Engaging with MIDFIELD Data

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

2021-07-26

Contents

Chapter 1

Introduction

1.1 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
- Explore and tell a story from MIDFIELD data
- Use midfieldr, an R package specifically designed for use with MIDFIELD
- Explain key features of effective data displays

1.2 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.

1.3 Pre-workshop 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 explained on the pages:

- Install everything
- [Setup a project]

1.4 Agenda

Our three hours are organized approximately as shown.

Min	Topic
15	Introduction
35	Exploring the data structure
35	Working with R
15	Break
20	Designing effective displays
50	Working with R (continued)
10	Next steps & assessing the workshop

1.5 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)

1.6 Licenses

1.7 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

Install everything

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.

Your pre-workshop homework:

- 2.1 Install R and RStudio
- 2.2 Install the remotes package
- 2.3 Install midfielddata
- 2.4 Install midfieldr
- 3.1 Create a project
- 3.2 Add some folders

If you already have R and RStudio installed, this is a great time to

- Make sure your R installation is current
- Make sure your RStudio installation is current
- Update your packages by running:

```
update.packages(ask = FALSE, checkBuilt = TRUE)
```

2.1 Install R and RStudio

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]

2.2 Install the remotes package

The fundamental unit of shareable code in R is the *package.* For the R novice, an R package is like an "app" for R—a collection of functions, data, and documentation for doing work in R that is easily shared with others [Wickham, 2014].

Most packages are obtained from the CRAN website [cra, 2018-04-22]. To install a CRAN package using RStudio:

- Launch RStudio

The RStudio interface has several panes. We want the *Files/Plots/Packages* pane.

- Select the *Packages* tab

Next,

- Click *Install* on the ribbon
- In the dialog box, type the name of the package. For our first package, type `remotes` to install the remotes package [Hester et al., 2021]
- Check the *Install dependencies* box
- Click the *Install* button

In the RStudio Console, you should see a message like this one,

```
    package 'remotes' successfully unpacked and MD5 sums checked
```

If successful, the package will appear in the Packages pane, e.g.,

2.3 Install midfieldr

Earlier we installed the remotes package. To access its functions we use the `library()` function.

```
# type in the RStudio Console
library("remotes")
```

Now we can use the `install_github()` function in the remotes package to install the midfieldr package from GitHub [Layton et al., 2021].

```
# type in the RStudio Console
install_github("MIDFIELDR/midfieldr")
```

You can confirm a successful installation by viewing the package help page. In the Console, run:

```
# type in the RStudio Console
library("midfieldr")
? midfieldr
```

If the installation is successful, the code chunk above should produce a view of the help page as shown here.

2.4 Install midfielddata

Because of its size, the data package is stored in a `drat` repository instead of CRAN. Installation takes time; please be patient and wait for the Console prompt ">" to reappear.

Type (or copy and paste) the following lines in the RStudio Console.

```
# type in the RStudio Console
install.packages("midfielddata",
                 repos = "https://MIDFIELDR.github.io/drat/",
                 type = "source")
# be patient
```

Once the Console prompt ">" reappears, you can confirm a successful installation by viewing the package help page. In the Console, run:

```
# type in the RStudio Console
library("midfielddata")
? midfielddata
```

If the installation is successful, the code chunk above should produce a view of the help page as shown here.

If this step is successful, please continue your homework by completing:

- 3.1 Create a project
- 3.2 Add some folders

top of page

Chapter 3

Set up a project

To begin any project, we create an RStudio *Project* file and directory. You can recognize an R project file by its *.Rproj* suffix.

3.1 Create a project

If you prefer your instructions with commentary,

- Start with a Project (02:34–04:50) by R Ladies Sydney [Richmond, 2018]

If you prefer basic written instructions,

- RStudio, *File > New Project… > New Directory > New Project*
- Or, click the *New Project* button in the Console ribbon,

In the dialog box that appears,

- Type the workshop name as the directory name, for example, `workshop`, or if you like more detail, `midfield-workshop-asee-2021`
- Use the browse button to select a location on your computer to create the project folder
- Click the *Create Project* button

Whenever you work with the workshop materials, launch the `workshop.Rproj` file (using the name you actually used) to start the session.

