

Data Visualization Literacy

Katy Börner @katycns

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



Computer Aided Drug Design, Gordon Research Conference

Mount Snow, West Dover, VT | July 17, 2019



Overview

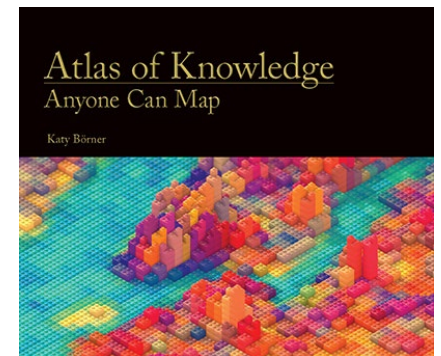
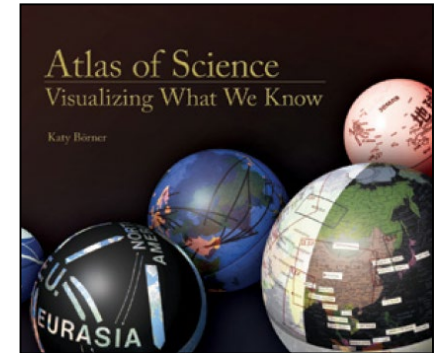
This talk will introduce a theoretical data visualization framework (DVL) meant to empower anyone to systematically render data into insights using temporal, geospatial, topical, and network analyses and visualizations.

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

The DVL was applied to map science and technology, see interactive data visualizations from the *Places & Spaces: Mapping Science* exhibit (<http://scimaps.org>) and recent *PNAS* special issue on *Modelling and Visualizing Science and Technology Developments* (<https://www.pnas.org/modeling>)

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

Börner, Katy. 2010. [Atlas of Science: Visualizing What We Know](#). Cambridge, MA: The MIT Press.



Atlas of Forecasts

Places & Spaces: Mapping Science Exhibit

1st Decade (2005-2014)

Maps

<p>Iteration I (2005) The Power of Maps</p>	<p>Iteration II (2006) The Power of Reference Systems</p>
<p>Iteration III (2007) The Power of Forecasts</p>	<p>Iteration IV (2008) Science Maps for Economic Decision Makers</p>
<p>Iteration V (2009) Science Maps for Science Policy Makers</p>	<p>Iteration VI (2010) Science Maps for Scholars</p>
<p>Iteration VII (2011) Science Maps as Visual Interfaces to Digital Libraries</p>	<p>Iteration VIII (2012) Science Maps for Kids</p>
<p>Iteration IX (2013) Science Maps Showing Trends and Dynamics</p>	<p>Iteration X (2014) The Future of Science Mapping</p>

2nd Decade (2015-2024)

Macroscopes

<p>Iteration XI (2015) Macroscopes for Interacting with Science</p>
<p>Iteration XIII (2017) Macroscopes for Playing with Scale</p>
<p>Iteration XII (2016) Macroscopes for Making Sense of Science</p>
<p>Iteration XIV (2018) Macroscopes for Ensuring our Well-being</p>

3rd Decade (2015-2034)

<p>101st Annual Meeting of the Association of American Geographers, Denver, CO. April 5th - 9th, 2005 (First showing of Places & Spaces)</p>	<p>University of Miami, Miami, FL. September 4 - December 11, 2014.</p>
<p>Duke University, Durham, NC. January 12 - April 10, 2015</p>	<p>The David J. Sencer CDC Museum, Atlanta, GA. January 25 - June 17, 2016.</p>

<http://scimaps.org>

The Structure of Science

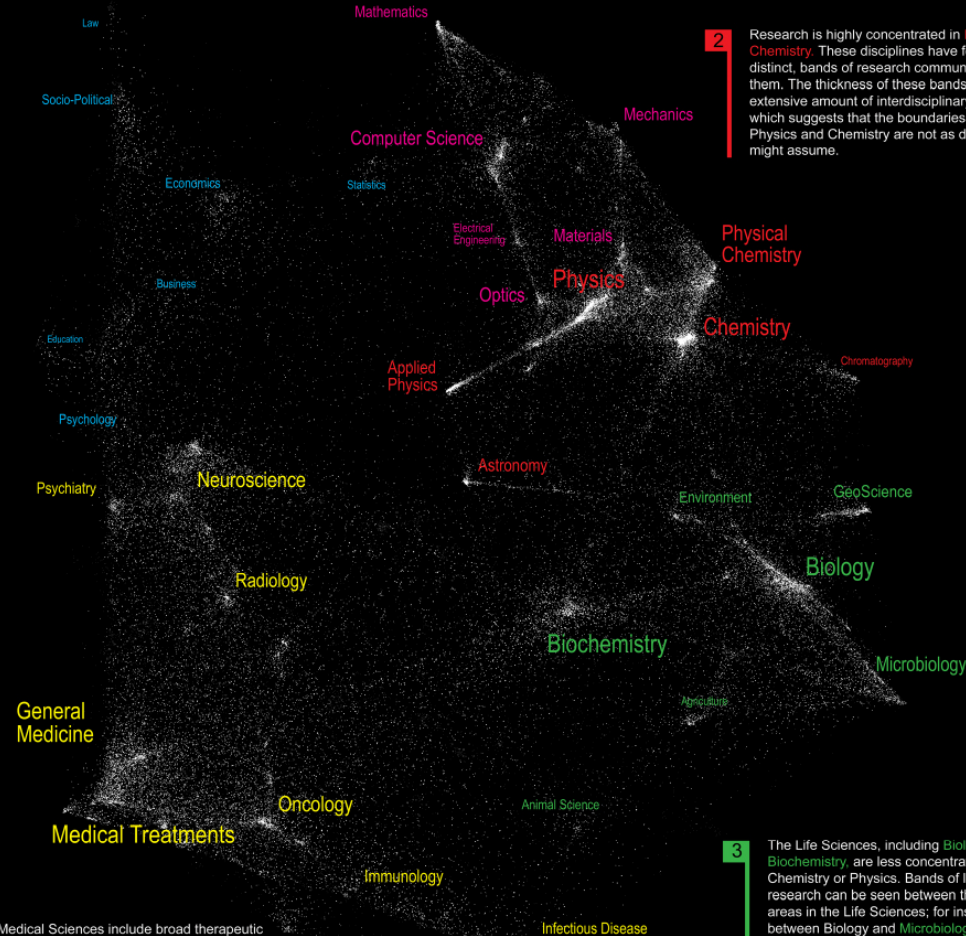
5 The Social Sciences are the smallest and most diffuse of all the sciences. **Psychology** serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. **Statistics** serves as the link with Computer Science and Mathematics.

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.

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.

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.



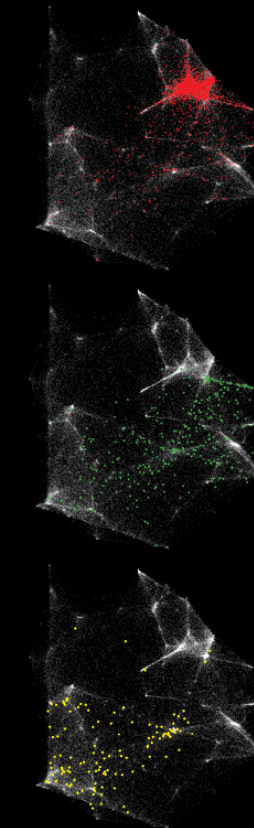
We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

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

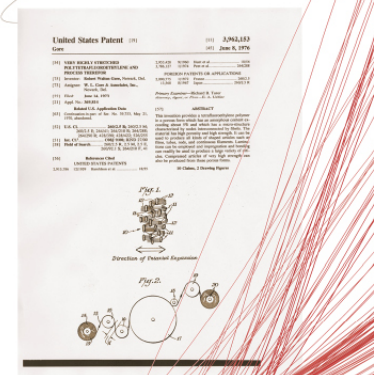
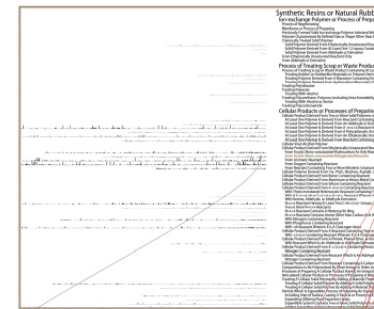
Pharmacogenomics is a relatively new field with most of its activity in **Medicine**. It also has many communities in **Biochemistry** and two communities in the Social Sciences.

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.

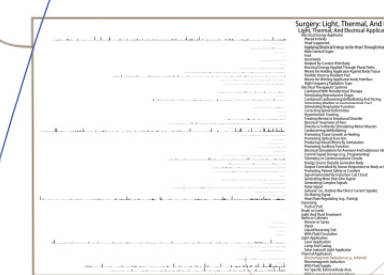
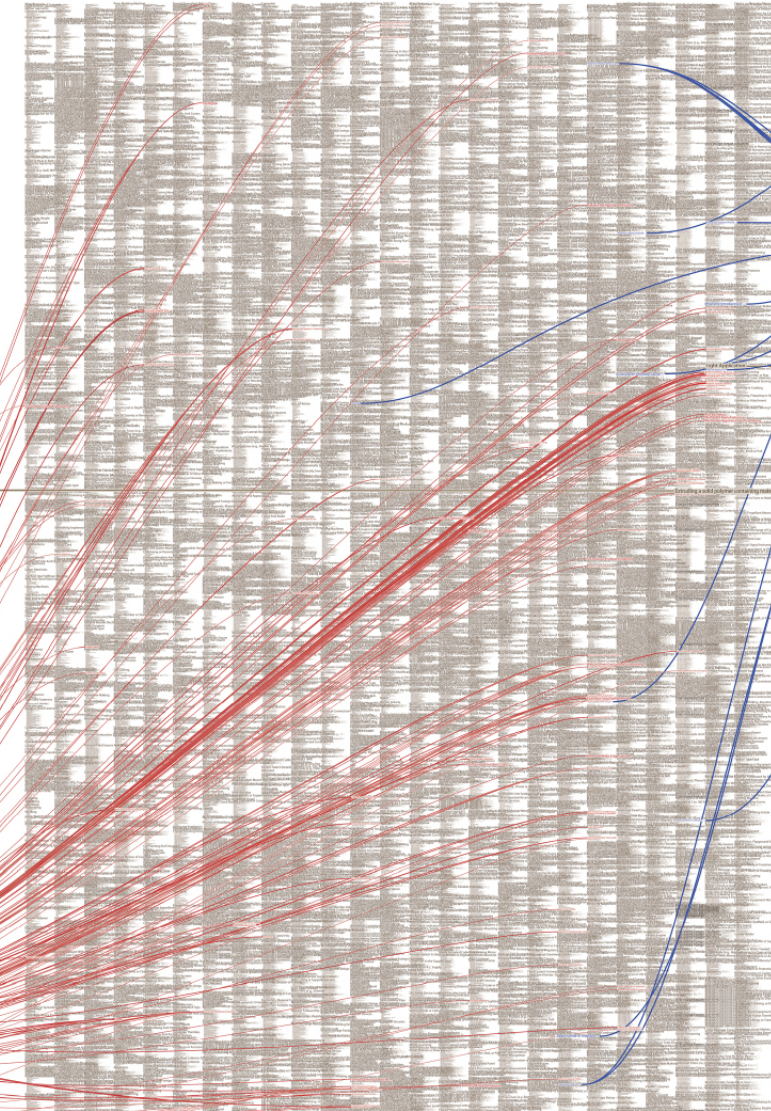
Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

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?

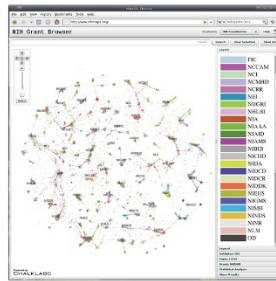
Something as simple as a bar chart helps people see how entities in a category relate to that category. Here, each bar encodes a "distance to prototype": how much each patent differs from an idealized "prototype patent" for that category. A measure like this can be based on statistics, computational 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.

