Information Visualization

Problem:

- HUGE Datasets: How to understand them?
- Data do not have a "natural dimension"

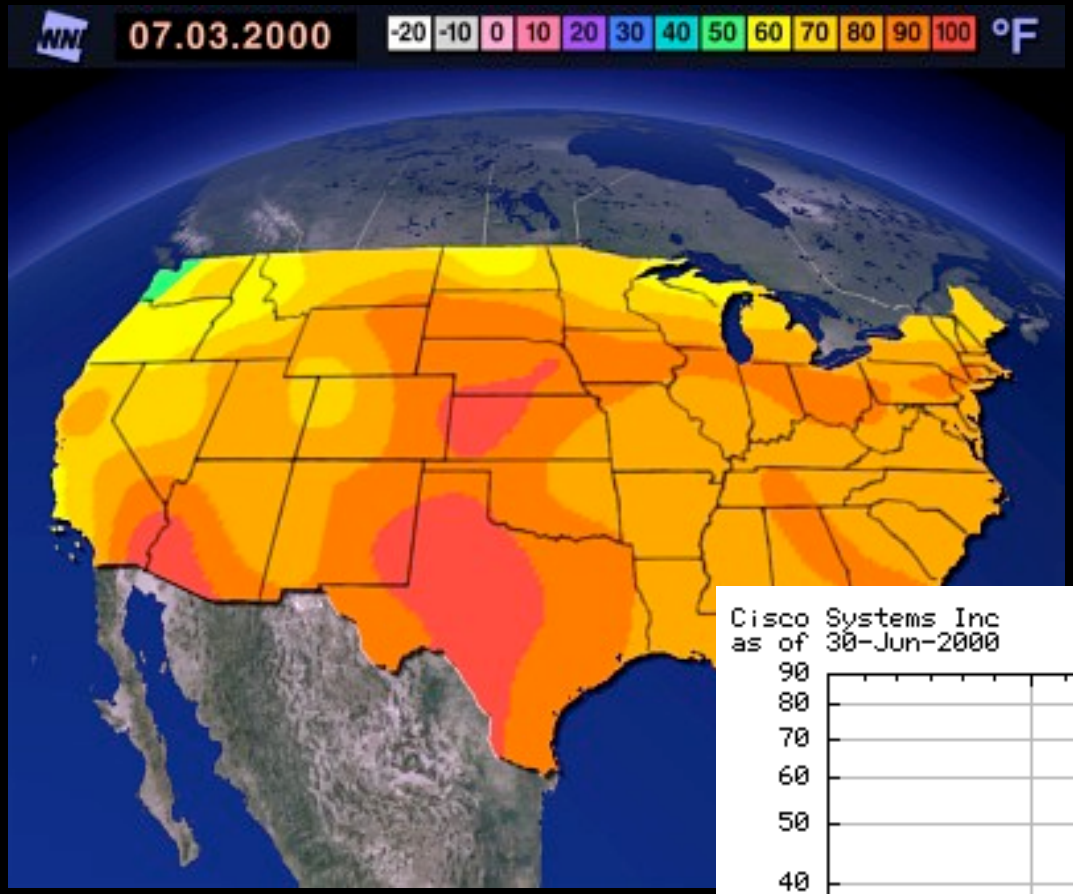
Solution

- 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

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 MASSACHU

6. Turn RIGHT onto RUSSELL ST.

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.

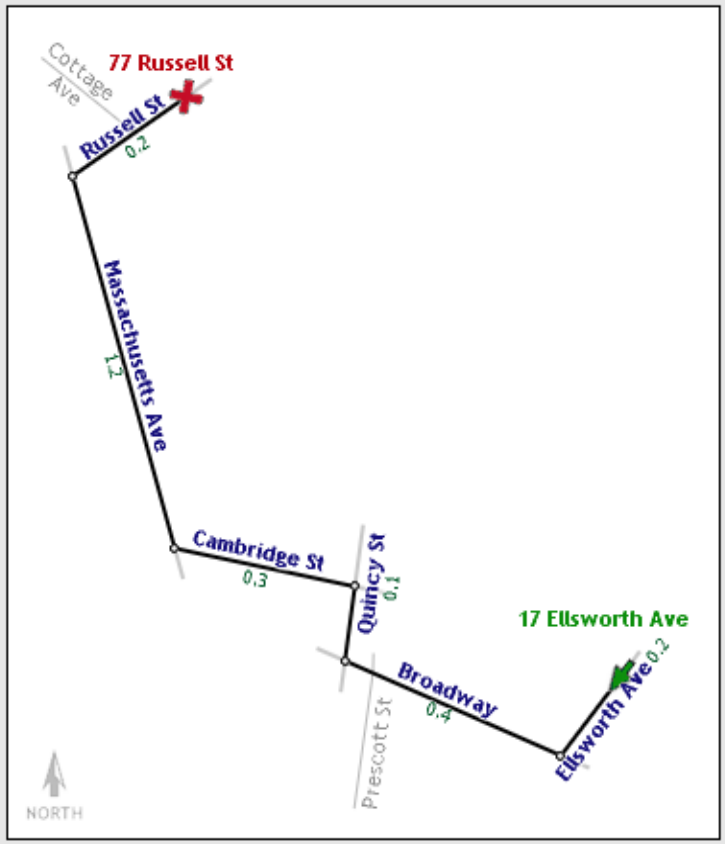
4. Turn LEFT onto CAMBRIDGE ST.

5. Turn SLIGHT RIGHT onto MASSACHU

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.

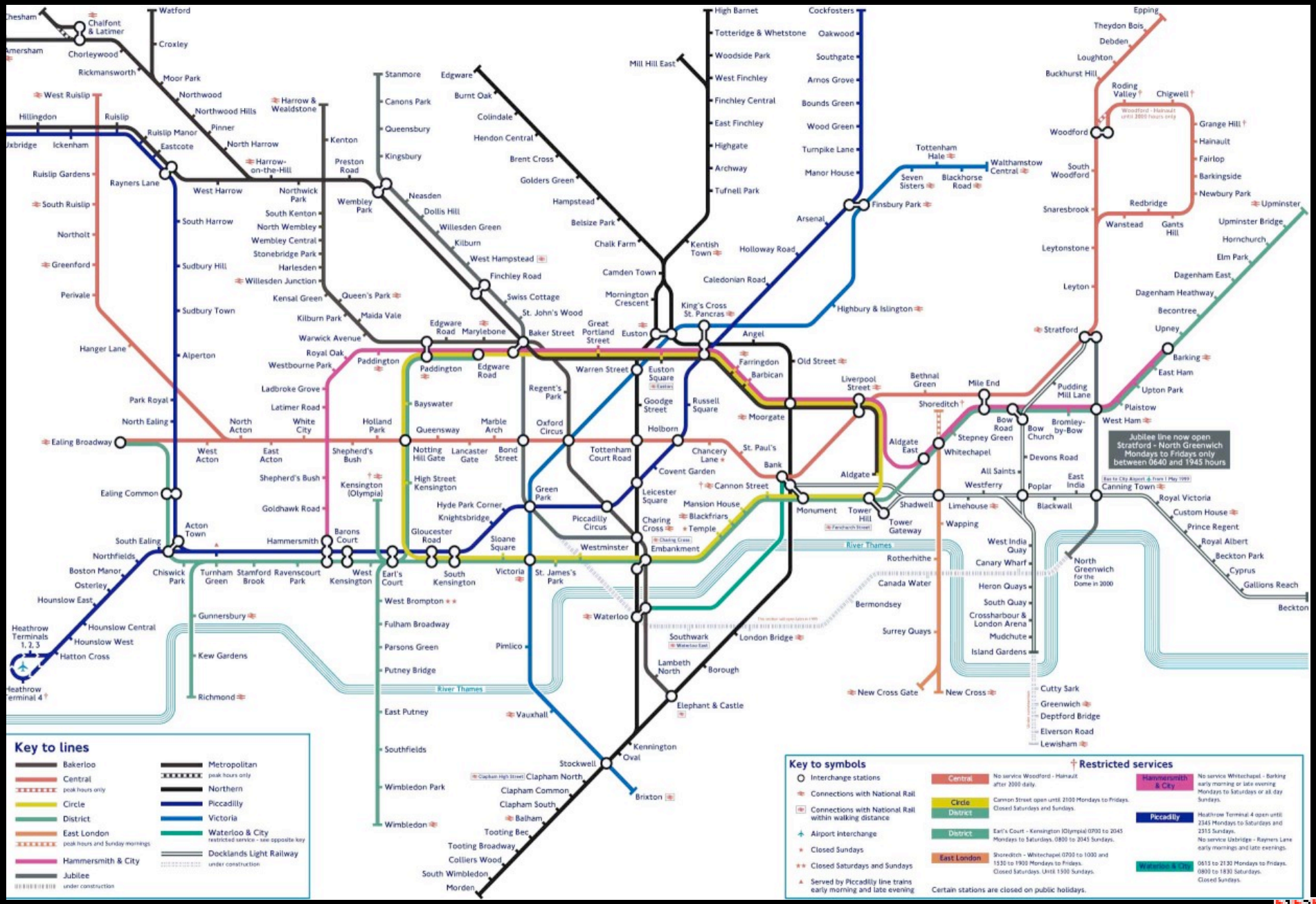


	Directions	Elapsed Distance
1	Begin at **17 Ellsworth Ave** on **Ellsworth Ave** and go Southwest for 500 feet	0.1
2	Turn right on **Broadway** and go Northwest for 0.4 miles	0.5
3	Turn right on **Quincy St** and go North for 200 feet	0.5
4	Turn left on **Cambridge St** and go West for 0.3 miles	0.8
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

Line drawing tool by Maneesh Agrawala http://graphics.stanford.edu/~maneesh/

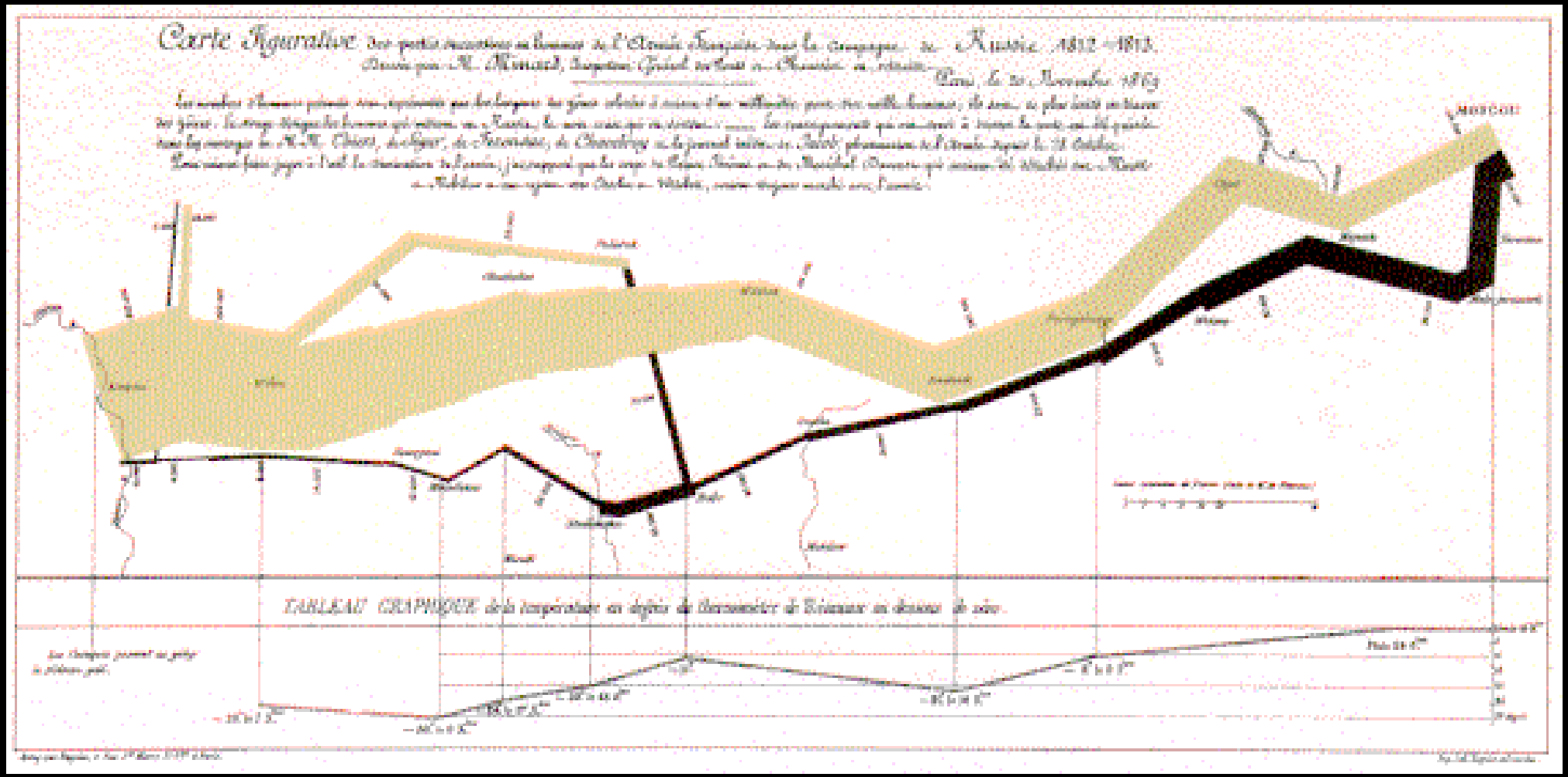
London Subway

www.londontransport.co.uk/tube



Napolean's March

From E. Tufte
The Visual Display of
Quantitative Information



Minard graphic

size of army
direction

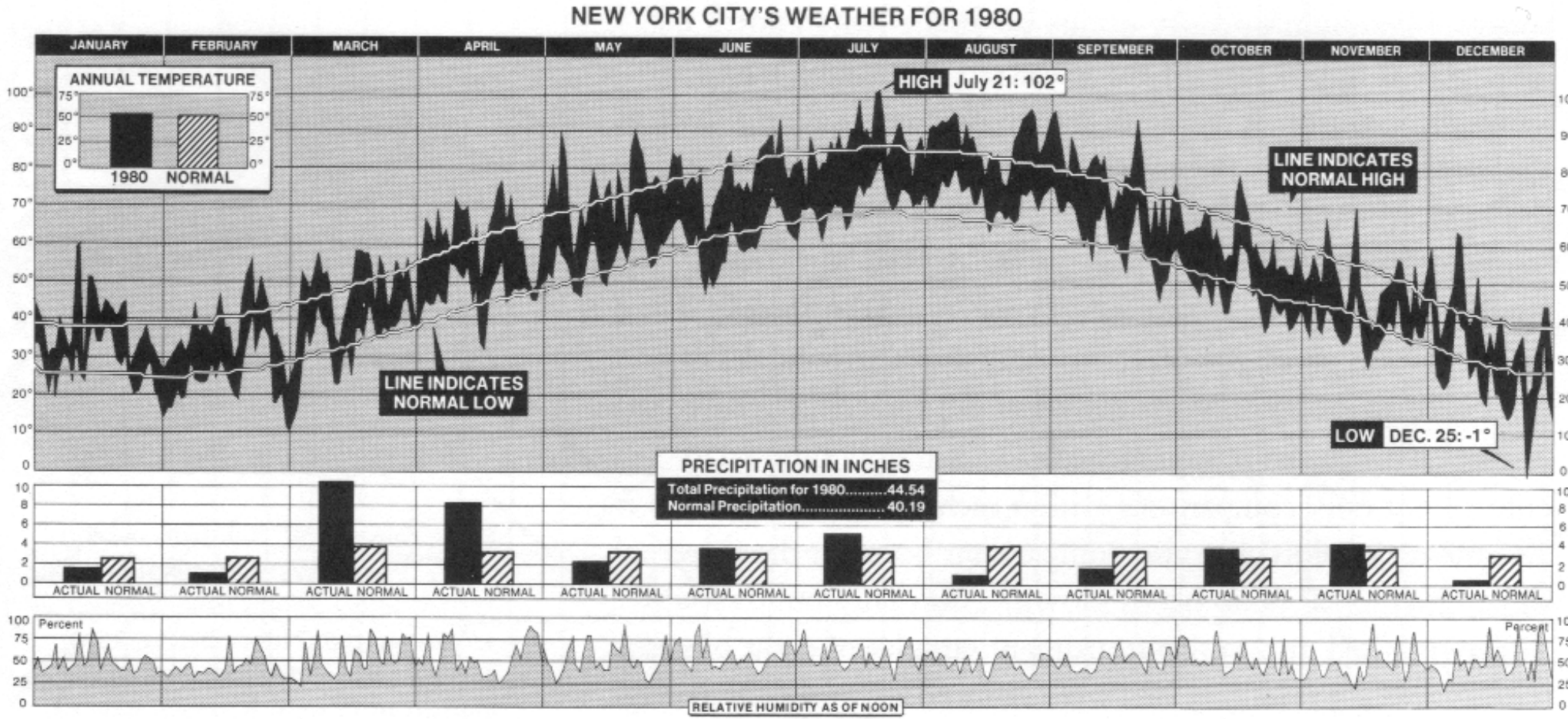
latitude
longitude

temperature
date

Example

NYC weather

2220 numbers



New York Times, January 11, 1981, p. 32.

Tufte, Vol. 1

Data Explosion

Society is more complex

- There simply is more "stuff"

Computers, internet and web give people access to an incredible amount of data

- news, sports, financial, purchases, etc...

How much data?

Between 1 and 2 exabytes of unique info produced per year

- 1000000000000000000 (10^{18}) bytes
- 250 meg for every man, woman and child
- Printed documents only .003% of total

Peter Lyman and Hal Varian, 2000
Cal-Berkeley, Info Mgmt & Systems
www.sims.berkeley.edu/how-much-info

The Challenge

Transform the data into information (understanding, insight) thus making it useful to people

One Approach

Provide tools that present data in a way to help people understand and gain insight from it

Information Visualization

What is "information"?

- Items, entities, things which do not have a direct geometric correspondence to space

- Baseball statistics, stock trends, connections between criminals, car attributes...

Information Visualization

What is "Information visualization"?

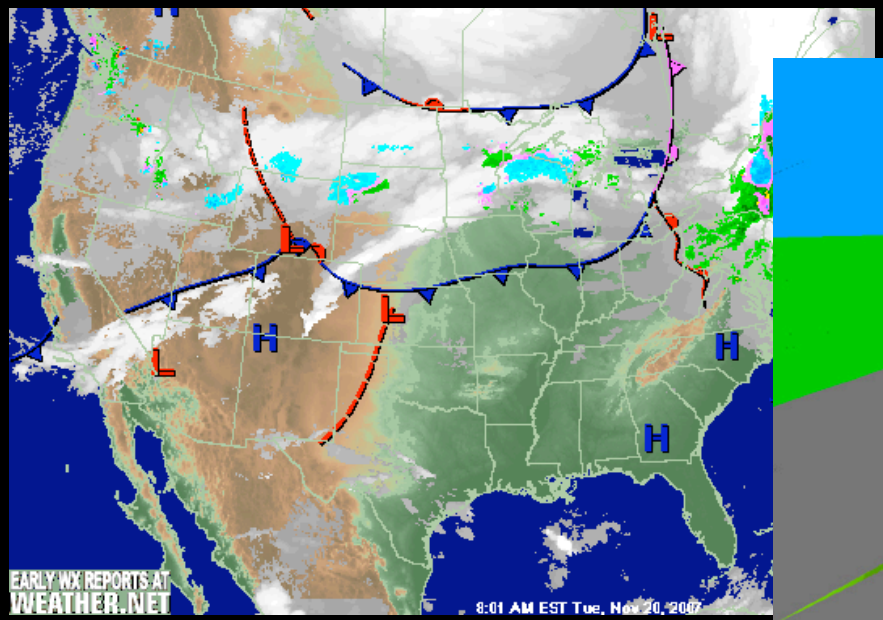
- The use of computer-supported, interactive visual representations of data to amplify cognition.

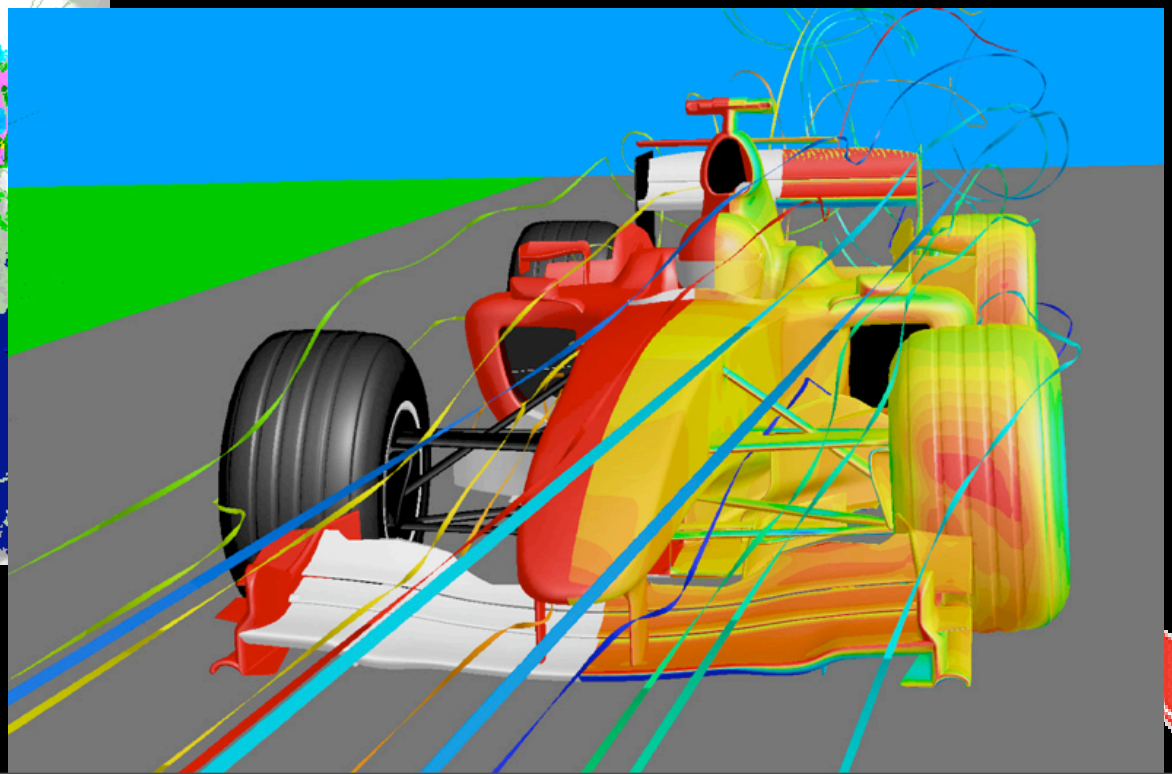
  —From [Card, Mackinlay Shneiderman '98]

What It's Not

Scientific Visualization

- Primarily relates to and represents something physical or geometric





Information Visualization

Components:

- Taking items without a direct physical correspondence and mapping them to a 2-D or 3-D physical space.

- Giving information a visual representation that is useful for analysis and decision-making

Two Key Attributes

Scale

- Challenge often arises when data sets become very large

Interactivity

- Want to show multiple different perspectives on the data

Key Tools

Using size to indicate quantity

Using color for distinguishing / selection

Brushing and Linking

Animation

Providing multiple views

The Need for Critical Analysis

We see many creative ideas, but they often bomb in use

The hard part:

how to apply it judiciously

- Inventors usually do not accurately predict how their invention will be used

Tasks in Info Vis

Search

- Finding a specific piece of information
  - —How many games did the Braves win in 1995?
  - —What novels did Ian Fleming author?

Browsing

- Look over or inspect something in a more casual manner, seek interesting information
  - —Learn about crystallography

Tasks in Info Vis

Analysis

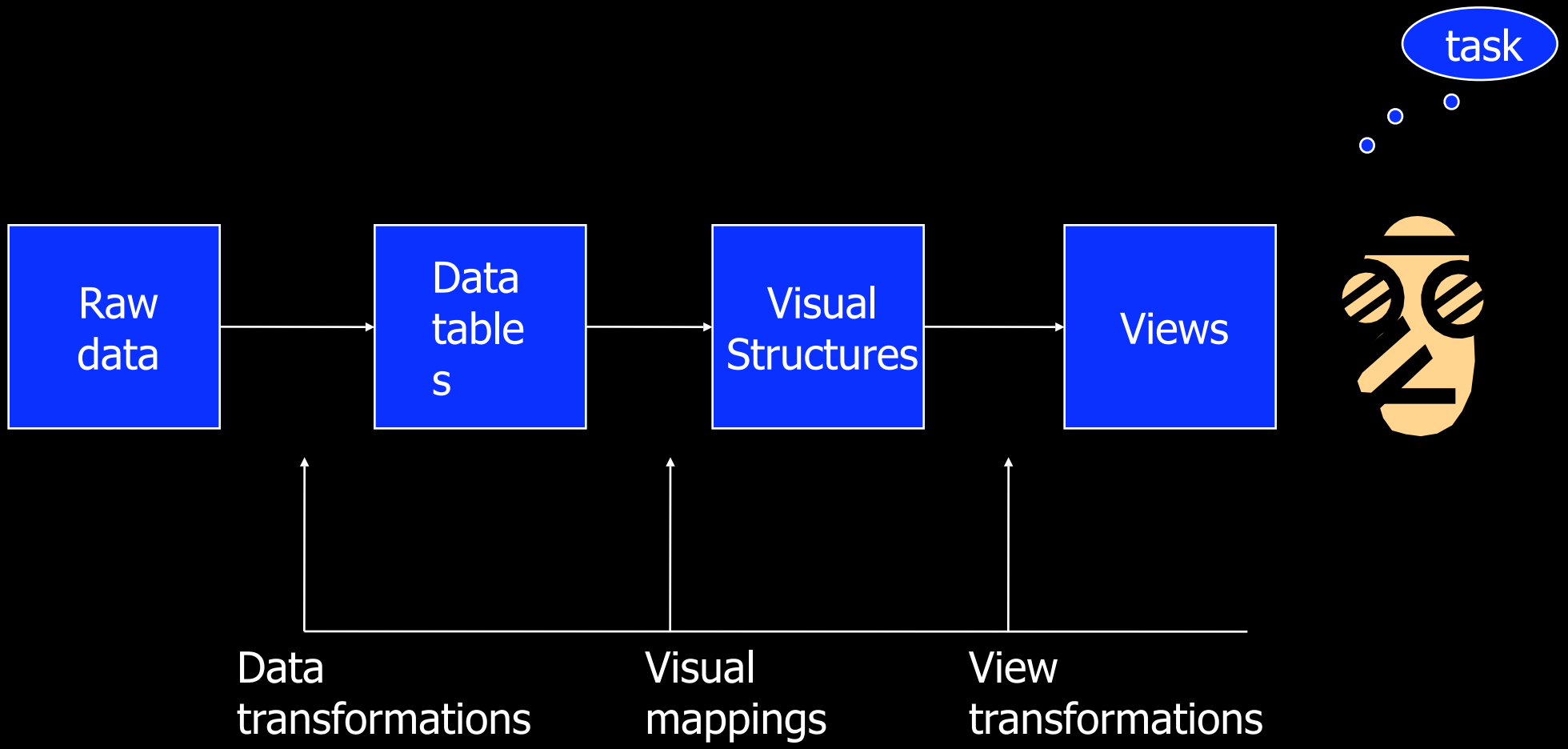
- Comparison-Difference
- Outliers, Extremes
- Patterns

Assimilation

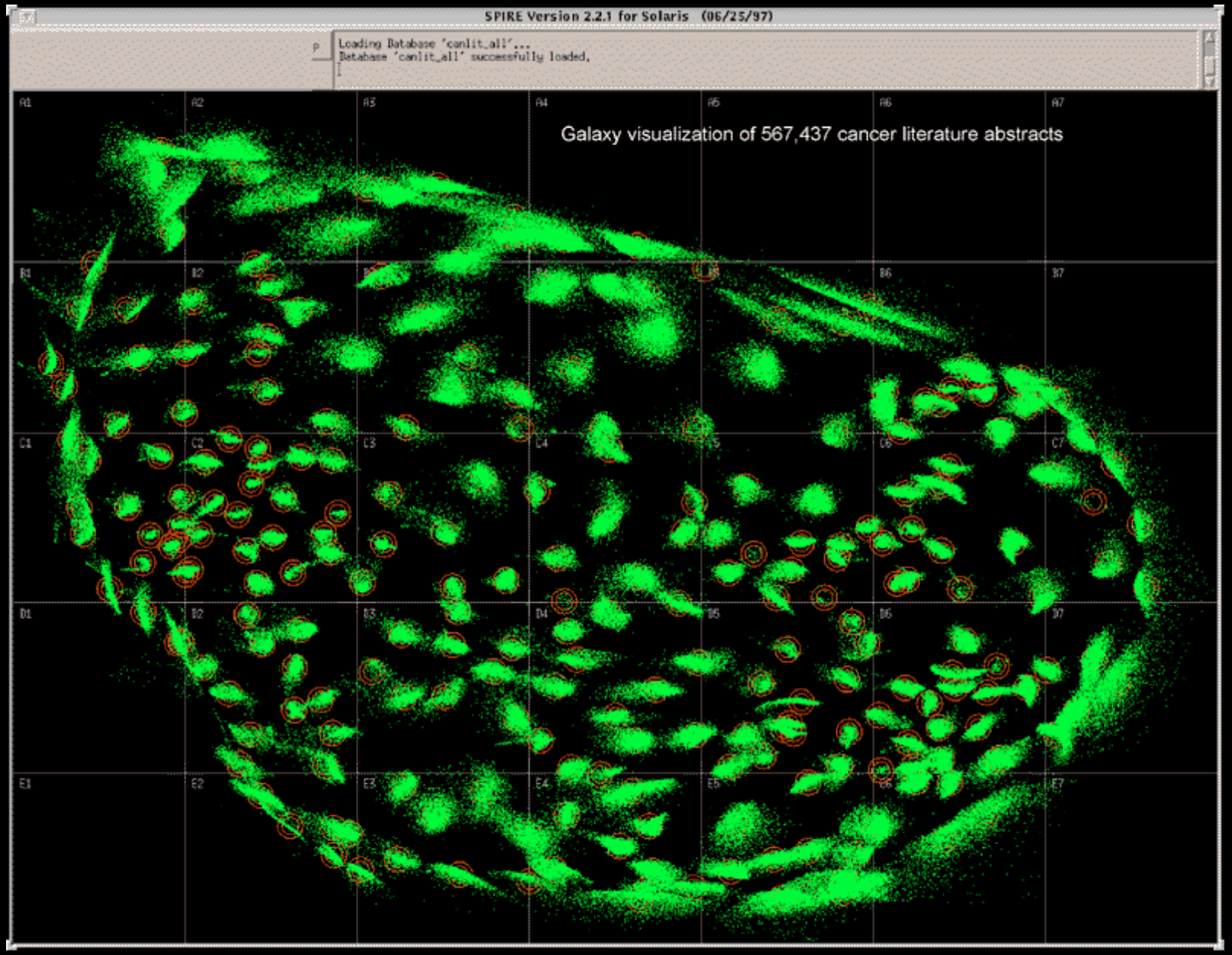
Monitoring

Awareness

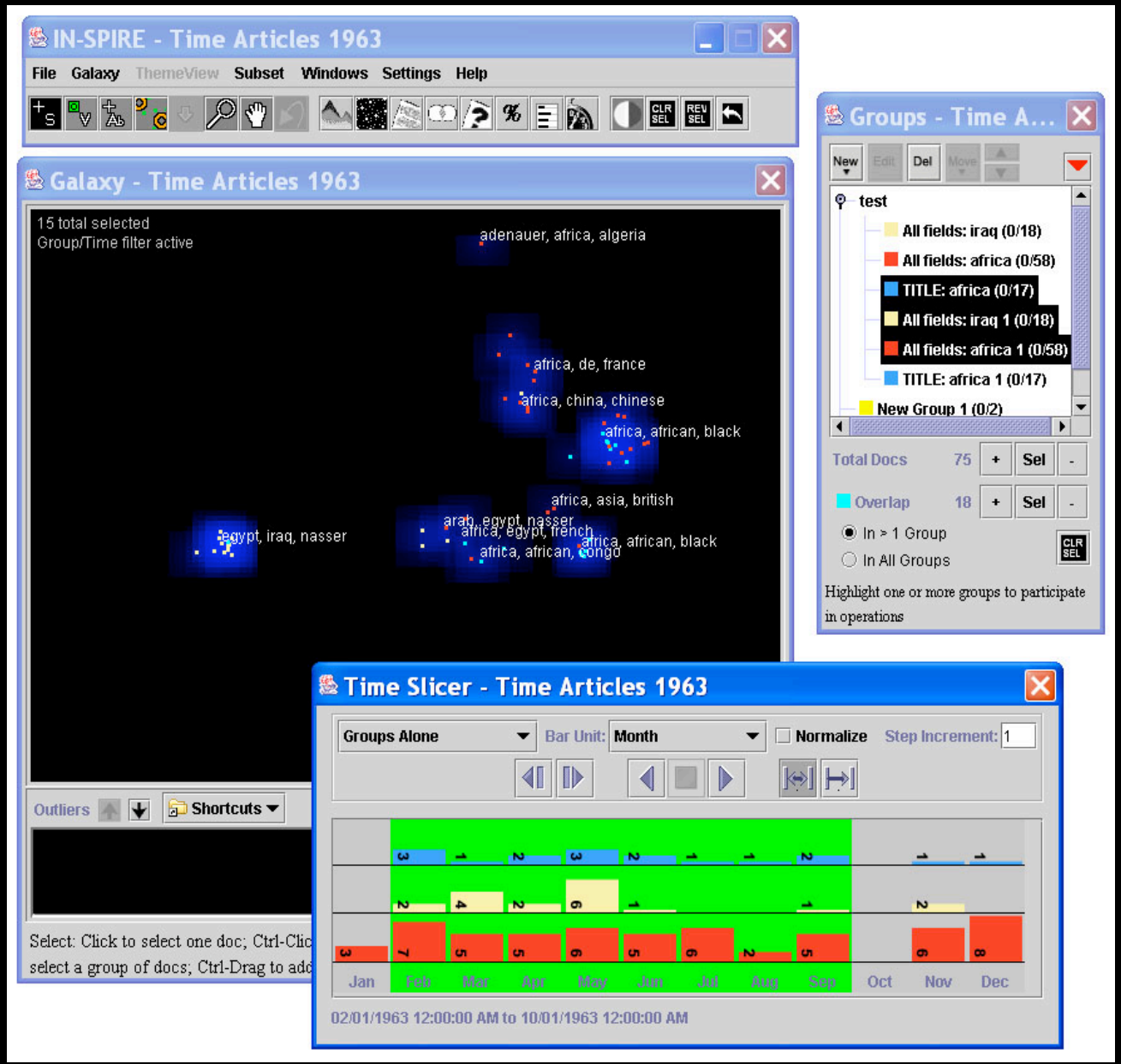
Process


task
Raw data
Data tables
Visual Structures
Views
Data transformations
Visual mappings
View transformations

Galaxy

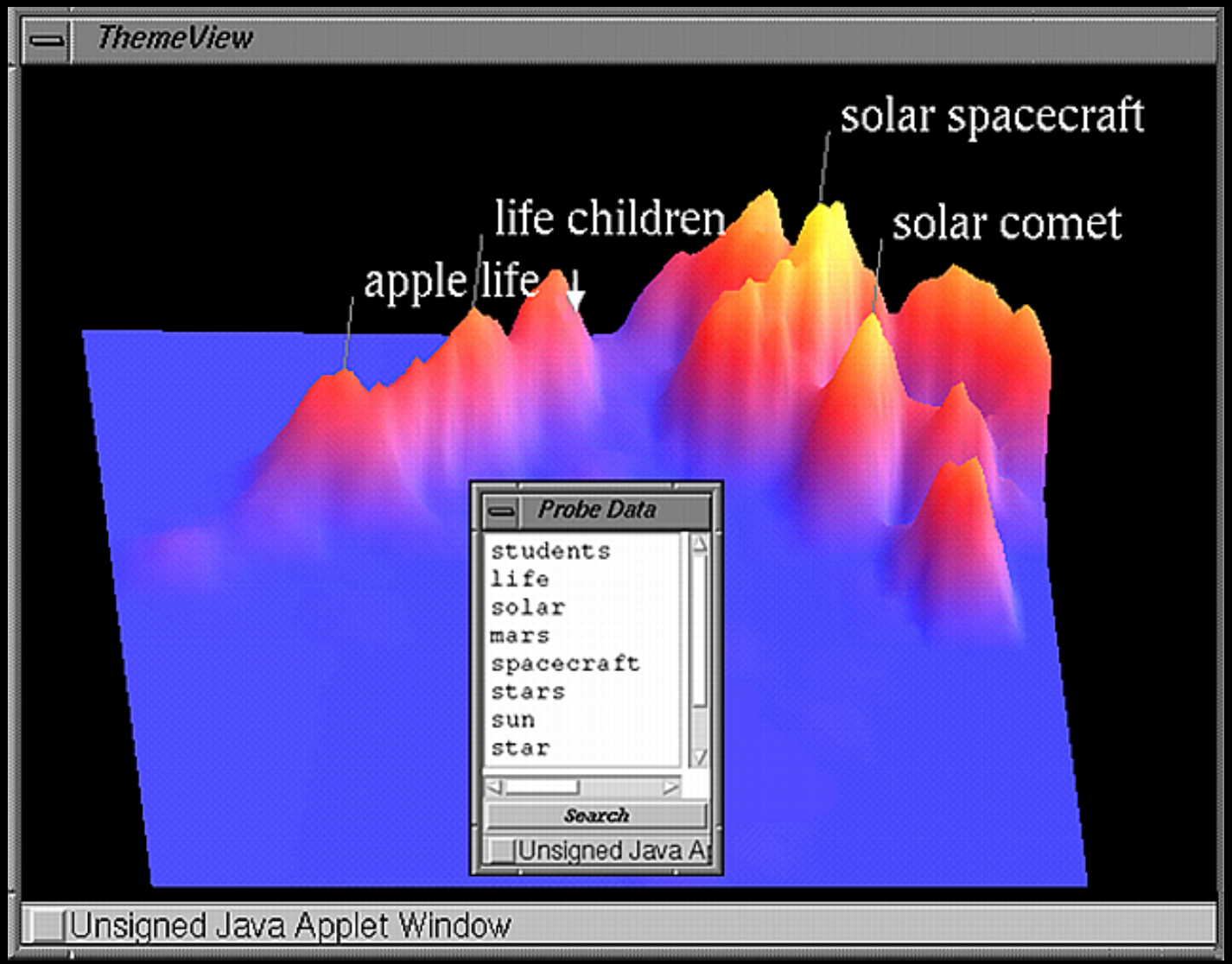

SPIRE Version 2.2.1 for Solaris (06/25/97)
Loading Database 'canlit_all'...
Database 'canlit_all' successfully loaded.
Galaxy visualization of 567,437 cancer literature abstracts

