TEACHING WITH DATA IN THE SOCIAL SCIENCES AT GEORGE MASON UNIVERSITY

October 1, 2021

George Mason University Libraries

Wendy Mann, Director, Digital Scholarship Center (DiSC), team lead
Kim MacVaugh, Policy and Government Librarian
Jasmine Spitler, Assessment Librarian
Andrew Lee, Social Sciences Librarian

Table of Contents

Study Description and Background

In February 2020, George Mason University Libraries began its participation in Ithaka S+R's Teaching with Data in the Social Sciences project. This project focuses 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] With guidance from Ithaka S+R (Ithaka), participating universities investigated these needs through interviews with undergraduate instructors. This report is the local report specific to George Mason University. A final national capstone report pulling in the findings from all participating institutions will be published by Ithaka in late 2021 or early 2022.

Ithaka works with colleges and universities, libraries, publishers, cultural institutions, and foundations to learn from research studies developed, designed, and led by their researchers. The data collected help institutions to improve performance.[2] In the Teaching with Data in the Social Sciences project, partner institutions carried out research designed by Ithaka. Throughout the entire process, Ithaka led workshops and check-ins, and provided expert advice for the participating universities. A call for partner institutions for this project was announced in summer 2019. 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 as a partner institution.

This study was carried out in tandem with nineteen other research universities in the United States. Five of the twenty universities involved are in the Washington, DC metro area and Virginia (American University, George Washington University, University of Richmond, and Virginia Tech).[3] Seven are recognized as SCHEV approved peer institutions for George Mason University (Boston University, Florida State University, George Washington University, Michigan State University, North Carolina State University, University of Massachusetts Amherst, and University of North Carolina Chapel-Hill).[4]

All institutions followed the same study timeline. In fall 2020, the authors conducted fourteen semi-structured interviews with instructors teaching undergraduate, data-intensive courses in the social sciences at George Mason University. The Methodology Section in Appendix A describes how the instructors were selected for interviews, which departments were targeted, and provides further details about the project timeline.

[1] https://sr.ithaka.org/blog/announcing-two-new-sr-projects-on-supporting-data-work/
[2] https://sr.ithaka.org/about/
[3] https://sr.ithaka.org/blog/launching-two-projects-on-supporting-data-work/
[4] https://research.schev.edu/policytools/peergroups.asp

For the purposes of this study, Ithaka S+R defined the social sciences using the following disciplines:

Anthropology
Archaeology
Area Studies
Communications
Cultural Studies
(e.g., African American, Gender)
Economics
Education
Environmental Science
Geography
History
Law/Criminal Justice
Linguistics
Political Science
Psychology
Public Health/Public Administration/Social Work
Sociology
Urban Studies/Urban Planning

Schools of Business were excluded from this study. Participating universities were instructed to look at the social sciences broadly and to include inter- or multi-disciplinary programs as appropriate.

The authors interviewed teaching faculty from nine different degree programs in the social sciences representing five colleges and schools at George Mason University.

College of Humanities and Social Sciences (CHSS)
Schar School of Policy and Government (Schar)
College of Health and Human Services (CHHS)
College of Science (COS)
Jimmy and Rosalynn Carter School for Peace and Conflict Resolution (Carter)

Five of the data courses that were taught by interviewees are part of the general education requirements for all undergraduates at the university, known as the Mason Core. The government, psychology, economics, and sociology research methods courses meet the requirements for quantitative reasoning, information technology, or capstone.

Degree Programs and Number of Instructors Interviewed

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

*Instructors interviewed teach a Mason Core course that also met the research project's criteria for a data-focused class.

Number of Instructors Interviewed by Rank (n=14)

Rank	Number Interviewed
Assistant Professor	4
Associate Professor	9
University Professor	1

About George Mason University

George Mason University is an R1 public research university with its main campus located in Fairfax, Virginia. The main campus is the principal center for undergraduate studies. The university has three smaller campuses in Arlington, VA, Manassas, VA, and Songdo, South Korea. Since this study addresses teaching undergraduates, most of the departments we targeted are associated with the Fairfax campus.

Fall 2020 enrollment totaled 38,542 degree seeking students, 26,515 of which were degree seeking undergraduates (US Campus Only).[5] 36% of undergraduate degree seeking students are first-generation college students.[6] As of 2020-2021, George Mason offers 78 undergraduate degree programs, 95 master's, 38 doctoral and a Law school (212 degree programs total).[7]

Undergraduate research is promoted and supported by the Office of Student Scholarship, Creative Activities, and Research (OSCAR) "students as scholars" programs. OSCAR is instrumental in connecting undergraduates with research experiences, including the opportunity to work on data-intensive research.

University Libraries

George Mason University Libraries is distributed across the three Virginia campuses. It comprises a large central library, Fenwick Library, at the Fairfax campus, Arlington Campus Library, serving the School of Policy & Government and the School for Peace and Conflict Resolution among other graduate programs, and Mercer Library at the Science and Technology campus in Manassas. Library collections total 1.5 million volumes and consortia memberships include the Washington Research Library Consortium (WRLC), Virtual Library of Virginia (VIVA), Association of Southeastern Research Libraries (ASERL), and LYRASIS.[8]

A library workforce of nearly 140 employees supports the research needs of the university by offering instruction and research guidance, by describing and maintaining physical and digital resources, and through preserving and making available valuable university collections and research output for the long-term.

Within University Libraries, the Digital Scholarship Center (DiSC) and subject librarian teams are the frontline resources for students and faculty seeking data research support. DiSC supports researchers

[5] https://oiep.gmu.edu/resources/fast-facts/mason-facts-and-figures-2020-2021/
[6] https://www2.gmu.edu/about
[7] https://oiep.gmu.edu/resources/fast-facts/mason-facts-and-figures-2020-2021/
[8] https://www2.gmu.edu/academics/university-libraries

working with data, whether quantitative, qualitative, GIS, or digital humanities. Subject librarians specialize in the variety of academic disciplines taught at the university and serve as liaisons to departmental faculty. See Appendix E for more details about library support for teaching with data.

Limitations

No term faculty, adjunct faculty, postdoctoral fellows, visiting professors, or graduate instructors of record were interviewed for this project. Although these instructor roles were acceptable within the project scope, there were difficulties identifying and reaching out to potential interviewee candidates. Of those that were contacted, none responded.

