Atlas of Knowledge: Anyone Can Map

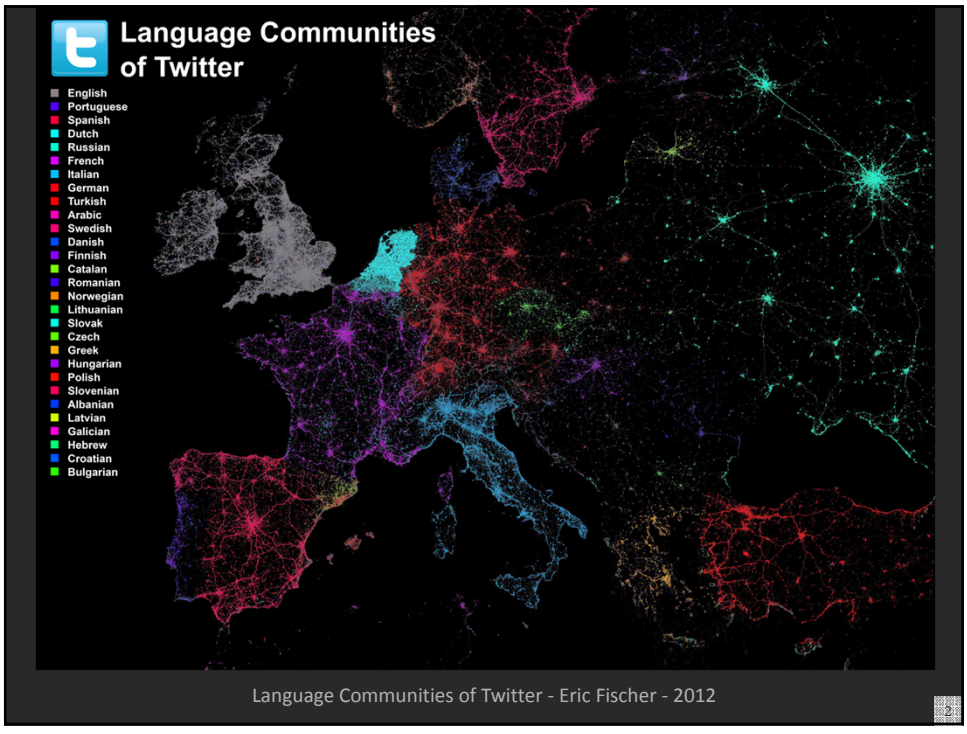
Katy Börner

Victor H. Yngve Professor of Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing, Indiana University, USA

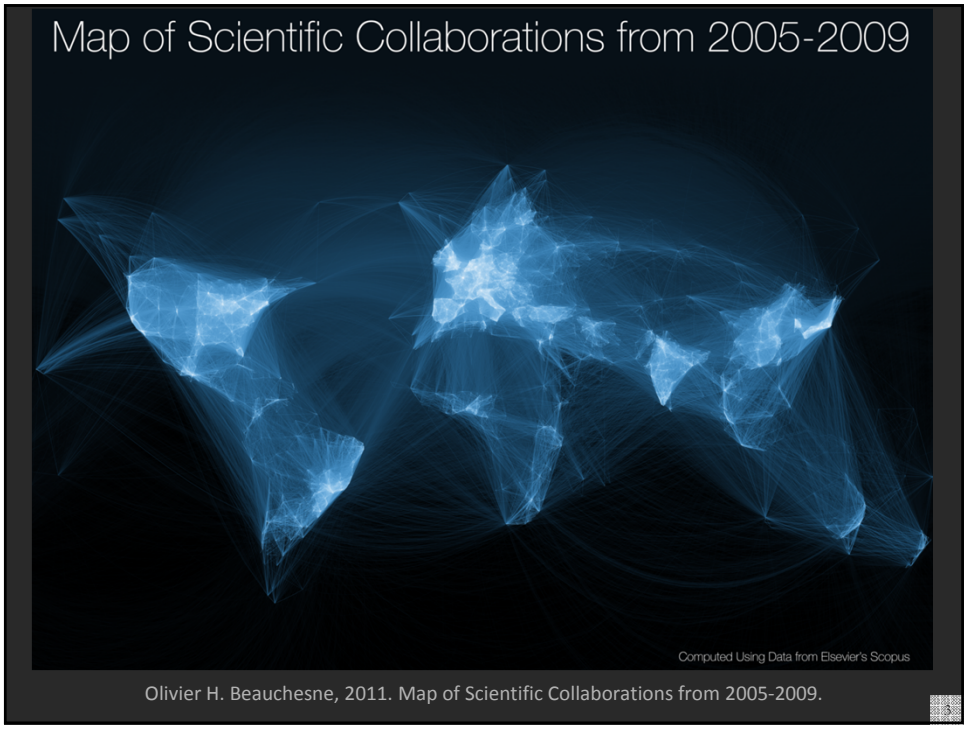
Mundaneum Online Art Talk
4pm CET April 22, 2015

Followed by a conversation with specialists
Dr. Andrea Scharnhorst (KNAW-DANS, The Netherlands) and Almila Akdag (KNAW-DANS, The Netherlands; UCLA, USA)

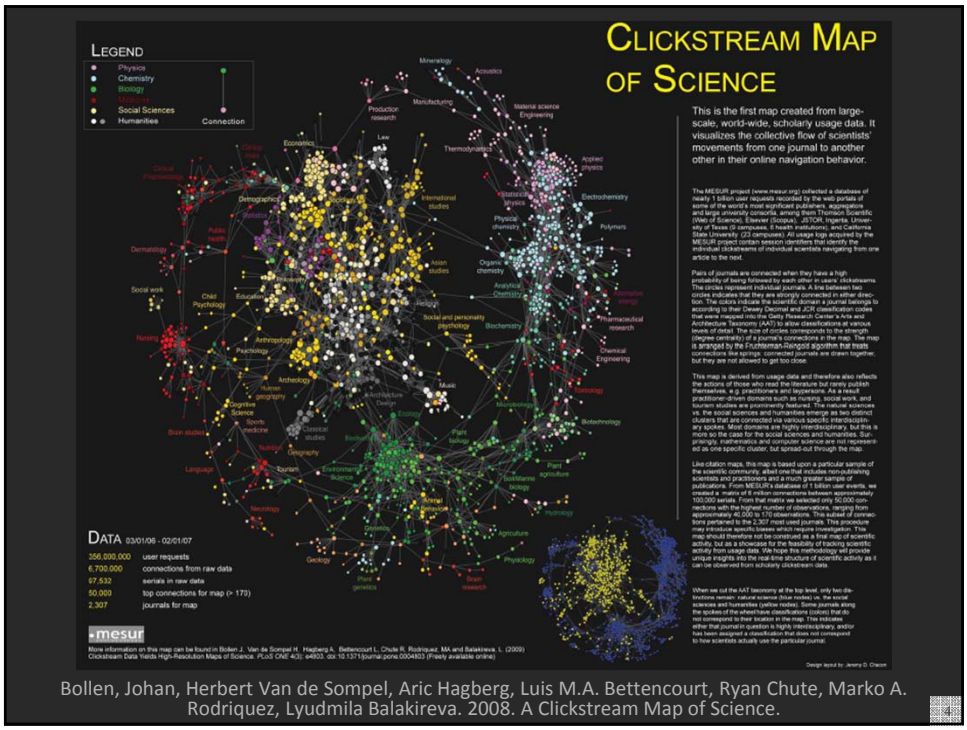
Language Communities of Twitter - Eric Fischer - 2012



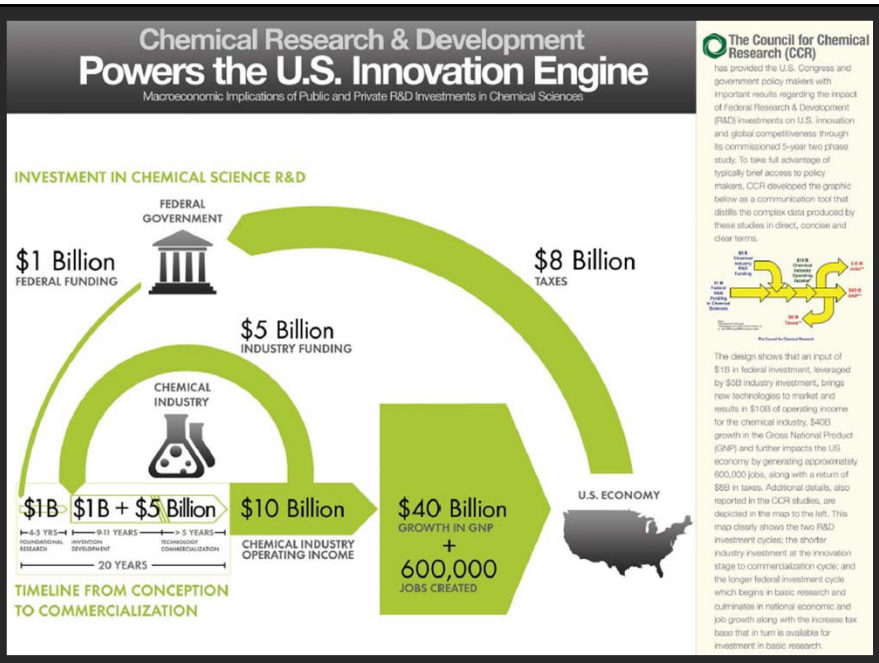
Language Communities of Twitter - Eric Fischer - 2012



