Visual Analytics & Learning Analytics in support of **Data-Driven Decision Making**

Michael Ginda, Naren Suri, Andreas Bueckle, and Katy Börner
Cyberinfrastructure for Network Science Center (CNS)
School of Informatics, Computing, and Engineering
Indiana University
@katycns

Knowledge Convergence Workshop at
9th International Learning Analytics and Knowledge Conference
https://lak19.solaresearch.org

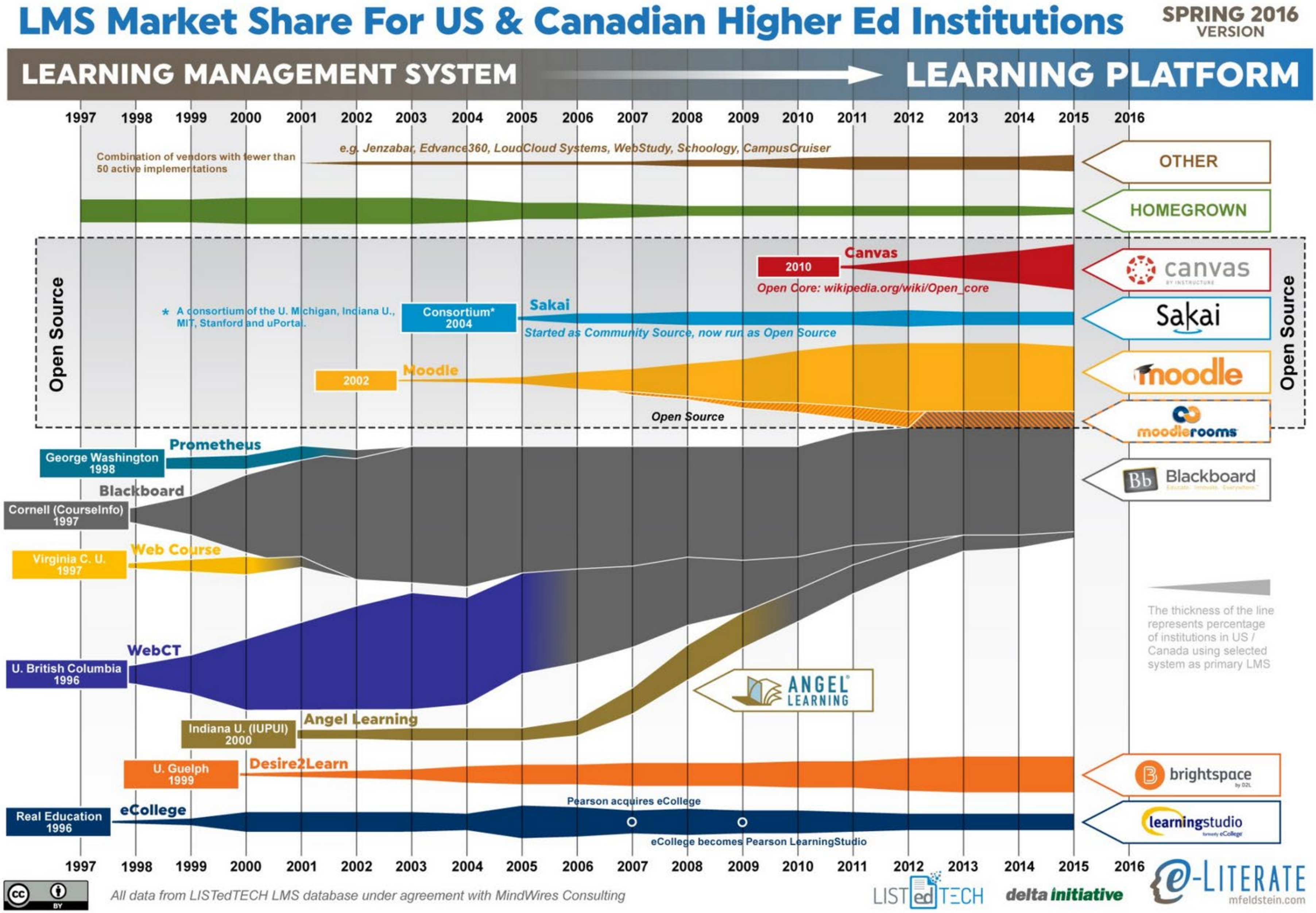
Tempe, Arizona

March 4, 2019

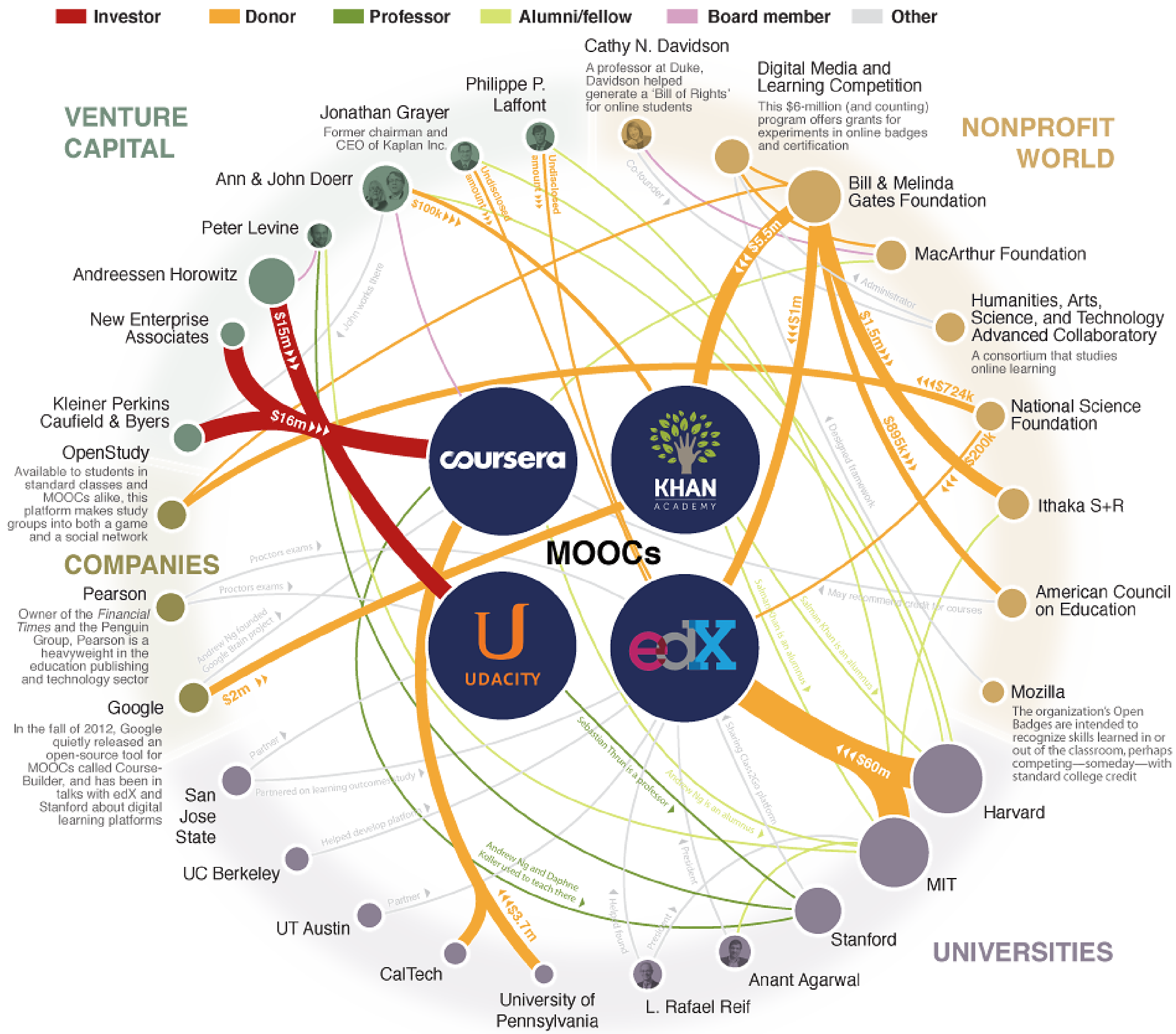
Visual Analytics & Learning Analytics in support of **Data-Driven Decision Making**

Outline:

Context
Data Driven Decision Making
Visual Analytics
Learning Analytics



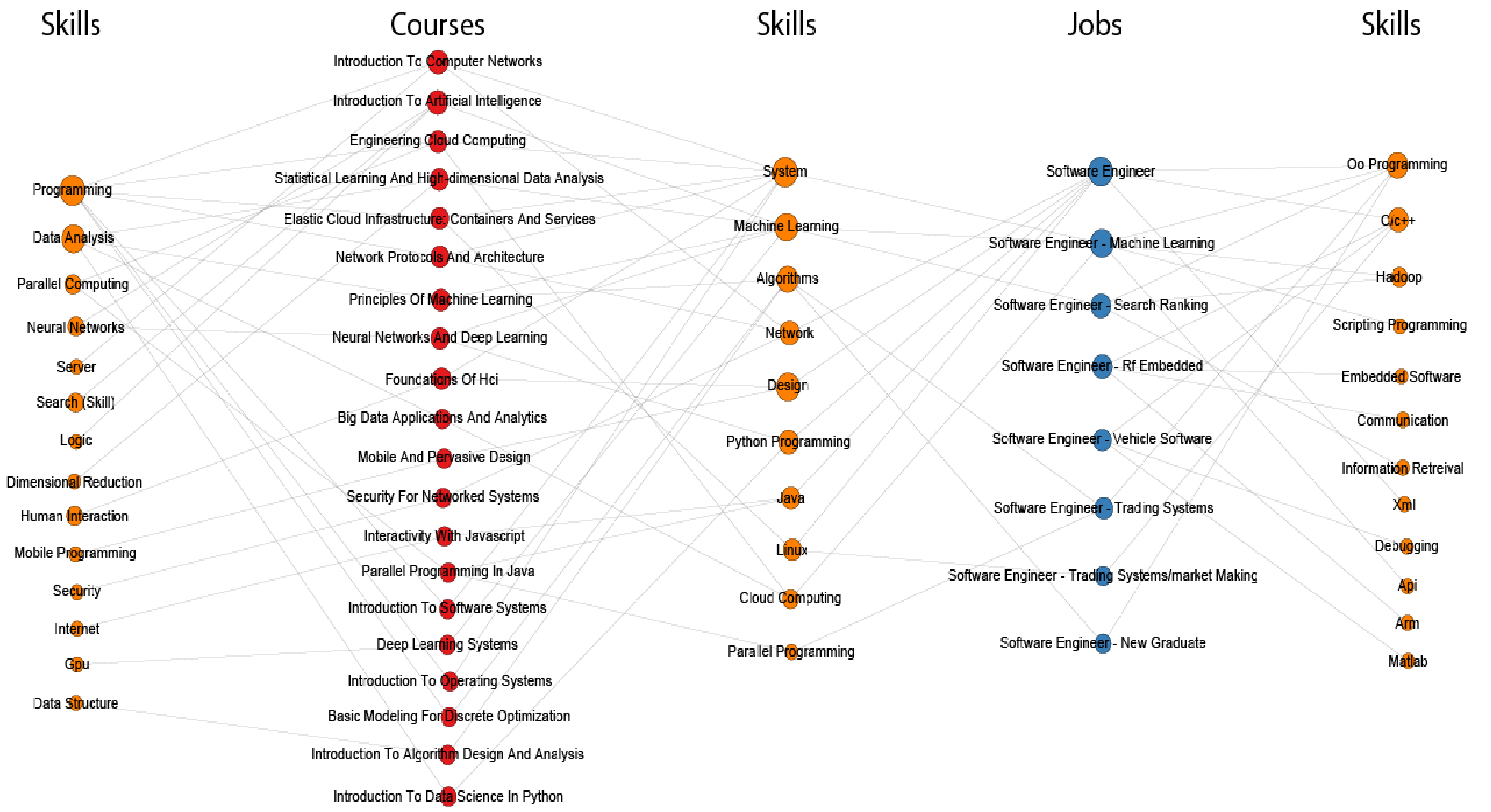
https://mfeldstein.com/state-higher-ed-lms-market-spring-2016



https://www.chronicle.com/article/Major-Players-in-the-MOOC/138817 - April 29, 2013