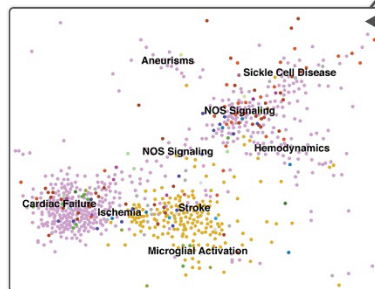
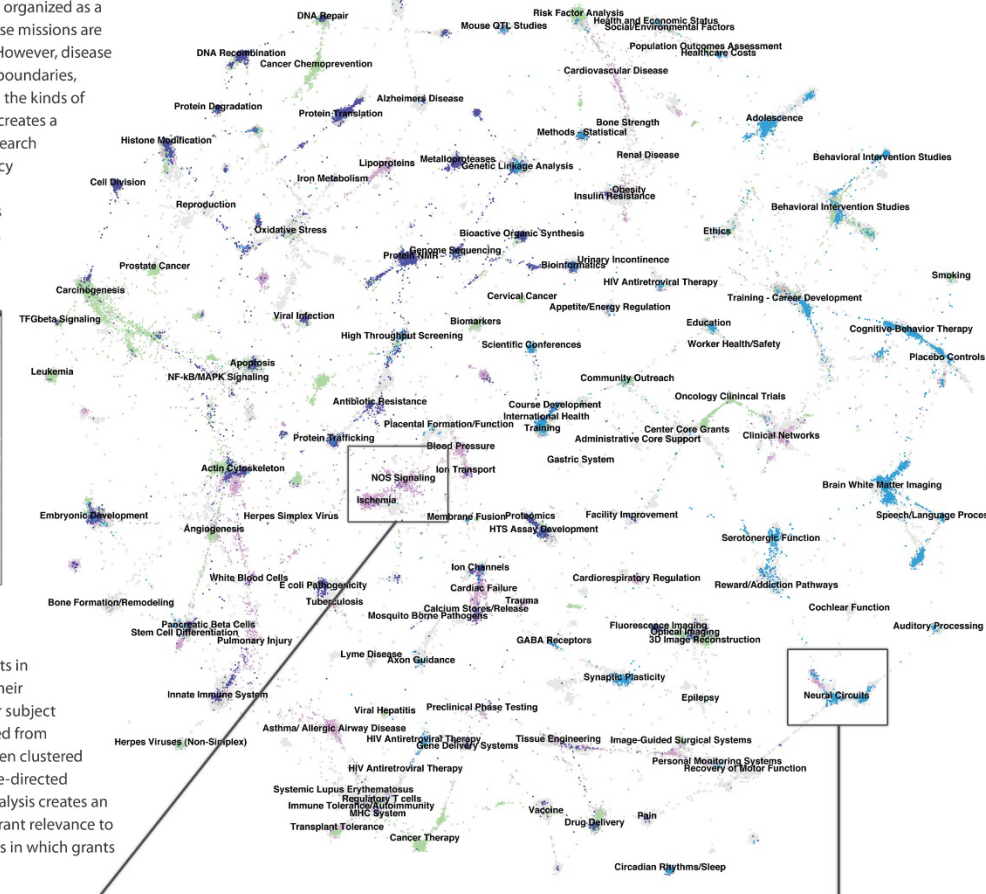
A Topic Map of NIH Grants 2007

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

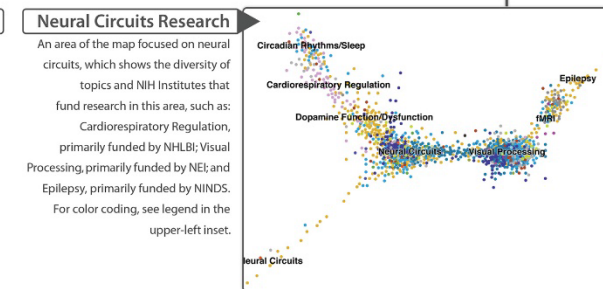
The National Institutes of Health (NIH) is organized as a multitude of Institutes and Centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies flout scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each 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.



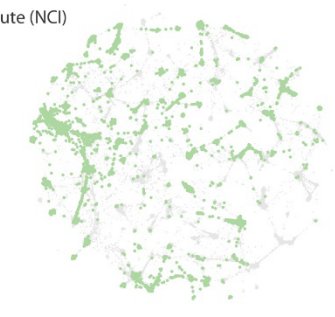
Cardiac Diseases Research
An area of the map focused on cardiovascular function and dysfunction. Cardiac Failure (primarily funded by NHLBI) is typically clustered next to Stroke (NINDS), since these are the two major medical emergencies associated with ischemia, which results from a restricted blood 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.

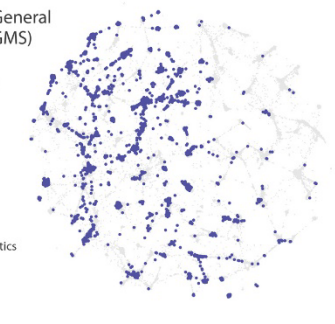
National Cancer Institute (NCI)

- 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



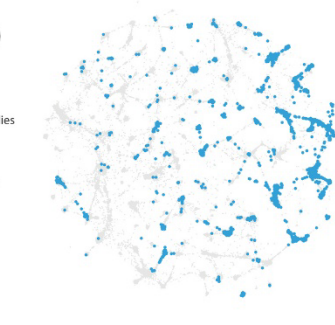
National Heart, Lung, and Blood Institute (NHLBI)

- 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





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

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

Top 5 Diseases

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

Top 5 Genes

1. TP53
2. PAK6
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 NETWORK 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 user filters diseases according to their selected genes, meaning the understanding of the roots of disease, and the functions of particular genes.

NETWORK RESOLUTION TECHNIQUES APPLIED

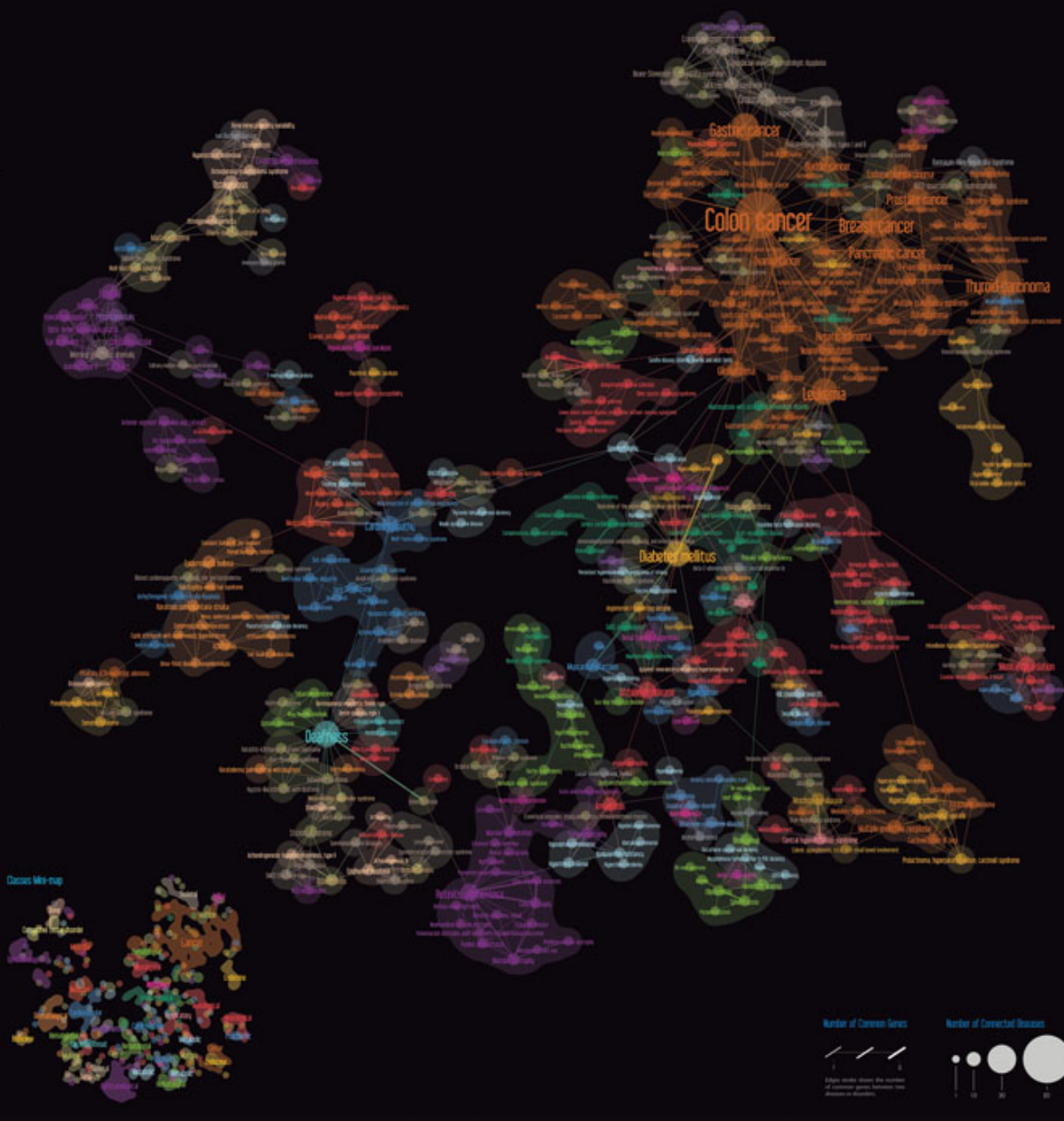
The map was done using the force-directed layout algorithm ForceAtlas in Gephi. Node sizes correspond to the number of genes that are implicated in both diseases and colored with the average color between source and target nodes. Isolated diseases are not shown and only the giant component has been kept. The Clusters Mini-map shows the most remarkable clusters and shows largest visual clusters.

The Disorder Class Interactions 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): e7000

Disorder Class Interactions



Acknowledgements

Exhibit Curators



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

Exhibit Advisory Board



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



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



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



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



Bonnie DeVarco
[Media X](#) Distinguished Visiting Scholar at Stanford University, Palo Alto, CA, USA



Olga Subirós
Curator of Big Bang Data and Founder of [Olga Subirós Studio](#) in Barcelona, Spain



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



Stephen Uzzo
Vice President of Science and Technology for the [New York Hall of Science](#)



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



Benjamin Wiederkehr
Founding Partner and Managing Director of [Interactive Things](#) in Zürich, Switzerland



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

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:

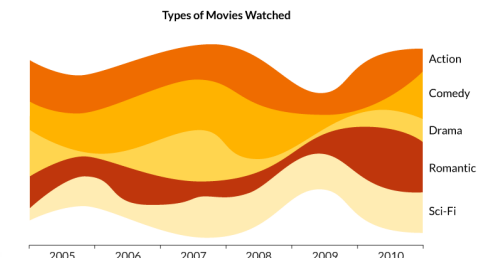
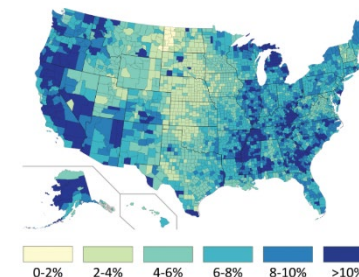
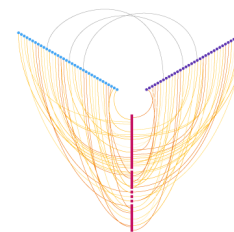
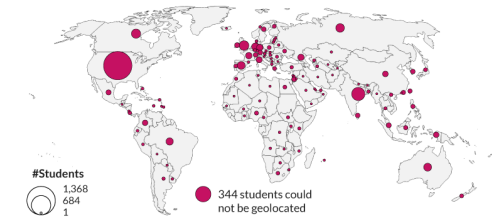
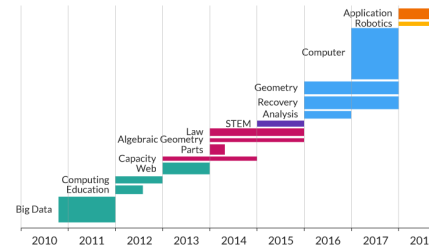
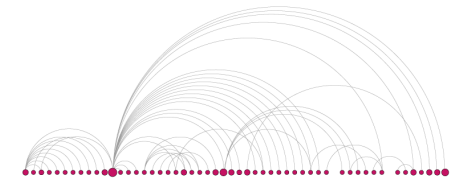
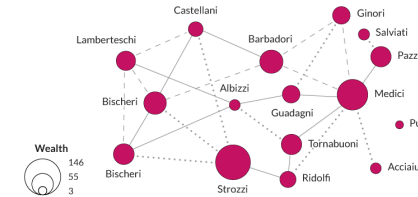
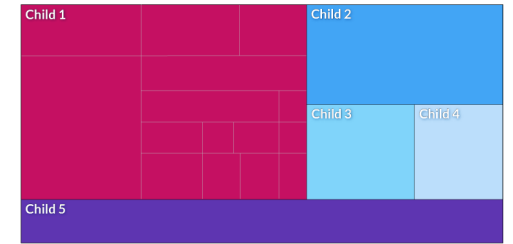
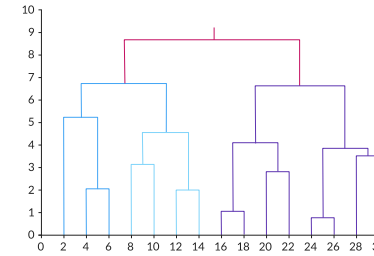
- literacy (ability to read and write text in titles, axis labels, legends, etc.),
- visual literacy (ability to find, interpret, evaluate, use, and create images and visual media), and
- mathematical literacy (ability to formulate, employ, and interpret math in a variety of contexts).

