Fostering Data Literacy Teaching with Quantitative Data in the Social Sciences

Dylan Ruedinger
Ithaka S+R

Daniella Miriam Cooper
Ithaka S+R

Samantha Minnis
Grand Valley State University

Gayle Schaub
Grand Valley State University, schaubg@gvsu.edu

Fostering Data Literacy

Teaching with Quantitative Data in the Social Sciences

Dylan Ruediger, Danielle Miriam Cooper

Angela Bardeen
Liesl Baum
Shmuel Ben-Gad
Shaun Bennett
Kathleen Berger
Laura Bonella
Ryan Brazell
Symphony Bruce
Louise Buckley
Trevor Burrows
Scout Calvert
Patricia Condon
Renata Curty
Hilary Davis
Eleta Exline
Julia Feerrar
Emily Finch
Elizabeth Foster
Melanie Gainey
Nikhat Ghouse
Joseph Goetz
Becca Greer
Kelly Grogg

Hannah Gunderman
Samantha Guss
Michele Hayslett
Yan He
Ann James
Erin Jerome
Barb Kern
Jessica Kleekamp
Jesse Klein
Stefan Kramer
Andrew Lee
Cynthia Levine
Ken Liss
Andrew Lundeen
Kimberly MacVaugh
Wendy Mann
Clarence Maybee
Steve McGinty
Bethany McGowan
Kayla McNabb
Samantha Minnis
Jennifer Moore
Shawn Nicholson

Kayla Olson
Christie Peters
Jeff Phillips
Julie Piacentine
Nathaniel Porter
Megan Potterbusch
Mallary Rawls
Miaomiao Rimmer
Gayle Schaub
Eric Schuler
Dorris Scott
Gang Shao
Emma Slayton
Kendra Spahr
Lisa Spiro
Jasmine Spitler
Ryan Splenda
Amanda Thomas
Amanda Tickner
Emily Treptow
Torin White
Jane Yatcill

ITHAKA S+R

Ithaka S+R provides research and strategic guidance to help the academic and cultural communities serve the public good and navigate economic, demographic, and technological change. Ithaka S+R is part of ITHAKA, a not-for-profit with a mission to improve access to knowledge and education for people around the world. We believe education is key to the wellbeing of individuals and society, and we work to make it more effective and affordable.

ITHAKA is interested in disseminating this brief as widely as possible. Please contact us with any questions about using the report: research@ithaka.org.

Table of Contents

Executive Summary

Quantitative literacy is an essential twenty-first century skill that universities are heavily invested in teaching to students. The social sciences play an important role in these efforts because they attract students who might otherwise avoid data and mathematically oriented courses and because they ground quantitative reasoning in political and social contexts that resonate with undergraduates. However, pedagogical best practices for social science instructors have been slow to emerge and the support needs of instructors and students remain difficult to discern. Ithaka S+R's Teaching Support Services program explores the teaching practices and support needs of collegiate instructors. Our most recent project in this program, "Teaching with Data in the Social Sciences," focused on identifying the instructional goals and practices of instructors in introductory and advanced social science courses and exploring strengths and weaknesses of existing institutional support services. As part of this study, we partnered with librarians from 20 colleges and universities in the United States, who conducted 219 interviews with social science faculty. These interviews form the basis of this report.

"Fostering Data Literacy: Teaching with Quantitative Data in the Social Sciences" explores why and how instructors teach with data, identifies the most important challenges they face, and describes how faculty and students utilize relevant campus and external resources. Full details and actionable recommendations for stakeholders are offered in the body of the report, which offers guidance to university libraries and other campus units, faculty, vendors, and others interested in improving institutional capacities to support data-intensive instruction in the social sciences.

Key Findings

- Career skills are emphasized across the curriculum and are important factors in the software and methods that many instructors teach.
- Instructors focus on the critical interrogation of quantitative information in introductory classes, while teaching students to conduct their own research and analysis in upper division courses.
- Teaching students to use analytical software is a hands-on process requiring a significant amount of valuable instructional time, sometimes at the cost of teaching discipline-specific perspectives.
- Instructors generally avoid asking students to locate data on their own because most students struggle to find appropriate datasets. Even instructors find it difficult to locate datasets of the right size and complexity for use in middle and upper division courses.
- Faculty rely heavily on teaching assistants and liaison librarians for support when teaching with data. Teaching assistants play a critical role in teaching students to clean data and use software, and librarians help students with data discovery as well as information and data literacy.

▪ Both faculty and staff rely more heavily on web tutorials and other informal instructional resources than on workshops and other services offered by campus units to learn new information.

Introduction

In today's data-driven world, quantitative reasoning skills are central to universities' missions of preparing well-rounded, adaptable citizens equipped with marketable career skills. Quantitative literacy—the capacity to interpret and evaluate numerical information—is closely related to data literacy (the capacity to access, manipulate, and present processed data) and information literacy (the ability to locate and evaluate sources of information).[1] These interconnected literacies have been incorporated into core learning outcomes on campuses across the country, written into strategic plans, codified in the undergraduate curricula, and integrated into inquiry-based learning initiatives.

The social sciences play an important role in many of these initiatives. Some social sciences disciplines have long been dominated by quantitative research. In others deeply-ingrained traditions of qualitative research have coexisted, with varying degrees of tension, with quantitative research methods for much of the twentieth century.[2] As a result, undergraduates enrolled in social science courses have significant opportunities to practice quantitative and data analysis at increasing levels of sophistication.

> Pedagogical best practices for data-driven social science instruction have been slow to emerge even as quantitative methods become more important to social science research and to undergraduate learning outcomes.

However, pedagogical best practices for data-driven social science instruction have been slow to emerge even as quantitative methods become more important to social science research and to undergraduate learning outcomes.[3] This is unfortunate because the social sciences use statistical

---

[1] Because these literacies are closely related, usage between the terms varies and is often indistinct, see Maria Spante et al., "Digital Competence and Digital Literacy in Higher Education Research: Systematic Review of Concept Use," ed. Shuyan Wang, *Cogent Education* 5, no. 1 (1 January 2018): 1519143, https://doi.org/10.1080/2331186X.2018.1519143. Our usage here is broadly aligned with standard usage: "Keeping Up with…Statistical Literacy," American Library Association, 14 June 2017, http://www.ala.org/acrl/publications/keeping_up_with/statistical_literacy; Iddo Gal, "Adults' Statistical Literacy: Meanings, Components, Responsibilities," *International Statistical Review / Revue Internationale de Statistique* 70, no. 1 (2002): 1–25, https://doi.org/10.2307/1403713; Milo Schield, "Information Literacy, Statistical Literacy and Data Literacy," *IASSIST Quarterly / International Association for Social Science Information Service and Technology* 28 (1 June 2004): 7–14, but makes no claims to avoid slippage between these closely related terms. For changing definitions of information literacy, see Angela Sample, "Historical Development of Definitions of Information Literacy: A Literature Review of Selected Resources," *The Journal of Academic Librarianship* 46, no. 2 (1 March 2020): 102116, https://doi.org/10.1016/j.acalib.2020.102116.

[2] Gary Goertz and James Mahoney, *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences* (Princeton: Princeton University Press, 2012); Roger E. Backhouse and Philippe Fontaine, eds., *The History of the Social Sciences since 1945* (Cambridge: Cambridge University Press, 2010);

[3] David M.J. Lazer et al., "Computational Social Science: Obstacles and Opportunities | Science," *Science* 369, no. 6507 (August 28, 2020), https://doi.org/10.1126/science.aaz8170; Melanie Nind and Sarah Lewthwaite, "Hard to Teach: Inclusive Pedagogy in Social Science Research Methods Education," *International Journal of Inclusive Education* 22, no. 1 (2 January 2018): 74–88,

methods to explore social issues. Introductory social science courses in particular provide opportunities to reach undergraduates who might otherwise avoid quantitatively-oriented classes. Finally, because many social science disciplines focus on the interpretation of social or socio-economic phenomena, they are well-positioned to address some of the most important information literacy issues of our time. For these reasons, supporting social science instructors as they teach with quantitative data is an important component in ensuring that institutional goals around information, data, and quantitative literacy are met.

Since 2012, Ithaka S+R has explored how data-intensive research is transforming disciplinary methodologies and research agendas and has tracked the development and efficacy of the support services that universities have developed to support them.[4] Focused on research practices in a wide range of disciplines, the resulting reports—based on over a thousand interviews with researchers—have made it clear that undergraduates contribute to many large-scale, data-intensive research projects. By exploring why instructors teach with quantitative data, how they incorporate those materials into their classrooms, and what learning outcomes they hope to achieve, the study sheds light on instructors' instructional goals and practices and provides insights into the challenges they and their students face.

