Quantitative Data Skills for Undergraduates

A Seminar Series for Social Science Students

Whitney Kramer, Industrial and Labor Relations Research and Data Librarian, Catherwood Library, Cornell University, wbk39@cornell.edu; Amelia Kallaher, Biomedical Data Manager, Sage Bionetworks, amelia.kallaher@sagebase.org

NUTRITION INFORMATION

This recipe is for a series of six 60-minute social science data-focused lessons that provide a broad overview of the social science research life cycle with respect to working with quantitative data. This seminar teaches undergraduate researchers to locate and evaluate quantitative data for a senior honors thesis or other undergraduate research project. It helps fill in the blanks for many undergraduates, who are often taught statistical analysis skills but not the basics of the research and research data management life cycles, including how to find the data that they need to analyze. Teaching these much-needed skills can help fill a void in both undergraduate knowledge and instructional offerings. This recipe can easily be adapted for librarians who would like to teach data literacy concepts, regardless of discipline, through a combination of synchronous and asynchronous methods.

TARGET AUDIENCE AND NUMBER SERVED

Our target audience for this seminar was upper-level undergraduate students who were completing a senior thesis or other in-depth research project. This can be modified based on your instructional needs.

This seminar was originally designed for 10–20 undergraduate participants, but can accommodate over 50 students through asynchronous learning. Teaching the seminar virtually eliminates physical space restrictions and enables the librarian to open up the seminar and associated online content to any interested students.

LEARNING OUTCOMES

Over the course of the six-week seminar, students will

- articulate the research process with emphasis on how to choose a research topic
- recognize sources for locating statistics, spatial data, and other quantitative data formats
- demonstrate the importance of data organization for collaborative group projects
- examine how data can be shared, reused, and replicated
- identify the elements of a data citation
- interpret and create a codebook

COOKING TIME

45–60 minutes per session, plus additional prep and planning time for each session based on the needs of the students, librarians, and optional guest speakers or panelists.

DIETARY GUIDELINES

This curriculum aligns with all six frames of ACRL's *Framework for Information Literacy for Higher Education*, especially Information Creation as a Process (e.g., collaborating around organizing data and how to create codebooks), Information Has Value (e.g., locating and properly citing data sets), and Research as Inquiry (e.g., understanding the iterative process of research and the research data life cycle). This recipe provides instructions for replicating an instructional seminar on data literacy in the social sciences, as well as more generalized instructions for utilizing Zoom and Canvas to construct a librarian-led seminar geared toward whatever data literacy topic you are hoping to address in your instructional session.

INGREDIENTS

- Librarians
- Zoom or other video software (or a space for in-person instruction)
- Access to Canvas or another learning management system
- Guest speakers or panelists (optional)

PREPARATION

1. *Determine the needs of your constituents:* Identify the research methods that are required for undergraduate honors

theses. For example, one-shot instruction sessions may not sufficiently support undergraduates who are looking for librarian assistance for their social science data needs. Creating a longer seminar to address multiple concepts such as how to successfully find and organize quantitative research data can provide more support than a single session. Librarians who wish to teach a similar seminar at their own institution should begin by considering the data literacy needs of their campus constituencies. Would any of the library's typical one-shot sessions benefit from being expanded into a longer seminar? Are you getting similar questions from different student groups that suggest there might be some kind of instructional gap that a library seminar could fill?

2. *Determine your target audience:* Are there particular student groups whose data needs are underserved by their school's curriculum or the library's typical instructional offerings? For example, your target audience of undergraduate students undertaking a senior thesis or other research project may not be enrolled in a for-credit research methods class. We want to target students who are starting from scratch and have not otherwise been exposed to a broad array of social science data topics.
3. *Create course content:* Once you have determined the data literacy needs of your constituents, curate the content of your seminar based on the needs of the groups on your campus. For example, we chose to focus on the theoretical elements of locating and working with social science data within the context of the research life cycle. These are not typically covered in statistics classes or in workshops offered by other groups on campus. Students are introduced to core social science research concepts and receive in-depth instruction on various elements of finding and organizing social science data sets for their research. (See Instructions for sample seminar topics.)
4. *Conduct outreach:* Targeted outreach works best for this type of instruction. Look at the programs or student populations that would benefit the most from the specific content of your seminar series, and focus on reaching out to these specific groups instead of performing blanket outreach. For example, try advertising directly to undergraduate honors programs. You may find that most of your synchronous live engagement will come from students in one specific honors program. They may be using the seminar to fulfill a research requirement for their program in a virtual space.

