Data Visualization Literacy: Reading and Making Data Visualizations

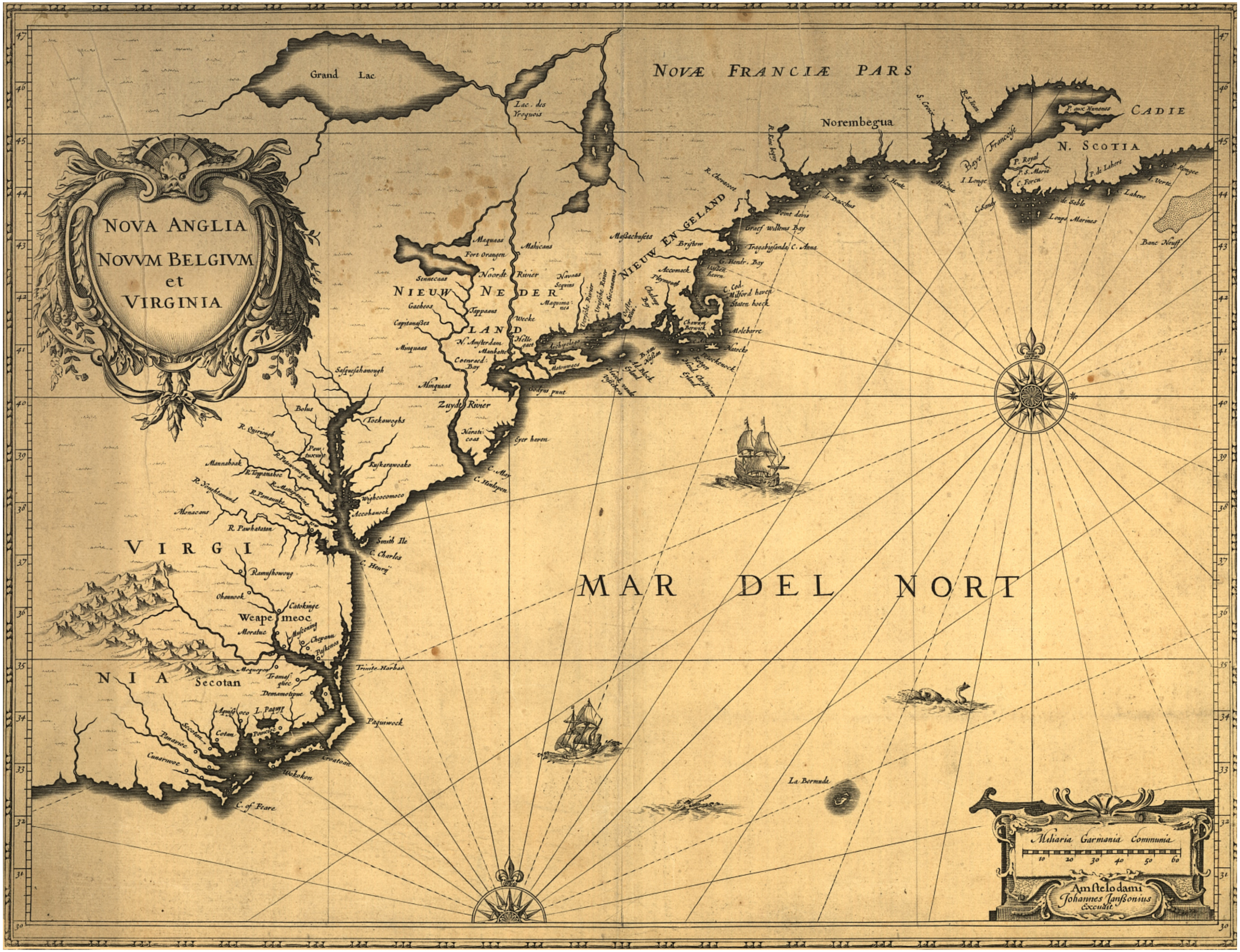
Katy Börner **@katycns**

Victor H. Yngve Distinguished Professor of
Intelligent Systems Engineering & Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing
Indiana University Network Science Institute
Indiana University, USA

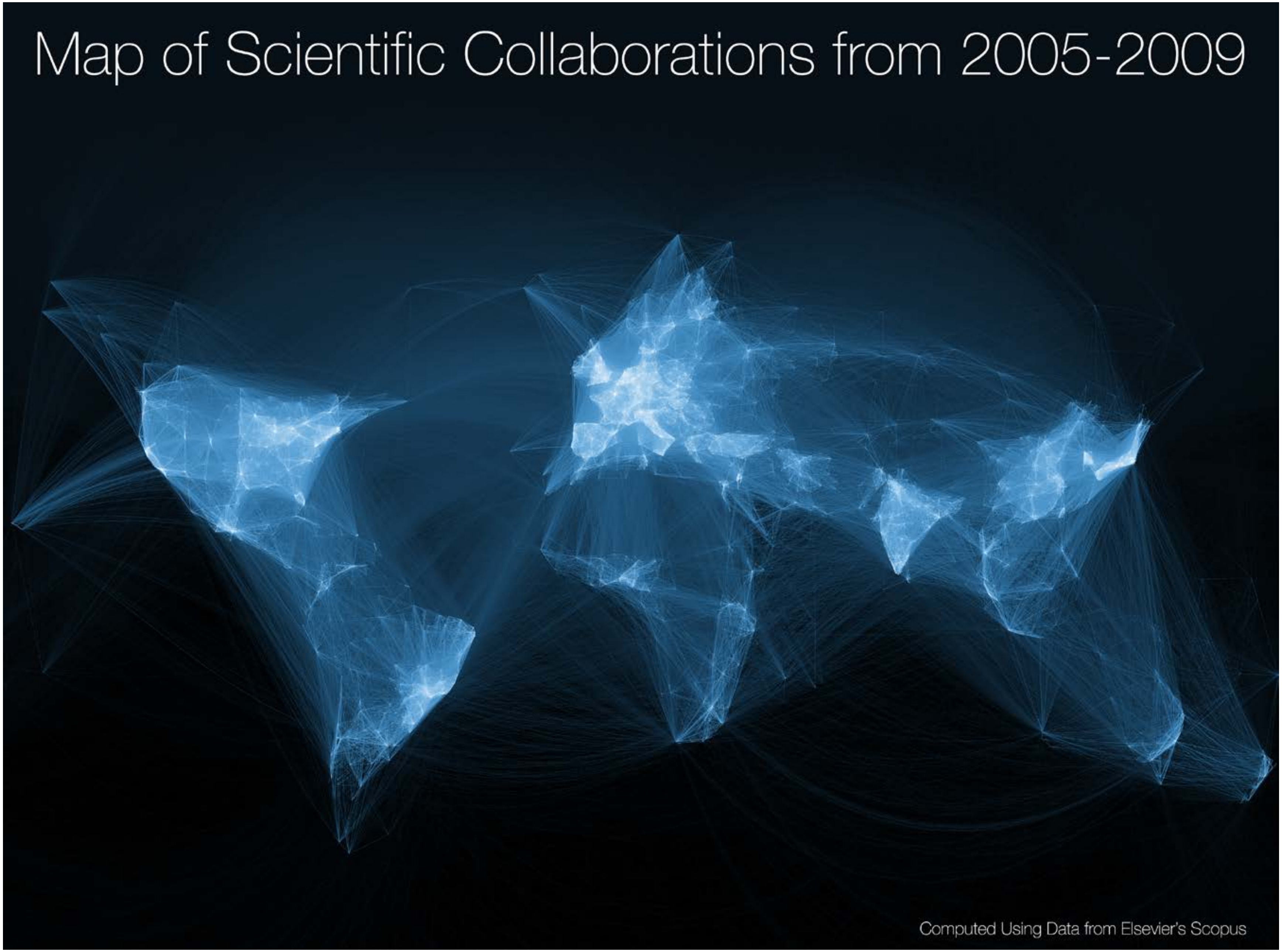
10th annual Midwest Undergraduate Cognitive Science Conference
DeVault Alumni Center, Bloomington, IN

April 7, 2018

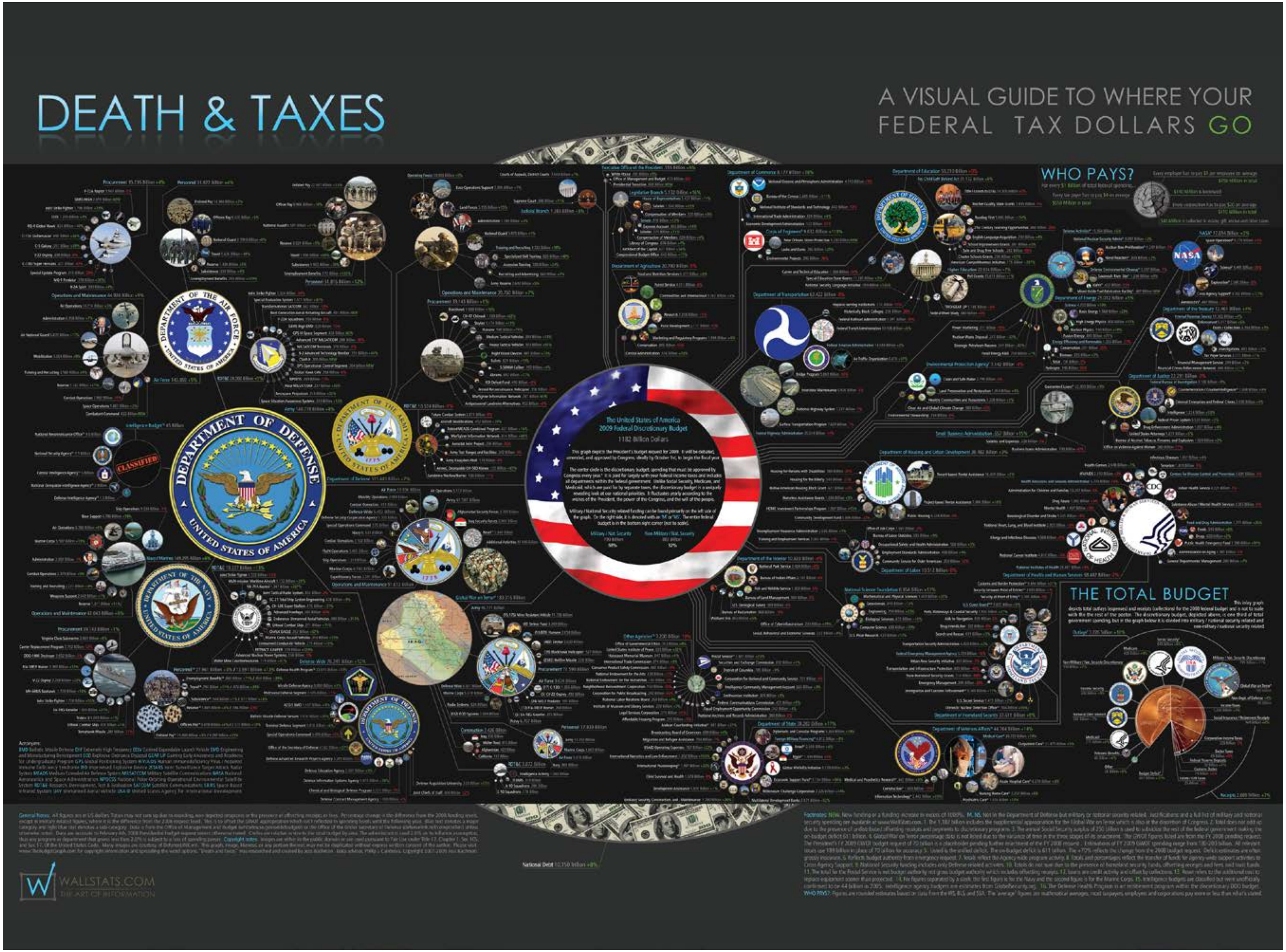
I.3 *A New Map of the Whole World with Trade Winds According to the Latest and Most Exact Observations* - Herman Moll - 1736



I.2 *Nova Anglia, Novvm Belgivm et Virginia* – Johannes Janssonius - 1642



Stream of Scientific Collaborations Between World Cities - Olivier H. Beauchesne - 2012



V.5 *Death and Taxes 2009* - Jess Bachman - 2009

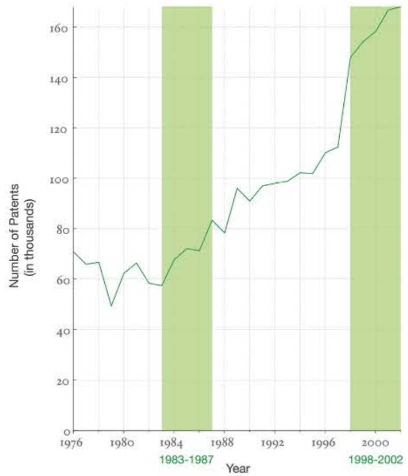
Examining the Evolution & Distribution of Patent Classifications

Managing Growing Patent Portfolios

Organizations, businesses, and individuals rely on patents to protect their intellectual property and business models. As market competition increases, patenting innovation and intellectual property rights becomes ever more important.

Managing the staggering number of patents demands new tools and methodologies. Grouping patents by their classifications offers an ideal resolution for better understanding how intellectual borders are established and change over time.

The charts below show the annual number of patents granted from January 1, 1976 to December 31, 2002 in the United States Patent and Trademark Office (USPTO) patent archive; slow and fast growing patent classes; the top 10 fast growing patent subclasses; and two evolving patent portfolios.



The Structure and Evolution of the Patent Space

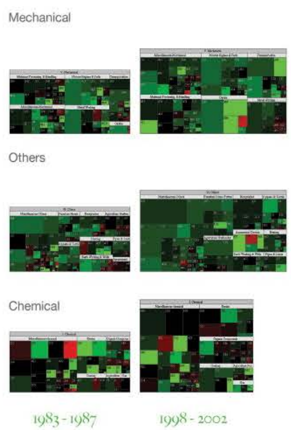
The United States Patent and Trademark Office assigns each patent to one of more than 450 classes covering broad application domains. For example, class 514 encompasses all patents dealing with 'Drug, Bio-Affecting and Body Treating Compositions.' Classes are further broken down by subclasses that have hierarchical associations. As one example, class 455 features subclass 99 entitled "with vehicle."

The top 10 fast growing patent classes for 1998–2002 are listed together with the number of patents granted. Most come from the 'Computer and Communications' and the 'Drugs and Medical' area.

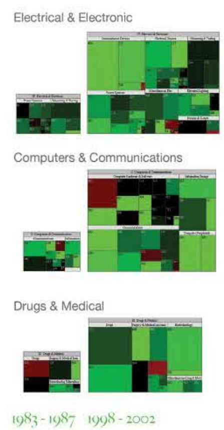
The evolving hierarchical structure of patent classes and their sizes is represented using treemaps, a space-filling visualization technique developed by Ben Shneiderman at the University of Maryland. A treemap presents a hierarchy as a collection of nested rectangles—demarcating a parent-child relationship between nodes by nesting the child within the parent rectangle. The size and color of each rectangle represent certain attributes of the nodes.

