2022

Diving into Data Literacy Instruction

Marybeth McCartin

Andrew Battista

Katherine Boss

DIVING INTO DATA LITERACY INSTRUCTION

MARYBETH MCCARTIN, ANDREW BATTISTA AND KATHERINE BOSS

INTRODUCTION

Requiring undergraduates to integrate data into projects is a growing trend. For students, diving into the data seas can be challenging, and many turn to the library for support. But librarians can also feel unsure in these waters. This paper describes a shared teaching model that encourages undergraduates to learn critical data literacy and visualization skills while also enabling librarians to increase their own competency and teach in a structured, supported environment. The model allows for effective learning to take place at scale, empowers librarians from many disciplines to grow as teachers, and enables NYU Libraries to affect the culture of undergraduate learning at our university.

The model originated with a collaboration between the NYU Libraries and an undergraduate communications program. The Communications Librarian and a Data Services Librarian partnered with the Media, Culture and Communications (MCC) department to develop an assignment involving data visualization (one of four in the course). The class is a first-year offering in the Media and Cultural Analysis major, one NYU's largest undergraduate programs. This core course averages 10 sections of 20 students per semester. The collaboration consists of a librarian-led class session that prepares students to evaluate data and construct a visualization that explores some element of media and political economy, or, the idea that the ownership of or access to media and communications infrastructure is intrinsically related to the well-being and development of countries around the world. Students are introduced to basic socio-demographic data and are invited to create a visualization in Google sheets that investigates any element of the political economy theory.

To support this class sustainably, the data and communications librarians needed to recruit other NYU librarians. Recruits were often data novices, but they viewed the invitation as an opportunity to learn data basics, expand their repertoire, and strengthen their practice. Calling on colleagues to teach outside their comfort zone is a big ask, which requires support and institutional buy-in. Hence, we provided those who volunteered with a fleshed-out lesson plan, a thorough hands-on training session, and the opportunity to shadow more experienced instructors before teaching the module solo (MCC-UE 14, 2019).

After a few iterations of the Libraries-MCC collaboration, the project became a replicable model. Students spend more time on the critical aspects of using data, rather than on the mechanics of finding data and integrating it into a visualization tool. Besides introducing data literacy basics, the model can be adapted to teach other forms of information seeking that incorporate audio, visual, or media production. We believe that this model bolsters student and librarian learning, thus contributing to the university's broader teaching mission. Below, we describe the MCC project in more detail and offer key takeaways. As a prelude, we briefly touch on two elements relevant to the discussion: the trend toward including data literacy in the undergraduate curriculum and the value of "learning by teaching" as a strategy for librarians to develop data literacy skills.

THE PUSH FOR UNDERGRADUATE DATA LITERACY

There have been several discussions in higher education pedagogy that recognize the need for students to develop fluidity with digital media and quantitative reasoning. In 2005, Association of American Colleges and Universities (AAC&U) began a decade-long initiative called Liberal Education and America's Promise (LEAP), which resulted in an inventory of 21st century learning outcomes for undergraduate education. Quantitative literacy is on the list (AAC&U, 2015). An AAC&U rubric statement asserts that "[v]irtually all of today's students [...] will need basic quantitative literacy skills such as the ability to draw information

from charts, graphs, and geometric figures, and the ability to accurately complete straightforward estimations and calculations." The rubric urges faculty to develop assignments that give students "contextualized experience" analyzing, evaluating, representing, and communicating quantitative information. (AAC&U, 2009, p. 1). Similarly, the Framework for Information Literacy in Higher Education (2016) stresses that information creation is a process and emphasizes knowledge practices congruent with the goals of data visualization. In particular, learners see data as a non-traditional form of information and "recognize that information may be perceived differently based on the format in which it is packaged" (ACRL, 2016).

Other organizations and disciplines also advocate for quantitative literacy in the undergraduate curriculum. Locke (2017) discusses the relevance of data in the humanities classroom and points to ways undergraduate digital humanities projects can incorporate data analysis and visualization. Similarly, research funded by the Knight Foundation and conducted by journalism scholars at Columbia and Stanford universities resulted in a report recommending that every journalism program provide a foundational data journalism course (Berret & Phillips, 2016). Each of these discussions played a role in our work with the Media and Cultural Communications department as we strove to create and support a meaningful learning experience with students.

THE POWER OF LEARNING-BY-TEACHING

In addition to creating meaningful learning for students, our collaborative model emerged with the conviction that structured, programmatic teaching can foster professional growth for librarians and library staff. Conventional wisdom says that the best way to learn is to teach someone else, an idea confirmed by research. Several studies published within the past decade indicate that learning with the intent to teach can lead to better understanding. One study, by Fiorella and Mayer (2013), finds that learners who were expecting to teach the material to which they were being introduced show better acquisition than learners who were expecting only to take a test. The authors believe this is because learning-by-teaching pushes the learner beyond *essential* processing to *generative* processing, which involves organizing content into a personally meaningful representation and integrating it with prior knowledge (Fiorella & Mayer, 2013).

