11-15-2021

Teaching with Data in the Social Sciences at the University of Richmond

Samantha Guss
University of Richmond, sguss@richmond.edu

Ryan Brazell
University of Richmond, rbrazell@richmond.edu

Teaching with Data in the Social Sciences at the University of Richmond

Samantha Guss, Social Sciences Librarian, Boatwright Library
Ryan Brazell, Technology Consultant, Pedagogy & Scholarship, Faculty Hub

November 15, 2021

Table of Contents

Boatwright Memorial Library Faculty Hub

Executive Summary

Study Overview

From Spring 2020 through Fall 2021, a team from UR participated in a multi-site study called "Teaching with Data in the Social Sciences" led by Ithaka S+R, a research and strategy organization that focuses on scholarly communication and libraries in higher education. Samantha Guss (Boatwright Library) and Ryan Brazell (Faculty Hub) interviewed 14 UR faculty, all of whom teach in social sciences disciplines or use social data, to learn more about faculty needs as they help their students build data literacy skills. Although this study focused on social sciences, we believe that many of its principles can be applied more broadly, especially considering the work to reform the General Education Curriculum that is currently underway. The primary objective for participating in this study was to better understand UR faculty needs so that the Library and Faculty Hub can better support faculty who teach with data across the curriculum. Our findings are presented below in four themes.

Context

Boatwright Library has a well-established information literacy program, through which our research and instruction librarians partner with faculty to provide instruction sessions and help design assignments to achieve information literacy learning outcomes. While information literacy is built into UR's first-year seminar (FYS), those courses are certainly not the only ones that benefit from librarian-faculty partnerships. The Faculty Hub offers a variety of programs to support faculty development, including developing teaching skills and fostering community and connections across disciplines. The Hub currently offer pedagogical support for faculty who incorporate data education into their teaching, including data visualization and working with the new campus HPC cluster.

Findings

1. **Literacy Across the Data Lifecycle**
   Students are often exposed to data analysis, and occasionally to data preparation and data collection. They are usually not asked to engage with things like data archiving, re-use, and description, while these pieces are key to understanding data as information.

2. **Supporting Students & Faculty on the Margins**
   While some faculty and students are well prepared and well supported for data-intensive pedagogy, others "on the margins" are hindered by differing levels of previous experience, expectations, or adapting to new modes and methods. Providing support for faculty and students on the margins will be key to advancing data curriculum initiatives.

3. **Teaching Data Ethics**
   Faculty across disciplines emphasized the importance of teaching students about ethical issues in working with data, but also described their own efforts to make sure they were teaching in an ethical way.

4. **Curriculum Sequencing & Alignment**
   A recurrent theme was the challenge of teaching data skills to students in social science majors who aren't required to take courses in a specific order. This makes it challenging to get beyond basics in a given course.

Recommendations for Boatwright Library

Library support for data education and pedagogy should be integrated into the new general education curriculum, just as information literacy is currently integrated into the first-year seminar program. This means:

- Deliberately extending our existing information literacy to include working with data.
- Partnering with faculty to focus on what libraries and archives can best contribute; for example, the parts of the lifecycle that are not frequently taught by faculty (like preservation and documentation).

Recommendations for the Faculty Hub

The Faculty Hub should expand its existing programming to incorporate support for faculty who are teaching with data and not using the HPC cluster. This means:

- Coordinating a community of practice for data instructors more broadly.
- Developing materials and programming regarding data literacy instruction that take into account the specific needs of faculty and students who are "on the margins."
- Work with the Academic Skills Center, TLC, and other relevant campus departments to minimize gaps in existing support structures for data-related teaching.

Introduction

From Summer 2020 to Fall 2021, Boatwright Memorial Library and the Faculty Hub at the University of Richmond participated in Ithaka S+R's *Teaching with Data in the Social Sciences* research project as part of a cohort of 20 colleges and universities across the United States. This project aims to learn about the needs of social sciences instructors at the undergraduate level as they seek to help their students build data literacy skills. Ithaka S+R, a non-profit research organization focused on innovation and impact in higher education and the arts, provided the study design, coordinated cross-institutional cooperation, and will produce and publish an aggregate report in Fall 2021.

UR's participation is a collaboration between the Library and the Hub, with the intent of gathering knowledge to inform future services and strategic directions for both units. UR is the smallest institution in the cohort by student body size and the only one focused primarily on undergraduate education.

