Behind the Scenes:

The Metadata of Data Visualization

Elizabeth "Jo" Klein

Geospatial and Data Visualization Librarian
UNC Greensboro University Libraries

Background & Thesis

Most data visualization processes are made up of four basic and often iterative steps:

1. Data and an appropriate visualization tool or method must be found and obtained,
2. Data must be processed and reorganized to fit the tool,
3. The visualization tool must be used to make a product, and
4. The product must be distributed and made accessible.

Metadata can be leveraged to help at any step in this process, from finding appropriate data and guidance for a tool using keyword searches to making a visualization findable by others once published. However, instruction typically addresses other facets of the process; Data Carpentry workshops offer support for data cleaning, organization, analysis, and topic-specific visualization, while hands-on library workshops instruct users on how to find data and use specific tools. Metadata, an important asset in making data findable, accessible, interoperable, and reusable (FAIR), is often left out of data visualization instruction and is sometimes seen as an optional step in the analysis and visualization process. Metadata can also be a frustrating stumbling block in the research process; in my experience both finding and producing data and data visualizations, I've found that metadata for data visualizations and associated resources are:

- Unpredictable and unstandardized,
- Largely proprietary or tool-dependent (especially in the case of geospatial resources),
- Specialized beyond typical metadata for cataloging purposes, and
- Often incomplete, inadequate, or incorrect.

This renders many data visualizations and the data and other resources needed to make them difficult to find, use, and get support for without going through specific and difficult to determine channels. In effect, they are inaccessible.

Identifying and addressing existing metadata-related challenges, advocating for and applying metadata standards that follow FAIR data principles to data visualization, and creating or supplementing metadata-specific instruction will improve usability of data, supplement scholarly communication of research, and help connect library users with the appropriate support for their data visualization needs.

Professional Context & Approach

As the geospatial and data visualization librarian at UNC Greensboro (UNCG), I provide support for geographic information systems (GIS) and data visualization through:

- Promoting GIS- and data visualization-related library services,
- Providing course and workshop support,
- Consulting with researchers and students on data discovery, use, and management,

- Leading GIS- and data visualization-oriented library professional development, and
- Advocating for openly discoverable, accessible, and reusable data among campus partners.

In carrying out my role, I will be able to directly use the results of this project to support data visualization and its instruction at UNCG. To identify existing metadata challenges and areas of improvement for instruction and resources at UNCG, I plan to utilize the following approach:

1. Survey existing library support and instruction at UNCG and other North Carolina universities (or similar institutions at national or regional level) – especially NC State University Libraries, which hosts the week-long Data Science and Visualization Institute for Librarians.

2. Conduct a literature review and use along with library survey findings to survey researchers and teaching faculty, graduate students, librarians and other learners at UNCG for their experiences finding and using data visualization techniques and associated metadata, as well as their perceived level of support from the library or other sources.

3. Collaborate with the symposia and other groups (including Geoscience Information Society (GSIS) and UNCG University Libraries teams) to identify or define metadata standard(s) related to data visualization and promote the use of these standard(s) at UNCG.

4. Address existing challenges and areas of improvement at the library by creating instruction and/or guides where appropriate and supplementing existing instruction (for example, an infographic on how to leverage metadata and apply metadata standards at every step of the data visualization process).

5. Provide internal instruction, especially for library departmental liaisons, librarians working with digital partners, and other to-be-determined groups.

6. Conduct follow-up surveys of library support and instruction, as well as of learners at UNCG to evaluate implementation of findings.

Impact

Bolstering instruction and support for metadata-related processes of data visualization is the first step towards filling an educational gap. Identifying concrete problems associated with incorrect, incomplete, or otherwise inadequate metadata associated with visualization resources and addressing them through creation and supplementation of instruction for librarians and teaching faculty will provide a leading example of how to improve instruction at other institutions, as well as a framework to better connect library users with data visualization resources and the support they need to utilize them.