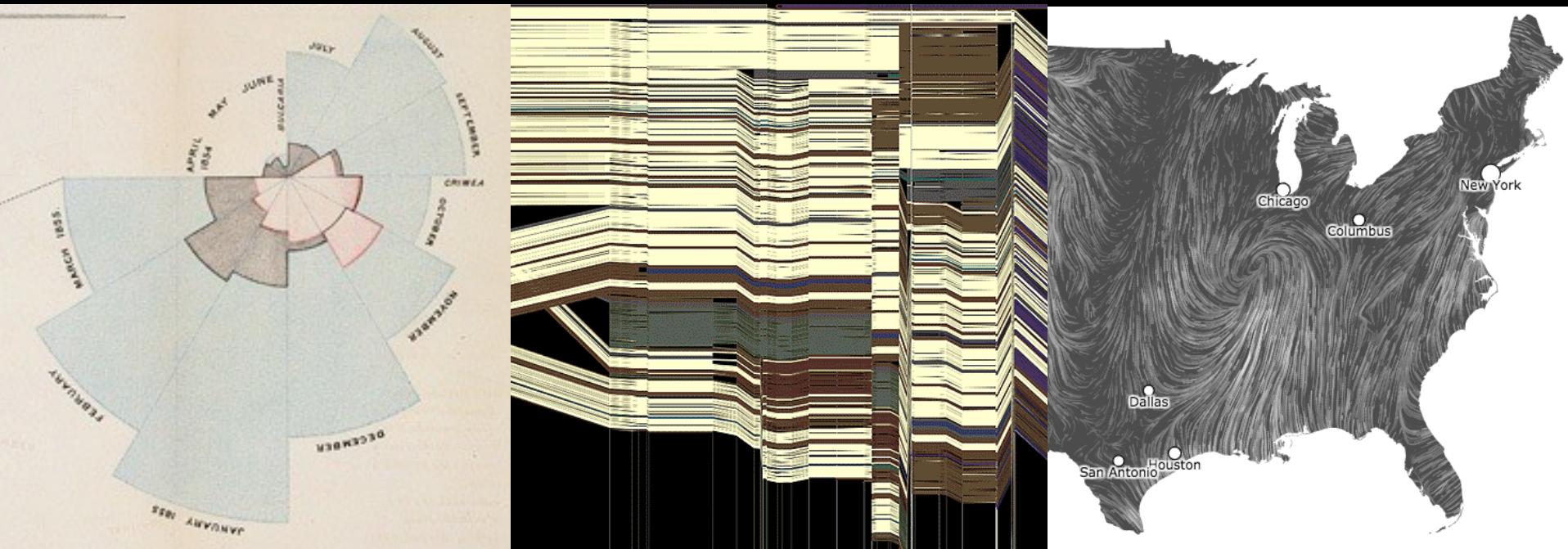


# CSE 442 - Data Visualization

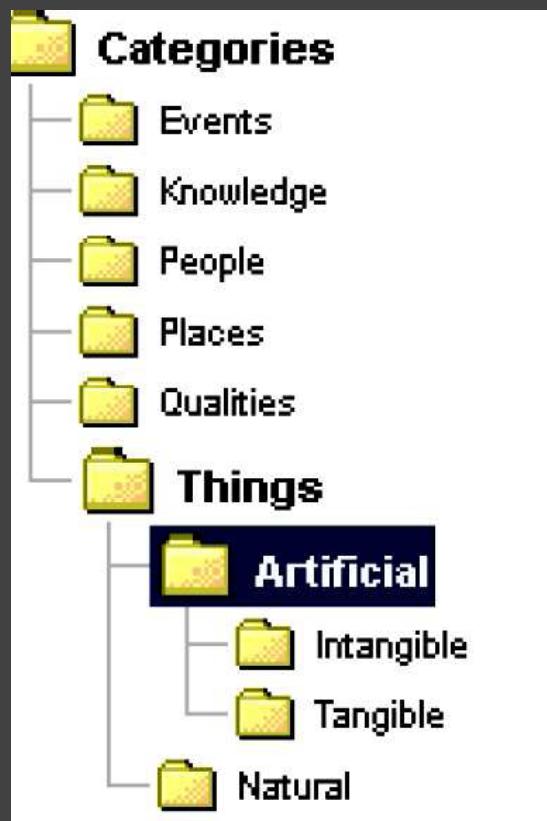
# Evaluation



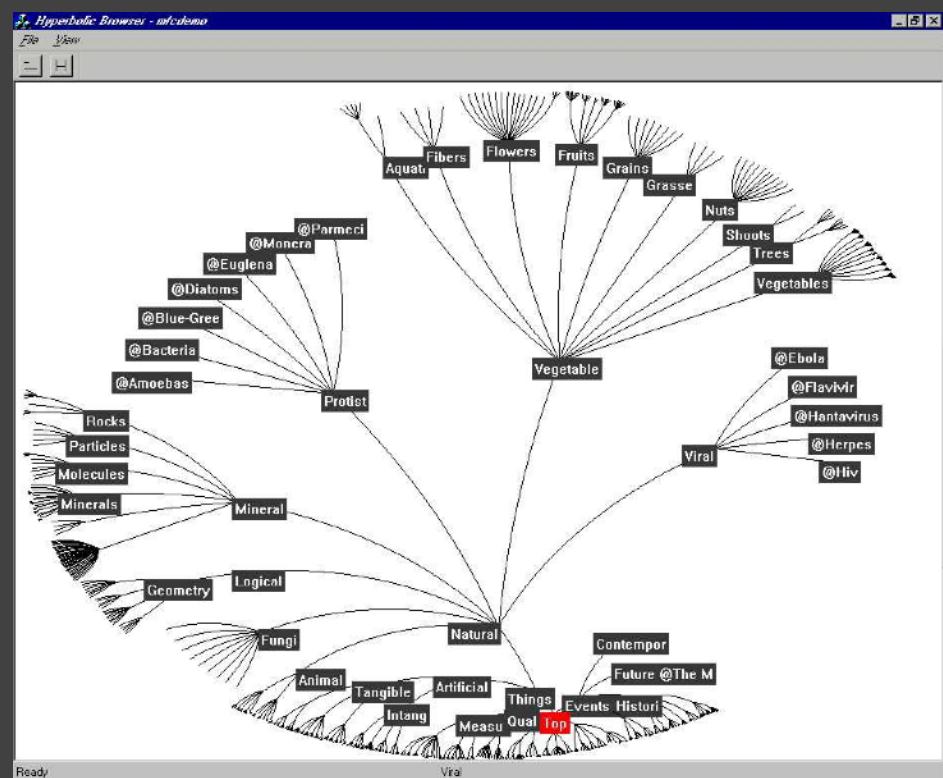
Jeffrey Heer University of Washington

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 (Trees, Spatial Navigation)

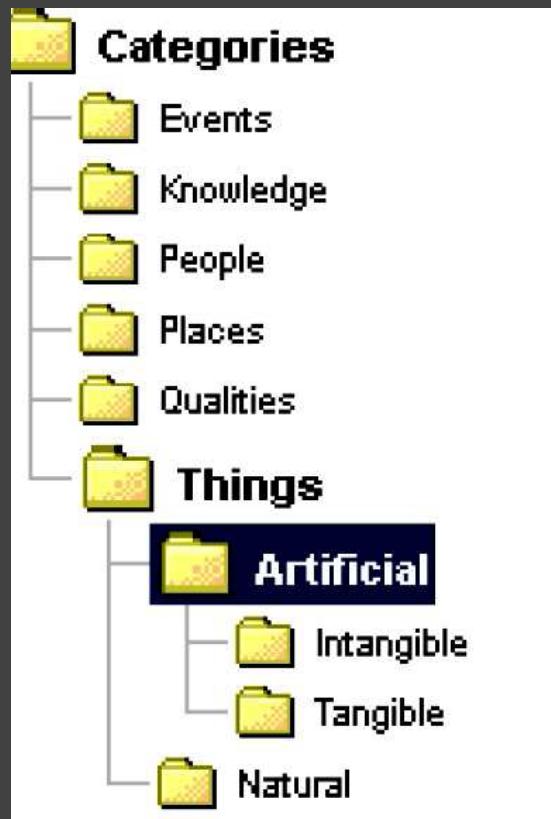
Data Density of Time Series

Perceptual Organization of Graphs

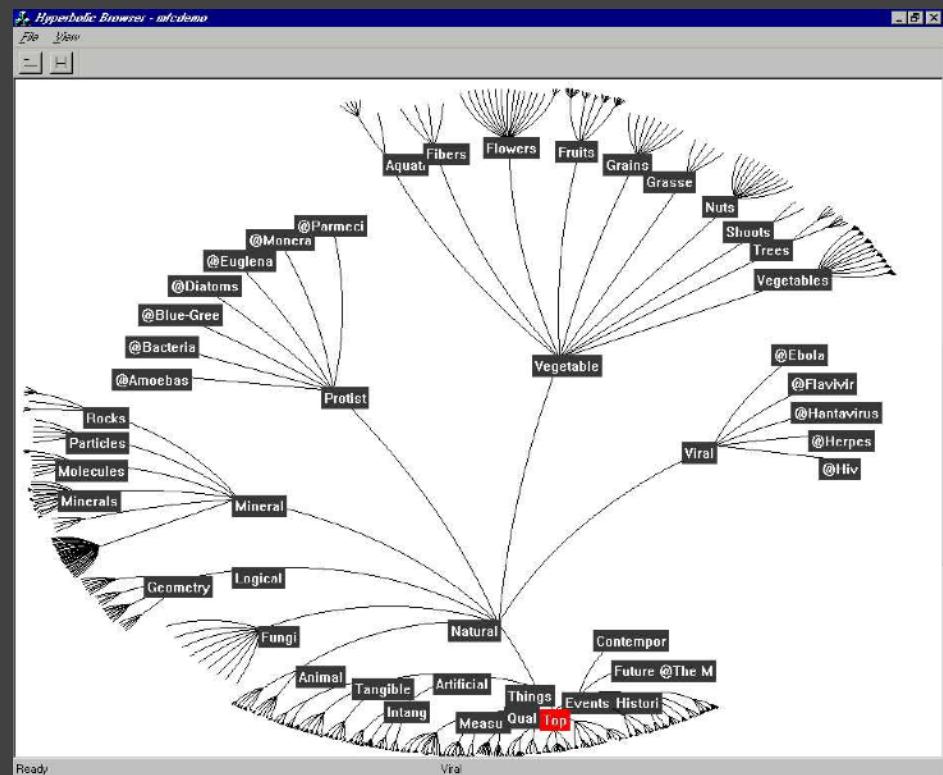
Discussion and Course Evaluation

# Trees

# The Great Browse-Off! [CHI 97]



VS.



Microsoft File Explorer

Xerox PARC Hyperbolic Tree

# Which visualization is better?

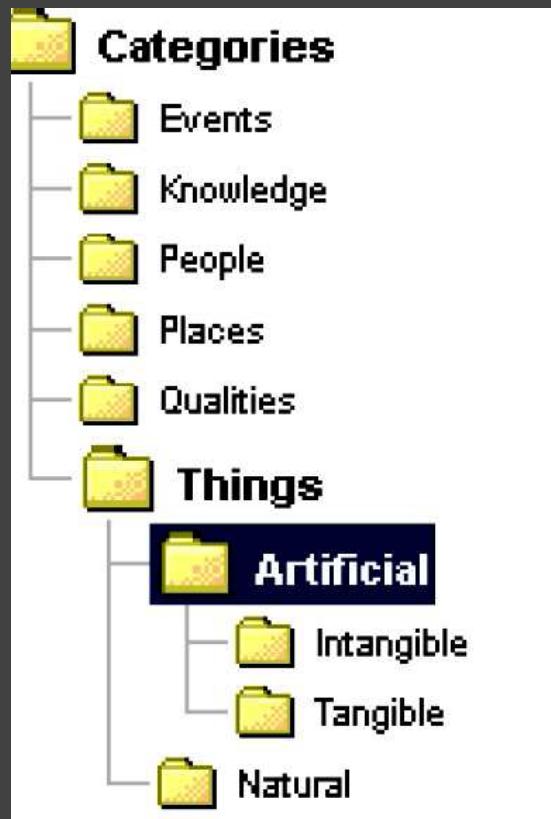
Xerox PARC researchers ran eye-tracking studies to investigate... [Pirolli et al 00]

Subjects performed both retrieval and comparison tasks of varying complexity.

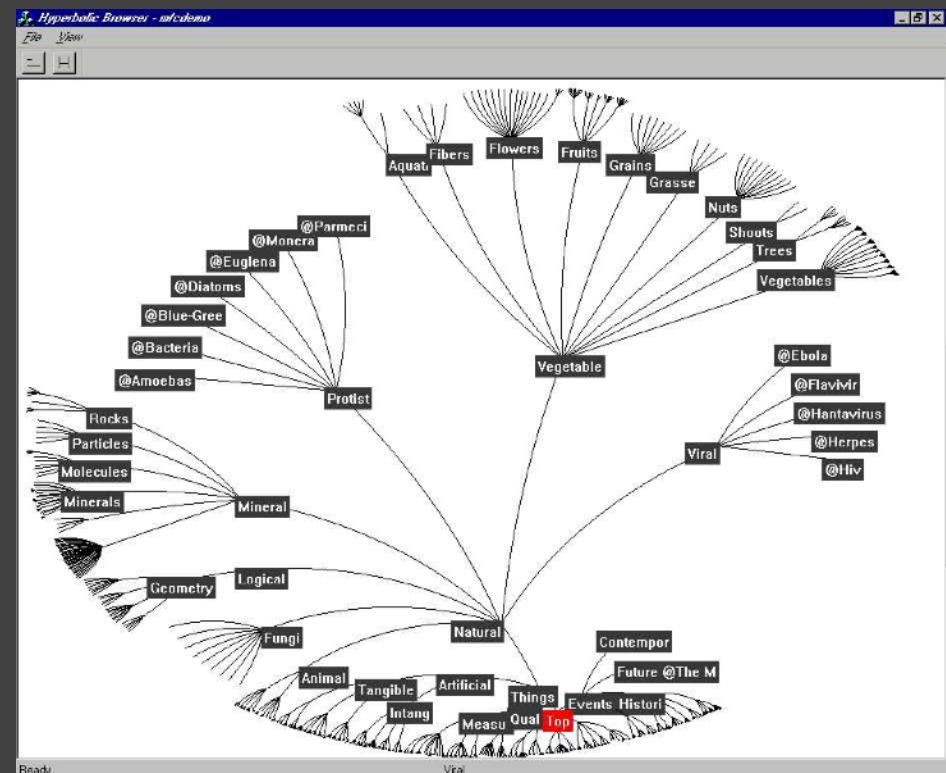
**No significant performance differences** were found across task conditions.

# How do users navigate the tree?

They read the labels!



VS.



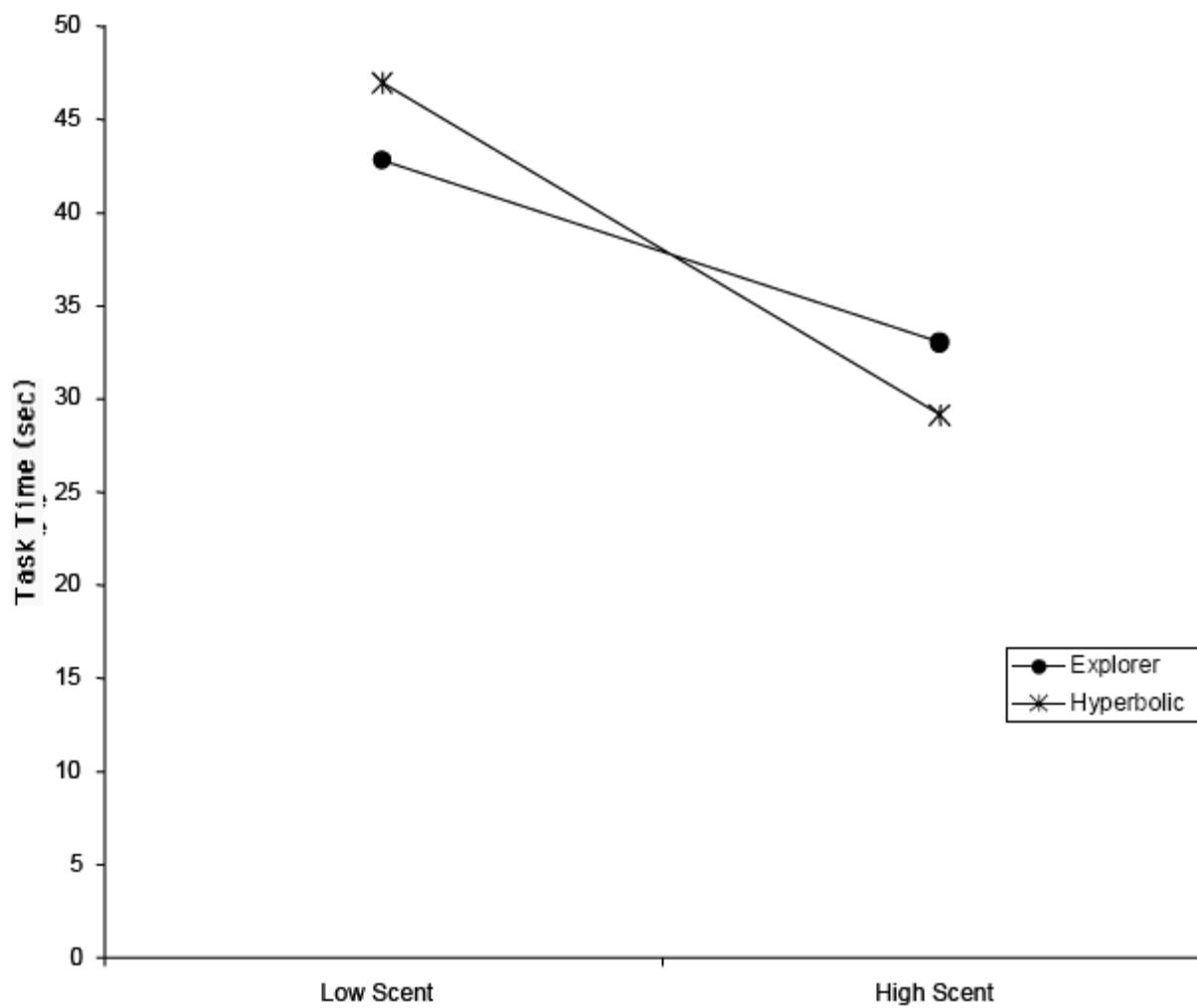
Microsoft File Explorer

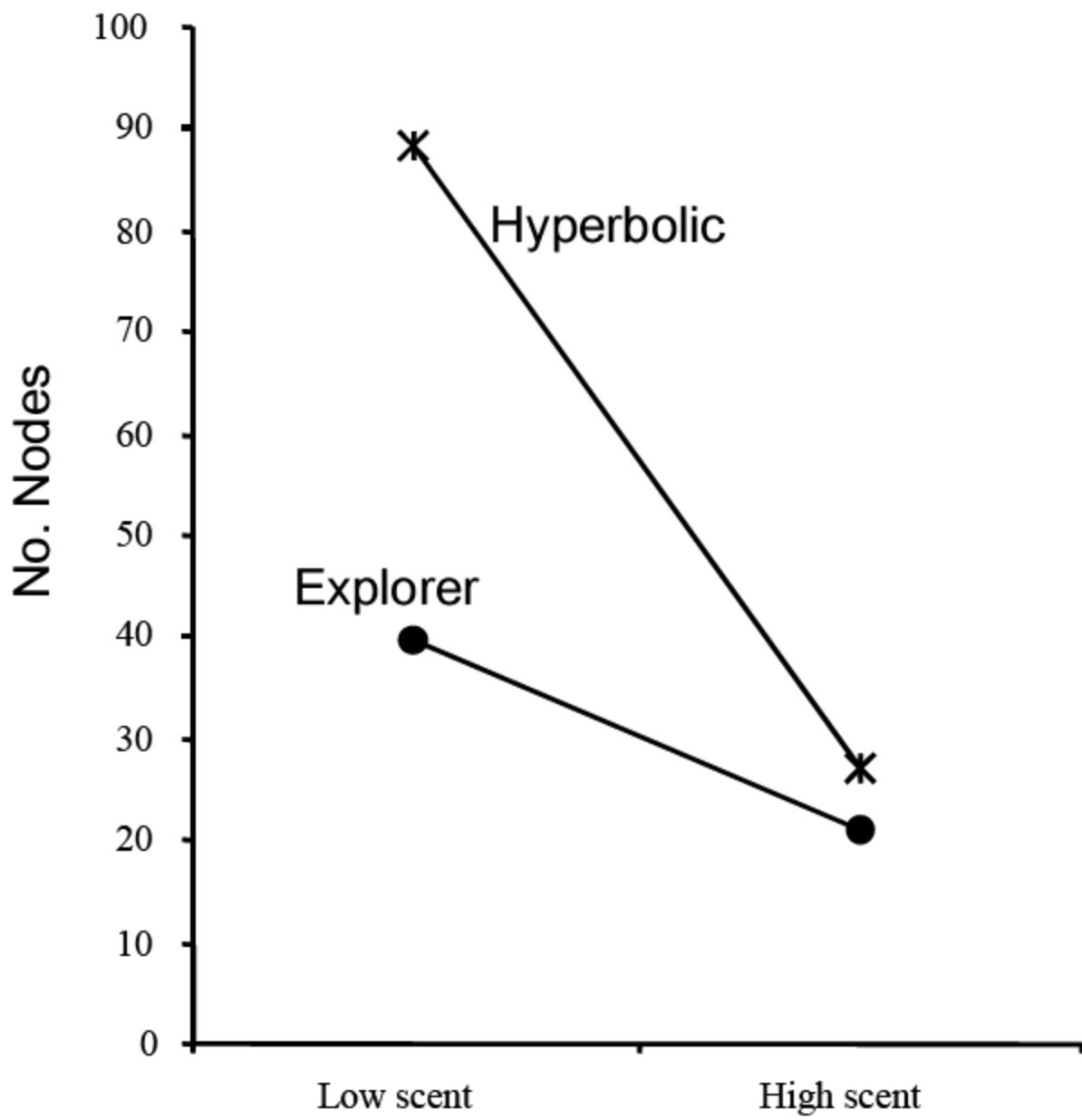
Xerox PARC Hyperbolic Tree

# How do users navigate the tree?

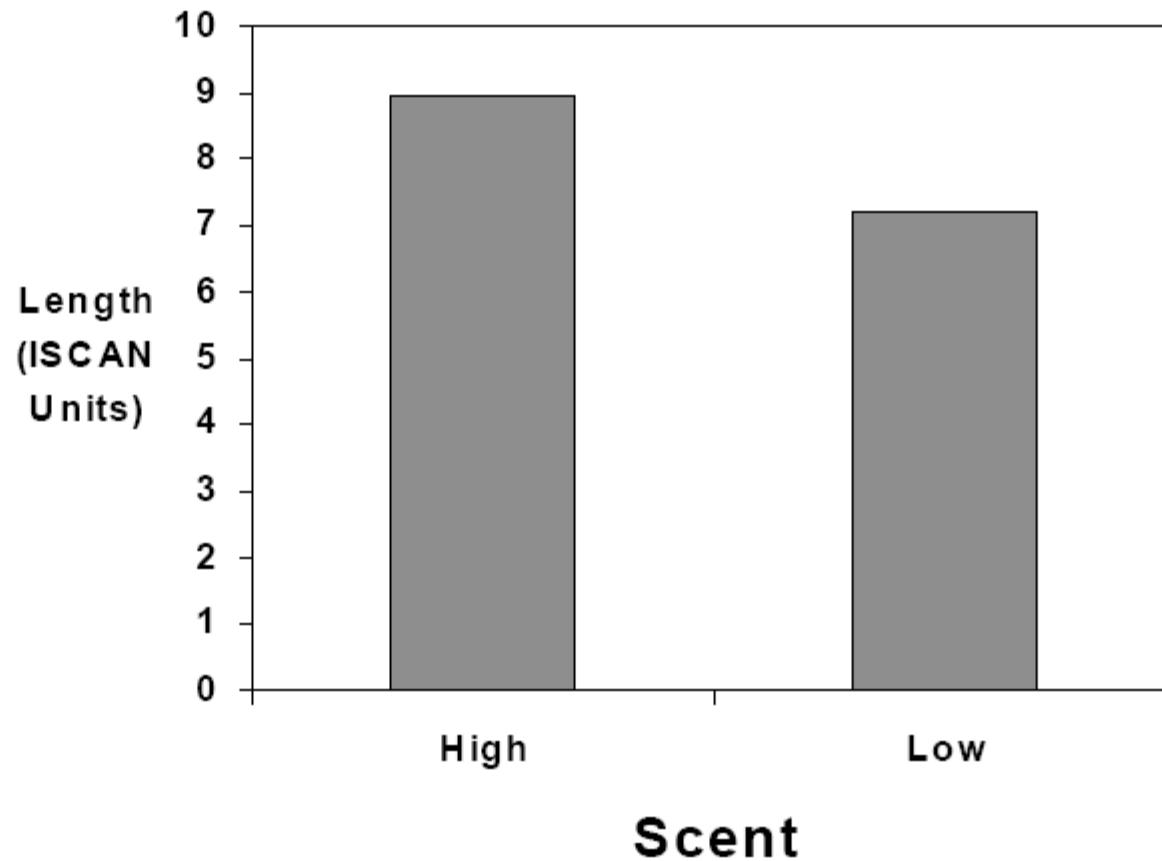
**Information Scent:** A user's (imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues. [Pirolli & Card 99]

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

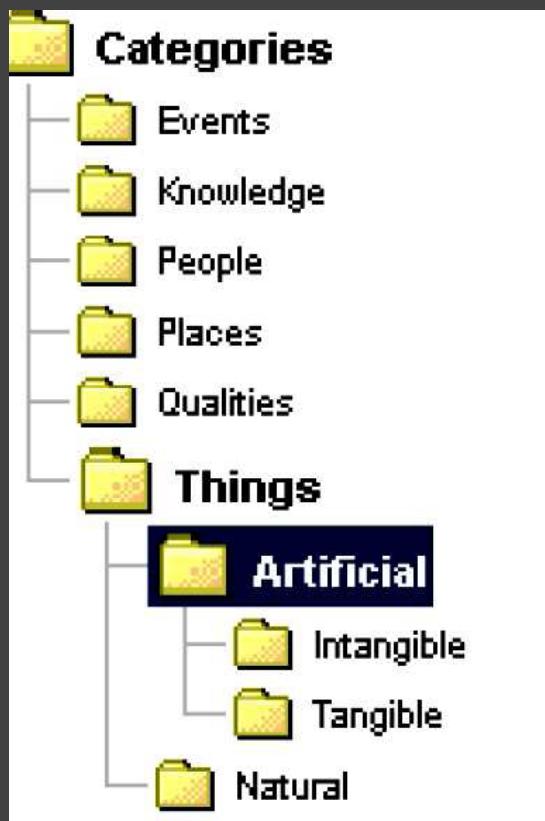




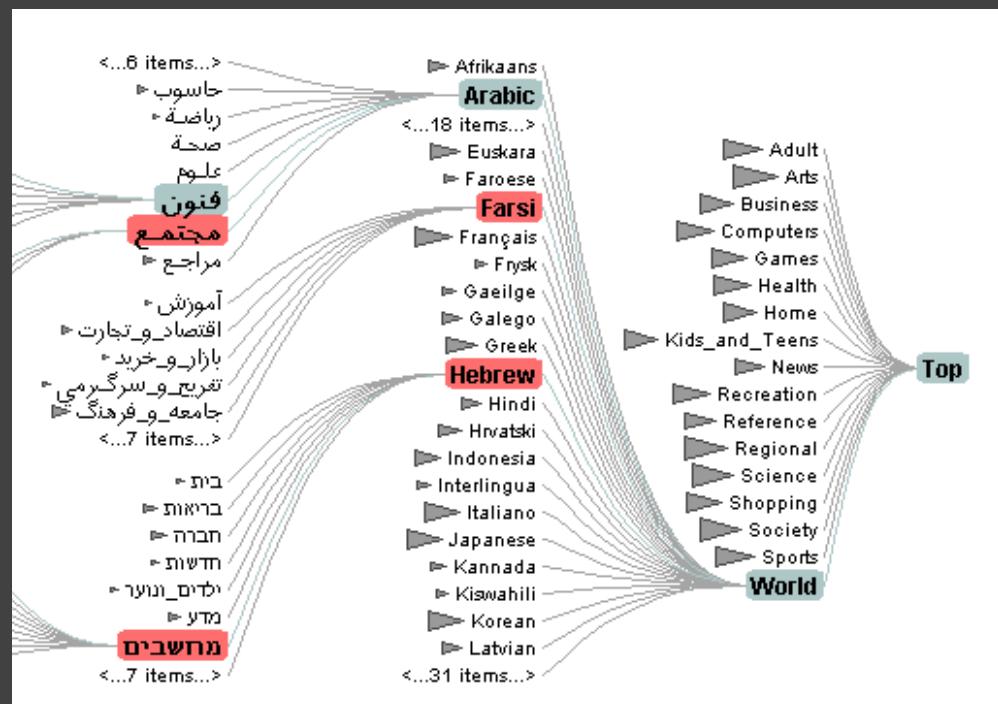
# Length of Eye Movements



# More Evaluations



VS.



