(Data) Visualization Literacy

Katy Börner  @katycns

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

Visualization Literacy for General Audiences - Can We Make A Difference? Panel at IEEE Vis2021 Conference http://ieeevis.org/year/2021/info/panels

October 27, 2021
Data Visualization Literacy


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.
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.
<table>
<thead>
<tr>
<th>Graphic Symbol Types</th>
<th>Geometric Symbols</th>
<th>Linguistic Symbols</th>
<th>Pictorial Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spatial Position</strong></td>
<td><strong>Point</strong></td>
<td><strong>Line</strong></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td><img src="image" alt="Point X" /></td>
<td><img src="image" alt="Line X" /></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td><img src="image" alt="Point Y" /></td>
<td><img src="image" alt="Line Y" /></td>
<td></td>
</tr>
<tr>
<td><strong>Form</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td><img src="image" alt="Size" /></td>
<td><img src="image" alt="Size" /></td>
<td><img src="image" alt="Size" /></td>
</tr>
<tr>
<td>Shape</td>
<td><img src="image" alt="Shape" /></td>
<td><img src="image" alt="Shape" /></td>
<td><img src="image" alt="Shape" /></td>
</tr>
<tr>
<td><strong>Value</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Value" /></td>
<td><img src="image" alt="Value" /></td>
<td><img src="image" alt="Value" /></td>
<td></td>
</tr>
<tr>
<td><strong>Color</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hue</td>
<td><img src="image" alt="Hue" /></td>
<td><img src="image" alt="Hue" /></td>
<td><img src="image" alt="Hue" /></td>
</tr>
<tr>
<td>Saturation</td>
<td><img src="image" alt="Saturation" /></td>
<td><img src="image" alt="Saturation" /></td>
<td><img src="image" alt="Saturation" /></td>
</tr>
<tr>
<td>Granularity</td>
<td><img src="image" alt="Granularity" /></td>
<td><img src="image" alt="Granularity" /></td>
<td><img src="image" alt="Granularity" /></td>
</tr>
<tr>
<td>Texture</td>
<td><img src="image" alt="Texture" /></td>
<td><img src="image" alt="Texture" /></td>
<td><img src="image" alt="Texture" /></td>
</tr>
<tr>
<td>Pattern</td>
<td><img src="image" alt="Pattern" /></td>
<td><img src="image" alt="Pattern" /></td>
<td><img src="image" alt="Pattern" /></td>
</tr>
<tr>
<td><strong>Motion Optics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blur</td>
<td><img src="image" alt="Blur" /></td>
<td><img src="image" alt="Blur" /></td>
<td><img src="image" alt="Blur" /></td>
</tr>
<tr>
<td>Speed</td>
<td><img src="image" alt="Speed" /></td>
<td><img src="image" alt="Speed" /></td>
<td><img src="image" alt="Speed" /></td>
</tr>
</tbody>
</table>


**Qualitative**
- Also called: Categorical Attributes
- Identity Channels

**Quantitative**
- Also called: Ordered Attributes
- Magnitude Channels
US Employers which have sent students include The Boeing Company, Eli Lilly, DOE, CDC, NSWC Crane.
Teaching Data Visualization Literacy in Science Museums
xMacroscopes in Science Museums

Data Visualization Literacy: Research and Tools that Advance Public Understanding of Scientific Data.
NSF AISL #1713567
Data Visualization Literacy
NSF AISL #1713567
Investigating Aspects of Data Visualization Literacy Using 20 Information Visualizations and 273 Science Museum Visitors


Abstract:

In the information age, a person's ability to read and make data visualizations is nearly as important as being able to read and write text. This paper reports the results of a multi-phase study conducted in informal learning environments in three U.S. science museums. The goal of the study was to determine the familiarity of youth and adult museum visitors with different visualization types. To address this, a total of 273 visitors were shown five out of 20 different visualizations that included two charts, five maps, eight graphs, and five network layouts. They were asked to judge the familiarity of the visualization, provide information on how to read it, and to provide a name, identify typical locations where they would encounter the data display and possible data sources that might be visualized in this way.

