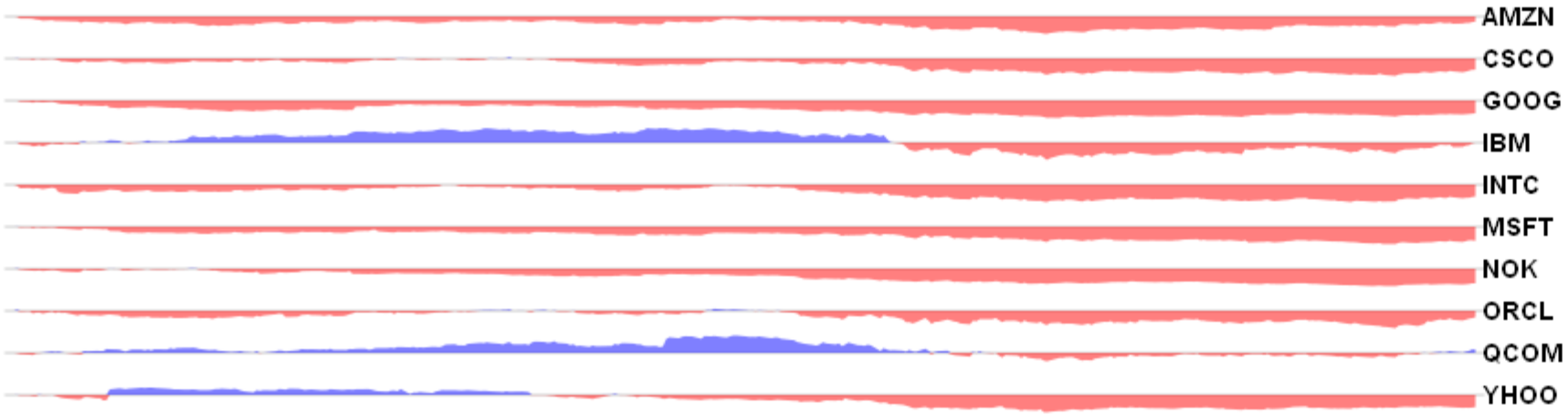

Relative Technology Stock Performance: Jan 2008 - Present
AMZN
CSCO
GOOG
IBM

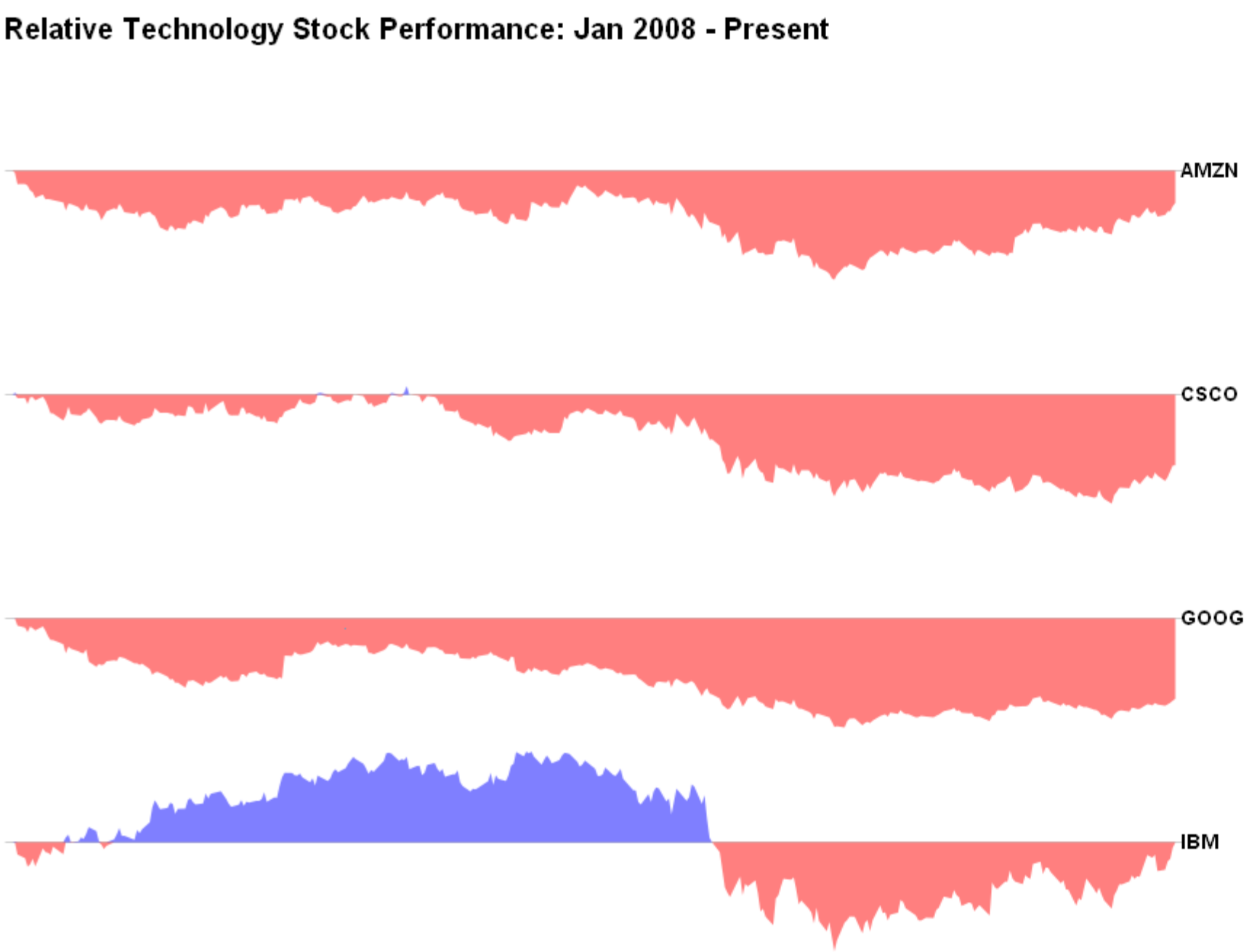

Relative Technology Stock Performance: Jan 2008 - Present
AMZN
CSCO
GOOG
IBM
INTC
MSFT
NOK
ORCL

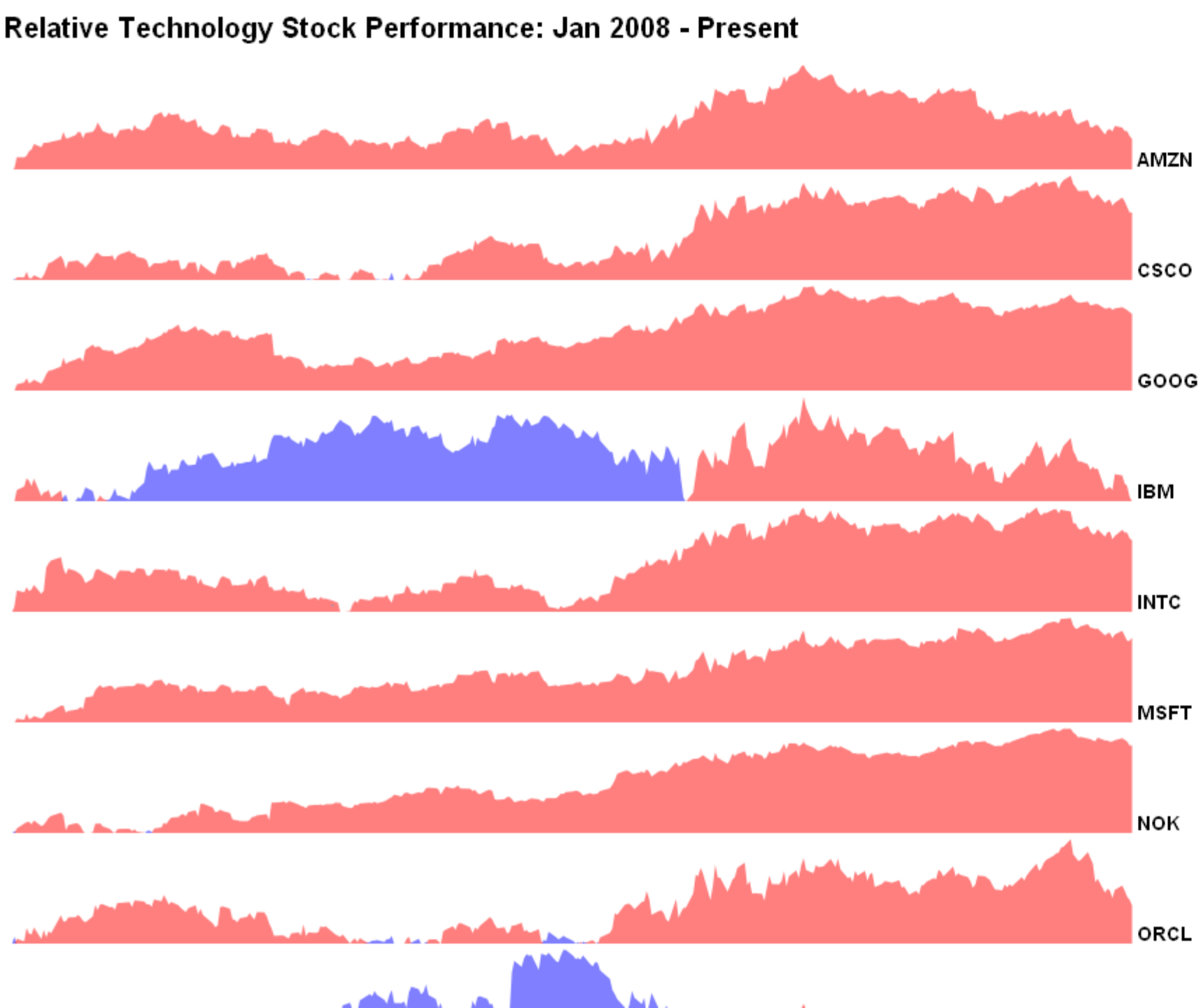
Relative Technology Stock Performance: Jan 2008 - Present



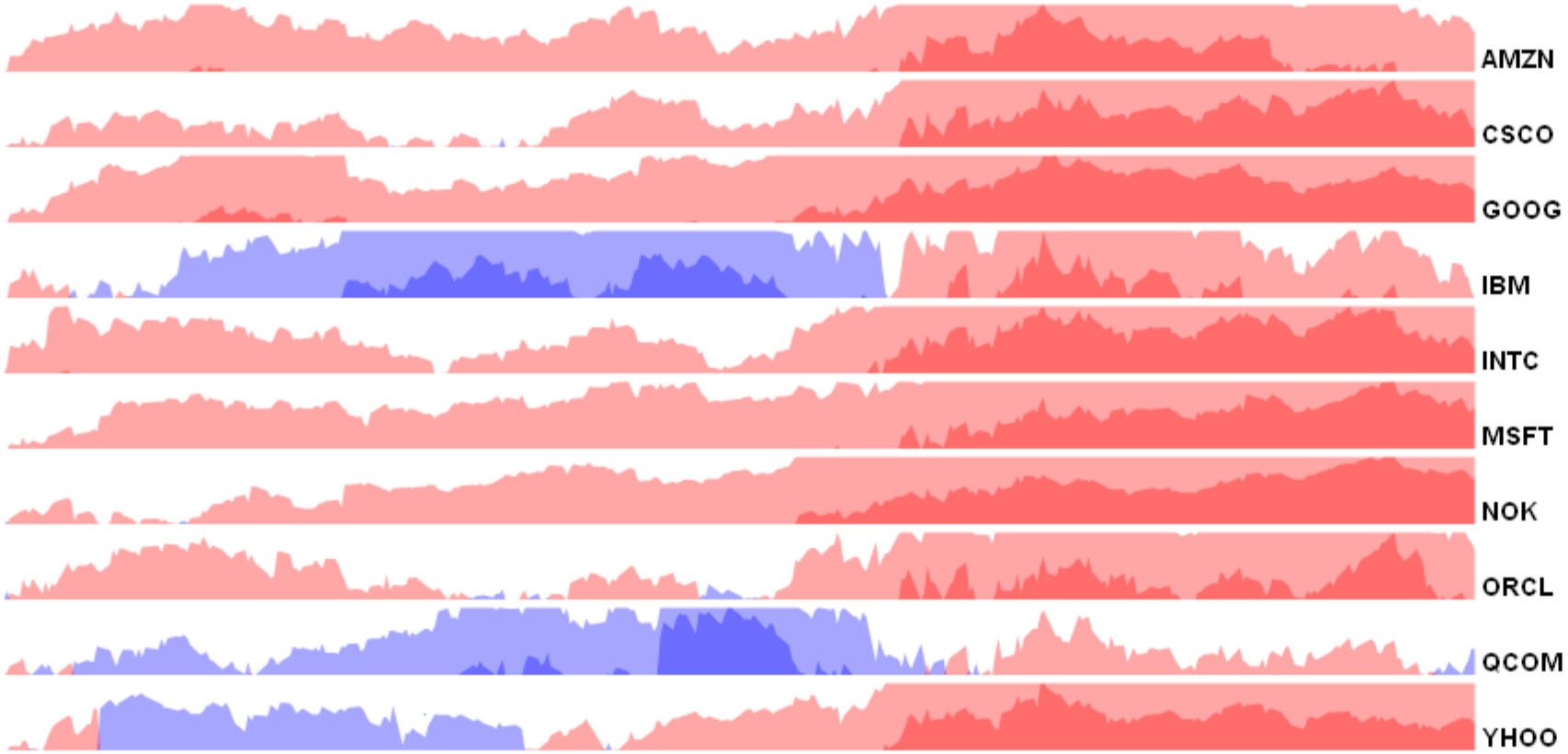
Relative Technology Stock Performance: Jan 2008 - Present



