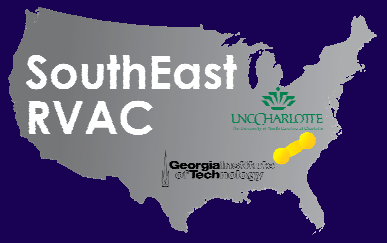
Sept. 17, 2008

Visualization for Information Analysis and Exploration

John Stasko

Information Interfaces Research Group
School of Interactive Computing
Georgia Institute of Technology



Georgia Tech College of Computing
School of Interactive Computing

Exercise

- Get out pencil & paper

information interfaces

Data Explosion

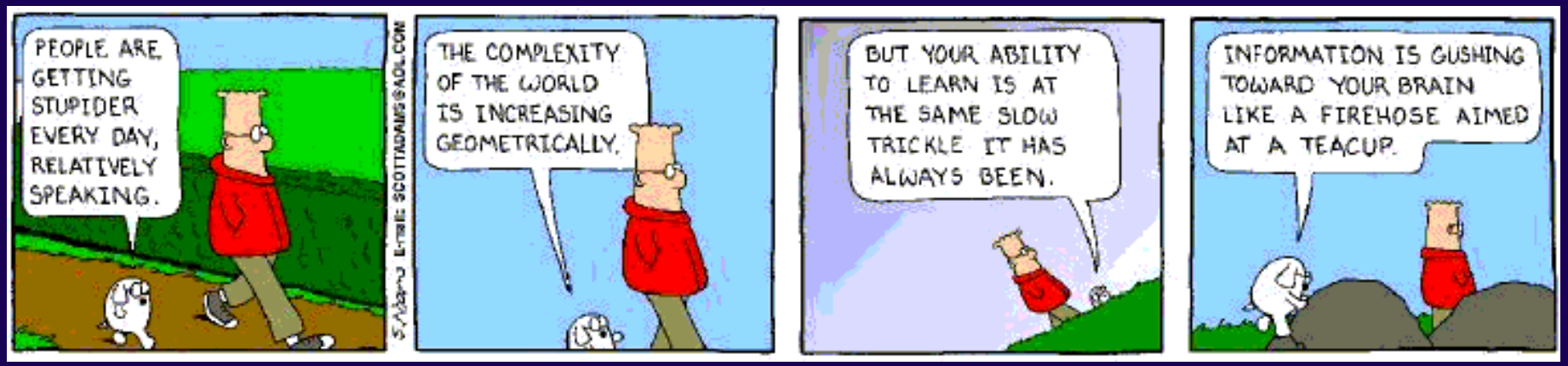
- Society is more complex
  - There simply is more "stuff"

- Computers, internet and web give people access to an incredible amount of data
  - news, sports, financial, purchases, etc...

information interfaces

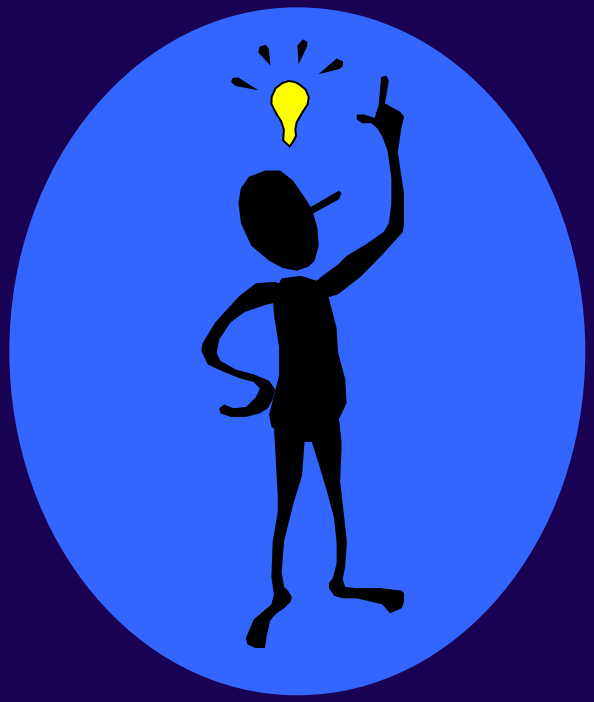
Data Overload

- Confound: How to make use of the data
  - How do we make sense of the data?
  - How do we harness this data in decision-making processes?
  - How do we avoid being overwhelmed?



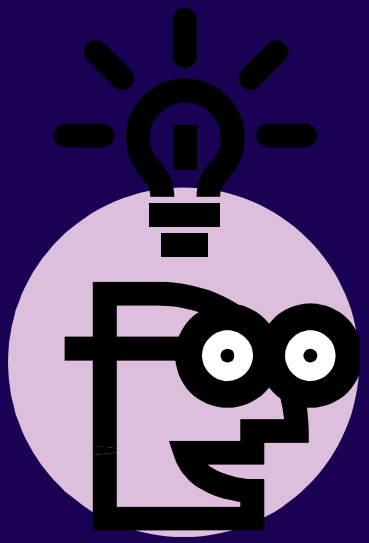
The Challenge

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



Premise of my Work

- Visualization of data helps people understand it better



Human Vision

- Highest bandwidth sense
  - ~100 MB/s
  - Parallel
  - Strong pattern recognition
  - Much done preattentively, ie, without thought

Visualization

- Definition
  - "The use of computer-supported, interactive visual representations of data to amplify cognition."
    - From [Card, Mackinlay Shneiderman '98]

Visualization

- Often thought of as process of creating a graphic or an image
- Really is a cognitive process
  - Form a mental image of something
  - Internalize an understanding
- "The purpose of visualization is insight, not pictures"
  - Insight: discovery, decision making, explanation, analysis, exploration, learning

Main Idea

- Visuals help us think
  - Provide a frame of reference, a temporary storage area
- Cognition $\rightarrow$ Perception
- Pattern matching
- External cognition aid
  - Role of external world in thinking and reason

Larkin & Simon '87
Card, Mackinlay, Shneiderman '98

When to Apply?

- Many other techniques for data analysis
  - Data mining, DB queries, machine learning…
- Visualization most useful in **exploratory data analysis**
  - Don't know what you're looking for
  - Don't have a priori questions
  - Want to know what questions to ask

Part of our Culture

- "I see what you're saying"
- "Seeing is believing"
- "A picture is worth a thousand words"

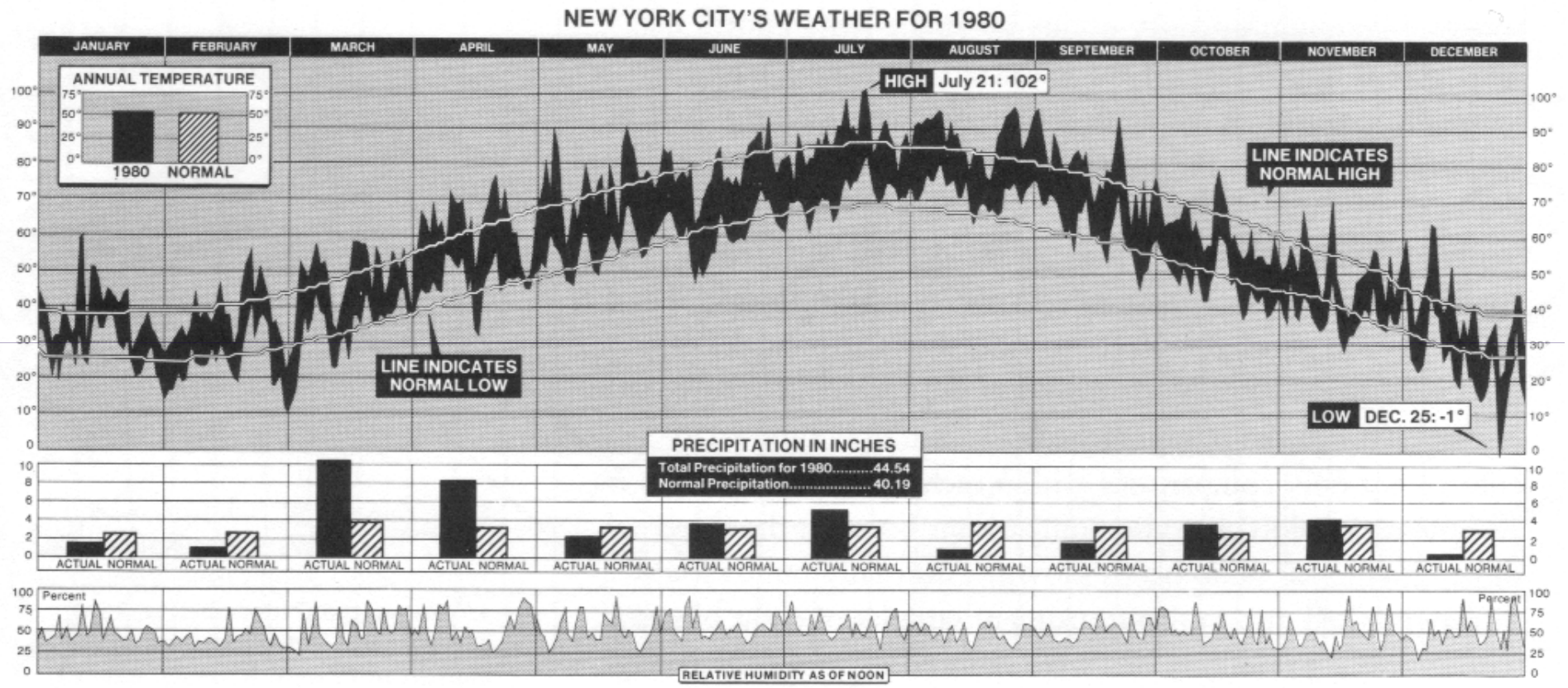
information interfaces

Some quick (static) examples…

information interfaces

NYC Weather

2220 numbers



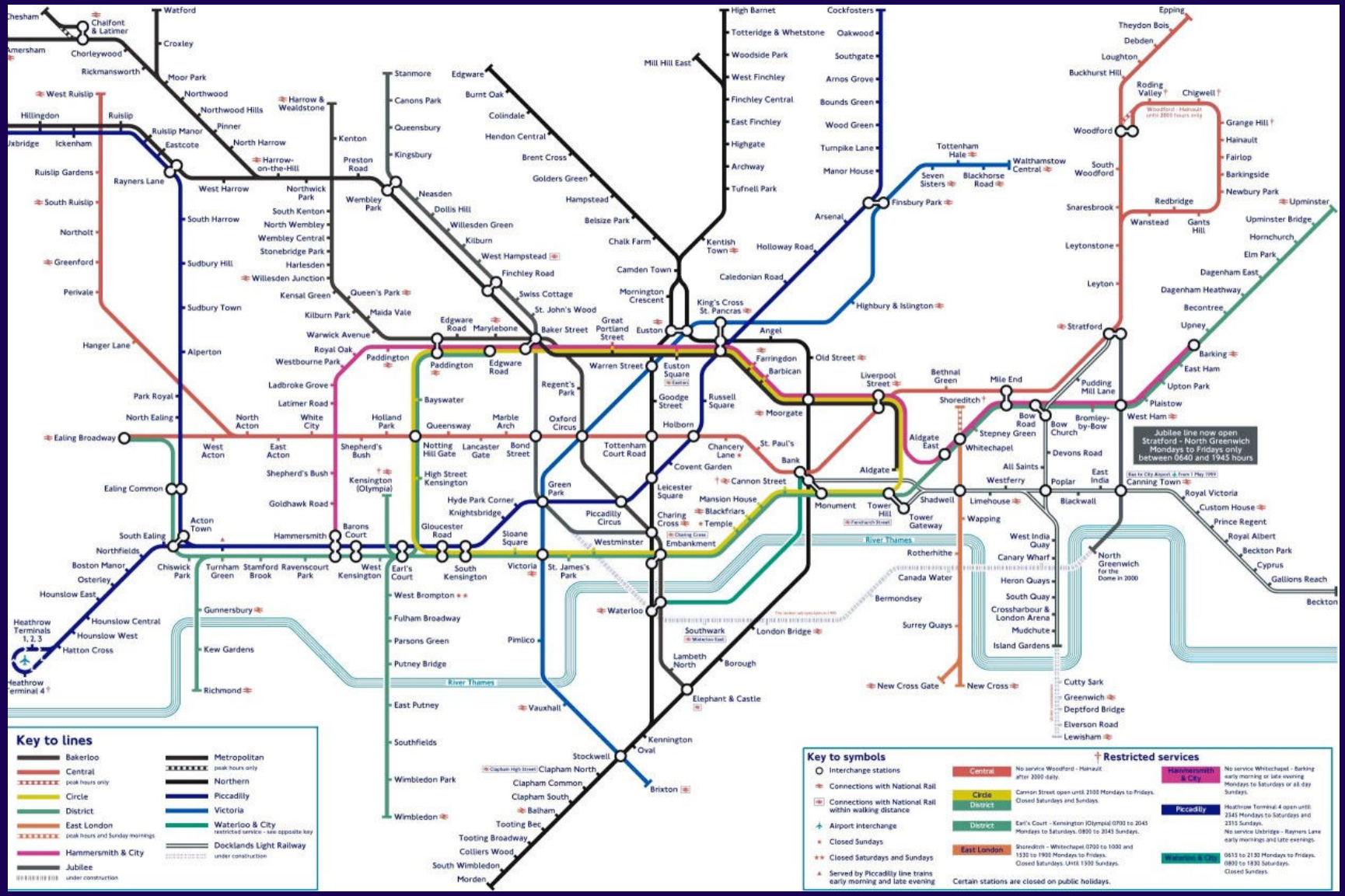
New York Times, January 11, 1981, p. 32.

