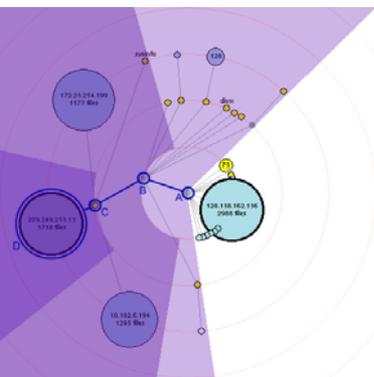


# Information Visualization: Principles, Promise, and Pragmatics

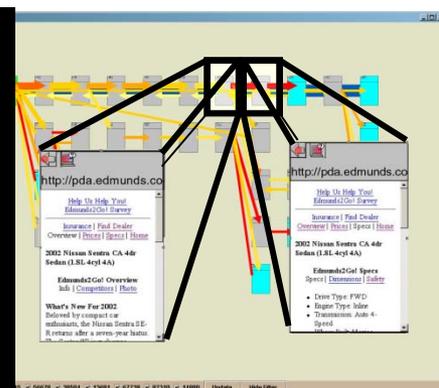
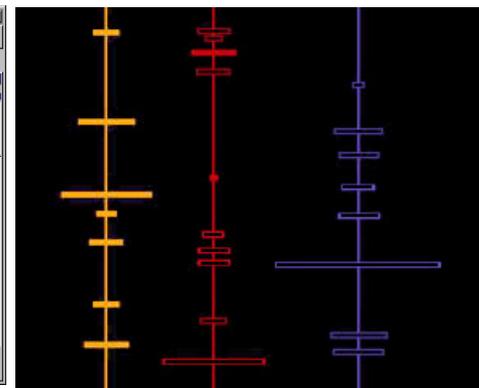
Marti Hearst

CHI 2003 Tutorial



A screenshot of a data visualization tool interface. The window title is "datazoom task\_o\_task77o". The interface includes a "Task Data" tab, "Task Parameters", and "Data Zoom" sections. It displays 761 records with 10 attributes. The attributes are listed on the left, and their values are shown in a grid format. The attributes are: ID, Sex, Age, Income, SocialStatus, Region, LifeIP, HomeIP, Car Insurance Policy, and Response. The values are color-coded: green for text, yellow for circles, and red for dots.

ID	Sex	Age	Income	SocialStatus	Region	LifeIP	HomeIP	Car Insurance Policy	Response
17225354 999 1577 864	female	36-37	40	42-4344	48	51-52-53-54	56-57-58	63	
138116163 136 2088 864	male			single	rural				
2982493153 13 3736 864				family	city				
161026191 1291 864									



# Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up

# Introduction

- Goals of Information Visualization
- Case Study: The Journey of the TreeMap
- Key Questions

# What is Information Visualization?

Visualize: to form a mental image or vision of ...

Visualize: to imagine or remember as if actually seeing.

American Heritage dictionary, Concise Oxford dictionary

# What is Information Visualization?

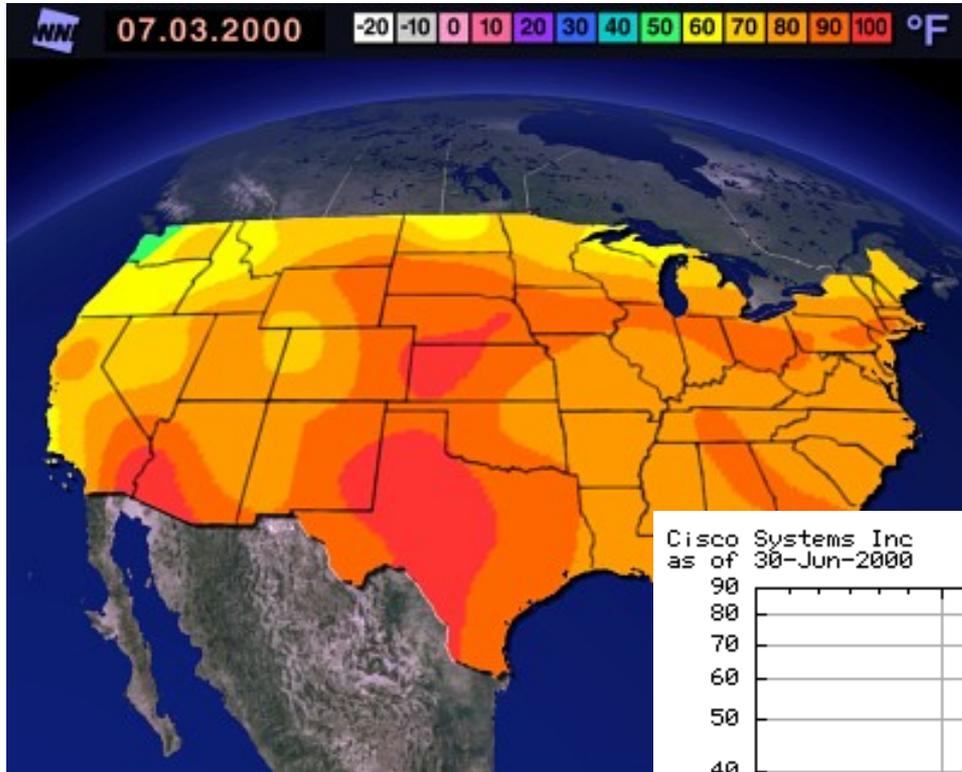
“Transformation of the symbolic into the geometric”  
(McCormick et al., 1987)

“... finding the artificial memory that best supports our natural means of perception.”  
(Bertin, 1983)

The depiction of information using spatial or graphical representations, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system.

# Information Visualization

- **Problem:**
  - HUGE Datasets: How to understand them?
- **Solution**
  - Take better advantage of human perceptual system
  - Convert information into a graphical representation.
- **Issues**
  - How to convert abstract information into graphical form?
  - Do visualizations do a better job than other methods?

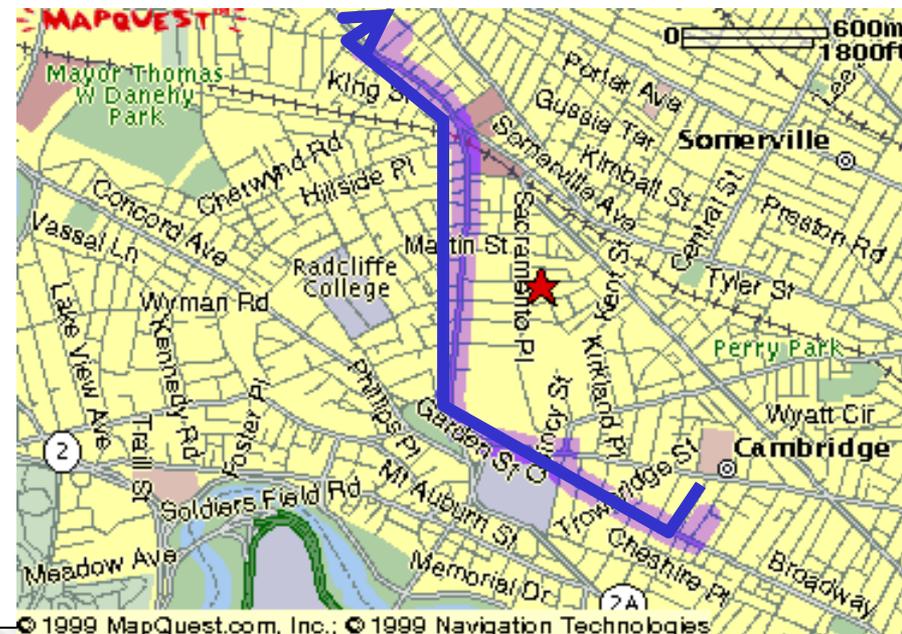


# Visualization Success Stories



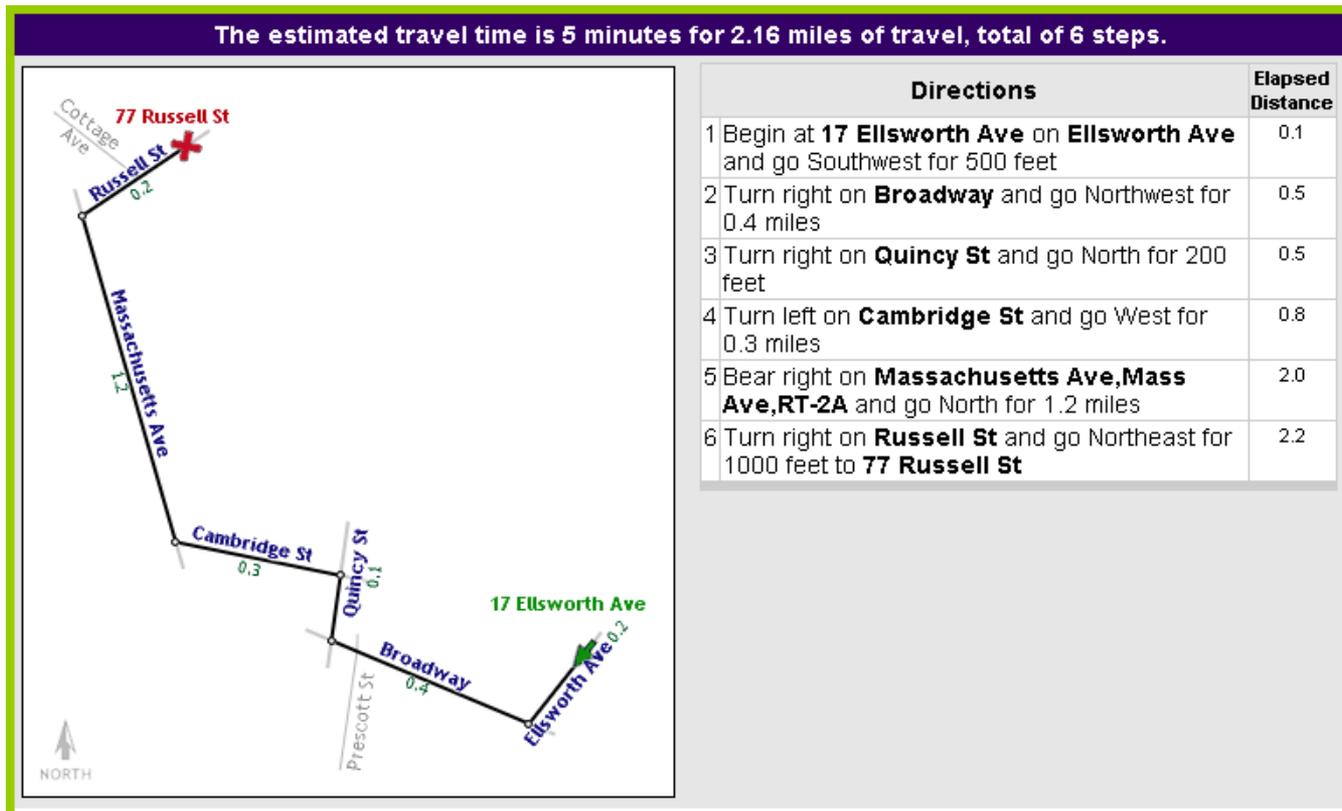
# The Power of Visualization

1. Start out going Southwest on ELLSWORTH AVE  
Towards BROADWAY by turning right.
- 2: Turn RIGHT onto BROADWAY.
3. Turn RIGHT onto QUINCY ST.
4. Turn LEFT onto CAMBRIDGE ST.
5. Turn SLIGHT RIGHT onto MASSACHUSETTS AVE.
6. Turn RIGHT onto RUSSELL ST.



© 1999 MapQuest.com, Inc.; © 1999 Navigation Technologies

# The Power of Visualization



# Visualization Success Story

Mystery: what is causing a cholera epidemic in London in 1854?

# Visualization Success Story



Illustration of John Snow's deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

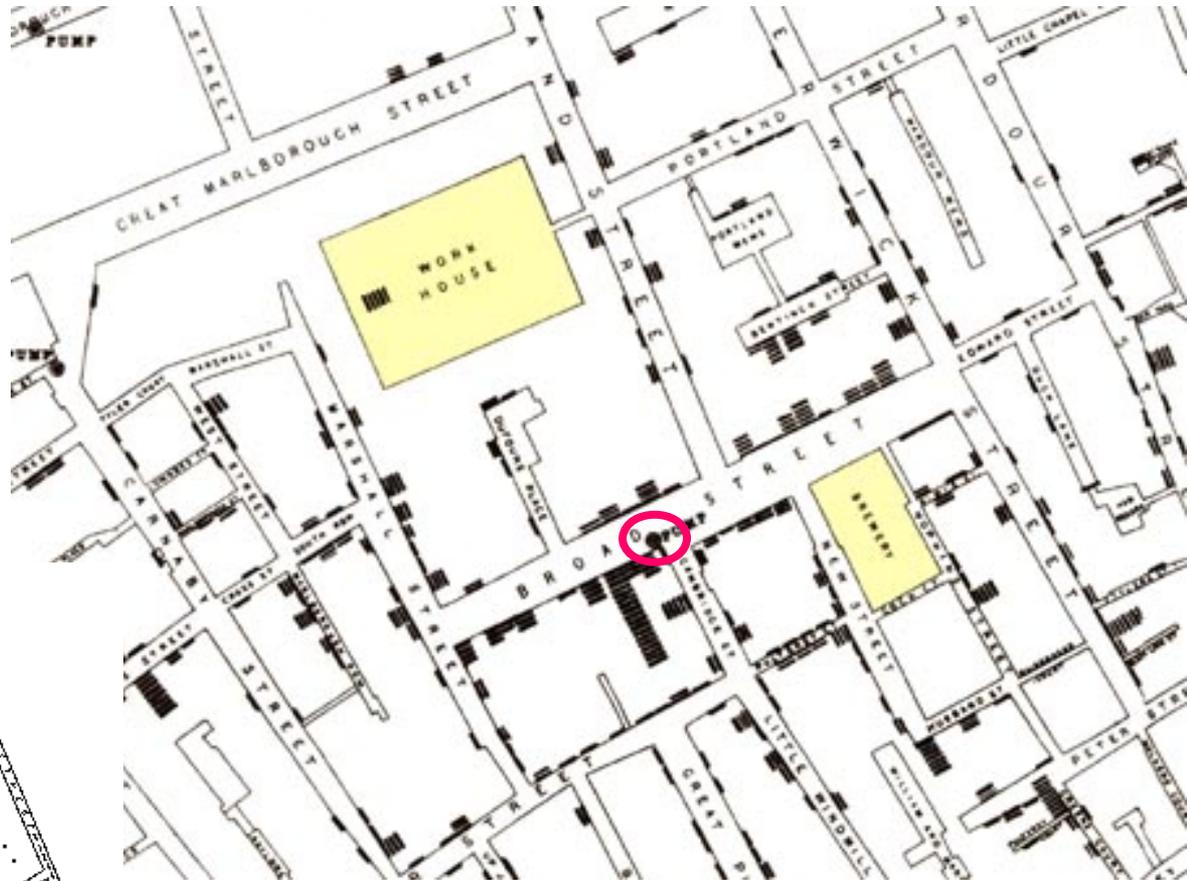
Horizontal lines indicate location of deaths.

From Visual Explanations by Edward Tufte, Graphics Press, 1997

# Visualization Success Story

Illustration of John Snow's deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

Horizontal lines indicate location of deaths.



From Visual Explanations by Edward Tufte, Graphics Press, 1997



# Purposes of Information Visualization

To help:

Explore

Calculate

Communicate

Decorate

# Two Different Primary Goals: Two Different Types of Viz

## **Explore/Calculate**

Analyze

Reason about Information

## **Communicate**

Explain

Make Decisions

Reason about Information

# Goals of Information Visualization

More specifically, visualization should:

- Make large datasets coherent  
(Present huge amounts of information compactly)
- Present information from various viewpoints
- Present information at several levels of detail  
(from overviews to fine structure)
- Support visual comparisons
- Tell stories about the data

# Why Visualization?

Use the eye for pattern recognition; people are good at  
scanning  
recognizing  
remembering images

Graphical elements facilitate comparisons via  
length  
shape  
orientation  
texture

Animation shows changes across time

**Color** helps make distinctions

Aesthetics make the process appealing

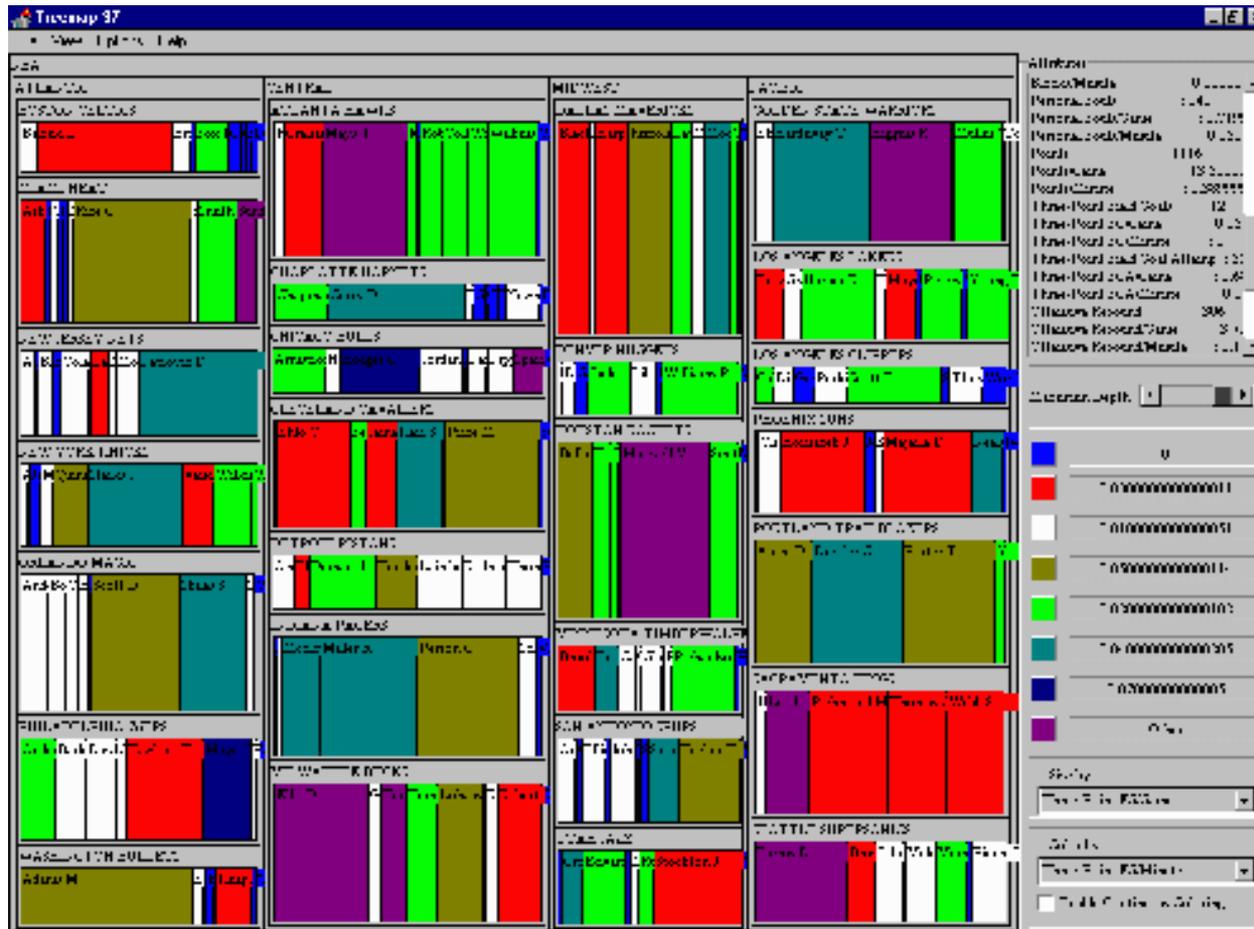
# The Need for Critical Analysis

- We see many creative ideas, but they often fail in practice
- The hard part: how to apply it judiciously
  - Inventors usually do not accurately predict how their invention will be used
- This tutorial will emphasize
  - Getting past the coolness factor
  - Examining usability studies

# Case Study: The Journey of the TreeMap

- The TreeMap (Johnson & Shneiderman '91)
- Idea:
  - Show a hierarchy as a 2D layout
  - Fill up the space with rectangles representing objects
  - Size on screen indicates relative size of underlying objects.

# Early Treemap Applied to File System



# Treemap Problems

- Too disorderly
  - What does adjacency mean?
  - Aspect ratios uncontrolled leads to lots of skinny boxes that clutter
- Color not used appropriately
  - In fact, is meaningless here
- Wrong application
  - Don't need all this to just see the largest files in the OS

# Successful Application of Treemaps

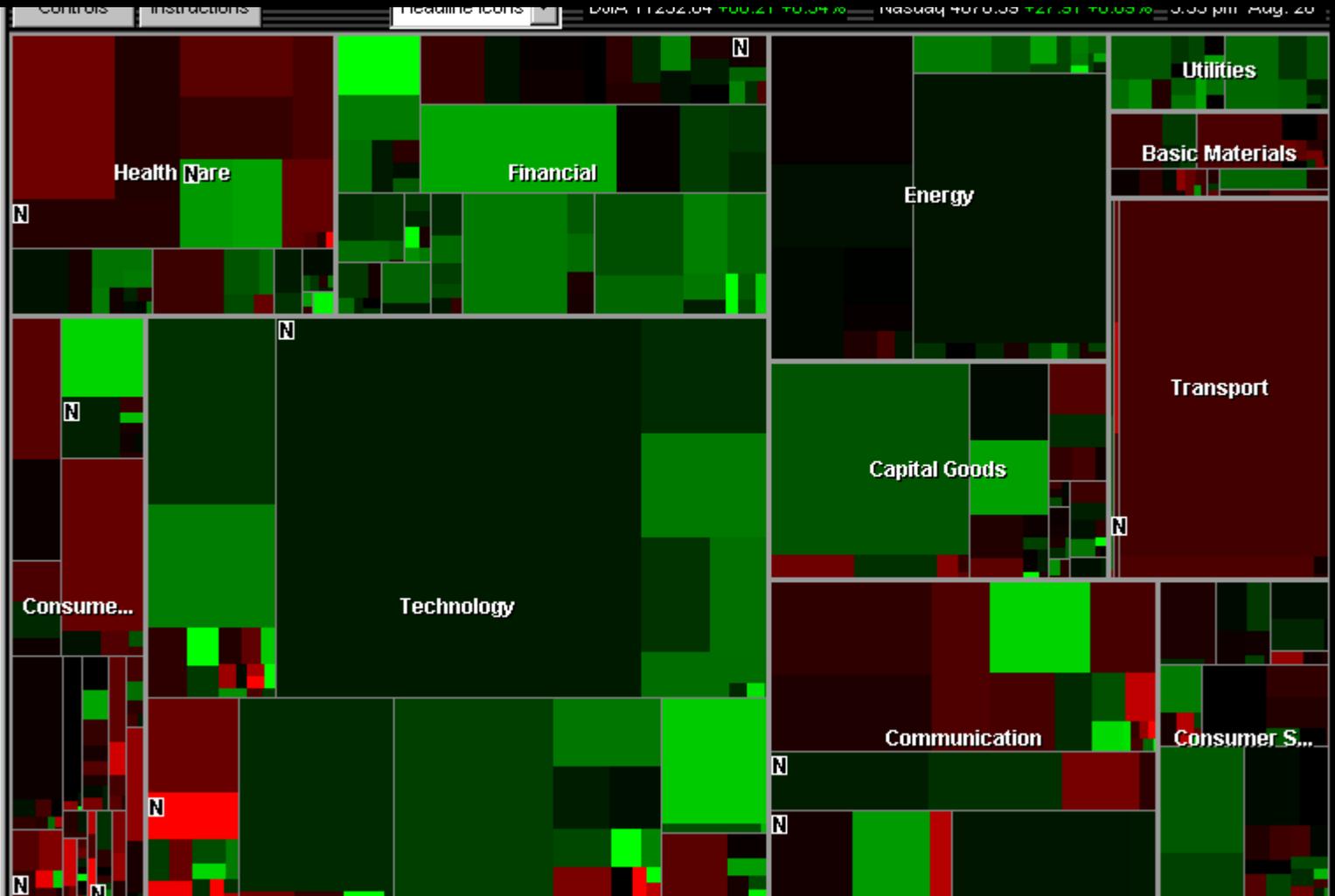
- Think more about the use
  - Break into meaningful groups
  - Fix these into a useful aspect ratio
- Use visual properties properly
  - Use color to distinguish meaningfully
    - Use only two colors:
      - Can then distinguish one thing from another
    - When exact numbers aren't very important
- Provide excellent interactivity
  - Access to the real data
  - Makes it into a useful tool

# TreeMaps in Action

<http://www.smartmoney.com/maps>

[http://www.peets.com/tast/11/coffee\\_selector.asp](http://www.peets.com/tast/11/coffee_selector.asp)

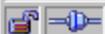
# A Good Use of TreeMaps and Interactivity



[www.smartmoney.com/marketmap](http://www.smartmoney.com/marketmap)

SmartMoney.com © 1999 SmartMoney.com

SmartMoney.com, Inc. All Rights Reserved



Document: Done



# Treemaps in Peets site

SHOP | ROASTING | FRESHNESS | TASTING | ABOUT US

COFFEE TASTING | COFFEE BREWING | TEA TASTING | TEA BREWING

**COFFEE SELECTOR**

CLICK HERE FOR HELP.

Peet's Blends

Decaffeinated

Africa and Arabia

The Pacific

The Americas

Dark Roast

PEET'S GUARANTEED FRESH

This treemap visualizes the product hierarchy. The root node is 'COFFEE SELECTOR', which branches into 'Peet's Blends' and 'Decaffeinated'. 'Peet's Blends' further divides into 'Africa and Arabia', 'The Pacific', and 'The Americas'. 'Decaffeinated' also divides into 'Africa and Arabia', 'The Pacific', and 'The Americas'. 'The Americas' further divides into 'Dark Roast' and another unlabeled sub-category.

Peet's Blends

Decaffeinated

**Gaia Organic Blend™**  
Good body and an earthy, nutty flavor. This certified organically grown coffee is a lively, rich blend of Indonesian and Central American coffees with balanced depth of flavor.  
\$12.95 / lb (Click for more)

Africa and Arabia

The Pacific

The Americas

Dark Roast

This treemap is identical to the one on the left, but with a tooltip overlaid on the 'The Americas' node. The tooltip provides detailed information about the 'Gaia Organic Blend', including its price and a link to learn more.

# Analysis vs. Communication

- MarketMap's use of TreeMaps allows for sophisticated analysis
- Peets' use of TreeMaps is more for presentation and communication
- This is a key contrast

# Open Issues

- Does visualization help?
  - The jury is still out
  - Still supplemental at best for text collections
    - A correlation with spatial ability
    - Learning effects: with practice ability on visual display begins to equal that of text
- Does visualization sell?
  - Jury is still out on this one too!
- This is a *hot* area! More ideas will appear!

# Key Questions to Ask about a Viz

1. What does it teach/show/elucidate?
2. What is the key contribution?
3. What are some compelling, useful examples?
4. Could it have been done more simply?
5. Have there been usability studies done?  
What do they show?

# What we are *not* covering

- Scientific visualization
- Statistics
- Cartography (maps)
- Education
- Games
- Computer graphics in general
- Computational geometry

# Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up

# Visual Principles

# Visual Principles

- Types of Graphs
- Pre-attentive Properties
- Relative Expressiveness of Visual Cues
- Visual Illusions
- Tufte's notions
  - Graphical Excellence
  - Data-Ink Ratio Maximization
  - How to Lie with Visualization

# References for Visual Principles

- Kosslyn: Types of Visual Representations
- Lohse et al: How do people perceive common graphic displays
- Bertin, MacKinlay: Perceptual properties and visual features
- Tufte/Wainer: How to mislead with graphs

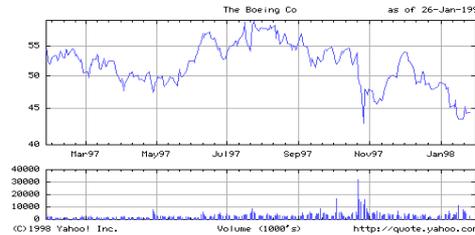
# A Graph is: (Kosslyn)

- A visual display that illustrates one or more relationships among entities
- A shorthand way to present information
- Allows a trend, pattern, or comparison to be easily apprehended

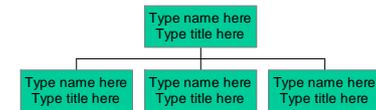
# Types of Symbolic Displays

(Kosslyn 89)

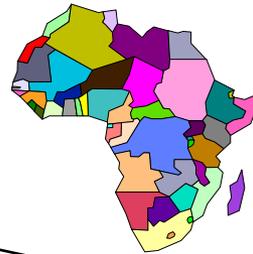
- Graphs



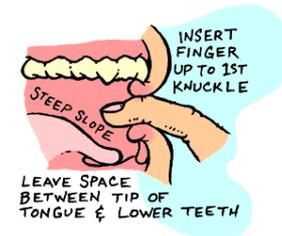
- Charts



- Maps



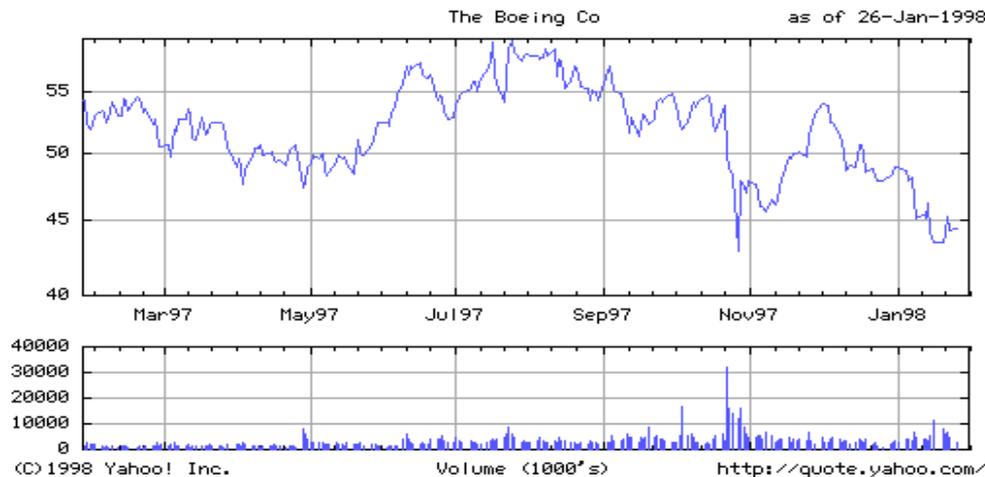
- Diagrams



# Types of Symbolic Displays

- Graphs

- at least two scales required
- values associated by a symmetric “paired with” relation
  - Examples: scatter-plot, bar-chart, layer-graph



# Types of Symbolic Displays

## Charts

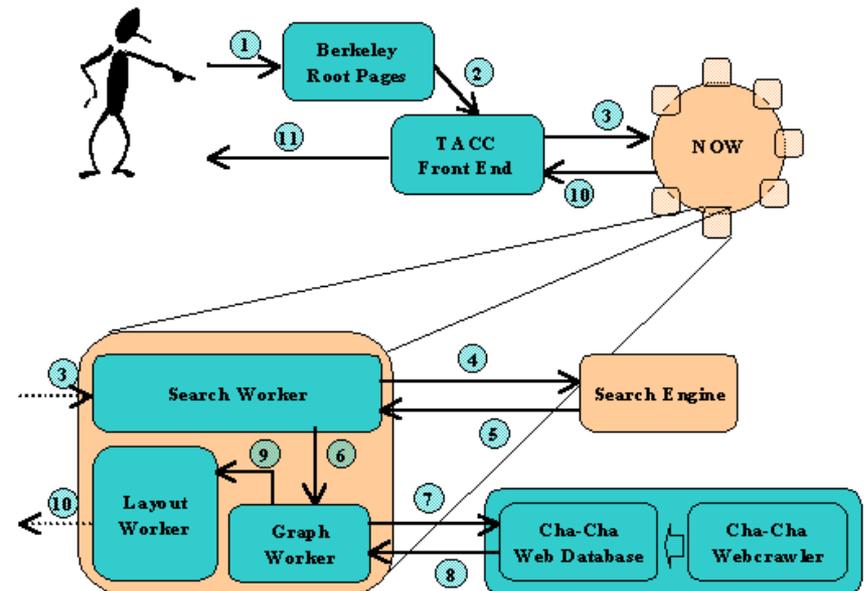
- discrete relations among discrete entities
- structure relates entities to one another
- lines and relative position serve as links

Examples:

family tree

flow chart

network diagram



# Types of Symbolic Displays

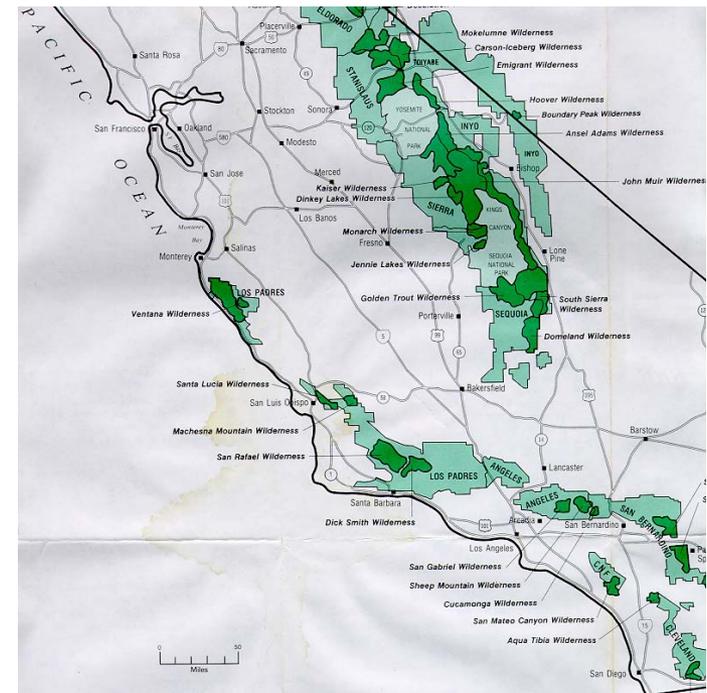
- Maps

- internal relations determined (in part) by the spatial relations of what is pictured
- labels paired with locations

Examples:

map of census data  
topographic maps

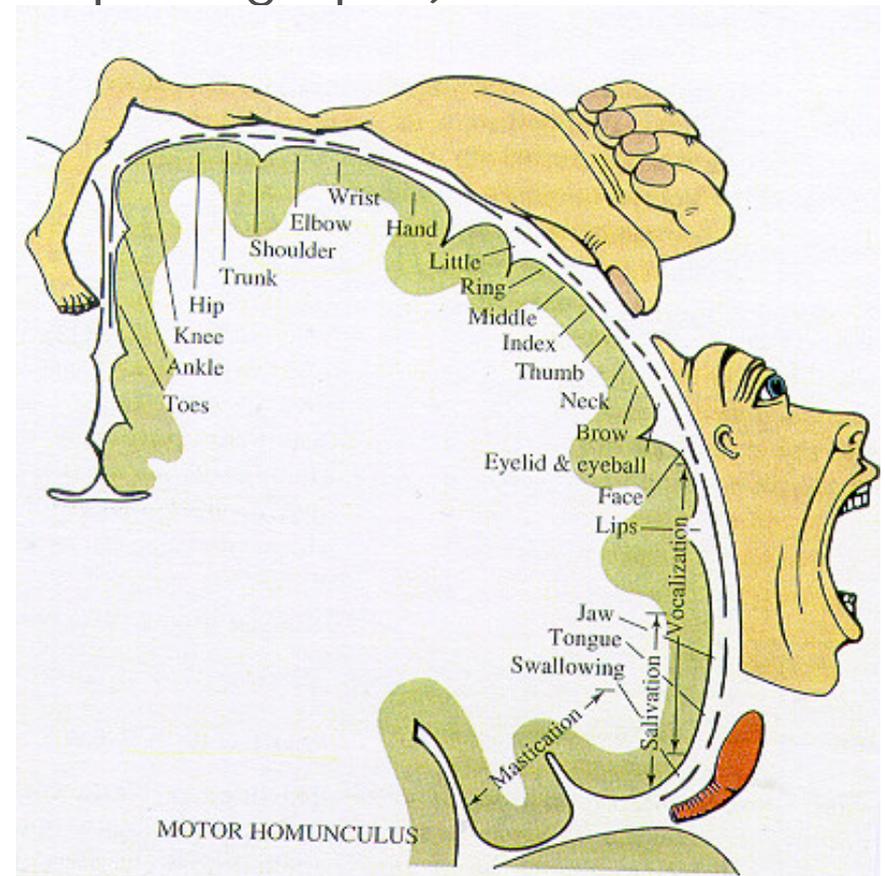
From [www.thehighsierra.com](http://www.thehighsierra.com)



# Types of Symbolic Displays

## Diagrams

- schematic pictures of objects or entities
- parts are symbolic (unlike photographs)
  - how-to illustrations
  - figures in a manual



From Gletman, Henry. Psychology.  
W.W. Norton and Company, Inc.  
New York, 1995

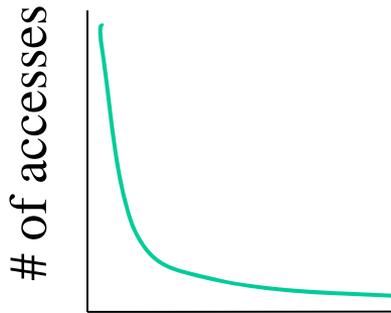
# Anatomy of a Graph (Kosslyn 89)

- Framework
  - sets the stage
  - kinds of measurements, scale, ...
- Content
  - marks
  - point symbols, lines, areas, bars, ...
- Labels
  - title, axes, tic marks, ...

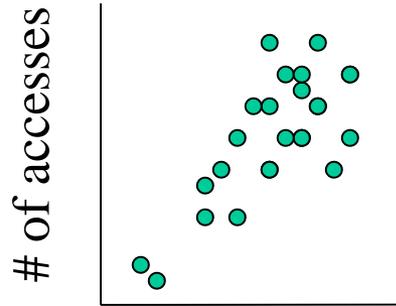
# Basic Types of Data

- **Nominal (qualitative)**
  - (no inherent order)
  - city names, types of diseases, ...
- **Ordinal (qualitative)**
  - (ordered, but not at measurable intervals)
  - first, second, third, ...
  - cold, warm, hot
- **Interval (quantitative)**
  - list of integers or reals

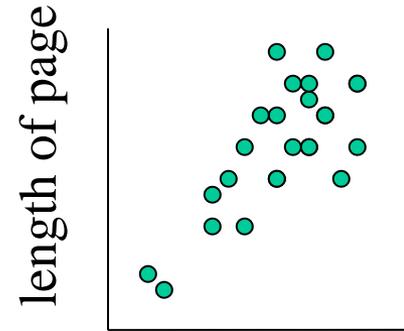
# Common Graph Types



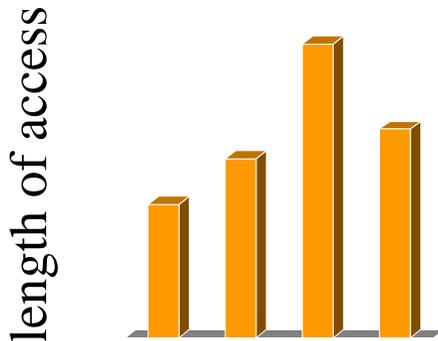
URL



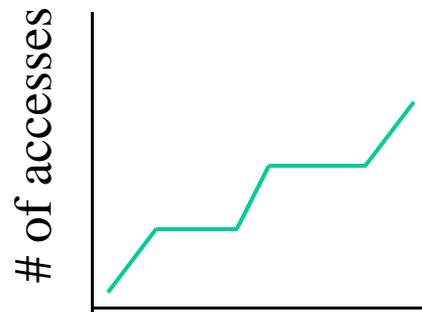
length of access



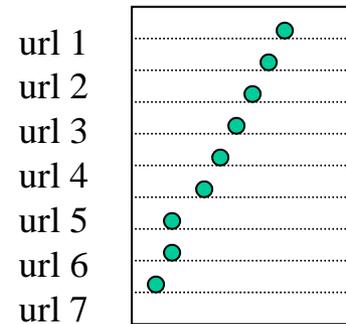
length of access



length of page



days



# of accesses

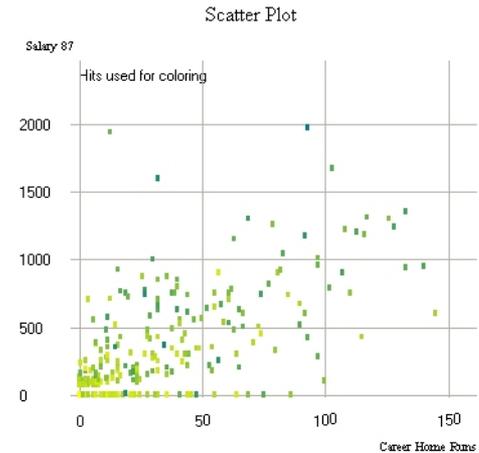
# Combining Data Types in Graphs

Examples?

