EDUC 263: Introduction to Programming and Data Management Using R

Ozan Jaquette

Fall 2019

Instructor: Ozan Jaquette
Pronouns: he/him/his
E-mail: ozanj@ucla.edu
Office Hours: Tues 2:30-4PM; and by appt
Office: Moore Hall 3038

Teaching Assistant: Xin Li
Pronouns: she/her/hers
E-mail: lixin12171217@gmail.com
Office Hours: Mon 4-5PM; Wed 4-5PM;
Location: Moore Hall 3120 (computer lab)

Class Room: Moore Hall 3027
Class Website: ozanj.github.io/rclass/

Class Hours: Fridays 12 - 4 pm
Class Discussion: piazza.com/ucla/fall2019/educ263/home

Course Description

This course has two foundational goals: (1) to develop core skills in "data management," which are important regardless of which programming language you use, and (2) to learn the fundamentals of the R programming language.

Data management consists of acquiring, investigating, cleaning, combining, and manipulating data. Most statistics courses teach you how to analyze data that are ready for analysis. In real research projects, cleaning the data and creating analysis datasets is often more time consuming than conducting analyses. This course teaches the fundamental data management and data manipulation skills necessary for creating analysis datasets.

The course will be taught in R, a free, open-source programming language. R has become the most popular language for statistical analysis, surpassing SPSS, Stata, and SAS. What differentiates R from these other languages is the thousands of open-source "libraries" created by R users. R is one of the most popular languages for "data science," because R libraries have been created for web-scraping, mapping, network analysis, etc. By learning R you can be confident that you know a programming language that can run any modeling technique you might need and has amazing capabilities for data collection and data visualization. By learning fundamentals of R in this course, you will be "one step away" from web-scraping, network analysis, interactive maps, quantitative text analysis, or whatever other data science application you are interested in.

Students will become proficient in data manipulation tasks through weekly "problem sets" that you complete in groups of three. These problem sets will account for 90% of your grade for the course. This course is a hybrid in-person/online course, meaning students will be responsible for watching online modules, attending lecture, and completing weekly problem sets. Each week a short video module will be posted on the class website where students will learn new material before meeting in person for lecture. The first half of lecture will consist of reviewing topics

mentioned in the video modules more deeply and the latter half of class will be more hands on application where students can get into their groups and work through code on their problem sets.

Course Learning Goals

1. Understand fundamental concepts of object oriented programming
    - What are the basic object types and how do they apply to statistical analysis
    - What are object attributes and how do they apply to statistical analysis
2. Become familiar with Base R approach to data manipulation and Tidyverse approach to data manipulation
3. Investigate data patterns
    - Sort datasets in ways that generate insights about data structure
    - Select specific observations and specific variables in order to identify data structure and to examine whether variables are created correctly
    - Create summary statistics of particular variables to diagnose errors in data
4. Create variables
    - Create variables that require calculations across columns
    - Create variables that require processing across rows
5. Combine multiple datasets
    - Join (merge) datasets
    - Append (stack) datasets
6. Manipulate the organizational structure of datasets
    - summarize and collapse observations by group
    - Reshape and "tidy" untidy data
7. Learn guidelines practical strategies for ensuring data quality when cleaning data and creating analysis variables

Prerequisite Requirements

1. Students must have taken at least a one-quarter introductory statistics course.
2. Students should have some very basic experience using statistical programming software (e.g., SPSS, Stata, R, SAS)
3. [General computer skills] Students should be able to download files from the internet, rename these files, save them to a folder of your choosing, and open this folder.
    - During this course we will often be downloading datasets, opening .Rmd files and .R scripts, changing directories to the folder where we stored the data, and then opening the dataset we just downloaded.

Course Website and Discussion Forum

Course website

Course Website can be found https://ozanj.github.io/rclass/.

In particular, we will use the "Class Resources" page (https://ozanj.github.io/rclass/resources/) to post weekly lecture slides (in pdf and .Rmd formats), datasets, weekly problem sets, etc. Additionally, the "Class Resources" page will list required and optional reading for each week.

Course discussion forum

We are using Piazza as our class discussion forum.

The Piazza website for our course is: https://piazza.com/ucla/fall2019/educ263/home

Piazza is a learning forum where folks can customize questions/comments to share with instructors or the entire class. Folks have the option to insert code, images, videos, tables, links, and text to their Q&A posts. Additionally, posts can be configured to be annonymous.

All questions related to course content should be posted on the Piazza discussion forum.

Communication with Instructor and Teaching Assistant

Email me directly if you have a question regarding any personal issue.

