PHP 1560/2560: Statistical Computing with R

Dr. Adam J Sullivan

Fall 2018

E-mail: adam_sullivan@brown.edu   Web: php2560.com

Office Hours: TBD.   Class Hours: Wednesday 1-4pm

Office: School of Public Health - Room 717   Class Room: School of Public Health - Room 375

Course Description

Statistical computing is an essential part of analysis. Statisticians need not only be able to run existing computer software but understand how that software functions. Students will learn fundamental concepts - Data Management, Data types, Data cleaning and manipulation, databases, graphics, functions, loops, simulation and Markov Chain Monte Carlo through working with various statistical analysis. Students will learn to write code in an organized fashion with comments. This course will use a variety of web-based material from:

- php2560.com
- DataCamp
- Online Book
- Github

Course Format

This course will be taught in a "flipped" format. Students will watch a series of videos and work through some simple coding examples before coming to class. The sequence of these will be displayed on the course website as well as a calendar for which videos need to be watched prior to attending class.

The classroom format will focus on labs and projects. You will be involved with computing and coding on a regular basis. Labs will form into projects which you will finish outside of class and turn in for grading.

Flipped Format and References

We need to take some time to discuss what it means to have a flipped format class. In this style of formatting the note taking portion of the course will be done via videos. Think of these as video textbooks in which you gather the basic details before practicing the material.

In the traditional format this course would have the instructor teaching 160 minutes a week with some hands on lab activities but most coding work done outside of class. In the flipped format

you will have access to these lectures as your work prior to class. You then will have 180 minutes of hands on coding projects, code sharing and time with the instructor.

This course will utilize a wealth of materials from many different resources aside from the textbooks:

- The majority of R videos will be hosted by https://www.datacamp.com/courses/1118 and created by the instructor.
- Some notes have been adapted from:
    - Shalizi, C. R. and Thomas, A. C. (2014), Statistical Computing 36-350: Beginning to Advanced Techniques in R
- Shiny is a product of R Studio and will be presented by RStudio with example created by the last 2 classes PHP 2560 class and the instructor.

Prerequisites

Students should have courses in probability and statistical inference at the level of PHP 1510 or PHP 2510.

Recommended Textbooks

- Matloff, Norman (2011). The Art of Programming. No Starch Press.
- Rizzo, Maria L. (2007). Statistical Computing with R. Chapman and Hall/CRC.
- Jones, Owens; Maillardet, Robert & Robinson, Andrew. (2011). An Introduction to Scientific Programming and Simulation Using R. Chapman and Hall/CRC.
- Teetor, Paul. (2011). R Cookbook. O'Reilly Media.

Course Objectives

After the successful completion of this course, you will understand and be able to implement the fundamental principles of statistical computing in R. In particular these include the following capabilities:

1. item Obtain and work with Data.
2. Clean and Transform data into usable dataframes.
3. Create Graphics.
4. Understand the writing and use of functions.
5. Working with larger dataframes efficiently.
6. Perform Statistical Optimizations
7. Code and run an MCMC.
8. Data visualization.
9. Relational Databases

Overall Course Expectations

Students in this course will be expected to do the following:

1. Attend all lectures and actively participate in in class sessions, for every class missed there will be a 5% reduction in overall grade.

2. Complete all assigned flipped material *prior* to coming to class and be prepared to work on in class lab.

3. Complete and turn in all assignments on time. All assignments will be graded on ability of code to work, quality of coding and quality of comments on code.

4. Demonstrate an understanding on material on all projects.

5. Respect each other, each others questions and each others discussion.

6. Peer review other students code.

Evaluation

Students will be evaluated based on:

Grade Category	Percentage
Participation	15%
Pre-Class Assignments	20%
In-Class Projects	20%
R Package	15%
Shiny App	30%

Differences between PHP 1560/2560

Given the nature of this course with multiple levels of students from Undergraduate to PhD, it is important to discuss the differences of expectations and how students will be graded.

