Teaching with Data in the Social Sciences at the University of Chicago: An Ithaka S+R Local Report

Elizabeth Foster
University of Chicago Library, ehfoster@uchicago.edu

Julie Piacentine
University of Chicago Library, jpia@uchicago.edu

Emily Treptow
University of Chicago Library, etreptow@uchicago.edu

Barbara Kern
University of Chicago Library, bkern@uchicago.edu

INTRODUCTION

In 2020-2021, the University of Chicago Library worked with Ithaka S+R, along with 20 peer institutions in North America, on the 'Teaching with Data in the Social Sciences' research study. This study involved teams of librarians at each of the participating institutions interviewing faculty in the social sciences to "examine social science instructors' practices in teaching undergraduates with data in order to understand the resources and services that instructors... need to be successful in their work."[1] The focus of the study was on undergraduate-level instruction with data, "student engagement with data that is tied to classroom instruction," and "teaching in which direct student engagement with data is built into the course structure."[2]

The goal of the study was to "contribute to the wider field of library and information studies, information literacy pedagogy, and the scholarship of teaching and learning in the social sciences, within the context of the evolving relationship between libraries and undergraduate teaching support."[3]

In this report, we cover four primary themes, with an overarching focus on undergraduate pedagogy, including: data discovery and creation, using data for teaching and research, undergraduate course and assignment design, and training and support opportunities.

[1] Ithaka S+R, "Ethics Review Instructions," Unpublished documentation for Teaching with Data in the Social Sciences, (Ithaka S+R, 2020).
[2] Ithaka S+R, "Project Scope and Recruiting," Unpublished documentation for Teaching with Data in the Social Sciences, (Ithaka S+R, 2020).
[3] Ithaka S+R, "Ethics Review Instructions."

At the University of Chicago Library, librarians engage with faculty, students, and staff about their research, teaching, and learning routinely, and this includes conversations around data. The Ithaka S&R study was, however, an excellent opportunity for us to conduct a deeper dive exploration into teaching data in the social sciences and to do so alongside colleagues at peer institutions. As the University of Chicago Library looks to further its role in data exploration, creation, dissemination, and use in research, teaching, and learning across all disciplines, this study and the insight it provides will help guide current and future strategic thinking and planning.

Social Sciences at the University of Chicago

At the University of Chicago, the Social Sciences Collegiate Division (SSCD) offers 16 programs that "immerse students in the full range of social scientific inquiry, exploring the conceptual frameworks, theories, and methodologies essential to understanding the economic, political, cultural, and psychological phenomena by which human communities organize themselves." While the SSCD was founded in 1930, individual departments date much earlier, such as the Department of Sociology, the Department of Political Science, and the Department of Economics, which were founded in 1892. Other majors include Chicago Studies, Comparative Race and Ethnic Studies, Computational Social Science, Public Policy Studies, and more.[4]

The cornerstone of the University of Chicago undergraduate curriculum, known as the Core, seeks to prepare students to explore and interrogate complex ideas as members of the University community. Course sequences across disciplines – humanities, social sciences, physical sciences, and biological sciences – and across the historical development of societies provide opportunities for students to engage with languages, concepts, and methods before engaging in specialized study within their majors. The Core curriculum is a hallmark of the University of Chicago undergraduate experience and an influential model of liberal arts education.

Participants

The University of Chicago Library team conducted interviews with ten faculty members from the Social Sciences Division. Participants held a range of position titles across four disciplines: Economics, Political Science, Sociology, and the Committee on Geographical Studies. No additional demographic information was collected. The interviews were conducted via Zoom.

ANALYSIS

[4] "Social Sciences Collegiate Division | The College," The University of Chicago, accessed October 1, 2021, https://college.uchicago.edu/academics/social-sciences-collegiate-division.

Data Discovery and Creation

Locating Data

Participants reported vastly different approaches to teaching data discovery and perceptions of students' need for instruction in this area. No one approach emerged as the most common. Some participants provide brief but detailed in-class demonstrations of data discovery, which may be expanded upon by teaching assistants in lab sessions or office hours. Others revisit data discovery instruction in multiple class meetings. Participants note the value of using multiple examples such as three options for accessing Census data to accommodate different types of learners. Some teach both quantitative and text-based data discovery.