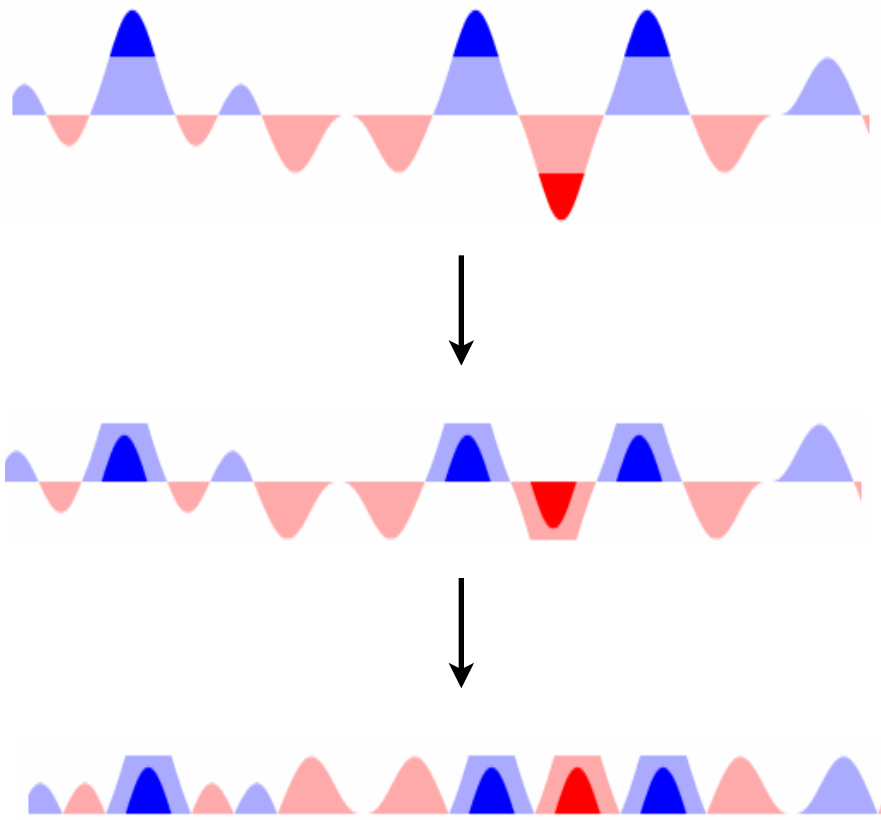

Relative Technology Stock Performance: Jan 2008 - Present
AMZN
CSCO
GOOG
IBM


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

Relative Technology Stock Performance: Jan 2008 - Present



Horizon Graphs

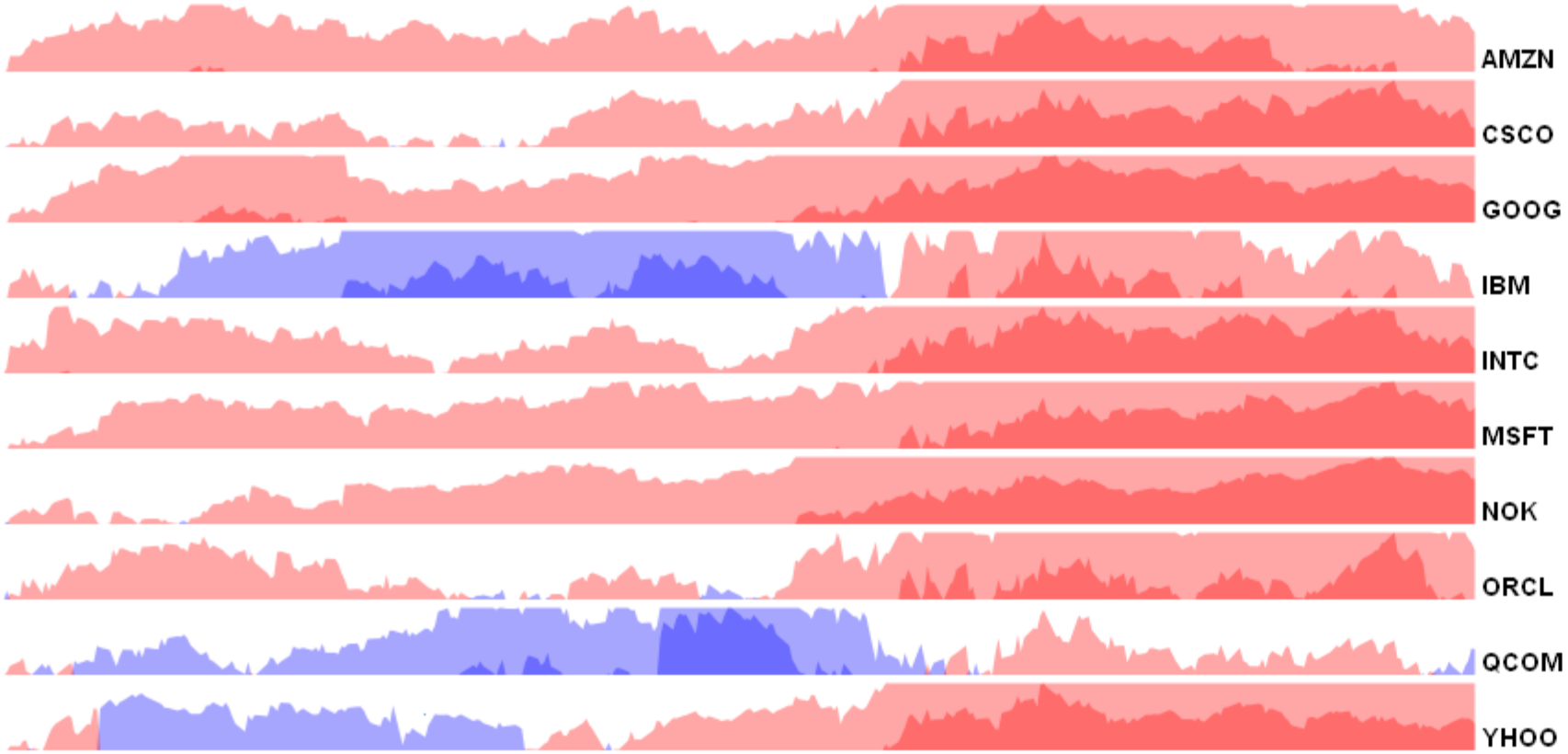


Segment Peaks

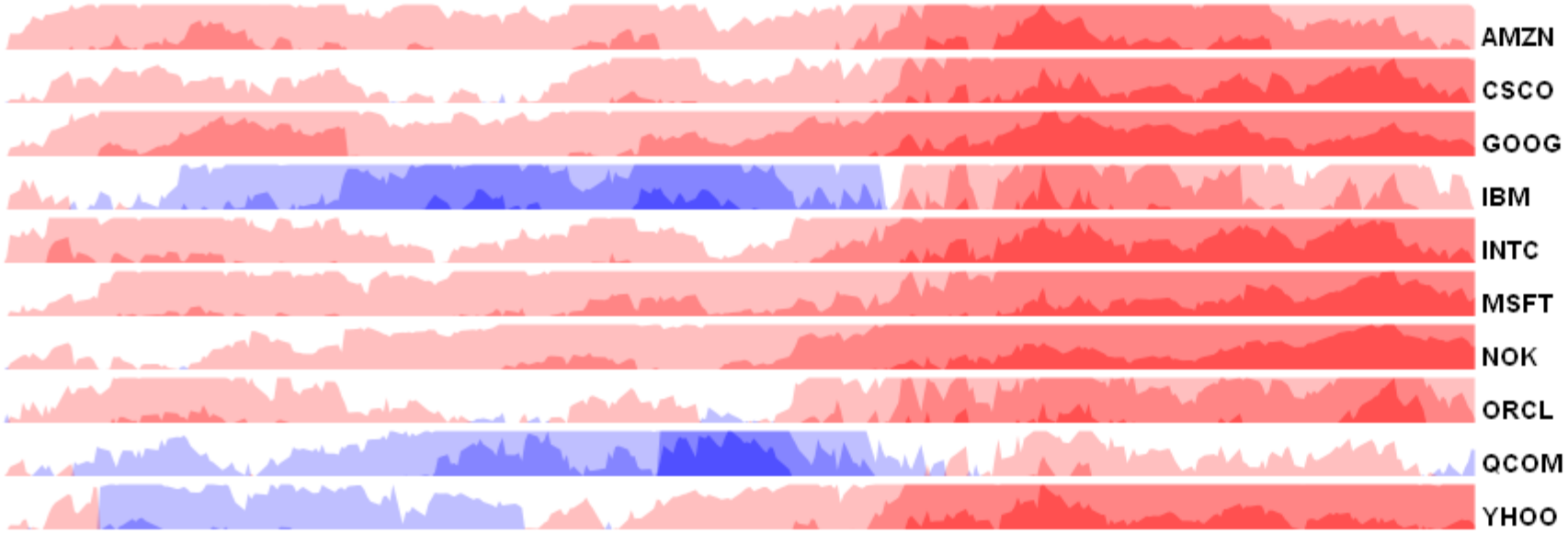
Layer Segments

Mirror Negative Values

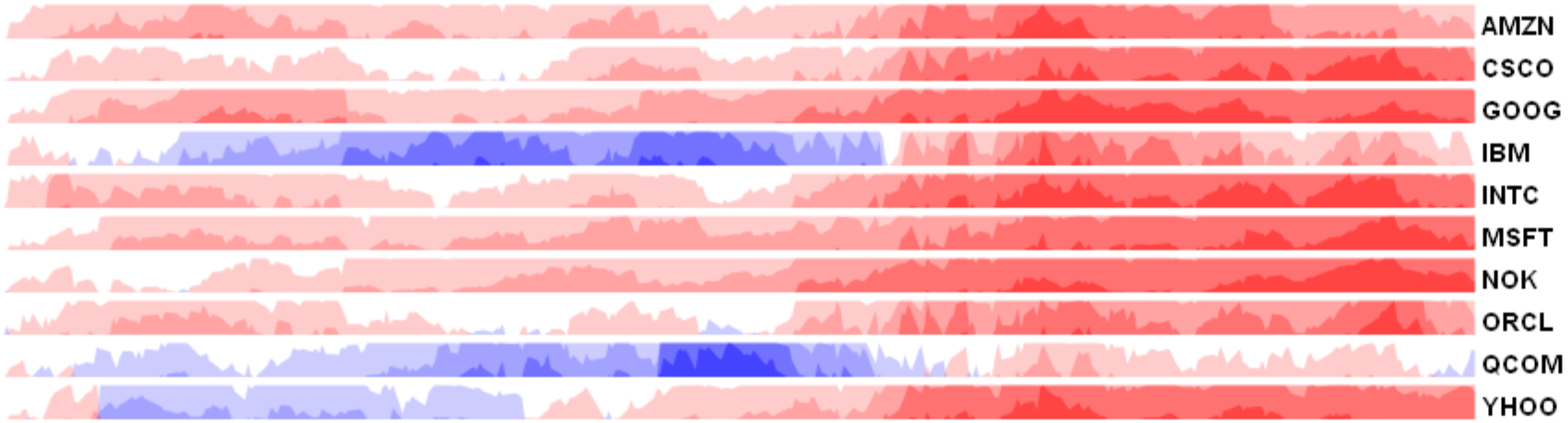
Relative Technology Stock Performance: Jan 2008 - Present



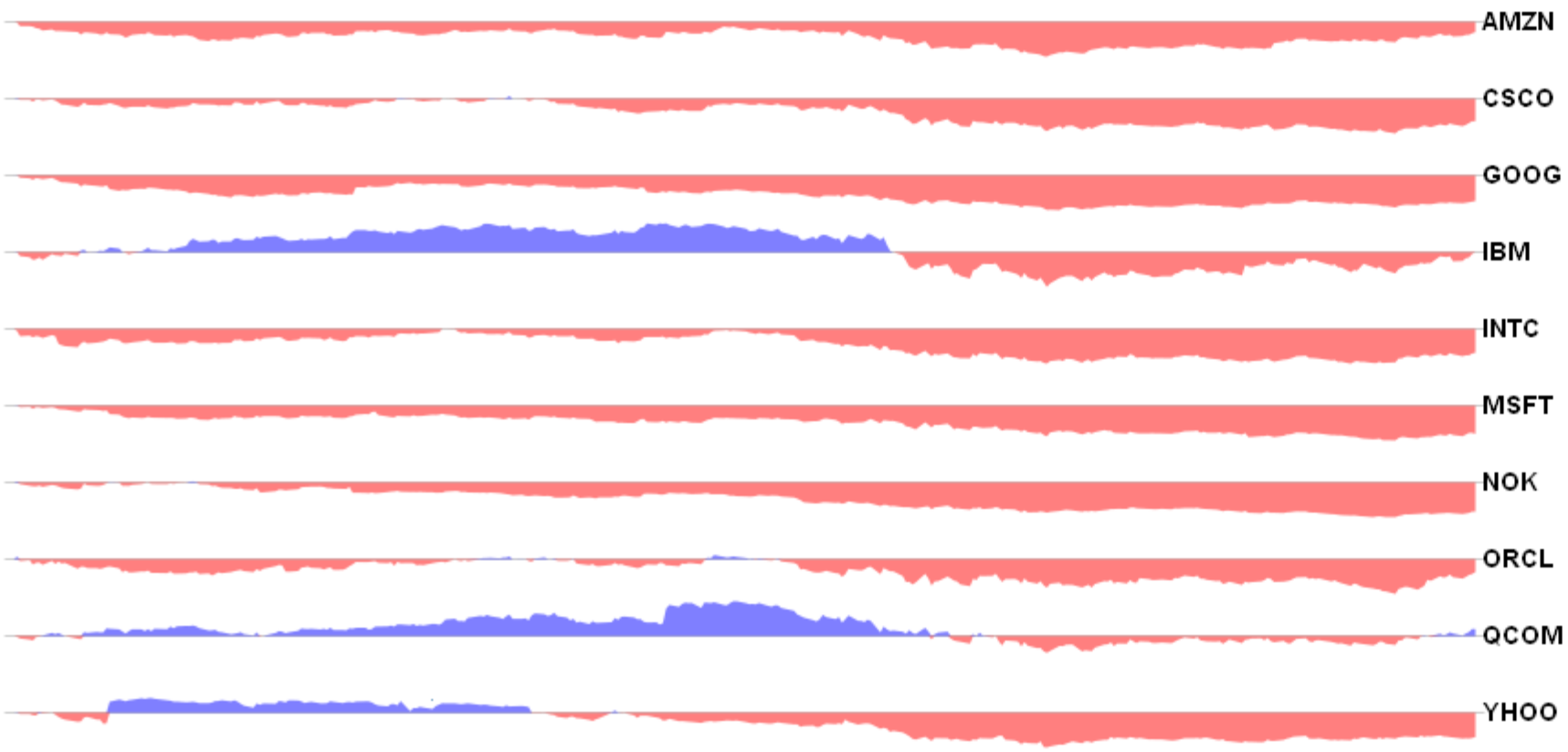
Relative Technology Stock Performance: Jan 2008 - Present



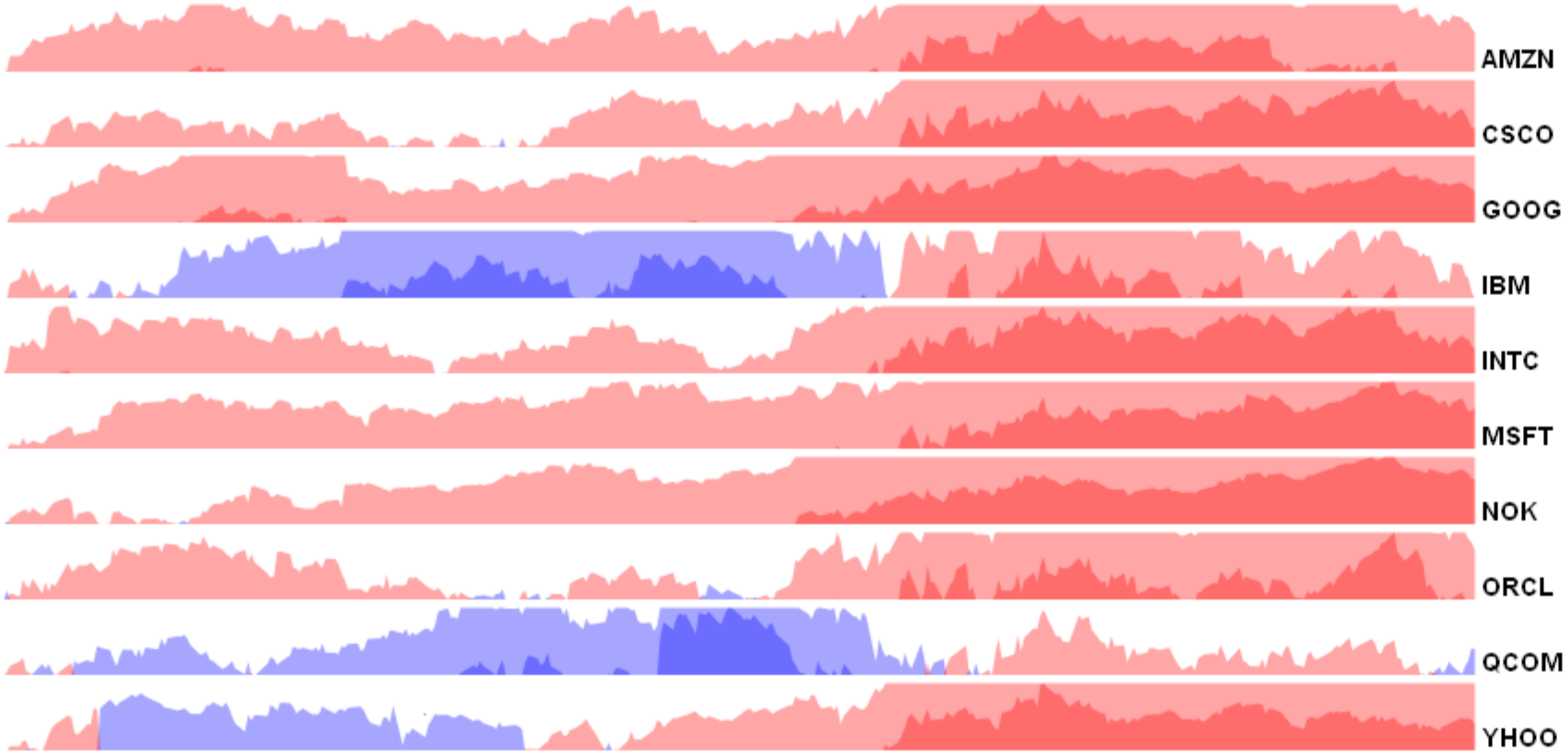
Relative Technology Stock Performance: Jan 2008 - Present

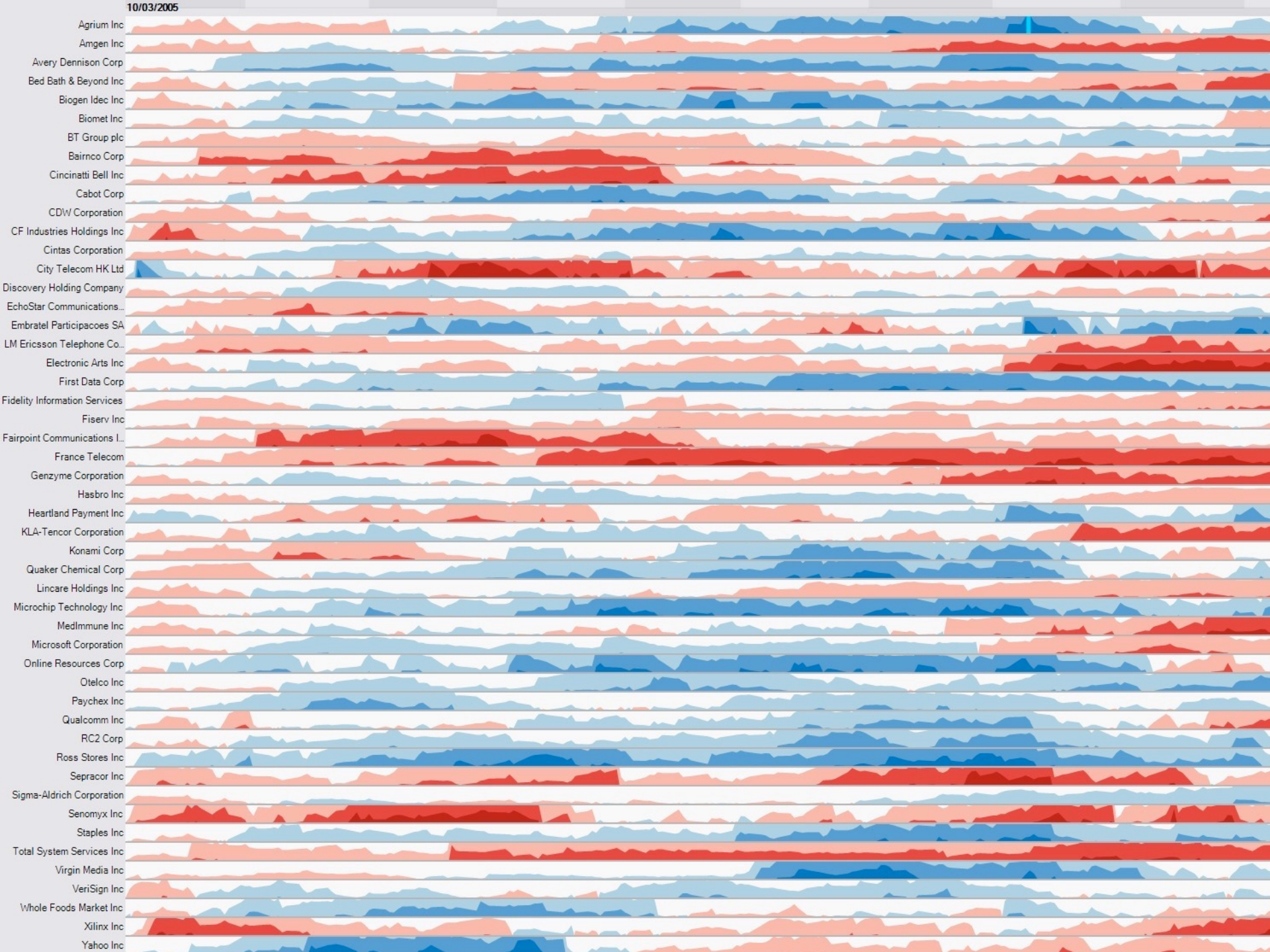


Relative Technology Stock Performance: Jan 2008 - Present



Relative Technology Stock Performance: Jan 2008 - Present

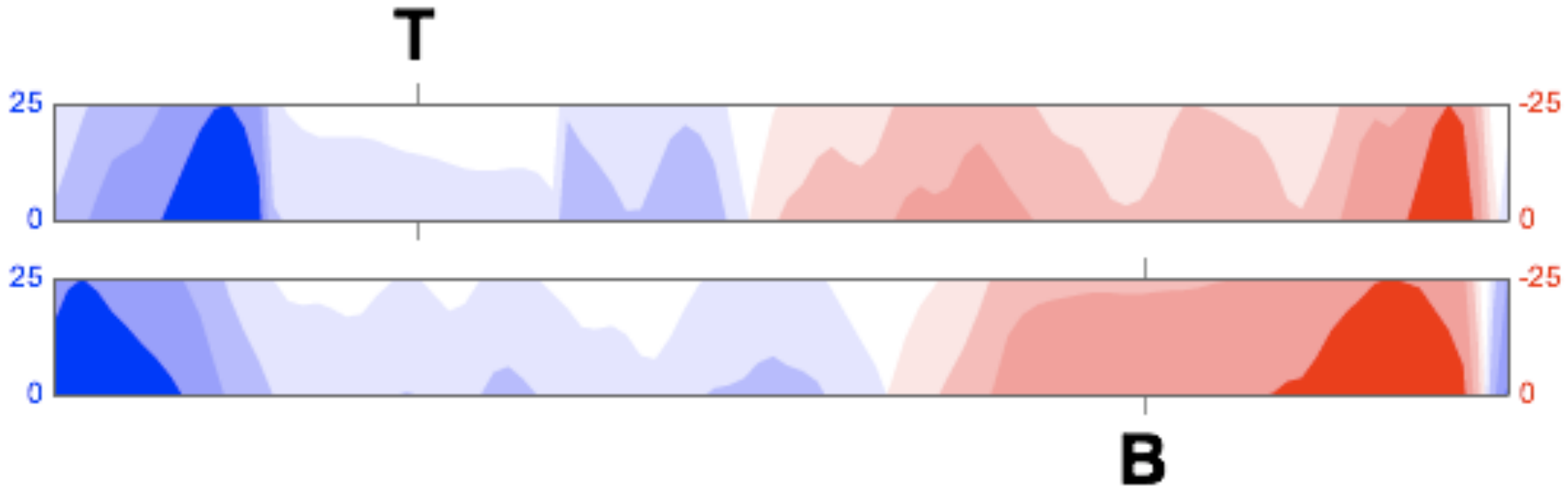



10/03/2005
Agrium Inc
Amgen Inc
Avery Dennison Corp
Bed Bath & Beyond Inc
Biogen Idec Inc
Biomet Inc
BT Group plc
Bairnco Corp
Cincinatti Bell Inc
Cabot Corp
CDW Corporation
CF Industries Holdings Inc
Cintas Corporation
City Telecom HK Ltd
Discovery Holding Company
EchoStar Communications...
Embratel Participacoes SA
LM Ericsson Telephone Co...
Electronic Arts Inc
First Data Corp
Fidelity Information Services
Fiserv Inc
Fairpoint Communications I...
France Telecom
Genzyme Corporation
Hasbro Inc
Heartland Payment Inc
KLA-Tencor Corporation
Konami Corp
Quaker Chemical Corp
Lincare Holdings Inc
Microchip Technology Inc
MedImmune Inc
Microsoft Corporation
Online Resources Corp
Otelco Inc
Paychex Inc
Qualcomm Inc
RC2 Corp
Ross Stores Inc
Sepracor Inc
Sigma-Aldrich Corporation
Senomyx Inc
Staples Inc
Total System Services Inc
Virgin Media Inc
VeriSign Inc
Whole Foods Market Inc
Xilinx Inc
Yahoo Inc

