Data Visualizations for Effective Exploration and Communication

Katy Börner @katycns

Victor H. Yngve Distinguished Professor of
Intelligent Systems Engineering & Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics, Computing, and Engineering
Indiana University Network Science Institute (IUNI)
Indiana University, Bloomington, IN, USA
+ 2018 Humboldt Fellow, TU Dresden, Germany



IBS Workshop Media Computing
IBS Haus (ehemaliges Verwaltungsgebäude für die Brikettfabrik Erika)
Lauta, Germany

February 5, 2019

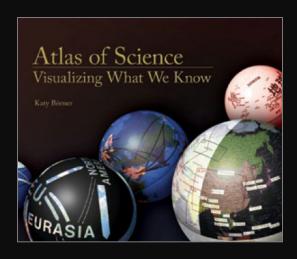
IBS Workshop Media Computing

Dienst	ag	5. Februar 2019	
ab 10 l	Uhr	Registrierung / Check In	
12:00	Uhr	Welcome Lunch	
13:30	Uhr	Session I: Auftakt Gegenseitige Vorstellung & Ausarbeitung im Duett	
15:30	Uhr	Kaffeepause	
16:00	Uhr	Session II: Mutual presentations im pitch your project Stil 1. Folie: Motivation/Anwendungsfall 2. Folie: Methodik 3. Ergebnisse	
18:30	Uhr	Dinner for all	
21:00	Uhr	Get together 2er Team-Fortführung im Entertainment-Format Science Slam: Fail-Session 2. Tag: Methodik	
Mittwo	och	6. Februar 2019	
ab 8:00	0 Uhr	Frühstück	
9:00 U	hr	Session III: Methodik & Analytik für Promotionsarbeiten (PIs) PhD-Struktur Argumentationsstruktur Experimentalsetup & Übertragbarkeit für PhD	

8-10a EST

Maps of Science & Technology

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



Maps of Science & Technology

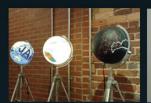
http://scimaps.org



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

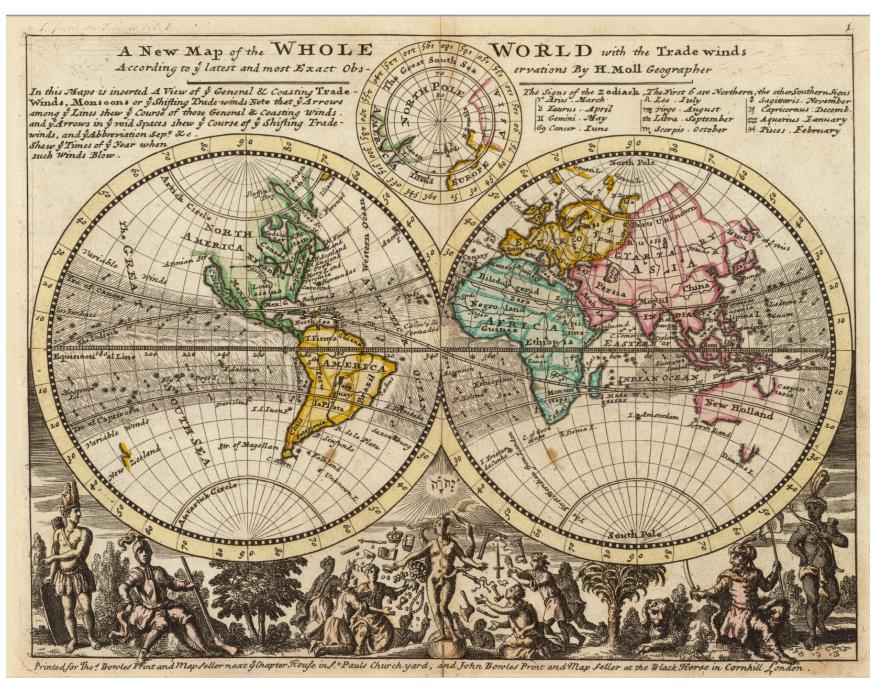




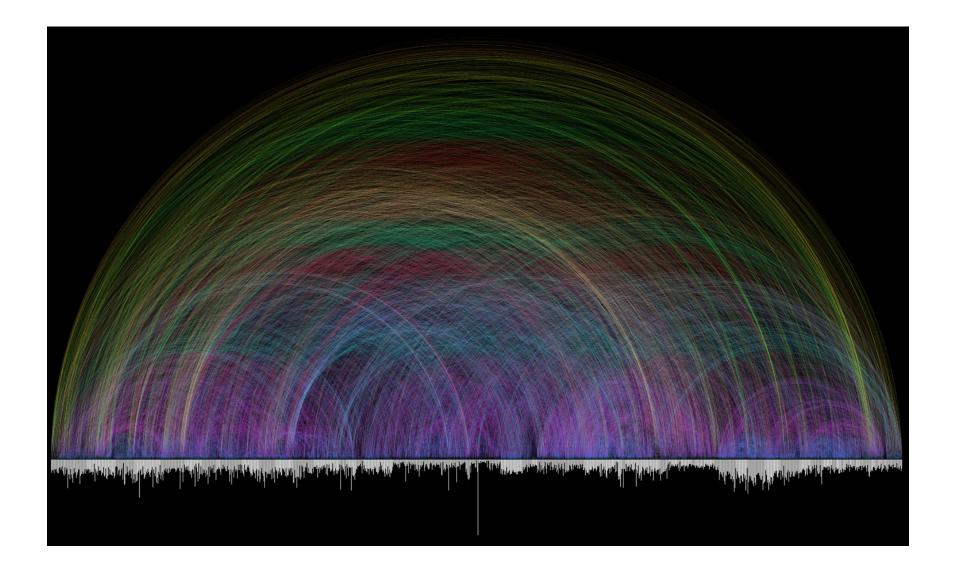


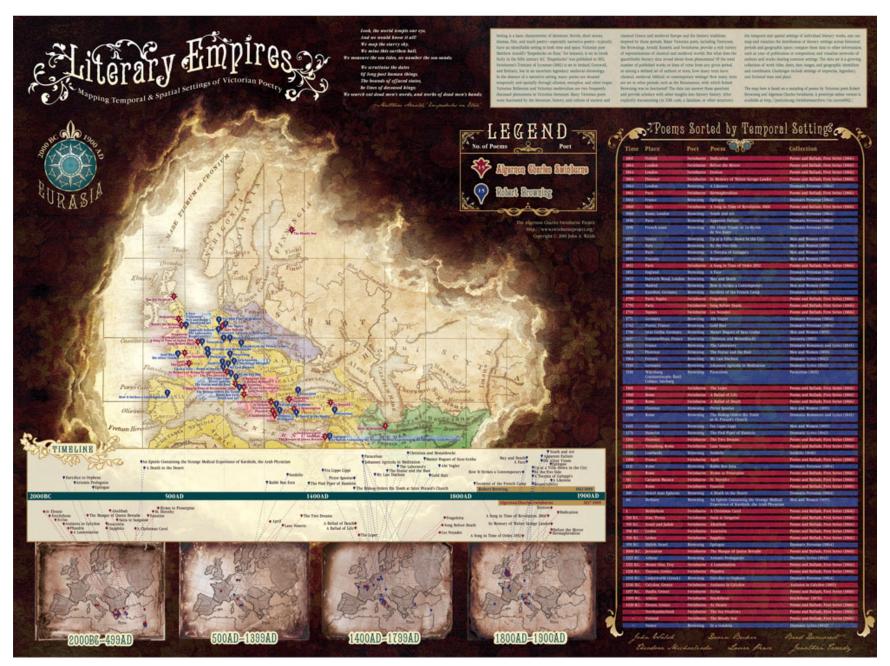
The David J. Sencer CDC Museum, Atlanta, GA. January 25 - June 17, 2016.

