Teaching with Data in the Social Sciences at the University of Massachusetts Amherst

Erin Jerome
University of Massachusetts Amherst

Stephen McGinty
University of Massachusetts Amherst

Teaching with Data in the Social Sciences at the University of Massachusetts Amherst

Erin Jerome, Open Access & Institutional Repository Librarian
Stephen McGinty, Social Sciences Research Services Librarian

Table of Contents

Background

Ithaka S+R Research Study

This report is an investigation of the pedagogical practices of social sciences instructors teaching with quantitative data at the undergraduate level at the University of Massachusetts Amherst. The study was conducted by the University Libraries, and was part of a larger suite of parallel studies coordinated by Ithaka S+R at 20 institutions of higher education in the U.S. The study's goal is to understand instructors' undergraduate teaching processes so that institutions can develop local resources to support them in their work.   While this local study will contribute to a larger cumulative capstone produced by Ithaka S+R, our recommendations will also inform the ways in which the UMass Amherst University Libraries can better support social science instructors who rely on quantitative data in their undergraduate instruction.

The Social Sciences at UMass Amherst[1]

The College of the Social and Behavioral Sciences at the University of Massachusetts at Amherst enrolls 3,936 undergraduate students, which comprises 16.24 per cent of the total undergraduate population of 24,233. Departments In the College include Anthropology, Communication, Data Analytics and Computational Social Science, Economics, Journalism, Labor Studies, Landscape Architecture, Latin American, Caribbean and Latino Studies, Legal Studies, Political Science, Public Policy and Administration, Regional Planning, Resource and Managerial Economics, Social Thought and Political Economy, Sociology, and Sustainable Community Development.

Methodology

We submitted our protocol to the campus' IRB on June 4, 2020 and received clearance on June 24, 2020. This list of possible participants that we compiled included instructors with whom we already had established relationships, recommendations from librarians and other campus colleagues, and instructors that we found in the course listings for undergraduate social science courses. The instructors contacted included lecturers, graduate students, assistant, associate, and full professors.

We began recruiting instructors in the middle of September 2020 and continued through the first month of the Spring 2021 semester. Some of our faculty connections also shared information about our study with their departmental listservs during our recruitment process. Of the 29 instructors invited to participate, 10 declined, 2 expressed interest but did not proceed with scheduling an interview, and 8 did not respond. 9 instructors were interviewed with the distribution of research areas summarized below.[2]

---

[1] U. of Massachusetts Amherst, Office of University Analytics and Institutional Research, UMass at a Glance, 2020-21. https://www.umass.edu/oir/sites/default/files/publications/glance/FS_gla_01.pdf

[2] We had a difficult time recruiting instructors to meet with us for interviews despite reassurances that the study only required them to meet with us for one 60-minute interview. We attribute our difficulties in

Social Sciences Research Area	Number of Participants	Number of Instructors sent Recruitment Emails
Sociology	2	8
Political Science	3	5
Environmental Conservation	1	2
Linguistics	1	5
Education	1	1
Economics	1	4

Interviews were conducted via Zoom between September 2020 and January 2021. Written consent was obtained from the participants (see Appendix B). Participants were given the option of audio only calls through Zoom, but all opted to keep their cameras on. The semi-structured interview guide (Appendix C) was developed by Ithaka S+R and explored 4 areas: 1) instructors' background, 2) getting data, 3) working with data, and 4) training and support. Zoom was set up to create transcripts of the calls, which were recorded to the cloud. The transcripts were sent to Transcript Divas for anonymization and cleanup. We then did additional anonymization and corrections of the transcripts that we received from Transcript Divas.

Coding and qualitative analysis

We did initial open coding on two interview transcripts. The codes were reviewed and discussed, and we created a list of themes and codes to use for the remaining seven interview transcripts. We then divided the codes and themes between us, and each coded all 9 transcripts for our assigned themes. Codes and key quotes were tracked in NVIVO and Google Docs.