The project team at UR conducted 14 hour-long, semi-structured interviews with faculty members during the Fall 2020 semester. The faculty interviewed were drawn from social science departments across the university based on the researchers' prior knowledge, exploration of course and department listings, and recommendations from interviewees. These included faculty from the School of Arts & Sciences, the Robins School of Business, and the Jepson School of Leadership Studies, where our traditional undergraduate student population is based. While this is qualitative research and the sample isn't intended to be directly representative, the researchers purposely invited faculty from a variety of ranks (Professor, Associate Professor, Assistant Professor, Visiting Professor, Adjunct Professor) and from all social sciences departments. There are also several faculty members represented who do not consider themselves "social scientists," but teach on social science topics and/or with social science data and, through discussion with the researchers, agreed that they fit the criteria for inclusion in the study. The interviews were conducted via Zoom and followed an interview guide (see Appendix A) consisting of questions about how instructors and students obtain and access data; how they work with data (including things like software, pedagogy, and ethics); and on training and support for instructors. The interviews were transcribed and coded, with the goal of identifying 3-4 themes that describe challenges and opportunities for UR in the area of teaching with data in the social sciences.

Data Pedagogy Initiatives at UR

This investigation was conducted amidst a number of ongoing and newer data initiatives on the UR campus; this context was considered critical to our project team as we conducted our interviews. We outline some of these initiatives and support services here, but this list is not meant to be exhaustive:

- Staff in the Faculty Hub, and previously in the (now defunct) Center for Teaching, Learning, & Technology (CTLT), have been working on data education for several years by providing workshops, pedagogy consultation, and technology support. The Faculty Hub hosted two Faculty Hub Associates for 2020-2021, Dr. Kristine Grayson and Dr. Kristine Nolin, whose project aims to "advance faculty training and expertise in data-

centric teaching practices." (https://facultyhub.richmond.edu/programs/faculty-hub-associates.html) Their activities included a book club and an R workshop for faculty.

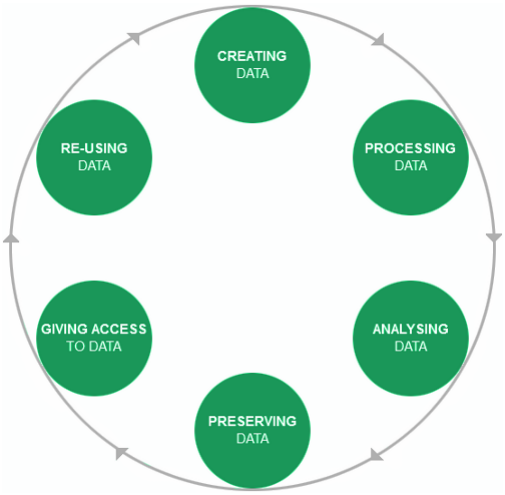
- Boatwright Library has had research and instruction librarians with data expertise for at least a decade. Librarians regularly consult with faculty and provide instruction sessions related to library-licensed data and on finding and using data. The librarians work through the liaison-librarian structure where each librarian works with designated departments and faculty, but one librarian also has specific responsibilities for data initiatives and often works with faculty across the university. Additionally, the Digital Scholarship Lab, which houses significant data expertise, is administratively part of the Library.
- The Data Analytics/Data Science Committee (university-wide; led by the Provost's Office) formed in early 2021 and has been meeting weekly ever since. The Committee thus far has worked on policies for UR's new high performance computing (HPC) shared resource, hired an administrator for the new HPC resource, discussed existing data programs across the university, and discussed strategies for implementing a proposal for a Data Science minor in the College of Arts & Sciences.
- The UR faculty are conducting a multi-year process to revise and update the undergraduate General Education curriculum. At the end of the 2020-2021 academic year, faculty voted to adopt the proposed "Web of Inquiry" curriculum, and it will now be sent to an implementation committee. The new curriculum requires all students to take a course with a "Quantitative Data Literacy" designation.
- The Spatial Analysis Lab (SAL) in the Department of Geography & the Environment (in the School of Arts & Sciences) has long been a leader in providing services and technology for teaching with data at UR. More recently, as demand for GIS services continued to grow beyond the department and the school, a steering committee was convened to advise on policy and strategy.

Structure of This Report

This report describes four themes that arose from the faculty interviews: **Literacy Across the Data Lifecycle**, **Supporting Students & Faculty on the Margins**, **Teaching Data Ethics**, and **Curriculum Sequencing & Alignment**. While these weren't the only themes that emerged, we chose to focus on themes that haven't been described or addressed elsewhere on our campus and those that seemed generally more actionable by those supporting teaching with data at UR. Some supporting quotations from faculty interviewees are included in the main text and Appendix B contains additional quotations. The report concludes with a Recommendations section.

