Stat 331/531: Statistical Computing with R
Winter 2025

Communication

Dr. Immanuel Williams

✉ Email: imwillia@calpoly.edu

▦ Office: Building 25 Office 113 (by Statistics Department Office)

Course Logistics

Class Meeting Times: Tuesdays/Thursdays

- Section 72: 4:10pm - 6:00pm

Room: 180-272 (Baker Ctr Sci & Math 180)

Learning Hours: are held in my office (25-113) during the following times:

Sun	Mon	Tue	Wed	Thu	Fri	Sat
	9:00am–11:00am		9:00am–12:00pm			

Note that learning hours are not just for when you have content questions. Stop by to introduce yourself, ask questions about the broader field of statistics, or share what you are working on!

Course Description

Stat 331/531 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.

Definition of ETV

ETV—Extraction, Transformation, and Visualization (or Analysis)—is a structured framework inspired by the ETL (Extraction, Transformation, Load) process commonly used in data warehousing. While ETL focuses on preparing data for storage and querying, ETV emphasizes workflows that culminate in actionable insights through visualization or analysis, reflecting real-world data analysis tasks.

1. Extraction: Acquiring and importing data from diverse sources (e.g., databases, files, APIs). This step focuses on ensuring data is properly retrieved, structured, and ready for subsequent processing.
2. Transformation: Cleaning, reshaping, and standardizing data to prepare it for meaningful interpretation. This may include filtering, aggregating, joining datasets, or handling missing values to ensure data quality and usability.
3. Visualization/Analysis:

   - Visualization: Creating impactful visual representations of data, enabling clear communication of insights.
   - Analysis: In cases where visualization is not the goal, conducting statistical or computational analyses to derive conclusions and support decision-making.

By adapting the foundational ETL process to include visualization or analysis, ETV provides a comprehensive framework for both technical and analytical data tasks, forming the backbone of this course.

———————————————————

Learning Objectives

This course will teach you the foundations of statistical computing principles in the language of R, guided by the ETV (Extraction, Transformation, Visualization) process.

After taking this course, you will be able to:

- Extraction:

  – Work with the RStudio Integrated Development Environment (IDE) and Quarto documents.
  – Import data from a wide variety of sources and formats (e.g., CSV, Excel, APIs).

- Transformation:

  – Manage and clean datasets, including handling missing data, filtering, and reshaping using tidy principles.
  – Write efficient, well-documented, and tidy R code for data manipulation.

- Visualization:

  – Create informative and effective visualizations to explore and present data.
  – Build Shiny applications to display interactive visualizations and summary tables.

Additionally, this course aims to help you:

- 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 as a tool for creativity, exploration, and fun!
- Contribute to and participate in the R Open Source Community.

Course Tools

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://gato365.github.io/stat331-calpoly-text/.

This text has been modified from material by Dr. Susan VanderPlas. See Statistical Computing using R and Python for her course book 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](#)
- [Introduction to Modern Statistics](#)
- [Advanced R](#)

🖥️ 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

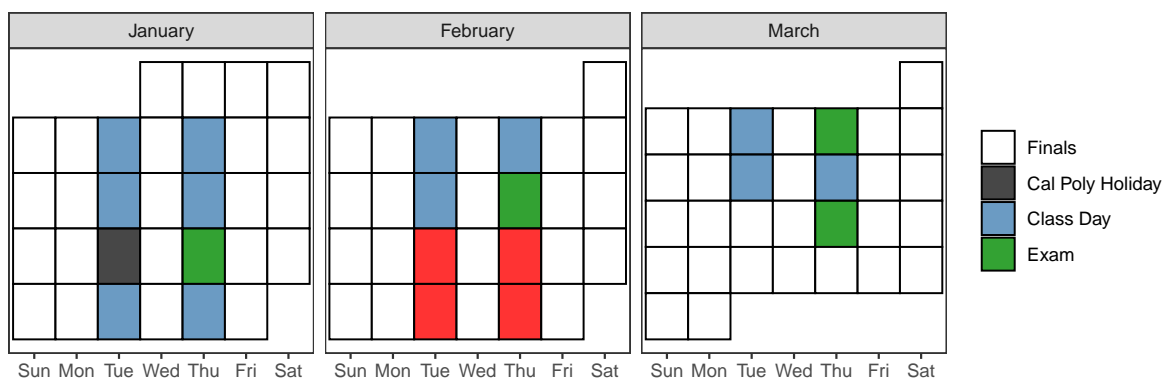This schedule is tentative and subject to change.

Figure 1: Note: Tuesday, January 21st will follow a Monday class schedule due to Martin Luther King's Day. Also, 18-20 and 25-27 of February are off due to Instructor not being in town.