Undergraduates

Grade Category	Comments
Participation	Graded the same as all students, Must be in class and prepared to work in groups.
Pre-Class Assignments	Students will be expected to complete a portion of the material with the exception of some more difficult problems which may be attempted but do not have to be complete. Peer Review will be the same.
In-Class Projects	Students will work on the same projects that all other students work on. They will be placed in groups with other students but will not be expected to contribute the same level of coding as graduate students.
R Package	Students will build an R package. Functions may be basic or simple given the amount of statistics taken by this point in time.
Shiny App	Shiny app coding as well as end result will be at an appropriate level for the understanding of statistics and data analysis of the students.

Graduate Students

Grade Category	Comments
Participation	Graded the same as all students, Must be in class and prepared to work in groups.
Pre-Class Assignments	Students will be expected to complete all parts of the assignments. Peer reviews will be thorough and well critiqued.
In-Class Projects	Students will work on the same projects that all other students work on. It is expected that graduate students will contribute more coding to the projects as well as leadership.
R Package	Students will build an R package. Functions are expected to be useful to the area of statistics students are working in. Graduate students will be expected to have more challenging methods and data incorporated into their R package.
Shiny App	Shiny app coding as well as end result will be at an appropriate level for the understanding of statistics and data analysis of the students.

Evaluation Category Details

Participation

Participation will be calculated by the successful completion of videos and practice coding done prior to class as well as being present and engaged during the in class portion.

With the class meeting once a week it is crucial that all students attend. Any student who needs to miss a class must inform the instructor by 9 am the morning of the class. Unexcused absences will result in a 5% reduction in overall course grade.

Pre-Class Assignments

At the end of the videos each week a preview of the in class lab will need to be completed. This will ensure that all students are prepared to work on the material in the lab. Once this assignment is turned in, each student will be required to peer review code of a number of other students. Code will be graded and commented on based on criteria given out in class. Each students grade will be a combination of there own work as well as their peer reviewing of other students code.

In-Class Project

Projects will be a culmination of in class labs with some extra parameters associated with them. Most of the work on Projects will be done in groups

R Package

Another useful skill with R is to take methods, data or other user created tools and turn them into a package. Students will work in groups to create an R Package as directed by the instructor.

Shiny App

An important part of statistics is the visualization and representation of data. Students will be expected to code and build their own Shiny Apps. You will also create a video presentation for your app, displaying the use for all in the class.

Semester Hours

Over the course of the semester students will spend at least the amounts of time shown below:

Task	Hours Spent on Task
Class Time	40
Online Content & Reading	90
Weekly Assignments	15
R Package	15
Shiny App	30
Total	190

Class Schedule

Important: Flipped material and readings are subject to change, contingent on mitigating circumstances and the progress we make as a class.

Week 00, 09/03 - 09/07: NO CLASS: First Class will be September 10th

- **NO CLASS**
- You are expected to complete work before the first class.
- Familiarize yourself with:
    - Course Website
    - Github
    - RStudio

Week 01, 09/10 - 09/14: Syllabus / Getting Started / Github

- *First Day of Class*
- Go over syllabus
- Learn about flipped course.
- Learn Use of Server.
- Learn Basics of R and RMarkdown
- Basic Data Retrieval and tracking Code

Week 02, 09/17 - 09/21: R Basics and Data Types / Working with Dplyr and Tidyr

- Vectors, Matrices, Arrays, Lists and Dataframes.
- *Required Reading*:
    - Matloff Chap 1-5
    - Jones Chap 1-2
    - Teetor Chap 7
    - Rizzo Chap 1
- Cleaning Data with Dplyr and Tidyr.
- Using Dplyr on MySql databases.
- *Required Readings*:

  - tidyr and dplyr from Stats Education
  - Chapters 9-16 in R for Data Science (short chapters)

Week 03, 09/24 - 09/28: Controlling the Flow

- Basics of Logic
- Loops and other Controls.
- *Required Reading*:
  - Matloff Chap 7
  - Jones Chap 5

