Actionable Data Visualizations of Science & Technology

Katy Börner  @katycns

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

Allen Institute for Brain Science, Seattle, WA  |  March 7, 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.


Exemplarily, the DVL is applied to

- map science and technology, see interactive data visualizations from the *Places & Spaces: Mapping Science* exhibit ([http://scimaps.org](http://scimaps.org)) and recent *PNAS* special issue on *Modelling and Visualizing Science and Technology Developments* ([https://www.pnas.org/modeling](https://www.pnas.org/modeling))


- design reference systems and user interfaces within the Human BioMolecular Atlas Program (HuBMAP) ([https://commonfund.nih.gov/hubmap](https://commonfund.nih.gov/hubmap)) that support the exploration and communication of single-cell data—from the subcellular to the whole body level.
Places & Spaces: Mapping Science Exhibit

1st Decade (2005-2014)
Maps

2nd Decade (2015-2024)
Macrosopes

3rd Decade (2015-2034)

http://scimaps.org
We are all familiar with traditional maps that show the relationships between country boundaries, roads, and cities. Similar relationships exist between the networked 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, to depict the flow of science from the discovery papers to the citation papers. That early map was intriguing, but it didn’t come close to science to accurately define its structure.

Things are different today. We have numerous competing papers and advanced visualization software that can analyze the structure of science. This galaxy map of science (left) was generated as part of the Network of National Laboratories using an advanced graph theory model. The data for this analysis comes from the Web of Science and is based on citations. The map represents one of the 14,000 research communities active in science in 2002. Another community is composed of papers that are recent and have been cited by the previous map. Over time, communities can grow, contract, split, merge, or die.

The map of science can be used in a number of ways. The visualization tool, which was used to create these maps, allows scientists and organizations to understand their field’s scientific capabilities. Additional information about the foundation and research support that supports the community is also provided. The map allows scientists to explore which areas are explored, exploited, abandoned, or ignored.

We also envision the map as a tool for education. For children, the theoretical relationships between areas of science can be mapped with a computer map. For teachers, biology, zoology, and social studies teachers. 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 Engineering. Science. However, many disciplines at 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 transgenic and medical areas. The balance of the proteomics community is expected to continue among the Life and Medical Sciences.

Pharmacogenomics

Pharmacogenomics is a relatively new field that sits in the middle of Medicine. It also has many connections in areas of life and communities in the Social Sciences.

For the central image, each article is size-coded based on the likelihood that it is math, science, or technology related. The size of a circle reflects the likelihood of a given article being related. The larger the size, the more likely it is that the article is related to a specific field. Different colors represent different fields: science, technology, and math.

For a detailed description of the image and its implications, refer to the original publication at http://scimaps.org.
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 flow 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.scimaps.org. Institute abbreviations can be found at www.nih.gov/nci.
Check out our **Zoom Maps** online!

Visit [scimaps.org](http://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
2,500 years of human history in 5 minutes

http://idemo.cns.iu.edu/macroscope-kiosk
Government, academic, and industry leaders discussed challenges and opportunities associated with using big data, visual analytics, and computational models in STI decision-making.

Conference slides, recordings, and report are available via http://modsti.cns.iu.edu/report
Modeling and Visualizing Science and Technology Developments
National Academy of Sciences Sackler Colloquium, December 4-5, 2017, Irvine, CA

Rankings and the Efficiency of Institutions
H. Eugene Stanley | Albert-László Barabási | Lada Adamic | Marta González | Kaye Husbands Fealing | Brian Uzzi | John V. Lombardi

Higher Education and the Science & Technology Job Market
Katy Börner | Wendy L. Martinez | Michael Richey | William Rouse | Stasa Milojevic | Rob Rubin | David Krakauer

Innovation Diffusion and Technology Adoption
William Rouse | Donna Cox | Jeff Alstott | Ben Shneiderman | Rahul C. Basole | Scott Stern | Cesar Hidalgo

Modeling Needs, Infrastructures, Standards
Paul Truffio | Sallie Keller | Andrew L. Russell | Guru Madhavan | Azer Bestavros | Jason Owen-Smith

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

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

Overview
This colloquium was held in Irvine, CA on December 4-5, 2017.