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.

Visualization Frameworks

MANY frameworks and taxonomies have been proposed to

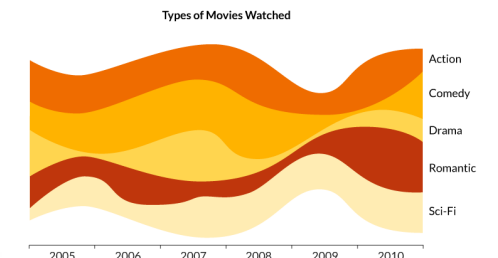
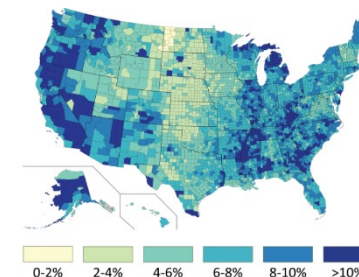
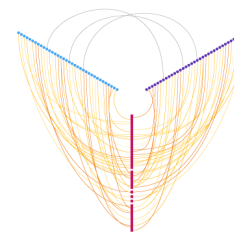
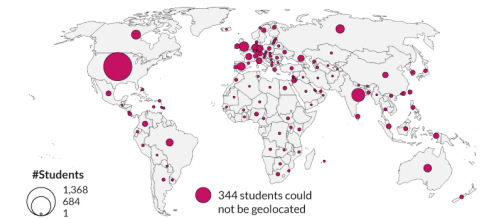
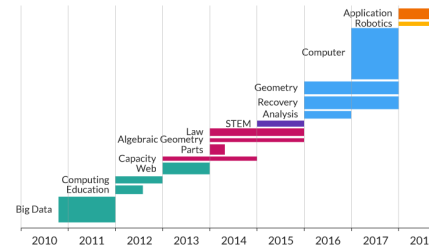
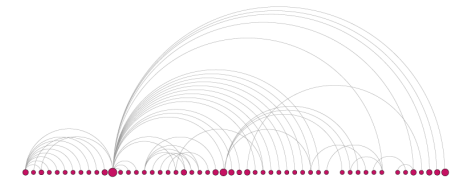
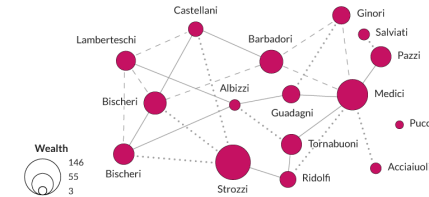
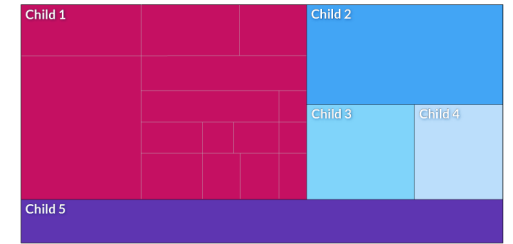
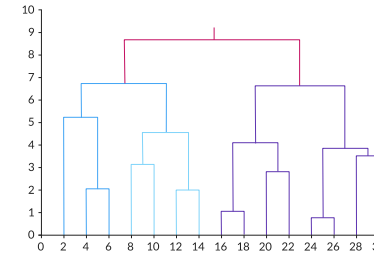
- help organize and manage the evolving zoo of 500+ different data visualization types,
- provide guidance when designing data visualizations, and
- facilitate teaching.



Existing Visualization Frameworks

Organize data visualizations by

- User insight needs
- User task types
- Data to be visualized
- Data transformations
- Visualization technique
- Visual mapping transformations
- Interaction techniques
- Deployment options
- and other features ...



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

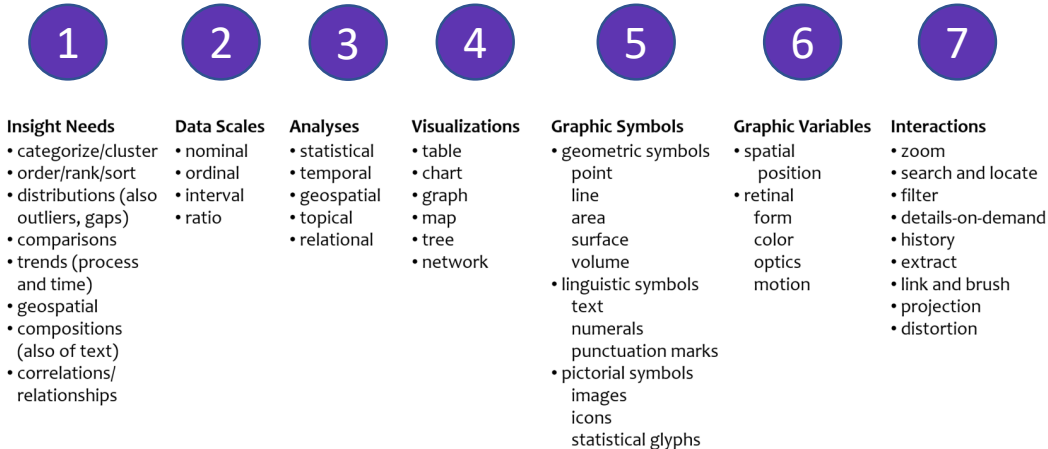
- The initial DVL-FW was developed via an extensive literature review.
- The resulting DVL-FW typology, process model, exercises, and assessments were then tested in the *Information Visualization* course taught for more than 15 years at Indiana University. More than 8,500 students enrolled in the IVMOOC 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:

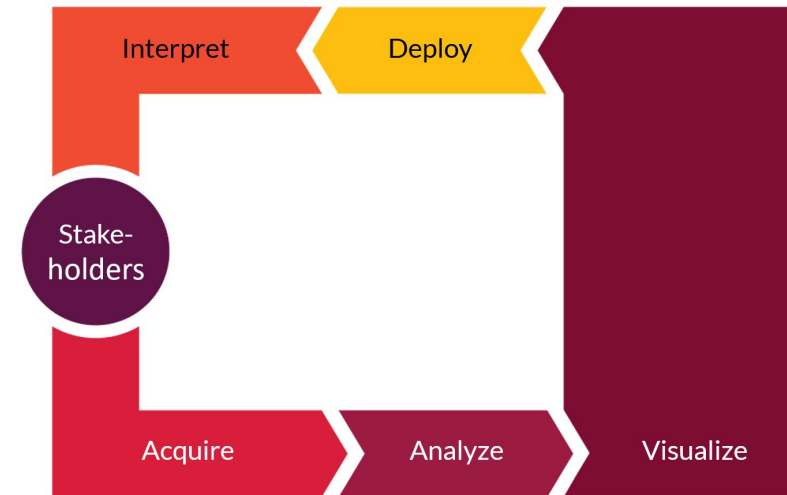
DVL Typology

Defines 7 types with 4-17 members each.



DVL Workflow Process

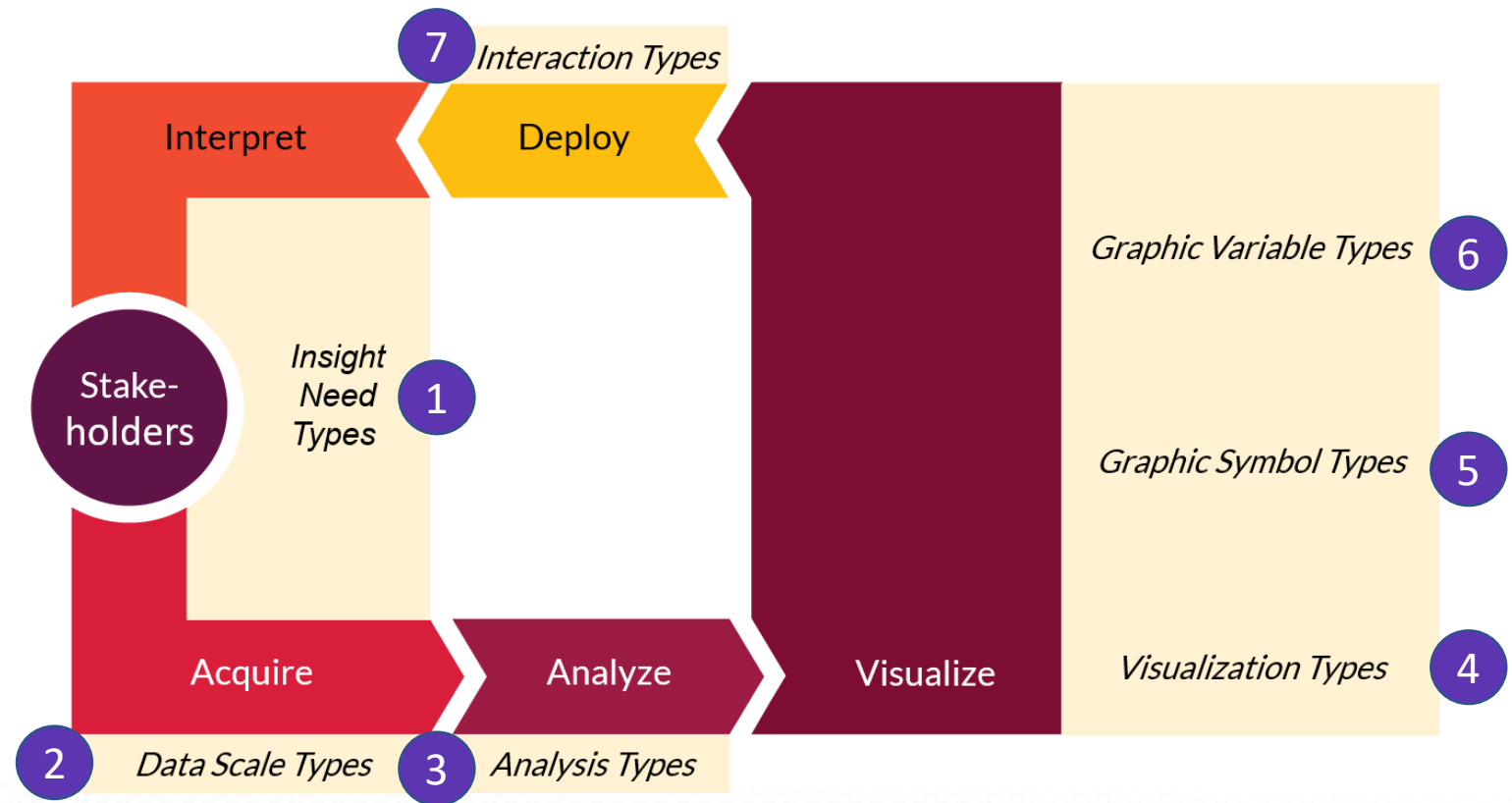
Defines 5 steps required to render data into insights.



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.

