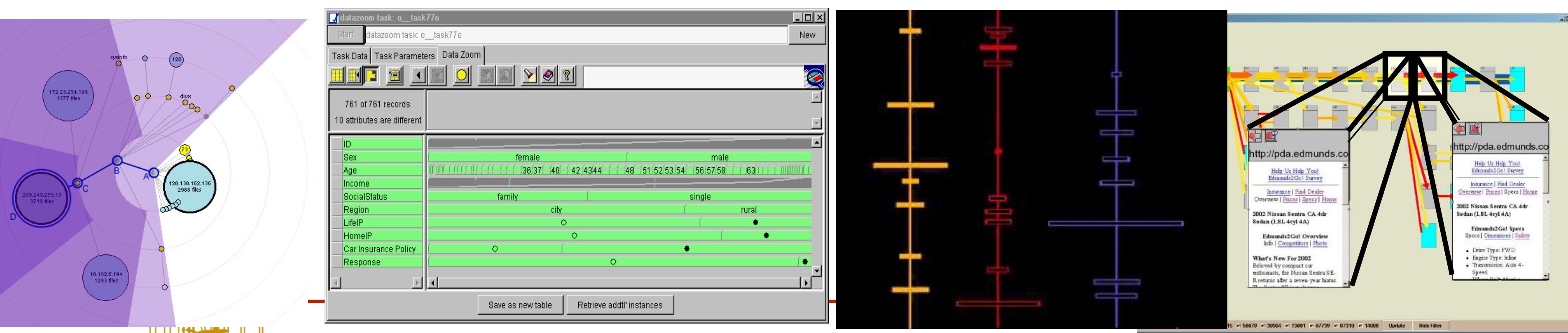
Quality of Visualization

Origins

Based on the CHI 2003 Tutorial by Marti Hearst and a Tutorial by Tamara Munzner



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 skill by making use of the visual system.

Information Visualization

Problem:

- HUGE Datasets: How to understand them?

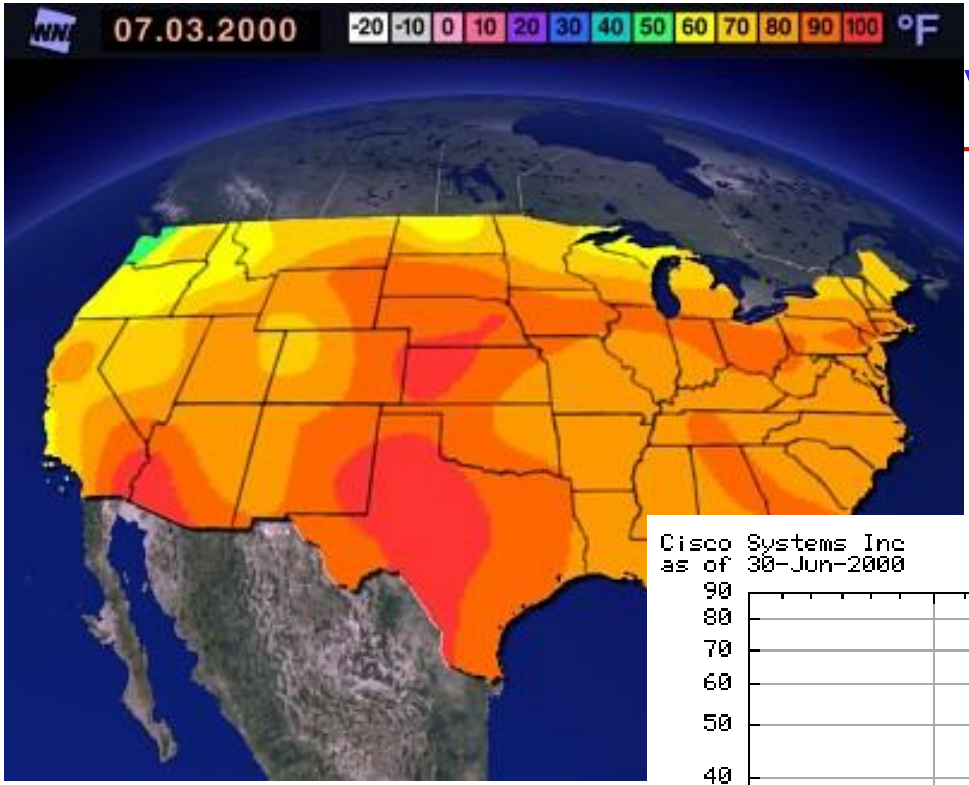
Solution

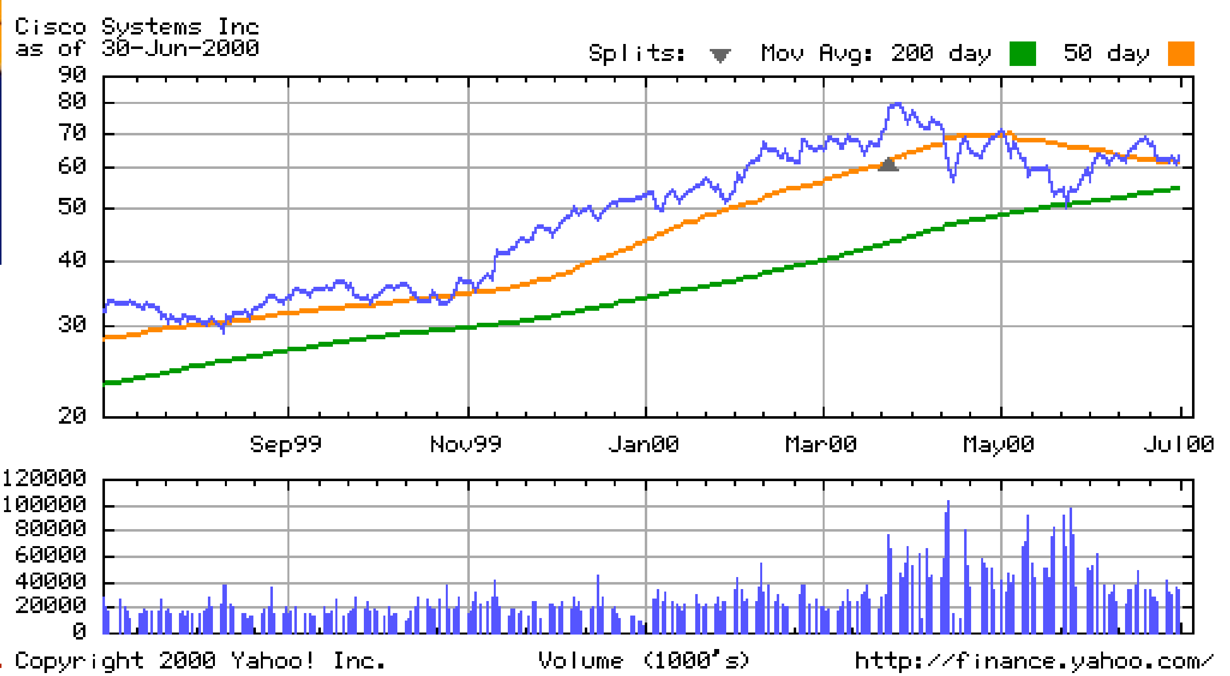
- Take better advantage of human perceptual system
- Convert information into a graphical representation.

Issues

- How to convert abstract information into graphical form?
- Do visualizations do a better job than other methods?

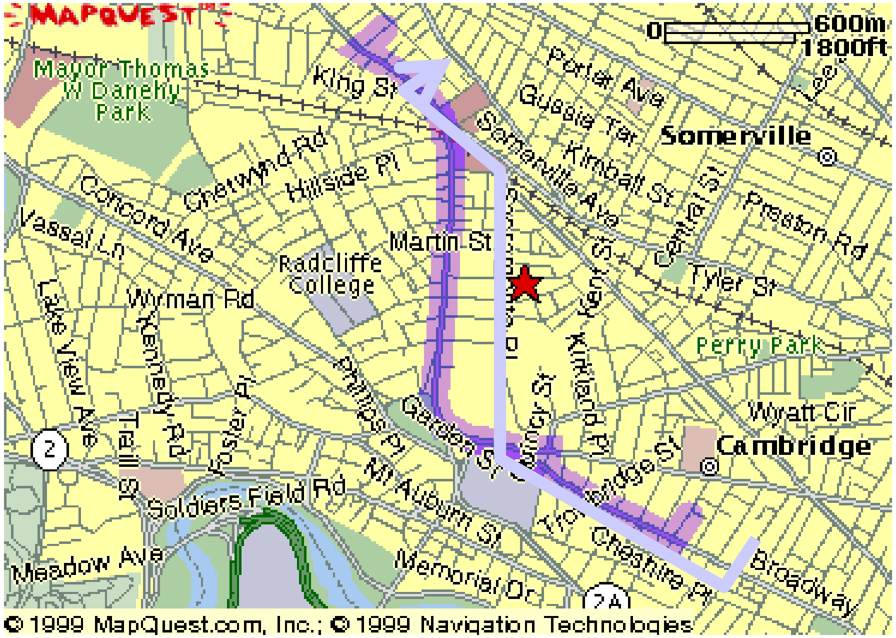
Visualization Success Stories





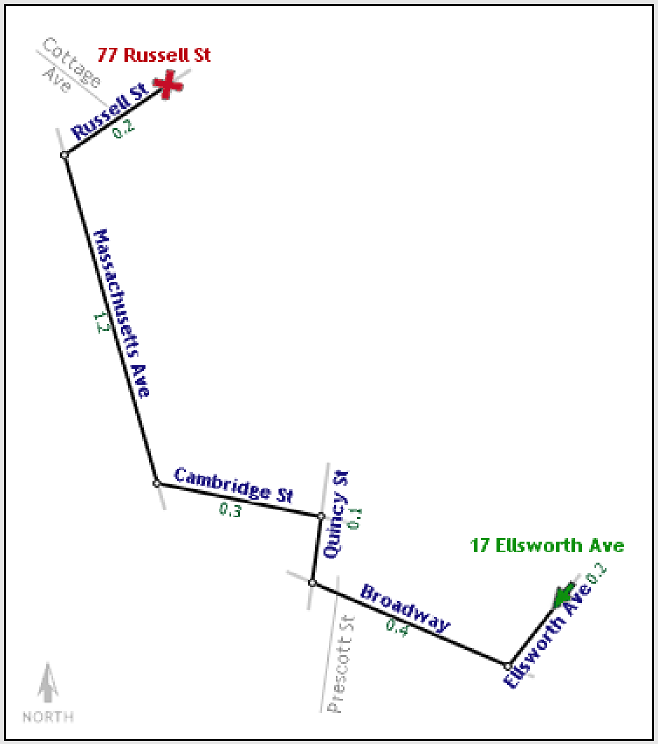
The Power of Visualization

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.



The Power of Visualization

The estimated travel time is 5 minutes for 2.16 miles of travel, total of 6 steps.



	Directions	Elapsed Distance
1	Begin at **17 Ellsworth Ave** on **Ellsworth Ave** and go Southwest for 500 feet	0.1
2	Turn right on **Broadway** and go Northwest for 0.4 miles	0.5
3	Turn right on **Quincy St** and go North for 200 feet	0.5
4	Turn left on **Cambridge St** and go West for 0.3 miles	0.8
5	Bear right on **Massachusetts Ave,Mass Ave,RT-2A** and go North for 1.2 miles	2.0
6	Turn right on **Russell St** and go Northeast for 1000 feet to **77 Russell St**	2.2

Visualization Success Story

Visualization Success Story

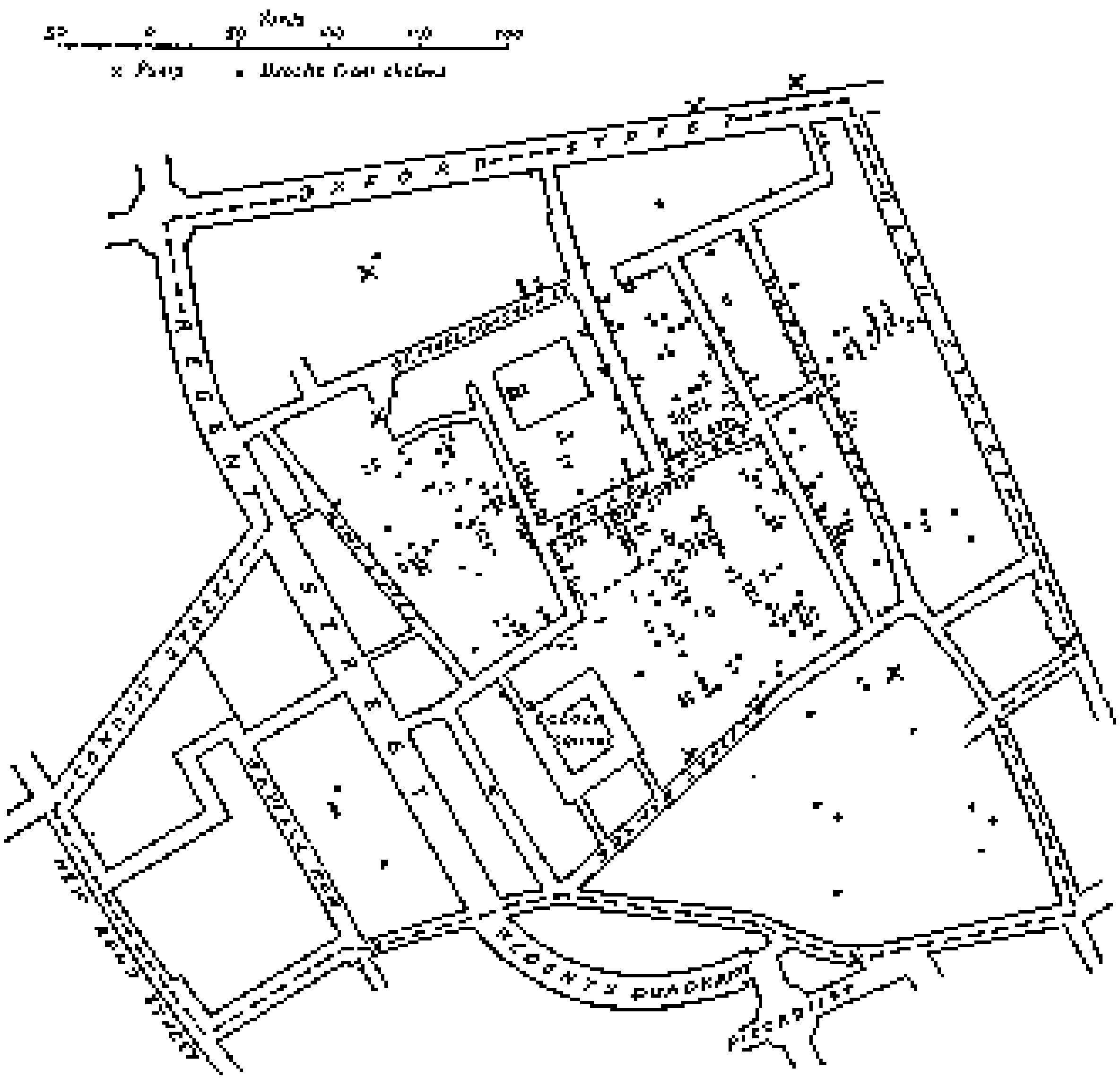


Illustration of John Snow's
deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

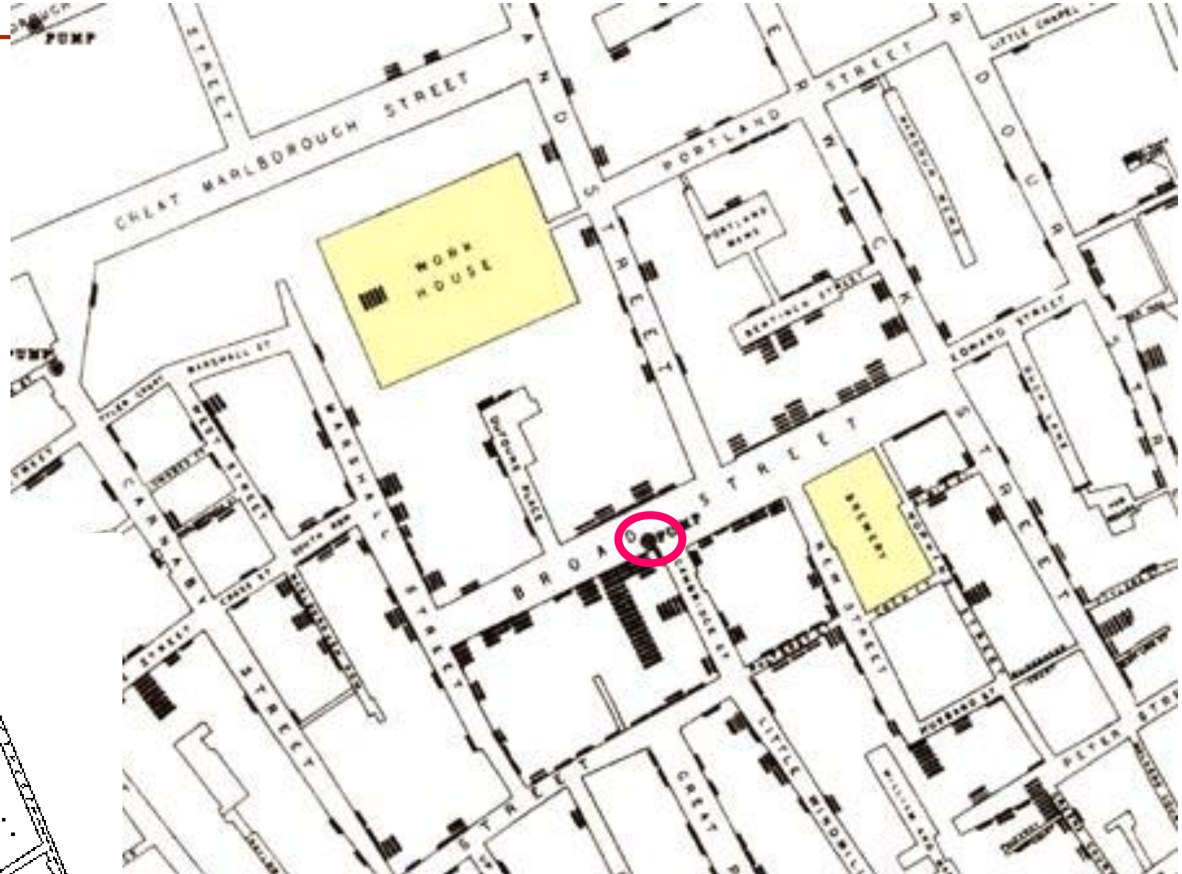
Horizontal lines indicate location of deaths.

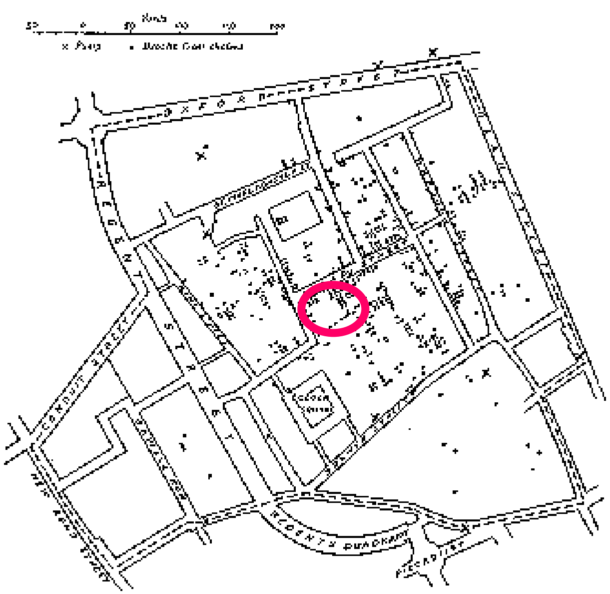
From Visual Explanations by Edward Tufte, Graphics Press, 1997

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 Vis

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

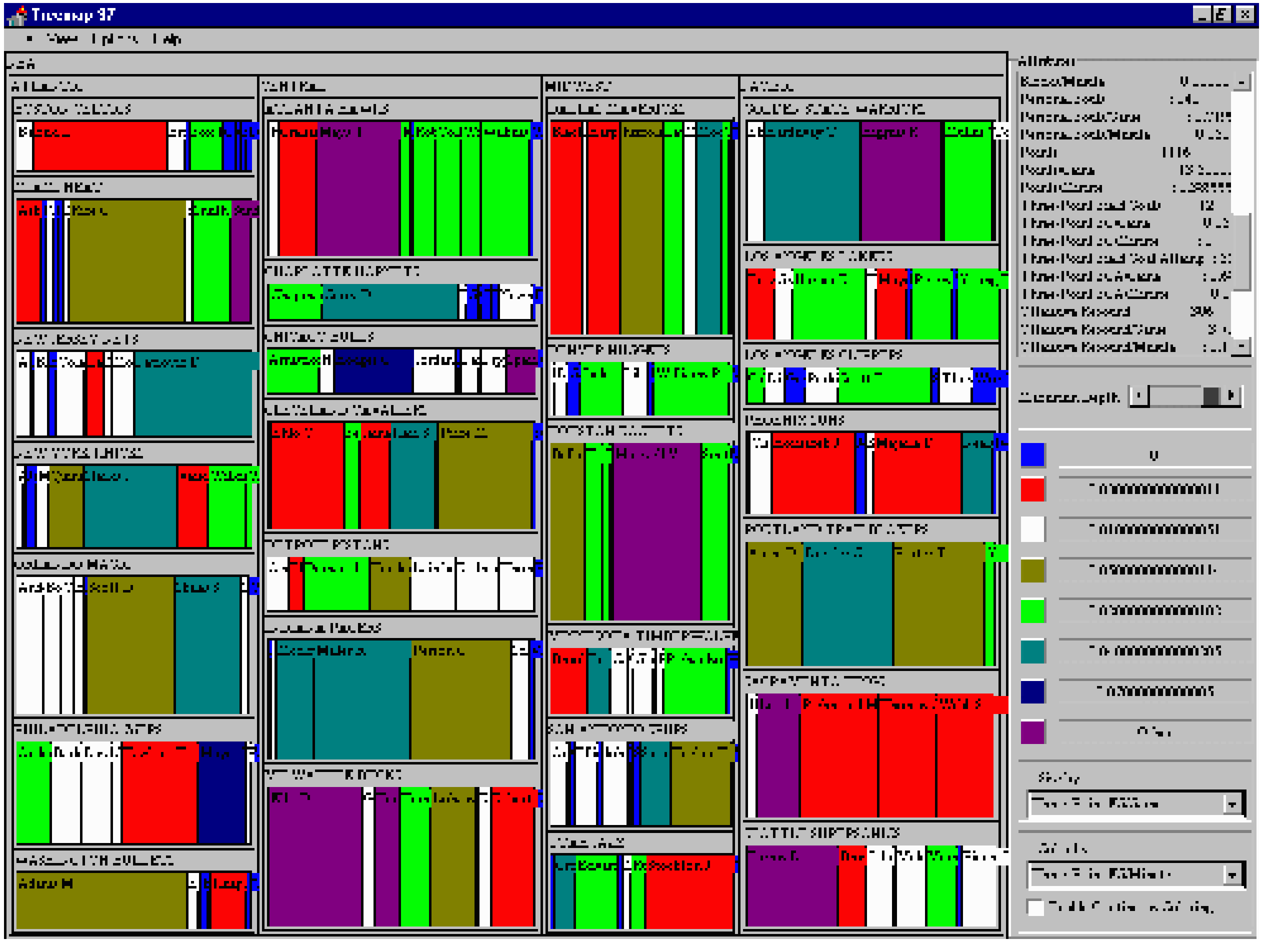
Case Study:
The Journey of the TreeMap

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

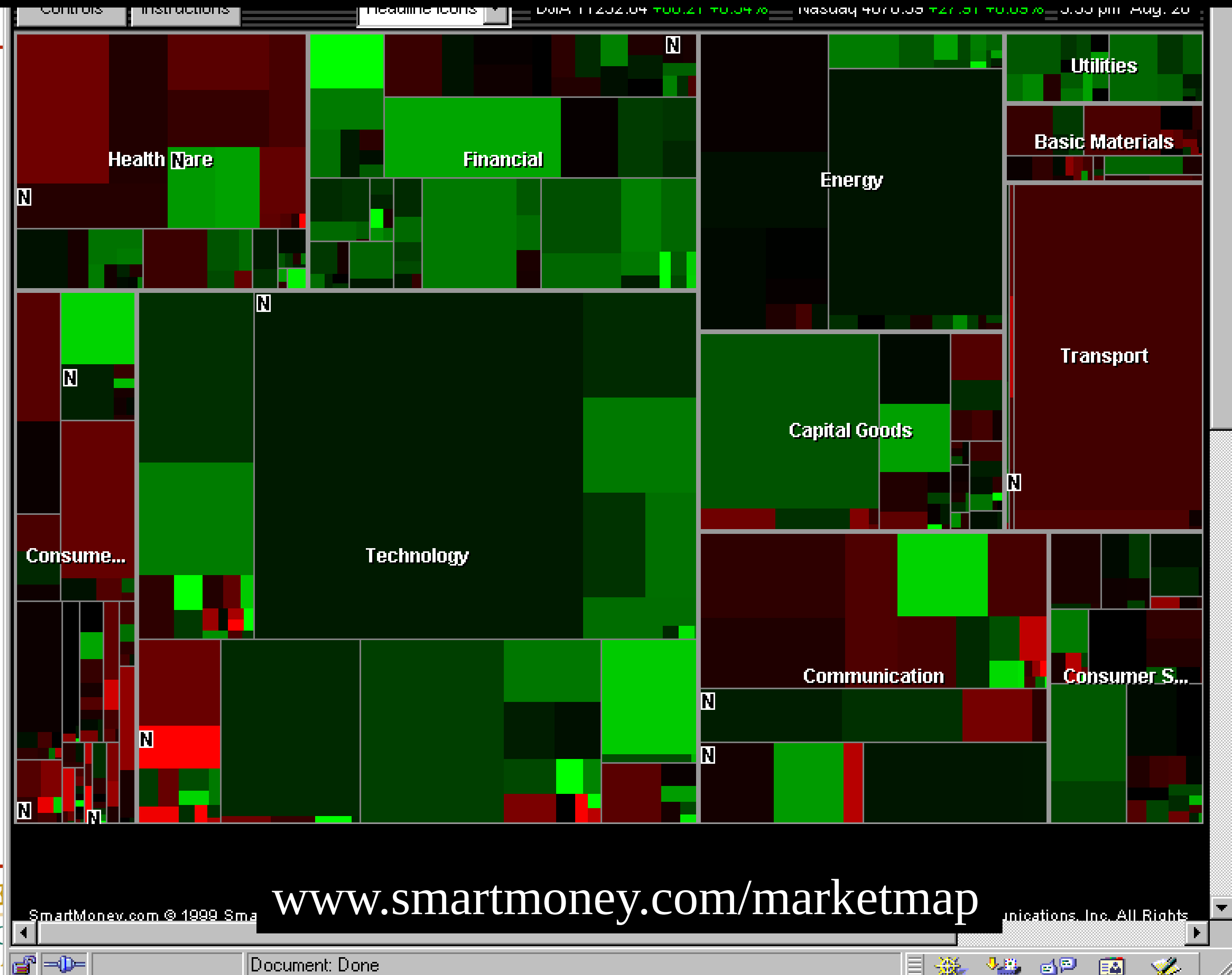
- Too disorderly
  - What does adjacency mean?
  - Aspect ratios uncontrolled leads to lots of skinny boxes that clutter
- Color not used appropriately
  - In fact, is meaningless here
- Wrong application
  - Don’t need all this to just see the largest files in the OS

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

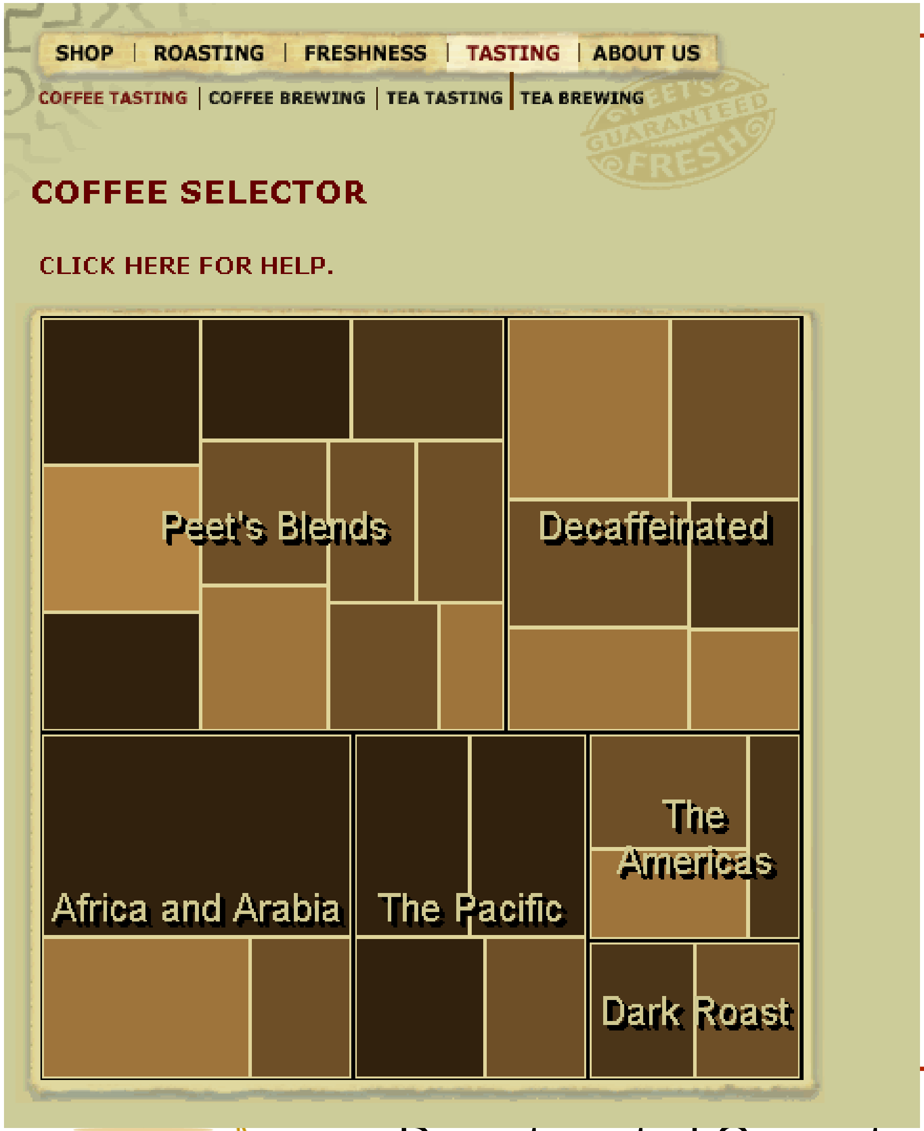
TreeMaps in Action

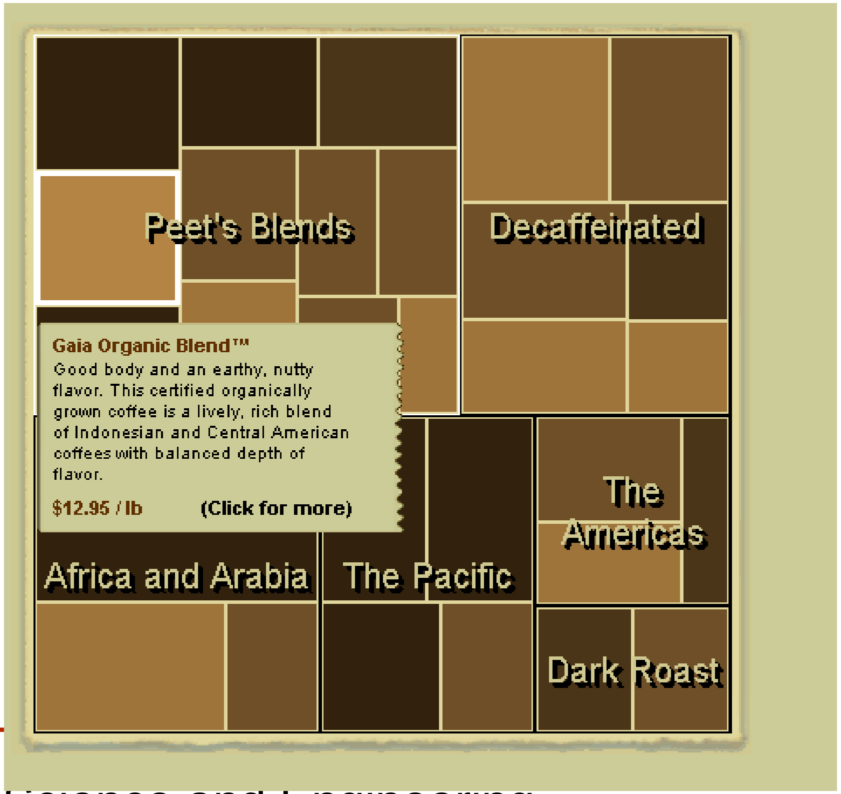
A Good Use of TreeMaps and Interactivity



www.smartmoney.com/marketmap

Treemaps in Peet's site





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

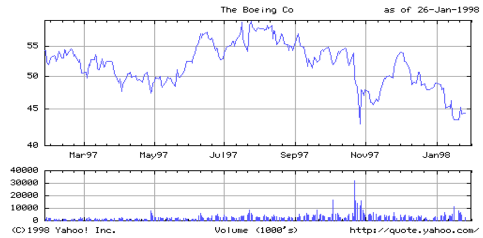
A shorthand way to present information

Allows a trend, pattern, or comparison to be easily apprehended

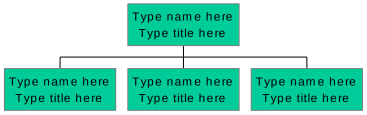
Types of Symbolic Displays

(Kosslyn 89)

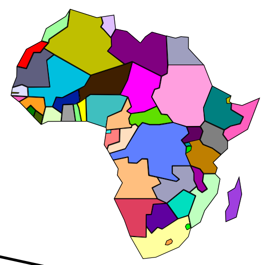
Graphs



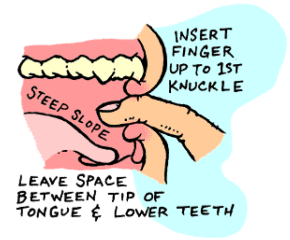
Charts



Maps



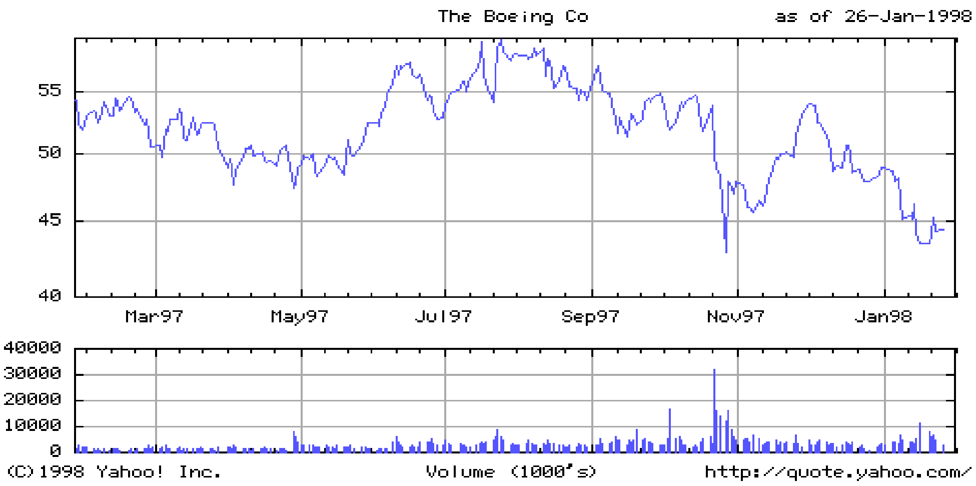
Diagrams



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



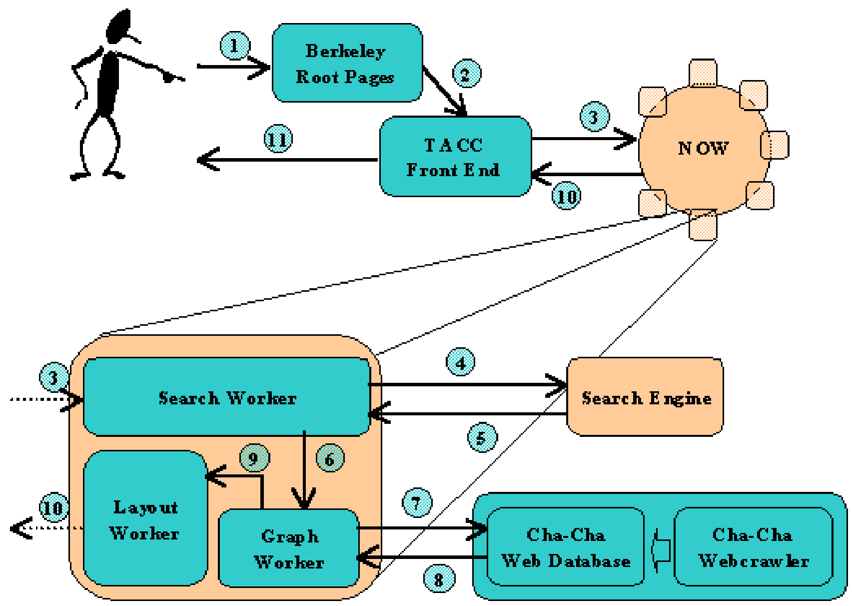
Types of Symbolic Displays

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

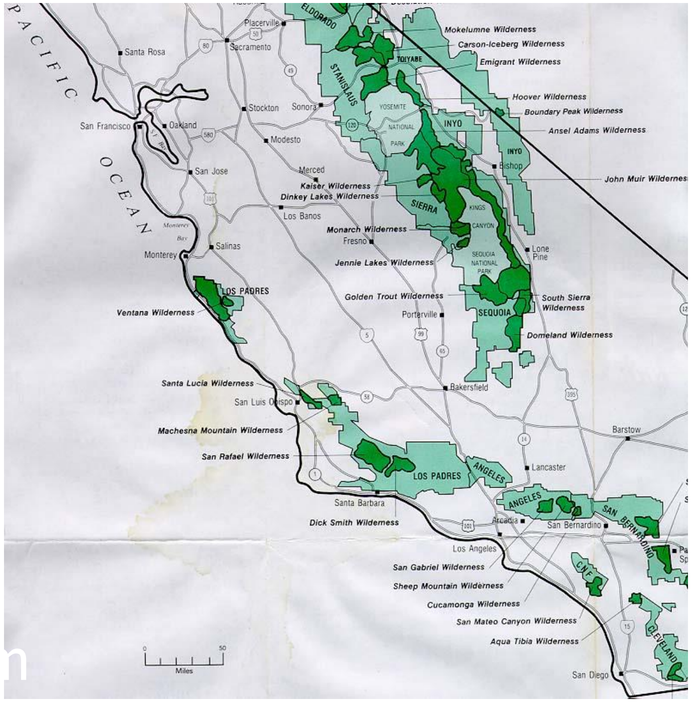
Types of Symbolic Displays

Maps

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

Examples:

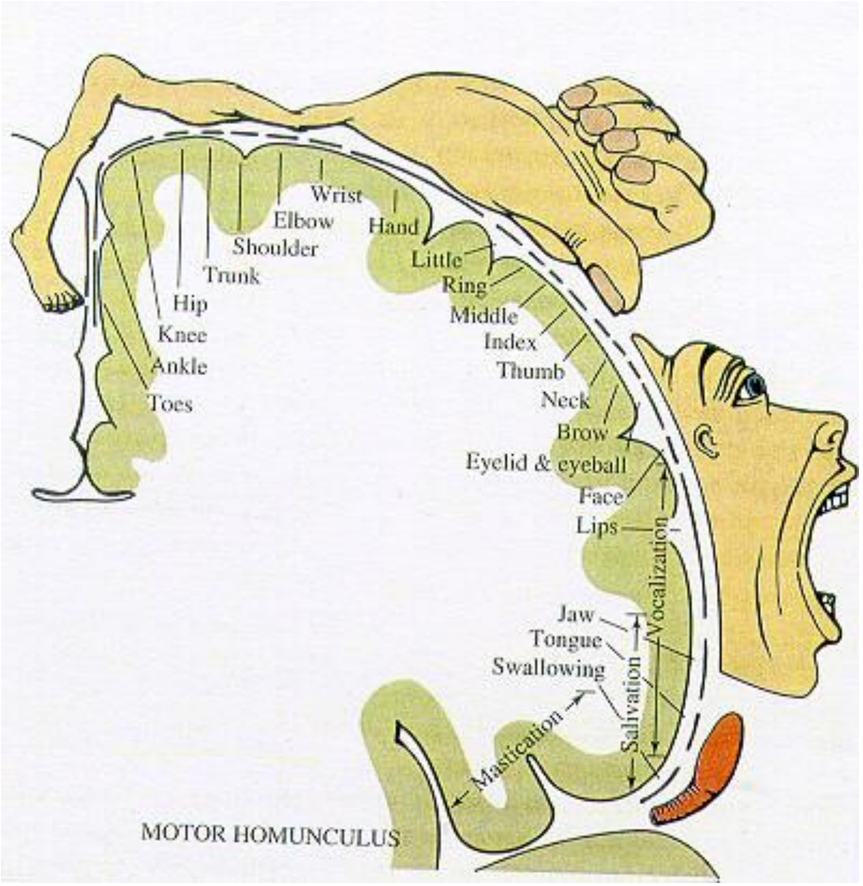
map of census data
topographic maps



Types of Symbolic Displays

Diagrams

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



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)

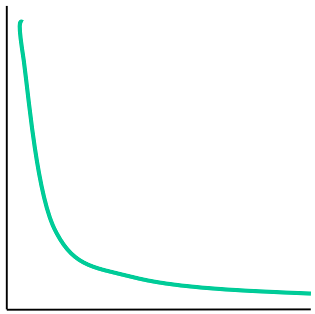
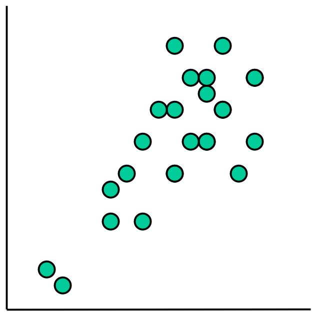
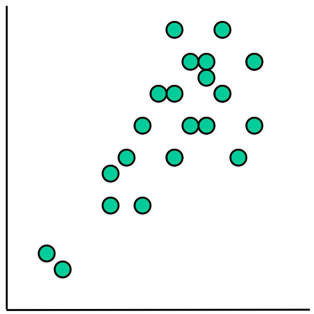
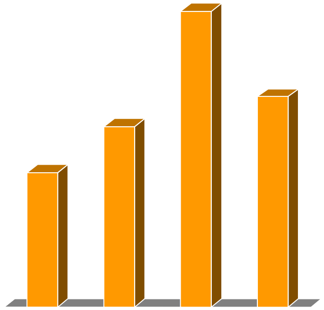
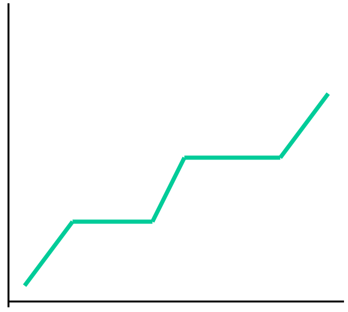
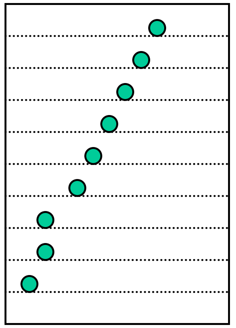
first, second, third, …

cold, warm, hot

Interval (quantitative)

list of integers or reals

Common Graph Types

Combining Data Types in Graphs

Examples?

Nominal	Nominal		
Nominal	Ordinal		
Nominal	Interval		
Ordinal	Ordinal		
Ordinal	Interval		
Interval	Interval		

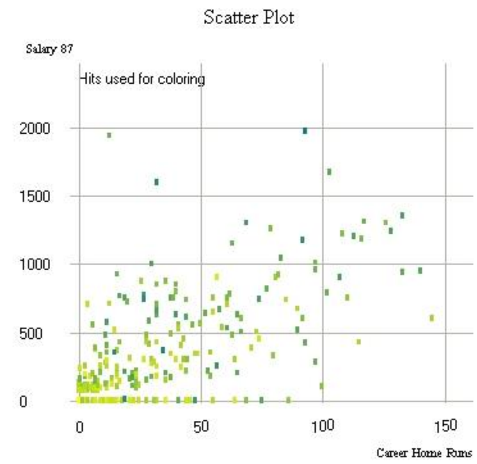
Scatter Plots

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

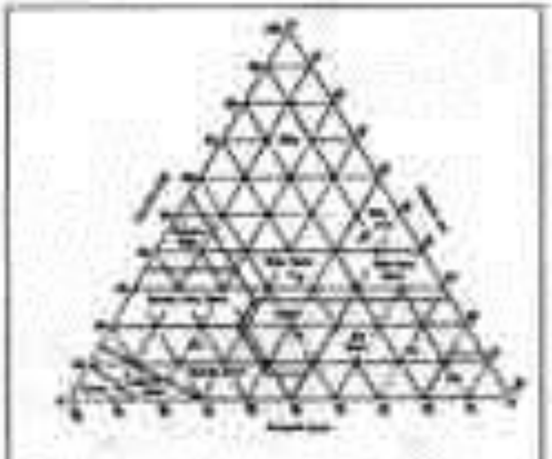
Lohse, G L; Biolsi, K; Walker, N and H H Rueter,
A Classification of Visual Representations
CACM, Vol. 37, No. 12, pp 36-49, 1994

Participants sorted 60 items into categories

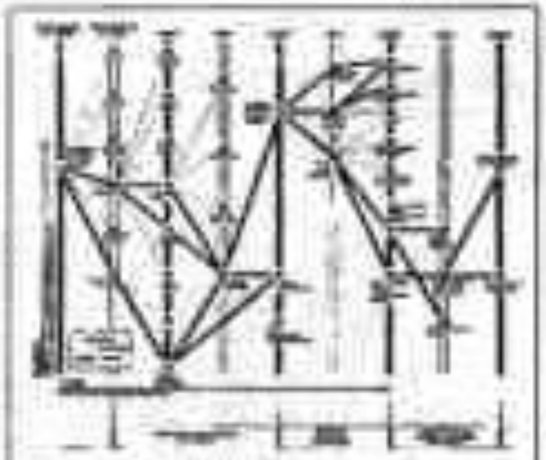
Other participants assigned labels from Likert scales

Experimenters clustered the results various ways.

Subset of Example Visual Representations From Lohse et al. 94



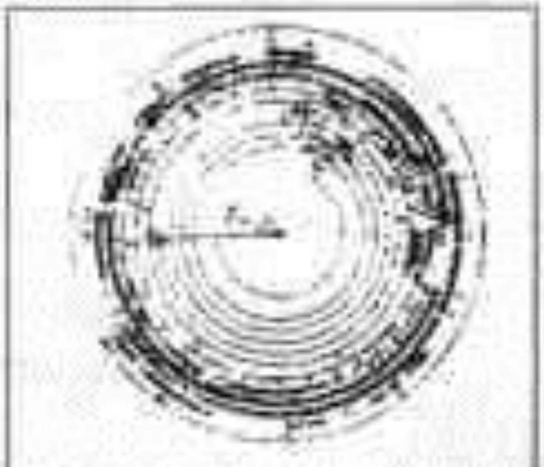
1. soil triangle



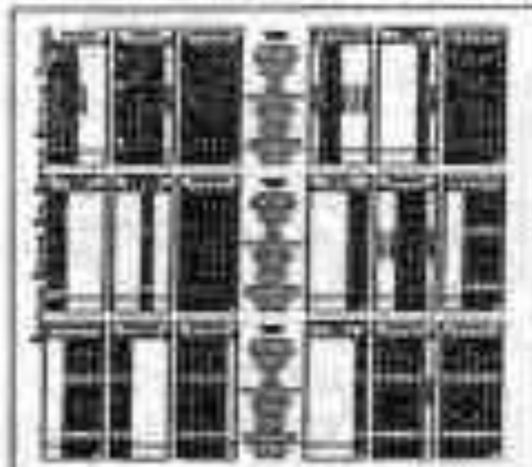
2. missile crisis

3. USA tornadoes

5. microscope



6. circular tree diagram



7. auto repair records


Lake

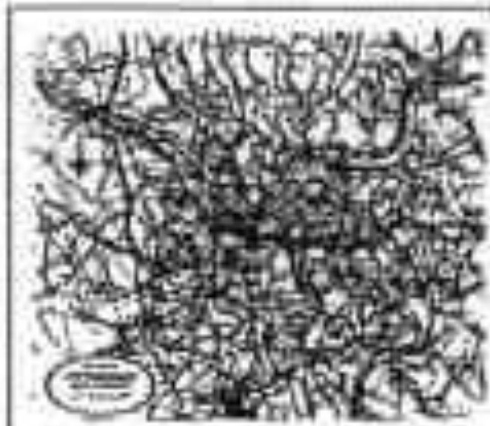

10. gibberellins


11. heart

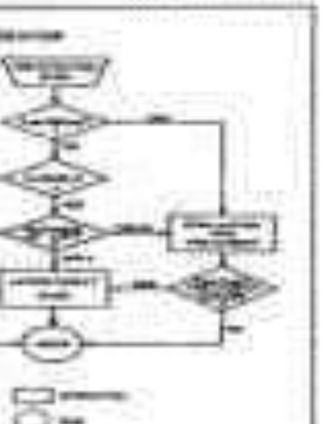

12. microbiolog


dsheet budget

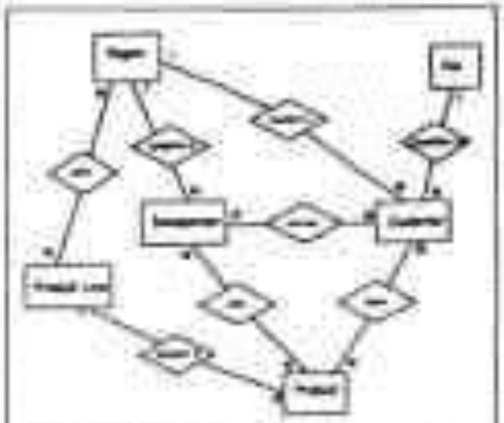

14. Gantt chart

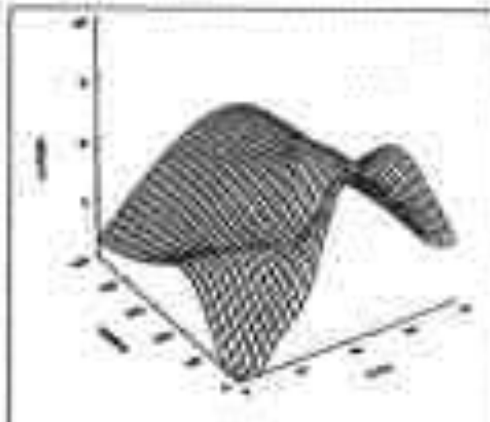

15. Pittsburgh map


16. list of integ



chart


18. data model


19. response surface


20. wheelbarro

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
  - Echos 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
  - Echos 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

All Preattentive Processing figures from Healey 97

Preattentive Processing

A limited set of visual properties are processed preattentively

(without need for focusing attention).

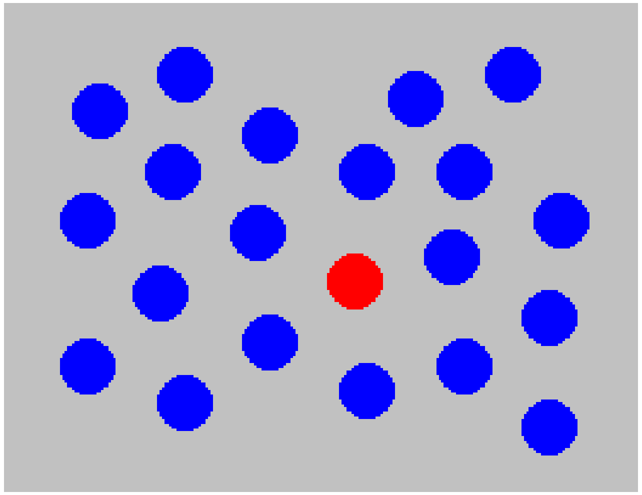
This is important for design of visualizations

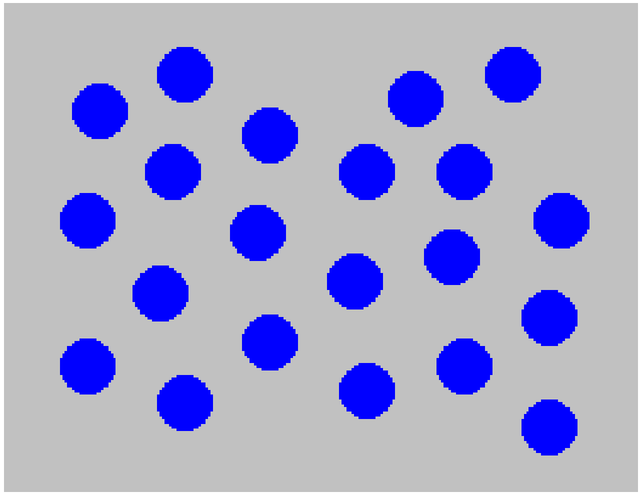
what can be perceived immediately

what properties are good discriminators

what can mislead viewers

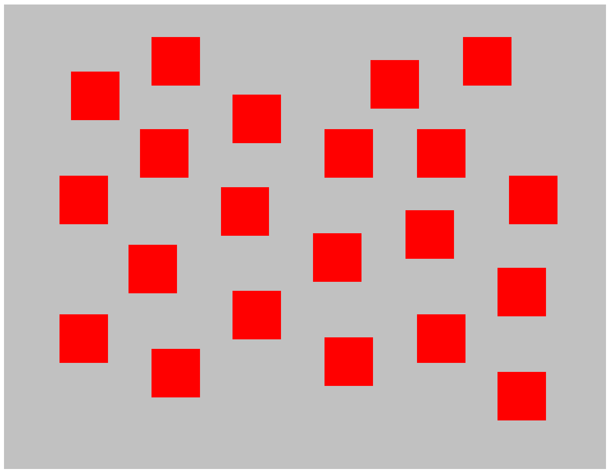
Example: Color Selection

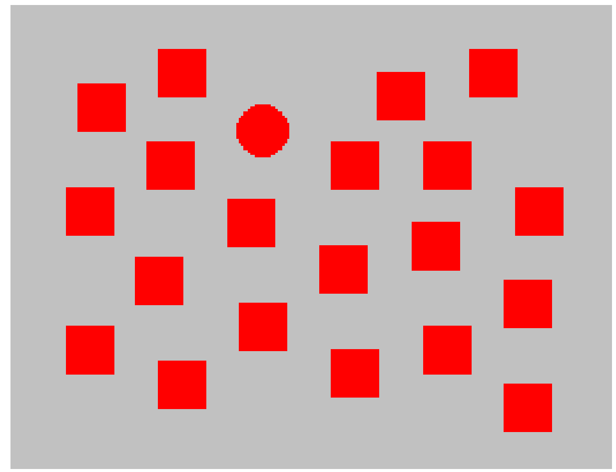




Viewer can rapidly and accurately determine whether the target (red circle) is present or absent. Difference detected in color.

Example: Shape Selection





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

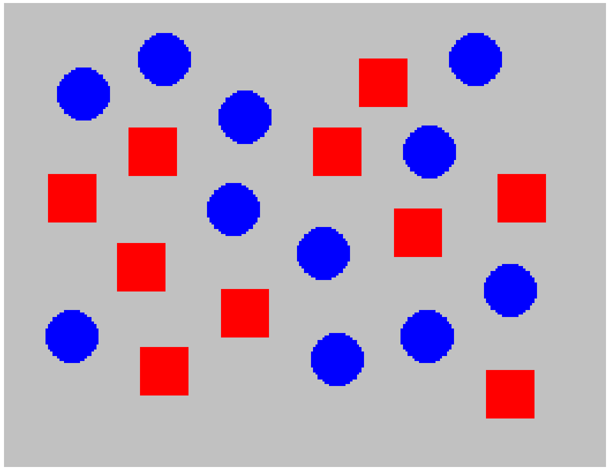
Pre-attentive Processing

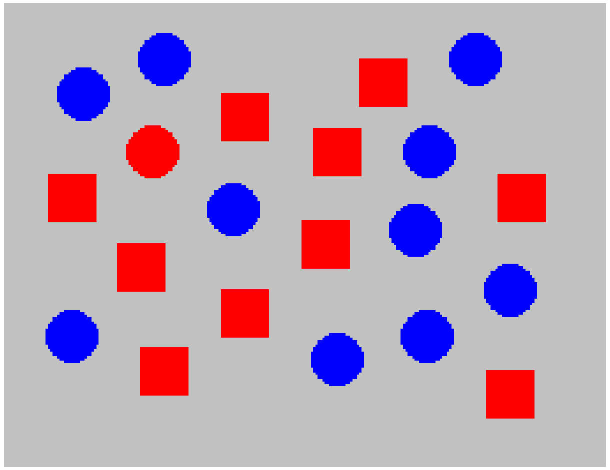
< 200 - 250ms qualifies as pre-attentive

- 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

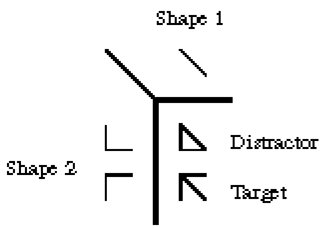


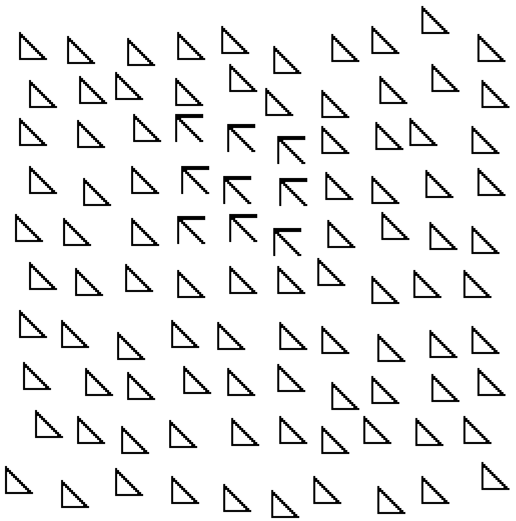


Viewer *cannot* rapidly and accurately determine whether the target (red circle) is present or absent when target has two or more features, each of which are present in the distractors. Viewer must search sequentially.

All Preattentive Processing figures from Healey 97

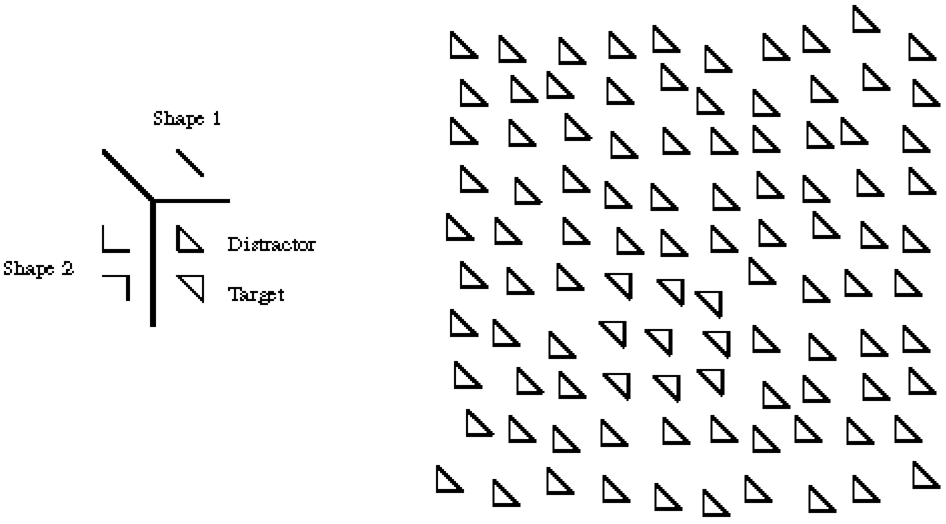
Example: Emergent Features





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.

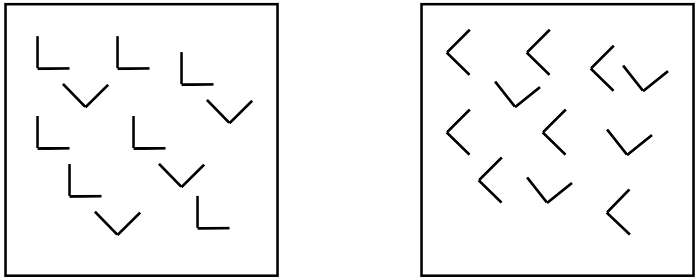
Asymmetric and Graded Preattentive Properties

Some properties are asymmetric

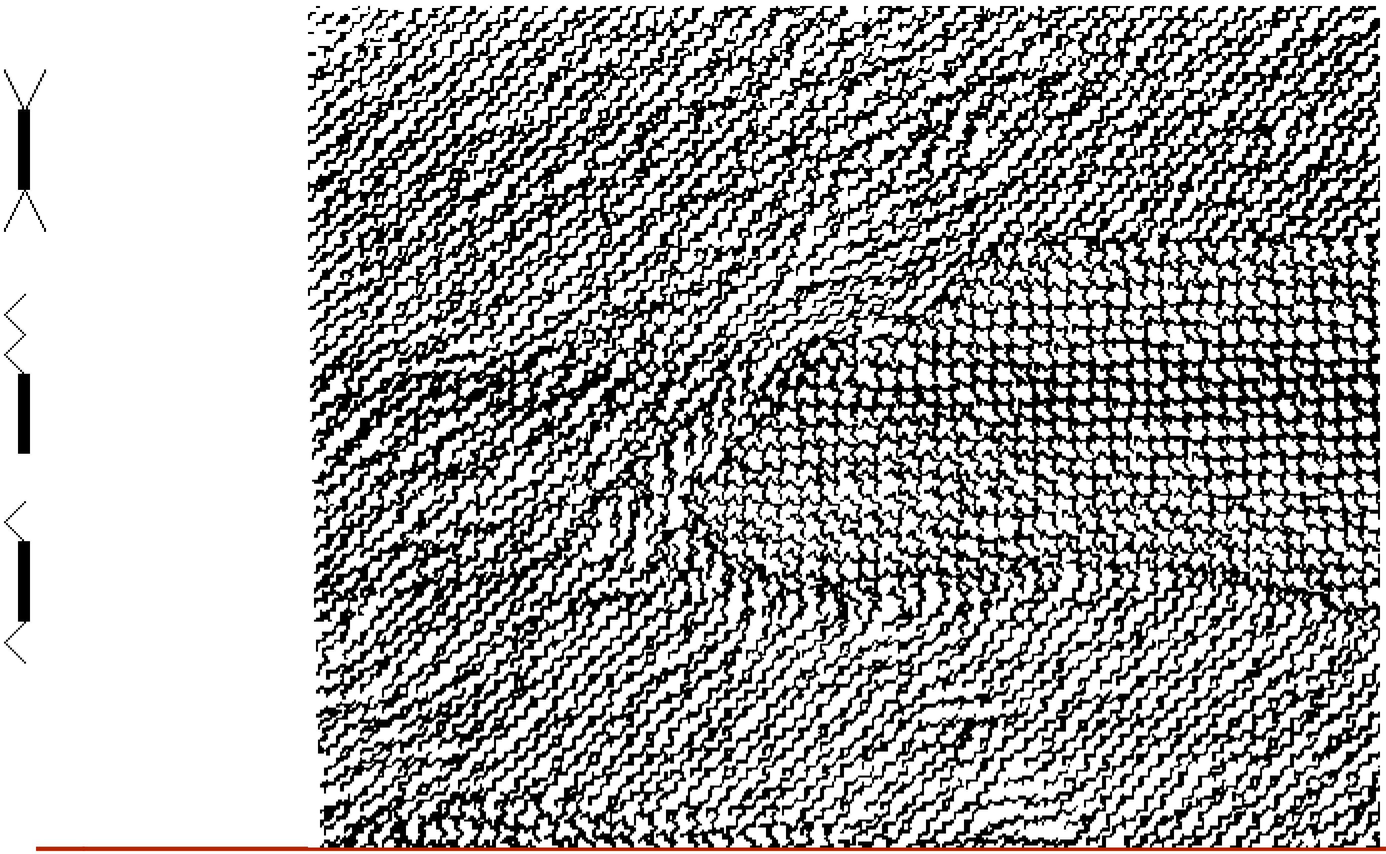
- a sloped line among vertical lines is preattentive
- a vertical line among sloped ones is not

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)

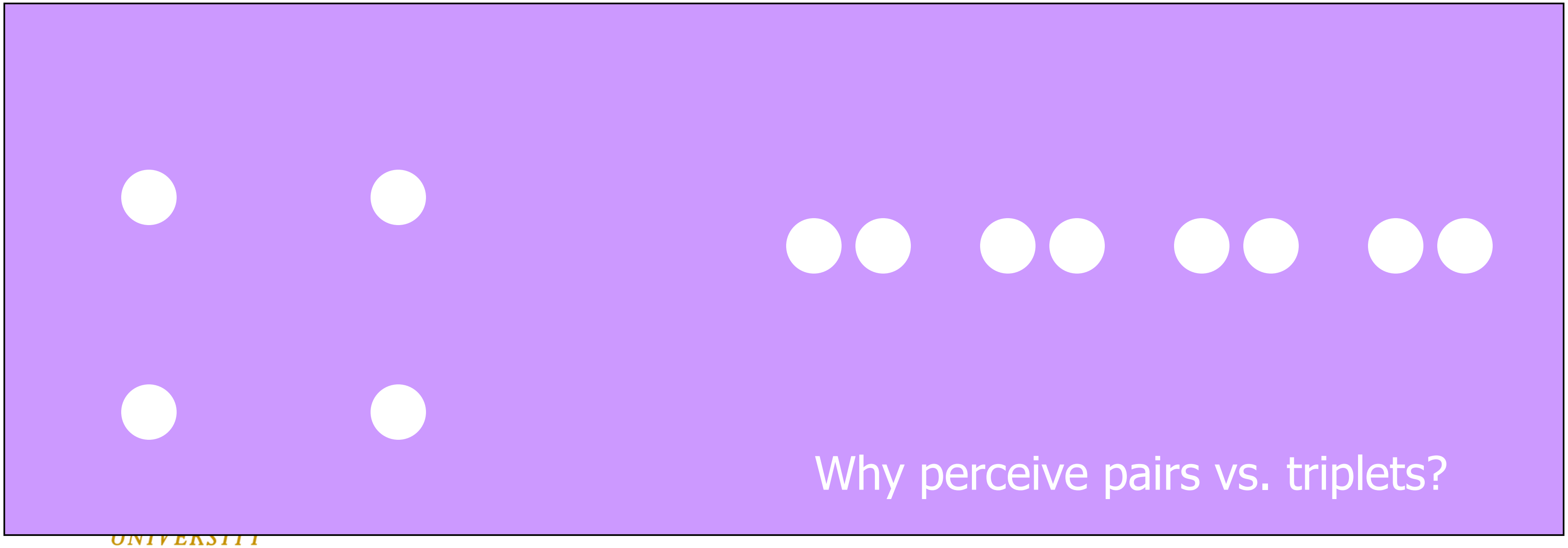
length	Triesman & Gormican [1988]
width	Julesz [1985]
size	Triesman & Gelade [1980]
curvature	Triesman & Gormican [1988]
number	Julesz [1985]; Trick & Pylyshyn [1994]
terminators	Julesz & Bergen [1983]
intersection	Julesz & Bergen [1983]
closure	Enns [1986]; Triesman & Souther [1985]
colour (hue)	Nagy & 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?



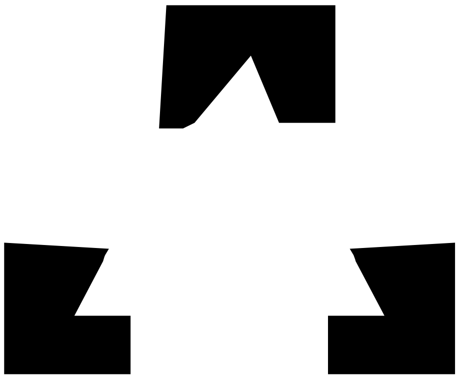
Gestalt Laws of Perceptual Organization

(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

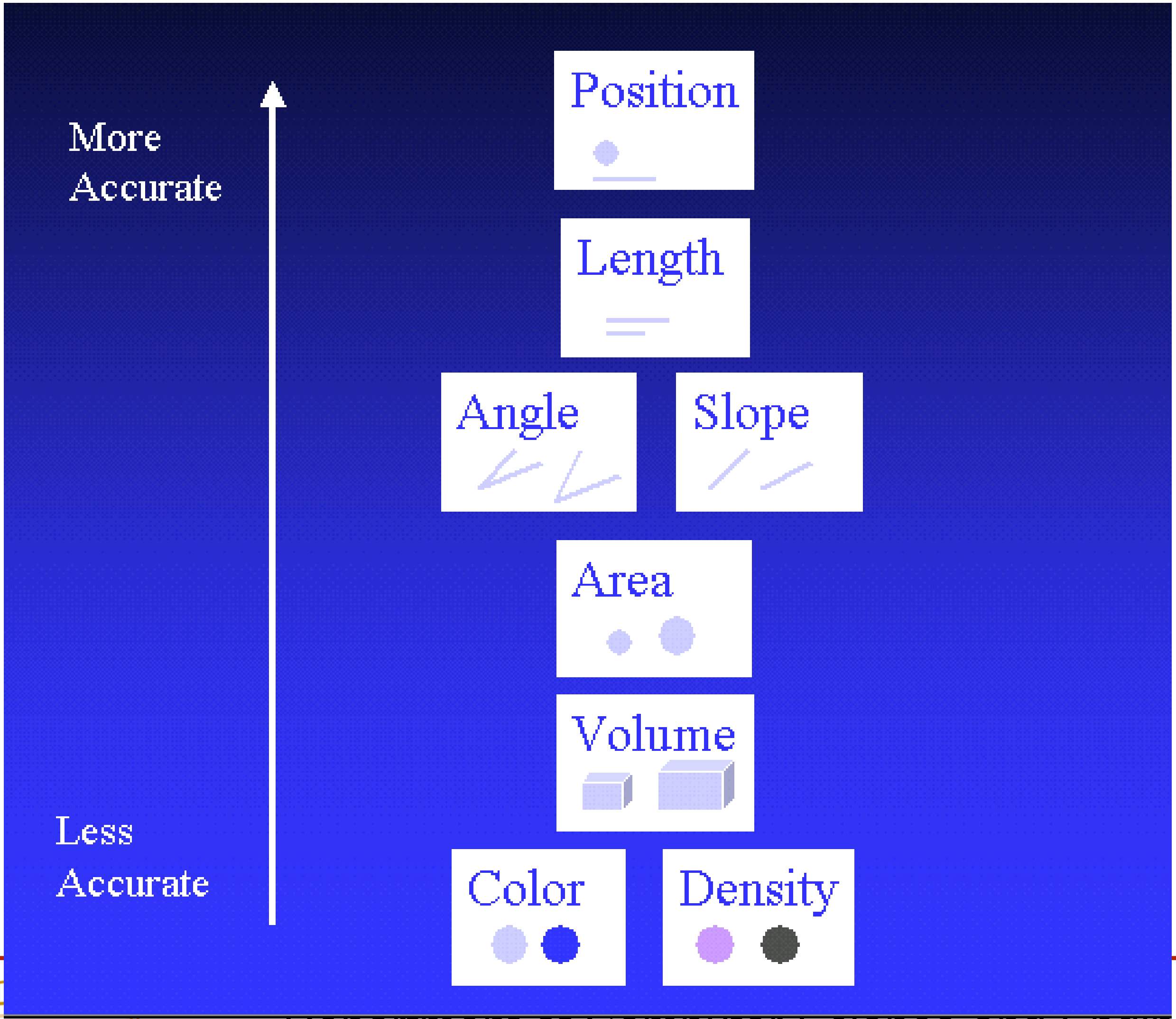
- like the preattentive processing examples

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

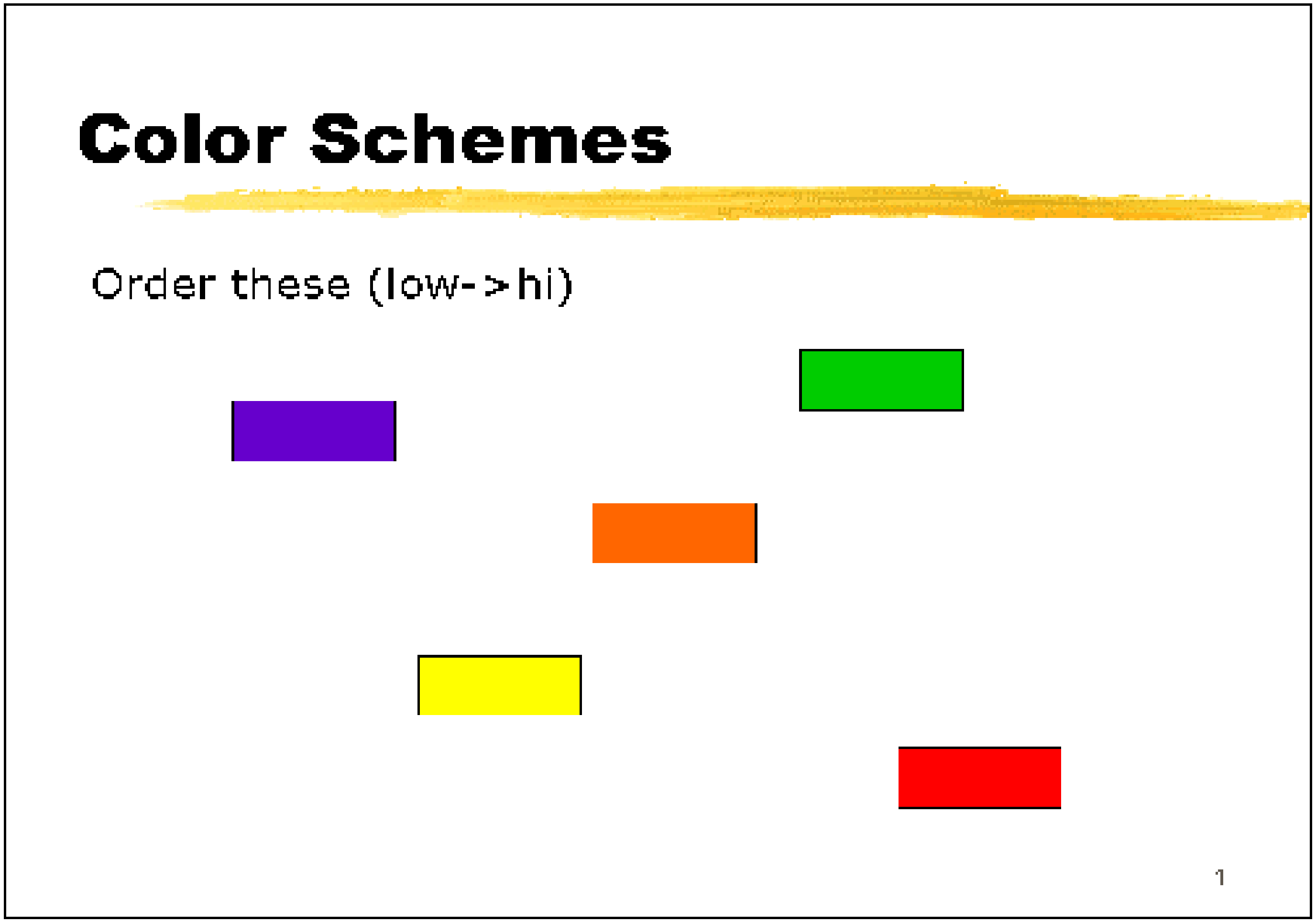
(Senay & Ingatious 97, Kosslyn, others)

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

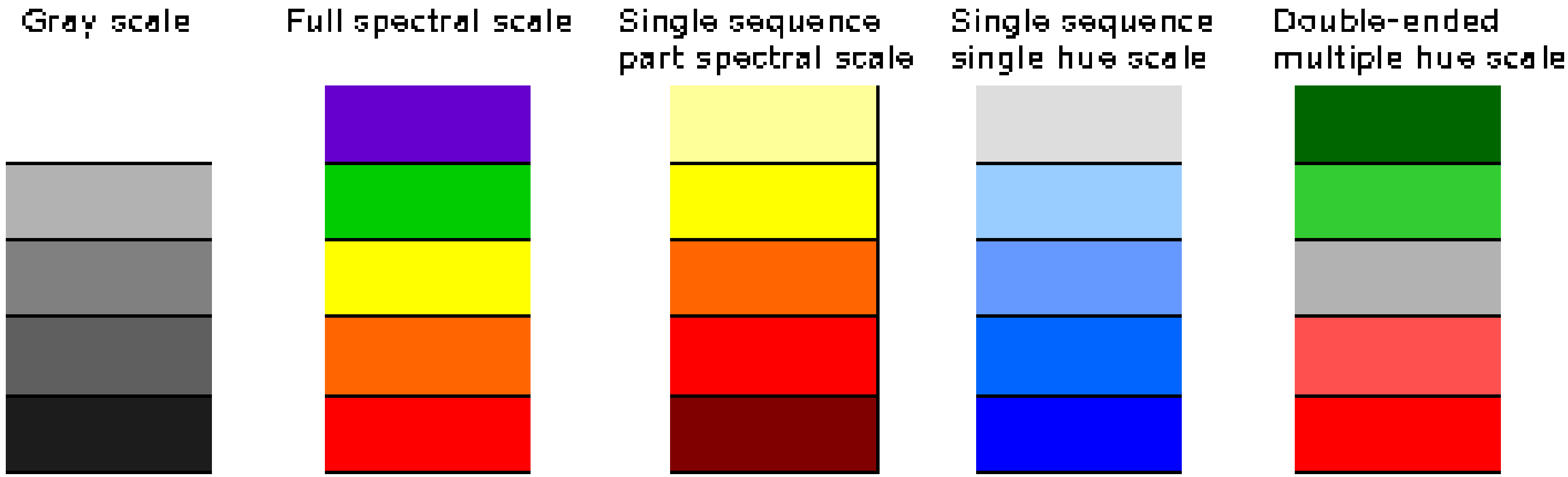
Ranking of Applicability of Properties for Different Data Types

(Mackinlay 88, Not Empirically Verified)

QUANTITATIVE	ORDINAL	NOMINAL
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)
1

Color Schemes



1

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 they way they "should".

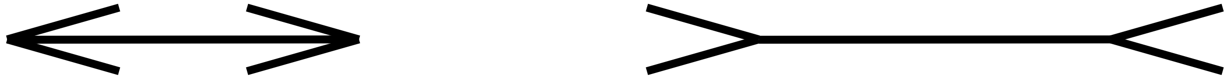
Some illusions have been reclassified as systematic perceptual errors

- 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

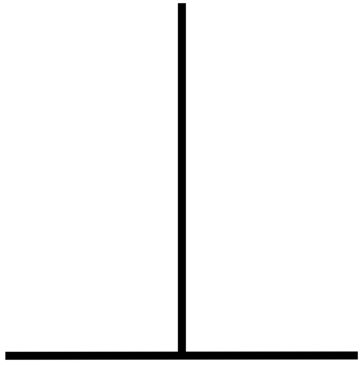
Nevertheless, the visual system does some really unexpected things.

Illusions of Linear Extent

Mueller-Lyon (off by 25-30%)

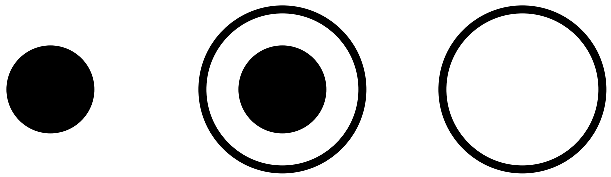


Horizontal-Vertical



Illusions of Area

Delboeuf Illusion



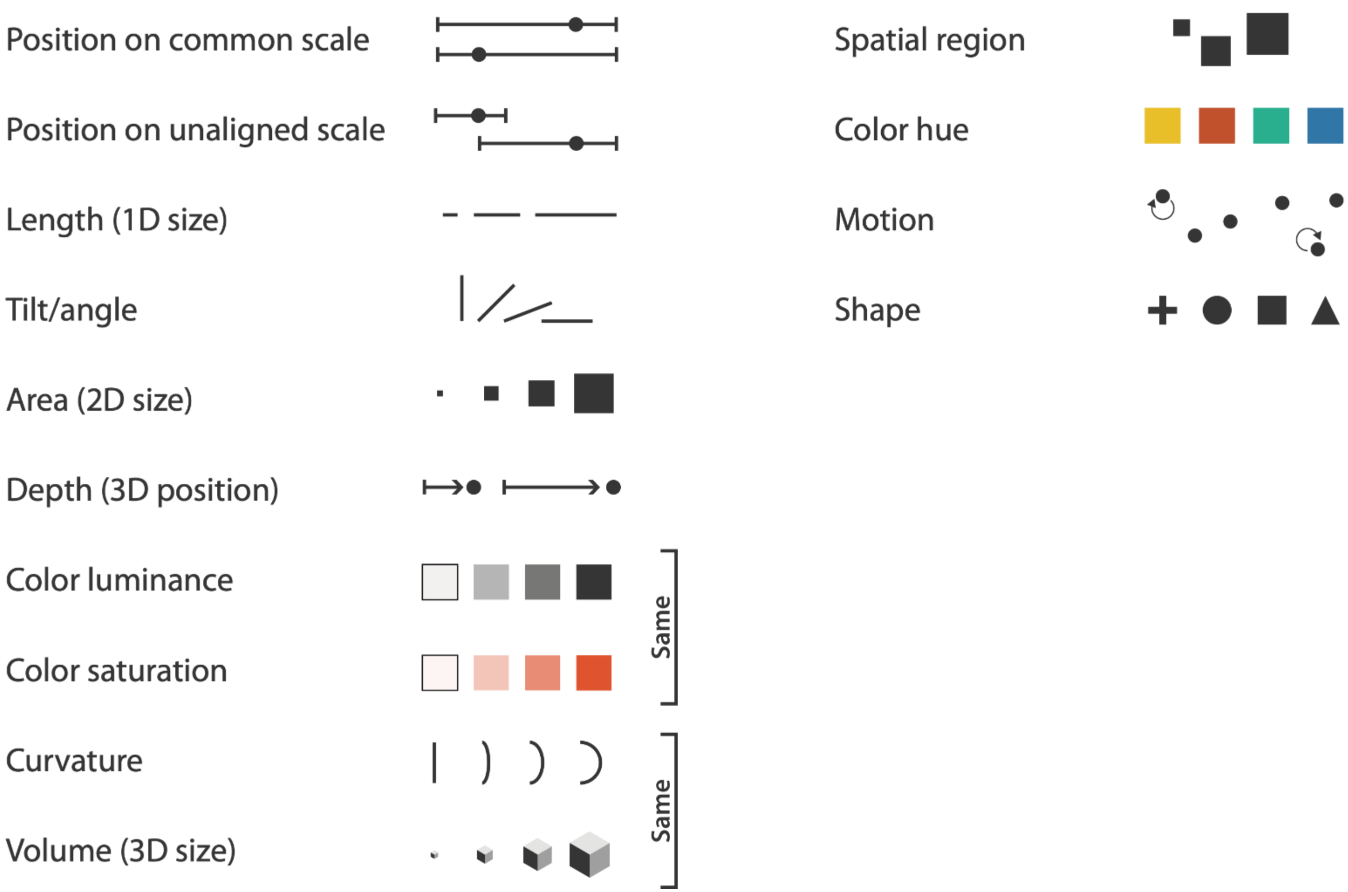
Height of 4-story building overestimated by approximately 25%

The two filled black circles are exactly the same size; however, the one on the left may *seem* larger or smaller.

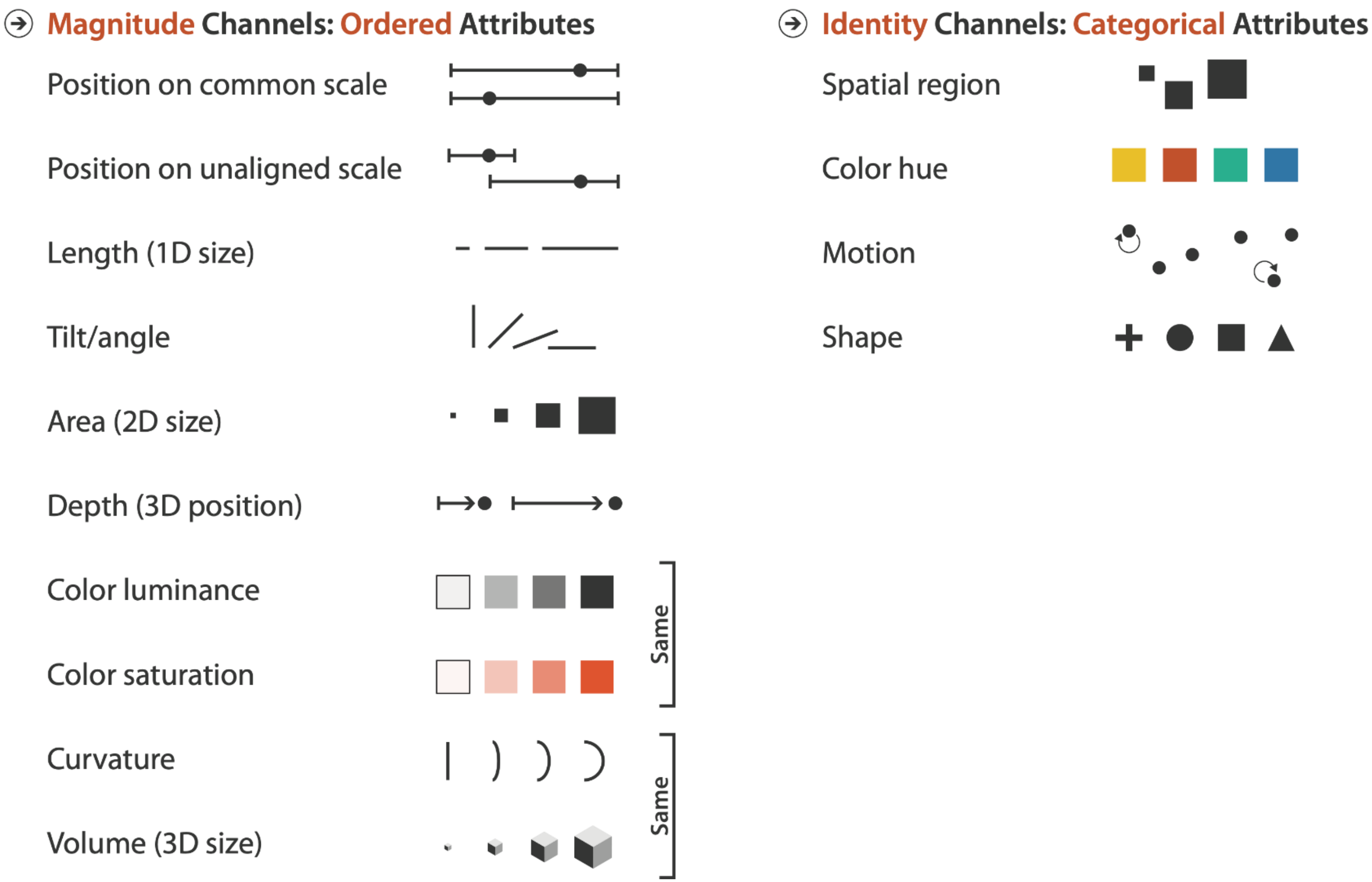
What are good guidelines for Infoviz?

