

Executive Summary

Teaching with Data in the Social Sciences

Wendy Mann, Kim MacVaugh, Jasmine Spitler, and Andrew Lee

Background

The **Teaching with Data in the Social Sciences** project focused on undergraduate courses and is based on the idea that *“instructors in the social sciences need support in locating appropriate datasets and identifying tools to help students manipulate, understand, and visualize data.”*¹ Understanding the importance of teaching with data, and the number of students that the University Libraries support in these classes, George Mason University Libraries signed on with Ithaka S+R (Ithaka) as a partner institution for this study. With guidance from Ithaka, four faculty in the University Libraries, the study authors, interviewed instructors teaching undergraduate, data-intensive courses in the social sciences to identify their needs, challenges, and teaching practices.

Methods

This qualitative study was carried out in tandem with nineteen other research universities in the United States. In addition to George Mason, five of the twenty universities are in the Washington, DC metro area and in Virginia. Seven are recognized as SCHEV approved peer institutions for George Mason University.²

| DC Area & Virginia | SCHEV Peer Institutions |
|---|--|
| <ul style="list-style-type: none">• American University• George Washington University• University of Richmond• Virginia Tech | <ul style="list-style-type: none">• Boston University• Florida State University• George Washington University• University of Massachusetts Amherst• Michigan State University• North Carolina State University• University of North Carolina - Chapel Hill |

Ithaka defines teaching with data as:

- gathering data through social science experiments, surveys, or other research,
- searching for appropriate datasets to address a particular research problem,
- cleaning, analyzing, mining, visualizing, or otherwise manipulating data,
- drawing narratives or conclusions from data,

¹ <https://sr.ithaka.org/blog/announcing-two-new-sr-projects-on-supporting-data-work/>

² <https://research.schev.edu/policytools/peergroups.asp>

- learning to use specific tools, software, or programming languages to work with data.³

Even though we used convenience sampling, we strived to find a mix of participants from a cross-section of undergraduate, social science courses.

- Subject librarians from relevant social science disciplines, instructional services, Digital Scholarship Center (DiSC) staff, and selected undergraduate education coordinators suggested faculty to interview.
- The George Mason University Course Catalog (<https://catalog.gmu.edu>) was mined to identify classes with an emphasis on quantitative data. We then filtered the information based on selected criteria.

All institutions followed the same study timeline. In fall 2020, using questions developed by Ithaka, we carried out fourteen semi-structured interviews with faculty from a cross-section of social sciences disciplines. Four of the classes targeted were designated as Mason Core courses.

| Department | School/College | Number Interviewed |
|----------------------------------|----------------|--------------------|
| Communication | CHSS | 1 |
| Conflict Analysis and Resolution | Carter | 1 |
| Economics | CHSS | 2 |
| Geography | COS | 1 |
| Global and Community Health | CHHS | 1 |
| Global Affairs | CHSS | 1 |
| Government | Schar | 4 |
| Psychology | CHSS | 2 |
| Sociology | CHSS | 1 |
| Total | | 14 |

Themes & Findings

After qualitative coding and analysis of the interview transcripts, four main themes were identified: **teaching**, **learning**, **working with data**, and **challenges**.

Theme One: Teaching

The teaching theme was professor-focused, and included:

- discussion of pedagogical techniques,
- teaching philosophy,
- lesson planning, textbooks, and teaching materials,
- teaching support from graduate teaching assistants (GTAs), and
- professional development or training.

³ Ithaka S+R. Project Scope and Recruiting Guide.

Social sciences professors generally enjoy teaching quantitative data courses to undergraduates, because they can bring their own research interests and data into the classroom to help students see real-world applications of the techniques and skills they teach. Instructors teach the methods and tools that they were taught in graduate school and have used in their own research. They may rely on colleagues with different expertise for teaching data but showed interest in learning new tools and programs they are less familiar with — especially new tools and software they haven't been exposed to before. Graduate teaching assistants (GTAs) also provide a valuable additional teaching resource. In courses where there are separate lab sections, GTAs teach the section independently, filling in gaps from the lecture section and offering extensive software support to students. GTAs are considered “invaluable aids” in these classes.

Theme Two: Learning

Data-centered undergraduate courses tend to attract a wide variety of students with different academic backgrounds and interests regarding data. These differences often make it difficult for professors to meet these varying needs in a single course. Students may express more reluctance to engage with quantitative data in some social science disciplines more so than others. However, professors in these areas believe it is important for students to have a basic understanding, so they can interpret and think critically about quantitative research presented in articles. There may be opportunities to supplement classroom learning with extracurricular activities or independent learning, but most interviewees were not aware of specific options or collaborations. As we were analyzing the interviews, it became clear that lack of students' perspectives is a limitation for this theme.

Theme Three: Working with Data

Many of the social science professors interviewed expressed having a shared pedagogical philosophy regarding teaching data. For projects and practice in their respective courses, approaches vary widely, especially between disciplines. In general, professors of one semester courses often provide datasets or specify a few sources to work from for practical reasons — to alleviate both time-constraints, as well as managing all of the students' data acquisition separately. The instructors make a concerted effort to choose data for their examples and assignments that they believe will be interesting to the students and will illustrate desired statistical concepts for their instruction. Year-long projects and capstones allow for more freedom for the students to generate or acquire data independently.

