

Data Visualization Best Practices

Katy Börner @katycns

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



*Seed Networks Computational Biology Meeting
Virtual Event*

April 6, 2021



2020 Flashback

Human Reference Atlas: Anatomical Structures, Cell Types & Biomarkers

Katy Börner | @katycns
Victor H. Yngve Distinguished Professor of
Intelligent Systems Engineering & Information Science
Luddy School of Informatics, Computing, and Engineering
Indiana University, Bloomington, IN



Seed Networks 2020 Annual Meeting
Virtual Event

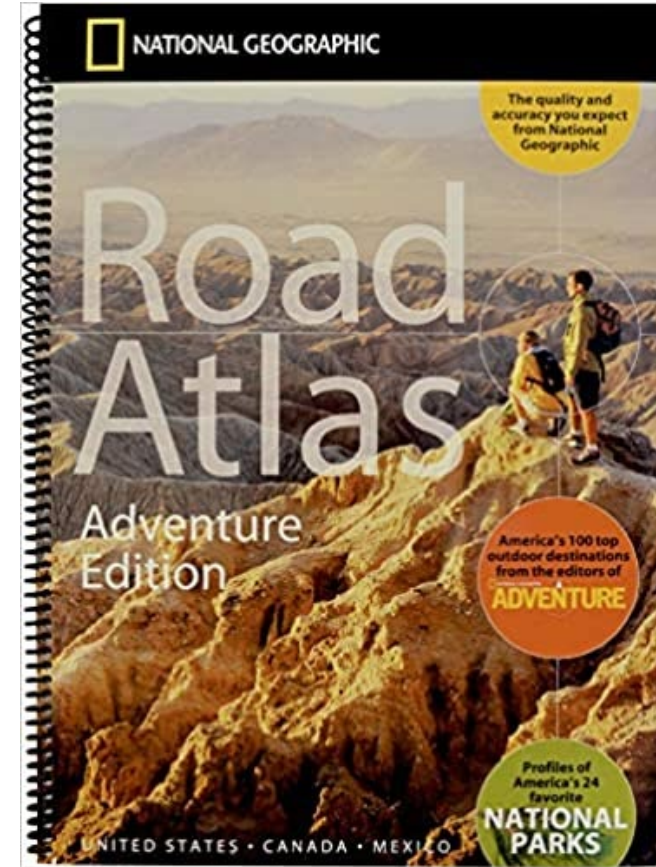
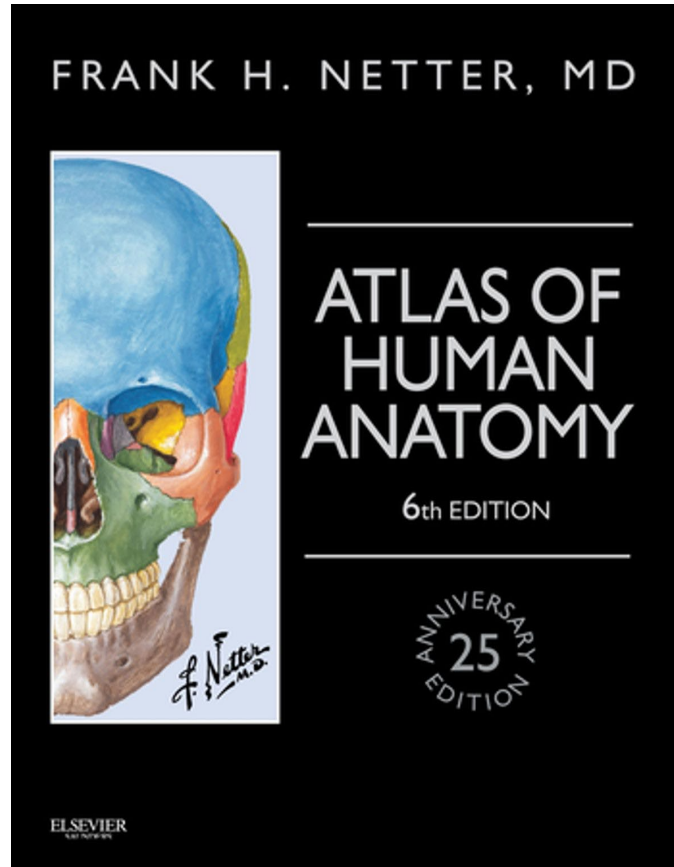
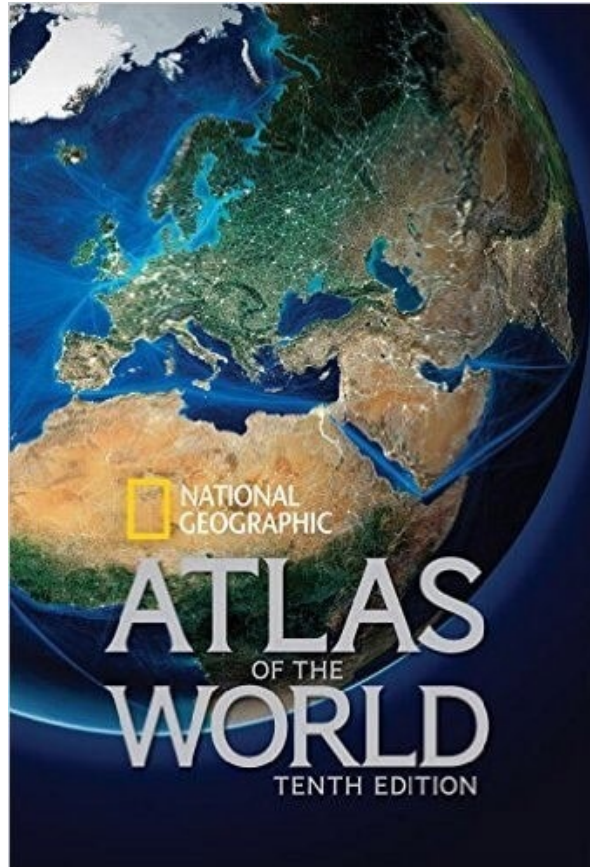
November 18, 2020



2020 Flashback

An **atlas** is an oversized, bound book of maps.

It has descriptive text, an index, possibly other data visualizations.





2020 Flashback

An **human cell atlas** might show a landscape of all cells, or

Maps of cells per tissue type/anatomical structure.

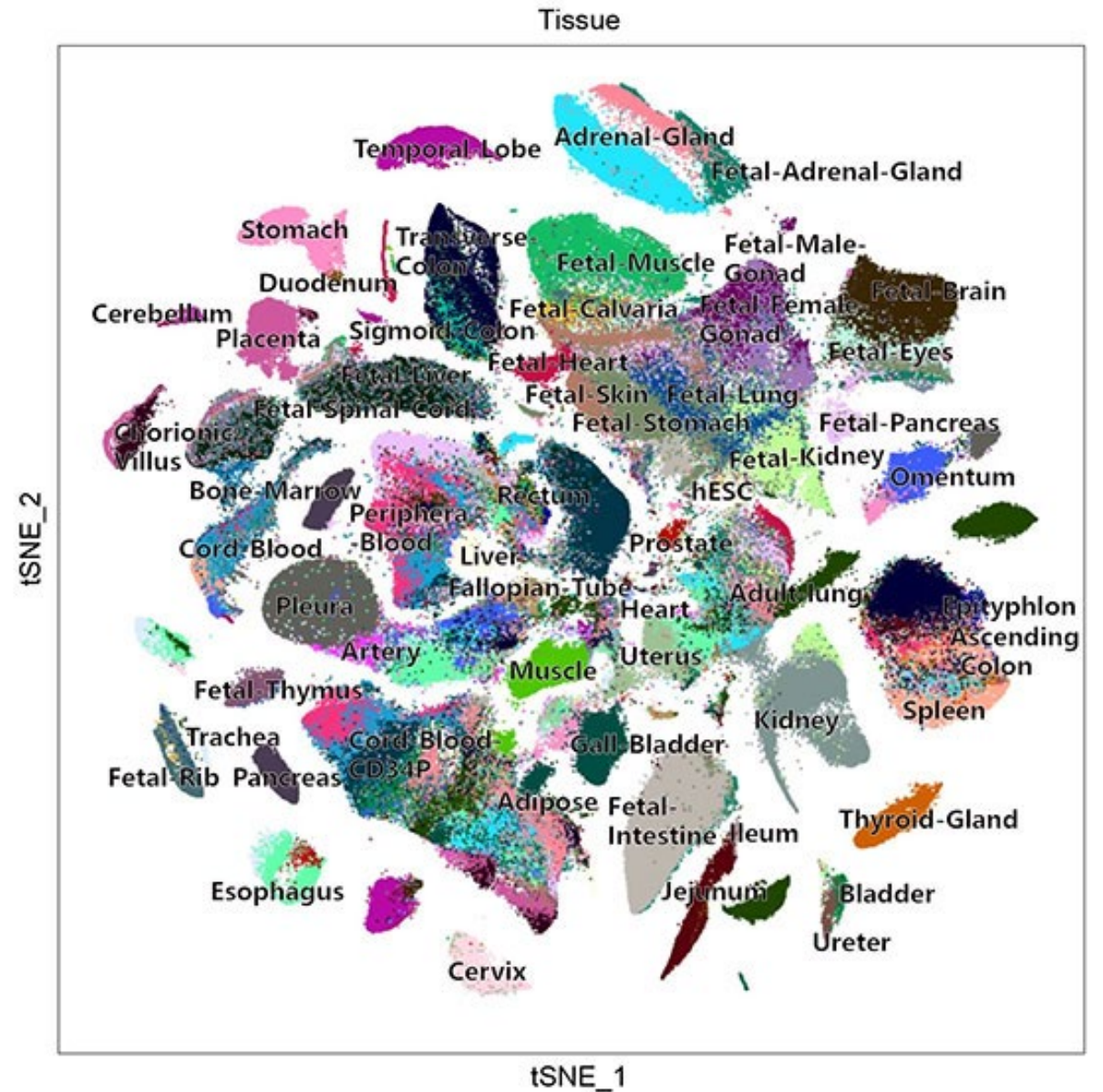
Article | Published: 25 March 2020

Construction of a human cell landscape at single-cell level

Xiaoping Han , Ziming Zhou, [...] Guoji Guo 

Nature **581**, 303–309(2020) | [Cite this article](#)

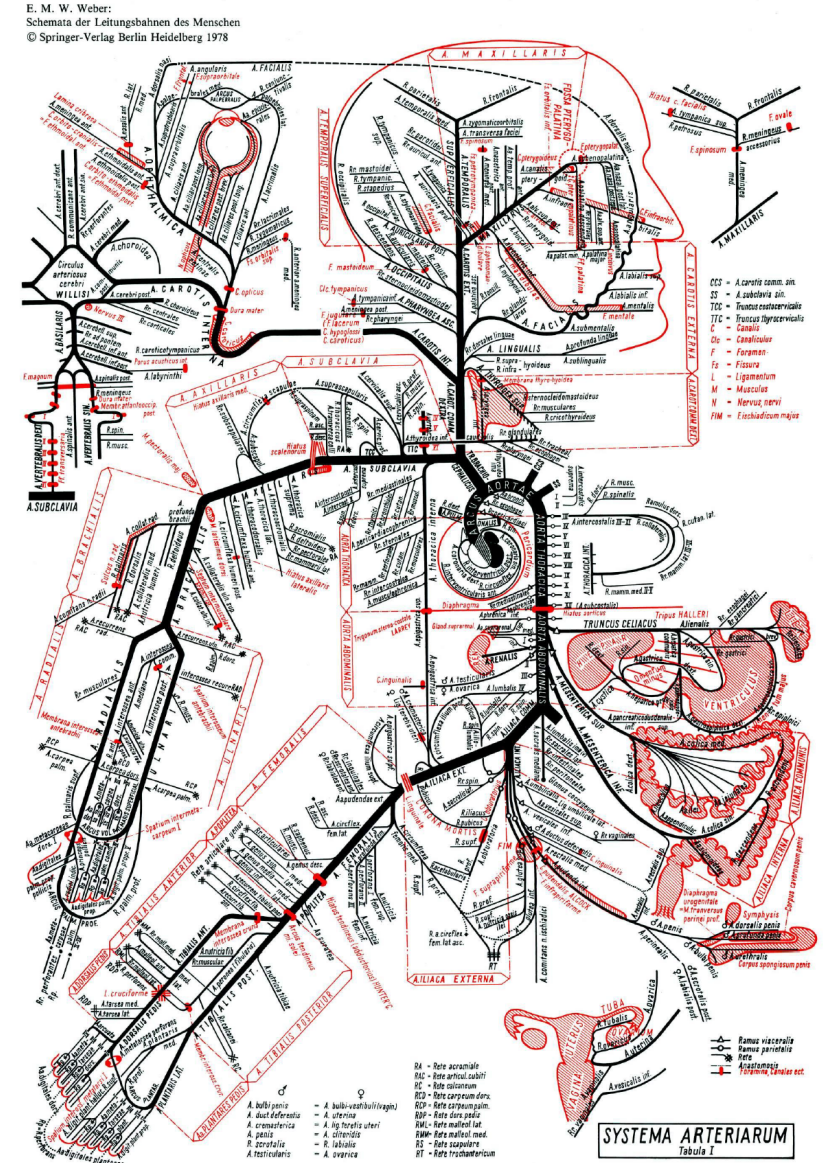
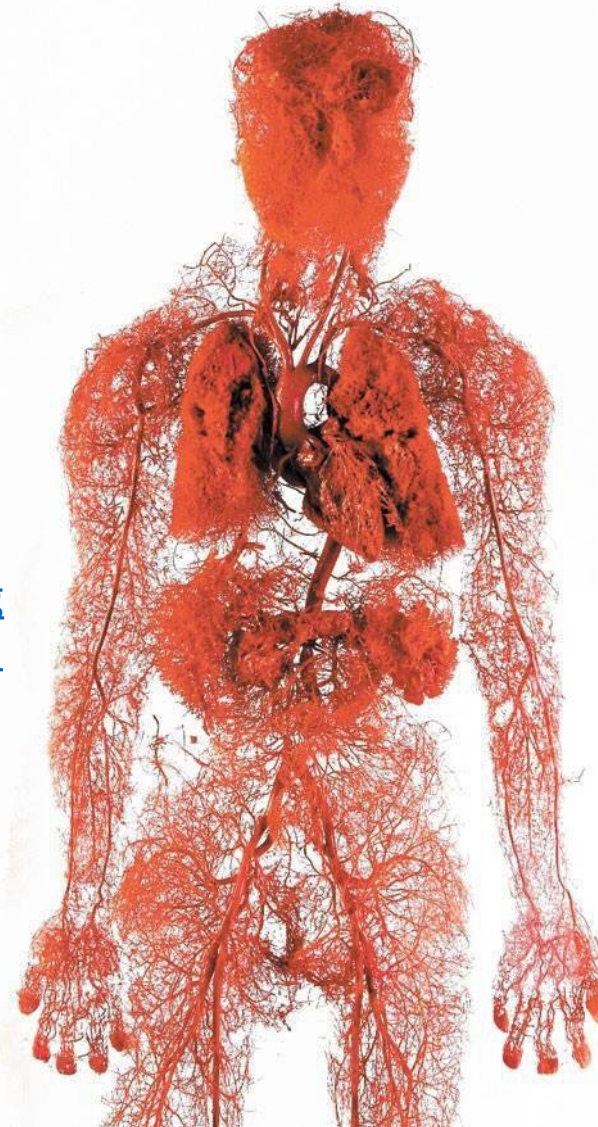
55k Accesses | 32 Citations | 409 Altmetric | [Metrics](#)



2020 Flashback

A human reference atlas might use human anatomy as a 'basemap,' or an abstract space.