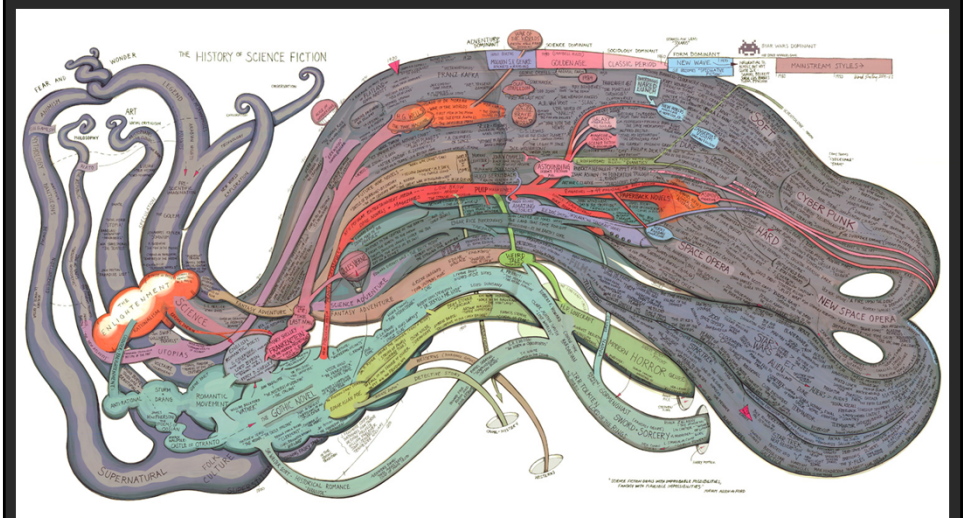
Olivier H. Beauchesne, 2011. Map of Scientific Collaborations from 2005-2009.



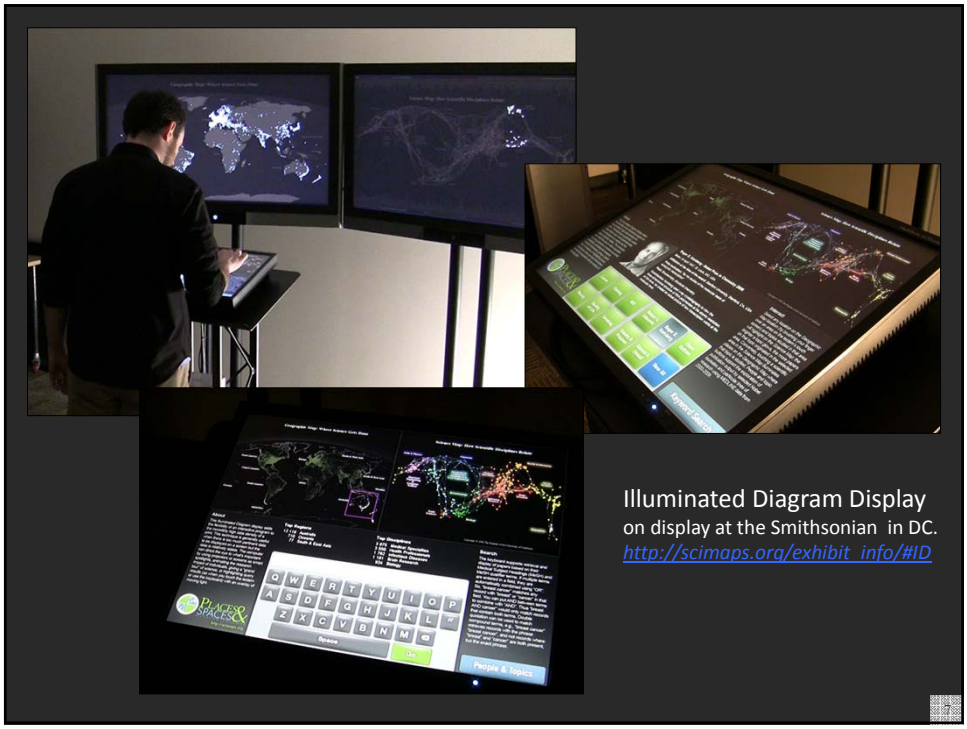
Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva. 2008. A Clickstream Map of Science.

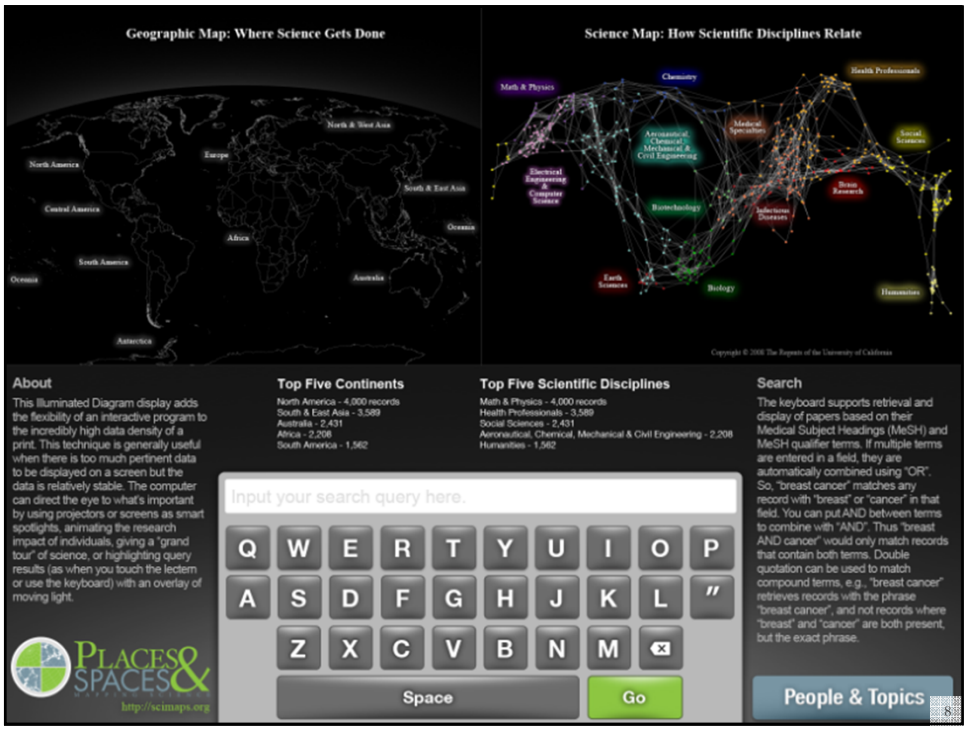


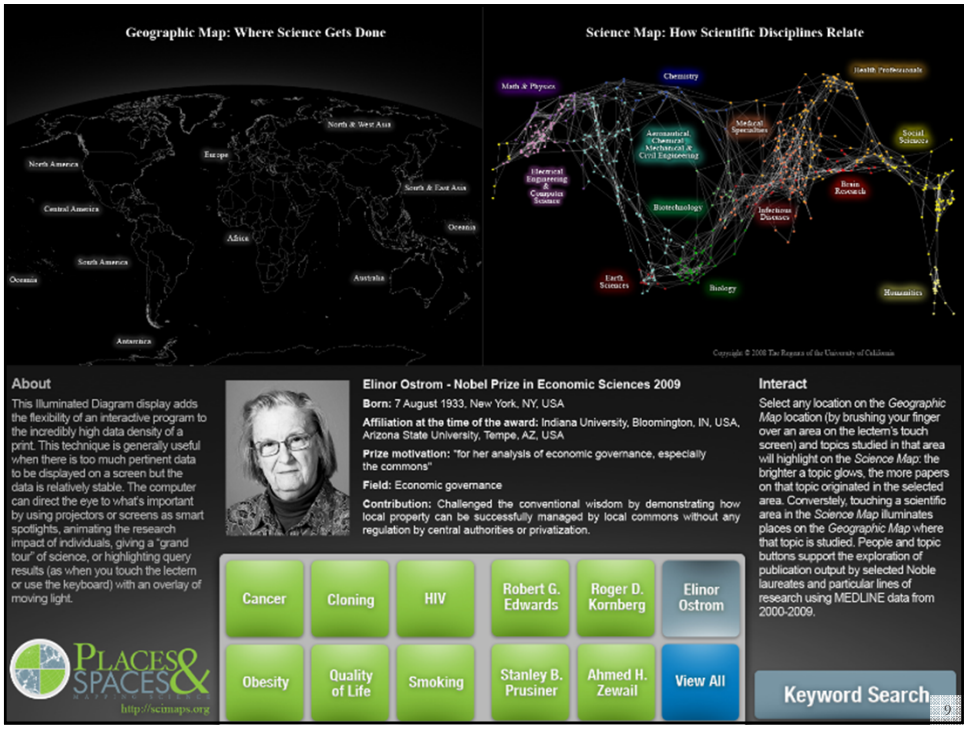
Council for Chemical Research. 2009. Chemical R&D Powers the U.S. Innovation Engine. Washington, DC. Courtesy of the Council for Chemical Research.