There were several strategies identified to facilitate student discovery of data. For one participant, using Google to identify and locate data was a successful strategy. Another puts a dataset for the students to access onto the course site within the University's learning management system, Canvas. Yet another provides a list of commonly used data sources on their web page for students to review and use. One participant pointed to Google Dataset Search, which they feel is a particularly good resource for discovering data.[5]

The classroom, whether in person or remote, provides opportunities to walk students through data discovery and downloading steps. Participants describe guiding students to websites in real time, showing them where to find data, and demonstrating how to download it. One participant points students to a particular dataset, but because of its size and complexity, will help students extract what they need and point them to key components, such as the code book. In this case, the students are actively engaged with the participant, who tells them, "Let's work together to set up a system by which you can extract the data." Some participants describe taking additional steps to make data easier for students to use. One shared an example in which they said,

> I'll clean it up for them so that there's no surprises and I'll deal with missing values and so forth. And so I'll kind of eliminate that first step for them and then just post those kind of clean files for them to use with their problem sets.

One participant described a progression during class from providing datasets to students ultimately finding their own. Some students opt to go to the source themselves to work with additional variables or simply to better understand the unmediated process. One participant

[5] "Dataset Search," Google, accessed October 1, 2021, https://datasetsearch.research.google.com/.

commented that “sometimes students go and get the original data with more detail, with more richness.”

Multiple participants described how information overload complicates students’ data discovery. Participants described making careful choices about course content, course materials, and assignment design to increase clarity. Information overload may also impact students’ use of library guides and resources. For example, one of the participants acknowledged that “no one person will know all of the data sources or data options available,” and another suggested that they actively identify only a small number of datasets, presumably not to overwhelm students with too much information. Another issue identified by a participant was the ease of using relatable versus unfamiliar datasets, stating that, “…. people like me who’ve never used a platform like [a stock trading app], it would take me days if not weeks to fully understand what is in the data.” In contrast, a bike sharing dataset used in class consists of data points that are familiar to more students: departure and destination stations, ride start and end times, etc. This enables students to engage more readily with the data and confidently hypothesize relationships between variables.

Part of data discovery includes being familiar with data hubs and repositories. This can entail a variety of strategies. One participant mentioned finding data hubs through their Twitter network. There was also an acknowledgment by one participant that data repositories are not necessarily as much the answer to data discovery as they had been in the past, stating, “Things are not held in neatly bound portals anymore,” while another commented that “sometimes some of the best datasets are ones that are not in these huge, big collections that are put together by academics and released publicly.”

Participants acknowledged that data discovery is not the same experience for all students, depending on the knowledge and skills they bring to the class. For one participant, the goal is to “level the playing field, as it were, across different kinds of skills” by pointing students to specific sources of data, creating a dataset to share with students, or having the students create datasets of their own. In the end, for this participant, about half of the students in the class are using existing datasets and the other half are collecting their own data. A different participant finds that datasets in some fields are plentiful and finding them is not difficult.

Some of the participants provide direct access to datasets. During the interviews we heard participants say, “I usually direct them,” or the students are “all directed to the same dataset, and so they have to do it themselves,” or simply, “I will provide datasets.” For one participant this made good sense since the dataset they point students to is “one that I know will work.” One participant shared that their approach is to focus on instructing students how to go about discovering data followed by sharing specific examples. One participant has students focus on

data related to one topic “because it helps to kind of keep everybody on the same page rather than having the students all kind of go out and get their own data.”

Participants also recognize that students need to work with different types of data formats. One faculty member specifically provides students with several different kinds of datasets, stating that they want students to “start learning how to work with different types of data formats because a lot of what I work with, the data might be available in, like, five or six different formats.”

Participants’ understanding of the Library's role in curating and facilitating access to data is unclear. Participants recognize and use library guides and resources, both free and subscription, for teaching with data in the social sciences, which is promising. However, decentralization and overlap across campus units create challenges. Participants report broad awareness of and satisfaction with library guides to data resources. However, library web pages foreground database names but obscure their descriptions, use cases, and other useful details. While participants value library resources, they see students struggle to make effective use of them. Students may be deterred by small barriers to entry or may require external motivation to explore library resources.