# 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

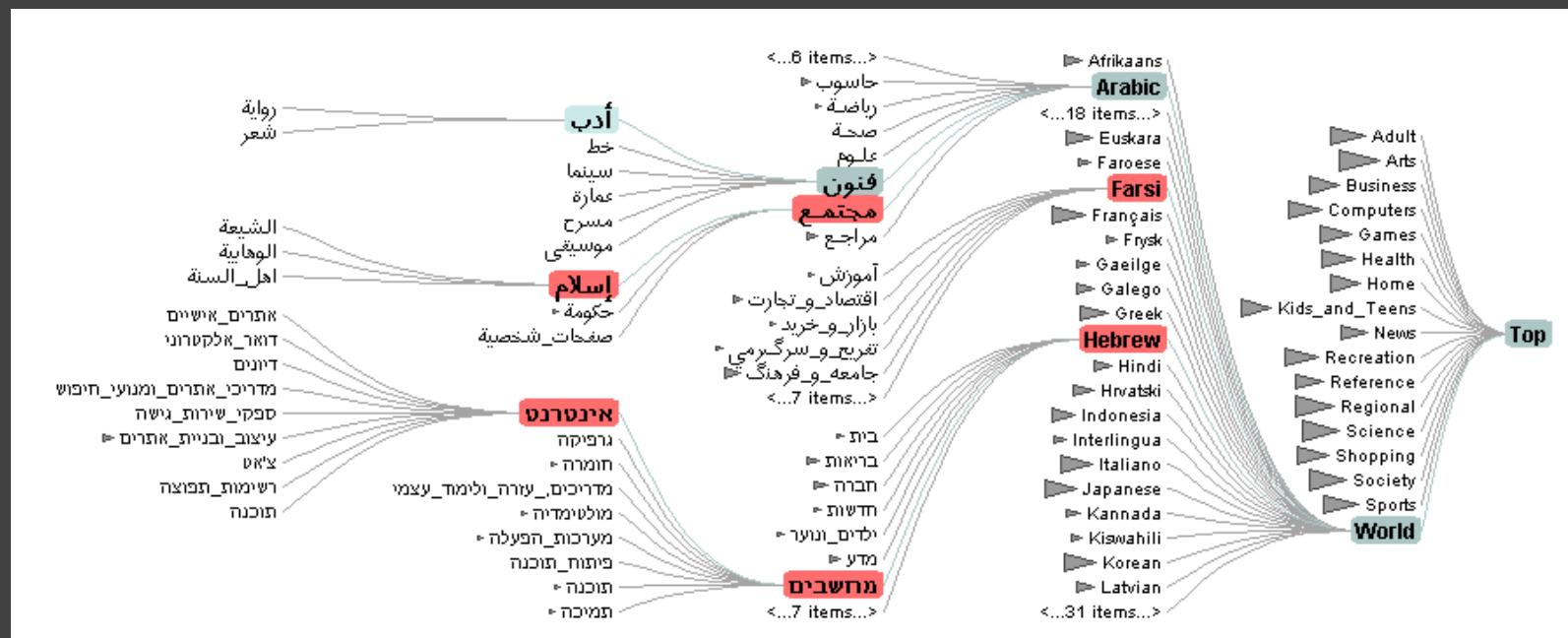
## DOI Tree vs. Google Directory [Pirolli, CHI 06]

DOI Tree has superior task knowledge transfer

# Design Guidelines

# 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.

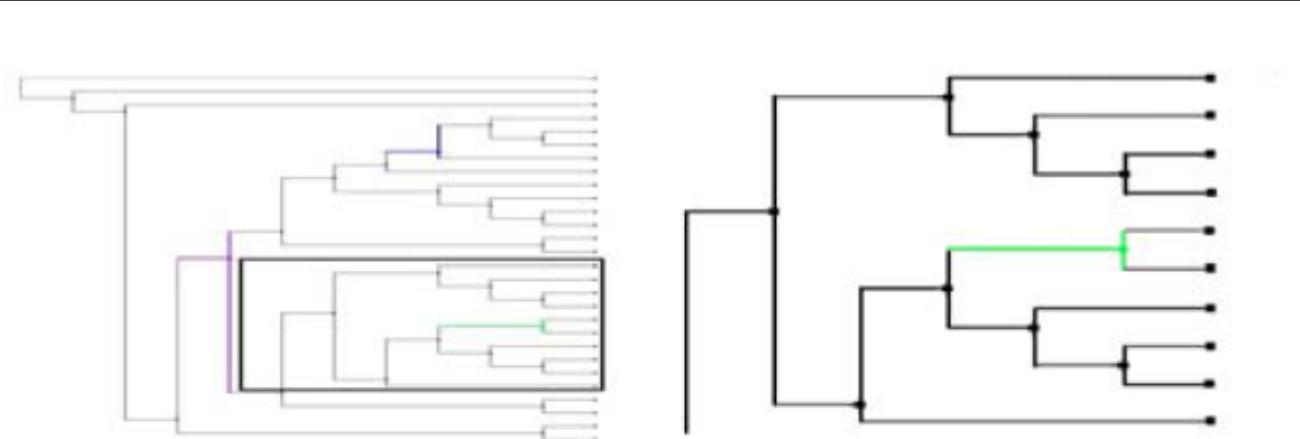
Equal **performance** in terms of accuracy or response time is **not the whole picture**. We often require more detailed study!

# Spatial Navigation

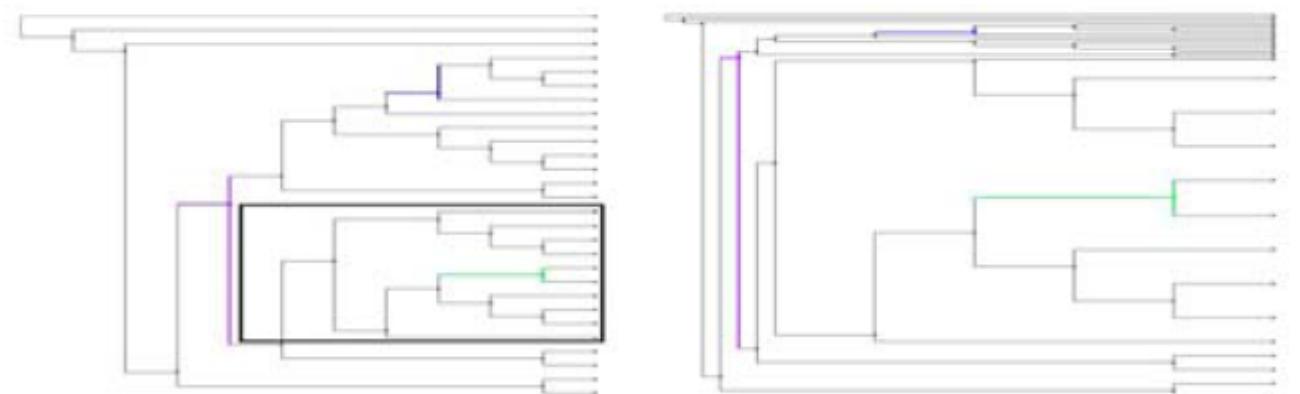
# **An Evaluation of Pan & Zoom and Rubber Sheet Navigation with and without an Overview**

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

# Pan & Zoom vs. Rubber Sheet



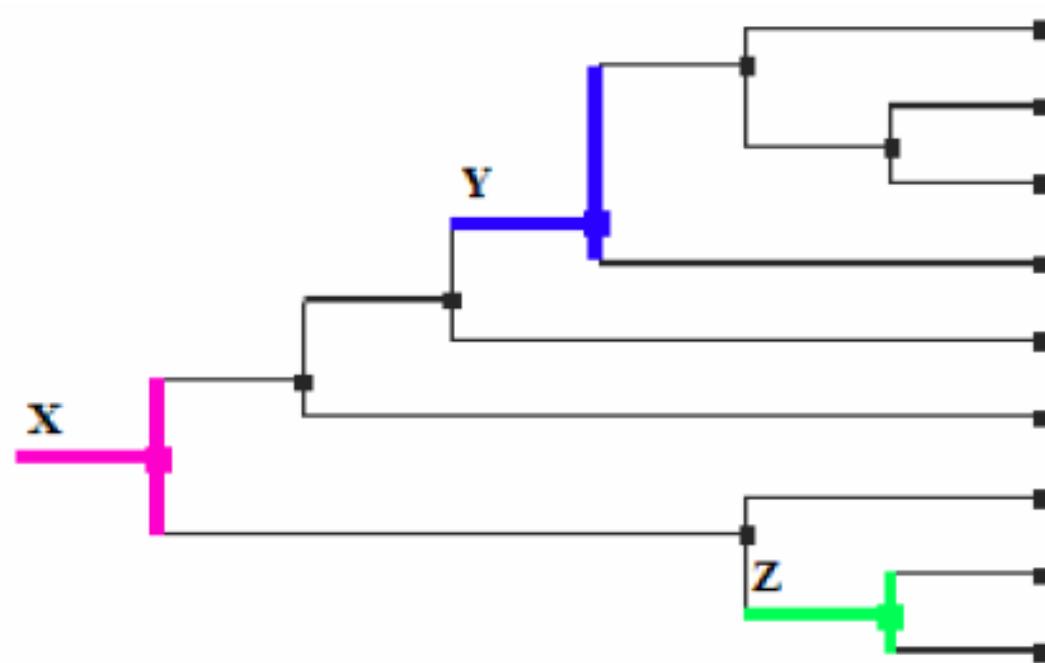
(i) PZN



(ii) RSN

# Experimental Task

Compare topological distance between nodes in a dendrogram.



# Experiment

Compare performance in 4 conditions:

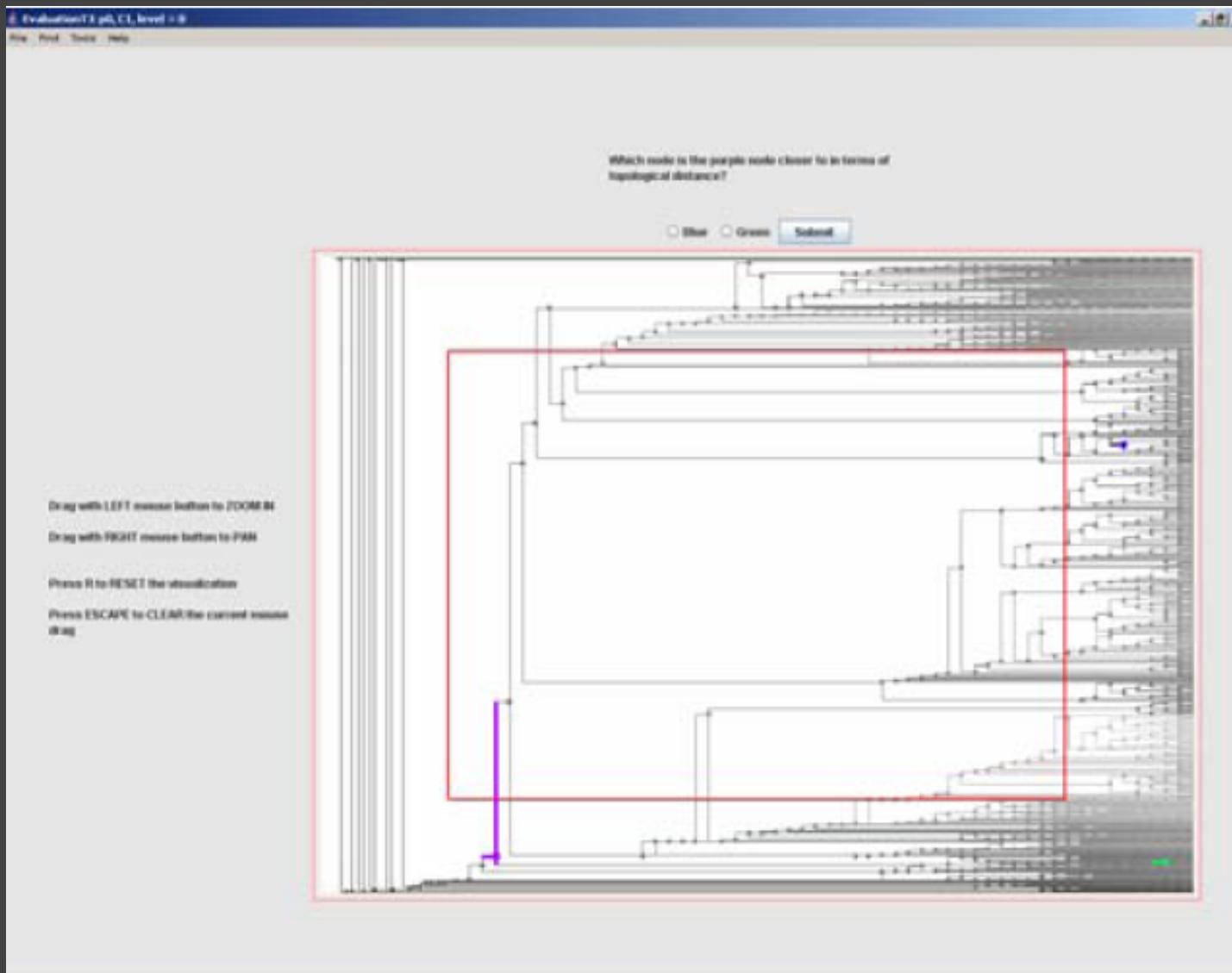
1. Pan & Zoom (no overview)
2. Pan & Zoom (with overview)
3. Rubber Sheet (no overview)
4. Rubber Sheet (with overview)

40 subjects (24F/16M), between 18-39 years old.

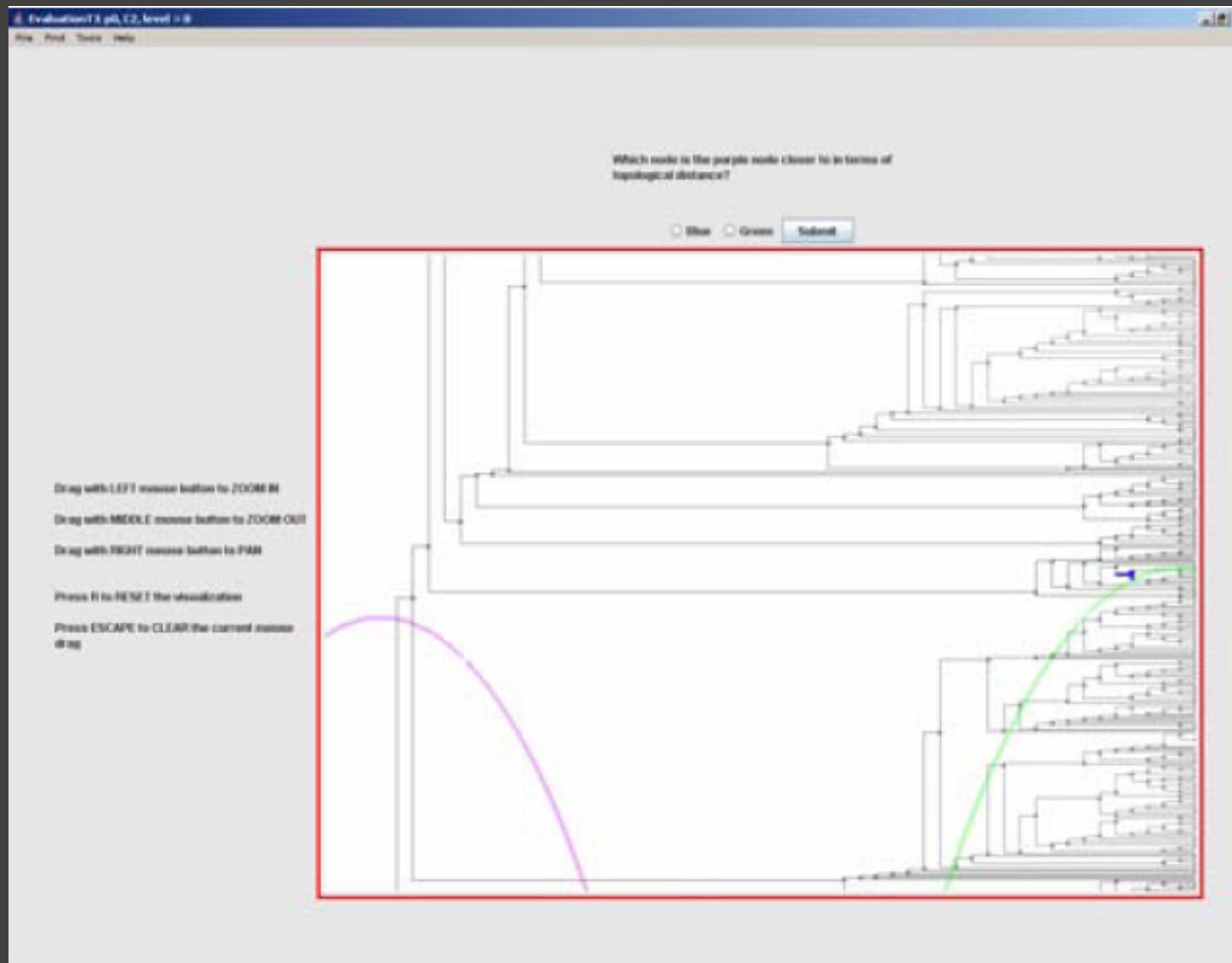
Right-handed, normal vision.

Between-subjects design.

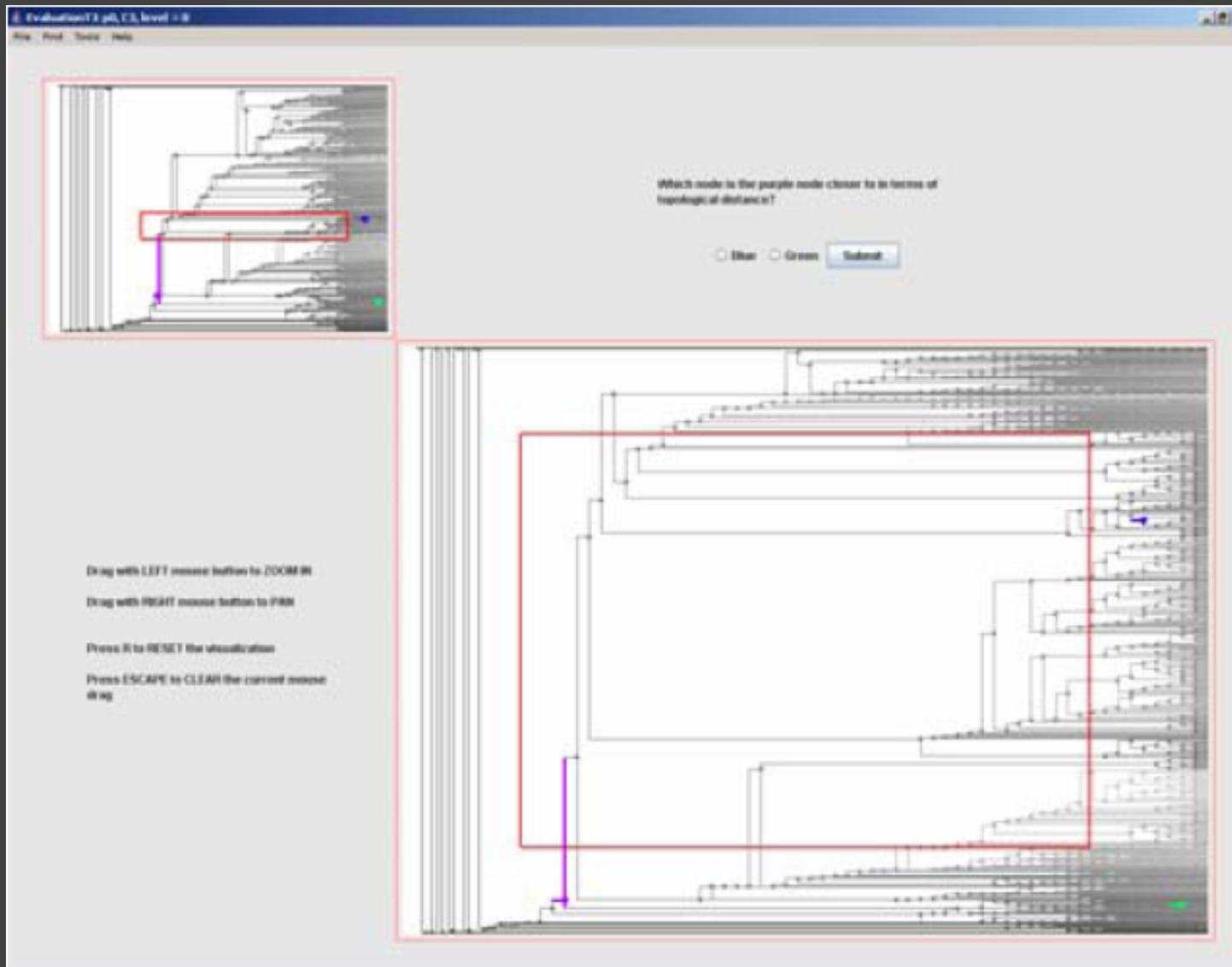
# 1. Rubber Sheet / No Overview



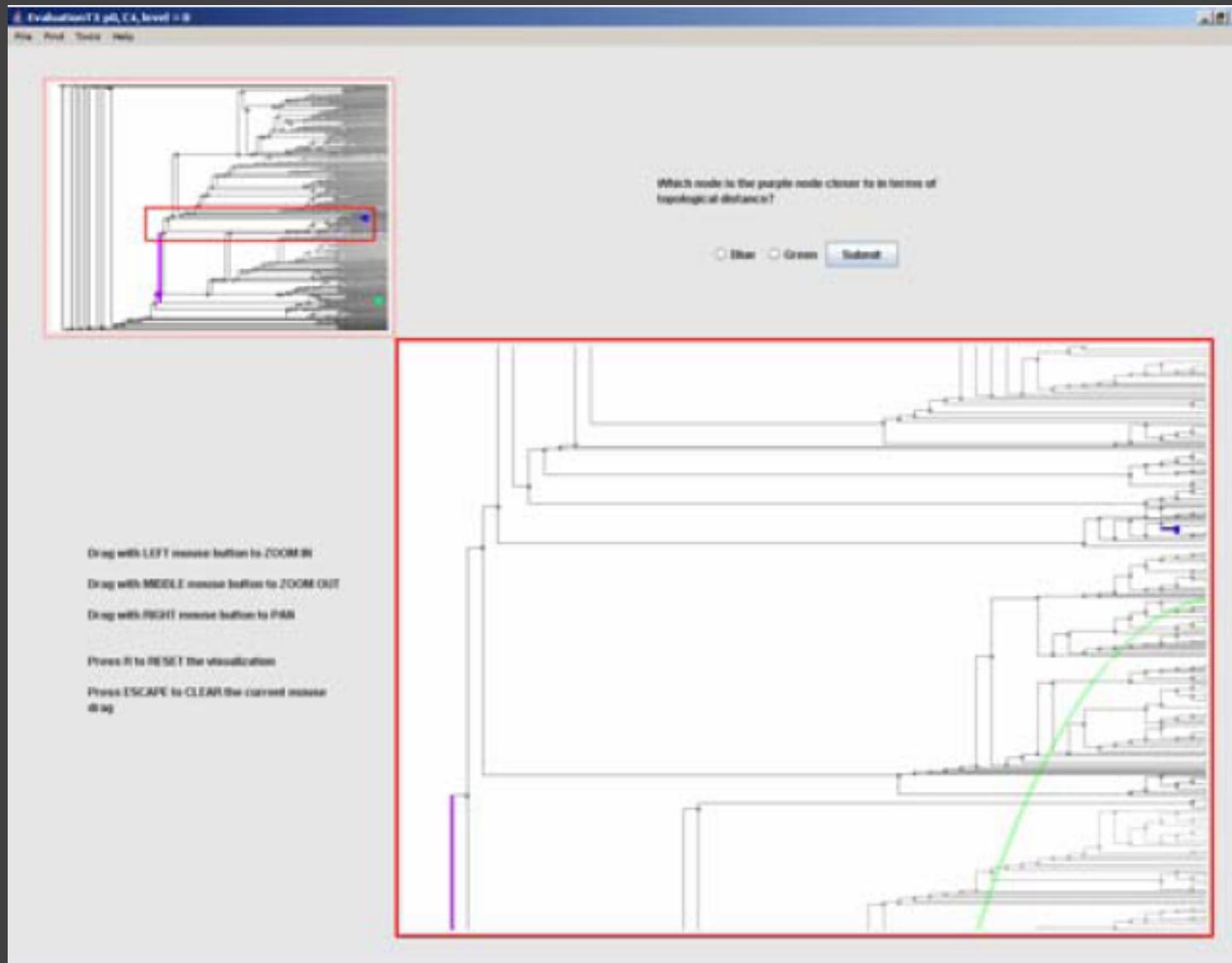
## 2. Pan & Zoom / No Overview



# 3. Rubber Sheet / Overview



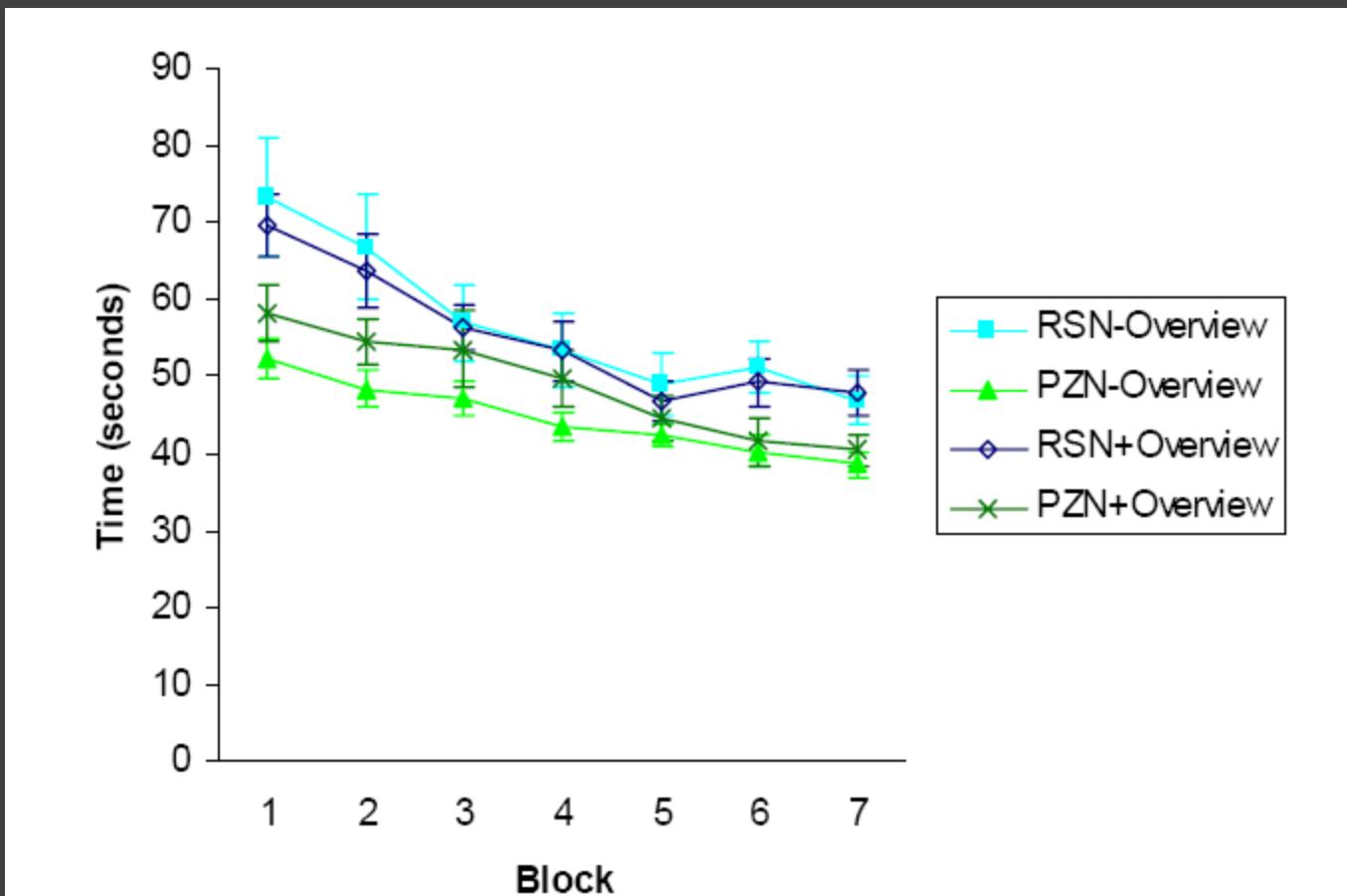
# 4. Pan & Zoom / Overview



# Hypotheses

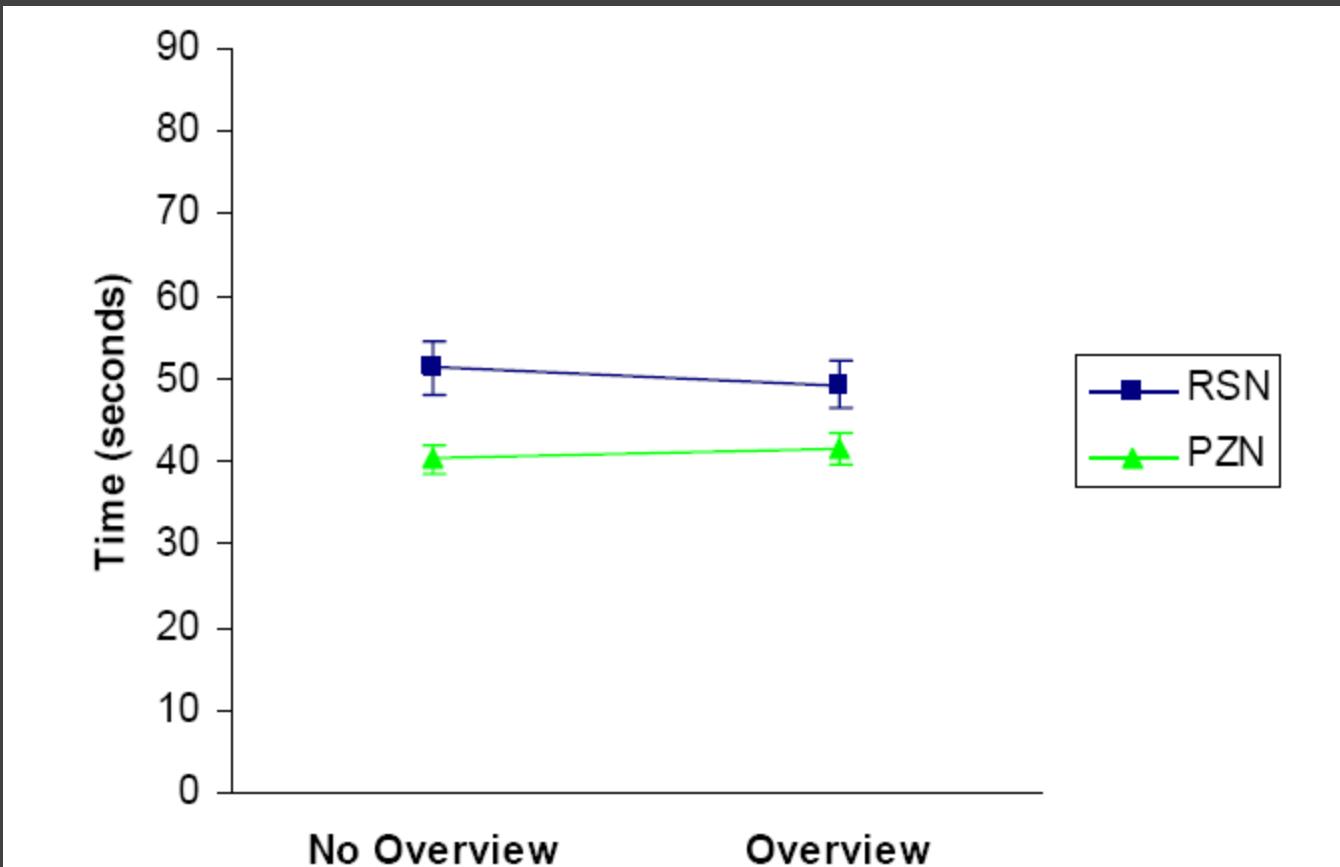
1. RSN interfaces perform better than PZN interfaces independently of the presence or absence of an overview.
2. For RSN, the presence of an overview does not result in better performance.
3. For PZN, the presence of an overview results in better performance.

