Visualizing Education, Science, and Technology

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

Keynote at SocInfo, https://socinfo2019.qcri.org Doha, Qatar

November 19, 2019
Overview

Data Visualization Literacy (DVL)


Skill Discrepancies


The 15th iteration of the *Places & Spaces: Mapping Science* exhibit ([http://scimaps.org](http://scimaps.org)).
Data Visualization Literacy Framework

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.

**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
Data Visualization Literacy Framework (DVL-FW)

Implemented in Make-A-Vis (MAV) to support learning via horizontal transfer, scaffolding, hands-on learning, etc.
Typology of the Data Visualization Literacy Framework

<table>
<thead>
<tr>
<th>Insight Needs</th>
<th>Data Scales</th>
<th>Analyses</th>
<th>Visualizations</th>
<th>Graphic Symbols</th>
<th>Graphic Variables</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• categorize/cluster</td>
<td>• nominal</td>
<td>• statistical</td>
<td>• table</td>
<td>• geometric symbols</td>
<td>• spatial position</td>
<td>• zoom</td>
</tr>
<tr>
<td>• order/rank/sort</td>
<td>• ordinal</td>
<td>• temporal</td>
<td>• chart</td>
<td>• line</td>
<td>• position</td>
<td>• search and locate</td>
</tr>
<tr>
<td>• distributions (also</td>
<td>• interval</td>
<td>• geospatial</td>
<td>• graph</td>
<td>• area</td>
<td>• retinal</td>
<td>• filter</td>
</tr>
<tr>
<td>outliers, gaps</td>
<td>• ratio</td>
<td>• topical</td>
<td>• map</td>
<td>• surface</td>
<td>• form</td>
<td>• details-on-demand</td>
</tr>
<tr>
<td>• comparisons</td>
<td></td>
<td>• relational</td>
<td>• tree</td>
<td>• volume</td>
<td>• color</td>
<td>• history</td>
</tr>
<tr>
<td>• trends (process and time)</td>
<td></td>
<td></td>
<td>• network</td>
<td>• linguistic symbols</td>
<td>• optics</td>
<td>• extract</td>
</tr>
<tr>
<td>• geospatial</td>
<td></td>
<td></td>
<td></td>
<td>• text</td>
<td>• motion</td>
<td>• link and brush</td>
</tr>
<tr>
<td>• compositions</td>
<td></td>
<td></td>
<td></td>
<td>• numerals</td>
<td></td>
<td>• projection</td>
</tr>
<tr>
<td>(also of text)</td>
<td></td>
<td></td>
<td></td>
<td>• punctuation marks</td>
<td></td>
<td>• distortion</td>
</tr>
<tr>
<td>• correlations/relationships</td>
<td></td>
<td></td>
<td></td>
<td>• pictorial symbols</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Typology of the Data Visualization Literacy Framework

