Information Visualization for Knowledge Discovery

Ben Shneiderman *ben@cs.umd.edu*

Founding Director (1983-2000), Human-Computer Interaction Lab
Professor, Department of Computer Science
Member, Institute for Advanced Computer Studies

UNIVERSITY OF MARYLAND

(Copyright Ben Shneiderman 2009)

HCIL

Interdisciplinary research community

- Computer Science & Info Studies
- Psych, Socio, Poli Sci & MITH

(www.cs.umd.edu/hcil)

Scientific Approach (~~beyond user friendly~~)

- Specify users and tasks
- Predict and measure
  - time to learn
  - speed of performance
  - rate of human errors
  - human retention over time
- Assess subjective satisfaction

  (Questionnaire for User Interface Satisfaction)
- Accommodate individual differences
- Consider social, organizational & cultural context

HCIL

Design Issues

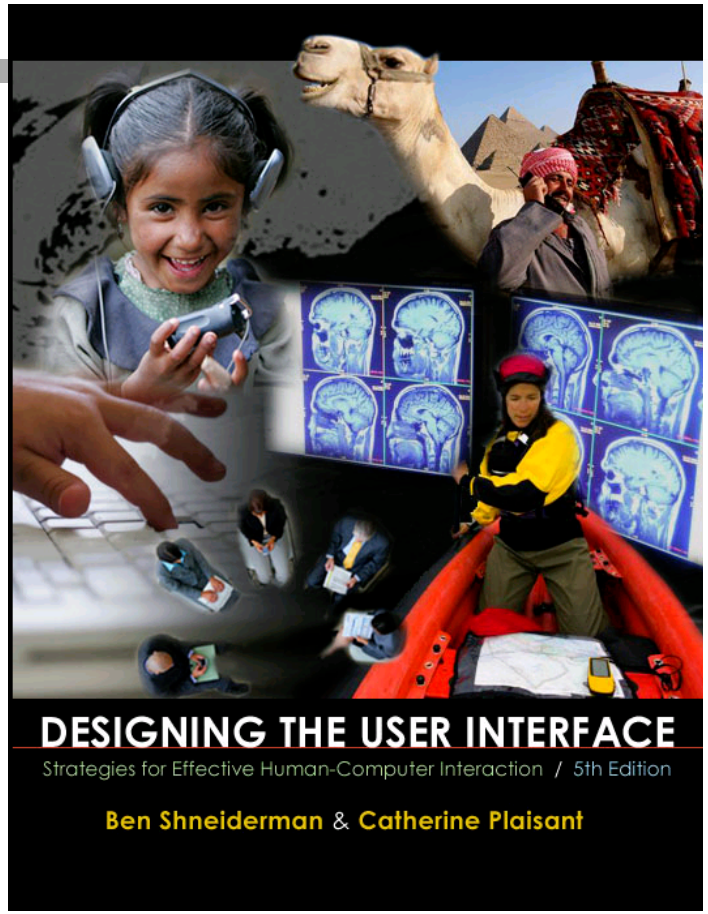
- Input devices & strategies
  - Keyboards, pointing devices, voice
  - Direct manipulation
  - Menus, forms, commands
- Output devices & formats
  - Screens, windows, color, sound
  - Text, tables, graphics
  - Instructions, messages, help
- Collaboration
- Help, tutorials, training

www.awl.com/DTUI

HCIL

Design Issues

- Input devices & strategies
  - Keyboards, pointing devices, voice
  - Direct manipulation
  - Menus, forms, commands
- Output devices & formats
  - Screens, windows, color, sound
  - Text, tables, graphics
  - Instructions, messages, help
- Collaboration & Social Media
- Help, tutorials, training
- Search & Visualization



www.awl.com/DTUI

Fifth Edition: March 2009

HCIL

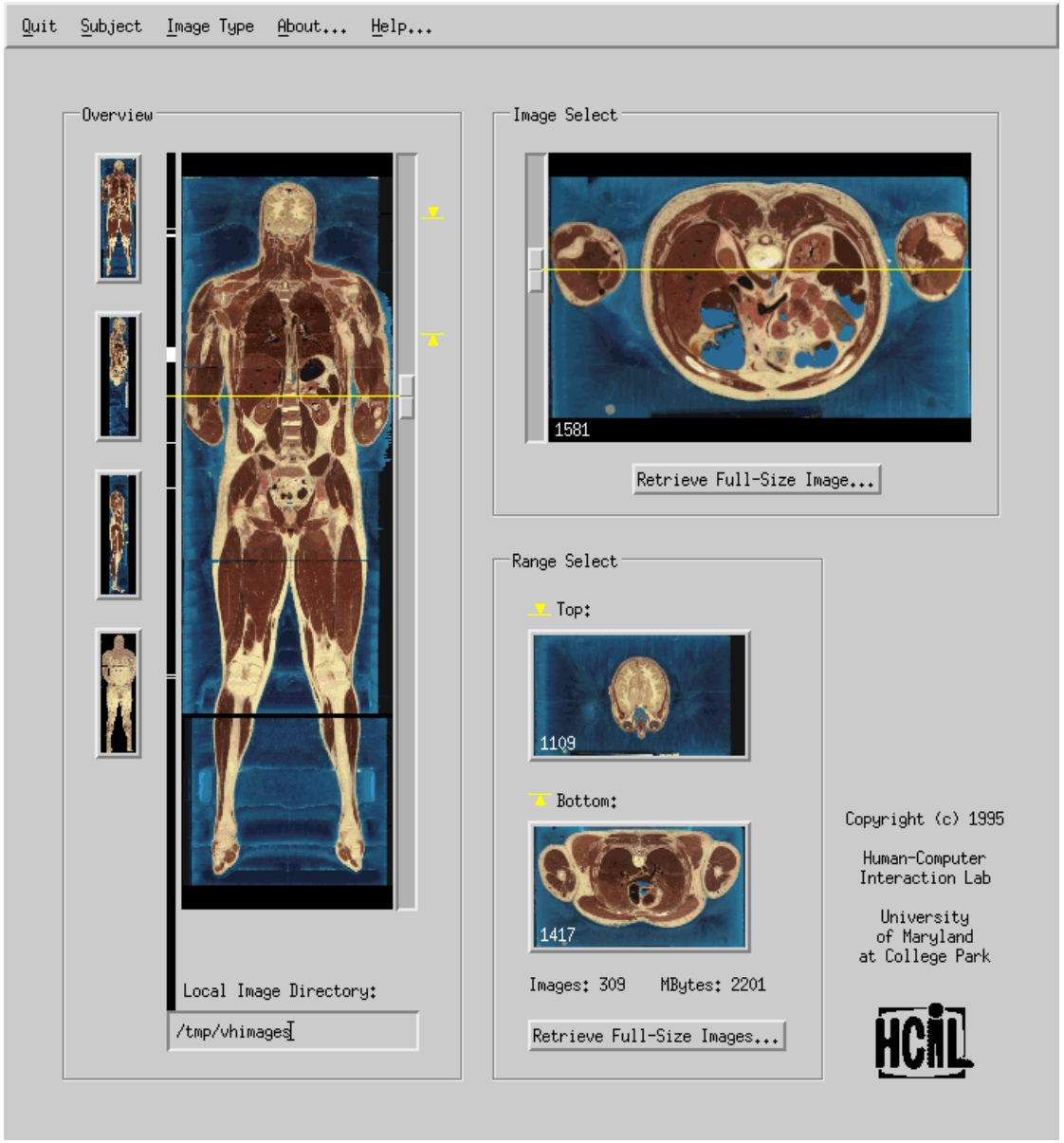
U.S. Library of Congress

- Scholars, Journalists, Citizens
- Teachers, Students

HCIL

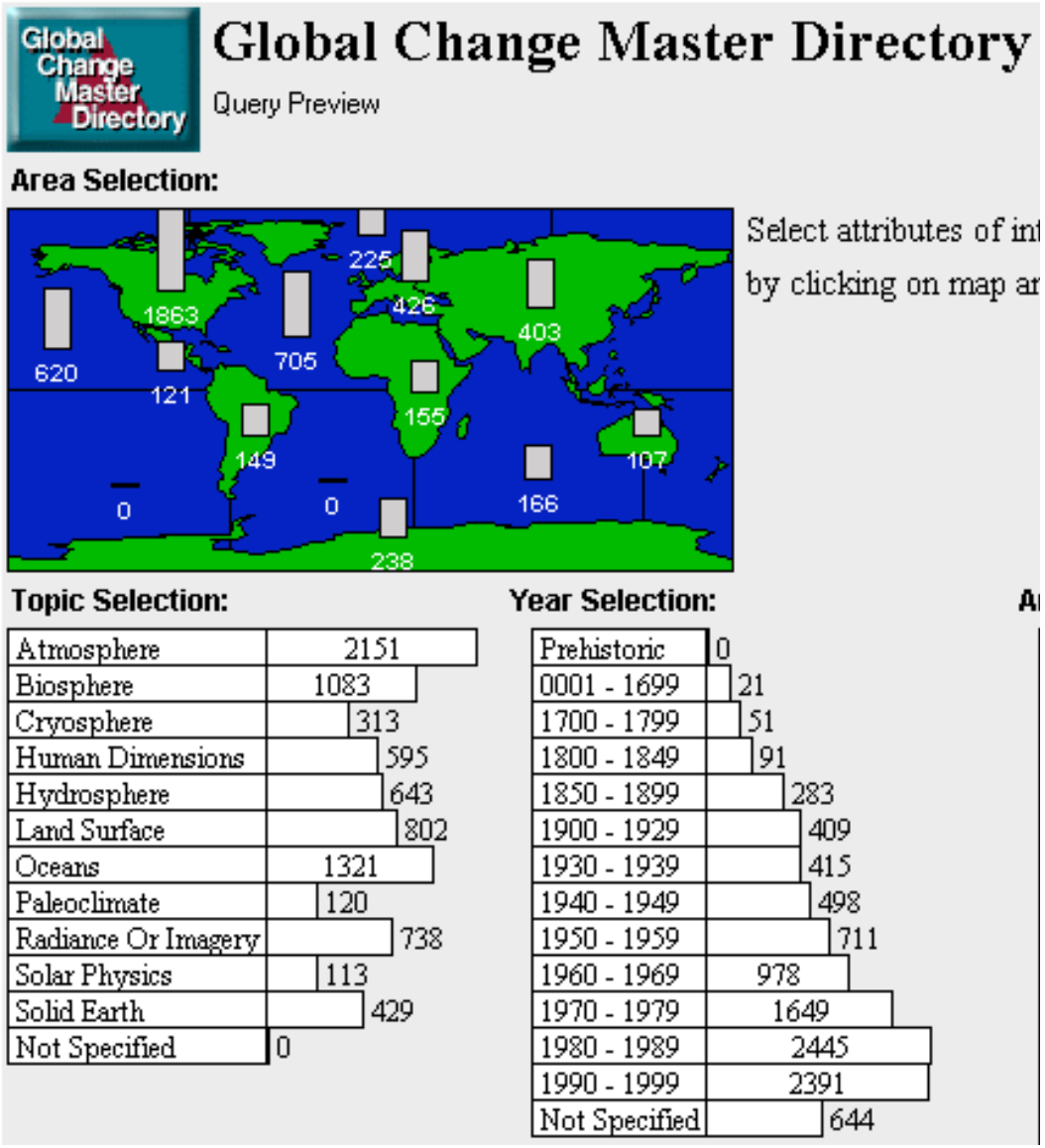
Visible Human Explorer (NLM)

- Doctors
- Surgeons
- Researchers
- Students



NASA Environmental Data

- Scientists
- Farmers
- Land planners
- Students



Bureau of the Census

- Economists, Policy makers, Journalists
- Teachers, Students

American FactFinder™

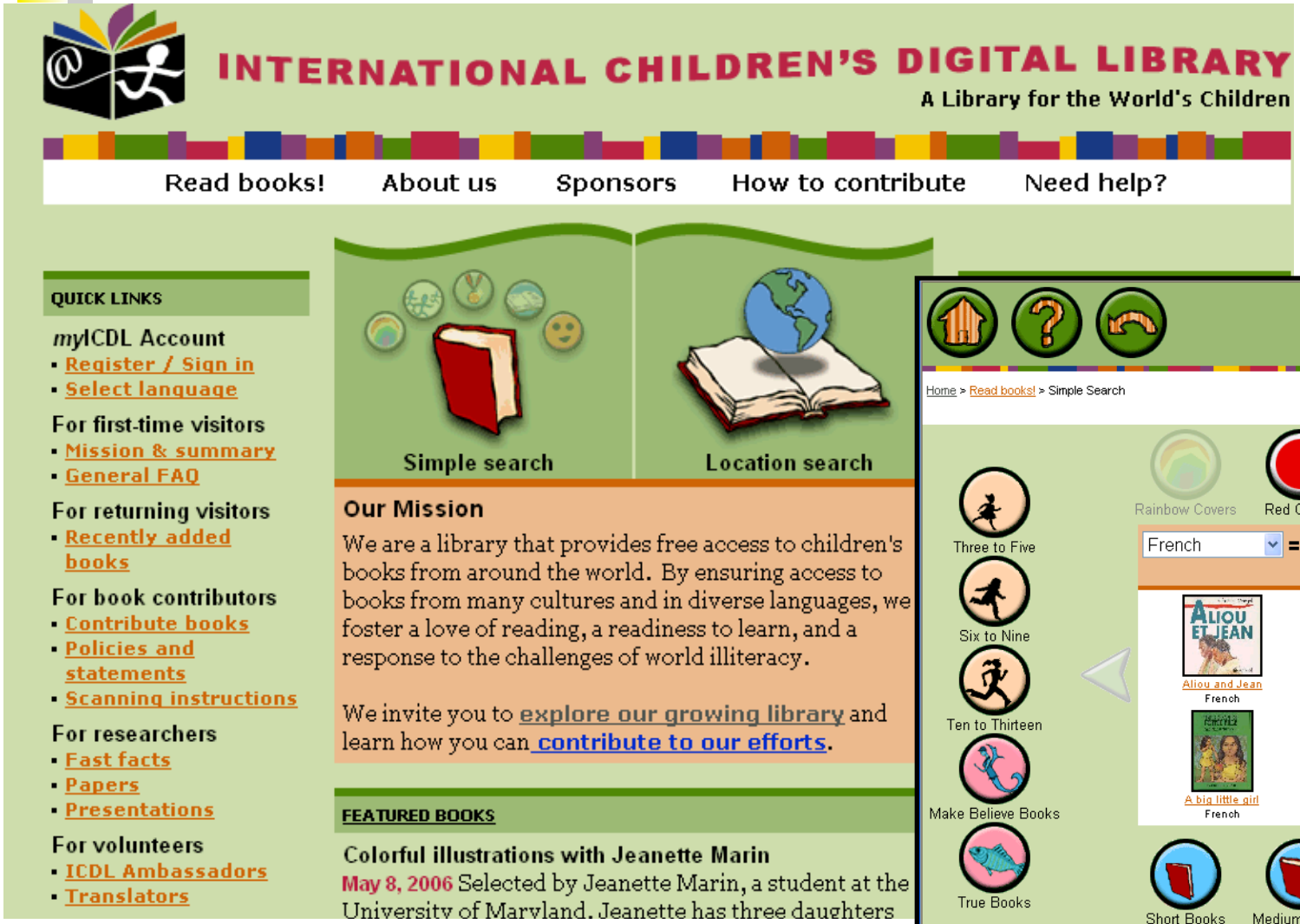
NSF Digital Government Initiative

Census,
NCHS,
BLS, EIA,
NASS, SSA

- Find what you need
- Understand what you Find

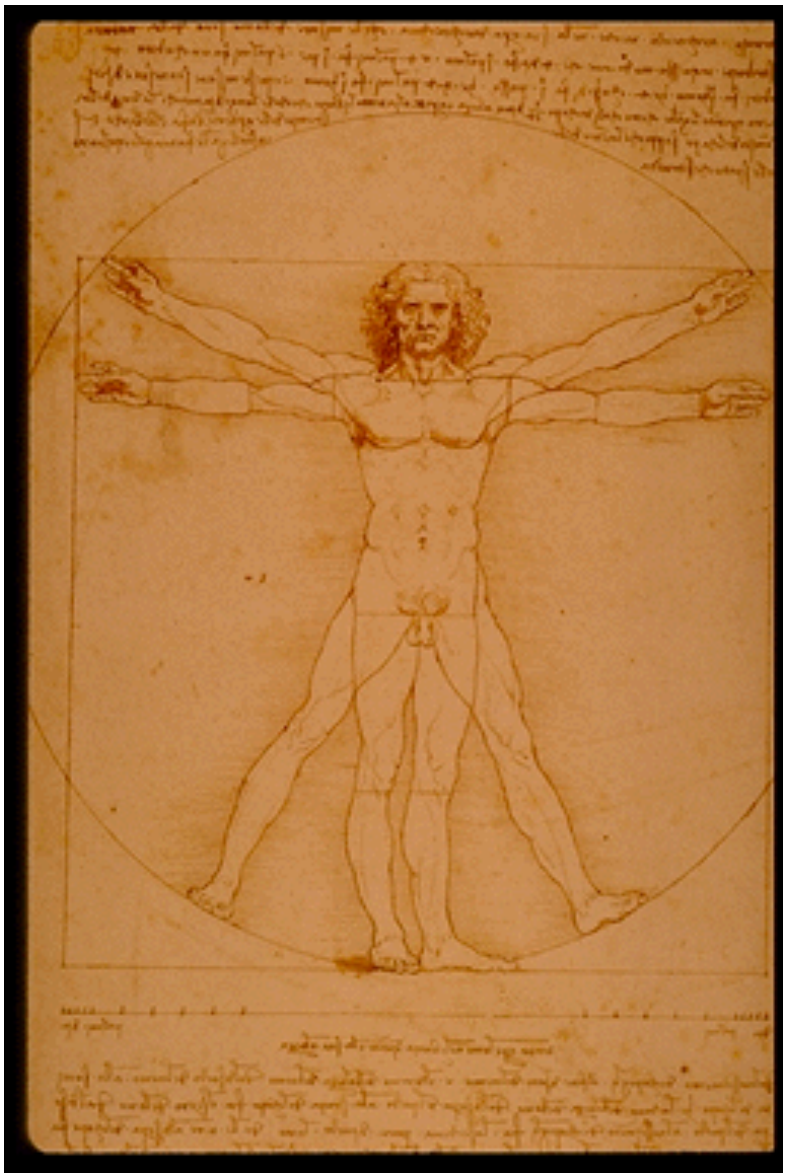
www.ils.unc.edu/govstat/

International Children's Digital Library



www.childrenslibrary.org

Information Visualization



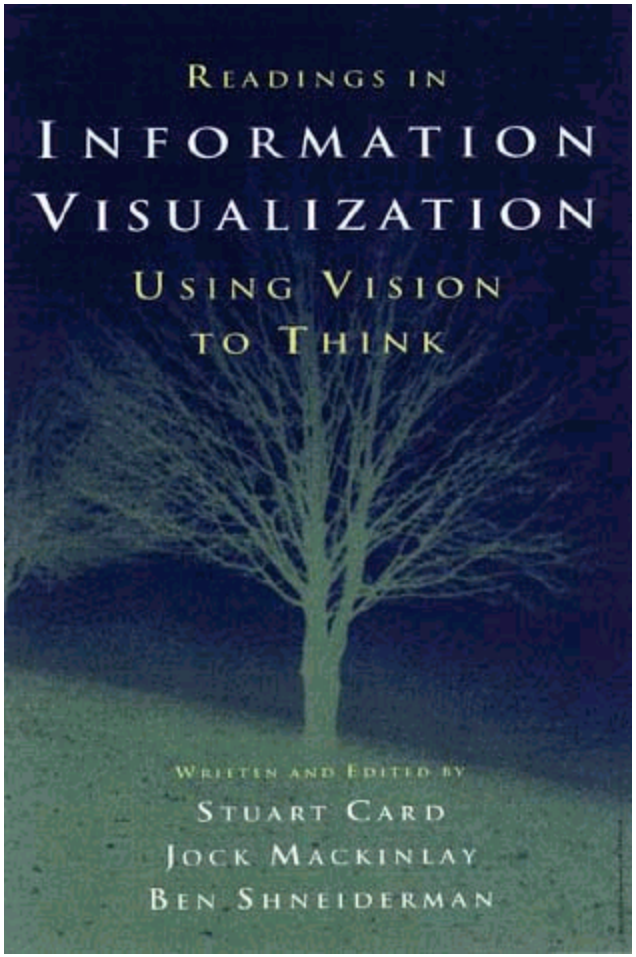
The eye…
the window of the soul,
is the principal means
by which the central sense
can most completely and
abundantly appreciate
the infinite works of nature.

