

Workshop: Information Visualization

Ashish Shendure & Katy Börner

Cyberinfrastructure for Network Science Center
School of Informatics and Computing and
Indiana University Network Science Institute
Indiana University, USA

CEWiT, Indiana University

*2:00-2:50 PM EST
February 26, 2016*



Please

- download the Sci2 Tool from <http://sci2.cns.iu.edu>
- these slides <http://cns.iu.edu/docs/presentations/2016-shendure-cewit.pdf>

CNS Tools and Services are used by hundreds of thousands around the globe



Our mission is to advance datasets, tools, and services for the study of biomedical, social and behavioral science, physics, and other networks. A specific focus is research on the structure and evolution of science and technology (S&T) and the communication of results via static and interactive maps of science. Learn more at cishell.org.



CNS Cyberinfrastructure for
Network Science Center
cns.iu.edu

The Sci2 Tool

Software, Datasets, Plugins, and Documentation

- These slides
<http://cns.iu.edu/docs/presentations/2016-shendure-cewit.pdf>
- Sci2 Tool Manual
<http://sci2.wiki.cns.iu.edu>
- Sci2 Tool v1.2 beta
<http://sci2.cns.iu.edu>
- Additional Datasets
<http://sci2.wiki.cns.iu.edu/2.5+Sample+Datasets>
- Additional Plugins
<http://sci2.wiki.cns.iu.edu/3.2+Additional+Plugins>

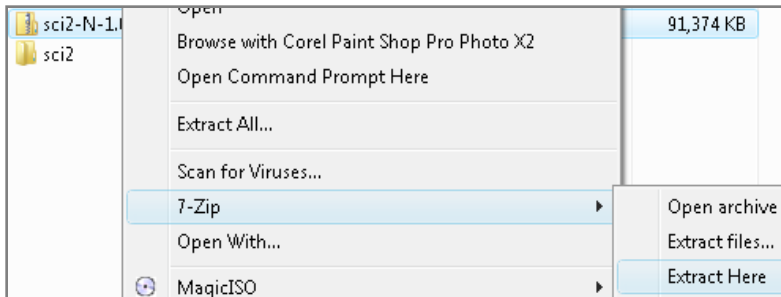
Make sure you have Java 1.6 (64-bit, if you selected 64-bit) or higher installed or download from <http://www.java.com/en/download>. To check your Java version, open a terminal and run 'java -version'.

Some visualizations are saved as Postscript files. A free Postscript to PDF viewer is at <http://ps2pdf.com> and a free PDF Viewer at <http://www.adobe.com/products/reader.html>.

Install and Run Sci2

Sci2 Tool runs on Windows, Mac, and Linux.

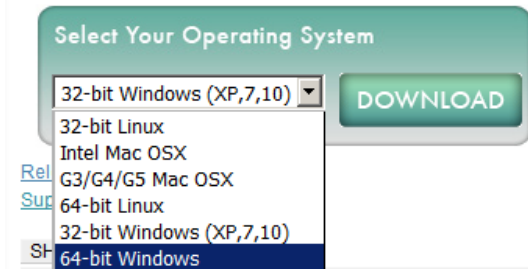
Unzip.



Run /sci2/sci2.exe

Download

Sci2 v1.2 beta
January 7th, 2015

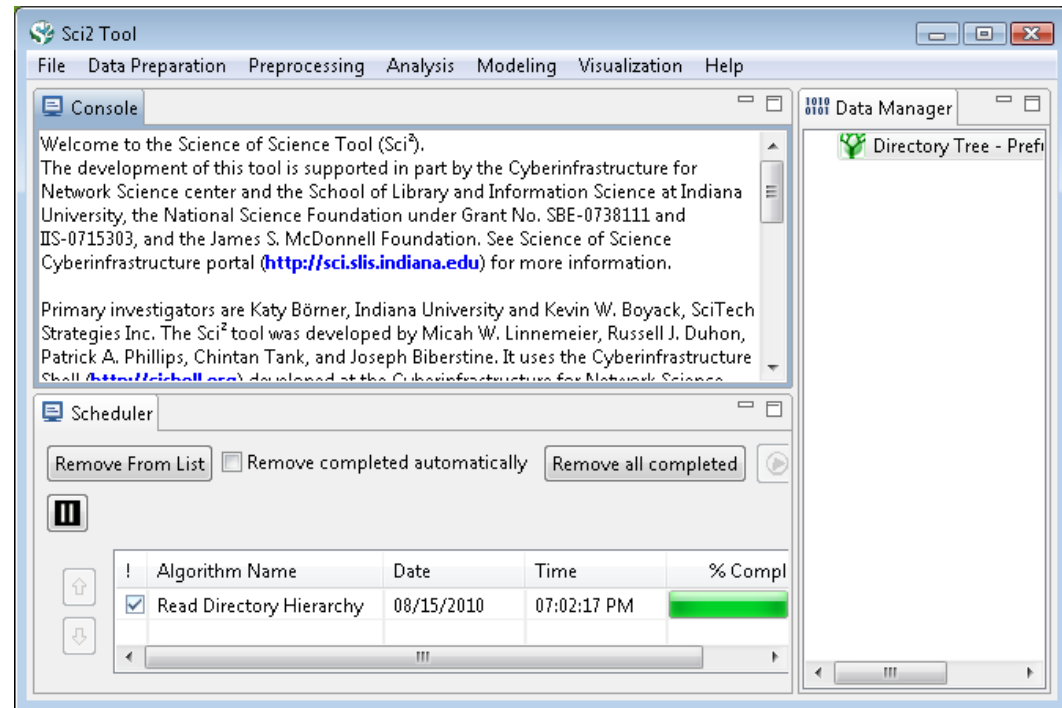


Sci2 Tool Interface Components

See also <http://sci2.wiki.cns.iu.edu/2.2+User+Interface>

Use

- **Menu** to read data, run algorithms.
- **Console** to see work log, references to seminal works.
- **Data Manager** to select, view, save loaded, simulated, or derived datasets.
- **Scheduler** to see status of algorithm execution.