> As quantitative data has gained stature as perhaps the most critical information literacy, the need for libraries to expand the support they provide for data and quantitative literacies has become acute.

Universities, funders, vendors, and scholarly societies are all deeply invested in supporting instructors and students in these goals, and our report includes detailed information to help institutions assess the effectiveness of existing support offerings and align new services and products to the needs of students and faculty. Within universities, social science departments, general education curricula committees, and centers for teaching and learning all have important roles in supporting students as they develop data literacy skills. University libraries are currently the most critical unit offering support in this area and are likely to continue to play a leading role in future offerings in this area. In part, this reflects libraries' traditional and robust role in promoting information literacies.[5] However, as quantitative data has gained stature as perhaps the most critical information literacy, the need for libraries to expand the

https://doi.org/10.1080/13603116.2017.1355413; Sarah Lewthwaite and Melanie Nind, "Teaching Research Methods in the Social Sciences: Expert Perspectives on Pedagogy and Practice," *British Journal of Educational Studies*, 64:4 (2016), https://doi.org/10.1080/00071005.2016.1197882.

[4] For recent examples, see Dylan Ruediger and Danielle Cooper, et. al, "Big Data Infrastructure at the Crossroads," *Ithaka S+R*, 1 December 2021, https://doi.org/10.18665/sr.316121; Jane Radecki and Rebecca Springer, "Research Data Services in US Higher Education: A Web-Based Inventory," *Ithaka S+R*, 18 November 2020, https://doi.org/10.18665/sr.314397; Danielle Cooper, Roger C. Schonfeld, Richard Adams, Matthew Baker, Nisa Bakkalbasi, John G. Bales, Rebekah Bedard, et al, "Supporting the Changing Research Practices of Religious Studies Scholars," *Ithaka S+R*, 8 February 2017, https://doi.org/10.18665/sr.294119; Danielle Cooper, Katherine Daniel, Jade Alburo, Deepa Banerjee, Tomoko Bialock, Hong Cheng, Su Chen, et al, "Supporting the Changing Research Practices of Asian Studies Scholars," *Ithaka S+R*, 21 June 2018, https://doi.org/10.18665/sr.307642.

[5] See, for example, the American Library Association's "Information Literacy Competency Standards for Higher Education, January 2000, https://alair.ala.org/handle/11213/7668, since replaced by the "Framework for Information Literacy for Higher Education," 9 February 9, 2015, http://www.ala.org/acrl/standards/ilframework.

support they provide for data and quantitative literacies has become acute. While we find that libraries are meeting these challenges in many respects, there are significant opportunities for improvement in these areas.

Methods

This project is part of Ithaka S+R's continuing research on how big data and technological change are reshaping higher education and has a specific goal of providing detailed insights into how social science instructors teach with data in both introductory and advanced courses. The research for this project was conducted in collaboration with librarians from 20 universities in the United States. (For a list of participating individuals, see Appendix 1). Participating institutions created individual research teams, then trained by Ithaka S+R, to conduct research on their respective campuses. Using a shared interview guide developed by Ithaka S+R, the local research teams completed a total of 219 interviews, resulting in a high volume of data collection from instructors from a range of career stages and a wide range of disciplines within the social sciences. A detailed description of the project methodology can be found in Appendix 2.

Why Instructors Teach with Quantitative Data

Social science instructors invest considerable time to teach students quantitative literacy and skills. Particularly in introductory courses, instructors focus on helping students think critically about numerical information, in the hopes of encouraging them to be active consumers of data and engaged citizens. Quantitative skills are also highly valued by employers, and instructors often make efforts to align the skills and software they teach their students to use with those with high demand in the private sector.

Data skills prepare students to be informed and engaged citizens

Instructors across ranks and disciplines consider information literacy a core learning objective, an essential part of what it means to provide students with what one geographer called "preparation for the real world." Some instructors consider this outcome the primary goal of their intro courses. "The number one goal for my students is critical thinking…getting them to become better critical consumers of information," said one faculty member whose class is specifically designed to teach students to be better able to "understand some of the things that they read in newspapers." In advanced courses, this instructor goes even deeper into working with data, but at the intro level, their hope is to simply help students become "better critical consumers of information that they come across in their everyday lives."

Instructors often described their students, especially those enrolled in introductory courses, as passive consumers of data. First-year students in particular, one instructor noted, have not yet learned to think about "data as anything other than an objective reflection of reality." The initial challenge instructors face is to foster student awareness that data are not "neutral and real and natural," but constructed products reflecting decision making that needs to be approached critically, "so that's where I start." Another instructor described telling students in an

introductory methods course that they want them to learn to be "critical consumer[s]" of data they encounter in media as well as in academic research, "because stats can lie and it's sort of about interpretation." An associate professor of sociology put it even more bluntly: "one of the challenging things for students is actually distinguishing between legitimate, good, data—scientifically collected data—and B.S. data." Students, the professor continued, have been raised in a social context that treats every issue as composed of competing, but equally valid sets of opinions. Encouraging students to see complexity and nuance, and perhaps as importantly to recognize when arguments are built on biased, distorted, or even malicious data, was repeatedly described as an essential pedagogical goal by instructors.

> "One of the challenging things for students is actually distinguishing between legitimate, good, data—scientifically collected data—and B.S. data."

Because social science disciplines intersect directly with many of the most hotly contested contemporary cultural and political issues, many instructors connect general information literacy to explicitly civic competencies. Students often have strong opinions on social and political topics, giving social science instructors a clear opportunity to contextualize data and statistical literacy in contexts that students can readily connect to their lives and interests.[6] Our interviews suggest that few instructors, particularly in disciplines such as political science, shy away from the opportunity to discuss politically charged events and ideas. Indeed, several include course sessions focused on contentious contemporary political issues, using them as examples to encourage students to think critically about the ways data and statistics are used by politicians and in the public sphere. For example, a political scientist at a private university described using public debates about COVID-19 and the racial patterns in police brutality as materials for classroom discussion about how to make "good and bad inference[s]" about data. Another emphasized that discussing political uses of data provided an important opportunity to spark students' interests in understanding "what's misleading, what's missing, what is conveyed really effectively" by different types of data.

When defined as a general capacity to critically interrogate data in civic or social contexts, these information literacy goals align closely with learning outcomes that many universities have incorporated into their general education curricula. Our interviews suggest that social science instructors are eager to support these goals, which they frequently described as essential, urgent, components of a liberal arts education that students should learn in introductory courses. However, when instructors described higher level learning outcomes appropriate for middle and upper division courses and for majors in their disciplines, their focus shifted. The

---

[6] This contextualized approach to teaching quantitative skills is associated with improved learning outcomes: Jackie Carter, Mark Brown, and Kathryn Simpson, "From the Classroom to the Workplace: How Social Science Students are Learning to do Data Analysis for Real," *Statistics Education Research Journal* 16, no. 1 (May 2017): 80–101; David L. Neumann, Michelle Hood, and Michelle M. Neumann, "Using Real-Life Data When Teaching Statistics: Student Perceptions of this Strategy in an Introductory Statistics Course," *Statistics Education Research Journal* 12, no. 2 (November 2013): 59–70, https://doi.org/10.52041/serj.v12i2.304. See also, Jeffrey R. Vittengl and Karen L. Vittengl, "Can Teaching Data Analysis In-House Improve Psychology Students' Skills?" *Teaching of Psychology*, February 8, 2021, 0098628321992842, https://doi.org/10.1177/0098628321992842, for indication that discipline specific exposure to statistics improves student learning.

early engagements with data drawn from popular and news media—go-to sources cited by many instructors—were also intended to lay the foundation for more complex engagements with larger datasets and more complex data analysis. As students progress from information literacy towards data literacy, instructors increasingly emphasize research-oriented skills, and devote more time to topics such as methodology and research ethics. These more advanced skills were frequently described as promoting two specific goals: preparing students for careers and mastering disciplinary ways of thinking. Together with civic competencies, these are the critical "three dimensions" of data literacy, the connected outcomes of preparing a student for "future coursework, in terms of being a good employee and professional worker, but also being a good citizen."

Data literacy is a marketable skill

In recent decades, higher education has become increasingly oriented towards career preparation, a trend propelled by student demands, state and federal policies, and complex cultural and market forces. In this environment, instructors in all disciplines face considerable pressure to equip students with marketable skills connected directly to career pathways. As a whole, the behavioral and social sciences—which in 2018 accounted for almost 14 percent of all bachelor's degrees completed in the United States—have held their own in the face of these trends.[7] However, these composite numbers can be misleading. Individual disciplines have had different levels of success appealing to career-oriented students. Fields such as economics seem to have benefited from being associated with hard skills and robust employment opportunities.[8] Others, such as anthropology, have declined due in part to the perception that they do not teach concrete job skills.[9] Instructors are acutely aware of these trends: one economist described their upper division courses as focused "100 percent on job skills." Few others were willing to so completely orient their teaching towards career skills, but across disciplines instructors reported having explicit conversations with students about the market value of the skills their discipline has to offer. Quantitative and data analysis skills figure prominently in these discussions. A political scientist at a private research university noted, for example, that the materials they teach are "of wide relevance to businesses and organizations...the skills and tools that we're teaching the students are the exact same skills and tools they use in almost every industry in the real world."