Weber, Griffin M, Yingnan Ju, and Katy Börner. 2020. "[Considerations for Using the Vasculature as a Coordinate System to Map All the Cells in the Human Body](#)". *Frontiers in Cardiovascular Medicine* 7 (29): doi: 10.3389/fcvm.2020.00029.



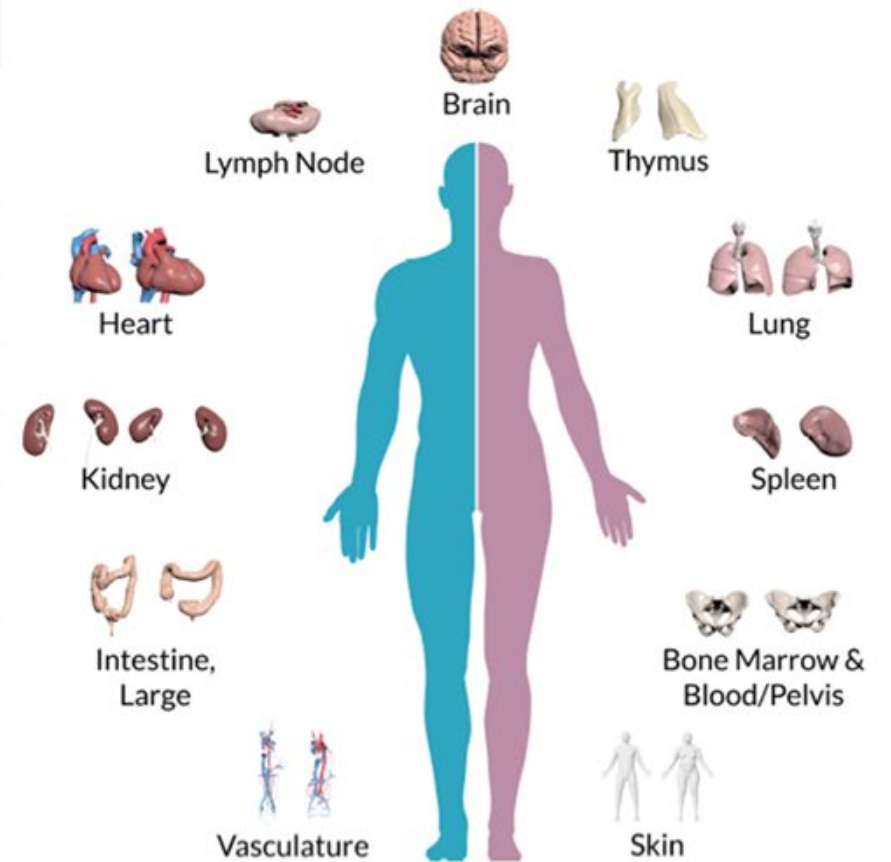
<https://bodyworlds.com>

Weber, 1978

2020 Flashback

ASCT+B tables are used to construct a **human reference atlas** across 16 consortia. Register at WG [Expert Registration](#) to receive invites/info.

Organ	#AS	#CT	#B Total	#BG	#BP	#AS-AS	#AS-CT	#CT-B
Bone Marrow & Blood/Pelvis	3	46	327	201	126	2	70	710
Brain	187	127	254	254	0	187	127	330
Heart	52	25	48	48	0	61	164	78
Intestine, Large	65	69	94	88	6	389	1,361	197
Kidney	68	63	152	152	0	67	59	257
Lung	161	92	176	172	4	1,633	12,094	286
Lymph Node	41	49	266	108	158	62	135	544
Skin	16	42	70	0	70	17	19	105
Spleen	46	66	255	80	145	68	172	414
Thymus	25	41	511	388	123	38	180	657
Vasculature	870	2	1	1	0	869	606	2
Totals:	1,534	622	2,154	1,492	632	3,393	14,987	3,580

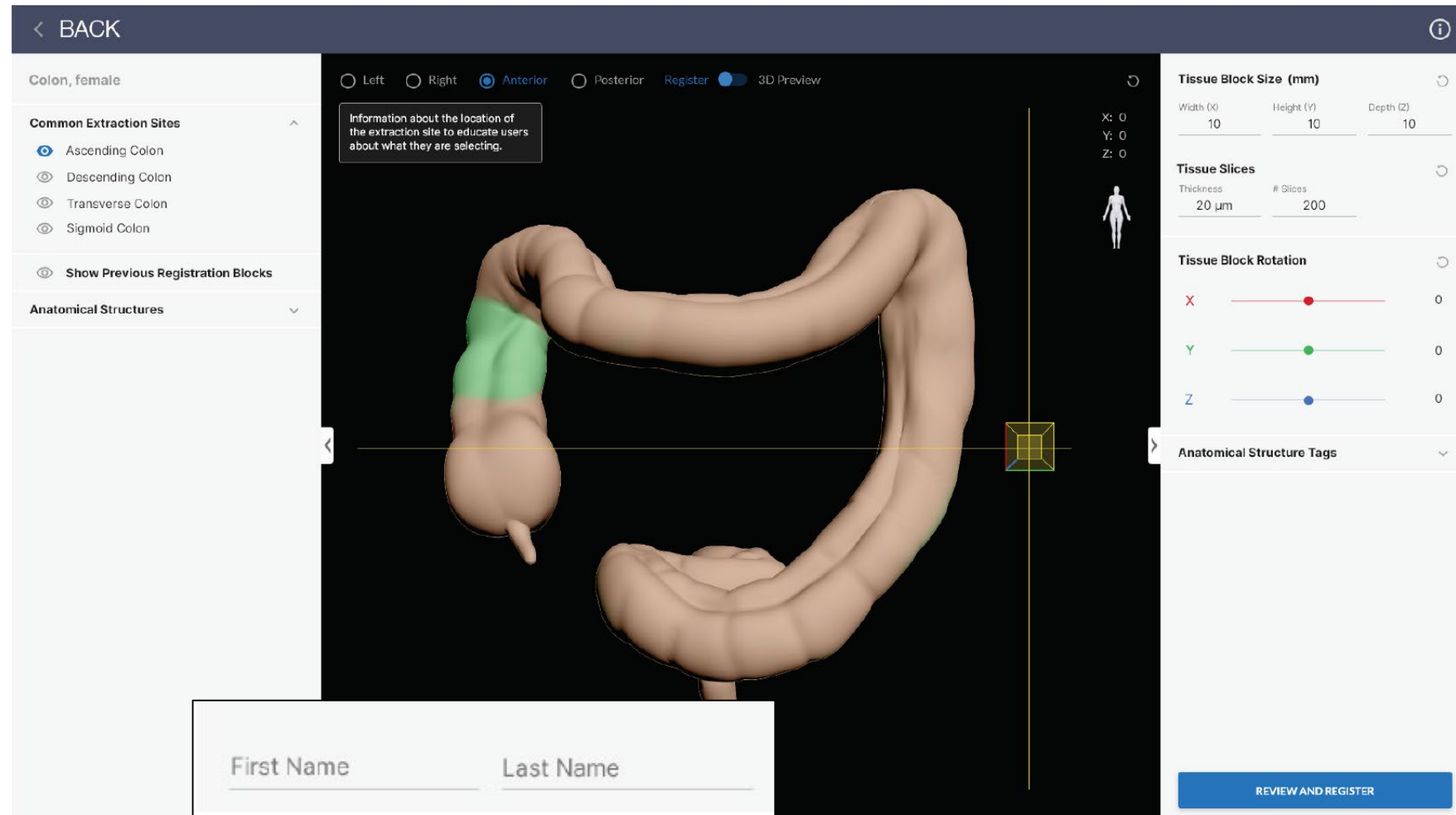


CCF Registration User Interface (RUI) v1.0.0






2020 Flashback

New Features:

- Organ carousel with 4 reference organs
- Support for tissue extraction sites
- Expanded ontology
- Semantic annotation via collision detection & manual annotation
- Support for non-HuBMAP usage



First Name _____ Last Name _____

<      >

Male Female

<https://hubmap-ccf-ui.netlify.app/rui/>

CCF Exploration User Interface (EUI)

2020 Flashback

HuBMAP Sex: Both Age: 1-110 BMI: 13-83

Search ontology terms ...

- body
 - heart
 - lung
 - kidney
 - right kidney
 - left kidney
 - kidney capsule
 - cortex of kidney
 - renal medulla
 - renal column
 - renal pyramid
 - hilum of kidney
 - kidney interstitium
 - kidney calyx
 - renal pelvis
 - ureter
 - renal papilla
 - renal fat pad
 - nephron

body

- 2 Centers
- 27 Donors
- 41 Samples

	Female, Age 14, BMI 14.7 HBM894.MPVN.828 TMC-Florida First case collected. Incomplete d...	
	Male, Age 18, BMI 27.1 HBM436.GHWX.449 TMC-Florida section is 190um from block surface	
	Male, Age 56, BMI 32.5 HBM696.XTVL.498 TMC-Vanderbilt Age 56, White Male	
	Male, Age 53, BMI 26.5 HBM652.VRLD.292 TMC-Vanderbilt Age 53, Black Male	
	Male, Age 58, BMI 22.0 HBM477.CJKM.888 TMC-Vanderbilt 107-111	
	Male, Age 18, BMI 25.5 HBM473.VKCM.878 TMC-Florida section is 255um from block surface	
	Male, Age 55, BMI 25.4 HBM824.BLXF.883 TMC-Vanderbilt 13-16	

<https://portal.hubmapconsortium.org/ccf-eui>

2020 Flashback

	HuBMAP	RBK	KPMP	SPARC	LungMAP	HTAN	HCA	GUDMAP	Gut Cell Atlas	BICCN	Allen Brain	TCGA	Wellcome	MRC	H2020	GTEx	Total
Kidney	1	1	1	0	0	0	1	1	0	0	0	1	1	1	0	1	9
Liver	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	3
Spleen	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	4
Heart	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	4
Lung	1	0	0	1	1	1	1	0	0	0	0	1	1	1	1	1	10
L intestine/Colon	1	0	0	1	0	1	1	0	1	0	0	1	0	0	0	1	7
S intestine	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
Bladder	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	5
Ureters	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2
Thymus	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2
Lymph nodes	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2
mediastinal lymph node	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Eye	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	3
Brain	0	0	0	0	0	0	1	0	0	1	1	1	0	0	1	1	6
Brain stem	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Cerebellum	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	3
Spinal cord	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	2
Pancreas	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	1	5
Breast	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	1	5
Skin	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	3
Pediatric systems	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	2
Ovaries	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	2
Testes	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	2
Cervix	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
Uterus	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	5
Blood	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	2
Bone	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Placenta	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Decidua	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Embryo	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
esophagus	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	3
hematopoietic system	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	2
immune system bulk	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
Stomach	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	3
Thyroid	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	2
Prostate	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	3
Adrenal gland	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	3
Totals	11	1	1	7	1	6	21	4	1	2	2	20	7	5	4	21	114

Table compiled for, during, and after the NIH-HCA Joint Meeting in March 2020, <https://hubmapconsortium.org/nihhca2020>

VH Massive Open Online Course (VHMOOC) 2020 Flashback

Goals

- Communicate tissue data acquisition and analysis,
- Demonstrate single-cell analysis and CCF mapping techniques, and
- Introduce major features of the HuBMAP portal.

Learning modules come with

- Videos (incl. interviews, tool demos)
- Hands-on exercises
- Self-quizzes



HuBMAP Visible Human MOOC (VHMOOC)
Started Aug 4, 2020
[GO TO CANVAS COURSE](#)
You are enrolled.



Course Introduction

This 10h course introduces the HuBMAP project which aims to create an open, global reference atlas of the human body at the cellular level. Among others, the course describes the compilation and coverage of HuBMAP data, demonstrates new single-cell analysis and mapping techniques, and introduces major features of the HuBMAP portal.

Delivered entirely online, all coursework can be completed asynchronously to fit busy schedules. If you have questions or experience issues during registration, please email cnsctr@indiana.edu.

Learning Outcomes

- Theoretical and practical understanding of different single-cell tissue analysis techniques.
- Expertise in single-cell data harmonization used to federate data from different individuals analyzed using different technologies in diverse labs.
- Hands-on skills in the design and usage of semantic ontologies that describe human anatomy, cell types, and biomarkers (e.g., marker genes or proteins).
- Knowledge on the design and usage of a semantically annotated three-dimensional reference system for the healthy human body.
- An understanding of how the HuBMAP reference atlas might be used to understand human health but also to diagnose and treat disease.

Module Topics Include

- HuBMAP Overview: Project Goals, Setup, and Ambitions
- Tissue Data Acquisition and Analysis
- Biomolecular Data Harmonization
- Ontology, 3D Reference Objects, and User Interfaces
- HuBMAP Portal Design and Usage

Meet the Instructors




Katy Börner, Victor H. Yingve Distinguished Professor of Engineering and Information Science. Founding Director of the Cyberinfrastructure for Network Science Center at Indiana University.





Ellen M. Quardokus, staff in the Chemistry Department and research scientist, Cyberinfrastructure for Network Science Center, SICE with expertise in molecular biology, microscopy, anatomy, and interdisciplinary communication.




Andreas Bueckle, PhD Candidate in Information Science, performing research on information visualization, specifically virtual and augmented reality.

 **Length:** 10 hours

 **Department:** Cyberinfrastructure Network Science

 **Credit:** None

 **Audience:** Biomedical students and professionals interested in single-cell tissue analysis and visualization

<https://expand.iu.edu/browse/sice/cns/courses/hubmap-visible-human-mooc>

Data Visualization Best Practices



Overview

Mapping Science: An Exhibit

Data Visualization Literacy Framework

Empower Yourself!

