

# Cõde Bootcamp 2019 - Data Science in R, v2.0

Collin M. McCabe, Ph. D. - [collinmichaelmccabe@gmail.com](mailto:collinmichaelmccabe@gmail.com)

## Schedule

- Monday, May 6th - TeaseR: A showcase of the awesome things we'll be doing this month
  - GitHub setup
  - Rstudio tour
  - Showcase of R products
- Tuesday, May 7th – Back to base-ics: Installing and getting familiar with base R
  - Math
  - Variables
  - Functions
- Wednesday, May 8th – Structuring your thoughts: Introduction to data structures in base R - *Guest Instructor: Matthew Osborne*
  - Vectors
  - Matrices
  - Data frames
  - Lists
- Thursday, May 9th – “if” only “for” a “while”: Control flow in base R - *Guest Instructor: Matthew Osborne*
  - Conditionals
  - If-else statements
  - Loops
- Monday, May 13th – Putting the data in “Data Science”: Importing from files, databases, web
  - “Flat” files
  - Relational databases
  - Web APIs
- Tuesday, May 14th – Tidying up a bit: Introduction to “tidy” data and cleaning data with tidyr
  - Tidy data and tibbles
  - Gathering/spreading
  - Data maintenance: type errors, NAs, etc
  - Factors as a special data class

- Wednesday, May 15th – Being manipulative: Using dplyr to manipulate data in R
  - Subsetting with select and filter
  - Reordering with arrange
  - Deriving variables with mutate
  - Linking with pipes
- Thursday, May 16th – Getting visual: Data visualization and exploratory analysis in R
  - Summary statistics and summarise
  - Visualizing distributions
  - Visualizing comparisons
- Monday, May 20th – Spot the differences: Statistical analyses of categorical data in R
  - Pure categorical: Chi-square test
  - Simple comparisons: t-test
  - Multiple comparisons: ANOVA
- Tuesday, May 21st – Spot the trends: Statistical analyses of continuous data in R
  - Correlation
  - Simple linear regression
  - Assumptions: normality and variance
  - Multiple linear regression
- Wednesday, May 22nd – Your cRystal ball: Predictive modeling and machine learning in R
  - Categorical predictions: Classification
  - Continuous predictions: Regression
  - Dimensionality reduction and clustering
- Thursday, May 23rd - wRap-up: Putting it all together to make production-ready analyses
  - Brief review of material
  - Pipelining
  - Automation
  - Publishing

## Course Material

Presentations, R code, datasets, and take-home exercises will be available from the course GitHub repository:

<https://www.github.com/collinmmccabe/bootcampr>

This will be updated before each day of class with new versions of the day's notes, code, and take-home exercises, 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 and into it from master to keep track of your progress throughout the course (Don't worry if you don't know what this means yet, you will after the first day of class).

## 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](mailto: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
- Cookbook for R
- R Programming Wikibook
- GitHub page for caret
- CRAN website
- RStudio website
- Tidyverse website
- Stack Overflow (for programming questions)
- Cross Validated (for statistics questions)