STA 141B Data & Web Technologies for Data Analysis

People

- Instructor: Randy Lai (rcslai@ucdavis.edu)

- Meeting time: 9:00 - 10:20 AM, TR

- Location: Wellman Hall 216

- Office hour: Tuesday/Thursday 2:00pm-3:00pm (or by appointment)

- TA: Franco Liang (frnliang@ucdavis.edu)

- Meeting time: 5:10 - 6:00 PM, W or 6:10 - 7:00 PM, W

- Office hour: to be announced

Site

- Canvas for grades
- Piazza for discussion
- GitHub for lecture notes

Tentative Schedule

Week	Topic
1	Introduction
2	Tidy data
3	Visualization
4	Regular Expressions and strings
5	Shiny
6	Databases and SQL
7	XML, JSON and YAML
8	Web Scraping and REST API
9	Text Mining
10	Show class

Grading

Category	Grade Percentage
Assignments	70%
Project	25%
Participation	5%

- There will be around 5/6 assignments
- Assignments must be turned in by the due date. No late assignments are accepted.
- Participation will be based on your involvement in class, discussion, or office hours. The most subjective way to earn participation points is to have some interactions on Piazza. (A+ will be only given to those

students with high participation)

How to ask questions

I and TA will not respond to any emails about general questions about assignments and course materials. Please use piazza in regard to this matter. For private or sensitive questions you can do private posts on Piazza or email the instructor or TA.

Learn how to ask a question. Asking a question is an art, stackoverflow.com has some good tips.

Group Project

Each group will be 3 to 4 people. Details of project will be announced later.

Resources

- J. Bryan, Data wrangling, exploration, and analysis with R (https://stat545.com/)
- J. Bryan, the STAT 545 TAs, J. Hester, Happy Git and GitHub for the useR (https://happygitwithr.com/)
- G. Grolemund and H. Wickhom, R for Data Science (https://r4ds.had.co.nz/)
- H. Wickham, Advanced R (https://adv-r.hadley.nz/)
- R. Peng, S. Kross, and B. Anderson, Mastering Software Development in R (https://bookdown.org/rdpeng/RProgDA/)

Assignment Rubric

(Adapted from Nick Ulle and Clark Fitzgerald)

Point values and weights may differ among assignments. This is to indicate what the most important aspects are, so that you spend your time on those that matter most. Check the homework submission page on Canvas to see what the point values are for each assignment.

The grading criteria are correctness, code quality, and communication. The following describes what an excellent homework solution should look like:

Correctness

The report does the following:

solves all the questions contained in the prompt makes conclusions that are supported by evidence in the data discusses efficiency and limitations of the computation cites any sources used The attached code runs without modification.

Code Quality

The code is idiomatic and efficient. Different steps of the data processing are logically organized into scripts and small, reusable functions. Variable names are descriptive. The style is consistent and easy to read.

Communication

Plots include titles, axis labels, and legends or special annotations where appropriate. Tables include only columns of interest, are clearly explained in the body of the report, and not too large. Numbers are reported in human readable terms, i.e. 31 billion rather than 31415926535. Writing is clear, correct English.

Inquisitiveness

The report points out anomalies or notable aspects of the data discovered over the course of the analysis. It discusses assumptions in the overall approach and examines how credible they are. It mentions ideas for extending or improving the analysis or the computation.