

Visual Analytics for Investigative Analysis and Exploration of Documents and Data

John Stasko

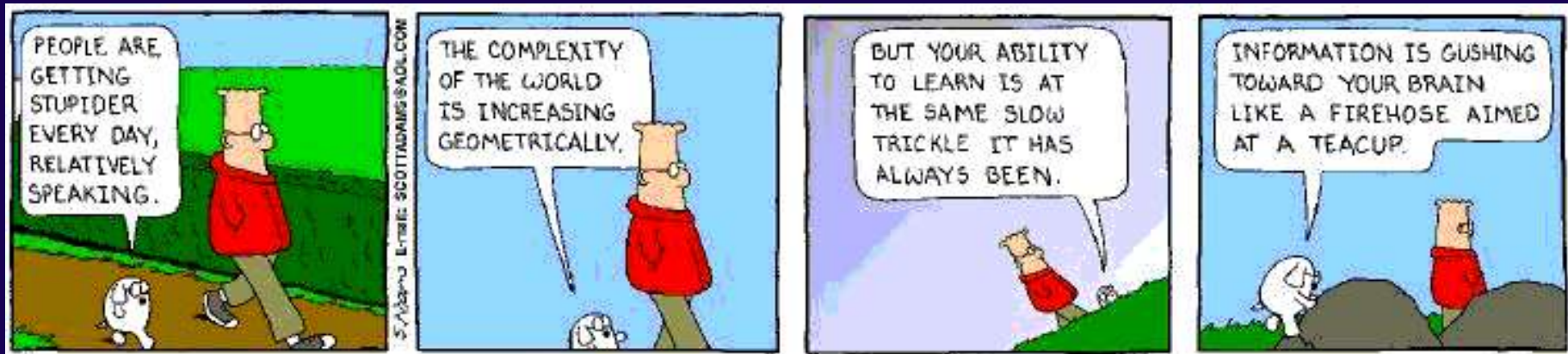
Information Interfaces Research Group
School of Interactive Computing
Georgia Institute of Technology

ISVC '10 Keynote



Data Overload

- How do we make use of the data rather than being overwhelmed by it?



Human Vision

- Highest bandwidth sense
- Fast, parallel
- Pattern recognition
- Pre-attentive
- Extends memory and cognitive capacity
- People think visually

Impressive. Lets use it!



Visualization

- “The use of computer-supported, interactive visual representations of data to amplify cognition.”
 - Card, Mackinlay, Shneiderman '98



Purpose

- Cognition, not graphics
- “The purpose of visualization is insight, not pictures”



How?

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

Larkin & Simon '87

Card, Mackinlay, Shneiderman '98



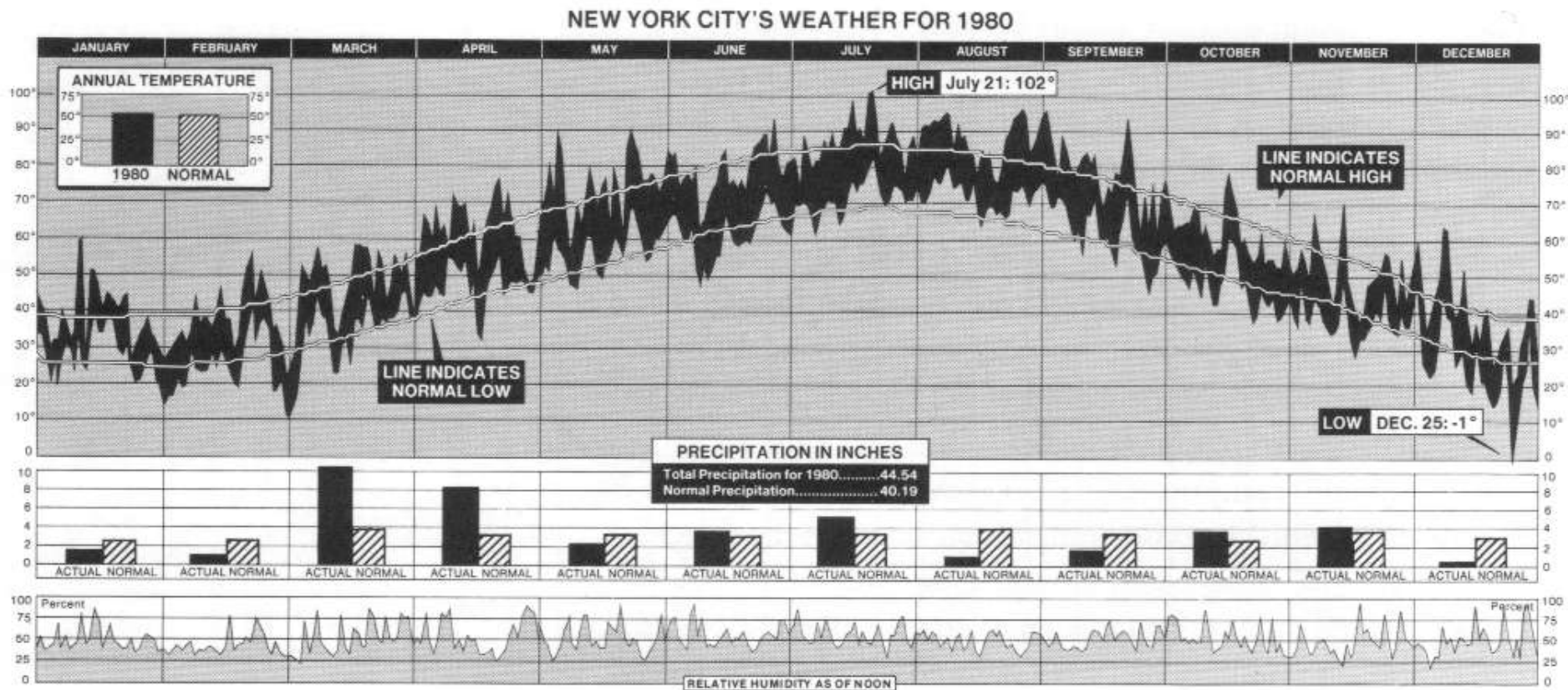
Some Examples

- InfoGraphics



NYC Weather

2220 numbers



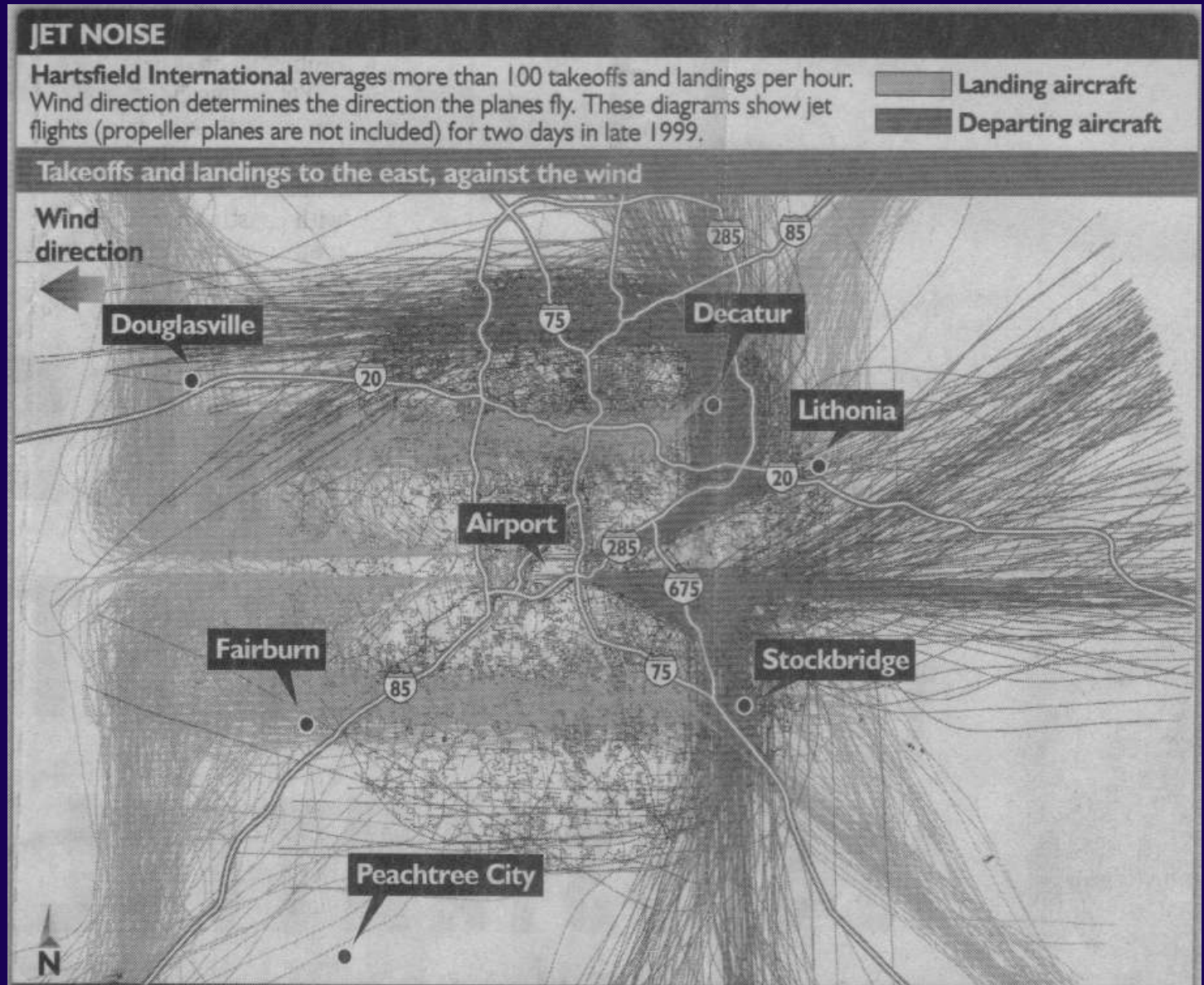
New York Times, January 11, 1981, p. 32.