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



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

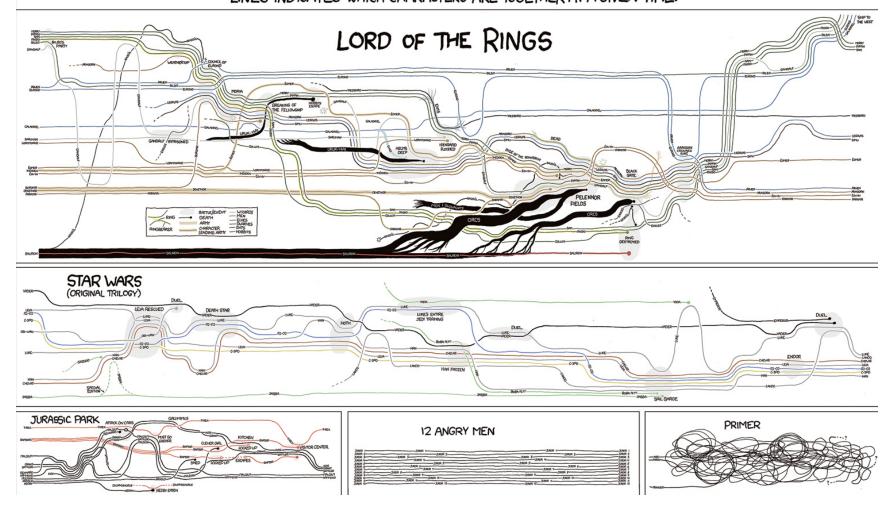


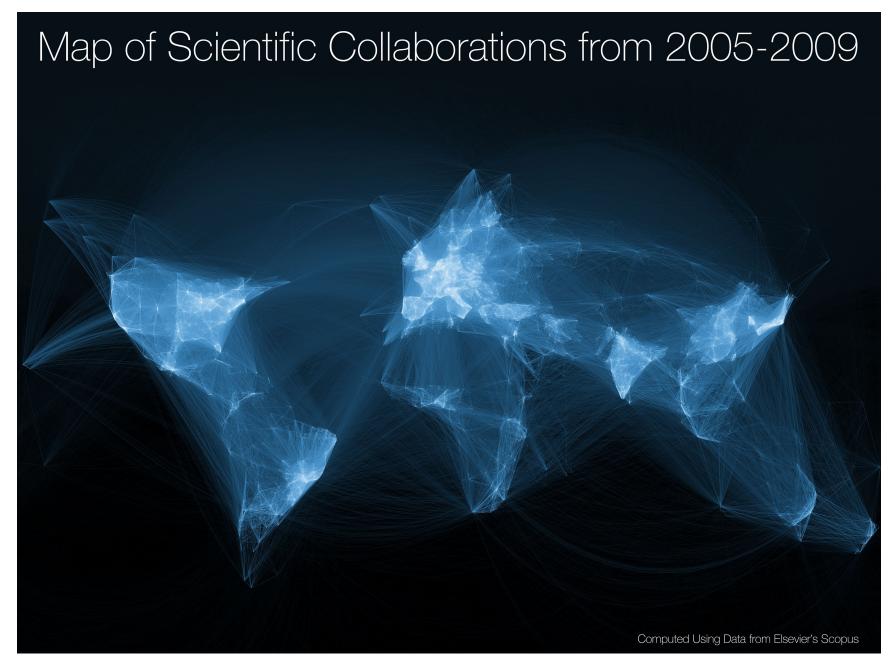


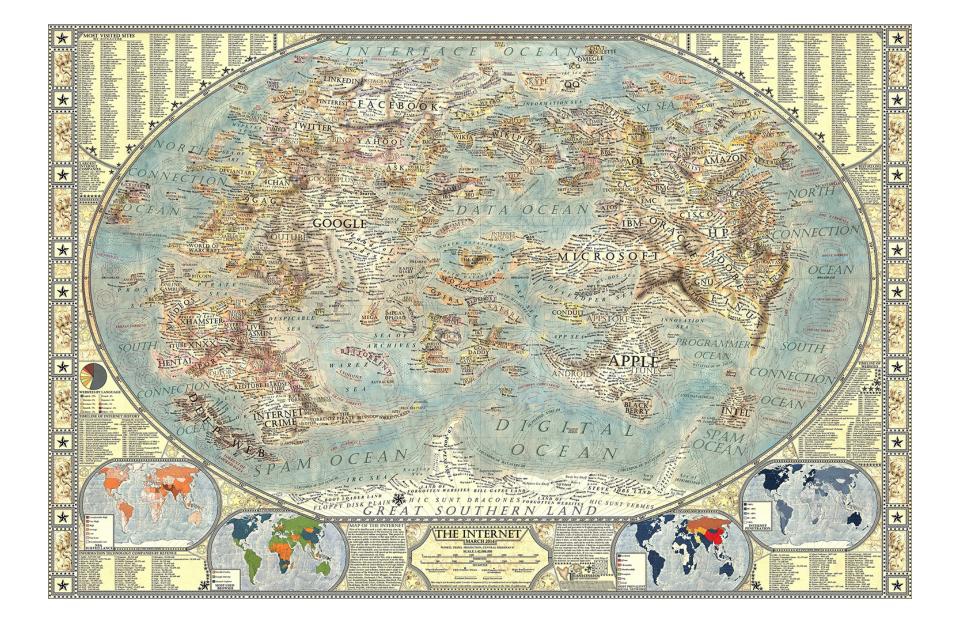
VI.7 Literary Empires: Mapping Temporal and Spatial Settings of Victorian Poetry

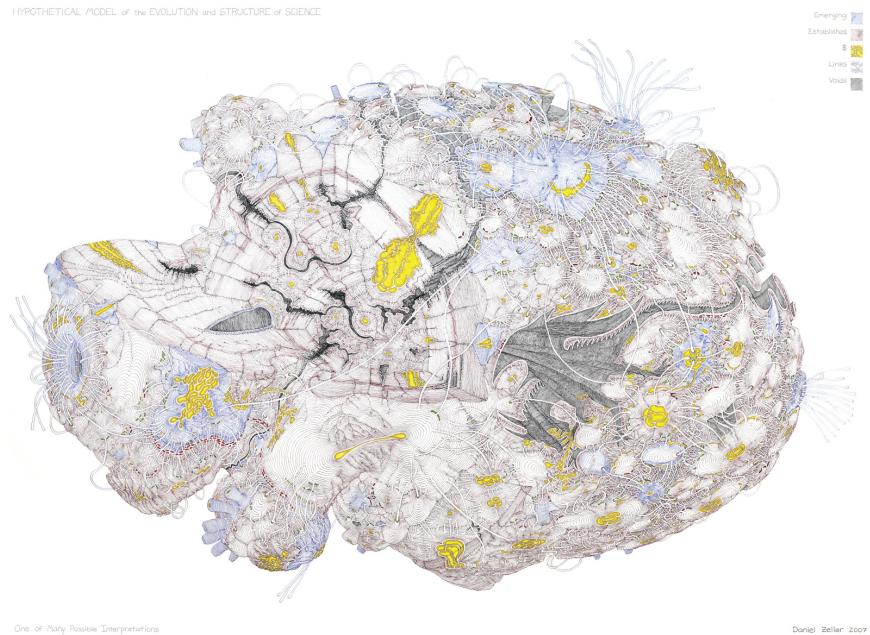
John A. Walsh, Devin Becker, Bradford Demarest, Jonathan Tweedy, Theodora Michaelidou, and Laura Pence - 2010