IU Data Science Program: Courses, Skills & Jobs

Katy Börner, Michael Ginda & Xiaozhong Liu, Indiana University

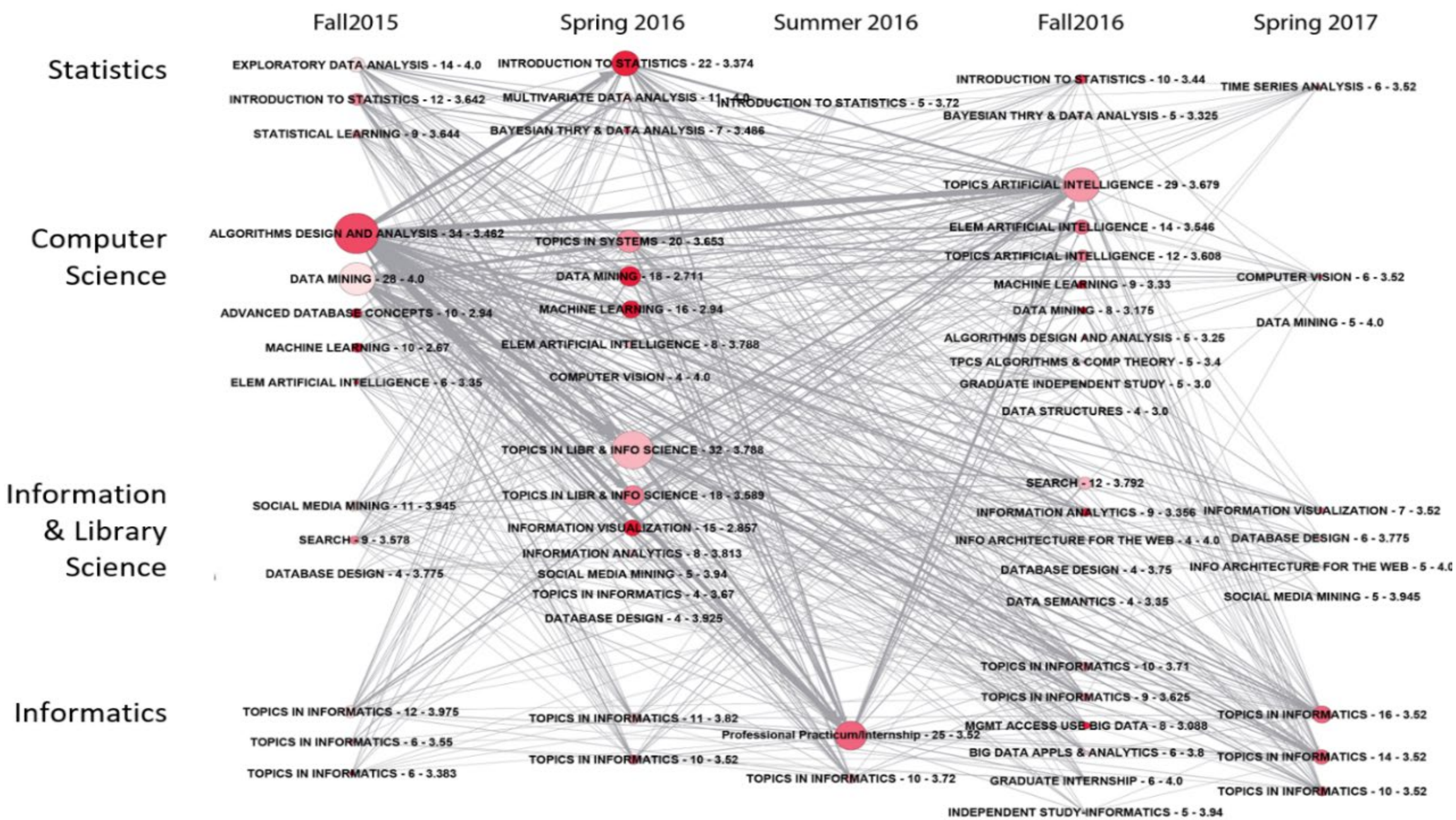


Exemplary set of IU Data Science courses, 'Software Engineering' jobs, and associated skills.

Job data was retrieved from LinkedIn and CareerBuilder and course data come from the IU course list. As can be seen, there are many skills (in orange) that are exclusively associated with courses or jobs; however, the skills in the middle interlink courses (in red) to jobs (in blue).

IU Data Science Program: Student Course Transition Network

Michael Ginda, Kayla Scroggins & Katy Börner, Indiana University



Empower students, teachers, and curriculum committee members to understand and discuss current and desirable student cohorts, key course trajectories, or the (gatekeeper) role that specific courses play.

Vertically, courses are arranged into four groups based on the department offering the course. Within each vertical grouping, the nodes are sorted by the total enrollment for the course with highest values on top. Node size encodes number of students enrolled; node color denotes overall GPA for the course.

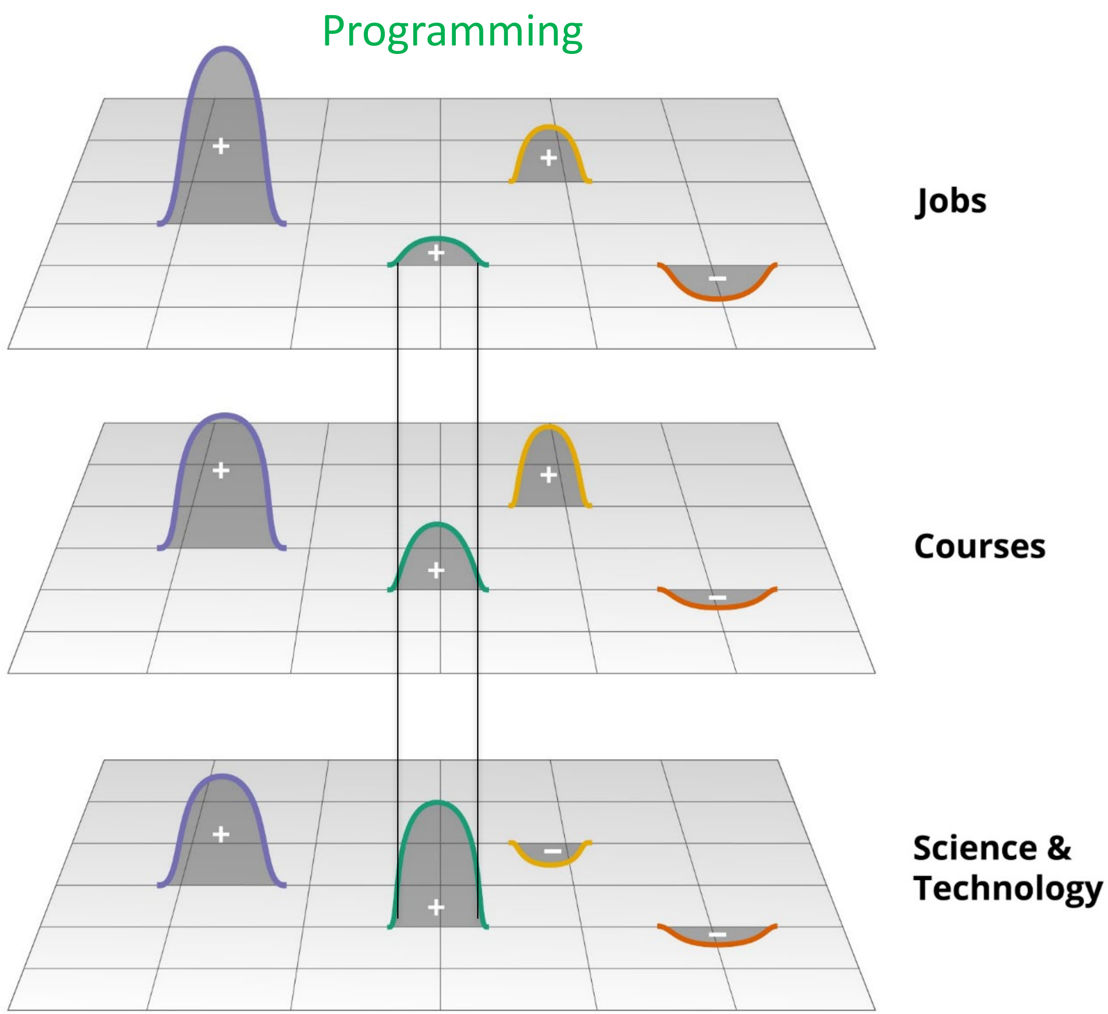
Science & Technology vs. Education/Training vs. Jobs

Katy Börner, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu and James A. Evans

Need to study the **(mis)match** and **temporal dynamics** of S&T progress, education and workforce development options, and job requirements.

Challenges:

- Rapid change of STEM knowledge
- Increase in tools, AI
- Social skills (project management, team leadership)
- Increasing team size



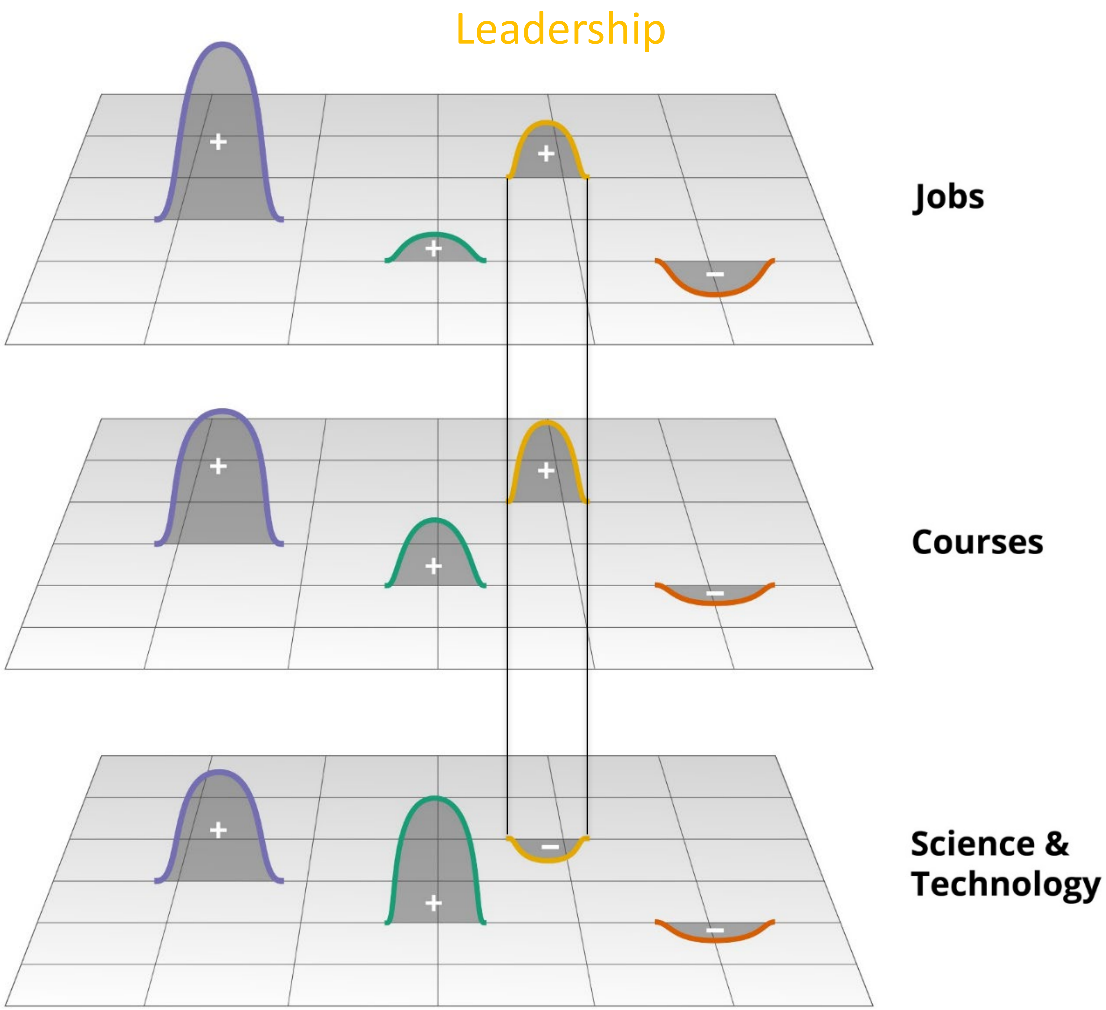
Science & Technology vs. Education/Training vs. Jobs

Katy Börner, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu and James A. Evans

Need to study the **(mis)match** and **temporal dynamics** of S&T progress, education and workforce development options, and job requirements.

Challenges:

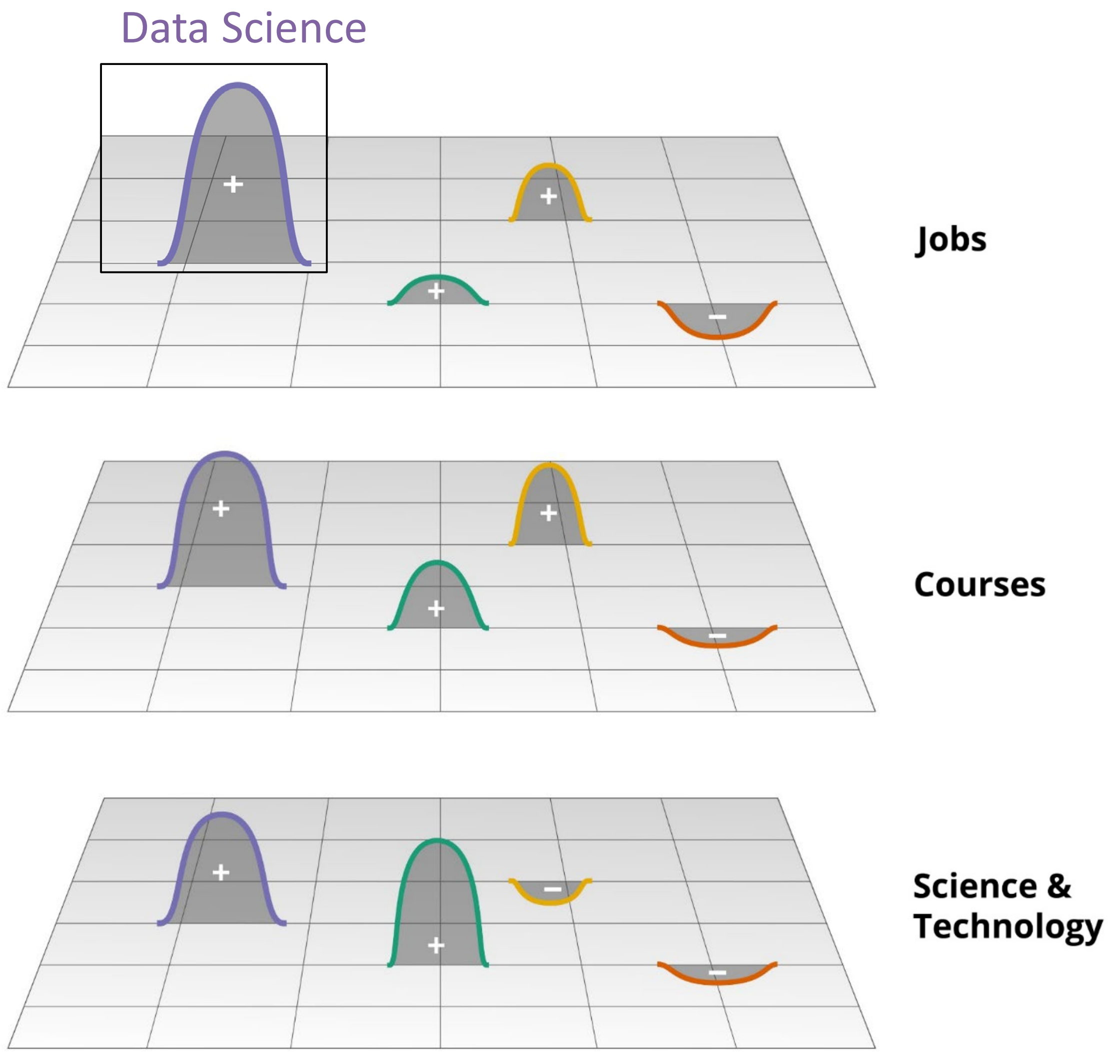
- Rapid change of STEM knowledge
- Increase in tools, AI
- Social skills (project management, team leadership)
- Increasing team size



Science & Technology vs. Education/Training vs. Jobs

Katy Börner, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu and James A. Evans

Need to study the **(mis)match** and **temporal dynamics** of S&T progress, education and workforce development options, and job requirements.



