

Tidyverse in R

Workshop Description

Brian Fogarty
Center for Social Science Research
Lucy Family Institute for Data & Society
University of Notre Dame
bfogart2@nd.edu

2022 Spring Semester

Overview

This 3-session workshop is designed as an introduction to the `tidyverse` packages in R. `Tidyverse` is a set of R packages that comprehensively deal with data wrangling and visualization. We will cover some of the most common packages and functions in `tidyverse` using a variety of social science data. It is possible that your particular research needs are not covered in this workshop, but the materials will provide you a solid foundation and understanding for learning on your own.

Prior Knowledge

The course assumes a working knowledge of R and RStudio. We will also be using R Markdown, but it is not critical for learning the material.

Software Details

Make sure you are using R 4.0 (the specific version should not matter) as `tidyverse` has been deprecating its packages for R 3.0. You should not need to update RStudio to the latest version as long as R 4.0 is working correctly in your version.

You should also install `tidyverse` and the other packages we are using prior to the first session. To do so, run this code:

```
install.packages(c("tidyverse", "haven", "readxl", "scales", "rmarkdown", "writexl"))
```

Workshop Plan and Dates

The workshop will be spread out over three weeks with one hour-long session a week. Each session will review the topic, cover the main R functions, and go through a few live demos. Following each session, an annotated lab handout will be provided with a small problem set for you to practice on your own.

Course Dates:

- Week 1: Wednesday, February 9, 2022, 1-2pm
- Week 2: Wednesday, February 16, 2022, 1-2pm
- Week 3: Wednesday, February 23, 2022, 1-2pm

Workshop Attendance

As this is a voluntary, informal workshop, there is no requirement to attend all the sessions. However, since each week builds off of the previous week(s), please only register if you expect to attend all three sessions.

Workshop Delivery

The workshop will be completely virtual using Zoom. All workshop materials will be provided in a shared folder in Google drive.

Topics

1. Introduction to Tidyverse & Data Set Management - Week 1

Along with an introduction to tidyverse, we will examine reading-in, merging, organizing, and saving datasets. The topics covered include:

- Reading-in and writing to different data formats (e.g., .csv, .dta, .sav)
- Pivoting datasets
- Merging and splicing different datasets
- Subsetting datasets

2. Variable Manipulation - Week 2

We will look at various ways to perform variable manipulation (i.e., recoding variables). One of the main benefits of using `tidyverse` for this is the ability to carry out numerous recoding tasks simultaneously thus allowing greater flexibility and precision. The topics covered include:

- Creating dummy variables
- Collapsing variables
- Labeling variable values
- Working with missing values
- Changing variables' functional form

3. Data Visualization - Week 3

Finally, we will pull together everything we learned in the previous sections and work through a variety of data visualization techniques. The topics covered include:

- Bar Plots
- Histograms
- Smoothed Density Plots
- Scatterplots
- Faceting & Combining Plots

Registration

To sign-up for the course, please fill in this Google form by **Monday, February 7th**: https://docs.google.com/forms/d/e/1FAIpQLScMHstr4EMgxSJO6vI_Idl0F3lu9fTMDHtJkKol1i0lpWrR7A/viewform