

A Proposal for a *Data Science in Education Using R* Course

Joshua Rosenberg, Ph.D.
Department of Theory and Practice in Teacher Education
Bailey Graduate School of Education
University of Tennessee, Knoxville

Background and Rationale

Across education, several new courses and degree programs have been developed that share a focus on applying new research methods to new data sources, such as data from students' interactions with digital technologies and teachers' participation in online professional learning communities. Though these courses and programs have different names (e.g., learning analytics, educational data mining, and educational data science) that reflect different assumptions about education, they share a number of common features and can collectively be considered in terms of the application of data science methods in education (Wise, 2020), or *data science in education* (Rosenberg et al., 2020).

One key feature of data science in education is an emphasis on the role of programming and computational tools in educational research (Mishra et al., 2015; Schneider et al., 2020; Siemens & Baker, 2012), which can allow students, analysts, and researchers to ask new questions or to solve problems in new ways. Programming allows researchers to access previously challenging (or impossible) to access data sources and to efficiently prepare these (often large and complex) data sets. Computational tools also allow for researchers to apply new methods, such as network analysis to explore the impact of peers' relationships on learning-related outcomes, or computational text analysis to evaluate the content of large numbers of responses to written assessment items. While courses and workshops designed to enable researchers to work with data in creative ways, they are not yet widespread, and their quality is uneven.

In addition to the utility of programming and computational tools, there is growing interest on the part of educational researchers in data science in education. Since 2018, I have presented workshops on topics related to data science in education at the American Educational Research Association, the International Conference of the Learning Sciences, the Association for Educational Communications and Technology, Florida State University, Michigan State University, and Middle Tennessee State University. I have also written a book, *Data Science in Education Using R*; since February 2020, the open access version of the book has been accessed more than 4,500 times.

In the Department of Theory and Practice in Teacher Education and across the Bailey Graduate School of Education (BGSE), students are increasingly interested in studying contexts such as educational social media use, games-based learning, and geospatial information systems (*GIS*). In the College of Education, Health, and Human Sciences, a survey of faculty conducted (in spring 2019) revealed a strong interest in greater support for advanced research methodologies (for graduate students and themselves). At the university level, computation and data sciences are "areas of strength,"¹ with multiple initiatives intended to broaden access to data science being undertaken, but with heretofore limited work being undertaken in TPTE and BGSE, which may lead to faculty and students being left out of conversations about how data relates to teaching, learning, and educational systems. While there is student interest in

¹ <https://research.utk.edu/funding/internal/seed-programs/>

(and faculty support for) these topics, no research methods course targeting the process of researching these and related contexts yet exists.

Goals of and Objectives for the Proposed Course

In response to the abovementioned needs, the overarching goal of the proposed course, *Data Science in Education Using R*, is to support graduate-level students across the Department of Theory and Practice in Teacher Education (and the Bailey Graduate School of Education) to be empowered to use new data sources and research methods in their research. Accordingly, the course will provide students with a foundation in data science capabilities, defined as those that integrate computation with statistics and substantive expertise. Also, the course will include opportunities for students to gain experience with specific data science techniques that may be widely used across research projects.

Students will have the opportunity to bring their own data from their research projects for use in this class. In this way, they will have immediate application for the concepts learned in the course. Bringing in one's own data will be highly encouraged, as the work done in this class could serve as the foundation of a future conference proposal or publication for the student. If no data is immediately available from the student's research, students can use one of hundreds of freely available datasets to complete coursework or students can use datasets provided to them.

My goal in teaching this course will be to catalyze students' interest in data science and to bolster their confidence in their abilities to use programming techniques to support their research programs. Many people who try to self-teach become overwhelmed by available resources. This course will provide scaffolding to help students become proficient in a few sophisticated data science techniques, and it will give students enough foundational knowledge to pick up new data science skills on their own after the course is through.

The objectives for the proposed course are for students to be able to:

1. Install, set up, and use R and RStudio
2. Use reproducible workflows (so that analyses can easily be modified and then carried out again by the analyst or others) with R Markdown
3. Prepare and explore complex data sources for analysis using the tidyverse suite of R packages
4. Create a social network data structure and create a network visualization
5. Carry out an automated text analysis
6. Access and analyze social media-based data related to a topic of interest
7. Understand how issues of equity, privacy, and ethics are central to data science in education
8. Develop a personal learning and development plan related to data science in education

These objectives will serve as a foundation for later data science in education-related courses, including data visualization, creating interactive web applications, and machine learning applications.

Textbook for the Proposed Course:

Estrellado, R. A., Freer, E. A., Mostipak, J., Rosenberg, J. M., & Velásquez, I. C. (in press). *Data science in education using R*. London, England: Routledge. Freely-available from:

<http://www.datascienceineducation.com/>

References

- Mishra, P., Koehler, M.J., & Greenhow, C. (2015). *The work of educational psychologists in a digitally networked world*. In L. Corno & E.M. Anderman (Eds.) *Handbook of Educational Psychology* (3rd edition) (pp. 29-40). New York: Routledge.
- Rosenberg, J. M., Lawson, M. A., Anderson, D. J., & Rutherford, T. (2020). *Making data science count in and for education*. In E. Romero-Hall (Ed.), *Research Methods in Learning Design & Technology*. Routledge: New York, NY.
- Schneider, B., Reilly, J., & Radu, I. (2020). Lowering barriers for accessing sensor data in education: Lessons learned from teaching multimodal learning analytics to educators. *Journal for STEM Education Research*, 1-34.
- Siemens, G., & Baker, R. S. D. (2012, April). *Learning analytics and educational data mining: towards communication and collaboration*. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254).
- Wise, A. F. (2020). Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences*, 29(1), 165-181.