Syllabus for STAT 450: Introduction to R for Data Science
Fall 2020

Instructor: Dr. Eric Fox
Office: SC-N 303A
Email: `eric.fox@csueastbay.edu`

Lecture: Tu/Th 11:00-12:15 over Zoom

Office Hours: Tu/Th 4:30-5:45 and Wed 1-3, or by appointment.
Zoom link: `https://csueb.zoom.us/j/502694714`

Website: Course materials will be posted on Blackboard.

Textbook: Garrett Grolemund and Hadley Wickham. *R for Data Science*. Free electronic version: `https://r4ds.had.co.nz/`

Additional Reference: Roger Peng. *R Programming for Data Science*. Free electronic version: `https://bookdown.org/rdpeng/rprogdatascience/`

Software:
R, can be downloaded here `https://www.r-project.org/`
RStudio, can be downloaded here `https://www.rstudio.com/`

Course Topics: We will start the course with an introduction to important data structures and some basic commands for data summaries and graphics in base R. Next we will cover the so-called `tidyverse`, which is a collection of modern R packages designed for data science. The two most important `tidyverse` packages we will discuss are `ggplot2` for data visualization, and `dplyr` for data wrangling. Packages in the `tidyverse` are designed to work well together, and share a common philosophy of data and programming. The end of the course will cover more advanced programming concepts (looping and functions), working with text data, and exploratory analysis of date-time data.

- Fundamental data structures in R: vectors, factors, and data frames
- Basic data summaries and graphics in base R: `summary()`, `table()`, `plot()`
- Exploratory data analysis using `ggplot2`

- Data wrangling using `dplyr`: `select()`, `filter()`, `mutate()`, `summarize()`
- Reading data into R, and writing data to files
- Combining data frames: `inner_join()`, `left_join()`
- Control structures: `if-else` statements, `for` loops, `apply()`
- Functions
- Strings and regular expressions
- Dates and times

Grading: There will be weekly homework assignments and two take-home exams. All assignments should be completed using R Markdown and submitted to Blackboard.

- 60% Homework
- 40% Two Exams (20% each)

Policy on Late Assignments and Exams: Late homework will generally not be accepted. However, your lowest scoring homework assignment will be dropped. I may agree to extensions on due dates if you are experiencing an emergency or illness.

Student Learning Outcomes: Upon successful completion of this course, students should be able to:

- Understand and apply fundamental R programming concepts: vectors, data frames, logical operators, `if-else` statements, `for` loops, and functions.
- Perform relevant transformations of data sets: subset rows and columns; create new columns, or variables, with functions of existing variables; merge data sets that share a common variable.
- Create and interpret meaningful visualizations of data.
- Communicate the results of a data analysis clearly and appropriately to others using reproducible research techniques (R Markdown).

Technology Requirements: This course will use the web conferencing software Zoom. To participate you will need a stable internet connection, and a laptop or desktop computer equipped with a webcam, microphone, and speakers. Please refer to the Zoom system requirements here.

Course Policies and Zoom Etiquette:

- All lectures will be delivered live during the scheduled class time, and attendance is highly recommended. Recordings of the sessions will be posted on Blackboard for students that cannot attend or have connectivity issues.
- Make sure that your audio is muted upon entry into the class.
- You may ask questions by using the chat function or by unmuting yourself. Please try to not disrupt the instructor or other students.

Common Syllabus Items: Items such as policies on academic dishonesty, disability, and handling emergency situations can be found under "University Policies" on Blackboard.