Mapping Science Exhibit

<http://scimaps.org>





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



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



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



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

<http://scimaps.org>

Places & Spaces: Mapping Science Exhibit

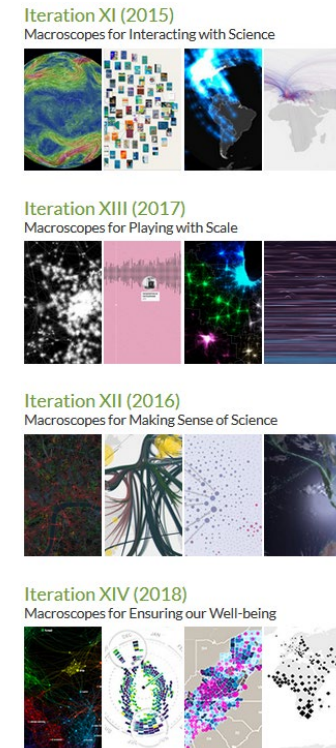
1st Decade (2005-2014)

Maps



2nd Decade (2015-2024)

Macroscopes

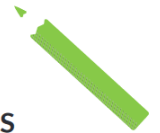


100

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

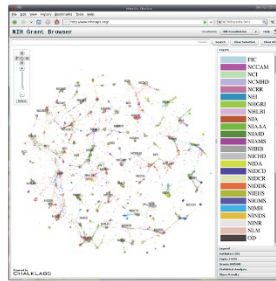


Computed Using Data from Elsevier's Scopus

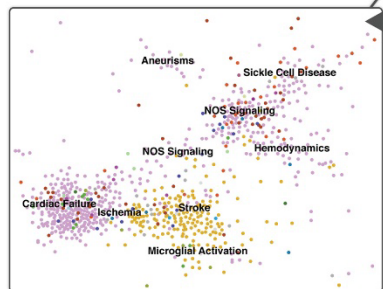
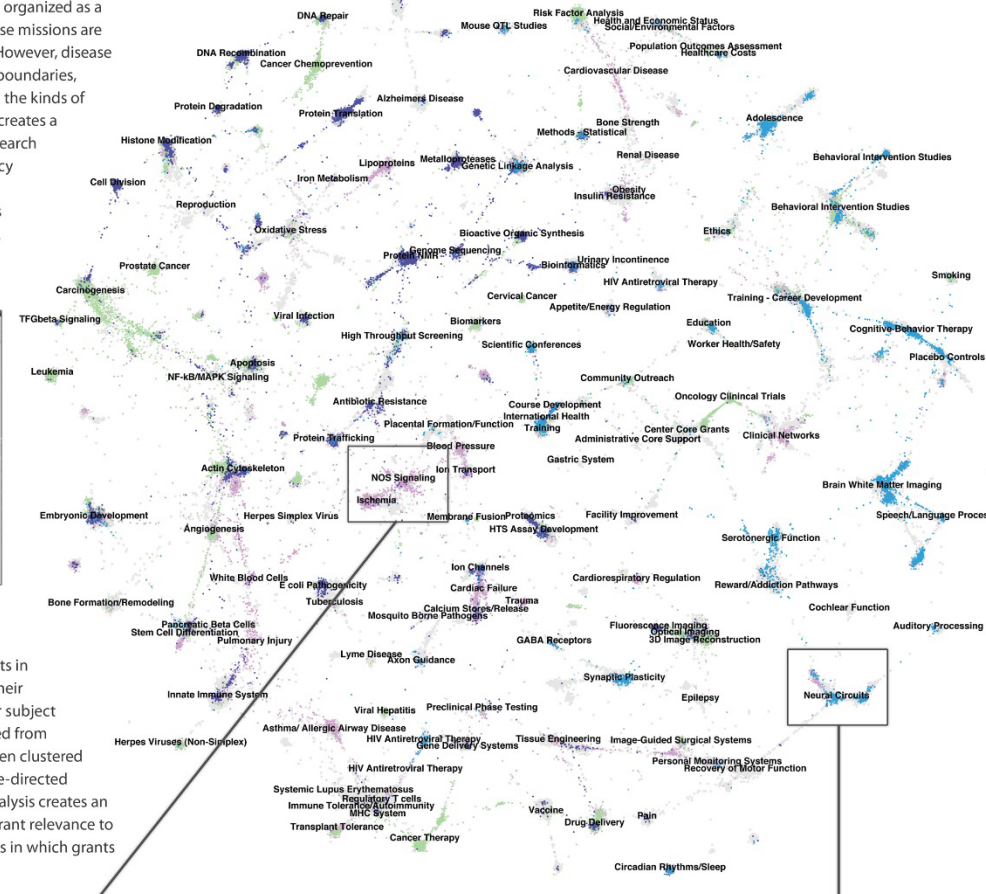
A Topic Map of NIH Grants 2007

Bruce W. Herr II (Chalklabs & IU), Gully Burns (ISI), David Newman (UCI), Edmund Talley (NIH)

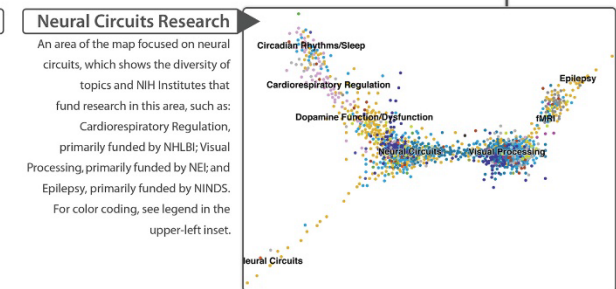
The National Institutes of Health (NIH) is organized as a multitude of Institutes and Centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies flout scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each Institute. This creates a daunting landscape for decisions on research directions, funding allocations, and policy formulations. Shown here is devised an interactive topic map for navigating this landscape, online at www.nihmaps.org. Institute abbreviations can be found at www.nih.gov/icd.



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.



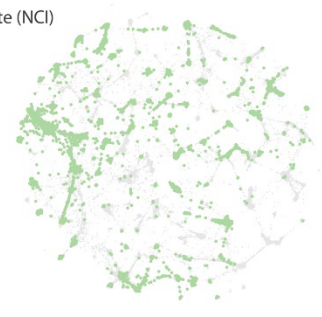
Cardiac Diseases Research
An area of the map focused on cardiovascular function and dysfunction. Cardiac Failure (primarily funded by NHLBI) is typically clustered next to Stroke (NINDS), since these are the two major medical emergencies associated with ischemia, which results from a restricted blood supply. Also localized in this area are grants focused on Nitric Oxide (NOS) Signaling, a major biochemical pathway for vasodilation, and grants on Hemodynamics, Sickle Cell Disease, and Aneurysms.



Neural Circuits Research
An area of the map focused on neural circuits, which shows the diversity of topics and NIH Institutes that fund research in this area, such as Cardiorespiratory Regulation, primarily funded by NHLBI; Visual Processing, primarily funded by NEI; and Epilepsy, primarily funded by NINDS. For color coding, see legend in the upper-left inset.

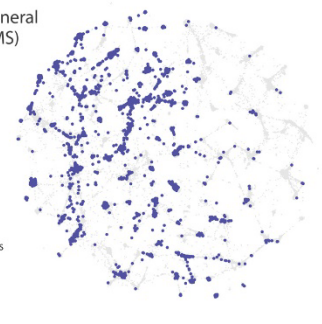
National Cancer Institute (NCI)

- TOP 10 TOPICS
- 1 Oncology Clinical Trials
 - 2 Cancer Treatment
 - 3 Cancer Therapy
 - 4 Carcinogenesis
 - 5 Risk Factor Analysis
 - 6 Cancer Chemotherapy
 - 7 Metastasis
 - 8 Leukemia
 - 9 Prediction/Prognosis
 - 10 Cancer Chemoprevention



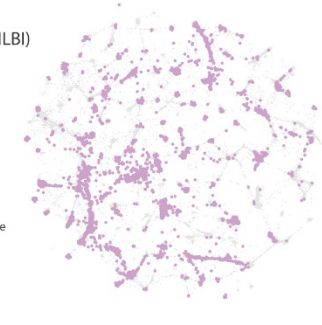
National Institute of General Medical Sciences (NIGMS)

- TOP 10 TOPICS
- 1 Bioactive Organic Synthesis
 - 2 X-ray Crystallography
 - 3 Protein NMR
 - 4 Computational Models
 - 5 Yeast Biology
 - 6 Metalloproteases
 - 7 Enzymatic Mechanisms
 - 8 Protein Complexes
 - 9 Invertebrate/Zebrafish Genetics
 - 10 Cell Division



National Heart, Lung, and Blood Institute (NHLBI)

- TOP 10 TOPICS
- 1 Cardiac Failure
 - 2 Pulmonary Injury
 - 3 Genetic Linkage Analysis
 - 4 Cardiovascular Disease
 - 5 Atherosclerosis
 - 6 Hemostasis
 - 7 Blood Pressure
 - 8 Asthma/ Allergic Airway Disease
 - 9 Gene Association
 - 10 Lipoproteins



National Institute of Mental Health (NIMH)

- TOP 10 TOPICS
- 1 Mood Disorders
 - 2 Schizophrenia
 - 3 Behavioral Intervention Studies
 - 4 Mental Health
 - 5 Depression
 - 6 Cognitive-Behavior Therapy
 - 7 AIDS Prevention
 - 8 Genetic Linkage Analysis
 - 9 Adolescence
 - 10 Childhood

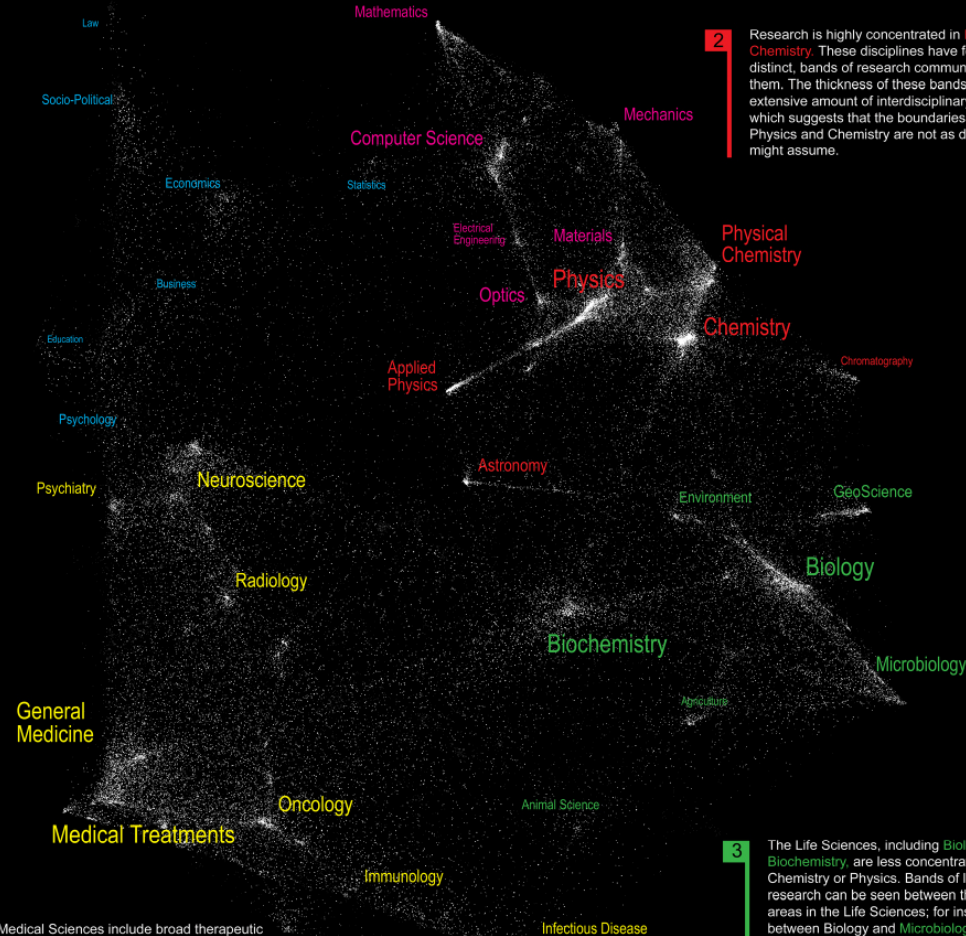


The Structure of Science

5 The Social Sciences are the smallest and most diffuse of all the sciences. **Psychology** serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. **Statistics** serves as the link with Computer Science and Mathematics.

1 **Mathematics** is our starting point, the purest of all sciences. It lies at the outer edge of the map. **Computer Science**, **Electrical Engineering**, and **Optics** are applied sciences that draw upon knowledge in Mathematics and Physics. These three disciplines provide a good example of a linear progression from one pure science (Mathematics) to another (Physics) through multiple disciplines. Although applied, these disciplines are highly concentrated with distinct bands of research communities that link them. Bands indicate interdisciplinary research.

2 Research is highly concentrated in **Physics** and **Chemistry**. These disciplines have few, but very distinct, bands of research communities that link them. The thickness of these bands indicates an extensive amount of interdisciplinary research, which suggests that the boundaries between Physics and Chemistry are not as distinct as one might assume.



4 The Medical Sciences include broad therapeutic studies and targeted areas of **Treatment** (e.g. central nervous system, cardiology, gastroenterology, etc.) Unlike Physics and Chemistry, the medical disciplines are more spread out, suggesting a more multi-disciplinary approach to research. The transition into Life Sciences (via Animal Science and Biochemistry) is gradual.

3 The Life Sciences, including **Biology** and **Biochemistry**, are less concentrated than Chemistry or Physics. Bands of linking 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 interesting in that it is a large discipline that has visible links to disciplines in many areas of the map, including Biology, Chemistry, Neuroscience, and General Medicine. It is perhaps the most interdisciplinary of the sciences.

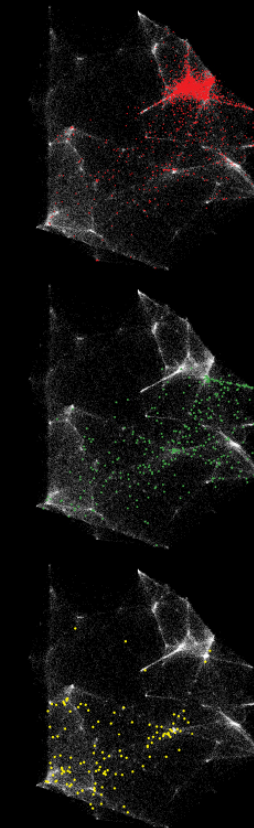
We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

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 17,000 scientific papers. That early map was intriguing, but it didn't cover enough of science to accurately define its structure.

Things are different today. We have enormous computing power and advanced visualization software that make mapping of the structure of science possible. This galaxy-like map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrd) from the citation patterns in 800,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 96,000 research communities active in science in 2002. A research community is a group of papers (9 on average) 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. This is the terrain in which organizations and institutions locate their scientific capabilities. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to explore, exploit, abandon, or ignore.

We also envision the map as an educational tool. For children, the theoretical relationship between areas of science can be replaced with a concrete map showing how math, physics, chemistry, biology and social studies interact. For advanced students, areas of interest 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 centered in **Biochemistry**. In addition, there is a heavy focus in the tools section of chemistry, such as **Chromatography**. The balance of the proteomics communities are widely dispersed among the Life and Medical Sciences.

Pharmacogenomics

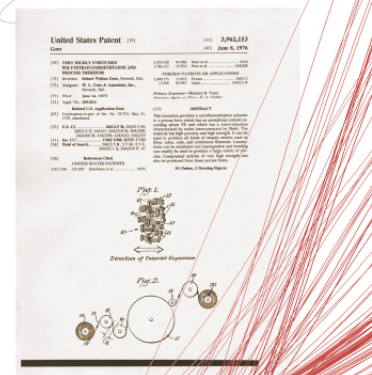
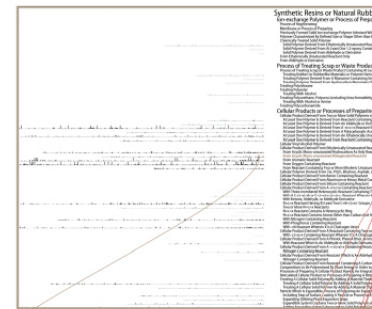
Pharmacogenomics is a relatively new field with most of its activity in **Medicine**. It also has many communities in **Biochemistry** and two communities in the Social Sciences.

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.