In addition, no students were interviewed, as students were not included within the scope of the study. For the future it may be useful to gather data from their perspectives, especially with respect to self-guided learning.

The COVID-19 global pandemic became a limitation to the project. The research study planning and early stages of the project occurred prior to the onset of the pandemic, and Ithaka S+R continued the project as scheduled, with training and interviews conducted virtually. The authors do not think that COVID-19 had a significant impact on the interview process, however they do not know whether the stress of that time period influenced responsiveness to the recruitment emails. The impact of COVID-19 on classroom instruction and access to technology are discussed throughout the interviews. In general, it seems likely that COVID-19 did have some effect on this project, though the scope is unknown.

Findings

After coding an initial sample of three interviews independently, the research team collectively identified the four major themes of teaching, learning, working with data, and challenges encountered by interviewees. Each researcher then coded all fourteen interviews for their designated theme. The teaching theme was professor-focused, and included discussion of pedagogical techniques, teaching philosophy, lesson planning, textbooks and teaching materials, as well as teaching support from graduate teaching assistants (GTAs), and professional development or training. The learning theme focused on the perspective of students and included learning objectives, the expectations versus reality faced by students, and anything related to student behaviors, interests, or needs. Working with data was a theme that referred to practical issues discussed such as software, technology, computational skills and any descriptions of obtaining, installing, accessing, manipulating, analyzing, or applying datasets and data software. Lastly, interviewees identified a variety of challenges that they encounter in their courses, which were coded together along with any recommendations they provided.

Theme One: Teaching

The professors interviewed for this study teach a wide variety of courses on using data in the social sciences, yet they share a common desire to help students understand how numerical data are applied in context and how to use statistical skills in real-life situations. Social science majors feature discipline-specific courses which are practically oriented and concerned with conducting research according to the standards of their respective fields. Although these quantitative research methods and data courses are

uniquely designed, the collective mantra was “numbers mean nothing, if you can't interpret them and understand what they mean.”

Pedagogy

All instructors interviewed embrace the pedagogical approach of project-based learning. This experimentation, “playing with the data,” was described by numerous professors as a key component of their teaching method. An interviewee explained that students will have a deeper understanding if they feel that they are “driving this car here instead of just being a passenger” and moreover, that working in teams with other students on projects is more beneficial than listening to lectures. Another professor expressed similar sentiments and stated that their goal was to “stimulate curiosity among the students” by using their interests to determine the sources of data and topics for their activities.

Active learning is a well-established principle in teaching, and the interviewees combine this technique with the use of real-world examples and data to offer students a look at how quantitative research may impact their own lives. This is perceived to increase student engagement with the material. The expectation among the professors interviewed is this practical instruction using real-world data in hands-on activities will improve students’ prospects on the job market as they learn desirable skills of their respective fields. Describing their teaching philosophy as “a practical training approach,” a professor explained that they are asking students to “do what professionals do.” Another professor elaborated on the idea they are teaching them skills and the ability to adapt and stay flexible as the job market and situation changes in the future. Perhaps the students will not retain everything from the course, but their professors hope they will be able to use the techniques of working with actual data and topics from the world outside the classroom to improve their career prospects.

Teaching students to think critically about data and to interpret it are high on the list of pedagogical concerns in these courses as well. “I want to get the students to a place where they can read or consume research competently, spot serious flaws with a piece of social science research,” an interviewee explained. Several honed in on the importance of teaching students how to discern when data was collected properly because it is necessary for valid analysis, otherwise they could fall victim to what is described inversely as the “garbage in, garbage out principle.” To put it another way, when secondary data is taken out of context, it’s not very useful—because we can’t change the questions asked after the fact. Further, if poor data collection methods are used to begin with, then the data become invalid or questionable. Another professor echoed the need to teach caution and clarity in conducting data analysis, saying “transparency is important.” Two interviewees stressed in almost the same language that they try to “really hammer home to the students the lesson that correlation is not causation.”

Professors often bring their own research interests and data into the classroom to help students see real world applications of the techniques and skills they teach, as in the words of one interviewee, to “show how the sausage is made.” One professor explained that they teach the way that they conduct their own research, which gives them “a good feel for what the impediments are for understanding certain concepts and statistical methods.” Most interviewees indicated the same, that they teach the quantitative methods that they use in the field themselves which adds authenticity but can also be limiting, if students are interested in different topics or methodologies.

For social science disciplines, one strategy for teaching data to a broad spectrum of students is to put students in contact with faculty whose interest and expertise in various research methods and

quantitative analysis software will best serve their individual projects. In other words, they connect the students to other faculty with similar interests to the students' cases where the instructor may not have those skills. This practice builds departmental relationships and students receive mentorship specific to their needs. This is especially valuable in the capstone research courses and projects required in each major, as a professor noted: "the point of the [undergraduate] thesis is for them to be connected with research community, so I encourage them to branch out if I don't have the expertise that they need."

Graduate Teaching Assistants

Along with leaning on their colleagues' expertise, graduate teaching assistants (GTAs) also provide a valuable additional teaching resource for many of the social science research methods courses at George Mason. In courses where there is a separate lab section, GTAs teach the section independently, filling in gaps from the lecture section and offering extensive software support to students. In courses without a lab component, they offer office hours and homework support and are perceived as more approachable for questions regarding statistical concepts and computer issues. Interviewees across the board acknowledged their GTAs as invaluable aides, describing their function as "the first point person," "the first line of defense," and "an extra set of hands." GTA positions are meant to both share the burden of intensive hands-on work for large classes and to train future professors in the field.

Professional Development

Taking graduate level statistics coursework and serving as GTAs is considered adequate preparation for entering professors to begin teaching undergraduate courses in research methods and data. In the words of one interviewee, "the average person coming out of graduate school with a PhD has probably taken at least a few graduate statistics courses, and that's probably sufficient training to teach an undergraduate psychology statistics course for the most part." Another faculty member mentioned that they give GTAs a wide berth to teach independently for their lab sections because "we're essentially also providing teaching training for TAs." In fact, multiple interviewees mentioned that the syllabi and lessons from the courses they taught in graduate school have formed the basis for the courses they taught on their own and still inform their teaching to this day.