Experiment: Chart Type & Size

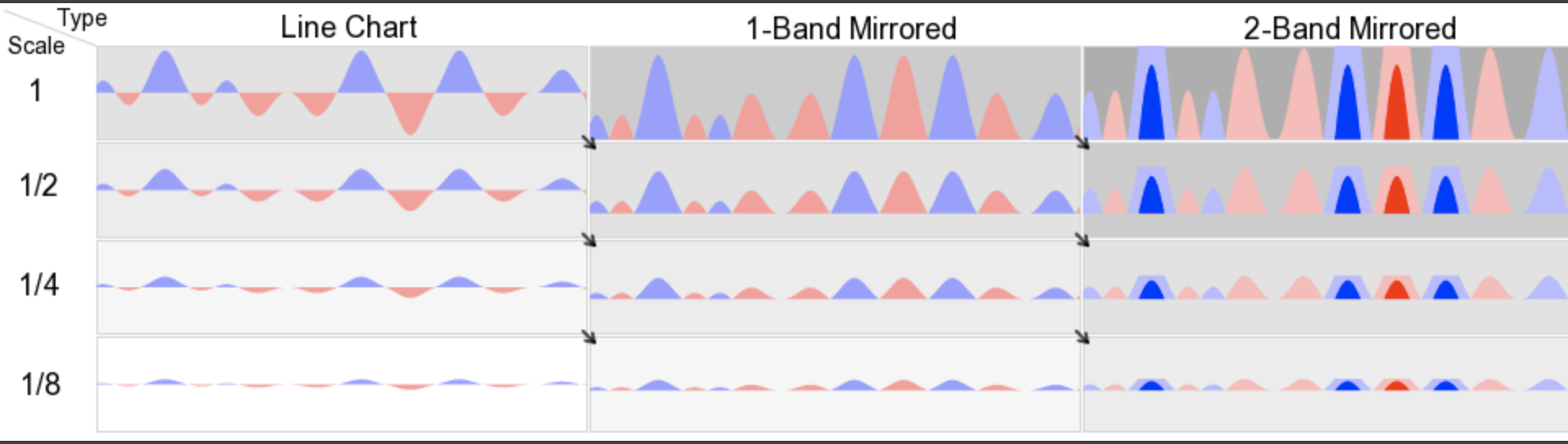
Q1: How do mirroring and layering affect estimation time and accuracy compared to line charts?

Q2: How does chart size affect estimation time and accuracy?



Estimate the difference between T and B (0-200) to within 5 values.

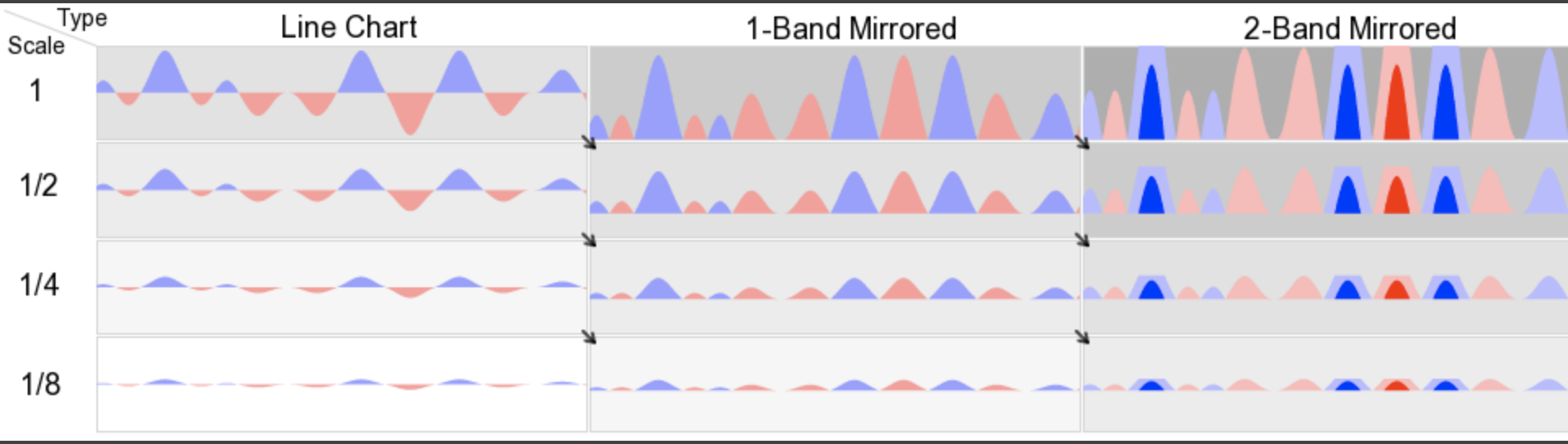
Experiment Design



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

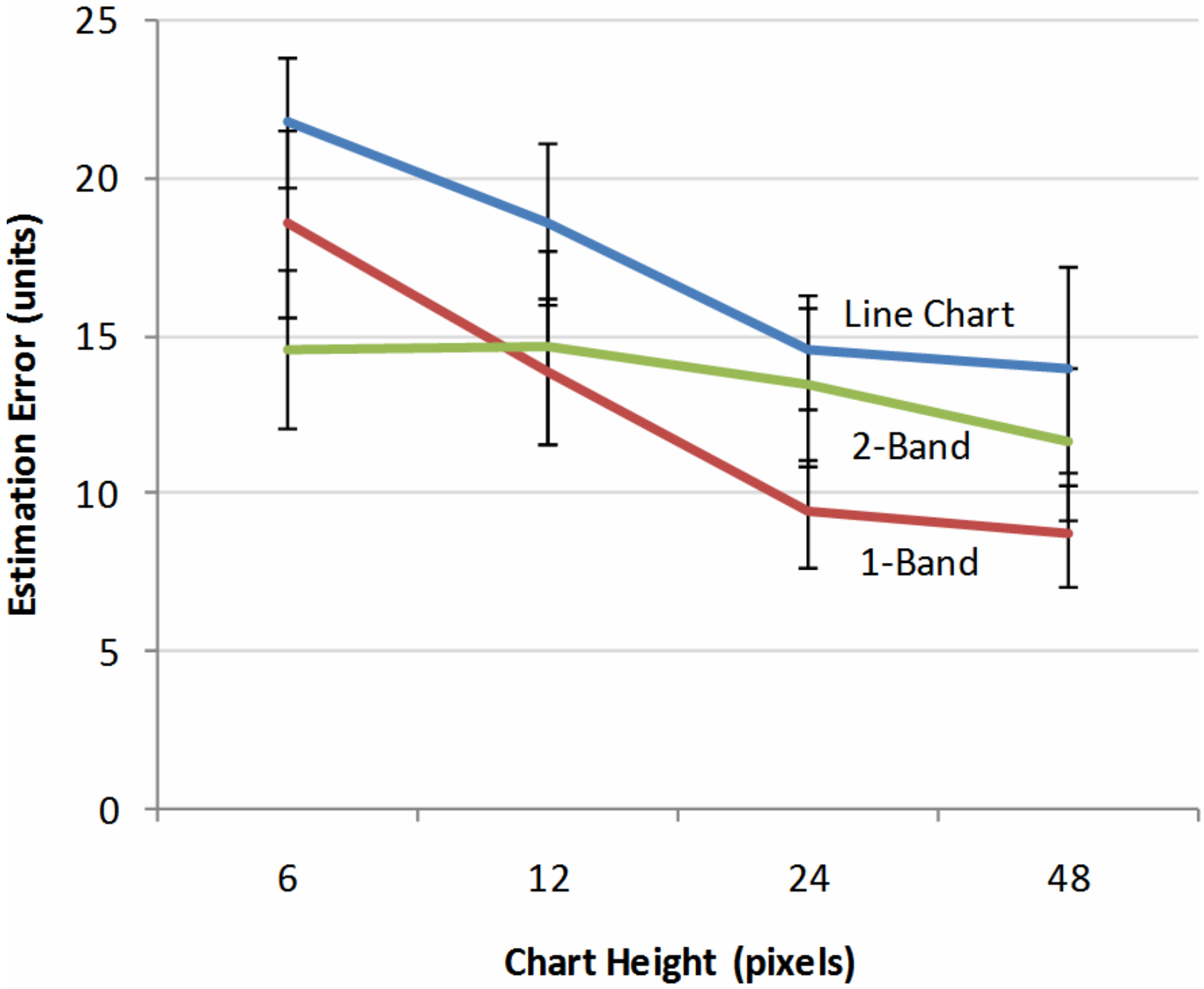


3 (type) x 4 (size) within-subjects design

N = 30 (17 male, 13 female), undergrads

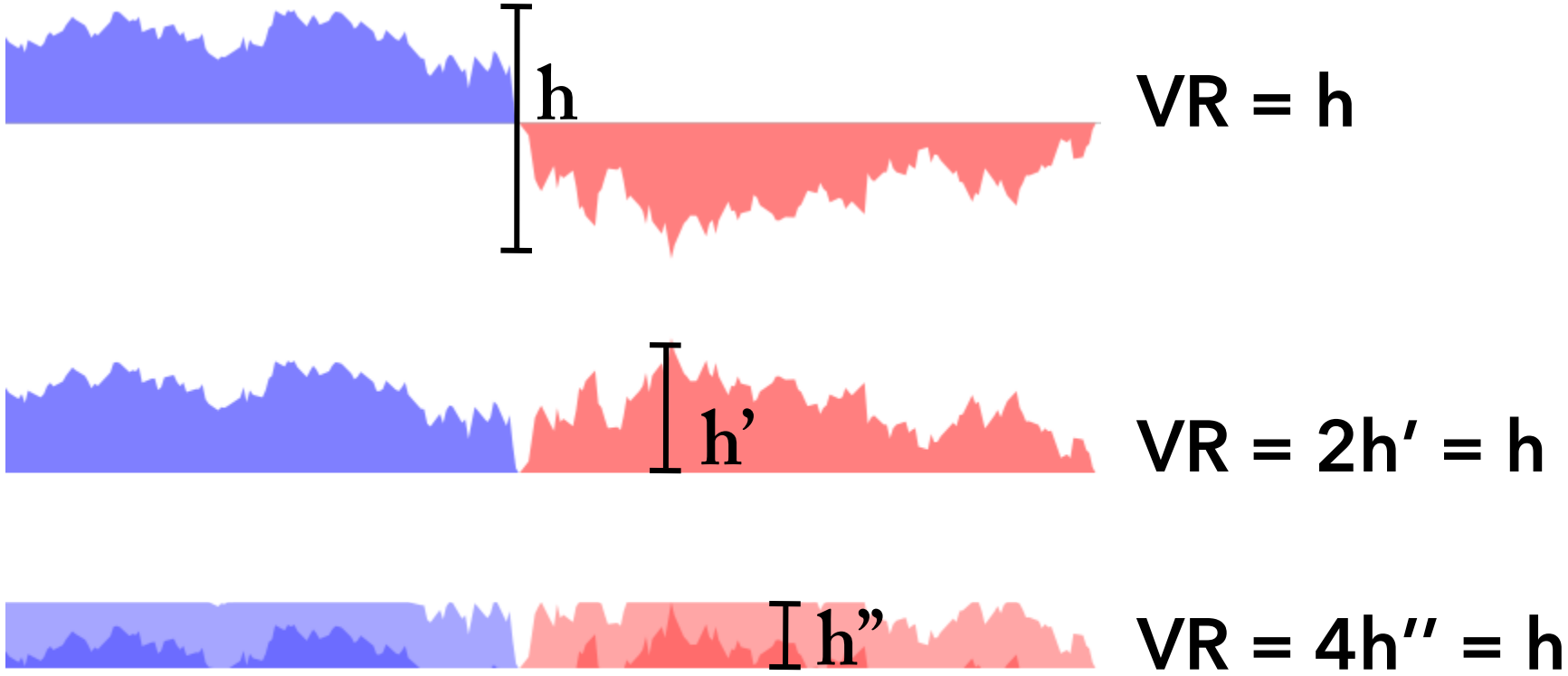
2 (type) x 3 (size:1/8, 1/12, 1/24) follow-up

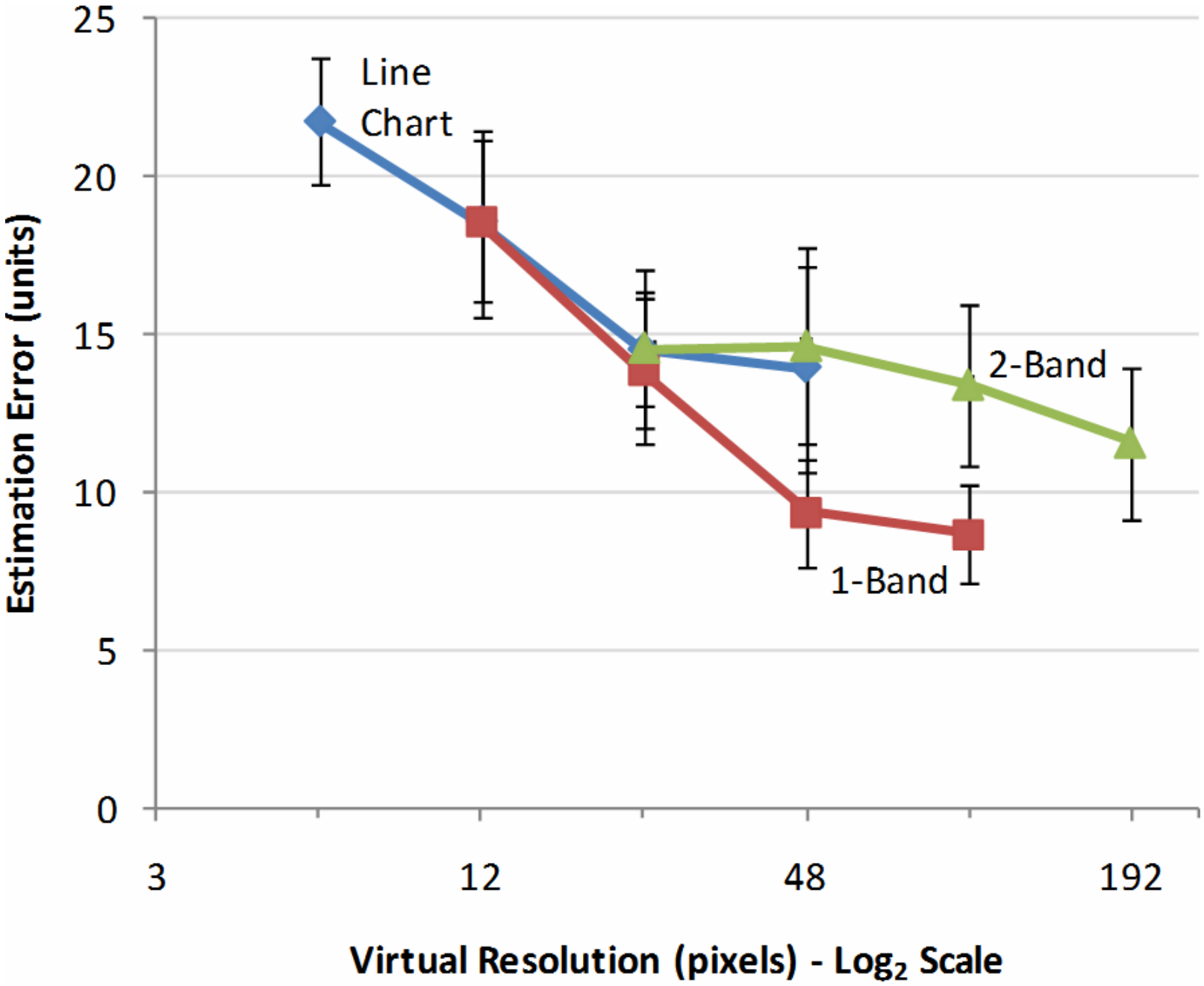
N = 8 (6 male, 2 female), engineering grads