Galaxy



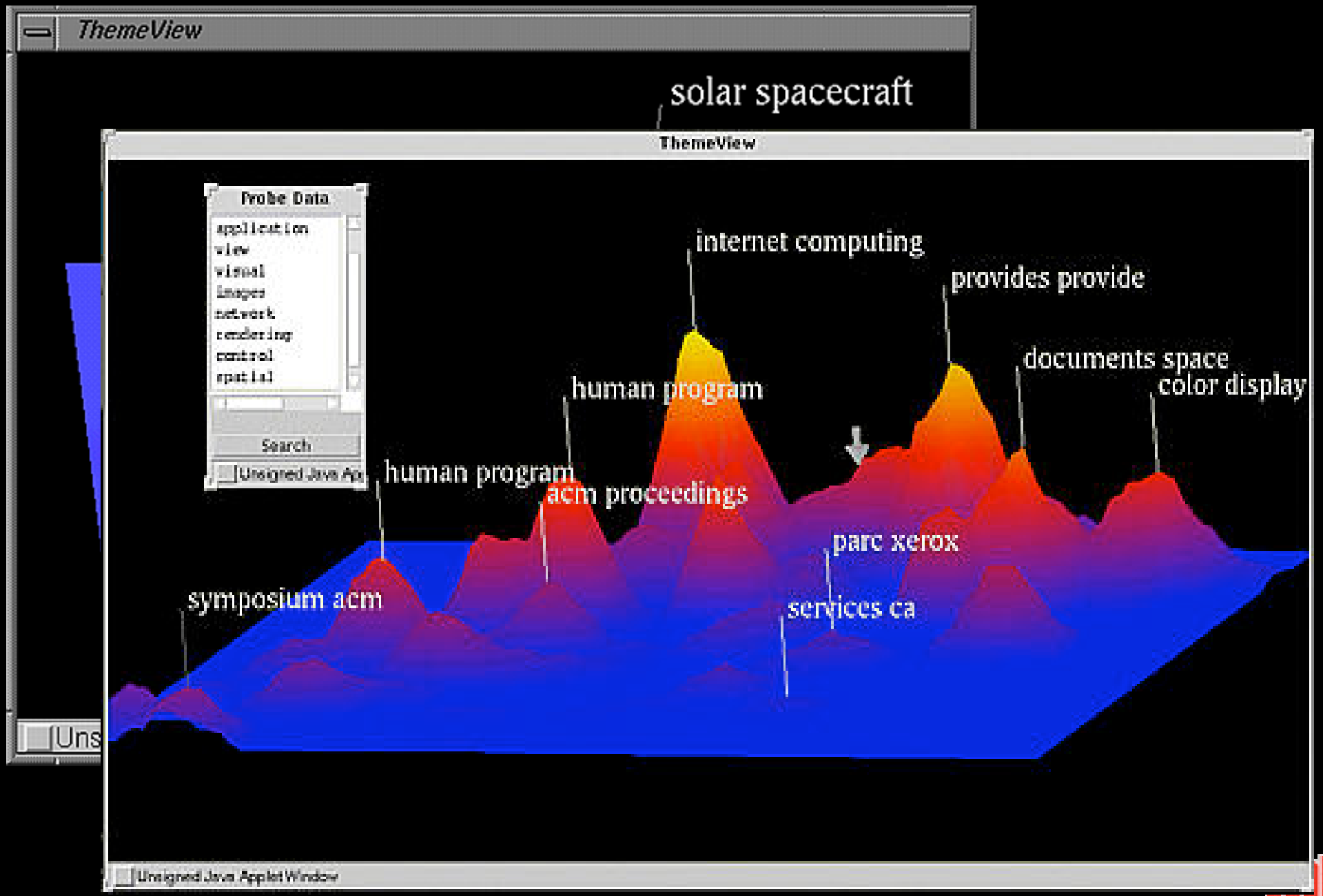
http://in-spire.pnl.gov/2shots/in-spire_tools.jpg

ThemeView



http://picturethis.pnl.gov/picturet.nsf/All/4J5Q3E?opendocument

ThemeView

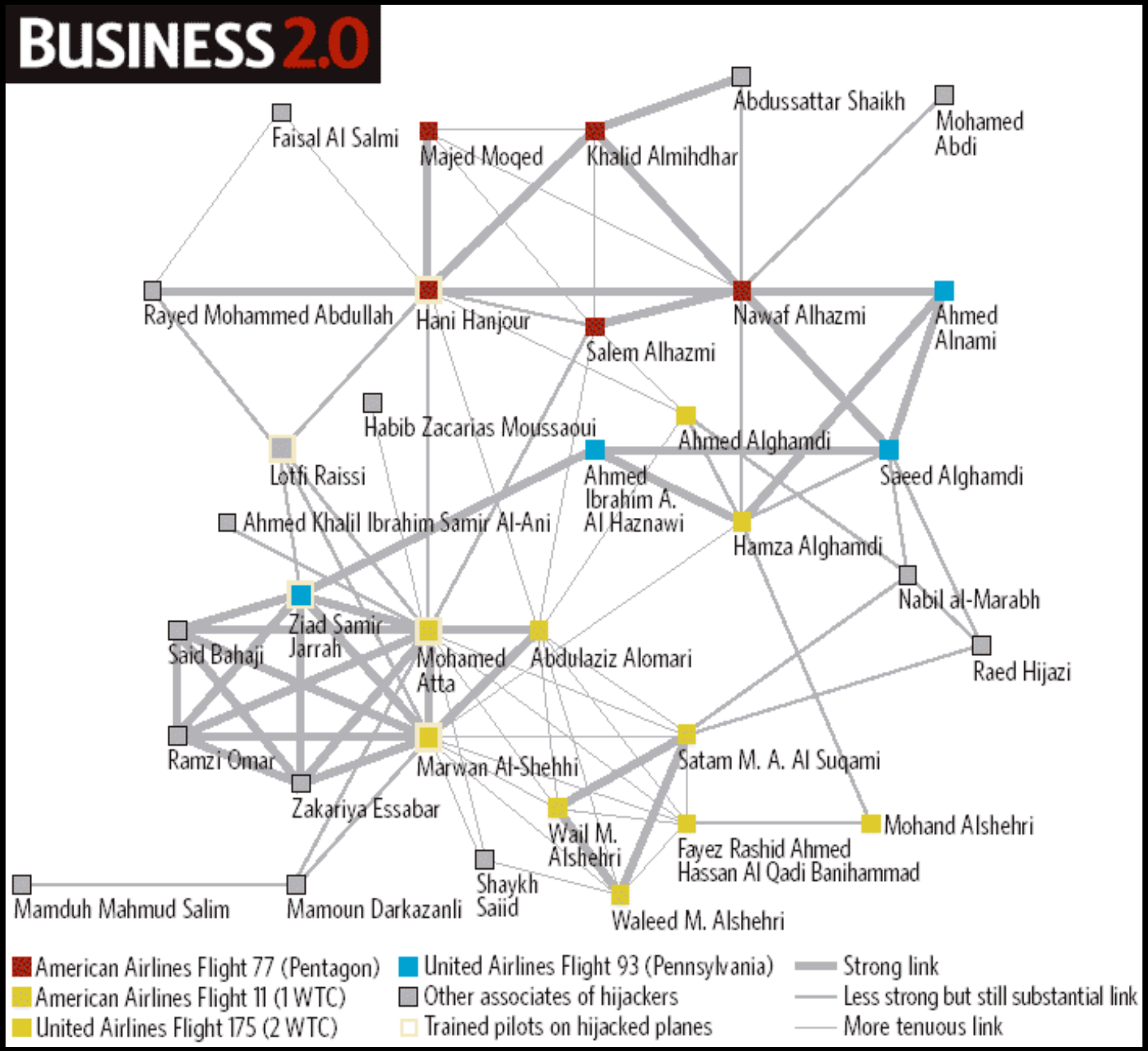


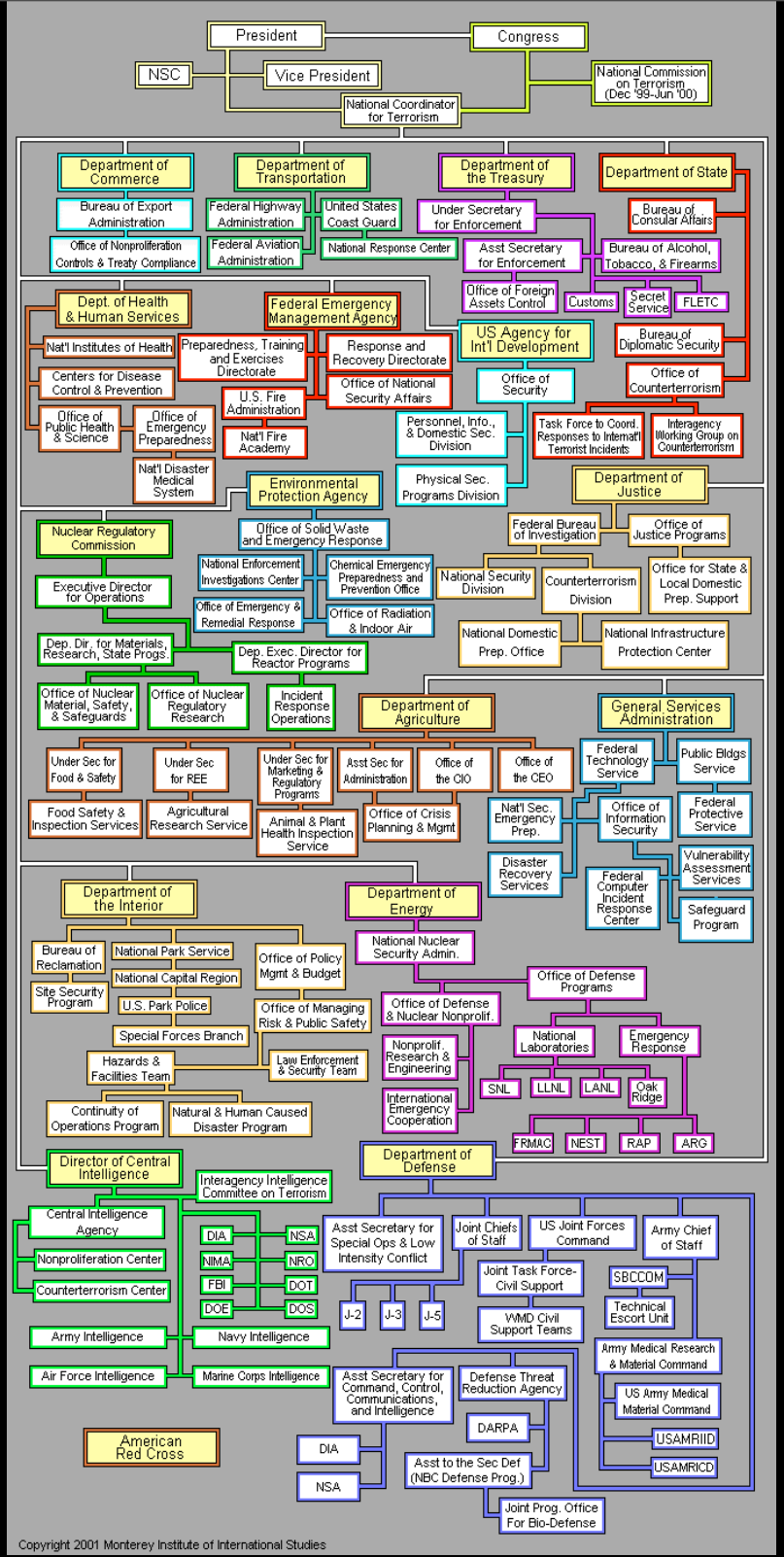
http://picturethis.pnl.gov/picturet.nsf/All/4J5Q3E?opendocument

Visualization in the Aftermath of 9/11

Six Degrees of Mohamed Atta

http://business2.com/articles/mag/0,1640,35253,FF.html





US Terrorism Response Org Chart

http://www.cns.miis.edu/research/cbw/domestic.htm#wmdchart

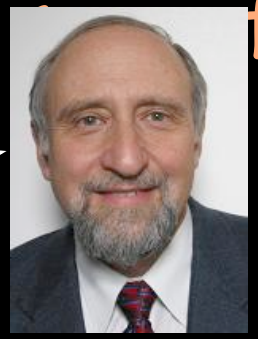
Example of finding the right application

The Treemap (Johnson & Shneiderman)

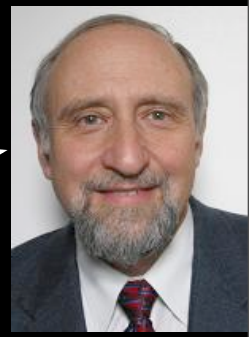
Idea:

- Show a hierarchy as a 2D layout
- Size on screen indicates relative [illegible] of underlying objects

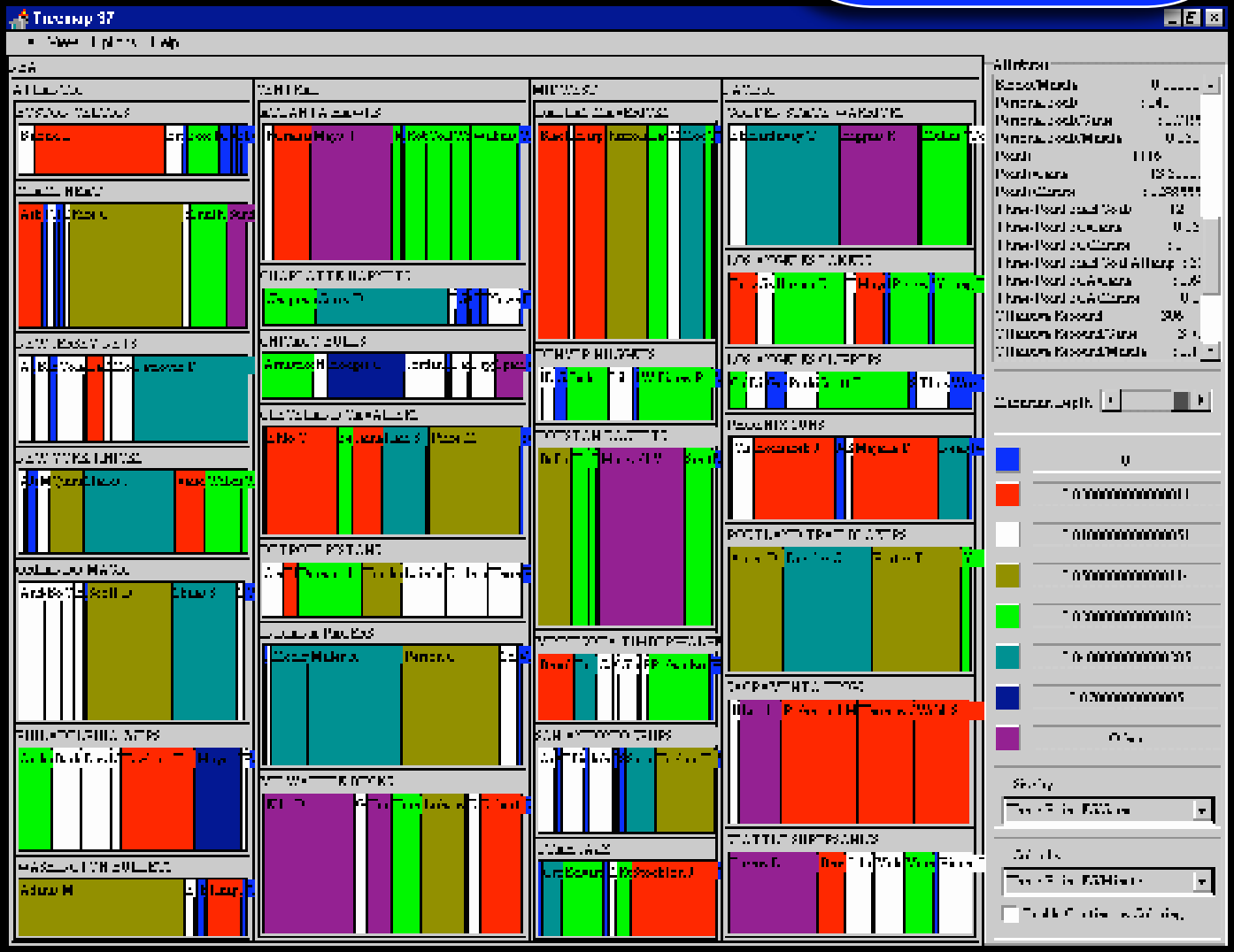
Overview, zoom & filter, details on demand



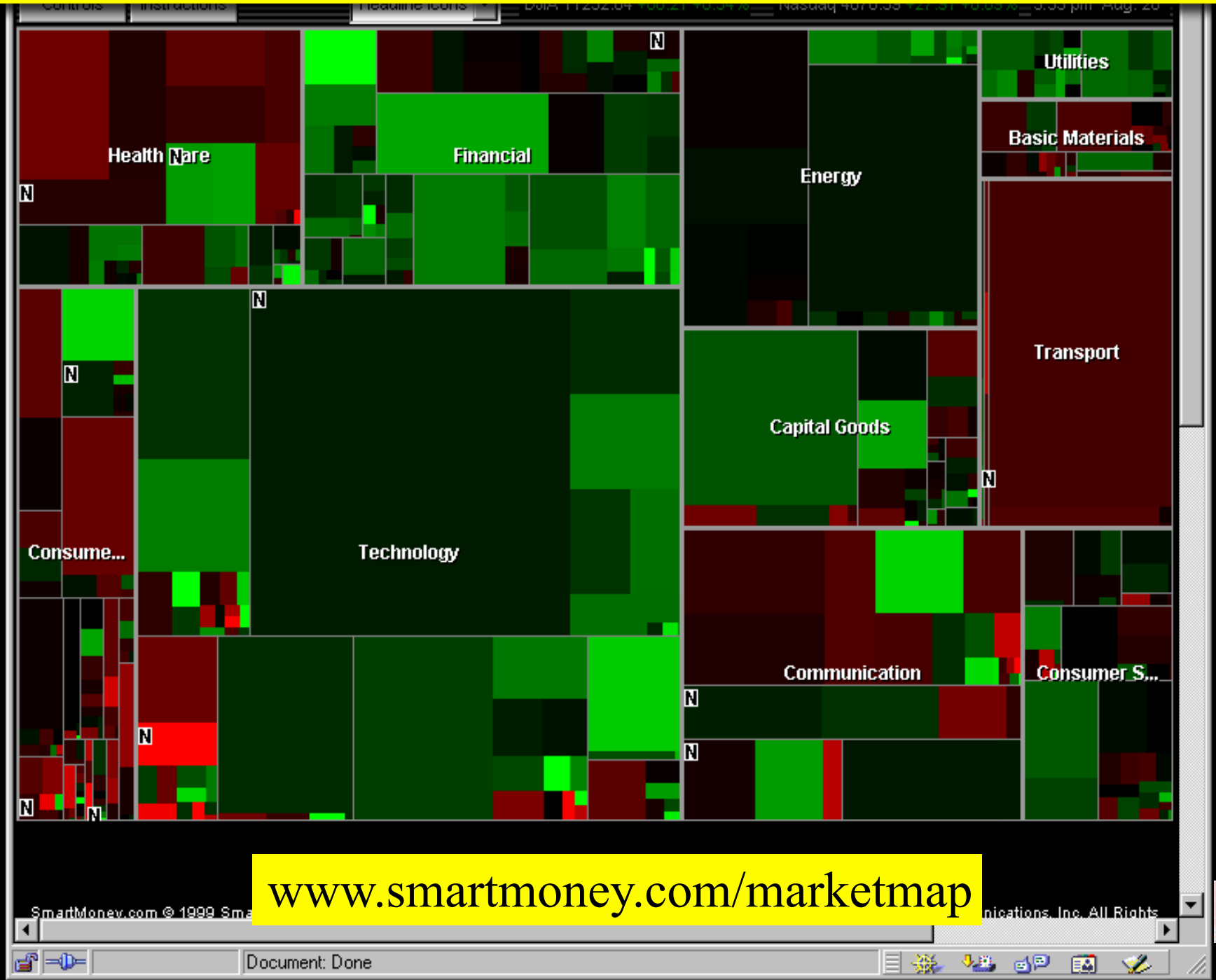
Treemap applied to File



Overview, zoom & filter, details on demand

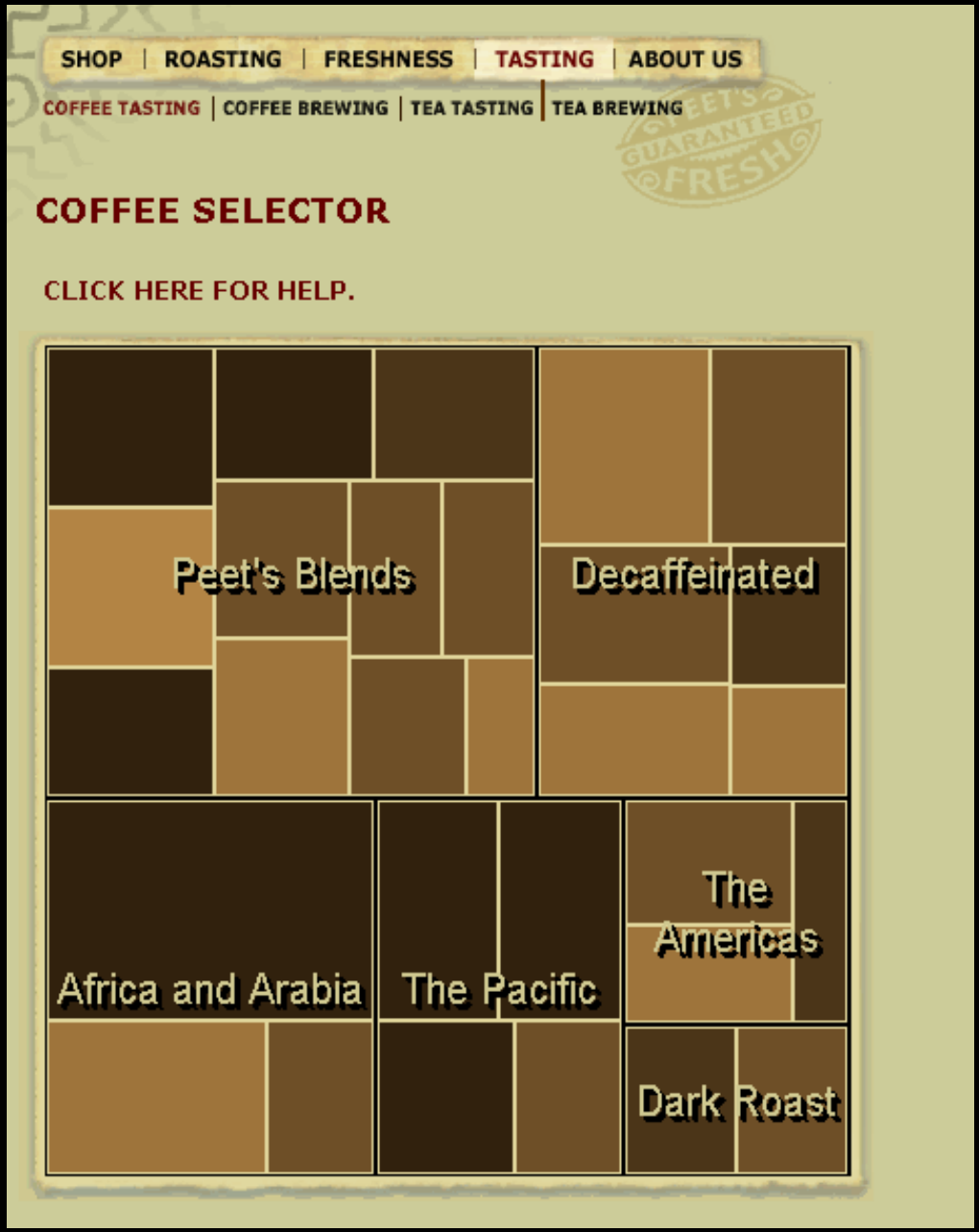


A Good Use of TreeMaps and

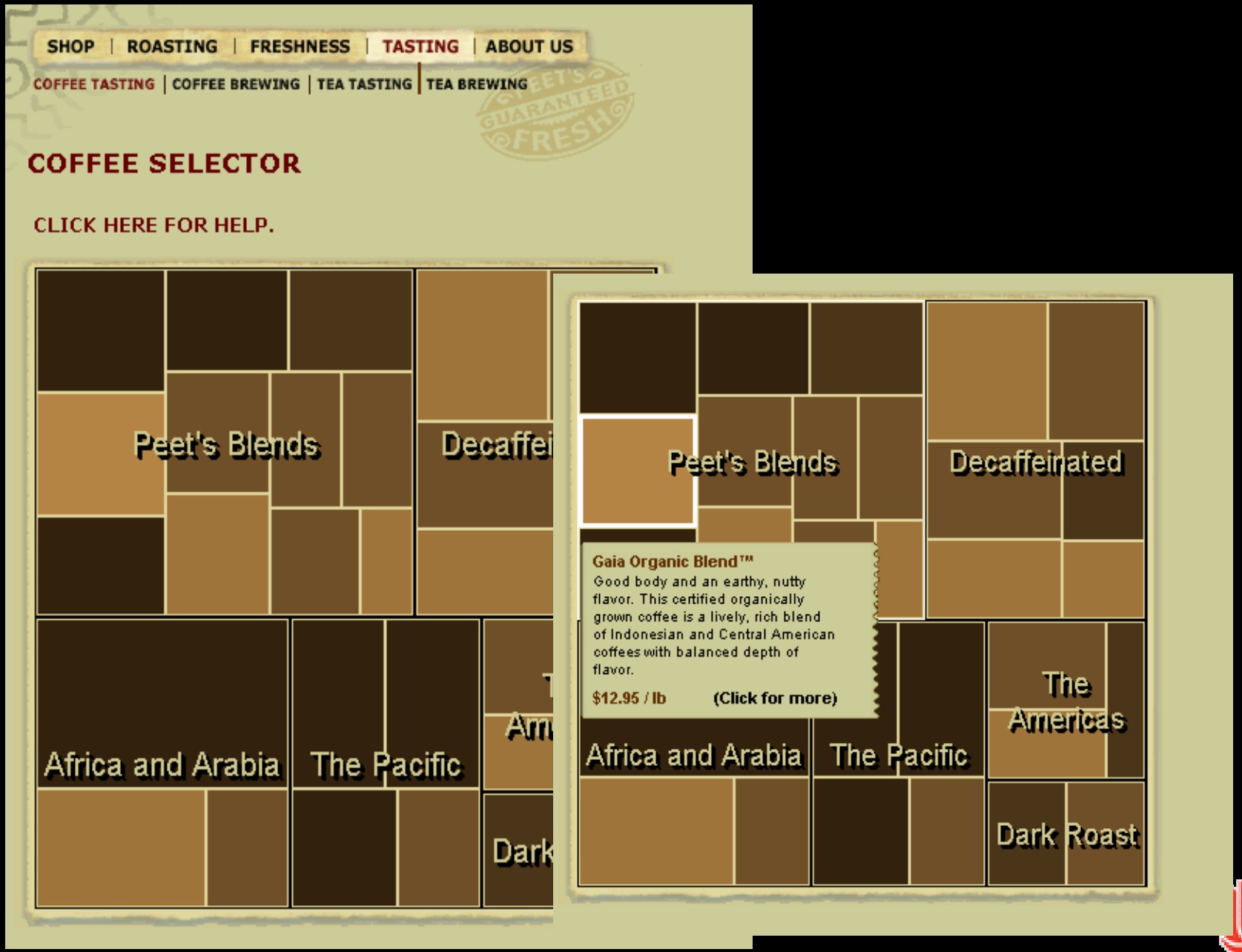


www.smartmoney.com/marketmap

Treemaps in Peets site



Treemaps in Peets site



TreeMaps in Action

http://www.smartmoney.com/maps

http://www.peets.com/selector_coffee/coffee_selector.asp

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
  - —Only two colors: can distinguish one thing from another
  - —Amount isn't very important

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

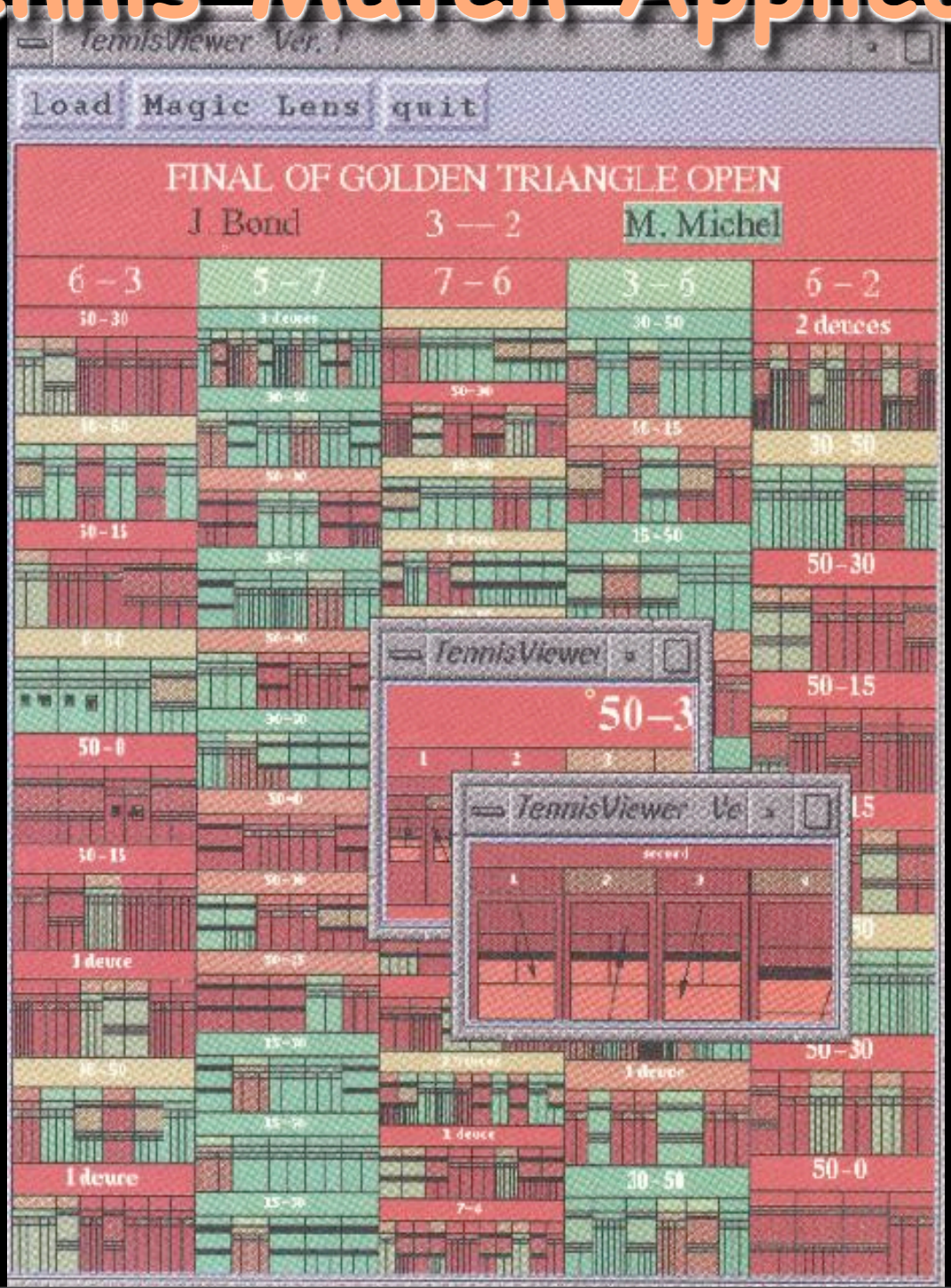
- Don't need all this to just see the largest files in the OS

Analysis vs. Communication

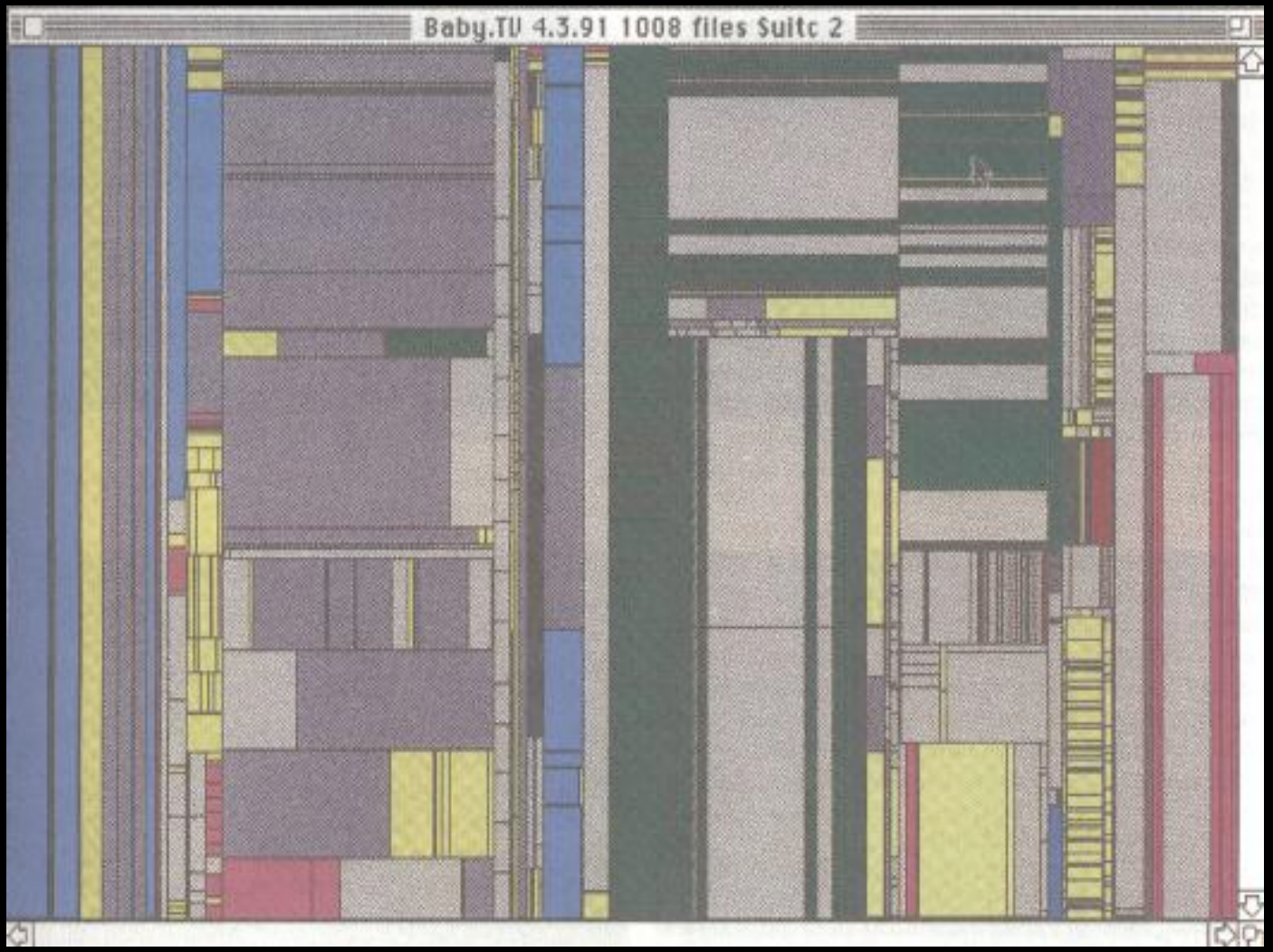
MarketMap use of treemaps allows for sophisticated analysis

Peets use of treemaps is more for presentation and communication

Tennis Match Application



File System Application



Most Widely Deployed InfoVis?

What is the most widely deployed information visualization tool (software?)

Example

Baseball statistics

Microsoft Excel - baseball

File Edit View Insert Format Tools Data Accounting Window Help

Arial 10 100%

A1 = Name

	A	B	C	D	E	F	G	H	I	J	
1	Name	At Bats	Hits	Home Run	Runs	Rbi	Walks	Years In M	Career At I	Career Hits	Car
2	STRING	INT	INT	INT	INT	INT	INT	INT	INT	INT	INT
3	Andy Allanson	293	66	1	30	29	14	1	293	66	
4	Alan Ashby	315	81	7	24	38	39	14	3449	835	
5	Alvin Davis	479	130	18	66	72	76	3	1624	457	
6	Andre Dawson	496	141	20	65	78	37	11	5628	1575	
7	Andres Galarra	321	87	10	39	42	30	2	396	101	
8	Alfredo Griffin	594	169	4	74	51	35	11	4408	1133	
9	Al Newman	185	37	1	23	8	21	2	214	42	
10	Argenis Salaza	298	73	0	24	24	7	3	509	108	
11	Andres Thomas	323	81	6	26	32	8	2	341	86	
12	Andre Thornton	401	92	17	49	66	65	13	5206	1332	
13	Alan Trammell	574	159	21	107	75	59	10	4631	1300	
14	Alex Trevino	202	53	4	31	26	27	9	1876	467	
15	Andy Van Slyk	418	113	13	48	61	47	4	1512	392	
16	Alan Wiggins	239	60	0	30	11	22	6	1941	510	
17	Bill Almon	196	43	7	29	27	30	13	3231	825	
18	Billy Beane	183	39	3	20	15	11	3	201	42	
19	Buddy Bell	568	158	20	89	75	73	15	8068	2273	
20	Buddy Biancala	190	46	2	24	8	15	5	479	102	
21	Bruce Bochte	407	104	6	57	43	65	12	5233	1478	

baseball

Ready

Data Tables

Often, we take raw data and transform it into a form that is more workable

Main idea:

- Individual items are called cases
- Cases have variables (attributes)

Data Table Format

	$Case_1$	$Case_2$	$Case_3$	...
$Variable_1$	$Value_{11}$	$Value_{21}$	$Value_{31}$	
$Variable_2$	$Value_{12}$	$Value_{22}$	$Value_{32}$	
$Variable_3$	$Value_{13}$	$Value_{23}$	$Value_{33}$	
...				