Results show that while most participants have a strong interest in science, math and art, many have a hard time naming and interpreting visualizations. Participants in this study commonly encounter visualizations in school, in books, at work, on the Internet, and in the news. Overall they were more familiar with basic charts, maps and graphs, but very few are familiar with network layouts and most have no ability in reading network visualizations. When asked how they would interpret the visualizations, most participants pointed to superficial features such as color, lines, or text as important to developing understanding. Overall, we found that participants were interested in the visualizations we presented to them, but had significant limitations in identifying and understanding them.

The results substantiate intuitions shared by many regarding the rather low level of data visualization literacy of general audiences. We hope they will help catalyze novel research on the development of easy-to-use yet effective visualizations with standardized names and guaranteed properties that can be readily used by those interested to understand and solve real world problems. Results also have implications for how information visualizations are taught and used in formal and informal education, the media, or in different professions.

**Data Visualization Literacy:** Research and Tools that Advance Public Understanding of Scientific Data. NSF AISL #1713567

Links:
- Data Collection Basics
- Instructions for Completing the Interview
- Data Collection Form
- 20 Visual Stimuli (see Figure 1)
- Refusal Log
- Blank Data Entry Spreadsheet

https://cns.iu.edu/2015-VisLit.html
Visualizations of the Scalable Precision Medicine Knowledge Engine (SPOKE)

https://spoke.ucsf.edu
The SPOKE network captures the essential structure of biomedicine and human health for discovery.

https://spoke.ucsf.edu
Lead Investigators

Sergio Baranzini, PhD
Principal Investigator

Sui Huang, MD, PhD (ISB)
Sharat Israni, PhD
Mike Keiser, PhD

Technical & Planning Team

Rafael Gonçalves, PhD (Stanford)
Adil Harroud, MD
Elaine Meng, MD
Scooter Morris, PhD
Charlotte Nelson, PhD
Boris Osokotsky, PhD
Angela Rizk-Jackson, PhD
Peter Rose, PhD (UCSD)
Brett Smith (ISB)
Karthik Soman, PhD
Xiaoyuan Zhou, PhD

Collaborators

Katy Börner, PhD (IU)
William Brown, PhD, DrPH
Ramanathan V. Guha, PhD (Google)
Mark Musen, MD, PhD (Stanford)
Camille Nebeker, EdD, MS (UCSD)
Roger Pearce, PhD (LLNL)

SPOKE investigative teams

The SPOKE team members are from the following organizations. Team members listed below are from UCSF, except when indicated.

- Google
- Indiana University (IU)
- Institute for Systems Biology (ISB)
- Lawrence Livermore National Lab (LLNL)
- Stanford University
- University of California, San Diego (UCSD)
- University of California, San Francisco (UCSF)
Envisioning SPOKE: 3M Nodes and 30M Edges

The Scalable Precision Medicine Oriented Knowledge Engine (SPOKE) graph federates about 19 open datasets into a public data commons of health-relevant knowledge. This site lets users explore the massive SPOKE knowledge graph.

The site was designed for two user groups: (1) novice users interested to understand the coverage and quality of SPOKE data and (2) expert users interested to analyze and optimize the interlinked knowledge graphs in SPOKE.

The overview visualization shows the different entity type and their diverse interlinkages. Detail

SPOKE is a fully interactive tool for exploring the interconnections between data.
HuBMAP: Mapping 30+ Trillion Cells

https://www.nature.com/articles/s41586-019-1629-x.pdf
HuBMAP

Vision
Catalyze the development of an open, global framework for comprehensively mapping the human body at 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

**Fig. 1 | The HubMAP consortium.** The TMCs will collect tissue samples and generate spatially resolved, single-cell data. Groups involved in TTD and RTI initiatives will develop emerging and more developed technologies, respectively; in later years, these will be implemented at scale. Data from all groups will be rendered useable for the biomedical community by the HuBMAP integration, visualization and engagement (HIVE) teams. The groups will collaborate closely to iteratively refine the atlas as it is gradually realized.
The Human Body at Cellular Resolution: 
The NIH Human Biomolecular Atlas Program. 

Fig. 2 | Key tissues and organs initially analysed by the consortium. 
Using innovative, production-grade (‘shovel ready’) technologies, 
HuBMAP TMCs will generate data for single-cell, three-dimensional maps 
of various human tissues. In parallel, TTD projects (and later RTI projects) 
will refine assays and analysis tools on a largely distinct set of human 
tissues. Samples from individuals of both sexes and different ages will be studied. The range of tissues will be expanded throughout the program.