Challenges:

- Rapid change of STEM knowledge
- Increase in tools, AI
- Social skills (project management, team leadership)
- Increasing team size

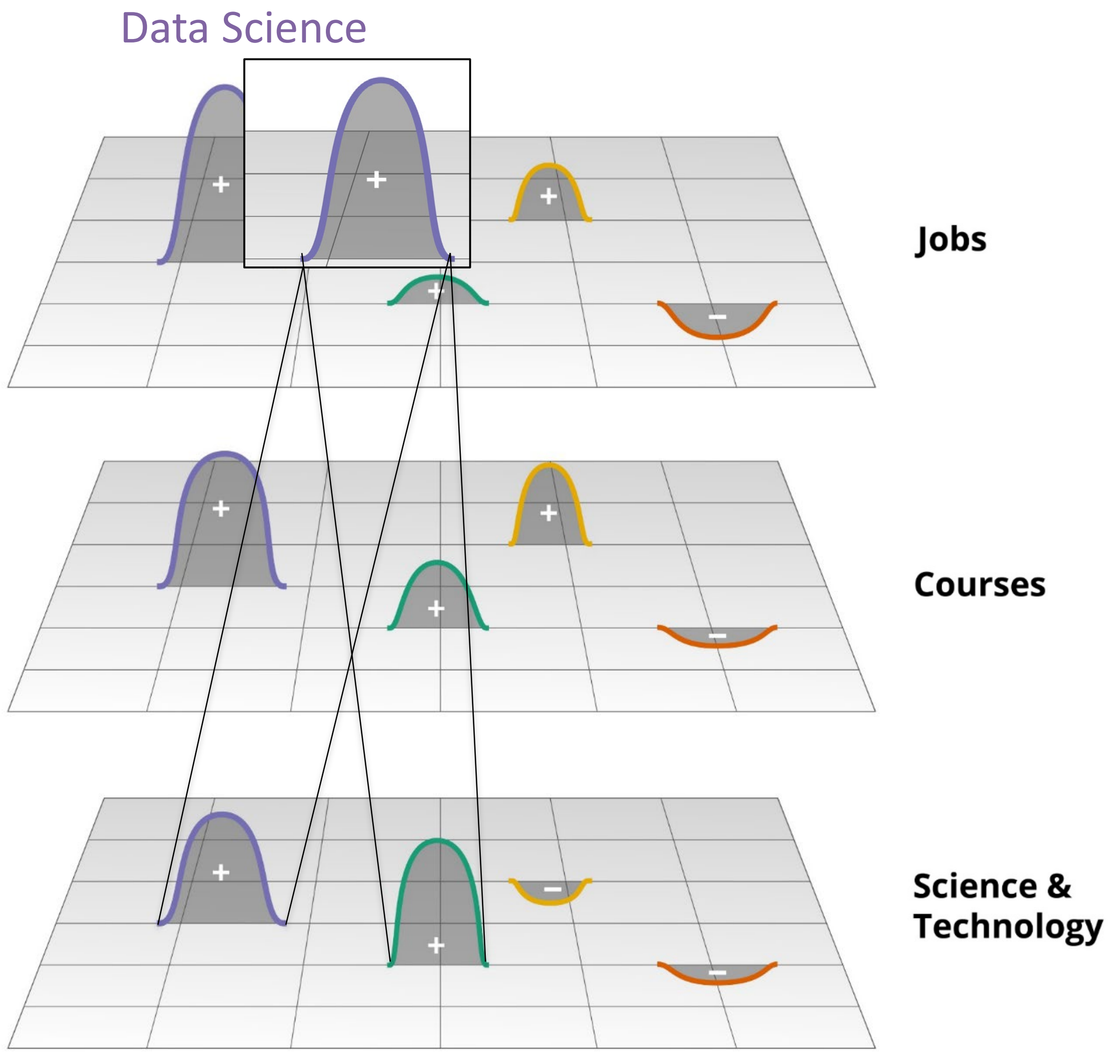
Science & Technology vs. Education/Training vs. Jobs

Katy Börner, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu and James A. Evans

Need to study the **(mis)match** and **temporal dynamics** of S&T progress, education and workforce development options, and job requirements.

Challenges:

- Rapid change of STEM knowledge
- Increase in tools, AI
- Social skills (project management, team leadership)
- Increasing team size

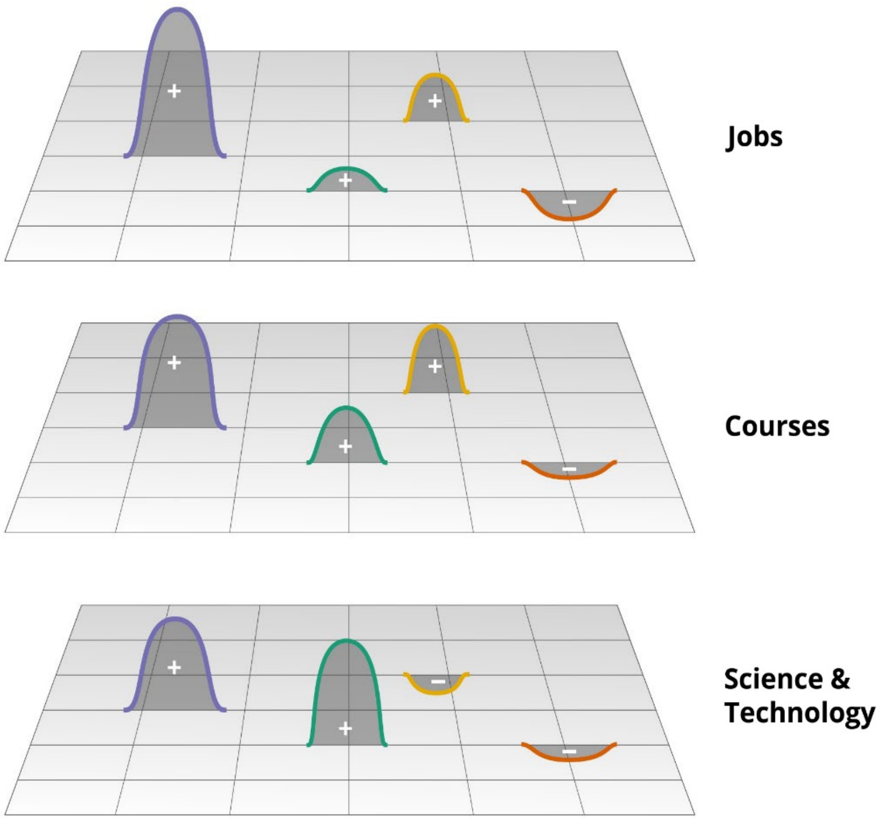


Science & Technology vs. Education/Training vs. Jobs

Katy Börner, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu and James A. Evans

Study results are needed by:

- **Students:** What jobs will exist in 1-4 years? What program/learning trajectory is best to get/keep my dream job?
- **Teachers:** What course updates are needed? What curriculum design is best? What is my competition doing? How much timely knowledge (to get a job) vs. forever knowledge (to be prepared for 80 productive years) should I teach? How to innovate in teaching and get tenure?
- **Employers:** What skills are needed next year, in 5 years? Who trains the best? What skills does my competition list in job advertisements? How to hire/train productive teams?



What is ROI of my time, money, compassion?

#SacklerModVisST

Arthur M. Sackler COLLOQUIA OF THE NATIONAL ACADEMY OF SCIENCES

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

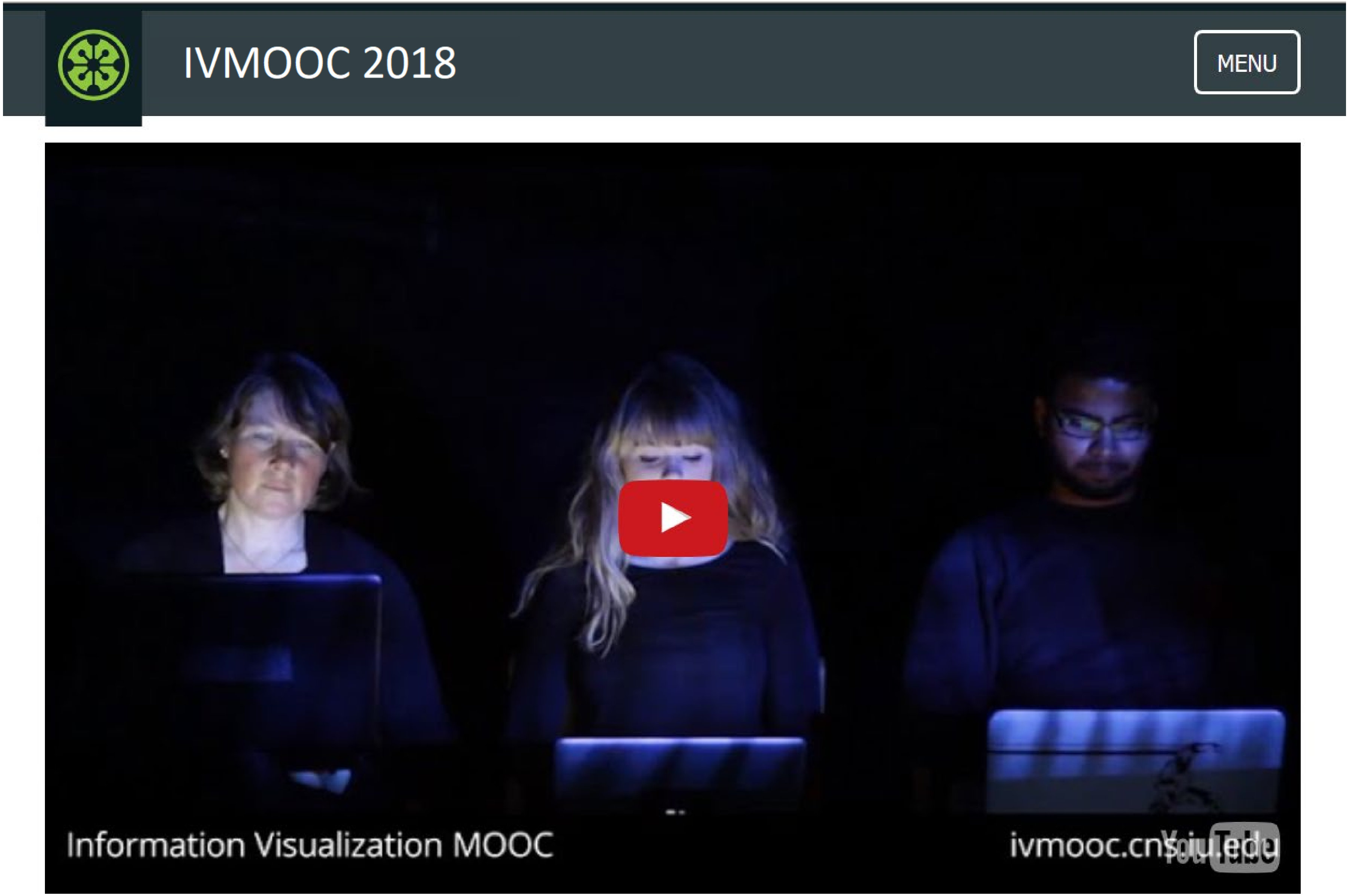
Modeling Needs, Infrastructures, Standards

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

nasonline.org/Sackler-Visualizing-Science

Arthur M. Sackler COLLOQUIA OF THE NATIONAL ACADEMY OF SCIENCES Clarivate Microsoft