> Very often, 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.

Participants report that the library is a great source for finding datasets but, because students are interested in specific events or policies, they may seek more tailored datasets than the ones found in academic resources. This leads students to rely on administrative or crowd sourced datasets.

Creating and Accessing Data

Few participants teach students how to create data that can be used for their studies or projects. One participant noted that some students, but not all, may know how to scrape the web. Even in upper-level and capstone courses, in which students are motivated to conduct novel research, this level of knowledge does not appear to be taught programmatically. That said, one participant implored colleagues to make data collection a part of their instruction:

> I really encourage instructors to try to get their students to do their own data collection and cleaning because that is ninety percent of the actual work and the thing that is the big barrier between them and doing new research. If they just go to the ANES or to ICPSR and download a pre-existing dataset or just do the R packages and do a pre-existing dataset, they’ll be completely lost when it comes to coming up with new projects. The

> experience of finding and obtaining and cleaning and discovering new data is both a positive thing and a necessary thing for them to know, to become users of this sort of material. So I encourage instructors everywhere to try to incorporate that aspect into the class even though it is the most difficult and the most painful and the most challenging part of the entire process. Don't skip it. That does mean – I'm not saying they should be doing this in a statistics course. It should be done in a course, not a statistics course. Like in statistics you need to be learning statistics but if you're doing data analysis that's the most important thing you should be doing is getting data.

Other participants identified challenges associated with students collecting their own data, such as the learning curve for tools like SurveyMonkey and Qualtrics. One participant stated that both are "cumbersome and neither of the companies that sponsor those really make it straight forward for students to be able to construct their own studies." It is not clear from this study whether instructors give students the baseline information they need to use such programs for data collection. These tools can also be cost prohibitive.

For some participants, pointing students to data created at the University is important. For one participant, data used in class came from their own departmental or lab data repository, created by individuals at the University that the participant felt were trustworthy. Another trusted resource was the General Social Survey created by NORC at the University of Chicago.

It is customary for students to use or modify existing data. This may be public use or proprietary data. Collecting data is more common in advanced courses and usually involves creating surveys, conducting interviews, or scraping the internet.

Participants, who hold extensive prior knowledge and experience, can overestimate students' facility for finding and accessing datasets. Participants may direct students towards data sources or repositories, list resources on students' topics, or provide data for student use. To ensure that datasets are suitable, some participants ask students to submit their dataset choices ahead of time. BA thesis students may receive additional mentoring from their advisor in locating data.

Data access is not challenging if students are using public use data. Accessing proprietary data and library subscriptions can be challenging due to additional requirements that participants may find easier to navigate than students do. These barriers include using the VPN, determining how to obtain data, and understanding permissions and rights. Additionally, bureaucratic obstacles, such as permissions and rights, prevent students from obtaining secondary or restricted data. The Library may have an important role to play in this by purchasing and facilitating access to data that undergraduate students need for their research and learning.

Finally, strategies for locating and using data can be discipline-specific, which may create additional challenges. For participants, data discovery practices within their discipline may be clear. This may not be true for students who are just learning disciplinary practices.

Using Data for Teaching and Research

Data Tools and Technology

Participants shared a wide range of approaches to teaching software. The primary tools taught by participants are Stata, R, and Excel. Some participants provide detailed, guided demonstrations over a single session or multiple class meetings. Some provide demonstrations during lecture and split hands-on software instruction into TA-led lab sessions or office hours. Regardless of approach, participants described not wanting students to get "bogged down in programming rather than engaging in social science inquiry." TAs were often described as the first line of defense for software and troubleshooting questions.

The ease of using software can influence participants' decisions regarding which software to adopt in a class. The University's remote PC system, vLab, which provides students access to Stata and other key data analysis tools, is antiquated and problematic. Participants avoid using it whenever possible due to its propensity for deleting students' files and being slow, especially when used with large datasets. This is a major issue because vLab is the only place where students can access Stata through a campus license.

Most participants prefer teaching Stata, with one noting that the syntax is simpler: "...[W]hat I want is [for] students to feel how easy it is to carry out empirical analysis." Some participants have considered asking students to purchase Stata rather than rely on access via vLab. One participant reported growing interest in teaching with R in their department. R is open source, but has a high learning curve. One participant noted that alignment regarding which tool is taught across all quarters of a course sequence would be ideal.