- 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

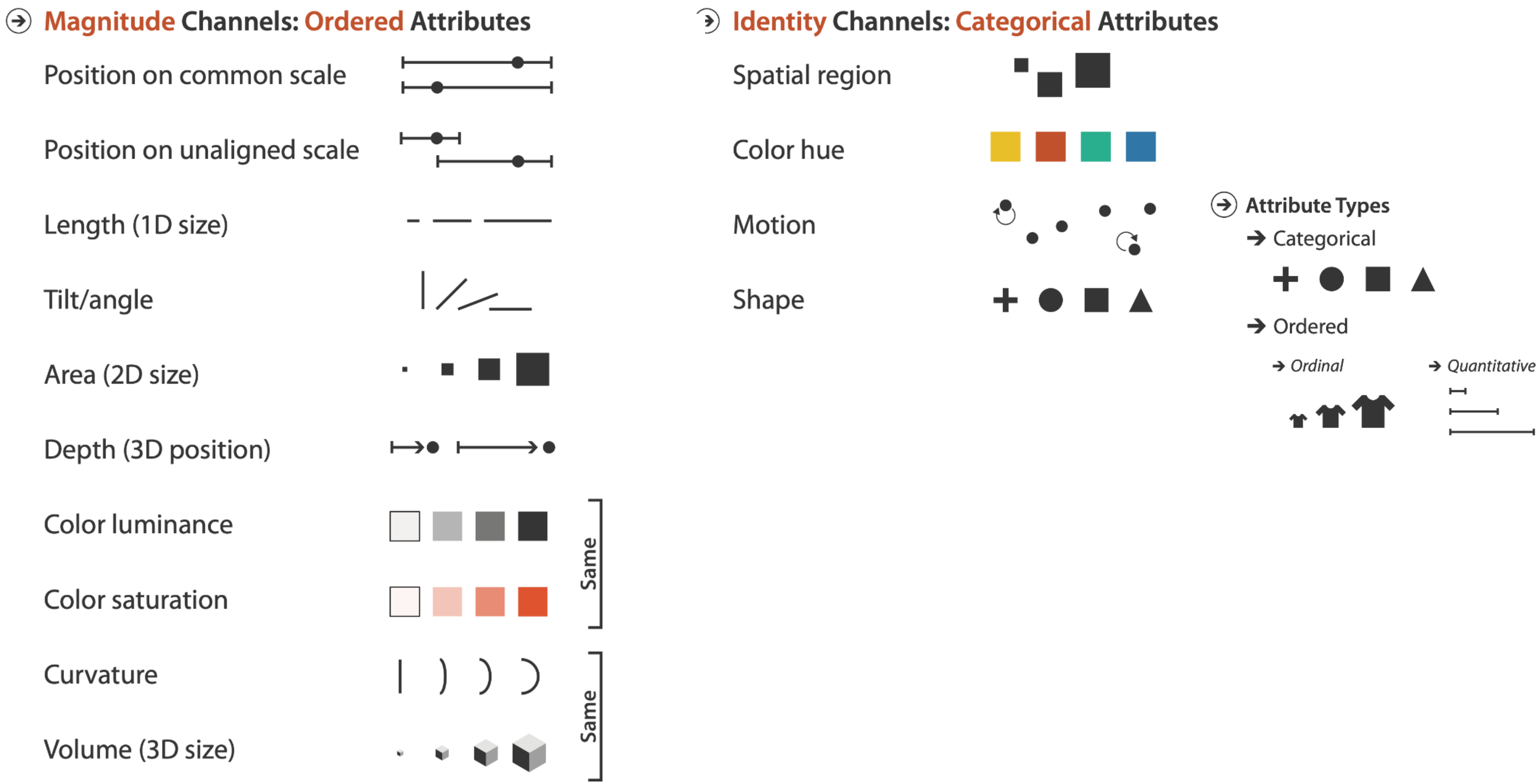
Channels: Rankings



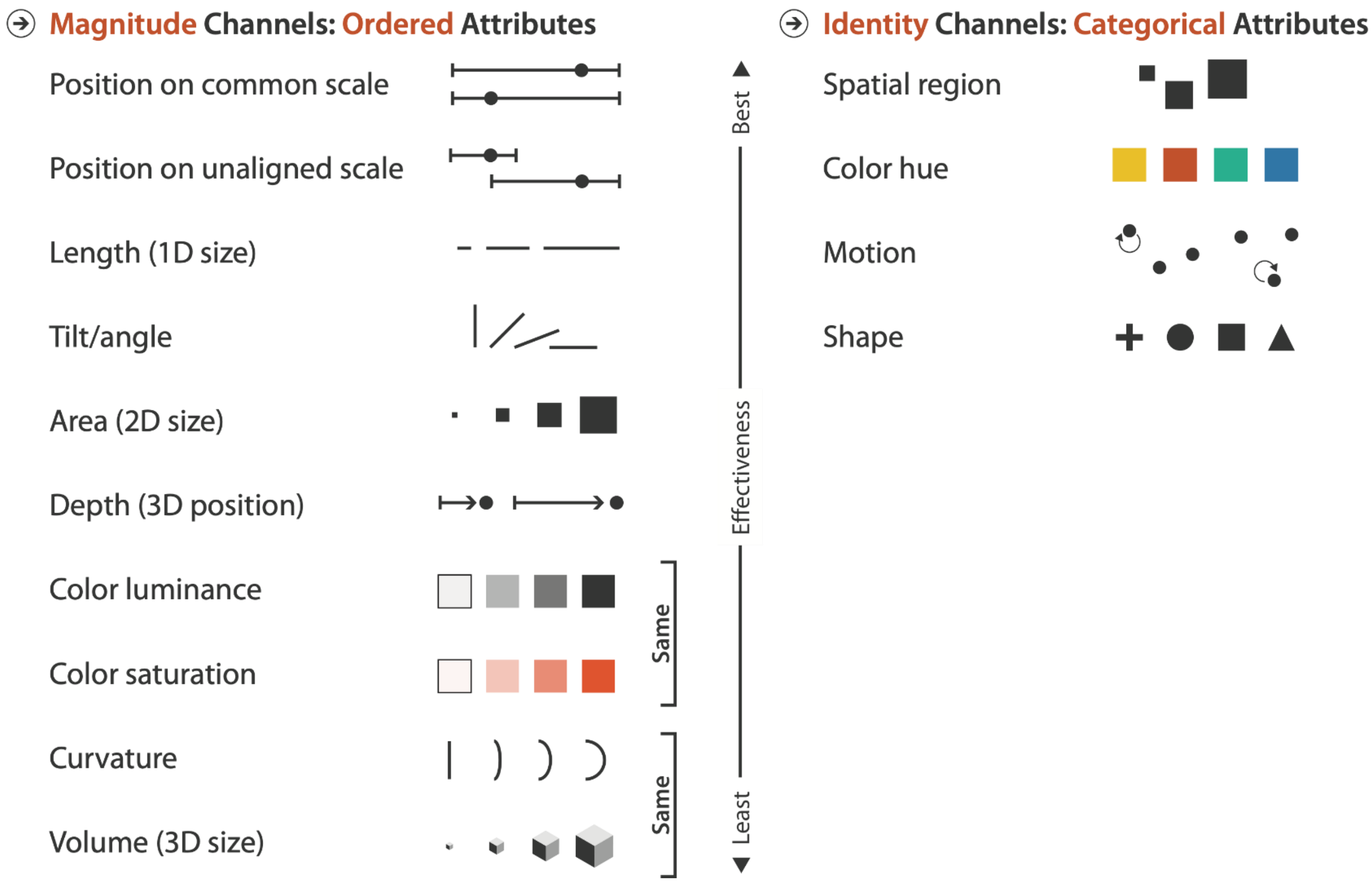
Channels: Rankings



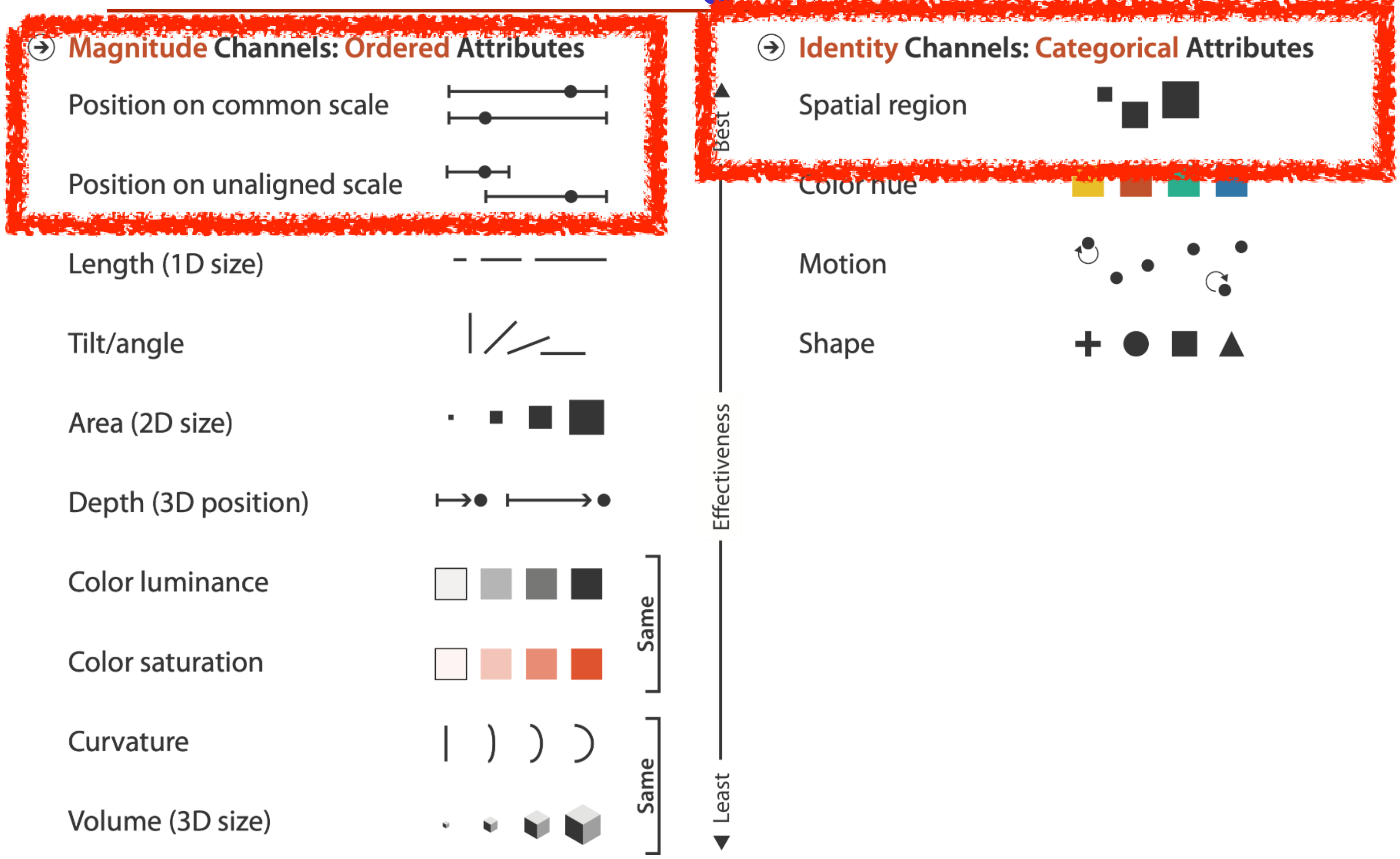
Channels: Rankings



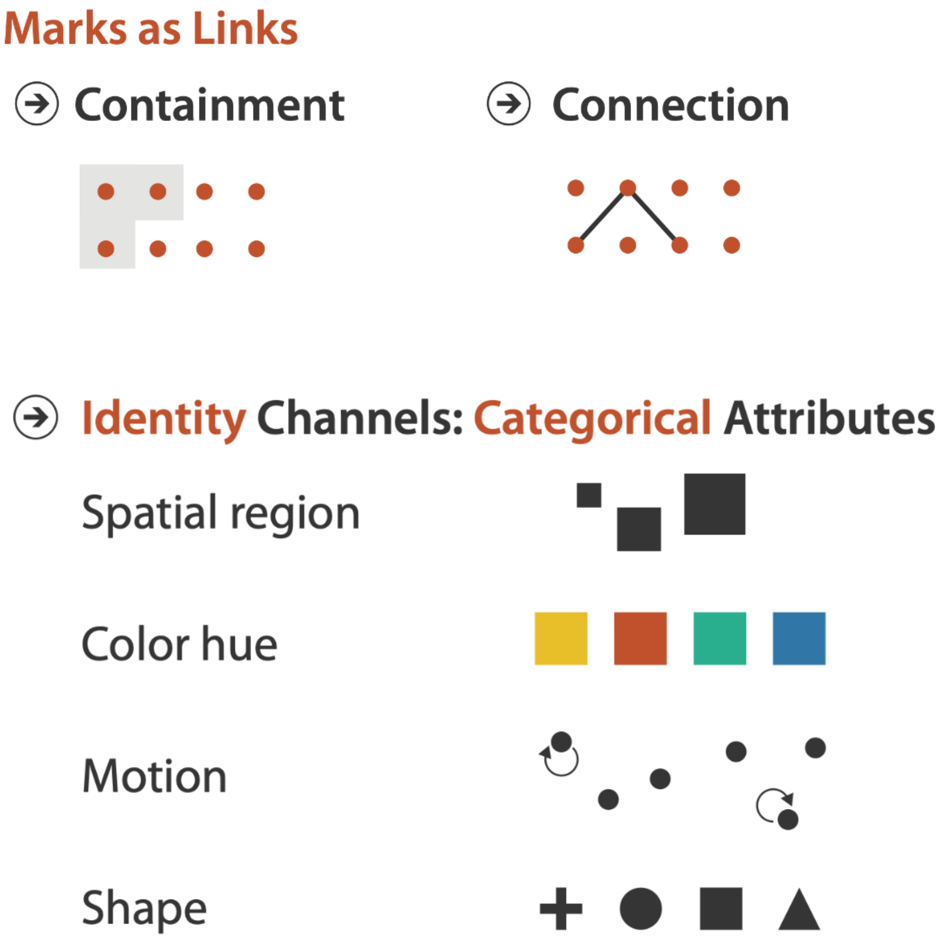
Channels: Rankings



Channels: Rankings



Grouping



containment

connection

proximity

– same spatial region

similarity

– same values as other categorical channels

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 greates 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 = $\dfrac{\text{size of effect in graph}}{\text{size of effect in data}}$

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

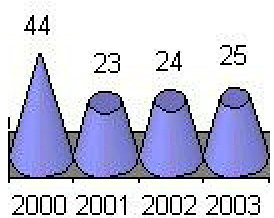
LIS 504 - Graphic displays of data - Netscape

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.

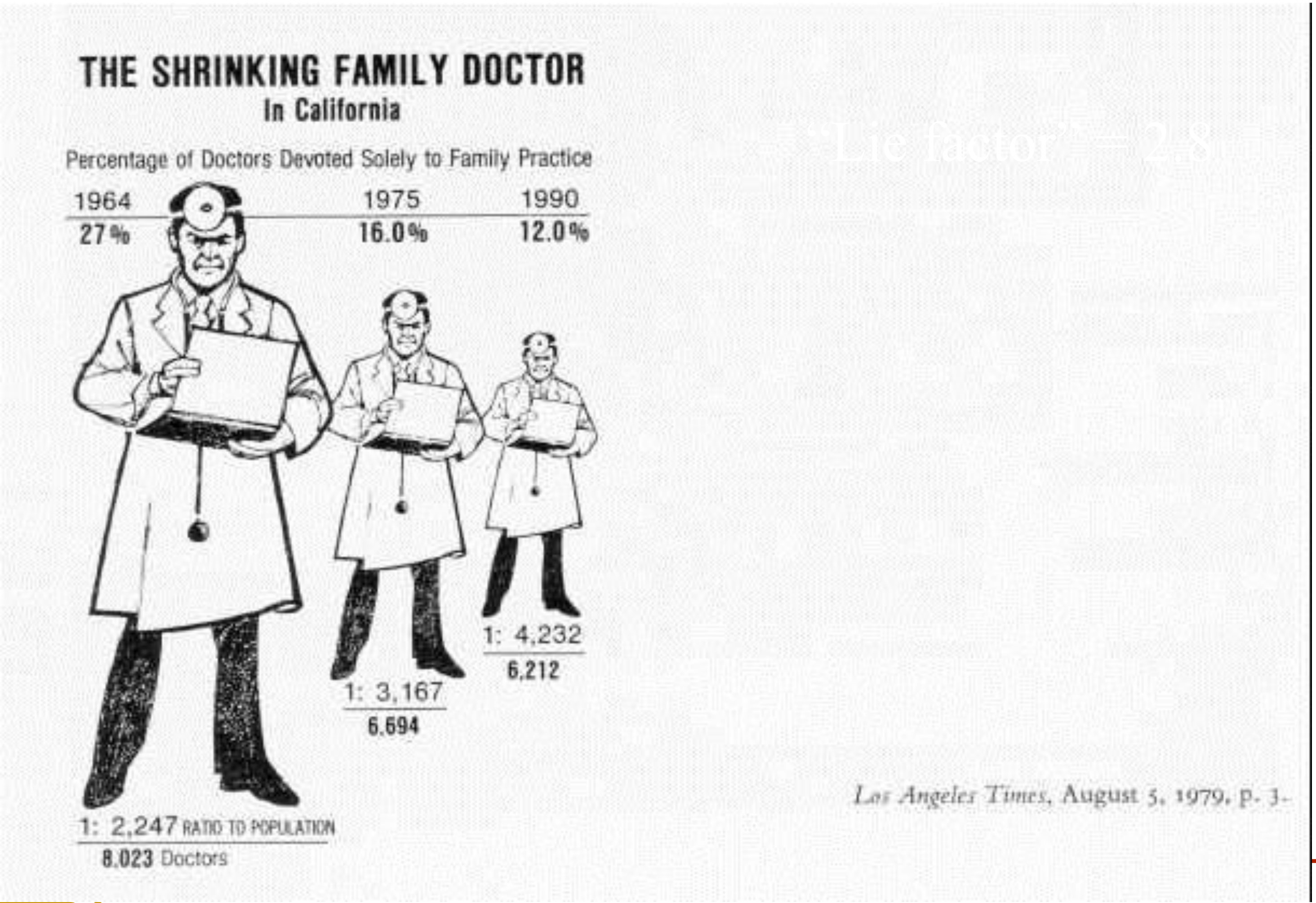


The heights of the (truncated) cones are proportional to the data, but their areas on the screen and their apparent volumes make the larger data values seem relatively small.

Document: Done

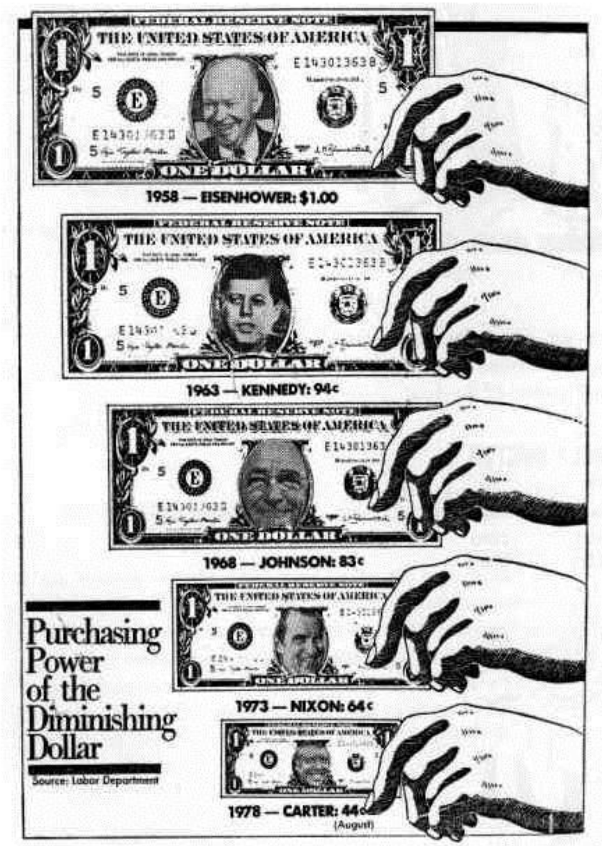
How to Exaggerate with Graphs

from Tufte '83



How to Exaggerate with Graphs

from Tufte '83



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

Howard Wainer
How to Display Data Badly (Video)

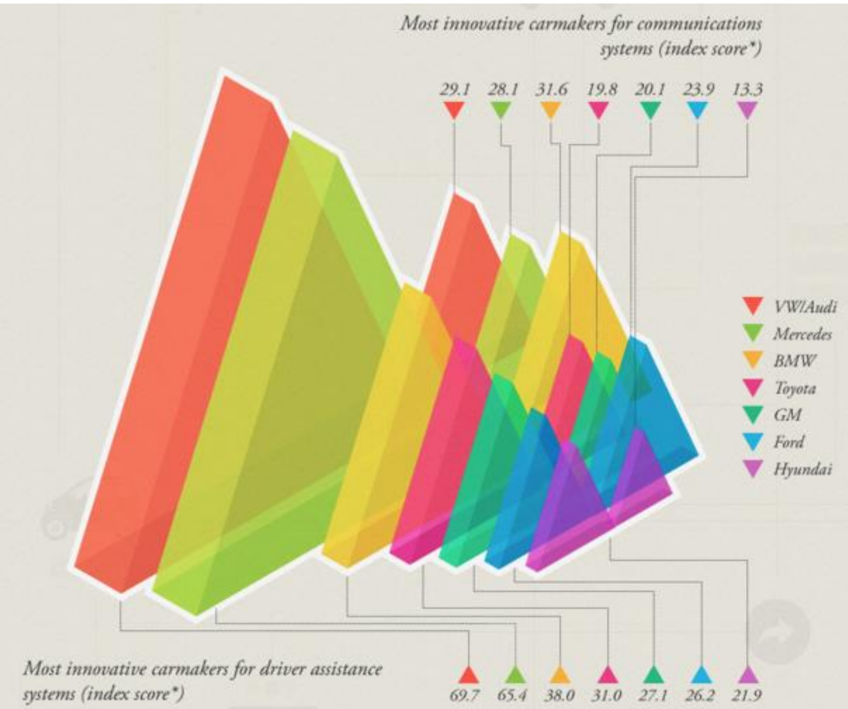
http://www.dartmouth.edu/~chance/ChanceLecture/AudioVideo.html

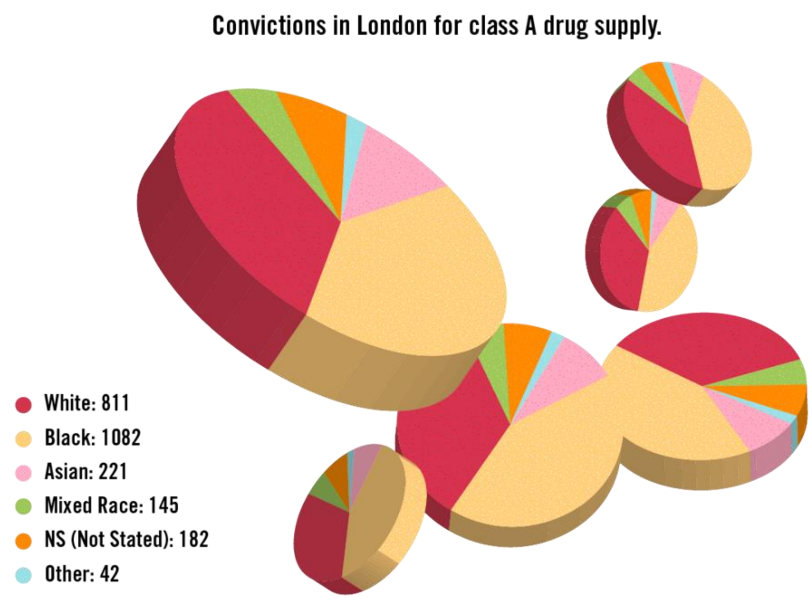
Rules of Thumb

Guidelines and considerations, not absolute rules

- when to use 3D? when to use 2D?
- when to use eyes instead of memory?
- when does immersion help?
- when to use overviews?
- how long is too long?
- which comes first, form or function?

Unjustified 3D all too common, in the news and elsewhere





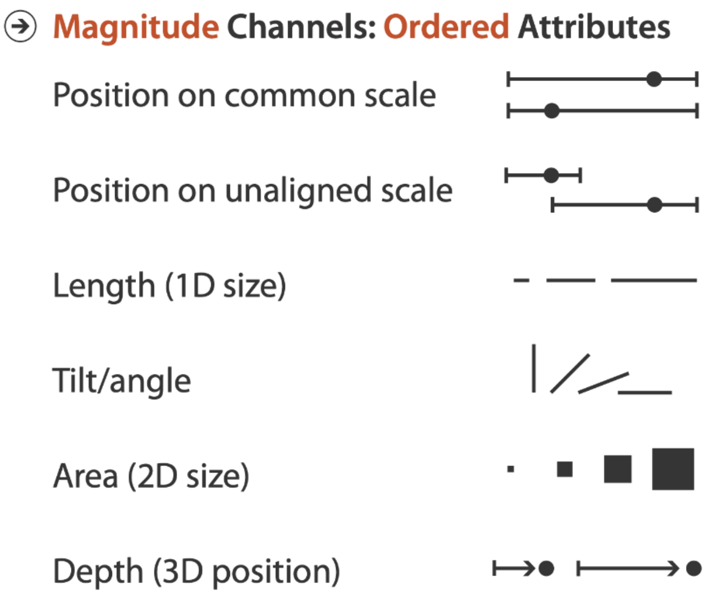
http://viz.wtf/post/137826497077/eye-popping-3d-triangles

http://viz.wtf/post/139002022202/designer-drugs-ht-ducqn

Depth vs power of the plane

high-ranked spatial position channels: **planar** spatial position

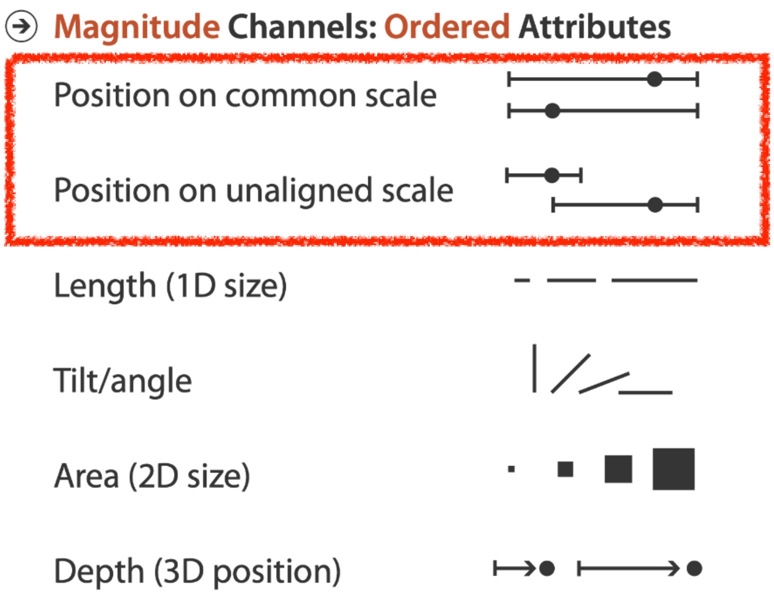
– not depth!



Depth vs power of the plane

high-ranked spatial position channels: **planar** spatial position

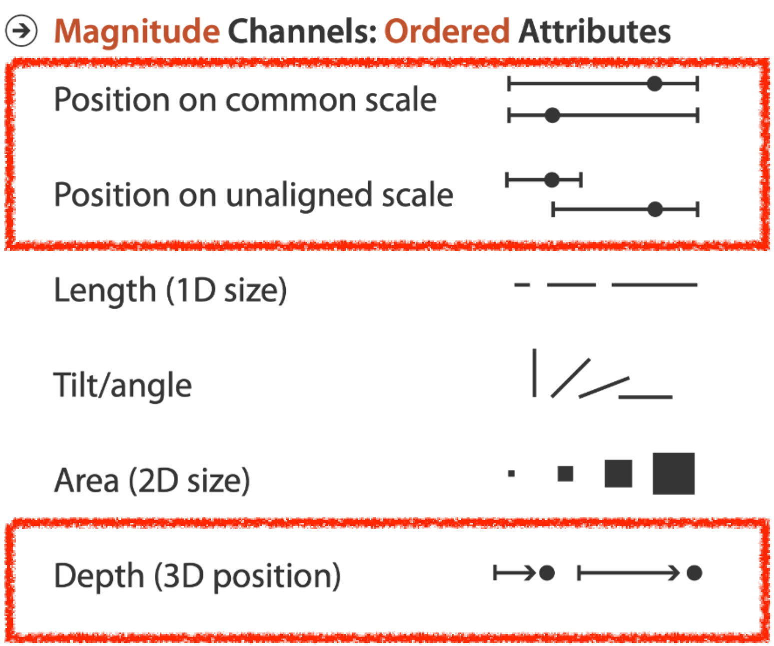
– not depth!



Depth vs power of the plane

high-ranked spatial position channels: **planar** spatial position

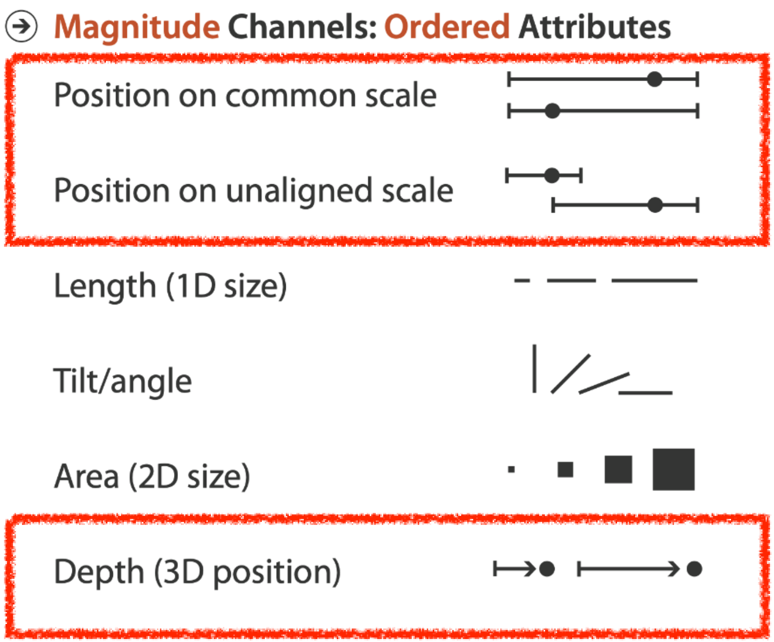
- not depth!

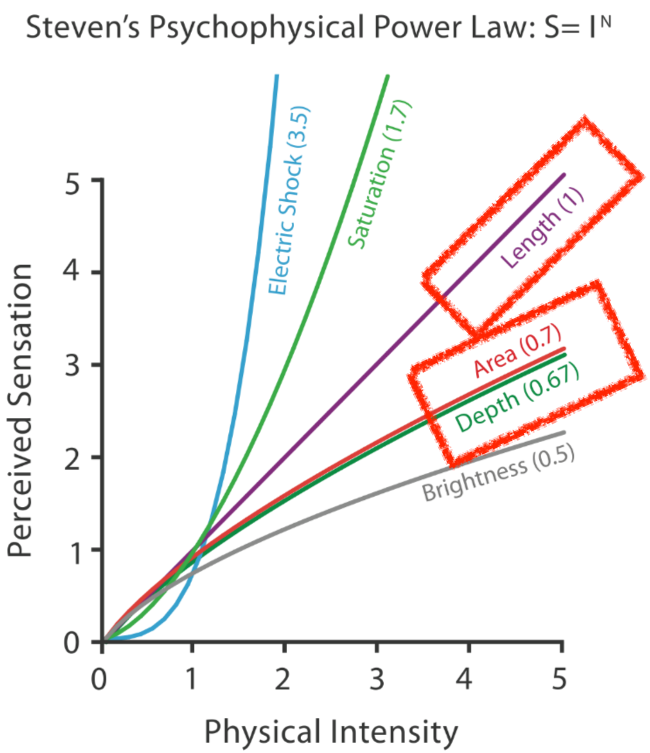


Depth vs power of the plane

high-ranked spatial position channels: **planar** spatial position

– not depth!

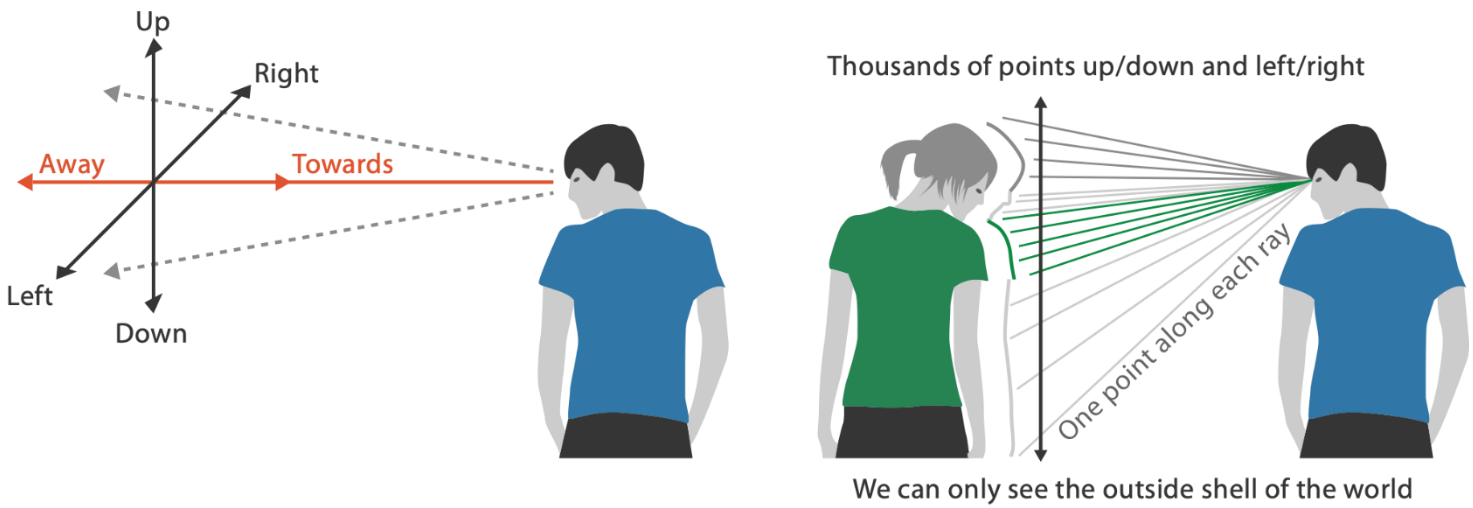




No unjustified 3D: Danger of depth

we don’t really live in 3D: we ***see*** in 2.05D

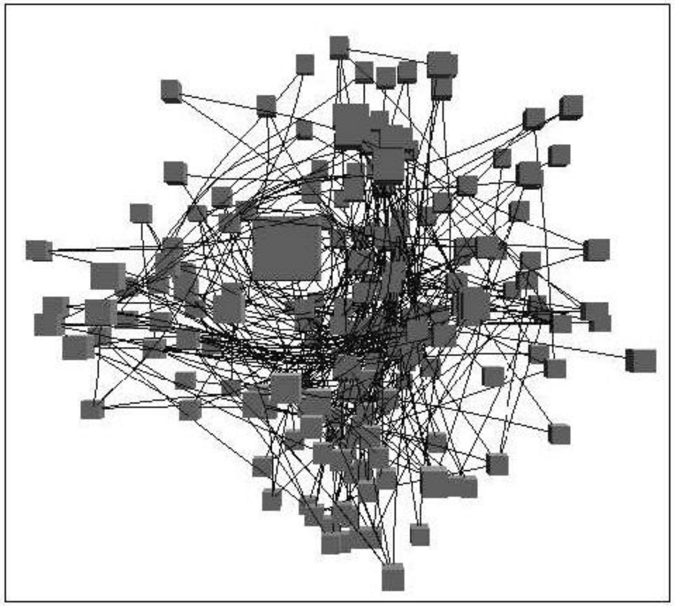
- acquire more info on image plane quickly from eye movements
- acquire more info for depth slower, from head/body motion



Occlusion hides information

occlusion

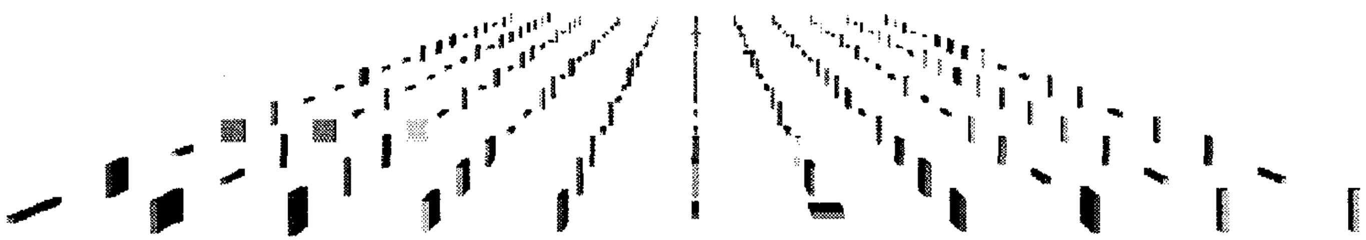
interaction can resolve, but at cost of time and cognitive load



Perspective distortion loses information

perspective distortion

- interferes with all size channel encodings
- power of the plane is lost!

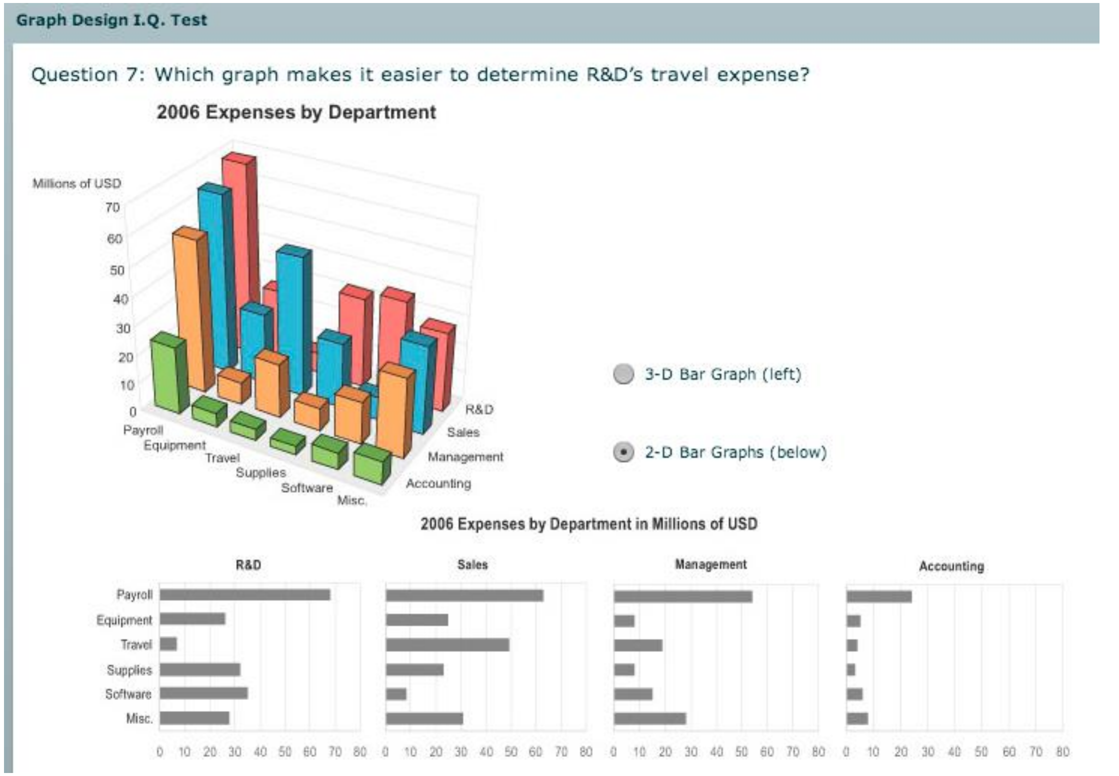


[Visualizing the Results of Multimedia Web Search Engines. Mukherjea, Hirata, and Hara. InfoVis 96]

3D vs 2D bar charts

3D bars:
very difficult to justify!
– perspective distortion
– occlusion

faceting into 2D almost always better choice



[http://perceptualedge.com/files/GraphDesignIQ.html]

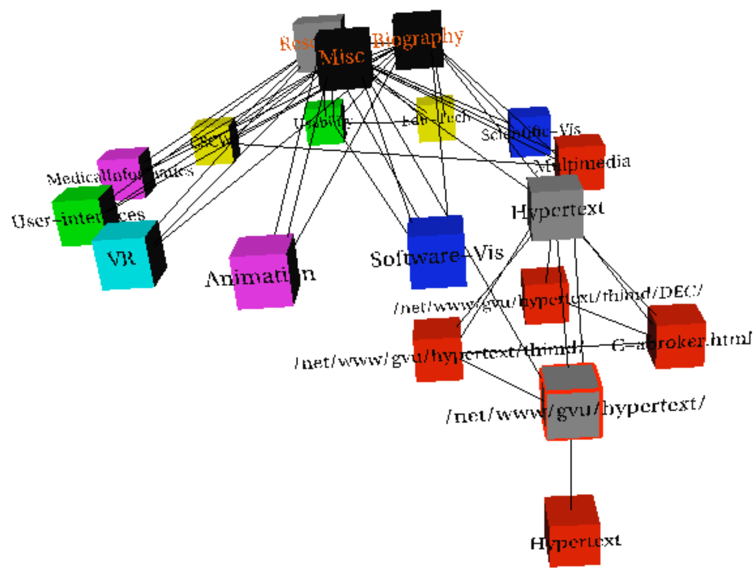
Tilted text isn't legible

text legibility

– far worse when tilted from image plane

further reading

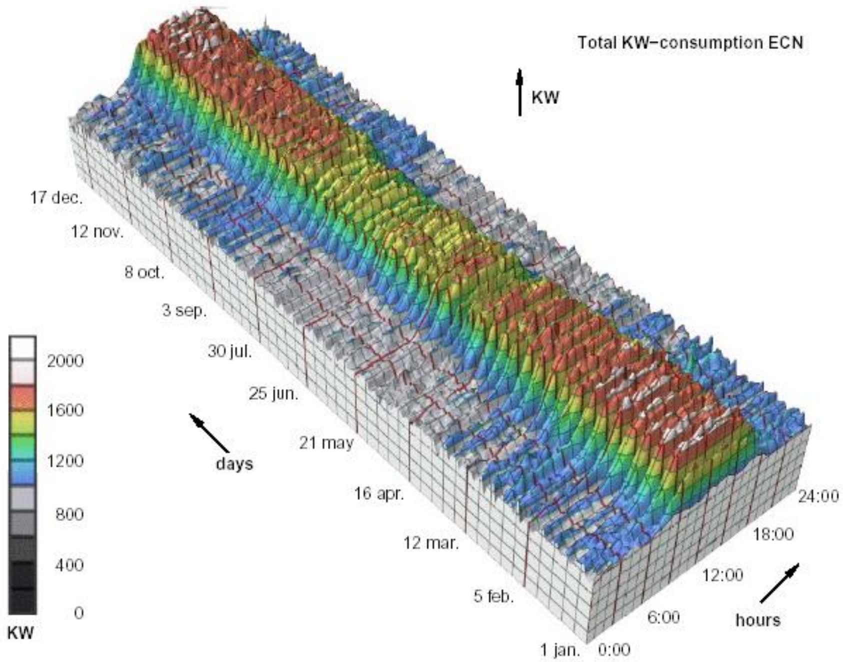
Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays. Grossman et al. CHI 2007



[*Visualizing the World-Wide Web with the Navigational View Builder. Mukherjea and Foley. Computer Networks and ISDN Systems, 1995.*]

No unjustified 3D example: Time-series data

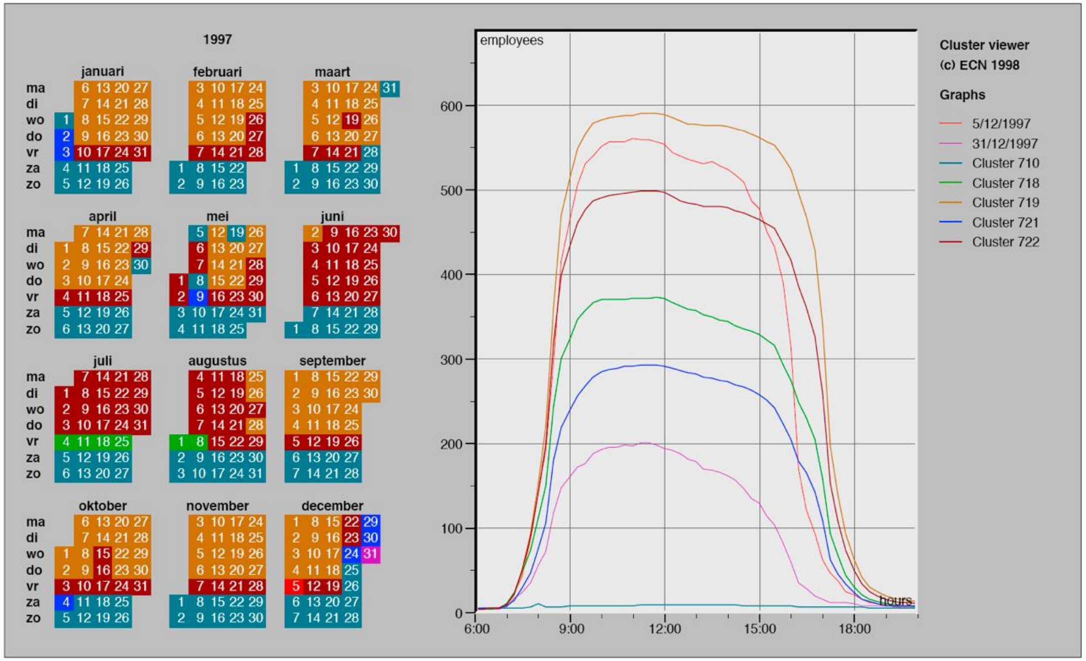
extruded curves: detailed comparisons impossible



No unjustified 3D example: Transform for new data abstraction

derived data: cluster hierarchy

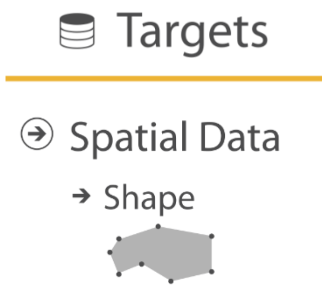
juxtapose multiple views: calendar, superimposed 2D curves

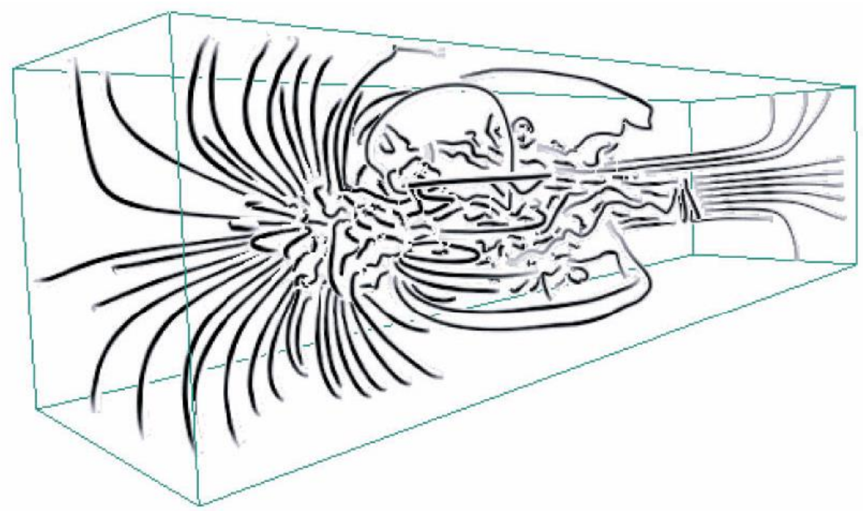


Justified 3D: shape perception

benefits outweigh costs when task is shape perception for 3D spatial data

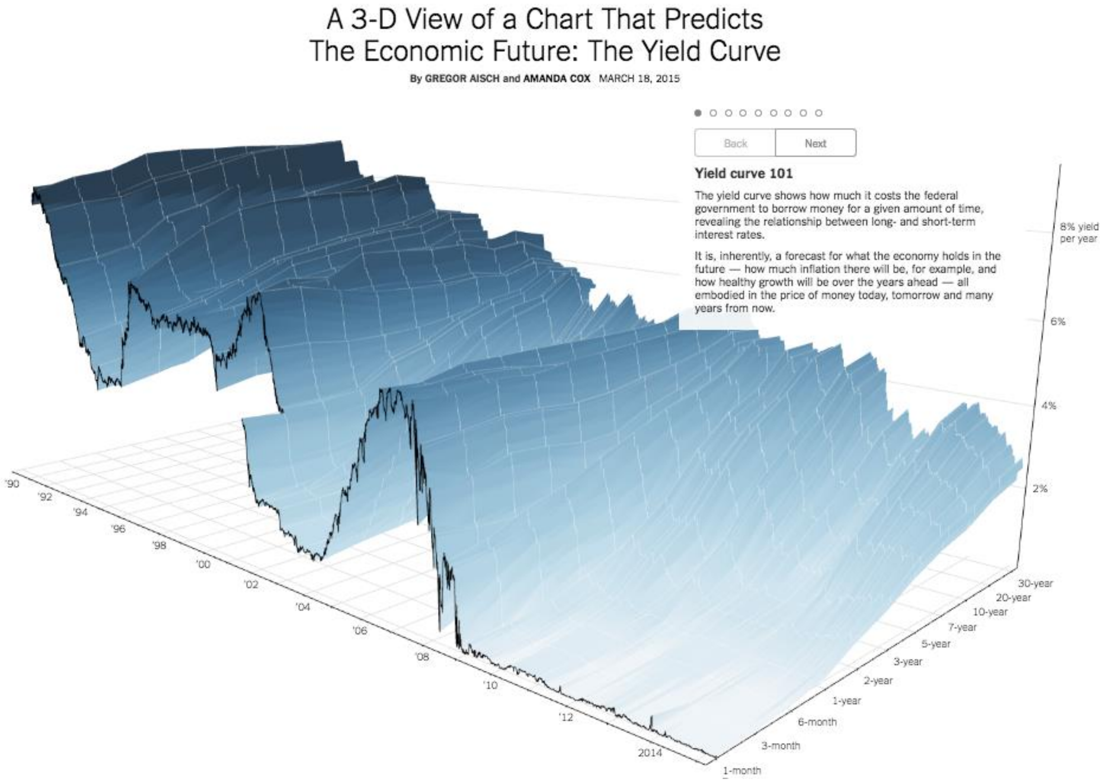
- interactive navigation supports synthesis across many viewpoints





Justified 3D: Economic growth curve

constrained navigation
- steps through carefully designed viewpoints



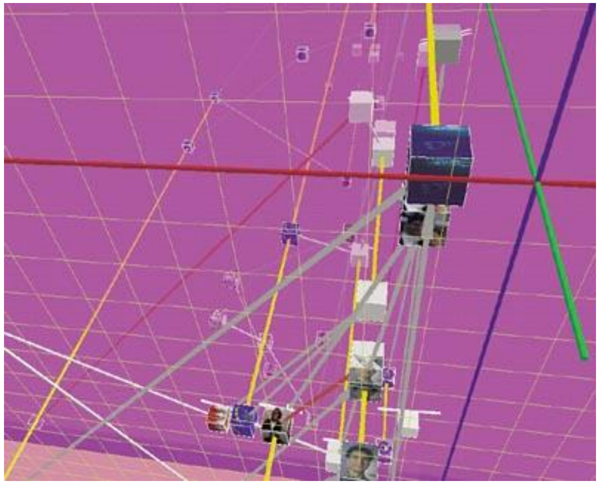
http://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html

No unjustified 3D

3D legitimate for true 3D spatial data

3D needs very careful justification for abstract data

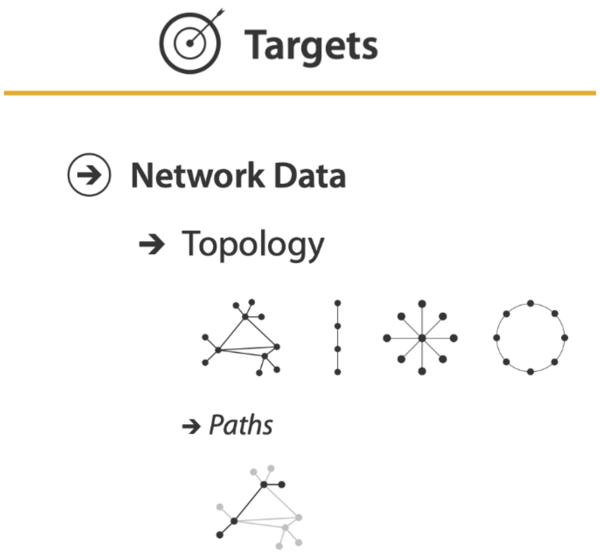
- enthusiasm in 1990s, but now skepticism
- be especially careful with 3D for point clouds or networks



No unjustified 2D

consider whether network data requires 2D spatial layout
- especially if reading text is central to task!
- arranging as network means lower information density and harder label lookup compared to text lists

benefits outweigh costs when topological structure/context important for task
- be especially careful for search results, document collections, ontologies



Eyes beat memory

principle: external cognition vs. internal memory

– easy to compare by moving eyes between side-by-side views

– harder to compare visible item to memory of what you saw

implications for animation

– great for choreographed storytelling

– great for transitions between two states

– poor for many states with changes everywhere

consider small multiples instead

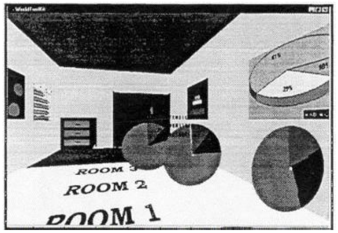
literal | abstract

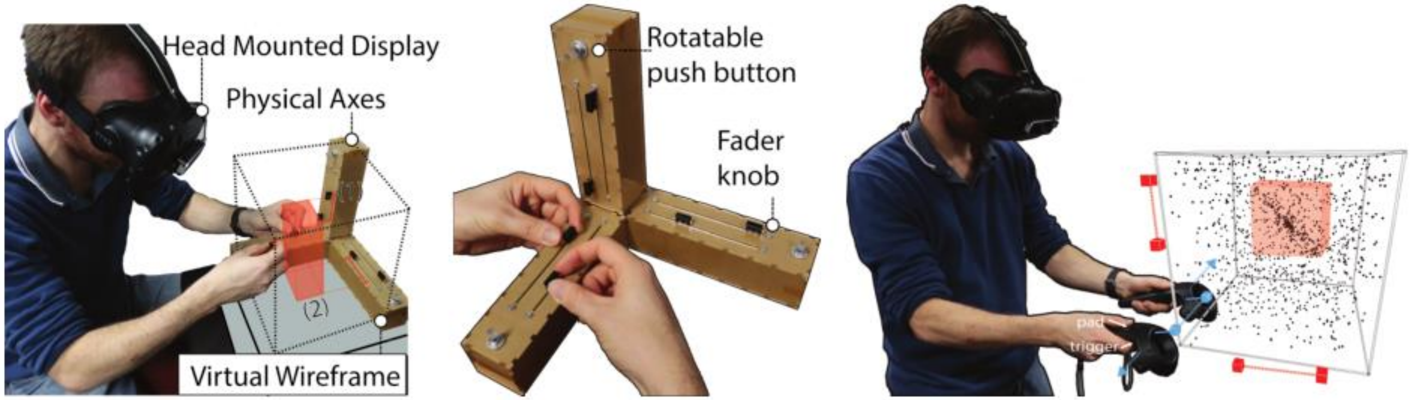
animation | small multiples

show time with time | show time with space

Resolution beats immersion

- immersion typically not helpful for abstract data
  - do not need sense of presence or stereoscopic 3D
  - desktop also better for workflow integration





data difficult to justify

second wave: AR/MR

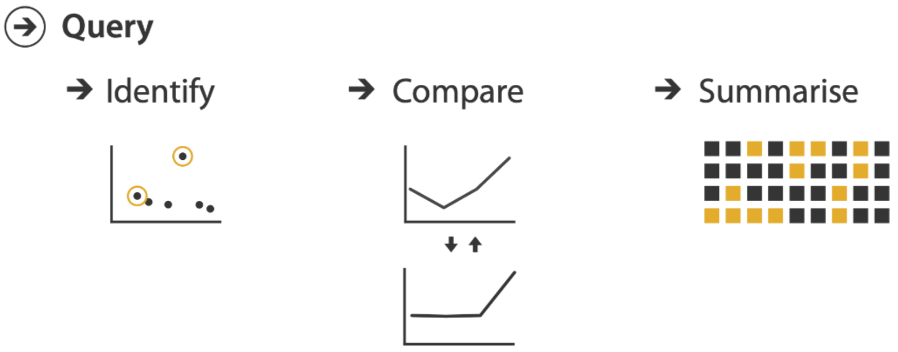
Overview first, zoom and filter, details on demand

influential mantra from Shneiderman

[The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Shneiderman. Proc. IEEE Visual Languages, pp. 336–343, 1996.]

overview = summary

- microcosm of full vis design problem



Rule of thumb: **Responsiveness is required**

visual feedback: three rough categories

– *0.1 seconds: perceptual processing*

subsecond response for mouseover highlighting - ballistic motion

– *1 second: immediate response*

fast response after mouseclick, button press - Fitts' Law limits on motor control

– *10 seconds: brief tasks*

bounded response after dialog box - mental model of heavyweight operation (file load)

scalability considerations

– highlight selection without complete redraw of view (graphics frontbuffer)
– show hourglass for multi-second operations (check for cancel/undo)
– show progress bar for long operations (process in background thread)
– rendering speed when item count is large (guaranteed frame rate)

Function first, form next

dangerous to start with aesthetics

- usually impossible to add function retroactively

start with focus on functionality

- possible to improve aesthetics later on, as refinement
- if no expertise in-house, find good graphic designer to work with
- aesthetics do matter! another level of function
  - visual hierarchy, alignment, flow
  - Gestalt principles in action

Form: Basic graphic design ideas

proximity

- do group related items together
- avoid equal whitespace between unrelated

alignment

- do find/make strong line, stick to it
- avoid automatic centering

repetition

- do unify by pushing existing consistencies

contrast

- if not identical, then very different
- avoid not quite the same

The Non-Designer's Design Book, 4th ed. Robin Williams, Peachpit Press, 2015.

- fast read, very practical to work through whole thing

Best practices: Labelling

make visualizations as self-documenting as possible

- meaningful & useful title, labels, legends

  axes and panes/subwindows should have labels

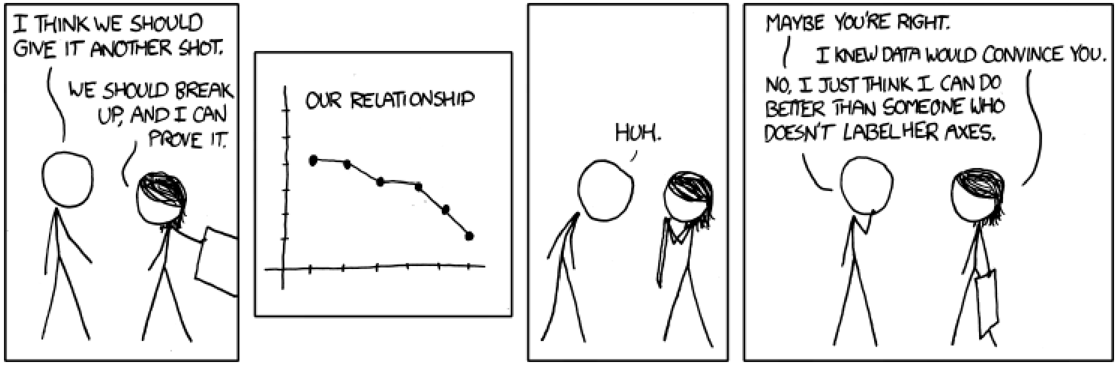
  - and axes should have good mix/max boundary tick marks

  everything that’s plotted should have a legend

  - and own header/labels if not redundant with main title

  use reasonable numerical format

  - avoid scientific notation in most cases



https://xkcd.com/833/

Rules of Thumb Summary

No unjustified 3D
- Power of the plane
- Disparity of depth
- Occlusion hides information
- Perspective distortion dangers
- Tilted text isn't legible

No unjustified 2D

Eyes beat memory

Resolution over immersion

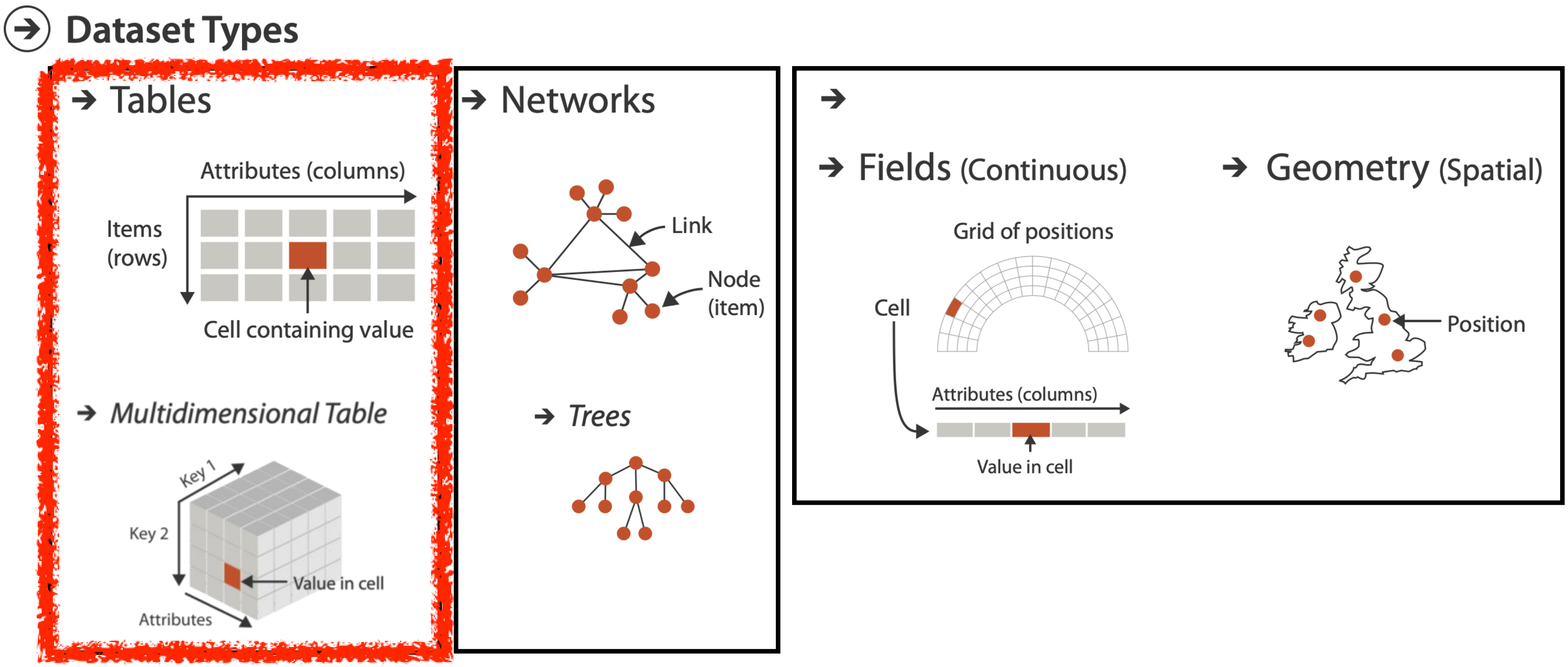
Overview first, zoom and filter, details on demand

Responsiveness is required

Function first, form next

Arranging Tables

Focus on Tables



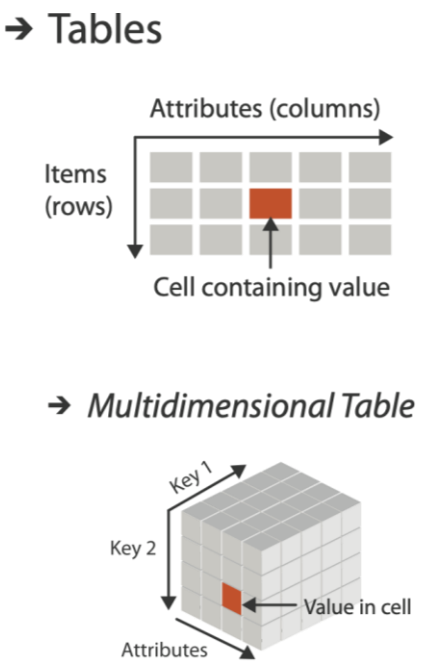
Keys and values

key

- independent attribute
- used as unique index to look up items
- simple tables: 1 key
- multidimensional tables: multiple keys

value

- dependent attribute, value of cell



Keys and values

key

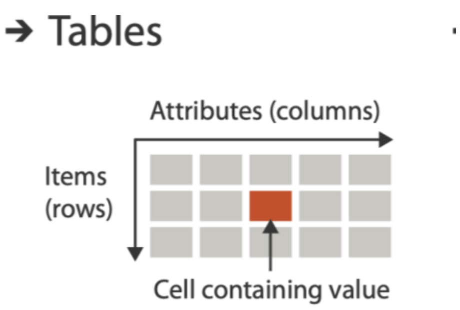
- independent attribute
- used as unique index to look up items
- simple tables: 1 key
- multidimensional tables: multiple keys

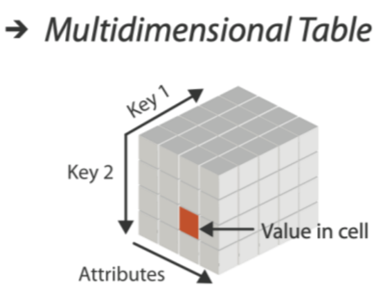
value

- dependent attribute, value of cell

classify arrangements by keys used

- 0, 1, 2, ...



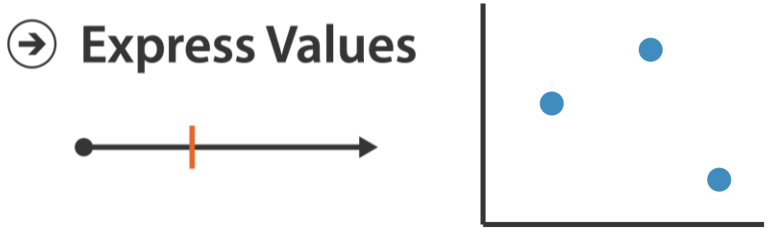


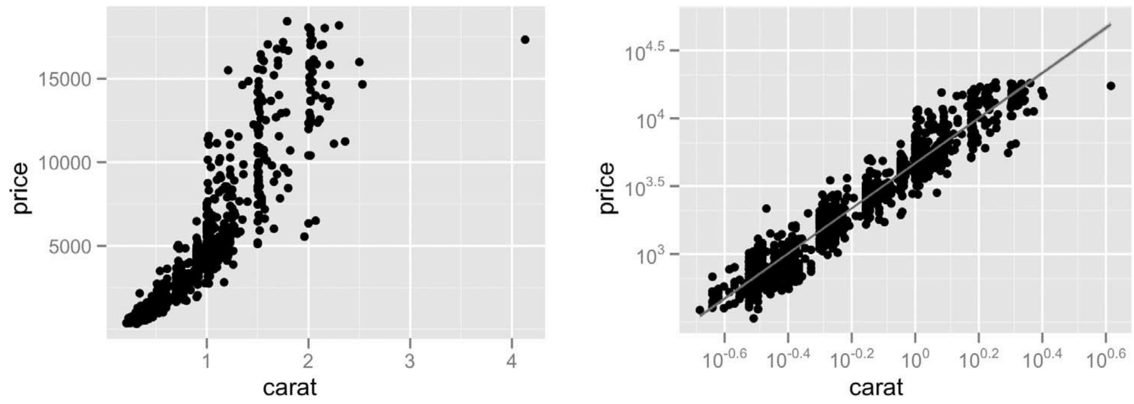
Idiom: **scatterplot**

express values (magnitudes)

- quantitative attributes

no keys, only values





[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

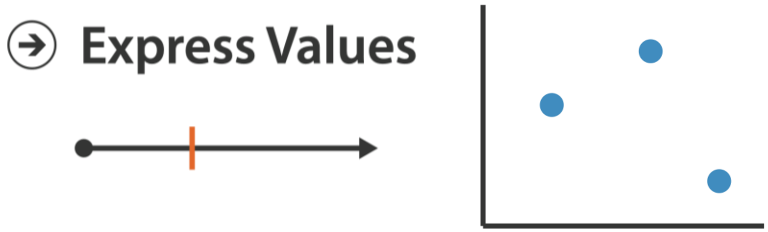
Idiom: **scatterplot**

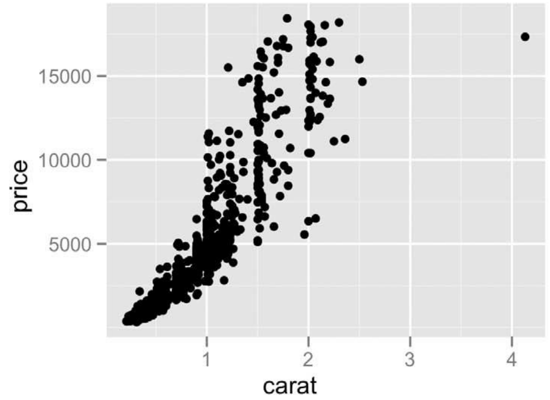
express values (magnitudes)

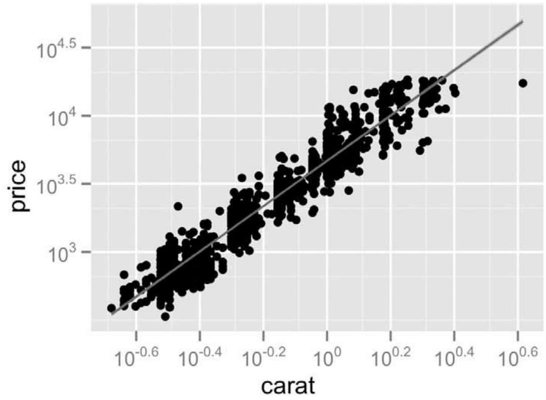
- quantitative attributes

no keys, only values

- data
  - 2 quant attribs
- mark: points
- channels
  - horiz + vert position







[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

Idiom: **scatterplot**

express values (magnitudes)

- quantitative attributes

no keys, only values

- data

  2 quant attribs

- mark: points
- channels

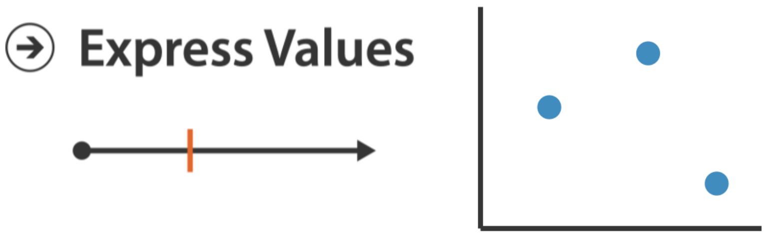
  horiz + vert position

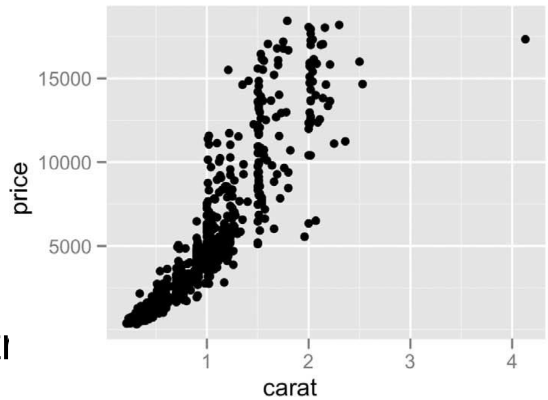
- tasks

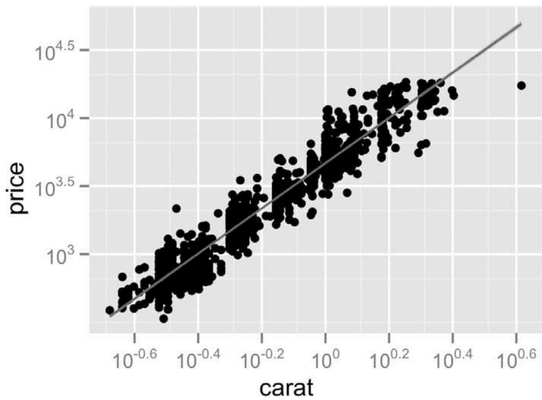
  find trends, outliers, distr

- scalability

  hundreds of items





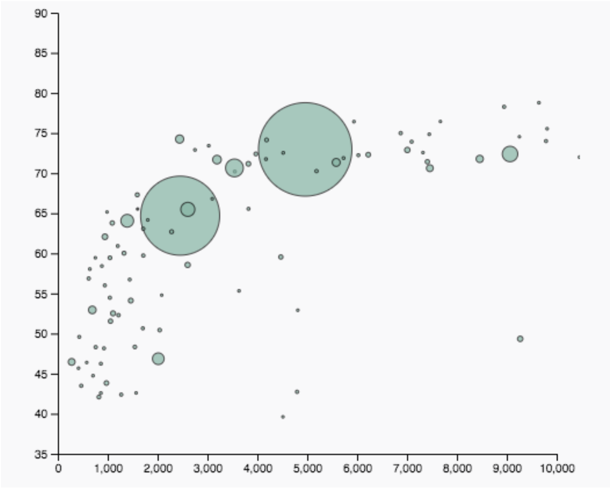


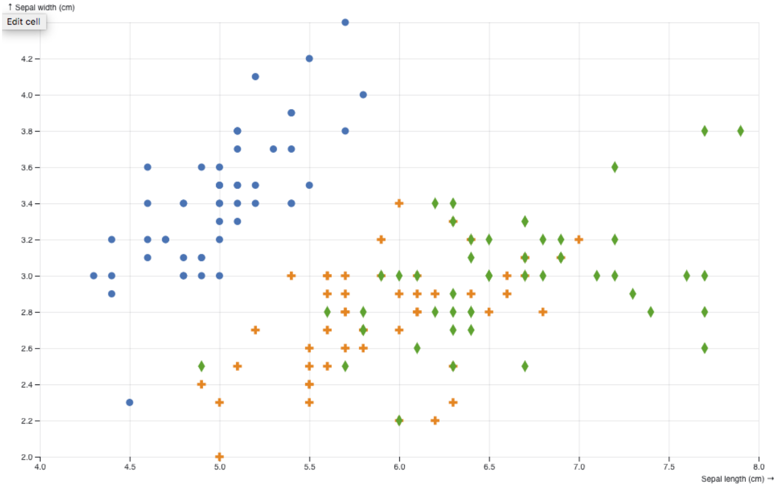
[A layered grammar of graphics. Wickham. Journ. Computational and Graphical Statistics 19:1 (2010), 3–28.]

Scatterplots: Encoding more channels

additional channels viable since using point marks

- color
- size (1 quant attribute, used to control 2D area)
  - note radius would mislead, take square root since area grows quadratically
- shape



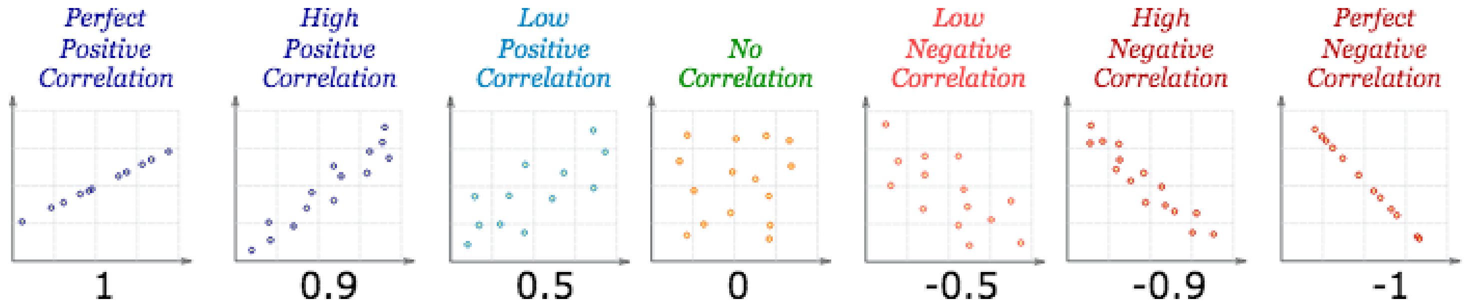


https://www.d3-graph-gallery.com/graph/bubble_basic.html

https://observablehq.com/@d3/scatterplot-with-shapes

Scatterplot tasks

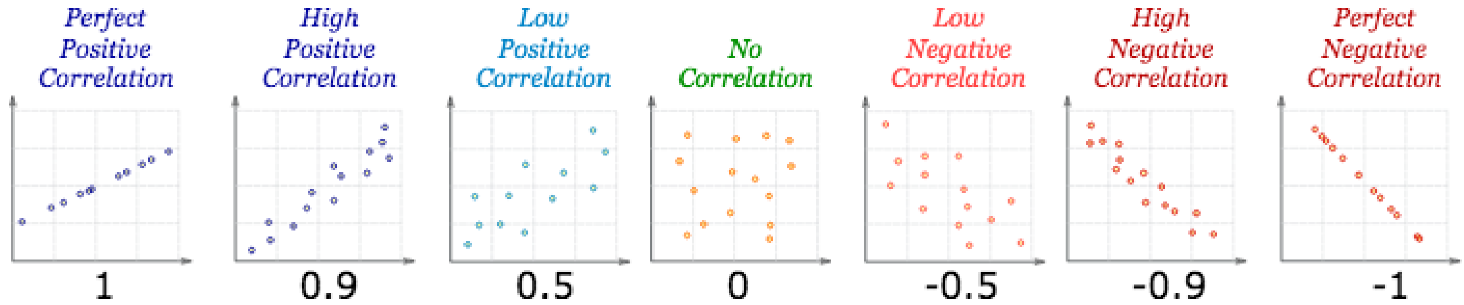
correlation



https://www.mathsisfun.com/data/scatter-xy-plots.html

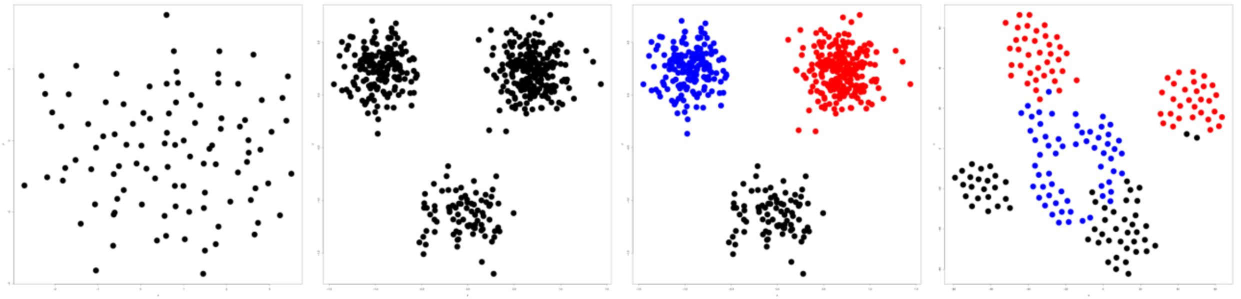
Scatterplot tasks

correlation



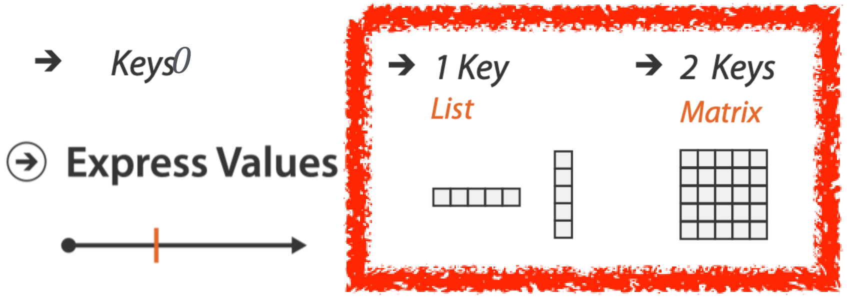
https://www.mathsisfun.com/data/scatter-xy-plots.html

clusters/groups, and clusters vs classes



https://www.cs.ubc.ca/labs/imager/tr/2014/DRVisTasks/

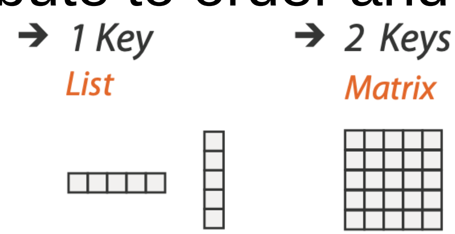
Some keys



Some keys: Categorical regions

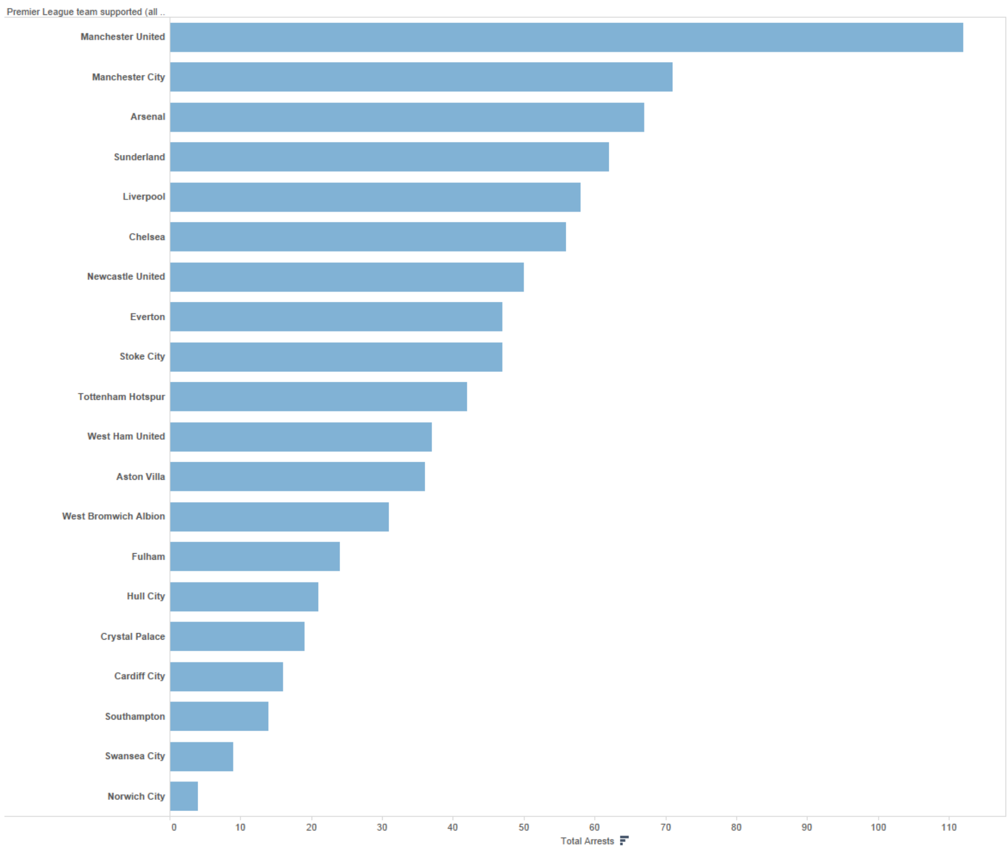
Regions: Separate, order, align

- regions: contiguous bounded areas distinct from each other
  - separate into spatial regions: one mark per region (for now)
- use categorical or ordered attribute to separate into regions
  - no conflict with expressiveness principle for categorical attributes
- use ordered attribute to order and align regions



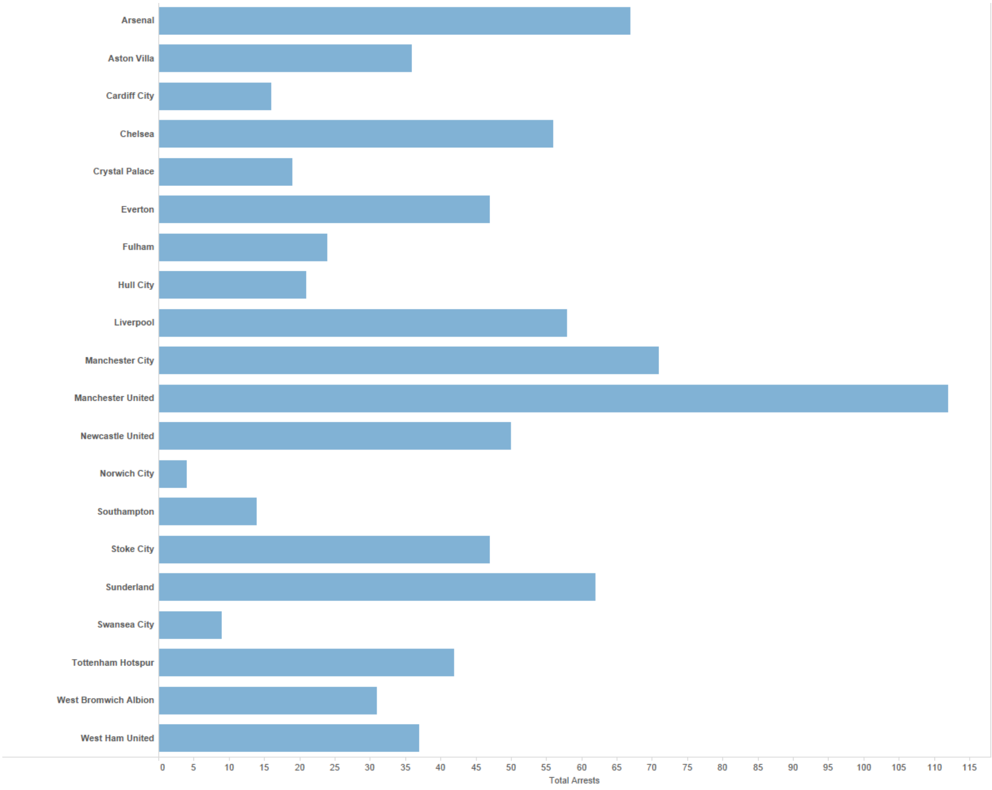
Separated and aligned and ordered

best case



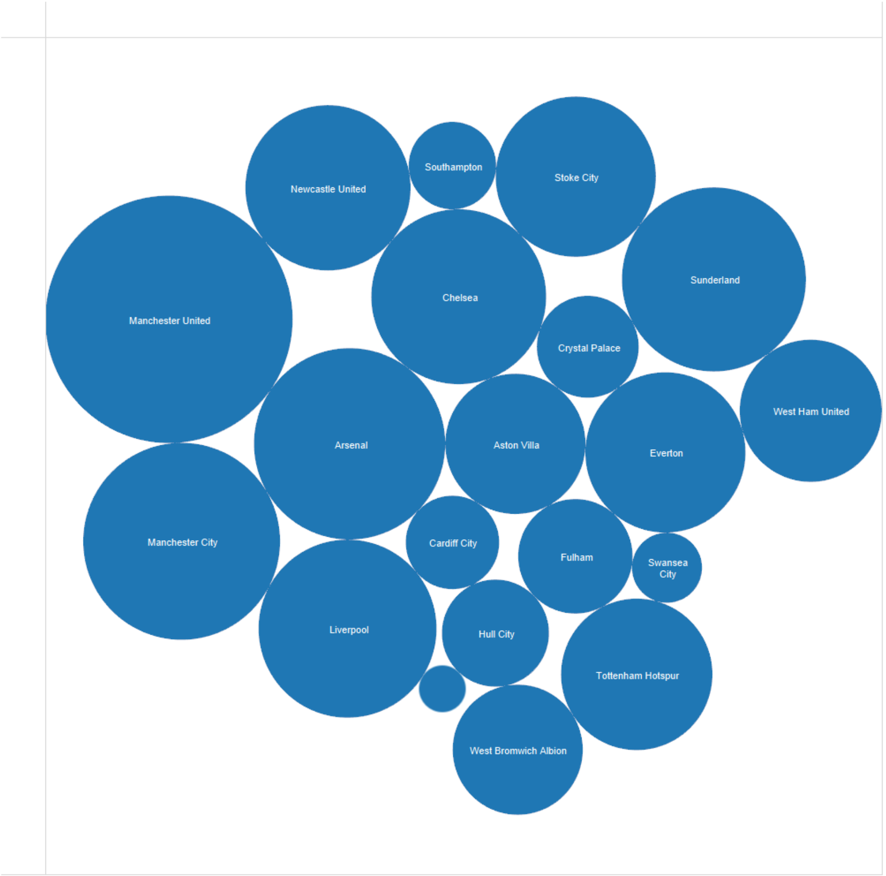
Separated and aligned but not ordered

limitation: hard to know rank. what's 4th? what's 7th?