# 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.**

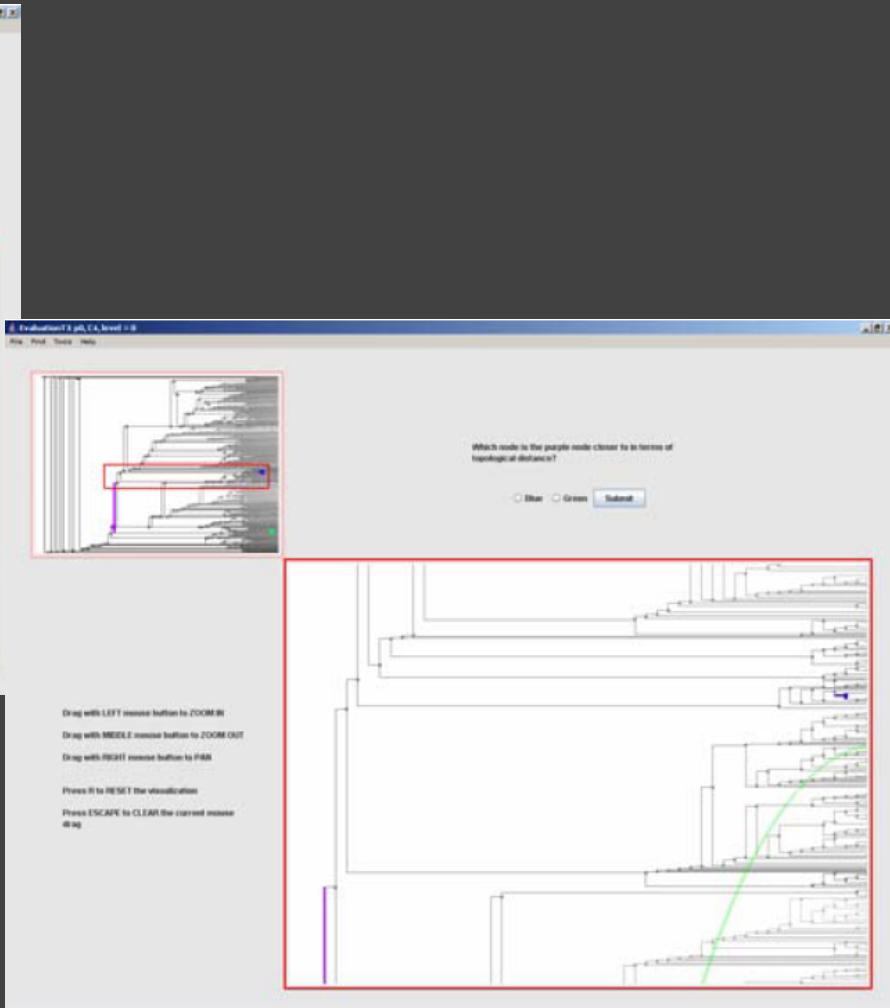
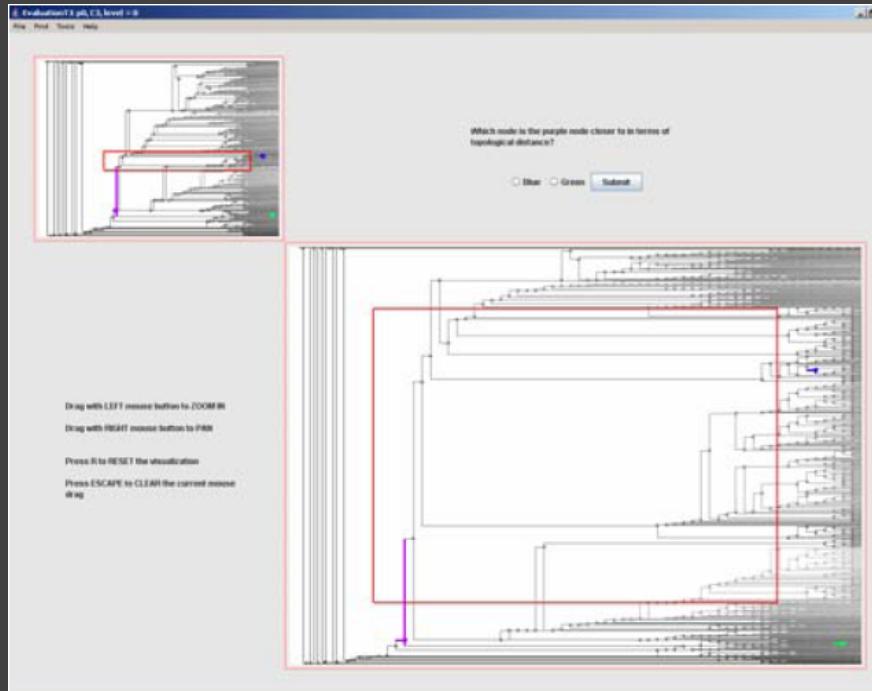
# Results

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.

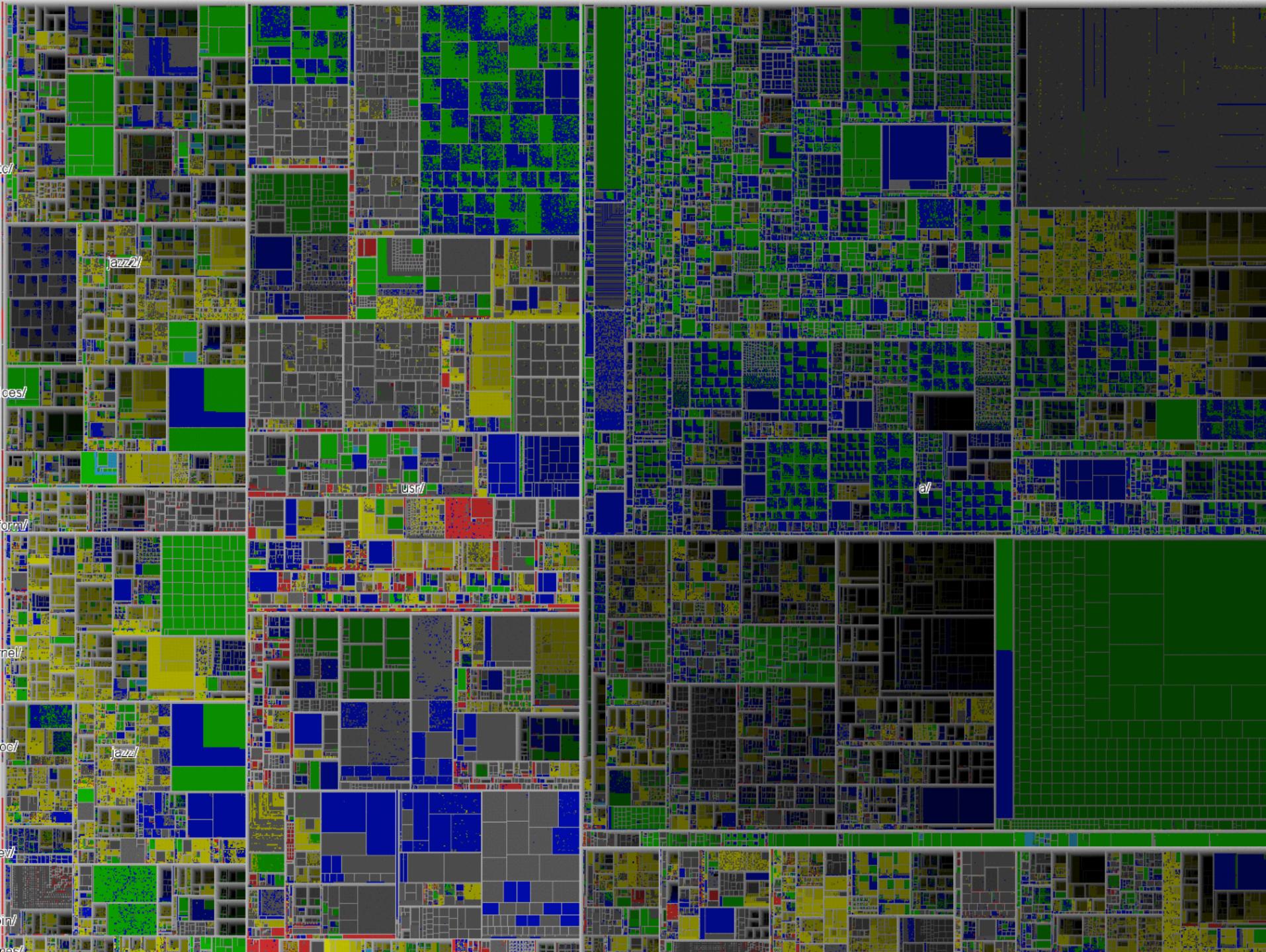
R3. Overview has no significant impact on pan & zoom navigation, though it was reported to reduce physical demand.

# Thoughts?



Does this generalize  
for overview displays?

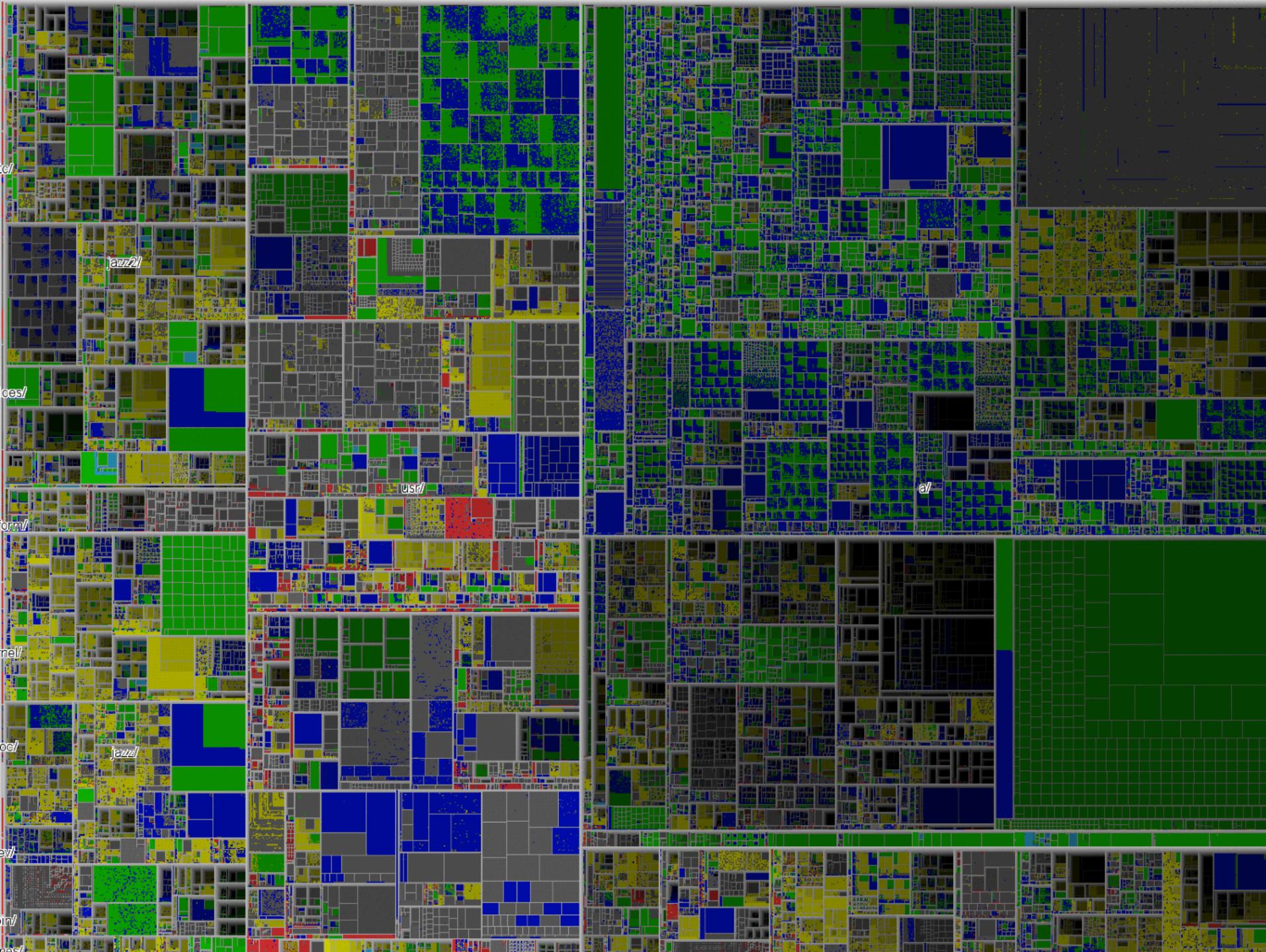
# Data Density

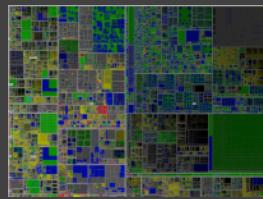


$$\text{Data Density} = \frac{(\# \text{ entries in data})}{(\text{area of graphic})}$$

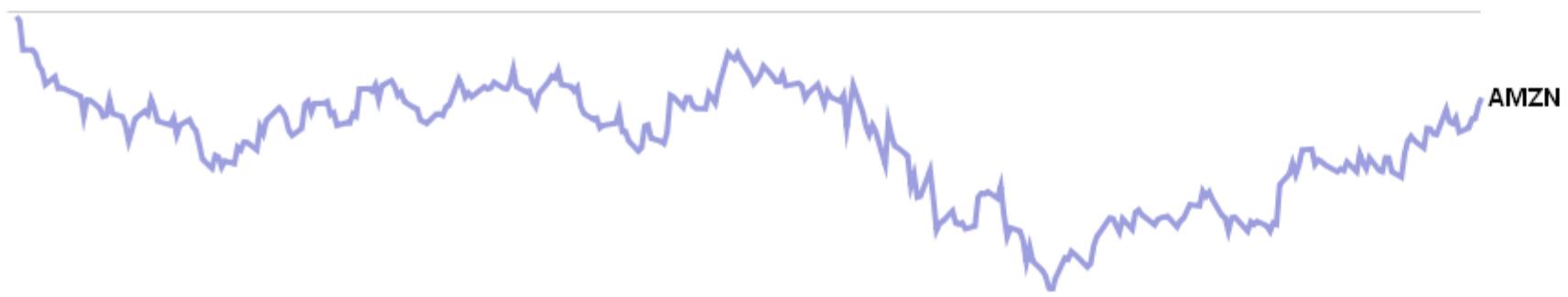
“Graphical excellence... gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space”

[Tufte 83]

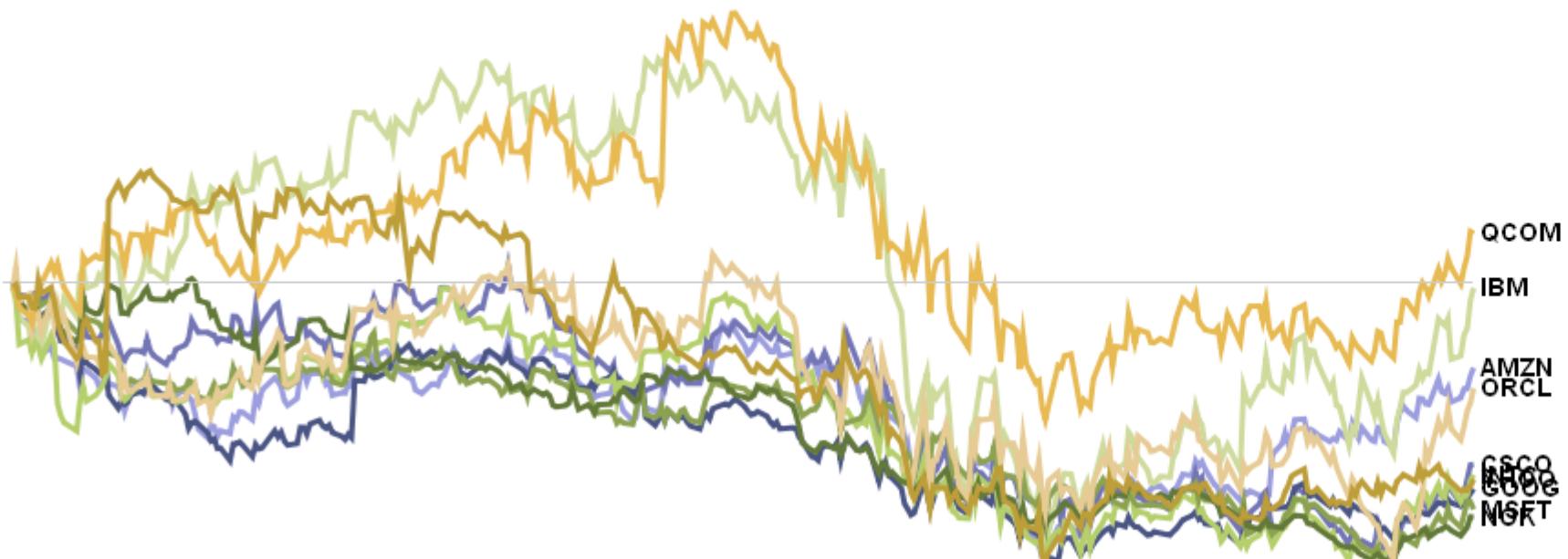




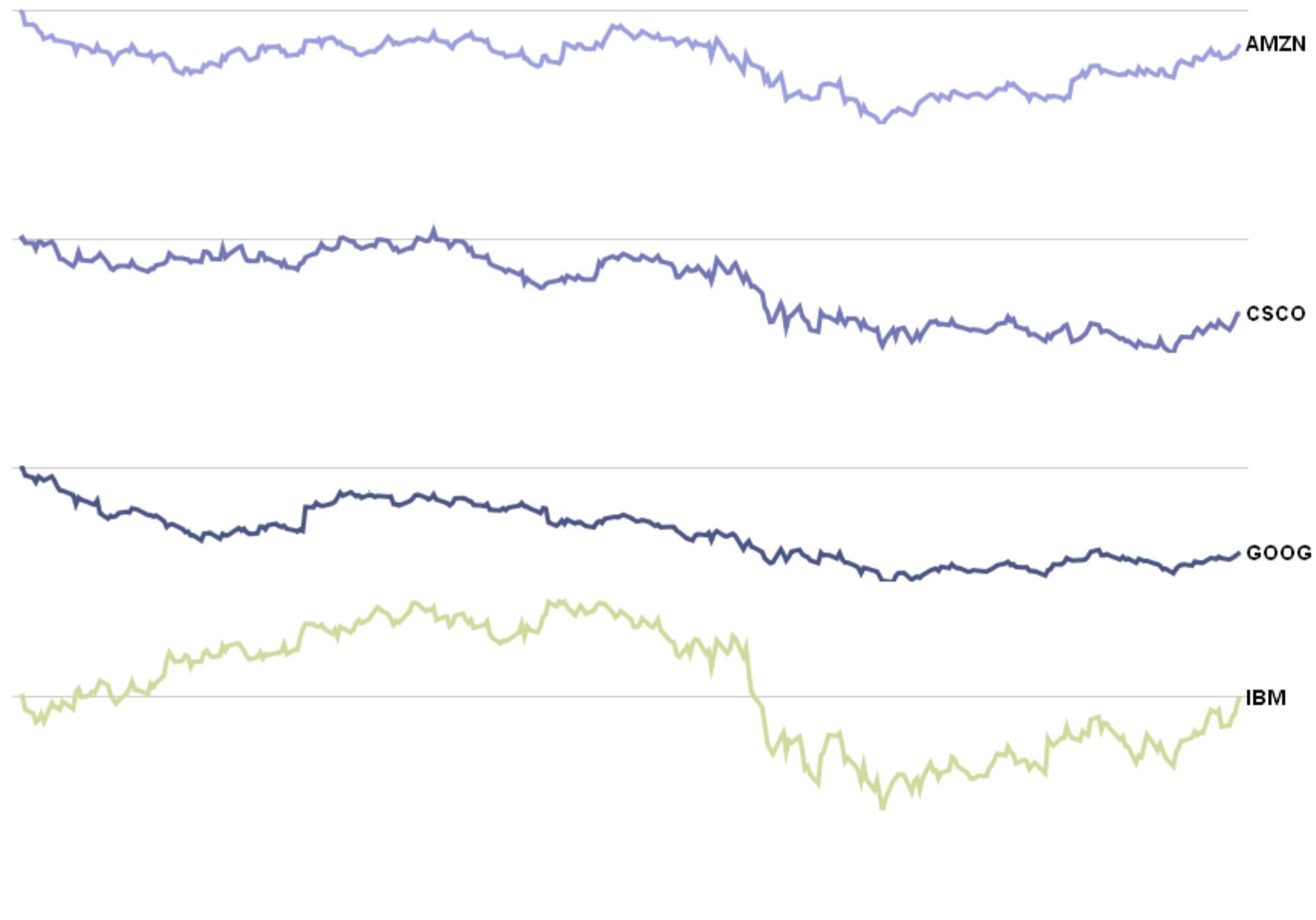
## Relative Technology Stock Performance: Jan 2008 - Present



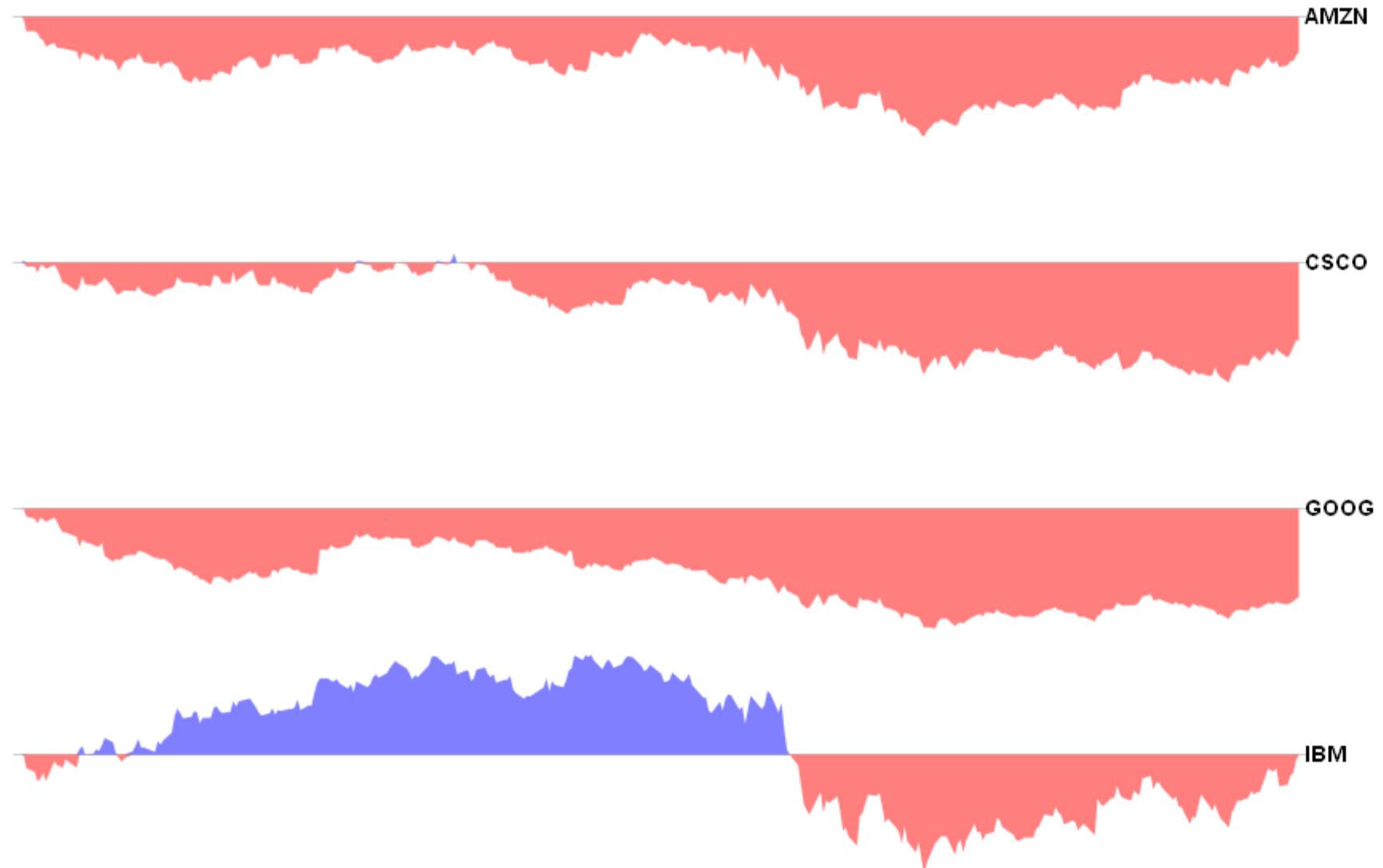
## Relative Technology Stock Performance: Jan 2008 - Present