Tufte, Vol. 1



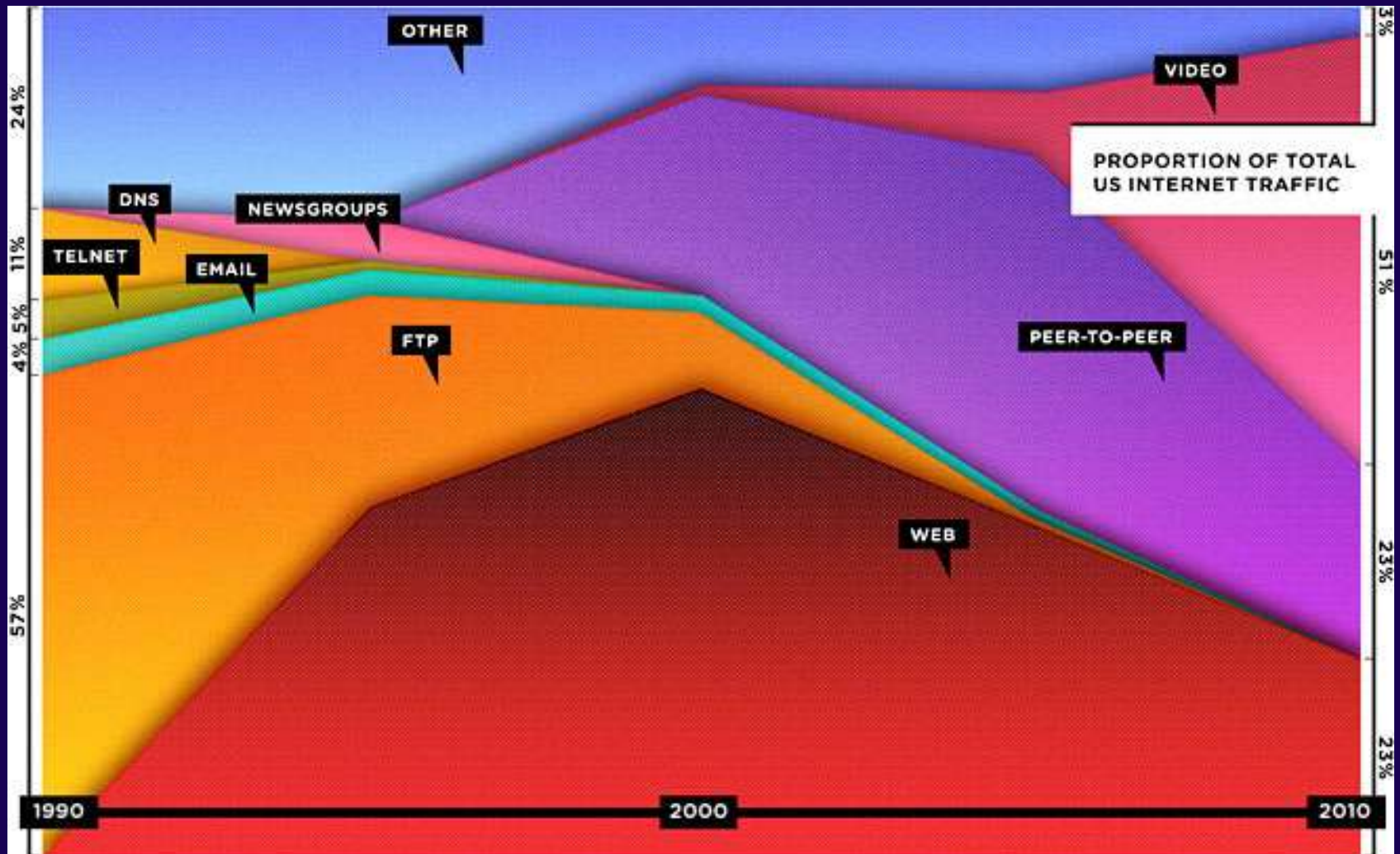
Atlanta Flight Traffic

Atlanta Journal
April 30, 2000



Internet Traffic

http://www.wired.com/magazine/2010/08/ff_webrip/all/1



Population

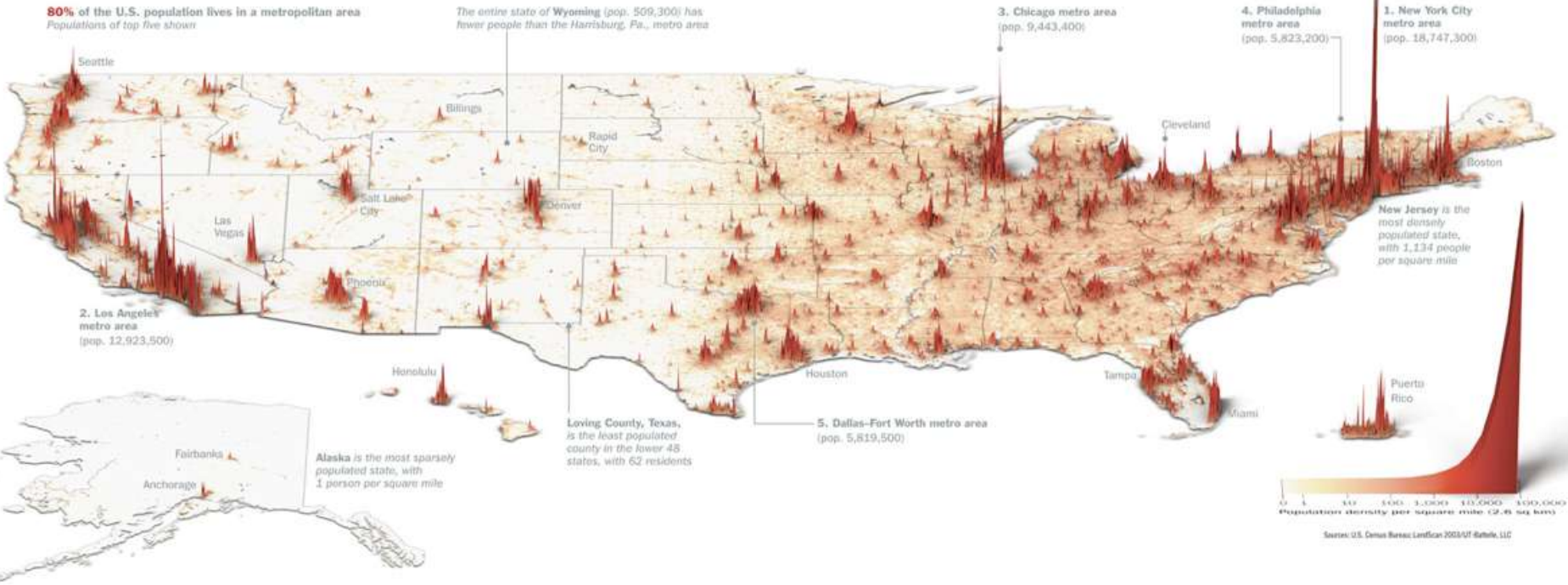
<http://infographicsnews.blogspot.com/2009/04/mantras-joe-lertolas-maps.html>

Where We Live...

Unlike many developed countries, the U.S. keeps growing. We are also moving south and west. But compared with China or India, the nation is a vast prairie

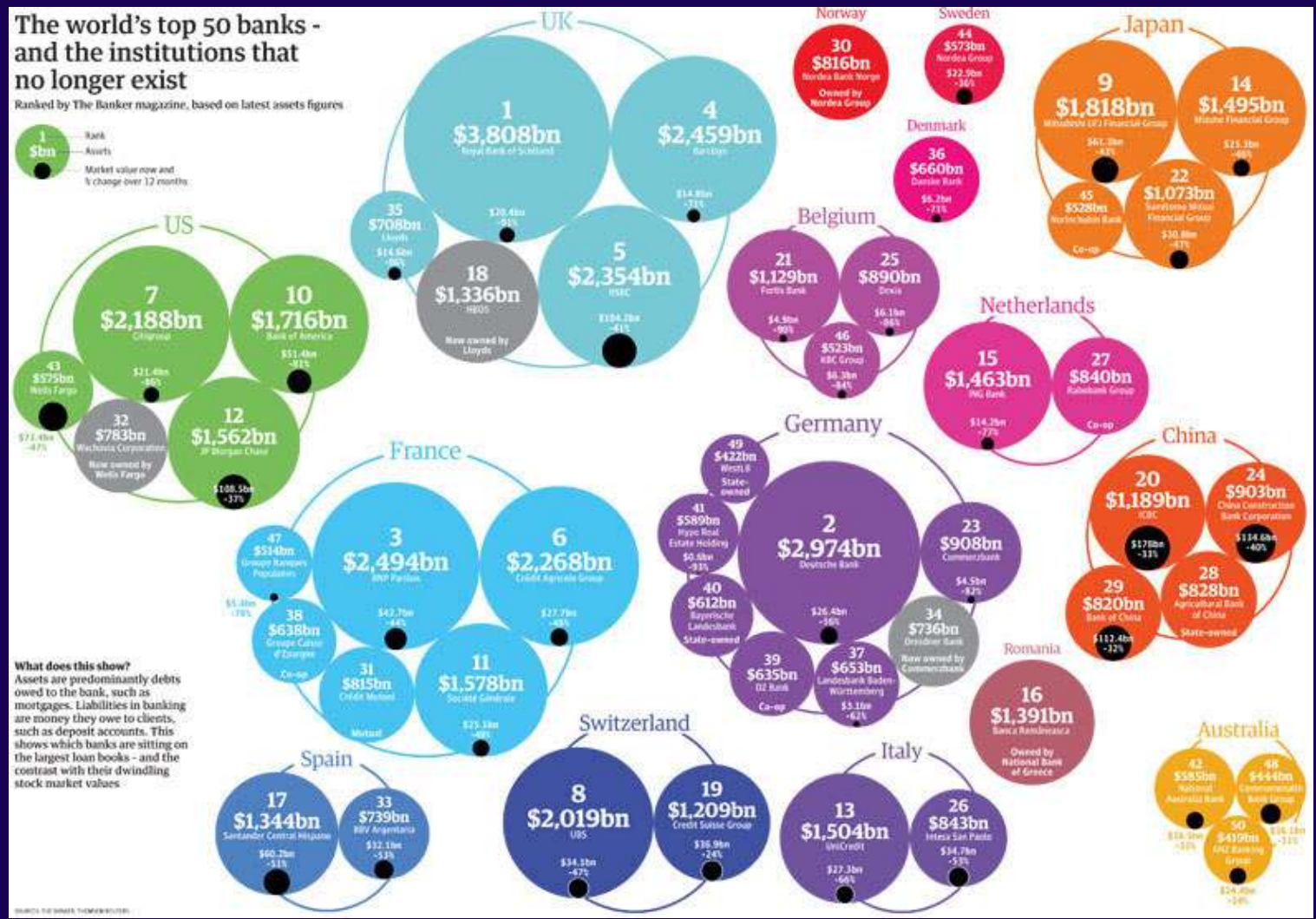
Our families are getting smaller—with one vital exception. Compared with those of Europe and Japan, the U.S. population is younger and more colorful because of the continued arrival of immigrants and their higher-than-average birthrates. Of the 100 million Americans who will join us in the next 37 years, half will be immigrants or their children. In the next few decades, 97% of the world's population growth will occur in the developing world; the U.S. is the largest developed country in the world that is still growing at a healthy clip. That matters, strategically, economical-

Ala.; Possum Trot, Ky; or Lonelyville, N.Y. But they are all probably close to someone's idea of paradise. —By Nancy Gibbs



Banking

<http://www.guardian.co.uk/news/datablog/2009/mar/25/banking-g20#>







http://infosthetics.com/archives/2008/09/funniest_pie_chart_ever.html



Purpose

- Two main uses of visualization
 - **Analysis** – Understand your data better and act upon that understanding
 - **Presentation** – Communicate and inform others more effectively



1. Analysis – 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
 - Trying to figure out what is interesting



Revisit the Definition

- “The use of computer-supported, interactive visual representations of data to amplify cognition.”
 - Card, Mackinlay, Shneiderman '98
- What does “amplify cognition” mean?



An Alternate Take

- Hutchins argues that tools don't amplify or scaffold cognition (a more traditional cognitive science view)
 - Eg, Our memory isn't amplified
- Instead, tools help transform the analytic process into another more doable one

Hutchins '94



Distributed Cognition

- Cognitive system is composed of people and the artifacts they use
 - Cognition isn't only internal
- Changes in external representation spur changes in internal representation and understanding
- It is **interaction** with the external representations that drives this process



Theoretical Basis for InfoVis

- Leverage Hutchins' theory of distributed cognition (DCog) to explain the value and utility of infovis
- Use DCog as a supporting theoretical framework for infovis