Leonardo da Vinci
(1452 - 1519)

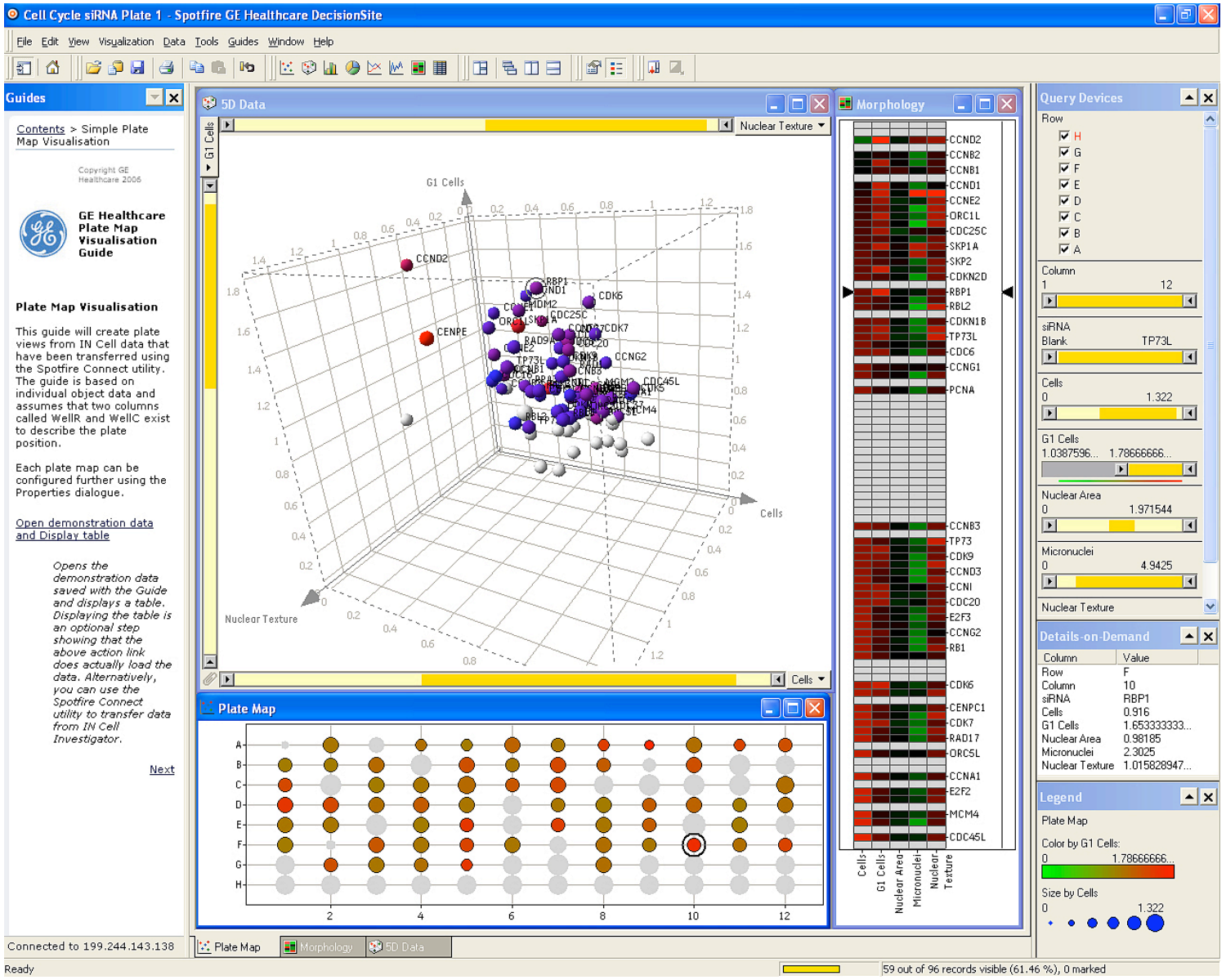
HCIL

Using Vision to Think

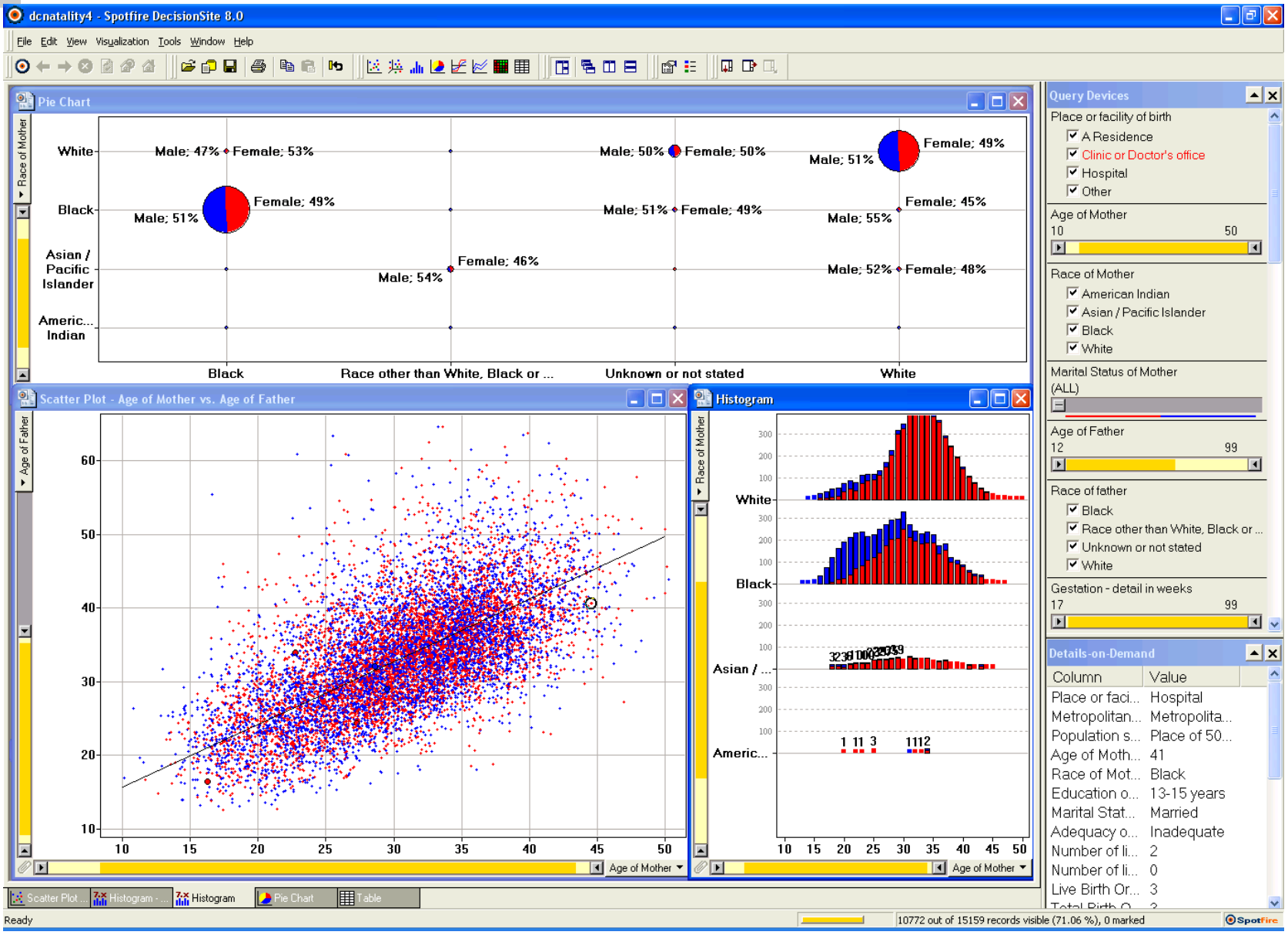
- **Visual bandwidth is enormous**
  - Human perceptual skills are remarkable
    - Trend, cluster, gap, outlier...
    - Color, size, shape, proximity...
  - Human image storage is fast and vast
- **Opportunities**
  - Spatial layouts & coordination
  - Information visualization
  - Scientific visualization & simulation
  - Telepresence & augmented reality
  - Virtual environments



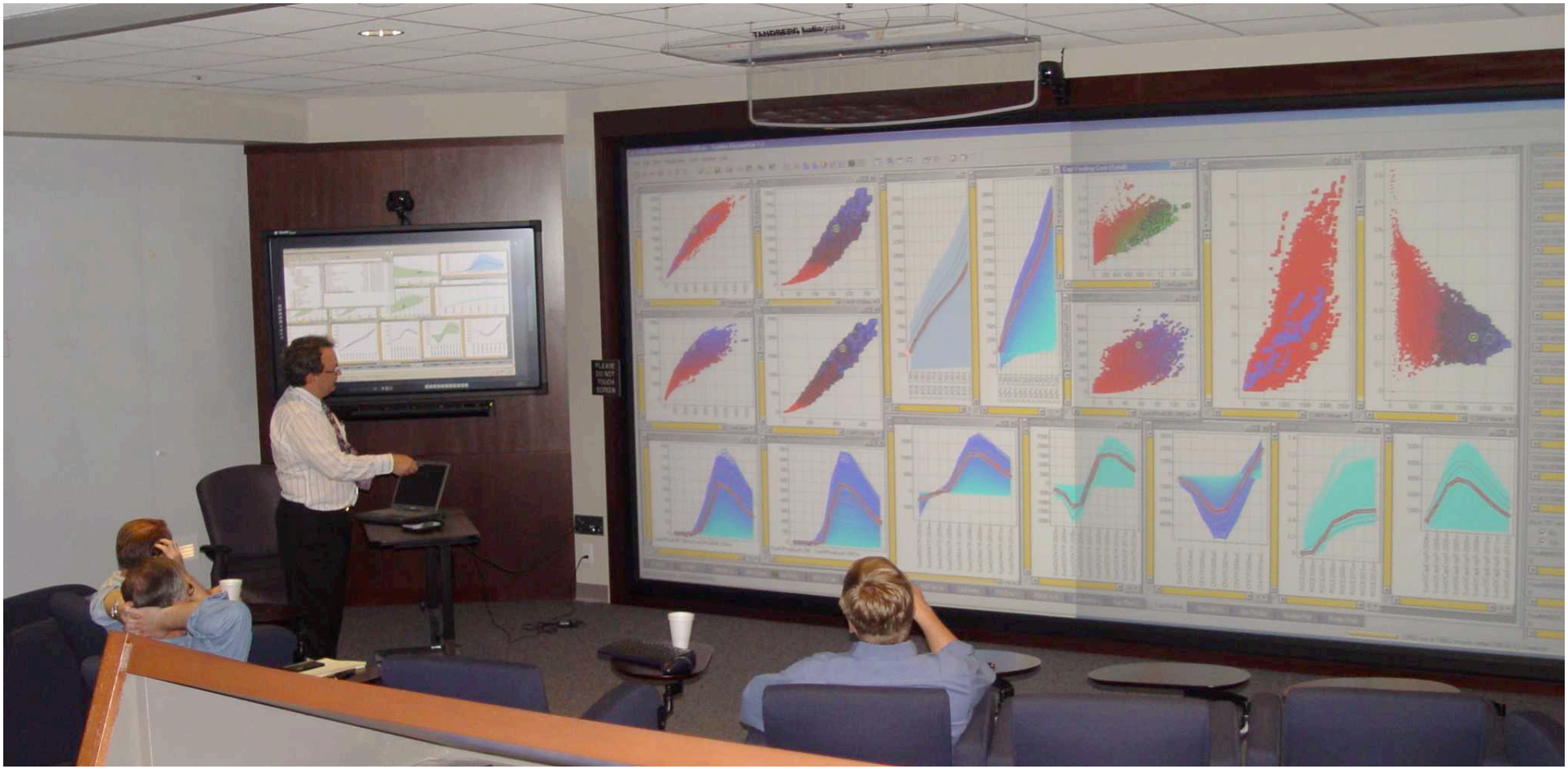
Spotfire: Retinol's role in embryos & vision



Spotfire: DC natality data

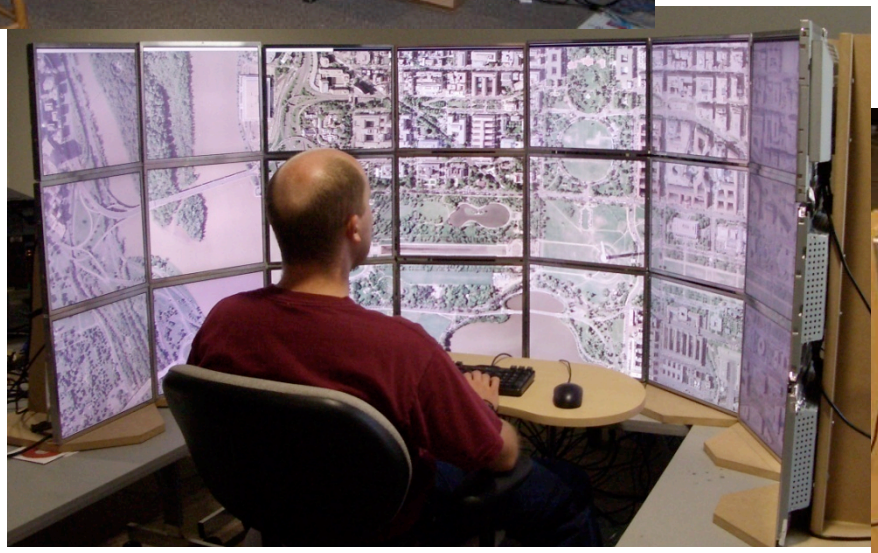


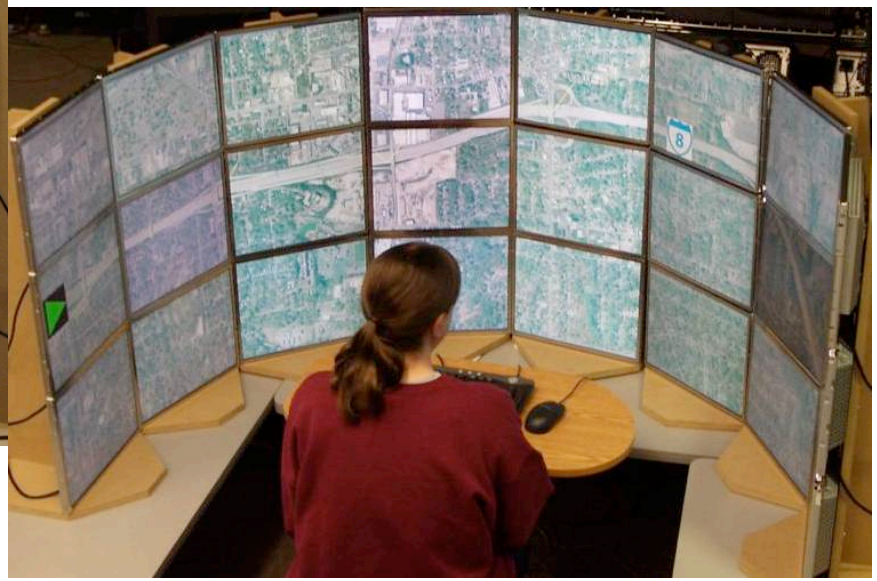
Spotfire



HCIL

100M-pixels & more





Large displays
for single users

infovis.cs.vt.edu/gigapixel

100M-pixels & more

Information Visualization: Mantra

- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand
- Overview, zoom & filter, details-on-demand

HCIL

Information Visualization: Data Types

SciViz

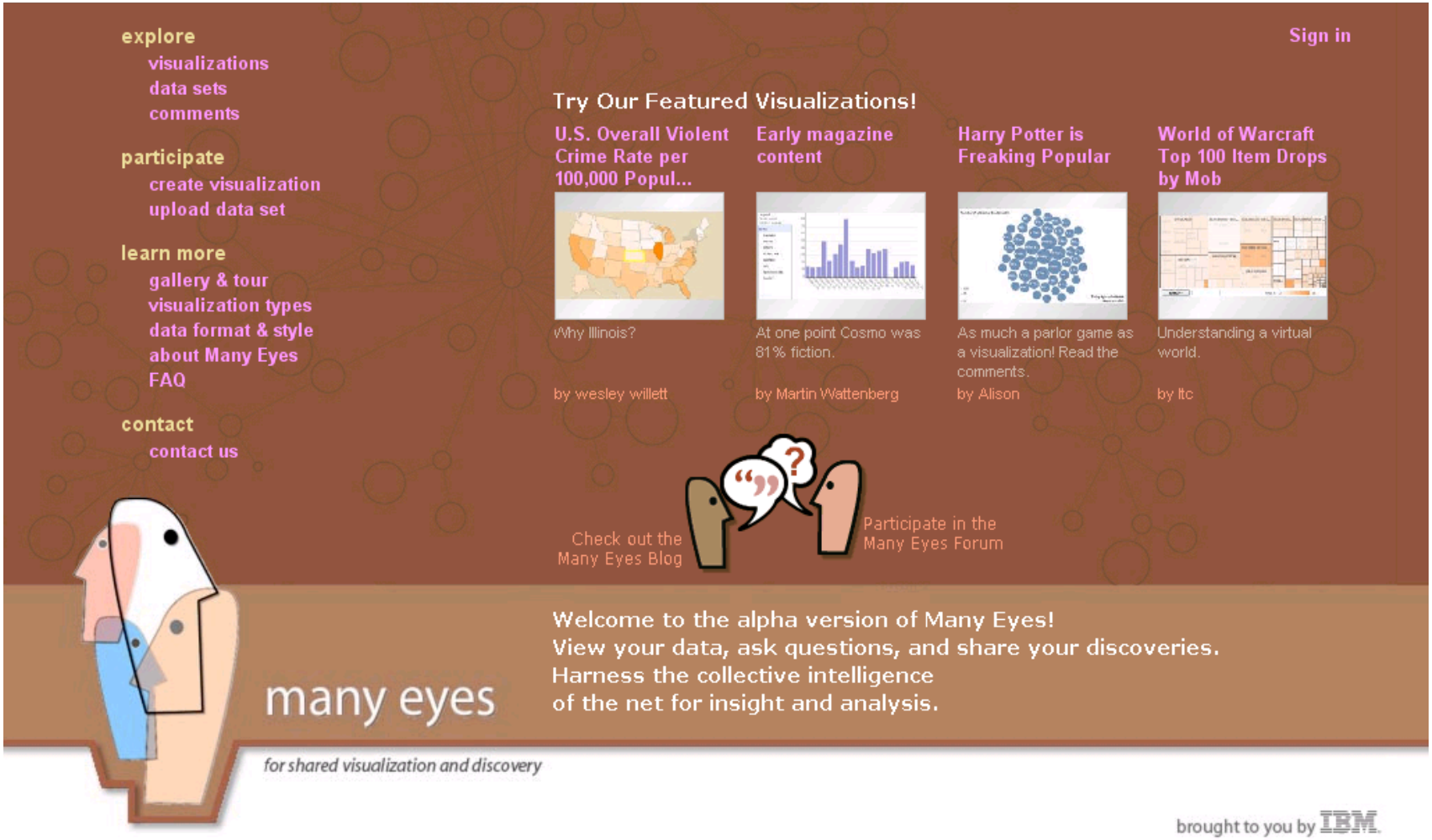
- **1-D Linear** — Document Lens, SeeSoft, Info Mural
- **2-D Map** — GIS, ArcView, PageMaker, Medical imagery
- **3-D World** — CAD, Medical, Molecules, Architecture