When asked about professional development on teaching with data, most professors interviewed indicated that they primarily rely on their graduate school training plus what they learn on the job. Colleagues and mentors in these departments provide both an informal network and formalized avenue for idea generation and knowledge transfer. One interviewee mentioned that their colleagues have casual "lunch and learn" type gatherings to share their skills and experiences, as well as "water cooler conversations." Peers within a department also directly share their syllabi, course content, textbooks, and assignments with one another to help new instructors find their way in teaching the standard research methods course in their discipline. In some cases, professors make materials available upon request; in others, collaboration is the norm and multiple instructors of a course build out lessons together. Where several professors teach sections of the same course, such as in Communication, they coordinate which textbook they are going to use, and past instructors have contributed to a "huge data dump of files" that instructors share. In some instances, whole courses are passed down from professor to professor, ready to go when a new person is brought in to teach the class.

While many interviewees offered examples of learning from and sharing with George Mason peers, almost no one mentioned taking any external classes, workshops, or other development opportunities to strengthen their teaching with data. The exception was an Economics professor, who noted that they

occasionally attend online workshops or check out internet resources by academics they are familiar with. Upon being asked, however, several instructors felt prompted to venture that professional development might be useful for this purpose. One professor summed up the difficulty with relying solely on one's own previous teaching experience in this way: "Usually when you're going to go teach a stats class, I think most people tend to teach it the way it was taught to them. And while that may be great, and that might work out well...at some point, that gets dated." Across the social sciences, the most frequently discussed professional development interest was to learn new data analysis programs to be able to teach students the current software in their fields.

Theme Two: Learning

Instructors who teach quantitative research courses at the undergraduate level expect their courses to ground students in the foundational set of technical and analytical skills that form the methodology of their discipline. They want students to learn what a quantitative study is, how it is conducted, and how data should be collected, analyzed and interpreted validly, accurately, and ethically. Yet, instructor success in imparting these fundamentals is dependent on students' readiness for learning, including a basic comfort level with statistics and computers.

Easing Into Learning

Many instructors interviewed teach 200- or 300-level quantitative research methods courses, which may be the sole methods course required for undergraduate students in their majors. Such courses are usually introductory in nature and do not anticipate earlier exposure to quantitative research. When asked what prior knowledge of working with data is expected of students, a Psychology professor said: "Basically nothing. I mean, to be able to use your computer and understand how a computer works, but that's nothing in this day. Pretty basic." Nevertheless, inexperienced students may not come to class with even that much awareness of how datasets are stored on computers, how to load data into software, or how to navigate data resources.

Instructors take a pragmatic approach, emphasizing their intention to expose students to the basics of quantitative research rather than expecting mastery of the tools or software programs so that all students can succeed in learning the main principles. Learning can be delayed from the start because some students are intimidated by the very nature of quantitative methods, often called "math anxiety." Professors respond by promising minimal mathematical requirements and design their courses with scaffolding and techniques intended to help students get over their psychological barrier. One Psychology professor explained that contextualizing the data within the familiarity of the discipline is a useful practice:

> "Some students just don't want to have anything to do with numbers at all. They're just intimidated. But I do find that in this class where we teach them the statistics in, in a context that has meaning to them and that is trying to answer psychology questions. I do think that helps and it removes some of the anxiety or at least lessen some of the anxiety that many of the students who have it will experience."

The professor also indicated that students may suffer from anxiety at the beginning of the course but they usually get past that at some point during the semester and start acting more comfortable as the

course progresses. Instructors also encourage students to form study groups or assign group projects so that they can learn from one another and alleviate apprehensiveness.

Engaging Reluctant and Motivated Learners

It should be noted that within the same class there are substantial discrepancies in students' interest, motivation, knowledge, and skill levels. Some students may be less motivated because of psychological barriers or lack of basic background knowledge, but some others show strong interest in quantitative methodology and welcome challenges—readily attempting hands-on experiments and self-directed projects. Instructors have to customize their instruction to meet the learning needs of students entering the classroom with different levels of knowledge, skills, and expectations. As a professor of public policy explained, "it's this careful balance of making sure that the basics are covered so that the students who are just trying to get the basics can get there, but then providing enough advanced materials so that the more advanced students don't get bored."

Skills-based Learning

Learning the particulars of the software used in statistical analysis of social science research is a necessity and in fact an objective in its own right, whether or not a student anticipates using the tools again in the future. One professor noted that it's "very important" for students who want to pursue graduate studies to develop their skills in using the dominant software of the field. But even if the student does not expect to go on to graduate school, technical skills in using tools like SPSS can be "very useful" in the job market. Some instructors encourage students to work with data and software outside of formal coursework if students have the drive for more challenging data analysis or if they want to develop their software skills. Some instructors mentioned that online videos, like YouTube tutorials, can be exceptionally helpful to students for a different perspective on content covered in class. Other instructors, however, are ambivalent about promoting such independent learning because it may conflict with the way they have taught the subject or statistical software and cause confusion for students.

Limitations

The challenge of teaching statistics, software, and research methods within the time constraint of a single semester-long course leads to instructors setting parameters which can impact student learning. Students are asked to complete small, manageable projects in the semester and learn how to enter variables or create a dataset on very small scale. Faculty usually decide to provide students with datasets, generally limited in size and scope, so that students can more easily practice opening them into certain software to analyze them. It can be difficult to allow students to find or collect their own data which would perhaps be ideal for learning. For example, it takes time to get IRB approval; if students' projects require IRB approval they may have to avoid such projects.

Even in upper-level or capstone courses, there are limits to what an undergraduate course can offer in terms of data collection and analysis. Students may overestimate the availability of certain kinds of data, and may not understand the economic and cultural limitations inherent in collecting sensitive or international data. For example, in a Global Affairs honors course, a student attempted to research women's rights issues in a foreign country, but she couldn't find the data she wanted. When students are not able to find immediate answers to their questions, they may become disappointed and lose

confidence or motivation to learn; instructors usually have to provide more guidance and direction when such issues arise. Interestingly, some instructors expressed that they view students' disappointment positively because it is a good opportunity for students to learn about the limitations as well as the extent of what data analysis can do and how hard it is to conduct quantitative research.

Ethical lessons

Not only do students learn how to generate, analyze and interpret data, they also need to make the connection between data and the social world—they need to go beyond numbers and formulas to grasp the meaning of the data in a real-world context. To teach students how data can be manipulated misleadingly, faculty use real-life data sources, for example, teaching students how to read graphs from the news to see how data can be interpreted and/or misrepresented in a variety of ways. Students also learn that it is not acceptable to fabricate data or alter calculations until they get the result they are looking for or expecting. As a Psychology professor said, "it's important to be transparent when we're writing up our data so people can understand the decisions," and for students, "understanding what goes behind the scenes, what gets put on paper, is something that they need to understand." Professors emphasize that learning the ethical considerations that underlie social science quantitative research is just as critical for the students as the nuts and bolts of statistical data analysis.

