

STAT 331/531 – Statistical Computing with R

Spring 2023

Section 71: MW 10:10-12:00 pm, 38-123

Section 72: MW 12:10-2:00 pm, 38-123

Instructor Information

Dr. Zoe Rehnberg

Pronouns: she, her, hers

Email: zrehnber@calpoly.edu

Office: 25-109 (Faculty Offices East)

Office Hours:

T 10:00-11:00

R 3:30-4:30

F 1:30-2:30

(Or by appointment!)

My office hours are for **YOU**. Come to discuss coursework, related concepts, or anything else I can help with. Outside of scheduled office hours, if my office door is (cracked) open, you are welcome to knock; if my door is closed, I am unavailable.

If you would like to set up an individual meeting, email is the best way to contact me. I will respond as quickly as possible and usually within 48 hours. Do not expect a response on weekends or after 7:00pm.

Course Description

Statistical software is critical for applying statistical techniques, especially when datasets are large. STAT 331/531 will provide you with an introduction to R programming for data and statistical analysis. The course covers basic programming concepts necessary for statistics, good computing practices, and the use of available functions to complete basic statistical analyses.

Prerequisites: Entrance to STAT 331/531 requires successful completion of a STAT II qualifying course **and** an introductory programming course.

Learning Objectives

This course will teach you the foundations of statistical computing principles in the language of R. **After taking this course, you will be able to:**

- Work with the RStudio Integrated development environment (IDE) and Quarto documents.
- Import, manage, and clean data from a wide variety of data sources.
- Visualize and summarize data for informative exploratory data analysis and presentations.
- Write efficient, well-documented, and tidy R code.
- Program random experiments and simulations from probability models.



Additionally, it is my hope that you will learn to:

- Extend your R skills independently through documentation and online resources.
- Be thoughtful, deliberate, and ethical in your use of R and similar tools.
- Use R to be playful, creative, and fun!
- Contribute to and participate in the R Open-Source Community.

In this class, and in every class, having a **growth mindset** will help us achieve our learning goals.

Course Materials

Course Webpage: <https://canvas.calpoly.edu/> (Canvas)

All course material will be made available on our Canvas page, including check-ins, practice activities, lab assignments, challenges, review questions, and selected solutions. Canvas will also contain detailed weekly schedules throughout the term. You are responsible for checking the Canvas page on a regular basis.

Canvas Discussion Page

For questions of general interest, such as course policy clarifications or conceptual questions, please use the **Canvas Discussion page**. I encourage you to give your post a concise and informative initial sentence, so that other people can find it. For example, *“How do I color bars in a barplot with ggplot?”* is a better opening sentence than *“help with plotting”*. **While your posts are not anonymous, there is no such thing as a bad question!**

I will likely be on the Discussion page throughout the workday when I am in my office. Should you have a question, post it and either myself or another student is likely to respond.

Textbook: <https://earobinson95.github.io/stat331-calpoly-text/>

There is an abundance of free online resources for learning programming and R. Therefore, the primary text for this course is a compilation of various resources - it is available online for free. It is a work in progress, so expect changes and updates throughout the course.

This text was constructed by [Dr. Emily Robinson](#), and has been modified from material by [Dr. Susan VanderPlas](#) with integration of content and videos from [Dr. Allison Theobald](#) and [Dr. Kelly Bodwin](#). In addition, you may find it useful to reference some of the following resources; most are available online for free: [R for Data Science \(2nd edition\)](#), [Modern Dive](#), [Advanced R](#).

Equipment

Although you may always work on the Studio computers, I strongly recommend that you use your own personal laptop for this course if you have one. **Note: Chromebooks and iPads will not be sufficient to use R.** If this requirement is limiting for you, please contact me ASAP.

Statistical Software: R and RStudio

We will (obviously) be using the R statistical software throughout this course. In addition, you are **required** to use RStudio, a companion integrated developer environment (IDE).

Both R and R Studio are freely available. Download instructions are available on Canvas.



