CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Visual Analytics in Support of Data-Driven Decision Making

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

Knowledge Network Talk, The Boeing Company

***February** 4, 2020*

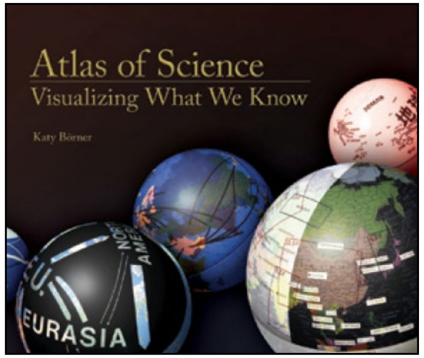
Overview

Intro to Data-Driven Decision Making Using (Interactive) Maps

- *Places & Spaces: Mapping Science* exhibit (http://scimaps.org).

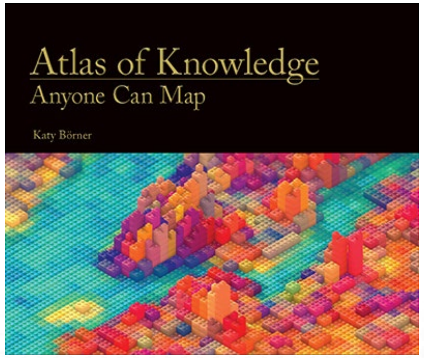
Data Visualization Literacy (DVL)

- Börner, Katy, Andreas Bueckle, and Michael Ginda. 2019. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *PNAS*, 116 (6) 1857-1864.
- Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press.
- Börner, Katy. 2010. *Atlas of Science: Visualizing What We Know*. Cambridge, MA: The MIT Press.



Visual Analytics Certificate

- Learn to render data into actionable insights in 6 weeks! Class begins **March 2, 2020**. Get course info at https://boeing.cns.iu.edu



Learning Analytics

- Ginda, Michael, Michael C. Richey, Mark Cousino, and Katy Börner. 2019. "Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design". *PLOS One* e0215964. doi: 10.1371/journal.pone.0215964.

CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Data-Driven Decision Making Using (Interactive) Maps

http://scimaps.org

Data-Driven Decision Making

Most decision makers prefer

- orderly, predictable conditions,
- little disruption, and
- sufficient resources (e.g., money, talent, compassion)

to invent and implement desirable futures.

They want to understand the likely impact of decisions (e.g., hiring, purchasing, strategy changes) BEFORE writing a check and/or starting implementation.

Data Access and Actionable Visualizations

Decision makers have a deep interest in—and are willing to pay for—

- easy-to-use,
- near-real-time access to
- data, models, and visualizations

that help them make sense of, communicate, and proactively manage science, technology, and education and increase return on investment (ROI).

Visualizations that show the structure and dynamics of science, technology, and education are actively being researched and developed.

Use Cases: Industry

- Determine how to utilize limited resources to increase innovation and labor productivity; improve inventory turnover, asset utilization, supply-chain management, traffic optimization, error and attack tolerance.
- Research collaboration and workforce development decisions require knowing where the most productive research is being done and the best experts are trained, as well as how that production has changed over time and across individuals and institutions.
- Data-driven strategic planning, hiring, and resource allocation.

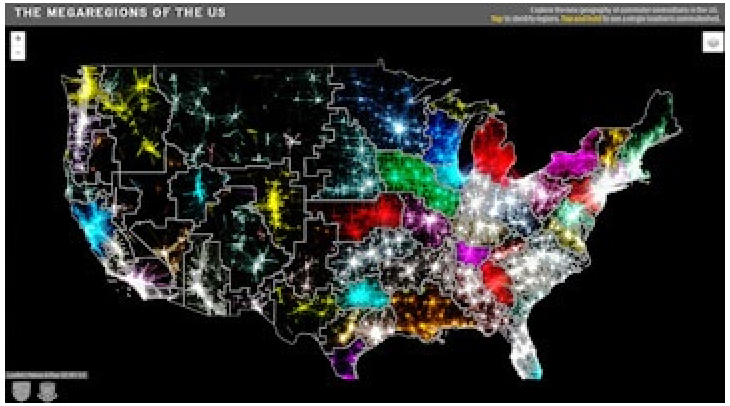
Börner, Katy, William B. Rouse, Paul Trunfio, and H. Eugene Stanley. 2018. "Forecasting Innovations in Science, Technology, and Education." *PNAS* 115(50): 12573-12581.

Use Cases: Government

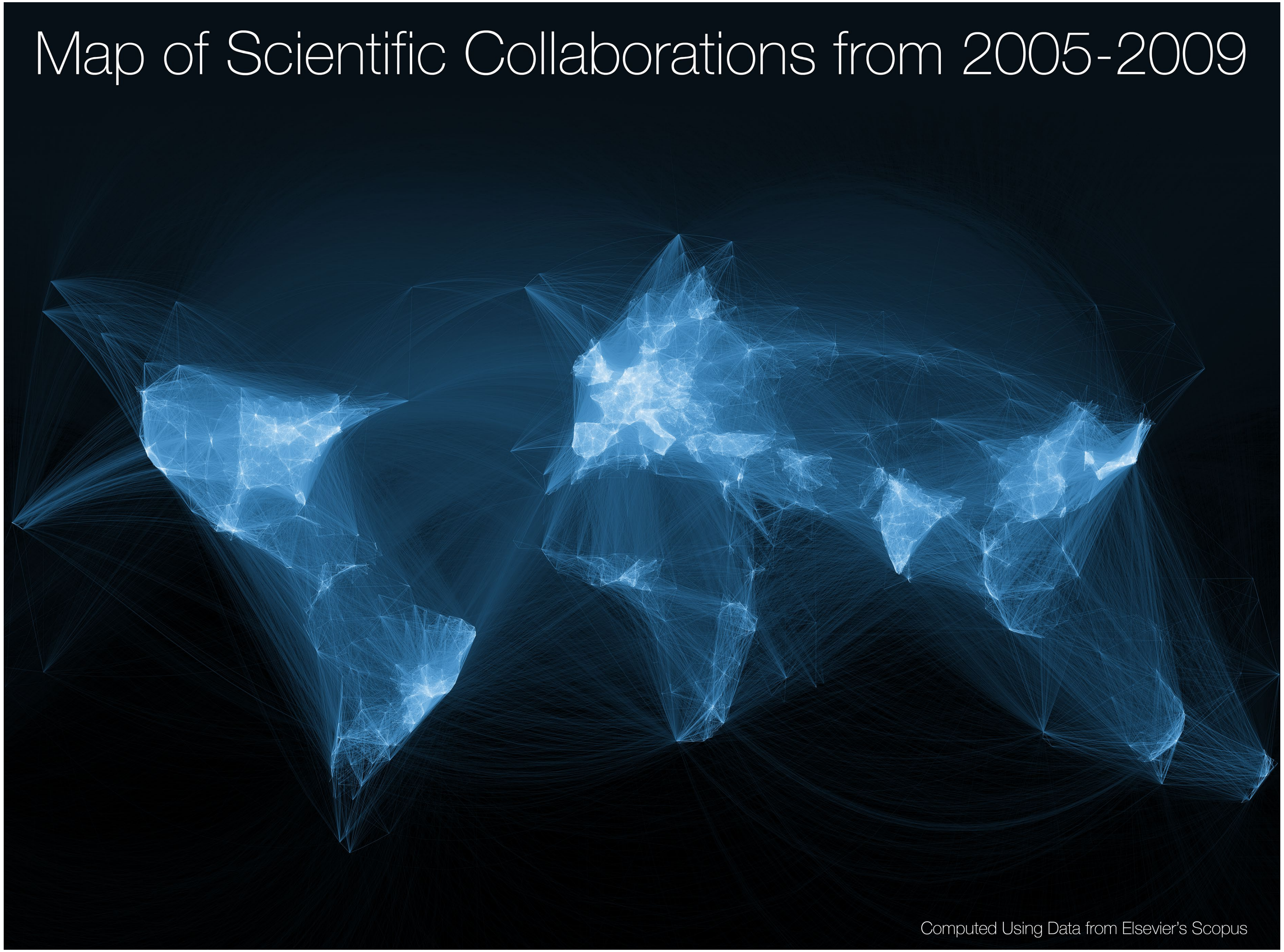
- *The Promise of Evidence-Based Policymaking* report from 2017 opens with this statement: "The American people want a government that functions efficiently and responsibly addresses the problems that face this country." The report argues for a future in which "rigorous evidence is created efficiently, as a routine part of government operations, and used to construct effective public policy."
- Short-term goal: Improve access to data to improve the quantity and quality of evidence that informs important program and policy decisions.
- Longer-term goal: Simulate the impact of different policy decisions (e.g., alternative retirement age, funding schemas, tax rates) BEFORE they are implemented.

Use Cases: Academia

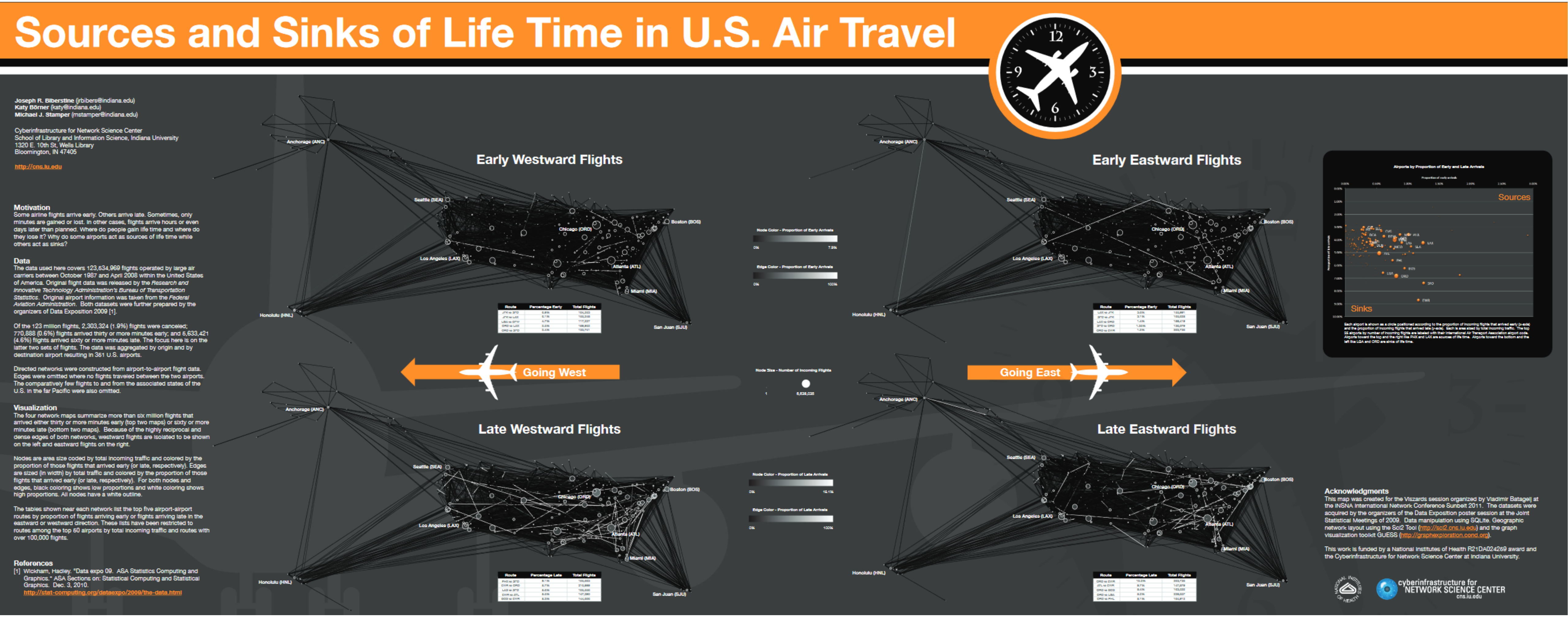
- Evaluate the scientific impact of scholars, journals, academic institutions, or nations.
- Quantify and predict scientific research, impact, and outcomes.
- Support the selection of candidate faculty members by universities.
- Identify the best reviewers.
- Prioritize the development of research fields in which a country or region should invest.