Think of as a function
$f(case_1) = <Val_{11}, Val_{12},...>$

Example

	Mary	Jim	Sally	Mitch	...
SSN	145	294	563	823	
Age	23	17	47	29	
Hair	brown	black	blonde	red	
GPA	2.9	3.7	3.4	2.1	
...					

People in class

Variable Types

Three main types of variables

- N-Nominal (equal or not equal to other values)
  - —Example: gender
- O-Ordinal (obeys < relation, ordered set)
  - —Example: fr,so,jr,sr
- Q-Quantitative (can do math on them)
  - —Example: age

Metadata

Descriptive information about the data

- Might be something as simple as the type of a variable, or could be more complex
- For times when the table itself just isn't enough
- Example: variable1 is "the current temperature measured on the South side of MEB". Or variable2 is the computed divergence of the vector field computed with this formula…. Or variable3 is the cost/benefit analysis based on :

How Many Variables?

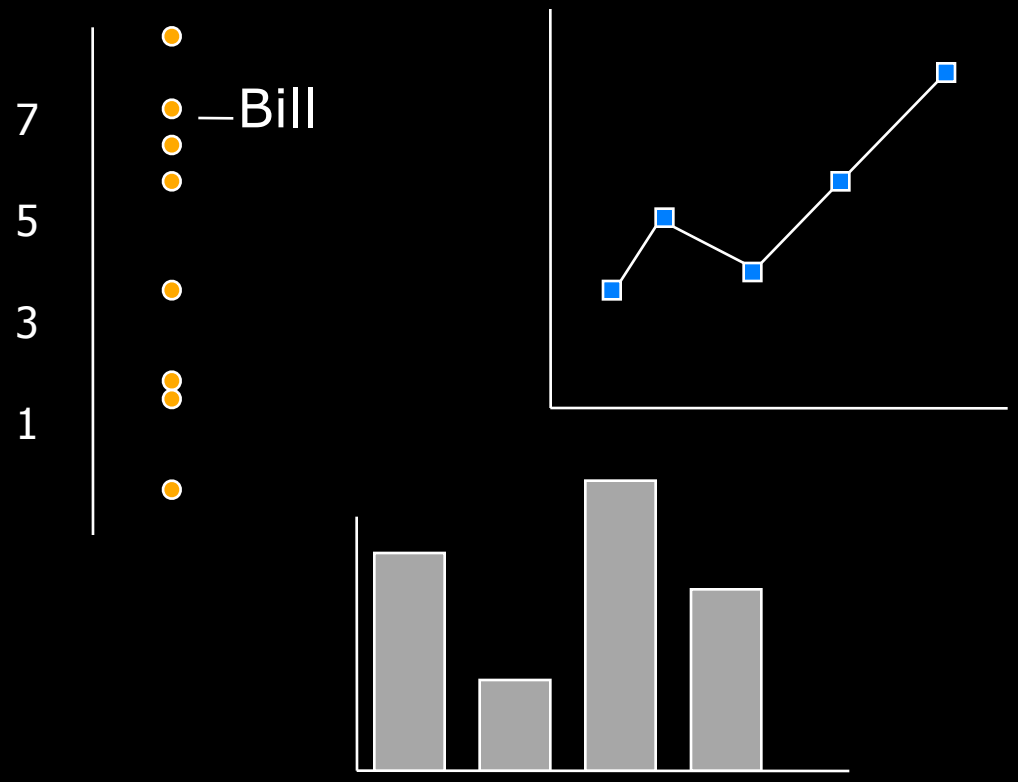
Data sets of dimensions 1,2,3 are common

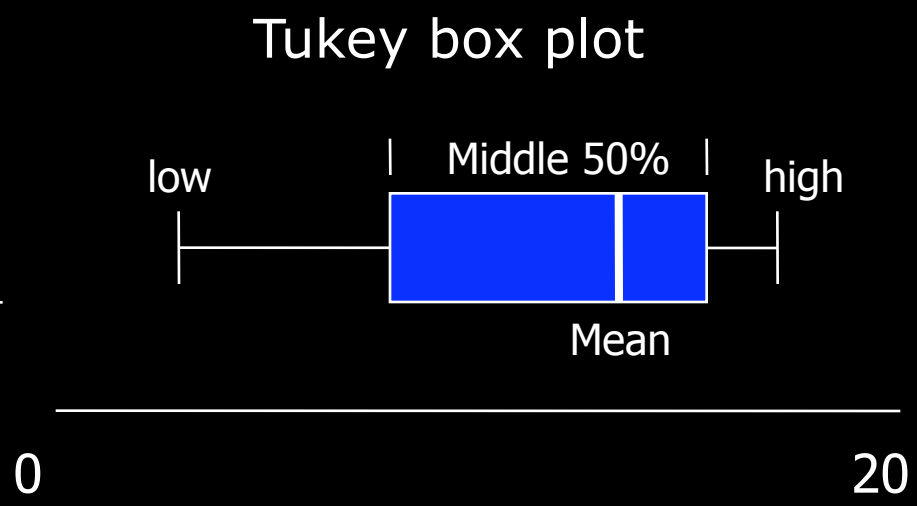
Number of variables per class

- 1 - Univariate data
- 2 - Bivariate data
- 3 - Trivariate data
- >3 - Hypervariate data

Univariate Data

Representations

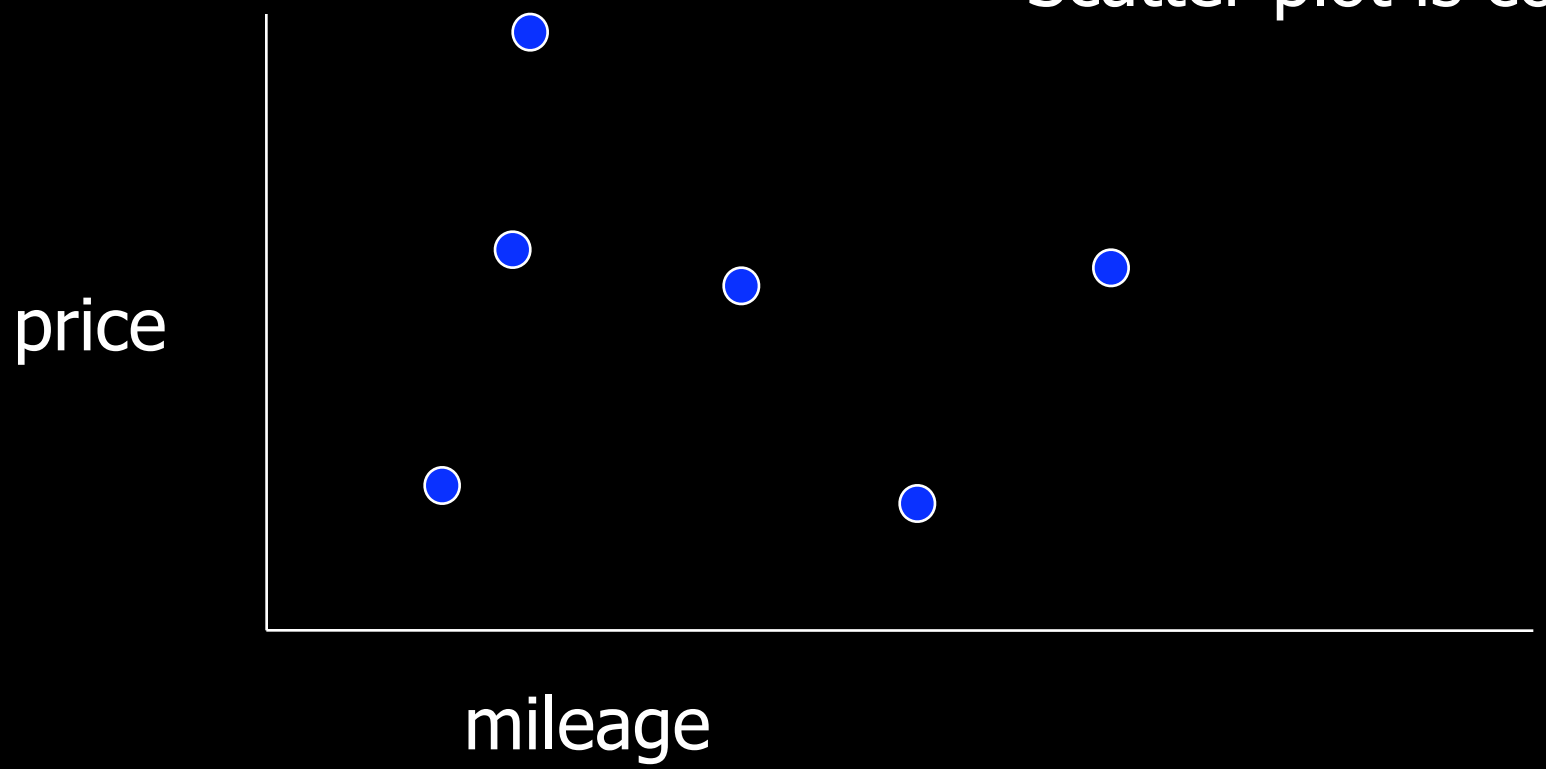




Bivariate Data

Representations

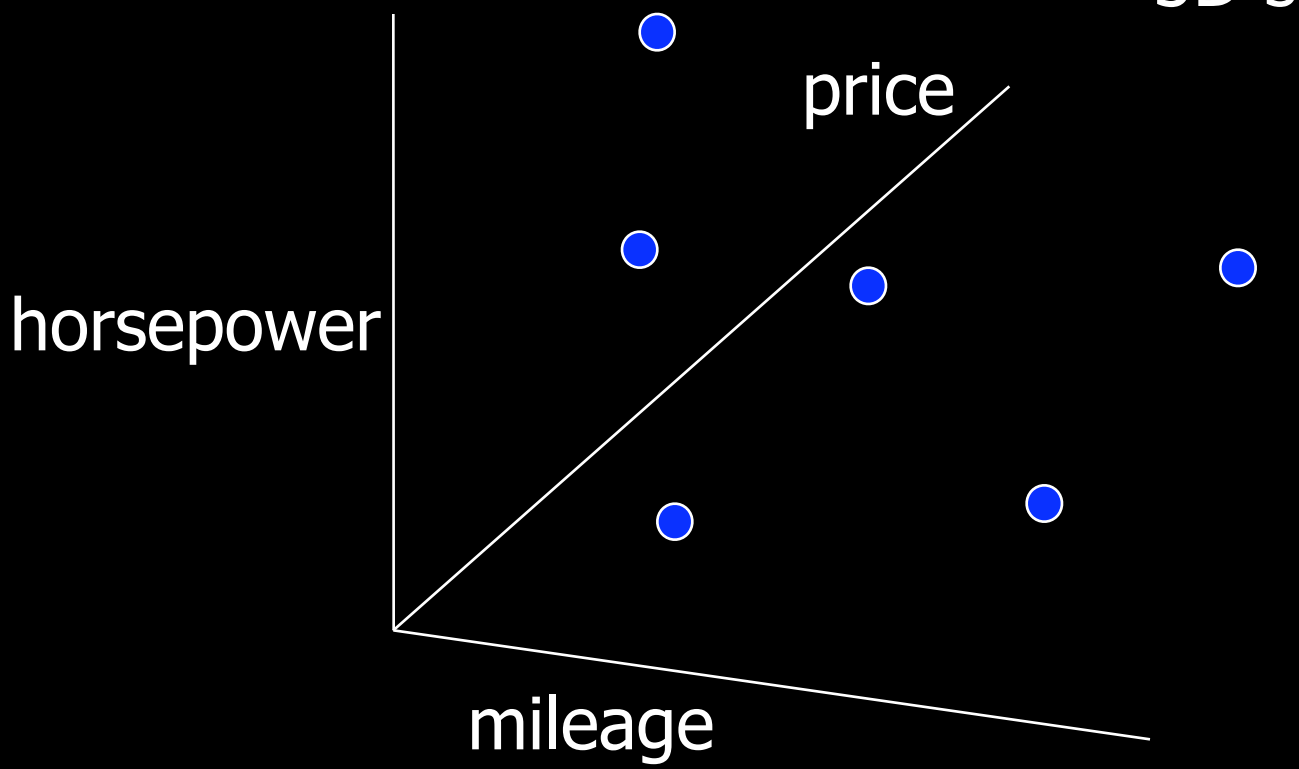
Scatter plot is common



Trivariate Data

Representations

3D scatter plot is possible



Hypervariate Data

Number of well-known visualization techniques exist for data sets of 1-3 dimensions

- line graphs, bar graphs, scatter plots OK
- We see a 3-D world (4-D with time)

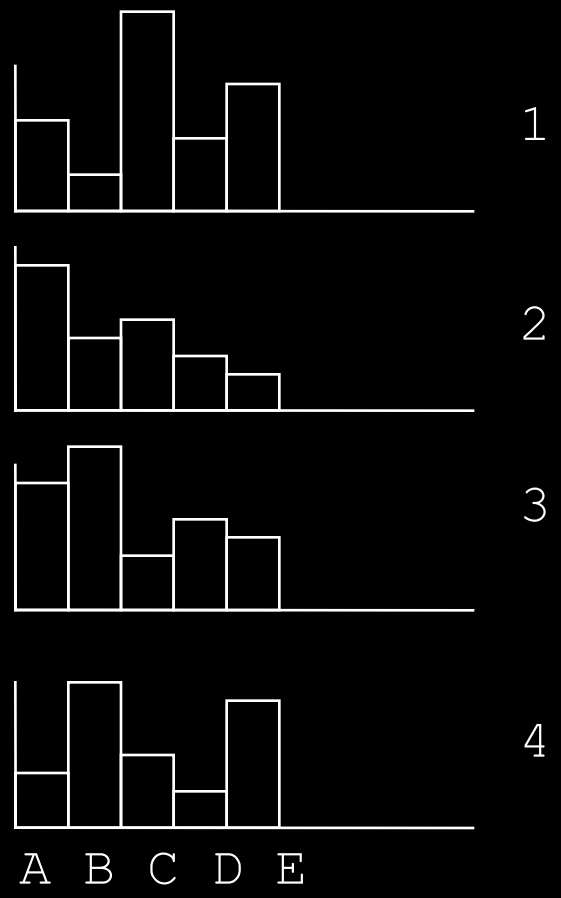
What about data sets with more than 3 variables?

- Often the interesting ones

Multiple Views

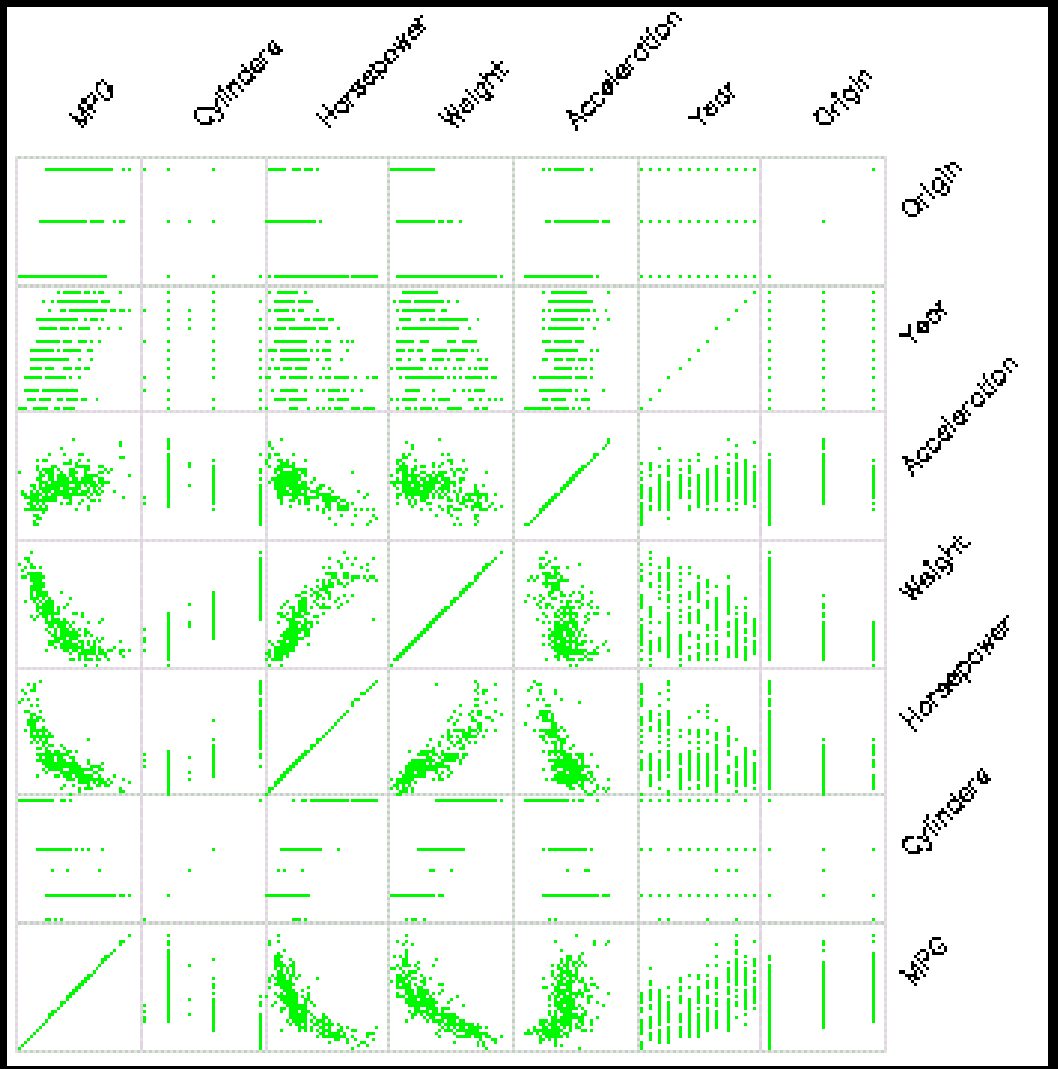
Give each variable its own display

	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5



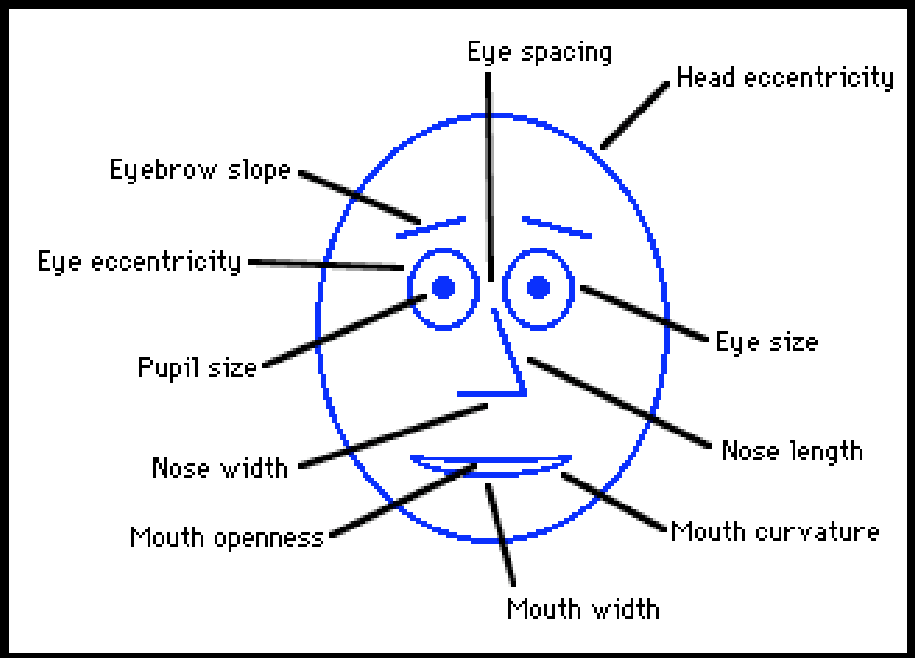
Scatterplot Matrix

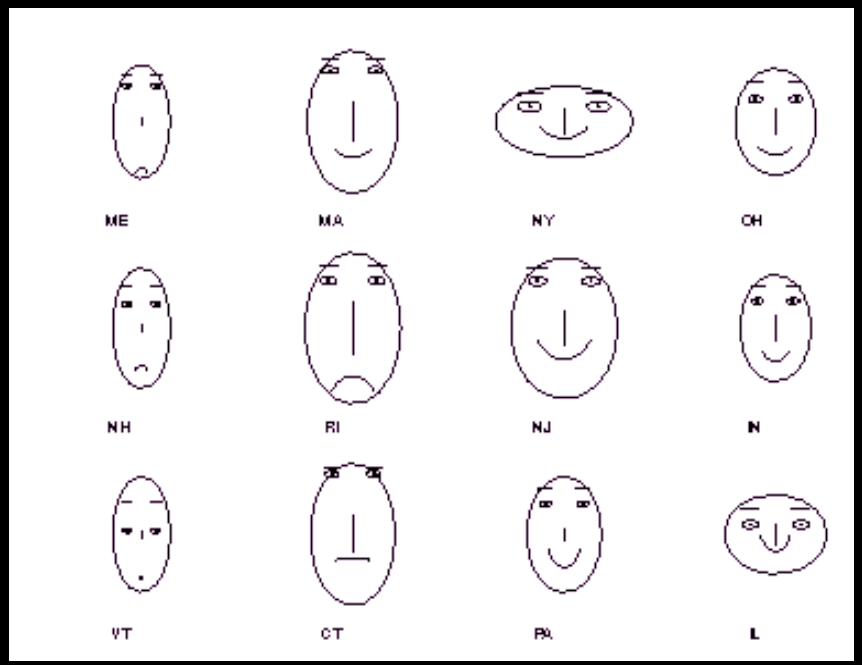
Represent each possible pair of variables in their own 2-D scatterplot



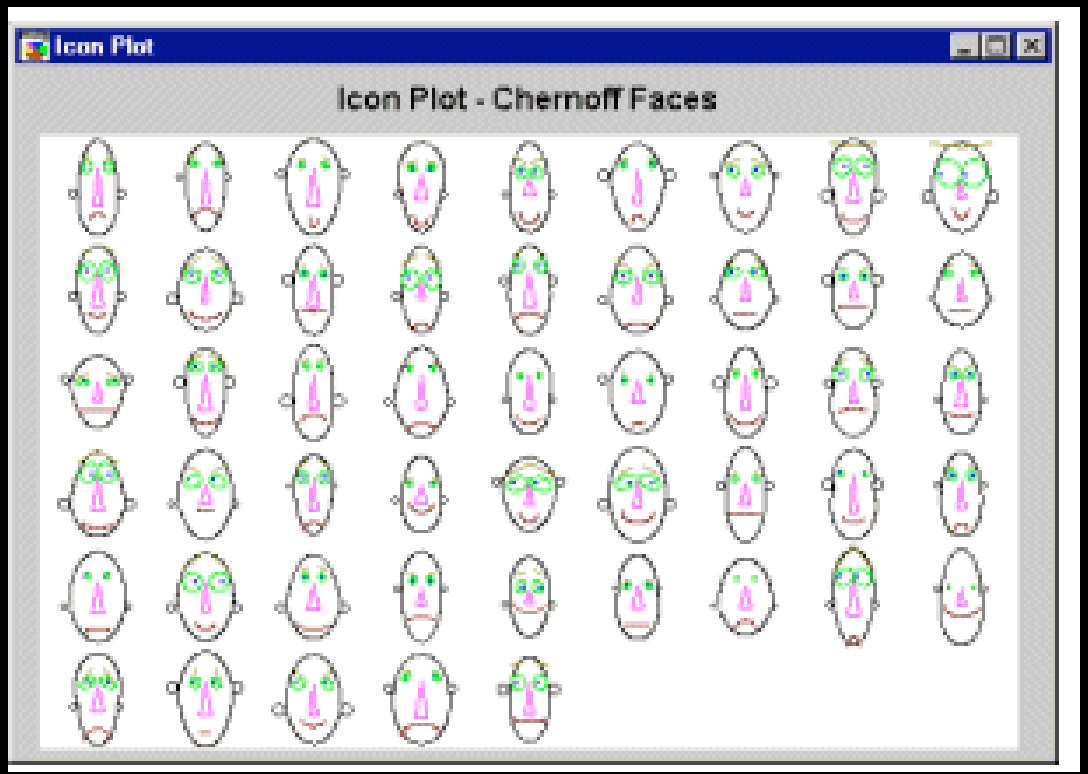
Chernoff Faces

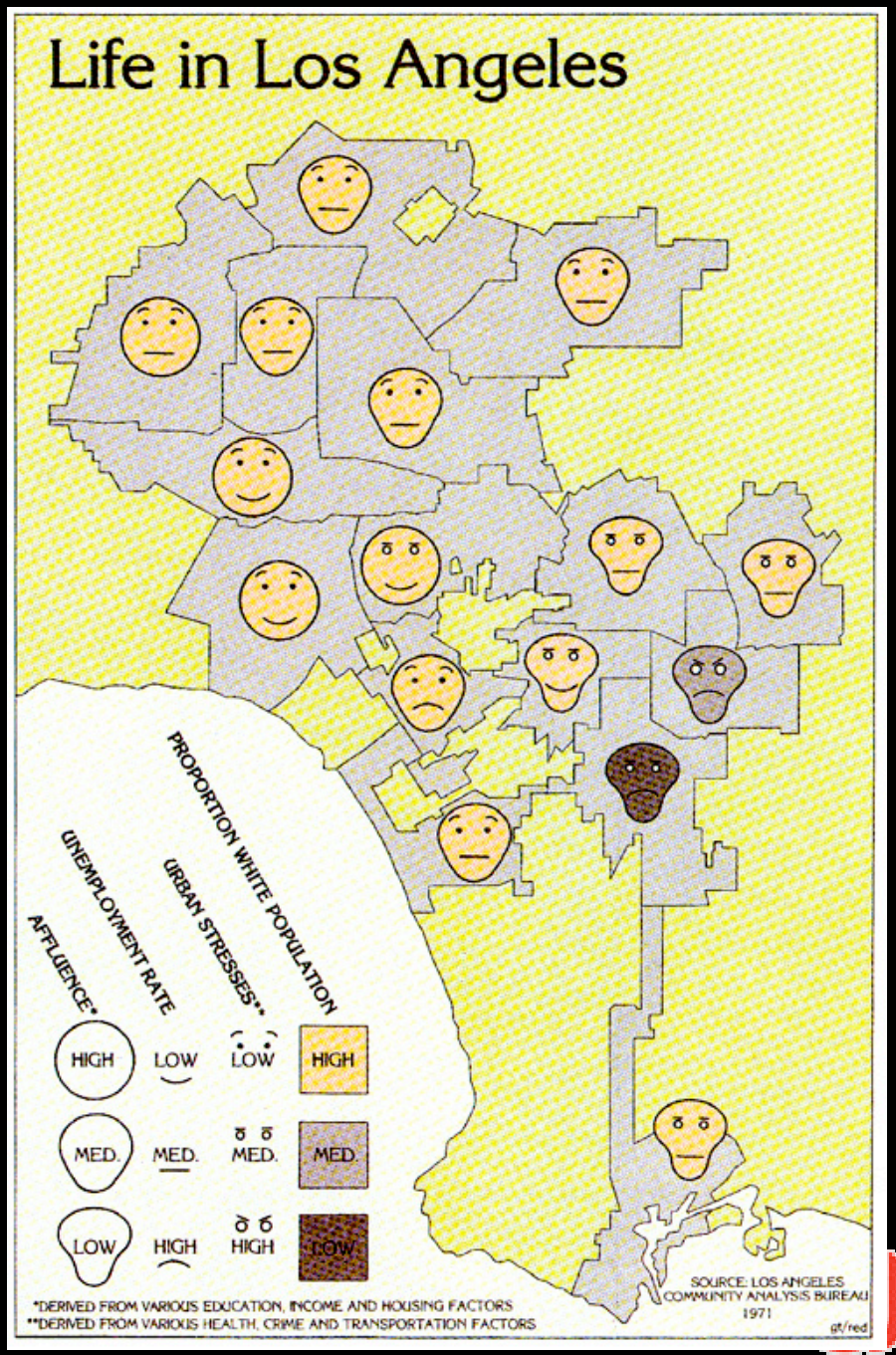
Encode different variables' values in characteristics of human face



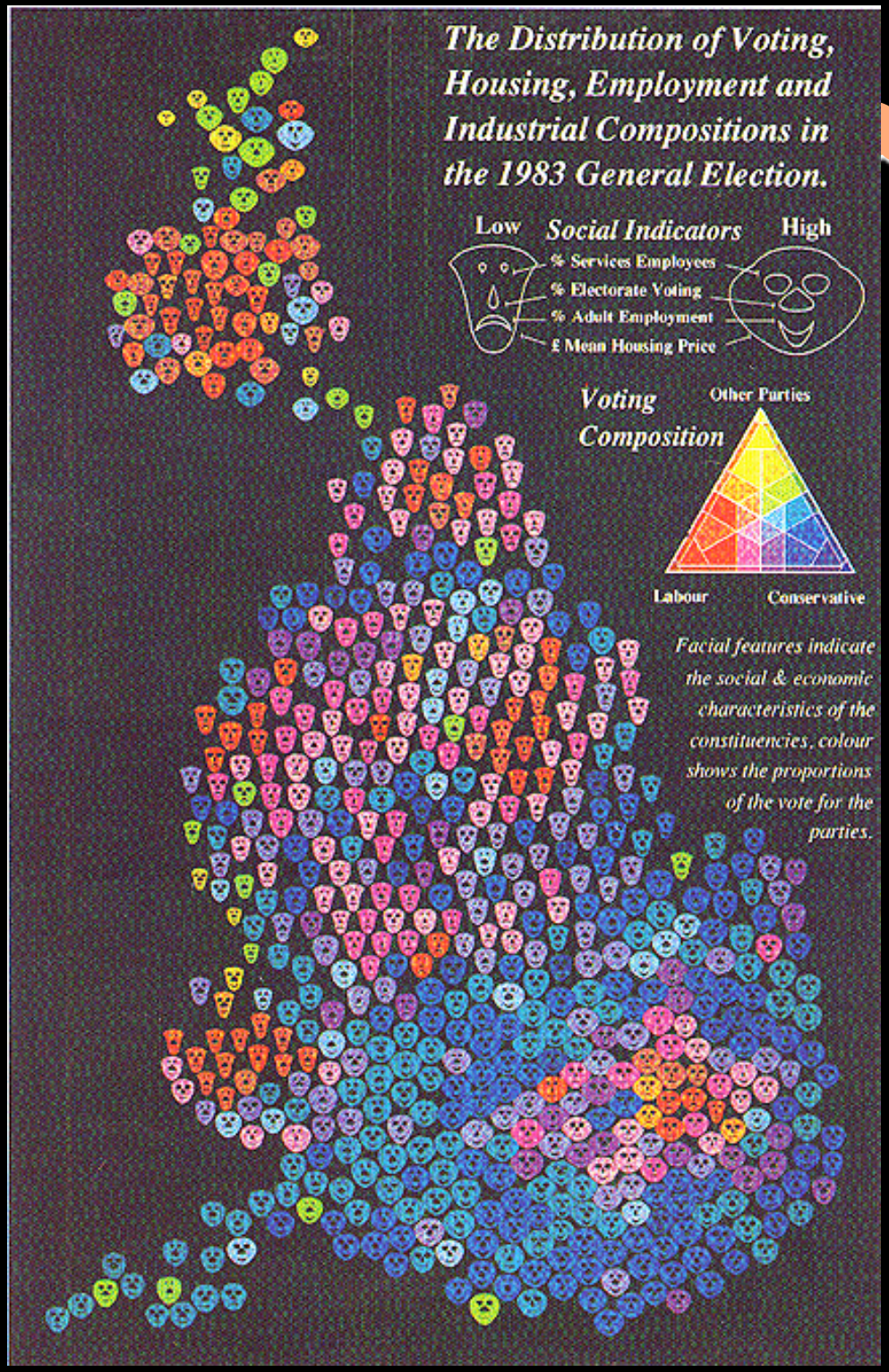


Chernoff Faces

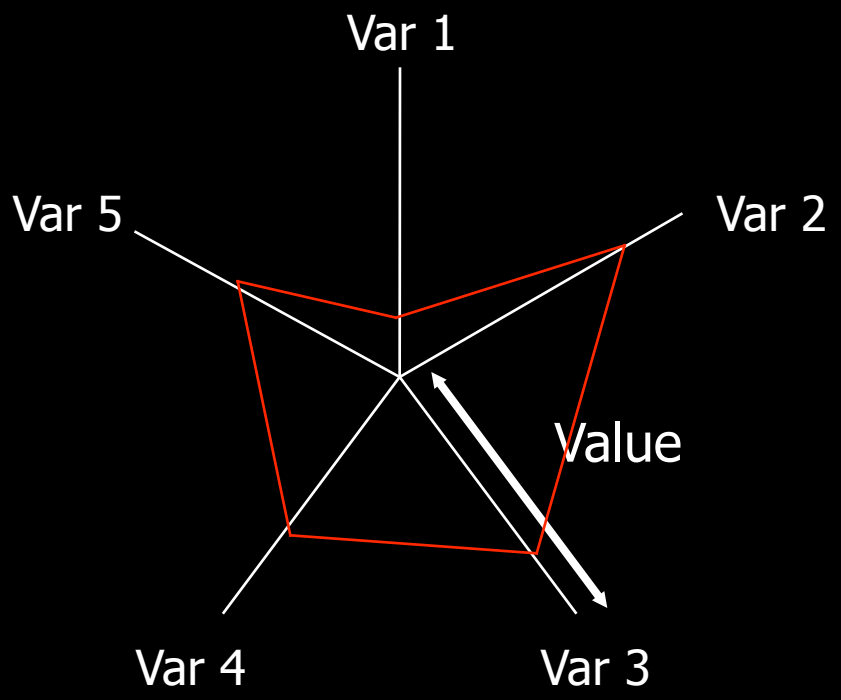




...ces


The Distribution of Voting, Housing, Employment and Industrial Compositions in the 1983 General Election.
Low
Social Indicators
High
% Services Employees
% Electorate Voting
% Adult Employment
£ Mean Housing Price
Voting Composition
Other Parties
Labour
Conservative
Facial features indicate the social & economic characteristics of the constituencies, colour shows the proportions of the vote for the parties.