Assessments

General Evaluation Criteria

In every assignment for this course, you are expected to demonstrate that you are **intellectually engaging with the material**. I will evaluate you based on this engagement, which means that technically correct but low effort answers which do not demonstrate engagement or understanding will receive no credit. When you answer questions in this class, your goal is to show that you understand the material and are actively engaging with it.

While this is not an English class, grammar and spelling are vital parts of your ability to communicate technical information in writing. Both of these will be used, in addition to assignment-specific rubrics, to evaluate your work.

Formative Assessments

1. **Check-ins.** Each week, you will be assigned a chapter to read from the online text. Accompanying that reading will be a short check-in quiz that you will complete on Canvas. These check-in quizzes will help ensure you are prepared for class that week. The Canvas check-in quizzes can be submitted **up to three times** without a penalty. Check-ins are (typically) due **Mondays at 8:00am**.
2. **Practice Activities.** Each week, you will complete a practice activity to get the hang of that week's R skills. These activities will always result in a single, straightforward answer that you will submit on Canvas. Practice activities are due **Wednesdays at 8:00am**; you get one attempt.
 - **Since these activities are intended to be your first attempt at new skills, they are meant to be done with help from me and your peers.** You will always be given time in class to work on them. As such, I strongly suggest you start the activities before class, so you can best take advantage of your in-class time.
3. **Lab Assignments.** Your "homework" will be weekly labs. You will follow each lab's instructions to complete tasks in R and submit a knitted .html Quarto document to Canvas. There will be class time most Wednesdays to work on lab assignments. Labs are due on **Fridays at 11:59pm**.
4. **Challenges.** Most lab assignment will come with a challenge that asks you to try skills beyond what is required that week. Challenges are individual submissions, worth 10 points each. Full credit is given for a good faith attempt. At the end of the quarter, the challenge points are taken out of 100. Scheduled challenges are due on **Saturdays at 11:59pm**.
 - **Note:** there are only 7 challenges on the course schedule! This means that if you only complete the challenges associated with each lab, you will receive 70/100 in this category. To achieve 100/100, you must complete some of the Bonus Challenges provided throughout the quarter. Challenge points beyond 100 earn you extra credit toward your overall course grade.

Evaluative Assessments

1. **Group Project.** There will be a data analysis project to be completed throughout the second half of the term. You will work in a group of 2-4 people. Each group will produce a written project report covering linear regression and model simulation. More information will follow.
2. **Exams.** There will be a midterm exam and a final exam. The midterm will have both in-class and take-home portions. The final will be entirely in-class during our scheduled final exam slot. More information will follow.

Note: The in-class exam is on Wednesday, May 10. The final exam is from 10:10-1:00pm on Friday, June 16 for section 71 and from 10:10-1:00pm on Wednesday, June 14 for section 72. Any known exam conflicts must be addressed with me and resolved at least three weeks before the exam date.

Your grade in STAT 331/531 will be based on the following components:

Check-Ins	5%	Midterm Exam	15%
Practice Activities	10%	Project	15%
Lab Assignments	25%	Final Exam	25%
Challenge Points	5%		

The Cal Poly grading scheme is as follows:

		B+	87	C+	77	D+	67
A	93	B	83	C	73	D	63
A-	90	B-	80	C-	70	D-	60

The final grade cutoffs for this course will not be higher than those listed above, but they *may* be lower (do not count on it).

Course Policies

Inclusive Classroom. Because data are collected by and about humans, data necessarily encode aspects of our proclivities and biases. As a result, this course may touch upon difficult topics related to race, gender, inequality, class, and oppression. We each come into this class with different perspectives that can be shared to enhance our understanding of these issues. I ask that you enter these conversations with respect, curiosity, and cultural humility. You should be open to alternative perspectives and be willing to revise beliefs that are based on misinformation. As a general rule, your ideas and experiences can always be shared during these conversations, but please refrain from dismissing the experiences of others. Personal attacks of any kind will not be tolerated.

Group Work. This course will rely heavily on in-class group work. At the beginning of Week 2, I will assign each student to a group of approximately 4. You will be expected to work with your group on in-class material. If you have concerns about random group assignment, please talk to me at the start of the term.