The choice of software for these data courses is discipline-specific, determined by prevailing norms in their respective industries and/or scholarship, and is also influenced by the instructors' familiarity or previous experience using the programs. Some interviewees desire to move towards free, open-source options for quantitative data analysis, but express reservations about learning a new tool while also being expected to teach it to students. Students, generally, are perceived as coming to the courses with more limited computer skills than instructors anticipated, making it very challenging to meet their numerous learning objectives.

Theme Four: Challenges

Five main challenges identified by these social sciences instructors were:

- the ethical issues inherent in collecting and generating data,

- the difficulty of designing a single course to cover statistical concepts, software skills, and research methods,
- students' math anxiety and psychological barriers,
- students' limited digital and data literacy, and
- general problems with accessing online software programs and resources.

Some of the professors interviewed were largely unaware of the extent of support and services available from George Mason University Libraries. Almost all the professors interviewed indicated that the course design for their research methods classes has inherent problems. The biggest concern is that there is, simply put, not enough time designated in their classes to teach students to work with data adequately. There is a divide between students who are struggling with the content, compared to students who are thriving. Many also said they struggle to find a balance between teaching social science concepts while also teaching how to use different data software programs.

Recommendations

- **Improve outreach and marketing specifically focused on the support the library provides for teaching data. DiSC, data sources, subject librarians, research software — target instructors.**
- **Connect relevant campus partners to improve students' access to software, technology, and data support.**
- **Work with campus stakeholders on data literacy initiatives.**
- **Address the lack of foundational computational skills needed to work with data.**

Improve outreach and marketing specifically focused on the support the library provides for teaching data. DiSC, data sources, subject librarians, research software — target instructors.

Our interviews evidenced a general lack of awareness of the services and support provided by George Mason University Libraries, including the Digital Scholarship Center, datasets and resources, subject librarians, and labs/software for quantitative research. This could be remedied by targeted outreach to instructors of research methods classes, particularly GTAs and professors teaching Mason Core courses related to data and quantitative reasoning.

Possible outreach mechanisms could include:

- direct emails to instructors and graduate assistants identified through the course catalog each term,
- tabling and/or marketing materials specific to “working with data” (i.e., DiSC) at all library orientation events,
- messaging templates for subject librarians to use during library instruction and departmental meetings,
- marketing materials and/or presentations at new faculty and adjunct orientations.

The interviewees indicated that they primarily learn new techniques and teaching practices from their peers informally, making word-of-mouth among instructors a critical component to outreach. Librarians

should encourage professors they have worked with to tell their colleagues about the resources available to them. The Stearns Center for Teaching and Learning, the Provost's Office, and individual departments and schools will also be key allies in communicating the breadth of library services available related to data and research methods.

Connect relevant campus partners to improve students' access to software, technology, and data support.

The University Libraries, Undergraduate Education, schools, colleges and academic units, and Information Technology Services (ITS), need to work together to help connect students taking data focused classes to the variety of support available across the university. We need to work together to create a seamless network which easily connects students to needed software, technology, and data. Not only would this support network aid students, but it would also help instructors teaching data courses. While data support exists in various departments, it would be ideal if it became common knowledge that DiSC and University Libraries is a central resource for help with data.

Work with campus stakeholders on data literacy initiatives.

The George Mason University Libraries focus on data literacy is paralleled by an emphasis on quantitative reasoning in the Mason Core—a university level commitment to data literacy for undergraduate students. Beyond outreach to core course instructors, efforts must be made to engage with other campus offices and organizations to improve students' skills in working with data. The Stearns Center for Teaching and Learning and the Mason Core committee are obvious partners for working on initiatives within the entire undergraduate curriculum, and the Office of Student Scholarship, Creative Activities, and Research (OSCAR) and the Honors College will be important collaborators on promoting data literacy for students with special interest in the research process.

Interviewees indicated that they are largely unaware of the extracurricular activities that students participate in, but student-led organizations that have an academic or technology focus may provide a unique avenue for reaching students struggling with data literacy. In addition, there are some university-wide programs such as Learning Communities (LCs), and departmental projects such as undergraduate research assistantship opportunities that would be optimal for integrating data literacy initiatives. University Career Services is another potential stakeholder to engage with to ensure that George Mason University students are prepared with the data skills needed in their job or internship.

Address the lack of foundational computational skills needed to work with data.

A common thread throughout many of the interviews was instructors' need to teach students basic computer skills. Precious class time (or time outside of class) ends up being dedicated to teaching students the basics of navigating their computers, (i.e., downloading files, locating downloaded files, creating folders, unzipping files and related). The University Libraries' Digital Scholarship Center (DiSC) has had similar experiences when teaching in-class sessions or workshops.

The library can work with academic departments and other units (such as Undergraduate Education and the Stearns Center) to address this lack of skills by developing in-person and online training modules. Further, we can rely on openly available curricula.

- Refer to other universities' experiences deploying the development of digital competencies, such as Bryn Mawr College's Digital Competency Skills
<https://www.brynmawr.edu/digitalcompetencies/what-are-digital-competencies>
- The Foundational Computer Skills training modules in development by The Carpentries'.
<https://carpentries.org/blog/2021/06/incubator-lesson-foundational-computer-skills/>
These lessons are in the pre-alpha stage, but should be useful once completed. The concepts on which they plan on building the curricula are the skills needed for working with the data.

While some of these information technology competencies are addressed by selected Mason Core Information Technology and Quantitative Reasoning courses, it was made clear throughout these interviews, and from DiSC staff experience, that foundational digital competencies need to be developed through other approaches.