E. Tufte, *Visual Display of Quant Info*

information interfaces

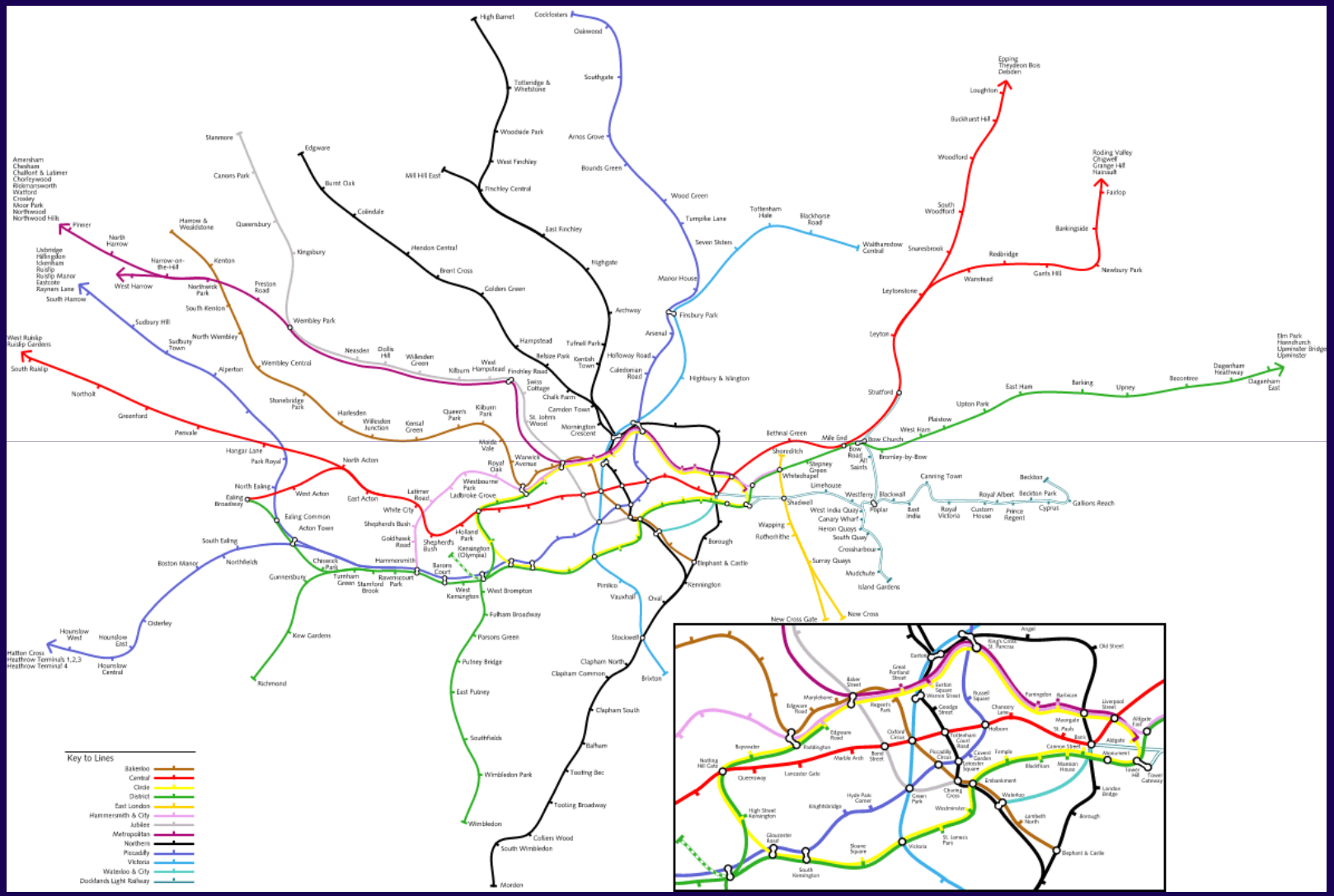
London Subway

www.thetube.com



www.kottke.org/plus/misc/images/tubegeo.gif

True Geography

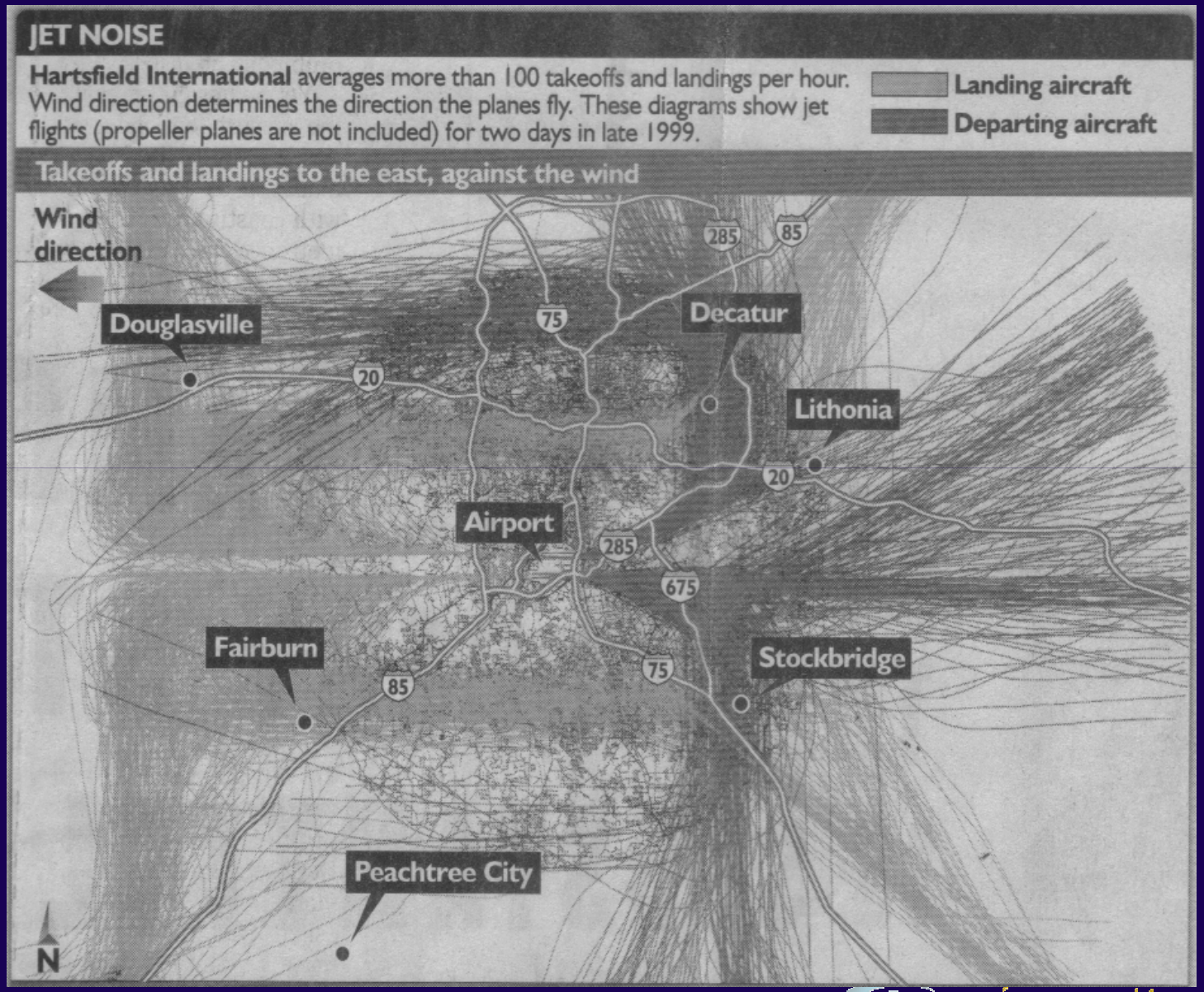


rodcorp.typepad.com/photos/art_2003/tube_walklines_final_lmfaint.html

Easy Walking Lines Added

information interfaces

Atlanta Flight Traffic



Atlanta Journal
April 30, 2000

InfoVis '07

Casual Information Visualization: Depictions of Data in Everyday Life

Zachary Pousman, John T. Stasko, *Member, IEEE*, and Michael Mateas

Abstract—Information visualization has often focused on providing deep insight for expert user populations and on techniques for amplifying cognition through complicated interactive visual models. This paper proposes a new subdomain for infovis research that complements the focus on analytic tasks and expert use. Instead of work-related and analytically driven infovis, we propose Casual Information Visualization (or Casual Infovis) as a complement to more traditional infovis domains. Traditional infovis systems, techniques, and methods do not easily lend themselves to the broad range of user populations, from expert to novices, or from work tasks to more everyday situations. We propose definitions, perspectives, and research directions for further investigations of this emerging subfield. These perspectives build from ambient information visualization [32], social visualization, and also from artistic work that visualizes information [41]. We seek to provide a perspective on infovis that integrates these research agendas under a coherent vocabulary and framework for design. We enumerate the following contributions. First, we demonstrate how blurry the boundary of infovis is by examining systems that exhibit many of the putative properties of infovis systems, but perhaps would not be considered so. Second, we explore the notion of insight and how, instead of a monolithic definition of insight, there may be multiple types, each with particular characteristics. Third, we discuss design challenges for systems intended for casual audiences. Finally we conclude with challenges for system evaluation in this emerging subfield.

Index Terms—Casual information visualization, ambient infovis, social infovis, editorial, design, evaluation.

1 Introduction

Much of the work in information visualization assumes a population of expert users who have knowledge and experience in analyzing problems in specific domains. Workers in widely varying domains from finance to government to journalism use information visualization tools to explore data, generate, refine and test hypotheses, and ultimately to produce insight. This user population of information workers and

Are these types of tools really infovis systems? The question arises, where are the limits of infovis with respect to the everyday uses of computational artifacts.

Card, Mackinlay, and Shneiderman define information visualization broadly: infovis is the use of computers to interactively amplify cognition, using visual representations [10]. Therefore if we take this as our definition, systems must be computer-based, interactive, pro-

