

Spring 2021 BDAA DataMastery R Syllabus

Course Description

Training in the foundational tools necessary for data analysis in R, such as data manipulation, and data visualization. This course will be focused on the application of the Tidyverse, as students will be equipped with the knowledge to use software tools such as dplyr, ggplot2, and readR. The course emphasizes the use of hands on learning through live workshops with code along exercises to aid in retention in material. The course is structured to target people from all different backgrounds, and with a wide variety of computer science experience, from beginner to expert.

Course Learning Outcomes

By the end of this course, students should successfully be able to:

- Understand what data science is, and the overall workflow in a project
- Use dplyr for data manipulation and data cleaning
- Explore data with data visualization techniques using ggplot2
- Be able to take a dataset, frame a question, and use the given tools to tell a story with data

Course Technology

- Computer: current Mac (OS X) or PC (Windows 10+) with high speed internet connection
- Webcam: built-in or external webcam, fully installed
- Microphone: built-in laptop or tablet mic or external microphone

Necessary Software

- This course will be using RStudioCloud as the R Environment for workshops.
- Members should make sure they have an RStudioCloud account made prior to workshops in order to access notebooks.

Course Delivery

Workshops will be held every Thursday on Zoom, from 7:30pm-8:30pm. The workshops will be recorded, and uploaded onto the Canvas page for any asynchronous learners. Further resources to dive deeper into topics covered in a workshop will be posted as a separate module, such as books and free courses. After 2 weeks of instruction, there will be a short, brief carmen quiz to check for understanding. Note: This is not something to stress over, rather it is just to make sure there is a general understanding of the processes and ideas that were covered in the workshop. These will usually be small coding snippets on a small dataset.

Course Content

Intro to Data Manipulation in R

- What is the Data Science Workflow?
 - How does Data Manipulation fit?
- Why do we do Data Manipulation and Data Cleaning?
- Introduction to the Tidyverse
- Introduction to dplyr for data manipulation
- Reading data with readR
- Live Coding Exercises
 - How to inspect a data frame
 - How to check properties of columns
 - Calculating summary statistics
 - Subsetting Data
 - Renaming Columns
 - Arranging Data
 - Distinct Column Values

Intro to Data Visualization in R

- How does Data Visualization fit in the workflow?
- Why is Data Visualization Important?
- Introduction to ggplot2
- The grammar of graphics of ggplot2
- Exploratory Data Analysis with ggplot2
- Live Coding
 - Histogram
 - Line Plots
 - Scatter Plots
 - Barcharts

Intermediate Data Manipulation in R

- Advanced dplyr verbs
 - mutate
 - filter
 - group_by
 - top_n
- Checking missing values
 - is.na
 - !is.na
 - Group selecting with %in%
- Chaining Multiple verbs together
- data manipulation exercises

Intermediate Data Visualization in R

- Advanced Plotting
 - Stacked Barcharts
 - Coordinate flipping
 - Facet Wrapping
 - Facet Grid
 - Plotting regression lines
 - Pie Charts
 - Themes
 - Legends

Advanced Data Manipulation in R

- Joining Data Frames in dplyr
- Methods for joining Data Frames
 - Joining on a key
 - Inner Joins
 - Left Joins
 - Right Joins
 - Outer Joins
 - Full Joins
- Concatenating DataFrames
 - Horizontally
 - bind_rows
 - Vertically
 - bind_cols