Use Piazza discussion forums for all questions related to course content. All students can then benefit from the response. Instructors will aim to respond within 24 hours of your post Monday through Friday and 48 hours on Saturday and Sunday. We may not be able to respond to questions asked after Thursday at 12PM (problem sets due Friday at 12PM)

I encourage students to answer questions your classmates post on Piazza discussion forums. Writing out explanations to student questions will improve your own knowledge and will benefit your classmates.

Course Reading

Course readings will be assigned from:

- Grolemund, G., & Wickham, H. (2018). *R for Data Science*. Retrieved from http://r4ds.had.co.nz/ [FREE!]
- Xie, Y., Allaire, J. j., & Grolemund, G. (2018). *R Markdown: The Definitive Guide*. Retrieved from https://bookdown.org/yihui/rmarkdown/ [FREE!]
- Other articles/resources we post

Required and optional readings for each week will be listed on the "Class Resources" page of the course website https://ozanj.github.io/rclass/resources/

Please do the reading! I have worked hard to keep required reading load light, focusing only on essentials, because weekly problem sets will be time consuming.

The reading schedule works as follows: I lecture on a topic in class, and then you do the reading about that topic and are required to complete a problem set about that topic. However, if you would prefer to the reading about a topic **prior** to me lecturing about that topic, feel free to do so.

Required Software and Hardware

Software [FREE!]

Instructions on downloading software can be found here

Please install the following software on your laptop

- R
- RStudio
- TinyTeX R package *OR* MikTeK/MacTeX

Hardware

- Bring in laptop with above software installed each week

Contact the teaching assistant beforehand if you cannot bring a laptop to class and we will work out a solution if we can.

Assignments & Grading

Your final grade will be based on the following components:

- Weekly problem sets (90 percent of total grade)
- Attendance and participation (10 percent of total grade)

Undergraduate grading scale

- A+, 95 - 100%
- A, 90 - 94%
- B, 80 - 89%
- C, 70 - 79%
- D, 60 - 69%
- F, 50 - 59%

Graduate grading scale

- A, 90 - 100%
- B, 80 - 89%
- C, 70 - 79%
- D, 60 - 69%

Weekly problem sets

Students will complete 10 problem sets (the last one due during finals week). Problem sets are due by 12PM each Friday, right before we start class.

Late submissions will lose 20% (i.e., max grade becomes 80%). Problem sets not submitted by 12PM the following Monday will not receive points because at that point we will post solutions on the course website. The lowest problem-set grade will be dropped from the calculation of your final grade.

You will not lose points for late submission if you cannot submit a problem set due to an unexpected emergency. But please contact the instructor by email as soon as you can so we can work out a plan.

NOTE ON PROBLEM SETS:
We believe that students learn by doing and therefore problem sets make up a majority of the grade for this class. Problem sets for both graduate and undergraduate students will cover the same topic each week. Students enrolled in the undergraduate section of this class will complete problem sets of shorter length to meet the rigor and expectations of an undergraduate course. Similarly, graduate students will be expected to complete longer problem sets each week to meet the rigor and expectations of a graduate course.

In general, each problem set will give your practice using the skills and concepts introduced during the previous lecture. For example, after the lecture on joining (merging) datasets, the problem

set for that week will require that students complete several different tasks involving merging data. Additionally, the weekly problem sets will require you to use data manipulation skills you learned in previous weeks.

With the exception of the first problem set, students will complete problem sets in groups of 3. We will form groups during class in week 2 and you will keep the same group throughout the quarter. However, each student will submit their own assignment. You are encouraged to work together and get help from your group. However, it is important that you understand how to do the problem set on your own, rather than copying the solution developed by group members.

Since you will be working together, it is understandable that answers for many questions will be the same as your group members. However, if I find compelling evidence that a student merely copied solutions from a classmate, I will consider this a violation of academic integrity and that student will receive a zero for the homework assignment.

A general strategy I recommend for completing the problem sets is as follows: (1) after lecture, do the reading associated with that lecture; (2) try doing the problem set on your own; (3) meet with your group to work through the problem set, with a particular focus on areas group members find challenging.

Finally, we strongly recommend using Piazza to ask questions you have about problem sets. Note that you have the option to post a question anonymously if you prefer. Instructors will do our best to reply quickly with helpful hints/explanations and we encourage members of the class to do the same.

Attendance and Participation (10 percent of total grade)

Students are expected to participate in the weekly class meetings by being attentive, supportive, by asking questions, or by answering questions posed by others. Additionally, students can receive strong participation grades by asking questions and answering questions on Piazza.