The screenshot shows the Make-A-Vis interface with three main sections:

- Data:** Contains two tables. The first table, 'ISI Publications: (CSV) Preprocessed-wos', has columns for Title, Authors, Journal, Year, and #Cites. The second table, 'Journals: (from ISI Publications)', has columns for Name, #Papers, #Cites, First Year, and Last Year.
- Make Visualization:** A panel with four visualization options: Scatter Graph, Geomap, Scimap, and Temporal Bar Graph. The Temporal Bar Graph is selected. Below these options are three dropdown menus: 'Select Graphic Symbol Type(s)', 'Select Graphic Variable Types', and 'Done'.
- Temporal Bar Graph:** A chart showing data from 1998 to 2017. The y-axis lists various categories, and the x-axis shows years. The bars represent the duration of each category's activity.

Annotations on the interface include:

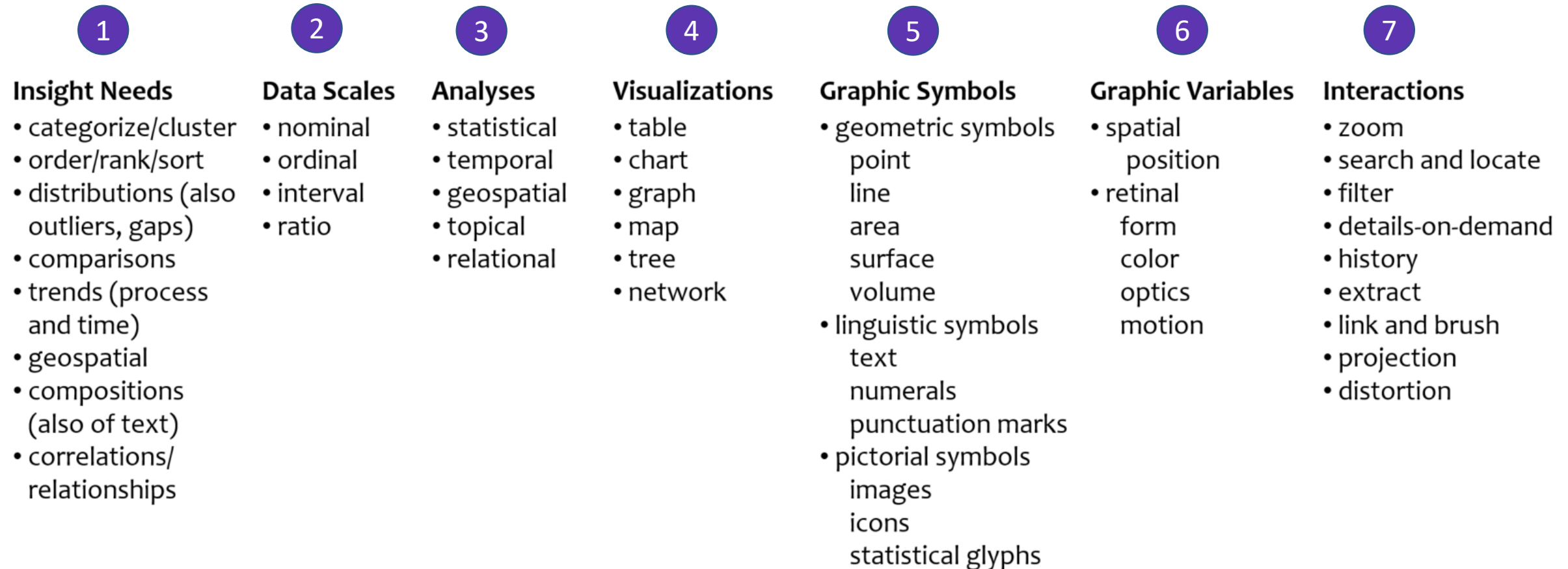
- A blue circle with the number '4' pointing to the 'Make Visualization' panel.
- A blue circle with the number '5' pointing to the 'Select Graphic Symbol Type(s)' dropdown.
- A blue circle with the number '6' pointing to the 'Select Graphic Variable Types' dropdown.

Title	Authors	Journal	Year	#Cites
[Redacted]				

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

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

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

2

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. 28-29.

Data Scale Types

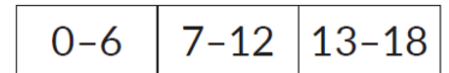
Nominal: A categorical scale, also called a nominal or category scale, is **qualitative**. Categories are assumed to be non-overlapping.



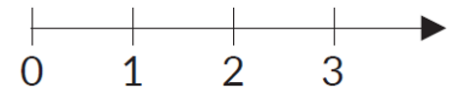
Ordinal: An ordinal scale, also called sequence or ordered, is **quantitative**. It rank-orders values representing categories based on some intrinsic ranking, but not at measurable intervals.



Interval: An interval scale, also called a value scale, is a **quantitative** numerical scale of measurement where the distance between any two adjacent values (or intervals) is equal, but the zero point is arbitrary.



Ratio: A ratio scale, also called a proportional scale, is a **quantitative** numerical scale. It represents values organized as an ordered sequence, with meaningful uniform spacing, and a true zero point.



Data Scale Types - Examples

Nominal: Words or numbers constituting the “categorical” names and descriptions of people, places, things, or events.

Ordinal: Days of the week, degree of satisfaction and preference rating scores (e.g., using a Likert scale), or rankings such as low, medium, high.

Interval: Temperature in degrees or time in hours. Spatial variables such as latitude and longitude are interval.

Ratio: Physical measures such as height, weight, (reaction) time, or intensity of light; number of published papers, co-authors, citations.

Data Scale Types - Examples

Nominal: Words or numbers constituting the “categorical” names and descriptions of people, places, things, or events.

Qualitative

Ordinal: Days of the week, degree of satisfaction and preference rating scores (e.g., using a Likert scale), or rankings such as low, medium, high.







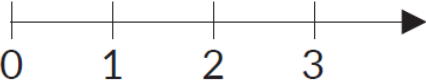
Quantitative

Interval: Temperature in degrees or time in hours. Spatial variables such as latitude and longitude are interval.

Ratio: Physical measures such as weight, height, (reaction) time, or intensity of light; number of published papers, co-authors, citations.

Data Scale Types - Mathematical Operations

This table shows the logical mathematical operations permissible, the measure of central tendency, and examples for the different data scale types.

Data Scale Types	Logical Mathematical Operations				Measure of Central Tendency	Examples			
	= ≠	< >	+ -	x ÷					
Nominal	y				mode	  			
Ordinal	y	y			median	  			
Interval	y	y	y		arithmetic mean	<table border="1" data-bbox="1396 1046 1819 1115"> <tr> <td>0-6</td> <td>7-12</td> <td>13-18</td> </tr> </table>	0-6	7-12	13-18
0-6	7-12	13-18							
Ratio	y	y	y	y	geometric mean				

Qualitative

Quantitative

Typology of the Data Visualization Literacy Framework

3

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. 25.

Analysis Types

- When: Temporal Data Analysis + Statistical
- Where: Geospatial Data Analysis
- What: Topical Data Analysis
- With Whom: Network Analysis

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

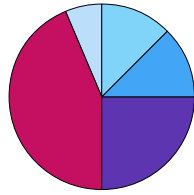
Interactions

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

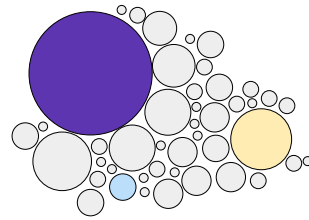
Börner, Katy. 2015. [Atlas of Knowledge: Anyone Can Map](#). Cambridge, MA: The MIT Press. 30-31.