While some instructors spoke generally about data analysis as a transferable skill "that will serve you well as you go out into this changing job market," many took a narrower approach and emphasized that the software they taught students to use—SPSS, R, Python, GIS, Excel—were widely used in industry. For instance, as a professor of political science who taught with R described, "the line I give them at the beginning of the semester is I'm going to teach you how to

---

[7] See the useful comparative chart of trends in bachelor's degrees published the American Academy of Arts and Sciences' Humanities Indicators program: https://www.amacad.org/humanities-indicators/higher-education/bachelors-degrees-humanities.

[8] John J. Siegfried, "Trends in Undergraduate Economics Degrees, 2001–2020," *The Journal of Economic Education* 52, no: 3 (30 May 2021): 264-267, https://doi.org/10.1080/00220485.2021.1925191.

[9] Daniel Ginsberg, "Trends in Anthropology Bachelor's Degrees: A Review of Federal Data," American Anthropological Association, September 2017, http://s3.amazonaws.com/rdcms-aaa/files/production/public/FileDownloads/pdfs/IPEDS%20anthro%20bachelor's%20degrees.pdf.

use a tool that's really powerful. And it's community based, it's open source, it's free. And it can do almost anything you can think of." The same instructor emphasized that knowledge of R would give students a leg up in work environments: "all of your competitors in the office are all going to be trotting out tried and true Excel spreadsheet graphs, and you are going to be able to wow the socks off your boss by being able to do stuff that your boss couldn't even imagine, and you'll put the rest of them to shame." A sociologist made similar observations about SPSS, telling students early and often that knowledge of the software would "look good to employers...[and] actually give you a leg up when you enter the job market."

Some instructors reported making decisions about what software to use based in part on their applicability to industry. A business management lecturer, for example, described this as a "main aim that I have in mind while introducing software—the question that I ask myself is, if the student goes into industry, will they be using this software?" A communications professor who teaches an upper-level research methods course for majors, many of them interested in careers in journalism, characterized their students as intimidated by and not especially interested in data analysis and statistics. To spur interest in their course, the professor pitched the professional value of knowing SPSS, "one of the most commonly used software packages in the industry." Similarly, several economists incorporated R into their classroom after discovering that it was more widely used in government and the private sector than STATA, the traditional statistical software used by academic economists. The overall trend away from commercial software in favor of R mirrors larger changes within higher education as a whole, which increasingly prioritizes open educational resources and tools.[10]

Upper division courses often focus on the development of research and analytic skills. These high-level skills, developed by asking students to conduct independent research and/or interpret real world datasets, require students to combine information literacy, software competencies, and quantitative literacy with emerging understandings of research methodologies and data analytic techniques. Instructors sometimes described these advanced skills as particularly important to students who wanted to attend graduate school. An associate professor of psychology, for example, reported encouraging students to gain experience working with quantitative data because having research experience as an undergraduate gave them the "best chance of getting into a graduate program later on." Others, however, recognized that research and analytic skills were important job skills. A sociologist who has taught many students who went on to careers in criminology or related legal fields described "research and research skills as well as data analysis tools and techniques" as laying the "foundation" for future careers.

---

[10] "Why Universities are Switching to R for Teaching Social Science," Sage Campus, 23 August 2019, https://campus.sagepub.com/blog/why-universities-are-switching-to-r-for-social-science; Robert A. Muenchen, "Is Scholarly Use of R Beating SPSS Already?" R-bloggers, 15 July 2019, https://www.r-bloggers.com/2019/07/is-scholarly-use-of-r-use-beating-spss-already/; Robert A. Muenchen,"The Popularity of Data Science Software," r4stats.com, https://r4stats.com/articles/popularity/.

Challenges in Teaching with Quantitative Data

Teaching with data requires synchronized development of students' analytic skills and mastery of software, a complex pedagogical challenge that is compounded by student's uneven levels of computer literacy and the relative scarcity of datasets of the right size and scale for undergraduate teaching. Scaffolding of departmental curricula to support progressive learning across a major can help alleviate these challenges, but many social science departments struggle to achieve this level of coordination.

Balancing skill development and domain knowledge

Working with data is a cluster of competencies rather than a single skill. Students need to progressively master basic computing concepts like file directories, gain competency with software tools, learn data-wrangling and analysis skills, and absorb disciplinary methodologies and perspectives. Teaching with data requires considerable investment by instructors to help students learn all these skills simultaneously, since falling behind in one area can lead to students falling behind in others. However, instructors frequently expressed concerns that the amount of instructional time they devoted to teaching software and basic computing skills impinged on their ability to focus on the content and domain knowledge they most want students to learn.

> Working with data is a cluster of competencies rather than a single skill.

Students enter social science courses with highly variable levels of familiarity with the basics of computing. They use a wide range of computer models and operating systems, making even installing necessary software a challenge for instructors that can require considerable amounts of valuable instructional time. Even more importantly, meaningful numbers of students are unfamiliar with the process of navigating file structures or downloading and extracting software packages. Others have never used essential software such as spreadsheets.[11] "I give them slow, step-by-step instructions on that, which I'm sure is remedial for some of them, but just teaching them what a CSV file is, for example, is new for many of them. And helping them understand what rows and columns are for," said an assistant professor. "That's time I wish I didn't have to spend, but somewhere someone along the way has to guide them through it."

Not all students need help with these basics, but levels of preparation are variable enough that instructors anticipate that every class will include true novices. For students who need the extra instruction, the class time instructors spend on fundamentals is well spent, and most instructors recognize their responsibility as a teacher to help these students succeed. However, students who have mastered the basics can become frustrated and bored as they wait for peers to catch

---

[11] Samuel J. Rubin and Binyomin Abrams, "Teaching Fundamental Skills in Microsoft Excel to First-Year Students in Quantitative Analysis," *Journal of Chemical Education* 92, no. 11 (10 November 2015): 1840–45, https://doi.org/10.1021/acs.jchemed.5b00122; Donna Grant, Alisha Malloy, and Marianne Murphy, "A Comparison of Student Perceptions of Their Computer Skills to Their Actual Abilities," *Journal of Information Technology Education: Research* 8, no. 1 (1 January 2009): 141–60, https://doi.org/10.28945/164.

up. The significant amount of class time devoted to teaching software extends well beyond the few class sessions. A linguist described spending weeks "just getting [students] comfortable with R," while a sociologist estimated that roughly a quarter of their instructional time was spent "going over aspects of SPSS." These heavy investments of scarce instructional time necessarily reduce the amount of content and domain knowledge instructors can teach. However, because higher level data literacy and disciplinary methodologies are so dependent on software skills, most instructors recognize the need to make this investment of time, and work to maximize the long-term value students will get from it.

> These heavy investments of scarce instructional time necessarily reduce the amount of content and domain knowledge instructors can teach.

One way instructors do so is by teaching students to use either free tools such as R or general use software such as Excel or Google Sheets. Part of the rationale for doing so stems from the larger movement within higher education towards open educational resources, which has emphasized limiting the costs to students of accessing course materials. However, instructors also described their preference for free or widely used tools as a way of ensuring that students learn tools they will likely encounter in their careers. On the other hand, some instructors shy away from R and other free tools because they believe they are more difficult to learn than commercial software. R, one political scientist noted, is clearly more versatile than SPSS or STATA, but because it has a steep learning curve, they still use commercial options. A government professor agreed, noting that they had drifted back to using SPSS because "it is more user friendly, and there is less of a learning curve."[12]

Ultimately, for most instructors, mastering tools is a secondary learning objective, a means to the end of developing analytical skills. "I'm always very wary if someone says, 'The only way to learn something is through Python,'" said a geography and GIS lecturer. "It's more important to think about data analysis as different sets of problems and then we have to take different software and tools to kind of adapt to that problem and then part of our responsibility is to be able to be fluent. I mean, maybe not fluent in multiple languages, but adaptable." Likewise, an associate professor of communication emphasized that "what I try to teach the students mostly is that you just need to know what the problem is you're trying to solve." Software tools come and go but the solutions "aren't going to change all the time." The challenge, of course, is that software is the medium through which students learn the underlying skills and perspectives that most instructors ultimately consider their primary learning objective. As an anthropologist noted, for all practical purposes data and software go "hand-in-hand."