As with data discovery skills, students have a range of abilities with coding and the use of data analysis tools. This makes course design challenging. Participants typically assume that students have no prerequisite knowledge. Students may default to using familiar tools like Excel, which may make some tasks such as data cleanup easier. Still, participants steer students to Stata for this and other functions because the work is then documented and systematic. Many participants note that software skills are transferable and that learning any one specific program or language is not a central concern. Some participants disagree, noting that students may face employer expectations to know a specific tool. Two participants pointed to employer expectations to know

Stata. Other participants report an expectation for students to have prior knowledge of R, particularly in graduate programs.

Interviews indicate there is a need to centralize data analysis resources in one place for students. The lack of software licenses for students and the brevity of trials is a hindrance to using software in courses. Centralized support would remove this barrier. An additional challenge is cost. For example, there are some great toolkits in Python, but this would mean adding funding for toolkit creation into research grants.

Additionally, there are the technical challenges of using specific software and knowing the right steps for different processes. Participants reported that some students are teaching themselves software and data tools using online courses; however, this is not across the board, creating inconsistent expertise in the classroom.

Undergraduate Course and Assignment Design

Instructors' Approaches to Prioritizing Course Content

Fundamentally, doing social science requires thinking critically about sources and uses of data. This is the crux of teaching with data in the social sciences. Students must learn how to think about problems and how to adapt their knowledge and skills to a variety of problems. To do this, participants teach students to interpret data by examining metadata, finding patterns in descriptive statistics, conducting statistical analyses, and evaluating which statistical analyses add meaning and which do not. Participants balance conceptual and practical course components over the span of ten-week quarters:

> [A]n instructor can have a hard time figuring out what to skip. There's always a time constraint problem. Time is a scarce resource, so what do you leave out? What do you put more emphasis on?

Participants grapple with the level of importance that they should place on coding skills and no approach emerged as most common. Some participants consider coding skills essential but do not want their courses to focus on coding alone; other participants explicitly emphasize analytical thinking skills over coding:

> But I don't want the students to get bogged down in programming, and this is one of the problems with working with something like Python and R is Python and R are powerful software and there are advantages to students learning about them. But it becomes

> extremely focused on the intricacies of programming rather than actually doing the thing that we care more about which is social science inquiry.

Along these lines, several participants spoke about the struggle with balancing different skill levels in one course:

> I'd say the range of abilities is what makes it challenging and that you have to kind of design instruction that's effective for someone that's already an expert at Stata or an expert at R and also someone that's never opened Excel before or doesn't even know how to navigate a Windows directory or file path. And so when you have that kind of variation and abilities, it's challenging to design a class that's effective for everyone, definitely.

A few participants design problem-based learning that engages students in topics of interest and may integrate the participation of community members and organizations:

> [I]ncreasingly I find that bringing in people from the community to kind of talk about some of the data challenges there – so what they need – and then also provide some kind of expertise or critique on what the students come up with is really helpful because that's something that is, rather than talking about communities, like, actually bringing people in from communities can be helpful in how we think about data.

In contrast to this applied approach, one participant shared that students are interested in looking at data through the lens of theory.

One participant specified that they focused their teaching to engage with students based on the major most represented in the class. Although several others noted that students are very creative and are invited to choose which research questions they want to answer. Additionally, a participant noted that their students appreciated seeing connections between faculty research and what they were learning in class.

As previously noted, the undergraduate Core curriculum is a distinguishing feature of teaching and learning at the University of Chicago. For some participants, this difference from other institutions is noteworthy. The University of Chicago's Core curriculum provides an additional foundation in inquiry that can be applied to data and social problems. Building on this strong general foundation, a new quantitative social analysis minor was developed to meet additional quantitative instruction needs. Because the minor is so new, its impact on quantitative skills has not yet been evaluated. Some duplication of content across the curriculum is necessary because different disciplines emphasize different aspects of data-intensive research. However, some course content is probably redundant and could be better organized across majors.