Visualization Types

Chart

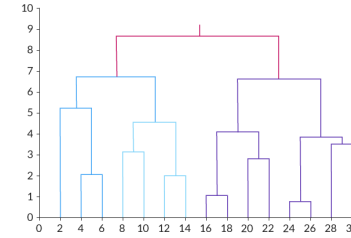


Pie Chart

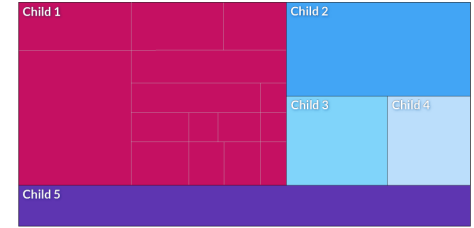


Bubble Chart

Tree

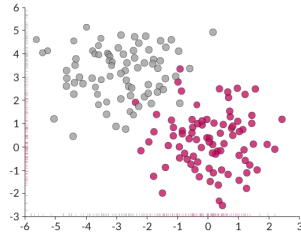


Dendrogram

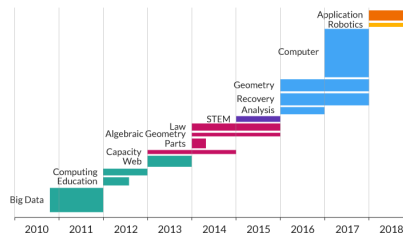


Tree Map

Graph

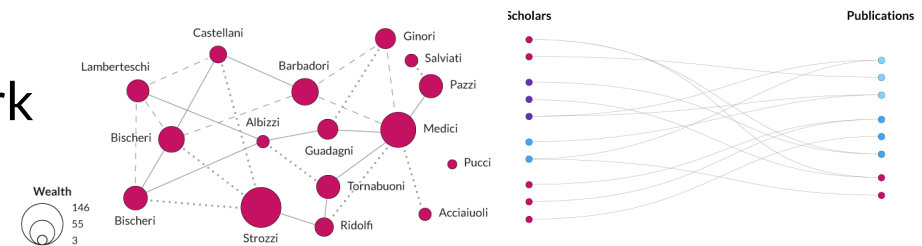


Scatter Graph



Temporal Bar Graph

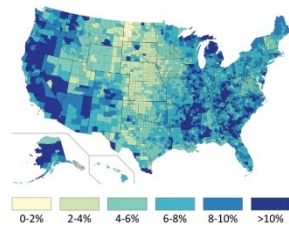
Network



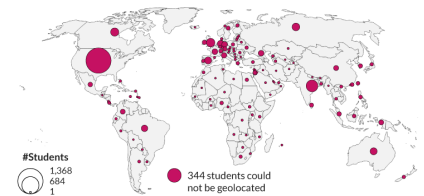
Dendrogram

Tree Map

Map



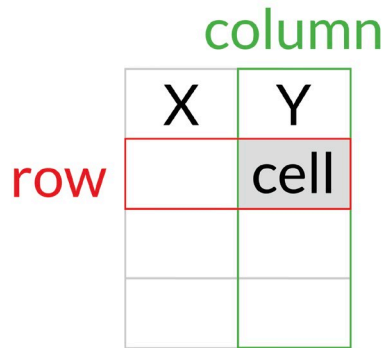
Choropleth Map



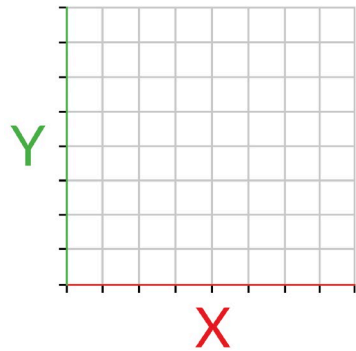
Proportional Symbol Map

Visualize: Reference Systems

Table
columns by rows



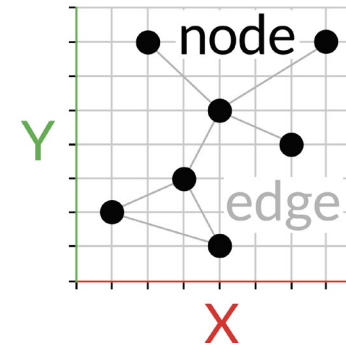
Graph
x-y coordinates



Map
latitude/
longitude



Network
local similarity

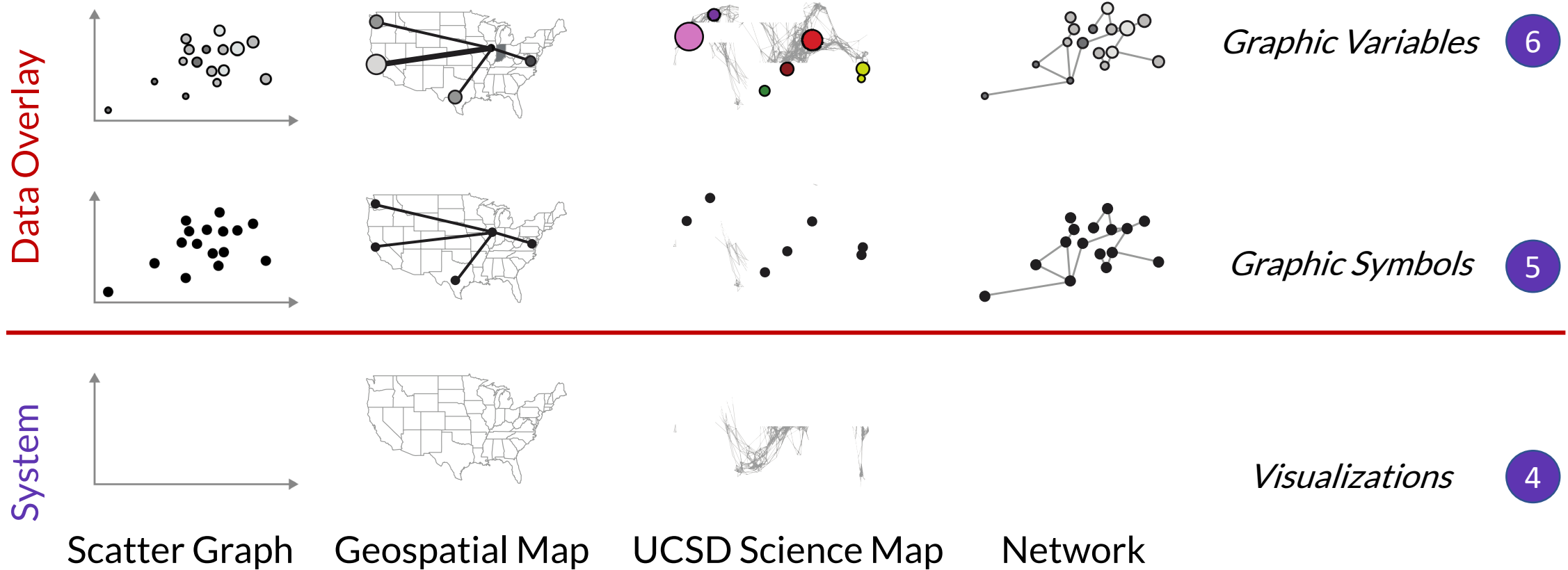


4

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

Quantitative

Form:

- Size
- Shape
- Rotation (Orientation)

Quantitative

Qualitative

Quantitative

Color:

- Value (Lightness)



Quantitative

- Hue (Tint)



Qualitative

- Saturation (Intensity)



Quantitative

Optics: Blur, Transparency, Shading, Stereoscopic Depth

Texture: Spacing, Granularity, Pattern, Orientation, Gradient

Motion: Speed, Velocity, Rhythm

Graphic Variable Types

Position: x, y; possibly z

Form:

- Size
- Shape
- Rotation (Orientation)

Color:

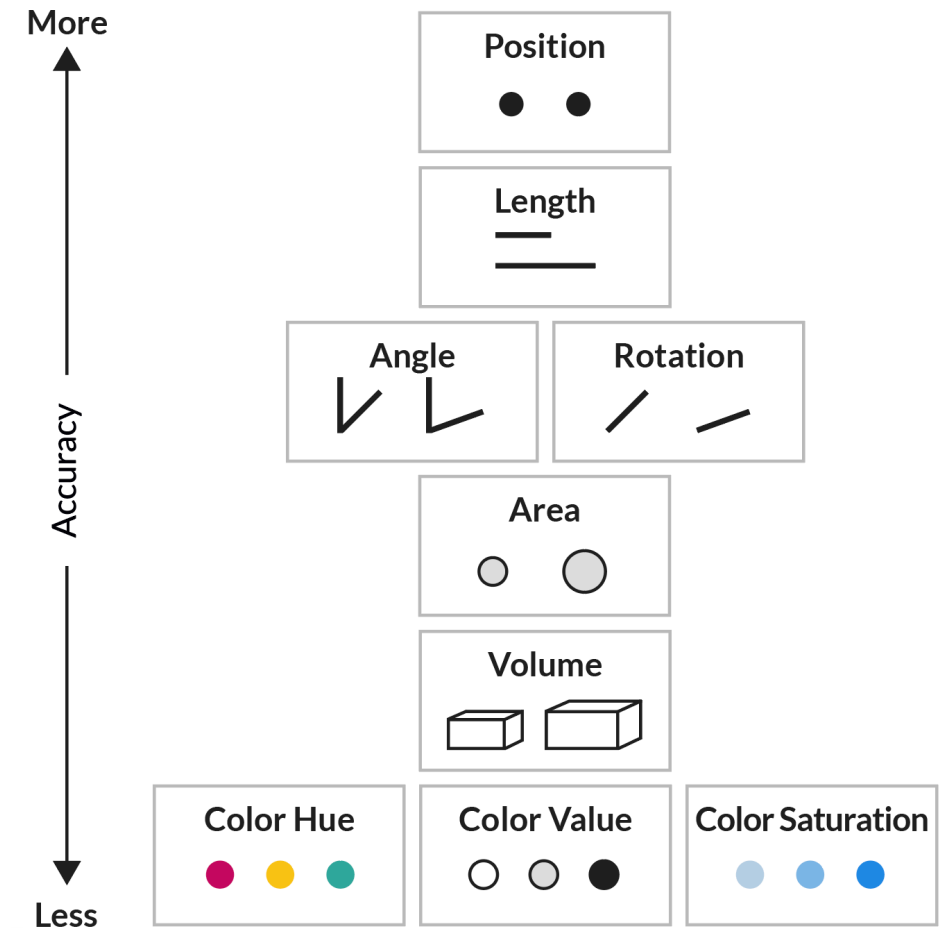
- Value (Lightness)
- Hue (Tint)
- Saturation (Intensity)



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.

Typology of the Data Visualization Literacy Framework

7

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

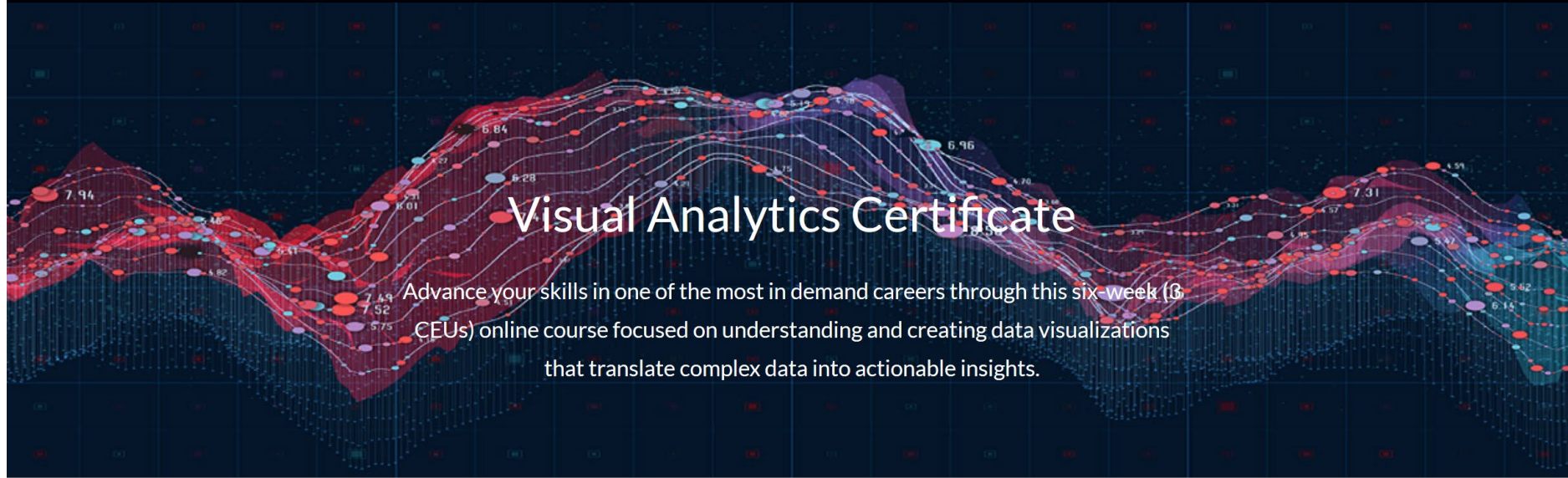
Graphic Variables

- 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, 68-69.



Visual Analytics Certificate

Advance your skills in one of the most in demand careers through this six-week (6 CEUs) online course focused on understanding and creating data visualizations that translate complex data into actionable insights.

DOWNLOAD FLYER

REGISTER FOR OCT 7-NOV 17, 2019



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>

Human BioMolecular Atlas Program (HuBMAP)

Mapping the Human Body at Cellular Resolution—
The NIH Common Fund Human BioMolecular Atlas Program
Snyder et al. <https://arxiv.org/abs/1903.07231>



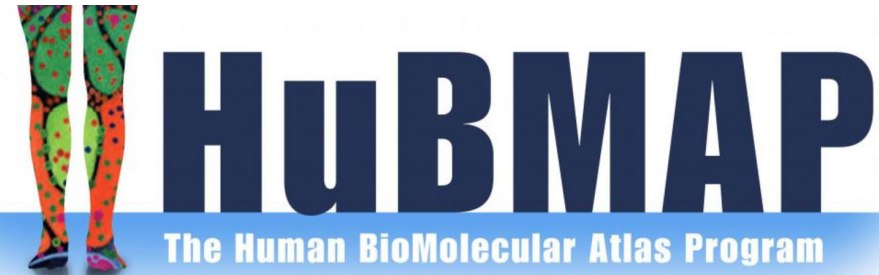
HuBMAP Goals

Vision

Catalyze the development of an open, global framework for comprehensively mapping the human body at a cellular resolution.

Goals

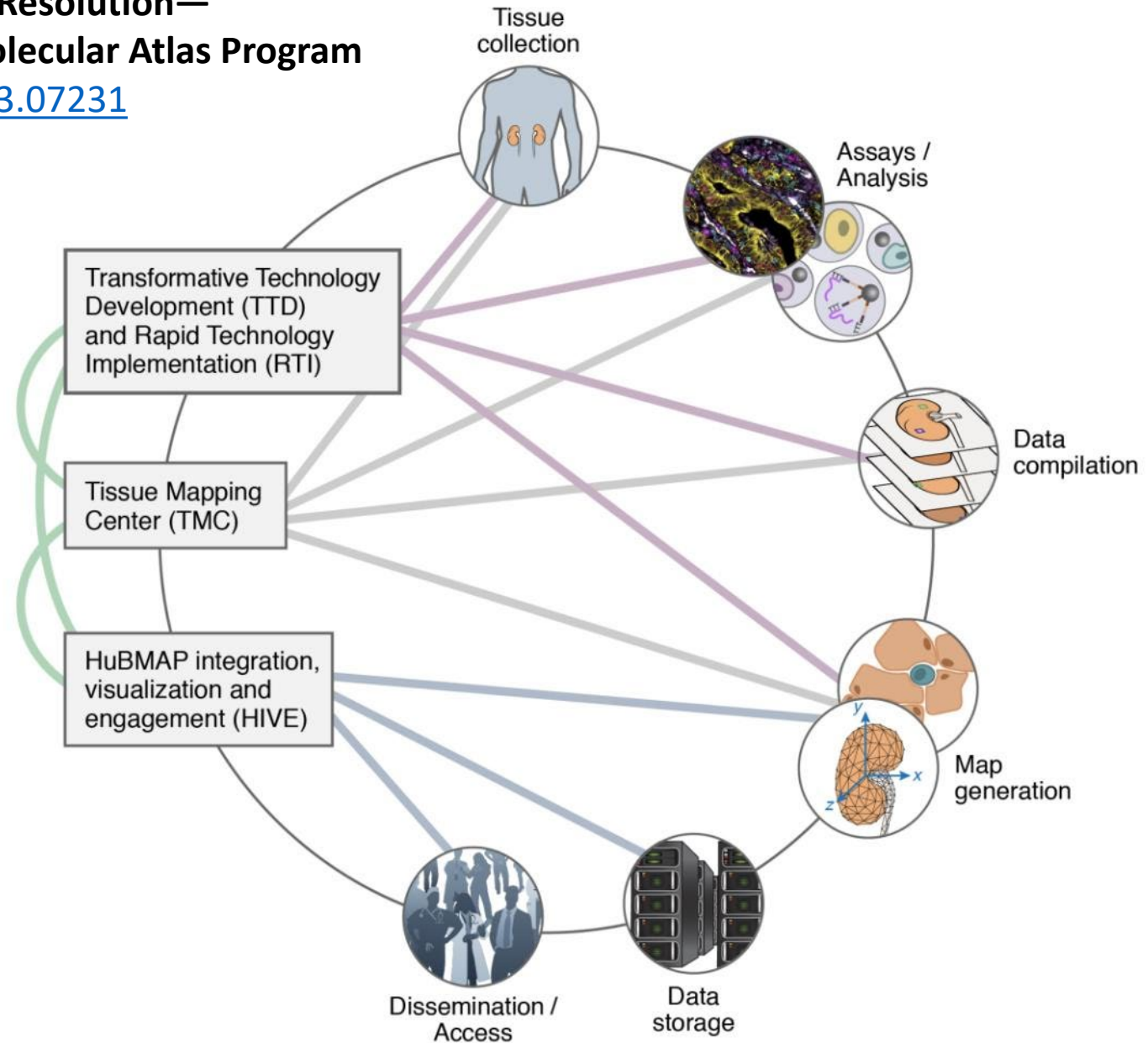
1. Accelerate the development of the next generation of tools and techniques for constructing high resolution spatial tissue maps
2. Generate foundational 3D tissue maps
3. Establish an open data platform
4. Coordinate and collaborate with other funding agencies, programs, and the biomedical research community
5. Support projects that demonstrate the value of the resources developed by the program