Here, each rectangle represents a class and the area size denotes the total number of patents in that class. The rectangle's color corresponds to percentage increase (**green**) or decrease (**red**) in the number of patents granted in that class from the previous interval.

Slow Growing Classes



Fast Growing Classes



Top-10 Subclasses

Class	Title	# of Patents
514	Drug, Bio-Affecting and Body Treating Compositions	18,778
438	Semiconductor Device Manufacturing:Process	17,775
435	Chemistry: Molecular Biology and Microbiology	17,474
424	Drug, Bio-Affecting and Body Treating Compositions	13,637
428	Stock Material or Miscellaneous Articles	13,314
257	Active Solid-State Devices (e.g., Transistors, Solid-State Diodes)	12,924
395	Information Processing System Organization	9,955
345	Computer Graphics Processing, Operator Interface Processing, and Selective Visual Display Systems	9,510
359	Optical: Systems and Elements	9,151
365	Static Information Storage and Retrieval	8,392
	Total	130,910

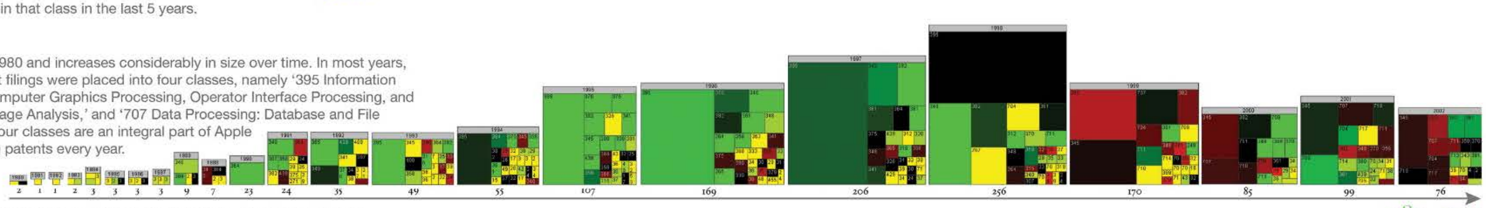
1998 - 2002

Patent Portfolio Analysis

A longitudinal analysis of portfolios reveals different patenting strategies. For each year (given in gray above each treemap), a treemap of all new patents granted to the assignee is shown. The number of patents is given below each treemap. The same size and color coding as above was used. In addition, yellow indicates that no patent has been granted in that class in the last 5 years.

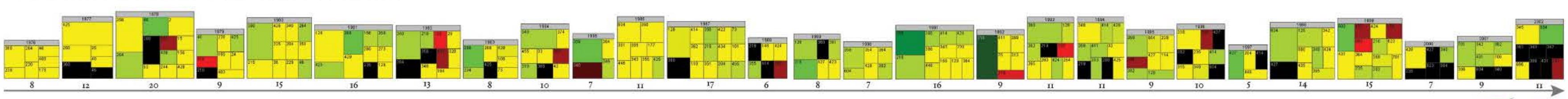
Apple Computer, Inc.

Apple Computer, Inc.'s portfolio starts in 1980 and increases considerably in size over time. In most years, more than half of Apple Computer's patent filings were placed into four classes, namely '395 Information Processing System Organization,' '345 Computer Graphics Processing, Operator Interface Processing, and Selective Visual Display Systems,' '382 Image Analysis,' and '707 Data Processing: Database and File Management or Data Structures.' These four classes are an integral part of Apple Computer, Inc.'s patent portfolio, receiving patents every year.

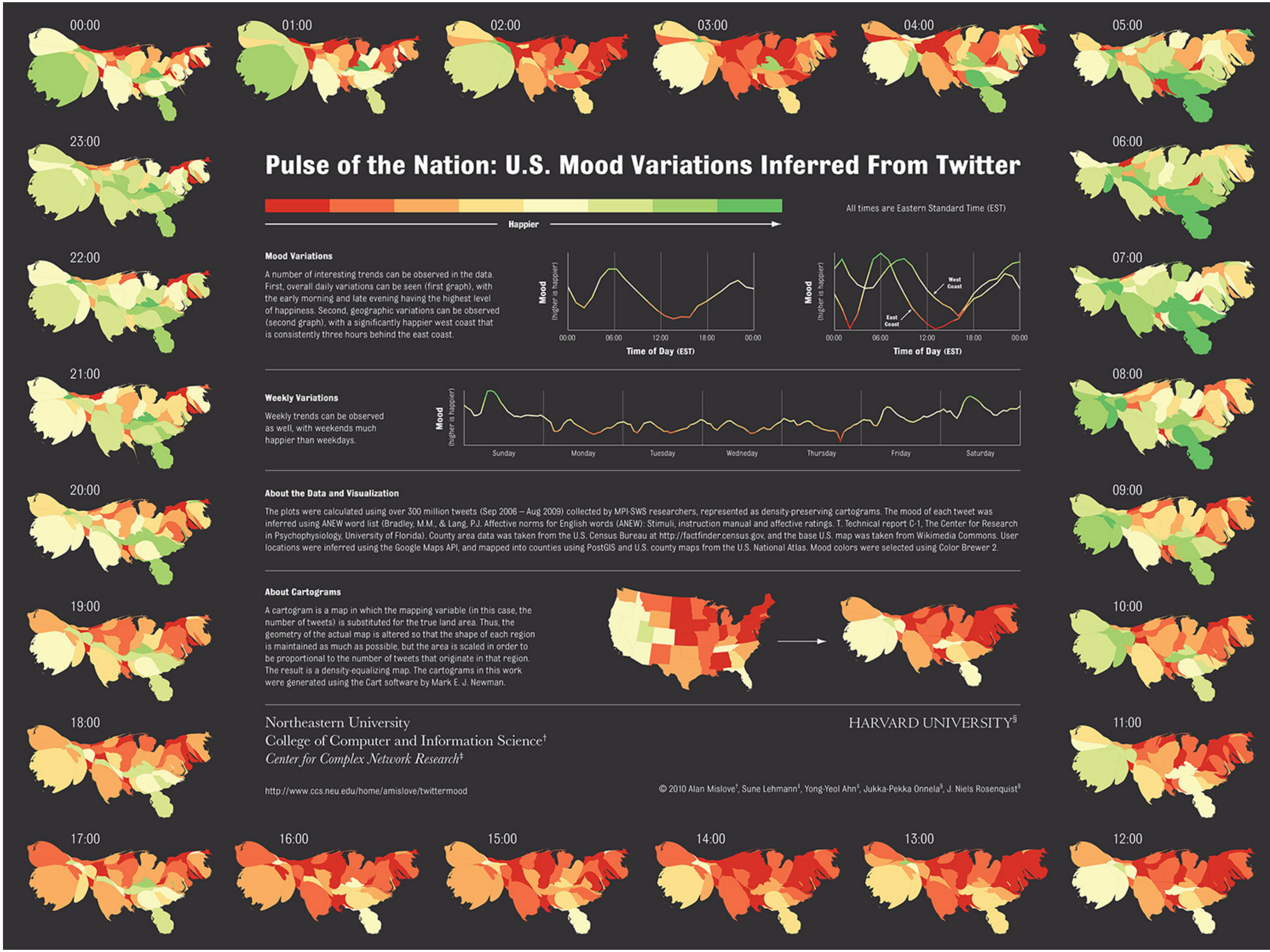


Jerome Lemelson

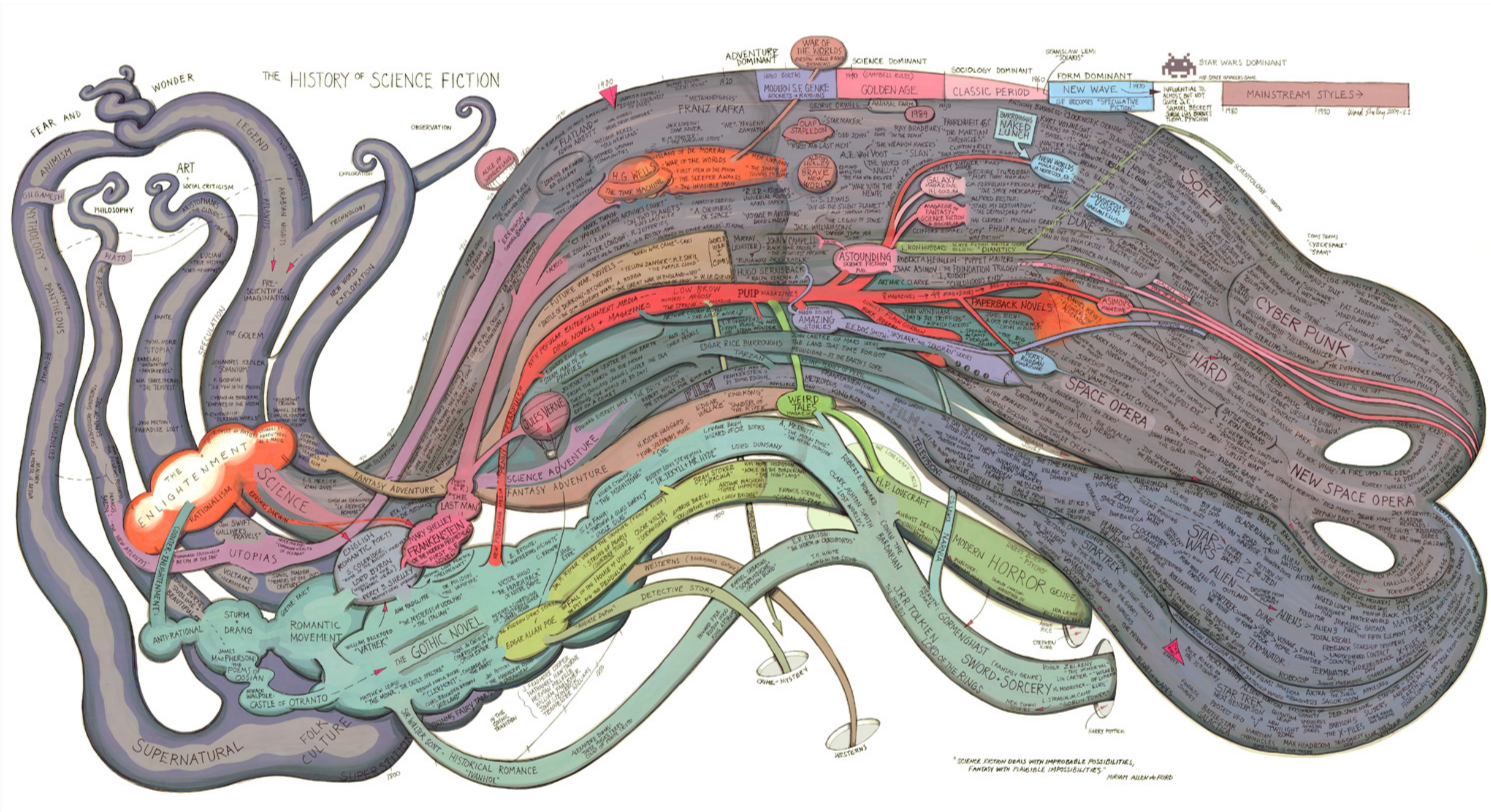
The patent portfolio of Jerome Lemelson shows a very different activity pattern. Starting in 1976, he publishes between 6–20 patents each year. However, the predominance of yellow shows that there is little continuity from previous years in regards to the classes into which patents are filed. No class dominates. Instead, more and more new intellectual space is claimed.



IV.5 *Examining the Evolution & Distribution of Patent Classifications* - Daniel O. Kutz, Katy Borner, and Elisha F. Hardy - 2004

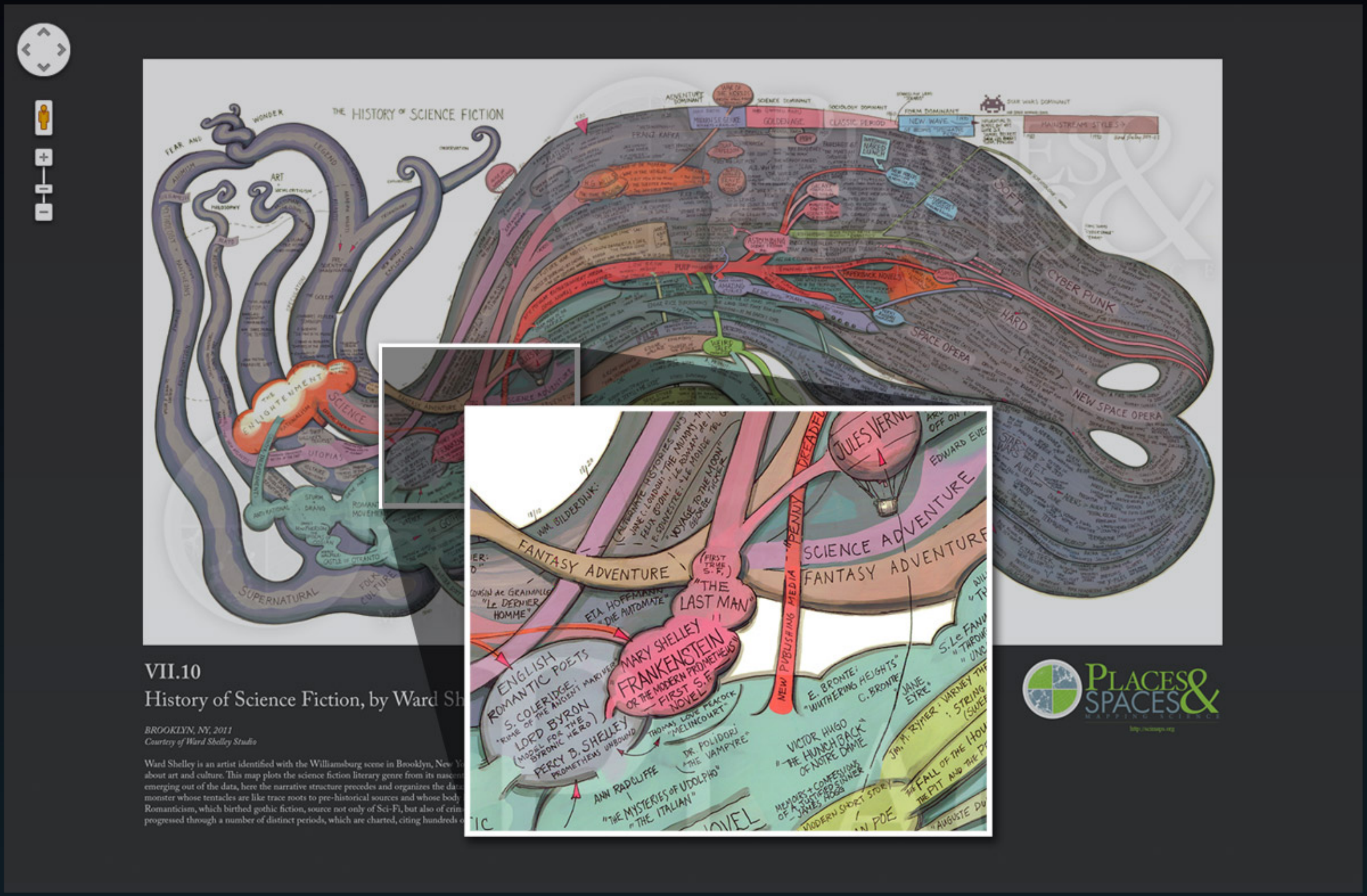


IX.4 *Pulse of the Nation* - Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and James Niels Rosenquist - 2010