Liu, Nersessian, Stasko
IEEE TVCG '08



Characterizing Interaction

- User intent is what matters

7 categories

Select

Explore

Reconfigure

Encode

Abstract/Elaborate

Filter

Connect

Yi et al
TVCG '07



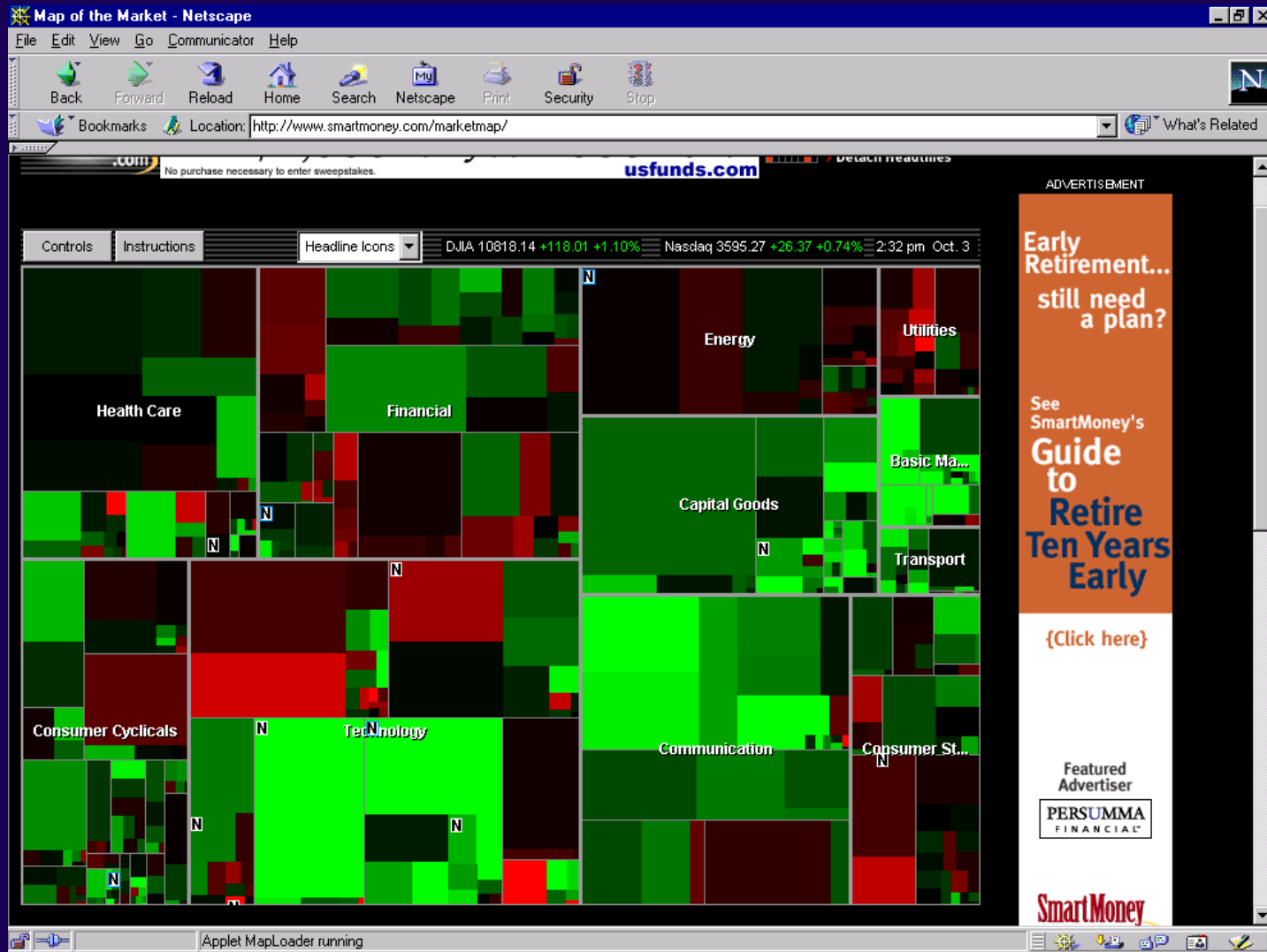
Some Examples

- Interactive visualizations



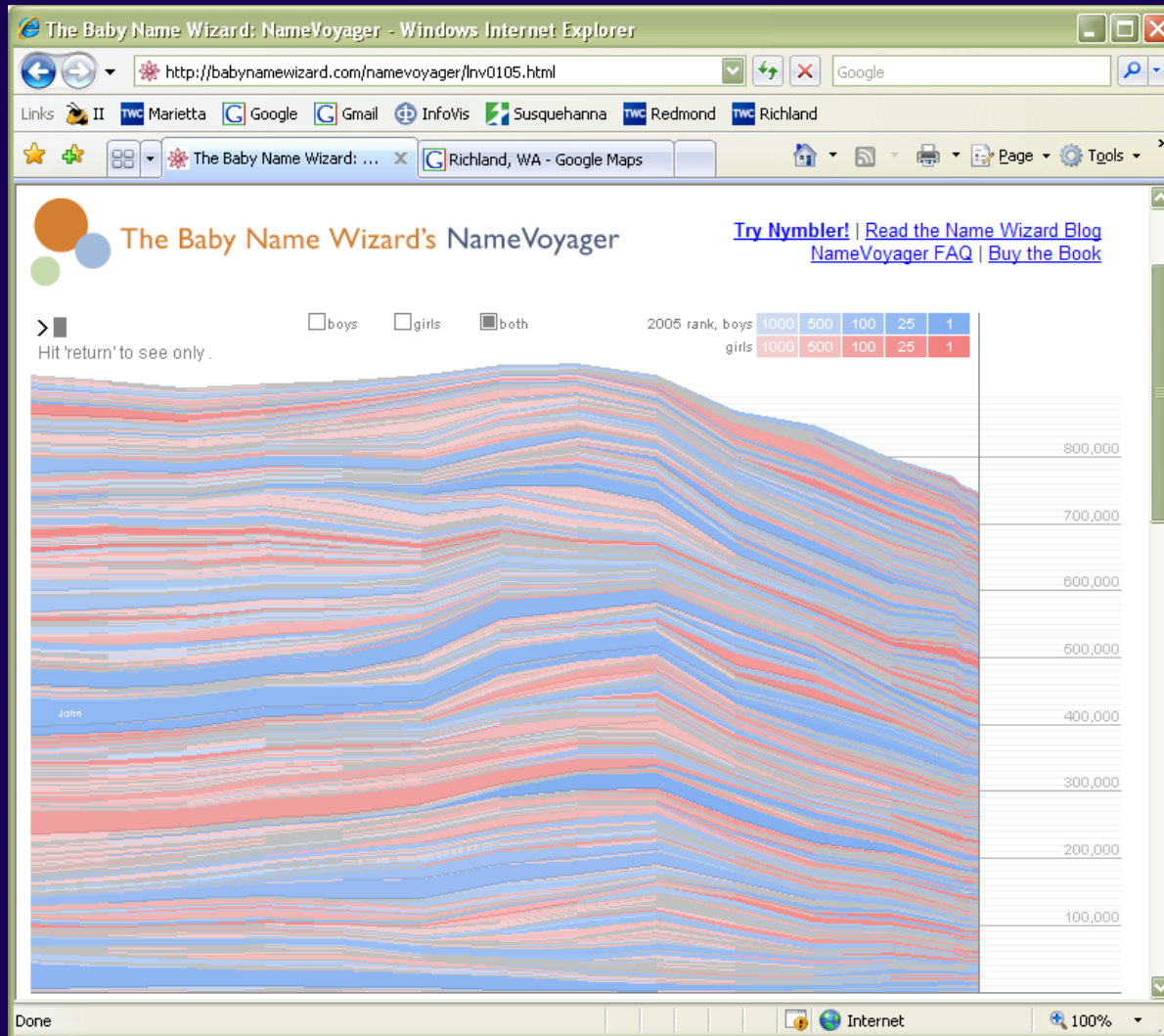
Map of the Stock Market

<http://www.smartmoney.com/marketmap>



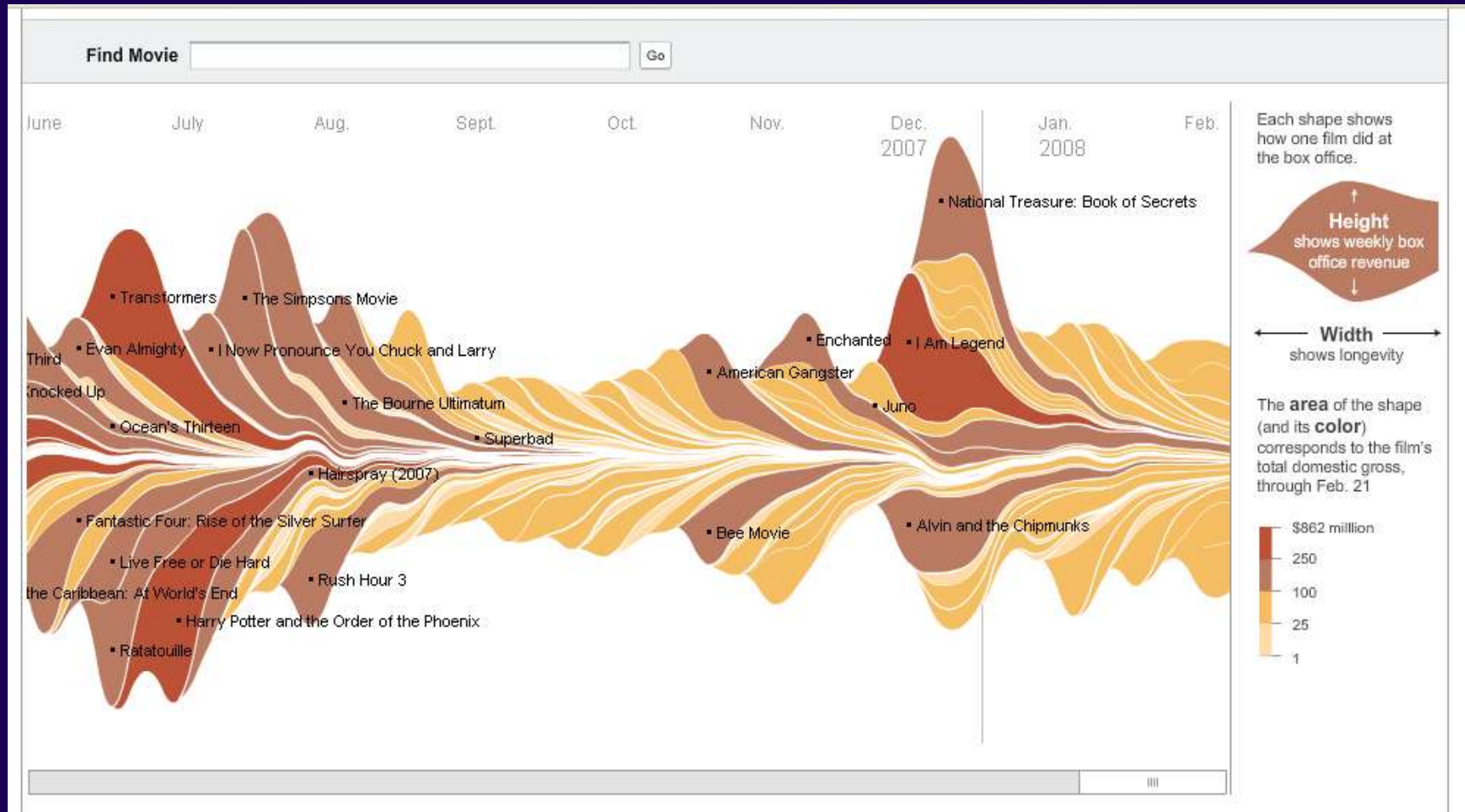
Baby Name Wizard

<http://babynamewizard.com/namevoyager/>



Movie Revenues

http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html



US Open Tennis

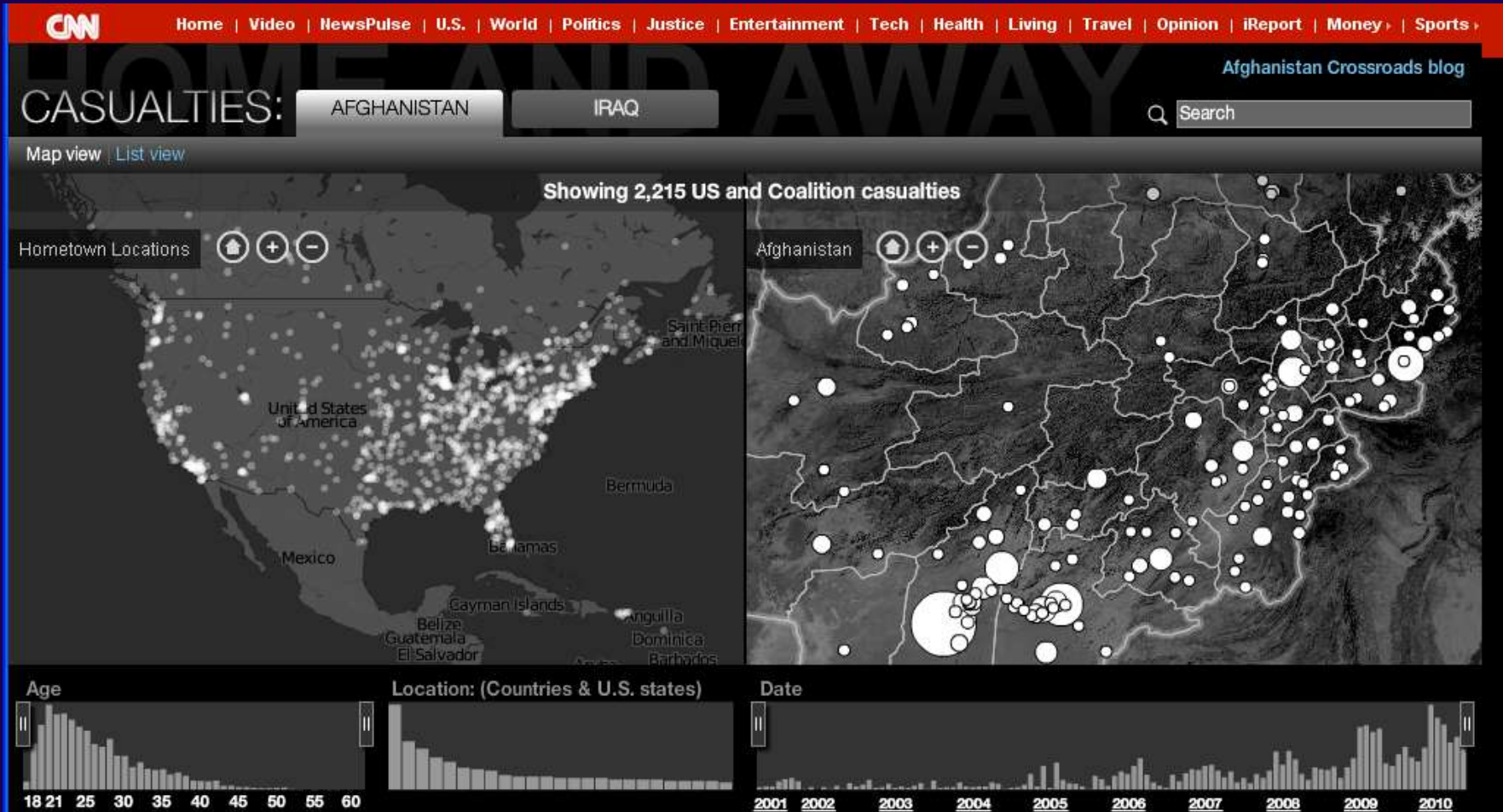
<http://www.usopen.org/ibm>



2. Presentation – Tell a story

- Use visualization to communicate ideas, influence, explain





Workshop at VisWeek

<http://thevcl.com/storytelling/>

TELLING STORIES WITH DATA



Using visualization to create narratives and engage audiences

A VisWeek 2010 Workshop

While visualization is an excellent tool for discovery and analysis, it is also a powerful medium for communication. The best information graphics do more than just present numbers: they tell a story, engage and convince their readers, invite them to make a personal connection to the data, and help them tell stories of their own.

This [VisWeek 2010](#) workshop examined the construction of narratives with visualization. We drew participants with interests in visualization, social media, journalism, and the humanities.

Time and Place

This workshop took place on **Tuesday, October 26th, 2010**, from 8:30AM to

WHAT ARE SOME EXAMPLES?

Data can speak to people in a variety of different ways; with the right audience in the right context, a [simple line graph](#) can be dramatic.

A good data narrative includes a strong perspective that tells a story; it may lead the reader through the data or encourage them to make discoveries. Its designers know their audience and invite them to form a personal connection to the data. And most importantly, it guides its audience

Done



Strongest Benefits of Visualization

- Facilitating awareness and understanding
- Helping to raise new questions and supply answers
- Generating insights
- Telling a story and making a point

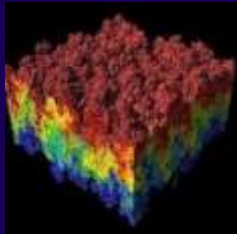


Key Challenges

- How to measure?
 - All those benefits are not easily quantifiable and measured
- Evaluation is perhaps primary open research challenge for visualization



Academic Areas & Conferences



Vis (SciVis)



InfoVis



VAST



1990

1995

2006



Scientific Visualization

- Primarily relates to and represents something physical or geometric
 - Often 3-D
 - Examples
 - Air flow over a wing
 - Stresses on a girder
 - Torrents inside a tornado
 - Organs in the human body
 - Molecular bonding

•Not the focus of this class



Information Visualization

- What is “information”?
 - Items, entities, things which do not have a direct physical correspondence
 - Notion of abstractness of the entities is important too
 - Examples: baseball statistics, stock trends, connections between criminals, car attributes...



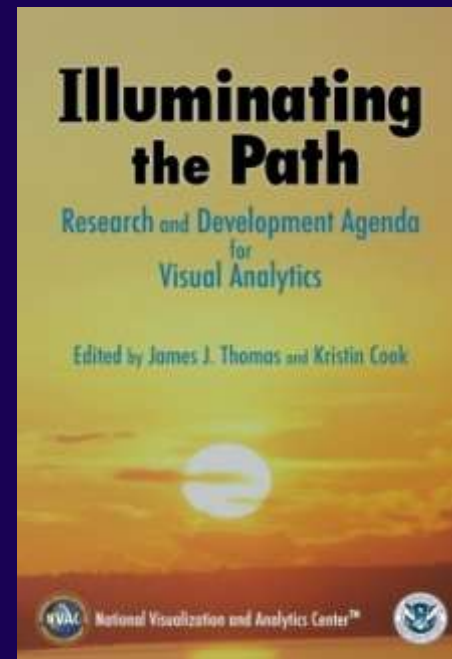
Example Domains for Info Vis

- Text
- Statistics
- Financial/business data
- Internet information
- Software
- ...