<table>
<thead>
<tr>
<th>Insight Needs</th>
<th>Data Scales</th>
<th>Analyses</th>
<th>Visualizations</th>
<th>Graphic Symbols</th>
<th>Graphic Variables</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>categorize/cluster</td>
<td>nominal</td>
<td>statistical</td>
<td>table</td>
<td>geometric symbols</td>
<td>spatial</td>
<td>zoom</td>
</tr>
<tr>
<td>order/rank/sort</td>
<td>ordinal</td>
<td>temporal</td>
<td>chart</td>
<td>point</td>
<td>position</td>
<td>search and locate</td>
</tr>
<tr>
<td>distributions (also outliers, gaps)</td>
<td>interval</td>
<td>geospatial</td>
<td>graph</td>
<td>line</td>
<td>retinal</td>
<td>filter</td>
</tr>
<tr>
<td>comparisons</td>
<td>ratio</td>
<td>topical</td>
<td>map</td>
<td>area</td>
<td>form</td>
<td>details-on-demand</td>
</tr>
<tr>
<td>trends (process and time)</td>
<td></td>
<td>relational</td>
<td>tree</td>
<td>surface</td>
<td>color</td>
<td>history</td>
</tr>
<tr>
<td>geospatial</td>
<td></td>
<td></td>
<td>network</td>
<td>volume</td>
<td>optics</td>
<td>extract</td>
</tr>
<tr>
<td>compositions (also of text)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>motion</td>
<td>link and brush</td>
</tr>
<tr>
<td>correlations/relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>projection</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>selection</td>
<td>categorize</td>
<td></td>
<td></td>
<td>category</td>
<td></td>
<td></td>
<td></td>
<td>categorize/cluster</td>
</tr>
<tr>
<td>order</td>
<td>rank</td>
<td>ranking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>order/rank/sort</td>
</tr>
<tr>
<td>distribution</td>
<td>distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>distributions (also outliers, gaps)</td>
</tr>
<tr>
<td>compare</td>
<td>nominal comparison &amp; deviation</td>
<td>differences</td>
<td>compare and contrast</td>
<td>compare data values</td>
<td>comparison</td>
<td>comparisons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time series</td>
<td>patterns over time</td>
<td>time</td>
<td>process and time</td>
<td>track rises and falls over time</td>
<td>trend</td>
<td>trends (process and time)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>geospatial</td>
<td>spatial relations</td>
<td>location</td>
<td></td>
<td></td>
<td></td>
<td>geospatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quantity</td>
<td>part-to-whole</td>
<td>proportions</td>
<td>form and structure</td>
<td>see parts of whole, analyze text</td>
<td>composition</td>
<td>compositions (also of text)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>association</td>
<td>correlate</td>
<td>correlation</td>
<td>relationships hierarchy</td>
<td>relations between data points</td>
<td>relationship</td>
<td>correlations/relationships</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Typology of the Data Visualization Literacy Framework

<table>
<thead>
<tr>
<th>Insight Needs</th>
<th>Data Scales</th>
<th>Analyses</th>
<th>Visualizations</th>
<th>Graphic Symbols</th>
<th>Graphic Variables</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• categorize/cluster</td>
<td>• nominal</td>
<td>• statistical</td>
<td>• table</td>
<td>• geometric symbols</td>
<td>• spatial position</td>
<td>• zoom</td>
</tr>
<tr>
<td>• order/rank/sort</td>
<td>• ordinal</td>
<td>• temporal</td>
<td>• chart</td>
<td>point</td>
<td>• search and locate</td>
<td>• filter</td>
</tr>
<tr>
<td>• distributions (also outliers, gaps)</td>
<td>• interval</td>
<td>• geospatial</td>
<td>• graph</td>
<td>line</td>
<td>• details-on-demand</td>
<td>• history</td>
</tr>
<tr>
<td>• comparisons</td>
<td>• ratio</td>
<td>• topical</td>
<td>• map</td>
<td>area</td>
<td>• extract</td>
<td>• extract</td>
</tr>
<tr>
<td>• trends (process and time)</td>
<td></td>
<td>• relational</td>
<td>• tree</td>
<td>surface</td>
<td>• link and brush</td>
<td>• projection</td>
</tr>
<tr>
<td>• geospatial</td>
<td></td>
<td></td>
<td>• network</td>
<td>volume</td>
<td>• projection</td>
<td>• distortion</td>
</tr>
<tr>
<td>• compositions (also of text)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• correlations/relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graphic Variable Types