information interfaces

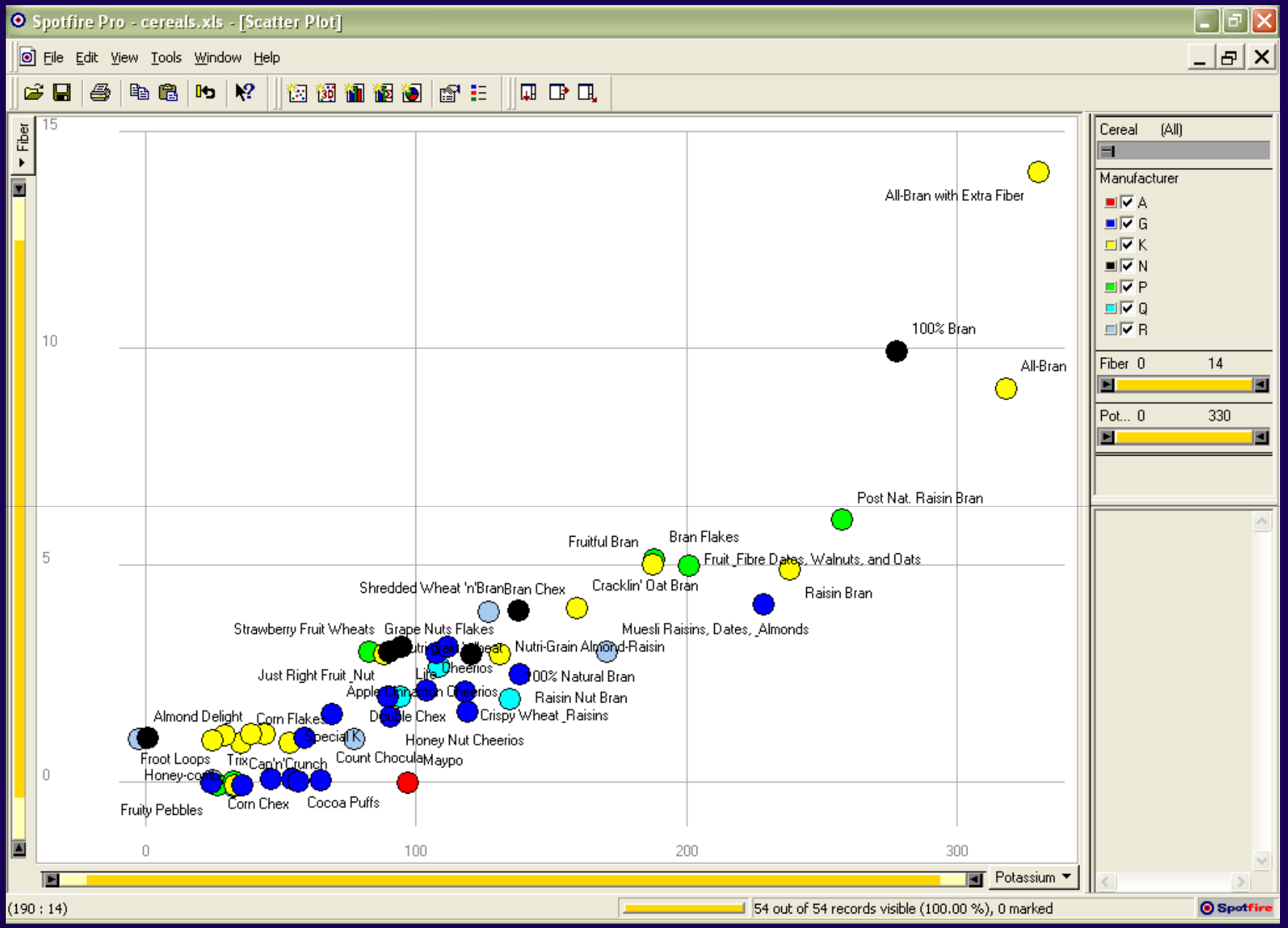
Reinforce my point with two examples

information interfaces

Questions:

Which cereal has the most/least potassium?
Is there a relationship between potassium and fiber?
If so, are there any outliers?
Which manufacturer makes the healthiest cereals?

	A	B	C	D
1	Cereal	Manufacturer	Fiber	Potassium
2	100% Bran	N	10	280
3	100% Natural Bran	Q	2	135
4	All-Bran	K	9	320
5	All-Bran with Extra Fiber	K	14	330
6	Almond Delight	R	1	0
7	Apple Cinnamon Cheeric	G	1.5	70
8	Bran Chex	R	4	125
9	Bran Flakes	P	5	190
10	Cap'n'Crunch	Q	0	35
11	Cheerios	G	2	105
12	Cocoa Puffs	G	0	55
13	Corn Chex	R	0	25
14	Corn Flakes	K	1	35
15	Count Chocula	G	0	65
16	Cracklin' Oat Bran	K	4	160
17	Cream of Wheat (Quick)	N	1	0
18	Crispy Wheat & Raisins	G	2	120
19	Double Chex	R	1	80
20	Froot Loops	K	1	30
21	Frosted Flakes	K	1	25
22	Fruit & Fibre Dates, Wal	P	5	200
23	Fruitful Bran	K	5	190
24	Fruity Pebbles	P	0	25
25	Golden Grahams	G	0	45
26	Grape Nuts Flakes	P	3	85
27	Honey Nut Cheerios	G	1.5	90
28	Honey-comb	P	0	35
29	Just Right Fruit & Nut	K	2	95
30	Life	Q	2	95
31	Lucky Charms	G	0	55
32	Maypo	A	0	95
33	Muesli Raisins, Dates, &	R	3	170
34	Multi-Grain Cheerios	G	2	90
35	Nutri-Grain Almond-Rais	K	3	130
36	Nutri-grain Wheat	K	3	90
37	Oatmeal Raisin Crisp	G	1.5	120
38	Post Nat. Raisin Bran	P	6	260
39	Product 19	K	1	45
40	Quaker Oatmeal	Q	2.7	110
41	Raisin Bran	K	5	240
42	Raisin Nut Bran	G	2.5	140
43	Rice Krispies	K	0	35
44	Shredded Wheat	N	3	95
45	Shredded Wheat 'n'Bran	N	4	140
46	Shredded Wheat spoon	N	3	120
47	Smacks	K	1	40
48	Special K	K	1	55
49	Strawberry Fruit Wheats	N	3	90
50	Total Corn Flakes	G	0	35
51	Total Raisin Bran	G	4	230
52	Total Whole Grain	G	3	110
53	Trix	G	0	25
54	Wheaties	G	3	110
55	Wheaties Honey Gold	G	1	60



Fiber

Potassium

Even Tougher?

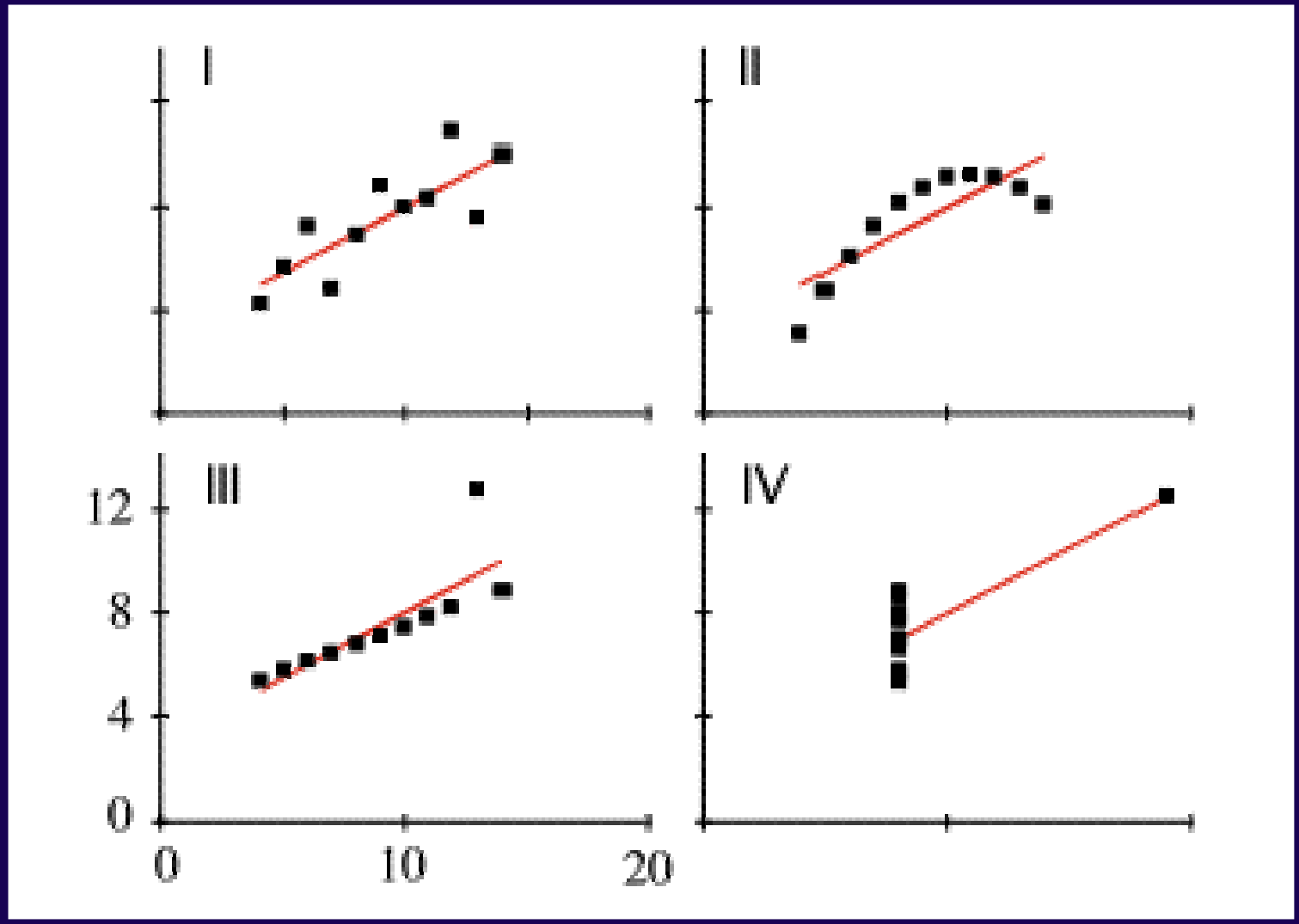
- What if you could only see one cereal's data at a time? (e.g. some websites)
- What if I read the data to you?

Four Data Sets

- Mean of the x values = 9.0
- Mean of the y values = 7.5
- Equation of the least-squared regression line is: y = 3 + 0.5x
- Sums of squared errors (about the mean) = 110.0
- Regression sums of squared errors (variance accounted for by x) = 27.5
- Residual sums of squared errors (about the regression line) = 13.75
- Correlation coefficient = 0.82
- Coefficient of determination = 0.67

http://astro.swarthmore.edu/astro121/anscombe.html

The Data Sets



information interfaces

The Values

1	2	3	4
10.0, 8.04	10.0,9.14	10.0, 7.46	8.0, 6.58
8.0, 6.95	8.0,8.14	8.0, 6.77	8.0, 5.76
13.0, 7.58	13.0,8.74	13.0,12.74	8.0, 7.71
9.0, 8.81	9.0,8.77	9.0, 7.11	8.0, 8.84
11.0, 8.33	11.0,9.26	11.0, 7.81	8.0, 8.47
14.0, 9.96	14.0,8.10	14.0, 8.84	8.0, 7.04
6.0, 7.24	6.0,6.13	6.0, 6.08	8.0, 5.25
4.0, 4.26	4.0,3.10	4.0, 5.39	19.0,12.50
12.0,10.84	12.0,9.13	12.0, 8.15	8.0, 5.56
7.0, 4.82	7.0,7.26	7.0, 6.42	8.0, 7.91
5.0, 5.68	5.0,4.74	5.0, 5.73	8.0, 6.89

Revisit Starting Exercise

- What did you put on paper?

information interfaces

Two Related Disciplines

- Information Visualization
- Visual Analytics

Information Visualization

- Using interactive computer visualizations to represent and communicate abstract data
  - Statistics, databases, software, …
- Area emerged approximately 1990

Information Visualization

- Recent research trends
  - InfoVis for the Masses
  - Challenges of evaluation
  - Interaction is crucial

Visual Analytics

- Informal: Using visual representations to help make decisions
- Formal: The science of analytical reasoning facilitated by interactive visual interfaces
- InfoVis++
- Area emerged approximately 2005

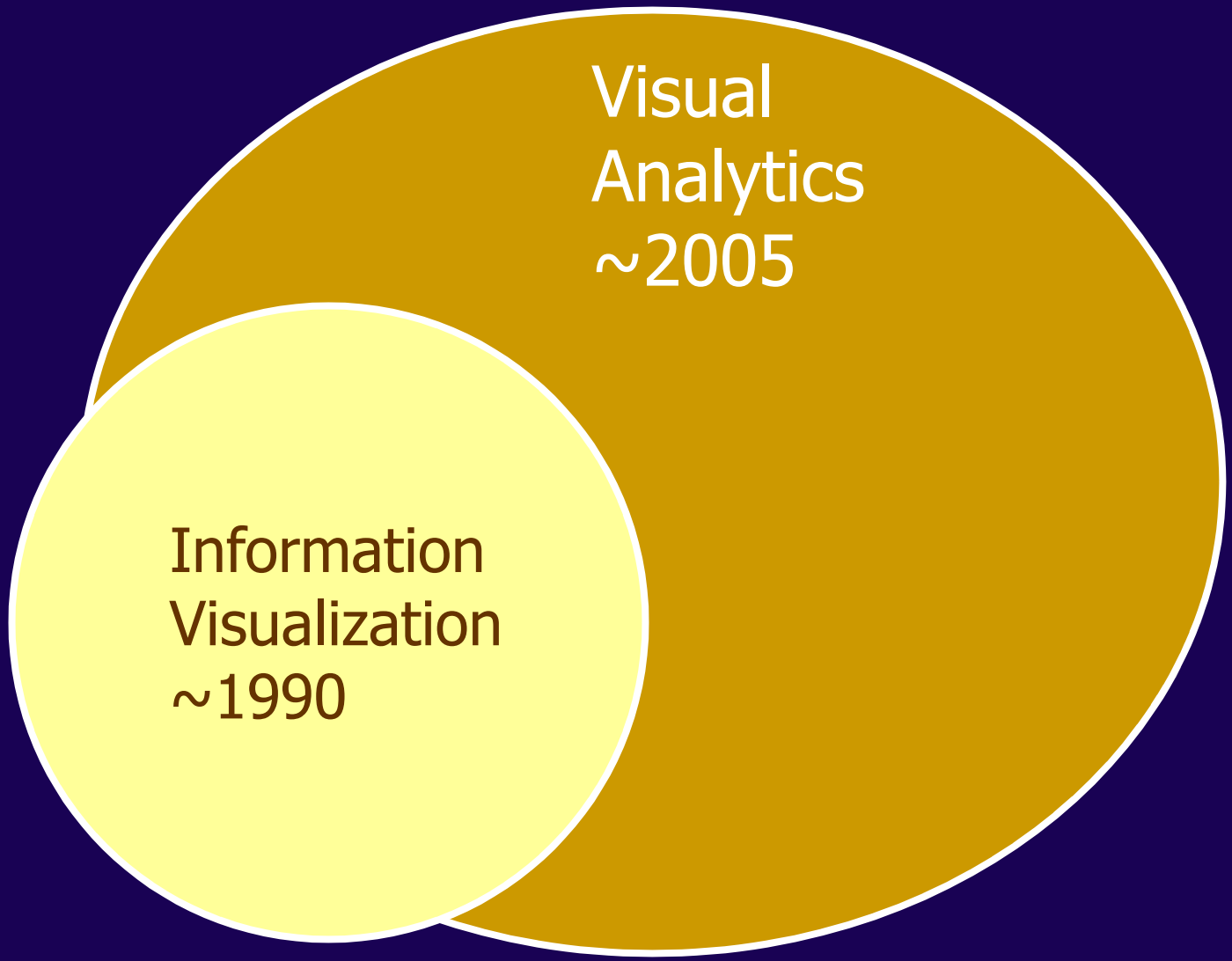
Overview of the R&D Agenda

- Challenges
- Science of Analytical Reasoning
- Science of Visual Representations and Interactions
- Data Representations and Transformations
- Production, Presentation, and Dissemination
- Moving Research Into Practice
- Positioning for an Enduring Success