Group member discussion is expected for the completion of all practice activities. You are also encouraged to discuss your ideas for lab assignments and challenges; **however, I expect that these collaborations are about ideas and no R code is shared.** Each person's lab assignment and challenge submissions are expected to reflect their own thinking, and thus copying the work of others does not provide me with any information about your learning.

The **community standards** for student-to-student and student-to-instructor interactions include the following:

- Listen actively and attentively.
- Ask for clarification if you are confused.
- Challenge one another respectfully.
- Gracefully accept constructive criticism.



- Take responsibility for the quality of the discussion and the work.
- Do not monopolize discussion.
- Acknowledge that everyone has something to offer.
- Speak from your own experience, without generalizing.

Accommodations and Mental Health. I am committed to supporting the learning of all students in my class. **If you are encountering any barriers to your learning that I can mitigate, please let me know.** You may also contact the **Disability Resource Center (DRC)** at <https://drc.calpoly.edu/> to request accommodations.

If you are experiencing mental health challenges that are affecting your learning or well-being, you can contact **Counseling Services** through **Campus Health & Wellbeing** at (805) 756-2511 and <https://chw.calpoly.edu/counseling> during and after hours, on weekends, and on holidays.

Academic Integrity. Our academic community is at its best when we treat each other with fairness, honesty, respect, and trust. Unfortunately, sometimes students slip up and do something that gives themselves (or someone else) an unfair advantage over other students. Such actions will not be tolerated.

In particular, paraphrasing or quoting another's work without citing the source is a form of academic misconduct. This includes the R code produced by someone else! Writing code is like writing a paper, it is obvious if you copied-and-pasted a sentence from someone else into your paper because the way each person writes is different.

Even inadvertent or unintentional misuse or appropriation of another's work (such as relying heavily on source material that is not expressly acknowledged) is considered plagiarism. If you are struggling with writing the R code for an assignment, please reach out to me. I would prefer that I get to help you rather than you spending hours Googling things and get nowhere!

If I suspect you have done something that violates academic integrity on any graded elements of the course, you will receive an email to tell you about my suspicion and how the situation will be handled. Typically, students earn a 0 for the assignment on which academic misconduct is found. University policy dictates that we must report every instance of suspected academic dishonesty to the [Office of Student Rights and Responsibilities](#), no matter how small. **You have the tools you need to be successful in this course without giving yourself (or someone else) an unfair advantage. Thank you in advance for being part of an academic community that is built on honesty, integrity, and respect.**

If you have any questions about using and citing sources, you are expected to ask for clarification.

Exam Conflicts. If you have a known conflict with an exam, please discuss it with me **at least three weeks prior to the exam date**. If an illness or a family emergency arises, please let me know as soon as possible, and we will work out a solution. This may involve taking the exam at a later date or replacing the missed exam score with the final exam score. If you are unable to take the final exam, you will likely receive an Incomplete in the course.

Late Assignments. For check-ins, practice activities, lab assignments, and challenges, you have a **bank of 5 late days** that you may use as you wish. You may use all 5 late days on one assignment, you may use one late day for five different assignments, or any other combination.

A few restrictions:

- **You must email me before the assigned due date indicating the assignment and the number of late days you will use.**
- An assignment submitted anytime between one minute after the assigned due date and 24 hours after the assigned due date will use one late day.
- You **may not** use late days for the final project report.

Email / Discussion Policy. Email should be used **only** for questions or concerns that *relate to you directly* (e.g., setting up an individual meeting). **Any other questions (e.g., about course material, clarification on course policies, etc.) should be posed to the Discussion board on Canvas.** I will not answer such questions over email. It is incredibly helpful for others in the course to see the questions you have and the responses to those questions. I will try to answer any questions posted to the Discussion boards within 4-5 hours (unless it is posted in the evening or on a weekend). If you can answer another student's question, please respond!