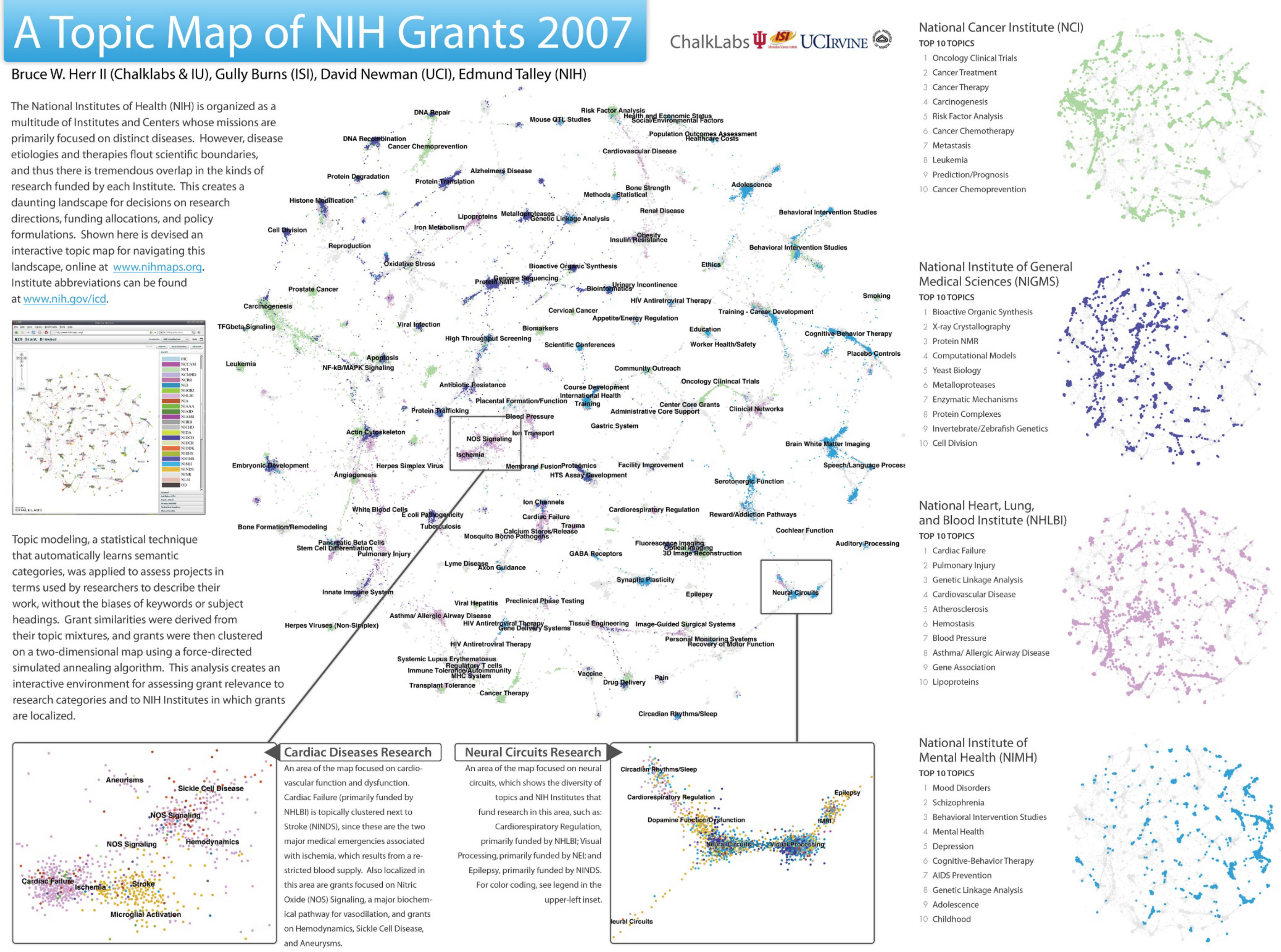
Börner, Katy, William B. Rouse, Paul Trunfio, and H. Eugene Stanley. 2018. "Forecasting Innovations in Science, Technology, and Education." *PNA* 115(50): 12573-12581.



VII.6 Stream of Scientific Collaborations Between World Cities - Olivier H. Beauchesne - 2012



https://cns.iu.edu/docs/research/visualizations/CNS_Sources_and_Sinks_of_Lifetime_Sunbelt_Viszards_Poster_2011.pdf



V.7 *A Topic Map of NIH Grants 2007* - Bruce W. Herr II, Gully A.P.C. Burns, David Newman, and Edmund Talley - 2009

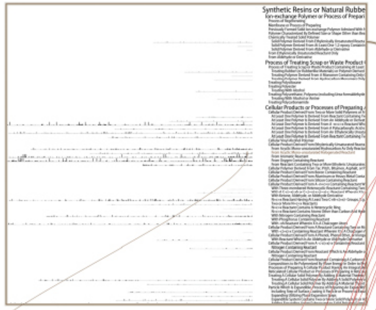
The US Patent Hierarchy

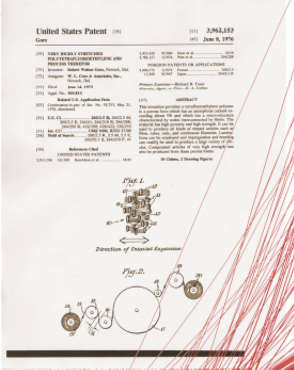
Impact

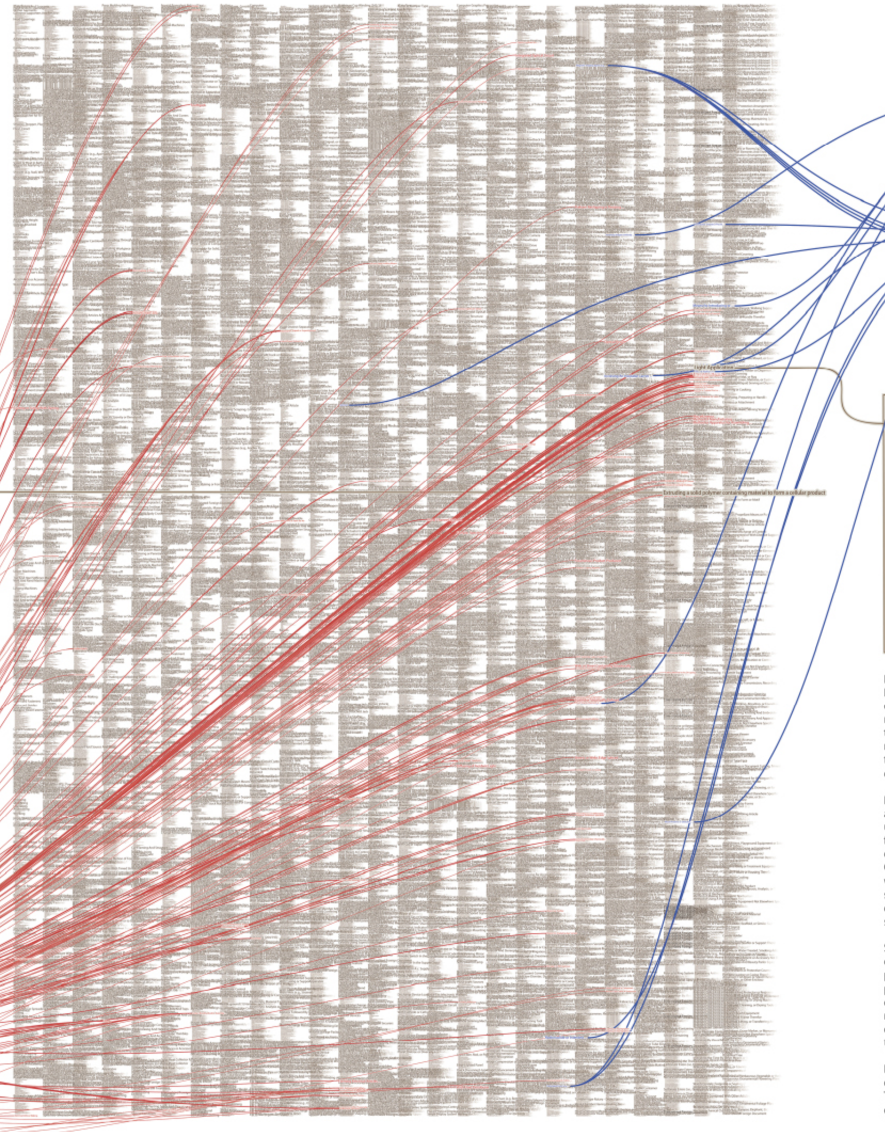
The United States Patent and Trademark Office does scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that groups patents by industry or use, proximate function, effect or product, and structure. At the time of this writing there are 160,523 categories in a hierarchy that goes 15 levels deep. We display the first three levels (13,529 categories) at right in what might be considered a textual map of inventions.

Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

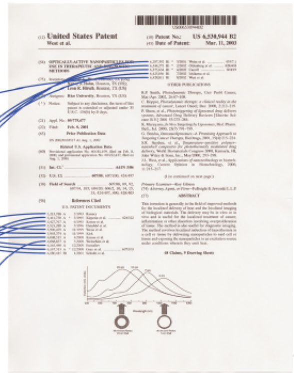
The patent on Goretex—a lightweight, durable synthetic fiber—is an example of one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1981 to 2006) point to the 130 categories that contain 182 patents, from waterproof clothing to surgical cosmetic implants, that mention Goretex as "prior art."