Theme 1: Literacy Across the Data Lifecycle

Data lifecycle models are commonly used in the library and information science (LIS) and archives professions to describe actions taken by a researcher or by an archive throughout the "life" of a data set or other digital object. There are many versions of these models, but for this report we'll refer to one from the UK Data Service[1] that focuses on the activities of a researcher. (For further explanation of this model, see Appendix C):



Just as there are many data lifecycle models, there are also many definitions of "data literacy," and this is where we make the connection to teaching with data. If we understand data literacy for our students as a holistic understanding of how data are produced, analyzed, visualized, used, and re-used, a data lifecycle model like this one becomes a useful heuristic as we think about our curriculum and individual classroom goals.

We chose to introduce the lifecycle model in this report because it helps us categorize many of the teaching activities we heard about in our interviews, but also showed us some things we did not hear about as much (or at all).

Our interviews showed that the majority of teaching emphasis is on Analyzing Data. This is completely logical, considering a lot of data use in the classroom is for the purpose of learning concepts (and not just using data for its own sake). Data analysis is also at the core of faculty research and expertise, so it's logical that it would be taught as a highly important piece of the lifecycle.

> "I try to tell them this, that I'm going to give you a nice, beautiful data set that's ready to go. And the reason is my class is really focused on methods. The collecting data, cleaning data, and getting it ready is excruciating. It's 1, it's not fun, but it's just an entire process that takes so much more. And so I say, 'look, I'm going to-- I'm focusing here. I'm going to bring you guys two steps into the process. And then this is where we're going.'"

[1] "Research Data Management," UK Data Service, accessed September 29, 2021, https://ukdataservice.ac.uk/learning-hub/research-data-management/

Several faculty members told us that they also incorporate Processing Data and Re-Using Data in their classes; that they don't only give their students tidy, clean data and instead ask them to find data and combine and clean it themselves, just as a researcher or data scientist might do. This was regularly described by faculty as an intentional pedagogical choice designed to give students more authentic, "real world" data experience.

> "Typically in an intro course, [in] at least the introductory exercises or beginning of the semester, I provide [data ready to use]. So everything they have is there already. But I don't like to do that continuously because that's not going to happen to them in the real world. So generally I start giving [data] to them in pieces, in different formats, and having them compile them into [usable data format] for their particular project. So that's certainly part of the learning objectives I have for them is that being able to not only discover data, but be able to get it into a usable form."

Since these steps are so time-intensive though, many faculty members told us that they don't have the time to spend and instead give their students clean (or fake) data that is ready to use for analysis.

> "As far as collecting data, you know, especially at the Intro level, they're so limited and we don't have time to get through an IRB and getting things approved that in dealing with human subjects, obviously. So we definitely rely on those kind of existing data sources more."

> "You don't want them to spend all this time learning a method and then get these really weird, bizarre results. So you want [good] data and I'm not shy about this: I make up data all the time. Like I just simulate it. I make up variables and, you know, I try to keep those fairly benign and in terms of like any interpretation, but it's really helpful and you can still create context, but I think it's more helpful for learning how to do things and how to interpret things with a nice sort of ready-to-go dataset."

Re-Using Data shows up often in discussions about where data come from, how to interpret how they were collected, and what they mean.

> "The first thing I tell them is that you are looking at data someone else collected, not for your purpose. So you need to understand that. So you need to figure out where you're getting the data from, and it's not all great. Even the government data isn't that great. It will be incomplete in some ways, you know, some county may not report their [American Community Survey] number in the state that you're in. So, just to understand that the collection is a process that may not always go that great. And to just be okay with that and be upfront about it."

> "I think university-wide it's I think it's a problem that, you know, that we need to tackle just increasing students' information literacy in terms of interpreting results of studies and understanding how knowledge is produced."

A few others described incorporating the Creating Data step in their classrooms; for example, having students create and execute surveys or collect observational data. In every case where students were collecting data, faculty talked about how much time must be dedicated to these

assignments and the struggle to give students an authentic, meaningful experience while still making time for other important topics. Faculty members who include data collection/creation in their courses also talked about how the experience was invaluable for their students, but they were always aware of the potential ethical issues involved with real subjects.

> "The logistical challenge was getting them to actually send out their questions and to know what to do when they receive the data back. So like the transferring into the actual SPSS package to analyze the data, that's a challenge as well. … When they're doing that process, they're cleaning the data. They're making sure everything that the coding is correct. So that's a logistical challenge as well, getting them to not only send out their surveys, but then what do you do once you receive the data? Like how do you make sure the data is clean, right?"