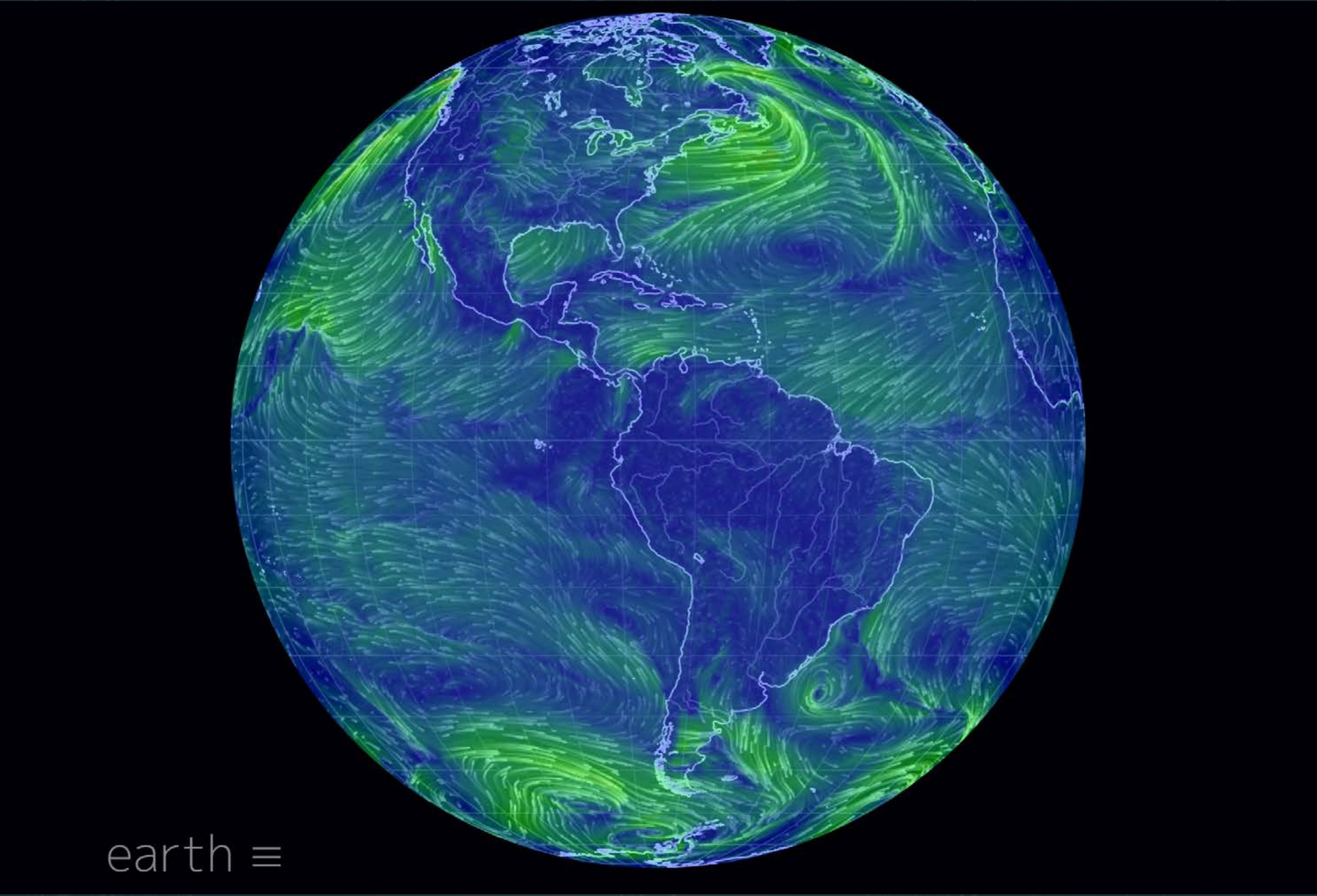


VII.10 *History of Science Fiction* - Ward Shelley - 2011

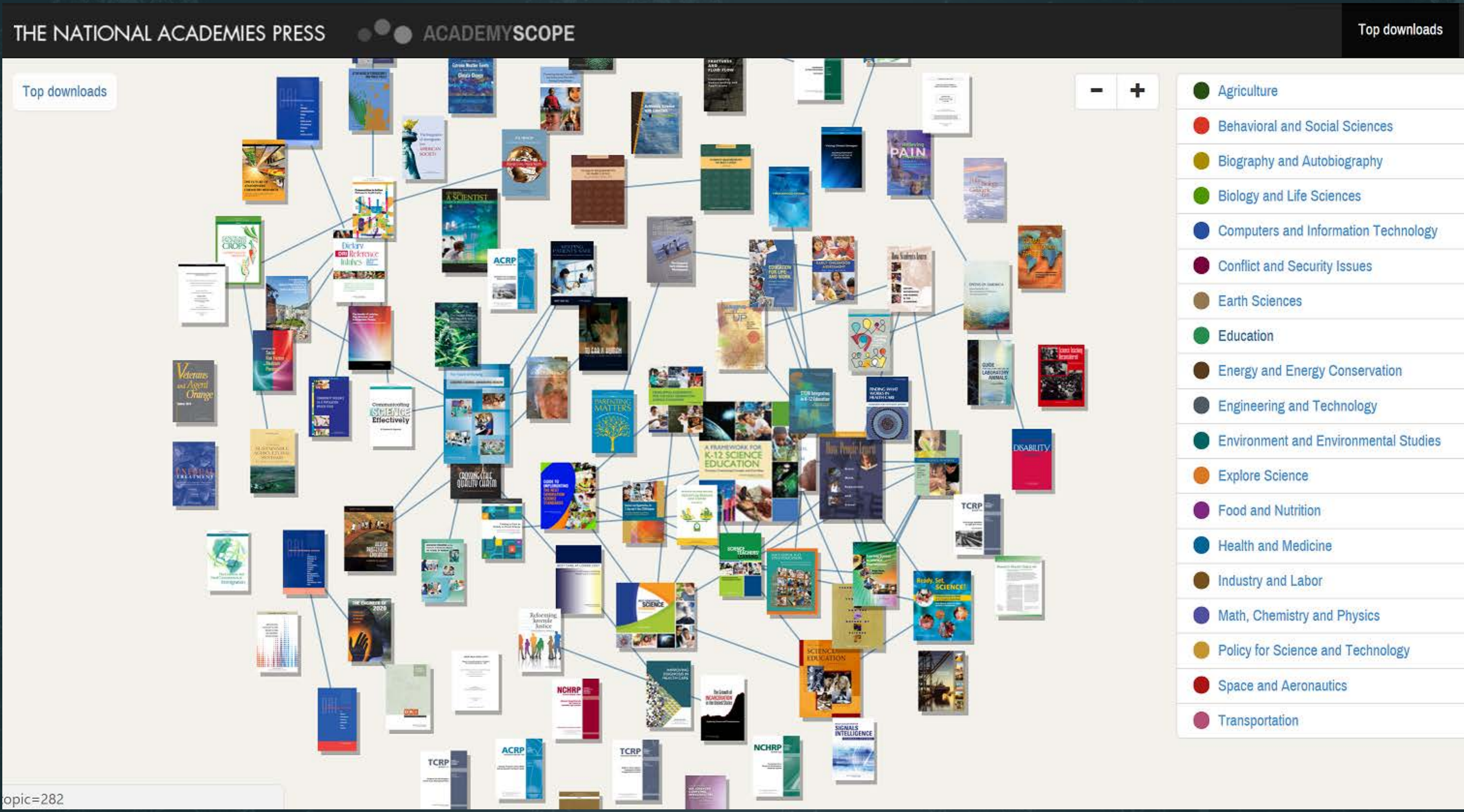
Check out our Zoom Maps online!



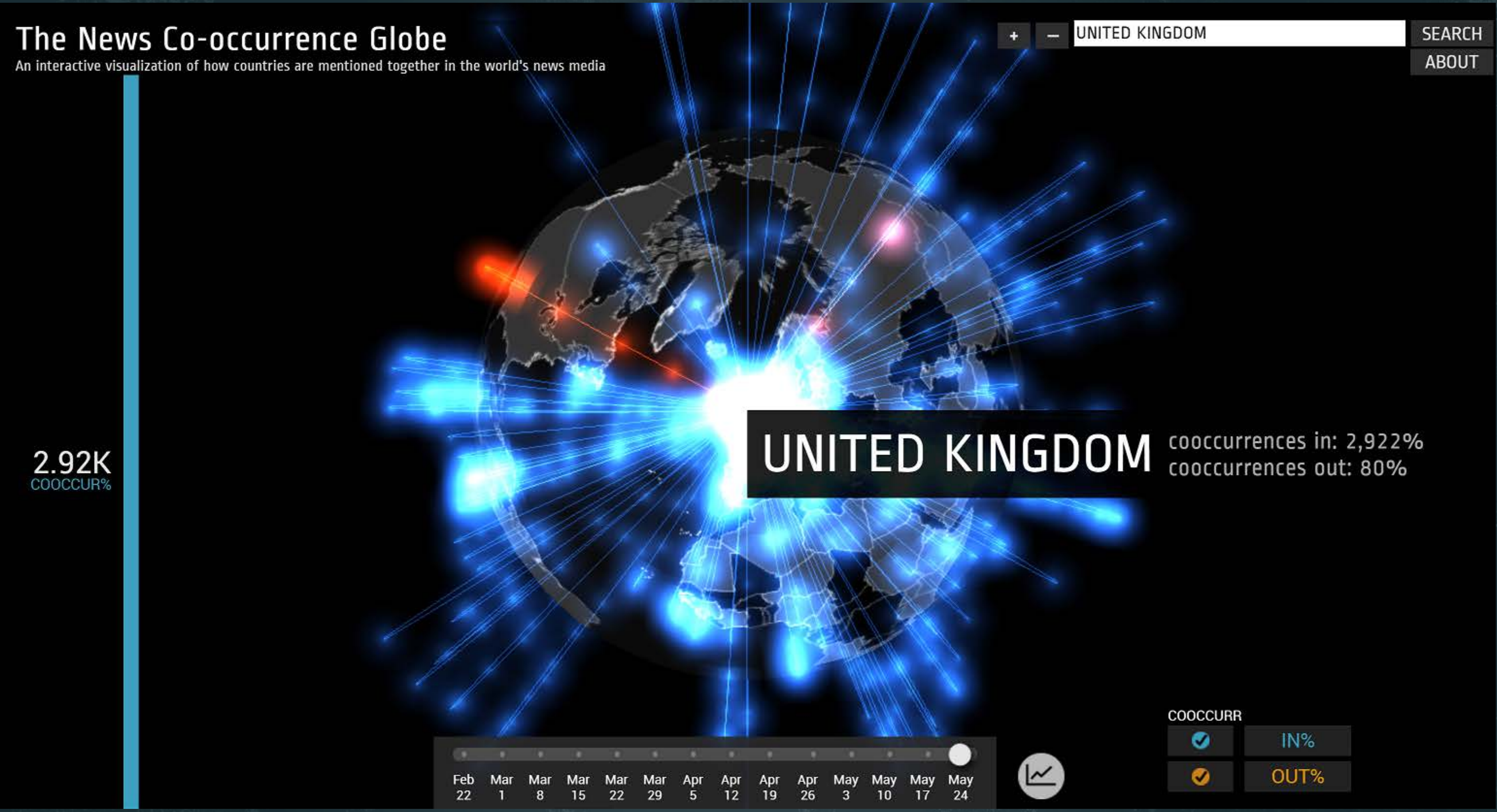
Visit scimaps.org and check out all our maps in stunning detail!



Earth – Cameron Beccario

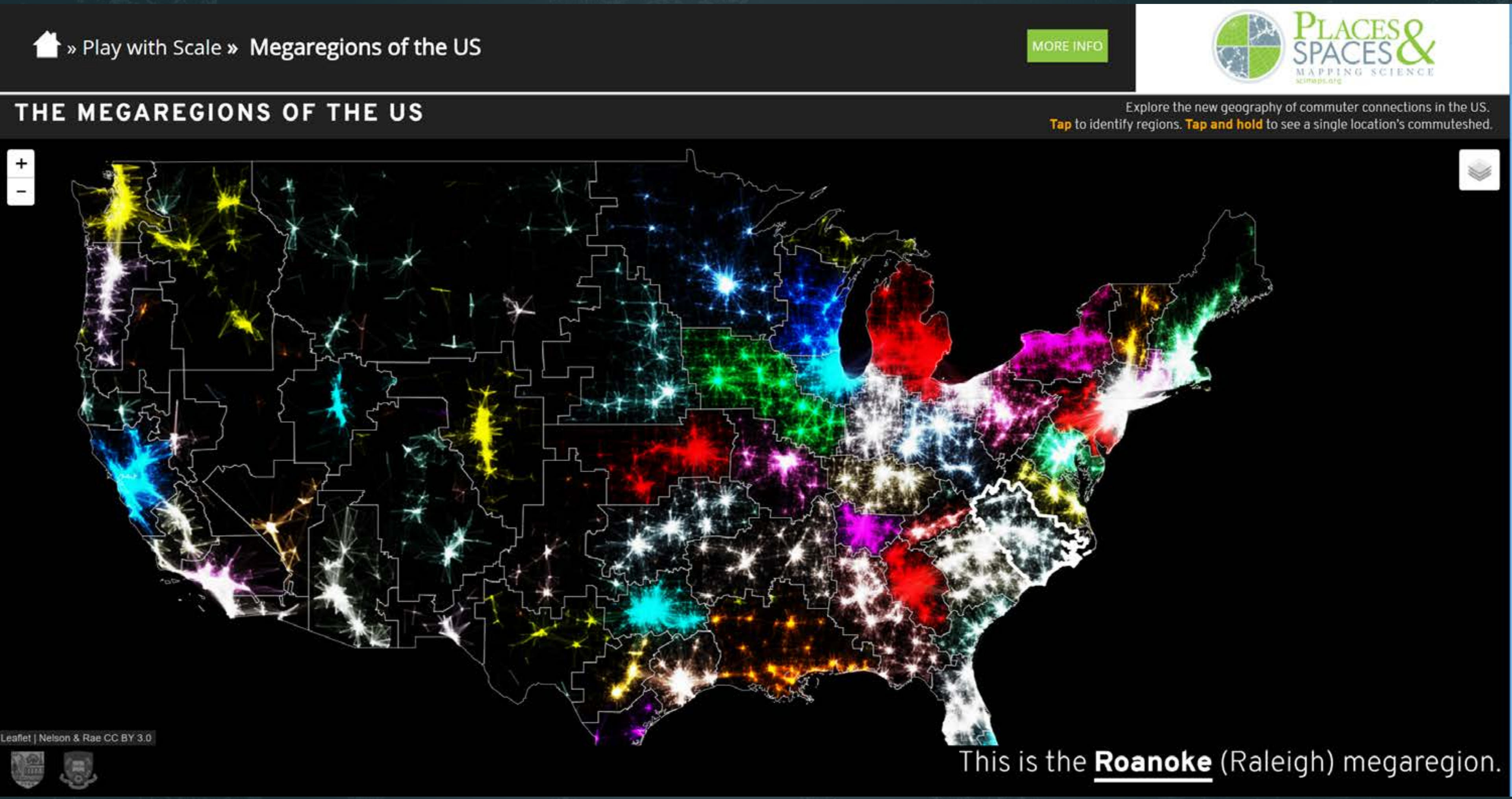


AcademyScope – National Academy of the Sciences & CNS



Mapping Global Society –Kalev Leetaru

Smelly Maps – Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello – 2015



Megaregions of the US –Garrett Dash Nelson and Alasdair Rae – 2016

Maps of Science & Technology

http://scimaps.org

101st Annual Meeting of the Association of American Geographers, Denver, CO.
April 5th - 9th, 2005 (First showing of Places & Spaces)

University of Miami, Miami, FL.
September 4 - December 11, 2014.