InfoViz

- **Multi-Var** — Spotfire, Tableau, GGobi, TableLens, ParCoords,
- **Temporal** — LifeLines, TimeSearcher, Palantir, DataMontage
- **Tree** — Cone/Cam/Hyperbolic, SpaceTree, Treemap
- **Network** — Pajek, JUNG, UCINet, SocialAction, NodeXL

HCIL

ManyEyes: A web sharing platform



http://services.alphaworks.ibm.com/manyeyes/app

HCIL

Business takes action

- General Dynamics **buys** MayaViz
- Agilent **buys** GeneSpring
- Google **buys** Gapminder
- Oracle **buys** (Hyperion **buys** Xcelsius)
- Microsoft **buys** Proclarity
- InfoBuilders **buys** Advizor Solutions
- SAP **buys** (Business Objects **buys** Infomersion & Inxight & Crystal Reports)
- IBM **buys** (Cognos **buys** Celequest) & ILOG
- TIBCO **buys** Spotfire

HCIL

Treemap: view large trees with node values

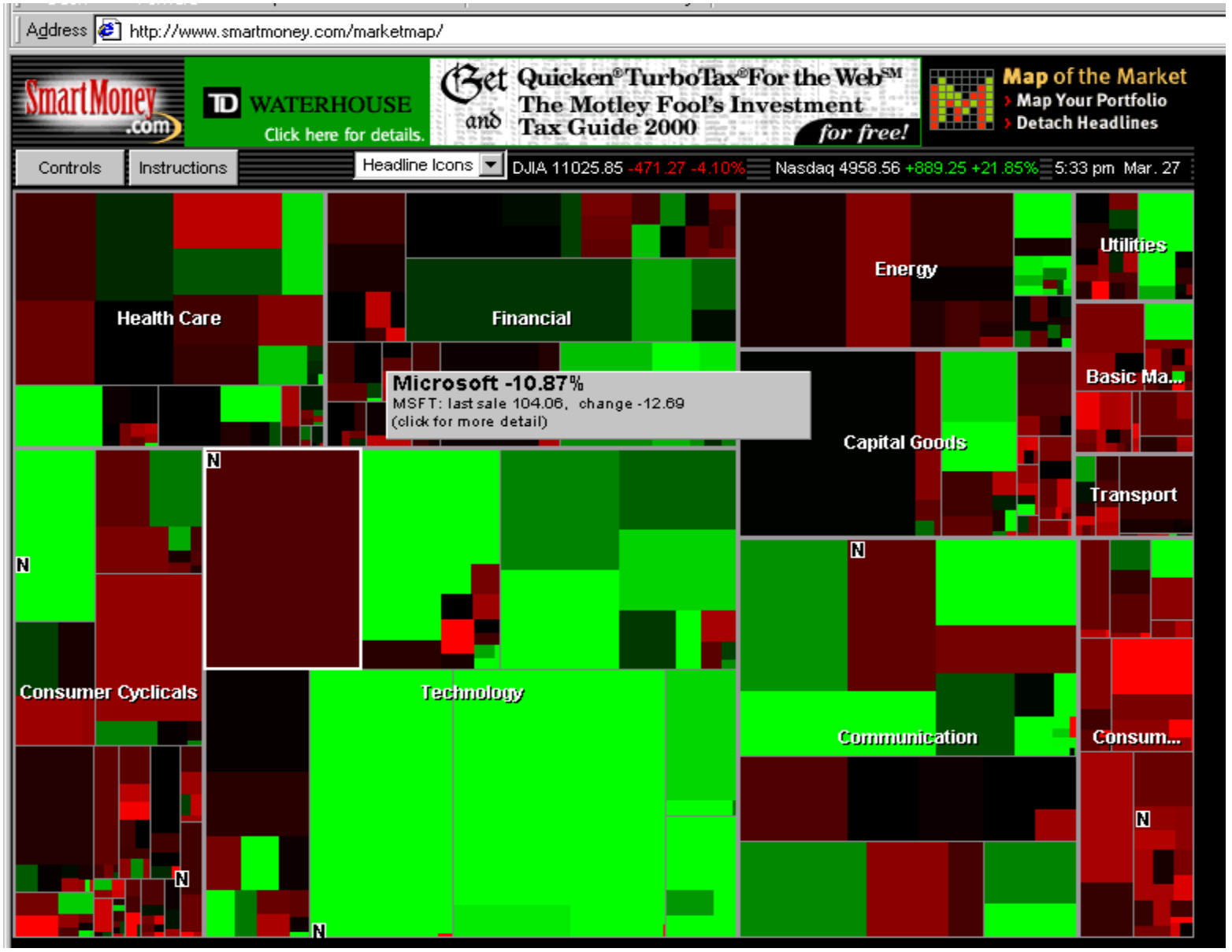
- \+ Space filling
- \+ Space limited
- \+ Color coding
- \+ Size coding
- \- Requires learning

TreeViz (Mac, Johnson, 1992)
NBA-Tree(Sun, Turo, 1993)
Winsurfer (Teittinen, 1996)
Diskmapper (Windows, Micrologic)
SequoiaView, Panopticon,
HiveGroup, Solvern
Treemap4 (UMd, 2004)

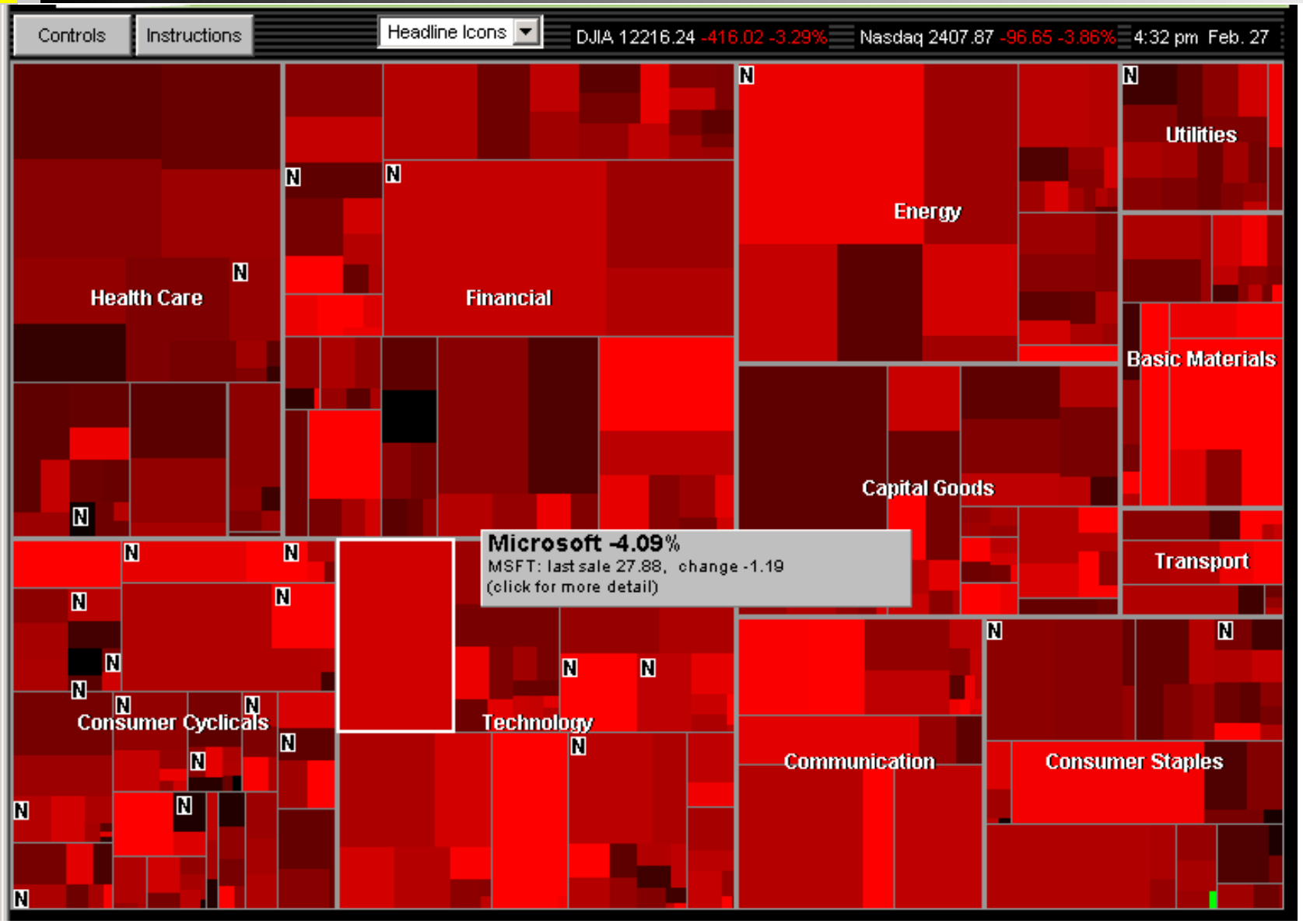
(Shneiderman, *ACM Trans. on Graphics*, 1992 & 2003)

HCIL

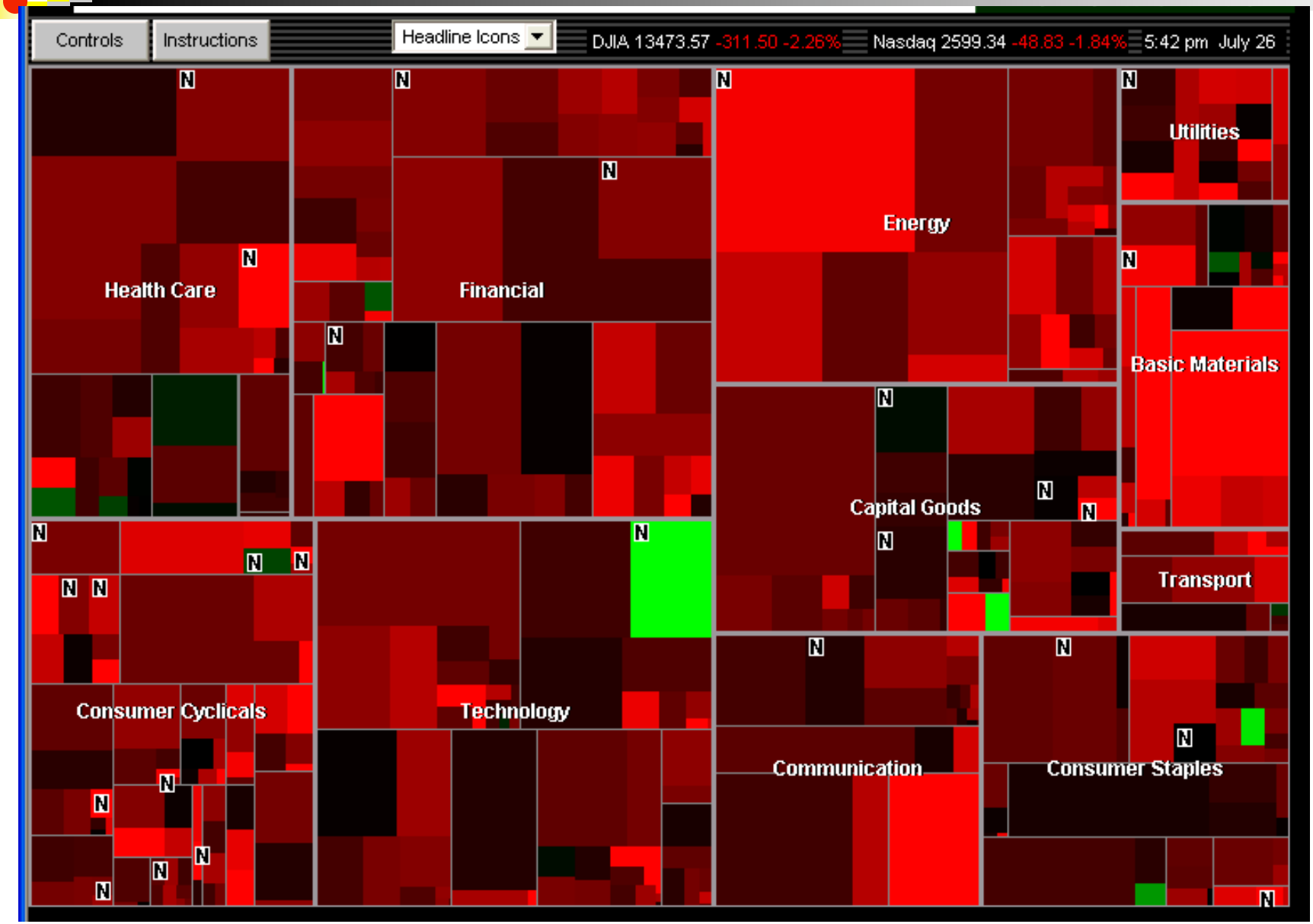
Treemap: Stock market, clustered by industry



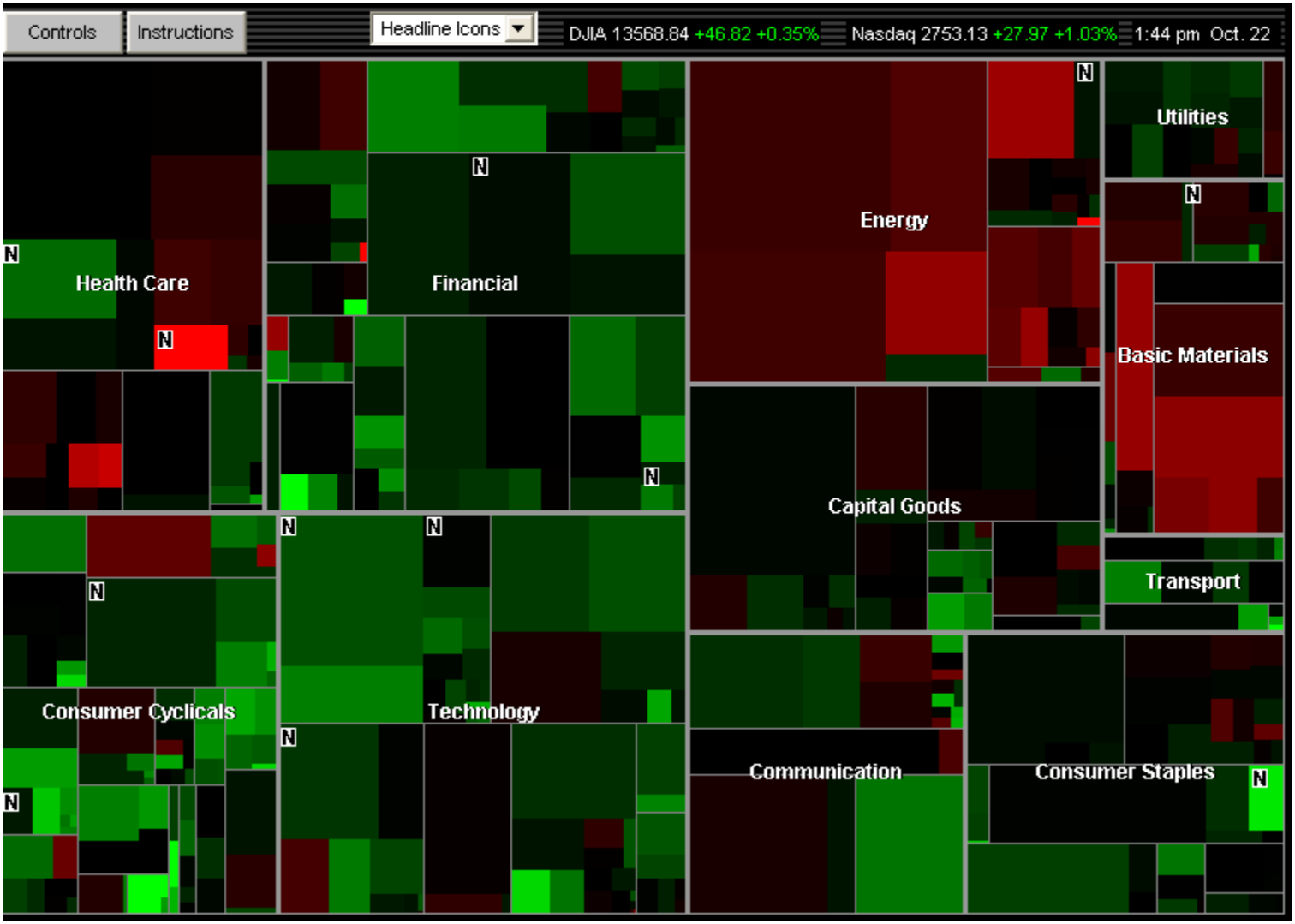
Market falls steeply Feb 27, 2007, with one exception