<b>Nominal</b>	<b>Nominal</b>
<b>Nominal</b>	<b>Ordinal</b>
<b>Nominal</b>	<b>Interval</b>
<b>Ordinal</b>	<b>Ordinal</b>
<b>Ordinal</b>	<b>Interval</b>
<b>Interval</b>	<b>Interval</b>

# Scatter Plots

- Qualitatively determine if variables
  - are highly correlated
    - linear mapping between horizontal & vertical axes
  - have low correlation
    - spherical, rectangular, or irregular distributions
  - have a nonlinear relationship
    - a curvature in the pattern of plotted points
- Place points of interest in context
  - color representing special entities



# When to use which type?

- Line graph
  - x-axis requires quantitative variable
  - Variables have contiguous values
  - familiar/conventional ordering among ordinals
- Bar graph
  - comparison of relative point values
- Scatter plot
  - convey overall impression of relationship between two variables
- Pie Chart?
  - Emphasizing differences in proportion among a few numbers

# Classifying Visual Representations

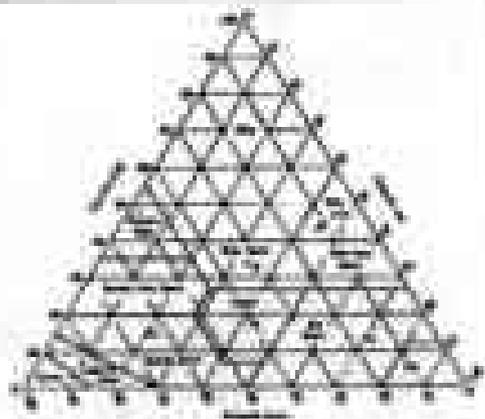
Lohse, G L; Biolsi, K; Walker, N and H H Rueter,  
A Classification of Visual Representations  
CACM, Vol. 37, No. 12, pp 36-49, 1994

Participants sorted 60 items into categories

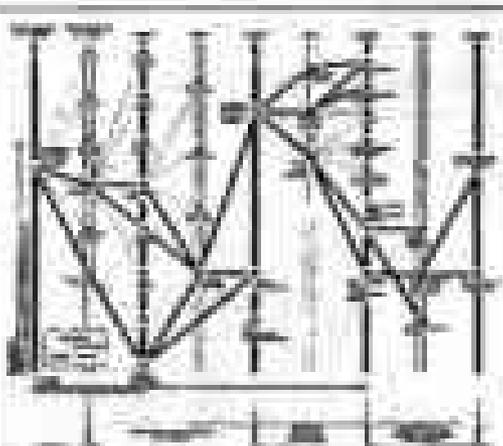
Other participants assigned labels from Likert scales

Experimenters clustered the results various ways.

# Subset of Example Visual Representations From Lohse et al. 94



1. soil triangle



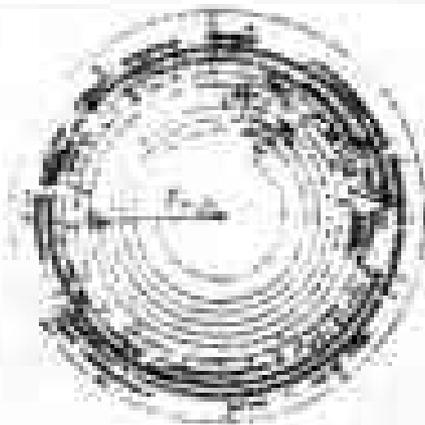
2. missile crisis



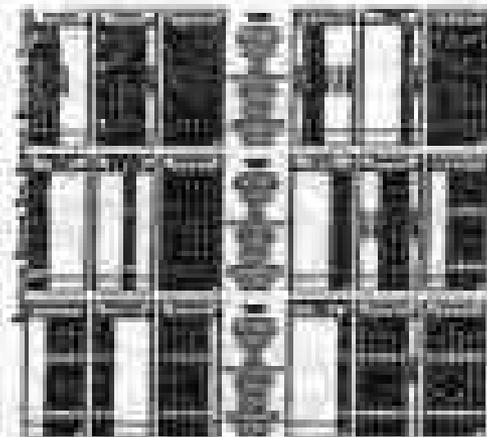
3. USA tomatoes



5. microscope



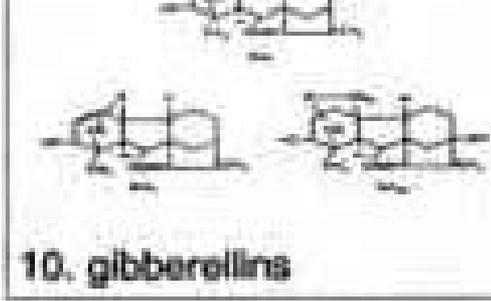
6. circular tree diagram



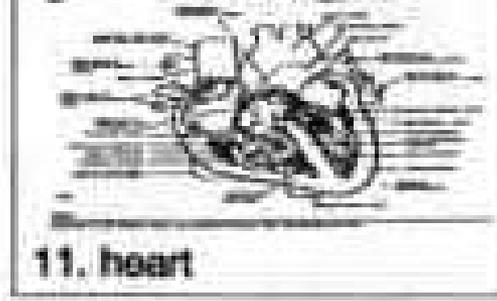
7. auto repair records



Lake



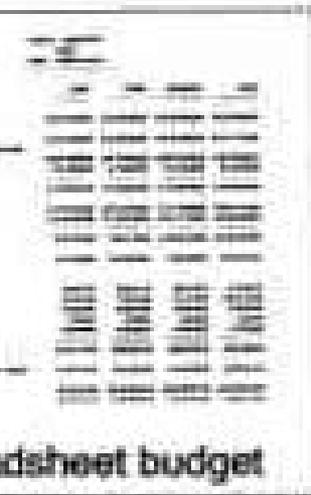
10. gibberellins



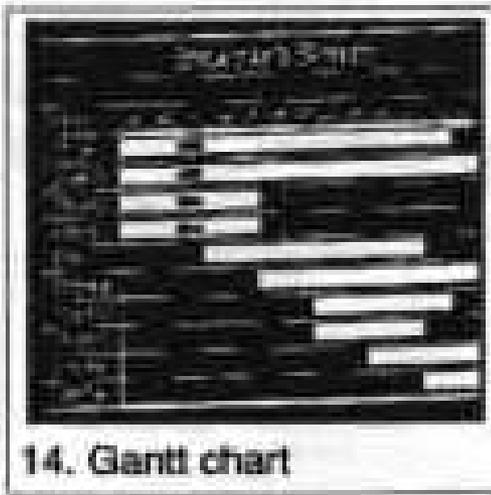
11. heart



12. microbiology



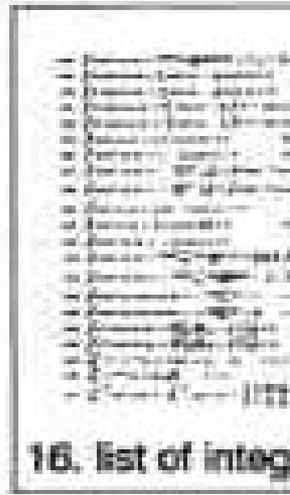
spreadsheet budget



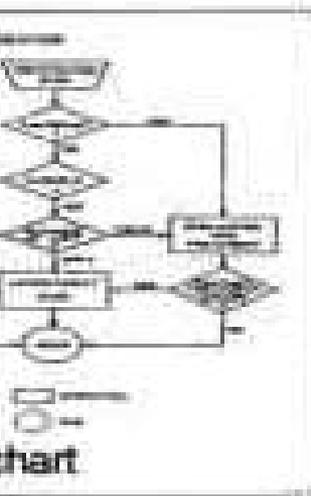
14. Gantt chart



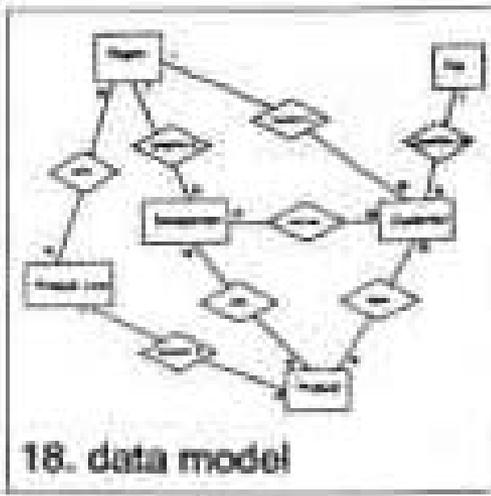
15. Pittsburgh map



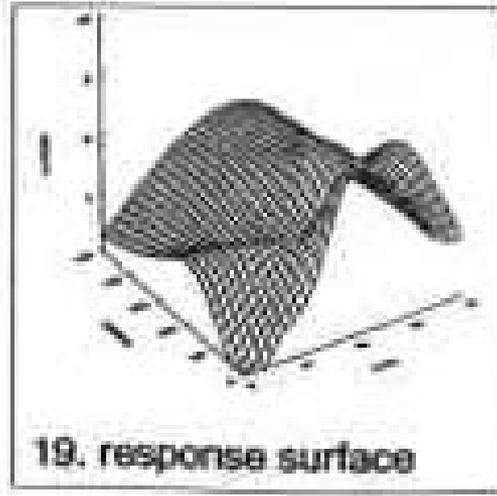
16. list of integers



flowchart



18. data model



19. response surface



20. wheelbarrow

# Likert Scales

(and percentage of variance explained)

16.0 emphasizes whole – parts

11.3 spatial – nonspatial

10.6 static structure – dynamic structure

10.5 continuous – discrete

10.3 attractive – unattractive

10.1 nontemporal – temporal

9.9 concrete – abstract

9.6 hard to understand – easy

9.5 nonnumeric – numeric

2.2 conveys a lot of info – conveys little

# Experimentally Motivated Classification (Lohse et al. 94)

- Graphs
- Tables (numerical)
- Tables (graphical)
- Charts (time)
- Charts (network)
- Diagrams (structure)
- Diagrams (network)
- Maps
- Cartograms
- Icons
- Pictures

# Interesting Findings

Lohse et al. 94

- Photorealistic images were least informative
  - Echoes results in icon studies – better to use less complex, more schematic images
- Graphs and tables are the most self-similar categories
  - Results in the literature comparing these are inconclusive
- Cartograms were hard to understand
  - Echoes other results – better to put points into a framed rectangle to aid spatial perception
- Temporal data more difficult to show than cyclic data
  - Recommend using animation for temporal data

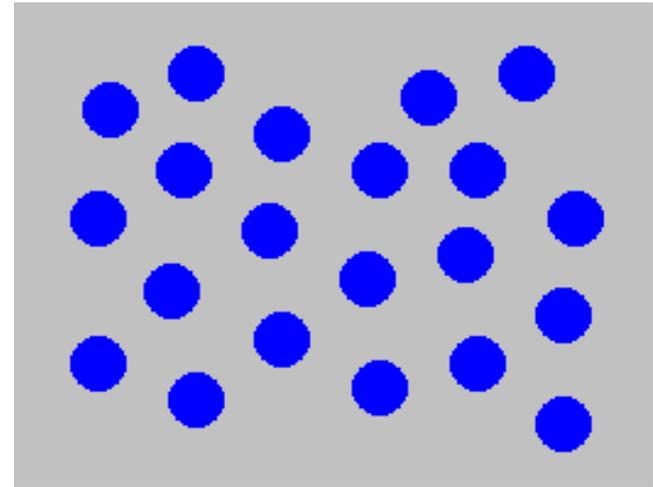
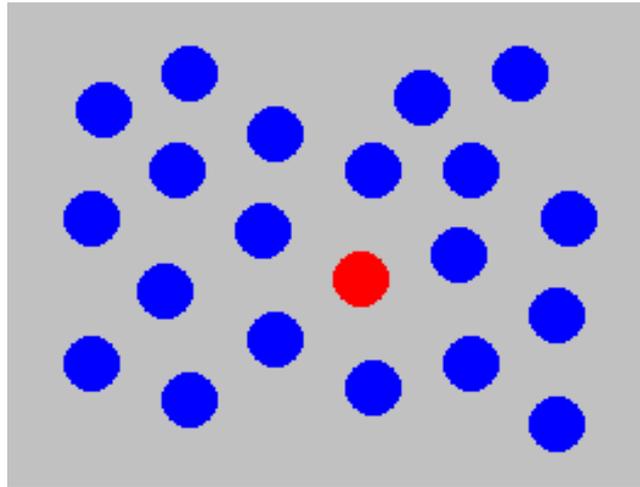
# Visual Properties

- Preattentive Processing
- Accuracy of Interpretation of Visual Properties
- Illusions and the Relation to Graphical Integrity

# Preattentive Processing

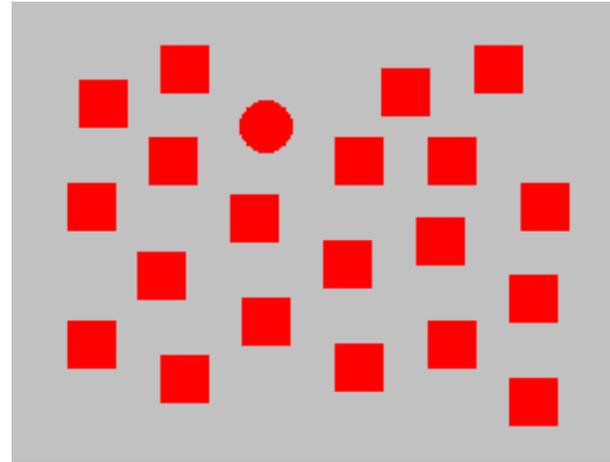
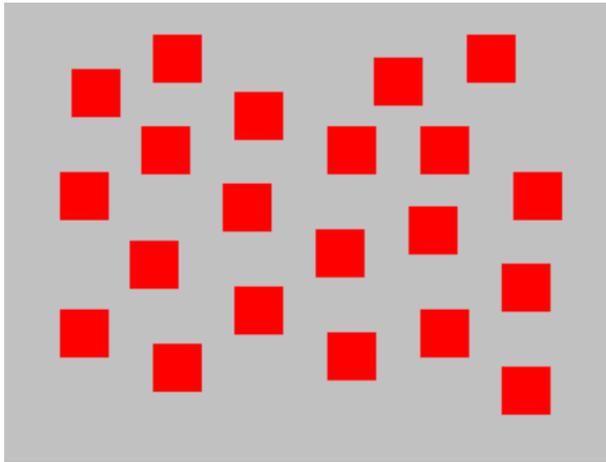
- A limited set of visual properties are processed preattentively
  - (without need for focusing attention).
- This is important for design of visualizations
  - what can be perceived immediately
  - what properties are good discriminators
  - what can mislead viewers

# Example: Color Selection



Viewer can rapidly and accurately determine whether the target (red circle) is present or absent. Difference detected in color.

# Example: Shape Selection

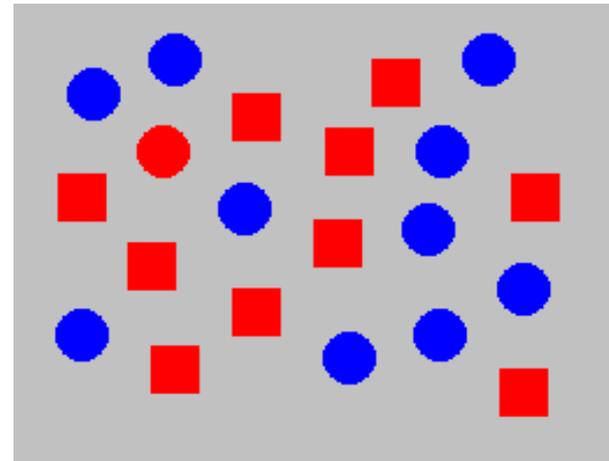
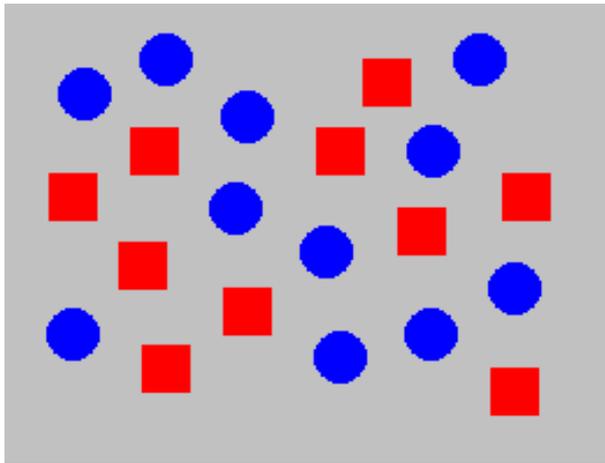


Viewer can rapidly and accurately determine whether the target (red circle) is present or absent. Difference detected in form (curvature)

# Pre-attentive Processing

- < 200 - 250ms qualifies as pre-attentive
  - eye movements take at least 200ms
  - yet certain processing can be done very quickly, implying low-level processing in parallel
- If a decision takes a fixed amount of time regardless of the number of distractors, it is considered to be preattentive.

# Example: Conjunction of Features

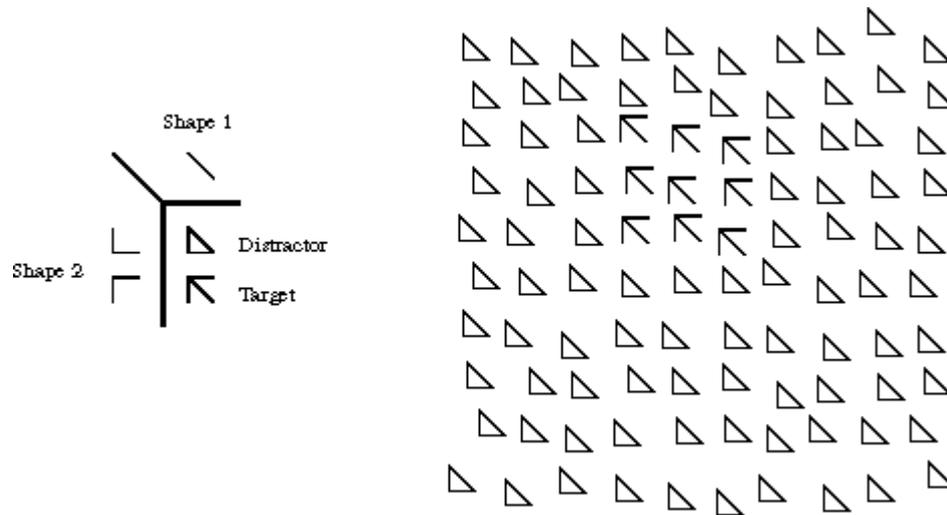


Viewer *cannot* rapidly and accurately determine whether the target (red circle) is present or absent when target has two or more features, each of which are present in the distractors. Viewer must search sequentially.

[All Preattentive Processing figures from Healey 97](http://www.csc.ncsu.edu/faculty/healey/PP/PP.html)

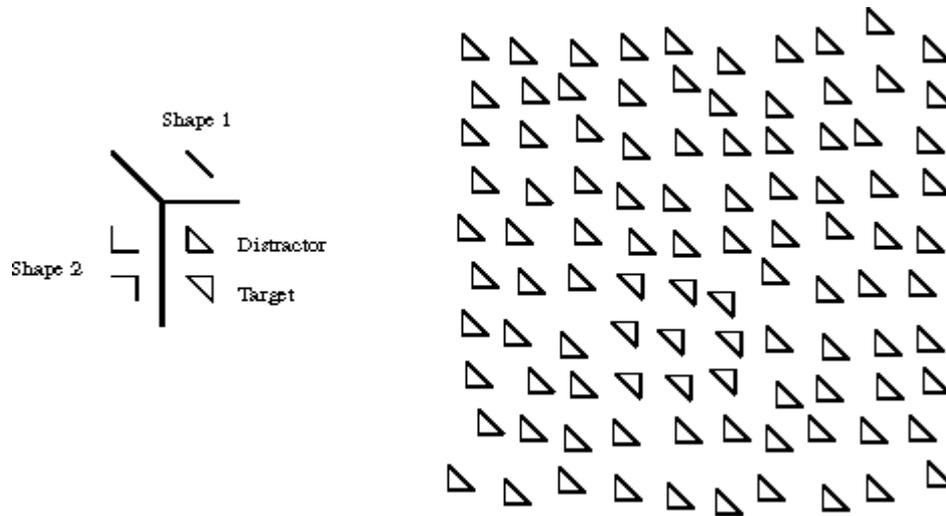
<http://www.csc.ncsu.edu/faculty/healey/PP/PP.html>

# Example: Emergent Features



Target has a unique feature with respect to distractors (open sides) and so the group can be detected preattentively.

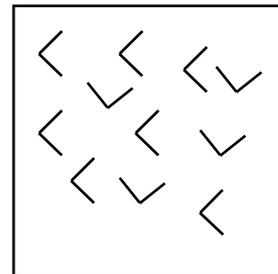
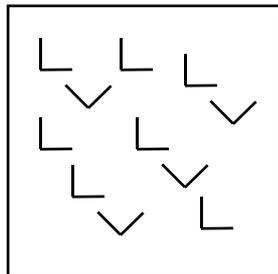
# Example: Emergent Features



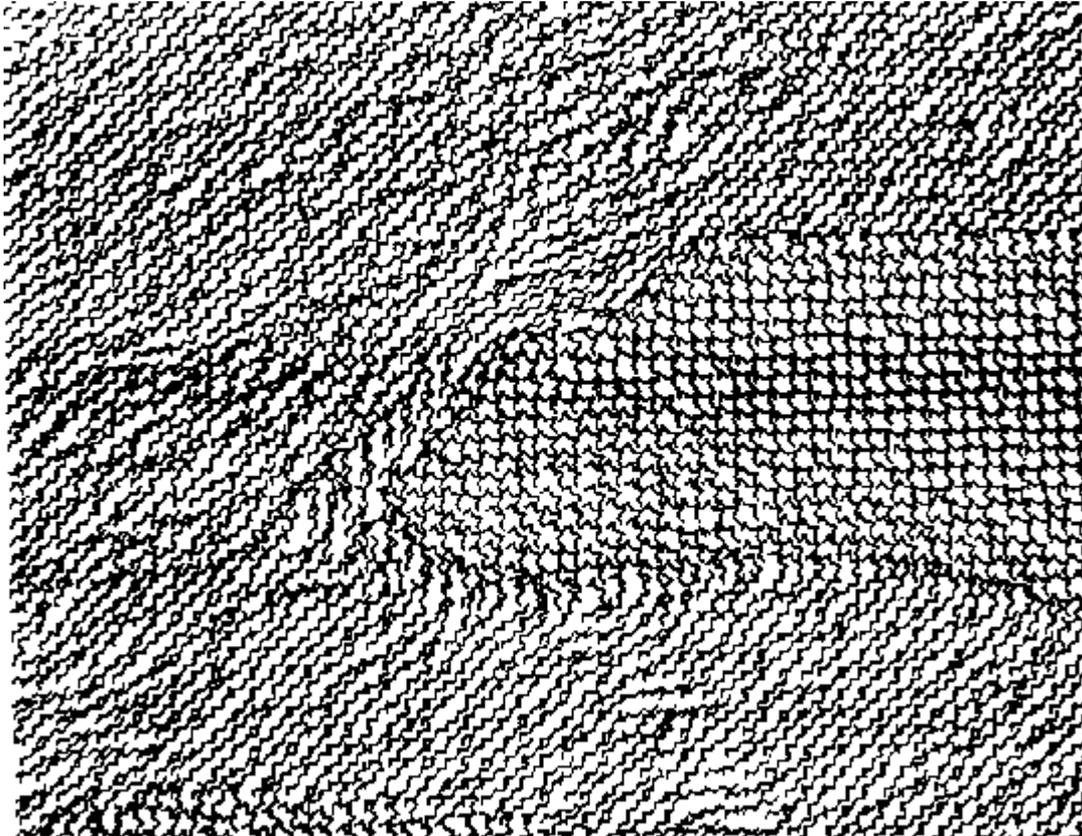
Target does not have a unique feature with respect to distractors and so the group cannot be detected preattentively.

# Asymmetric and Graded Preattentive Properties

- Some properties are asymmetric
  - a sloped line among vertical lines is preattentive
  - a vertical line among sloped ones is not
- Some properties have a gradation
  - some more easily discriminated among than others



# Use Grouping of Well-Chosen Shapes for Displaying Multivariate Data



SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO  
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM  
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC  
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM  
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM  
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM  
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC  
SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO  
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM  
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC

# Text NOT Preattentive

SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO  
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM  
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC  
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM  
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM  
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM  
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC  
SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO  
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM  
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC

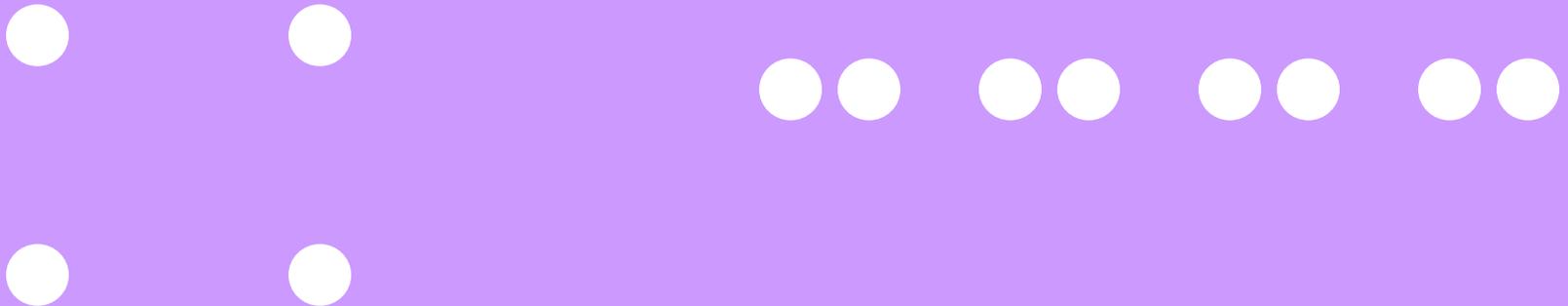
# Preattentive Visual Properties

(Healey 97)

length	Triesman & Gormican [1988]
width	Julesz [1985]
size	Triesman & Gelade [1980]
curvature	Triesman & Gormican [1988]
number	Julesz [1985]; Trick & Pylyshyn [1994]
terminators	Julesz & Bergen [1983]
intersection	Julesz & Bergen [1983]
closure	Enns [1986]; Triesman & Souther [1985]
colour (hue)	Nagy & Sanchez [1990, 1992]; D'Zmura [1991] Kawai et al. [1995]; Bauer et al. [1996]
intensity	Beck et al. [1983]; Triesman & Gormican [1988]
flicker	Julesz [1971]
direction of motion	Nakayama & Silverman [1986]; Driver & McLeod [1992]
binocular lustre	Wolfe & Franzel [1988]
stereoscopic depth	Nakayama & Silverman [1986]
3-D depth cues	Enns [1990]
lighting direction	Enns [1990]

# Gestalt Properties

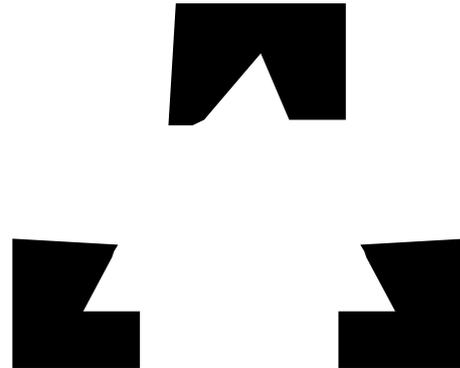
- *Gestalt*: form or configuration
- Idea: forms or patterns transcend the stimuli used to create them.
  - Why do patterns emerge?
  - Under what circumstances?



Why perceive pairs vs. triplets?

# Gestalt Laws of Perceptual Organization (Kaufman 74)

- Figure and Ground
  - Escher illustrations are good examples
  - Vase/Face contrast
- Subjective Contour

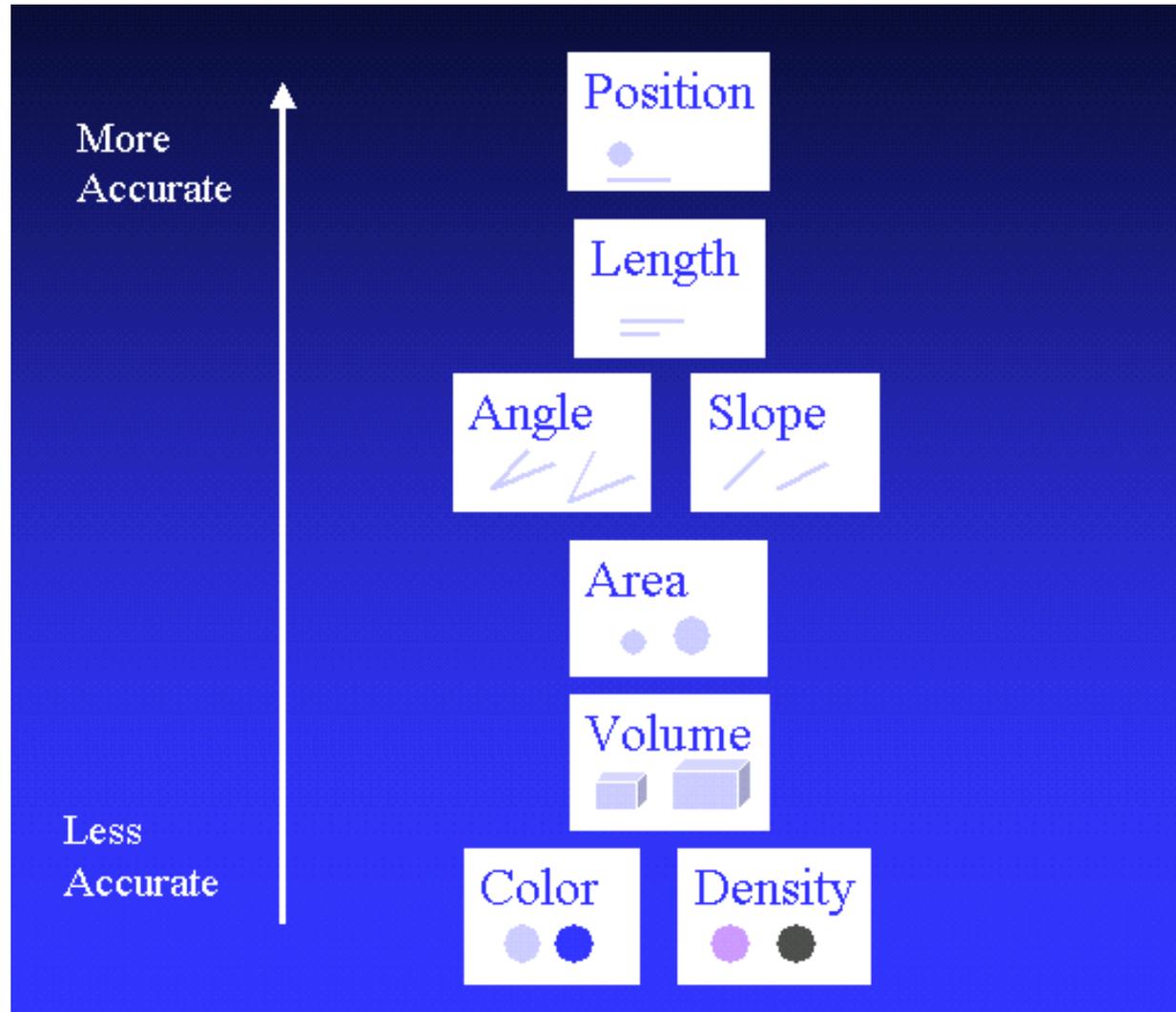


# More Gestalt Laws

- Law of Proximity
  - Stimulus elements that are close together will be perceived as a group
- Law of Similarity
  - like the preattentive processing examples
- Law of Common Fate
  - like preattentive motion property
    - move a subset of objects among similar ones and they will be perceived as a group

Which Properties are  
Appropriate for Which  
Information Types?

Accuracy Ranking of Quantitative Perceptual Tasks  
Estimated; only pairwise comparisons have been validated  
(Mackinlay 88 from Cleveland & McGill)



# Interpretations of Visual Properties

Some properties can be discriminated more accurately but don't have intrinsic meaning

(Senay & Ingatious 97, Kosslyn, others)

- Density (Greyscale)  
Darker -> More
- Size / Length / Area  
Larger -> More
- Position  
Leftmost -> first, Topmost -> first
- Hue  
??? no intrinsic meaning
- Slope  
??? no intrinsic meaning

# Ranking of Applicability of Properties for Different Data Types

(Mackinlay 88, Not Empirically Verified)

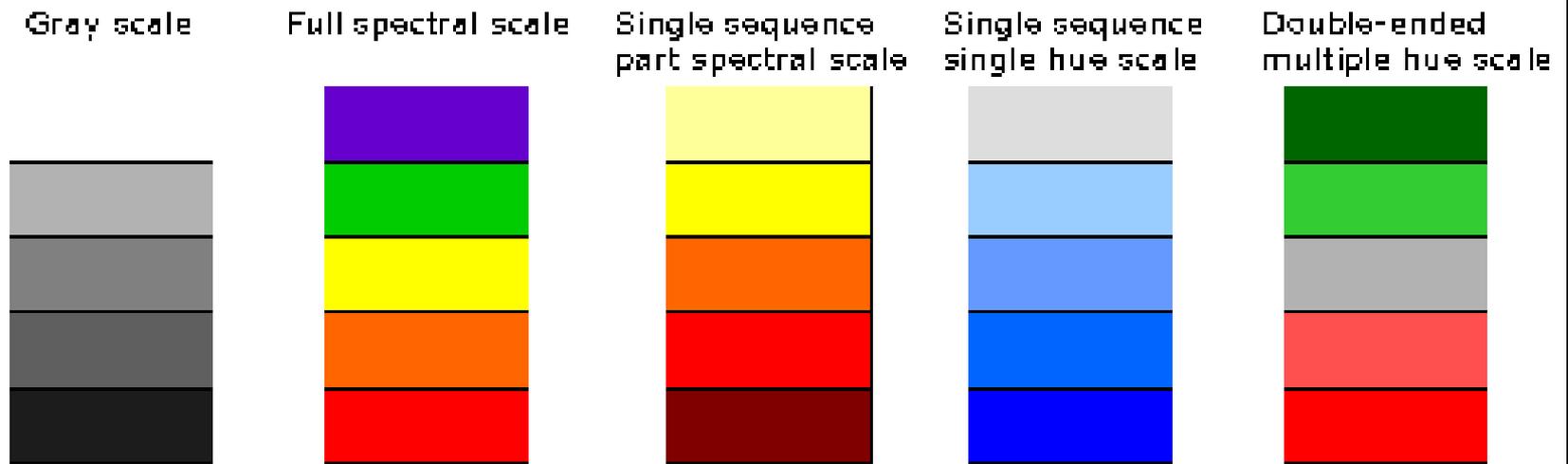
<b><u>QUANTITATIVE</u></b>	<b><u>ORDINAL</u></b>	<b><u>NOMINAL</u></b>
Position	Position	Position
Length	Density	Color Hue
Angle	Color Saturation	Texture
Slope	Color Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Color Saturation
Color Saturation	Length	Shape
Color Hue	Angle	Length

# Color Schemes

Order these (low->hi)



# Color Schemes



# Color Purposes

- Call attention to specific items
- Distinguish between classes of items
  - Increases the number of dimensions for encoding
- Increase the appeal of the visualization

# Using Color

- Proceed with caution
  - Less is more
  - Representing magnitude is tricky
- Examples
  - Red-orange-yellow-white
    - Works for costs
    - Maybe because people are very experienced at reasoning shrewdly according to cost
  - Green-light green-light brown-dark brown-grey-white works for atlases
  - Grayscale is unambiguous but has limited range

# Visual Illusions

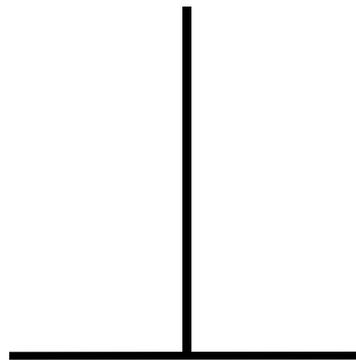
- People don't perceive length, area, angle, brightness the way they "should".
- Some illusions have been reclassified as systematic perceptual errors
  - e.g., brightness contrasts (grey square on white background vs. on black background)
  - partly due to increase in our understanding of the relevant parts of the visual system
- Nevertheless, the visual system does some really unexpected things.

# Illusions of Linear Extent

- Mueller-Lyon (off by 25-30%)

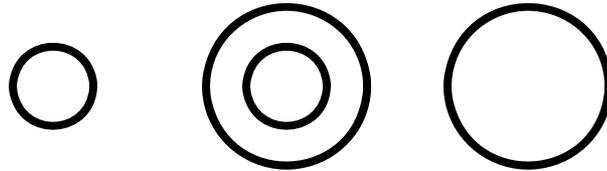


- Horizontal-Vertical



# Illusions of Area

- Delboeuf Illusion



- Height of 4-story building overestimated by approximately 25%

# What are good guidelines for Infoviz?

- Use graphics appropriately
  - Don't use images gratuitously
  - Don't lie with graphics!
    - Link to original data
  - Don't conflate area with other information
    - E.g., use area in map to imply amount
- Make it interactive (feedback)
  - Brushing and linking
  - Multiple views
  - Overview + details
- Match mental models

# Tufte

- Principles of Graphical Excellence
  - Graphical excellence is
    - the well-designed presentation of interesting data – a matter of substance, of statistics, and of design
    - consists of complex ideas communicated with clarity, precision and efficiency
    - is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
    - requires telling the truth about the data.

# Tufte's Notion of Data Ink Maximization

- What is the main idea?
  - draw viewers attention to the substance of the graphic
  - the role of redundancy
  - principles of editing and redesign
- What's wrong with this? What is he really getting at?

# Tufte Principle

Maximize the data-ink ratio:

$$\text{Data-ink ratio} = \frac{\text{data ink}}{\text{total ink used in graphic}}$$

Avoid “chart junk”

# Tufte Principles

- Use multifunctioning graphical elements
- Use small multiples
- Show mechanism, process, dynamics, and causality
- High data density
  - Number of items/area of graphic
  - This is controversial
    - White space thought to contribute to good visual design
    - Tufte's book itself has lots of white space

# Tufte's Graphical Integrity

- Some lapses intentional, some not
- Lie Factor =  $\frac{\text{size of effect in graph}}{\text{size of effect in data}}$
- Misleading uses of area
- Misleading uses of perspective
- Leaving out important context
- Lack of taste and aesthetics

From Tim Craven's LIS 504 course

[http://instruct.uwo.ca/fim-lis/504/504gra.htm#data-ink\\_ratio](http://instruct.uwo.ca/fim-lis/504/504gra.htm#data-ink_ratio)

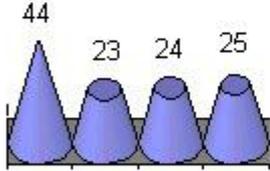
LIS 504 - Graphic displays of data - Netscape

File Edit View Go Communicator Help

A common example of a high lie factor occurs when both dimensions of a two-dimensional figure are made proportional to the same data, so that the size of the figure is proportional to the square of the data; for instance,

Year	Books circulated
2001	100 
2002	141 
2003	200 

An example of a **low** lie factor can be seen in the "Cones" custom chart format in Microsoft Excel.



44	23	24	25
2000	2001	2002	2003

The heights of the (truncated) cones are proportional to the data, but their areas on the screen and their apparent volumes make the larger data values seem relatively small.