Separated but not aligned or ordered

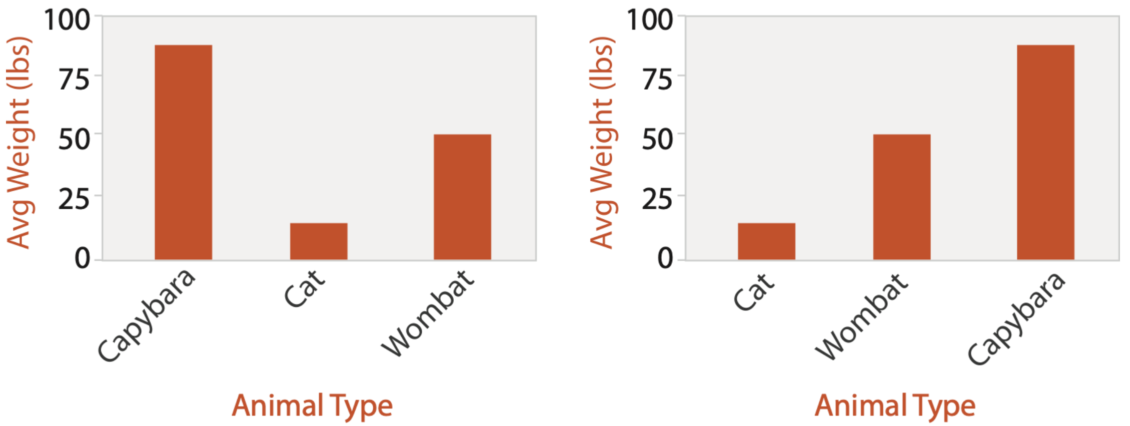
limitation: hard to make comparisons with size (vs aligned position)



Idiom: **bar chart**

one key, one value

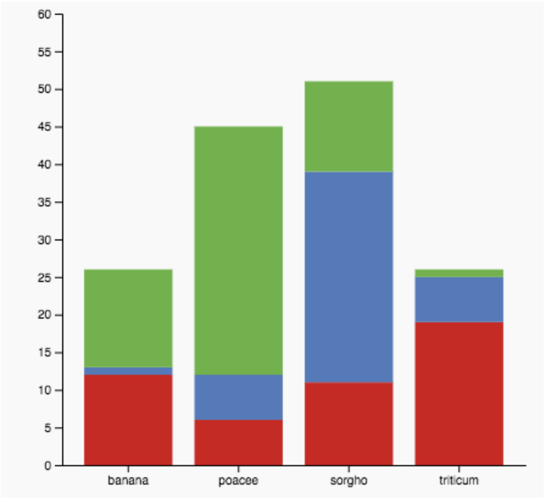
- data
  - 1 categ attrib, 1 quant attrib
- mark: lines
- channels
  - length to express quant value
  - spatial regions: one per mark
    - separated horizontally, aligned vertically
    - ordered by quant attrib
      - » by label (alphabetical), by length attrib (data-driven)
- task
  - compare, lookup values
- scalability
  - dozens to hundreds of levels for key attrib [bars], hundreds for values



Idiom: **stacked bar chart**

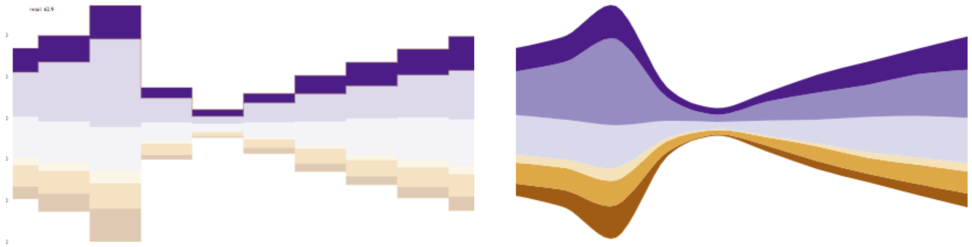
one more key

- data
  - 2 categ attrib, 1 quant attrib
- mark: vertical stack of line marks
  - **glyph**: composite object, internal structure from multiple marks
- channels
  - length and color hue
  - spatial regions: one per glyph
    - aligned: full glyph, lowest bar component
    - unaligned: other bar components
- task
  - part-to-whole relationship
- scalability: asymmetric
  - for *stacked* key attrib, 10-12 levels [segments]
  - for *main* key attrib, dozens to hundreds of levels [bars]



https://www.d3-graph-gallery.com/graph/barplot_stacked_basicWide.html

Idiom: **streamgraph**



[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]

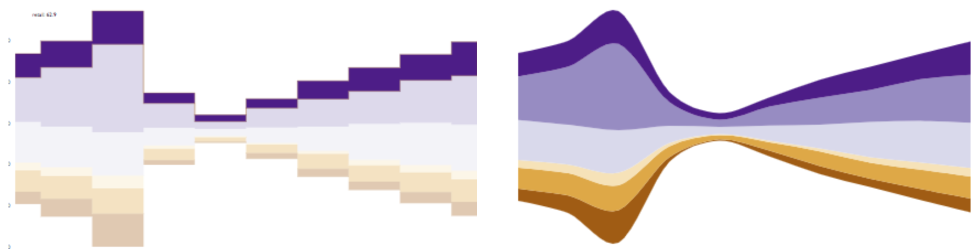
generalized stacked graph

- emphasizing horizontal continuity
  - vs vertical items
- data
  - 1 categ key attrib (movies)
  - 1 ordered key attrib (time)
  - 1 quant value attrib (counts)
- derived data
  - geometry: layers, where height encodes counts
  - 1 quant attrib (layer ordering)

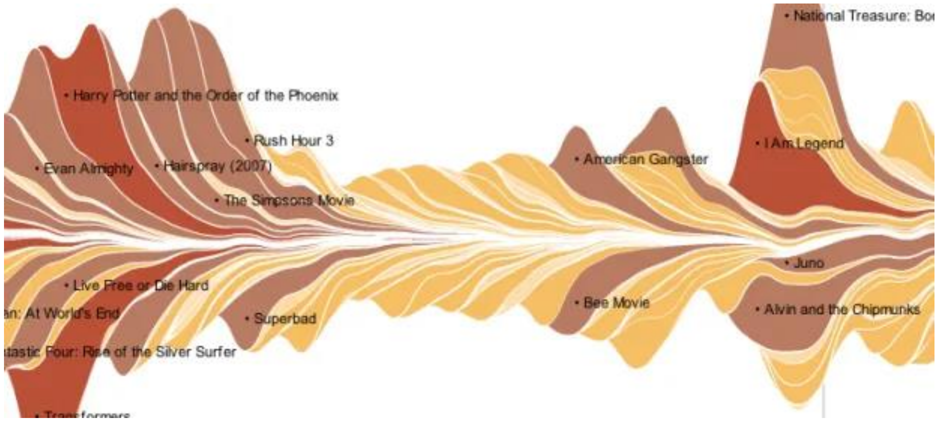
Idiom: **streamgraph**

generalized stacked graph

- emphasizing horizontal continuity
  - vs vertical items
- data
  - 1 categ key attrib (movies)
  - 1 ordered key attrib (time)
  - 1 quant value attrib (counts)
- derived data
  - geometry: layers, where height encodes counts
  - 1 quant attrib (layer ordering)
- scalability
  - hundreds of time keys
  - dozens to hundreds of movies keys
    - more than stacked bars: most layers don't extend across whole chart



[Stacked Graphs Geometry & Aesthetics. Byron and Wattenberg. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2008) 14(6): 1245–1252, (2008).]

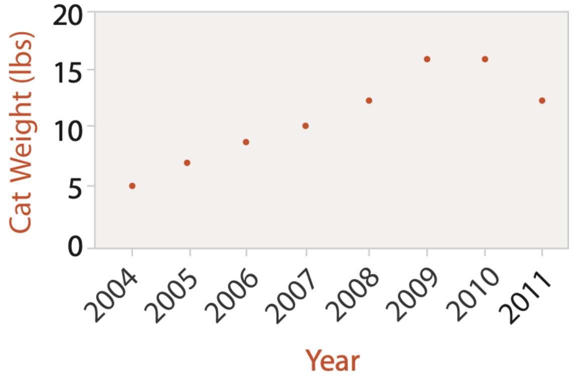


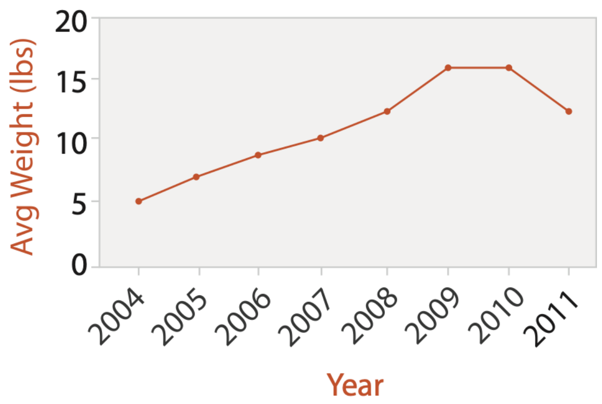
https://flowingdata.com/2008/02/25/ebb-and-flow-of-box-office-receipts-over-past-20-years/

Idiom: **dot / line chart**

one key, one value

- data
  - 2 quant attribs
- mark: points
  AND line connection marks between them
- channels
  - aligned lengths to express quant value
  - separated and ordered by key attrib into horizontal regions
- task

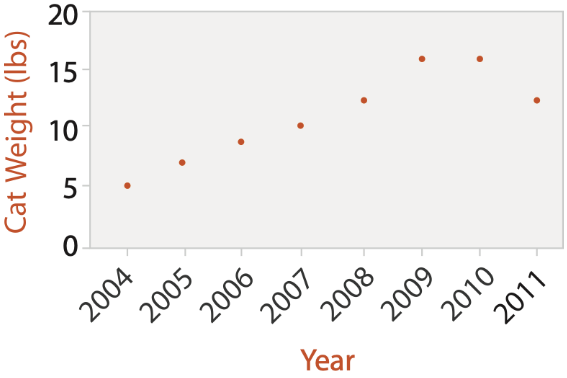


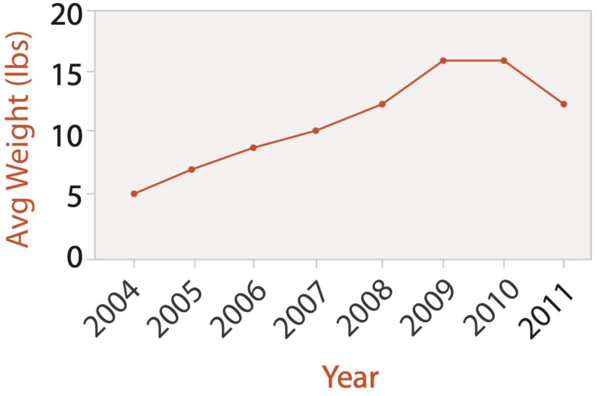


Idiom: **dot / line chart**

one key, one value

- data
  - 2 quant attribs
- mark: points
  AND line connection marks between them
- channels
  - aligned lengths to express quant value
  - separated and ordered by key attrib into horizontal regions
- task
  - find trend
    - connection marks emphasize ordering of items along key axis by explicitly showing relationship between one item and the next
- scalability
  - hundreds of key levels, hundreds of value levels





Choosing bar vs line charts

- depends on type of key attrib
  - bar charts if categorical
  - line charts if ordered
- do not use line charts for categorical key attribs
  - violates expressiveness principle
    - implication of trend so strong that it overrides semantics!
      - “The more male a person is, the taller he/she is”

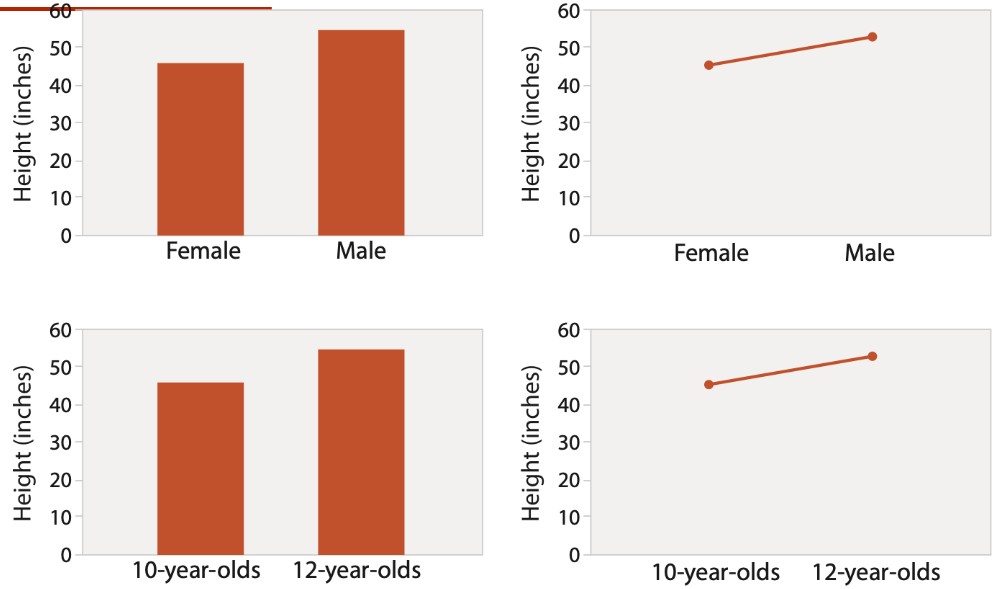
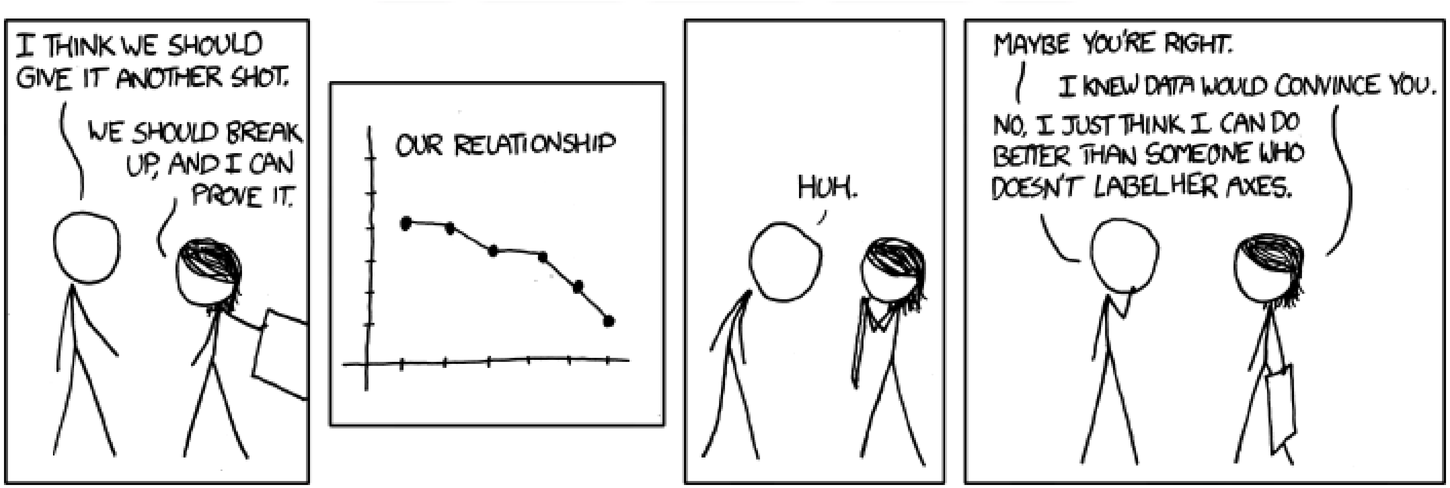


Chart axes: label them!

best practice to label

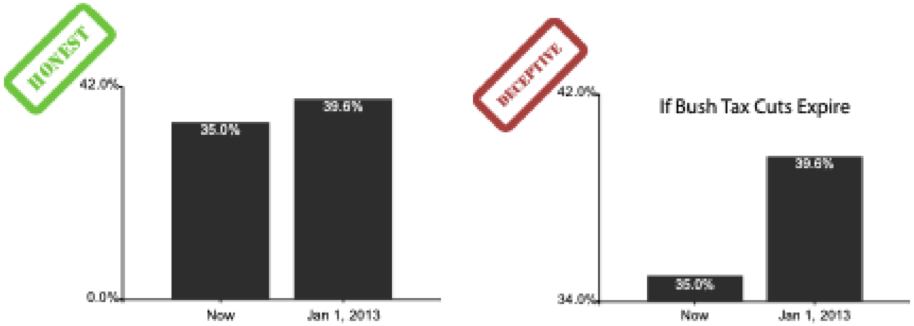
- few exceptions: individual small multiple views could share axis label



https://xkcd.com/833/

Chart axes: avoid cropping y axis

include 0 at bottom left or slope misleads

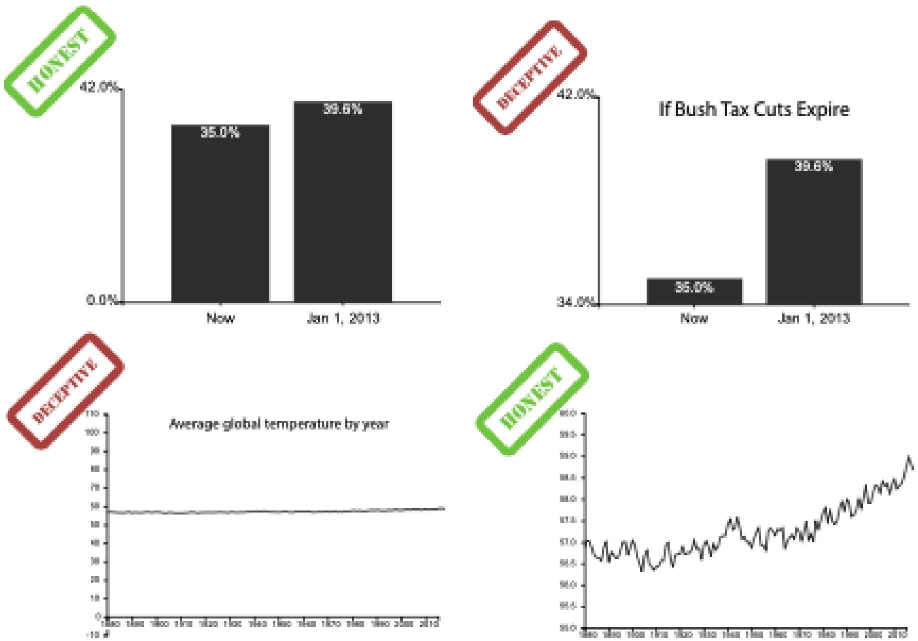


[Truncating the Y-Axis: Threat or Menace? Correll, Bertini, & Franconeri, CHI 2020.]

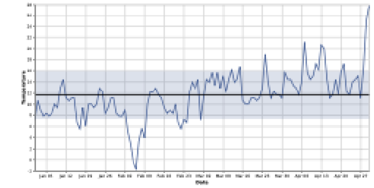
Chart axes: avoid cropping y axis

include 0 at bottom left or slope misleads

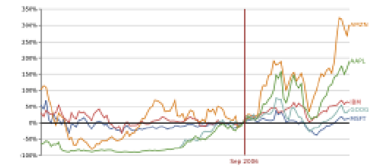
– some exceptions (arbitrary 0, small change matters)



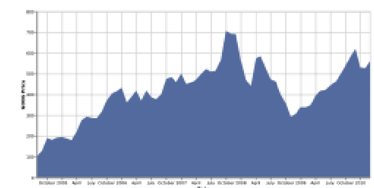
[Truncating the Y-Axis: Threat or Menace? Correll, Bertini, & Franconeri, CHI 2020.]



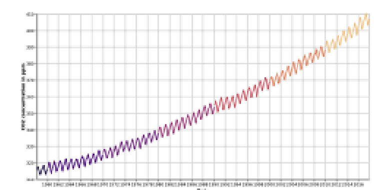
(a) Statistical process charts rely on comparison to an expected value, and so deviations from that value, not from zero, are important



(b) Index charts compare to an indexed value rather than zero.



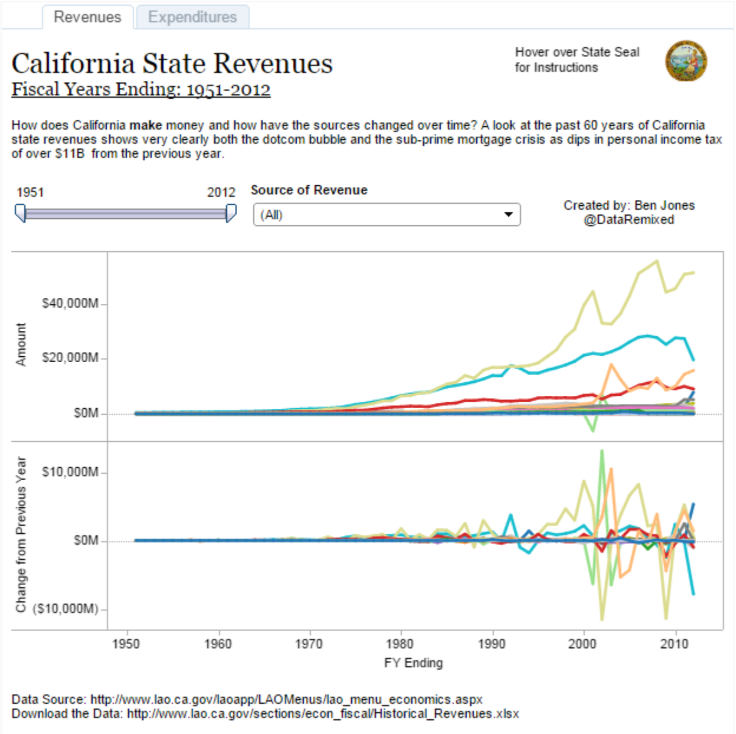
(c) Stock charts must show small differences in stock value, as these can translate to enormous monetary gains or losses.



(d) Climate Anomaly charts rely on both highlighting deviation from a non-zero expected value but also emphasize the potentially disastrous impact of even minute changes in climate.

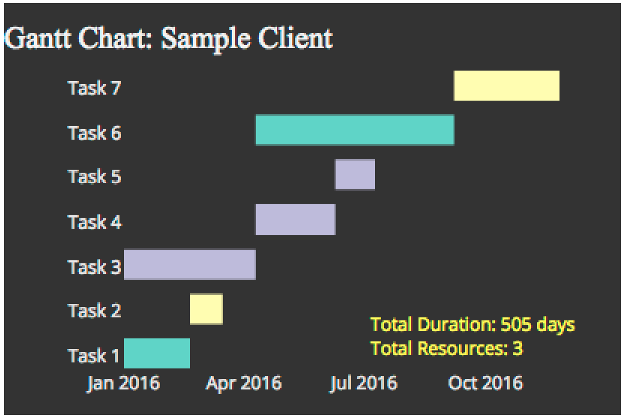
Idiom: **Indexed line charts**

- data: 2 quant attribs
  - 1 key + 1 value
- derived data: new quant value attrib
  - index
  - plot instead of original value
- task: show change over time
  - principle: normalized, not absolute
- scalability
  - same as standard line chart



https://public.tableau.com/profile/ben.jones#!/vizhome/CAStateRevenues/Revenues

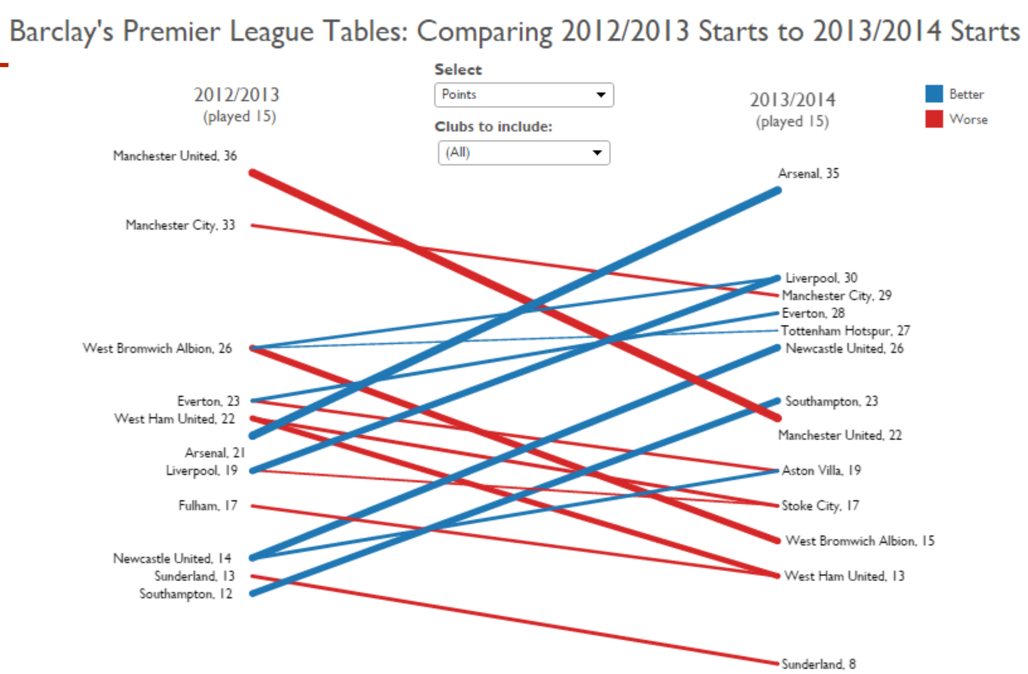
Idiom: **Gantt charts**



one key, two (related) values

- data
  - 1 categ attrib, 2 quant attribs
- mark: line
  - length: duration
- channels
  - horiz position: start time (+end from duration)
- task
  - emphasize temporal overlaps & start/end dependencies between items
- scalability
  - dozens of key levels [bars]
  - hundreds of value levels [durations]

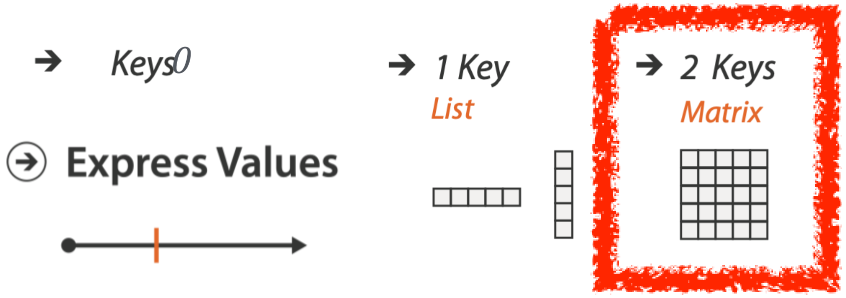
Idiom: **Slopegraphs**



https://public.tableau.com/profile/ben.jones#!/vizhome/Slopegraphs/Slopegraphs

- two values
  - data
    - 2 quant value attribs
    - (1 derived attrib: change magnitude)
  - mark: point + line
    - line connecting mark between pts
  - channels
    - 2 vertical pos: express attrib value
    - (linewidth/size, color)
  - task
    - emphasize changes in rank/value
  - scalability
    - hundreds of value levels
    - dozens of items

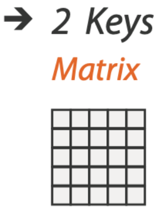
2 Keys

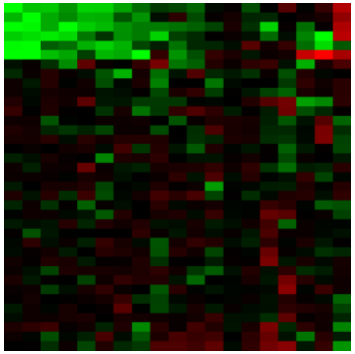


Idiom: **heatmap**

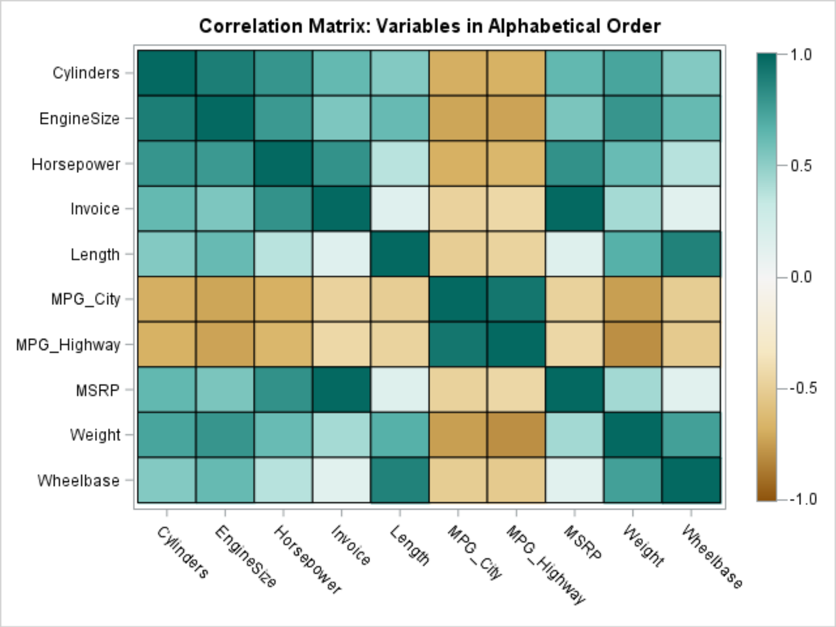
two keys, one value

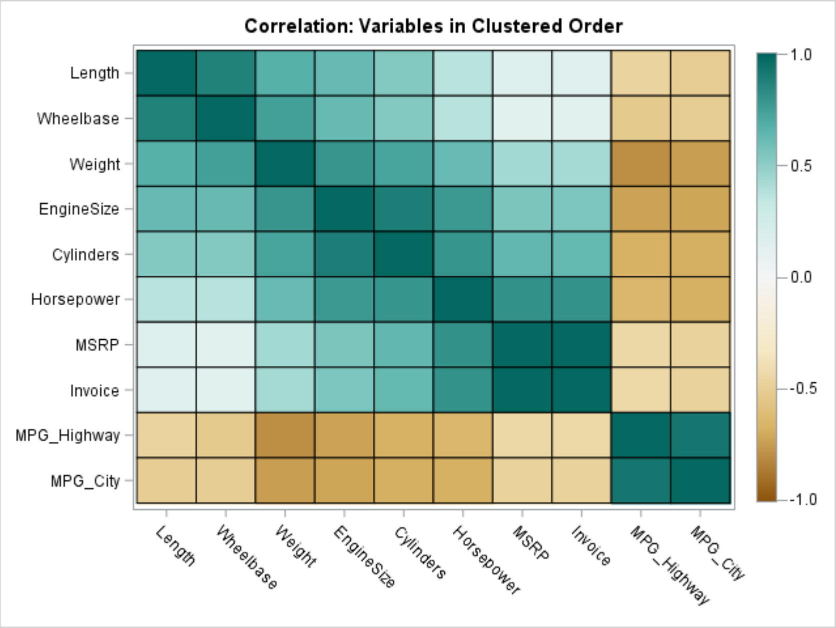
- data
  - 2 categ attribs (gene, experimental condition)
  - 1 quant attrib (expression levels)
- marks: point
  - separate and align in 2D matrix
    - indexed by 2 categorical attributes
- channels
  - color by quant attrib
    - (ordered diverging colormap)
- task
  - find clusters, outliers
- scalability
  - 1M items, 100s of categ levels, ~10 quant attrib levels





Heatmap reordering



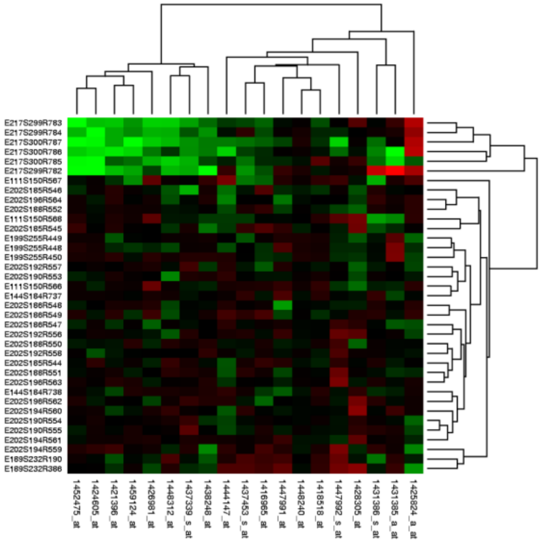


https://blogs.sas.com/content/iml/2018/05/02/reorder-variables-correlation-heat-map.html

Idiom: **cluster heatmap**

in addition

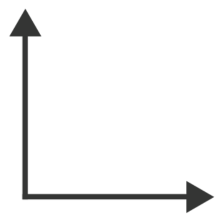
- derived data
  - 2 cluster hierarchies
- dendrogram
  - parent-child relationships in tree with connection line marks
  - leaves aligned so interior branch heights easy to compare
- heatmap
  - marks (re-)ordered by cluster hierarchy traversal
  - task: assess quality of clusters found by automatic methods



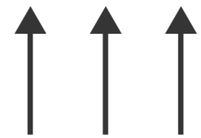
Tables

Axis Orientation

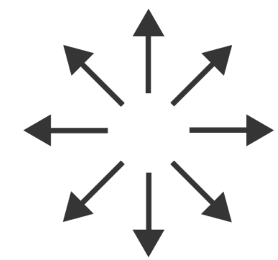
Rectilinear



Parallel



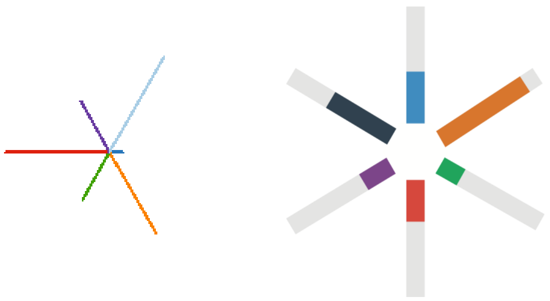
Radial



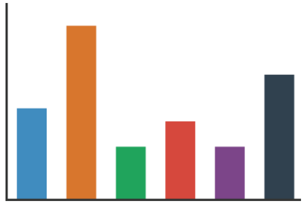
Idioms: **radial bar chart, star plot**

star plot
- line mark, radial axes meet at central point

radial bar chart
- line mark, radial axes meet at central ring
- channels: length, angle/orientation



bar chart
- rectilinear axes, aligned vertically



accuracy
- length not aligned with radial layouts

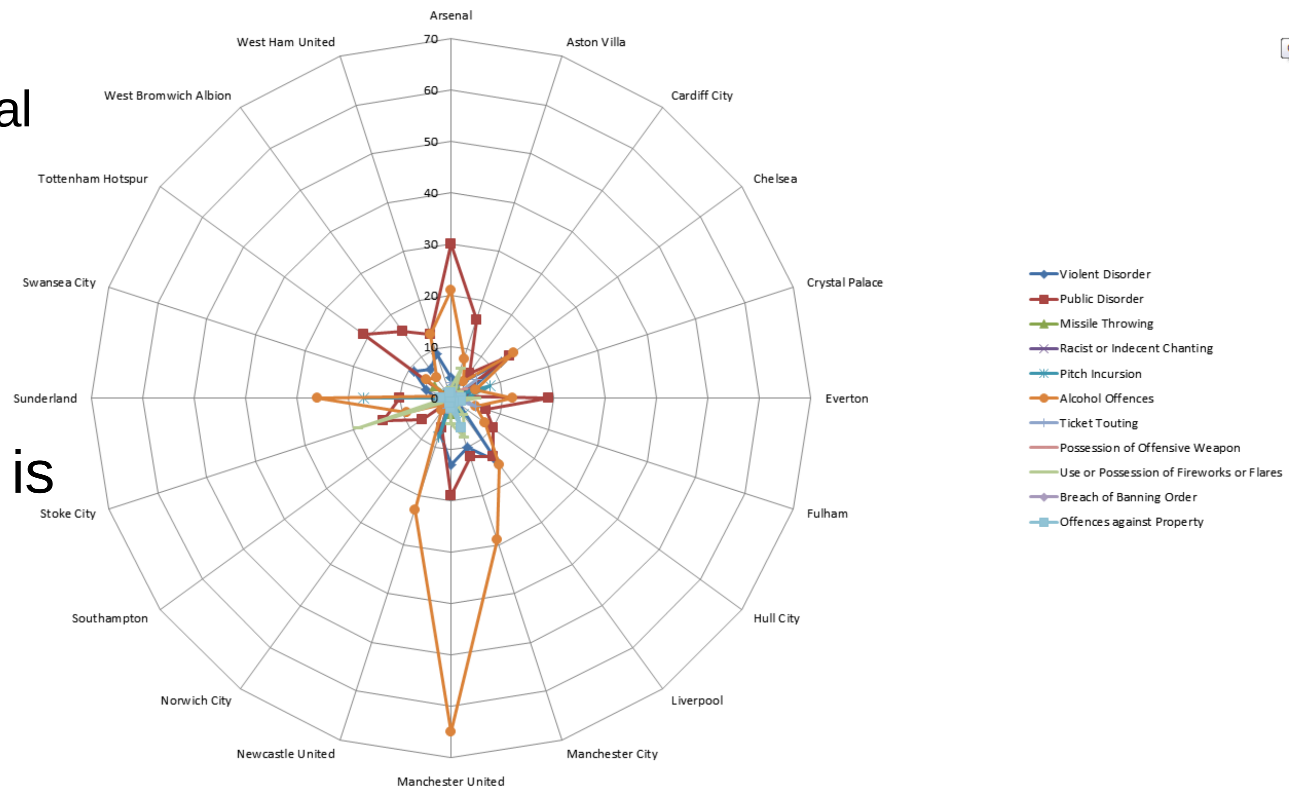
  less accurately perceived than rectilinear aligned

Idiom: **radar plot**

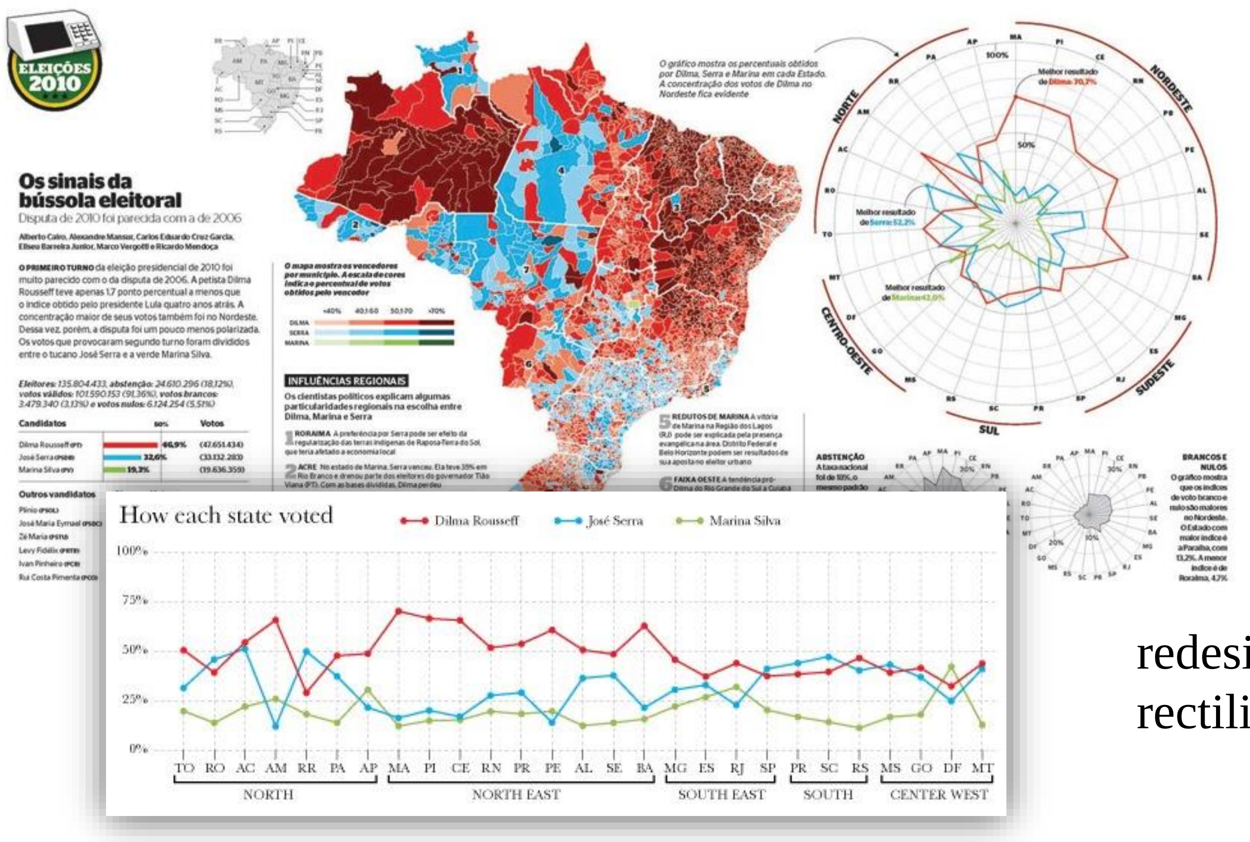
radial line chart

- point marks, radial layout
- connecting line marks

avoid unless data is cyclic



“Radar graphs: Avoid them (99.9% of the time)”

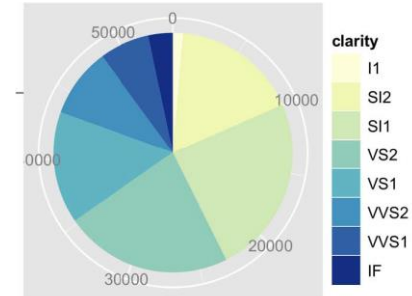


original difficult to interpret

redesign for rectilinear

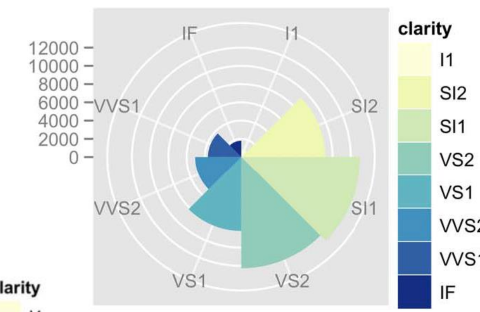
http://www.thefunctionalart.com/2012/11/radar-graphs-avoid-them-999-of-time.html

Idioms: **pie chart, coxcomb chart**



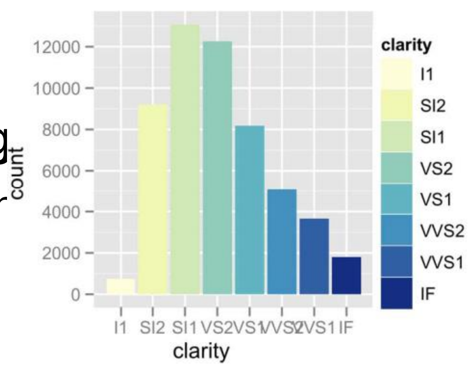
pie chart

- **interlocking area** marks with angle channel: **2D area varies**
  - separated & ordered radially, uniform height
- accuracy: area less accurate than rectilinear aligned line length
- **task: part-to-whole judgements**



coxcomb chart

- line marks with length channel: **1D leng**
  - separated & ordered radially, uniform widt
- direct analog to radial bar charts



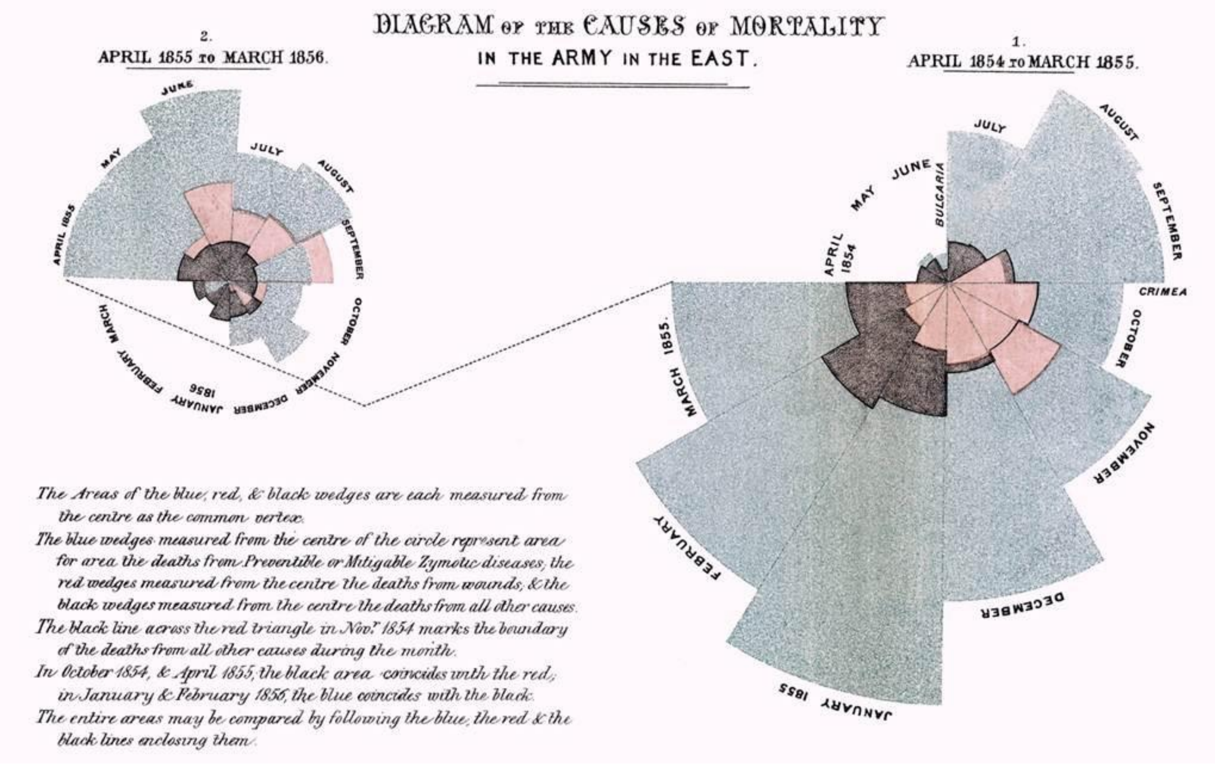
data

- 1 categ key attrib, 1 quant value attrib

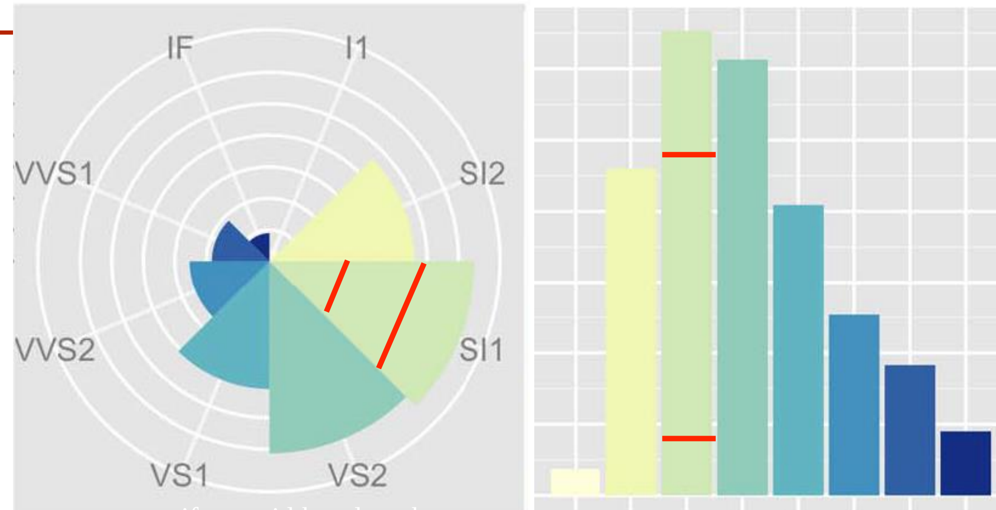
Coxcomb / nightingale rose / polar area chart

invented by Florence Nightingale:

Diagram of the Causes of Mortality in the Army in the East



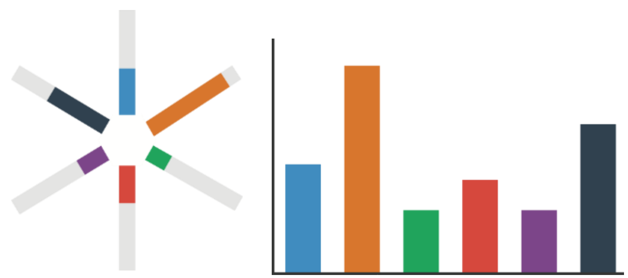
Coxcomb: perception



encode: 1D length

decode/perceive: 2D area

- nonuniform line/sector width as length increases
  - so area variation is nonlinear wrt line mark length!



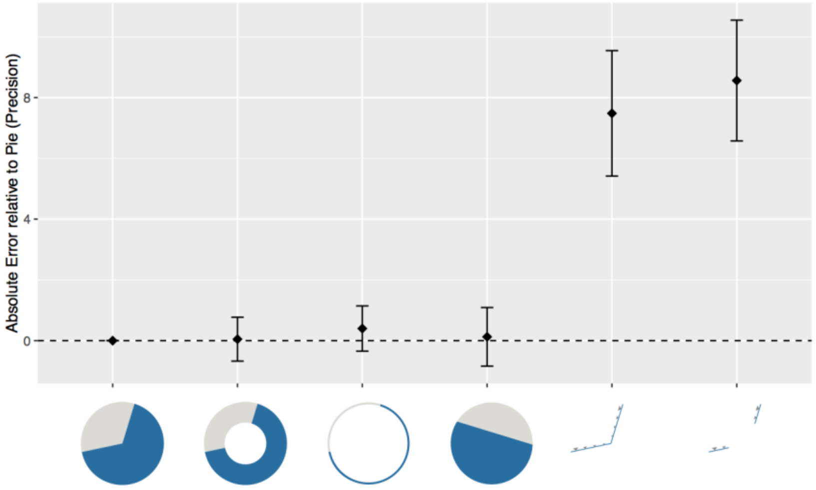
- bar chart safer: uniform width, so area is linear with line mark length
  - both radial & rectilinear cases

Pie charts: perception

some empirical evidence that people respond to arc length

- decode/perceive: not angles
- maybe also areas?…

donut charts no worse than pie charts

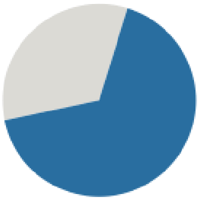


[Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts. Skau and Kosara. Proc. EuroVis 2016.]

https://eagereyes.org/blog/2016/an-illustrated-tour-of-the-pie-chart-study-results

Pie charts: best practices

not so bad for two (or few) levels, for part-to-whole task

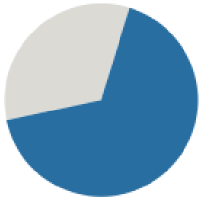


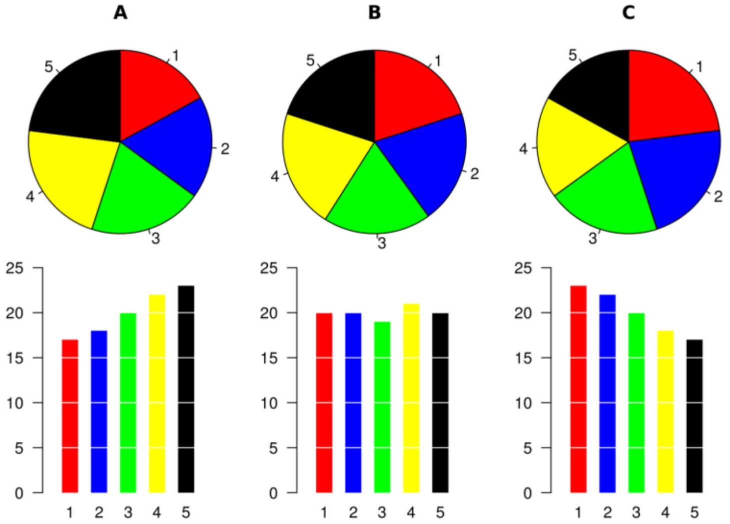
https://eagereyes.org/pie-charts

Pie charts: best practices

not so bad for two (or few) levels, for part-to-whole task

dubious for several levels if details matter





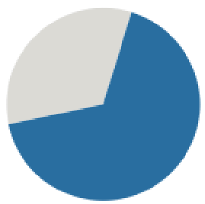
https://eagereyes.org/pie-charts

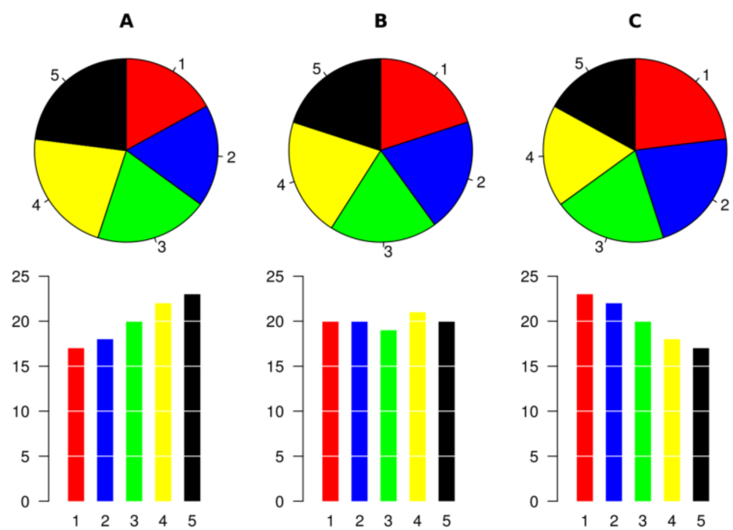
Pie charts: best practices

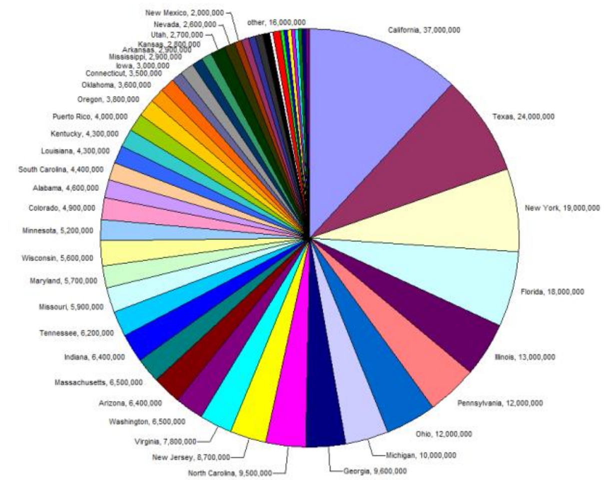
not so bad for two (or few) levels, for part-to-whole task

dubious for several levels if details matter

terrible for many levels







https://eagereyes.org/pie-charts

Idioms: **normalized stacked bar chart**

task

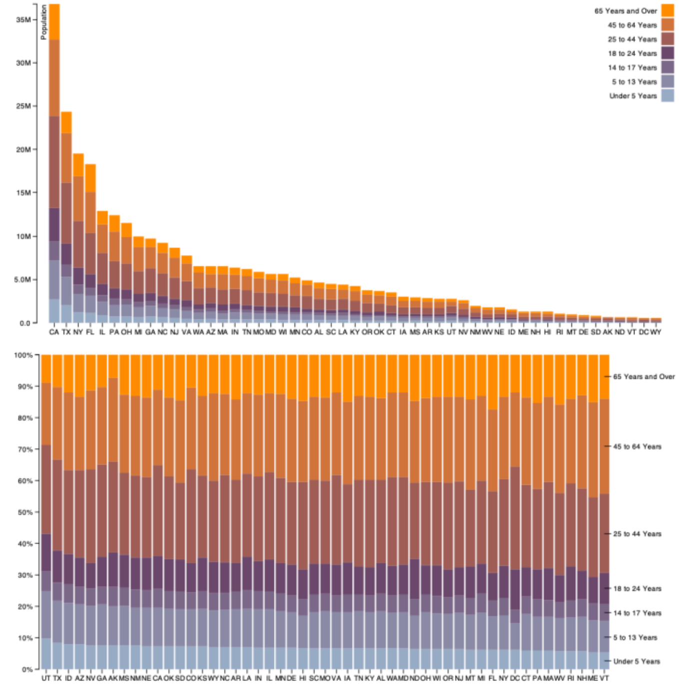
- part-to-whole judgements

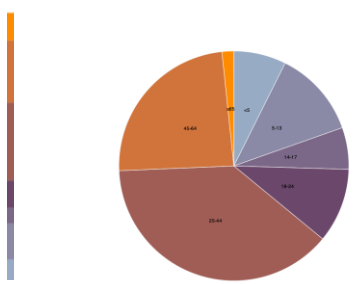
normalized stacked bar chart

- stacked bar chart, normalized to full vert height
- single stacked bar equivalent to full pie
  - high information density: requires narrow rectangle

pie chart

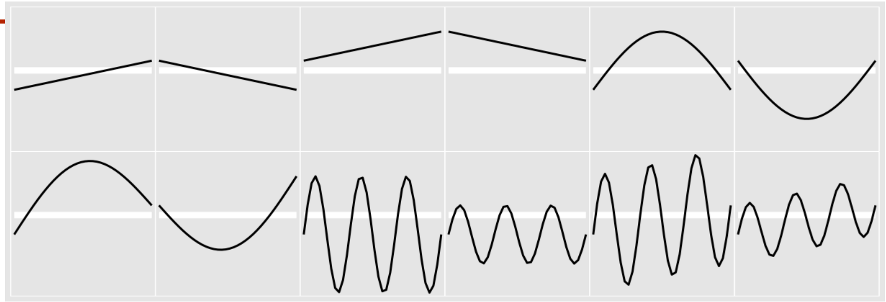
- information density: requires large circle



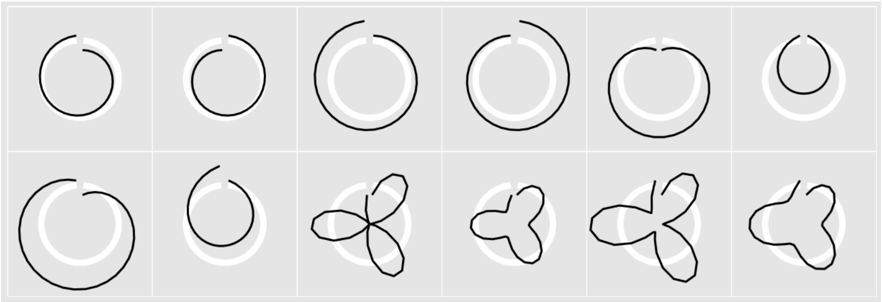


http://bl.ocks.org/mbostock/3886208
http://bl.ocks.org/mbostock/3887235
http://bl.ocks.org/mbostock/3886394

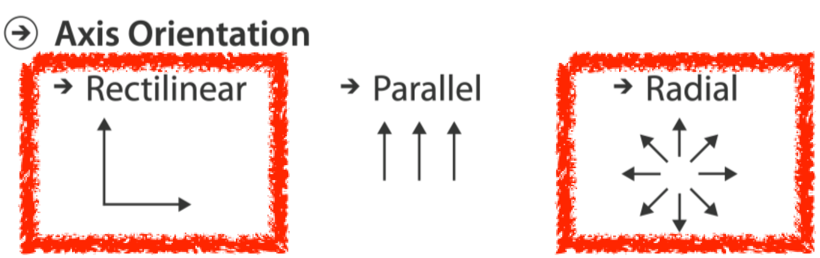
Idiom: **glyphmaps**



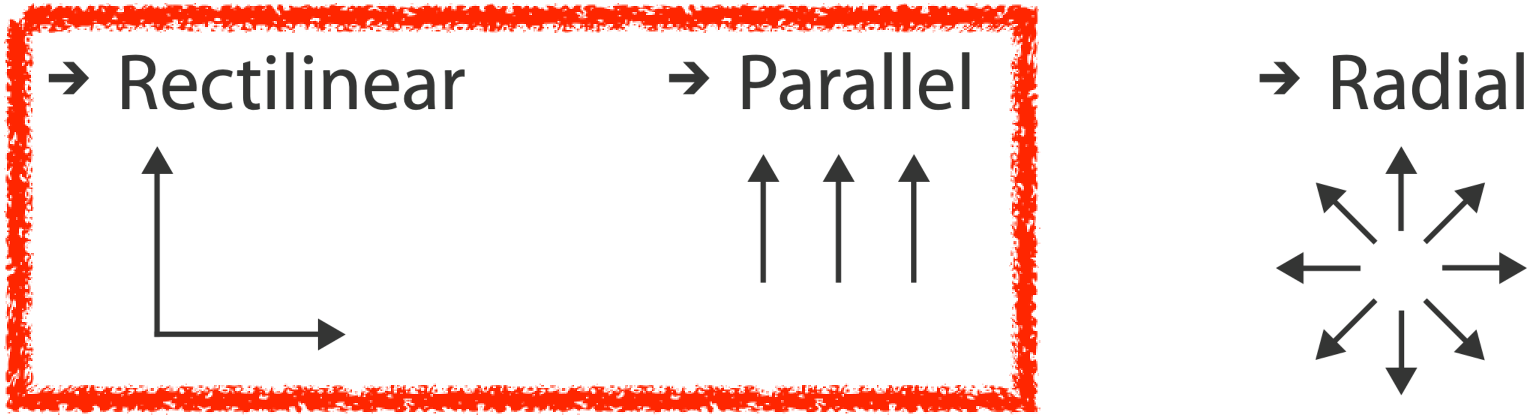
rectilinear good for
- linear vs nonlinear trends



radial good for cyclic patterns
- evaluating periodicity



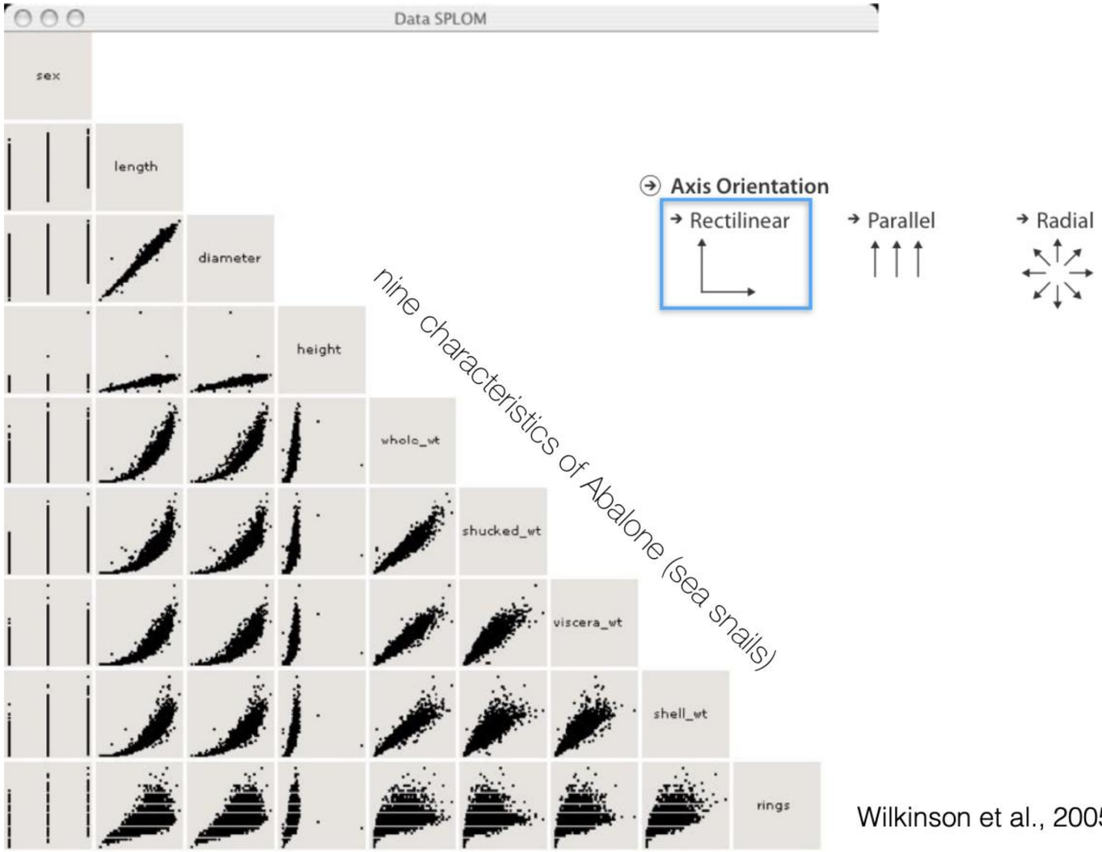
Axis Orientation



Idiom: **SPLOM**

scatterplot matrix (SPLOM)

- rectilinear axes, point mark
- all possible pairs of axes
- scalability
  - one dozen attribs
  - dozens to hundreds of items



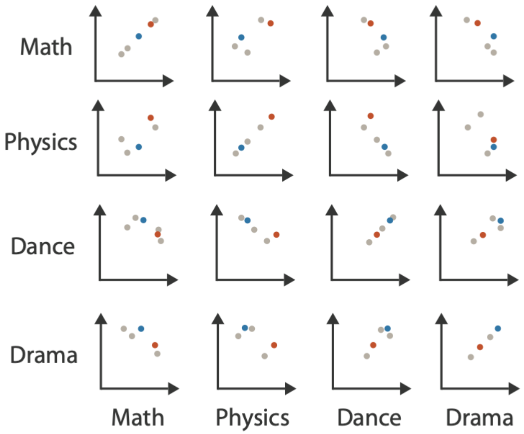
Wilkinson et al., 2005

Idioms: **parallel coordinates**

scatterplot limitation

- visual representation with orthogonal axes
- can show only two attributes with spatial position channel

Scatterplot Matrix



Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

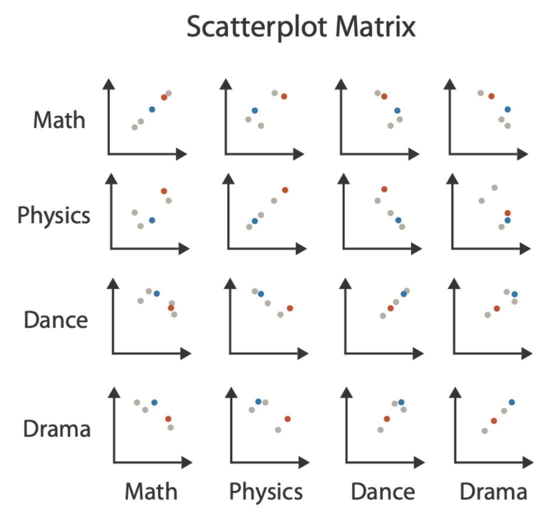
http://www.michaelmcguffin.com/courses/vis/

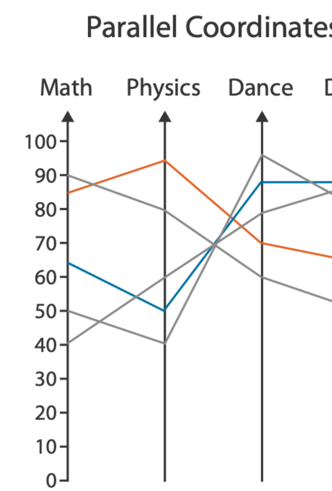
Idioms: **parallel coordinates**

scatterplot limitation
- visual representation with orthogonal axes
- can show only two attributes with spatial position channel

alternative: line up axes in parallel to show many attributes with position
- item encoded with a line with n segments
- n is the number of attributes shown

parallel coordinates
- parallel axes, jagged line for item
- rectilinear axes, item as point
  axis ordering is major challenge
- scalability
  dozens of attribs
  hundreds of items





Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

http://www.michaelmcguffin.com/courses/vis/

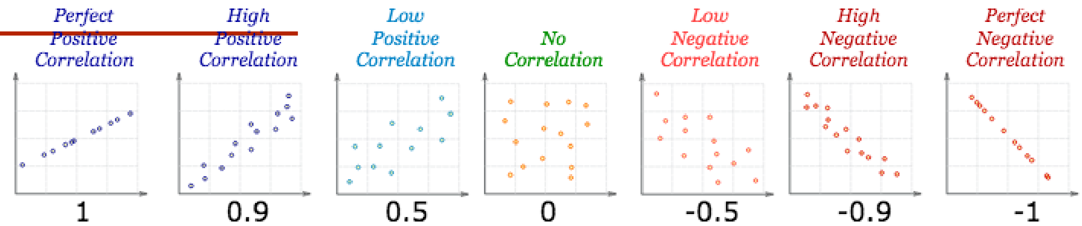
Task: Correlation

scatterplot matrix

- positive correlation
  - diagonal low-to-high
- negative correlation
  - diagonal high-to-low
- uncorrelated: spread out

parallel coordinates

- positive correlation
  - parallel line segments
- negative correlation
  - all segments cross at halfway point
- uncorrelated
  - scattered crossings



https://www.mathsisfun.com/data/scatter-xy-plots.html

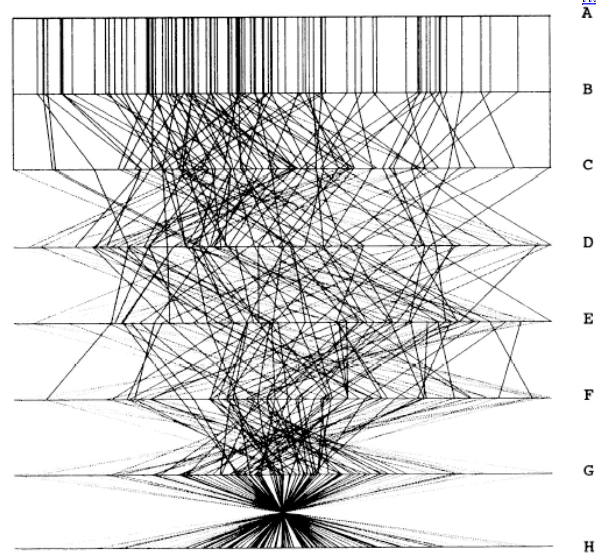


Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $\rho = 1, .8, .2, 0, -.2, -.8$, and -1.

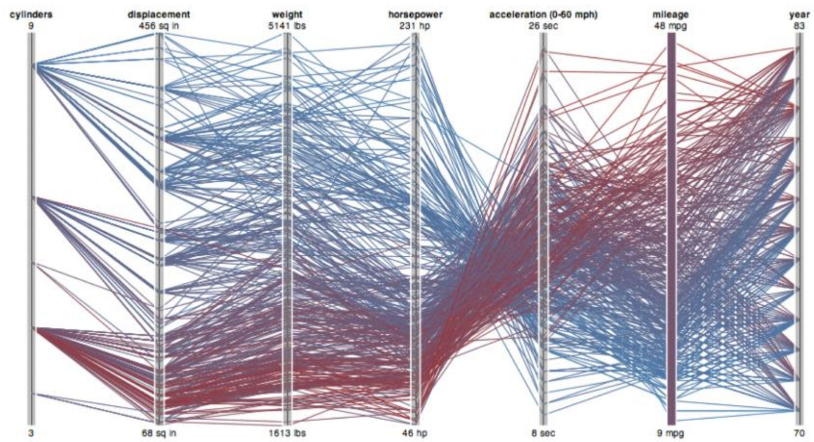
Parallel coordinates, limitations

visible patterns only between neighboring axis pairs

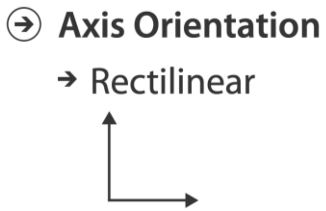
how to pick axis order?

- usual solution: reorderable axes, interactive exploration
- same weakness as many other techniques

  downside of interaction: human-powered search
- some algorithms proposed, none fully solve



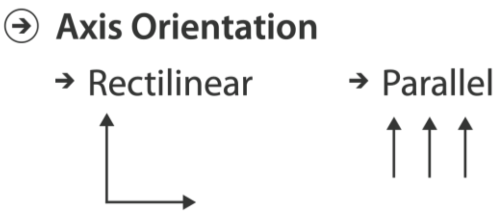
Orientation limitations



rectilinear: scalability wrt #axes

2 axes best, 3 problematic, 4+ impossible

Orientation limitations

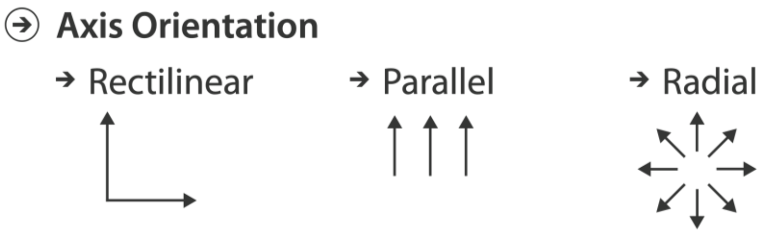


rectilinear: scalability wrt #axes

- 2 axes best, 3 problematic, 4+ impossible

parallel: unfamiliarity, training time

Orientation limitations



rectilinear: scalability wrt #axes

2 axes best, 3 problematic, 4+ impossible

parallel: unfamiliarity, training time

radial: perceptual limits

– polar coordinate asymmetry

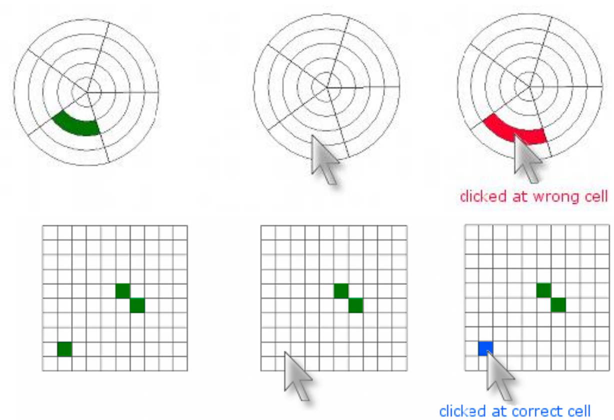
angles lower precision than length

nonuniform sector width/size depending on radial distance

– frequently problematic

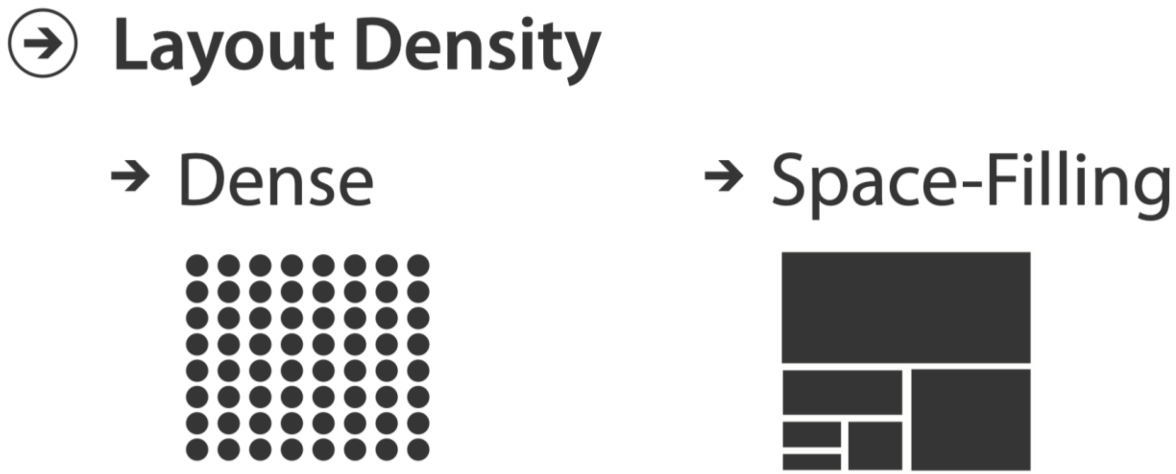
but sometimes can be deliberately exploited!

– for 2 attribs of very unequal importance



[Uncovering Strengths and Weaknesses of Radial Visualizations - an Empirical Approach. Diehl, Beck and Burch. IEEE TVCG (Proc. InfoVis) 16(6):935--942, 2010.]

Layout density



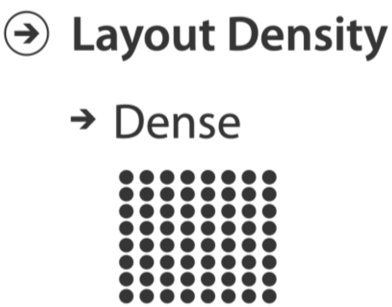
Idiom: Dense software overviews

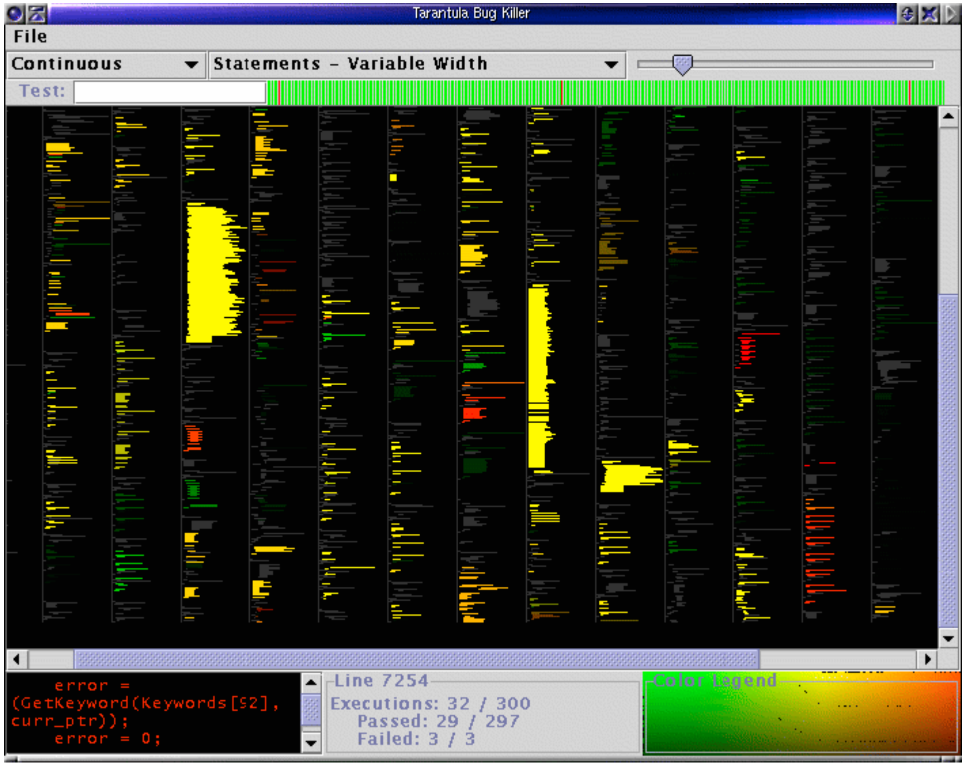
data: text
- text + 1 quant attrib per line

derived data:
- one pixel high line
- length according to original

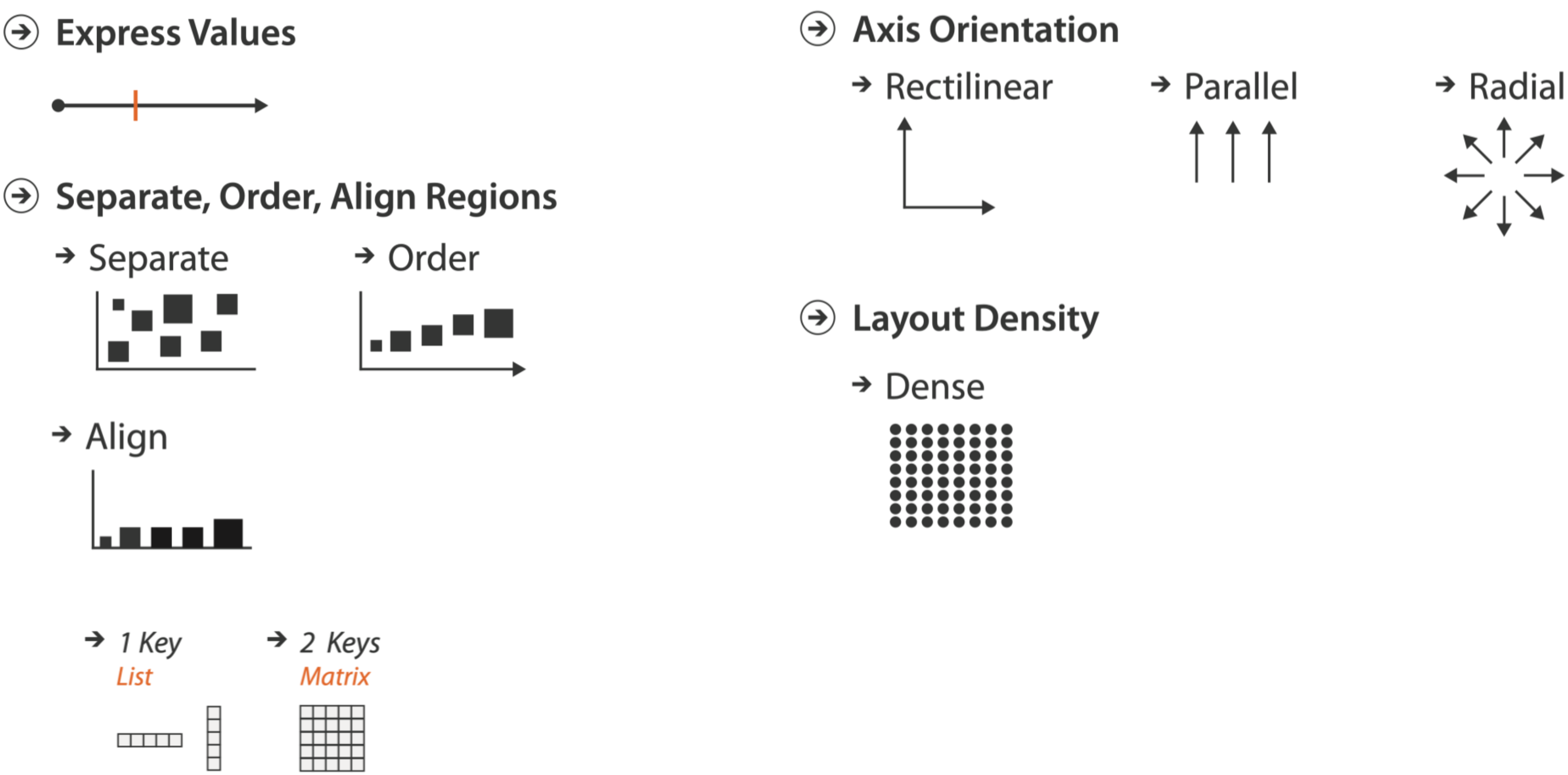
color line by attrib

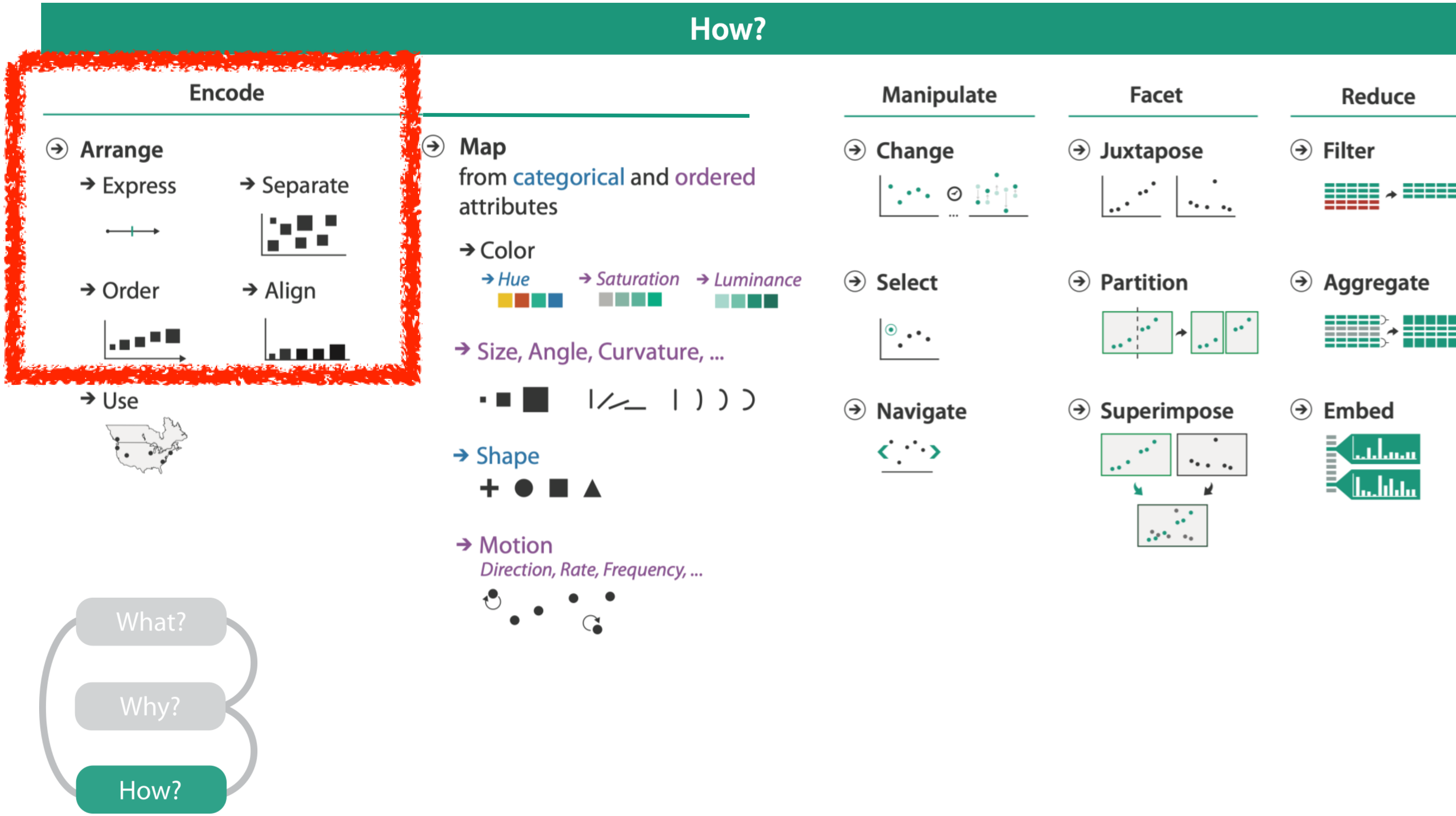
scalability
- 10K+ lines





Arrange tables




How?
Encode
Arrange
Express
Separate
Order
Align
Use
Map
from categorical and ordered attributes
Color
Hue
Saturation
Luminance
Size, Angle, Curvature, ...
Shape
Motion
Direction, Rate, Frequency, ...
Manipulate
Change
Select
Navigate
Facet
Juxtapose
Partition
Superimpose
Reduce
Filter
Aggregate
Embed
What?
Why?
How?