**Fig. 3** | Map generation and assembly across cellular and spatial scales. HuBMAP aims to produce an atlas in which users can refer to a histological slide from a specific part of an organ and, in any given cell, understand its contents on multiple ‘omic levels—genomic, epigenomic, transcriptomic, proteomic, and/or metabolomic. To achieve these ends, centres will apply a combination of imaging, ‘omics and mass spectrometry techniques to specimens collected in a reproducible manner from specific sites in the body. These data will be then be integrated to arrive at a high-resolution, high-content three-dimensional map for any given tissue. To ensure inter-individual differences will not be confounded with collection heterogeneity, a robust CCF will be developed.
What is a Human Reference Atlas?

The Human Reference Atlas (HRA)

1. defines the 3D space and shape of anatomical structures and cell types that are of biomedical relevance plus the biomarkers used to characterize them. Anatomical structures, cell types and biomarkers are validated and represented in/added to ontologies (Uberon/FMA, CL, HGNC).

2. defines how new datasets can be mapped to the HRA, e.g., spatially using the Visible Human CCF and/or Vasculature CCF, via ASCT+B ontology terms/IDs such as gene or protein biomarkers, or via gene expression data as in Azimuth.

3. it is
   a. authoritative (there exists expert agreement and it was validated by data),
   b. computable (supports API queries, UIs, linkages; see slides #8 and #9),
   c. published as LOD (connecting to disease and other ontologies and data),
   d. open (anyone can use the HRA data and code), and
   e. continuously evolving (e.g., as new technologies become available).

Anatomical Structures (AS), Cell Types (CT), and Biomarkers (B) or ASCT+B tables aim to capture the partonomy of anatomical structures, cell types, and major biomarkers (e.g., gene, protein, lipid or metabolic markers).

### ASCT Table

<table>
<thead>
<tr>
<th>Structure/Region</th>
<th>Substructure/Subregion</th>
<th>Cell Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renal Corpuscle</td>
<td>Bowman's (glomerular) Capsule/parietal layer</td>
<td>Parietal epithelial cell</td>
</tr>
<tr>
<td></td>
<td>Bowman's (glomerular) Capsule/visceral layer</td>
<td>Podocyte</td>
</tr>
<tr>
<td></td>
<td>Glomerular Tuft</td>
<td>Capillary Endothelial cell</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mesangial cell</td>
</tr>
<tr>
<td>Tubules</td>
<td>Proximal Tubule</td>
<td>Proximal Tubule Epithelial Cell (general)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proximal Convoluted Tubule Epithelial Cell Segment 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proximal Tubule Epithelial Cell Segment 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Proximal Tubule Epithelial Cell Segment 2</td>
</tr>
<tr>
<td>Loop of Henle, Thin Limb</td>
<td>Descending Thin Limb Cell (general)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ascending Thin Limb Cell (general)</td>
</tr>
<tr>
<td>Loop of Henle, Thick Limb</td>
<td>Thick Ascending Limb Cell (general)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cortex-TAL Cell</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medulla-TAL Cell</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TAL-Macula Densa Cell</td>
</tr>
<tr>
<td>Distal Convolution</td>
<td>Distal Convoluted Tubule Cell (general)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DCT Type 1 Cell</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DCT Type 2 Cell</td>
</tr>
<tr>
<td>Connecting Tubule</td>
<td>Connecting Tubule Cell (general)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CNT Principal Cell</td>
</tr>
</tbody>
</table>

### Ontology

- **Anatomical Structures Partonomy**
  - kidney
  - kidney capsule
  - cortex of kidney
  - outer cortex of kidney
  - renal medulla

- **Cell Types Ontology**
  - connective tissue cell
  - pericyte cell
  - mesangial cell
  - extraglomerular mesangial cell
  - glomerular mesangial cell

### 3D Reference Object Library

- Kidney structure
- Kidney capsule
- Cortex of kidney
- Outer cortex of kidney
- Renal medulla

- Glomerulus structure
- Mesangial cell
- Extra glomerular mesangial cell
- Glomerular mesangial cell
An Atlas describes & names 2/3D entities