Document: Done

# How to Exaggerate with Graphs

from Tufte '83

## THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%



"Lie factor" = 2.8

*Los Angeles Times*, August 5, 1979, p. 3-

# How to Exaggerate with Graphs

from Tufte '83



Washington Post, October 25, 1978, p. 1.

Purchasing Power of the Diminishing Dollar  
Source: Labor Department

# Howard Wainer

## How to Display Data Badly (Video)

<http://www.dartmouth.edu/~chance/ChanceLecture/AudioVideo.html>

# Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up

# Promising Techniques

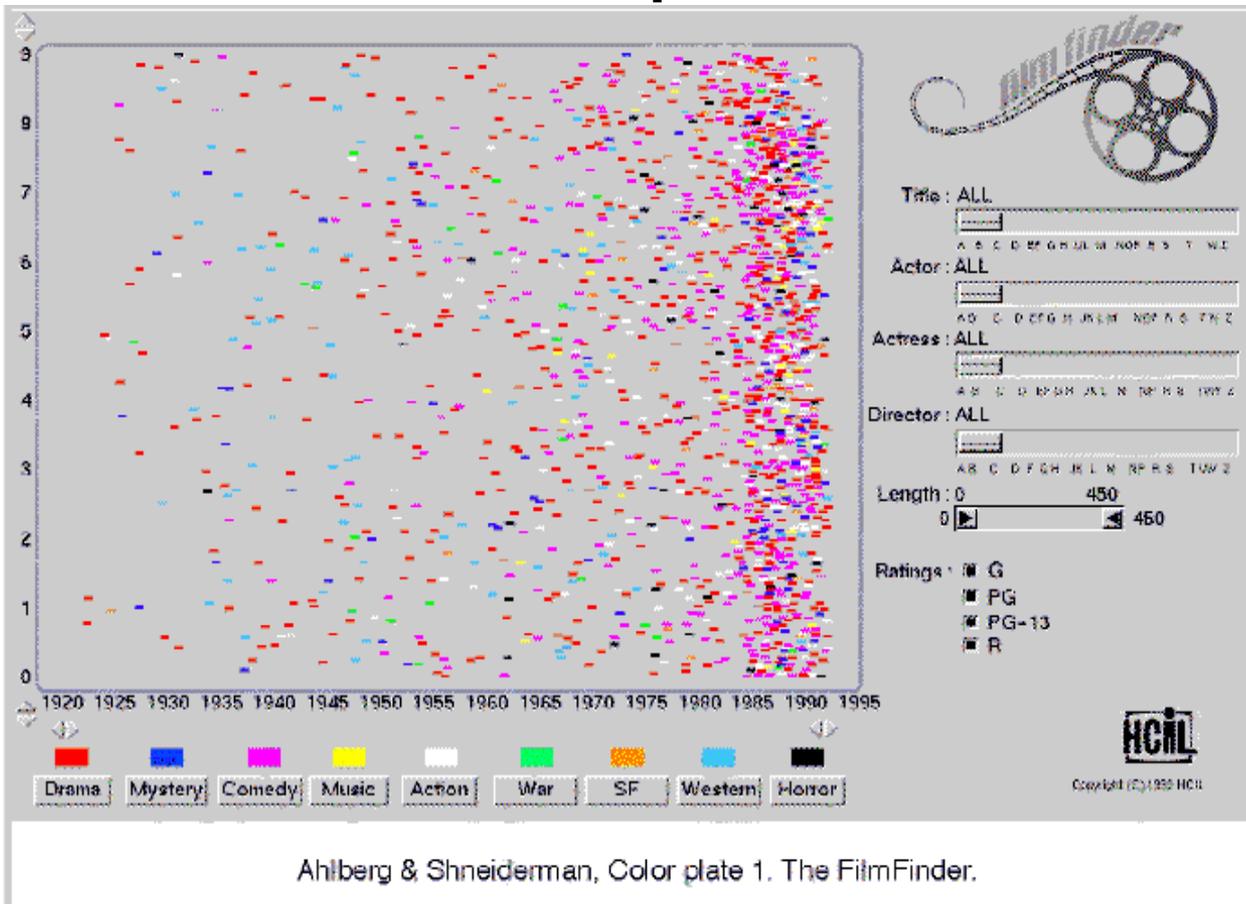
# Promising Techniques & Approaches

- Perceptual Techniques
  - Animation
  - Grouping / Gestalt principles
  - Using size to indicate quantity
  - Color for Accent, Distinction, Selection
    - NOT FOR QUANTITY!!!!
- General Approaches
  - Standard Techniques
    - Graphs, bar charts, tables
  - Brushing and Linking
  - Providing Multiple Views and Models
  - Aesthetics!

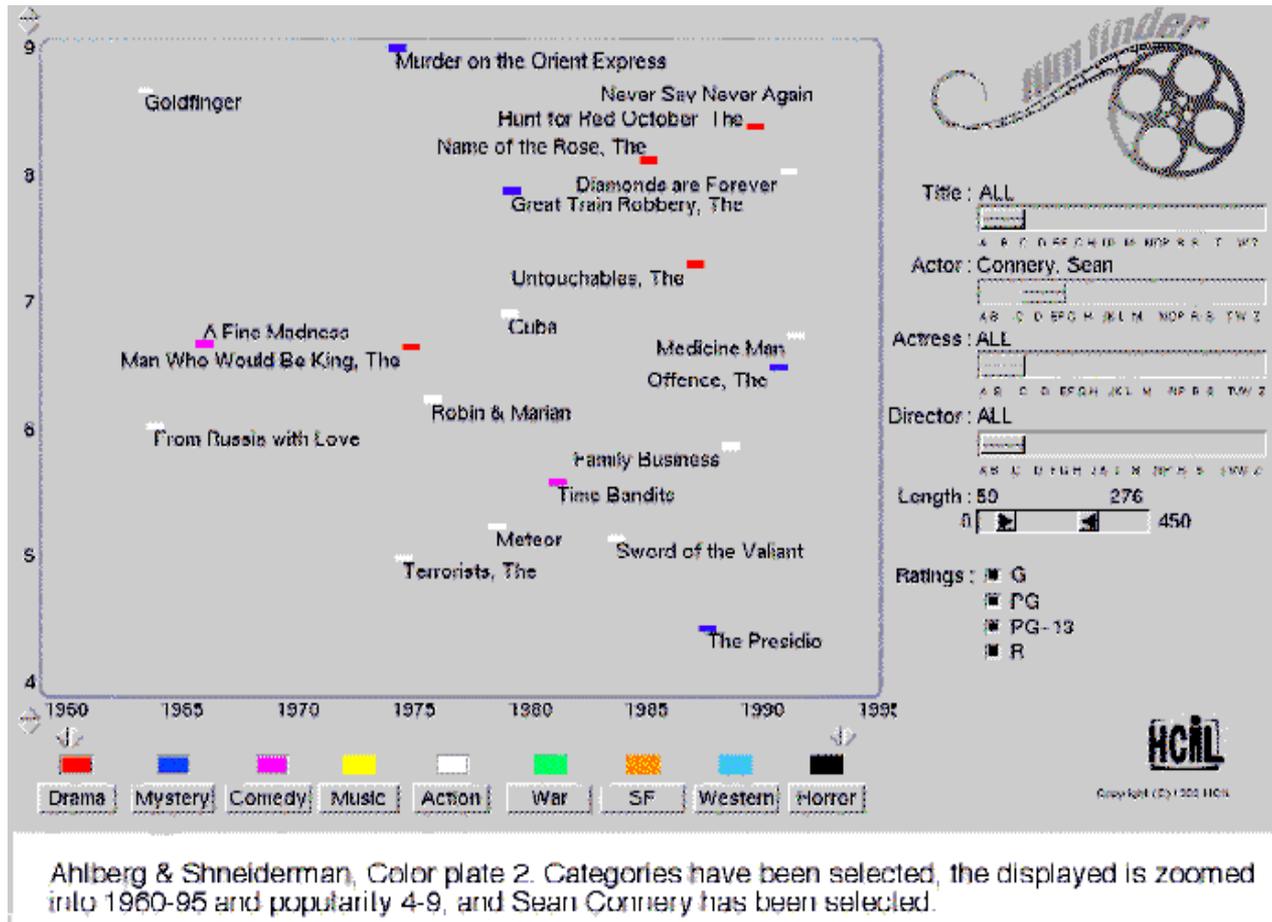
# Standard Techniques

- It's often hard to beat:
  - Line graphs, bar charts
  - Scatterplots (or Scatterplot Matrix)
  - Tables
- A Darwinian view of visualizations:
  - Only the fittest survive
  - We are in a period of great experimentation; eventually it will be clear what works and what dies out.
- A bright spot:
  - Enhancing the old techniques with interactivity
  - Example: Spotfire
    - Adds interactivity, color highlighting, zooming to scatterplots
  - Example: TableLens / Eureka
    - Adds interactivity and length cues to tables

# Spotfire: Integrating Interaction with Scatterplots



# Spotfire/IVEE: Integrating Interaction with Scatterplots

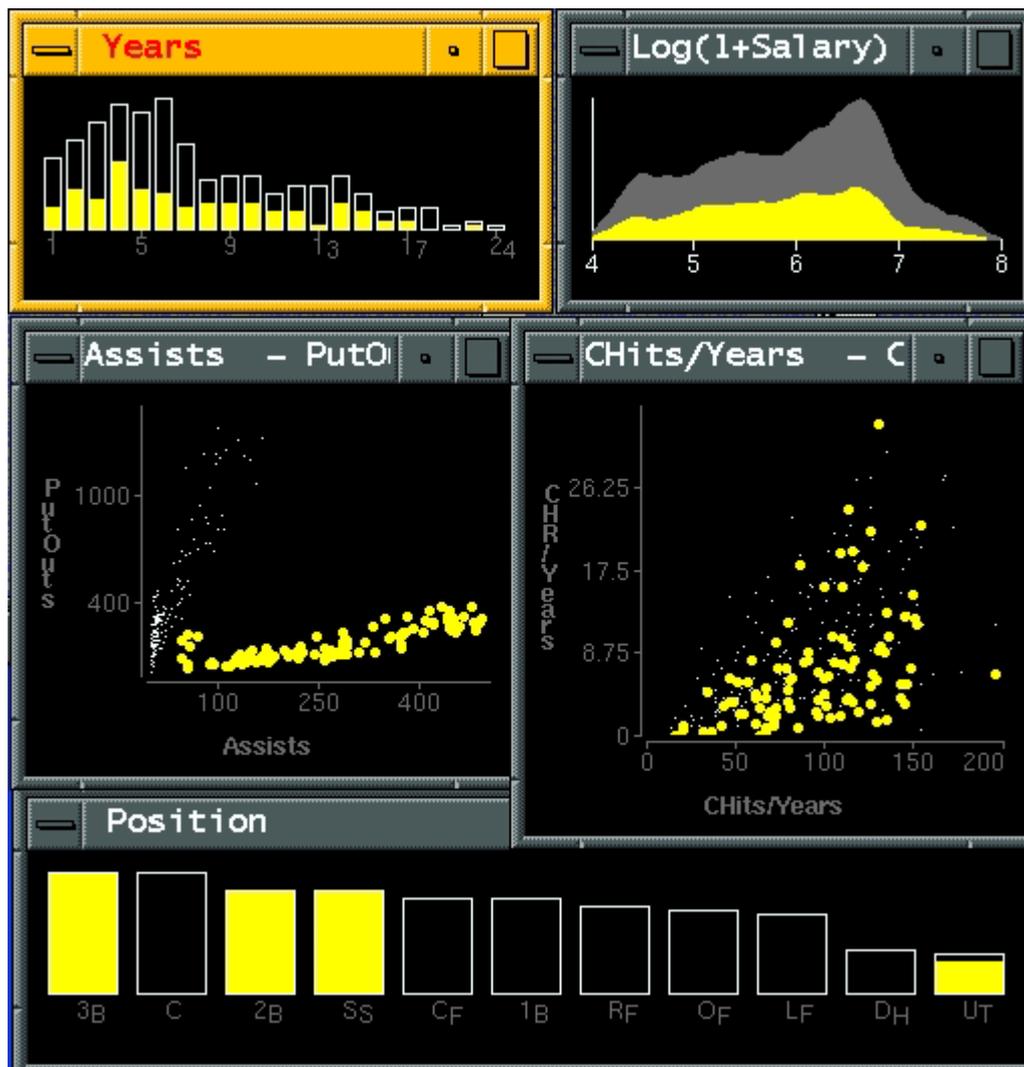


Ahlberg & Shneiderman, Color plate 2. Categories have been selected, the displayed is zoomed into 1960-95 and popularity 4-9, and Sean Connery has been selected.

# Brushing and Linking

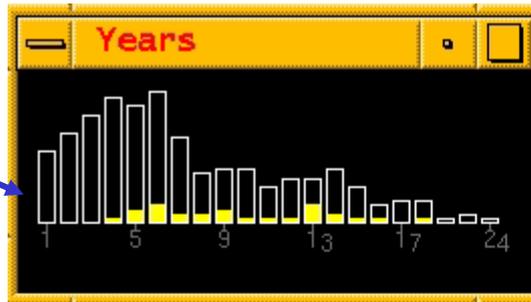
- Interactive technique
  - Highlighting
  - Brushing and Linking
- At least two things must be linked together to allow for brushing
  - select a subset of points
  - see the role played by this subset of points in one or more other views
- Example systems
  - Graham Will's EDV system
  - Ahlberg & Sheiderman's IVEE (Spotfire)

# Linking types of assist behavior to position played (from Eick & Wills 95)

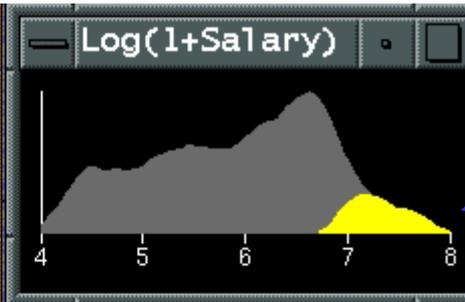


# Baseball data: Scatterplots and histograms and bars (from Eick & Wills 95)

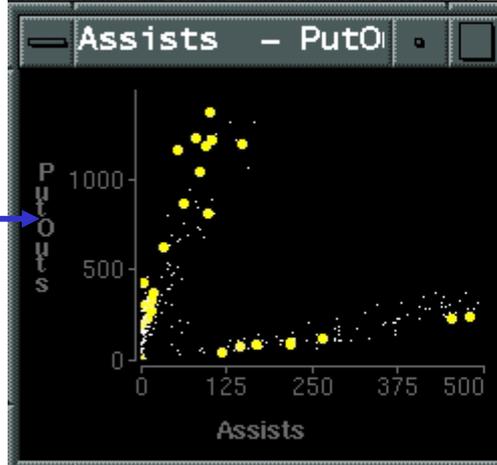
how long  
in majors



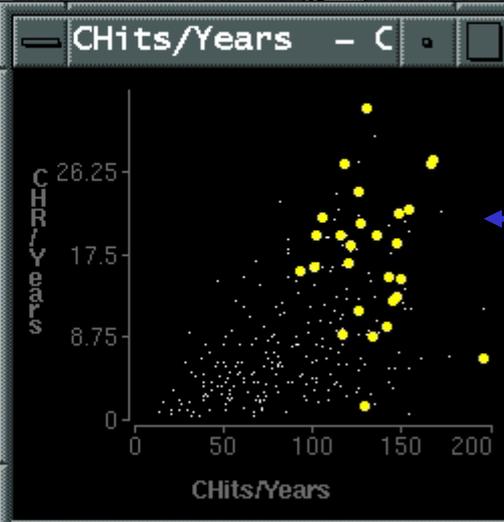
select high  
salaries



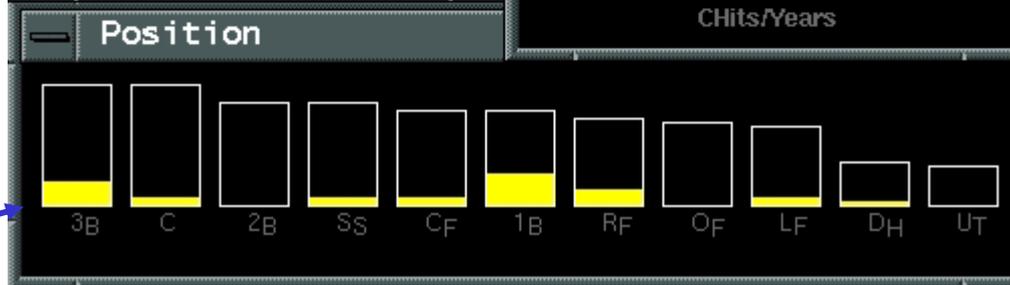
avg assists vs  
avg putouts  
(fielding ability)



avg career  
HRs vs avg  
career hits  
(batting ability)



distribution  
of positions  
played



# What was learned from interaction with this baseball data?

- Seems impossible to earn a high salary in the first three years
- High salaried players have a bimodal distribution (peaking around 7 & 13 yrs)
- Hits/Year a better indicator of salary than HR/Year
- High paid outlier with low HR and medium hits/year. Reason: person is player-coach
- There seem to be two differentiated groups in the put-outs/assists category (but not correlated with salary) Why?



# Animation

- “The quality or condition of being alive, active, spirited, or vigorous” (dictionary.com)
- “A dynamic visual statement that evolves through movement or change in the display”
- “... creating the illusion of change by rapidly displaying a series of single frames” (Roncarelli 1988).

# We Use Animation to...

- Tell stories / scenarios: cartoons
- Illustrate dynamic process / simulation
- Create a character / an agent
- Navigate through virtual spaces
- Draw attention
- Delight



# Cartoon Animation Principles

- Chang & Unger '93
- Solidity (squash and stretch)
  - Solid drawing
  - Motion blur
  - Dissolves
- Exaggeration
  - Anticipation
  - Follow through
- Reinforcement
  - Slow in and slow out
  - Arcs
  - Follow through

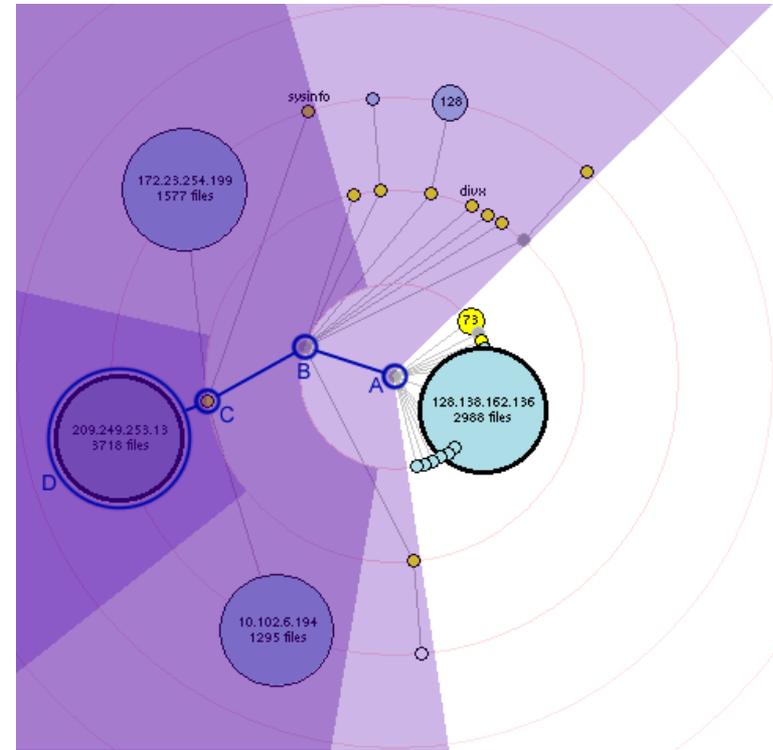
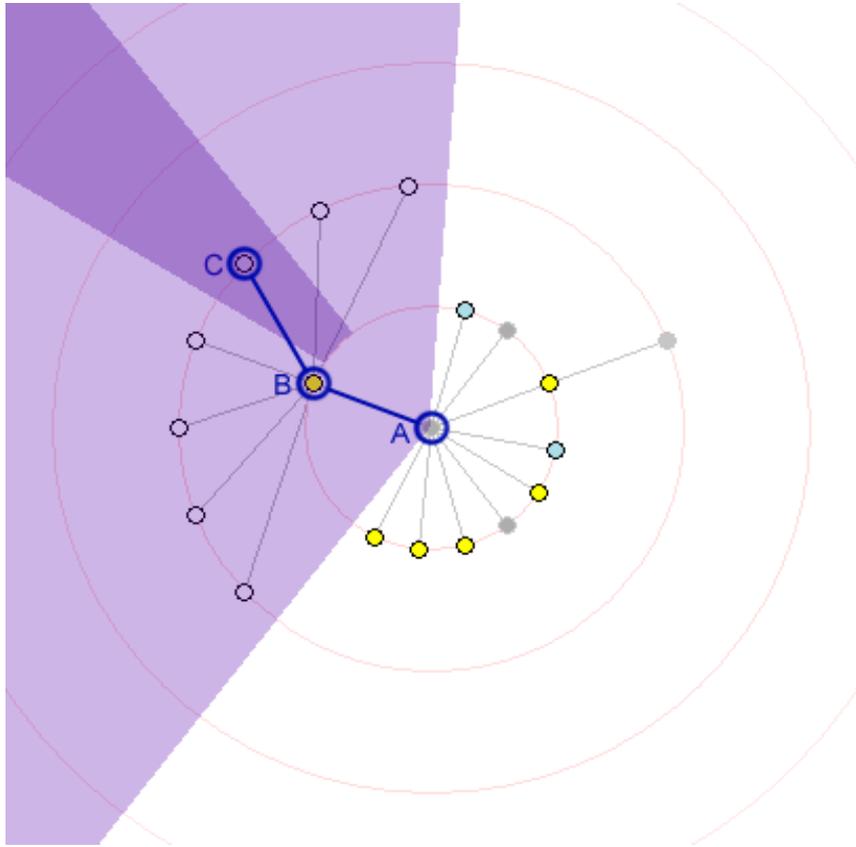
# Why Cartoon-Style Animation?

- Cartoons' theatricality is powerful in communicating to the user.
- Cartoons can make UI engage the user into its world.
- The medium of cartoon animation is like that of graphic computers.

# Application using Animation: Gnutellavision

- Visualization of Peer-to-Peer Network
  - Hosts (with color for status and size for number of files)
  - Nodes with closer network distance from focus on inner rings
  - Queries shown; can trace queries
- Gnutellavision as exploratory tool
  - Very few hosts share many files
  - Uneven propagation of queries
  - Qualitative assessment of queries (simple)

# Layout - Illustration



# Animation in Gnutellavision

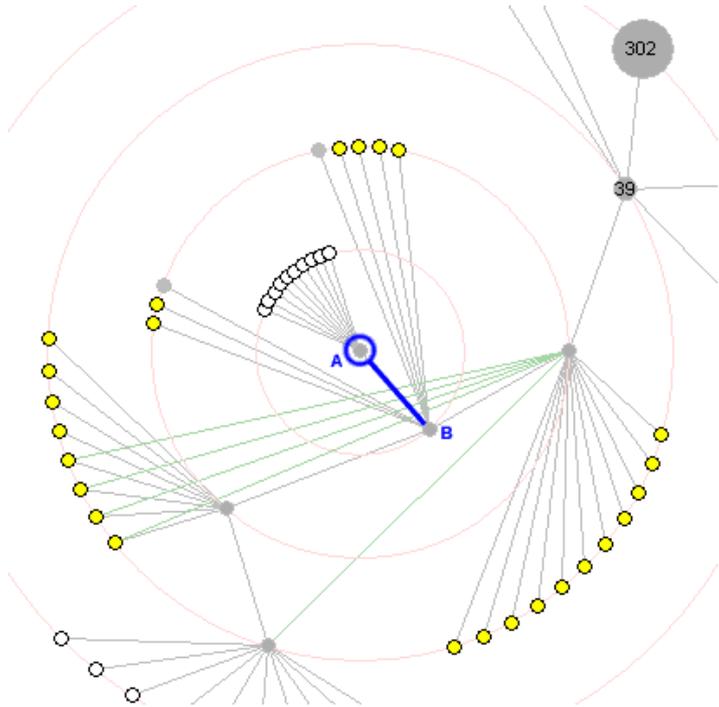
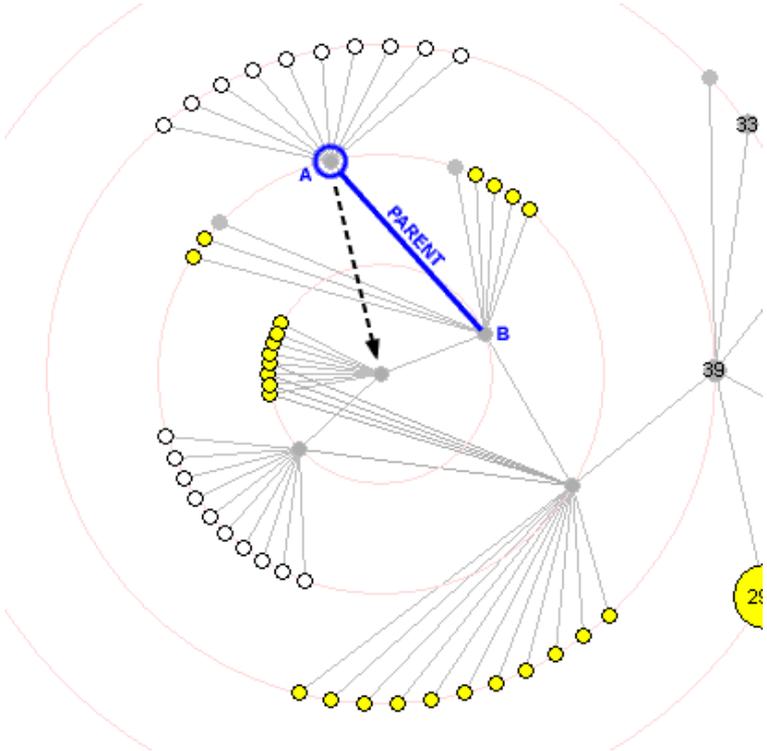
Goal of animation is to help maintain context of nodes and general orientation of user during refocus

- Transition Paths
  - Linear interpolation of polar coordinates
  - Node moves in arc not straight line
  - Moves along circle if not changing levels (like great circles on earth)
  - Spirals in or out to next ring

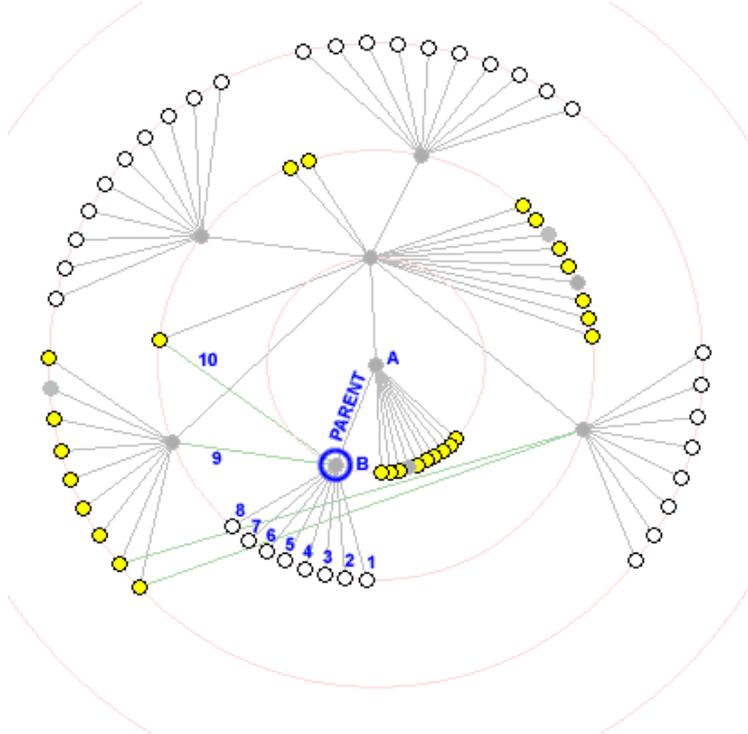
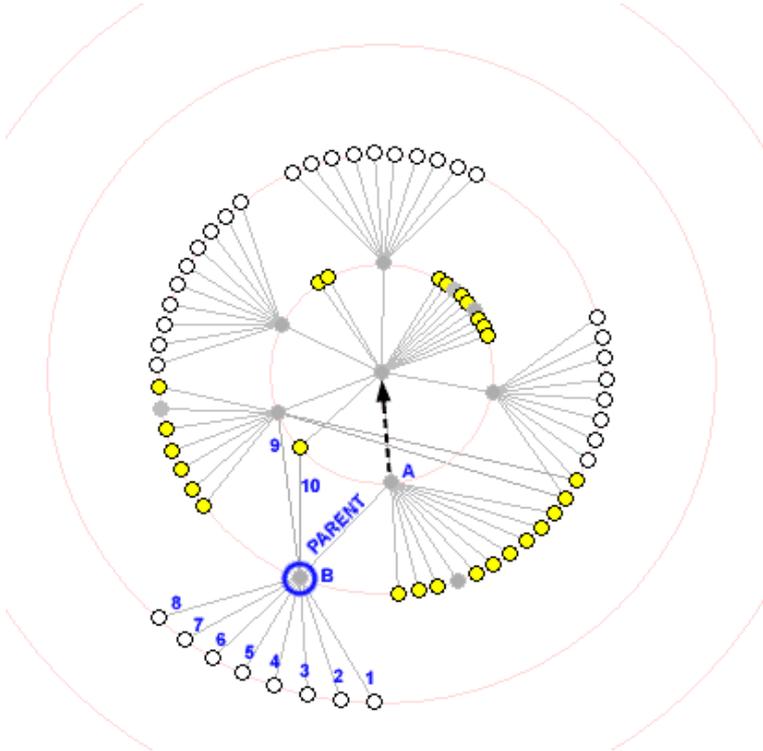
# Animation (continued)

- Transition constraints
  - Orientation of transition to minimize rotational travel
  - (Move former parent away from new focus in same orientation)
  - Avoid cross-over of edges
  - (to allow users to keep track of which is which)
- Animation timing
  - Slow in Slow out timing (allows users to better track movement)

# Transition Constraint - Orientation



# Transition Constraint - Order



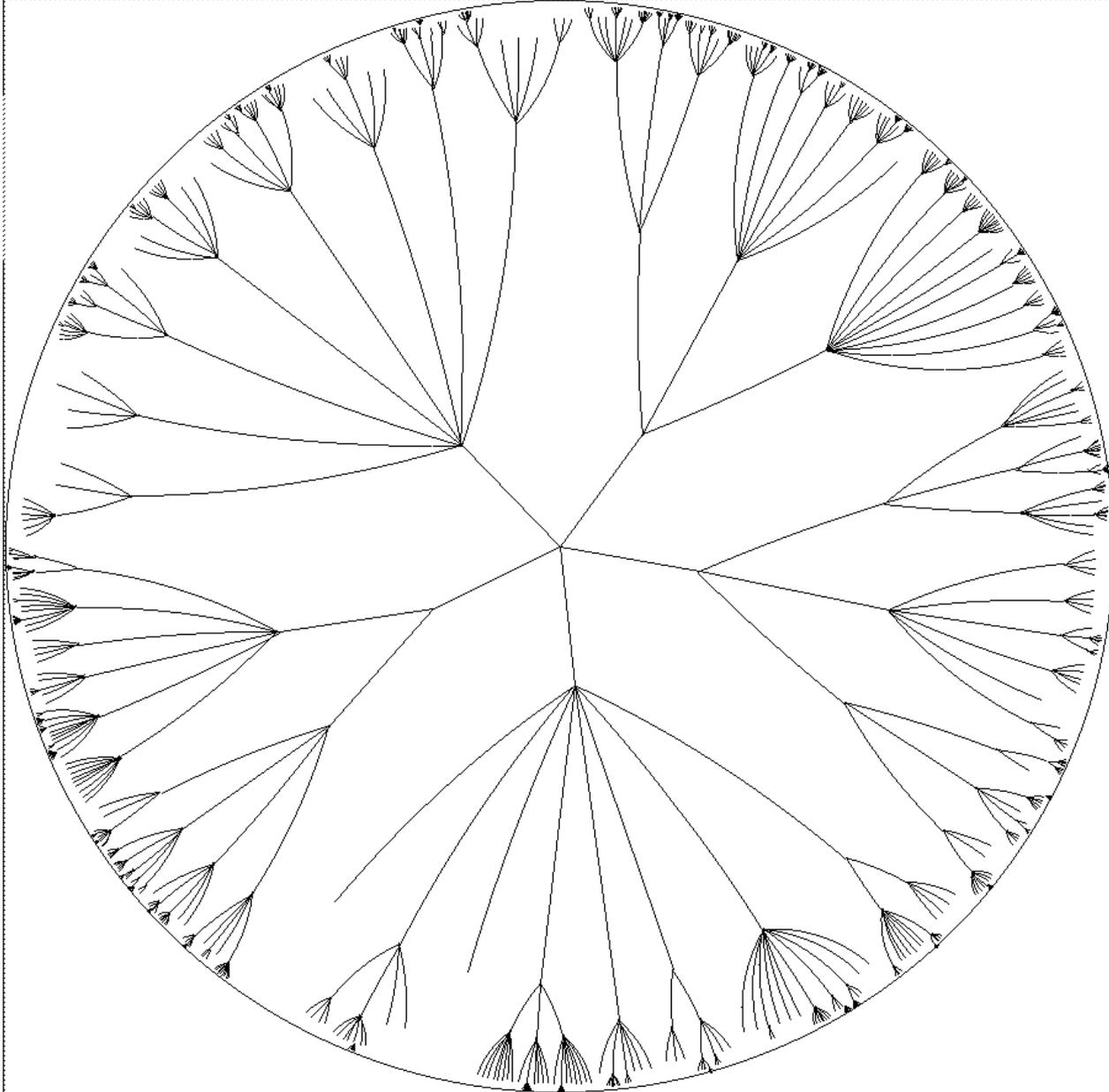
# Usability Testing

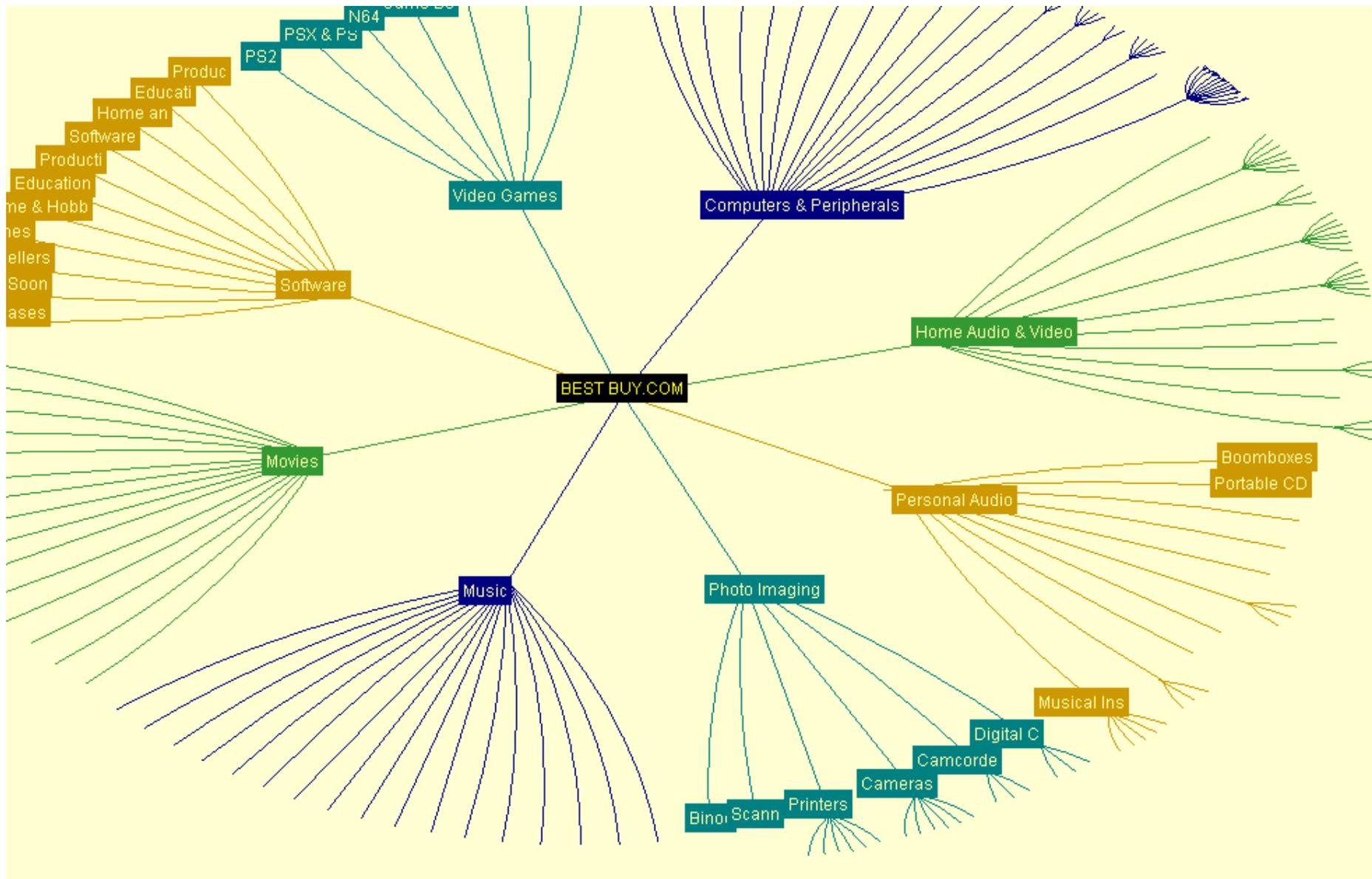
- In general, users appreciated the subtleties added to the general method when the number of nodes increased.
- Perhaps the most interesting result is that most people preferred rectangular movement for the small graph and polar coordinate movement for the large one.