Market falls 311 points July 26, 2007, with a few exceptions

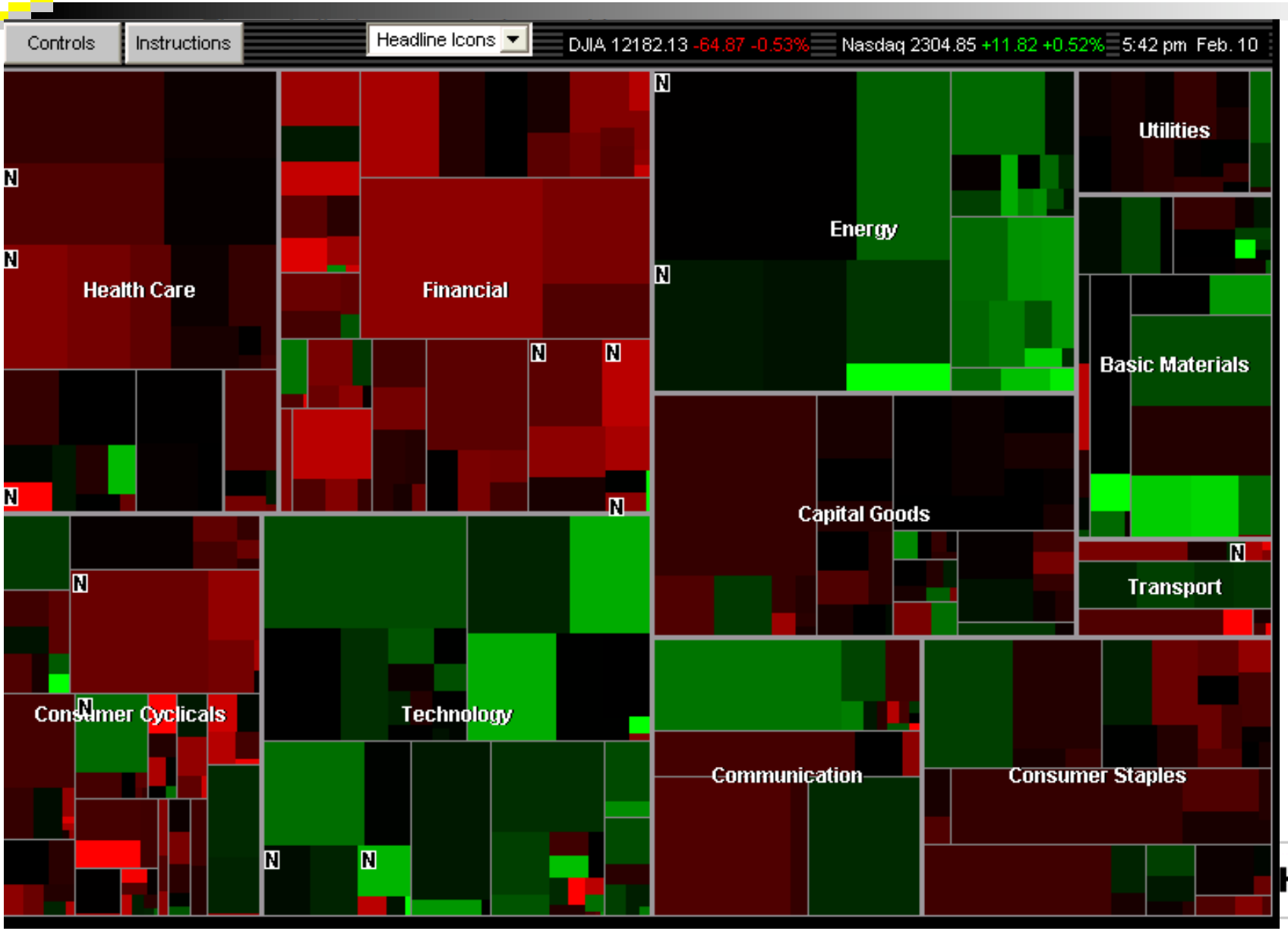


Market mixed, October 22, 2007,
Energy & Basic Material are down

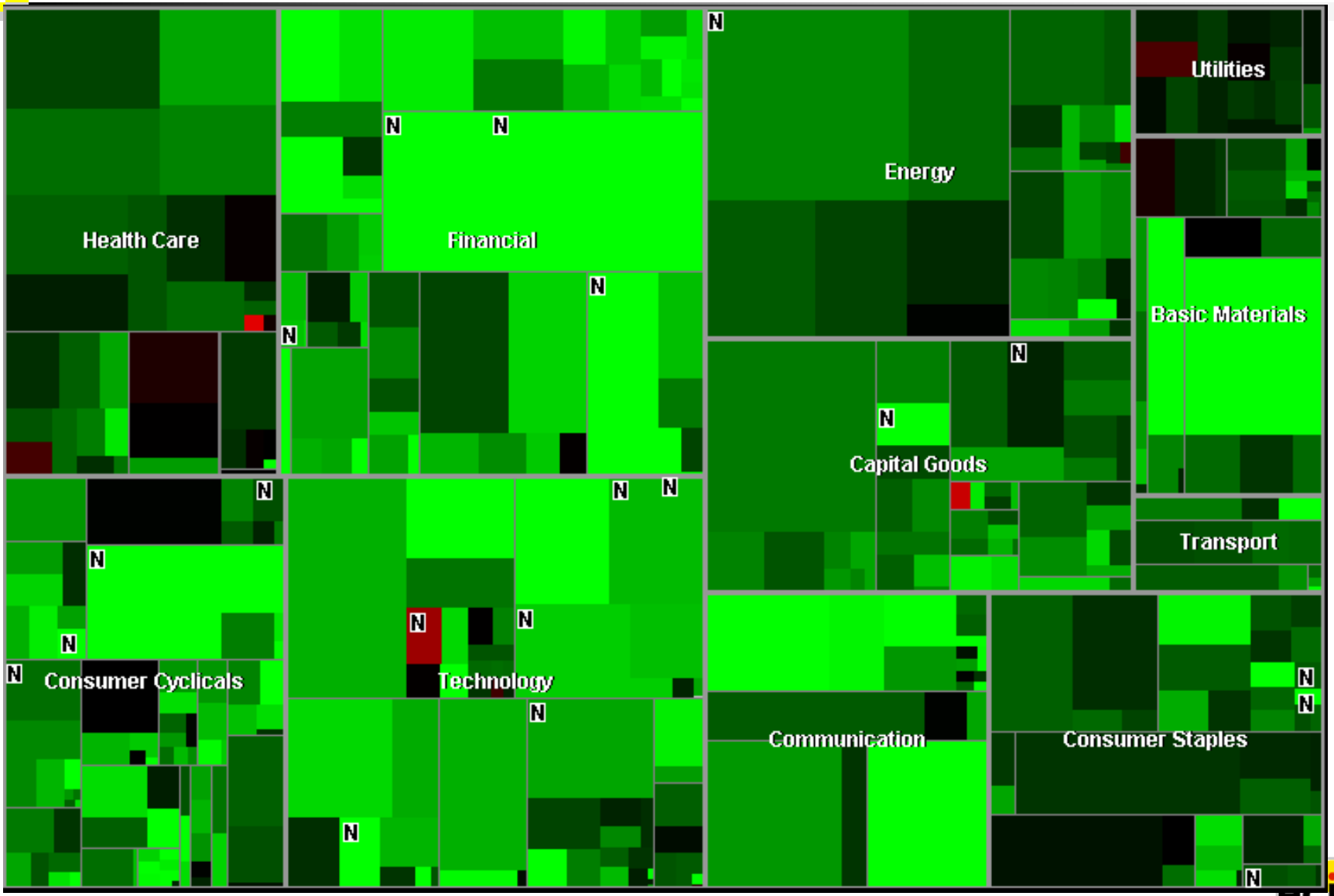


Market mixed, February 8, 2008

Energy & Technology up, Financial & Health Care down

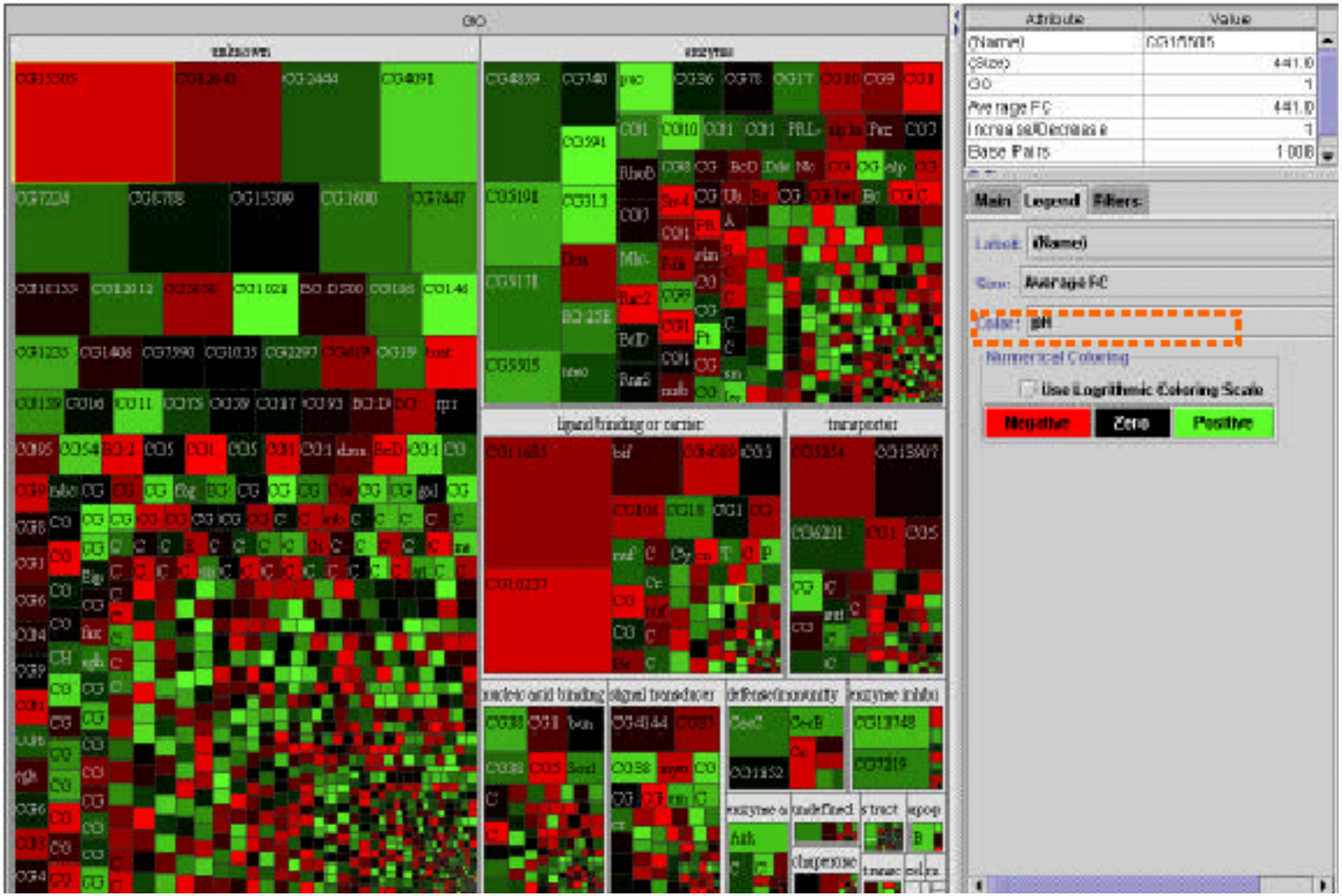


Market rises 319 points, November 13, 2007, with 5 exceptions



Treemap: Newsmap

Treemap: Gene Ontology

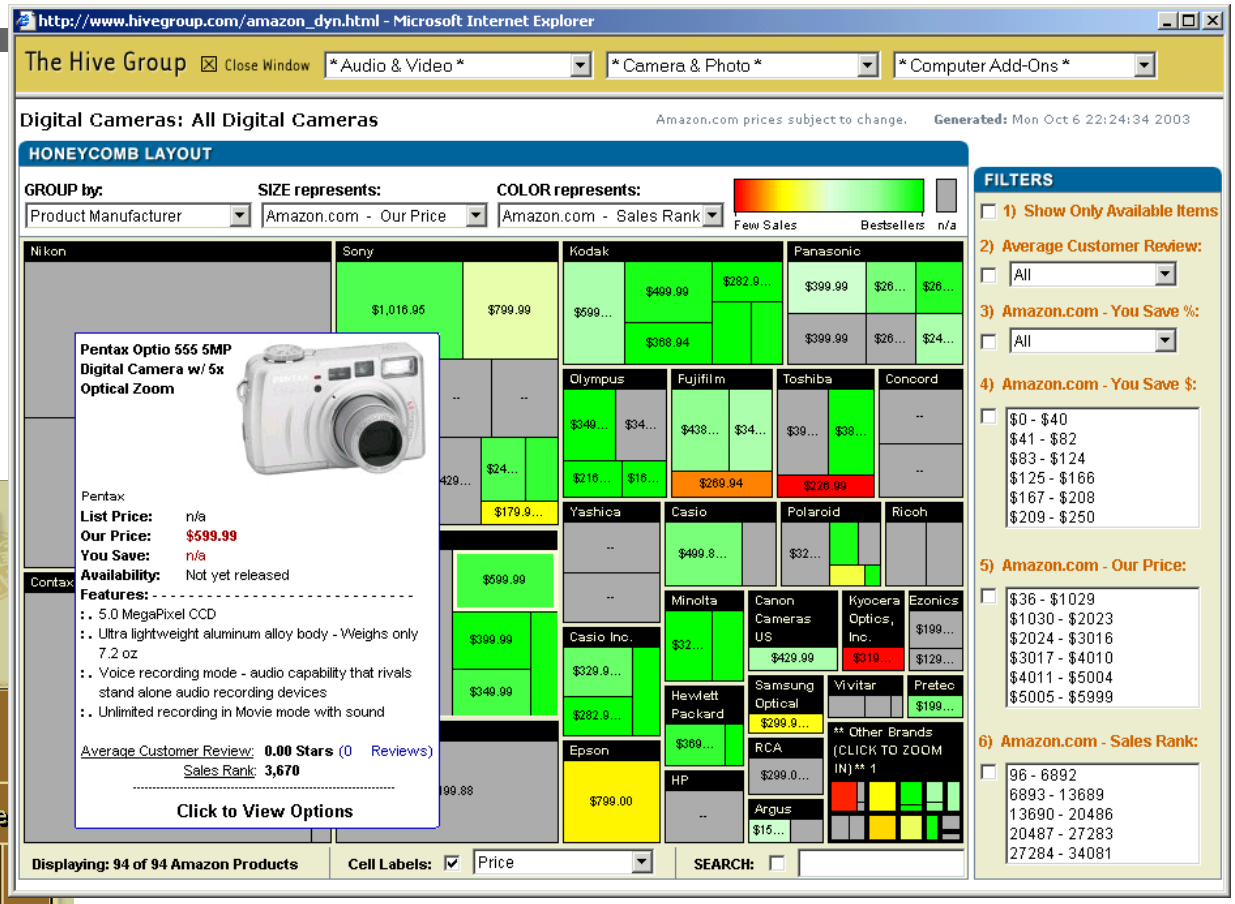


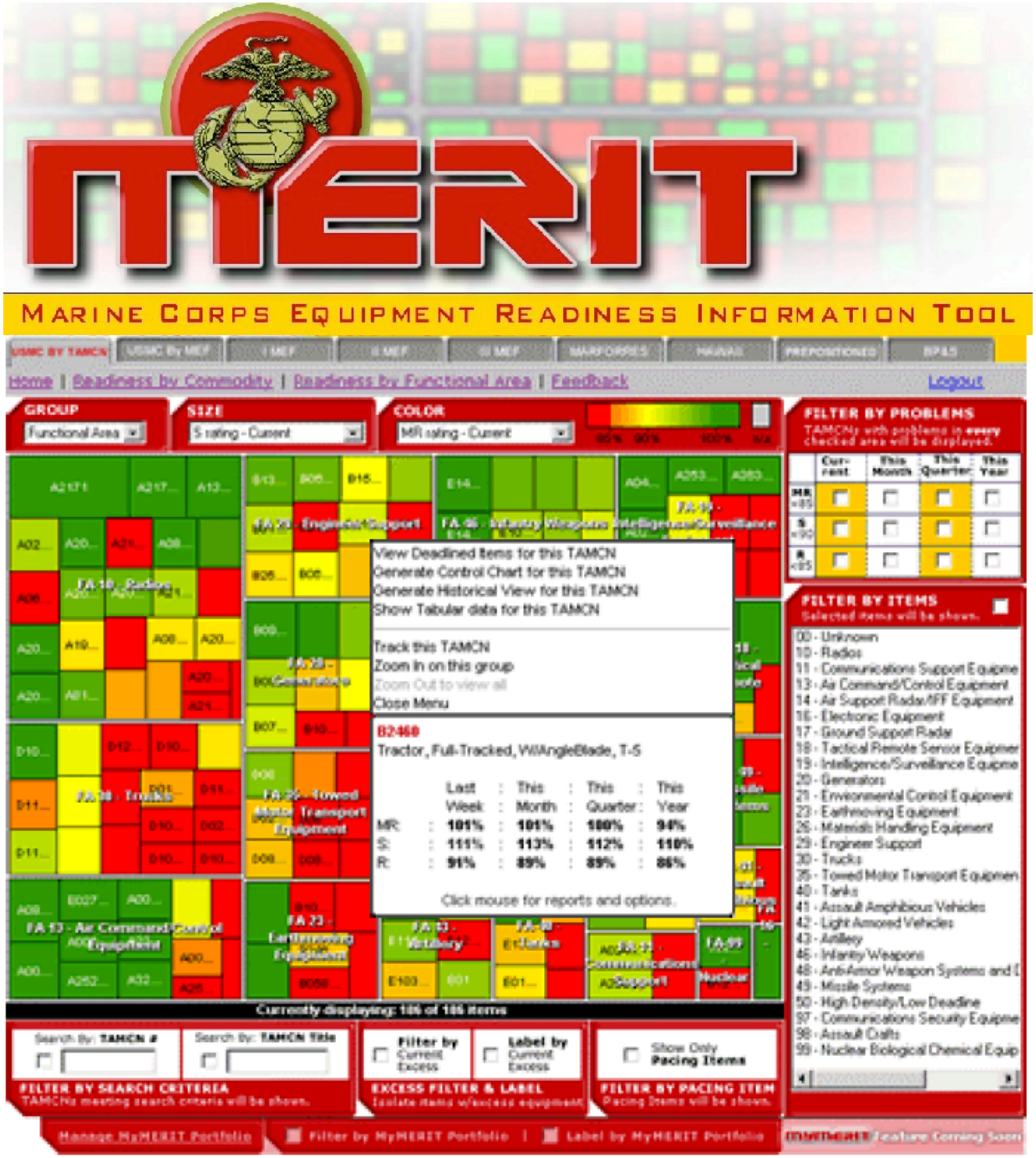
http://www.cs.umd.edu/hcil/treemap/

HCIL

Treemap: Product catalogs

www.hivegroup.com

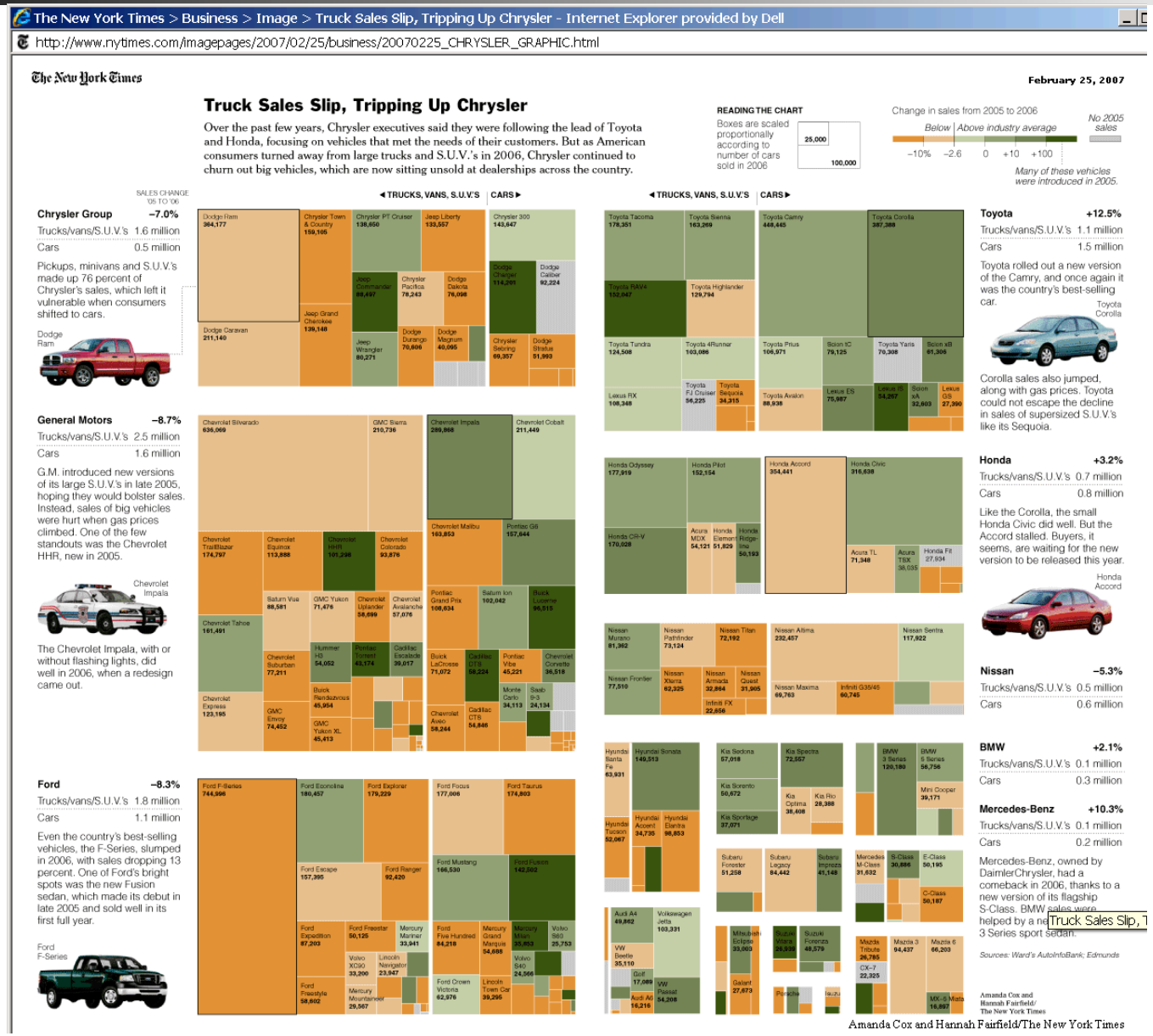



MERIT
MARINE CORPS EQUIPMENT READINESS INFORMATION TOOL
Home | Readiness by Commodity | Readiness by Functional Area | Feedback
Logout
GROUP
Functional Area
SIZE
S rating - Current
COLOR
MR rating - Current
FILTER BY PROBLEMS
TAMCNs with problems in every checked area will be displayed.
FILTER BY ITEMS
Selected items will be shown.
00 - Unknown
10 - Radios
11 - Communications Support Equipme
13 - Air Command/Control Equipment
14 - Air Support Radar/IFF Equipment
16 - Electronic Equipment
17 - Ground Support Radar
18 - Tactical Remote Sensor Equipmer
19 - Intelligence/Surveillance Equipme
20 - Generators
21 - Environmental Control Equipment
23 - Earthmoving Equipment
25 - Materials Handling Equipment
29 - Engineer Support
30 - Trucks
35 - Towed Motor Transport Equipmen
40 - Tanks
41 - Assault Amphibious Vehicles
42 - Light Armored Vehicles
43 - Artillery
46 - Infantry Weapons
48 - Anti-Armor Weapon Systems and
49 - Missile Systems
50 - High Density/Low Deadline
