Statistics 133: Concepts in Computing with Data
University of California, Berkeley, Summer 2015

Instructor: Gaston Sanchez
Email: gaston.stat *at* gmail.com
Office: 387 Evans Hall
Office Hours: Mon 10:00-11:00am, or by appointment

GSI: Donovan Lieu
Office Hours: to be announced

Lectures: Mon-Fri 9-10:00 am, 342 Computer Lab
Lab: Tu-Th 10:00-11:00am, 342 Computer Lab

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	Jun 8 - Jun 12	Introduction, Getting Started with R
2	Jun 15 - Jun 19	R Fundamentals (data structures, control flow)
3	Jun 22 - Jun 26	Graphics and Data Visualization
4	Jun 29 - Jul 2	Data import, manipulating data tables
5	Jul 6 - Jul 10	Exploratory Data Analysis
6	Jul 13 - Jul 17	Working with text, Regular Expressions
7	Jul 20 - Jul 24	Unix utilities
8	Jul 27 - Jul 31	Data Technologies and markup languages
9	Aug 3 - Aug 7	Databases and SQL
10	Aug 10 - Aug 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
Introduction to Data Technologies (Paul Murrell) https://www.stat.auckland.ac.nz/~paul/ItDT/

Grading
10% participation
10% lab
20% homework
30% midterm (Jul-6) and final exams (Aug-14)
30% final project

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. Late homeworks will be accepted only for 24 hours after the due date and time, for 50% of possible points. 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.