Engaging with MIDFIELD Data

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

2021-07-26

Contents

Chapter 1

Introduction

Objectives

The main goal of the workshop is making 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
- Select academic programs and populations to study
- Use midfieldr, an R package specifically designed for use with MIDFIELD, to compute persistence metrics
- 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.

Agenda

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 midfieldr
10	Next steps & assessing the workshop

Before you arrive

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

```
htmltools::img(src =
knitr::include_graphics("figures/NSF_4-Color_bitmap_Logo.png"),
               alt = 'NSF logo',
               width = 90)
```

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 (localmgr) before installing the software.

- Install R for your operating system

- Install RStudio, a user interface for R

References

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/.

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.