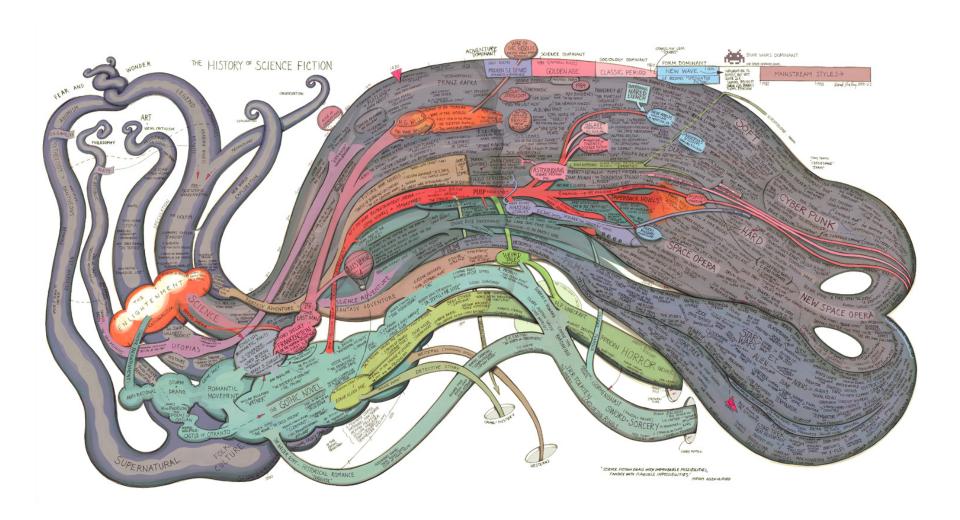
THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS. THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.



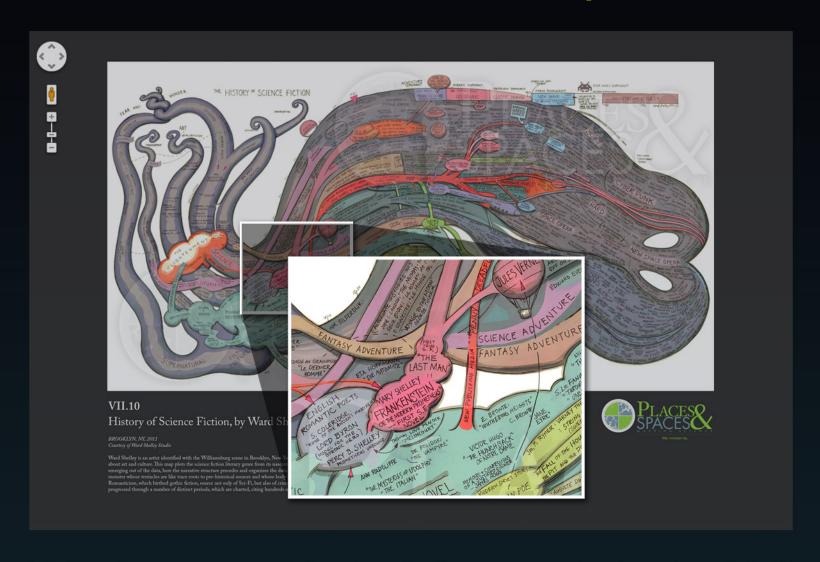








Check out our Zoom Maps online!



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





100 Maps of Science on Display in Hörsaalzentrum (Augustusplatz), Uni Leipzig, July 13-30, 2018



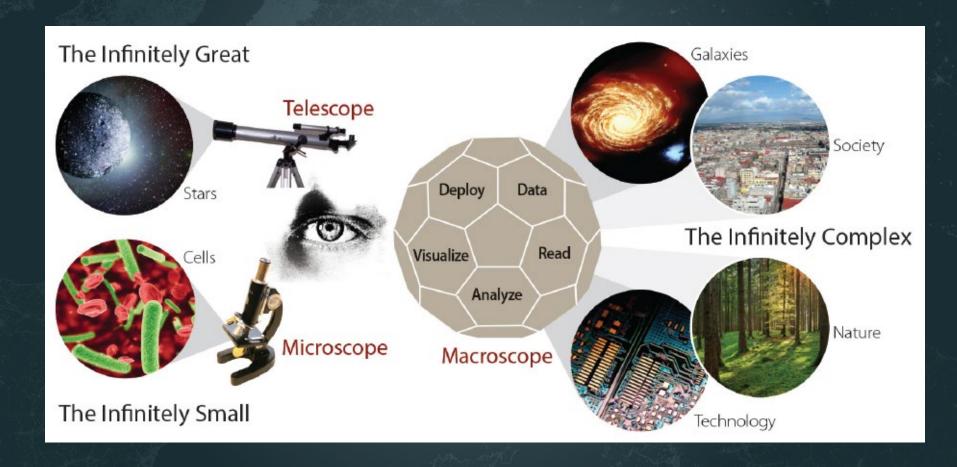




MAPS
vs.
MACROSCOPES

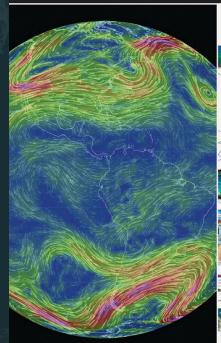


Microscopes & Telescopes vs. MACROSCOPES



(i) MACROSCOPES FOR INTERACTING WITH SCIENCE











Earth

Weather on a worldwide scale

AcademyScope

Exploring the scientific landscape

Mapping Global Society

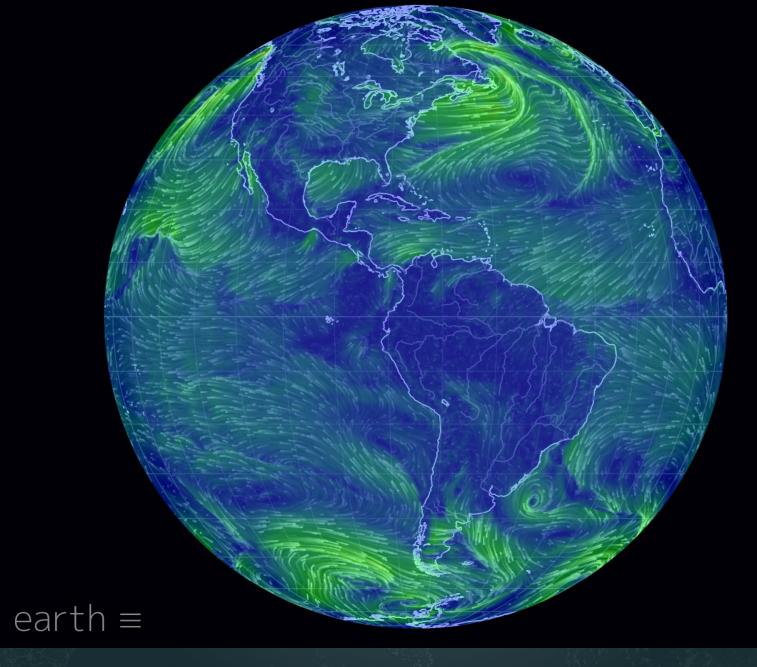
Local news from a global perspective

Charting Culture

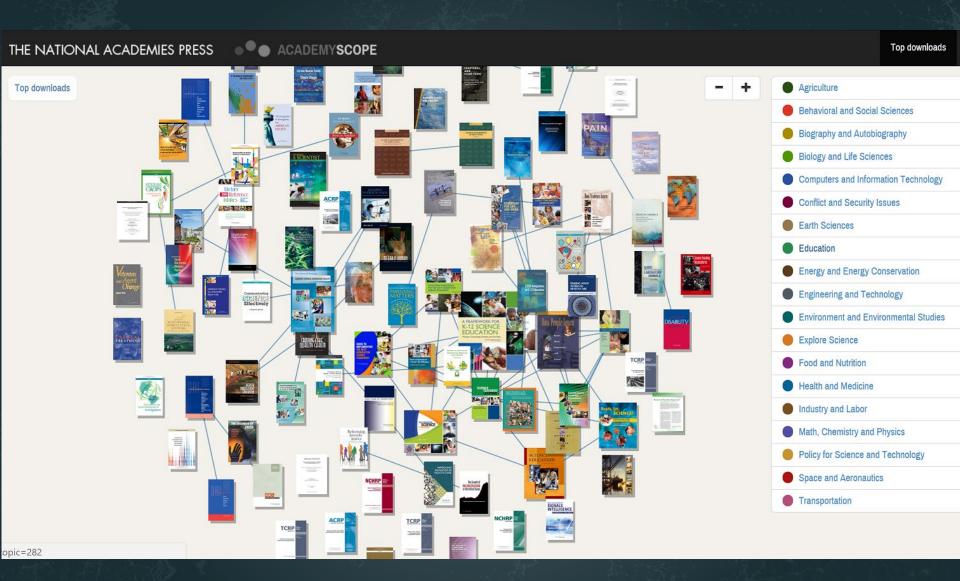
2,600 years of human history in 5 minutes