Theme Three: Working with Data

Data Acquisition

While the social science professors involved in the study share much of the same pedagogical philosophy of teaching data, when it comes to the actual projects and practice in these courses, approaches vary widely. They may differ by timing concerns, by instructional design, or by the discipline itself. The Psychology professors interviewed, for example, provide data directly to students, typically from their textbook or datasets handed down by course instructors through the years in order to streamline the classes. The data courses for undergrads in Psychology focus on teaching statistics and it is important for them to provide data that applies to the statistical topic or method being taught each class. Similarly, Government professors are concerned with illustrating different types of variables and scenarios through preselected datasets. In many of the classes described by interviewees, the instructors both provide some data and allow students to find their own data in certain cases. This approach to data acquisition varies depending on the length of assignment, with GTAs and professors providing data for homework and lab assignments, and students gathering data on their own for semester-or-longer projects. For instance, in honors classes, where long-term projects might occur over a year timeframe rather than a single semester, students usually are responsible for finding or gathering their own data.

In addition to time constraints, many professors indicated that a major reason that they provide data is to make instruction more straightforward and the students' lives easier for their projects. One Government professor explained that, "I've tried it all of those ways [search, collect, provide data] and with undergraduates…my preference is for me to find data that I know are clean data and ready to be analyzed." The students are already facing hurdles in learning statistics and working with data. Adding data discovery makes working on the final projects more challenging and more time-consuming. So, providing one single dataset to work from, such as the General Social Survey (GSS) or the World Values Survey, and allowing students to select variables that interest them, or providing a few curated datasets

to choose from, alleviates this problem and allows the students to focus on applying what they had learned. The data supplied by professors tends to be "go-to" secondary datasets, such as GSS, World Bank data, or other popular surveys, or data from an instructor's own research.

Instructors try to take students' interests into consideration and provide data that they think the students would find appealing and want to explore. As one professor put it, "one of my main goals is to stimulate curiosity among the students. So I take what their pre-existing interests are….I help them feel through the detective work that is so common among professional researchers and finding some data that would be relevant to the question." On the other hand, it can be tricky to deal with data for social science questions that could be personally or ethically sensitive. For example, a Global and Community Health professor noted the difficulty of choosing appropriate data on current issues that might affect their classes, like obesity or racial inequality of health outcomes, saying "I try to navigate the waters of personal feeling and human health challenges."

Instances where students are permitted to find their own data depends largely on the discipline and its research methods. For Sociology courses, the Inter-university Consortium for Political and Social Research (ICPSR) is often recommended as a data source for students. For students who want to collect their own data, they are typically advised to limit their survey scope to a handful of variables with about 50 respondents. Likewise in Economics, secondary and replication data are usually preferred; if honors students need to collect data themselves, they are encouraged to work with faculty who specialize in survey research. For the Global Affairs capstone, a little more than half of the students generate data for their research from conducting interviews or surveys. Professors noted that students tend to overestimate the availability and ease of acquiring of data for their projects, whether looking for their own secondary data or collecting data directly. International data for example, is notoriously difficult to acquire reliably for comparative purposes. Project ideas that may require IRB review can be an issue with respect to the limited timeframe of a semester-based class. Geography is unique among the social science disciplines in its requirement for both quantitative and geospatial data, but the professor interviewed also uses a combination of providing data and allowing students to search for data on their own. Several of the instructors interviewed were aware that the Digital Scholarship Center (DiSC) of George Mason University Libraries, formerly known as Data Services, provides assistance with: searching for data for projects, accessing and using software, and working with data in general. Some instructors were recruited for interviews because they teach courses that DiSC has worked with over the years. However, some of the instructors were not aware that DiSC exists and that it can provide this type of support. For some of the classes, the use of the expertise provided by DiSC wasn't necessary, e.g., Psychology.

Software and Technology

The most popular software programs used in the quantitative data courses are Excel, SPSS, Stata, and to a lesser degree, R. For Geography, which has distinct software needs from the social sciences, ArcGIS is the main program taught. In addition, Geography uses the widest variety of software depending on the students' project needs and includes other tools such as QGIS, Google Sketchup, ArcGIS StoryMaps, MapTiler, and MultiSpec. Each discipline tends to use certain software packages, because those are the preferred packages of their fields and interviewees explained that it is important for students to have exposure to these programs to prepare them for future careers. Some professors interviewed noted that they would like to move away from software that requires paid licensing, such as SPSS, and toward something open-source and free like R. For some of the social sciences, such as Sociology or Global

Affairs, students are given the option to take a qualitative or mixed methods approach to their research and in those cases, software such as NVivo or MAXQDA are used. Additional software programs mentioned include Qualtrics (which is licensed at the university level) for conducting surveys, UCINET for conducting network analysis, and alternative spreadsheet programs such as Google Sheets or Apple Numbers. For those instructors who have used the Virtual Computing Lab (VCL) for course access to software packages, this was a major pain point. The older version of the VCL put students at a disadvantage, especially students who used Macs since the VCL was not optimized for Apple products. In Fall 2020, Mason transitioned to a new VCL platform with more seamless access than the earlier version; it is now fully functional with Mac products. Most of the instructors were aware of the transition, but others were not. In addition, not all research software needed to teach data courses is available on the VCL (e.g., NVivo is not, and it was mentioned by interviewees). One instructor taught with MAXQDA because they were able to get a teaching license and the students could use it for free for the semester. Even though this project focuses on quantitative data, there are a few undergraduate social science classes where both quantitative and qualitative methods are taught in the same course.

While access to proprietary software such as SPSS and Stata is alleviated by the VCL, several faculty interviewed expressed concern about relying on licensed software. A professor in the Sociology department explained that they wanted to get away from teaching using SPSS, a costly program that has fewer applications outside the classroom, towards using the open-source R coding language, but they did not know how to accomplish that. In their own words, "I have not had time and I don't have a sense of where I would go to best learn." An instructor in Global Affairs mentioned that data visualization tools are becoming the norm, and said "I personally, I would love to teach them how to use Tableau or something equivalent, but I'd have to teach myself how to use it first." Perhaps the most urgent example is from the geography department, where an interviewee reported that most of the faculty are continuing to teach with ArcGIS Desktop, that they were trained in, and have not yet transitioned to ArcGIS Pro. The professor acknowledged:

> "At some point we're going have to do that and it becomes a pretty heavy lift for a faculty member to learn this new software and then also update every single lab that they have, and every single homework assignment, and all that. It's like teaching the class brand new again while also teaching it to yourself."