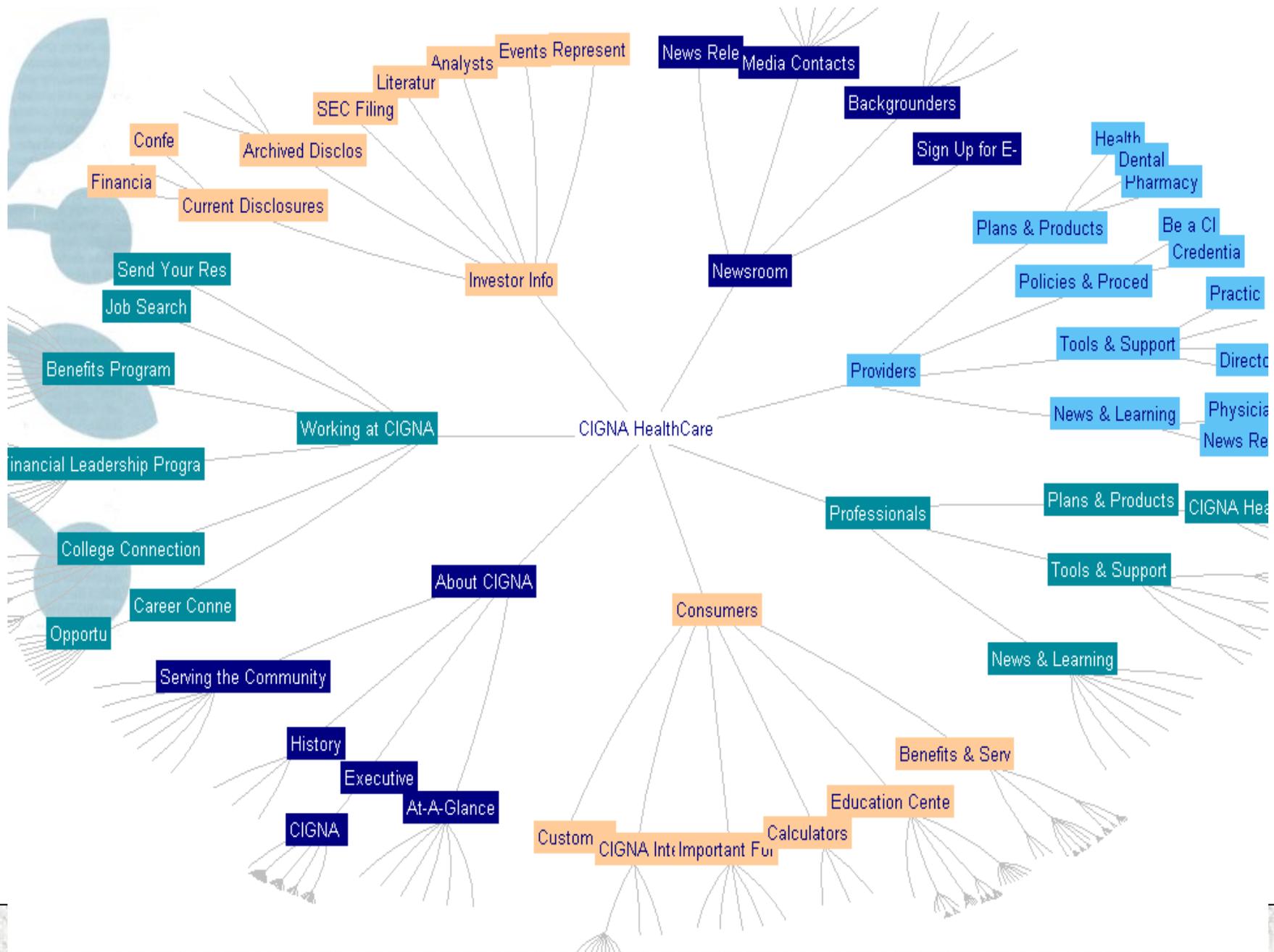
Overall Preference of Users		
	No Features	All Features
Small Graph	5	5
Large Graph	1	9

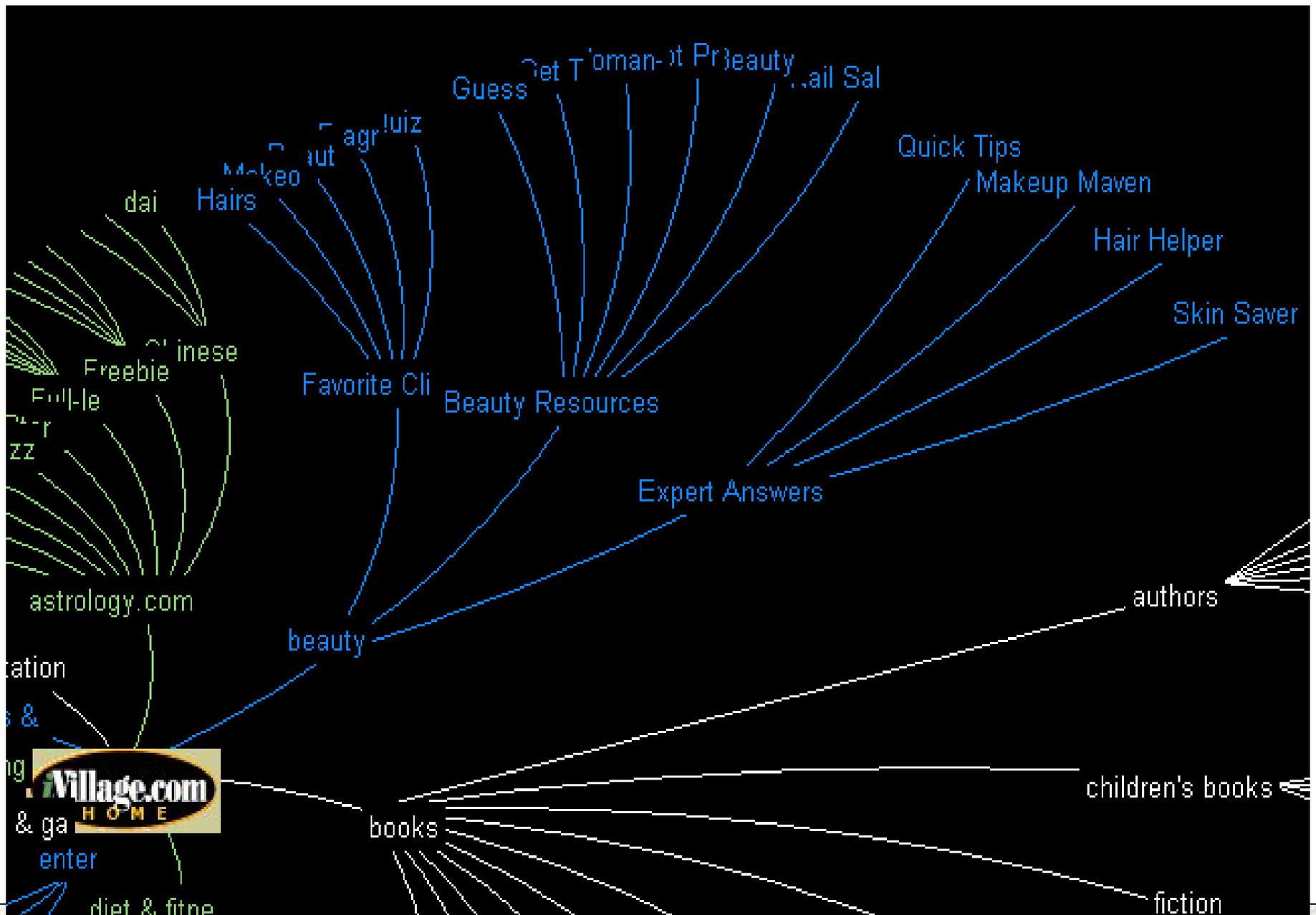
# Hyperbolic Tree

- A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies (1995) John Lamping, Ramana Rao, Peter Pirolli Proc. ACM Conf. Human Factors in Computing Systems, CHI
- Also uses animation
- Tree-based layout; leaves stretch to infinity
- Only a few labels can be seen at a time









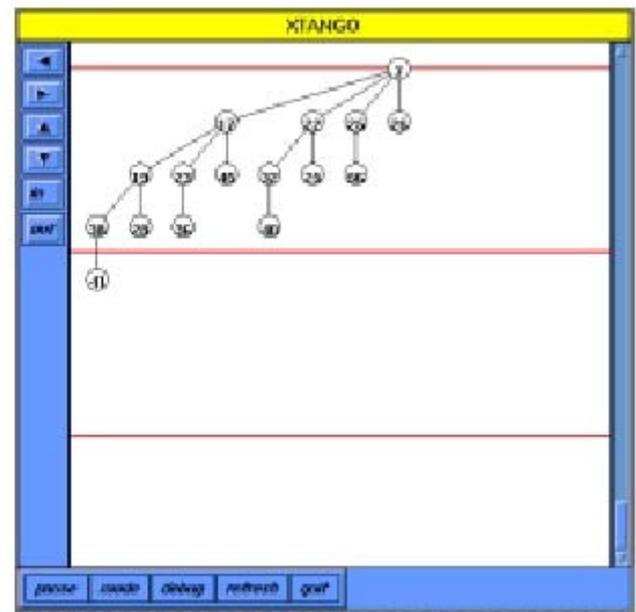
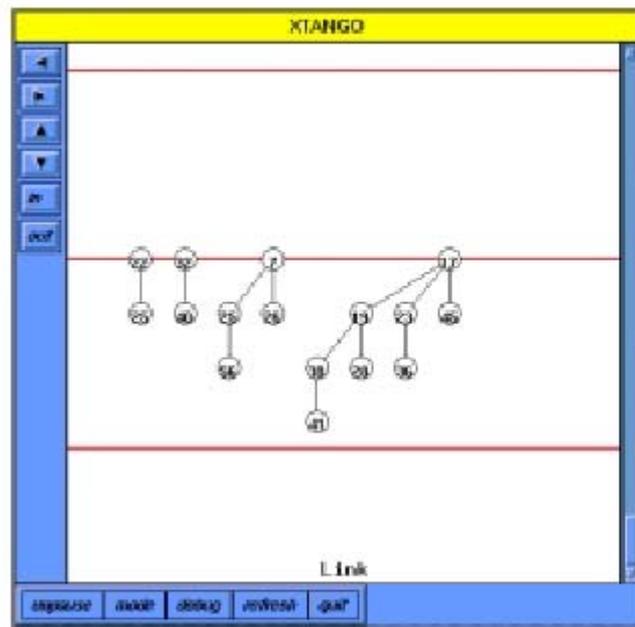
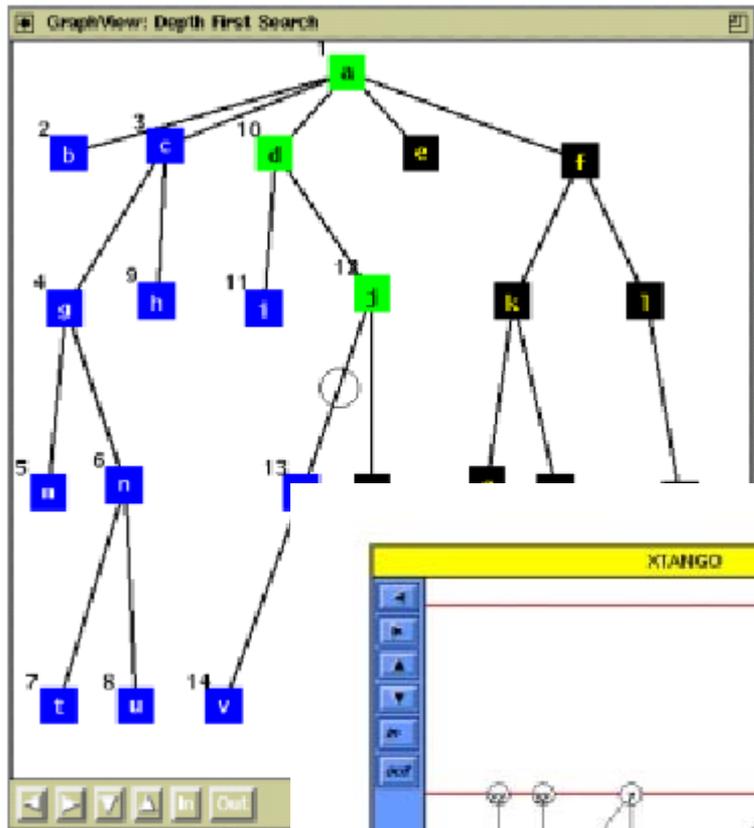
# Issues

- Displaying text
  - The size of the text
    - Works good for small things like directories
    - Not so good for URLs
- Only a portion of the data can be seen in the focus at one time
- Only works for certain types of data - Hierarchical
- Not clear if it is actually useful for anything.



# Animating Algorithms

- Kehoe, Stasko, and Taylor, “Rethinking Evaluation of Algorithm Animations as Learning Aids”
- Why previous studies present no benefits:
  - No or limited benefits from particular animations
  - Benefits are not captured in measurements
  - Design of experiments hides the benefits
- Methods for this study:
  - Combination of qualitative & quantitative
  - More flexible setting
  - Metrics: score for each type of questions, time used, usage of materials, qualitative data from observations & interviews





# Findings

- Value of animation is more apparent in interactive situations
- Most useful to learn procedural operations
- Makes subject more accessible & less intimidating → increase motivation

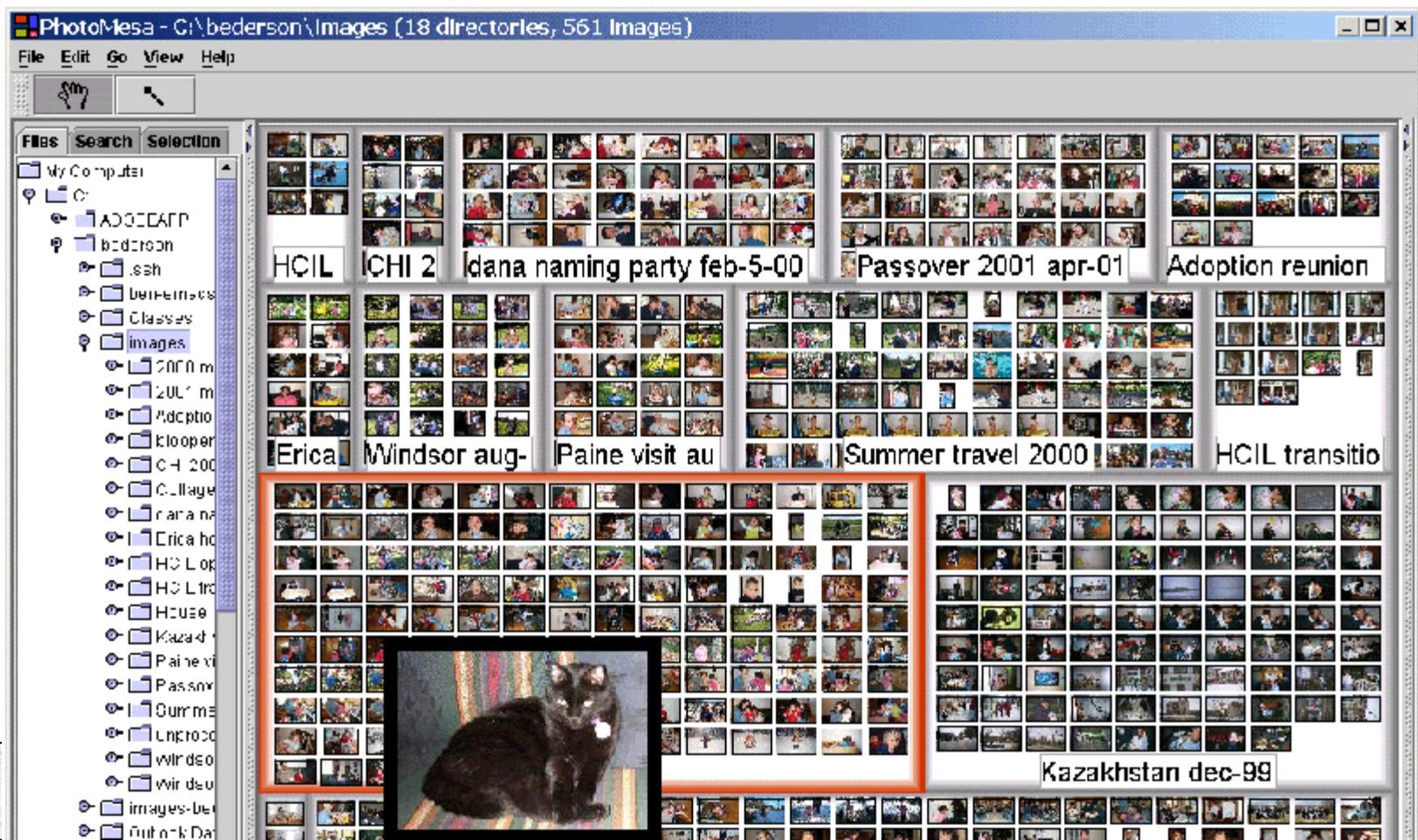
# What Isn't Working?

The existing studies indicate that we don't yet know how to make the following work well for every-day tasks:

- Pan-and-Zoom
- 3D Navigation
- Node-and-link representations of concept spaces

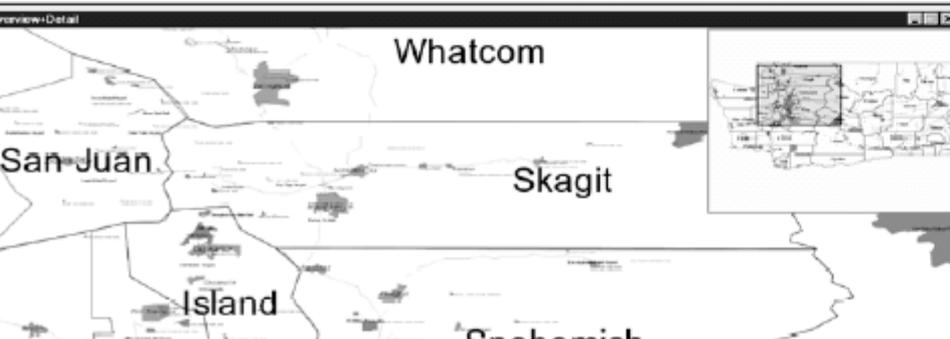
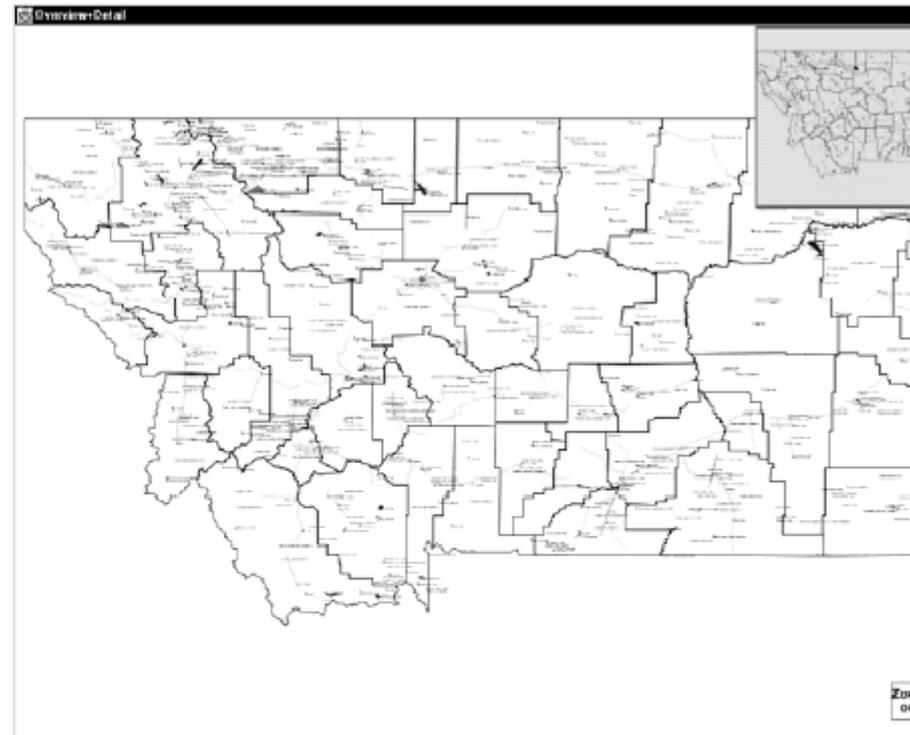
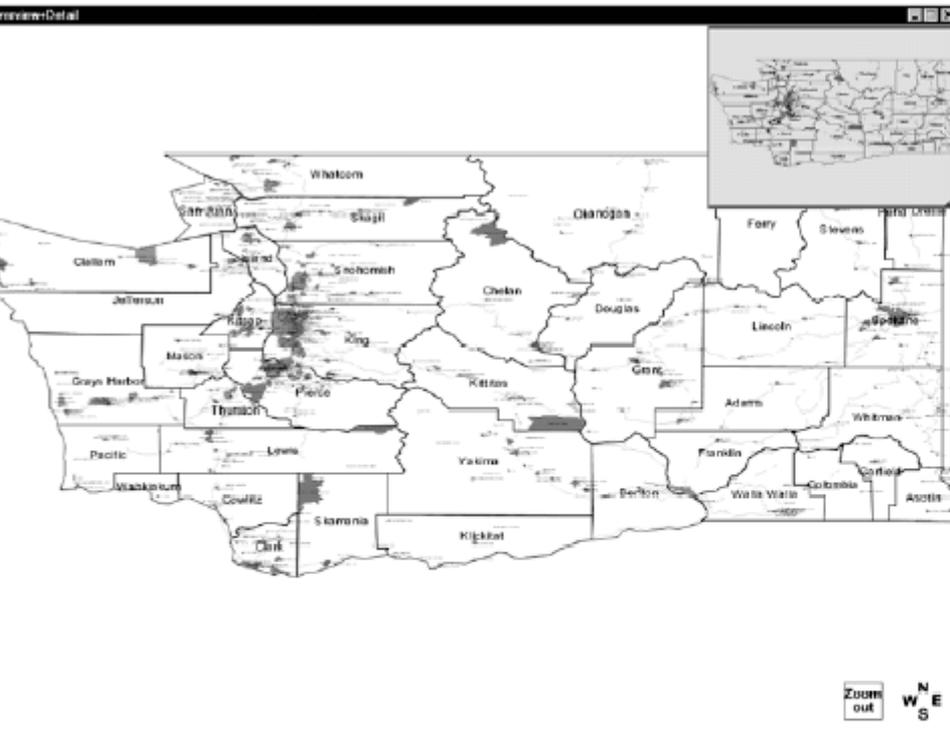
# Zoom, Overview + Detail

- An exception, possibly:
  - Benjamin B. Bederson: PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps. [UIST 2001](#): 71-80



# Overview + Detail

- K. Hornbaek et al., Navigation patterns and Usability of Zoomable User Interfaces with and without an Overview, ACM TOCHI, 9(4), December 2002.



# Overview + Detail

- K. Hornbaek et al., Navigation patterns and Usability of Zoomable User Interfaces with and without an Overview, ACM TOCHI, 9(4), December 2002.
- A study on integrating Overview + Detail on a Map search task
  - Incorporating panning & zooming as well.
  - They note that panning & zooming does not do well in most studies.
- Results seem to be
  - Subjectively, users prefer to have a linked overview
  - But they aren't necessarily faster or more effective using it
  - Well-constructed representation of the underlying data may be more important.
- More research needed as each study seems to turn up different results, sensitive to underlying test set.

# Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up

# Problem Solving

# Problem Solving

- A Detective Tool for Multidimensional Data
  - Inselberg on using Parallel Coordinates
- Analyzing Web Clickstream Data
  - Brainerd & Becker, Waterson et al.
- Information Visualization for Pattern Detection
  - Carlis & Konstan on Periodic Data
- Visualization vs. Analysis
  - Comments by Wesley Johnson of Chevron

# Multidimensional Detective

A. Inselberg, Multidimensional Detective, Proceedings of IEEE Symposium on Information Visualization (InfoVis '97), 1997.

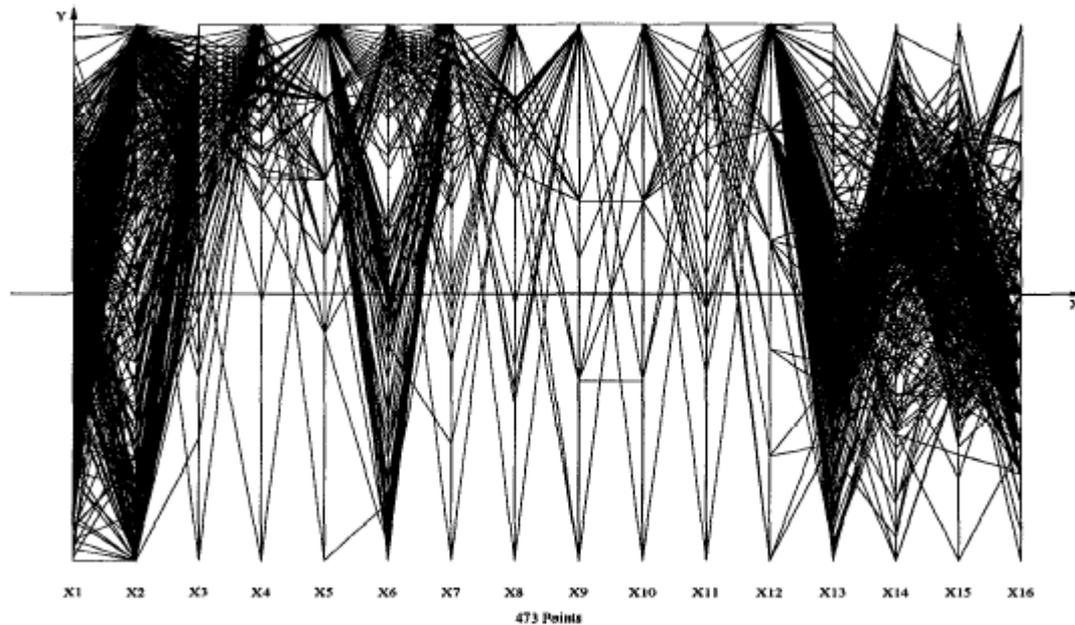


Figure 1: The full dataset consisting of 473 batches

# A Detective Story

A. Inselberg, Multidimensional Detective, Proceedings of IEEE Symposium on Information Visualization (InfoVis '97), 1997

## Inselberg's Principles for analysis using visualizations:

1. Do not let the picture scare you
2. Understand your objectives
  - Use them to obtain visual cues
3. Carefully scrutinize the picture
4. Test your assumptions, especially the "I am really sure of's"
5. You can't be unlucky all the time!

# A Detective Story

A. Inselberg, *Multidimensional Detective*, Proceedings of IEEE Symposium on Information Visualization (InfoVis '97), 1997

- **The Dataset:**
  - Production data for 473 batches of a VLSI chip
  - 16 process parameters
  - The yield: % of produced chips that are useful
    - X1
  - The quality of the produced chips (speed)
    - X2
  - 10 types of defects (zero defects shown at top)
    - X3 ... X12
  - 4 physical parameters
    - X13 ... X16
- **The Objective:**
  - Raise the yield (X1) and maintain high quality (X2)

# Multidimensional Detective

A. Inselberg, Multidimensional Detective, Proceedings of IEEE Symposium on Information Visualization (InfoVis '97), 1997.

Do Not Let the Picture Scare You!!

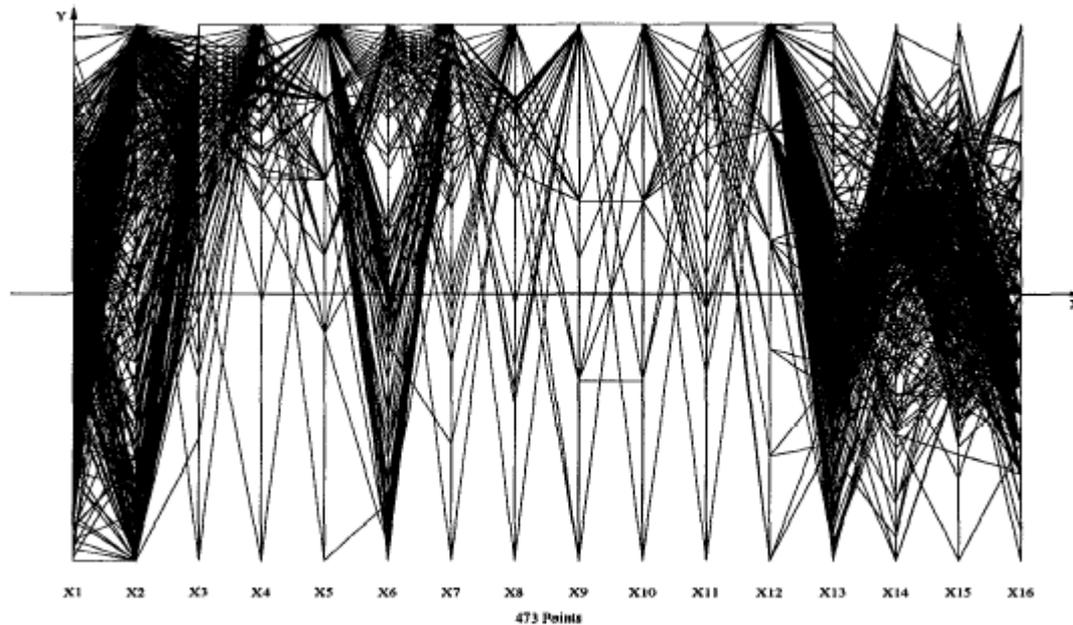


Figure 1: The full dataset consisting of 473 batches

# Multidimensional Detective

- Each line represents the values for one batch of chips
- This figure shows what happens when only those batches with both high X1 and high X2 are chosen
- Notice the separation in values at X15
- Also, some batches with few X3 defects are not in this high-yield/high-quality group.

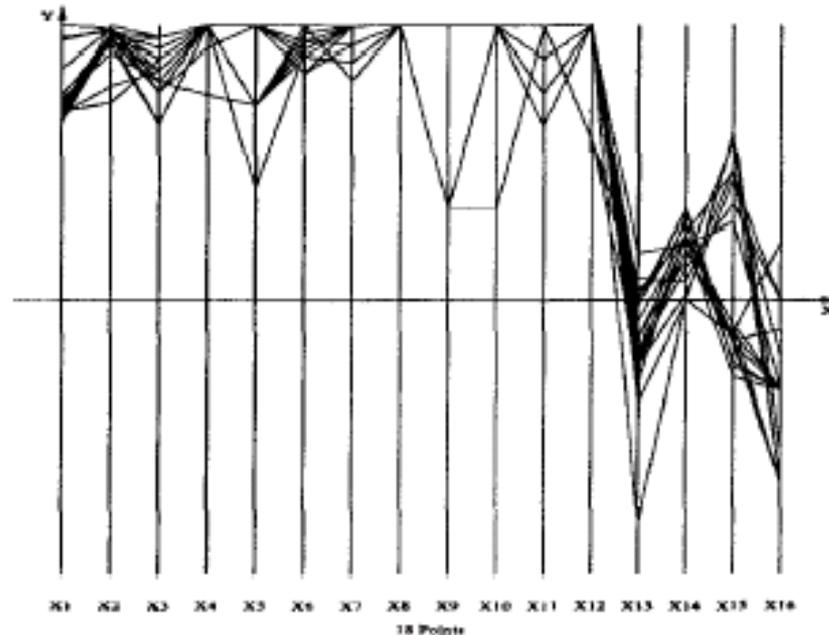


Figure 2: The batches high in Yield, X1, and Quality, X2.

# Multidimensional Detective

- Now look for batches which have *nearly* zero defects.
  - For 9 out of 10 defect categories
- Most of these have low yields
- Surprising because we know from first diagram that some defects are ok.
- Go back to first diagram, looking at defect categories
- Notice that X6 behaves differently than the rest
- Allow two defects, where one defect in X6
- This results in the very best batch appearing

# Multidimensional Detective

- Fig 5 and 6 show that high yield batches don't have non-zero values for defects of type X3 and X6
  - Don't believe your assumptions ...
- Looking now at X15 we see the separation is important
  - Lower values of this property end up in the better yield batches

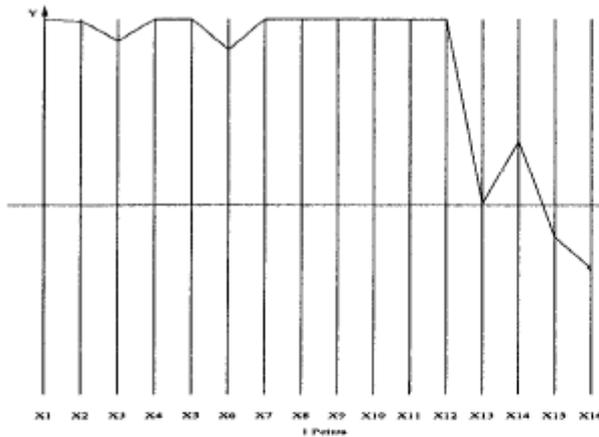


Figure 5: The best batch. Highest in Yield, X1, and very high in Quality, X2.

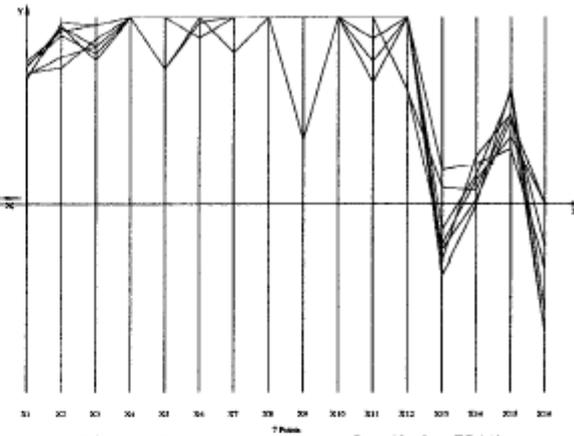
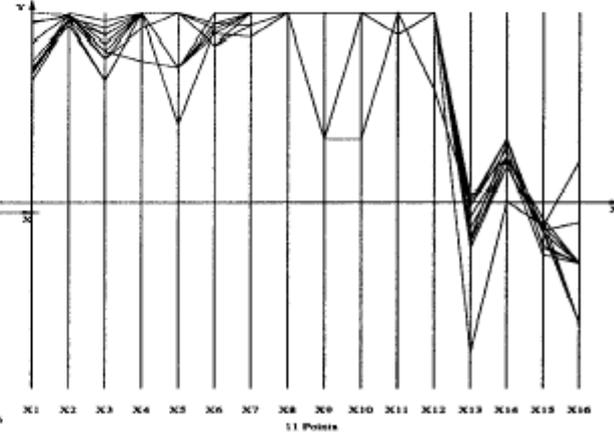
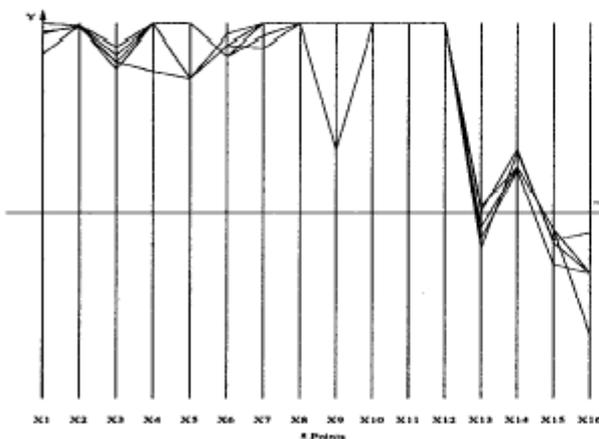
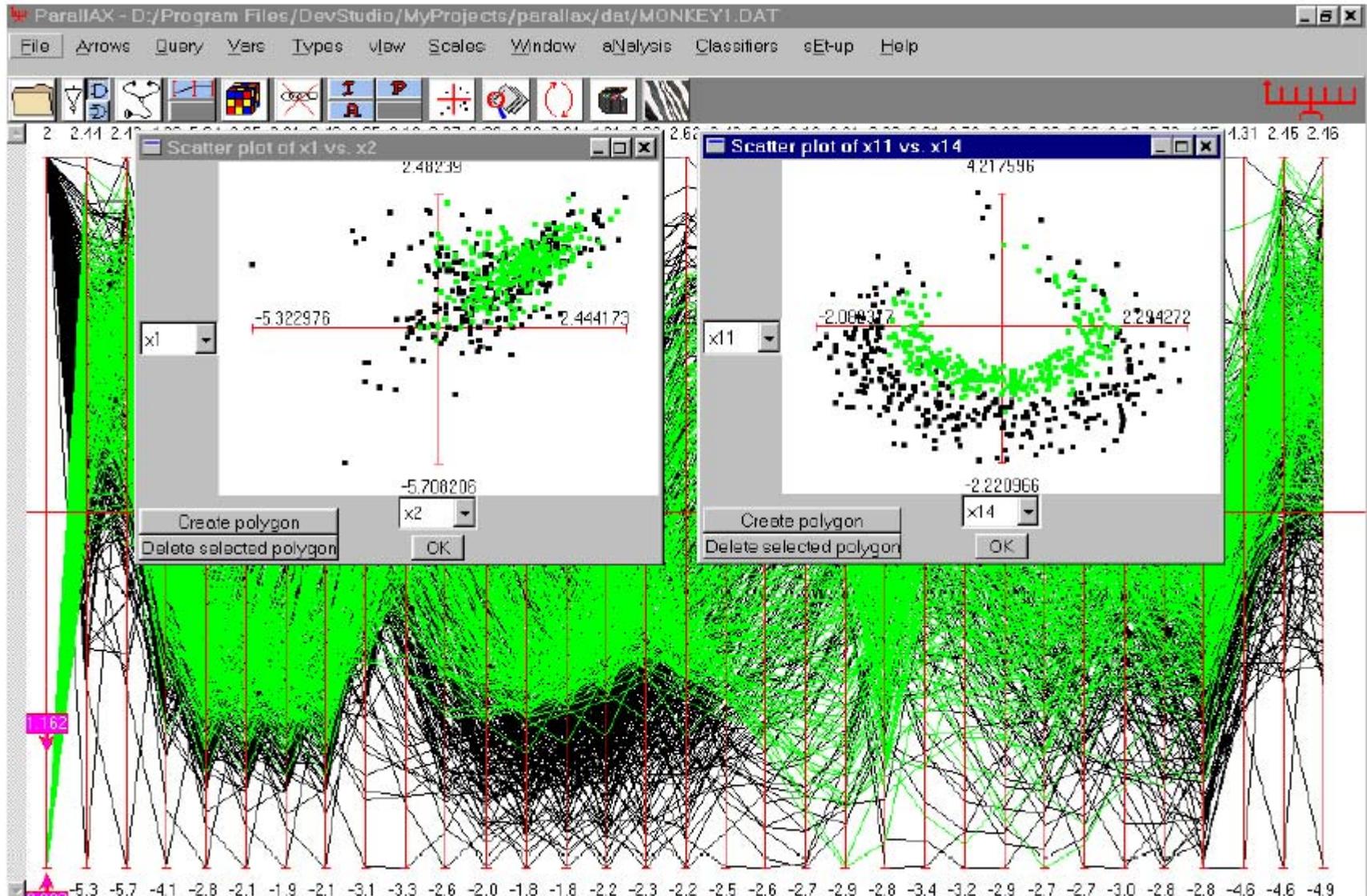


Figure 7: Upper range of split in X15



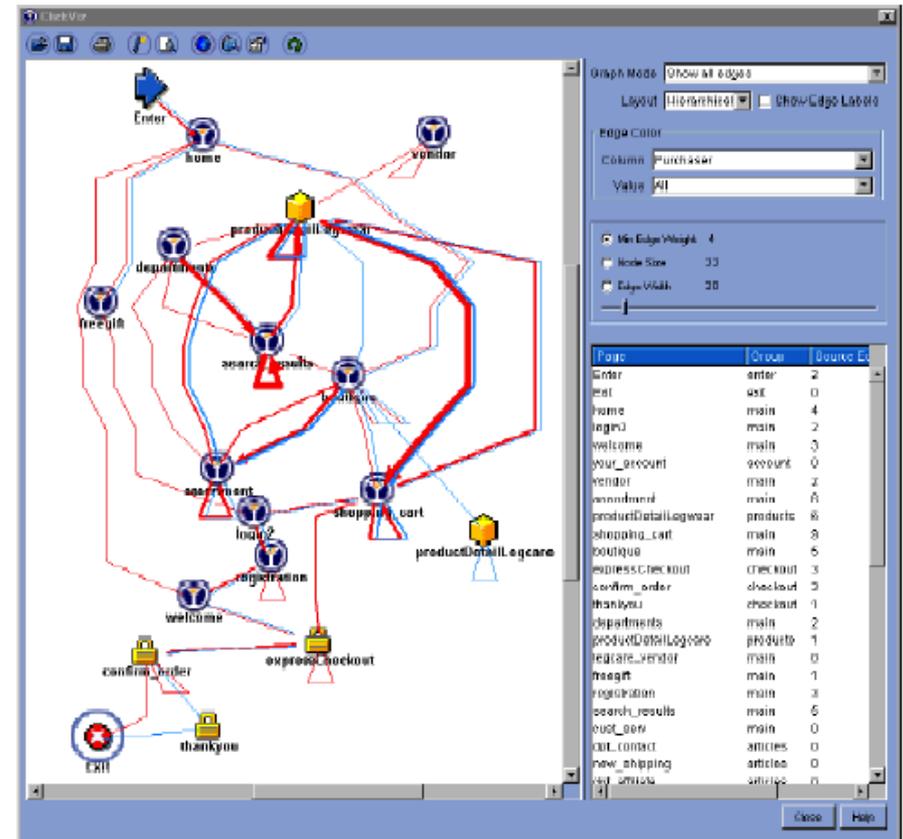
# Automated Analysis

A. Inselberg, Automated Knowledge Discovery using Parallel Coordinates, INFOVIS '99



# Case Study: E-Commerce Clickstream Visualization

- Brainerd & Becker, IEEE Infovis 2001
- Aggregate nodes using an icon (e.g. all the checkout pages)
- Edges represent transitions
  - Wider means more transitions