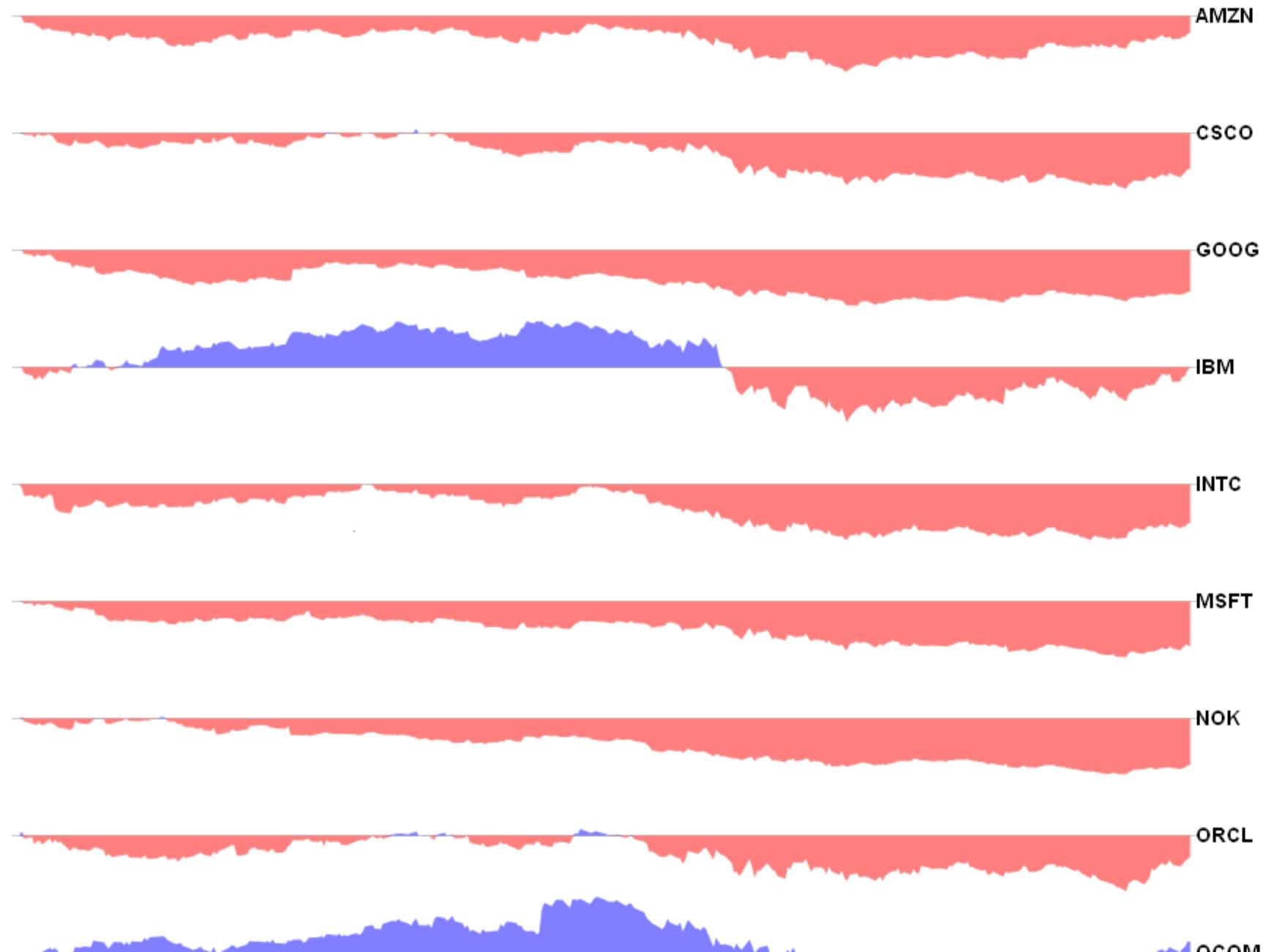
## Relative Technology Stock Performance: Jan 2008 - Present



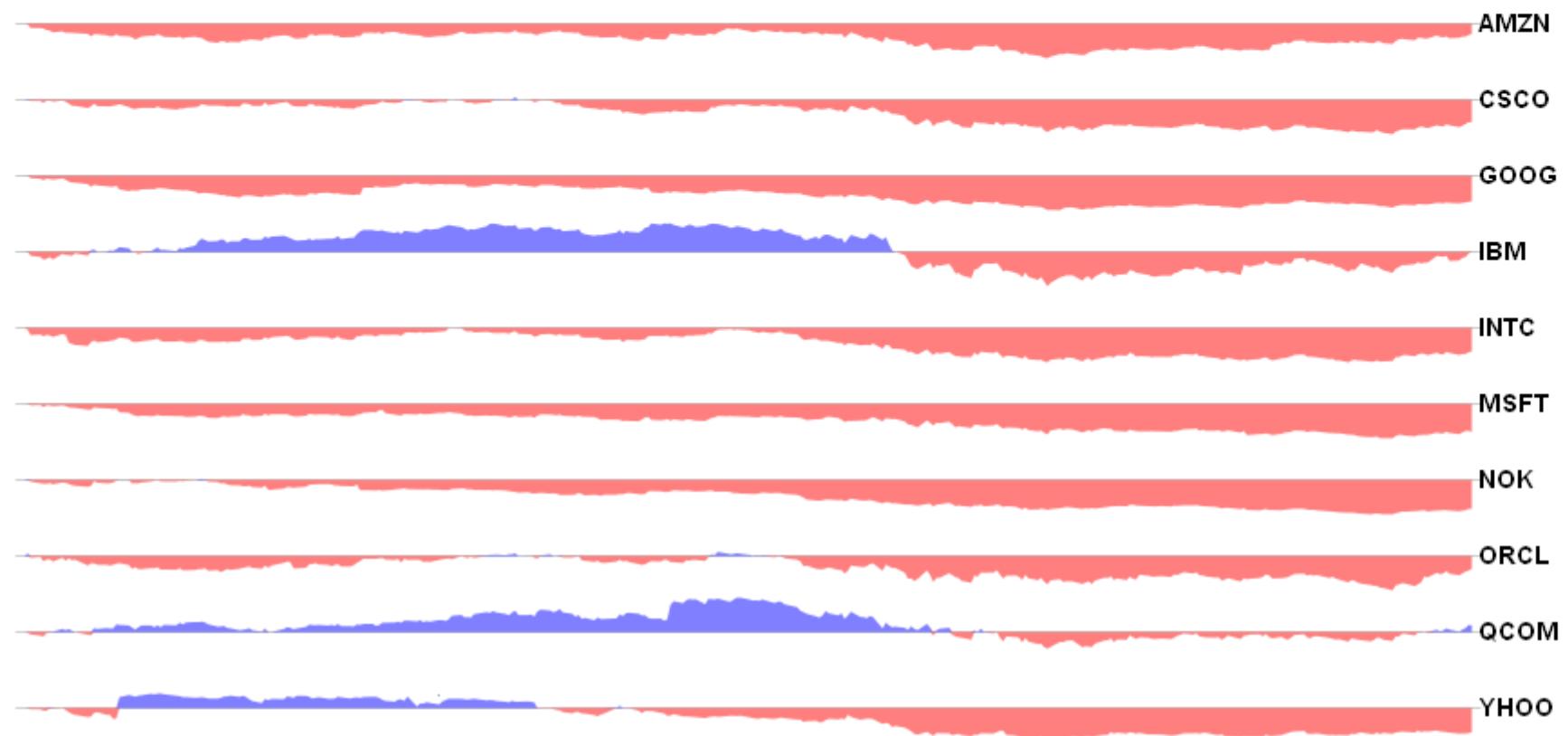
## Relative Technology Stock Performance: Jan 2008 - Present



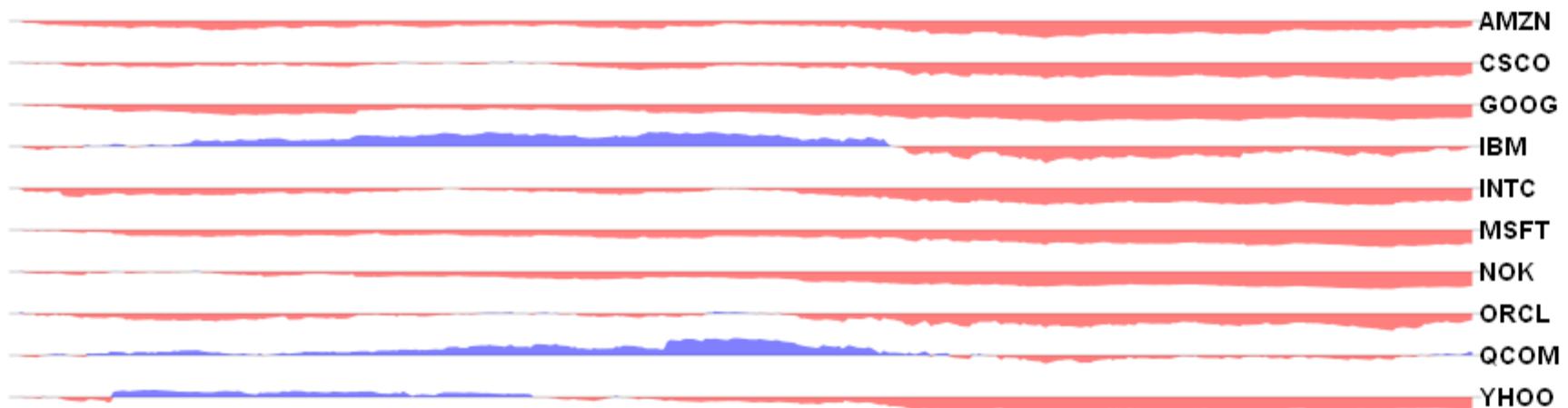
## Relative Technology Stock Performance: Jan 2008 - Present



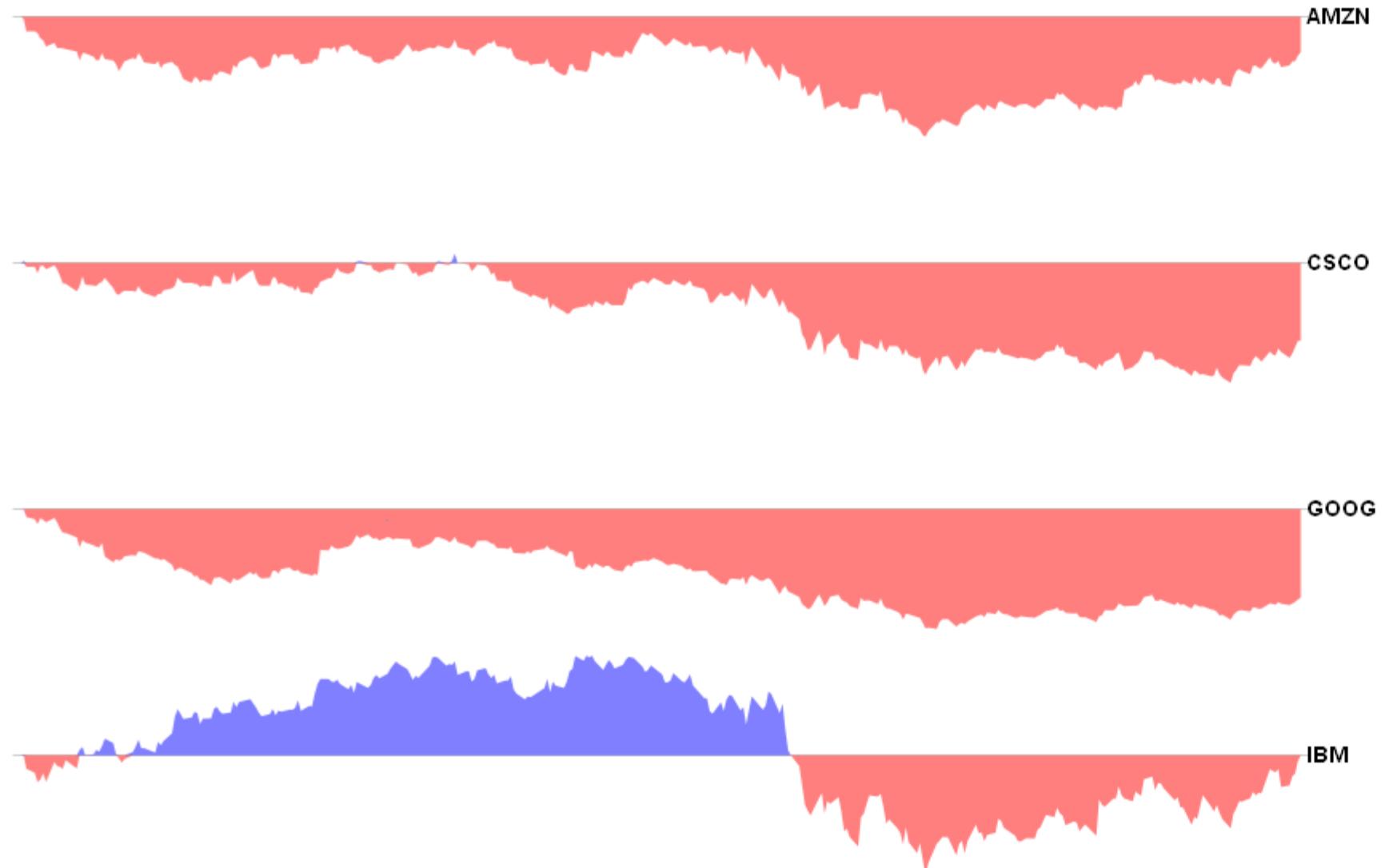
## Relative Technology Stock Performance: Jan 2008 - Present



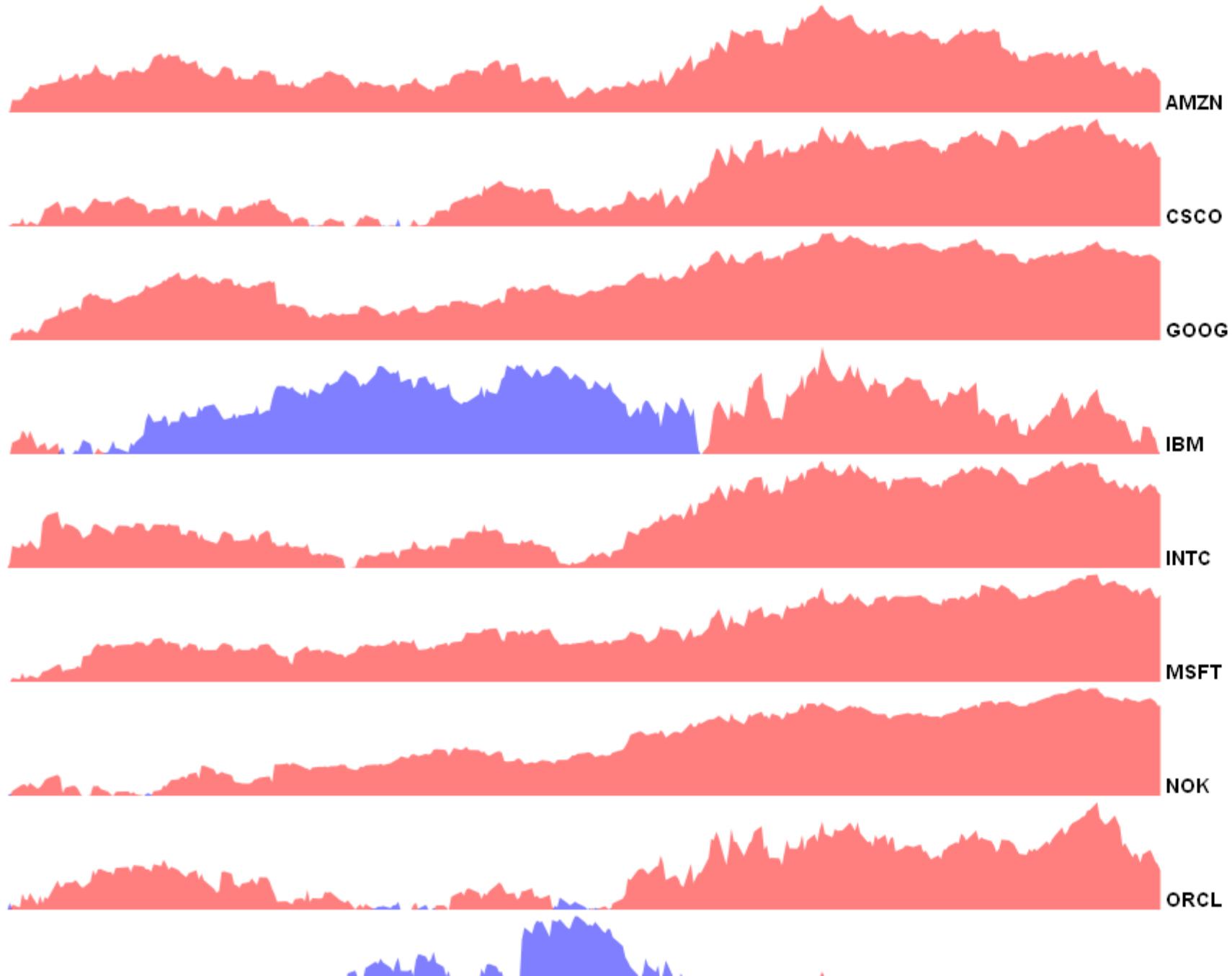
## Relative Technology Stock Performance: Jan 2008 - Present



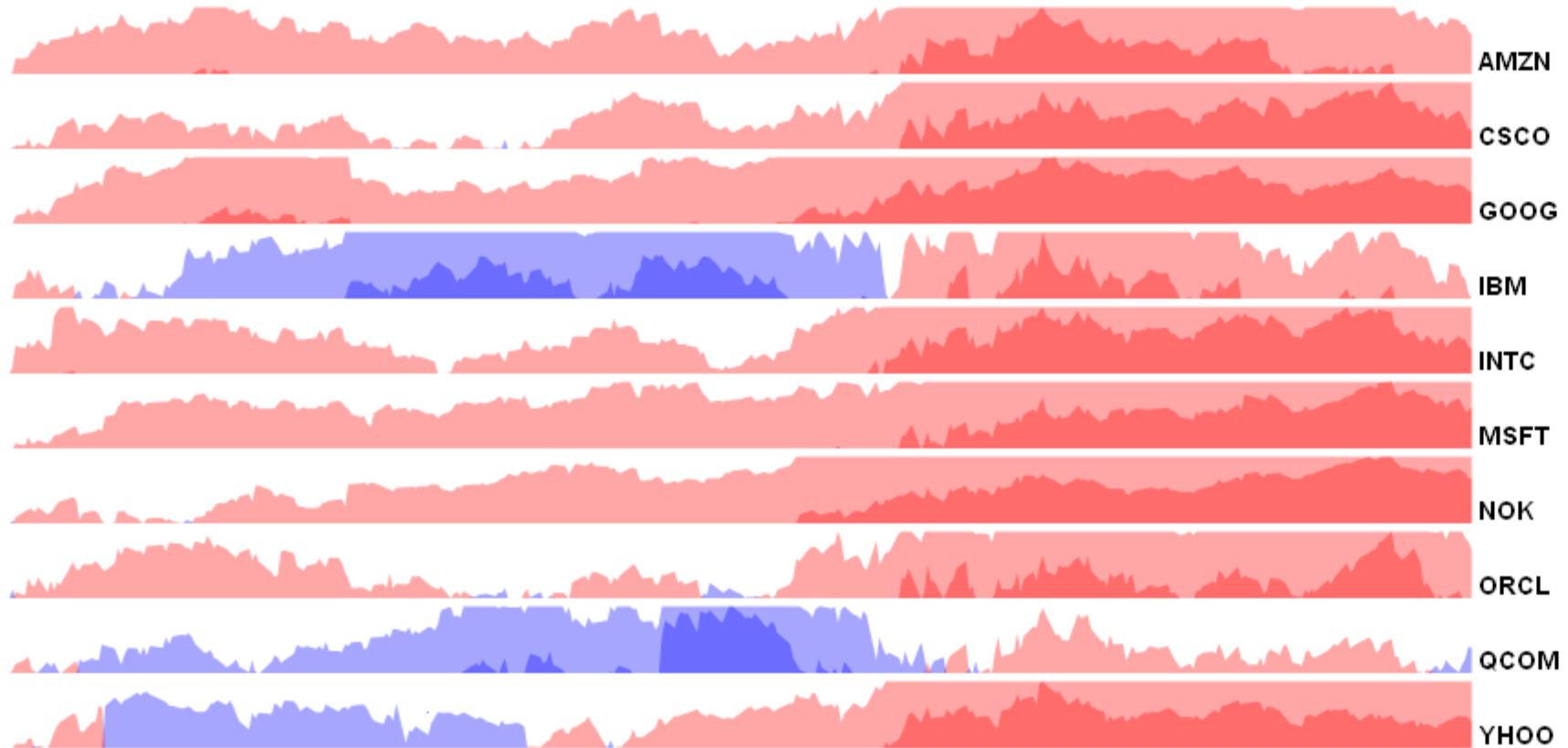
## Relative Technology Stock Performance: Jan 2008 - Present



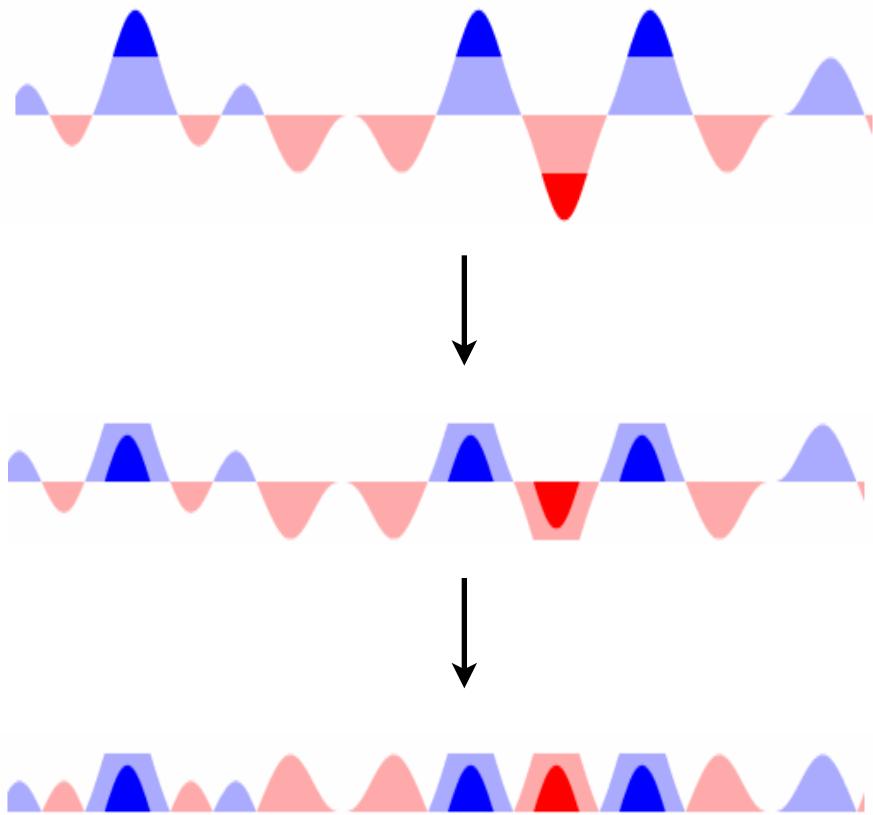
## Relative Technology Stock Performance: Jan 2008 - Present



## Relative Technology Stock Performance: Jan 2008 - Present



# Horizon Graphs

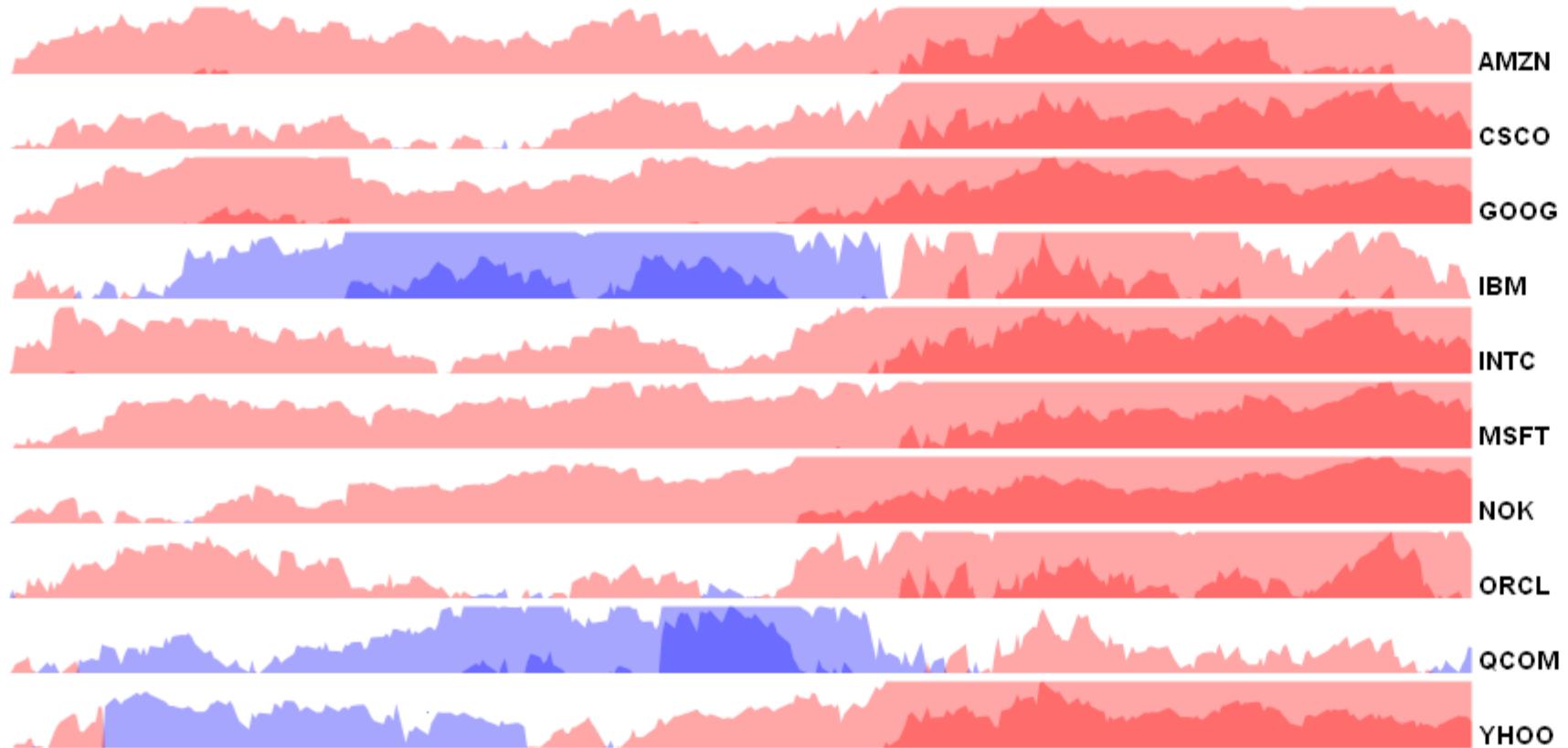


**Segment** Peaks

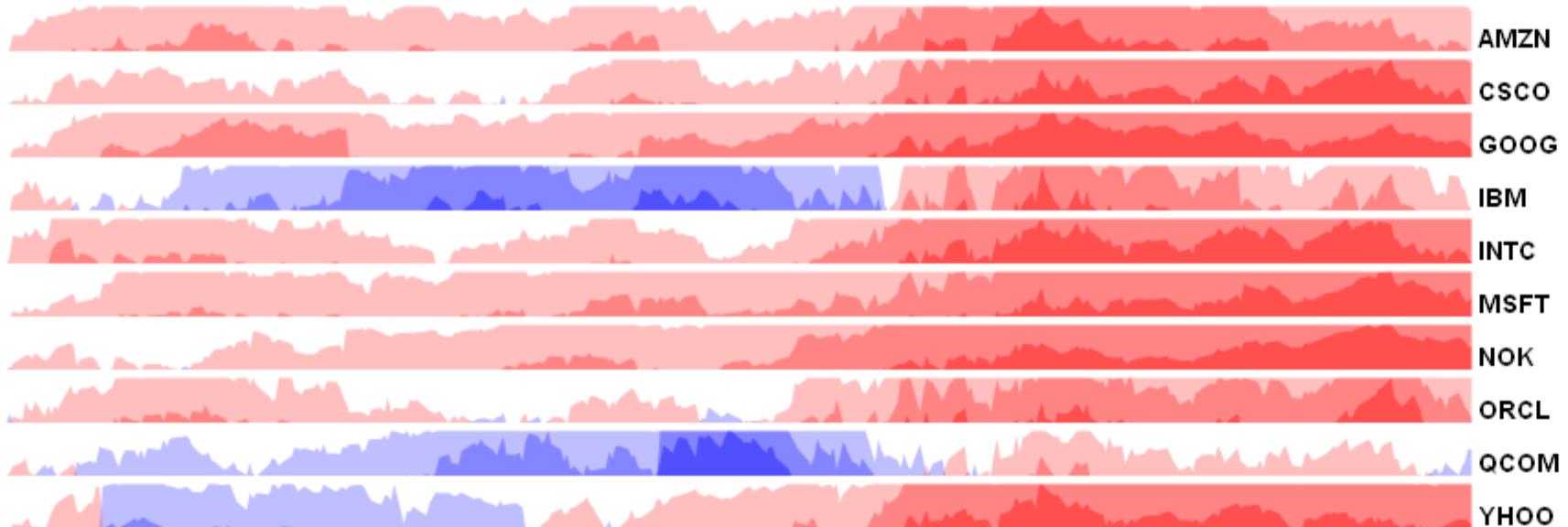
**Layer** Segments

**Mirror** Negative Values

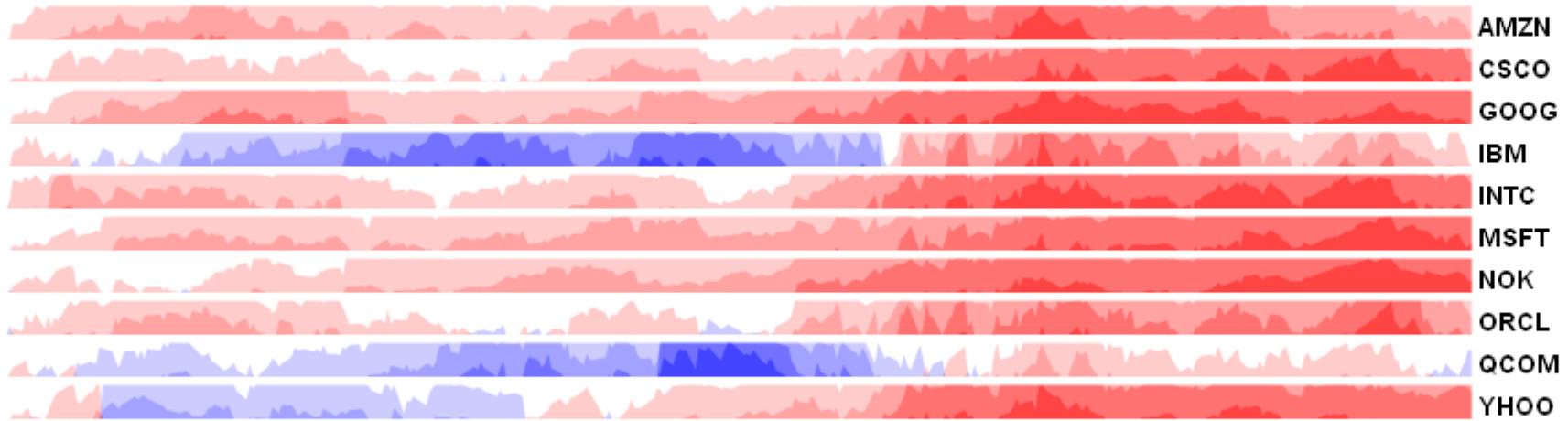
## Relative Technology Stock Performance: Jan 2008 - Present



