New Insights from Data: the Future of Data Visualization: Integrating Visualization and Analysis or
Important Future Research Areas for Information Visualization

IEEE Computer Society and GBC/ACM CSAIL November 8, 2012
Georges Grinstein, University of Massachusetts Lowell

Today

- Today there are many visualization systems
  - web/cloud-based, desktop, free, partially free, in analysis packages
  - user-programmable, simple to use, some very complex
  - some browser agnostic and some not, some 2D and some 3D
  - Some in Javascript and some in languages new to many of us.
- This is the decade for visualization
- What is making it exciting are all its new instances involving phones, pads, projectors, and bionics
- I am going to discuss old/new Grand Challenges resulting from new theories, new devices, and new capabilities
- There are still many problems ... Very exciting!

Outline

- Who am I? What do I do? Motivation
- Some intuition for visualization and analysis
- Grand challenges and key research areas
- Discuss three of my favorites
  - Scale and Theory, Visualization for the Masses
- Weave
  - Web-based Analysis and Visualization Environment

Visualization (today)

- a.k.a.
  - Data Visualization
  - Scientific Data Visualization
  - Information Visualization
  - Visual Analytics

and visualization in almost all applications

Software, biology, math, sensors, devices, …
Text, knowledge, images, videos, …
Networks, space, physics, …

We do visualization, analysis, software

- NVAC
- NSF
- CDC
- USAF

SensAble
MGH
Pfizer
IARPA

- Open Indicators Consortium

and BBN, AGM, Biogen, AstraZeneca, …

charles river analytics

Institute for Visualization and Perception Research

UMASS

Visual Analytics

The science of analytical reasoning facilitated by interactive visual interfaces

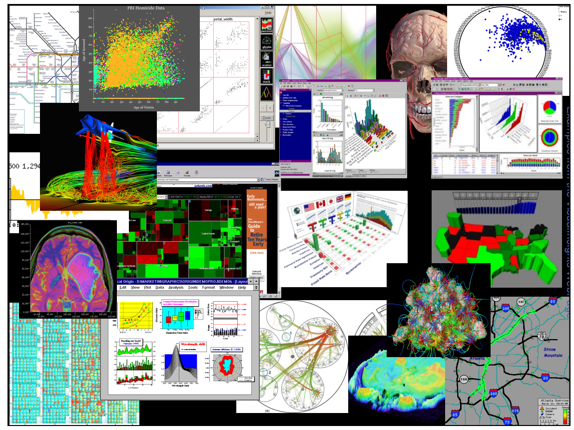
information visualized
knowledge discovered
decisions made

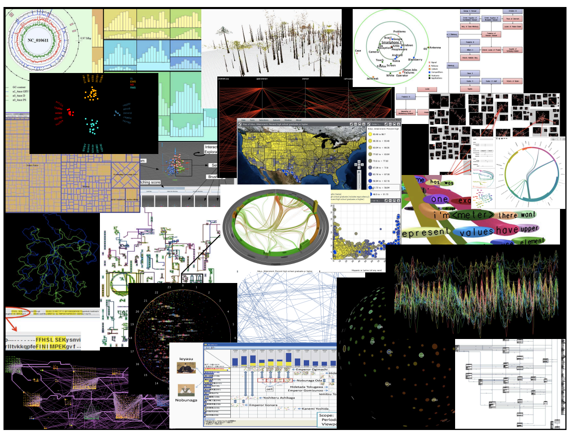
Institute for Visualization and Perception Research

My Visualization Research

Doctoral students (20-25), MS (2), UG (5), HS (1)

information visualized
knowledge discovered
decisions made

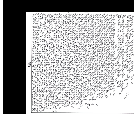


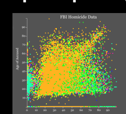


Our Visual Analytics Research

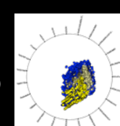
- **Theory**

  Enhance data perception (1980s)





  Visualize thousands of variables (1990s)



- **Reasoning**

  Evaluate insight (2000s) ← IEEE VAST Challenges

  Measure cognition (2010 on) ← USAF Leverage Points

Our Visual Analytics Research

- **Systems**

  Integrated analysis and visualization

  Session history

  Collaboration

  Visualization for the masses

  Performance with large data (millions of variables)

  ADA Compliance

- **We deployed many systems**

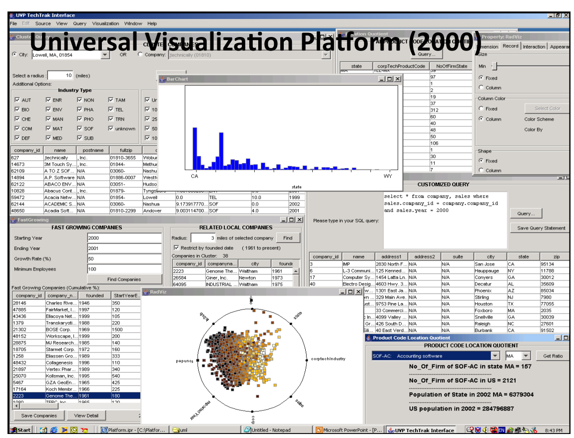
  Lockheed's cockpit graphics in Fighter Jets (80s)

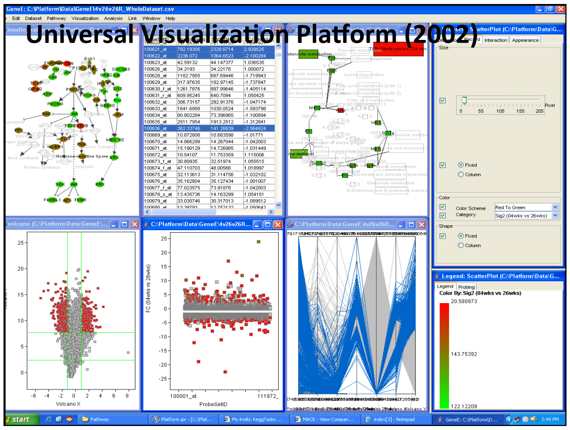
  Intel's i860 graphics libraries (90s)

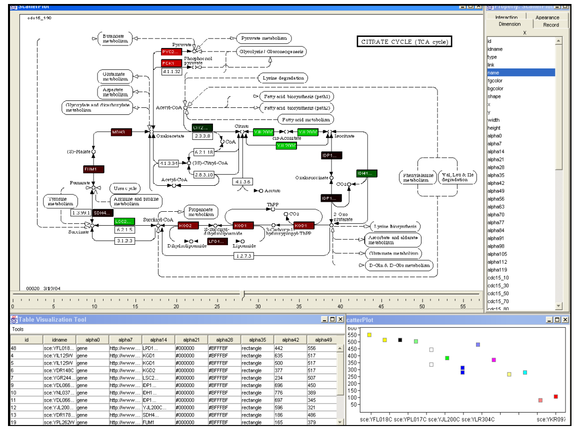
  Universal Visualization Platform (2000s)

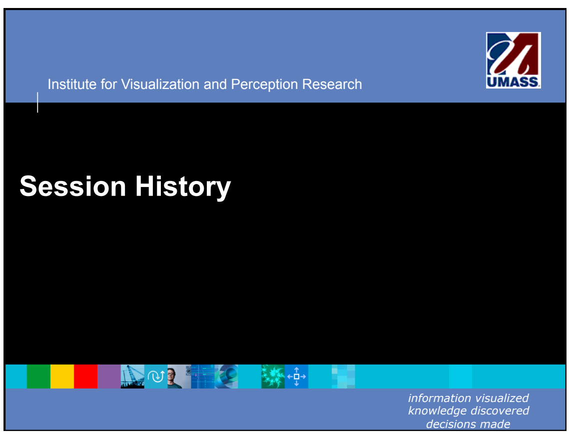
  Open Indicators Consortium's Weave (2010s)

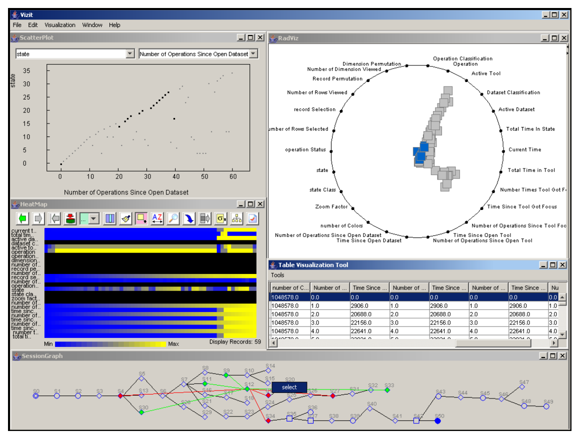
Universal Visualization Platform (2000)