**Position:** x, y; possibly z

**Form:**
- Size
- Shape
- Rotation (Orientation)

**Color:**
- Value (Lightness)
- Hue (Tint)
- Saturation (Intensity)

**Optics:** Blur, Transparency, Shading, Stereoscopic Depth

**Texture:** Spacing, Granularity, Pattern, Orientation, Gradient

**Motion:** Speed, Velocity, Rhythm
Graphic Symbol Types

<table>
<thead>
<tr>
<th>Graphic Variable Types</th>
<th>Geometric Symbols</th>
<th>Linguistic Symbols</th>
<th>Pictorial Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Position</td>
<td><img src="image1" alt="Point" /></td>
<td><img src="image2" alt="Line" /></td>
<td><img src="image3" alt="Point" /></td>
</tr>
<tr>
<td>Form</td>
<td><img src="image4" alt="Size" /></td>
<td><img src="image5" alt="Shape" /></td>
<td><img src="image6" alt="Size" /></td>
</tr>
<tr>
<td>Value</td>
<td><img src="image7" alt="Value" /></td>
<td><img src="image8" alt="Value" /></td>
<td><img src="image9" alt="Value" /></td>
</tr>
<tr>
<td>Hue</td>
<td><img src="image10" alt="Hue" /></td>
<td><img src="image11" alt="Hue" /></td>
<td><img src="image12" alt="Hue" /></td>
</tr>
<tr>
<td>Saturation</td>
<td><img src="image13" alt="Saturation" /></td>
<td><img src="image14" alt="Saturation" /></td>
<td><img src="image15" alt="Saturation" /></td>
</tr>
<tr>
<td>Granularity</td>
<td><img src="image16" alt="Granularity" /></td>
<td><img src="image17" alt="Granularity" /></td>
<td><img src="image18" alt="Granularity" /></td>
</tr>
<tr>
<td>Pattern</td>
<td><img src="image19" alt="Pattern" /></td>
<td><img src="image20" alt="Pattern" /></td>
<td><img src="image21" alt="Pattern" /></td>
</tr>
<tr>
<td>Blur</td>
<td><img src="image22" alt="Blur" /></td>
<td><img src="image23" alt="Blur" /></td>
<td><img src="image24" alt="Blur" /></td>
</tr>
<tr>
<td>Speed</td>
<td><img src="image25" alt="Speed" /></td>
<td><img src="image26" alt="Speed" /></td>
<td><img src="image27" alt="Speed" /></td>
</tr>
</tbody>
</table>

Qualitative
- Also called: Categorical Attributes
- Identity Channels

Quantitative
- Also called: Ordered Attributes
- Magnitude Channels

Visual Analytics Certificate

Advance your skills in one of the most in demand careers through this six-week (3 CEUs) online course focused on understanding and creating data visualizations that translate complex data into actionable insights.

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
Modelling Science, Technology, Education & Innovation


See also [https://www.pnas.org/modeling](https://www.pnas.org/modeling)
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 Trufio | Sallie Keller | Andrew L. Russell | Guru Madhavan | Azer Bestavros | Jason Owen-Smith

nasonline.org/Sackler-Visualizing-Science
Modeling and Visualizing Science and Technology Developments

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

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

https://www.pnas.org/modeling
Arthur M. Sackler Colloquium on Modeling and Visualizing Science and Technology Developments

- **Twin-Win Model: A human-centered approach to research success**
  Ben Shneiderman

- **Forecasting innovations in science, technology, and education**
  FROM THE COVER
  Katy Börner, William B. Rouse, Paul Trunnio, and H. Eugene Stanley
  PNAS December 11, 2018 115 (50) 12573-12581; first published December 10, 2018. https://doi.org/10.1073/pnas.1818750115

- **How science and technology developments impact employment and education**
  Wendy Martinez

- **Scientific prize network predicts who pushes the boundaries of science**
  Yifang Ma and Brian Uzzi

- **The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms**
  C. Jara-Figueroa, Bogang Jun, Edward L. Glaeser, and Cesar A. Hidalgo
  PNAS December 11, 2018 115 (50) 12646-12653; first published December 10, 2018. https://doi.org/10.1073/pnas.1800475115
Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy
Katy Börner, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James A. Evans

Changing demographics of scientific careers: The rise of the temporary workforce
Staša Milojević, Filippo Radicchi, and John P. Walsh

The chaperone effect in scientific publishing
Vedran Sekara, Pierre Deville, Sebastian E. Ahnert, Albert-László Barabási, Roberta Sinatra, and Sune Lehmann
PNAS December 11, 2018 115 (50) 12603-12607; first published December 10, 2018. https://doi.org/10.1073/pnas.18047115

Modeling research universities: Predicting probable futures of public vs. private and large vs. small research universities
William B. Rouse, John V. Lombardi, and Diane D. Craig
PNAS December 11, 2018 115 (50) 12582-12589; first published December 10, 2018. https://doi.org/10.1073/pnas.1807174115

and more ...
Skill Discrepancies Between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy

- Data and Crosswalks
- Causal Analyses
- MaxMatch for NLP
- Visualizations

Börner, Katy; Scrivner, Olga; Gallant, Mike; Shutian Ma, Xiaohong Liu, Keith Chewning, Lingfei Wue, and James A. Evans. 2018. “Skill Discrepancies Between Research, Education, and Jobs: Revealing the Critical Need to Supply Soft Skills for the Data Economy.” PNAS 115(50): 12630-12637.
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

Fig. 1. The interplay of job market demands, educational course offerings, and progress in S&T as captured in publications. Color-coded mountains (+) and valleys (−) indicate different skill clusters. For example, skills related to Biotechnology might be mentioned frequently in job descriptions and taught in many courses, but they may not be as prevalent in academic publications. In other words, there are papers that mention these skills, but labor demand and commercial activity might be outstripping publication activity in this area. The numbers of jobs, courses, and publications that have skills associated and are used in this study are given on the right.
Biotechnology

Jobs

Courses

Science & Technology
Biotechnology

Jobs

Courses

Science & Technology
Stakeholders and Insight Needs

• **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 balance of timely and timeless knowledge (to get a job vs. learn how to learn) should I teach? How to innovate in teaching and maintain job security or tenure?

• **Universities**: What programs should be created? What is my competition doing? How do I tailor programs to fit local needs?

• **Science Funders**: How can S&T investments improve short- and long-term prosperity? Where will advances in knowledge also yield advances in skills and technology?

• **Employers**: What skills are needed next year and in 5 and 10 years? Which institutions produce the right talent? What skills does my competition list in job advertisements?

• **Economic Developers**: What critical skills are needed to improve business retention, expansion, and recruitment in a region?

  What is ROI of my time, money, compassion?
Urgency

• 35% of UK jobs, and 30% in London, are at high risk from automation over the coming 20 years.
  https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/uk-futures/london-futures-agiletown.pdf

• The aerospace industry and NASA have a disproportionately large percentage of workers aged 50 and older compared to the national average, and up to half of the current workforce will be eligible for retirement within the coming five years.
  https://www.aiaa.org/uploadedFiles/Issues_and_Advocacy/Education_and_Workforce/Aerospace%20Workforce-%20030112.pdf

Skill Discrepancies Between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy

- Data and Crosswalks
- MaxMatch for NLP
- Causal Analyses
- Visualizations

Datasets Used

Job advertisements by Burning Glass posted between Jan 2010-Dec 2016.

Web of Science publications published Jan 2010-Dec 2016.

Course descriptions from the Open Syllabus Project acquired in June 2018 for courses offered in 2010-2016.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>#Records</th>
<th>#Records with skills</th>
<th>#Records without skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Courses</td>
<td>3,062,277</td>
<td>2,744,311</td>
<td>54,733</td>
</tr>
<tr>
<td>All Jobs</td>
<td>132,011,926</td>
<td>121,073,950</td>
<td>10,937,976</td>
</tr>
<tr>
<td>DS/DE Jobs</td>
<td>69,405</td>
<td>65,944</td>
<td>3,461</td>
</tr>
<tr>
<td>All Publications</td>
<td>15,691,162</td>
<td>1,048,575</td>
<td>14,642,587</td>
</tr>
<tr>
<td>DSDE Publications</td>
<td>1,048,575</td>
<td>807,756</td>
<td>240,819</td>
</tr>
</tbody>
</table>
Fig. 2. Basemap of 13,218 skills. In this map, each dot is a skill, triangles identify skill clusters, and squares represent skill families from the Burning Glass (BG) taxonomy. Labels are given for all skill family nodes and for the largest skill cluster (NA) to indicate placement of relevant subtrees. Additionally, hard and soft skills are overlaid using purple and orange nodes, respectively; node area size coding indicates base 10 log of skill frequency in DS/DE jobs. Skill area computation uses Voronoi tessellation.
**Fig. 3.** Basemap of 13,218 skills with overlays of skill frequency in jobs, courses, and publications. This figure substantiates the conceptual drawing in Fig. 1 using millions of data records. Jobs skills are plotted in blue, courses are in red, and publications are in green. Node area size coding indicates base 10 log of skills frequency. The top 20 most frequent skills are labeled, and label sizes denote skill frequency.
Fig. 4. Burst of activity in DS/DE skills in jobs and publications. Each burst is rendered as a horizontal bar with a start and an end date; skill term is shown on the left. Skills that burst in jobs are blue; skills bursting in publications are green. Seven skills burst in both datasets during the same years and are shown in gray. HRMS stands for human resources management system, and Maximo is an IBM system for managing physical assets.
Kullback-Leibler divergence

