# Stat 331/531: Statistical Computing with R

## Contact

### Dr. Emily Robinson

- Email: [erobin17@calpoly.edu](mailto:erobin17@calpoly.edu)
- Office: Building 25 Office 103 (by Statistics Department Office)

### Course Discord:

For questions of general interest, such as course clarifications or conceptual questions, please use the course Discord page (you will join this Week 1). 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, in this case there is no such thing as a bad question!

## Course Info

### Class Meeting Times:

Mondays/Wednesdays

- Section 70: 10:10am - 12:00pm
- Section 71: 12:10pm - 2:00pm

**Room:** 38-0123 (Math & Science)

## Office Hours

| Day | Time |
| --- | --- |
| Tuesdays | 10:10am - 11:00am, in-person (25-103) |
| Thursdays | 10:10am - 11:00am, in-person (25-103) |
| Fridays | 1:15pm - 2:15pm, by appt ([https://calendly.com/calpoly-emily-robinson/15min](https://calendly.com/calpoly-emily-robinson/15min)) |

Office hours by appointment are required to be scheduled at least 1-hour prior to the meeting. You can schedule either Zoom or in-person office hour appointments through Calendly.

I will likely be on Discord throughout the workday when I am in my office. Should you have a question, post in one of the discord channels and either myself or another student is likely to respond.

# Course Description

Stat 335/531 provides you with an introduction to programming for data and statistical analysis. The course covers basic programming concepts necessary for statistics, good computing practice, and use of built-in 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.

# Course Resources

**Textbook**

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 for free at https://earobinson95.github.io/stat331-calpoly-text/. It is under construction/a work in progress, so it may be hard to work more than a week ahead in this class using the primary textbook.

This text has been modified from material by Dr. Susan VanderPlas. See UNL Stat 151: Introduction to Statistical Computing and UNL Stat 850: Computing Tools for Statisticians for her course books with integration of content and videos from Dr. Allison Theobold and Dr. Kelly Bodwin.

In addition, you may find it useful to reference some of the following resources that I have consulted while assembling the text. 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.

Chromebooks and iPads will **not** be sufficient to use R. If this requirement is limiting for you, please contact me ASAP.

# Class Schedule & Topic Outline

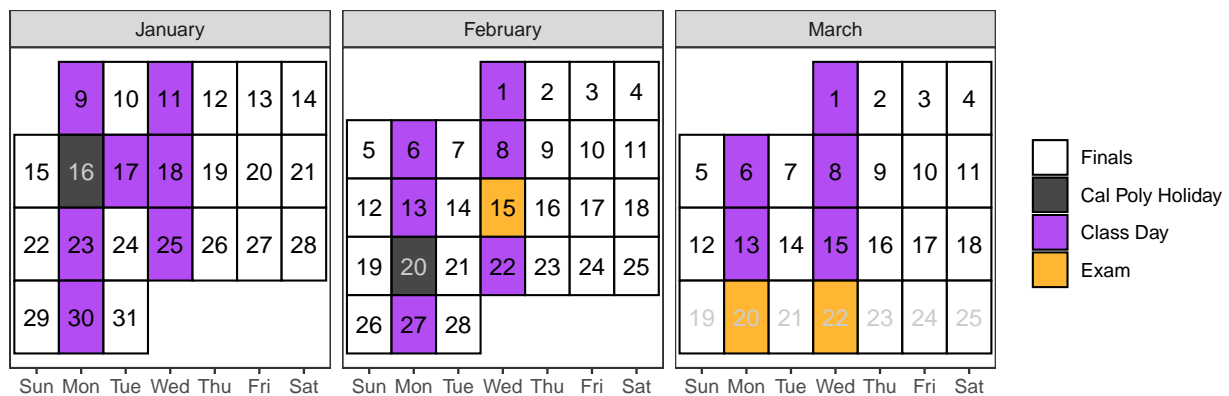This schedule is tentative and subject to change.



Figure 1: Note: Tuesday, January 17th will follow a Monday class schedule.