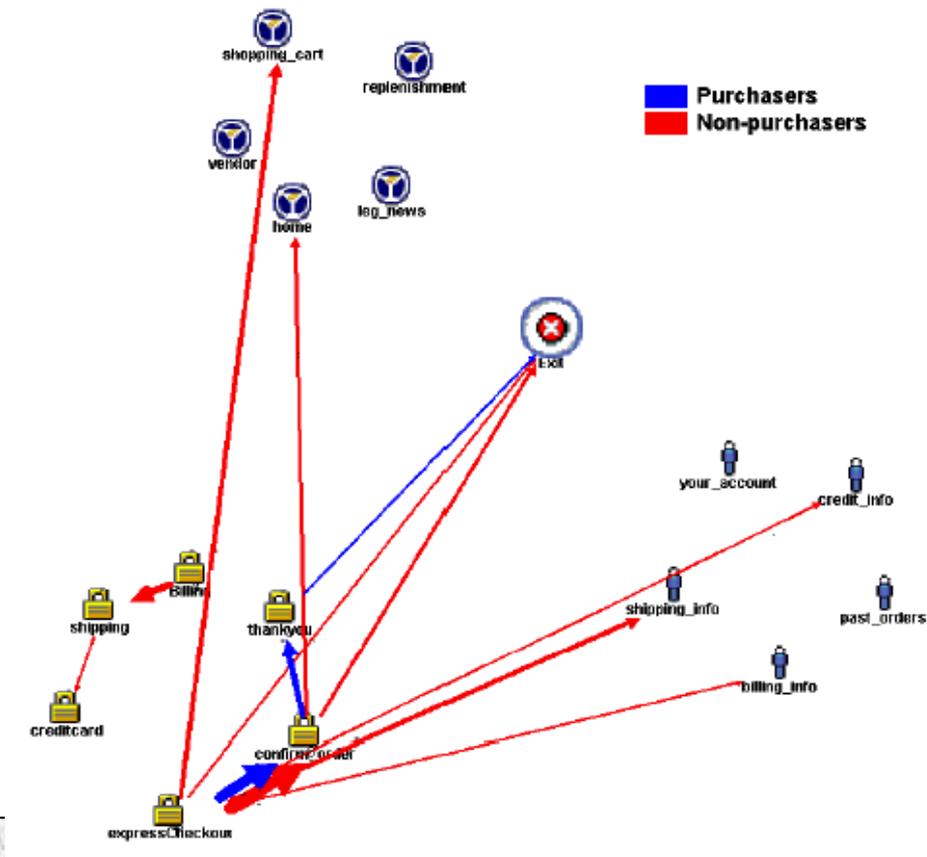


# Customer Segments

- Collect
  - Clickstream
  - Purchase history
  - Demographic data
- Associates customer data with their clickstream
- Different color for each customer segment

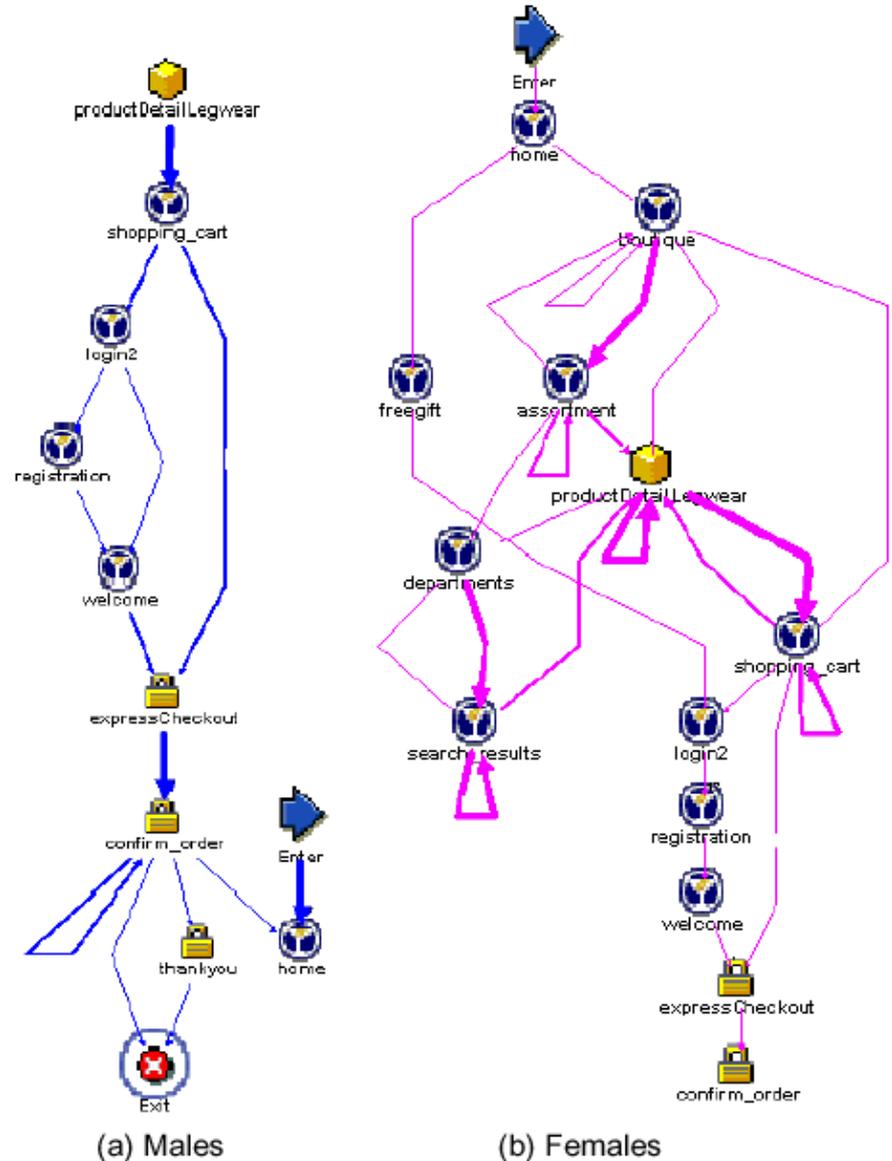
# Layout

- Aggregation based on file system path



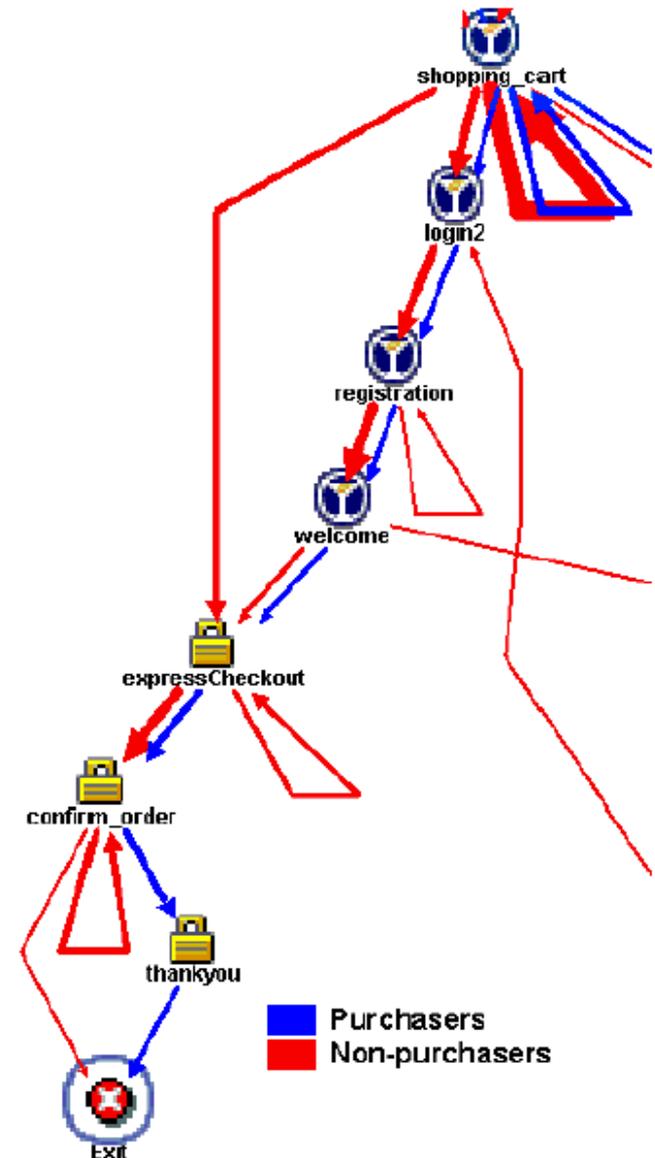
# Initial Findings

- Gender shopping differences



# Initial Findings (cont)

- Checkout process analysis
- Newsletter hurting sales

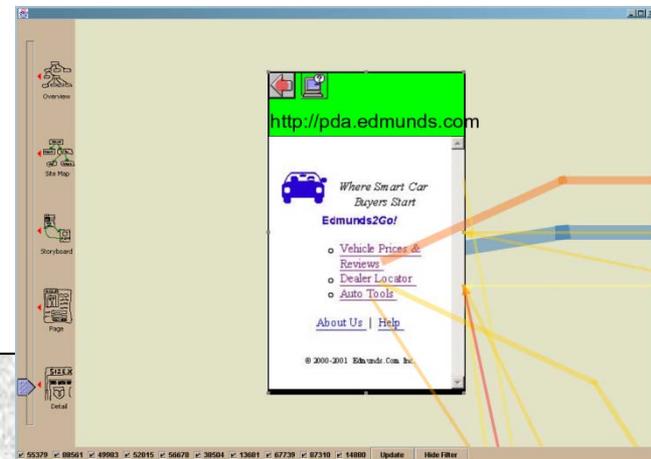
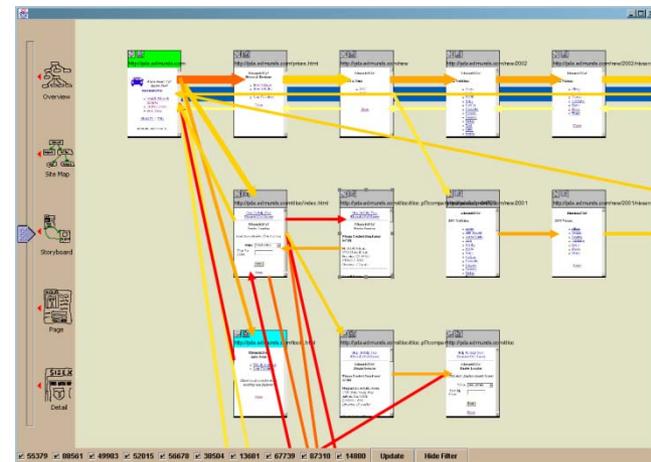
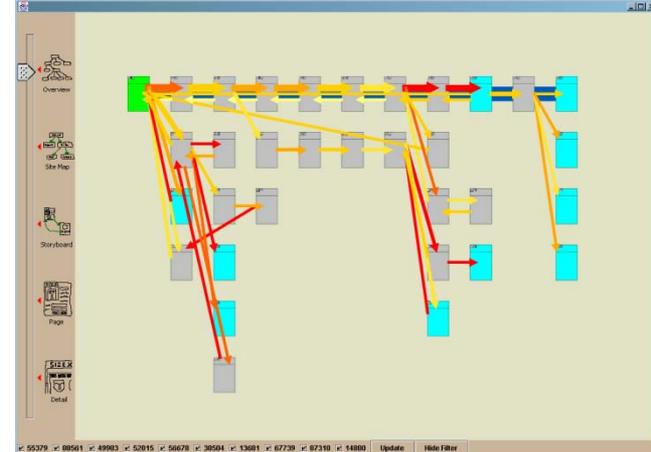


# WebQuilt

Interactive, zoomable directed graph

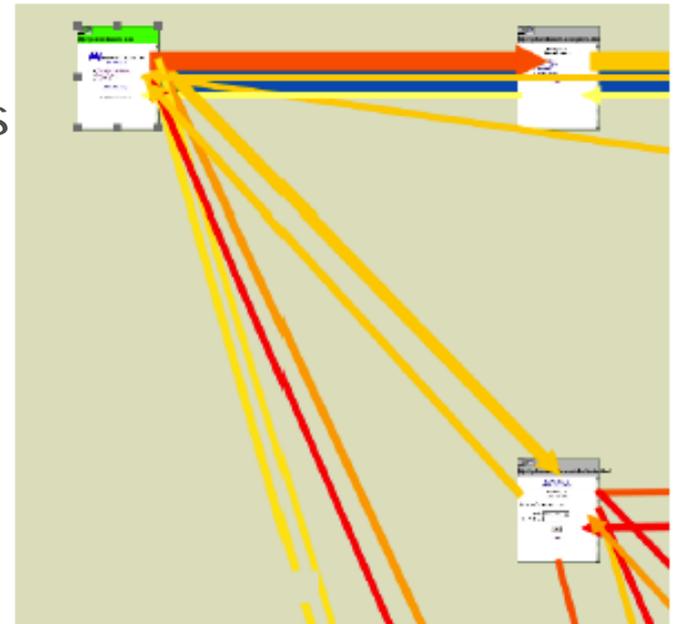
- Nodes = web pages
- Edges = aggregate traffic between pages

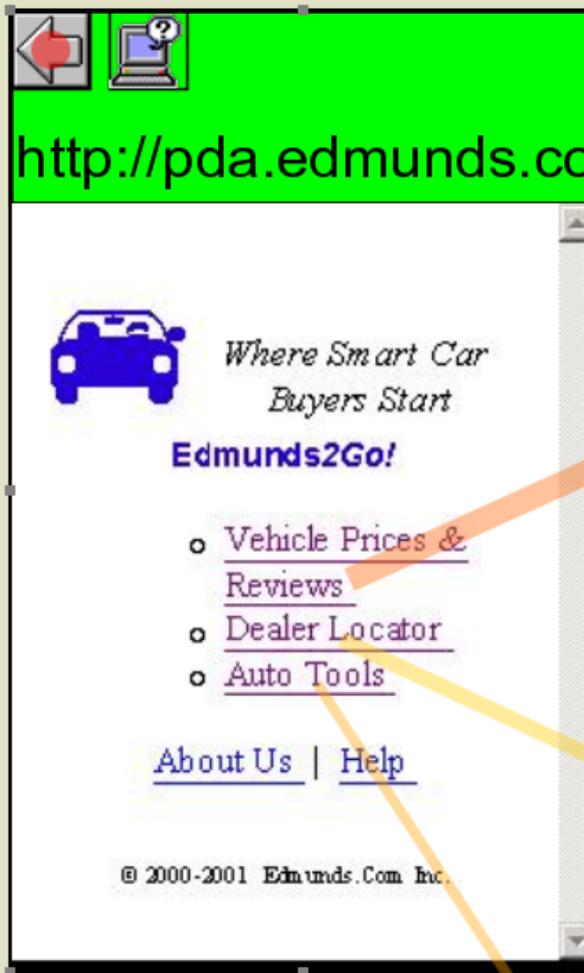
Waterson et al., "What Did They Do? Understanding Clickstreams with the WebQuilt Visualization System." in AVI 2002.



# Directed graph

- Nodes: visited pages
  - Color marks entry and exit nodes
- Arrows: traversed links
  - Thicker: more heavily traversed
  - Color
    - Red/yellow: Time spend before clicking
    - Blue: optimal path chosen by designer





http://pda.edmunds.com



Where Smart Car  
Buyers Start

**Edmunds2Go!**

- o [Vehicle Prices & Reviews](#)
- o [Dealer Locator](#)
- o [Auto Tools](#)

[About Us](#) | [Help](#)

© 2000-2001 Edmunds.Com Inc.



Overview



Site Map



Storyboard



Page



Detail

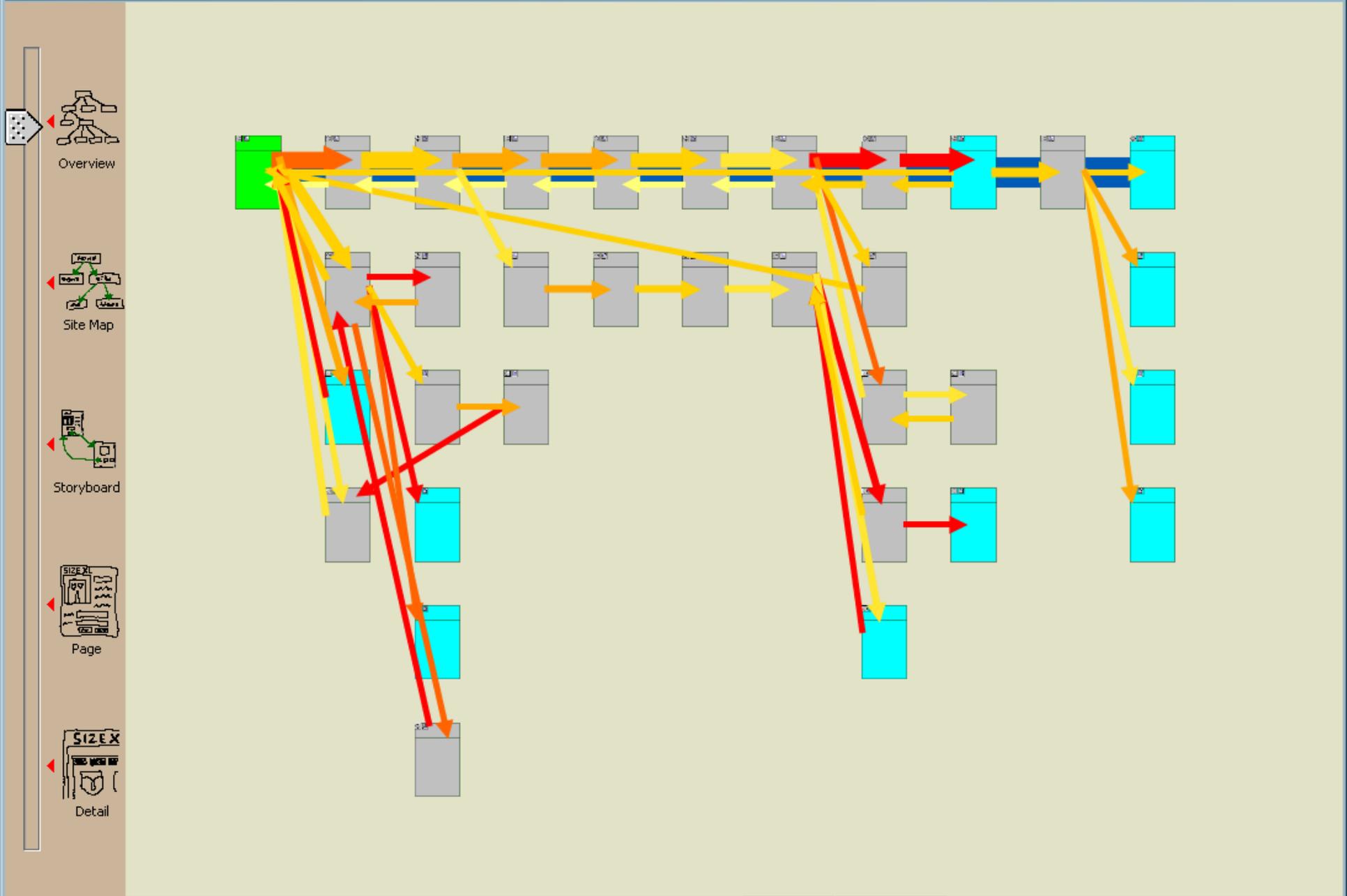
# Pilot Usability Study

- Edmunds.com PDA web site
- Visor Handspring equipped with a OmniSky wireless modem
- 10 users asked to find...
  - Anti-lock brake information on the latest Nissan Sentra model
  - The Nissan dealer closest to them.

# In the Lab vs. Out in the Wild

Comparing in-lab usability testing with WebQuilt remote usability testing

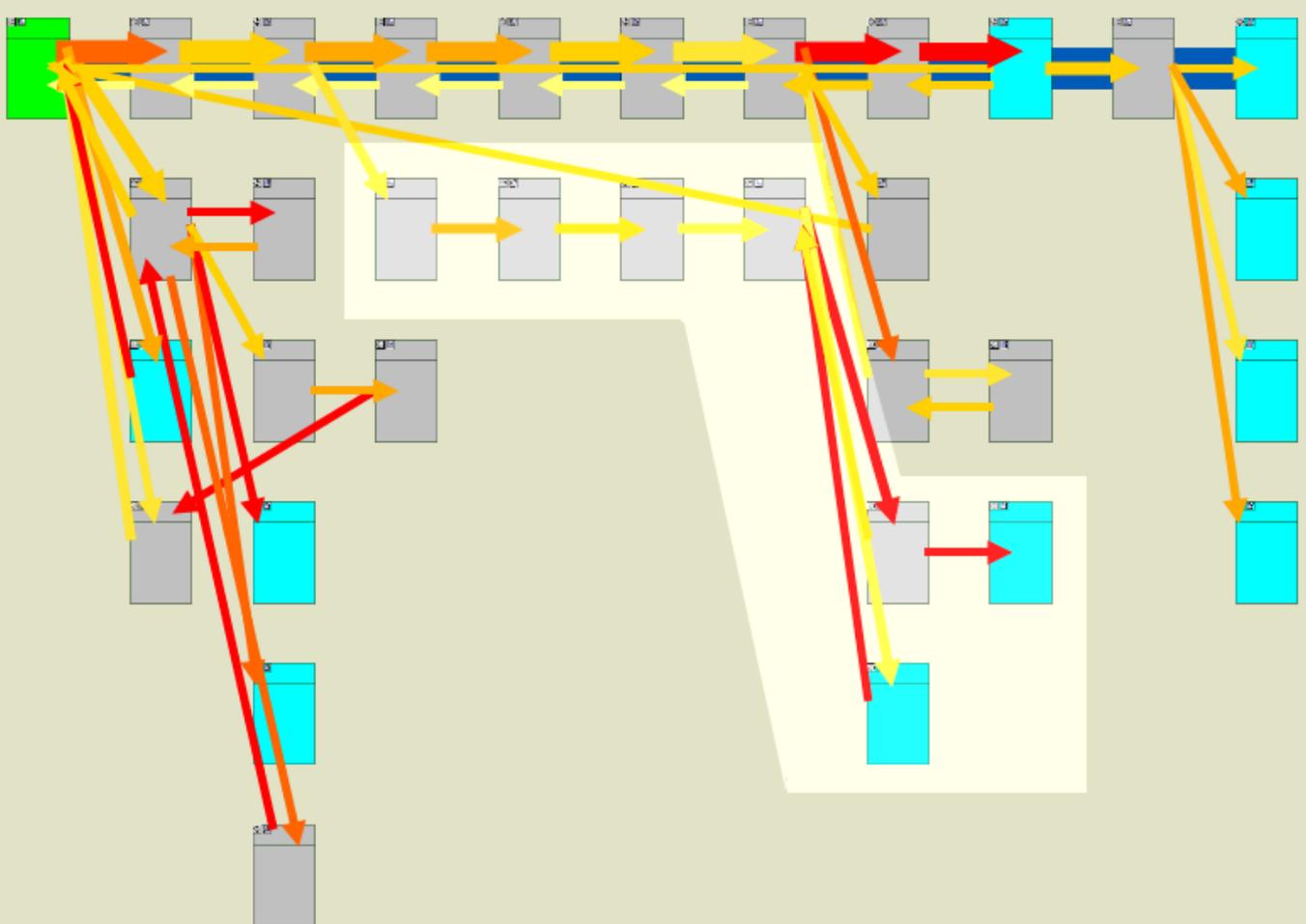
- 5 users were tested in the lab
- 5 were given the device and asked to perform the task at their convenience
- All task directions, demographic data, and follow up questionnaire data was presented and collected in web forms as part of the WebQuilt testing framework.



55379 88561 49983 52015 56678 38504 13681 67739 87310 14880

Update Hide Filter

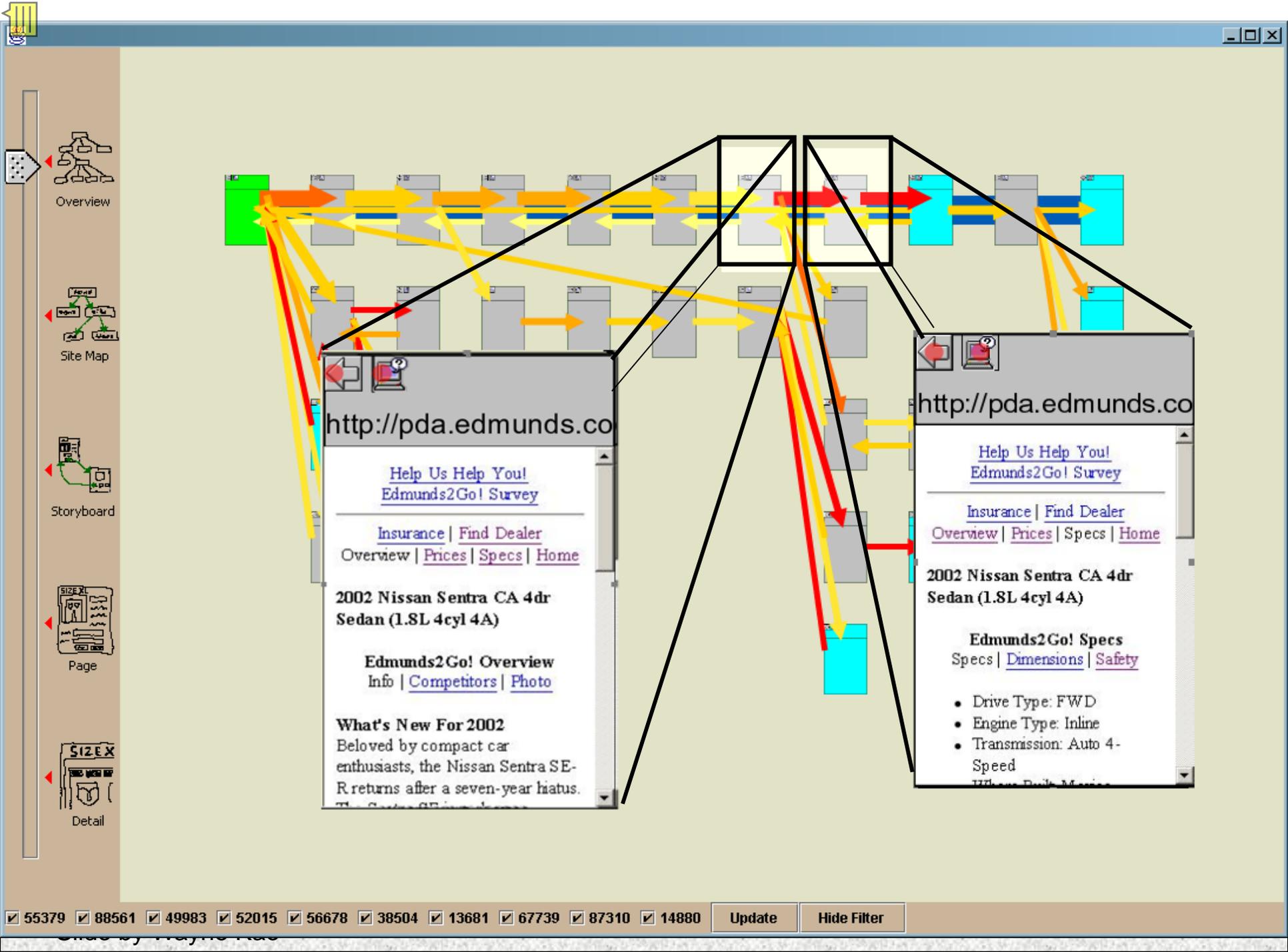
- Overview
- Site Map
- Storyboard
- Page
- Detail



55379  88561  49983  52015  56678  38504  13681  67739  87310  14880

Update

Hide Filter



Overview

Site Map

Storyboard

Page

Detail

http://pda.edmunds.co

[Help Us Help You!](#)  
[Edmunds2Go! Survey](#)

[Insurance](#) | [Find Dealer](#)  
[Overview](#) | [Prices](#) | [Specs](#) | [Home](#)

**2002 Nissan Sentra CA 4dr Sedan (1.8L 4cyl 4A)**

**Edmunds2Go! Overview**  
[Info](#) | [Competitors](#) | [Photo](#)

**What's New For 2002**  
Beloved by compact car enthusiasts, the Nissan Sentra SE-R returns after a seven-year hiatus.

http://pda.edmunds.co

[Help Us Help You!](#)  
[Edmunds2Go! Survey](#)

[Insurance](#) | [Find Dealer](#)  
[Overview](#) | [Prices](#) | [Specs](#) | [Home](#)

**2002 Nissan Sentra CA 4dr Sedan (1.8L 4cyl 4A)**

**Edmunds2Go! Specs**  
[Specs](#) | [Dimensions](#) | [Safety](#)

- Drive Type: FWD
- Engine Type: Inline
- Transmission: Auto 4-Speed

55379 88561 49983 52015 56678 38504 13681 67739 87310 14880

Update Hide Filter



# Findings

Browser	Device
<ul style="list-style-type: none"><li>• Interact before load (3)</li><li>• No forward button (2)</li></ul>	<ul style="list-style-type: none"><li>• Difficulty with input in questionnaire (3)</li><li>• <b>Difficulty scrolling</b> (2)</li><li>• Device errors unrelated to testing (1)</li><li>• Tried writing on screen (0)</li></ul>
Site Design	Test Design
<ul style="list-style-type: none"><li>• <b>Falsely completed task</b> (4)</li><li>• <b>Long download times</b> (4)</li><li>• <b>Ping-pong behavior</b> (3)</li><li>• Interact before load (3)</li><li>• <b>Too much scrolling</b> (2)</li><li>• Save address functionality not clear (1)</li><li>• <b>Back button navigation</b> (0)</li><li>• Would like more features (0)</li><li>• Finds site useful (0)</li></ul>	<ul style="list-style-type: none"><li>• <b>Falsely completed task</b> (4)</li><li>• <b>Difficulty remembering task description</b> (3)</li><li>• Difficulty with input in questionnaire (3)</li><li>• Questionnaire wording problems (3)</li><li>• Forgot how to end task (1)</li><li>• Confusing task description (1)</li></ul>

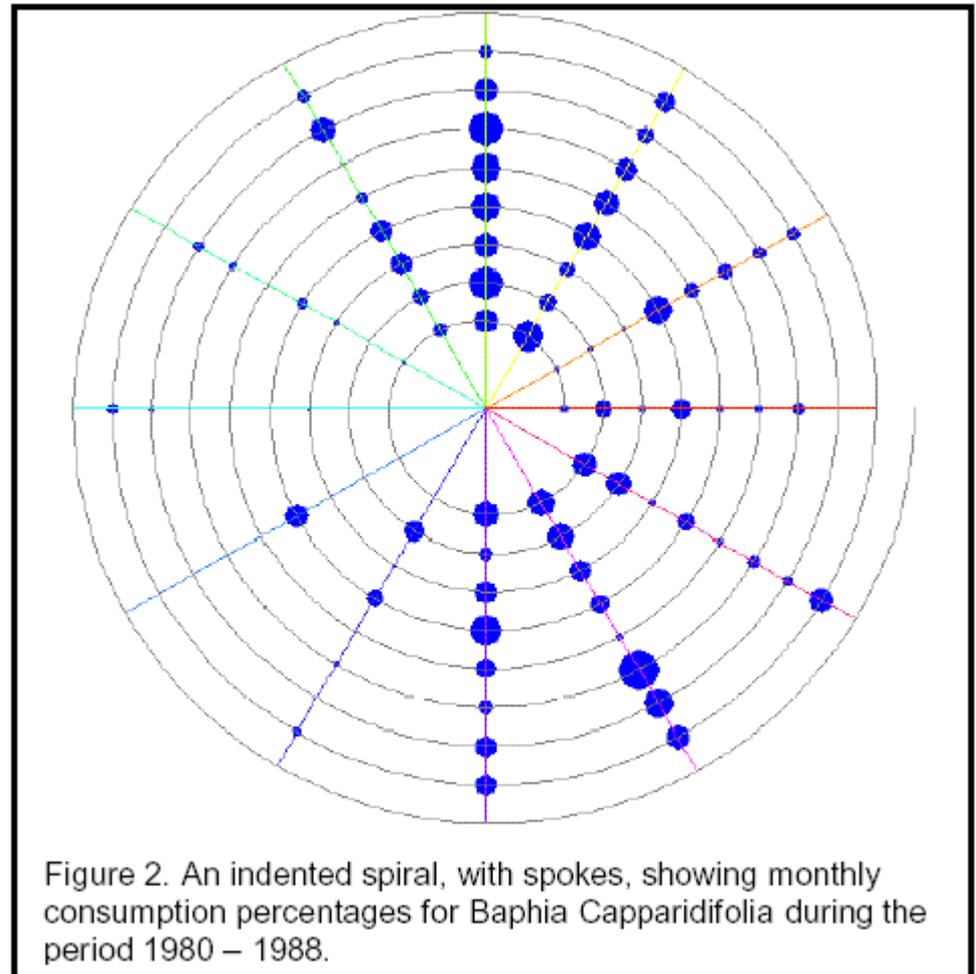
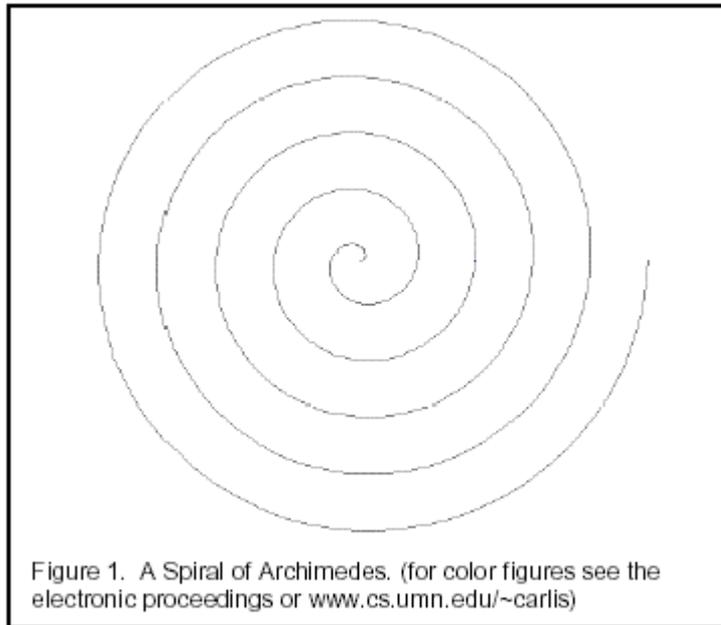
# Findings

- WebQuilt methodology is promising for uncovering site design related issues.
- 1/3 of the issues were device or browser related.
  - Browser and device issues can not be captured automatically with WebQuilt unless they cause an interaction with the server
  - Can be revealed via the questionnaire data.

# Visualization for Analysis

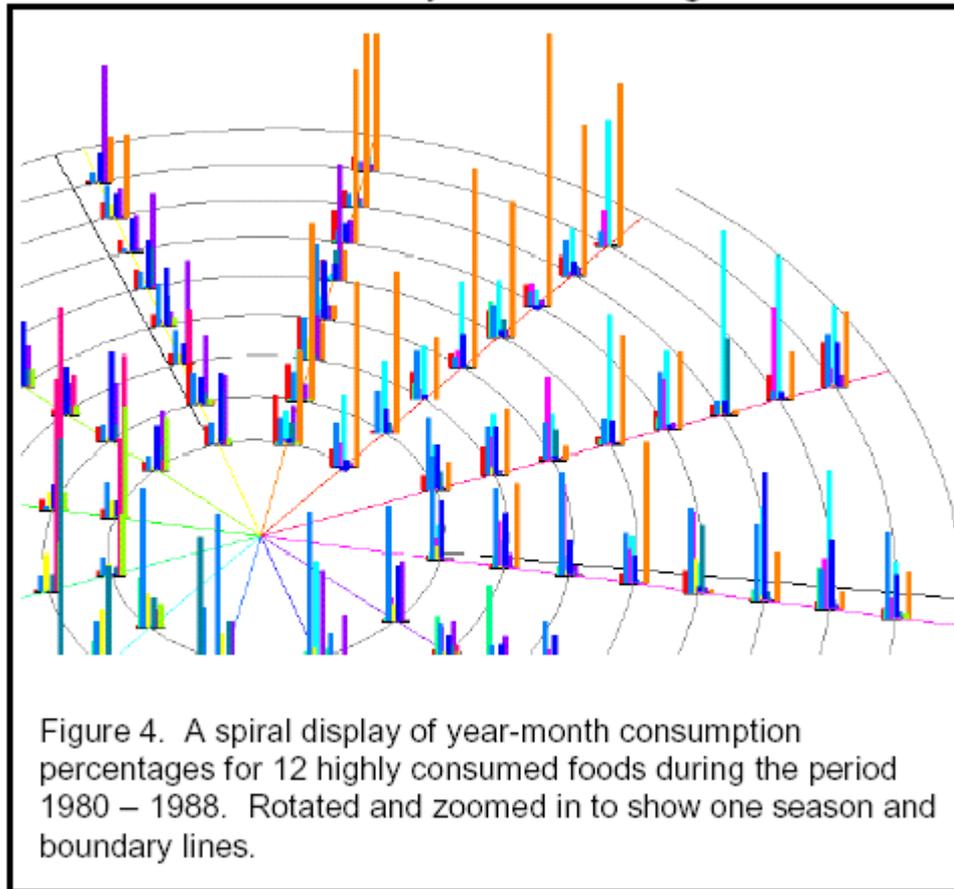
- Carlis & Konstan, UIST 1998
- Problem: data that is both periodic and serial
  - Time students spend on different activities
  - Tree growth patterns
    - Time: which year
    - Period: yearly
  - Multi-day races such as the Tour de France
  - Calendars arbitrarily wrap around at end of month
  - Octaves in music
- How to find patterns along both dimensions?

# Analyzing Complex Periodic Data



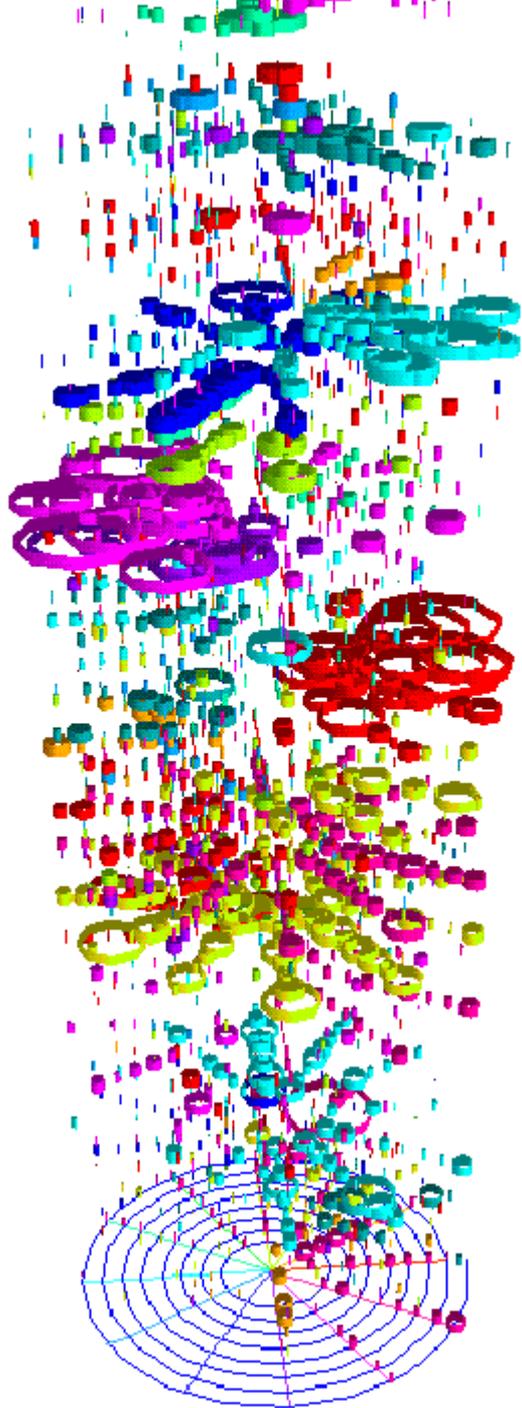
Carlis & Konstan, UIST 1998.

# Analyzing Complex Periodic Data



- Consumption values for each month appear as spikes
- Each food has its own color
- Boundary line (in black) shows when season begins/ends

Carlis & Konstan, UIST 1998.



Carlis & Konstan, UIST 1998.