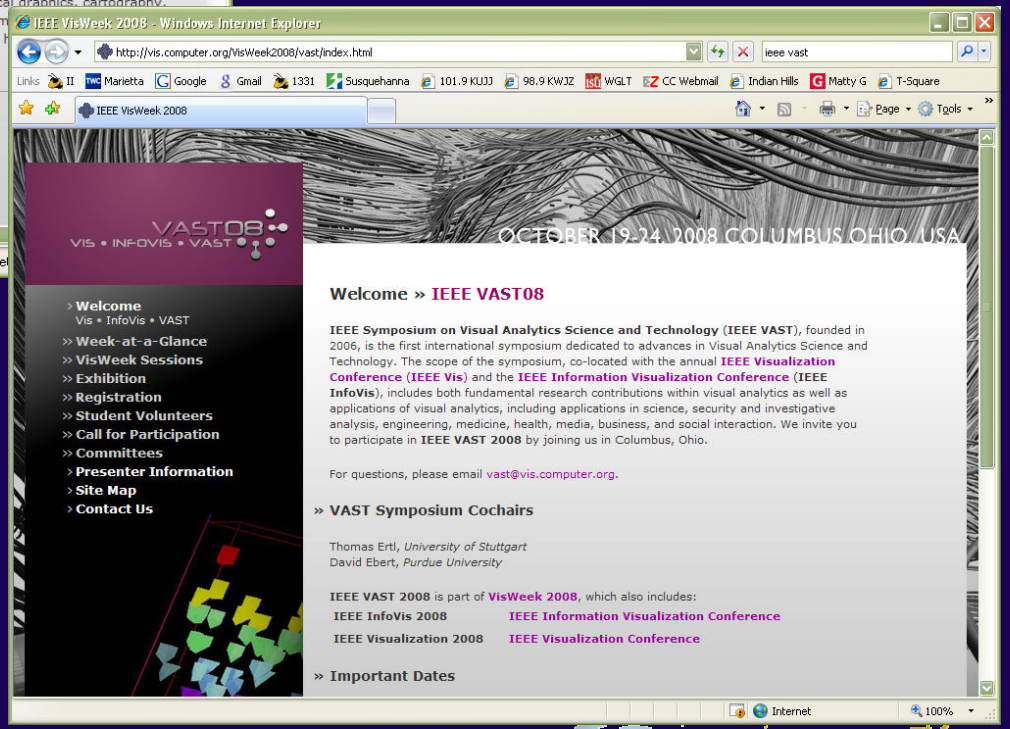
Visual Analytics: Beyond InfoVis

- Statistics, data representation and statistical graphics
- Geospatial and Temporal Sciences
- Applied Mathematics
- Knowledge representation, management and discovery
- Ontology, semantics, NLP, extraction, synthesis, …
- Cognitive and Perceptual Sciences
- Comunications: Capture, Illustrate and present a message
- Decision sciences

Academic Context



IEEE InfoVis



IEEE VAST

Sensemaking

"A motivated , continuous effort to understand connections (which can be among people, places, and events) in order to anticipate their trajectories and act effectively."

– Klein, Moon and Hoffman

Jigsaw

- Visualization for Investigative Analysis across Document Collections

information interfaces

The Jigsaw Team

Carsten Görg
Zhicheng Liu
Vasili Pantazopoulos
+ 4 new students

Gennadiy Stepanov
Sarah Williams
Neel Parekh
Kanupriyah Singhal

information interfaces

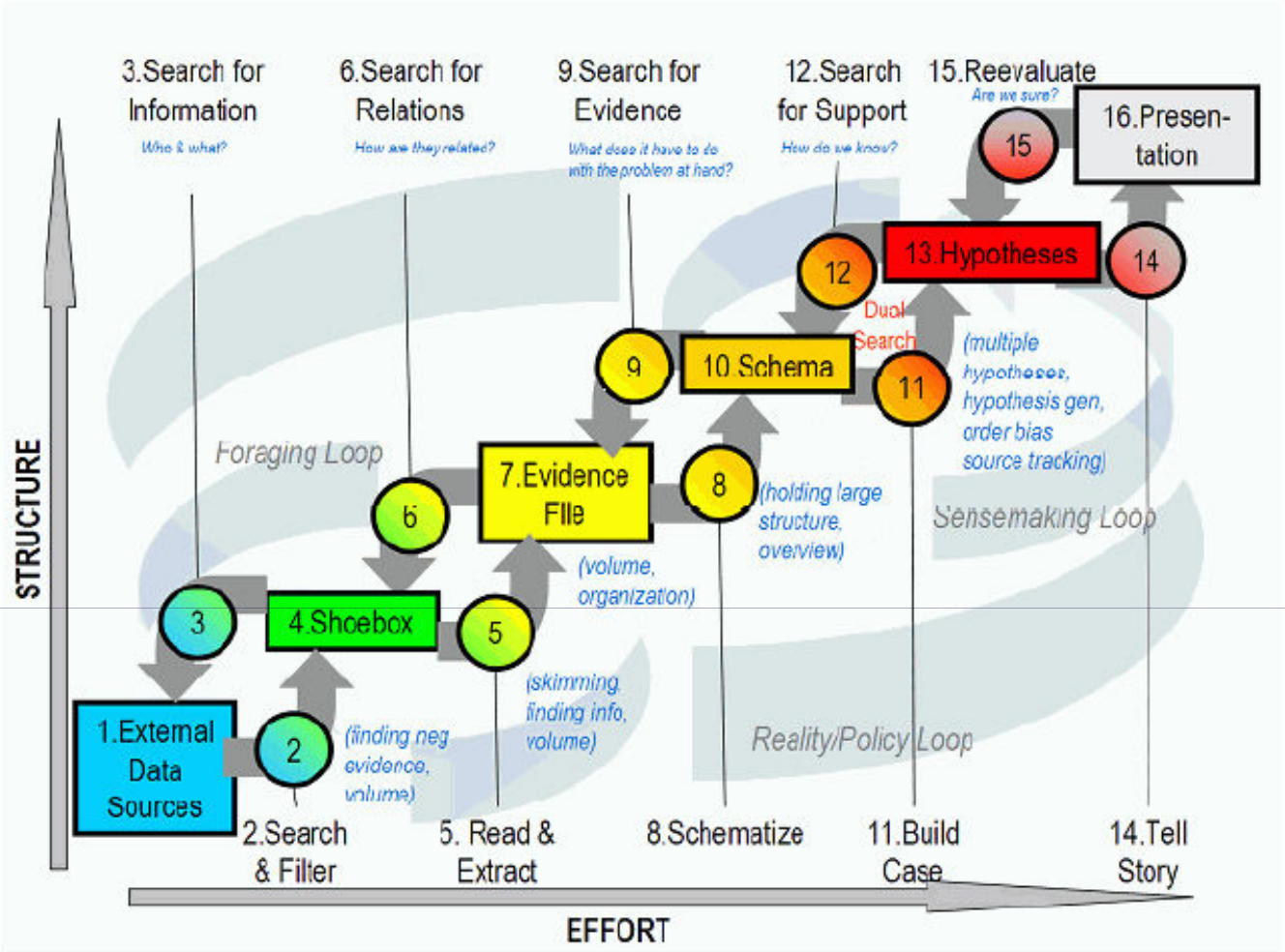


Figure 2.1. Notional model of sensemaking loop for intelligence analysis derived from CTA.

Pirolli & Card, ICIA '05

information interfaces

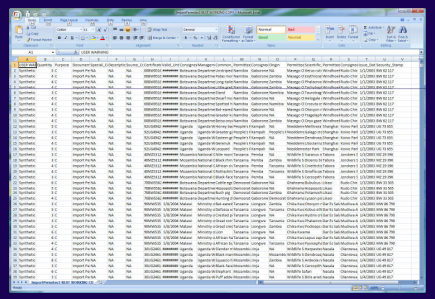
Pain Points

- Cost structure of scanning and selecting items for further attention
- Analysts' span of attention for evidence and hypotheses

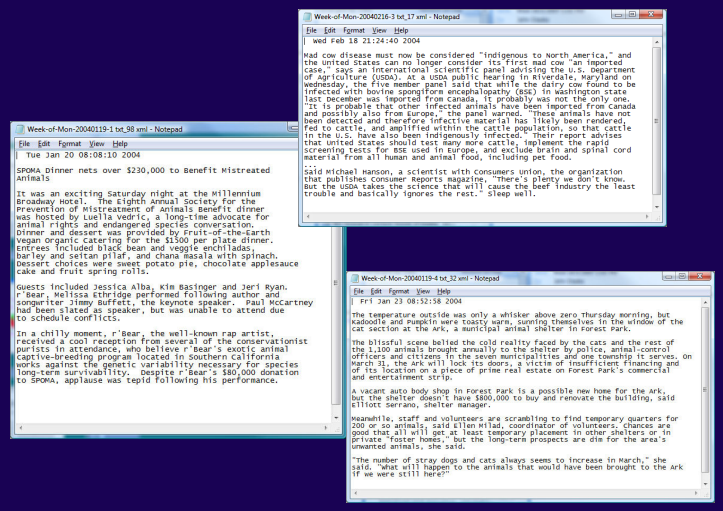
Problem Addressed

•Analogy

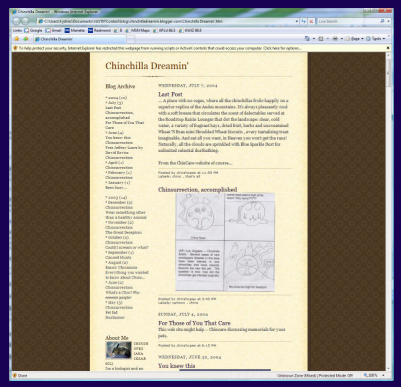
- Help investigative analysts discover plans, plots and threats embedded across the individual documents in large document collections



DBs



Documents/
case reports



Blogs

information interfaces

Example Document

Report: 20040510-4_16
May 14 2004

VANCOUVER, British Columbia - A Canadian immigration panel is considering whether accused environmental saboteur Tre Arrow can apply for refugee status in Canada.

Arrow, 30, who is wanted for fire bombing logging and cement trucks in Oregon, asked the Canadian authorities to remain in Canada as a political refugee at a hearing in Vancouver on Tuesday.

A key issue will be whether Arrow is affiliated with a terrorist group, which would immediately disqualify him from receiving refugee status in Canada, authorities said.

The Immigration and Refugee Board is scheduled to decide by May 31 whether Arrow is affiliated with the Earth Liberation Front, a group the FBI considers a terrorist organization responsible for scores of attacks on property over the past dozen years.

Our Focus

- Entities within the documents
  - Person, place, organization, phone number, date, license plate, etc.
- Thesis: A plot/threat within the documents will involve a set of entities in coordination

Entity Identification

- Must identify and extract entities from plain text documents
  - Crucial for our work
- Not our main research focus – Collaborate with or use tools from others

information interfaces

Entities Identified

Source:
Date: May 14, 2004

VANCOUVER, British Columbia - A Canadian immigration panel is considering whether accused environmental saboteur Tre **Arrow** can apply for refugee status in Canada.