Week 04, 10/01 - 10/05: Functions in R

- Writing and Debugging Functions in R
- *Required Reading*:
  - Matloff Chap 12-14
  - Rizzo Chap 4
  - Jones Chap 7-9
  - Teetor Chap 13

Week 04, 10/01 - 10/05: Functions in R: Part 2

- Writing and Debugging Functions in R
- *Required Reading*:
  - Matloff Chap 12-14
  - Rizzo Chap 4
  - Jones Chap 7-9
  - Teetor Chap 13

Week 06, 10/15 - 10/19: Simulation

- Basics of Simulation
- Simulating Distributions and MCMC.
- *Required Reading*:
  - Matloff Chap 8
  - Rizzo Chap 3
  - Jones Chap 20

Week 07, 10/22 - 10/26: SQL

- Basic SQL Commands.
- Accessing MySQL on a server.
- *Required Readings*:
  - TBD

Week 08, 10/29 - 11/02: Text Mining in R

- tidytext
- Text mining in R.
- Sentiment Analysis
- *Required Readings*:
  - Text Mining with R

Week 09, 11/05 - 11/09: Graphics in R

- `ggplot2` in R
- *Required Readings*:
  - Graphics from Stats Education

Week 10, 11/12 - 11/16: Webscraping in R

- Webscraping in R
- *Required Readings*:
  - TBD

Week 11, 11/19 - 11/23: Thanksgiving Break

No Classes - Thanksgiving Break

Week 12, 11/26 - 11/30: Package Building in R

- Creating packages in R
- *Required Readings*:
  - R Packages

Week 13, 12/03 - 12/07: Shiny Apps

- Developing a Shiny App in R
- *Required Readings*:
  - TBD

Week 14, 12/10 - 12/14: Shiny Apps Continued

- Creating packages in R
- *Required Readings*:
  - TBD

Brown Undergraduate Competencies

Primary Competencies

1. Be familiar with high speed and high throughput computing.
2. Apply Computationally intensive statistical methods (e.g., iterative methods, optimization, resampling, and simulation/Monte Carlo methods)

Refresher Competencies

1. Be able to manipulate data (possibly "big") using software in a well-documented and reproducible way.
2. Apply basic programming concepts (e.g., breaking a problem into modular pieces, algorithmic thinking, structured programming, debugging, and efficiency)
3. Be able to use of one or more professional statistical software environment

Brown ScM/AM Competencies

Primary Competencies

1. Identify and implement statistical techniques and models for analysis of data.
2. Attain proficiency in management, documentation of study data for use in practical statistical analysis.

Refresher Competencies

1. Acquire knowledge and skills in research methodologies to collaborate with substantive investigators

2. Apply programming skills to analyze data and develop simulation studies
3. Develop proficiency in making oral, written and poster presentations of work to statistical and non-statistical colleagues

Brown PhD Competencies

Refreshed Competencies

1. Identify and implement advanced statistical models for the purposes of estimation, comparison, prediction, and adjustment in non-standard settings.
2. Apply programming skills to analyze data and develop simulation studies.
3. Develop proficiency in making oral, written and poster presentations of work to statistical and non-statistical colleagues
4. Generate original computer code for new statistical techniques
5. Determine the statistical properties of new methods using mathematical and computer tools

Students with Special Needs

Brown University is committed to full inclusion of all students. Students who, by nature of a documented disability, require academic accommodations should contact the professor during office hours. Students may also speak with Student and Employee Accessibility Services at 401-863-9588 to discuss the process for requesting accommodations.

Diversity Statement

This course is designed to support an inclusive learning environment where diverse perspectives are recognized, respected and seen as a source of strength. It is our intent to provide materials and activities that are respectful of various levels of diversity: mathematical background, previous computing skills, gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture.

English Language Learners

Brown University welcomes students from around the world, and the unique perspectives international students bring enrich the campus community. To empower students whose first language is not English, an array of ELL support is available on campus including language and culture workshops and individual appointments. For more information about English Language Learning at Brown, contact the ELL Specialists at ellwriting@brown.edu.