Table 1: Tentative schedule of class topics and important due dates

Date	Topic
Jan 23	Midterm Exam 1
Feb 13	Midterm Exam 2
Mar 6	Midterm Exam 3
Mar 20	Final Exam
Jan 7, Jan 9, Jan 14, Jan 16, Jan 28, Jan 30, Feb 4, Feb 6, Feb 11, Mar 4, Mar 11, Mar 13	NA

Note: Please note that our textbook is currently being updated to include additional chapters that are essential for our course. I appreciate your patience and understanding as these updates are made throughout the quarter.

> 💡 Final Exam on Thursday, March 20:
>
> - Section 72 at 7:10pm –10:00pm

Course Policies

Assessment/Grading

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

Grading Breakdown

Assignments	Weight
Tuesday Quizzes	5%
Friday Quizzes	15%
Homework Assignments	10%
Midterm Exams	35%
Final Exam	35%

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 social media 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

This course focuses on developing a thoughtful and analytical problem-solving approach, emphasizing the ETV process—Extraction, Transformation, and Visualization—a framework inspired by the widely-used ETL (Extract, Transform, Load) process in data science and data warehousing. The goal of ETV is to simulate real-world data workflows, enabling students to handle data from acquisition to actionable insights effectively.

1. Extraction: Students must demonstrate their ability to acquire data from diverse sources. This step involves understanding the structure of raw data and preparing it for further processing, mirroring real-world challenges in data integration.

2. Transformation: Beyond technical correctness, students are expected to actively engage with tasks such as cleaning, filtering, reshaping, and normalizing data. Intellectual engagement is demonstrated through the clarity and rationale behind their transformations, reflecting a deep understanding of the material.

3. Visualization: Students will learn to create meaningful visual representations of data to communicate insights effectively. This step emphasizes not only technical proficiency but also the ability to think critically about how best to present information.

In every task, engagement with the material is paramount. This means that technically correct but low-effort answers that do not demonstrate understanding or effort will not receive full credit. Similarly, while simplicity is valued, it must not come at the expense of depth or accuracy.

Assignment Breakdown

Tuesday Quizzes

- Purpose: These quizzes assess your understanding of material introduced in pre-class videos and covered during the previous week's classes. They are designed to ensure that you've grasped foundational concepts and are prepared for deeper application in subsequent tasks.

- Format: Each quiz consists of 3 short questions that focus on recall and basic application of concepts. You will have a single attempt, and these quizzes are open-note, emphasizing the importance of preparation and active engagement with class materials.

- What to Expect: Questions may include definitions, identification of key functions, or simple applications to reinforce your learning. Do not use any AI software or any resources outside of the notes for these quizzes.

- Students will have 1 attempt to complete the quiz. You will not be given time to complete the quiz in class.

- Opens on Every Tuesday at 8:00am and closes on Wednesday at 11:59pm.

Homework Assignments

- Purpose: Homework provides structured practice with key functions and concepts discussed in class, often through the lens of the ETV (Extraction, Transformation, Visualization) framework. Assignments also introduce new tasks or functions not covered in class, with clear instructions on their use, encouraging independent learning and exploration.

- Format: Each assignment includes a series of coding exercises and applied tasks. These may involve completing an ETV cycle, solving real-world data problems, or using specific functions to manipulate and visualize data. There will 12 questions per homework assignment.

- What to Expect: Homework emphasizes problem-solving and creativity, often requiring you to integrate multiple concepts into a cohesive workflow. Use AI and other resources to help you problem solve, but do not use them to complete the homework.

- Students will have 5 attempts to complete the homework.

- Opens on Thursday at 8:00am and closes on the following Wednesday at 11:59pm.

Friday Quizzes

- Purpose: These quizzes focus on evaluating your ability to apply the concepts, functions, and techniques covered during the week's classes. They are an opportunity to demonstrate mastery of the material in a practical context.

- Format: Each quiz includes 10 questions, blending conceptual understanding with applied problem-solving. Unlike Tuesday quizzes, these are more comprehensive and designed to integrate ideas from the week's tasks.

- What to Expect: Questions may involve interpreting code, applying multiple steps of the ETV framework, or solving problems based on class discussions and tasks. Do not use any AI software or any resources outside of the notes for these quizzes.

- Students will have 1 attempt to complete the quiz. You will not be given time to complete the quiz in class.

- Opens on Every Friday at 8:00am (the days you do not have a midterm) and closes on Saturday at 11:59pm.