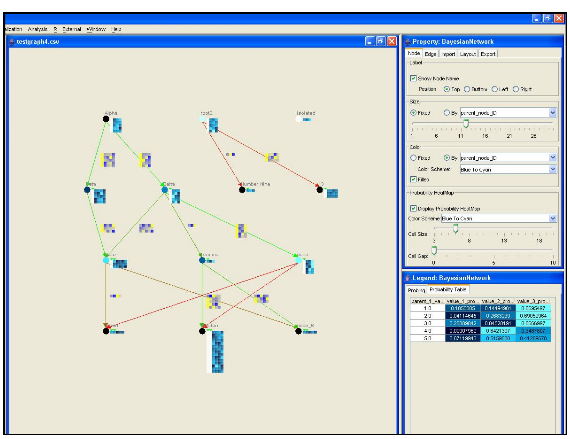


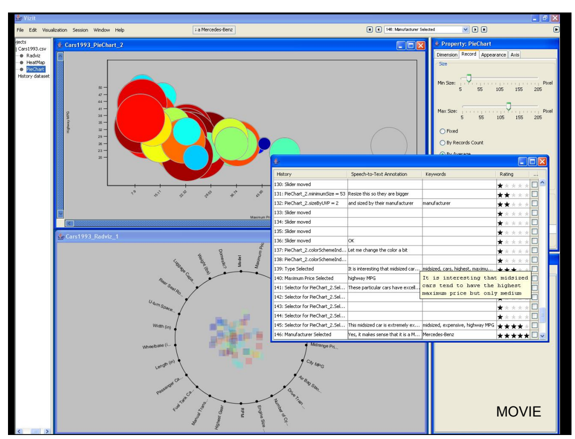

Universal Visualization Platform (2002)


CITRATE CYCLE (TCA cycle)
Table Visualization Tool

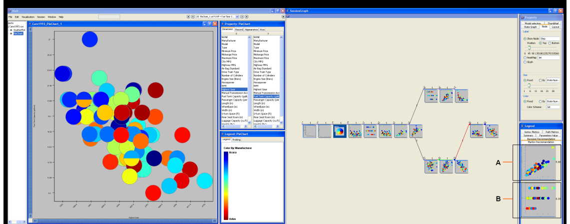

Institute for Visualization and Perception Research
UMASS
Session History
information visualized
knowledge discovered
decisions made


ScatterPlot
RadViz
HeatMap
Table Visualization Tool
SessionGraph


testgraph4.csv
Property: BayesianNetwork
Legend: BayesianNetwork


Cars1993_PieChart_2
Property: PieChart
Cars1993_Radviz_1
MOVIE

Recommendation System (Chiang 2006)

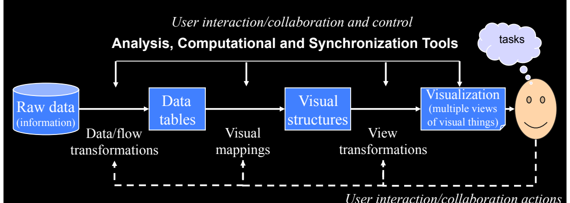


Institute for Visualization and Perception Research

A Visualization Pipeline

information visualized
knowledge discovered
decisions made

Visualization } Pipeline
Visual Analytics }



User interaction **everywhere**
Algorithms **everywhere**

Institute for Visualization and Perception Research

Lots Visualization Issues

- the data
- the transformations/computations
- the presentation
- the human

information visualized
knowledge discovered
decisions made

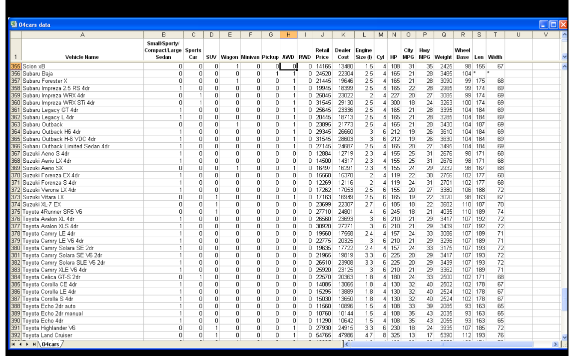
Institute for Visualization and Perception Research