In the social sciences inquiry Core course sequence specifically, students learn basic statistical analysis and how to draw inferences from data. It is a foundation for learning more advanced topics. In other courses, participants teach students to produce a variety of deliverables, including maps, simple charts and graphs, more complex visualizations, linear regressions, and work-in-progress shared through GitHub repositories. Deliverables of all kinds help students demonstrate learning and showcase their skills.

Participants stressed the need for a secondary data course sequence before the BA thesis research methods course. Intertwining undergraduates' first independent application of their data skills with their first major independent research project – the BA thesis – creates pressure for data to prove an anticipated result. Participants noted that there is not enough time to cover everything students need to know before they begin their thesis research. There are few paths to develop skills because students must take courses in a specific order due to the duration of the Core curriculum. Furthermore, there may not be enough opportunities for students to learn skills within their departments. In an attempt to solve this issue, one participant opened a graduate level course to interested undergraduates. Participants need to find ways to provide additional training within their departments.

Teaching Data Skills from Beginning to Advanced

While some courses introduce students to discovering data, creating data, or both, participants described courses which integrated little to no use of data. These courses introduced students to evaluating data as evidence, writing about models, incorporating tables or charts into their writing, and learning theory. In an ideal world, students will leave introductory courses with basic knowledge on how to program and use analysis software. Not everyone will gain advanced knowledge, but they will be able to perform foundational tasks. A lack of experience with prerequisites that are not covered in the course can be challenging for students.

Interviews demonstrated that participants have deep expertise working with social sciences data. This can present participants with a challenge to identify their own tacit knowledge and develop awareness of their students' expertise, which cannot match their knowledge. With awareness, participants can develop instruction to address skills gaps, such as teaching advanced search syntax in Google to find data.

Skills that move students from beginner to advanced learners include:

- Creating their own or adding variables vs. using an existing dataset: Students actively create new knowledge by adding new variables, such as the distance between data points in two datasets.

- Cleaning data since students will frequently encounter messy datasets: When students lack the ability to clean data, they face barriers to doing new research. Students who lack the skills to work with data have to rely on clunky web interfaces.
- Creating appealing, meaningful, and informative data visualizations: Critiquing visualizations from the media helps students identify elements of good data visualizations. Learning how to visualize data helps students showcase their work. Visualization is also important for checking datasets for issues before conducting analysis.
- Learning how to make decisions with data, a key element of learning data analysis: Many tasks require students to clean, manipulate, analyze, and interpret data. Faculty teach students how to navigate the decision-making process.

Bridging Varied Student Backgrounds and Cultivating Student Engagement

Many participants spoke about the role student confidence plays in teaching with data in the social sciences. One reported that students' confidence in their own capability is critical to overcoming their first problem, which then reinforces their confidence and contributes to their ongoing success. Related to confidence, a participant shared that students conflate standard data science practices such as using snippets, functions, and documentation with cheating. Multiple participants described designing assignments to offset differing levels of confidence and expertise. Sequenced assignments with defined tasks require students to evaluate data and complete projects in the same way. This reduces the advantage of being familiar with the material prior to taking the class. Some participants increase clarity by providing resources, such as handouts with commands, instructions, and specific data sources to use. Providing data also helps students focus on empirical methods and bypass locating and cleaning data. Five study participants described additional strategies used to empower students: working through process steps together, reflecting after each step, reminding students that frustration is a sign of learning, and assuring students explicitly of their ability to succeed.

One participant finds that students may perceive practical training as a requirement to be met rather than a rich learning experience. On the other hand, numerous participants find that emphasizing practical applicability and career skills can activate students' motivation to learn.

Many participants emphasized the importance they place on teaching students to interrogate data. Participants describe teaching students to find patterns in data, whether through regressions or other statistical analyses, and interpret those patterns, but also to interrogate how concepts such as neighborhoods, conflicts or war, and unemployment are abstracted into data. Determining which data are best suited to address a specific problem, and how different methods of analysis

may hide or highlight different interpretations of given data, provides students with a foundation to "assess the weight of evidence behind causal claims."

Interviewees reported that students are interested in applying their data skills for community service both within and beyond their coursework. Other ways to engage students include teaching with datasets related to topics of interest or to which students can relate, such as bike sharing data, data related to structural racism, or data which participants have used in their published research. Guest lectures are another strategy to enhance student engagement. Furthermore, coursework and extracurricular work with data can build on each other. Data skills acquired through coursework help students gain internships, research assistant positions, and other extracurricular opportunities, which then enable students to acquire more advanced skills.

