

XXX: Course title here

Fall 2024

Table of contents

1	Course information	2
2	Course description	2
2.1	Extended description	2
3	Office hours	3
3.1	Instructional team	3
3.2	Teaching assistants	4
4	Course learning goals	4
4.1	Course structure	5
5	Assignments and grading	6
5.1	Problem sets (90% of total grade)	6
5.2	Participation (10 percent of total grade)	8
5.3	Grading scale	8
6	Course Schedule	9
7	Course policies	10
7.1	Online collaboration/netiquette	10
7.2	Academic accomodations	11
7.3	Academic integrity	11
8	Campus resources	12
8.1	Campus maps	14
8.2	Title IX Resources	14

1 Course information

Resource	Link
Weekly meetings (online)	Day(s), Time, Timezone
Class website (public)	Link to website here
Questions, discussion, announcements (private)	Link to class repository for student issues here
Class Zoom link	Zoom link here

2 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 is used, 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. 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 arguably the most popular language for statistical analysis and one of the most popular languages for graphics and data visualizations (e.g., web-scraping, interactive maps, network analysis) compared to other open-source languages. Students will become proficient in data management and R programming through weekly problem sets, which will be completed in groups.

This class is created in collaboration with [UCLA’s DataX initiative](#) in an effort to make data science accessible to the the UCLA community. The DataX initiative encourages interdisciplinary research, scholarship, and innovation centering on the application of “data science and critical social, ethical, and data justice research and pedagogy.” The course aims to provide a welcoming space for all learners interested in working with data and the implications of how it is used and for what purpose.

2.1 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 particularly in social science research in asking questions that inform our social world (e.g., social relationships, policies impacting communities, administrative data for organizational change, etc.).

The course will use 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 the 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 uses data and examples from social science research and 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.**

Recommended prerequisites (encourage, **but not required**)

- One prior introductory statistics course (e.g., undergraduate-level stats course)
- Proficiency in general computer skills is helpful, (e.g., downloading files from the internet, renaming files, saving them to a folder of your choosing, finding this folder on your computer, etc).

3 Office hours

The instructor’s and teaching assistants’ weekly office hours are listed below. If you would like to schedule a 1:1 appointment with someone from the instructional team, please email them 48 hours in advance.

3.1 Instructional team

Instructor Name

- Pronouns:
- Office:
- Email:

- Office hours:
 - Zoom office hours: DAY, TIME, zoom link
 - And by appointment

3.2 Teaching assistants

Name

- Pronouns:
- Email:
- Office hours:
 - Zoom office hours:
 - And by appointment

Name

- Pronouns:
- Email:
- Office hours:
 - Zoom office hours:
 - And by appointment

4 Course learning goals

You will become proficient in data management and R programming language. What does that look like? You will be able to:

1. Understand fundamental concepts of object-oriented programming
 - Understand basic object types and how they apply to statistical analysis?
 - What are object attributes and how do they apply to statistical analysis?
2. Become familiar with Base R approach 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. Visualize data
 - Create plots using the ggplot2 library
 - Customize plots through color palettes, labels, line shapes, etc.
 6. Combine multiple datasets
 - Join (merge) datasets
 - Append (stack) datasets
 7. Manipulate the organizational structure of datasets
 - Summarize and collapse observations by group
 - Reshape and “tidy” untidy data
 8. Learn guidelines and practical strategies for ensuring data quality when cleaning data and creating analysis variables
 9. In addition, you will become proficient at using GitHub issues– the industry standard platform used by programmers to collaborate on projects– to ask questions about course material and to collaborate with your classmates

4.1 Course structure

Overview. The course structure consists of weekly **asynchronous course materials** and weekly **synchronous meetings**. Each week the class focuses on a particular topic (e.g., creating variables; writing functions). For each weekly topic, you will complete a problem set. Problem sets will be completed in groups and focus on practical applications of concepts/skills from the topic of the week.