---

[12] Some research suggests that while SPSS is initially easier for students to learn than R, over time students will adapt equally well to R. Jacob B. Rode and Megan M. Ringel, "Statistical Software Output in the Classroom: A Comparison of R and SPSS," *Teaching of Psychology* 46, no. 4 (October 1, 2019): 319–27, https://doi.org/10.1177/0098628319872605; Ruoxi Li, "Teaching Undergraduates R in an Introductory Research Methods Course: A Step-by-Step Approach," *Journal of Political Science Education* 17, no. 4 (October 2, 2021): 653–71, https://doi.org/10.1080/15512169.2019.1667811.

Finding data for classroom use

In order to teach with data, instructors need to locate datasets appropriate for classroom use. In today's environment, instructors' primary difficulties are born from abundance rather than scarcity: wading through datasets readily at hand to locate ones suitable for undergraduate teaching was often reported as a major challenge. There are important exceptions—for example, international data can be more difficult to access than data from the United States. But as one instructor said, "I don't face challenges at all finding data. I can always find data to work with." Many of their colleagues, regardless of field, agreed. One described an "embarrassment of riches of data that's available." However, the decentralized and disorganized state of those data can quickly turn this bounty into a liability rather than an asset. Many instructors described the difficulty of finding data scaled for classroom use, citing disorganized repositories and difficulty navigating websites to discern where data can be accessed and downloaded.

Instructors often rely on their existing knowledge of common datasets in their field as a starting point when looking for materials to use for teaching. As an economist explained, "most of our research is empirical, depending on whether we are health economists, labor economists, there are several data sources that we all are aware of that we tap into." Faculty often share datasets with colleagues in their department or discover promising data resources via domain listservs, social media, and word of mouth. On occasion, faculty share data from their personal research with students for instructional purposes. An associate professor of communications noted giving students in a methods class "data that I've collected" for use in practice sessions. Another described making use of "things I'm working on, and I just bring it right in and make the lab about that. Other times it's something that I wish I had time to work on, but I couldn't, and so I write a small short lab around it."

Particularly in introductory courses, the topic of a dataset is usually less important than its size and usefulness for teaching basic concepts or learning software. For this reason, instructors often use datasets included as tutorials in software packages or as part of course textbooks. These datasets are sometimes composed of "made-up" data, designed purely for instructional purposes. "I'm using a workbook that comes with the textbook and in the workbook they have some made-up data—some very small data sets," said an assistant professor of political science at a private university. In addition to relying on textbook and tutorial datasets, as well as data offered by vendors like Kaggle, instructors frequently make up their own simulated datasets to give to students.

Libraries play little role in faculty data discovery. When faculty make open-ended searches for datasets, they overwhelmingly use Google or other search engines as discovery tools. As a full professor of political science noted, "back in the old days, the library was always helpful in being able to, you know, get canned data sets and stuff like that. But I mean, the world has moved beyond that, that's been true for 15 years."

Teaching students to locate data

Instructors, particularly in introductory level courses, usually avoid having students find their own data, preferring to eliminate that variable from the complex mix of skills and competencies involved in teaching with data. In fact, instructors at all course levels often prioritize teaching students to analyze data at the expense of teaching them to find or generate their own data. The most significant exceptions to this tendency are upper division capstone courses, where students typically write research papers based on data they locate or create.

But in lower and mid-level courses, where the focus is on software and methods, most instructors provide students with datasets to work with. As one graduate student who taught as an instructor of record said, "at the beginning of the semester or towards the beginning of the course, I'll be providing them their own data. But maybe later on in the semester or as part of a final project maybe they need to find their own data." Instructors in disciplines as different as statistics, sociology, psychology, and geography agreed that except in final projects, they preferred to give students preselected data, in many cases simplified or reduced in size. The reason that instructors provide data is because data discovery is an advanced skill with a steep learning curve, among the most difficult components of data literacy. An assistant professor at a private university noted that students consistently struggle to find their own data because "they are not necessarily accustomed to sussing out the quality of a data set just by looking at it." Students face several barriers to effective discovery.

The most common obstacle students face around discovery is their lack of familiarity with the process of locating data. "Very often," said an assistant professor at a private university, "they've never used a library before. Many of the students that I interact with don't know how to search for information either online or, you know, using basic library resources, so that's a challenge. They're not used to doing research, basically at all." The instructor, and many others interviewed for this project said their students were "easily deterred by small entry barriers." Library discovery platforms were often full of such barriers. A statistician noted that their university library holds a wide range of useful data resources, but that accessing them requires "like 15 clicks." Faculty are sympathetic with their students on this point: some even report that they, too, find the library catalog difficult to navigate. For example, a lecturer in international affairs described their experience trying to access a subscription database of public opinion research through the library and finding the process "very, very unclear." While they eventually figured it out, they recognized that their students would find the process impossible.

Like faculty, students preferred Google to the library catalog and even to curated lists of resources provided by their instructor. As one instructor said, "the biggest hurdle that I have when students have to find their own data is still that they don't refer to any of my reference materials that I've worked so hard to cultivate, and they just started Google searching." Here, too, they encountered a range of difficulties. Many instructors noted that students tend to select the first data they found, usually at the top of search rankings, instead of taking the extra time to search for data that truly fits their needs or the parameters of an assignment. This search strategy rarely works well, often leading students to use "bad datasets from weird sites." Students may also run into paywalls, particularly for high quality datasets, or only be able to

access partial data. As one instructor noted, "not all data sets are free. Sometimes the data set might look really good, but it's always behind a paid service. So I think that's one of the biggest challenges. Data sets that are free, they're often not complete. So there might be gaps within those data sets. But really good data sets I feel you always have to pay a fee for." Another noted that "finding freely available sources" was the "number one challenge" their students faced.

Developing students' research skills

Because finding useful data is so difficult, instructors often provide extensive guidance to students throughout the discovery process. As a lecturer in anthropology noted, "even with fairly comprehensive explanations, trying to sort of give them tutorials, and walkthroughs on how you actually go about finding data and things like that sometimes when push comes to shove, they do not necessarily always find the data on their own. They require a little bit of extra hand holding, or they just don't do it." In upper division courses, this might involve literally weeks of back and forth with students.

Interviewees had developed many strategies for helping students locate research data. One of the most common was to provide resource guides to students that include links to websites with useful data. Government data and datasets from major nonprofits and think tanks figured prominently on such lists. Even when provided with links, students still typically required additional support. A professor of economics, for example, mentioned that they had learned that "If I just gave them the assignment and the assignment clearly says go to this link, this table, download this data. A lot of them would struggle. So they struggle less when you go over it in the classroom, at least walk them through the process." An associate professor of geography concurred. They described providing students with a list of relevant data sources and an overview of useful search terms and strategies as a first step in the process of guiding students through the discovery process. They also devoted an entire class session to "finding data, and how to do it, and where to look, and what to look for," and a tutorial about "what different data looks like, and different formats of interacting with things like web-based queries versus FTP sites." A third instructor remarked that they had learnt the hard way that taking time to demonstrate "Google searching and troubleshooting" was an important component of teaching students to work with data. This significant level of instructional support did not solve student's discovery challenge, but instructors believe that it did improve the quality and appropriateness of the data their students found.

The difficulties instructors described are by no means unique to discovery of quantitative datasets. Previous research from Ithaka S+R and others has described analogous problems when students seek out qualitative data and primary sources, making it clear that student discovery is a challenge across disciplines.[13] However, discovery is essential to learning, so instructors—especially when teaching upper division students—are often willing to invest the time required to help students acquire this skill. As several instructors mentioned, discovery is an important part of data literacy and a necessary part of the process of learning to conduct research. Across the social sciences, instructors agreed that developing student's capacities to

---

[13] Kurtis Tanaka, Danielle Cooper, and Dylan Ruediger, et al, "Teaching with Primary Sources: Looking at the Support Needs of Instructors," *Ithaka S+R*, 23 March 2021, https://doi.org/10.18665/sr.314912.

ethically collect and properly structure research data was considered a core learning outcome of the major, the culmination of the progressively more complex encounters with data across what one faculty member called the "full lifecycle from finding data to presenting results."

Instructors agreed that asking students to collect primary research data or perform original research with secondary datasets was uniquely challenging. One particular difficulty students face is understanding how to work with data at scale, even within the relatively modest confines of an undergraduate research project. Students often start with ideas that are overly ambitious, sometimes by orders of magnitude. As a lecturer in the geospatial sciences remarked, "students tend to underestimate how difficult the whole process is, and it really limits the scale of what they can do." Instructors often invest considerable time helping them reign in their expectations and juggle the difficult task of finding a small-scale project that is both interesting and possible to complete in a single semester.