Lots Visualization Issues

- the data
- the transformations/computations
- the presentation
- the human

information visualized
knowledge discovered
decisions made

Data: 2004 cars and trucks

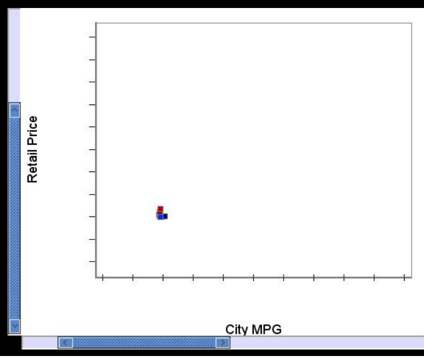


04cars data

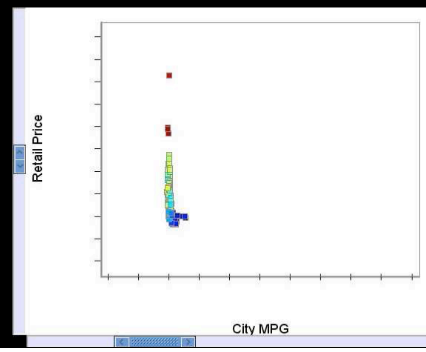
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Vehicle Name	Small/Sporty/ Compact/Large Sedan	Sports Car	SUV	Wagon	Minivan	Pickup	AWD	RWD	Retail Price	Dealer Cost	Engine Size (l)	Cyl	HP	City MPG	Hwy MPG	Weight	Wheel Base	Len	Width
355	Scion xB	0	0	0	1	0	0	0	0	14165	13480	1.5	4	108	31	35	2425	98	155	67
356	Subaru Baja	0	0	0	0	0	1	1	0	24520	22304	2.5	4	165	21	28	3485	104	*	*
357	Subaru Forester X	0	0	0	1	0	0	1	0	21445	19646	2.5	4	165	21	28	3090	99	175	68
358	Subaru Impreza 2.5 RS 4dr	1	0	0	0	0	0	1	0	19945	18399	2.5	4	165	22	28	2965	99	174	69
359	Subaru Impreza WRX 4dr	0	1	0	0	0	0	1	0	25045	23022	2	4	227	20	27	3085	99	174	69
360	Subaru Impreza WRX STi 4dr	0	1	0	0	0	0	1	0	31545	29130	2.5	4	300	18	24	3263	100	174	69
361	Subaru Legacy GT 4dr	1	0	0	0	0	0	1	0	25645	23336	2.5	4	165	21	28	3395	104	184	69
362	Subaru Legacy L 4dr	1	0	0	0	0	0	1	0	20445	18713	2.5	4	165	21	28	3285	104	184	69
363	Subaru Outback	0	0	0	1	0	0	1	0	23895	21773	2.5	4	165	21	28	3430	104	187	69
364	Subaru Outback H6 4dr	1	0	0	0	0	0	1	0	29345	26660	3	6	212	19	26	3610	104	184	69
365	Subaru Outback H-6 VDC 4dr	1	0	0	0	0	0	1	0	31545	28603	3	6	212	19	26	3630	104	184	69
366	Subaru Outback Limited Sedan 4dr	1	0	0	0	0	0	1	0	27145	24687	2.5	4	165	20	27	3495	104	184	69
367	Suzuki Aeno S 4dr	1	0	0	0	0	0	0	0	12884	12719	2.3	4	155	25	31	2676	98	171	68
368	Suzuki Aerio LX 4dr	1	0	0	0	0	0	0	0	14500	14317	2.3	4	155	25	31	2676	98	171	68
369	Suzuki Aerio SX	0	0	0	1	0	0	1	0	16497	16291	2.3	4	155	24	29	2932	98	167	68
370	Suzuki Forenza EX 4dr	1	0	0	0	0	0	0	0	15568	15378	2	4	119	22	30	2756	102	177	68
371	Suzuki Forenza S 4dr	1	0	0	0	0	0	0	0	12269	12116	2	4	119	24	31	2701	102	177	68
372	Suzuki Verona LX 4dr	1	0	0	0	0	0	0	0	17262	17053	2.5	6	155	20	27	3380	106	188	72
373	Suzuki Vitara LX	0	0	1	0	0	0	1	0	17163	16949	2.5	6	165	19	22	3020	98	163	67
374	Suzuki XL-7 EX	0	0	1	0	0	0	0	0	23699	22307	2.7	6	185	18	22	3682	110	187	70
375	Toyota 4Runner SR5 V6	0	0	1	0	0	0	0	0	27710	24801	4	6	245	18	21	4035	110	189	74
376	Toyota Avalon XL 4dr	1	0	0	0	0	0	0	0	26560	23693	3	6	210	21	29	3417	107	192	72
377	Toyota Avalon XLS 4dr	1	0	0	0	0	0	0	0	30920	27271	3	6	210	21	29	3439	107	192	72
378	Toyota Camry LE 4dr	1	0	0	0	0	0	0	0	19560	17558	2.4	4	157	24	33	3086	107	189	71
379	Toyota Camry LE V6 4dr	1	0	0	0	0	0	0	0	22775	20325	3	6	210	21	29	3296	107	189	71
380	Toyota Camry Solara SE 2dr	1	0	0	0	0	0	0	0	19635	17722	2.4	4	157	24	33	3175	107	193	72
381	Toyota Camry Solara SE V6 2dr	1	0	0	0	0	0	0	0	21965	19819	3.3	6	225	20	29	3417	107	193	72
382	Toyota Camry Solara SLE V6 2dr	1	0	0	0	0	0	0	0	26510	23908	3.3	6	225	20	29	3439	107	193	72
383	Toyota Camry XLE V6 4dr	1	0	0	0	0	0	0	0	25920	23125	3	6	210	21	29	3362	107	189	71
384	Toyota Celica GT-S 2dr	0	1	0	0	0	0	0	0	22570	20363	1.8	4	180	24	33	2500	102	171	68
385	Toyota Corolla CE 4dr	1	0	0	0	0	0	0	0	14085	13065	1.8	4	130	32	40	2502	102	178	67
386	Toyota Corolla LE 4dr	1	0	0	0	0	0	0	0	15295	13889	1.8	4	130	32	40	2524	102	178	67
387	Toyota Corolla S 4dr	1	0	0	0	0	0	0	0	15030	13650	1.8	4	130	32	40	2524	102	178	67
388	Toyota Echo 2dr auto	1	0	0	0	0	0	0	0	11560	10896	1.5	4	108	33	39	2085	93	163	65
389	Toyota Echo 2dr manual	1	0	0	0	0	0	0	0	10760	10144	1.5	4	108	35	43	2035	93	163	65
390	Toyota Echo 4dr	1	0	0	0	0	0	0	0	11290	10642	1.5	4	108	35	43	2055	93	163	65
391	Toyota Highlander V6	0	0	1	0	0	0	1	0	27930	24915	3.3	6	230	18	24	3935	107	185	72
392	Toyota Land Cruiser	0	0	1	0	0	0	1	0	54765	47986	4.7	8	325	13	17	5390	112	193	76

04cars

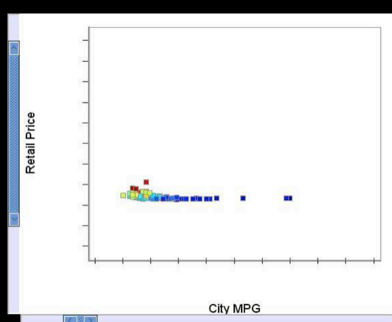
Large x and y scale



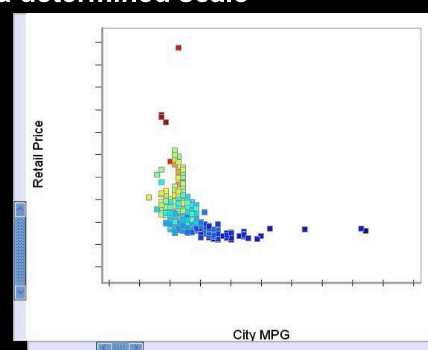
Large x scale



Large y scale



Data determined scale



Institute for Visualization and Perception Research

UMASS

Lots Visualization Issues

- the data
- the transformations/computations
- the presentation
- the human

information visualized
knowledge discovered
decisions made

Depends on Data Mining Tasks & Techniques

Major Data Mining Tasks
- Summarization
- Association
- Classification
- Prediction
- Clustering

using →

Major Techniques
- Linear Regression Trees
- Non-Linear Regression
- MARS
- Naïve Bayes
- K-Means and K-Median
- Neural Networks
- Association Rules
- Decision Trees
- Principal Curve Analysis
- Support Vector Machines
- Genetic Algorithms

based on ←

Statistical Tools
- Missing Value Imputation
- Normalizations
- Error & Variational Analysis
- Confidence Estimates

Why are there so many?

- Most problems are NP-Hard
- High dimensional distributions are elusive

300 dimensional data set of only binary values

$2^{300} = (2^{10})^{30} = 1024^{30} \approx (10^3)^{30} = 10^{90}$

>> number of particles in the universe

There are so many!

- KDD2001 CUP

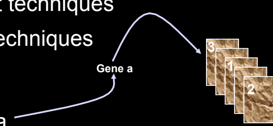
  Thrombosis data set

  Over 200 submissions

  Over 100 different techniques

  Many combined techniques
- KDD2002 CUP

  Document corpora

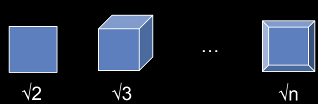


- IEEE Visualization and VAST contests

  Text, networks, videos, large data, …

Why are there so many? Really?

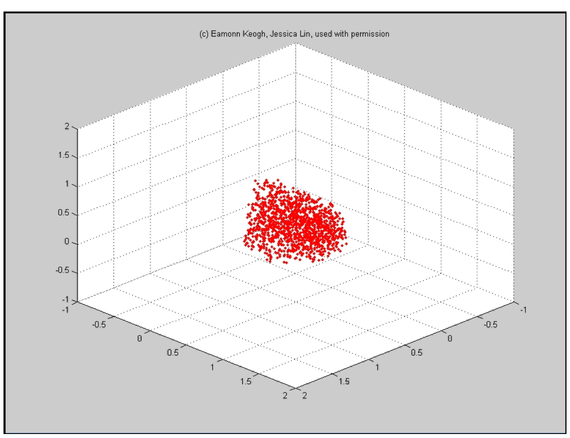
- Metrics in high dimensional spaces are non-intuitive

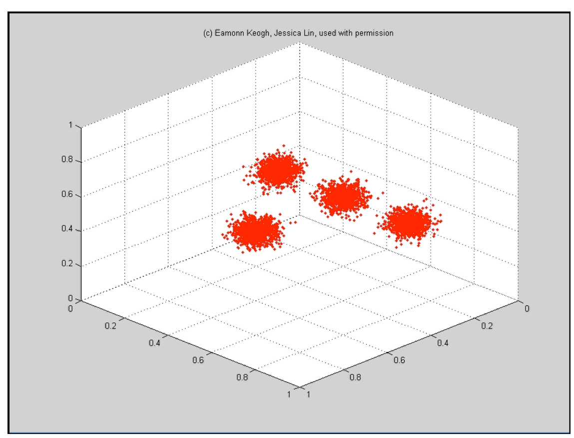


- For 36 dimensions the cube (1,1,1, …,1) corner is 6 units away from the origin!