Students are required to attend the weekly class meetings. Each unexcused absence results in a loss of 20% from your attendance/participation grade. Three or more unexcused absences will result in a failing grade for the course.

An excused absence is a professional opportunity that you discuss with me beforehand or a medical, or family emergency. Excused absences will not result in a loss of attendance points. However, you will be responsible for all material covered in that class and you will be expected to turn in homework assignments on time.

Course Policies

Classroom environment

With respect to the course material, learning programming and the essential skills of data manipulation is hard! This stuff feels overwhelming to me all the time. So it is important that we all create an environment where students feel comfortable asking questions and talking about what they did not understand. Discomfort is part of the learning process. Unburdern yourself from the weight of being an "expert" and just focus on improving, helping your classmates improve, and helping your instructors improve.

With respect to classroom environment, let's work together to create an environment that is relaxed, supportive, and where students feel comfortable voicing any concerns they have. Be mind-

ful that words and body language affect people. Express your thoughts in a way that doesn't make people feel excluded and does not make disparaging generalizations about any group. As an instructor, I am responsible for setting an example through my own conduct.

Online Collaboration/Netiquette

You will communicate with instructors and peers virtually through a variety of tools such as discussion forums, email, and web conferencing. The following guidelines will enable everyone in the course to participate and collaborate in a productive, safe environment.

- Be professional, courteous, and respectful as you would in a physical classroom.
- Online communication lacks the nonverbal cues that provide much of the meaning and nuances in face-to-face conversations. Choose your words carefully, phrase your sentences clearly, and stay on topic.
- It is expected that students may disagree with the research presented or the opinions of their fellow classmates. To disagree is fine but to disparage others' views is unacceptable. All comments should be kept civil and thoughtful.

Academic accomodations

Students needing academic accommodations based on a disability should contact the Center for Accessible Education (CAE) at (310)825-1501 or in person at Murphy Hall A255. When possible, students should contact the CAE within the first two weeks of the term as reasonable notice is needed to coordinate accommodations. For more information visit https://www.cae.ucla.edu/.

Academic Honesty:

- UCLA policy
  - UCLA is a community of scholars. In this community, all members including faculty, staff and students alike are responsible for maintaining standards of academic honesty. As a student and member of the University community, you are here to get an education and are, therefore, expected to demonstrate integrity in your academic endeavors. You are evaluated on your own merits. Cheating, plagiarism, collaborative work, multiple submissions without the permission of the professor, or other kinds of academic dishonesty are considered unacceptable behavior and will result in formal disciplinary proceedings.
- This class
  - Given that 90% of course grade is based on weekly problem sets, the primary academic honesty concern that could come up in this class is copying problem set solutions from somebody else and passing this in as your own work.

Course Schedule

Course schedule is subject to change at the discretion of the instructor. If there are any changes, it will be to go slower and cut subsequent topics.

Lecture 1, 09/27: Introduction; objects in R

- Reading (after class):
  - Wickham and Grolemund 2018 (W&G) 1; W&G 2; W&G 4; W&G 20.1 - 20.3

Lecture 2, 10/04: Investigating objects and data patterns

- Problem set due (before class): Yes
- Reading (after class):
  - Wickham and Grolemund 2018 (W&G) 5.1-5.4
  - Xie, Allaire & Grolemund (XAG) 3.1 LINK HERE
  - Spend 15 minutes studying the "R Markdown Reference Guide" LINK HERE

Lecture 3, 10/11: Investigating data patterns using Base R

- Problem set due (before class): Yes
- Reading (after class):
  - Wickham, H. (2014). *Advanced R*. Retrieved from https://adv-r.hadley.nz/: 4.1-4.3
  - Nicholls, A., Pugh, R., & Gott, A. (2015). *Sams Teach Yourself R in 24 Hours*. (pg. 236-237) LINK

Lecture 4, 10/18: Pipes and variable creation

- Problem set due (before class): Yes
- Reading (after class):
  - W&G 5.5 (creating variables)
  - XAG 3.3 (R Markdown, creating PDF documents) LINK HERE
    - note: sections 3.3.5 through 3.3.8 will feel somewhat cryptic and are not required for this course; so just do the best you can with those

Lecture 5, 10/25: Processing across rows

- Problem set due (before class): Yes
- Reading (after class):
  - W&G 5.6 - 5.7 (grouped summaries and mutates)
  - XAG 4.1 (R Markdown, ioslides presentations) LINK HERE and 4.3 (R Markdown, Beamer presentations) LINK HERE