<table>
<thead>
<tr>
<th>Organ</th>
<th>#AS</th>
<th>#CT</th>
<th>#B Total</th>
<th>#BG</th>
<th>#BP</th>
<th>#AS-AS</th>
<th>#AS-CT</th>
<th>#CT-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone Marrow &amp; Blood/Pelvis</td>
<td>3</td>
<td>46</td>
<td>327</td>
<td>201</td>
<td>126</td>
<td>2</td>
<td>70</td>
<td>710</td>
</tr>
<tr>
<td>Brain</td>
<td>187</td>
<td>127</td>
<td>254</td>
<td>254</td>
<td>0</td>
<td>187</td>
<td>127</td>
<td>330</td>
</tr>
<tr>
<td>Heart</td>
<td>52</td>
<td>25</td>
<td>48</td>
<td>48</td>
<td>0</td>
<td>61</td>
<td>164</td>
<td>78</td>
</tr>
<tr>
<td>Intestine, Large</td>
<td>85</td>
<td>69</td>
<td>94</td>
<td>88</td>
<td>6</td>
<td>389</td>
<td>1,361</td>
<td>197</td>
</tr>
<tr>
<td>Kidney</td>
<td>88</td>
<td>63</td>
<td>152</td>
<td>152</td>
<td>0</td>
<td>67</td>
<td>59</td>
<td>257</td>
</tr>
<tr>
<td>Lung</td>
<td>161</td>
<td>92</td>
<td>176</td>
<td>172</td>
<td>4</td>
<td>1,633</td>
<td>12,094</td>
<td>286</td>
</tr>
<tr>
<td>Lymph Node</td>
<td>41</td>
<td>49</td>
<td>286</td>
<td>108</td>
<td>158</td>
<td>62</td>
<td>135</td>
<td>544</td>
</tr>
<tr>
<td>Skin</td>
<td>16</td>
<td>42</td>
<td>70</td>
<td>70</td>
<td>0</td>
<td>17</td>
<td>19</td>
<td>105</td>
</tr>
<tr>
<td>Spleen</td>
<td>46</td>
<td>66</td>
<td>255</td>
<td>80</td>
<td>145</td>
<td>68</td>
<td>172</td>
<td>414</td>
</tr>
<tr>
<td>Thymus</td>
<td>25</td>
<td>41</td>
<td>511</td>
<td>388</td>
<td>123</td>
<td>38</td>
<td>180</td>
<td>657</td>
</tr>
<tr>
<td>Vasculature</td>
<td>870</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>869</td>
<td>606</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>1,534</td>
<td>622</td>
<td>2,154</td>
<td>1,492</td>
<td>632</td>
<td>3,393</td>
<td>14,987</td>
<td>3,580</td>
</tr>
</tbody>
</table>


CCF Registration User Interface (RUI)

**New Features:**

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

https://hubmap-ccf-ui.netlify.app/rui/
Register your data via [https://hubmap-ccf-ui.netlify.app/rui/](https://hubmap-ccf-ui.netlify.app/rui/) so it can be spatially/semantically explored in EUI.
VH Massive Open Online Course (VHMOOOC)

Goals

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

Learning modules come with

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

https://expand.iu.edu/browse/sice/cns/courses/hubmap-visible-human-mooc
Thanks go to all the patients that agreed to volunteer healthy tissue and open use of their data.
Upcoming Events & Books
The July/Aug 2022 special issue in *IEEE Computer Graphics and Applications* on “Multi-Level Graph Representations for Big Data in Science”

**Articles due for review:**
**December 29, 2021**

**Guest editors:**
- Katy Börner, Indiana University, Bloomington, US
- Stephen G. Kobourov, University of Arizona, Tucson, US

Indiana University Bloomington will host the International Society of Scientometrics & Informetrics Conference (ISSI) in Summer 2023.
Atlas Trilogy

Atlas of Science
Visualizing What We Know
Katy Börner
2010

Atlas of Knowledge
Anyone Can Map
Katy Börner
2015

Atlas of Forecasts
Modeling and Mapping Desirable Futures
Katy Börner
2021

https://mitpress.mit.edu/books/atlas-forecasts
Jannette L. Finch, MLIS, is a librarian in the College of Charleston Libraries system. Her research interests include information design and the effect of technology on student learning, online learning and teaching, effective teaching through experiential learning activities, constructivist techniques in the teaching and learning environment, visualizing data, library service models, the library role in the scholarly community, assessment, and planning. Contact her at finchj@cofc.edu.