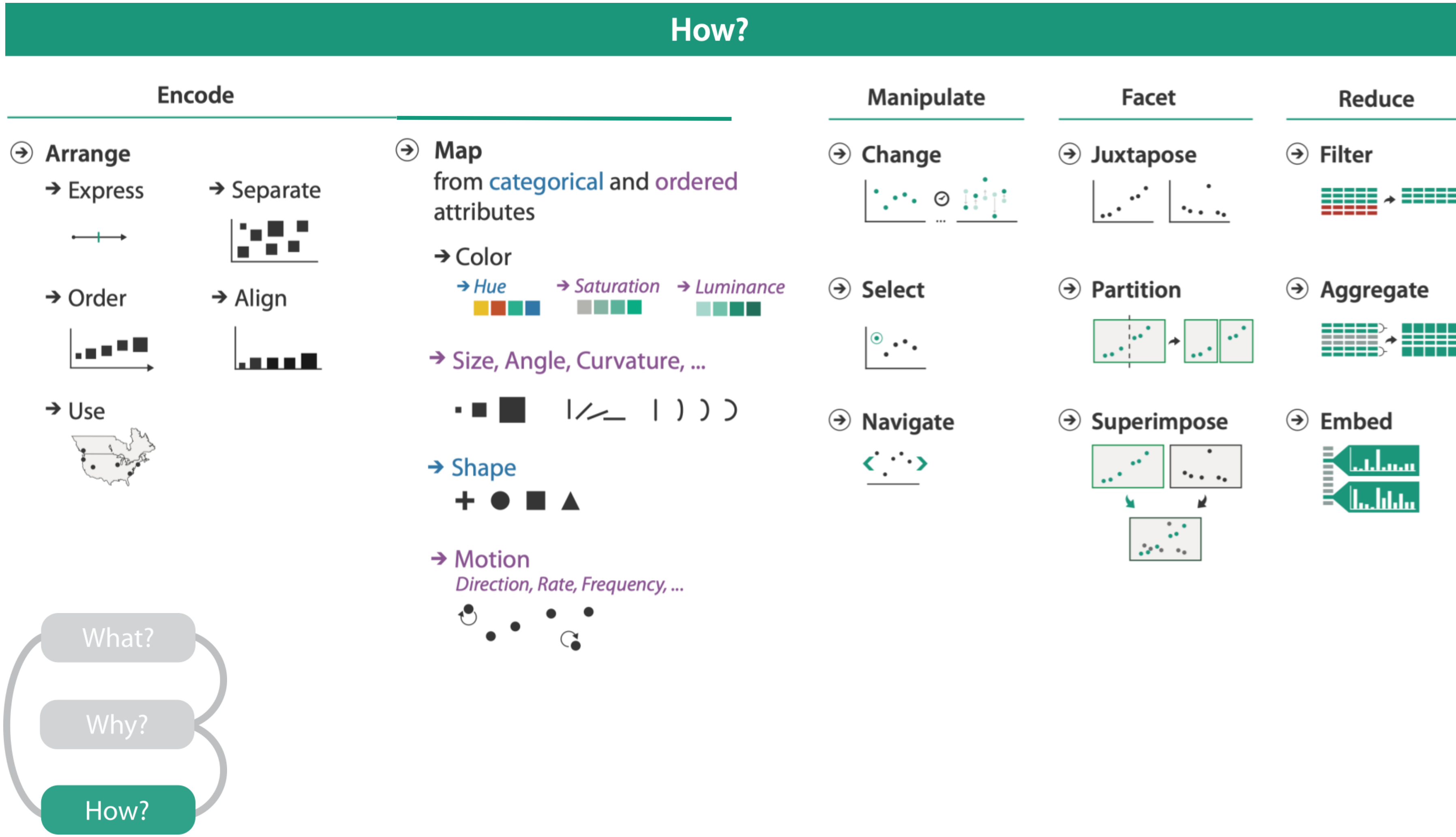
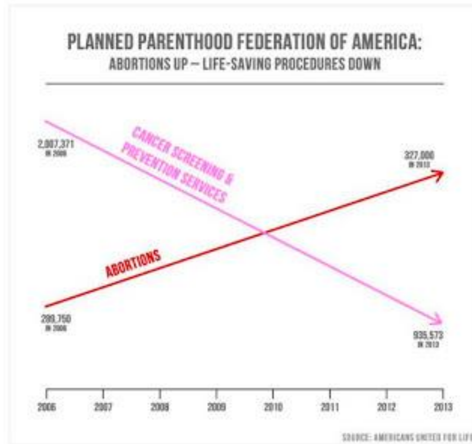

How?
Encode
Manipulate
Facet
Reduce
Arrange
Express
Separate
Order
Align
Use
Map
from categorical and ordered attributes
Color
Hue
Saturation
Luminance
Size, Angle, Curvature, ...
Shape
Motion
Direction, Rate, Frequency, ...
Change
Select
Navigate
Juxtapose
Partition
Superimpose
Filter
Aggregate
Embed
What?
Why?
How?

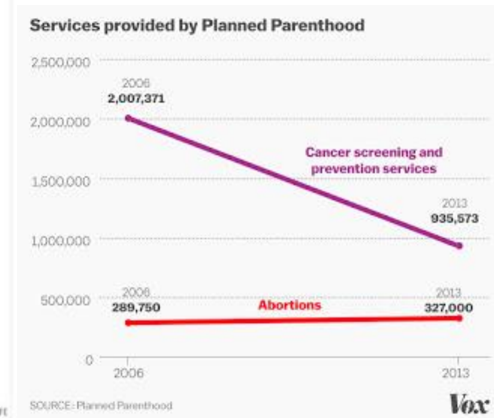
Chart axes

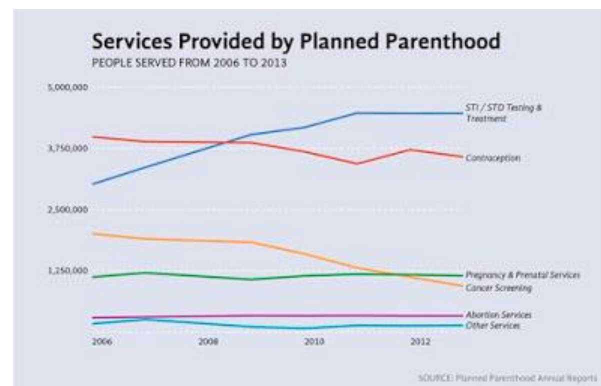
labelled axis is critical

avoid cropping y-axis

- include 0 at bottom left
- or slope misleads





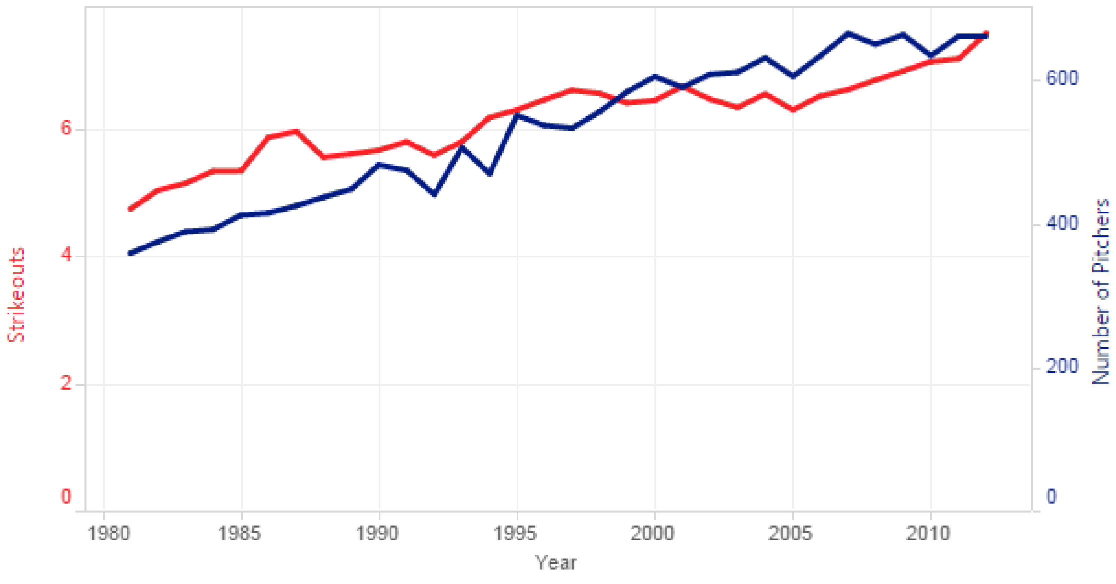


http://www.thefunctionalart.com/2015/10/if-you-see-bullshit-say-bullshit.html

Idiom: **dual-axis line charts**

controversial

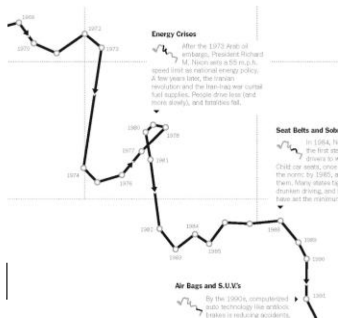
- acceptable if commensurate
- beware, very easy to mislead!

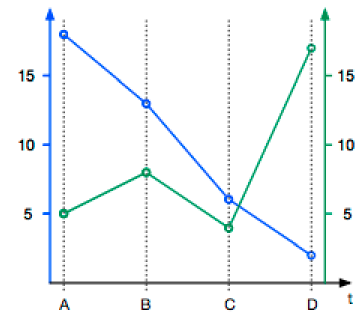


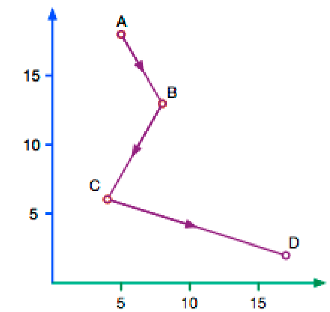
Source | http://www.baseball-reference.com/leagues/MLB/pitch.shtml Ben Jones (@DataRemixed) | 5/4/2013

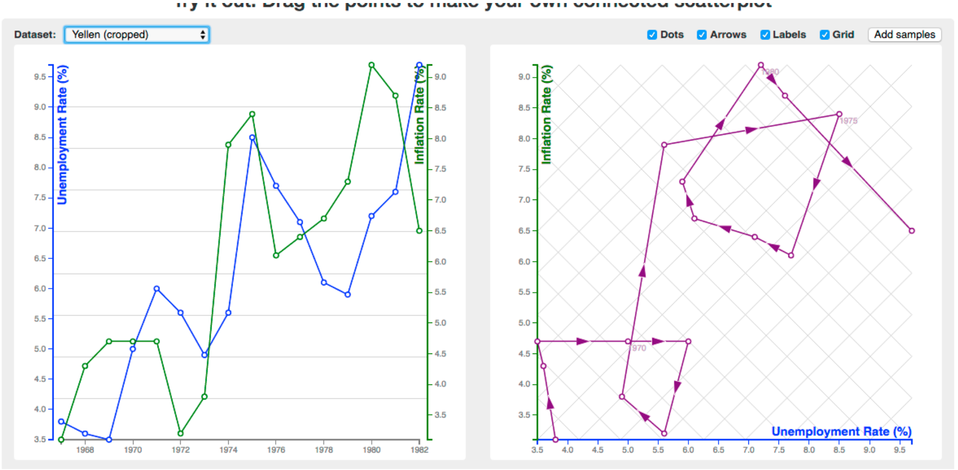
Idiom: **connected scatterplots**

- scatterplot with line connection marks
  - popular in journalism
  - horiz + vert axes: value attri
  - line connection marks: temporal order
  - alternative to dual-axis charts
    - horiz: time
    - vert: two value attribs
- empirical study
  - engaging, but correlation unclear









http://steveharoz.com/research/connected_scatterplot/

Choosing line chart aspect ratios

1: banking to 45 (1980s)

- Cleveland perceptual argument: most accurate angle judgement at 45

Fig 7.1 Sunspot Data: Aspect Ratio 1

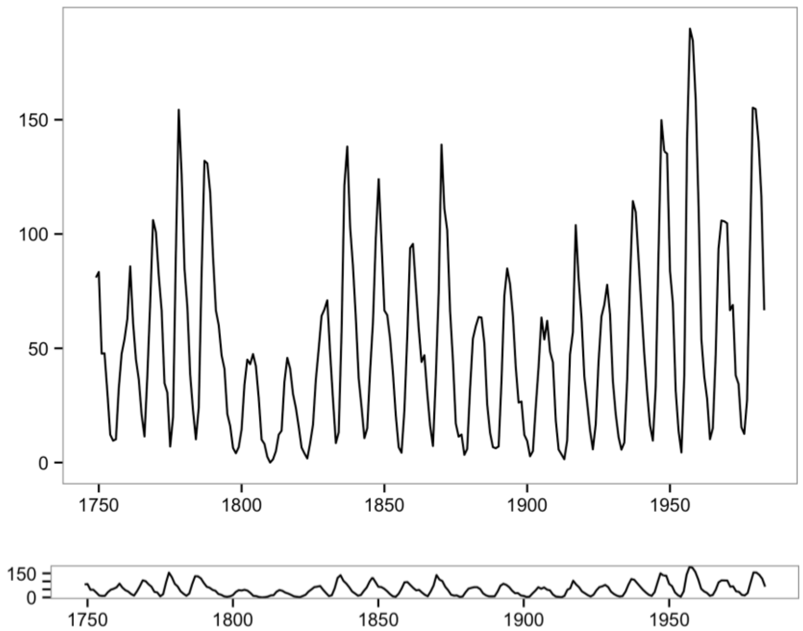
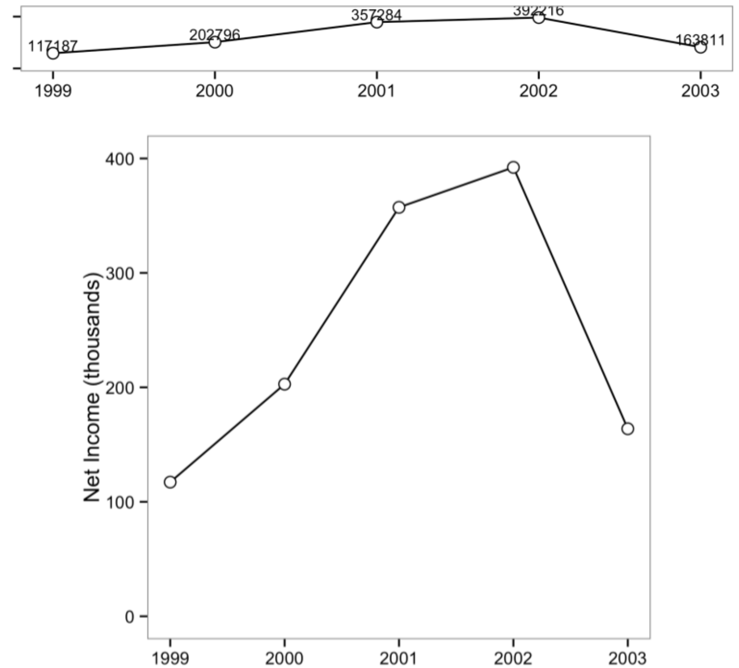


Fig 7.2 Annual Report: Aspect Ratio 2



Choosing line chart aspect ratios

2: multi scale banking to 45 (2006)

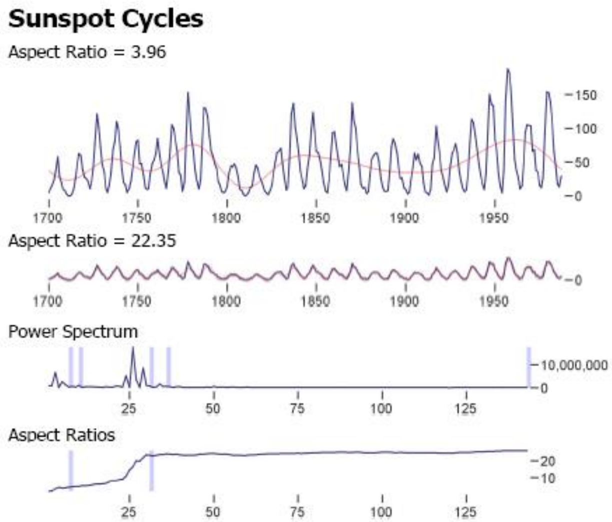
– frequency domain analysis to find ratios

FFT the data, convolve with Gaussian to smooth

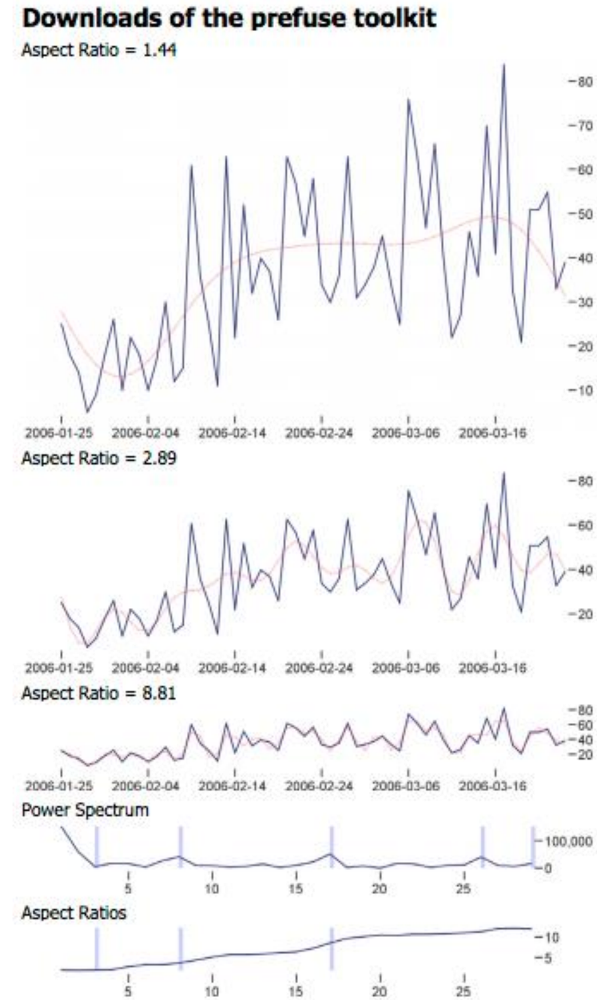
– find interesting spikes/ranges in power spectrum

cull nearby regions if similar, ensure overview

– create trend curves (red) for each aspect ratio



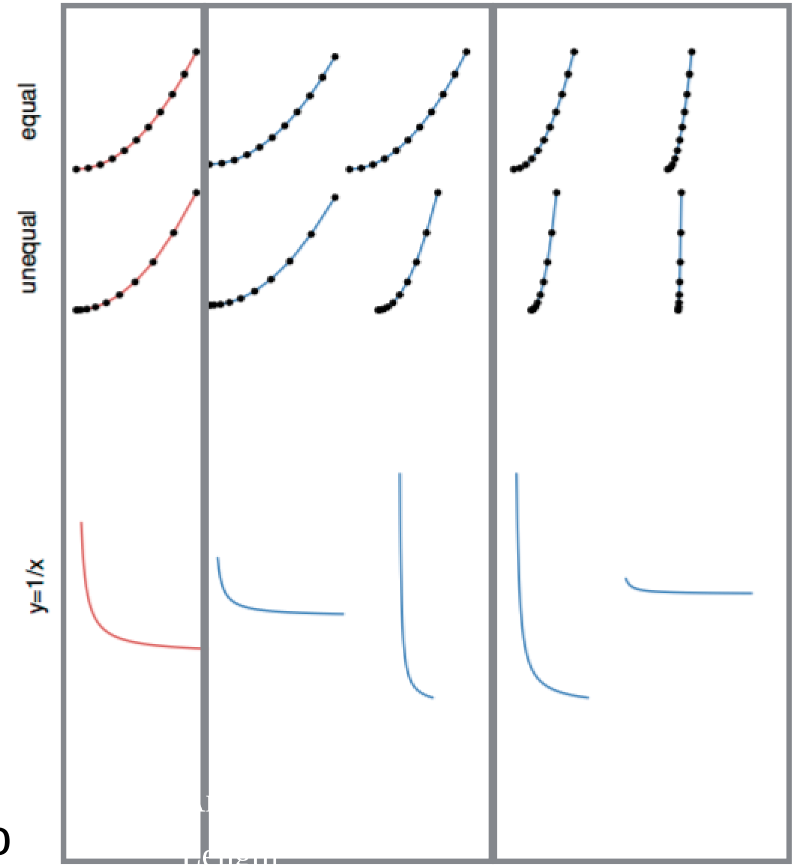
[Multi-Scale Banking to 45 Degrees. Heer and Agrawala, Proc InfoVis 2006]



Choosing line chart aspect ratios

3: arc length based aspect ratio (2011)

- minimize the arc length of curve while keeping the area of the plot constant
- parametrization and scale invariant
- symmetry preserving
- robust & fast to compute



meta-points from this progression

- young field; prescriptive advice changes rapidly
- reasonable defaults required deep dive into perception meets math

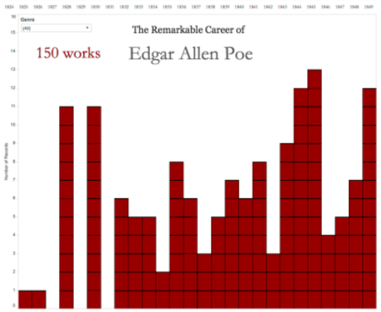
[Arc Length-Based Aspect Ratio Selection. Talbot, Gerth, and Hanrahan. Proc InfoVis 2011]

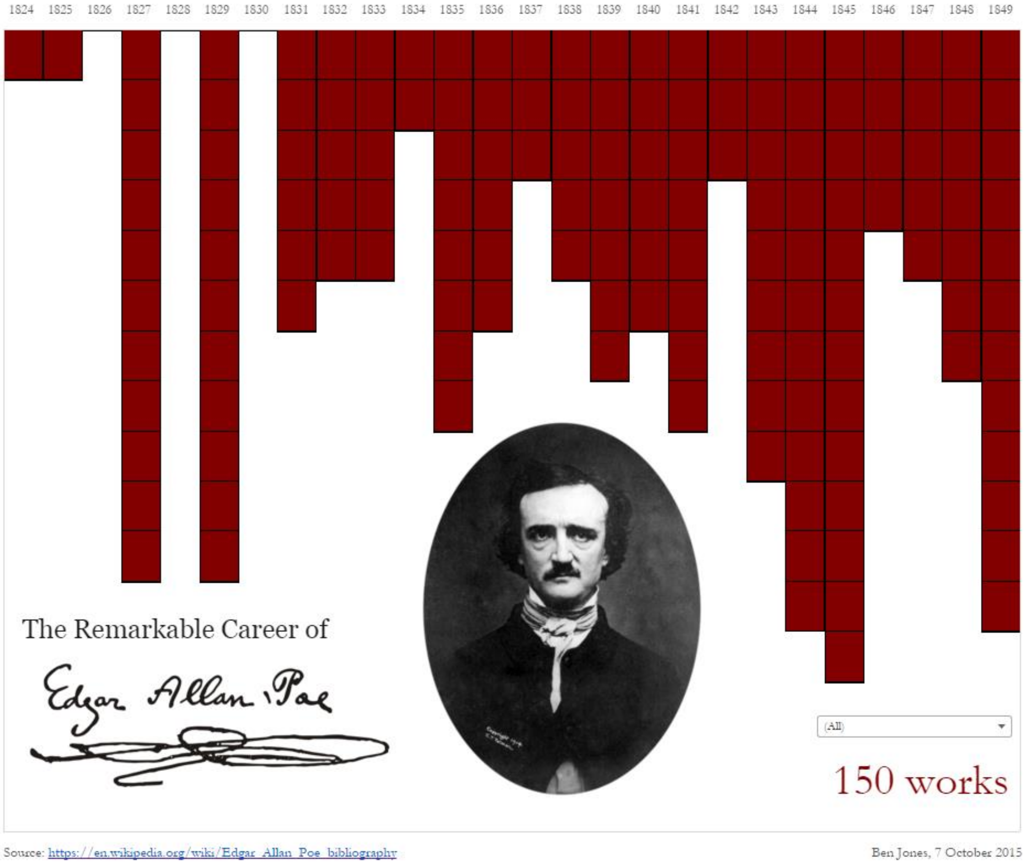
Breaking conventions

presentation vs exploration

- engaging/evocative
- inverted y axis

  blood drips down on Poe





https://public.tableau.com/profile/ben.jones#!/

Network Data

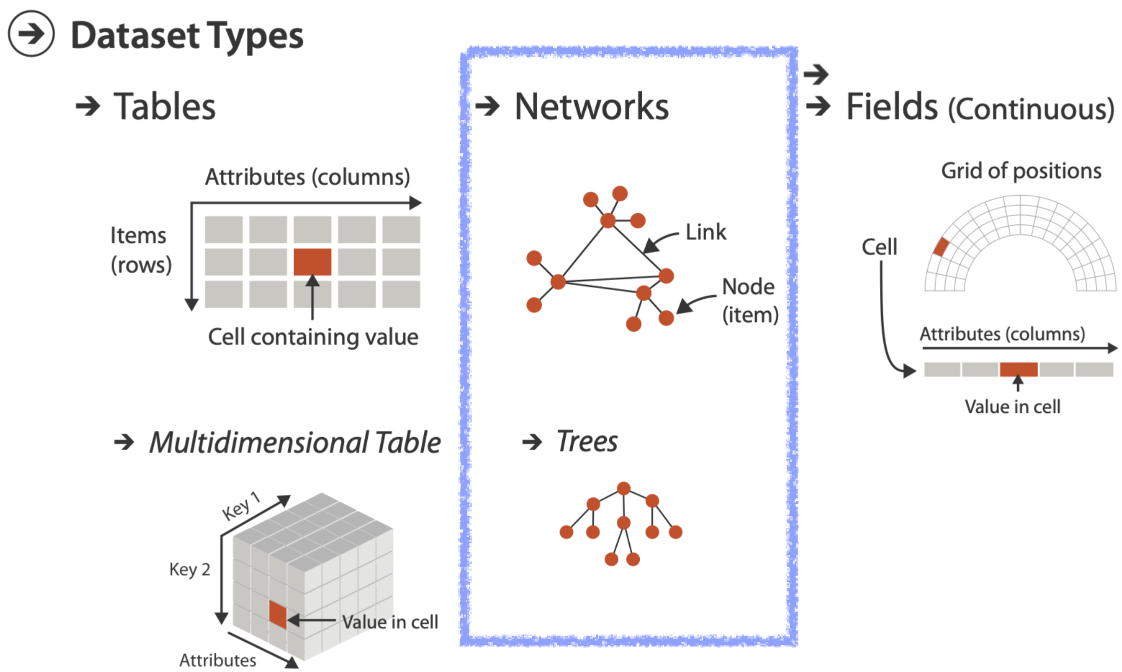
Network data

networks

- model relationships between things
  - aka graphs
- two kinds of items, both can have attributes
  - nodes
  - links

tree

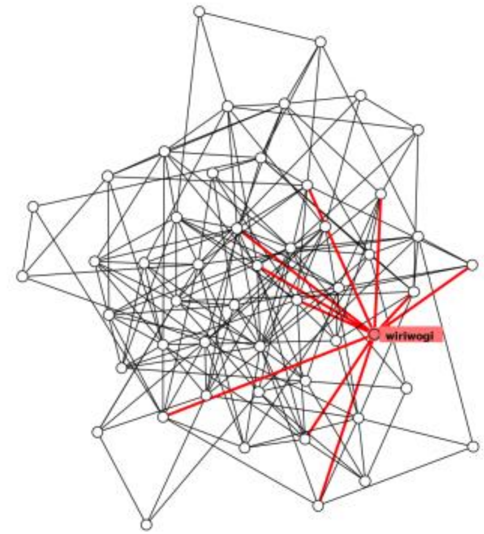
- special case
- no cycles
  - one parent per node



Network tasks: topology-based and attribute-based

topology based tasks

- find paths
- find (topological) neighbors
- compare centrality/importance measures
- identify clusters / communities



attribute based tasks (similar to table data)

- find distributions, ...

combination tasks, incorporating both

- example: find friends-of-friends who like cats

  topology: find all adjacent nodes of given node

  attributes: check if has-pet (node attribute) == cat

Node-link diagrams

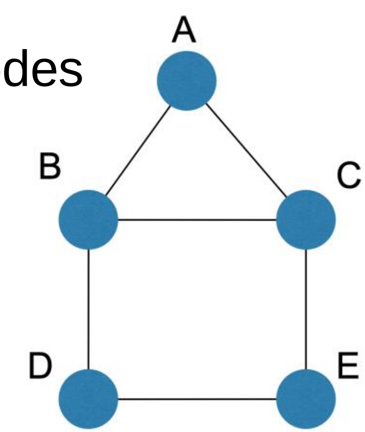
nodes: point marks

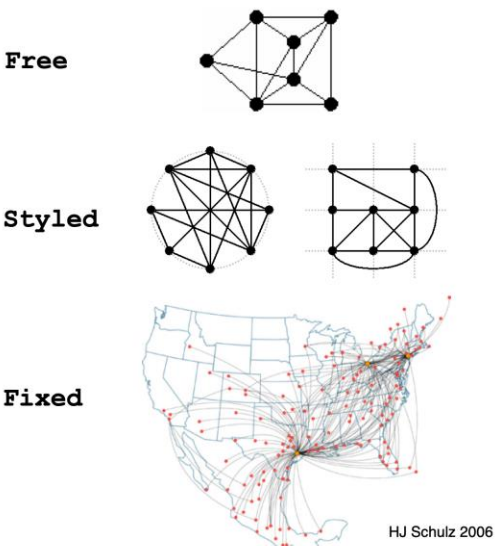
links: line marks

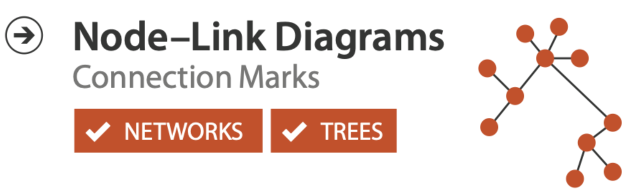
- straight lines or arcs
- connections between nodes

intuitive & familiar

- most common
- many, many variants



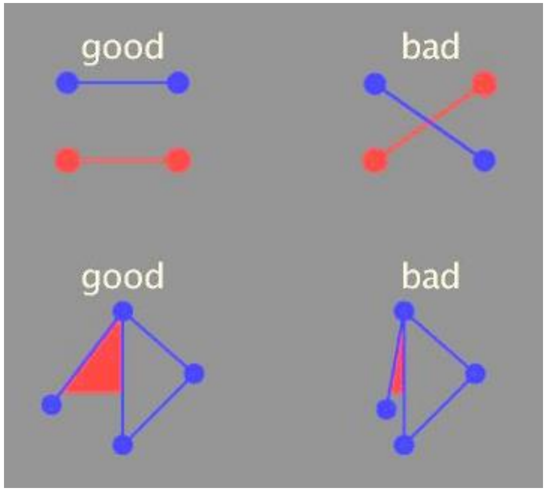




Criteria for good node-link layouts

minimize

- edge crossings, node overlaps
- distances between topological neighbor nodes
- total drawing area
- edge bends



maximize

- angular distance between different edges
- aspect ratio disparities

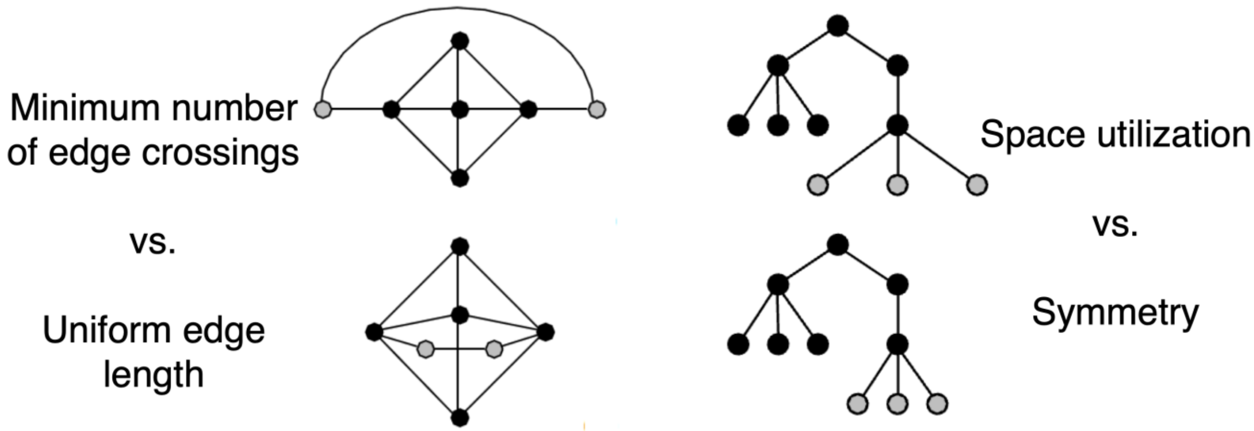
emphasize symmetry

- similar graph structures should look similar in layout

Criteria conflict

most criteria NP-hard individually

many criteria directly conflict with each other



Schulz 2004

Optimization-based layouts

formulate layout problem as optimization problem

convert criteria into weighted cost function

- F(layout) = a*[crossing counts] + b*[drawing space used]+...

use known optimization techniques to find layout at minimal cost

- energy-based physics models
- force-directed placement
- spring embedders

Force-directed placement

physics model

- links = springs pull together
- nodes = magnets repulse apart

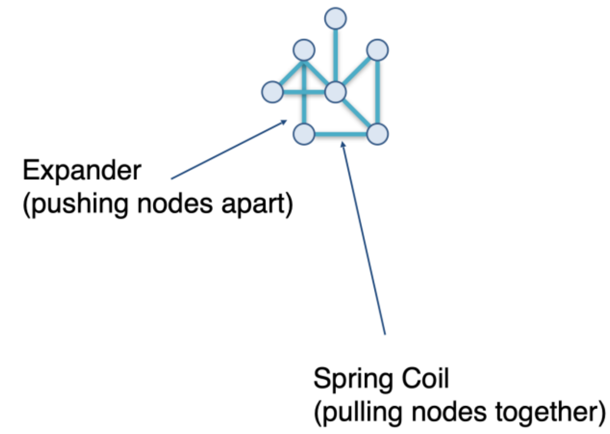
algorithm

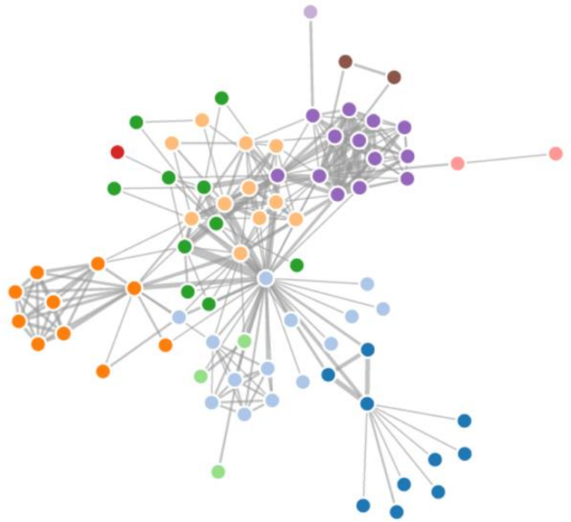
- place vertices in random locations
- while not equilibrium

  calculate force on vertex

  - sum of
    - » pairwise repulsion of all nodes
    - » attraction between connected nodes

  move vertex by c * vertex_force





http://mbostock.github.com/d3/ex/force.html

Force-directed placement properties

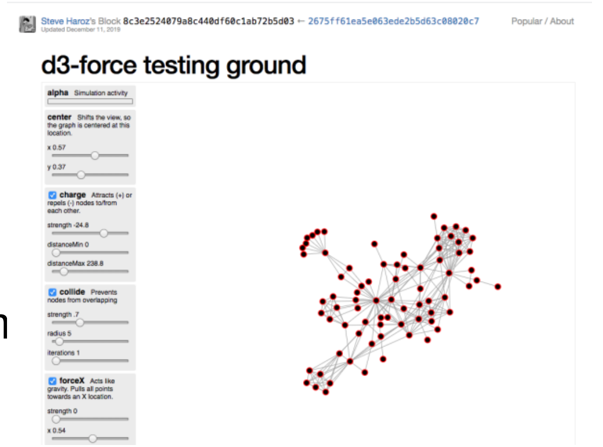
strengths

- reasonable layout for small, sparse graphs
- clusters typically visible
- edge length uniformity

weaknesses

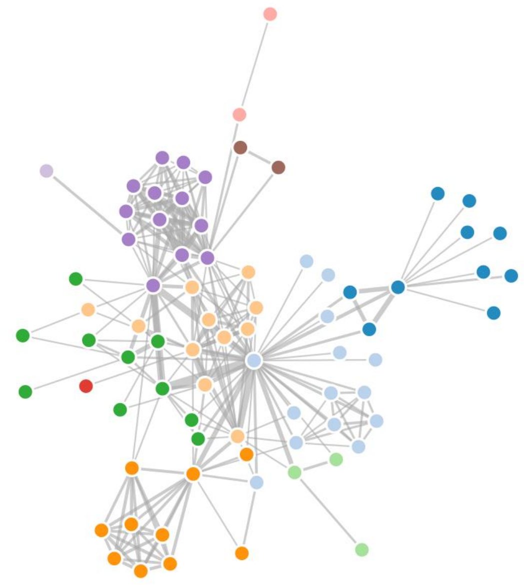
- nondeterministic
- computationally expensive: O(n^3) for n nodes

  each step is n^2, takes ~n cycles to reach equilibrium
- naive FD doesn't scale well beyond 1K nodes
- iterative progress: engaging but distracting



https://bl.ocks.org/steveharoz/8c3e2524079a8c440df60c1ab72b5d03

Idiom: **force-directed placement**



http://mbostock.github.com/d3/ex/force.html

visual encoding

- link connection marks, node point marks

considerations

- spatial position: no meaning directly encoded
  - left free to minimize crossings
- proximity semantics?
  - sometimes meaningful
  - sometimes arbitrary, artifact of layout algorithm
  - tension with length
    - long edges more visually salient than short

tasks

- explore topology; locate paths, clusters

scalability

- node/edge density E < 4N

Idiom: **circular layouts / arc diagrams (node-link)**

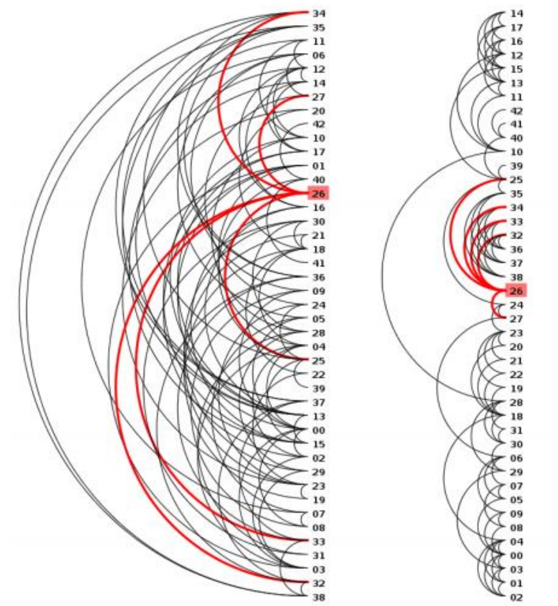
restricted node-link layouts: lay out nodes around circle or along line

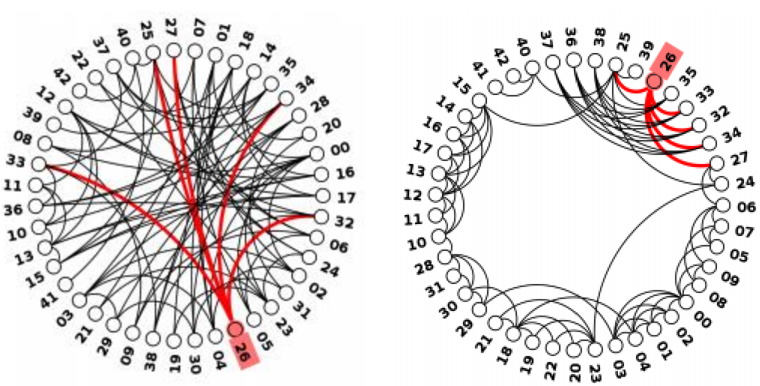
data

- original: network
- derived: node ordering attribute (global computation)

considerations: node ordering crucial to avoid excessive clutter from edge crossings

- examples: before & after barycentric ordering

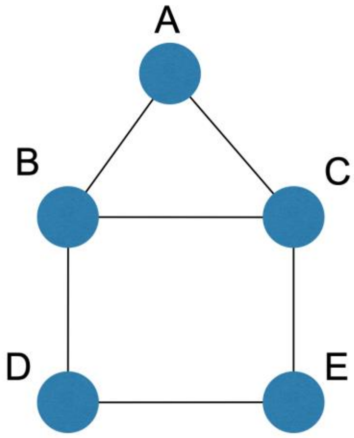


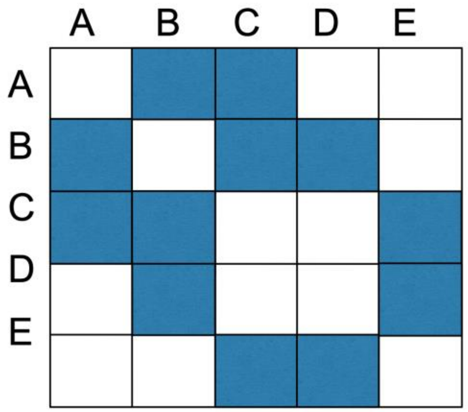


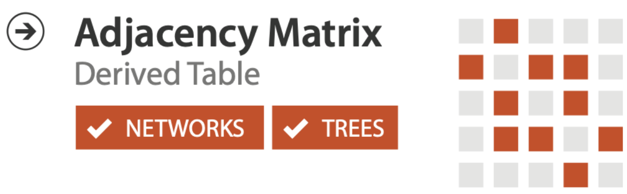
http://profs.etsmtl.ca/mmcguffin/research/2012-mcguffin-simpleNetVis/mcguffin-2012-simpleNetVis.pdf

Adjacency matrix representations

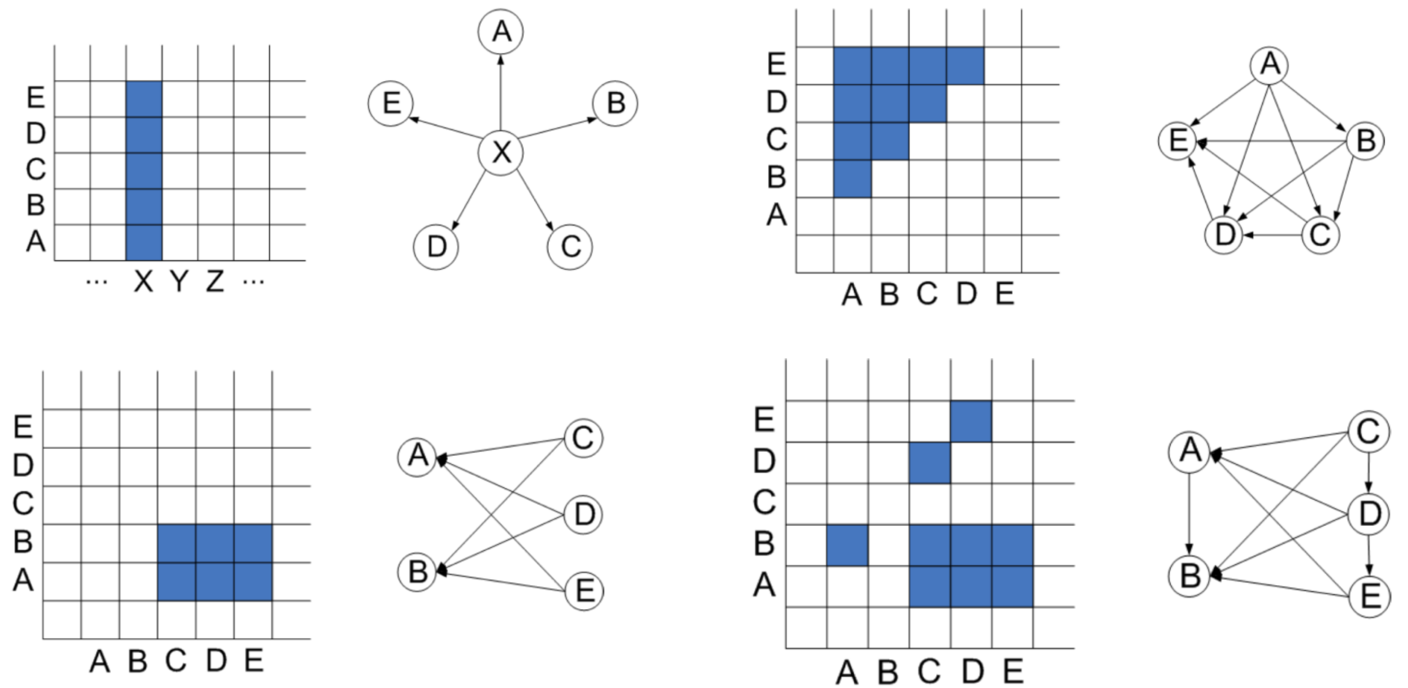
derive adjacency matrix from network





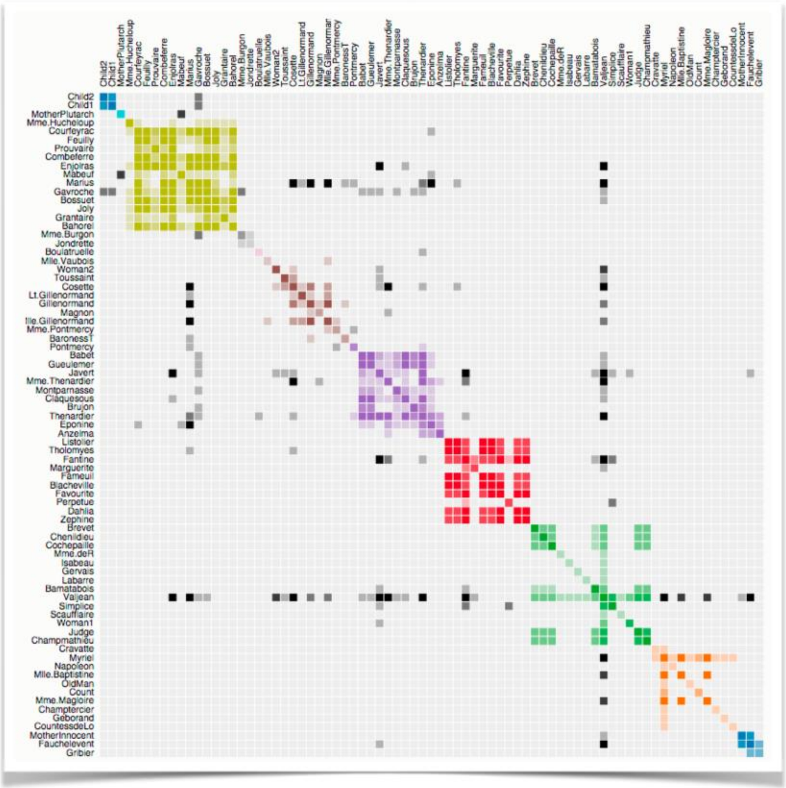


Adjacency matrix examples



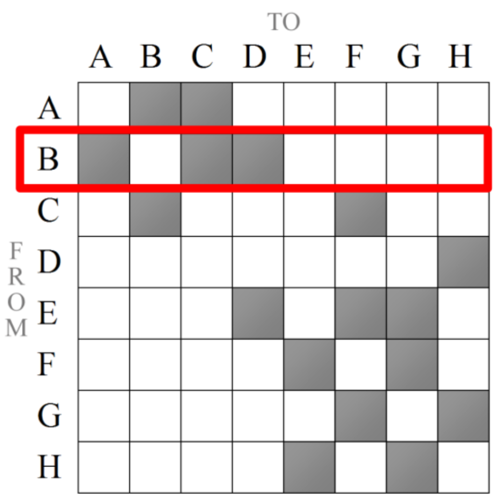
HJ Schulz 2007

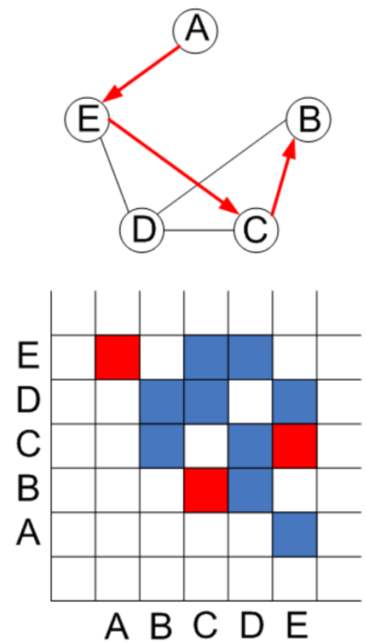
Node order is crucial: Reordering



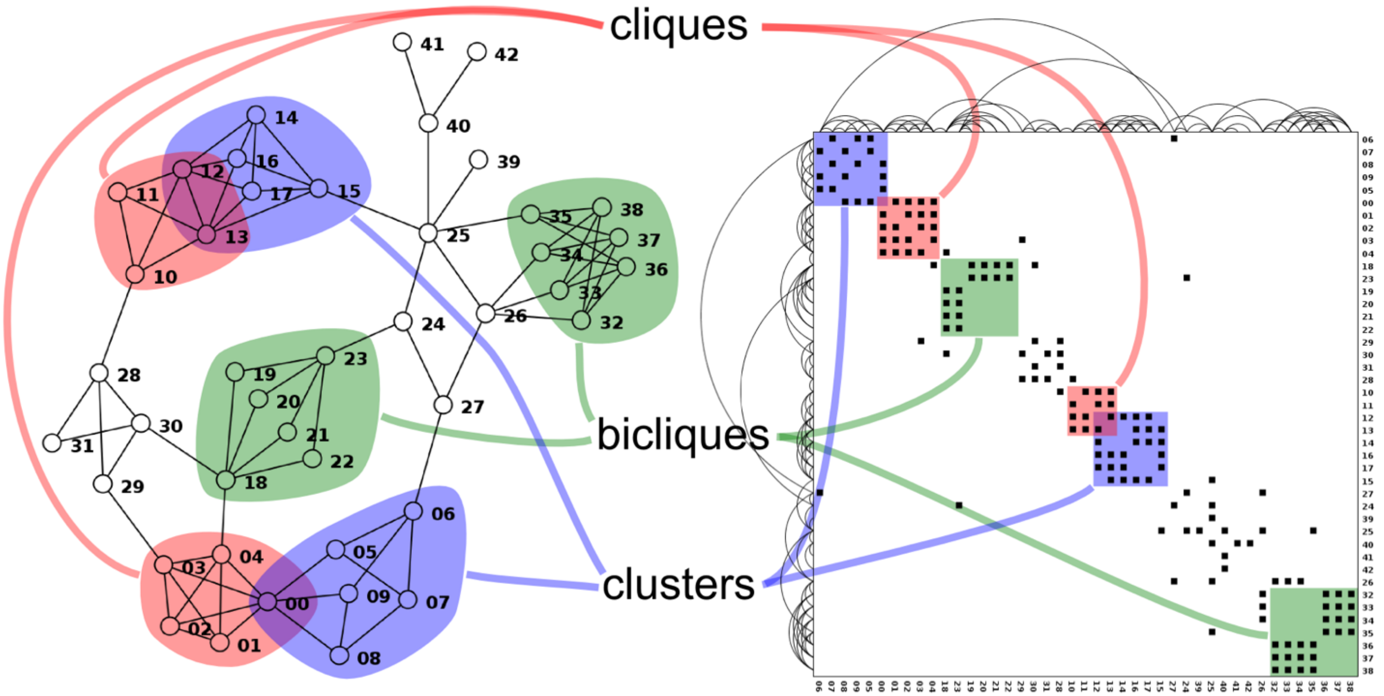
https://bost.ocks.org/mike/miserables/

Adjacency matrix





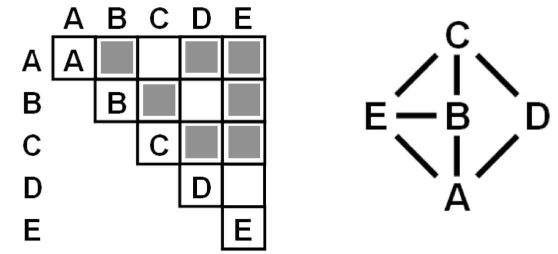
Structures visible in both

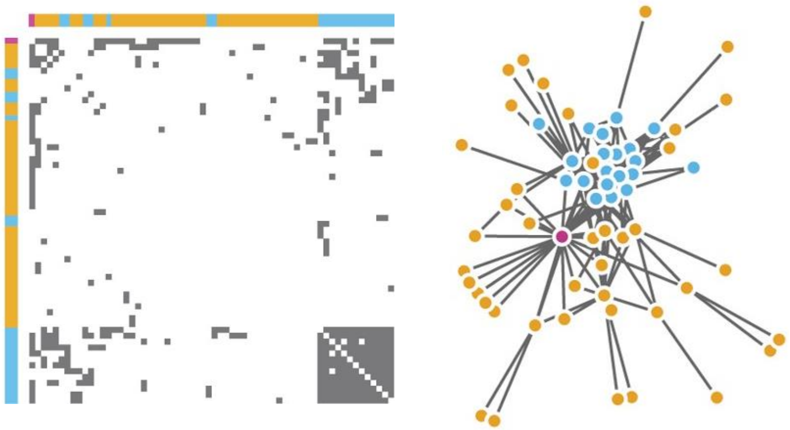


http://www.michaelmcguffin.com/courses/vis/patternsInAdjacencyMatrix.png

Idiom: **adjacency matrix view**

- data: network
  - transform into same data/encoding as heatmap
- derived data: table from network
  - 1 quant attrib
    - weighted edge between nodes
  - 2 categ attribs: node list x 2
- visual encoding
  - cell shows presence/absence of edge
- scalability
  - 1K nodes, 1M edges

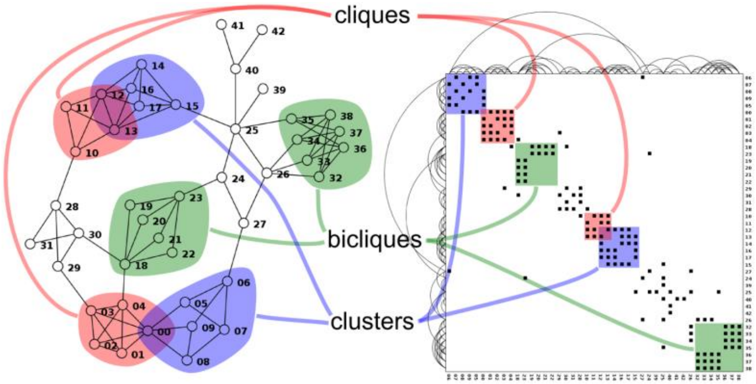




Node-link vs. matrix comparison

node-link diagram strengths

– topology understanding, path tracing
– intuitive, flexible, no training needed



adjacency matrix strengths

– focus on edges rather than nodes
– layout straightforward (reordering needed)
– predictability, scalability
– some topology tasks trainable

http://www.michaelmcguffin.com/courses/vis/patternsInAdjacencyMatrix.png

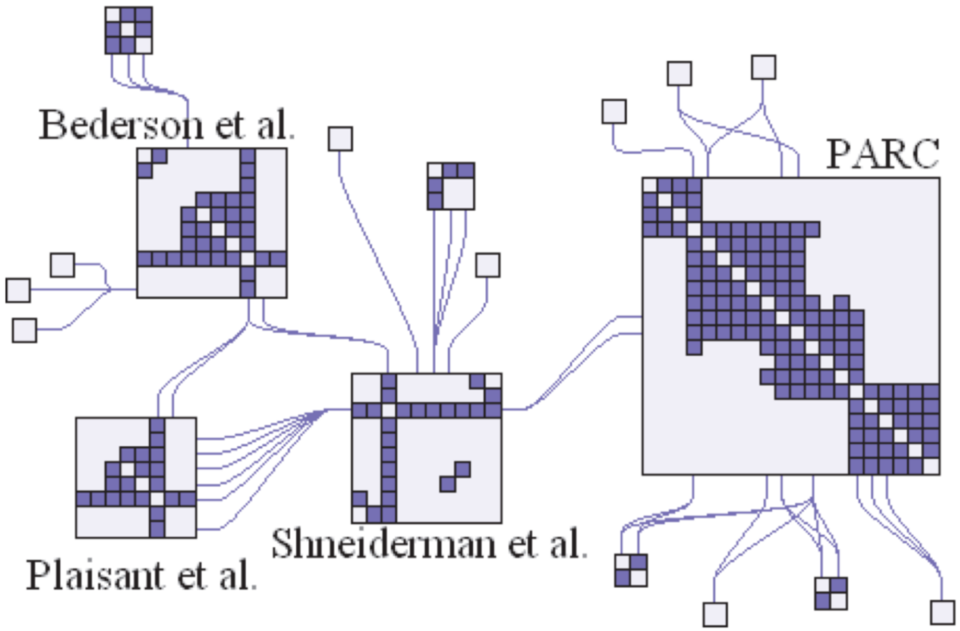
empirical study

– node-link best for small networks
– matrix best for large networks
if tasks don’t involve path tracing!

[On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. Ghoniem, Fekete, and Castagliola. Information Visualization 4:2 (2005), 114–135.]

Idiom: **NodeTrix**

- hybrid nodelink/matrix
- capture strengths of both

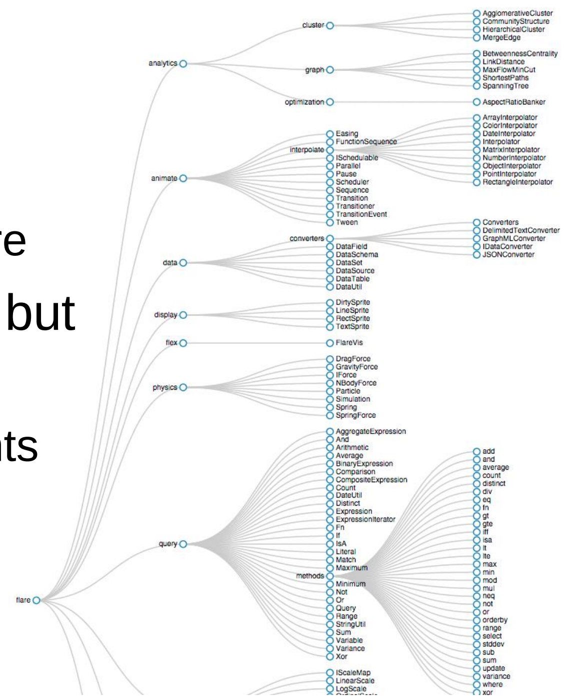


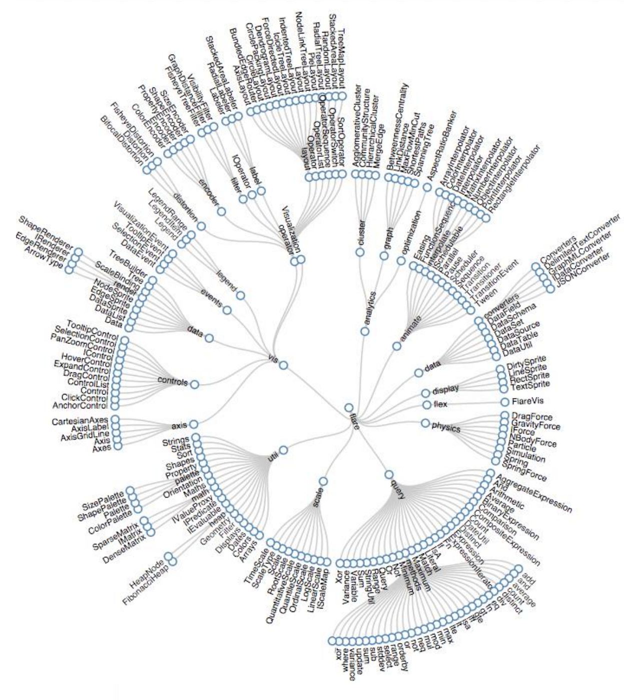
Trees

Node-link trees

Reingold-Tilford

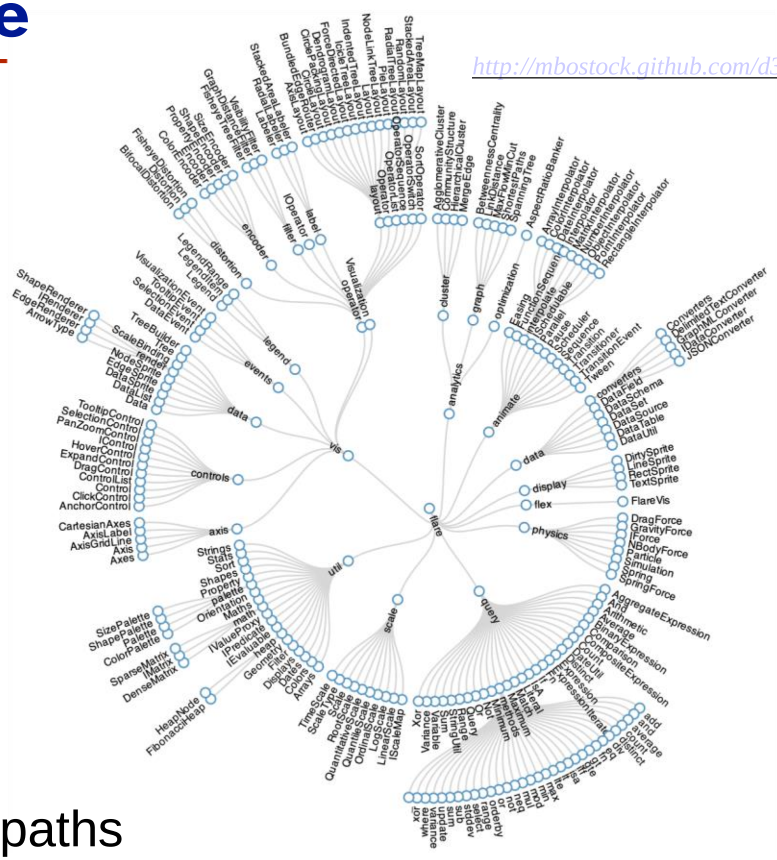
- tidy drawings of trees
  - exploit parent/child structure
- allocate space: compact but without overlap
  - rectilinear and radial variants
- nice algorithm writeup





Idiom: **radial node-link tree**

http://mbostock.github.com/d3/ex/tree.htm



data

- tree

encoding

- link connection marks
- point node marks
- radial axis orientation
  - angular proximity: siblings
  - distance from center: depth in tree

tasks

- understanding topology, following paths

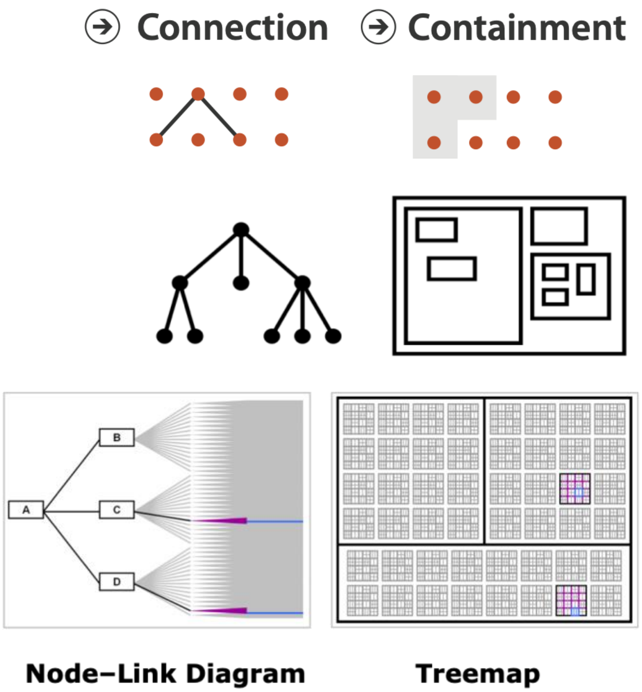
scalability

- 1K - 10K nodes (with/without labels)

Link marks: Connection and containment

marks as links (vs. nodes)

- common case in network drawing
- 1D case: connection
  - ex: all node-link diagrams
  - emphasizes topology, path tracing
  - networks and trees
- 2D case: containment
  - ex: all treemap variants
  - emphasizes attribute values at leaves (size coding)
  - only trees



[Elastic Hierarchies: Combining Treemaps and Node-Link Diagrams. Dong, McGuffin, and Chignell. Proc. InfoVis 2005, p. 57-64.]

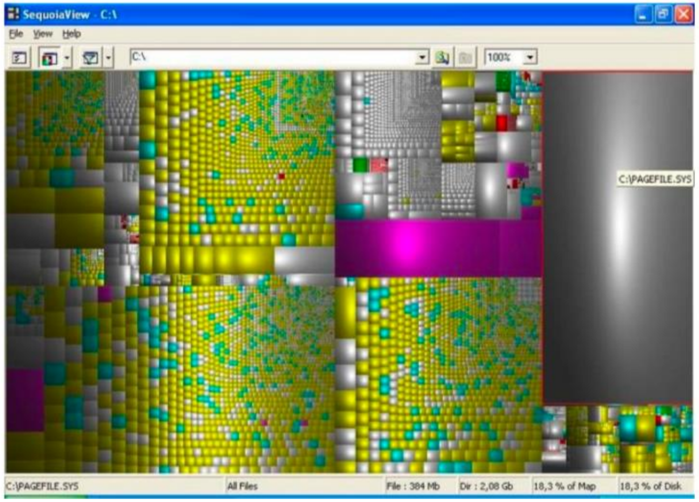
Idiom: **treemap**

data
- tree
- 1 quant attrib at leaf nodes

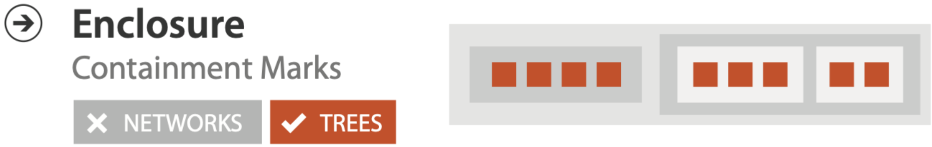
encoding
- area containment marks for hierarchical structure
- rectilinear orientation
- size encodes quant attrib

tasks
- query attribute at leaf nodes
- ex: disk space usage within filesystem

scalability
- 1M leaf nodes



https://www.win.tue.nl/sequoiaview/



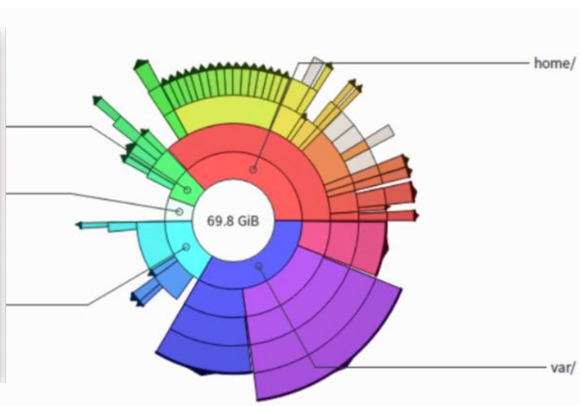
Idiom: implicit tree layouts (sunburst, icicle plot)

alternative to connection and containment: position

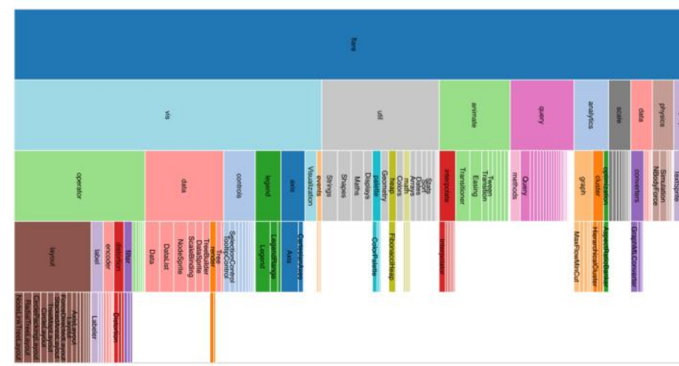
– show parent-child relationships only through relative positions

Treemap

Sunburst



Icicle Plot



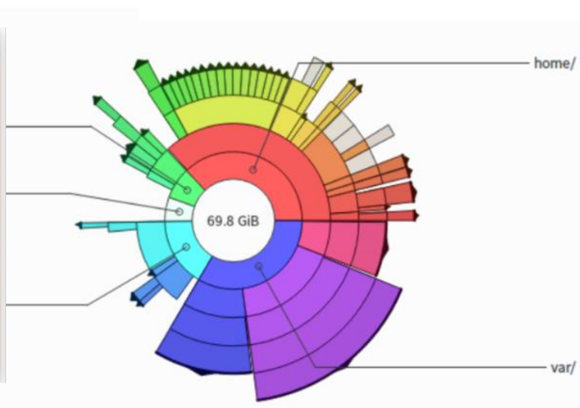
Idiom: implicit tree layouts (sunburst, icicle plot)

alternative to connection and containment: position

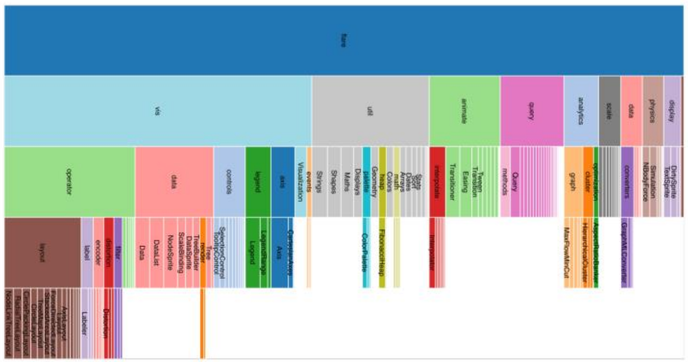
- show parent-child relationships only through relative positions

Treemap

Sunburst



Icicle Plot



Idiom: implicit tree layouts (sunburst, icicle plot)

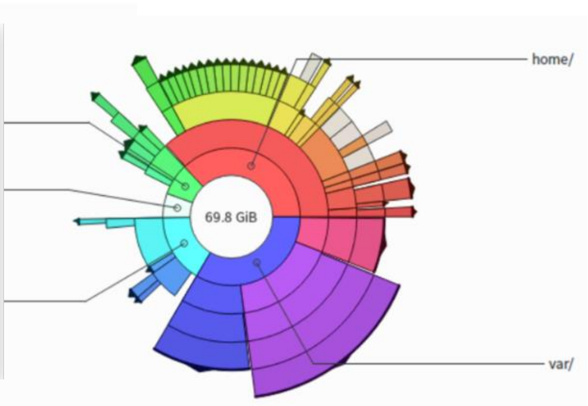
alternative to connection and containment: position

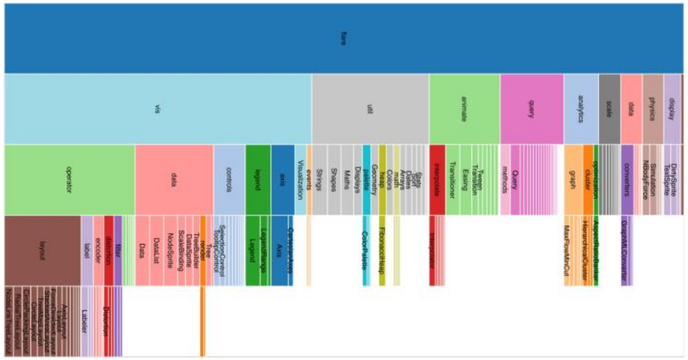
- show parent-child relationships only through relative positions

Treemap

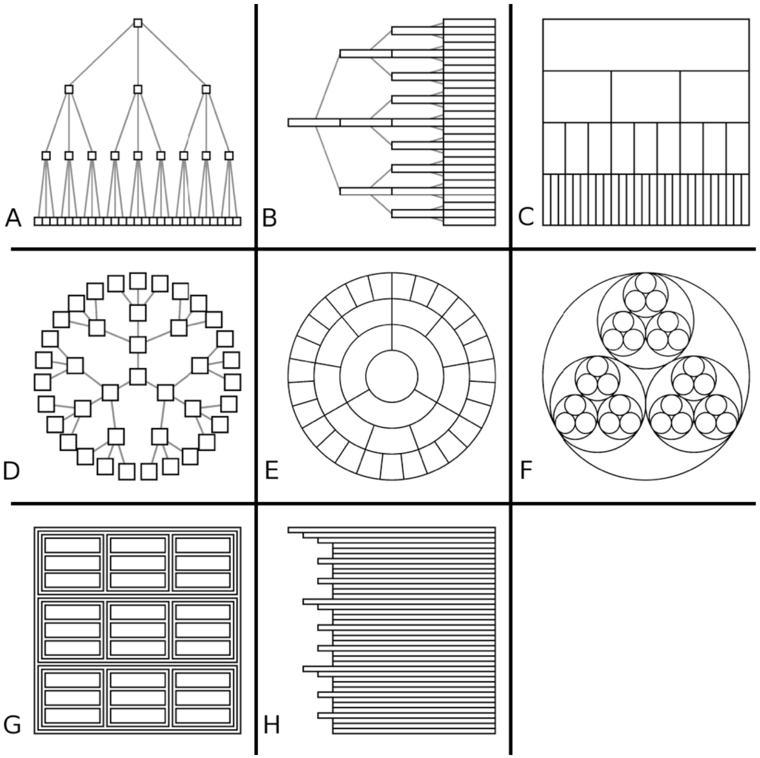
Sunburst

Icicle Plot





Tree drawing idioms comparison

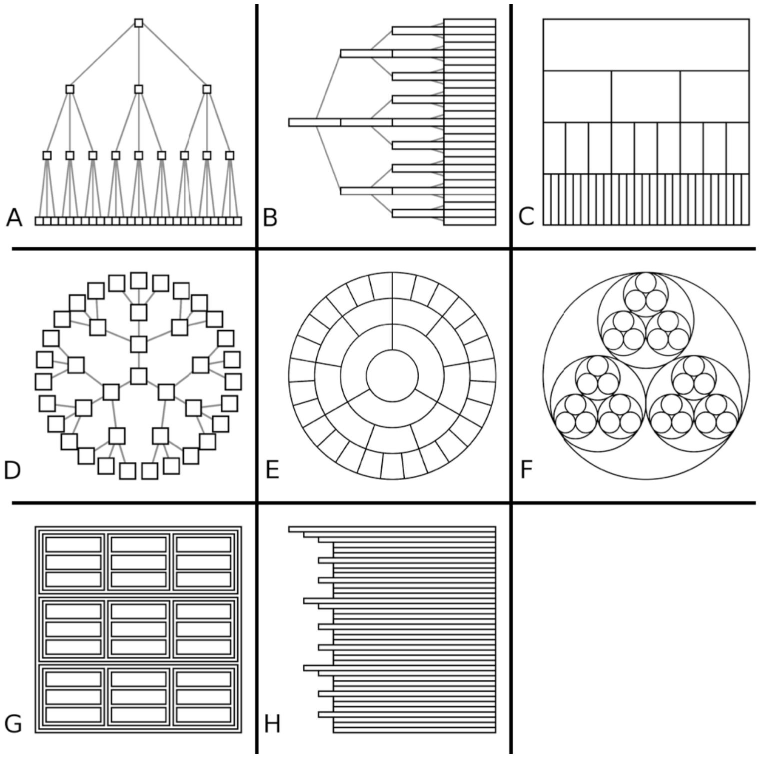


[Quantifying the Space-Efficiency of 2D Graphical Representations of Trees. McGuffin and Robert. Information Visualization 9:2 (2010), 115–140.]

Comparison: tree drawing idioms

data shown

- link relationships
- tree depth
- sibling order



[Quantifying the Space-Efficiency of 2D Graphical Representations of Trees. McGuffin and Robert. Information Visualization 9:2 (2010), 115–140.]

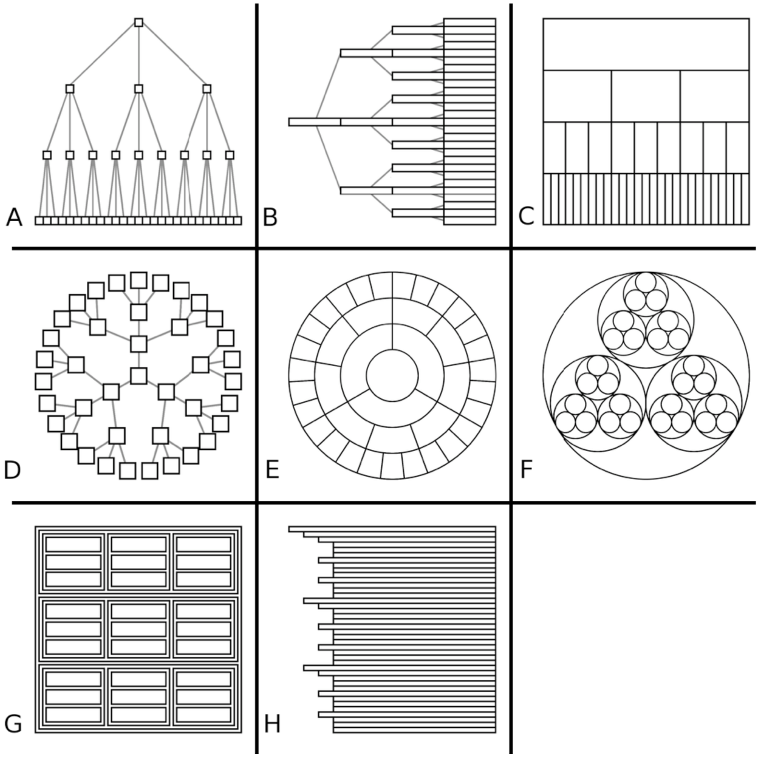
Comparison: tree drawing idioms

data shown

- link relationships
- tree depth
- sibling order

design choices

- connection vs containment link marks
- rectilinear vs radial layout
- spatial position channels



Comparison: tree drawing idioms

data shown

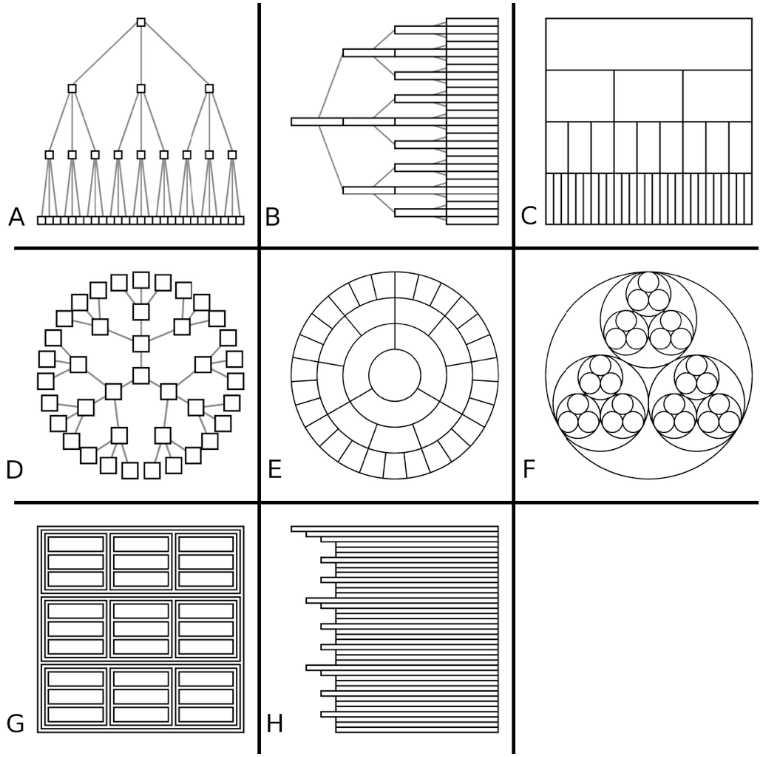
- link relationships
- tree depth
- sibling order

design choices

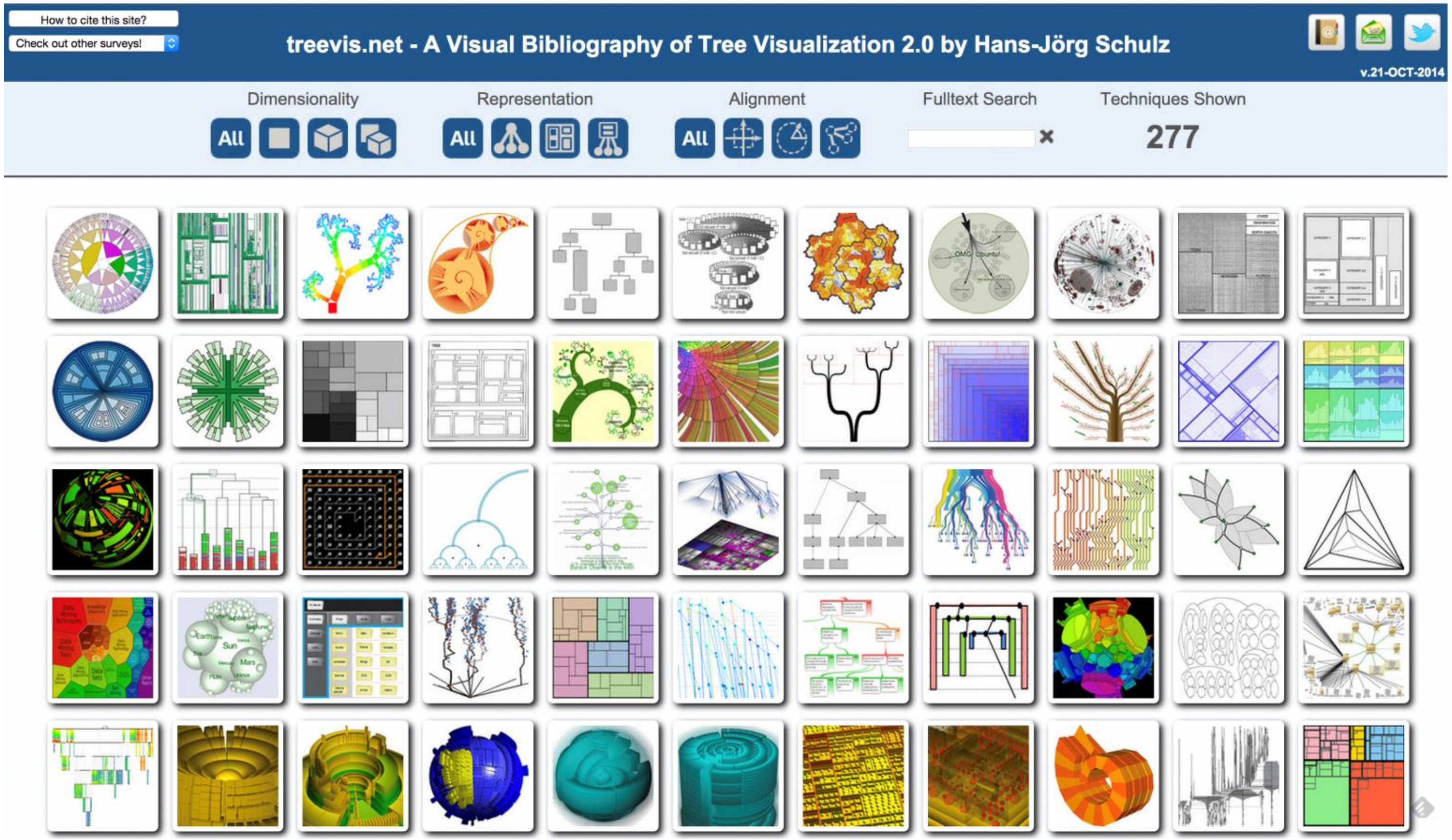
- connection vs containment link marks
- rectilinear vs radial layout
- spatial position channels

considerations

- redundant? arbitrary?
- information density?
  - avoid wasting space
  - consider where to fit labels!



treevis.net: Many, many options!