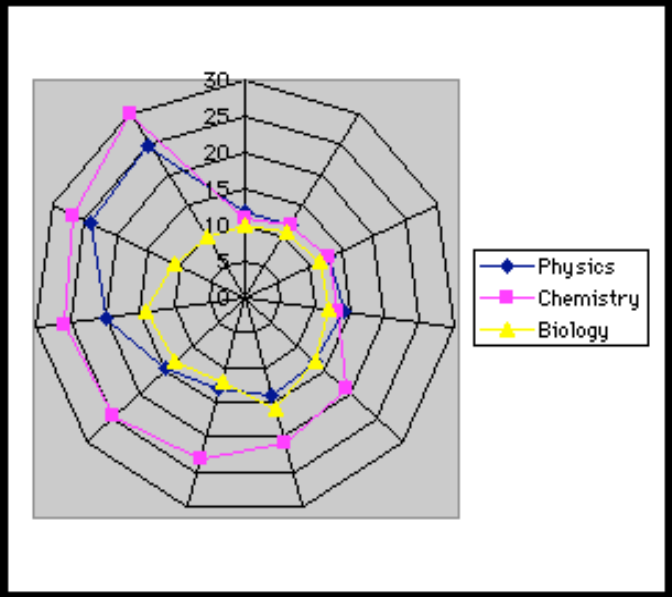
Star Plots

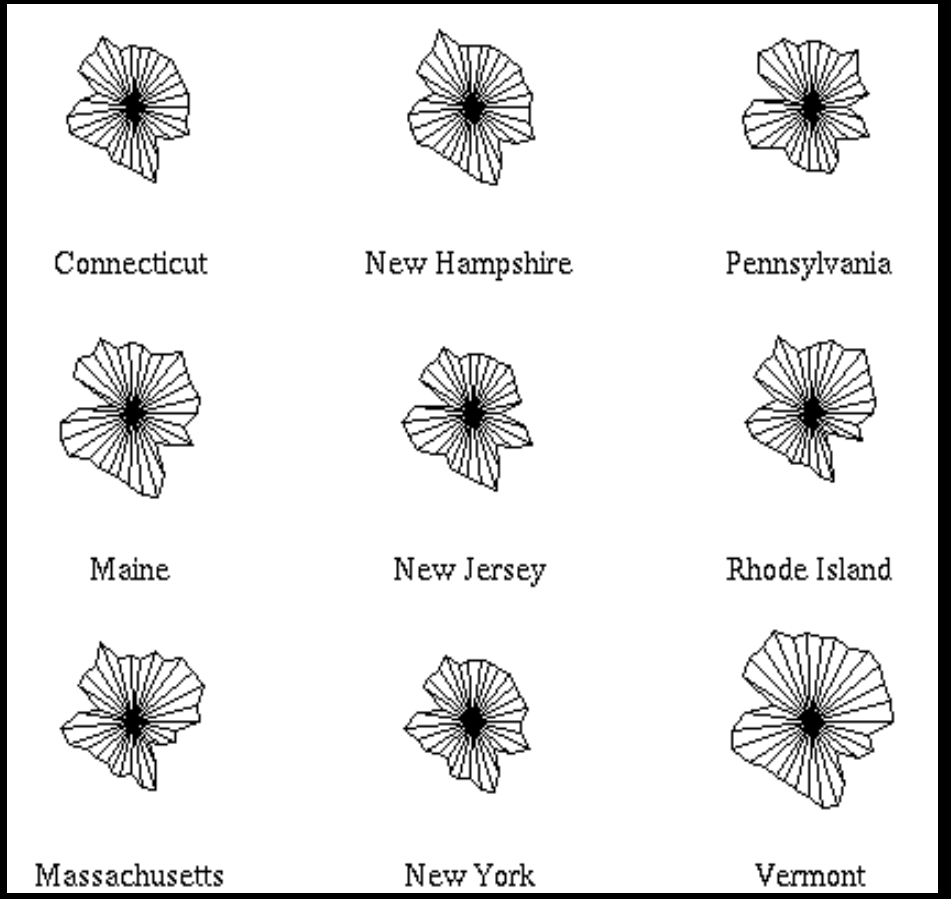


Space out the n variables at equal angles around a circle

Each "spoke" encodes a variable's value

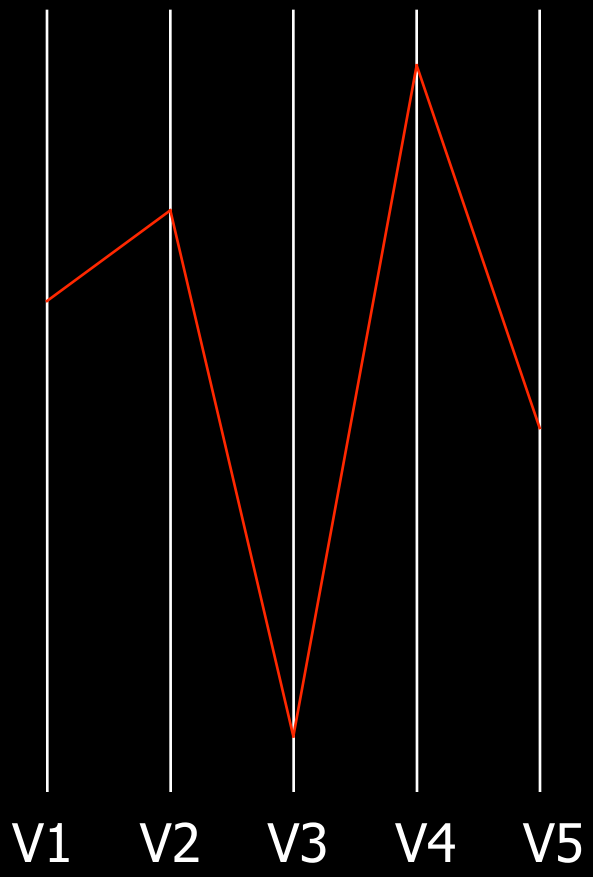
Star Plot examples





http://seamonkey.ed.asu.edu/~behrens/asu/reports/compre/comp1.html

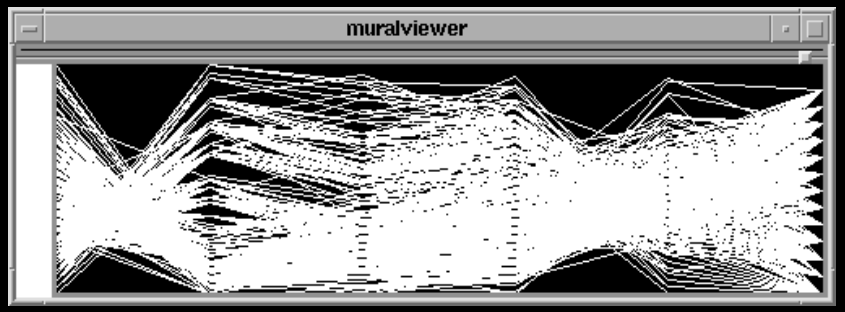
Parallel Coordinates



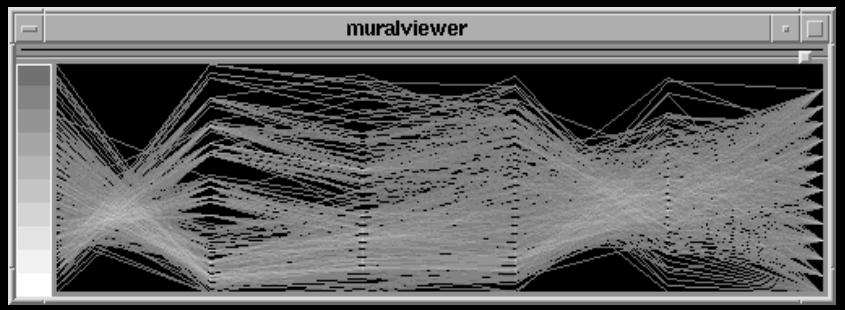
Encode variables along a horizontal row

Vertical line specifies values

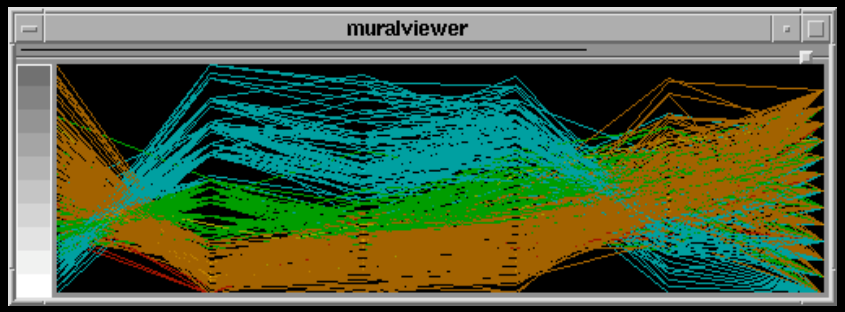
Parallel Coords Example



Basic



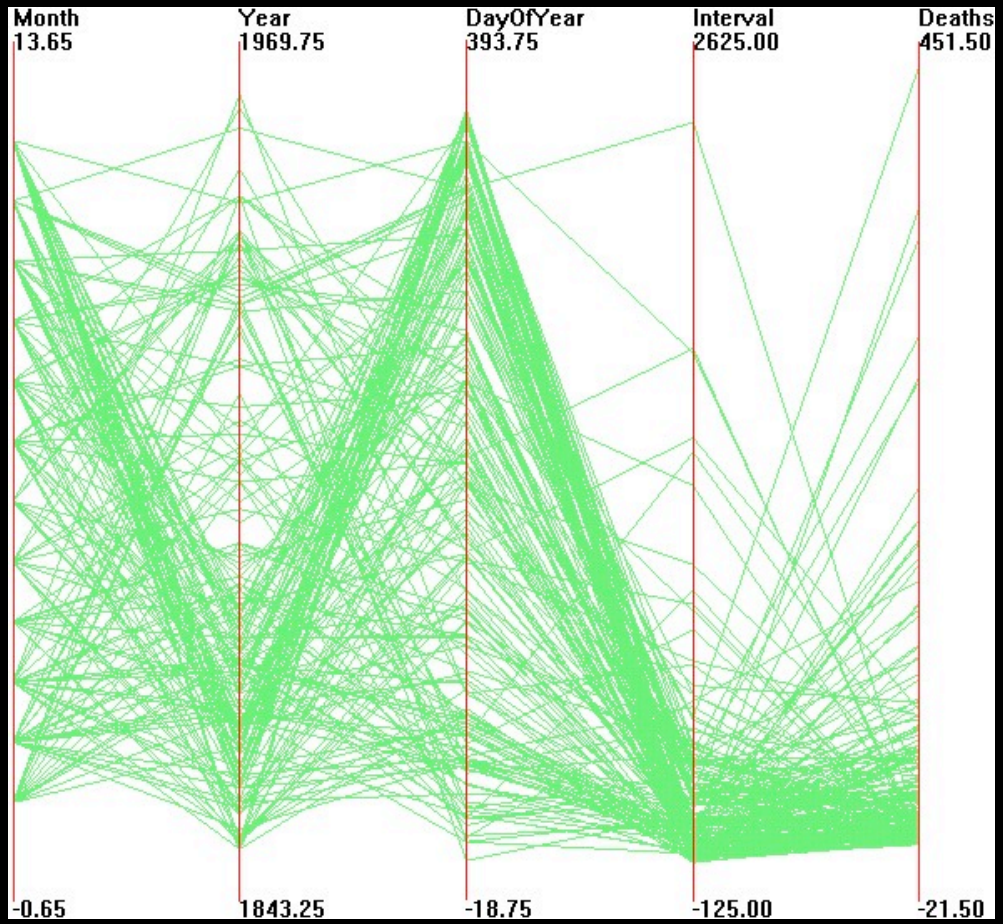
Grayscale



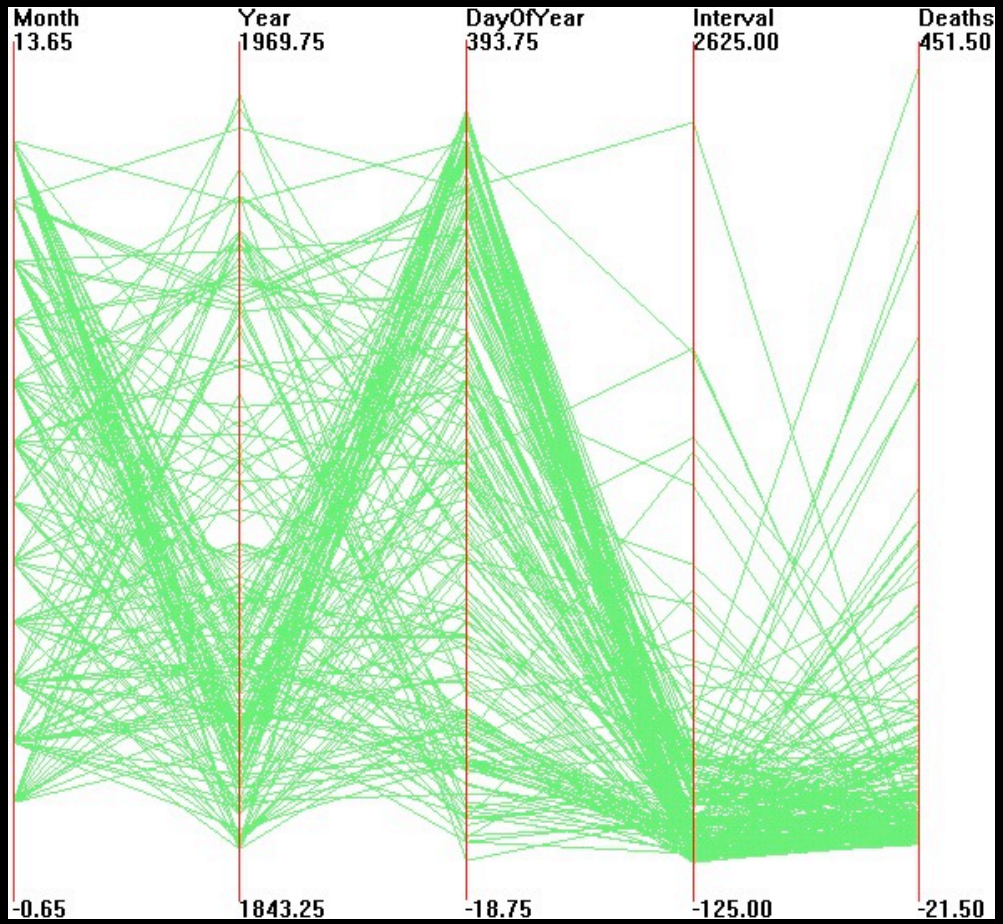
Color

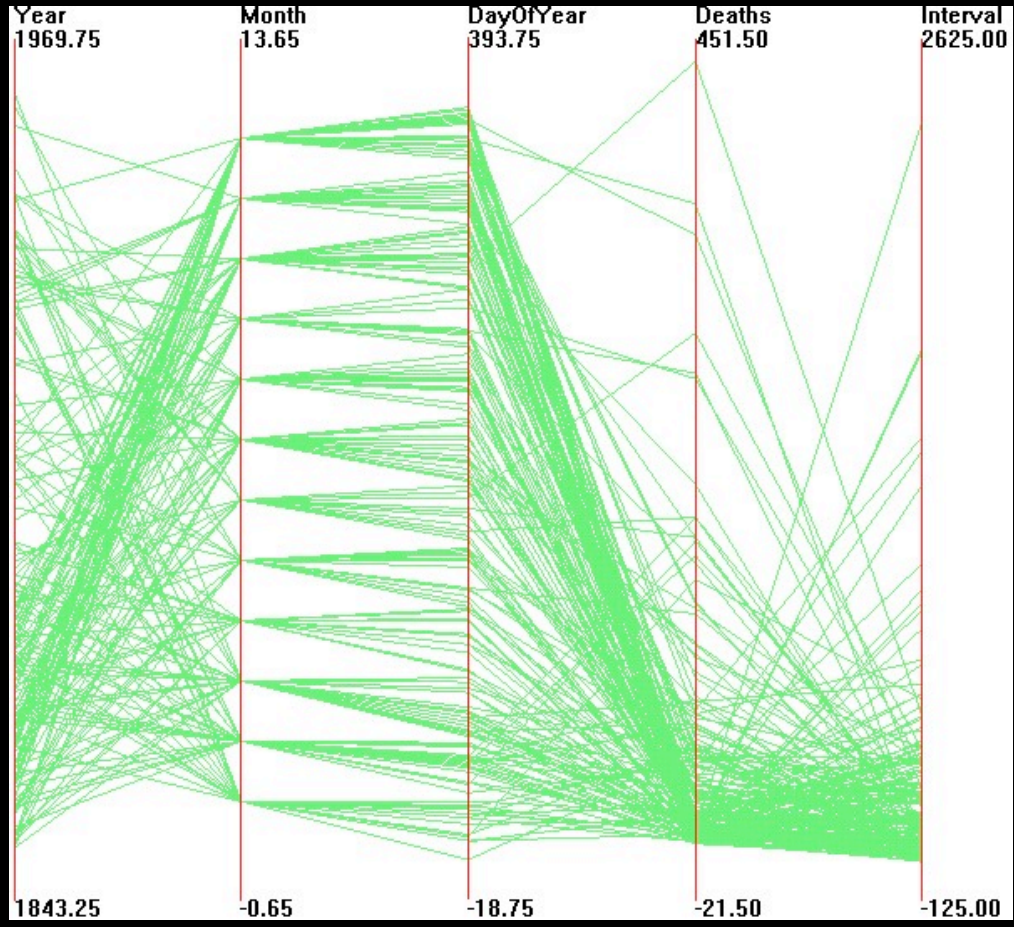
Why Does Order Matter?

Why Does Order Matter?

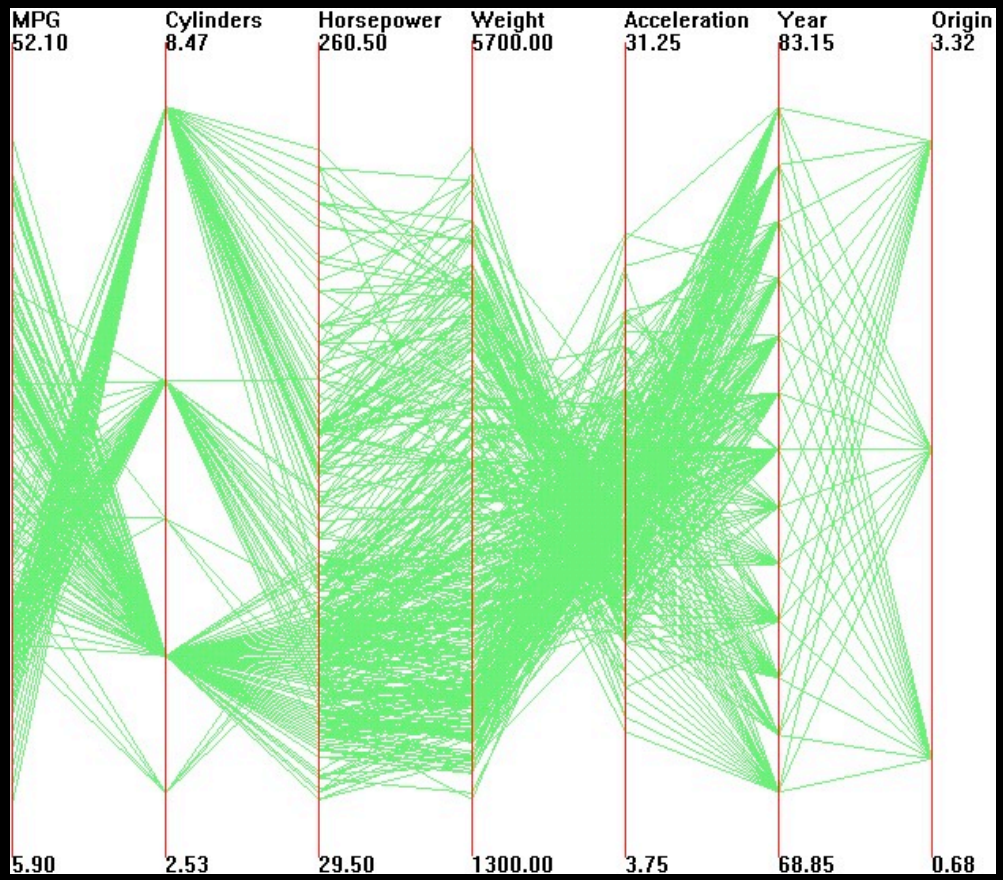


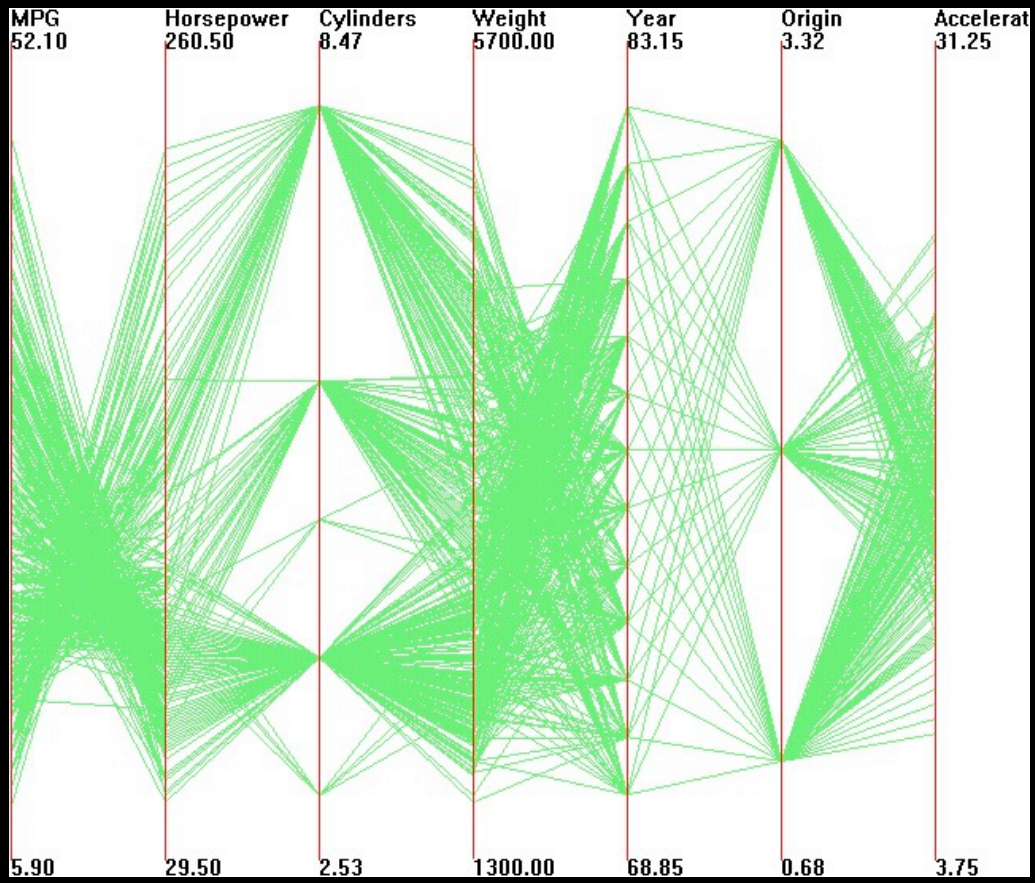
Why Does Order Matter?



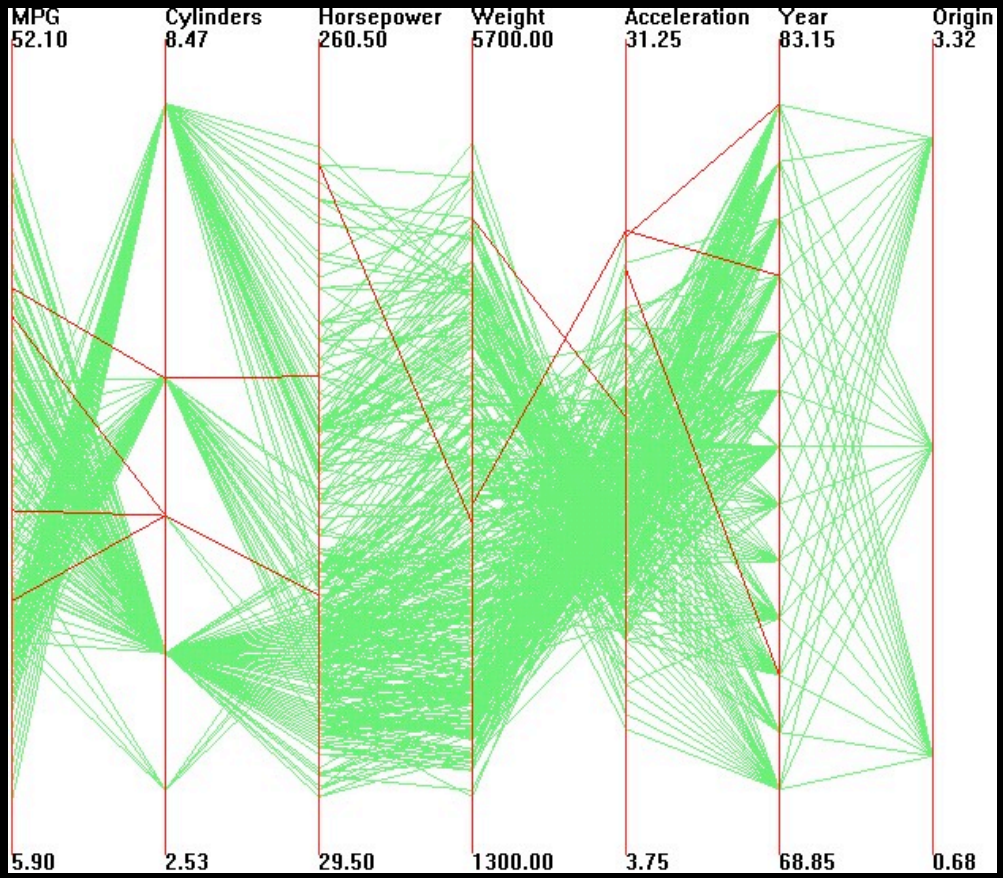


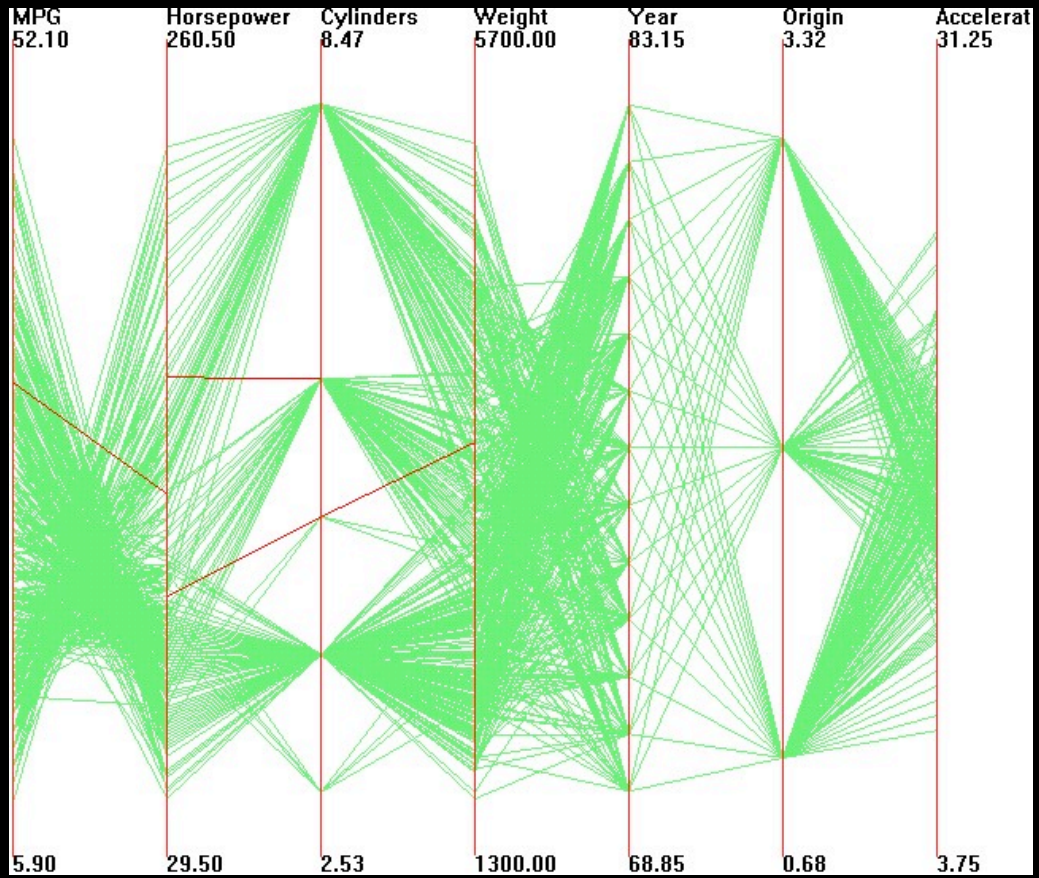
Example


MPG
52.10
Cylinders
8.47
Horsepower
260.50
Weight
5700.00
Acceleration
31.25
Year
83.15
Origin
3.32
5.90
2.53
29.50
1300.00
3.75
68.85
0.68


MPG
52.10
Horsepower
260.50
Cylinders
8.47
Weight
5700.00
Year
83.15
Origin
3.32
Accelerat
31.25
5.90
29.50
2.53
1300.00
68.85
0.68
3.75

Outliers Highlighted





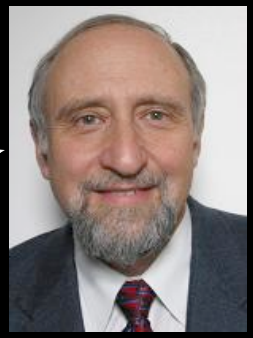
||-coords demo

http://www.cs.uta.fi/~hs/pce/

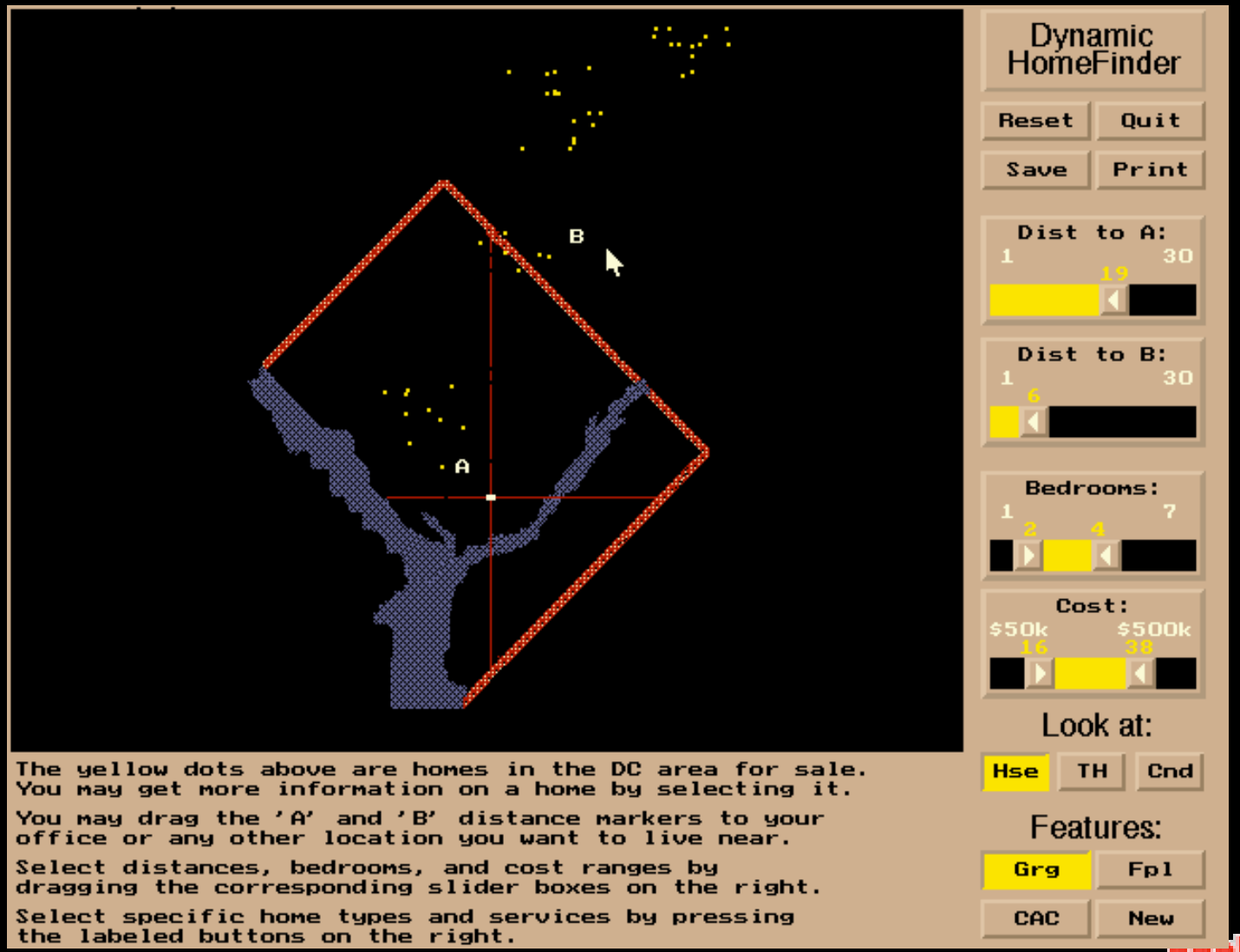
http://www.stat.sc.edu/~west/bradley/

HomeFinder

Overview, zoom & filter, details on demand



HCIL
Univ. Maryland



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!

Hierarchies (Trees)

Definition

- Data repository in which cases are related to subcases

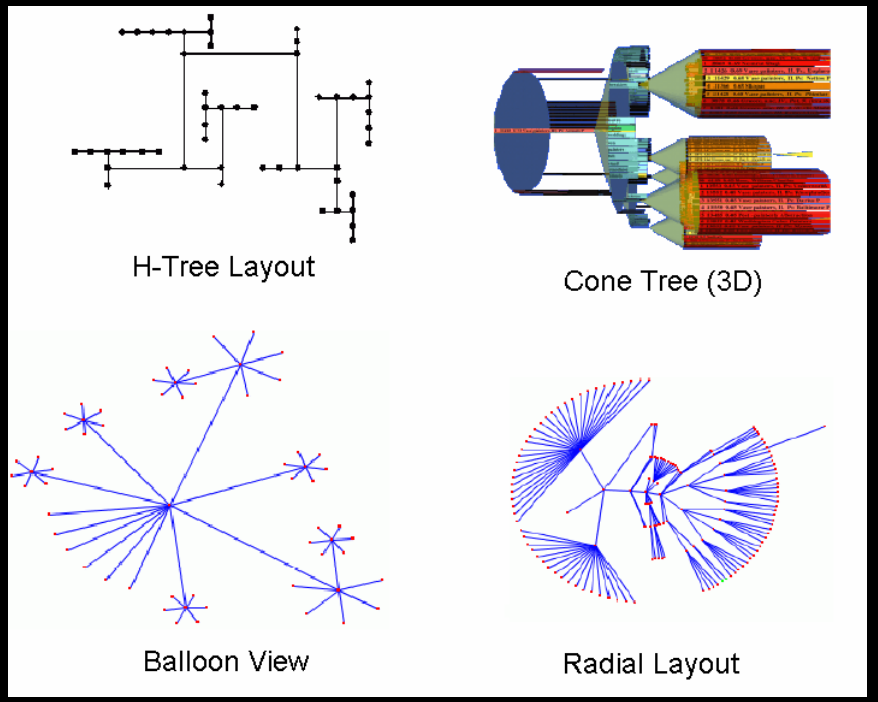
Pervasive

- Family histories, ancestries
- File/directory systems on computers
- Organization charts
- Animal kingdom: Phylum,…, genus,…
- Object-oriented software classes

Trees

Two main representation schemes

- Node-link
- Space-filling



Approaches to scale:

- Complex representation
- Navigation
- Elide (don't show) some nodes
- Show nodes at different sizes

Tasks

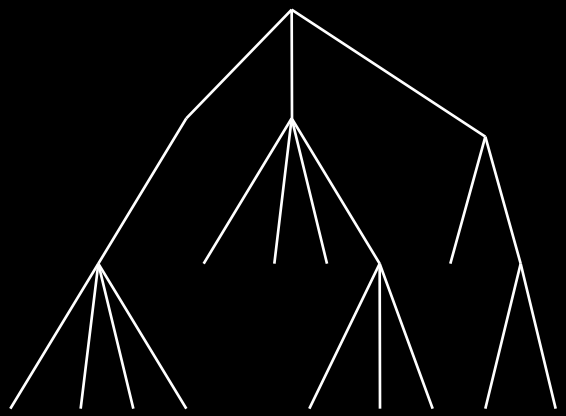
Help understand node characteristics or tree structure?

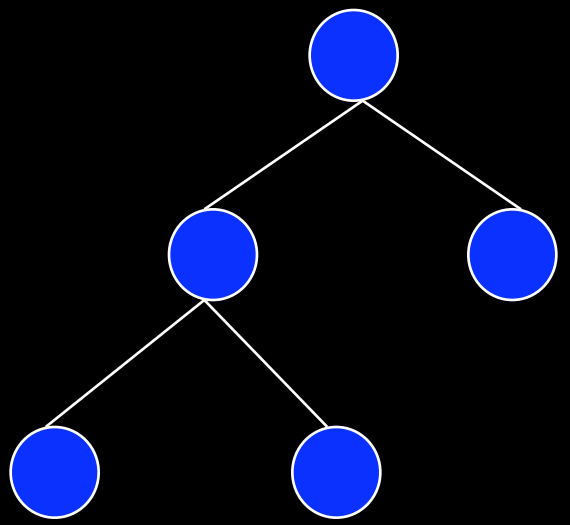
Some kinds of tasks:

- Find a node
- Revisit node
- List node ancestors
- Understand local topology
- Understand global topology

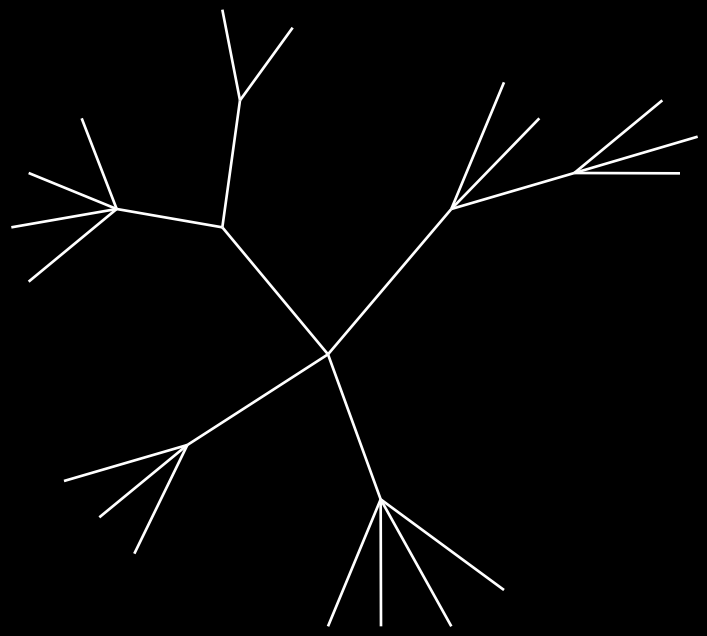
Node-Link Diagrams

Root at top, leaves at bottom is very common





Why Put Root at Top?

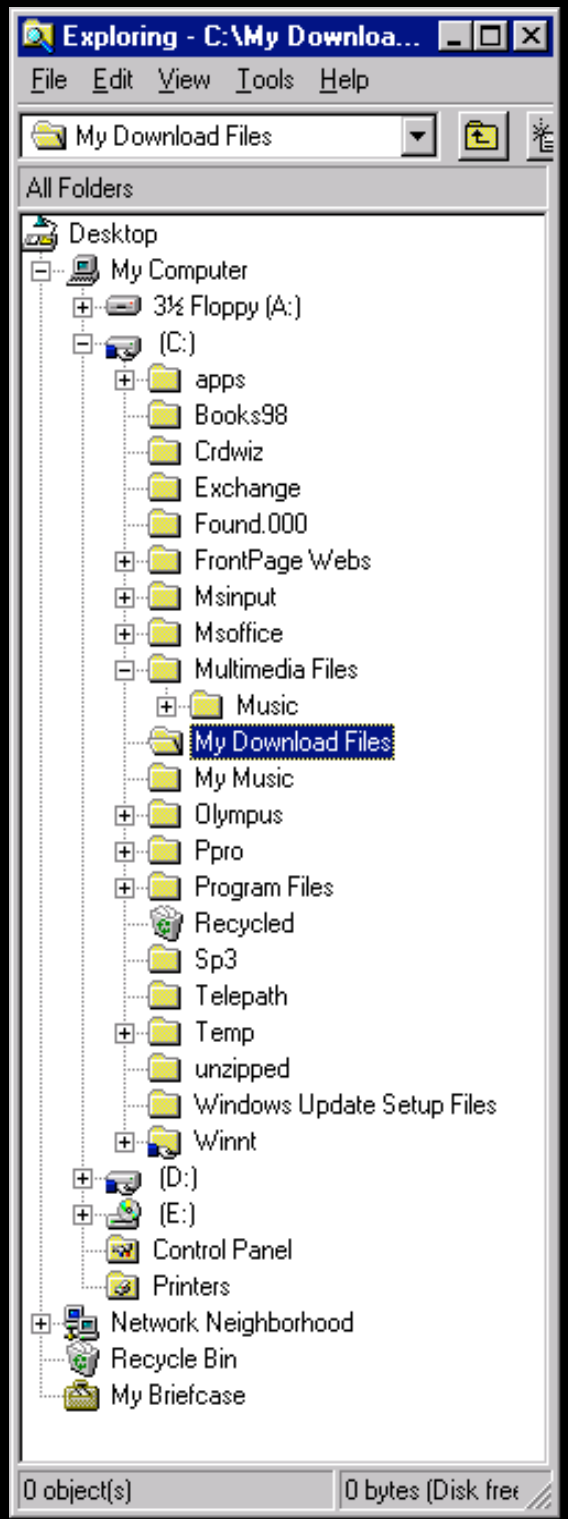


Root can be at center with levels growing outward too

Examples

Good for?

Bad for?