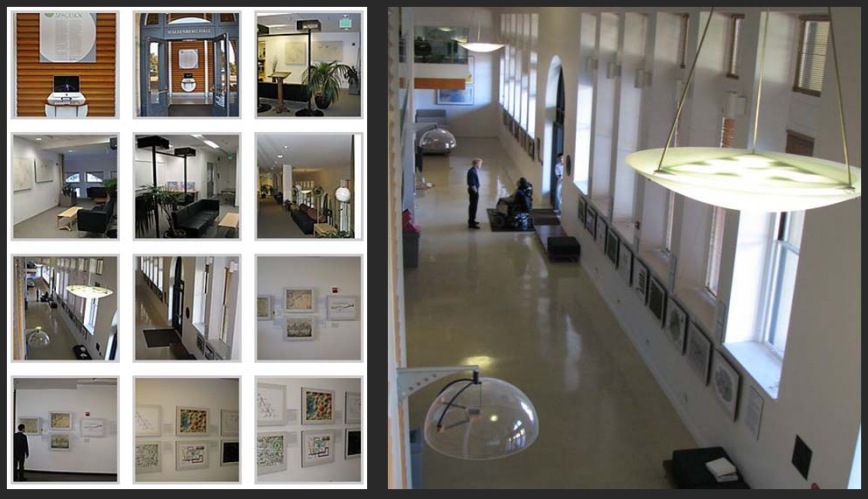


Ward Shelley . 2011. History of Science Fiction.


Illuminated Diagram Display
on display at the Smithsonian in DC.
http://scimaps.org/exhibit_info/#ID

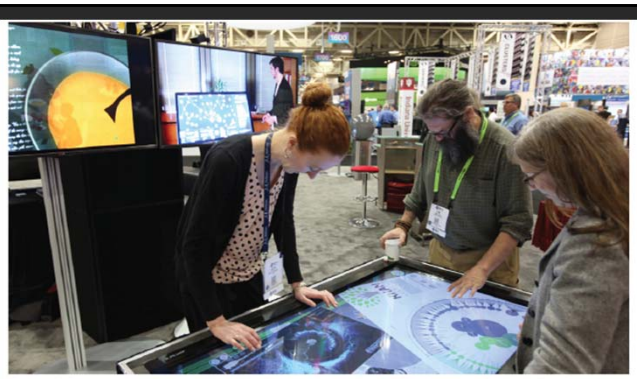

Geographic Map: Where Science Gets Done
Science Map: How Scientific Disciplines Relate
About
This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.
Top Five Continents
North America - 4,000 records
South & East Asia - 3,589
Australia - 2,431
Africa - 2,208
South America - 1,562
Top Five Scientific Disciplines
Math & Physics - 4,000 records
Health Professionals - 3,589
Social Sciences - 2,431
Aeronautical, Chemical, Mechanical & Civil Engineering - 2,208
Humanities - 1,562
Search
The keyboard supports retrieval and display of papers based on their Medical Subject Headings (MeSH) and MeSH qualifier terms. If multiple terms are entered in a field, they are automatically combined using "OR". So, "breast cancer" matches any record with "breast" or "cancer" in that field. You can put AND between terms to combine with "AND". Thus "breast AND cancer" would only match records that contain both terms. Double quotation can be used to match compound terms, e.g., "breast cancer" retrieves records with the phrase "breast cancer", and not records where "breast" and "cancer" are both present, but the exact phrase.
Input your search query here.
Space
Go
People & Topics
http://scimaps.org



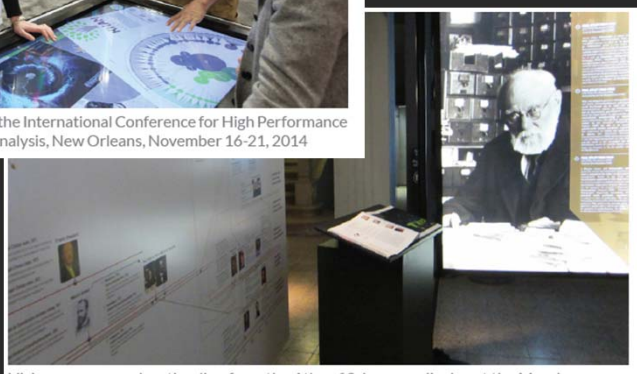


Mapping Science Exhibit on display at MEDIA X, Stanford University
http://mediax.stanford.edu, http://scaleindependentthought.typepad.com/photos/scimaps

10



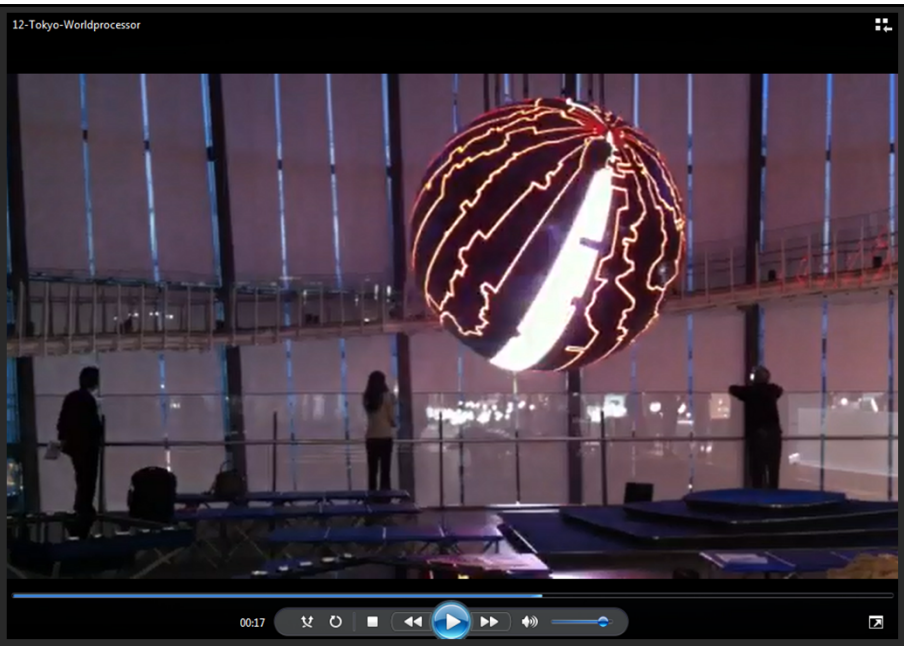
Places & Spaces maps on a touch table at the International Conference for High Performance Computing, Networking, Storage, and Analysis, New Orleans, November 16-21, 2014



Visionary approaches timeline from the *Atlas of Science* on display at the Mundaneum Museum, Mons, Belgium

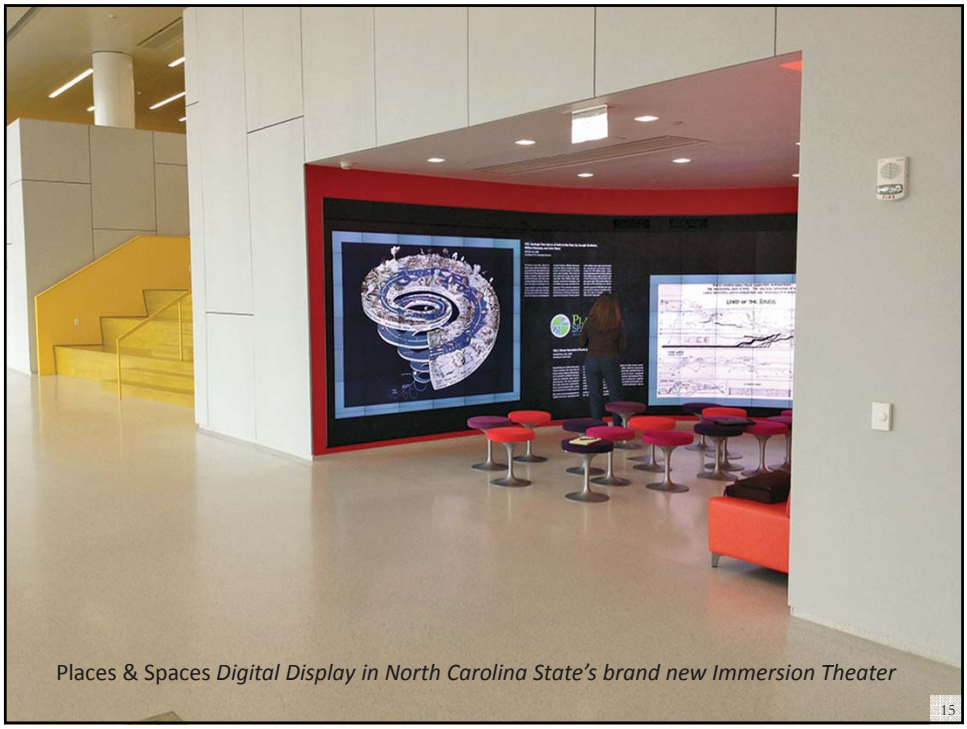
Katy Börner and Norbert Heber attend a screening of *Humanexus* at the Festival de Cannes, Cannes, France, May 20, 2014

International Science & Creativity Conference, Seoul, Republic of Korea, December 3-7,



Ingo Gunther's Worldprocessor globe design on display at the Giant Geo Cosmos OLED Display at the Museum of Emerging Science and Innovation in Tokyo, Japan

Science Maps in “Expedition Zukunft” science train visiting 62 cities in 7 months 12 coaches, 300 m long Opening was on April 23rd, 2009 by German Chancellor Merkel http://www.expedition-zukunft.de



Places & Spaces *Digital Display in North Carolina State's brand new Immersion Theater*

15

Places & Spaces: Mapping Science Exhibit

http://scimaps.org

Maps are available for sale and the exhibit can be hosted by anyone.