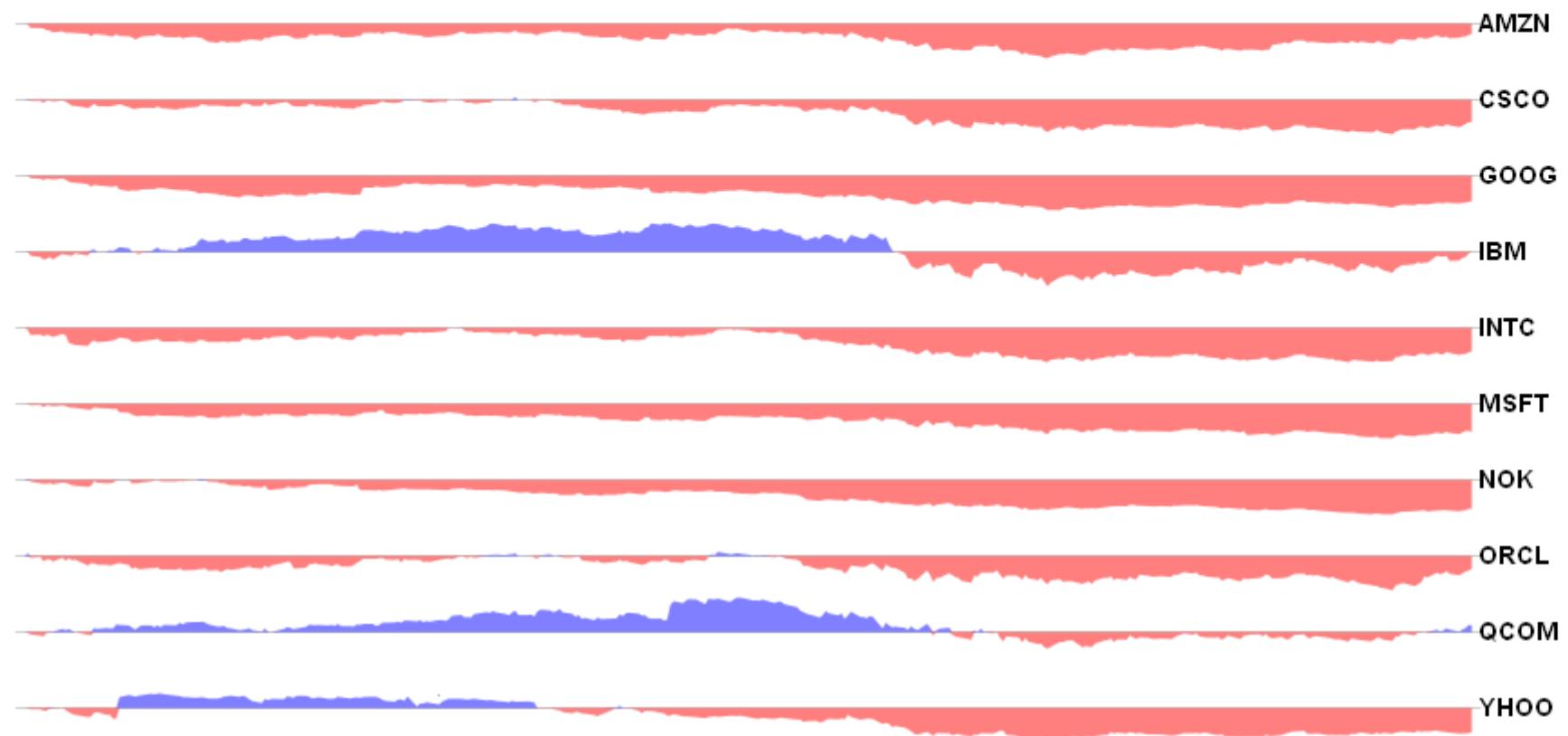
## Relative Technology Stock Performance: Jan 2008 - Present



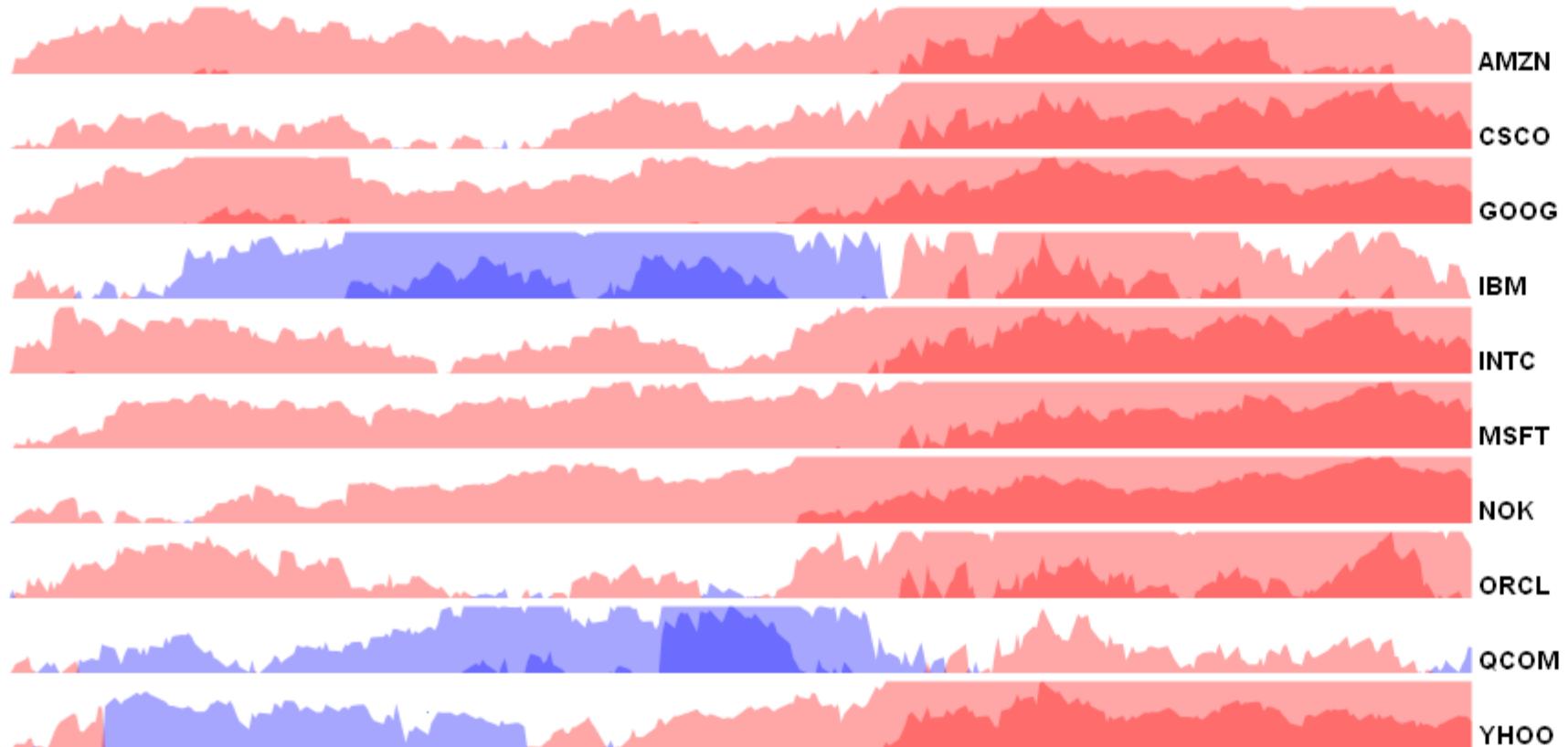
## Relative Technology Stock Performance: Jan 2008 - Present



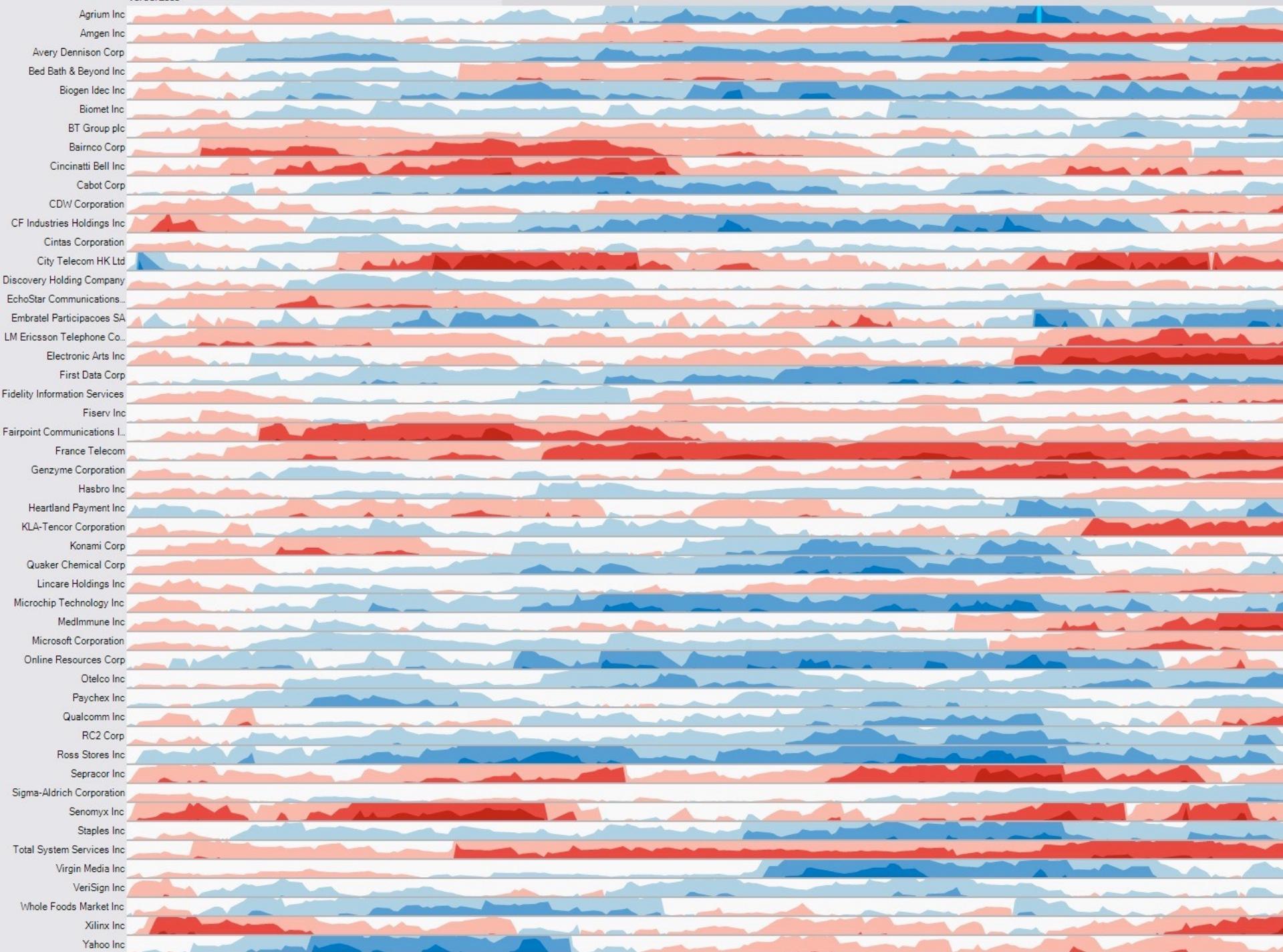
## Relative Technology Stock Performance: Jan 2008 - Present



## Relative Technology Stock Performance: Jan 2008 - Present



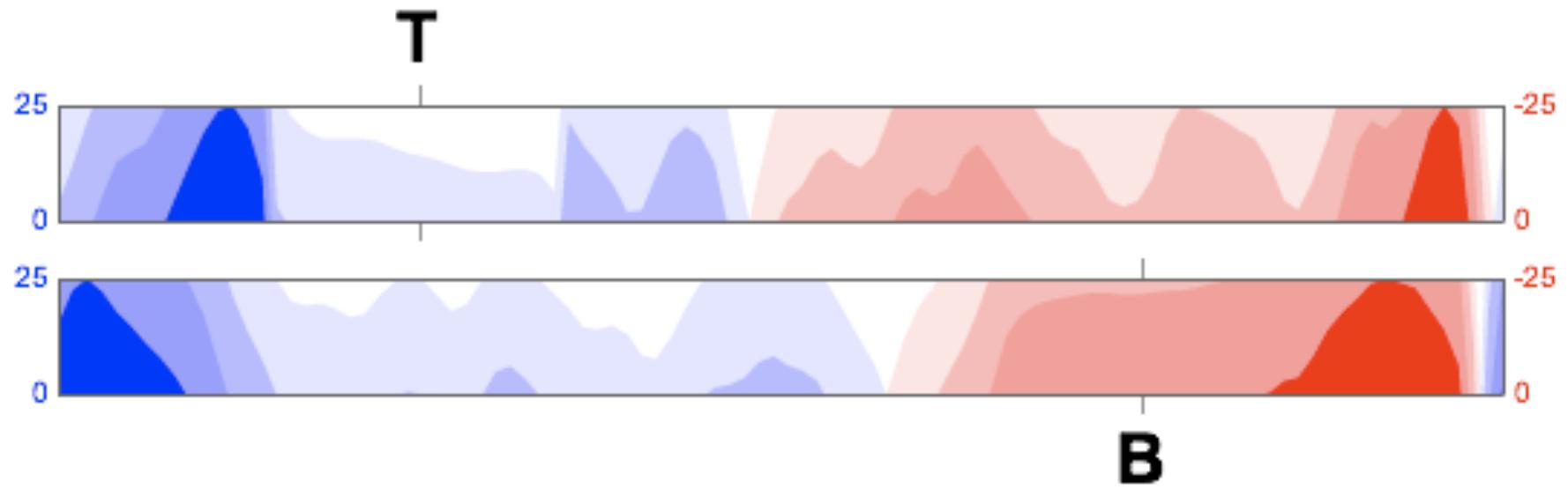
10/03/2005



# Experiment: Chart Type & Size

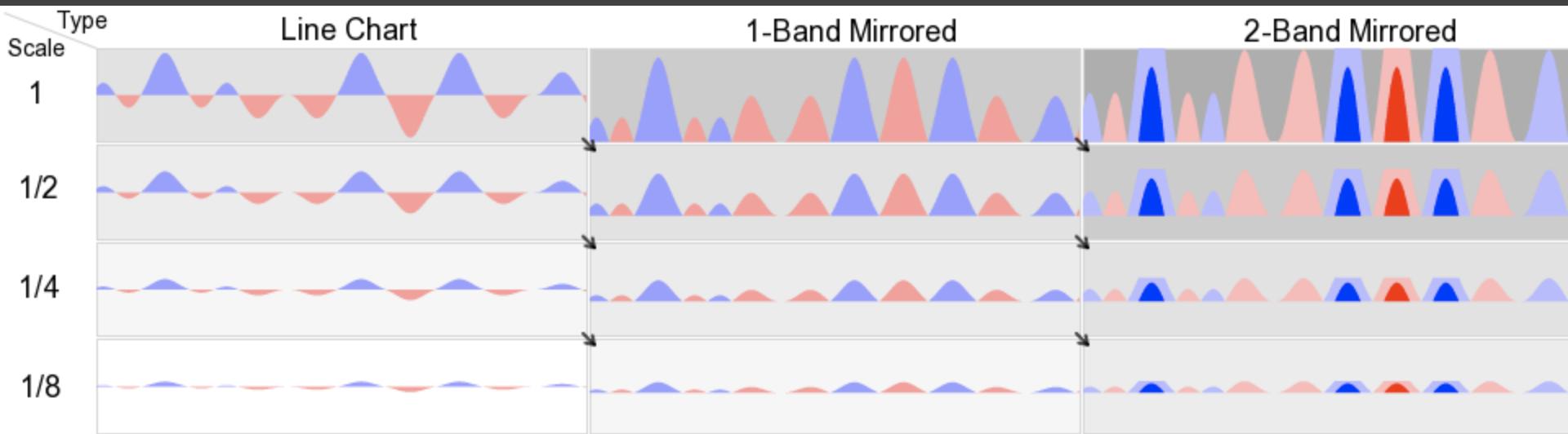
**Q1:** How do mirroring and layering affect estimation time and accuracy compared to line charts?

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



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

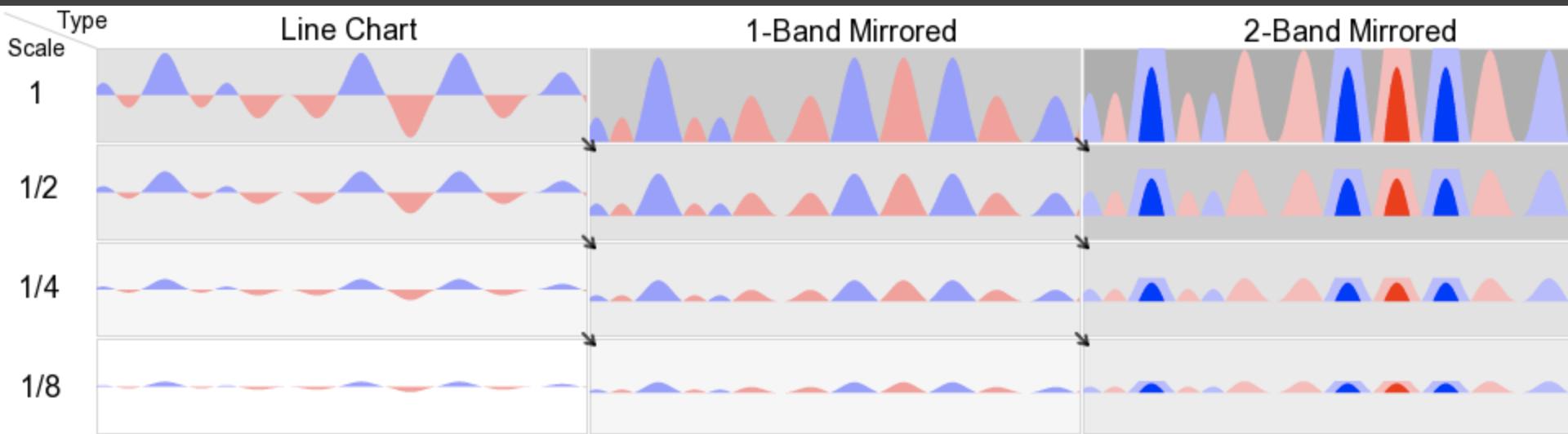
# Experiment Design



3 (chart type)  $\times$  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

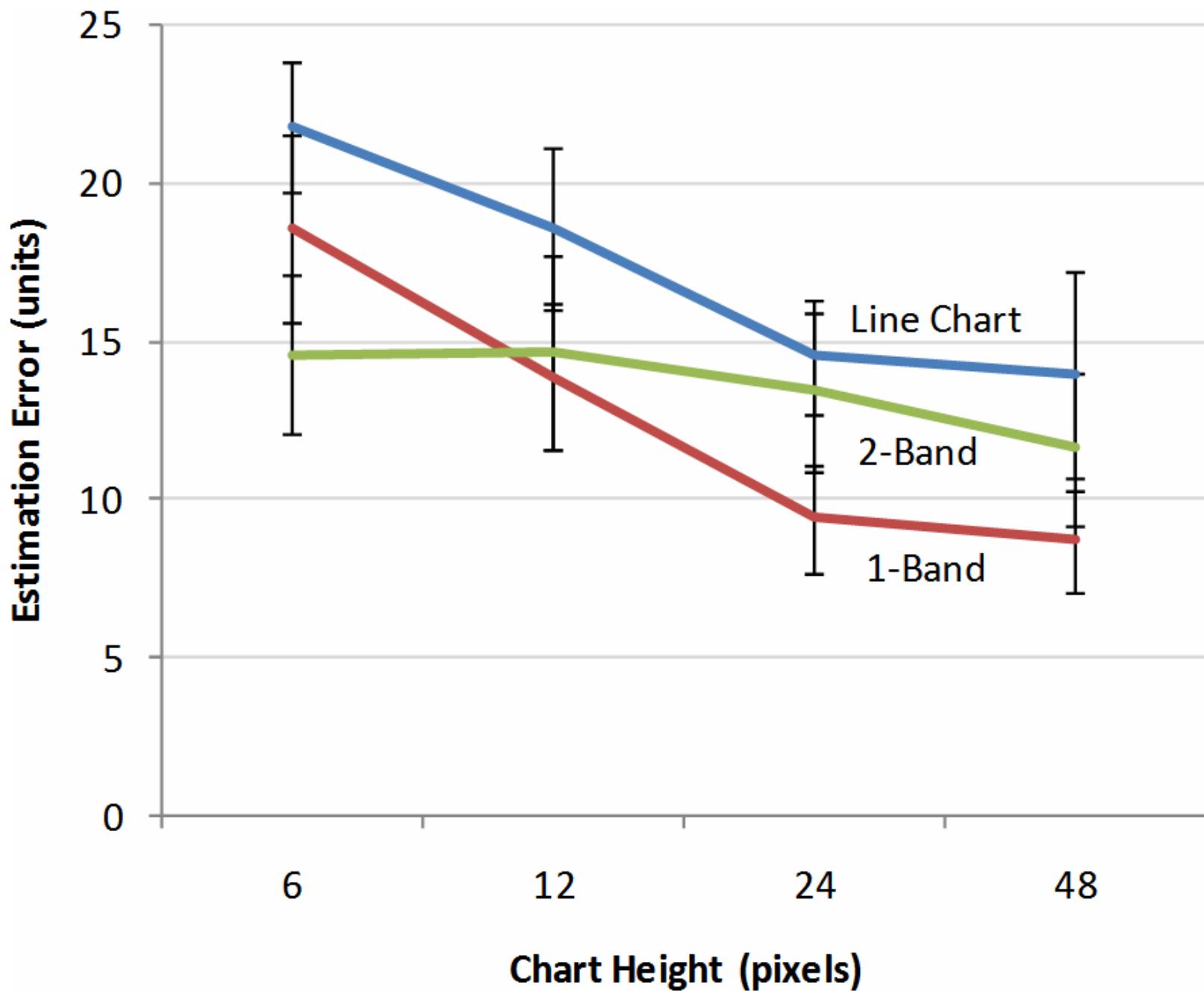


3 (type) x 4 (size) within-subjects design

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

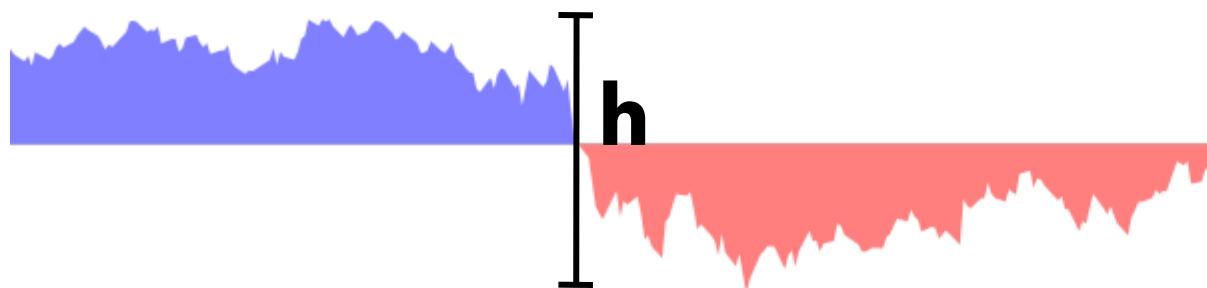
2 (type) x 3 (size: 1/8, 1/12, 1/24) follow-up