# Visualization vs. Analysis?

- Applications to data mining and data discovery.
- Wesley Johnson '02:
  - Visualization tools are helpful for exploring hunches and presenting results
    - Examples: scatterplots
  - They are the WRONG primary tool when the goal is to find a good classifier model in a complex situation.
  - Need:
    - Solid insight into the domain and problem
    - Tools that visualize several alternative models.
    - Emphasize “model visualization” rather than “data visualization”

# Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up

# Visualizing Documents and Search

# Documents and Search

- Why Visualize Text?
- Why Text is Tough
- Visualizing Concept Spaces
  - Clusters
  - Category Hierarchies
- Visualizing Retrieval Results
- Usability Study Meta-Analysis

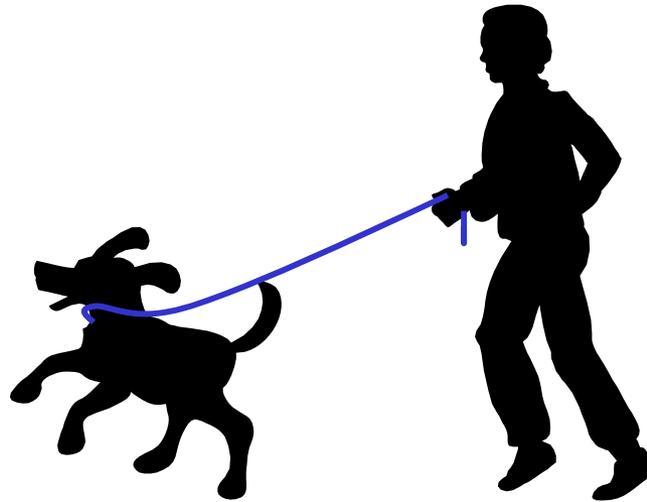
# Why Visualize Text?

- To help with Information Retrieval
  - give an overview of a collection
  - show user what aspects of their interests are present in a collection
  - help user understand why documents retrieved as a result of a query
- Text Data Mining
  - Mainly clustering & nodes-and-links
- Software Engineering
  - not really text, but has some similar properties

# Why Text is Tough

- Text is **not** pre-attentive
- Text consists of abstract concepts
  - which are difficult to visualize
- Text represents similar concepts in many different ways
  - space ship, flying saucer, UFO, figment of imagination
- Text has very high dimensionality
  - Tens or hundreds of thousands of features
  - Many subsets can be combined together

# Why Text is Tough



As the man walks the cavorting dog, thoughts arrive unbidden of the previous spring, so unlike this one, in which walking was marching and dogs were baleful sentinals outside unjust halls.

# Why Text is Tough

- Abstract concepts are difficult to visualize
- Combinations of abstract concepts are even more difficult to visualize
  - time
  - shades of meaning
  - social and psychological concepts
  - causal relationships

# Why Text is Tough

- Language only hints at meaning
- Most meaning of text lies within our minds and common understanding
  - “How much is that doggy in the window?”
    - how much: social system of barter and trade (not the size of the dog)
    - “doggy” implies childlike, plaintive, probably cannot do the purchasing on their own
    - “in the window” implies behind a store window, not really inside a window, requires notion of window shopping

# Why Text is Easy

- Text is highly redundant
  - When you have lots of it
  - Pretty much any simple technique can pull out phrases that seem to characterize a document
- Instant summary:
  - Extract the most frequent words from a text
  - Remove the most common English words
- People are very good at attributing meaning to lists of otherwise unrelated words

# Guess the Text:

10 PEOPLE  
10 ALL  
9 STATES  
9 LAWS  
8 NEW  
7 RIGHT  
7 GEORGE  
6 WILLIAM  
6 THOMAS  
6 JOHN  
6 GOVERNMENT  
5 TIME  
5 POWERS  
5 COLONIES  
4 LARGE  
4 INDEPENDENT  
4 FREE  
4 DECLARATION  
4 ASSENT  
3 WORLD  
3 WAR  
3 USURPATIONS  
3 UNITED  
3 SEAS  
3 RIGHTS

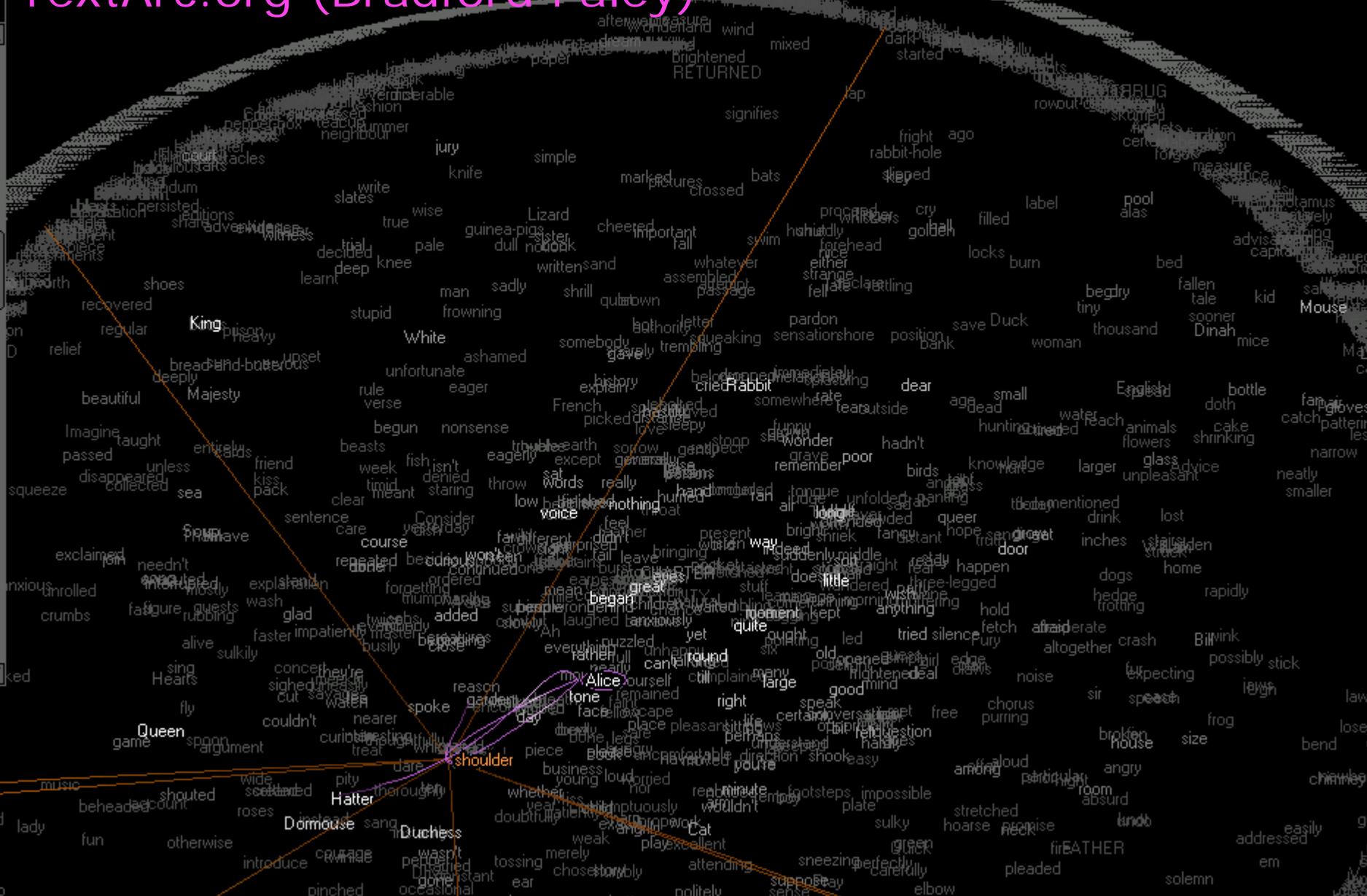
# Visualization of Text Collections

- How to summarize the contents of hundreds, thousands, tens of thousands of texts?
- Many have proposed clustering the words and showing points of light in a 2D or 3D space.
- Examples
  - Showing docs/collections as a word space
  - Showing retrieval results as points in word space

# Alice's Adventures in Wonderland

## TextArc.org (Bradford Paley)

Associations







# general weather

## Winter storm dumps more snow on weary region, storm brings landslides, slows traffic

general weather, weather, severe weather, trouble,

**NEW YORK (UPI)** -- Winter-weary metropolitan area residents battered down the hatches Tuesday and readied their rock salt as another snowstorm threatened to dump more than a foot of snow over the region.

The storm swept over northern New Jersey and southeastern New York including New York City and Long Island. Moderate to heavy snow spread across the entire area, with fast-falling snow piling up at a rate of two inches to three inches an hour in northern suburbs, the city and western Long Island.

By Monday afternoon, snow banks of nearly a foot were in some areas, blizzards were feared and schools and offices were closed.

severe weather, trouble, general weather, weather

**LOS ANGELES (UPI)** -- Rain and gusty winds hit southern California Monday, snarling traffic with flooding, landslides and avalanches on highways from the beach to the mountains.

Flooding closed portions of Pacific Coast Highway throughout the morning, and by mid-afternoon a torrent of mud and rock slid onto the roadway in Malibu, closing a 4-mile stretch of PCH in both directions between Topanga Canyon and Los Angeles Canyon.

## Weather Satellite Photo, Feb. 8, 1994

photo, weather, hazards

**US (AP)** -- The weather satellite photo taken at 5:30 a.m. EST, Tuesday, February 8, 1994 shows cloud cover over much of the country. A

## Winter storm dumps more snow on weary region

general weather, weather, severe weather, trouble,

**NEW YORK (UPI)** -- Winter-weary metropolitan area residents battered down the hatches Tuesday and readied their rock salt as another snowstorm threatened to dump more than a foot of snow over the region.

The storm swept over northern New Jersey and southeastern New York including New York City and Long Island. Moderate to heavy snow spread across the entire area, with fast-falling snow piling up at a rate of two inches to three inches an hour in northern suburbs, the city and western Long Island.

By Monday afternoon, snow banks of nearly a foot were in some areas, blizzards were feared and schools and offices were closed.

## Ol' man winter strikes again

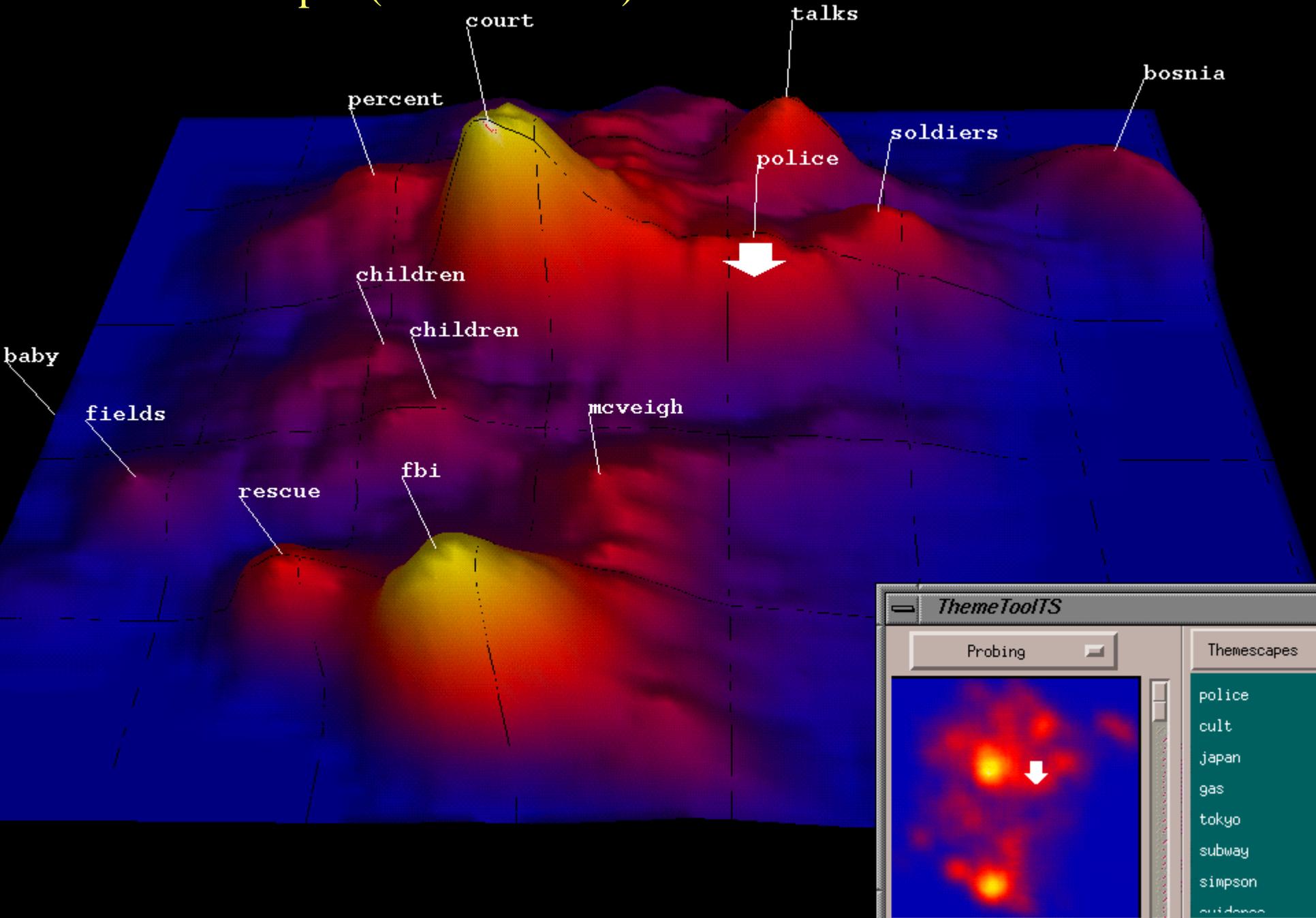
general weather, weather

were canceled at schools and colleges Tuesday.

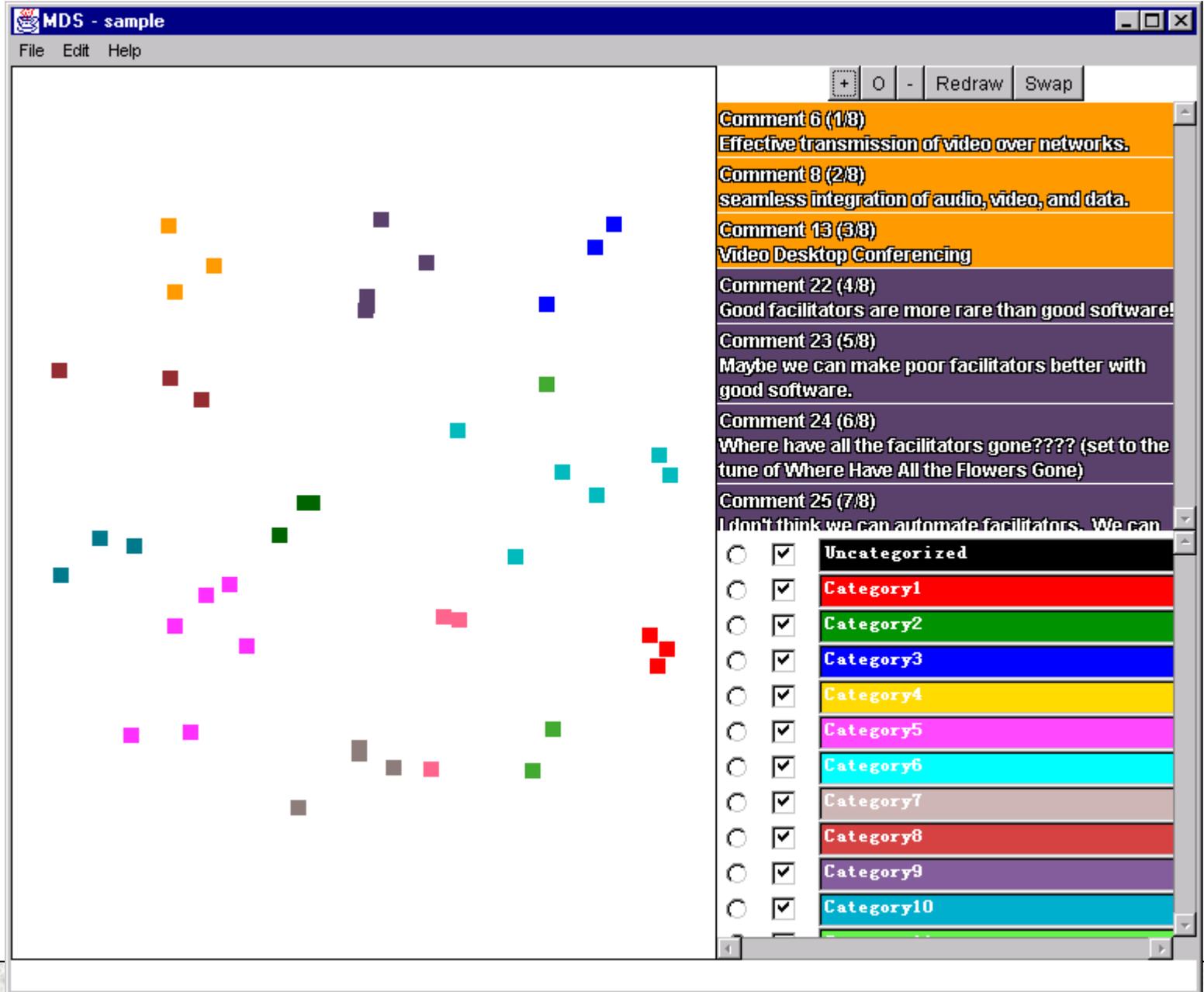
# weather

Galaxy of News  
Rennison 95

# Themescapes (Wise et al. 95)

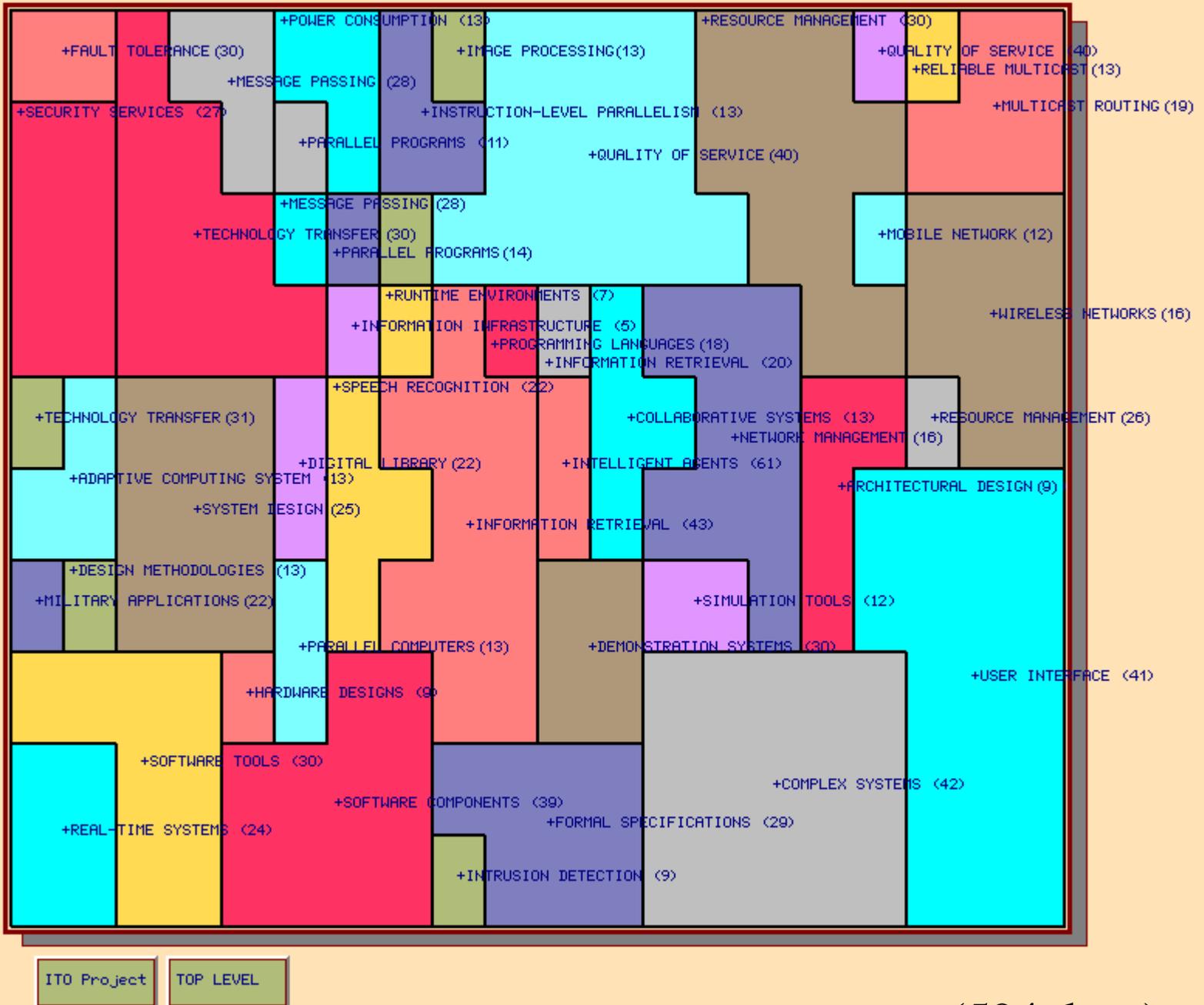


# ScatterPlot of Clusters (Chen et al. 97)



# Kohonen Feature Maps

(Lin 92, Chen et al. 97)



(594 docs)

# Clustering for Collection Overviews

- Two main steps
  - cluster the documents according to the words they have in common
  - map the cluster representation onto a (interactive) 2D or 3D representation
- Since text has tens of thousands of features
  - the mapping to 2D loses a tremendous amount of information
  - only very coarse themes are detected

# Scatter/Gather

Cluster 1 Size: 8 key army war francis spangle banner air song scott word poem british

- Star–Spangled Banner, The
- Key, Francis Scott
- Fort McHenry
- Arnold, Henry Harley
- Nilsen, Arthur

Cluster 2 Size: 68 film play career win television role record award york popular stage p

- Burstyn, Ellen
- Stanwyck, Barbara
- Berle, Milton
- Zukor, Adolph
- Deakins, Ted

Cluster 3 Size: 97 bright magnitude cluster constellation line type contain period spectr

- star
- Galaxy, The
- extragalactic systems
- interstellar matter
- cluster star

Cluster 4 Size: 67 astronomer observatory astronomy position measure celestial telescop

- astronomy and astrophysics
- astrometry
- Agena
- astronomical catalogs and atlases
- Hubble, Sir William

Cluster 5 Size: 10 family specie flower animal arm plant shape leaf brittle tube foot hor

- blazing star
- brittle star
- bishop's–cap
- feather star

# How Useful is Collection Cluster Visualization for Search?

Three studies find negative results

# Study 1

Kleiboemer, Lazear, and Pedersen. Tailoring a retrieval system for naive users. In Proc. of the 5th Annual Symposium on Document Analysis and Information Retrieval, 1996

- This study compared
  - a system with 2D graphical clusters
  - a system with 3D graphical clusters
  - a system that shows textual clusters
- Novice users
- Only textual clusters were helpful (and they were difficult to use well)

# Study 2: Kohonen Feature Maps

H. Chen, A. Houston, R. Sewell, and B. Schatz, *JASIS* 49(7)

- Comparison: Kohonen Map and Yahoo
- Task:
  - “Window shop” for interesting home page
  - Repeat with other interface
- Results:
  - Starting with map could repeat in Yahoo (8/11)
  - Starting with Yahoo unable to repeat in map (2/14)

# Study 2 (cont.)

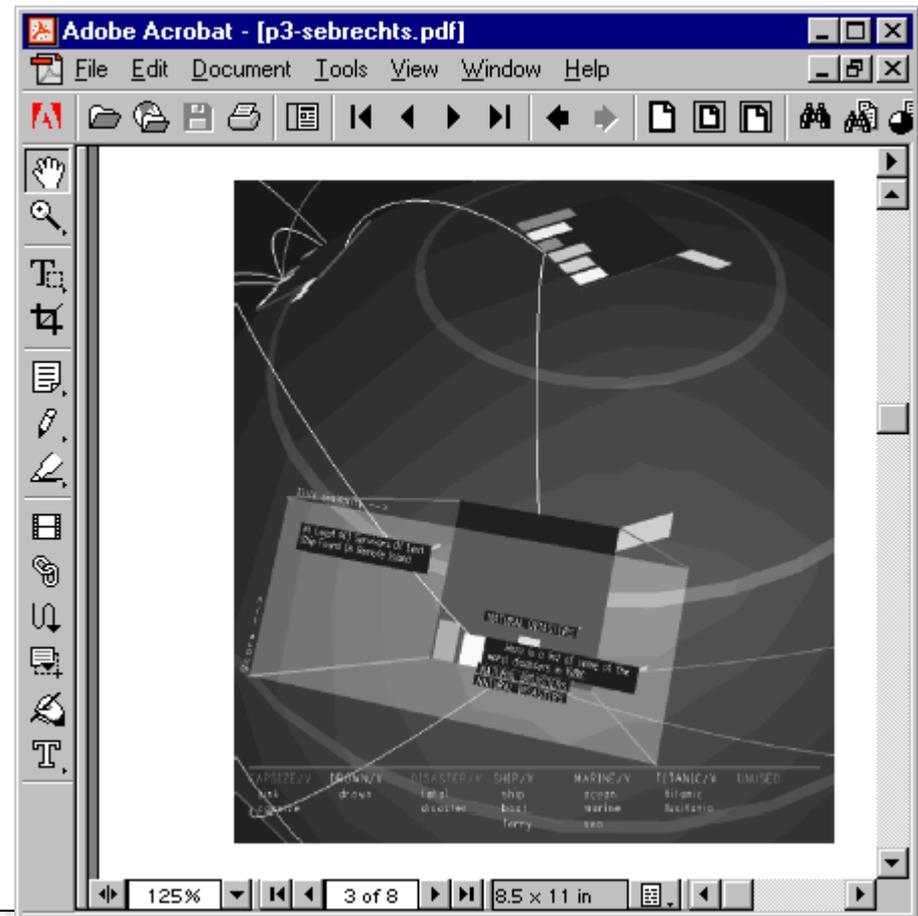
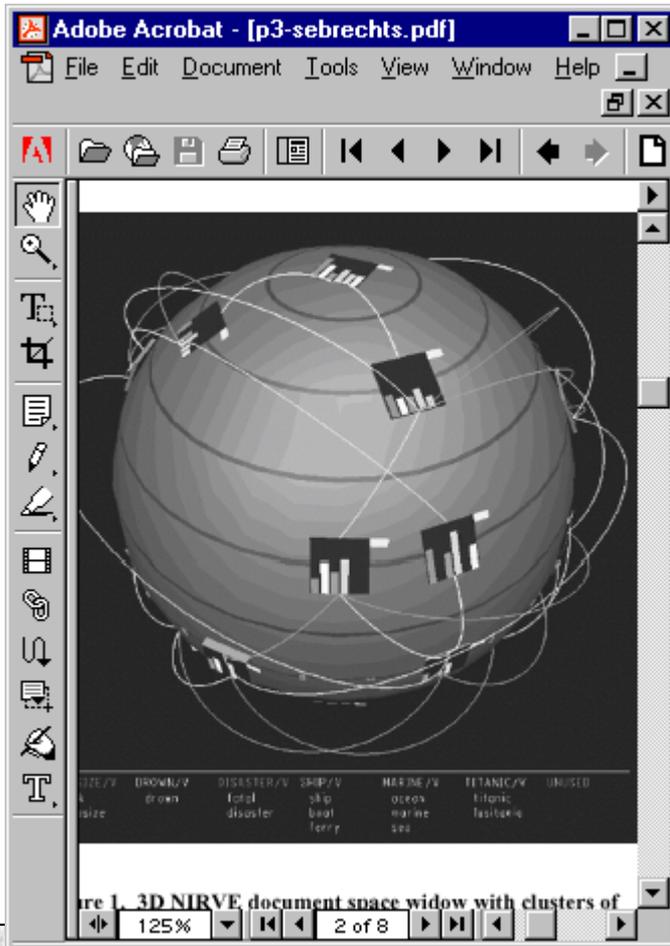
- Participants liked:
  - Correspondence of region size to # documents
  - Overview (but also wanted zoom)
  - Ease of jumping from one topic to another
  - Multiple routes to topics
  - Use of category and subcategory labels

# Study 2 (cont.)

- Participants wanted:
  - hierarchical organization
  - other ordering of concepts (alphabetical)
  - integration of browsing and search
  - correspondence of color to meaning
  - more meaningful labels
  - labels at same level of abstraction
  - fit more labels in the given space
  - combined keyword and category search
  - multiple category assignment (sports+entertain)

# Study 3: NIRVE

NIRVE Interface by Cugini et al. 96. Each rectangle is a cluster. Larger clusters closer to the "pole". Similar clusters near one another. Opening a cluster causes a projection that shows the titles.



# Study 3

Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces Sebrechts, Cugini, Laskowski, Vasilakis and Miller, Proceedings of SIGIR 99, Berkeley, CA, 1999.

- This study compared:
  - 3D graphical clusters
  - 2D graphical clusters
  - textual clusters
- 15 participants, between-subject design
- Tasks
  - Locate a particular document
  - Locate and mark a particular document
  - Locate a previously marked document
  - Locate all clusters that discuss some topic
  - List more frequently represented topics

# Study 3

- Results (time to locate targets)
  - Text clusters fastest
  - 2D next
  - 3D last
  - With practice (6 sessions) 2D neared text results; 3D still slower
  - Computer experts were just as fast with 3D
- Certain tasks equally fast with 2D & text
  - Find particular cluster
  - Find an already-marked document
- But anything involving text (e.g., find title) much faster with text.
  - Spatial location rotated, so users lost context
- Helpful viz features
  - Color coding (helped text too)
  - Relative vertical locations

# Summary: Visualizing Clusters

- Huge 2D maps may be inappropriate focus for information retrieval
  - cannot see what the documents are about
  - space is difficult to browse for IR purposes
  - (tough to visualize abstract concepts)
- Perhaps more suited for pattern discovery and gist-like overviews

# IR Infovis Meta-Analysis

(Empirical studies of information visualization:  
a meta-analysis, Chen & Yu IJHCS 53(5),2000)

- **Goal**
  - Find invariant underlying relations suggested collectively by empirical findings from many different studies
- **Procedure**
  - Examine the literature of empirical infoviz studies
    - 35 studies between 1991 and 2000
    - 27 focused on information retrieval tasks
    - But due to wide differences in the conduct of the studies and the reporting of statistics, could use only 6 studies

# IR Infovis Meta-Analysis

(Empirical studies of information visualization:  
a meta-analysis, Chen & Yu IJHCS 53(5),2000)

- **Conclusions:**
  - IR Infoviz studies not reported in a standard format
  - Individual cognitive differences had the largest effect
    - Especially on accuracy
    - Somewhat on efficiency
  - Holding cognitive abilities constant, users did better with simpler visual-spatial interfaces
  - The combined effect of visualization is not statistically significant

# So What Works?

- Yee, K-P et al., Faceted Metadata for Image Search and Browsing, to appear in *CHI 2003*. Hearst, M, et al.; Chapter 10 of *Modern Information Retrieval*, Baeza-Yates & Ribiero-Neto (Eds).
- Color highlighting of query terms in results listings
- Sorting of search results according to important criteria (date, author)
- Grouping of results according to well-organized category labels.
  - Cha-cha
  - Flamenco
- Only if highly accurate:
  - Spelling correction/suggestions
  - Simple relevance feedback (more-like-this)
  - Certain types of term expansion
- Note: most don't benefit from visualization!

# Cha-Cha

- Chen, M., Hearst, M., Hong, J., and Lin, J. Cha-Cha: A System for Organizing Intranet Search Results in the *Proceedings of the 2nd USENIX Symposium on Internet Technologies and SYSTEMS (USITS)*, Boulder, CO, October 11-14, 1999

SEARCH RESULTS BY: **Cha-Cha**

1-25 of 2239 matches [List View](#)

## Table of Contents

[SIMS: School of Information Management and Systems, ...](#)

- ▼ [SIMS: Site Map](#)
  - ▼ [SIMS: Fall 2000 Course Schedule](#)
    - ▼ [SIMS 247 Home Page](#)
      - ▼ [Homepage for Marti \*\*Hearst\*\*](#)
        -  [Selected Recent Publications by Marti \*\*Hearst\*\* and ...](#)

[BAILANDO Project Homepage](#)

-  [BAILANDO Project: Publications](#)

[UC Berkeley Seismo and Earth and Planetary ...](#)

- ▼ [UC Berkeley Seismological Laboratory Home Page](#)
  - ▼ [Geological Tour of UC - Berkeley](#)
    -  [Geological Tour of UC - Berkeley: \*\*Hearst\*\* ...](#)
    -  [Geological Tour of UC - Berkeley: \*\*Hearst\*\* ...](#)

[Phoebe A. \*\*Hearst\*\* Museum of Anthropology, Home ...](#)

- ▼ [News and Events Archive](#)
  -  [\*\*Hearst\*\* Foundation Augments Museum Endowment](#)
  -  [The \*\*Hearsts\*\* New Director](#)

# Teoma: appears to combine categories and clusters

(this version before it was bought by askjeeves)

TEOMA™

Search

text mining

Search

Find this phrase

Help

## Search Results for **text mining**

### WEB PAGES GROUPED BY TOPIC:

- Text Data Mining
- Computer, Science
- Text Mining Software
- Workshop, Knowledge
- Colorado History Videos
- Machine Learning
- Research, Papers

### WEB PAGES

1-10 of 142000 matched:

[Next>>](#)

- **WEBSOM - A novel SOM-based approach to free-text mining**  
WEBSOM - Self-Organizing Maps for Internet Exploration  
Welcome to test the document exploration tool WEBSOM. An ordered map of the information space...  
<http://websom.hut.fi/websom/>
- **Data Mining, Text Mining and Web Mining Software**  
Megaputer offers data mining, **text mining**, and web data mining software tools for e-commerce, database marketing, and CRM; seminars, training and...  
<http://www.megaputer.com/> --Related Topic & Experts' Links

### EXPERTS' LINKS ([What are these?](#))

1-10 matched:

- **Resources**  
<http://textmining.krdl.org.sg/reso...>
- **The Data Warehousing Information Center -**  
<http://www.dwinfocenter.org/docum...>
- **Software for Text Analysis**  
[http://www.kdnuggets.com/software/...](http://www.kdnuggets.com/software/)
- **text mining and web-based info...**  
<http://www-personal.umich.edu/~wfa...>

# Teoma: Now in prime time



[Search Tips](#)

- [Advanced Search](#)
- [Preferences](#)

 Find this phrase

---

## Results

Relevant web pages

Showing 1-10 of about 291,000

[TEMIS - Text Mining Solutions](#)  
Text Mining Technology and Consulting company offering software components ("Insight Discoverer") for the efficient analysis of large document...  
[www.temis-group.com/](http://www.temis-group.com/)

[Untangling Text Data Mining](#)  
Defines data **mining**, information access, and corpus-based computational linguistics, and then discusses the relationship of these to **text** data...  
[www.sims.berkeley.edu/~hearst/papers/acl99...](http://www.sims.berkeley.edu/~hearst/papers/acl99...)  
[\[Related Pages\]](#)

[Hearst, Marti](#)  
Untangling **Text Data Mining**, ACL'99.  
[www.sims.berkeley.edu/~hearst/](http://www.sims.berkeley.edu/~hearst/)  
[\[Related Pages\]](#)  
[\[More results from www.sims.berkeley.edu\]](#)

[Text Mining at Waikato](#)  
The **Text Mining** group at the University of Waikato in New Zealand. With a focus on Viterbi search and entropy-based methods the group has a...  
[www.cs.waikato.ac.nz/~nzdl/textmining/](http://www.cs.waikato.ac.nz/~nzdl/textmining/)  
[\[Related Pages\]](#)  
[\[More results from www.cs.waikato.ac.nz\]](#)

[KRDL Text Mining Home Page](#)  
Text Mining Research Group at Kent Ridge Digital Labs in Singapore;  
charset=iso-8859-1...  
[textmining.krdl.org.sg/](http://textmining.krdl.org.sg/)  
[\[Related Pages\]](#)

## Refine

Suggestions to narrow your search

- [Text Data Mining](#)
- [Computer Linguistics](#)
- [Text Mining Workshop](#)
- [Plain Text](#)
- [Text Mining Research](#)
- [Computer Science](#)

[\[Show All Refinements\]](#)

## Resources

Link collections from experts and enthusiasts

- [A Roadmap to Text Mining and Web Mining](#)  
[www.cs.utexas.edu/...](http://www.cs.utexas.edu/...)
- [Resources](#)  
[textmining.krdl.org.sg/...](http://textmining.krdl.org.sg/...)
- [Mary D. Taffet's Home Page: WWW Sites for Students...](#)  
[web.syr.edu/...](http://web.syr.edu/...)
- [Text Analysis, Text Mining and...](#)



# Better to reduce the viz

- Flamenco – allows users to steer through the category space
- Uses
  - Dynamically-generated hypertext
  - Color for distinguishing and grouping
  - Careful layout and font choices
- Focused first on the users' needs

# Flamenco Image Search

## Media

[aquatint](#) (2025)  
[basketry](#) (44)  
[book](#) (666)  
[ceramic](#) (1008)  
[costume](#) (660)  
[decorative box](#) (163)  
[domestic object](#) (176)

[drawing](#) (2624)  
[drypoint](#) (2143)  
[etching](#) (9507)  
[furnishing](#) (127)  
[glass](#) (651)  
[more...](#)

## Nature

[animal material](#) (515)  
[birds](#) (1437)  
[bodies of water](#) (3604)  
[creatures](#) (801)  
[fish](#) (219)  
[flowers](#) (1220)  
[geological formations](#) (2122)

[heavens](#) (2353)  
[hoofed mammals](#) (2480)  
[invertebrates and arthropods](#) (330)  
[mammals](#) (2116)  
[plant material](#) (788)  
[more...](#)

## Location

[Africa](#) (463)  
[Asia](#) (1325)  
[Australia](#) (21)  
[Central America](#) (134)  
[Europe](#) (23331)

[Middle East](#) (78)  
[North America](#) (11111)  
[Oceania](#) (111)  
[Roman Empire](#) (4)  
[South America](#) (453)

## Places and Spaces

[bridges](#) (592)  
[building parts](#) (3088)  
[buildings](#) (2393)  
[dwellings](#) (1709)

[lawn](#) (20)  
[open spaces](#) (1732)  
[roads](#) (1480)  
[workplaces](#) (753)

## Date

[1 - 1000 A.D.](#) (138)  
[12th century](#) (3)  
[13th century](#) (1)  
[14th century](#) (3)  
[15th century](#) (76)  
[16th century](#) (1225)  
[17th century](#) (3058)

[18th century](#) (2287)  
[19th century](#) (7552)  
[20th - 21th century](#) (18)  
[20th century](#) (14295)  
[21st century](#) (12)  
[more...](#)

## People

[aristocrats](#) (974)  
[children](#) (2501)  
[men](#) (7372)

[occupations](#) (715)  
[women](#) (5906)

## Shapes, Colors, and Materials

[colors](#) (5861)  
[decorations](#) (1441)  
[fabrics](#) (345)  
[metal](#) (273)

[paper](#) (457)  
[shapes](#) (2752)  
[visual framing](#) (5911)

## Themes

# Flamenco Image Search

Refine your search further within these categories:

## Media

[aquatint](#) (3), [drawing](#) (15), [drypoint](#) (2), [etching](#) (29), [lithograph](#) (12), [mezzotint](#) (14), [painting](#) (1), [photograph](#) (1), [print](#) (10), [wood engraving](#) (4), [more...](#)

## Location [\(group results\)](#)

[Asia](#) (7), [Europe](#) (69), [North America](#) (13)

## Date: [all](#) > 19th century [\(group results\)](#)

[1800 - 1809](#) (3), [1810 - 1819](#) (8), [1820 - 1829](#) (7), [1830 - 1839](#) (6), [1840 - 1849](#) (2), [1850 - 1859](#) (6), [1860 - 1869](#) (8), [1870 - 1879](#) (7), [1880 - 1889](#) (8), [1890 - 1899](#) (10), [more...](#)

## Themes [\(group results\)](#)

[military](#) (5), [mortality](#) (4), [music, writing, and sport](#) (21), [nautical](#) (12), [religion](#) (18)

## Objects [\(group results\)](#)

[clothing](#) (4), [containers](#) (2), [food](#) (5), [furnishings](#) (4), [jewelry and riches](#) (1), [vehicles](#) (2)

## Nature: [all](#) > [heavens](#) > [cloud](#) [\(group results\)](#)

## Places and Spaces [\(group results\)](#)

[bridges](#) (6), [building parts](#) (7), [buildings](#) (9), [dwellings](#) (9), [open](#)

These terms define your current search. Click the  to remove a term.