This colloquium brought together researchers and practitioners from multiple disciplines to present, discuss, and advance computational models and visualizations of science and technology (S&T). Existing computational models are being applied by academia, government, and industry to explore questions such as: What jobs will exist in ten years and what career paths lead to success? Which types of institutions will likely be most innovative in the future? How will the higher education cost bubble burst affect these institutions? What funding strategies have the highest return on investment? How will changing demographics, alternative economic growth trajectories, and relationships among nations impact answers to these and other questions? Large-scale datasets (e.g., publications, patents, funding, clinical trials, stock market, social media data) can now be utilized to simulate the structure and evolution of S&T. Advances in computational power have created the possibility of implementing scalable, empirically validated computational models. However, because the databases are massive and multidimensional, both the data and the models tend to exceed human comprehension. How can advances in data visualizations be effectively employed to communicate the data, the models, and the model results to diverse stakeholder groups? Who will be the users of next generation models and visualizations and what decisions will they be addressing.

Videos of the talks are available on the Sackler YouTube Channel

https://www.pnas.org/modeling
Science Forecast
S1:E1
Data Visualization Literacy Framework

Data Visualization Literacy (DVL)

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

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

Being able to “read and write” data visualizations is becoming as important as being able to read and write text. Understanding, measuring, and improving data and visualization literacy is important to strategically approach local and global issues.
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.

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

**DVL Workflow Process**
Defines 5 steps required to render data into insights.

1. **Interpret**
2. **Acquire**
3. **Analyze**
4. **Visualize**
5. **Deploy**

- **Stakeholders**
  - **Interactions**:
    - • zoom
    - • search and locate
    - • filter
    - • details-on-demand
    - • history
    - • extract
    - • link and brush
    - • projection
    - • distortion
Data Visualization Literacy Framework (DVL-FW)

Consists of two parts that are interlinked:

DVL Typology +
DVL Workflow Process
Data Visualization Literacy Framework (DVL-FW)

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

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

2. Data Scales
   - nominal
   - ordinal
   - interval
   - ratio

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

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

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

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

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

Typology of the Data Visualization Literacy Framework

**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

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

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

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.  
**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 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.

<table>
<thead>
<tr>
<th>Data Scale Types</th>
<th>Logical Mathematical Operations</th>
<th>Measure of Central Tendency</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>yes</td>
<td>mode</td>
<td><img src="https://via.placeholder.com/150" alt="Qualitative" /></td>
</tr>
<tr>
<td>Ordinal</td>
<td>yes yes</td>
<td>median</td>
<td><img src="https://via.placeholder.com/150" alt="Qualitative" /></td>
</tr>
<tr>
<td>Interval</td>
<td>yes yes yes</td>
<td>arithmetic mean</td>
<td><img src="https://via.placeholder.com/150" alt="Quantitative" /></td>
</tr>
<tr>
<td>Ratio</td>
<td>yes yes yes yes</td>
<td>geometric mean</td>
<td><img src="https://via.placeholder.com/150" alt="Quantitative" /></td>
</tr>
</tbody>
</table>
Typology of the Data Visualization Literacy Framework

**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

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

**Insight Needs**
- categorize/clustering
- order/rank/sorting
- 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

Visualization Types

Chart
- Pie Chart
- Bubble Chart

Graph
- Scatter Graph
- Temporal Bar Graph

Map
- Choropleth Map
- Proportional Symbol Map

Tree
- Dendrogram
- Tree Map

Network
- Dendrogram
- Tree Map
Visualize: Reference Systems

**Table**
columns by rows

**Graph**
x-y coordinates

**Map**
latitude/longitude

**Network**
local similarity

**Visualization Types**
- table
- chart
- graph
- map
- network layout
Visualize: Reference Systems, Graphic Symbols and Variables

Data Overlay

Reference System

Graphic Symbols

Graphic Variables

Visualizations

Scatter Graph  Geospatial Map  UCSD Science Map  Network
## Typology of the Data Visualization Literacy Framework