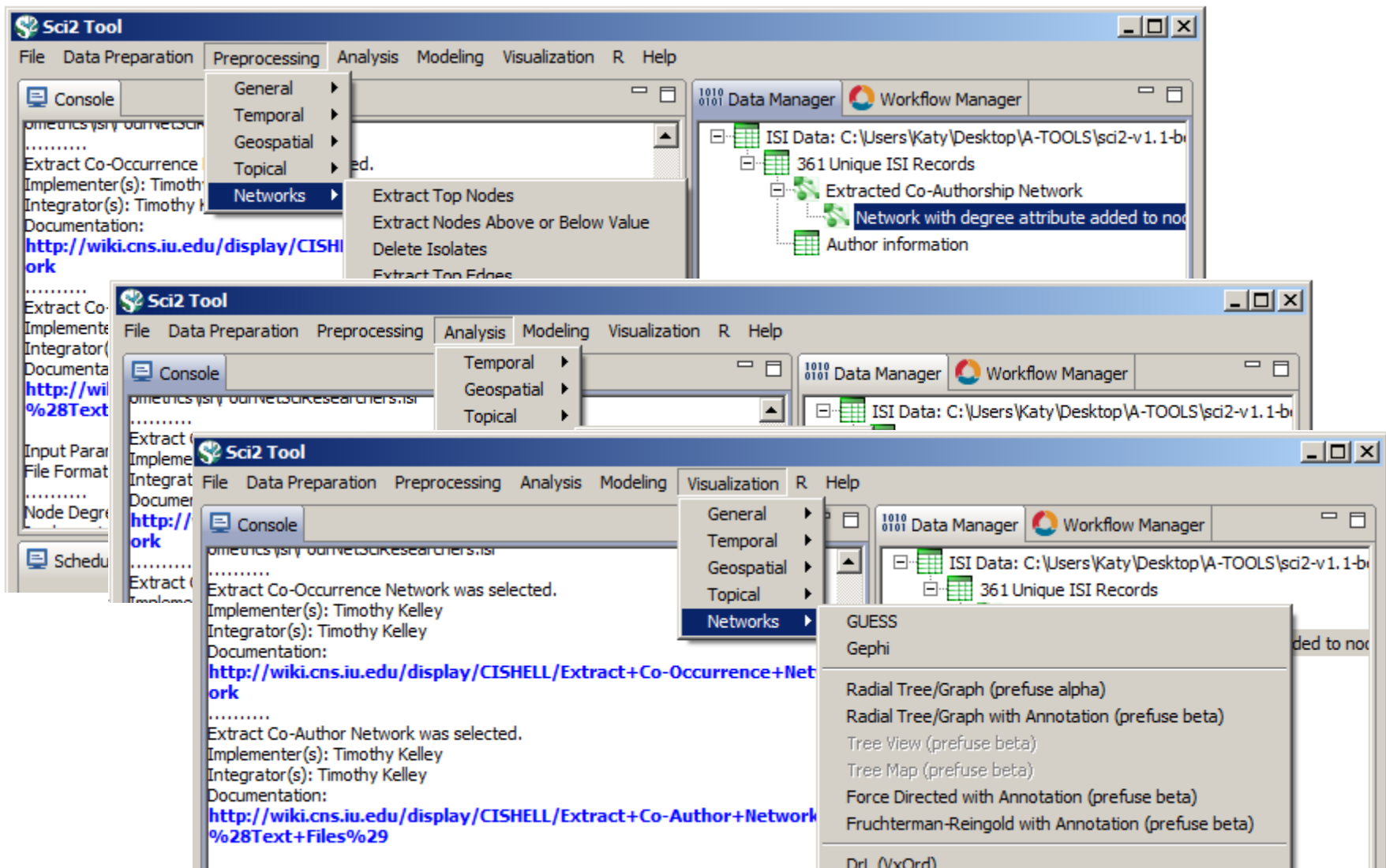


All workflows are recorded into a log file (see /sci2/logs/...), and can be re-run for easy replication. If errors occur, they are saved in a error log to ease bug reporting.

All algorithms are documented online; workflows are given in Sci2 Manual at <http://sci2.wiki.cns.iu.edu>

Sci2 Tool Interface Components

Download for free at <http://sci2.cns.iu.edu>



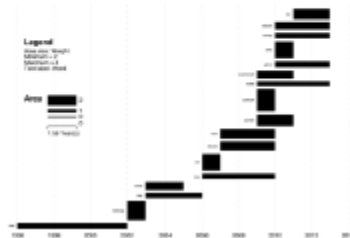
Load **One** File and Run **Many** Analyses and Visualizations

| Times Cited | Publication Year | City of Publisher | Country | Journal Title (Full) | Title | Subject Category | Authors |
|-------------|------------------|-------------------|---------|--|---|---|--|
| 12 | 2011 | NEW YORK | USA | COMMUNICATIONS OF THE ACM | Plug-and-Play Macroscopes | Computer Science | Borner, K |
| 18 | 2010 | MALDEN | USA | CTS-CLINICAL AND TRANSLATIONAL SCIENCE | Advancing the Science of Team Science | Research & Experimental Medicine | Falk-Krzesinski, HJ Borner, K Contractor, N Fiore, SM Hall, KL Keyton, J Spring, B Stokols, D Trochim, W Uzzi, B |
| 13 | 2010 | WASHINGTON | USA | SCIENCE TRANSLATIONAL MEDICINE | A Multi-Level Systems Perspective for the Science of Team Science | Cell Biology Research & Experimental Medicine | Borner, K Contractor, N Falk-Krzesinski, HJ Fiore, SM Hall, KL Keyton, J Spring, B Stokols, D Trochim, W Uzzi, B |

Statistical Analysis—p. 44

| Location | Count | # Citations |
|----------------|-------|-------------|
| Netherlands | 13 | 292 |
| United States | 9 | 318 |
| Germany | 11 | 36 |
| United Kingdom | 1 | 2 |