Date: 19th century

Nature: [heavens](#) > [cloud](#)

Search

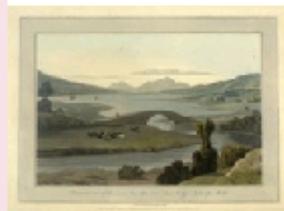
all items  within current results

[start a new search](#)

89 items (grouped by media)

[view ungrouped items](#)

## [aquatint](#) 3



Distant View of ...  
Daniell  
1817



Verlassen (Aband...  
Klinger  
1884



Village au Bord ...  
Hervier  
19th century

## [drawing](#) 15



A Lady at a Shrine  
Anonymous



A View of the No...  
Eustache



Alsatian Landscape  
Anonymous



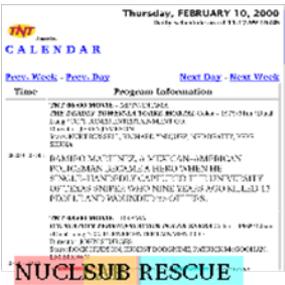
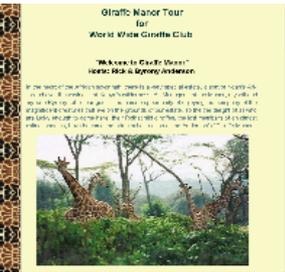
Boar Avatar of V...  
Anonymous

# Using Thumbnails to Search the Web

A. Woodruff, R. Rosenholtz, J. Morrison, A. Faulring, & P. Pirolli, A comparison on the use of text summaries, plain thumbnails, and enhanced thumbnails for web search tasks. *JASIST*, 53(2), 172-185, 2002.; A. Woodruff, A. Faulring, R. Rosenholtz, J. Morrison, & P. Pirolli, Using thumbnails to search the web. SIGCHI 2001

## Design Goals

- Enhance features that help the user decide whether document is relevant to their query
  - Emphasize text that is relevant to query
    - Text callouts
  - Enlarge (make readable) text that might be helpful in assessing page
    - Enlarge headers



# Text and Image Summaries

- Text summaries
  - Lots of abstract, semantic information
- Image summaries (plain thumbnails)
  - Layout, genre information
  - Gist extraction faster than with text
- Benefits are complementary
- Create textually-enhanced thumbnails that leverage the advantages of both text summaries and plain thumbnails

# Putting Callouts in a Separate Visual Layer

- Transparency
- Occlusion

Junctions indicate the occurrence of these events.



# Design Issues:

- Color Management

- Problems: Callouts need to be both readable and draw attention
- Solution: Desaturate the background image, and use a visual search model to choose appropriate colors
- Colors look like those in highlighter pens

- Resizing of Text

- Problem: We want to make certain text elements readable, but not necessarily draw attention to them
- Solution: Modify the HTML before rendering the thumbnail

# Examples

**HOG WILD?**

**New Hybrid electric car tops EPA mileage**

**Car**

The new model for the hybrid is presented by a sleek, white, of-the-century design, which is a perfect fit for a car that is designed to be a hybrid of a car and a car.

**liquidprice**

**MiniDisc Player**

Product Description

Choose from more than 100 items. The MiniDisc Player is a portable, high-quality audio player with 100MB capacity.

**MiniDisc Player**

**Jose Hellerstein**

Researcher

**Keyword Research (Publication Title)**

**Selected Papers**

- Factorial Control in the Single-Delay Queue Problem (Full Text) (1993) [PDF]
- Factorial Control in the Single-Delay Queue Problem (Full Text) (1993) [PDF]
- Factorial Control in the Single-Delay Queue Problem (Full Text) (1993) [PDF]
- Factorial Control in the Single-Delay Queue Problem (Full Text) (1993) [PDF]

**HALCION**

**side effects**

**Halcioside**

**fects**

**Halcion**

# Tasks

- Criteria: tasks that...
  - Are representative of common queries
  - Have result sets with different characteristics
  - Vary in the number of correct answers
- 4 types of tasks
  - Picture:* "Find a picture of a giraffe in the wild."
  - Homepage:* "Find Kern Holoman's homepage."
  - Side-effects:* "Find at least three side effects of halcion."
  - E-commerce:* "Find an e-commerce site where you can buy a DVD player. Identify the price in dollars."

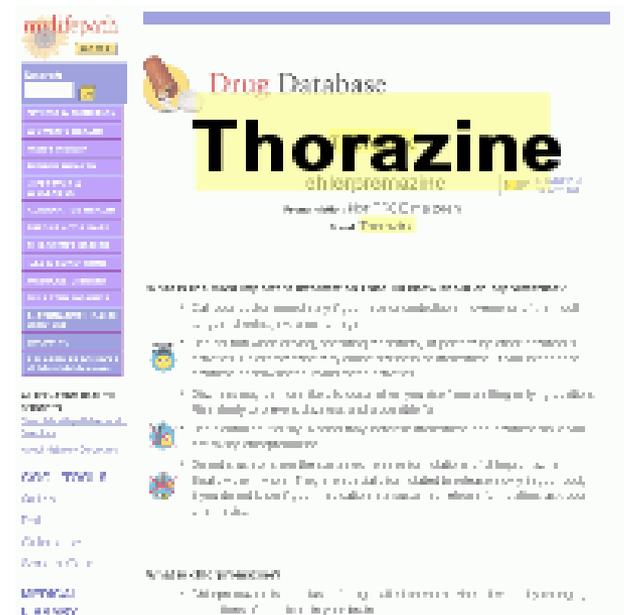
# Conditions

- Text summary
  - Page title
  - Extracted text with query terms in bold
  - URL
- Plain thumbnail
- Enhanced thumbnail
  - Readable H1, H2 tags
  - Highlighted callouts of query terms
  - Reduced contrast level in thumbnail

## [The Lycaenum -- GHB](#)

.....seems to be without serious **side effects**." His almost off-hand...  
...recovery with no long-term **side effects** is universal." They...

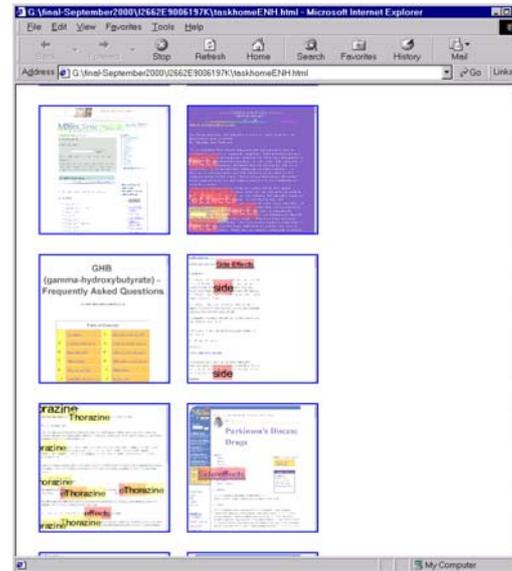
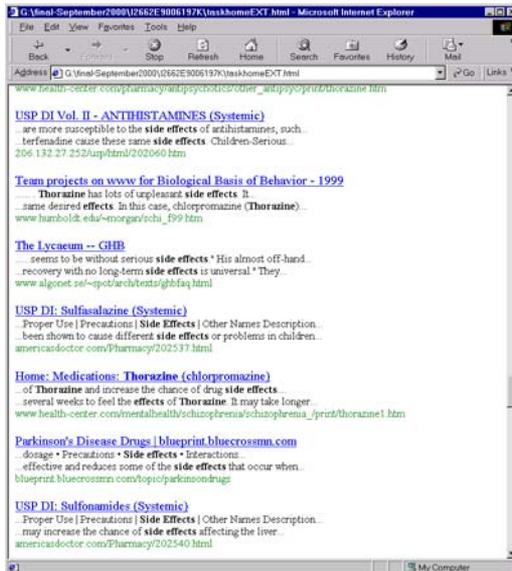
[www.algonet.se/~spot/arch/texts/ghbfaq.html](http://www.algonet.se/~spot/arch/texts/ghbfaq.html)



# Collections of Summaries

- 100 results in random order

Approximately same number of each summary type on a page



# Method

**18 questions, with 100 query results each**

**Entire process took about 75 minutes**

- Procedure

- 6 practice tasks
- 3 questions for each of the 4 task types
  - e.g., each participant would do one E-commerce question using text, one E-commerce question using plain thumbnails, and one E-commerce question using enhanced thumbnails
- Questions blocked by type of summary
- WebLogger recorded user actions during browsing
- Semi-structured interview

- Participants

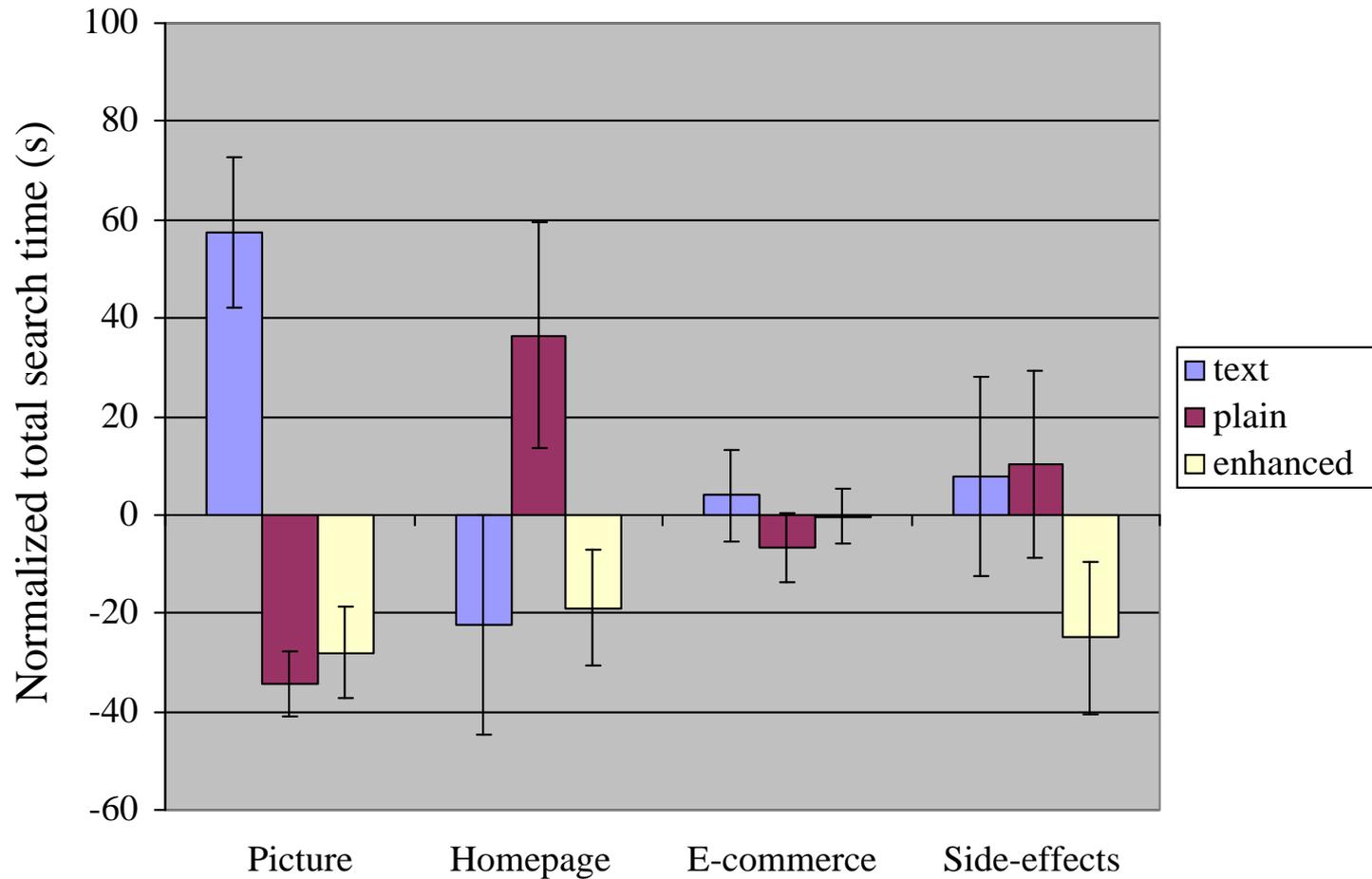
- 12 members of the PARC community



# Results

- Average total search times, by task:
  - Picture: 61 secs
  - Homepage: 80 secs
  - E-commerce: 64 secs
  - Side effects: 128 secs
- Results pooled across all tasks:
  - Subjects searched 20 seconds faster with enhanced thumbnails than with plain
  - Subjects searched 30 seconds faster with enhanced thumbnails than with text summaries
  - Mean search time overall was 83 seconds

# Results



# Results: User Responses

- Participants preferred enhanced thumbnails
  - 7/12 preferred overall
  - 5/12 preferred for certain task types
- Enhanced thumbnails are intuitive and less work than text or plain thumbnails
  - One subject said searching for information with text summaries did not seem hard until he used the enhanced thumbnails.
- Many participants reported using genre information, cues from the callouts, the relationship between search terms, etc.

# Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up

# Comparing Approaches

# Comparing 3 Commercial Systems

## Alfred Kobsa, An Empirical Comparison of Three Commercial Information Visualization Systems, INFOVIS'01.



Figure 1a. A screenshot from Eureka that shows how a user might solve the question "Did males cheat more on their girlfriends than females on their boyfriends?" (an actual question used in the experiment). After grouping the attribute "Gender" and sorting the column "Did you cheat?," a user can compare the number of "Yes" entries and thus find that more females than males indicated having cheated. (One male and one female gave no answers.)

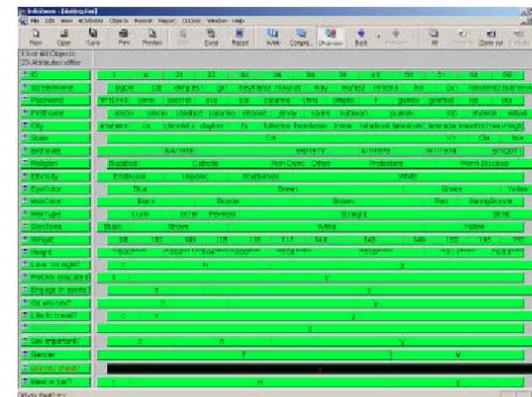


Figure 1b. This screenshot shows one possible way for solving the same problem in InfoZoom, specifically in its overview mode. After clicking at, and thereby zooming into, the "Yes" entries in the attribute "Did you cheat?," users can see from the length of the bars in the Gender category that females indicated more frequently having cheated than males.

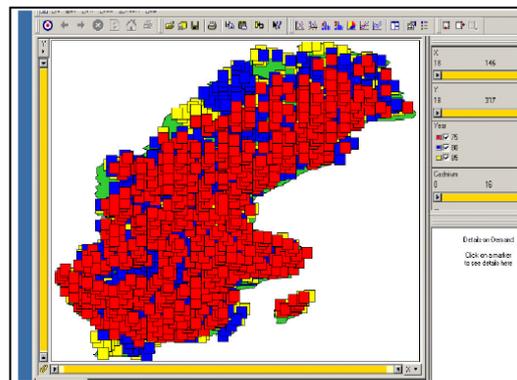


Figure 2. Spotfire's geographical representation of heavy metal concentrations through a scatterplot diagram.

# Comparing 3 Commercial Systems

## Eureka (InXight)

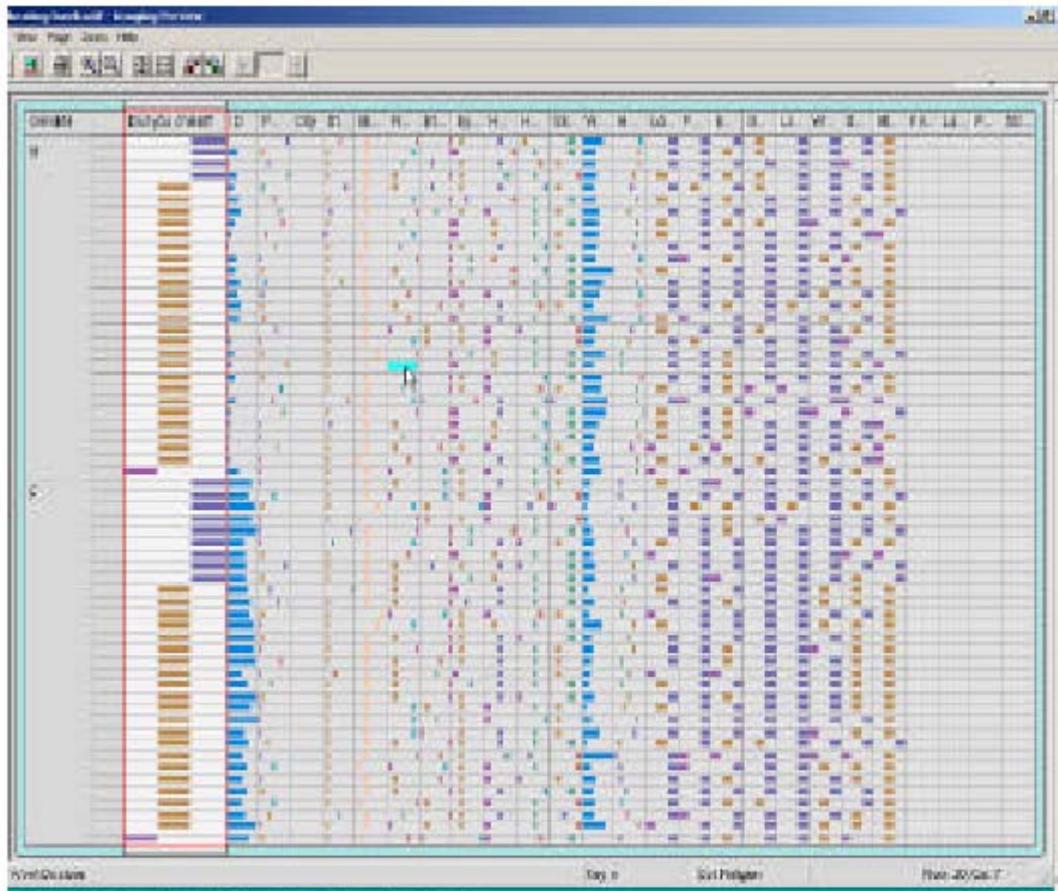


Figure 1a. A screenshot from Eureka that shows how a user might solve the question “Did males cheat more on their girlfriends than females on their boyfriends?” (an actual question used in the experiment). After grouping the attribute “Gender” and sorting the column “Did you cheat?,” a user can compare the number of “Yes” entries and thus find that more females than males indicated having cheated. (One male and one female gave no answers.)

# Comparing 3 Commercial Systems

## InfoZoom (HumanIT)

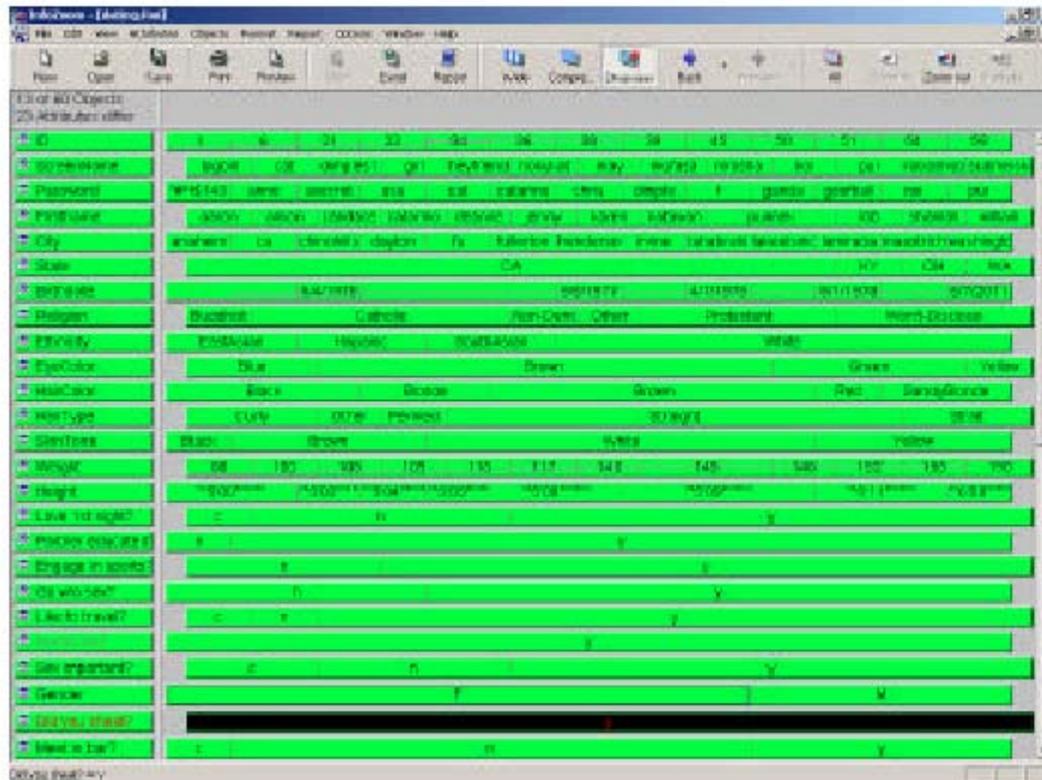
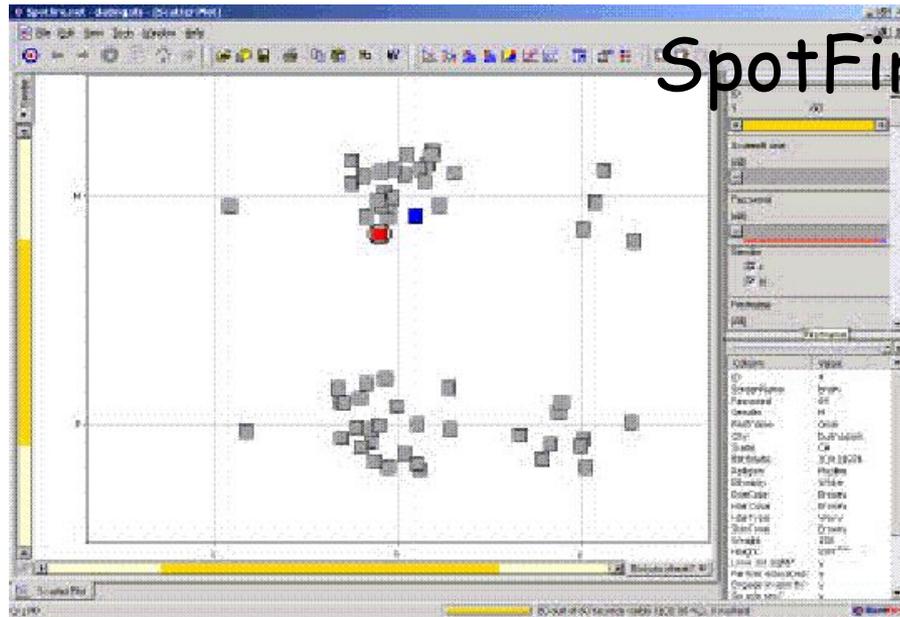
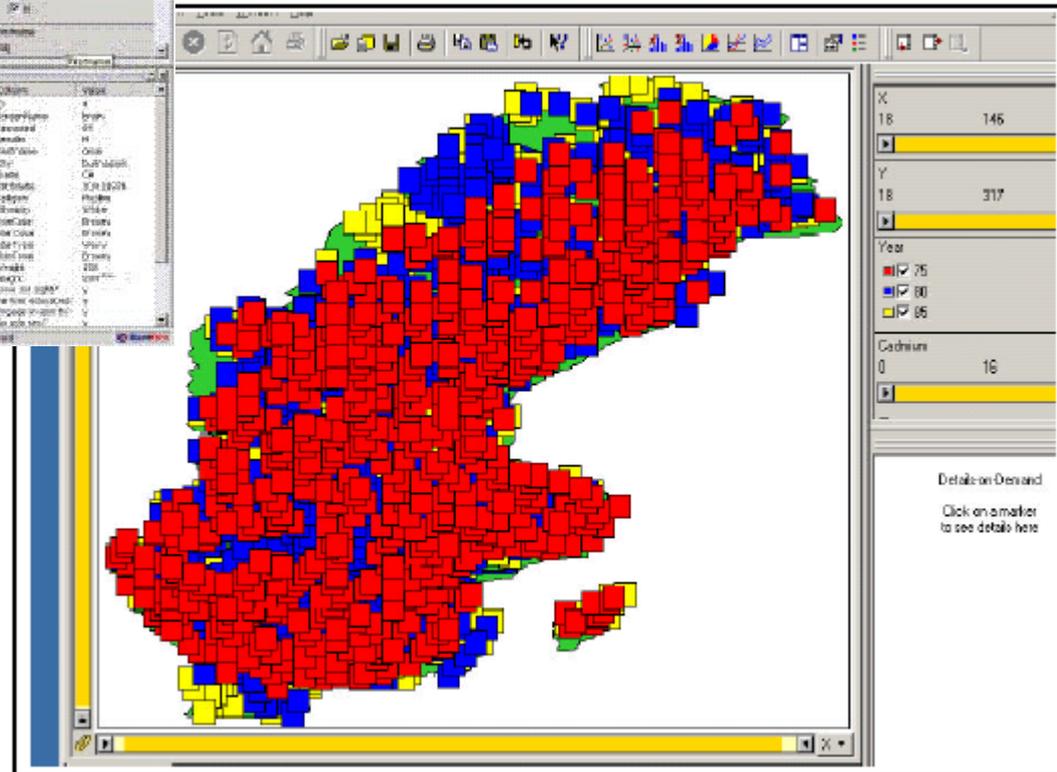


Figure 1b. This screenshot shows one possible way for solving the same problem in InfoZoom, specifically in its overview mode. After clicking at, and thereby zooming into, the “Yes” entries in the attribute “Did you cheat?”, users can see from the length of the bars in the Gender category that females indicated more frequently having cheated than males.

# Comparing 3 Commercial Systems



SpotFire



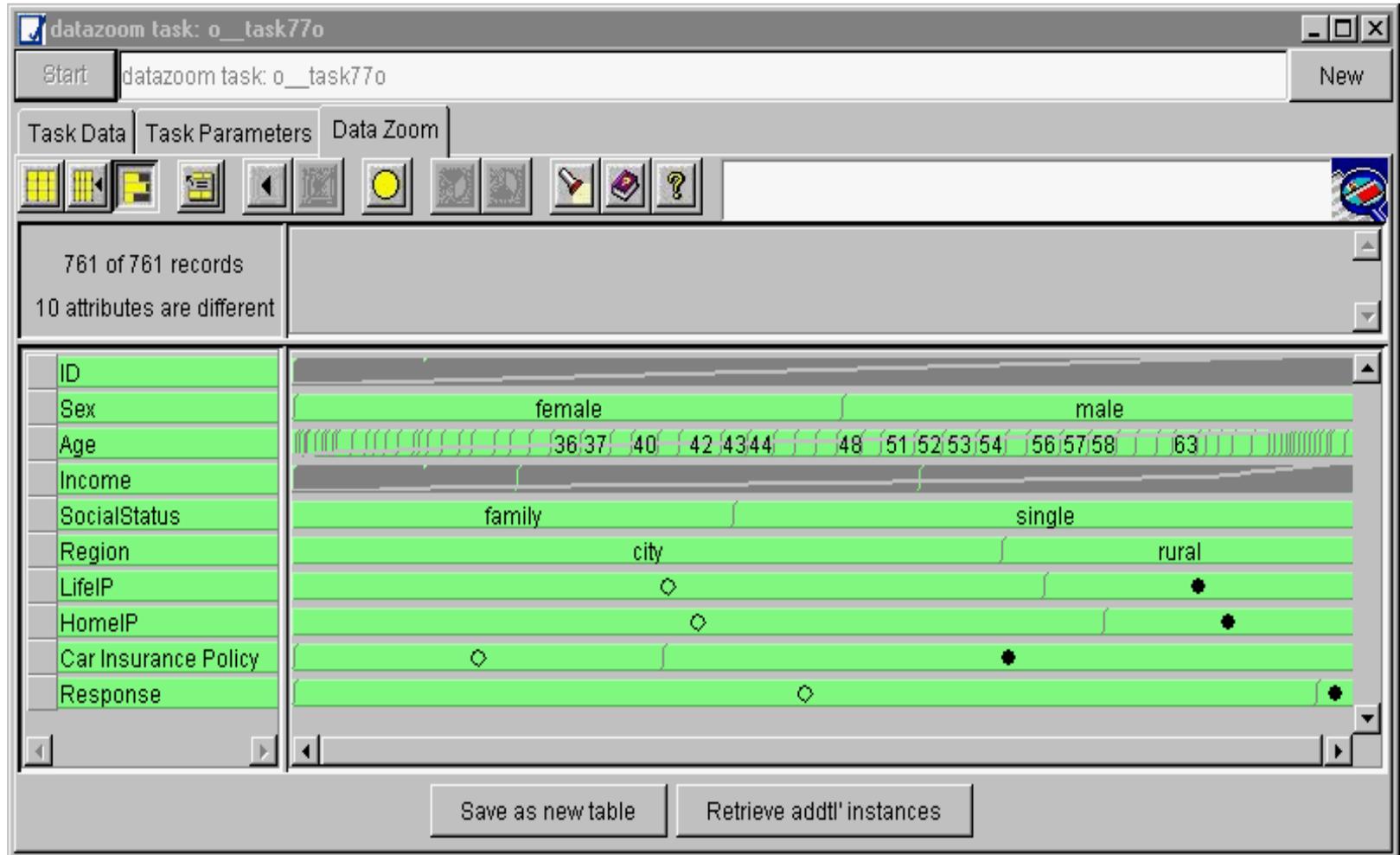
# Infozoom Overview

- Presents data in three different views.
  - Wide view shows data set in a table format.
  - Compressed view packs the data set horizontally to fit the window width.
  - Overview mode has all attributes in ascending or descending order and independent of each other.

# InfoZoom Overview View

The screenshot shows the InfoZoom application window titled "InfoZoom - [Films.fox]". The menu bar includes File, Edit, View, Attributes, Objects, Format, Report, Options, Window, and Help. The toolbar contains icons for New, Open, Save, Print, Preview, Chart, Excel, Report, Wide, Compr..., Overvi..., Back, Forward, All, Zoom in, Zoom..., Exclude, and Search. Below the toolbar is a query input field containing "The actors with the most films" and a "Perform" button. The main area displays a data grid with filters on the left. The filters include Film, Title, Year, Length, Subject, Actor, Actress, Director, Popularity, and Awards. The data grid shows columns for letters A through W, and rows for various attributes like Year (1986-1992), Length (60, 90, 91, 94, 95, 96, 97), Subject (Action, Comedy, Drama, Horror, Mystery, Sci-Fi, Westerns), Actor, Actress, Director, Popularity (2-8), and Awards (No, Yes).

# InfoZoom Overview View



# InfoZoom Compressed Table View

The screenshot displays the InfoZoom application window titled "InfoZoom - [Candy Corporations.fop]". The menu bar includes File, Edit, View, Attributes, Objects, Format, Report, Options, Window, and Help. The toolbar contains icons for New, Open, Save, Print, Preview, Chart, Excel, Report, Wide, Compr..., Overvi..., Back, Forward, All, Zoom in, Zoom..., Exclude, and Search. Below the toolbar is a Queries section with a Perform button. The main area shows a table with 220 objects and 42 attributes that differ. The left sidebar lists attributes such as Key, Notes, Company Short Name, Company Name, Location, Structure of Corporation (with sub-attributes like CEO, CEO (Photo), Upper Management, Director of Staff, Director of Staff (Photo), Division Head), Type of Company Relation, Markets, Lines of Business, and Parent Company. The table data is compressed, showing only the first few characters of each cell. A "Browse/Open column" button is overlaid on the table. The Windows taskbar at the bottom shows the Start button, several open applications (InfoZoom340EN, 3 1/2 Floppy (A:), InfoZoom.ppt, Intro), and the system tray with the time 1:31 AM.

# InfoZoom Wide Table View

InfoZoom - [Films.fox]

File Edit View Attributes Objects Format Report Options Window Help

New Open Save Print Preview Chart Excel Report

WideTable... Overvi... Back Forward All Zoom in Zoom... Exclude Search

The actors with the most films

Queries Perform

1740 of 1740 Objects  
10 Attributes differ

	0	1	2	3	4	5	6	7	8
<b>Film</b>	0	1	2	3	4	5	6	7	8
<b>Title</b>	Wild at Heart	Goodbye Again	Hunt for Red October, The	Terminator, The	Terminator 2	John Cleese on How to Irritate People	Au Revoir les Enfants	The Ballad of Narayama	Cyrano De Bergerac
<b>Year</b>	1990	1961	1990	1984	1991	1993	1987	1983	1999
<b>Length</b>	125	120	135	108	136	65	103	128	134
<b>Subject</b>	Drama	Drama	Drama	Action	Action	Comedy	Drama	Drama	Drama
<b>Actor</b>	Cage, Nicolas	Perkins, Anthony	Connery, Sean	Schwarzenegger A.	Schwarzenegger A.	Cleese, John	Manesse, Gaspard	---	Depardieu Gerard
<b>Actress</b>	Dern, Laura	Bergman, Ingrid	---	Hamilton, Linda	Hamilton, Linda	Booth, Connie	Racette, Francine	Missing	Brochet, Catherine
<b>Director</b>	Lynch, David	Litvak, Anatole	McTiernan, J.	Cameron, J.	Cameron, J.	---	Malle, Louis	Imamura, Shohei	Rappeneau Jean-Paul
<b>Popularity</b>	6	6	8	17	8	62	35	15	86
<b>Awards</b>	No	No	No	No	No	No	No	No	No

Show table in uncompressed mode

# Datasets

- Multidimensional data: three databases were used
  - Anonymized data from a web based dating service (60 records, 27 variables)
  - Technical data of cars sold in 1970 - 82 (406 records, 10 variables)
  - Data on the concentration of heavy metals in Sweden (2298 records, 14 variables)

# Sample Questions

- Do more women than men want their partners to have a higher education?
- What proportion of the men live in California?
- Do all people who think the bar is a good place to meet a mate also believe in love at first sight?
- Do heavier cars have more horsepower?
- Which manufacturer produced the most cars in 1980?
- Is there a relationship between the displacement and acceleration of a vehicle?

# Experiment Design

- The experimenters generated 26 tasks from all three data sets.
- 83 participants. Between-subjects design.
- Each was given one visualization system and all three data sets.
- Type of visualization system was the independent variable between them.
- 30 mins were given to solve the tasks of each data set i.e 26 tasks in 90 mins.

# Overall Results

- Mean task completion times:
  - Infozoom users: 80 secs
  - Spotfire users: 107 secs
  - Eureka users: 110 secs
- Answer correctness:
  - Infozoom users: 68%
  - Spotfire users: 75%
  - Eureka users: 71%
- Not a time-error tradeoff
  - Spotfire more accurate only 6 questions

# Eureka - problems

- Hidden labels: Labels are vertically aligned, max 20 dimensions
- 3+ Attributes: Problems with queries involving three or more attributes
- Correlation problems: Some participants had trouble answering questions correctly that involved correlations between two attributes.

# Spotfire - problems

- Cognitive setup costs: Takes participants considerable time to decide on the right representation and to correctly set the coordinates and parameters.
- Biased by scatterplot default: Though powerful, many problems cannot be solved (well) with it.

# Infozoom - problems

- Erroneous Correlations
  - Overview mode has all attributes sorted independent of each other
  - Narrow row height in compressed view
  - Participants did not use row expansion and scatterplot charting function which shows correlations more accurately

# Geographic Questions

- Spotfire should have done better on these
  - Which part of the country has the most copper
  - Is there a relationship between the concentration of vanadin and that of zinc?
  - Is there a low-level chrome area that is high in vanadim
- Spotfire was only better only for the last question (out of 6 geographic ones)

# Discussion

- Many studies of this kind use relatively simple tasks that mirror the strengths of the system
  - Find the one object with the maximum value for a property
  - Count how many of certain attributes there are
- This study looked at more complex, realistic, and varied questions.

# Discussion

- Success of a visualization system depends on many factors:
  - Properties supplied
    - Spotfire doesn't visualize as many dimensions simultaneously
  - Operations
    - Zooming easy in InfoZoom; allows for drill-down as well
    - Zooming in Eureka causes context to be lost
    - Column view in Eureka makes labels hard to see

# Information Exploration “Shootout”

- <http://ivpr.cs.uml.edu/shootout/about.html>
- Data Mining Applications
- One component focuses on visualization

# Comparing Tree Views

- T. Barlow and P. Neville, Comparison of 2D Visualizations of Hierarchies, INFOVIS'01.
- Problem
  - Organization Chart is de facto standard for visualizing decision trees. Is there a better compact view of the tree for the overview window?
- Solution
  - Two usability studies to determine which tree works best.