Fig. 5. Structural and dynamic differences between skill distributions in jobs, courses, and publications for 2010–2013 and 2014–2016. (A) Poincaré disks comparing the centrality of soft skills (orange) and hard skills (purple) across jobs, courses, and publications. (B) KL divergence matrix for jobs, courses, and publications in 2010–2013 and 2014–2016. (C) The most surprising skills in publications and jobs; R is a scripting language, VTAM refers to the IBM Virtual Telecommunication Access Method application, VS is the integrated development environment Visual Studio, and SAS is a data analytics software.
Fig. 6. Strength of influence mapping. Top 200 most frequent skills in jobs (blue) and in publications (green) plotted on the skills basemap from Fig. 2. Arrows represent skills with significant Granger causality (P value < 0.05). Line thickness and label size indicate skill frequency. The direction and thickness of each arrow indicate the F-value strength and direction.
Fig. 7. Multivariate Hawkes Process influence network of DS/DE skills within job advertisements 2010–2016. Each of the 45 nodes represents a top-frequency skill (29 soft and 16 hard skills) with a strong influence edge from/to other skill(s) in job advertisements between 2010 and 2016. Node and label size correspond to the number of times that the skill appeared in a job advertisement. Thickness of the 75 directed edges indicates influence strength.
Fig. 7. Hawkes influence network of DS/DE skills within job advertisements 2010–2016. Each of the 45 nodes represents a top-frequency skill (29 soft and 16 hard skills) with a strong influence edge from/to other skill(s) in job advertisements between 2010 and 2016. Node and label size correspond to the number of times that the skill appeared in a job advertisement. Thickness of the 75 directed edges indicates influence strength.
Results

• Novel cross-walk for mapping publications, course offerings, and job via skills.

• Timing and strength of burst of activity for skills (e.g., Oracle, Customer Service) in publications, course offerings, and job advertisements.

• Uniquely human skills such as communication, negotiation, and complex service provision are currently underexamined in research and undersupplied through education for the labor market in an increasingly automated and AI economy.

• The same pattern manifests in the domain of DS/DE where teamwork and communication skills increase in value with greater demand for data analytics skills and tools.

• Skill demands from industry are as likely to drive skill attention in research as the converse.
NSF RAISE: C-Accel Pilot - Track B1: Analytics-Driven Accessible Pathways To Impacts-Validated Education (ADAPTIVE)

**Goal:** Development of data-driven tools to support the tens of millions of US workers whose jobs are being transformed by Artificial Intelligence (AI) and automation.

The project will demonstrate how labor market and course syllabi data, learning analytics, and insights on transferability of learned skills can be combined and visualized in novel ways to support a learner’s decision-making about, sustained engagement in, and application to their job of professional skills acquired through education and job-related training.

**Team B-6656:** Katy Börner, Indiana University, Ariel Anbar, Arizona State University, Kemi Jona, Northeastern University, Martin Storksdieck and Heather Fischer, Oregon State University
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

Places & Spaces: Mapping Science Exhibit

1st Decade (2005-2014)

Maps

2nd Decade (2015-2024)

Macrosopes

100
MAPS in large format, full color, and high resolution.

248
MAPMAKERS from fields as disparate as art, urban planning, engineering, and the history of science.

43

MACROSCOPE MAKERS including one whose job title is "Truth and Beauty Operator."