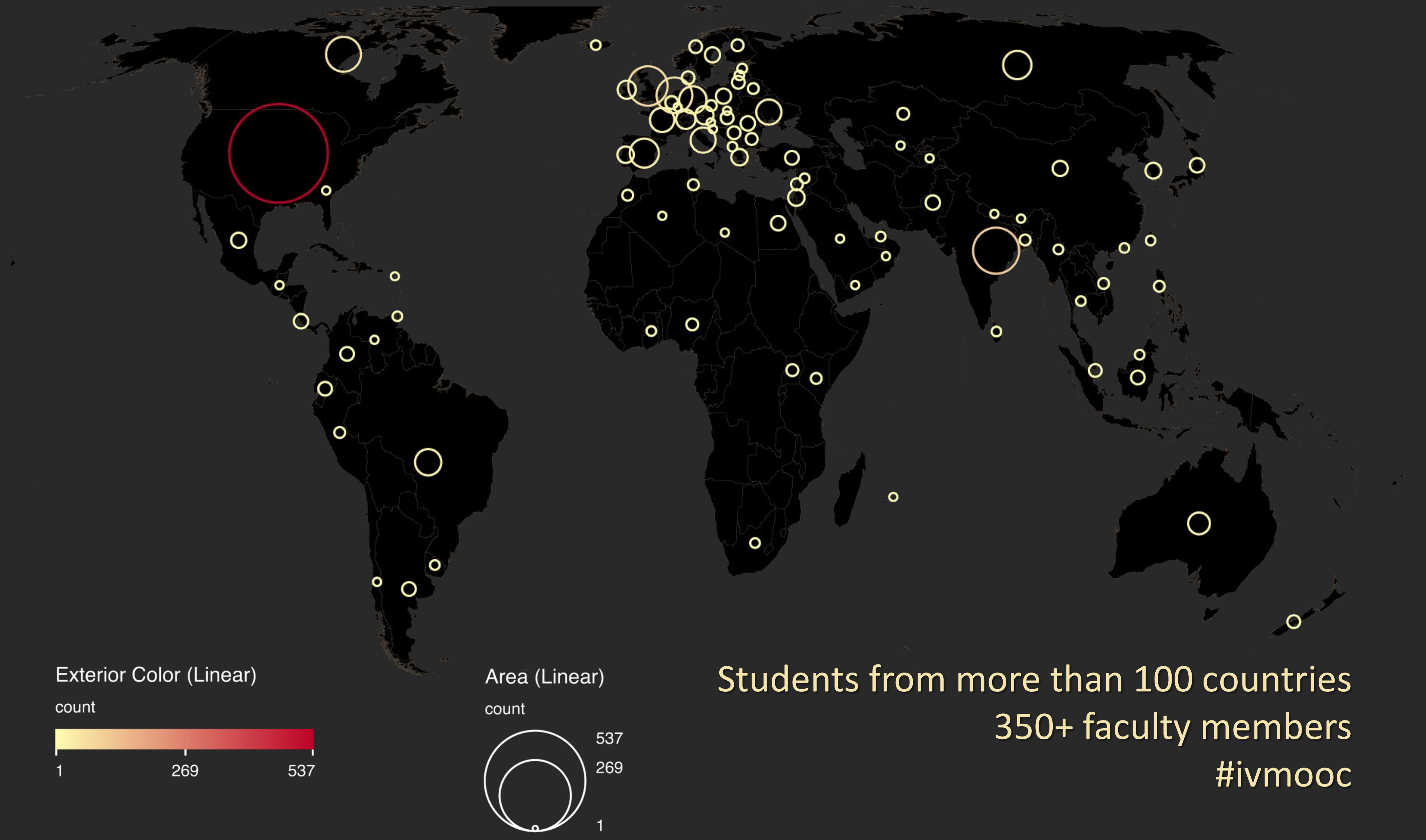
Visual Analytics - IVMOOC



Register for free: http://ivmooc.cns.iu.edu

The Information Visualization MOOC

ivmooc.cns.iu.edu



Students from more than 100 countries

350+ faculty members

#ivmooc

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.

Course Schedule

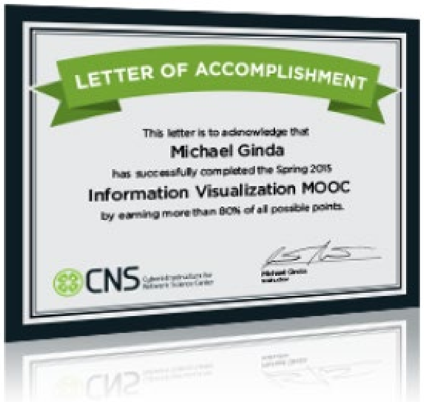
Part 1: Theory and Hands-On

- **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

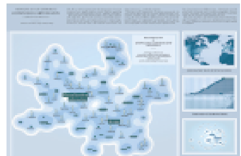
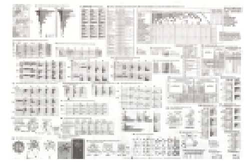
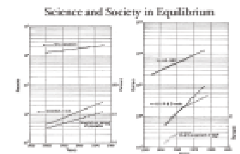
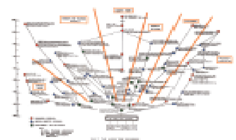
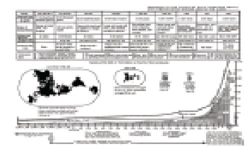
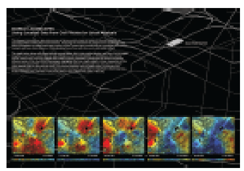
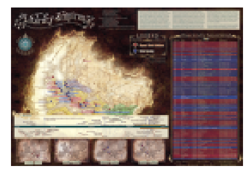
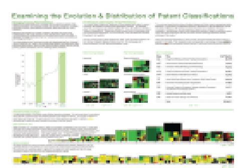
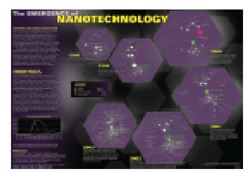
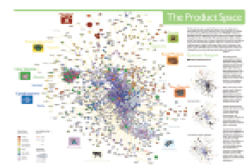
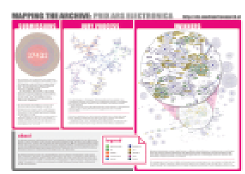
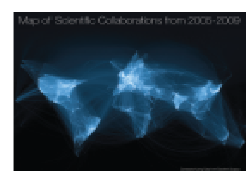
Final Exam

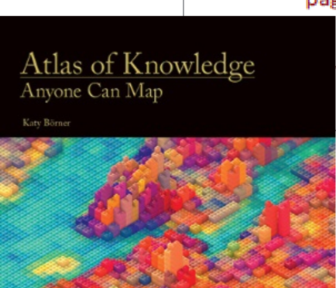


Part 2: Students work in teams on client projects.