Asynchronous course materials. Asynchronous course materials will focus on the topic for that week (e.g., processing across rows). Course materials will consist of three types of resources:

1. Detailed lecture slides (HTML) with sample code
2. Pre-recorded video lecture of the instructor working through these slides
3. The “.qmd” file that created the HTML lecture slides.
 - The .qmd file will contain all “code chunks” and links to all data utilized in the lecture. Thus, you will “learn by doing” in that you will run *R* code on your own computer while you work through lecture materials on your own.

Synchronous meetings. Synchronous class meetings will be on Zoom. Attendance during the entire period is required, but you may ask the instructor/TAs for exceptions due to scheduling conflicts.

During synchronous class time, you will have the option of (A) attending live lectures from the instructor or (B) working through lecture materials/problem sets in Zoom breakout rooms in small groups (e.g., problem set groups) or on their own. For the first three weeks of class, you will not have the option of working in Zoom breakout rooms.

For those who decide to work in Zoom breakout rooms, you will use this time to work through course materials (e.g., lecture slides, video lectures) and/or the associated problem set as you see fit. The synchronous workshops are also a great time to ask questions about course material or practical applications. TAs will be moving from one breakout room to the next, providing help. Each group can develop their own approach to how they want to use the synchronous workshop time. Some groups may work relatively independently, while others may work collaboratively. Some groups may agree to work through all asynchronous lecture materials beforehand so they can devote all workshop time to making progress on the problem set. A recommendation is to work through the lecture material before getting started on the problem set.

5 Assignments and grading

Course grade will be based on the following components:

- Weekly problem sets (90% of total grade)
- Participation (10% of total grade)

5.1 Problem sets (90% of total grade)

You will complete 10 problem sets (the last one due during finals week). Problem sets are due by [TIME] on [DAY], right before the start of class. In general, each problem set will give you practice using the skills and concepts introduced in the course materials for that week. For example, after the lecture on joining (merging) datasets, the problem set for that week will require that you 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. Link to problem set expectations and helpful resources [HERE](#).

Problem set groups

- Except for the first problem set, you will complete problem sets in groups of 3. If you are abroad, you are encouraged to form your own group to set aside time to work on the problem sets together.
- You have the option of not being part of a problem set group.

- The instructional team will assist you in forming groups during the second synchronous class, and you will keep the same group throughout the quarter. However, you will each submit your own assignment. You are encouraged to work together and get help from your group. However, you must 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 to many questions will be the same as those of your group members. However, if there is compelling evidence that a student merely copied solutions from a classmate, this could be considered a violation of academic integrity. That student will receive a zero for the homework assignment.

A general strategy recommended for completing the problem sets is as follows: (1) after the 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.

Grading policies

- For students working in a problem set group, one submission from each problem set group will be chosen at random. The grade on that problem set submission will be the grade for all members of the group.
 - If a member of a problem set group has not submitted the problem set by the time the TAs conduct grading, that submission will be graded separately once it is submitted
 - The lowest problem-set grade will be dropped from the calculation of your final grade.
- Students who are not part of a problem set group will have their problem sets graded individually. A random subset of 4 or 5 problem sets will be graded. For students who work individually, the lowest problem set grade will not be dropped from the calculation of final grade.
- Weekly required participation on GitHub will be part of your problem set grade
- Policy on late assignments
 - Problem sets submitted after 11:59PM on [CLASS DATE] will lose one percentage point (e.g., max grade becomes 99% instead of 100%)
 - Starting at 12AM [THREE DAYS AFTER CLASS SCHEDULED TIME] morning, problem sets will lose an additional percentage point for each week-day it is not submitted
 - * e.g., for a problem set submitted at 10AM on Monday, the max grade becomes 98%
 - * e.g., for a problem set submitted at 10AM on Tuesday, the max grade becomes 97%

- For late submissions due to an unexpected emergency, you will not lose points. Please contact the instructor and/or TAs and we will work it out together.

5.2 Participation (10 percent of total grade)

Broadly, you are expected to participate by being attentive, and supportive of classmates, by asking questions, and by answering questions posed by classmates.

Practically speaking, the vast majority of your participation grade will depend on weekly participation on GitHub. Each week, you are required to post one communication on GitHub. This could be asking a question about the problem set, answering a question posed by a classmate, or a post describing something you learned while working through the week's material/problem set. If you post at least one communication on GitHub each week, you will earn an "A" for participation for the quarter.