> "If they want to collect data [for their project], then they can, but I don't really push it because collecting data is very time consuming and you have to understand how to actually design the instrument and things like that. And then you have to find respondents. [And this is in addition to having only one] semester where they also have to learn the substance of their project."

> "The time consumption and getting them to develop questions, who they're going to talk to, doing the interviews. And then also …we were limited on time [and] I'm not having them actually code interviews and … the analysis always feels really light. … But also thinking about kind of how you pull that in and use that, weaving that into your arguments. …but then there's always that hesitation, right? Just, I'm just always scared to like release my students to go bug people like that."

We didn't find evidence that social sciences faculty at UR are teaching their students explicitly about Preservation and Access/Description issues regarding data. We acknowledge that this absence may have been because of the specific questions asked and there may be more happening than we heard about in this study. We believe that most social sciences faculty would agree that these concepts are important to their own practice and to research, but they are secondary concerns in the context of teaching data skills to students. As a result, we see this as an opportunity for collaboration with the library and others.

Theme 2: Supporting Students & Faculty on the Margins

Data literacy has been identified as a cross-curricular priority at UR; in an April 2021 report, the General Education Curriculum Improvement Committee (GECIC) proposed that all undergraduate students should take at least one course with a focus on quantitative data literacy (QDL) as part of their general education requirements. This requirement could potentially bring a significant number of students into QDL courses for the first time, many of whom would not opt into such a course willingly. The GECIC's proposed change to the general education curriculum would only increase the number of students "on the margins" (those with little or no background in data literacy; those who have forgotten the data literacy skills they previously learned; or who come with a mindset that they cannot learn data literacy skills) in QDL courses, making the need to address this issue both urgent and timely.

Our respondents reported that one of the biggest challenges for students is less about the mechanics of working with data, and more about grasping the underlying concepts: how to start building a research project, why a project needs data that at first glance may not be directly relevant, how to manage multiple tasks within a project, and why they need to learn the fundamentals before skipping directly to more advanced techniques.

> "Some of them definitely get the message and collect all of the information that I asked them to collect. And, others...don't. They'll just leave parts of the form blank, you know, or I'll ask them to do a very detailed sketch of the lay of the land around the marker or the plot and anywhere where plantings are. And some of them are just like, why? What is this?"

> "And so really in many ways it's like getting them to think differently, and then to think differently essentially in another language and in computer code, if you will. And then on top of that, getting them to understand statistical analysis. [...] So it's actually a very huge undertaking, for students when they're collecting their own data, because they have to figure out all of these different pieces, all of these different moving pieces. And sometimes I think it's hard for them to do."

> "The other cultural shift is just the term 'data analytics,' right. They hear that so much. And then they come into stats class and I have to teach them sampling, really fundamental, theoretical things that you need to do the most basic analysis. And they want to go 50 steps ahead and predict these huge terabyte level, predictive algorithms cause that's what they see [in the world]. That's what they ultimately want to do. So that makes it a little bit more difficult to walk them through the thousand steps you need before you actually get there, because they see that right in front of them."

Faculty are also challenged when students in a single course have different amounts of experience, gained their experience in a different discipline than the faculty member, or even just built their foundational knowledge with a different faculty member teaching the same class in the same department.

> “I think the challenging thing I think I've run into is just the spread of prior knowledge [...] For some of them, to go on [a data repository's website] and know what they're seeing and to think through how variable code talk is and what they're seeing in those headings and stuff is really easy. [...] Some of them are pretty set up to where it feels like it's too easy. And then for some of them, it's still a huge struggle.”

> “And that's been a little problematic at the University of Richmond, not because the students aren't up to the task, but because at least in [my school] a lot of them come without any formal training in statistics or social science. So that's one thing that, you know, a challenge for all of us is that we come from different disciplines, and we want to impart discipline-specific or even social science general approaches. It requires a lot of background work with the students.”

> “The most fantastic thing is when you get the same, when you get repeat students. And again, this is not a knock at the other people that teach the course, but when I teach advanced [topic], the students who took intro [topic] with me, we're on the same page, right. They know the process, they know how I teach. I know exactly what I showed them. I know how I taught them. And it makes such a difference to sort of carry them along.”