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



# Virtual Resolution (VR)

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



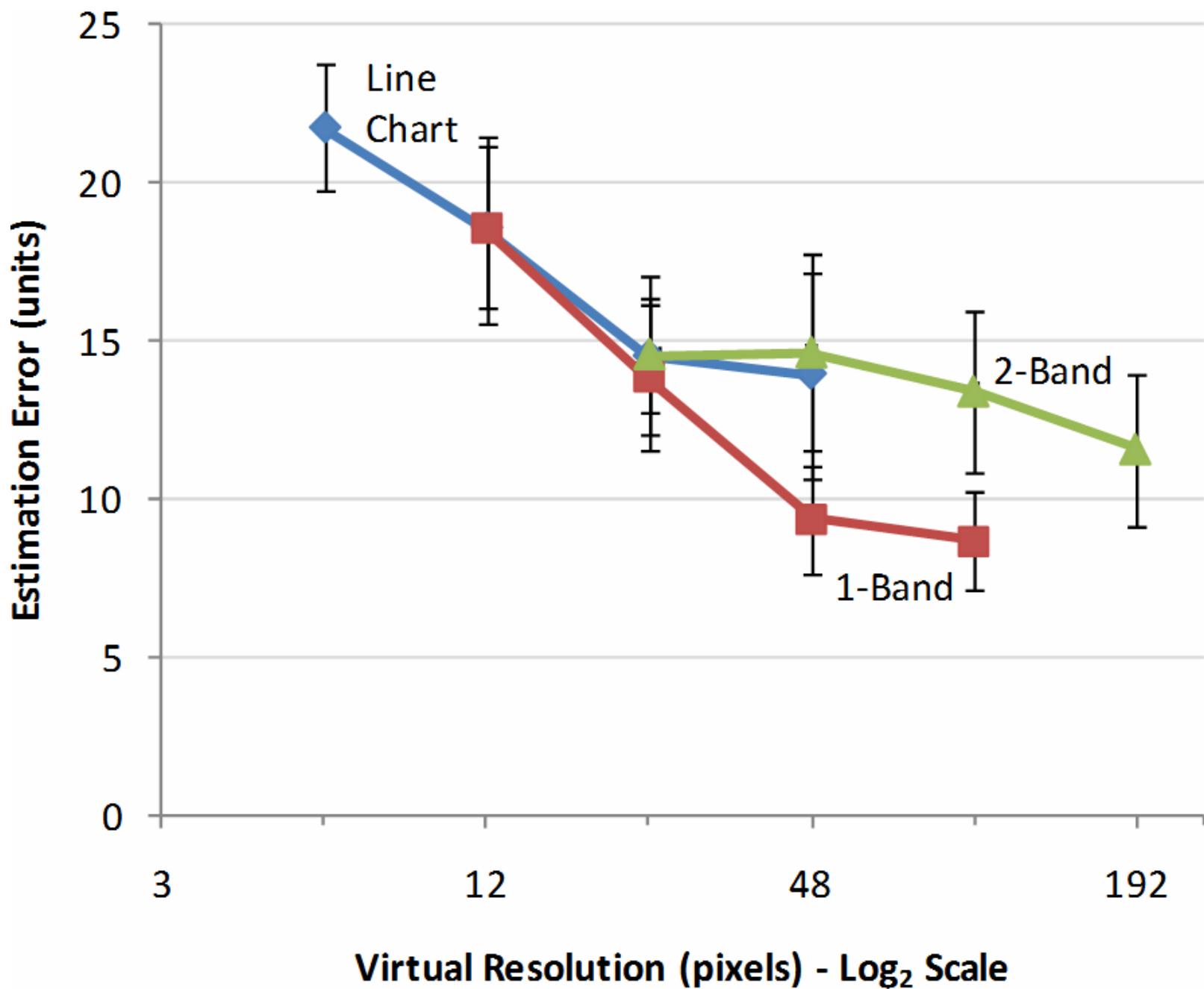
$$\mathbf{VR = h}$$

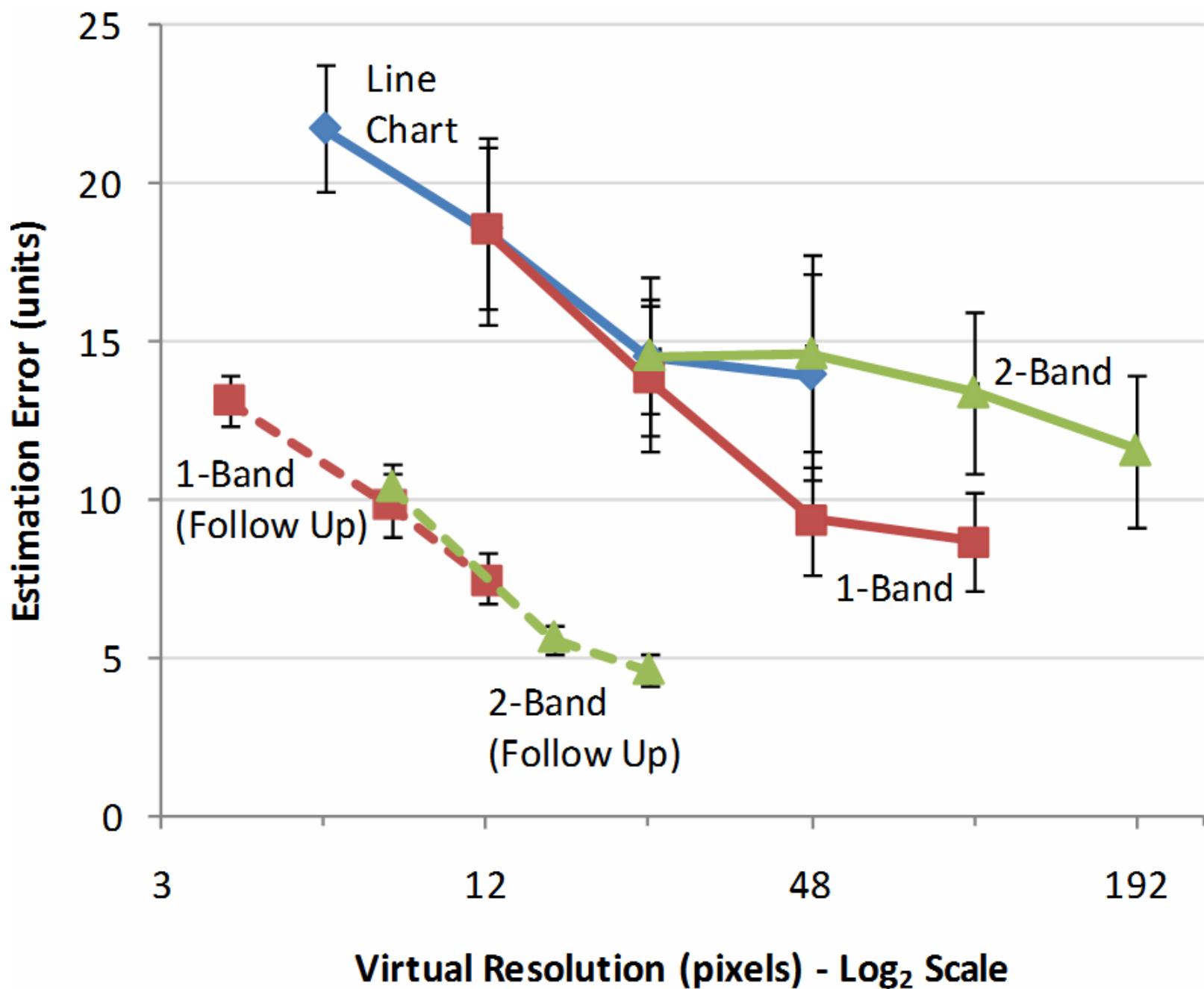


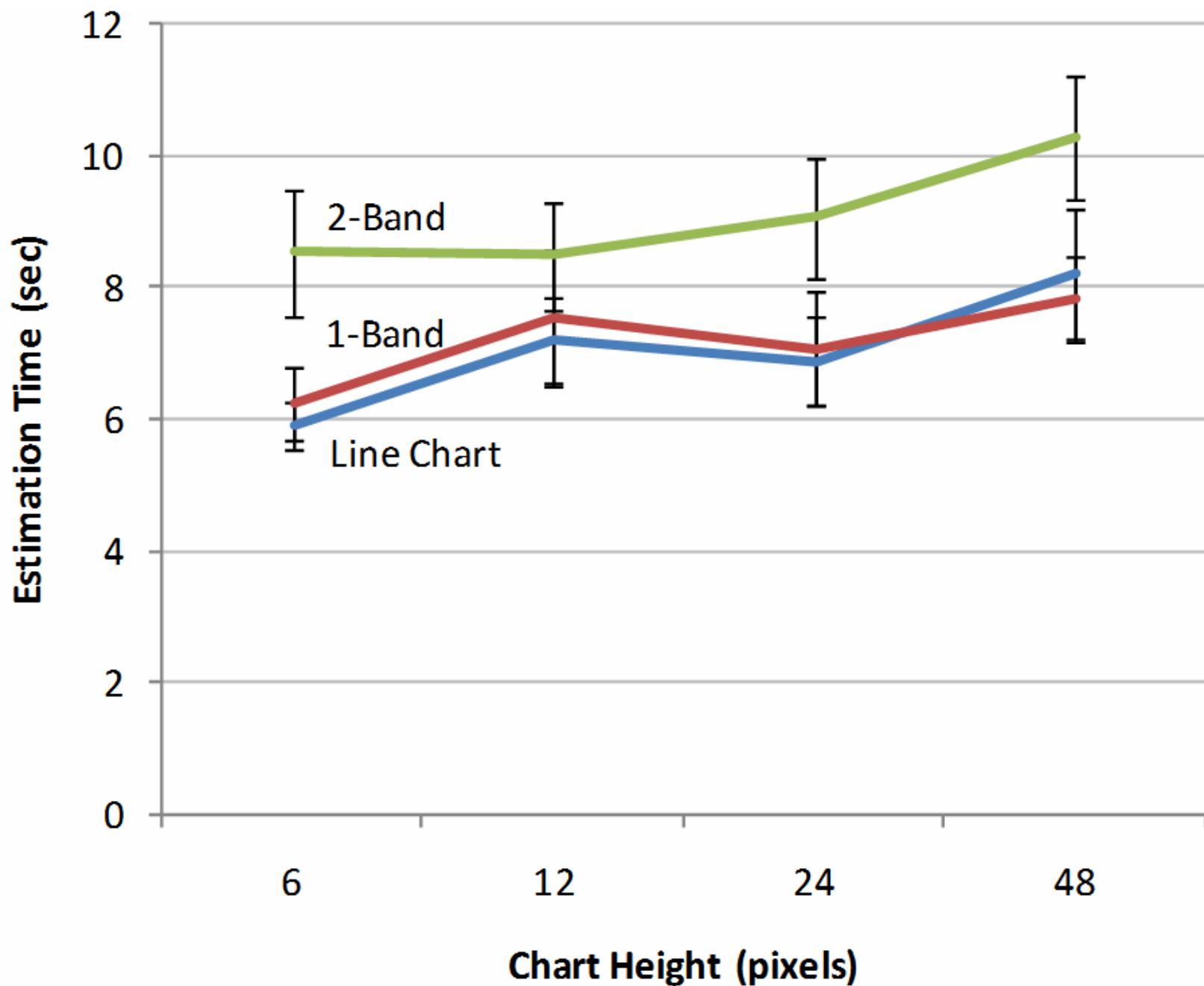
$$\mathbf{VR = 2h' = h}$$



$$\mathbf{VR = 4h''' = h}$$







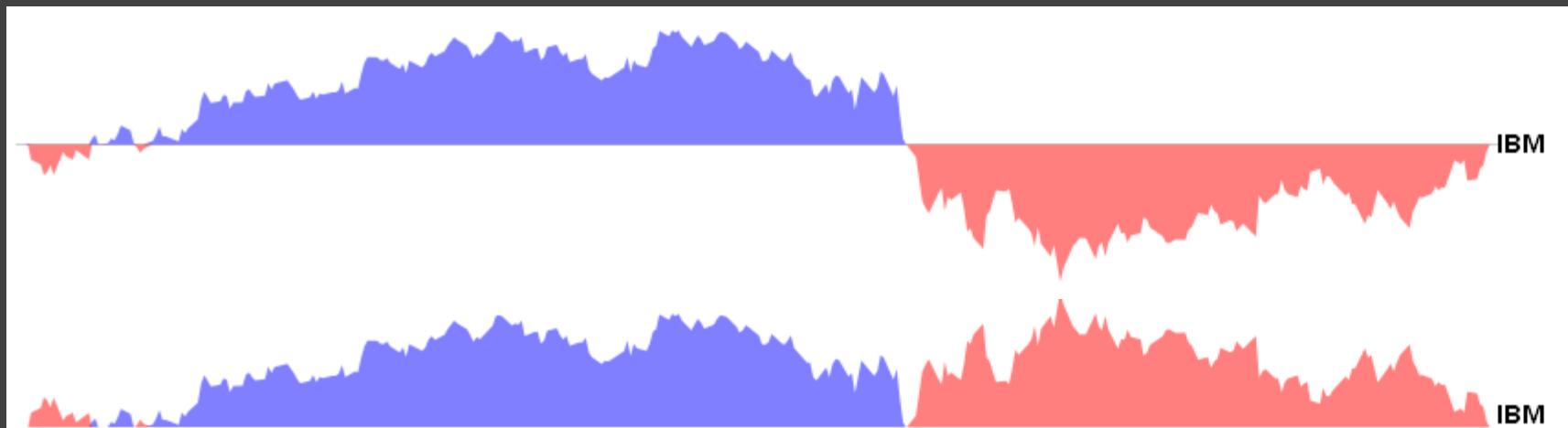
# Experiment Results

**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



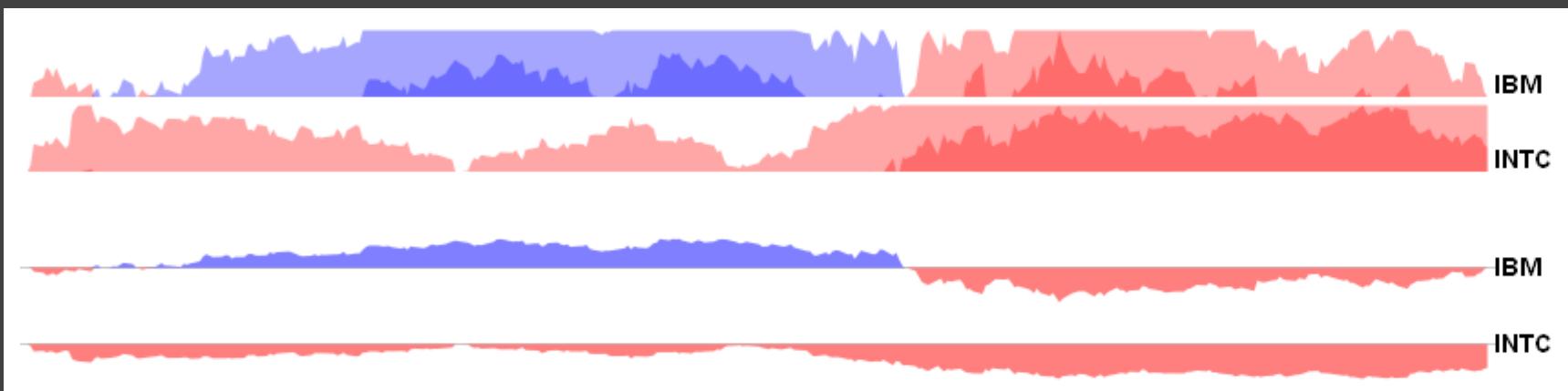
# Design Guidelines

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

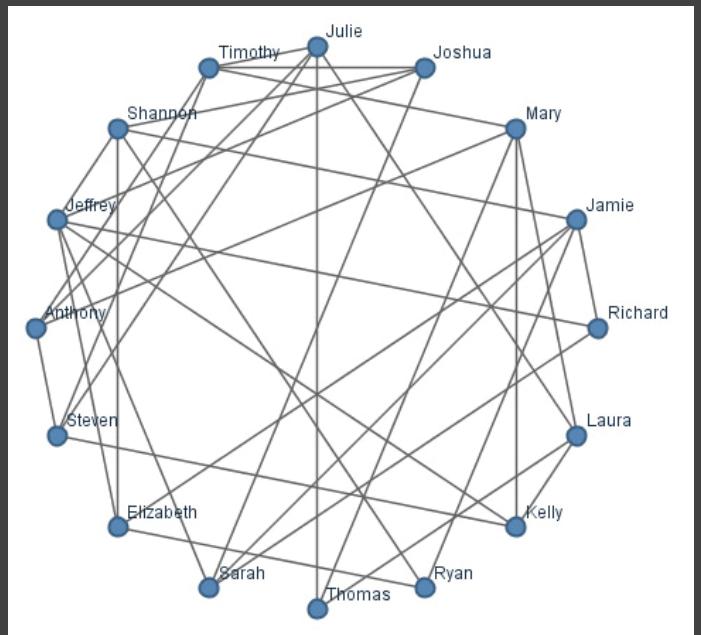
6.8mm (24 px) for line chart & mirrored chart

3.4mm (12 px) for 2-band horizon graph

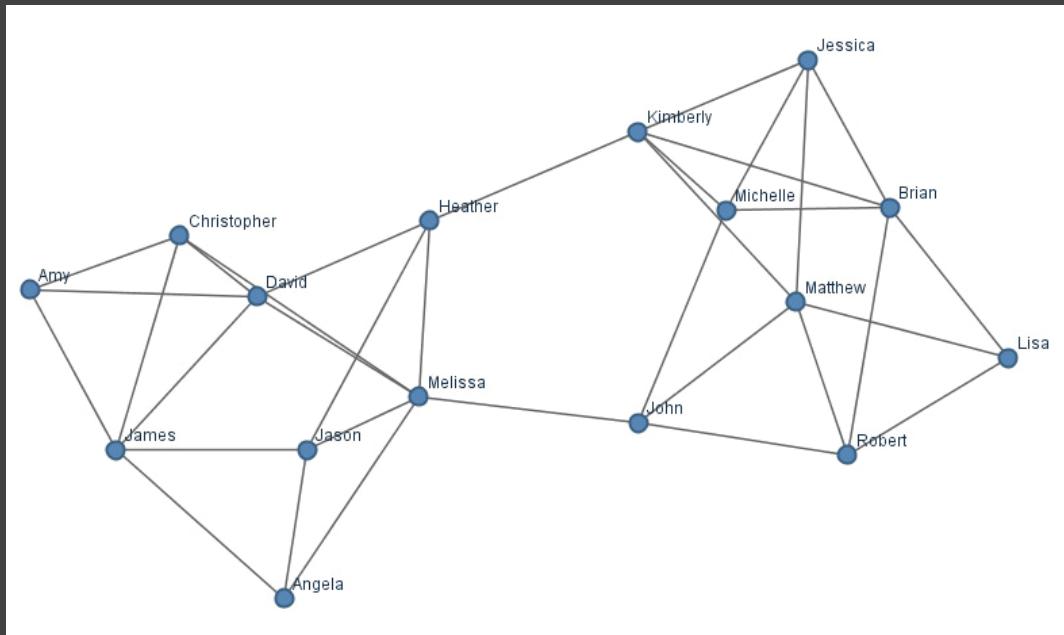
FOLLOW-UP QUESTION:  
What other **tasks** and  
**performance measures**  
should one test?

# 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

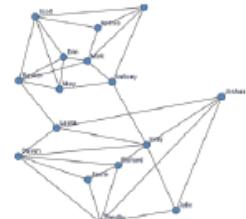
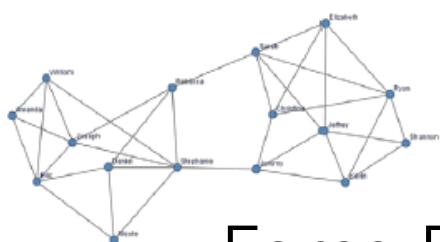
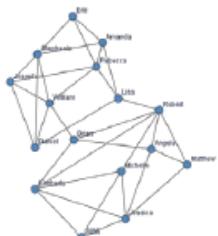
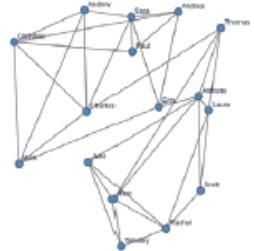
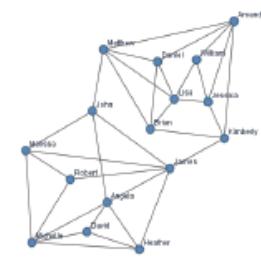
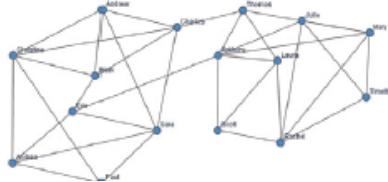
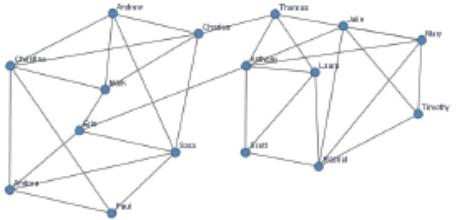
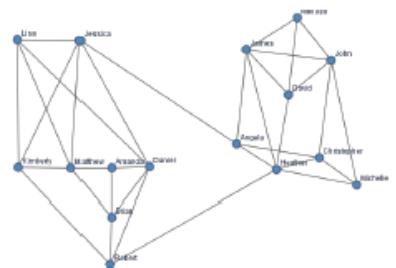
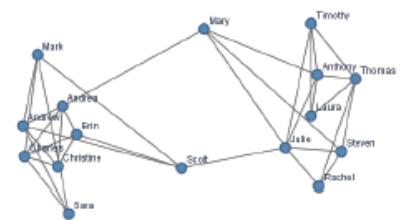
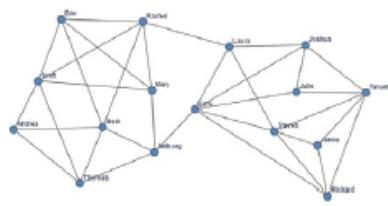
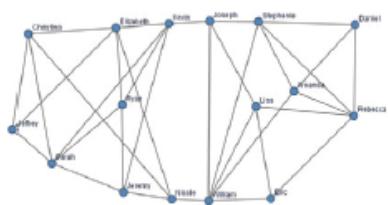
## Measures

# of Edge Crossings

Average Edge Length

Average Node Distance

within or between clusters



Force-Directed

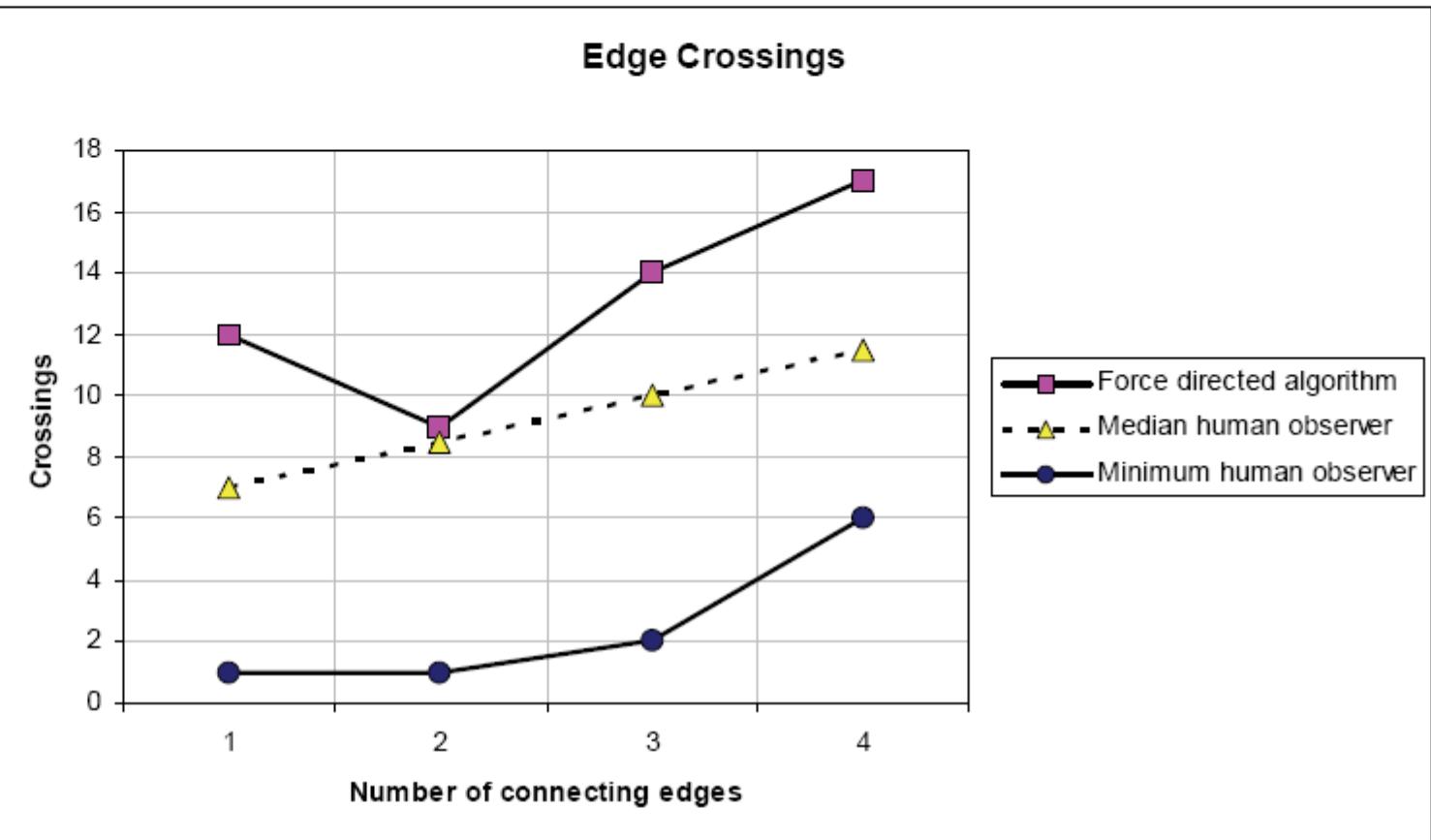


Figure 4. Edge Crossings. Human observers produced graph layouts with fewer edge crossings than the force-directed graph algorithm.

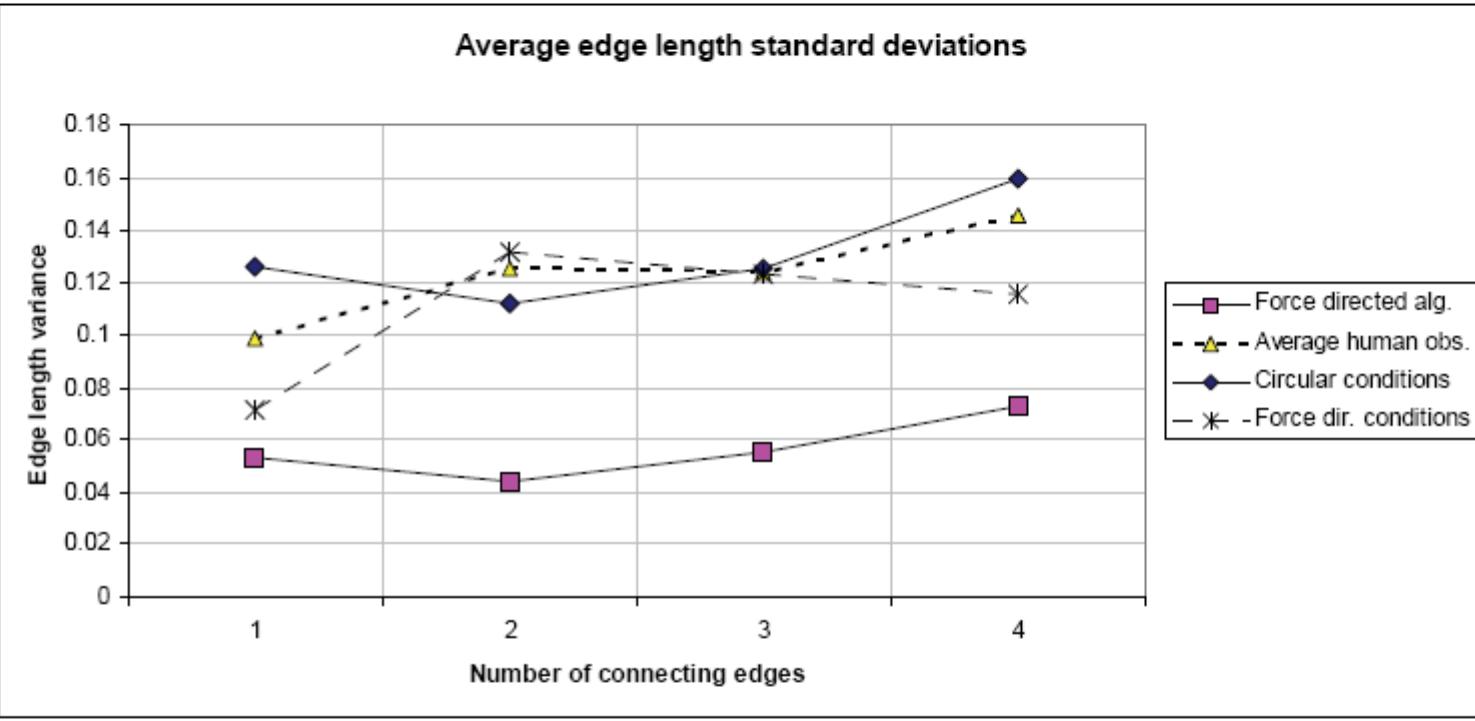


Figure 5. Edge Length Distribution. Human observers did not focus on maintaining equal edge length as much as the force directed algorithm.