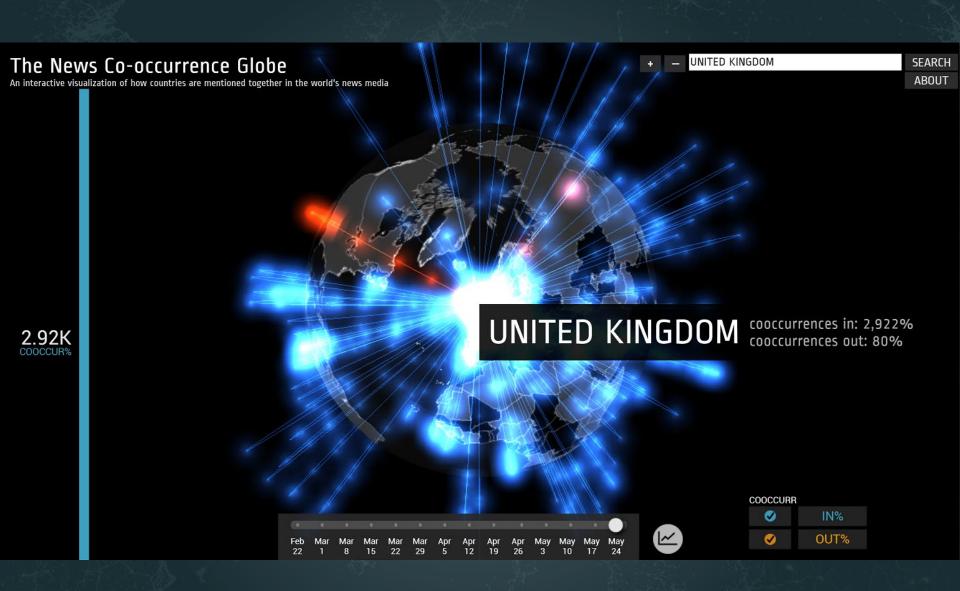
Iteration XI (2015): Macroscopes for Interacting with Science

http://scimaps.org/iteration/11

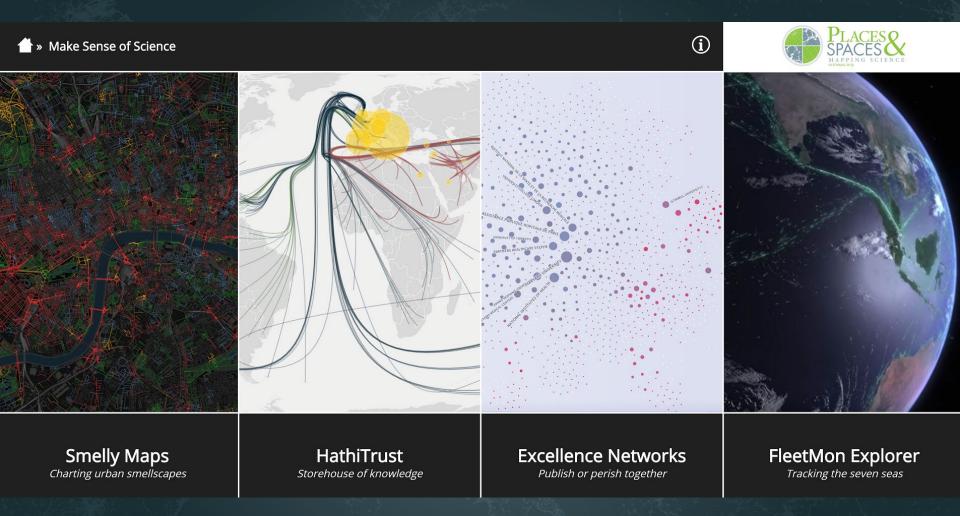


Earth – Cameron Beccario





Mapping Global Society - Kalev Leetaru



Iteration XII (2016): Macroscopes for Making Sense of Science http://scimaps.org/iteration/12





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

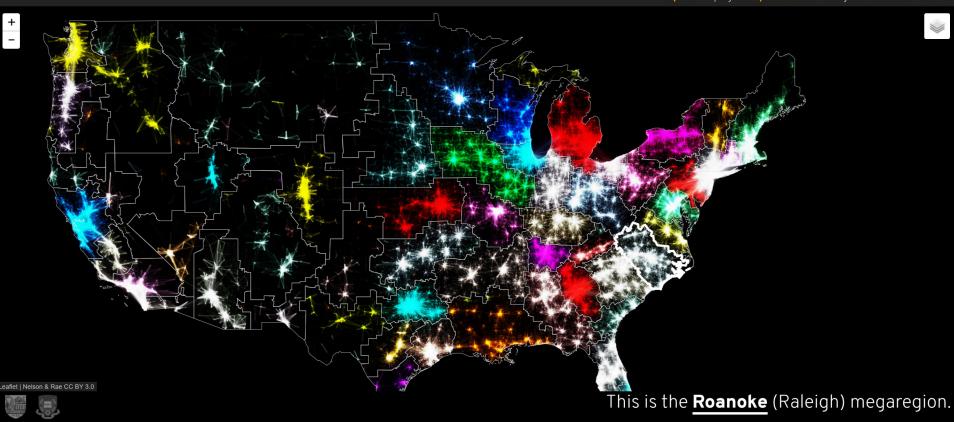




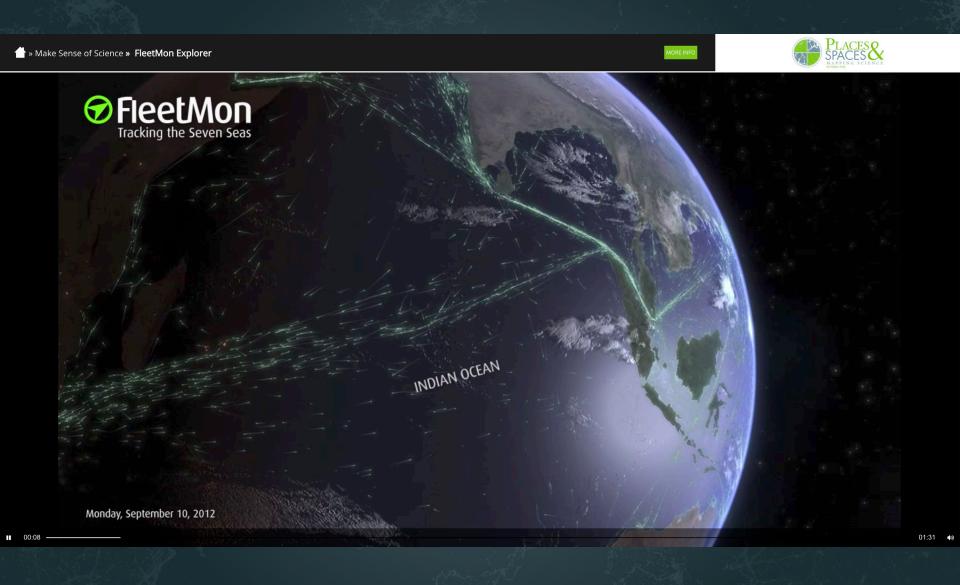


THE MEGAREGIONS OF THE US

Explore the new geography of commuter connections in the US. **Tap** to identify regions. **Tap and hold** to see a single location's commuteshed.