Another study by Nestojko, Bui, Kornell, and Bjork (2014) finds that learners who were expecting to teach show better organizational output and recall of main points than those who were not expecting to teach. According to the authors, learners who anticipate teaching tend to put themselves "into the mindset of a teacher," leading them to use preparation techniques—such as concept organizing, prioritizing, and structuring—that double as enhancements to a learner's own encoding processes (Nestojko, et al., 2014, p. 1046). This evidence boosts our belief that learning-by-teaching is a good strategy for librarians to build foundational data literacy skills, and it informed the development of our program at NYU.

DEVELOPMENT OF THE LIBRARIES-MCC MODULE

First Attempt

The collaborative teaching model between NYU Libraries and the Media and Cultural Analysis program began in 2015. Situated in NYU's Steinhardt School of Culture, Education, and Human Development, the MCC program covers areas such as global and transcultural communication, media institutions and politics, and technology and society. Looking to interject practical elements into a theory-heavy course, MCC program administrators approached the Data Services Librarian to co-develop instructional content that would expose students to applied data literacy. Together, they designed a module to pilot in two sections of the core course Media and Cultural Analysis, which had 45 students per section. The module consisted of a class session that covered selected data basics and prepared students for an assignment that would ask them to develop a research question and make a visualization related to media and political economy. The students would develop a research question, find a relevant data set to speak to that question, choose a software tool (from an array of tools including ESRI's ArcMap, Carto, Plot.ly, Raw, or Tableau) with which to analyze the data, arrive at some conclusions about their research question, and communicate their results visually.

The assignment's ambitious learning outcomes and fairly loose parameters turned out to be problematic, and the 75-minute class session could not provide sufficient preparation. Students struggled with developing viable research questions, finding data sets, cleaning data, and learning new software tools. To compound the difficulties, the MCC instructors (unbeknownst to us) required that their 90 students make individual appointments with both Data Services *and* the Communications Librarian. Requests began to flood in, which created an overwhelming situation for the Data Services department and the lone Communications Librarian. Library staff did their best to deal with the situation but ultimately realized the module would need significant adjustment going forward, especially since the MCC department was interested in expanding the pilot to all 10 sections of the course.

Second Iteration

For the second round, members of the Data Services team and the Communications Librarian collaborated to revise instructional content and refine the assignment. Previously, students had struggled with finding, cleaning, and manipulating data without error. They also had trouble connecting this exercise to the broader ideas of media and political economy intrinsic to the assignment. Furthermore, library instructors needed to correct these issues and prioritize what the students could reasonably learn. With this in mind, they worked to refine the learning outcomes, create a companion LibGuide, and distribute the teaching load by having Data Services specialists join the instruction pool. The revised learning objectives and lesson plan elements were to:

- Become familiar with the principles, concepts, and language related to data visualization
- Investigate the context and creation of a given dataset, and think critically about the process of creating data
- Emphasize how online visualization platforms allow users to make aesthetic choices, which are part and parcel of the rhetoric of visualization

These rearticulated learning outcomes relieved students of the onus of finding a dataset, which had been a point of frustration and anxiety. Instead, a handful of selected data sets were prescribed. Identifying, cleaning, and preparing these data sets meant extra work on the Libraries' part, but it would allow students to focus on the higher order activity of investigating the relationship between visualizing information and examining social or political culture. Further, the lesson plan was adjusted to include a comparison of four example data visualizations created from the same data, questions for eliciting a discussion about the origins and constructions of data, and practice scenarios for visualizing data in Google sheets and Carto. To provide a better lead-in for students, a preparatory lesson plan was developed for the MCC instructors to present in the class prior to the library visit.

The revisions saw mixed success. Scaling down the learning outcomes helped, though students still struggled with using the data sets, especially when using Carto. One of the fundamental objectives of the assignment still eluded many students, namely, connecting their data analysis back to the larger themes of media and political economy. In addition, there was variability in the way individual MCC instructors described, interpreted, and assessed the assignment. A lack of communication between the librarians and MCC faculty resulted in some tension in the classroom and miscommunication over the goals and purpose of the library instruction.

Third Iteration

In this iteration, the Data and Communications librarians resolved to think more programmatically and work toward strengthening their partnership with MCC departmental administrators. In round two, all ten library sessions were taught by the Data Services staff and Communications Librarian, which still proved onerous. It became clear that sustainability depended on enlisting help from a broader group of librarians. To achieve a level of consistency for this number of sections and library instructors, the Data and Communications librarians recognized the importance of standardizing the lesson plan—both on the Libraries side and the MCC side—in order to avoid conflicting interpretations of the workshop and goals of library instruction. Thus, they consulted with the Instruction Design Librarian to better assess the strengths and weaknesses of the current module. They distributed a survey to MCC faculty asking about the software tools and datasets the students were using to complete the assignment, what aspects of the assignment students struggled with, and what aspects they enjoyed. Survey results informed conversations with the MCC administrators about how to move forward.