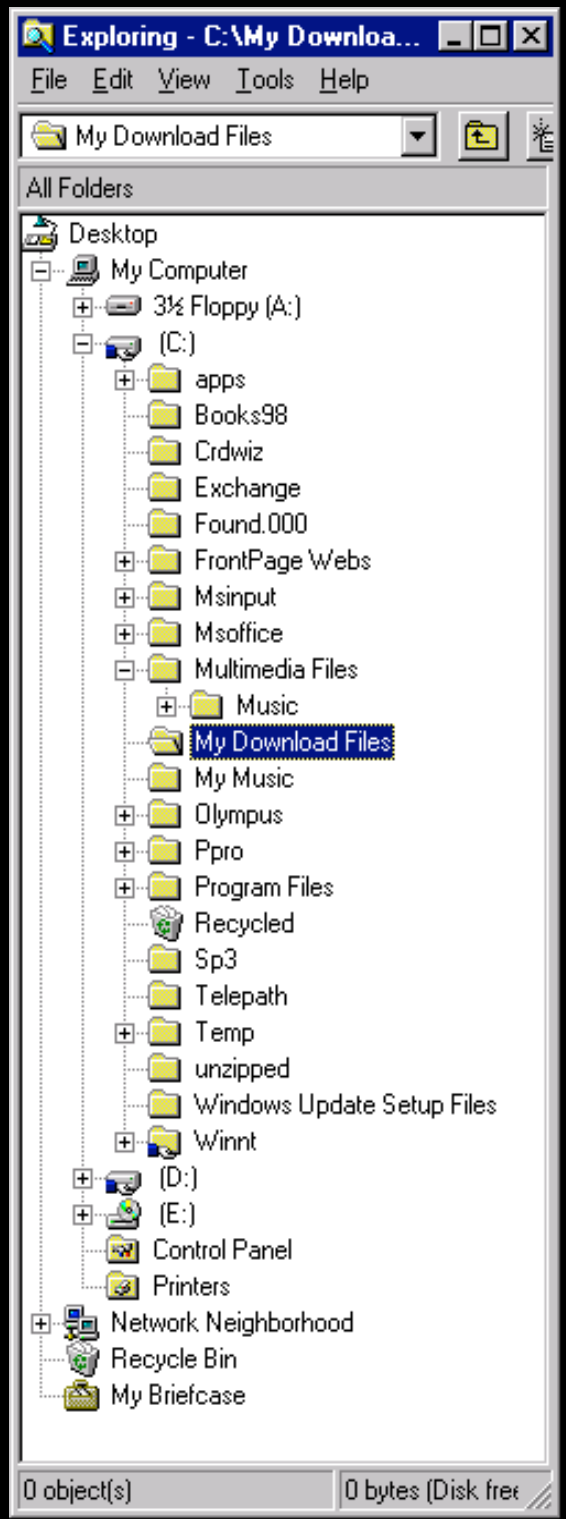
Examples

Good for?

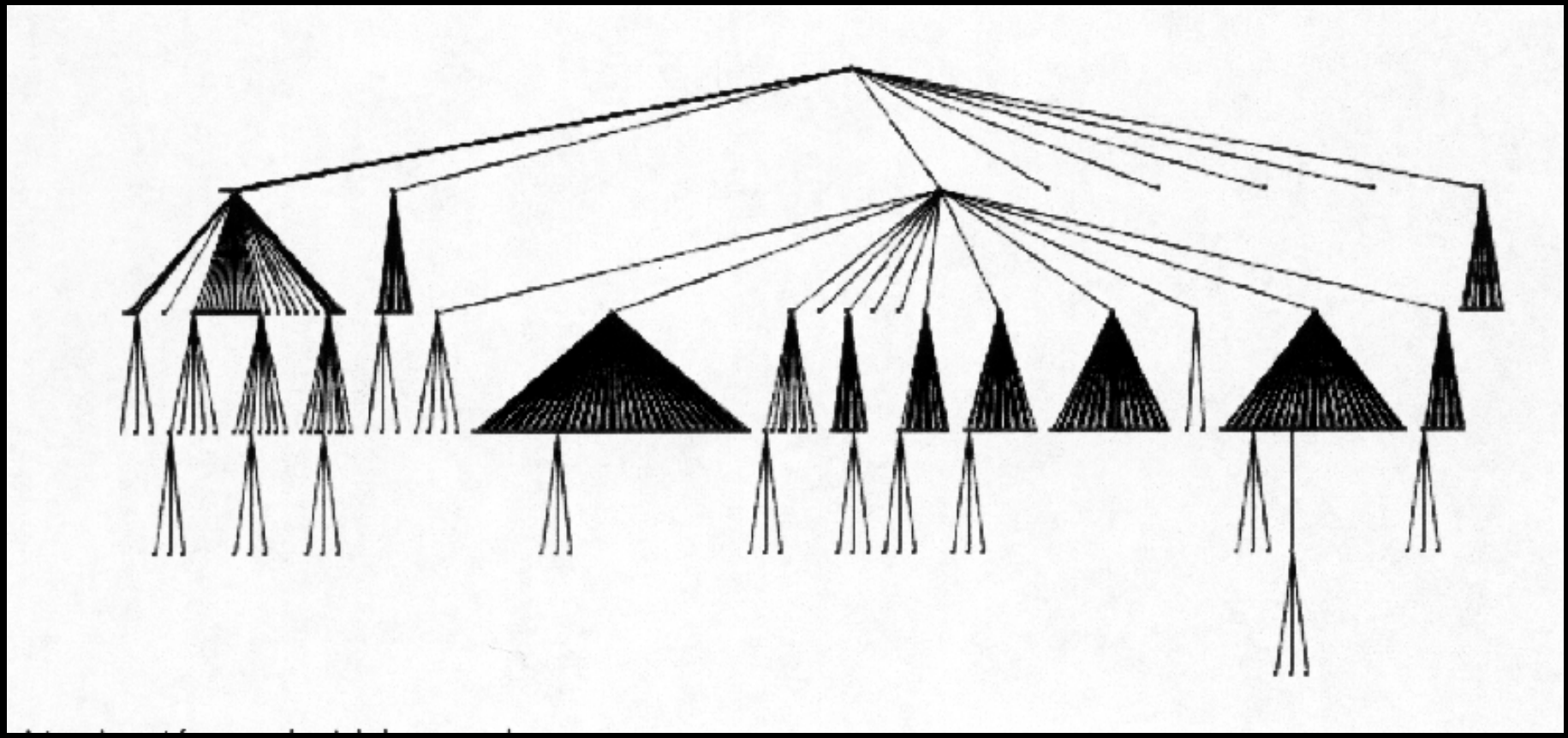
Search

Bad for?

Understanding structure



Drawing a Tree

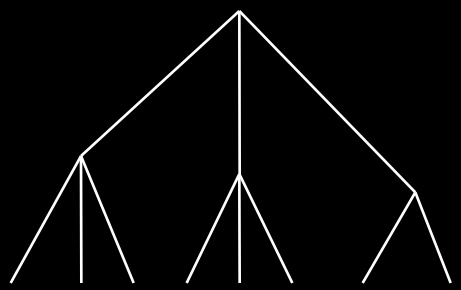


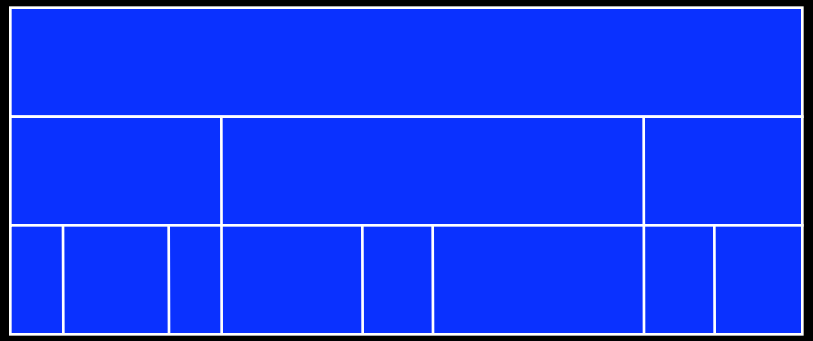
How does one draw this?

Space-Filling Representation

Each item occupies an area

Children are “contained” under parent





One example

Potential Problems

For top-down, width of fan-out uses up horizontal real estate very quickly

- At level n, there are 2^n nodes

Tree might grow a lot along one particular branch

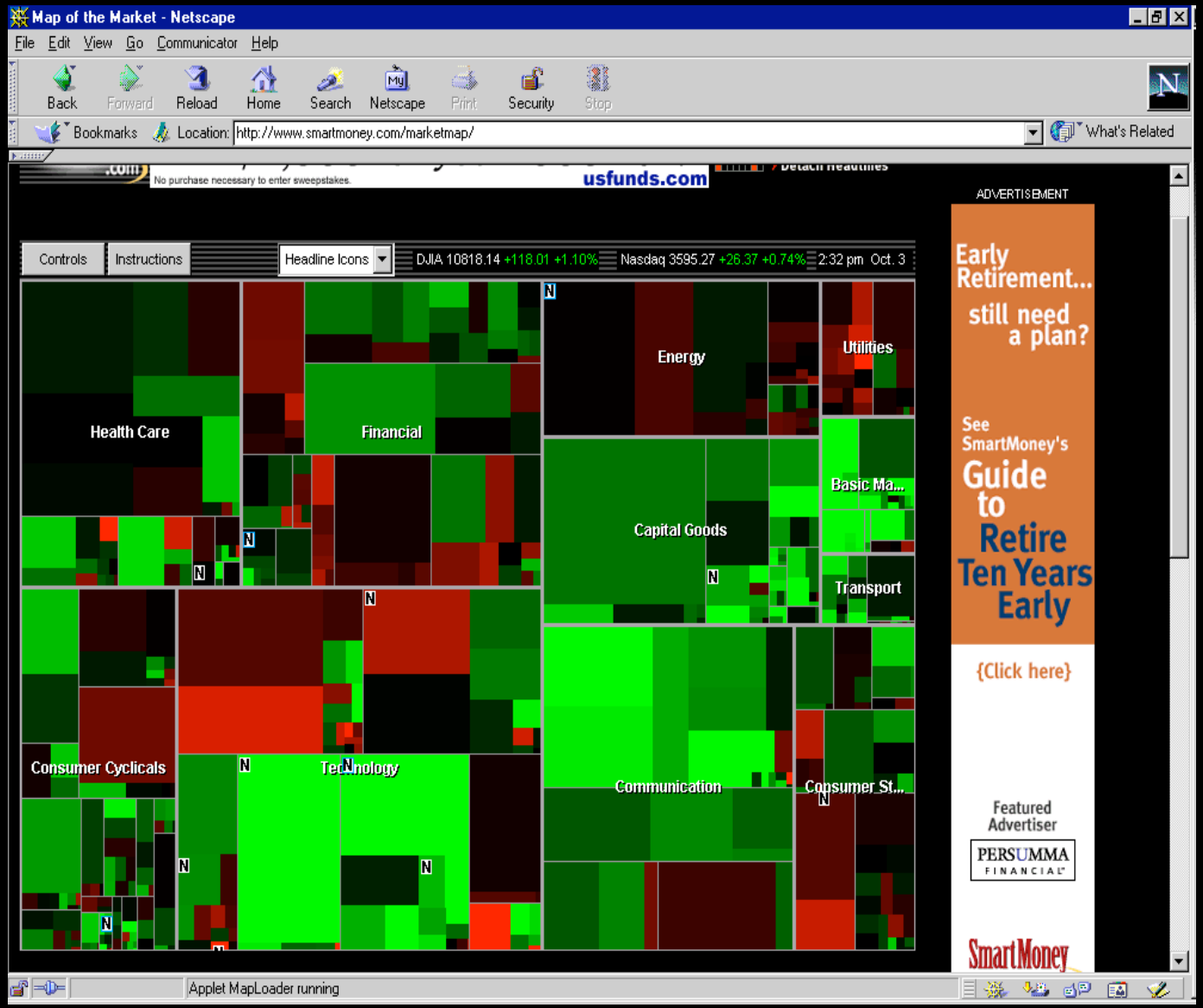
- Hard to draw it well in view without knowing how it will branch

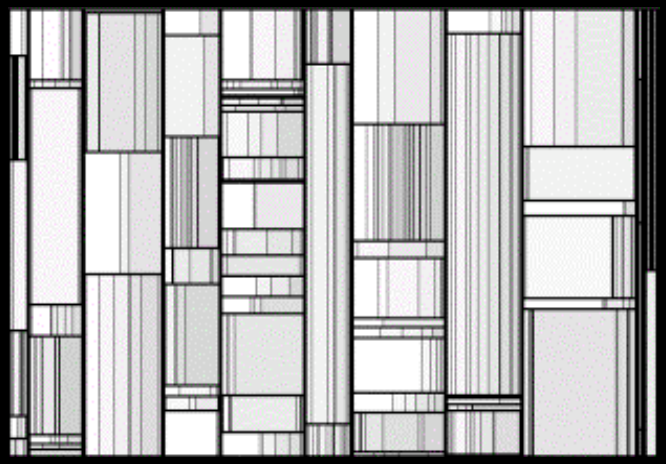
InfoVis Solutions

Techniques developed in Information Visualization largely try to assist the problems identified in the last slide

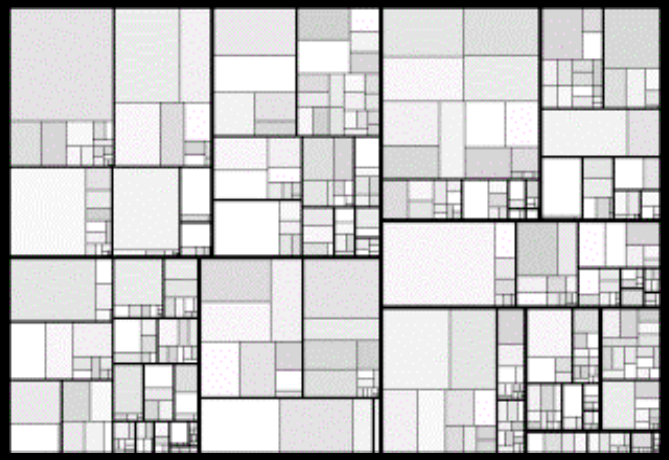
Alternatively, Information Visualization techniques attempt to show more attributes of data cases in hierarchy or focus on particular applications of trees

Treemap - Shneiderman et. al.

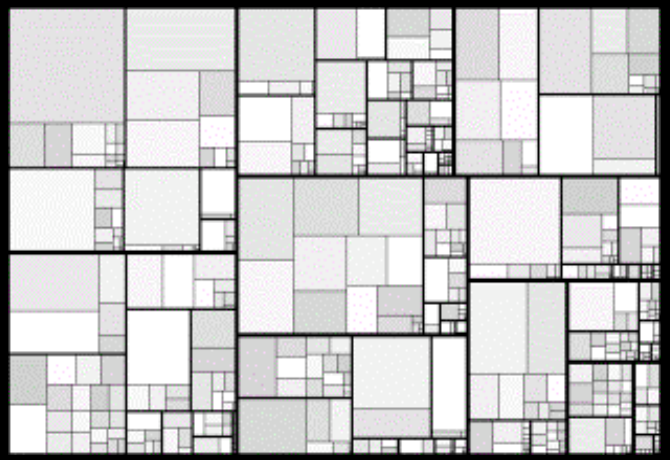




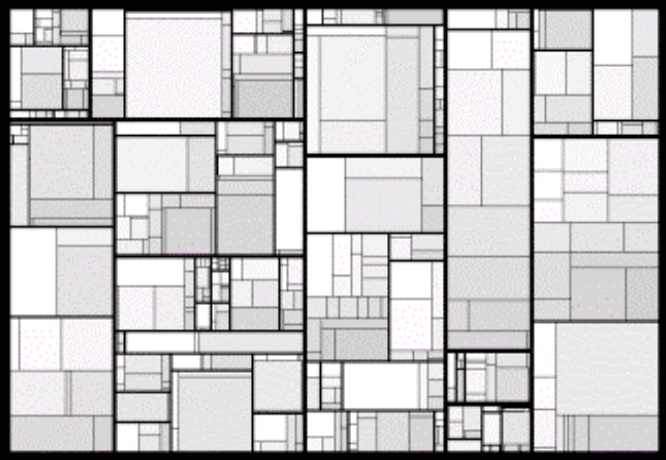
Slice-and-dice



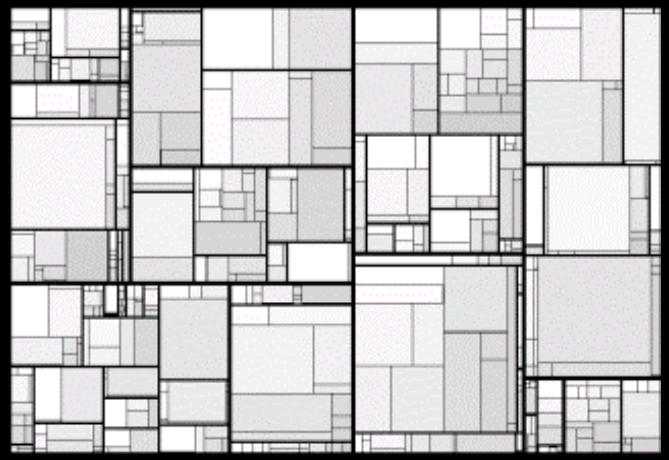
Cluster



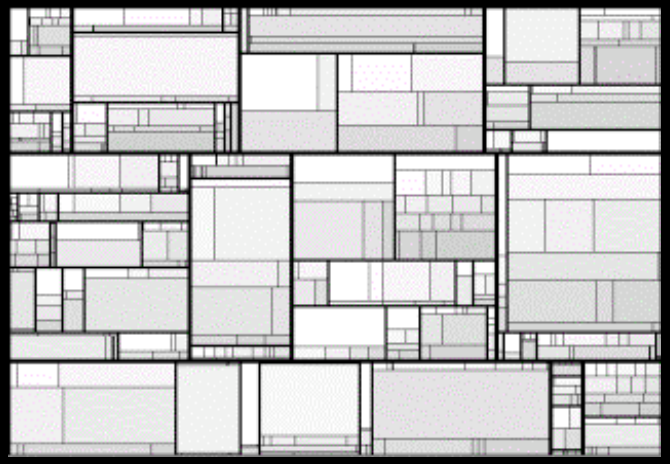
Squarified



Pivot-by-middle



Pivot-by-size

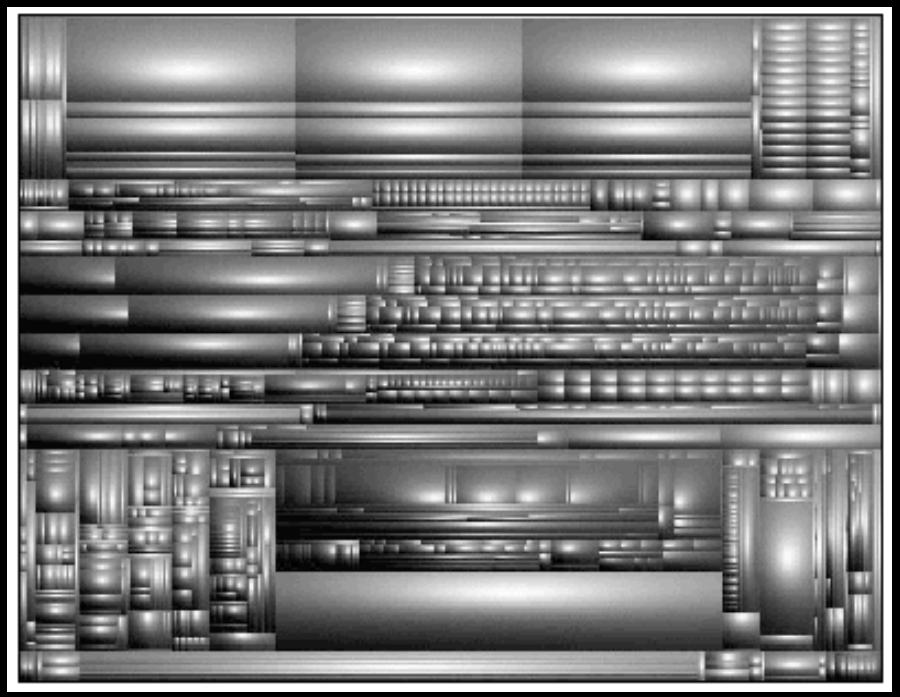


Strip

Variation: Cushion Treemap

Add shading and texture to help convey structure of hierarchy

Van Wijk '99



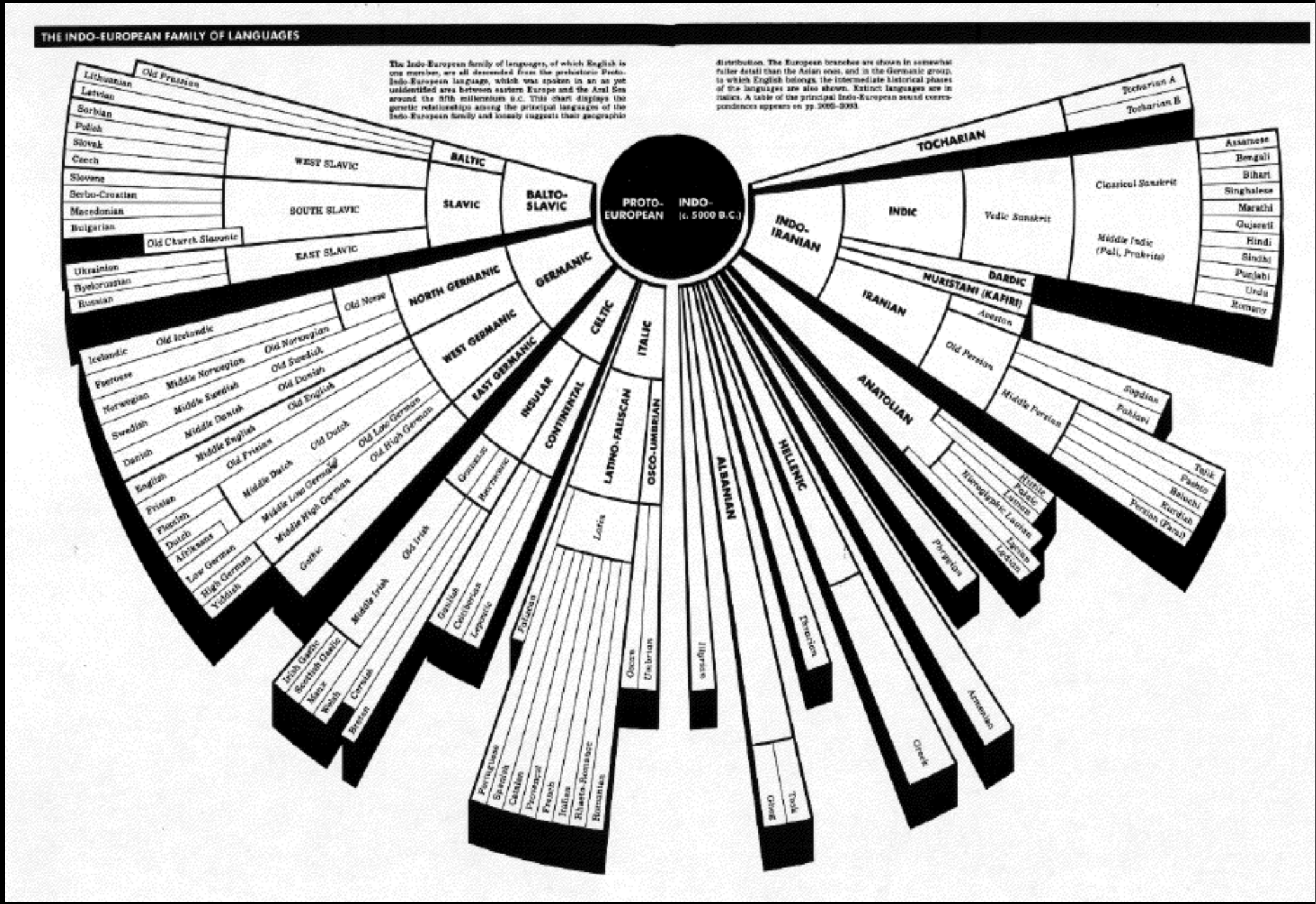
Another Technique

What if we used a radial rather than a rectangular space-filling technique?

- We saw node-link trees with root in center and growing outward already...

Make pie-tree with root in center and children growing outward

- Radial angle now corresponds to a variables rather than area



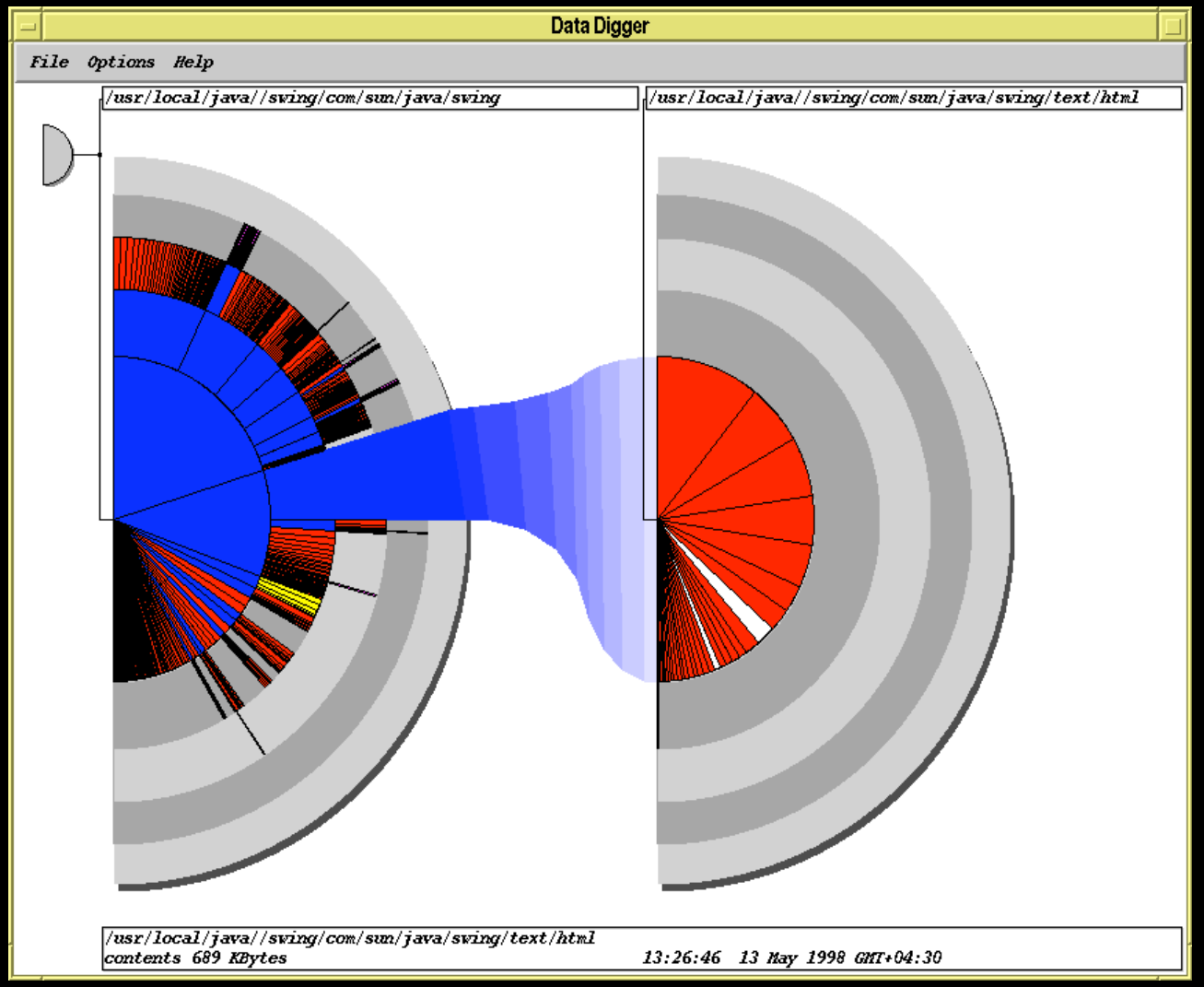
Appears in:
American Heritage Dictionary, 3rd Ed. Houghton Mifflin, 1992

Radial Space-Filling

Chuah

Andrews & Heidegger →

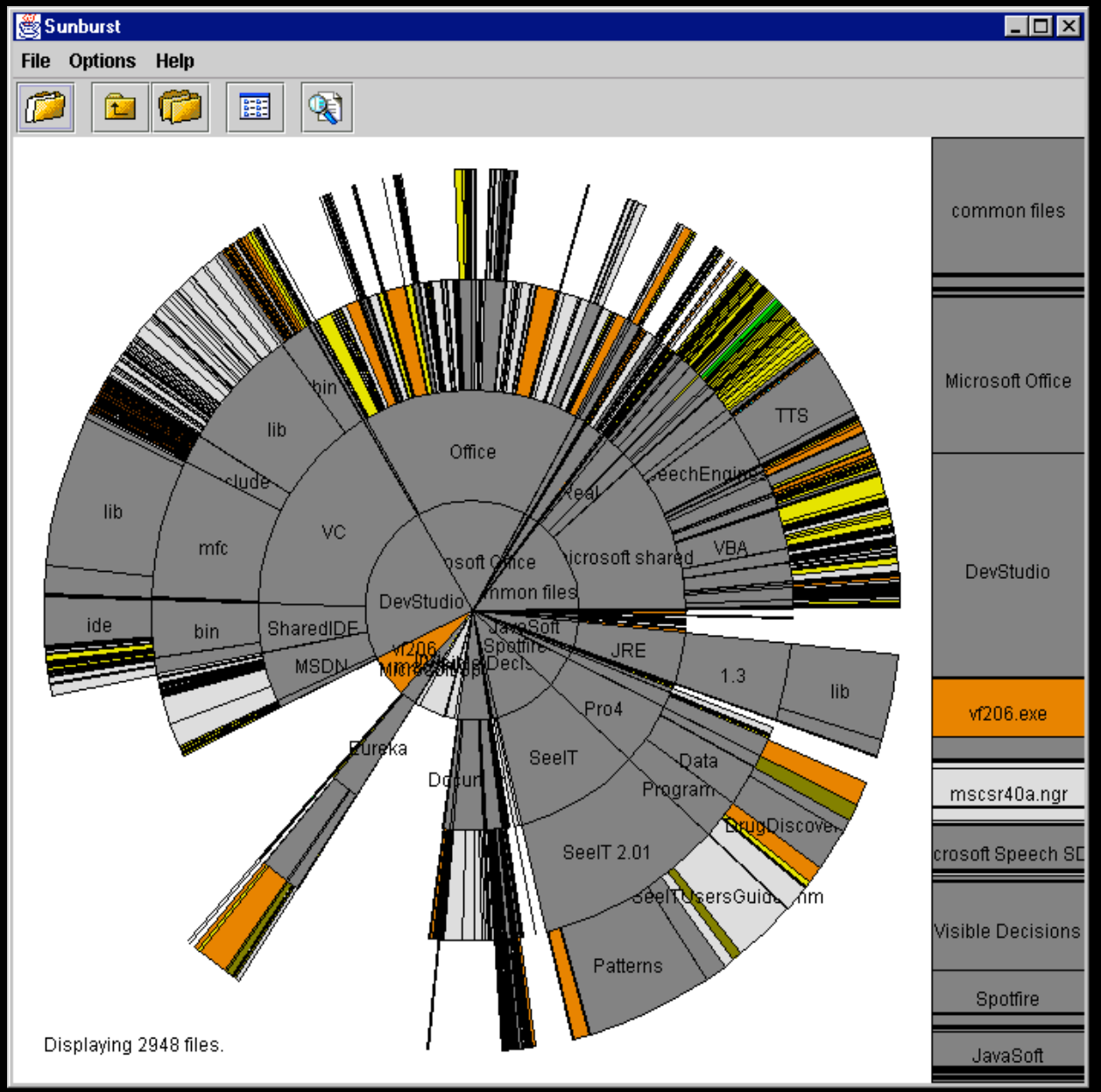
InfoVis '98



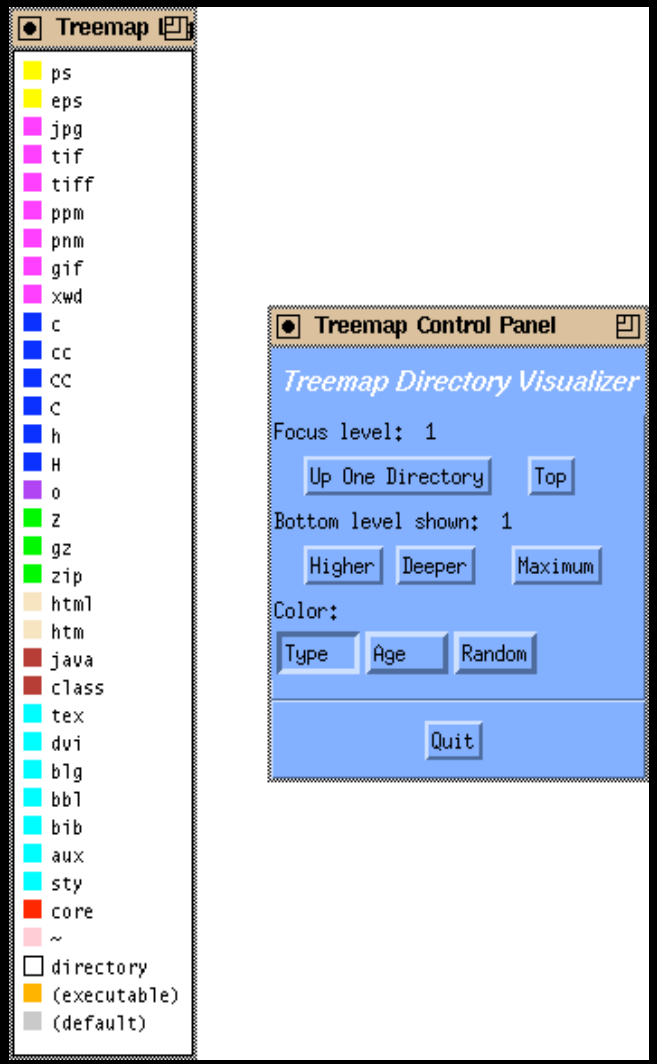
SunBurst

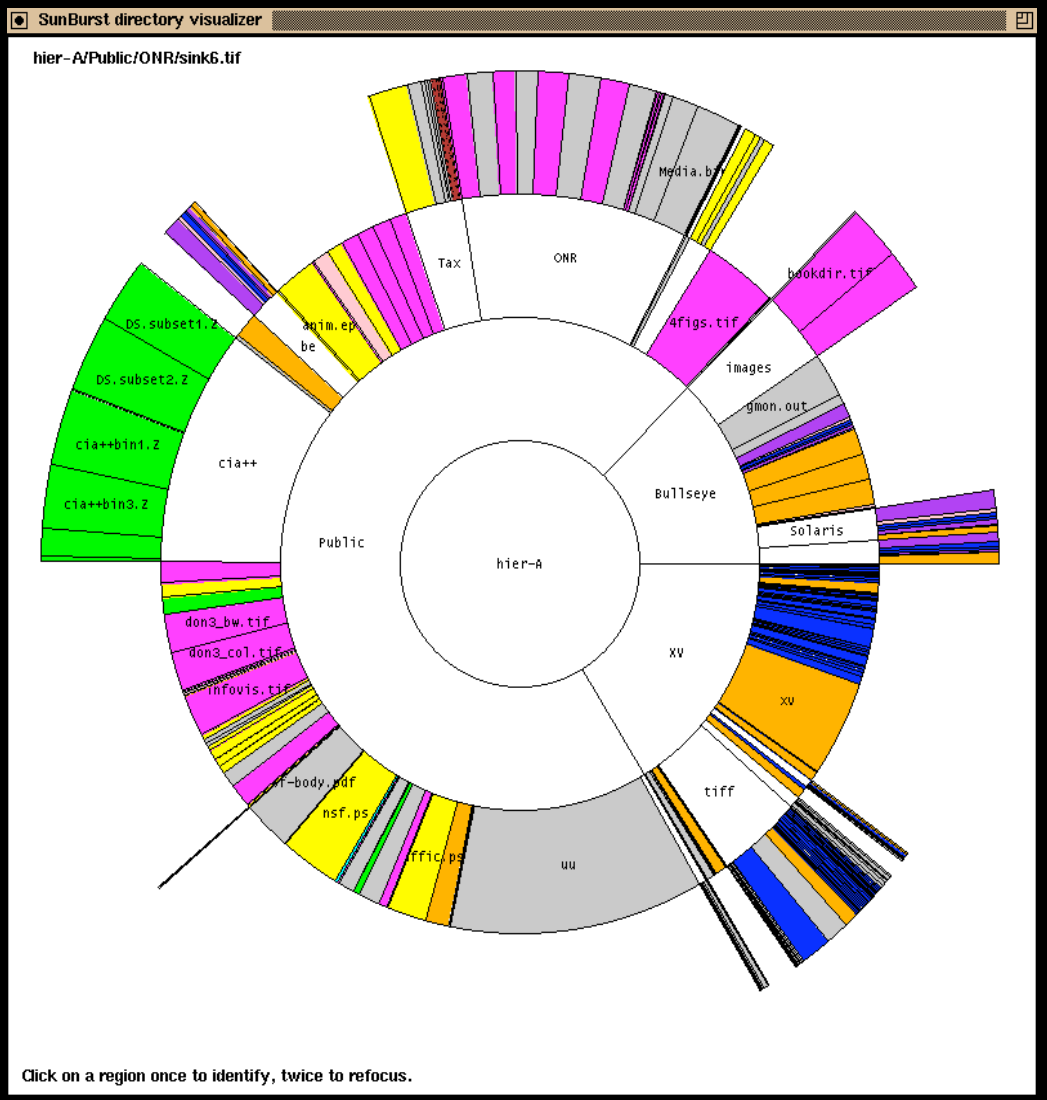
www.cc.gatech.edu/gvu/ii/sunburst

File browser



SunBurst


Treemap
ps
eps
jpg
tif
tiff
ppm
pnm
gif
xwd
c
cc
CC
C
h
H
o
z
gz
zip
html
htm
java
class
tex
dvi
blg
bbl
bib
aux
sty
core
~
directory
(executable)
(default)
Treemap Control Panel
Treemap Directory Visualizer
Focus level: 1
Up One Directory
Top
Bottom level shown: 1
Higher
Deeper
Maximum
Color:
Type
Age
Random
Quit


SunBurst directory visualizer
hier-A/Public/ONR/sink6.tif
Media.bak
Tax
ONR
bookdir.tif
4figs.tif
images
gmon.out
Bullseye
Solaris
XV
tiff
uu
nsf.ps
tf-body.pdf
don3_bw.tif
don3_col.tif
infovis.tif
Public
hier-A
cia++
DS.subset1.Z
DS.subset2.Z
cia++bin1.Z
cia++bin3.Z
Click on a region once to identify, twice to refocus.

