

PM 566 Introduction to Health Data Science

Units:	4
Term:	Fall 2020
Time:	Wednesdays 1-5pm
Location:	USC HSC, Soto Building 114 & Online
Instructors:	Meredith Franklin, George Vega-Yon, Emil Hvitfeldt, Abigail Horn
Office:	SSB 202A
Office Hours:	By Appointment
Contact Info:	meredith.franklin@usc.edu

Course Description

This course serves as an introduction to data science with focus on the acquisition and analysis of real-life data. Students will learn the toolsets needed to 1) create workable and reproducible datasets by accessing, scraping, sampling and cleaning data; 2) conduct exploratory data analysis and data visualizations; 3) apply statistical tools to learn from data; and 4) build functions and basic apps. Coding languages R and Python will be used.

Learning Objectives

Through this course, students will become familiar with the techniques used in Data Science, applied to health-related datasets. Students will learn:

- Programming in R, and associated tools Markdown, Git
- Data visualization – summarizing data through interpretable summaries
- Data collection – data scraping, wrangling, cleaning, and sampling
- Exploratory data analysis – generating hypotheses and building intuition
- Basic statistical algorithms
- Building software packages and apps

Prerequisite(s): None

Recommended Preparation: Undergraduate course in statistics and programming

Course Notes

Lecture notes presented in class will be posted on GitHub.

Technological Proficiency and Hardware/Software Required

Computation using R (downloaded from <http://cran.r-project.org>), and development tools including Git (<https://github.com/>) and Markdown will be used throughout the semester.

Readings and Supplementary Materials

- 1) **R Programming for Data Science**, 2019. Roger Peng.
<https://bookdown.org/rdpeng/rprogdatascience/>

Supplementary References

- 1) **R for Data Science**, 2017 Garrett Grolemund and Hadley Wickham.
<http://r4ds.had.co.nz/>
- 2) **Exploratory Data Analysis with R**, 2020 Roger Peng
<https://bookdown.org/rdpeng/exdata/>
- 3) **Mastering Software Development in R**, 2017 Roger Peng, Sean Kross, Brooke Anderson
<https://bookdown.org/rdpeng/RProgDA/>

Description and Assessment of Assignments

Assignments: There will be 5-6 assignments given throughout the semester, approximately every week. Students may discuss the problems with one another, however, individual solutions must be submitted and copying will not be tolerated. All assignments must be completed in R Markdown, and submitted through the Github classes portal of the course. Late assignments will be penalized by 20% for each day past the due date.

Exams: There will be one in-class exam (midterm 2hrs). It will be conducted on your laptop using the tools learned up to that point in the semester.

Final Project: The final project will be to develop a reproducible R package, Shiny app, or pipeline for analysis applied to a real-world dataset.

Labs: Lab attendance is mandatory and participation in the lab is required and counts as part of the overall lab grade.

Grading Breakdown

<u>Assignment</u>	<u>% of Grade</u>
Labs	10%
Homework (6)	30%
In-Class Quizzes (3)	10%
Midterm Exam	20%
Final Project	30%
TOTAL	100%

Assignment Submission Policy

Assignments shall be submitted on the Github classroom portal of the course. Late homework assignments will not be accepted without penalty, except when verifiable extenuating circumstances can be demonstrated.

Course Schedule: A Weekly Breakdown

	Topics/Weekly Activities	Due Dates
Week 1 August 19	Introduction to Data Science tools: R, Python, Markdown, Git, command line tools	
Week 2 August 26	Version Control & Reproducible Research	
Week 3 September 2	Exploratory Data Analysis	
Week 4 September 9	Data visualization	HW1 Due
Week 5 September 16	Data cleaning and wrangling	
Week 6 September 23	Regular Expressions, Big Data, Data scraping, using APIs, data security	HW2 Due
Week 7 September 30	Regular Expressions, Big Data, Data scraping, using APIs, text mining	
Week 8 October 7	Midterm Exam	Midterm Exam
Week 9 October 14	High performance computing, cloud computing	
Week 10 October 21	Managing big data, SQL and non-SQL languages, Google BigQuery	HW4 Due
Week 11 October 28	Interactive visualization and effective data communication I	
Week 12 November 4	Final project workshop: review project progress, preliminary presentations	HW5 Due
Week 13 November 11	Interactive visualization and effective data communication II	
Week 14 November 18	Final Project	

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Part B, Section 11, “Behavior Violating University Standards” <https://policy.usc.edu/scampus-part-b/>. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, <http://policy.usc.edu/scientific-misconduct>.

Support Systems:

Student Counseling Services (SCS) - (213) 740-7711 – 24/7 on call

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention. <https://engemannshc.usc.edu/counseling/>

National Suicide Prevention Lifeline - 1-800-273-8255

Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week. <http://www.suicidepreventionlifeline.org>

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 - 24/7 on call

Free and confidential therapy services, workshops, and training for situations related to gender-based harm. <https://engemannshc.usc.edu/rsvp/>

Sexual Assault Resource Center

For more information about how to get help or help a survivor, rights, reporting options, and additional resources, visit the website: <http://sarc.usc.edu/>

Office of Equity and Diversity (OED)/Title IX compliance – (213) 740-5086

Works with faculty, staff, visitors, applicants, and students around issues of protected class. <https://equity.usc.edu/>

Bias Assessment Response and Support

Incidents of bias, hate crimes and microaggressions need to be reported allowing for appropriate investigation and response. <https://studentaffairs.usc.edu/bias-assessment-response-support/>

The Office of Disability Services and Programs

Provides certification for students with disabilities and helps arrange relevant accommodations. <http://dsp.usc.edu>

Student Support and Advocacy – (213) 821-4710

Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic. <https://studentaffairs.usc.edu/ssa/>

Diversity at USC

Information on events, programs and training, the Diversity Task Force (including representatives for each school), chronology, participation, and various resources for students. <https://diversity.usc.edu/>

USC Emergency Information

Provides safety and other updates, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible, <http://emergency.usc.edu>

USC Department of Public Safety – 213-740-4321 (UPC) and 323-442-1000 (HSC) for 24-hour emergency assistance or to report a crime

Provides overall safety to USC community. <http://dps.usc.edu>