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

## Typology of the Data Visualization Literacy Framework

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

Graphic Variable Types

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

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

<table>
<thead>
<tr>
<th></th>
<th>Geometric Symbols</th>
<th>Linguistic Symbols</th>
<th>Pictorial Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td>Point</td>
<td>Line</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Form</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><strong>Shape</strong></td>
<td>•</td>
<td>△</td>
<td>•</td>
</tr>
<tr>
<td>Value</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><strong>Color</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hue</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Saturation</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td><strong>Granularity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pattern</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Motion Optics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blur</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>Speed</td>
<td>•</td>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

**Qualitative**

Also called:
- Categorical Attributes
- Identity Channels

**Quantitative**

Also called:
- Ordered Attributes
- Magnitude Channels

*See Atlas of Knowledge pages 36-39 for complete table.*
# Typology of the Data Visualization Literacy Framework

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

Visual Analytics Certificate

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

Learn from Experts
Connect with industry professionals and leading researchers.

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

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

https://visanalytics.cns.iu.edu
High-Resolution, Functional Mapping of Voxel, Vector, and Meta Datasets

within

Human BioMolecular Atlas Program (HuBMAP)

https://commonfund.nih.gov/HuBMAP
High-Resolution, Functional Mapping of Voxel, Vector, and Meta Datasets within the Human BioMolecular Atlas Program (HuBMAP)

MC-Indiana Team:

Katy Börner, Bruce Herr II, Paul Macklin, Randy Heiland & Ellen Quardokus
Intelligent Systems Engineering, SICE, Indiana University, Bloomington, IN

Griffin Weber, Harvard Medical School, Boston, MA

Samuel Friedman, Opto-Knowledge Systems, Inc.
HuBMAP: HIVE

- **Tools Component 1** – Gehlenborg (Harvard)
- **Tools Component 2** – Bar-Joseph, Paten, Teichmann (Pittsburgh)
- **Infrastructure Component** – Nystrom (PSC)
- **Mapping Component 1** – Borner (Indiana)
- **Mapping Component 2** – Satija, Marioni (NYGC)
- **Coordination Component** – Mabee, Lenhardt (South Dakota)

---

**Coordinate and promote collaboration in the HuBMAP Consortium and the wider research community**

**Integrate TMC data with existing data to build spatial maps of biomolecular distribution for the human body**

**Build and manage the data platform and support the internal and external facing IT tools for the Consortium**

**Build and manage suite of search, analysis and visualization tools for HuBMAP data**
HuBMAP: HIVE Mapping Components (MC)

Mapping Component 2 – Satija, Marioni (NYGC)

Mapping Component 1 – Borner (Indiana)

Meet in the middle

Figure 1: Overview of MC strategy to iteratively generate a reference coordinate system. We will work with TMC to annotate initial datasets with key features and ontologies, which will serve as 'landmarks' to align images across individuals. We will modify and adapt our strategy to diverse tissues, retaining a probabilistic framework that represents uncertainty due to measurement, and inter-individual variation.

Fig. 5. CCF concept. Navigating through the global anatomical coordinate system to insert a synthetic tissue sample (from PhysiCell®) into the left liver lobe with a local coordinate system.
User Needs Analysis: Gathering User Stories

- Data was collected during the HIVE kickoff meeting held October 11-12, 2018.
- 26 user stories were provided from NIH and HIVE award project participants.
- Data collection took place via a Google Form.
- Please note that many stakeholders are not represented here. Additional data collection continues.

As a computational biologist, I would like to use HubMAP to integrate data across platforms, modalities, and organs so that I can identify key genes, pathways, cell types, and common principles.

As a cell biologist, I would like to use HubMAP to analyze data so that I can develop research in individual labs.