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



FleetMon Explorer – FleetMon – 2012

Models of Science & Technology

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













Government, academic, and industry leaders discussed challenges and opportunities associated with using big data, visual analytics, and computational models in STI decision-making.

Conference slides, recordings, and report are available via http://modsti.cns.iu.edu/report

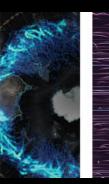












Modeling and Visualizing Science and Technology Developments

National Academy of Sciences Sackler Colloquium, December 4-5, 2017, Irvine, CA

Rankings and the Efficiency of Institutions

H. Eugene Stanley | Albert-László Barabási | Lada Adamic | Marta González | Kaye Husbands Fealing | Brian Uzzi | John V. Lombardi

Higher Education and the Science & Technology Job Market

Katy Börner | Wendy L. Martinez | Michael Richey | William Rouse | Stasa Milojevic | Rob Rubin | David Krakauer

Innovation Diffusion and Technology Adoption

William Rouse | Donna Cox | Jeff Alstott | Ben Shneiderman | Rahul C. Basole | Scott Stern | Cesar Hidalgo

Modeling Needs, Infrastructures, Standards

Paul Trunfio | Sallie Keller | Andrew L. Russell | Guru Madhavan | Azer Bestavros | Jason Owen-Smith







PROGRAMS

Sackler Colloquia

- About Sackler Colloquia
- Upcoming Colloquia
- Completed Colloquia
- » Sackler Lectures
- » Video Gallery
- Connect with Sackler Colloquia
- » Give to Sackler Colloquia

Cultural Programs

Distinctive Voices

Kavli Frontiers of Science

Keck Futures Initiative

LabX

Sackler Forum

Science & Entertainment Exchange



Modeling and Visualizing Science and Technology Developments



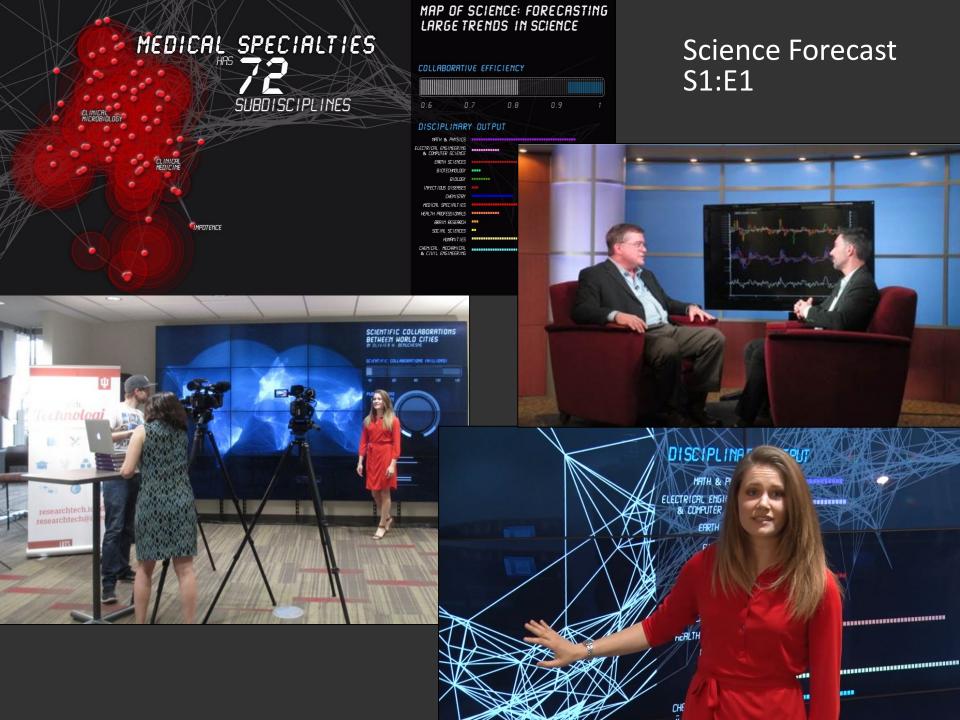
December 4-5, 2017; Irvine, CA Organized by Katy Börner, H. Eugene Stanley, William Rouse and Paul Trunfio

Overview

This colloquium was held in Irvine, CA on December 4-5, 2017.

This colloquium brought together researchers and practitioners from multiple disciplines to present, discuss, and advance computational models and visualizations of science and technology (S&T). Existing computational models are being applied by academia, government, and industry to explore questions such as: What jobs will exist in ten years and what career paths lead to success? Which types of institutions will likely be most innovative in the future? How will the higher education cost bubble burst affect these institutions? What funding strategies have the highest return on investment? How will changing demographics, alternative economic growth trajectories, and relationships among nations impact answers to these and other questions? Large-scale datasets (e.g., publications, patents, funding, clinical trials, stock market, social media data) can now be utilized to simulate the structure and evolution of S&T. Advances in computational power have created the possibility of implementing scalable, empirically validated computational models. However, because the databases are massive and multidimensional, both the data and the models tend to exceed human comprehension. How can advances in data visualizations be effectively employed to communicate the data, the models, and the model results to diverse stakeholder groups? Who will be the users of next generation models and visualizations and what decisions will they be addressing.

Videos of the talks are available on the Sackler YouTube Channel.

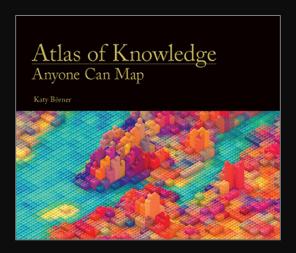


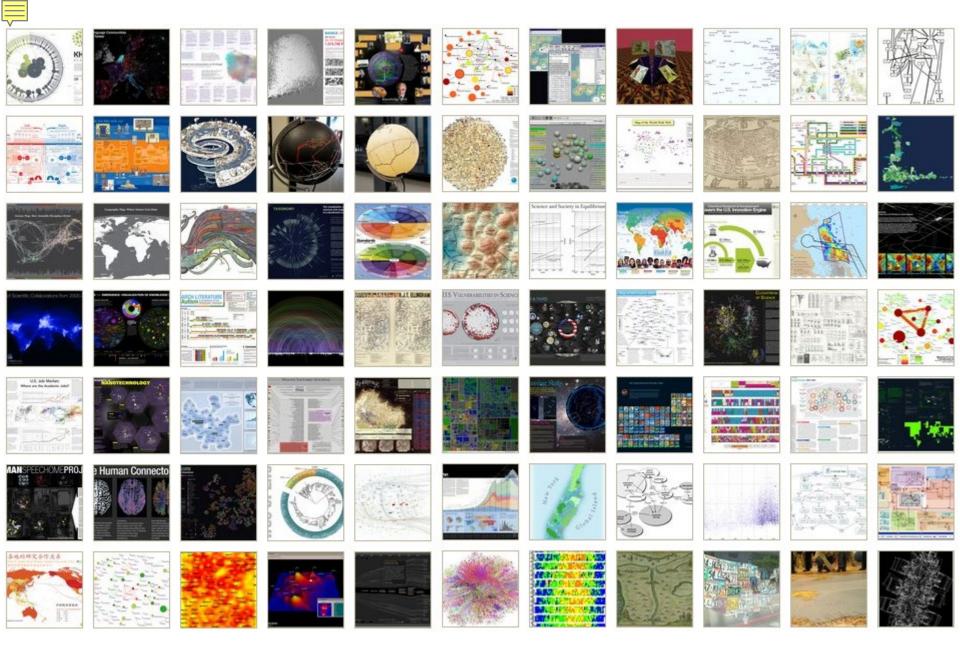
Science Forecast S1:E1



Making Science & Technology Visualizations

Using a theoretically grounded visualization framework that defines key terminology and processes together with valid workflows and data mappings.