In addition, you can work towards a 100% participation grade for the quarter by asking/answering questions during synchronous lectures (e.g., zoom chat) or by consistently being helpful/supportive to your classmates on GitHub.

5.3 Grading scale

Letter Grade	Percentage
A	93<=100%
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%

6 Course Schedule

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

Week 1: Introduction to R

- Introduction to R and R data structures
- Execute R commands, understand R objects and data structures, use R functions
- Introduce atomic vectors, lists, and functions for investigating objects (e.g., `length`, `type`, `str`)

Week 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

Week 3: Enter the Tidyverse Part I: Pipes & Dplyr

- Data investigation and manipulation using tidyverse
- Select, filter, and sort data using tidyverse functions, chain functions together using pipes (`%>%`)

Week 4: Enter the Tidyverse Part II: variable creation

- Create new variables using `mutate()`
- Create new variables conditionally using `if_else()`, `recode()`, and `case_when()`

Week 5: Processing across rows

- Calculate aggregate statistics from multiple rows of data
- Group rows of data using `group_by()`, create aggregate statistics using `summarize()`

Week 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

Week 7: 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

Week 8: Create plots w/ ggplot

- Understand the layered grammar of graphics for visualizing data with ggplot2
- Make plots with the `ggplot` function (e.g., bar plots, scatter plots)
- Customize plots through color palettes, labels, legends, etc.

Week 9: 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

Week 10: Joining data

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

7 Course policies

7.1 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 you 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:** You are not required to turn on your video during synchronous lectures, but we encourage you to do so if you feel comfortable— particularly during small group breakout rooms.
- **Audio:** We are that you mute your microphones when you 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, we will email the class with instructions 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 us.

7.2 Academic accommodations

Center for Accessible Education

If you need academic accommodations please contact the Center for Accessible Education (CAE). When possible, you 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/>

7.3 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.

AI in the classroom

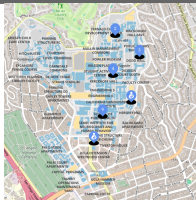
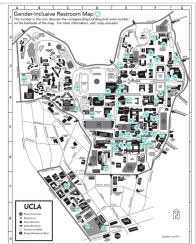
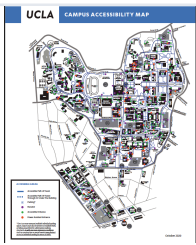
-

8 Campus resources

Resource	Description	Contact
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 that may lead to diminished academic performance or reduce your ability to participate in daily activities.	To speak directly with a counselor 24/7 at (310) 825-0768, if an emergency call 911 Located in the Wooden Center West Website: https://www.caps.ucla.edu
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.	Email: lgbt@lgbt.ucla.edu Located in the Student Activities Center Website: 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.	Website: https://www.internationalcenter.ucla.edu

Resource	Description	Contact
Undocumented Students 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.	Email: usp@saonet.ucla.edu Website: https://www.usp.ucla.edu/
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	Website: https://www.swd.ucla.edu/
Be Well Program	UCLA Be Well Bruin is committed to increasing students' access and awareness of health and well-being resources on campus. They provide a search tool to find resources related to physical, academic, emotional, financial, social, and basic needs.	Website: https://bewellbruin.ucla.edu/
Data Squad	UCLA Library DataSquad is a team of undergraduate students offering consultations in coding, data cleaning and manipulation, data visualization, and statistical analysis in R, Python, Tableau, and SQL.	Website: https://www.library.ucla.edu/about/program/library-datasquad/
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	Website: http://www.studentlegal.ucla.edu/index.p
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.	Website: https://equity.ucla.edu/report-an-incident/

8.1 Campus maps

Map	Resource	Link
	Lactation Rooms	https://chr.ucla.edu/benefits/support-nursing-mothers-at-work
	Gender-Inclusive Restrooms	https://lgbtq.ucla.edu/file/1500b7f1-7c6b-4c1b-9bbe-12674dd23f67
	Campus Accessibility	https://map.ucla.edu/downloads/pdf/Access_Oct_2020.pdf

8.2 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, ti-

tleix@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.