Following fruitful discussions with MCC, several refinements were made to the module. The preparatory lesson plan, delivered by MCC instructors the week before the library visit, was strengthened to include a brief set of slides introducing fundamental principles of data visualization. The assignment was re-envisioned, based on an article on Israel's Digital Occupation of Palestine, which ushers in key concepts of media and political economy (Tawil-Souri, 2015). Students attended the lead-in session having read the article, and the MCC instructors spent time introducing and explaining the assignment.

Improvements to the lead-in session ensured that the students were prepared with context for the library workshop and an understanding of why the library was supporting the assignment. Basing the assignment on a specific article made it possible for librarians to model a way of bridging the theoretical concepts of the class to a question that could be asked of data. There was also more time for two pair-and-share discussions and group work in Google Sheets and Carto, which addressed a fundamental and recurring frustration in the students' understanding of the assignment—the ability to ask an original question of a dataset, and to ask a question that would address a larger theme of media and political economy.

Figure 1: Sample Section of Library Lesson Plan
<Placeholder; Editors will place Table here in final doc>

Once these changes had been folded into the library lesson plan and the assignment given to the MCC instructors, the workshop felt ready to be fully scaled. The Data and Communications librarians invited a wide range of librarians to attend a "train the trainer" informational session. Everyone who attended was also invited to shadow an experienced librarian on the workshop.

Then, when they felt comfortable, each person would teach a section individually. This outreach has been incredibly successful and has built up a pool of librarians able to support and teach this out-of-the-box instruction section. It has allowed a community of practice to grow at NYU Libraries around data literacy for undergraduates.

Conclusions and Takeaways

Educators now recognize the need to provide foundational data literacy to undergraduates, and many teaching faculty look to the library for support. One approach libraries can take is to develop an adaptable introductory data literacy module that can be applied in various disciplinary contexts. Sustainability of this model depends on equipping a wide range of librarians with necessary data literacy skills, which can be achieved with a learning-by-teaching approach. As we moved through iterations of the module, we gained some important insights. First, keeping module content streamlined and assignment requirements clear-cut will lead to more effective learning for both students and library instructors. Second, when recruiting library instructors, emphasize that volunteers will not only build their data literacy skill set, but will also expand their pedagogical knowledge and teaching range. Third, to ensure that volunteer instructors have a successful experience, provide support mechanisms such as a tight lesson plan, an intensive training session, opportunities to observe and team-teach before going solo, and a point person to contact with questions and concerns.

REFERENCES

Association of American Colleges and Universities (AAC&U. (2015) *An Introduction to LEAP: Liberal Education & America's Promise Excellence for Everyone as a nation Goes to College*. Retrieved from https://www.aacu.org/leap

Association of American Colleges and Universities (AAC&U). (2009). Quantitative VALUE Rubric. Retrieved from https://www.aacu.org/value/rubrics/inquiry-analysis

Association of College & Research Libraries (ACRL). (2016). *Framework for Information Literacy for Higher Education*. Retrieved from: http://www.ala.org/acrl/standards/ilframework

Berret. C. & Phillips C. (2016). *Teaching Data and Computational Journalism*. New York: Columbia Journalism School. Retrieved from: https://journalism.columbia.edu/system/files/content/teaching_data_and_computational_journalism.pdf

Fiorella, L. & Mayer, R.E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, *38*(4), 281-288.

Locke, B.T. (2017). Digital Humanities Pedagogy as Essential Liberal Education: A Framework for Curriculum Development. *Digital Humanities Quarterly*, *11*(3). Retrieved from: http://www.digitalhumanities.org/dhq/vol/11/3/000303/000303.html

Nestojko, J.F., Bui, D.C., Kornell, N., & Bjork, E.L. (2014). Expecting to teach enhances learning and organization of knowledge in free recall of text passages. *Memory & Cognition*, *42*(7), 1038–1048.

MCC-UE 14 Media & Cultural Analysis (2019). Retrieved from https://guides.nyu.edu/mediaandcommunication/mccue14/.

Tawil-Souri, H. (2012). Digital occupation: Gaza's high-tech enclosure. *Journal of Palestine Studies, 41*(2): 27-43.

Images for Tables and Figures (Editor will put in body of the text later)

Figure 1

Pair & Share: asking questions of data | 5 min.		
1	**Have students look at the UN Millennium Development data set and brainstorm a question to ask of the data**	What are some questions they might ask of this data? Have each group share their most promising question. This is a good time to highlight that asking questions about data is a circular process that is wrapped up in the task of manipulating it. The first question they ask of the data might not necessarily be the most compelling, but that's ok.