

EDUC 260A: Introduction to Programming and Data Management

Course information

Resource	Link
Class website (public)	https://anyone-can-cook.github.io/rclass1/
Questions & Discussion (private)	https://github.com/anyone-can-cook/rclass1_student_issues
Announcements (private)	https://github.com/orgs/anyone-can-cook/teams/rclass1_announcements
Class Zoom link	TBD

Course description

The primary goals of this course are (1) to teach fundamental skills of “data management,” which are important regardless of which programming language you use, and (2) to develop a strong foundation in the R programming language. The course is designed for students who never thought they would become programmers and no prior experience with R is required. For goal (1), most statistics courses teach you how to analyze data that are ready for analysis. In real research projects, data management – the process of cleaning, manipulating, and integrating datasets in order to create analysis datasets – is often more challenging than conducting analyses. For goal (2), R is a free, open-source, object-oriented programming language. R is the most popular language for statistical analysis and one of the most popular languages for “data science.” Students will become proficient in data management and R programming through weekly problem sets, which will be completed in groups. Course format consists of weekly asynchronous lectures and weekly synchronous workshop-style class sessions on Zoom.

Extended description

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 using 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.

The data management and programming skills you learn in this course will transfer to other object-oriented programming languages (e.g., Python).

The course primarily use data and examples from education research. However, the course is designed to teach skills that are important for social science research more broadly and also for computational research within the humanities. **We welcome students from across the university.**

Prerequisite Requirements

- Students must have taken at least a one-semester introductory statistics course
- Students should have some very basic experience using statistical programming software (e.g., SPSS, Stata, R, SAS)
- [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

Instructor and teaching assistants

Instructor

Ozan Jaquette

- Pronouns: he/him/his
- Office: Moore Hall, Room 3038
- Email: ozanj@ucla.edu
- Office hours:
 - Zoom office hours: TBD
 - And by appointment

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 and practical strategies for ensuring data quality when cleaning data and creating analysis variables
8. Become proficient at using GitHub – the industry standard platform used by programmers to collaborate on projects – to ask questions about course material and to collaborate with your classmates

Another broad goal of the course is for students to begin developing practical proficiency in “computational thinking.” The [California Computer Science Standards](#) define computational thinking as “the human ability

to formulate problems so that their solutions can be represented as computational steps or algorithms to be executed by a computer.” This course will encourage students to work on the following elements of computational thinking:

- Before you start writing code to accomplish some task, write out the individual steps that must be completed to accomplish the task
- When a particular piece of code is not working, develop a problem-solving approach where you change one element of the code at a time in order to systematically isolate and fix the problem
- For when you conceptually understand what you need to do but you don’t know the code to accomplish the task, develop a set of “go to” practices to help you figure it out, for example:
 - Ask Google
 - Post a question on the course GitHub “issues” page
 - Become proficient at searching the course lecture slides and course textbook for answers
 - When you know the right function, but not how to use it, become proficient at reading function documentation

Course structure

We will have asynchronous (pre-class) lectures and a synchronous workshop-style class that meets on TBD. Weekly homework will consist of students working through the lectures on their own, a modest amount of required reading, and weekly problem sets completed in groups of three.

1. **Asynchronous (pre-class) lectures.** Weekly asynchronous lectures will be posted on the course website with the expectation that students work through the lecture in advance of our weekly synchronous class meeting. Lecture materials will consist of three types of resources: first, detailed lecture slides (PDF or HTML) with sample code; second, short videos (e.g., 15 to 30 total minutes per week) that provide a high-level discussion of important and/or challenging concepts from the lecture slides, but not a line-by-line recitation of the lecture; and, third, the .Rmd file that created the PDF/HTML lecture slides. This .Rmd file will contain all “code chunks” and links to all data utilized in the lecture. Thus, students will be able to “learn by doing” in that they can run R code on their own computer while they work through lecture materials on their own.
2. **Synchronous workshop-style class meetings.** We will have one synchronous class meeting per week. Typically, these meetings will begin with a discussion of concepts students found difficult or confusing from the lecture materials. The bulk of class time will be devoted to students working in groups on a substantive activity posed by the instructor. While students work in groups, the instructor and teaching assistants will visit each group to answer questions and talk through ideas. We will not be recording these synchronous sessions to respect student privacy.