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

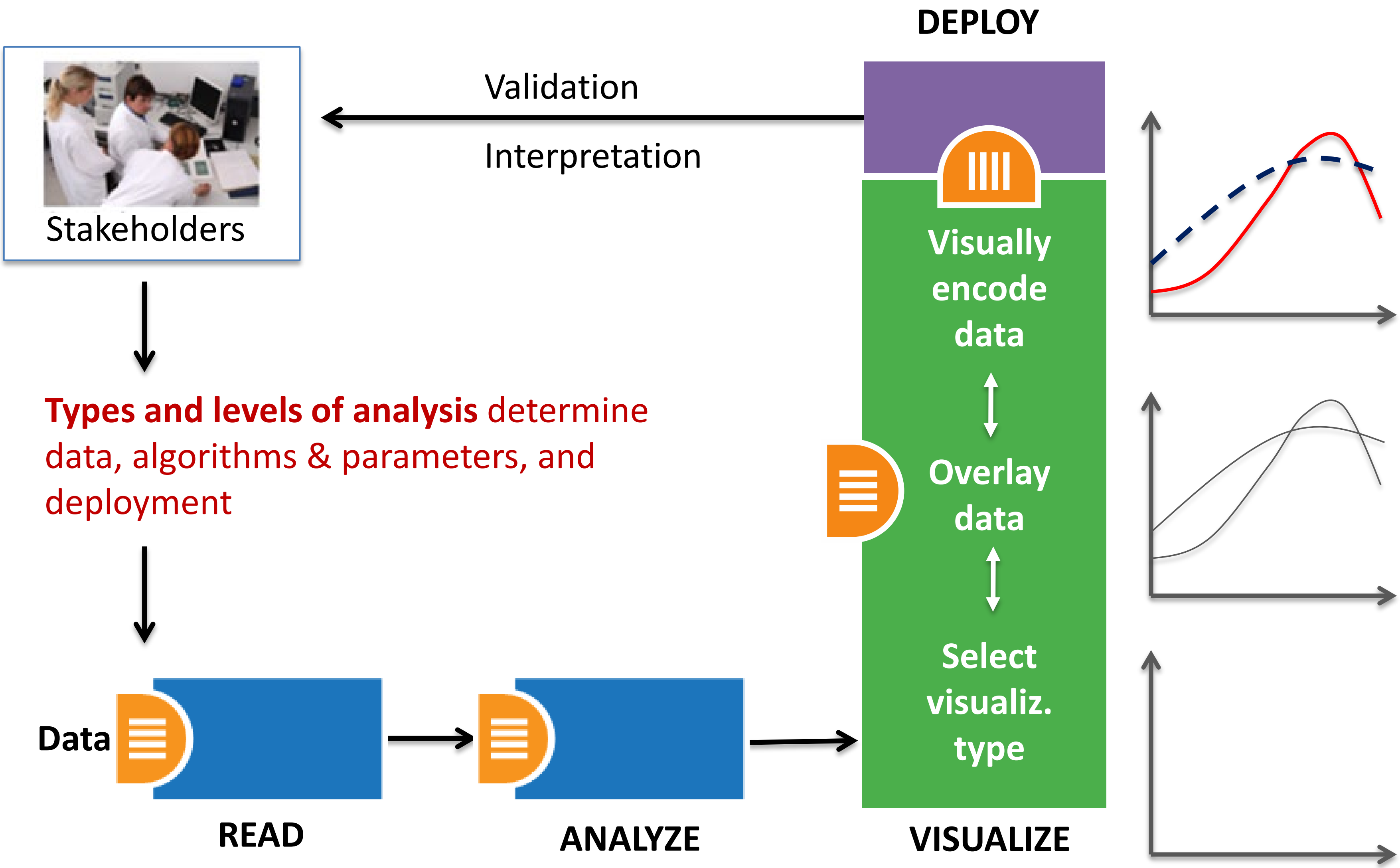
Tasks

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

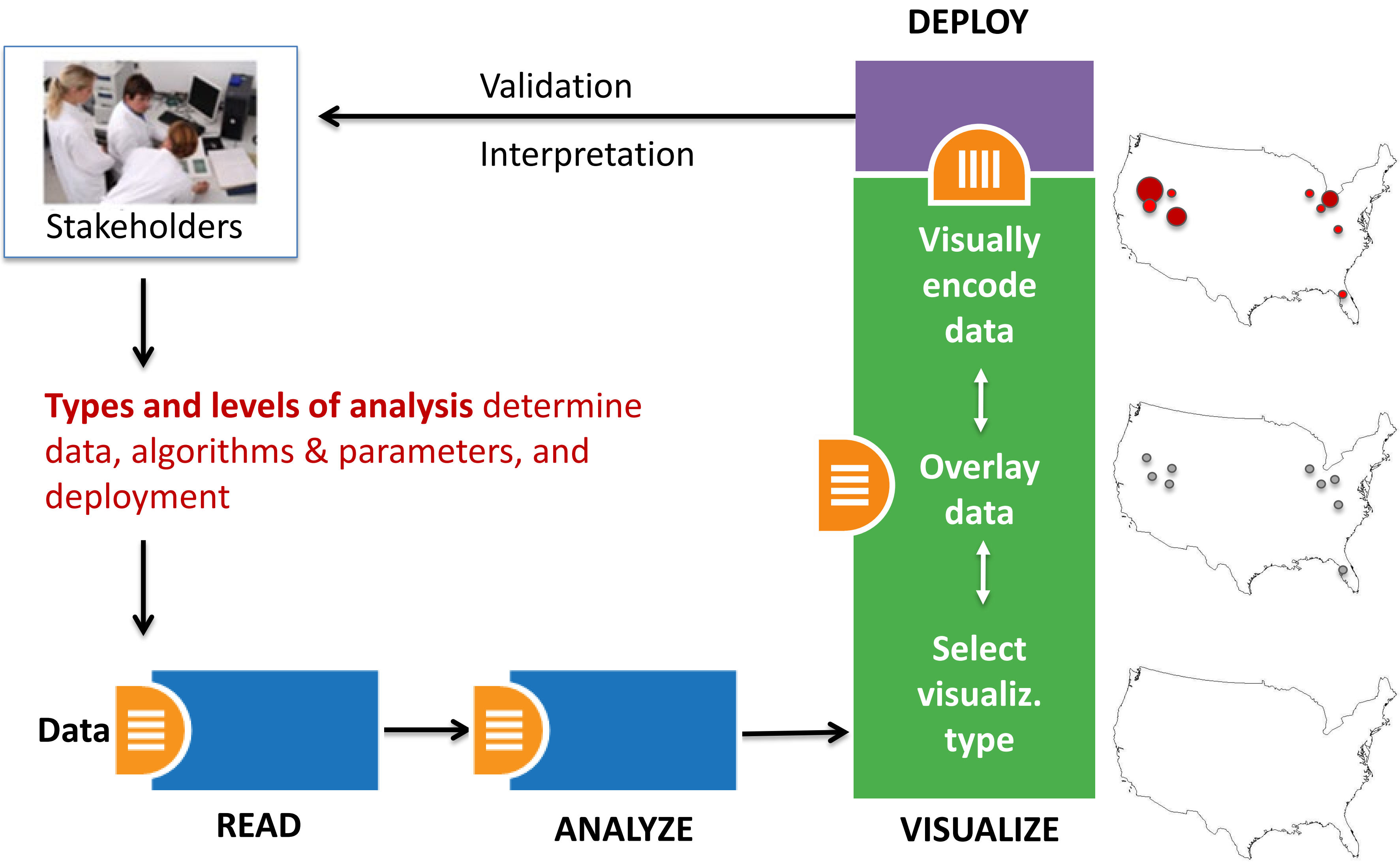


See *Atlas of Science: Anyone Can Map*, page 5

Needs-Driven Workflow Design

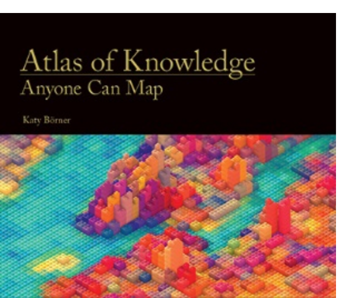


Needs-Driven Workflow Design



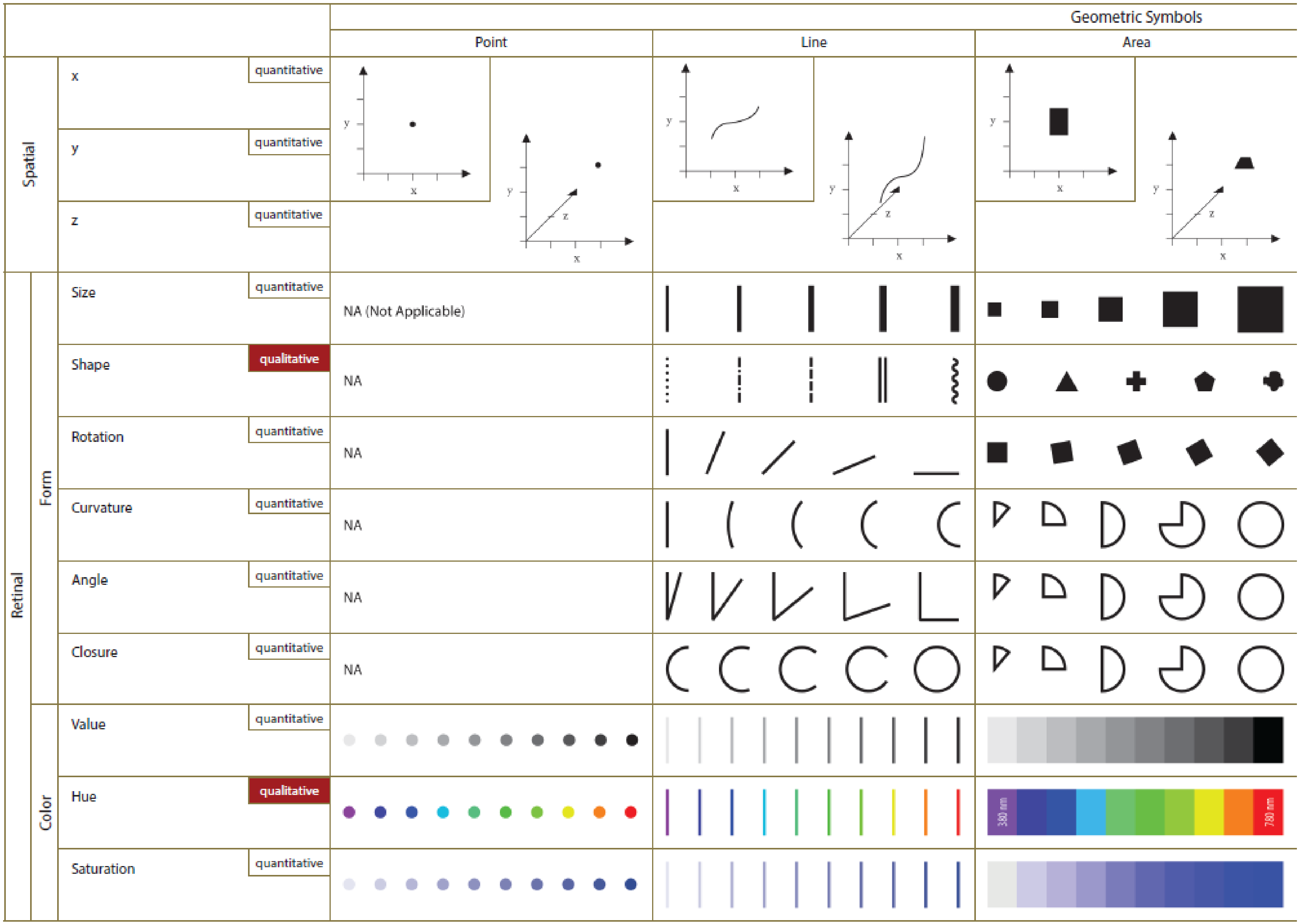
Visualization Framework

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

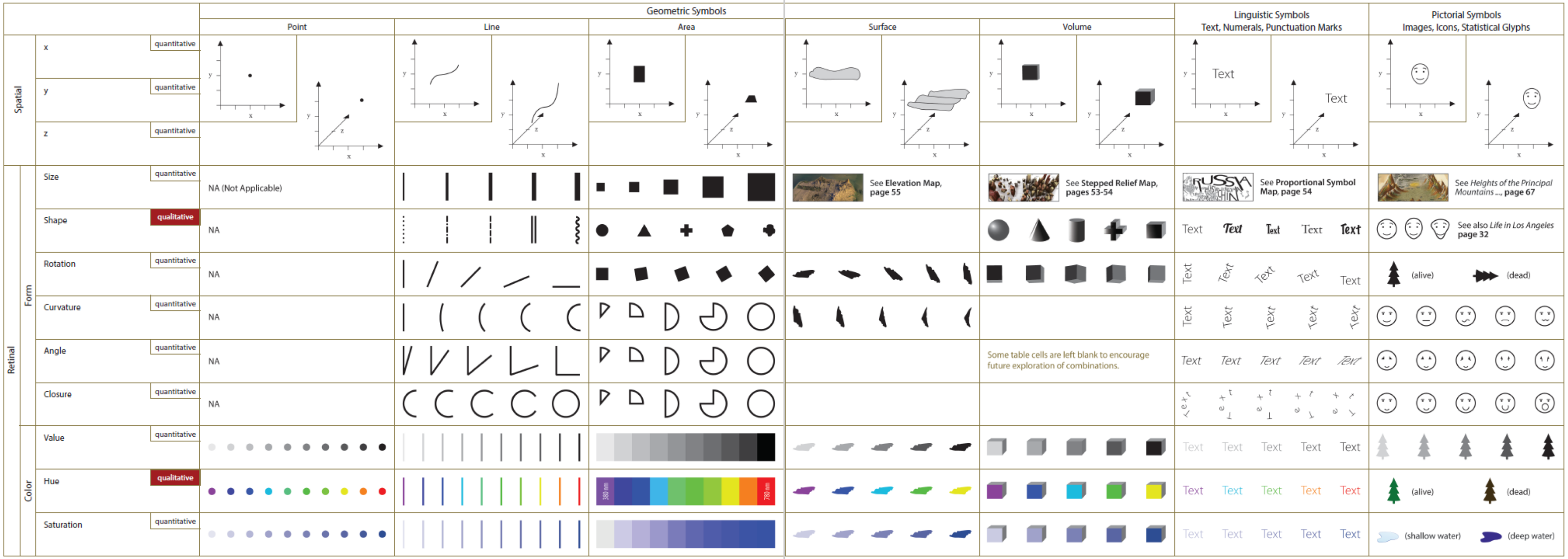


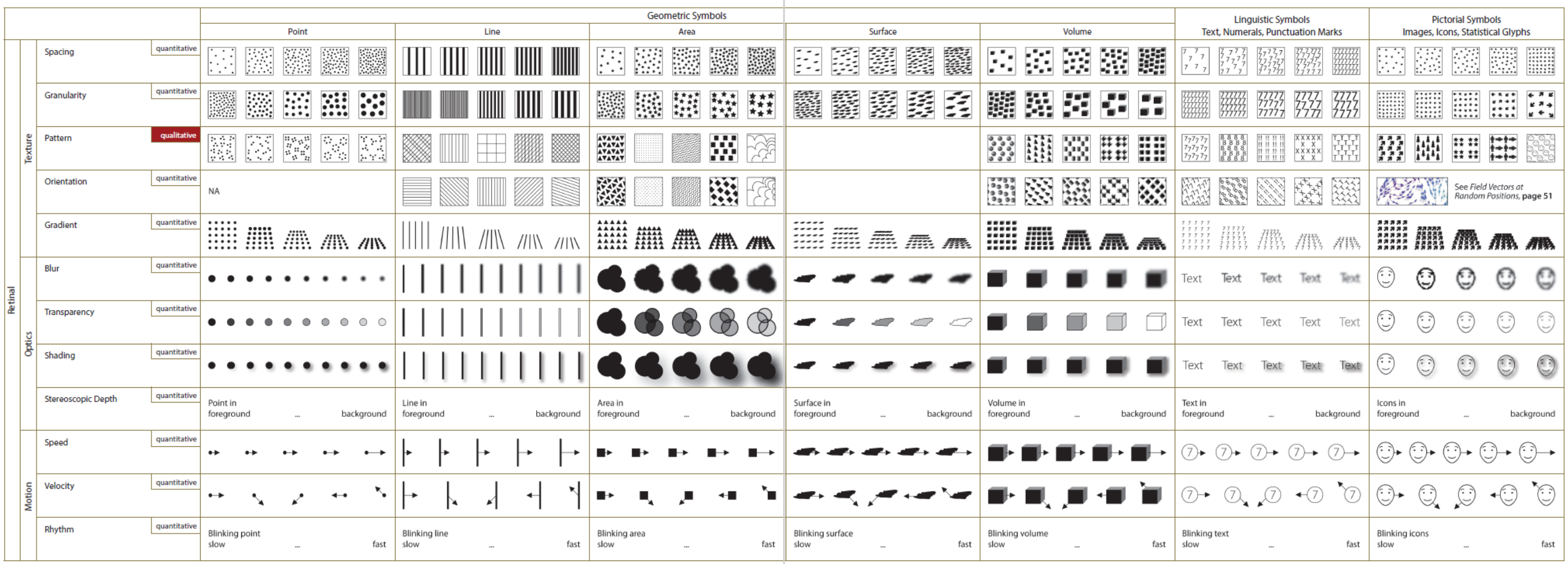
See *Atlas of Science: Anyone Can Map*, page 24

Graphic Variable Types Versus Graphic Symbol Types



Graphic Variable Types Versus Graphic Symbol Types





Learning Analytics

Learning Analytics

Empowering Teachers: How to make sense of the activities of thousands of students? How to guide them?

Empowering Students: How to navigate learning materials and develop successful learning collaborations across disciplines and time zones?

Empowering Researchers: How do people learn? What pedagogy works (in a MOOC) and when?

Empowering MOOC Platform Designers: What technology helps and what hurts?

Visualizing IVMOOC Data

Data was collected from different sources:

- 1,901 students registered via GCB (1215 male/557 female)
- 52,557 slide downloads from our server
- 18,893 video views via YouTube
- 193 accounts made 730 tweets
- 134 students took 183 exams in GCB
- 674 remarks on 215 different forum threads in Drupal
- 64 students submitted projects via Drupal

Learning Analytics

Empowering Teachers: How to make sense of the activities of thousands of students? How to guide them?

Empowering Students: How to navigate learning materials and develop successful learning collaborations across disciplines and time zones?

Empowering Researchers: How do people learn? What pedagogy works (in a MOOC) and when?

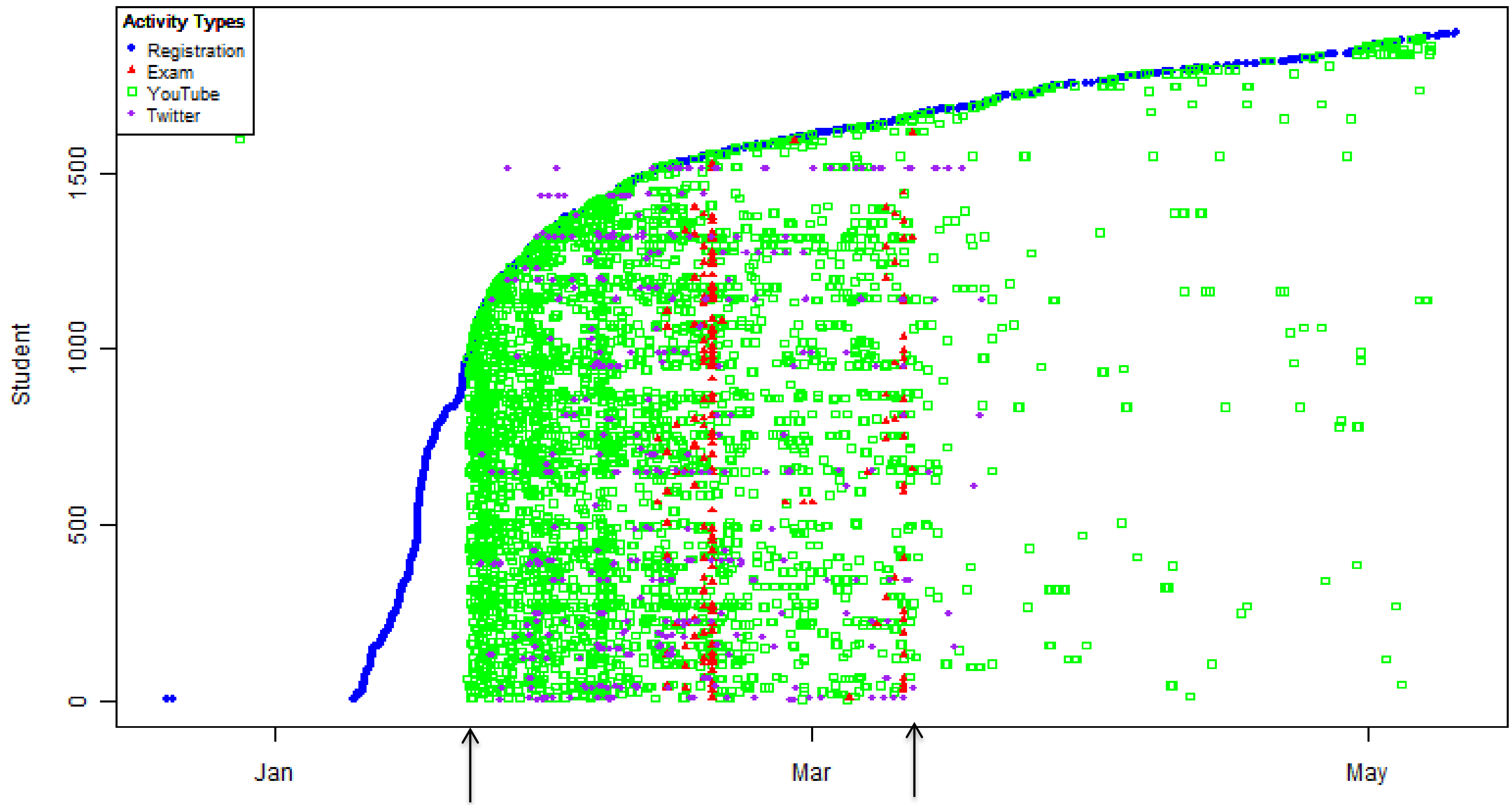
Empowering MOOC Platform Designers: What technology helps and what hurts?