3.2 Add some folders

While file organization is a matter of personal preference, we ask that you use the directory structure shown here for your work in the workshop. Assuming we

called our project `workshop`, the minimal directory structure has three folders in it plus the `.Rproj` file at the top level.

```
\workshop
    \data
    \results
    \scripts
    workshop.Rproj
```

We use the folders as follows:

- `data` data files
- `results` finished graphs and tabulated data formatted for display
- `scripts` R scripts that operate on data to produce results

If you prefer your instructions with commentary,

- Make some folders (04:50–06:08) by R Ladies Sydney [Richmond, 2018]

If you prefer basic written instructions,

- use your usual method of creating new folders on your machine
- or you can use the *New Folder* button in the Files pane

You finished your homework!

top of page

Chapter 4

Workshop materials

4.1 Stuff

4.2 More stuff

top of page

Chapter 5

Exploring the data structure

5.1 Stuff

5.2 More stuff

top of page

Chapter 6

Working with 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.
- R packages provide effective tools for data analysis and visualization.
- More than 17,750 open-source R packages are available (Jun 2021). Many are cutting-edge tools.

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].

6.1 Prerequisites

Before proceeding, you should have completed

- Install everything
- Set up a project
- Launched your workshop project—`workshop.Rproj` or other name that you selected—to start the R session
- We suggest you start a new R script for each tutorial and save it to the `scripts` directory. For example, at the end of the workshop, your scripts directory might contain the following files:

```
        \scripts
            \R-basics.R
            \getting-started.R
            \case-study-programs.R
            \case-study-students.R
            etc.
```

6.2 New to R?

Prerequisites should be completed before proceeding. By the end of the workshop, our R beginners will have made progress on two or possibly three tutorials:

- R basics An introduction to R, generally less than an hour.
- Getting started: An introduction to the MIDFIELD practice data tables.
- Case study programs Construct a data frame of program CIP codes and program names for four engineering programs (Civil, Electrical, Industrial, and Mechanical)

If you complete these tutorials and there is still time remaining, please consider moving on the to tutorials listed in the After the workshop section.

6.3 Familiar with R?

Prerequisites should be completed before proceeding. By the end of the workshop, our more experienced R users will have made substantive progress on two or possibly three tutorials:

- Getting started: An introduction to the MIDFIELD practice data tables
- Case study programs Construct a data frame of program CIP codes and program names for four engineering programs (Civil, Electrical, Industrial, and Mechanical)
- Case study students Develop a data frame of the case study students who pass the data sufficiency criterion.

If you complete these tutorials and there is still time remaining, please consider moving on the to tutorials listed in the After the workshop section.

6.4 After the workshop

At his point, the learning is self-directed. Choose the skills you want to continue working on. We have a set of tutorials for

- Developing R skills
- Continuing the case study
- Exploring midfieldr functions

6.4.1 Developing R skills

The basic skills tutorials take about 50 minutes each.

- R basics
- Graph basics

- Data basics

6.4.2 Continuing the case study

The case study is a quick tour of a typical workflow using student unit record data. This is a "big picture" development—functions are used without detailed explanations or development so that we can get to the results with as little distraction as possible. Anyone wanting more detail will find it in the detailed vignettes (links below in Exploring midfieldr functions).

- Case study programs
- Case study students
- Case study stickiness
- Case study graduation rate

6.4.3 Exploring midfieldr functions

Deep dive into the midfieldr functionality. The work flow follows the same general pattern as the quicker case study, but pauses to explore each function in more detail, exploring the arguments and strategies for use. In general, each tutorial is self-contained so you may enter at almost any point.

- Program codes and names Practice strategies of searching `cip` for programs we want to study.
- Subsetting MIDFIELD data Use programs codes to subset the MIDFIELD data tables.
- Data sufficiency What it is and how it is applied to student unit-record (SUR) data.
- Timely completion What it is and how it is applied to SUR data.
- FYE programs What they are and how they are accommodated with SUR data.
- Multiway graphs How to graph and interpret a common data structure encountered when working with SUR data.
- Tabulating data How to tabulate multiway data for publication.

6.5 R basics

text

top of page

Chapter 7

Designing effective displays

7.1 Stuff

7.2 More stuff

top of page

Bibliography

The Comprehensive R Archive Network, 2018-04-22. URL https://cran.r-project.org/.

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

Jim Hester, Gábor Csárdi, Hadley Wickham, Winston Chang, Martin Morgan, and Dan Tenenbaum. *remotes: R Package Installation from Remote Repositories, Including GitHub*, 2021. URL https://CRAN.R-project.org/package=remotes. R package version 2.4.0.

Richard Layton, Russell Long, and Matthew Ohland. *midfieldr: Tools for Studying MIDFIELD Student Unit Record Data in R*, 2021. https://github.com/MIDFIELDR/midfieldr, https://engineering.purdue.edu/MIDFIELD.

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.