

Statistics 133: Concepts in Computing with Data

Fall 2015, Lec 001 and 002

University of California, Berkeley

Instructor: Gaston Sanchez

Email: gaston.stat@gmail.com

Office: 341 Evans Hall

Office Hours: by appointment

Lec 001: MWF 11-12:00 pm, 2060 VLSB

Lec 002: TuTh 9:30-11 am, 10 Evans

About This Course

This course focuses on how to use the computer to conduct a statistical analysis of data. The goal of this course is to introduce you to a variety of programs and technologies that are useful for acquiring, cleaning, manipulating, organizing, exploring, visualizing and analyzing data, as well as report findings. We'll begin with an overview of the R language, which you will use extensively throughout this course. Then we'll move on to other useful tools, including working with regular expressions, basic UNIX tools, XML, and SQL. Because this course is so computer-oriented, it's very important that you take the time outside of class to learn by doing – to explore the software we'll be talking about in class, and to try out your skills on real datasets in the homework assignments and individual project.

Tentative Schedule and Topics

Week	Content
1	Introduction, Getting Started with R
2	R Fundamentals
3	Data objects: vectors, matrices, lists, data.frame
4	Data import, manipulating data tables
5	Data Visualization
6	Graphics with ggplot2
7	Functions
8	Control flow
9	Exploratory Data Analysis
10	Working with text, Regular Expressions
11	Unix utilities
12	Data Technologies and markup languages
13	Databases and SQL
14	Reporting and Communication

Resources

There is no required textbook for this course. Instead, there are various resources that you can check (see list below). I will also point out other appropriate resources during the course.

Please make sure you can access the online and free resources above remotely.

Stat 133 Class Notes by Phil Spector: <http://www.stat.berkeley.edu/~s133/resources.html>

Data Manipulation with R: <http://www.springer.com/us/book/9780387747309>

Grading

10% participation

15% lab

25% homework

25% midterm (Oct 22/23) and final exam (Dec 14/15)

25% projects

Attendance and behavior in class

You are expected to attend all lectures. Any known or potential extracurricular conflicts should be discussed in person with the instructor during the first week of classes, or as soon as they arise. Cell phones are to be turned off during class time, unless otherwise specified by the instructor. Any cell phone activity will cause you to lose participation points.

Submission of assignments

Assignments, both homework and individual projects, will be accepted by electronic submission to the designated location. No late submissions will be accepted. Changes to this policy are possible only by prior arrangement with the instructor in person.

Academic Honest Policy

Homework and projects must be done independently. You may discuss the HW with other students, but you must independently write your code and solutions. For example, suggesting a function to another student is acceptable, whereas simply giving him or her your own code is not. If you are not clear about the expectations for completing an assignment or taking an exam, be sure to seek clarification from the instructor or GSI beforehand.

For exams, cheating includes, but is not limited to, bringing written or electronic materials into an exam or quiz, using written or electronic materials during an exam or quiz, copying off another person's exam or quiz, allowing someone to copy off of your exam or quiz, and having someone take an exam or quiz for you. Any evidence of cheating will result in a minimum penalty of a score of zero (0) on the assignment or examination. Depending on the severity of the infraction, cheating may result in an F for the course grade.

Disability

If you need accommodations for any physical, psychological, or learning disability, please speak to me after class or during office hours. Please make arrangements in a timely manner (through DSP) so that I can make the appropriate accommodations.