How to succeed in this class

In just a few words, the keys to success in this class are: **start early, ask for help, help others**

Here are some substantive tips to help you succeed:

- Work through weekly asynchronous lecture materials as soon as you can
 - The weekly asynchronous lecture materials (lecture PDF/HTML, lecture .Rmd file with code, short video lecture) are the core of this course. Lecture materials are designed for you to run the code on your computer as you work through the lecture. Therefore, treat each lecture as an active learning experience rather than passively reading slides.
- Start the weekly problem set early so that have time to seek help on questions you are struggling with
- If you can’t figure something out, ask for help!
 - Discuss with your problem set group
 - Ask a question on GitHub
 - Come to office hours

- Be supportive of your classmates and together we will create a classroom environment where we all help each other succeed!

Classroom environment

We all have a responsibility to ensure that every member of the class feels valued and safe. Be mindful that our words and body language affects others in ways we might not fully understand. We have a responsibility to express our ideas in a way that doesn't make disparaging generalizations and doesn't make people feel excluded. As an instructor, I am responsible for setting an example through my own conduct.

Learning data management, while trying to get a handle on R and unfamiliar data, can feel overwhelming! We must create an environment where students feel comfortable asking questions and talking about what they did not understand. Discomfort is part of the learning process. Unburden yourself from the weight of being an "expert." Focus your energy on improving and helping your classmates improve.

Towards an anti-racist, anti-heteronormative learning experience

This course teaches data management and R programming, tools that are often perceived as objective, independent of context and content. We must acknowledge that racism, white supremacy, and heteronormative ideas of gender identity and sexual orientation are rooted in every aspect of data. These seemingly objective rules (e.g., "the right way to handle data") affect the way data are gathered, how variables are created, the questions asked (or not asked), etc.

In this course we will utilize data that reflect systemic gaps based on race, ethnicity, immigration status, and gender identity, among other aspects of identity. It is critical to acknowledge that the processes used to create these data (e.g., how data collected, the categories chosen to represent identity) are often based on notions of white supremacy and heteronormativity. When you encounter a data management strategy that may cause harm, we encourage you to raise concerns. It may be that your instructor/TAs may need to think more critically about strategies they have been using for a long time! It is also critical that we acknowledge that the social and economic marginalization reflected in data is rooted in systemic oppression that upholds white supremacy and heteronormativity. We should all be reflecting about our own role in upholding these systems.

Course website and communication

Course Zoom link

Time: TBD

Link: TBD

Course website

All course related material can be found on the [course website](#). Pre-recorded lecture videos, lecture slides (PDF/HTML), and .Rmd files will be posted on the class website under the associated sections. Additional resources (e.g., syllabus) may also be posted on the class website.

Course discussion

We will be using GitHub teams for class announcements [HERE](#).

- **GitHub teams:** The teaching team will post all class announcements using GitHub teams. The GitHub team discussions feature allows for quick and seamless communication to all members of an organization or team – in this case, to all students with a GitHub account enrolled in the course. Some features include:
 1. The class team can be viewed and [@mentioned](#) by all students enrolled in the class and part of the organization.
 2. Posts can include code snippets, links, images, and references to issues which make them ideal for this class discussion and participation.

We will be using GitHub issues for questions and class discussion [HERE](#).

- **GitHub issues:** GitHub issues are traditionally used by collaborators of a repository for managing tasks for a project. Our rationale for using issues is twofold: 1) help track and organize questions related to course material and problem sets and 2) promote classroom participation. Students are encouraged to contribute to issues by posting questions, sharing helpful resources, and/or taking a stab at answering questions posted on issues. Some features include:
 1. Adding labels
 2. Assigning or mentioning users to an issue
 3. Referencing other issues

Communication with instructor and TA

If you have a personal question or issue, you can email the instructor or TA directly. Additionally, we are available for office hours or by appointment if there is anything you would like to discuss with us in private.