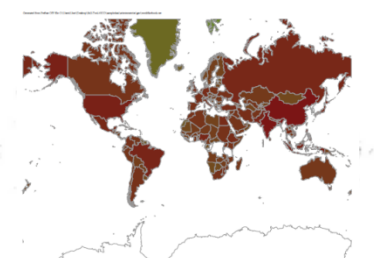
Temporal Burst Analysis—p. 48



Geospatial Analysis—p. 52



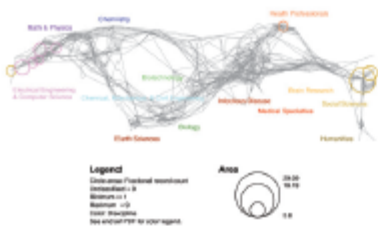
Geospatial Analysis—p. 52



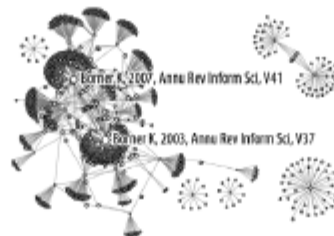
Load **One** File and Run **Many** Analyses and Visualizations

| Times Cited | Publication Year | City of Publisher | Country | Journal Title (Full) | Title | Subject Category | Authors |
|-------------|------------------|-------------------|---------|--|---|---|--|
| 12 | 2011 | NEW YORK | USA | COMMUNICATIONS OF THE ACM | Plug-and-Play Macroscopes | Computer Science | Borner, K |
| 18 | 2010 | MALDEN | USA | CTS-CLINICAL AND TRANSLATIONAL SCIENCE | Advancing the Science of Team Science | Research & Experimental Medicine | Falk-Krzesinski, HJ Borner, K Contractor, N Fiore, SM Hall, KL Keyton, J Spring, B Stokols, D Trochim, W Uzzi, B |
| 13 | 2010 | WASHINGTON | USA | SCIENCE TRANSLATIONAL MEDICINE | A Multi-Level Systems Perspective for the Science of Team Science | Cell Biology Research & Experimental Medicine | Borner, K Contractor, N Falk-Krzesinski, HJ Fiore, SM Hall, KL Keyton, J Spring, B Stokols, D Trochim, W Uzzi, B |

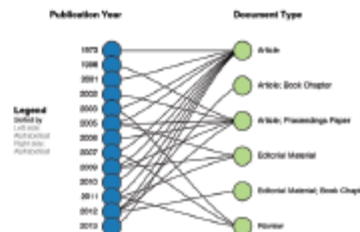
Topical Analysis—p. 56



Paper Citation Network—p. 60



Bi-Modal Network—p. 60



Co-author and many other bi-modal networks.

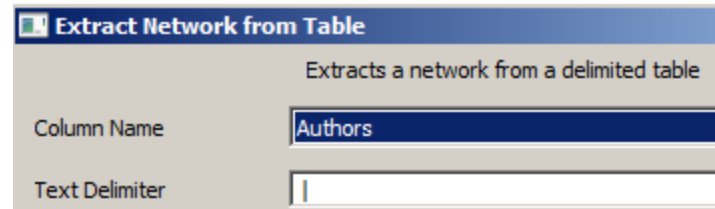
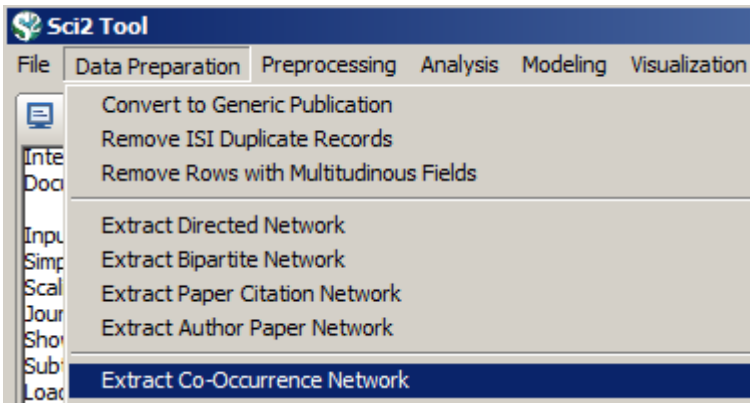
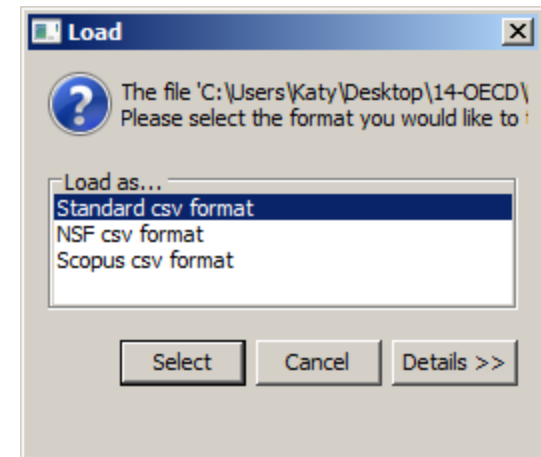
Load **One** File and Run **Many** Analyses and Visualizations

Download 20publications.csv from

<http://wiki.cns.iu.edu/download/attachments/1245848/20publications.csv?version=1&modificationDate=1403450235951>

In Sci2, use 'File > Load' and load file as 'Standard csv format'.

Run 'Data Preparation > Extract Co-Occurrence Network' with parameters:



Co-author network will appear in **Data Manager**.

Load **One** File and Run **Many** Analyses and Visualizations

Run 'Analysis > Network Analysis Toolkit (NAT)' to get basic properties:

Nodes: 65

Isolated nodes: 0

Edges: 404

No self loops were discovered.

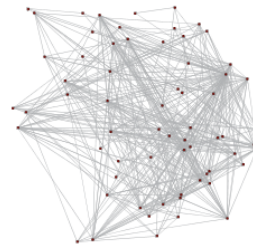
Average degree: 12.4308

The largest connected component consists of 65 nodes.

Density (disregarding weights): 0.1942

Select 'Extracted Network on Column Authors' network in Data Manager and run 'Visualization > GUESS' to open GUESS with file loaded.