SunBurst

Root directory at center, each successive level drawn farther out from center

Sweep angle of item corresponds to size

Color maps to file type or age

Interactive controls for moving deeper in hierarchy, changing the root, etc.

Double-click on directory makes it new root

Empirical Study

Stasko, Catrambone, Guzdial & McDonald
International Journal of Human-Computer Studies, 2000

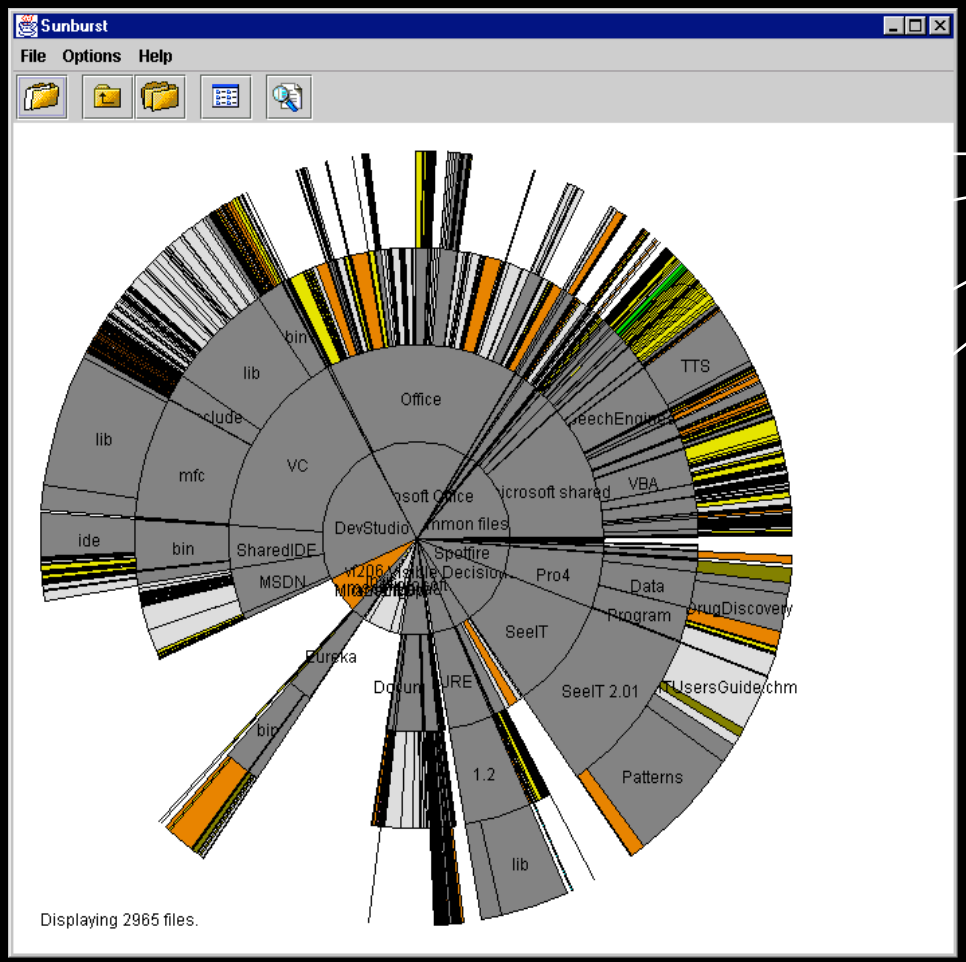
Compared SunBurst to Treemap (borderless) on a variety of file browsing tasks

- SunBurst performed as well (or better) in task accuracy and time
- Learning effect - Performance improved with Treemap on second session
- Strong subjective preference (51-9) for SunBurst
- Participants cited more explicit depiction of structure as an important reason

More to come on evaluation...

SunBurst Negative

In large hierarchies, files at the periphery are usually tiny and very difficult to distinguish



examples

Fix: Objectives

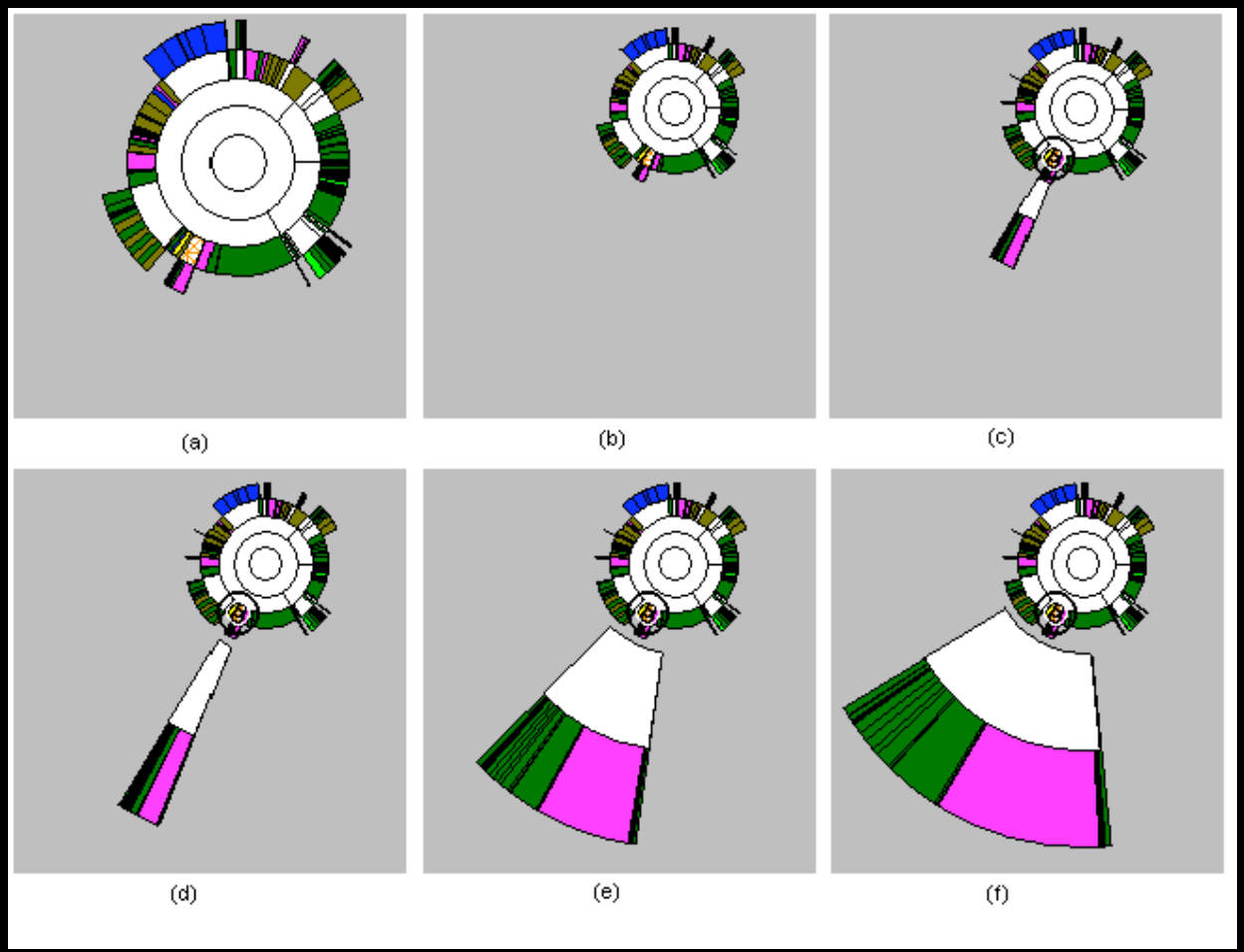
- Make small slices bigger
- Maintain full circular space-filling idea
- Allow detailed examination of small files within context of entire hierarchy
- Don't alter ratios of sizes
- Avoid use of multiple windows or lots of scrollbars
- Provide an aesthetically pleasing interface in which it is easy to track changes in focus

3 Attempts

Three visualization+navigation techniques developed to help remedy the shortcoming

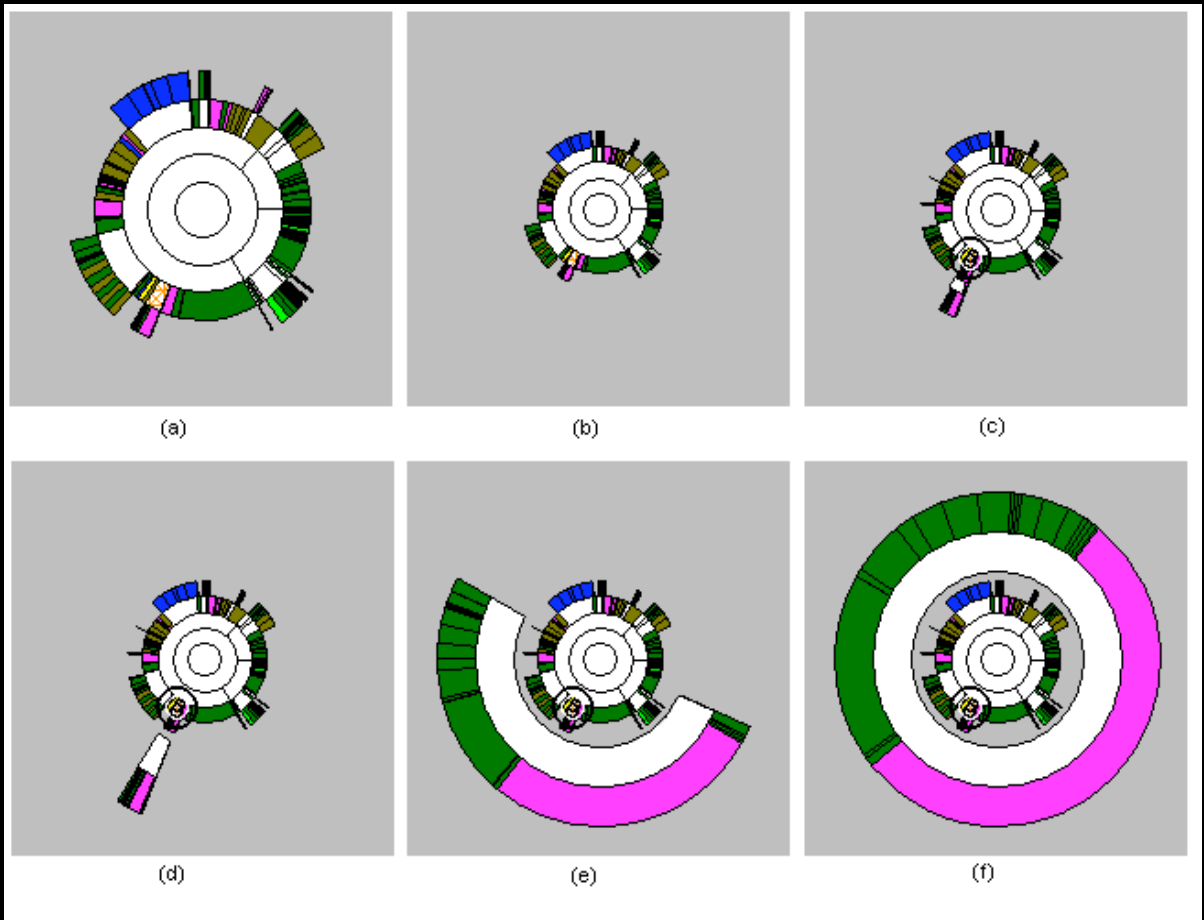
- Angular detail
- Detail outside
- Detail inside

Angular Detail



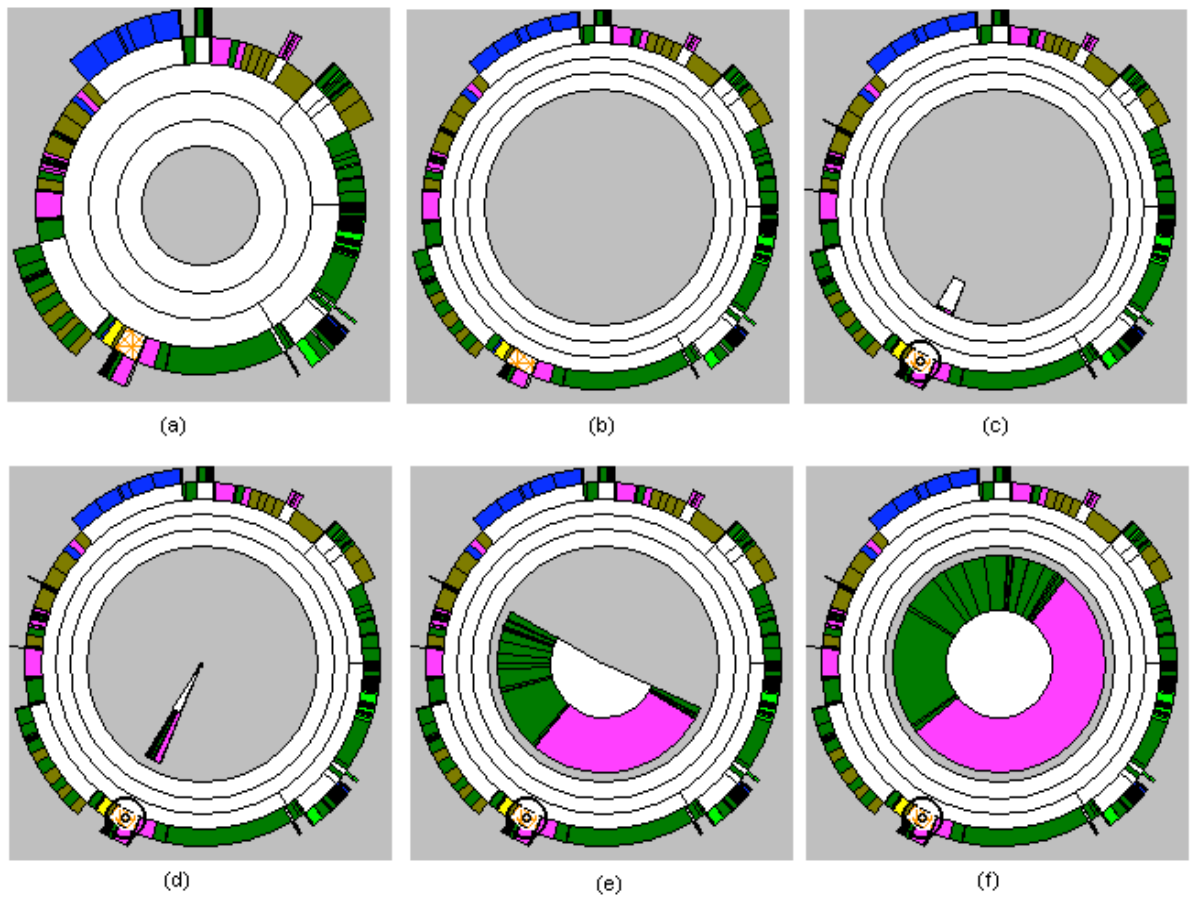
- Most "natural"
- Least space-efficient
- Most configurable by user

Detail Outside



- Exhibits non-distorted miniature of overview
- Somewhat visually disconcerting
- Focus is quite enlarged (large circumference and 360°)
- Relatively space efficient

Detail Inside



- Perhaps least intuitive and most distorting
- Items in overview are more distinct (larger circumference)
- Interior 360° for focus is often sufficient

3D Approaches

Add a third dimension into which layout can go

Compromise of top-down and centered techniques mentioned earlier

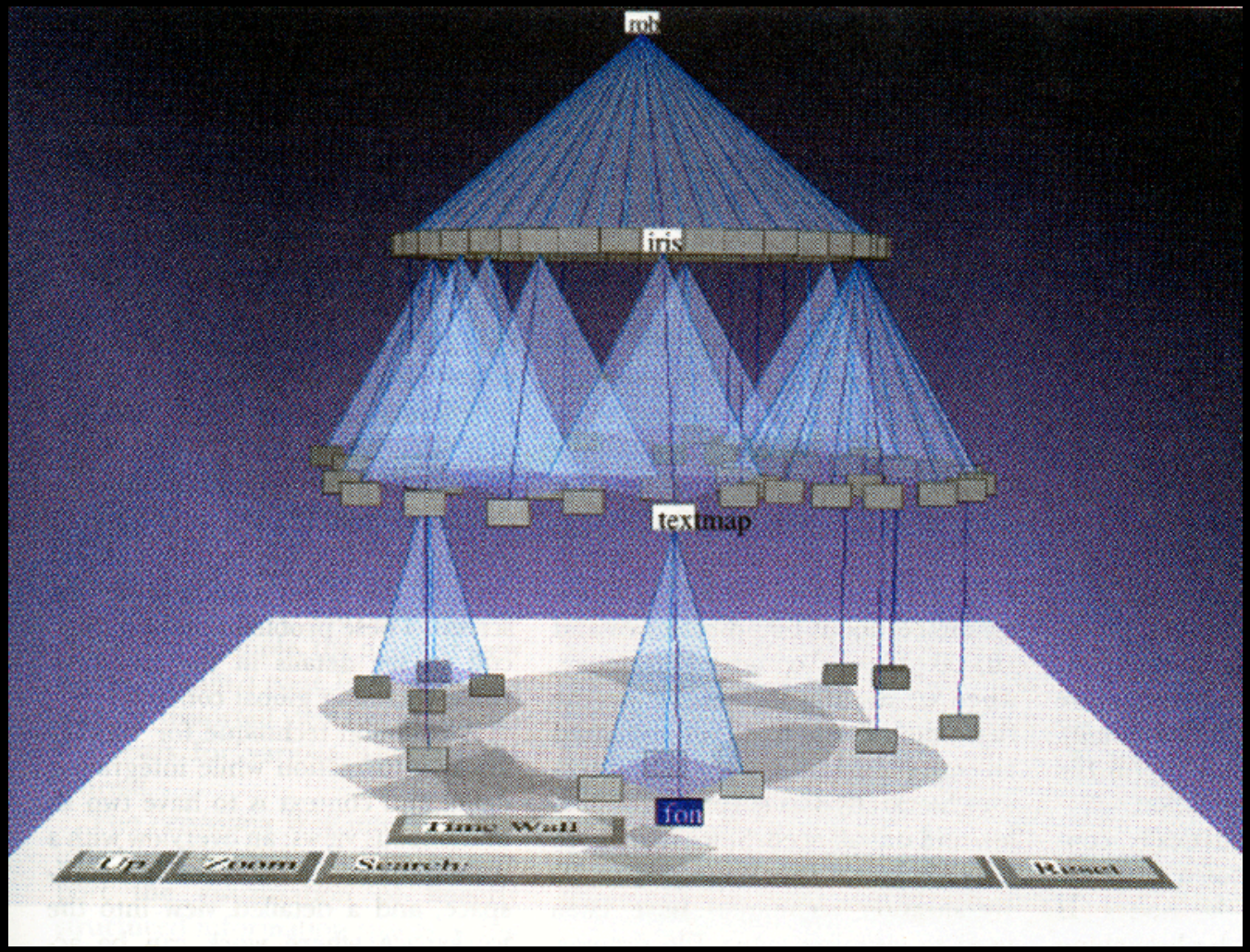
Children of a node are laid out in a cylinder "below" the parent

- Siblings live in one of the 2D planes

Cone Trees

Developed at Xerox PARC

3D views of hierarchies such as file systems



Robertson, Mackinlay, Card '91

Alternate Views

Cone Trees

Positive

- More effective area to lay out tree
- Use of smooth animation to help person track updates
- Aesthetically pleasing (NOT)

Negative

- As in all 3D, occlusion obscures some nodes
- Non-trivial to implement and requires some graphics horsepower

Alternative Solutions

Change the geometry

Apply a hyperbolic transformation to the space

Root is at center, subordinates around

Apply idea recursively, distance decreases between parent and child as you move farther from center, children go in wedge rather than circle

Hyperbolic Browser

Focus + Context Technique

- Detailed view blended with a global view

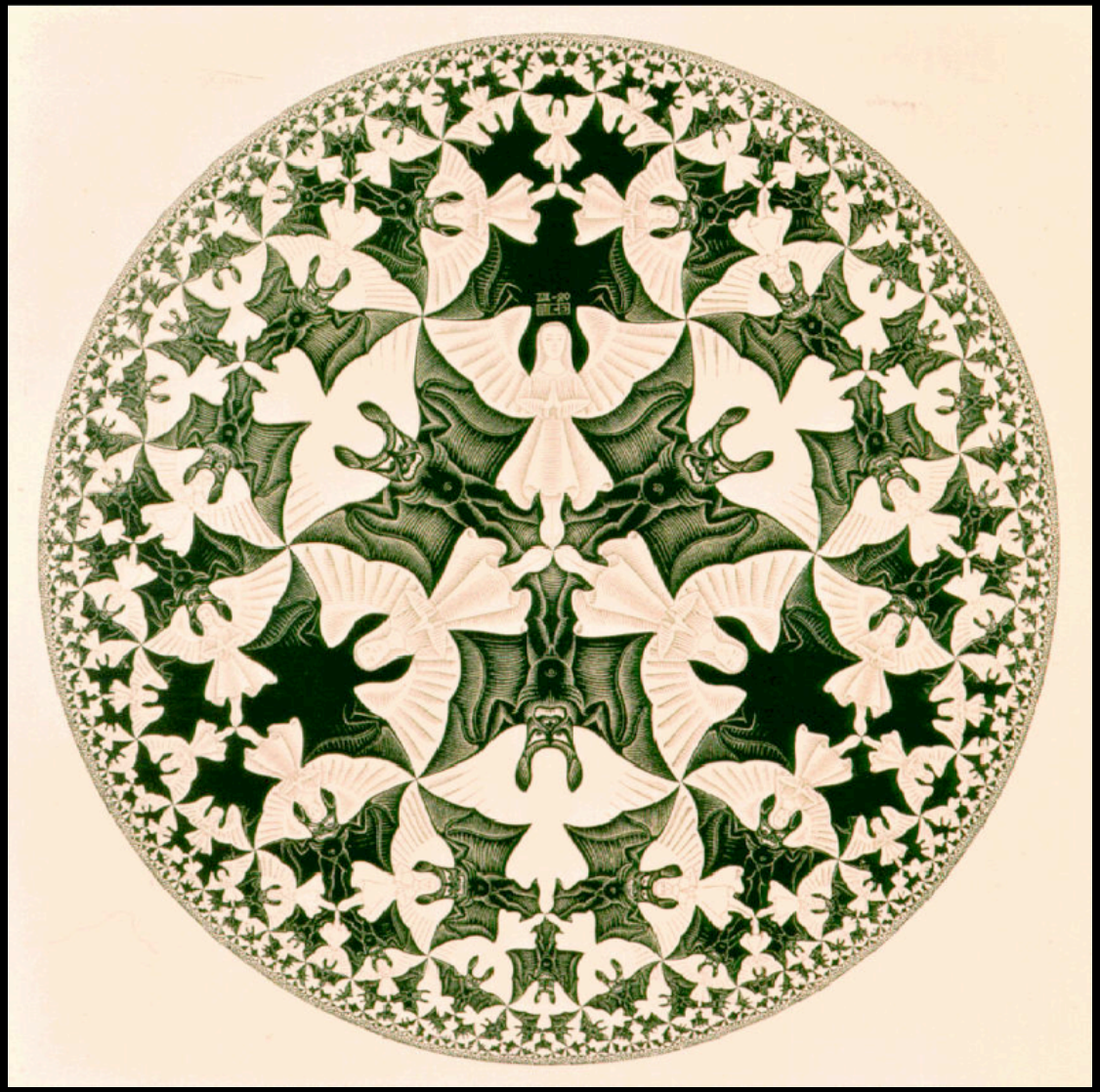
First lay out the hierarchy on the hyperbolic plane

Then map this plane to a disk

Start with the tree's root at the center

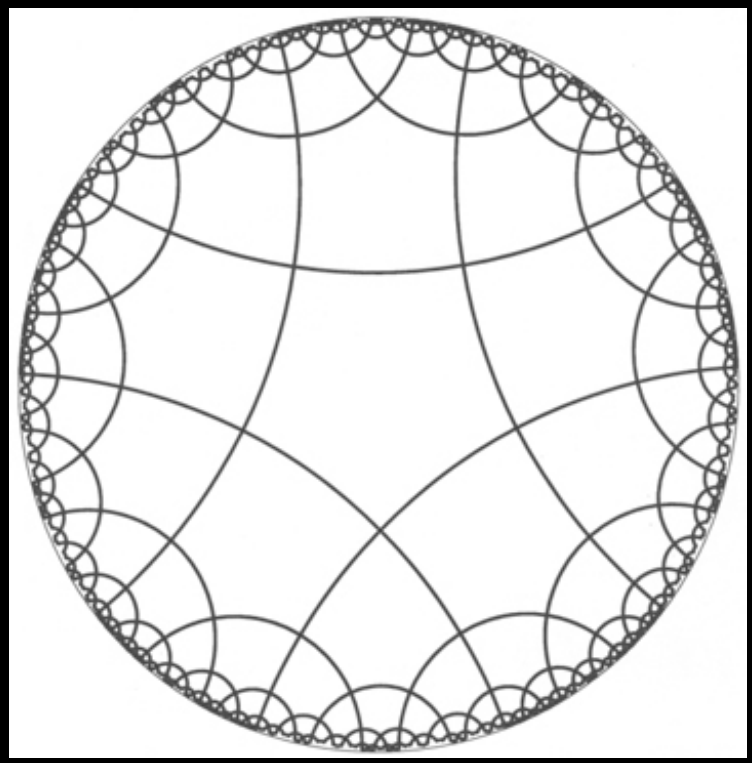
Use animation to navigate along this representation of the plane Lamping and Rao, '94

Hyperbolic Space



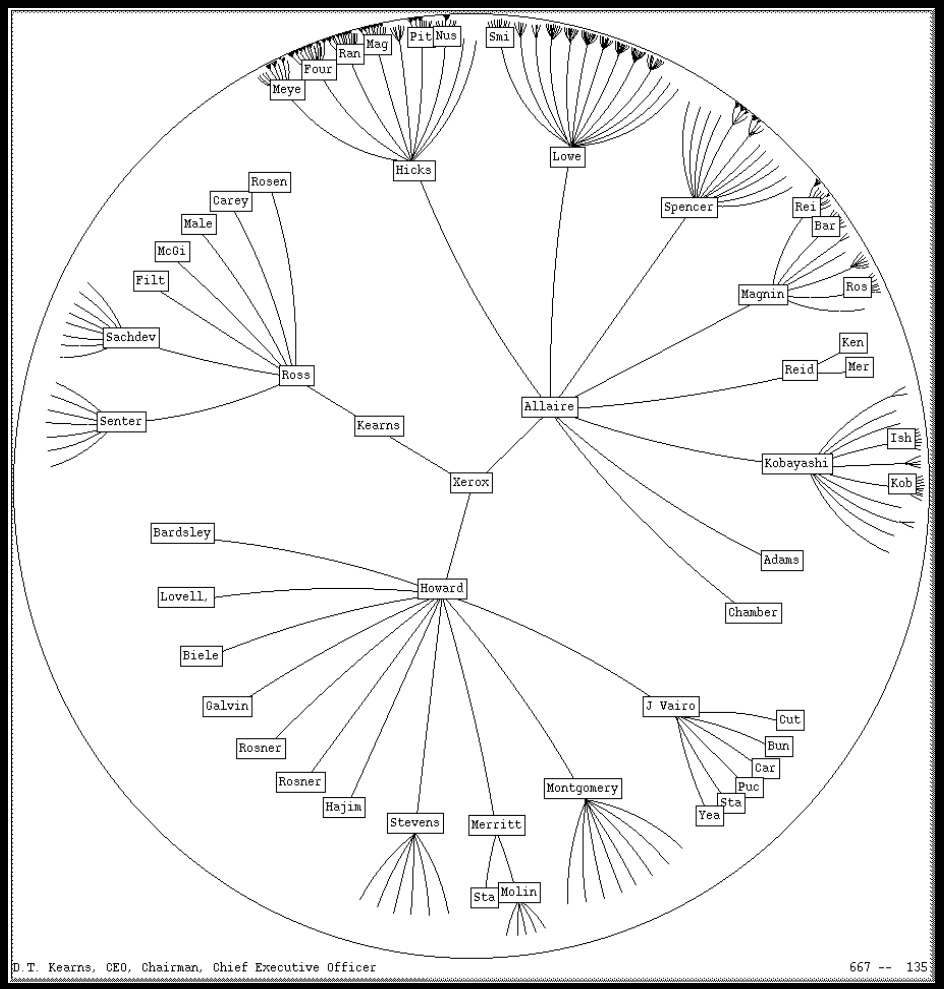
M. C. Escher, Circle Limit IV

Tiling space with pentagons!



All (an infinite number of them) pentagons have
“straight lines”,
“same area”

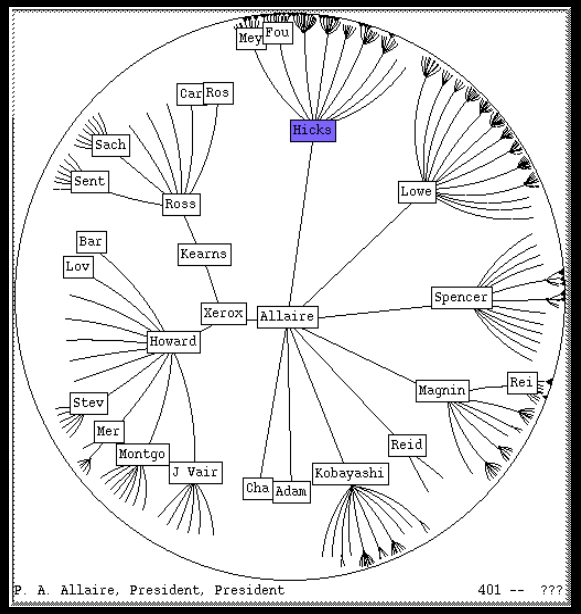
2D Hyperbolic Browser



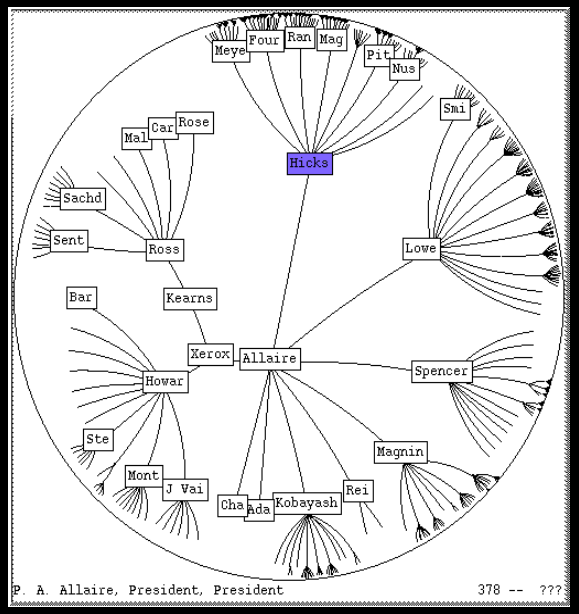
Approach: Lay out the hierarchy on the hyperbolic plane and map this plane onto a display region.

Comparison

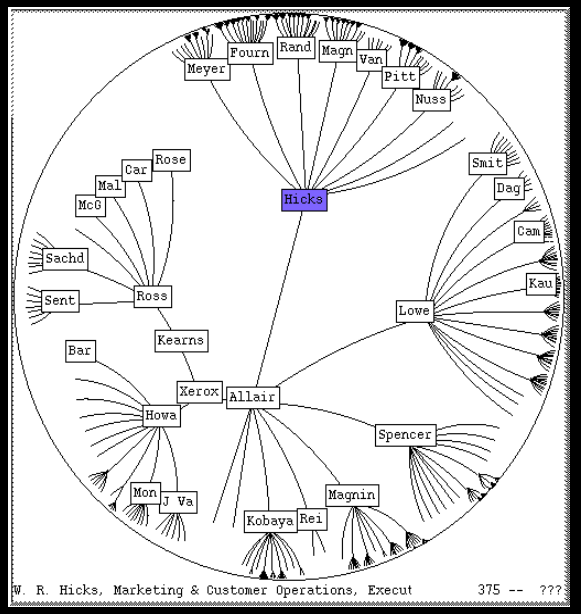
- A standard 2D browser: 100 nodes (w/3 character text strings)
- Hyperbolic browser: 1000 nodes, about 50 nearest the focus can show from 3 to dozens of characters



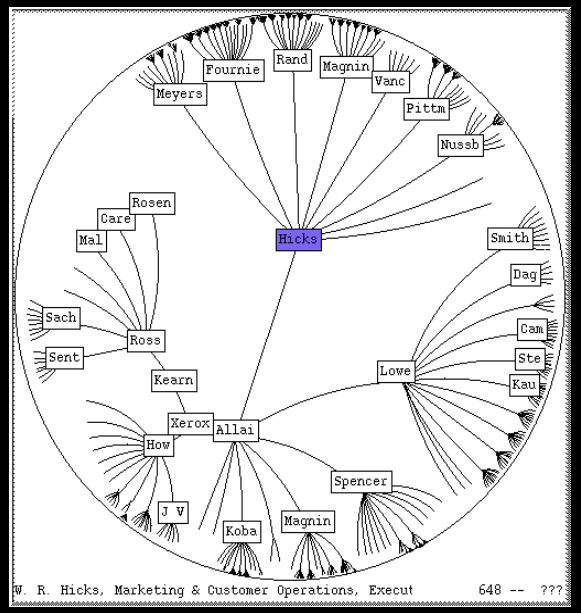
1



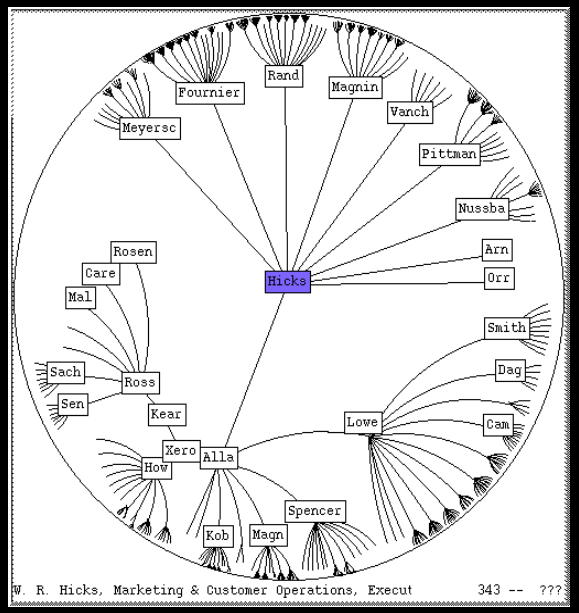
2



3



4



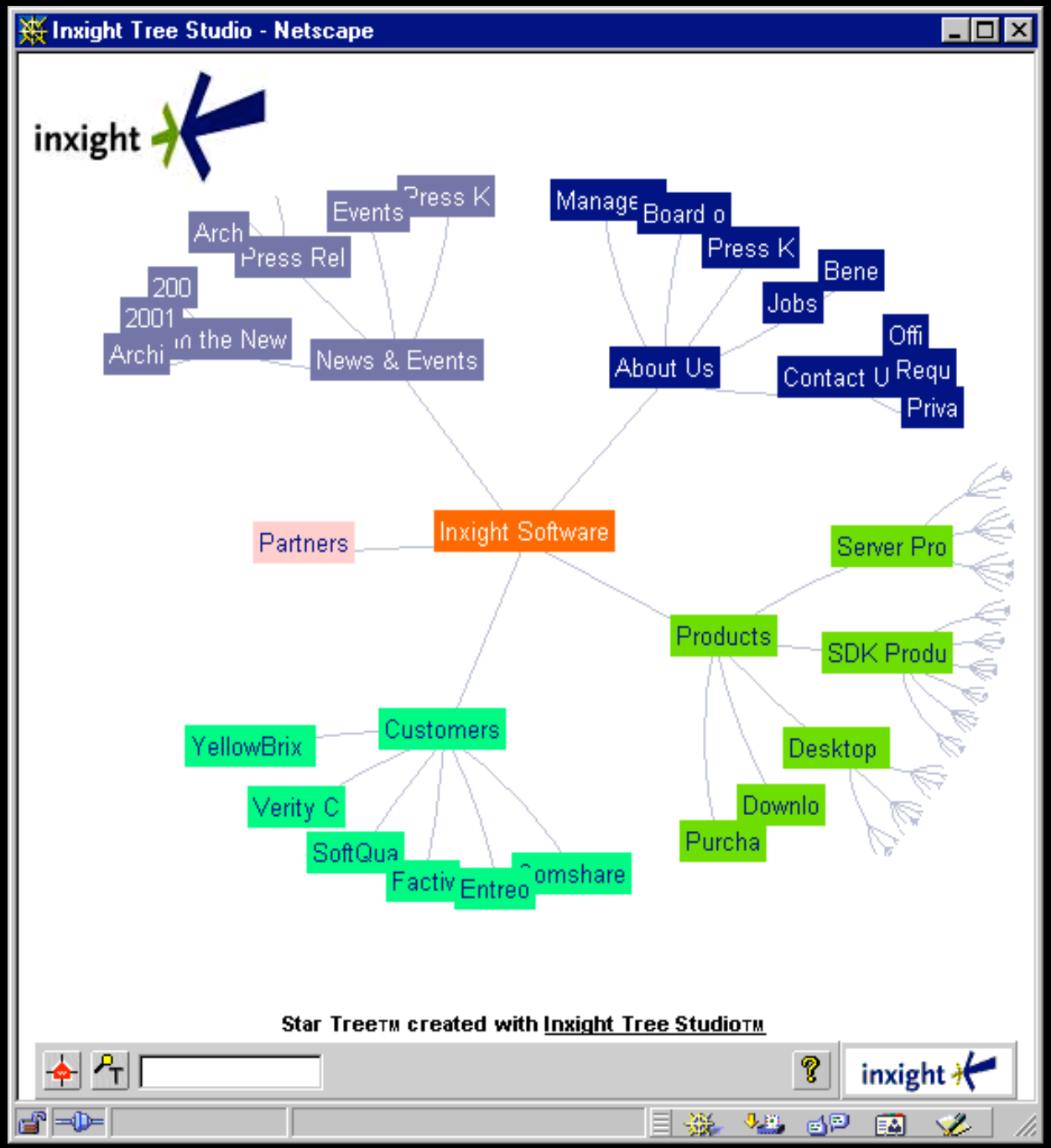
5

Clicking on the blue node brings it into focus at the center

StarTree

Hyperbolic tree

www.inxight.com



Key Attributes

Natural magnification (fisheye) in center

Layout depends only on 2-3 generations from current node

Smooth animation for change in focus

Don't draw objects when far enough from root (simplify rendering)

Problems

Orientation

- Watching the view can be disorienting
- When a node is moved, its children don't keep their relative orientation to it as in Euclidean plane, they rotate
- Not as symmetric and regular as Euclidean techniques, two important attributes in aesthetics

How about 3D?