A second challenge students face is aligning their research questions to data that either exists or which they can collect. In certain disciplines, notably international relations—a field in which a student's ability to generate meaningful research data is sharply limited—this can be a particularly thorny issue. As one international relations instructor noted, their students will often start with what are excellent, but impossible topics, such as exploring human trafficking on a global scale. Even when they can help students to re-scope their projects to a manageable scale, they still frequently run into dead ends: "data availability," they said, is "a constant problem, a really big issue." To try and head off these kinds of problems, the instructor provides students with a pre-selected list of research topics that are germane to the course and for which the instructor knows data is available. This approach, used by instructors in other disciplines as well, tries to find a middle ground that provides students with the option to research a range of topics while directing them towards projects for which meaningful data exists.

Instructors who did require students to generate original research data were highly likely to be teaching capstone courses, which were typically the first course in which social science students were expected to conduct research.[14] Instructors guided students to a variety of data collection methods. For example, some reported that their students used data mining or scraping tools, and others indicated that their students had digitized existing data so it could be processed with course software. In a few instances, faculty integrated students into their personal research projects, though this practice is much less common in the social sciences than many STEM fields. The most common way that students collected data was through surveys, a research instrument that is relatively easy to scale to undergraduate level work, even if it places limits on the types of research questions students can pursue. Survey based research projects can require students to obtain IRB approval. Though this introduces yet another hurdle, instructors sometimes welcomed this because it forced students to engage deeply with issues relating to ethics and privacy.

---

[14] A finding consistent with Theresa Burress et al., "Data Literacy in Undergraduate Education," in *Data Literacy in Academic Libraries: Teaching Critical Thinking with Numbers*, ed. Julia Bauder (Chicago: ALA Editions, 2021): 11, which found that "faculty in social science disciplines tended to reserve the collection of original data to senior or capstone-level courses."

Another challenge is that students often misjudge the labor involved in cleaning data, which is perhaps understandable given that they have most often used only very small, pre-cleaned data, in their previous courses. While sympathetic to the time required to clean data, instructors welcome the opportunity to give students practice with this part of data analysis. "The experience of finding and obtaining and cleaning and discovering new data is both a positive thing and a necessary thing," said one instructor, "even though it is the most difficult and the most painful and the most challenging part of the entire process."

Data cleaning requires students to make a series of small but consequential intellectual decisions and interpretations of their data. As one instructor put it, creating clean, processed data, requires numerous layers of human decision making and expertise. "So a lot of it is just kind of hand-holding students, holding their hands along the way and it is very frustrating for them to hit a lot of these brand new topics in ways that they haven't experienced before. And so my job is to kind of remind them that that's a good sign. That means they're learning." Another noted wanting to teach their students that "80 percent of any analysis is just working with the data," and that "making a choice about [how to fix] those errors is a huge part of what analysis is all about." For these reasons (as well as the obvious fact that raw data collected by students needs to be cleaned before it can be used), instructors allot students significant time in the research process to clean their data, even at the expense of time spent gathering and analyzing it.

> The cumulative impacts of decisions by other instructors about how to scaffold increasingly complex encounters with data can meaningfully impact students' ability to succeed.

Undergraduate research projects provide a unique opportunity for students to put together the various competencies they have developed over the course of their major, to iterate and synthesize at a high level of difficulty. This is a significant challenge that can be exacerbated if a student's capacities with the requisite methodological, analytic, and technical skills involved in quantitative research have developed unevenly. Our interviews make it clear that finding ways to ensure that students are making consistent gains in all these areas is perhaps the major challenge instructors face. For instructors in the upper division courses where these competencies are tested as students design and conduct independent research projects, the cumulative impacts of decisions by other instructors about how to scaffold increasingly complex encounters with data can meaningfully impact students' ability to succeed.

Instilling a sense of ethics

Ethics are an important aspect of data literacy that many instructors emphasize in their teaching. As such they warrant special attention here. Faculty approach the topic of ethics in several ways, treating it as an aspect of critical thinking or as a component of sound research practices depending on the context.

In the first instance, instructors consider ethics as substantively overlapping with questions of interpretation and information literacy. One faculty member described beginning each semester

by encouraging students to consider what they described as ethical questions such as how data has been collected, by whom, and for what purpose. A lecturer in the environmental sciences said that they teach students to consider how biases can be built into the GIS software they will be using in the class. "I try to be very explicit that GIS is a non-neutral software…it is built with biases and it can do biased things." A colleague in a geography department emphasized the importance of raising student's awareness of the "encoded biases" that lie hidden within algorithms and visualization software, while a graduate student teaching introductory economics devoted a course session to the "idea of misrepresentation of data," in which they show how same data set can be used to create wildly different conclusions supporting different unspoken agendas "by just altering things slightly."

In middle and upper-level courses, ethics are often introduced as part and parcel of sound research design. Within academic settings, these ethical considerations are usually governed by IRB boards that supervise research conducted on human subjects. The historical development of IRBs is something at least a few instructors specifically discuss with students. A statistician, for example, takes students "through just a brief history of human subjects research," because "not all of them know of the Tuskegee experiments or the Stanford prison experiment or some of the early stuff." The instructor believes that it is "really important for them to learn to think about, especially in this world of big data." Faculty regularly familiarize students with disciplinary ethics standards and with the basics of the IRB process, often drawing on their personal experience to do so. A professor of political science at a private university, for example, discusses their experience as a member of the university's IRB committee with students. Another political scientist "shared my IRB proposal from last year" with students and explained the modifications to their research plan that the IRB mandated. That, the instructor said, "was a perfect teaching opportunity for me, to actually share with the students my proposal, and have them analyze it, and imagine that they are IRB reviewers. Would they approve of the proposal or not?"

Understanding the IRB procedures is especially important if students will be asked to collect their own primary data on human subjects, as it will inform their research plans. Most often, student research conducted for educational purposes does not require IRB approval unless students plan to publish their findings, but instructors often want them to understand how the IRB process informs research and protects the rights of research subjects. In at least one instance, an instructor required their students to complete CITI research ethics training as part of the course.

Scaffold learning within a course

One constant across our interviews was that instructors recognized the difficulty of teaching data literacy and acknowledged their uneven success in achieving their goals. Instructors frequently expressed the value of greater coordination across the curricula of a major. While some described their departments as having made progress on this front, most did indicate that efforts to standardize learning across classes was difficult to accomplish, and that their department's efforts to do so were at best modestly successful. As a result, instructors focused much of their attention on scaffolding learning within their courses.

Instructors routinely built their courses around a calibrated and iterative exposure to data analysis and software, starting with very small exercises and progressing towards more complex assignments. This approach was seen as most likely to successfully allow students' skills to develop in tandem. Ideally, very simple assignments at the beginning of a course provide the basis for students to perform more difficult tasks and work with more complex data over time.

> Instructors were largely in agreement that active learning and hands-on instruction were the most effective way to teach students to work with data and become proficient with software.

Instructors were largely in agreement that active learning and hands-on instruction were the most effective way to teach students to work with data and become proficient with software. Some instructors, especially those with relatively small enrollments, reported that they flipped their classes and frequently provided opportunities for students to engage in collaborative learning. For example, an archaeologist enlisted undergraduates to name files and create metadata for their ongoing archaeological research. Because that work is exacting and can be tedious, the professor had students do that work together in class, a practice that increased student interest in the work and facilitated co-learning. "If somebody has a question," the professor said, "if I can't answer it usually there is a computer whiz in the class that could answer it…And so it was a kind of a group learning process and, you know, if they applied themselves, they could typically get the project done during class time." A professor of political science used similar techniques, saying that "the class is taught as kind of a flipped class. I don't lecture. They spend a lot of time in their groups during class time and I work with them and I'll bounce from group to group figuring out who's got what kinds of problems." In large introductory courses, this kind of intensive, hands-on learning is often delegated to teaching assistants who lead lab sections focused on developing software and data skills.

Building a coherent curriculum

Even with carefully designed instruction, students are unlikely to master data skills during a single course. Our interviews focused on course level instruction but yielded occasional insights into the challenges of coordinating learning across a major. As noted above, many instructors recognized that more consistent reinforcement of learning outcomes across courses would help their students but had little faith in their department to enact the measures necessary to accomplish this goal. As one faculty member said, integrating student's skills with data required "at least some thought to progression and building on foundations." Unfortunately, neither their department nor their university's general education curricula provided this kind of coordinated learning experience.

Our interviews turned up evidence that some departments are attempting to better scaffold learning across a major. However, relatively few have enacted policies and processes to support this goal. The autonomy that faculty feel towards their class design and pedagogical choices was the most prominent explanation for why departments struggled to effectively scaffold student

learning across the curricula of a major. Some faculty, including those at regional comprehensive universities with teaching focused missions, said that colleagues in their department avoided even talking with each other about instructional practices. "We basically believe very fervently in everybody's right to teach their courses the way they want," said an associate professor. "So we don't discuss what we teach with each other at all." In this kind of environment, making decisions about standardizing software across courses or steadily building student discovery skills so that they are ready to conduct their own research in capstone courses is all but impossible.