Arrow, 30, who is wanted for fire bombing logging and cement trucks in Oregon, asked the Canadian authorities to remain in Canada as a political refugee at a hearing in Vancouver on Tuesday.

A key issue will be whether **Arrow** is affiliated with a terrorist group, which would immediately disqualify him from receiving refugee status in Canada, authorities said.

The Immigration and Refugee Board is scheduled to decide by May 31 whether **Arrow** is affiliated with the Earth Liberation Front, a group the FBI considers a terrorist organization responsible for scores of attacks on property over the past dozen years.

Connections

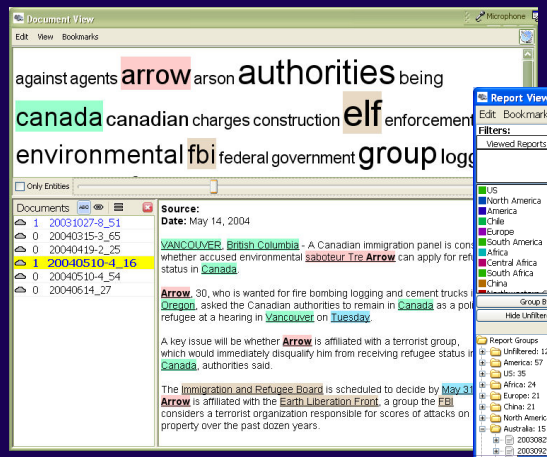
- Entities relate/connect to each other to make a larger "story"
- Connection definition:
  - Two entities are connected if they appear in a document together
  - The more documents they appear in together, the stronger the connection

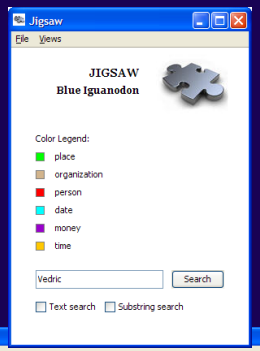
Jigsaw

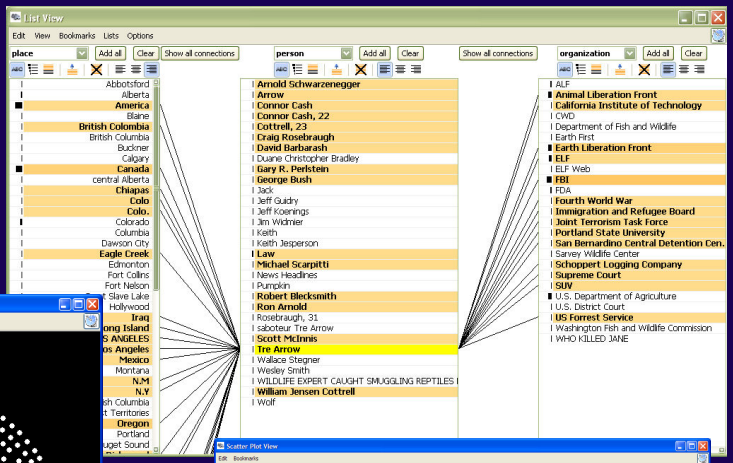
"Putting the pieces together"

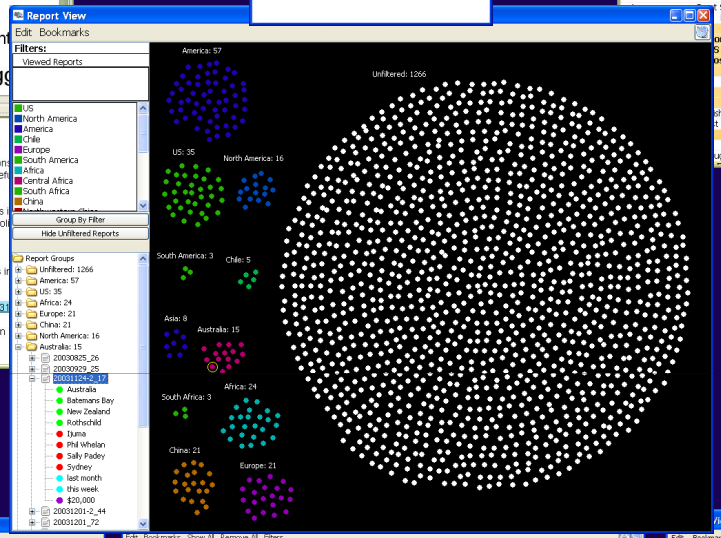
- Multiple visualizations (views) of documents, entities, & their connections
- Views are highly interactive and coordinated
- User actions generate events that are transmitted to and (possibly) reflected in other views