Course materials

Course readings will be assigned from:

- [R for Data Science](#) by Garrett Grolemund and Hadley Wickham [**FREE!**]
- [R Markdown: The Definitive Guide](#) by Yihui Xie, J. J. Allaire, and Garrett Grolemund [**FREE!**]
- Other articles/resources we post

Required software we will be using:

- [R](#), statistical programming language [**FREE!**]
- [RStudio](#), integrated development environment for R [**FREE!**]
- Link to tips for software installation [HERE](#).

Assignments and grading

Course grade will be based on the following components:

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

Problem sets (90 percent of total grade)

Students will complete 10 problem sets (the last one due during finals week). Problem sets are due by 10am 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.

In general, each problem set will give you 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) communicate with your group to work through the problem set, with a particular focus on areas group members find challenging.

Link to problem set expectations and helpful resources [HERE](#).

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 on Zoom. Additionally, students can receive strong participation grades by asking questions and answering questions on GitHub issues.

Students are required to attend the weekly class meetings (unless you have talked to the instructors about this beforehand). 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.

Grading scale

Letter Grade	Percentage
A+	99-100%
A	93<99%
A-	90<93%
B+	87<90%
B	83<87%
B-	80<83%
C+	77<80%
C	73<77%
C-	70<73%
D	60<70%
F	0<60%

Course topics and schedule

Below is an overview of course topics and schedule. Topics and schedule are subject to change at the discretion of the instructor. Topics may be cut if we need to devote more time to learning the most central topics. It is unlikely that additional topics will be added. The official course schedule, including weekly required reading and optional reading, will be posted on the [course website](#).

1. Introduction to R
 - Introduction to R and R data structures
 - Execute R commands, understand R objects and data structures, use R functions
2. Investigating data patterns in Base R
 - Data investigation and manipulation using Base R
 - Investigate R object type and structure, isolate elements using Base R subset operators and the `subset()` function, create new variables in Base R
3. Enter the tidyverse: pipes and dplyr
 - Data investigation and manipulation using tidyverse
 - Select, filter, and sort data using `tidyverse` functions, chain functions together using pipes (`%>%`), create new variables conditionally using `if_else()`, `recode()`, and `case_when()`
4. Enter the tidyverse: processing across rows
 - Calculate aggregate statistics from multiple rows of data
 - Group rows of data using `group_by()`, create aggregate statistics using `summarize()`
5. Strings and dates
 - Work with strings and date/datetime objects
 - Understand string basics, manipulate strings using `stringr` functions, work with dates and times using the `lubridate` package
6. Attributes and class
 - Understand the class and attributes of R objects
 - Investigate R object class and attributes, work with factor variables, label variables and values of a dataframe using the `labelled` package
7. Data quality
 - Tools and guidelines for exploratory data analysis (EDA)
 - Learn common approaches for exploring and analyzing data, understand skip patterns in survey data
8. Tidy data
 - Understand tidy data structure and reshaping data
 - Define tidy data and how to reshape untidy data into tidy form, reshape data from wide to long using `pivot_longer()`, reshape data from long to wide using `pivot_wider()`, handle missing values during reshaping

9. Joining data

- Combine data from multiple datasets using joins
- Merge datasets using mutating joins, check quality of merge using filtering joins, append datasets by stacking rows

Course policies

Learning during a global pandemic

With the ongoing spread of the COVID-19 pandemic, we understand that right now is a challenging time for everybody. Many of us may be experiencing added stress or responsibilities that make learning and completing classwork difficult. If you are having trouble keeping up with the class, please reach out to the teaching team and we will help work out a plan with you. We understand that right now is a precarious time and in the event that you or someone in your family and/or shared living space gets sick, we ask that you please reach out to us as soon as you are able to. We want to be accommodating to everyone's unique situation and hope to make this class an enjoyable learning experience for all.

Online collaboration/netiquette

You will communicate with instructors and peers virtually through a variety of tools such as GitHub, email, and Zoom 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.
- It is imperative that we respect one another in this course, and all other spaces. One way to gain/show respect is to actively listen to one another. Please do not text, tweet, email, Facebook, LinkedIn, browse the internet, and such during class.
- In the unlikely event that Zoom is down, please be sure to check your email often for instructions on how we will complete that class session in an asynchronous manner.