> Departments would likely benefit from coordinated efforts to integrate quantitative analysis into more courses across a major, instead of relying—as at least some departments do—on a single methodology class to do much of the heavy lifting.

Many instructors noted that students frequently find the transition from the initial, highly structured exposure to data and software to relatively open-ended expectations that they can conduct meaningful research to be a difficult one. Mid-level courses do not seem, on the whole, to effectively bridge the gap between the skills taught in introductory courses and those required to succeed in upper-level courses. Departments would likely benefit from coordinated efforts to integrate quantitative analysis into more courses across a major, instead of relying—as at least some departments do—on a single methodology class to do much of the heavy lifting.[15] Another is to consider adopting more rigid sequencing of courses. As a political scientist noted, their department does not require students to take a mid-level political science course focused on methodology at any particular time, and students often wait to take it until their final semester in college. As a result, they miss the opportunity to use that methodological knowledge in other mid-level courses and are frequently learning methods in the same semester that they are taking capstone courses that require them to conduct original research. Rigidly sequenced programs of study are uncommon in social sciences disciplines, but other fields—such as chemistry and nursing—provide models for departments to consider. Faculty also expressed interest in changes to the university's general education curriculum that would, ideally, help equalize basic computing skills and levels of information literacy among students.[16]

---

[15] Unislawa Williams et al., "Teaching Data Science in Political Science: Integrating Methods with Substantive Curriculum," *PS: Political Science & Politics* 54, no. 2 (April 2021): 336–39, https://doi.org/10.1017/S1049096520001687.

[16] Brian Kim, "Scaling Up Data Science for the Social Sciences," *Harvard Data Science Review* 3, no. 2 (June 7, 2021), https://doi.org/10.1162/99608f92.d3f14ea4. For similar concerns in the behavioral sciences: Ronald R. Hoy, "Quantitative Skills in Undergraduate Neuroscience Education in the Age of Big Data," *Neuroscience Letters* 759 (August 10, 2021): 136074, https://doi.org/10.1016/j.neulet.2021.136074.

Supporting Faculty Teaching

Faculty rely heavily on teaching assistants and liaison librarians for support when teaching with data. Especially in introductory courses, teaching assistants often take primary responsibility for teaching software and data cleaning. Librarians are used primarily for assistance with data discovery, but occasionally collaborate with faculty in more extensive ways. The most frequently mentioned areas where faculty would like more support from their institutions include pedagogical training specific to teaching with data, tutorials on new software, and opportunities to share resources and ideas with colleagues.

Collaborators

As we have seen, social science instructors tend to be highly protective of their autonomy on matters relating to course design and content. Unsurprisingly, many of the instructors interviewed for this project did not appear to seek opportunities for meaningful collaboration with others on issues relating to teaching with data. Across the interviews, many potentially valuable campus resources such as writing centers and centers for teaching and learning were barely mentioned. Nor did we turn up significant evidence that instructors routinely discuss instructional practices with colleagues in their department. There are two specific exceptions to these trends: teaching assistants and librarians, both of whom frequently collaborate with faculty.

Graduate students are the most important of these exceptions. Particularly in introductory courses, the common lecture and lab structure promotes a division of labor in which faculty focus on "abstract" or "big picture" instruction, with hands-on instruction handled by teaching assistants. Instructors often described teaching assistance as the front line of teaching with data in the social sciences. A professor of political science noted that graduate students served as "the go-to person when students are having problems with R." A tenured geographer required that their TAs "hold offer hours every week" and expected them to "help the students find data and work through any issues they're having" with discovery and access of data. Likewise, a professor in a psychology department also reported that teaching assistants, rather than the instructor of record, were responsible for helping students learn to work with SPSS. Because graduate students often serve as graders, they also play important roles in assessing student learning outcomes related to data as well.

> Many faculty said that they direct students to library collections and LibGuides or encourage them to contact librarians for assistance with data discovery.

The importance of the library in supporting students' development of data literacy skills was a recurrent theme throughout the project interviews, with many faculty speaking highly of the assistance librarians provided. Many faculty said that they direct students to library collections and LibGuides or encourage them to contact librarians for assistance with data discovery. Less often, librarians introduced students to new software or helped them troubleshoot and problem

solve coding challenges. One faculty member spoke of their liaison librarian as someone who "walks on water as far as I'm concerned." They had her give presentations to students of all levels, from "intro courses all the way up to PhD level classes." "I always tell my students," the instructor noted, that "she is my gift to them because she is instrumental in helping them define their projects, in helping them find and access data, and also in helping them identify appropriate literature." Many instructors valued the expertise of librarians, especially liaison librarians, who they frequently invited to guest lecture about library resources or the general topic of information literacy.[17] When available, instructors incorporated modules or assignments designed by librarians about information literacy into their syllabi. As an associate professor of geography put it, the campus library is an "amazing" resource. "They have great links to data, tutorials—super helpful...I send students there all the time." Overall, the project interviews demonstrate that the library, and especially subject librarians who are prepared to conduct meaningful outreach and able to provide discipline specific support, is valued by instructors across the social sciences.

In a few instances, instructors reported engaging in sustained and extensive collaboration with librarians. For example, an assistant professor described having a "partner librarian." "She works with me to teach students how to access sources," they said, and "teaches two class periods every semester. The faculty member continued by saying that they often turn to the librarian when they have questions about datasets and "she'll send me things that she thinks my students might be interested in relating to researching and writing with data." Another faculty member meets with their subject librarian "multiple times a semester." "We communicate very often," the faculty said, "and we expressly plan lessons together."

Ongoing pedagogical training needs

Many instructors have learned to teach with data on the job through trial and error. A handful of faculty had taken advantage of opportunities to gain skills as teachers, and even specifically around teaching with data, while working on their PhD. Most often, these opportunities were available outside their department and its curriculum, requiring students to seek them out. A sociologist, for example, benefited from working at a campus data center, providing assistance to faculty and students about how to analyze data using different software. In the process, they picked up important skills in coding, data management, and data discovery. Some interviewees described their experience as teaching assistants as giving them early exposure to teaching with data. Even so, the predominant opinion voiced by faculty was that, as a sociologist put it, "I've really been kind of self-taught in my pedagogy." A faculty member in an education department described having no "formal training in teaching students about data in my graduate degree"

---

[17] This finding contrasts with earlier research on the topic. Karen Hogenboom, Carissa M. Holler Phillips, and Merinda Kaye Hensley, "Show Me the Data! Partnering With Instructors to Teach Data Literacy" (ARCL 15th National Conference, Philadelphia, Association of College & Research Libraries, 2011), https://hdl.handle.net/2142/73409, found that instructors identified having in-class data literacy trainings offered by librarians as their least favored format for including them in student learning, a fact they attribute to the scarcity of instructional time. Perhaps the growing relative importance of data literacy has led instructors to be more open to using class time in this fashion.

program. Another tenured faculty member noted, despite having spent ten years in graduate school, "nobody ever taught me anything about teaching."

Once they enter the ranks of postsecondary instructors, many interviewees at least occasionally take advantage of structured opportunities to improve their pedagogical practices. Workshops were one prominent means of continuing education in this space: many instructors used them to learn new software or methodologies, and—less often—reported taking workshops specifically focused on teaching with data. Those who had done so seem to have found these resources useful, but our impression is that faculty take advantage of these resources irregularly and prioritize their research agendas and other high stakes aspects of faculty life over developing pedagogical skills. Most instructors recognized that they and their students would benefit from additional training but also indicated that they had little time to devote to this pursuit.

Keeping up with new software

Throughout the interviews, instructors articulated a need for continual training in software, since the tools used to analyze data change frequently. Many instructors were trained on STATA or SPSS and feel less confident with the open source programming languages such as Python and R that they recognize are eclipsing commercial options. Online resources such as tutorials and instructional videos were widely used by faculty to develop software skills. According to one instructor, "YouTube is just loaded with videos on different things. There's always something you can learn about how you present [any] particular material...even though you know the basic idea...there are good examples you can pick up." Some instructors had even created tutorials to share on the web. In some cases, such as GIS, commercial vendors like ESRI offer free training modules and tutorials that are useful to faculty.