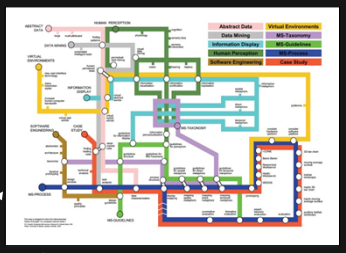
16

Empowering Anyone to Visualize STI

The Information Visualization MOOC

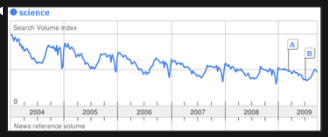
Terra bytes of data

Descriptive & Predictive Models

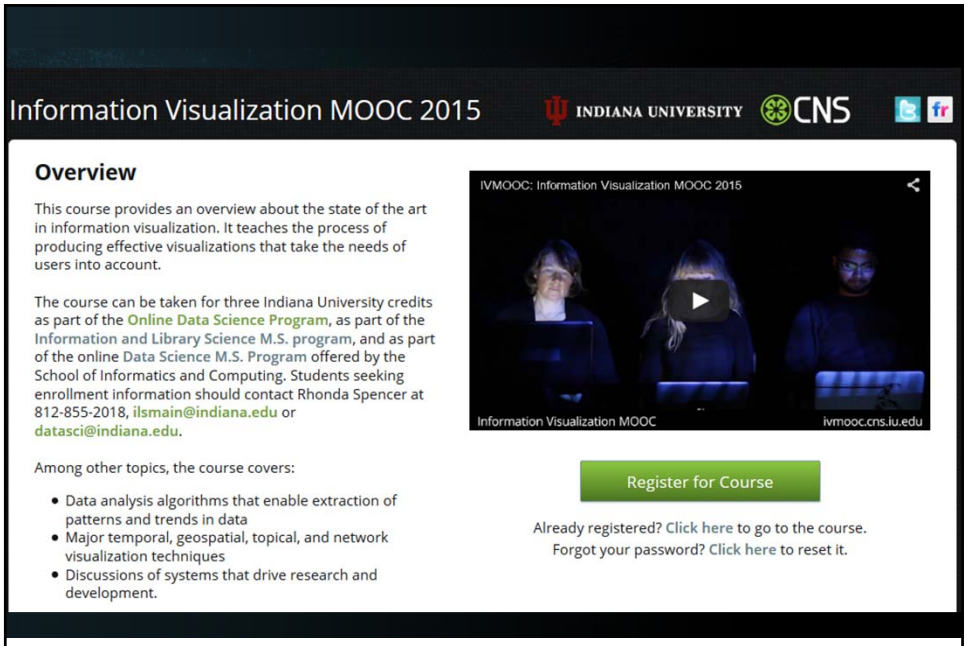


Find your way

Find collaborators, friends

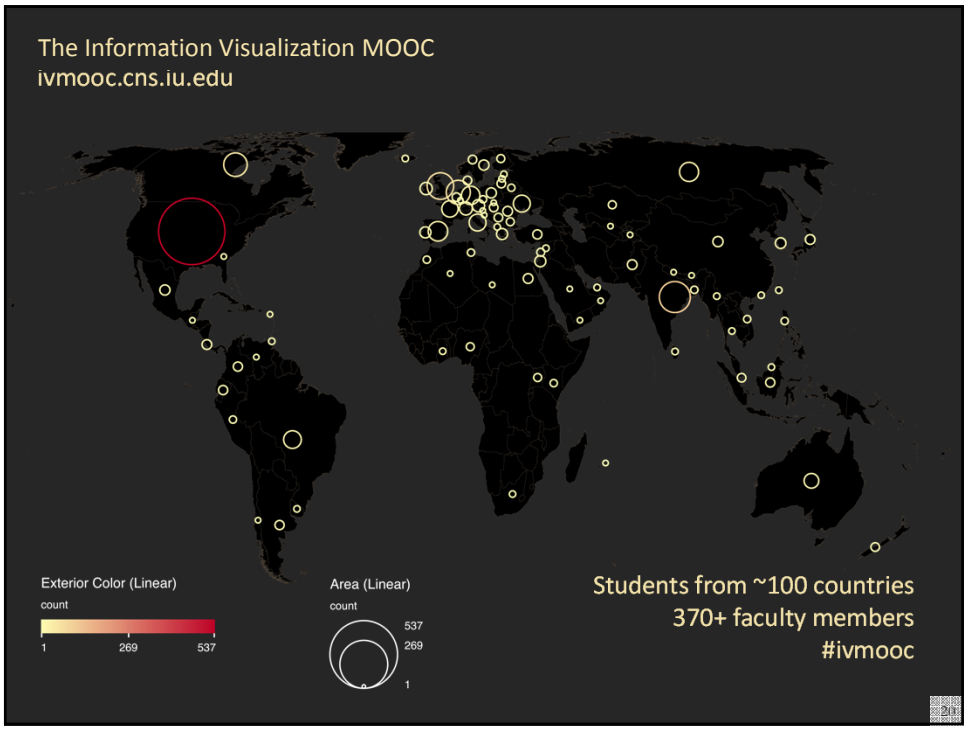


Identify trends



Register for free at http://ivmooc.cns.iu.edu. Class restarted in January 13, 2015.

19



20

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

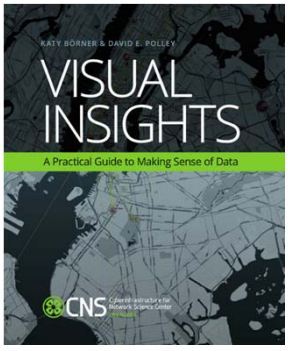
Final Exam

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%).

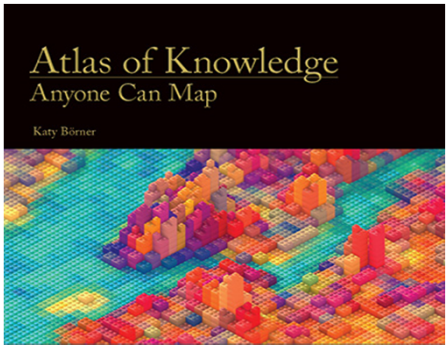
21

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.

22

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%).

23

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

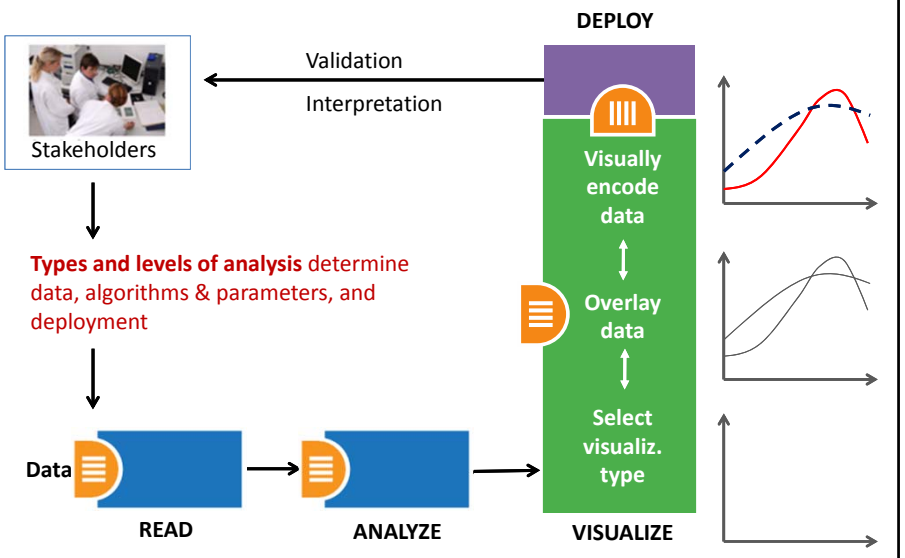
Final Exam

Part 2: Students work in teams on client projects.