97 - Communications Security Equipme
98 - Assault Crafts
99 - Nuclear Biological Chemical Equip
View Deadlined Items for this TAMCN
Generate Control Chart for this TAMCN
Generate Historical View for this TAMCN
Show Tabular data for this TAMCN
Track this TAMCN
Zoom In on this group
Zoom Out to view all
Close Menu
B2460
Tractor, Full-Tracked, W/AngleBlade, T-5
Last Week
This Month
This Quarter
This Year
MR: 101% 101% 100% 94%
S: 111% 113% 112% 110%
R: 91% 89% 89% 86%
Click mouse for reports and options.
Currently displaying: 186 of 186 items
Search By: TAMCN #
Search By: TAMCN Title
FILTER BY SEARCH CRITERIA
TAMCNs meeting search criteria will be shown.
Filter by Current Excess
Label by Current Excess
EXCESS FILTER & LABEL
Isolate items w/excess equipment
Show Only Pacing Items
FILTER BY PACING ITEM
Pacing items will be shown.
Manage MyMERIT Portfolio
Filter by MyMERIT Portfolio
Label by MyMERIT Portfolio
MyMERIT Feature Coming Soon

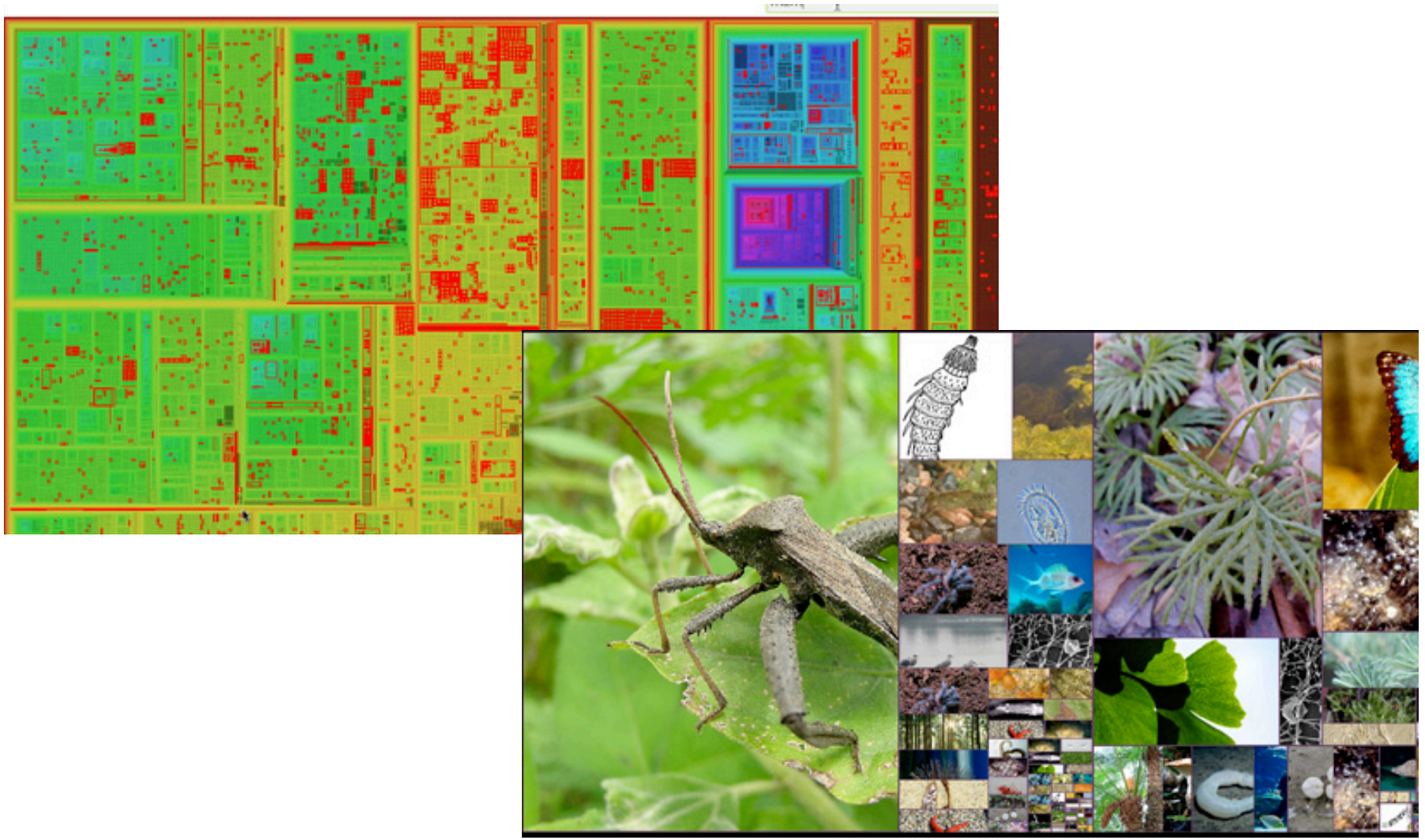
UNITED STATES MARINE CORPS
SEMPER FIDELIS

MARINE CORPS EQUIPMENT READINESS INFORMATION TOOL
254
MERIT
AWARDS:
AFEI EXCELLENCE IN INTEGRATION
GOVERNMENT COMPUTER NEWS
NAVY EGOV
DOD CIO
DNE WATCH

Treemap: NY Times – Car&Truck Sales

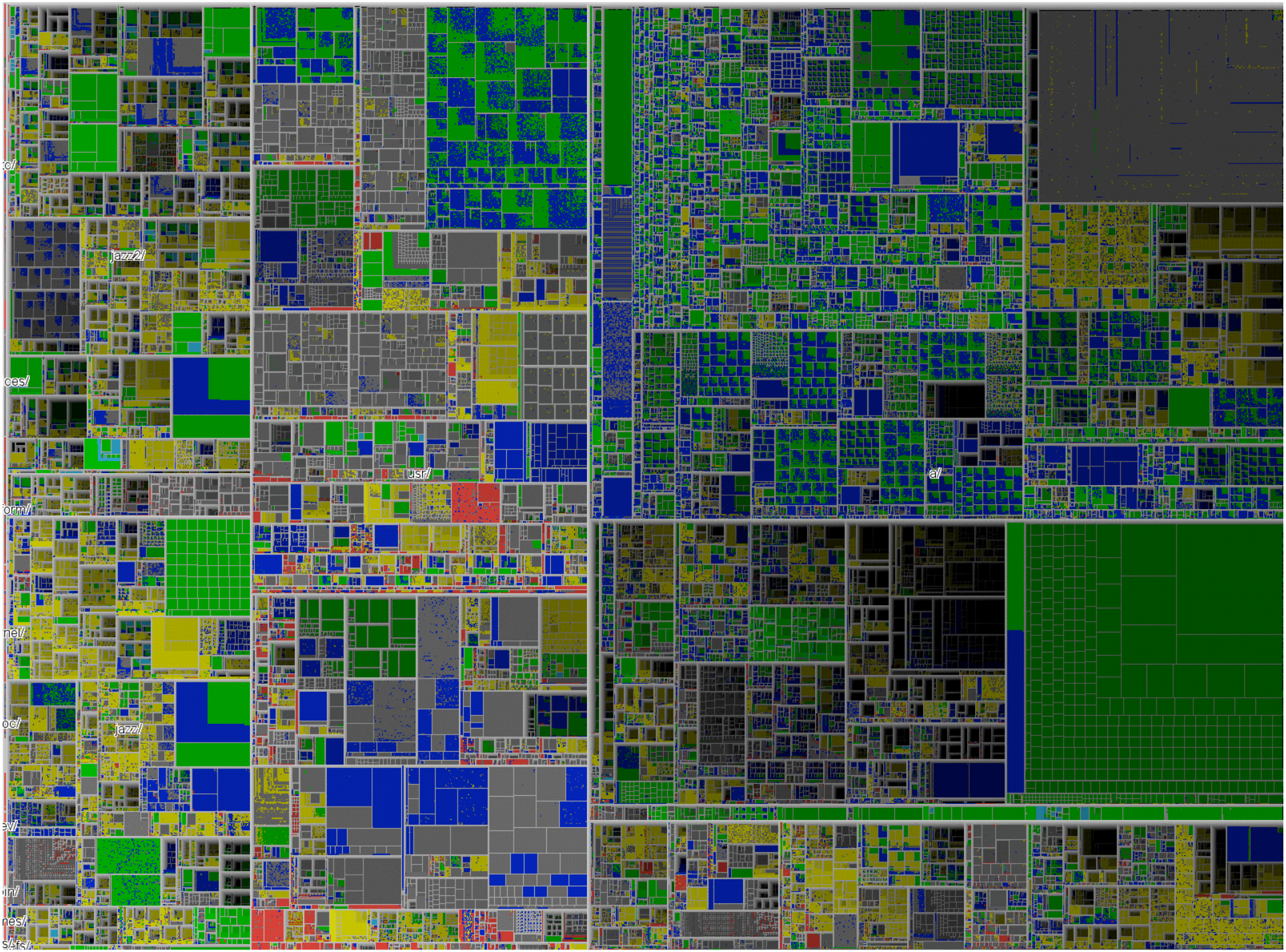


Treemap: Tree of Life (1.8M species)



http://eol.org

HCIL

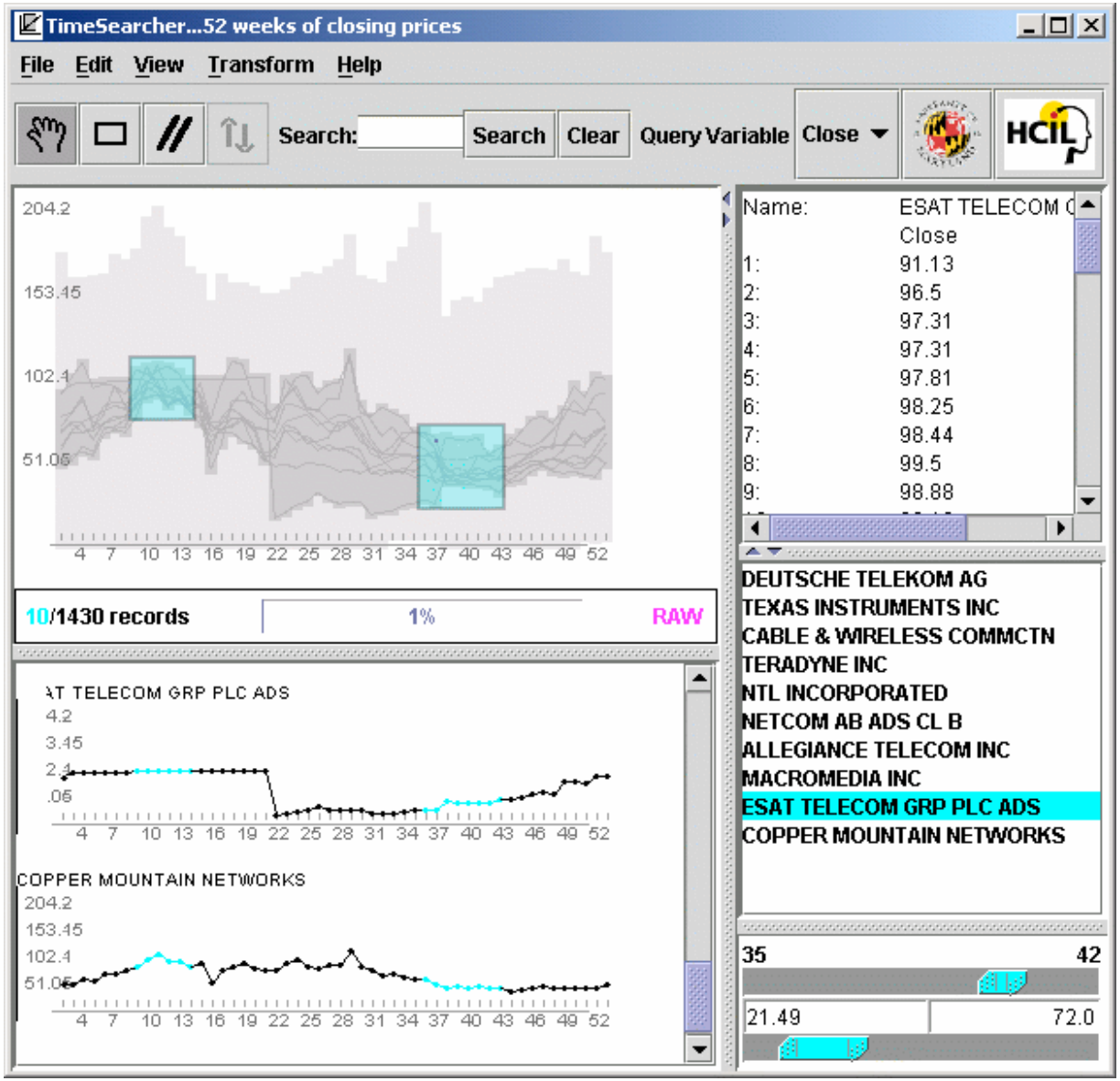


Speed: 5.917fps/0.169spf 971061 items

.root.xml.gz/a/kronos/raid/ 'a/users/hsh/src/ ld/build.sh

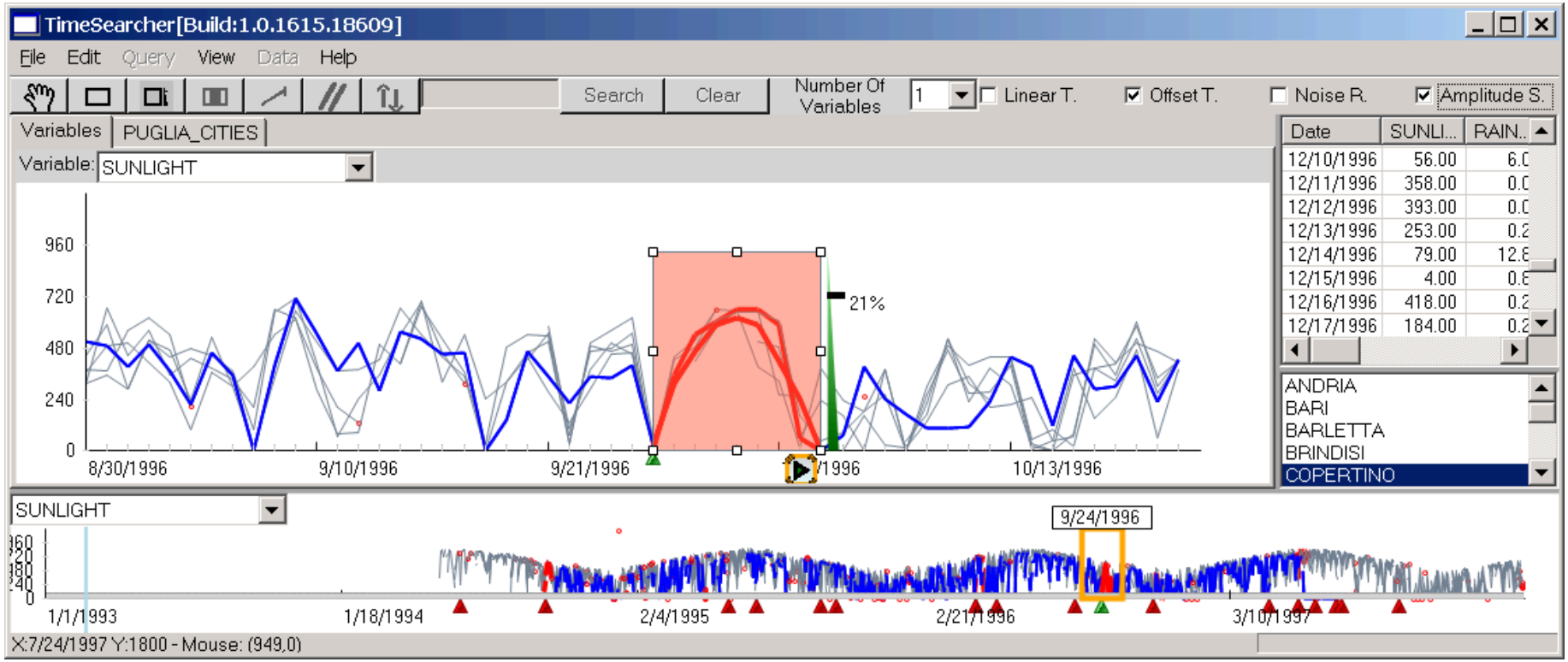
Temporal Data: TimeSearcher 1.3

- Time series
  - Stocks
  - Weather
  - Genes
- User-specified patterns
- Rapid search