Faculty value these kinds of resources because they can be accessed on-demand and often address more specific issues than are possible to cover in the general training workshops organized by libraries, scholarly societies, and similar entities. Perhaps most importantly, these resources were seen as placing minimal demands on faculty members' time. As an associate professor of sociology put it, "in the last couple of years I've found tutorials on YouTube to be extremely helpful. So, I'll be thinking, like, I'd really love to be able to animate this series of graphs in this particular way and I don't even know if it is possible." YouTube allows them to "spend maybe 10 or 15 minutes, search, and find something that's exactly what I need. It walks [you] through it step by step."

Sharing knowledge and resources

Few instructors identified their departments as a meaningful resource for professional development relating to pedagogy or software skills. The few instances that departments were seen as relevant were usually in departments that had made a decision to standardize the tools used in their courses, and thus had incentives to make sure that all faculty were trained in that software or programming language. Pedagogical training in departments was usually dependent on personal interactions rather than department policy. Many departments seem to lack what

one political scientist described as a "culture of sharing," even though faculty occasionally recognized that sharing teaching resources would save them time and benefit their students.

Most instructors responded favorably to the idea that their institution might offer additional training and support opportunities that would either increase their own data skills or help them teach students more effectively. Many expressed concerns about blind spots in their competencies with newer software tools or relative lack of knowledge about how to find data sets that would be useful in pedagogical contexts. Even simple opportunities for in-depth discussions about teaching with colleagues was seen as a welcome innovation by many instructors. As a linguist noted, the chance to share "anecdotes of things that work really well or don't work really well from other instructors is really helpful for me, and I can assume probably for everyone." Ideally, these conversations would be discipline specific, because learning outcomes and even the use of common tools such as R vary from field to field. A geographer who had attended an informal working group of faculty who used R for teaching, ultimately quit because "what's really helpful for one field may not be for another field." Relatively few instructors had found forums at their institutions.

Supporting Student Learning

Perhaps the most common hurdle students face is their anxiety around math, a barrier that instructors tackle with a combination of empathy and carefully designed assignments that allow students to develop confidence over time. When seeking assistance outside the classroom, students favor informal resources such as web tutorials and data competitions to the formal resources offered by various university units.

Overcoming anxieties about math

Many students are terrified of working with quantitative data.[18] This anxiety was noted by instructors across disciplines and at selective as well as access-oriented institutions. I hear so often," said an associate professor of psychology at a selective institution, "I'm not a stats person. I'm not a math person. I'm really bad at stats." An assistant professor of urban planning at a large public university agreed: "not everybody is comfortable with numbers. Some of them just start seeing a big spreadsheet and start freaking out." A communications professor frequently heard "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,'" from students, who "kind of cringe and smirk all at the same time," as they described their relationship to math. A faculty member who taught a

---

[18] Nind and Lewthwaite, "Hard to Teach: Inclusive Pedagogy in Social Science Research Methods Education"; Stefanie S Boswell, "Undergraduates' Perceived Knowledge, Self-Efficacy, and Interest in Social Science Research," *The Journal of Effective Teaching* 13, no. 2 (2013), https://files.eric.ed.gov/fulltext/EJ1092119.pdf; Brien L. Bolin et al., "Impact of Research Orientation on Attitudes Toward Research of Social Work Students," *Journal of Social Work Education* 48, no. 2 (1 April 2012): 223–43, https://doi.org/10.5175/JSWE.2012.200900120; Gail Markle, "Factors Influencing Achievement in Undergraduate Social Science Research Methods Courses: A Mixed Methods Analysis," *Teaching Sociology* 45, no. 2 (1 April 2017): 105–15, https://doi.org/10.1177/0092055X16676302.

statistics course aimed at social science majors, described it as a "class for students who have an absolutely terrible fear of math and a fear of quantitative data."

Obviously, social science instructors are not the only faculty members whose students are intimidated by statistics. However, students often enroll in social science classes—both at the introductory level and even in upper division courses as majors—with the impression that they are qualitatively oriented fields of inquiry. Indeed, quite a few instructors believed their courses disproportionately attracted students who were seeking to avoid math and statistics. A sociologist, for example, described students as "com[ing] in thinking that sociology is very much opinion based and looks at history and stuff like that." Students often enroll in linguistics courses, said another instructor, expecting to be "talking mostly about language and identity."

Student's perceptions that the social sciences are qualitatively focused are not entirely inaccurate. Many social science disciplines, including sociology, political science, anthropology, and psychology, have historical roots in qualitative research and many remain open to them. However, in most social science disciplines, quantitative methods are now either the dominant mode of inquiry or are so sufficiently entrenched as to require that students be familiar with them.

> In most social science disciplines, quantitative methods are now either the dominant mode of inquiry or are so sufficiently entrenched as to require that students be familiar with them.

The linguist quoted above was one of many faculty who described regularly having to acclimatize students to conceptualizing "language as data" and embracing the quantitative focus of the discipline. The sociologist sought to demonstrate that sociology was a science, based on gathering and analyzing quantifiable, empirical data through sound methodological procedures. Developing students' skills in these areas are, many instructors agree, fundamental goals of their teaching. As an assistant professor of political science argued, the discipline is now so rooted in quantification that without exposure to statistics and data analysis, students would be "hard-pressed to even say what political science is."

To develop their students' confidence, instructors often expressed empathy, offered encouragement, and emphasized the value of quantitative knowledge. They frequently told students about their own fear and discomfort with statistics. As one instructor put it, they hoped to "reframe[e]" their student's fears by telling them that "I kind of felt the same way, but then once I've learned how to apply those stats to address questions that were of interest to me and once you practice it, it's much easier when you're using it in a meaningful way. And so, I think just trying to convince students that this is not their enemy, this is a tool and that it's, you know, it's fun." Others drew connections between numerical data and meaning, describing statistics and quantitative data as "tools to really see what's going on in the society," a way of telling stories and making meaning. This tactic, they believed, "helped them feel more comfortable." "Some students," said an associate professor of psychology, "just don't want to have anything to do with numbers at all. It just intimidates them." The professor worked hard to make sure that

they presented numerical data within the context of psychological questions and issues, as a way of exploring topics that were of great interest to their students. "I do think that helps and it removes some of the anxiety," they concluded. Instructors also take pains to carefully scale the level of statistical analysis they require of students to their abilities and fears. Keeping early examples and exercises simple built students' confidence for subsequently tackling more difficult concepts later in the course.

Supplemental campus resources

Instructors agreed that they and their teaching assistants bore primary responsibility for teaching their students to work with quantitative data. Even so, they often encouraged students to supplement their classroom instruction with outside resources. The two resources that instructors most frequently suggested to students were internet tutorials and campus resources. Instructors believe that students seek out the former on their own, but use the latter reluctantly, and only after prompting from faculty.

Students seem to value open web resources for reasons similar to those that cause many faculty to turn to them—they can be accessed at any time and are specialized enough that students can reasonably hope to obtain help with even very specific problems. A political science professor, for example, indicated that many of their students use "YouTube to learn additional things that we don't teach them about R because it's a great way to quickly pick up the bits and pieces that you need to know." The wide variety of tutorials available on the web seem to be more useful to students than the more general subscription resource, such as LinkedIn Learning, that many universities make available to students.

In general, instructors find their students reluctant to use campus resources without extensive prompting, despite that students have access to a wide range of resources to help them learn software, research methodologies, statistics, and data analysis. Perhaps the most useful of these are campus writing centers, data labs or camps, statistics consulting services, and library services. Some instructors do not believe that students are aware of these resources, but even those who regularly tell students about them and encourage students to use them had little sense that students regularly did so. In this respect, student behavior seems to closely mirror faculty behavior. Students seem more likely to take advantage of extracurricular opportunities such as internship programs and the growing number of student clubs that host or participate in data science competitions. What students learn from those experiences is difficult to judge based on our interviews, but instructors associated them with improved student performance.

Conclusion

Across the US, social science instructors are working to ensure that college graduates acquire data literacy skills and familiarity with quantitative analysis. These skills are widely considered to be important to twenty-first century civic life and workplace needs. However, teaching with data is difficult. Many social science students have deep-seated fears of math and limited experience with basic computer skills, let alone advanced programming languages like R. They also need to master the domain content and research methodologies that allow the social

sciences to make unique contributions to knowledge. To help students acquire the cluster of skills they need in order to understand, interpret, and manipulate data, instructors rely heavily on scaffolding assignments and active learning. As students advance from introductory to upper-level courses, they are increasingly asked to develop the capacity to independently locate useful datasets and to conduct ethical research of their own. Our interviews suggest that, despite the efforts of instructors, the transition from the simplified and carefully orchestrated encounters with data common in intro courses to the more open-ended and active task of creating their own knowledge is often very difficult on students.

Recommendations

Libraries
- Develop curated lists of tutorials and open web resources for both instructors and students.
- Offer discipline-specific data bootcamps or workshops aligned to specific undergraduate courses.
- Sponsor data analysis competitions in partnership with student organizations.
- Develop user-friendly guides to the library's data support resources for students and instructors.
- Prioritize licensing collections that include datasets suitable for use in undergraduate classes.
- Offer basic computing and software courses to ensure students have the foundational knowledge required for success.

