A Bootcamp for Reproducible Data Analysis using R

Robin Jeffries

2015-07-10

Introduction to the class

Goals

- 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 management and debugging.
- 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 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 likely still relevant.

- General guides
- 102 University based tutorials

A couple interactive tutorials.

- Data Camp
- Try R

And some YouTube videos.

Structure of this class

- Content is fully online
- Semi structured. Self-paced but with quiz deadlines.
- Weekly R problem solving session (Time/Loc TBD)
- Labs - 3 per week.
- Interactive coding sessions to learn R in R using the Swirl package.
- One quiz per lab. All due by the following Sunday at midnight.

Time Committment

- Fast paced 1 unit course conducted over 3 weeks.
- Learning a programming language is not trivial.
- Expect to spend 10-12 hours per week for these three weeks if you have never programmed before.

Getting Help

- Use the forum and problem solving sessions.
- Your first resource is each other.
- I will **not** answer emails that are in essence "*My code won't run*". Check the forum for a solution first.
- 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. Working through each lab fully.
2. Completing each Swirl lesson.
3. Completing quizzes for each lab.
4. Staying on track.
5. Using the class forum and attending the problem solving sessions to get help.

Syllabus

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

Getting started!

- Once you enroll in Math 199 check the course Blackboard Learn site. Use the link provided there to enroll in the Piazza class forum.
- Review this set of slides to install R and R Studio.
- Introduce yourself the class forum board.
- Download Lab1 and get started coding!