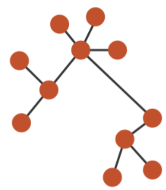


https://treevis.net/

Arrange networks and trees

Node–Link Diagrams
Connection Marks

✔ NETWORKS ✔ TREES



Spatial Position

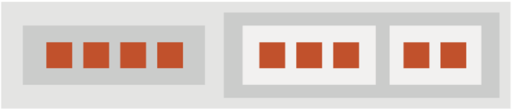
✘ NETWORKS ✔ TREES

Adjacency Matrix
Derived Table

✔ NETWORKS ✔ TREES

Enclosure
Containment Marks

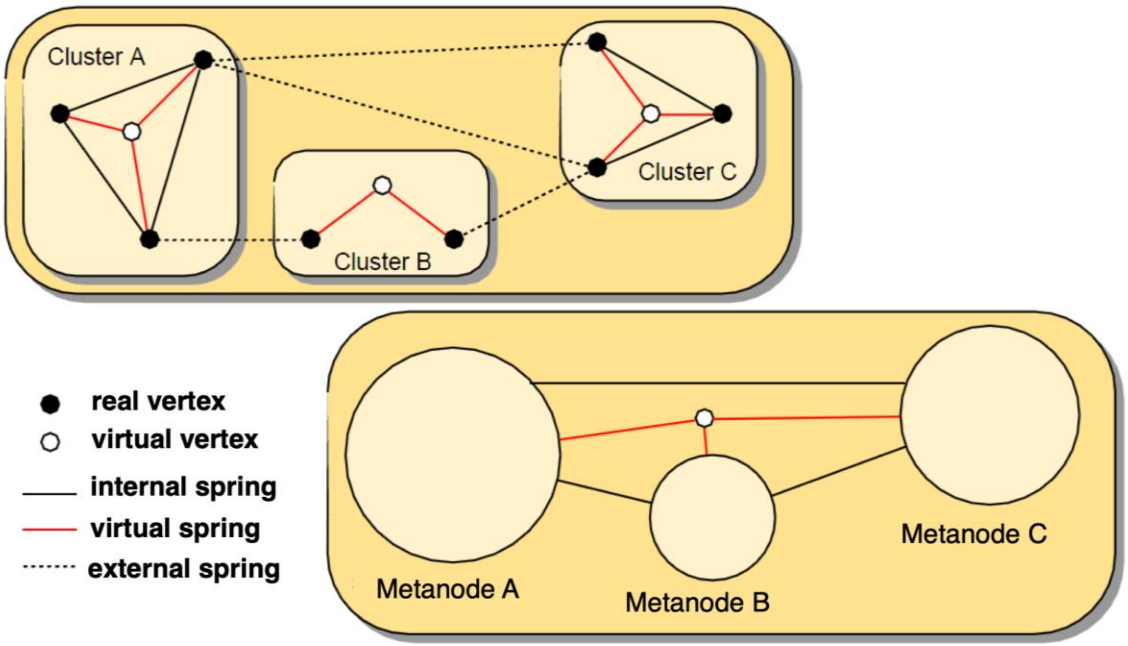
✘ NETWORKS ✔ TREES

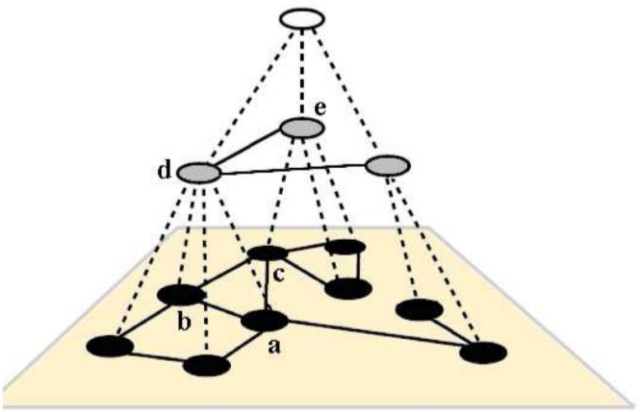


Network Data with Trees

Multilevel networks

derive cluster hierarchy of metanodes on top of original graph nodes





[Schulz 2004]

Idiom: **GrouseFlocks**

data: compound network

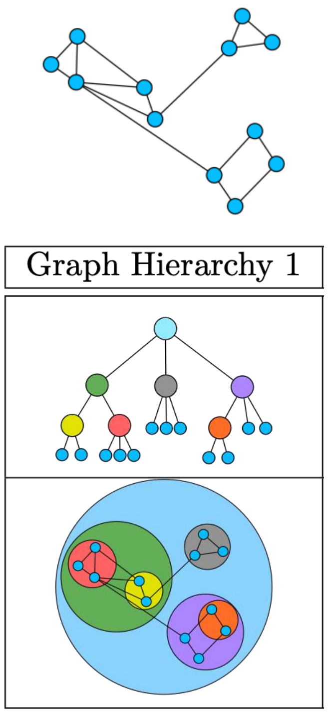
- network
- cluster hierarchy atop it
  - derived or interactively chosen

visual encoding

- connection marks for network links
- containment marks for hierarchy
- point marks for nodes

dynamic interaction

- select individual metanodes in hierarchy to expand/contract



GrouseFlocks: Steerable Exploration of Graph Hierarchy Space. Archambault, Munzner, Auber

Idiom: **sfdp** (multi-level force-directed placement)

data: compound graph

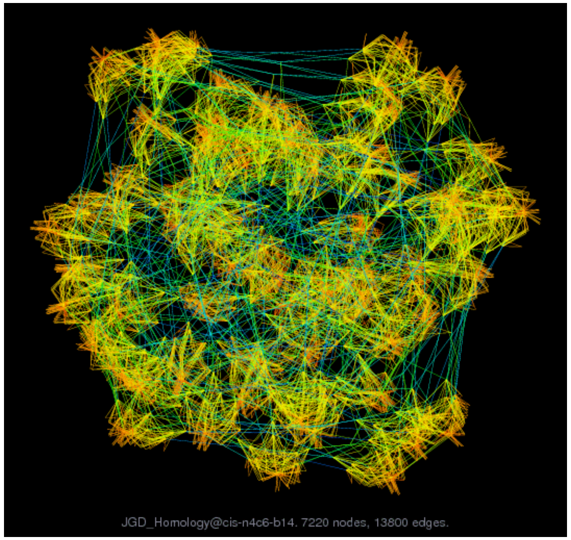
- original: network
- derived: cluster hierarchy atop it

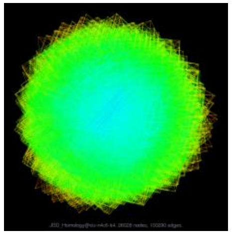
considerations

- better algorithm for same encoding technique
  - same: fundamental use of space
  - hierarchy used for algorithm speed/quality but not shown explicitly

scalability

- nodes, edges: 1K-10K
- hairball problem eventually hits





http://www.research.att.com/yifanhu/GALLERY/GRAPHS/index1.html

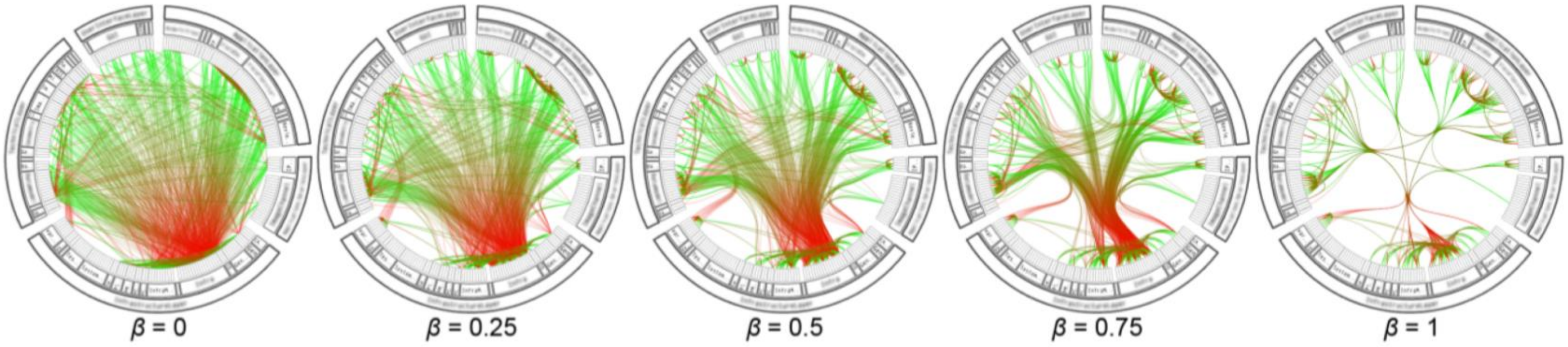
Idiom: **hierarchical edge bundling**

data

- any layout of compound network
  - network: software classes (nodes), import/export between classes (links)
  - cluster hierarchy: class package structure
- derived: bundles of edges with same source/destination (multi-level)

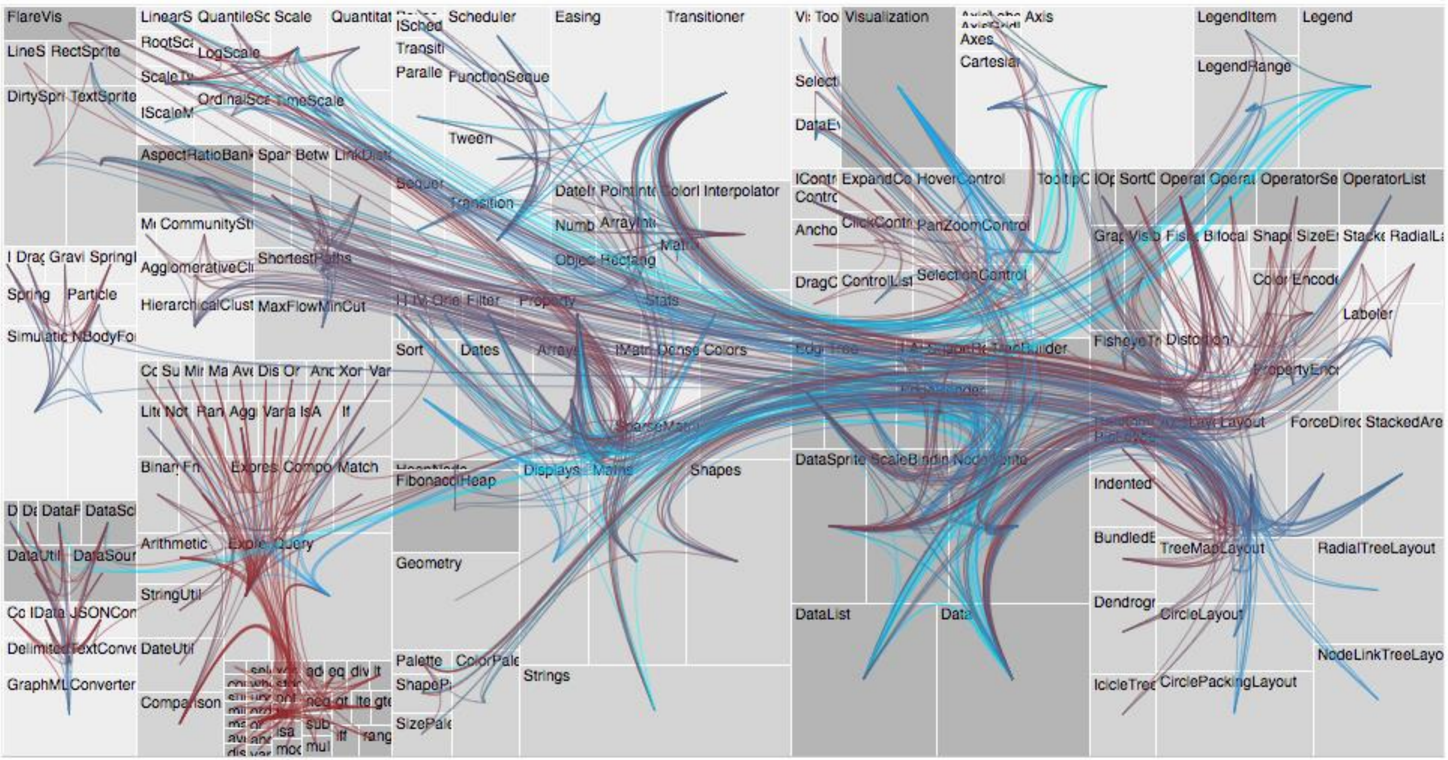
idiom: curve edge routes according to bundles

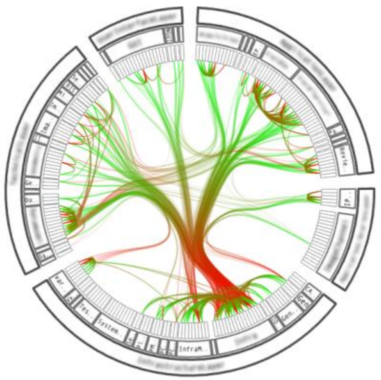
task: edge clutter reduction



Hierarchical edge bundling

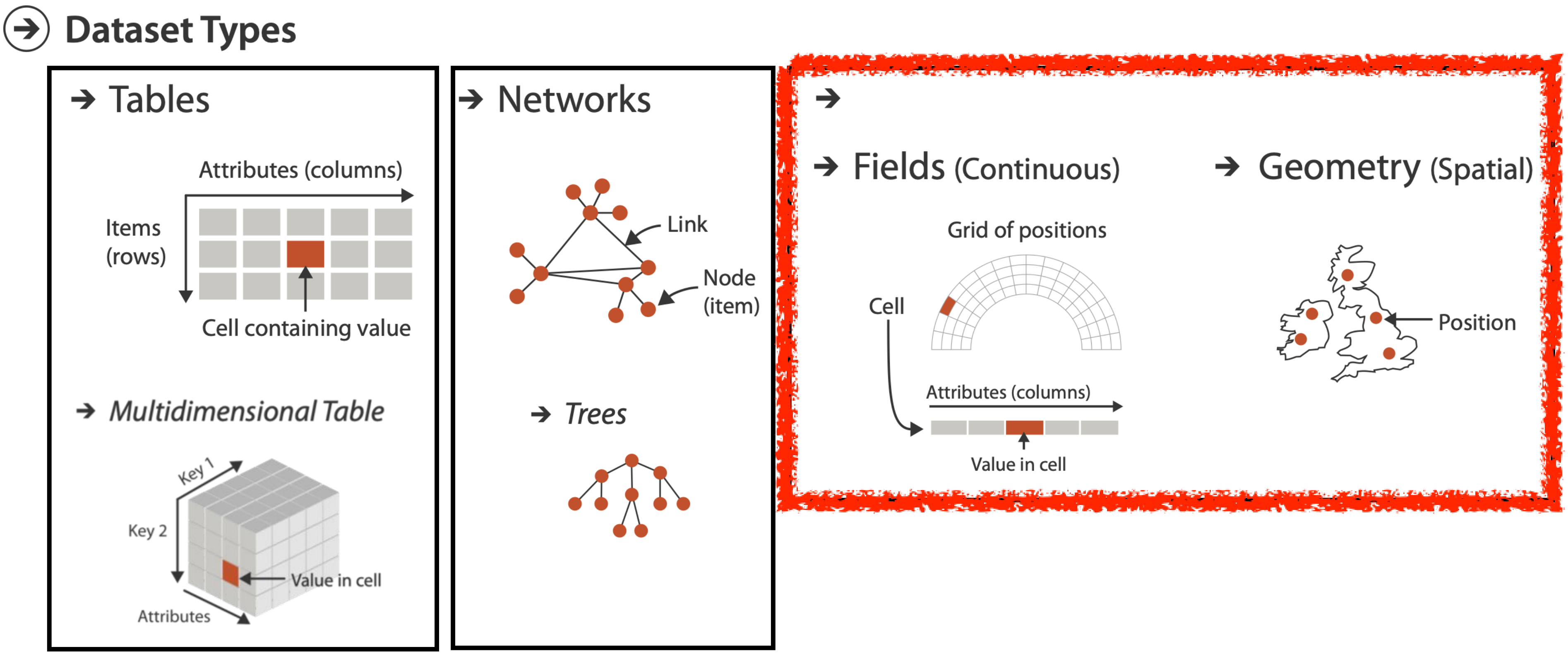
works for any layout: treemap vs radial

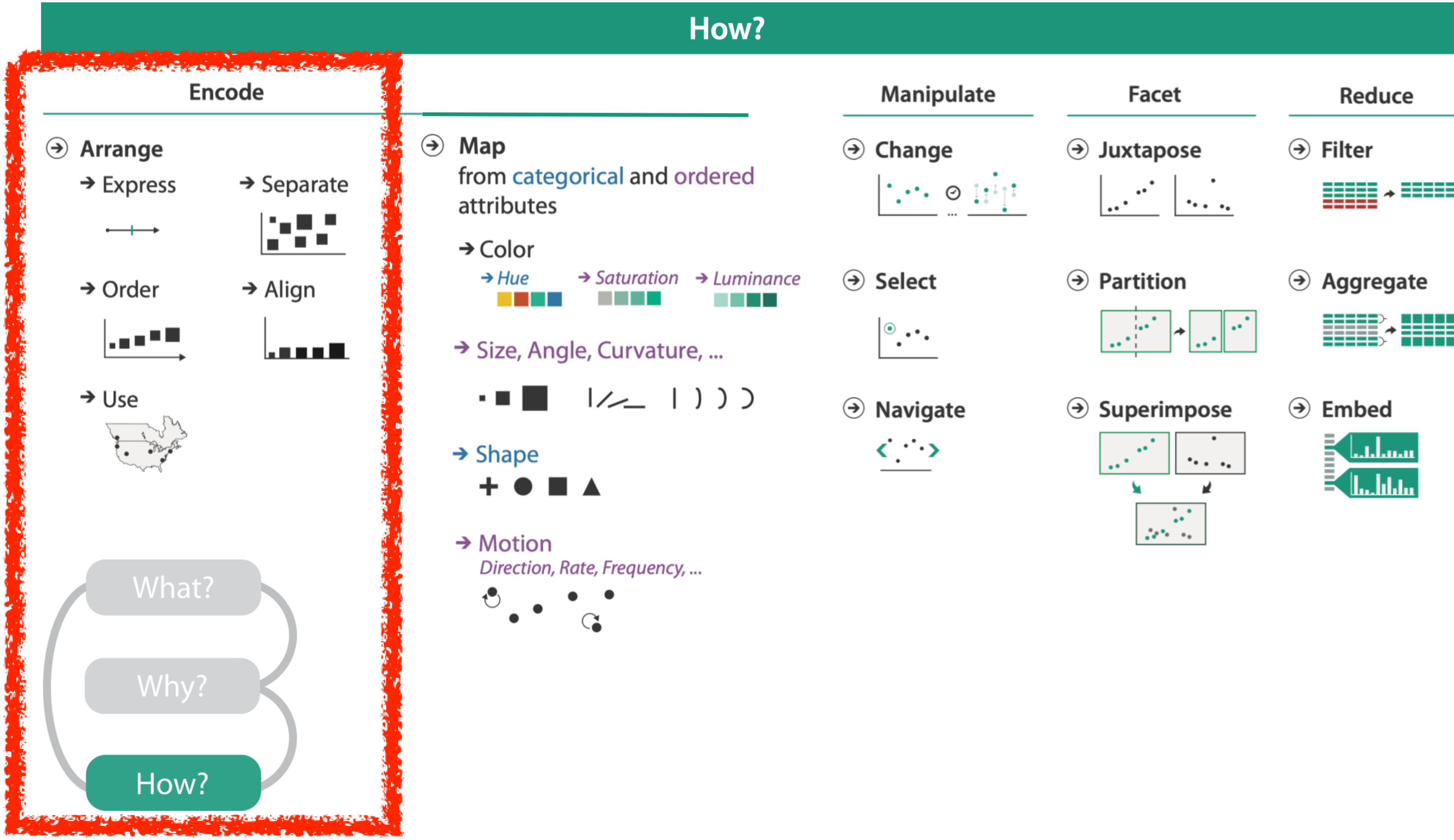




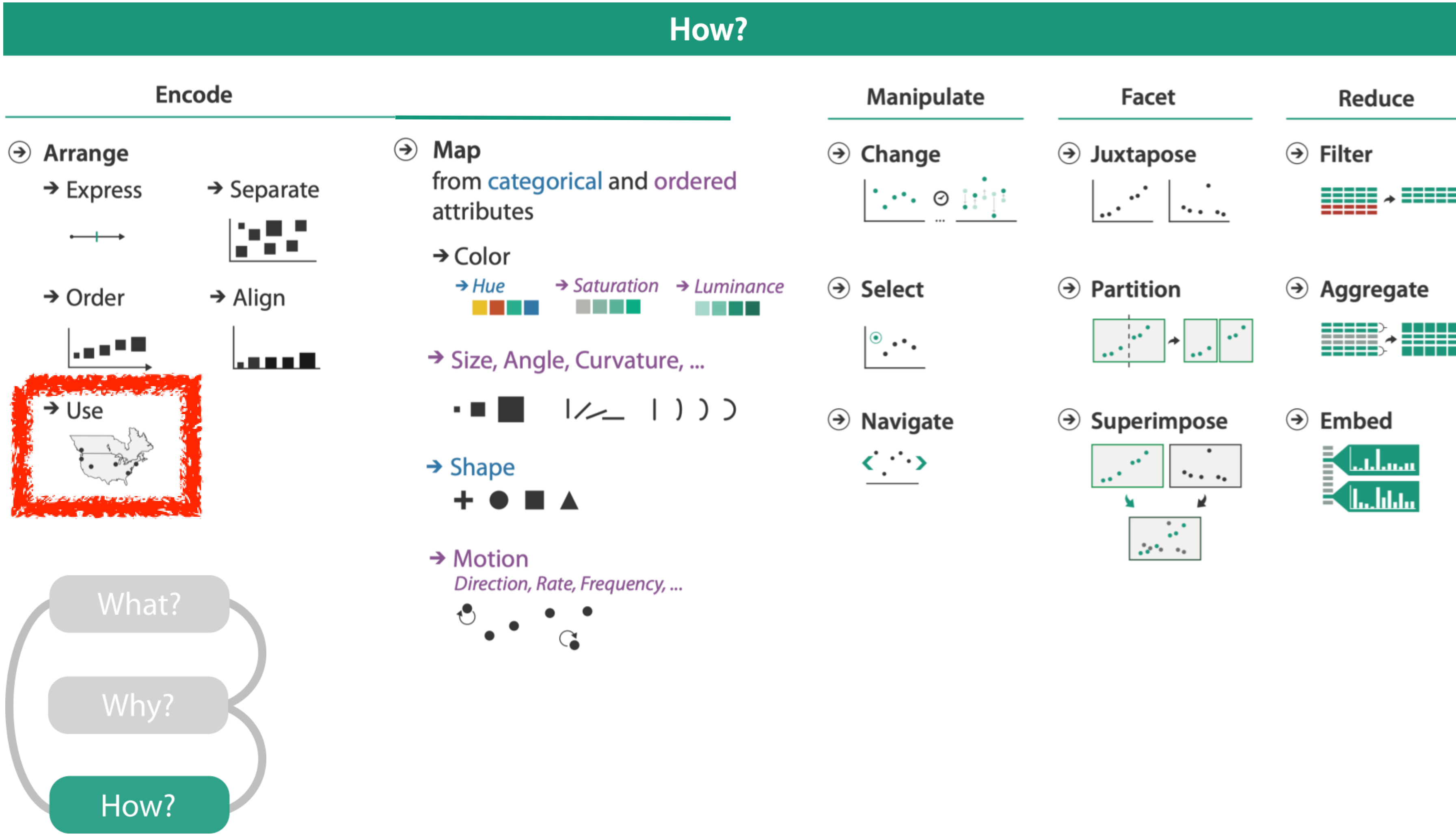
Spatial Data

Focus on Spatial





How?
Encode
Arrange
Express
Separate
Order
Align
Use
What?
Why?
How?
Map
from categorical and ordered attributes
Color
Hue
Saturation
Luminance
Size, Angle, Curvature, ...
Shape
Motion
Direction, Rate, Frequency, ...
Manipulate
Change
Select
Navigate
Facet
Juxtapose
Partition
Superimpose
Reduce
Filter
Aggregate
Embed


How?
Encode
Arrange
Express
Separate
Order
Align
Use
Map
from categorical and ordered attributes
Color
Hue
Saturation
Luminance
Size, Angle, Curvature, ...
Shape
Motion
Direction, Rate, Frequency, ...
Manipulate
Change
Select
Navigate
Facet
Juxtapose
Partition
Superimpose
Reduce
Filter
Aggregate
Embed
What?
Why?
How?

Spatial data

use given spatial position

when?

- dataset contains spatial attributes and they have primary importance
- central tasks revolve around understanding spatial relationships

examples

- geographical/cartographic data
- sensor/simulation data

Geographic Maps

Geographic Map

Interlocking marks

shape coded

area coded

position coded

- cannot encode another attribute with these channels, they're "taken"

Thematic maps

show spatial variability of attribute ("theme")

- combine geographic / reference map with (simple, flat) tabular data
- join together
  - region: interlocking area marks (provinces, countries with outline shapes)
    - also could have point marks (cities, locations with 2D lat/lon coords)
  - region: categorical key attribute in table
    - use to look up value attributes

major idioms

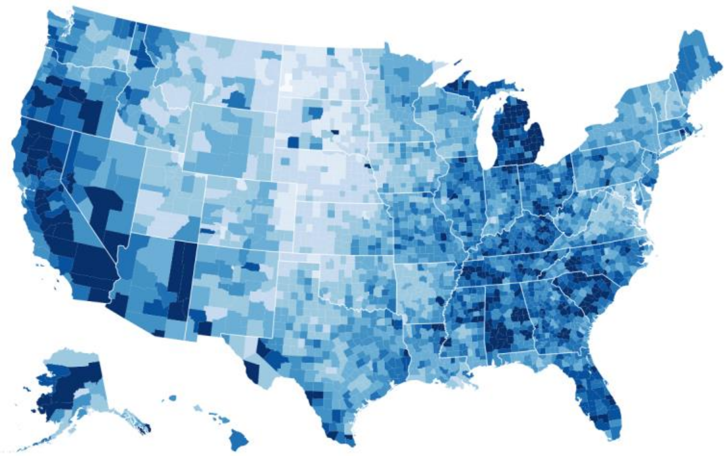
- choropleth
- symbol maps
- cartograms
- dot density maps

Idiom: **choropleth map**

use given spatial data
- when central task is understanding spatial relationships

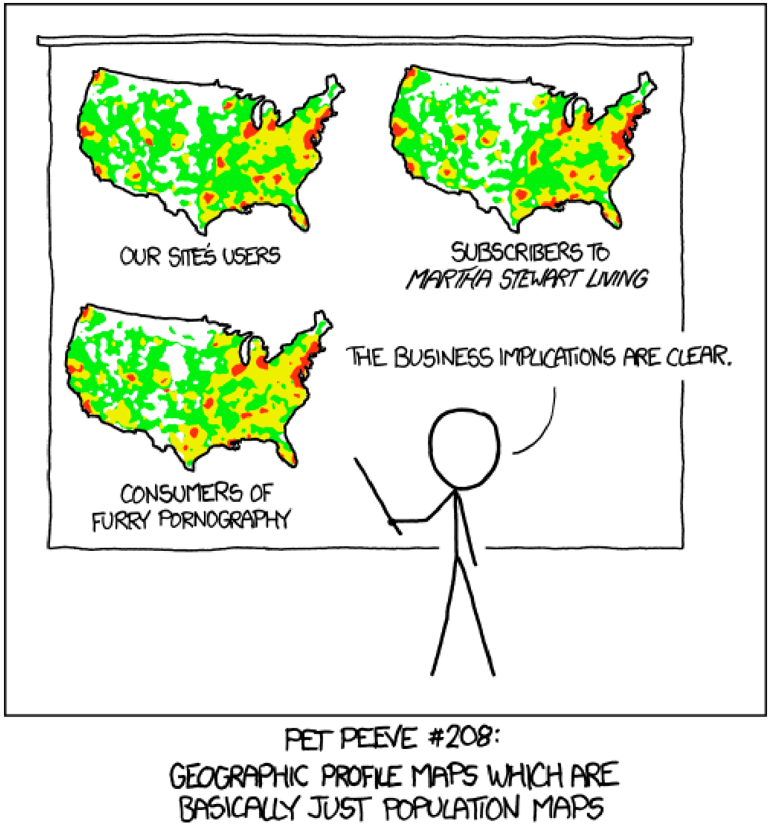
data
- geographic geometry
- table with 1 quant attribute per region

encoding
- position:
  use given geometry for area mark boundaries
- color:
  sequential segmented colormap



http://bl.ocks.org/mbostock/4060606

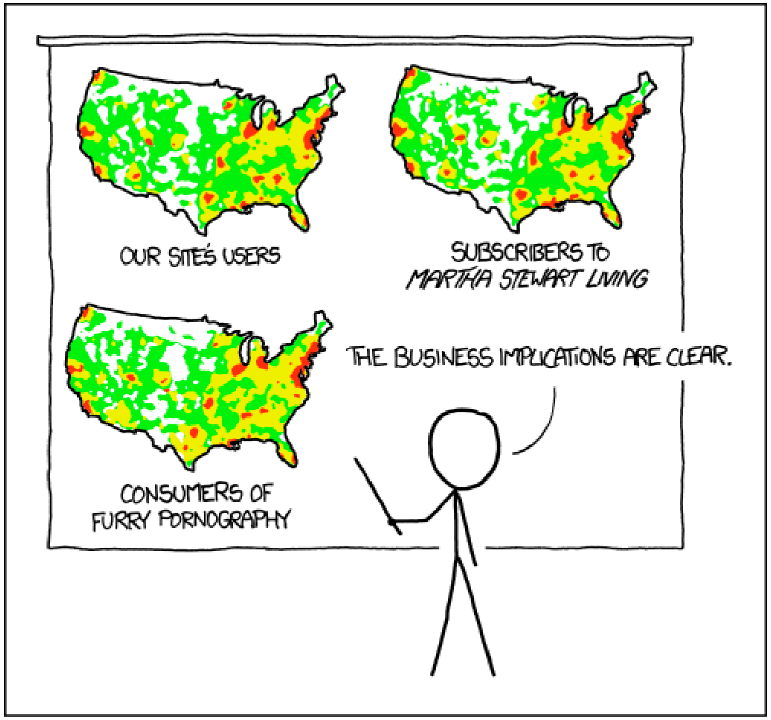
Beware: Population maps trickiness!



[https://xkcd.com/1138]

Beware: Population maps trickiness!

spurious correlations: most attributes just show where people live

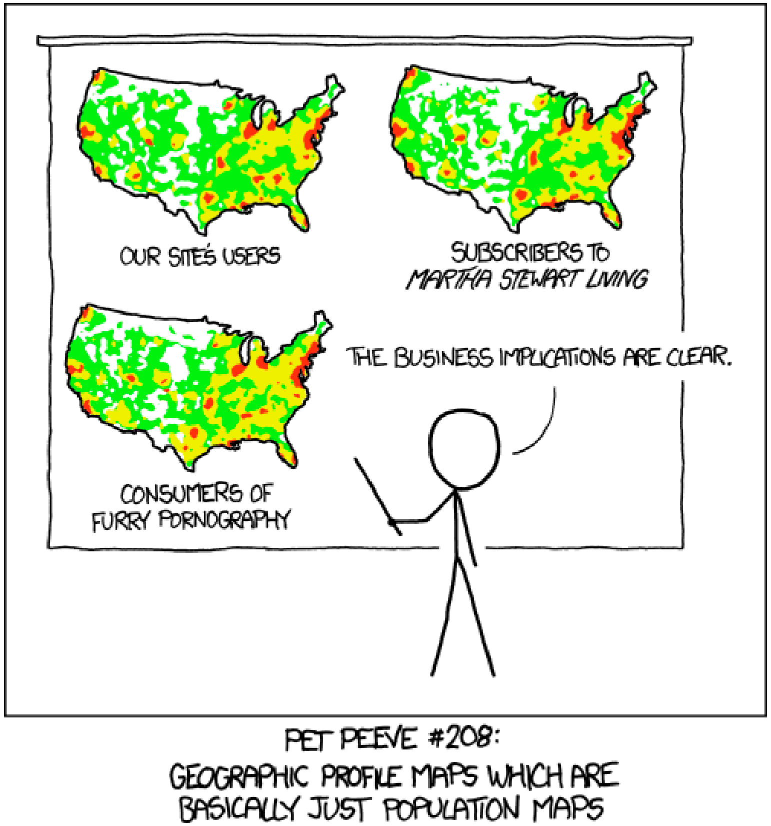


PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

[https://xkcd.com/1138]

Beware: Population maps trickiness!

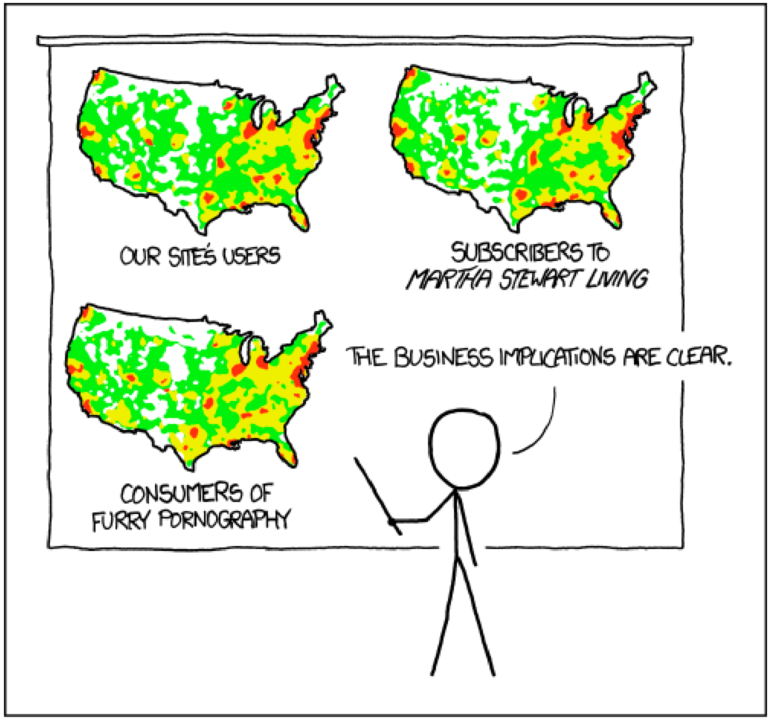
- spurious correlations: most attributes just show where people live
- consider when to normalize by population density
  - encode raw data values
    - tied to underlying population
  - but should use normalized values
    - unemployed people per 100 citizens, mean family income



[https://xkcd.com/1138]

Beware: Population maps trickiness!

- spurious correlations: most attributes just show where people live
- consider when to normalize by population density
  - encode raw data values
    - tied to underlying population
  - but should use normalized values
    - unemployed people per 100 citizens, mean family income
- general issue
  - absolute counts vs relative/normalized data
  - failure to normalize is common error



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE BASICALLY JUST POPULATION MAPS

[https://xkcd.com/1138]

Choropleth maps: Recommendations

only use when central task is understanding spatial relationships

show only one variable at a time

normalize when appropriate

be careful when choosing colors & bins

best case: regions are roughly equal sized

Choropleth map: Pros & cons

pros

- easy to read and understand
- well established visualization (no learning curve)
- data is often collected and aggregated by geographical regions

cons

- most effective visual variable used for geographic location
- visual salience depends on region size, not true importance wrt attribute va
  - large regions appear more important than small ones
- color palette choice has a huge influence on the result

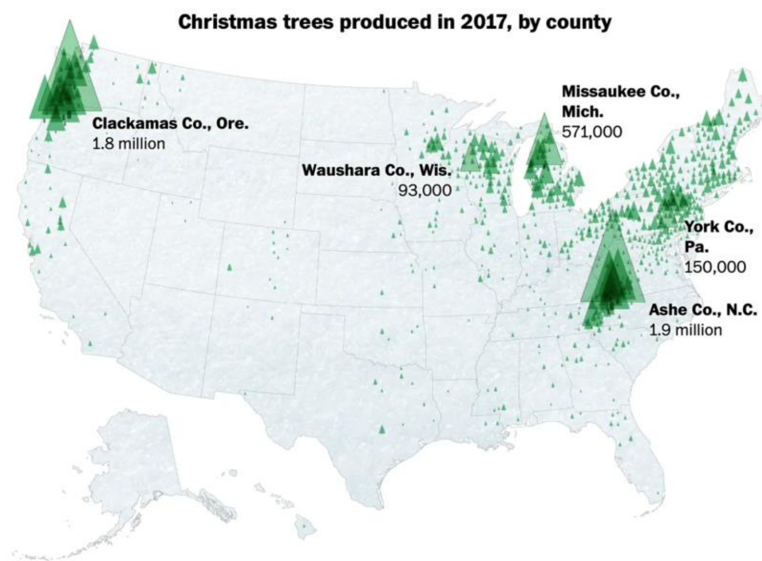
Idiom: **Symbol maps**

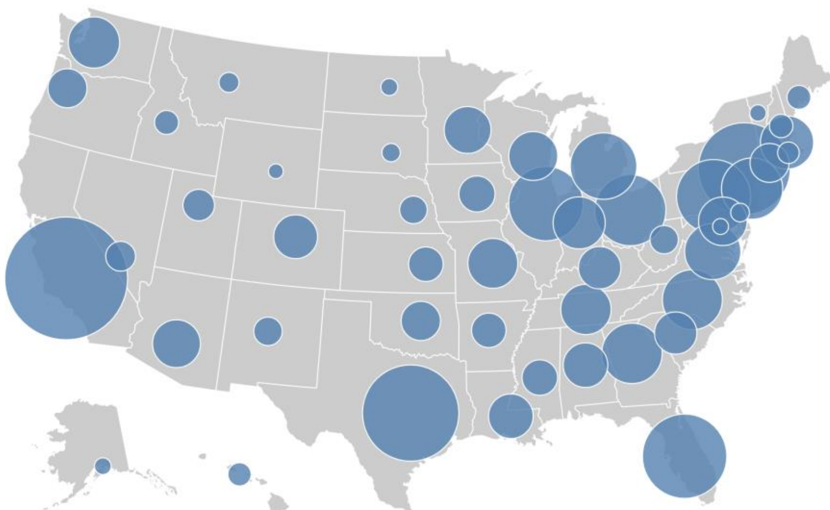
symbol is used to represent aggregated data (mark or glyph)

- allows use of size and shape and color channels
  - aka proportional symbol maps, graduated symbol m

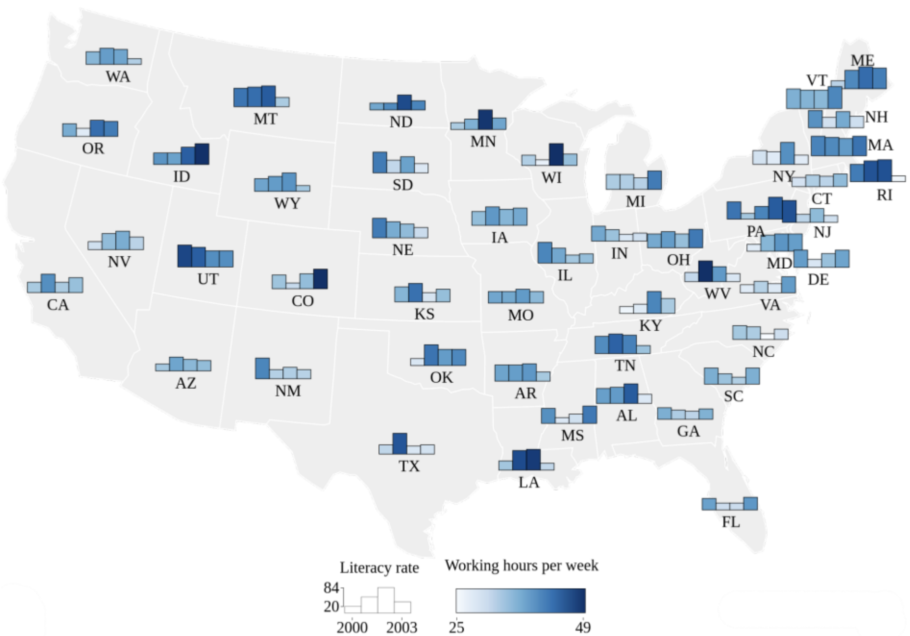
keep original spatial geometry in the backg

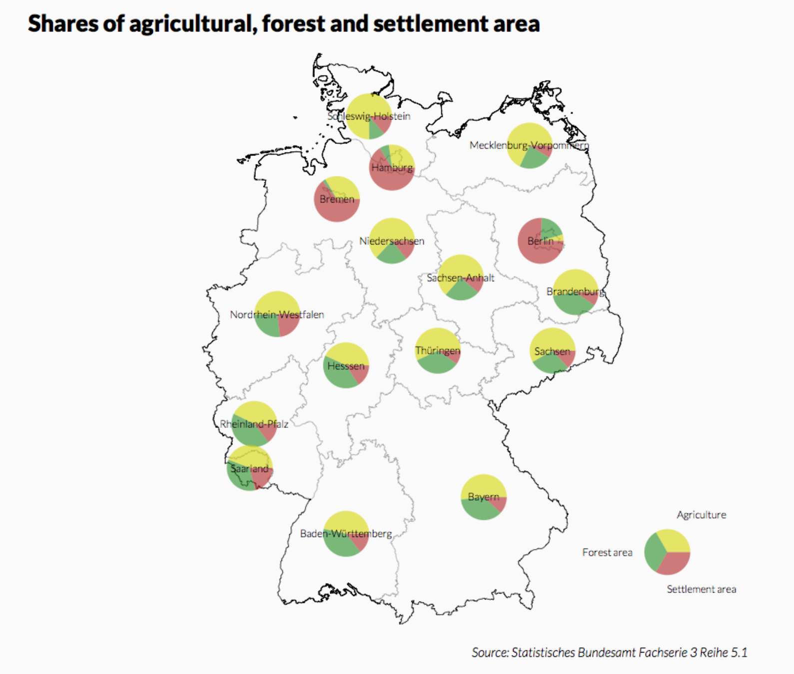
often a good alternative to choropleth map





Symbol maps with glyphs





Symbol map: Pros & cons

pros

- somewhat intuitive to read and understand
- mitigate problems with region size vs data salience
  - marks: symbol size follows attribute value
  - glyphs: symbol size can be uniform

cons

- possible occlusion / overlap
  - symbols could overlap each other
  - symbols could occlude region boundaries
- complex glyphs may require explanation / training

Idiom: **Contiguous cartogram**

interlocking marks:
shape, area, and position coded

derive new interlocking marks

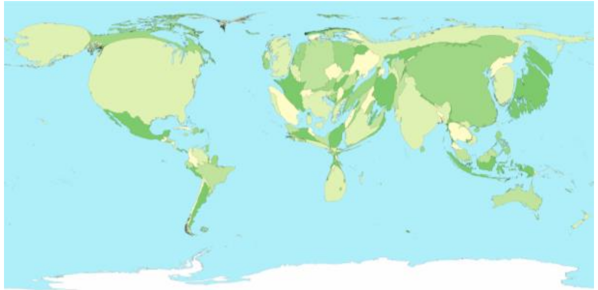
- based on combination of original interlocking marks and new quantitative attribute

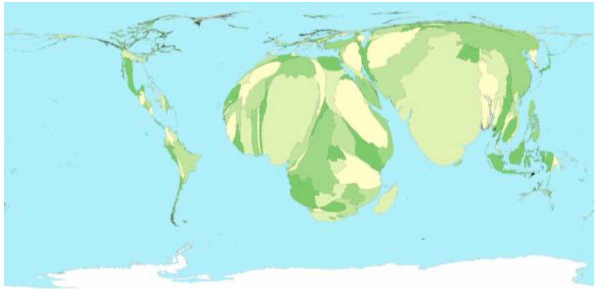
algorithm to create new marks

- input: target size
- goal: shape as close to the original as possible
- requirement: maintain constraints

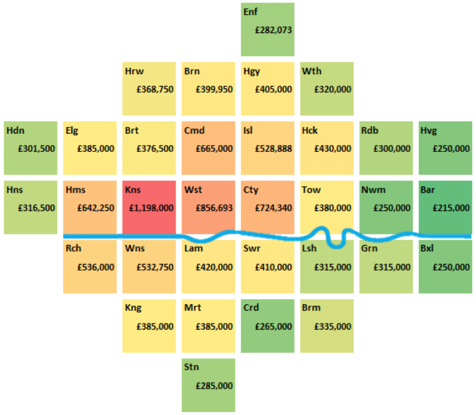
relative position

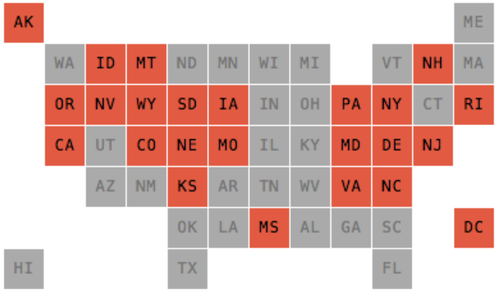
contiguous boundaries with their neighbors

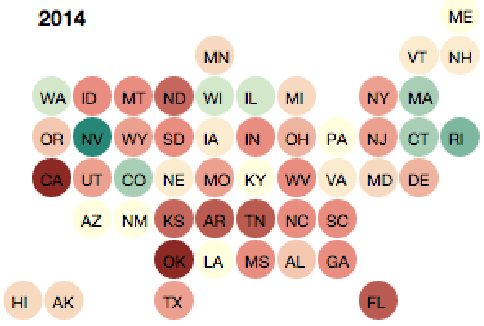




Idiom: **Grid Cartogram**







uniform-sized shapes arranged in rectilinear grid

maintain approximate spatial position and arrangement

Cartogram: Pros & cons

pros

- can be intriguing and engaging
- best case: strong and surprising size disparities
- non-contiguous cartograms often easier to understand

cons

- require substantial familiarity with original dataset & use of memory
  - compare distorted marks to memory of original marks
  - mitigation strategies: transitions or side by side views
- major distortion is problematic
  - may be aesthetically displeasing
  - may result in unrecognizable marks
- difficult to extract exact quantities

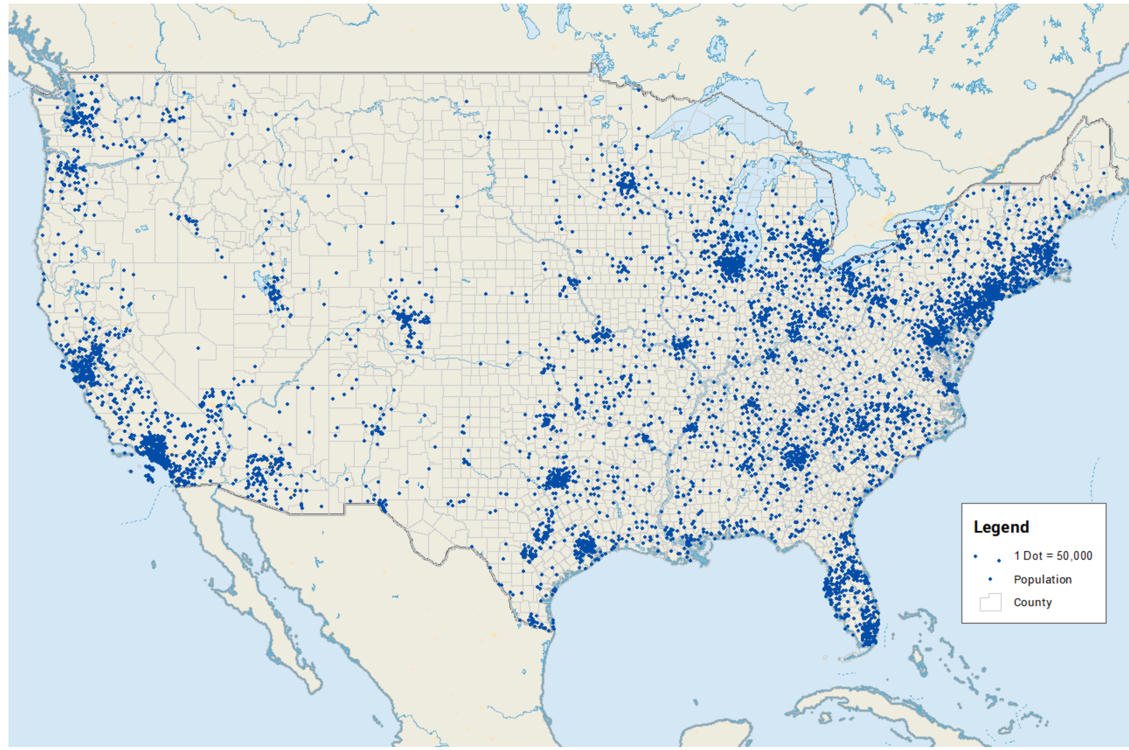
Idiom: **Dot density maps**

visualize distribution of a phenomenon by placing dots

one symbol represents a constant number of items

- dots have uniform size & shape
- allows use of color channel

task: show spatial patterns, clusters



Dot density maps: Pros and cons

pros

- straightforward to understand
- avoids choropleth non-uniform region size problems

cons

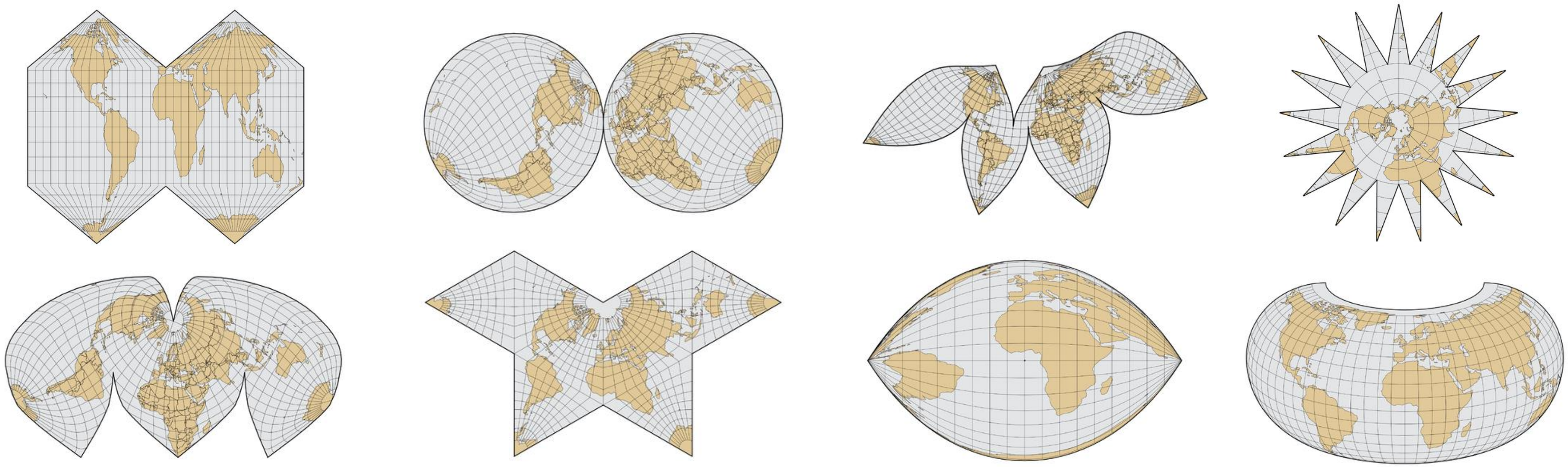
- challenge: normalization, just like choropleths
  - show population density (correlated with attribute), not effect of interest
- perceptual disadvantage:
  difficult to extract quantities
- performance disadvantage:
  rendering many dots can be slow

Map Projections

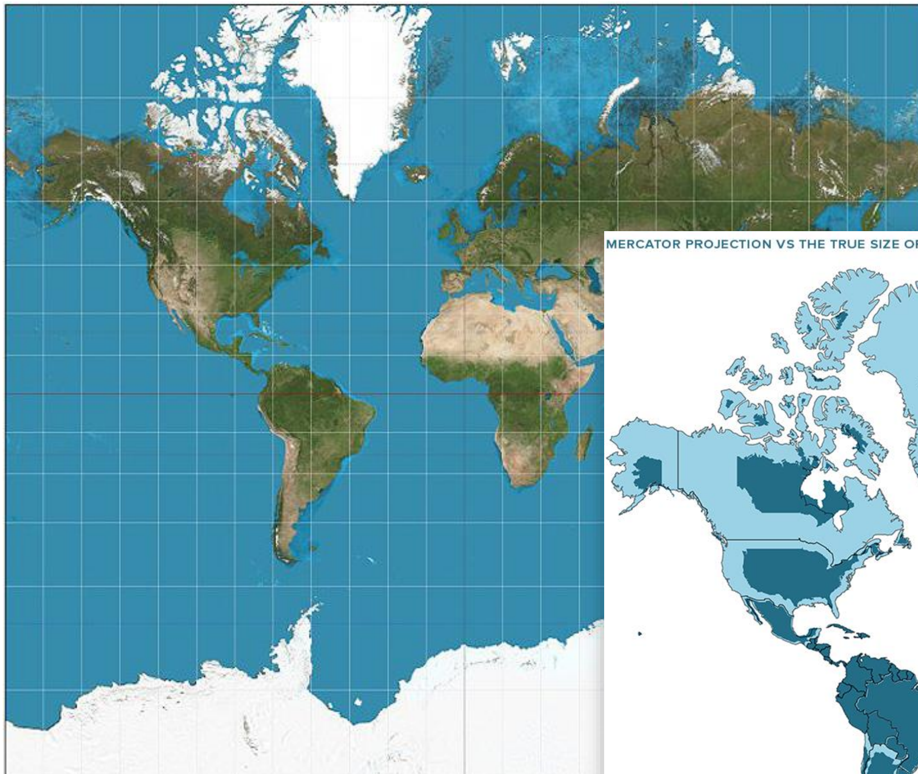
mathematical functions that map 3D surface geometry of the Earth to 2D maps

all projections of sphere on plane necessarily distort surface in some way

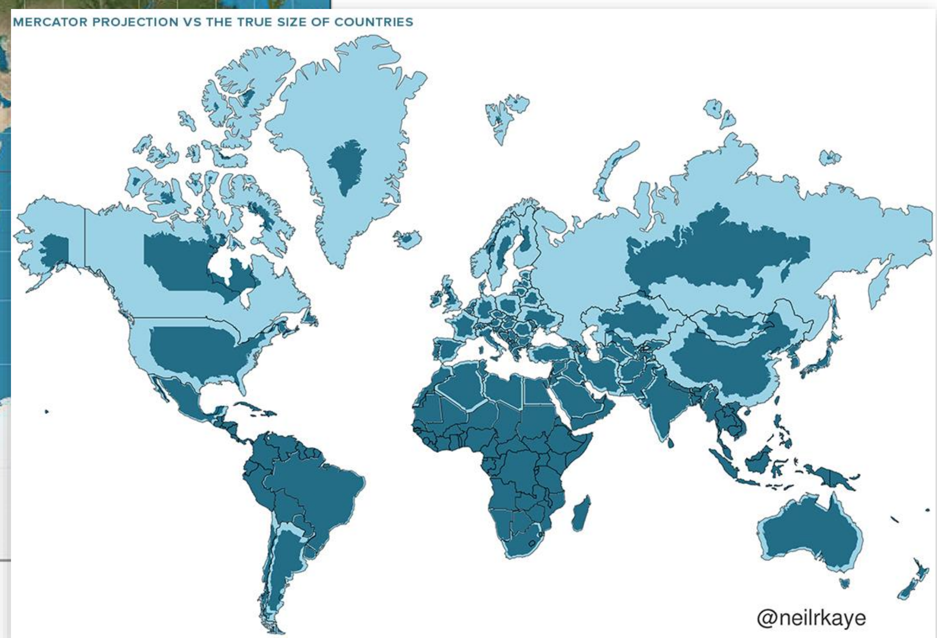
interactive: philogb.github.io/page/myriahedral/ and jasondavies.com/maps/



Mercator Projection

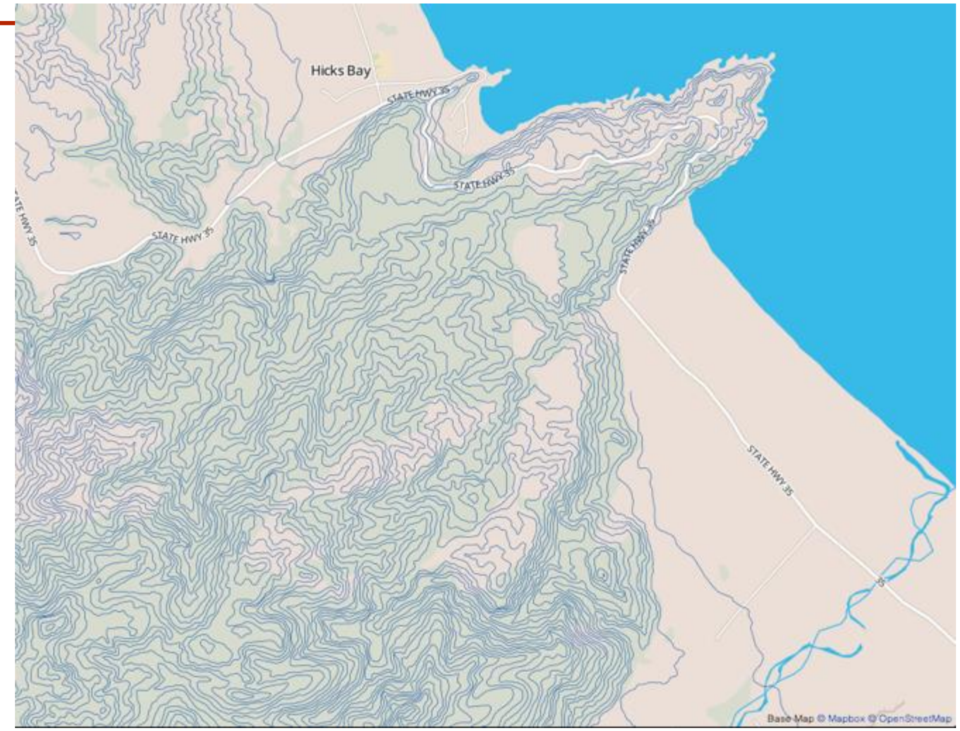


» Heavily distorts country sizes; particularly close to the poles.



Spatial Fields

Idiom: **topographic map**



Land Information New Zealand Data Service

data

- geographic geometry
- scalar spatial field
  - 1 quant attribute per grid cell

derived data

- isoline geometry
  - isocontours computed for specific levels of scalar values

task

- understanding terrain shape
  - densely lined regions = steep

pros

- use only 2D position, avoid 3D challenges
- color channel available for other attributes

cons

- significant clutter from additional lines

Idioms: **isosurfaces, direct volume rendering**

data

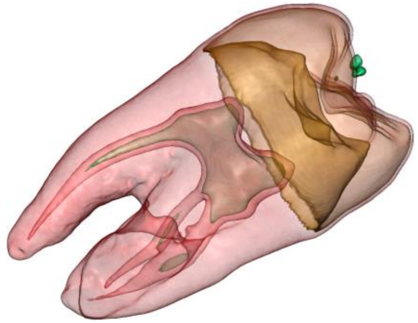
- scalar spatial field (3D volume)
  - 1 quant attribute per grid cell

task

- shape understanding, spatial relationships

[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]
Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]

Idioms: **isosurfaces, direct volume rendering**



data

- scalar spatial field (3D volume)
  - 1 quant attribute per grid cell

task

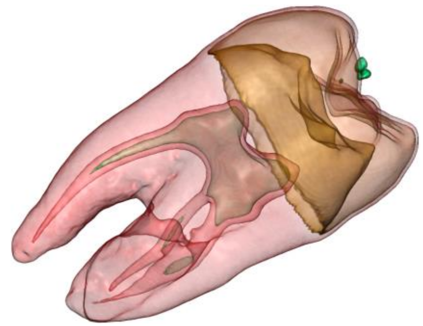
- shape understanding, spatial relationships

isosurface

- derived data: isocontours computed for specific levels of scalar values

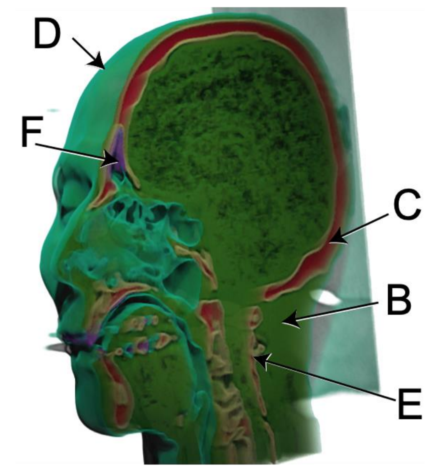
[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]
Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]

Idioms: **isosurfaces, direct volume rendering**



data
- scalar spatial field (3D volume)
  - 1 quant attribute per grid cell

task
- shape understanding, spatial relationships



isosurface
- derived data: isocontours computed for specific levels of scalar values

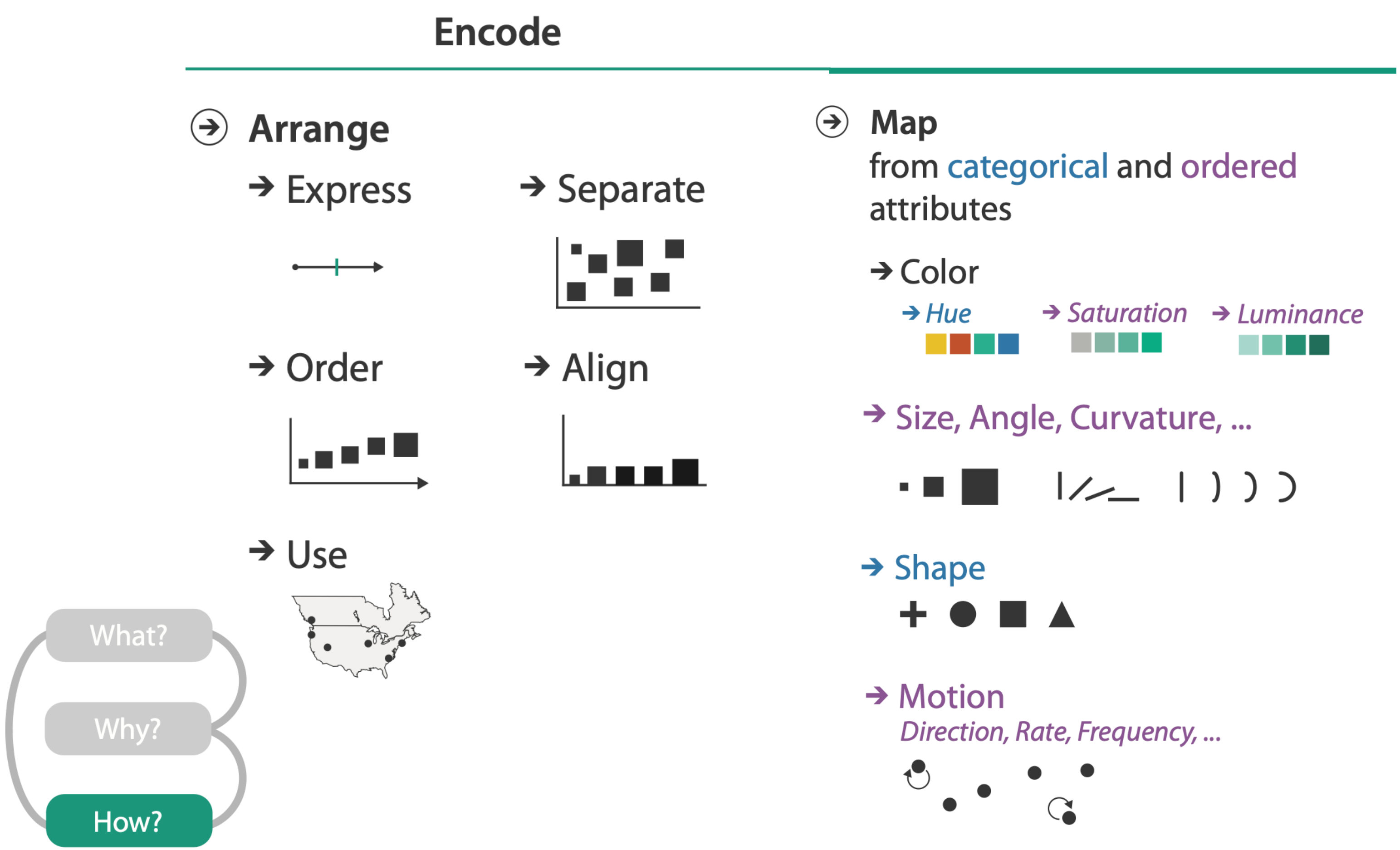
direct volume rendering
- transfer function maps scalar values to color, opacity
  - no derived geometry

[Interactive Volume Rendering Techniques. Kniss. Master's thesis, University of Utah Computer Science, 2002.]

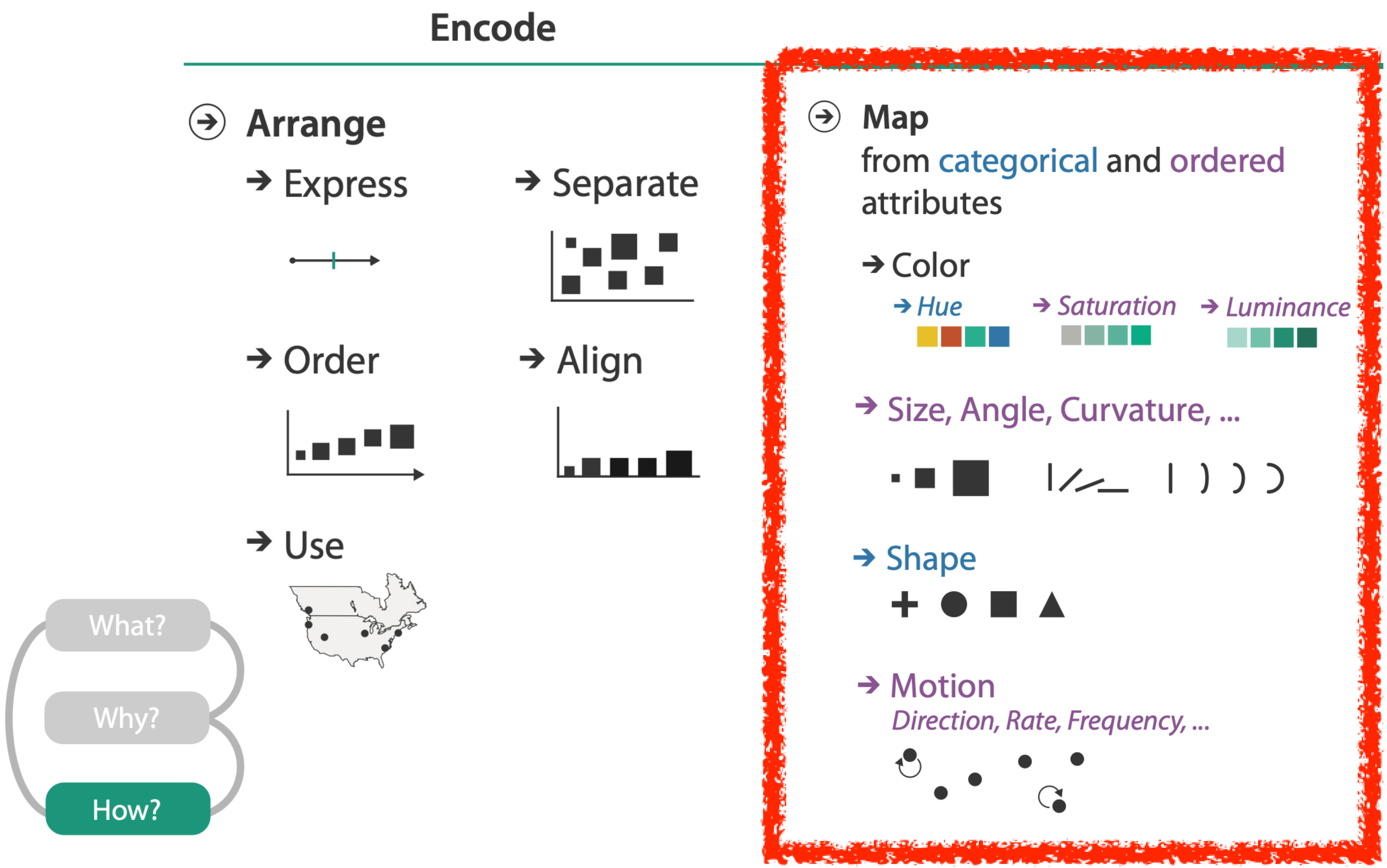
Multidimensional Transfer Functions for Volume Rendering. Kniss, Kindlmann, and Hansen. In The Visualization Handbook, edited by Charles Hansen and Christopher Johnson, pp. 189–210. Elsevier, 2005.]

Color

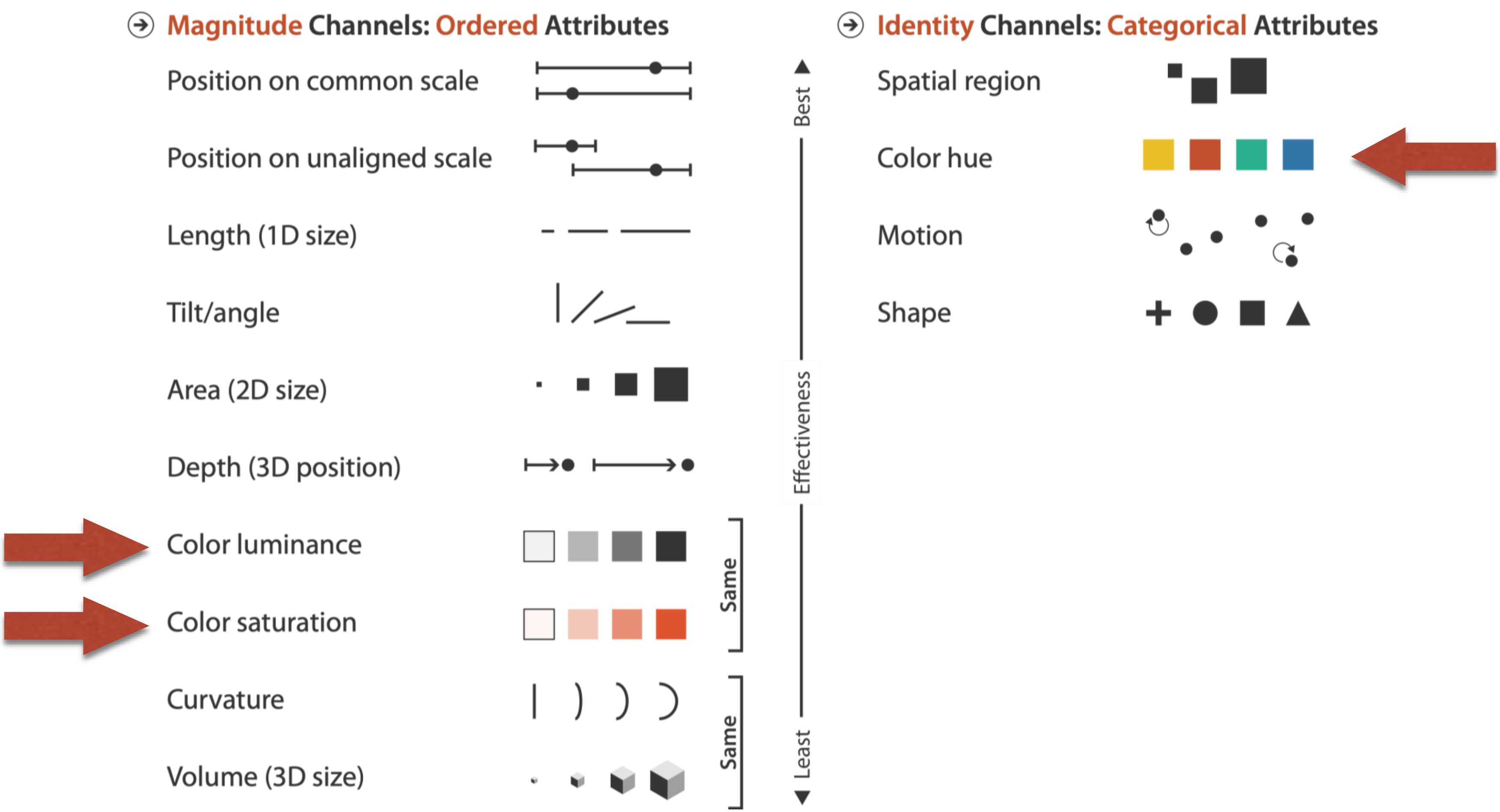
Idiom design choices: Visual encoding



Idiom design choices: Beyond spatial arrangement



Channels: What's up with color?



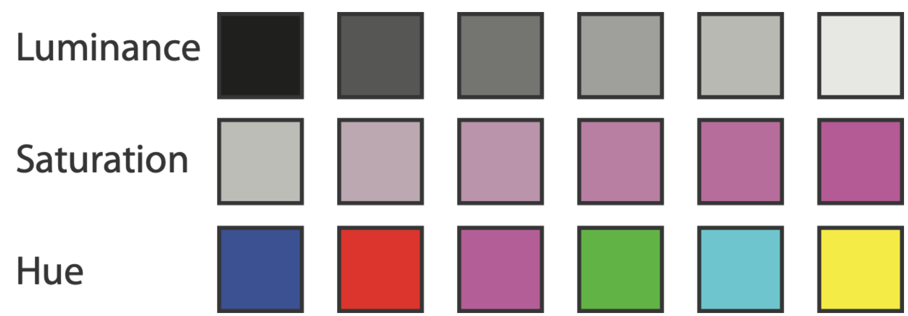
Decomposing color

first rule of color: do not (just) talk about color!

- color is confusing if treated as monolithic

decompose into three channels

- ordered can show magnitude
  - **luminance**: how bright (B/W)
  - **saturation**: how colourful
- categorical can show identity
  - **hue**: what color

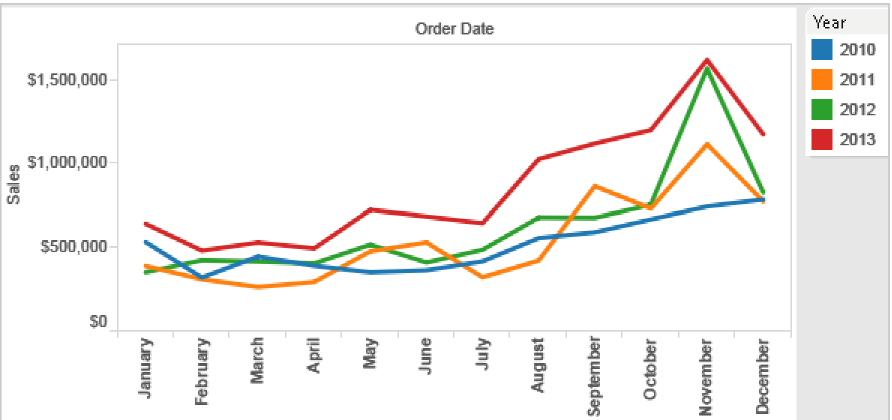


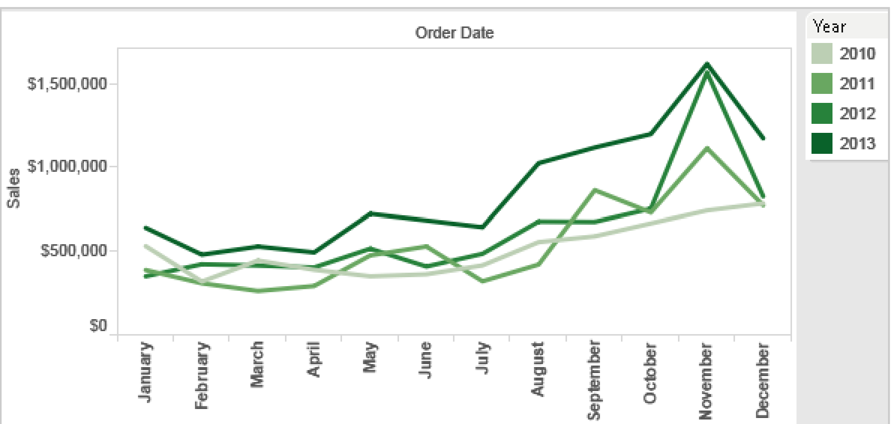
channels have different properties

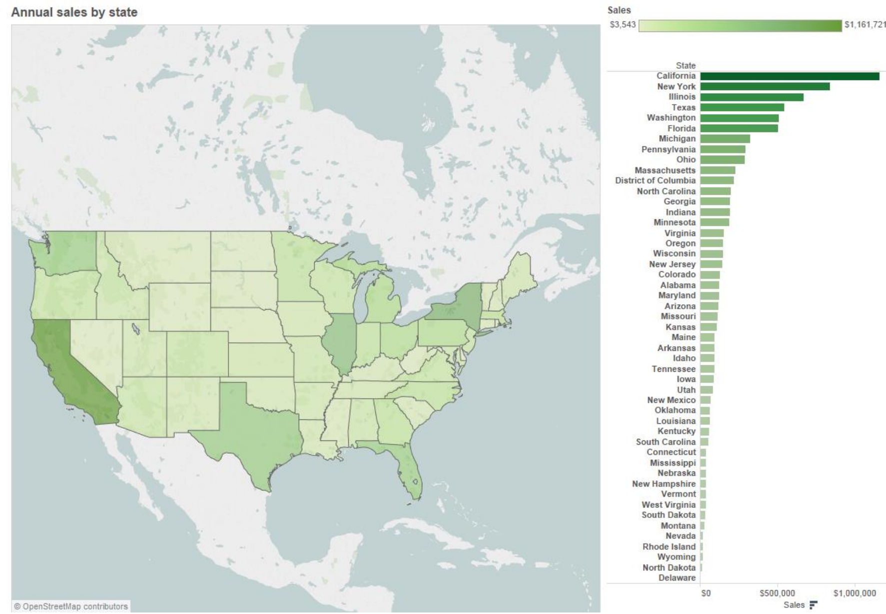
- what they convey directly to perceptual system
- how much they can convey
  - how many discriminable bins can we use?

Color Channels in Visualization

Categorical vs ordered color

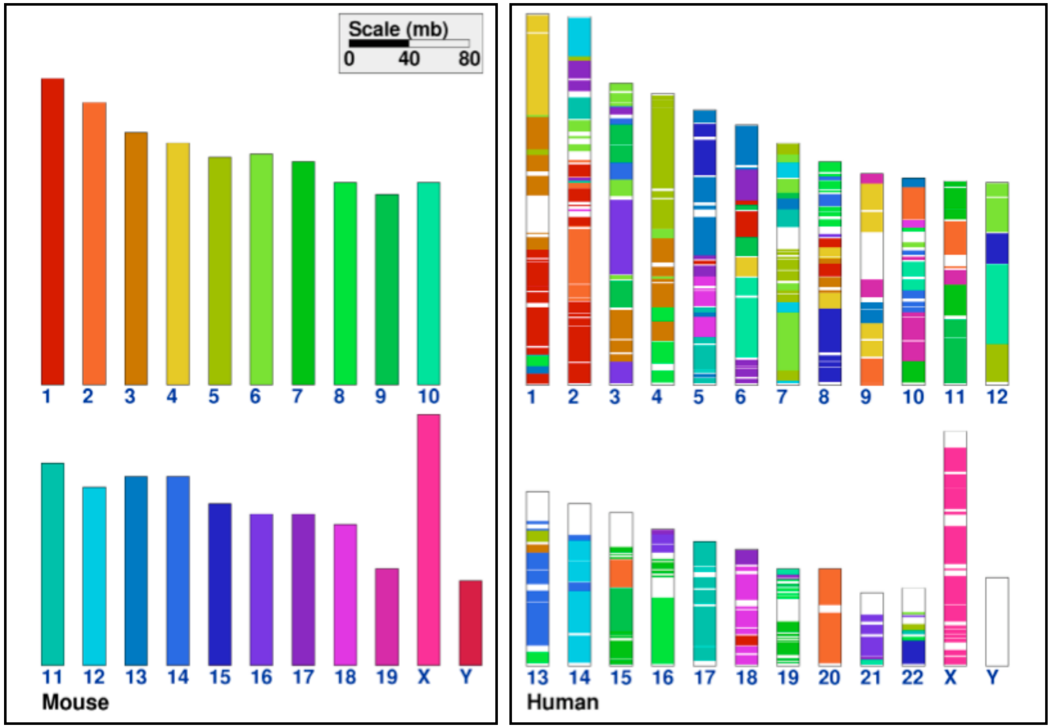






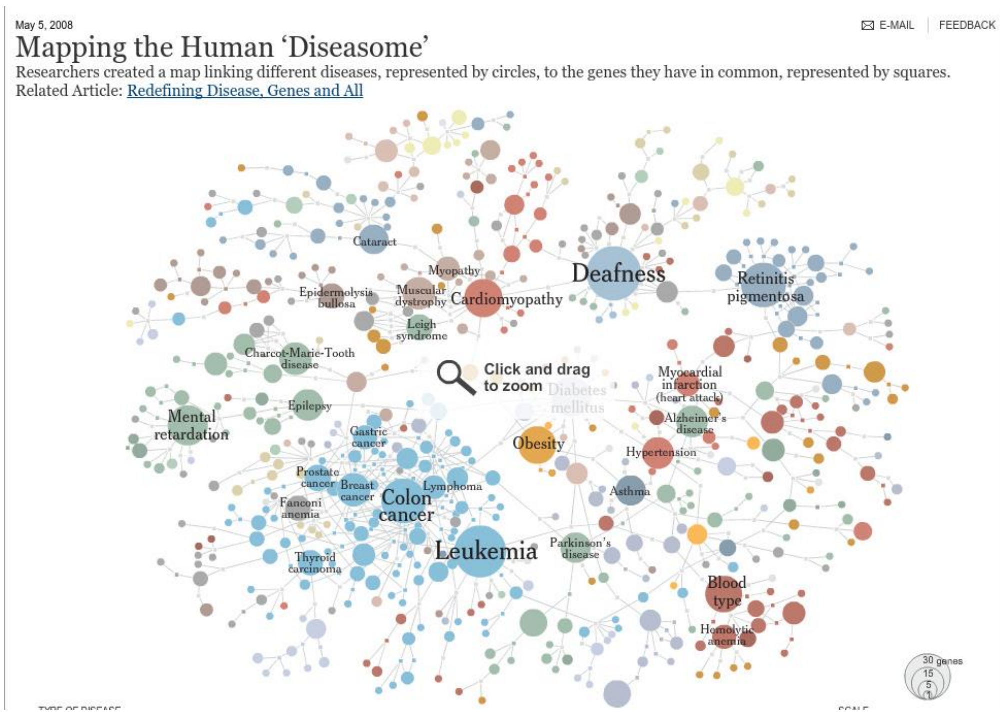
Categorical color: limited number of discriminable bins

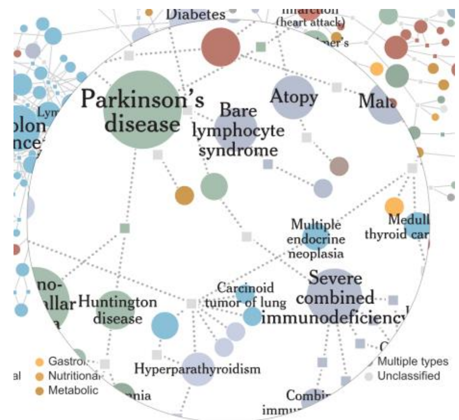
- human perception built on relative comparisons
  - great if color contiguous
  - surprisingly bad for absolute comparisons
- noncontiguous small regions of color
  - fewer bins than you want
  - rule of thumb: 6-12 bins, including background and highlights



[Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. Sinha and Meller. BMC Bioinformatics, 8:82, 2007.]