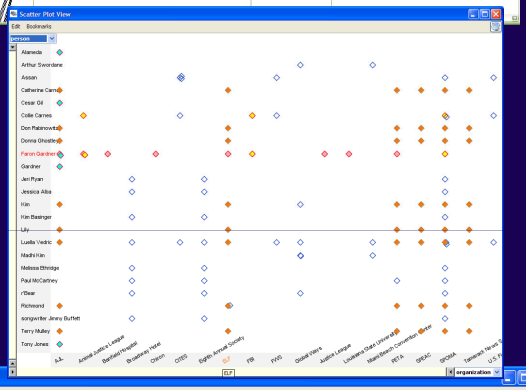
System Views

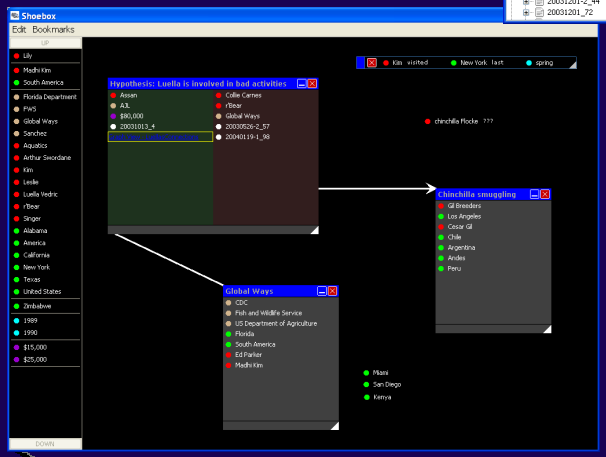


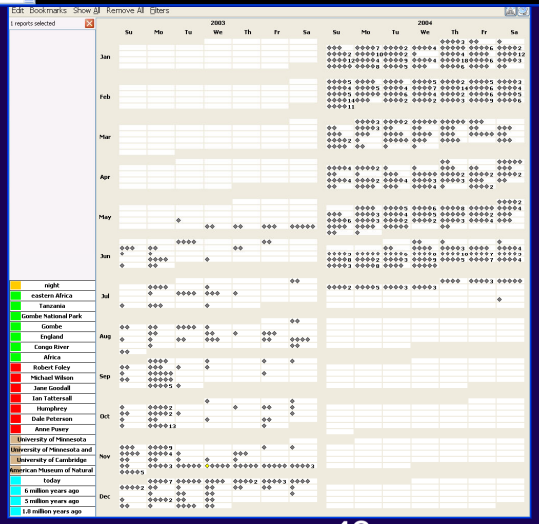


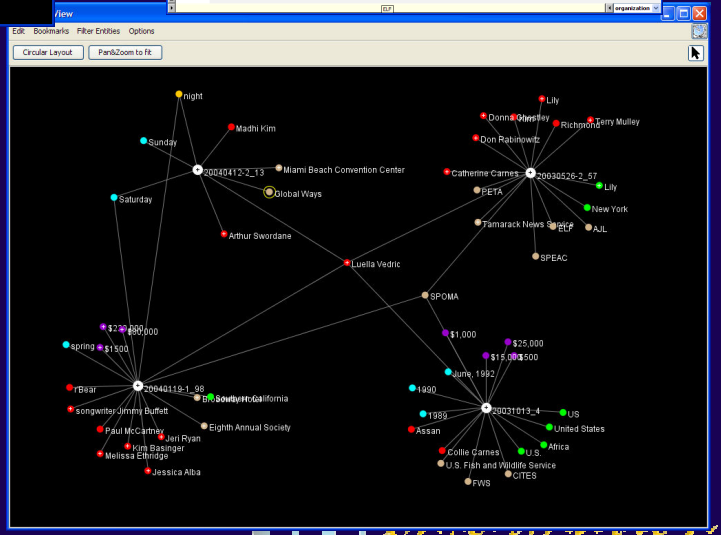




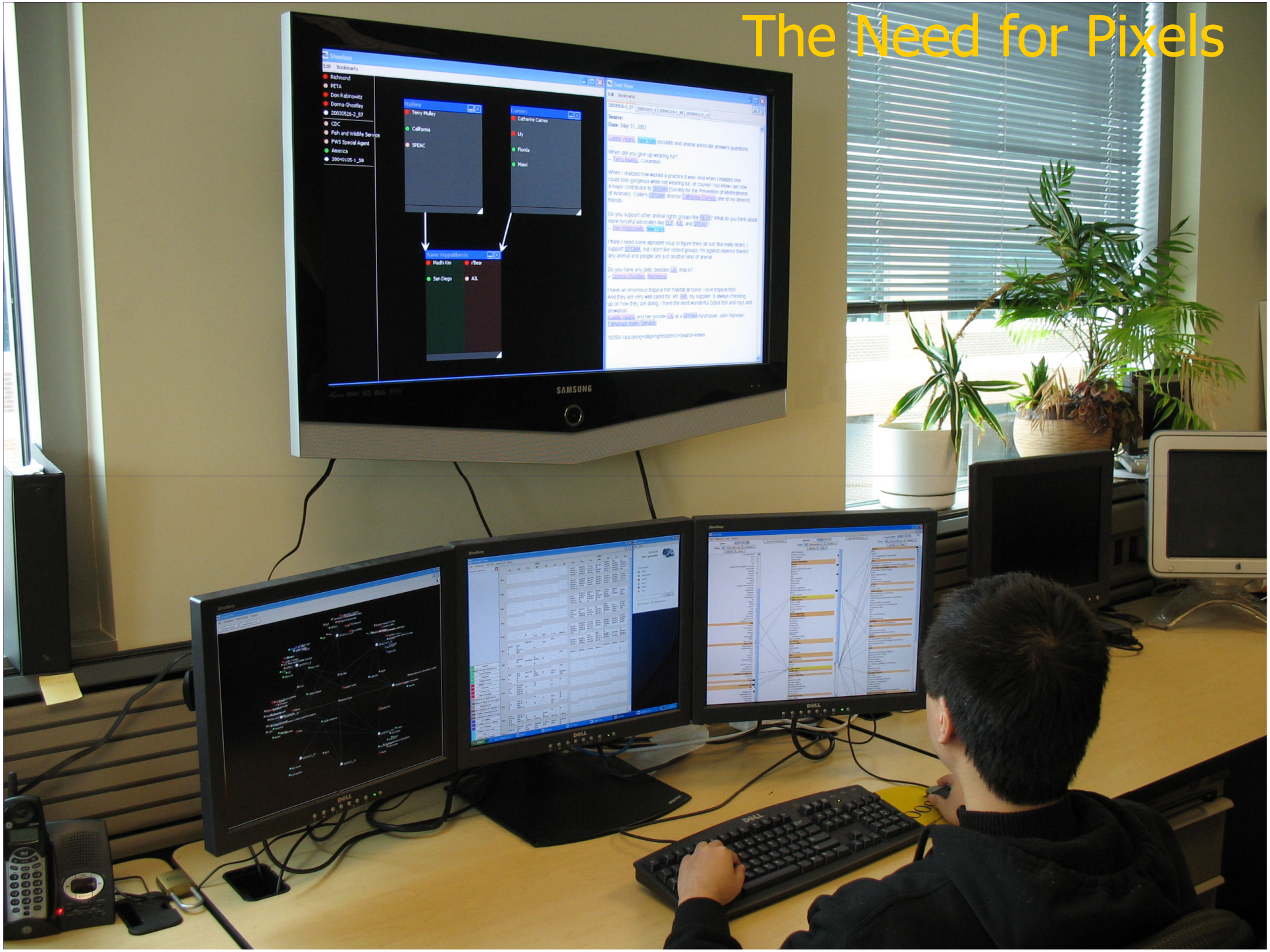








The Need for Pixels



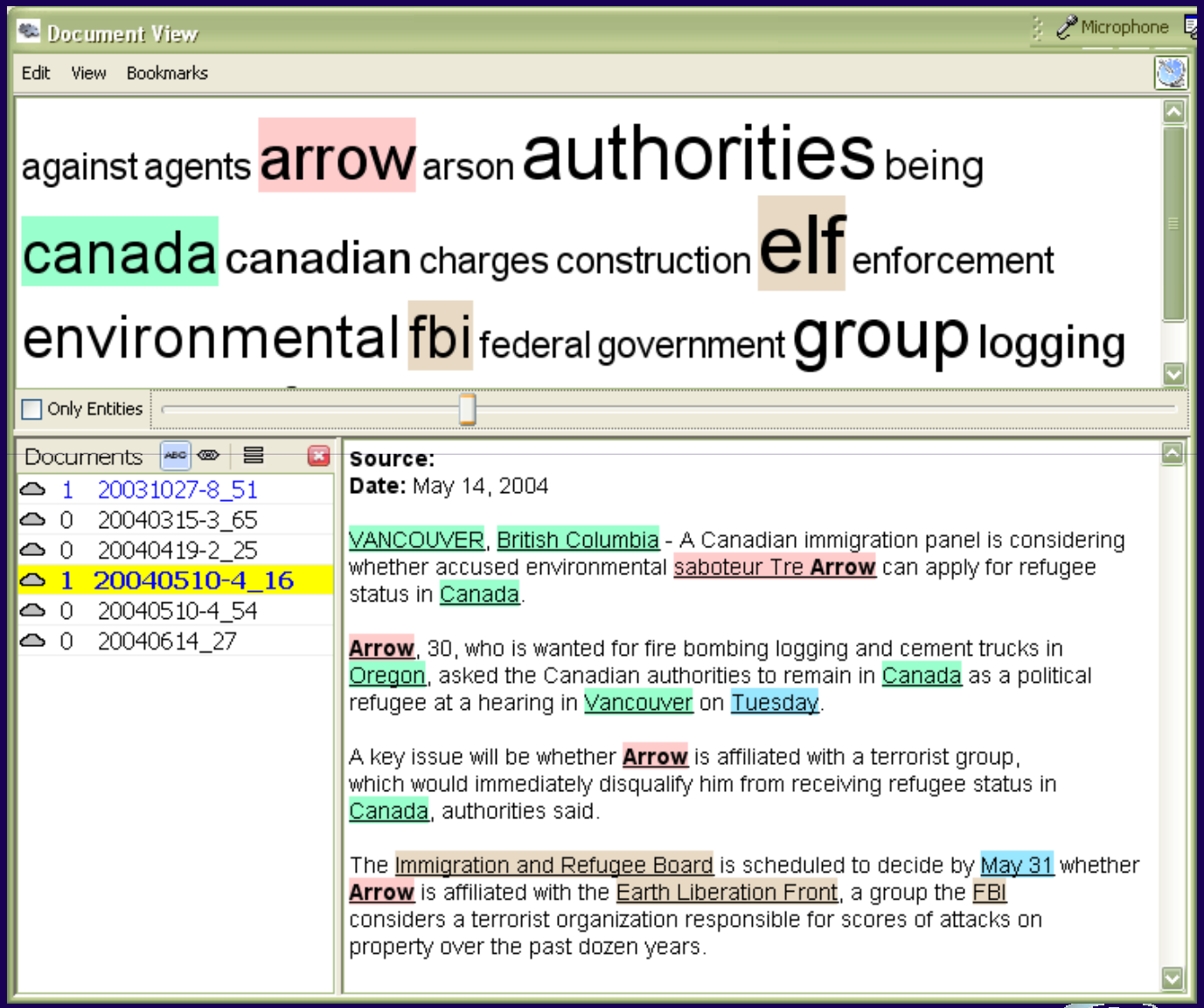
Demo

information interfaces

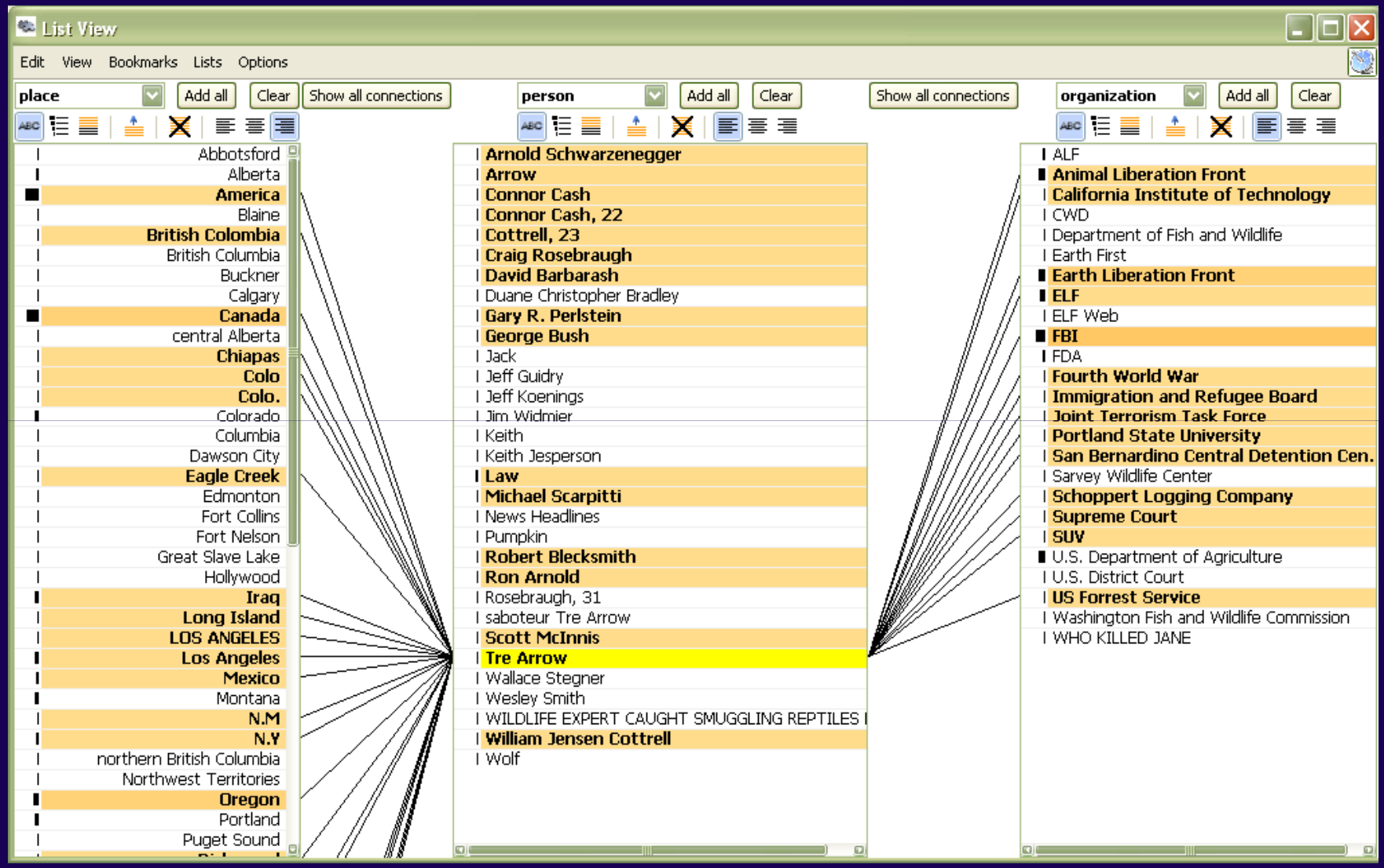
Console

information interfaces

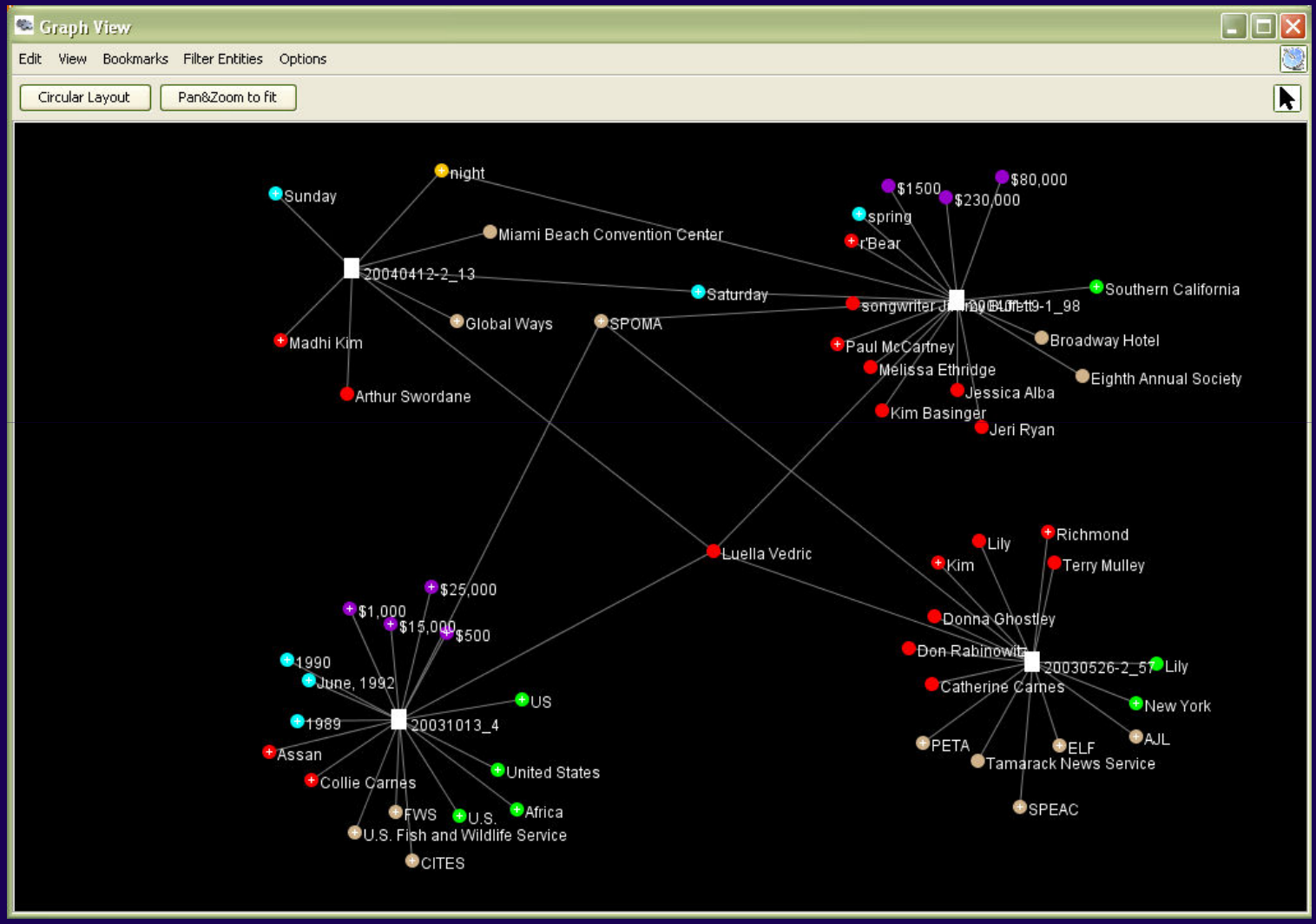
Document View



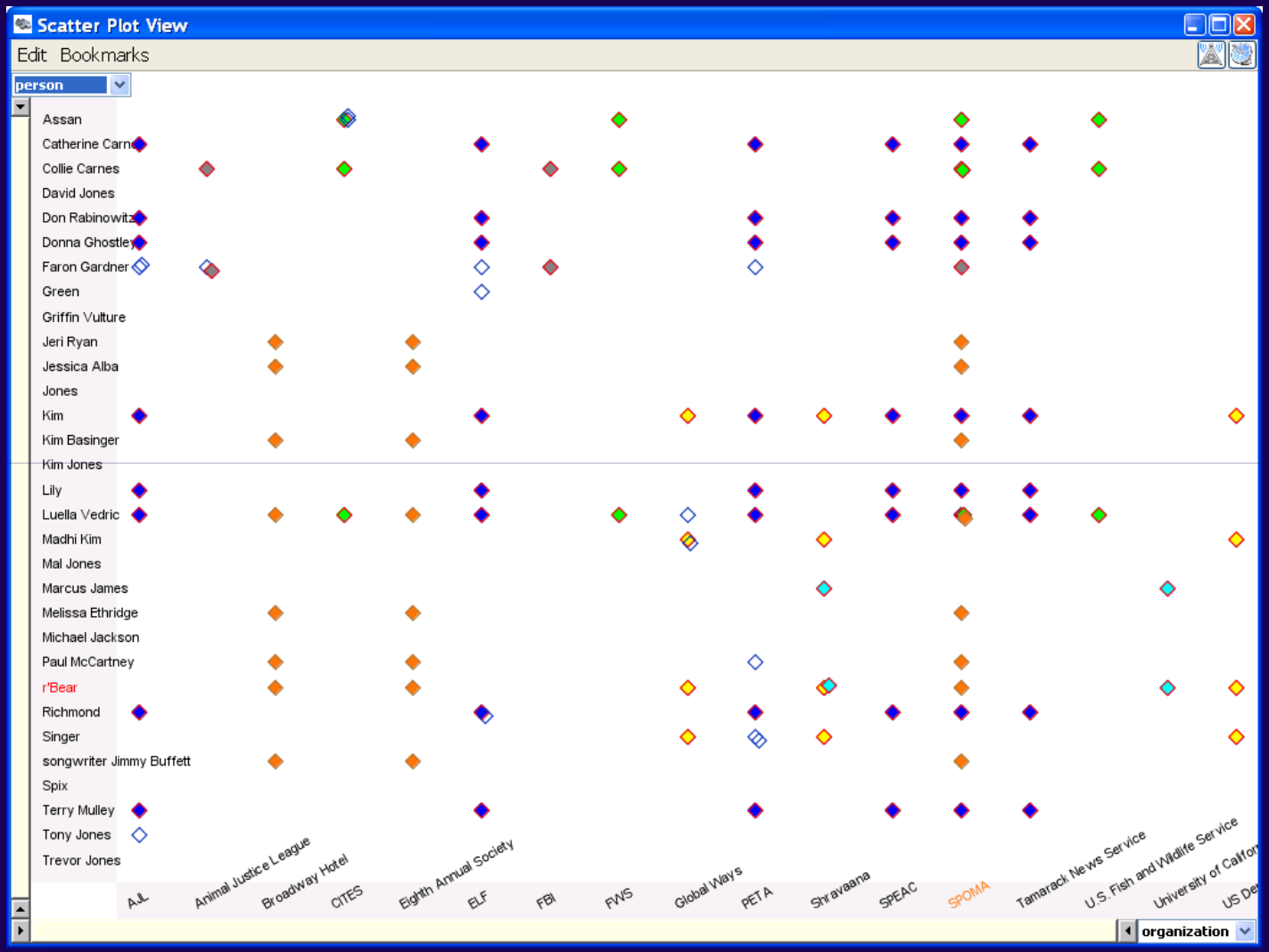
List View


List View
Edit View Bookmarks Lists Options
place
Add all
Clear
Show all connections
person
Add all
Clear
Show all connections
organization
Add all
Clear
Abbotsford
Alberta
America
Blaine
British Colombia
British Columbia
Buckner
Calgary
Canada
central Alberta
Chiapas
Colo
Colo.
Colorado
Columbia
Dawson City
Eagle Creek
Edmonton
Fort Collins
Fort Nelson
Great Slave Lake
Hollywood
Iraq
Long Island
LOS ANGELES
Los Angeles
Mexico
Montana
N.M
N.Y
northern British Columbia
Northwest Territories
Oregon
Portland
Puget Sound
Arnold Schwarzenegger
Arrow
Connor Cash
Connor Cash, 22
Cottrell, 23
Craig Rosebraugh
David Barbarash
Duane Christopher Bradley
Gary R. Perlstein
George Bush
Jack
Jeff Guidry
Jeff Koenings
Jim Widmier
Keith
Keith Jesperson
Law
Michael Scarpitti
News Headlines
Pumpkin
Robert Blecksmith
Ron Arnold
Rosebraugh, 31
