

R bootcamp - August 2021: Syllabus/schedule

August 21, 2021

The bootcamp will be held August 21-22, so Day 1 is August 21 and Day 2 is August 22.

Unless otherwise noted, modules are about 75 minutes long: 45 minutes for presentation, 20 minutes for breakout and 10 minutes for discussion of solutions.

- Day 1 morning (8:30-12:15) (learning R)
 - Module 0: Introduction, what is R, starting R, why R? why not R? (Chris) (15 minutes)
 - Module 1: Basics of R, with Rstudio (Chris)
 - * R as a calculator
 - * helpful shortcuts: tab-complete, up arrow, Ctrl-{up arrow}
 - * vectors and indexing and subset assignment
 - * some basic functions; help()
 - * vectorized calculations, comparisons
 - * basic R objects: vectors, matrices, dataframes, lists
 - * basic graphics
 - * breakout problems
 - Break (15 minutes)
 - Module 2: Managing R and your analyses (Chris) (45 minutes)
 - * managing R objects, the R workspace
 - * using packages (installing, loading, namespaces)
 - * the working directory and basic file reading/writing
 - * Git, Github and version control
 - * getting R help online

- * breakout problems
- Module 3: Working with data (Corrine) (45 minutes)
 - * dataframes/matrices
 - * attributes, missing values and factors
 - * subsetting
 - * strings
- Lunch (on your own) (12:00-1:30)
- Day 1 afternoon (1:30-5:00) (data processing and manipulation)
 - Module 3: Working with data, continued (Corrine) (40 minutes)
 - * more on reading data
 - * breakout problems
 - Module 4: Calculations (Chris)
 - * vectorized calculations and efficiency
 - * apply, lapply
 - * tabulation, stratified analyses, aggregation, merging data
 - * breakout problems
 - Break (15 minutes)
 - Module 5: Programming in R (Chris)
 - * loops, if-else
 - * writing your own functions, function arguments, functions as objects
 - * basic scoping and environments
 - * breakout problems
- Day 2 morning (9-12:45) (programming and data analysis)
 - Module 6: Data manipulation using the tidyverse (Corrine)
 - * stratified analyses: groupwise operations and split-apply-combine using dplyr
 - * reshaping and tidying data
 - * breakout problems/homework
 - Break (15 minutes)

- Module 7: Data analysis (Chris)
 - * regression, GLMs
 - * smoothing
 - * optimization
 - * simulation, sample()
 - * dates and times
 - * breakout problems
- Module 8: Graphics (Chris)
 - * exporting graphics (vector/raster formats)
 - * lattice graphics
 - * ggplot2
 - * breakout problems
- Lunch (on your own) (12:45-2:00)
- Day 2 afternoon (2:00-4:30) (more advanced topics)
 - Module 9: Workflows, coding practices, and project management (Chris) (60 minutes)
 - * debugging, timing, memory use
 - * scripting, source(), batch jobs
 - * good coding practices
 - * reproducible research
 - Break (fill out feedback forms) (20 minutes)
 - Module 10: Advanced topics morsels (Chris) (60 minutes)
 - * object-oriented programming (S3, S4, ReferenceClasses)
 - * computing on the language (using R to write and evaluate R code)
 - * errors and try-catch
 - * encodings
 - * working with databases
 - * parallel processing: foreach, parApply, RNG issues
 - Module 11: Wrapping up (Chris) (15 minutes)
 - * R inconsistencies and different ways to do things
 - * Where to learn more (campus and non-campus resources)