Curse of Dimensionality

- Very costly to compute on large data sets (even if not NP, often > order n^2)
- In a "real" data set, a large number of dimensions precludes identifying precisely its distribution model making the selection of an algorithm somewhat heuristic)
- But we still reduce (subset, reduce dimensions, sample, …)





Institute for Visualization and Perception Research

UMASS

Lots Visualization Issues

- the data
- the transformations/computations
- **the presentation**
- the human

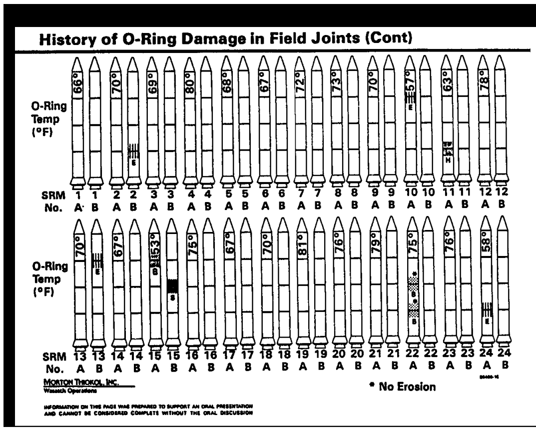
information visualized
knowledge discovered
decisions made

Visualization Techniques

Pure
- 2D and 3D Scatterplots
- Matrices of Scatterplots
- Statistical Charts
- Line and Multi-line Graphs
- Parallel Coordinates
- Circle Segment
- Polar Charts
- Survey Plots
- Heatmaps
- Height Maps
- Tree Maps
- Graphs
- Iconographic/Glyph Displays
- Radial Visualizations
- …

Interactions
- Selection
- Probing, Querying
- Zoom (Linear and Non-linear)
- Grand Tours
- …

Integrated with Analysis
- Projection Pursuit
- Dimensional Stacking
- Sammon Plots
- Multi-Dimensional Scaling
- PCA and Principal Curves
- Self Organizing Maps
- …



Identical patient data presented 4 different ways

Preliminary results from the comparison of two hypothetical clinical trials both treating the same condition

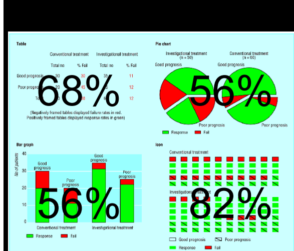
a generic conventional treatment

a generic investigational treatment

Would clinicians decide to stop the trial because the data show **one of the treatments is obviously better than the other?**

From Elting et al., BMJ 1999, 318:1527-31

Identical patient data presented 4 different ways



The decision to stop varied significantly depending on the presentation of the data

Correct decisions were
- 56% with bar charts
- 56% pie charts
- 68% with tables
- 82% with icon displays

In actual clinical practice, up to 25% of the patients treated based on the bar or pie charts would have received inappropriate treatment

From Elting et al., BMJ 1999, 318:1527-31

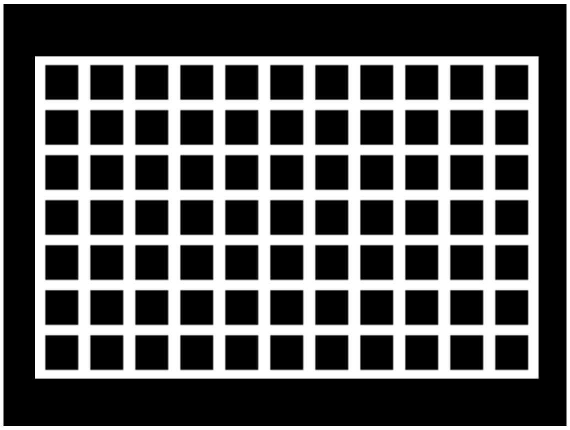
Institute for Visualization and Perception Research

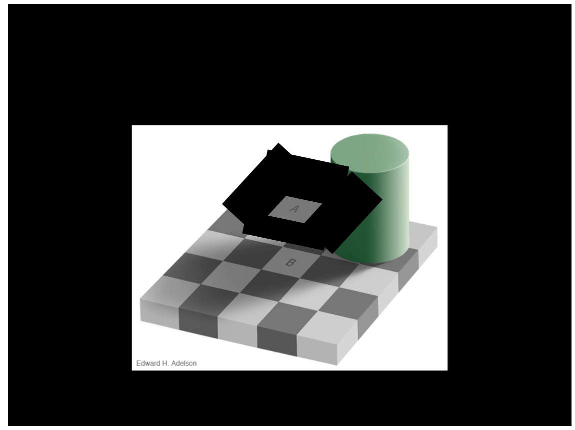
UMASS

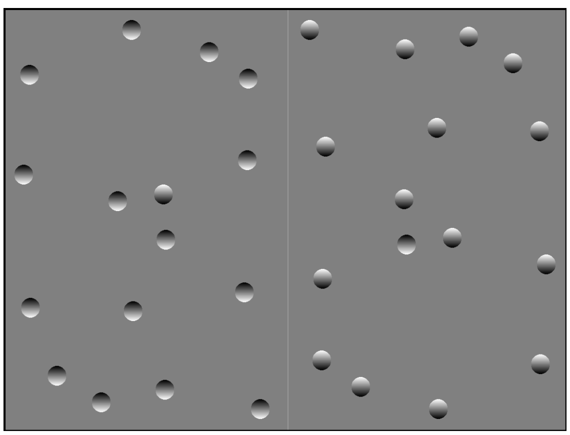
Lots Visualization Issues

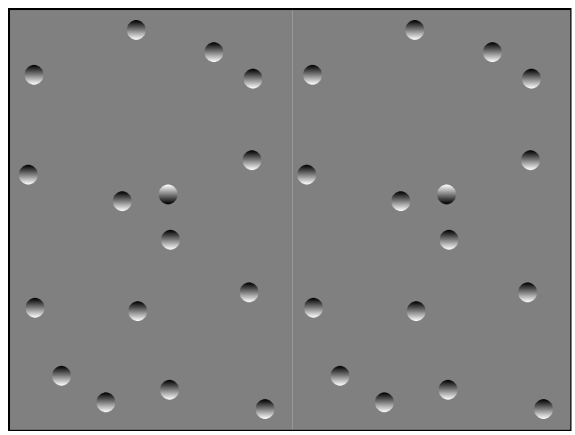
- the data
- the transformations/computations
- the presentation
- **the human**

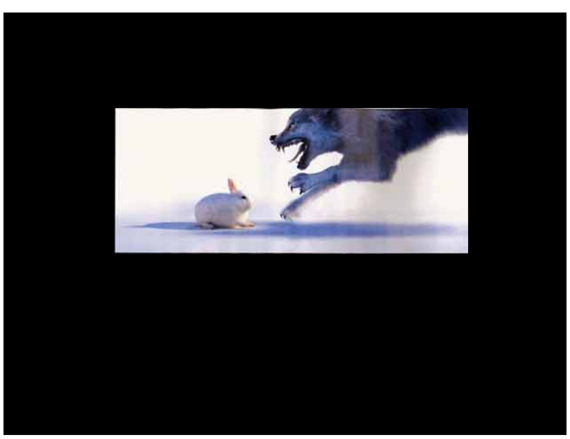
information visualized
knowledge discovered
decisions made











Institute for Visualization and Perception Research

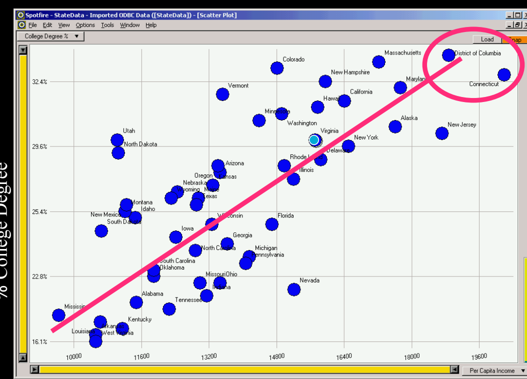
UMASS

So when/why do visualizations?