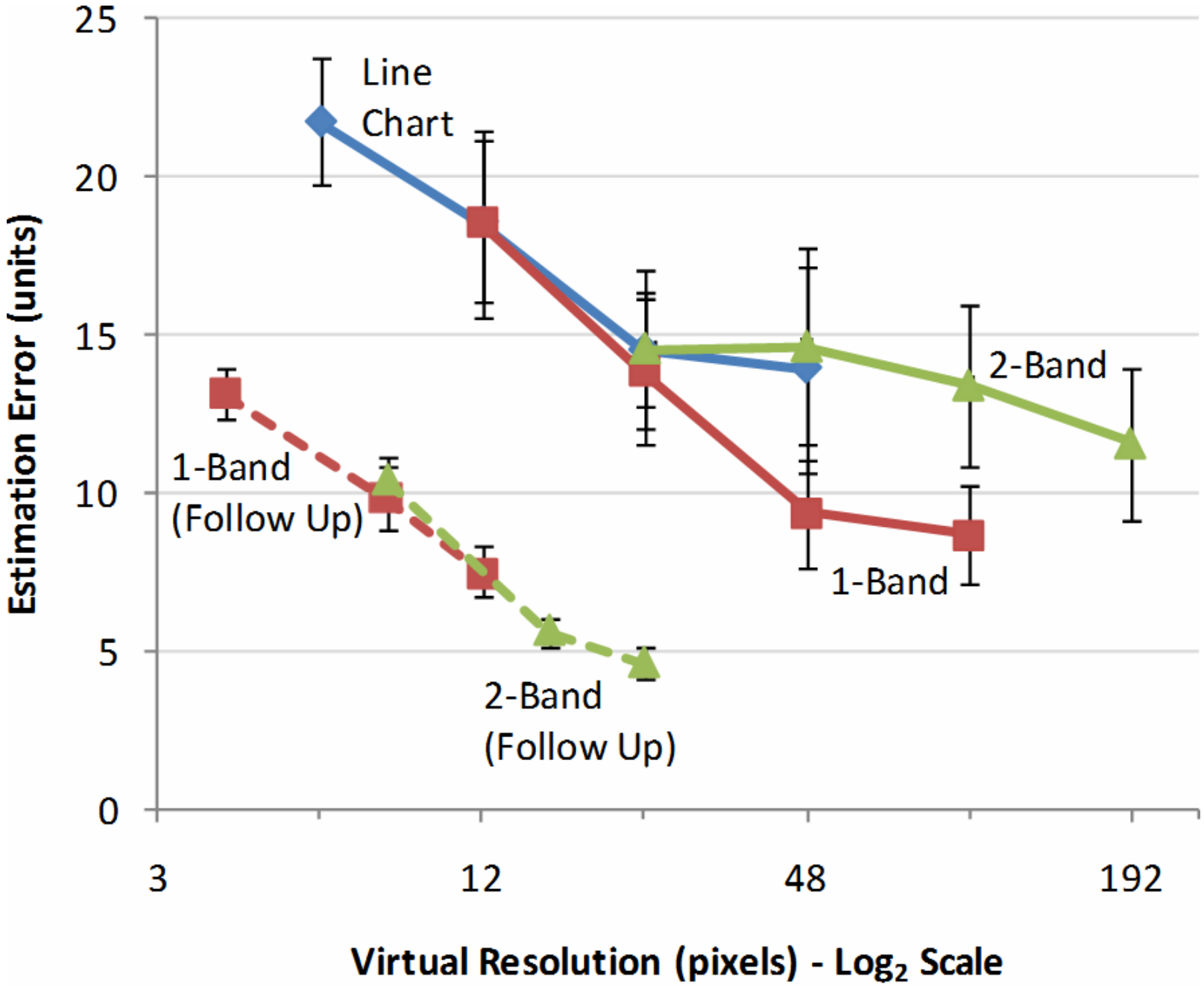

25
20
15
10
5
0
Estimation Error (units)
Line Chart
2-Band
1-Band
6
12
24
48
Chart Height (pixels)

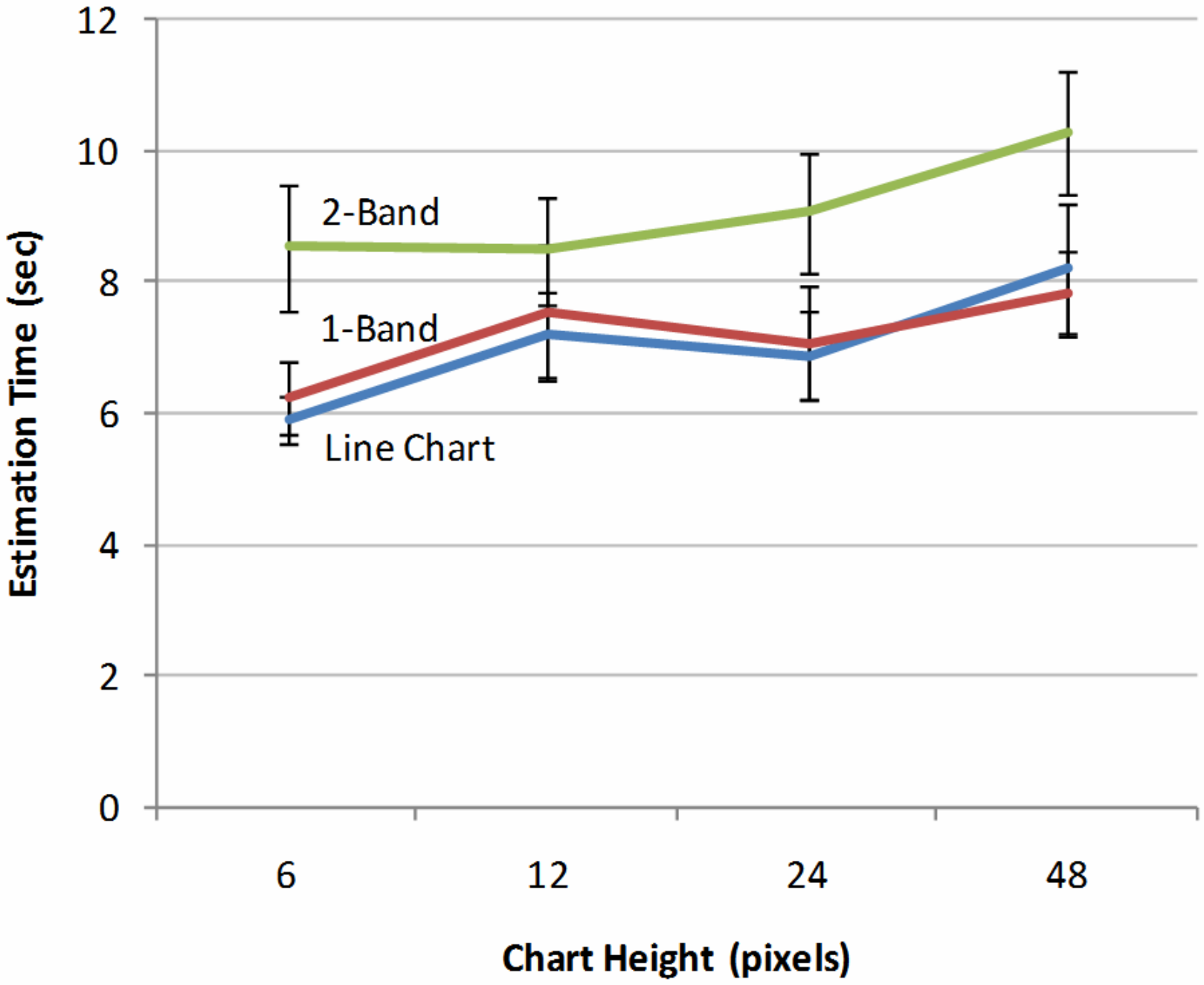
Virtual Resolution (VR)

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




25
20
15
10
5
0
Estimation Error (units)
Line
Chart
2-Band
1-Band
3
12
48
192
Virtual Resolution (pixels) - Log2 Scale


25
20
15
10
5
0
Estimation Error (units)
Line
Chart
2-Band
1-Band
(Follow Up)
1-Band
2-Band
(Follow Up)
3
12
48
192
Virtual Resolution (pixels) - Log2 Scale


12
10
8
6
4
2
0
Estimation Time (sec)
2-Band
1-Band
Line Chart
6
12
24
48
Chart Height (pixels)

Experiment Results

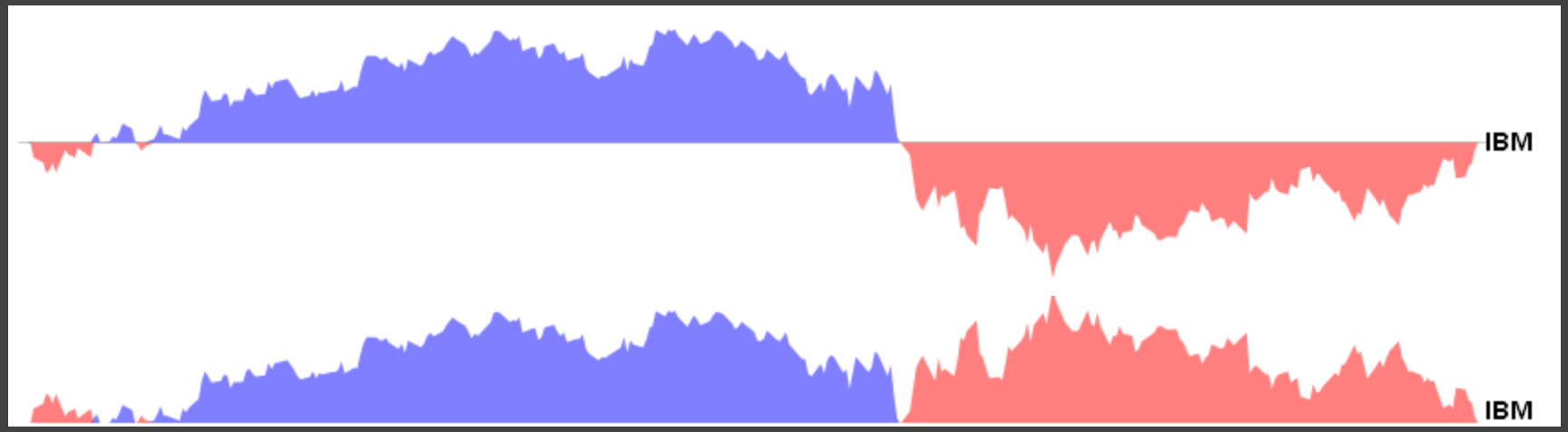
Q1: 2-band horizon graph (but not mirrored graph) has higher baseline estimation time and error.

Q2: Estimation error increases as the *virtual resolution* decreases.

Estimation time decreases as the *physical height* decreases.

Design Guidelines

Mirroring does not hamper perception

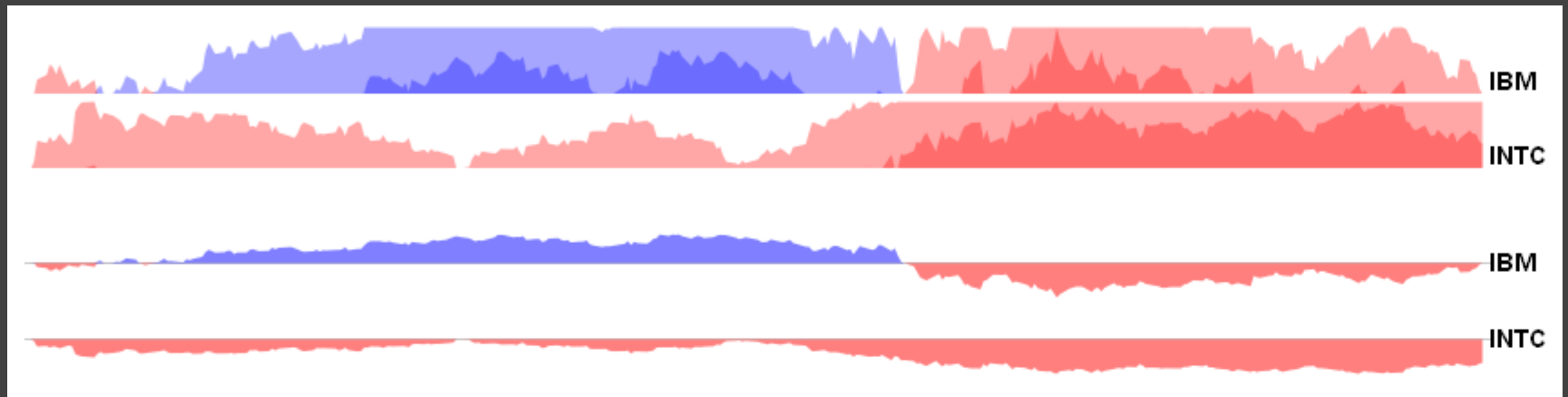


Design Guidelines

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?

Administrivia

Final Project

Poster Presentations

Session is **Tue Jun 7 5-8pm** in CSE Atrium
Bring **Poster** + **Laptop/Device** for demos
Arrive *early* to setup!

Post Webpage on GitHub Pages

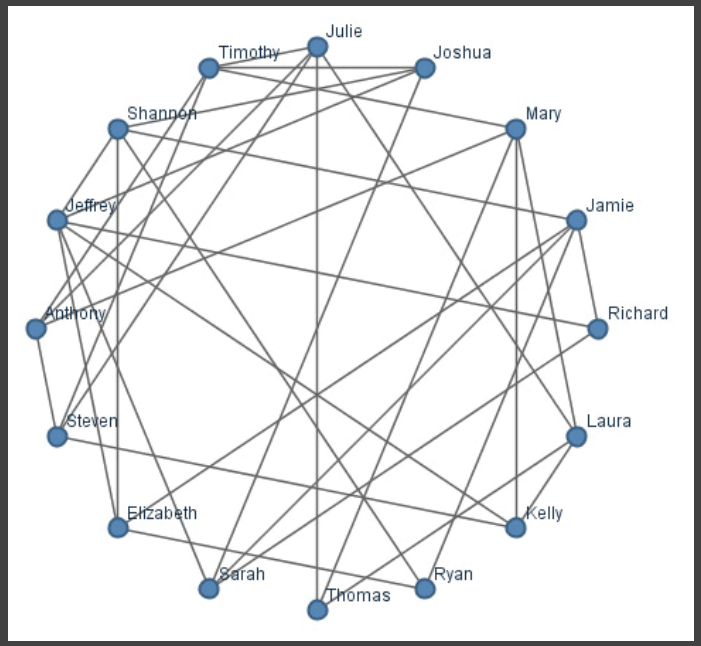
List team members, title, abstract, link to paper
Include summary image for project!

Final Project Reports

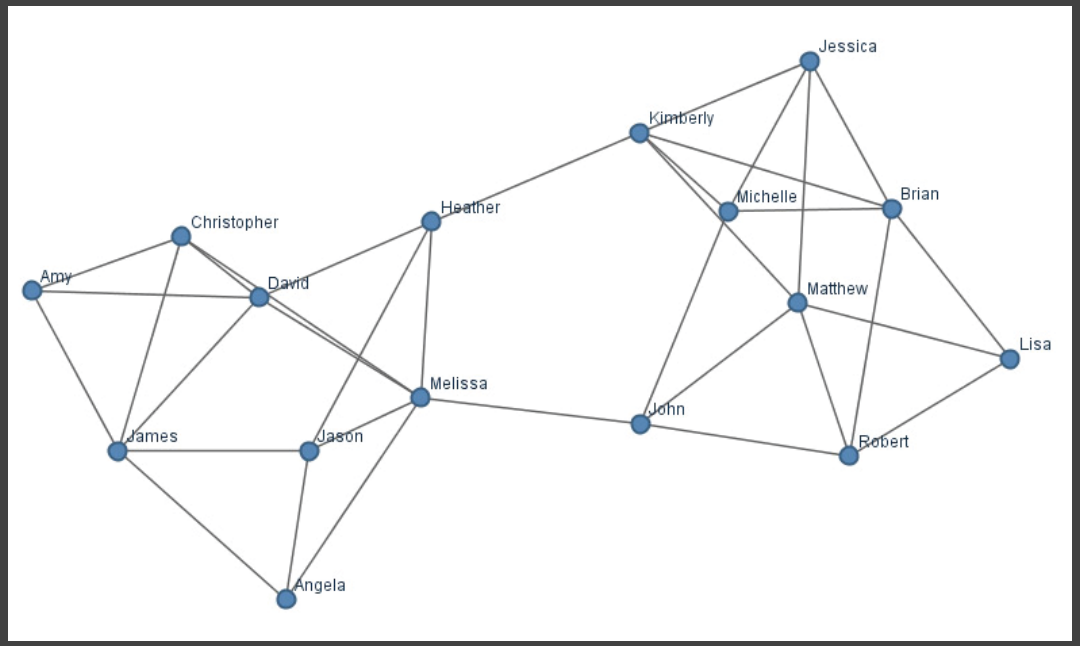
Due **Thu Jun 9**, by **8am**, posted to GitHub
4-6 pages in ACM or IEEE TVCG format