A number of faculty respondents spoke to us about the challenge presented when students have little experience learning about or working with data at the most basic levels, oftentimes despite their facility with other technologies. Some students come to these courses not only with a lack of experience, but a mindset that working with data isn't something they can learn to do. Helping these students on the margins to overcome these challenges takes a significant amount of work and time on the part of faculty, limiting their ability to engage students in more advanced (and perhaps more engaging) data-related topics.

> “I'm always surprised, you know, they're digital natives, but there's so much tech support we have to do, like, they're great at Instagram, but there are a lot of other things that they're really lousy at. And that takes a lot of work for the faculty.”

> “Some people come into a stats class and it's, you know, they're only 18, but they've already written this off as something that they don't do. It's just not how they think. And it's a real struggle, it's a struggle to teach them. And it's a struggle for them as well, because I think they just haven't had a lot of practice thinking like that.”

> “I don't bring a lot of data now. I try to do a little bit, but you don't know the background of what the students have seen in terms of methods. I like to bring in some examples, but we can't do a lot of real hands-on data analysis stuff in those classes without knowing that everybody's on the same page.”

Faculty at UR address these challenges in a number of different ways. Some start their teaching with the very basics, even when students report prior learning, and even if that means that some students end up getting very little from the course lessons.

> "When I start the course, I ask them if they've had any experience with qualitative data and I get maybe one or two hands. And then even with those one or two people, I come to realize that they really don't know what they're doing. And so I teach it as if they don't know."

> "I know a lot of our students actually double major in [departments that have quantitative methods courses], so that by the time they are taking this course, they are already fairly advanced in statistics. So I have a couple of students in my class right now who, I mean, I feel bad for them. I'm sure they understand what we're learning very well already, but they still have to sit through the classes. That's just unfortunate in terms of how the curriculum, they overlap."

Other faculty allow students with greater experience to chart a slightly different path while still moving in tandem with the rest of the group. Sometimes, this means the more advanced students have to take greater responsibility for their own learning. Other times, the faculty member uses the differences in student projects as a lesson, to help students understand the broader scope of what they are learning.

> "It's not nearly as full featured as Excel is, at least for students who don't program. If students can program in Python or R they can actually do quite a bit, but I don't generally teach that. Occasionally I get a student in there that has some facility with those, and then let them go with that."

> "I'm teaching them that even though everybody in the class is doing different things, y'all are all doing the same thing. You're all recoding, you're all regressing. You're all making correlations. And so it's more about teaching them that the actual process is the same, even if what they want to do is a little bit different."

We also heard from some faculty that they themselves feel like they're "on the margins" of teaching with data. They know the importance of having their students learn about data issues, but don't feel like they have the training themselves to sufficiently teach those lessons. These examples came from disciplines where faculty use many different methods (therefore an individual faculty member teaching a course might be trying to expose students to methods outside their own expertise) or from disciplines in which using data for certain types of analysis or research problems is still relatively new or cutting edge. Several faculty members expressed that teaching students at the undergraduate level meant they had to incorporate so much into their courses that the data pieces were necessarily minor and not really worth them putting a lot of time into learning themselves. In a few cases, these faculty relied on others at the university to help with their data lessons, either professionals in the Spatial Analysis Lab (for GIS "labs" or assignments) or in the former Center for Teaching, Learning, & Technology.

Theme 3: Teaching Data Ethics

The theme of "Teaching Data Ethics" included discussion on teaching students about ethical issues associated with using data, but also about ethical issues of using data in the classroom.

Many of the faculty who said they incorporate discussions of ethics into their data-related teaching did so in the context of data collection. How are we treating the participants from whom we are actively collecting data? Why do we need to take historical relationships between researchers and community into account when collecting data? What steps need to be taken to ensure data collected "anonymously" does not inadvertently identify a respondent? Faculty talked about their own ethical challenges as they teach students about collecting data and also about the discussions they have with their students about these topics:

> "If we look historically, I mean right here at [redacted], bodies were... people's graves were dug in order collect research without the parents' permission, without the family's permission. So a lot of my ethical dilemmas and considerations around how we treat participants and how we treat the data and whether we are respecting them or perpetuating the same thing that has been done."

> "The other thing that we do, that I often think about, is going into communities that are disproportionately impacted by environmental hazards. And going in there in a responsible way that's not treating it like as a tourist stop or something like that for our students."

> "You're doing these interesting little studies in the class and you're just using your class [as subjects]. And it's not even really a study, it's just a demonstration. But I'm asking all these kinds of sensitive questions and there are only two people in X demographic group in the class and I asked something and then like, I'm giving you this dataset. [...] And that's something that we talk about, and it could be an issue."