Initial layout is random:



In GUESS, apply 'Layout > GEM':



Information Visualization MOOC (IVMOOC)

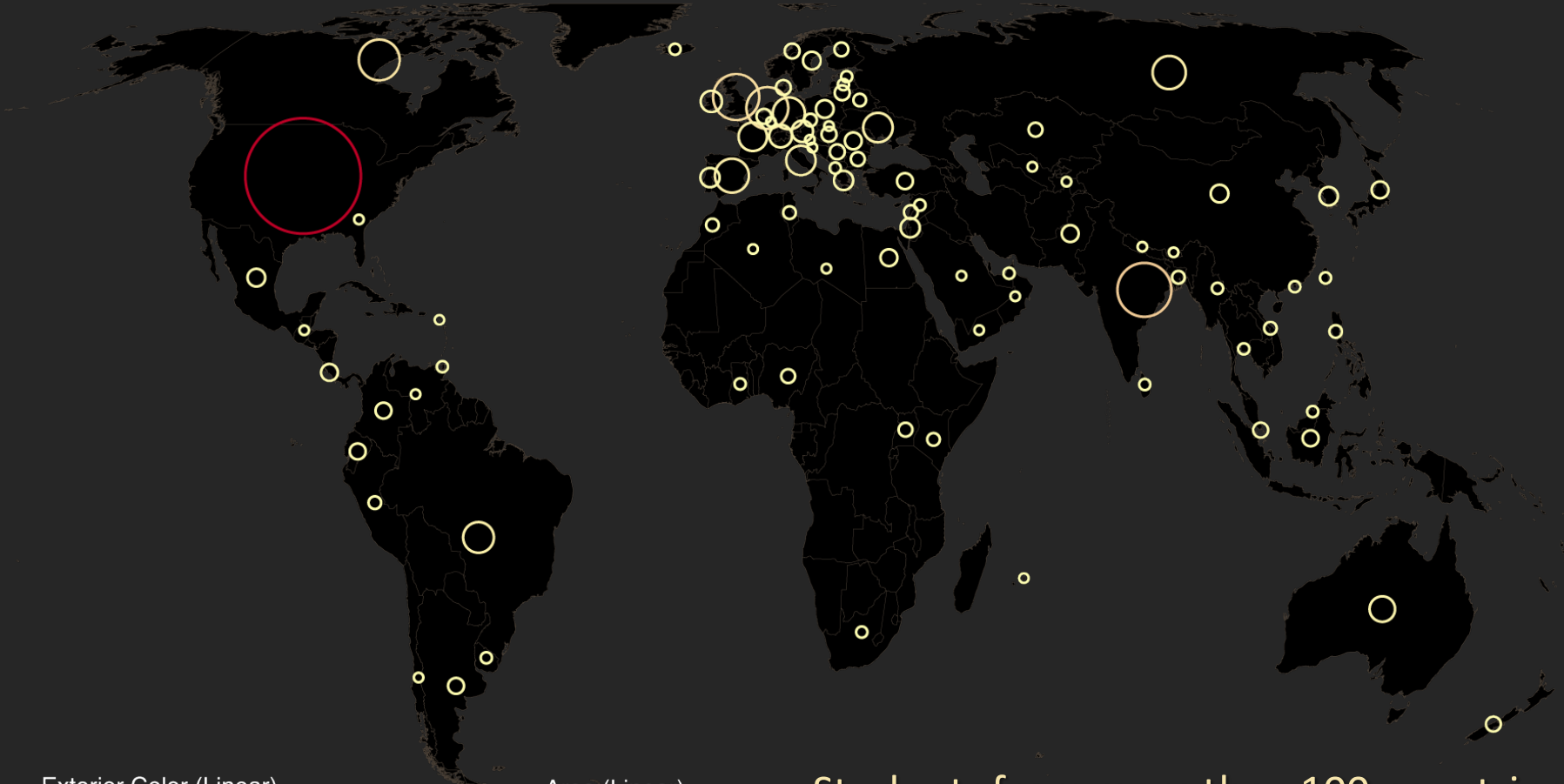
Teaches the Sci2 Tool



Register for free: <http://ivmooc.cns.iu.edu>. Class started Jan 12, 2016.

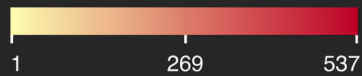
The Information Visualization MOOC

ivmooc.cns.iu.edu



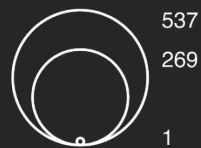
Exterior Color (Linear)

count



Area (Linear)

count



Students from more than 100 countries

350+ faculty members

#ivmooc

Course Schedule

Part 1: Theory and Hands-On

- **Session 1** – Workflow Design and Visualization Framework
- **Session 2** – “When:” Temporal Data
- **Session 3** – “Where:” Geospatial Data
- **Session 4** – “What:” Topical Data

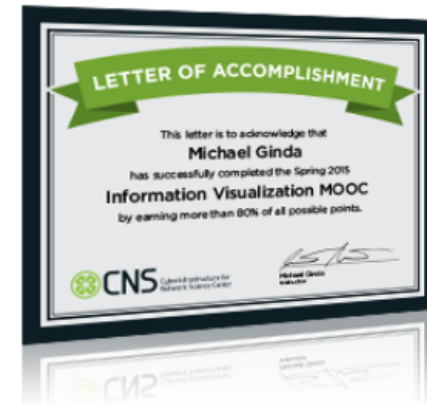
Mid-Term

- **Session 5** – “With Whom:” Trees
- **Session 6** – “With Whom:” Networks
- **Session 7** – Dynamic Visualizations and Deployment

