

Statistics 450: Introduction to R for Data Science

Fall 2019

Prof. Suess, Department of Statistics and Biostatistics, CSU East Bay

Lecture:

- Section 1: MW 6:30 to 7:45, NSc 104

Instructor: Prof. Eric A. Suess **Office:** NSc 319 **Phone:** 510-885-3879 **e-mail:** eric.suess@csueastbay.edu

Office Hours: MW 2 to 3pm, or by appointment

Class Website: cox.csueastbay.edu/~esuess/stat450

Required Text:

- Wickham, Golemund R for Data Science, O'Reilly, 2017.

Further References:

- Ismay, Kim, ModernDive
- Phillips, Yarr
- Wickham, Advanced R
- Kross, Unix Workbench

Material To Be Covered:

This is the first course in the sequence of Data Science courses offered by the Department of Statistics and Biostatistics for the B.S. Data Science Concentration.

The sequence of courses are:

1. Stat. 450 - Introduction to R for Data Science
2. Stat. 451 - Introduction to Data Visualization
3. Stat. 452 - Introduction to Statistical Learning

These courses are intended to be taken in order as they build upon each other, but you can discuss taking the courses out of order with instructor approval.

The topics of the course will follow the topics presented in the R for Data Science book. The focus of the computing in R is with the use of the Tidyverse.

The main topics for Statistics 450: Introduction to R for Data Science

- Exploring data - tibble: `data_frame()`, `ggplot`:
- Wrangling data - dplyr: `select()`, `mutate()`, `filter()`, `group_by()`, `summarize()`
- Writing programs in R
- Modeling data - `lm()`
- Communicating results - RMarkdown

Homework: A list will be on the website. Homework will be assigned weekly. Homework will be “due” on Mondays, which means you should complete the homework and come to class prepared to ask questions. Homework will be “collected” though Blackboard and needs to be submitted by the end of the week the homework is due.

Quizzes and Exams: Practice quizzes will be given in class (no credit), two short quiz, one midterm will be given and the final.

Grading:

- Project 30%
- Homework 15%
- Quizzes 5%
- Midterm 25%
- Final 25%

Policy on Make-up Exams: You are expected to take the quizzes and exams at the scheduled times. In case of genuine emergency, illness or hardship, for which you can present written documentation I may agree to arrange for a make-up exam. Make-up exams must always be arranged BEFORE the regular exam is given and always take place AFTER the regular exam. Quizzes may not be made up!

Statistics 450 SLOs**Student Learning Outcomes (SLO's):**

Students graduating with an B.S. in Statistics from Cal State East Bay will be able to:

1. Apply basic computational skills in descriptive statistics and data visualization, hypothesis testing, confidence intervals, modeling and error analysis, including the use of large data sets.
2. Analyze data using appropriate software, including cloud-based software, and to interpret results covering descriptive statistics and data visualization, hypothesis testing, confidence intervals, modeling and error analysis, including the use of large data sets.
3. Communicate to others results involving descriptive statistics and data visualization, hypothesis testing, confidence intervals, modeling and error analysis using reproducible research best practices.
4. Acquire data using methods of design of experiments, survey sampling, or observational data, including data scraping and data wrangling from open source data and free data sources.