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

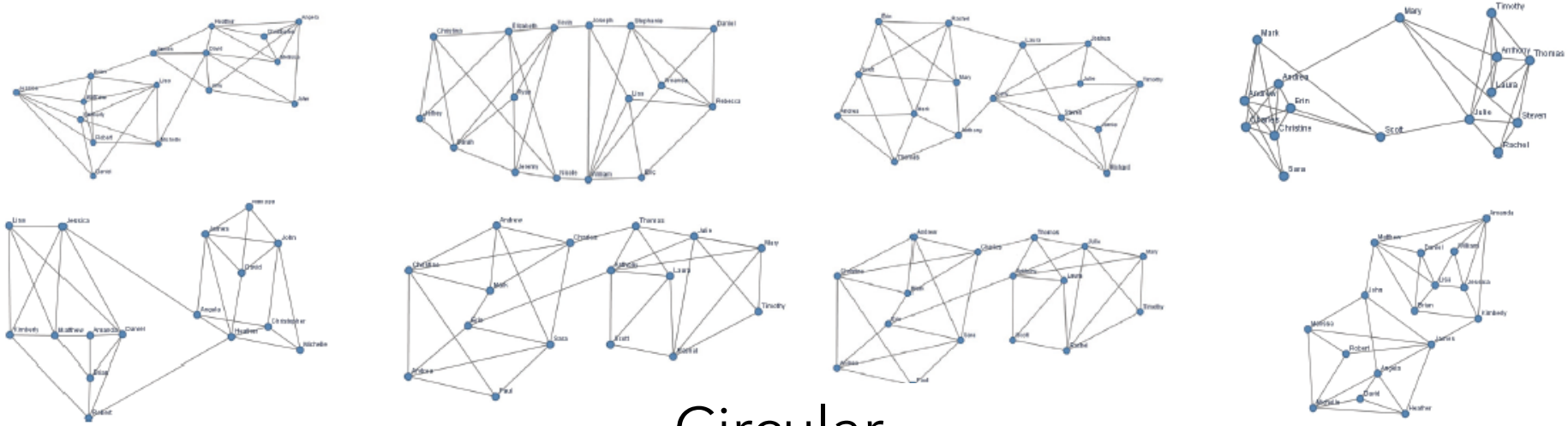
Measures

# of Edge Crossings

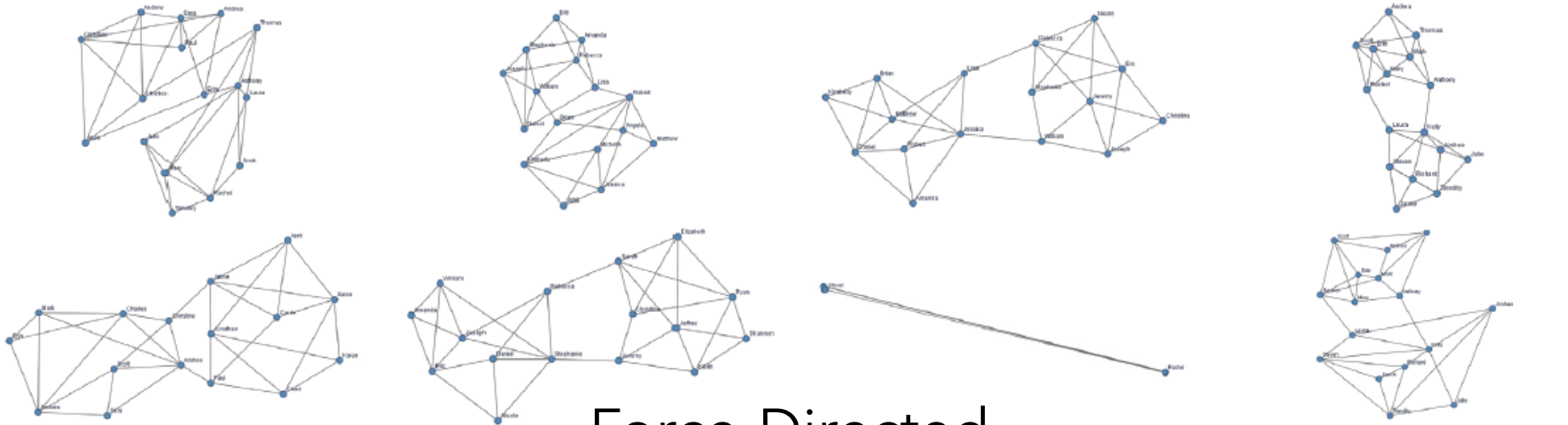
Average Edge Length

Average Node Distance

within or between clusters



Circular



Force-Directed

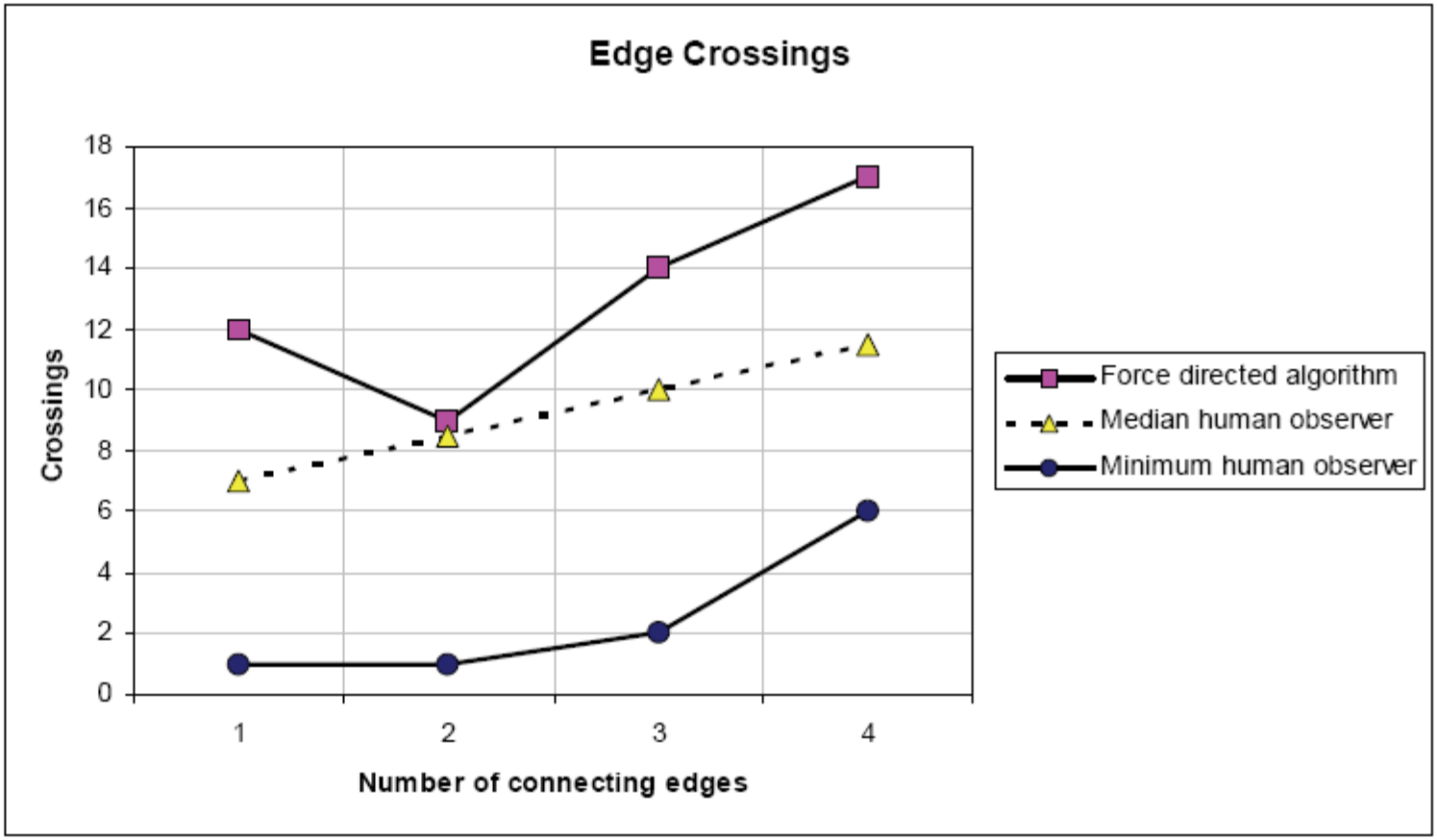


Figure 4. Edge Crossings. Human observers produced graph layouts with fewer edge crossings than the force-directed graph algorithm.

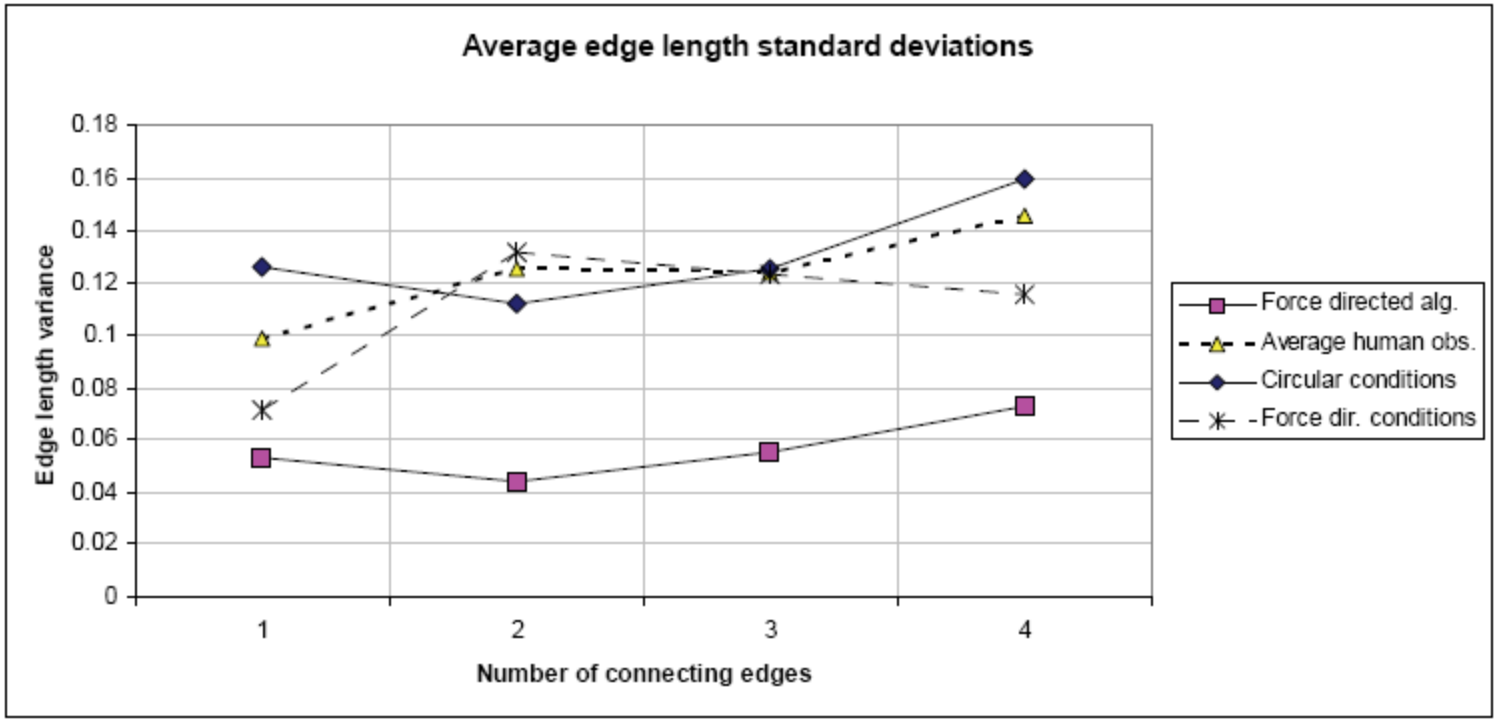


Figure 5. Edge Length Distribution. Human observers did not focus on maintaining equal edge length as much as the force directed algorithm.

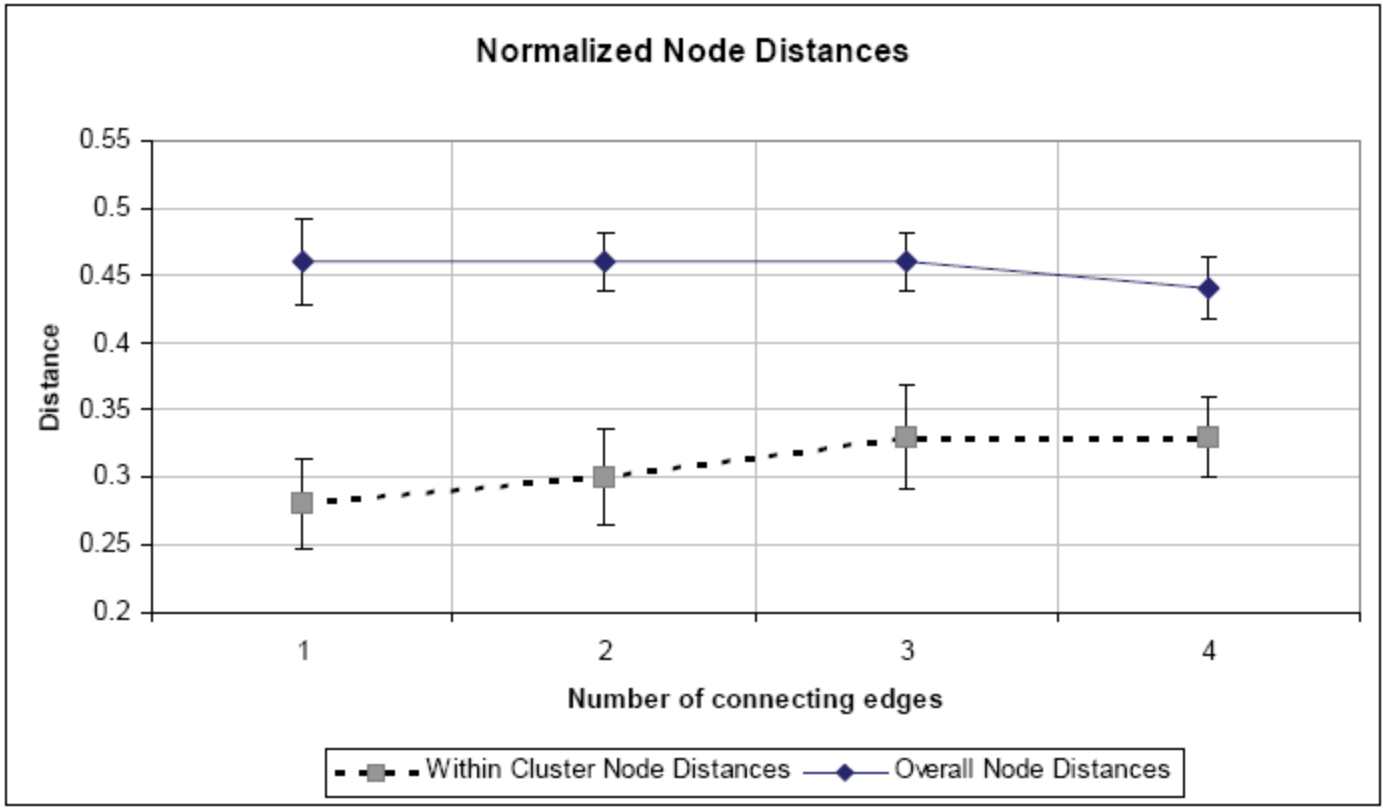


Figure 7. Cluster Extraction. For all levels of masking, the distance between nodes within a cluster is significantly smaller than the overall inter-node distance, demonstrating perceptual grouping. Error bars show 95% confidence intervals

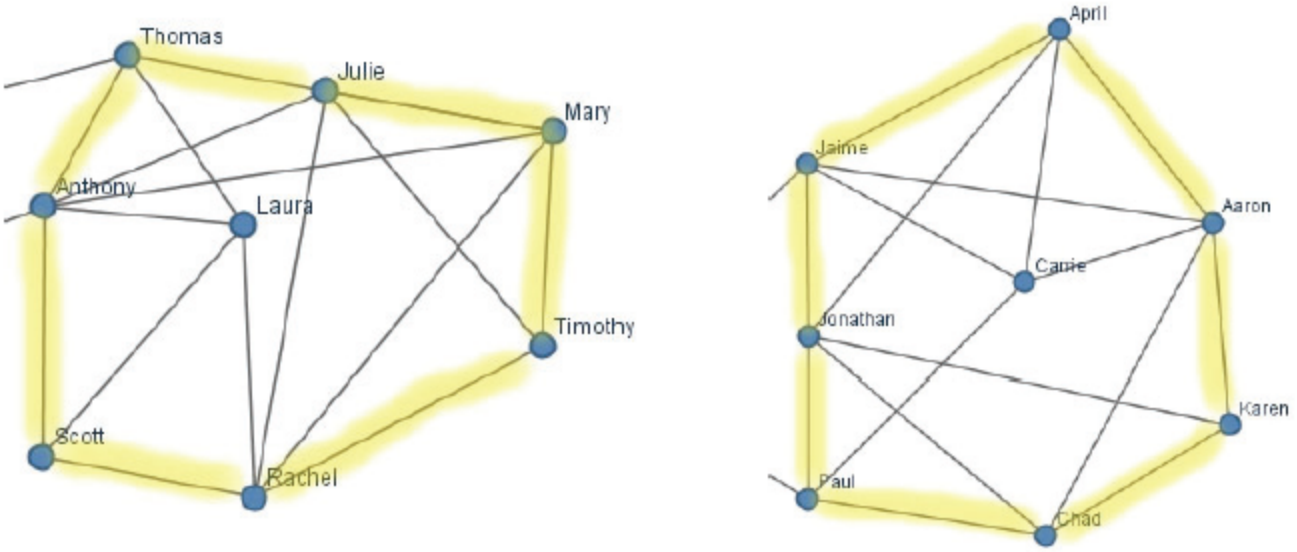
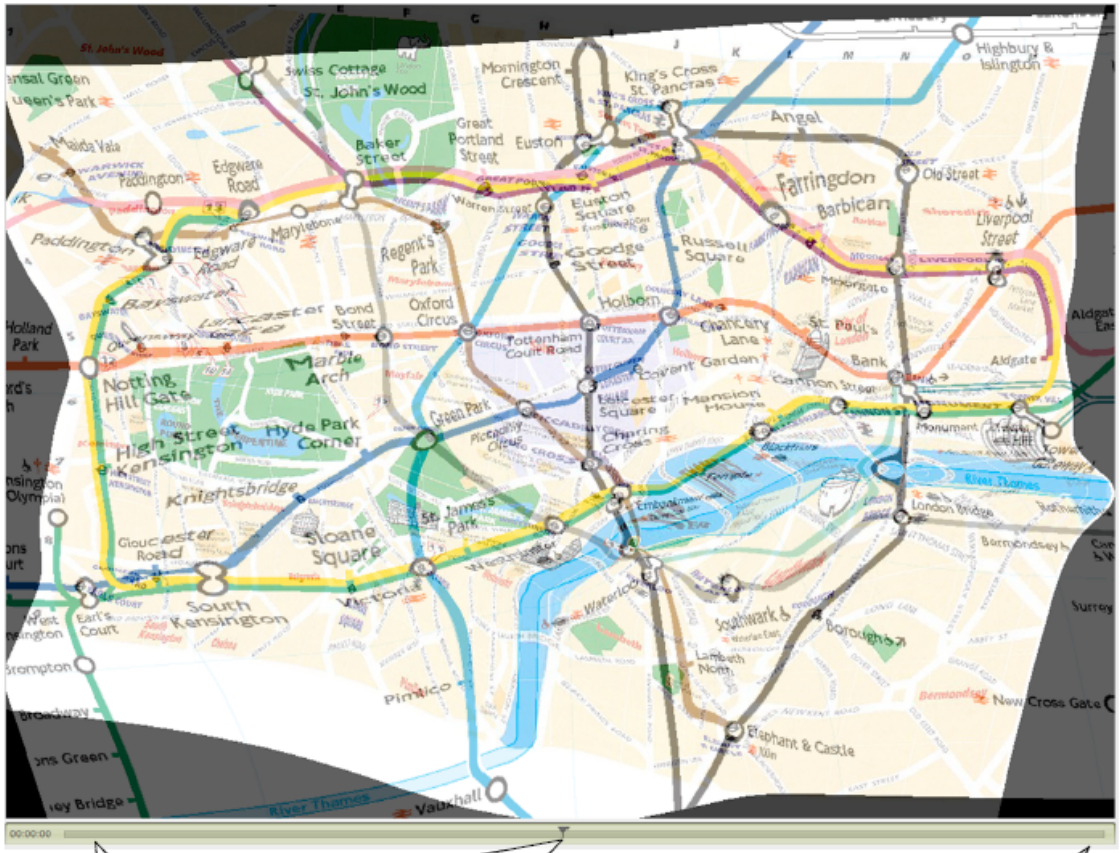


Figure 9. Cluster Hulls. Two examples of user-generated layouts where cluster edges formed a hull enclosing the cluster, organizing it into a single perceptual group.