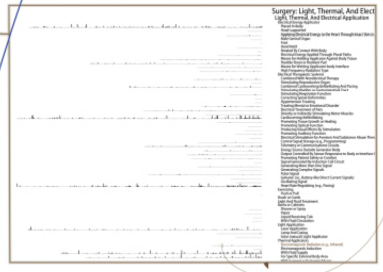






Prior Art





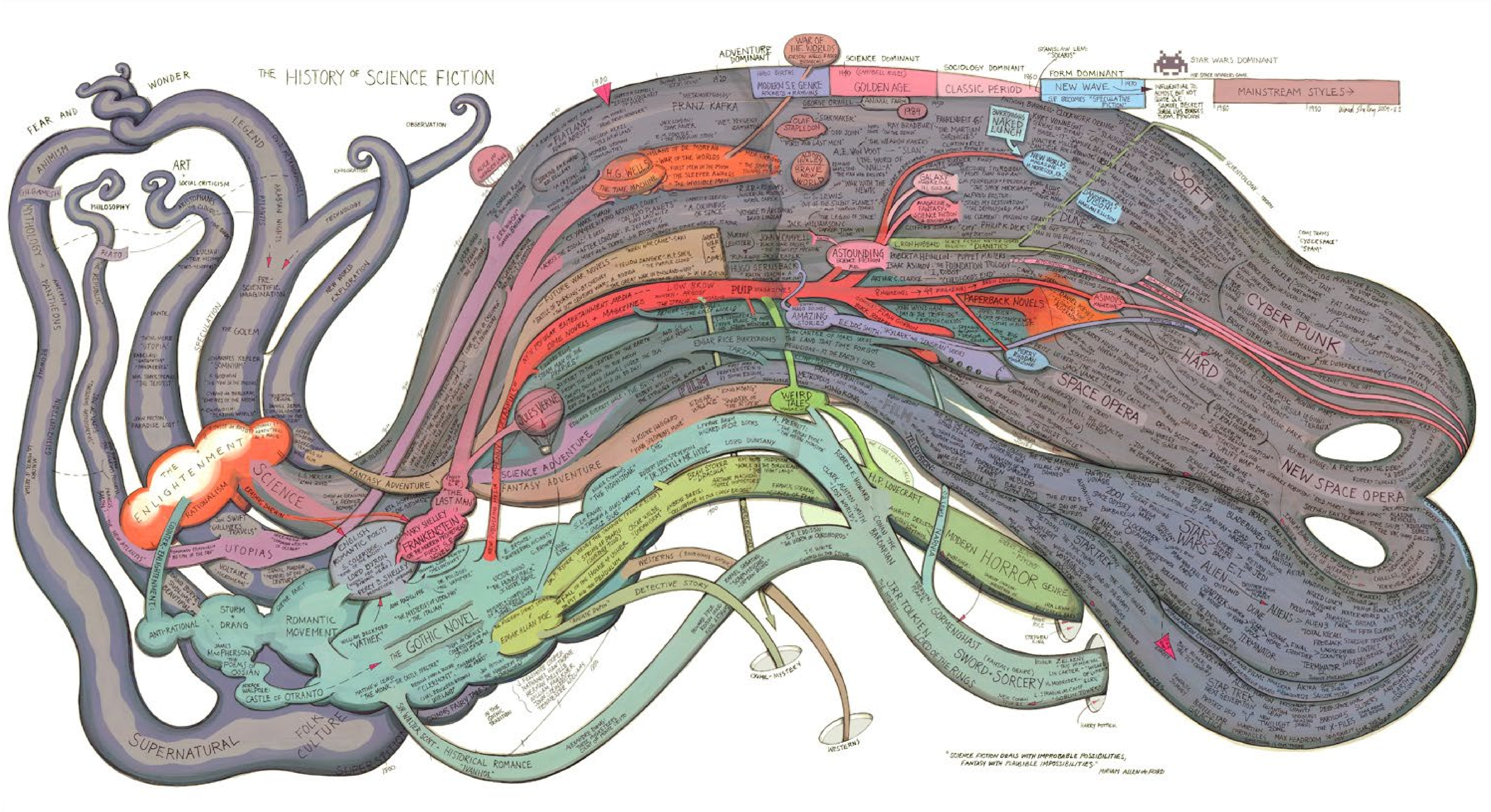
New patents often build on older ideas from many different categories. Here, blue lines originate in the sixteen categories that contain patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny gold spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infra-red scans; they have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for this invention.

Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements that comfortably fit the definition of the category. The box above shows tiny bar charts, part of a *Taxonomy Validator* that reveals whether elements fit their categories. Categories may need to be redefined, and sometimes need to be split when they get too vague or large; a problem shared by many classification systems in this information-rich century. But how can we tell which ones to eliminate, add or revise—or how to revise them—in the complex, abstract sociolinguistic spaces we partition into ontologies?

Something as simple as a bar chart helps people see how entities in a category relate to that category. Here, each bar encodes a "distance to prototype": how much each patent differs from an idealized "prototype patent" for that category. A measure like this can be based on statistics, computational linguistics, or even human insight. Thus a category with mostly small bars is a good one, and a generally ragged one needs scrutiny or reorganization; but one that has only two or three tall bars may mean that only those few elements don't belong.

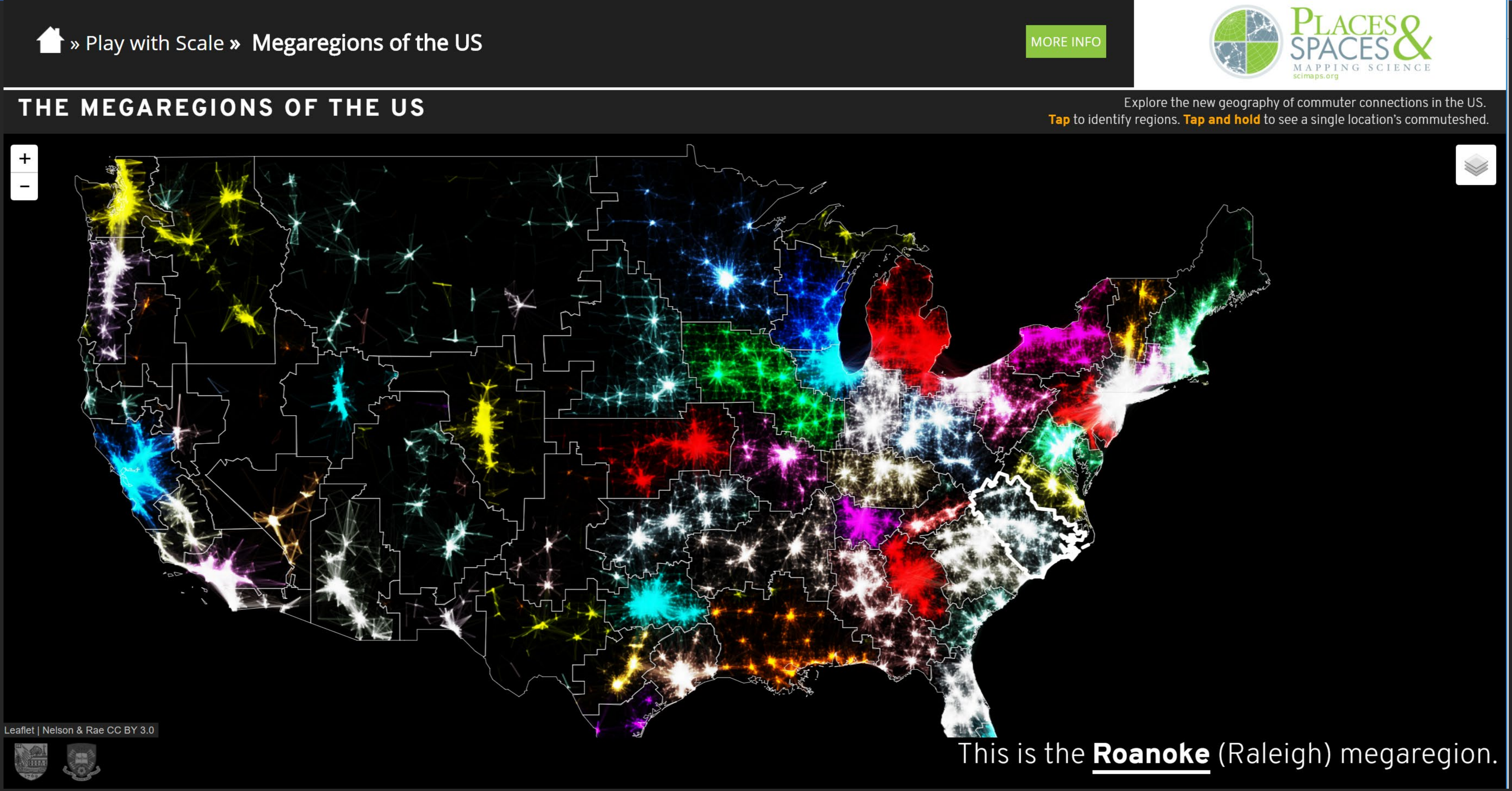
Even simple visuals can make thinking easier by providing better distilled data to the eye: vastly more data than working memory can hold as words. They focus people on exactly the right issues, and support them with the comprehensive overviews they need to make more informed judgements.

II.8 *Taxonomy Visualization of Patent Data* - Katy Börner, Elisha F. Hardy, Bruce W. Herr II, Todd Holloway, and W. Bradford Paley - 2006



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


MACROSCOPES FOR INTERACTING WITH SCIENCE
PLACES & SPACES
MAPPING SCIENCE
scimaps.org
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

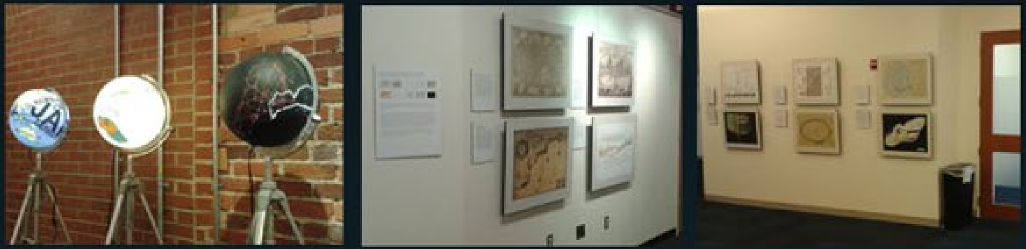


Megaregions of the US –Garrett Dash Nelson and Alasdair Rae – 2016

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

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

http://scimaps.org

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

CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Data Visualization Literacy Framework

Börner, Katy, Andreas Bueckle, and Michael Ginda. 2019. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *PNAS*, 116 (6) 1857-1864.

Data Visualization Literacy (DVL)

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

- literacy (ability to read and write text in titles, axis labels, legends, etc.),
- visual literacy (ability to find, interpret, evaluate, use, and create images and visual media), and
- mathematical literacy (ability to formulate, employ, and interpret math in a variety of contexts).