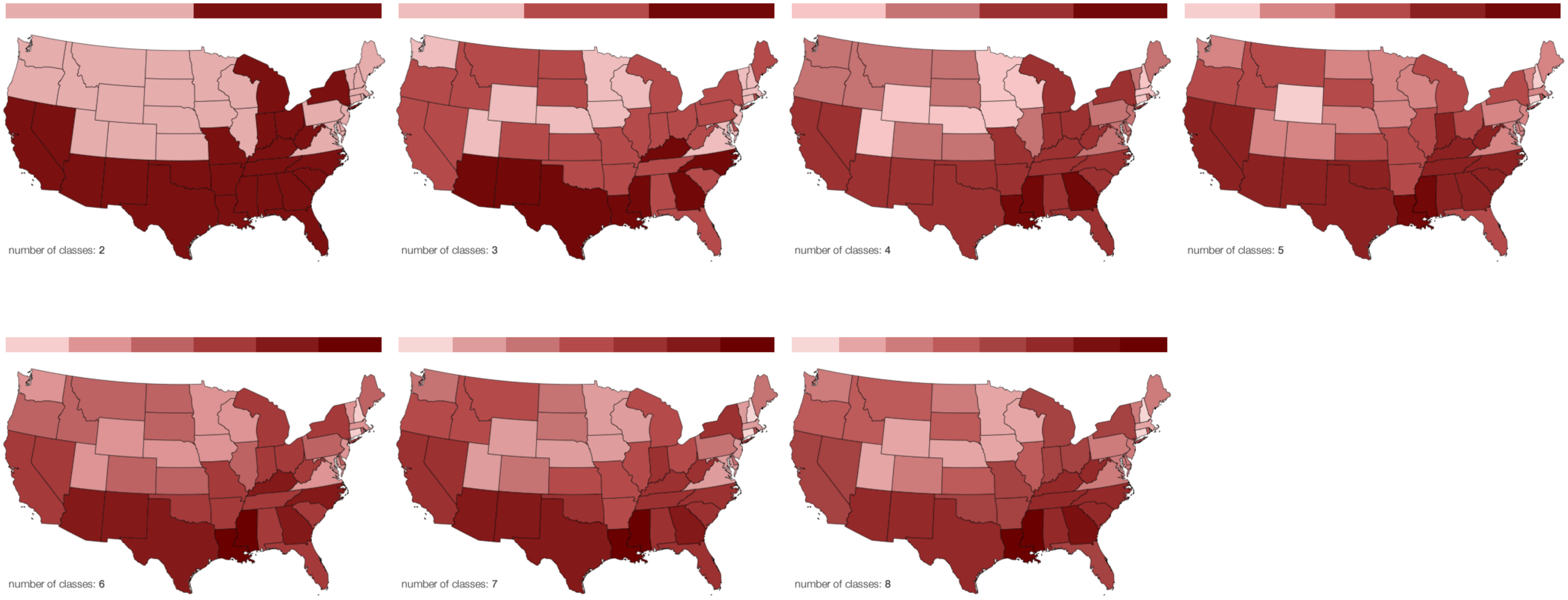
Categorical color: limited number of discriminable bins





https://archive.nytimes.com/www.nytimes.com/interactive/2008/05/05/science/20080506_DISEASE.html

Ordered color: limited number of discriminable bins



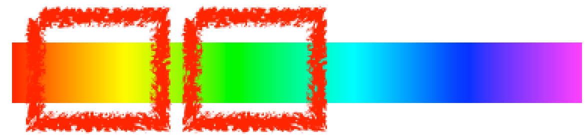
[Gregor Aisch, vis4.net/blog/posts/choropleth-maps/](vis4.net/blog/posts/choropleth-maps/)

Ordered color: Rainbow is poor default

problems

- perceptually unordered
- perceptually nonlinear

Ordered color: Rainbow is poor default



problems

- perceptually unordered
- perceptually nonlinear

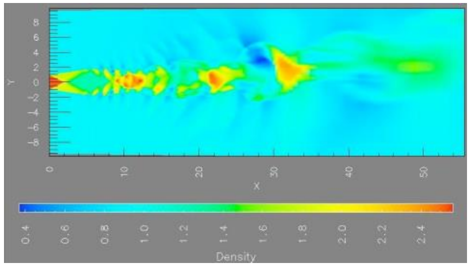
Ordered color: Rainbow is poor default

problems

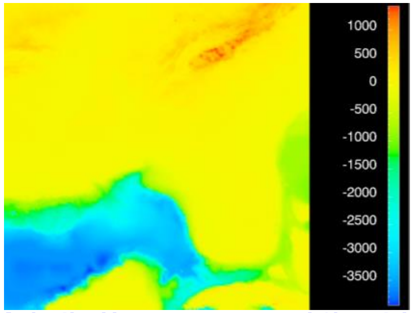
- perceptually unordered
- perceptually nonlinear

benefits

- fine-grained structure visible and nameable



[A Rule-based Tool for Assisting Colormap Selection. Bergman,. Rogowitz, and. Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

Ordered color: Rainbow is poor default

problems

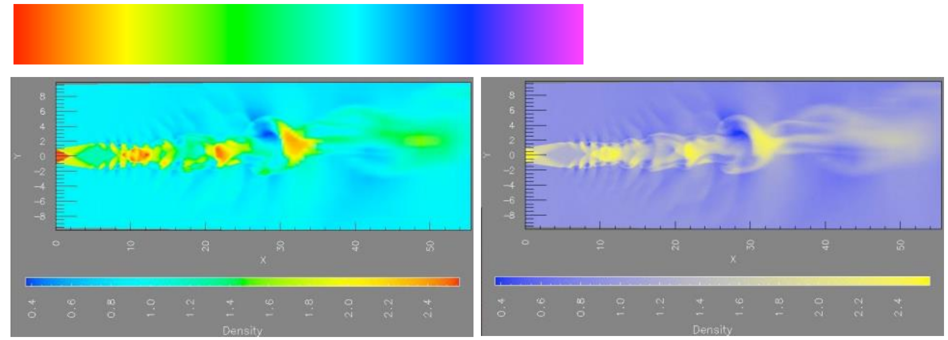
- perceptually unordered
- perceptually nonlinear

benefits

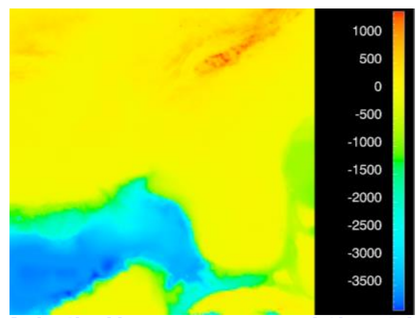
- fine-grained structure visible and nameable

alternatives

- large-scale structure: fewer hues



[A Rule-based Tool for Assisting Colormap Selection. Bergman,. Rogowitz, and. Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

Ordered color: Rainbow is poor default

problems

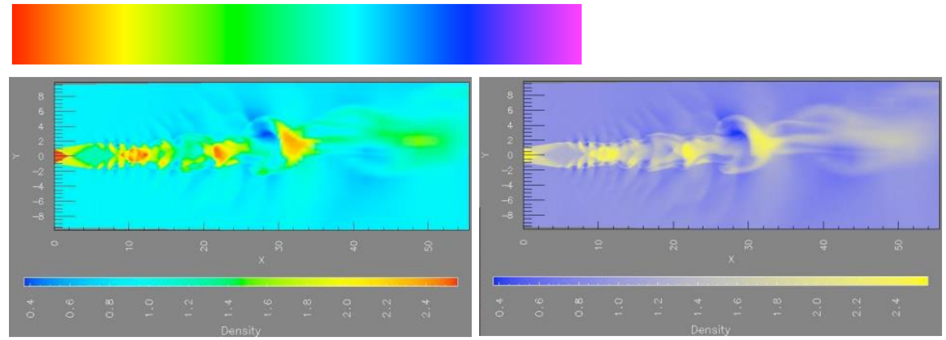
- perceptually unordered
- perceptually nonlinear

benefits

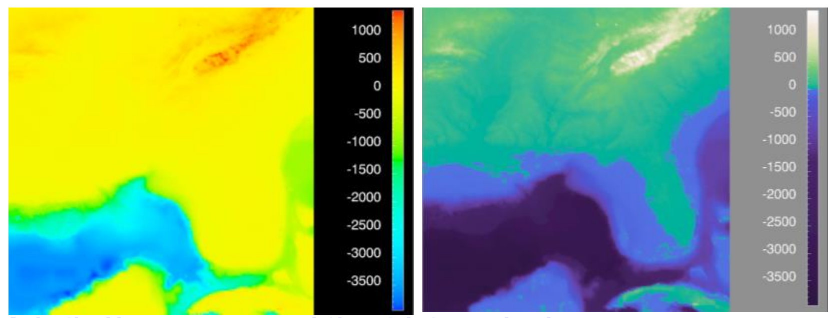
- fine-grained structure visible and nameable

alternatives

- large-scale structure: fewer hues
- fine structure: multiple hues with monotonically increasing luminance [eg viridis]



[A Rule-based Tool for Assisting Colormap Selection. Bergman,. Rogowitz, and. Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



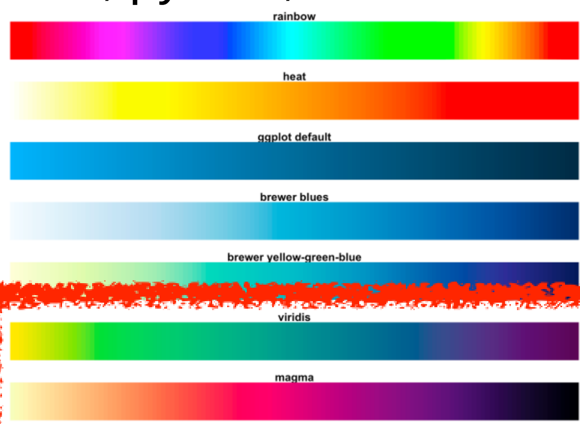
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

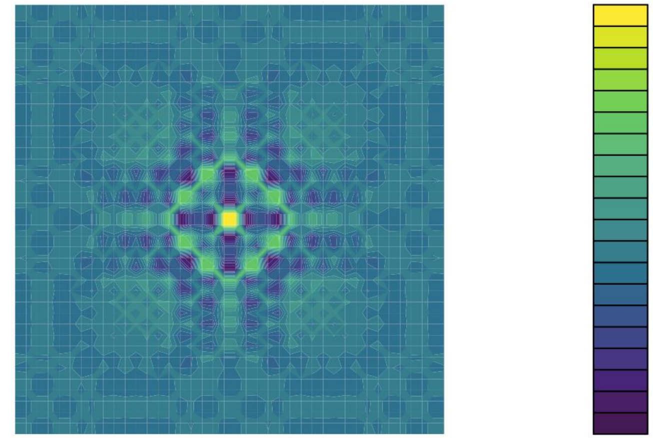
Viridis / Magma: sequential colormaps

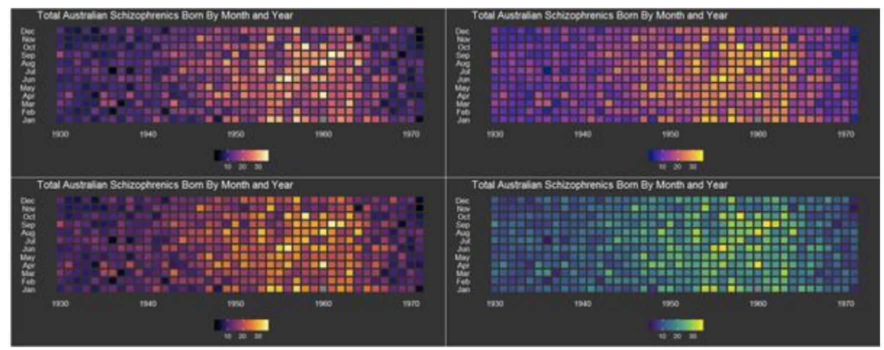
monotonically increasing luminance, perceptually uniform

colorful, colorblind-safe

- R, python, D3







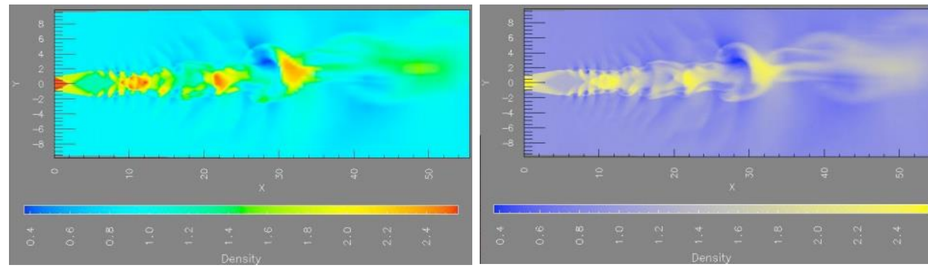
Ordered color: Rainbow is poor default

problems
- perceptually unordered
- perceptually nonlinear

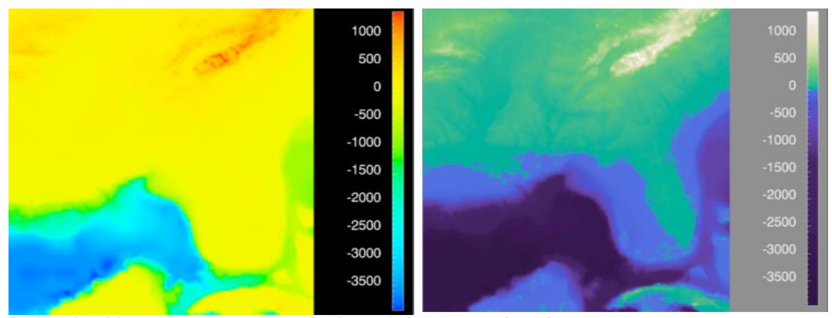
benefits
- fine-grained structure visible and nameable

alternatives
- large-scale structure: fewer hues
- fine structure: multiple hues with monotonically increasing luminance [eg viridis]

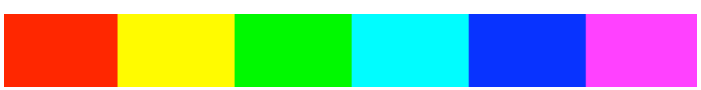
legit for categorical
- segmented saturated rainbow is good!



[A Rule-based Tool for Assisting Colormap Selection. Bergman,. Rogowitz, and. Treinish. Proc. IEEE Visualization (Vis), pp. 118–125, 1995.]



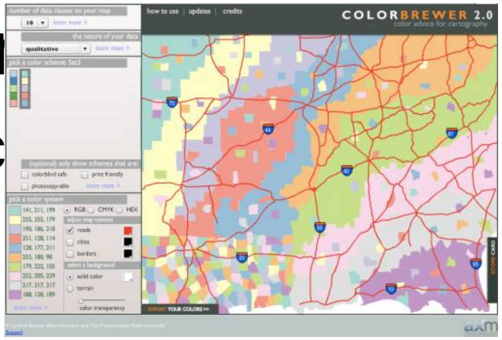
[Why Should Engineers Be Worried About Color? Treinish and Rogowitz 1998. http://www.research.ibm.com/people/l/lloydt/color/color.HTM]

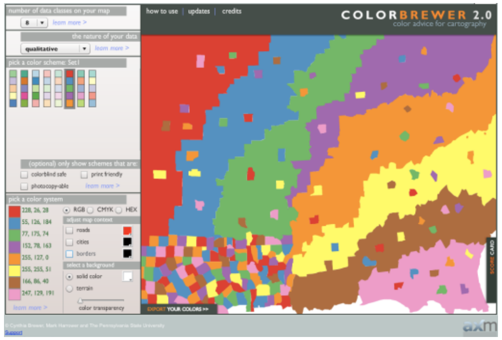


[Transfer Functions in Direct Volume Rendering: Design, Interface, Interaction. Kindlmann. SIGGRAPH 2002 Course Notes]

Interaction between channels: Not fully separable

color channel interactio





– size heavily affects salienc
– small regions need high saturation
– large regions need low saturation

http://colorbrewer2.org/

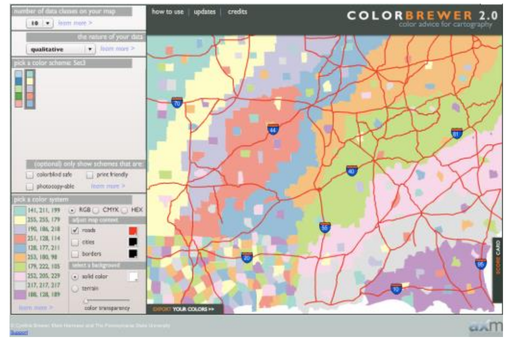
Interaction between channels: Not fully separable

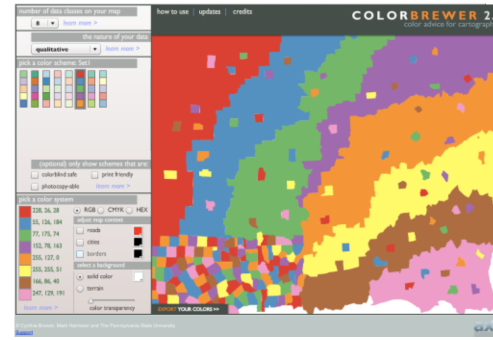
color channel interactions

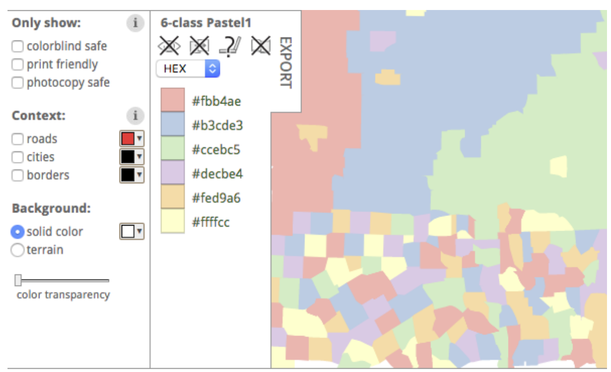
- size heavily affects salience
- small regions need high saturation
- large regions need low saturation

saturation & luminance:

- not separable from each other!
- also not separable from transparency

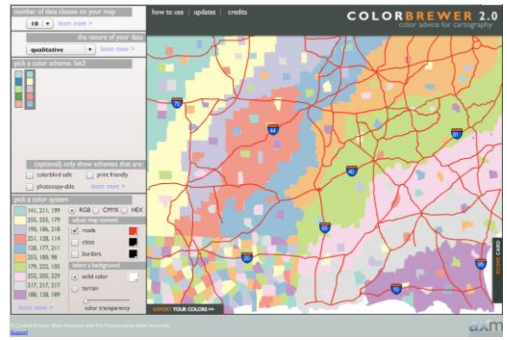


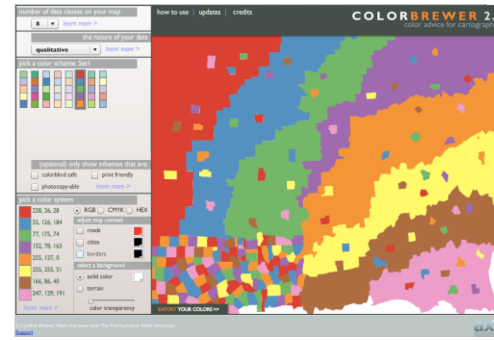




http://colorbrewer2.org/

Interaction between channels: Not fully separable



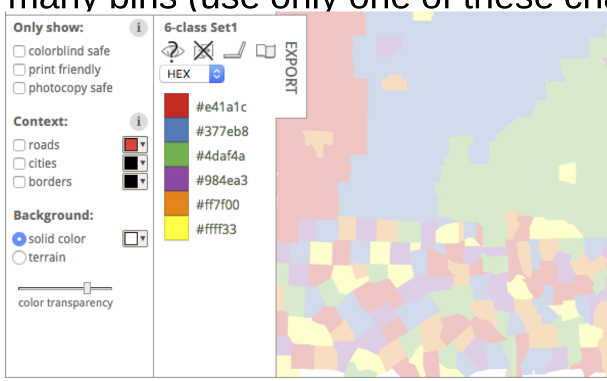


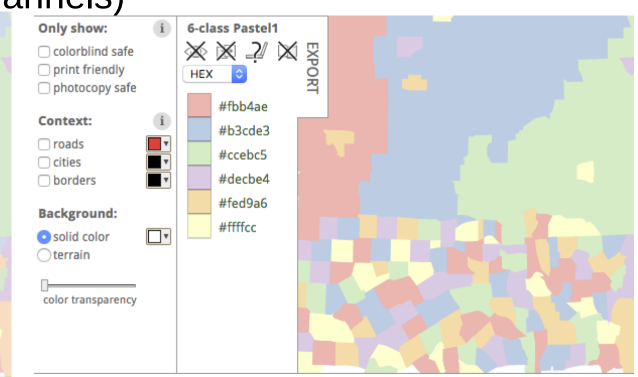
color channel interactions

- size heavily affects salience
- small regions need high saturation
- large regions need low saturation

saturation & luminance:

- not separable from each other!
- also not separable from transparency
- small separated regions: 2 bins safest (use only one of these channels), 3-4 bins max
- contiguous regions: many bins (use only one of these channels)





http://colorbrewer2.org/

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

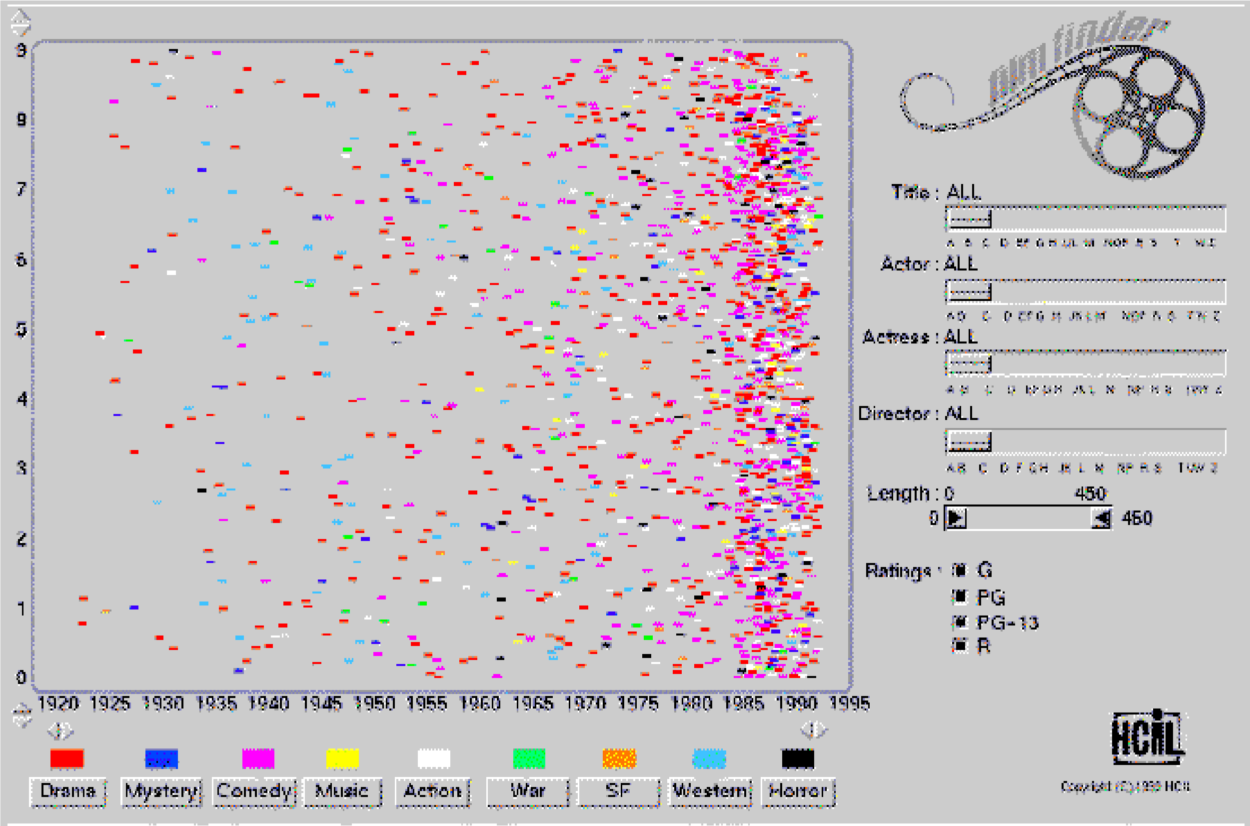
A Darwinian view of visualizations:

- 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:

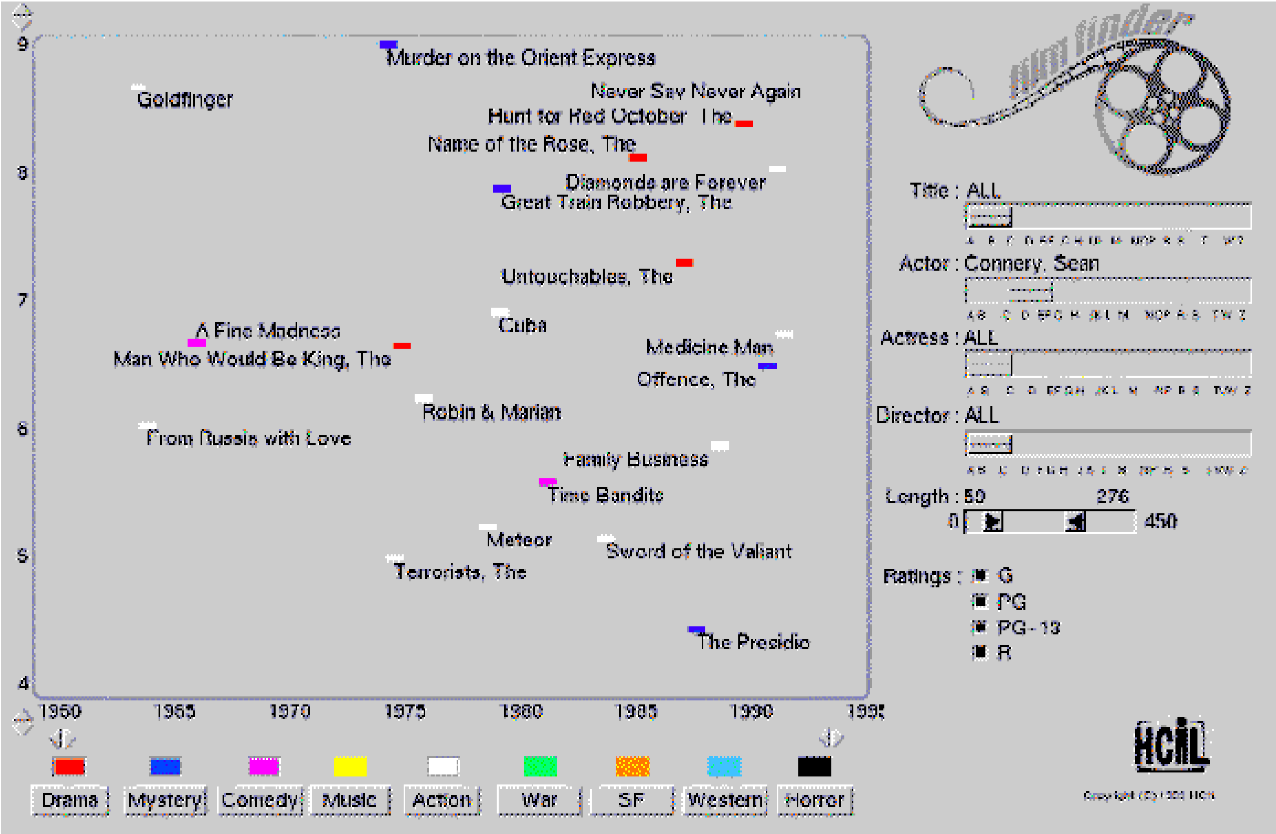
- Enhancing the old techniques with interactivity
- Example: Spotfire
  - Adds interactivity, color highlighting, zooming to scatterplots
- Example: TableLens / Eureka
  - Adds interactivity and length cues to tables

Spotfire: Integrating Interaction with Scatterplots



Ahlberg & Shneiderman, Color plate 1. The FilmFinder.

Spotfire/IVEE: Integrating Interaction with Scatterplots

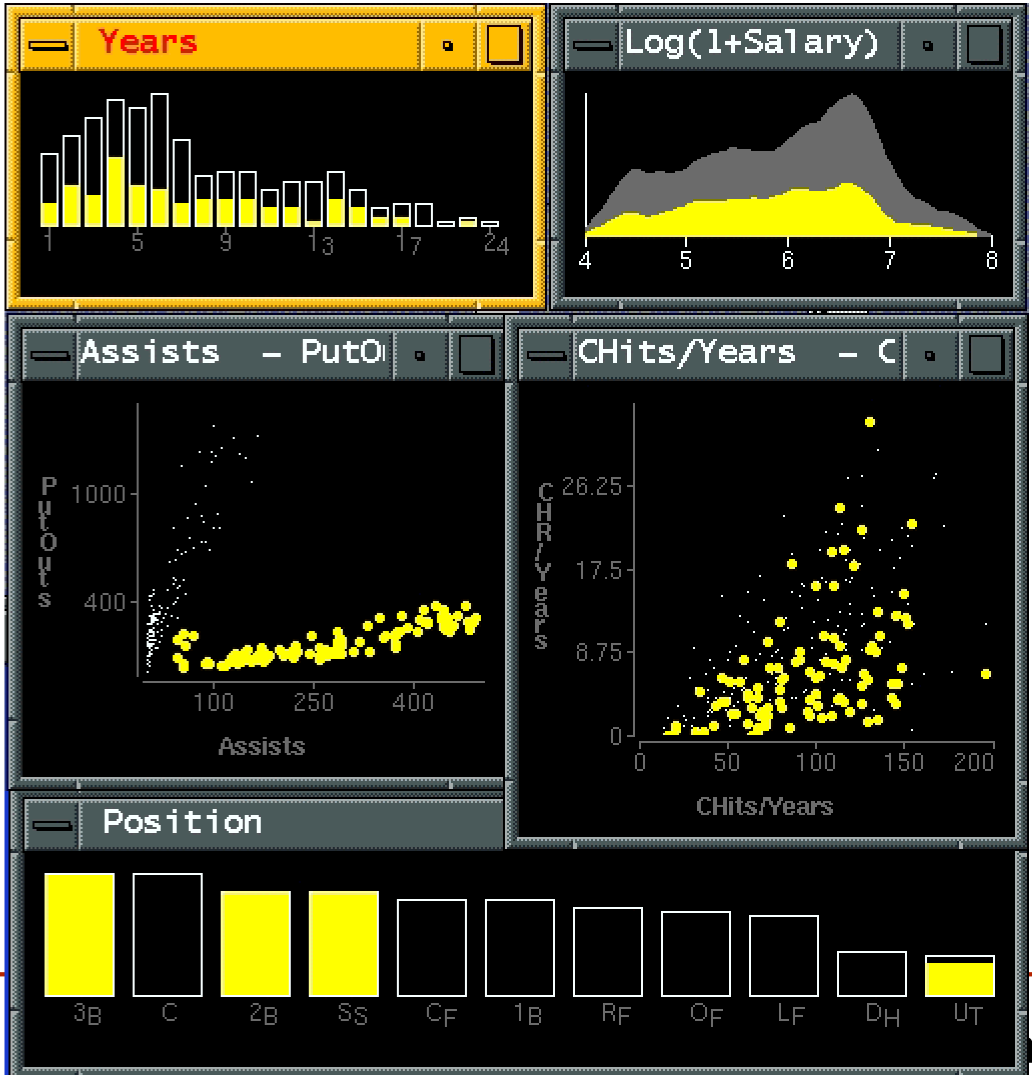


Ahlberg & Shneiderman, Color plate 2. Categories have been selected, the displayed is zoomed into 1960-95 and popularity 4-9, and Sean Connery has been selected.

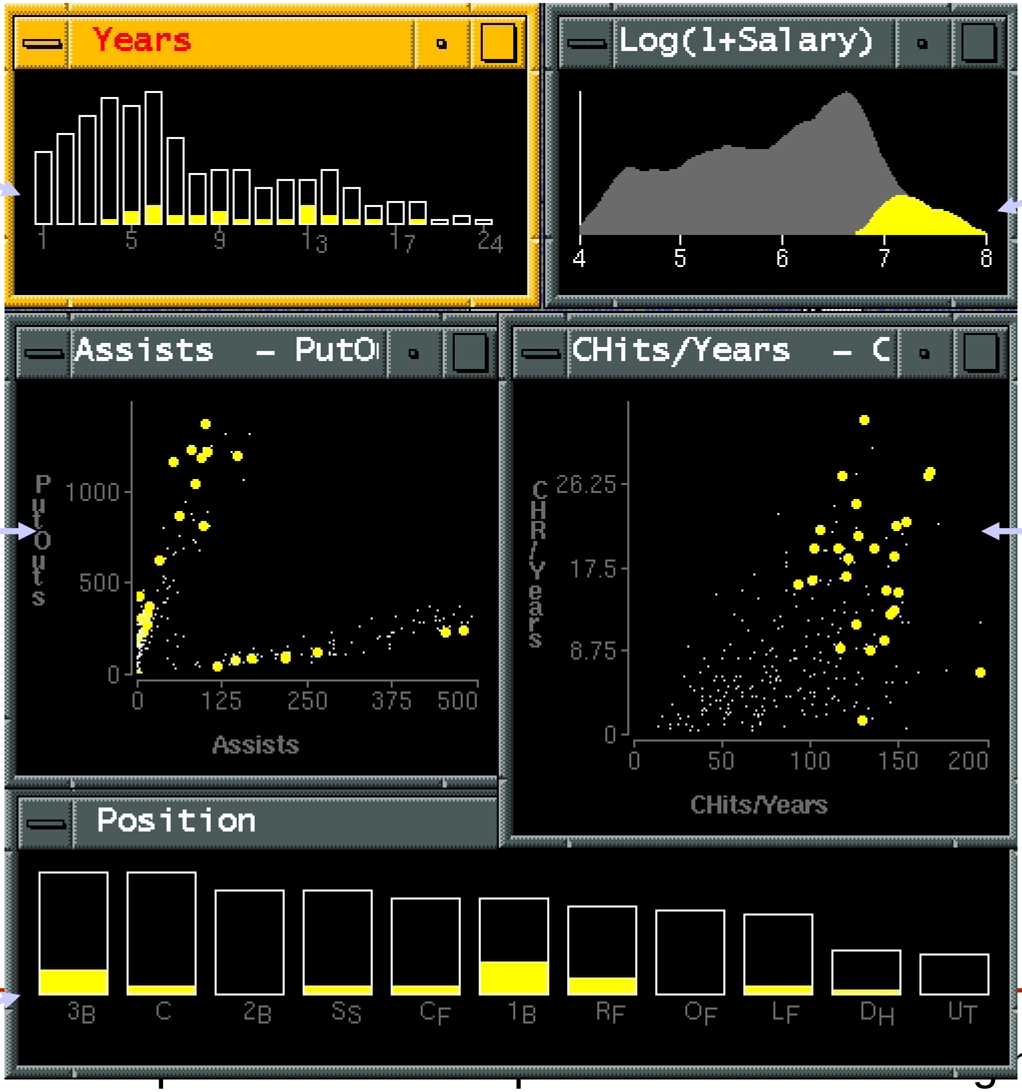
Brushing and Linking

- 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)



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

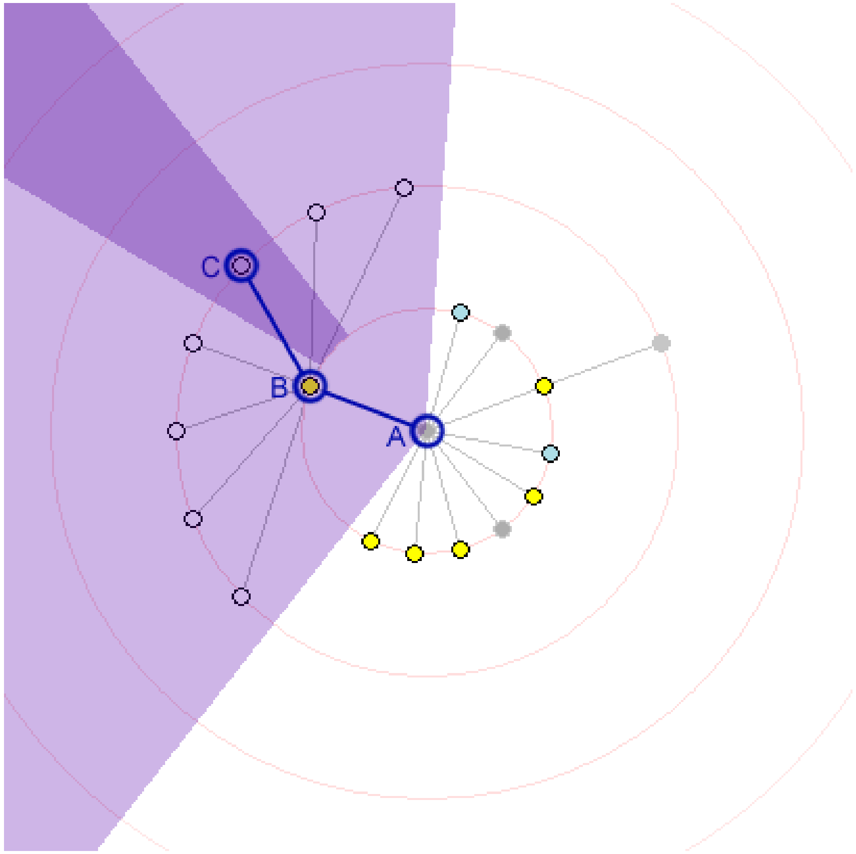
Visualization of Peer-to-Peer Network

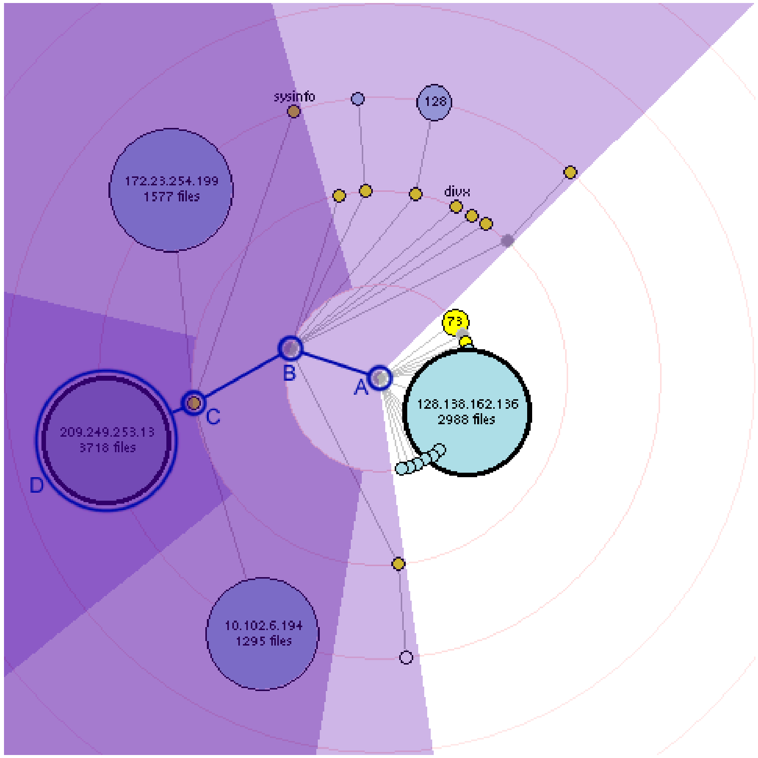
- Hosts (with color for status and size for number of files)
- Nodes with closer network distance from focus on inner rings
- Queries shown; can trace queries

Gnutellavision as exploratory tool

- Very few hosts share many files
- Uneven propagation of queries
- Qualitative assessment of queries (simple)

Layout - Illustration





Each node is placed at the center of the angular sector allocated to it defined by the width of that subtree

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)

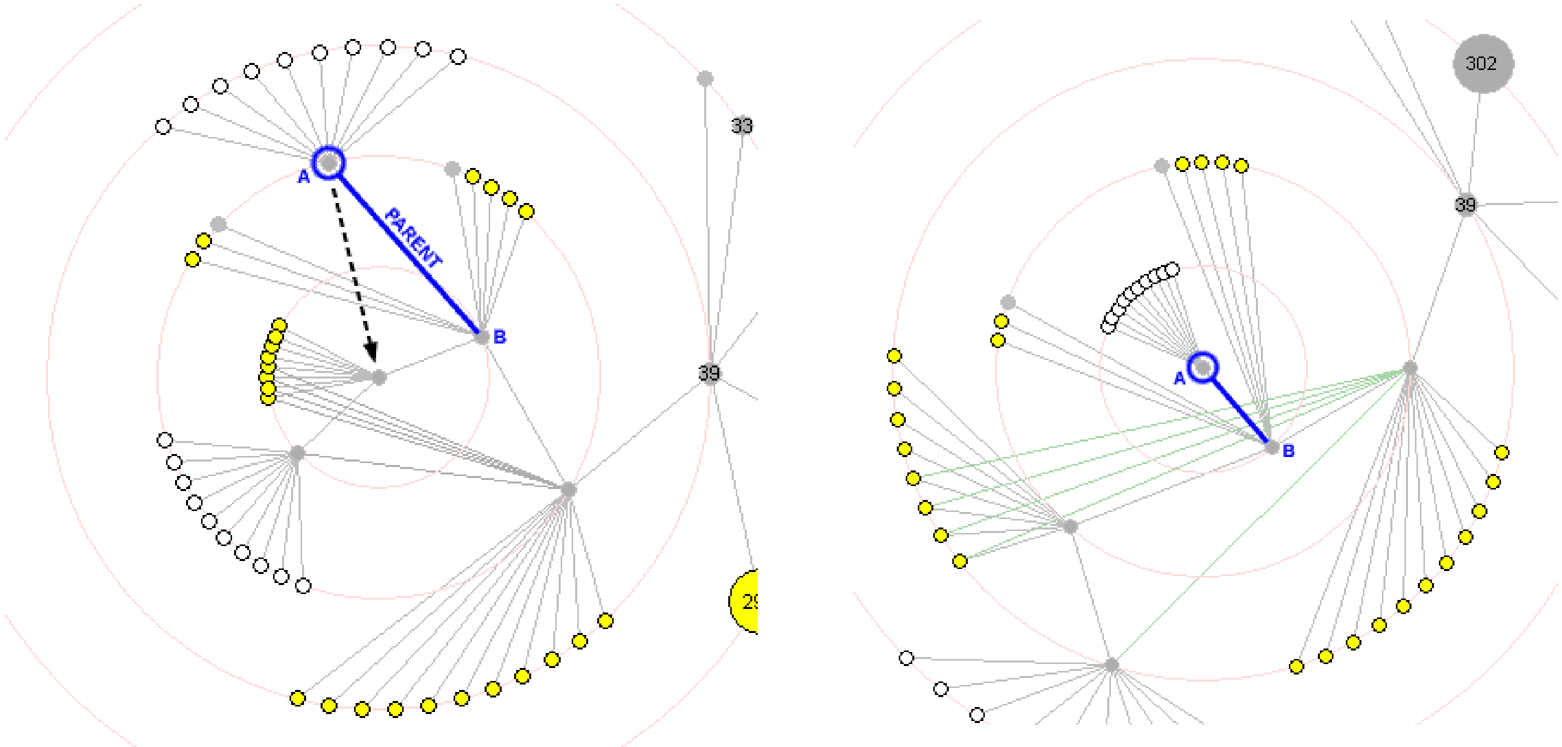
Avoid cross-over of edges

(to allow users to keep track of which is which)

Animation timing

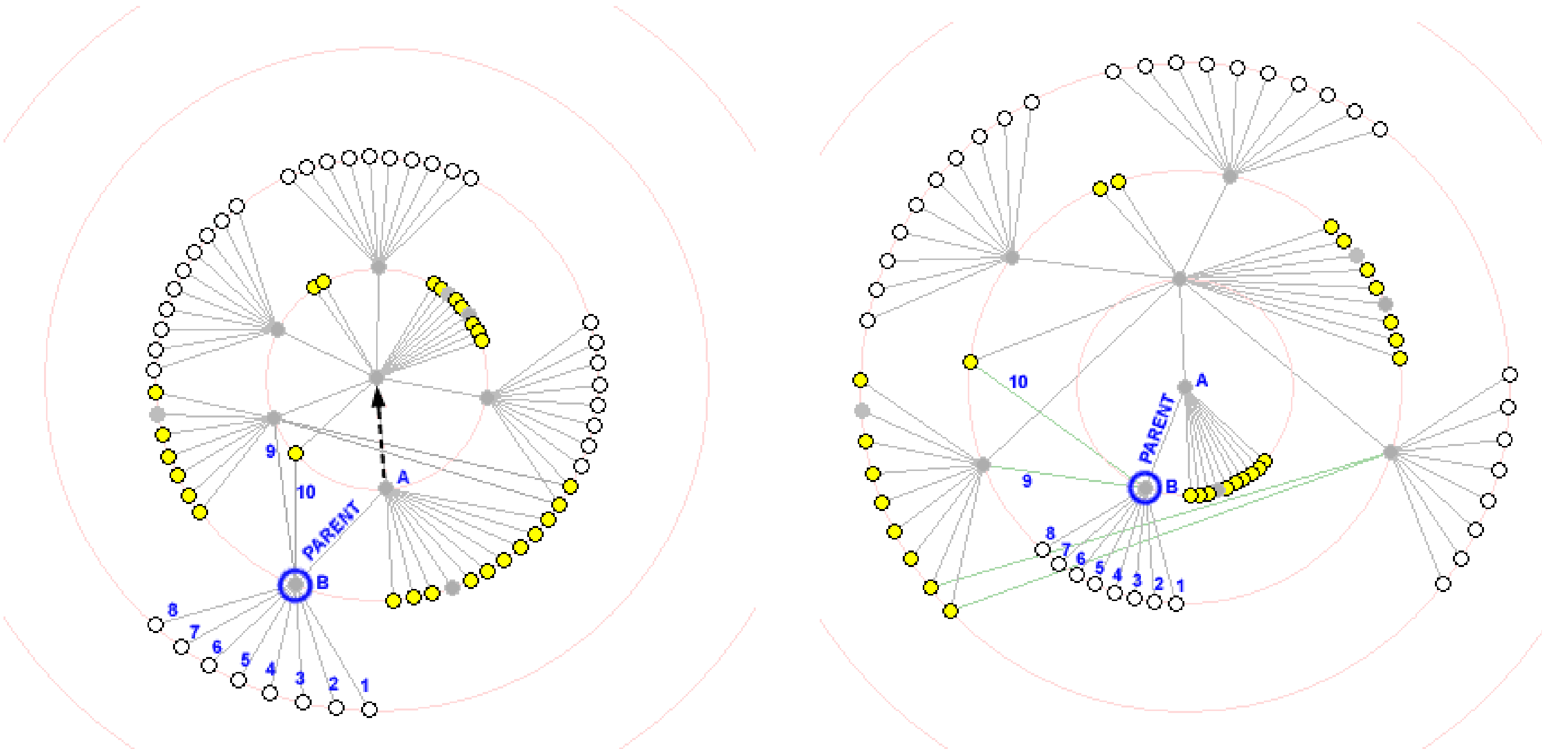
Slow in Slow out timing (allows users to better track movement)

Transition Constraint - Orientation



https://www.youtube.com/watch?v=AvfIlcE2xZw

Transition Constraint - Order



https://www.youtube.com/watch?v=AvfIlcE2xZw

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

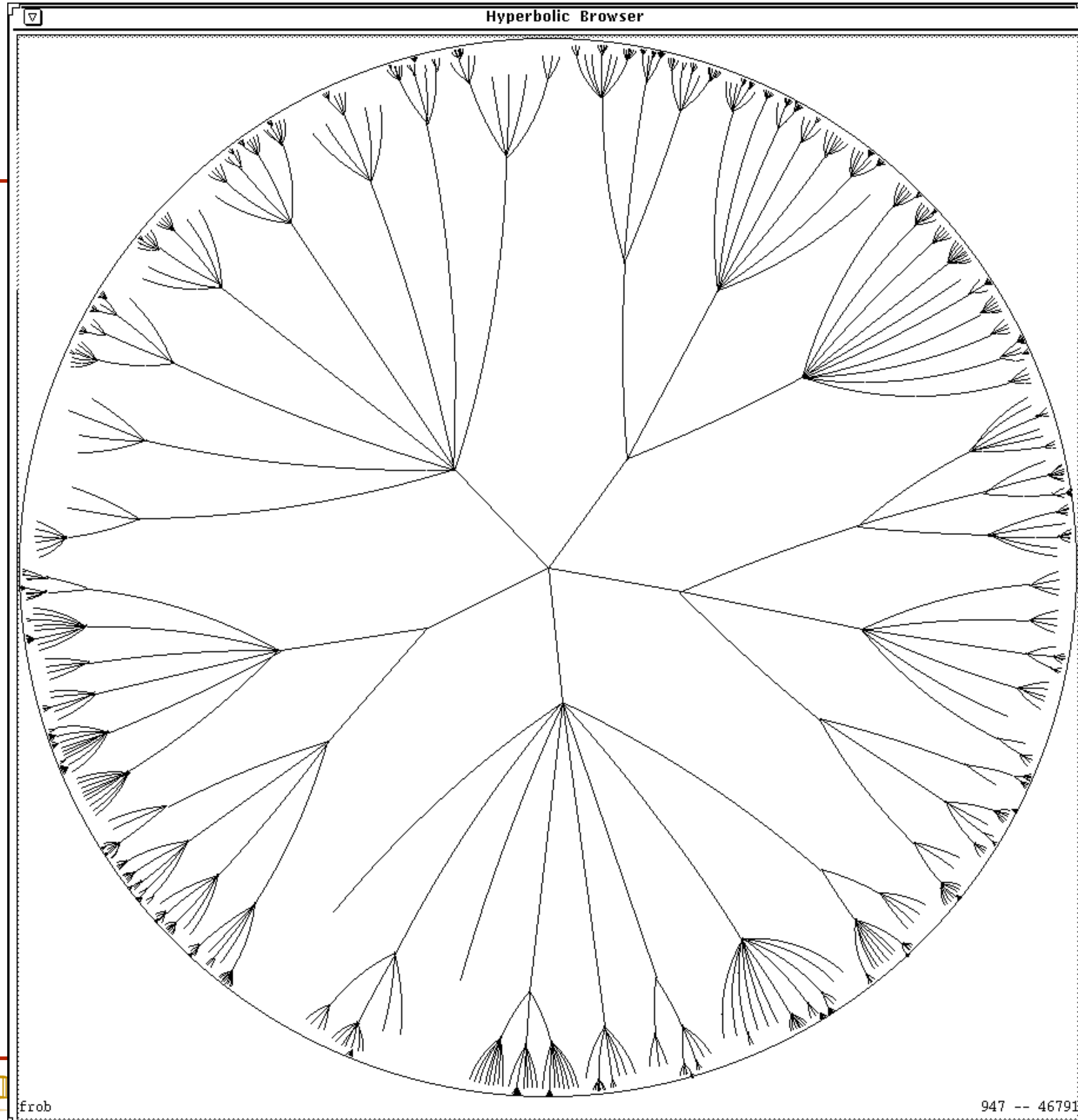
A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies (1995) John Lamping, Ramana Rao, Peter Pirolli Proc. ACM Conf. Human Factors in Computing Systems, CHI

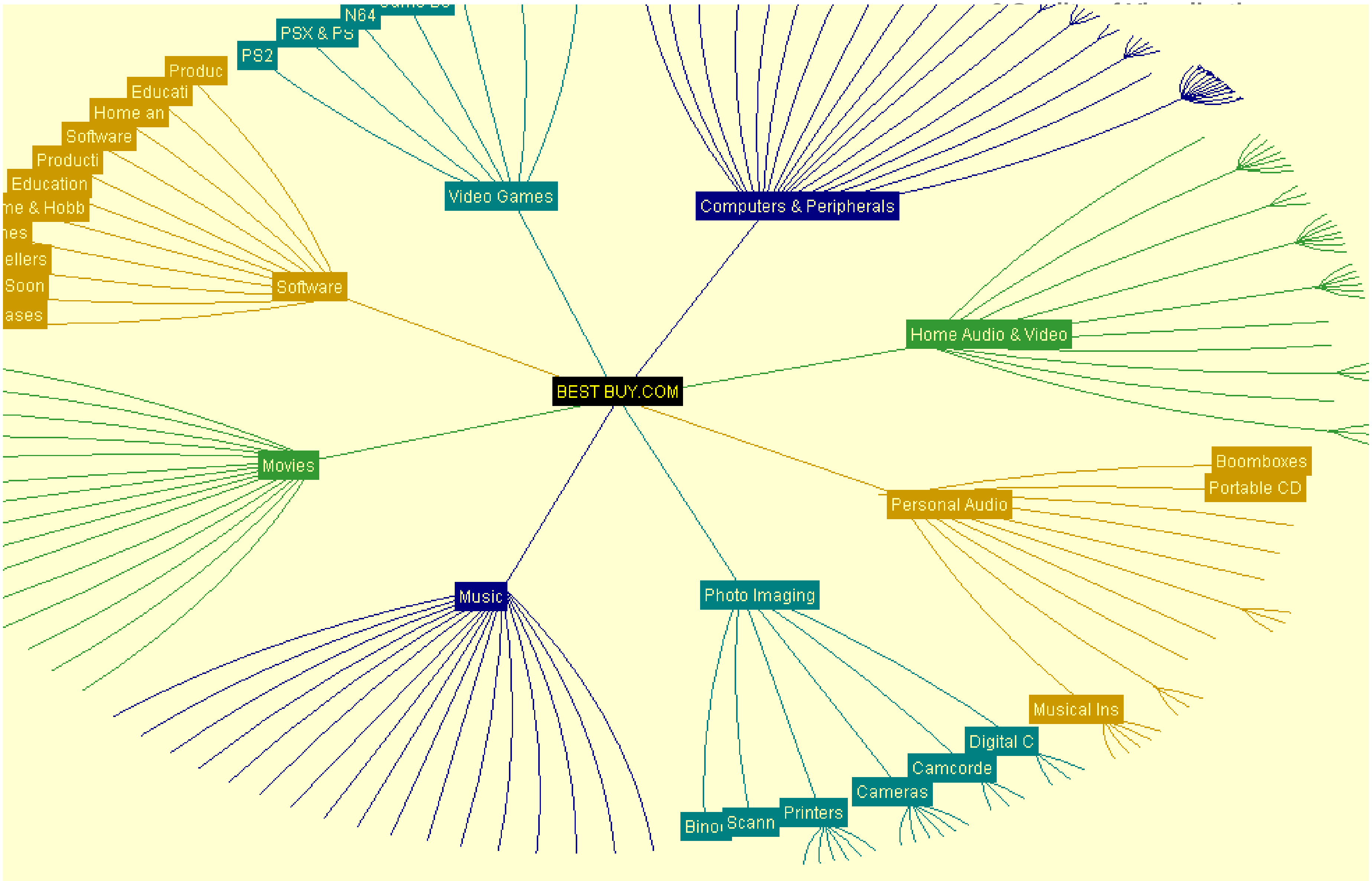
Also uses animation

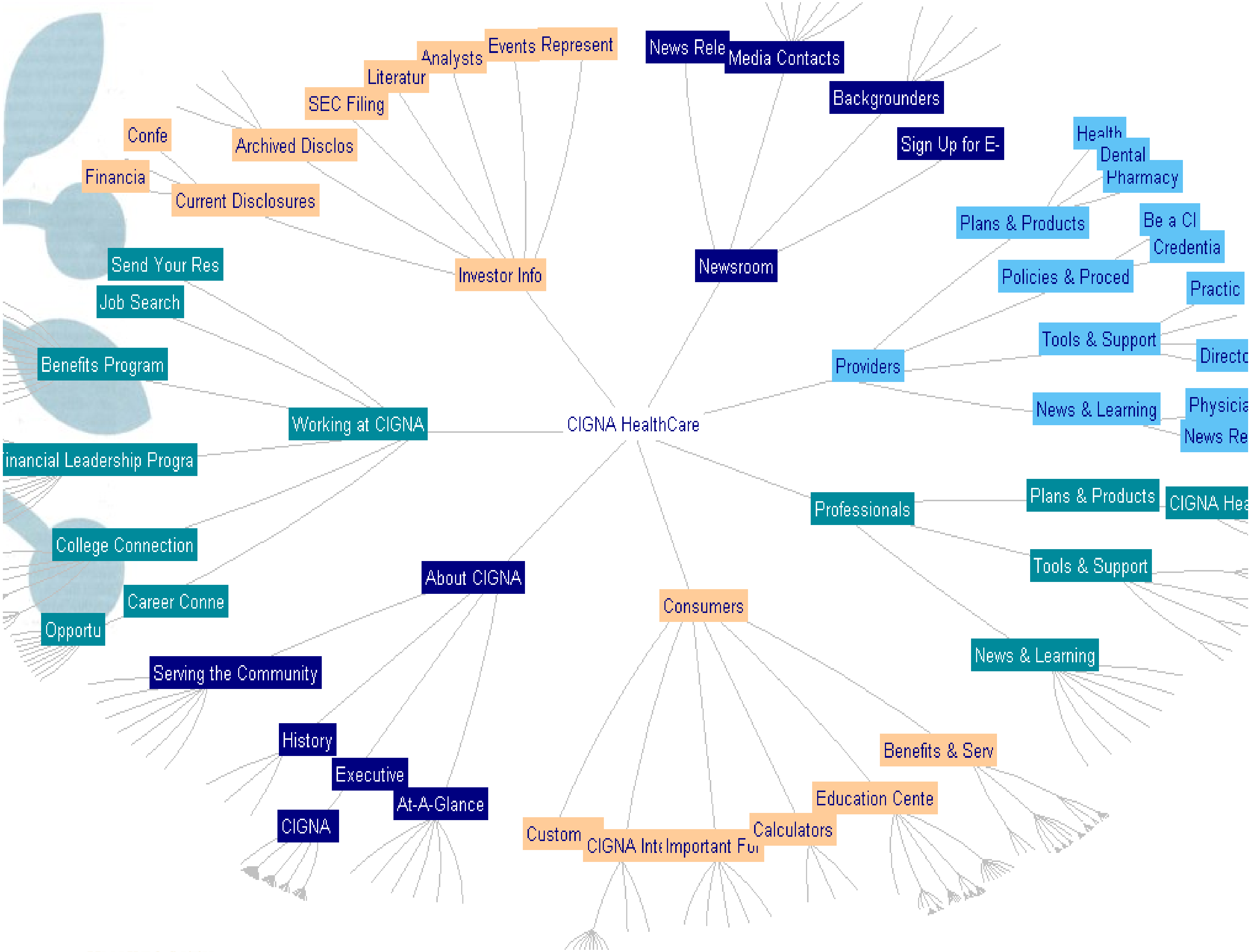
Tree-based layout; leaves stretch to infinity

Only a few labels can be seen at a time

https://www.youtube.com/watch?v=ZUwpqzDNhME

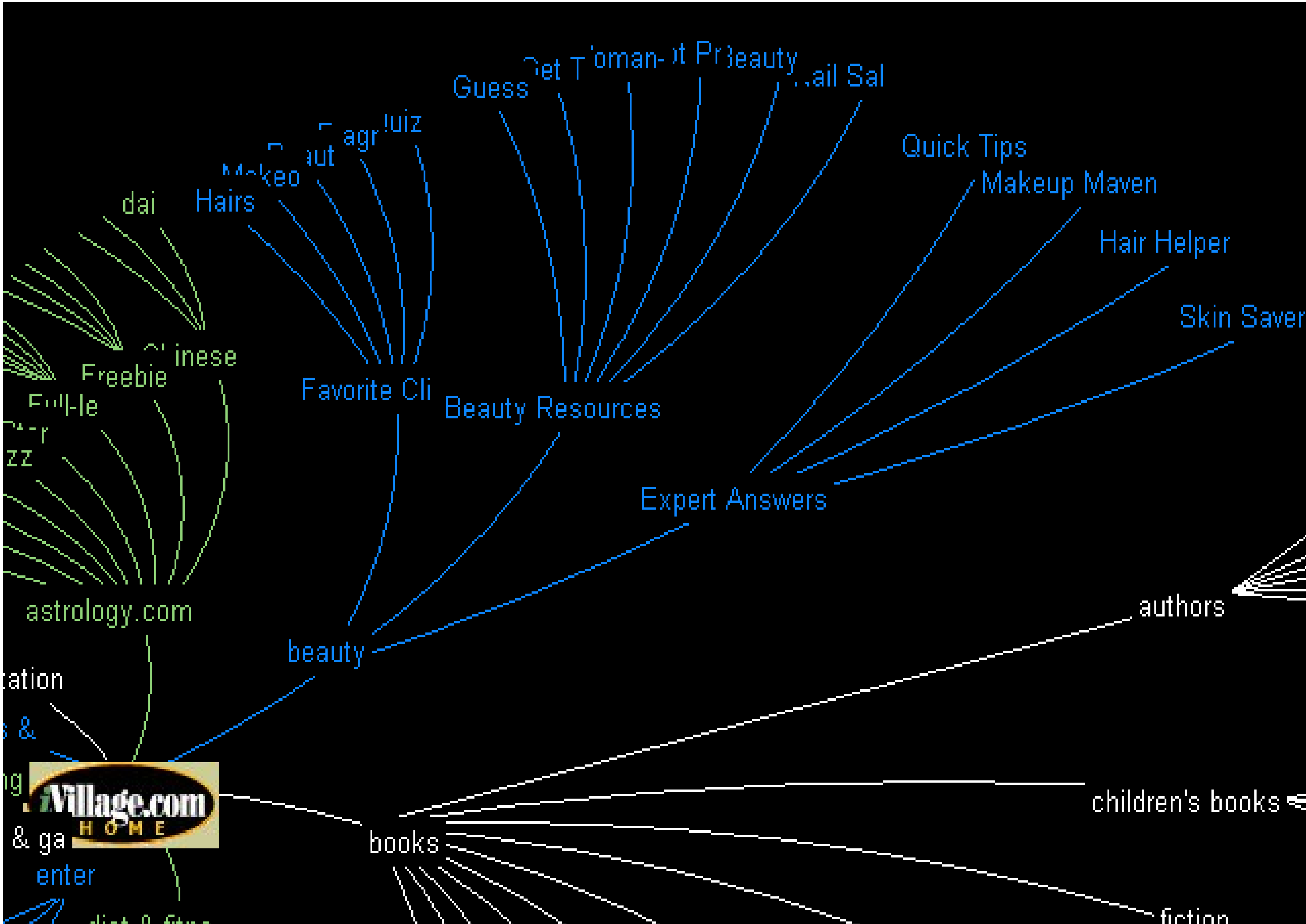

Hyperbolic Browser
frob
947 -- 46791


BEST BUY.COM
Video Games
PS2
PSX & PS
N64
Computers & Peripherals
Home Audio & Video
Software
Produc
Educati
Home an
Software
Producti
Education
me & Hobb
nes
ellers
Soon
ases
Movies
Personal Audio
Boomboxes
Portable CD
Musical Ins
Music
Photo Imaging
Digital C
Camcorde
Cameras
Printers
Scann
Binoı



CIGNA HealthCare
Investor Info
Events
Represent
Analysts
Literatur
SEC Filing
Archived Disclos
Current Disclosures
Confe
Financia
Newsroom
News Rele
Media Contacts
Backgrounders
Sign Up for E-
Providers
Plans & Products
Health
Dental
Pharmacy
Policies & Proced
Be a CI
Credentia
Tools & Support
Practic
Directo
News & Learning
Physicia
News Re
Working at CIGNA
Send Your Res
Job Search
Benefits Program
inancial Leadership Progra
College Connection
Career Conne
Opportu
Professionals
Plans & Products
CIGNA Hea
Tools & Support
News & Learning
About CIGNA
Serving the Community
History
Executive
At-A-Glance
CIGNA
Consumers
Benefits & Serv
Education Cente
Calculators
Custom
CIGNA Intε
Important Fu

UNIVERSITY


Guess
Quick Tips
Makeup Maven
Hair Helper
Skin Saver
dai
Hairs
Freebie
Favorite Cli
Beauty Resources
Expert Answers
astrology.com
beauty
authors
iVillage.com
HOME
books
children's books
enter
fiction

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

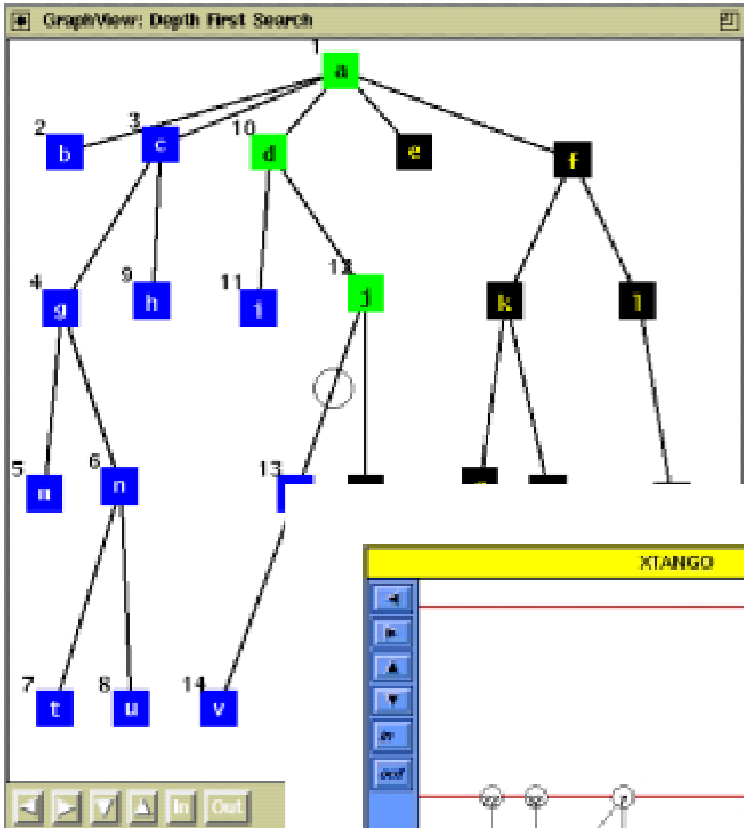
Kehoe, Stasko, and Taylor, "Rethinking Evaluation of Algorithm Animations as Learning Aids"

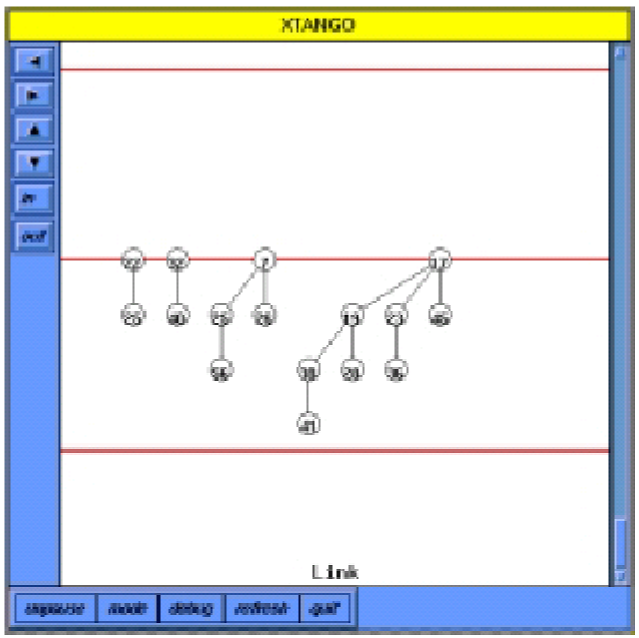
Why previous studies present no benefits:

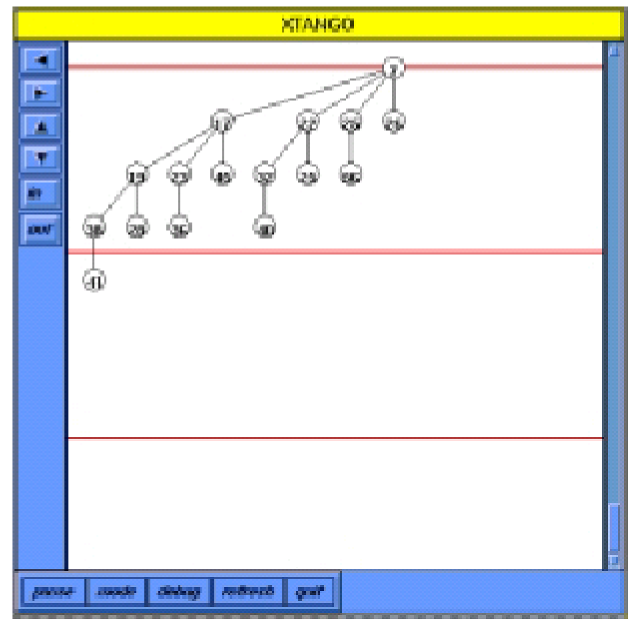
- No or limited benefits from particular animations
- Benefits are not captured in measurements
- Design of experiments hides the benefits

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


GraphView: Depth First Search


XTANGO


XTANGO

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

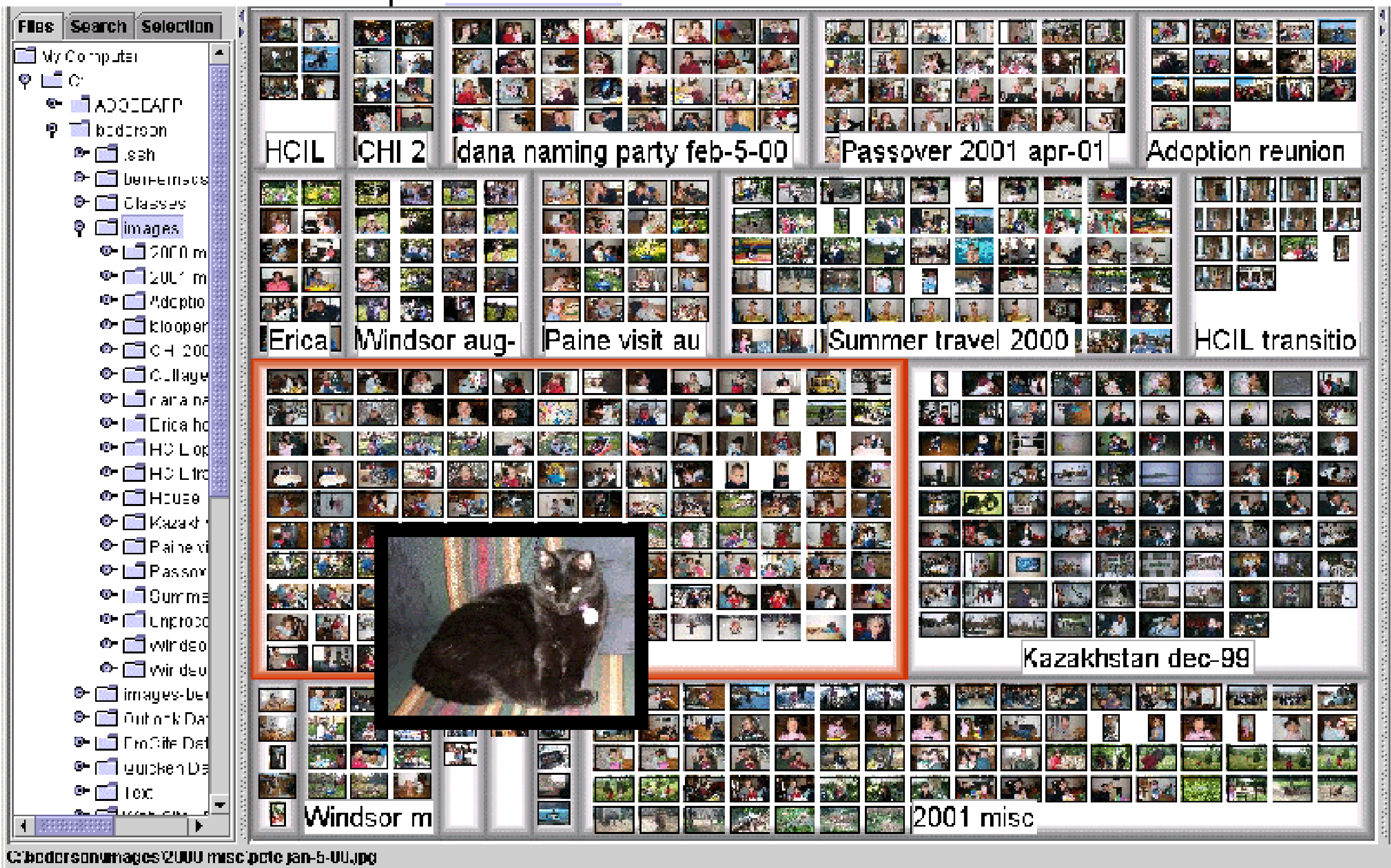
3D Navigation

Node-and-link representations of concept spaces

Zoom, Overview + Detail

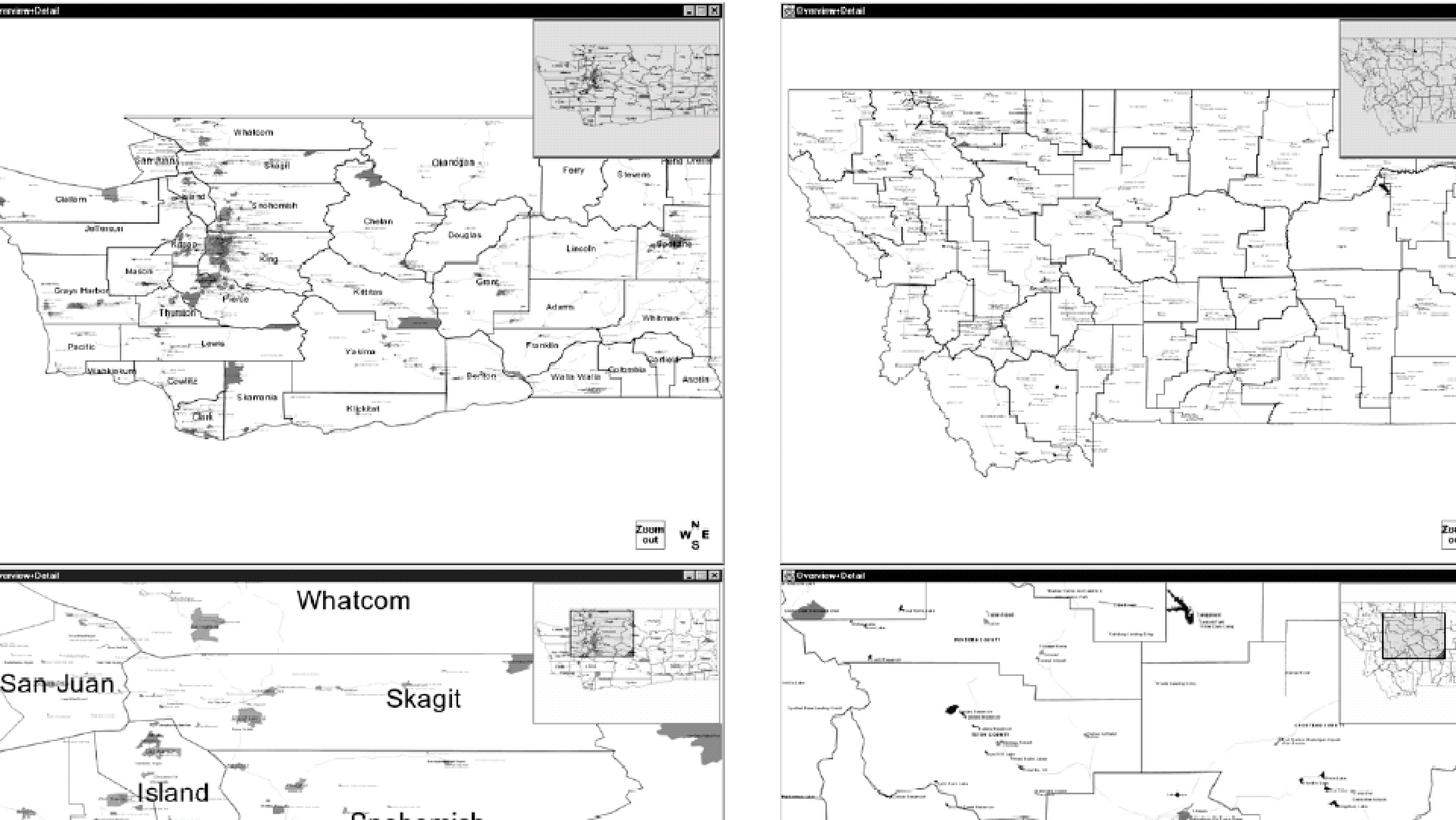
An exception, possibly:

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.

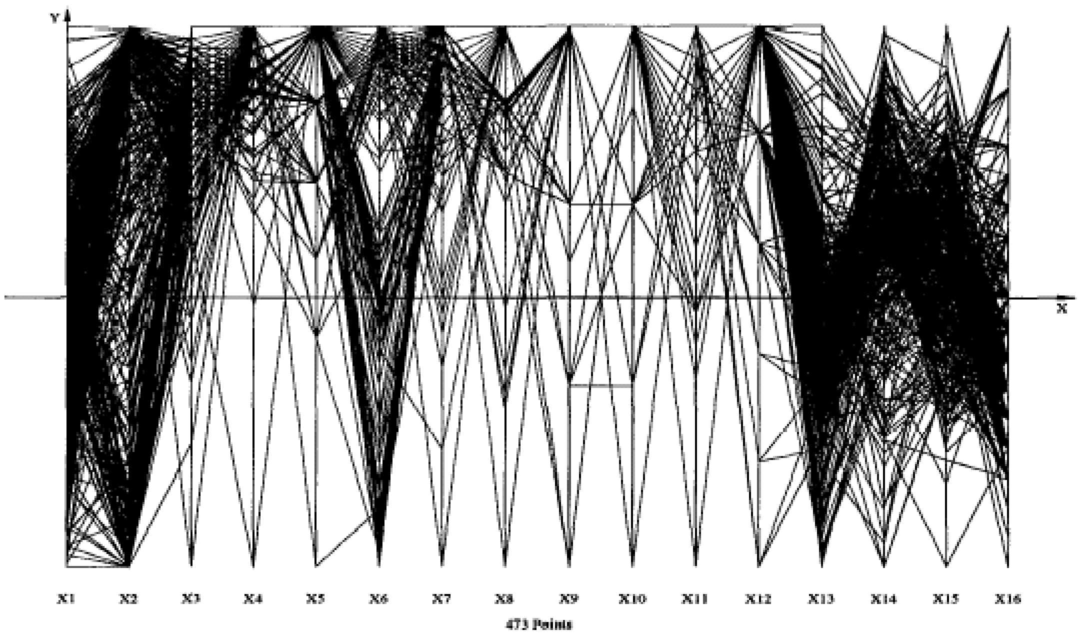


Figure 1: The full dataset consisting of 473 batches

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!!

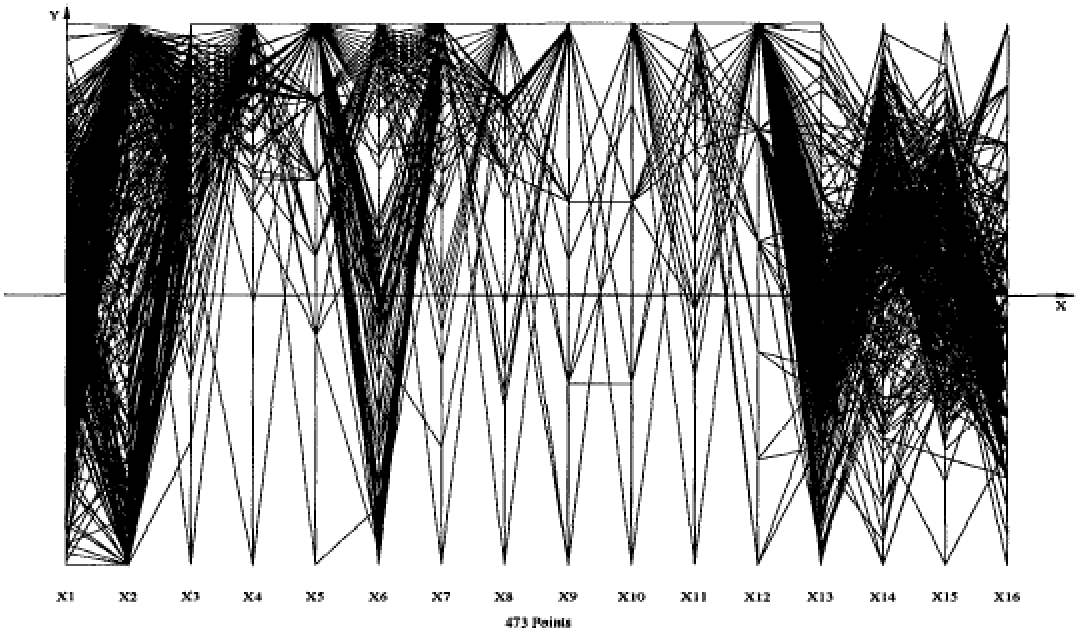


Figure 1: The full dataset consisting of 473 batches

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.

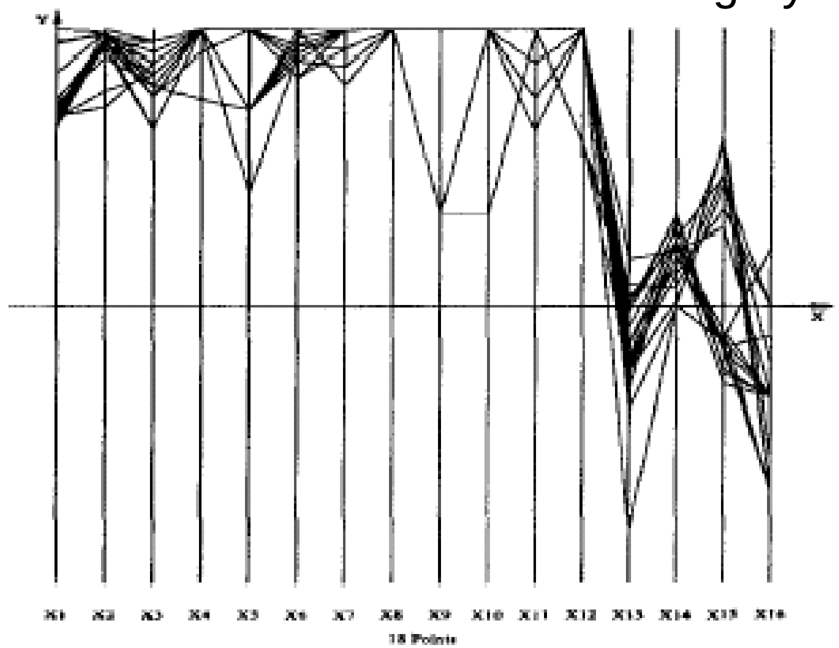


Figure 2: The batches high in Yield, $X1$, and Quality, $X2$.

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

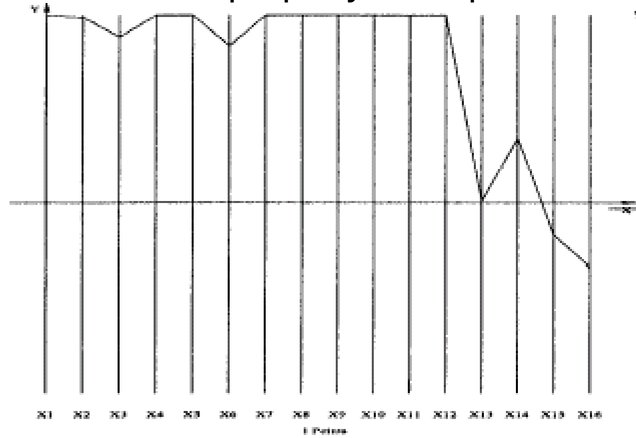


Figure 5: The best batch. Highest in Yield, $X1$, and very high in Quality, $X2$.