Midterm Exams

- Purpose: Midterms assess your comprehensive understanding of the material covered up to the exam date. They test your ability to demonstrate mastery through problem-solving, practical application, and reflection on how tasks connect to broader data analysis principles.

- Format: Exams may include multiple-choice questions, coding-based tasks, and short-answer explanations. You will be asked to apply the ETV framework and other concepts to solve structured problems and demonstrate critical thinking. There will be 8 questions per midterm.

- What to Expect: Midterms challenge you to show how well you can perform tasks learned in class and connect them to previous concepts.

- Students will have 1 attempt to complete the midterm. You will complete midterms in class.

- Thursdays Week 3, 6 and 9. These exams can between 50-75 minutes.

Final Exam

- Purpose: The final exam evaluates your cumulative understanding of the entire course. It ensures you can integrate and apply the concepts, techniques, and tools learned over the quarter in a coherent and meaningful way.

- Format: The exam will consist of a mix of conceptual questions, coding tasks, and applied scenarios requiring a complete ETV workflow. You may also be asked to reflect on how your problem-solving skills have developed throughout the course. There will be 26 questions on the final exam.

- What to Expect: Expect questions that ask you to apply all stages of ETV to complex, real-world datasets, along with explanations of your approach and decision-making process.

- Thursday, March 20.

Just because I like numbers, I just want to let you know there will be 270 questions in total for the quarter.

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

Given the importance of maintaining consistency and fairness for all students, this course has a strict policy regarding deadlines. It is essential that you manage your time effectively and meet all assignment deadlines.

No Late Work Accepted

- Strict Deadlines: Assignments must be submitted by their specified deadlines. Late submissions will not be accepted under any circumstances.

- Responsibility: It is your responsibility to stay on top of your work and adhere to the schedule. Plan ahead to ensure you can meet all deadlines.

- Exceptions: There will be no "grace days" or penalty-free extensions. This policy is in place to encourage timely submission and to ensure that all students are treated equally.

Managing Your Workload

- Plan Ahead: Keep track of all due dates and plan your work schedule accordingly. Avoid last-minute submissions to ensure the quality of your work.
- Stay Informed: Regularly check course announcements and email communications for any updates on assignments or deadlines.

Extenuating Circumstances

- While no late work is accepted, I understand that unforeseen circumstances can occur. If you have a serious issue that might affect your ability to meet a deadline, please contact me as soon as possible before the due date to discuss your situation.

Extra Credit Opportunity

If you see me on campus, you know me. Ask me a statistical question. You will fill out this survey an I will give points to you at the end of quarter.

I currently do not know how many points.

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."

> 💡 Sign Contract
>
> - Take time to read this document from time to time to feel empowered by this class and our effort to make this class an experience.

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.

Learning Environment and Support

I am committed to creating a safe and inclusive learning environment where all students feel respected and supported. If there are any ways I can improve the classroom environment to make it more welcoming for you, please don't hesitate to let me know.

If you have a disability and require accommodations to fully participate in the course, please contact me as soon as possible to discuss how I can best support you. I also encourage you to register with Cal Poly's Disability Resource Center (Building 124, Room 119 or at 805-756-1395) to explore additional accommodations that may be available to you.

If you are experiencing food insecurity, housing instability, or other challenges that may impact your ability to succeed in this course, please refer to the resources listed on Canvas under "Student Support Services at Cal Poly." These resources provide a range of essential support services, including emergency financial assistance, counseling, and academic support.

I am committed to working with you to ensure that you have the resources and support you need to succeed in this course. Let's work together to create a positive and inclusive learning environment for all students.

Academic Integrity and Class Conduct

Academic integrity is a fundamental value of this course and of the university. Simply put, I will not tolerate cheating, plagiarism, or any other form of academic dishonesty.

Any incident of academic misconduct, including dishonesty, copying, 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.

It is important to note that paraphrasing or quoting another's work without proper citation 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. This includes using Chat GPT, which should only be used to help you problem solve just as a tutor or peer would, and not as a substitute for your own work. 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.

To ensure academic integrity, please be sure to cite all sources appropriately and only use Chat GPT in an ethical manner. For more information on academic misconduct and what constitutes cheating and plagiarism, please see academicprograms.calpoly.edu/content/academicpolicies/Cheating.

California Faculty Association

The California Faculty Association (the labor union of Lecturers, Professors, Coaches, Counselors, and Librarians across the 23 CSU campuses) is in a difficult contract dispute with

California State University management. It is possible that we will call a strike or other work stoppage this term. I promise to promptly inform you of any schedule disruption. Our working conditions are your learning conditions; we seek to protect both. For further information go to www.CFAbargaining.org.