Syllabus

Cőde Bootcamp
Data Science in R

Collin M. McCabe, Ph. D.
collinmichaelmccabe@gmail.com

Schedule:

1. Wednesday, May 9th – Back to base-ics: Installing and getting familiar with base R
    a. Math
    b. Variables
    c. Functions
2. Thursday, May 10th – Structuring your thoughts: Introduction to data structures in base R
    a. Vectors
    b. Matrices
    c. Data frames
    d. Lists
3. Friday, May 11th – "if" only "for" a "while": Control flow in base R
    a. Conditionals
    b. If-else statements
    c. Loops

--- WEEKEND ---

4. Monday, May 14th – Putting the data in "Data Science": Importing from files, databases, web
    a. "Flat" files
    b. Relational databases
    c. Web APIs
5. Tuesday, May 15th – Tidying up a bit: Introduction to "tidy" data and cleaning data with tidyr
    a. Tidy data and tibbles
    b. Gathering/spreading
    c. Data maintenance: type errors, NAs, etc
    d. Factors as a special data class
6. Wednesday, May 16th – Being manipulative: Using dplyr to manipulate data in R
    a. Subsetting with select and filter
    b. Reordering with arrange
    c. Deriving variables with mutate
    d. Linking with pipes
7. Thursday, May 17th – Getting visual: Data visualization and exploratory analysis in R
    a. Summary statistics and summarise
    b. Visualizing distributions
    c. Visualizing comparisons
8. Friday, May 18th – Spot the differences: Statistical analyses of categorical data in R
    a. Pure categorical: Chi-square test
    b. Simple comparisons: t-test
    c. Multiple comparisons: ANOVA

--- WEEKEND ---

9. Monday, May 21st – Spot the trends: Statistical analyses of continuous data in R
    a. Assumptions: normality and variance
    b. Correlation
    c. Simple linear regression
    d. Multiple linear regression
10. Tuesday, May 22nd – Your cRystal ball: Predictive modeling and machine learning in R
    a. Dimensionality reduction and clustering
    b. Categorical predictions: Classification
    c. Continuous predictions: Regression

Course Material:

Presentations, R code, and datasets will be available from the course GitHub repository:

https://www.github.com/collinmmccabe/bootcampR

This will be updated before each day of class with the day's notes and code, so make sure to pull from the master branch before each class. You will also be free to create your own branch on this repository to keep track of your progress throughout the course.

Instructor Availability:

I will do my best to be around during the hackathon and code sprint to answer any questions or solve any problems that might arise. If, however, I'm not around and you can't figure something out, feel free to email your questions to me at collinmichaelmccabe@gmail.com – I'll respond to your query within 24 hours. If your question involves some sort of bug in your code, please attach the code in question along with any data you're using to the email.

Useful Resources:

R for Data Science - http://r4ds.had.co.nz/index.html

Cookbook for R - http://www.cookbook-r.com/

R Programming Wikibook - https://en.wikibooks.org/wiki/R_Programming

CRAN website - https://cran.r-project.org/

RStudio website - https://www.rstudio.com/

Tidyverse website - https://www.tidyverse.org/

Stack Overflow (for programming questions) - https://stackoverflow.com/questions/tagged/r

Cross Validated (for statistics questions) - https://stats.stackexchange.com/questions/tagged/r