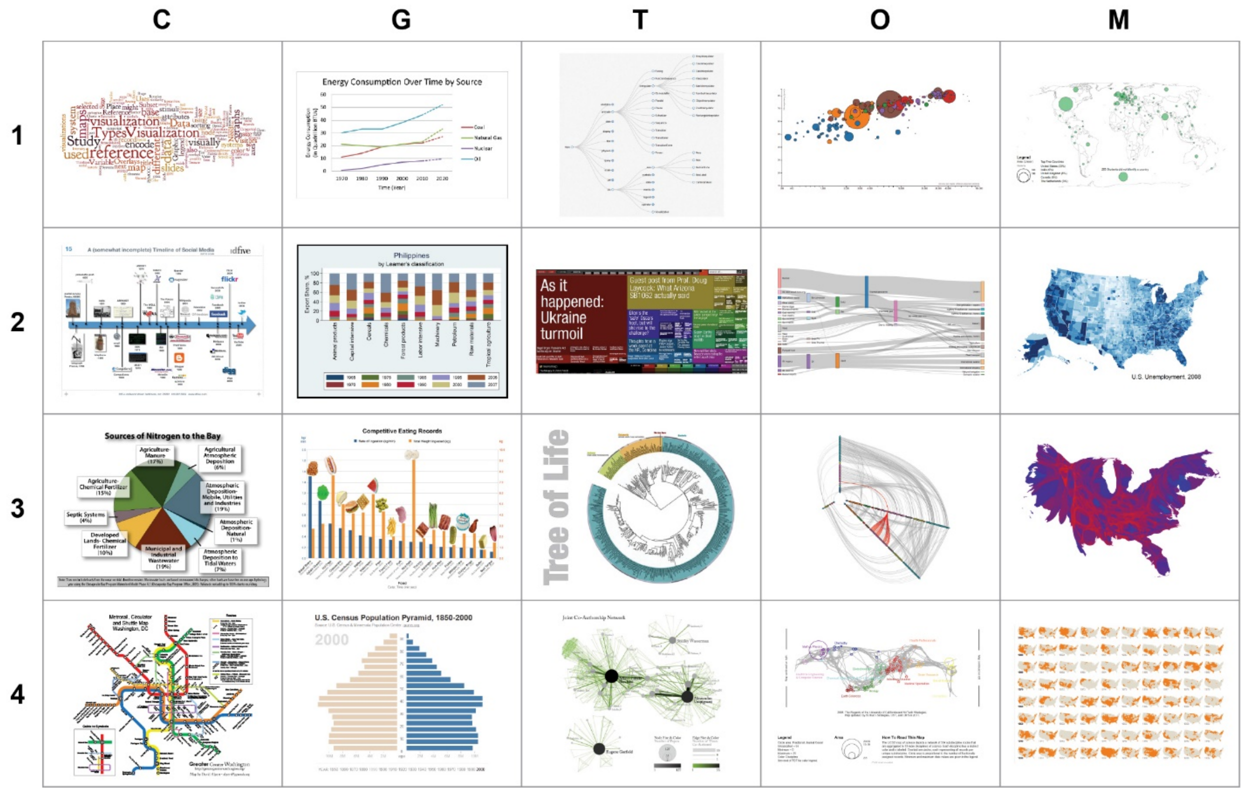
Duke University, Durham, NC.
January 12 - April 10, 2015

The David J. Sencer CDC Museum, Atlanta, GA.
January 25 - June 17, 2016.

100 maps and 12 macroscopes by 215 experts on display at 354 venues in 28 countries.

Problem: Data Visualization Literacy is Low

Most science museum visitors in the US cannot name, read, or interpret common data visualizations.



Börner, Katy, Joe E. Heimlich, Russell Balliet, and Adam V. Maltese. 2015. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization 1-16.* [*http://cns.iu.edu/docs/publications/2015-borner-investigating.pdf*](http://cns.iu.edu/docs/publications/2015-borner-investigating.pdf)

Data Visualization Literacy:
Research and Tools that Advance Public Understanding of Scientific Data

10TH ANNUAL
MUCSC
MIDWEST UNDERGRADUATE
COGNITIVE SCIENCE CONFERENCE

Data Visualization Literacy

Data visualization literacy (ability to read, make, and explain data visualizations) requires

- *literacy* (ability to read and write text, e.g., in titles, axis labels, legend),
- *visual literacy* (ability to find, interpret, evaluate, use, and create images and visual media), and
- *data literacy* (ability to read, create, and communicate data).

Being able to “read and write” data visualizations is becoming as important as being able to read and write text. Understanding, measuring, and improving data and visualization literacy is important for understanding STEAM developments and to strategically approach global issues.

How to Classify (Name & Make) Different Visualizations?

By

- User insight needs?
- User task types?

- Data to be visualized?
- Data transformation?

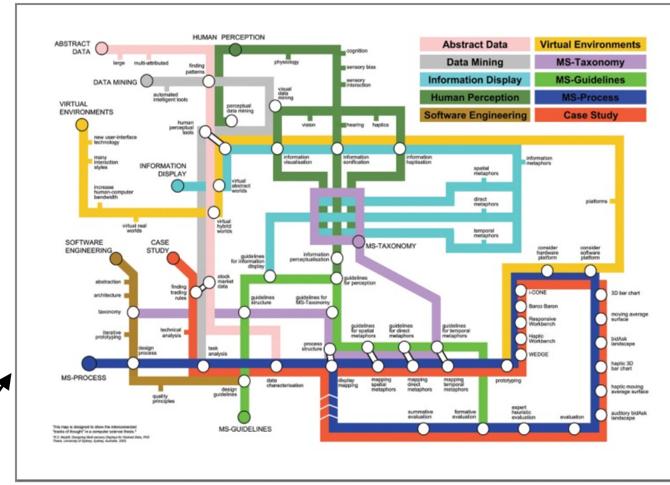
- Visualization technique?
- Visual mapping transformation?
- Interaction techniques?

- Or ?

Different Question Types

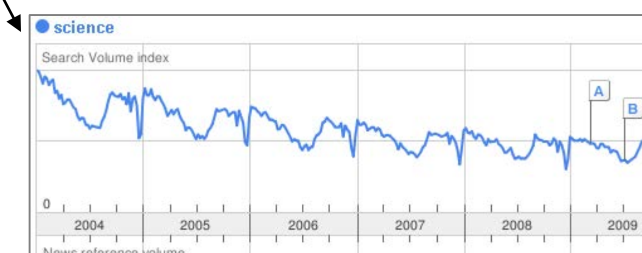
Terabytes of data

Descriptive & Predictive Models



Find your way

Find collaborators, friends

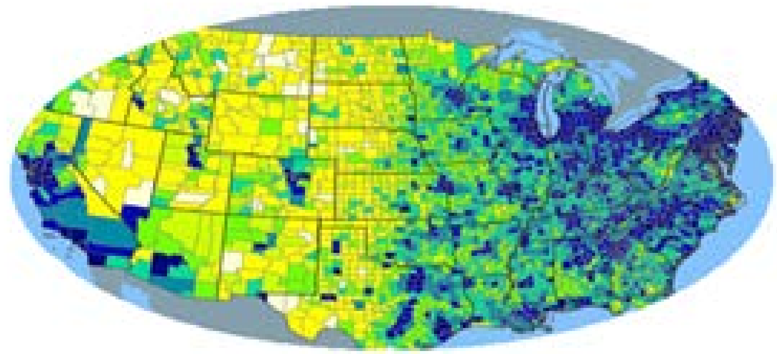


Identify trends

Different Levels of Abstraction/Analysis

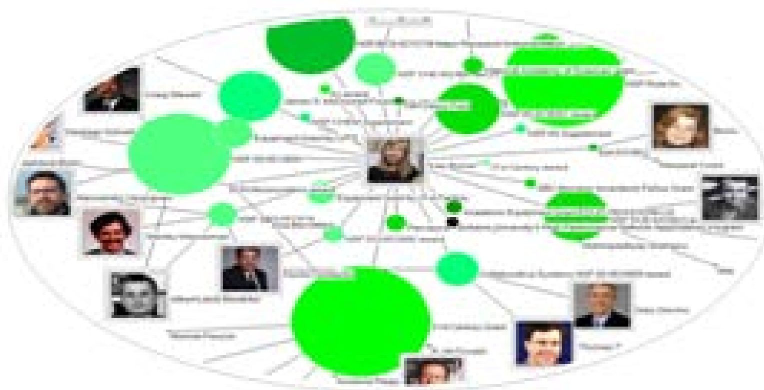
Macro/Global

Population Level



Meso/Local

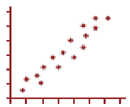
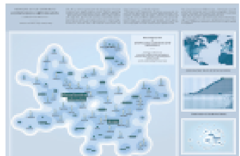
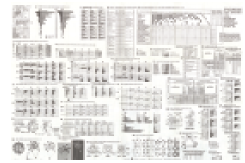
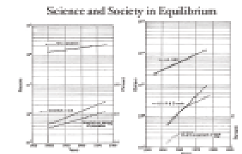
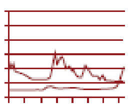
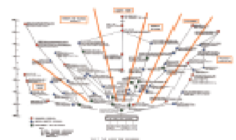
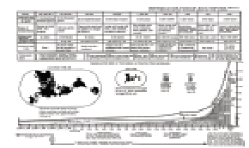
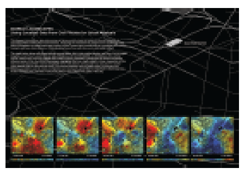
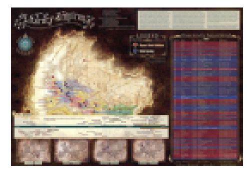
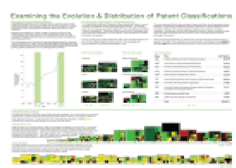
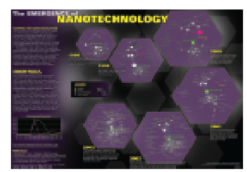
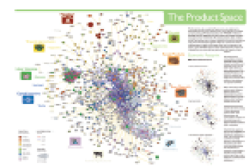
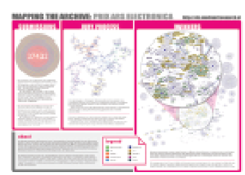
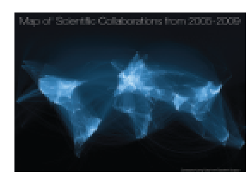
Group Level