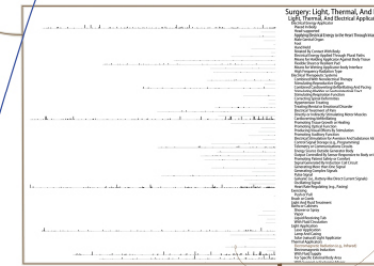
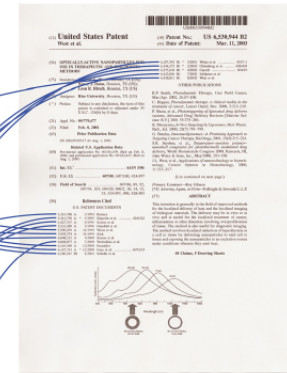
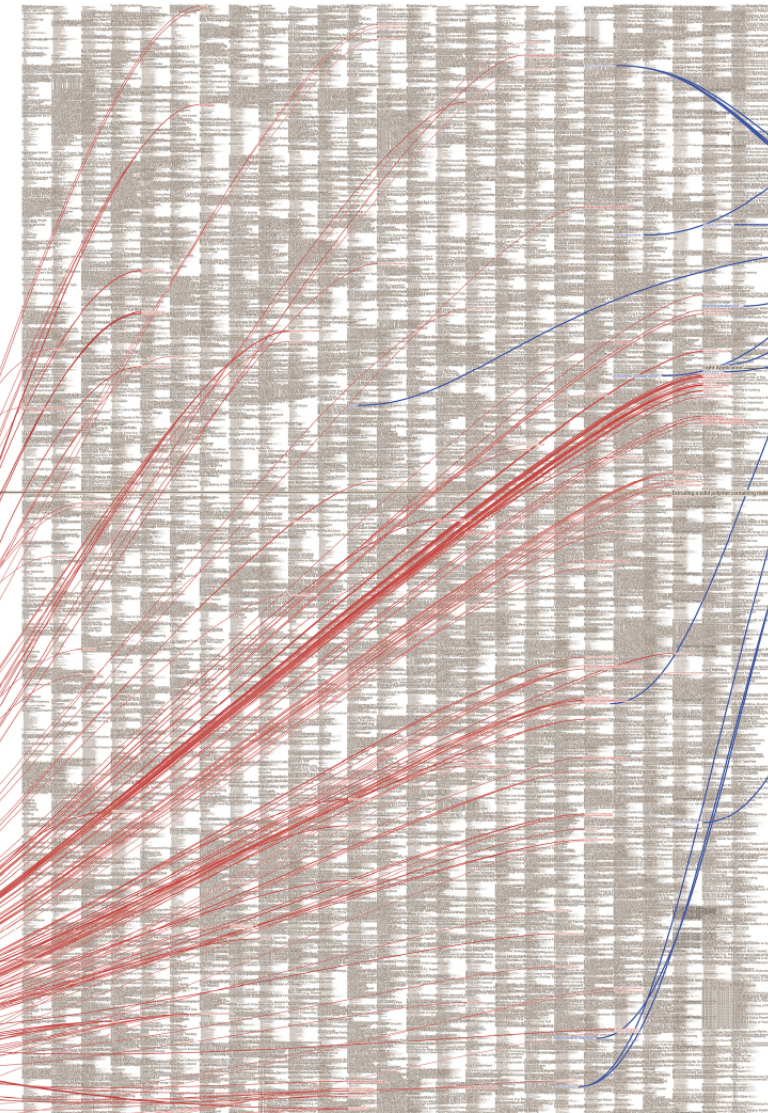
Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

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."



The US Patent Hierarchy

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 infrared 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.

Science related Wikipedian ACTIVITY

This visualization explores the activity of science, math, and technology (SMT) related articles in the English-language Wikipedia (<http://en.wikipedia.org>). The central image shows 659,388 articles (circles). Overlaid is a 37 x 37 grid of relevant half-inch sized images.

Blue, green, and yellow circles represent the 3,599 math, 6,474 science, and 3,164 technology related articles respectively. The larger the size of a circle the higher the likelihood it is that type of article. The four corners show activity patterns of SMT articles.

Article Edit Activity
Articles are size coded based on how frequently they have been edited from Feb. 6, 2001 to April 6, 2007. More consideration is given to current and major edits. Larger circles have been edited more frequently than smaller circles.

2007 Major Edits
Articles are size coded based on how many major edits they received from January 1st, 2007 to April 6th, 2007. Larger circles have received more edits than smaller circles. The highest number of major edits was 2,627.

For the central image, each article is size coded based on the likelihood that it is math, science, or technology related.

- 0%
- 50%
- 100%

All five images are color coded based on type. Transparency is used for legibility, and creates different colors when nodes overlap.

- Math
- Science
- Technology

Article Popularity
Articles are size coded based on the number of Wikipedia articles referencing it. Larger circles are receiving more links from other articles than smaller circles. The highest number of references to an article was 142,602.

Number of Bursts
Articles are size coded based on the number of bursts, i.e. sudden increases, of edit activity that occurred during the article's lifetime. Larger circles have had more bursts in activity than smaller circles. The most bursts an article had was 9.



Diseasome

The Human Disease Network

Explore online at <http://diseasome.eu>

Statistics

of Nodes: 516
 # of Edges: 1188
 Density: 0,0089
 Average Degree: 9,20
 Diameter: 15
 Average Shortest Path: 6,5

Top 5 Diseases

1. Deafness
2. Leukemia
3. Colon Cancer
4. Retinitis Pigmentosa
5. Diabetes Mellitus

Top 5 Genes

1. TP53
2. PAX6
3. FGFR2
4. RTN
5. MSH2

Description

The map presents a network of 516 diseases linked by 1188 known disorder-gene associations, indicating the common genetic origin of many diseases.

GENE NETWORKS CLUES

The map offers a rapid visual reference of the genetic links between disorders and a valuable global perspective for physicians, genetic counselors, and biomedical researchers alike. This view appears only when the network is zoomed, revealing to their associated genes, together the understanding of the roots of disease, and the functions of particular genes.

NETWORK VISUALIZATION TECHNIQUES APPLIED

The map was done using the force-directed layout algorithm ForceAtlas in Gephi. Node sizes correspond to the number of genes to which the disease belongs, and the size is proportional to its node degree, the overall number of links. Link's width is proportional to the number of genes that are implicated in both diseases and colored with the average color between source and target nodes. Isolated diseases are not shown and only the giant component has been kept. The Clusters Mini-map shows more readable disorder classes and shows largest visual clusters.

The Disorder Class Interactivity graph below shows the interaction level between disorder classes, representing the number of shared genes, up to 80.

References

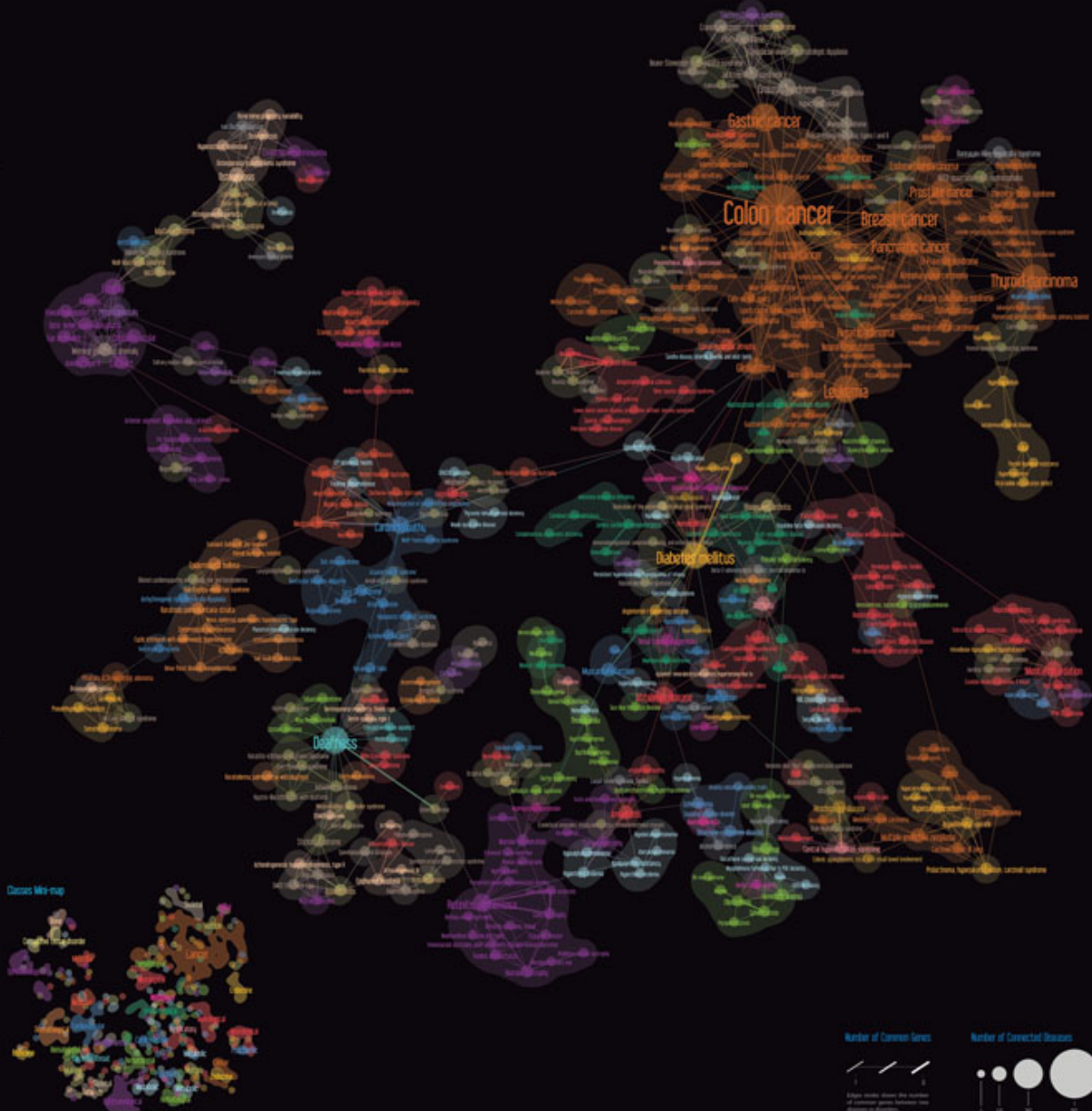
The Human Disease Network
 Bastin & Heymann 2009, *PLoS ONE*, 4(10): e7048
 PLoS ONE 4(10): e7048. doi:10.1371/journal.pone.0070488

Disorder Class Interactions

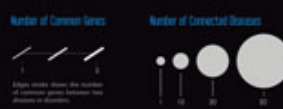


Disorder Class

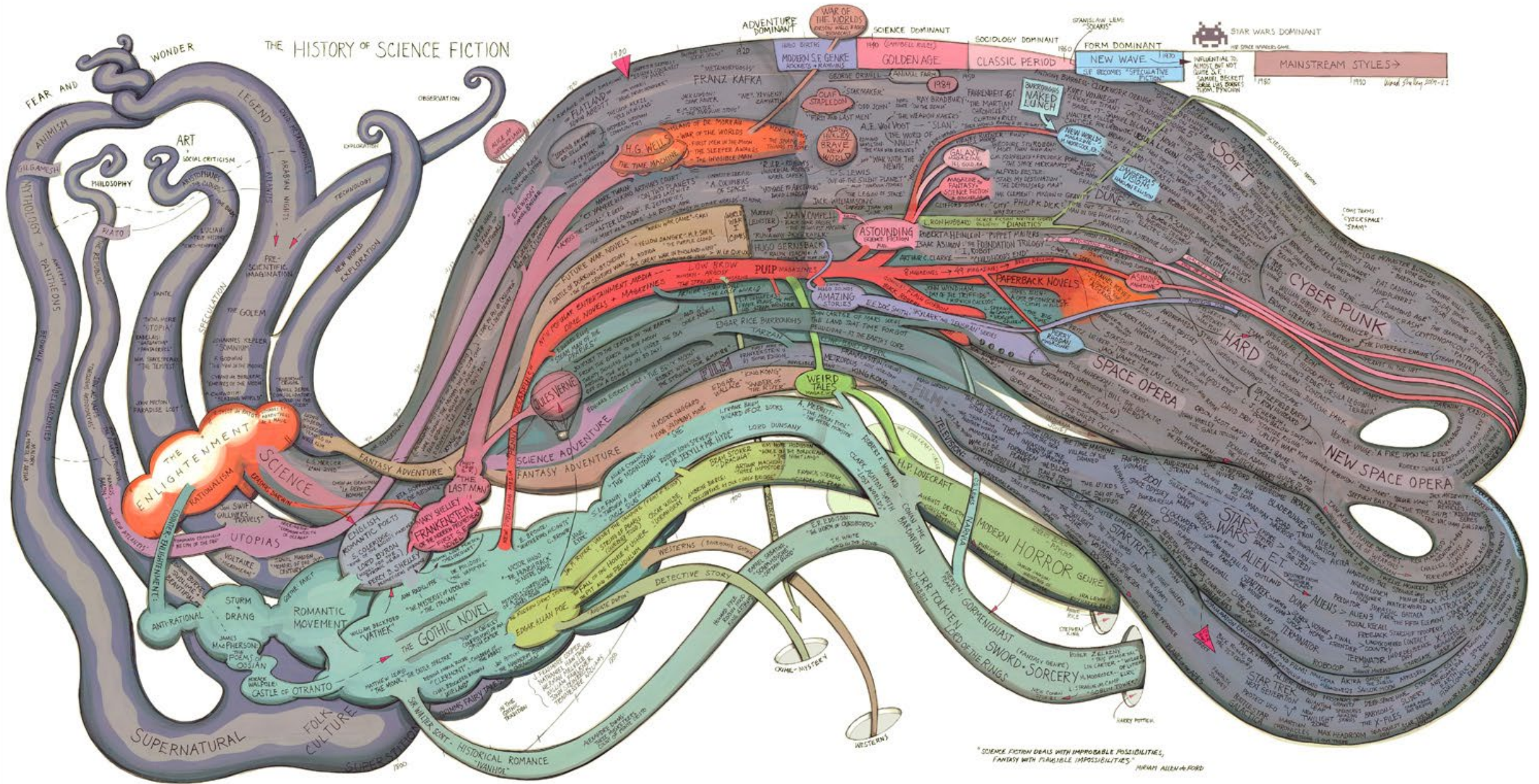
- Cancer
- Endocrine
- Ear, Nose, Throat
- Ophthalmological
- Neurological
- Hematological
- Cardiovascular
- Muscular
- Immunological
- Dermatological
- Nutritional
- Connective Tissue Disorder
- Renal
- Psychiatric
- Metabolic
- Bone
- Skeletal
- Developmental
- Gastrointestinal
- Respiratory
- Multiple
- Unclassified



Disorder Mini-map



VI.3 Diseasome: The Human Disease Network - Mathieu Bastian and Sébastien Heymann - 2009



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

Check out our **Zoom Maps** online!

VII.10
History of Science Fiction, by Ward Shulley

BROOKLYN, NY, 2011
Courtesy of Ward Shulley Studios

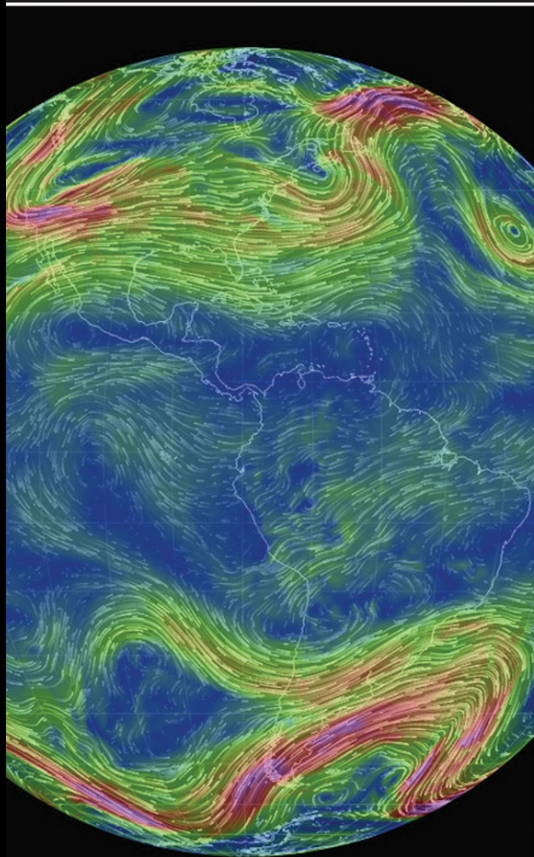
Ward Shulley is an artist identified with the Williamsburg scene in Brooklyn, New York. He is a writer, artist, and curator. This map plots the science fiction literary genre from its nascent beginnings in the late 18th century, through the Romantic period, and into the modern era. Emerging out of the data, here the narrative structure perceives and organizes the data. The map's structure is like a tree, tracing roots to pre-historical sources and whose body, the branches, are like a tree's canopy, showing the progression of the genre. The map is a complex network of colored lines and text, representing the evolution of science fiction literature. It includes various sub-genres like 'SCIENCE ADVENTURE', 'FANTASY ADVENTURE', 'CYBER PUNK', 'SPACE OPERA', and 'HARD'. It also lists numerous authors and works, such as Jules Verne, Mary Shelley, Frankenstein, and the first S.F. novel.