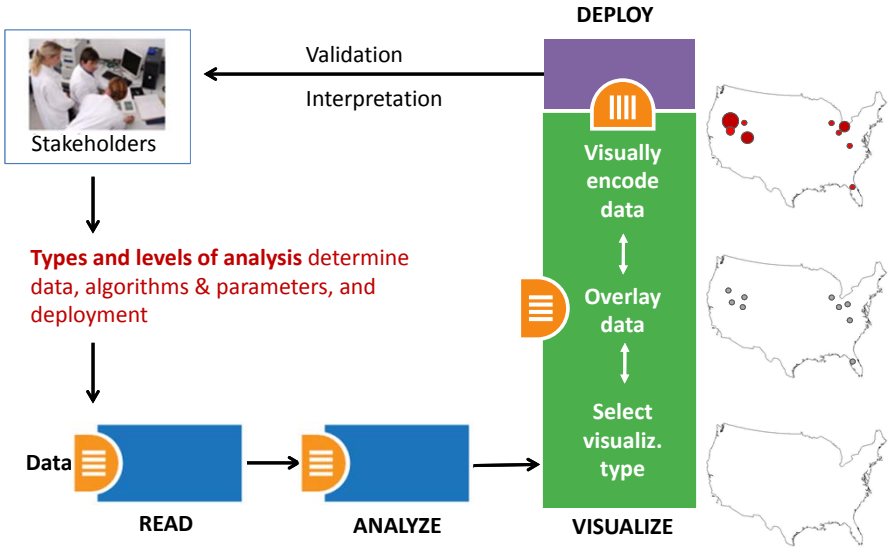
Final grade is based on Class Participation (10%), Midterm (30%), Final Exam (30%), and Client Project(30%).

24

Needs-Driven Workflow Design



Needs-Driven Workflow Design



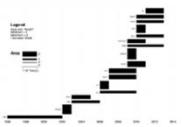
Load **One** File and Run **Many** Analyses and Visualizations

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

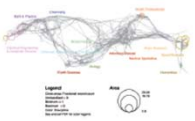
Geospatial Analysis–p. 52

27

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

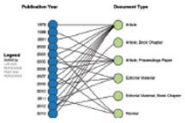
Times Cited	Publication Year	City of Publisher	Country	Journal Title (Full)	Title	Subject Category	Authors
12	2011	NEW YORK	USA	COMMUNICATIONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA	CTS-CLINICAL AND TRANSLATIONAL SCIENCE	Advancing the Science of Team Science	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.

28

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%).

29

Visualization Frameworks

How to Classify Different Visualizations?

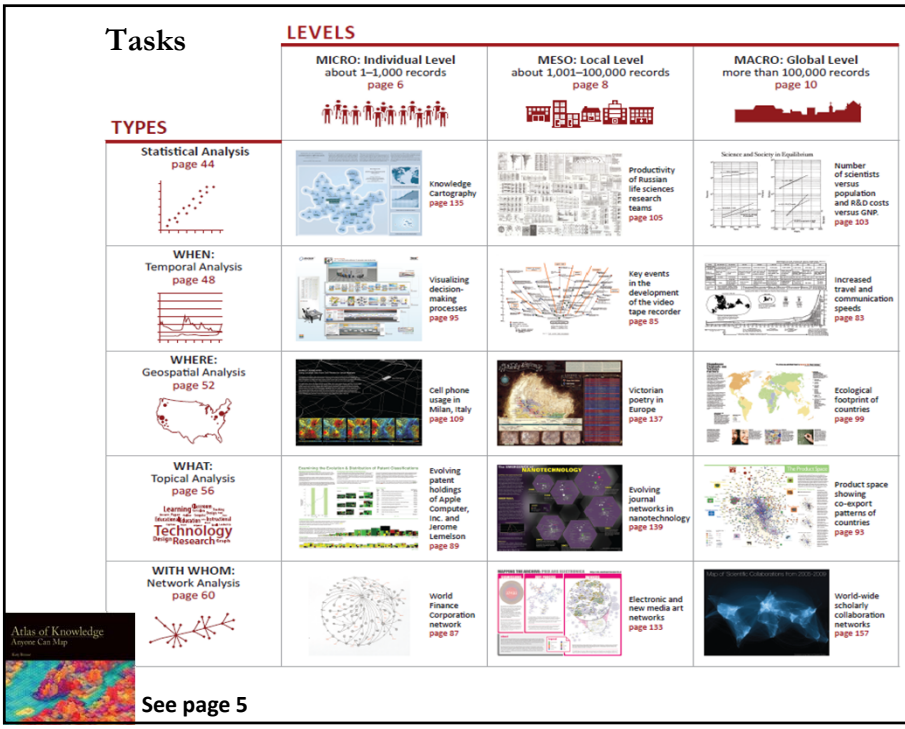
By

- User insight needs?
- User task types?

- Data to be visualized?
- Data transformation?

- Visualization technique?
- Visual mapping transformation?
- Interaction techniques?

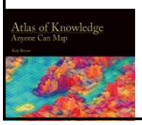
- Or ?



See page 5

32

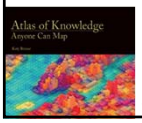
Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships	• nominal • ordinal • interval • ratio	• table • chart • graph • map • network layout	• geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs	• spatial position • retinal form color optics motion	• overview • zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion



See page 24

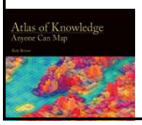
Basic Task Types

Bertin, 1967	Wehrend & Lewis, 1996	Few, 2004	Yau, 2011	Rendgen & Wiedemann, 2012	Frankel, 2012	Tool: Many Eyes	Tool: Chart Chooser	Börner, 2014
selection	categorize			category				categorize/ cluster
order	rank	ranking					table	order/rank/ sort
	distribution	distribution					distribution	distributions (also outliers, gaps)
	compare	nominal comparison & deviation	differences		compare and contrast	compare data values	comparison	comparisons
		time series	patterns over time	time	process and time	track rises and falls over time	trend	trends (process and time)
		geospatial	spatial relations	location		generate maps		geospatial
quantity		part-to-whole	proportions		form and structure	see parts of whole, analyze text	composition	compositions (also of text)
association	correlate	correlation	relationships	hierarchy		relations between data points	relationship	correlations/ relationships



See page 26

Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships	• nominal • ordinal • interval • ratio	• table • chart • graph • map • network layout	• geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs	• spatial position • retinal form color optics motion	• overview • zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion



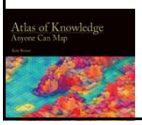
See page 24

Visualization Types (Reference Systems)

1. **Charts:** No reference system—e.g., Wordle.com, pie charts
2. **Tables:** Categorical axes that can be selected, reordered; cells can be color coded and might contain proportional symbols. Special kind of graph.
3. **Graphs:** Quantitative or qualitative (categorical) axes. Timelines, bar graphs, scatter plots.
4. **Geospatial maps:** Use latitude and longitude reference system. World or city maps.
5. **Network layouts:** Node position might depends on node attributes or node similarity. **Trees:** hierarchies, taxonomies, genealogies. **Networks:** social networks, migration flows.

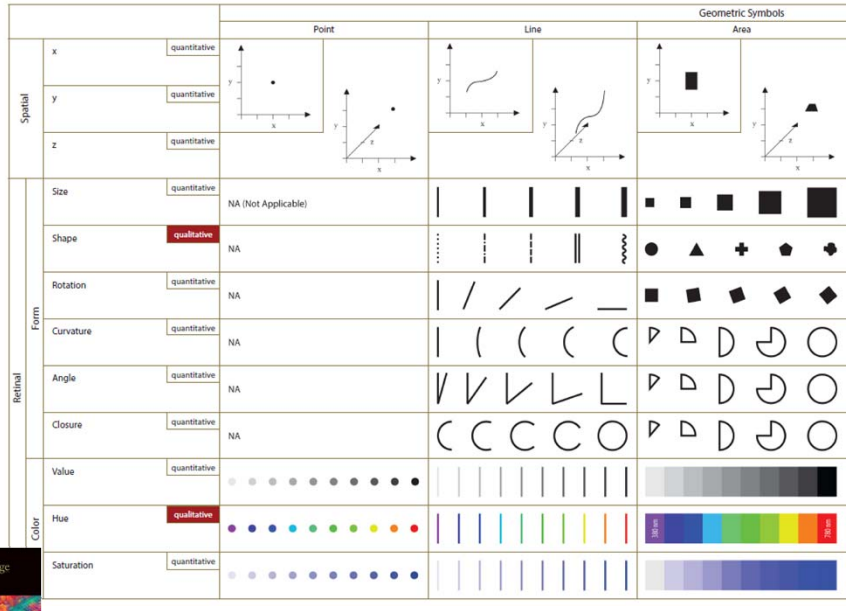
Types

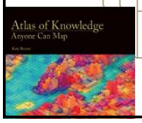
Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships	• nominal • ordinal • interval • ratio	• table • chart • graph • map • network layout	• geometric symbols point line area surface volume • linguistic symbols text numerals punctuation marks • pictorial symbols images icons statistical glyphs	• spatial position • retinal form color optics motion	• overview • zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion