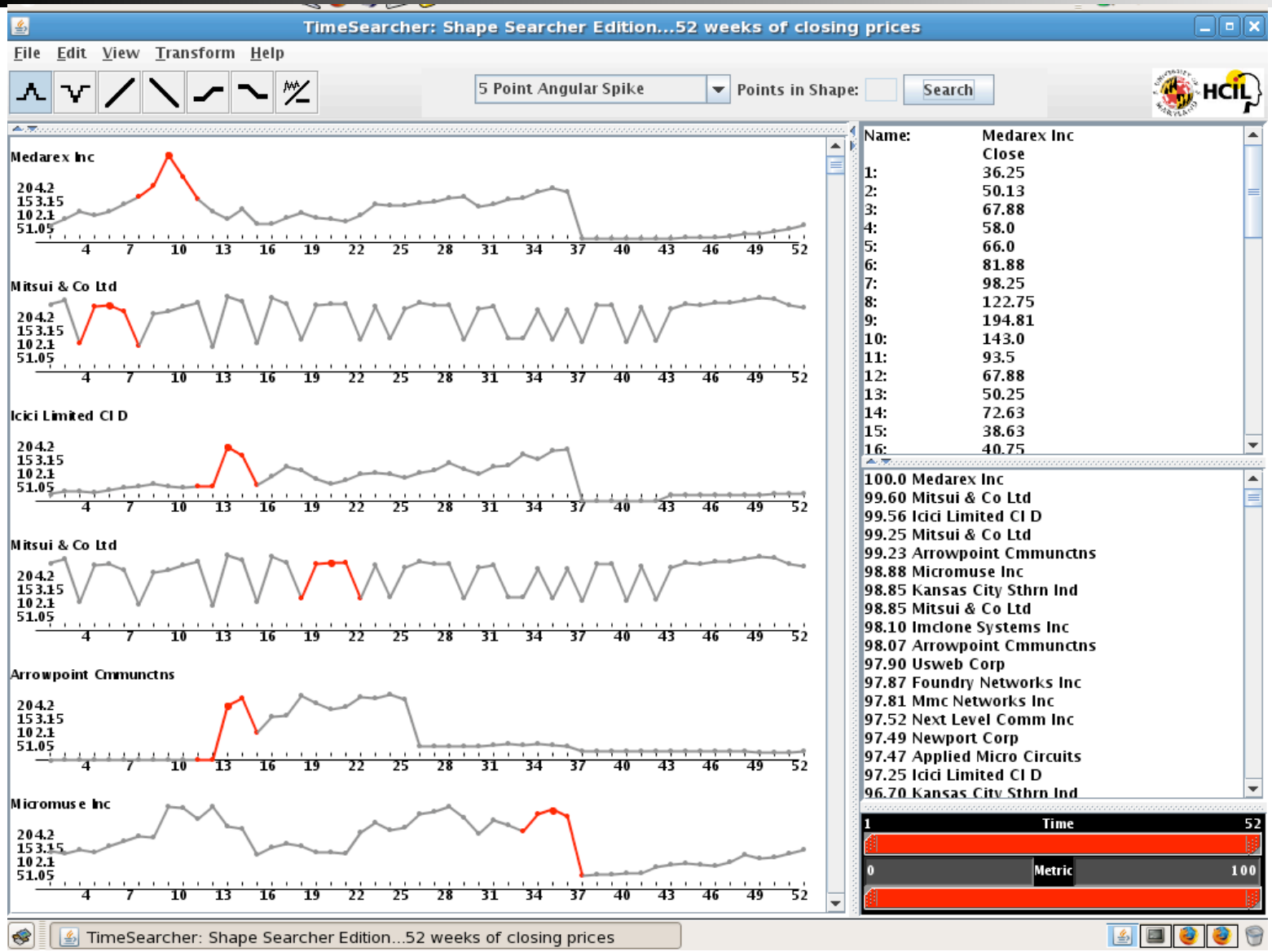


Temporal Data: TimeSearcher 2.0

- Long Time series (>10,000 time points)
- Multiple variables
- Controlled precision in match
  (Linear, offset, noise, amplitude)

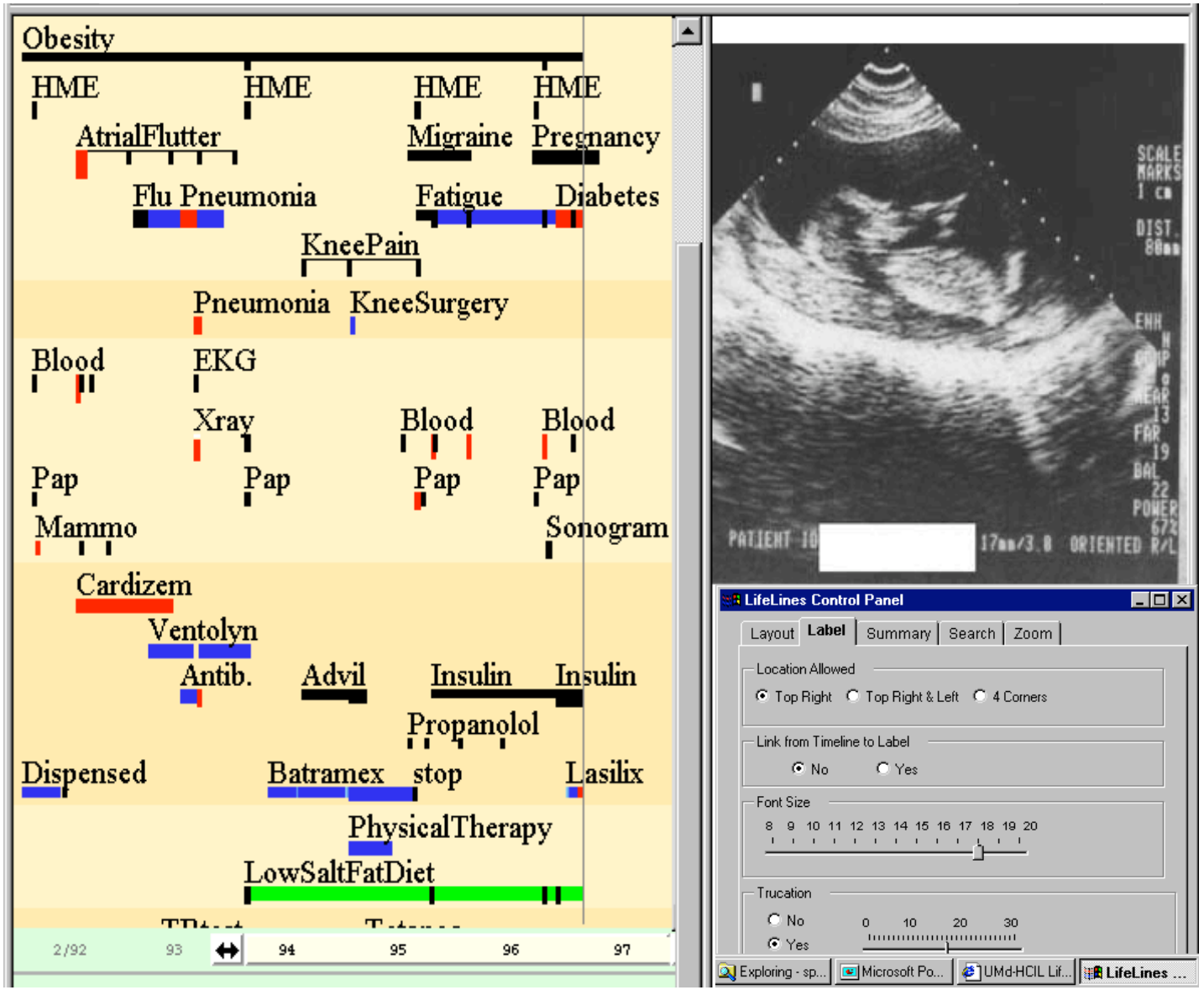


TimeSearcher: Shape Searcher Edition

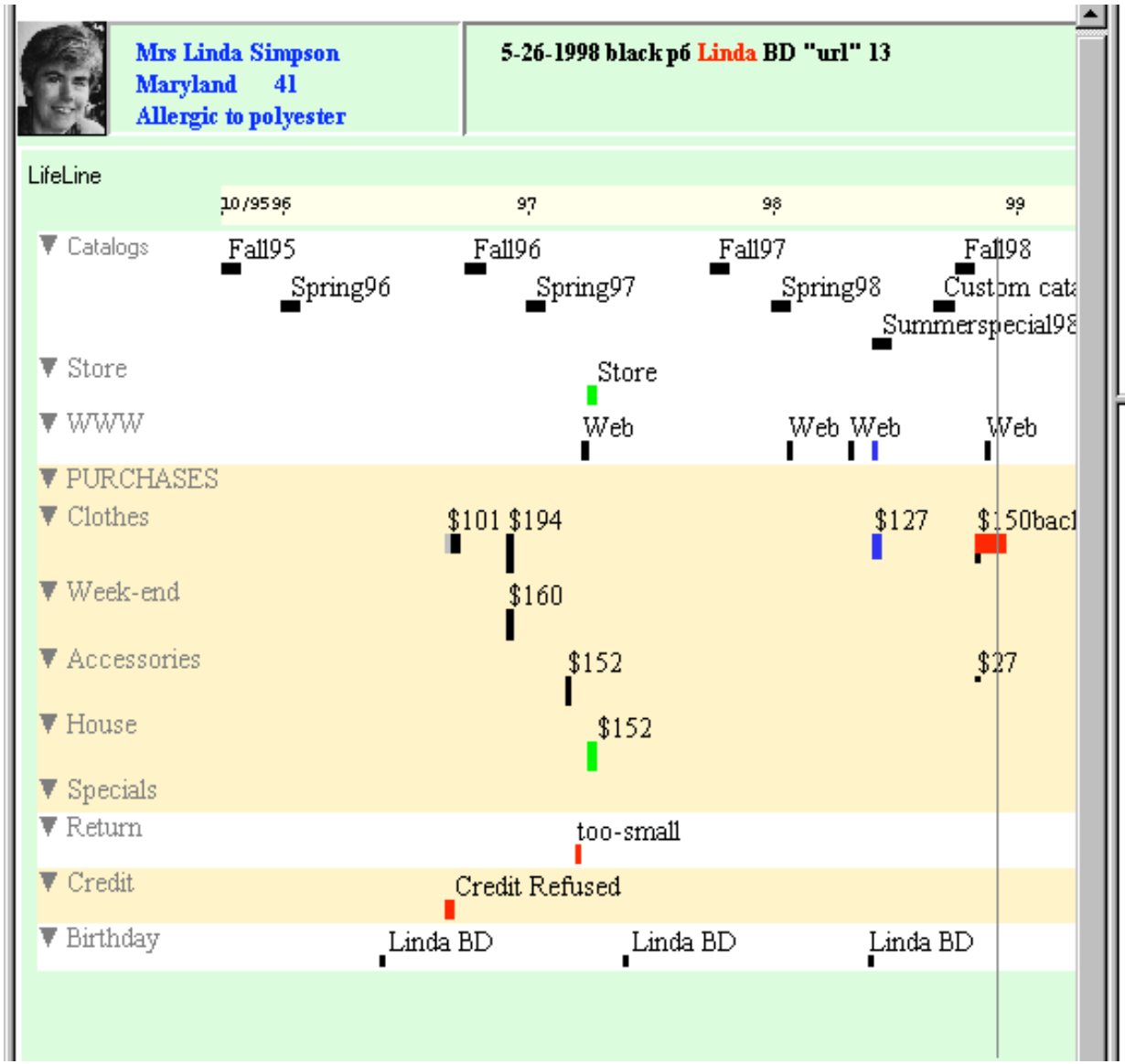


HCIL

LifeLines: Patient Histories



LifeLines: Customer Histories

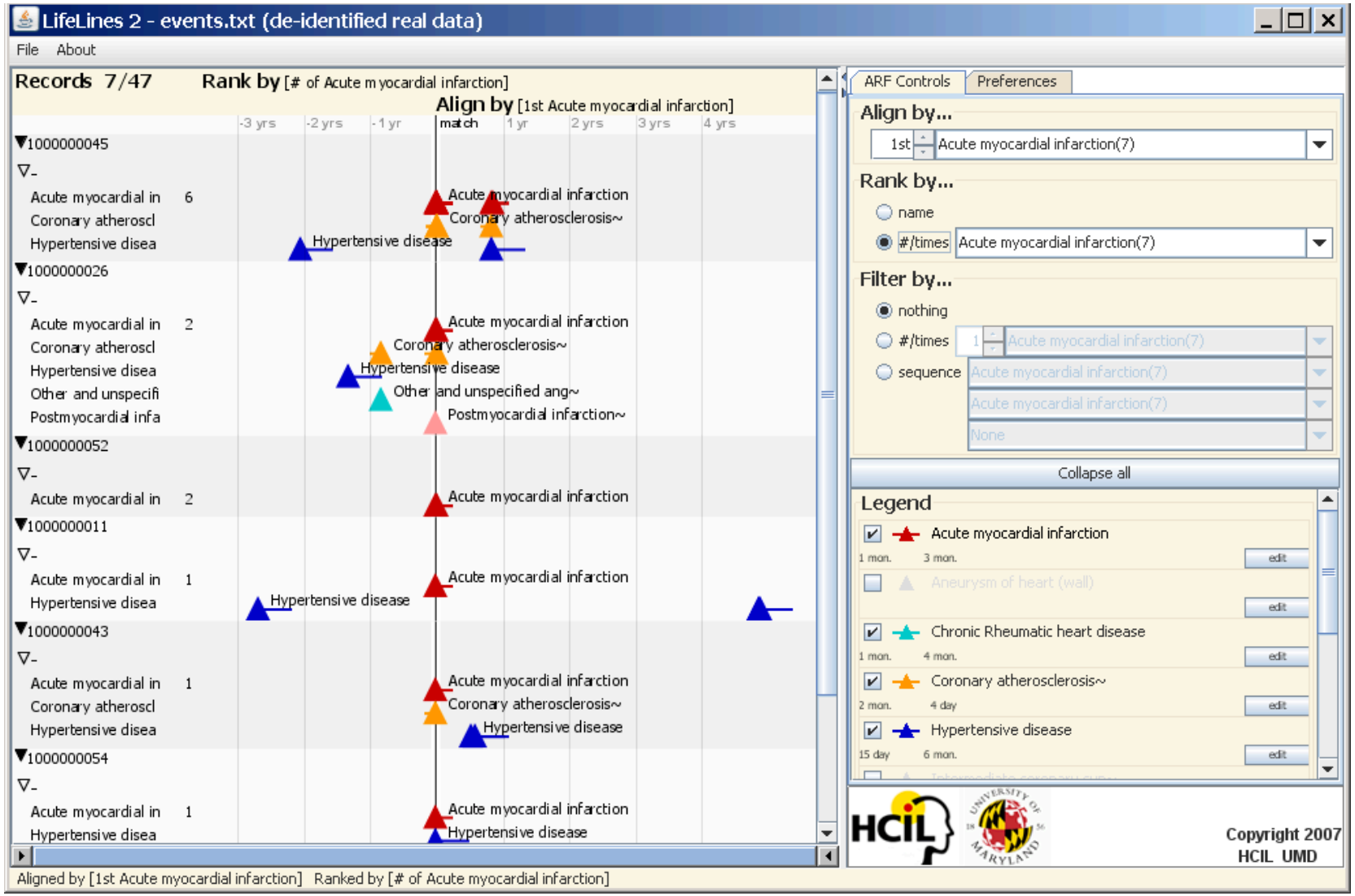


Linda Simpson ORDER#614345 (MOST RECENT)
11-1-1997,$150, Visa card, order by phone

D45F224: Boots - 4 - Black
----> **BACKORDER - estimated shipping 12-15**
D45F236: Gloves - Black - M
A78156: Corduroy pants - Tan - L
157826: Hat - Grey - M

Kids Backcountry hiking boots
Leather boots with super grip soles.