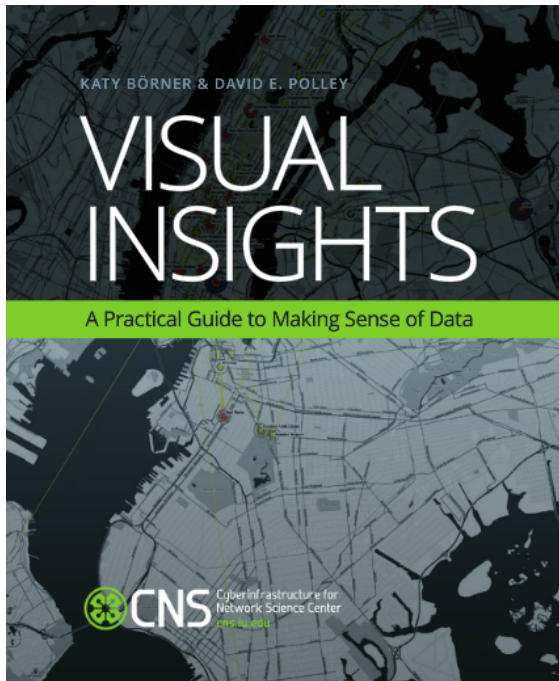
Final Exam

Part 2: Students work in teams on client projects.

Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

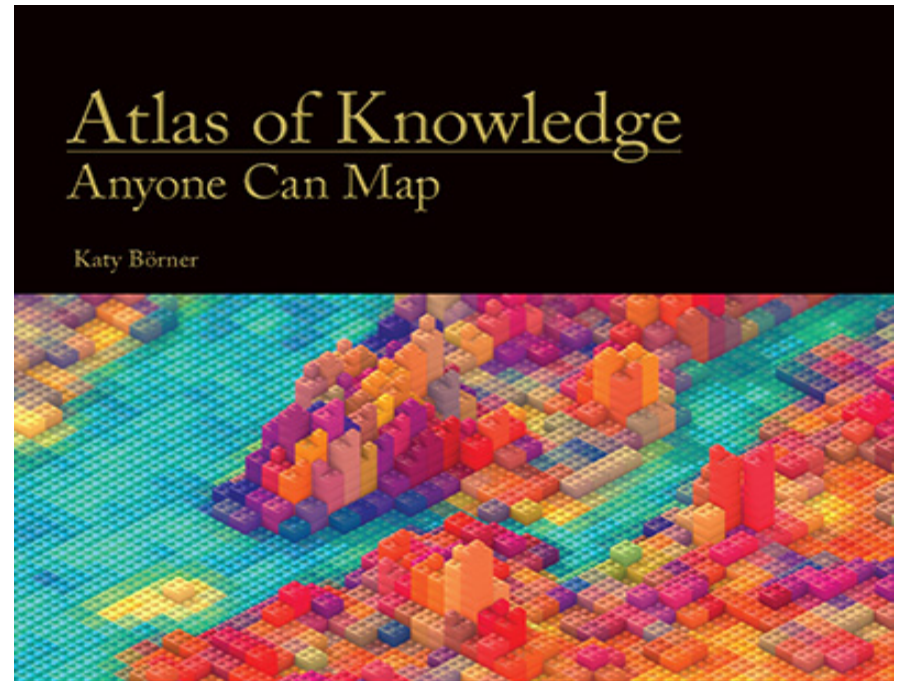


Books Used in the IVMOOC



Teaches timely knowledge:

Advanced algorithms, tools, and hands-on workflows.



Teaches timeless knowledge:

Visualization framework—exemplified using generic visualization examples and pioneering visualizations.

IVMOOC App

The “IVMOOC Flashcards” app can be downloaded from Google Play and Apple iOS stores.

The image displays two screenshots of the IVMOOC Flashcards app listing. The left screenshot is from the Google Play store, showing the app's icon (a green circular logo with a network diagram) and a map of the United States. The right screenshot is from the iTunes Preview page, providing detailed information about the app.

Google Play Store Listing:

- Search bar: ivmooc
- App Name: IVMOOC Flashcards
- Developer: Studio Cypher
- Category: Education
- Age Rating: 3+
- Compatibility: This app is compatible with all of your devices.
- Thumbnail: A map of the United States with a color-coded overlay.

iTunes Preview Page:

- App Name: Information Visualization MOOC Flashcards
- Developer: Studio Cypher
- Category: Education
- Updated: Feb 01, 2016
- Version: 1.1.5
- Size: 50.2 MB
- Language: English
- Seller: Studio Cypher, LLC
- © Studio Cypher
- Rated 4+
- Compatibility: Requires iOS 6.0 or later. Compatible with iPhone, iPad, and iPod touch.
- Description: Embrace the power of information visualization and render data into actionable insights. These flashcards show more than 50 visualizations taught in the Information Visualization MOOC (IVMOOC) at [Information Visualization MOOC Flashcards Support](#).
- What's New in Version 1.1.5: Small text bug fixes.
- Screenshots: IVMOOC Flashcard Example showing three people at a podium.

How to Classify Different Visualizations?

By

- User insight needs?
- User task types?
- Data to be visualized?
- Data transformation?
- Visualization technique?
- Visual mapping transformation?
- Interaction techniques?
- Or ?