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 to strategically approach local and global issues.

DVL Framework: Desirable Properties

- Most existing frameworks focus on **READING**. We believe that much expertise is gained from also **CONSTRUCTING** data visualizations.
- Reading and constructing data visualizations needs to take human perception and cognition into account.
- Frameworks should build on and consolidate prior work in cartography, psychology, cognitive science, statistics, scientific visualization, data visualization, learning sciences, etc. in support of a de facto standard.
- Theoretically grounded + practically useful + easy to learn/use.
- Highly modular and extendable.

DVL Framework: Development Process

- The initial DVL-FW was developed via an extensive literature review.
- The resulting DVL-FW typology, process model, exercises, and assessments were then tested in the *Information Visualization* course taught for more than 17 years at Indiana University. More than 8,500 students enrolled in the IVMOOC version (http://ivmooc.cns.iu.edu) over the last six years.
- The FW was further refined using feedback gained from constructing and interpreting data visualizations for 100+ real-world client projects.
- Data on student engagement, performance, and feedback guided the continuous improvement of the DVL-FW typology, process model, and exercises for defining, teaching, and assessing DVL.
- The DVL-FW used in this course supports the systematic construction and interpretation of data visualizations.

Data Visualization Literacy Framework (DVL-FW)

Consists of two parts:

DVL Typology

Defines 7 types with 4-17 members each.

1 Insight Needs
- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/ relationships

2 Data Scales
- nominal
- ordinal
- interval
- ratio

3 Analyses
- statistical
- temporal
- geospatial
- topical
- relational

4 Visualizations
- table
- chart
- graph
- map
- tree
- network

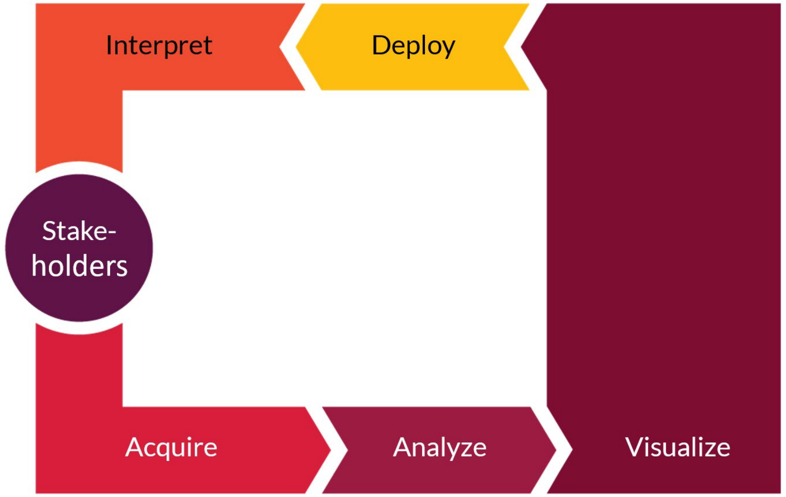
5 Graphic Symbols
- geometric symbols
  point
  line
  area
  surface
  volume
- linguistic symbols
  text
  numerals
  punctuation marks
- pictorial symbols
  images
  icons
  statistical glyphs

6 Graphic Variables
- spatial
  position
- retinal
  form
  color
  optics
  motion

7 Interactions
- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

DVL Workflow Process

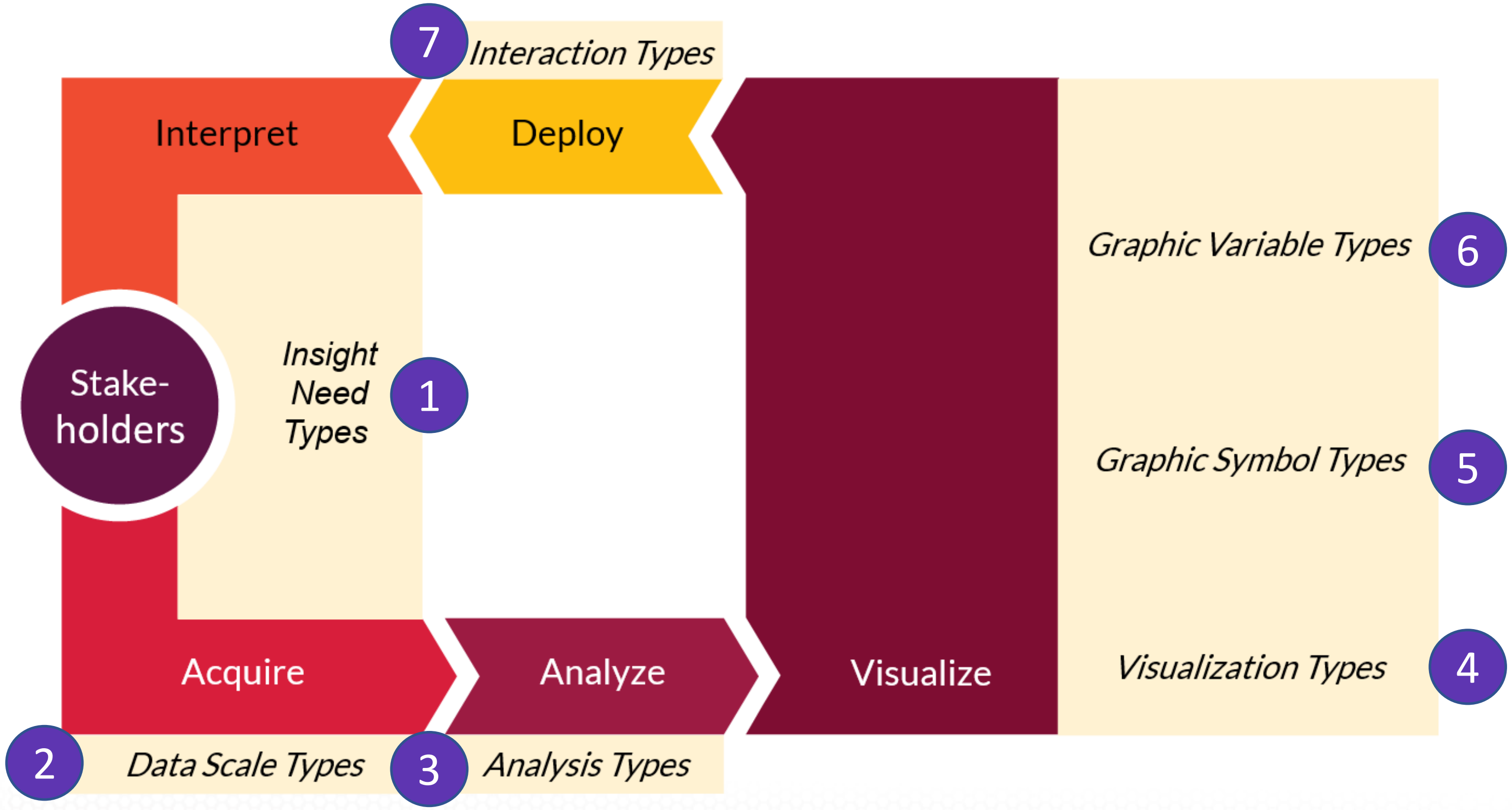
Defines 5 steps required to render data into insights.



Data Visualization Literacy Framework (DVL-FW)

Consists of two parts that are interlinked:

DVL Typology +
DVL Workflow Process



CNS Cyberinfrastructure for Network Science Center

Typology of the Data Visualization Literacy Framework

1 Insight Needs
- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/ relationships

2 Data Scales
- nominal
- ordinal
- interval
- ratio

3 Analyses
- statistical
- temporal
- geospatial
- topical
- relational

4 Visualizations
- table
- chart
- graph
- map
- tree
- network

5 Graphic Symbols
- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

6 Graphic Variables
- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

7 Interactions
- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 25.

Typology of the Data Visualization Literacy Framework

1

Insight Needs	Data Scales	Analyses	Visualizations	Graphic Symbols	Graphic Variables	Interactions
• 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	• statistical • temporal • geospatial • topical • relational	• table • chart • graph • map • tree • network	• 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	• zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion

Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 26-27.

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

Typology of the Data Visualization Literacy Framework

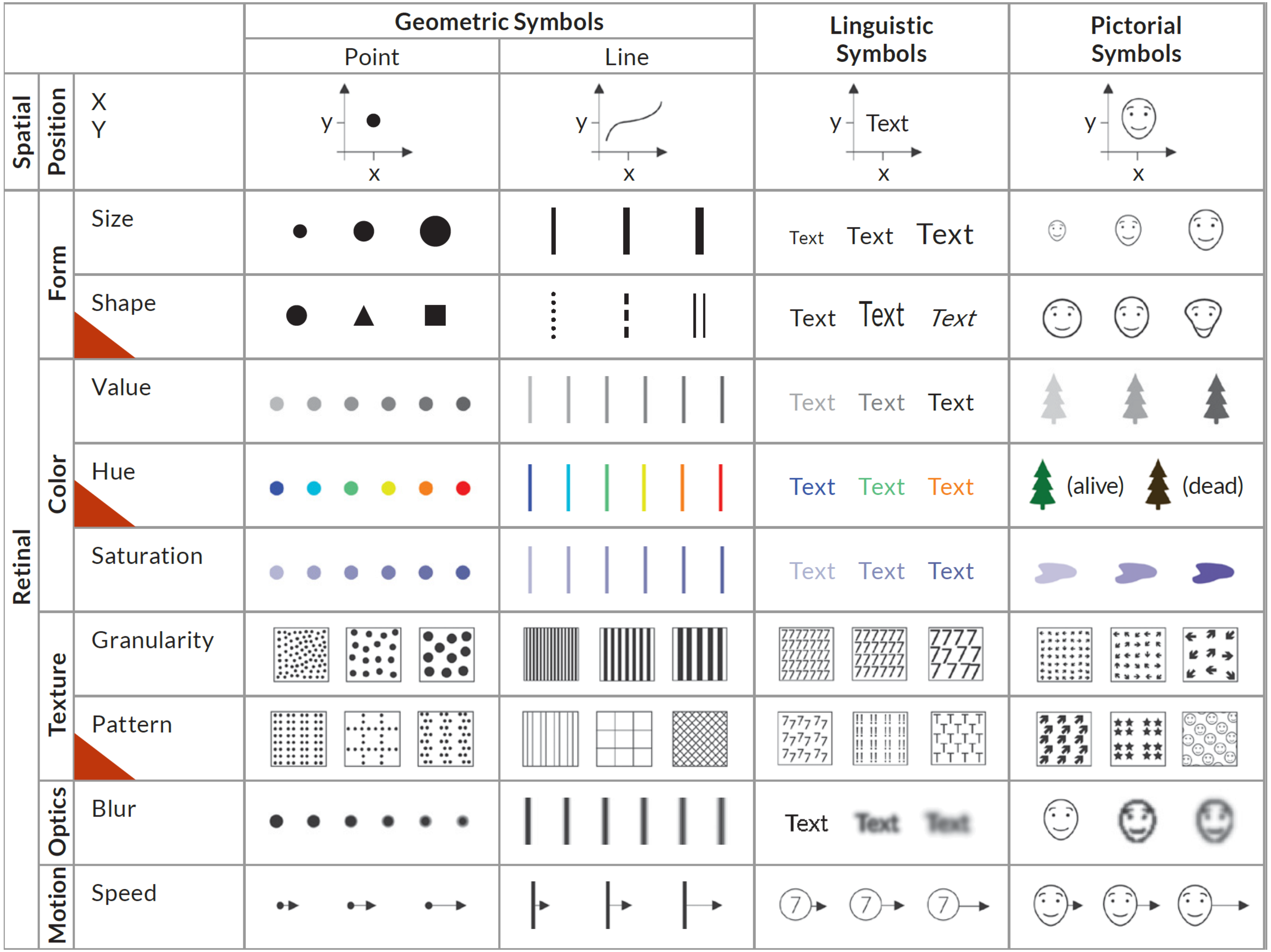
6

Insight Needs	Data Scales	Analyses	Visualizations	Graphic Symbols	Graphic Variables	Interactions
• 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	• statistical • temporal • geospatial • topical • relational	• table • chart • graph • map • tree • network	• 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	• zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion

Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 34-35.