# Goal: Compact View of Tools

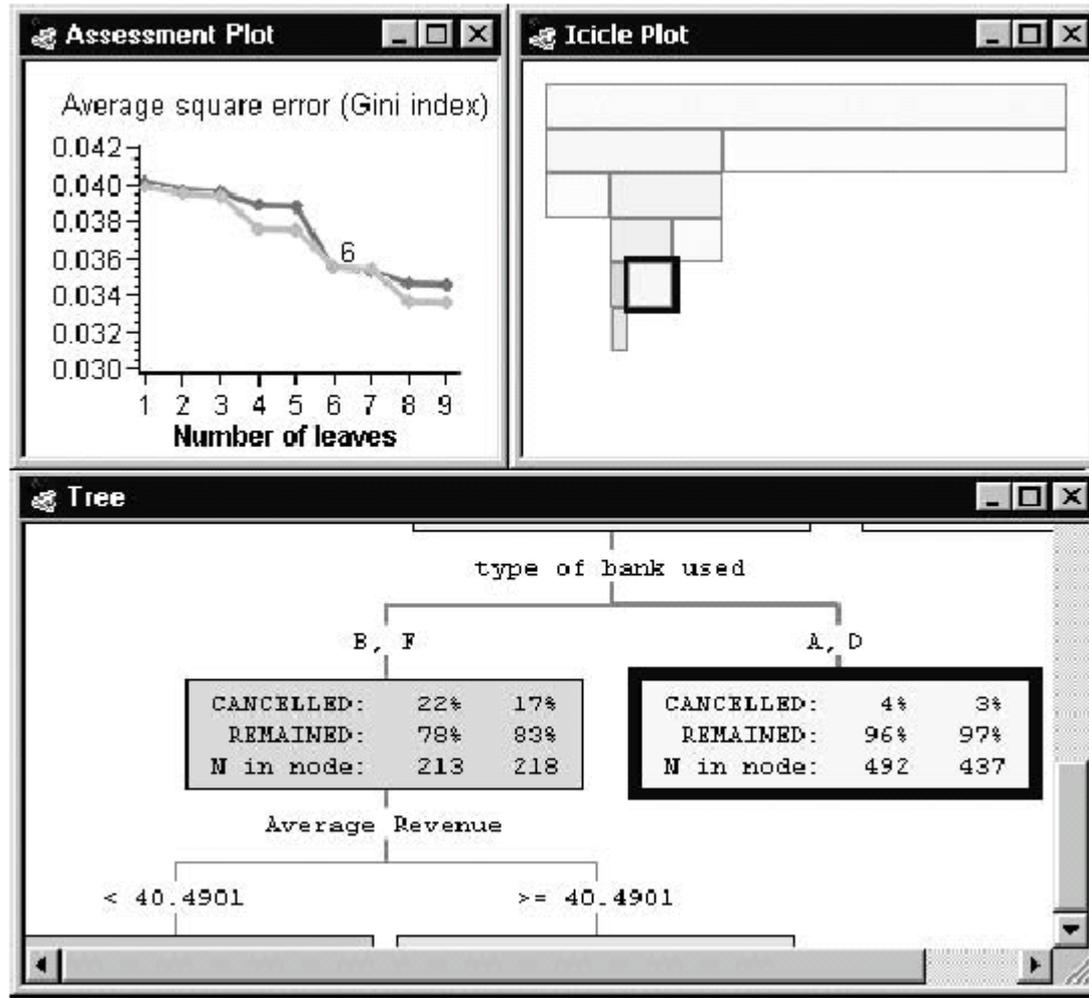
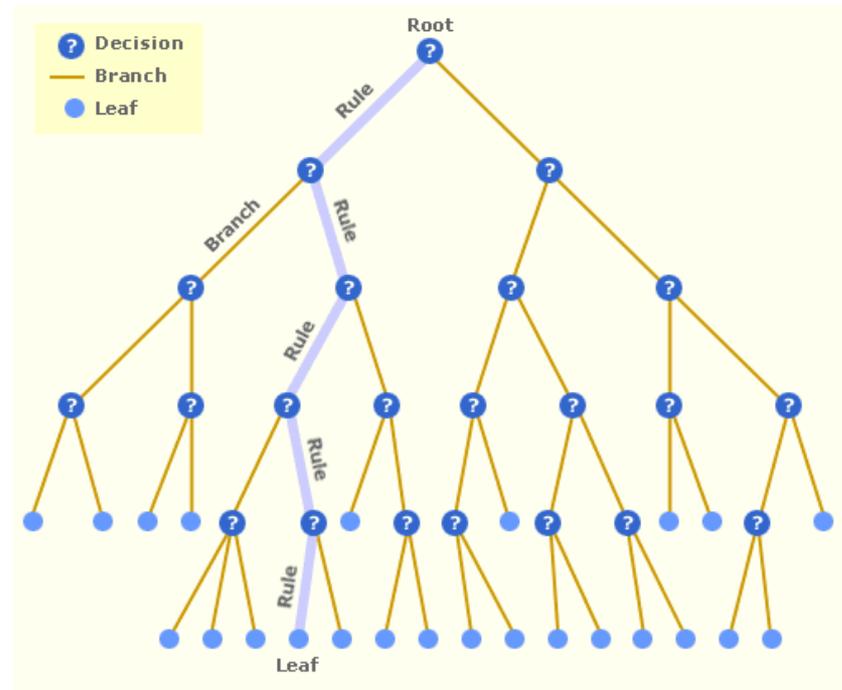


Figure 1. Example of compact view in data mining

# Decision Trees

- Each split constitutes a rule or variable in predictive model
- Begin Splitting into nodes
- Often hundreds of leaves



# Decision Trees – What makes a good visualization

- Uses
  - For novice-helps them understand models
  - Experts-initial evaluation of decisions tree without looking at models
- Criteria for usability in study
  - Ease of Interpretation of Topology (Parent Child Sibling relations)
  - Comparison of Node Size
  - User preference

# Different views examined in study

Org Chart Tree Ring Icicle Plot

TreeMap

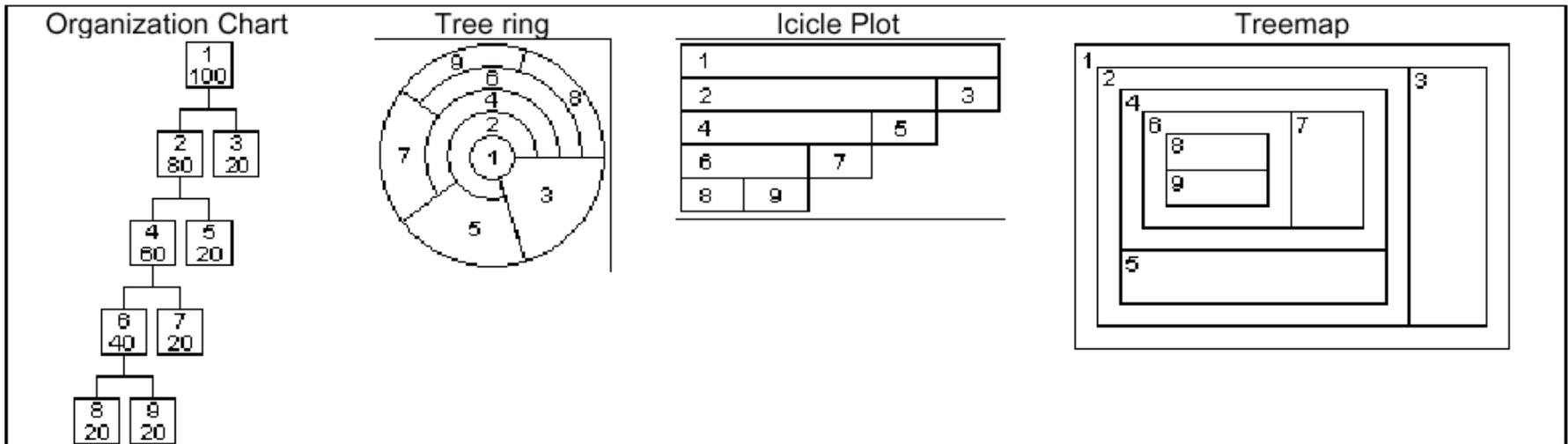


Figure 2. Different views of the same tree



# Usability Test 1:

- Users:
  - 15 colleagues familiar with org chart but not others
- Tasks
  - Is the tree binary or n-ary?
  - Is the tree balanced or unbalanced?
  - Find deepest common ancestor of two nodes
  - Number of levels?
  - Find three largest leaves (excluding org chart)
- Data: Created 8 trees for analysis
- Study Design
  - Randomized order of tasks
  - 4X5 design (almost)
  - Timed task from appearance on screen until spacebar tap

# Results

- Response Time
  - TreeMap slowest; no statistical difference between others
- Response Accuracy
  - No significant difference
- User Preference
  - Prefer icicle map and org chart (faster)
  - Dislike tree map

# Discussion

- Org chart served as benchmark
- Icicle plot favored amongst others
  - Hypothesis: Same left to right / top to bottom structure
- TreeRing did well
- TreeMap suffered from poor accuracy
  - Offset of rectangles required because of off (which is needed for selection)

# Usability Test II: Tree implementation

- Three views:
  - TreeMap eliminated from this round
- Tasks
  - Node Description
    - Four versions – select those nodes or leaves that meet certain criteria
  - Node Analysis:
    - Memorize a highlighted node – find again after tree redrawn in different position

# Results

- Tree rings slower for description but fast and accurate for memory tasks
- Perhaps due to unique geometric forms / spatial clues

# Conclusions

- TreeMap not useful for this type of task
- Org Chart/Icicle seem to be best overall
- TreeRing has merits for certain tasks
  
- Icicle chosen for implementation
  - Best design considering Org Chart could not be used for node size tasks
- However:
  - Didn't seem to actually do tests on trees as large as the ones they describe as typical of datamining

# Visualizing Conversations

# Text-Based Chat

Katesmiles1 enters  
You tell Horse\_99 me too  
Horse\_99 says Real.  
Speci\_Man\_98 says Where you from Kim?  
Soapbox\_7 leaves, heading for the Gen-X Love #19  
Horse\_99 says On here!  
Horse\_99 says Lets go private and find out.  
Muta4 leaves heading for another room  
Muta4 leaves  
Horse\_99 says Sure.  
Speci\_Man\_98 says Ever been to new York?

Katesmiles1 enters  
You tell Horse\_99 me too  
Horse\_99 says Real.  
Speci\_Man\_98 says Where you from Kim?  
Soapbox\_7 leaves, heading for the Gen-X Love #19  
Soapbox\_7 says  
GODESS\_OFLOVE leaves, heading for the Gen-X Love #19  
Kim\_24\_98 says connecticut  
Horse\_99 How old are you Mika?  
Sycam leaves  
Kim\_24\_98 says where are you from  
Muta4 says 16  
You tell Horse\_99 are you talking to me when you say Mika?  
Speci\_Man\_98 says Paducah, ky  
Horse\_99 says 24/m/a  
Horse\_99 says Yes  
You say 26  
Kim\_24\_98 says cool, I'm 25f  
Horse\_99 says On here!  
Horse\_99 says Lets go private and find out.  
Muta4 leaves heading for another room  
Muta4 leaves  
Horse\_99 says Sure.  
Speci\_Man\_98 says Ever been to new York?





# Chat Circles



- “Chat Circles is a graphical interface for synchronous communication that uses *abstract shapes* to convey *identity* and *activity*.”
- Each participant appears as a colored circle, which is accompanied by the user name
- Location of circles will also identify participants (important for many users having similar colors associated)
- Participants’ circles become larger when posting occurs (circle adapts to text length)
- Circle appears bright when posting occurs
- Circles of inactive users fade in the background

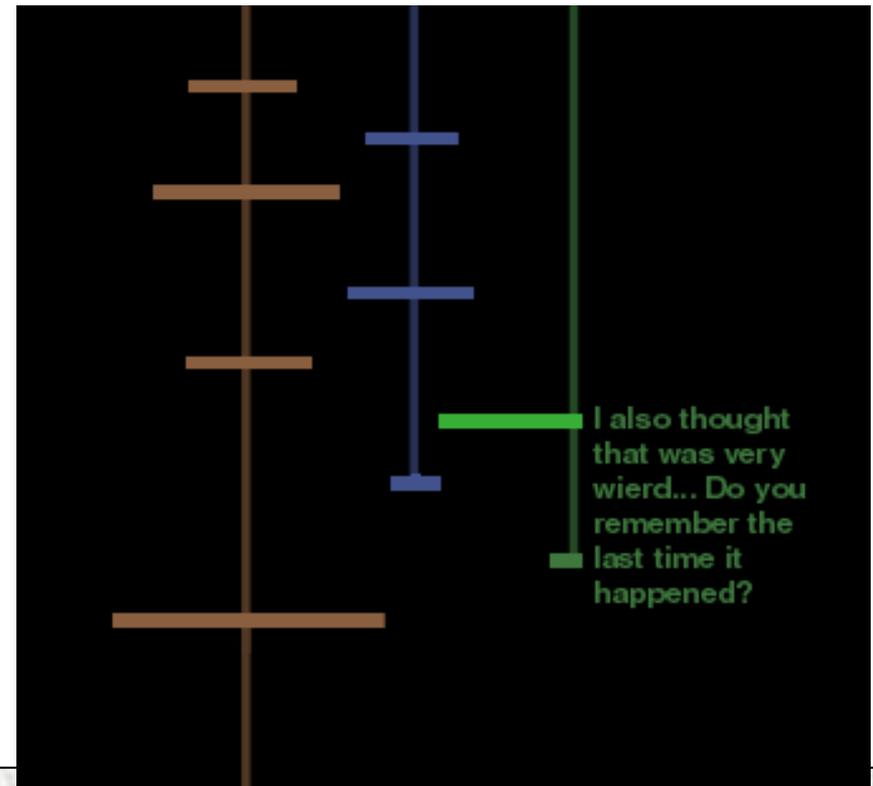
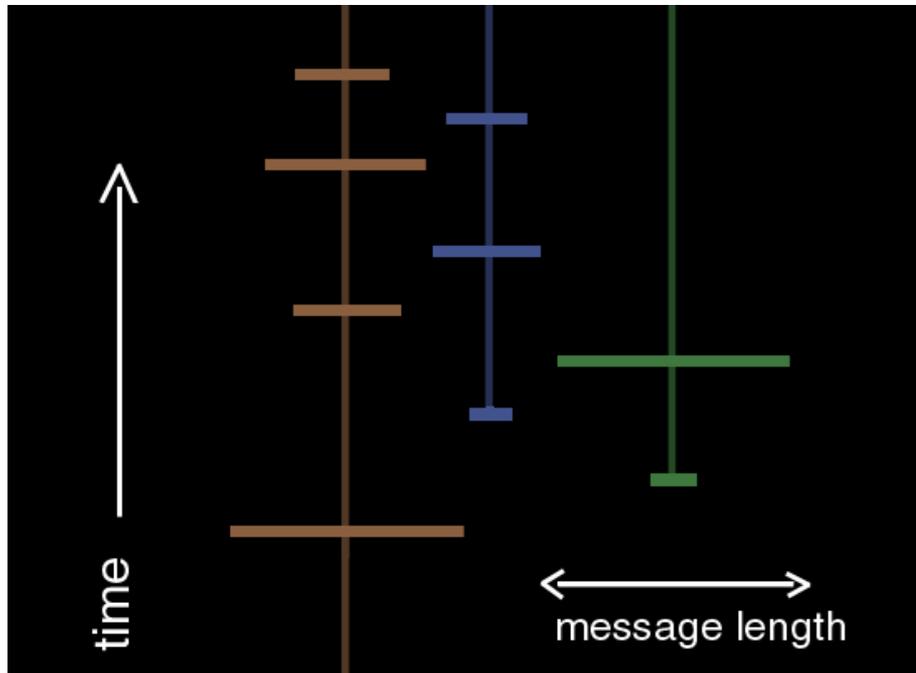
# Chat Circles – Conversational Groupings



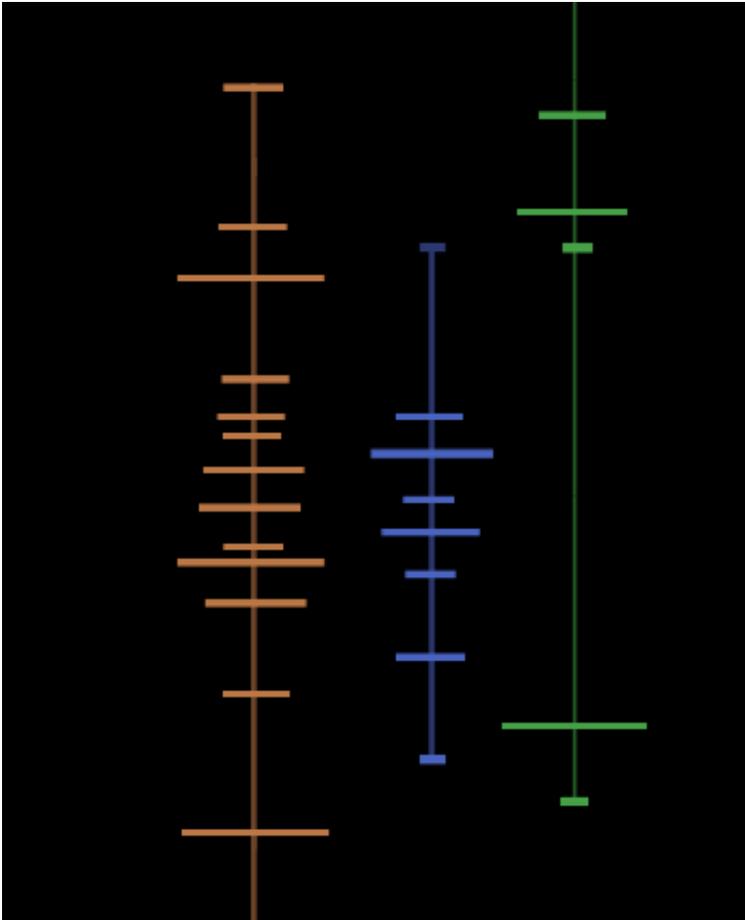
- There is only ONE room in Chat Circles
- Groupings are achieved by moving closer to other participants
- At any time, a participant can view all other participants
- A participant can also detect interesting conversations in different areas of the room by looking at how many circles are gathered and how often circles become larger
- Overview panel in Chat Circles II nice example of focus + context



# Chat Circles History

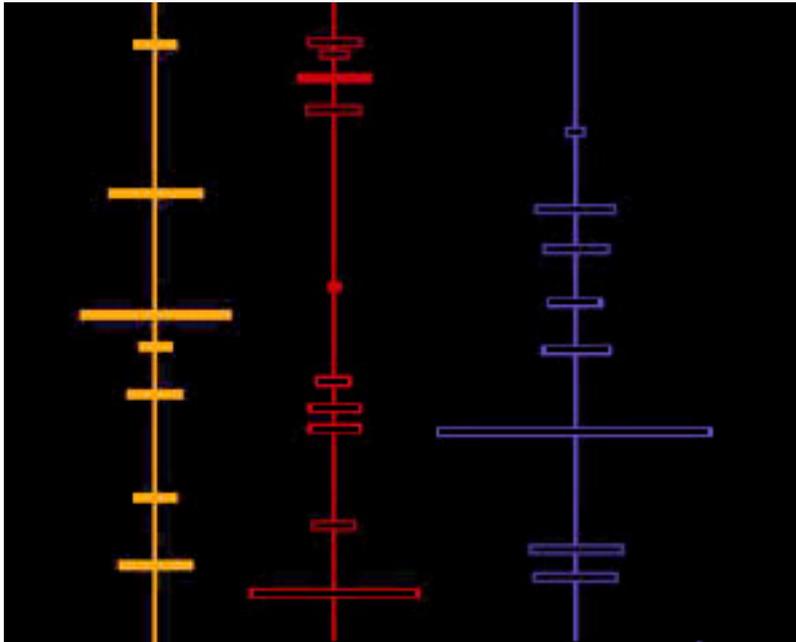


# History Log Patterns



- + Easy to see “lurkers”
- + Sequence and size of messages quickly visible
- Not very scalable

# History Log Patterns



- +/- User-centric: only 1 point of view represented
- Impossible to see all the text at once – requires individual mouse rollovers
- Easy to see “out of range” conversations – *but why would you want to?*

# Agenda

- Introduction
- Visual Principles
- What Works?
- Visualization in Analysis & Problem Solving
- Visualizing Documents & Search
- Comparing Visualization Techniques
- Design Exercise
- Wrap-Up

# Design Exercise

# Design Exercise

- **BreakingStory**

(Reffel, Fitzpatrick, Ayedelott SIMS final project, at CHI 2003)

- Create an application that supplies a visualization for trends over time in web-based news. The primary purpose is to provide an overview, but it should also be possible to view text from individual news sources on specific days. Its goal is to inform, inspire, and enlighten, and also to make people want to look more deeply at the news.



## What is BreakingStory?

BreakingStory is a tool to help you explore online news. Curious about when a phrase was first mentioned? Wondering if references to a current event vary by geographic region? Here you can find out!

## How do you use it?

You can use this site to search for words or phrases that have appeared in online news. The results will be displayed in one or more charts that show the history of references.

**Single Chart View** allows you to examine references from one geographic area or news site.

**Multiple Chart View** shows a number of small charts next to each other, allowing you to compare references from different geographic areas or news sites.

You can also view the **full text** of the front pages that matched your search.

Don't know where to begin? Try looking at an example: [World Aids Day](#)

## Want More?

- [Help](#) - suggestions, advice for using BreakingStory
- [Site List](#) - the news sites we collect and their locations
- [FAQ](#) - the who what when where how of BreakingStory
- [Project Page](#) - detailed information on the project
- [E-mail us](#) - we want to hear your comments, really we do

### Step 1: Enter Search Term(s) [\[Need help?\]](#)

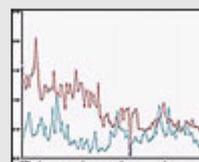
First word or phrase:

Add another term for comparison

### Step 2: Select Date Range [\[Need help?\]](#)

to

### Step 3: Choose One of Two Views [\[Need help?\]](#)



Single Chart View

Choose a news site or collection of news sites, by selecting **one** of the following:

Individual News Site:

All News Sites in Country:

All News Sites in Region:



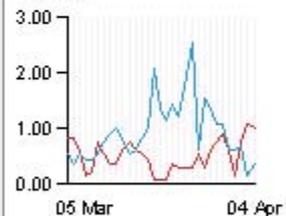
## Front page hits for:

oil peace

## Selected date range:

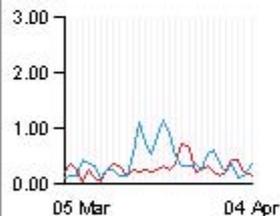
05 March 2003 to 04 April 2003 (31 days total)

[World >](#)  
Africa



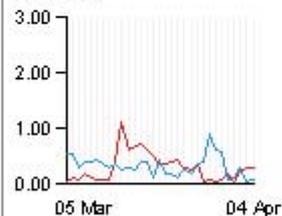
[\(view large chart\)](#)

[World >](#)  
Asia



[\(view large chart\)](#)

[World >](#)  
Europe

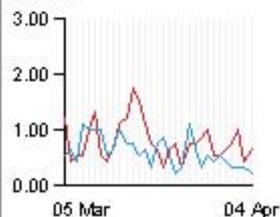


[\(view large chart\)](#)

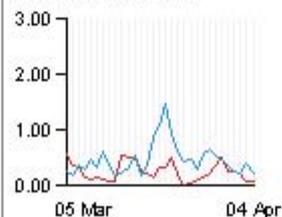
[World >](#)  
Latin America and the Caribbean



[World >](#)  
Middle East / North Africa



[World >](#)  
North America



## Change Search Term(s)

[\[Need help?\]](#)

First word or phrase:

oil

Second word or phrase:

peace

Third word or phrase:



Add another term for comparison

## Change Date Range

[\[Need help?\]](#)

05 Mar 03

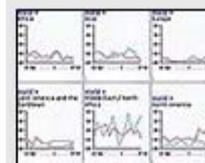
to

04 Apr 03

## Change View

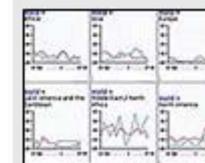
[\[Need help?\]](#)

### Redisplay Multiple Charts



Redisplay Charts

### Select New Multiple Charts



Select New Charts

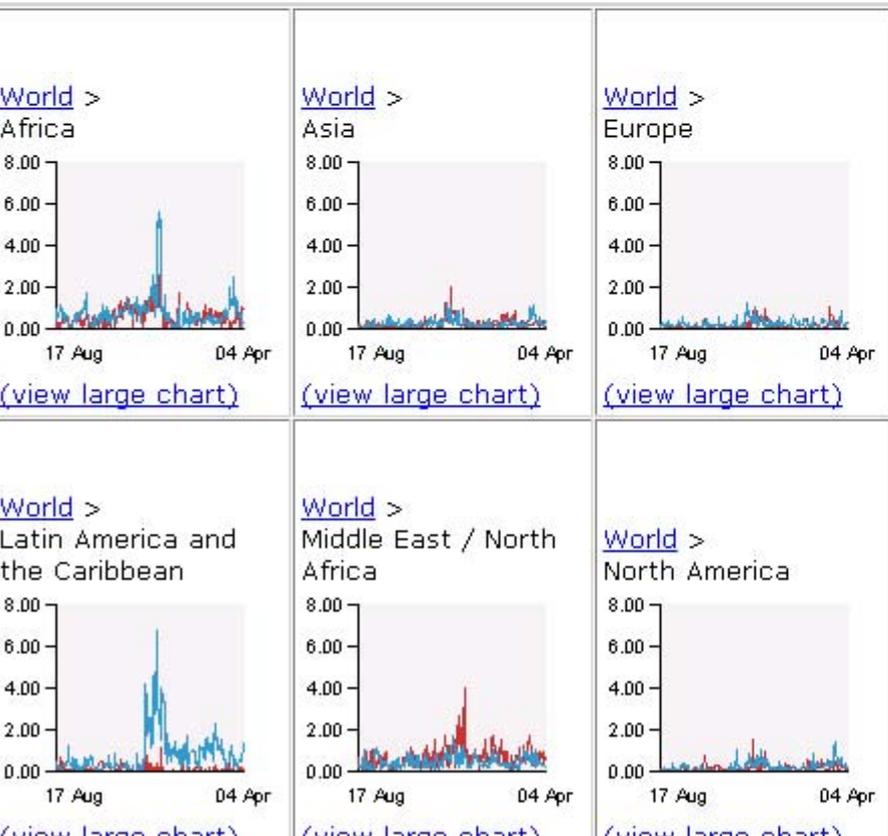


## Front page hits for:

oil peace

## Selected date range:

17 August 2002 to 04 April 2003 (231 days total)



## Change Search Term(s)

[\[Need help?\]](#)

First word or phrase:

Second word or phrase:

Third word or phrase:

Fourth word or phrase:

Add another term for comparison

## Change Date Range

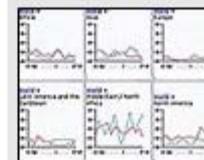
[\[Need help?\]](#)

to

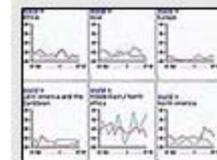
## Change View

[\[Need help?\]](#)

### Redisplay Multiple Charts



### Select New Multiple Charts



## Front page hits:

Average references per page (total references)

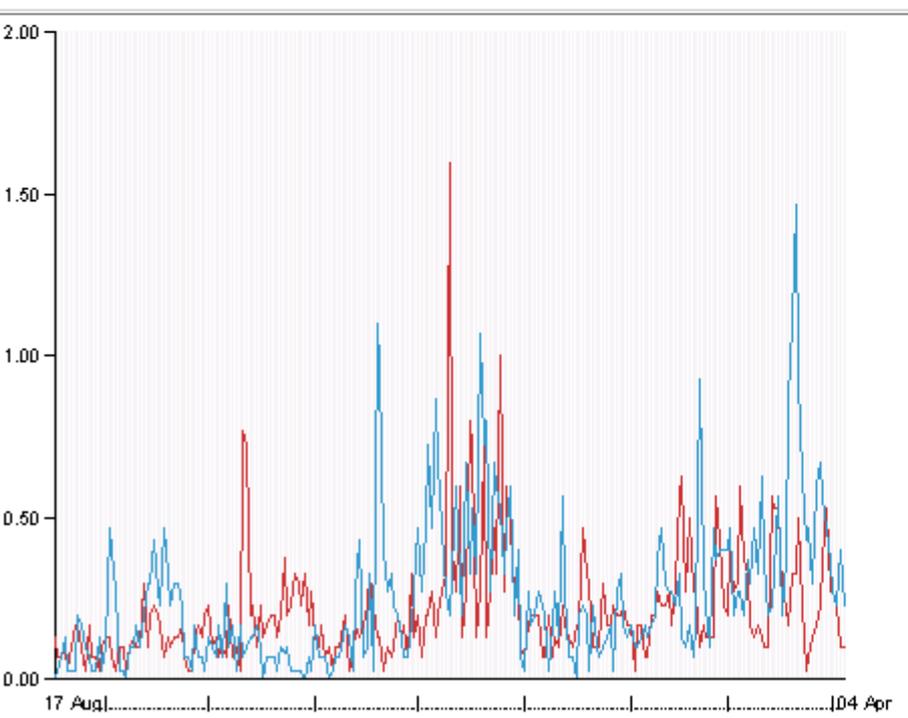
**oil** : 0.26 (1794 total references in 1305 pages)

**peace**: 0.22 (1494 total references in 1163 pages)

## Selected date range:

17 August 2002 to 04 April 2003 (231 days total)

[World](#) > North America



Follow a link below to view a chart for an individual country. From there,

## Change Search Term(s)

[\[Need help?\]](#)

First word or phrase:

Second word or phrase:

Third word or phrase:

Fourth word or phrase:



Add another term for comparison

## Change Date Range

[\[Need help?\]](#)

to

## Change News Sites(s)

[\[Need help?\]](#)

To choose a different news site or collection of news sites, select **one** of the following:

Individual News Site:

All News Sites in Country:

All News Sites in Region:

## Change View

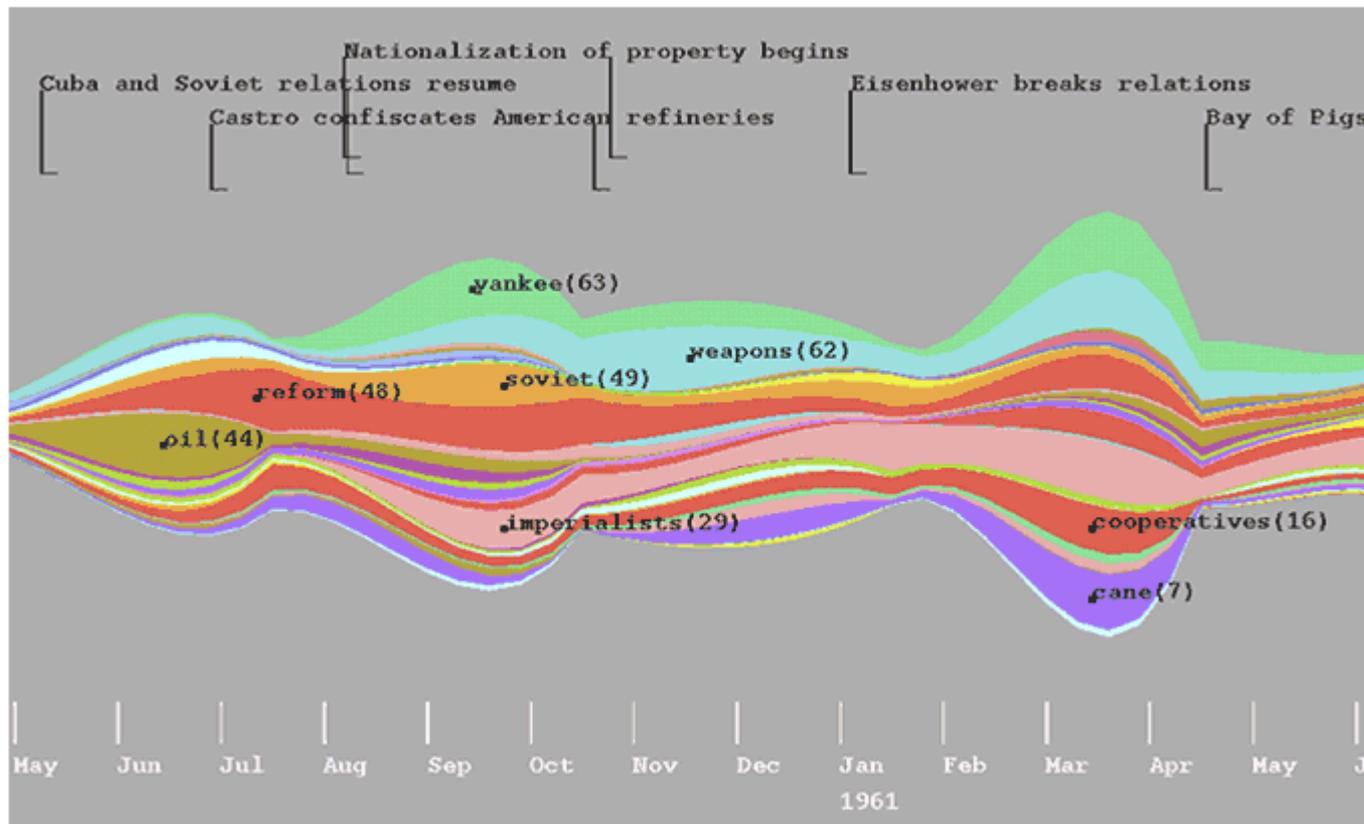
[\[Need help?\]](#)

**Redisplay  
Single Chart**

**Select  
Multiple Charts**

# Another Approach: ThemeRiver

- S. Havre, B. Hetzler, L. Nowell, "ThemeRiver: Visualizing Theme Changes over Time," Proc. IEEE Symposium on Information Visualization, 2000



# Wrap-up: Guidelines for Success

# Key Questions to Ask about a Viz

1. Is it for analysis or presentation?
2. What does it teach/show/elucidate?
3. What is the key contribution?
4. What are some compelling, useful examples?
5. Could it have been done more simply?
6. Have there been usability studies done?  
What do they show?

# Holistic Design Goals for Information Visualization

- Tailor to the application and the domain
- Create highly interactive and integrated systems
- Embed the visualization within a larger application
- Provide alternative views

# Visualization with a Light Touch: Orbitz.com

Address <http://www.orbitz.com/>

**ORBITZ** VISIT PLANET EARTH

home flights hotels cars

the **Orbitz** quick search

flights hotels cars

round trip [one-way](#) [multi-city](#)

\*city name (e.g. Boston), or [airport code](#)

from

to

I need to travel on **specific dates**

My dates are **flexible** (US & Canada)

**NEW**

leave Apr 11

return Apr 18

select up to 4 travelers

1 adult (12-64) 0 child (2-11)

0 senior (65+)

[12-17 traveling alone](#)

**Need a hotel?**  
save up to 75% on your hotel reservation

[full search options](#)

Orbitz - Microsoft Int...

close window

April 2003

Su	Mo	Tu	We	Th	Fr	Sa
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	1	2	3

May 2003

Su	Mo	Tu	We	Th	Fr	Sa
27	28	29	30	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

from\*

- \$209
- \$83
- \$183
- \$277
- \$395
- \$234
- \$239
- \$79

see all flight deals

other travel deals

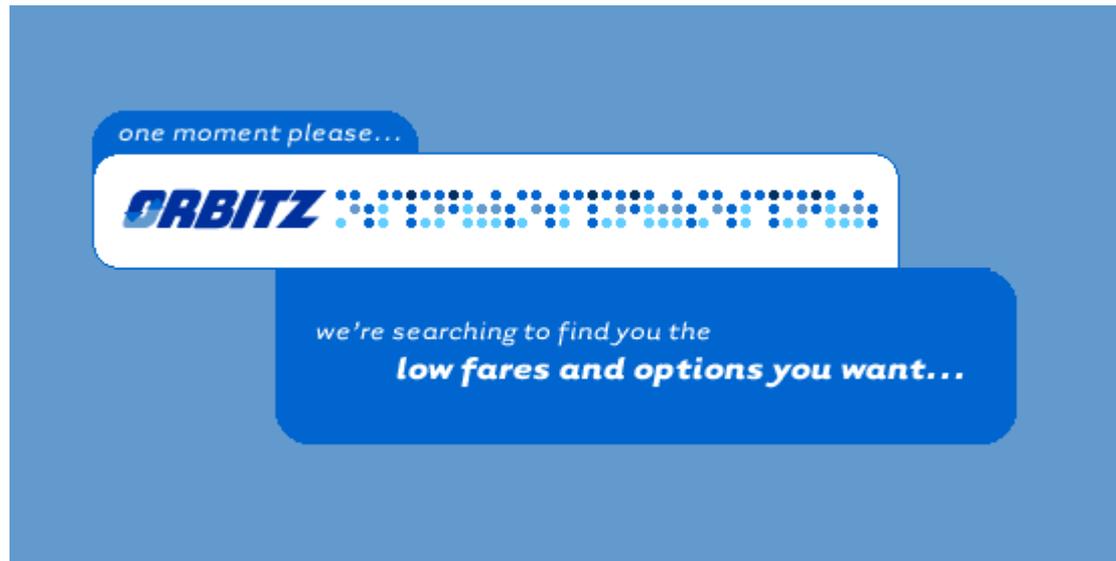
- Orbitz Rewards - FREE airport parking! \$G0\$
- Air/Hotel/Car Deals with NEAT E-Packs \$SAVE\$
- Last Minute Cruise Deals \$249
- Homeland Cruises-Sales Close to Home \$SAVE\$
- Vacation Picks Just For You! \$SAVE\$
- Popular sunny destinations on sale! \$SAVE\$

\* Some taxes, fees additional. [Learn more.](#)

about Orbitz | Orbitz for Hotels | Orbitz for Business | site map | contact us |



# Visualization with a Light Touch: Orbitz.com



# Visualization with a Light Touch: Orbitz.com

The screenshot shows the Orbitz.com website interface. At the top, there are navigation tabs for 'home', 'flights', 'hotels', 'cars', 'cruises', 'packages', 'deals', and 'customer care'. Below these are 'TRAVEL WATCH', 'MY STUFF', and 'REWARDS'. The main header features the 'ORBITZ MATRIX DISPLAY' logo and 'SIGN OUT' and 'HELP' links. A search bar contains the criteria: 'Fri, Apr 11 anytime Oakland (All Locations), CA (OAK) New York (All Locations), NY (NYC)'. A 'change search' button is visible. Below the search bar, there are icons for various airlines: American Airlines, Multiple Carriers, United Airlines, Continental Airlines, America West, Delta Air Lines, Frontier Airlines, US Airways, and Northwest Airlines. A table shows the lowest fares for 0, 1, and 2+ stops. A tip box indicates that the lowest fare found beats the 30-day average by \$81. At the bottom, there is a 'list flights by' section with radio buttons for 'lowest price', 'departure times', and 'shortest flights'. Below this is a detailed flight list table.

home flights hotels cars cruises packages deals customer care TRAVEL WATCH MY STUFF REWARDS

ORBITZ MATRIX DISPLAY SIGN OUT HELP

help with this page

[AD-MUNCHER]

change search

Fri, Apr 11 anytime Oakland (All Locations), CA (OAK)  
New York (All Locations), NY (NYC)

Fri, Apr 18 anytime New York (All Locations), NY (NYC)  
Oakland (All Locations), CA (OAK)

SELECT STOPS

SELECT STOPS	American Airlines	Multiple Carriers	United Airlines	Continental Airlines	America West	Delta Air Lines	Frontier Airlines	US Airways	Northwest Airlines
0									
1	\$229+ see below	\$332+	\$345+	\$347+	\$482+	\$733+	\$745+	\$892+	\$1334+
2+		\$638+	\$471+	\$354+				\$1564+	

prices above are per person and may not be purchased on Orbitz without applicable [service fees](#)

**tip** \$ We checked for better fares at nearby airports, and your search found our lowest prices. Lowest fare below beats this trip's 30-day average by \$81

list flights by:  lowest price  departure times  shortest flights

PRICE (USD)	AIRLINE	TIMES	FROM (airport codes)	TO (airport codes)	STOPS
<b>SELECT</b> Web Fare \$229 airfare \$5 service fee \$234 trip cost	<b>American Airlines 862</b>	<b>8:20a-1:44p</b>	Oakland (OAK)	Dallas/Fort Worth (DFW)	1
	Fri, Apr 11	plane change			
	<b>American Airlines 590</b>	<b>5:28p-10:03p</b>	Dallas/Fort Worth (DFW)	New York (JFK)	
	Fri, Apr 11			total duration: 10h 43min	
<b>SELECT</b> Web Fare \$229 airfare \$5 service fee	<b>American Airlines 1309</b>	<b>5:53p-9:04p</b>	New York (JFK)	Dallas/Fort Worth (DFW)	1
	Fri, Apr 18	plane change			
	<b>American Airlines 1463</b>	<b>10:01p-11:45p</b>	Dallas/Fort Worth (DFW)	Oakland (OAK)	
	Fri, Apr 18			total duration: 8h 52min	
<b>SELECT</b> Web Fare \$229 airfare \$5 service fee	<b>American Airlines 554</b>	<b>10:00a-3:24p</b>	Oakland (OAK)	Dallas/Fort Worth (DFW)	1
	Fri, Apr 11	plane change			
	<b>American Airlines 590</b>	<b>5:28p-10:03p</b>	Dallas/Fort Worth (DFW)	New York (JFK)	
	Fri, Apr 11			total duration: 9h 3min	

Internet

269

# Visualization with a Light Touch: Orbitz.com

ORBITZ **MATRIX**™ DISPLAY

search location: **Fort Lauderdale (Fort Lauderdale Beach), Florida, United States**

showing all hotels

0-2 miles		from \$63	from \$59		
2-5 miles	from \$55	from \$47	from \$69	from \$182	
5-15 miles	from \$43	from \$45	from \$59	from \$99	from \$270

look below to find: **All hotels at all distances from Fort Lauderdale (Fort Lauderdale Beach)**

list hotels by:  price  distance

**from 119** **Holiday Inn FT L**

999 Fort Lauderdale  
**0.4 Miles East of Ft**

Each of our 240 guestrooms has safe and refrigerator, irons, data p TV and in room movies and voice [hotel and room details, photos and m](#)

avg. nightly rate

<b>Orbitz Saver</b>	<b>118.80 USD</b>
<b>Orbitz Saver</b>	<b>118.80 USD</b>
<b>Orbitz Saver</b>	<b>118.80 USD</b>

[more rates and rooms](#)

**from 209** **Marriott Fort Lauderdale Marina**

1881 Southeast 17th Street, Fort Lauderdale, FL 33316  
**2.7 Miles South of Fort Lauderdale (Fort Lauderdale Beach)** ★★★

Located two miles south of downtown, the Fort Lauderdale Marriott Marina is nestled in an upscale marina on Florida's beautiful Intracoastal Waterway; a series of residential canals leading to the ocean. The Broward County Convention Center and Port...  
[hotel and room details, photos and maps](#)

avg. nightly rate	room description
<b>209.00 USD</b>	Spectacular view of our tropically landscaped...
<b>219.00 USD</b>	Watch luxury yachts cruising the intracoastal...

[more rates and rooms](#)

**from 134** **Doubletree Hotel Ft. Lauderdale-Oceanfront**

440 Seabreeze Boulevard, Ft Lauderdale, FL 33316  
**1.4 Miles South of Fort Lauderdale (Fort Lauderdale Beach)** ★★★

Ideally located on beautiful Ft. Lauderdale Beach, bordered by the sparkling Atlantic and the Intracoastal Waterway, the Doubletree Ft. Lauderdale Oceanfront is your perfect

# For more information

- My course:
- <http://www.sims.berkeley.edu/courses/is247/s02/Lectures.html>
- Atlas of Cyberspaces:
- <http://www.geog.ucl.ac.uk/casa/martin/atlas/atlas.html>
- Gallery of Data Visualization; The Best and Worst of Statistical Graphics
- <http://www.math.yorku.ca/SCS/Gallery/>
- Tamara Munzner's collection:
- <http://graphics.stanford.edu/courses/cs348c-96-fall/resources.html>

Thank you!