Which state has the highest income? Most college degrees? (from Chris North)

Table - StateData ()

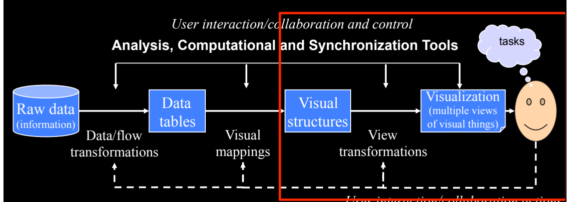
State	College Degree %	Per Capita Income
Alabama	20.6%	11486
Alaska	30.3%	17610
Arizona	27.1%	13461
Arkansas	17.0%	10520
California	31.3%	16409
Colorado	[illegible]	[illegible]
Connecticut	33.8%	20189
Delaware	[illegible]	[illegible]
District of Columbia	36.4%	18881
Florida	24.9%	14698
Georgia	24.3%	13631
Hawaii	31.2%	15770
Idaho	25.2%	11457
Illinois	26.8%	15201
Indiana	20.9%	13149
Iowa	24.5%	12422
Kansas	26.5%	13300
Kentucky	17.7%	11153
Louisiana	19.4%	10635
Maine	25.7%	12957
Maryland	31.7%	17730
Massachusetts	34.5%	17224
Michigan	24.1%	14154
Minnesota	30.4%	14389
Mississippi	19.9%	9648
Missouri	22.3%	12989
Montana	25.4%	11213
Nebraska	26.0%	12452
Nevada	21.5%	15214
New Hampshire	32.4%	15959
New Jersey	30.1%	18714
New Mexico	25.5%	11246
New York	29.6%	16501
North Carolina	24.2%	12885
North Dakota	28.1%	11051
Ohio	22.3%	13461
Oklahoma	22.8%	11893
Oregon	27.5%	13418
Pennsylvania	23.2%	14068
Rhode Island	27.5%	14981
South Carolina	23.0%	11897
South Dakota	24.6%	10661
Tennessee	20.1%	12255
Texas	25.5%	12904
Utah	30.0%	11029
Vermont	31.5%	13527
Virginia	30.0%	15713
Washington	30.9%	14923
West Virginia	16.1%	10520
Wisconsin	24.9%	13276
Wyoming	25.7%	12311



% College Degree

Per Capita Income

Involving the User actually involves all previous steps!



Four Types of Visualizations

- Exploratory
  Have no hypotheses about the data
  Explore data interactively as undirected searches
- Confirmatory
  Have specific hypotheses about the data
  Goal-oriented examination of the hypotheses
- Presentation
  Facts to be presented are fixed a priori
  Select appropriate presentation techniques
- Interactive Presentation
  Interactions with a pre-defined animation

Institute for Visualization and Perception Research

UMASS

What is a Grand Challenge?

Operational Definition

- A Grand Challenge is a goal recognized one or two decades in advance, where by its achievement represents a major milestone in the advance of knowledge or technology
- The problem is hard
- There is a specific goal (clear when the problem is solved)
- Its solution often generates a “wow”

Grand Challenge Examples

Lots of Grand Challenge Studies

VAST
My Survey

Institute for Visualization and Perception Research

Past Grand Challenges in Visualization

Not really Grand Challenges but Nice Problems

information visualized
knowledge discovered
decisions made

Grand Challenge Problems

- NSF Monumental research challenges
  Cognitive partners for humans
  Personalized lifelong learning environments
  Unfailingly reliable systems
  Making information technology less complex*
- More modern
  Architecture the brain, mind and computer
  Memories for life
  In silico life simulation
  Non-classical computation

Note: BIG DATA is NOT a grand challenge

Institute for Visualization and Perception Research

Grand Challenges in Visualization and …

information visualized
knowledge discovered
decisions made

Email Survey

- I contacted well over 200 visualization researchers, some CHI, graphics, and related others, in two passes and asked simply

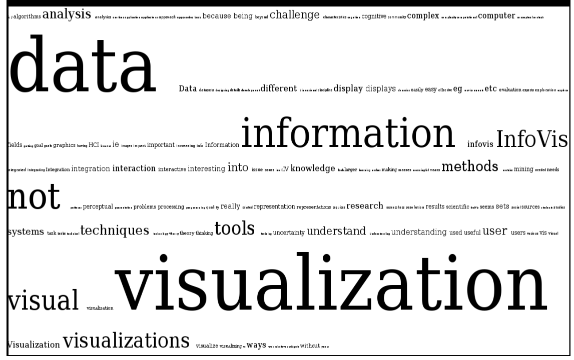
  Do me a favor: tell me what you think are the five grand challenges of InfoVis

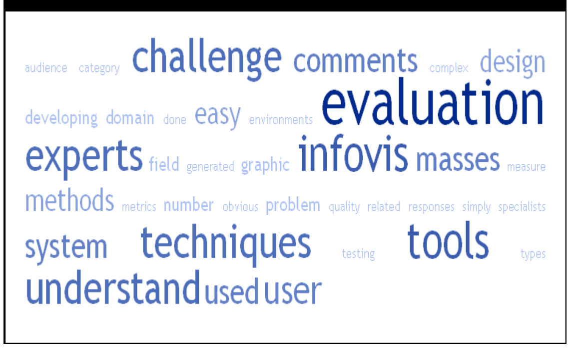
- There was a tremendous return
- And the work began

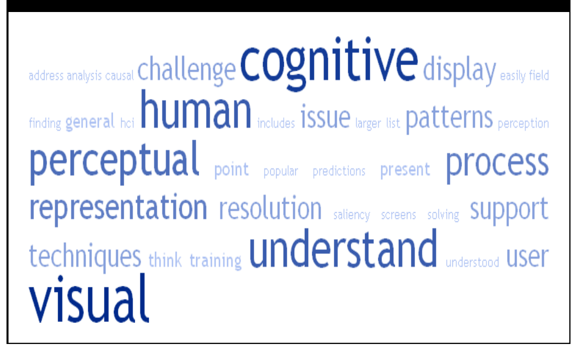
Email Survey

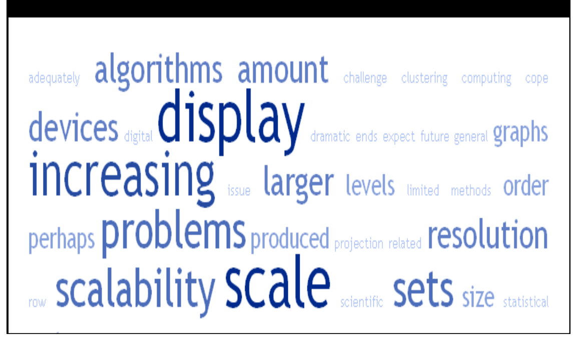
- Resulted in the identification of about 30 different but related topics
- These match the previous grand challenges

All unprocessed emails









The top 11 (organized)

- Evaluation (30)
- Human Cognition (24)
- Interaction (19)
- Non-expert (30)
- Market InfoVis (21)
- Knowledge Capture (21)
- Data Integration (21)
- Visual Analytics (22)
- Scalability (27)
- Uncertainty (15)
- Theory (18)

Institute for Visualization and Perception Research

Integration of Visualization and Analysis

The One Grand Challenge

information visualized
knowledge discovered
decisions made

Institute for Visualization and Perception Research

Selecting the Visual Mapping (or Selecting the Visualization)

information visualized
knowledge discovered
decisions made

The big question

Given data set *D*, Task *T*, Medium *M* ,User *U* *(and some other things)*

What is the **best** visual display?

"best" must to be defined

Rephrased

- This is really **THE** grand challenge
- Given a **task**

  Determine the **best/optimal visualization** to support the user in the decision-making process

- Requires measures of tasks, cognition, …

Example of a theory's use

- **Conjecture**:
  Given a data set ***D***, given a task ***T***,
  given a medium ***M***, given a user ***U***
  Then there exists a visualization ***V*** for which the perceived information ***I*** is such that task ***T*** is optimally perceptually and cognitively "resolved"
- There are lots of unknowns, undefined terms and measures ***(D, T, M, V, U, I)*** and a number of dependencies
  For example the perceived information ***I*** depends on the user ***U*** and the medium ***M***

$$I = F(I_S, I, M)$$

Applications

- With such a Conjecture we can (attempt) to solve a large number of difficult problems
- For example, we could build an **average perceived information measure across all users** or classes of users or visualizations or media or …

Institute for Visualization and Perception Research

Integration of Visualization and Analysis

2 Big Challenges

Scalability (theory, engineering)

Visualization for the masses (engineering)

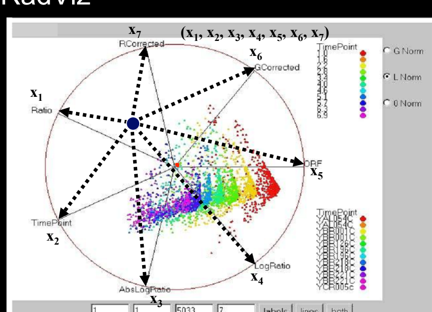
information visualized
knowledge discovered
decisions made