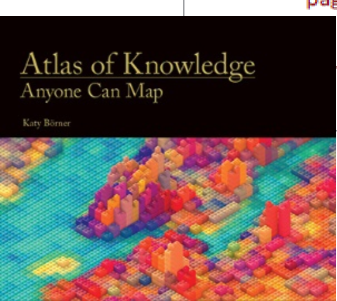


Micro

Individual Level

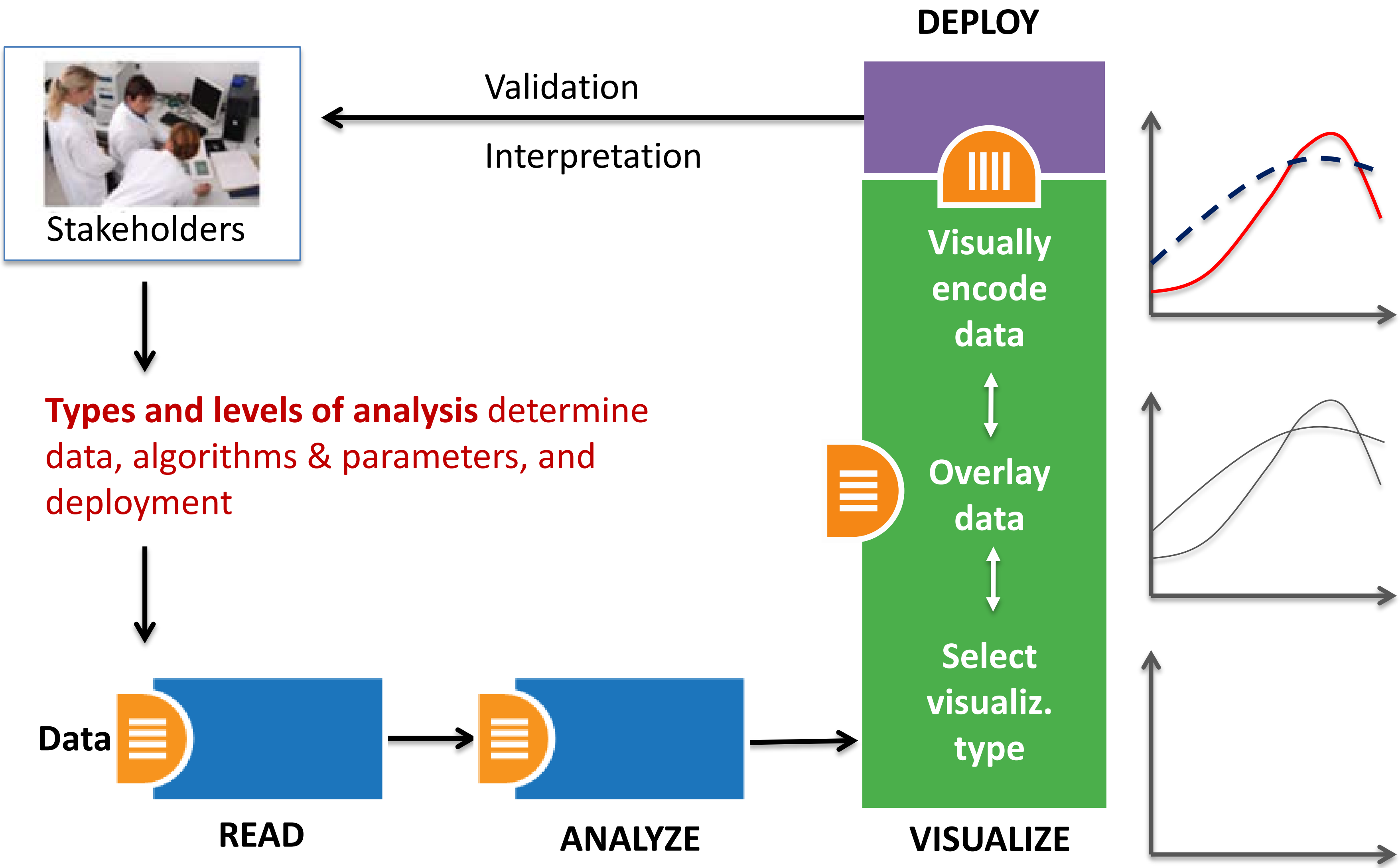
Tasks

TYPES \ LEVELS	MICRO: Individual Level about 1–1,000 records page 6		MESO: Local Level about 1,001–100,000 records page 8		MACRO: Global Level more than 100,000 records page 10	
Statistical Analysis page 44 		Knowledge Cartography page 135		Productivity of Russian life sciences research teams page 105		Number of scientists versus population and R&D costs versus GNP. page 103
WHEN: Temporal Analysis page 48 		Visualizing decision-making processes page 95		Key events in the development of the video tape recorder page 85		Increased travel and communication speeds page 83
WHERE: Geospatial Analysis page 52 		Cell phone usage in Milan, Italy page 109		Victorian poetry in Europe page 137		Ecological footprint of countries page 99
WHAT: Topical Analysis page 56 		Evolving patent holdings of Apple Computer, Inc. and Jerome Lemelson page 89		Evolving journal networks in nanotechnology page 139		Product space showing co-export patterns of countries page 93
WITH WHOM: Network Analysis page 60 		World Finance Corporation network page 87		Electronic and new media art networks page 133		World-wide scholarly collaboration networks page 157

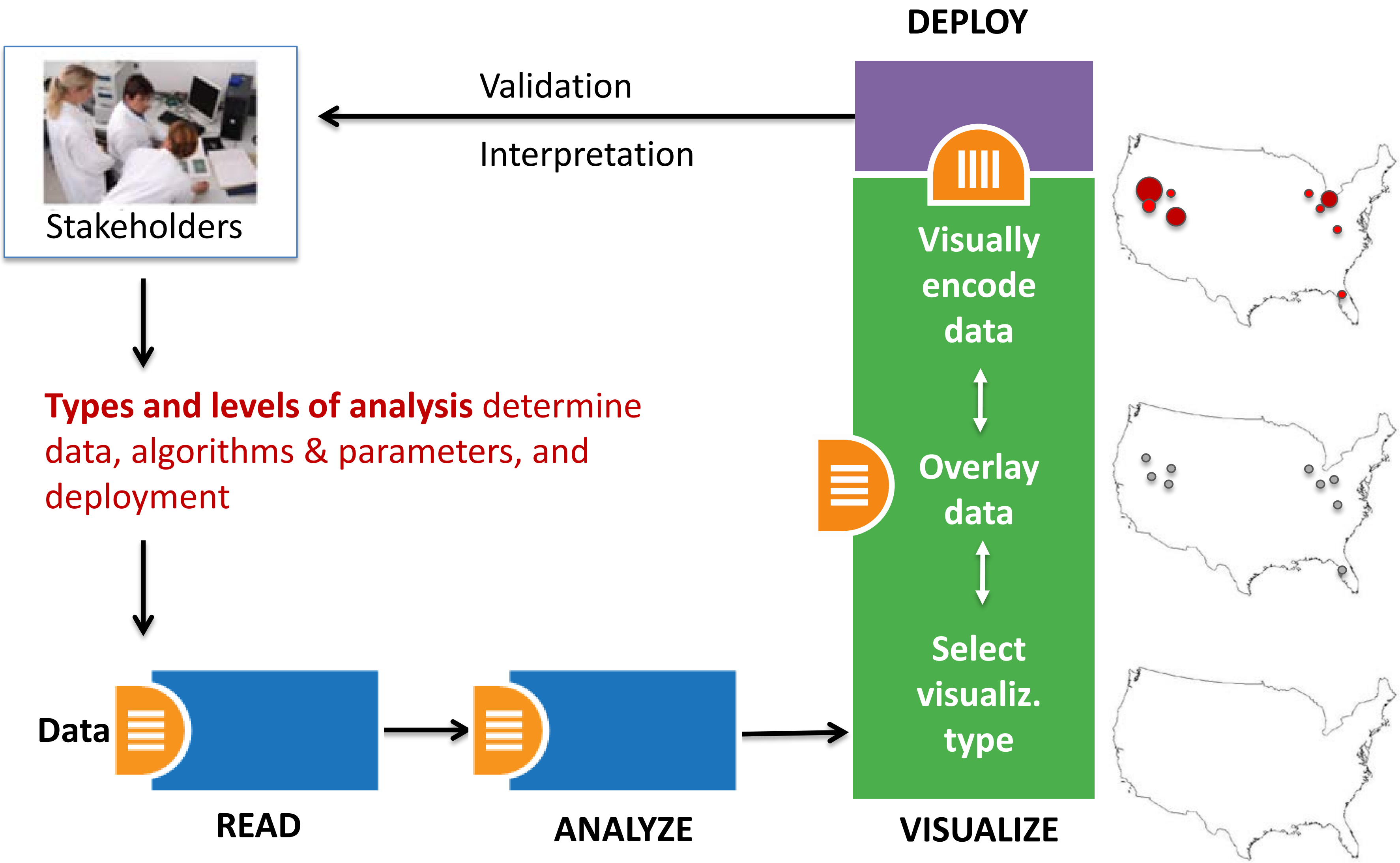


See *Atlas of Science: Anyone Can Map*, page 5

Needs-Driven Workflow Design

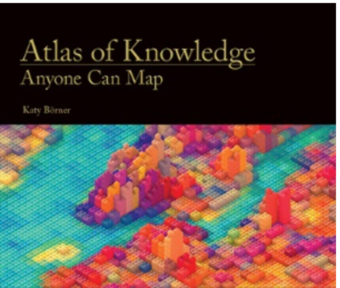


Needs-Driven Workflow Design



Visualization Framework

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	• nominal	• table	• geometric symbols	• spatial	• overview
• order/rank/sort	• ordinal	• chart	point	position	• zoom
• distributions	• interval	• graph	line	• retinal	• search and locate
(also outliers, gaps)	• ratio	• map	area	form	• filter
• comparisons		• network layout	surface	color	• details-on-demand
• trends			volume	optics	• history
(process and time)			• linguistic symbols	motion	• extract
• geospatial			text		• link and brush
• compositions			numerals		• projection
(also of text)			punctuation marks		• distortion
• correlations/relationships			• pictorial symbols		
			images		
			icons		
			statistical glyphs		



See *Atlas of Science: Anyone Can Map*, page 24

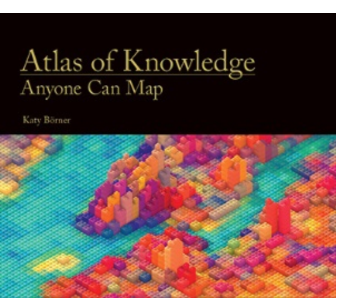
Visualization Framework

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

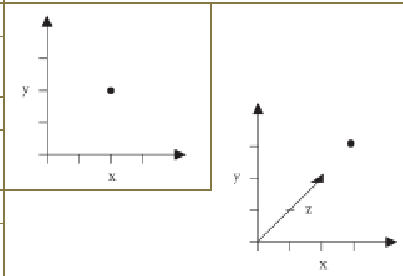
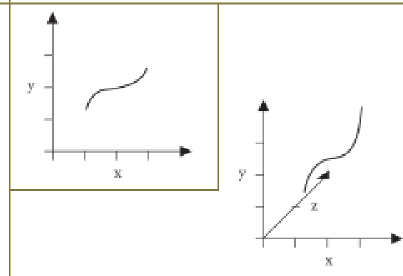
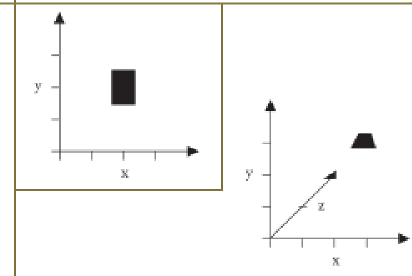
Visualization Framework

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	• nominal	• table	• geometric symbols	• spatial	• overview
• order/rank/sort	• ordinal	• chart	point	position	• zoom
• distributions	• interval	• graph	line	• retinal	• search and locate
(also outliers, gaps)	• ratio	• map	area	form	• filter
• comparisons		• network layout	surface	color	• details-on-demand
• trends			volume	optics	• history
(process and time)			• linguistic symbols	motion	• extract
• geospatial			text		• link and brush
• compositions			numerals		• projection
(also of text)			punctuation marks		• distortion
• correlations/relationships			• pictorial symbols		
			images		
			icons		
			statistical glyphs		

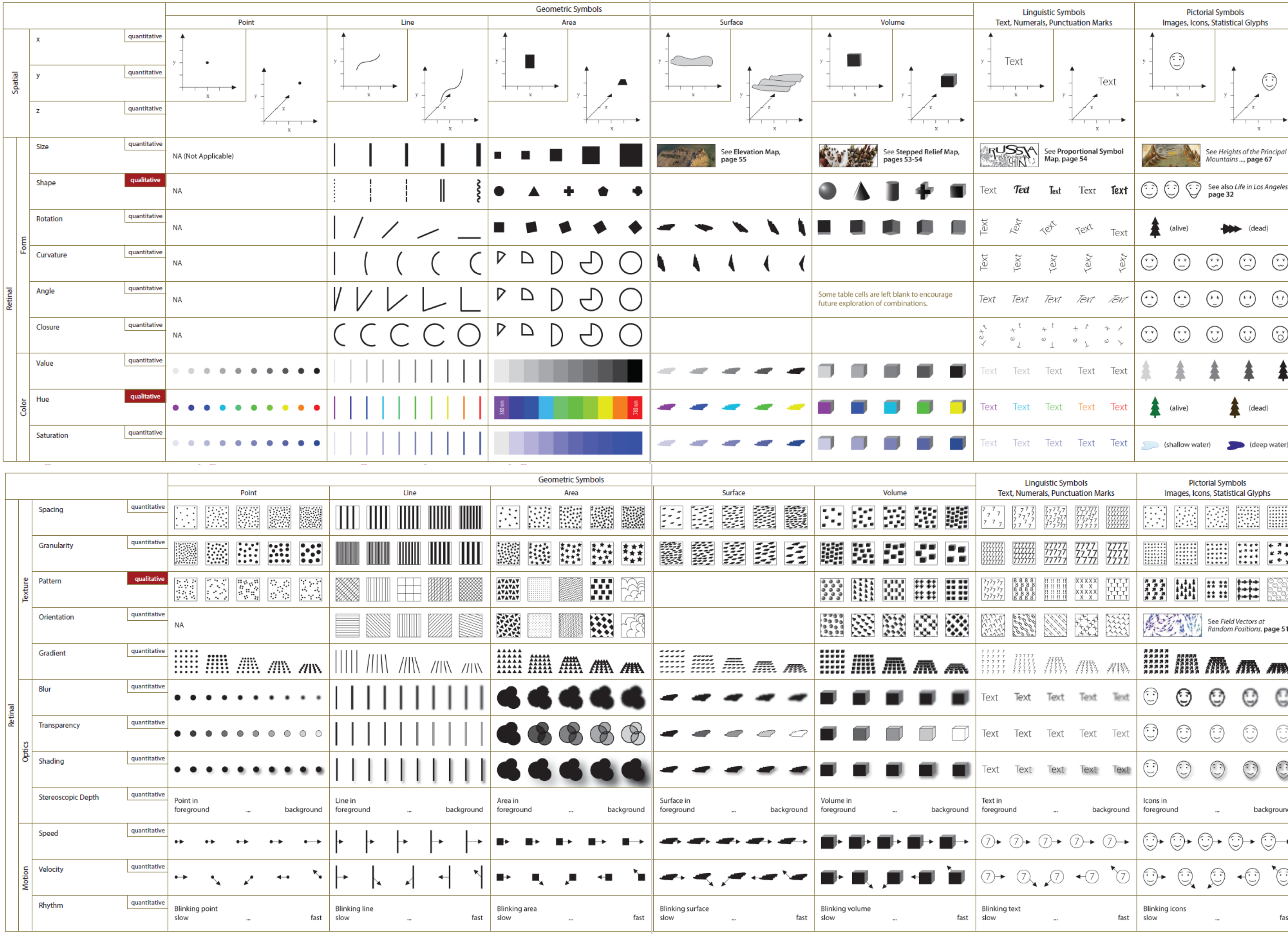


See *Atlas of Science: Anyone Can Map*, page 24

Graphic Variable Types Versus Graphic Symbol Types

				Geometric Symbols		
				Point	Line	Area
Spatial		x	quantitative			
		y	quantitative			
		z	quantitative			
Retinal	Form	Size	quantitative	NA (Not Applicable)		
		Shape	qualitative	NA		
		Rotation	quantitative	NA		
		Curvature	quantitative	NA		
		Angle	quantitative	NA		
		Closure	quantitative	NA		
	Color	Value	quantitative			
		Hue	qualitative			380 nm … 780 nm
		Saturation	quantitative			

Graphic Variable Types Versus Graphic Symbol Types



Data Visualization Literacy: Research and Tools that Advance Public Understanding of Scientific Data

+ Methods

10TH ANNUAL MUCSC

MIDWEST UNDERGRADUATE COGNITIVE SCIENCE CONFERENCE

Register for free: http://ivmooc.cns.iu.edu. Class restarted Jan 9, 2018.