How to Succeed

1. **Come to class and ask questions.** You will be responsible for all material covered in class. If you miss a day, get notes from a classmate so you know what you missed. When you are in class, please stop me to ask questions as they arise. If you are confused, chances are one of your classmates is, too.
2. **Stay on top of the material.** The material in this course builds upon itself, so you will find it is far easier to stay on top of the material than it is to catch up. Staying on top of the material involves not only coming to class and completing your assignments, but also reading the text and trying out the textbook examples on your own. The text contains information (both in written and video form) that may not be directly discussed in lecture. It also offers many opportunities for you to practice the new material on your own time.
3. **Come to office hours and ask questions.** Office hours are the best time to discuss and clarify any confusing class concepts. I am here to help!
4. **Make mistakes.** Programming is the process of making a series of silly mistakes, and then slowly fixing each one (while adding a few more). The only way to know how to fix mistakes (and avoid them in the future) is to make them. You might have to make the same mistake a few dozen times before you can avoid making it again. At some point during the course, you will find that you've spent 30 minutes staring at an error caused by a typo, a space, or a parenthesis in the wrong place -- this is part of programming, it is normal, and you shouldn't feel embarrassed (unless you put no effort into troubleshooting the problem on your own).
5. **Ask questions!**



Hello Statistics 331/531 Students!

Welcome to a new term! I'm so excited that I get to spend the next 10 weeks sharing the beauty and importance of statistical computing with each and every one of you. (Well, almost as excited as I am about snow.) Whether you are new to statistics or to coding (or to neither), I hope that this course will get you excited about the power and possibility that arise when you combine them!



My first experience with statistics and statistical computing was through a research program during my first summer of college. I had never taken a statistics course nor used a statistical programming language before, and I was way behind most of my peers. Intimidated, I reached out to the instructors and mentors, and they happily answered my most basic questions and guided me on my way to learning how to code. At the end of the summer, I was able to present work I had done on factors affecting heart attacks, and more importantly, I was hooked on statistical computing.

As with those of us in that research program, students taking Statistics 331/531 have widely varying backgrounds and academic goals. That is part of why I love statistics -- it applies to practically everything! No matter your background, your academic journey, or your anxieties about math and programming, **you belong in this class**. I'm here to challenge and support you along this journey.

-- Dr. Rehnberg

Tentative Course Schedule

CI = check-in PA = practice activity LA = lab assignment CH = challenge

Wk	Day	Date	Class Topic	Due
1	M	4/3	Intro to the course and R	
	W	4/5	Intro to R	CI1; PA1
	F	4/7	LA1
	Sa	4/8	CH1
2	M	4/10	Tidy Data + Basics of Graphics	CI2
	W	4/12	Tidy Data + Basics of Graphics	PA2
	F	4/14	LA2
	Sa	4/15	CH2
3	M	4/17	Data Cleaning + Manipulation	CI3
	W	4/19	Data Cleaning + Manipulation	PA3
	F	4/21	LA3
	Sa	4/22	CH3
4	M	4/24	Data Joins + Transformations	CI4
	W	4/26	Data Joins + Transformations	PA4
	F	4/28	LA4
	Sa	4.29	CH4
5	M	5/1	Strings + Factors + Dates	CI5
	W	5/3	Strings + Factors + Dates	PA5
	F	5/5	LA5

6	M	5/8	Debugging + Version Control	CI6; PA6
	W	5/10	Midterm Exam	
7	M	5/15	Writing Functions	CI7
	W	5/17	Functional Programming	PA7
	F	5/19	LA7
	Sa	5/20	CH7
8	M	5/22	Functional Programming	CI8
	W	5/24	Regression + Simulation	PA8
	F	5/26	LA8
	Sa	5/27	CH8
9	M	5/29	No class - Memorial Day	
	W	5/31	Regression + Simulation	CI9; PA9
	F	6/2	LA9
	Sa	6/3	CH9
10	M	6/5	Predictive Checks	CI10
	W	6/7	Predictive Checks + Project Work	PA10
	F	6/9	LA10
11	M	6/12	Final project report due at 11:59pm	
	W	6/14	Final Exam for Section 72 (10:10-1:00pm)	
	F	6/16	Final Exam for Section 71 (10:10-1:00pm)	