New Area Emerging

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces

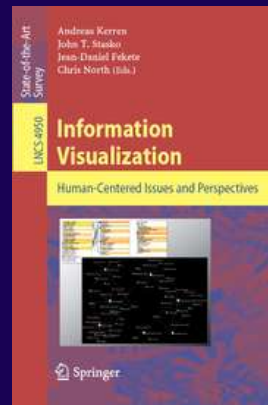


Available at <http://nvac.pnl.gov/>
in PDF form



Alternate Definition

Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets



Keim et al
chapter in *Information Visualization: Human-Centered Issues and Perspectives*, 2008



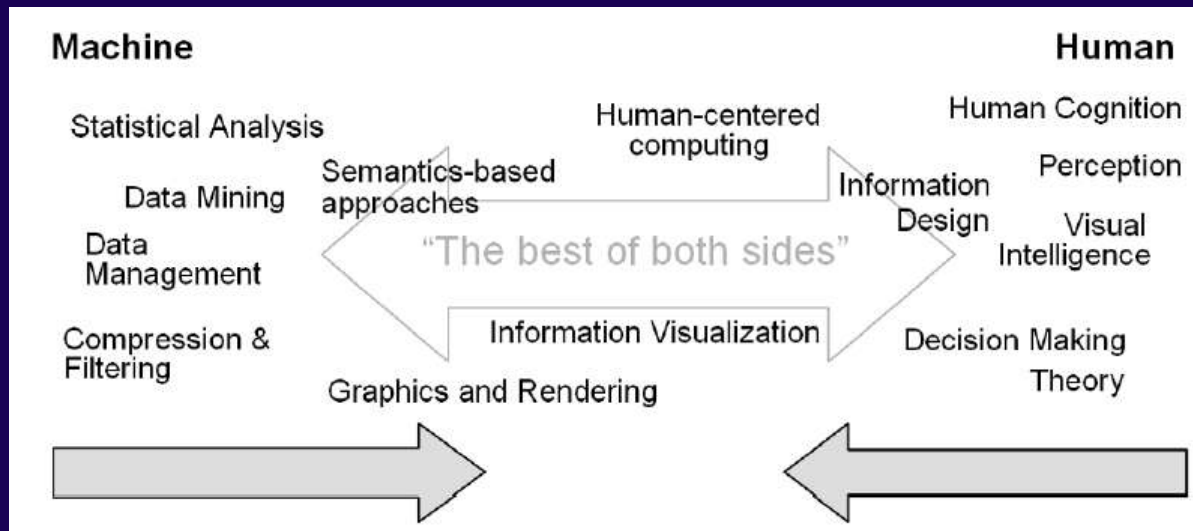
Encompassing Notion

- VA not really an “area” per se
 - More of an “umbrella” or encompassing notion
 - Combines multiple areas or disciplines
- Ultimately about using data to improve our knowledge and help make decisions



Human-Machine Synergy

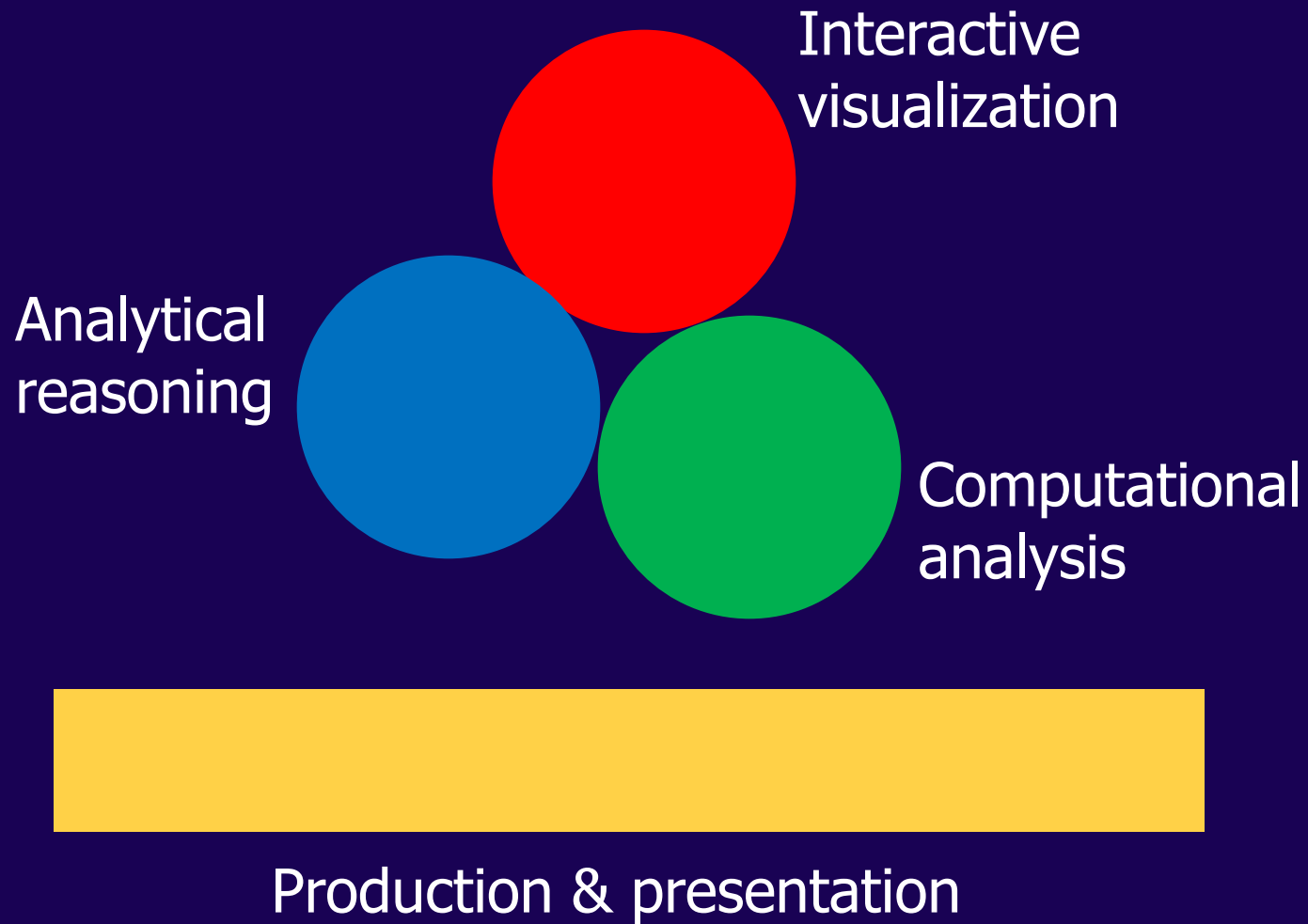
- Combine strengths of both human and electronic data processing
 - Gives a semi-automated analytical process
 - Leverage what each does best



From Keim



Main Components



Visual Analytics

- Encompassing, integrated approach to data analysis
 - Use computational algorithms where helpful
 - Use human-directed visual exploration where helpful
 - Not just “Apply A, then apply B” though
 - Integrate the two tightly



Going Beyond InfoVis

- Larger data, more heterogeneous
- Emphasis on sense-making and analytical reasoning
- Focus on complete applications



VA-related Areas

- Visualization
 - InfoVis, SciVis, GIS
- Data management
 - Databases, information retrieval, natural language
- Data Analysis
 - Knowledge discovery, data mining, statistics
- Cognitive Science
 - Analytical reasoning, decision-making, perception
- Human-computer interaction
 - User interfaces, usability



Transition

- My main research project of the past few years...



Jigsaw

Visualization for Investigative Analysis across Document Collections

- Law enforcement & intelligence community
- Fraud (finance, accounting, banking)
- Academic research
- Journalism & reporting
- Consumer research

“Putting the pieces together”



The Jigsaw Team

Current:

Carsten Görg

Zhicheng Liu

Youn-ah Kang

Jaeyeon Kihm

Jaegul Choo

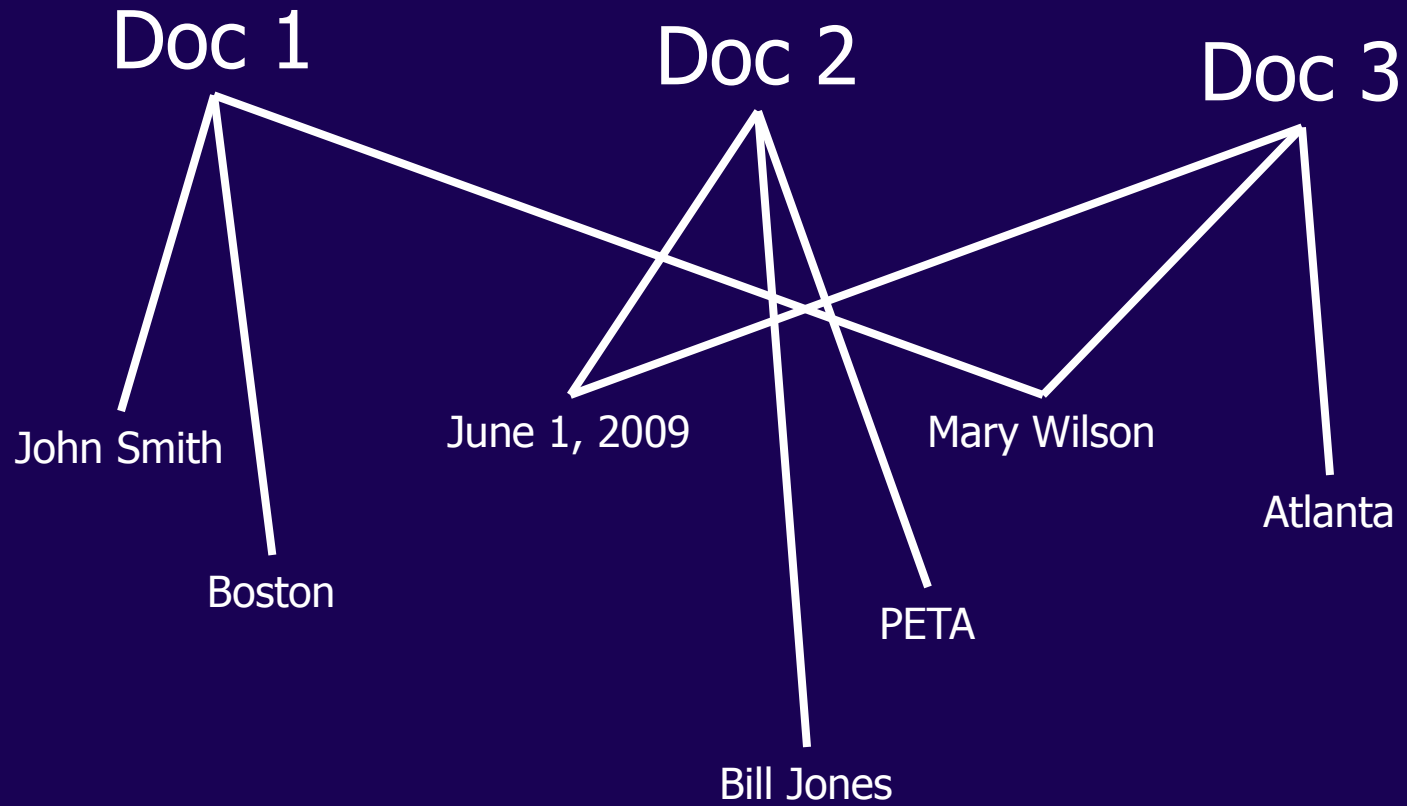
and many alumni



Our Focus

- Entities within the documents
 - Person, place, organization, phone number, date, license plate, etc.
- Thesis: A story/narrative/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 – We use tools from others



Sample 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.



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.



Sample Document 2

Title: Proving Columbus was Wrong

Abstract: In this work, we show the world is really flat. To do this, we build a bunch of ships. Then we...

PI: Amerigo Vespucci

Co-PI: Vasco de Gama, Ponce de Leon

Organization: Northwest Central Univ.

Amount: 123,456

Program Mgr: Ephraim Glinert

Division: IIS

ProgramElementCode: 2860



Entities Already Identified

Title: Proving Columbus was Wrong

Abstract: In this work, we show the world is really flat. To do this, we build a bunch of ships. Then we...

PI: Amerigo Vespucci

Co-PI: Vasco de Gama, Ponce de Leon

Organization: Northwest Central Univ.