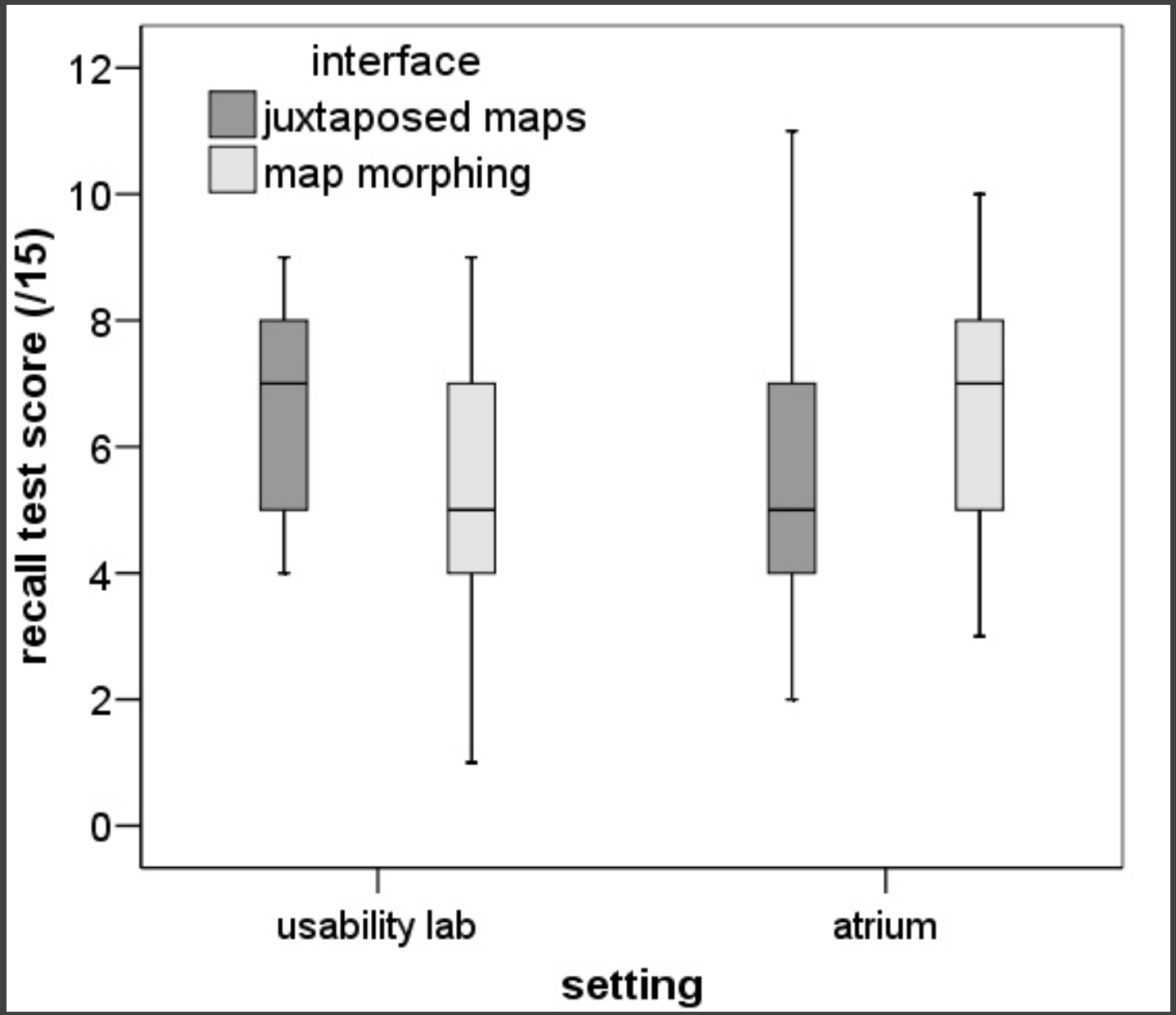
Role of Environment [Reilly '07]



The slider was used to control the bidirectional morph. Shape blending (distortion) and alpha blending occurred simultaneously, in proportion to the distance of the slider from each endpoint.

The Tube map was at the left endpoint of the morph.

The right endpoint showed the tourist map.

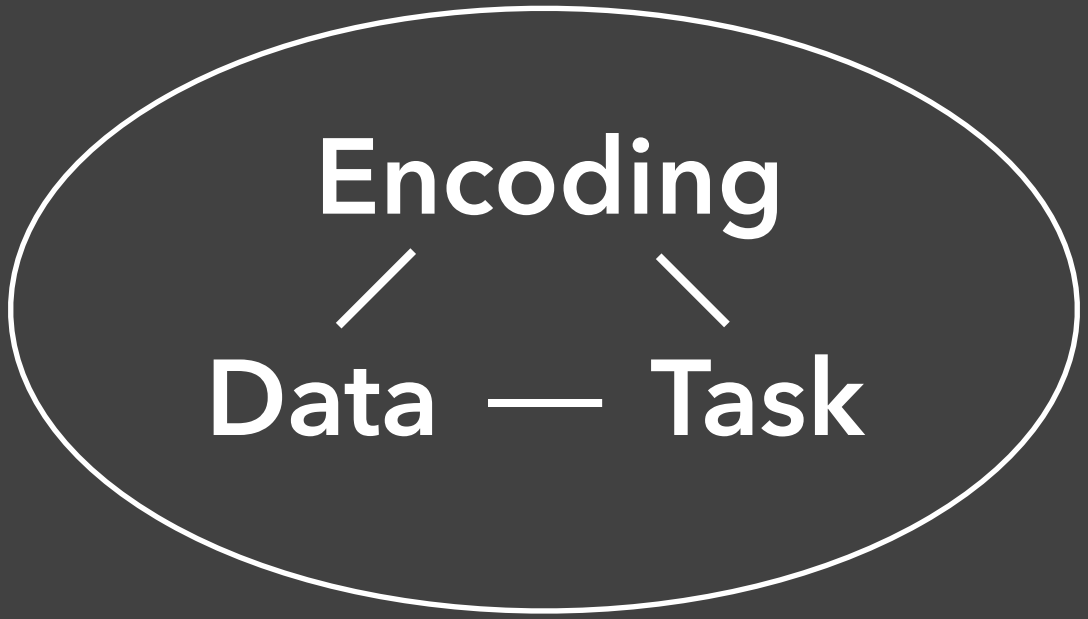

interface
juxtaposed maps
map morphing
recall test score (/15)
12
10
8
6
4
2
0
usability lab
atrium
setting

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.

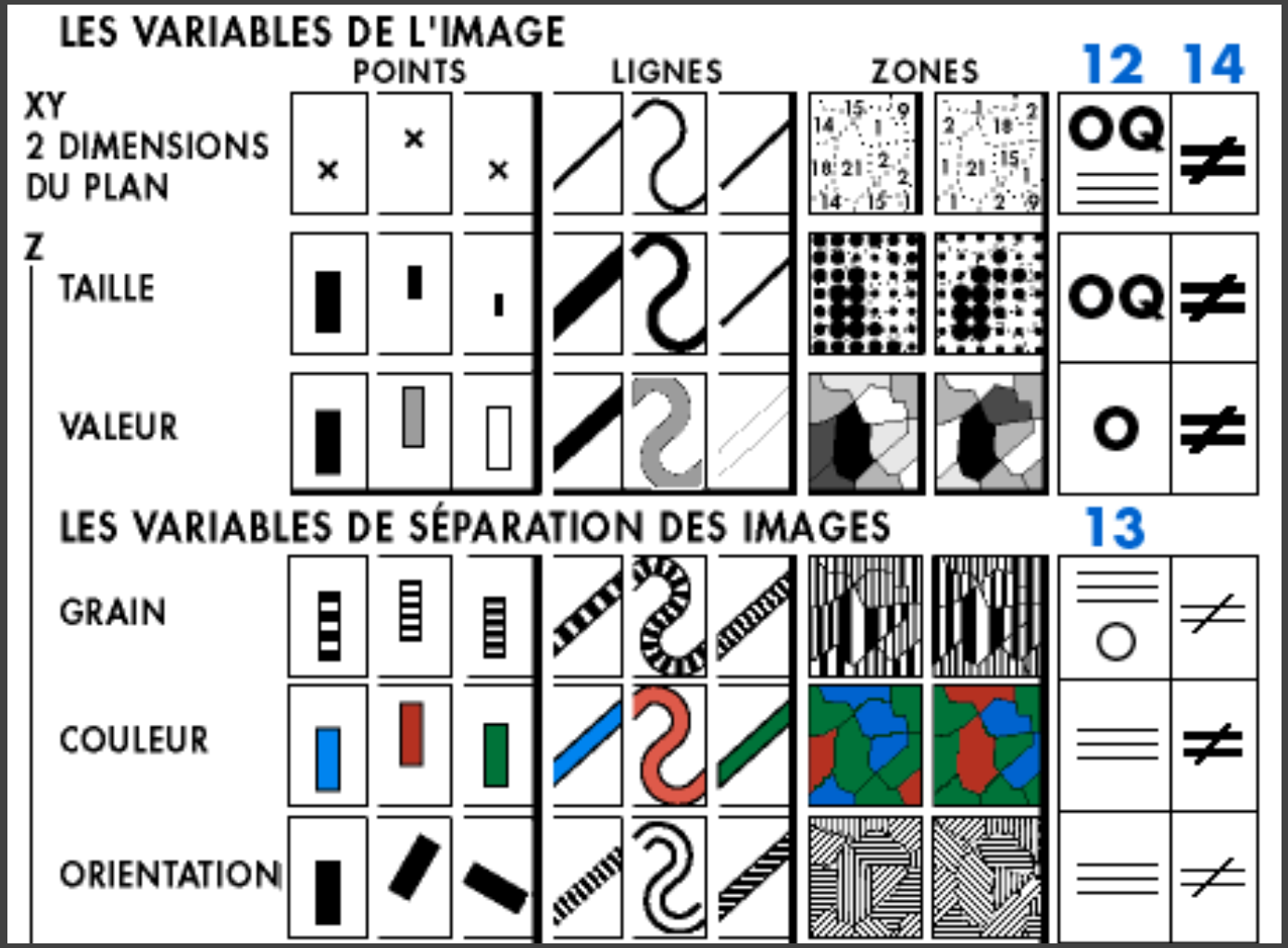
Performance measures may be more suited to serious analysis than casual use?



Users & Domain

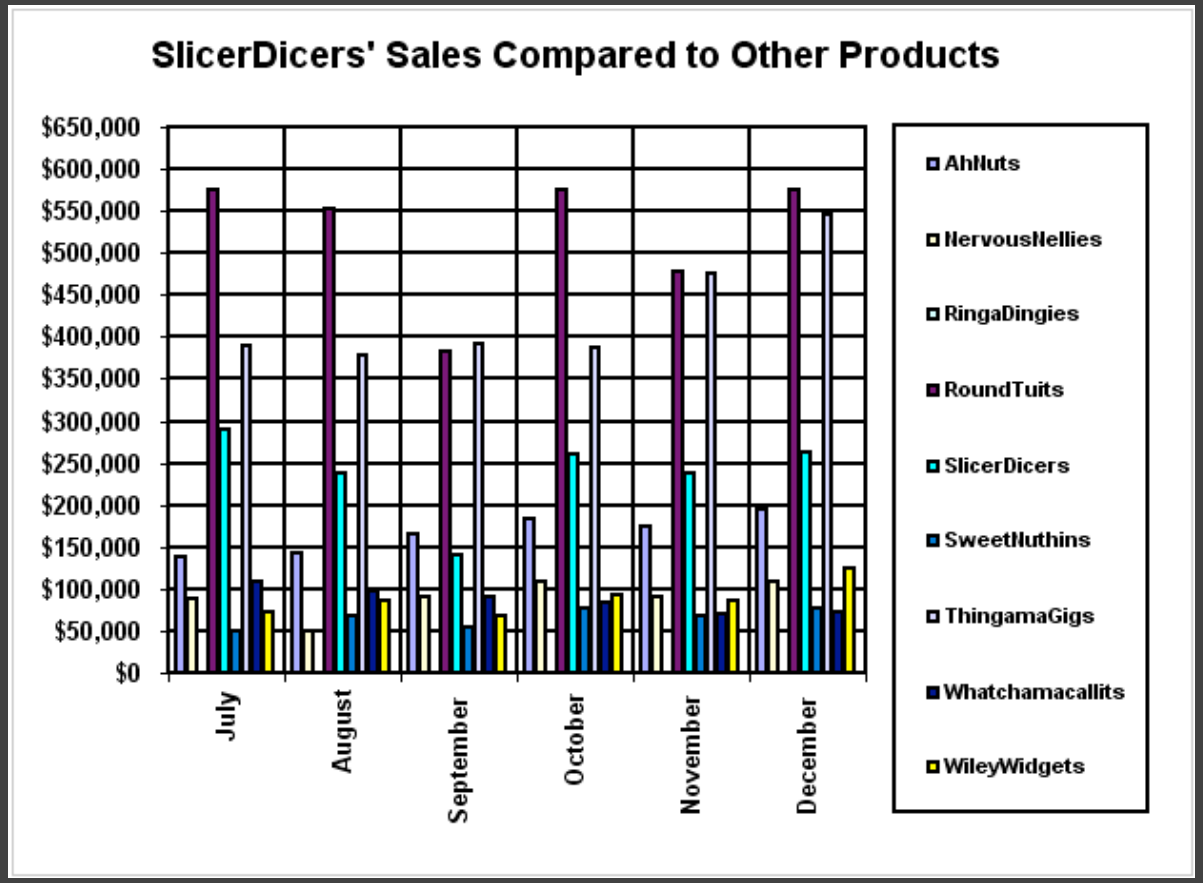
Course Summary

Data and Image Models

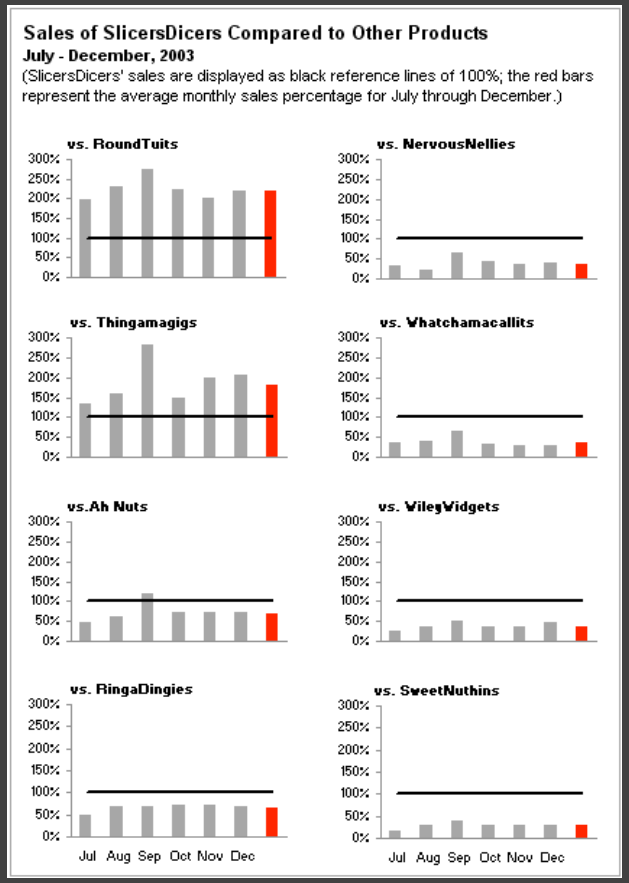


Sémiologie Graphique [Bertin 67]

