ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

ASEE ROCKY MOUNTAIN SECTION CONFERENCE 2023

The Graduate Student Research Data Bootcamp: A Work in Progress

Elizabeth Novosel

Elizabeth Novosel is the Computer Science, Mathematics, and Social Sciences Librarian at the University of Colorado, Boulder. She supports the Computer Science, Mathematics, Applied Mathematics, Information Science, Data Science, Economics, and German Studies departments on the Boulder campus.

The Graduate Student Research Data Bootcamp: A Work in Progress

Abstract:

Before the start of Spring semester 2023, the Center for Research Data and Digital Scholarship at the University of Colorado Boulder offered its second data bootcamp for graduate students. Although overall a successful event, the instructors hope to improve the bootcamp in future iterations to better meet student needs. Some of the lessons learned include making events exclusively virtual or in-person, rather than hybrid; more clearly stating the curricular goals; and more clearly defining the intended audience. All of these changes could help manage student expectations for learning outcomes and improve attendance and overall satisfaction. In addition, more instructors with expertise in all areas would be beneficial to help expand this program.

ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION
Rocky Mountain Section
MINES

INTRODUCTION

Increasingly, students in higher education do not receive adequate training in skills that are essential for conducting research and completing technical degrees, and which can make them more marketable and competitive for future career opportunities [1], such as jobs involving data management and analysis. To help fill this gap, academic libraries and other entities often offer short-format instruction on such topics, as well as in related topics, such as conducting literature reviews, understanding the research lifecycle, and issues regarding scholarly publishing [2].

Although commonly associated with STEM (science, technology, engineering, and math) disciplines, data analysis and related skills and knowledge have become an important component of research in many academic fields, including humanities and social sciences [3]. If training in these skills is not integrated into their program's curriculum, students may be expected to learn such skills on their own [4]. Many students in this situation turn to instruction offered in workshop series, drop-in hours, and concentrated educational events such as bootcamps which strive to "teach learners from diverse backgrounds foundational data competencies that they can immediately apply to their work and research" [5].

THE BOOTCAMP EVENT 2023

The week before the beginning of Spring semester 2023, the Center for Research Data and Digital Scholarship (CRDDS) at the University of Colorado at Boulder held its second annual *Graduate Student Bootcamp* event. The bootcamp curriculum was advertised as providing a "gentle introduction" for "incoming and continuing graduate students" to data-related topics that "will increase their access to cutting-edge methods as they begin their programs and research." The specific topics to be covered included: "reproducible research and data management; tools for collaboration and sharing (using GitHub and GitHub Pages); data analysis and visualization fundamentals in Python and R; and an overview of interdisciplinary data practices and ethics."

The instructors for the bootcamp consisted of CRDDS faculty and staff as well as librarians affiliated with the center. The CRDDS was developed in 2017 as a partnership between Research

Computing and the University Libraries and consists of a collaborative team of data and information professionals who provide education, support, and community for data-focused research, students, and scholars on campus. Much like the first iteration of the data bootcamp that was held in 2022, the 2023 bootcamp included instruction sessions in data management, discovery, and analysis; an introduction to GitHub for collaborative work; how to make use of the institutional repository for data storage and dissemination; and introductory sessions on R and Python. Additionally, librarians taught sessions on data ethics, how to find and evaluate datasets, and an introduction to digital humanities, scholarly publishing, and qualitative research methods. Participants were able to earn a digital badge for their participation.

This three day event was scheduled before the start of the semester in an attempt to provide support to graduate students, who are often very busy with many commitments and are unable to attend workshops scheduled during the semester [4]. To accommodate the greatest number of students, the bootcamp was planned as a hybrid event; there were twenty-four in-person seats available and one hundred virtual seats. The actual number of participants was much lower; the number of attendees fluctuated over the three days but averaged around 12 in-person and 30 online participants. The in-person participants were provided with beverages and snacks throughout the day and offered a daily catered lunch.

Once planned, the bootcamp was advertised in November via email to the entire CU Boulder campus in an effort to include students from all disciplines and not preferentially invite students from certain disciplines. From their experience from the first bootcamp, the planners anticipated that students from a variety of disciplines would sign up. All seats were filled, and a huge waitlist had formed within 24 hours of the announcement.

Overall, the bootcamp was successful. Only ten participants offered feedback, most of whom had attended virtually. The majority of the feedback centered around the Zoom webinar format of the bootcamp; participants found it inadequate because they were not able to interact with each other in breakout rooms and some found that it was hard to get their questions answered due to ineffective chat monitoring. A few participants reported that they would have preferred more time to learn R or Python; others found the sessions to be a good opportunity for review.

From the perspective of the instructors, the bootcamp went fine but could be improved. The amount of demand indicated that there are many students who feel they would benefit from more data-related instruction. The format, number of presenters, advertising, and broad scope are all components that may be changed in future iterations.