Amount: 123,456

Program Mgr: Ephraim Glinert

Division: IIS

ProgramElementCode: 2860

Entities



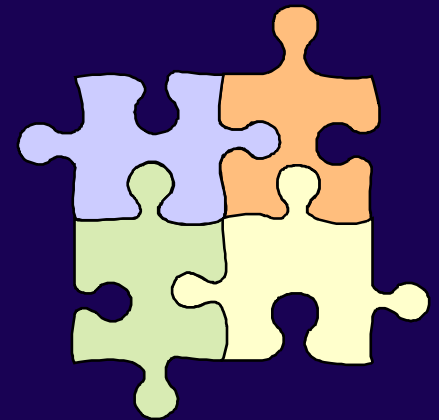
Connections

- 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

- 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



Document View

analysis [analysts](#) [analytic](#) [animation](#) [based](#) [cognition](#) [design](#) [discuss](#) [display](#) [evaluation](#)

framework [Information](#) [infovis](#) [interaction](#) [level](#) [localization](#) [paper](#) [research](#)

systems [tasks](#) [techniques](#) [video](#) [visual](#) [visualization](#) [visualizations](#)

Documents: infovis01-963277, infovis02-1249027, infovis04-1382902, infovis05-1532136, infovis07-4376144, infovis08-4658127, infovis09-4658129, infovis09-4658146, infovis09-520708, infovis09-520685, vast07-4389006, vast07-4389013, vast09-5332596, **vast09-5333878**

Summary: Evaluating visual analytics systems for investigative analysis. Deriving design principles from a case study. Despite the growing number of systems providing visual analytic support for investigative analysis, few empirical studies of the potential benefits of such systems have been conducted, particularly controlled, comparative studies. **Source:** Visual Analytics Science and Technology, 2009. VAST 2009. ISTE Symposium on. **Date:** Oct 12, 2009

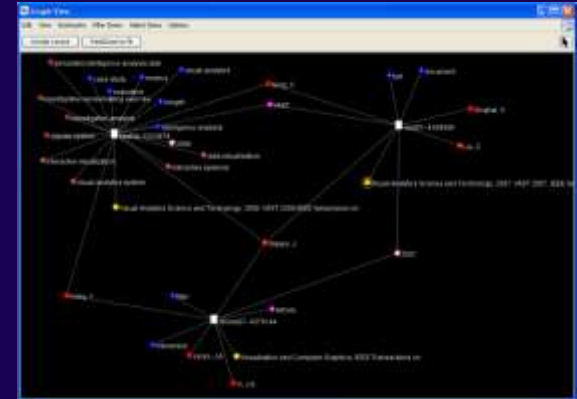
Abstract: Evaluating [visual analytics](#) systems for [investigative analysis](#). Deriving design principles from a case study. Despite the growing number of systems providing visual analytic support for investigative analysis, few empirical studies of the potential benefits of such systems have been conducted, particularly controlled, comparative studies. Determining how such systems foster insight and sensemaking is important for their continued growth and study, however. Furthermore, studies that identify how people use such systems and why they benefit (or not) can help inform the design of new systems in this area. We conducted an [evaluation of the visual analytics system](#) [ajgaw](#) employed in a small [investigative sensemaking exercise](#), and we compared its use to three other more traditional methods of analysis. Sixteen participants performed a simulated intelligence analysis task under one of the four conditions. Experimental results suggest that [ajgaw](#) assisted participants to analyze the data and identify an embedded threat. We describe different analysis strategies used by study participants and how computational support (or the lack thereof) influenced the strategies. We then illustrate several

Hit View

Concept: Interaction, Insight, Small analytics, Case study, Copression, Color, Visualization, Information, Cultural, Document, Video query, Filter, Video context, Hierarchy, Intelligence analysis, Metadata, Description, Social software visualization, Text, Theory, Time series, Timeline, Awareness, Interactions, Learning, Business, Cluster, Collaboration, Data mining, Database, Location, Filter, Geographic, Graph, Network, High-dimensional data

Author: I. Sencos, B., I. Sencos, D. W., I. Sencos, T. C., I. Sencos, H., I. Sencos, P. F., I. Sencos, B., I. Sencos, C. A., I. Sencos, C., I. Sencos, K., I. Sencos, A., I. Sencos, M. A. D., I. Sencos, T., I. Sencos, D., I. Sencos, T., I. Sencos, K. J., I. Sencos, M., I. Sencos, D., I. Sencos, A., I. Sencos, A., I. Sencos, T., I. Sencos, K. L., I. Sencos, M., I. Sencos, L. B., I. Sencos, P., I. Sencos, R., I. Sencos, V., I. Sencos, T., I. Sencos, S., I. Sencos, D., I. Sencos, S., I. Sencos, M.

Year: 1999, 1998, 1997, 1996, 1995, 2000, 2001, 2002, 2003, 2003, 2004, 2005, 2005, 2007, 2008, 2009



Document Grid View

Document to compare similarity to: infovis01-963275.0.25

Sort by: Sentiment, Color by: Sentiment

Compute Similarity based on: Text

Calendar View

Use Date Data

Granularity: 2004, 2005, 2006, 2007, 2008, 2009, 2010

Begin: 2004, End: 2010

Keywords: independent systems, distributed system, computer graphics, network (3), user interface (3), visualization (3), data (3), interaction (3), organization (3), multiple technological U.S.

Visual Analytics

Techniques: network, used, structure, 20, 34, used, user, 30, interaction, framework, user, 9, abstract, viewing, user, 25, visualization, exploring, interface, 36, analyzing, analytics, insight, support, traffic, technique, 14, design, collaboration, support, schema, video, magnitude, network, viewing, user, 25, code, using, source, 8, visual, flow, graphs, 35, set, trends, time, 13, clustering, brushing, code, 11, analytic, design, systems, 45, query, exploring, interface, 36, analyzing, analytics, insight, support, traffic, technique, 14, design, collaboration, support, schema, video, magnitude, network, viewing, user, 25, code, using, source, 8, new, video, stories, 17, visual, flow, graphs, 35, set, trends, time, 13, clustering, brushing, code, 11

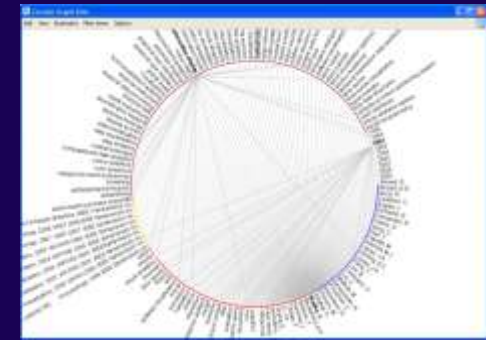
Document Cluster View

Highlight Viewed Documents

Filters: All Filters

Group by Filters, Clear Filters, Hide Unlabeled

Clusters: All Documents, info reduction, methods, 15, document, text, color, 26, network, used, structure, 20, 34, used, user, 30, interaction, framework, user, 9, multiple, exploring, design, design, collaboration, support, schema, video, magnitude, network, viewing, user, 25, code, using, source, 8, new, video, stories, 17, visual, flow, graphs, 35, set, trends, time, 13, clustering, brushing, code, 11, analytic, design, systems, 45, query, exploring, interface, 36, analyzing, analytics, insight, support, traffic, technique, 14, design, collaboration, support, schema, video, magnitude, network, viewing, user, 25, code, using, source, 8



The Need for Pixels





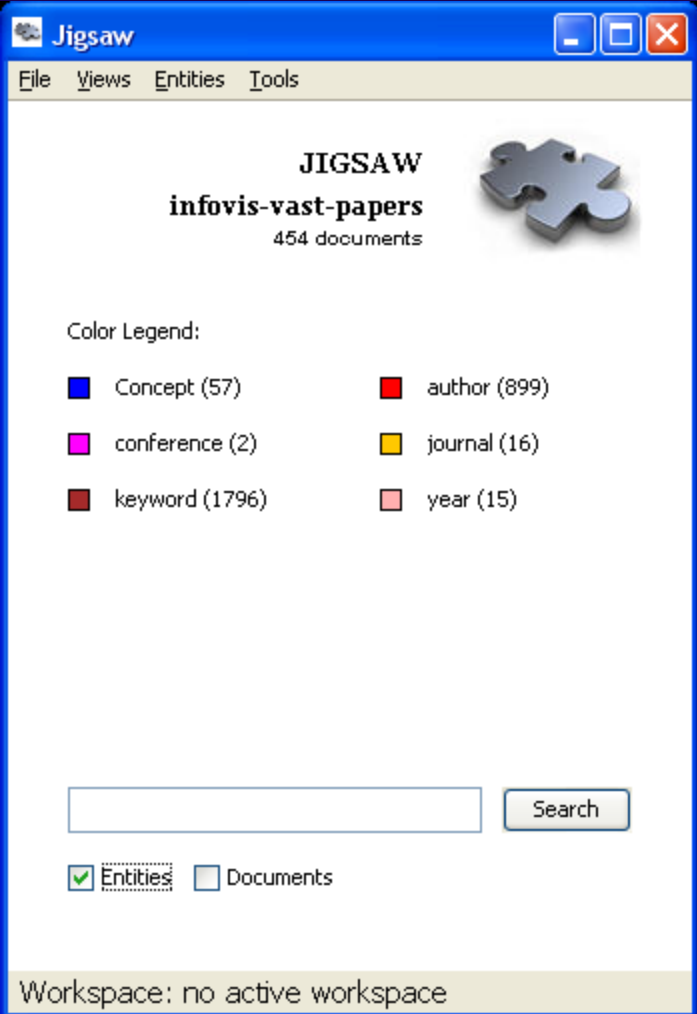
Demo

- NSF grant info
 - Text: Title + abstract
 - Entities: PI, co-PI, organization, PM, keywords, amount, program, ...
- NSF > CISE > IIS
- 2005-2010
 - 2,070 awards

Thanks to Remco Chang (UNCC/Tufts)



Console



The screenshot shows the Jigsaw application window. The title bar reads "Jigsaw" and the menu bar includes "File", "Views", "Entities", and "Tools". The main content area displays the project name "JIGSAW" and "infovis-vast-papers" with "454 documents". A 3D puzzle piece icon is shown to the right. Below this is a "Color Legend:" section with six entries, each with a colored square and text: "Concept (57)" (blue), "author (899)" (red), "conference (2)" (magenta), "journal (16)" (yellow), "keyword (1796)" (brown), and "year (15)" (pink). At the bottom of the legend area is a search bar and a "Search" button. Below the search bar are two checkboxes: "Entities" (checked) and "Documents" (unchecked). The status bar at the bottom of the window reads "Workspace: no active workspace".

Jigsaw

File Views Entities Tools

JIGSAW
infovis-vast-papers
454 documents

Color Legend:

- Concept (57)
- author (899)
- conference (2)
- journal (16)
- keyword (1796)
- year (15)

Search

Entities Documents

Workspace: no active workspace



Document View

The screenshot shows a 'Document View' window with a menu bar (Edit, View, Bookmarks) and a search bar. The main content area displays a list of terms: analysis, analysts, analytic, animation, based, cognition, design, discuss, display, evaluation, framework, information, infovis, interaction, level, localization, paper, research, systems, tasks, techniques, video, visual, visualization, visualizations. Below this is a 'Documents' list on the left and a detailed view of a selected document on the right.

Documents

- 1 infovis00--885091
- 0 infovis01--963277
- 0 infovis03--1249027
- 1 infovis04--1382902
- 1 infovis05--1532136
- 1 infovis07--4376134
- 0 infovis07--4376144
- 2 infovis08--4658127
- 0 infovis08--4658139
- 2 infovis08--4658146
- 0 infovis09--5290708
- 0 infovis95--528685
- 0 vast07--4389006
- 0 vast07--4389013
- 0 vast09--5332596
- 1 vast09--5333878

Summary: Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study Despite the growing number of systems providing visual analytic support for investigative analysis, few empirical studies of the potential benefits of such systems have been conducted, particularly controlled, comparative evaluations.

Source: Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on

Date: Oct 12, 2009

Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study

Despite the growing number of systems providing visual analytic support for investigative analysis, few empirical studies of the potential benefits of such systems have been conducted, particularly controlled, comparative evaluations. Determining how such systems foster insight and sensemaking is important for their continued growth and study, however. Furthermore, studies that identify how people use such systems and why they benefit (or not) can help inform the design of new systems in this area. We conducted an evaluation of the visual analytics system Jigsaw employed in a small investigative sensemaking exercise, and we compared its use to three other more traditional methods of analysis. Sixteen participants performed a simulated intelligence analysis task under one of the four conditions. Experimental results suggest that Jigsaw assisted participants to analyze the data and identify an embedded threat. We describe different analysis strategies used by study participants and how computational support (or the lack thereof) influenced the strategies. We then illustrate several