Institute for Visualization and Perception Research

Scalability and Theory: RadViz and Millions of Variables

information visualized
knowledge discovered
decisions made

RadViz

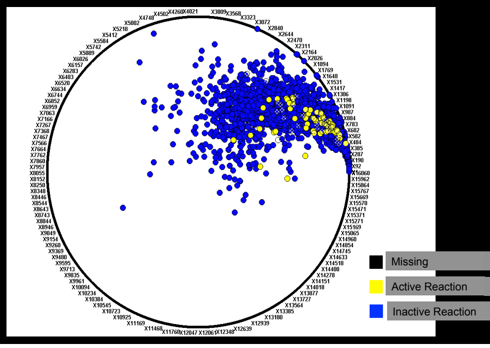


RadViz Mathematical Formulation

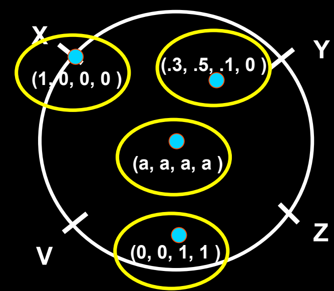
$$x_i = \frac{\left(\sum_{j=1,\ldots,d} a_{i,j}\cos\theta_j\right)}{\left(\sum_{j=1,\ldots,d} a_{i,j}\right)},\ y_i = \frac{\left(\sum_{j=1,\ldots,d} a_{i,j}\sin\theta_j\right)}{\left(\sum_{j=1,\ldots,d} a_{i,j}\right)}, i = 1,\ldots,n \quad \boxed{O(n)}$$

x_i and y_i are the coordinates for record *I*
θ_j is the angle for variable *j*
$a_{i,j}$ is the value for dimension j of record *i*
d = # of dimensions and *n* = # of records.

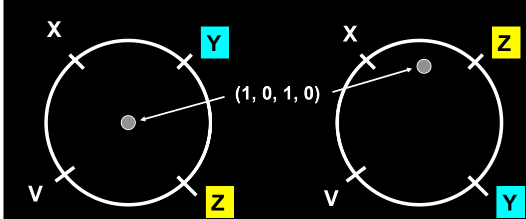
Chemical Compound Reactions Associations Between 16,000 Variables



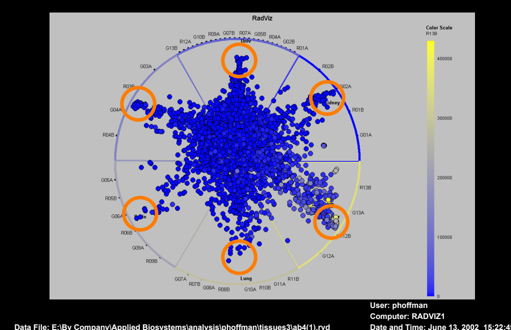
Example: 4d - (x, y, z, v)



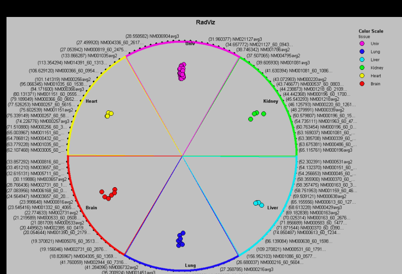
Order of Dimensional Anchors is Critical



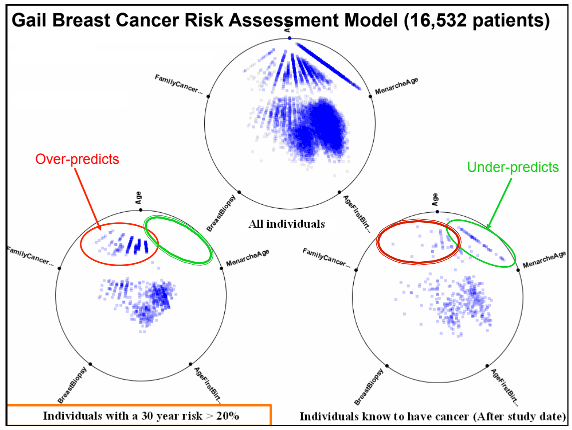
RadViz Tissue Analysis

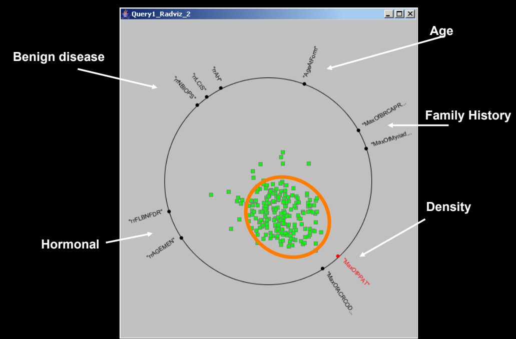


RadViz^{TM} Gene Analysis



Gail Breast Cancer Risk Assessment Model (16,532 patients)





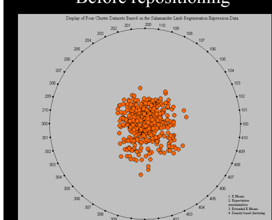
After selecting cancer patients and arranging anchors, visualization highlights that density-based models are needed

Multiple Clustering

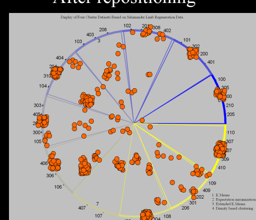
- Run many clustering algorithms on the same data (MANY)
- Make the dimensions in RadViz the clusters
- This generates displays that are indirectly incorporate fuzzy clustering

Repositioning Anchor Points Using a Class Discrimination Layout Algorithm on Newt Gene Expression Data

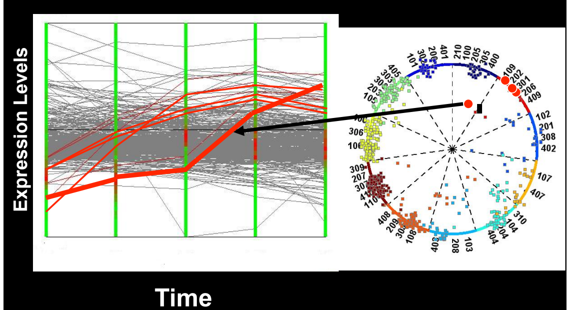
Before repositioning



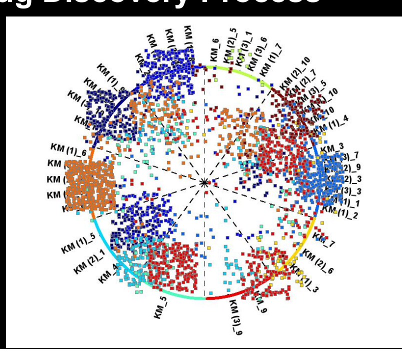
After repositioning



Salamander Expression Levels



Analysis of Compounds Used in Drug Discovery Process

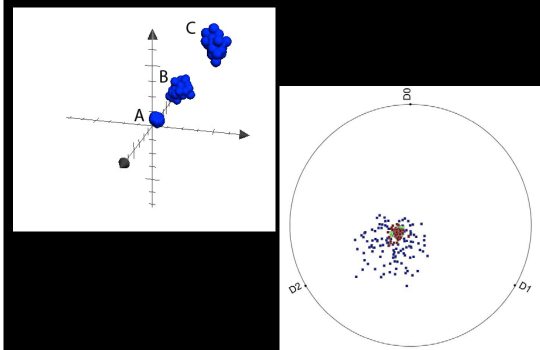


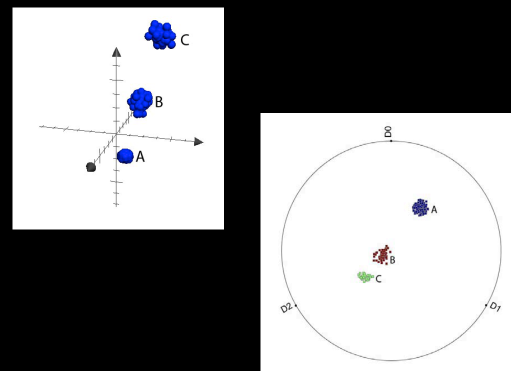
Normalized Radial Visualizations

A Normalized Radial Visualization (NRV) is a projective transformation VNΨ where V is a data vector, N a normalization mapping (perspective and hence projective), and Ψ, an affine dimensional anchor mapping