Inclusive teaching practices described by participants ranged from specific measures like ensuring the accessibility of visualizations for students with visual disabilities to broad measures that foster a more supportive environment for students from underrepresented backgrounds and groups. One participant noted, "There are many women and students of color who think, 'This isn't my thing.' And it *is* their thing," adding that such barriers can be overcome by creating a supportive environment.

Participants may expect and accept that students will achieve different skill levels through a course, some acquiring fluency to write code, others gaining familiarity with basic programming functions, such as importing data and saving work. Group work can make it hard for participants to gauge individual student skill levels.

Learning with peers to develop varied data and software skills as a part of group projects is important. One participant noted that this even extends to formal student organizations, such as Oeconomica, an undergraduate economics research society.

Participants noted varied student experience and expertise. One participant compared learning coding languages to learning another spoken language, a parallel which they felt could be helpful for ESL students. Another noted that there were fewer issues with basics with their undergraduate students compared to other institutions, suggesting that students may be coming to the University with greater exposure or experience with the skills they need for working with data.

Training and Support Opportunities

For Students

Thinking about training and support opportunities for students, one participant thought that online training tools, such as LinkedIn Learning are, for the most part, ineffectual. This participant thought that in-person workshops and institutes would allow students and faculty to learn skills in a more meaningful way.

More specifically, students need opportunities, whether online or through campus resources and workshops, where they can learn the basics of spreadsheets, Excel, and statistical analysis. One such opportunity could be participation in summer institutes currently offered by departments at the University. There are departments at the University of Chicago that run summer institutes for learning relevant skills and networking with other students. Additionally, students can get assistance via the Department of Statistics, but it is not always the right level of support and lacks important disciplinary context. One participant noted that students had to go outside their major to get certain training. They stated it would be better if the major could provide training in-house.

For Instructors

Overall, participants expressed a willingness to share their teaching materials and there is a culture of sharing class materials (and being asked to share) at least in certain departments. A few participants noted not wanting to share problem sets because they felt they would need to create new ones if those were used by others.

As noted earlier, there was interest expressed in turning teaching materials into a textbook which the Library would encourage to be open access. The participant noted that they would like support in doing this which might be an important role for the Library.

In terms of training, a participant noted that they felt like leaning through their network was enough for them. Another expressed interest in formal software training but noted that time was an issue.

Working with Librarians

Study participants diverge in their perceptions of the role of librarians. Some participants collaborate with librarians on instruction and encourage their students to seek consultations with librarians for help. Others, due to the nature of the assignments they design, see little or no role for librarians to play. Some see students' needs related to data as so discipline-specific that a librarian who does not conduct research within the discipline would be ill-equipped to provide support. One participant overestimated the library staffing available to support teaching with

data. One participant noted that the library could serve as a hub for locating data services on campus.

Campus IT Support

Software support is essential. At the University of Chicago, drop-in support is available via the "TechBar" help desk, located in the largest campus library, Regenstein Library. Having this type of support allows faculty and TAs to focus on other essential elements of teaching and learning.

CONCLUSION

The undergraduate Core curriculum provides key context for teaching with data in the social sciences at the University of Chicago. The Core equips students to engage in rigorous inquiry but restricts the time available for advanced, disciplinary course sequences and applied skills instruction. Faculty teaching with data in the social sciences make careful decisions about course design and pedagogy in order to achieve a balance between teaching concepts and theory versus applied coding and data analysis skills.

Findings from this study will inform new and continued Library initiatives to support undergraduate teaching with data in the social sciences. Recommendations include, but are not limited to:

- Normalize the range of experiences and expertise that students enter courses with by expanding general data workshops and ensuring all students have a baseline knowledge of specific software and resources.
- Address known knowledge gaps throughout the school year by developing extensible instruction that highlights, as examples, core datasets and data sources, teaching advanced web search syntax for data discovery, and clarifying academic integrity as it relates to data.
- Ensure equal access to software and resources by students through continued partnerships and conversations with departments and offices on campus about software licenses.
- Create tailored training opportunities for teaching assistants as well as students, allowing faculty to focus on specialized, disciplinary instruction.