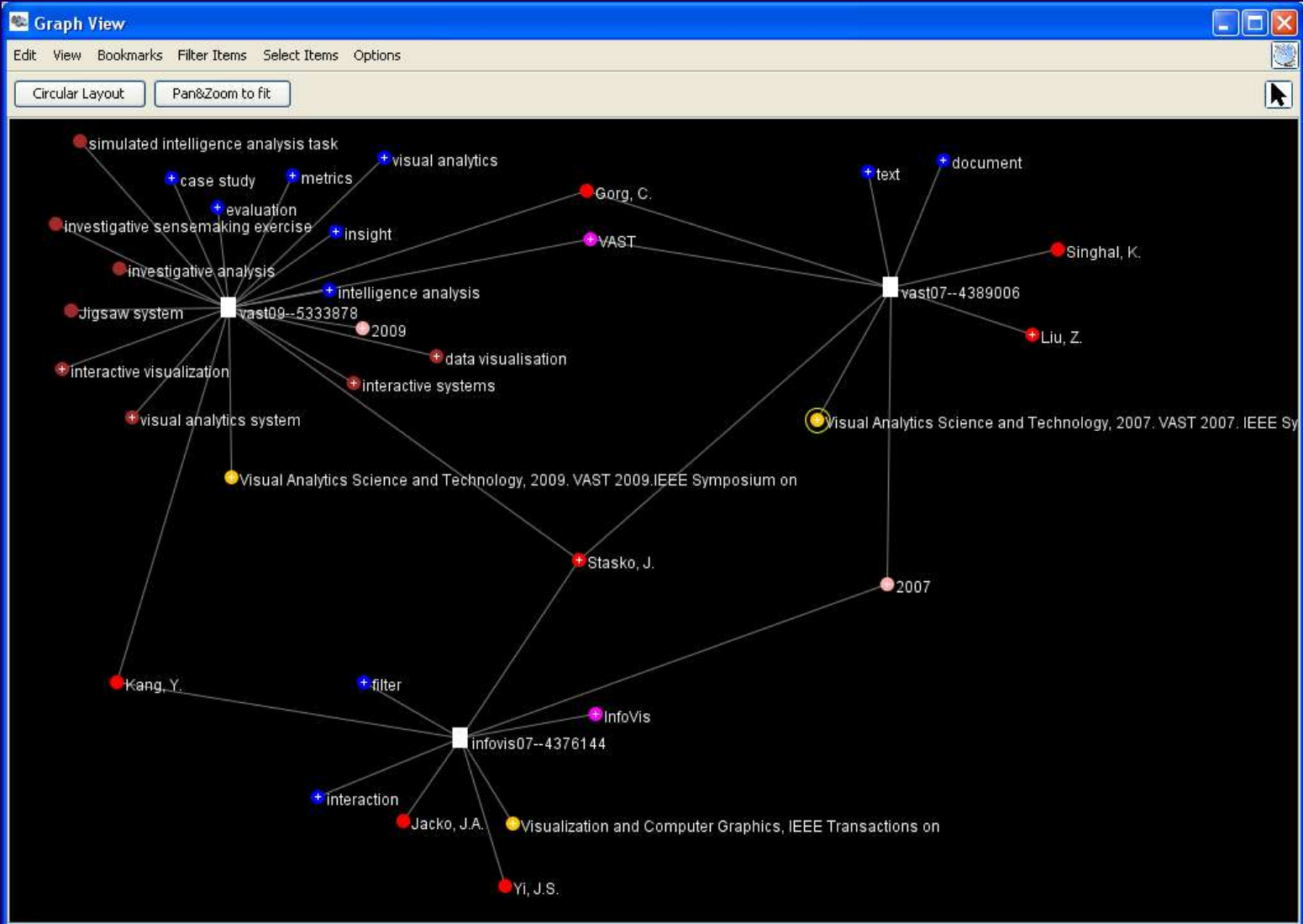


List View

The screenshot shows a 'List View' application window with four main panels. The 'Concept' panel on the left contains a list of categories such as 'interaction', 'evaluation', 'insight', 'visual analytics', 'case study', 'cognition', 'color', 'navigation', 'animation', 'categorical', 'document', 'dynamic query', 'filter', 'focus+context', 'hierarchy', 'intelligence analysis', 'metrics', 'perception', 'social', 'software visualization', 'text', 'theory', 'time series', 'treemap', 'aesthetics', 'awareness', 'bioinformatics', 'brushing', 'business', 'cluster', 'collaboration', 'data mining', 'database', 'education', 'fisheye', 'geographic', 'geospatial', 'glyph', 'graph', 'hardware', and 'high-dimensional data'. The 'author' panel in the center lists authors including Spence, B., Sprague, D.W., Sprenger, T.C., Spring, N., Stadler, P.F., **Stasko, J.**, Steed, C.A., Stein, C., Stockinger, K., Stoffel, A., Stolz, C., Storey, M.-A.D., Strasser, T., Strayer, D., Strobel, H., Strollo, P.J., Stuckey, P., Stukes, F., **Stuntebeck, E.P.**, Sturtz, D., Su, H., Sudjanto, A., Suh, B., **Sullivan, T.**, Suma, E., Summers, K.L., **Summet, J.**, Swan, J.E., Swindells, C., Syroid, N., Takeshima, Y., Tal, A., Talbot, J., Tan, D.S., Tan, R., Tanasse, T., Tandon, S., Tang, D., Tanin, E., Tatu, A., and Tavanti, M. The 'year' panel on the right lists years from 1995 to 2009, with **2000** highlighted. The 'conference' panel on the far right lists 'InfoVis' and 'VAST'. Lines connect the selected 'Stasko, J.' in the author list to the selected '2000' in the year list.



Graph View



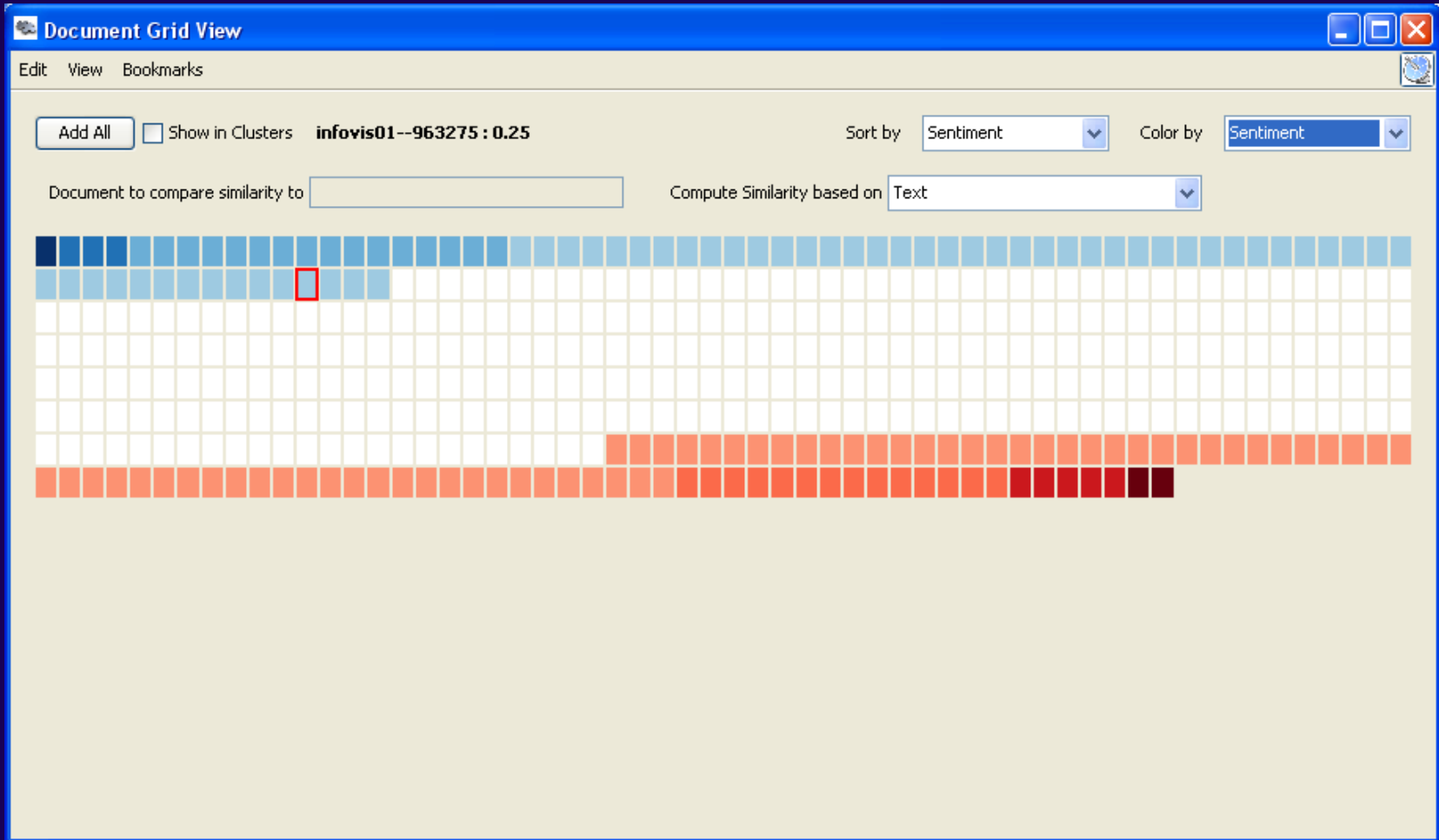
Document Cluster View

The screenshot displays the 'Document Cluster View' application window. The interface is divided into several sections:

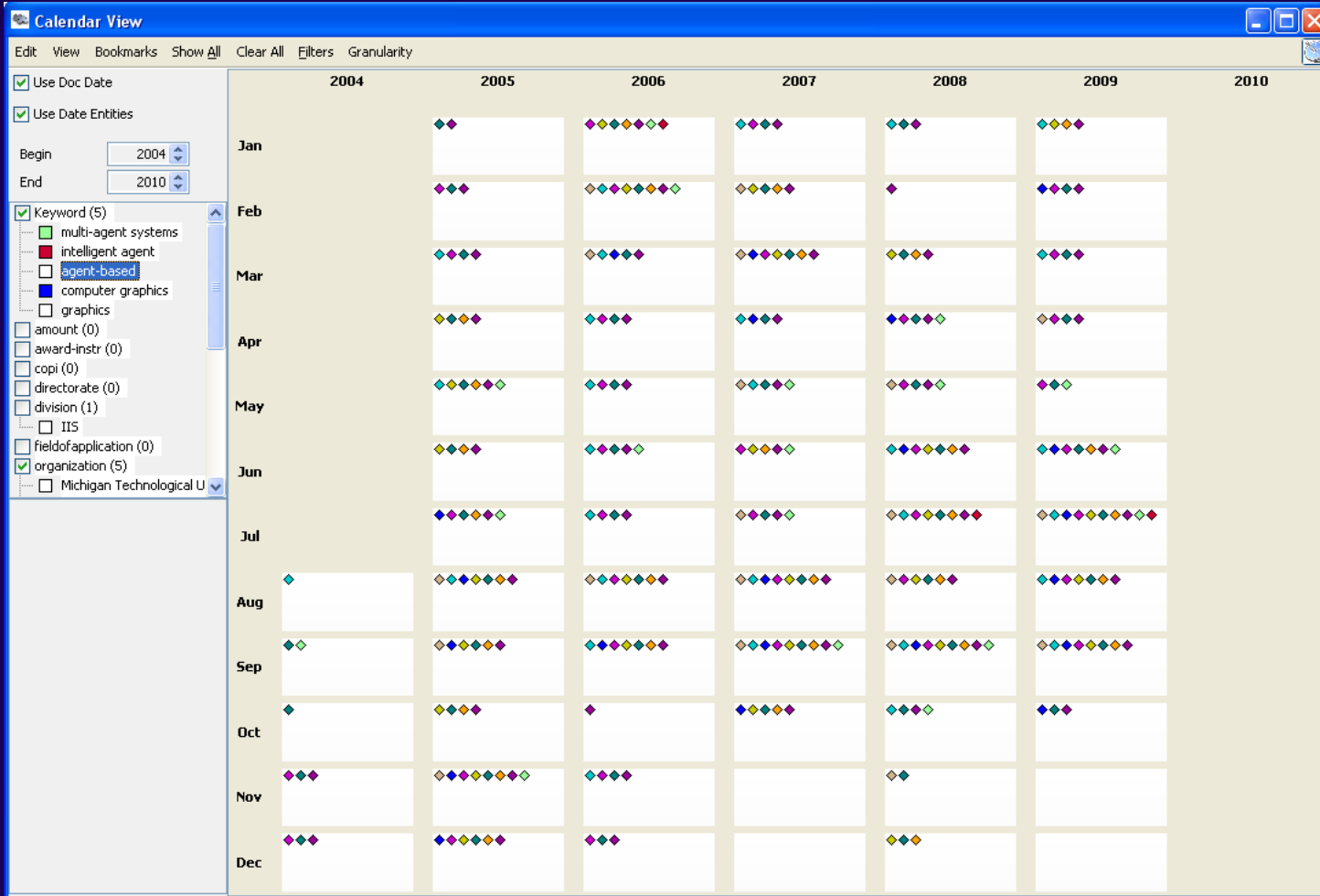
- Header:** 'Document Cluster View' title bar with standard window controls (minimize, maximize, close).
- Menu:** 'Edit View Bookmarks'.
- Left Sidebar:**
 - Highlight:** 'Viewed Documents' button.
 - Filters:** A list containing 'All Filters' with a checked checkbox. Below are buttons for 'Group by Filters', 'Clear Filters', and 'Hide Unfiltered'.
 - Clusters:** A scrollable list of document clusters with expandable icons. The selected cluster is 'Cluster (Text)'. Below the list are buttons for 'Cluster (Entity)' and 'Uncluster'.
- Main View:** A grid of 20 document clusters, each represented by a set of colored squares or circles. Each cluster is labeled with its name and count:
 - Row 1: 'mag, reduction, methods: 15' (blue squares), 'document, text, color: 26' (magenta squares), 'network, used, structure: 28' (red squares), '3d, used, user: 30' (pink squares), 'interaction, treemaps, user: 50' (grey squares).
 - Row 2: 'multivariate, exploring, dimensions: 10' (green squares), 'metaphors, state, using: 14' (purple squares), 'design, collaboration, support: 21' (cyan squares), 'schemas, vanish, magnification: 8' (dark blue squares), 'abstract, viewing, uses: 25' (yellow squares).
 - Row 3: 'code, using, sources: 8' (red squares), 'new, video, stories: 17' (light blue squares), 'visual, tree, graphs: 35' (magenta squares), 'set, trends, time: 13' (blue squares), 'visualizing, brushing, model: 11' (yellow circles).
 - Row 4: 'analytic, design, systems: 45' (light green squares), 'querying, exploring, interfaces: 36' (dark blue squares), 'visualizing, analytics, insights: 10' (orange squares), 'internet, traffic, technique: 14' (purple squares), 'display, graph, time: 25' (white squares).