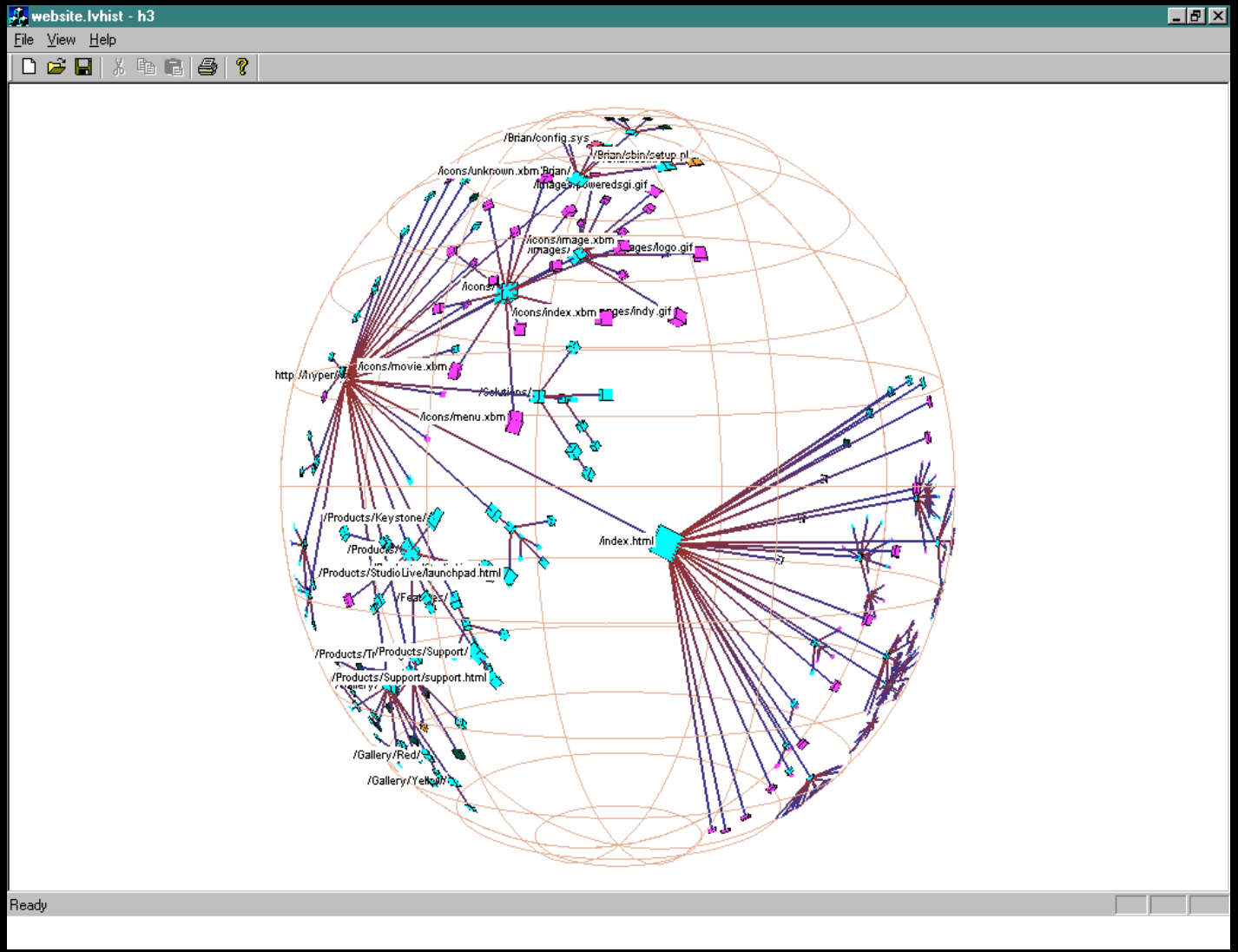
Can same hyperbolic transformation be applied, but now use 3D space?

Sure can

Have fun with the math!

(not that bad)

H3Viewer



Munzner, '98

Layout

Find a spanning tree from an input graph

- Use domain-specific knowledge

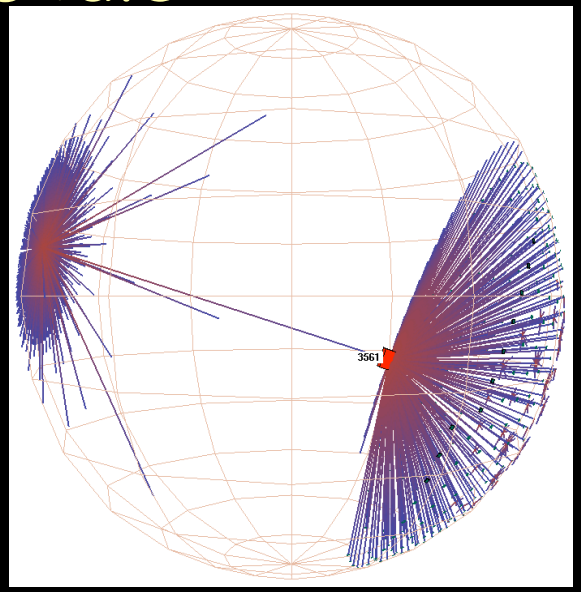
Layout algorithm

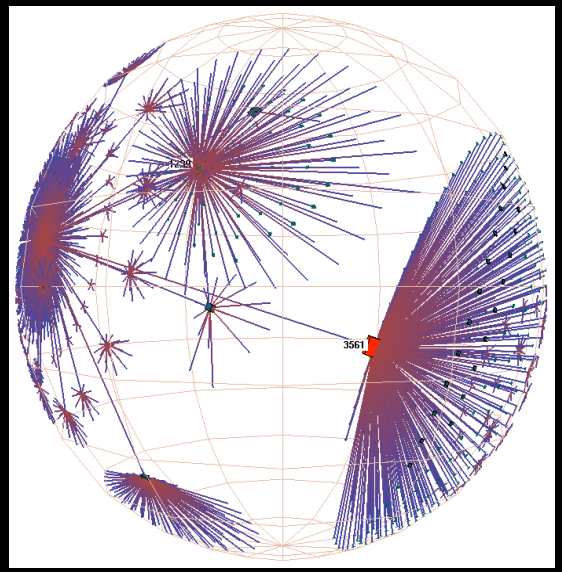
- Nodes are laid out on the surface of a hemisphere
- A bottom-up pass to estimate the radius needed for each hemisphere
- A top-down pass to place each child node on its parental hemisphere's surface

Drawing

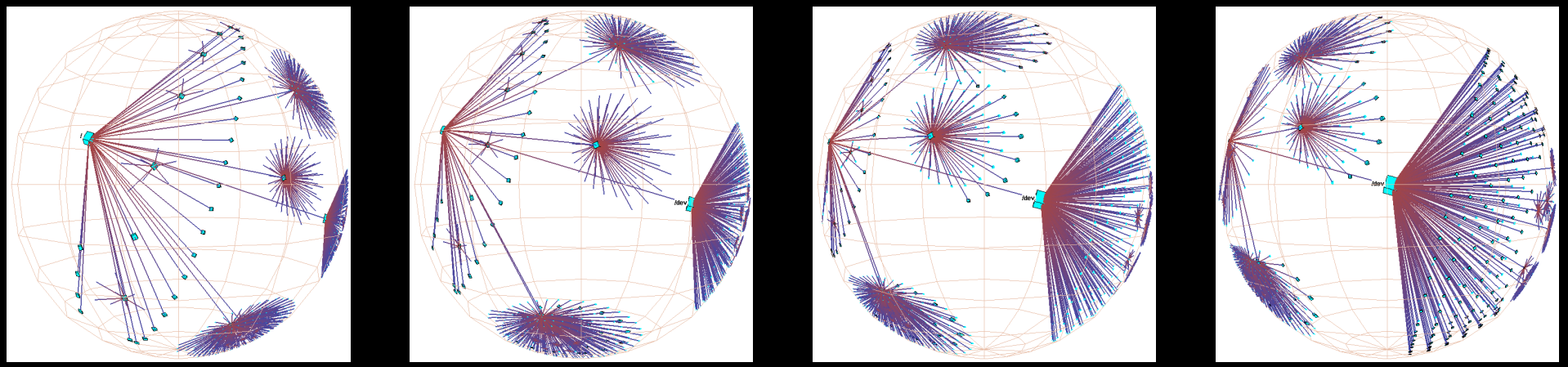
Maintain a target frame by showing less of the context surrounding the node of interest during interactive browsing

Fill in more of the surrounding scene when the user is idle

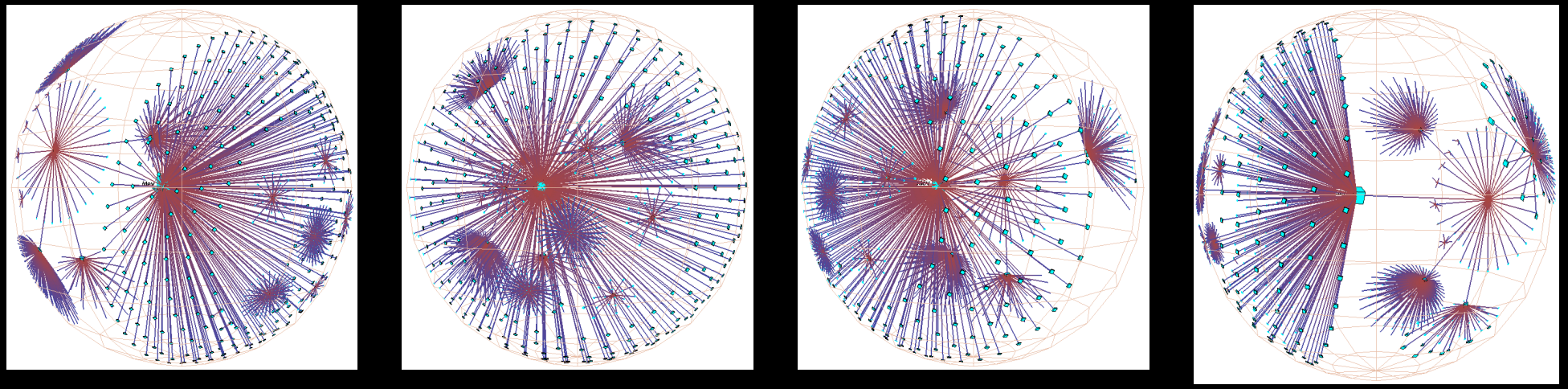




Navigation



Translation of a node to the center



Rotation around the same node

Performance

- Handle much larger graphs, i.e. >100,000 edges
- Support dynamic exploration & interactive browsing
- Maintain a guaranteed frame rate

http://graphics.stanford.edu/~munzner/

Focus + Context

Scale - Many data sets are too large to visualize on one screen

- May simply be too many cases
- May be too many variables
- May only be able to highlight particular cases or particular variables, but viewer's focus may change from time to time

Challenge

Have context/overview seamlessly and smoothly co-exist with focus/detail

Why?

- Easier to move between the two, helps assimilate view updates, less jarring, …

Not all overview and detail techniques are good at this

Focus + Context Views

Same idea as overview and detail, with one key difference:

- Typically, the overview and the detail are combined into a single display
- Mimics our natural vision systems more closely

How?

What techniques have we seen so far that would help accomplish focus+context?

Possible Methods

Filtering

Selective aggregation

Micro-macro readings

Highlighting

Distortion

Prototypical Example

When people think about focus+context views, they typically think of the Fisheye View (distortion)

Introduced by George Furnas in 1981 report, more famous article is 1986 SIGCHI paper

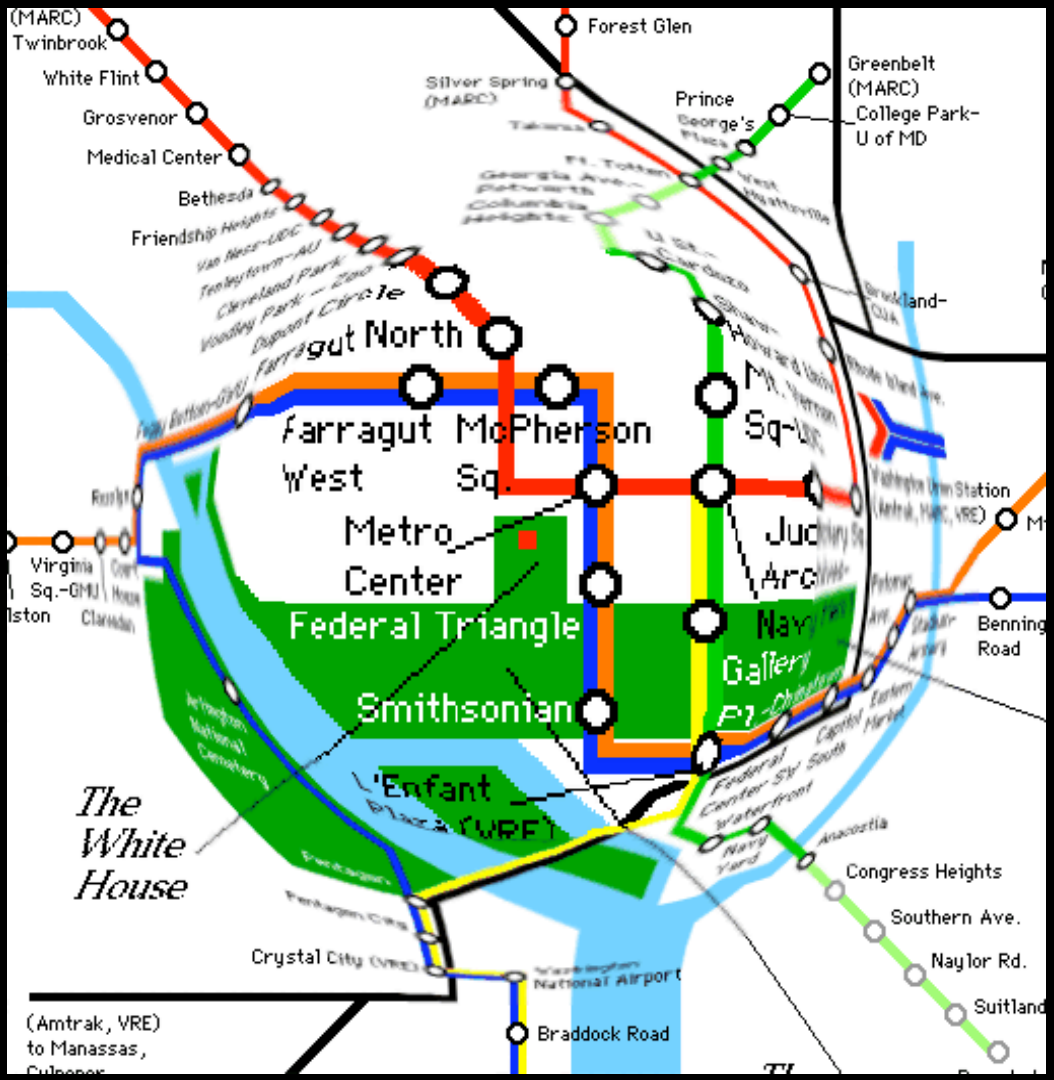
Definition

Fisheye View -

"Provide[s] detailed views (focus) and overviews (context) without obscuring anything…The focus area (or areas) is magnified to show detail, while preserving the context, all in a single display."

-(Shneiderman, DTUI, 1998)

Example



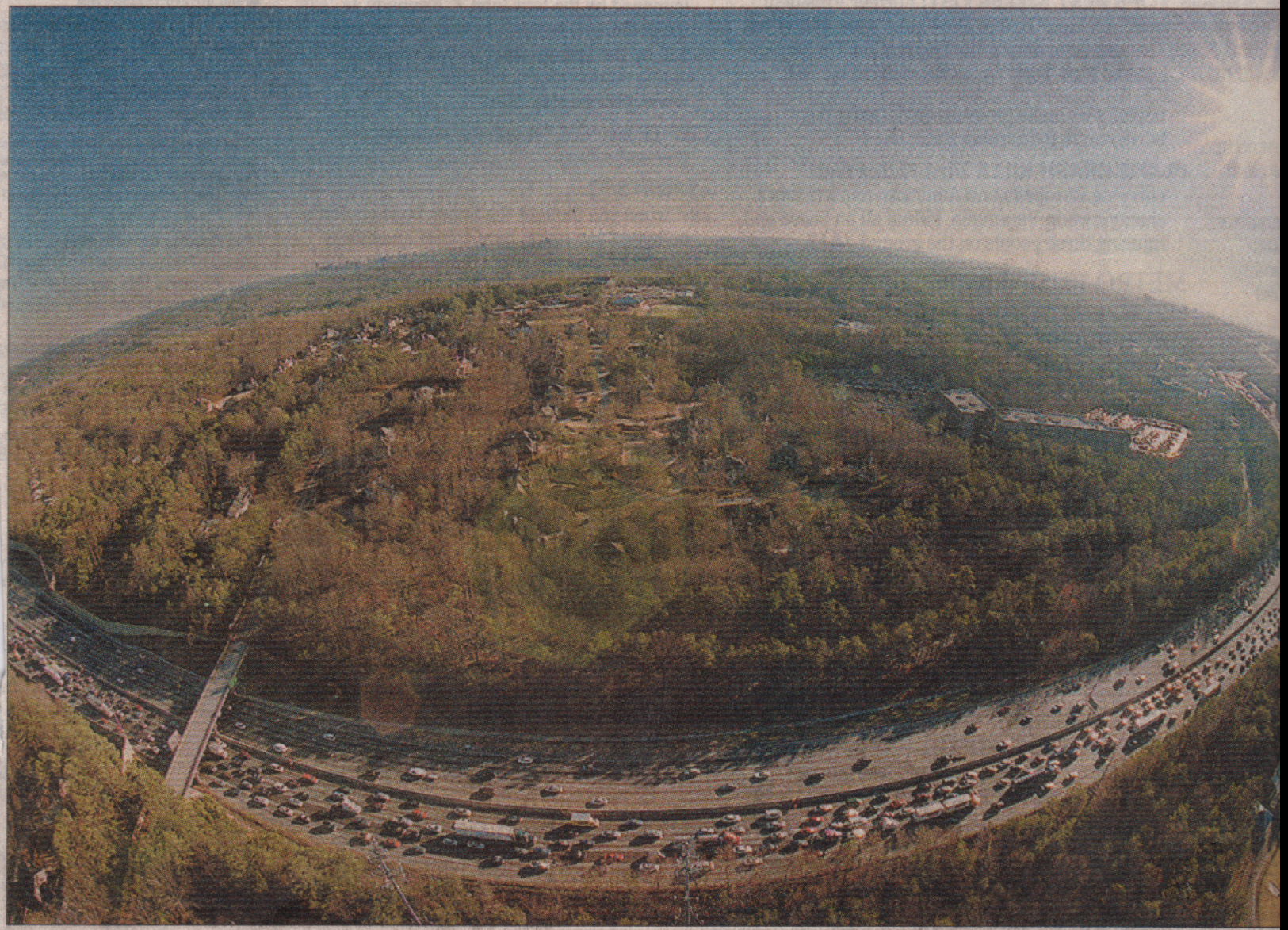
(Not really a good one. Can you see why?
---just a Photoshop trick---)

Everyday Life Example

Real fisheye camera lens

Atlanta Journal

On I-285, another Perimeter maul



Traffic creeps along I-285 westbound during the evening rush hour Thursday after a hole developed in a bridge over the Chattahoochee River and th
closed. Officials say recent construction work may have caused the problem. This photo, taken with a fisheye lens, is looking south toward downtow

Fisheye Terminology

Focal point

Level of detail

Distance from focus

Degree of interest function

DoI Function

Can take on various forms

- Continuous - Smooth interpolation away from focus
- Filtering - Past a certain point, objects disappear
- Step - Levels or regions dictating rendering
  0<x<.3 all same, .3<x<.6 all same
- Semantic changes - Objects change rendering at different levels

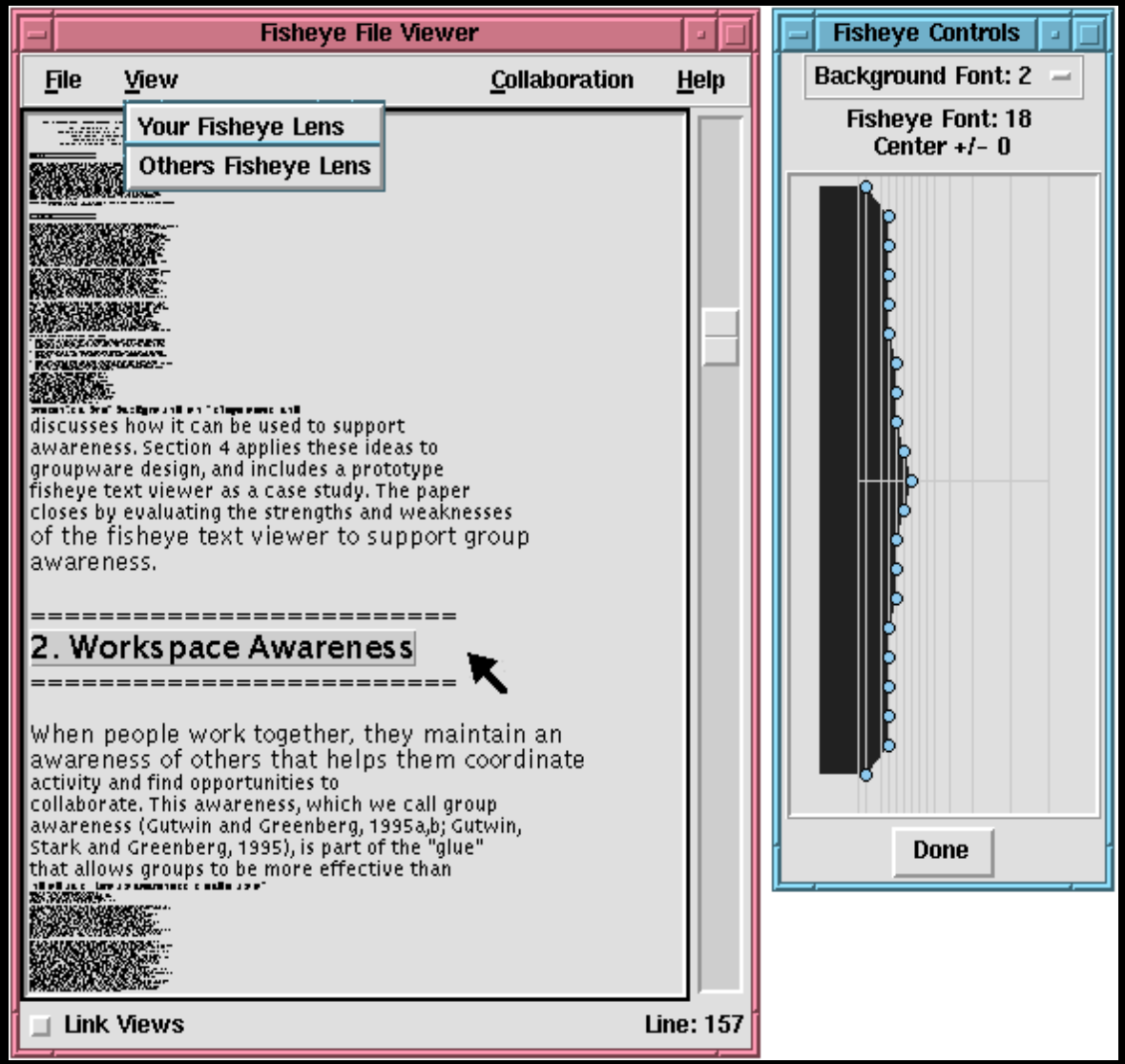
Applications

Text/program viewing

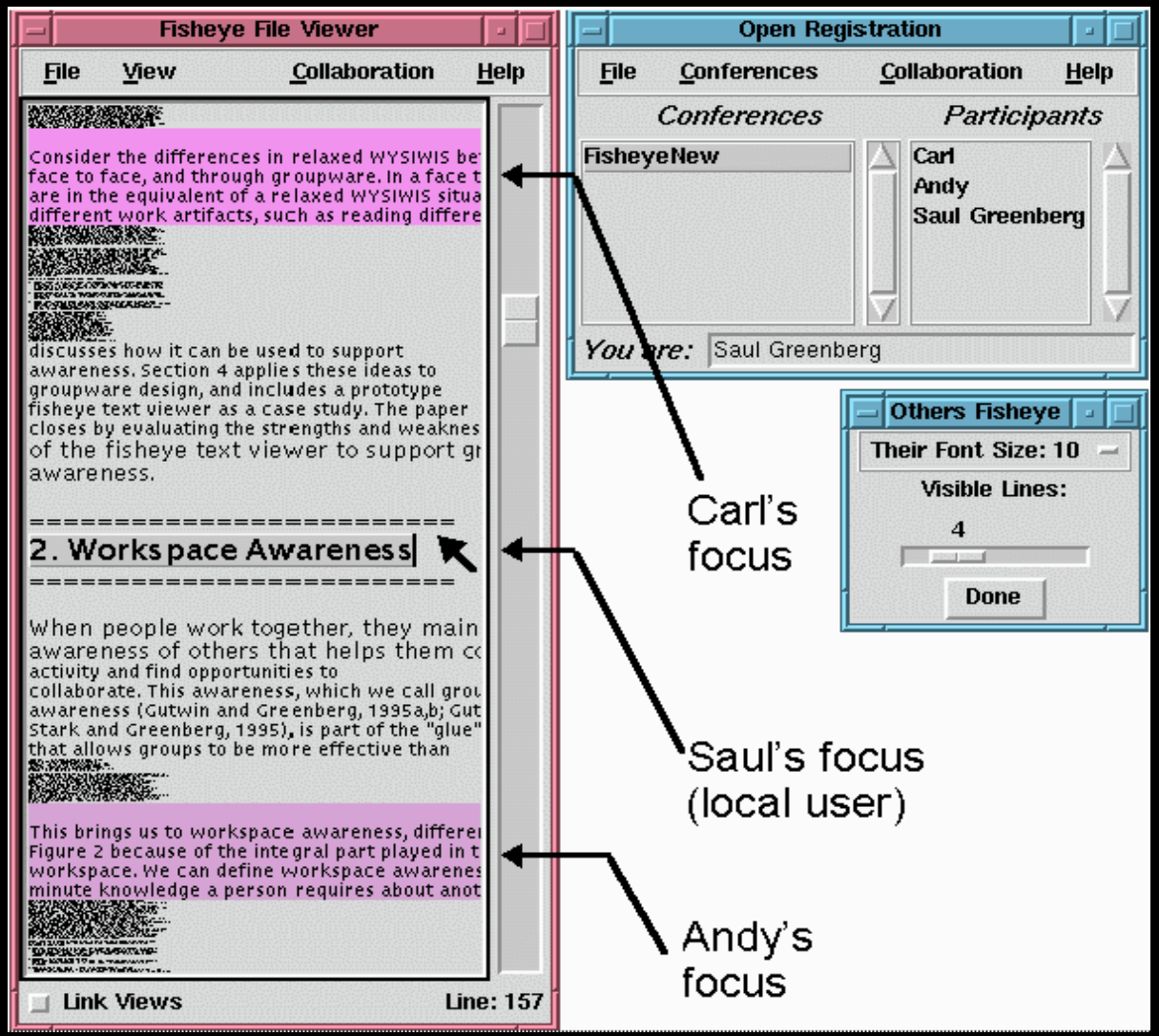
Furnas' original example

Shown here are examples from Gutwin and Greenberg

Step function



Applications



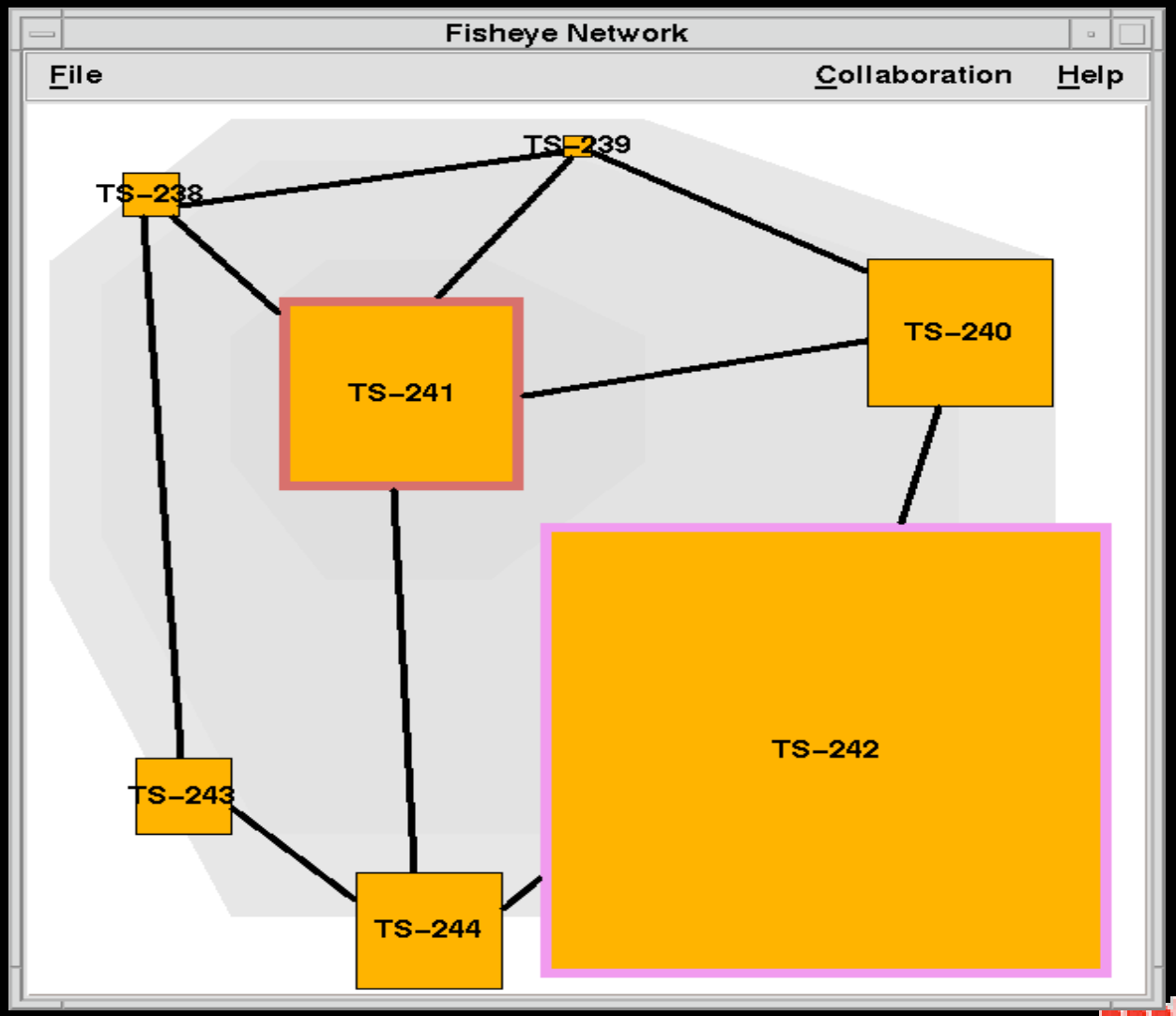
Shared text editor for CSCW

Gutwin and Greenberg, '96

Applications

Viewing nodes in networks

Gutwin and Greenberg



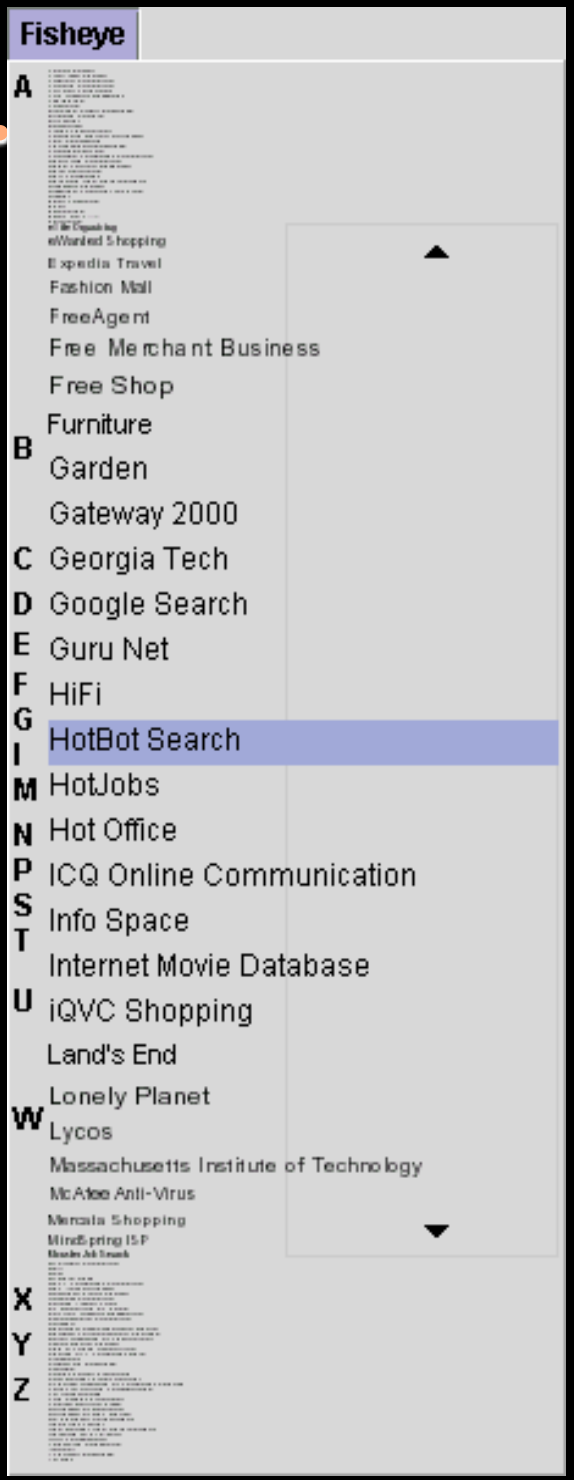
Applicat

Fisheye menu

Uses a focus-lock mode
(move cursor to right)

Demo:
`www.cs.umd.edu/hcil/fisheyemenu`

Bederson '00



Graphical Fisheye Views

- Apply fisheye techniques to 2D graph
- Experiment with a variety of distortion factors
- Interactive tool that allows user to browse display and change focus

Sarkar and Brown ‘94

Graphical Fisheye Views

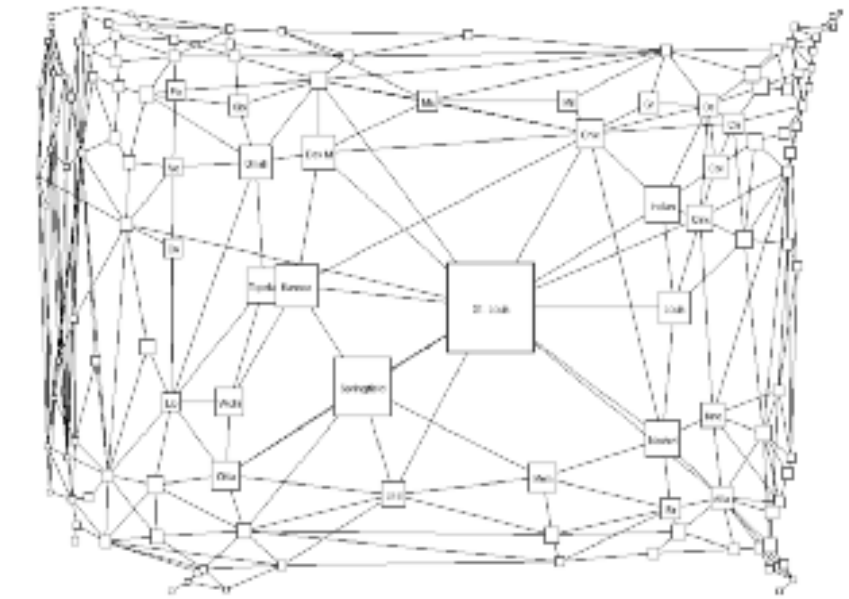


Figure 2: A fisheye view of the graph in Figure 1. The focus is on St. Louis. (The values of the fisheye parameters are $\beta = 5$, $\gamma = 0$, $\epsilon = 0$, $VWcutoff = 0$; the meanings of these parameters are explained in Sections 4 and 6.)