Visualizing IVMOOC Data

Data was collected from different sources:

- 1,901 students registered via GCB (1215 male/557 female)
- 52,557 slide downloads from our server
- 18,893 video views via YouTube
- 193 accounts made 730 tweets
- 134 students took 183 exams in GCB
- 674 remarks on 215 different forum threads in Drupal
- 64 students submitted projects via Drupal

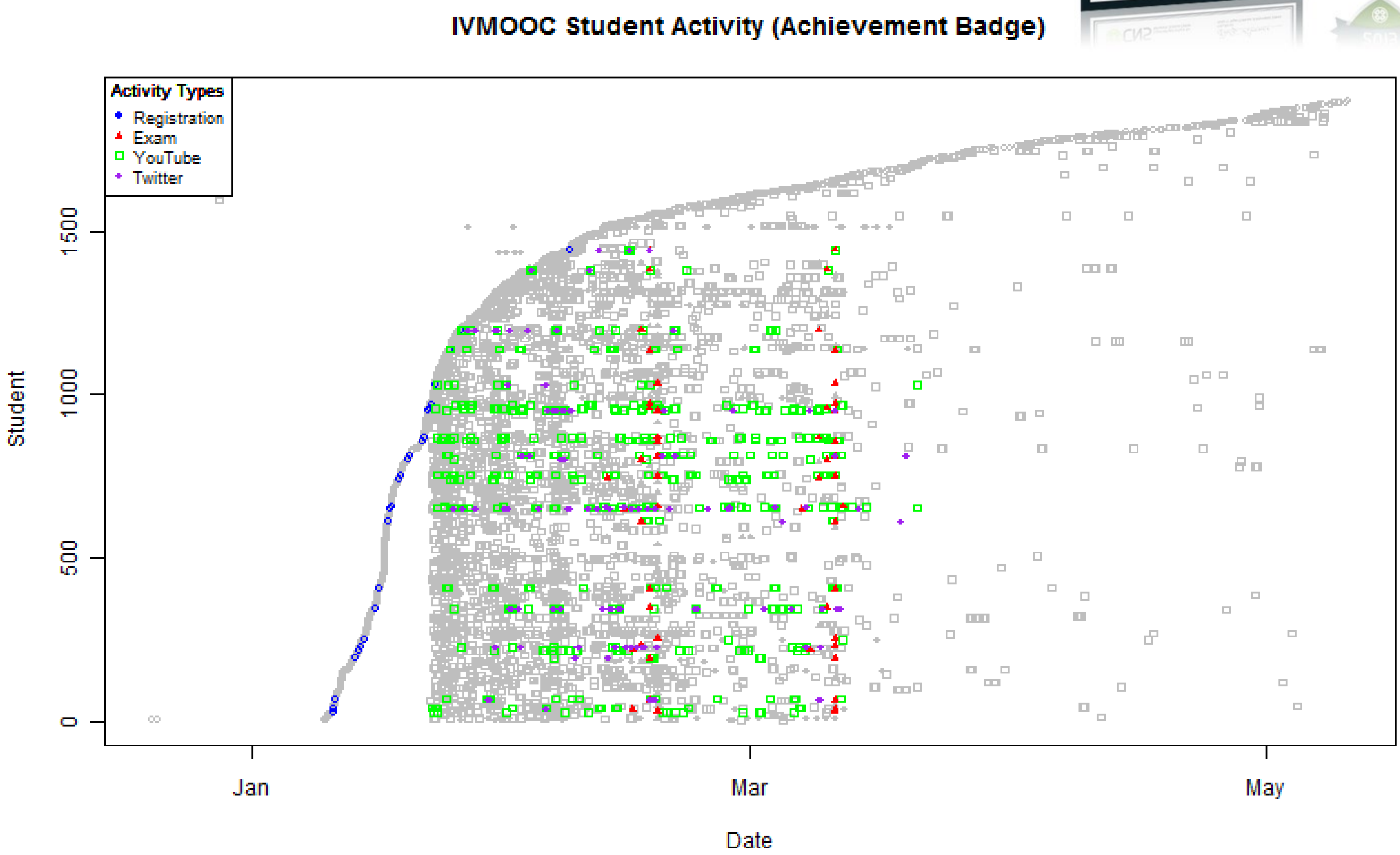
Student Registration and Activity



Jan. 22: Course Starts

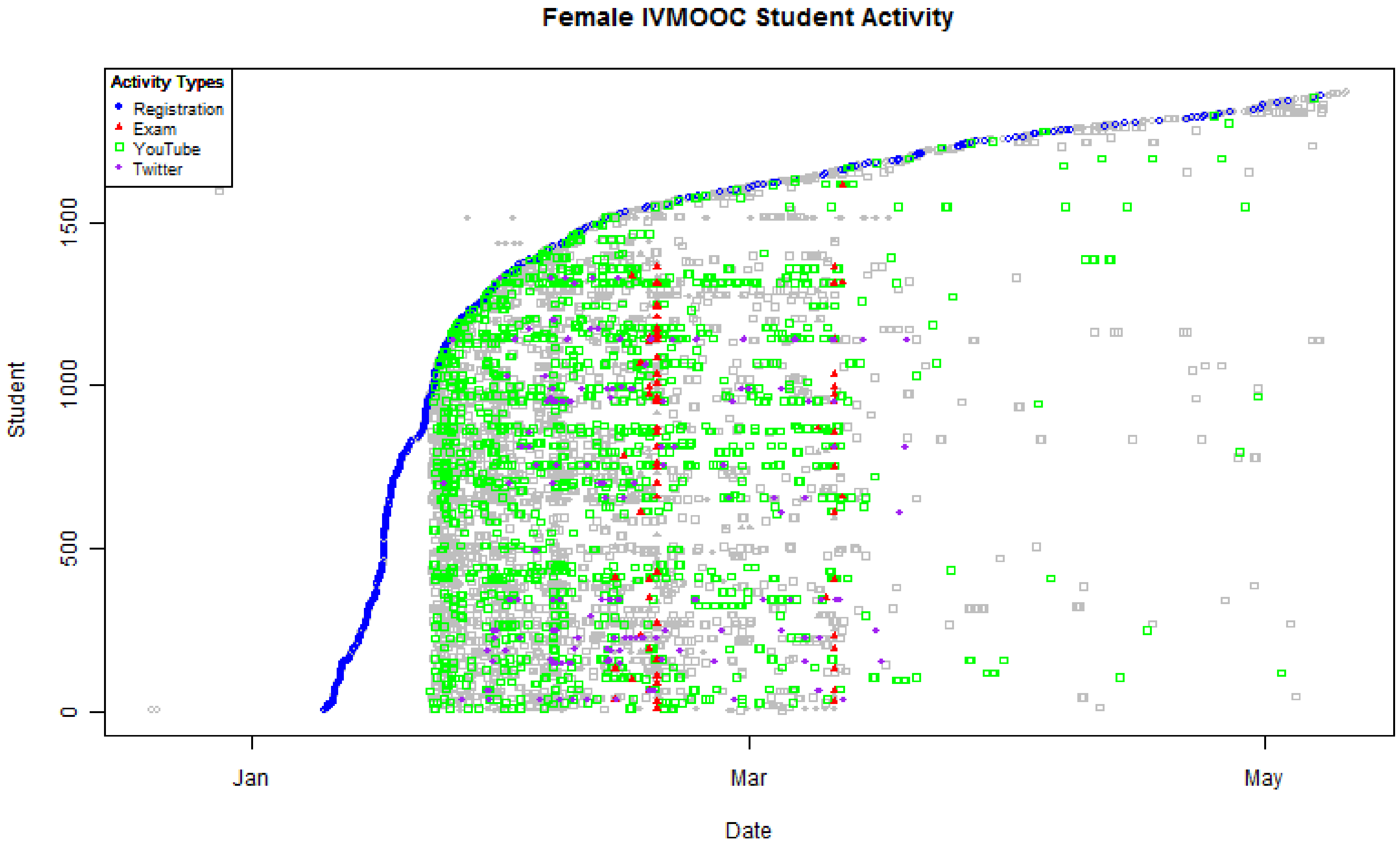
March 11: Final Exam Deadline

Student Registration and Activity



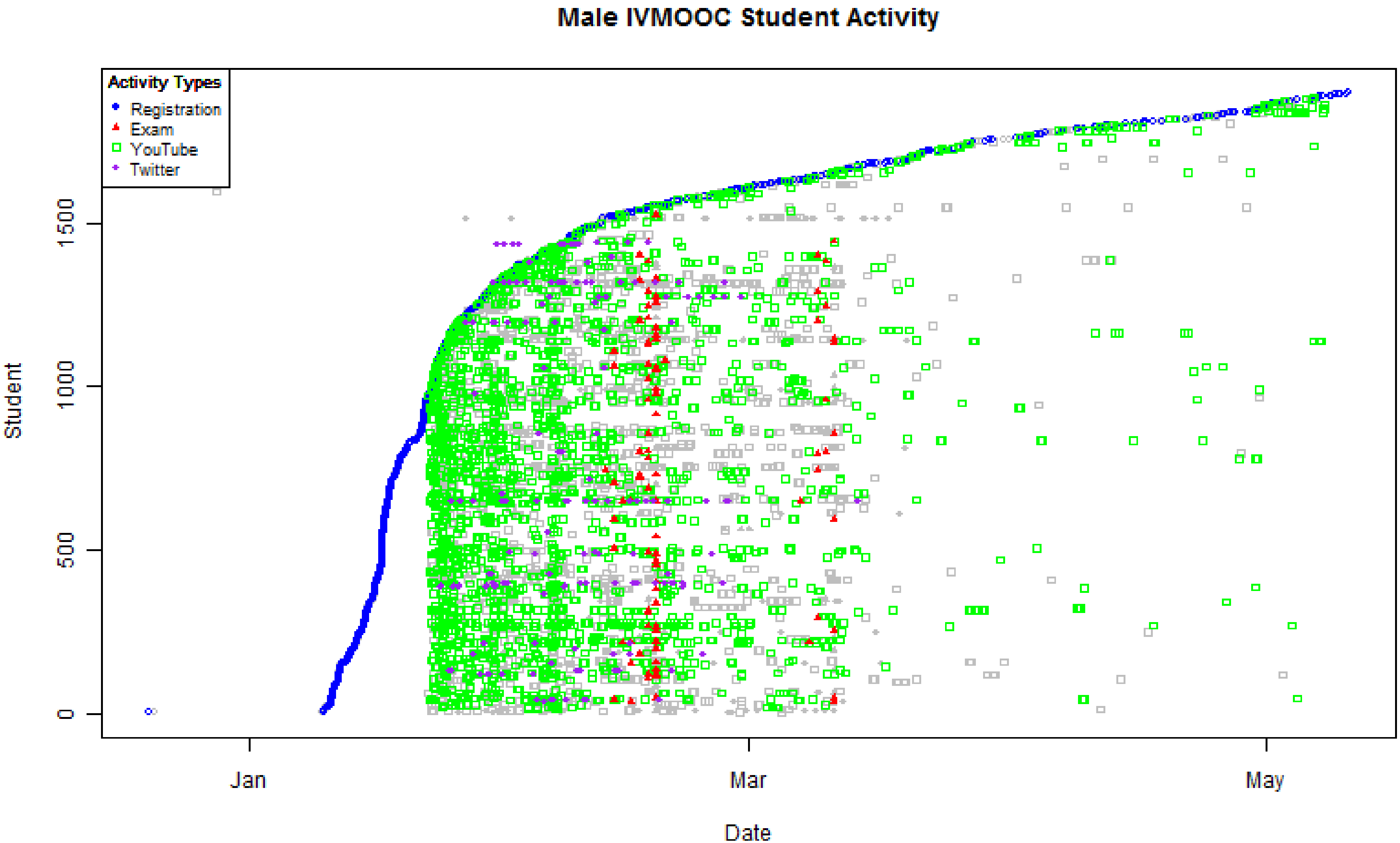
Student Registration and Activity

1215 male students
557 female students

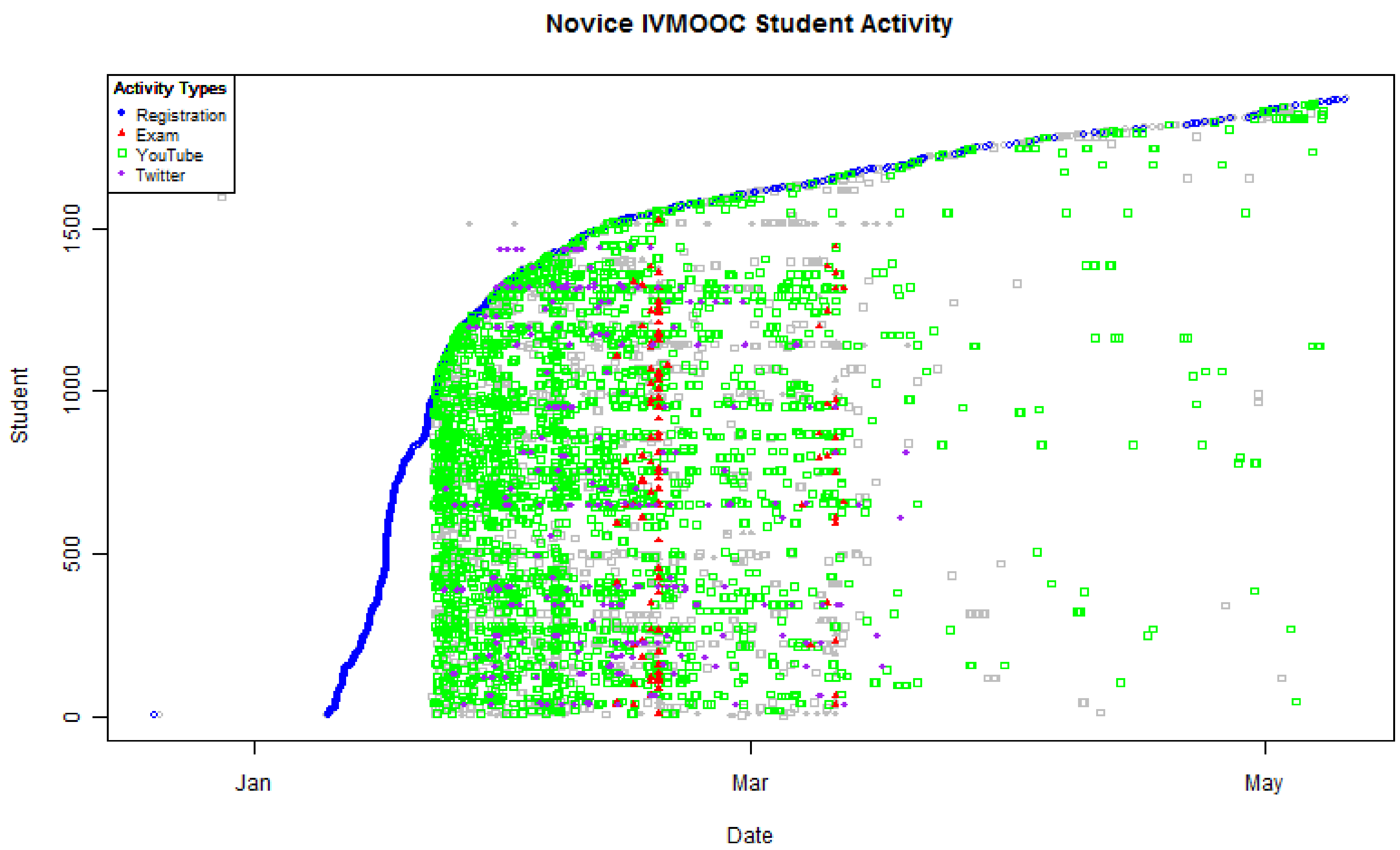


Student Registration and Activity

1215 male students
557 female students

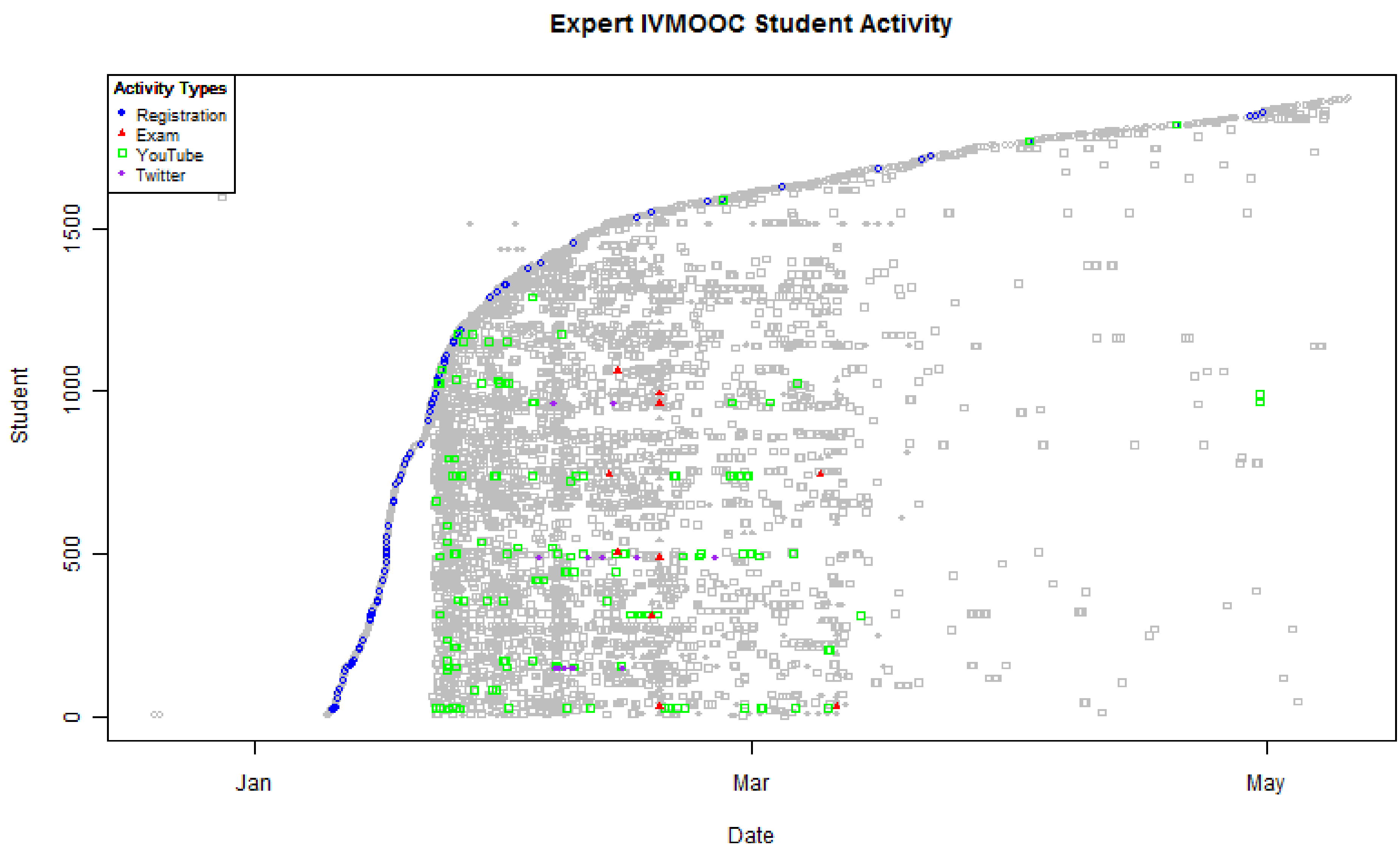


Student Registration and Activity



IVMOOC

Student Registration and Activity



IVMOOC

Student Client Projects: All Interactions

Student Engagement and Performance

Learning Analytics

IVMOOC 2015 Student Group Engagement and Scores

	Pre-Course	Week 1	Week 2	Week 3	Week 4	Midterm	Week 5	Week 6	Week 7	Week 8	Week 9	Final	Curr. Score
IVMOOC	26.05%	38.32%	31.32%	29.96%	27.1%	28.34%	31.07%	24.28%	16.86%	18.23%	13.08%	13.41%	20.87%
Z637-29374	33.01%	52.91%	49.89%	59.22%	50.89%	82.56%	65.04%	49.99%	39.59%	61.63%	54.91%	82.25%	82.4%