Document Grid View



Calendar View



WordTree View

Word Tree View

Edit Bookmarks

<- Back Forward -> interaction Search

110 matches

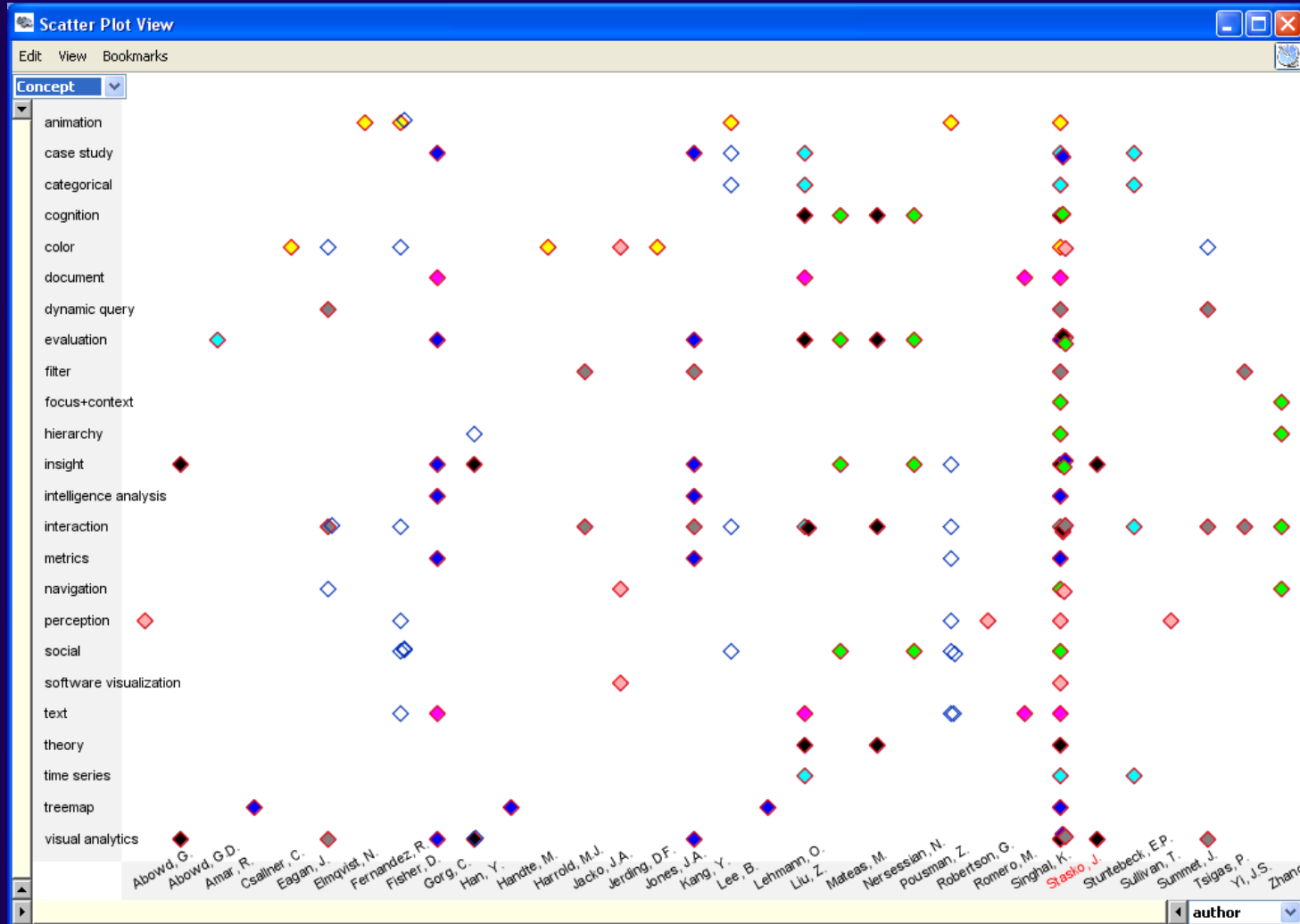
interaction

- techniques**
 - create algorithms to execute techniques efficiently.
 - illustrate their use in interactive graph visualization.
 - : a geometric-semantic zoom that provides smooth transition between overview and detail, and a fisheye distortion that displays the focus area (focus strength changing, radial rotation, level highlighting, secondary foci, animated transitions and node information) to assist in the exploration allowing semantic analysis of the underlying data sets.
 - and hopefully lay an initial foundation toward a deeper understanding and a science of interaction.
 - exist, but they typically focus on low-level operations and do not address the variety of benefits interaction provides.
 - for Selecting and Manipulating Subgraphs in Network Visualizations We present a novel and extensible set of interaction techniques for manipulating displayed nodes.
 - systems were made.
- with**
 - a 2D modified scatterplot computed from two different metrics calculated over the elements of a network.
 - information from text documents The paper describes an approach to IV that involves spatializing text content for enhanced visual browsing and analysis.
 - non-axis dimensions, and unknown values, are addressed by this method.
 - automated analytical methods to support analysts in discovering the special properties and relations of identified patterns.
- ,**
 - it is hard to provide rich visual feedback at interactive rates for datasets containing millions of entries.
 - temporal flexibility, spatial flexibility, and changing collaboration styles.
 - and through tests with security experts, several ameliorations over the standard techniques have been provided.
- and**
 - financial visual analysis it has been widely accepted that interactive visualization techniques enable users to more effectively form hypotheses and identify areas.
 - provide a way to improve interactive navigation in treemaps.
 - reasoning both operationally and strategically.
- is**
 - an important part of information visualization (Infovis), it has garnered a relatively low level of attention from the Infovis community.
 - by drilling down through many successive layers.
- between**
 - visualizations.
 - system and task type were found.
- tools**
 - for dimension hierarchy manipulation.
 - has been developed.
- are not immediately clear.
- Costs in Information Visualization Interaction cost is an important but poorly understood factor in visualization design.
- Data Spatial interactions (or flows), such as population migration and disease spread, naturally form a weighted location-to-location network (graph).
- graph model.
- in complex social networks is a critical component of visual tools for intelligence analysis, consumer behavior analysis, and human geography.
- on a desktop computer.
- technique for expanding or collapsing subtrees to any depth with a single mouse drag.
- environment by providing direct visual and algorithmic support for the coordination of data analysis actions over shared large displays.
- for graphs with visual nodes Graph and tree visualization techniques enable interactive exploration of complex relations while communicating topology.
- techniques for manipulating visualizations of networks by selecting subgraphs and then applying various commands to modify their layout or graphical properties.
- capabilities in SellTrend provides an innovative approach to this problem and to other similar types of multivariate, temporally driven transaction data analysis.
- context.

Showing 38 / 100 (38.000%) leaves in current branch, of 100 leaves in tree. Prune Percentage: 38 Prune to Window



Scatterplot View



Computational Analyses

- Document summarization
- Document similarity
- Document clustering by content
 - Text or entities
- Sentiment analysis



Recommend Related Entities

Entities recommended for: Los Angeles Times, Cesar Gil

date	money	organization	person	place	time
today last year Wednesday last week Thursday Friday 2003 last month Last year this month last fall		PETA Humane Society FBI USDA U.S. Department of Ag... ELF Creutzfeldt Jakob Fund for Animals Chiron FDA SPOMA	Dennis Kucinich Kucinich Faron Gardner Michael Markarian Collie Carnes Robert L. Ehrlich Jr. Frans de Waal Sarah Brosnan	United States U.S. Washington America Texas California Europe Africa Los Angeles San Francisco Ohio	last night late night

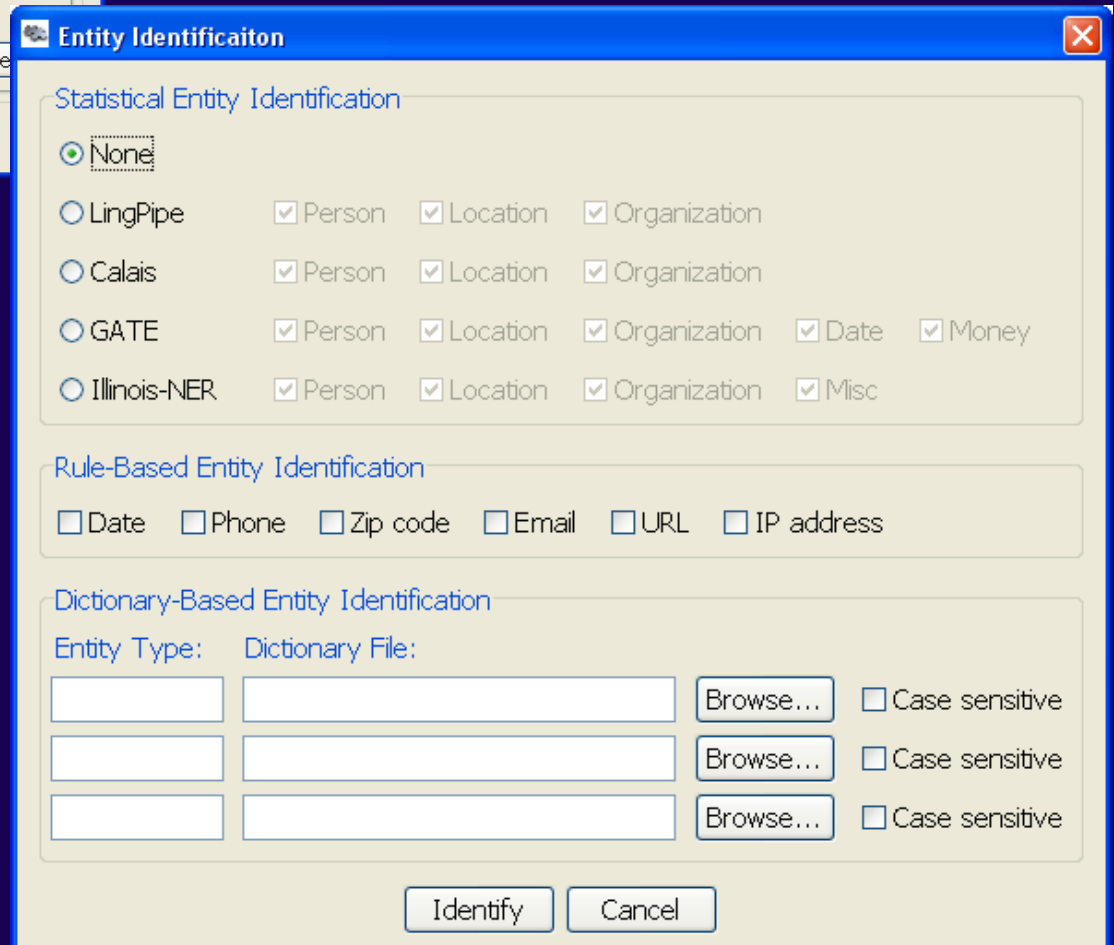
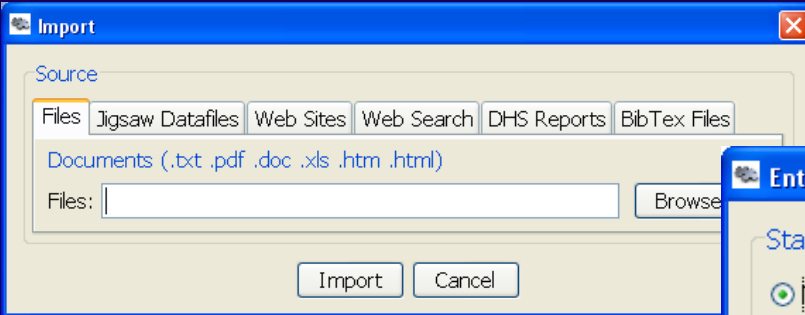
Recommended entity: **Faron Gardner**

Path: **Los Angeles Times** - (20030714-2_25) - **Animal Justice League** - (20030602-1_66 / 20030818_23) - **Faron Gardner**

Path: **Cesar Gil** - (20030609_4) - **Faron Gardner**



Document Import



Various document formats with entity identification



Input Data Formats

- Text, pdf, Word, html, Excel
- Jigsaw data file format
 - Our own xml
- DB?
 - Go to Excel
 - Go to text, transform to Jigsaw data file



```

<award>
<awardnumber>0640291</awardnumber>
<title>SGER: Distributed Spatial Partitioning Algorithms for Scalable Processing of Mobile
<nsfororganization>IIS </nsfororganization>
<programs>DATA MANAGEMENT SYSTEMS</programs>
<startdate>September 1, 2006</startdate>
<lastamendmentdate>September 12, 2007</lastamendmentdate>
<principalinvestigator>Liu, Ling</principalinvestigator>
<state>GA</state>
<organization>GA Tech Research Corporation - GA Institute of Technology </organizatio
<awardinstrument>Standard Grant </awardinstrument>
<programmanager>Le Gruenwald </programmanager>
<expirationdate>February 29, 2008</expirationdate>
<awardedamounttodate>65502</awardedamounttodate>
<co_pinames></co_pinames>
<piemailaddress>lingliu@cc.gatech.edu
<organizationstreetaddress>Office of Sponsored Programs </organizationstreetaddress>
<organizationcity>Atlanta </organizationcity>
<organizationstate>GA</organizationstate>
<organizationzip>30332</organizationzip>
<organizationphone>4048944819</organizationphone>
<nsfdirectorate>CSE </nsfdirectorate>
<programelementcodes>7485</programelementcodes>
<programreferencecodes>HPCC|9218|7484</programreferencecodes>
<fieldofapplications>0104000 Information Systems |</fieldofapplicati
<awardnumber>0640291</awardnumber>
<abstract>IIS-0640219 Ling Liu &lt;lingliu@cc.gatech.edu&gt; Georgia Institute of Instit
</award>

```

Scraped XML



```

<document>
  <docID>0808863</docID>
  <docDate>July 1, 2008</docDate>
  <docSource></docSource>
  <docText>FODAVA-Lead: Dimension Reduction and Data Reduction: Foundations for Visualization

  FODAVA-Lead: Dimension Reduction and Data Reduction: Foundations for Visualization The FODAVA (Foundations of
  Data Analysis and Visualization) Lead research team at the Georgia Institute of Technology provides unified
  expertise in the critical areas for providing leadership of the FODAVA effort, including machine learning and
  computational statistics, information visualization, massive-dataset algorithms and data structures, and
  optimization theory. The team is focused on the fundamental theory and approaches to make breakthroughs in data
  representations and transformations. The work is directed along the two main axes of scale reduction, data reductio
<directorate>CSE</directorate>
<award-instr>Continuing grant</award-instr>
<programreferencecode>HPCC</programreferencecode>
<programreferencecode>9218</programreferencecode>
<keyword>visualization</keyword>
<keyword>algorithms</keyword>
<fieldofapplication>0000912 Computer Science</fieldofapplication>
<state>GA</state>
<organization>GA Tech Research Corporation - GA Institute of Technology</organization>
<keyword>data analysis</keyword>
<keyword>information visualization</keyword>
<keyword>machine learning</keyword>
<amount>1200000</amount>
<pi>Park, Haesun</pi>
<copi>John Stasko</copi>|
<copi>Alexander Gray</copi>
<copi>Renato D. C. Monteiro</copi>
<copi>Vladimir Koltchinskii</copi>
<progmgr>Lawrence Rosenblum</progmgr>
<division>CCF</division>
<keyword>visual analytics</keyword>
<programelementcode>I114</programelementcode>
<programelementcode>H194</programelementcode>
</document>

```

Jigsaw Datafile Format



EI Correction

The screenshot shows a 'Document View' window with a menu bar (Edit, View, Bookmarks) and a toolbar. The main text area contains the following words: 'analysis approach based design exploration graph information interaction interactive large layout method network paper patterns present results space technique techniques users using visual visualization visualizations'. The words 'graph', 'information', and 'interaction' are highlighted in blue. Below the text is a 'Documents' list on the left, with '1 infovis01--963...' selected. A context menu is open over the selected document, listing options: 'Add as Concept entity', 'Add as author entity', 'Add as conference entity', 'Add as journal entity', 'Add as keyword entity', 'Add as year entity', 'Add as new entity type', and 'Add Text to Tablet'. The main text area also shows a 'Summary' and 'Source' section.

