A Bootcamp for Reproducible Data Analysis using R

Robin Donatello

Math 199 Fall 2016

Introduction to the class

- Target audience: Anyone who wants to do their own data analysis!
- Primer to get the complete novice up and running with the basic knowledge of how to use the statistical programming language R.
- Topics include: R programming basics, importing data, properties of tidy data, visualizing data, reproducible research with Markdown and basic data wrangling.
- Get you up and running with basic knowledge of `R` ASAP.
- MATH 315 uses R heavily. This course is designed as a co- or pre-requisite.

Why use R?

- Free!
- Cross platform.
- Tons of free tutorials.
- The R-project http://www.r-project.org/ is a free open-source programming language that can easily create dynamic graphics for data visualization.
- It is also a flexible statistical analysis toolkit, and provides access to powerful cutting edge analytics.
- The fastest growing statistical analysis program in the Natural Sciences according to the Journal Nature.
- The R community is a robust, vibrant community of users that has grown rapidly in the past few years.
- As has the number of companies who rely on R as their data science platform.

Why use R Studio

- Also free!
- Customizable workspace that docks all windows together.
- 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.
- One button publishing of reproducible documents such as reports, interactive visualizations, presentations (like this one!)

Other ways to learn R

There are literally hundreds of free tutorials on how to learn R.

These lists were created in 2012 but still mostly relevant.

- General guides

- 102 University based tutorials
  - One of note UCLA R Starter kit

A couple interactive tutorials.

- Data Camp: https://www.datacamp.com/
- Try R: http://tryr.codeschool.com/

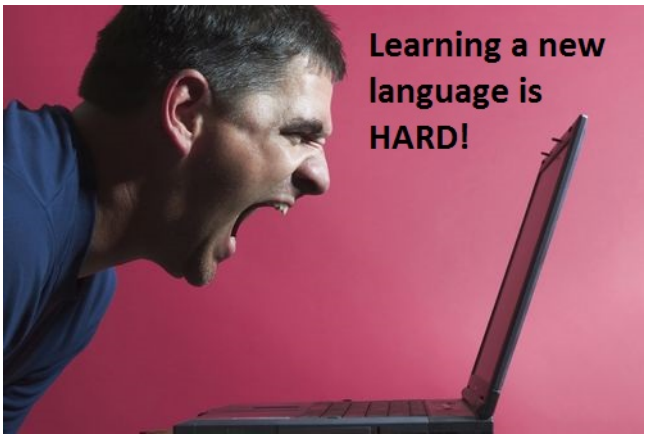
And hundreds of YouTube videos.

Structure of this class

- Content is fully online and available to you now.
- We will meet TR 5-7pm for you to work through content.
- You can work at your own pace, but there are quiz deadlines.
- Weeks 1-3: Interactive coding sessions and non-interactive labs.
- Week 4: Putting it together in an analysis project.

Time Committment

- Fast paced 1 unit course conducted over 7 days.
- Expect to spend 5-7 hours per week if you have never programmed before.



Getting Help

- We will be contributing to the Chico State Statistics Students Google Group. - This is a place to get help from students and instructors at all levels outside class time.
  - Use tags on your posts liberally to make searching for answers easier.
- Your first resource is each other. Your second resource is Google.
- I will **not** answer emails about any coding problems. Post your question on the discussion forum and I or someone else will answer it there.
- Another great resource is StackOverflow using the `r` or `markdown` tags.
- Prepare a minimal working example of your problem on the forum, it may be helpful to include a screenshot of the problem.

You will be responsible for

1. Understanding the requirements of the class.
2. Working through each lab fully.
3. Completing each Swirl lesson.
4. Completing quizzes for each lab.
5. Staying on track.
6. Using the class forum and attending the problem solving sessions to get help.
7. Completing a data analysis project.
8. Evaluating three of your peers on said data analysis project.

Syllabus

- The syllabus for the class can be found on the class website.
- This covers course details such as grading, office hours and required materials.

Getting started!

1. Enroll in the Google Group class forum.
2. Introduce yourself the class on the forum board. State your background and why you are enrolled in this bootcamp.
3. These notes cover how to setup R and R Studio.