

Engaging with MIDFIELD Data

Susan Lord, Matthew Ohland, Marisa Orr, Richard Layton, and Russell Long

2021-07-26

Contents

| | | |
|----------|-----------------------------------|----------|
| 1 | Introduction | 5 |
| | Objectives | 5 |
| | Description | 5 |
| | Pre-conference homework | 6 |
| | Agenda | 6 |
| | Facilitators | 6 |
| | Licenses | 7 |
| | Acknowledgement | 7 |
| 2 | Get started with R | 9 |
| | Why R? | 9 |
| | Install R and RStudio | 9 |
| | Install packages | 10 |
| | References | 10 |

Chapter 1

Introduction

Objectives

MIDFIELD—The *Multiple-Institution Database for Investigating Engineering Longitudinal Development*—is a partnership of higher education institutions with engineering programs. MIDFIELD contains student record data from 1988–2017 for approximately one million undergraduate, degree-seeking students at the partner institutions.

The goal of this workshop is to make MIDFIELD more accessible to the ASEE community. The workshop introduces *midfieldr* (a package in the R software environment) that provides access to a MIDFIELD student-record data sample and tools to analyze and graph persistence metrics such as graduation rates. The workshop is designed for R beginners.

By the end of the workshop, participants should be able to:

- Describe key variables in MIDFIELD data tables
- Select academic programs and populations to study
- Use *midfieldr*, an R package specifically designed for use with MIDFIELD
- Explore and tell a story from MIDFIELD data
- Explain key features of effective data displays

Description

The robustness of the MIDFIELD data allows us to emphasize an intersectional approach to the study of student records, permitting multiple categories of inequity such as race/ethnicity and sex to be considered simultaneously.

To introduce beginners to R, participants work through a self-paced tutorial covering basic elements of the R computing language and environment. To

introduce `midfieldr` and using it to work with student record data, participants work through a “Get started” tutorial in which they determine the numbers of students ever enrolled in two programs, group and summarize the data, and graph the results.

For more experienced R users or anyone working at a faster pace, we offer a series of self-paced tutorials that introduce key features of `midfieldr` and how they are applied to compute persistence metrics and graph results.

We also discuss the merits of the multiway graph design that is recommended for displaying results of this type. The agenda includes an interactive session to demonstrate contemporary principles of effective data display.

Pre-conference homework

To get the most out of the workshop, you should have the essential software installed and running several days before the workshop to give you time to contact us with questions if anything goes amiss.

Your homework is to install R and RStudio and the MIDFIELD R packages `midfielddata` and `midfieldr`. Instructions are given on the Get started with R page.

Agenda

Our three hours are organized approximately as shown.

| Min | Topic |
|-----|---|
| 10 | Introduction |
| 35 | Finding stories in the data |
| 35 | Getting started with R |
| 15 | Break |
| 20 | Designing effective displays |
| 55 | Getting started with <code>midfieldr</code> |
| 10 | Next steps & assessing the workshop |

Facilitators

Susan Lord Director of the MIDFIELD Institute and Professor and Chair of Integrated Engineering at the University of San Diego. She is a Fellow of the IEEE and the ASEE. Dr. Lord has considerable experience facilitating workshops including the National Effective Teaching Institute (NETI) and special sessions at FIE. (slord@sandiego.edu)

Matthew Ohland MIDFIELD Director and Principal Investigator. He is Professor and Associate Head of Engineering Education at Purdue University and a Fellow of IEEE, ASEE, and AAAS. Dr. Ohland has considerable experience facilitating workshops including the NETI and CATME training. (ohland@purdue.edu)

Marisa Orr MIDFIELD Associate Director and Associate Professor in Engineering and Science Education with a joint appointment in Mechanical Engineering at Clemson University. She received the 2009 Helen Plants Award for the best nontraditional session at FIE, “Enhancing Student Learning Using SCALE-UP Format.” (marisak@clemson.edu)

Richard Layton MIDFIELD Data Visualization Specialist and Professor Emeritus of Mechanical Engineering at Rose-Hulman Institute of Technology. He is the lead developer of the R packages used in this workshop. Dr. Layton has considerable experience facilitating workshops, including FIE workshops on data visualization (2014) and midfieldr (2018). (graphdoctor@gmail.com)

Russell Long MIDFIELD Managing Director and Data Steward. He developed the stratified data sample for the R packages used in this workshop. Mr. Long is a SAS expert with over twenty years of experience in institutional research and assessment. (ralong@purdue.edu)

Licenses

The following licenses apply to the text, data, and code in these workshops. Our goal is to minimize legal encumbrances to the dissemination, sharing, use, and re-use of this work. However, the existing rights of authors whose work is cited (text, code, or data) are reserved to those authors.

- CC-BY 4.0 for all text
- GPL-3 for all code
- CC0 for all data

Acknowledgement

Funding provided by the National Science Foundation Grant 1545667 “Expanding Access to and Participation in the Multiple-Institution Database for Investigating Engineering Longitudinal Development.”

top of page

Chapter 2

Get started with R

If you already have R and RStudio installed, please update to the most recent releases and update your R packages as well.

If you are trying R for the first time, it is vital that you attempt to set up your computer with the necessary software in advance or it will be difficult to keep up.

Why R?

R is an open source language and environment for statistical computing and graphics [R Core Team, 2021], ranked by IEEE in 2020 as the 6th most popular programming language (Python, Java, and C are the top three) [Cass, 2020]. If you are new to R, some of its best features, paraphrasing Wickham [2014], are:

- R is free, open source, and available on every major platform, making it easy for others to replicate your work.
- More than 17,500 open-source R packages are available (Jun 2021). Many are cutting-edge tools.
- R packages provide deep-seated support for data analysis, e.g., missing values, data frames, and subsetting.

RStudio, an integrated development environment (IDE) for R, includes a console, editor, and tools for plotting, history, debugging, and workspace management as well as access to GitHub for collaboration and version control [RStudio Team, 2016].

Install R and RStudio

Unless noted otherwise, we assume the reader is an R novice. Thus the first steps are to install R and RStudio. Windows users may have to login as an

Administrator before installing the software.

- Install R for your operating system
- Install RStudio, a user interface for R

Once the installation is complete, you can take a 2-minute tour of the RStudio interface.

- Let's start (00:57–02:32) by R Ladies Sydney [Richmond, 2018]

The same video includes a longer (7 minute) tour of the four quadrants (panes) in RStudio if you are interested.

- The RStudio quadrants (07:21–14:40) by R Ladies Sydney [Richmond, 2018]

Install packages

References

[top of page](#)

Bibliography

Stephen Cass. Top programming languages. *IEEE Spectrum*, July 22 2020. URL <https://spectrum.ieee.org/at-work/tech-careers/top-programming-language-2020>.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://www.R-project.org/>.

Jenny Richmond. Basic Basics Lesson 1: An opinionated tour of RStudio, 2018. URL <https://youtu.be/kfcX5DEMAp4>.

RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2016. URL <http://www.rstudio.com/>.

Hadley Wickham. *Advanced R*. Chapman & Hall/CRC The R Series. Taylor & Francis, 2014. ISBN 9781466586963.