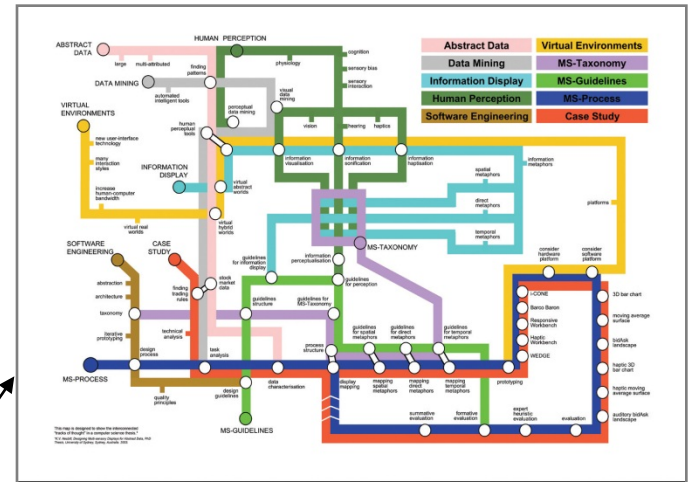


Different Question Types



Terabytes of data

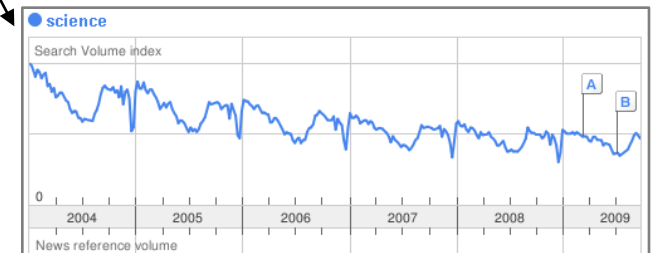
Descriptive & Predictive Models



Find your way



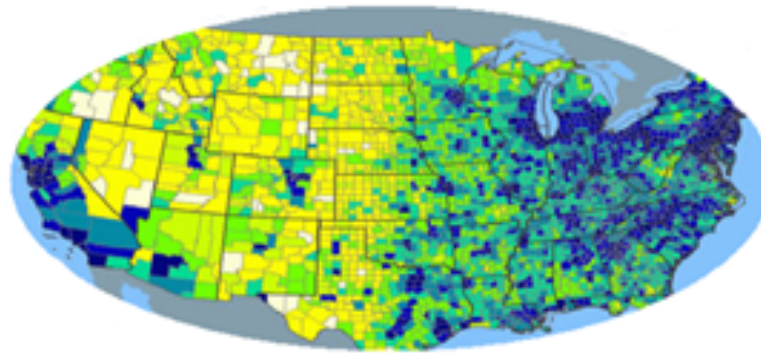
Find collaborators, friends



Identify trends

Different Levels of Abstraction/Analysis

Macro/Global
Population Level



Meso/Local
Group Level



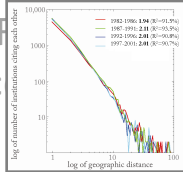

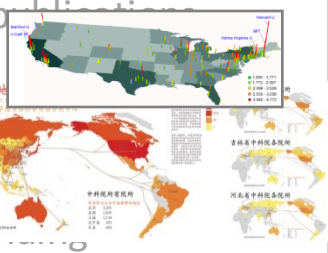

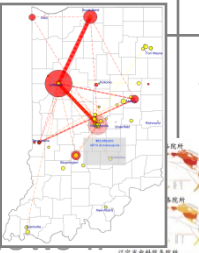

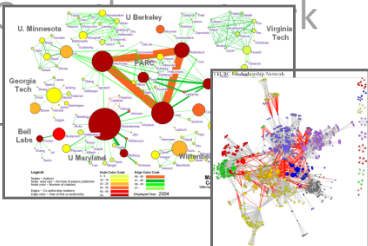
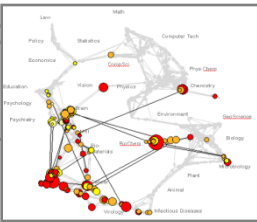
Micro
Individual Level



Type of Analysis vs. Level of Analysis

| | <i>Micro/Individual (1-100 records)</i> | <i>Meso/Local (101–10,000 records)</i> | <i>Macro/Global (10,000 < records)</i> |
|--|--|---|--|
| <i>Statistical Analysis/Profiling</i> | Individual person and their expertise profiles | Larger labs, centers, universities, research domains, or states | All of NSF, all of USA, all of science. |
| <i>Temporal Analysis (When)</i> | Funding portfolio of one individual | Mapping topic bursts in 20 years of <i>PNAS</i> | 113 years of physics research |
| <i>Geospatial Analysis (Where)</i> | Career trajectory of one individual | Mapping a state's intellectual landscape | <i>PNAS</i> publications |
| <i>Topical Analysis (What)</i> | Base knowledge from which one grant draws. | Knowledge flows in chemistry research | VxOrd/Topic maps of NIH funding |
| <i>Network Analysis (With Whom?)</i> | NSF Co-PI network of one individual | Co-author network | NIH's core competency |

Type of Analysis vs. Level of Analysis

| | Micro/Individual (1-100 records) | Meso/Local (101–10,000 records) | Macro/Global (10,000 < records) |
|---------------------------------------|---|--|---|
| Statistical Analysis/Profiling | Individual person and their expertise profiles | Larger labs, centers, universities, research domains, or states | All of NSF, all of science  |
| Temporal Analysis (When) | Funding portfolio of one individual | Scientific bursts of PNAS | 113 years of PNAS research  |
| Geospatial Analysis (Where) | Career trajectory of one individual | Wrapping a state intellectual landscape | PNAS publications  |
| Topical Analysis (What) |  | Knowledge from chemistry research |  |
| Network Analysis (With Whom?) | NSF network of one  |  | NIH's network  |

Clients

List of Clients



Project Title: Isis: 100 Years

Client Name: Jay Malone

Project goal/scientific or practical value: A visual representation Isis' contributors and locales over the past 100 years. Isis is the journal of the History of Science Society. This representation will provide a dynamic picture of how scholarship in the history of science has shifted over the past century.

Information on dataset(s) to be used: Citation information, author locale, and issue number for Isis publications.

Relevant publications, websites, etc: <http://www.press.uchicago.edu/ucp/journals/journal/isis.html>

Conditions under which students can publish results and/or add project results to their resume: Client would like to approve results.



Project Title: e-Xploration

Client Name: Luyi

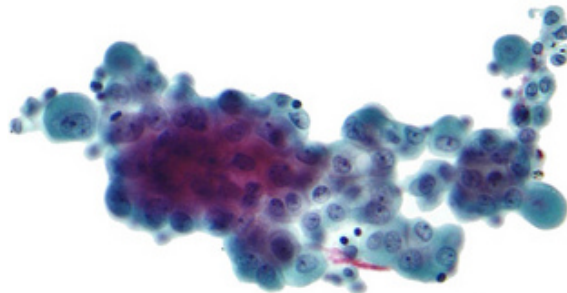
Project goal/scientific or practical value: e-Xploration is an agent-based model for the ethnographic observation and the registry, analysis, and interpretation of social practices in virtual communities for intervention in the development of collaboration and cooperation. This project will analyze the interactions between subjects and objects in a platform collaborative community called OYCIB, a project based on e-Xploration (e-crick.net).