Z637-32593	25.08%	54.54%	43.58%	50.67%	53.63%	77.67%	65.7%	59.48%	52.19%	65.71%	47.27%	72.59%	75.13%
Z637-33781	29.33%	55.38%	49.26%	62.18%	77.47%	85%	87.4%	69.8%	55.56%	57.6%	45.69%	70.89%	77.94%

IVMOOC 2015 Student Group Engagement for Midterm

	Midterm	Final	Curr. Score	Overall Engagemer
Student 198	100%	85.33%	92.67%	30.34%
Student 210	100%	84%	92%	33.91%
Student 242	97.14%	98.67%	97.9%	55.89%
Student 265	95.71%	92%	93.86%	82.64%
Student 216	95.71%	24%	59.86%	34.92%
Student 257	94.29%	98.67%	96.48%	68.25%
Student 264	94.29%	89.33%	91.81%	80.47%
Student 262	94.29%	85.33%	89.81%	79.65%

Legends

Engagement

Inactive Active Very Active

0% 25% 50% 75% 100%

Score

F D C B A

0% 60% 70% 80% 100%

Description

The heat map visualization is a representation of student engagement (magenta to blue color scale) and performance (red to green color scale) throughout a course. The visualization has two levels. The top level provides an overview of engagement and performance for groups of students, while the bottom level provides a detailed break out of student engagement statistics for individuals with an identified group.

Custom interactive visualizations of IVMOOC student engagement and performance data, explore functionality online at http://goo.gl/TYixCn

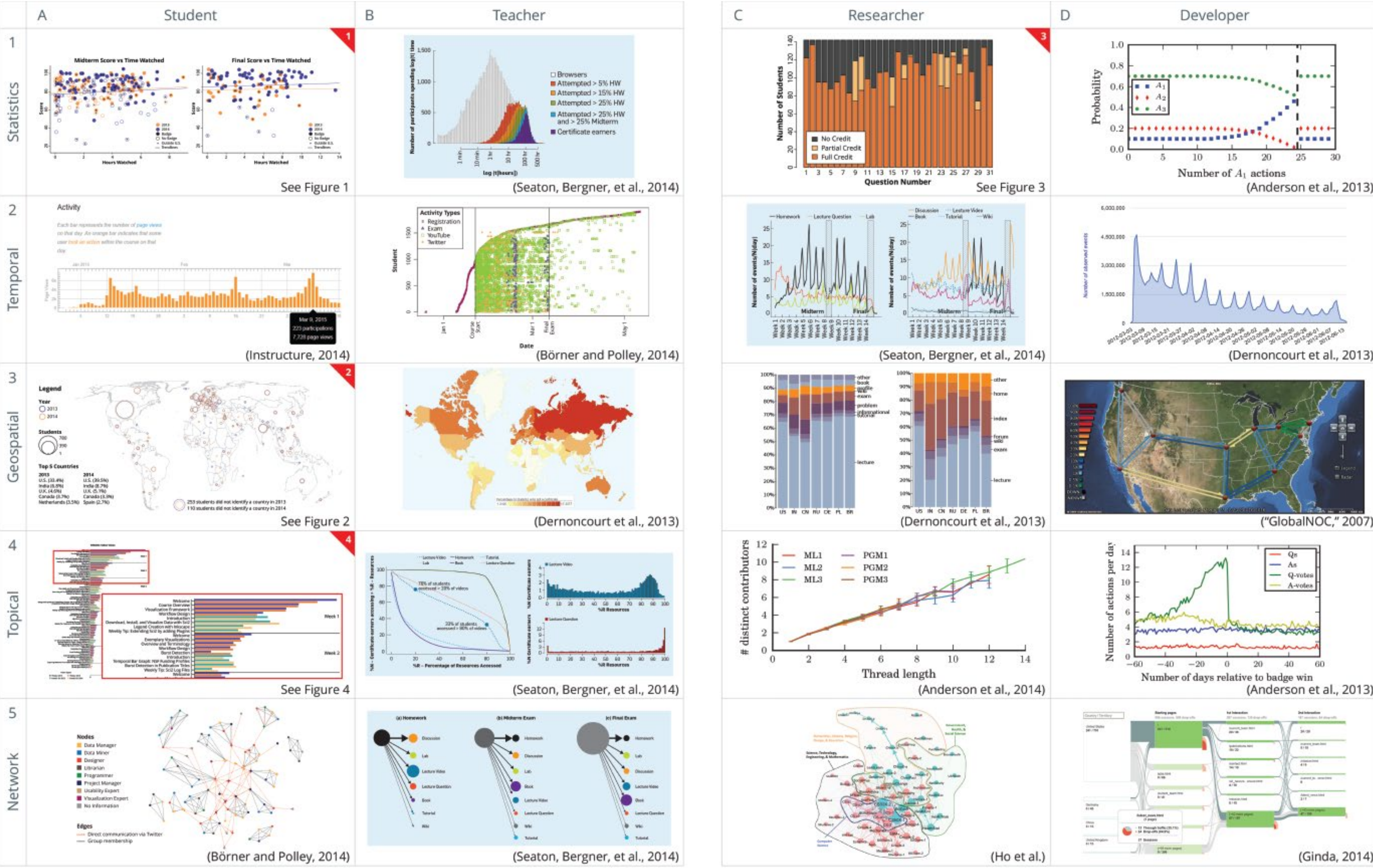


Figure 1: Analysis types vs. user needs.