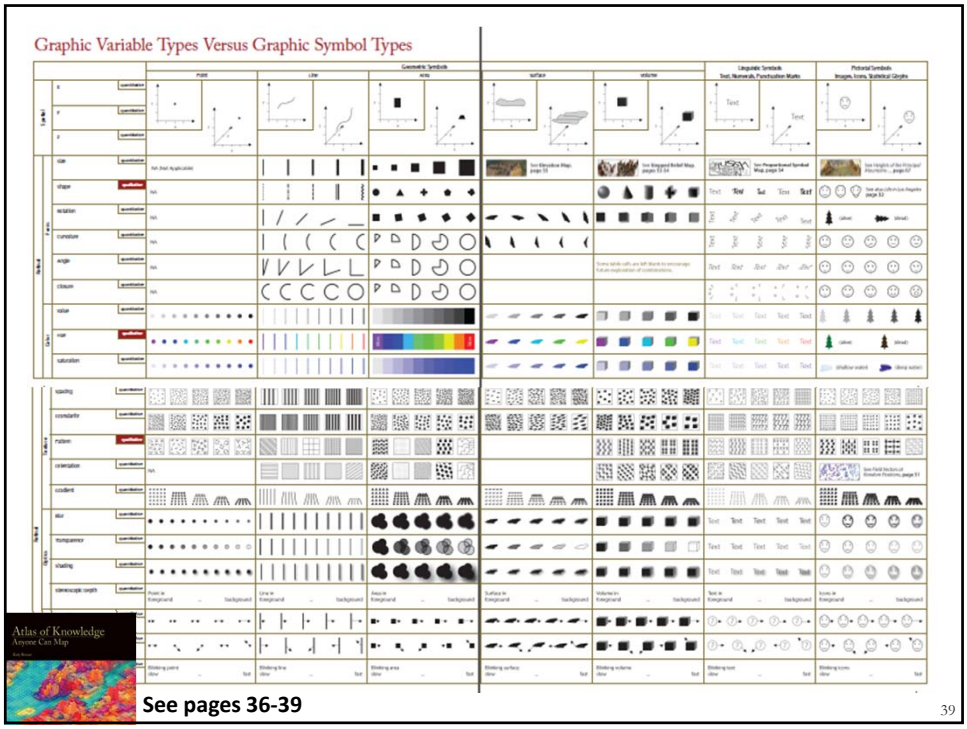
See page 25

Graphic Variable Types Versus Graphic Symbol Types





See page 36



39

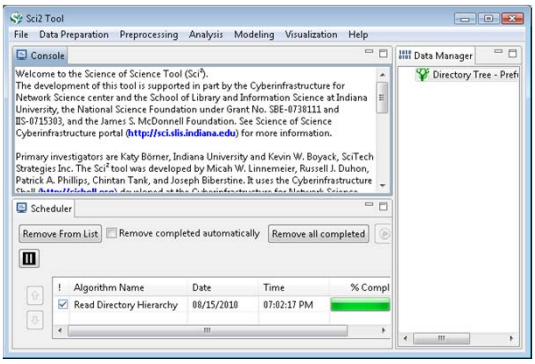
CNS Cyberinfrastructure for Network Science Center

Sci2 Tool – OSGi/CIShell-based Macroscope

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

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.

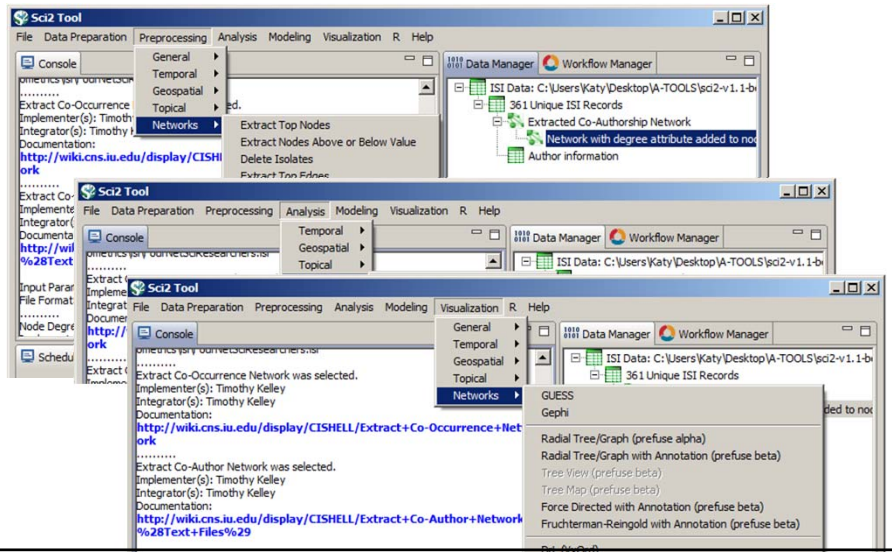
All algorithms are documented online; workflows are given in Sci2 Manual at http://sci2.wiki.cns.iu.edu

40

CNS Cyberinfrastructure for Network Science Center

Sci2 Tool Interface Components

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

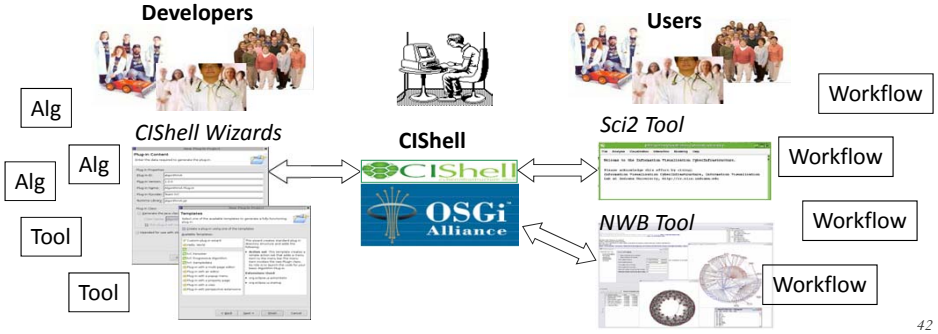


41

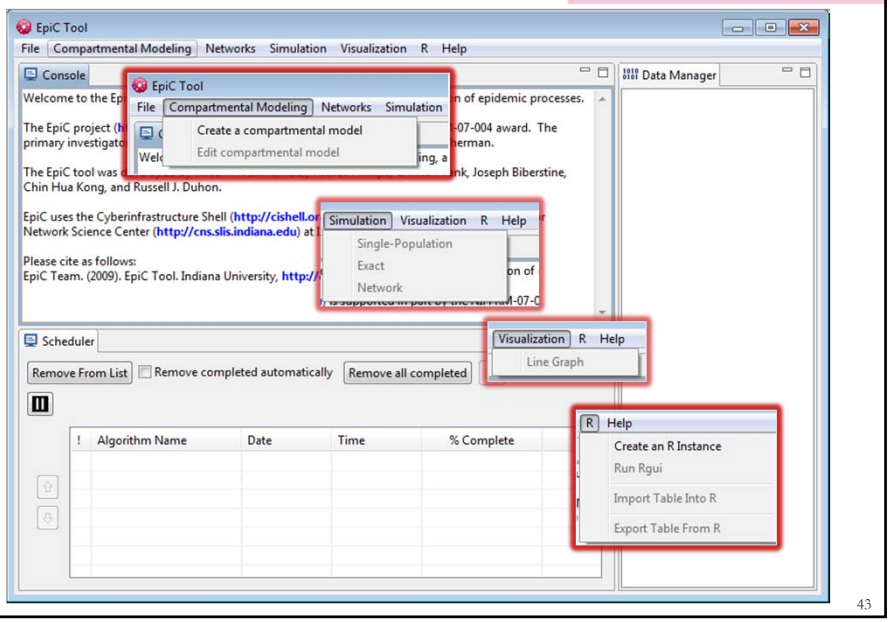
CNS Cyberinfrastructure for Network Science Center

OSGi & CIShell Support the Plug-and-Play of Algorithms

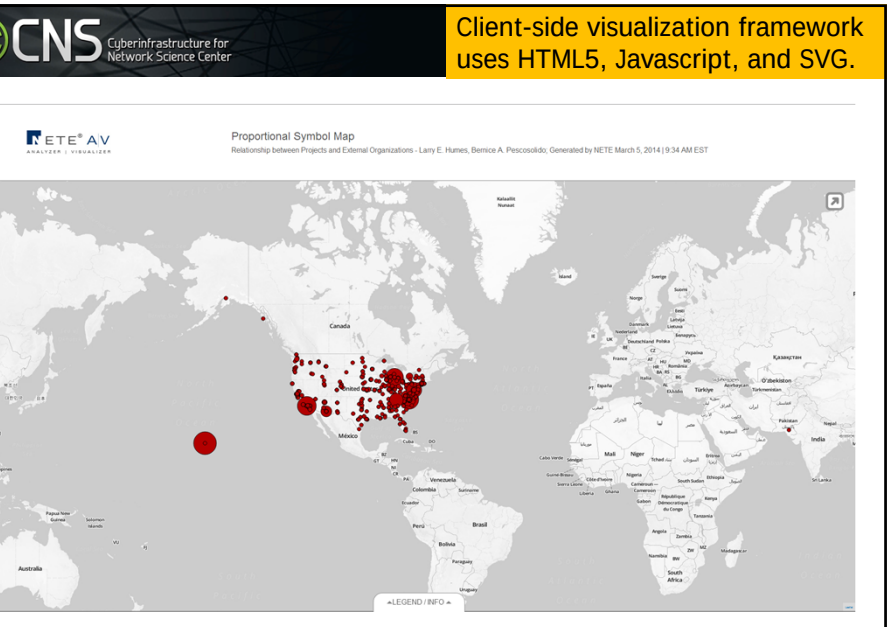
- CIShell (http://cishell.org) is an open source software specification for the integration and utilization of datasets, algorithms, and tools.
- It extends the Open Services Gateway Initiative (OSGi) (http://osgi.org), a standardized, component oriented, computing environment for networked services widely used in industry since more than 10 years.
- Specifically, CIShell provides "sockets" into which existing and new datasets, algorithms, and tools can be plugged using a wizard-driven process.