Others chose to discuss ethics in a broader context, following the practices of their specific discipline. One faculty member uses official documentation sourced from their professional organization and the university's IRB website to have a discussion prior to any work taking place for the project.

> "I have them read the ethical statement from my professional society, read through the IRB stuff. Even though if they're just doing ethnographic oral history projects within a class, that's no longer bound by IRB, but it's worth reading because it gets them thinking about the issues of representation and the fear of losing one's control over one's representation in an ethnographic interaction. So they do all of that at the top."

Another addresses the issue of ethics in order to guide decision-making near the end of the research process:

> "We talk about, you know, what does evidence mean? And then how do you use that information to make ethical decisions? So for example, today we talked about [redacted] in the class, so they ran a simulation, they got some [redacted] answers and, you know, it's who do you care about more? Do you care about the young, but you care about the old and what

are the trade-offs you face? So, you know, we talk about the results, not so much the ethics of data itself."

A third approaches ethical implications as intermediate knowledge, to be considered after understanding a model's function, but not necessarily used to make decisions.

> "I think our goal is to take those ethical challenges and just sort of place them right here and say, 'here they are,' know that they're there, and not really get into them. [...] I think the general sense is if there are problems with doing this, if there are ethical implications, you'll figure them out once you realize exactly how this model is working."

Overall, most faculty who introduced discussions of ethics into their data-related teaching did so in an applied fashion, using very specific situations to raise larger questions about how and why research is conducted. This was especially true for research which included populations that are relatively small (e.g. transgender respondents) or vulnerable (e.g. youth).

> "So if you only ask on your survey, 'are you male [or] female?,' you're missing out on a population, right? But someone who's gender non-conforming, someone who's trans, they may see that and say, 'wait a minute, I'm not represented. And so I feel a particular way because [...] the perception might be the survey is intentionally leaving me out.'"

> "A lot of the research I do asks children to think about topics that can be upsetting, and so I really have to think carefully about how children will think about these questions and could it cause them any discomfort. And if so, it's just not something I can do for purposes of a class exercise. So I think about it a lot with children and particularly because I feel like I need to model it for my students."

> "The goal always has to be, I think, the, restoration of, of information for the community and not what our students can do with it. [...] We can never let that outweigh the main purpose of what we're doing, making the information available."

> "A lot of the ethical challenge in dealing with youth data is setting those boundaries with the adults that think they have every right to all of the data that you're now collecting on their child or their students or their whatever. So we confront some of those theoretical ethical issues more head-on."

Theme 4: Curriculum Sequencing & Alignment

One of the challenges that was mentioned again and again in our interviews is that of curriculum sequencing and alignment. Data literacy, like other literacies, is not something that can be achieved in a single course, so necessarily must be interwoven through many courses. The theme of "not enough time" came up often and is related to many of the other themes expressed here. While this study focused mainly on individual faculty and courses as units of analysis, we want to acknowledge that this significant challenge is also present at the curriculum level.

Most of the social sciences majors at UR don't have the defined structure of prerequisites that dictate how students in sciences majors move through their courses in a certain order. As a result, social sciences majors often don't take their major courses in a specific, predictable order and

instructors of each course can't assume prior data knowledge (except in some specific cases, such as Econometrics and Advanced Econometrics).

> "I have to assume that you had nothing because we do not require it in our curriculum. And so that can be hard, especially when you have such people with divergent backgrounds. Some are like, 'Oh, I'm pretty proficient' and others, 'I've never seen [this]. I don't know what I mean and a mode is.' Okay, all right. I have to assume that, and I'm going to work with you."

> "It would be great if [they] had prior knowledge, but for the way our major is set up, the [capstone course for seniors] is really the only time that they actually do like real-deal original research, like actual analysis. And in my class, [they do] statistical analysis. They do a lot of critical thinking in other classes, but not the actual analysis."

Faculty expressed frustration about this dynamic for several reasons: they cannot move beyond the basics to incorporate real-world, hands-on analysis work as they could if students reliably entered their course with a common methodological foundation; the necessity of starting with the basics often leads to boredom for the more advanced students; and that the issue cannot easily be addressed at the individual course or faculty level because it's a function of how the curriculum is configured. At least one faculty member noted that teaching the same students in both intro and advanced courses can mitigate the issue somewhat, as repeat students are more accustomed to a faculty member's teaching style, and the faculty member knows exactly what and how the students have learned. Others found this to be an unreliable fix, though, because students do not always quickly recall their learning from previous semesters due in part to lack of use and significant time lapses between courses:

> "If you don't use it, you lose it. And so I approach it as if I obviously can't go into the depth that I would if I were teaching a stats course, but I approach it as if they've never heard of this analysis. And then I just try to read the room and if students are remembering, then I respond to that and pick up the pace. But in some, in most cases, they're the majority of students just, 'I have no recollection.'"