PLACES & SPACES
MAPPING & DESIGN

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



MACROSCOPES FOR INTERACTING WITH SCIENCE



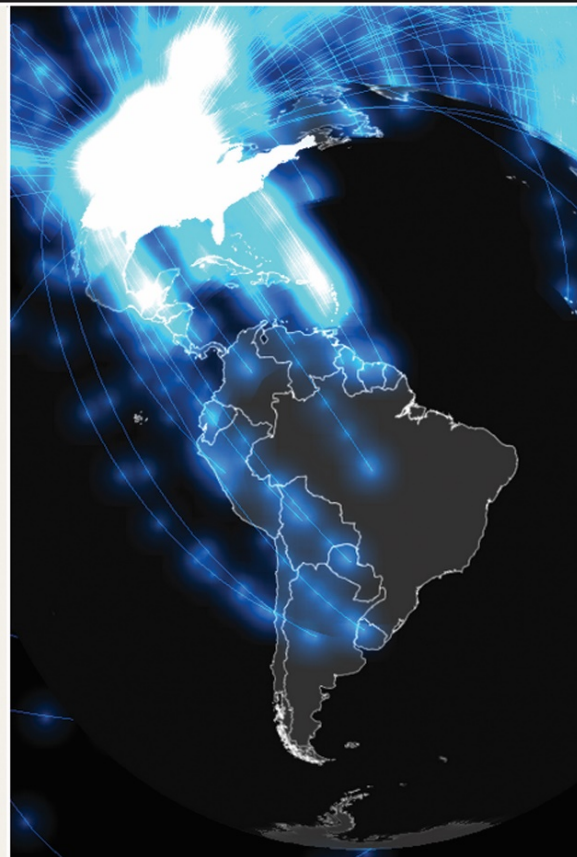
Earth

Weather on a worldwide scale



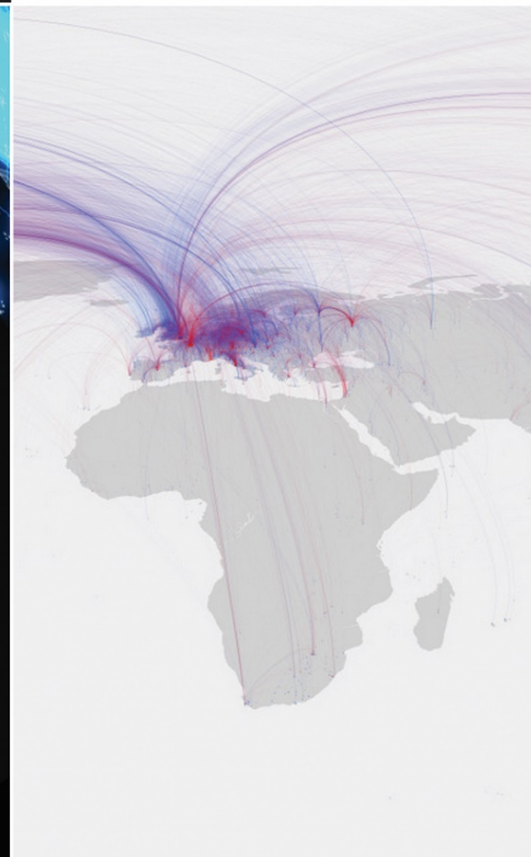
AcademyScope

Exploring the scientific landscape



Mapping Global Society

Local news from a global perspective

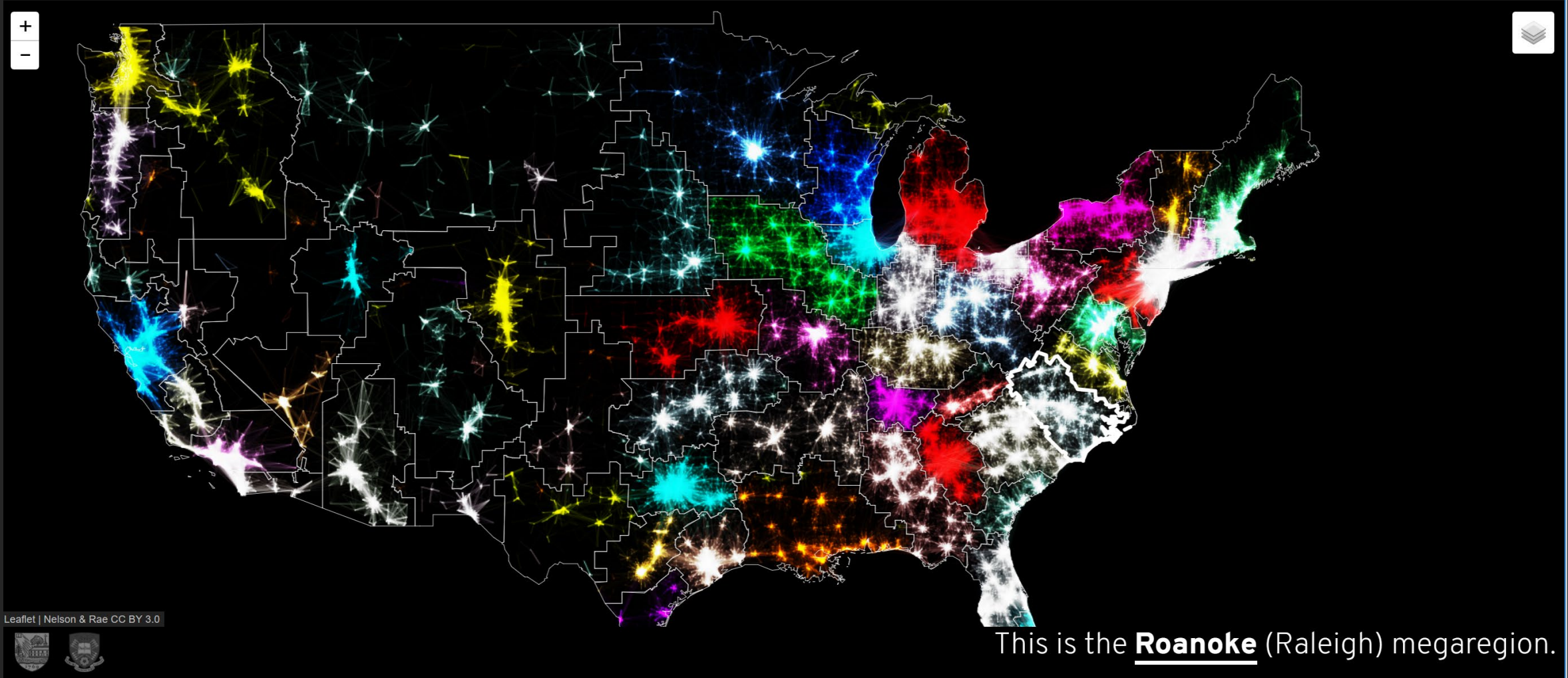


Charting Culture

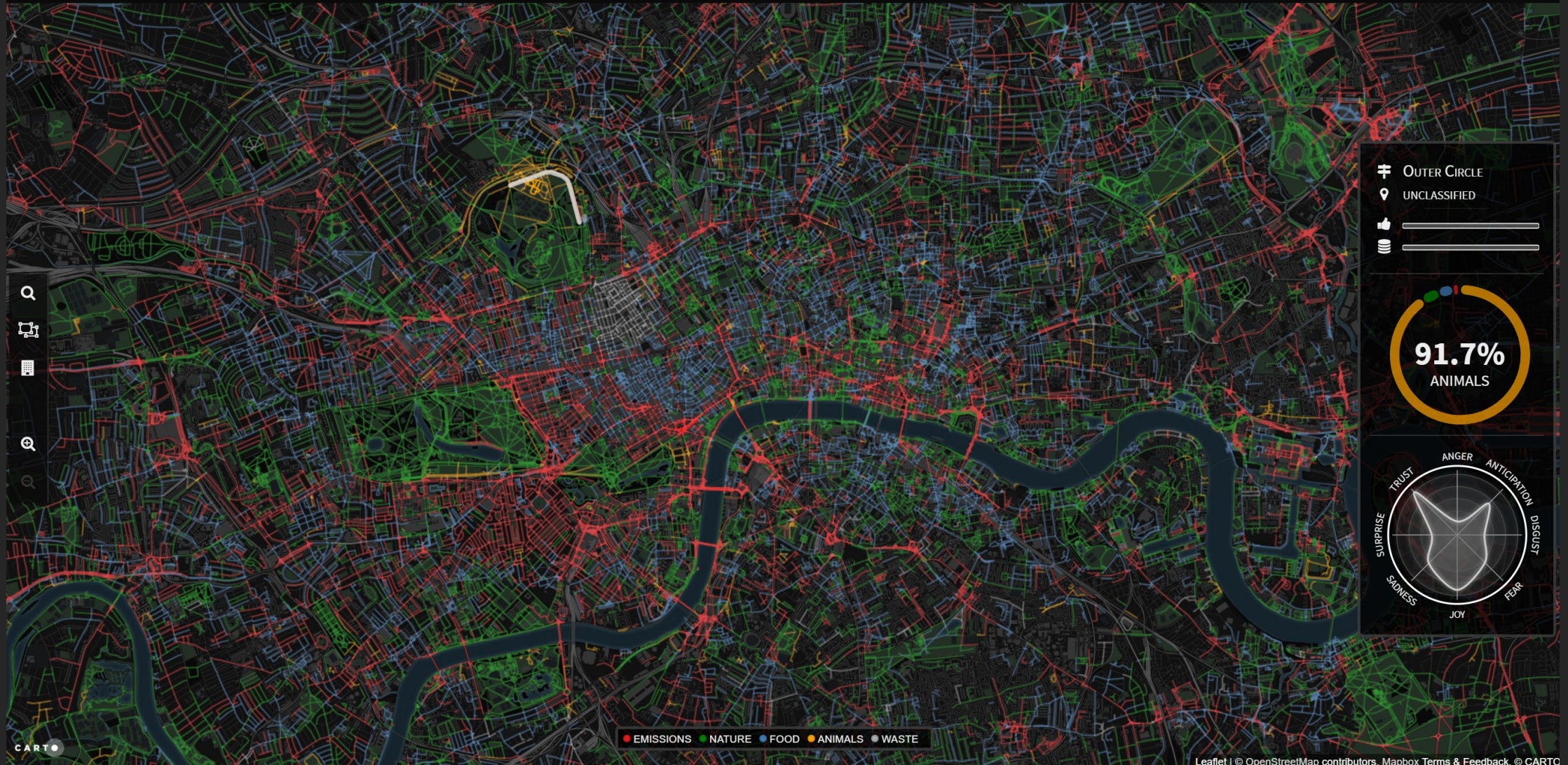
2,600 years of human history in 5 minutes

THE MEGAREGIONS OF THE US

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



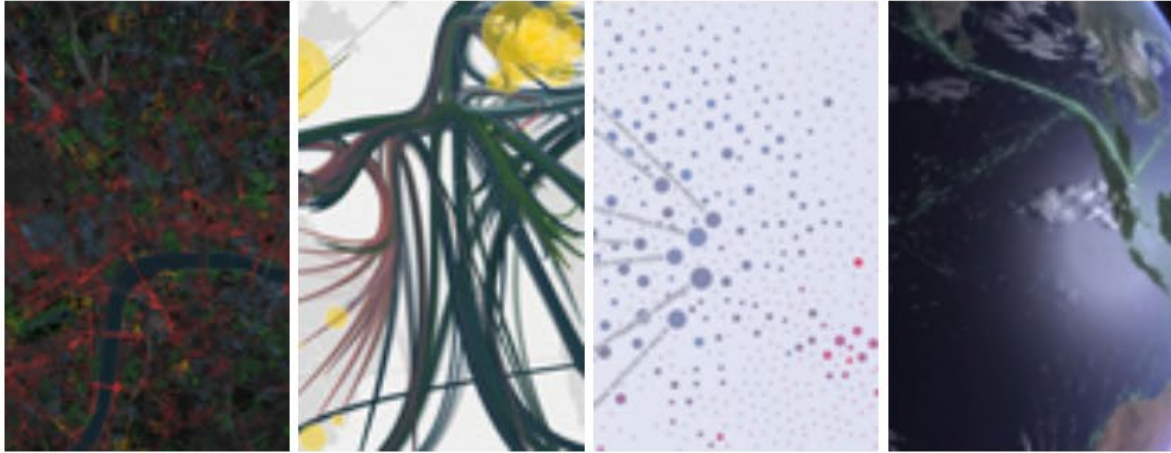
SMELLY MAPS



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

Iteration XII (2016)

Macrosopes for Making Sense of Science



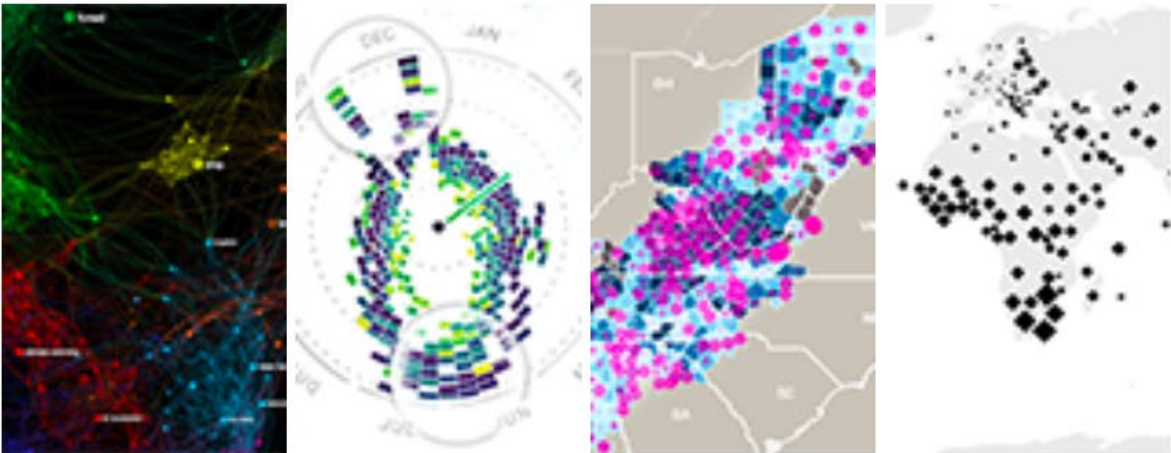
Iteration XIII (2017)