Major Software Used by Department

Software	Departments
ArcGIS	Geography
Excel	Geography, Global and Community Health, Global Affairs, Economics, Government, Sociology
R/R Studio	Economics, Government, Psychology
SPSS	Communication, Global and Community Health, Psychology, Sociology
Stata	Economics, Government

The table above summarizes the major software packages used, as discussed in the interviews, by department. Note that the software discussed during the interviews applies to the individual interviewed and what they use when teaching undergraduate data classes.

Computer Skills

Fundamental skills gaps exist in what students know and understand about using software on their computers to analyze data. One professor noted that ten years ago, they would expect the students to know how computers work and how file directories are structured, but says, "I expect nothing these days. I found out the hard way that the students have very little sense of how information is stored on computers and how to navigate their own computer besides point and click." Other interviewees made similar remarks, and indicated they have to start instruction at the very basic steps of opening files and saving files to set locations before students can engage with the data files. This creates a "hard balancing act between teaching data and statistical concepts and software," in the words of one professor. This balancing act occurs because it takes additional time to establish comfort and familiarity with the students' devices and takes away time from the lessons they had planned. The technological gaps can prevent students and instructors from exploring more advanced tools that they may have interest in.

Data visualization was broached in two interviews. In a Global Affairs course for example, students expressed interest but realized they didn't have the time or ability to learn software for this purpose. The instructor noted that students realize their own limitations with respect to technology, learning software, and working with data, and that they tend to choose the path of least resistance. The same instructor mentioned that a few students did create visualizations using GIS software because they had previous experience from another class. Geography students, on the other hand, spend more time throughout their major using visualization tools. Using social media data was mentioned by a Communication professor as of great interest to students, but they noted that this advanced research work would require too much for an undergraduate course.

Data Literacy and Quantitative Work

While all the interviewees teach quantitative research methods courses in the social sciences, the courses in the disciplines of Geography, Economics, Psychology, and Government place the heaviest emphasis on students applying statistical concepts to analyze and interpret data themselves. In addition, Economics and Government both offer an undergraduate minor in data analysis for students who want to pursue their interest further. Alternatively, majors like Global Affairs, Communication, and Sociology prioritize basic "quantitative literacy" as one professor called it. For these courses, students often express reluctance to engage with quantitative data, but the professors believe it's important for students to at least be able to interpret quantitative research when they read it in articles, and to think critically about data when it is presented. Research methods is a required course for Communication so that students develop the ability to analyze communication, propose research questions, and collect empirical data.

Theme Four: Challenges

Five clear challenges were highlighted over the course of the interviews. These include: 1) ethical issues and collecting and/or generating datasets; 2) class design; 3) math anxiety and psychological barriers; 4) digital and data literacy; 5) COVID complications involving accessibility and lack of support for faculty.

Ethical Concerns

There were numerous mentions of ethical issues and collecting and/or generating data in the interviews. Faculty mentioned a few ethical topics they seek to convey to classes including: the importance of

critical thinking while examining data, identifying misleading data reporting, examining individual biases before collecting data and reporting it, being transparent in data collection and reporting, avoiding taking data out of context, as well not changing variables in the data to get needed results.

Although they endeavor to teach ethical practices within their data courses, most social science professors reported that they prefer to provide datasets for their students, rather than allow students to collect their own data, because of the ethical challenges. The Institutional Review Board (IRB) training and application process is perceived as too onerous for undergraduates. As one professor said, “Rarely do I have them collect data because that just becomes too problematic…if they're going to actually do something that will add to the literature or that they want to write it up or publish or something like that, you have to go through all the IRB things which can take a long time.” Even if their project is IRB approved, potential ethical issues or accidents could still occur, if students do not follow the protocols carefully, or don’t have the resources to maintain confidentiality. One interviewee stated: “If people are collecting their own data, then you have to be a little worried about where they're getting the data and what they're going to do with the data.” The university’s Honors program requires independent research and thus must engage with the ethical considerations of data collection more deeply.

Class Design

Almost all the professors interviewed indicated that the class design for their research methods courses has inherent problems. The biggest concern is that there is, simply put, not enough time designated in their courses to teach students to work with data adequately. There are various underlying reasons for this described in the interviews. First, there is a divide between students who are struggling with the content, compared to students who are thriving. One professor said, “In terms of their prior knowledge or preparation and that makes designing the classes a little bit challenging because you want to push those that, you know, can perform at a higher level, but you don't want to push too much because you sort of miss the ones who are perhaps struggling.” Another said, “What we're teaching is kind of so fundamental and basic and some of the students, they really grasp it and move forward and don't seem to be challenged, where some of the students really struggle.”

Professors responded that it is difficult to juggle these two groups, both of which require extra attention. Students who are struggling really need one-on-one support, while those who are doing well with the material also need encouragement to go above and beyond. These courses are usually required for their major, which contributes to the vast discrepancy in student interest and ability to master the material. Professors are keenly aware of the emotional toll this can bring: “I have to be sure that those students that are math phobic and who are struggling…don't fall way behind because the datasets are just too overwhelming and intimidating. And they need it for their major so…they're all stressed out.”

Professors critiqued other aspects regarding class design. For instance, many feel that there aren’t enough prerequisites before their course, so they are burdened by having to review the basics of math, computer skills, and data instruction, on top of their already heavy syllabi. One professor said: “I often can't assume that my students come in with stats skills or programming skills or research skills and if I give them a project in which they really need those things in order to do the project, I wind up spending so much time handholding through it that it becomes a frustrating experience.”

Many also said they struggle to find a balance between teaching social science concepts while also teaching how to use different data software programs. Both are necessary, but time is limited for providing quality instruction in each area and they feel rushed. In the words of one professor, it’s a

“hard balancing act” where “we have to teach these concepts, but also you do end up having to teach the software which…I think it can be difficult for some of the undergrads learning a software they've never used.”