Lecture 6, 11/01: Augmented vectors, working with survey data

- Problem set due (before class): Yes
- Reading (after class):
  - W&G 15.1 - 15.2 (factors) [this is like 2-3 pages]
  - W&G 20.6 - 20.7 (attributes and augmented vectors)
  - W&G 10 (tibbles)
  - [OPTIONAL] W&G 15.3 - 15.5 (remainder of "factors" chapter)

Lecture 7, 11/08: Exploratory data analysis and guidelines for data quality

- Problem set due (before class): Yes

- Reading (after class):
  - Data cleaning guidelines PDF

Lecture 8, 11/15: Tidy data

- Association for the Study of Higher Education (ASHE) Conference; Xin Li will lead class
- Reading (after class):
  - W&G chapter 12 (tidy data)
  - [OPTIONAL] Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10), 1-23. doi:10.18637/jss.v059.i10

    - This is the journal article that introduced the data concepts covered in W&G chapter 12 and created the packages related to tidying data
    - Link to article here: LINK

Lecture 9, 11/22: Joining multiple datasets

- Problem set due (before class): Yes
- Reading (after class): W&G 13

Thanksgiving, 11/29: No class

Lecture 10, 12/06: Acquiring data

- Problem set due (before class): Yes
- Reading (after class): W&G 11

Finals Week, 12/13: No class

- Problem set due (before class): Yes

Campus Resources

Counseling and Psychological Services (CAPS)
This is a multidisciplinary student mental health center for the UCLA campus. CAPS offers an array of free services including individual counseling. If you suspect you are experiencing mental health problems or just need someone to talk to, you can make an appointment at John Wooden Center West, facing Drake Stadium, second floor, 310-825-0768. http://www.counseling.ucla.edu

Report Discrimination UCLA is committed to maintaining a campus community that provides the stronget possible support for the intellectual and personal growth of all its members- students, faculty, and staff. Acts intended to create a hostile climate are unacceptable. To file an online incident report, visit: https://equity.ucla.edu/report-an-incident/

Sexual harassment
Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. If you have experienced sexual harassment or sexual violence, you can receive confidential support and advocacy at the CARE Advocacy Office for Sexual and Gender-Based Violence. 1st Floor Wooden Center West, via email CARE.advocate@careprogram.ucla.edu or by phone (310) 206-2465. In addition, Counseling and Psychological Services (CAPS) provides confidential counseling to all students and can be reached 24/7 at (310) 825-0768. You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator, 2241 Murphy Hall, via email at mcato@equity.ucla.edu, or via phone at (310) 206-3417. Reports to law enforcement can be made to UCPD at (310) 825-1491. *Faculty and TAs are required under the UC Policy on Sexual Violence and Sexual Harassment to inform the Title IX Coordinator should they become aware that you or any other student has experienced sexual violence or sexual harassment.*

LGTBQ Resource Center This resource center provides a range of education and advocacy services supporting intersectional identity development. It fosters unity; wellness; and an open, safe, inclusive environment for lesbian, gay, bisexual, intersex, transgender, queer, asexual, questioning, and same-gender-loving students, their families, and the entire campus community. Find it in the Student Activities Center, or via email lgbt@lgbt.ucla.edu. Visit their website for more information: https://www.lgbt.ucla.edu/

International Students
The Dashew Center provides a range of programs to promote cross-cultural learning, language improvement, and cultural adjustment. Their programs include trips in the LA area, performances, and on-campus events and workshops. Visit their website for more information: https://www.internationalcenter.ucla.edu/

Undocumented Student Program
This program provides a safe space for undergraduate and graduate undocument students. USP supports the UndocuBruin community through personalized services and resources, programs, and workshops. Visit their website for more information: https://www.usp.ucla.edu/

Student legal services
UCLA student legal services provides a range of legal support to all registered and enrolled UCLA students. Some of their services include:

- Landlord/Tenant Relations
- Accident and Injury Problems
- Domestic Violence and Harassment
- Divorces and Other Family Law Matter

For more information visit their website: http://www.studentlegal.ucla.edu/index.php

Students with dependents
UCLA Students with Dependents provides support to UCLA studens who are parents, guardians, and caregivers. Some of their services include:

- Information, referrals, and support to navigate UCLA (childcare, family housing, financial aid)
- Access to information about resources within the larger community
- On-site application and verification for CalFresh (food stamps) & MediCal and assistance with Cal Works/GAIN
- A quiet study space
- Family friendly graduation celebration in June

For more information visit their website: https://www.deanofstudents.ucla.edu/Portals/16/Documents/studentsdependents.pdf

Lactation Rooms
Map to lactation rooms on campus

Gender Inclusive restrooms
Map to gender inclusive restrooms

Campus accessibility
Campus accessibility map