The Information Visualization MOOC

ivmooc.cns.iu.edu

Students from more than 100 countries

350+ faculty members

#ivmooc

Course Schedule

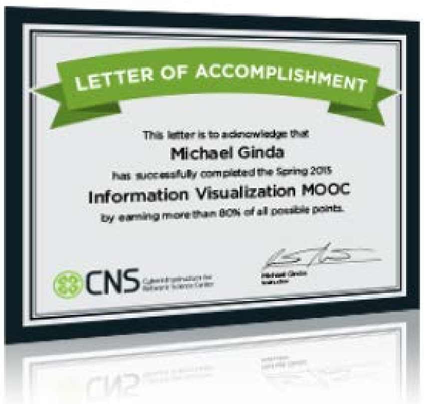
Part 1: Theory and Hands-On

- **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.

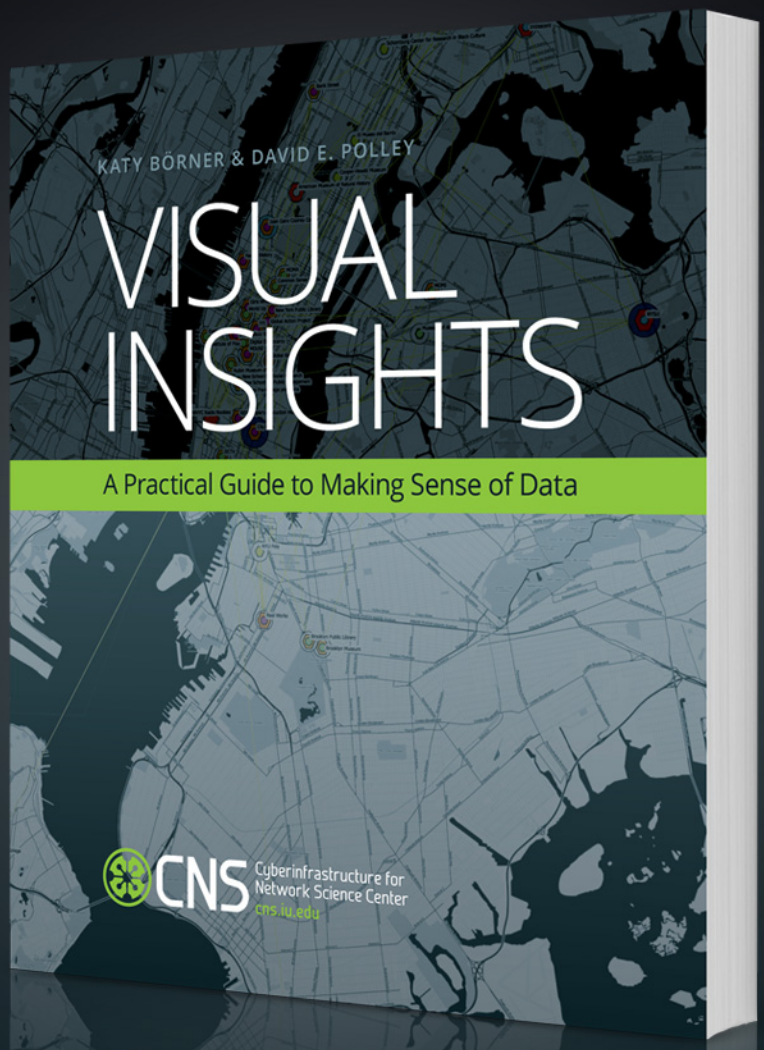
Final grade is based on Homework and Quizzes (**10%),** Midterm (**20%**), Final (**30%**), Client Project (**30%**), and Class Participation (**10%**).

The IVMOOC Companion Textbook

This textbook offers a gentle introduction to the design of insightful visualizations. It seamlessly blends theory and practice, giving readers both the theoretical foundation and the practical skills necessary to render data into insights.

The book accompanies the Information Visualization MOOC that attracted students, scholars, and practitioners from many fields of science and more than 100 different countries.

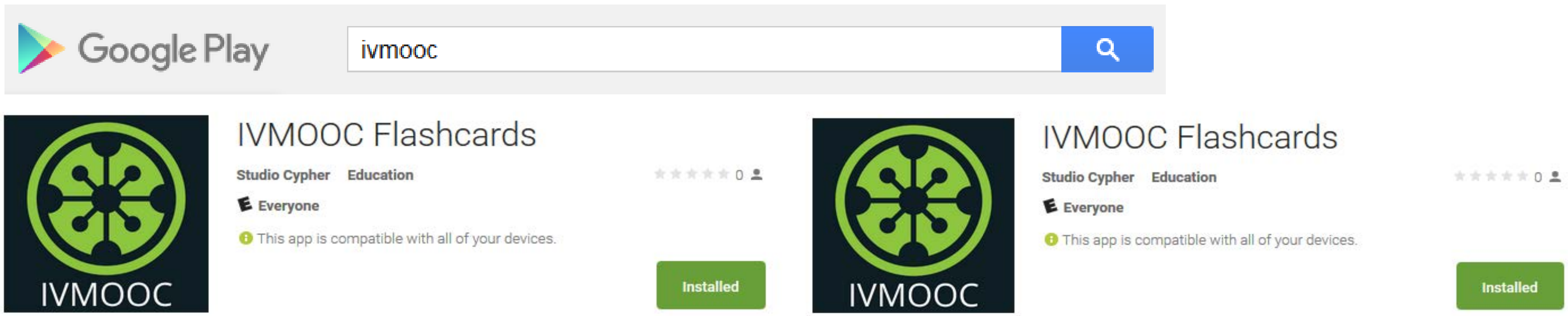
http://ivmooc.cns.iu.edu

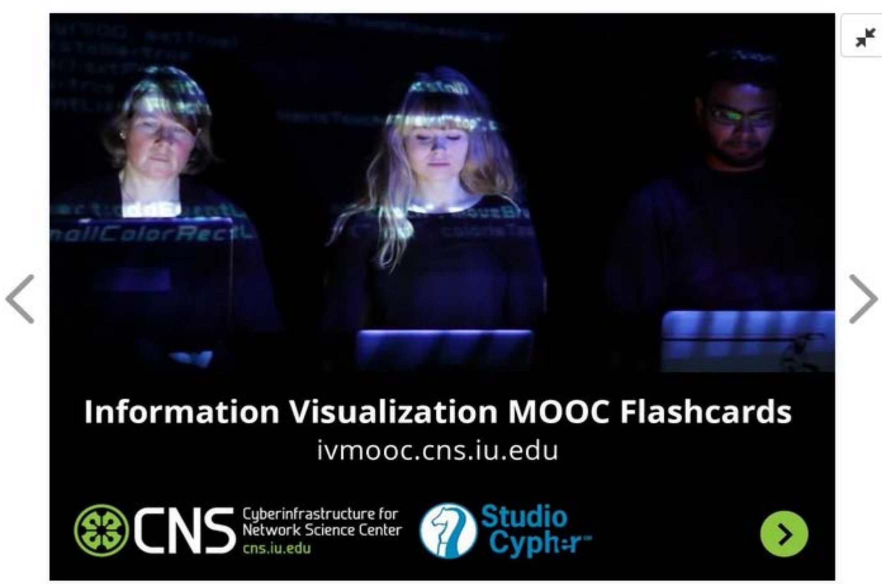


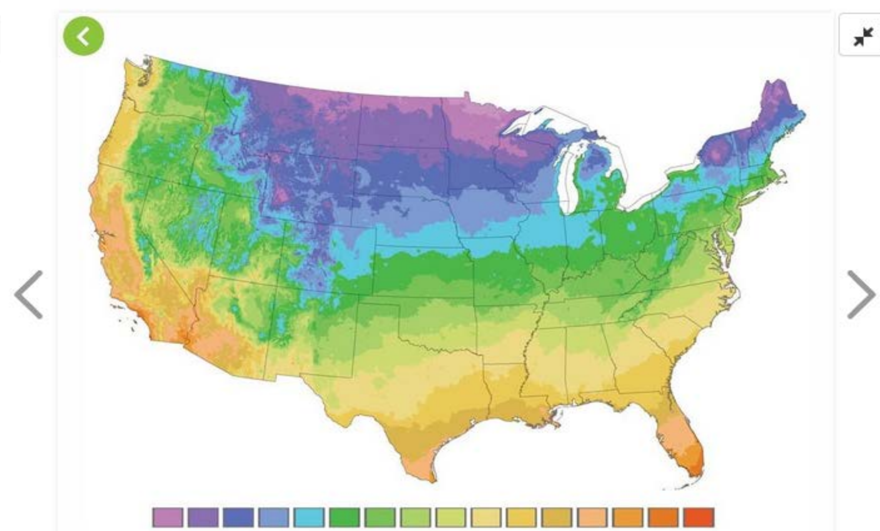
cns.iu.edu/ivmoocbook14.html

IVMOOC App

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

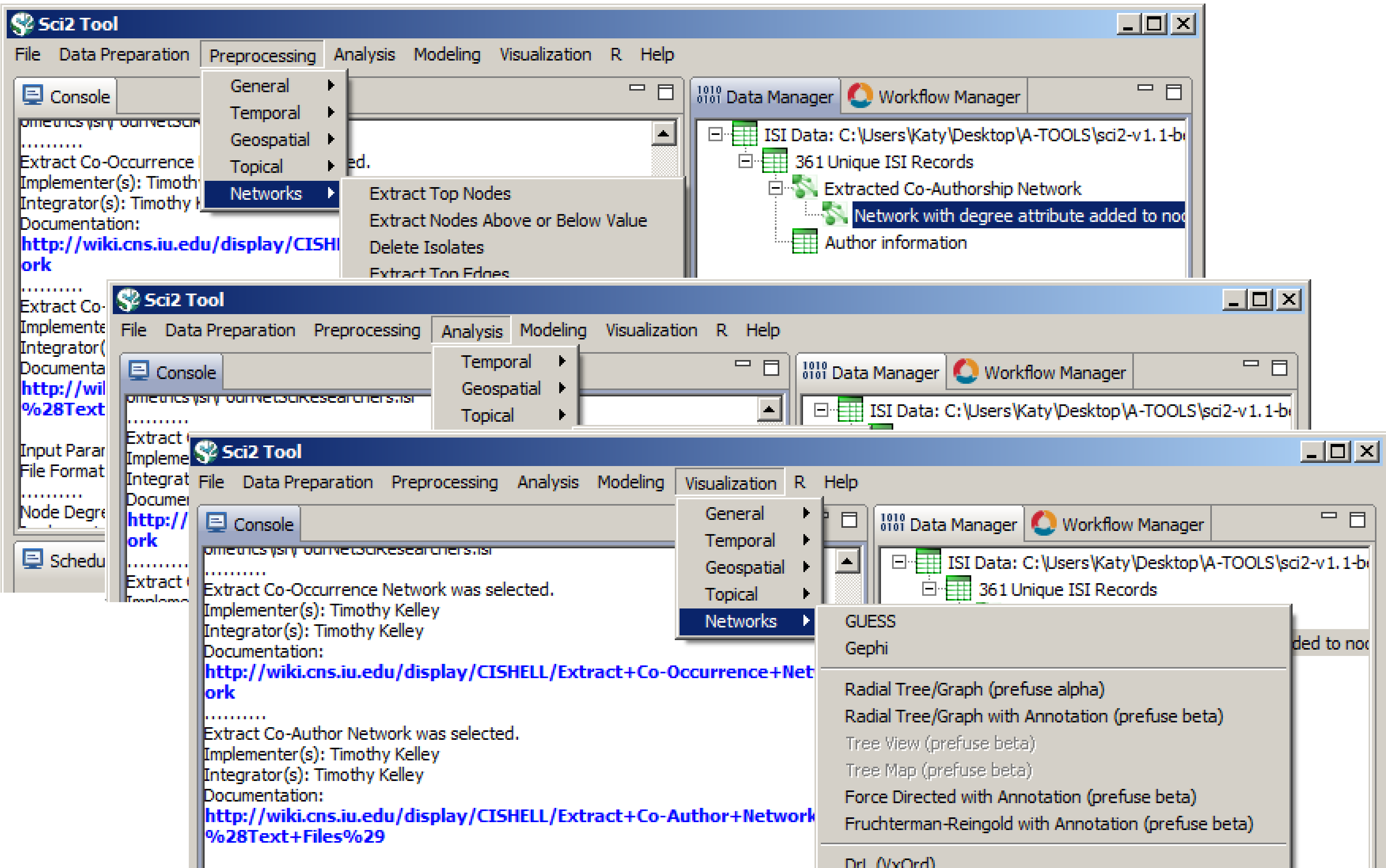






Sci2 Tool Interface Components Implement Vis Framework

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



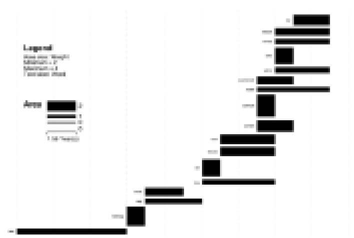
Load **One** File and Run **Many** Analyses and Visualizations

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

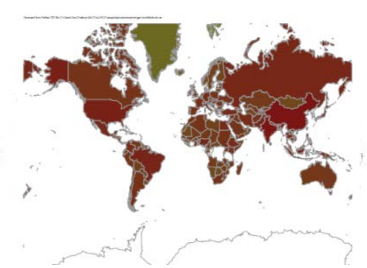
Location	*Count*	*# Citations*
Netherlands	13	292
United States	9	318
Germany	11	36
United Kingdom	1	2

Temporal Burst Analysis—p. 48



Geospatial Analysis—p. 52

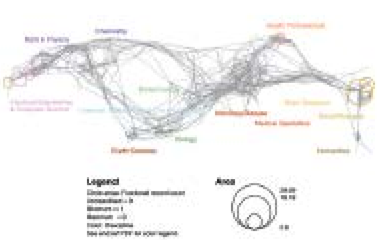
Geospatial Analysis—p. 52



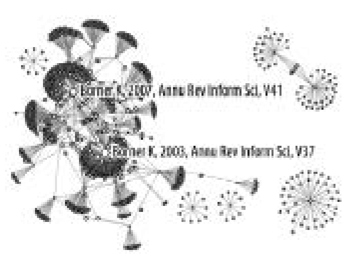
Load **One** File and Run **Many** Analyses and Visualizations