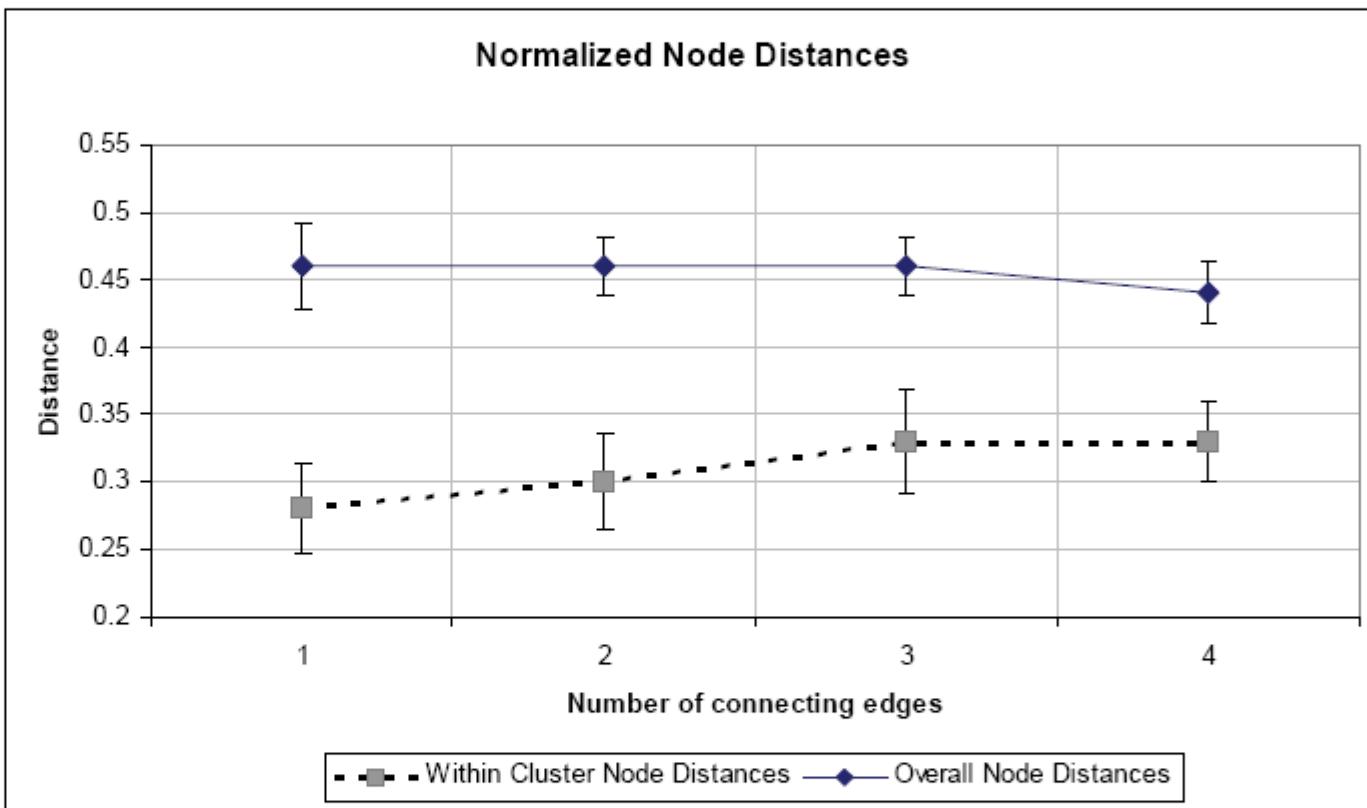


Figure 7. Cluster Extraction. For all levels of masking, the distance between nodes within a cluster is significantly smaller than the overall inter-node distance, demonstrating perceptual grouping. Error bars show 95% confidence intervals

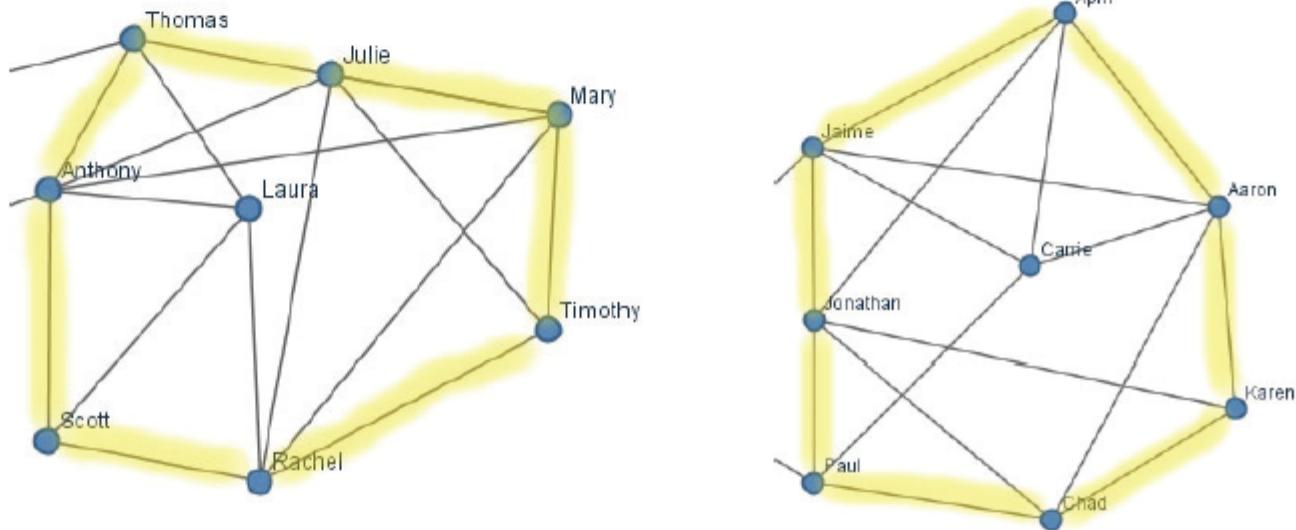


Figure 9. Cluster Hulls. Two examples of user-generated layouts where cluster edges formed a hull enclosing the cluster, organizing it into a single perceptual group.

# 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

# Administrivia

# Final Project Deliverables

## **Interactive Web Page**

Post online by midnight Thu 12/6.

## **Demonstration Video** (<= 2 min)

Due on Vimeo & Canvas by midnight Thu 12/6.

## **Final Project Showcase**

We will show demo videos in class, Fri 12/7.

Read assignment description for more!

# Course Evaluation

**Official course evaluation, due by 12/9**

*Counts as part of course participation.*

Your opinion is valued!

<https://uw.iasystem.org/survey/200289>

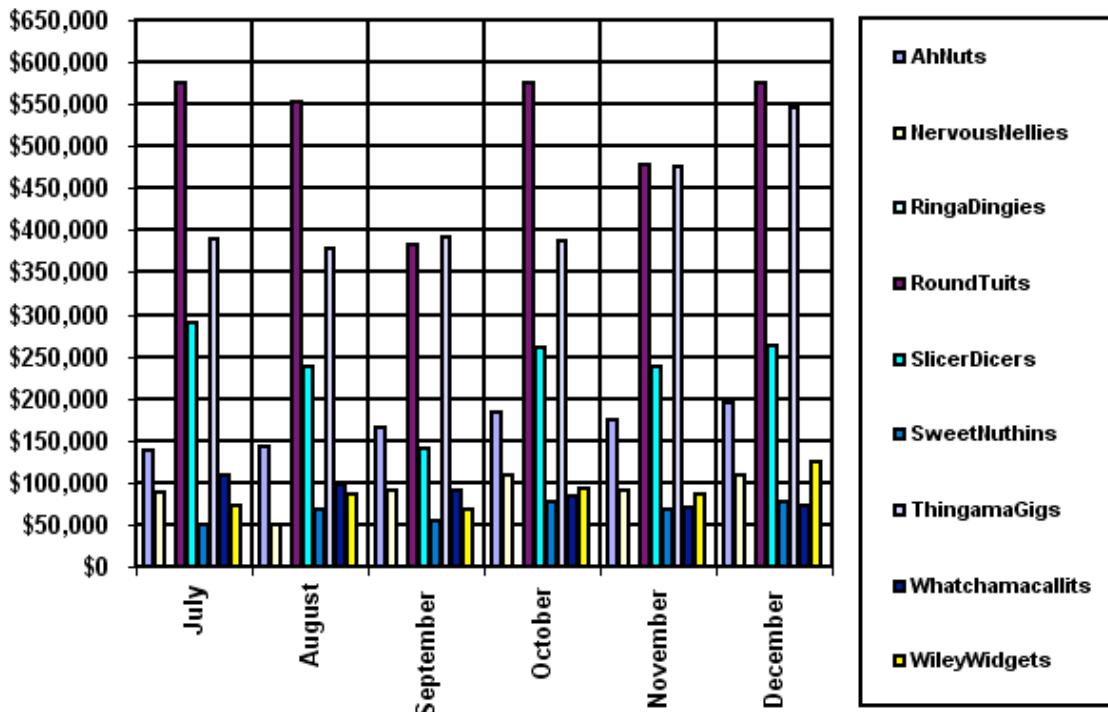
# Course Summary

# Data and Image Models

LES VARIABLES DE L'IMAGE							
	POINTS	LIGNES	ZONES	12	14		
XY 2 DIMENSIONS DU PLAN							
Z TAILLE							
VALEUR							
LES VARIABLES DE SÉPARATION DES IMAGES							
GRAIN							
COULEUR							
ORIENTATION							

# Visualization Design

SlicerDicers' Sales Compared to Other Products

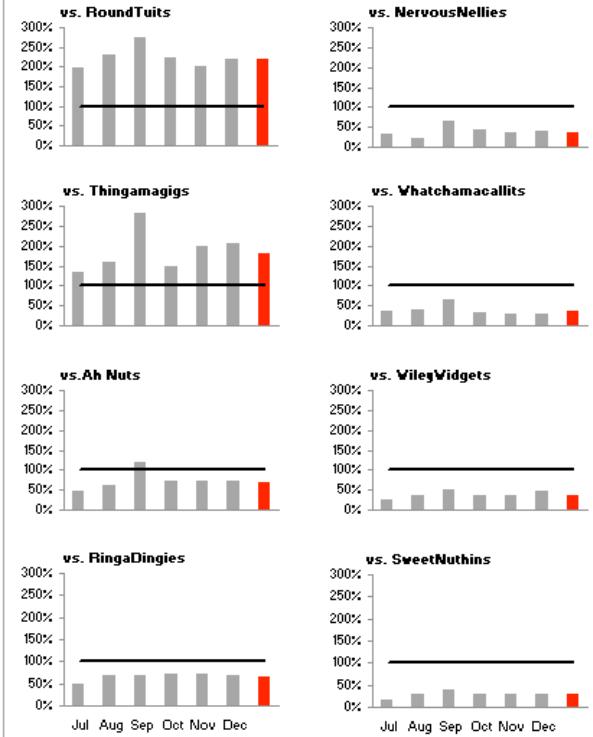


Problematic design

Sales of SlicerDicers Compared to Other Products

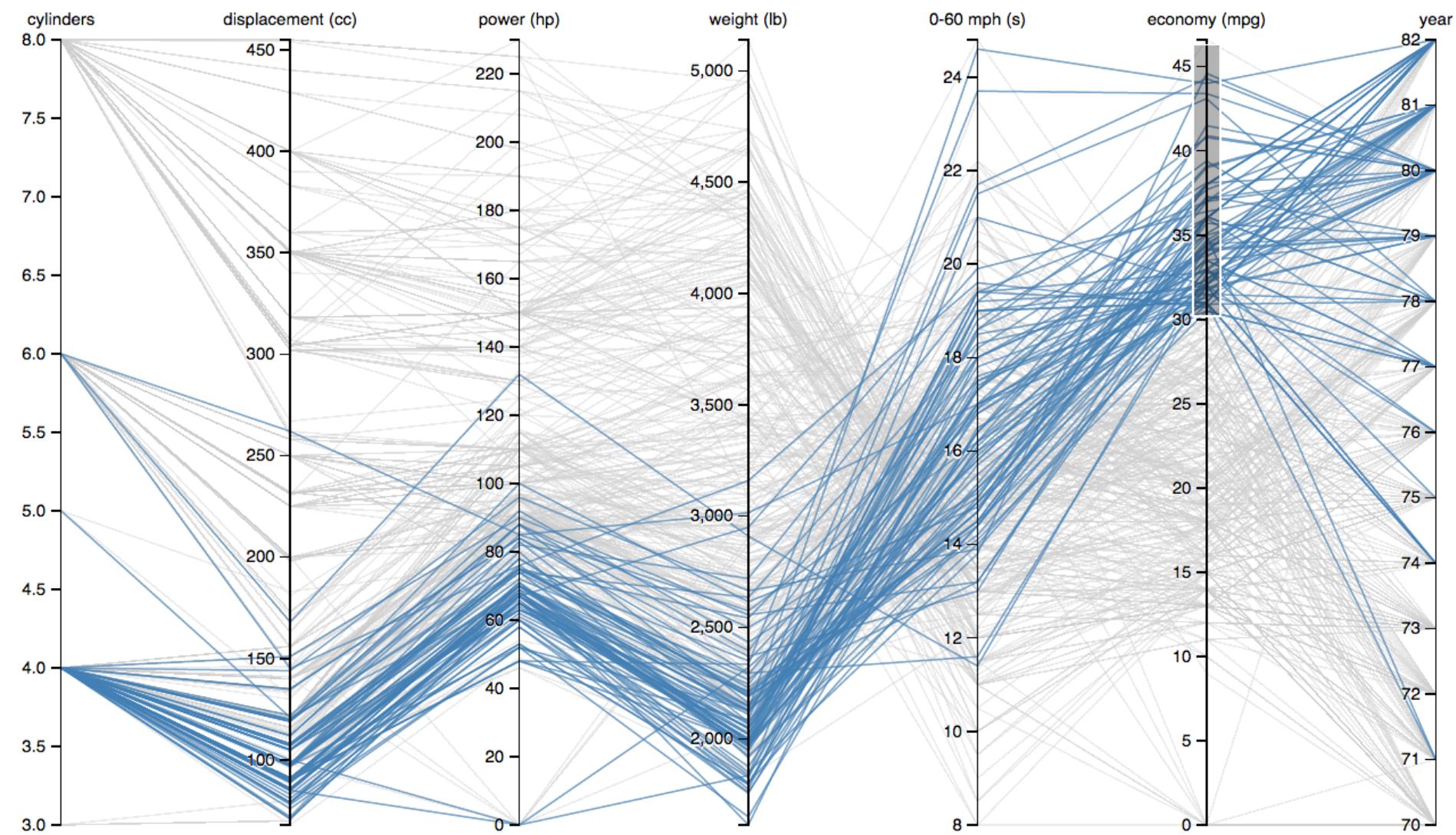
July - December, 2003

(SlicerDicers' sales are displayed as black reference lines of 100%; the red bars represent the average monthly sales percentage for July through December.)

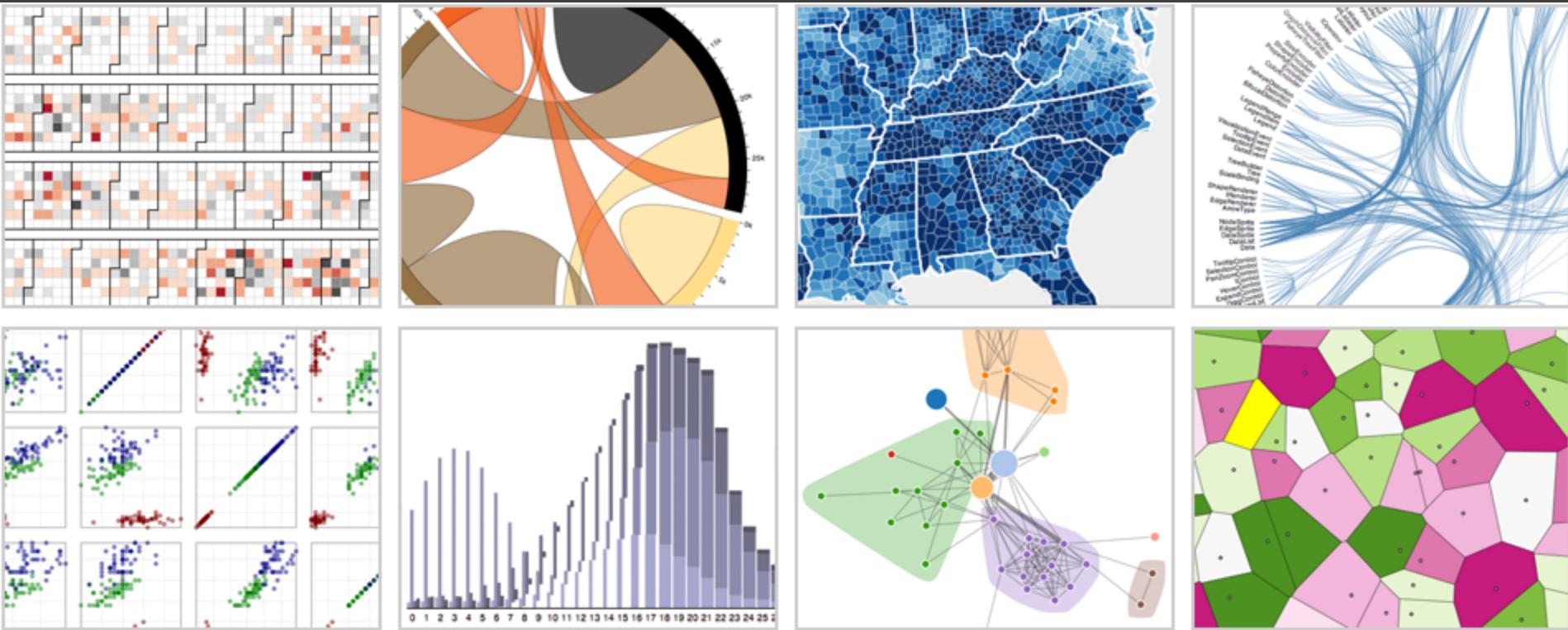


Redesign

# Exploratory Data Analysis

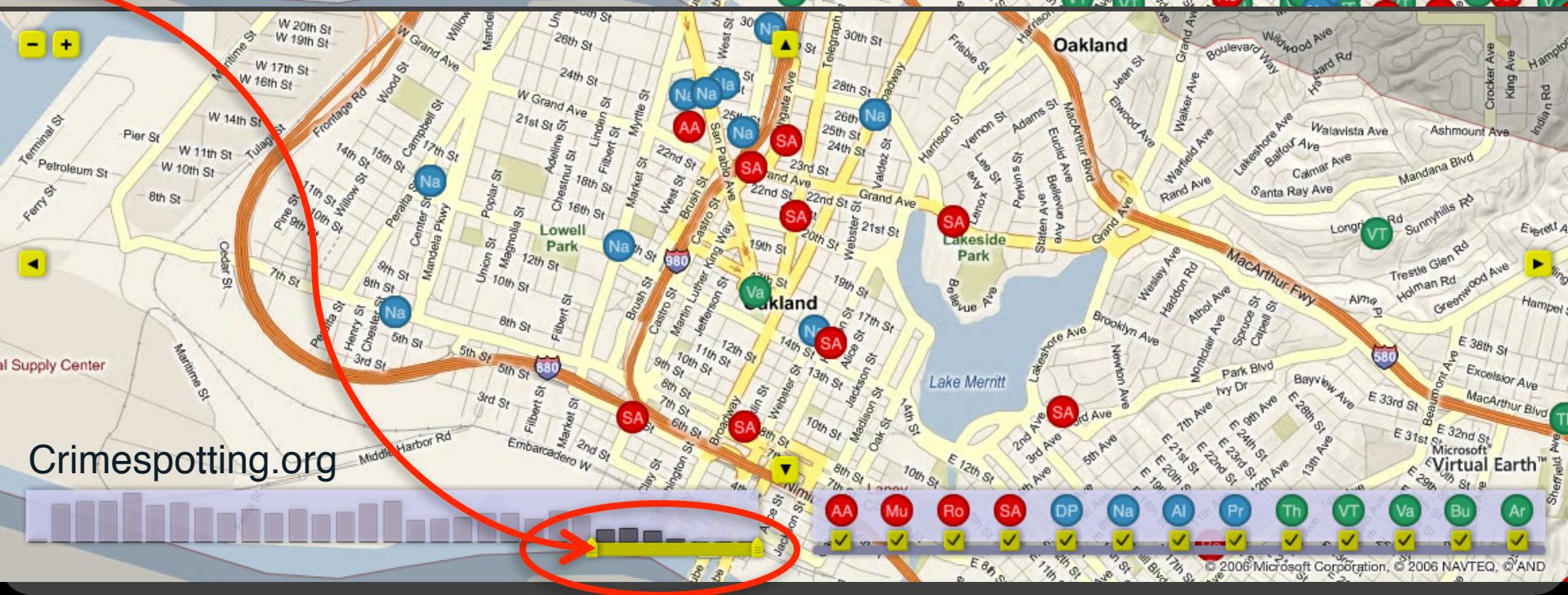
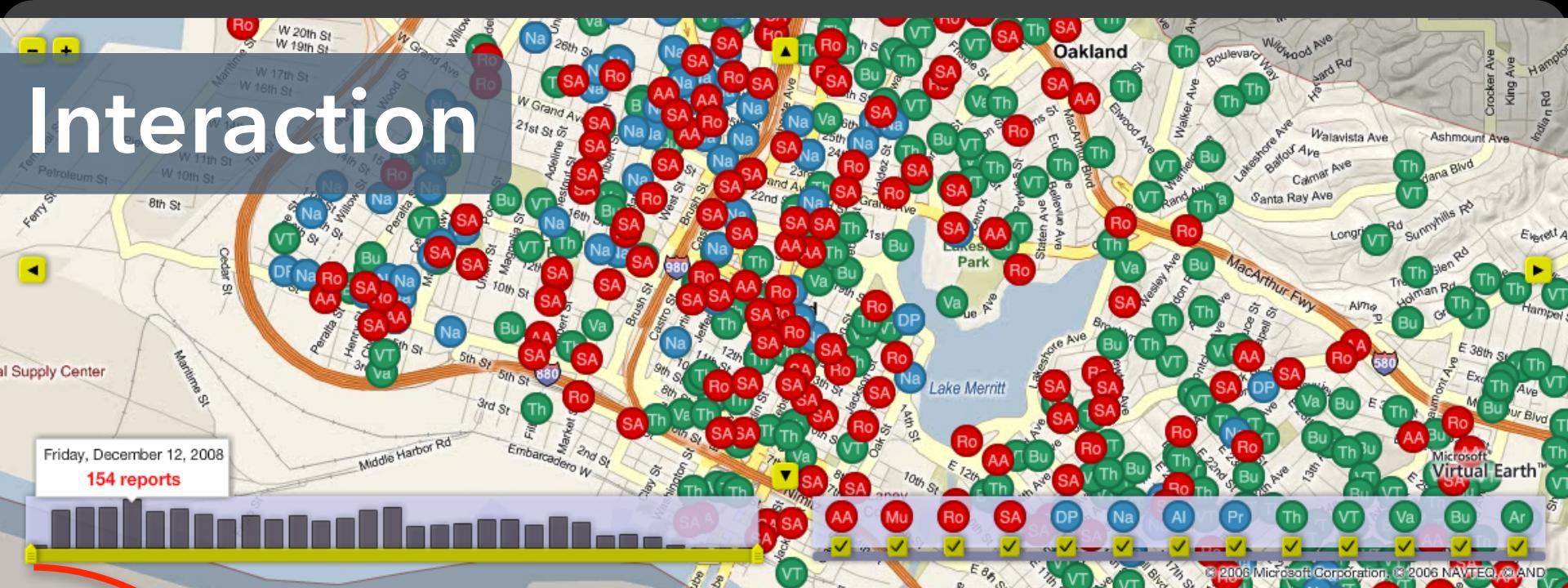


# Visualization Software

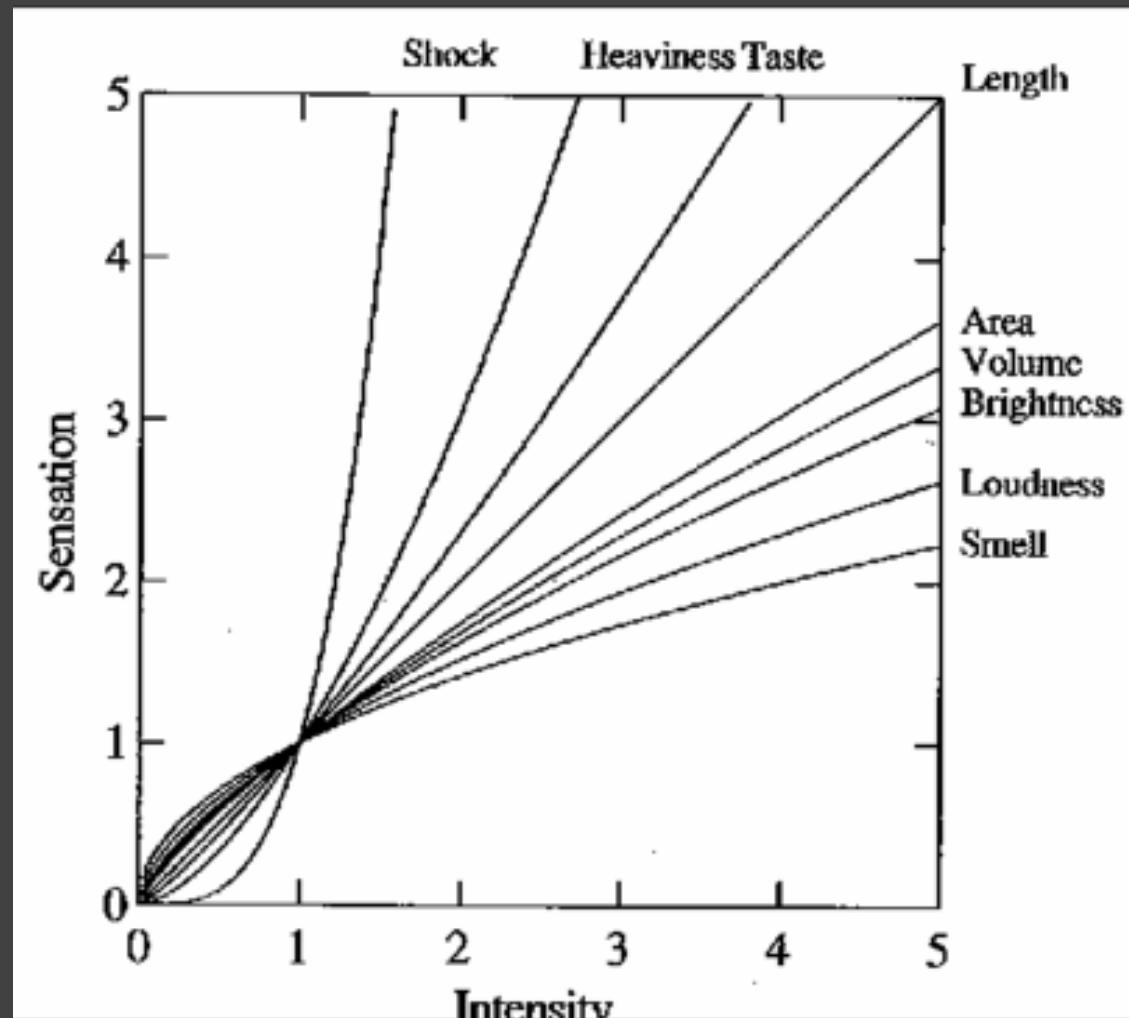


D3: Data-Driven Documents

# Interaction

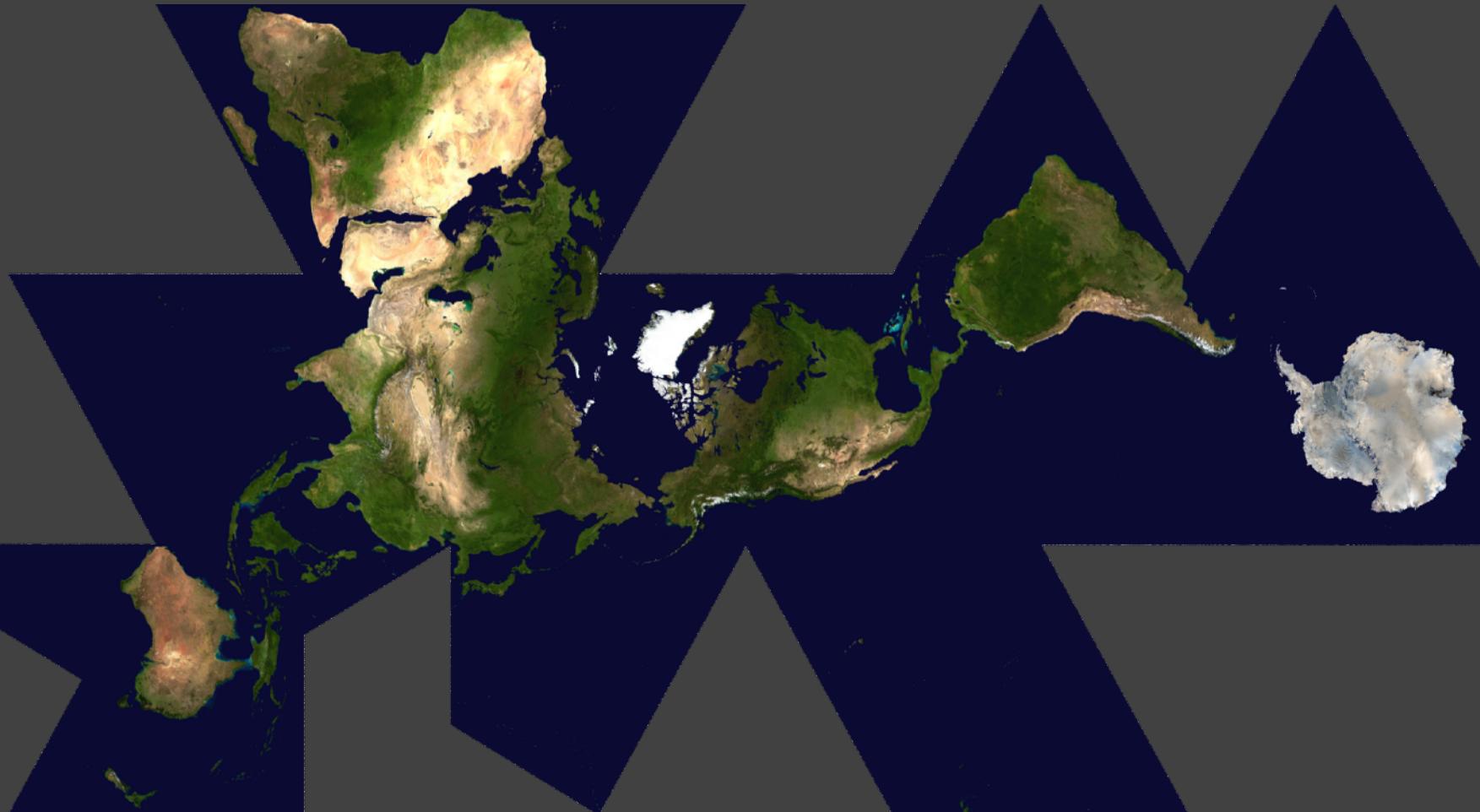


# Graphical Perception



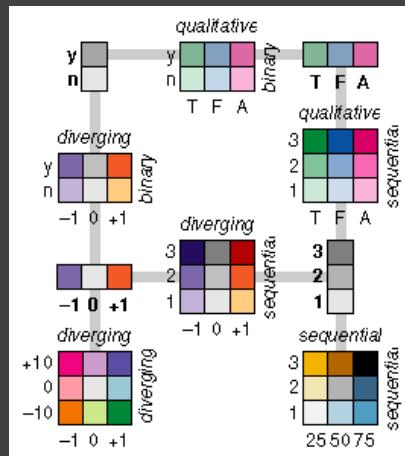
The psychophysics of sensory function [Stevens 61]