20
MACROSCOPES for touching all kinds of data.

382
DISPLAY VENUES from the Cannes Film Festival to the World Economic Forum.

354
PRESS ITEMS including articles in Nature, Science, USA Today, and Wired.

http://scimaps.org
Map of Scientific Collaborations from 2005-2009

VII.6 Stream of Scientific Collaborations Between World Cities - Olivier H. Beauchesne - 2012
Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then clustered on a two-dimensional map using a force-directed simulated annealing algorithm. This analysis creates an interactive environment for assessing grant relevance to research categories and to NIH Institutes in which grants are localized.
The Structure of Science

1. Mathematics is the starting point, the pivot of all sciences. It lies at the corner of the map. Computer Science, Electrical Engineering, and others are applied sciences that draw upon knowledge in Mathematics and Physics. These three disciplines make up a small portion of a large universe from pure science (Mathematics) to another (Physics) through multiple disciplines. Although they are specialists in different fields of research, they are connected to the mainstream of research communities that link them. Bonds indicate interdisciplinary research.

2. Research is highly concentrated in Physics and Chemistry. These disciplines have fewer, but very distinct, bonds of research communities. In this way, the thickness of these bonds indicates an active area of interdisciplinary research, which suggests that the boundaries between Physics and Chemistry are not as distinct as one might assume.

3. The Life Sciences, including Biology and Biochemistry, are less concentrated than Chemistry and Physics. Results of this research can be seen between the larger areas in the Life Sciences. For instance, between Biology and Microbiology, and between Biology and Environmental Science. Biochemistry is very important in that it is a large discipline that few studies link to disciplines from other areas of the map, including Physics, Chemistry, Neuroscience, and General Medicine. It is perhaps the most interdisciplinary of the sciences.

We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between various disciplines and research topics in science. This allows us to map the structure of science.

One of the first maps of science was developed at the Institute for Scientific Information over 30 years ago. It identified 41 areas of science from the citation patterns in 1,500 scientific papers. That effort was not entirely successful since it didn’t cover enough of science to accurately define its structures.

Things are different today. We have enormous computing power and advanced visualization software that make mapping the structure of science possible. This galaxy map of science, which was generated at the National Laboratories using an advanced graphical display (VizMap) from the citation patterns in 90,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 707 research communities whose science was studied. A research community is a group of papers that are written on the same research topic in a given year. Over time, communities can be born, continue, split, merge, or die.

The map of science can be used as a tool for science strategy. It is the terrain in which organizations and nations must navigate their scientific capabilities. Information about the scientific and economic impact of their research community allows policy makers to decide where to allocate funds, equipment, scientists, and resources.

We have shown the map as an educational tool. For children, the visualized relationships and areas of science can be explored with a concrete map showing how math, physics, chemistry, biology, and social sciences interact. Advanced areas of science can be located and neighboring areas can be explored.

Nanotechnology

Most research communities in nanotechnology are concentrated in Physics, Chemistry, and Materials Science. However, many disciplines in the Life and Medical Sciences also have nanotechnology applications.

Proteomics

Research communities in proteomics are concentrated in Biochemistry. In addition, there is a heavy focus in the look across the other disciplines, such as Cell Biology and Microbiology. Most of the research communities are widely dispersed among the Life and Medical Sciences.

Pharmacogenomics

Pharmacogenomics is relatively new, but it is trying to use information in Medicine and Medical Genomics to improve the ability to be related in Medicine and two communities in the Social Sciences.
VII.10 History of Science Fiction - Ward Shelley - 2011
Check out our **Zoom Maps** online!

Visit [scimaps.org](http://scimaps.org) and check out all our maps in stunning detail!
THE MEGAREGIONS OF THE US

This is the **Roanoke** (Raleigh) megaregion.

Iteration XII (2016)
Macroscopes for Making Sense of Science

Iteration XIII (2017)
Macroscopes for Playing with Scale

Iteration XIV (2018)
Macroscopes for Ensuring our Well-being

Iteration XV (2019)
Macroscopes for Tracking the Flow of Resources
Acknowledgments

Exhibit Curators

The exhibit team: Lisel Record, Katy Börner, and Todd Theriault.

Exhibit Advisory Board

http://scimaps.org

Plus, we thank the more than 250 authors of the 100 maps and 16 interactive macrosopes.
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
Indiana University Bloomington will host the
International Society of Scientometrics & Informetrics Conference (ISSI)
in Summer 2023
References