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

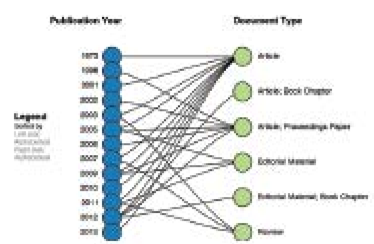
Topical Analysis–p. 56



Paper Citation Network–p. 60



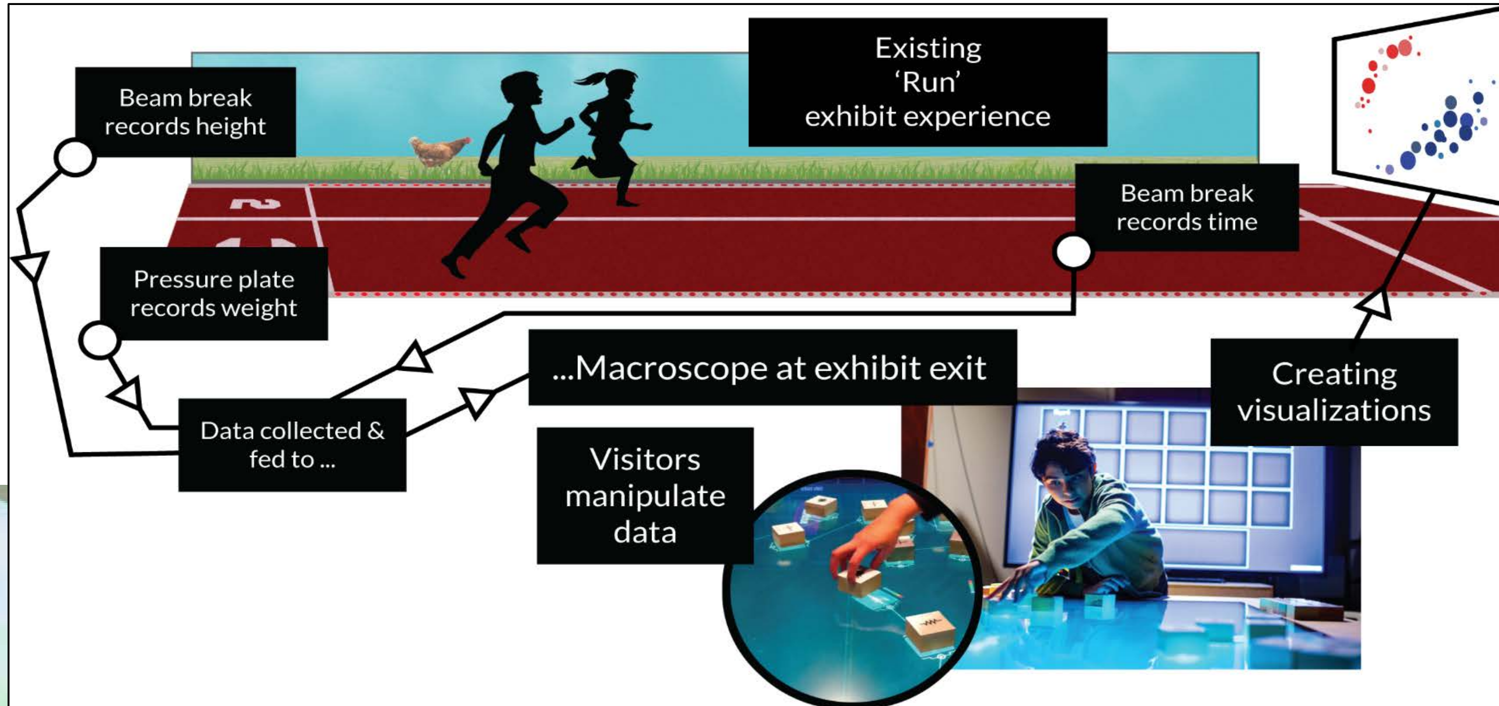
Bi-Modal Network–p. 60

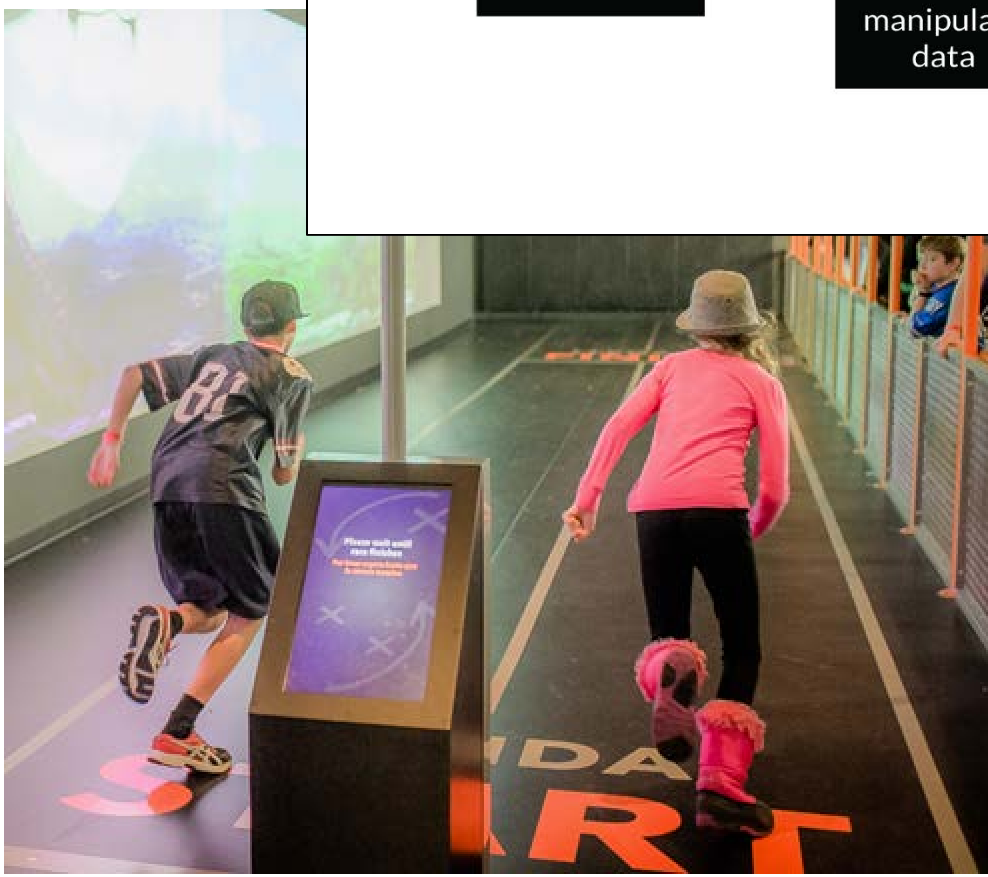


Co-author and many other bi-modal networks.

Data Visualization Literacy:
Research and Tools that Advance Public Understanding of Scientific Data

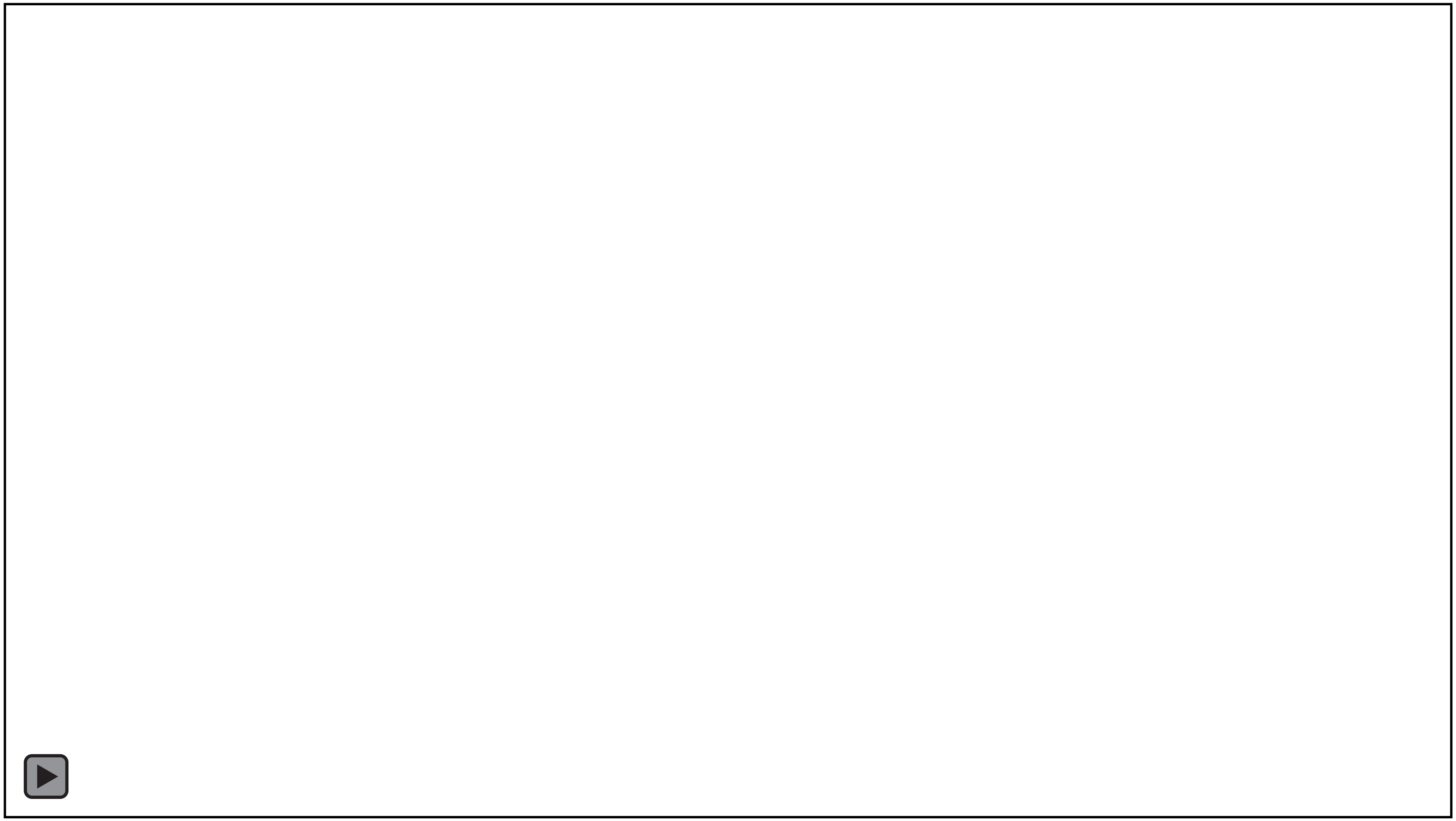
Data Visualization Literacy: Research and Tools that Advance Public Understanding of Scientific Data.
Katy Borner & Kylie Peppler (IU), Bryan Kennedy (SMM), Stephen Uzzo (NYSCI), Joe Heimlich (COSI).
NSF AISL award #1713567.



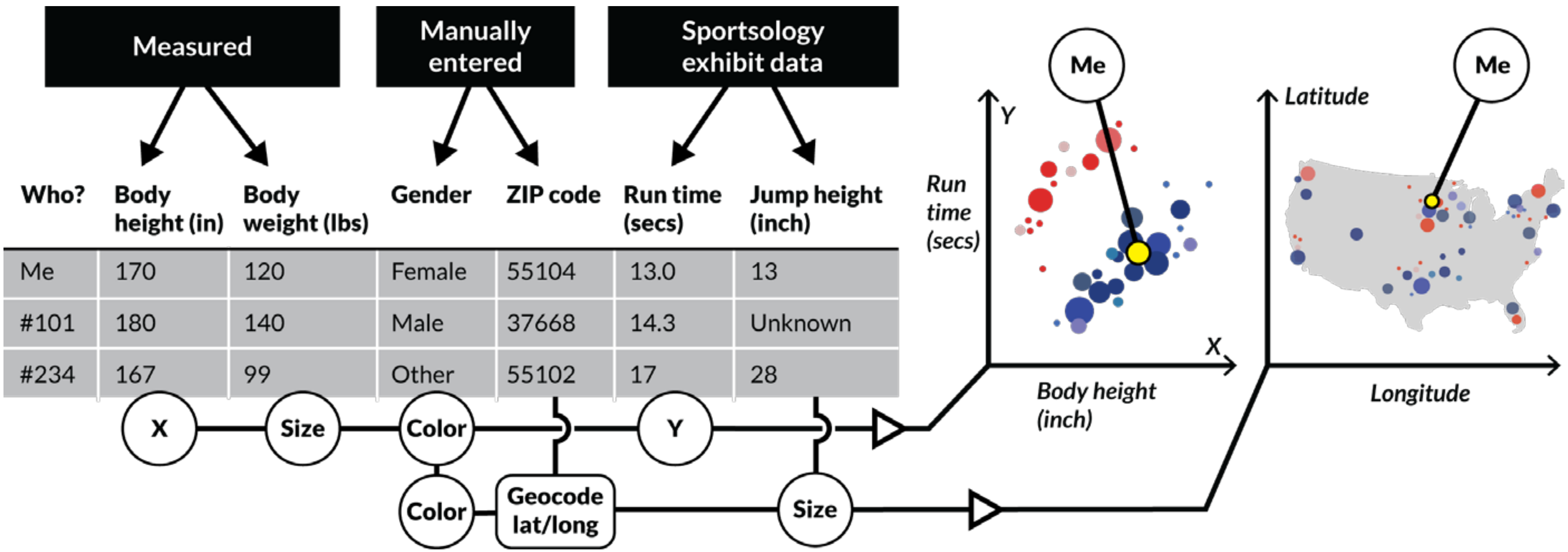


Sketch of the *Run* exhibit including data collection (top) and macroscope add-on that lets interested visitors explore more complex data visualizations using table-top displays.