Information on dataset(s) to be used: I can provide a data base in .graphml format for the students. The file .graphml contains the interactions between subjects and objects in a platform collaborative community called OYCIB. In the level of practice, it is not necessary that students know agent-based models for using the database. But, in another level, for example: the collaborate level for the OYCIB development, it is necessary to have basic knowledge in AMS or MAS and another competences like PHP and MySQL.

Relevant publications, websites, etc: <http://www.e-crick.net/logs>

Conditions under which students can publish results and/or add project results to their resume: If any person or institution use my dataset or another info about eXploration (e-crick.net, oycib.net), I need to approve the results and appear as co-author.

<http://ivmooc.cns.iu.edu/clients.html>



Mesothelioma

Main title topics in Medline papers

Mesothelioma (or, more precisely, malignant mesothelioma) is a rare form of cancer that develops from transformed cells originating in the mesothelium, the protective lining that covers many of the internal organs of the body. It is usually caused by exposure to asbestos.

The most common anatomical site for the development of mesothelioma is the pleura (the outer lining of the lungs and internal chest wall), but it can also arise in the peritoneum (the lining of the abdominal cavity), and the pericardium (the sac that surrounds the heart), or the tunica vaginalis (a sac that surrounds the testis).

Most people who develop mesothelioma have worked in jobs where they inhaled asbestos, or were exposed to asbestos dust and fibers in other ways. It has also been suggested that washing clothes of a family member who worked with asbestos increases their risk for developing mesothelioma. Unlike lung cancer, there seems to be no association between mesothelioma and tobacco smoking, but smoking greatly increases the risk of other asbestos-induced cancers. Some people who were exposed to asbestos have collected damages for asbestos-related disease, including mesothelioma. Compensation via asbestos funds or class action lawsuits is an important issue in law practices regarding mesothelioma.

MALIGNANT PLEURAL CYSTIC BENIGN DIAGNOSIS

1982 1987 1992 1997 2002 2007

Asbestos use greatly increased during World War II. Since the early 1940s, millions of American workers have been exposed to asbestos.

In the town of Wittenoom, asbestos-containing mine waste was used to cover schoolyards and playgrounds.

In 1960 an article published by Wagner et al. was seminal in establishing mesothelioma as a disease arising from exposure to asbestos.

In 1962 McNulty reported the first diagnosed case of malignant mesothelioma in an Australian asbestos worker.

In 1965 an article in the British Journal of Industrial Medicine established that people who lived in the neighbourhoods of asbestos factories and mines, but did not work in them, had contracted mesothelioma.

In 1974 the first public warnings of the dangers of blue asbestos were published in a cover story called "The Blue Killer in Your Home?" in Australia's Bulletin.

By 1979 the first writs for negligence related to Wittenoom were issued against CSR and its subsidiary ABA, and the Asbestos Diseases Society was formed to represent the Wittenoom victims.

From 1980 to the late 1990s, the death rate from mesothelioma in the USA increased from 2,000 per year to 3,000, with men four times more likely to acquire it than women.

It has been estimated that incidence may have peaked at 95 per 1,000,000 in the United States in 2004.

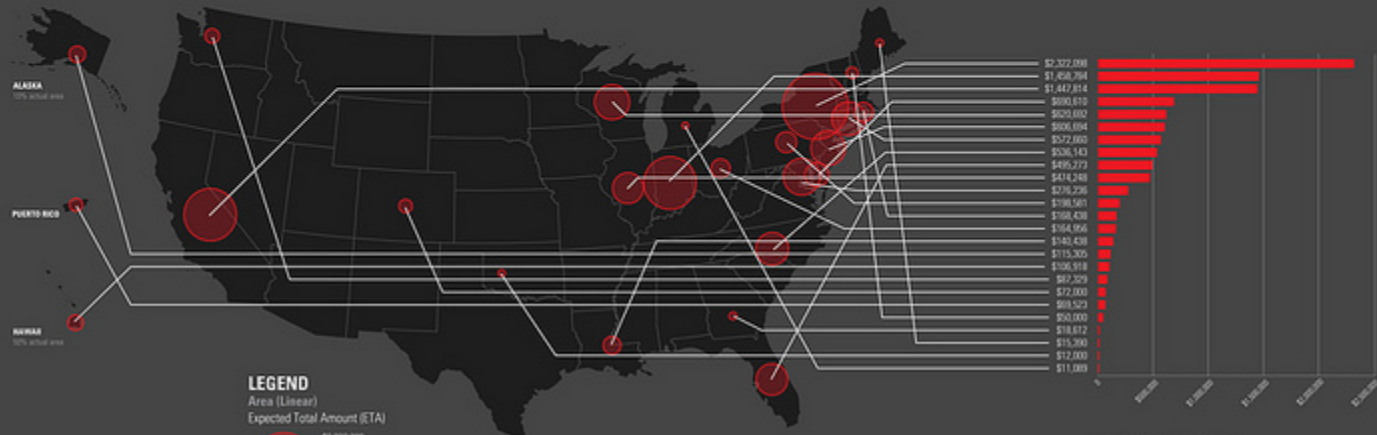
1932 1937 1942 1947 1952 1957 1962 1967 1972 1977



How To Read This Map
This temporal bar graph visualization represents each record as a horizontal bar with a specific start and end date and a text label on its left side. The area of each bar encodes surges of interest - and surges' magnitude - in the topics identified in the labels (original words were stemmed).