<https://commonfund.nih.gov/HuBMAP>

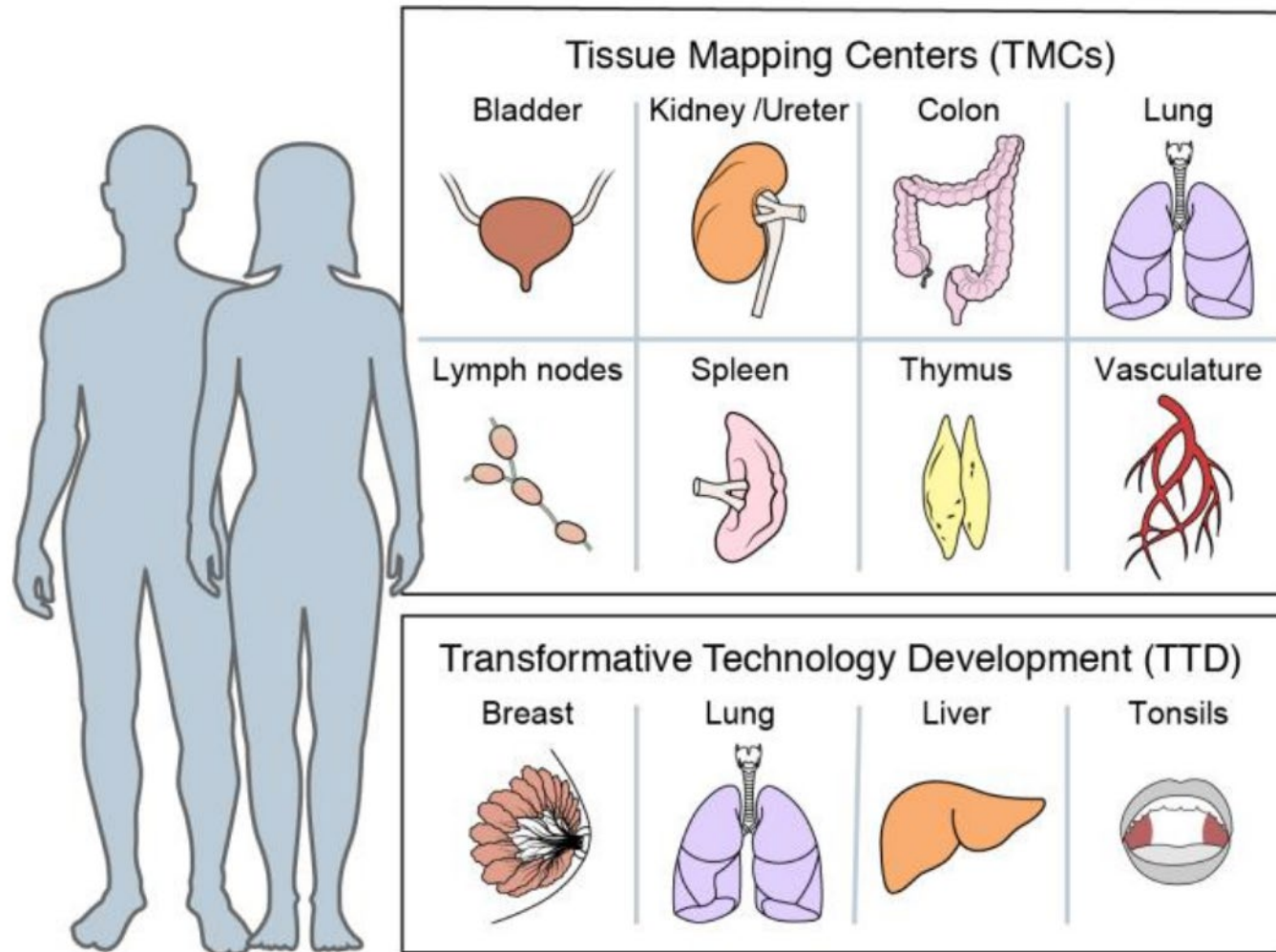
Mapping the Human Body at Cellular Resolution— The NIH Common Fund Human BioMolecular Atlas Program

Snyder et al. <https://arxiv.org/abs/1903.07231>



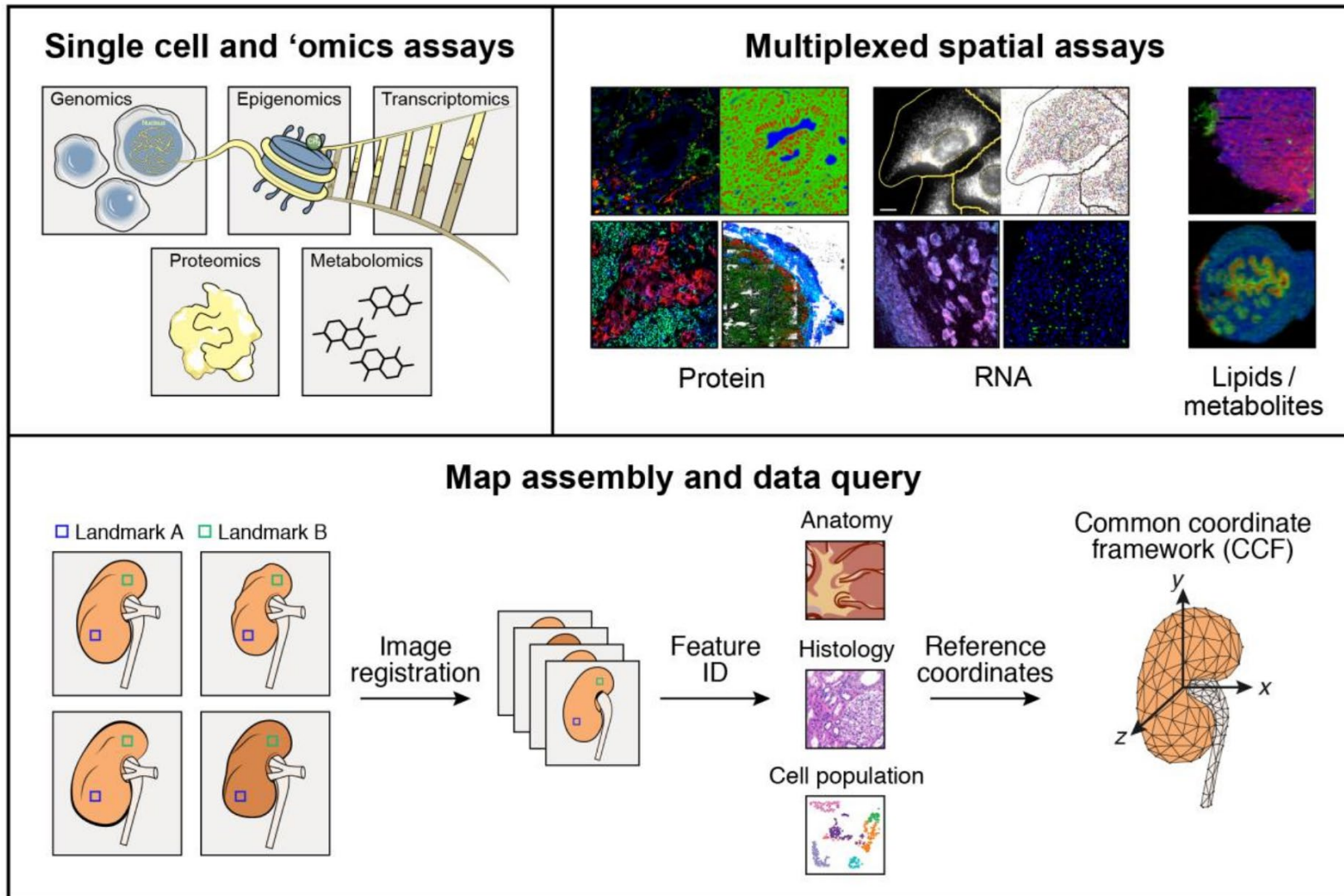
Mapping the Human Body at Cellular Resolution— The NIH Common Fund Human BioMolecular Atlas Program

Snyder et al. <https://arxiv.org/abs/1903.07231>

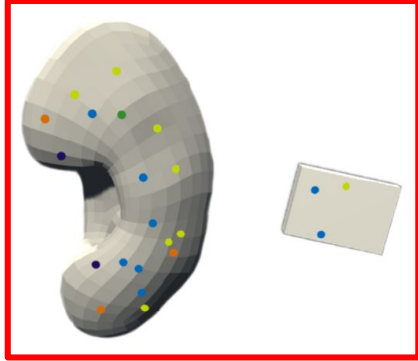


Mapping the Human Body at Cellular Resolution— The NIH Common Fund Human BioMolecular Atlas Program

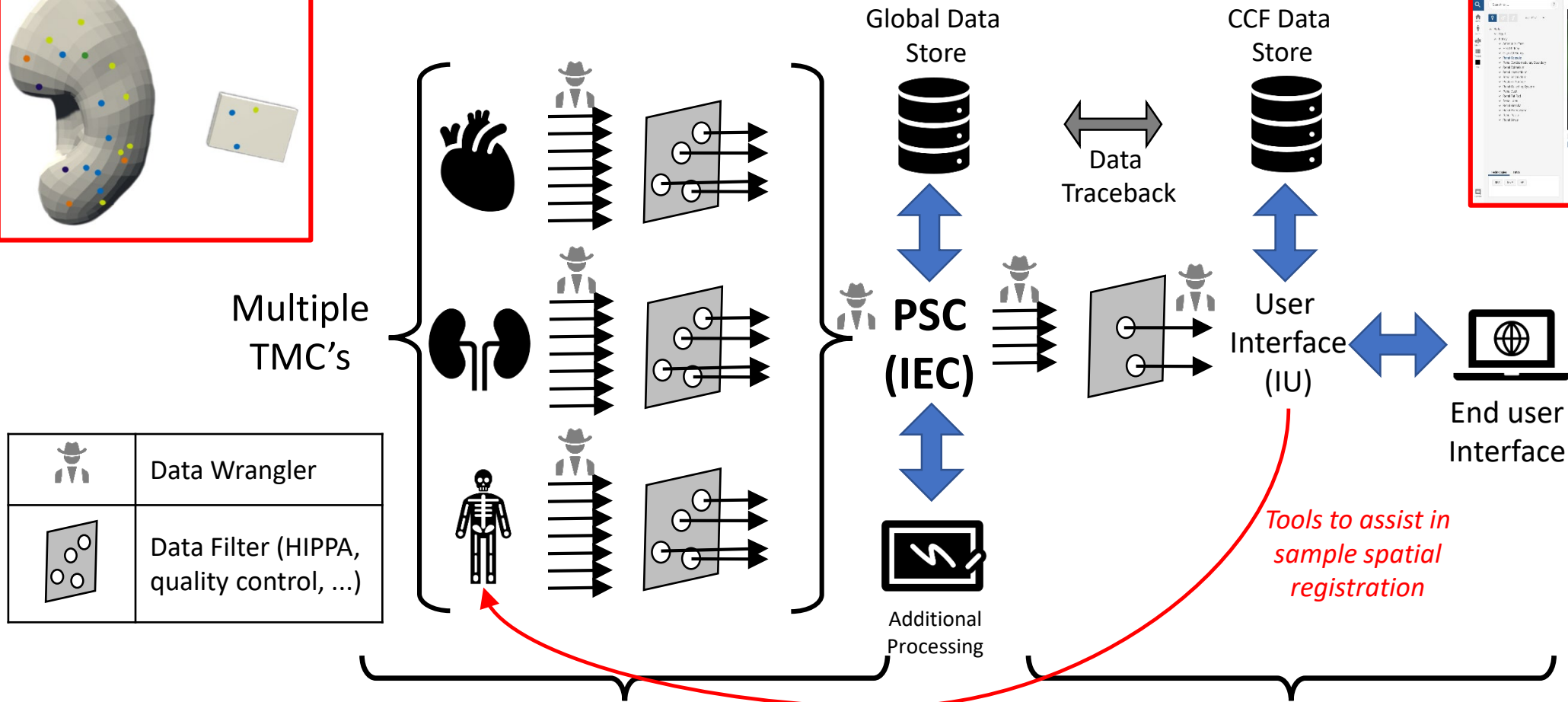
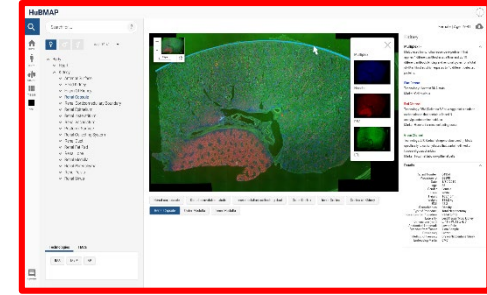
Snyder et al. <https://arxiv.org/abs/1903.07231>