Places & Spaces: Mapping Science Exhibit, online at http://scimaps.org



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 ?

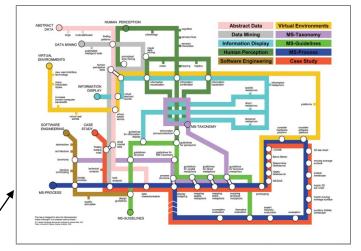




Different Question Types



Terabytes of data



Find your way

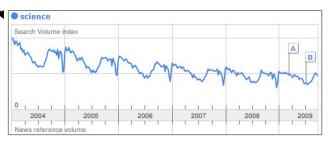
Descriptive &

Predictive

Models



Find collaborators, friends



Identify trends



Different Question Types

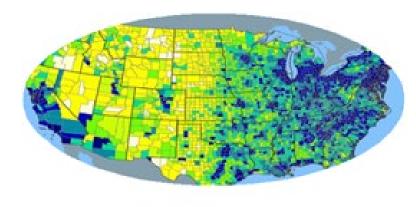
- Monitor Data Quality
- Customer Complaints
- Customer/Supplier/Learner Churn
- Optimizing Supplier Chains
- Improving (Traffic/Communication) Network Resilience
- Optimizing Traffic/Communication Flows
- Optimizing Work Cells/Built Process
- Workforce Development

And other WHEN, WHERE, WHAT, WITH WHOM questions.



Different Levels of Abstraction/Analysis

Macro/Global Population Level



Meso/Local Group Level The state of the s

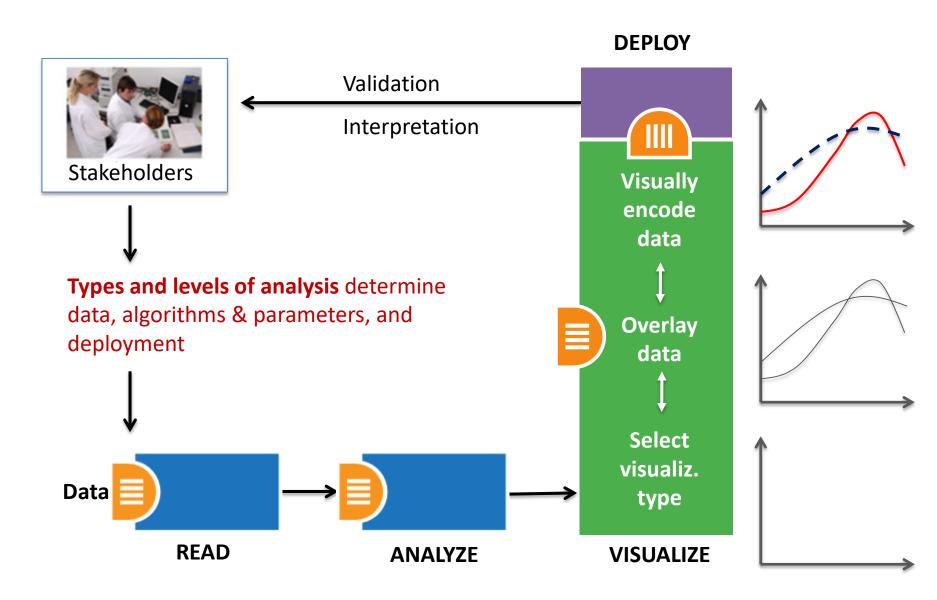
Micro Individual Level



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

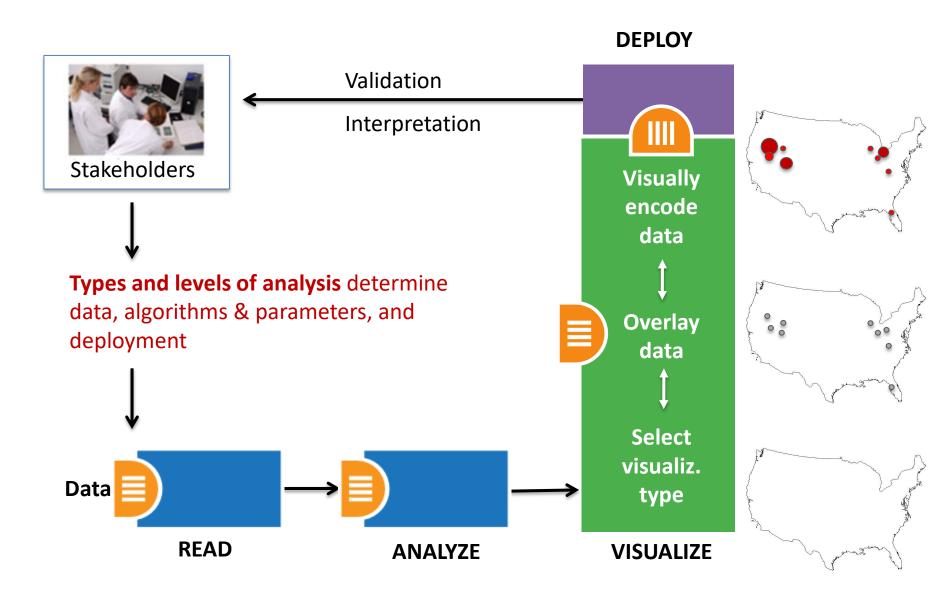


Needs-Driven Workflow Design





Needs-Driven Workflow Design





Visualization Framework (Vis-FW)

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

Visualization Framework

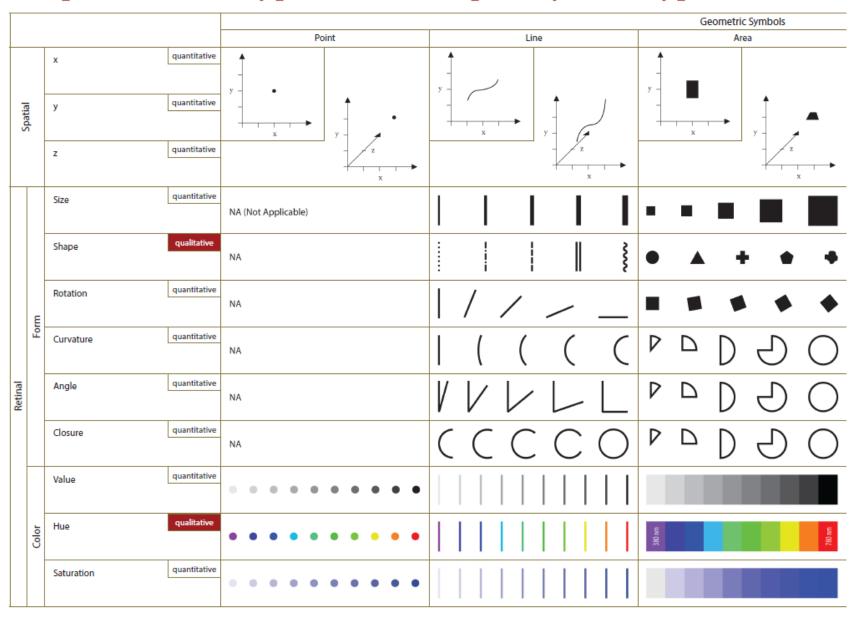
Basic Task Types								
Bertin, 1967	Wehrend & Lewis, 1996	Few, 2004	Yau, 2011	Rendgen & Wiedemann, 2012	Frankel, 2012	Tool: Many Eyes	Tool: Chart Chooser	Börner, 2014
selection	categorize			category				categorize/ cluster
order	rank	ranking					table	order/rank/ sort
	distribution	distribution					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