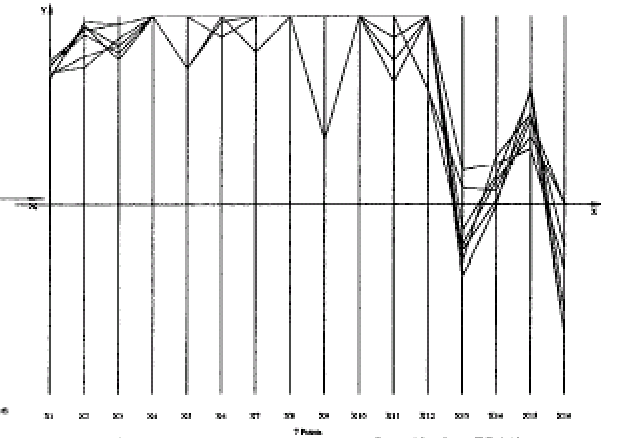
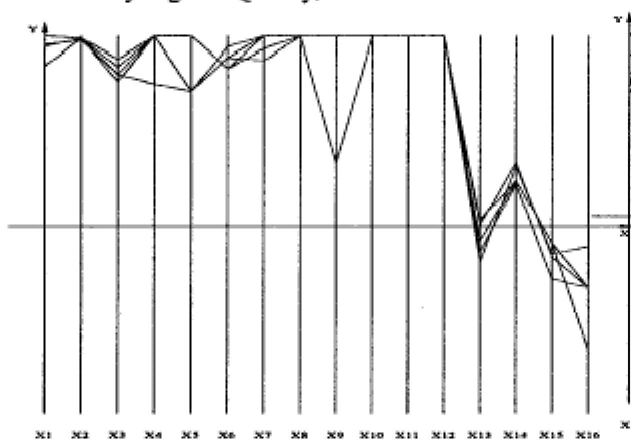
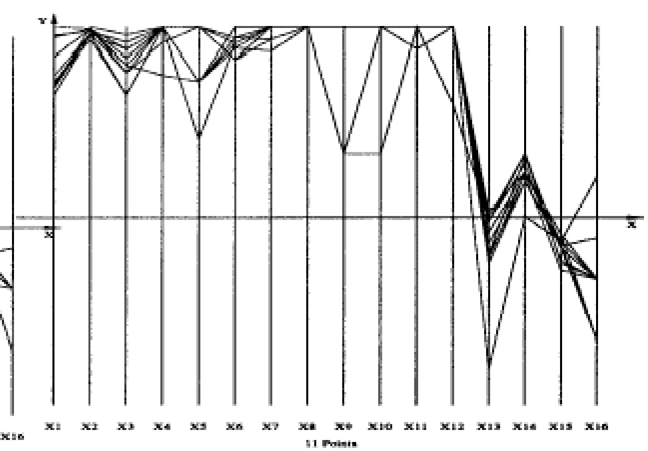


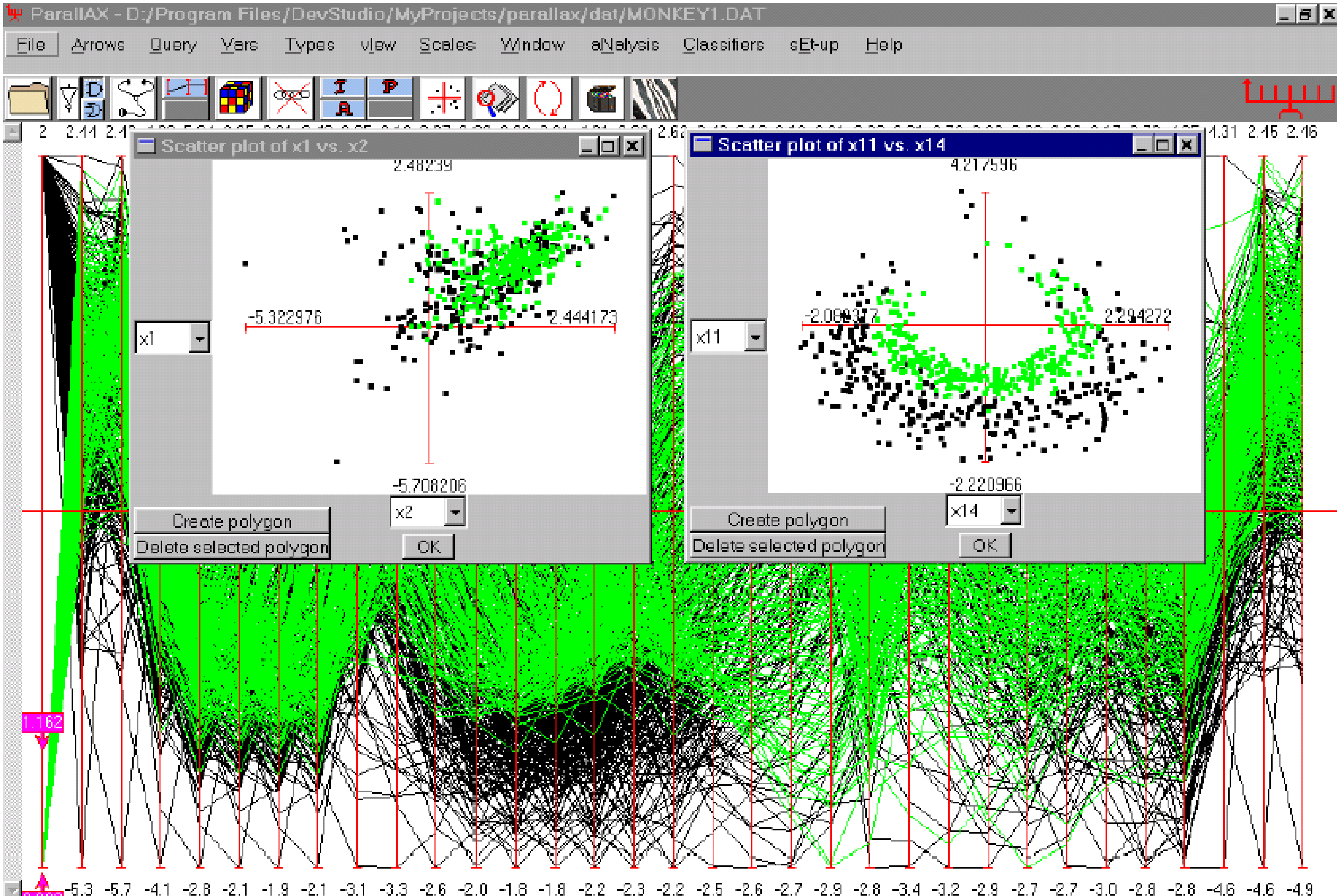
Figure 7: Upper range of split in $X15$





Automated Analysis

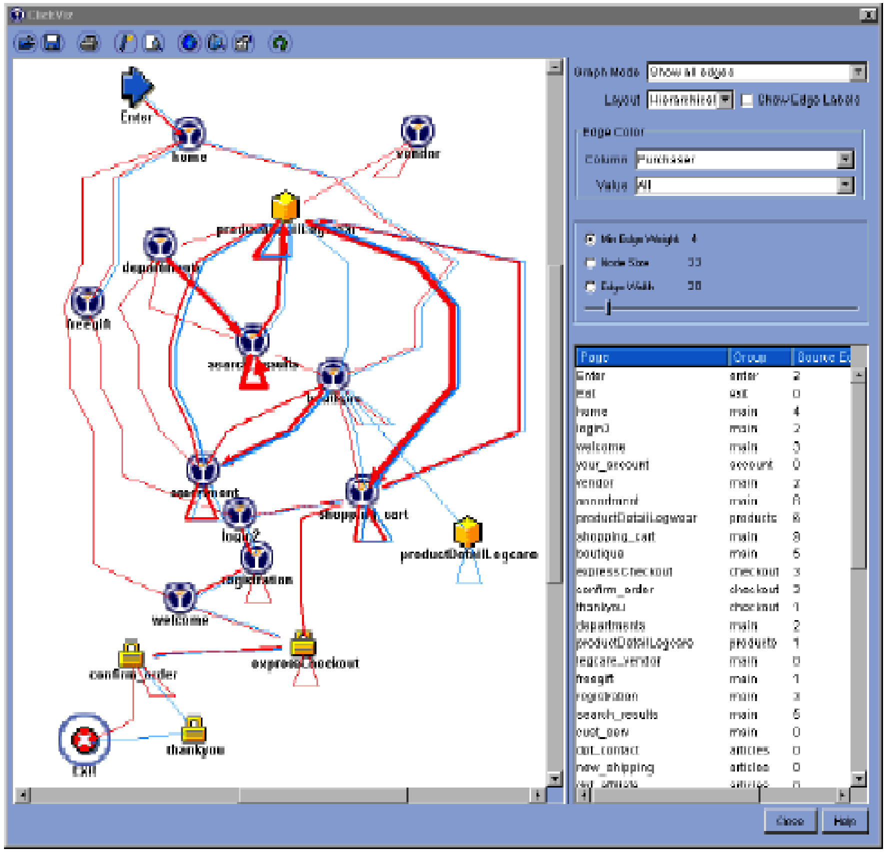
A. Inselberg, Automated Knowledge Discovery using Parallel Coordinates, INFOVIS '99



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

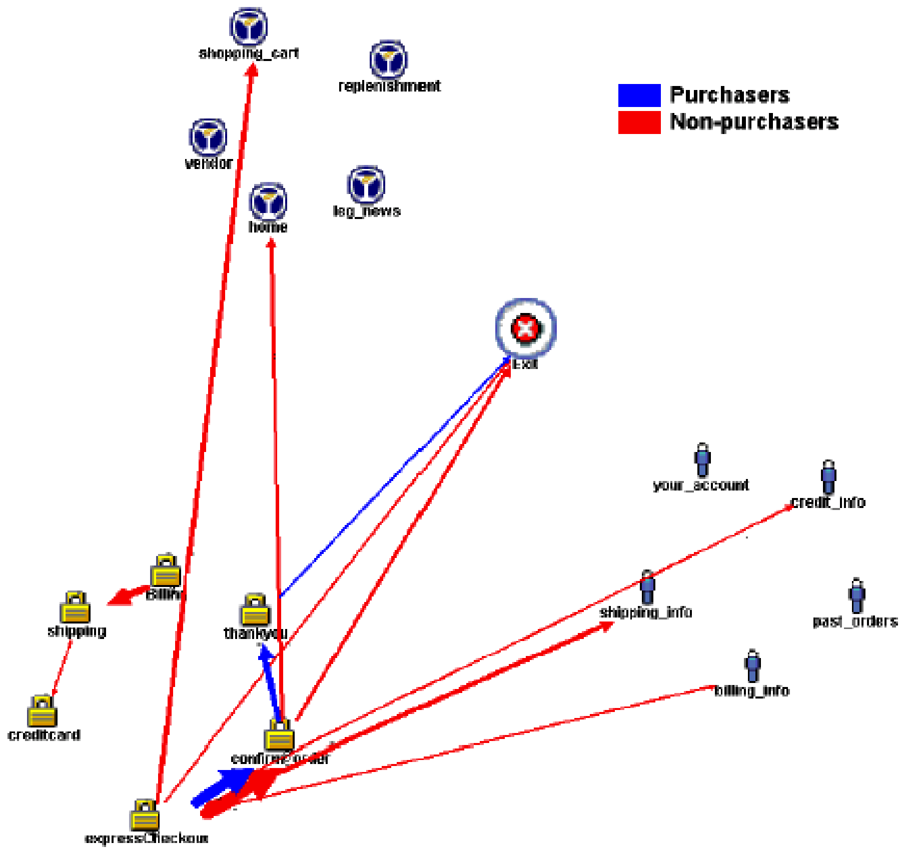
- Clickstream
- Purchase history
- Demographic data

Associates customer data with their clickstream

Different color for each customer segment

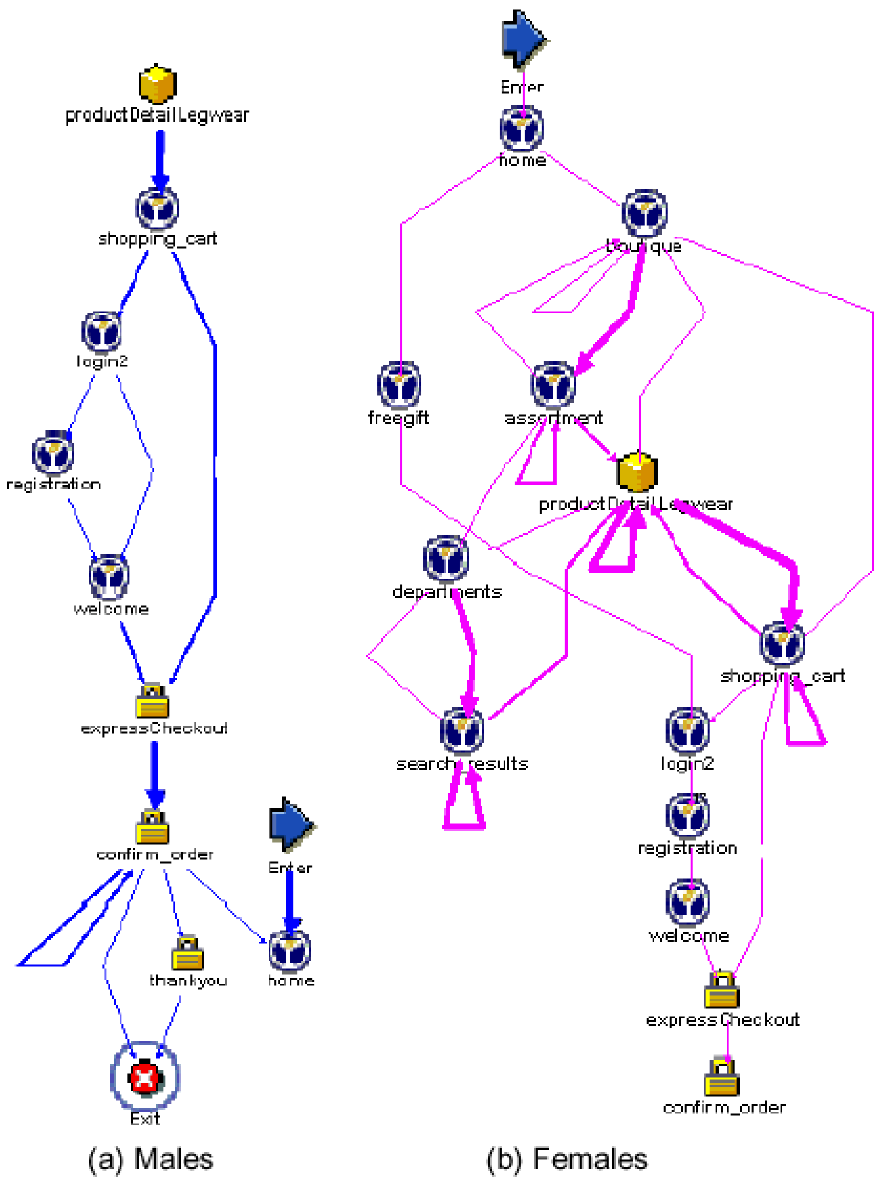
Layout

Aggregation based on file system path



Initial Findings

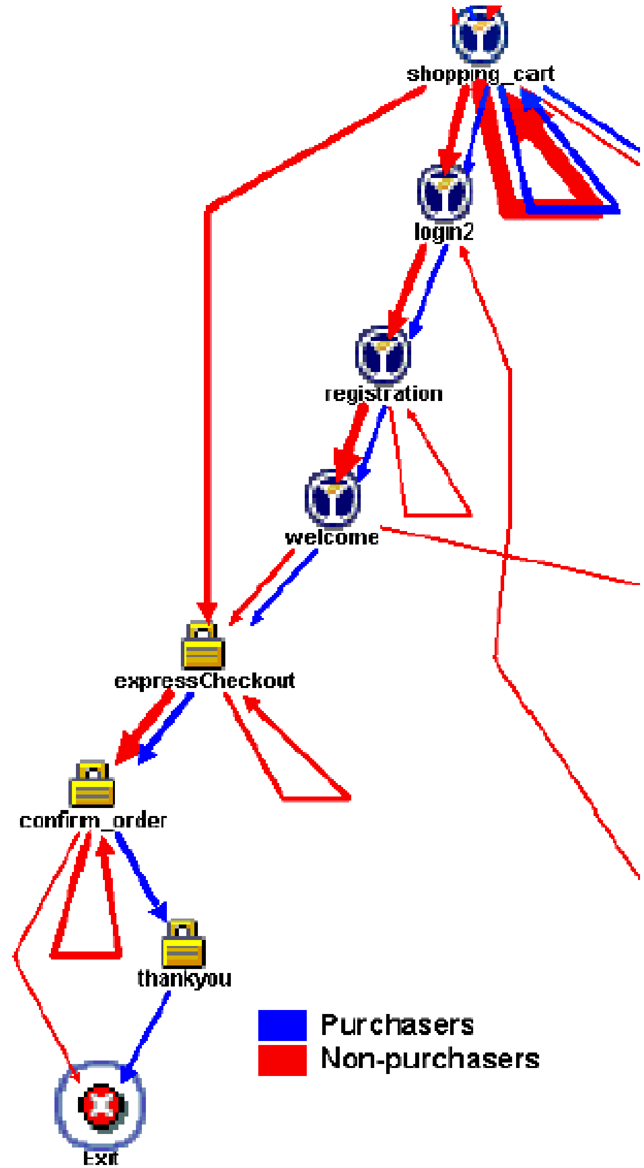
Gender shopping
differences



Initial Findings (cont)

Checkout process analysis

Newsletter hurting sales

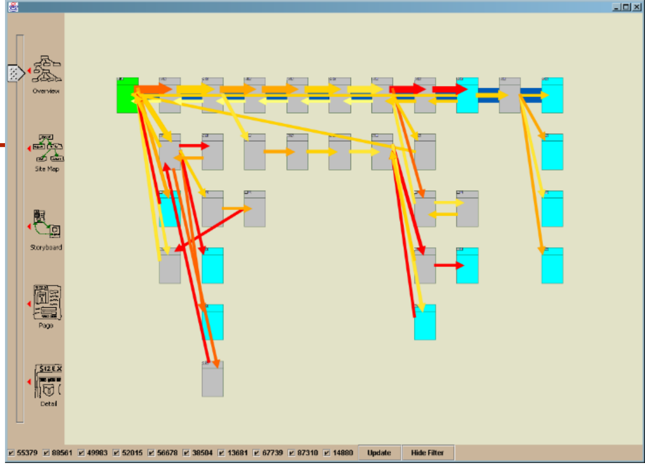


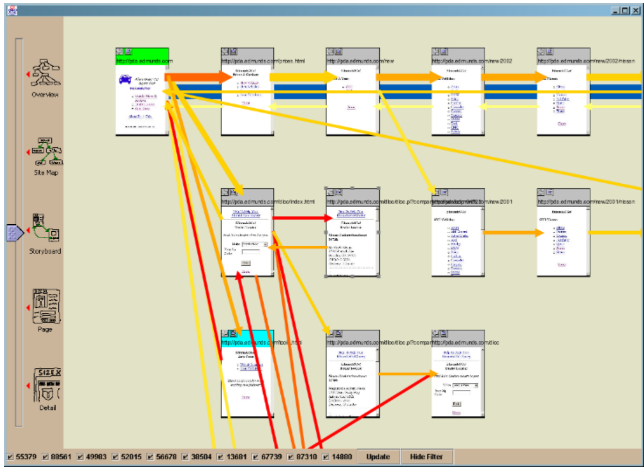
WebQuilt

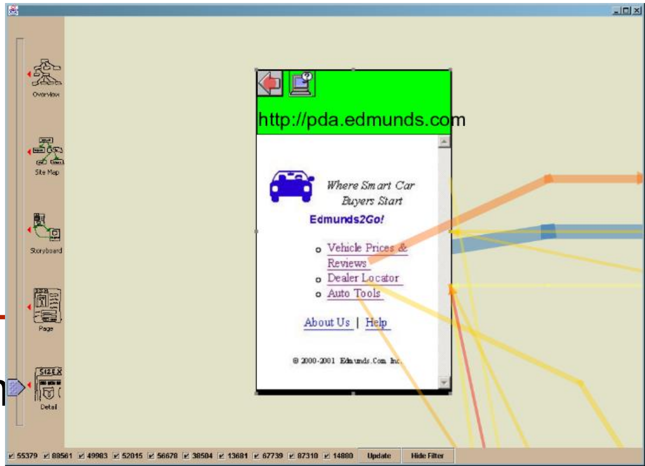
Interactive, zoomable directed graph

Nodes = web pages

Edges = aggregate traffic between pages







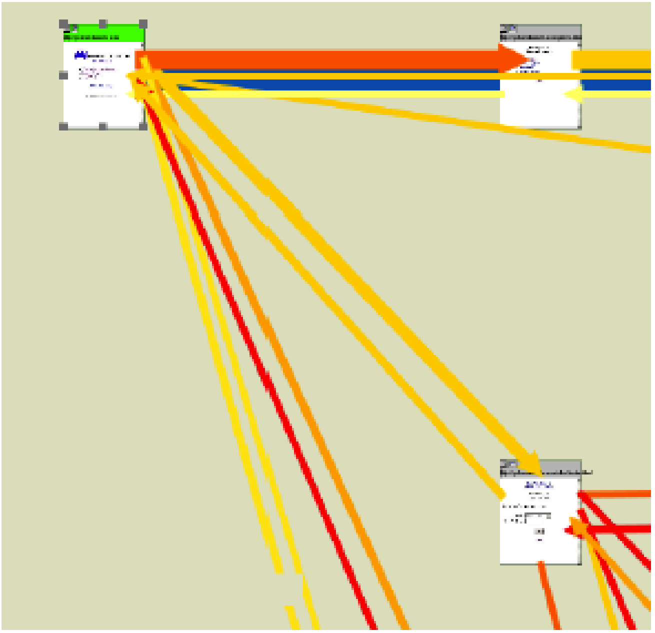
Directed graph

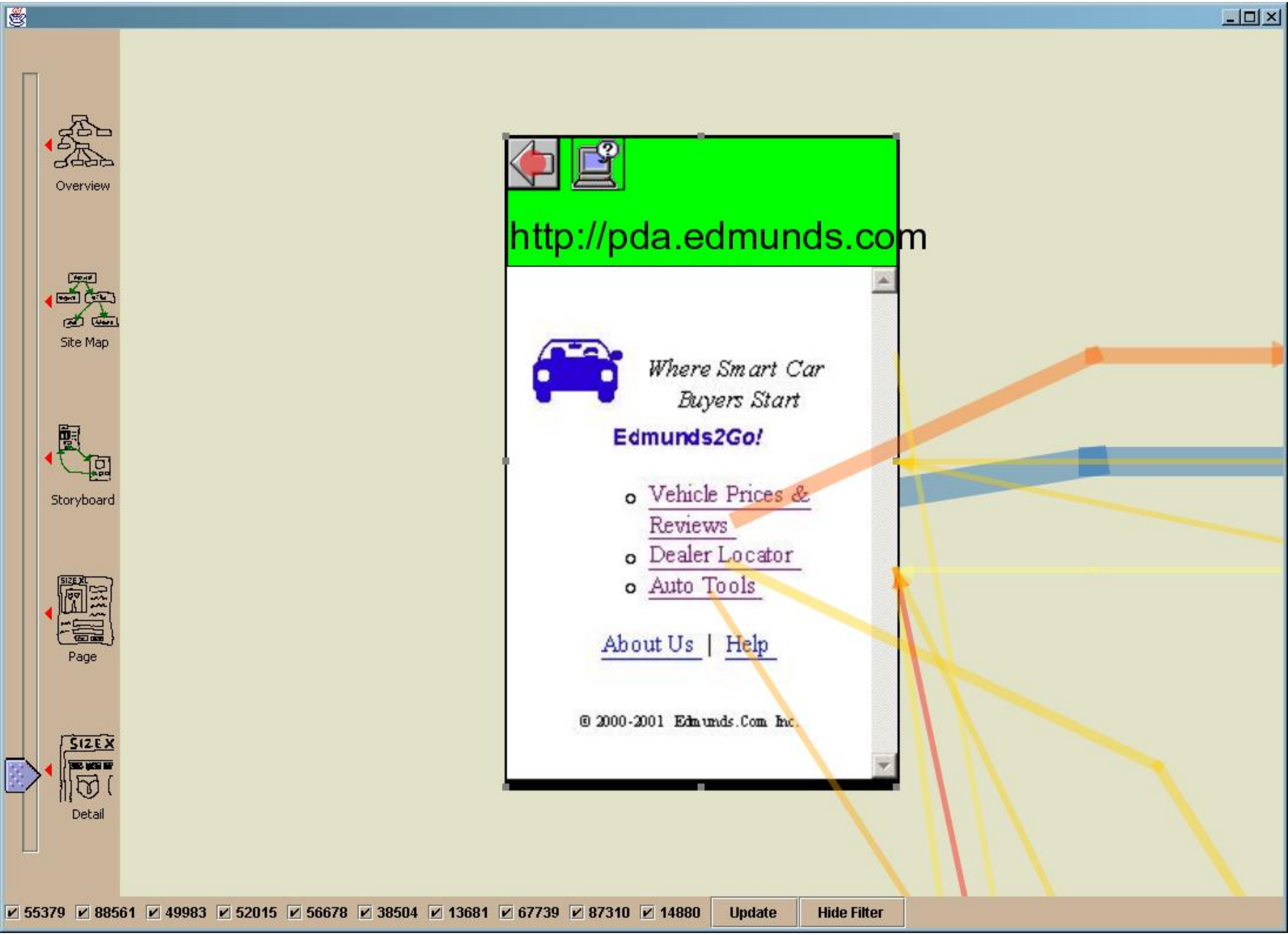
Nodes: visited pages

- 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




Overview
Site Map
Storyboard
Page
Detail
http://pda.edmunds.com
Where Smart Car
Buyers Start
Edmunds2Go!
Vehicle Prices &
Reviews
Dealer Locator
Auto Tools
About Us | Help
© 2000-2001 Edmunds.Com Inc.
55379
88561
49983
52015
56678
38504
13681
67739
87310
14880
Update
Hide Filter

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.

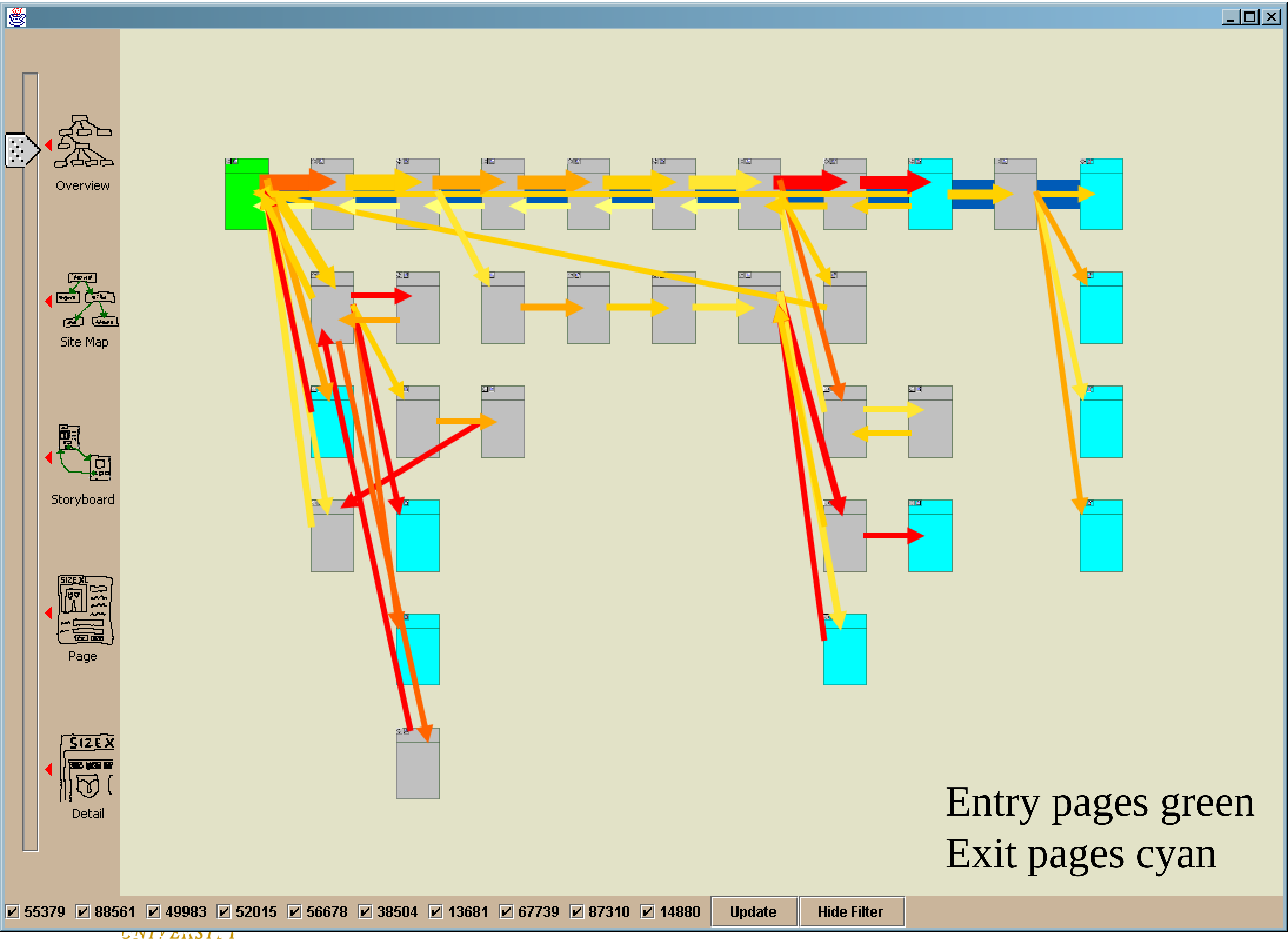
In the Lab vs. Out in the Wild

Comparing in-lab usability testing with WebQuilt remote usability testing

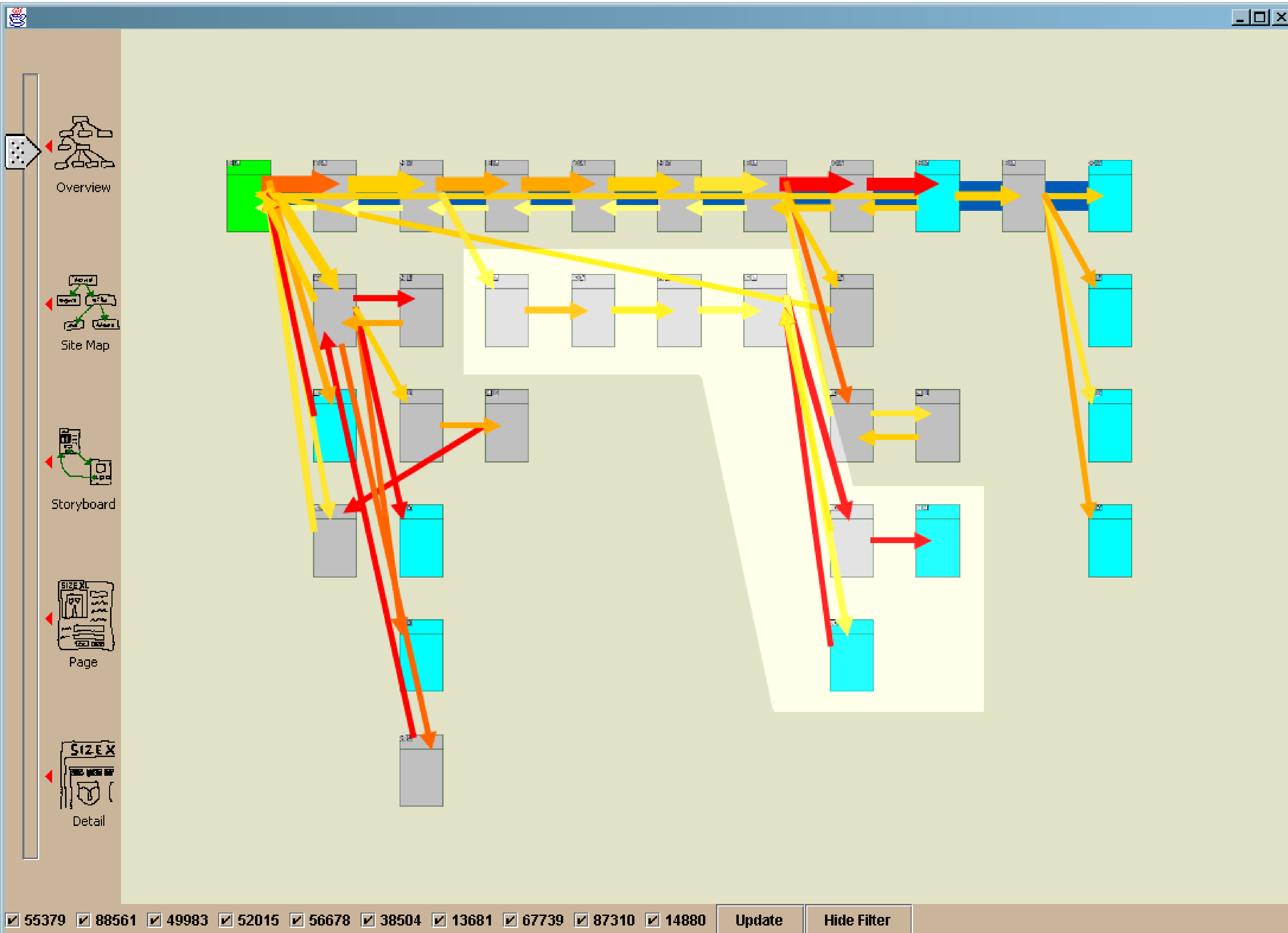
5 users were tested in the lab

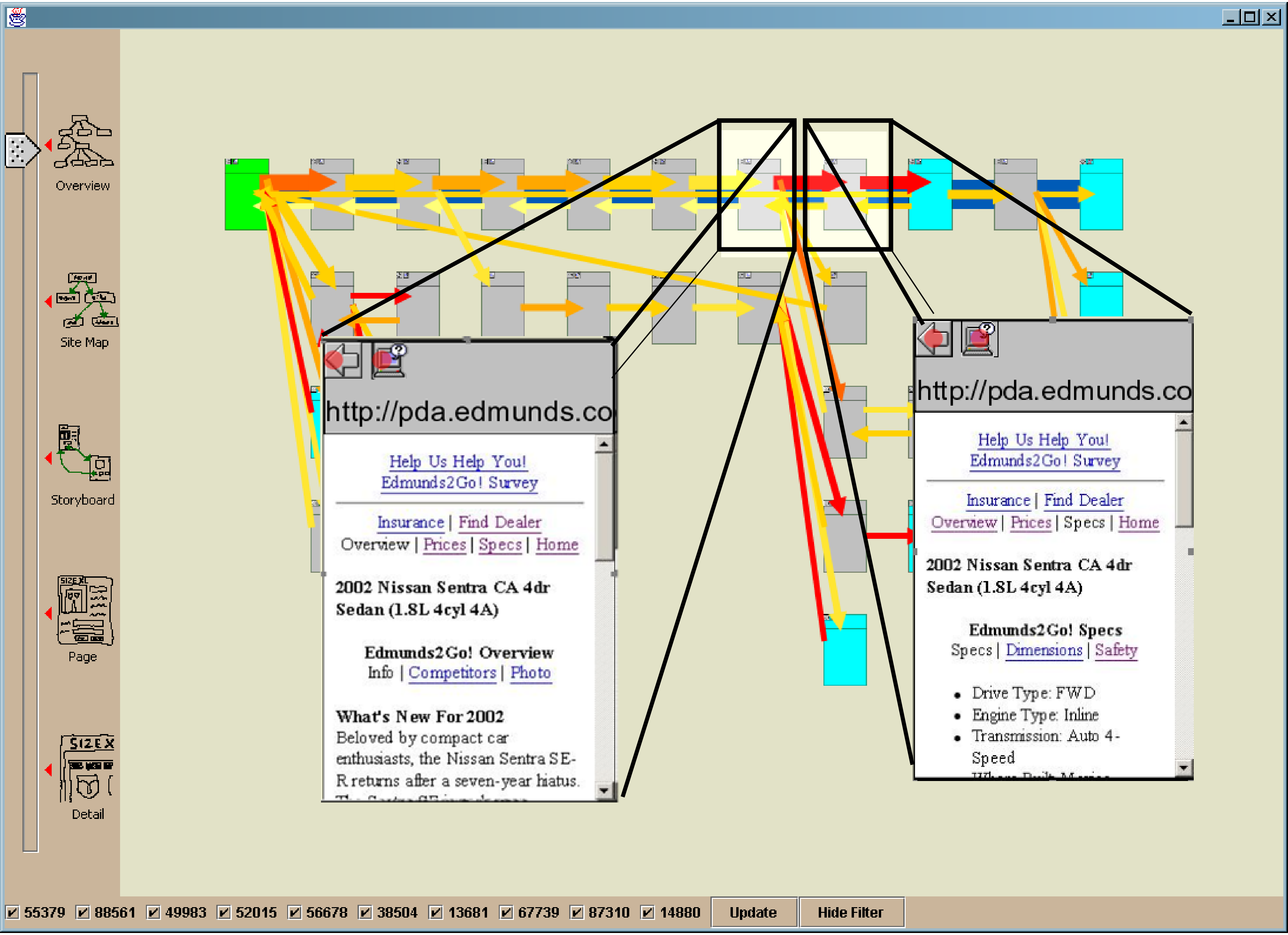
5 were given the device and asked to perform the task at their convenience

All task directions, demographic data, and follow up questionnaire data was presented and collected in web forms as part of the WebQuilt testing framework.



Entry pages green
Exit pages cyan


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
• Interact before load (3) • No forward button (2)	• Difficulty with input in questionnaire (3) • **Difficulty scrolling** (2) • Device errors unrelated to testing (1) • Tried writing on screen (0)
Site Design	**Test Design**
• **Falsely completed task** (4) • **Long download times** (4) • **Ping-pong behavior** (3) • Interact before load (3) • **Too much scrolling** (2) • Save address functionality not clear (1) • **Back button navigation** (0) • Would like more features (0) • Finds site useful (0)	• **Falsely completed task** (4) • **Difficulty remembering task description** (3) • Difficulty with input in questionnaire (3) • Questionnaire wording problems (3) • Forgot how to end task (1) • Confusing task description (1)

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

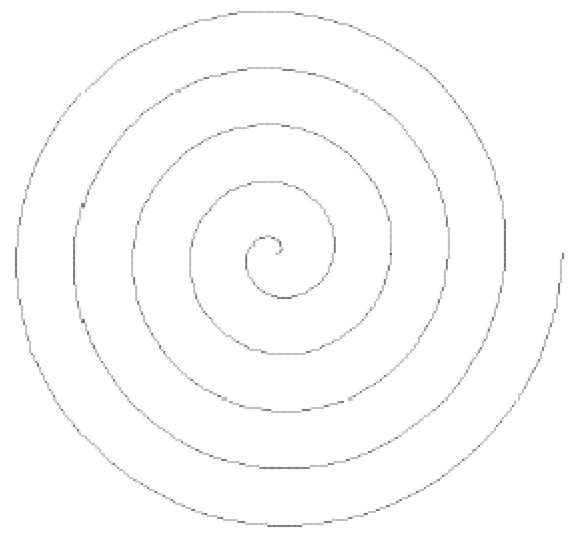


Figure 1. A Spiral of Archimedes. (for color figures see the electronic proceedings or www.cs.umn.edu/~carlis)

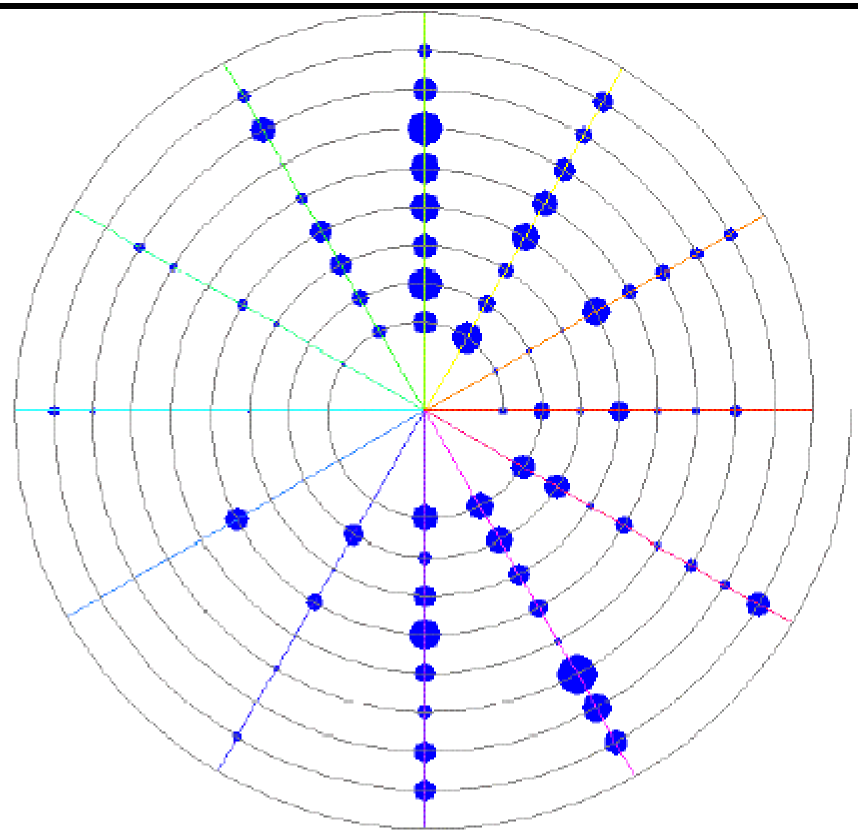


Figure 2. An indented spiral, with spokes, showing monthly consumption percentages for Baphia Capparidifolia during the period 1980 – 1988.

Baphia Cappardifolia: tropical plant

Carlis & Konstan, UIST 1998.

Analyzing Complex Periodic Data

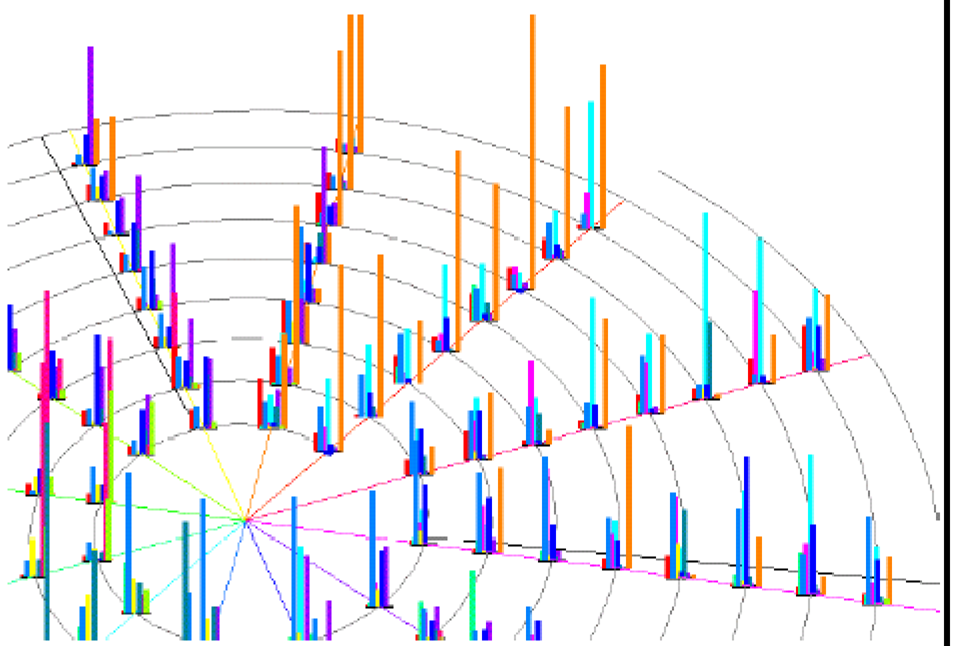


Figure 4. A spiral display of year-month consumption percentages for 12 highly consumed foods during the period 1980 – 1988. Rotated and zoomed in to show one season and boundary lines.

- 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.

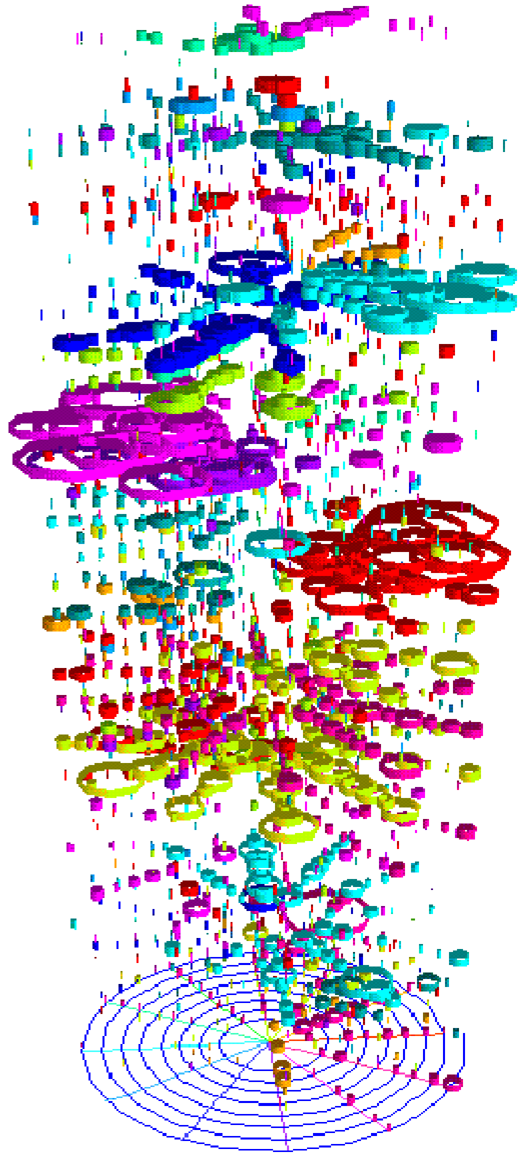


Figure 3. A spiral display of monthly consumption percentages for all 112 foods during the period 1980 – 1988.

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

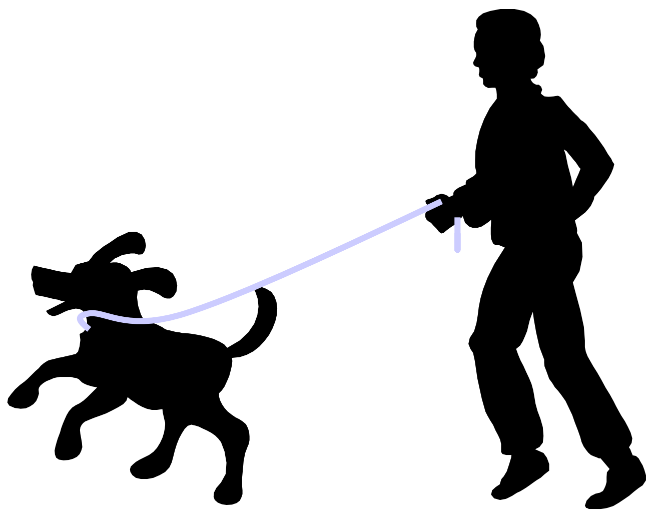
Text represents similar concepts in many different ways

- 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

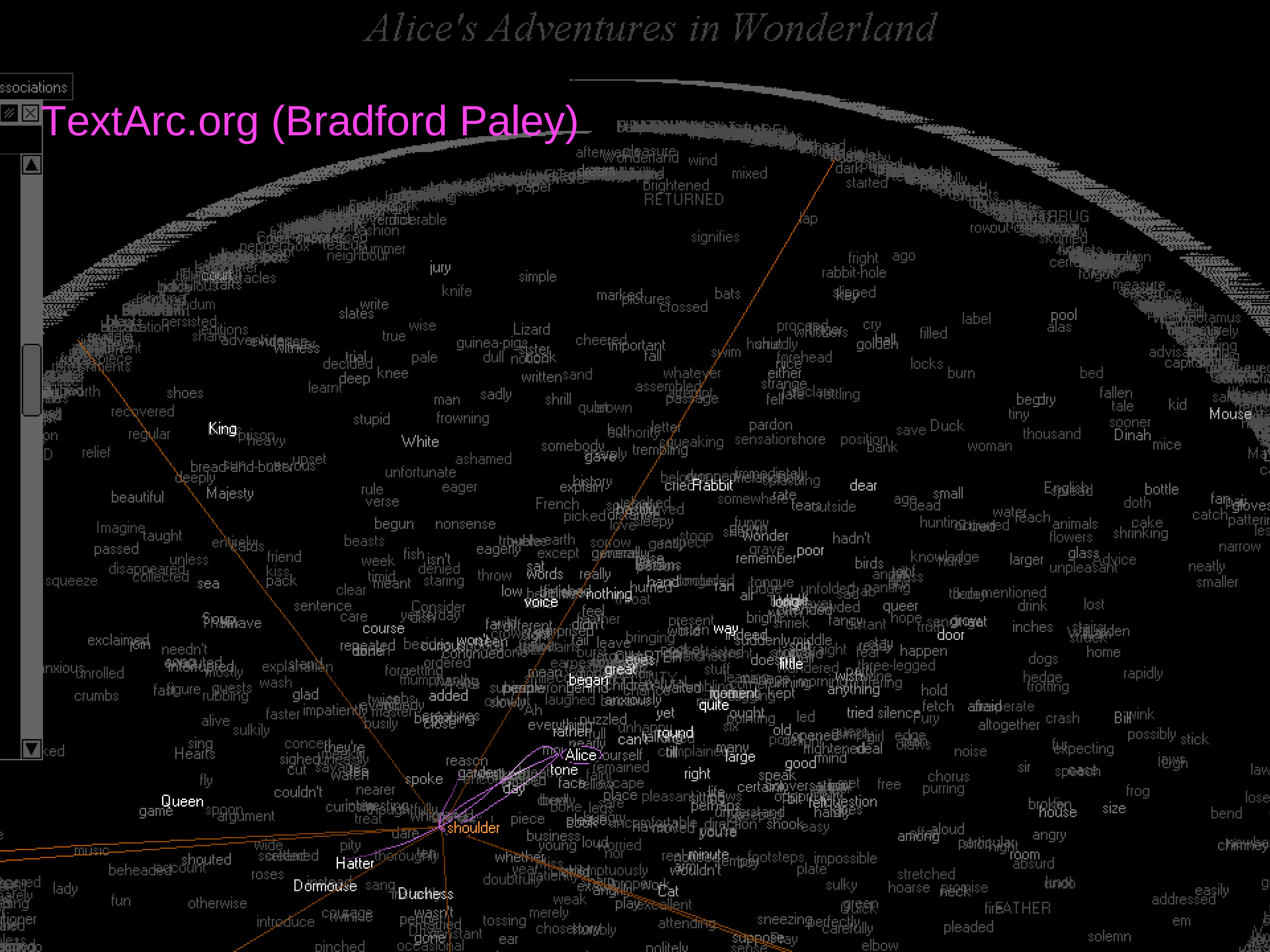
Visualization of Text Collections

How to summarize the contents of hundreds, thousands, tens of thousands of texts?

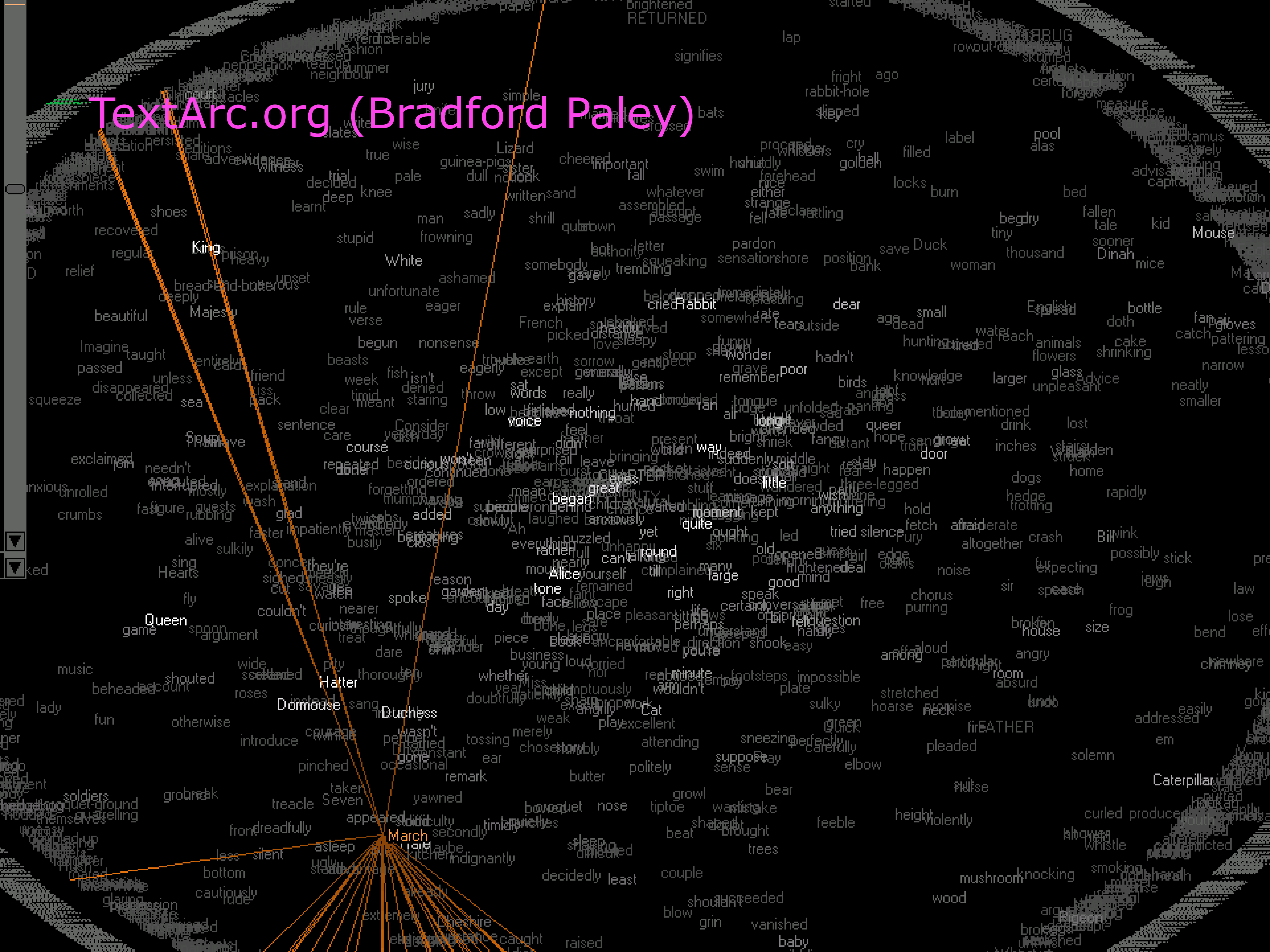
Many have proposed clustering the words and showing points of light in a 2D or 3D space.

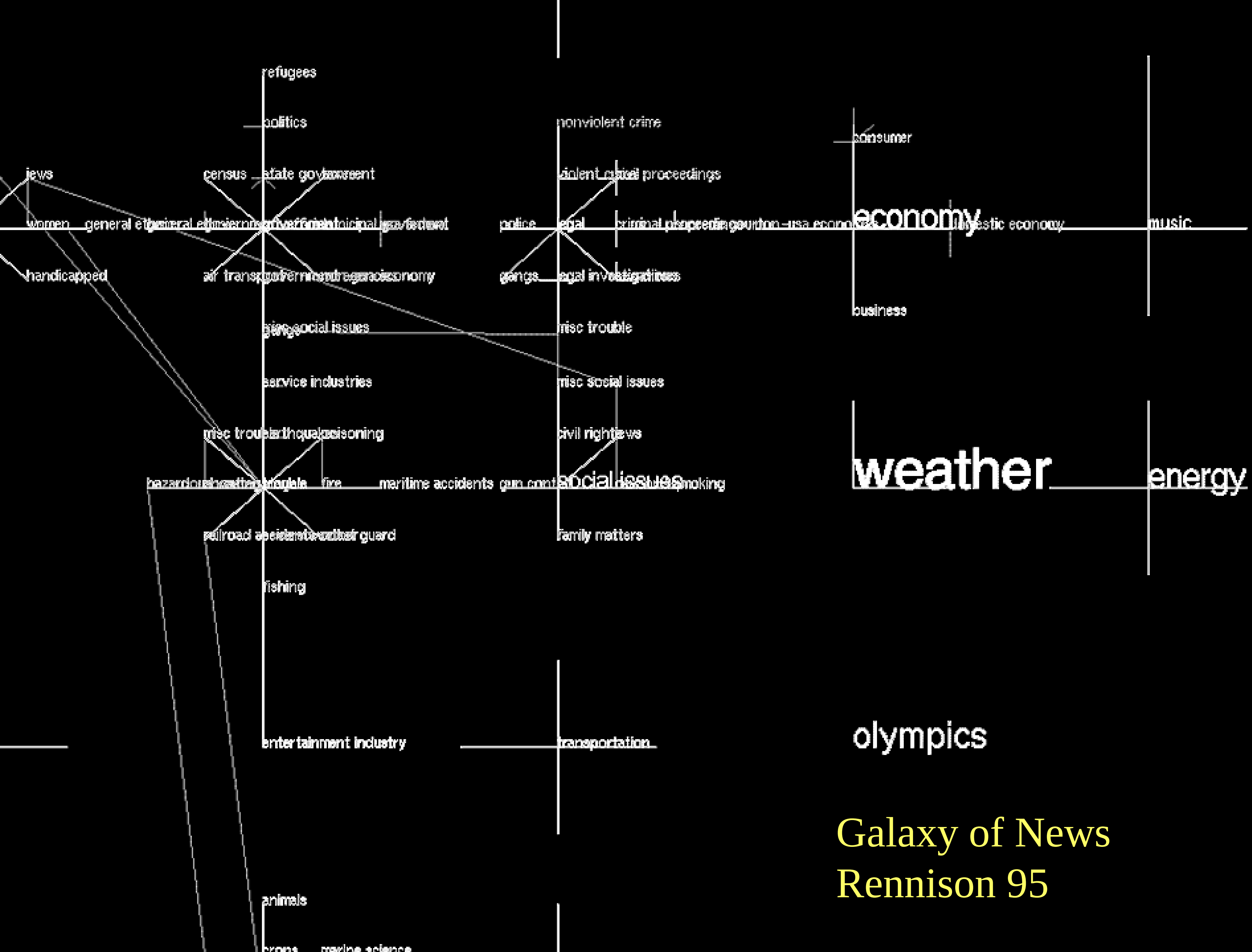
Examples

- Showing docs/collections as a word space
- Showing retrieval results as points in word space

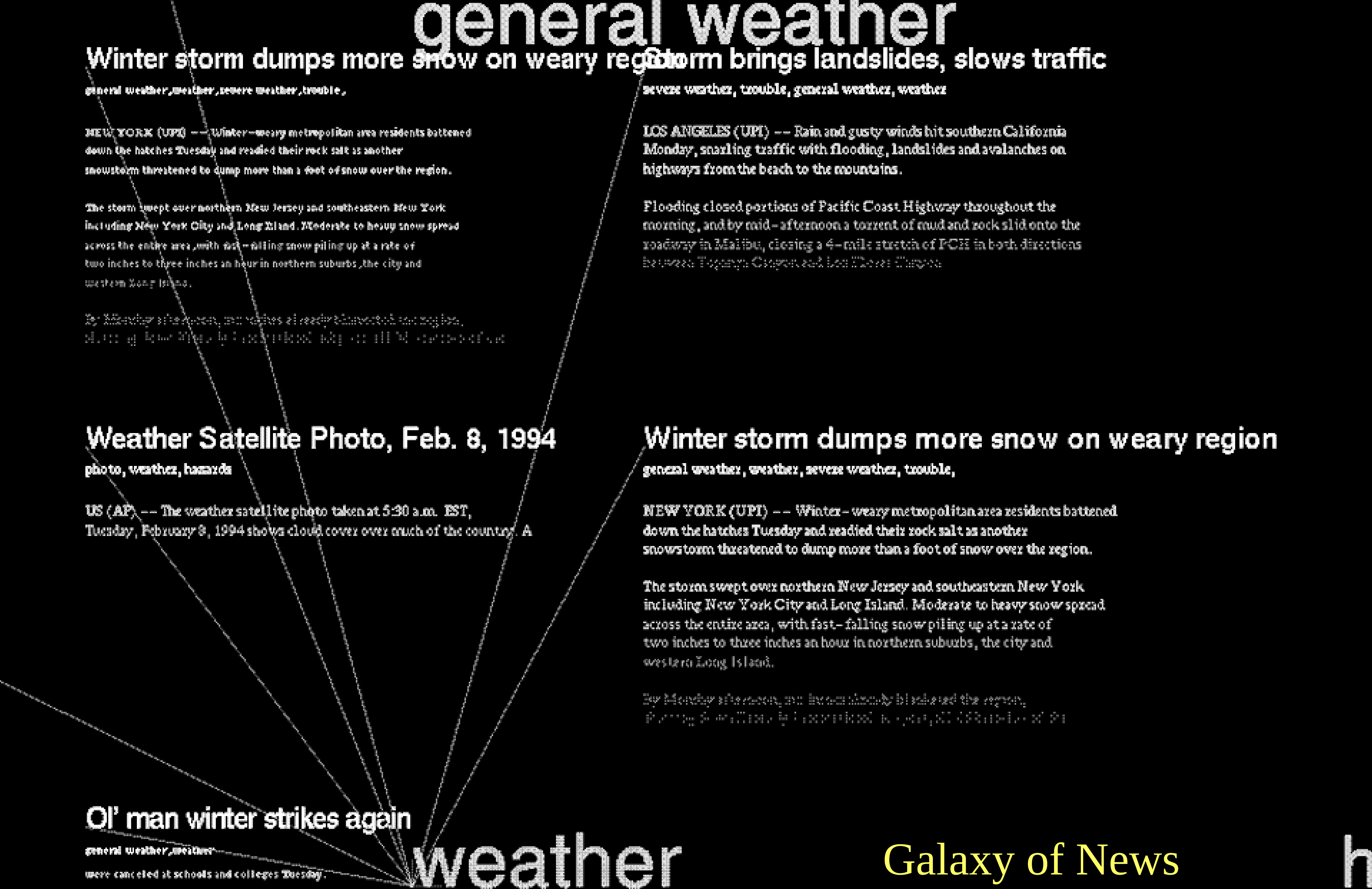


TextArc.org (Bradford Paley)


TextArc.org (Bradford Paley)
King
White
Majesty
Rabbit
dear
Mouse
Dinah
Duck
voice
nothing
course
began
great
quite
Alice
tone
right
large
good
Queen
Hatter
Dormouse
Duchess
Cat
Caterpillar
March
little
door
way

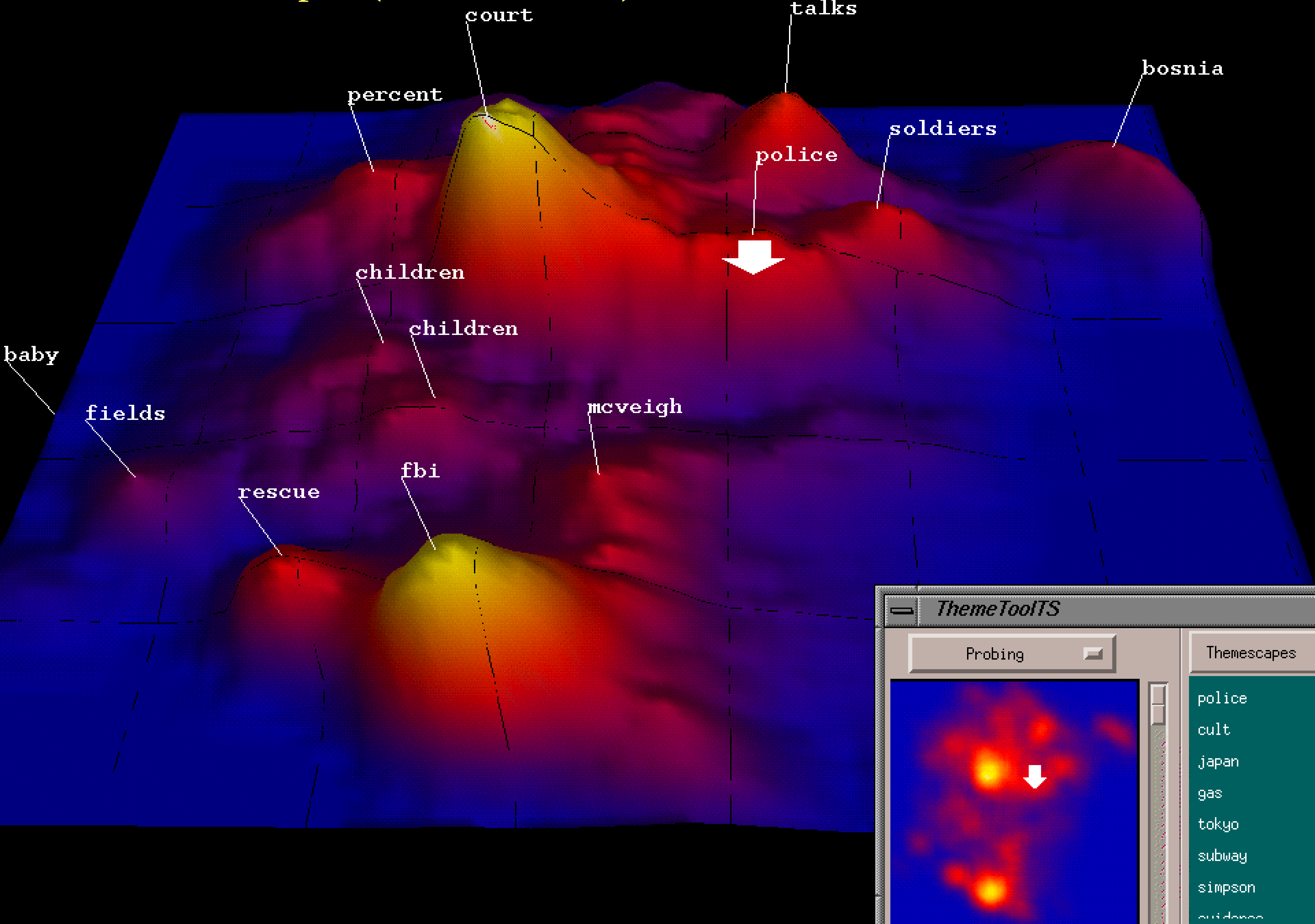


Galaxy of News
Rennison 95



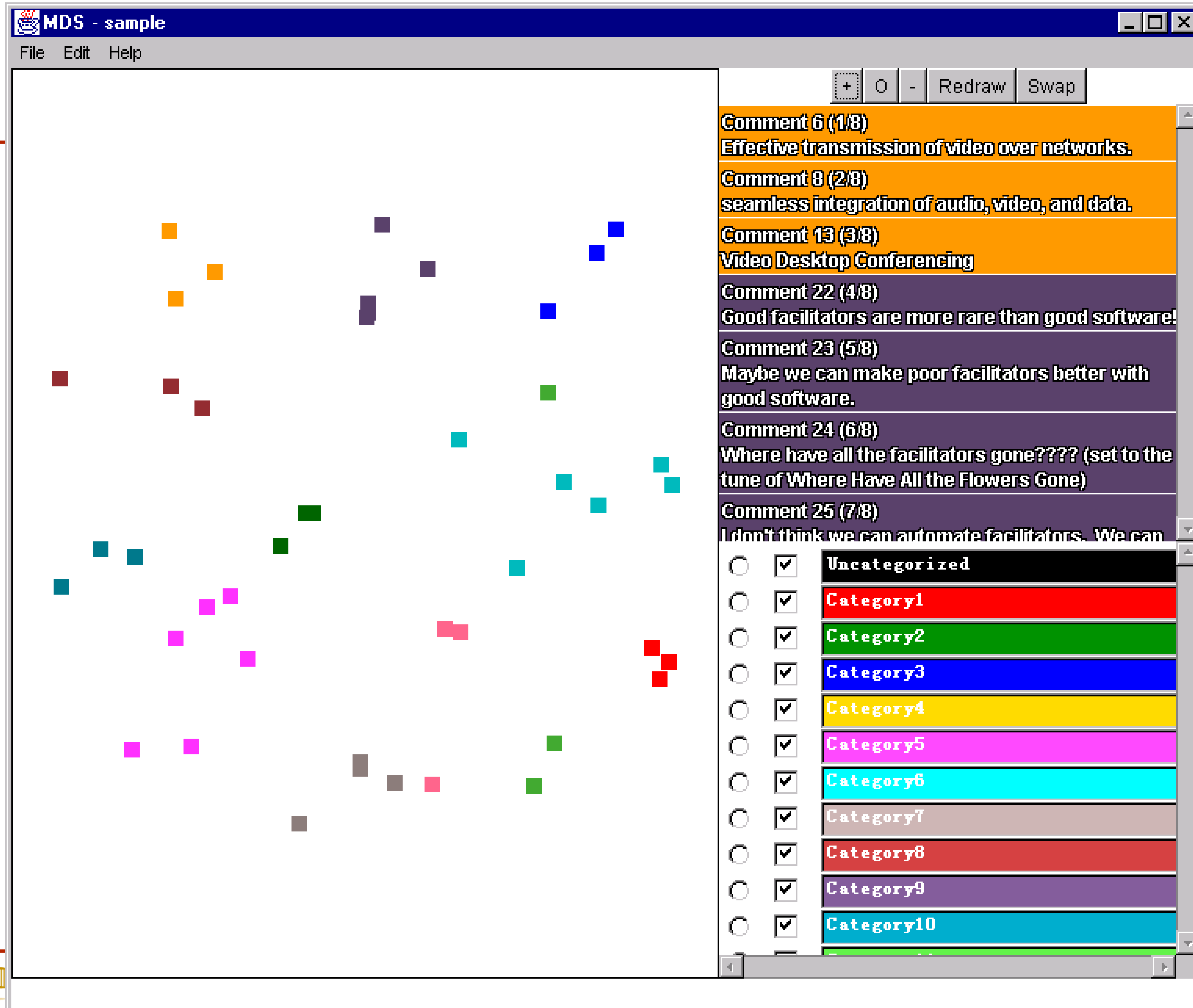
Galaxy of News
Rennison 95

Themescapes (Wise et al. 95)

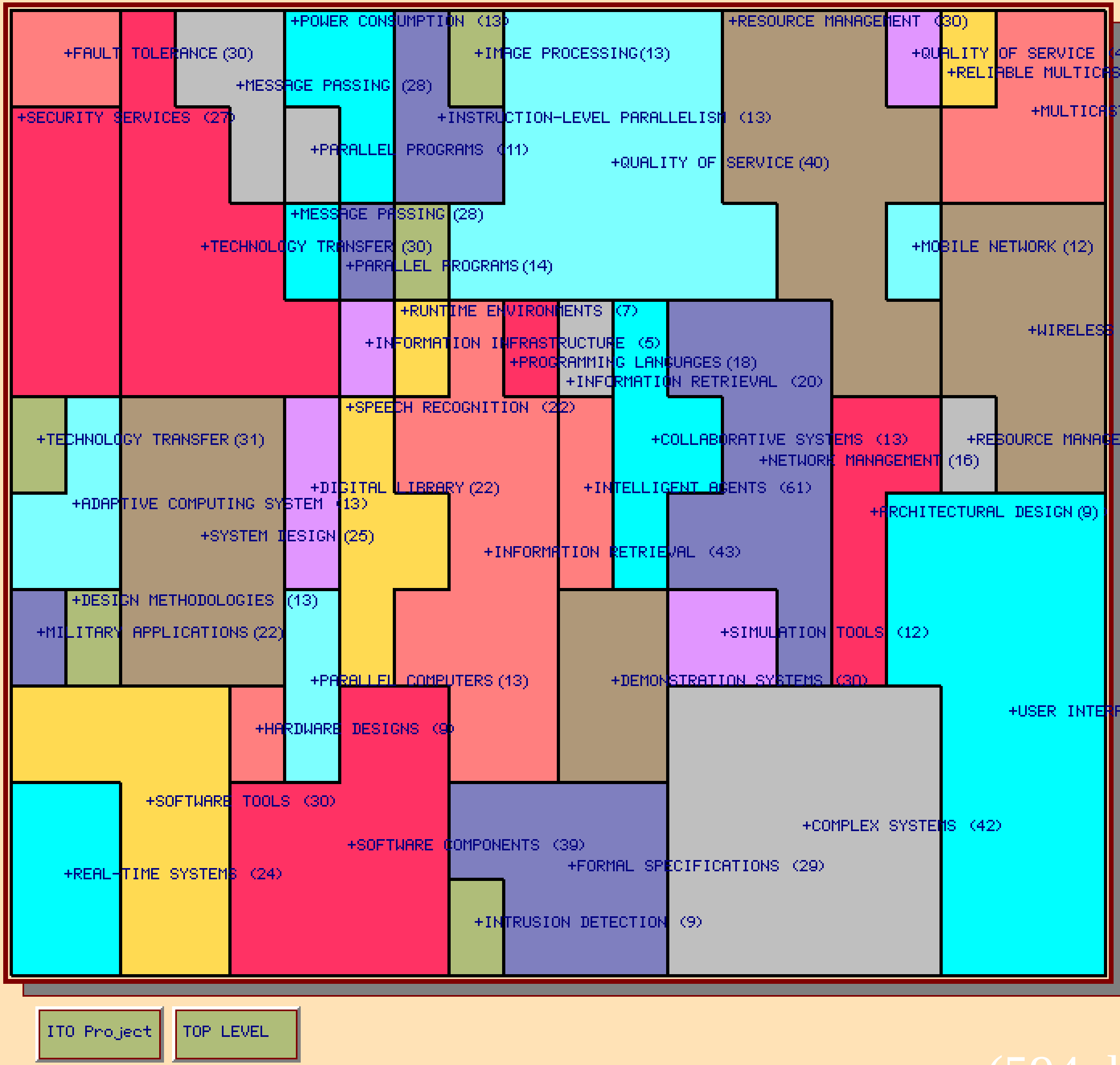


ScatterPlot of Clusters

(Chen et al. 97)



Kohonen Feature Maps
(Lin 92, Chen et al. 97)



(594 docs)

WRIGHT UNIVER

Clustering for Collection Overviews

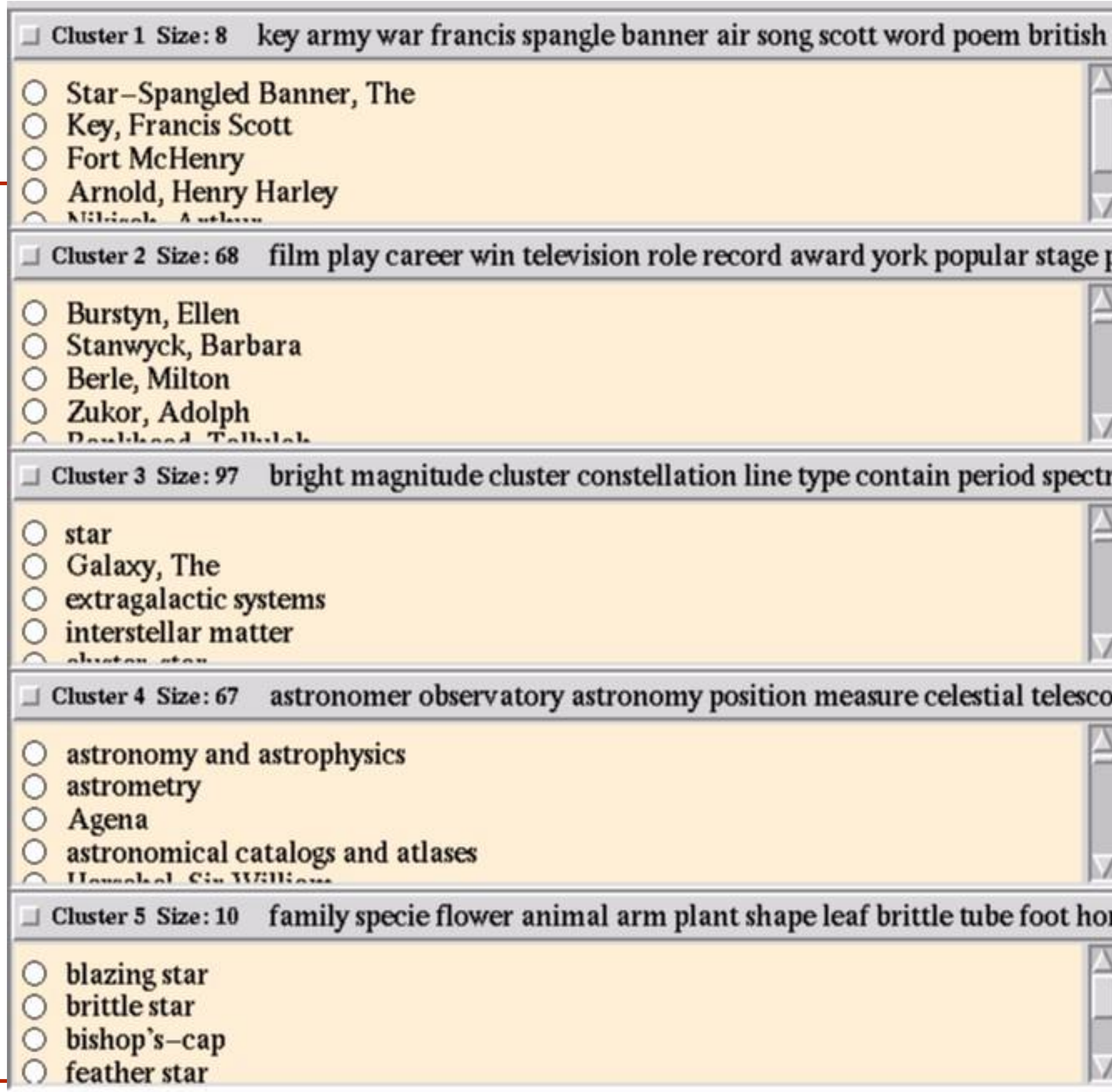
Two main steps

- 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



Cutting, Pedersen, Tukey & Karger 92, 93, Hearst & Pedersen 95

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

i.e. self organizing map for entertainment sub-category of 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)

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.128&rep=rep1&type=pdf

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.)

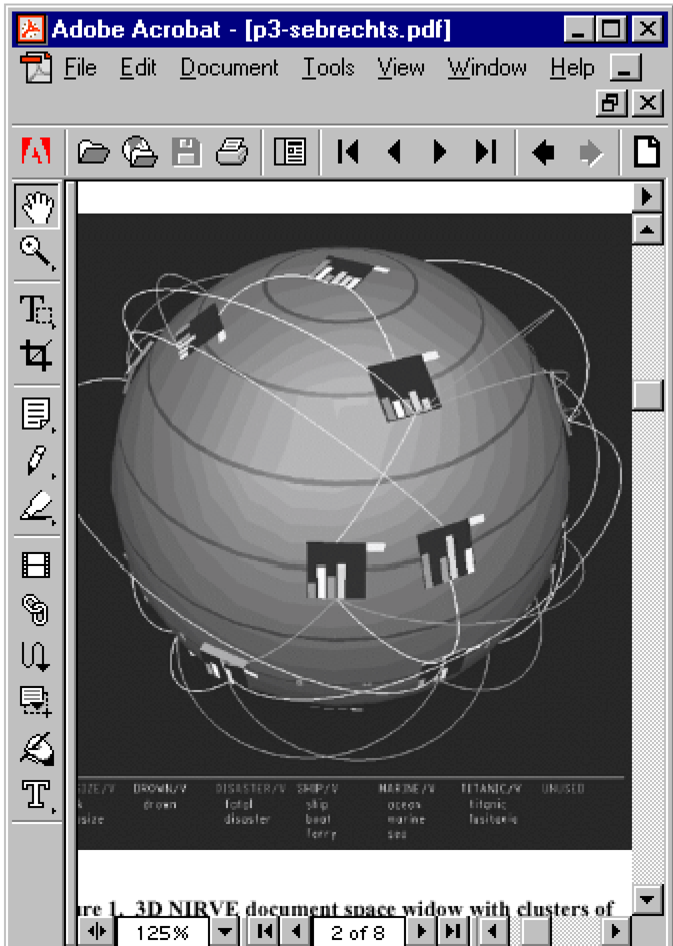
Participants wanted:

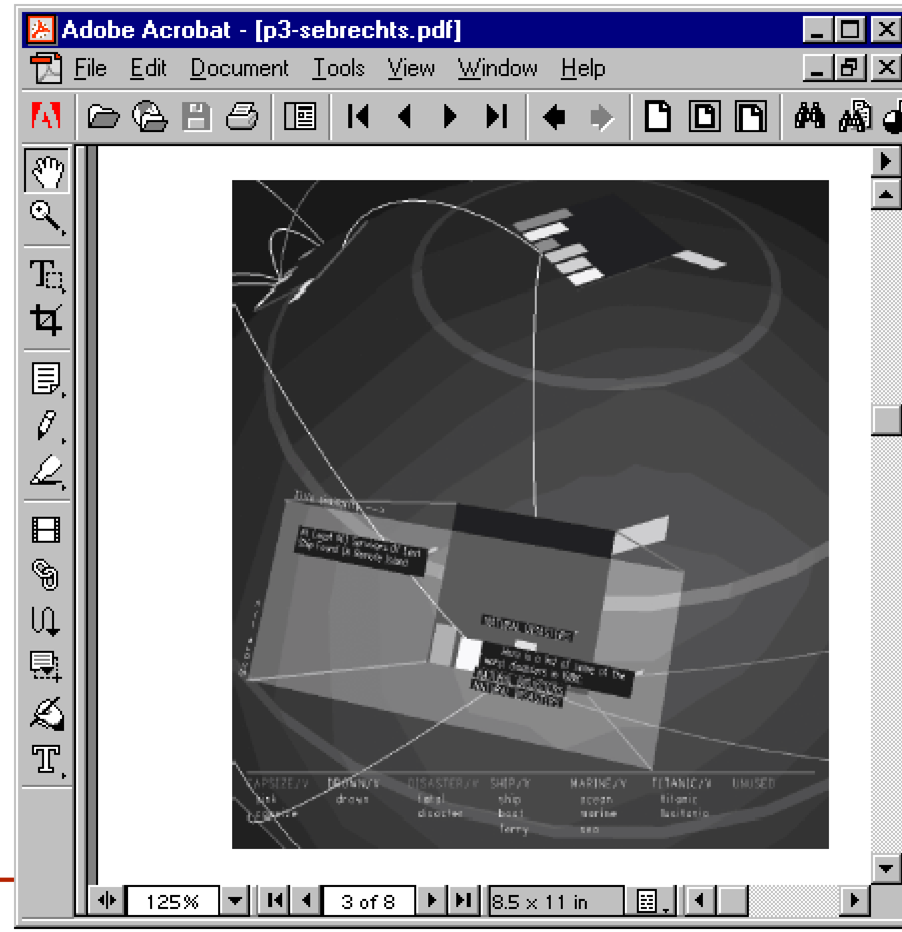
- hierarchical organization
- other ordering of concepts (alphabetical)
- integration of browsing and search
- correspondence of color to meaning
- more meaningful labels
- labels at same level of abstraction
- fit more labels in the given space
- combined keyword and category search
- multiple category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

http://zing.ncsl.nist.gov/cugini/uicd/nirve-paper.html





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

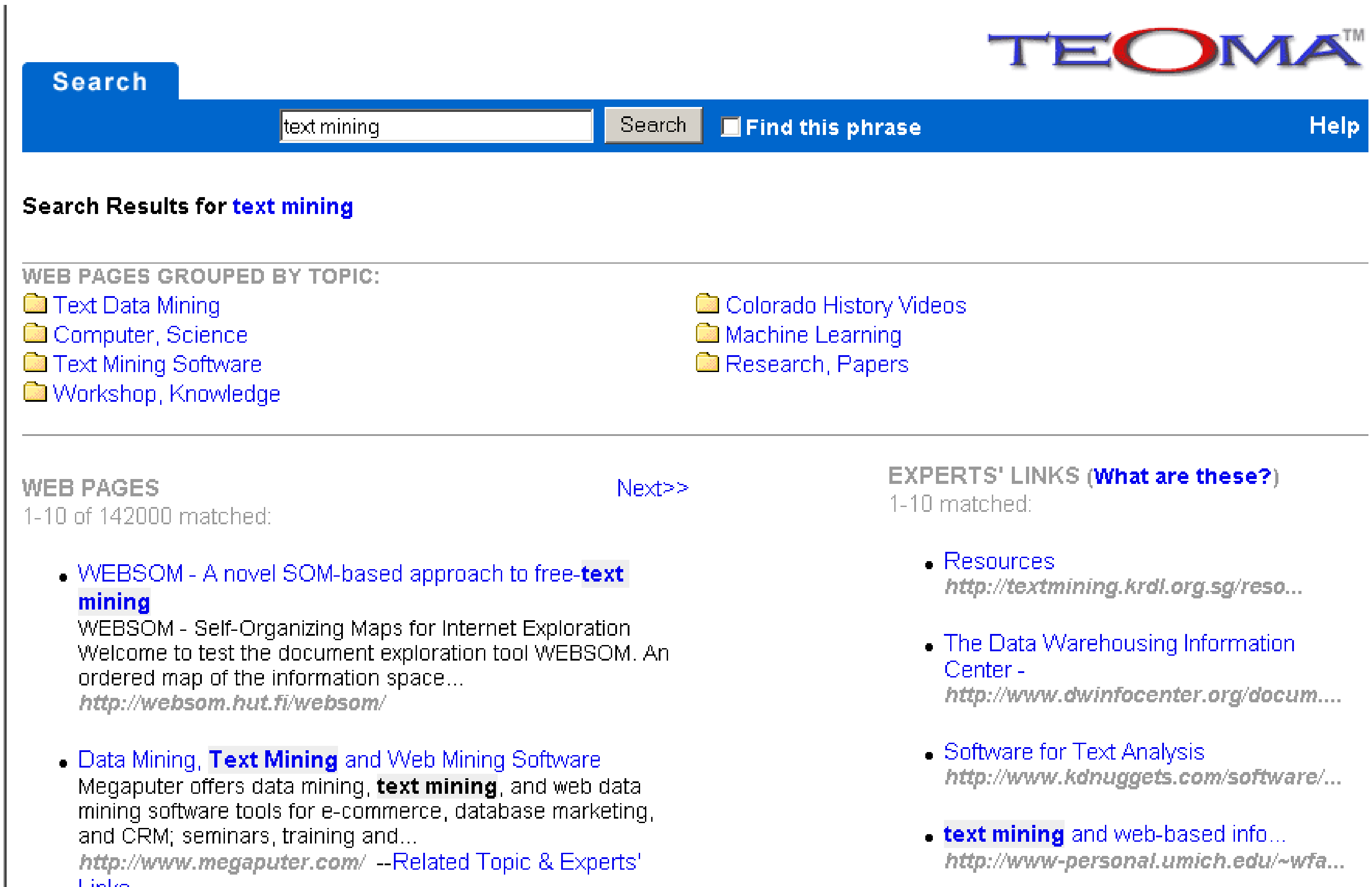
Note: most don't benefit from visualization!

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

Teoma: appears to combine categories and clusters

(this version before it was bought by askjeeves)



Teoma: Now in prime time

TEOMA

Search Tips

text mining | Search

• 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/

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...
[Related Pages]

Hearst, Marti
Untangling **Text** Data **Mining**, ACL'99.
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/
[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/
[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

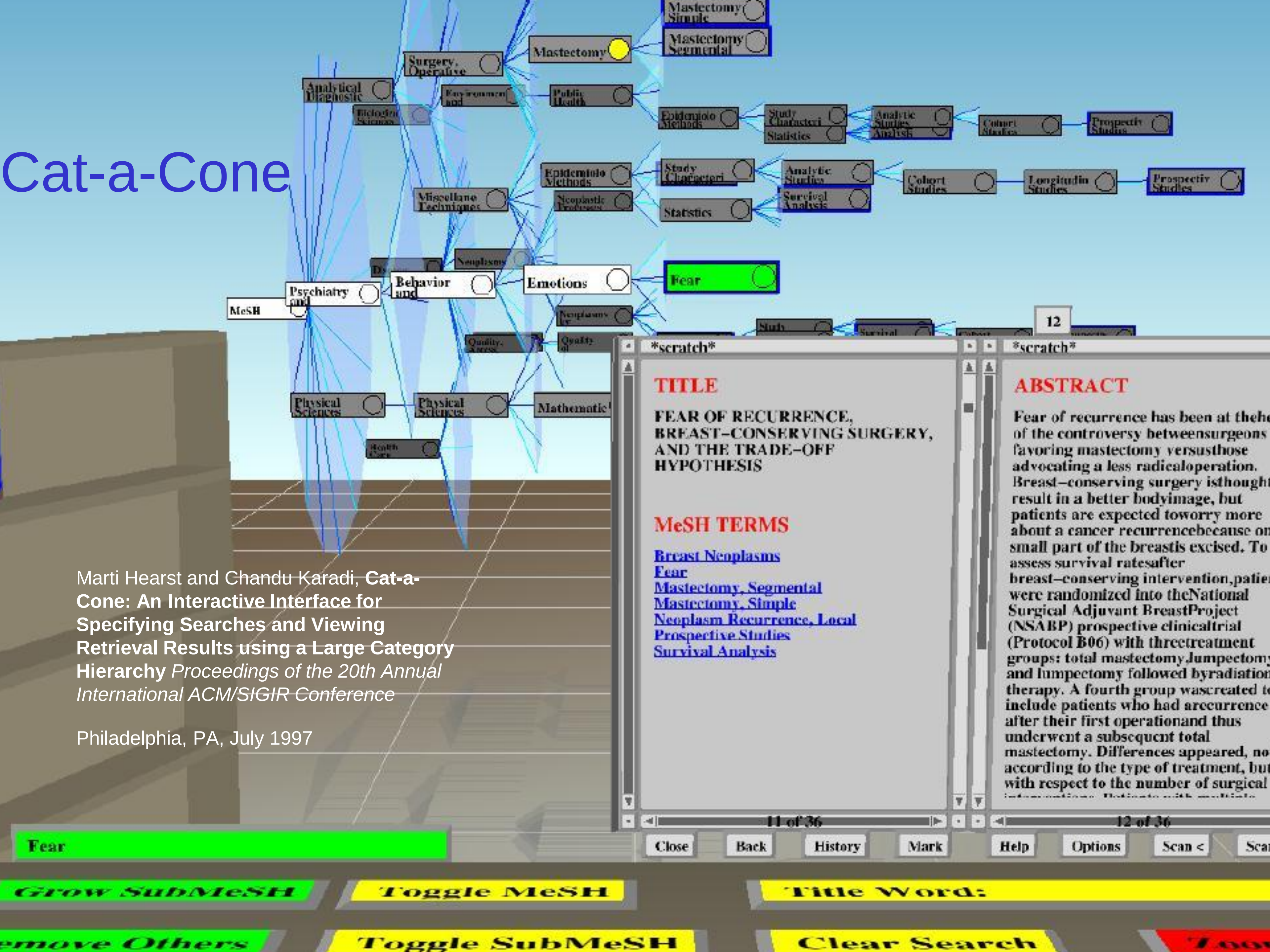
A Roadmap to **Text Mining** and Web **Mining**
www.cs.utexas.edu/...

Resources
textmining.krdl.org.sg/...

Mary D. Taffet's Home Page: WWW Sites for Students...
web.syr.edu/...