Math Anxiety

Professors pinpointed one of the major difficulties students face when taking a teaching with data class—math anxiety. As one professor explained, “There will be students who have never used a programming language in their life and have avoided every math class that they possibly could, and you know have really very little experience and frankly not a great attitude about gaining experience.” Another said, “I've heard from many students over the years that they that were some version of ‘I’m a Comm major and I don't do math,’ or ‘I’m a Comm major, that means I don't do math,’ and they kind of cringe and smirk all at the same time.” Professors reported having to repeatedly reassure students that they can manage the work, which takes time away from their planned curriculum, with one saying, “a lot of it is sort of assuaging the worries of the students and letting them understand that statistics—it's not the same thing as math and the amount of math that goes on really isn't that difficult… A lot of times their math knowledge is very, very limited.”

Digital Literacy

Although students may use computers regularly, professors have come to realize that students may not know how to find information on their computers. As discussed in the Working with Data section above, students “have very little sense of how information is stored on computers and how to navigate their own computer.” An interviewee said, “I think the number one issue, perhaps exclusive issue, is this a shift of technology awareness,” elaborating:

> “The biggest challenge has been honestly that a lot of them don't understand that there is a folder structure behind the computer where files are stored. Another problem is that they don't understand that sometimes you can click on a file on a website like Blackboard and sometimes you don't want to open the file or display it in the browser, you want…to download it and save it somewhere, and a lot of them have absolutely no clue about that, which was really very surprising to me.”

Once students have moved on to learning the software they will be using in the class, some may still not understand how to use data files with the program— “even doing something as simple as loading the data into the software” can create difficulties according to one professor.

COVID Complications and Online Accessibility

The impact of the COVID-19 pandemic has been detrimental for data and research methods instruction, particularly as courses have moved to online modalities. Some professors felt that the online experience was not as beneficial as in-person learning and can be particularly hard on struggling students. One professor said, “In all honesty it's very disconnected with students in the class, with it being completely online.” Another said, “I think online is perfectly suitable, as a vehicle for delivering coursework, but I think it increases a disparity between strong and weak students.” As mentioned in the Working with Data section, professors relied on the Virtual Computing Lab to access expensive software programs like SPSS and STATA that students can’t afford to buy for themselves (even the discounted student versions), but its performance was inconsistent. As one professor put it, “if that online lab isn't working, that puts students at a dramatic disadvantage…that has got to be rock solid in order to foster a true online degree program for using data.”

Conclusions

George Mason University social sciences professors enjoy teaching quantitative data courses to undergraduates, because they are able to bring their own research interests and data into the classroom to help students see real-world applications of the techniques and skills they teach. Interviewees expressed the hope that active, problem-based learning will help students in their future academic pursuits and their diverse career paths, as the ability to understand and interpret quantitative data is an important life skill. Instructors teach the methodologies and tools that they have been taught in graduate school and have used in their own research, and rely on peer faculty with different methods expertise if students need specialized help in areas they are not as familiar with. They do not usually seek out professional development on teaching data but would be interested, especially in learning new tools and programs they haven't been exposed to before.

The wide variety of students in the data-centered courses with different learning styles and backgrounds makes it more difficult to meet everyone's needs in one course. Some students experience significant math anxiety and fear regarding work with quantitative data and software programs. Others excel and can learn the content much more quickly than their peers, requiring extra challenges to keep them interested. 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.

While the social science professors involved in the study share much of the same pedagogical philosophy of teaching data, when it comes to the actual projects and practice in these 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 time-constraints and the difficulty of managing all 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 also 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.

The 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. George Mason University Libraries offers services and resources that can address many of these concerns, however some the professors interviewed were largely unaware of the extent of support available.

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.

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, Information Technology Services (ITS), schools, colleges and academic units, and student services need to work together to help connect students taking data focused classes to the variety of support available across the university. We must collaborate to create a seamless network which easily connects students to needed software, technology, and data. Not only will this support network aid students, but it will also help instructors teaching data courses. While data support exists in various departments, it would be ideal if it became common knowledge that DiSC 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 Libraries should work with academic departments and other units (such as the Office of Undergraduate Education and the Stearns Center) to address this lack of skills by developing in-person and on-line training modules. Further, we can rely on openly available curricula. For instance,

- 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/ should be useful when they are completed.

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.

Appendices

Appendix A: Methodology

The authors received IRB approval for this project in June 2020 [IRB Number 1614567-1]. In July and August, we worked on researching undergraduate courses in the social sciences that met the criteria laid out by Ithaka. Ithaka defined 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
- Learning to use specific tools, software, or programming languages to work with data.[9]

To help identify classes and participants, we also solicited suggestions from subject librarians and Digital Scholarship Center staff. Two members of this project group are subject librarians representing social science disciplines and one member is a staff member of the Digital Scholarship Center (the bulk of their work is helping people working with data). We also talked to two faculty who are involved with undergraduate education at the university and school level to discuss this project, and to ask for feedback and suggestions for people to contact.

Interview Selection (Summer 2020)

We selected interviewees the following ways:

- Contacted subject librarians from relevant social science disciplines, instructional services, and DiSC staff asking them for suggestions.
- Mined the George Mason University Course Catalog (https://catalog.gmu.edu), filtering by undergraduate courses and keywords (data, statistics, research methods, quantitative) to identify classes with an emphasis on data.
- Reviewed the filtered course descriptions and consulted course syllabi (if available) for classes where we needed more detail on what was being taught.
- In the cases of suggested contacts that we received, we also reviewed the course descriptions for the courses they taught to ensure that the classes met our criteria.
- After we developed a list of classes, we identified instructors by using a combination of departmental websites, the course catalog, and PatriotWeb (George Mason's course registration system). When searching PatriotWeb, we identified instructors for courses taught during the Fall 2020, Summer 2020, Spring 2020, and Fall 2019 semesters.
- Developed a spreadsheet where we ranked courses by how well they matched Ithaka's criteria. Many of the courses had multiple sections and instructors. We decided on who to contact first based on feedback from library staff and suggestions from interviews with undergraduate education coordinators.
- Ithaka's goal was for each site to interview fifteen instructors. They indicated that we needed to send emails starting with fifteen and working from there. We strived to find a mix of participants

[9] Ithaka S+R. Project Scope and Recruiting Guide.

from a cross-section of social science courses. However, Ithaka indicated that in a study of this type, it would be difficult to put together a diverse group of interviewees.