Table 2: Tentative schedule of class topics and important due dates

| Date | Topic |
|------|-------|
| Jan 9, Jan 11 | Introduction to R |
| Jan 17, Jan 18 | Tidy Data + Basics of Graphics |
| Jan 23, Jan 25 | Data Cleaning and Manipulation (dplyr) |
| Jan 30, Feb 1 | Data Transformations (tidyr) |
| Feb 6, Feb 8 | Special Data Types: Strings + Factors + Dates |
| Feb 13 | Debugging + Version Control |
| Feb 15 | Midterm Exam |
| Feb 22 | Reproducibility & Professional Communication |
| Feb 27, Mar 1 | Functions & Functional Programming |
| Mar 6, Mar 8 | Simulation |
| Mar 13, Mar 15 | Statistical Modeling |
| Mar 20 | Final Exam |
| Mar 22 | Final Exam |

# Course Policies

## Assessment/Grading

Your grade in STAT 331/531 will contain the following components:

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

Lower bounds for grade cutoffs are shown in the following table. I will not "round up" grades at the end of the quarter. See this 🐦 Twitter thread for advice on *"Playing the lines. Don't be there."*

| Letter grade | X + | X | X - |
|---|---|---|---|
| A | . | 93 | 90 |
| B | 87 | 83 | 80 |
| C | 77 | 73 | 70 |
| D | 67 | 63 | 60 |
| F | <60 | | |

Interpretation of this table:

- A grade of 85 will receive a B.
- A grade of 77 will receive a C+.
- A grade of 70 will receive a C-.
- Anything below a 60 will receive an F.

## General Evaluation Criteria

In every assignment, discussion, and written component of this class, 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 either understand the material or are actively engaging with it. If you did not achieve this goal, then your answer is incomplete, regardless of whether or not it is technically correct. This is not to encourage you to add unnecessary complexity to your answer - simple, elegant solutions are always preferable to unwieldly, complex solutions that accomplish the same task.

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

## Assignment Breakdown

### Check-ins

Each week, you will find short Check-In questions or tasks throughout the text to make sure you are prepared for class that week. Make sure you submit your answers to these on Canvas to get credit for your efforts. Note that the Canvas Check-in quizzes can be submitted up to three times without a penalty - so you should get 100% on this part of the course!

- All responses to Check-ins are due **Mondays at 8:00am**.

### Practice Activities

Most Monday's, you will be given a Practice Activity to complete, to get the hang of the week's necessary R skills. These activities will always result in a single, straightforward correct answer, that you will submit via Canvas (one attempt). Therefore, there is no reason you should not get full credit in this category!

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. Therefore, you will always be given some time in class to work on them. I strongly suggest that you attempt to start the activities before class, so you can maximize the utility of your in-class time.

- Practice Activities are due **Wednesdays at 8:00am.**

### Lab Assignments

Your typical homework assignments 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.

Most weeks, there will be class time on Wednesdays dedicated to working on completing lab assignments.

- Labs are due on **Fridays at 11:59pm**.

### Challenges

With each Lab Assignment will come a Challenge, asking you to try skills beyond what is required that week. Challenges are individual submissions, worth 10 points each. Full credit is given for any good faith attempt.

As these are extensions to the lab assignments, they are a great opportunity to discuss your ideas with your classmates. However, I do expect that these collaborations are about ideas and no R code is shared between individuals. Each person's Challenge submission is expected to reflect their own thinking, and thus copying the work of others does not provide me with any information about your learning.

At the end of the quarter, the Challenge points are taken out of 100. However, there are only 8 lab assignments! This means that if you only complete (in good faith attempts) the challenges associated with each lab, you will receive 80/100 (or 80%) in this category. In order to achieve 100/100, you must submit impressive challenge submissions that earn bonus points and/or complete optional Challenge point opportunities provided throughout the quarter.

Extra bonus points beyond 100 earn you extra credit toward your overall course grade.

- Challenges associated with labs are due on **Saturdays at 11:59pm**.
- Watch Canvas for additional/optional Challenge point opportunities and deadlines.

### Attendance & Participation

I do not take formal attendance in this class. However, it is my expectation that you remain in class and on task until you have finished all your activities and assignments. Consistent, repeated failure to attend class or actively participate in portions of the course will affect the demonstration of your engagement with the course.

If you are feeling ill, please **do not come to class**. Instead, email me, review the material and work on the participation activity and weekly lab assignment; then schedule an appointment with me to meet virtually.

### Late Policy

There are no extensions on any assignments (this includes Check-ins, Practice Activities, Labs, and Challenges).

