Syllabus for Statistics 133: Concepts in Computing with Data, University of California, Berkeley, Spring 2016

MWF 1-2 pm Lecture 1, 2040 Valley LSB
MWF 2-3 pm Lecture 2, 2040 Valley LSB

Instructor: Adam Lucas
Office: 449 Evans Hall
Office Hours: MWF 10:00-11 Evans 449 and by appointment
e-mail: alucas@berkeley.edu

GSIs:
Andrew Do do@berkeley.edu
Hao Lyu haolyu@berkeley.edu
Luyun Zhao lynnzhao@berkeley.edu
Ye Zhi yezhi@berkeley.edu
Dilek Uz, dilekuz@berkeley.edu
Office Hours: TBA

About This Course

This course will introduce you to the skills for data wrangling using the programming language `R`. Data wrangling is the process of preparing data for visualization and other modern techniques of statistical interpretation. We will focus on using the `R` packages dplyr and ggplot 2 developed by Hadley Wickham to accomplish data wrangling tasks in a concise and expressive way. We will also focus on applications of `R` to statistical learning. Be aware that this is a time consuming, upper-division course. Much of your learning will be by doing, taking the time outside of lectures to experiment with what we have discussed, and to do the assignments.

Textbooks and Other Resources

There are two resources required for the class:

1) Data camp: https://www.datacamp.com
2) Daniel Kaplan (2015) Data Computing: An Introduction to Wrangling and Visualization with R, Project Mosaic Books.

Exams

There will be one 50 minute in-class midterm, and one 50 minute final (room to be announced). The test dates are:

M: Wednesday March 16

F: Friday May 13, 11:30am-12:30pm

Grading

Your final grade will be a weighted average of scores in the following areas:

- 44% assignments
- 15% group project - end of semester
- 15% midterm
- 15% final exam
- 10% lab work
- 1% i-clickers

Submission of assignments: Weekly assignments and group projects, will be accepted by electronic submission to b-courses only. Late homework will not be accepted. You are allowed to drop your lowest assignment. Changes to this policy are possible only by prior arrangement with the instructor. *Be sure to allow sufficient time to upload your files.*

Labs

Lab work will consist of doing chapters in DataCamp and miniprojects that supplement the material covered in the class. Lab work is due at 10pm the following Monday in case you don't finish during lab. Late labs will not be accepted. You are allowed to drop your lowest lab. Accounts for the computer lab will be given out during the first week of classes starting Wednesday January 20 but you are encouraged to bring your own laptop to lab. If you can't make this first session, be sure to contact the GSI to get your account. *Get an account even if you plan to use your laptop; you will need it later in the course.*

Other Course Policies

Laptops in class: I encourage you to bring your laptops to class and lab. Using a laptop for unrelated tasks is a distraction to other students (and yourself!) and is prohibited. Cell phones should be put away during class.

Academic integrity: Any test or assignment submitted by you and that bears your name is presumed to be your own original work that has not previously been submitted for credit in another course. While you are encouraged to work together on homework assignments, the code and writeup must be your own. 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. Any evidence of cheating or plagiarism will be subject to disciplinary action, at minimum a failing grade on the assignment in question.

Students with disabilities: If you need accommodations, please get in touch with me so that we can make the necessary arrangements.