# Maps

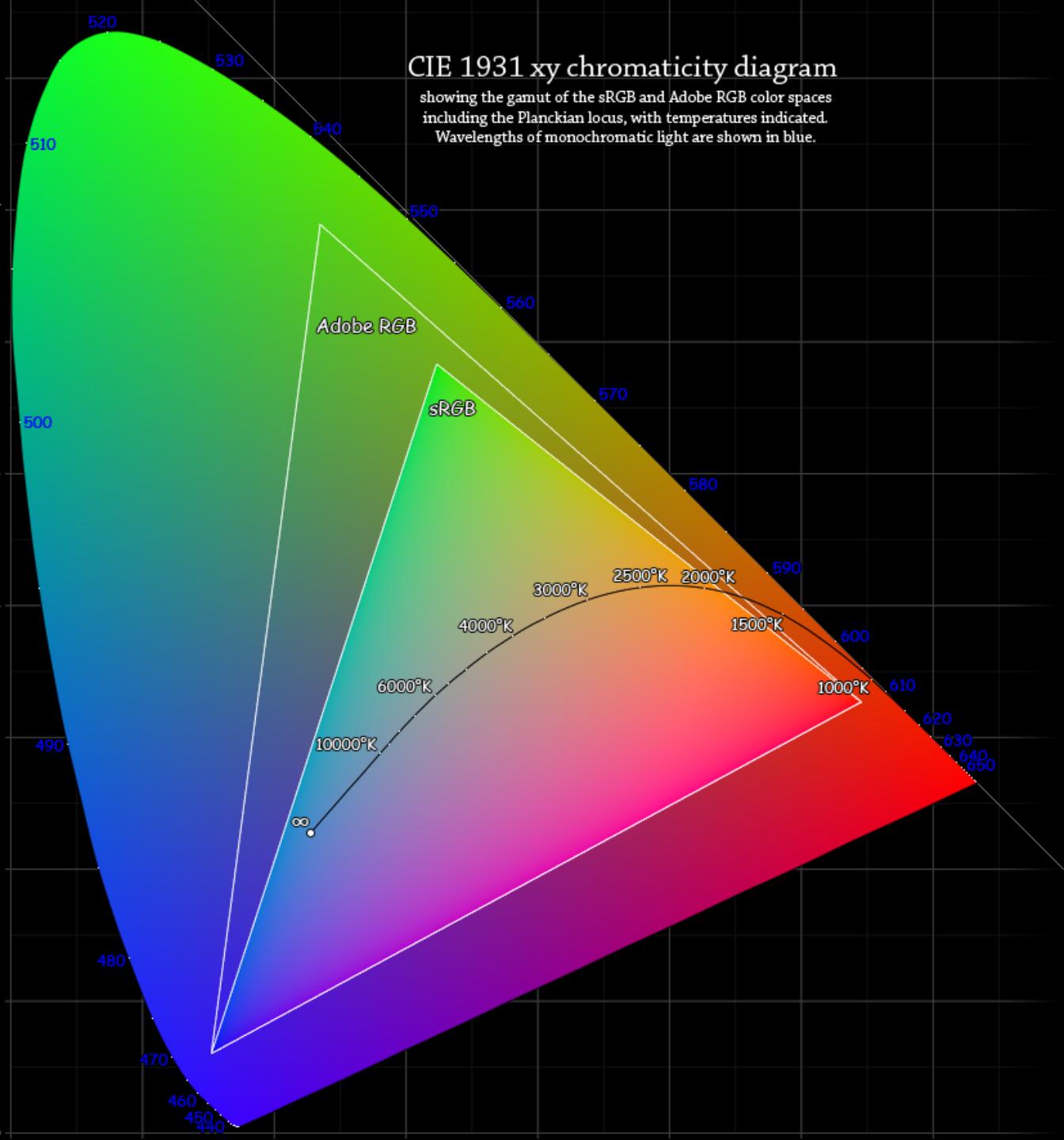


Dymaxion Maps [Fuller 46]

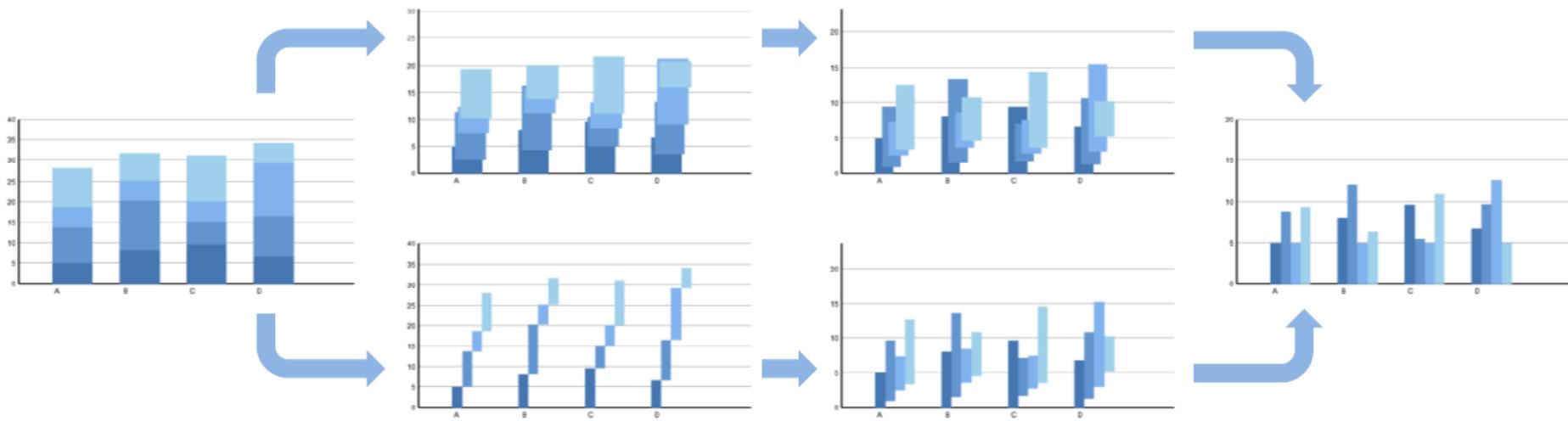
# Color



Color Brewer



# Animation



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

Recent elections have placed a heavy emphasis on “swing states” — Ohio, Florida and the other competitive states. In the past, many more states shifted between the Democratic and Republican parties. A look at how the states stuck to one side or another, and how they have shifted over past elections.

Each box represents a state sized by number of electoral votes.

Each curve shows how much it shifted left or right between elections

# Narrative

Chart  
Size of Lead

Chart  
Electoral Votes

≥50% +40% +30% +20% +10% ← MORE DEMOCRATIC → MORE REPUBLICAN → +10% +20% +30% +40% ≥50%

## Obama Re-elected

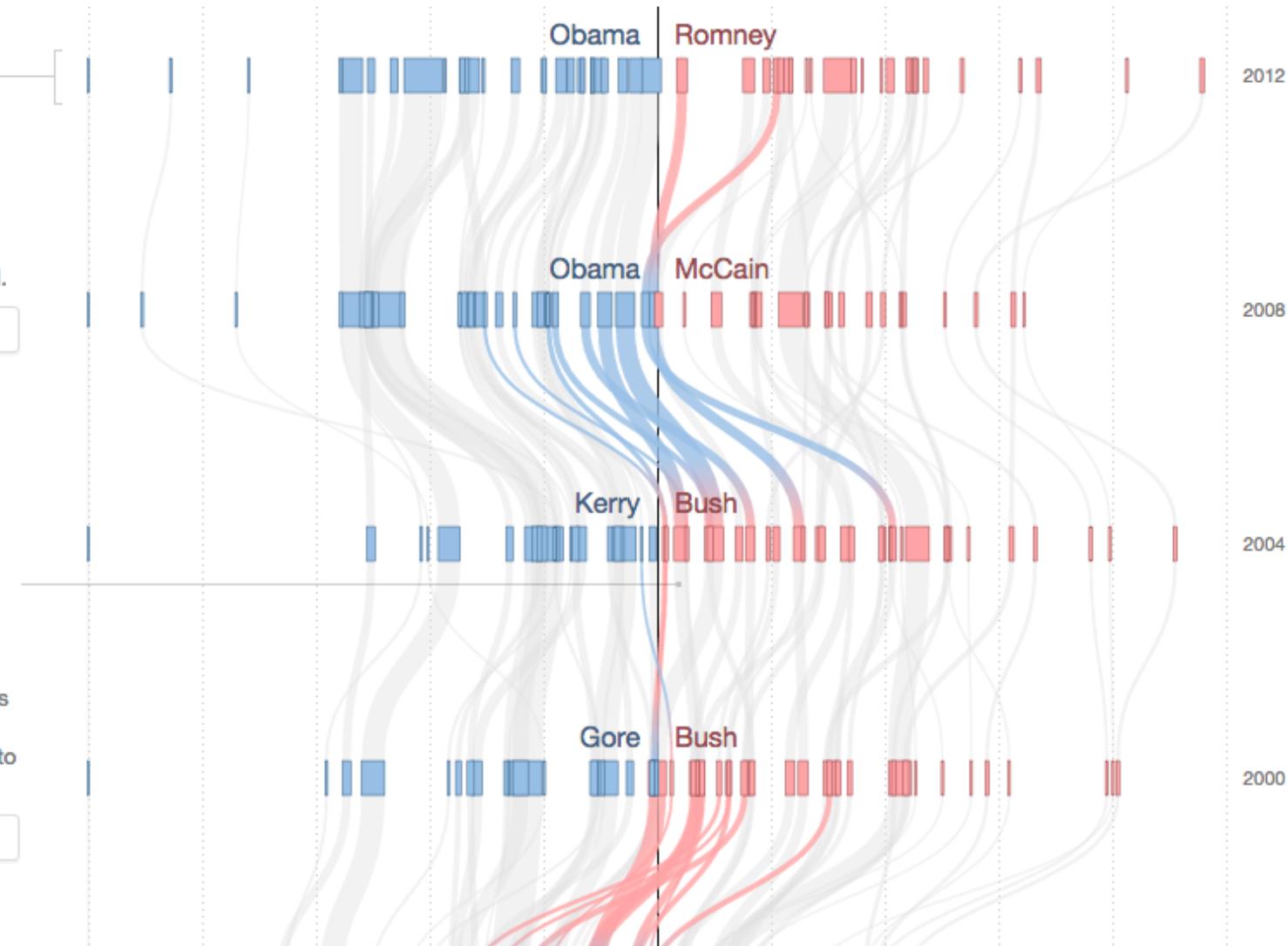
The country voted about 5 percentage points more Republican in 2012 than in 2008. Obama lost North Carolina and Indiana, but won every tossup except Florida, which remains too close to call.

Highlight Tossups

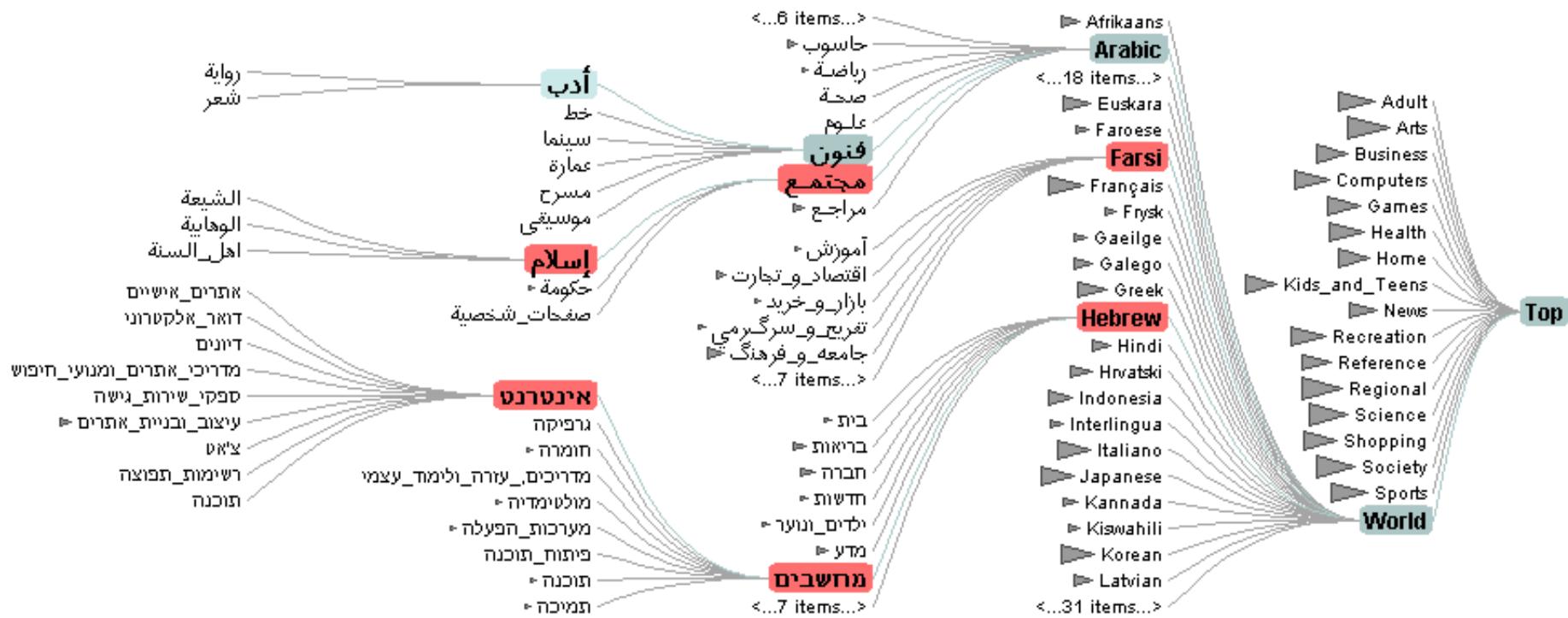
## As Goes Ohio

Ohio, which has voted for the winner in every election since 1964, provided the decisive electoral votes in 2004, and it is the state likeliest to play that role again this year, according to the FiveThirtyEight model.

Highlight Ohio

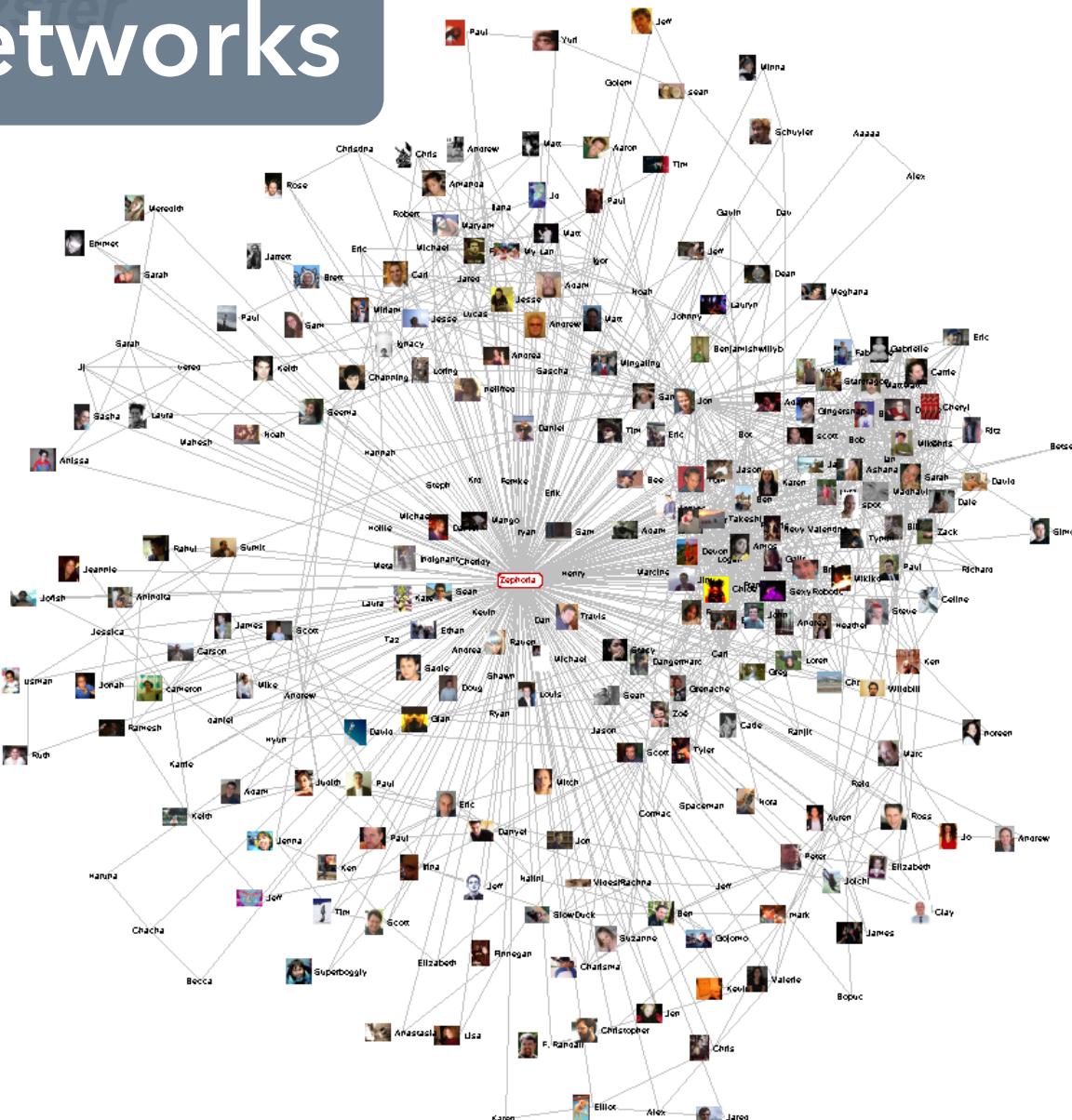


# Hierarchies



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

# Vizster Networks



community &gt;&gt;



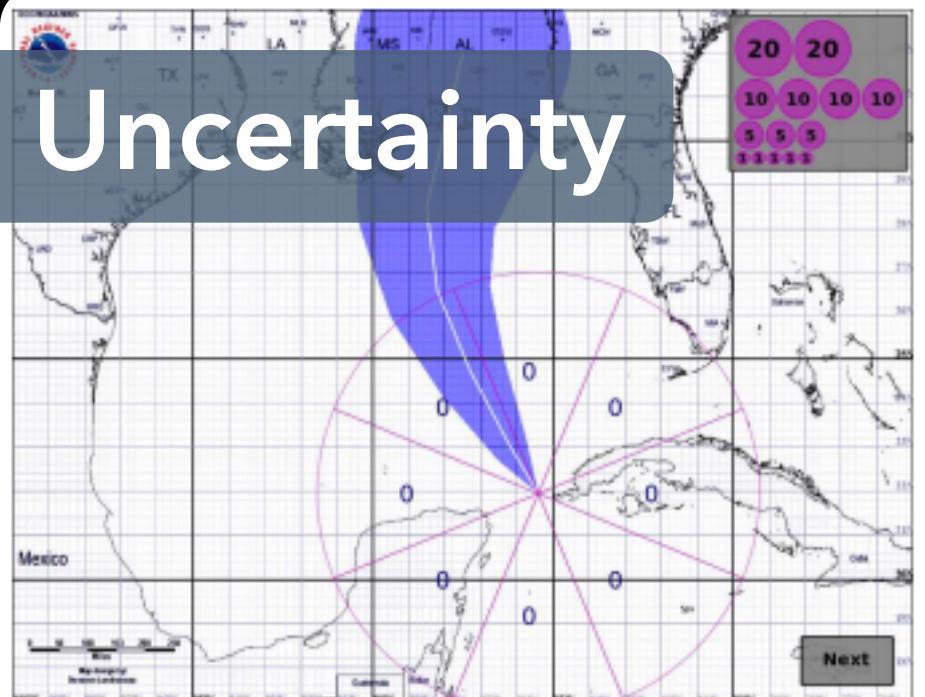
Enable

search &gt;&gt;

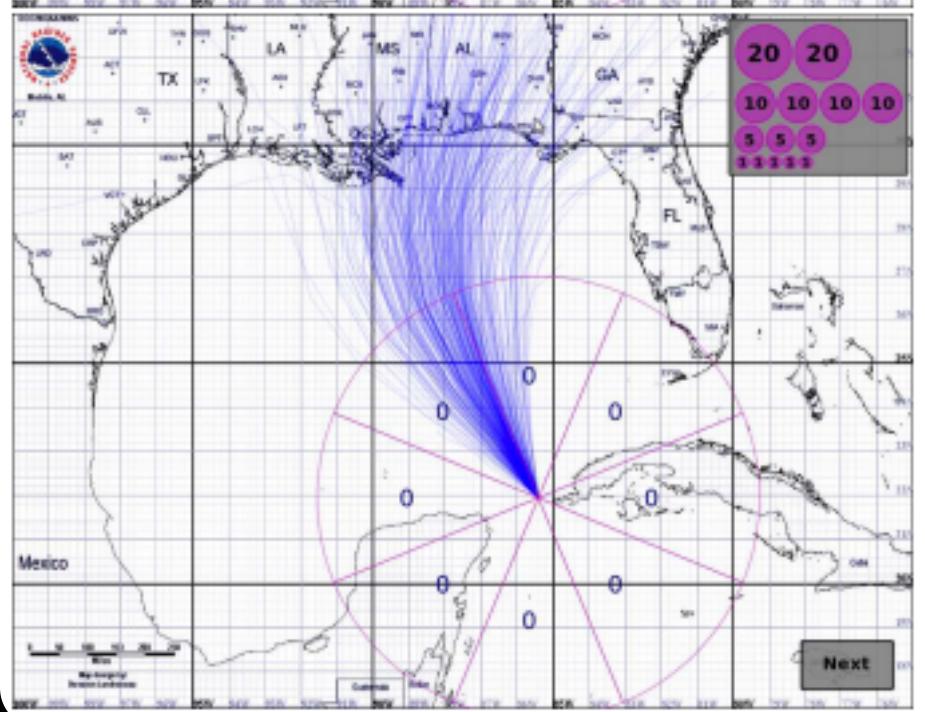
## Zephoria

User ID	21721
Friends	266
Age	??
Gender	Female
Status	Single
Location	San Francisco, CA
Hometown	Lancaster, PA
Occupation	researcher: social networks, identity, context
Interests	apophenia, observing people, culture, questioning power, reading, buddhism, ipseity, computer-mediated communication, social networks, technology, anthropology, stomping, psytrance/goa/trance [Infected Mushroom, Son Kite... Iboga/Digital Structures], Ani Difranco, downtempo, Thievery Corporation, Beth Orton, Morcheeba, Ween, White Stripes
Music	Authors: Erving Goffman, Stanley Milgram, Jeanette Winterson, Eric Schlosser, Leslie Feinberg, Dorothy Allison, Italo Calvino, Hermann Hesse
Books	??
TV Shows	Koyaanisqatsi, Amelie, Waking Life, Tank Girl, The Matrix, Clockwork Orange, American Beauty, Fight Club, Boys Don't Cry
Movies	??
Member Since	2003-10-21
Last Login	2003-10-21
Last Updated	2003-10-21
About	[Some know me as danah...]
	I'm a geek, an activist and an academic, fascinated by people and society. I see life as a very large playground and enjoy exploring its intricacies. I revel in life's chaos, while simultaneously providing my own insane element.
	My musings: <a href="http://www.zephoria.org/thoughts/">http://www.zephoria.org/thoughts/</a>
Want to Meet	Someone who makes life's complexities seem simply elegant. A partner in crime with an

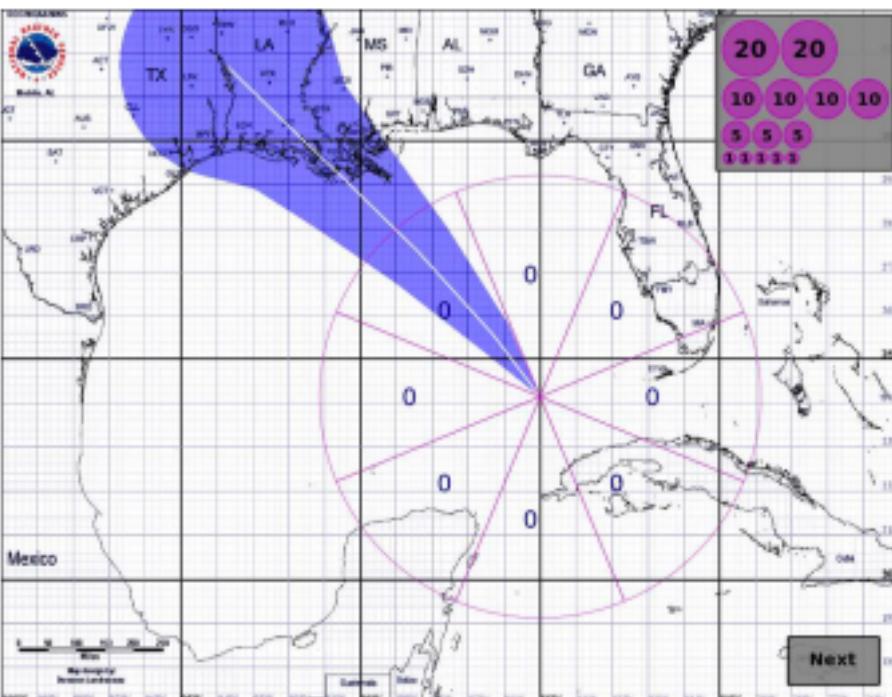
# Uncertainty



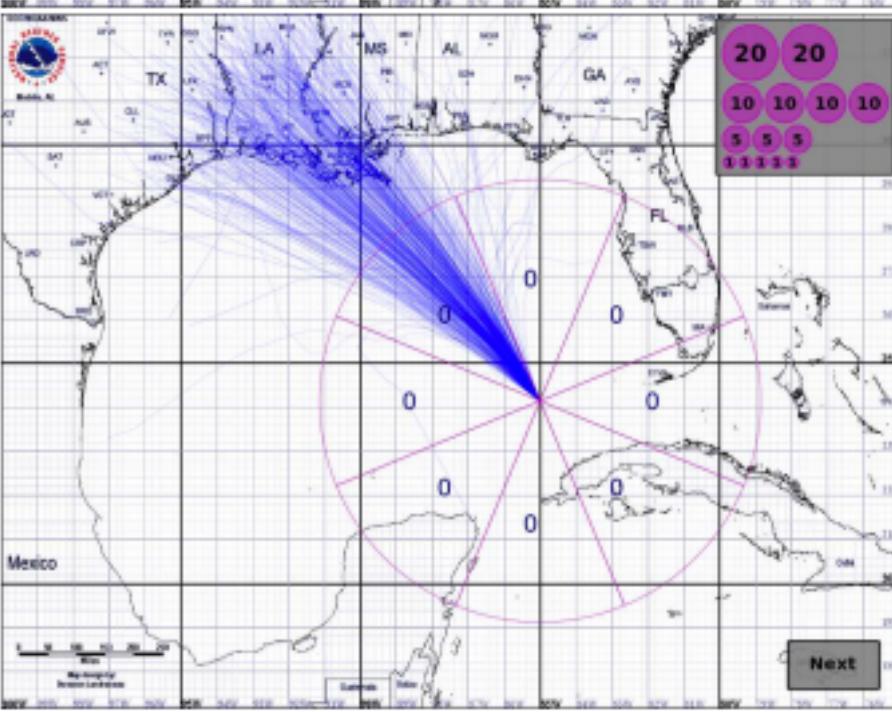
Next



Next



Next



Next

# 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?

# Thank You!