Findings

Finding Data

The process of finding and presenting data to the students was identified as a high priority by the instructors. Many felt concerned about striking a balance between the learning objectives of the class and the availability of usable data online. When it came to relying on particular online data sources, persistent availability was cited as an issue. Data sources that participants relied upon to present data occasionally prove to be short lived. Other online sources remained active

---

recruitment directly to the pandemic fatigue and burnout that many in our campus community were experiencing due to COVID-19.

but updated their interfaces in a way that made them almost unusable for students in an introductory level course. One instructor specifically mentioned that they encountered this issue with the General Social Survey (GSS).[3] The GSS had previously allowed users to recode data online, but that functionality is no longer available, and the instructor had to redesign many of their GSS-related assignments in the fall semester. As a result, instructors note that they must continuously check their sources to ensure that they still exist, search for new data online each semester, or check with colleagues for new useful sources.

Some instructors do allow students to find their own data in order to ensure that they will be engaged and interested in their assignments:

> *"*And so, I feel like I feel more free to tell students to find stuff that speaks to them particularly in terms of other learning material. But it also adds an extra responsibility to help them identify what the good ones are, which was probably always an issue. But it feels a little bit different.*"*

However, this strategy caused apprehension, as seen in the quotation above. Instructors face an "extra responsibility" to make sure that the data discovered by the students will fit the goals of the course assignments. Several instructors felt they could not take on that role for reasons of both time and pedagogy and therefore opted to provide students with data instead. Some instructors had previously experimented with allowing students to find their data, but found that method to be more time consuming than they had expected. The majority of instructors have opted to provide their students with data in their courses.

Government sites, both state and federal, are used as data sources by some instructors. Data collection procedures for these sites are consistent and timely. This is tremendously important for instructors because it allows them to concentrate on the learning objectives of the class instead of spending a lot of time on checking for data reliability and clean up. The latter skills are important, but participants noted that they often feel conflicted about how much of these topics they should cover when a course is using data analysis as a teaching tool (vs. a course that is specifically about data and data management).

Frustrations still occurred for some instructors who relied on government sites. Some distinct data presentations were changed by the agencies, like the example of the GSS noted above, which led to instructors changing their assignments. These sites are often managed by government agencies whose missions do not include supporting college course assignments.

Participants who provide data noted an advantage in having their students begin learning about computation and data manipulation from repositories whose data has been vetted and prescreened. This is both a time saver and a skill development technique used by the instructors. Flexibility can come later in the course after students have learned the necessary fundamental concepts. Once students learn these preliminary skills and concepts, they can work independently with a variety of data sets.

---

[3] https://gss.norc.org/

Challenges for Students Finding Their Own Data

> "So, this isn't a class on data, this is a class on analyzing data. But yeah, it was an incredibly frustrating process. So, I think moving forward, I'm just going to have clean data."

As described above, some instructors choose to have students find their own data as part of their pedagogical approach, but most of our participants opted to provide students with data to work with throughout the course of the semester. For some, this was a response to lessons learned after letting students search for and find their own data in previous iterations of the course. Instructors whose choice to provide data was a response to previous iterations of the class noted that allowing students to find their own data ended up being more work than they'd anticipated while also leaving students feeling frustrated.

> *"It's just an exercise in frustration and some of them found data that weren't really raw or even helpful data. Many of them were looking for data that simply doesn't exist in the world."*

Finding usable data requires expertise that a lot of the students do not have at the start of the semester, especially in introductory courses. And most of the courses that our participants teach are not data classes, but rather classes that focus on data analyses. Instructors noted the following common scenarios--a lack of relevant data for the topic the student wanted to focus on, the students would find data that needed extensive cleaning, or students would try to force answers from the data that they couldn't possibly get.

> *"And so I would spend just an inordinate amount of office hours in emailing back and forth with students getting the data just set up in a form that would work for them throughout the course of the semester. And then they'd get frustrated with the project and hop off it or something like that."*

In introductory classes where students do not necessarily have the skills required to make the data that they've found usable for course assignments, our participants generally prioritized reducing the levels of student frustration over the amount of time they might have to spend preparing their course materials. In some cases, cleaning and finding data becomes the responsibility of graduate student teaching assistants, when applicable, but most of our instructors did not have this option.

Even with subject expertise, our participants said that collecting data for classes is not without its own issues. Course preparation certainly becomes more labor intensive. While a lot of big data is publicly available, that does not mean it is immediately usable. Instructors have to take the intermediary steps of cleaning the data or creating usable datasets from larger data collections. As one participant noted, they are often faced with taking messy data and creating something that's less messy to study.

Survey Design and Data Collection

A small number of instructors, particularly in Political Science, did focus on survey design as a means through which students could collect and use their own data rather throughout the semester. Again, instructors focused on course goals when determining whether or not to include specific projects such as survey design.

Instructors who chose to include the use of surveys in their courses expressed the desire for students to see the entire process of survey design through data collection. In these cases, students develop the questions, distribute the survey, and collect the data over the course of the semester. The student-collected data is then used for any related data analysis and research. Students use tools such as Survey Monkey, Facebook Groups, and Excel to complete these tasks. Instructors report using these tools, which students are usually already familiar with, as a stepping stone to introducing new concepts. Once the data has been collected and arranged in the instructor's preferred software for the course, students will then learn about the types of queries that can be made of the data. Despite the apparent differences in teaching methods between those who point to instructor-provided data and those who have students develop and collect their own data, there are still similarities. In both cases, instructors establish a preliminary base of knowledge that allows students some autonomy to explore the data on their own as they progress through the semester.

Instructors reported that they wanted to show students what a survey that has been used for decades looks like in order to inform their own survey development. Participants noted that it is important to draw a distinction between the data collected by their students for the class, and data that is collected, cleaned and displayed in online sources. This is an extra step for instructors who include data collection and survey design in their courses.

Other original sources of data mentioned by participants are Twitter, YouTube, and other social media sites. These data sources were favored by instructors in Linguistics or Political Science in particular. Students who use these sources for their research must learn how to clean their data so that it is usable for analysis. This cleanup process allows students to begin building their own corpus of data.

Participants noted that working with text data comes with its own unique issues and challenges. Newspaper headlines, song lyrics, and presidential speeches all can be analyzed through various software. However, instructors report the difficulties occasionally encountered when accessing news-oriented text data. Some mainstream sources, such as the *New York Times,* have extensive archives, but require the use of an API to interact with their data. This means that instructors either must incorporate teaching students how to work with APIs into their syllabi or run specific queries for students to get them access to the necessary data. Other datasets, that might be considered less mainstream than a source like the *New York Times,* must be purchased and such purchases are not always possible due to budgetary constraints. Under-represented groups are often better represented in smaller local media resources and these resources are not often part of a library's subscription databases. Instructors reported frustration with the market-driven changes in text-based databases that make it difficult for students and

instructors to rely on. Title lists are often dynamic and change over time, making it more difficult to rely on databases as text-data sources.

Student Knowledge, Experience, and Abilities

Most of the instructors interviewed taught with data in introductory classes that lacked prerequisites, so they often do not know what kind of knowledge or experience their students have until the semester starts. They reported devoting significant class time to making sure the students have a basic level of understanding of particular tools or skills that will be required for the rest of the semester (i.e., Algebraic expressions, square roots, and hand calculations). Only then do the instructors feel comfortable in going further with the material. Participants noted great differences in what students already knew, having made assumptions that many basic principles would be common knowledge.

Instructors noted that students often assume introductory courses will be easy and undemanding. Some students have specific expectations about homework assignments and the amount of time an assignment might require in these courses.

> "...some of our majors...are used to the idea that they will spend a fixed amount of time on an assignment. And they'll write something, and then it'll be done, right? Programming assignments are not like that, you write something, and that's where it started, right? And it might take you many hours to get it to work."

With such a variety of skills and levels of expertise in their courses, participants noted that it is difficult to continue to challenge the high performing students while also making sure that the material is accessible to everyone enrolled in the course. That being said, one instructor noted that they try to target their course materials, "for people who are comfortable with analytical reasoning", but who might lack programming experience. Participants noted a tension between the desire to teach students critical thinking skills while also ensuring that all students had the technical skills necessary to successfully complete the course. For professional development opportunities, instructors want students to be able to demonstrate experience with specific software. Yet on the other hand, instructors were wary of granting a kind of "technical certificate" that would be obsolete in a few years.

Assignments and classroom activities are not focused on the workings of the various analytical tools for many of the respondents. These instructors view the software they use as a tool. The tool is not the object of study in the data course, but will help students achieve a successful outcome in their assignments. For some instructors, students don't need to understand the exact mechanics of a particular tool, as long as they can use it well enough to complete their assignments.

Computer Literacy

As one participant noted, "there is this assumption that students understand how a computer works really, but it's not that way." One of the biggest challenges faced by participants is that, in their experience, most of their students lack a basic understanding of how files are saved and organized on their computers regardless of operating system. Students need to have a rudimentary understanding of file organization as well as working directories and file paths in

order to complete assignments in R and Python, the two primary programming languages used by our participants. ("...no matter what computing system you use you need to know where your data is stored and how it functions and stuff like that.") Many participants indicated that this lack of knowledge has only increased over time.

Respondents blame the prevalence of smartphones as part of the issue. Students are mainly digitally fluent through their phones, but data work must be done on a computer. Noting that students cannot find files once they download them, one respondent held up their iPhone, saying, "And you download something and it goes on the home screen and you tap it, well 'where' doesn't matter as much does it?" Many instructors shared similar stories about students who could not find their data files only to discover that those files, along with almost every other file the student had ever downloaded or opened, were all stored in the downloads folders.

In addition to issues with downloads, our respondents noted that their students also do not know how to name their files or organize them locally. Students do not think about using an organizational structure to track where all of their files are. One participant mentioned a student who struggled to import a data file onto their computer. She did not know what folder anything was in because "she had never made a new folder on her computer before, ever." Students do not seem to understand the "save as" function and this becomes problematic when it comes time to work with the data that they have downloaded for their assignments.

This problem is further exacerbated by the fact that instructors cannot stipulate the operating system that their students must use for a specific course. Mac and Windows machines approach working directories and file management in different ways which can make teaching basics to students with little knowledge rather difficult.

This lack of basic computer literacy also impacts the success of a student's installation of Python or R. After experiencing the frustration of ensuring that Python was installed properly on each students' machine, one instructor has made the shift to Google's cloud Python option, Colab.[4]

Most of our respondents who teach introductory courses without prerequisites once assumed that their students would enter their classes with this basic level of understanding or minimum computer proficiency. Some instructors have adjusted their courses so that working directories, file paths and file management are covered in the first few weeks of the semester, but most expressed the desire to have some kind of computer literacy prerequisite that would allow them to devote less course time to this information.

Programming Languages and Software Choices

Participants either used R or Python as their primary programming language when teaching with data in the social sciences. As with all the other choices that instructors must make when

---

[4] https://research.google.com/colaboratory/

planning their courses, they must decide how much time they want to devote to teaching their students a programming language. They are aware that their students will struggle and noted that "it's a bit of a leap" for many. As one instructor said, "The programming language is not the object of the study, it's just a tool."

Licensing issues are a primary concern when it comes to choosing which software will be used for a specific course. Instructors must ask themselves some of the following questions: Is the software free regardless of enrollment or affiliation, like RStudio, or will students only have access to it until they graduate? Can the software be installed locally or are the students required to use a lab computer for their assignments? What do the students already know? Instructors want to ensure that their students will continue to have access to the necessary software throughout their coursework and, in some cases, in their professional careers. This all factors into the process of deciding which software to use and teach in a specific course. If a student will no longer have access to Stata after the semester ends, is it worth devoting valuable course time to teaching students how to use it?

In terms of software that is not free, participants typically choose between Stata and SPSS, noting that SPSS requires students to use lab computers. Excel was mentioned in passing in a few interviews, but not as a primary tool in the way that Stata/SPSS are. There is definitely a push towards using RStudio which is free regardless of enrollment status although some students were unable to install RStudio locally and had to rely on the web-based version.

Conclusions

Instructors teaching with data in the social sciences are operating in a congested landscape. Hardware development, software development, student attitudes, reliability of data sources, teaching assistants, software licensing, course load and the variety of computer operating systems are just some of the issues explored by this research. Libraries and larger institutions can play a role in mitigating some of these concerns.

The disconnection between instructors and students emerged, in part, through their mutual experience in introductory classes with few prerequisites. Instructor expectations have been adjusted over time, but there is still more that schools or colleges can do to address this issue.

The availability of reliable, usable data was noted as an area of concern. Instructors had difficulty locating data that consistently fit the pedagogical requirements of the course. Formats change. Dirty data and clean data are intermingled on the same site. This requires more work for the instructors. To tackle this problem, usable data could be collected into a repository that would provide access to arranged data from which the students could begin assignments.

Few instructors knew that the University Libraries could provide support when it comes to teaching research data management principles or that there is a Data Services Librarian on campus. One of the colleges has a Slack channel for instructors teaching with data, but the university libraries can do a lot more to support instructors who are doing this kind of work. Ultimately, instructors teaching with data in the social sciences would benefit from a greater level of support from the

university (cloud-solutions, VPNs) and as well as further collaboration with the University Libraries (workshop series, libguides, instruction sessions).

Recommendations

Cloud computing and VPN options for students

Increased cloud computing would greatly benefit instructors teaching with data in the social sciences. Installing Python and other software on students' computers can be time consuming and, while some of our participants have opted to use Google's Colab, for example, as a way to avoid this, others expressed the desire to be able to VPN into a resource such as the Massachusetts Green High Performance Computing Center.[5] Students could then login, run scripts, interact with the data, and send proof of assignment completion to their instructor without having to install anything locally on their own machines. This would definitely decrease the number of technological hiccups that both students and instructors encounter.

Furthermore, troubleshooting code and file issues while teaching remotely proved to be difficult for most participants. As the university continues to focus on multiple modes of teaching and learning, it should focus on what support for any courses that involve coding or data analysis will look like.

Repository of usable data

The interviews demonstrated that finding usable data for teaching is a consistent pain point for instructors even when students are encouraged to find their own. Locating suitable data that supports a course's learning outcomes requires a lot of effort. When asked for ways in which the university can better support their teaching, several participants mentioned the possibility of developing some kind of repository of data for teaching. This repository could contain messy data, raw data, cleaned data, and data from a variety of sources, including instructors on campus, and subject areas.

The University Libraries could pilot a teaching data repository over the course of a year, perhaps preparing in the fall by identifying the best data sets to start with and doing a soft launch in the spring.

1st year experience that covers computer literacy

Every participant noted that their students' lack of computer literacy is a problem that is only growing in prevalence on campus. The authors recommend adding some sort of computer literacy prerequisite that students must pass before enrolling in any kind of entry level data analysis course. The authors realize that this is a rather complex recommendation that might not be entirely feasible given the way changes in the curriculum are currently approved.

---

[5] https://www.mghpcc.org/about/about-the-mghpcc/

The Libraries could host a workshop series based on the Carpentries curriculums as a number of librarians are certified Library Carpentry instructors.[6] The College of Social and Behavioral Sciences (SBS) hosts a workshop series (SBS 191) targeted at first year students that covers "how to be a college student". It might be possible to add some computer literacy basics into this series. The Libraries can collaborate with SBS to add this content to the series or to possibly integrate one-shot instruction sessions into specific courses on an as-needed basis.

---

[6] https://carpentries.org/

Appendix A: Individual email recruitment text sent by the authors

<u>First version</u>

Subject: UMass Amherst's study on teaching with data in the social sciences

Dear [Name],

The University Libraries are 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 UMass study is part of a suite of parallel studies at 19 other institutions of higher education in the US, coordinated by Ithaka S+R, a not-for-profit research and consulting service. The information gathered at UMass will also be included in a landmark capstone report by Ithaka S+R and will be essential for UMass 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.

Subject: UMass Amherst's study on teaching with data in the social sciences

Dear [Name]

The University Libraries are 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. We have a flexible timeline, our data is not due until late **February 2021**. Would you be willing to participate in **one 60-minute interview** to share your unique experiences and perspective?

Our local UMass study is part of a suite of parallel studies at 19 other institutions of higher education in the US, coordinated by Ithaka S+R, a not-for-profit research and consulting service. The information gathered at UMass will also be included in a landmark capstone report by Ithaka S+R and will be essential for UMass 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.

Consent Form for Participation in a Research Study
University of Massachusetts Amherst

Researcher(s): *Stephen McGinty, Social Sciences Librarian, University Libraries; Erin Jerome, Open Access & Institutional Repository Librarian, University Libraries*

Study Title: *Teaching with Data in the Social Sciences*

1.   WHAT IS THIS FORM?

This form is called a Consent Form. It will give you information about the study so you can make an informed decision about participation in this research. We encourage you to take some time to think this over and ask questions now and at any other time. If you decide to participate, you will be asked to sign this form and you will be given a copy for your records.

2.   WHAT ARE SOME OF THE IMPORTANT ASPECTS OF THIS RESEARCH STUDY THAT I SHOULD BE AWARE OF?

1)   The fact that consent is being sought for research and that participation is voluntary;

2)   This study seeks to examine social science instructors' practices in teaching undergraduates with data in order to understand the resources and services that instructors at the University of Massachusetts Amherst need to be successful in their work. We expect that your participation will be one interview lasting approximately sixty minutes.

3)   There are minimal risks associated with this research study; however, a risk of breach of confidentiality always exists and we have taken the steps to minimize this risk as outlined in section 9 below.

4)   While there are no direct benefits to you for taking part in this study, you may experience increased insight and awareness into your teaching practices.

3.   WHY ARE WE DOING THIS RESEARCH STUDY?

We are conducting this research study to examine the pedagogical practices of social sciences instructors teaching with quantitative data at the undergraduate level. The goal of the study is to understand instructors' undergraduate teaching processes toward developing resources and services at the University of Massachusetts Amherst to support them in their work. The study contributes 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.

Further, the study at the University of Massachusetts is connected to a suite of parallel studies being developed locally at other higher education institutions. The anonymized, aggregate data shared with the coordinating organization, Ithaka S+R, will be used to compose a comprehensive report written and made publicly available by Ithaka S+R.

4.   WHO CAN PARTICIPATE IN THIS RESEARCH STUDY?

We are seeking instructors who teach undergraduate social science courses at the University of Massachusetts Amherst including tenured and tenure-track faculty, graduate students, adjunct instructors, and staff.

5.   WHERE WILL THIS RESEARCH STUDY TAKE PLACE AND HOW MANY PEOPLE WILL PARTICIPATE?

Interviews will take place either in person or virtually. In person interviews will take place somewhere private, like your office or a room designated for meetings. Virtual interviews will take place over Zoom.

We anticipate 15-20 instructors to take part in this study.

6.   WHAT WILL I BE ASKED TO DO AND HOW MUCH TIME WILL IT TAKE?

If you agree to take part in this study, you will be asked to participate in one 60-minute, audio recorded interview about your experiences teaching undergraduates with data. Your participation in all or part of the study is completely voluntary. You are free to withdraw consent and discontinue participation in the interview at any time for any reason.

Please note that if you do not wish to be audio recorded, we cannot conduct an interview, since note taking introduces variability in data collection. However, we can have an informal discussion with you once our data collection is complete and we have findings to report.

7.   WILL BEING IN THIS RESEARCH STUDY HELP ME IN ANY WAY?

You may not directly benefit from this research; however, we hope that your participation in the study may increase your insight and awareness into your undergraduate instruction practices. More broadly, your participation in this study will help develop resources and services in support of your undergraduate instruction at the University of Massachusetts Amherst.

8.   WHAT ARE MY RISKS OF BEING IN THIS RESEARCH STUDY?

We believe that there are minimal risks associated with this research study; however, a risk of breach of confidentiality always exists and we have taken the steps to minimize this risk as outlined in section 9 below.

9.   HOW WILL MY PERSONAL INFORMATION BE PROTECTED?

Your privacy and confidentiality is important to us. The following procedures will be used to protect the confidentiality of your study records.

Participants will sign informed consent forms, either in person or remotely via DocuSign but these forms will in no way be linked to the collected data because there will be no key that corresponds the participants to their pseudonyms. Informed consent forms will be stored as paper copies in a locked file cabinet only accessible to the investigator(s) and/or as digital files by the investigator(s) in a password protected Box folder. The informed consent forms will be destroyed three years after the completion of the research project. More information regarding the DocuSign process can be found here: https://www.umass.edu/it/news/20191218/docusignreplaceadobesigncampuselectronicsignatureservice.

Interviews will be recorded in order to create transcripts for analysis. We immediately apply a pseudonym to the transcript of our interview. There is no key to link you to your pseudonym. Once transcription of your interview is complete, we delete the original audio recording. We anticipate transcription to take approximately one month from the date of your interview.

At the conclusion of this study, the results of the research will be publicly disseminated, such as through conference presentations, scholarly articles and as part of publicly available reports published online through ScholarWorks@UMass, the University of Massachusetts Amherst's dedicated institutional repository, and the Ithaka S+R website.

Information will be presented in summary format, demographic or contextual information will not be used in public reports of the research findings, and you will not be identified in any publications or presentations.

10. WILL I BE GIVEN ANY MONEY OR OTHER COMPENSATION FOR BEING IN THIS RESEARCH STUDY?

Participants in this study will not receive payment or other compensation for being in this research study.

11. WHO CAN I TALK TO IF I HAVE QUESTIONS?

Take as long as you like before you make a decision. We will be happy to answer any question you have about this study. If you have further questions about this project or if you have a research-related problem, you may contact the researchers:

Stephen McGinty, smcginty@library.umass.edu, 413-545-1871
Erin Jerome, ewjerome@library.umass.edu, 413-545-2174

If you have any questions concerning your rights as a research subject, you may contact the University of Massachusetts Amherst Human Research Protection Office (HRPO) at (413) 545-3428 or humansubjects@ora.umass.edu.

12. WHAT HAPPENS IF I SAY YES, BUT I CHANGE MY MIND LATER?
You do not have to be in this study if you do not want to. If you agree to be in the study, but later change your mind, you may drop out at any time. There are no penalties or consequences of any kind if you decide that you do not want to participate.

13. SUBJECT STATEMENT OF VOLUNTARY CONSENT

When signing this form, I am agreeing to voluntarily enter this study. I have had a chance to read this consent form, and it was explained to me in a language which I use. I have had the opportunity to ask questions and have received satisfactory answers. I have been informed that I can withdraw at any time. A copy of this signed Informed Consent Form has been given to me.

_____   _____   _____
Participant Signature: Print Name: Date:

By signing below, I indicate that the participant has read and, to the best of my knowledge, understands the details contained in this document and has been given a copy.

_____   _____   _____
Signature of Person   Print Name: Date:
Obtaining Consent

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.

- o How does your teaching relate to your current or past research?
- o 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).

- o 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.

- o Do you provide instruction to students in how to find and/or select appropriate datasets to work with?
- o 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*

- o How do you find and obtain datasets to use in teaching?
- o 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)?

- o 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?