1. A line segment projects into a single line
2. Betweenness of collinear points is preserved
3. Point-line incidence is preserved
4. Convex sets map to convex sets
5. Hyperspheres and ellipsoids map to ellipsoids (or degenerate to lower dimensional features)

Daniels and Grinstein, Journal of Information Visualization 2012





Institute for Visualization and Perception Research

UMASS

Visualization for the Masses (Democratizing Visualization)

Weave (Web-based Analysis and Visualization Environment)

information visualized
knowledge discovered
decisions made

Weave's History

- William Mass
- Charlotte Kahn, Director Boston Indicator Project at the Boston Foundation
- Grants from National Academy of Sciences and DataHavens to explore technologies for web-based visualization system
- Built prototype and estimate cost ($1Million

- Established Open Indicator Consortium and first 7 members provide $150K each for 2 years

Open Indicators Consortium (OIC) Formed to fund first years of Weave

1. Arizona
2. Atlanta
3. Boston
4. Chicago
5. Connecticut
6. Grand Rapids
7. Kansas City
8. MA EEC
9. Ohio
10. Portland
11. Rhode Island
12. Rockford
13. San Antonio
14. Seattle
15. South Florida

…

The James L. Knight, Barr and The Boston Foundations
The Greater Lowell, Boston, New Haven and Rhode Island Community Foundations

More members and others

- MA Dept of Higher Education
- US Dept of Labor
- Associated Grant Makers
- Centers for Disease Control

- In use in the Census Bureau, UN, …

Many more across the world

iWeave.org
Numerous downloads

The Fundamental Mission

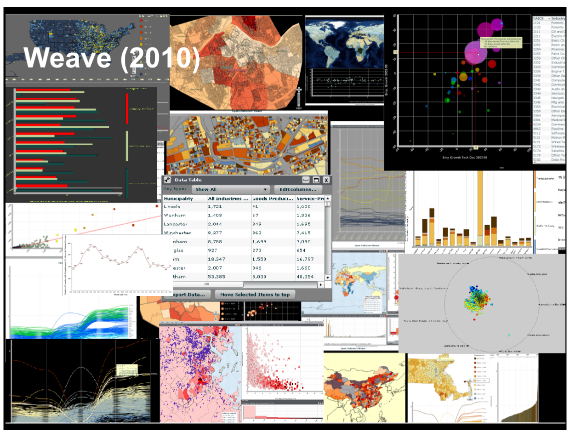
Enable data visualization and analysis
of any available data
anywhere
by anyone
for any purpose

Under administrative and user controls

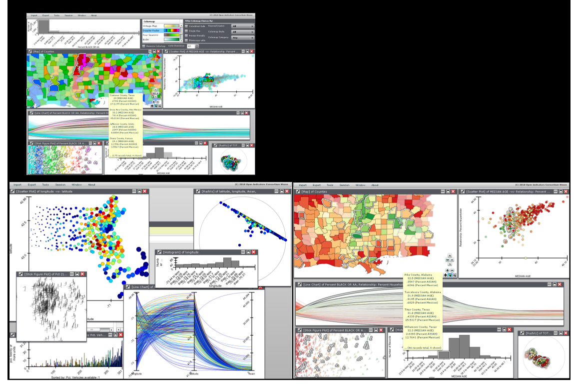
With source code and software freely available

Four Types of Visualizations

- Exploratory
  Have no hypotheses about the data
  Explore data interactively as undirected searches
- Confirmatory
  Have specific hypotheses about the data
  Goal-oriented examination of the hypotheses
- Presentation
  Facts to be presented are fixed a priori
  Select appropriate presentation techniques
- Interactive Presentation
  Interactions with a pre-defined animation

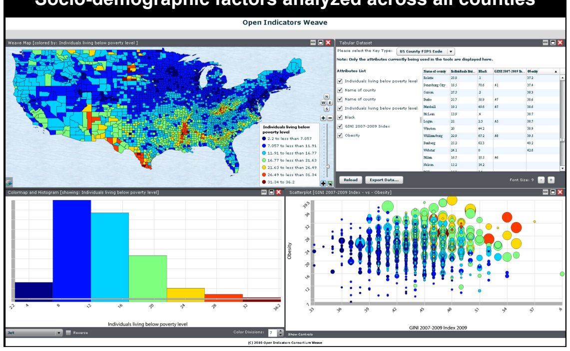


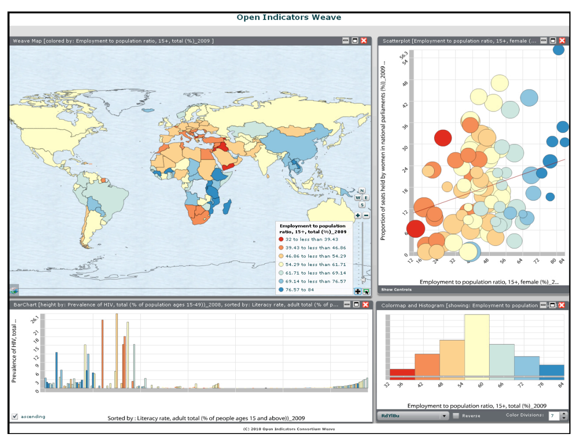
Weave: Data Visualization Software

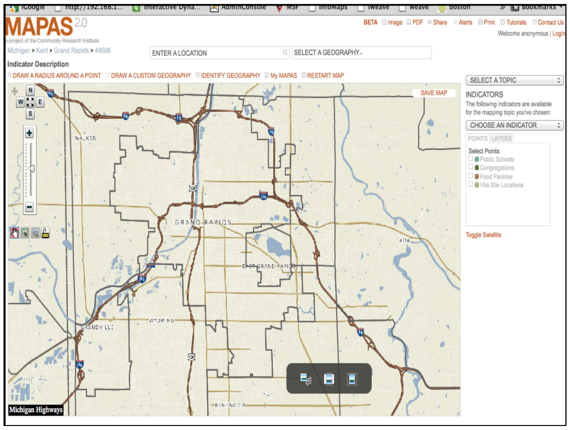


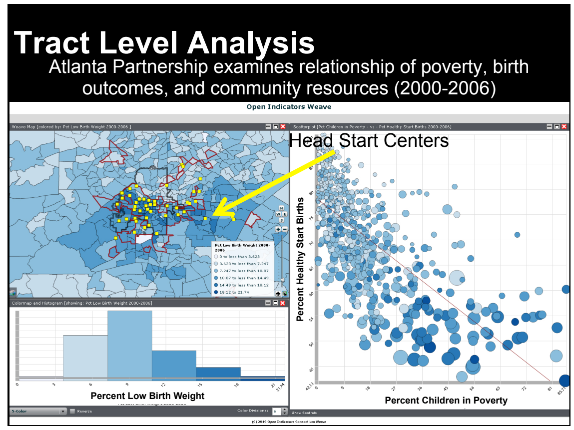
County Health Status

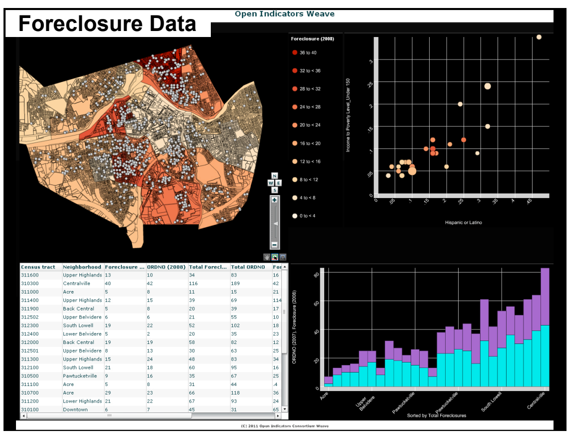
Correlating and Benchmarking – Obesity, Diabetes, and Socio-demographic factors analyzed across all counties