Visualization (Re-)Design

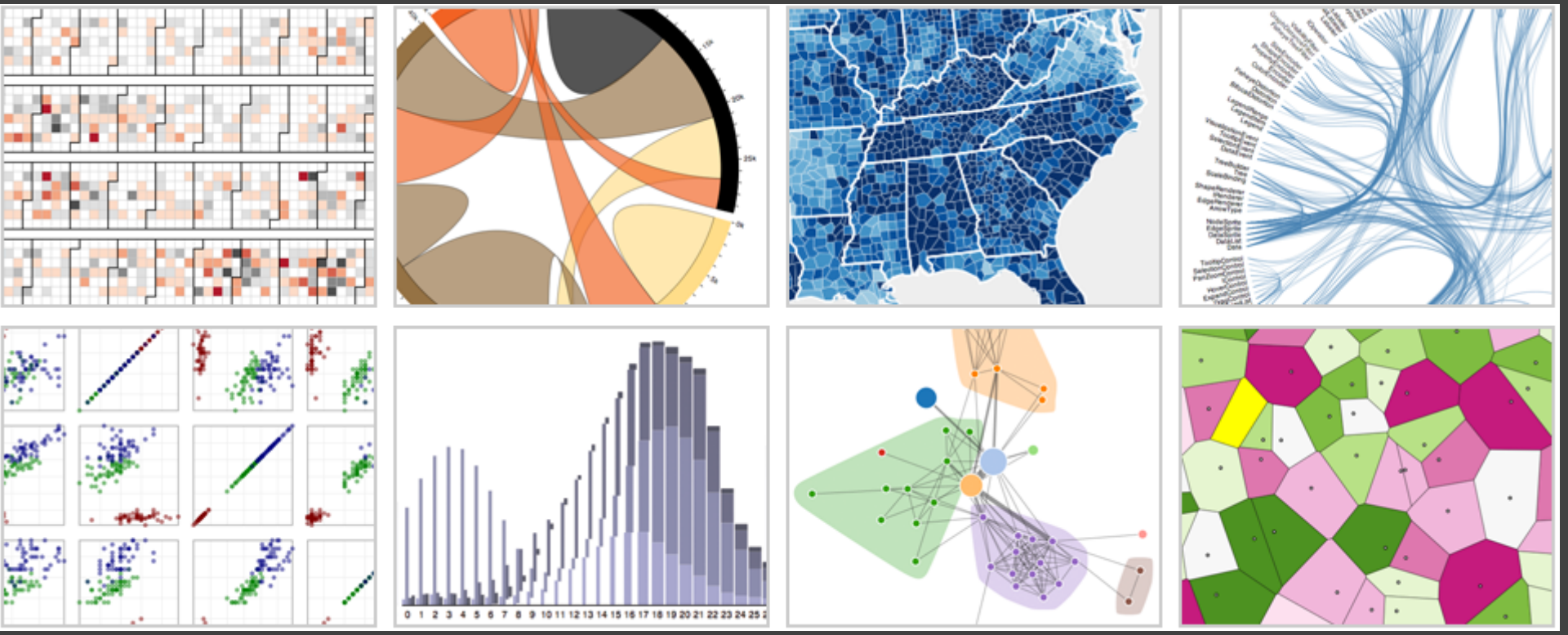


Problematic design



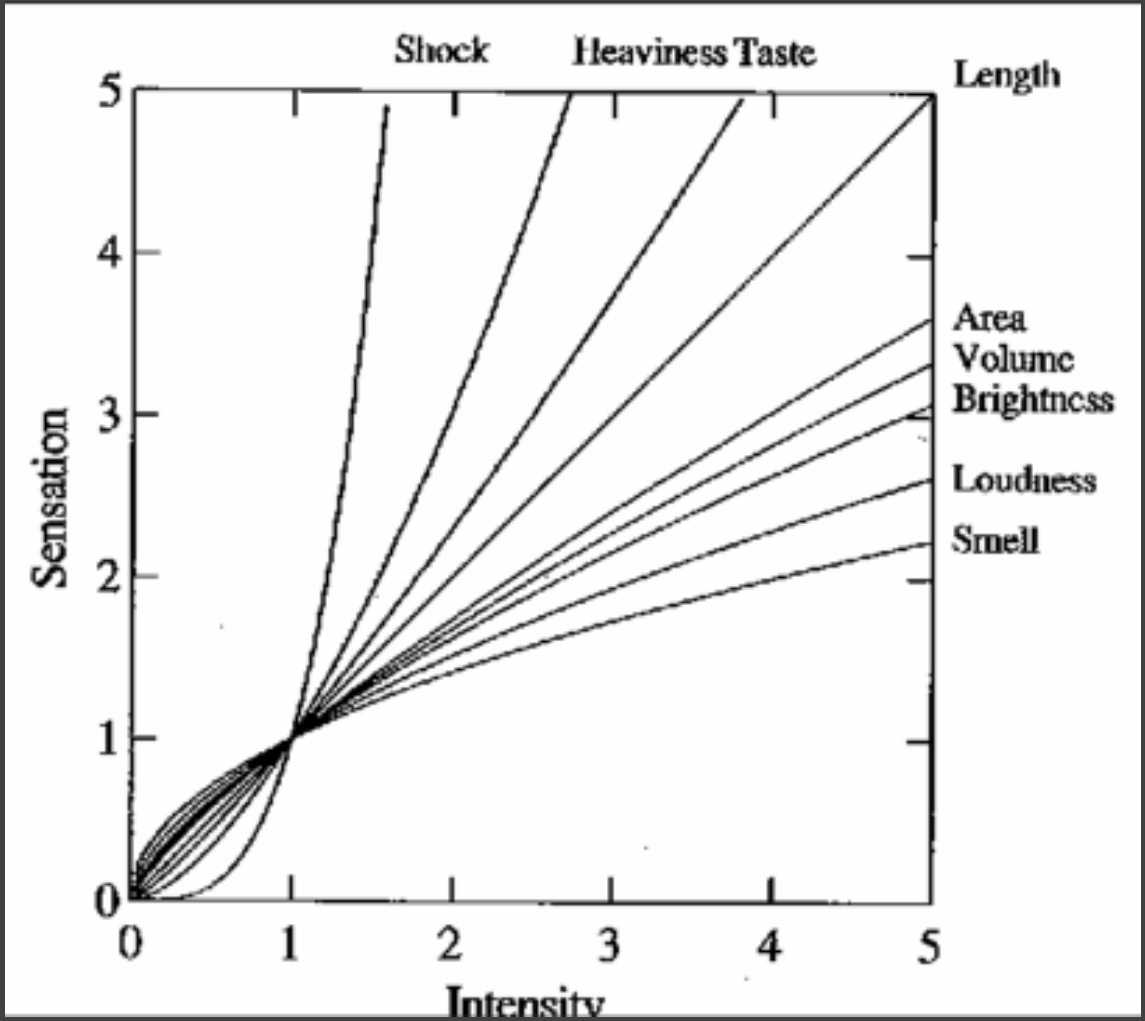
Redesign

Visualization Software



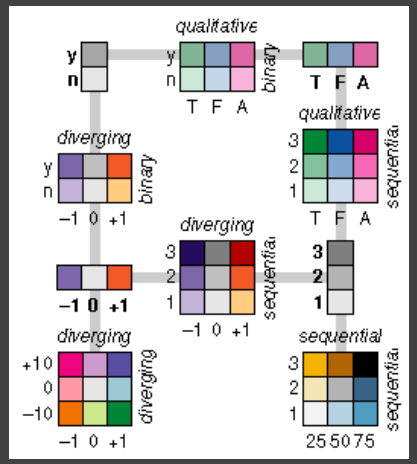
D3: Data-Driven Documents

Graphical Perception

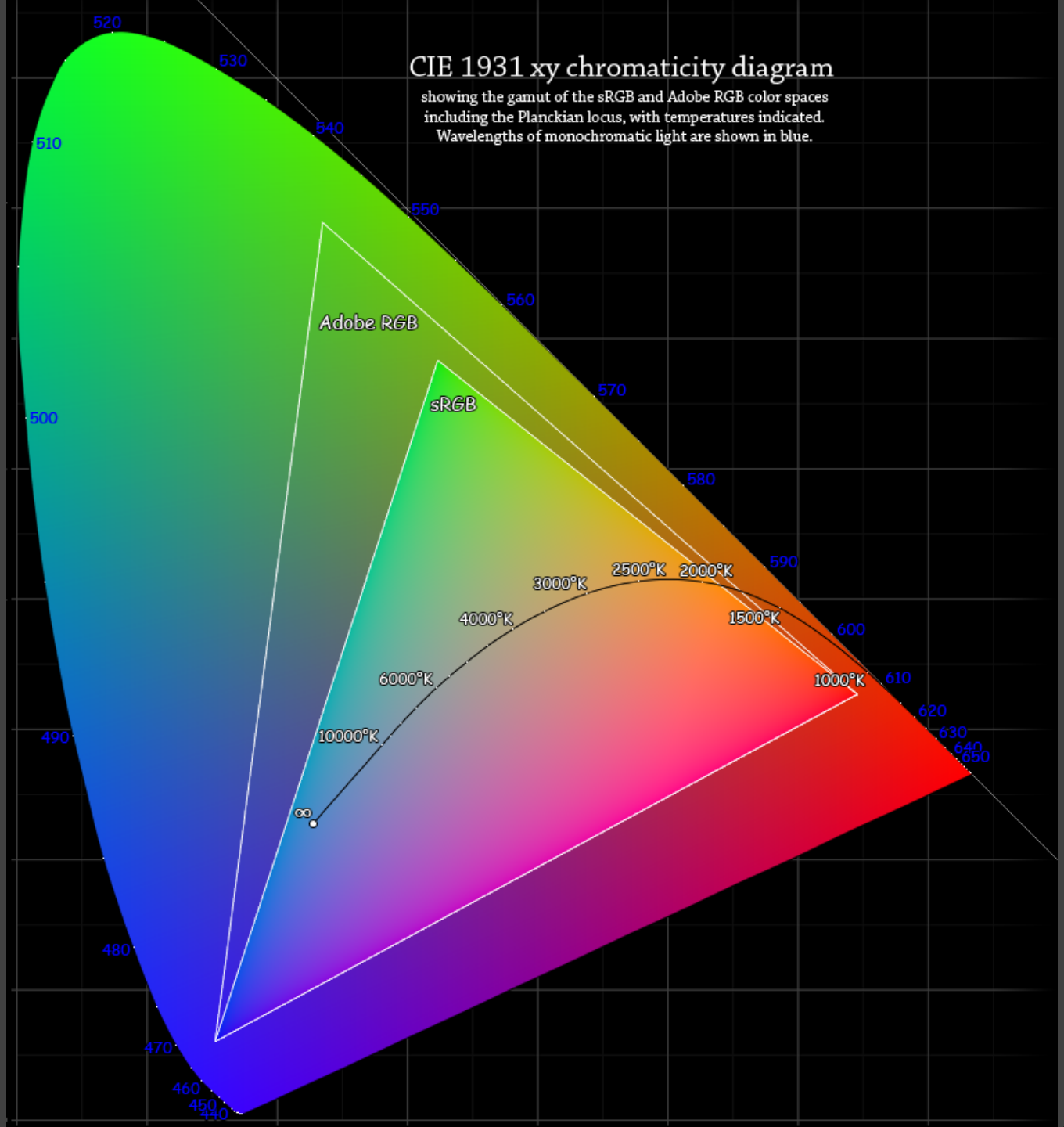


The psychophysics of sensory function [Stevens 61]