Document View
Edit View Bookmarks

Only Entities

analysis approach based design exploration **graph** information **interaction**
interactive large layout method **network** paper patterns present results space technique
techniques users using visual **visualization** visualizations

Documents ABC [Icons]

- 2 infovis00--885091
- 1 infovis01--963...**
- 1 infovis03--1249027
- 2 infovis04--1382902
- 2 infovis05--1532136
- 3 infovis07--4376134
- 2 infovis07--4376144
- 2 infovis08--4658127
- 1 infovis08--4658139
- 2 infovis08--4658146
- 1 infovis09--5290708
- 1 infovis95--528685
- 3 vast07--4389006
- 1 vast07--4389013
- 3 vast09--5332596
- 6 vast09--5333878
- 0 infovis01--963288
- 0 infovis02--1173144
- 0 infovis04--1382888
- 0 infovis05--1532148
- 0 infovis06--4015426
- 0 infovis07--4376129
- 0 infovis00--5290604

Summary: It also provides a flexible user interface for examining different perspectives that show the behavior of the source code on test sets, ranging from individual tests, to important subsets such as the set of failed tests, to the entire test suite.

Source: Information Visualization, 2001. INFOVIS 2001. IEEE Symposium on
Date: Oct 22, 2001

Technical note: visually encoding program test information to find faults in software

Large test suites are frequently used to evaluate software systems and to locate errors. Unfortunately, this process can generate a huge amount of data that is difficult to interpret manually. We have created a system, **TARANTULA**, that visually encodes test data to help find program errors. The system uses a prior...
represent how source lines act in passed and failed test...
interface for examining different perspectives that show...
on test sets, ranging from individual tests, to important s...
tests, to the entire test suite.

Add as Concept entity
Add as author entity
Add as conference entity
Add as journal entity
Add as keyword entity
Add as year entity
Add as new entity type
Add Text to Tablet

Add all



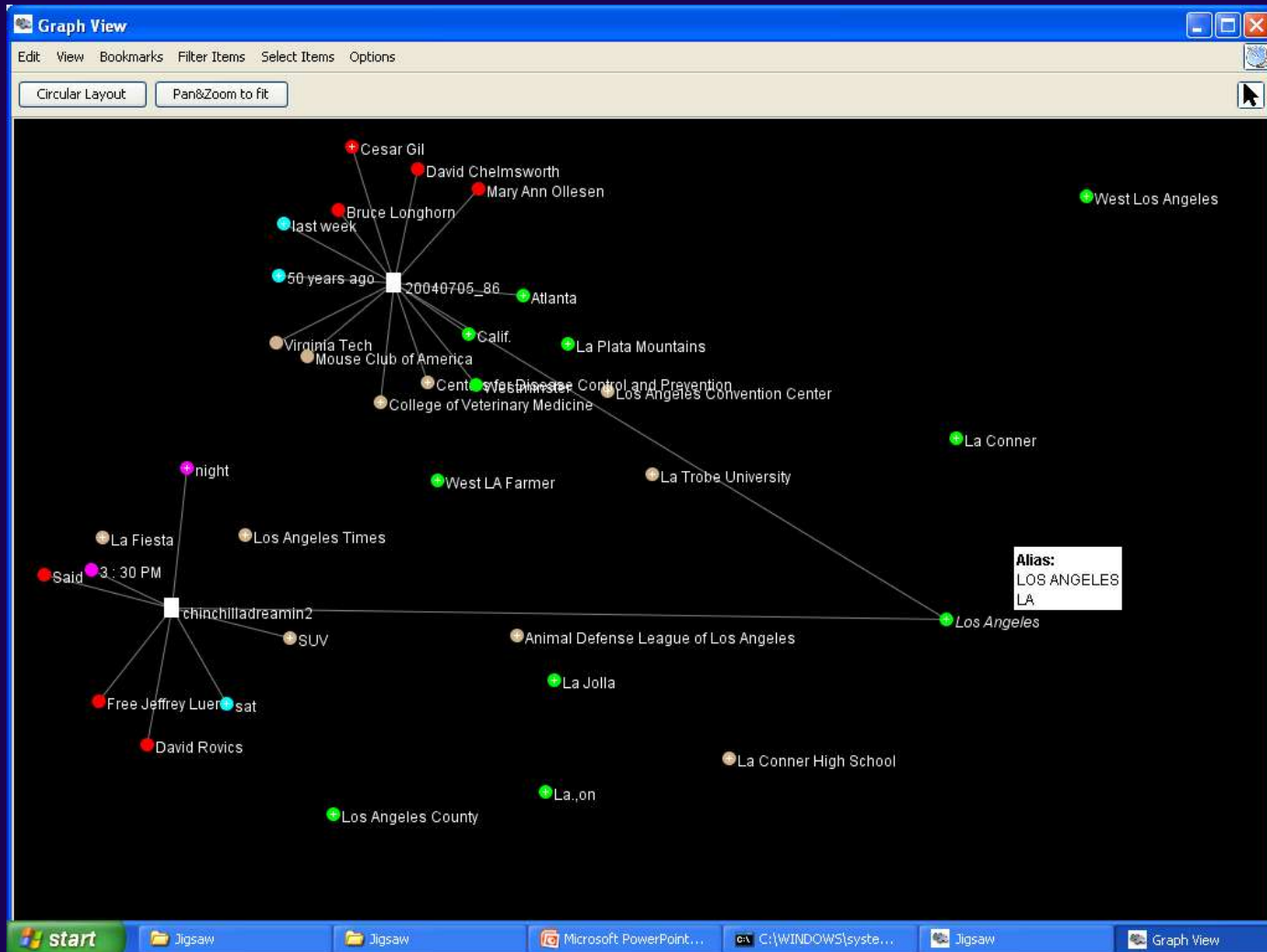
Entity Aliasing

The screenshot shows a 'List View' application with four columns: Concept, author, year, and keyword. Each column has a search filter and 'Add all' and 'Clear' buttons. The 'author' column is filtered to show authors like Stasko, J., Cependale, S., Roth, S.F., and Chuah, M.C. The 'year' column is filtered to show years from 1995 to 2009. The 'keyword' column is filtered to show terms like 'graphical user interface', 'information visualization', and 'interactive systems'. A context menu is open over the 'interactive visualisation' keyword, with 'Make Aliases' selected.

Concept	author	year	keyword
insight	Keim, D.A.	1995	graph visualization
interaction	Stasko, J.	1996	graphical user interface
visual analytics	Munzner, T.	1997	graphical user interfaces
business	Ward, M.O.	1998	grayscale shading
case study	Wattenberg, M.	1999	historical trend analysis
database	Hanrahan, P.	2000	image processing
evaluation	Rundensteiner, E.A.	2001	image visualization
graph	Shneiderman, B.	2002	immediate context graph
intelligence analysis	van Ham, F.	2003	indoor radio
metrics	van Wijk, J.J.	2004	information display
aesthetics	Cependale, S.	2005	information mural
animation	Heer, J.	2006	information navigation
awareness	Ribarsky, W.	2007	information visualization
bioinformatics	Yang, J.	2008	inspection techniques
brushing	Ebert, D.S.	2009	inter-attribute visual analysis
categorical	North, C.		interactive systems
cluster	Thomas, J.		interactive visualisation
cognition	Wong, P.C.		interactive visualisations
collaboration	Fekete, J.-D.		investigative
color	Robertson, G.		investigative
data mining	Chi, E.H.		investment
document	Fisher, D.		investment vis...
dynamic query	Ma, K.-L.		Jigsaw system
education	MacEachren, A.M.		large graph dat...
filter	Roth, S.F.		large graph exp...
fisheye	Stolte, C.		large informatic...
focus+context	Viegas, F.B.		large software
geographic	Weaver, C.		legal citation
geospatial	Zhou, M.X.		map visualization
glyph	Agrawala, M.		mathematics computing
hardware	Card, S.K.		matrix visualizations
hierarchy	Chuah, M.C.		medical information systems
high-dimensional data	Dayal, U.		multi-variate temporal event se...
machine learning	Eick, S.G.		multilevel call matrices
matrix	Footo, H.		multilevel visualizations
multiple views	Hao, M.C.		mutual fund portfolios
navigation	Hauser, H.		navigation
network	Hetzler, E.		node link diagrams
nominal	Keahey, T.A.		object-oriented programming
overview	Kosara, R.		object-oriented programs
parallel coordinates	Lee, B.		optimal animations



Alias Representation



Other Domains

- Intelligence & law enforcement
 - Stasko et al, *Information Visualization* '08
- Academic papers, PubMed
 - Görg et al, KES '10
- Consumer reviews
 - Görg et al, HCIR '10
- Investigative reporting
- Topics on the web (medical condition)

- Software
 - Ruan et al, SoftVis '10



To Learn More & Availability

<http://www.gvu.gatech.edu/ii/jigsaw>

Available for (free)
trial use

Send email to:
stasko@cc.gatech.edu



Conclusion

- Visualization is about fostering new insights
 - Analysis
 - Presentation
- Measuring the effects is challenging
- Text/documents is a fascinating new area for visual analytics research



Acknowledgments

- Work conducted as part of the Southeastern Regional Visualization and Analytics Center, supported by DHS and NVAC and the DHS Center of Excellence in Command, Control & Interoperability (VACCINE Center)



- Supported by NSF IIS-0414667, CCF-0808863 (FODAVA lead), NSF IIS-0915788



Thanks!

<http://www.gvu.gatech.edu/ii>

The screenshot shows a Mozilla Firefox browser window displaying the website for the Information Interfaces Research Group. The browser's address bar shows the URL <http://www.cc.gatech.edu/gvu/ii/>. The website features a navigation menu on the left with links for About, People, Projects, Publications, Resources, Talks, and Videos. The main content area is divided into several sections: a header with the group's logo and name, a 'Hot News' section with three news items, a 'Projects' section listing five projects with icons and status indicators (Active, Downloads, Papers), and a 'People' section with small portraits of group members. The browser's status bar at the bottom shows 'Done'.

Information Interfaces Research Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

[http://www.cc.gatech.edu/gvu/ii/](#)

Most Visited Getting Started TWC Marietta Traffic 1331 7450 Zimbra t2 gmail Facebook ihnga Indian Hills FOX FoxSoccer SRM If you do this in an em...

Information Interfaces Research Gr...

information interfaces

About
People
Projects
Publications
Resources
Talks
Videos

The Information Interfaces Group, an HCI research group in the School of Interactive Computing at Georgia Tech, develops computing technologies that help people take advantage of information to enrich their lives. More about the lab approach

Hot News

John is giving a **Keynote lecture** at the **Intl. Symposium on Visual Computing** in Las Vegas at the end of November.

We presented a paper at **InfoVis 2010** about our work on the theory of interaction in information visualization, and we won two awards in the **2010 VAST Challenge**.

We presented a paper about adding new initial computational text analysis capabilities to Jigsaw at the **HCIR '10 Workshop** in August.

John co-organized a **Dagstuhl seminar on Information Visualization** this June in Germany.

We are part of the new **VACCINE** (Visual Analytics for Command, Control, and Interoperability) DHS Center of Excellence led by **Purdue University**.

Projects

Project Name	Description	Active	Downloads	Papers
Jigsaw	Using visualization and visual analytics to help analysis and sensemaking on document collections.	ACTIVE	DOWNLOADS	PAPERS
Visual Analytics for Intelligence Analysis	Informing the design of visual analytics systems for intelligence analysis.	ACTIVE	PAPERS	
Dotlink360	Analyzing and exploring converging business ecosystems through visual analytics.	ACTIVE	PAPERS	
Information Visualization and Visual Analytics	Helping people explore, analyze, and understand data through interactive visualization techniques and systems.	ACTIVE	DOWNLOADS	PAPERS
Imprint	Empowering workers to use visualizations of printer data as a basis for conversation and reflection.	ACTIVE	PAPERS	

People

- John Skasko
- Mengdie Hu
- Youn ah Kang
- Zhicheng Liu
- Zach Pousman
- Jaeyeon Kihm
- Tanyoung

Done

II Lab - Visit our lab
CS 7450 - InfoVis Class
VACCINE Center
FODAVA Center