A few faculty members mentioned wanting to teach more advanced courses in their department that would have prerequisites, but cited staffing as a road block. Advanced courses (by their nature) could only be taught by one or a few specific faculty members and adding new courses would mean fewer faculty to teach sections of existing courses, including the very prerequisites that would prepare students for the advanced level of study. The staffing challenge also prevents departments from splitting existing courses into two-semester sequences, which would provide students more time to learn, engage with, and hopefully retain methodological skills.

> "Some suggestions [were] floated around in terms of how about we have a second in a series of methods courses, but I think that's still, you know, very early stage of conversation because then we have to think about who [would] teach it."

> "When teaching [the methods course], I only have about two classes to actually get into statistical inference and to get into actual regressions. … These are [research] questions that the students have, and I just don't have time to teach them. And so we're kind of thinking

about maybe getting a quant methods course, just designed to teach them those advanced stats."

The draft proposal for a Data Science minor in A&S is an exciting development that addresses this challenge, enabling students to sequence data-intensive classes across disciplines and develop more advanced data skills and literacy. The writers of this proposal also acknowledge that ongoing conversations will be necessary to align courses across departments so that students can move between courses in different departments with minimal friction. In addition, a minor doesn't fully address the challenge of building data literacy across all students in the social sciences.

In our interviews, several faculty members expressed interest in hearing about how other instructors, especially those outside their own departments, approach teaching with data. While it was also clear that most faculty don't feel like they have time for formal training, this kind of cross-disciplinary information sharing would be desirable.

Recommendations

Our suggestions here are intended to guide the support provided by the Library, Faculty Hub, and potentially other units at UR. They originate from what we heard from faculty in our interviews, and are adapted based on our knowledge of faculty support. These recommendations build on existing strengths and staffing models.

A. Recommendations for the Library's data education program.
    1. Library support for data education and pedagogy should be integrated into the new general education curriculum, just as information literacy is currently integrated into the first-year seminar program.
    2. This support for the general education curriculum should acknowledge differences in disciplinary approaches, but work toward a common experience for all students.
    3. All library data education support in the social sciences should be in close consultation with the course instructor and fulfilling the needs of the course. In addition, it should focus on:
        - Deliberately extending our existing information literacy to the context of working with data. This includes helping faculty integrate resources (both library-purchased resources and external resources) into their courses.
        - The idea of a data lifecycle and what libraries and archives can best contribute, which will often mean the pieces of the lifecycle that are not frequently taught by faculty (like preservation and documentation).

B. Recommendations for the Faculty Hub.
    1. Coordinate a community of practice for data instructors broadly, in addition to existing programming for data researchers and instructors using the new HPC cluster.
    2. Develop materials and programming regarding data literacy instruction that takes into account the specific needs of faculty and students who are "on the margins."
    3. Work with the Academic Skills Center, TLC, and other relevant campus departments to minimize gaps in existing support structures for data-related teaching. For example, Hub technology consultants could co-develop a curriculum with the TLC manager, using information learned from working with faculty new to data literacy instruction, that would prepare TLC consultants for supporting students who need varying level of assistance.

Appendix A: Semi-Structured Interview Guide

The following guide was provided by Ithaka S+R and is standard across all institutions in the *Teaching with Data in the Social Sciences* cohort. While the general topics and question areas were set, we were given leeway to ask follow-up or clarifying questions as needed.

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 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 B: Supplemental Faculty Quotations

We are including more direct quotations from faculty interviewees here to provide additional supporting information and context.

Theme 1: Data Lifecycle

"They want to just Google, you know, and find a dataset that's right. And I'm telling you, it just, it doesn't exist. Life would be great if it did, but you can't do that. So you gotta, you know, you gotta mix and match. You gotta dig a little bit and that's really hard. … I give them a little bit of guidance on, you know, I sort of warn them that they're not looking for a data set, [but] that they're looking for a bunch of data. And then you gotta combine it and that's, again, there's some methods there that I don't get into, but, you know, making it clear that this is a process in itself, not a singular task where you just go find it."

"Oh, I have to construct a hypothesis. I might've gotten that in fourth or fifth grade, but now I'm doing independent and dependent variables. ... Then on top of that, I have to do this coding. I have to do this inputting of data. I have to make sure my data is clean. So it's actually a very huge undertaking, for students when they're collecting their own data, because they have to figure out all of these different pieces."