University leadership
- Invest in the coordination of data support resources across campus.
- Fund the development of data fellowship programs in the library.
- Support the development of comprehensive, user-friendly, quick-reference guides to data support resources available to students across campus.
- Invest in the development and/or expansion of data labs, especially those focused on peer-to-peer learning.
- Increase investments in data-related extracurricular opportunities (e.g., clubs, competitions, and internship programs).

Departments
- Foster faculty peer-learning groups about instructional practices related to teaching with data.
- Continue to emphasize experiential learning and integrating undergrads into scholarly research projects.

- Explore opportunities to build more structured curricula to support the transition to conducting research and gathering data.
- Offer workshops for teaching assistants, either alone or in partnership with the teaching and learning center, on how to teach software skills as well as data cleaning, curation, and analysis.
- Work in partnership with the counseling center to identify strategies to alleviate students' anxieties around math.

Vendors

- Develop collections of social science datasets suitable for undergraduate use in consultation with faculty or create metadata in existing collections to assist with discovery of such material.

Appendix 1: Teams and Local Reports

American University
Symphony Bruce, Eric Schuler, Nikhat Ghouse, and Stefan Kramer

Boston University
Kathleen Berger and Ken Liss

Carnegie Mellon University
Emma Slayton, Hannah Gunderman, Melanie Gainey, and Ryan Splenda. "Understanding the Practices and Challenges of Teaching with Data in Undergraduate Social Science Courses at Carnegie Mellon University." https://doi.org/10.1184/R1/16693072.v1

Florida State University
Jesse Klein, Jeff Phillips, and Mallary Rawls. "Teaching with Data in the Social Sciences: An Ithaka S+R Local Report."
https://purl.lib.fsu.edu/diginole/FSU_libsubv1_scholarship_submission_1635530615_fad7b9e
a

George Mason University
Wendy Mann, Kim MacVaugh, Jasmine Spitler, and Andrew Lee. "Teaching with Data in the Social Sciences at George Mason University." http://hdl.handle.net/1920/12109

George Washington University
Ann James, Kelly Grogg, Megan Potterbusch, Shmuel Ben-Gad, and Yan He. "Teaching With Data in Undergraduate Social Science Courses at George Washington University." https://scholarspace.library.gwu.edu/work/t148fh96k

Grand Valley State University
Gayle Schaub and Samantha Minnis. "Ithaka S+R: Teaching with Data in the Social Sciences." https://scholarworks.gvsu.edu/library_sp/69

Kansas State University
Laura Bonella, Emily Finch, and Kendra Spahr. "Teaching With Data in the Social Sciences at Kansas State University: How Can K-State Libraries Support Undergraduate Instruction?" https://hdl.handle.net/2097/41696

Michigan State University
Scout Calvert, Andrew Lundeen, Shawn Nicholson, and Amanda Tickner. "Teaching with Data in the Social Sciences at Michigan State University." https://hcommons.org/deposits/item/hc:42419

North Carolina State University
Hilary Davis, Cynthia Levine, and Shaun Bennett. "Teaching with Data in the Social Sciences

Ithaka S+R study - NC State University Libraries Report."
https://doi.org/10.17605/OSF.IO/2X4CN

Purdue University
Jane Yatcilla, Clarence Maybee, Bethany McGowan, Gang Shao, and D. Trevor Burrows.
"Teaching with Data in the Social Sciences: The Purdue Report."
https://docs.lib.purdue.edu/libreports/6

Rice University
Lisa Spiro, Miaomiao Rimmer, Joseph E. Goetz, and Amanda Thomas. "Teaching with
Quantitative Data in Undergraduate Social Science Classes at Rice University: An Ithaka S+R
Local Report." https://hdl.handle.net/1911/111595

University of California Santa Barbara
Renata Curty, Rebecca Greer, and Torin White, T. "Teaching Undergraduates with Quantitative
Data in the Social Sciences at University of California Santa Barbara: A Local Report."
http://dx.doi.org/10.25436/E2101H

University of Chicago
Barb Kern, Elizabeth Foster, Emily Treptow, and Julie Piacentine

University of Massachusetts – Amherst
Erin Jerome and Stephen McGinty. "Teaching with Data in the Social Sciences at the University
of Massachusetts Amherst. https://doi.org/10.7275/py7y-yc52

University of New Hampshire
Patricia Condon, Eleta Exline, and Louise Buckley. "Teaching with Quantitative Data in the
Social Sciences at the University of New Hampshire: An Ithaka S+R Local Report."
https://dx.doi.org/10.34051/p/2021.39

University of North Carolina at Chapel Hill
Michelle Hayslett, Angela Bardeen, and Kayla Olson

University of Richmond
Samantha Guss and Ryan Brazell. "Teaching with Data in the Social Sciences at the University
of Richmond." https://scholarship.richmond.edu/university-libraries-publications/45/

Virginia Polytechnic Institute and State University
Liesl Baum, Julia Feerrar, Kayla B. McNabb, and Nathaniel D. Porter. "Teaching with Data in
the Social Sciences at Virginia Tech: An Ithaka S+R Local Report."
http://hdl.handle.net/10919/105135

Washington St. Louis University
Jennifer Moore, Christie Peters, Dorris Scott, and Jessica Kleekamp. "Teaching with Data in the
Social Sciences at Washington University in St. Louis: An Ithaka S+R Local Report."
https://openscholarship.wustl.edu/lib_papers/31

Appendix 2: Methodology

Participating institutions created individual research teams to conduct research on their respective campuses. The teams were composed mainly of library faculty and staff, primarily subject liaison and data librarians, though some institutional teams included staff from centers for teaching and learning and other university units. Each team was trained by Ithaka S+R to conduct semi-structured interviews with social science instructors about their instructional practices, using a shared interview guide developed by methodological experts at Ithaka S+R (See Appendix 3 for the interview guide). Due to ongoing disruptions caused by COVID-19, interviews were conducted remotely.

Each local research team conducted approximately 11 interviews, with a range of seven to 16. Together, they completed a total of 219 interviews, resulting in a high volume of data collection from instructors from a range of career stages and a wide range of disciplines within the social sciences. While Ithaka S+R provided a suggested list of disciplines or departments to consider for the study, universities define the social sciences in many ways. Each local research team used their discretion when recruiting individuals for interviews, providing them with flexibility to match institutional definitions and to interview instructors from interdisciplinary departments or majors, such as women's, gender, and sexuality studies or African American studies, who teach social science methodologies in their courses.

After completing their interviews, project teams coded their anonymized interview transcripts using a grounded coding approach to produce local reports encapsulating their findings and recommendations to campus leaders. Ithaka S+R provided support and guidance throughout the coding and report writing process, including arranging for the opportunity for participating institutions to engage in open peer-review by sharing their draft reports with colleagues from another participating university. Many of the local reports produced during the course of this study are now publicly available, although local teams had the option of keeping them private if that choice was conducive to consensus building at their institutions. These local reports are important complementary resources to this capstone report. Appendix 1 includes links to the reports that local teams chose to make publicly available.

Ithaka S+R developed a sample of 40 transcripts from the 219 interviews conducted by project teams. Tables 1 and 2 summarize our sample, which was developed to be representative of the overall population's distribution as measured by subject field, academic rank of the interviewee, and institution. The sampled transcripts were coded for analysis in NVivo using a grounded approach. In some instances, key word searches were conducted across the entire data set to explore specific topics in greater depth.

Table 1: Academic Rank of Interviewees Included in Sample

Academic Rank	Percentage of Sample
Assistant Professor, Tenure-Track	23%
Associate Professor, Tenure-Track	30%
Professor, Tenure-Track	23%
Non-Tenure-Track	20%
Grad Student	3%
Other	3%
Total	**100%**

Table 2: Departmental Affiliation of Interviewees Included in Sample

Department	Count
Anthropology/Archaeology/Classics	8%
Business/Leadership Studies	3%
Communication	8%
Economics	13%
Education/Information Sciences	3%
Geography/Geospatial Sciences	8%
Government/Public Policy/Planning	8%
International Affairs	5%
Linguistics	3%
Political Science	10%
Behavioral Sciences	13%
Sociology	15%
Statistics	5%
Unknown	3%
Total	100%

To protect their privacy, the identities of interviewees were not shared with Ithaka S+R, and they remain anonymous in this report. However, we thank the interviewees for their participation. Above all, we wish to thank the 20 project teams that took part in this research, without whom this report would not have been possible.

Appendix 3: Semi-Structured Interview Guide

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 preexisting 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 LMS, 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?