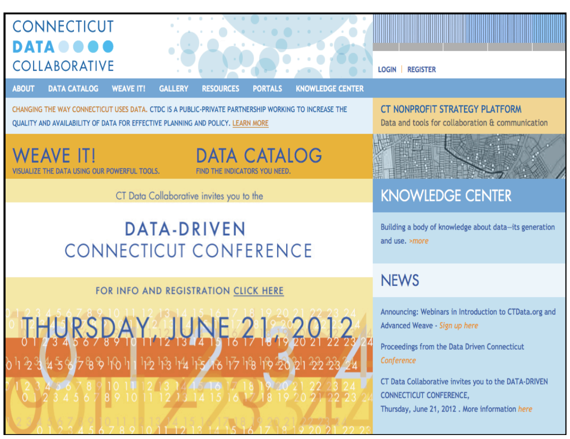


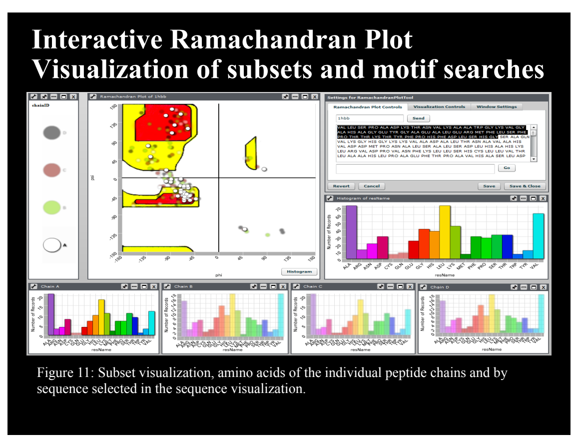

MAPAS 2.0
ENTER A LOCATION
SELECT A GEOGRAPHY
SELECT A TOPIC
INDICATORS
CHOOSE AN INDICATOR
SAVE MAP
Toggle Satellite
Michigan Highways


Tract Level Analysis
Atlanta Partnership examines relationship of poverty, birth outcomes, and community resources (2000-2006)
Open Indicators Weave
Head Start Centers
Percent Healthy Start Births
Percent Low Birth Weight
Percent Children in Poverty

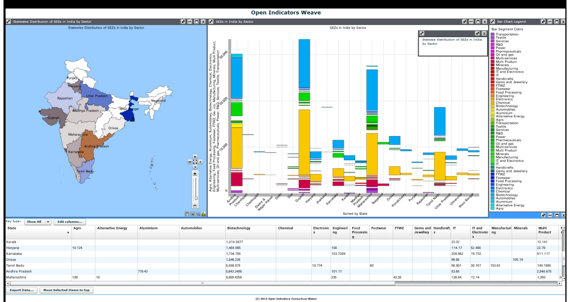

Foreclosure Data
Open Indicators Weave
Hispanic or Latino
Sorted by Total Foreclosures


ABOUT THE PROJECT
MetroBoston DataCommon
EXPLORE DATA:
COMMUNITY SNAPSHOTS:
DO-IT-YOURSELF:
RESOURCES:
LEARNING COMMUNITY:
Join Us at Data Day 2012!
A free one-day conference on January 27
START A DATA SEARCH:
Explore Data by Topics:
Introducing WEAVE...
Map: Children in Families below Poverty (Census Tracts)
Hands-on Training Available

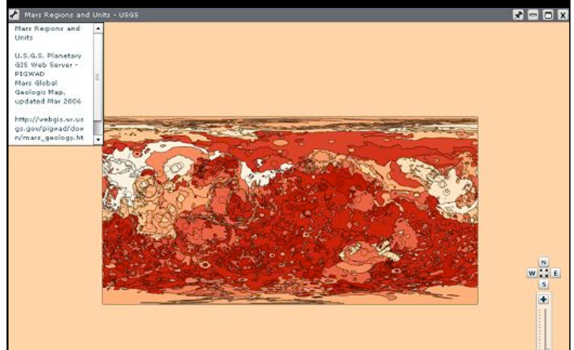

CONNECTICUT
DATA
COLLABORATIVE
LOGIN | REGISTER
ABOUT
DATA CATALOG
WEAVE IT!
GALLERY
RESOURCES
PORTALS
KNOWLEDGE CENTER
CHANGING THE WAY CONNECTICUT USES DATA. CTDC IS A PUBLIC-PRIVATE PARTNERSHIP WORKING TO INCREASE THE QUALITY AND AVAILABILITY OF DATA FOR EFFECTIVE PLANNING AND POLICY. LEARN MORE
CT NONPROFIT STRATEGY PLATFORM
Data and tools for collaboration & communication
WEAVE IT!
VISUALIZE THE DATA USING OUR POWERFUL TOOLS.
DATA CATALOG
FIND THE INDICATORS YOU NEED.
KNOWLEDGE CENTER
CT Data Collaborative invites you to the
DATA-DRIVEN
CONNECTICUT CONFERENCE
Building a body of knowledge about data--its generation and use. +more
FOR INFO AND REGISTRATION CLICK HERE
NEWS
THURSDAY, JUNE 21, 2012
Announcing: Webinars in Introduction to CTData.org and Advanced Weave - Sign up here
Proceedings from the Data Driven Connecticut Conference
CT Data Collaborative invites you to the DATA-DRIVEN CONNECTICUT CONFERENCE,
Thursday, June 21, 2012 . More information here


Interactive Ramachandran Plot
Visualization of subsets and motif searches
Figure 11: Subset visualization, amino acids of the individual peptide chains and by sequence selected in the sequence visualization.

Weave: India Free Zones



Weave: Mars



http://demo.oicweave.org/weave.html?defaults=Mars1.xml

Version 1.0 Release

- Multiple visualizations in web browser (a full system)
- Highly interactive
  necessary expectations (goal: 20s download, 1s interaction)
- Client and host APIs
  facilitate integration with web frameworks
  equation and R-script editor
- Data and Geometry can be on server OR client OR hybrid
- Metadata support (Dublin Core)
- R-Project, Weka, OpenMap, and Geoserver peer services
- Administrative control over layout, content and interactions

- All based on session state/history

Institute for Visualization and Perception Research

UMASS

Demonstrations

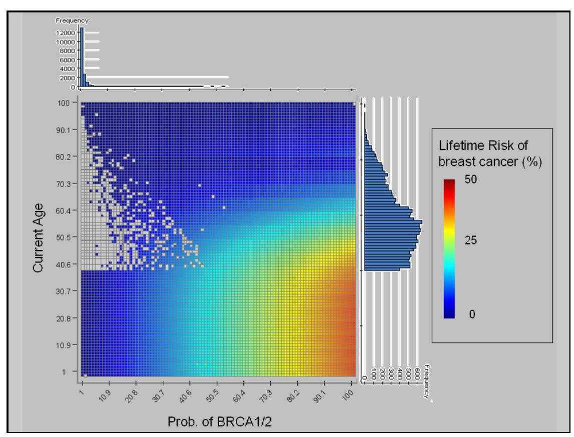
information visualized
knowledge discovered
decisions made

Institute for Visualization and Perception Research

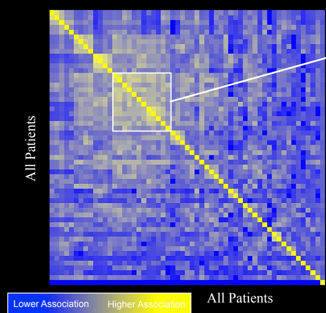
UMASS

Modern Custom Visualizations

information visualized
knowledge discovered
decisions made



Identification of Patient Groups



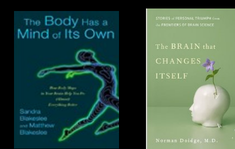
Drug Correlation Cost Function (over all drugs)

$C(p_i, p_j) = \Sigma f(drug_i)$

Institute for Visualization and Perception Research

Futures Areas for Visualization

Plasticity
- **brain**
- **body**



information visualized
knowledge discovered
decisions made

Institute for Visualization and Perception Research

Slow Physical Plasticity: Genetics
Rapid Physical Plasticity: Bionics

information visualized
knowledge discovered
decisions made

Institute for Visualization and Perception Research

Last New Grand Challenge

Anonymity – protecting privacy for open data yet allowing analyses

information visualized
knowledge discovered
decisions made

Protecting Privacy

- Grand Challenge
- Given a data set D
- Generate a data set D’ where

stat (D’) = stat (D)
pattern (D’) = pattern (D)
mineAlgorithm (D’) = mineAlgorithm (D)
if x in D’ then cannot identify x in D

Institute for Visualization and Perception Research

Questions?

information visualized
knowledge discovered
decisions made