Graphic Symbol Types

Graphic Variable Types



See *Atlas of Knowledge* pages 36-39 for complete table.

Qualitative

Also called:

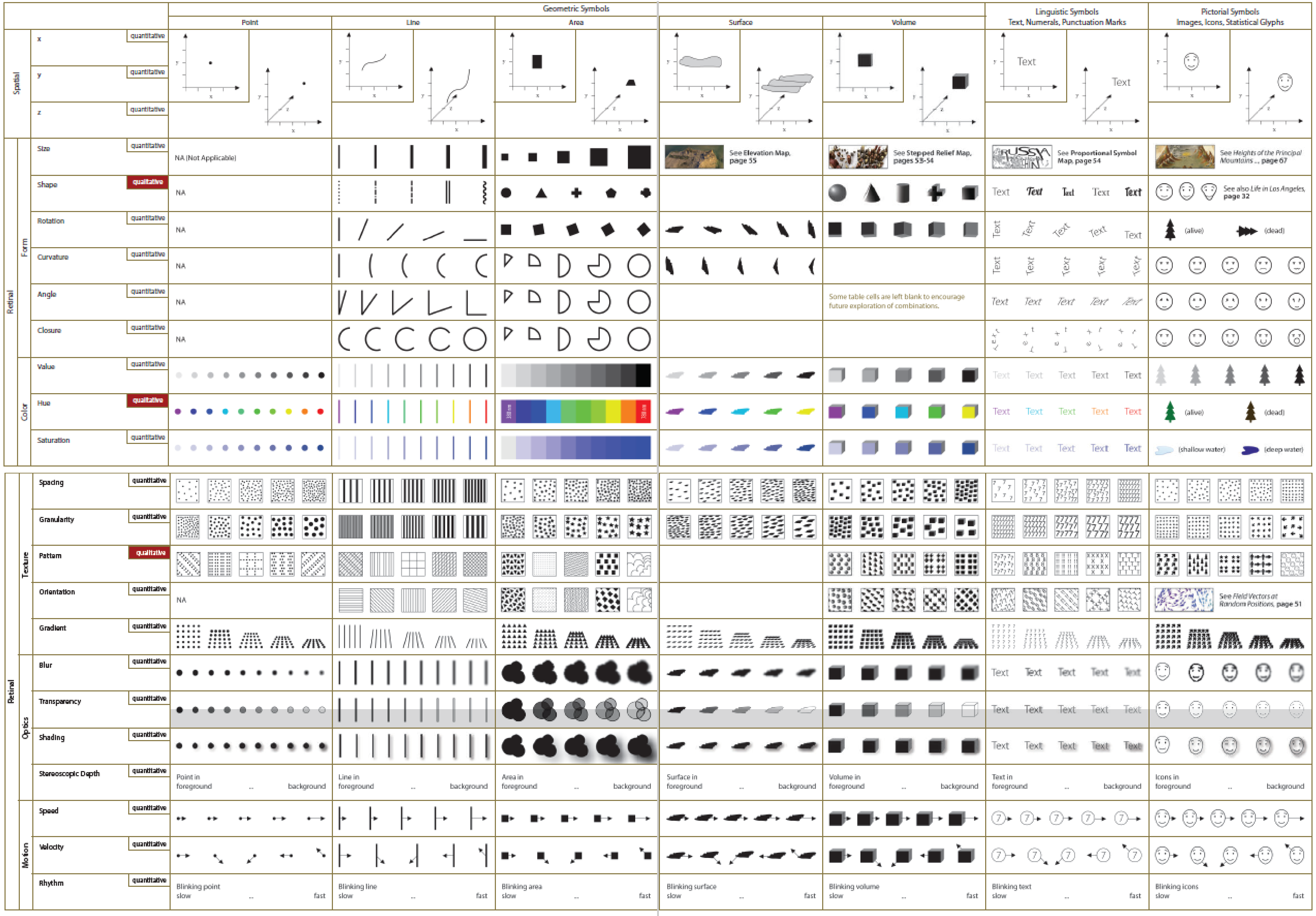
Categorical Attributes
Identity Channels

Quantitative

Also called:

Ordered Attributes
Magnitude Channels

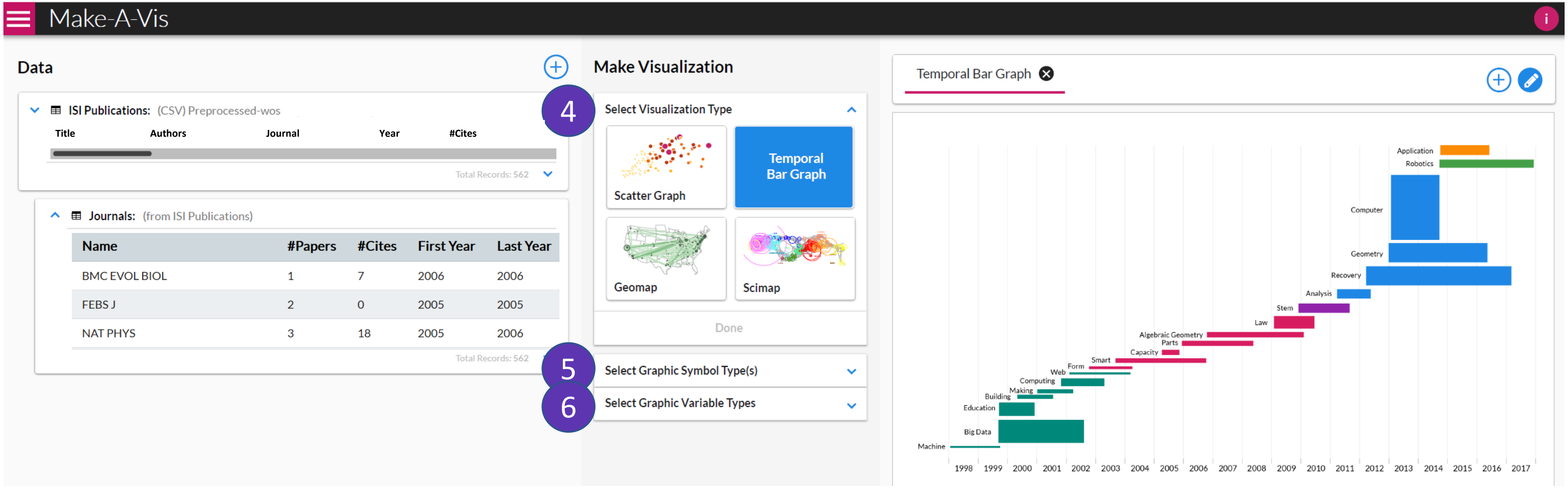
Graphic Variable Types Versus Graphic Symbol Types



See *Atlas of Knowledge* pages 36-39 for complete table.

Data Visualization Literacy Framework (DVL-FW)

Implemented in Make-A-Vis (MAV) to support learning via horizontal transfer, scaffolding, hands-on learning, etc.



CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Visual Analytics Certificate

https://boeing.cns.iu.edu

Visual Analytics Certificate

Duration:	6 weeks x 5 hours = 30 hours (3 CEUs)
Format:	Online using Canvas
Type:	Executive Education

Learn to render data into actionable insights in 6 weeks!
Class begins March 2, 2020. Get course info at https://boeing.cns.iu.edu

For US-based Boeing employees:

Begin the enrollment process by completing the pre-enrollment form (https://iu.co1.qualtrics.com/jfe/form/SV_ekXpn97sf6gvZAx). NOTE: You will need your voucher to complete the form. Over the days following submission, look for emails about next steps and course information to finalize your enrollment.

For non-US based Boeing employees and others not using the voucher system:

Register for the course online through Indiana University's IU Expand portal. Go to: https://expand.iu.edu/browse/sice/cns/courses/visual-analytics-boeing-2. Payment is via credit card.

Instructors

Katy Börner
Instructor

Victor H. Yngve Distinguished Professor of Engineering and Information Science at the School of Informatics, Computing, and Engineering. Founding Director of the Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana University.

- Research focus on development of data analysis and visualization techniques for information access, understanding, and management.
- Cyberinfrastructures development for large-scale scientific collaboration and computation.

Michael Ginda
Assistant Instructor

Data analyst and research assistant with the Cyberinfrastructure Center for Network Science. He holds a Master's degree in Library Science from Indiana University.

- Research focus on knowledge representation and organization, metadata, and information networks.
- Lead instructional designer.

Andreas Bueckle
Assistant Instructor

PhD student in Information Science at Indiana University focused on information visualization.

- Research focus on information visualization, specifically interactive and augmented reality.
- Videography and photography.

Schedule

- Week 1 – Introduction to Visual Analytics

- Week 2 – When: Temporal Data Analysis and Visualization
- Week 3 – Where: Geospatial Data Analysis and Visualization
- Week 4 – What: Topical Data Analysis and Visualization
- Week 5 – With Whom: Network Analysis and Visualization

- Week 6 – Value Creation via Data-Driven Decision Making and Future Developments

Schedule

- Highly structured course design.
- Each week features the same modules that build on each other to support **scaffolding**.
- Materials are designed to support **transfer**—building on previous knowledge and expertise to promote deeper levels of learning.
- Students immediately apply theory to gain hands-on knowledge and expertise.
- Social learning is supported.