Text Analysis, **Text Mining** and Information Retrieval

Cat-a-Cone



Marti Hearst and Chandu Karadi, **Cat-a-Cone: An Interactive Interface for Specifying Searches and Viewing Retrieval Results using a Large Category Hierarchy** *Proceedings of the 20th Annual International ACM/SIGIR Conference*

Philadelphia, PA, July 1997

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

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 ☒

start a new search

Search ◉ all items ○ within current results

89 items (grouped by media)

view ungrouped items

aquatint 3

Distant View of ...
Daniell
1817

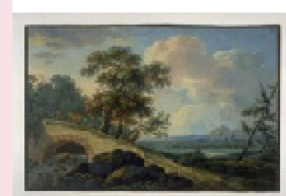
Verlassen (Aband...
Klinger
1884

Village au Bord ...
Hervier
19th century

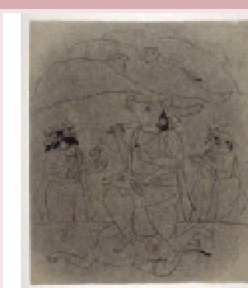
drawing 15

A Lady at a Shrine
Anonymous

A View of the No...
Eustache



Alsatian Landscape
Anonymous



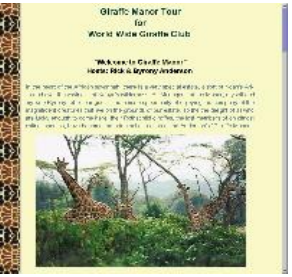
Boar Avatar of V...
Anonymous

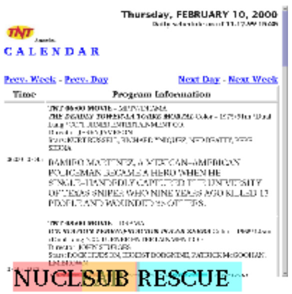
Using Thumbnails to Search the Web

A. Woodruff, R. Rosenholtz, J. Morrison, A. Faulring, & P. Pirolli, A comparison on the use of text summaries, plain thumbnails, andenhanced 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

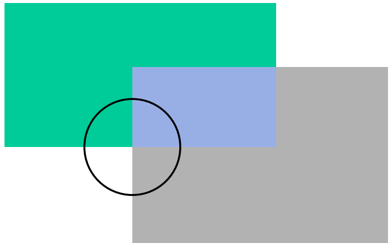
- Text summaries
  - Lots of abstract, semantic information
- Image summaries (plain thumbnails)
  - Layout, genre information
  - Gist extraction faster than with text
- Benefits are complementary
- Create textually-enhanced thumbnails that leverage the advantages of both text summaries and plain thumbnails

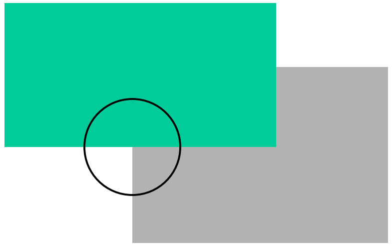
Putting Callouts in a Separate Visual Layer

Transparency

Occlusion

Junctions indicate the occurrence of these events.





Design Issues:

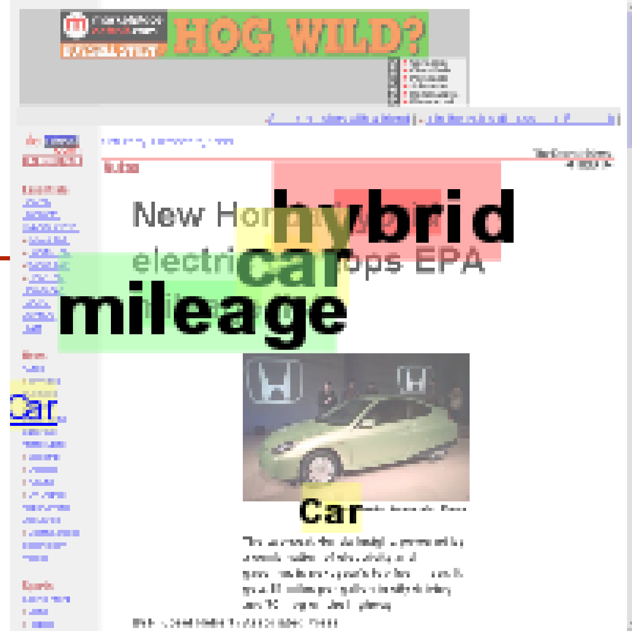
Color Management

- 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

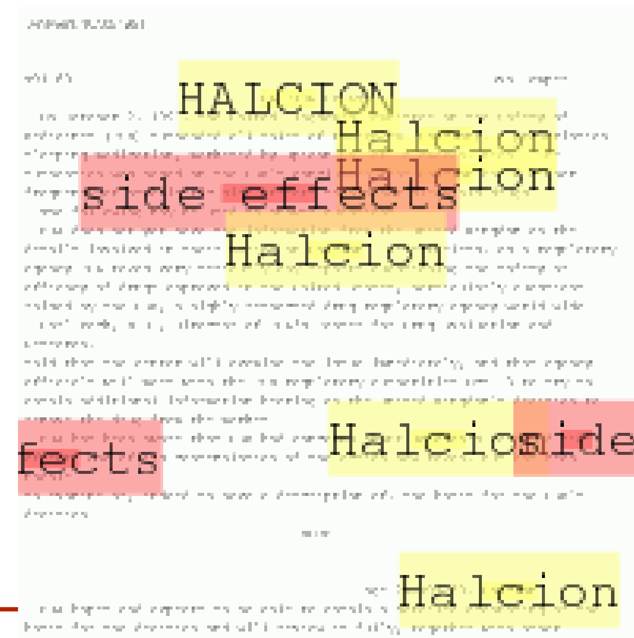
- Problem: We want to make certain text elements readable, but not necessarily draw attention to them
- Solution: Modify the HTML before rendering the thumbnail

Examples


HOG WILD?
hybrid
car
mileage
Car
Car


MiniDisc Player
MiniDisc Player


Hellerstein


HALCION
Halcion
Halcion
side effects
Halcion
Halcion
side
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

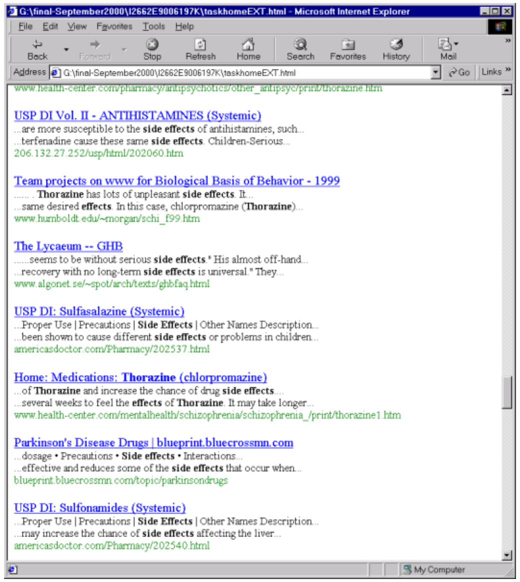
The Lycaeum -- GHB
......seems to be without serious **side effects**." His almost off-hand...
...recovery with no long-term **side effects** is universal." They...
www.algonet.se/~spot/arch/texts/ghbfaq.html

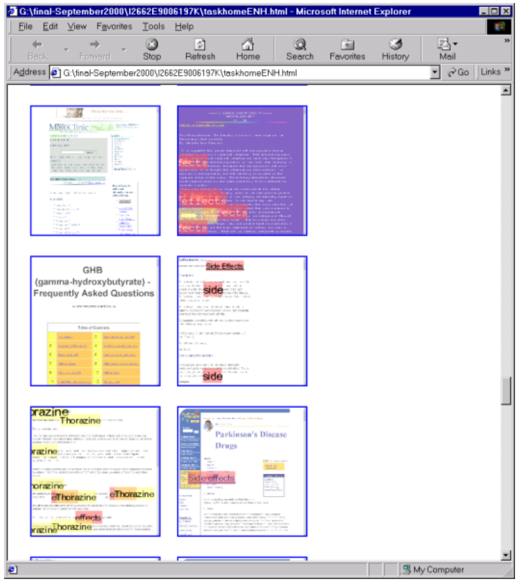
- Text summary
  - Page title
  - Extracted text with query terms in bold
  - URL
- Plain thumbnail
- Enhanced thumbnail
  - Readable H1, H2 tags
  - Highlighted callouts of query terms
  - Reduced contrast level in thumbnail

Collections of Summaries

100 results in random order

Approximately same number of each summary type on a page





Method

Procedure

18 questions, with 100 query results each

Entire process took about 75 minutes

- 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

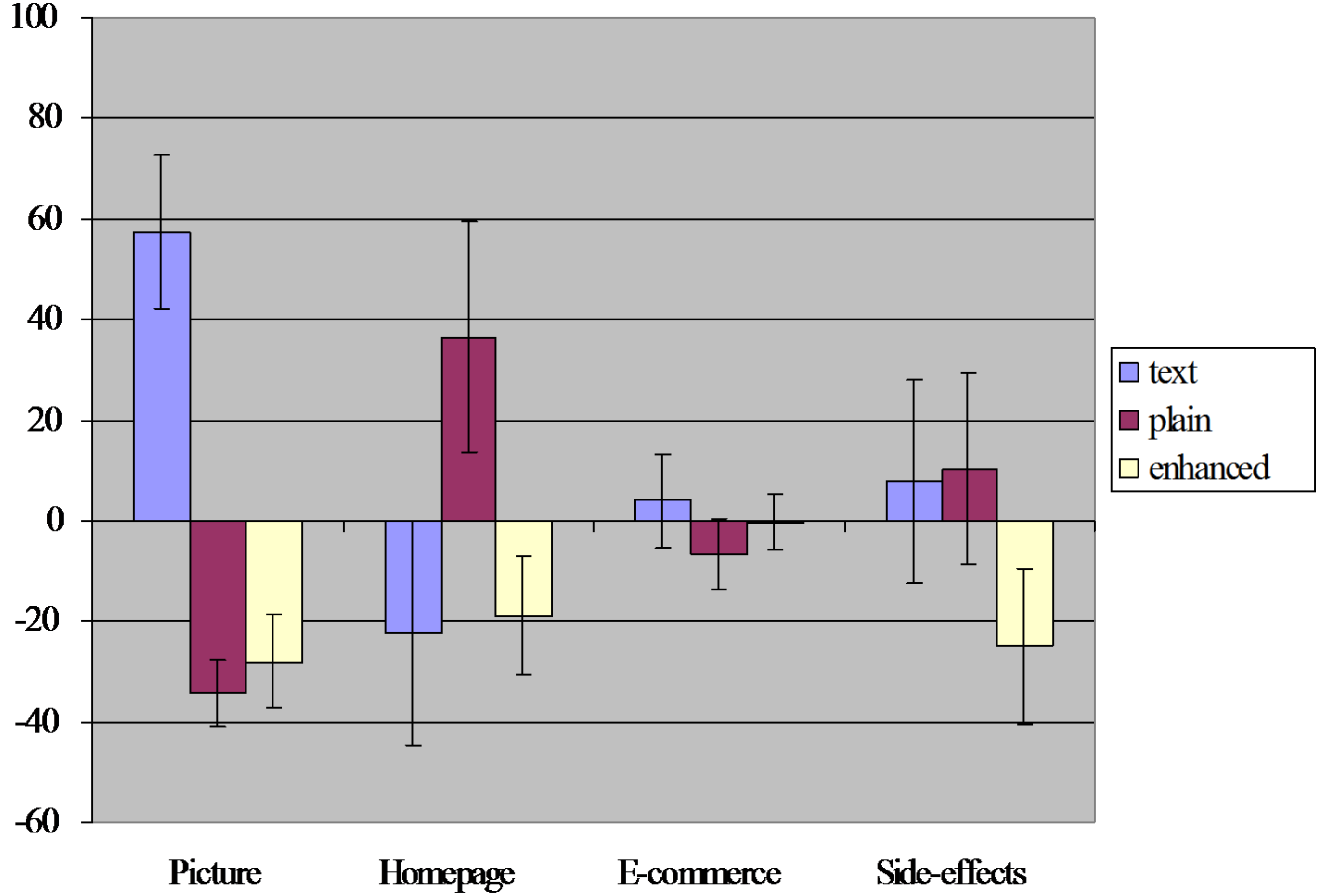
Average total search times, by task:

- 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



100
80
60
40
20
0
-20
-40
-60
Picture
Homepage
E-commerce
Side-effects
text
plain
enhanced

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 boy-friends?" (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.)

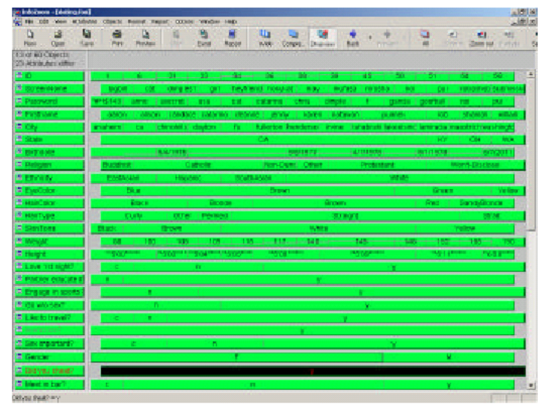


Figure 1b. This screenshot shows one possible way for solving the same problem in InfoZoom, specifically in its overview mode. After clicking at, and thereby zooming into, the "Yes" entries in the attribute "Did you cheat?", users can see from the length of the bars in the Gender category that females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

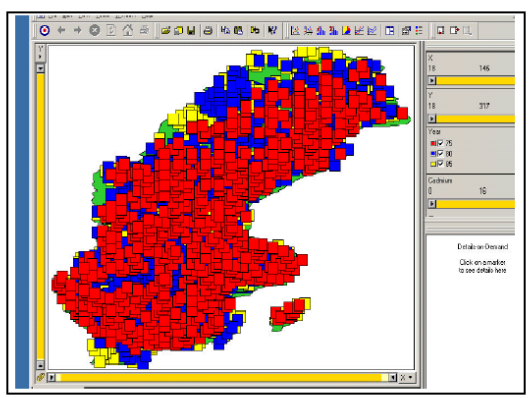


Figure 2. Spotfire's geographical representation of heavy metal concentrations through a scatterplot diagram.

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 boy-friends?" (an actual ques-tion used in the experi-ment). After grouping the attribute "Gender" and sort-ing the column "Did you cheat?," a user can com-pare 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)

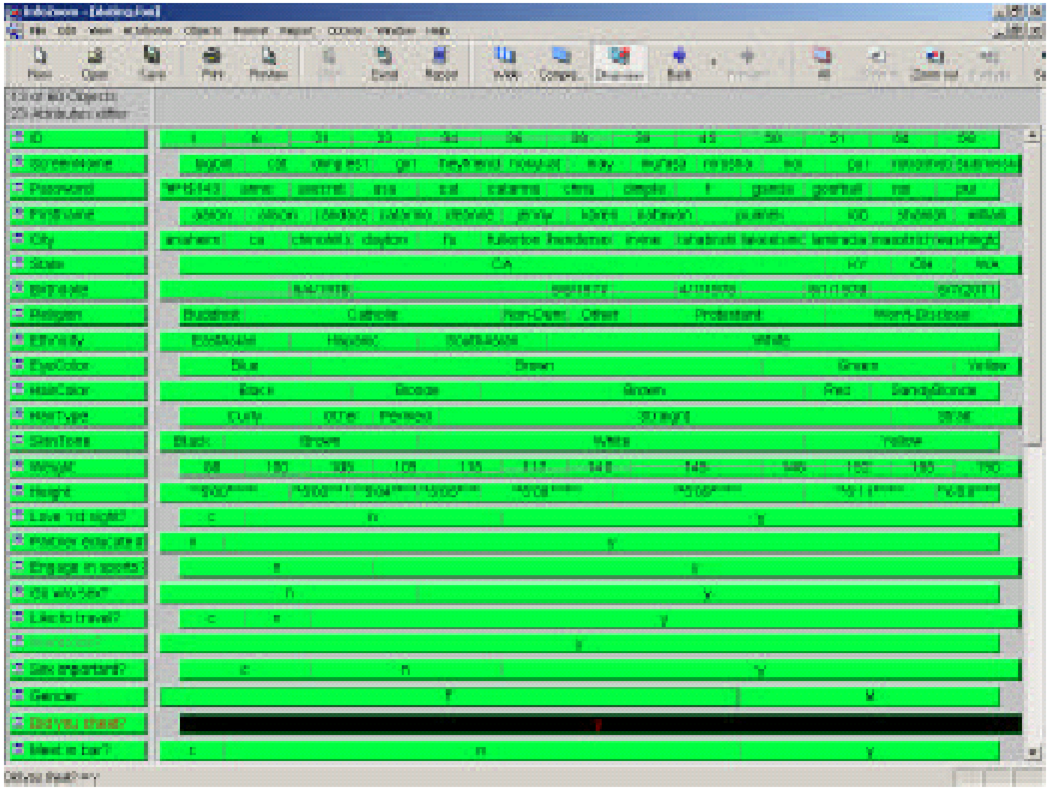
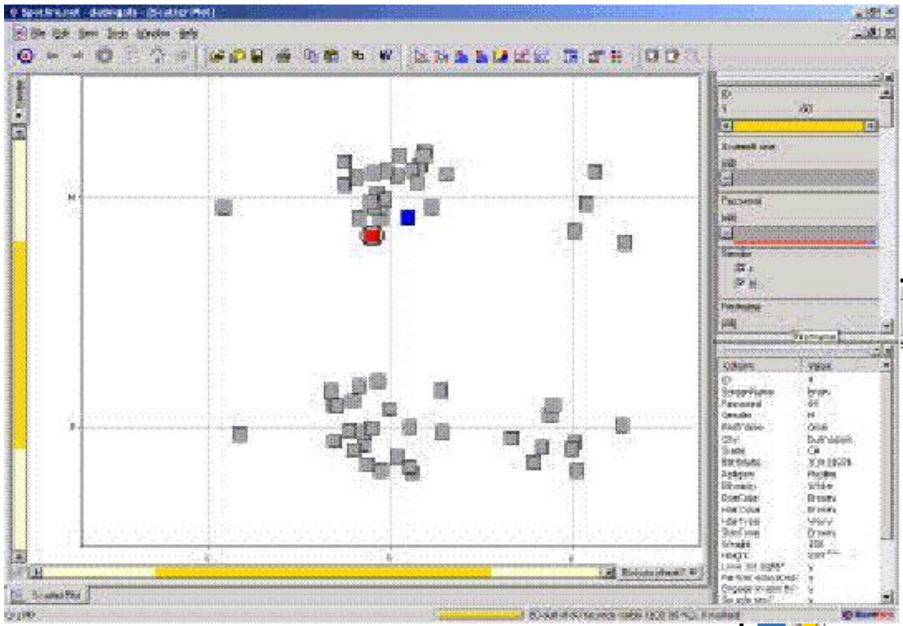
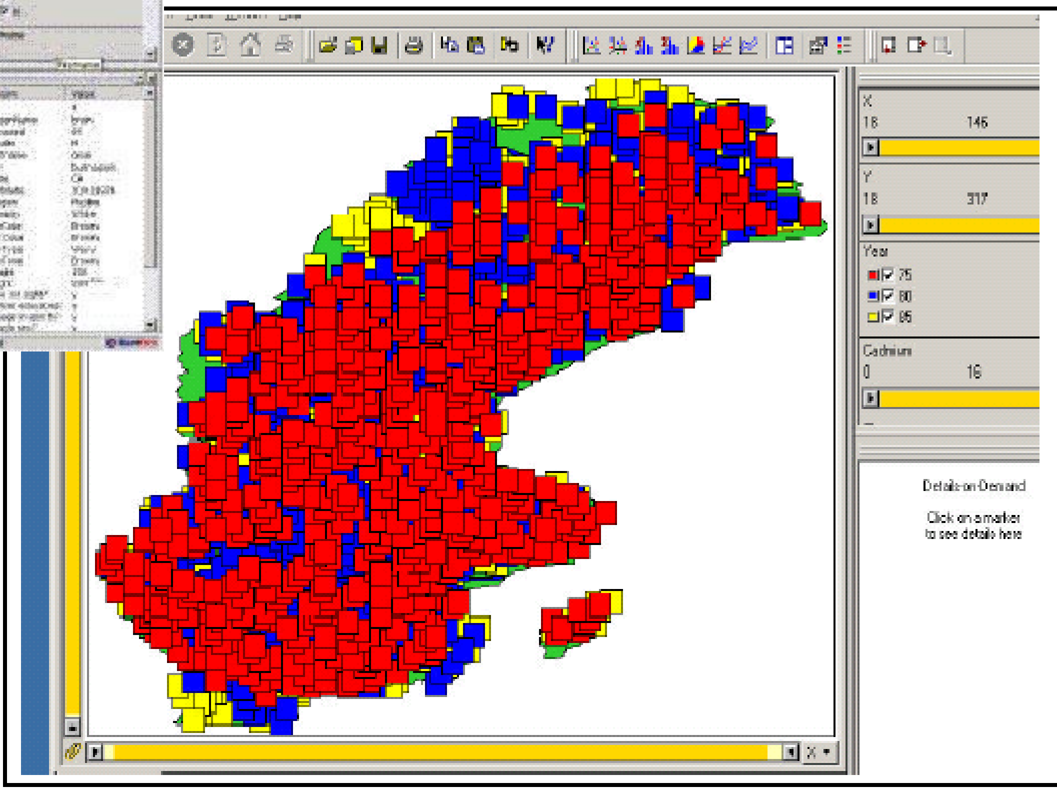


Figure 1b. This screenshot shows one possible way for solving the same problem in InfoZoom, specifically in its overview mode. After clicking at, and thereby zooming into, the "Yes" entries in the attribute "Did you cheat?", users can see from the length of the bars in the Gender category that females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

Comparing 3 Commercial Systems

SpotFire

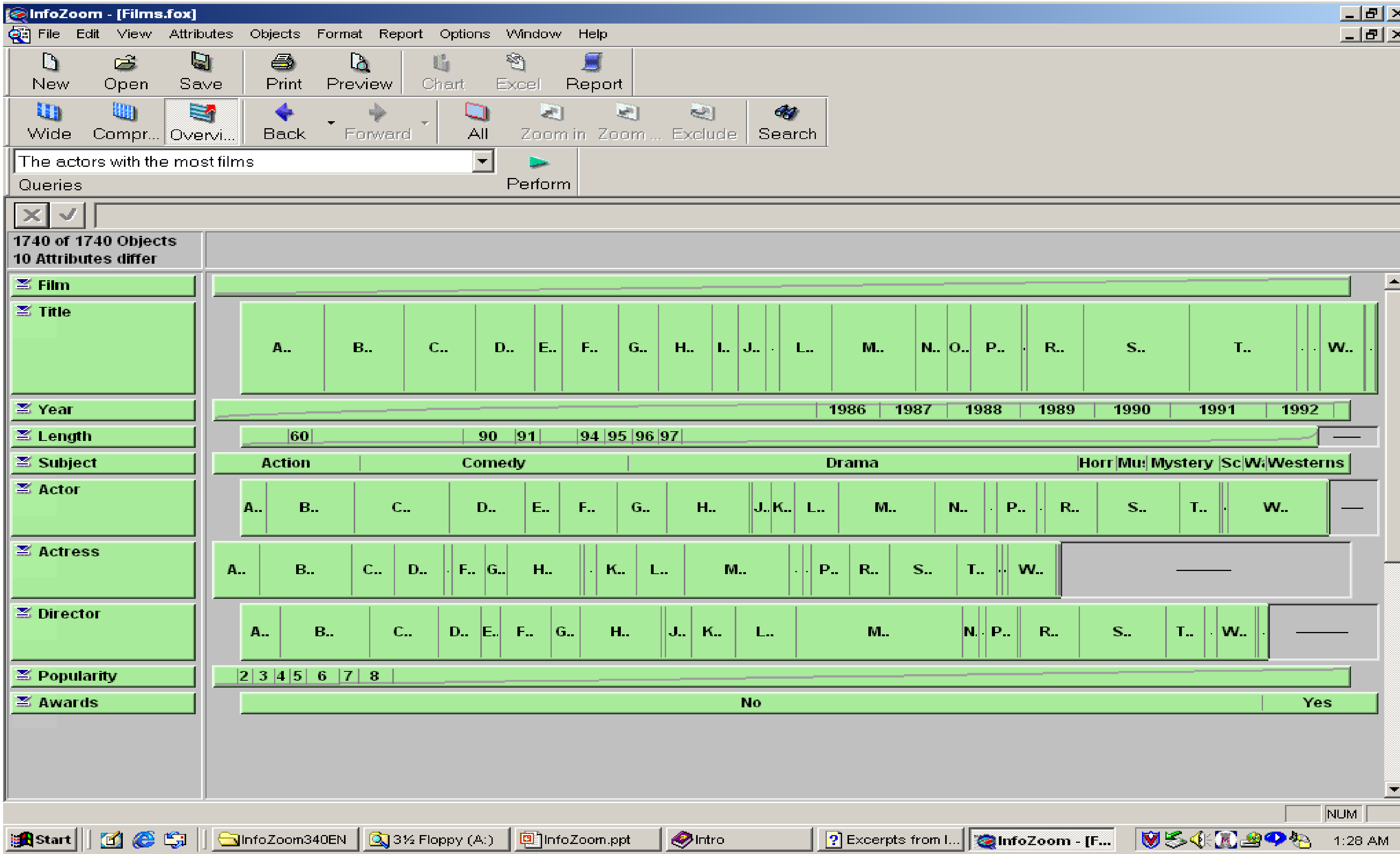




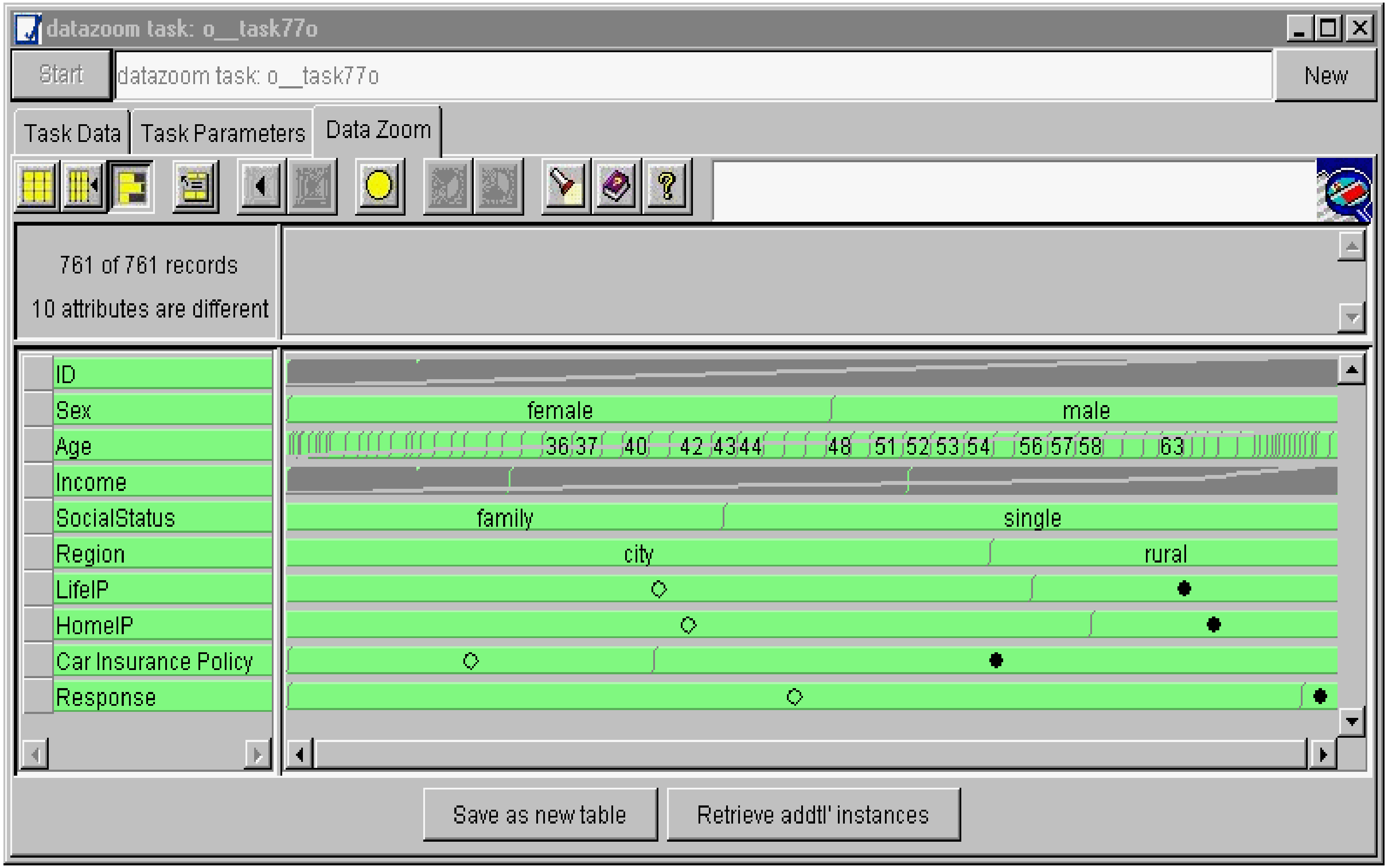
Infozoom Overview

- 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.

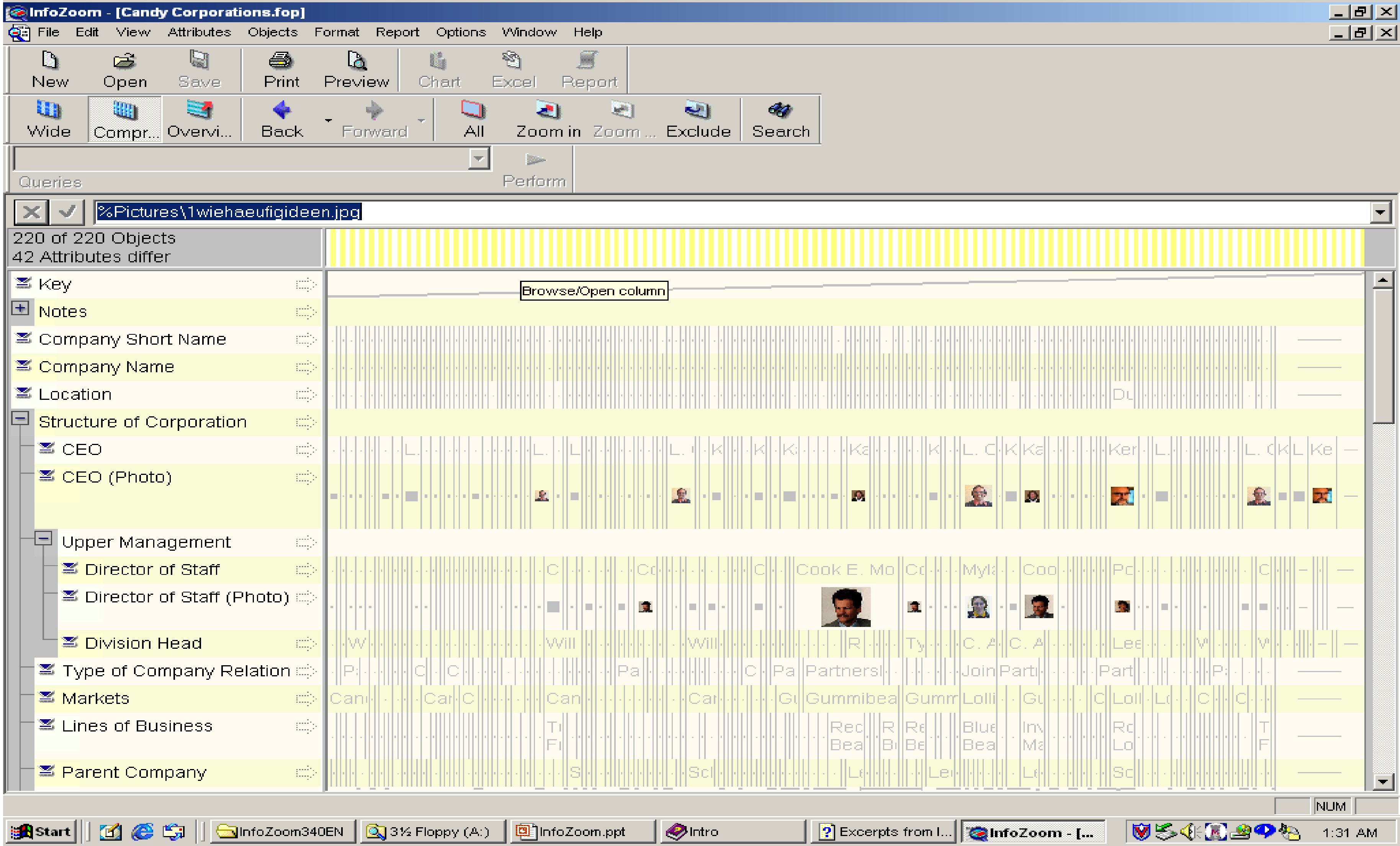
InfoZoom Overview View



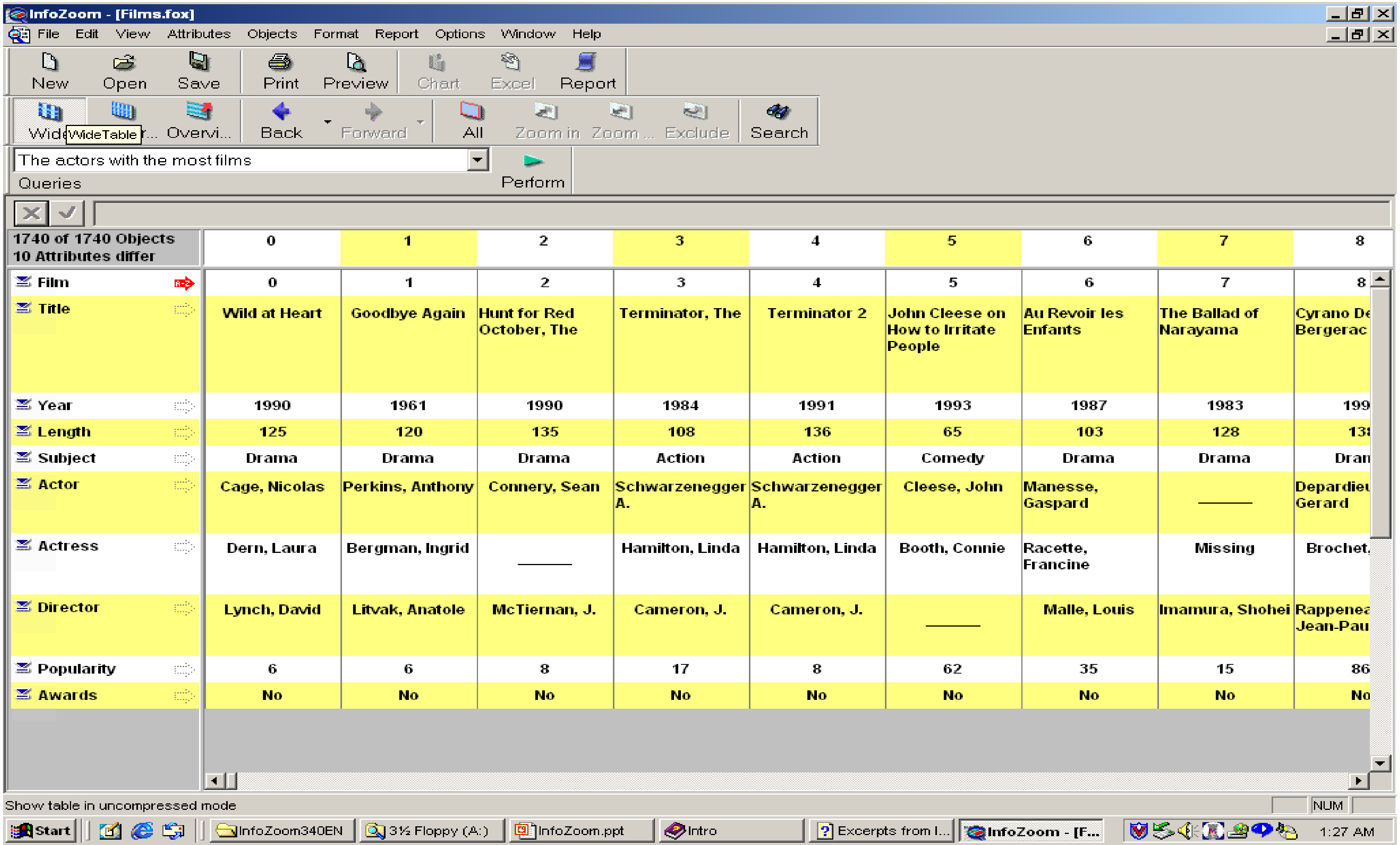
InfoZoom Overview View



InfoZoom Compressed Table View



InfoZoom Wide Table View



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 site?

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 vanadium and that of zinc?
  - Is there a low-level chrome area that is high in vanadium
- 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

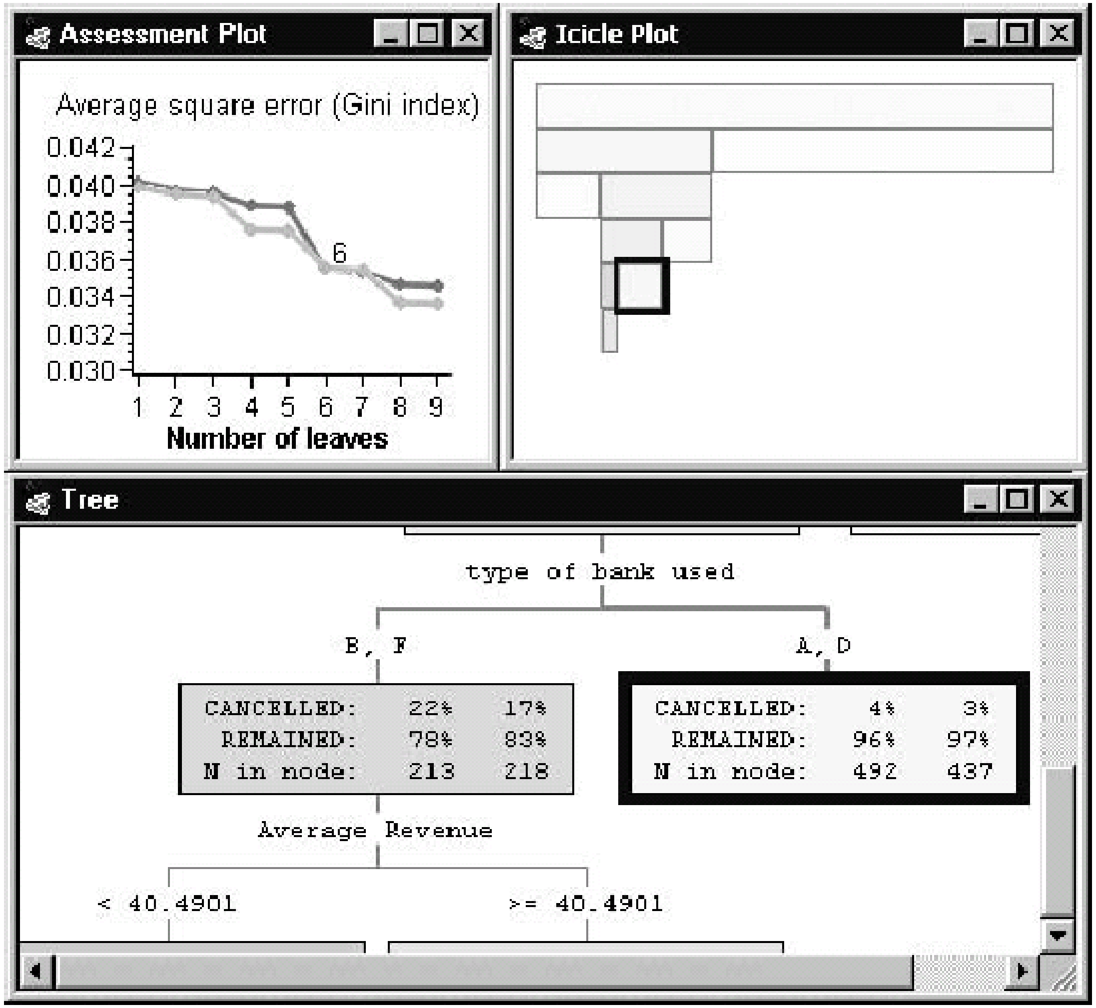
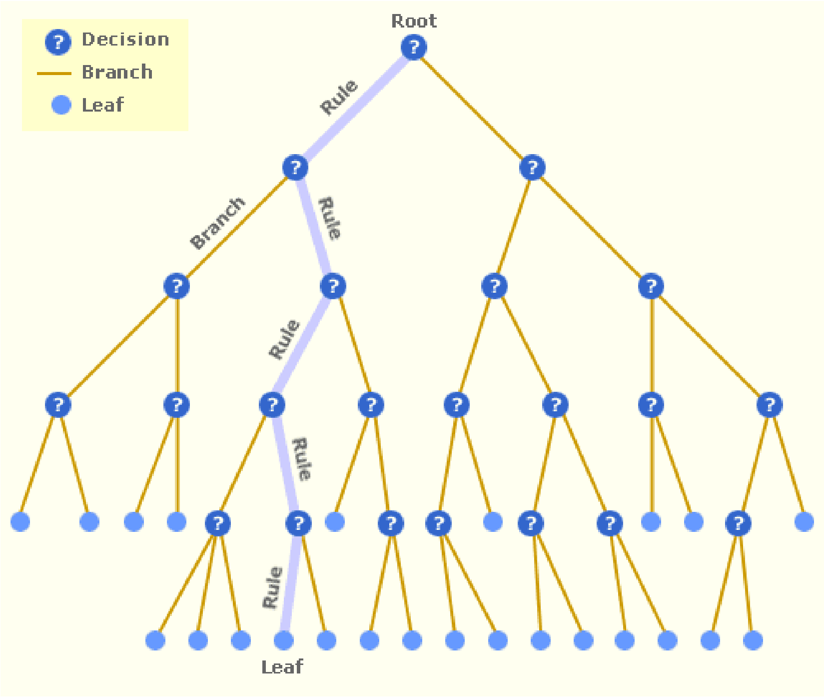


Figure 1. Example of compact view in data mining

Comparison of 2D Visualizations of Hierarchies, INFOVIS'01.

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

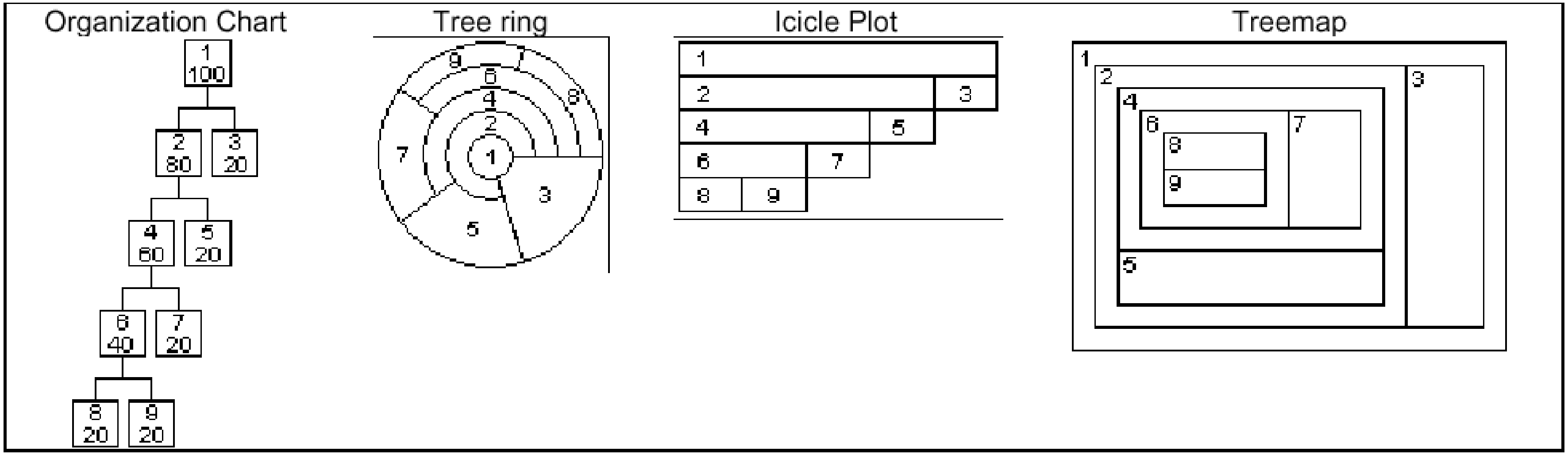


Figure 2. Different views of the same tree

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 larges 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

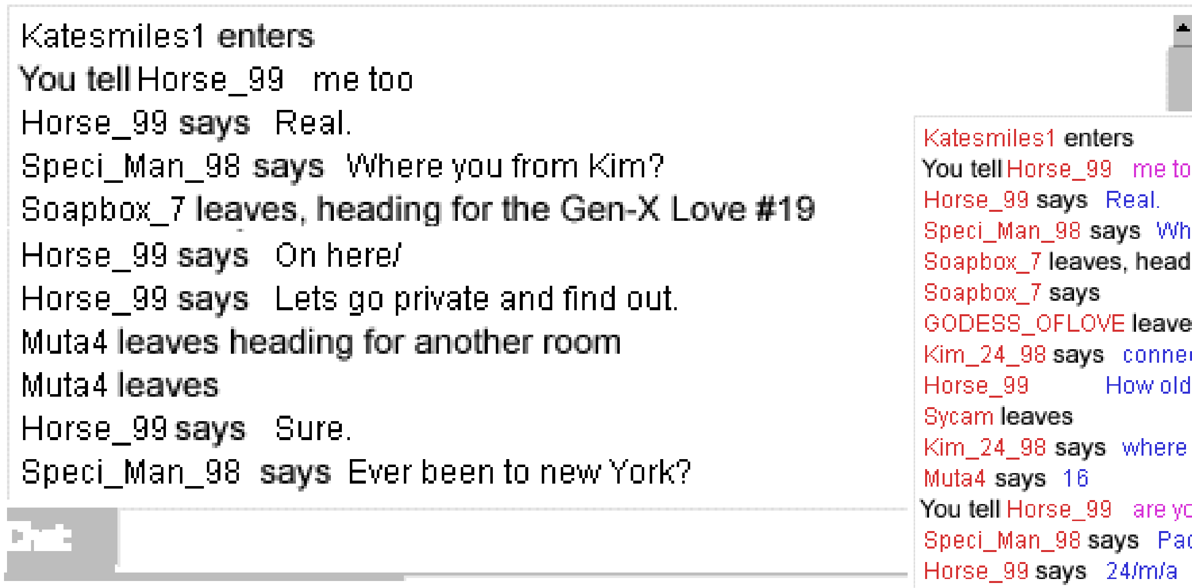
- Best design considering Org Chart could not be used for node size tasks

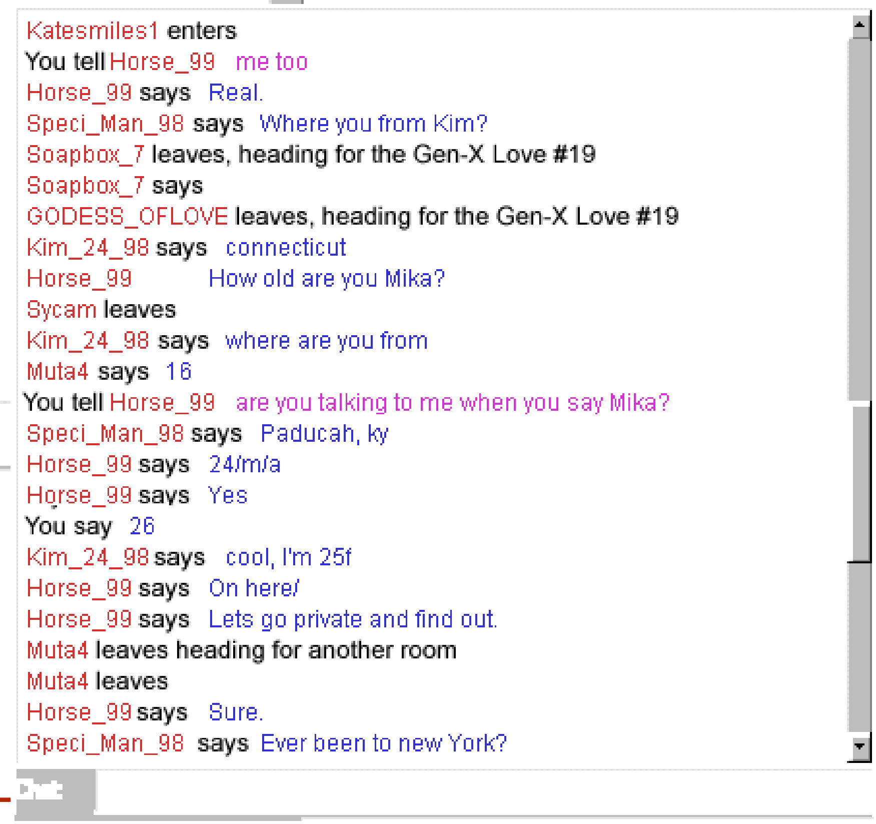
However:

- Didn't seem to actually do tests on trees as large as the ones they describe as typical of datamining

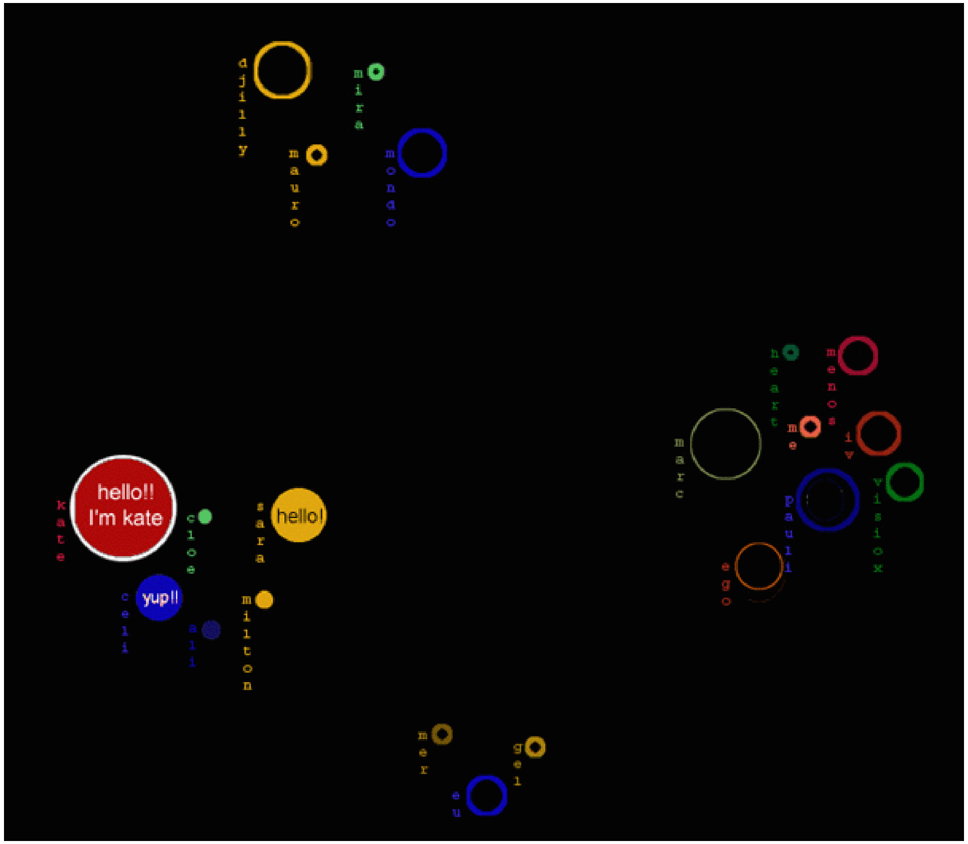
Visualizing Conversations

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

"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)

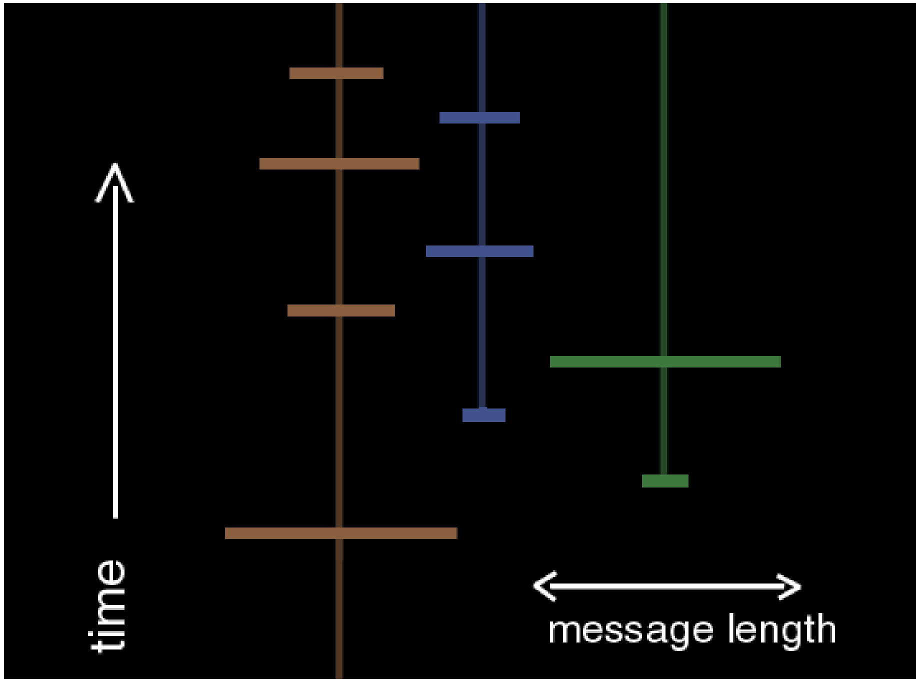
Circle appears bright when posting occurs

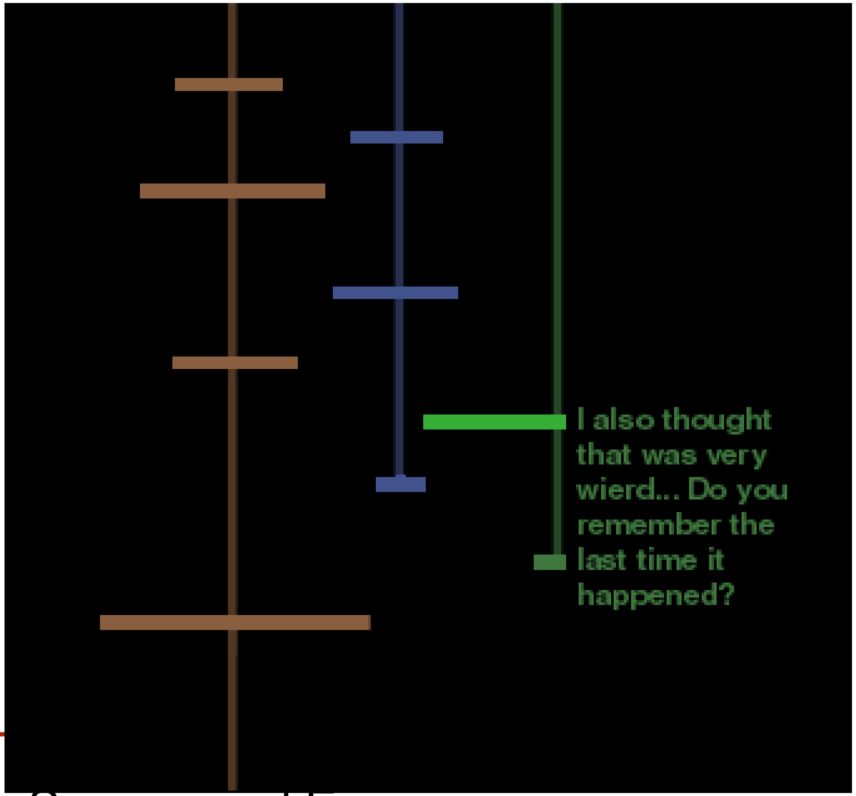
Circles of inactive users fade in the background

Chat Circles – Conversational Groupings

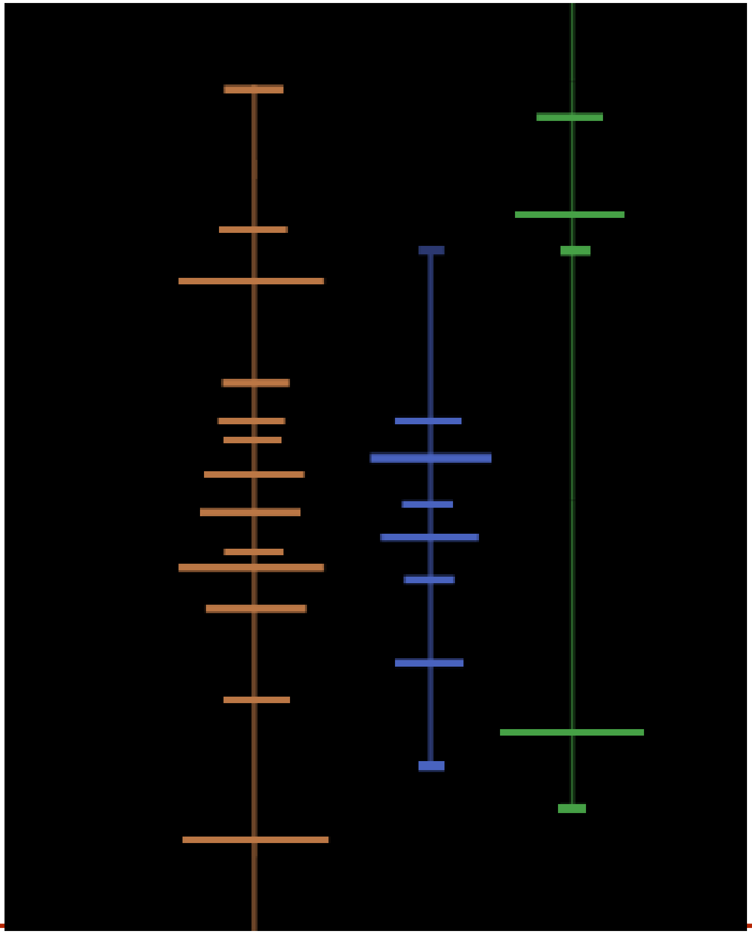
- There is only ONE room in Chat Circles
- Groupings are achieved by moving closer to other participants
- At any time, a participant can view all other participants
- A participant can also detect interesting conversations in different areas of the room by looking at how many circles are gathered and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

Chat Circles History





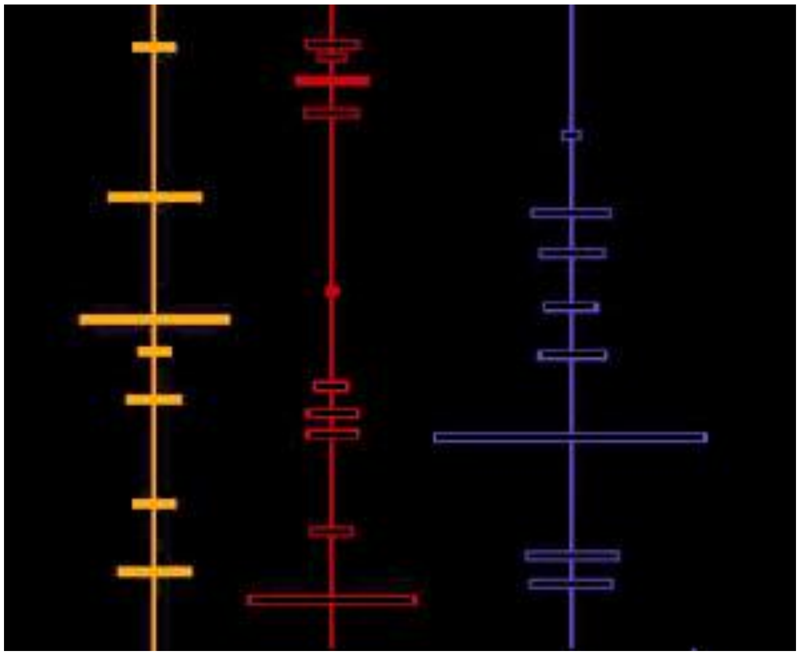
History Log Patterns



+ 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.

BreakingStory

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 seach for words or phrases that have appeared in online news. The results wlll 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

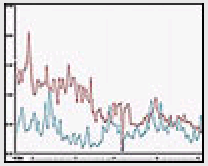
Step 2: Select Date Range

[Need help?]

05 Mar 03 to 04 Apr 03

Step 3: Choose One of Two Views

[Need help?]



Single Chart View

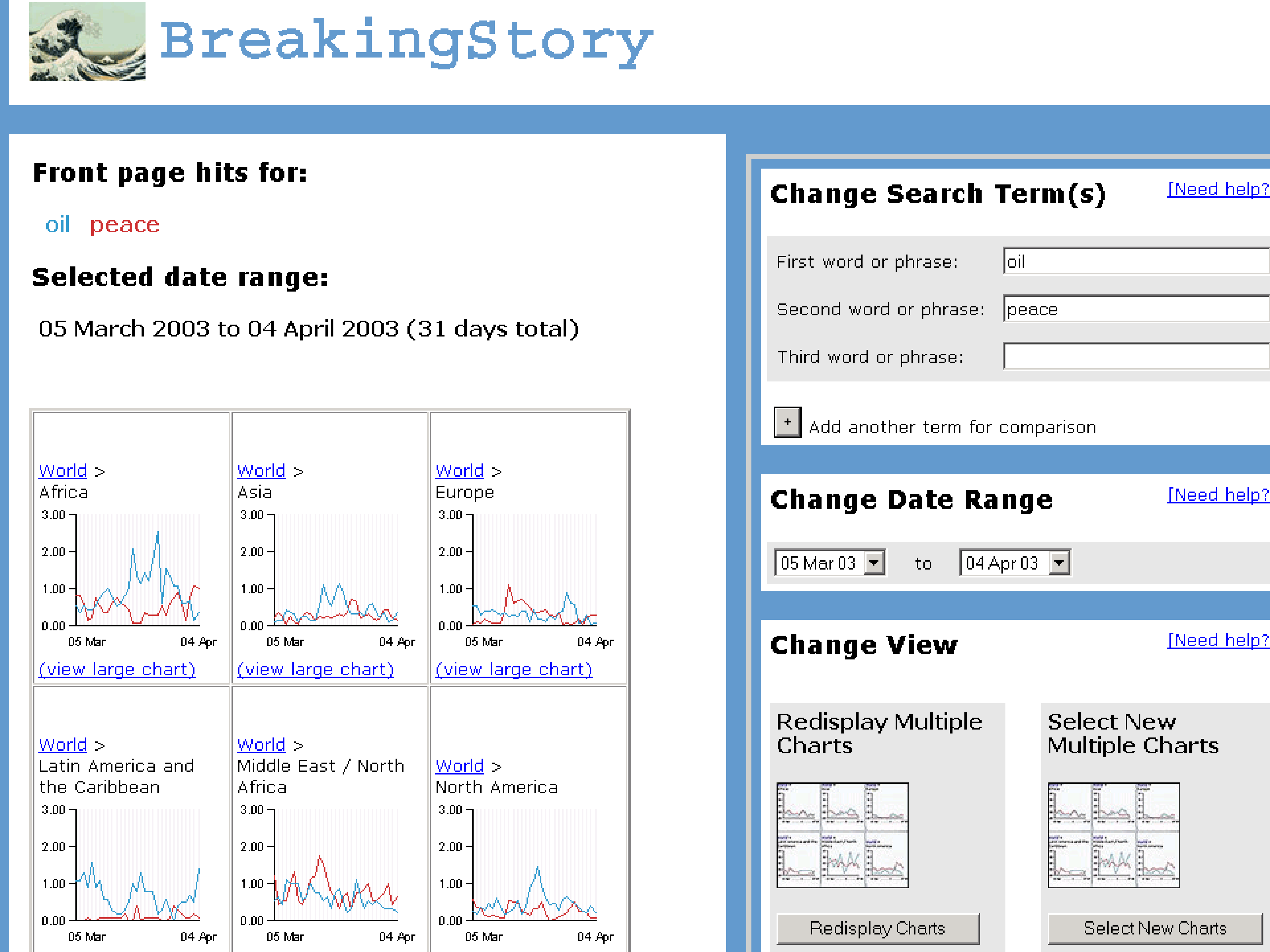
Choose a news site or collection of news sites, by selecting **one** of the following:

Individual News Site: –

All News Sites in Country: –

All News Sites in Region: All news sites

View a Single Chart


BreakingStory
Front page hits for:
oil peace
Selected date range:
05 March 2003 to 04 April 2003 (31 days total)
World > Africa
World > Asia
World > Europe
3.00
2.00
1.00
0.00
05 Mar
04 Apr
(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
05 Mar 03 to 04 Apr 03
Change View
Redisplay Multiple Charts
Select New Multiple Charts
Redisplay Charts
Select New Charts

BreakingStory

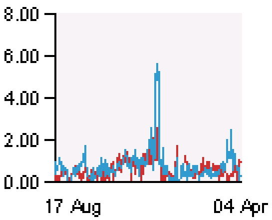
Front page hits for:

oil peace

Selected date range:

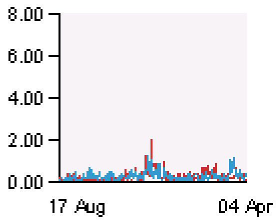
17 August 2002 to 04 April 2003 (231 days total)

World >
Africa



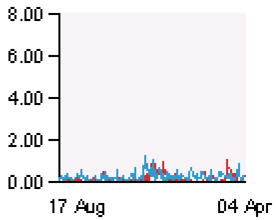
(view large chart)

World >
Asia



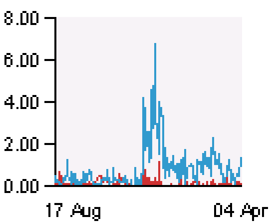
(view large chart)

World >
Europe



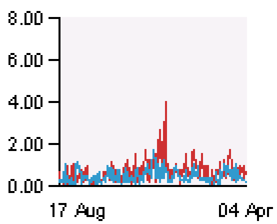
(view large chart)

World >
Latin America and the Caribbean



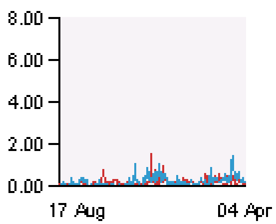
(view large chart)

World >
Middle East / North Africa



(view large chart)

World >
North America



(view large chart)

Change Search Term(s)

[Need help?]

First word or phrase: oil

Second word or phrase: peace

Third word or phrase:

Fourth word or phrase:

+ Add another term for comparison

Change Date Range

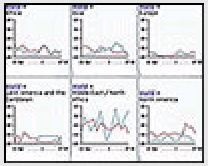
[Need help?]

17 Aug 02 to 04 Apr 03

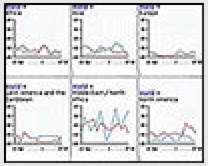
Change View

[Need help?]

Redisplay Multiple Charts



Select New Multiple Charts



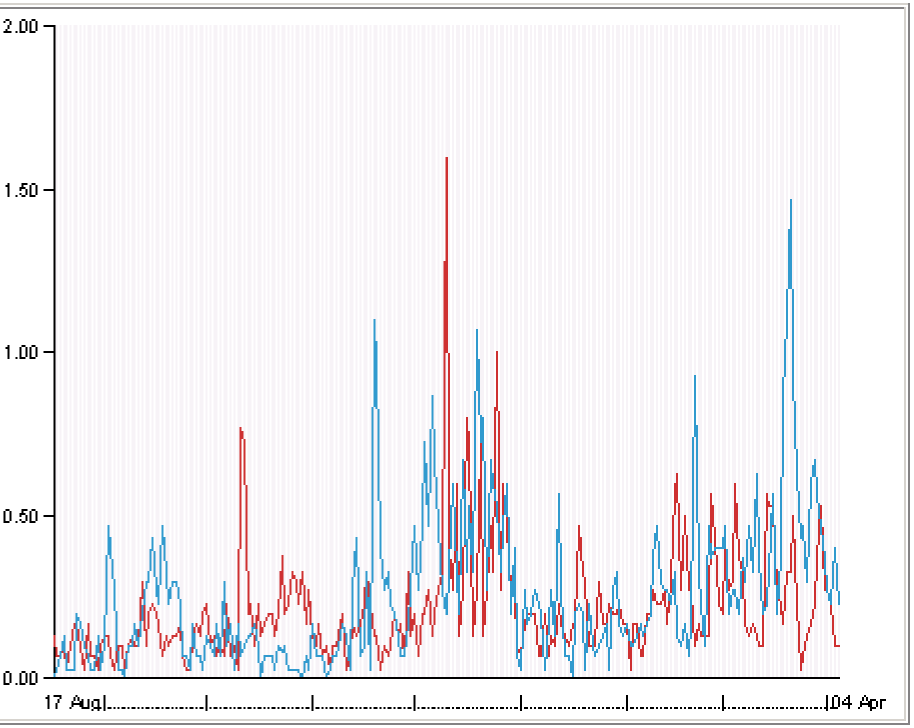
Front page hits:

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: oil
Second word or phrase: peace
Third word or phrase:
Fourth word or phrase:

\+ Add another term for comparison

Change Date Range

[Need help?]

17 Aug 02 to 04 Apr 03

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: North America

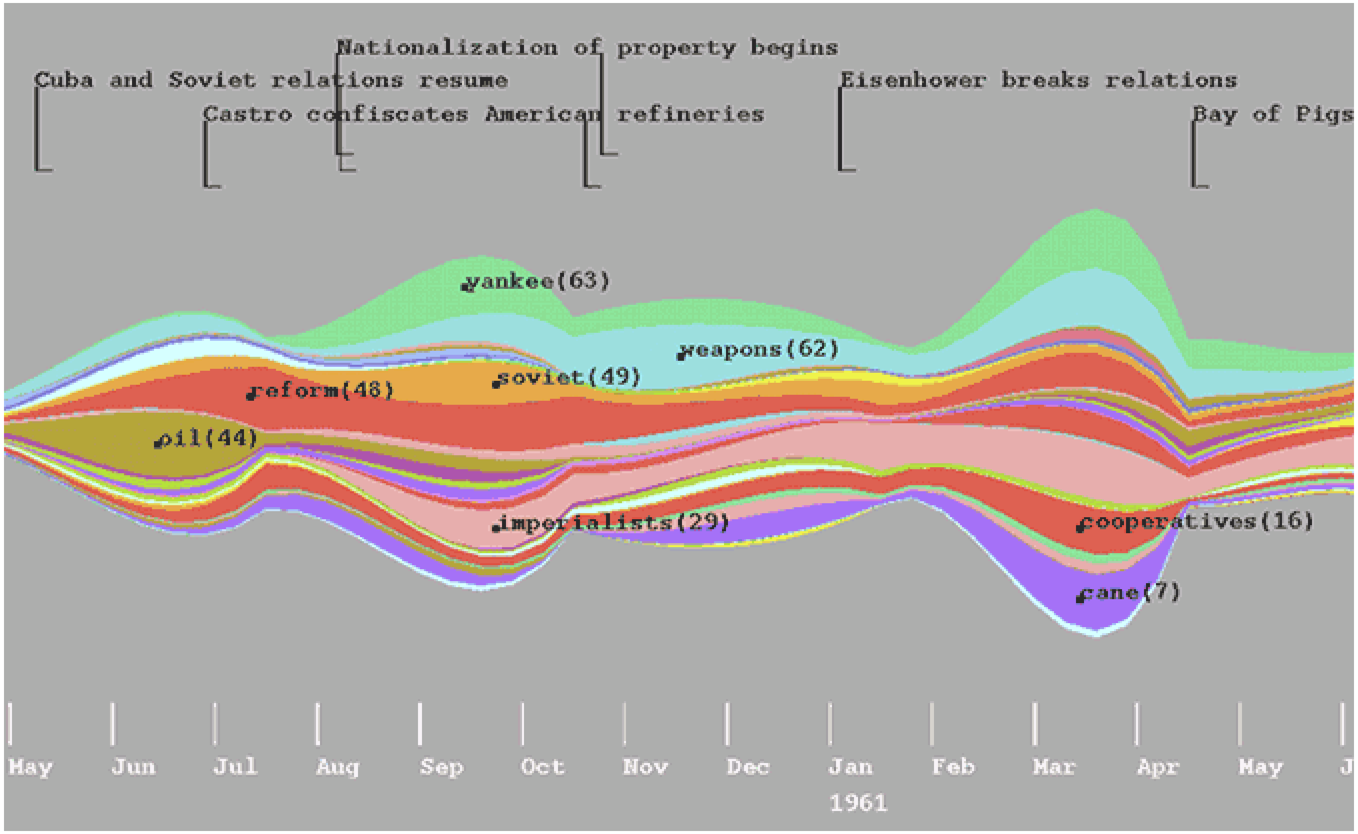
Change View

[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

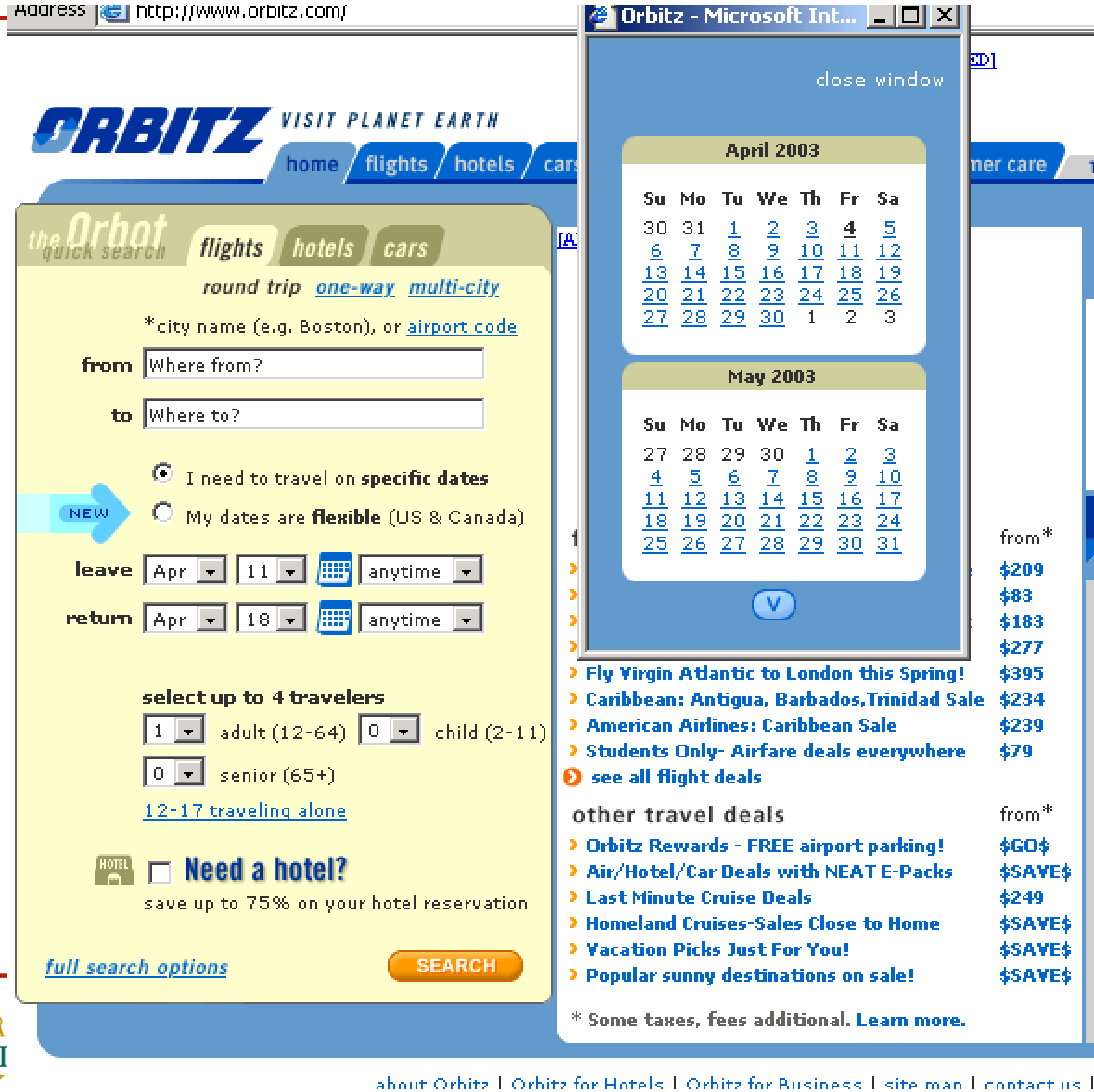
Tailor to the application and the domain

Create highly interactive and integrated systems

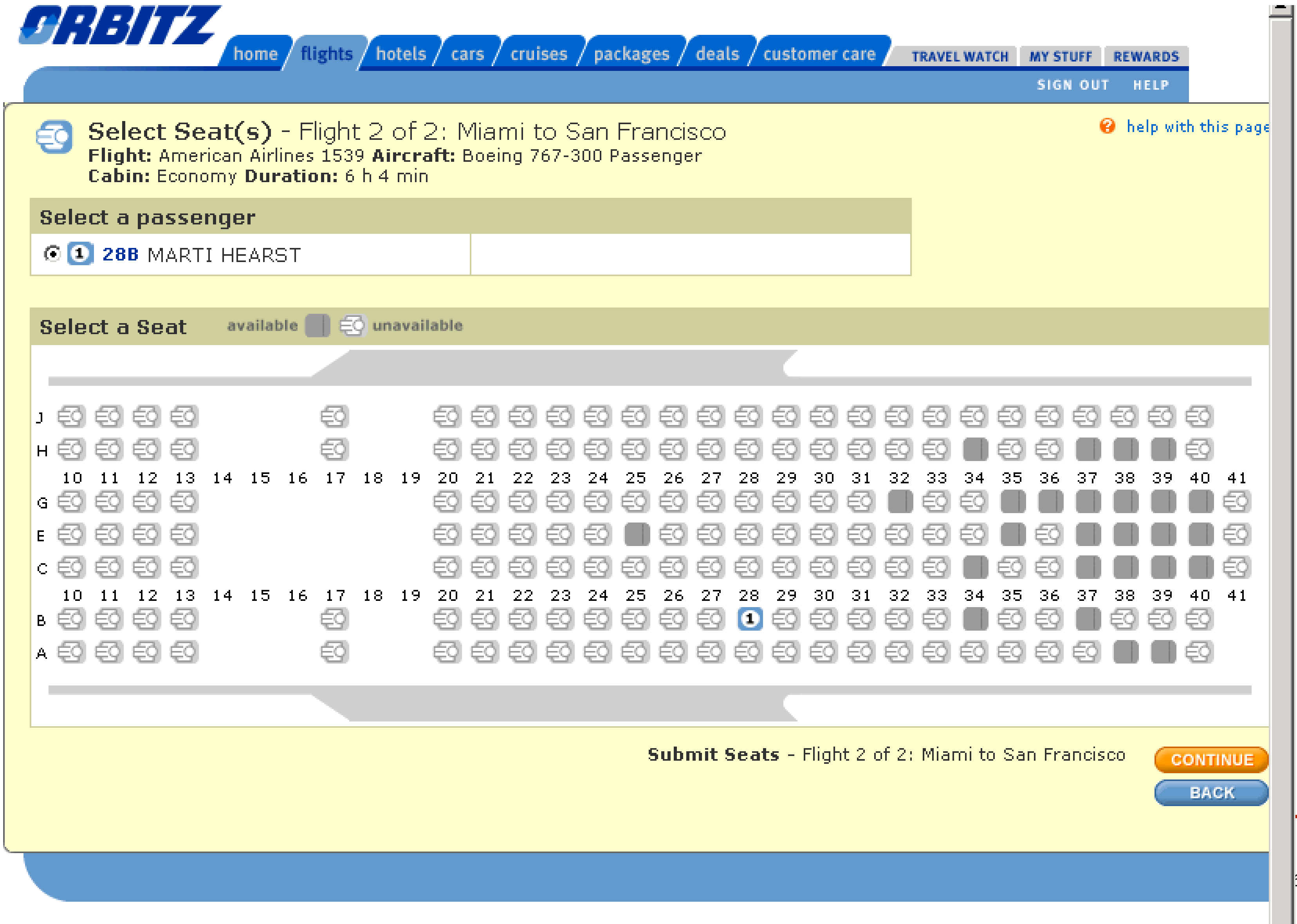
Embed the visualization within a larger application

Provide alternative views

Visualization with a Light Touch: Orbitz.com

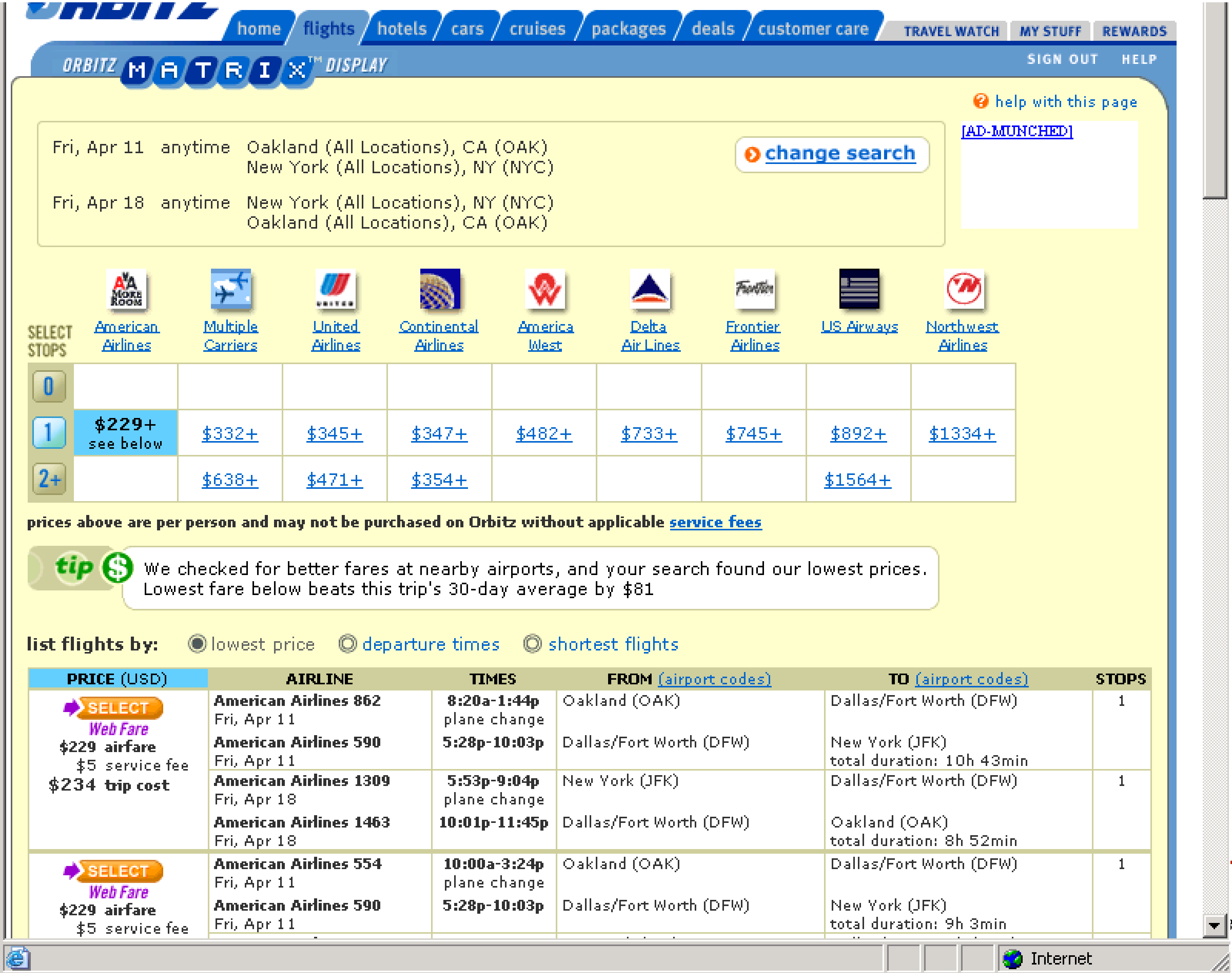


Visualization with a Light Touch: Orbitz.com

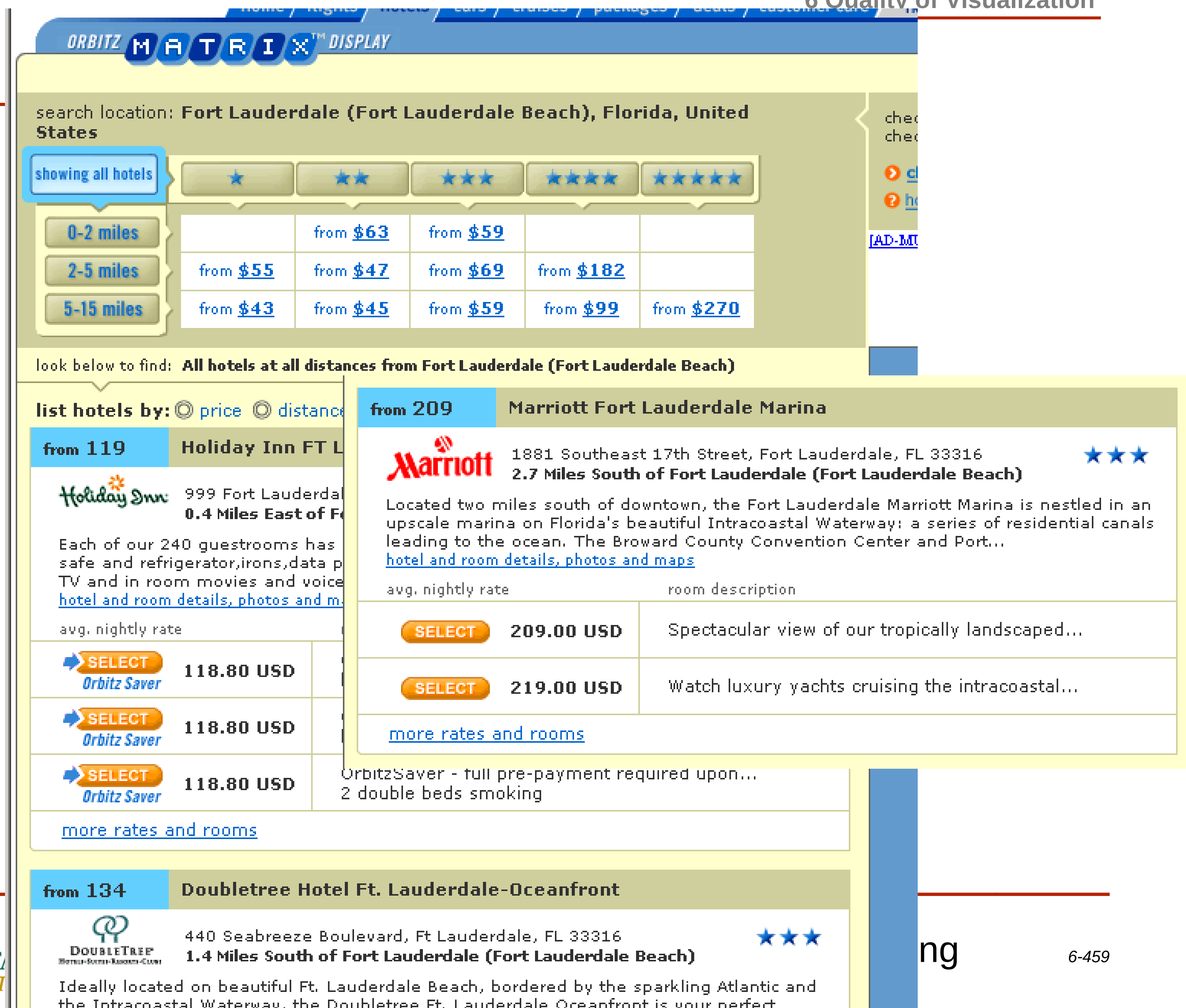


Visualization with a Light Touch: Orbitz.com

Visualization with a Light Touch: Orbitz.com



Visualization with a Light Touch: Orbitz.com



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