Emmons, Light, and Börner. "MOOC Visual Analytics: Empowering Teachers, Students, Researchers, and Developers of Massively Open Online Courses". *Journal of the Association for Information Science and Technology (in press).*

Future Work

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

Katy Börner[1], Andreas Bueckle[1], Michael Ginda[1]

[1]Indiana 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

Submission PDF

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

https://www.pnas.org/content/early/2019/01/29/1807180116

Next Generation IVMOOC

Instructor: Victor H. Yngve Distinguished Professor Katy Börner & CNS Team, ISE, SICE, IUB

Duration: *6 weeks x 5 hours = 30 hours (3 CEUs)*

Format: *Online | Theory and Hands-on Instruction, Concept Questions, Graded Assignments, Case Studies, Discussions*

Next Run Starts: *March 25, 2019*

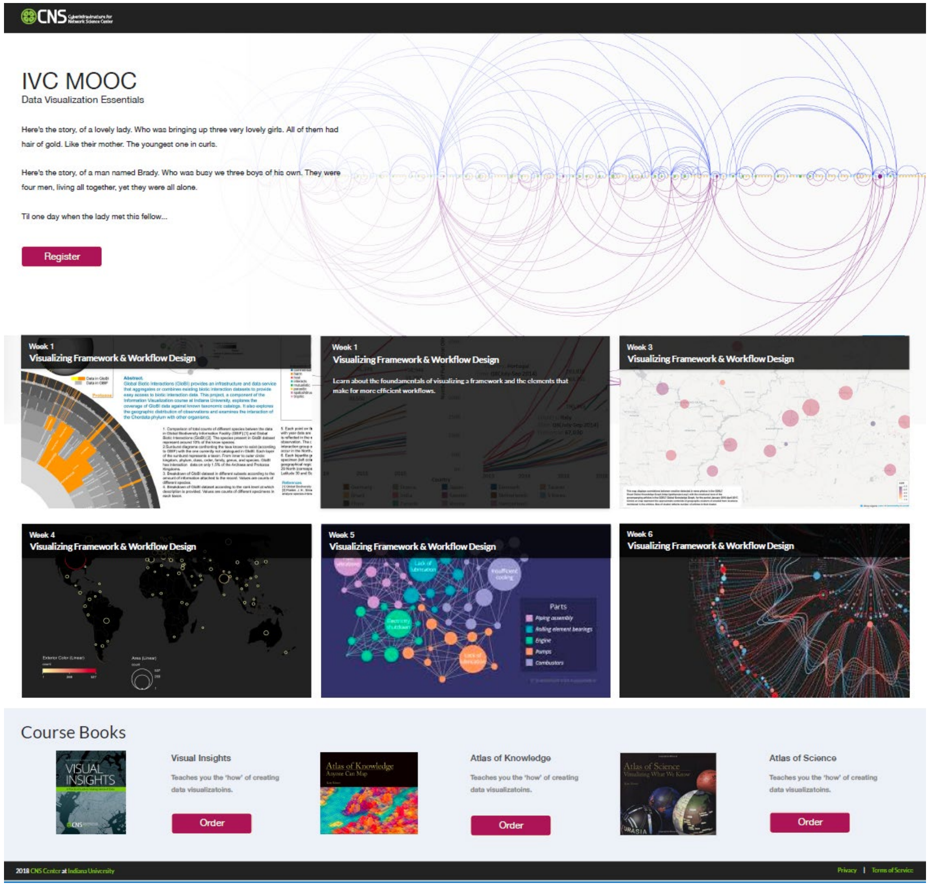
Covers:

Temporal, geospatial, topical (linguistic), network analyses and 60+ visualization types

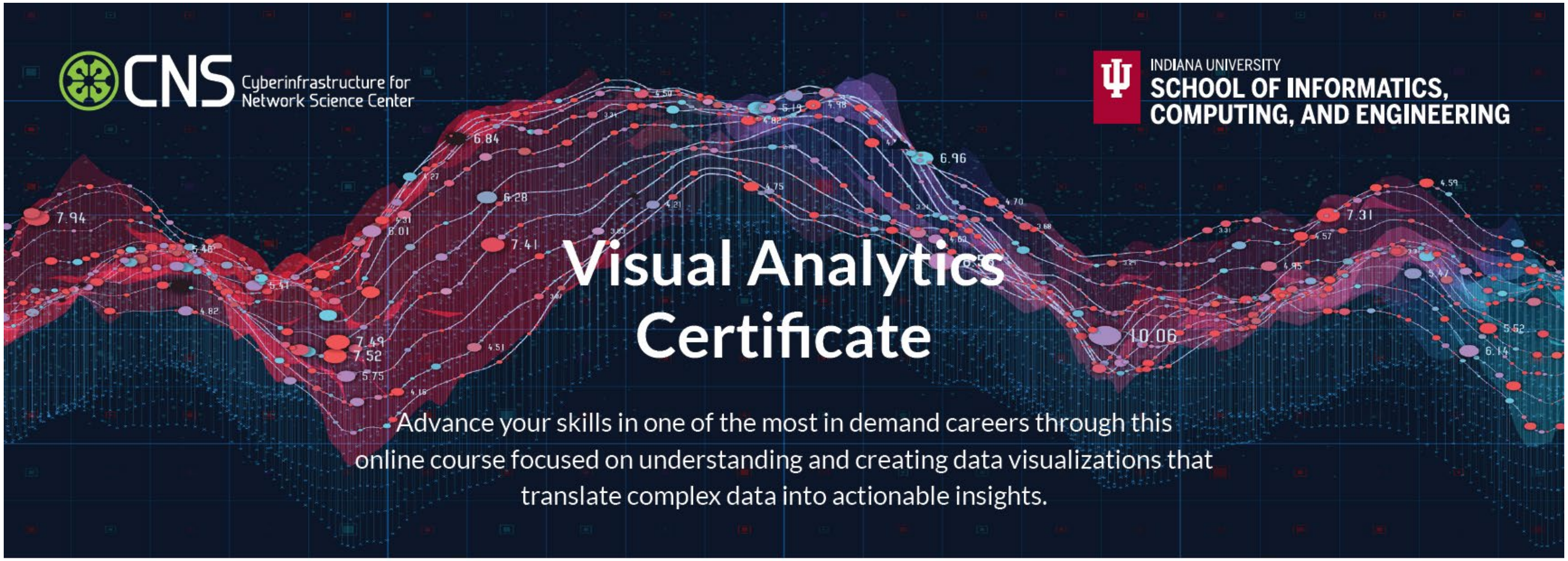
Tools: Tableau, Gephi, BI,

Industry case studies such as

- Acting on customer complaints data.
- Improving communication/traffic flows.
- Understanding web page usage.
- Visualizing online shopping behavior.
- Optimizing supply chains.
- Reducing customer/supplier churn.
- Monitoring emerging R&D areas.
- Workforce development planning.



https://visanalytics.cns.iu.edu



Register: tinyurl.com/VACRegister

Learn from Experts

Connect with industry professionals and leading researchers.

Evolve Yourself

Gain *forever* knowledge and skill-up in powerful data visualization tools.

Make a Difference

Embrace data-driven decision-making in your personal and professional life.

https://visanalytics.cns.iu.edu

Next Generation IVMOOC

Systematic study of how different student cohorts learn best—using **Mechanical Turk formal user studies**, e.g., to optimize horizontal transfer:

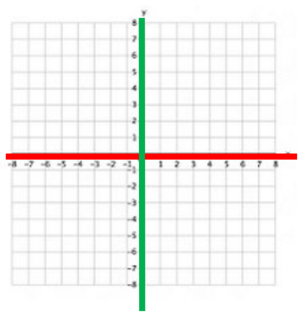
Table
Columns by rows

column row

x	y
0	3
2	11
4	19
6	27
8	35

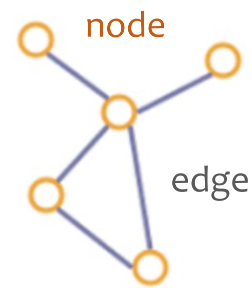
cell

Graph
x-y coordinates
linear/log scale



Map
Latitude/
longitude

Network
Local
similarity



Horizontal Transfer

Next Generation IVMOOC

Systematic study of how different student cohorts learn best—using **Learning Analytics** to optimize scaffolding and learning trajectories:

Scaffolding

Insights
Trends, clusters, outliers

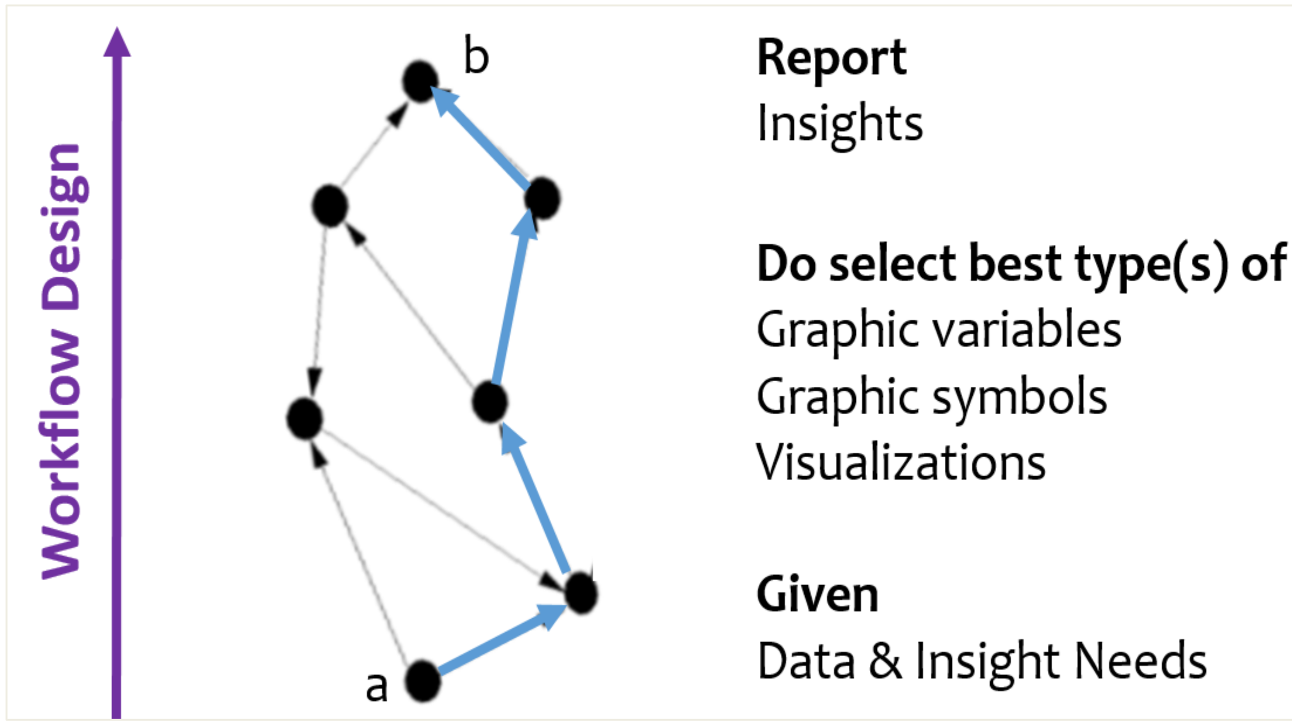
Interactivity
Zoom, filter, details-on-demand

Graphic variables
Position, color, size, shape code

Graphic symbols
Geometric, linguistic, or pictorial

Visualizations
Table, graph, map, network

Insight Needs & Data



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

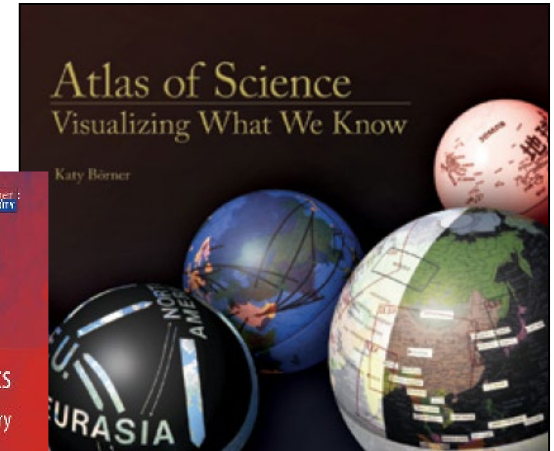
Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1

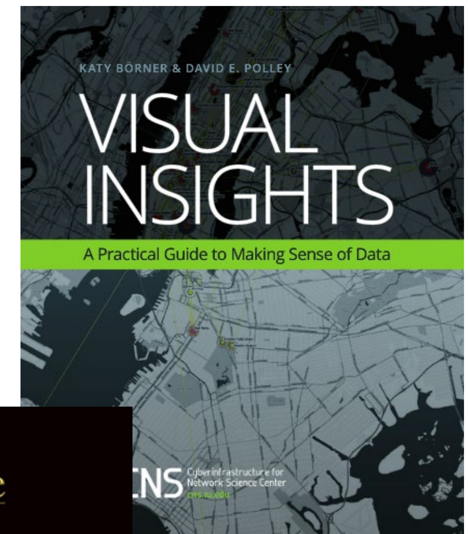
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

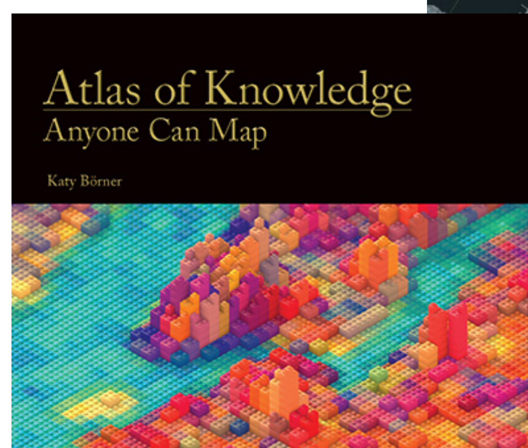
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

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







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