Macrosopes for Playing with Scale



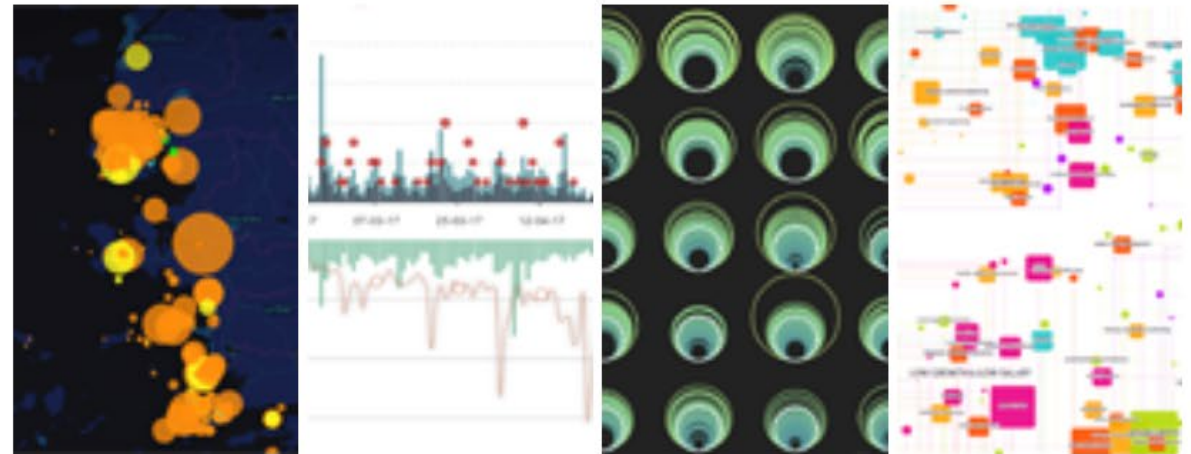
Iteration XIV (2018)

Macrosopes for Ensuring our Well-being



Iteration XV (2019)

Macrosopes for Tracking the Flow of Resources

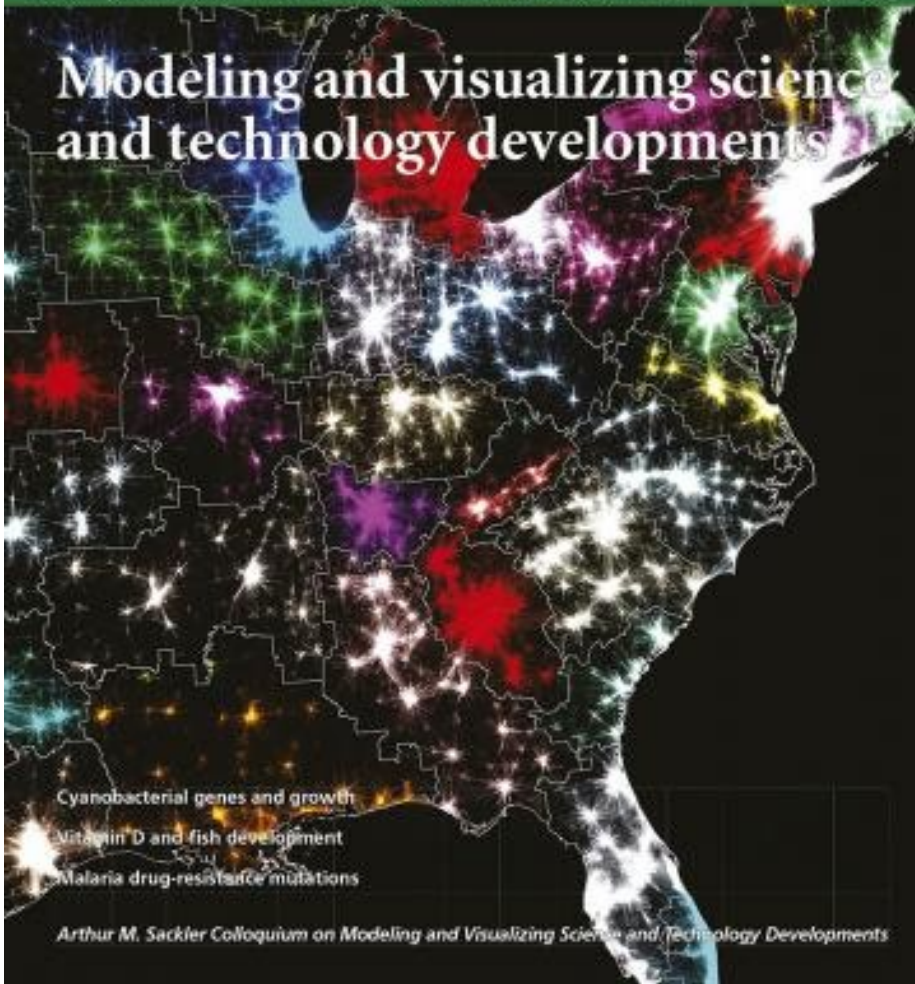


December 11, 2018 | vol. 115 | no. 50 | pp. 12537-12828

PNAS

Proceedings of the National Academy of Sciences of the United States of America www.pnas.org

Modeling and visualizing science and technology developments



Cyanobacterial genes and growth
Vitamin D and fish development
Malaria drug-resistance mutations

Arthur M. Sackler Colloquium on Modeling and Visualizing Science and Technology Developments

<https://www.pnas.org/modeling>

Atlas of Forecasts

Modeling and Mapping Desirable Futures

Katy Börner



<https://mitpress.mit.edu/books/atlas-forecasts>

Acknowledgments

Exhibit Curators



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

<http://scimaps.org>

Plus, we thank the more than 250 authors of the 100 maps and 16 interactive macroscopes.

Exhibit Advisory Board



Gary Berg-Cross
Cognitive psychologist (PhD, SUNY-Stony Brook). Potomac, MD, USA



Donna J. Cox, MFA, Ph.D.
Director of the **Advanced Visualization Laboratory** at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, IL, USA



Bonnie DeVarco
Media X Distinguished Visiting Scholar at Stanford University, Palo Alto, CA, USA



Peter A. Hook
Head of Digital and Scholarly Services and LawArXiv Administrator, Cornell Law Library, Ithaca, NY, USA



Francis Harvey
Professor of Visual Communication in Geography at the Leibnitz Institute for Regional Geography, Leipzig University, Germany



Lev Manovich
Professor, **The Graduate Center**, City University of New York; Director, **Software Studies Initiative** (big data, digital humanities, visualization)



André Skupin
Associate Professor of Geography at San Diego State University, California



Moritz Stefaner
Freelance designer on the crossroads of data visualization, information aesthetics, and user interface design in Germany



Olga Subirós
Curator of Big Bang Data and Founder of **Olga Subirós Studio** in Barcelona, Spain



Stephen Uzzo
Vice President of Science and Technology for the **New York Hall of Science**



Benjamin Wiederkehr
Founding Partner and Managing Director of **Interactive Things** in Zürich, Switzerland

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

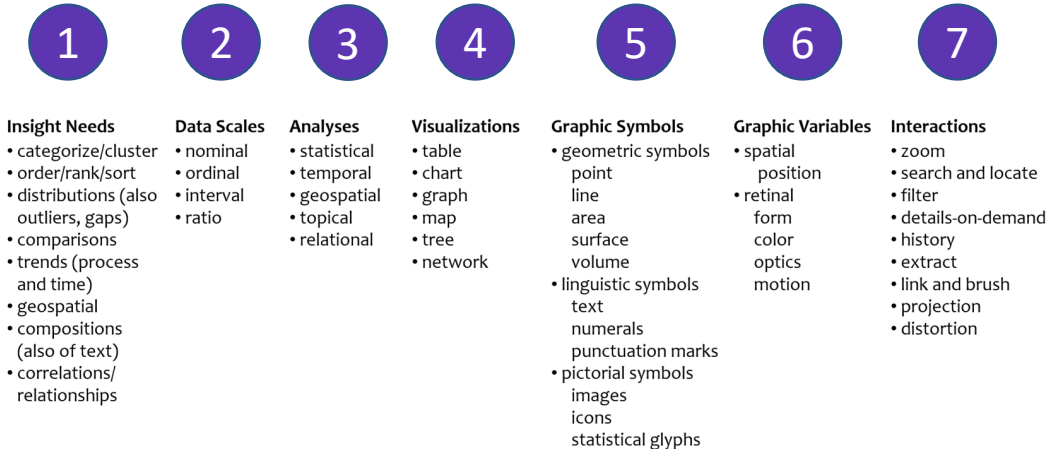
- The initial DVL-FW was developed via an extensive literature review.
- The resulting DVL-FW typology, process model, exercises, and assessments were then tested in the *Information Visualization* course taught for more than 17 years at Indiana University. More than 8,500 students enrolled in the IVMOOC 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:

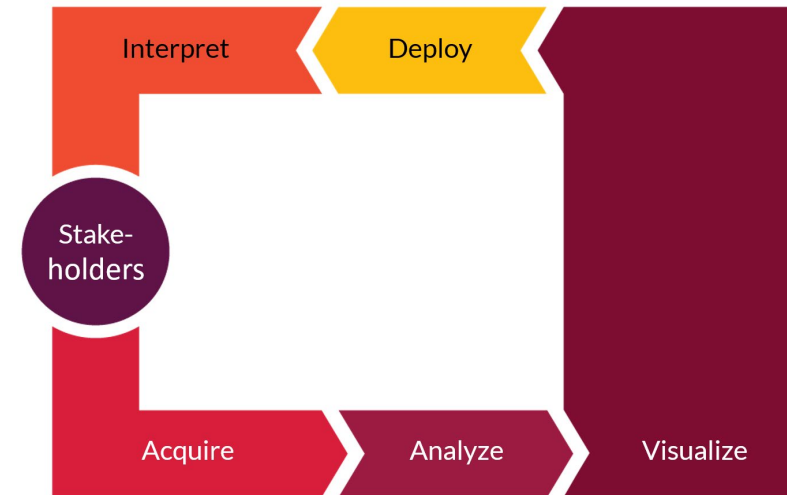
DVL Typology

Defines 7 types with 4-17 members each.



DVL Workflow Process

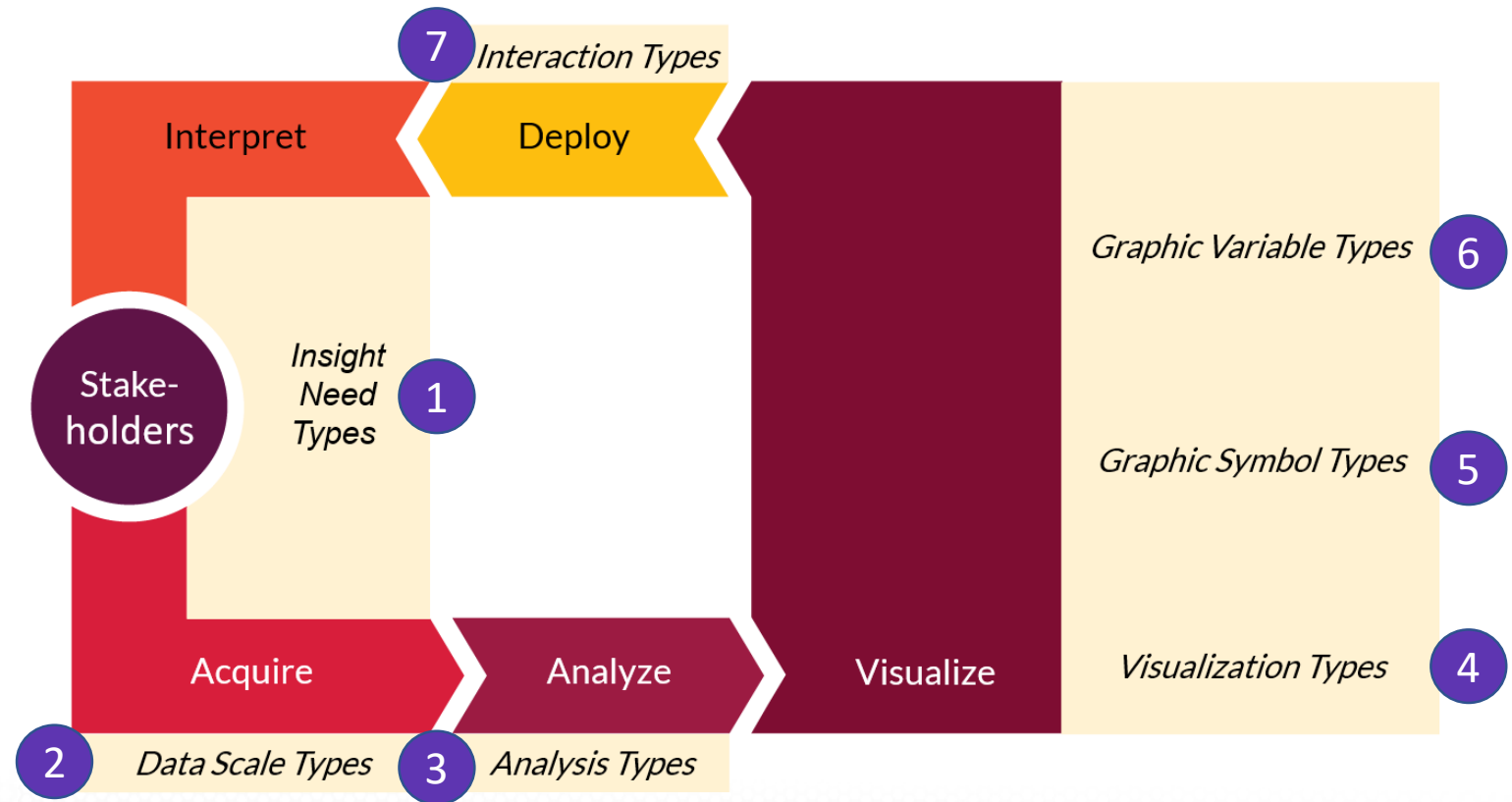
Defines 5 steps required to render data into insights.