At what level would you like to enter the map:
26 responses
Mapping Components (MC):
Spatial maps of biomolecular data

Given anatomical and molecular data, develop and validate

1.) Terminologies/Ontologies (Semantics)
   • Reference concepts, e.g., organs, organ parts, cell types, cell states
   • Fiduciary concepts: Well defined landmarks that can be provided by TMCs and used by MC to spatially orient data with respect to 3D structures

2.) 3D Spatial Models interlinked with terminology/ontology
   • Across levels (gross anatomy/organ, tissue, cell level) using hierarchical containment to localize the sample within the body
   • Make landmarks visible in 3D models

3.) Interface for semantic and spatial search, filter, review, download of data.
   • Use ontology for query expansion (semantic search), semantic browsing, and as controlled vocabulary (e.g., turning on/off male/female or different cell states).
   • Use 3D models for spatial browsing, confirmation of proper tissue registration, exploring cell context, etc.
1.) Terminologies/Ontologies (Semantics)

Goals/Needs

• Create an ontologically based coordinate system, i.e., semantic terminology/ontology works with a 3D spatial model

• Respect existing ontologies and data standards that are machine readable, e.g., UBERON or FMA. NIH also has Common Data Elements (CDE) and other standards for clinical data, but elements need structure and to be machine readable.

• Metadata: Aim to unify across TMCs with common terms and usage

• Landmarks data: Identify what annotations are feasible for different organs/TMCs

• Molecular “image” data: Work with any any interoperable data standards (X, Y data; slice # or Z; possibly time; many well defined “color” channels), e.g., OMETiff (see work by Data Science WG)

Challenges & Possible Solutions

• Using massive ontologies—Create multiple “slim” ontologies or subsets of ontologies (an approach used in GO), annotate terms with “slim” annotations, and merge the slim ontologies for our ontology.

• Many ontologies not suitable for UI navigation—Use full ontology for query expansion/filter but show only a “slim” version containment/navigation-tree in user interface

• Ever expanding list of terms to include—Create software to accommodate automatically growing term list. Ontology should work with tools like https://www.mbfbioscience.com/tissue-mapper to permit ontology and 3D spatial model integration; will need panel of anatomists/pathologists to agree on names for new 3D structures discovered by humans or via ML
1.) Terminologies/Ontologies *(Semantics)* cont.

**Relevant Ontologies**

**Anatomic/Phenotypic**
- Uberon
- Foundational Model of Anatomy (FMA) (has anatomical terms NOT in Uberon)
- Human Phenotype Ontology (HPO)
- Phenotype and Trait Ontology (PATO)
- Organ specific: Kidney Tissue Atlas Ontology (KTAO) and LungMAP

**Tissue/Data Collection**
- Biological Spatial Ontology (BSPO)
- Ontology of Biomedical Investigations (OBI)
- EDAM (Bioinformatics concepts)

**(Sub-)Cellular**
- Cell Ontology (CL)
- Gene Ontology (GO)
- Chemical Entities of Biological Interest (ChEBI)
- RNA Ontology (RNAO)
- Protein Ontology (PR)
- Cell Behavior Ontology (CBO)

**Metadata**
- Basic Formal Ontology (BFO)
- Information Artifact Ontology (IAO)
- Ontology of units of Measure (OM)
- Provenance, Authoring and Versioning ontology (PAV)
- VIVO (Identifying researchers)

**MeSH and NCI Thesaurus**
2.) **3D Spatial Models** interlinked with terminology/ontology

**Kidney: Jeff Spraggins et al., VU**

See data on Globus, BIOMIC_patient-64354

![Kidney image]

**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

![Heart image]
3.) Interface for **semantic** and **spatial** search, filter, review, download

CCF UI Spec v0.5.0 v2

The initial ‘user story’ features a researcher interested to search for, filter out, review, and download biomolecular data in the context of the whole human body or in spatial relationship to specific organ(s), tissues, or cell types. The researcher is also able to learn more about how the data was acquired, to connect with data authors, and to submit questions and comments on the CCF UI.

Consequently, the UI will support

- Search (ontology supported semantic search) and filter by ontology, anatomy, and metadata
- Visual browsing of tissue samples and metadata at the whole body, organ, tissue, and cell level
- Connect with data authors to inquire about technology details.
- Data download at the whole body, organ, tissue, and cell level.
- Submit questions and comments on the CCF UI.

In the initial 9m, proof of concept versions of the whole body, organ, tissue and cell level views will be implemented.
References