Interview Process (Fall 2020)

We sent out the first call for participants on September 8 to the top fifteen/first round of contacts on our list. Personalized reminder messages were sent on September 17. The initial emails were "b'cc'd" to the group and only four people responded, with one declining. We decided to send the reminder messages to each individual potential participant rather than a generic email. We received a much better response to the personalized reminder messages (six out of eleven). Overall, nine people agreed to be interviewed, with a 60% response rate for the first round of contacts.

We proceeded to schedule these interviews, providing the interviewees with the Informed Consent document for them to read through before the interview date. We had the entire Fall semester to conduct the interviews.

On October 13 we sent the second round of emails to the second group of contacts. We proceeded to send out emails to each individual on the list since that worked well. We only needed to solicit six more interviews to reach our goal of fifteen. Note that as we moved through our contacts, we removed anyone who did not respond to initial and reminder emails from each round from the list. In the second round, five out of the eight people contacted agreed to be interviewed (62.5% response rate).

We were instructed by Ithaka that fourteen interviews were a good place to stop and that we no longer needed to reach out to any additional faculty. **Our response rate was 60.87%** (fourteen out of the twenty-three instructors contacted agreed to be interviewed).

Instructors Contacted by Department

Department	Number Contacted	Responded Yes
Sociology/Anthropology	2	1
Global Affairs	1	1
Communication	1	1
Conflict Analysis & Resolution	1	1
Criminology, Law & Society	1	0
Economics	2	2
Global & Community Health	2	1
Geography & Geoinformation Science	4	1
Government & Policy	5	4
Psychology	3	2
Totals	**23**	**14**

Because of the COVID-19 pandemic, all interviews were held virtually via Zoom. Instructors who agreed to participate were sent the Informed Consent document prior to the interview day. At the time of the interview, participants were then asked to verbally consent and asked if they agreed to be audio recorded. No one declined. The audio recording and transcript were saved for each interview. When

they were ready, the recordings and transcripts were immediately downloaded and deleted from Zoom storage and saved to a secure, shared folder on a server managed by the university (MESA).

Since these were semi-structured interviews, the interview questions (Appendix C) were used as a guideline. The major sections and points of the interview questions were addressed at each interview.

Transcription (January/February 2021)

In January and February 2021 each interviewer verified the auto-transcription accuracy and corrected the transcripts by reading through them and listening to the audio recordings. Once the transcripts were corrected, the audio recordings were deleted. No audio recordings were saved. The file names used for the transcripts were anonymized and any identifying information was removed from the transcripts. In January 2021, generalized metadata containing information about the interviewees' department and faculty rank were sent to Ithaka. We sent copies of the anonymized final transcripts to Ithaka at the end of February.

Qualitative Coding and Analysis (March/April/May 2021)

Each local research group member was instructed to code the same three interviews. Once that was completed, we compared codes that each group member used to identify four major themes and to develop our coding structure. Once the four themes were identified and defined, each group member selected a theme to code and then coded all of the interviews based on their selected theme. Next, each group member synthesized the data and wrote the analysis for their respective theme. NVivo was used for the coding process.

Appendix B: Recruitment Emails

Recruitment Email

Subject: George Mason University's study on teaching with data in the social sciences

Dear [*first name of instructor*],

George Mason University Libraries is conducting a study on the practices of social science instructors in order to improve support services for their work (IRB number 1614567-1). We are interviewing instructors whose undergraduate students engage with quantitative data, such as by conducting research using quantitative methods, analyzing or visualizing datasets, or learning to use specific tools or software to work with data. Would you be willing to participate in a one-hour interview to share your unique experiences and perspective?

Our local George Mason University study is part of a suite of parallel studies at 19 other institutions of higher education in the US, coordinated by Ithaka S+R, a not-for-profit research and consulting service. The information gathered at George Mason will also be included in a landmark capstone report by Ithaka S+R and will be essential for George Mason to further understand how the support needs of social science instructors are evolving more broadly.

If you have any questions about the study, please don't hesitate to reach out. Thank you so much for your consideration.

Sincerely,

[*name of investigator listed on this protocol*]

Recruitment Follow-Up Email

Dear [*first name of instructor*],

Thank you for expressing your interest in participating in this study (IRB number 1614567-1). I would love to set up a time to interview you at your convenience. Please advise me of your availability in [*time frame*].

Before the interview begins, I will ask you to please review the Informed Consent. At the beginning of the interview, you can provide verbal consent in order to ensure that you understand the study and are willing to participate in it. I am attaching the Informed Consent document to this email so that you can review it before the interview.

Please feel free to reach out with any questions or concerns.

Sincerely,

[*name of investigator listed on this protocol*]

Appendix C: Informed Consent

Teaching with Data in the Social Sciences

INFORMED CONSENT FORM

RESEARCH PROCEDURES
This research is being conducted to examine social science instructors' practices in teaching undergraduates with data in order to understand the resources and services that instructors at George Mason University need to be successful in their work. This research is being funded internally by the University Libraries. If you agree to participate, the study involves a 60-minute, audio-recorded interview about research practices, instruction and teaching with data. The interviews will take place using a private WebEx or Zoom connection. If you consent to be audio-recorded, the audio recording will start at the beginning of the interview session and will be ended once the interview is completed. While it is understood that no computer transmission can be perfectly secure, reasonable efforts will be made to protect the confidentiality of your transmission. Your participation in all or part of this study is completely voluntary. You are free to withdraw consent and discontinue participation in the interview at any time for any reason.

RISKS
There are no foreseeable risks for participating in this research study.

BENEFITS
The benefits to you may include increased insight and awareness into research and instructional practices as well as data support.

CONFIDENTIALITY
The data in this study will be confidential.

The audio recordings will be deleted from WebEx or Zoom and stored on Mason Enterprise Services Architecture (MESA), a password protected, secure shared drive. The project folder will only be accessible by the co-investigators. Digital audio file names will be coded and will not include your name or other personal identifiers. Pseudonyms will be immediately applied to the interview transcripts and the metadata associated with the transcripts. Once transcription is complete, the digital audio files will be deleted from MESA.

Participants may review Webex's website for information about their privacy statement https://www.cisco.com/c/en/us/about/legal/privacy-full.html. Participants may review Zoom's website for information about their privacy statement https://zoom.us/privacy.

The de-identified and aggregated data could be used for future research without additional consent from participants.

The Institutional Review Board (IRB) committee that monitors research on human subjects may inspect study records during internal auditing procedures and are required to keep all information confidential.