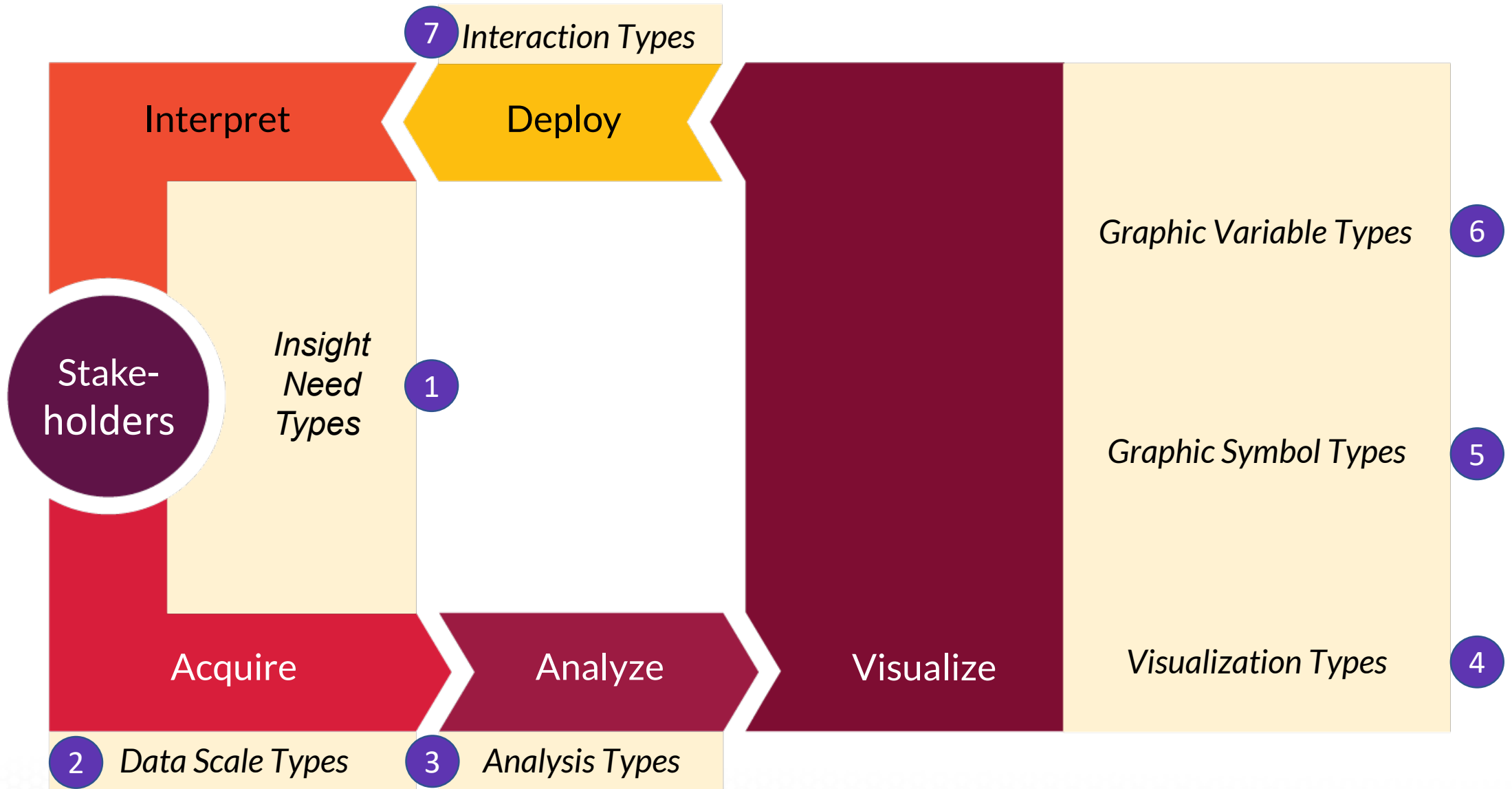


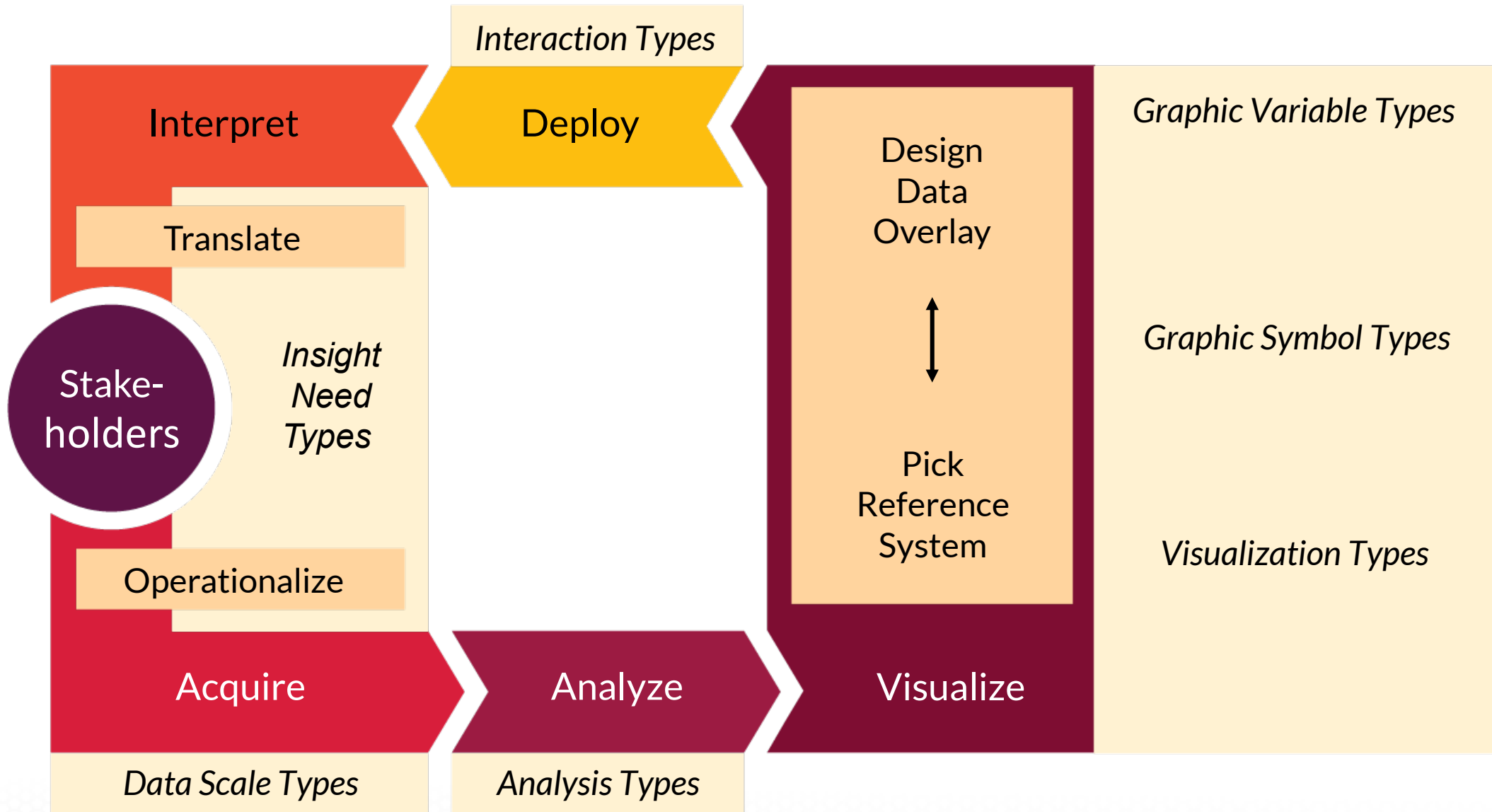
Data Visualization Literacy Framework (DVL-FW)

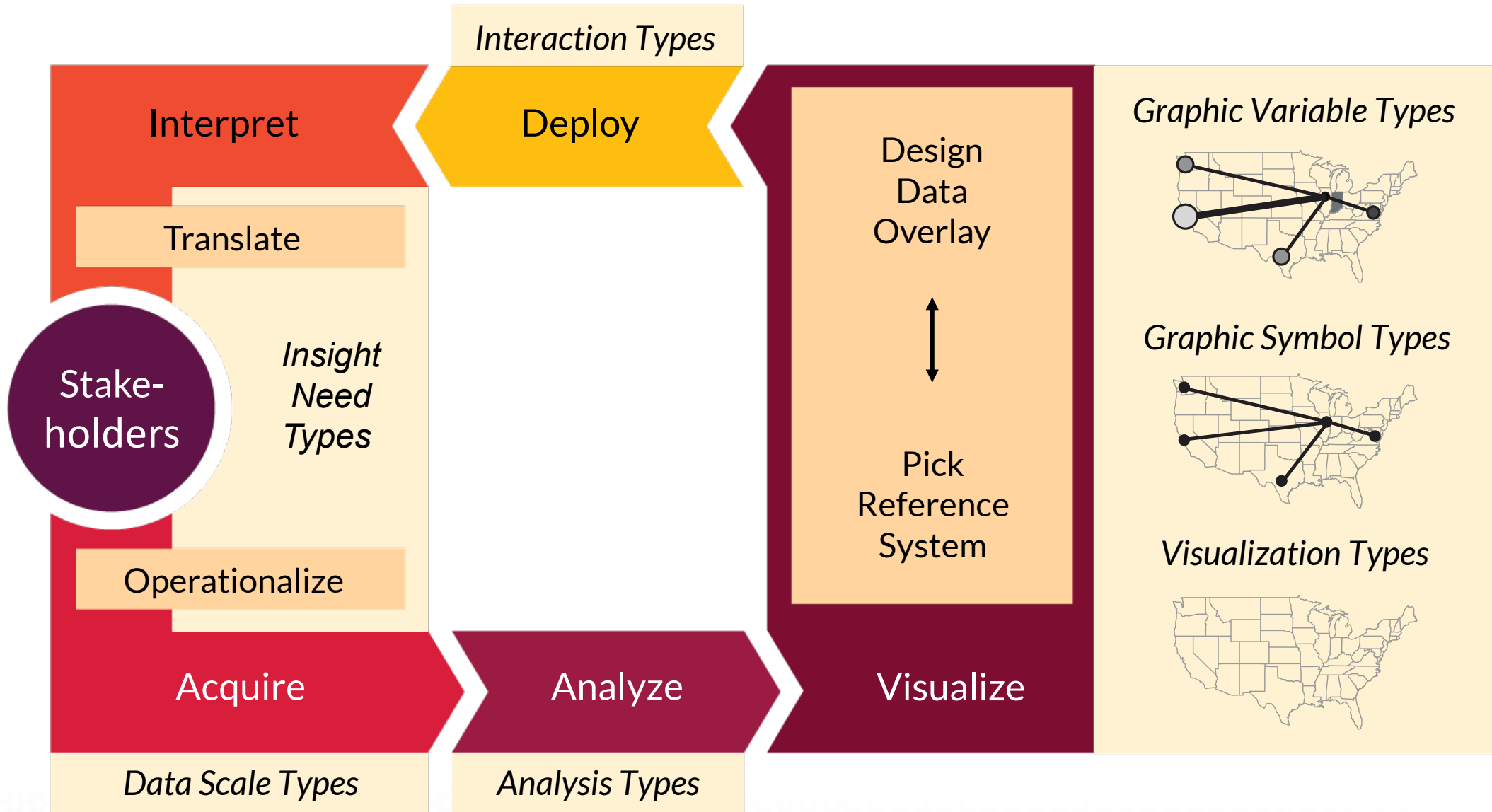
Consists of two parts *that are interlinked*:

**DVL Typology +
DVL Workflow Process**









Audience Poll

How much time (out of project total) do you spent on

- Data acquisition?
- Data cleaning?
- Data analysis?
- Data visualization?
- Data interpretation?

Data Visualization Literacy Framework (DVL-FW)

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

The screenshot shows the Make-A-Vis interface with three main sections: Data, Make Visualization, and a visualization preview.

Data Section:

- ISI Publications: (CSV) Preprocessed-wos**

Title	Authors	Journal	Year	#Cites
[Progress bar]				

Total Records: 562
- Journals: (from ISI Publications)**

Name	#Papers	#Cites	First Year	Last Year
BMC EVOL BIOL	1	7	2006	2006
FEBS J	2	0	2005	2005
NAT PHYS	3	18	2005	2006

Total Records: 562

Make Visualization Section:

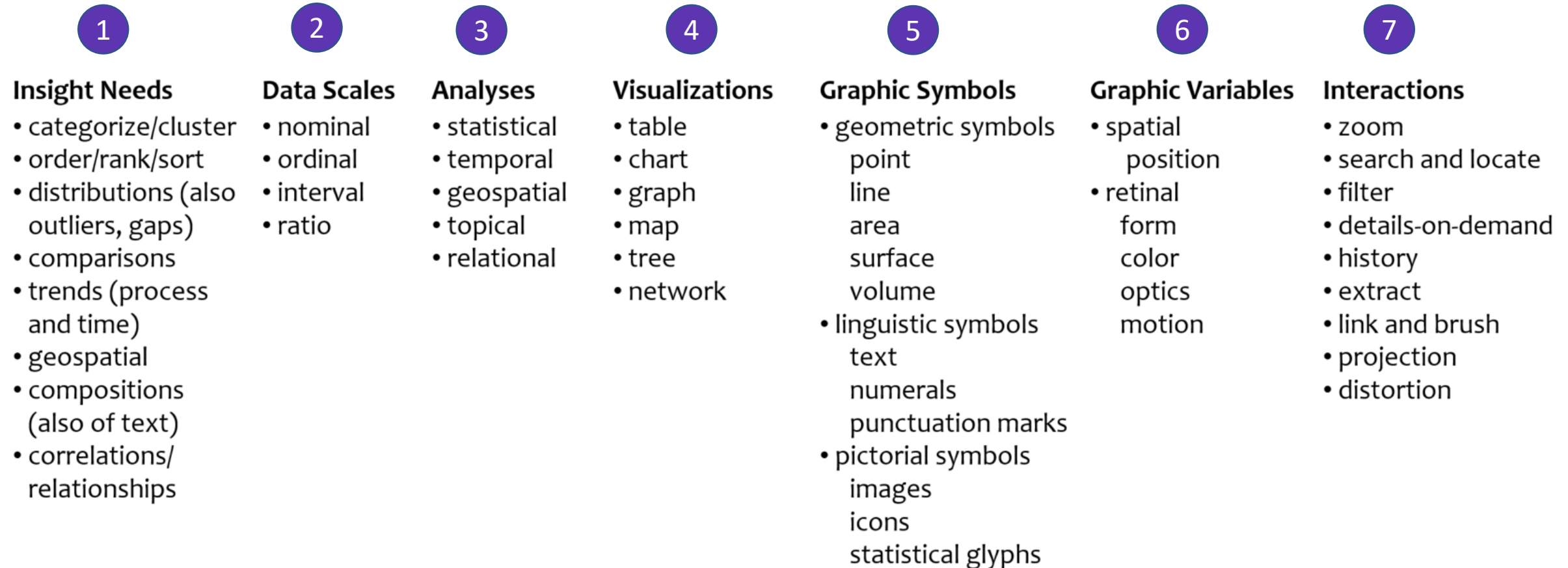
- Select Visualization Type:** Scatter Graph, Geomap, Scimap, Temporal Bar Graph (selected).
- Select Graphic Symbol Type(s):** (Dropdown menu)
- Select Graphic Variable Types:** (Dropdown menu)

Temporal Bar Graph Preview:

Temporal Bar Graph

Year	Category
1998	Machine
1999	Machine
2000	Education
2001	Building
2002	Making
2003	Computing
2004	Web
2005	Form
2006	Smart
2007	Capacity
2008	Algebraic Geometry
2009	Parts
2010	Law
2011	Stem
2012	Analysis
2013	Recovery
2014	Geometry
2015	Computer
2016	Application
2017	Robotics

Typology of the Data Visualization Literacy Framework



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

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

Data Scales

- nominal
- ordinal
- interval
- ratio

Analyses

- statistical
- temporal
- geospatial
- topical
- relational

Visualizations

- table
- chart
- graph
- map
- tree
- network

Graphic Symbols

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

Graphic Variables

- spatial
 - position
- retinal
 - form
 - color
 - optics
 - motion

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. 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

4

Insight Needs

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

Data Scales

- nominal
- ordinal
- interval
- ratio

Analyses

- statistical
- temporal
- geospatial
- topical
- relational

Visualizations

- table
- chart
- graph
- map
- tree
- network

Graphic Symbols

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

Graphic Variables

- spatial
 - position
- retinal
 - form
 - color
 - optics
 - motion

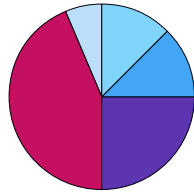
Interactions

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

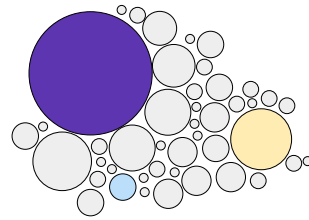
Börner, Katy. 2015. [Atlas of Knowledge: Anyone Can Map](#). Cambridge, MA: The MIT Press. 30-31.

