Old Dominion University

ODU Digital Commons

5-2021

Review of: Estrellado, R. A., Freer, E. A., Motsipak, J., Rosenberg, J. M., & Velásquez, I. C. (2021). Data science in education using R. London, England: Routledge

Esther Michela
University of Tennessee, Knoxville, esther.michela@gmail.com

Robert L. Moore
Old Dominion University, robmoorephd@gmail.com

Follow this and additional works at: https://digitalcommons.odu.edu/stemps_fac_pubs

Part of the Educational Assessment, Evaluation, and Research Commons, Educational Methods Commons, Educational Technology Commons, Online and Distance Education Commons, and the Other Education Commons

Original Publication Citation

Review of: Estrellado, R. A., Freer, E. A., Motsipak, J., Rosenberg, J. M., & Velásquez, I. C. (2021). Data science in education using R. London, England: Routledge

*Esther Michela – esther.michela@gmail.com | ORCID: http://orcid.org/0000-0002-9080-3530

Robert L. Moore – robmoorephd@gmail.com | Twitter: http://www.twitter.com/robmoore3 | ORCID: http://orcid.org/0000-0002-5645-9297

Citation:

Abstract: Data use in education is continually expanding in scope (Romero and Ventura 2020), and though programming resources are plentiful, those that truly scaffold the learning process are scarce. *Data Science in Education Using R* by Estrellado, Freer, Motsipak, Rosenberg, and Velásquez is a combination primer and workbook. This format allows the volume to serve as both an introductory R resource aimed to help readers overcome the daunting initial roadblocks to proficiency and familiarize them with educational data uses. We reviewed this book from the perspective of a beginning R user with little other programming background for whom this book is an introduction to the field. As an open book offering, it is available in physical form and digitally at: https://DataScienceInEducation.com.

Review of: Estrellado, R. A., Freer, E. A., Motsipak, J., Rosenberg, J. M., & Velásquez, I. C. (2021). *Data science in education using R*. London, England: Routledge.

Esther Michela & Robert L. Moore

Introduction

Data use in education is continually expanding in scope (Romero and Ventura 2020), and though programming resources are plentiful, those that truly scaffold the learning process are scarce. *Data Science in Education Using R* (Estrellado et al., 2021) is a combination primer and workbook. This format allows the volume to serve as both an introductory R resource aimed to help readers overcome the daunting initial roadblocks to proficiency and familiarize them with educational data uses. We reviewed this book from the perspective of a beginning R user with little other programming background for whom this book is an introduction to the field. As an open book offering, it is available in physical form and digitally at: https://DataScienceInEducation.com.

Overview of Book Organization

Three objectives provide structure to the organization of the chapters: 1) introduce the field of data science in education, 2) provide exercises to teach data science methods using R, and 3) address the practical application of data science in the workplace. The first four chapters constitute an introduction and overview of professional considerations for education data scientists. The next eight chapters comprise the main section of the book, including two chapters on getting started with R and RStudio. Each of these eight chapters includes a walkthrough scenario using freely available data with step-by-step instructions and screenshots, each focusing on a different concept within R. The final three chapters return to topics of data science application, teaching data science, and continued personal learning.

Content Review

The four introductory chapters focus on the larger field of data science and its application to the field of education. The book uses an informal, first-person tone, which feels like a coach walking you through each step of the process. A repeated disclaimer throughout these chapters is that this book is not intended to teach statistics or how to use R for other purposes. Instead, this book is meant to give the reader the foundational skills necessary to leverage R for their research purposes. In the third chapter, the authors provide several examples of ways data science can look and be approached in education. Chapter four focuses on the complexity of data science in education, from institutional application to ethical concerns regarding using students' data. It is important for anyone using such data to be aware of these issues, and though this book cannot address them in depth, it does suggest other resources for further study. Chapters five and six address the practical steps for setting up R and RStudio and the fundamental principles of using R.

Chapters seven through 14 consist of data science modeling through detailed walkthroughs. Each walkthrough is organized to loosely follow Wickham and Grolemund's (2018) recommended steps for performing data science: import data, tidy data, transform data, visualize data, model data, and communicate results. These steps will be useful for readers looking for more information on data science principles. The beginning of each walkthrough provides an overview of the topics, functions and vocabulary before delving into the content. Detailed background explanations on the format and type of data used also clarify the applied functions. Following the introduction and background, each walkthrough follows a similar structure: details on the methods; instructions for how to import data, load packages and libraries, process data, run analyses, and model the data; and a discussion of the data analysis.

The walkthrough chapters draw upon three sources of education data: "Student Perceptions of Learning," "Student Performance Data," and "Publicly Available Data." These datasets are freely available and representative of data that might commonly be encountered by the intended users. This is practical in that it alerts readers to the many datasets publicly available for practice and orients them to the kinds of information that are relevant in an educational context. The final chapters of the book (15 - 19) delve into other issues of data science in education, such as applying skills to your job and teaching data science to others.

A key feature of this book is its tone. The authors seem to acknowledge beginning programmers' stress and anxiety and speak directly to the readers using first-person language. The result is feeling like a coach or teacher is explaining the concepts and helping readers overcome early learning hurdles. Other helpful elements include screenshots of code in RStudio, listing common coding errors and solutions, and posing reflective questions to help readers think critically about the process and what they are seeing in the data.

With beginners in mind, the authors could have more effectively addressed the large amount of vocabulary and programming concepts that must be learned. The first author found herself trying to scan through the chapters in the print copy to find the definitions of words listed at the beginning of each walkthrough, which was not efficient. In the digital version, terms are searchable using the "find" command, but adding definitions or page numbers to the vocabulary list would make a more complete reference in chapter intros.

Conclusion

As in the field of instructional design, many people come to data science via varying career paths. The work required in each data science job will be different, which makes the broad reach of this book applicable to people in different stages. This book is most useful for beginners in R or those who want to learn more about how data science applies in education. Any programming knowledge will be helpful to get started, as the language and process of programming is

complicated. The authors attend to the needs of beginning learners by addressing common pitfalls, but undeniably, learning to code will still involve much trial and error. The walkthroughs remove much of the potential for error by illustrating each step precisely, but applying the skills to other data is the reader's responsibility. A static book cannot answer every question or debug code for you, but it can and does model ways of effectively approaching data science questions in education. While it will not be the only data science in R resource a beginner will need, either version of this book will be a convenient and straightforward guide for those interested in using open educational resources. The curated exercises provide the scaffolding beginning R users need to get started on their journey.

References

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIRES Data Mining and Knowledge Discovery*. https://doi.org/10.1002/widm.1355

Wickham, H. & Grolemund, G. (2018). *R for data science.* O'Reilly Media.