LifeLines2: Align-Rank-Filter



HCIL

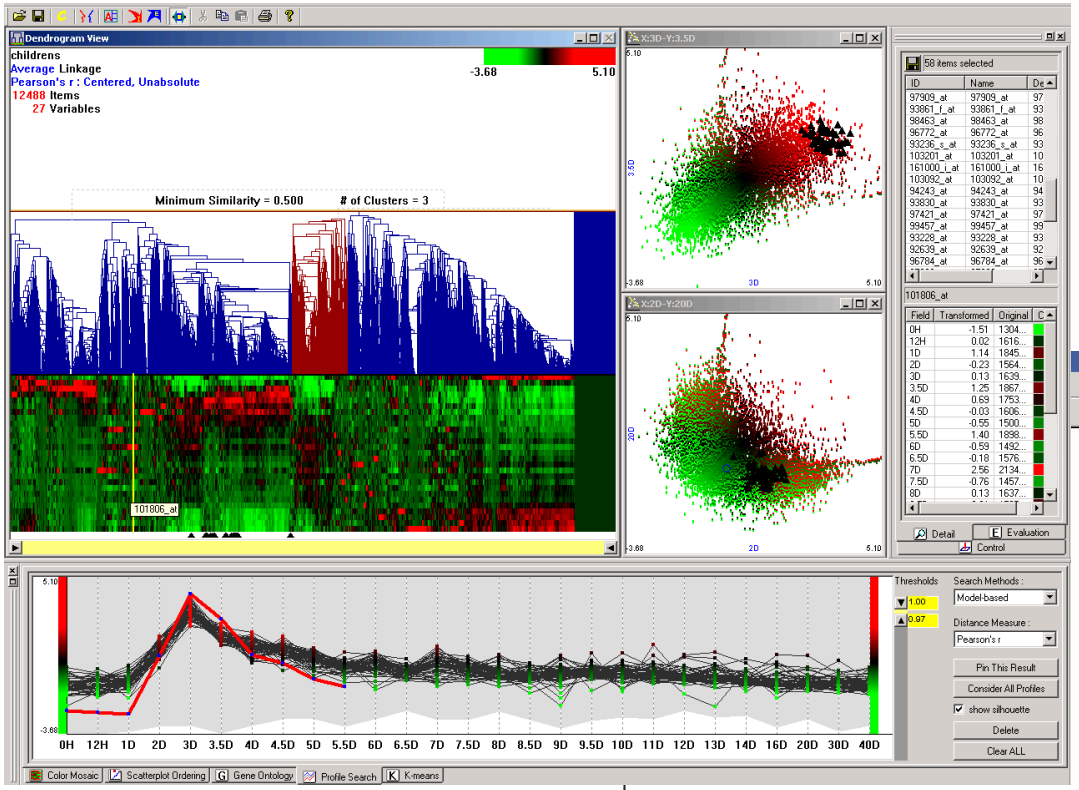
Goal: Find Features in Multi-Var Data

- Clear vision of what the data is
- Clear goal of what you are looking for

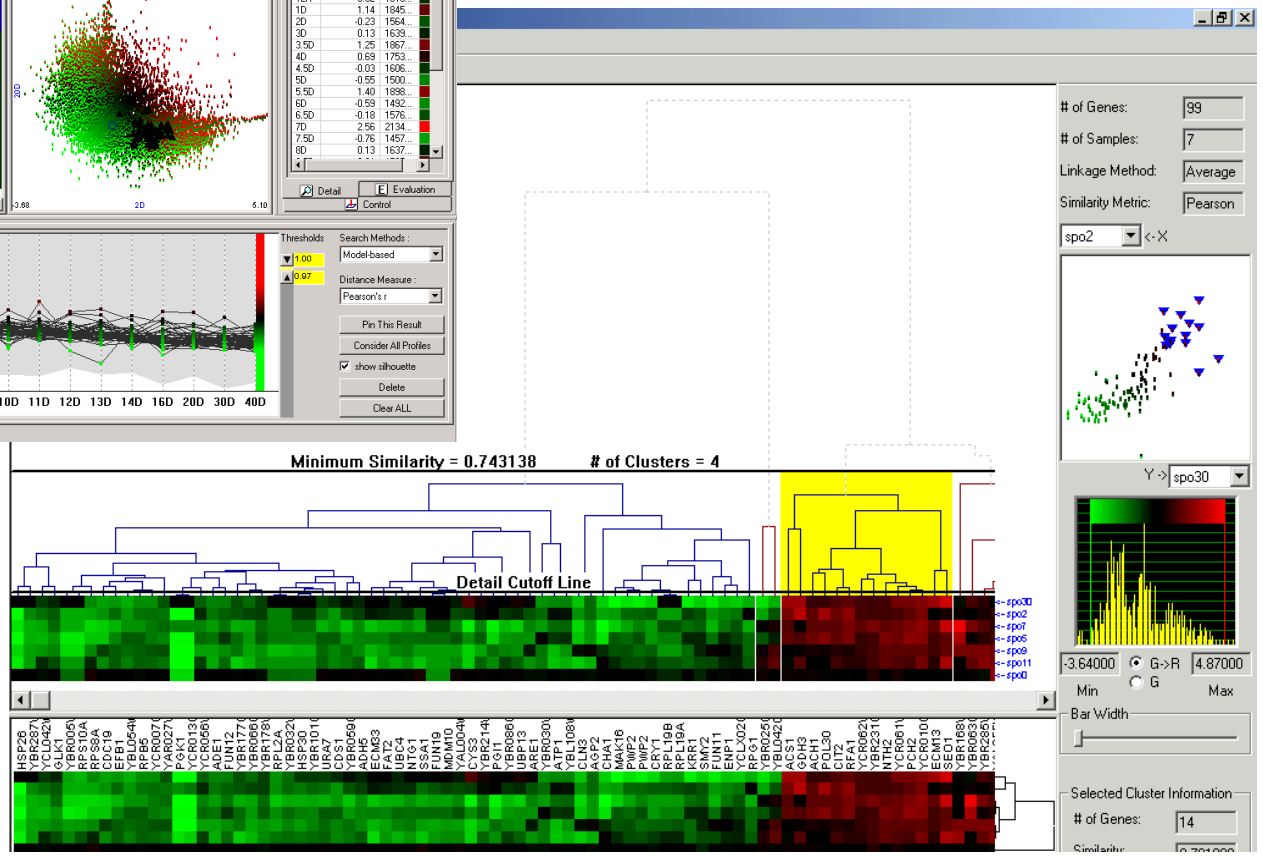
- Systematic strategy for examining all views
- Ranking of views to guide discovery
- Tools to record progress & annotate findings

HCIL

Multi-V: Hierarchical Clustering Explorer



www.cs.umd.edu/hcil/hce/



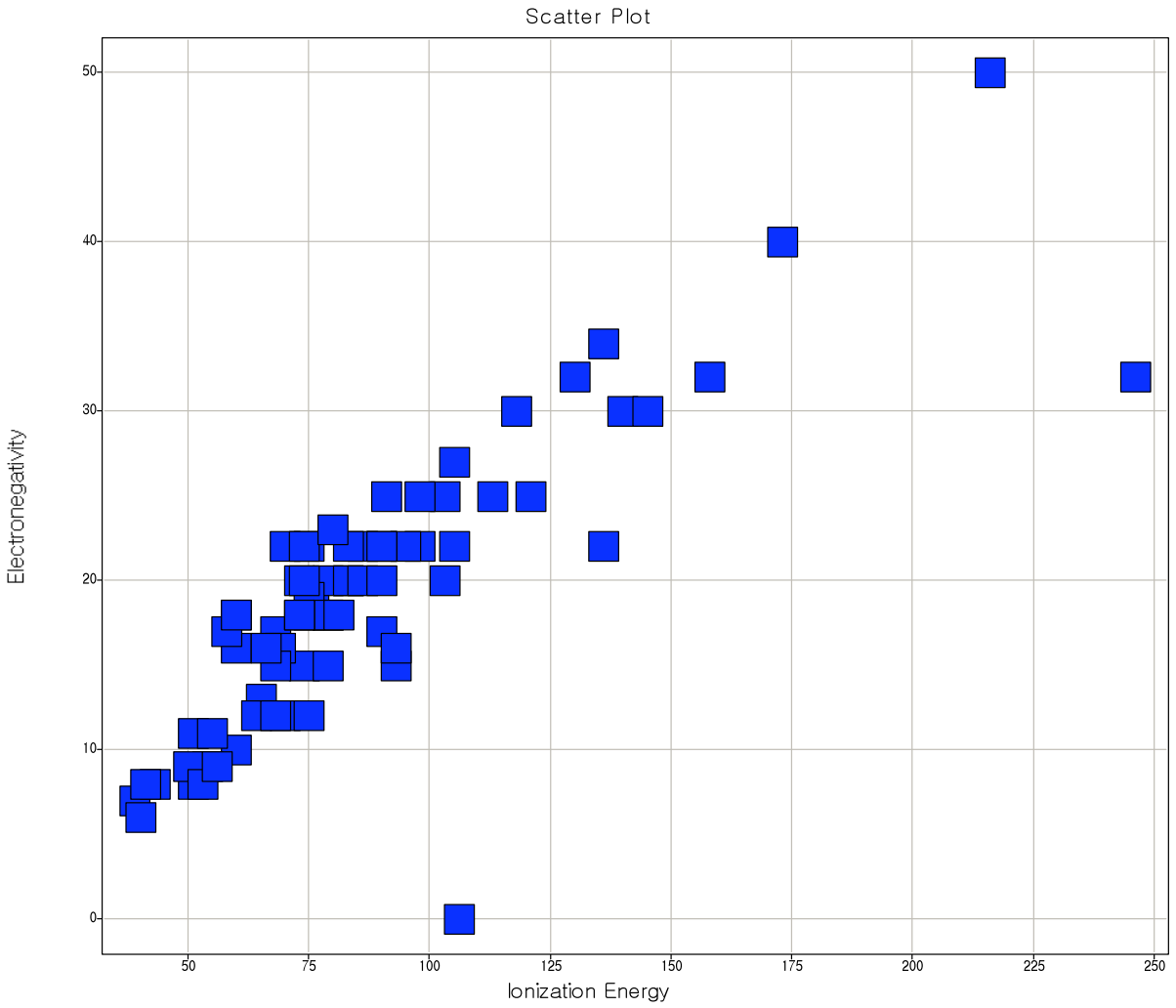
"HCE enabled us to find important clusters that we didn't know about."
- a user

Do you see anything interesting?

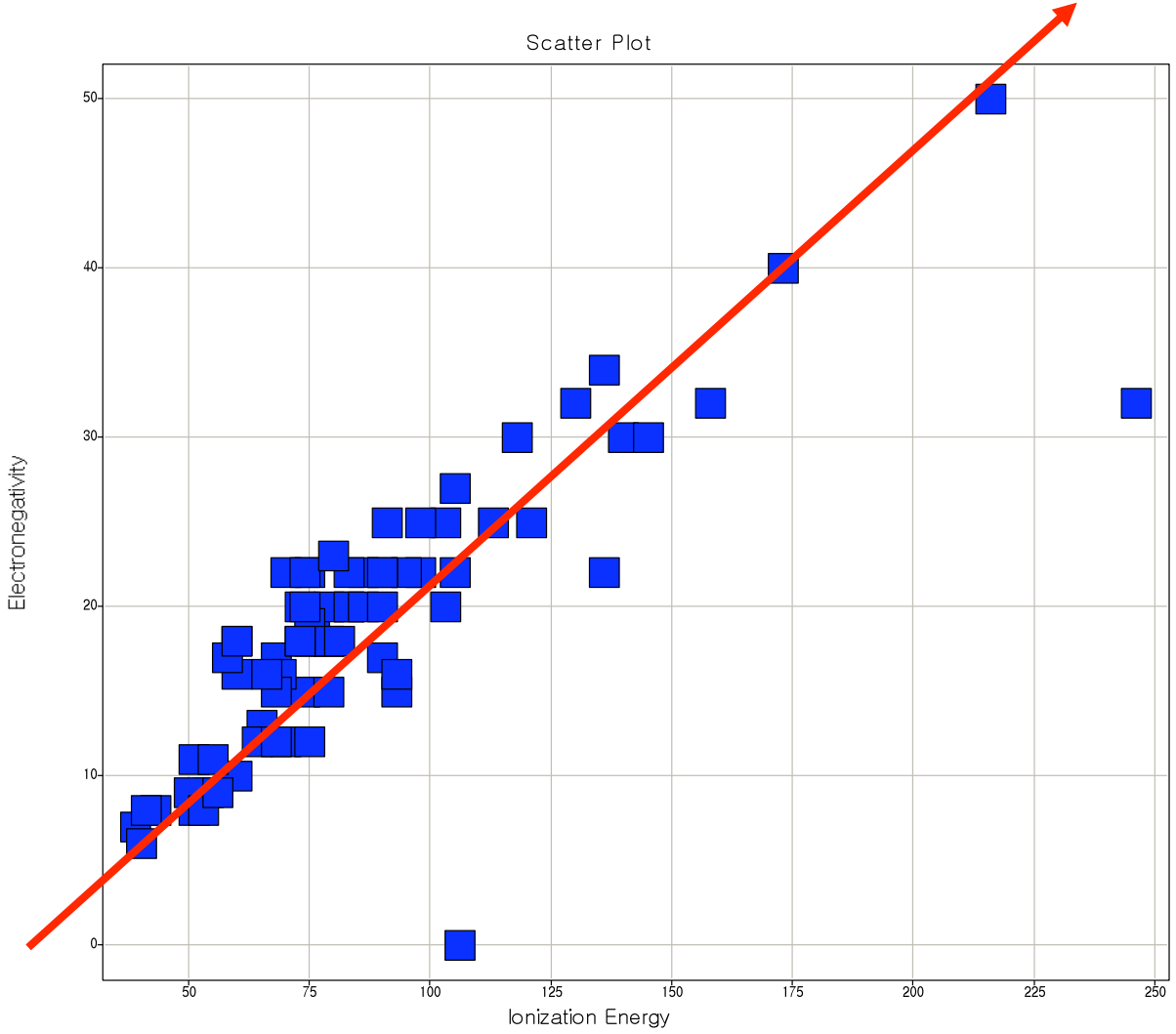
	A	B	C	D	E	F	G	H	I	J	K
1	Element	*P1	*P2	Atomic Num	Atomic Mas	Atomic Radi	Ionic Radius	Ionization E	Electronega	*C1	*C2
2	Ac	140	0	89	227	200	126	51	11	62	56
3	Ag	630	80	47	107	144	129	75	18	124	40
4	Al	750	160	13	27	143	67	60	16	28	25
5	Ar	1050	160	18	39	98	154	158	32	176	51
6	As	870	120	33	75	120	72	98	22	115	33
7	At	990	40	85	210	140	76	95	22	119	22
8	Au	630	40	79	197	144	99	91	25	131	22
9	B	750	200	5	10	85	41	83	20	101	8
10	Ba	80	40	56	137	222	149	51	8	46	56
11	Be	80	200	4	9	112	59	93	15	82	15
12	Bi	870	40	83	209	150	117	73	20	140	27
13	Br	990	120	35	79	114	182	118	30	161	44
14	C	810	200	6	12	77	30	113	25	82	1
15	Ca	80	120	20	40	197	114	60	10	70	51
16	Cd	690	80	48	112	151	109	90	17	113	43
17	Cl	990	160	17	35	100	167	130	32	173	47
18	Co	500	120	27	59	125	83	79	18	120	30
19	Cr	320	120	24	52	128	75	68	17	91	28
20	Cs	20	40	55	132	265	181	39	7	7	56
21	Cu	630	120	29	63	128	87	76	19	118	32
22	F	990	200	9	19	72	119	173	40	39	1
23	Fe	440	120	26	55	126	83	79	18	115	32
24	Fr	20	0	87	223	269	194	40	6	1	56
25	Ga	750	120	31	69	135	76	60	18	89	31
26	Ge	810	120	32	72	122	87	79	20	118	33
27	H	20	240	1	1	32	0	136	22	40	1
28	He	1050	240	2	4	31	93	246	32	1	1
29	Hf	200	40	72	178	159	85	70	12	95	44
30	Hg	690	40	80	200	151	116	103	20	147	27
31	I	990	80	53	126	133	206	105	27	153	44
32	In	750	80	49	114	167	94	58	17	93	42
33	Ir	500	40	77	192	136	82	90	22	116	25
34	K	20	120	19	39	227	152	43	8	37	56
35	Kr	1050	120	36	83	112	169	140	30	163	47

periodic

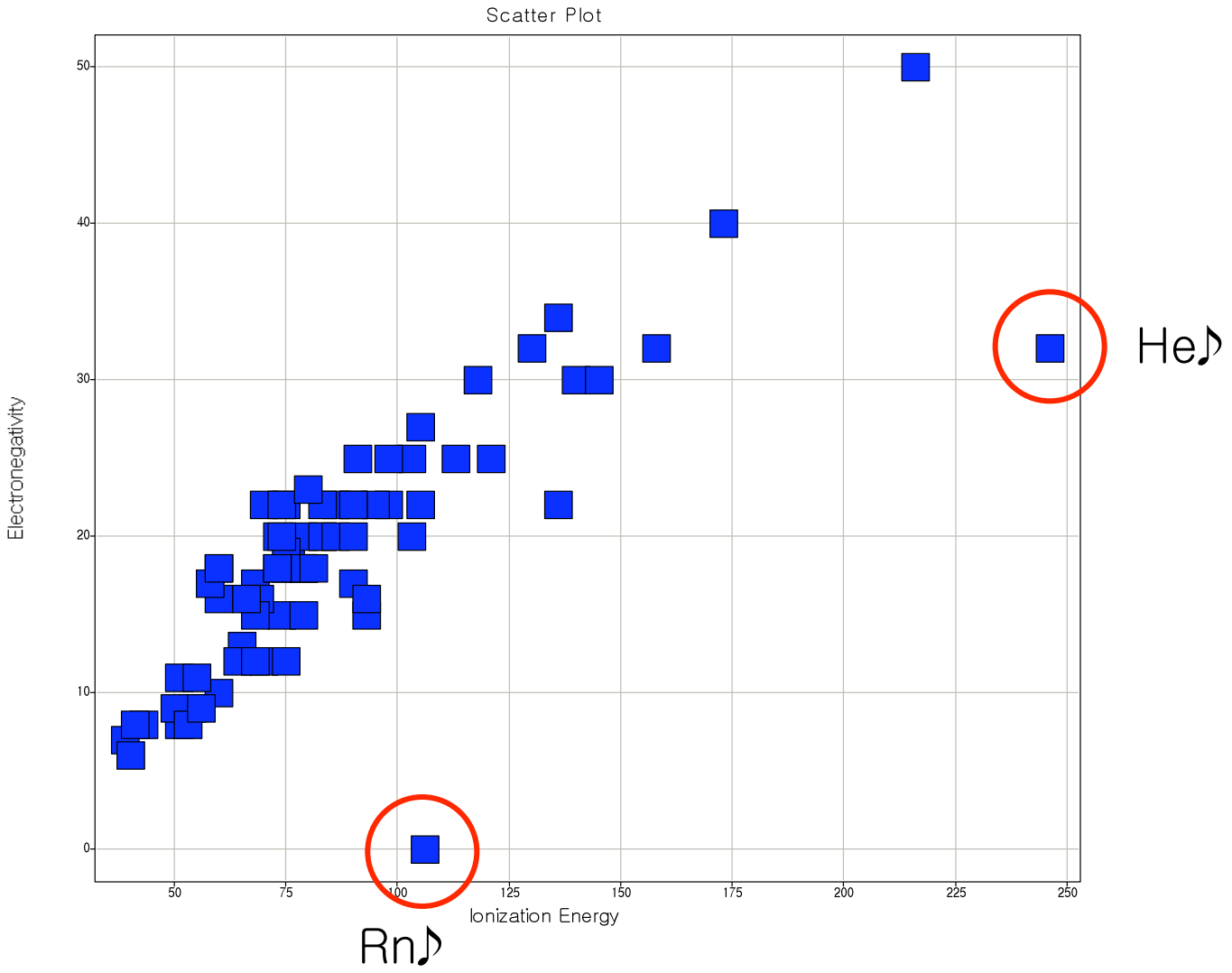
What features stand out?



Correlation…What else?