Tissue Registration UI



CCF User Interface (UI)



Tools to assist in sample spatial registration

- Provenance
- Patient
- Sample
- Sample Processing
- Technology (MS, IH, ...)
- Analysis
- Etc.

Propagate needs back to TMC's

- Only the data needed for the GUI

TMC: Tissue Mapping Center
PSC: Pittsburgh Supercomputing Center

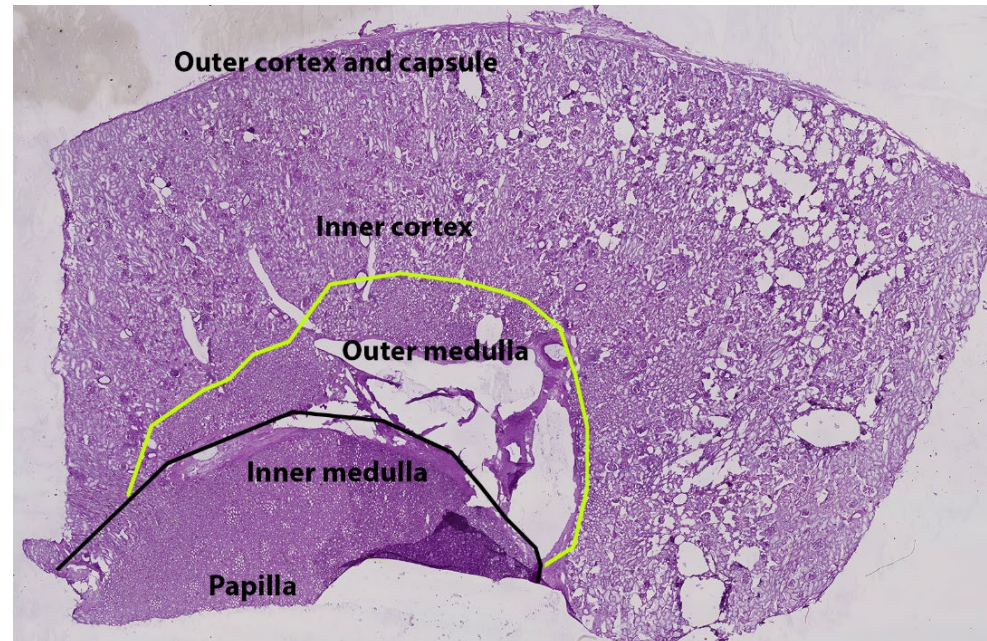
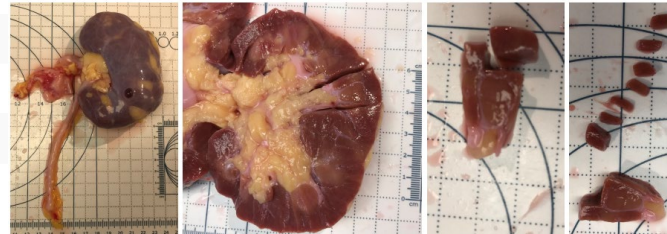
Tissue from 9 Organs | 25 Assays

Kidney: Jeff Spraggins et al., VU

See data on Globus, BIOMIC_patient-64354

- 📄 BIOMIC_patient-64354_clinical_and_spatial_metadata.xlsx
- 📄 BIOMIC_patient-64354_data_guide.pptx
- 📄 BIOMIC_patient-64354_overview.png
- 📄 BIOMIC_patient-64354_Sample-20-Histology.tif
- 📁 neg_ion_mode_section
- 📁 pos_ion_mode_section

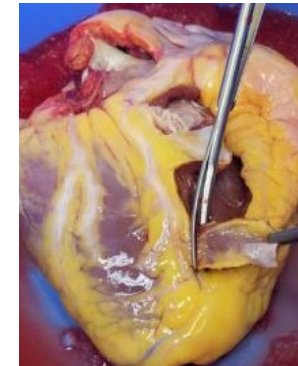
BUKMAP, Zhang Group



Heart: Shin Lin, UW

Year 1: Tissue data for 1-2cm cubed volumes from 9 sites for 1 heart from 1 individual.

Terminology; Coordinates and photos to spatialize



Sites	<u>Distinctive features</u>
1. LV, apex	
2. LV, free wall 3 cm from apex	
3. septum, 3cm from apex including LAD	major arterial vessel, Purkinje fiber CM
4. RV, free wall 3 cm from apex	
5. RA appendage	
6. RA, SA node to AV node	pacemaker CM
7. LA, appendage	
8. LA, PV inflow	
9. Posterior, adjacent to coronary sinus	major venous vessel

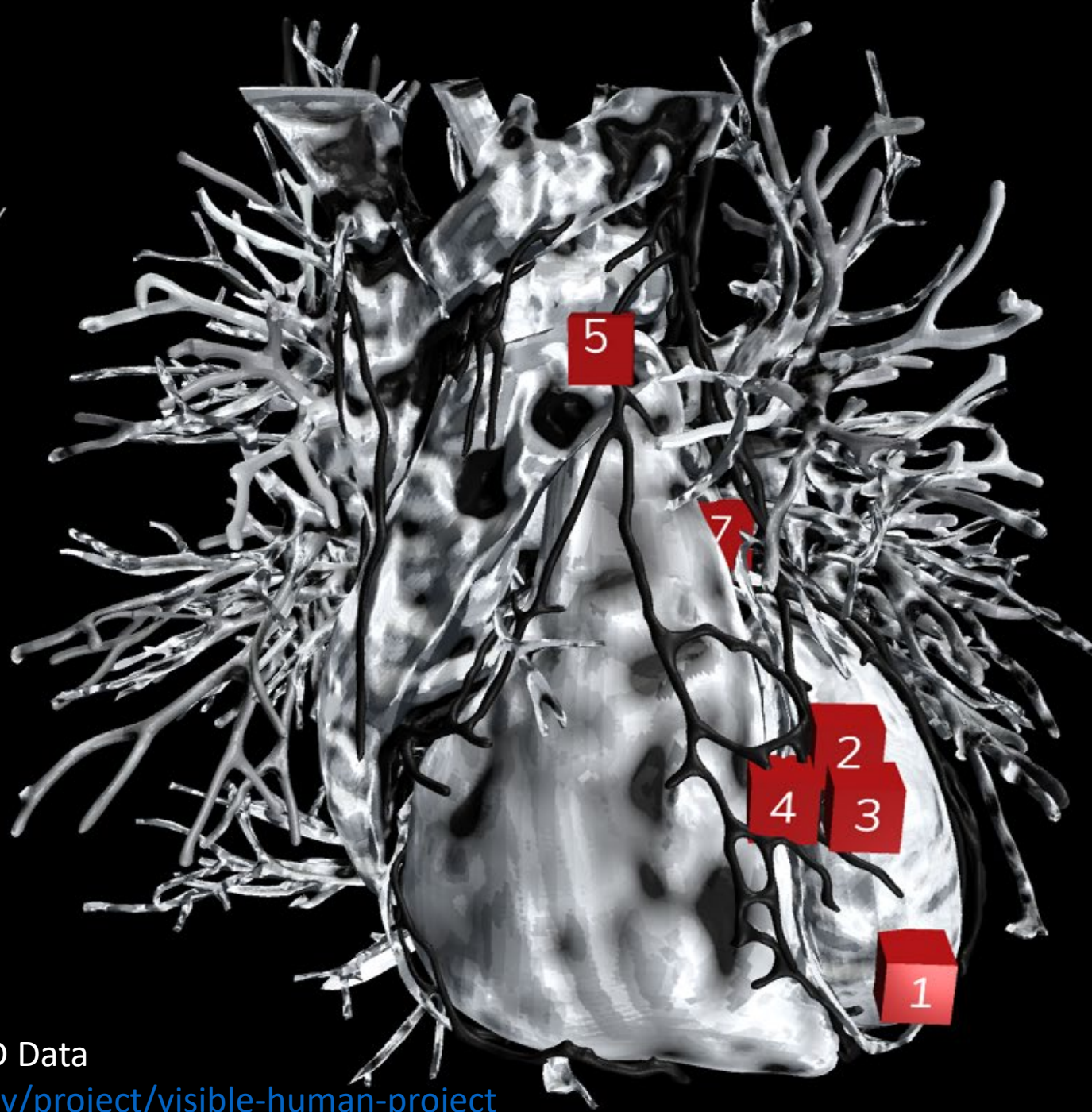
Heart

Human heart with data overlay
Developer: Andreas Bueckle

Show/hide

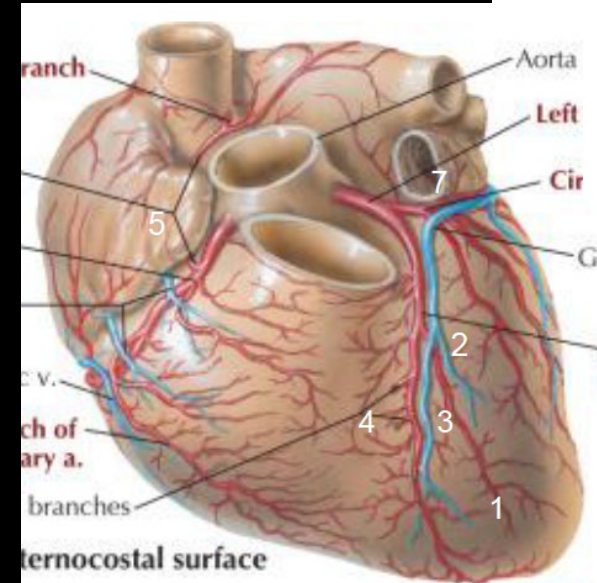
- Coronary arteries
- Coronary veins
- Left atrium
- Left ventricle
- Right atrium
- Right ventricle
- Markers

Adjust camera speed



Currently Selected

Please click any of the red markers!



Heart model from NLM3D Data

<https://lhncbc.nlm.nih.gov/project/visible-human-project>

CCF Tissue Registration UI: Kidney

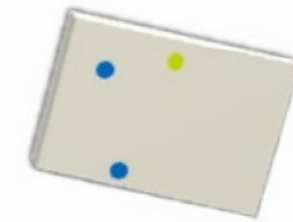
- Exploit human pattern recognition and fine motor skills (by surgeons) to register tissue in organs.
- Add info on anatomical landmarks, cell types, molecular data to support alignment.
- LATER: Use human alignment data as training data for machine learning algorithms, to better support manual alignment OR to possibly fully automatize alignment.