We are looking forward to implementing these recommendations in the near future and to the Library continuing to be a strong and active partner at the University in teaching with data in the social sciences.

Acknowledgements

Thanks to Elizabeth Edwards, Assessment Librarian at the University of Chicago for her assistance with this project.

Bibliography

Google. “Dataset Search,” Google, accessed October 1, 2021, https://datasetsearch.research.google.com/

Ithaka S+R, “Ethics Review Instructions,” Unpublished documentation for Teaching with Data in the Social Sciences, (Ithaka S+R, 2020).

Ithaka S+R, “Project Scope and Recruiting,” Unpublished documentation for Teaching with Data in the Social Sciences, (Ithaka S+R, 2020).

University of Chicago. “The Core Curriculum | The College,” Chicago: University of Chicago, accessed October 1, 2021, https://college.uchicago.edu/academics/core-curriculum

University of Chicago. “Social Sciences Collegiate Division | The College,” Chicago: The University of Chicago, accessed October 1, 2021, https://college.uchicago.edu/academics/social-sciences-collegiate-division.

Appendix I: Semi-Structured Interview

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

Background

Briefly describe your experience teaching undergraduates.

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

Getting Data

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

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

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

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

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

If students work with datasets the instructor provides Describe the process students go through to access the datasets you provide. *Examples: link through 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?

Appendix II: Recruitment Email

Subject. University of Chicago Library's study on teaching with data in the social sciences

Dear [first name of instructor],

The Library is conducting a study on the practices of social science instructors in order to improve support services for their work. We are interviewing instructors whose undergraduate students engage with quantitative data, such as by conducting research using quantitative methods, analyzing, or visualizing datasets, or learning to use specific tools or software to work with data.

Would you be willing to participate in a one-hour interview to share your unique experiences and perspective?

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

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

Sincerely,

Barb, Elizabeth, Emily and Julie

Appendix III: Resources Named in Interviews

Core Coding and Analysis Tools

- Stata
- R
- RStudio
- Excel

Other Coding and Analysis Tools

- Fortran
- MATLAB
- PostgreSQL

- Python
- SAS
- SPSS
- SQL

GIS Tools

- ArcGIS
- ESRI
- FME
- GeoDa
- QGIS

Census Tools

- Census Bureau APIs
- National Historical Geographic Information System (NHGIS)
- Social Explorer

Survey Tools

- Census Bureau APIs
- National Historical Geographic Information System (NHGIS)
- Social Explorer

Other Tools

- ABBYY Finereader - Optical Character Recognition (OCR) software
- DataCamp - Data analysis training resource
- Foreign Affairs - Journal
- GitHub - Code and data repository
- vLab - University of Chicago's virtual/remote computer lab

Core Teaching Datasets and Data Sources

- American National Election Studies (ANES)
- Bureau of Economic Analysis
- Correlates of War
- Eurobarometer
- Federal Reserve Economic Data (FRED)
- Foreign Relations of the United States
- Gallup
- General Social Survey (GSS)
- Integrated Public Use Microdata Series (IPUMS)

- Inter-university Consortium for Political and Social Research (ICPSR)
- International Monetary Fund (IMF)
- Organisation for Economic Co-operation and Development (OECD)
- Panel Study of Income Dynamics (PSID)
- The Harris Poll
- UN Human Development Data Center
- Uppsala Conflict Data Program (UCDP)
- World Trade Organization (WTO)
- World Values Survey
- World Bank

Other Datasets and Data Sources

- Airbnb
- Chicago Bears schedule
- Chicago Police Reports Dataset
- Chicago Public Schools
- CNN
- *Companion to Applied Regression Analysis*, Car Package, Prestige dataset
- Compustat
- CQ Magazine
- Divvy
- Freedom House
- Gravity database
- HathiTrust
- Iris dataset
- National Security Archive
- New York Times COVID dataset
- Penn World Tables
- Pickle Packers International, Inc. website, ilovepickles.org
- Regulation.gov
- Supreme Court Database
- The Field Poll
- Open Secrets
- Twitter
- USAID Demographic and Health Surveys (DHS)
- Yelp
- YouTube