42



43



Developed for NIH by CNS and NETE. Responsive design.

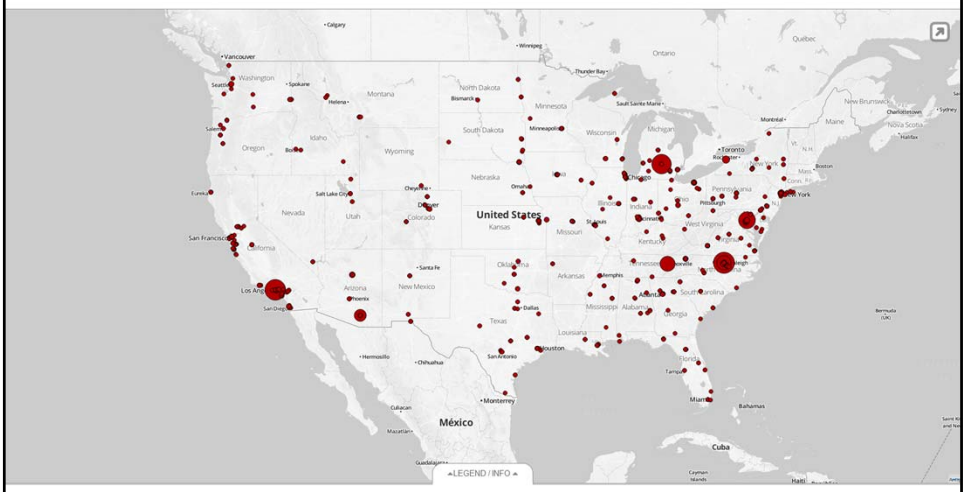
44

CNS Cyberinfrastructure for Network Science Center

NETE® AV ANALYZER | VISUALIZER

Proportional Symbol Map

Relationship between Projects and External Organizations - Larry E. Humes, Bernice A. Pescosolido; Generated by NETE March 5, 2014 | 9:34 AM EST



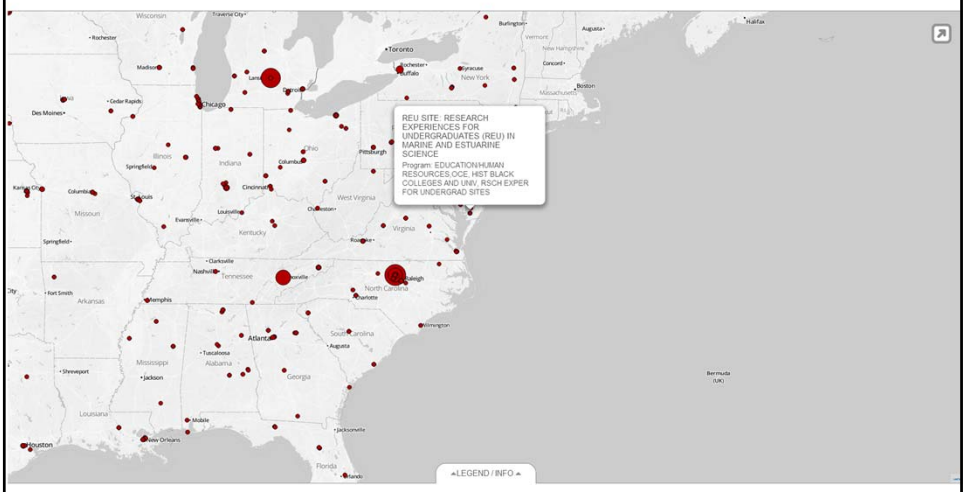
45

CNS Cyberinfrastructure for Network Science Center

NETE® AV ANALYZER | VISUALIZER

Proportional Symbol Map

Relationship between Projects and External Organizations - Larry E. Humes, Bernice A. Pescosolido; Generated by NETE March 5, 2014 | 9:34 AM EST



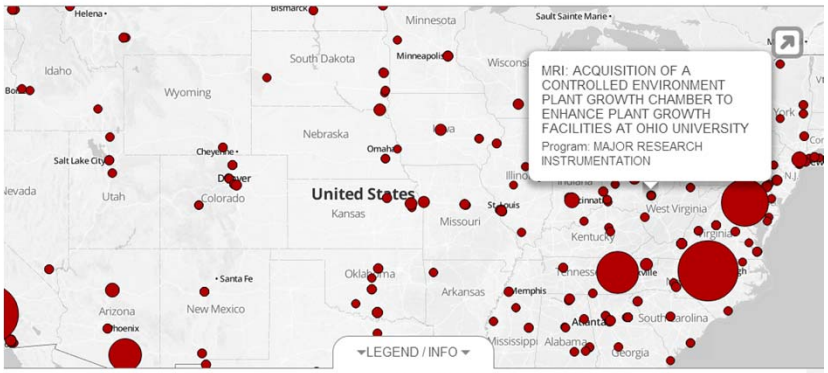
46

CNS Cyberinfrastructure for Network Science Center

NETE® AV ANALYZER | VISUALIZER

Proportional Symbol Map

Relationship between Projects and External Organizations - Larry E. Humes, Bernice A. Pescosolido; Generated by NETE March 5, 2014 | 9:34 AM EST



Amount Awarded

How To Read This Map

This proportional symbol map shows 52 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.

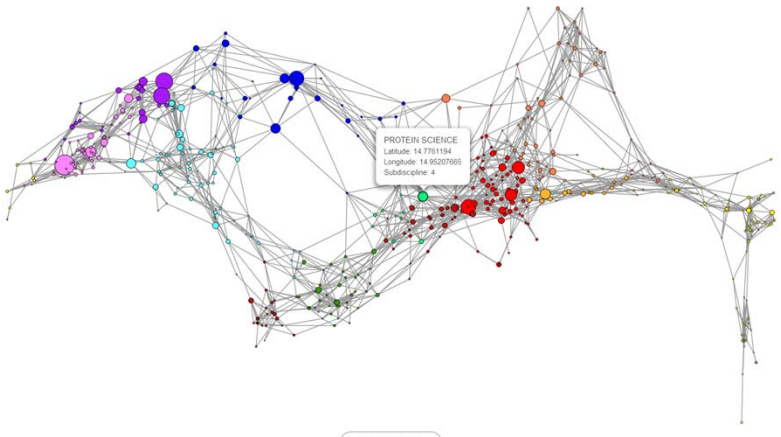
47

CNS Cyberinfrastructure for Network Science Center

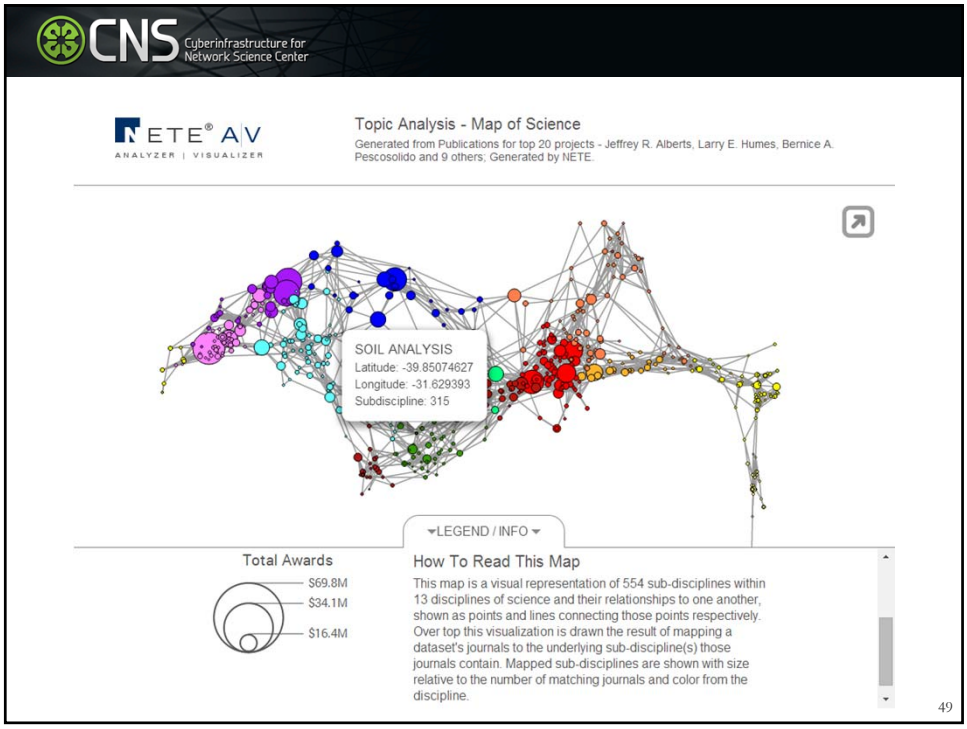
NETE® AV ANALYZER | VISUALIZER

Topic Analysis - Map of Science

Generated from Publications for top 20 projects - Jeffrey R. Alberts, Larry E. Humes, Bernice A. Pescosolido and 9 others; Generated by NETE.