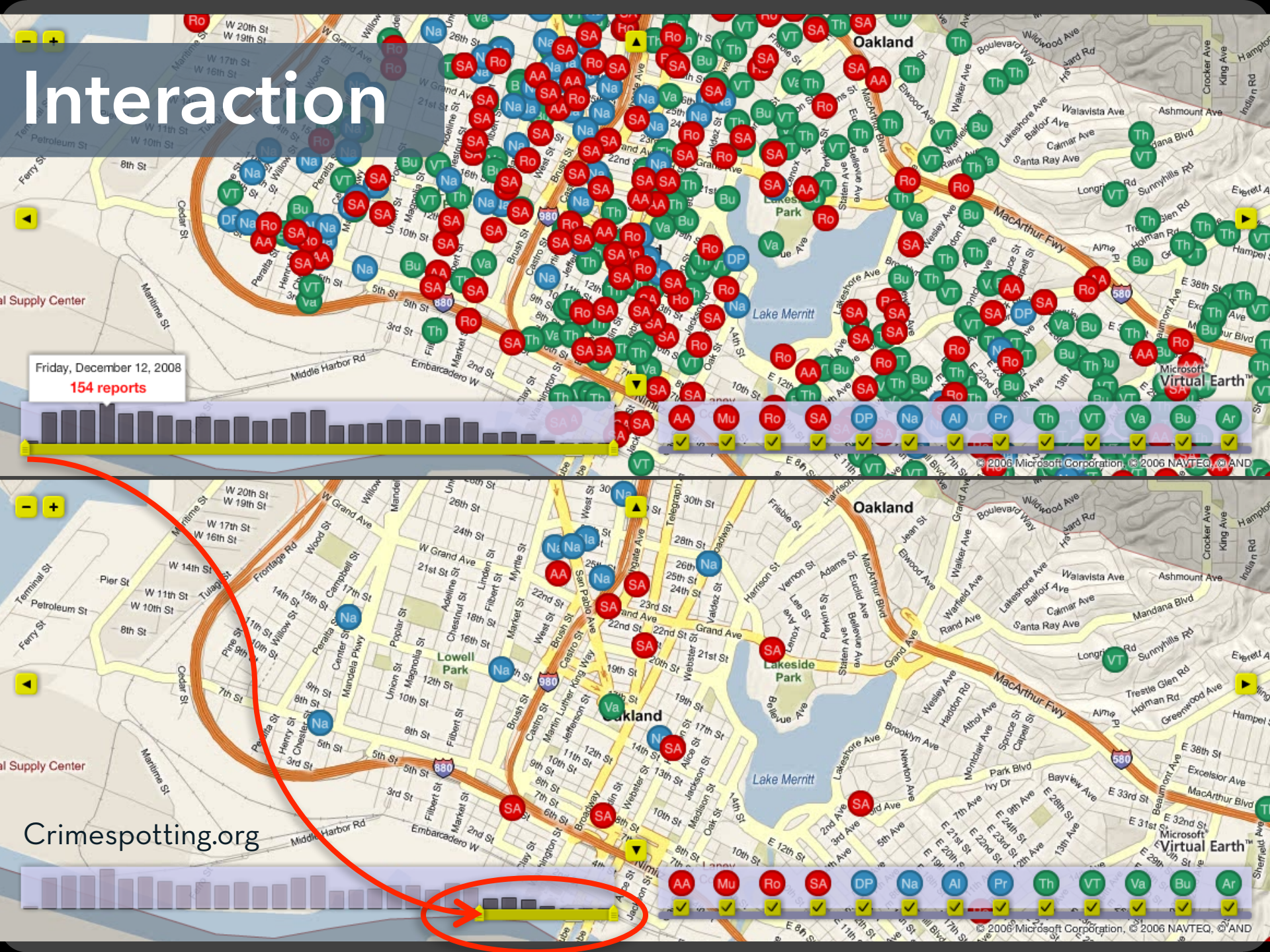
Color



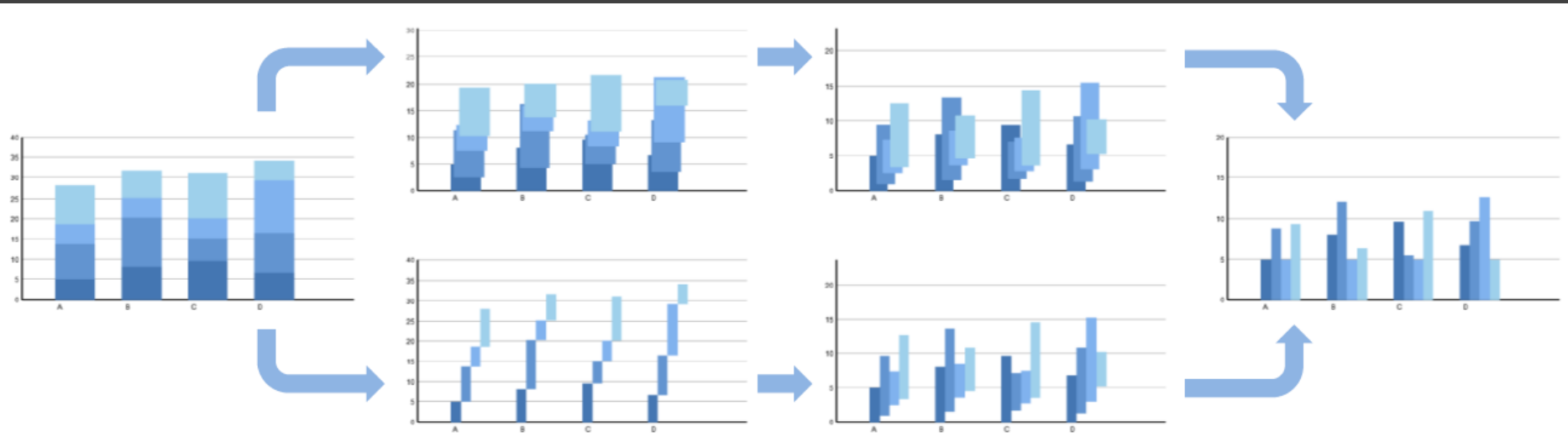
Color Brewer



Interaction

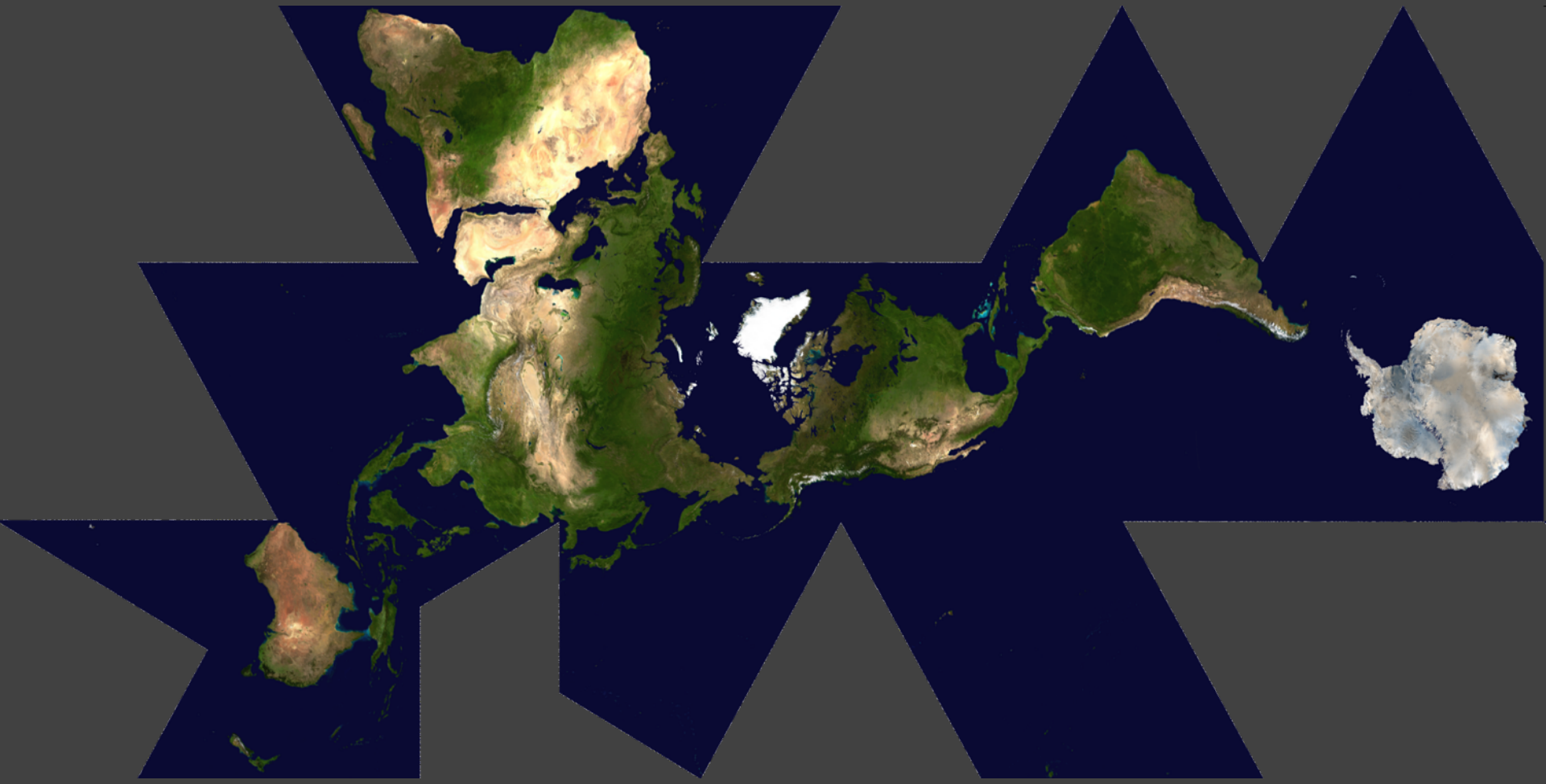


Animation



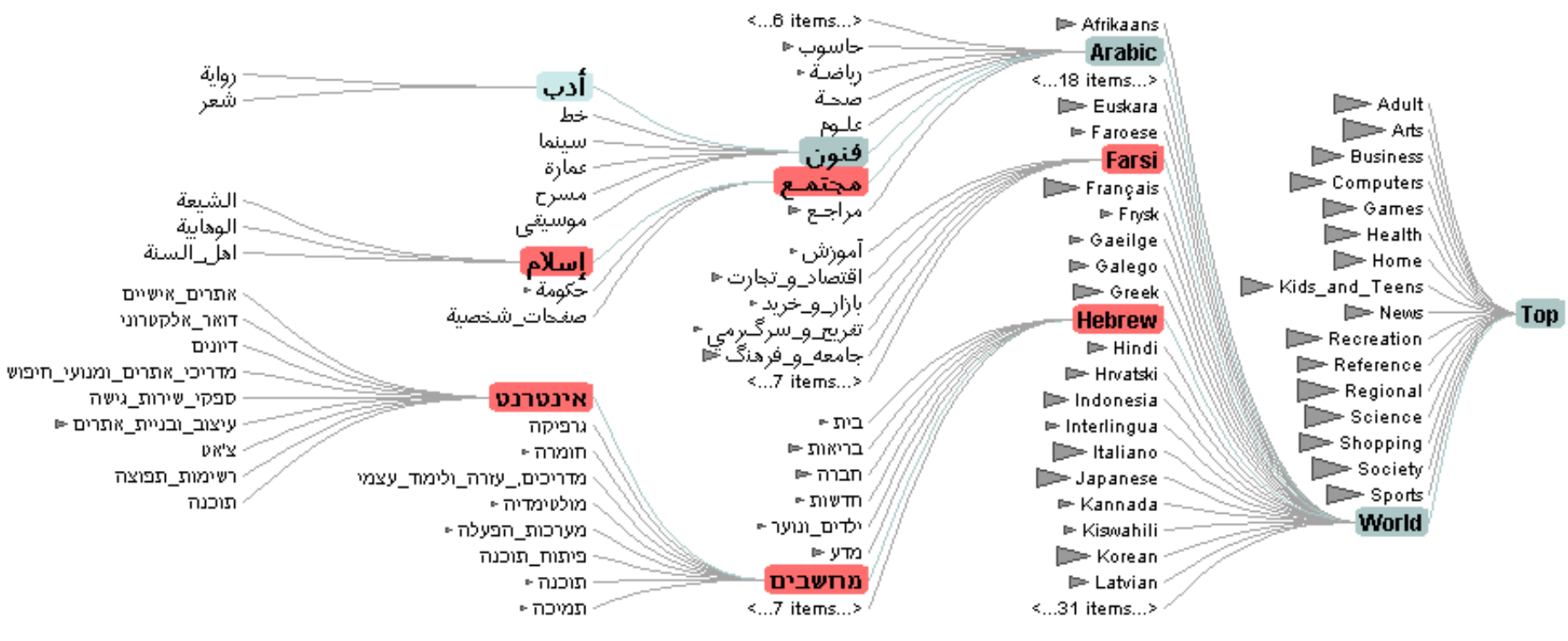
Animated transitions in statistical data graphics [Heer & Robertson 07]

Mapping & Cartography



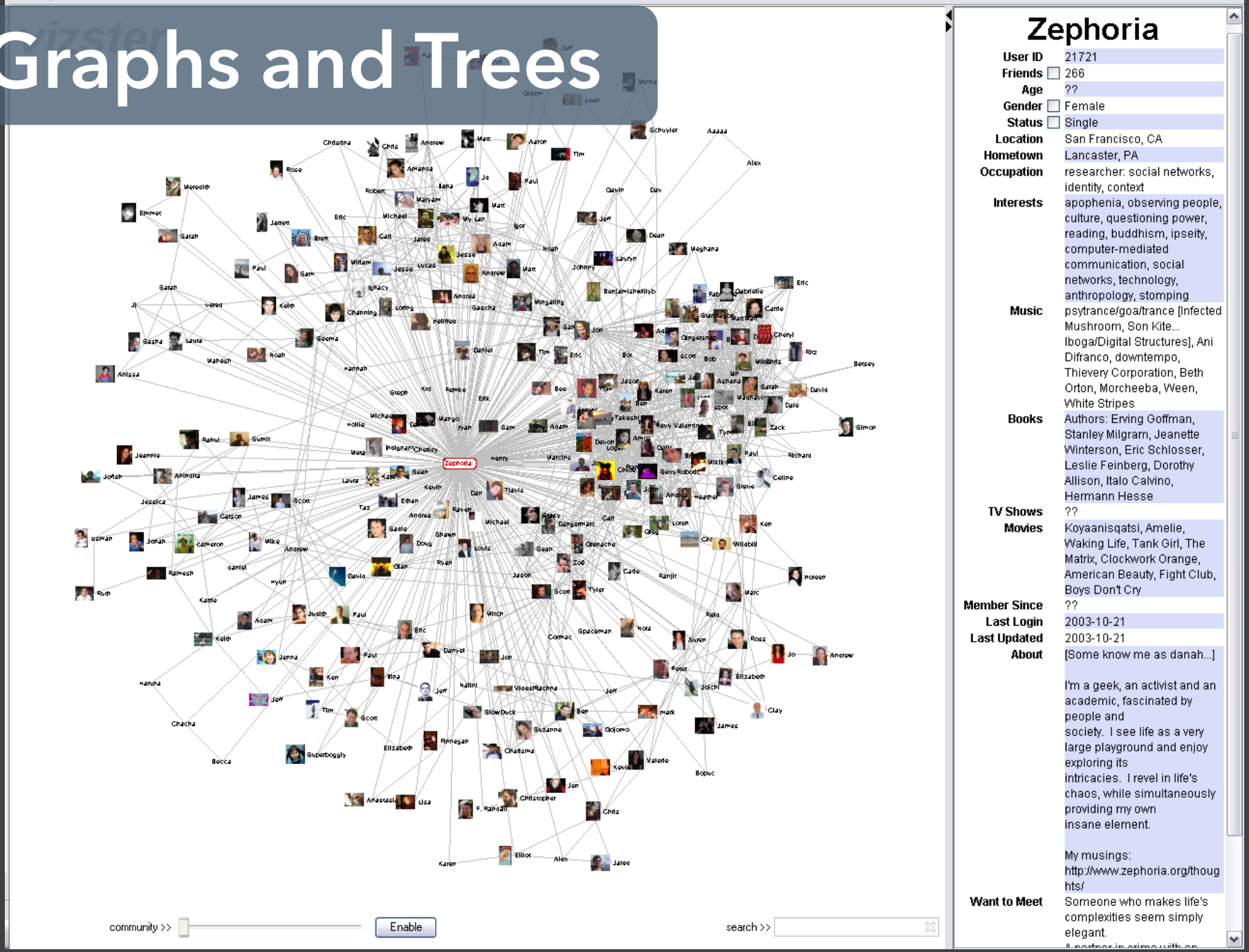
Dymaxion Maps [Fuller 46]

Graphs and Trees



Degree-Of-Interest Trees [Heer & Card 04]

Graphs and Trees



Text Visualization

Visualizations : Word tree / Alberto Gonzales

Creator: Martin Wattenberg
Tags:

explore
visualizations
data sets
comments
topic hubs

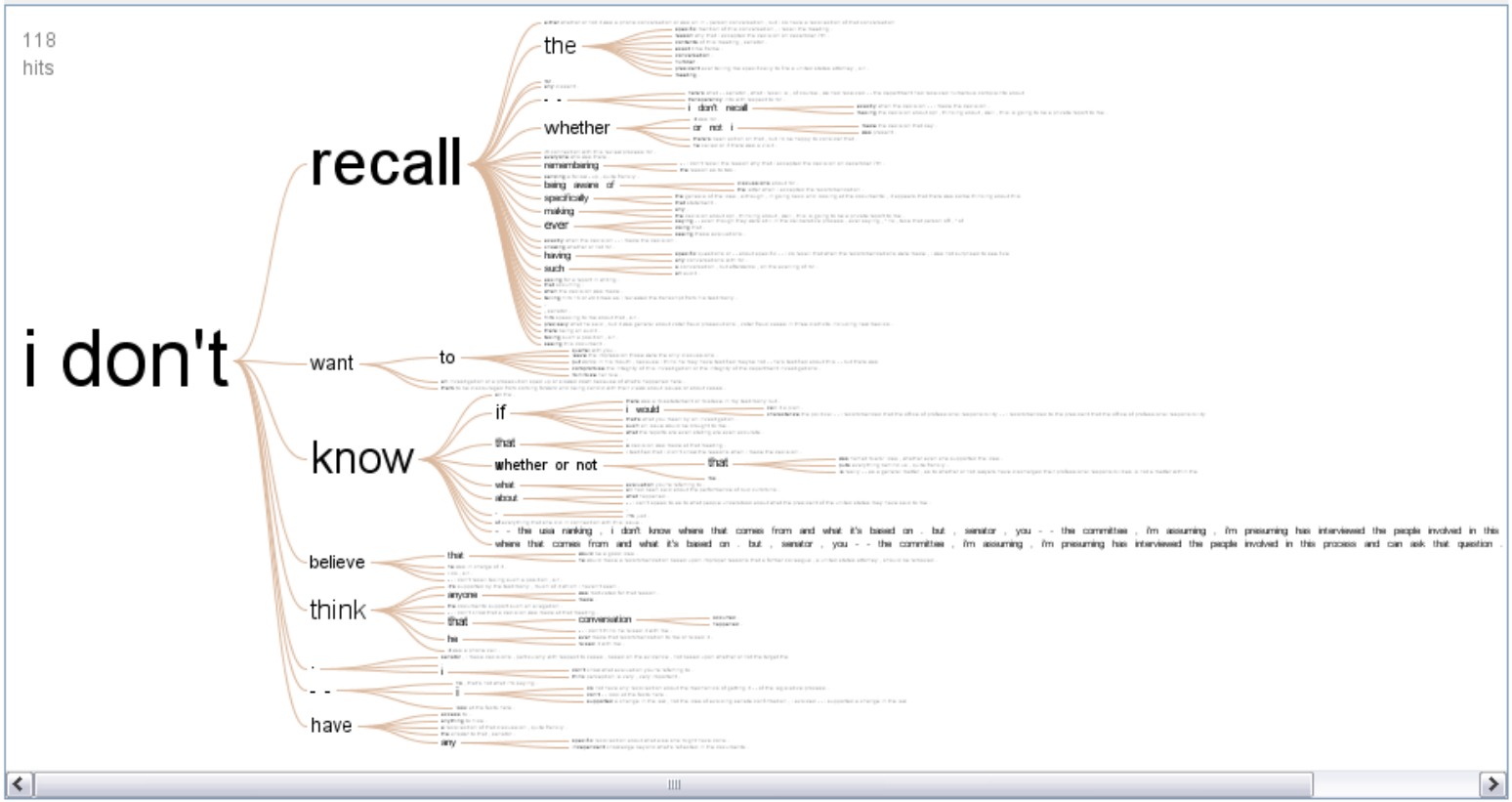
participate
create visualization
upload data set
create topic hub
register

learn more
quick start
visualization types
data format & style
about Many Eyes
FAQ
blog

contact Us
contact
report a bug

legal
terms of use

Search i don't | Back | Forward | Start | End | Occurrence Order | Clicks Will Zoom



Data file: Word in testimony from Gonzales, 4/19/2007 Data source: CQ Transcript Wire via the Washington Post This data set has not yet been rated

full image | share this | watch this | add to topic hub | rate this

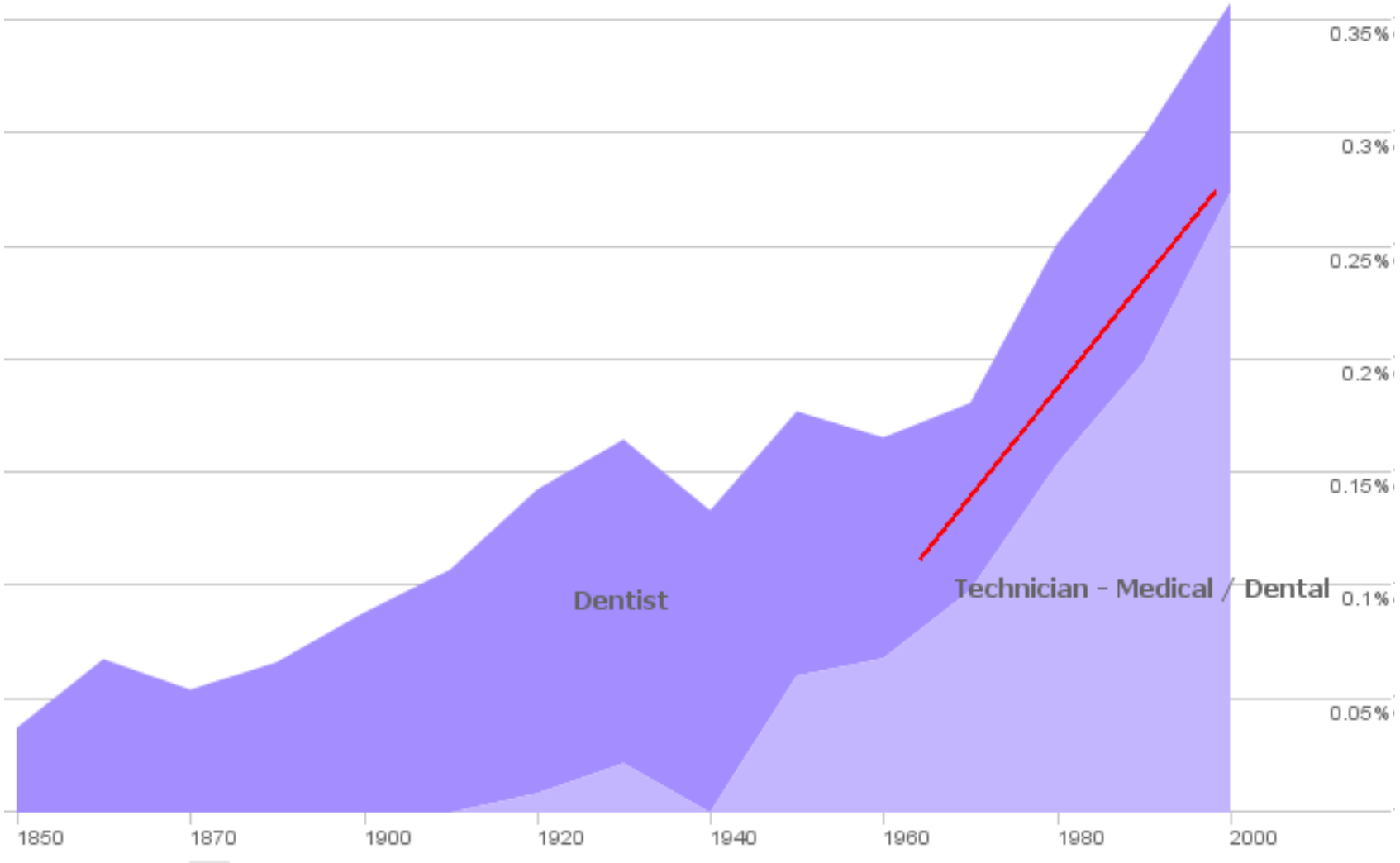
Comments (4)

currently showing

This visualization has 4 positive and 0 negative

Collaborative Analysis

Where have all the dentists gone?



The Future of Visualization

Where is more work required?

What emerging technologies and societal trends will impact visualization design?

What did you find most difficult in creating visualizations and designing techniques?

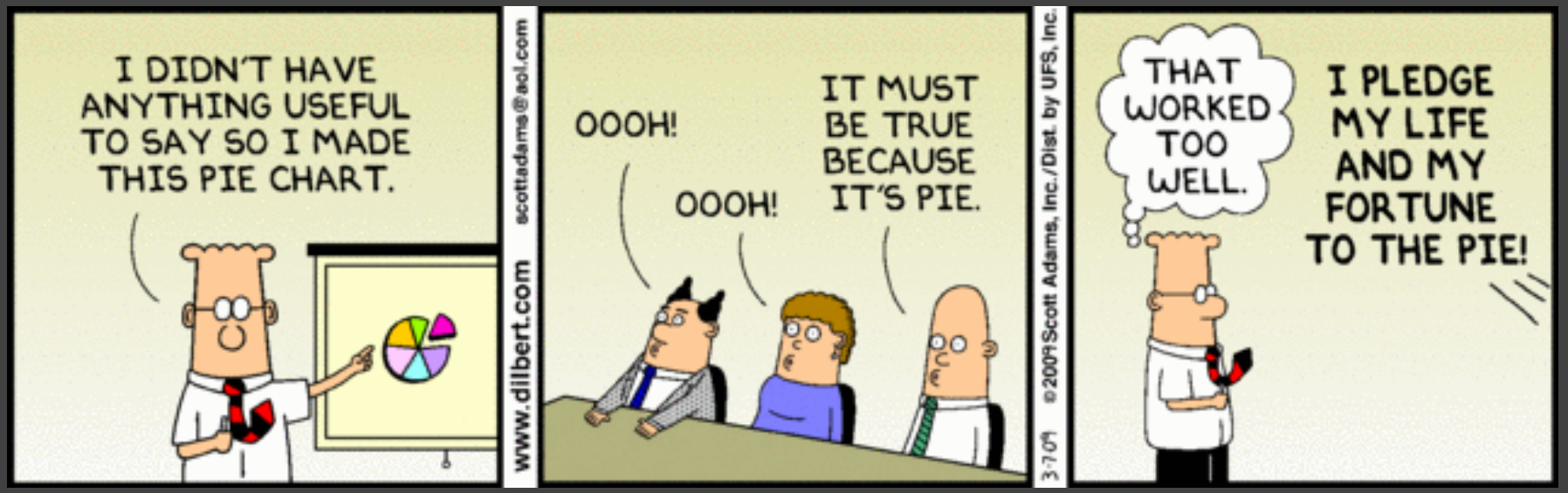
Course Evaluation

Course Evaluation

Official campus course evaluation
Distributed and completed online.
Your opinion is valued!
https://uw.iasystem.org/survey/161042

Thanks for a great quarter!

Thank You!


I DIDN'T HAVE ANYTHING USEFUL TO SAY SO I MADE THIS PIE CHART.
www.dilbert.com scottadams@aol.com
OOOH!
OOOH!
IT MUST BE TRUE BECAUSE IT'S PIE.
3-7-09 ©2009 Scott Adams, Inc./Dist. by UFS, Inc.
THAT WORKED TOO WELL.
I PLEDGE MY LIFE AND MY FORTUNE TO THE PIE!