Scaffolding

Theory | Hands-on

Transfer

Pre | Week 1 | Week 2 | Week 3 | Week 4 | Week 5 | Week 6 | Post

Week	Theory	Atlas 2 *(Low Floor)*	Optional Reading *(High Ceiling)*	Flashcards, DVL Typology+Process *(Horizontal Transfer)*	Hands-On MAV	MAV-Task *(Scaffolding)*	Demo	T-Translate *(Analogical Transfer)*	Your Project *(Create Your Own Visualization)*	Participation *(Social Learning)*
Supp. Doc.	18-IVC-T-Slides-TOC**.docx	Give full citations	https://cns.iu.edu/publications		MAV-Workflows-Run1**.docx	IVC-Data+Code**.docx MAV-Workflows-Run1**.docx				
Pre	**Welcome** • V-Katy and Andreas, VIS wall shows GDELT • Web content similar to https://visanalytics.cns.iu.edu and IUExpand • Links to Case Studies like on https://visanalytics.cns.iu.edu • Q-Entrance-Survey: Why do you take this course?									D-Introduce yourself
1	**Data-Driven Decision Making, Visual Analytics Framework, Workflow Design** • V-Overview (VIS wall shows A2 DVL FW table) • V-Data-Driven Decision Making • V-DVL-FW ❖ Data Scale Types ❖ Visualization Types ❖ Graphic Symbol Types – Geometric Symbols ❖ Graphic Variable Types • V-Workflow Design	21-39	PNAS-DVL	50 flashcard vis in 50 secs animation	Introduction to Make-a-Vis Intro to ISE WOS and NSF Data, see MAV-Workflows-Run1**.docx Scatter Graph of cites over year, see 4.1 in MAV-Workflows-Run1**.docx but use AI data	Students see a perfect vis and are asked to make minor changes. Also explain logging via YML files, https://github.com/cns-iu/make-a-vis/tree/develop/projects/dvl-fw/src/lib/examples	Introduction to CNS	There exist many tools	Identify key insight needs in your organization, e.g., **Case Studies:** • Monitor S&T Dev • Manage Communication Flows • Optimize Traffic Flows • Manage Customer Feedback • Develop Workforce Sketch possible visualization.	D-MAV Help Discussion Forum D-Post your comments on A3 and Optional Reading. D-Post your favorite and most insightful vis, and brief explanation why it is important to you. Rate insightfulness of two vis posted by others.
2	**"When": Temporal Data Analysis and Visualization** • V-Exemplary Visualizations (Fleetmen) • V-Terminology • V-DVL Framework ❖ Insight Need Types • V-Workflow Design ❖ V-Trends & Seasonality ❖ V-Burst Detection & Visualization ❖ V-Types of Dynamics • User Needs Analysis • Algorithm Details: Burst Analysis and Visualization	40-51	PNAS-Topic Bursts	**Flashcards:** Charts and Graphs—identify correct name and type **DVL-Typology:** Given some insight needs, identify others **DVL-Process:** Fill in all steps.	Temporal Bar Graph	Students see an imperfect vis, are asked to make minor changes. NSF AI funding over years.	Boeing Learning Trajectories = Develop Workforce	Excel	Identify key temporal insight needs in your organization. Sketch possible visualization.	D-Post your comments on A3 and Optional Reading. D-Post your favorite and most insightful temp vis or tool. Rate insightfulness of two temp vis/tool posted by others.
3	**"Where": Geospatial Data Analysis and Visualization** • V-Exemplary Visualizations (Megaregions) • V-Terminology • V-DVL Framework & Terminology ❖ Graphic Symbol Types – Ling+Pictorial Symbols ❖ Color ❖ Interactivity Types • V-Color • V-Workflow Design ❖ Proportional Symbol Maps ❖ Choropleth Maps ❖ Traffic Network Maps • V-Algorithm Details: Geocoding and Aggregation	52-53	Food Web	**Flashcards:** Maps—identify correct name and type **DVL-Typology:** Given some graphic variables, identify others **DVL-Process:** Given steps, fill in all types.	US Map	Students see an imperfect vis, are asked to make major changes. ISE AI co-author network—Adam is rendering it with the new data.	ERC Geospatial Map with Co-Author Network = Manage Traffic / Communication Flows	Tableau	Identify key geo insight needs in your organization. Sketch possible visualization. Explain data variables to graphical variables mapping.	D-Post your comments on A3 and Optional Reading. D-Post your favorite and most insightful geo vis or tool. Rate insightfulness of two geo vis/tool posted by others.
4	**"What": Topical Data Analysis and Visualization** • V-Exemplary Visualizations (SmellyMaps) • V-Terminology • V-DVL Framework ❖ Data Layers • V-Workflow Design ❖ Wordle ❖ UCSD Map • V-Algorithm Details: UCSD Map Visualization	56-59	UCSD paper Text vs. Linkage comparison paper	**Flashcards:** Charts, Graphs, Maps—identify correct name and type **DVL-Typology:** Given 3-5 Flashcards, identify what graphic symbols and variables are used **DVL-Process:** Given partially filled, fill in other steps and types.	UCSD Map	Students see one vis, are asked to make minor changes AND to create a 2nd vis Students start with UCSD map of ISE AI data—Haici is rendering it. They then render it for first and second decade. Can divide records by pub year in Excel. Compare first and second decade topic map.	ERC UCSD Map = Monitor S&T Dev	Sci2 UCSD Map	Identify key topical insight needs in your organization. Sketch possible visualization. Detail data variables to graphical variables mapping.	D-Post your comments on A3 and Optional Reading. D-Post your favorite and most insightful topic vis or tool. Rate insightfulness of two topic vis/tool posted by others.
5	**"With Whom": Network Data Analysis and Visualization** • V-Exemplary Visualizations (GDELT) • V-Terminology • V-DVL Framework & Terminology ❖ Horizontal Transfer ❖ Trees ❖ Networks • V-Workflow Design • V-Algorithm Details: TreeMap Visualization	60-63	PNAS-CJobs	**Flashcards:** Trees and networks—identify correct name and type **DVL-Typology:** Given some vis from the Demo videos, identify what interactivity types are supported. **DVL-Process:** Given a type, list all members (multiple choice).	Co-Author Network	Students see no vis, are asked to create 1-2 vis Students start with ISE AI co-author network—Haici is rendering it. They then load NSF AI file and render co-PI network. Compare both.	ERC Co-Author Network = Manage Traffic / Communication Flows	Gephi	Identify key network insight needs in your organization. Sketch possible visualization. Explain what data, workflow, deployment should be used.	D-Post your comments on A3 and Optional Reading. D-Post your favorite and most insightful network vis or tool. Rate insightfulness of two nw vis/tool posted by others.
6	**Value Creation via Data-Driven Decision Making, Future Developments** • V-Value Creation via Data-Driven Decision Making (Amatria) (also team composition and costs) • V-Usability Studies • V-Future Developments	64-73	PNAS- Intro	**Flashcards:** All 50. Get 5 names and types correct. **DVL-Typology + Process:** Given 'Your Project,' list all used steps and members of each type.	How to load your own data into MAV. + How to read YML files, also work logs + How to expand MAV.	Students create a vis solution for their own project, see second to last column—with or without using MAV.	Tavola AISL MAV = Exploit IoT	None	Identify key insight needs in your organization. Sketch possible visualization. Explain what data, workflow, data mapping, deployment should be used. Estimate ROI.	D-Post your comments on A3 and Optional Reading. D-Post brief info on 'your project' and how it helps you make DDD. Rate two projects posted by others.
Post	• V-Katy and Andreas in front of VIS wall (Earth) congratulating all, welcoming future collaborations on Case Studies; foreshadowing Courses on Visual AI and Visual IoT • Exit-Survey—what did you like/dislike? • What was most valuable/actionable? • What topics should be covered in more detail? • Certificate via IU Expand									

CNS Cyberinfrastructure for Network Science Center

Schedule: Your Project

- Week 1 – Introduction to Visual Analytics

- Week 2 – When: Temporal + Identify your very own "When" questions
- Week 3 – Where: Geospatial + Identify your "Where" questions
- Week 4 – What: Topical + Identify your "What" questions
- Week 5 – With Whom: Network + Identify your "With Whom" questions

- Week 6 – Value Creation via Data-Driven Decision Making and Future Developments + Submit Visualization (Specification)

Grading and Completion Criteria

Final grade is based on :

- Data Visualization Literacy Framework Quizzes (30%)
- Make-A-Vis Tasks (30%)
- “Your Project” Work (20%)
- Participation (20%)

Students must achieve a score of 70% or higher to pass the course.

Case Studies

- Guiding Professional Training Choices
- Optimizing Career Trajectories within Boeing
- Visualizing Business Process Outsourcing/Supply Chain Management

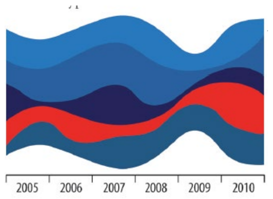
Based on MyProject submissions in Beta Run of the Visual Analytics Certificate we are also exploring case studies on

- Airspace and Operational Efficiency Visualizations
- Safety and Health Visualizations
- Visualizing Software Development Progress Using GitHub Activity Data

Case Study 1: Career Trajectories within Boeing

Many employees at Boeing have been with the company for several years if not decades. During their tenure at Boeing, many held diverse jobs, with different associated skill sets. Using anonymized Boeing HR data, we will plot the trajectories within Boeing over the last 20 years.

- Different types of "reference systems" might be valuable, including
- WHEN-Time: A stream graph, color coded by job types, and thickness coded by the number of employees
- WHERE-Geo: A geospatial map of the Boeing campus
- WHAT-Topic: A topic map of skills, see https://www.pnas.org/content/115/50/12630
- NETWORK: A network layout of all major career trajectories.



"Data overlays" of interest would help reveal typical salary, mean age, types of skills, typical number of years active, etc. associated with certain job types. They would make visible entry jobs that effectively start a successful career at Boeing. They would show major "career highways" that a large number of employees have travelled. Visualizations would identify jobs (and associated skills) that slow down or accelerate career growth; or lead to employees resigning.

Different combinations of reference systems and data overlays will be explored with Boeing experts, see below list of most relevant experts. The 2-3 best will be prototyped and one will be selected for use in the course. Associated exercises and assessments will be tested in Run1-Boeing and refined and optimized for Run2-Boeing.

Experts most relevant: HR experts and those involved in hiring and promotion decision making + optimizing the Boeing workforce for maximum competitiveness.

Case Study 2: Guiding Professional Training within Boeing

Employees at Boeing have access to 50,000 courses and degrees that take between 30mins to 3 years (for a degree) to complete and are provided by IEEE, Coursera, and others. Most employees are overwhelmed by this rich set of options. Few explore the space of training opportunities and a rather small percentage completes any of the courses.

Boeing does own access log files for all courses that Boeing employees and employees of Boeing suppliers can take. Using these log files and metadata for all courses, the landscape of training opportunities can be plotted and data on course quality, costs, time invested can be overlaid.

Different types of "reference systems" might be valuable, including

- WHEN-Time: A stream graph, color coded by course types, and thickness coded by the number of employees taking the course OR the skills trained.
- WHERE-Geo: A geospatial (world) map of work locations of Boeing/supplier employees that took the course.
- WHAT-Topic: A topic map of skills, see https://www.pnas.org/content/115/50/12630
- NETWORK: A network layout of all major course taking trajectories.