"And I'm always telling them, just put it in 'notes and clarification.' Like, if you have the question, somebody will have the question down the line and you've got to just put it in there. And as long as it's in there, you're good."

Theme 2: Supporting Students & Faculty on the Margins

"I think methods and survey design and just figuring out a place to jump in is a challenge for them. They don't have a lot of models, don't have a lot of experience and don't really have the experience even to know where to look to get the experience, to find useful models to mimic their work after."

Theme 3: Ethics

"I have a lot of ethical dilemmas are around how we are treating participants, whether we're collecting the participants' data in an ethical way, whether we are perpetuating things that have been done for a long time to participants."

"We'll turn to the ethics of [the discipline ...] and then, you know, personal information about people from the past. It's one thing to think about people who lived 10,000 years ago [...] but when they're recent, where do you draw the line?"

Theme 4: Curriculum Sequencing & Alignment

“I try to do a little bit, but just, you don't know the background of what the students have seen in terms of methods. So it's, you know, I like to bring in some examples, but we can't do a lot of real hands-on data analysis stuff in those classes without knowing that everybody's on the same page.”

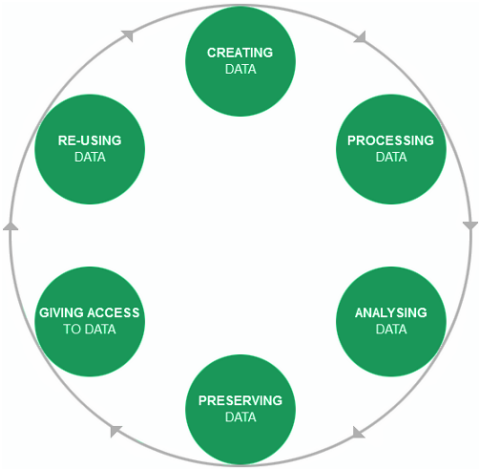
“It's the most fantastic thing is when you get repeat students. And again, this is not like a knock at the other people that teach the course, but when I teach advanced, the students who took intro with me, we're just on the same page, right. They know the process, they know how I teach. I know exactly what I showed them. I know how I taught them.”

“Even just doing basic like, frequency statistics and doing basic like mean, median, mode, but I also take those where I know that probably 85, at least, percent of students, they're like, ‘yeah, we did that as freshmen in high school, what are you talking about.’ And then, and then I've got 10% that are like, ‘what, wait, what is that? I don't know what that is.’ And I, because of my own background, my teaching history, I tend to err on the side of, you know, ‘We're covering it. If you're bored cause you know, this, that's great.’ I'm glad, right. But my job is also to make sure everyone knows this.”

“There are classes that we want to teach that we can't teach because we have to staff. … So yeah, I would like to learn how to teach with data, but frankly, I don't know if the students even have the skills to get to that point where they're going to be able to actually do [advanced] stuff.”

Appendix C: Explanation of the Data Lifecycle Model

Research Data Lifecycle, UK Data Service[2]



The UK Data Services' Data Lifecycle Model focuses on the activities of a researcher (or research group) and describes the cyclical nature of data use and reuse. For example, a researcher may start by creating a data set using data collected with a survey instrument (Creating Data), or they may use an existing public or semi-public data set or combine public data with data they've collected themselves, which is common in many social sciences disciplines (Re-Using Data). They might then use a software program to clean the data, create some imputed variables, do calculations and regressions, create tables and visuals (Processing Data), and decide what they mean and what is significant (Analyzing Data). Of course these steps aren't always done in this sequence or in a linear way. After the analysis steps are complete and a publication is finished, researchers might see this as the end of their data set's life and turn to other projects, but most are actually doing the final two steps in this model as well. "Giving Access to Data" describes a variety of reasons and methods a researcher would want to access "old" data to use again at a later time – for their own purposes, to give to others for use in another project, for a reproducibility study, or made public through an archive for replicability and reuse purposes, at which point the cycle starts again. Access is made possible by the actions taken before the data set is needed to make sure it remains usable, understandable, findable, and uncorrupted (Preserving Data). This last step can be time-intensive, require specialized knowledge and resources, and is rarely any researcher's top priority, so is often a point in the data lifecycle where a library or archive is called upon to step in.

[2] "Research Data Management," UK Data Service, accessed September 29, 2021, https://ukdataservice.ac.uk/learning-hub/research-data-management/