saboteur Tre Arrow
Scott McInnis
Tre Arrow
Wallace Stegner
Wesley Smith
WILDLIFE EXPERT CAUGHT SMUGGLING REPTILES
William Jensen Cottrell
Wolf
ALF
Animal Liberation Front
California Institute of Technology
CWD
Department of Fish and Wildlife
Earth First
Earth Liberation Front
ELF
ELF Web
FBI
FDA
Fourth World War
Immigration and Refugee Board
Joint Terrorism Task Force
Portland State University
San Bernardino Central Detention Cen.
Sarvey Wildlife Center
Schoppert Logging Company
Supreme Court
SUV
U.S. Department of Agriculture
U.S. District Court
US Forrest Service
Washington Fish and Wildlife Commission
WHO KILLED JANE

Graph View

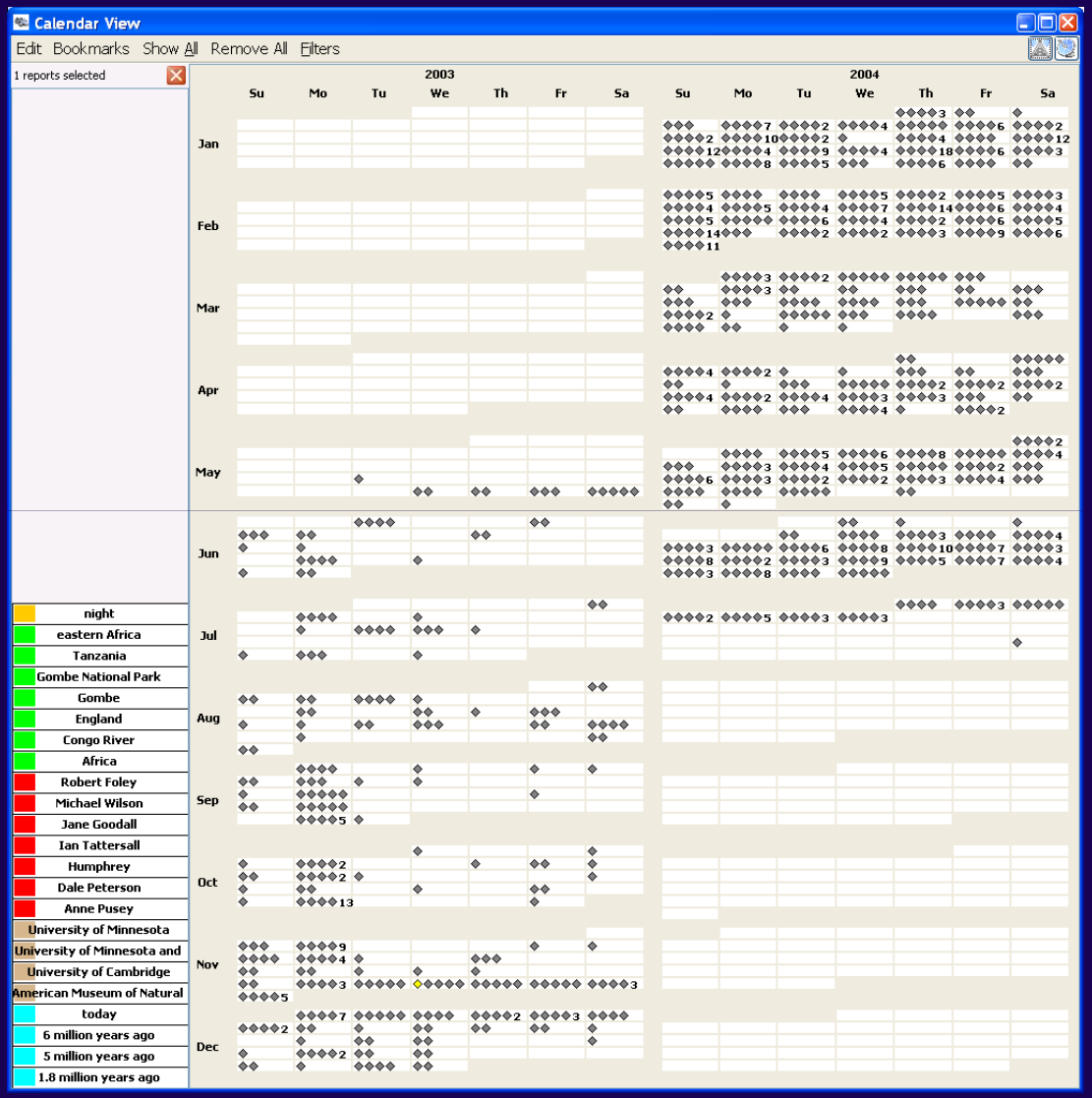

Graph View
Edit
View
Bookmarks
Filter Entities
Options
Circular Layout
Pan&Zoom to fit
night
Sunday
Miami Beach Convention Center
20040412-2_13
Saturday
Global Ways
SPOMA
Madhi Kim
Arthur Swordane
$1500
$80,000
$230,000
spring
r'Bear
Southern California
Broadway Hotel
Paul McCartney
Melissa Ethridge
Eighth Annual Society
Jessica Alba
Kim Basinger
Jeri Ryan
Luella Vedric
$25,000
$1,000
$15,000
$500
1990
June, 1992
1989
Assan
Collie Carnes
US
United States
Africa
FWS
U.S.
U.S. Fish and Wildlife Service
CITES
Richmond
Lily
Kim
Terry Mulley
Donna Ghostley
Don Rabinowitz
20030526-2_57
Lily
Catherine Carnes
New York
AJL
PETA
ELF
Tamarack News Service
SPEAC