"Data overlays" of interest would help reveal course quality, costs, time commitment, #employees that took the course. They would make visible courses that have high/low completion rates. They would show major "learning highways" that a large number of employees have travelled. If data on employee job positions/salaries can be linked, visualizations would help identify courses (and associated skills) that slow down or accelerate career growth; or lead to employees resigning.

Different combinations of reference systems and data overlays will be explored with Boeing experts, see below list of most relevant experts. The 2-3 best will be prototyped and one will be selected for use in the course. Associated exercises and assessments will be tested in Run1-Boeing and refined and optimized for Run2-Boeing.

Experts most relevant: HR experts and others to guide employees in selecting best training options. Employees interested to invest their family time wisely to advance their expertise and career options. Plus, anyone interested to optimize employee attention usage and in developing "signals" that employees can use to maximize ROI for workforce training.

Case Study 3: Optimizing Business Process Outsourcing

This case study builds on insights gained from *Case Study: Supply Chain Dreams and Nightmares* by Gerard Chick. Kogan Page. (c) 2016.

Advanced business process outsourcing (BPO) is needed to procure a complex product system such as a commercial airliner that involves xx suppliers in xx countries producing xx parts that all need to fit together in highest quality—and within time and budget constraints. For the 787 Dreamliner, the BPO model can be easily as complex as the model of the aircraft itself. A detailed understanding and optimization of the BPO is required to meet quality goals as well as budget and time constraints.

Given the maxim that the performance of the prime manufacturer (here Boeing) can never exceed the capabilities of the least proficient supplier, and the fact that delays in the earlier parts of the supply chain percolate through the entire supply chain, it seems desirable to visualize the network of all suppliers and intermediate products together with associated costs, current delays, any quality issues that arose, etc. so all involved gain a more complete understanding of this multi-level socio-technical systems problem and its current solution.

Boeing might not be able to release the complete supply BPO model for the 787 Dreamliner. However, an example such as all parts needed for an important subcomponent of an aircraft that is not produced any more or a toy example such as all ingredients needed for a an amazing recipe (that employees might like to cook with their families) could serve to illustrate how to analyze and visualize BPO data. Together with information on how to read in Boeing data, employee-students could re-run the visualizations with Boeing internal data.

- Different types of “reference systems” might be valuable, including
- WHEN-Time: Gantt chart
- WHERE-Geo: A geospatial (world) map of Boeing suppliers
- NETWORK: A tree visualizations showing part-of relationships of all airplane parts.

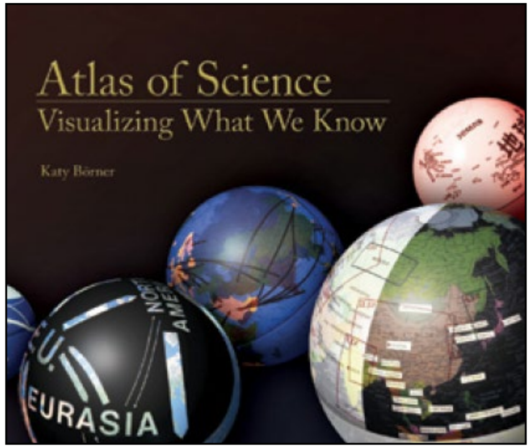
“Data overlays” of interest would help reveal costs, delays, quality for each supplier/subcomponent. They would make visible suppliers that cause delays, increase costs, and are the reason for quality problems. In addition, costs for transporting components and managing teams that work on outsourced components should be shown.

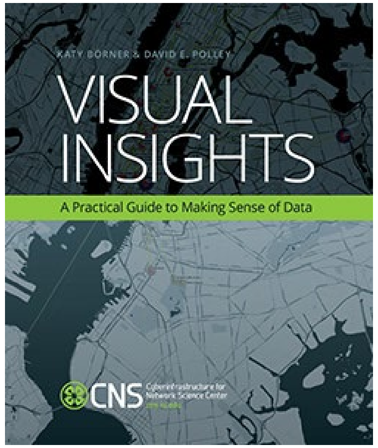
Different combinations of reference systems and data overlays will be explored with Boeing experts, see below list of most relevant experts. The 2-3 best will be prototyped and one will be selected for use in the course. Associated exercises and assessments will be tested in Run1-Boeing and refined and optimized for Run2-Boeing.

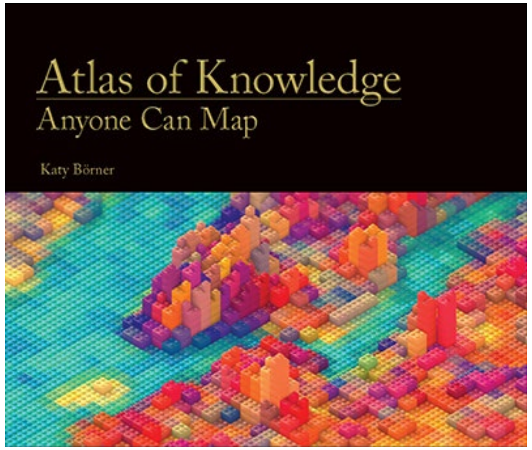
Experts most relevant: System engineering experts that develop, optimize, and ensure compliance to BPO.

Resources

- Börner, Katy. 2010. *Atlas of Science*. Cambridge, MA: The MIT Press. http://scimaps.org/atlas1.
- Börner, Katy and David E. Polley. 2014. *Visual Insights.* Cambridge, MA: The MIT Press. http://cns.iu.edu/ivmoocbook14.html.
- Börner, Katy. 2015. *Atlas of Knowledge*. Cambridge, MA: The MIT Press. http://scimaps.org/atlas2.







Visual Analytics Certificate

Learn to render data into actionable insights in 6 weeks!

Class begins **March 2, 2020**. Get course info at https://boeing.cns.iu.edu

For US-based Boeing employees:

Begin the enrollment process by completing the pre-enrollment form (https://iu.co1.qualtrics.com/jfe/form/SV_ekXpn97sf6gvZAx). NOTE: You will need your voucher to complete the form. Over the days following submission, look for emails about next steps and course information to finalize your enrollment.

For non-US based Boeing employees and others not using the voucher system:

Register for the course online through Indiana University's IU Expand portal. Go to: https://expand.iu.edu/browse/sice/cns/courses/visual-analytics-boeing-2. Payment is via credit card.

CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Learning Analytics

Ginda, Michael, Michael C. Richey, Mark Cousino, and Katy Börner. 2019. "Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design". *PLOS One* e0215964. doi: 10.1371/journal.pone.0215964.

Background and Motivation

In Fall 2017, our team began working with The Boeing Company to leverage our expertise in visual analytics to study data produced by students taking MITxPro online courses to understand

- the structure of courses resources,
- student engagement and learner trajectories, and
- student performance

Ginda, Michael, Michael C. Richey, Mark Cousino, and Katy Börner. 2019. "Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design". *PLOS One* e0215964. doi: 10.1371/journal.pone.0215964.

MIT, Boeing, NASA, and edX to launch online architecture and systems engineering program

Four-course program will train professionals in latest practices on models and methods to manage complex systems

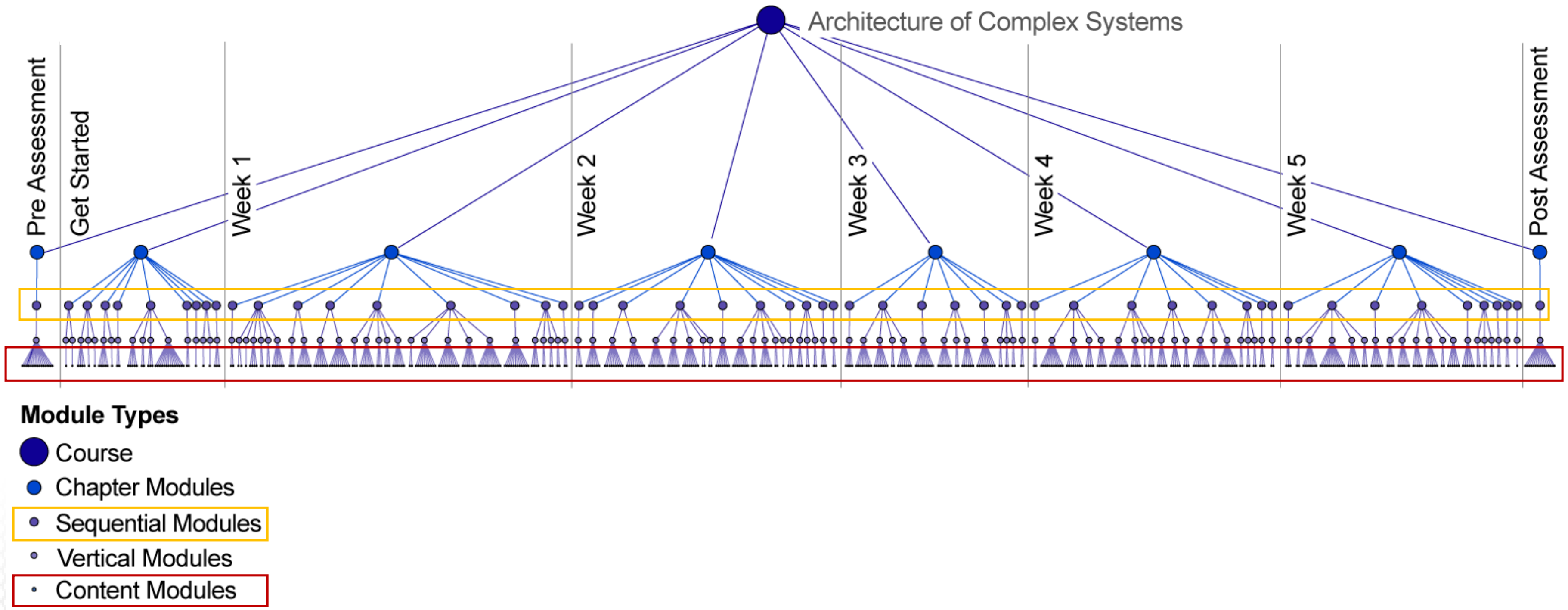
1,611 Boeing engineers registered; 1,565 were active and generated nearly **31 million click event records** while accessing videos, projects, and assessments. Some students generated over 100,000 separate events.

All but 255 engineers passed the course, resulting in a completion rate of 84.1%.

Course Structure

Course Structure Tree Diagram shows 5-level hierarchical structure of the *Architecture of Complex Systems* course. Nodes are ordered based on the sequence of learning modules presented to learners in the course.