VIVE™ | VIVE Virtual Reality System
vive.com

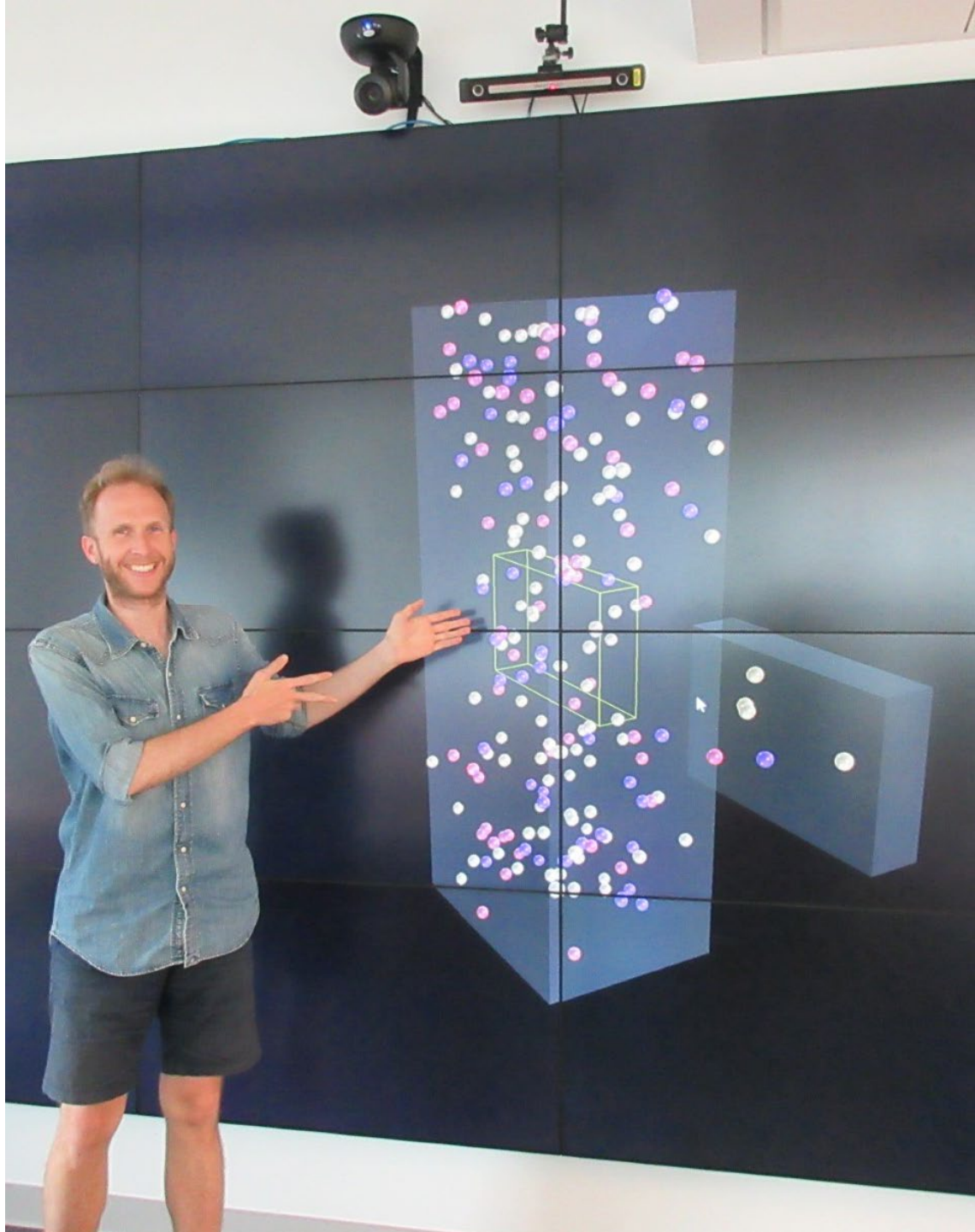


**How many of you have used
a VIVE or space mouse?**



Kidney model from NLM3D Data

<https://lhncbc.nlm.nih.gov/project/visible-human-project>



Kidney model from NLM3D Data

<https://lhncbc.nlm.nih.gov/project/visible-human-project>



CCF User Interface (UI)

HuBMAP Semantic Search

Search for ...

Female | Age 19-41

Semantic Browsing

- Body
 - Heart
 - Kidney
 - Anterior Surface
 - Head Kidney
 - Hilum Of Kidney
 - Renal Capsule
 - Renal Corticomedullary Boundary
 - Renal Epithelium
 - Renal Interstitium
 - Renal Vasculature
 - Posterior Surface
 - Renal Collecting System
 - Renal Duct
 - Renal Fat Pad
 - Renal Lobe
 - Renal Medulla
 - Renal Parenchyma
 - Renal Pelvis
 - Renal Sinus

Semantic Filters

Technologies TMCs

IMS MxIF AF

Kidney

Multiplex IF

Multiplex

Blue Channel

Technology: Hoechst DNA stain
Marker: Cell nucleus

Red Channel

Technology: DBA (Dolichos Biflorus Agglutinin) binds to carbohydrates that contain a-linked N-acetylgalactosamine residues
Marker: General for renal collecting ducts.

Green Channel

Technology: LTO (Lotus tetragonolobus Lectin) binds specifically to carbohydrates that contain α-linked L-fucose oligosaccharides
Marker: Proximal tubule epithelial cells

Details

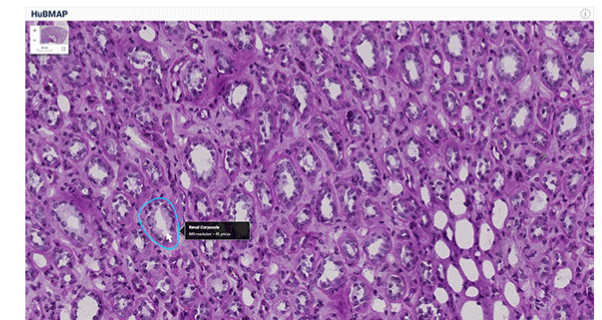
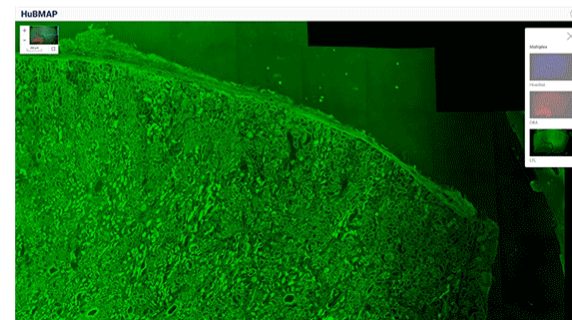
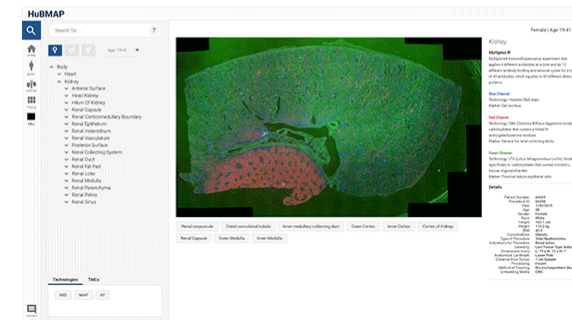
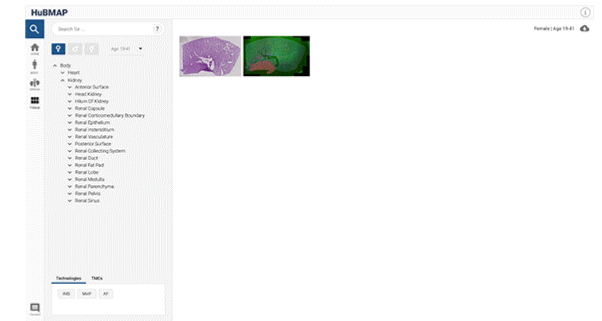
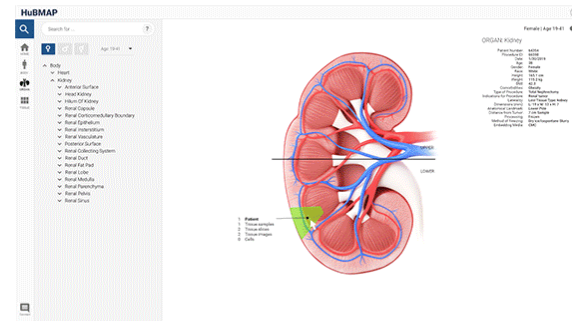
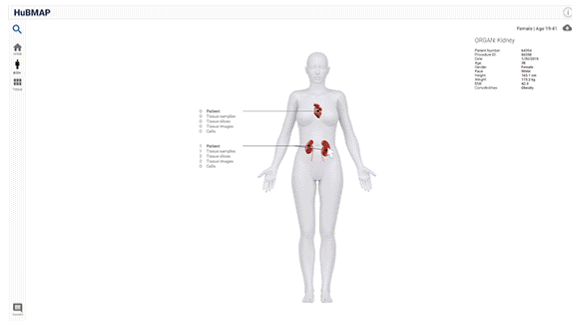
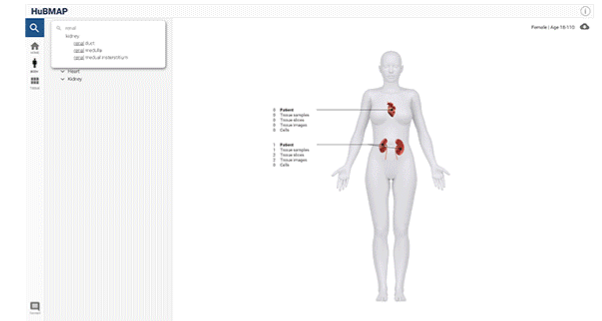
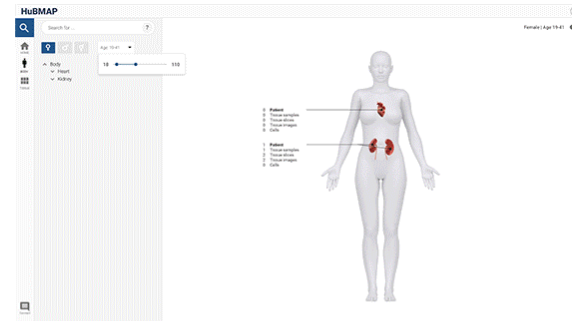
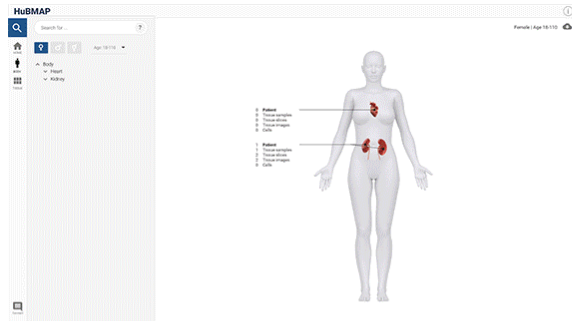
Patient Number:	64354
Procedure ID:	66598
Date:	1/30/2019
Age:	38
Gender:	Female
Race:	White
Height:	165.1 cm
Weight:	115.2 kg
BMI:	42.3
Comorbidities:	Obesity
Type of Procedure:	Total Nephrectomy
Indications for Procedure:	Renal tumor
Laterality:	Left
Tissue Type:	kidney
Dimensions (mm):	L: 19 x W: 13 x H: 7
Anatomical Landmark:	Lower Pole
Distance from Tumor:	7 cm
Processing:	Frozen
Method of Freezing:	Dry ice/Isopentane Slurry
Embedding Media:	CMC

Renal corpuscule Distal convoluted tubule Inner medullary collecting duct Outer Cortex Inner Cortex Cortex of Kidney

Renal Capsule Outer Medulla Inner Medulla

GitHub demo site: <https://hubmapconsortium.github.io/ccf-ui/>

CCF User Interface (UI)



Acknowledgements



<https://hubmapconsortium.org>

Plus, **patients** that agreed to volunteer healthy tissue and to use their data openly.



Jeffrey Spraggins
TMC-Vanderbilt
Vanderbilt University



Sanjay Jain
TMC-UCSD
Washington University, St. Louis



MC-IU HIVE Team at IU

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>