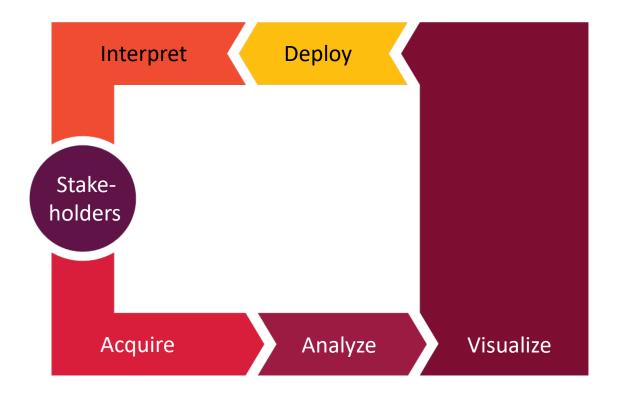
Insight Need Types page 26	Data Scale Types page 28	* *	Graphic Symbol Types page 32	-	Interaction Types page 26
 categorize/cluster order/rank/sort distributions (also outliers, gaps) comparisons trends (process and time) geospatial compositions (also of text) correlations/relationships 	nominal ordinal 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

Graphic Variable Types Versus Graphic Symbol Types

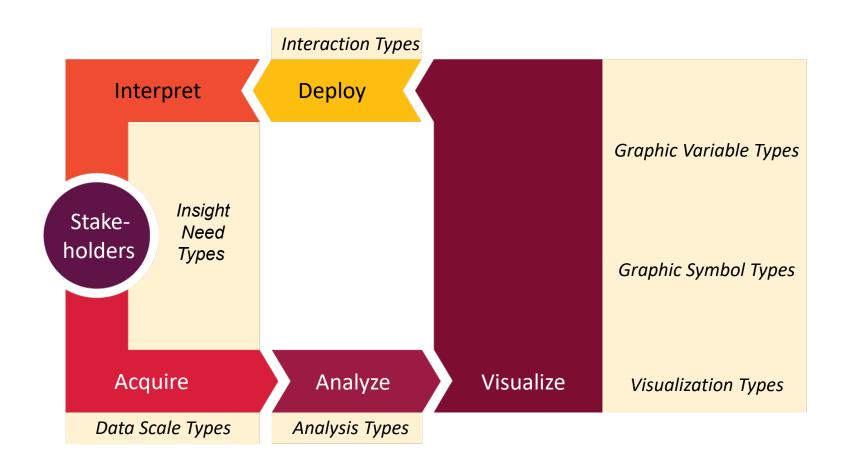


Graphic Variable Types Versus Graphic Symbol Types Geometric Symbols Linguistic Symbols Pictorial Symbols Text, Numerals, Punctuation Marks Images, Icons, Statistical Glyphs quantitative quantitative quantitative quantitative See Stepped Relief Map, pages 53-54 See Heights of the Principal Mountains ..., page 67 NA (Not Applicable) Text See also Life in Los Angele Text quantitative Rotation Text Text Angle Some table cells are left blank to encourage future exploration of combinations. quantitative Value quantitative quantitative Saturation Linguistic Symbols Pictorial Symbols Spacing quantitative Granularity Pattern quantitative Orientation quantitative Gradient Blur quantitative Transparency quantitative Shading quantitative Stereoscopic Depth background quantitative Speed 7,7 quantitative Rhythm Blinking point Blinking area Blinking volume Blinking text Blinking icons Blinking line Blinking surface

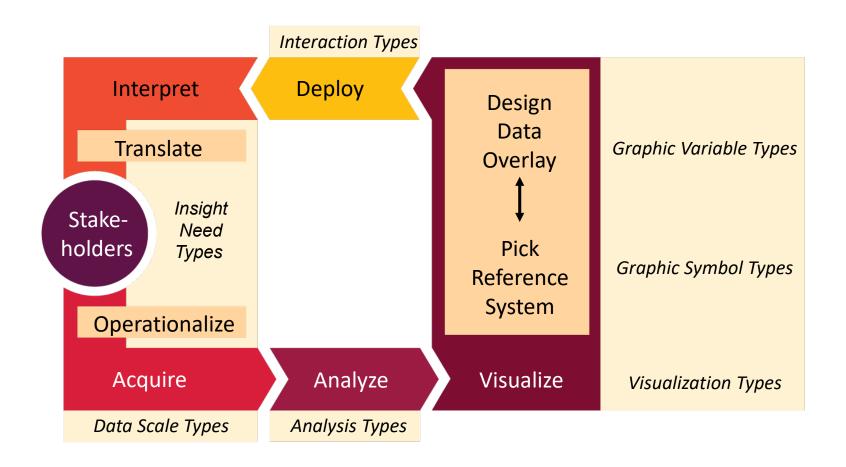
Workflow Design



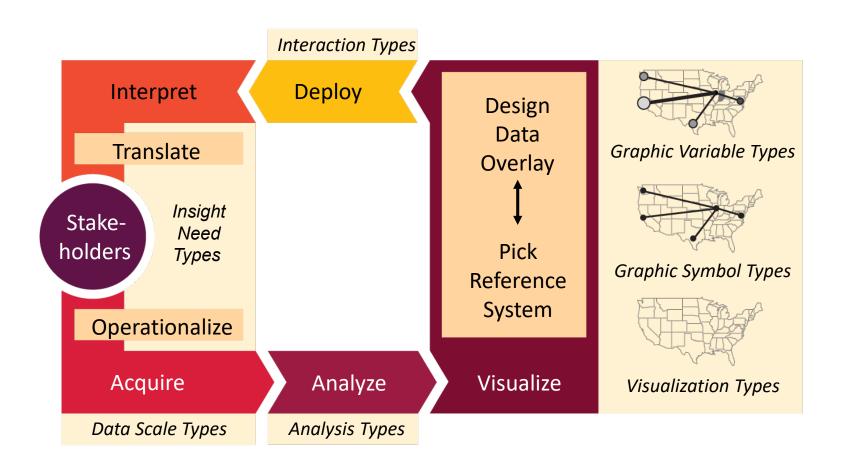
Workflow Design + Visualization Framework



Workflow Design + Visualization Framework



Workflow Design + Visualization Framework



Data Visualization Literacy: Definitions, Conceptual Frameworks, Exercises, and Assessments

Katy Börner¹, Andreas Bueckle¹, Michael Ginda¹

¹Indiana University

Submitted to Proceedings of the National Academy of Sciences of the United States of America

In the information age, the ability to read and construct data visualizations becomes as important as the ability to read and write text. However, while standard definitions and theoretical frameworks to teach and assess textual, mathematical, and visual literacy exist, current data visualization literacy (DVL) definitions and frameworks are not comprehensive enough to guide the design of DVL teaching and assessment. This paper introduces a conceptual framework (DVL-FW) that was specifically developed to define, teach, and assess DVL. The holistic DVL-FW promotes both the reading and construction of data visualizations, a pairing analogous to that of both reading and writing in textual literacy and understanding and applying in mathematical literacy. Specifically, the DVL-FW defines a hierarchical typology of core concepts and details the process steps that are required to extract insights from data. Advancing the state of the art, the DVL-FW interlinks theoretical and procedural knowledge and showcases how both can be combined to design curricula and assessment measures for DVL. Earlier versions of the DVL-FW have been used to teach DVL to more than 8,500 residential and online students, and results from this effort have helped revise and validate the DVL-FW presented here.