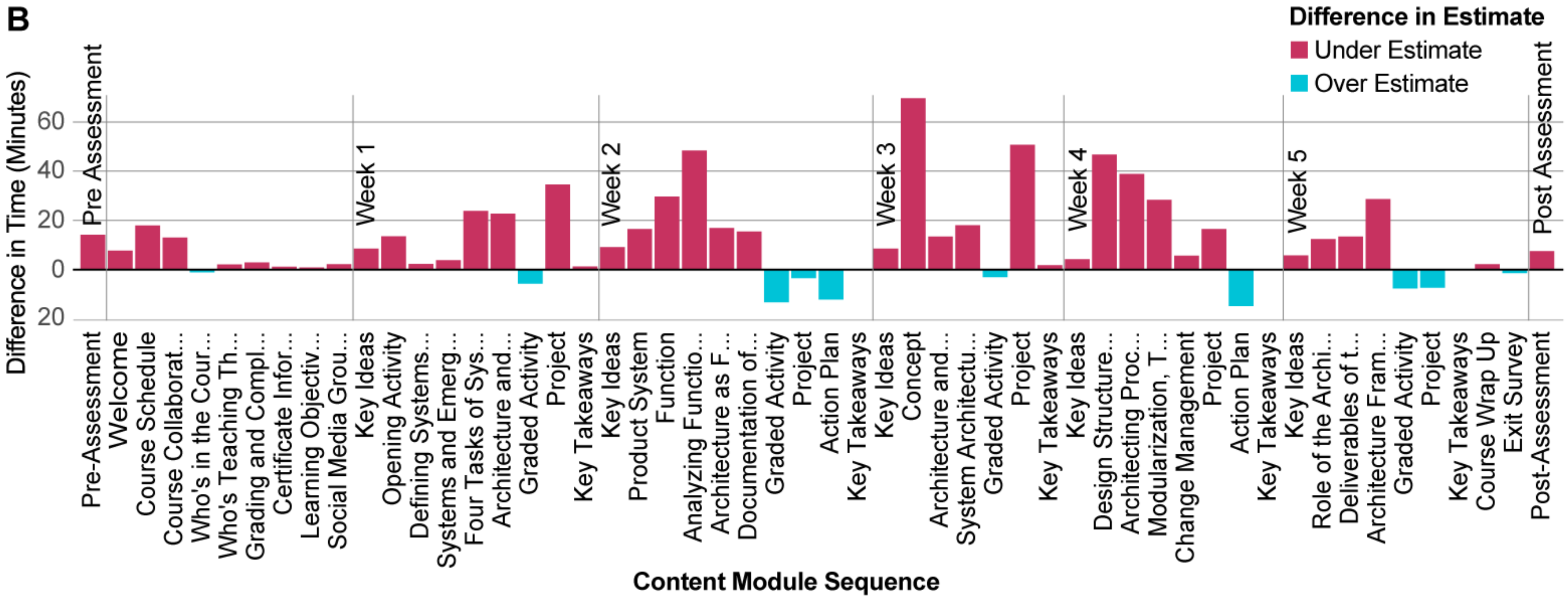
Insights: Course structure allows for analysis and visualizations at multiple levels of granularity, temporality. Modules presented to students share similar lengths.



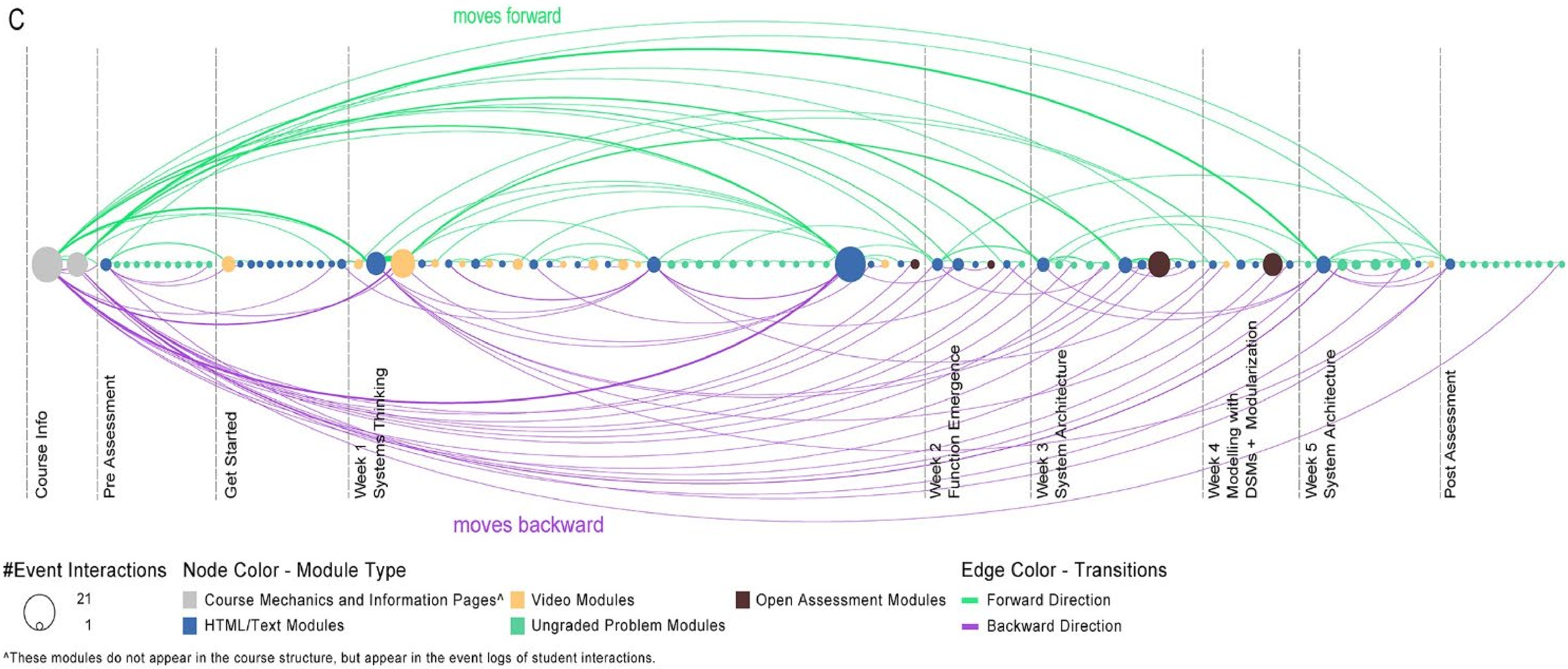
Student Engagement Predictions

Instructors Temporal Predictions are represented in a temporal bar graph that compares course instructors *estimated* time learners would need to complete course materials, and the average time taken by learners in the course computed from data.

Insights: Instructor's temporal estimates are accurate but did not account for studying activity of students in their estimates.



Learning Trajectories

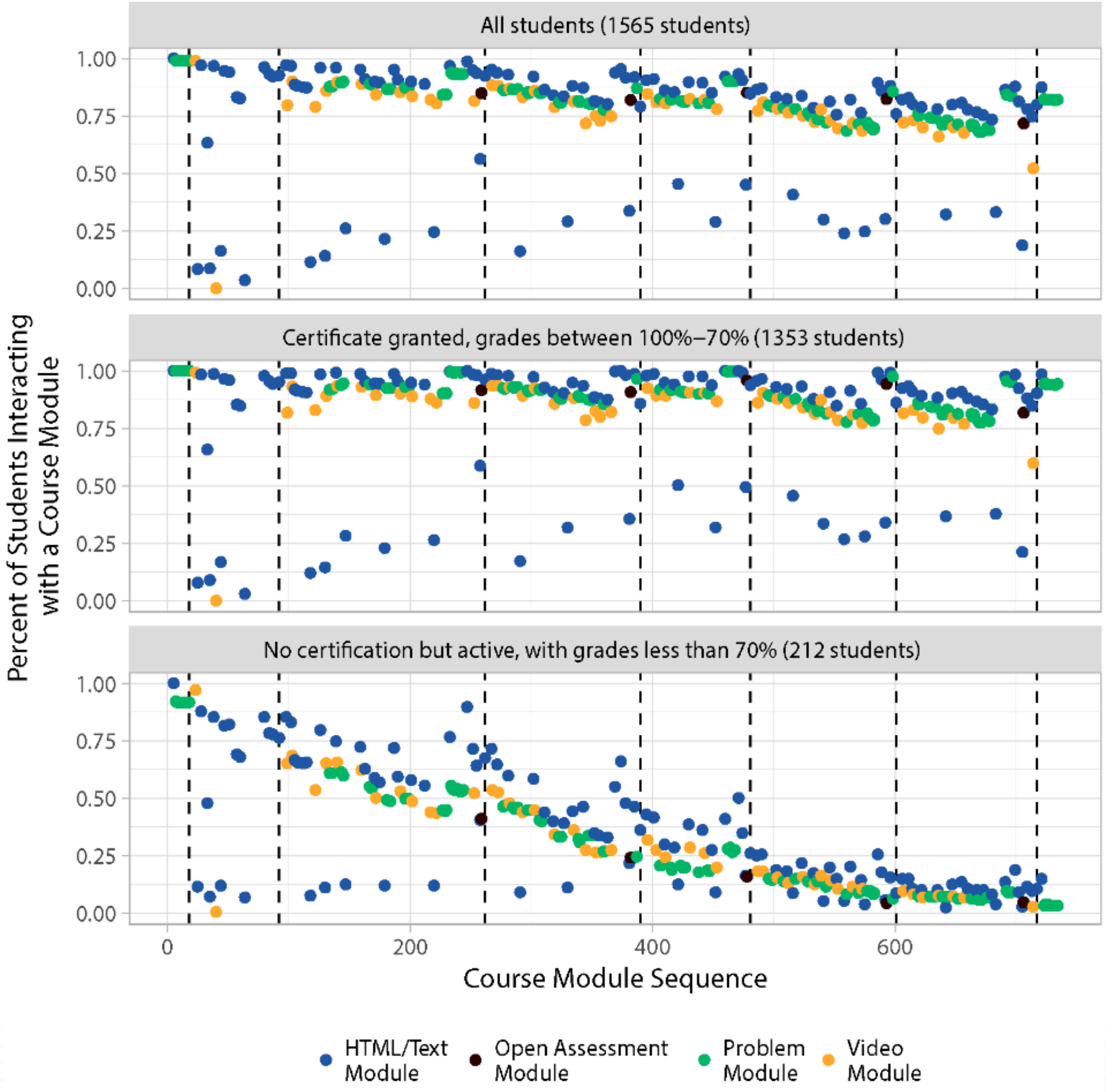


Student Engagement and Performance

Students Interacting with a Course Module.

Scatter graph showing the percentage of the learners in the *Architecture of Complex System* course accessing modules **by certificate group** and module type.

Insights: Clear difference in access patterns by students across the course by certificate and non certificate earners, as well as subtle differences between module types. Most notably, few of the students that do not earn a certificate do access the Open Assessment Modules.



CNS Cyberinfrastructure for Network Science Center

Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Thank You

Q&A

https://boeing.cns.iu.edu