48



Course Schedule

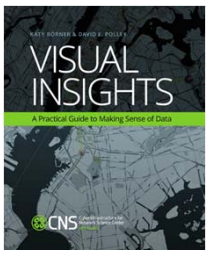
Part 1: Theory and Hands-On

- **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

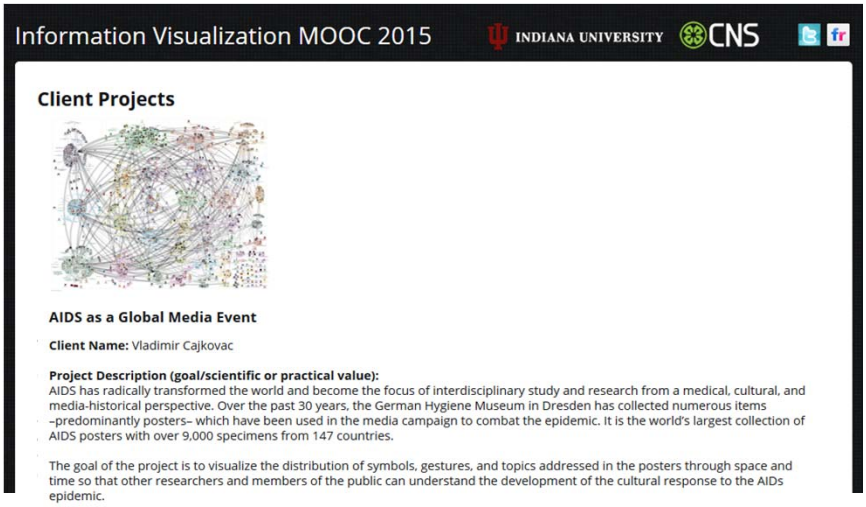


Part 2: Students work in teams on client projects.

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

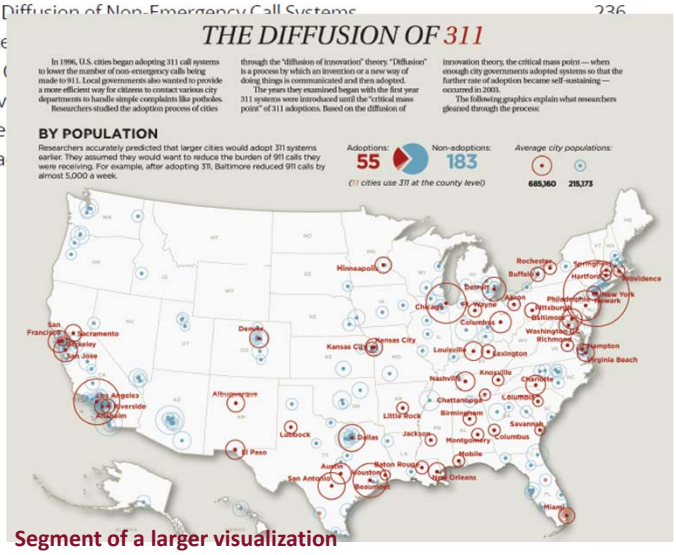
50

Last 7 Weeks: Students Work in Teams With Clients

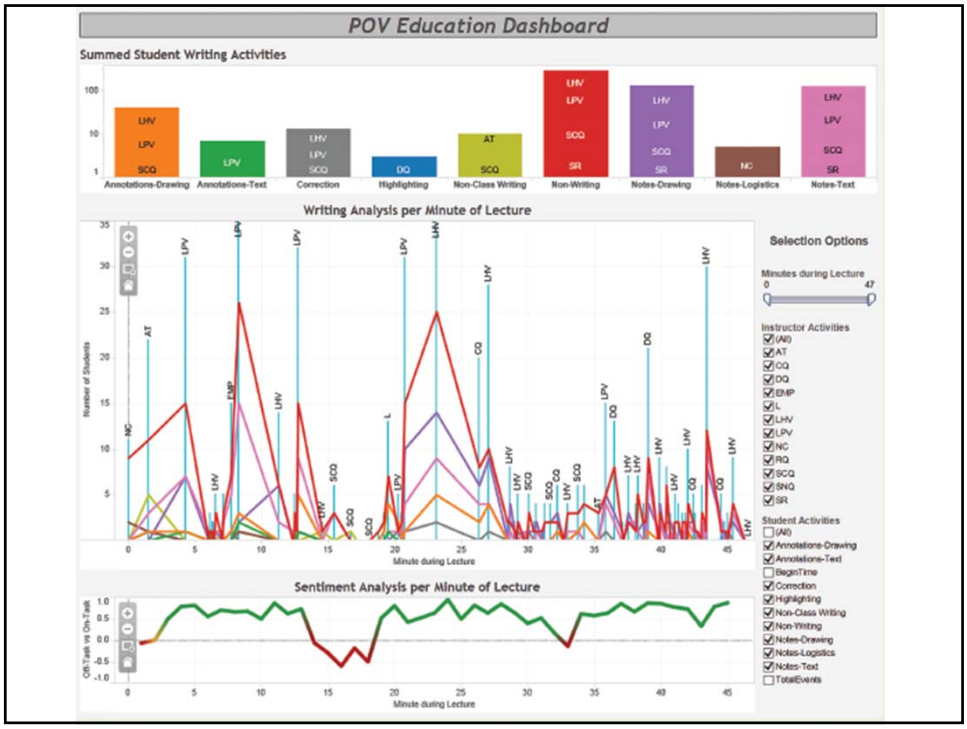


http://ivmooc.cns.iu.edu/clients.html

Results of 2013 client projects can be found in the *Visual Insights* book here.



Segment of a larger visualization



One Hundred Years of *Isis*

David E. Hubbard (Texas A&M University), Anouk Lang (University of Strathclyde), Kathleen Reed (Vancouver Island University), Anelise Hanson Shrout (Davidson College) and Lyndsay D. Troyer (Colorado State University)

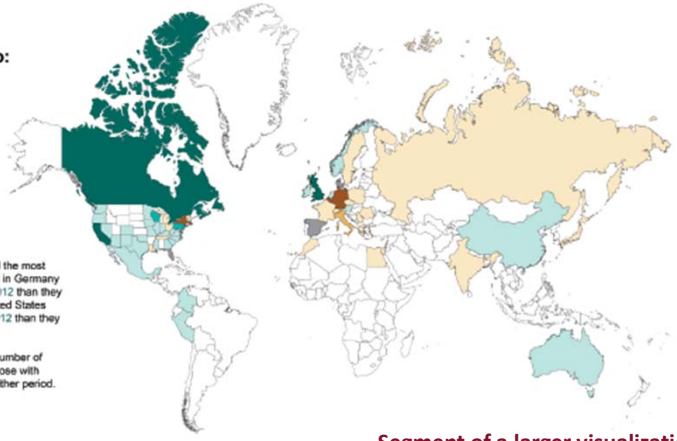
Changes in *Isis* Authorship: 1913-1937 to 1988-2012.

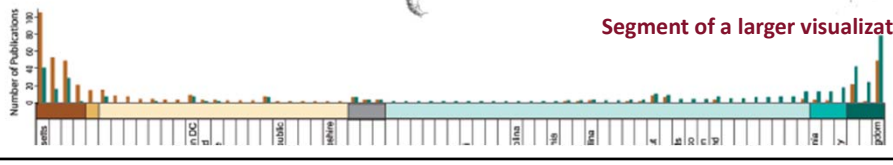
This map and chart show geographic shifts in *Isis* authors' locations across two time periods of 25 years each: 1913-1937 and 1988-2012.

Countries and states tinted brown had a greater number of articles published in the 1913-1937 period, while those tinted green had a greater number published in the 1988-2012 period.

Germany and the United States experienced the most extreme shifts in authorship. Authors located in Germany published 37 fewer articles between 1988-2012 than they did in 1913-1937. Authors located in the United States published 31 more articles between 1988-2012 than they did in 1913-1937.

Countries and states in gray had the same number of of published articles in both periods, while those with no color (i.e., white) had no publications in either period.





Segment of a larger visualization

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/

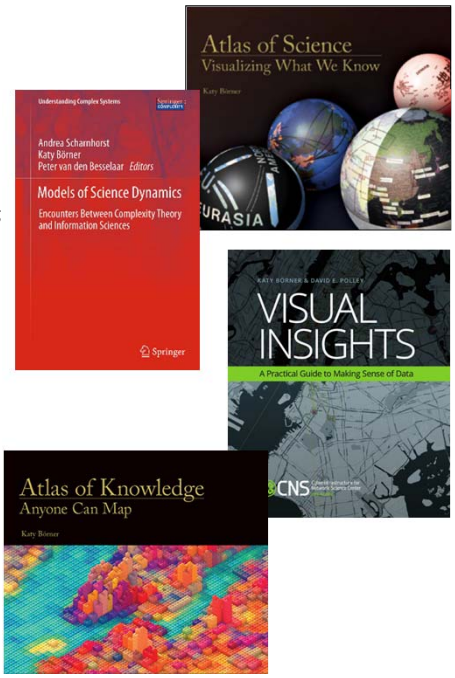
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

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



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