Scatterplot View



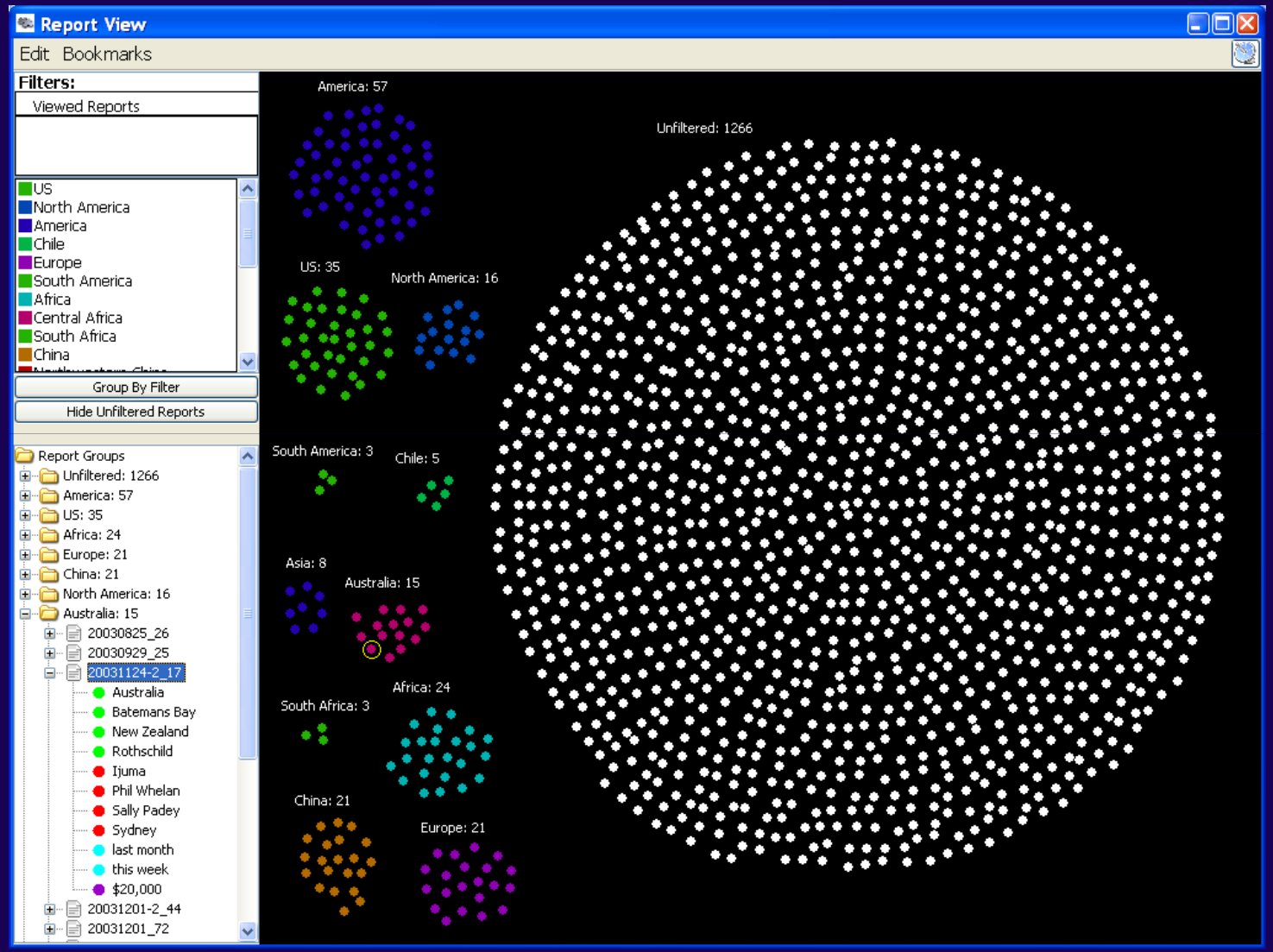
information interfaces

Calendar View

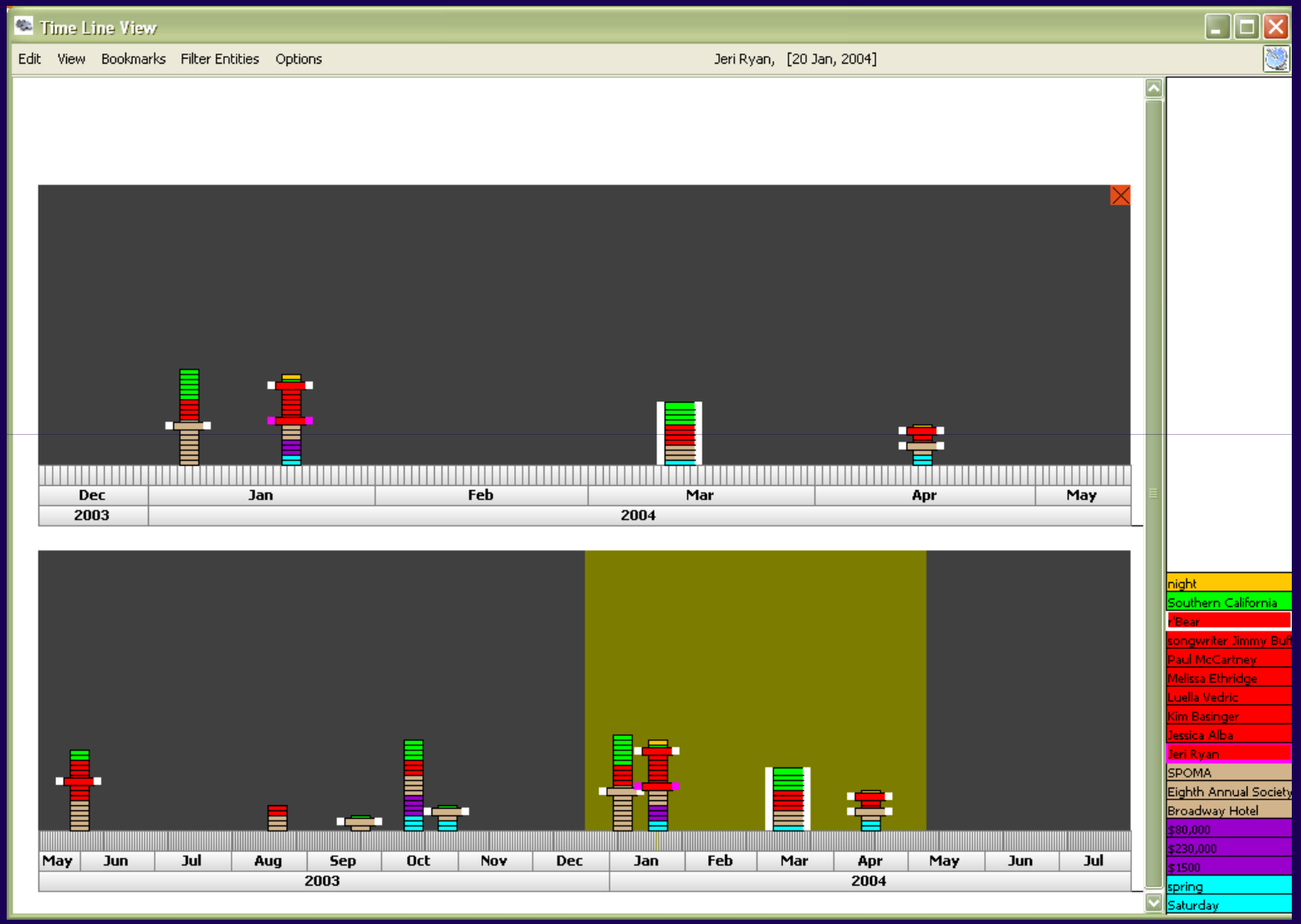


information interfaces

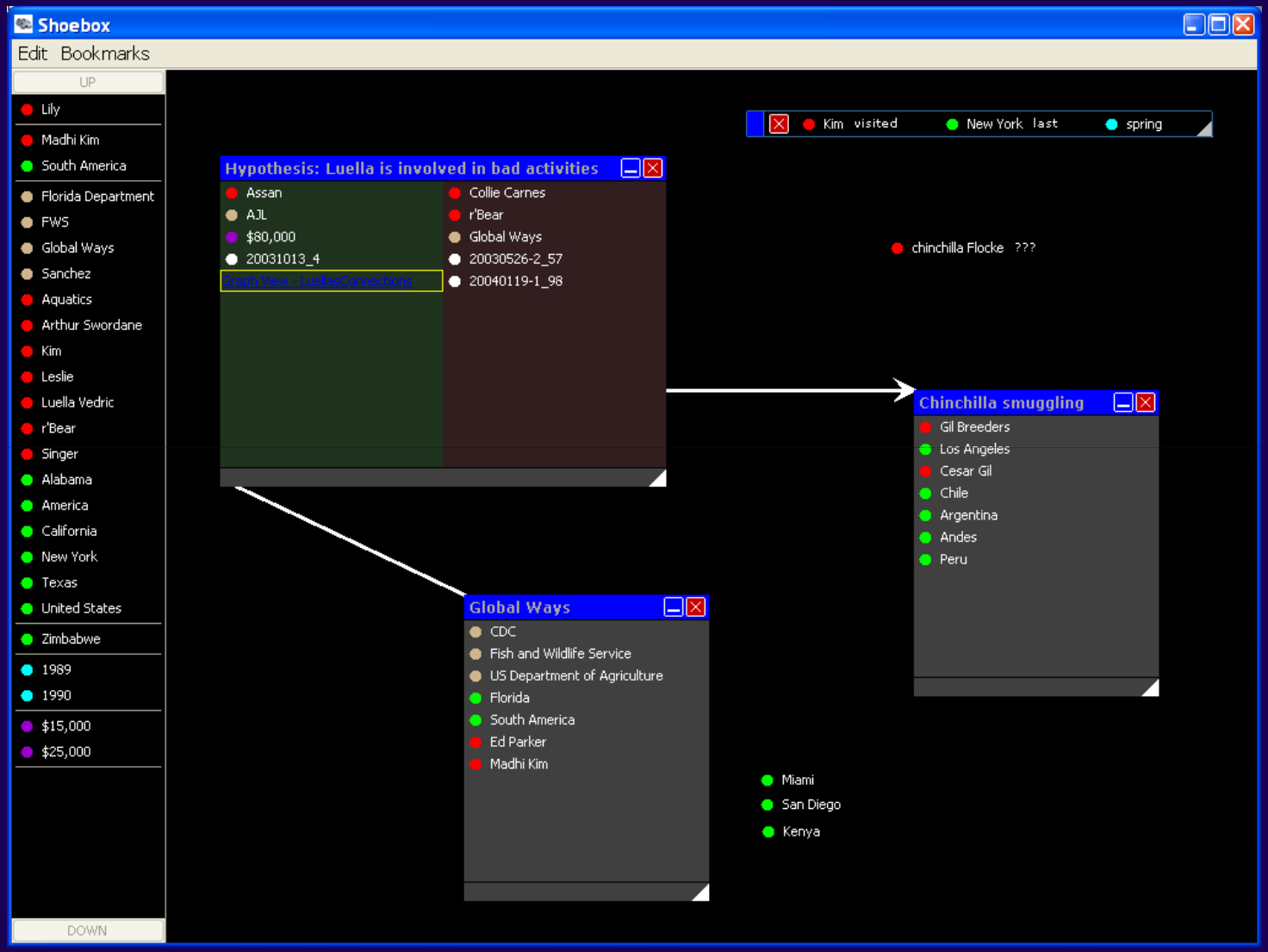
Report Cluster View



Timeline View



Shoebox


Shoebox
Edit Bookmarks
UP
Lily
Madhi Kim
South America
Florida Department
FWS
Global Ways
Sanchez
Aquatics
Arthur Swordane
Kim
Leslie
Luella Vedric
r'Bear
Singer
Alabama
America
California
New York
Texas
United States
Zimbabwe
1989
1990
$15,000
$25,000
DOWN
Kim visited
New York last
spring
Hypothesis: Luella is involved in bad activities
Assan
AJL
$80,000
20031013_4
Graph View - LuellasConnections
Collie Carnes
r'Bear
Global Ways
20030526-2_57
20040119-1_98
chinchilla Flocke ???
Chinchilla smuggling
Gil Breeders
Los Angeles
Cesar Gil
Chile
Argentina
Andes
Peru
Global Ways
CDC
Fish and Wildlife Service
US Department of Agriculture
Florida
South America
Ed Parker
Madhi Kim
Miami
San Diego
Kenya

Trial Use

- Transitioning system to real clients

information interfaces

Future Work

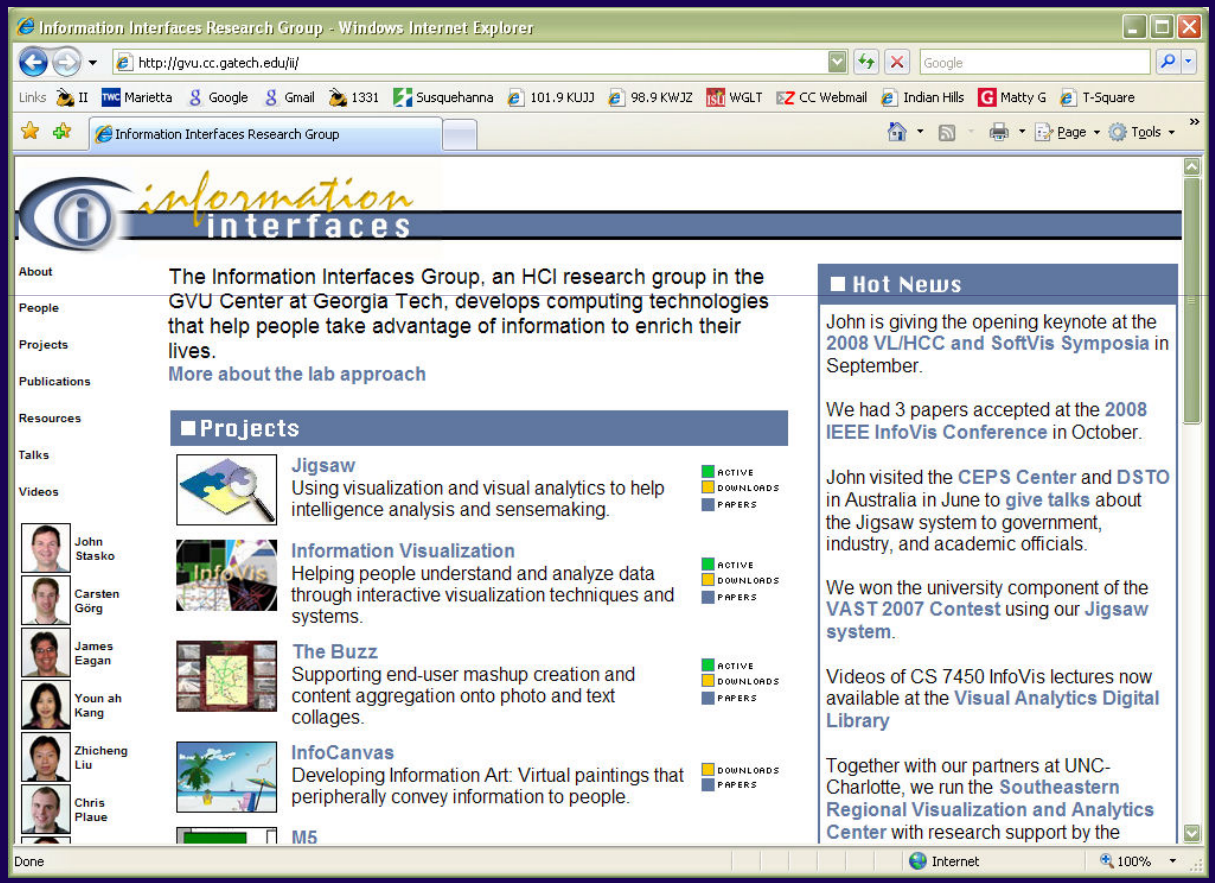
- Entity Identification
- Evaluation
- Collaborative version
- Themes/concepts
- Enhanced evidence marshalling
- Present/browse investigation history
- Scalability issues
- Wikipedia & Intellipedia
- Geospatial View
- Connectivity search
- Reliability/uncertainty
- Other types of data
- Web search & situational awareness
- Display wall?
- Deployment

Take Away Point

- Design your visualization systems and tools to facilitate analysis and exploration
  - Not to just illustrate and reconfirm existing knowledge
- Including flexible, useful interaction is one of the best ways to do this

To Learn More

- http://www.gvu.gatech.edu/ii



information interfaces

Acknowledgment

- Some slides in this presentation borrowed from overviews of visual analytics by Jim Thomas, NVAC Director

information interfaces

Acknowledgments

- Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC

- Supported by NSF IIS-0414667

information interfaces

End

- Thanks for your attention!
- Questions?

information interfaces