AAHSL Data Scholarship Report: Kerry Sewell
July 7, 2018

This letter describes the Data Science & Visualization Institute for Librarians held at North Carolina State University from June 4th-8th, 2018, which I was able to attend through the generous support provided by AAHSL's Data Scholarship.

The Data Science & Visualization Institute for Librarians (DSVIL) was a week-long, intensive training that covered the fundamentals of data, choosing appropriate data visualizations, specific tools and resources for gathering, cleaning, analyzing, and presenting data, and a discussion of data sharing. The workshop involved a lot of hands-on learning so that concepts were applied using the most appropriate tools. To reiterate the content of the previous day, each morning began with a discussion of how the skills from the previous day could be used to support research and researchers at participants' home institutions.

The week included training in:

- Fundamentals of Data, covering the most common challenges with statistics, and methods for determining the appropriate method of analysis. We used Excel and STATA to explore datasets.
- Data Visualization, which covered how well-designed data visualizations serve to help people see patterns and tell stories with data. Excel, Tableau Public, Plot.ly, CartoDB, and Raw were used to create and manipulate data visualizations.
- Web scraping, using webscraper.io to gather data, following a discussion of ethical and legal concerns when scraping data from websites.
- Data Cleaning and preparation, using OpenRefine, followed by using HTML and JSON to reconcile or enrich data. We also gathered Twitter data for network analysis.
- Bibliometric Network Analysis, using Sci2 and Gephi to create network analyses for an individual researcher's records from Web of Science.
- Data Description, Sharing, and Reuse, covering reproducibility, the relationship between reproducibility and data management, funder requirements for data sharing, and issues with data security when sharing datasets. We used the DataCite Metadata Generator and Stata
- Data Use Agreements
- Introduction to Data Visualization with R

The training provided the hard skills needed for engaging in data science, as well as an understanding of what types of data services are offered by libraries. I will be using what I learned to design data services for my library over the next year. In the short term, I will be using bibliometric network analysis skills to demonstrate research impact at ECU. I will also teach some of the skills to my library colleagues and incorporate some of the training into data-related classes offered to health sciences faculty this fall.

I came away from the DSVIL inspired by the role that librarians can have in data science. The training I received through the DSVIL was among the most enriching of my career. I am grateful to AAHSL for the Scholarship that allowed me to attend the DSVIL training. The support from AAHSL to attend this training will have a long-term impact on my professional development, my library, and my institution.

Respectfully,

Kerry Sewell
Laupus Health Sciences Library, East Carolina University
browderk@ecu.edu