POSSIBLE IMPROVEMENTS

In terms of format, the original idea was that a hybrid format would allow participants to choose either in-person or virtual to best accommodate their schedules and preferences, since many people continue to be uncomfortable in large groups congregating inside due to COVID-19 and other illnesses and the bootcamp was taking place in January. Although many students appreciated the option, in practice, it was hard to manage a hybrid format. The small number of instructors and organizers had to teach sessions, manage physical spaces and food, and attend to chat questions as well as questions from the in-person attendees. Relatedly, there were not enough instructors with expert knowledge in R and Python who helped teach to give enough support to students who were struggling. This has led to the idea of eventually offering tiered courses to give all participants better support based on their technical skill-level.

Other aspects of the bootcamp in need of improvement included: the scope of the bootcamp, a clearer definition of the intended audience in terms of experience, and a more limited advertising strategy. Although the announcement tried to convey that the content would be introductory, many experienced and even expert data users signed up and likely were among those who stopped attending. Other participants were very inexperienced, displayed evidence of feeling confused or lost during the technical sessions, and could have benefitted from a slower, more basic approach to the material. In the future, the bootcamp might be more effective if it were offered for more specific topics, geared toward specific disciplines, or focused on specific research methodologies. It would also be advantageous to clearly define the activities, topics, and tools that would be covered, such as in the form of a simple syllabus. Additionally, it might be helpful to have participants fill out an application for the bootcamp so that the organizers could

group or accept participants by knowledge level and ask questions about the specific learning needs and desires of applicants.

Although many educational settings are returning to in-person instruction, it might be easier to keep a bootcamp in a virtual format, but allows for the use of breakout rooms or digital spaces for students to interact, work together in groups, or ask in-depth questions and receive individual help; therefore a Zoom webinar format, which does not allow this functionality, should not be used. Physical instruction has benefits, such as social interaction and relationship building, but it is also much more expensive if food and beverages are provided, which could impact the sustainability of this learning model. Additionally, it is much easier for more instructors and students to attend if they are not required to come to a specific physical space. This flexibility is important in terms of the timing of this bootcamp; feedback indicates that before the beginning of the semester is an optimal time to attend a learning event, however, the weather may be prohibitive, and students may still be in other locations while on break. It is also important to recognize that increasing numbers of programs are completely virtual and an online bootcamp would be inclusive of students in such programs.

Perhaps the most significant issue the organizers encountered was the limited number of instructors, especially those with the expertise to teach the technical sessions. Luckily, the instructors for each session were able to teach their sessions (and were not sick or absent.) However, it would have served the students better, particularly those attending virtually, to have had more expert help available to them in small groups or individually. Some of the chat moderators could not answer the technical questions participants asked, so at times those questions were delayed until another instructor could answer. This issue is substantial in that the bootcamp format is relatively new to the CRDDS and is therefore not a part of anyone's job description, so instructors participated voluntarily in addition to their regular duties. Given the level of interest across many disciplines as well as for undergraduate students, more instructors would need to be involved to meet the high level of demand.

FUTURE ITERATIONS

There are many other aspects of the bootcamp that could be improved, but the issues addressed in this short paper outline the organizers' highest priorities. The next iteration of the bootcamp is scheduled for May 2023, and the organizers are piloting an application process. This next session will be completely in-person; there is discussion of holding the bootcamp several times per year in alternating formats. The CRDDS also plans to continue to investigate similar approaches at peer institutions for further inspiration and to participate in the national conversation of how institutions of higher education are providing students and researchers with adequate data-centered skill development to support their academic, research, and career aspirations.

REFERENCES

[1] "Broader training sought for STEM students," *American Psychological Association Monitor on Psychology*, vol. 48, no. 7, Aug. 2017, Accessed: Feb. 12, 2023. [Online]. Available: https://www.apa.org/monitor/2017/07-08/stem-students

[2] D. F. Feldon *et al.*, "Null effects of boot camps and short-format training for PhD students in life sciences," *Proceedings of the National Academy of Sciences*, vol. 114, no. 37, pp. 9854–9858, Sep. 2017, doi: 10.1073/pnas.1705783114.

[3] T. Arnold and L. Tilton, *Humanities data in R : exploring networks, geospatial data, images, and text / Taylor Arnold, Lauren Tilton. - Record details - EBSCO Discovery Service*. Springer, 2015. Accessed: Feb. 12, 2023. [Online]. Available: https://discovery.ebsco.com/c/3czfwv/details/qbv7eduwev?limiters=FT1%3AY&q=humanities%20and%20data%20analysis

[4] B. L. Fong, "Boot Camps for Graduate Student Success: A Collaborative Initiative," *Journal of Library Administration*, vol. 59, no. 4, pp. 373–394, May 2019, doi: 10.1080/01930826.2019.1593710.

[5] A. Ranganath and J. Wrigley, "The Data Bootcamp as a Platform for Data Literacy Education: Reflections from the University of Colorado-Boulder," University of Colorado Boulder, Jun. 01, 2022. Accessed: Feb. 12, 2023. [Online]. Available: https://zenodo.org/record/6633777#.Y-leVBPMJqy