… and Outliers

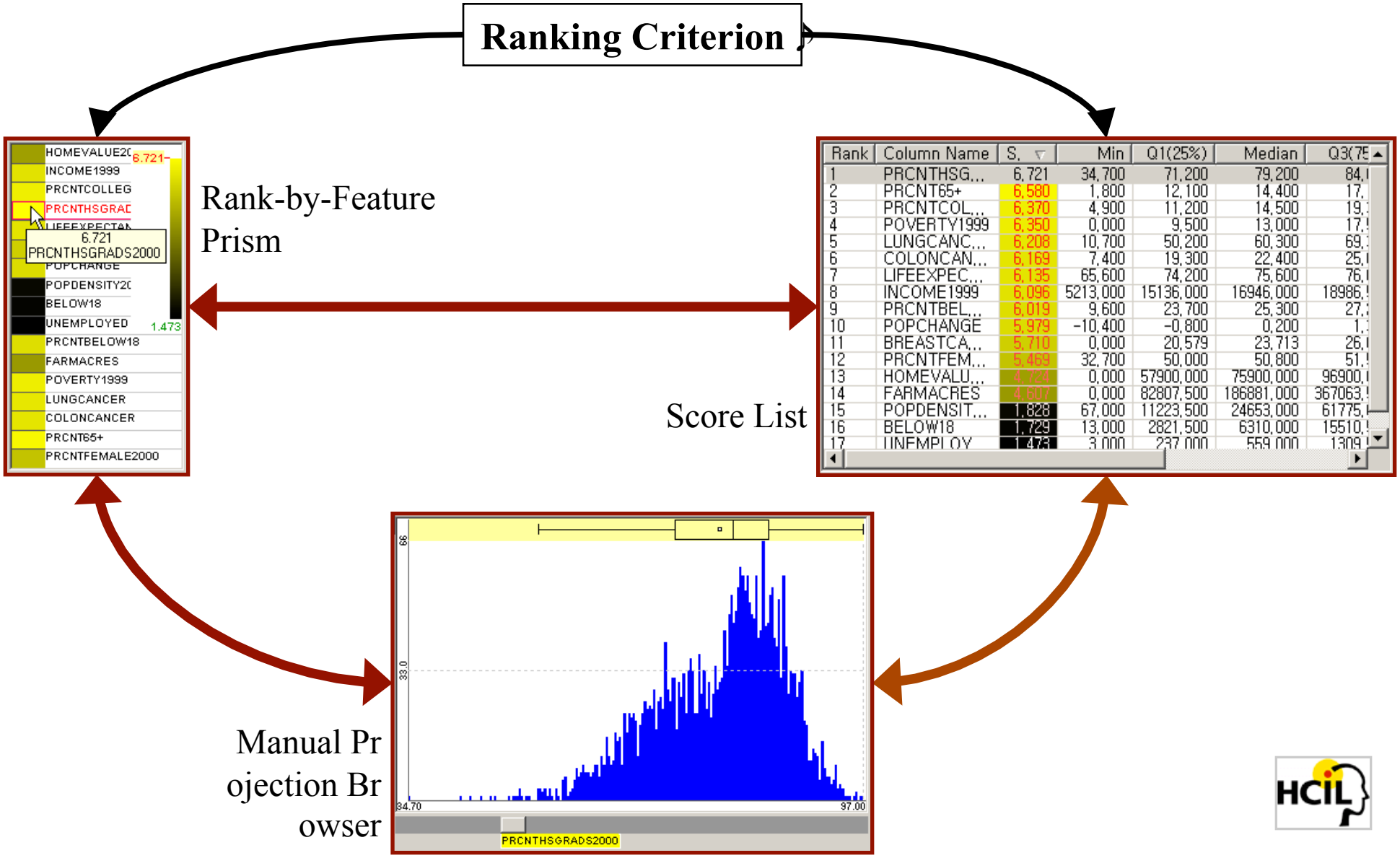


Demonstration

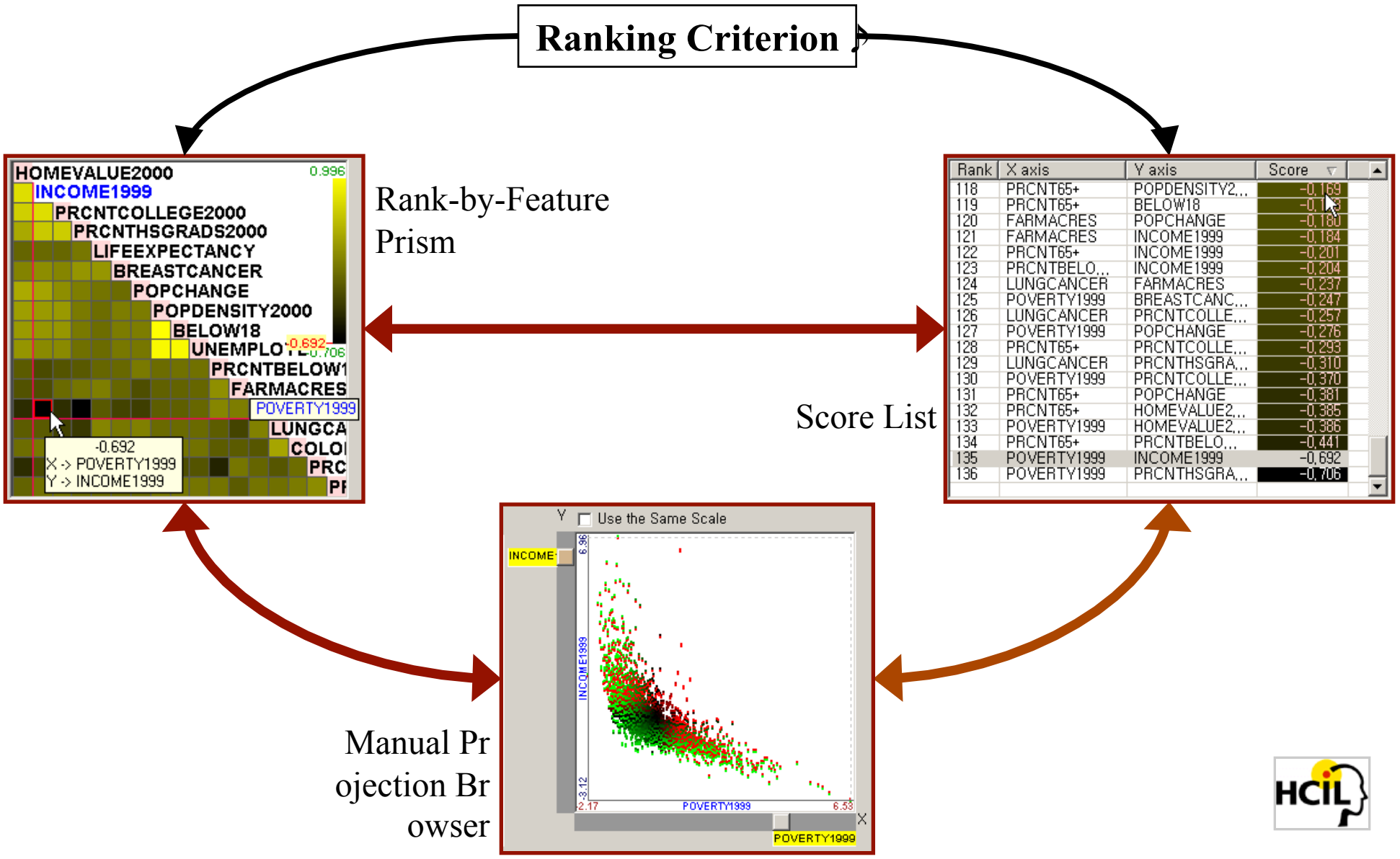
- US counties census data
  - 3138 counties
  - 14 dimensions : population density, poverty level, unemployment, etc.

HCIL

Rank-by-Feature Framework: 1D

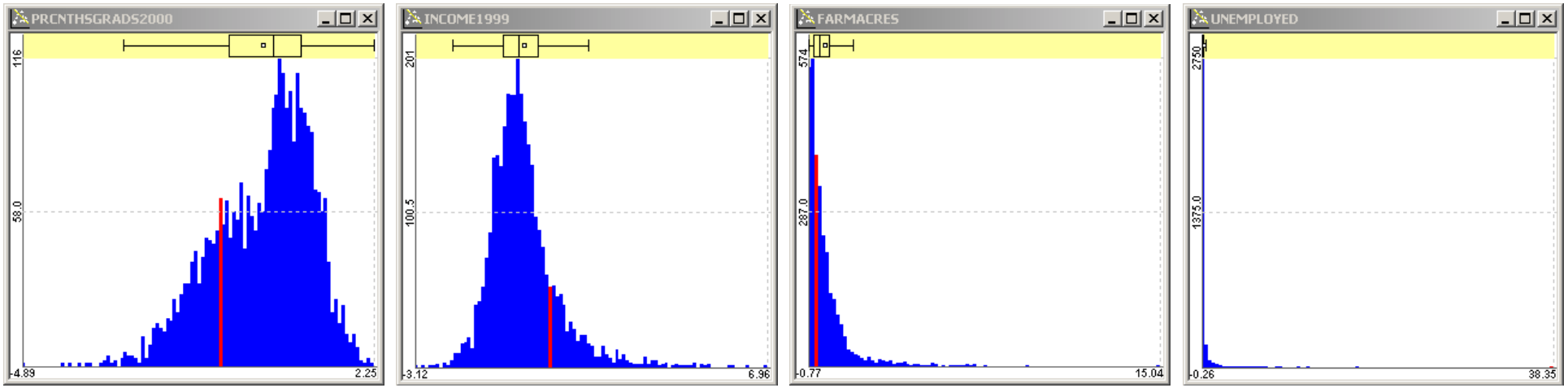


Rank-by-Feature Framework: 2D

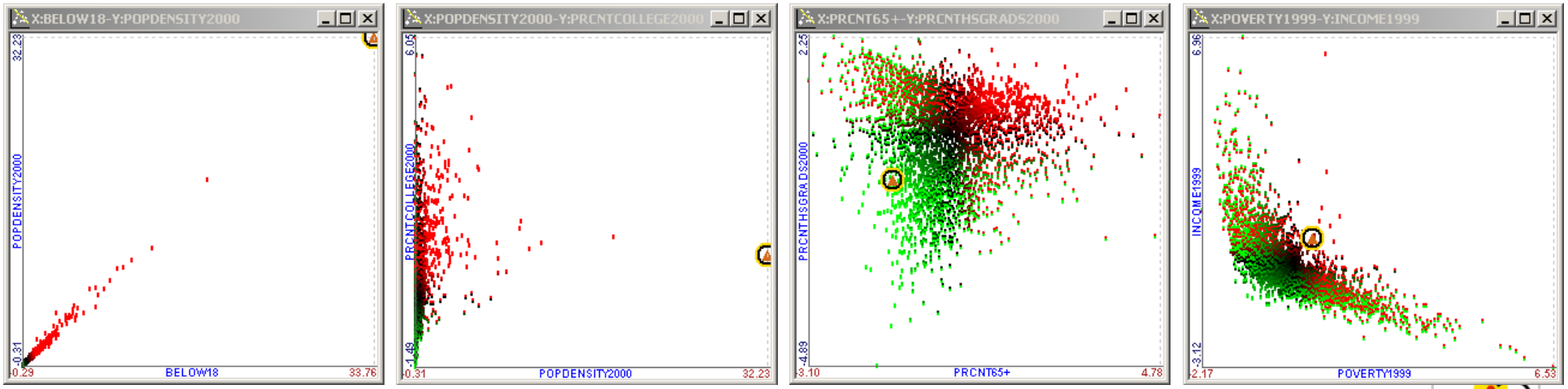


A Ranking Example

3138 U.S. counties with 17 attributes



Ranking Criterion: Uniformity (entropy) (6.7, 6.1, 4.5, 1.5)



Ranking Criterion: Pearson correlation (0.996, 0.31, 0.01, -0.69)

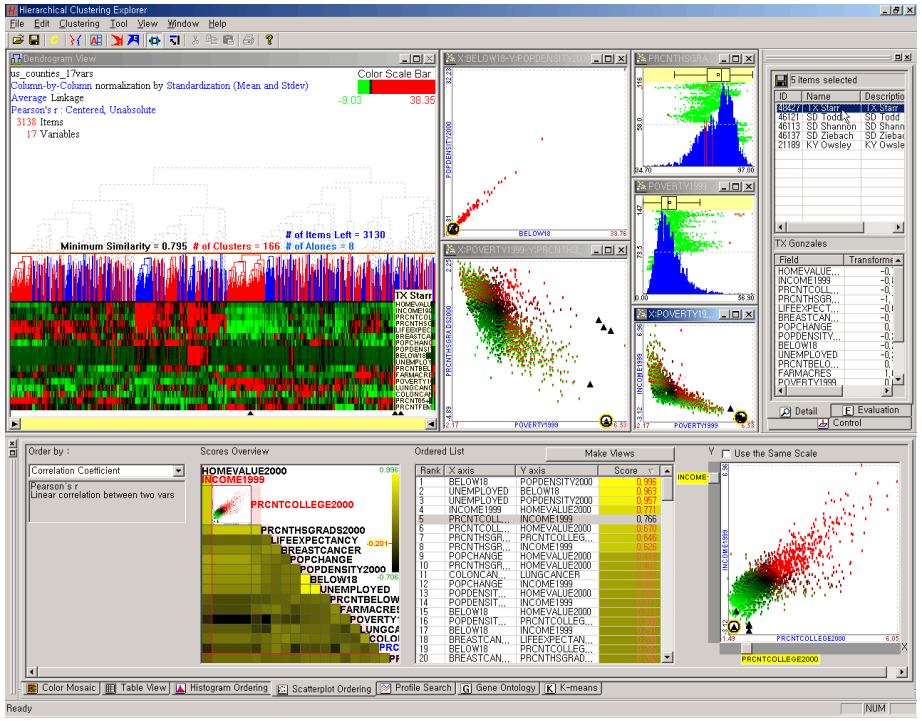
HCIL

HCE Status

- In collaboration and sponsored by Eric Hoffman: Children's National Medical Center
- Phd work of Jinwook Seo

- 72K lines of C++ codes
- 8,000+ downloads since April 2002

- www.cs.umd.edu/hcil/hce



Evaluation Methods

Ethnographic Observational Situated

- Multi-Dimensional
- In-depth
- Long-term
- Case studies

Evaluation Methods

Ethnographic Observational Situated

- Multi-Dimensional
- In-depth
- Long-term
- Case studies

  Domain Experts
  Doing Their Own Work
  for Weeks & Months

HCIL

Evaluation Methods

Ethnographic Observational Situated

- Multi-Dimensional
- In-depth
- Long-term
- Case studies

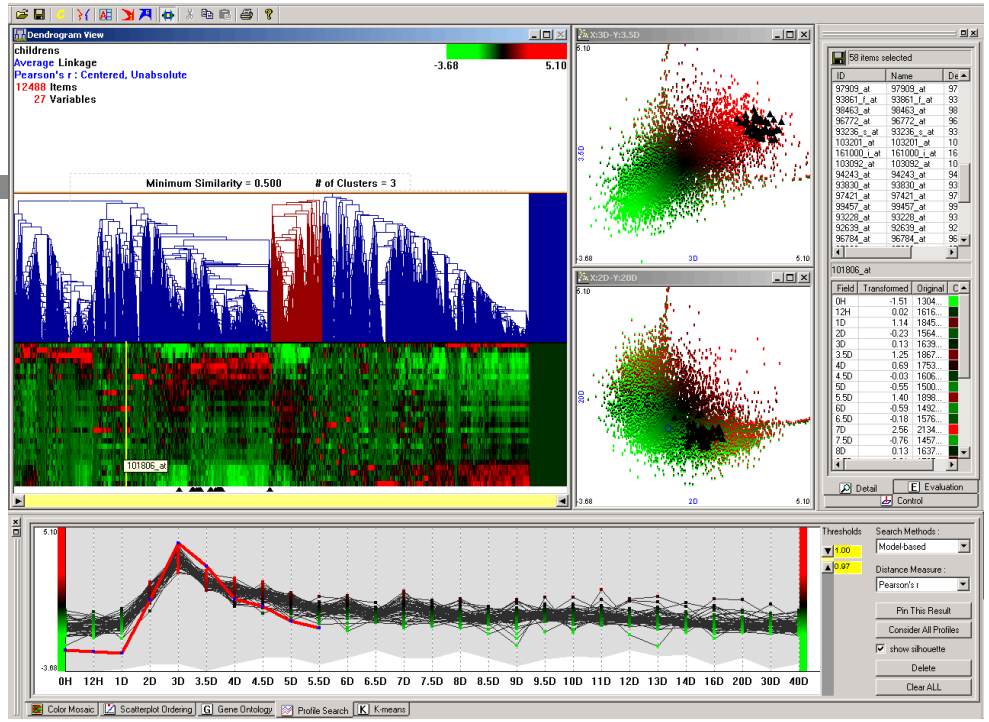
MILCs

Shneiderman & Plaisant, *BeLIV workshop*, 2006

HCIL

MILC example

- Evaluate Hierarchical Clustering Explorer



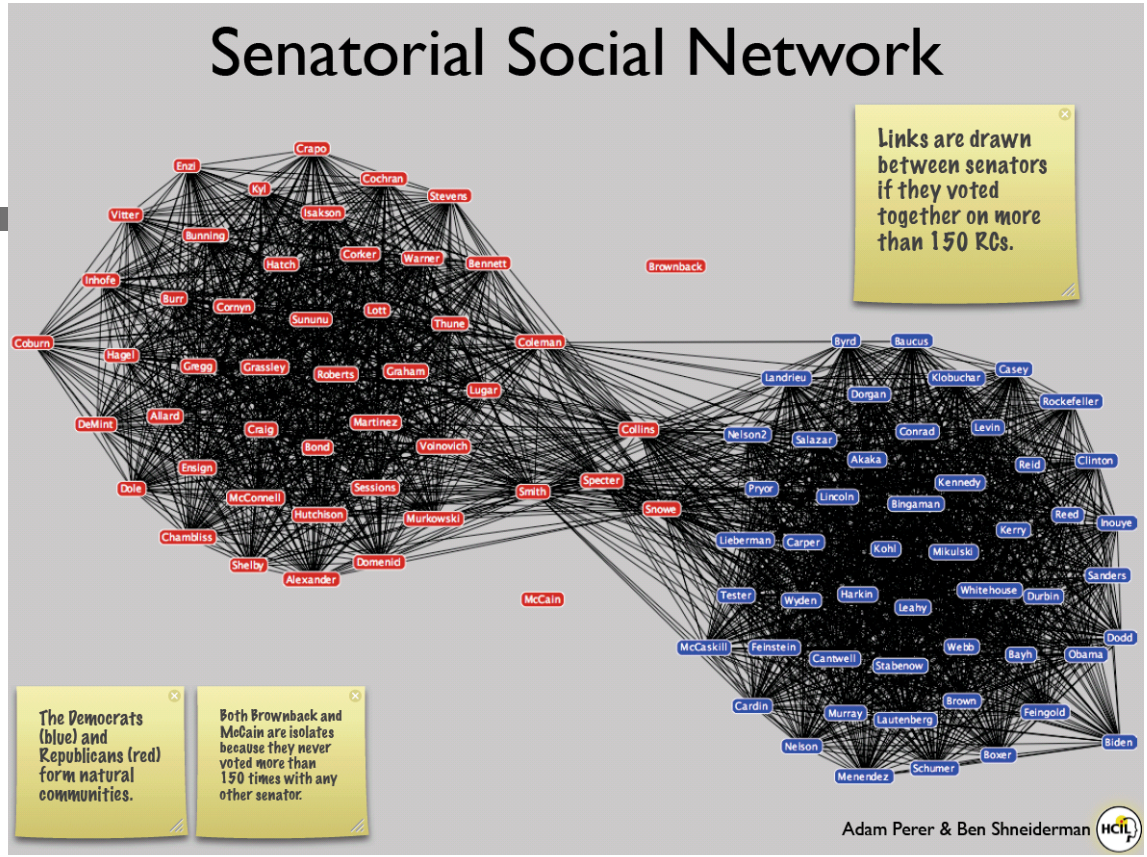
- Focused on rank-by-feature framework
- 3 case studies, 4-8 weeks (molecular biologist, statistician, meteorologist)
- 57 email surveys
- Identified problems early, gave strong positive feedback about benefits of rank-by-feature

Seo & Shneiderman, *IEEE TVCG* 12,3, 2006

HCIL

MILC example

- Evaluate SocialAction



- Focused on integrating statistics & visualization
- 4 case studies, 4-8 weeks (journalist, bibliometrician, terrorist analyst, organizational analyst)
- Identified desired features, gave strong positive feedback about benefits of integration

Perer & Shneiderman, CHI2008

Case Study Methodology

1) Interview (1 hr)
2) Training (2 hr)
3) Early Use (2-4 weeks)
4) Mature Use (2-4 weeks)
5) Outcome (1 hr)

Take Away Message

Rank-by-Feature Framework

- Decomposition of complex problems into multiple simpler problems wins
- Ranking guides discovery
- Systematic strategies

www.cs.umd.edu/hcil/hce

HCIL

26th Annual Symposium
May 28-29, 2009

www.cs.umd.edu/hcil