- Canvas will automatically input a 0% for missing assignments as an incentive for you to still turn the assignment in with a penalty described below.
- Canvas will automatically apply a 10% grade deduction for each day past the due date with a minimum grade of 50%.
  - *Take note that Canvas sees 12:01am as a new day. Again, "Don't Play the Lines."*
- Solutions to Practice Activities will be posted. Therefore, no Practice Activities will be accepted after the solution has been posted.

If you find yourself with extenuating circumstances, please email me **prior** to the due date. You will typically have multiple days to complete assignments, make sure to plan ahead for any unforeseen circumstances.

## Course Expectations

You will get out of this course what you put in. The following excerpt was taken from Rob Jenkins' article "Defining the Relationship" which was published in The Chronicle of Higher Education (August 8, 2016). This accurately summarizes what I expect of you in my classroom (and also what you should expect of me).

*"I'd like to be your partner. More than anything, I'd like for us to form a mutually beneficial alliance in this endeavor we call education.*

*I pledge to do my part. I will:*

- *Stay abreast of the latest ideas in my field.*
- *Teach you what I believe you need to know; with all the enthusiasm I possess.*
- *Invite your comments and questions and respond constructively.*
- *Make myself available to you outside of class (within reason).*
- *Evaluate your work carefully and return it promptly with feedback.*
- *Be as fair, respectful, and understanding as I can humanly be.*
- *If you need help beyond the scope of this course, I will do my best to provide it or see that you get it.*

*In return, I expect you to:*

- *Show up for class each day or let me know (preferably in advance) if you have some good reason to be absent.*
- *Do your reading and other assignments outside of class and be prepared for each class meeting.*
- *Focus during class on the work we're doing and not on extraneous matters (like whoever or whatever is on your phone at the moment).*
- *Participate in class discussions.*
- *Be respectful of your fellow students and their points of view.*
- *In short, I expect you to devote as much effort to learning as I devote to teaching.*

*What you get out of this relationship is that you'll be better equipped to succeed in this and other college courses, work-related assignments, and life in general. What I get is a great deal of professional and personal satisfaction. Because I do really like you [all] and want the best for you."*

### Make Mistakes!

Programming is the process of making a series of silly or stupid mistakes, and then slowly fixing each mistake (while adding a few more). The only way to know how to fix these mistakes (and avoid them in the future) is to make them. (Sometimes, you have to make the same mistake a few dozen times before you can avoid it in the future). At some point during the class, you will find that you've spent 30 minutes staring at an error caused by a typo, a space, a parenthesis in the wrong place. You may ask for help debugging this weird error, only to have someone immediately point out the problem… it is always easier to see these things in someone else's code. This is part of programming, it is normal, and you shouldn't feel embarrassed or sorry (unless you put no effort into troubleshooting the problem before you asked for help)

If you manage to produce an error I haven't seen before, then congratulations. You have achieved something special, and that achievement should be celebrated. Each new and bizarre error is an opportunity to learn a bit more about the programming language, the operating system, or the interaction between the two.

## University Policies

See [academicprograms.calpoly.edu/content/academicpolicies](academicprograms.calpoly.edu/content/academicpolicies).

### Learning Environment and Support

I believe everyone is capable of learning statistics and programming with proper support. It is my goal for everyone to feel safe and comfortable in my classroom. If there is any way I can make the course more welcoming for you, please do not hesitate to ask.

In particular, if you have a disability, I will gladly work with you to make this class accessible.

I encourage you to also contact the Disability Resource Center (Building 124, Room 119 or at 805-756-1395), who can help you register for extra accommodations such as extended exam time.

If you are having difficulty affording groceries, lacking a safe & stable place to live, or needing additional essential supports, please see Canvas for a list of Student Support Services at Cal Poly.

## Academic Integrity and Class Conduct

Simply put, I will not tolerate cheating or plagiarism.

Any incident of dishonesty, copying, exam cheating, or plagiarism will be reported to the Office of Student Rights and Responsibilities.

Cheating will earn you a grade of 0 on the assignment and an overall grade penalty of at least 10%. In circumstances of flagrant cheating, you may be given a grade of F in the course.

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

For more information about what constitutes cheating and plagiarism, please see academicprograms.calpoly.edu/content/academicpolicies/Cheating.