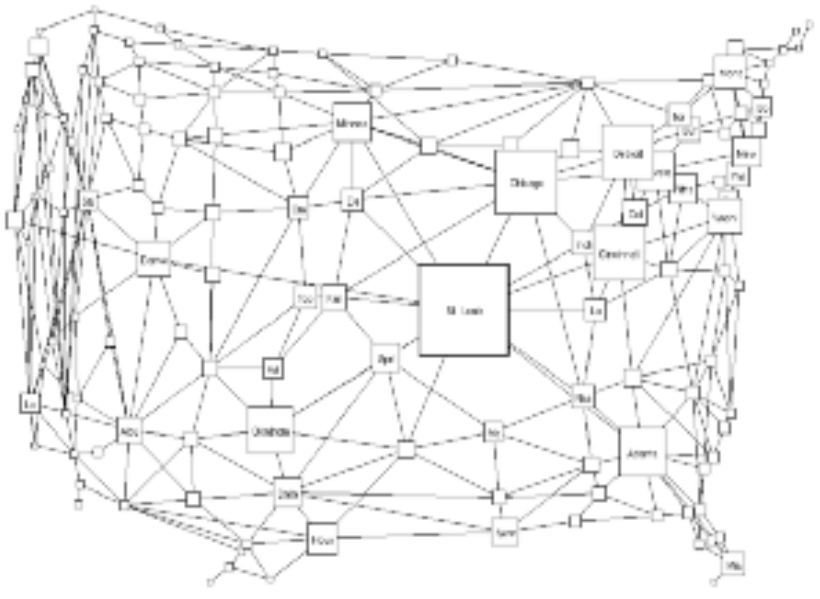
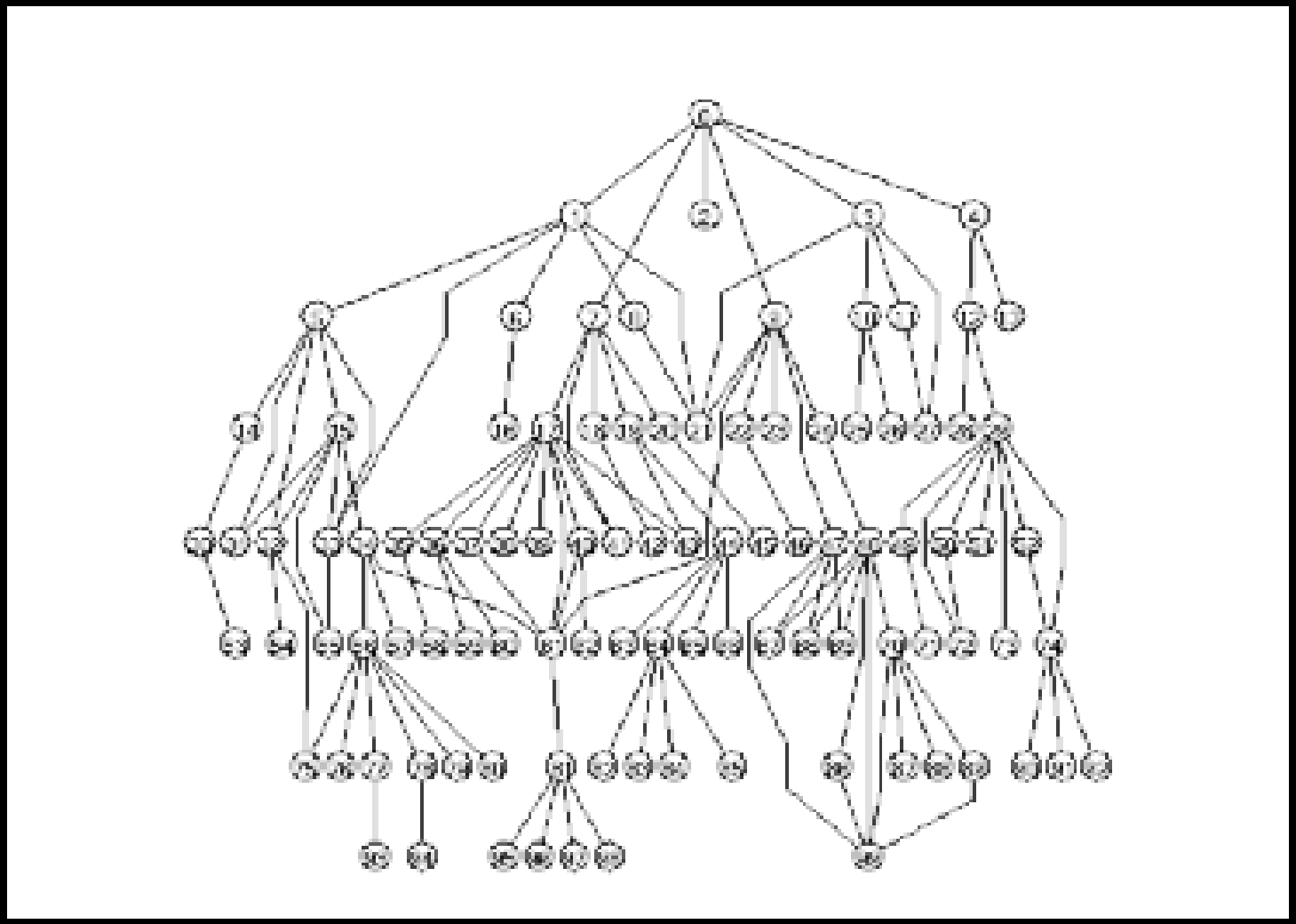


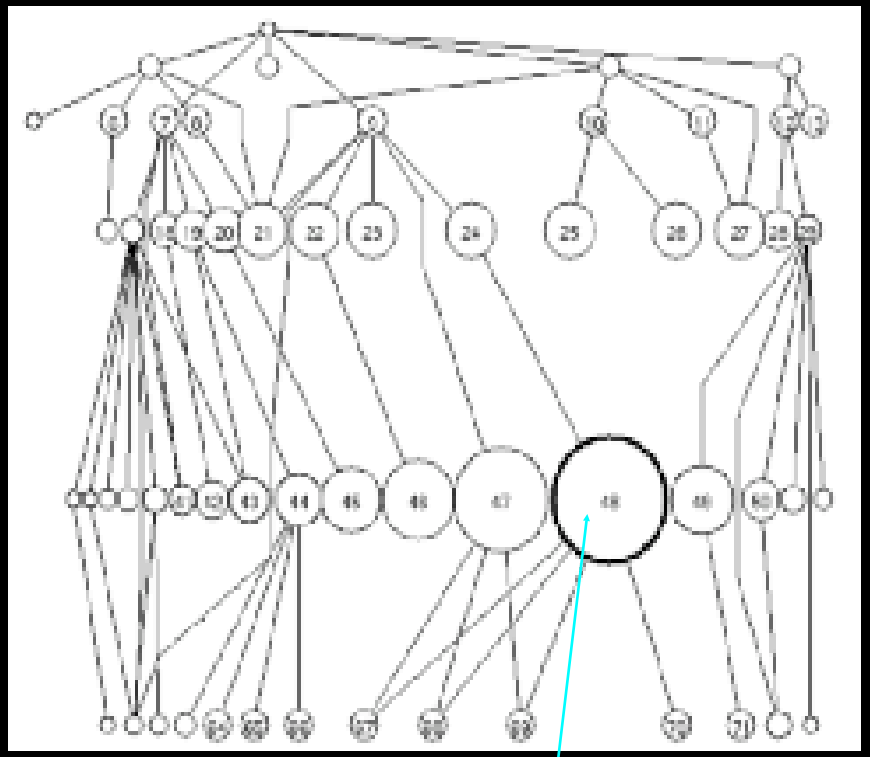
Figure 3: A fisheye view of the graph in Figure 1, with less distortion than in Figure 2. The values of the fisheye parameters are $\beta = 2$, $\gamma = 0.5$, $\epsilon = 0.5$, $VWcutoff = 0$.

Example

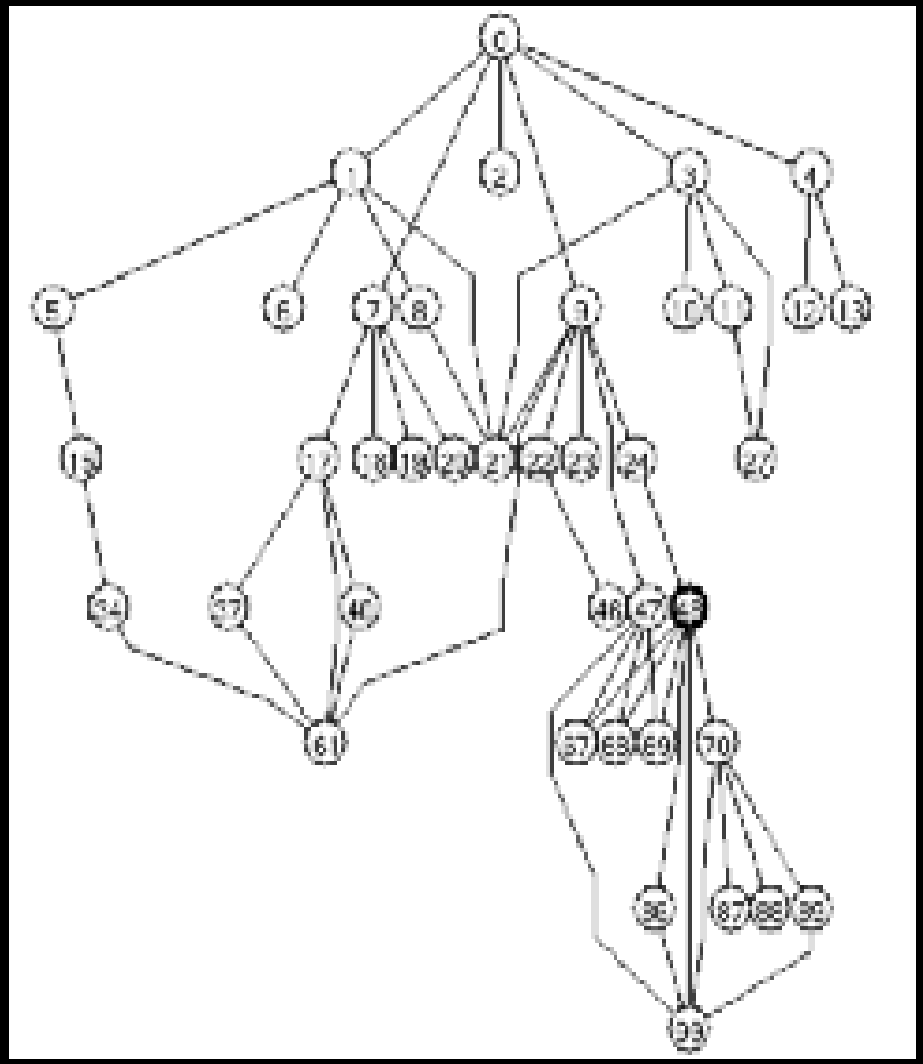


Original

Example



Focal point



Constraining Changes

Continuous zoom

- Can change focal point smoothly in graph
- Other nodes give up space
- Bartram et al '95

Constrained fisheye

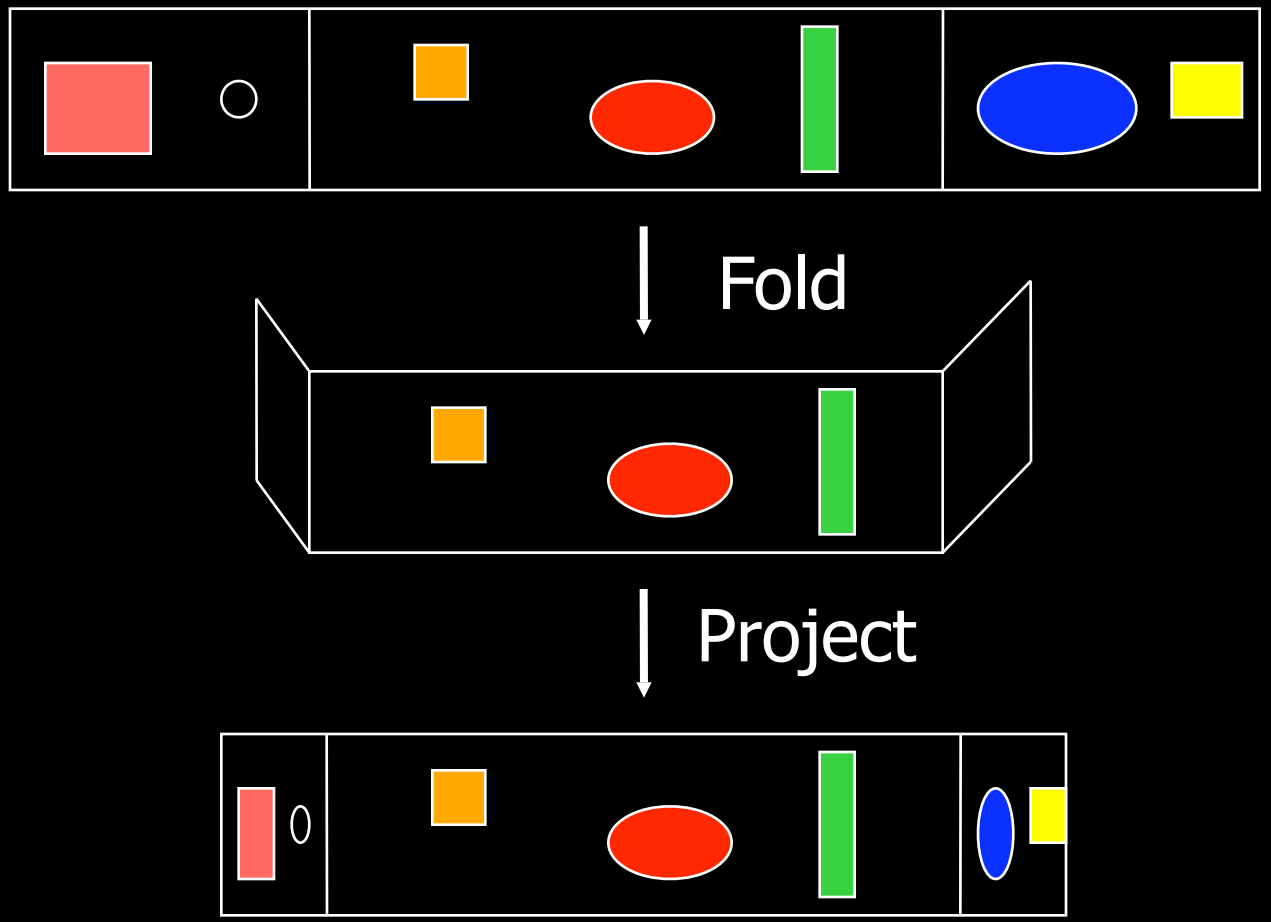
- Make transitions in focus more aesthetically pleasing and easier to track
- Storey et al '99

Bifocal Display

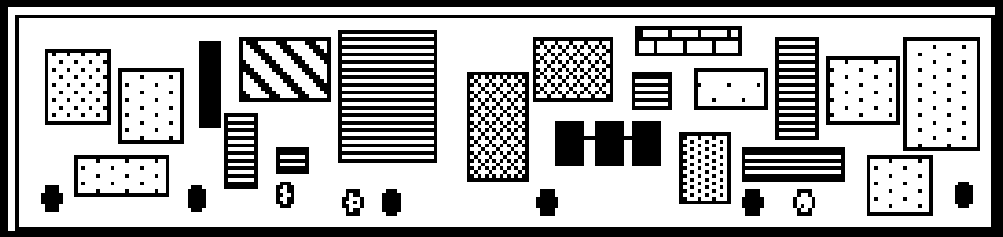
- Interesting application of fisheye view
- View office documents
- Take items in periphery and fold back in 3-space
- Project onto front viewing screen

Spence & Apperly '82

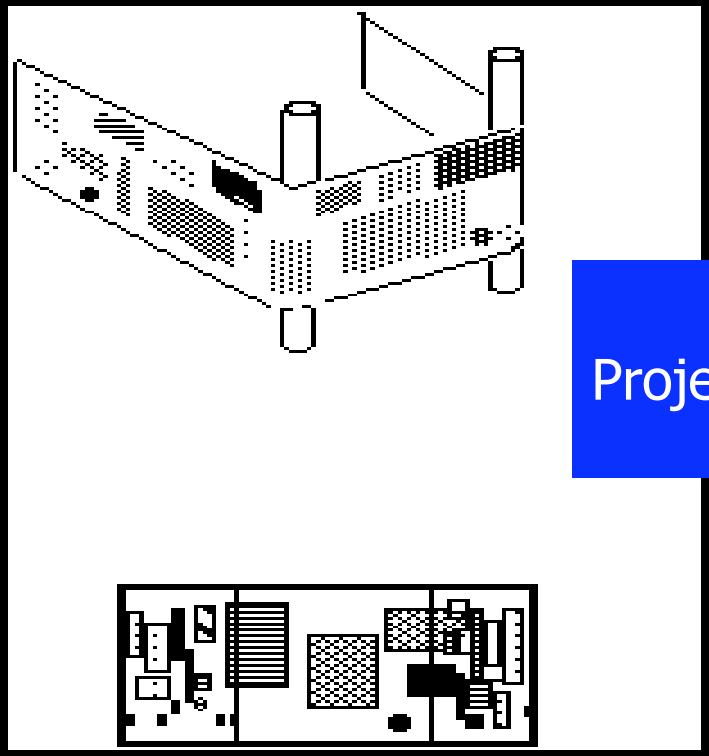
Bifocal Display



Bifocal Display

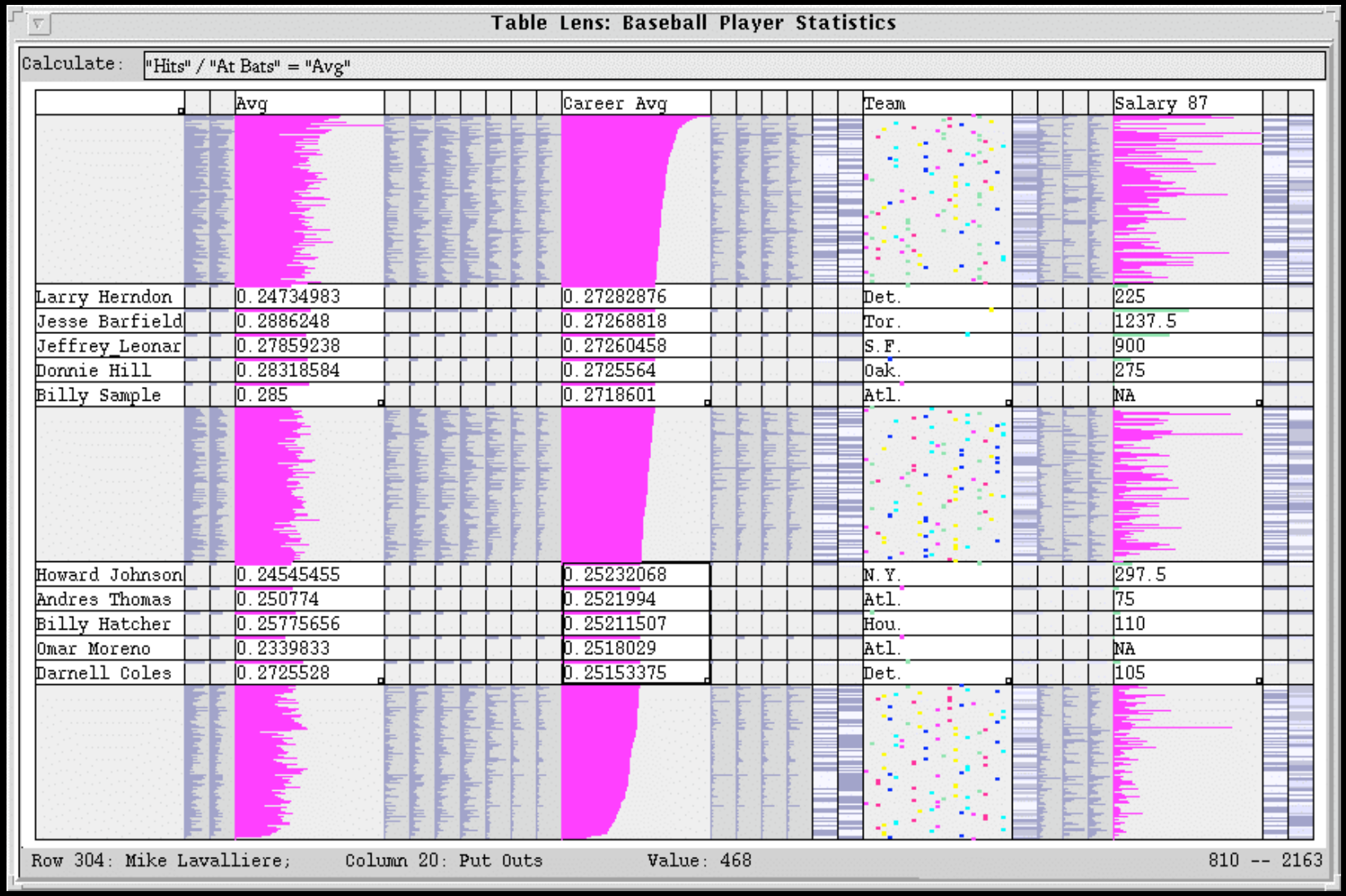


Fold



Project

Table Lens



From Xerox PARC and Inxight

A bifocal display

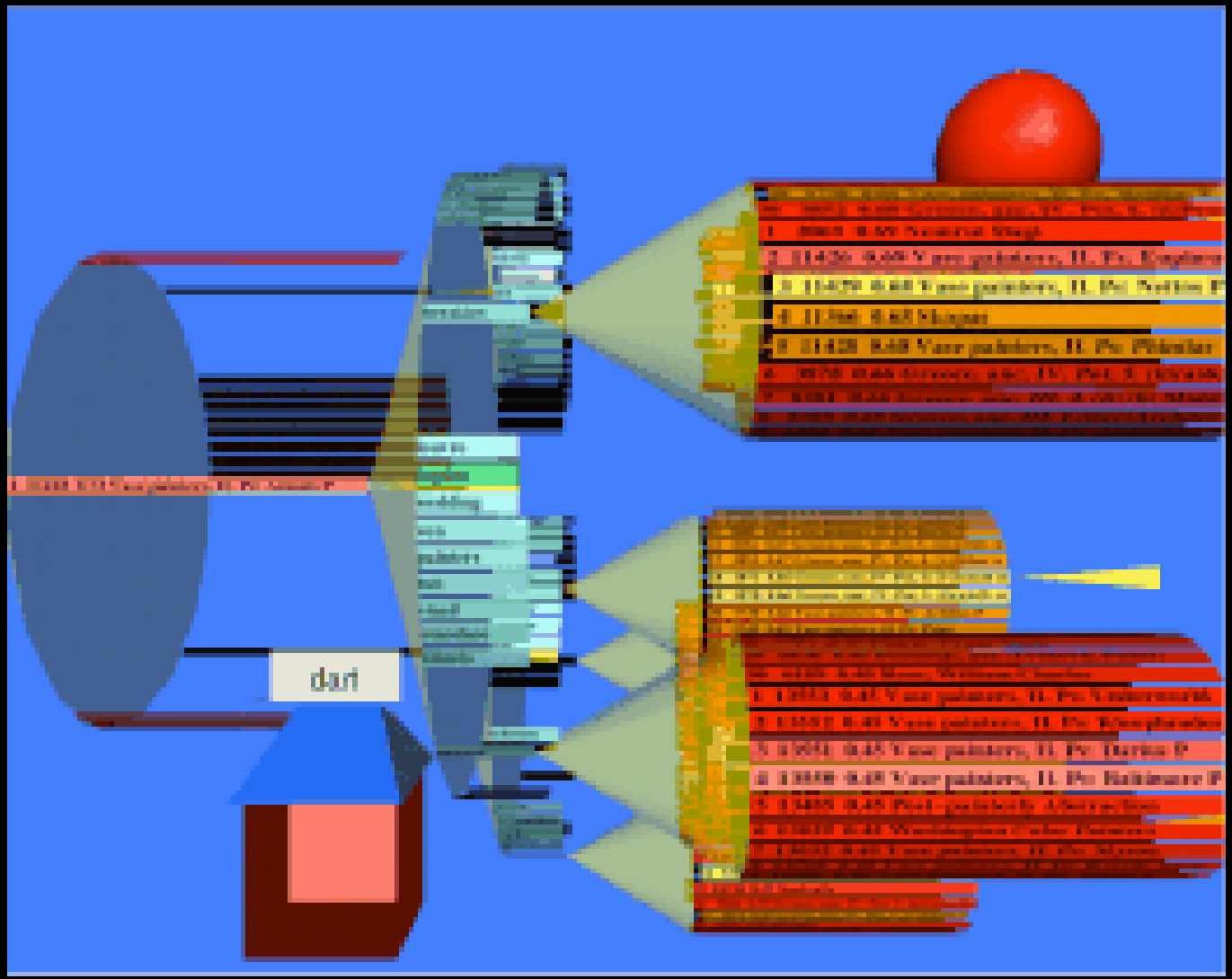
Perspective Wall

Computerized, automated 3D implementation of Bifocal display

Map work charts onto diagram, x-axis is time, y-axis is project

Mackinlay, Robertson, Card '91

Another Cone Tree Example



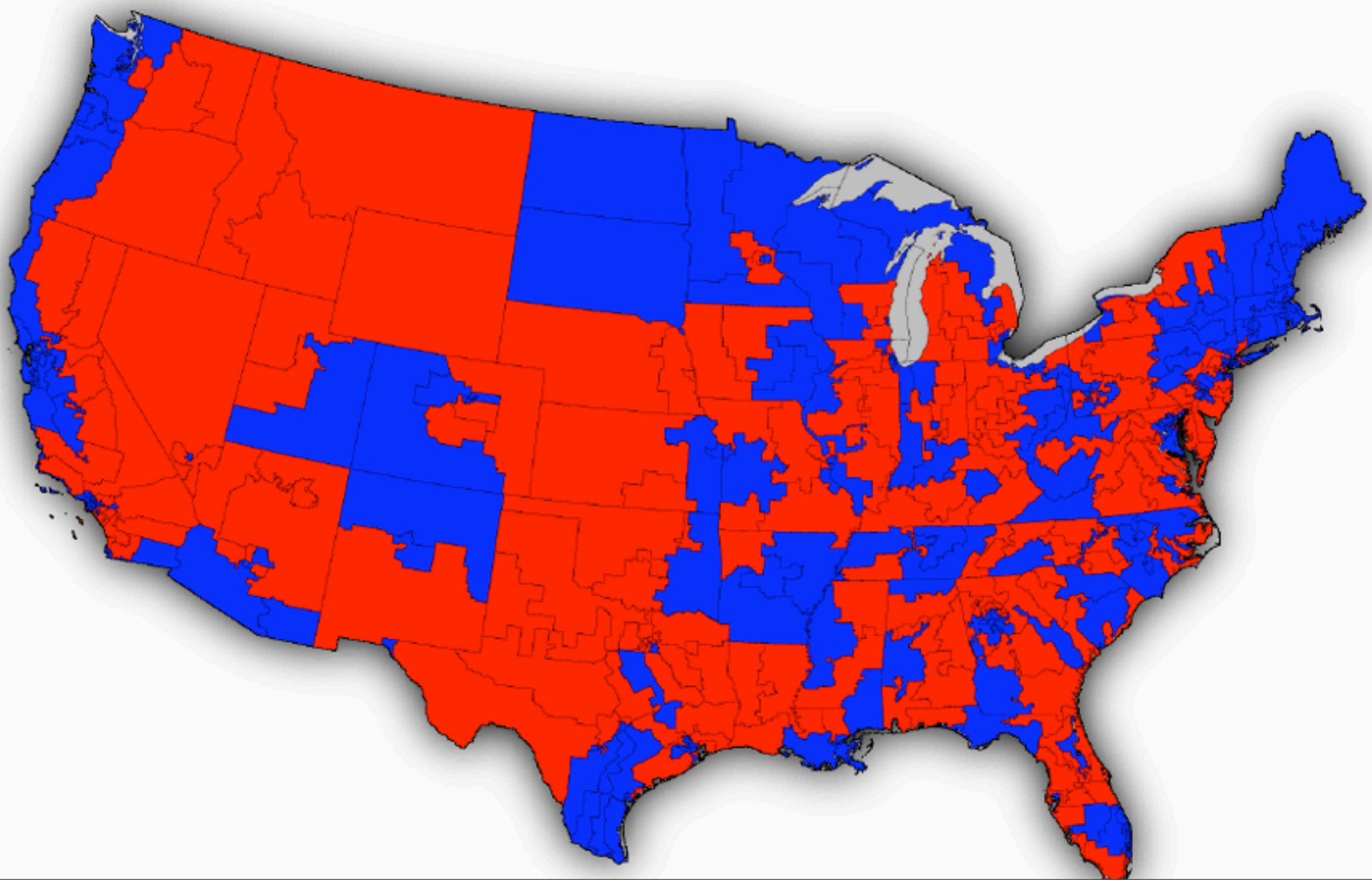
Disadvantages

Distortion can be annoying

Can be very difficult to implement

Any change in focal point potentially requires recalculation of DoI for all objects and hence re-rendering of all objects -> Expensive!

Cartograms



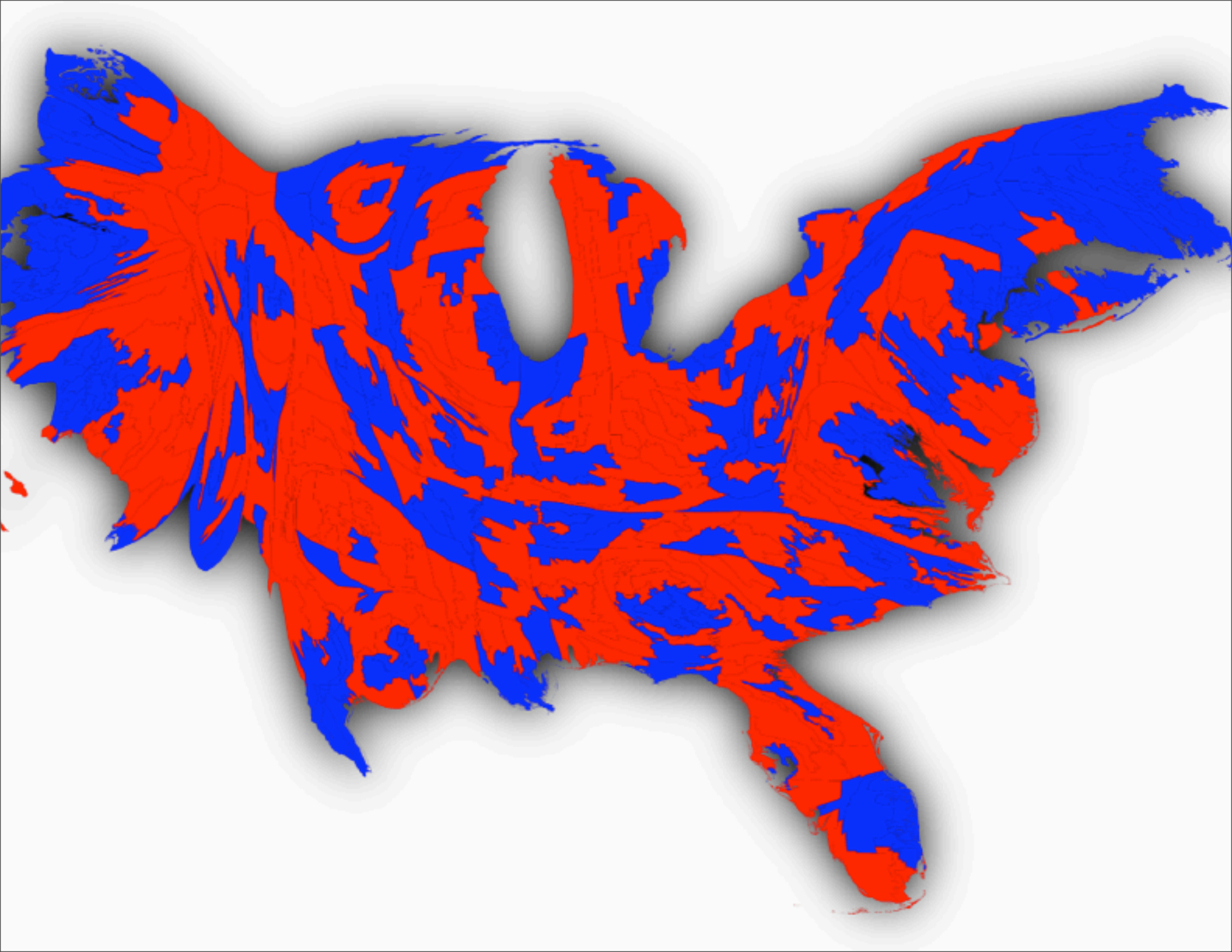
Cartograms

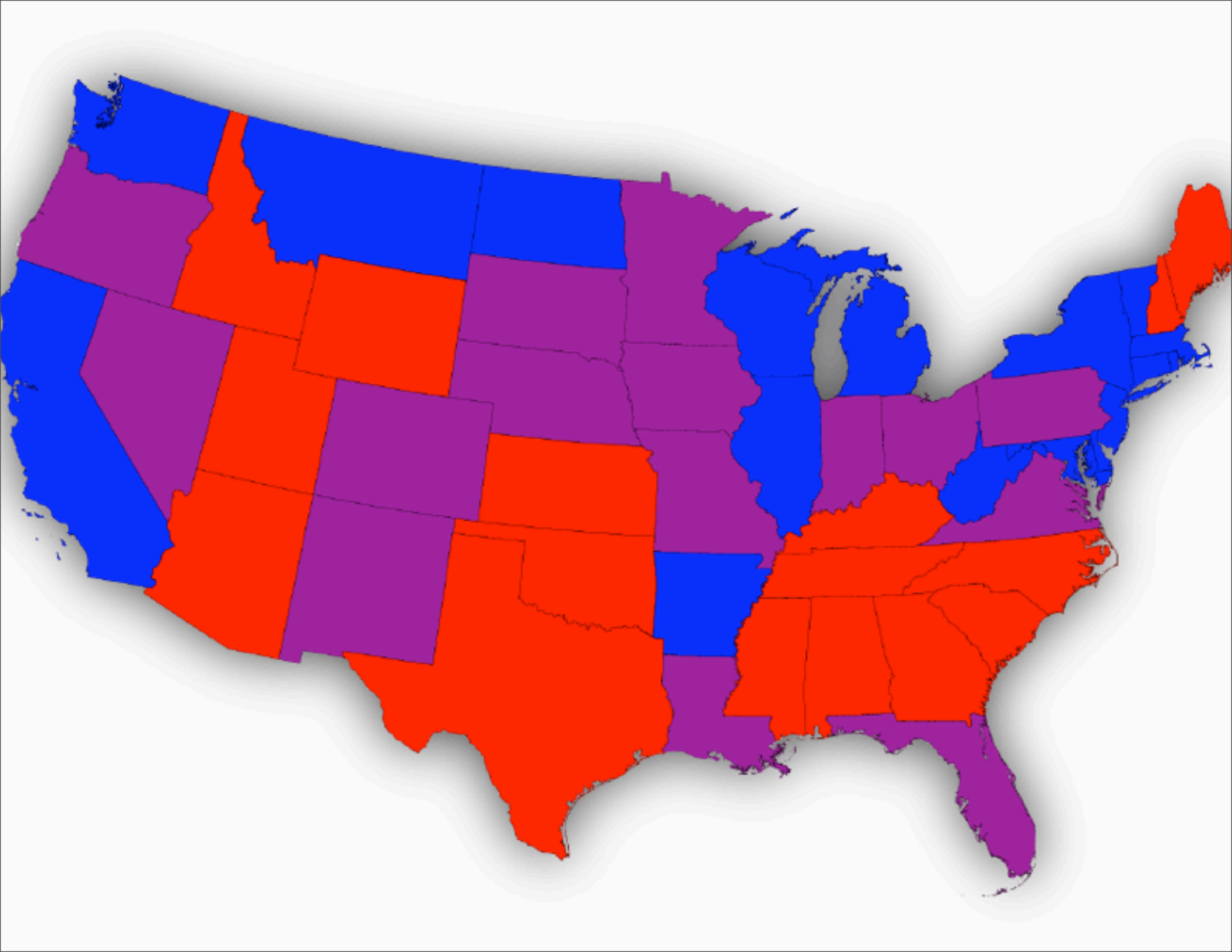
Cartograms

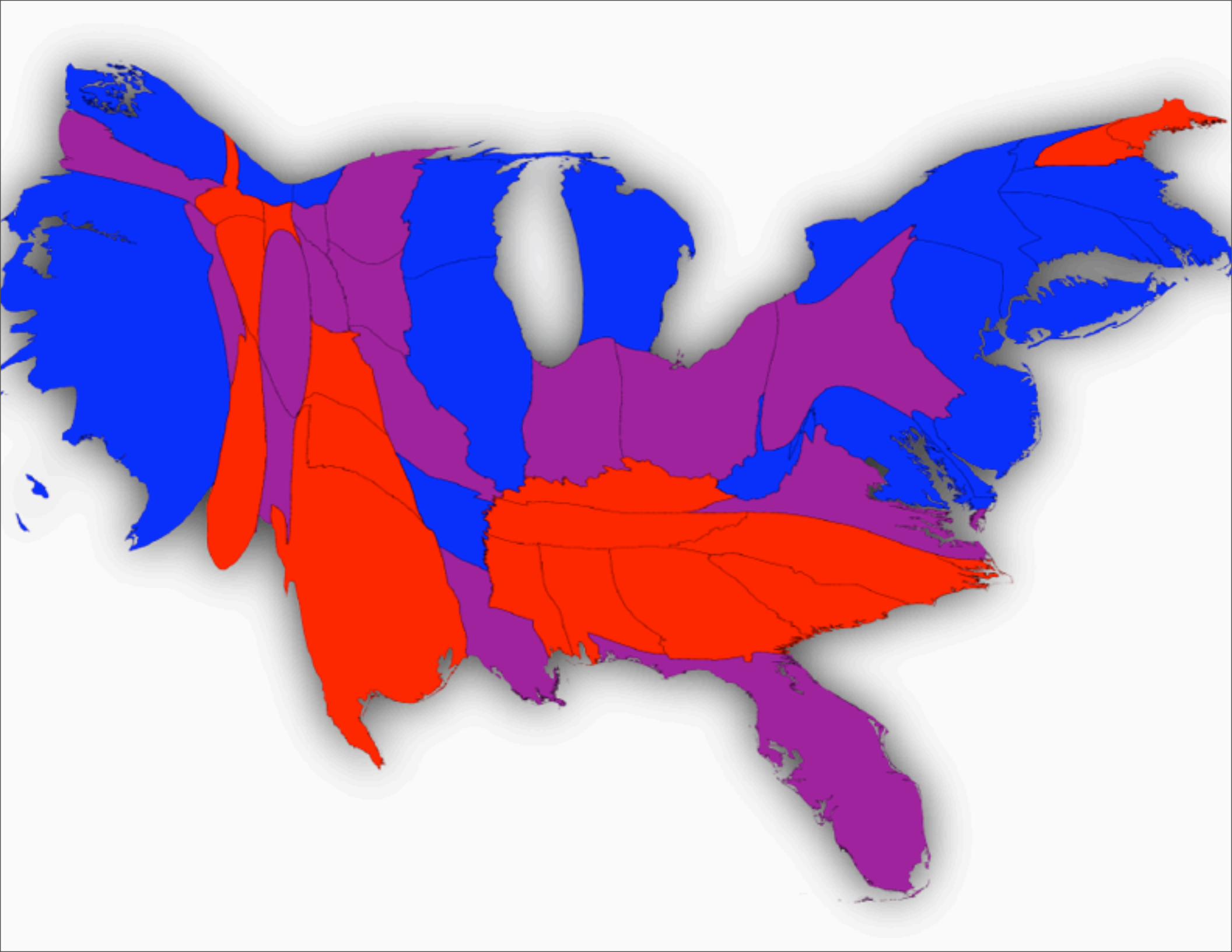
Vis or Infovis?

There is a 2D natural space, but it is misleading!

Solution: distort the space!







http://www-personal.umich.edu/~mejn/cartograms/