INSTRUCTIONS

Each session of the seminar consists of a 20-to-30-minute presentation from a librarian on that week's topic and a 20-to-30-minute presentation by a guest speaker or panel, with time for questions at the end. For virtual instruction, we found that it was helpful to have one librarian teaching the class session and one providing technological assistance to participants and monitoring the chat for questions.

For each session, the general outline follows:

1. Welcome students to the seminar and outline the content of that day's lesson.
2. Review contents of previous sessions (if necessary).
3. Teach the lesson through a lecture and accompanying slide deck (or similar).
4. Any optional interactive elements of your presentation can be included here.
5. Introduce guest speakers and have them present (if applicable).
6. Question-and-answer period (students may unmute themselves or interact via the chat).
7. Wrap up—introduce next week's seminar and remind students who to contact if they have any questions.

Sample Topics

We focused on the following six topics for our seminar series, which was loosely structured around both the research life cycle and the research data management life cycle:

1. *Choosing Your Research Topic: How to Plan for a Successful Research Project:* This session focused on the research life cycle concepts of developing a research question and beginning to plan a research project. Because we didn't know how much research experience our attendees had, we wanted to start off with an overview of the research process. Our guest speakers for this session were a panel of three graduate students in the social sciences who spoke about their own research and paths to pursuing a PhD and took questions from our primarily undergraduate audience.

2. *Telling a Story with Data: How to Locate Data for Your Research:* This session focused on locating social science data to help support (or refute) a research question. In this session, we reviewed the different ways one can structure one's research and the different types of social science data, along with common sources for social science data. We also discussed data visualization and introduced students to Tableau. Our guest speaker for this session was the GIS and geospatial data librarian, who presented an overview of geospatial data and showed where students can locate social science–focused geospatial data.
3. *Collaborating with Data: Tools for Data-Focused Group Projects:* This session focused on working with data in groups and the concepts and tools needed for a successful collaboration. We discussed how to use collaboration tools that our students have access to, to collaborate with other researchers, particularly with respect to research data. Our guest speaker was the Research Data Management Services Group (RDMSG) coordinator, who discussed best practices for collaboration and workflow management and provided an introduction to the Open Science Framework (OSF).
4. Organizing Your Data: How to Organize, Share, and Replicate Your Research: We introduced the research data management life cycle, with a focus on data replication and validation and why this is important in social science research. While undergraduates often work with cleaned and packaged data, it is important for them to understand what it takes to get data to that point. Our guest speaker was a research associate from the Cornell Center for Social Sciences, who runs the center's Results Replication service.
5. *Citing Data: How to Cite Data for Your Research Project:* This session delved into data citation. Students are not commonly taught how to cite data, so we devoted a session to the concept. Our guest speaker was the data editor for the American Economic Association, who is also a professor at our institution.
6. Codebooks: How to Find, Read, and Write a Codebook: In our final session, we discussed codebooks. Whether students use already existing data or are performing original research, knowing how to read and write a codebook is an important part of the research process. We discussed best practices for codebook creation. Our guest speaker was the librarian from the Cornell Center for Social Sciences, who discussed her experiences with creating codebooks and led students in a simple exercise of generating a codebook from a provided sample data set.

Post-session Follow-up

After a live session is complete, post the recording of the instruction session and accompanying materials to Canvas or your learning management system so that students who attended the live session can review materials and students who are attending asynchronously have access to the recording and materials as soon as possible. If you are teaching the seminar in person, it is a best practice to post materials as soon as possible for students to access while it is still fresh in their minds.

REVIEWS/ASSESSMENT STRATEGY

For virtual instruction, it can be challenging to assess the seminars only by attendance metrics. There can often be a significant difference between the number of registrants and the amount of actual engagement with the live webinars and the asynchronous materials on Canvas. We found that students engaged more with the materials on Canvas than during the live webinars. However, conducting targeted outreach may encourage students to attend the live seminar to engage with the material and ask questions in real time.

ADAPTING THE RECIPE

The structure of this series can be adapted depending on the library instruction needs of your institution. The combination of live instruction (whether over Zoom or in person) with the asynchronous components in the accompanying Canvas course can be utilized for instructional scenarios that are not suited for a one-shot instruction session or a for-credit course. We highlighted guest speakers or panelists at each session of our workshop instead of hands-on activities. If you do not include guest speakers, you can add in hands-on activities (such as using a particular data tool and software package) or allow more time for discussion. Keep in mind that students who don't attend the live lectures may miss out on participating in the

activities. You may want to consider recording separate demos or linking to self-paced tutorials for students to view the content asynchronously.

ADDITIONAL INFORMATION

For additional information on the original seminar, including course outlines, please contact the authors at wbk39@cornell.edu or amelia.kallaher@sagebase.org.