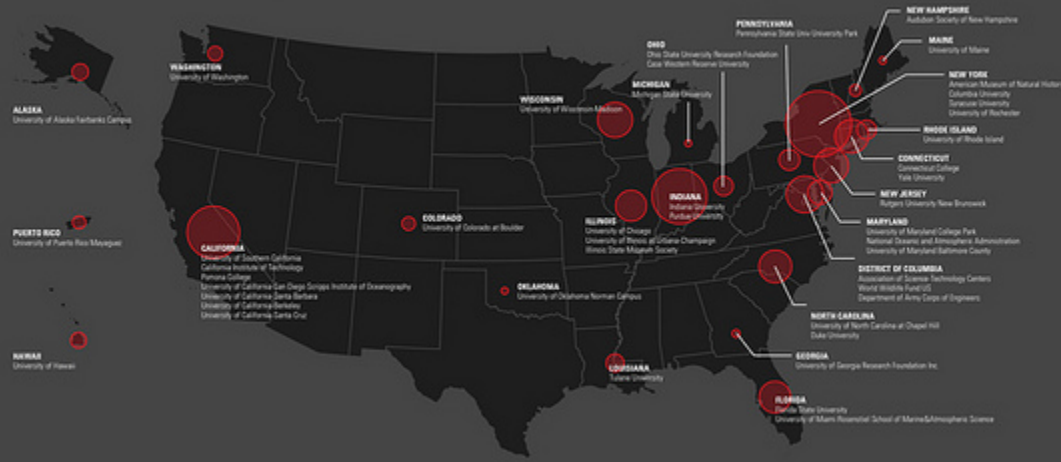
National Science Foundation Awards to Institutions Conducting "Global Warming" Research from 1979-2009

Data from Scholarly Database (sdb.cns.iu.edu/) | Visualized with Sci2 Tool (sci2.cns.iu.edu/)
Feb 08, 2013 | 04:44:14 PM CST | by Michael J. Stammer



ABOUT THIS MAP
This map was created by searching through NSF data for titles that contain the term "global warming" on the Scholarly Database (<http://sdb.cns.iu.edu/>). The results shown here are organized by State and Award Amount (top) and State and Institutions within them receiving the awards (bottom).

HOW TO READ THIS MAP
This proportional symbol map shows 51 U.S. states and other jurisdictions using the Albers equal-area conic projection with Alaska, Puerto Rico, and Hawaii inset. Each dataset record is represented by a circle centered at its geolocation. The area, interior color, and exterior color of each circle may represent numeric attribute values. Minimum and maximum data values are given in the legend.



References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255. <http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1/

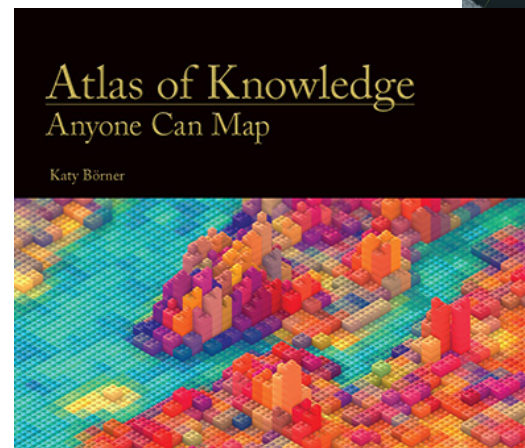
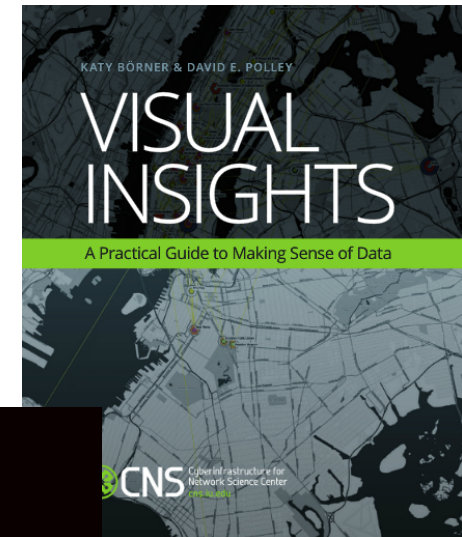
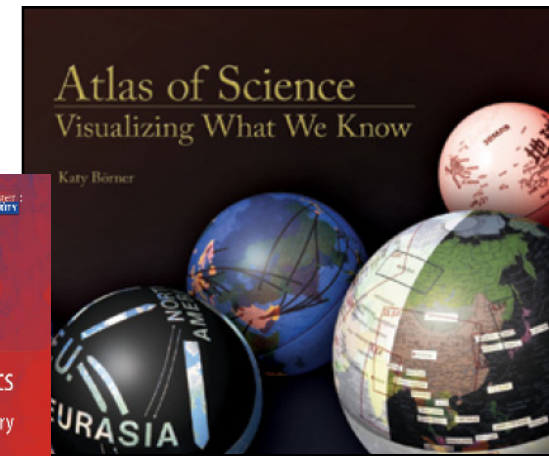
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. <http://scimaps.org/atlas>

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

Börner, Katy (2015) **Atlas of Knowledge: Anyone Can Map**. The MIT Press. <http://scimaps.org/atlas2>






We work closely with clients to provide custom-made data, visualization, and software solutions

Research

 Open Data and Open Code for Big Science of Science Studies


Latest News

 Put your money where your citations are: a proposal for a new funding system (website accessed 9/05/13)

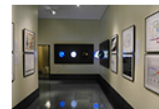
Upcoming Events

- OCT 1** Katy Börner attends PIUG 2013 Northeast Conference
- 10.13** Katy Börner presents Mapping Science Exhibit at WSSF
- 10.15** Ted Polley & Google Team present IVMOOC at EDUCAUSE
- 10.22** Katy Börner presents at the SciELO 15 Years Conference


Development

 Behind the scenes of the design and development of *AcademyScope*


Outreach

 See some of the most fascinating data visualizations in the world.


Videos

 Watch Katy Börner's full presentation from TEDxBloomington

Teaching

 Successful IVMOOC will be offered again in January of 2014

Our Products

 We work closely with clients to provide custom-made data, visualization, and software solutions

All papers, maps, tools, talks, press are linked from <http://cns.iu.edu>

These slides will soon be at <http://cns.iu.edu/docs/presentations>

CNS Facebook: <http://www.facebook.com/cnscenter>

Mapping Science Exhibit Facebook: <http://www.facebook.com/mappingscience>