Data visualization | literacy | assessment | learning sciences

measurement, and estimation," as well as an "understanding of ratio concepts, notably fractions, proportions, percentages, and probabilities" (6). PISA defines it as "an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts," including "reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena." PISA administers standardized tests for math, problem-solving, and financial literacy (7). The PISA 2015 Draft Mathematics Framework (8) explains the theoretical underpinnings of the assessment, the formal definition of mathematical literacy, the mathematical processes which students undertake when using mathematical literacy, and the fundamental mathematical capabilities that underlie those processes.

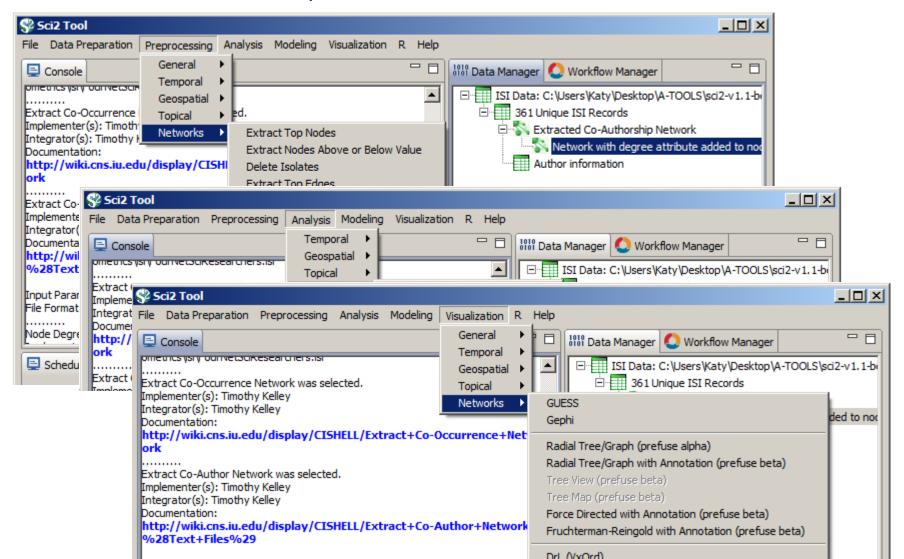
Visual literacy was initially defined as a person's ability to "discriminate and interpret the visible actions, objects, and symbols natural or man-made, that he encounters in his environment" (9). In 1978, it was defined "as a group of skills which enable an individual to understand and use visuals for intentionally communicating with others" (10). More recently, the Association of College and Research Libraries (ACRL) defined standards, performance indicators, and learning outcomes for visual literacy (11, 12). In the academic setting, Avgerinou (13) developed and validated a visual literacy index by running focus groups of visual

PNAS, 2019



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	COMMUNICATI ONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA	CTS-CLINICAL AND TRANSLATIONA L 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		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	COMMUNICATI ONS OF THE ACM	Plug-and-Play Macroscopes	Computer Science	Borner, K
18	2010	MALDEN	USA		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	TRANSLATIONA	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



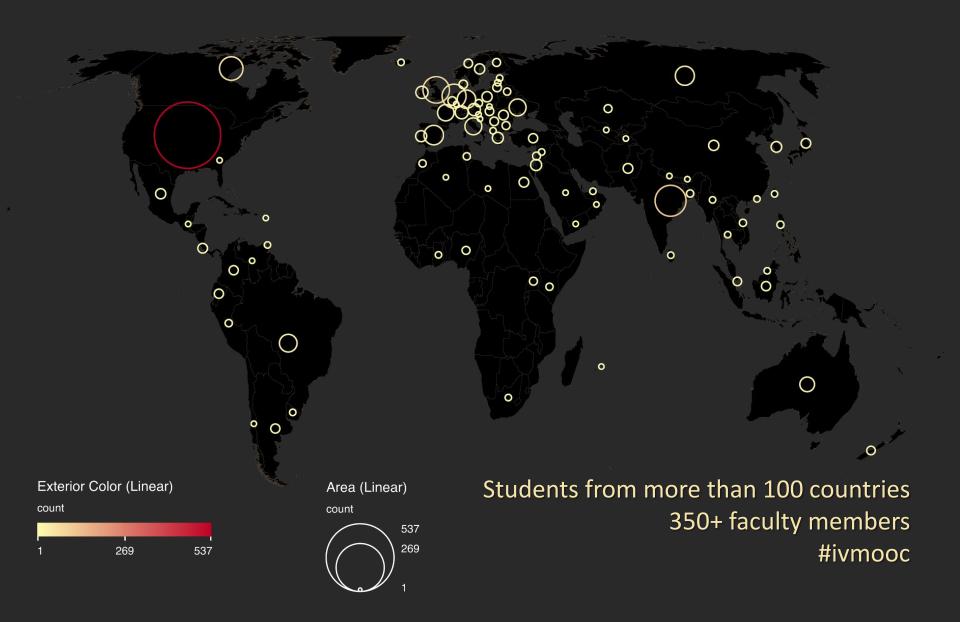
Co-author and many other bi-modal networks.



Register for free: http://ivmooc.cns.iu.edu. Class restarts Jan 7, 2019.



The Information Visualization MOOC ivmooc.cns.iu.edu



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

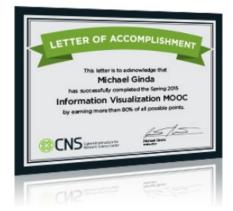
Mid-Term

- 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%).

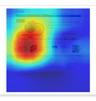






Client Projects





Visualizing the Evolution of Website Design

With over 25 years of history, the web itself has become a significant cultural artifact. We are studying how website design has changed over time, and how these changes reflect changes

Read more...



Visualizing Research Silos in Ecological Interaction datasets

Open access to high quality and integrated evaluated detects is important to better

Read more...



ChaCha Menopause queries

The ChaCha menopause query data is the foundation for building intervention modules to improve people's knowledge and problem solving skills related to menopause. For this project

Read more...



Text-Mining of User-Generated Queries on Menstrual Pain

Maneteral nate to highly appropriate amount against a frame direction and District estates of time

Read more...



BioSimmer

BioSim is a participatory simulation where young students (grades K-3) enact the roles of ants and biological systems through the assistance of electronically-enhanced e-puppets. It is

Read more...

2018 projects are at https://ivmooc.cns.iu.edu/clients.html
Submit your own 2019 client project via FORM by Dec 6, 2018.

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

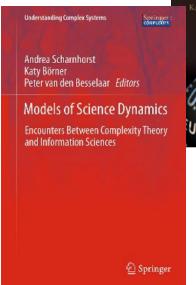
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

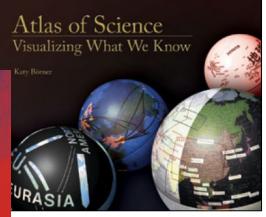
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

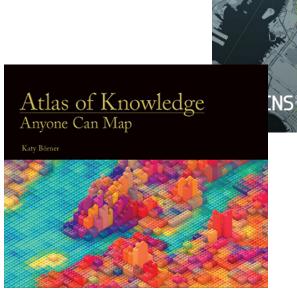
Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) VIVO: A Semantic Approach to Scholarly Networking and Discovery. Morgan & Claypool.

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







Interdisciplinary Training in **Complex Networks and Systems**

RESOURCES / TWITTER



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.



https://cns-nrt.indiana.edu

Open Positions at CNS

- Postdoc as part of NIH: The Human Body Atlas
 https://commonfund.nih.gov/HuBMAP
 https://projectreporter.nih.gov/project_info_description.cfm?aid=9687220
- Postdoc as part of NSF: Collaborative Research:
 Multi-Level Graph Representation for Exploring Big
 Data (with Stephen Kobourov)
 https://www.nsf.gov/awardsearch/showAward?AW
 D_ID=1839167



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

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

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