Visualization Types

Chart

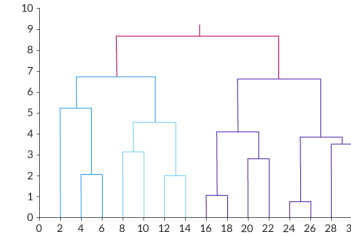


Pie Chart

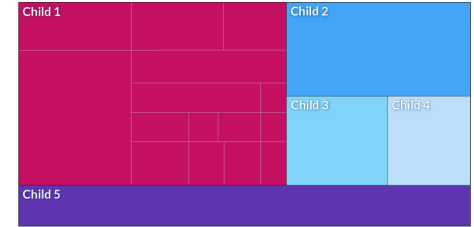


Bubble Chart

Tree

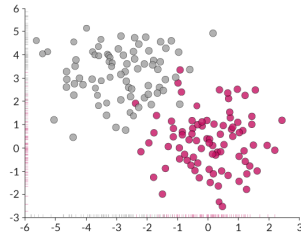


Dendrogram

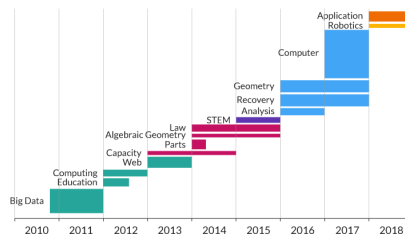


Tree Map

Graph

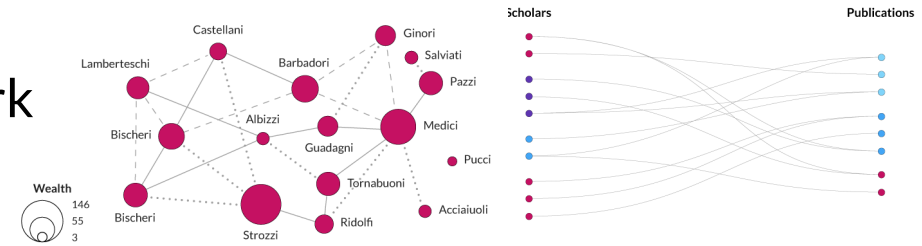


Scatter Graph



Temporal Bar Graph

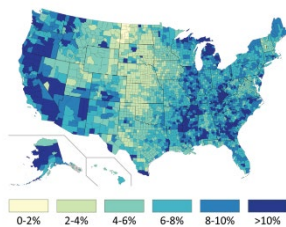
Network



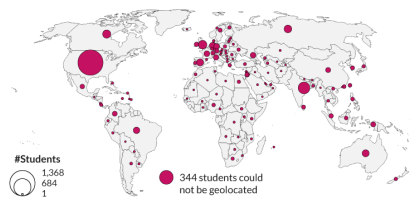
Force-Directed Network Layout

Bimodal Network Layout

Map



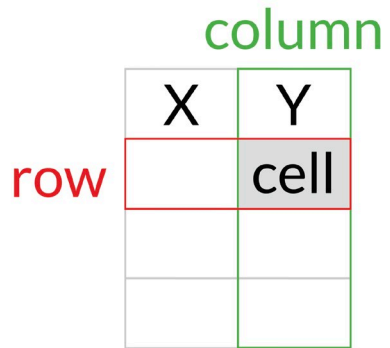
Choropleth Map



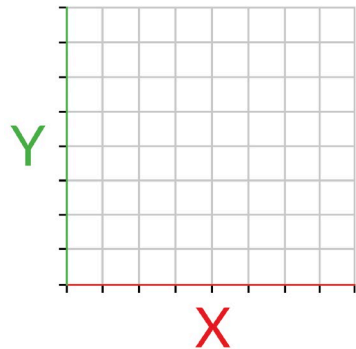
Proportional Symbol Map

Visualize: Reference Systems

Table
columns by rows



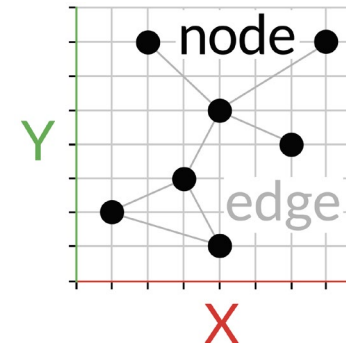
Graph
x-y coordinates



Map
latitude/
longitude



Network
local similarity

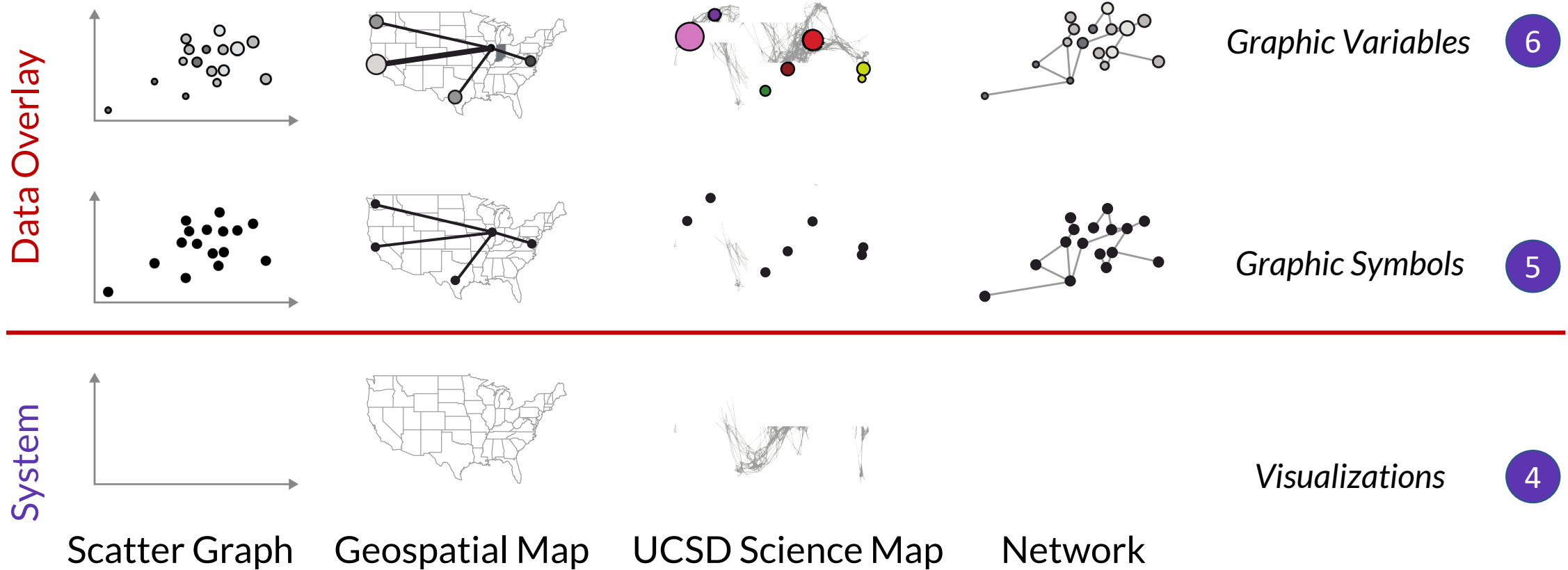


4

Visualization Types

- table
- chart
- graph
- map
- network layout

Visualize: Reference Systems, Graphic Symbols and Variables



Typology of the Data Visualization Literacy Framework

5

Insight Needs

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

Data Scales

- nominal
- ordinal
- interval
- ratio

Analyses

- statistical
- temporal
- geospatial
- topical
- relational

Visualizations

- table
- chart
- graph
- map
- tree
- network

Graphic Symbols

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

Graphic Variables

- spatial
 - position
- retinal
 - form
 - color
 - optics
 - motion

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. 32-33.

Typology of the Data Visualization Literacy Framework

6

Insight Needs

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

Data Scales

- nominal
- ordinal
- interval
- ratio

Analyses

- statistical
- temporal
- geospatial
- topical
- relational

Visualizations

- table
- chart
- graph
- map
- tree
- network

Graphic Symbols

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

Graphic Variables

- spatial
 - position
- retinal
 - form
 - color
 - optics
 - motion

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. 34-35.

Graphic Variable Types

Position: x, y; possibly z

Form:

- Size
- Shape
- Rotation (Orientation)

Color:

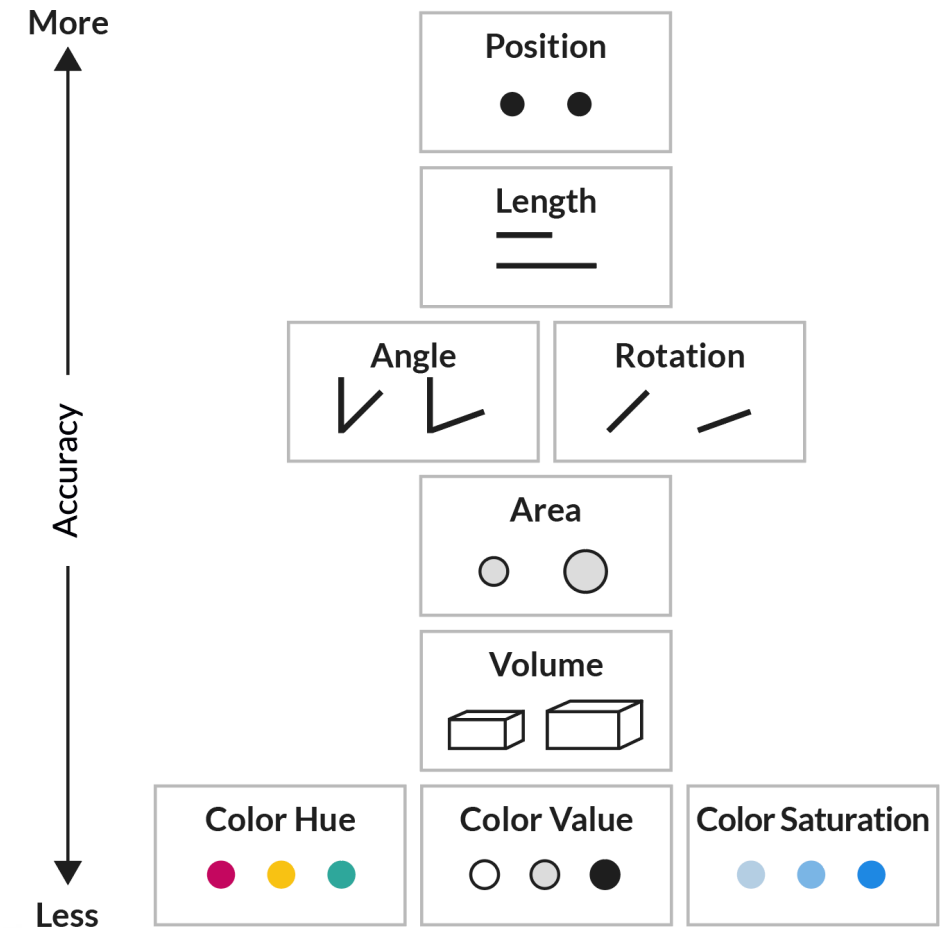
- Value (Lightness)
- Hue (Tint)
- Saturation (Intensity)



Optics: Blur, Transparency, Shading, Stereoscopic Depth

Texture: Spacing, Granularity, Pattern, Orientation, Gradient

Motion: Speed, Velocity, Rhythm



Graphic Symbol Types

			Geometric Symbols		Linguistic Symbols	Pictorial Symbols
			Point	Line		
Spatial	Position	X Y				
		Retinal	Form	Size		
Shape					Text Text Text	
Color	Value				Text Text Text	
	Hue				Text Text Text	
	Saturation				Text Text Text	
Texture	Granularity					
	Pattern					
Motion Optics	Blur				Text Text Text	
	Speed					

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

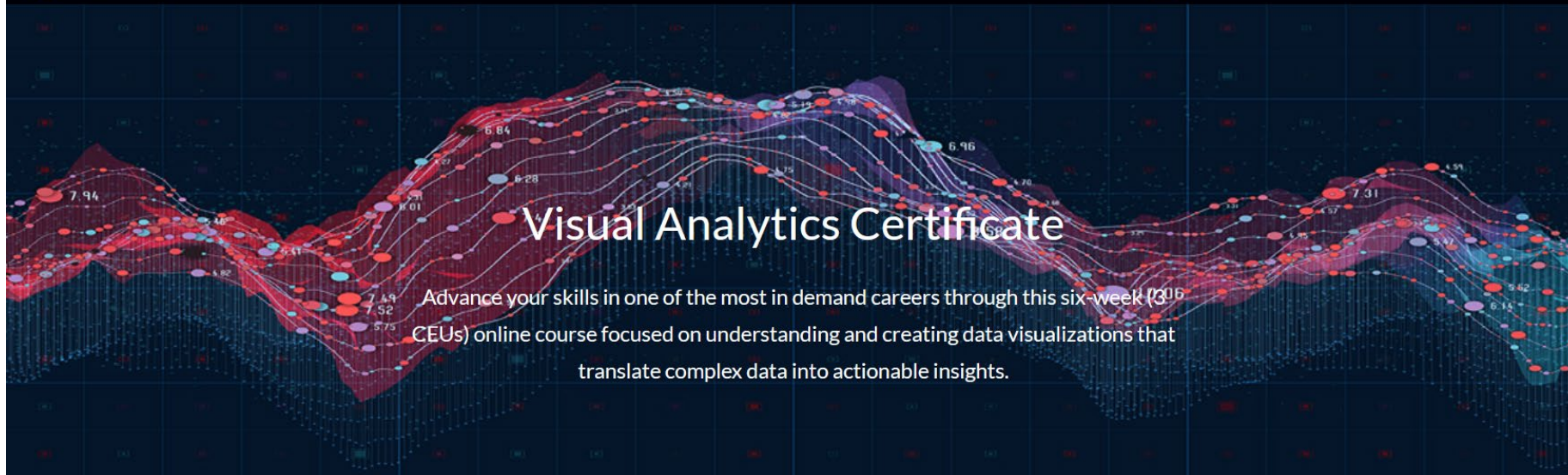
Graphic Variable Types Versus Graphic Symbol Types

			Geometric Symbols					Linguistic Symbols Text, Numerals, Punctuation Marks					Pictorial Symbols Images, Icons, Statistical Glyphs					
			Point	Line	Area	Surface	Volume											
Spatial	x	quantitative																
	y	quantitative																
	z	quantitative																
Retinal	Form	Size	quantitative	NA (Not Applicable)														
		Shape	qualitative	NA														
		Rotation	quantitative	NA														
		Curvature	quantitative	NA														
	Angle	quantitative	NA															
	Closure	quantitative	NA															
	Value	quantitative																
	Color	Hue	qualitative															
Saturation	quantitative																	
Retinal	Texture	Spacing	quantitative															
		Granularity	quantitative															
		Pattern	qualitative															
		Orientation	quantitative	NA														
		Gradient	quantitative															
	Optics	Blur	quantitative															
		Transparency	quantitative															
		Shading	quantitative															
	Motion	Stereoscopic Depth	quantitative	Point in foreground .. background	Line in foreground .. background	Area in foreground .. background	Surface in foreground .. background	Volume in foreground .. background	Text in foreground .. background					Icons in foreground .. background				
		Speed	quantitative															
Velocity		quantitative																
Rhythm	quantitative	Blinking point slow .. fast	Blinking line slow .. fast	Blinking area slow .. fast	Blinking surface slow .. fast	Blinking volume slow .. fast	Blinking text slow .. fast					Blinking icons slow .. fast						

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

Empower Yourself and Others! Data Visualization Literacy

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



Visual Analytics Certificate

Advance your skills in one of the most in demand careers through this six-week (30 CEUs) online course focused on understanding and creating data visualizations that translate complex data into actionable insights.

FLYER

REGISTER FOR MAY 17-JUNE 27, 2021

FAQS



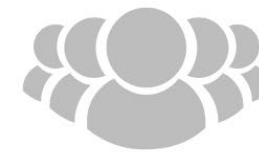
Learn from Experts

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>

US Employers which have sent students include
The Boeing Company, Eli Lilly, DOE, CDC, NSWC Crane.

References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains**. In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255.

<http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf>

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1).

http://www.pnas.org/content/vol101/suppl_1

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

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

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

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

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



Q&A