PARTICIPATION
Your participation is voluntary, and you may withdraw from the study at any time and for any reason. If you decide not to participate or if you withdraw from the study, there is no penalty or loss of benefits to which you are otherwise entitled. There are no costs to you or any other party.

CONTACT
This research is being conducted by Wendy Mann, Kimberly MacVaugh, Andrew Lee, and Jasmine Spitler from University Libraries at George Mason University. Wendy Mann may be reached at 703-993-5272 or wmann@gmu.edu for questions or to report a research-related problem. You may contact the George Mason University Institutional Review Board office at 703-993-4121 or IRB@gmu.edu if you have questions or comments regarding your rights as a participant in the research (IRB number 1614567-1).

This research has been reviewed according to George Mason University procedures governing your participation in this research.

CONSENT
I have read this form, all of my questions have been answered by the research staff, and I agree to participate in this study.

If you agree to participate in this study you will be asked before the interview begins if you agree or do not agree to audio recording.

Semi-Structured Interview Guide

Before starting the recording, verify that the interviewee agrees to being recorded. Point out that only the audio will be recorded.

Do you agree to the Informed Consent?

Note regarding COVID-19 disruption I want to start by acknowledging that teaching and learning has been significantly disrupted in the past year due to the coronavirus pandemic. For any of the questions I'm about to ask, please feel free to answer with reference to your normal teaching practices, your teaching practices as adapted for the crisis situation, or both.

Background

Briefly describe your experience teaching undergraduates.

- How does your teaching relate to your current or past research?
- In which of the courses that you teach do students work with data?

Getting Data

In your course(s), do your students collect or generate datasets, search for and select pre-existing datasets to work with, or work with datasets that you provide to them?

If students collect or generate datasets themselves Describe the process students go through to collect or generate datasets in your course(s).

- Do you face any challenges relating to students' abilities to find or create datasets?

If students search for pre-existing datasets themselves Describe the process students go through to locate and select datasets.

- Do you provide instruction to students in how to find and/or select appropriate datasets to work with?
- Do you face any challenges relating to students' abilities to find and/or select appropriate datasets?

If students work with datasets the instructor provides Describe the process students go through to access the datasets you provide. *Examples: link through Blackboard, instructions for downloading from database*

- How do you find and obtain datasets to use in teaching?
- Do you face any challenges in finding or obtaining datasets for teaching?

Working with Data

How do students manipulate, analyze, or interpret data in your course(s)?

- What tools or software do your students use? *Examples: Excel, online platforms, analysis/visualization/statistics software*
- What prior knowledge of tools or software do you expect students to enter your class with, and what do you teach them explicitly?
- To what extent are the tools or software students use to work with data pedagogically important?
- Do you face any challenges relating to students' abilities to work with data?

How do the ways in which you teach with data relate to goals for student learning in your discipline?

- Do you teach your students to think critically about the sources and uses of data they encounter in everyday life?
- Do you teach your students specific data skills that will prepare them for future careers?
- Have you observed any policies or cultural changes at your institution that influence the ways in which you teach with data?

Do instructors in your field face any ethical challenges in teaching with data?

- To what extent are these challenges pedagogically important to you?

Training and Support

In your course(s), does anyone other than you provide instruction or support for your students in obtaining or working with data? *Examples: co-instructor, librarian, teaching assistant, drop-in sessions*

- How does their instruction or support relate to the rest of the course?
- Do you communicate with them about the instruction or support they are providing? If so, how?

To your knowledge, are there any ways in which your students are learning to work with data outside their formal coursework? *Examples: online tutorials, internships, peers*

- Do you expect or encourage this kind of extracurricular learning? Why or why not?

Have you received training in teaching with data other than your graduate degree? *Examples: workshops, technical support, help from peers*

- What factors have influenced your decision to receive/not to receive training or assistance?
- Do you use any datasets, assignment plans, syllabi, or other instructional resources that you received from others? Do you make your own resources available to others?

Considering evolving trends in your field, what types of training or assistance would be most beneficial to instructors in teaching with data?

Wrapping Up

Is there anything else from your experiences or perspectives as an instructor, or on the topic of teaching with data more broadly, that I should know?

Appendix E: Library Support Services Specific to Teaching Data to Undergraduates

The University Libraries supports students and instructors teaching with data through the following resources:

A. Digital Scholarship Center (DiSC)
   DiSC provides research support for the following creating, finding, and using data; geographic information systems (GIS); digital project planning and management; and data management and archiving. More specifically, it facilitates access to data, statistical, qualitative, GIS & mapping, text, and more, provides individual consultations, and offers training through in-person and virtual workshops and instruction and online tutorials. DiSC works with instructors to help them to develop course materials and can suggest software alternatives if issues arise with software of choice.

   a. Digital Scholarship Center Website: https://dsc.gmu.edu
   b. Research software available at the DiSC Lab and remotely: https://dsc.gmu.edu/software/

B. Subject librarians
   In addition to providing expert advice on research in their subject areas, the University Libraries' Subject Librarians also provide consultations and instruction on locating data sources for their disciplines. Often, the subject librarians and DiSC work together to help students and faculty. https://library.gmu.edu/research

C. Subscriptions (e.g., Data-Planet, Statista) and memberships (ICPSR*, Roper, HathiTrust)
   The University Libraries purchases data. Access to data sources is available through the Libraries' databases links: https://infoguides.gmu.edu/az.php?s=27306. In addition to data subscriptions

and memberships, we also purchase individual data sets. How these individual datasets are hosted depends on the license agreement.

Besides numerical and geospatial data, University Libraries also purchases or subscribes to text corpora designated for use in text and data mining projects.

*ICPSR membership is paid for by the Provost's Office.

D. Data-related Research Guides (InfoGuides)
In addition to access to data sources, University Libraries provides guidance on locating and using data and data software and digital tools. Our data-focused research guides are excellent places to start. https://infoguides.gmu.edu/dsc

Appendix F: University Resources

a. Citrix Virtual Computing Lab:
https://its.gmu.edu/service/citrix-virtual-lab/

b. Campus Computing labs:
https://its.gmu.edu/service/computer-labs/
https://its.gmu.edu/knowledge-base/what-it-resources-are-available-at-mason/

c. Software Licensed by the University for Campus-wide Use
https://its.gmu.edu/service/application-software/#software-offerings

If no campus-wide software license is available, individual departments purchase their own access. The University Libraries licenses selected research software, accessible in DiSC, to help support the data needs of campus researchers.