Class Zoom guidelines

All synchronous class sessions will be held online, via Zoom. Below, we have outlined some general guidelines about Zoom learning. As we continue learning together, we can add to and change the below list. I'm open to your feedback and your experiences as we continue to learn how to learn via Zoom.

- **Video:** We will not require students to turn on their video during synchronous lectures. We encourage students to turn on their video only if they feel comfortable doing so – particularly during small group breakout rooms.
- **Audio:** We ask students to mute their microphones when they are not speaking. We encourage the use of earphones or headphones if you are in a space with background noise.
- **Zoom outage:** In the unlikely event that Zoom is down, the instructors will email the class with instruction for completing the class section in an asynchronous manner. Therefore, if Zoom is not functioning properly during the class period, be sure to check your email often.

- **Internet connectivity:** We understand that having access to a stable internet connection and/or electronic equipment is a privilege. With that in mind, we want to provide a space where everyone has the resources they need to do well in the class. If you have any issues with your internet connection and/or don't have access to electronic equipment, please reach out to the instructors.

Academic accomodations

Center for Accessible Education

Students needing academic accommodations based on a disability should contact the Center for Accessible Education (CAE). 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/>.

Located in A255 Murphy Hall: (310) 825-1501, TDD (310) 206-6083; <http://www.cae.ucla.edu/>

- Due to COVID-19, the CAE office is closed for in-person meetings
- CAE counselor, resources, and services are still available via email / virtual appointment
- Stay up-to-date with CAE newsletters & announcements at <https://www.cae.ucla.edu/announcements-events/student>

Academic integrity

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 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.

Campus resources

Counseling and Psychological Services (CAPS)

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, depression, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. UC offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from any of the aforementioned conditions, consider utilizing the confidential mental health services available on campus.

Students in distress may speak directly with a counselor 24/7 at (310) 825-0768, or may call 911; located in Wooden Center West; <https://www.caps.ucla.edu>

- CAPS is open and has transitioned to Telehealth services ONLY

- Open Mon – Thurs: 8am-6pm and Fri: 8am-5pm
- As always, 24/7 crisis support is always available by phone at (310) 825-0768

Discrimination

UCLA is committed to maintaining a campus community that provides the strongest 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/>

LGBTQ resource center

The LGBTQ 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/> and virtual upcoming events

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.

- Due to COVID-19, the Dashew Center has transitioned its operations to a remote setting
- Visit their website for more information: <https://www.internationalcenter.ucla.edu/>
- For COVID updates, visit <https://www.internationalcenter.ucla.edu/covid-19-updates>

UCLA Undocumented Student Program

This program provides a safe space for undergraduate and graduate undocumented 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/>
- You can reach USP at usp@saonet.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 (Including challenges during COVID)
- Accident and Injury Problems
- Domestic Violence and Harassment
- Divorces and Other Family Law Matter

Due to COVID, Student legal Services is closed to walk-ins.

- All services are by appointment only

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

Students with Dependents

UCLA Students with Dependents provides support to UCLA students 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.swd.ucla.edu/>

Campus maps

Lactation Rooms

- [Map to lactation rooms on campus](#)

Gender Inclusive restrooms

- [Map to gender inclusive restrooms](#)

Campus accessibility

- [Campus accessibility map](#)

Title IX Resources

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, there are a variety of resources to assist you.

- **CONFIDENTIAL RESOURCES:** You can receive confidential support and advocacy at the CARE Advocacy Office for Sexual and Gender-Based Violence, A233 Murphy Hall, CAREadvocate@careprogram.ucla.edu, (310) 206-2465. Counseling and Psychological Services (CAPS) also provides confidential counseling to all students and can be reached 24/7 at (310) 825-0768.
- **NON-CONFIDENTIAL RESOURCES:** You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator, 2255 Murphy Hall, titleix@conet.ucla.edu, (310) 206-3417. Reports to law enforcement can be made to UCPD at (310) 825-1491. These offices may be required to pursue an official investigation.

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.