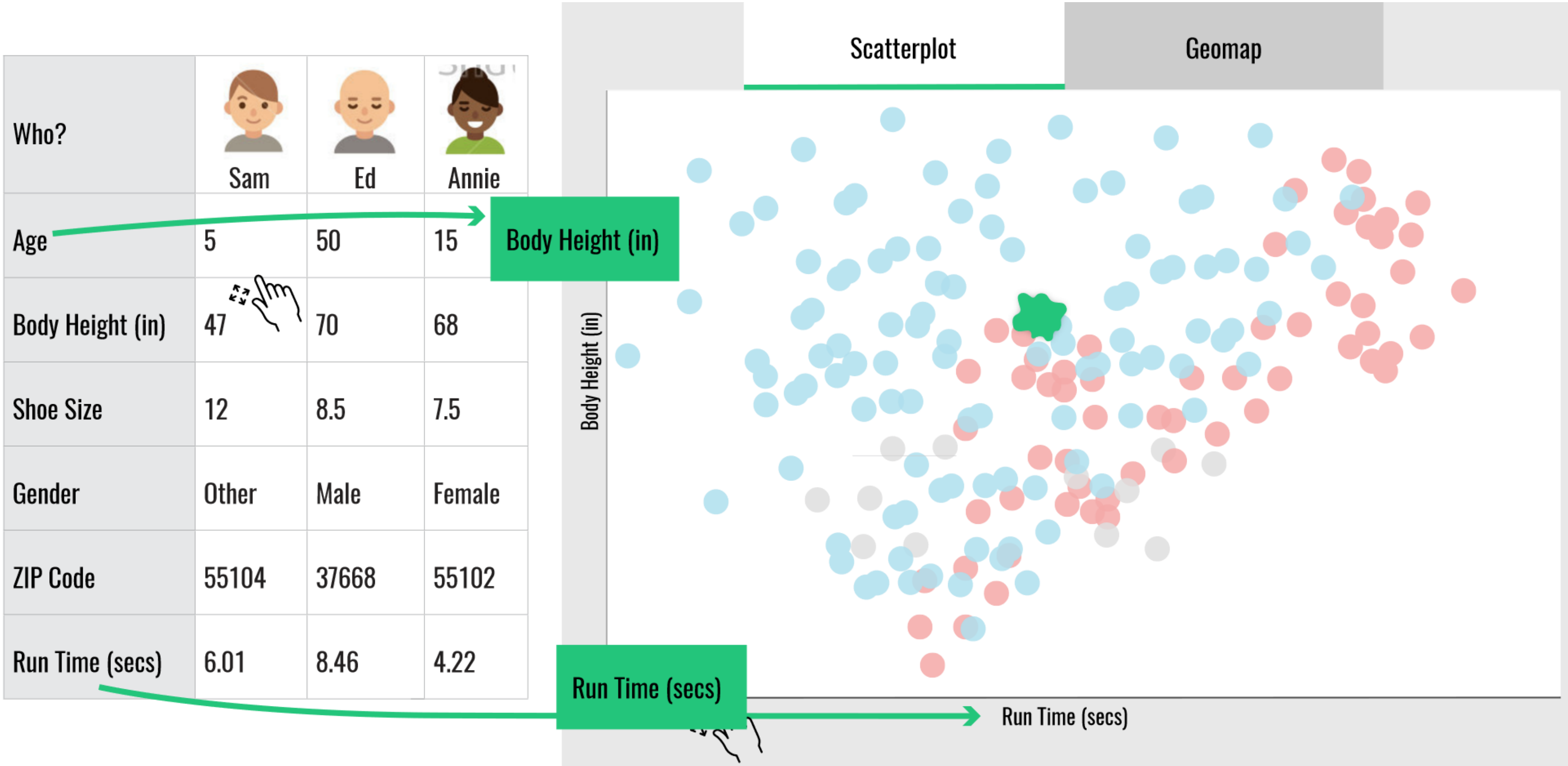
Sportsology @ Science Museum of Minnesota



https://www.youtube.com/watch?v=oy34R45EfBg



xMacroscope general setup and activity—Raw data on left is converted to visualization on right by dragging and dropping (or connecting) column headers to axes, paint buckets, size, and shape.



xMacroscope general setup and activity—Raw data on left is converted to visualization on right by dragging and dropping (or connecting) column headers to axes, paint buckets, size, and shape.

Data Visualization Literacy: Research and Tools that Advance Public Understanding of Scientific Data

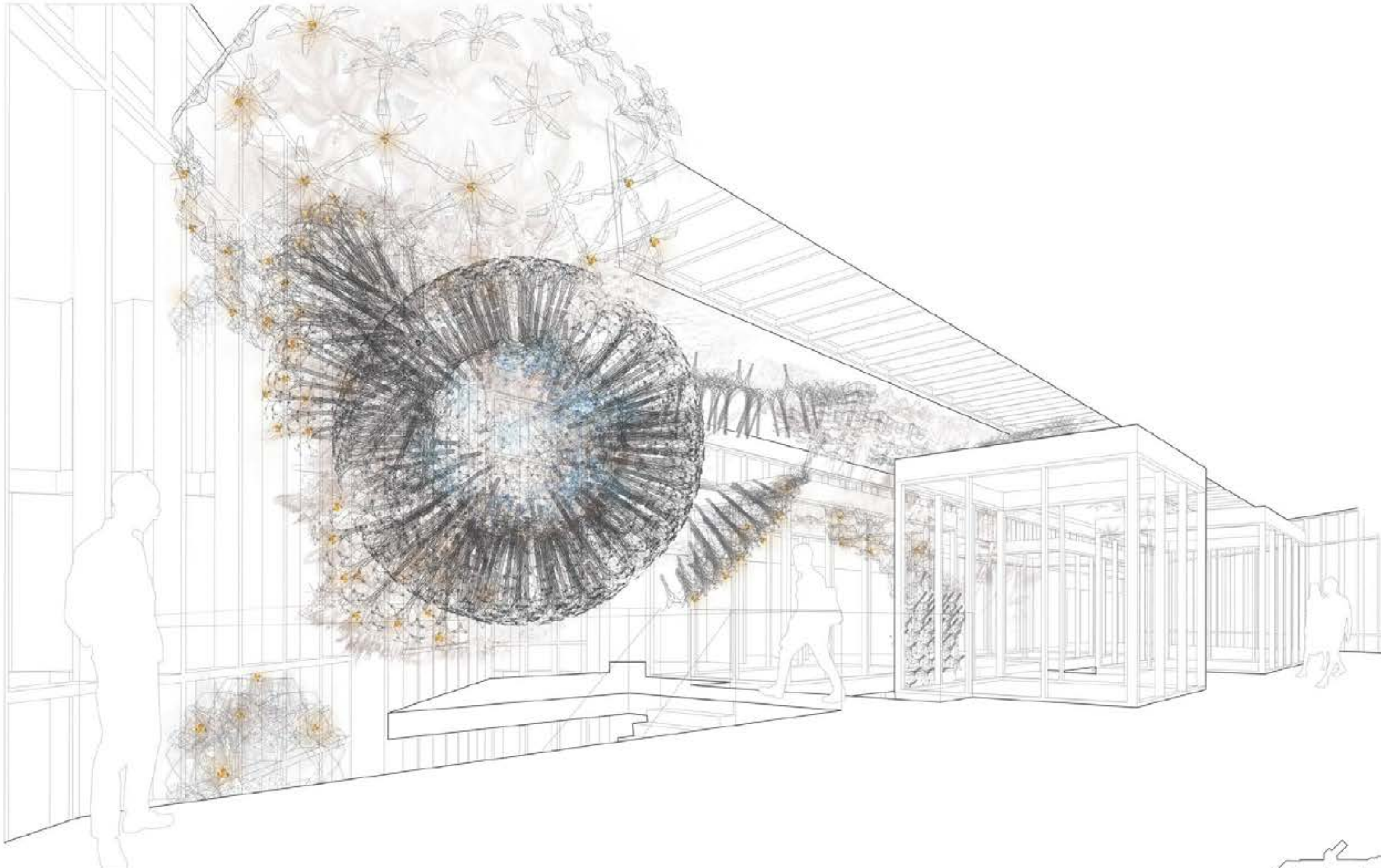
10TH ANNUAL
MUCSC
MIDWEST UNDERGRADUATE COGNITIVE SCIENCE CONFERENCE

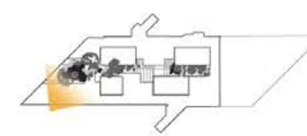
Visualizing the Internet of Things (IoT)

Using large scale datasets, advanced data mining and visualization techniques, and substantial computing resources.

Work by Philip Beesley | www.philipbeesley.ca | www.lasg.ca

Sentient Chamber, National Academy of Sciences, Washington, D.C. (2016)





Luddy Hall Installation
Indiana University Bloomington
April 29 2017

UPPER ATRIUM

Philip Beesley • Living Architecture Systems

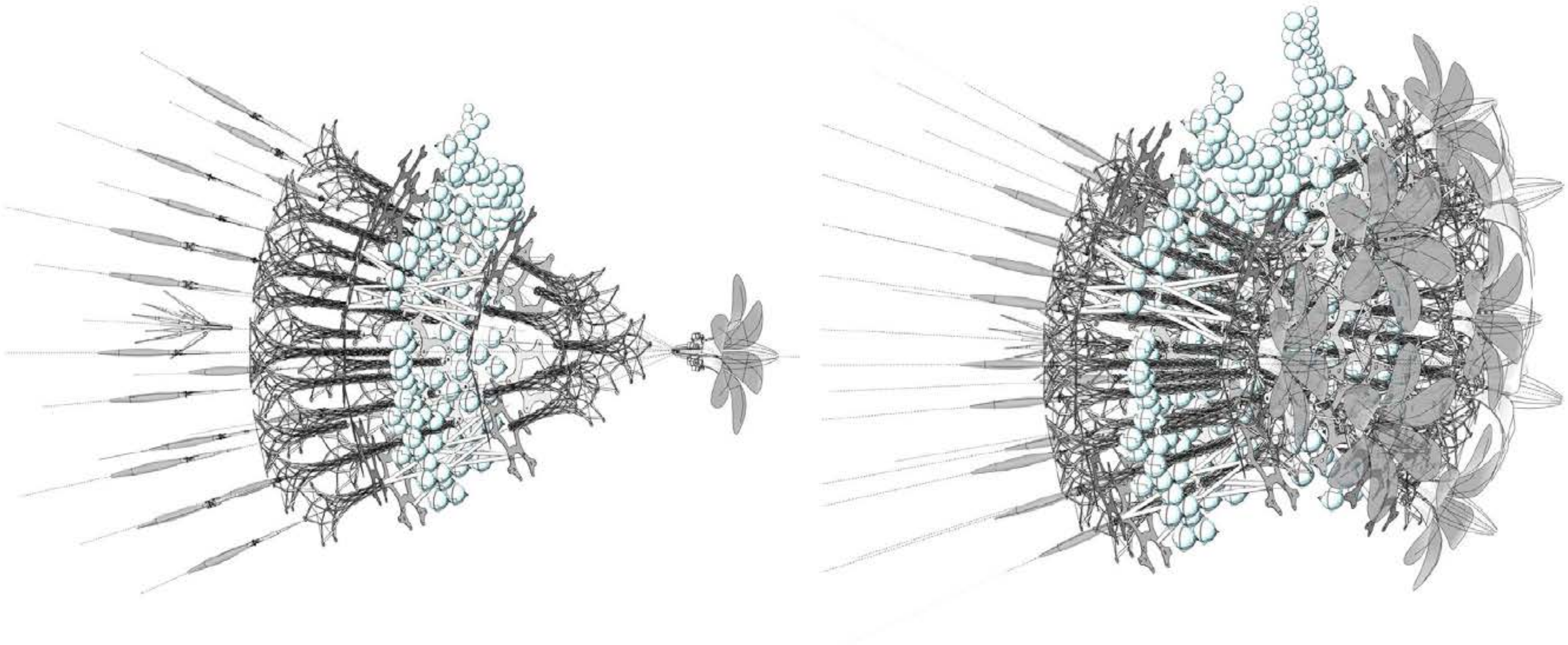


Photo shared by Bruce Shei

Photo by Tracey Theriault

Amatria Unveiled by Andreas Bueckle et al. Data visualizations of sensor/actuator positions and types, energy and communication flows, and emergent behavior of smart environments.

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

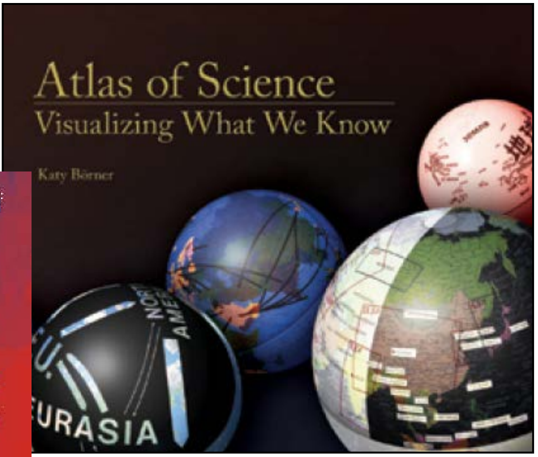
Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1

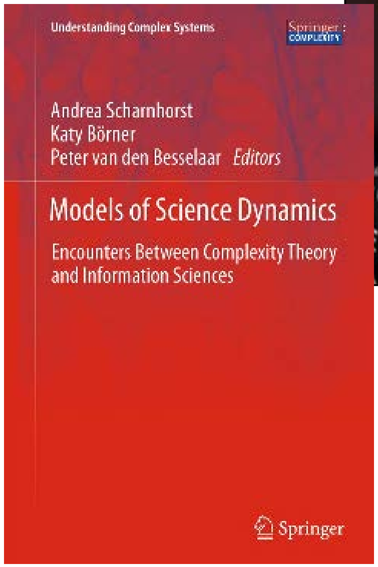
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

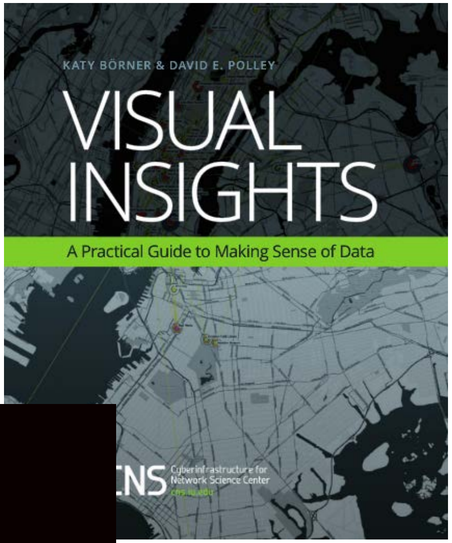
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

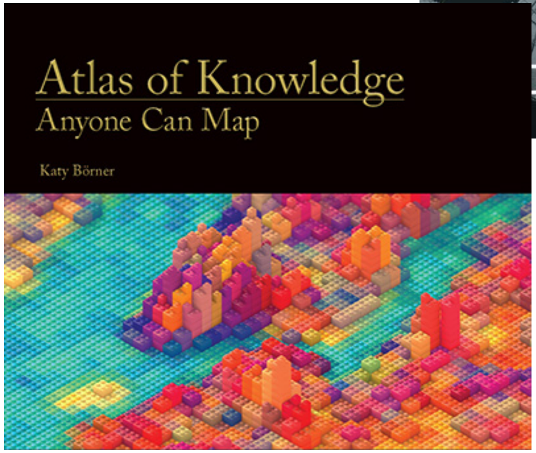
Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

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









INDIANA UNIVERSITY BLOOMINGTON

RESOURCES / TWITTER

Interdisciplinary Training in **Complex Networks and Systems**

THE PROGRAM | RESEARCH | HOW TO APPLY | STUDENTS | COLLOQUIA | NEWS

Share Tweet Email

The program

Understanding complex networked systems is key to solving some of the most vexing problems confronting humankind, from discovering how dynamic brain connections give rise to thoughts and behaviors, to detecting and preventing the spread of misinformation or unhealthy behaviors across a population. Graduate training, however, typically occurs in one of two dimensions: experimental and observational methods in a specific area such as biology and sociology, or in general methodologies such as machine learning and data science.

CNS NRT

https://cns-nrt.indiana.edu

All papers, maps, tools, talks, press are linked from http://cns.iu.edu
These slides are at http://cns.iu.edu/presentations.html

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience