Chapter 11
An Approach to Supporting Teaching With Data in the Social Sciences

Michele Hayslett
https://orcid.org/0000-0001-8783-2763
University of North Carolina at Chapel Hill, USA

Angela Bardeen
https://orcid.org/0000-0003-0740-5548
University of North Carolina at Chapel Hill, USA

Kayla Olson
University of North Carolina at Chapel Hill, USA

ABSTRACT

The University of North Carolina at Chapel Hill is moving rapidly toward integrating data literacy and data science in the undergraduate curriculum. This study interviewed instructors across the social sciences to identify ways the university libraries could improve and expand support for teaching undergraduates with data. While the library already offers an extensive suite of data support services, instructors were not always aware of all the services available. In addition to expanded outreach, the study identified three areas in which to expand the library's services: helping instructors obtain data for teaching, extending course-specific and direct-to-student support, and expanding workshop instruction for working with data. It includes advice for libraries wanting to build new data services.

DOI: 10.4018/978-1-7998-9702-6.ch011

INTRODUCTION

Most librarians are familiar with information literacy. The phrase began to appear in journal article titles as early as the 1980s, and Association of College & Research Libraries (ACRL) published its *Information Literacy Competency Standards for Higher Education* in 2000 (ACRL, 2000). These standards cite the definition from the Presidential Committee on Information Literacy:

a set of abilities requiring individuals to "recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information." (ALA, 1989)

Data literacy is an analogous framework with actually a longer history (albeit not in *library* literature) and a host of labels: Walker wrote about "statistical literacy" in 1951. "Numeracy" articles appear in the 1960s and 70s. And the American Statistical Association field-tested its "Quantitative Literacy" Project in 1984 (Scheaffer, 1986). Different authors define these terms differently, of course, but the parallel with information literacy is clear. Schield (2005) identified *evaluating* materials as the common denominator between information literacy and data literacy (as well as statistical literacy which he defines separately), pointing out the similarities of each of these concepts despite their respective unique complexities. Mandinach and Gummer (2013) define data literacy as the ability to understand and use data effectively to inform decisions. This definition is deceptively simple. Just as there are many facets of and ways to approach information literacy, students must gain facility in a wide array of skills to use data effectively. They need to learn how to locate, collect, or create appropriate data; employ ethical protections for populations being studied; evaluate the strengths and deficiencies of a given dataset for a given research purpose; clean and manipulate ("wrangle") data to prepare for analysis; understand the context of data; and analyze the data with methods suitable to the study purpose. Although these terms originated with non-library fields, in the past twenty years more and more librarians have begun to understand the foundational role they can play in offering students and faculty extensive assistance with data literacy in addition to information literacy.

This realization has been spurred by a growing emphasis in higher education in recent years on developing students' data skills . Through the 1990s and early 2000s, a number of institutions built quantitative assessments for incoming students, or dedicated attention to data literacy by offering specific quantitative skills courses for credit or building quantitative practice into existing subject-domain courses. The Quantitative Reasoning Program at Wellesley College has been in place since 1997 and was prefaced by faculty concern about students' data literacy as far back as 1991 (National Numeracy Network Newsletter, 2005). Carleton College studied student use of quantitative information in senior capstone papers (Grawe and Rutz, 2009). Upon finding low use of statistics to support arguments in these papers, the Quantitative Inquiry, Reasoning, and Knowledge (QuIRK) initiative joined efforts with the Writing program on that campus to improve students' quantitative reasoning skills. Miami University of Ohio faculty instituted a core competency requirement for quantitative literacy and developed full-credit courses between 2006 and 2013 to build that competency among their students (Bailey, [n.d.]).

Despite this long history of concern about data literacy in certain circles, systematic campus programs for building data literacy among undergraduates across all disciplines have been slower to appear, particularly at large institutions. A search in June 2021, using a common academic database, EBSCO's Academic Search Premier, with the terms "data literacy" and "undergraduates," yielded only eight articles, with the first published in Fall 2015. Growing national attention for big data and data science

are opening new opportunities, though, for promoting data literacy as a first steppingstone for students into those endeavors. The National Academies of Science, Engineering, and Medicine (NASEM) in its report, *Data Science for Undergraduates: Opportunities and Options*, uses the term "data acumen" rather than literacy, but the concepts are clearly similar: "A new generation of tool developers and tool users will require the ability to understand data, [and] to make good judgments about and good decisions with data … (NASEM, 2018, p. 12)." And ACRL's 2015 update, *Framework for Information Literacy for Higher Education*, specifically notes data as one of the things, "… [s]tudents have a greater role and responsibility in creating … understanding … and using … ethically. (p. 7)."

This chapter describes interviews librarians at UNC-Chapel Hill conducted with social science professors who already teach undergraduates with and about data. .Understanding instructors' teaching processes informed the University Libraries on waysto improve support for data literacy more broadly. While only social science instructors were interviewed for this study, understanding the needs of those in the social sciences may also illuminate or predict the needs of those in other disciplines. In addition, individual social science disciplines use diverse types of data and in quite varied ways. That diversity may offer insights that are applicable to both the sciences and humanities.

The study at UNC-Chapel Hill is part of a suite of parallel studies developed locally at multiple higher education institutions to interview social science instructors about their instruction of undergraduate students hands-on with data. Ithaka S+R, a not-for-profit research and consulting organization that works to benefit the academic, cultural, and publishing communities, worked with the researchers across all of the participating institutions to coordinate this collective effort and to provide guidance on research methodology and data analysis. The research project at UNC-Chapel Hill was implemented exclusively by three Carolina investigators. Subsequently, the anonymized aggregated data, together with the UNC-Chapel Hill findings, will also be used towards a comprehensive national report written and made publicly available by Ithaka S+R sometime in 2022.

This chapter will describe the findings of the interviews at UNC-Chapel Hill and relate them to possible service enhancements to meet the needs instructors described, as well as to point out ways other libraries might develop services to support data instruction in any discipline.

BACKGROUND

The University Libraries serves UNC-Chapel Hill, an R1 university of just over 30,000 students (as of Fall 2020), 64.5% of whom are undergraduates (Office of Institutional Research and Assessment, UNC-Chapel Hill, 2020). The Library has long offered its users data expertise through librarians. The first Geographic Information System (GIS) librarian was hired in 2002, but assistance for locating and understanding data was available well before that through business and government documents librarians and other subject specialists. The current array of data services in the Library includes a wide variety of functional and subject specialties: assistance locating and acquiring numeric and other data, GIS mapping, cleaning and preparing data, data visualization, data analysis, programming, bioinformatics, business and economics, linguistics, physical sciences, and more. Services can always be improved, though, and Library staff were eager to participate in this Ithaka study to identify ways to enhance current services.

METHODOLOGY

Because the goal of the project was to generate insights that can be used locally, the study was designed to be exploratory and small-scale. It did not aim to be statistically representative, nor are the recommendations meant to be prescriptive. Rather, the report and its recommendations provide insight into the methods used by UNC-Chapel Hill social science instructors and the support they need to be successful.

Baker and Edwards (2012) explain that qualitative researchers should consider both methodology and practical issues (time available, intended audience) when determining the sample size of an interview-based study. Although Creswell (1998) recommends 15 to 20 interviews, Guest et al. (2006) demonstrate that data saturation can be achieved with about 12 interviews. Based on this, the researchers aimed to schedule between 12 and 15 interviews, and ultimately interviewed 13 instructors.

In the Fall of 2020, a team of three UNC-Chapel Hill librarians invited instructors across the University's social science departments to participate in individual one-hour interviews via Zoom. Any instructors of record who teach an undergraduate social science course at UNC-Chapel Hill were eligible to be interviewed, whether they were tenured, tenure-track faculty, teaching faculty, graduate students, adjunct instructors, or staff, as long as they had their students use data hands-on. Instructors were identified from personal knowledge and recommendations from subject librarians and departmental directors of undergraduate studies. The interviews were conducted between October 2020 and January 2021 (see https://doi.org/10.17615/at8g-hf24 for the semi-structured interview guide, the email invitation to participate, and the consent to participate form). The recordings were submitted to a contractor for transcription. Upon receipt, the transcripts were de-identified and the recordings destroyed. Participants represented a wide variety of departments across the social sciences: Anthropology, Economics, Geography, Information & Library Science, City & Regional Planning, Political Science, Public Policy, and Sociology. Subjects who agreed to participate were all faculty (mostly tenure-track) and were fairly evenly distributed by rank (see Table 1).

Table 1. Ranks of participating instructors

Professor	4
Associate Professor	5
Assistant Professor	3
Teaching Associate Professor	1

The collected data were analyzed using grounded theory methodology, as per Corbin and Strauss (1990), meaning there were no pre-existing codes; rather, a coding structure was developed by the UNC-Chapel Hill team while reviewing transcripts. During coding and analysis, attention was focused primarily on teaching support needs mentioned by participants to identify ways to improve library services, but cross-campus implications were also noted.

FINDINGS

Two themes emerged from analysis of the transcripts: the challenges educators face in doing this type of instruction, and their ideal outcomes for teaching with data.

Challenges

Participants were invited to discuss their teaching practices before the COVID-19 pandemic, the effects of the pivot to online instruction during the crisis, or both. While instructors did describe challenges related to the pandemic, most felt their online teaching was fairly successful, despite the drastic changes experienced during the emergency. The main impediment of the pandemic was the upheaval throughout the year: time lost when UNC-Chapel Hill had to send students home in both the spring and fall semesters of 2020.

Challenges that were unrelated to the pandemic but more "standard" to teaching with data included how to handle differing levels of knowledge among students; how to address a common apprehension among students about working with data; and issues with technology, even when campus is operating normally.

Student Knowledge

As might be expected, instructors observed students enter their classes with a range of backgrounds and experiences with both tools and content. At one end of the spectrum, instructors had students who lacked basic computer literacy, from *"I had one student who didn't realize she had a right-click button on her Mac,"* to working with files and directories:

They'll try to open up their project, but in order to do that, you have to give it the path name. You've got to know the path name…. So to tell you the truth … there's a lot of things that I think that students don't understand about data and working with data, like the whole idea of folders and paths and variable names and all that sort of stuff….

Instructors also noted they get students at the other end of the spectrum who come with more technical skills and experience, *"Some of them come in with programming skills* [emphasis added by interviewee]. *Some of them know R better than I do, and they're 18 years old."* Some noted that these more advanced students have expectations of learning multiple tools.

Several instructors expressed that the absence of prerequisites provided an opportunity for students to come to their courses with "a clean slate" and saw their classes as an opportunity to get students excited about data science. *"No, the whole point of that course is that they come in completely new to data science. We're evangelizing."*

Other instructors noted that a lack of a consistent baseline knowledge (and sometimes a lack of a foundational class for all students) posed challenges that made it difficult for them to scaffold their students' learning effectively. One noted, *"I think that if they had basic classes and just data structure, it could really help everyone out."* Another stated,

I don't need them to start in the same position, place. I need them to all have a foundation. My experience is not that they come in with that similar foundation and experience with data.

Student Mindset

Students' mindset was mentioned as a challenge separate from their knowledge, often in terms of their fear of data or math. A few instructors noted that students coming to their classes might not expect that they must work with data. Several noted that students may select social sciences classes with the assumption that math will not be required, and they can avoid quantitative work.

Some students select … the humanities and social sciences because they're afraid of STEM topics not realizing that [my discipline] is a STEM social science at this point, or at least it is at UNC. And so they're a little freaked out, and they look at the requirements of the course and the deliverables they're going to have to produce, and they're just like, "I'm going to fail this course."

Getting students into a mindset where they feel less intimidated by math and also where they connect quantitative methods to studying big topics, like war or cancer, takes scaffolding and a willingness to adjust teaching, as one instructor remarked: *"So we have to provide the scaffolding and the confidence for them until they're ready to do it on their own. And it usually takes six weeks or so. … [W]e get students that have just never thought of these things in terms of quantified scores. There's a leap there."* Another instructor restructured their course to eliminate the midterm because they noticed it took time to get students *"back into the math mindset."* The midterm came too soon and students who did poorly on it struggled to get their course grade back up. Eliminating that test gave the students more time to acclimate to the work and succeed in the long run.

Some instructors observed that scaffolding and a concentrated learning experience with data helped students over the initial discomfort:

I think it's like learning another language. And I think that they have to sort of immerse themselves in the language. And for many of them, it doesn't click for a while, but for most of them, they figured out in the end, which I'm pretty impressed with their ability. It's just a different way of thinking.

Another participant echoed this, noting the way they structure their class is more basic than they would like due to students' varying comfort levels with math, but that the structure ultimately helps students gain confidence in working with data.

Technology

While technology is key to working with data, many instructors noted the challenges it posed for teaching: Software is necessary for wrangling and analyzing data, but learning to use it is a barrier for students.

[The software programs are] very important because they stand between the question and the answer…. [W]ithout the software, [students] don't know how to answer the question. So it is definitely an important tool, [but] it feels like an obstacle to many of them, instead of feeling like a new skill.

One instructor noted that getting students comfortable with statistical programs can be time-consuming, requiring a week of class time to teach the basics and help students overcome their intimidation.

In addition to challenges with using various programs and packages, several instructors noted the challenges associated with using technology to *access* a program. Campus offers access to a myriad of virtualized applications through the Virtual Lab. While the platform enables students to use many programs without having to pay for their own copy or install it on their own computer, users first must learn how to use the Virtual Lab platform. A number of participants noted that not all students understand how to use the Virtual Lab, and that it operates differently with Macs versus PCs. One instructor wished for a standard training around the Virtual Lab:

If there could be some kind of training when students first come to the campus, where they have to go onto the Virtual Lab and have it set up … because … the most frustrating thing that I found is two months after the classes started in several assignments is getting emails from people, "Oh, I haven't been able to use STATA this entire time, because I don't know how to use the Virtual Lab."

Instructors seemed to choose which software they have students use based on one of two factors: either 1) what they would be expected to use in the discipline generally (or in the workforce), or 2) its availability for students if not in an academic setting; that is, whether it would be freely available to them after graduation.

Instructors noted that while some students are intimidated by learning statistical software packages, others have expectations about which programs they will learn. Participants conveyed challenges they face in selecting which tools to teach students to prepare them for careers in specific domains.

But then … the … advanced students … often will push back saying, "No, but I already know the better tool." The point that I make to them is, "You're going to be working with colleagues who don't know your better tools. Everybody in [my discipline] will be using Excel as the common language. So you'll be speaking a language that even though it might be more elegant and faster, nobody else understands."

Another instructor debated choosing the tool most widely used in the workforce (with which the students wanted to be trained) versus one that was open-source and would therefore be available post-graduation regardless of place of employment or student finances.

… I'm thinking about it though because the ArcGIS … software is proprietary. It's not available. Once people leave the university, they can get a one-year license after they leave, but it's not open source, it's not open access. And we've had a lot of interesting discussions in my department about, should we even be teaching [that] … software? Every time I have brought up that to students that I'm not going to teach it though, I get a lot of pushback saying, "Hey, this is what is used, so you should probably be training us for that." And so that's what I've actually been trying to expose the students to different software packages other than just that.

A different perspective was offered by another instructor who noted some students welcome the chance to learn multiple programs.

Several instructors cited the importance of flexible, technology-enabled classrooms in facilitating student learning with data, particularly in classes with large enrollments. Classrooms with desks on

wheels and multiple large screens enabled instructors to use small-group exercises to engage students in hands-on work, even in large classes. Students could move their desks and display their laptops on a screen large enough for their whole group to see. One instructor felt their classroom was so vital to their course that in handing off the class during a teaching hiatus, they wanted to tell the instructor filling in for them that they could teach the material in whatever way they liked, *"As long as you don't lose the room, you're fine. Just don't lose the room."*

Ideal Outcomes

Questions about ethics, finding and working with data, and training and support inevitably led participants to talk about what they see as ideal outcomes, both in terms of their students' learning in the classroom and their eventual achievements in the workforce. These outcomes fell into four main categories: scaffolded learning; finding data; ethics; and data as a workforce/life skill.

Scaffolded Learning

Participants mentioned incorporating scaffolding in several ways: scaffolding throughout a single course, scaffolding across a sequence of courses, and scaffolding throughout students' undergraduate careers. This scaffolding helped instructors work towards goals ranging from helping students get over their apprehension around working with data to increasing the depth of their data and analytical skills.

Several instructors indicated it would be helpful to understand students' exposure to and skill level with data and tools. One instructor suggested having an assessment for all incoming first-year students to obtain this information and indicated it would help in scaffolding classes, *"We could be more dynamic in terms of adjusting our approach and our ambitions in the classroom if we really understood what the students can do."* As noted in the Student Knowledge section, some instructors expressed a desire for students to come to their classes with similar baseline knowledge and for there to be a campus-wide quantitative "core competency" or prerequisite required of all students.

The College of Arts and Sciences at UNC-Chapel Hill is offering the Ideas, Information, and Inquiry (Triple-I) program as part of the new undergraduate curriculum. It provides a wide variety of cross-disciplinary exploration courses for first-year students (see Appendix 1 for more detail), which all include a 1-credit-hour Data Literacy Lab. In addition to students gaining greater proficiency working with data generally, the Lab specifically requires them to use Excel to work with data. Eventually this may help students get to a more consistent level of knowledge. In the meantime, individual instructors as well as some departments are finding their own ways to scaffold data learning for students.

One instructor talked about the importance of advancing students' understanding throughout their course and walking them through the process of understanding how usable the data are.

Because one thing that comes up in all of my classes is that I basically say, data that you don't have downloaded or know how to download exactly is data that doesn't exist…. And so I build that in for the final project … we go from topic to problem and question to data that they think is going to be useful or usable for it. And then … me or the TA … make[s] sure that it is actual usable data that they can download and work with in the class.

One instructor discussed scaffolding as a way to help students get over their initial apprehension about working with data and indicated that once they had some experience, students often like it and want to continue learning with data. *"For a lot of them, they said they were really nervous about a class that was all statistics, but they liked a class where they could learn a little. Get their feet wet."*

Another participant talked about sequencing across two courses with the first introducing students to a statistical package and focusing on developing comfort working with data. The optional second course then builds on the first (which is a prerequisite) and is for students *"who want to go to the next level"* with data analysis.

Another perspective on scaffolding concerned development over the course of students' undergraduate careers.

I think that's also something to think about if you truly want students to graduate with a competency in data. It has to be a sustained engagement with data. To get that sustained engagement, there has to be, if not integration, at least some thought to progress.

Two departments use an innovative teaching model that utilizes Undergraduate Learning Assistants, where students who take a course one year become eligible to apply to help teach it in subsequent years. These students are paid in academic credit rather than money. The department may provide extra training for those students, or the course instructor may simply work with them individually to convey expectations and assign tasks. In one department, the move to this model was driven by students who wished they had more data-driven classes to move on to after taking a first one and that they had been given the opportunity to take such classes earlier in their undergraduate career. This model provides those students with an additional scaffolded learning opportunity and gives both instructors and earlier-career students more support for hands-on work.

Finding Data

Methods for discovering and acquiring data were linked to instructors' learning outcomes for their courses. Participants were generally self-sufficient in finding data for their students to use. Their specific choice of data was driven by each instructor's particular learning objectives for their students. Many provide their students with pre-selected datasets, either to prevent glitches (as a quotation in the previous section pointed out) where students believe they can get data but then are stymied when a contact falls through, or to save time and angst, freeing students to focus on learning data analysis instead of data wrangling. One participant expressed: *"… it's just not ideal for me to give them a data set with eight thousand variables and they're overwhelmed by how to even handle this thing in a principled way."*

Others have students find their own data or deal with messy data on purpose in order for students to learn how to evaluate data sources and/or wrangle data. Some have students collect their own data by conducting surveys to learn about wording questions and survey methodology. A few offer flexibility in allowing students to choose whether they want to use the instructors' datasets or data they find themselves based in part on the students' prior experience. Some instructors provide students with data from their own research; others are unable to do so due to license restrictions or subjects' privacy.

Ethics

Participants were asked what ethical challenges instructors in their field face in teaching with data, but they also discussed ethics as being crucial to teach. They interpreted "ethics" in quite a variety of ways.

Some addressed the importance of teaching methodological ethics: protecting respondent confidentiality and the risks posed by data collection *"How do you get data like this? How do you get data on violence without [putting] the people you are talking to or collecting data from … in danger, right?"* Other topics mentioned included not mistaking correlation for causation; considering possible bias in the data and adjusting samples to be representative; and interpreting data contextually, in relation to the whole population rather than in isolation. One comment summarized this well:

I think, in particular, for classes that are delving into social stratification, without training that many professors … can very easily have the data reinforce negative stereotypes and characterizations.… That is my number one concern of how I've seen data taught in some classes. That they don't tell larger stories of historic divestment, of historic genocide, in the case of many indigenous populations. They'll show current diabetes rates in a classroom. It becomes, "Oh, individuals of this particular population," which in addition to not telling the full story of how systems affect the individual, it places students who might share elements of that identity on the spot in a way that I work really hard to not do. I hope that I'm achieving that goal. But that is one of my main concerns.

Because data can strip away that context, if it's not presented with that context, which is why I push so much to have both the quant and qual data talked about together is that there's often a push that the "data" will tell the story. What they're often talking about [is, the] quantitative will tell the story versus "data" discussed more broadly. And then, when the quant data tells you of the magnitude or the frequency without the hows and whys, it can really do a lot of damage that I then see and sometimes have the occasion to have to undo.

Others addressed various ethical facets of teaching: the importance of teaching students with uncomfortable topics, even if students might feel personally challenged or offended by the material; the importance of teaching students about the Institutional Review Board process even though much of student classwork is exempt from that; the need to help students navigate the question of whether to use available data merely because they can versus what is ethical, and in the same vein but from a different perspective, how researchers must make ethical decisions about how to present their data, i.e., the possibility of lying with data visualizations. *"I emphasize about how people can lie with maps. Just because you see this map … well, there's a lot of different ways they could have visualized this data, but they did it in this very particular way."* One person described teaching students how to deal with messy data as an ethical issue, to prepare students for the real-life issues of recognizing problems and knowing how to deal with them.

Other ethical issues mentioned were from more personal but no less important perspectives, such as conveying the importance of giving contributors credit for their work, even when their contribution might be "merely" creating a visualization; and taking care to adhere to sharing restrictions on data they themselves license.

Finally, one instructor commented on how the emphasis on ethics prepares students for the workforce:

Sometimes the students think it's just the skillset, but it's not. I actually think my students are better suited for jobs [in coding] than, say, [someone] who may know R and Python better, but just doesn't know how to process human data in an ethical and competent way.

Data As a Workforce/Life Skill

This theme came through strongly. Participants repeatedly pointed out how students' critical thinking skills are greatly enhanced by knowing how to interpret and work with data, noting it as "*another transferable skill[set] that can be used in a wide variety of areas after graduation*." Instructors discussed the ability to understand data as a life skill: *"We have a responsibility to make sure that, like before, students leave as critical consumers of this data,"* and the ability to use data to understand the world around them:

… for me, every year it becomes more true that data skills are a central thing in [my discipline], and social science more generally. We want students … when they are confronted with new developments or with some claim … to have the socialization and the tools and the habits of mind to be able to answer those things in an analytical way.

Another linked the ability to use data in critical thinking to students' eventual employment:

We don't teach students what to think. We just teach them how to think and how to make their own decisions.... The more comfortable they are with data, the better their job's going to be when they get out of here, and we know that.

Other instructors also talked about how data-related skills are directly applicable in the workforce, improving students' prospects when job-hunting in a wide variety of ways. One instructor when referring to tools and methods to analyze data stated, *"And so I always tell the students, learn as much of this stuff as you can, because it'll help you get a job*." Some instructors talked about preparing students for careers in specialized areas such as GIS mapping: *"… undergraduates in that class have gone on to get GIS positions over the years because they know these different types of data and their different formats and they can do stuff with them and that's a pretty rare skillset."*

Others noted that data-related skills like coding could translate broadly to the workforce: *"… a general understanding of how to code, that is valued in the market no matter what."* Several shared specific feedback they received from past students over the years: *"Either one of those [R or STATA] really … translates into jobs and cash pretty directly. And that's what our alumni tell us. And [what] I've seen anecdotally."* Another relayed feedback from past students about the value of other data-related skills to their work after college.

In fact, I've had a lot of students that I've bumped into over the years say to me, "Oh, that research methods class, I never knew how handy that was going to be. At my job, they required me to design surveys for customers," or some other things.

And another instructor relayed how analytical skills translate not just to getting a job, but to career advancement.

The one that they thank me for down the road is the analytic skillset. I have this email…. It was from a student working for an agency…. And she said, "I just want to thank you. I rolled out the … theoretic analysis we did in your class, and they put me in charge of the whole team."

Finally, one instructor expressed the hope that UNC-Chapel Hill would build data competencies across majors so that *"regardless of your major, regardless of your minor, you're going to get this level of skill that will serve you well as you go out into this changing job market."*

SOLUTIONS AND RECOMMENDATIONS

When the conversation touched on library services, several participants had overwhelmingly positive comments about how the Library currently serves them. Some mentioned specific librarians and/or departments, while others expressed appreciation for the Library overall:

- *[Our subject librarian] is an important resource because she gives them the ... It's like a gut check, like a legitimacy moment for the students. They're like, "Okay, the librarian's here. And she's excited. So maybe these professors are not crazy."*
- *… the Research Hub at Davis Library is amazing. They are lifesavers sometimes. They have great links to data, tutorials, super helpful. Every time I've ever had a question about things, about asking for help finding this, I send students to there all the time. I'm like, I know how to find a lot of data, but [our subject librarian] and the people at the Research Hub are professionals. And so I can't say enough good stuff about them….*
- *… to throw some praise at our library is the way our library celebrates data, all of it from the books in the stacks to the most advanced sophisticated quantitative resources. It goes a long way towards bringing that core language into the fiber of the student experience. And it's just super important. It allows me to do things in the classroom that I'd have a much harder time doing otherwise…. The libraries are a campus-wide identity. They don't represent a specific group of us. They serve everybody. And so to celebrate data science, whether you're a humanist or a natural scientist, that matters. It weaves data science into the fiber of the Carolina identity in a very positive, celebratory way. I love the way the library does that.*

This feedback indicates the valuable role libraries can play in helping students improve data skills and grow a data mindset. Academic librarians everywhere work hard to develop good relationships with faculty and that effort lays the foundation for libraries meeting needs around teaching with data. Much can be done even if functional specialists (in GIS mapping, data visualization, etc.) are not in the budget.

Advice for Libraries New to Data Services

Although UNC-Chapel Hill's Library is already offering an extensive suite of support for teaching with data, the interviews suggest additional ways to expand UNC's current efforts in four main areas: helping instructors obtain data for teaching; improving course-specific and direct-to-student support; expanding workshop instruction for working with data; and growing data collections and support for new research techniques like text and data mining. These might look quite different for another institution, depending on available resources and constituents' needs, but might also inform initial efforts for a library just

embarking on data services. Before noting ways UNC might further develop its services, though, it may be helpful to describe how a library might start out.

Experience over time building data support in the UNC Library suggests the following assessment framework. This common-sense approach can help libraries decide both whether to begin offering such services, and what specific support might be best for their campus. These decisions necessarily involve actions key for planning any new service.

- Assess staff capabilities in terms of skills, interests, and time.
- Assess resources for further developing staff capabilities and within what timeframe.
- Assess sustainability and scalability of user support efforts for the long term.
- Assess faculty and/or student needs.
- Assess the campus landscape—what other offices, centers, institutes, etc., are working in this area? Could you collaborate to have their staff to teach workshops hosted by the library?
- Assess the community landscape—are there businesses or vendors (especially if local) that might be interested in co-sponsoring a data-related program like the UNC/Adobe Course Development Grants, https://cfe.unc.edu/initiatives/cfe-grants/?
- Assess what other services can be ended to enable the launch of new services.
- Test the rollout of new services with a small audience before promoting more broadly.
- Consider carefully how and to whom to promote new services.
- Employ staff reflection and participant feedback to improve content immediately.
- Promote successful services carefully but assiduously.
- After some time (and thereafter at regular intervals), assess how well the implementation of the new service meets user needs, and revise or end the service in response.

Even at a large R1 university like UNC-Chapel Hill, data services within the Library have grown by a deliberate, gradual process. At each step, the needs of users were assessed and weighed against other staff needs across the Library and in reference to the Library's strategic plan.

Leveraging existing resources to build new services is obviously necessary in lean budget times like these, but libraries' human resources are rich with knowledge. Subject librarians have established relationships with faculty in their respective areas and can identify and help triage requests for data or data instruction with data services staff (if available). They also often have at least some knowledge of important data sources in their subject areas and may be interested in developing that further to contribute to building data services in the absence of specialized data staff. Likewise, information sharing between regional academic libraries has been extremely valuable at UNC-Chapel Hill over time, resulting in everything from data workshop tip-sharing to collaborative data acquisitions. Student workers are another possible resource. Leveraging graduate-level and/or advanced-career undergraduate student workers for less complex questions at the Davis Library Research Hub service desk has saved an enormous amount of time for staff, enabling them to focus on more complex questions. Developing instructional materials within a platform like GitHub also gives users access to learning resources after hours. During the 2020 academic year, UNC librarians compared instruction formats to determine whether patrons preferred live workshops or an asynchronous learning option whereby participants received access to instructional materials to work through on their own, followed by a live question-and-answer session. Librarians found the live workshop format was more popular, but that live workshop attendees appreciated having access

to the workshop recordings and asynchronous materials (see Davis Library Research Hub, 2021 in the Additional Reading section at the end of the chapter for a link to the Research Hub GitHub repository).

Resources often emerge in unexpected places. Initial support for numeric data at UNC-Chapel Hill Library grew out of government information staff who received many questions about government data like the U.S. Census. Over time that position was eventually rewritten as a data librarian. Perhaps staff in non-public service departments have capacity and interest to contribute to data support. Data staff here also collaborate regularly with staff in research institutes and administrative offices across campus to build services and workshops. Often places like research computing offices, offices for research, student success offices, and data or research centers have goals in common with libraries and are happy to partner to achieve them.

As with any new service, it may be advisable to start small and promote a service to a particular class or department before advertising it more broadly. Invite regular feedback from users to fine-tune its utility. Gauge its impact on library resources, both in terms of staff and funding. Expand as demand increases and resources allow. Table 2 shows possible staging of data services depending on available resources. If a small library wants to begin building data services but can add no functional specialists, the lower intensity services may still be achievable with existing funding. As staff gain experience or budgets allow specialists to be hired, higher intensity services may be added. Higher intensity services may be much more difficult to offer without devoted funding. Such funding may: allow for existing staff to pursue new knowledge or building on existing skills through continuing education, enable the hiring additional staff, and may enable the purchase of new or improved computing equipment to support student use of specialized computing programs with demanding base requirements. Even without devoted funding, there may be free or low-cost continuing education opportunities including things like campus licenses for online learning services such as LinkedIn Learning, and open educational resources.

Table 2. Possible levels of data support by intensity of staff involvement

Area of Support	Low Intensity	Medium Intensity	High Intensity
Instruction	Link to outside tutorials from statistical agencies, other libraries, etc.	Build tutorials to address frequently asked questions. Offer class instruction for finding and using data.	Teach workshops for data applications and techniques. Offer data visualization or GIS mapping support to students and faculty. Work with faculty to build class assignments.
Collections	Offer research guides to open data sources. License databases that provide numeric data or mapping capabilities.	License datasets that come with their own documentation and can be cataloged easily.	License more complex data sources with little or no metadata. Create metadata and original catalog records.
Data Wrangling	Offer data assistance to a particular class.	Offer an open data service desk staffed by student workers.	Consult on faculty research projects as co-author(s).

Every library will be different, and these services may occur in a different order than presented here. Because the UNC-Chapel Hill Library first hired a GIS mapping specialist, services offering data assistance to particular classes and departments began at the same time as consulting with faculty on research projects. The "medium-intensity" service of offering a service desk staffed by student workers came later.

Outreach

At UNC-Chapel Hill, outreach may be the key to improving the way the Library supports the use of data in the classroom. As the interviews progressed, many instructors were found to be unaware of the full range of existing services. Getting the word out about the support already available may provide immediate relief to instructors and students alike. Libraries newly offering data services should not neglect this important activity. Library staff can work with library communications to target social media outreach advertising not only specific events like workshops, but also services and resources like databases and subject guides. Also, emails from subject librarians to specific student or departmental groups can effectively promote opportunities to specific audiences. Recently at UNC, registrations for a new library workshop targeting humanities scholars were substantially increased by working with subject librarians to promote the workshop to specific researchers and student groups. Also workshop participants here are added to an email list for monthly emails promoting services and upcoming workshops. Emails are also sent to student services staff in targeted departments to promote specific workshops or services.

In the course of the Ithaka project, the UNC-Chapel Hill Library identified a number of tools and services that could be better promoted. In addition, some services can be expanded or new services offered to address other needs expressed in the interviews. The following three sections offer suggestions for libraries to incorporate from the start if they are establishing new data services, or pointers for existing data services to tweak their offerings in similar ways.

Helping Instructors Prepare for Teaching with Data

Library support for data literacy may provide an especially valuable resource to instructors new to teaching with data. To date, UNC Library has promoted data-related instruction primarily to those faculty already using data in their classrooms since those instructors have been the majority asking for librarian input. The University's new and growing emphasis on data science and data literacy, however, is expected to begin encouraging other instructors who have not previously taught with data to incorporate such teaching methods into their classes. This offers a prime opportunity for data librarians to broaden outreach to a wider audience. The following strategies can be used within libraries' typical outreach methods: brief stories on the library's home page highlighting particular services or resources; social media posts; physical or digital posters or signs; etc.

- Focus on helping instructors who are just starting to work with data.
    - While most of the faculty interviewed for this project did not indicate they needed help finding data to use in the classroom, assistance finding data for class use is a service that may benefit instructors who want to *begin* teaching with data. Identifying instructors in this category may be challenging unless they self-identify by contacting their subject librarian, but libraries might also notify offices that support teaching (such as UNC's Center for Faculty Excellence) of the availability of such services.
- Create or link to an existing guide such as UNC's Using Data in the Classroom (https://guides.lib.unc.edu/teachingdata). Consider soliciting input from faculty already teaching with data about resources they have found useful.
- Further promote existing subscription data sources such as Social Explorer and Data Planet that offer ready-to-use data. Most libraries have one or two subscription databases that offer statistical

data, and there are plenty available for free from government sources, trade associations, nonprofits, think tanks, etc.

- Create a service for faculty of evaluating data assignments in advance of using them in class. Sometimes all this has to be is working through an assignment the way a student would and identifying any spots where students might have trouble. UNC has offered this service for some time and it only has small uptake so it could be better promoted.
- Promote campus services in- and outside the library that support course design with data. Libraries will want to identify such services as part of their initial campus assessment before initiating data services.
    - UNC's Center for Faculty Excellence offers competitive grants for faculty who need support to reformulate a course.
    - The Library's Scholarly Communications Officer is available to consult on legal questions, such as whether a dataset licensed to a faculty member allows student access.
- Offer faculty support with cleaning and/or wrangling raw data into a format more manageable for students or suitable for an instructor's purpose. This service will depend on having functional specialists on staff, or at least librarians comfortable with such applications as Excel, OpenRefine, or code such as Python or R.

Extending Course-Specific and Direct-to-Student Support

Several faculty members noted the valuable role UNC librarians already play in introducing data to classes through instruction and course-specific research guides. Such instruction and related guides customized to the class topic introduce students to some basics on how to think about and locate data and also note applicable ethical issues that students need to keep in mind. Libraries may employ the following techniques to establish or extend similar success in a variety of ways.

- Poll instructors who teach with data to identify specific operations students have difficulty with and create tools or workshops to address those specific issues that would be marketed directly to those classes. This can be valuable outreach simply in making more instructors aware of library services.
- Promote existing asynchronous materials and/or create additional resources.
    - Reposition existing documentation that outlines how to connect to virtual computing platforms to make it more visible. UNC offers two of these platforms and although campus IT maintains them and provides a good bit of documentation, some aspects could benefit from additional how-tos.
    - Record videos for navigating virtual computing platforms and understanding file directory structure. Bear in mind appealing to different learning styles—text how-tos can be extended with video tutorials to help illustrate the steps.
    - Create or link to asynchronous tutorials that coach users in how to use subscribed databases' features for creating datasets, comparing data, and creating data visualizations that can be exported for use in reports. Database vendors often provide helpful documentation that librarians can just link into a guide. Likewise many librarians have already recorded or constructed help around particular resources or processes and are generous about allowing

others to link to them. Particularly in the case of coding, developers have a culture of openly sharing chunks of code online and have no expectation of asking permission to reuse or link.

- Promote (or further promote) availability of functional specialist consultations to assist undergraduates with wrangling their data and setting up files to conduct their analyses.
- Promote the availability of walk-in assistance with data projects at the service desk staffed by students.

These last two actions depend on offering these services, of course, but may be targeted to specific classes. For promoting the availability of functional specialists, faculty teaching undergraduate honors classes may be most interested to hear of such services. For public service desks oriented to data support, classes that employ particular software in which those desk staff are knowledgeable may likewise be enthusiastic about the opportunity to refer their students for help. Library staff may want to invite those instructors to provide their lab assignments to desk staff. This enables the staff (whether students or library staff) to practice the specific activities to be ready for students' questions.

Workshop Instruction for Working with Data

The UNC Library offers a long-standing workshop series covering a variety of data-related topics: GIS mapping, data visualization, and coding, among others. The coding workshops are offered in collaboration with staff from a campus research institute (for the beginR series) and the campus Research Computing group (for the Python series).

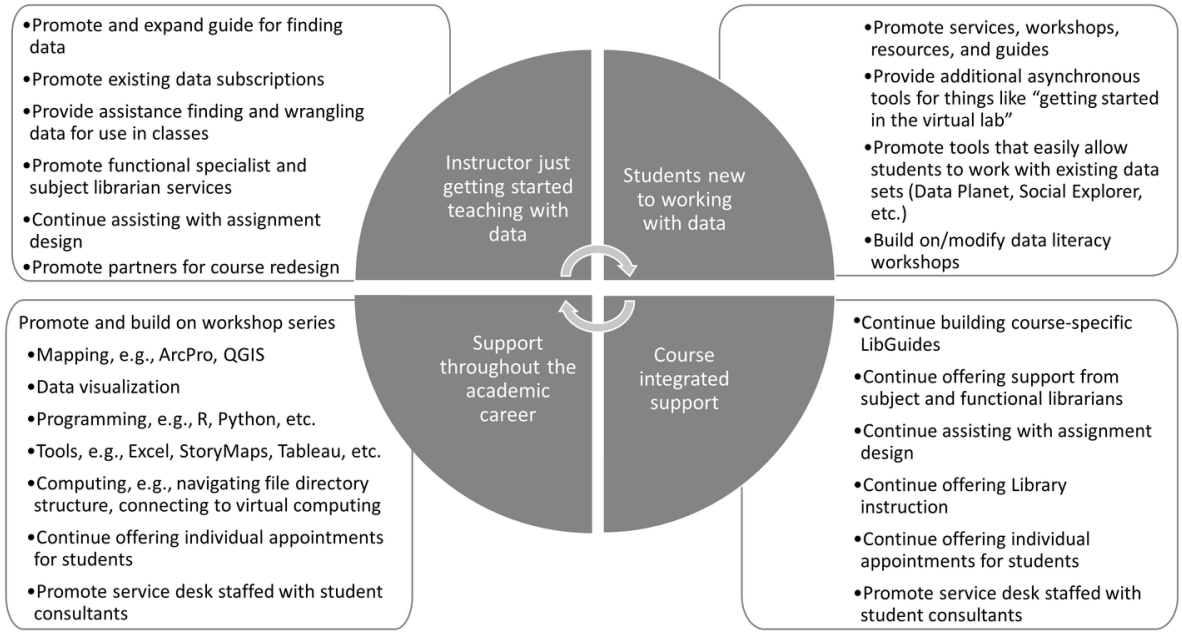
These offerings give students additional opportunities for learning outside of class, providing an opportunity for less experienced students to level up their skills to what their peers may have learned before arriving at UNC, and helping more experienced students strengthen or diversify existing skills. There are a variety of ways the program can be improved, though. Other libraries may take inspiration from UNC's topics and course outlines in planning their own workshops.

- Offer (or expand) Excel workshops. While one of the goals of the Triple-I classes at UNC is for students to gain greater proficiency working with data, specifically using Excel to work with data, these courses are only open to first-year students. The Library already offered a recurring workshop in Excel prior to the development of the Triple-I program that is open to anyone. This one-shot workshop could be expanded to a two- or three-session series to cover more material, especially data cleaning and wrangling tasks. Other libraries may want to avoid describing Excel as an easy tool with which to manipulate data. Excel was created as an accounting program and although it can be useful for some types of data wrangling, advanced data users are justifiably critical of it as a sophisticated data tool. UNC's workshop content is available from https://go.unc.edu/ExcelBasics.
- Consider introductory workshops on very basic topics like opening and navigating statistical programs used in campus classes like STATA and SAS. A Library partner at UNC already offers in-depth workshops on more advanced use of such programs so it is worth investigating whether such a partner might collaborate. If librarians know instructors who require using such programs, interviewing those faculty about what they want their students to know would be both a good way to plan content and a good way to let them know they could send their students to the workshop.

- Continue to add new workshop topics to meet changing needs. New sessions added at UNC for Fall 2021 included research file management strategies, and tools and topics from The Carpentries (https://carpentries.org/) such as OpenRefine and Tidy Data (three librarians have become certified Carpentries instructors).

Figure 1 demonstrates how service revisions support the growth of data literacy at different stages of need.

Figure 1. Point of Need Library Services for Data Instruction Support



FUTURE RESEARCH DIRECTIONS

UNC-Chapel Hill believes all graduates should be data literate and is expanding opportunities in data science education. Participants in these interviews expressed the profound importance of data literacy across all social science disciplines—and across the University as a whole—to ensure that students have the skills needed to be engaged citizens and successful professionals. Within the Library, Data Science is a strategic priority, and supporting student success is one of the pillars in the Library's Strategic Framework. To support student success, the Library seeks to:

- empower learners to navigate the complex digital world as information consumers and creators
- build students' research competence and confidence and contribute to student performance and retention, and
- focus resources on students most in need of support. (University Libraries, [2019])

All three of these objectives are relevant to Library efforts to support teaching and learning with data and guide service choices moving forward. In doing these things, library staff can ensure that all UNC-Chapel Hill students have support to achieve data literacy, and that those who want to pursue more in-depth learning have the resources and skills to take advantage of new data science programs, regardless of their experience and skills upon entering Carolina.

This trend at UNC-Chapel Hill reflects a national effort to educate undergraduates more thoroughly with and about data. There are many ways libraries can play to their traditional strengths in supporting teaching with data. Future research directions might include studying what outreach methods are most effective; how data and/or digital research methods are being used in humanities classrooms in order to understand more specifically the support needed therein; and the most effective support methods for non-social science disciplines. Measuring the impact of library workshops on student outcomes is another area for possible future research. While participant evaluations of Library workshops are consistently positive, this is not a substitute for empirical evidence of student learning.

CONCLUSION

Library efforts at UNC-Chapel Hill offer a model from which others can learn. Librarians can help inexperienced students gain skills quickly through live workshops and asynchronous material, improving equity for students. They can also assist instructors in preparing to use data in the classroom, in evaluating assignments, in instructing students about data sources and basic concepts, and by providing a service point where students can get additional assistance using data. Library staff should expect to synchronize their work with that of other campus entities so respective efforts are complementary and build on each other. Services like these extend the support libraries have traditionally provided to instructors and affirm the importance of libraries in the twenty-first century. Prado and Marzal (2013) argue for data literacy to be embedded within information literacy programs. They both outline core competencies of data literacy to this end and describe significant case studies of such integration.

As with data analysis, context is key for successful support for teaching with data. As one study participant noted, support must be tailored to students' needs:

So, for example, I teach a class … for data scientists … at [another institution] [where] the data scientists are very good … at data wrangling, but not very good at data interpretation. My … students [here at UNC] are the other way…. [W]hen they have some data literacy … [they're] very good at interpretation, but not at the data wrangling. So, if I'm teaching [here] I'm trying to plug that hole, [but when] I'm teaching [at the other institution] I'm plugging the other hole.

There are many ways for academic libraries to support their instructors and students, but the most success will be realized if librarians customize support to the specific needs of their own campus. Moreover, library staff should expect campus needs to change over time. Librarians need to reassess how well their services are serving their constituents periodically and be prepared to adapt as needed.

ACKNOWLEDGMENT

This research study received funding from the University Libraries of UNC at Chapel Hill for participation in the Ithaka project as well as for the cost of transcribing interviews.

The authors would also like to acknowledge and thank Tia Francis, the DRS Digital Research Support Specialist, for her help with formatting chapter images; and Amanda Henley, head of Digital Research Services (DRS), and Jacqueline Solis, head of Research and Instructional Services, in UNC's University Libraries, for editorial help and for their encouragement and support in this project.

REFERENCES

Bailer, A. J. (n.d.). *Implementing a Quantitative Literacy Core Competency Requirement in the College of Arts & Science at Miami University*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.640.1172&rep=rep1&type=pdf

Baker, S. E., & Edwards, R. (2012). *How many qualitative interviews is enough?* National Centre for Research Methods Review Discussion Paper. https://eprints.ncrm.ac.uk/2273/4/how_many_interviews.pdf

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri*, *63*(2), 123–134. doi:10.1515/libri-2013-0010

College of Arts & Sciences, University of North Carolina at Chapel Hill. (2018). *New Ideas, Information and Inquiry Courses Explore Broad Topics Across Disciplines*. https://global.unc.edu/news-story/new-ideas-information-and-inquiry-courses-explore-broad-topics-across-disciplines/

College of Arts & Sciences, University of North Carolina at Chapel Hill. (2019). *IDEAs in Action Curriculum: General Education at the University of North Carolina at Chapel Hill*. https://curricula.unc.edu/wp-content/uploads/sites/332/2020/01/IDEAS-in-Action-Curriculum-April-2019.pdf

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*(1), 3–21. doi:10.1007/BF00988593

Creswell, J. W. (1998). *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Sage Publications.

Framework for Information Literacy for Higher Education. (2015, February 9). *Association of College & Research Libraries*. https://www.ala.org/acrl/standards/ilframework

Grawe, N. D., & Rutz, C. A. (2009). Integration with Writing Programs: A Strategy for Quantitative Reasoning Program Development. *Numeracy*, *2*(2), 1–18. doi:10.5038/1936-4660.2.2.2

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82. doi:10.1177/1525822X05279903

Interview: The QR Program at Wellesley (2005, November 15). *National Numeracy Network Newsletter*, (2), 1-3. https://math.dartmouth.edu/~nnn/newsletter/002/Interview-Wellesley.pdf

Mandinach, E. B., & Gummer, E. S. (2013). A Systemic View of Implementing Data Literacy in Educator Preparation. *Educational Researcher*, *42*(1), 30–37. doi:10.3102/0013189X12459803

National Academies of Sciences, Engineering, and Medicine. (2018). *Data Science for Undergraduates: Opportunities and Options*. The National Academies Press. doi:10.17226/25104

Office of Institutional Research and Assessment, University of North Carolina at Chapel Hill. (2020). *Enrollment Fall 2020* [infographic]. https://oira.unc.edu/

Office of Undergraduate Curricula, University of North Carolina at Chapel Hill. (2021). *Triple-I: Ideas, Information, & Inquiry*. https://ideasinaction.unc.edu/first-year-foundations/ideas-information-and-inquiry/

Office of Undergraduate Curricula, University of North Carolina at Chapel Hill. (n.d.a). *Ideas, Information, and Inquiry (Triple-I) Course Development—Start the Process*. https://curricula.unc.edu/iii-course-development/#pane-0-1

Office of Undergraduate Curricula, University of North Carolina at Chapel Hill. (n.d.b). *Ideas, Information, and Inquiry (Triple-I) Courses*. https://curricula.unc.edu/students/ideas-information-and-inquiry-iii-courses/

Presidential Committee on Information Literacy. (1989). *Information Literacy Competency Standards for Higher Education* (Final Report). Association of College & Research Libraries. https://www.ala.org/acrl/publications/whitepapers/presidential

Scheaffer, R. L. (1986). The Quantitative Literacy Project. *Teaching Statistics*, *8*(2), 34–38. doi:10.1111/j.1467-9639.1986.tb00623.x

Schields, M. (2005). Information literacy, statistical literacy, data literacy. *IASSIST Quarterly*, *28*(2-3), 6–6. doi:10.29173/iq790

University Libraries, University of North Carolina at Chapel Hill. (2019). *Strategic Framework*. https://library.unc.edu/about/strategic-framework/

University of North Carolina at Chapel Hill. (2020). *Carolina Next: Innovations for Public Good*. https://carolinanext.unc.edu/wp-content/uploads/sites/1105/2020/01/CarolinaNext.pdf

Walker, H. M. (1951). Statistical literacy in the social sciences. *The American Statistician*, *5*(1), 6–12. doi:10.1080/00031305.1951.10481912

ADDITIONAL READING

Burton, M., Lyon, L., Erdmann, C., & Tijerina, B. (2018). *Shifting to Data Savvy: The Future of Data Science in Libraries. Project Report*. University of Pittsburgh. http://d-scholarship.pitt.edu/33891/

Carlson, J., & Johnston, L. R. (2015). *Data information literacy: Librarians, data, and the education of a new generation of researchers*. Purdue University Press. doi:10.26530/OAPEN_626975

Dai, Y. (2020). How many ways can we teach data literacy? *IASSIST Quarterly*, *43*(4), 1–11. doi:10.29173/iq963

Davis Library Research Hub, UNC-Chapel Hill Libraries. (2021). *University Libraries Research Hub Workshops*. https://unc-libraries-data.github.io/

Federer, L. M., & Joubert, D. J. (2018). Providing Library Support for Interactive Scientific and Biomedical Visualizations with Tableau. *Journal of Escience Librarianship*, *7*(1), e1120. doi:10.7191/jeslib.2018.1120

Harp, M. R., & Ogborn, M. (2019). Collaborating Externally and Training Internally to Support Research Data Services. *Journal of Escience Librarianship*, *8*(2), e1165. doi:10.7191/jeslib.2019.1165

The Carpentries. (n.d.). *About the Carpentries Curricula*. https://carpentries.org/workshops-curricula/

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(10), 1–23. doi:10.18637/jss.v059.i10

KEY TERMS AND DEFINITIONS

Course-Specific Support: Resources and instruction a librarian provides to a specific course, usually customized to the instructor's particular assignments and learning goals.

Data: The plural form of "datum"; commonly used to refer to research material. While data can assume many forms depending on the particular field—texts for humanities researchers, images for astronomers, videos for psychiatrists, etc.—the term is used in this chapter to refer specifically to *numeric* data.

Data Wrangling: Researchers often receive data in a format that will not work for their analyses. In such circumstances they must clean and manipulate them into a more usable form: remove unnecessary fields or tags; reformat from columns to rows or vice versa; create new fields by concatenation or calculations; etc. This process is referred to as data wrangling. Good guidelines to use as a goal during data wrangling are Hadley Wickham's Tidy Data principles. An article about these standards is available in the Additional Reading section.

Functional Specialist: An expert in a functional area like GIS mapping, data visualization, or text mining, as opposed to a subject librarian.

Geographic Information System (GIS): Software that combines data with maps to convey information. Where reference maps generally portray only geographic features and boundaries, spatial datasets also include characteristics about the features. For example, GIS maps may describe the population living in a place by such things as race, age, income, etc., or the amounts and types of ore mined there, or the amounts and types of vegetation that grow there.

Scaffolded Learning: Sequencing instruction to build learners' skills and confidence in a logical, manageable way, avoiding gaps that might cause confusion.

Tidy Data: Principles described by Hadley Wickham for organizing data to prevent problems during data wrangling and analyses. Principles include such things as replacing functions/formulae with their resulting values, recording only one data point per column or cell, and never using formatting (e.g., bolding, highlighting, etc.) to convey information. These principles are particularly important when moving from a statistical program with a graphical user interface into a programming language like R or Python. An article about these standards is available in the Additional Reading section.

APPENDIX. IDEAS, INFORMATION, AND INQUIRY (TRIPLE-I) PROGRAM

From the Office of Undergraduate Curricula web site:

Ideas, Information, and Inquiry ("Triple-I") courses bring together three outstanding professors from different fields to examine common themes from multiple perspectives. Each course is a rare opportunity to join some of UNC's top scholars as they investigate big ideas. Triple-I courses are reserved for first-year students only. (Office of Undergraduate Curricula, UNC-Chapel Hill, n.d.b)

The Ideas, Information, and Inquiry ("Triple-I") program was created in response to the UNC's ten-year strategic plan, Carolina Next: Innovations for Public Good, which was formally approved in early 2020. Within the strategic initiative to strengthen student success, the plan outlines the objective to "Facilitate learning that is experiential and collaborative, develops individual strengths and encourages the understanding, ethical use and application of data" (UNC-Chapel Hill, 2020). Two of the component strategic opportunities noted there include:

- Implement the IDEAs in Action General Education Curriculum; and
- Assure that every graduate of UNC-Chapel Hill is data literate.

The University's new General Education curriculum, Identify—Discover—Evaluate—Act (IDEAs) in Action, had been under development since 2016, and was approved by the Faculty Council in April 2019 (College of Arts & Sciences, UNC-Chapel Hill, 2019). The Triple-I program was proposed as one of its key components, "… designed to introduce students to four key capacities that they will develop in future study: data literacy, global orientation, principles of evidence and collaboration" (College of Arts & Sciences, UNC-Chapel Hill, 2018).

These classes intentionally showcase how different disciplines can inform a given topic from their respective viewpoints, challenging assumptions and contributing new insights. These are 3-credit courses with an added 1-credit data literacy component. Department chairs or departmental curriculum chairs must affirm that a proposed Triple-I course will be taught at least three times in the subsequent five years, beginning when the course is first taught (Office of Undergraduate Curricula, UNC-Chapel Hill, n.d.a). The eventual aim is to offer 19 Triple-I classes per year, enough seats for almost the entire first-year class to participate. Students may only receive credit for one Triple-I class.

The first pilot Triple-I courses launched in the spring of 2019. Since they are proposed by the teaching faculty members, the topics and course designs vary. The first offerings included Death and Dying; Ethics, Economics, & Public Policy; The Idea of Race; The Environment, Intersectionality, and Science Fiction; and Happiness and Well Being. The interdisciplinary nature of the courses is immediately apparent. Death and Dying was taught by faculty from Psychology & Neuroscience, Anthropology, and American Studies. The Environment, Intersectionality, and Science Fiction faculty were from Environmental Sciences and Engineering; African, African American and Diaspora Studies; and Women's and Gender Studies. These are large classes, usually 250 students, but are intended not to be merely lecture-based. Several have incorporated unique learning experiences in addition to engaging with data, from visiting museums to analyze artwork on class themes, to conducting personal interviews about class-related topics, to assigning projects that encourage 3-D printing and other hands-on fabrication.

Because these were experimental efforts (the program is to be fully implemented in Fall 2022), the data literacy component was not required for the initial iterations. And because any given three faculty members are likely to have different approaches to data and data analysis, and since individual professors have varying levels of experience with Excel (deemed a key piece of the data literacy component), in the Fall 2021 semester the program is piloting an online Data Literacy Lab that will provide all Triple-I students with consistent data instruction using Excel (Office of Undergraduate Curricula, UNC-Chapel Hill, 2021).