Lesson 01 - Introduction to the class and R

Robin Donatello

Last Updated 01-21-2019

Course Description

This course is designed as a primer to get the complete novice up and running with the basic knowledge of how to use the statistical programming language R in an environment that emphasizes reproducible research and literate programming for data analysis.

The target audience is anyone who wants to do their own data analysis. The course will cumulate with an peer-evaluated exploratory data analysis on either a pre-specified data set or your data set of choice.

Schedule of Topics

- Intro to the R language and it's packages
- R Studio suite of awesomeness
- Reproducible Research with R Markdown
- Functions
- Getting data into R
- Managing the spreadsheet-like data frame
- Data Visualization
- Data Manipulation and Aggregation
- Exploratory Data Analysis

Logistics

- All content is fully online on the course website: https://norcalbiostat.github.io/MATH130/
- We will also use Data Camp (www.datacamp.com) is a free online learning tool we will use to start learn R.
- Class time is spent expanding on ideas and concepts, working through labs as a class or in pairs, and working on individual assignments.

Grading

- Credit / No Credit. There are 100 points available in this course through attendance, assignments, and a final project.
- You must earn 70 points to receive credit for the course.
- The syllabus provides specific details on points per assignment.

Time Committment

- Fast paced, we're hitting the ground running and only have 5 weeks.
- Not designed to teach you everything about R, just enough to make you dangerous.

- Daily practice will pay off.
- But can be very rewarding
- Garbage in = garbage out
- Effort in = satisfaction

We've all been there. It will get easier

Tool choices

Why use R?

- Open source, and free
- Rapid improvement
- Tons of learning resources
- Currently R is used in Chico's upper division Applied Statistics and Data Science courses.
- Very popular in Data Science, Economics, Natural Sciences, Psychology and many other fields.

Why use R Studio?

- Customizable workspace that docks all windows together.
- Notebook formats that allow for easy sharing of code and output, and integration with other languages (Python, C++, SQL, Stan)
- Syntax highlighting, warning errors when missing a closing parentheses.
- Cross-platform interface. Also works on Windows/iOS/Linux.
- Tab completion for functions. Forget the syntax or a variable name? Popup helpers are available.
- Free training videos available from the developers directly.
- Built in version control
- One button publishing of reproducible documents such as reports, interactive visualizations, presentations (like this one!), websites.

Programming is scary!

Learning to program has other benefits

- Improves your logical skills and critical problem solving
- Increases your attention to detail
- Increases your self reliance and empowers you to control your own research.
- Your PI will love your awesome graphics and reports.
- Some people think what you do is magic.
- Thinking graduate school? [expect to learn this on your own]
- [A few] [other lists] [of reasons]

No joke, this is the key to learning the details of how things work

Why no point and click?

Because it's not reproducible.

- Which boxes did you click last time?
- New data? Gotta do it all over.
- Need to expand your model? Gotta do it all over.
- Made a mistake in the data coding? Gotta do it all over. . .

Getting Help

In Person

- Data Science Initiative
    - http://datascience.csuchico.edu What the Data Science community at Chico is up to.
    - Community Coding. TR 2-4pm THMA 116.
        * Dedicated space to work on coding projects.
        * Collaborate with your peers and learn from experts.
        * Need 1 unit? Enroll in Math 290-02 and attend at least 10 times.
        * You DON'T need to be enrolled to attend!
    - Weekly seminars and workshops in a variety of languages and topics.
- The R Users Meetup group useful if you want to stay connected to the community and learn about upcoming events.
    - https://www.meetup.com/Chico-R-Users-Group/

Sometimes a second pair of eyeballs is all you need

Online

- In RStudio go to `Help` –> `Cheatsheets`
- Chico R Users Google Group. This is our discussion forum to ask (and answer) questions outside of class.
    - https://groups.google.com/forum/#!forum/chico-rug
- Googling error messages: https://www.google.com/search?q=common+R+error+messages
    - Good blog by expert R contributor http://varianceexplained.org/courses/errors/
- Stack Overflow tag [r] https://stackoverflow.com/questions/tagged/r
- R Studio community https://community.rstudio.com/
- Twitter: #rstats @Rstudio

Written

If you're a book kinda person, there is plenty of help available as well. Many have online versions or free PDF's.

- R Markdown, the Definitive Guide: https://bookdown.org/yihui/rmarkdown/
- R for Data Science https://r4ds.had.co.nz/
- Cookbook for R http://www.cookbook-r.com/
- R Graphics Cookbook (I use this all the time) – Chapter 8 in the above link
- The Art of R Programming https://nostarch.com/artofr.htm
- R for... http://r4stats.com/
    - Excel Users https://www.rforexcelusers.com/
    - SAS and SPSS Users http://r4stats.com/books/r4sas-spss/
    - STATA Users http://r4stats.com/books/r4stata/