

Statistical Programming with R Section 01

MATH 167R

Spring 2024 3 Unit(s) 01/24/2024 to 05/13/2024 Modified 02/03/2024

Contact Information

Instructor: Dr. Peter A. Gao

Email: peter.gao@sjsu.edu

Office: MH 311

Website: peteragao.github.io

Email or the course discussion board is the best way to reach me. Email should be used for administrative/logistics issues and the discussion board should be used for questions on course material or homework. Questions about the homework submitted by email will generally be redirected to the discussion board. I will make every effort to respond to emails/discussion posts within a day, but please feel free to send me a reminder after 48 hours have passed. Please include [MATH 167R] in your subject line.

Office Hours

Monday, Wednesday, 10:30 AM to 11:30 AM, MH 426

or by appointment

Course Description and Requisites

Introduction to the R programming language. Topics include data structures, reading and writing data, databases, data visualization, accessing and creating packages, programming structures and functions.

Prerequisite(s): MATH 161A with a grade of "C-" or better. Allowed Declared Major: Mathematics major/minor, MS Statistics, MS Data Science, or MS Bioinformatics.

Letter Graded

Classroom Protocols

Collaboration

On most assignments, collaboration is allowed and encouraged. You may discuss problems, approaches, and solutions with your classmates. Acceptable collaboration is limited to your classmates in this course and you must clearly include on any collaborative work the name(s) of anyone with whom you worked. Additionally, all submitted work must be your own; you should not submit code or answers copied from any resource including your classmates. Plagiarism and cheating is easy to detect and can lead to serious negative consequences for you. If you have any questions regarding this policy, please ask for clarification.

Online resources

Students may use external online resources including discussion forums (ex. StackOverflow) large language model-based chatbots (ex. ChatGPT) **as aids for learning and understanding course material**. However, submitting code or answers for course labs, assignments, exams, and projects obtained using external resources like StackOverflow or ChatGPT is not permitted. **When external resources are consulted for an assignment, they must be cited clearly at the top of your submission.**

Discussion

You are encouraged to participate on the discussion forum by posting questions about assignments and answering questions from other students. Posts may not include substantial amounts of code that can be used for a solution to any problem, but may include code snippets within reason. Participation, in the form of both questions and answers, can earn you up to 2% extra credit for your final grade. Posts will be evaluated based on how substantive and helpful they are to the class.

Course Goals

Upon successful completion of this course, students will be able to:

1. Understand the structures of R objects
2. Import data from a variety of sources.
3. Save data in formats that can be used by other programs.
4. Create publication quality graphs.
5. Download and install packages.
6. Create reusable functions.
7. Perform statistical analysis on R.

This course aims to help you build a foundation of computational skills for data analysis. Data encountered in real world applications are usually messy, breaking many of the assumptions we make in typical statistics courses. Throughout this course, we will practice using computers to help us understand, summarize, visualize, and model complex data in a reproducible way.

Course Materials

The primary text for the course will be the lecture slides, which will be posted to the course website. The following textbooks may be useful resources.

- Golemund, G. (2014). Hands-On Programming with R. Available [here](#) for free.
- Wickham, H. and Golemund, G. (2023). R for Data Science. Available [here](#) for free.
- Healy, K. (2019). Data Visualization: A Practical Introduction. Draft version available [here](https://socviz.co/) (<https://socviz.co/>).
- Çetinkaya-Rundel, M. and Hardin, J. (2021) Introduction to Modern Statistics. Available [here](#) for free.
- Lee, B.L. (2016). An Introduction to Computational Probability and Statistics with R (Draft).

The main way we will interact with R is through the software RStudio, both of which can be downloaded and installed for free. The students will be required to have access to a computer with R and RStudio. The computer lab in MacQuarrie Hall 221 contains computers with all of the software that will be used during the semester. All of the coursework may be completed on a personal computer and the software is freely available to students.

Course Requirements and Assignments

- **Check-ins:** Throughout the semester, you will be given short check-in assignments. These are designed to be completed during class or shortly after and will usually be due at the start of the next class. At the end of the quarter, your lowest check-in grade will be dropped. They will be graded on the following two point scale:
 - 0: indicates incomplete or unacceptable work
 - 1: represents demonstrated effort towards completing at least 75% of the assignment
 - 2: represents demonstrated effort towards completing the entire assignment.
- **Labs:** Throughout the semester, you will also be given lab assignments. These are extended, more complicated assignments that you will likely not be able to complete during class. You will typically have two weeks for a lab assignment. At the end of the quarter, your lowest lab grade will be dropped.
- **Final exam:** There will be one final exam during finals week. Practice questions will be provided in advance of the exams. There will be no make-up exams.
- **Final project:** During the semester, you will complete a class project that requires you to apply the data manipulation, visualization, and analysis skills covered in this course to a real-world dataset of your choice. Throughout the semester, you will submit a project proposal and multiple checkpoint assignments that will help you to prepare a final report and presentation.

- **Late Work:** In general, the late policy is as follows: Any assignment that is received late but less than 24 hours late will receive a grade penalty of 25%. Any assignment that is received 24 to 48 hours late will receive a grade penalty of 50%. Assignments will not be accepted more than 48 hours late. Extensions for academic purposes (ex. conference presentation or job interview) or extreme circumstances (ex. illness, emergency) will generally be granted. If you would like to request an extension, please do your best to email me at least 24 hours before an assignment is due, along with the reason.
- Success in this course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally three hours per unit per week) for instruction, preparation/studying, or course related activities, including but not limited to internships, labs, and clinical practica. Other course structures will have equivalent workload expectations as described in the syllabus.

✓ Grading Information

Your final grade will be calculated as follows:

- 10%: Check-ins
- 40%: Labs
- 20%: Final exam
- 30%: Final project

There may be the opportunity for extra credit.

Letter Grade	Raw Percentage
A plus	96 to 100%
A	93 to 95%
A minus	90 to 92%
B plus	86 to 89%
B	83 to 85%
B minus	80 to 82%
C plus	76 to 79%

C	73 to 75%
C minus	70 to 72%
D plus	66 to 69%
D	63 to 65%
D minus	60 to 62%
F	0 to 60

University Policies

Per [University Policy S16-9 \(PDF\)](http://www.sjsu.edu/senate/docs/S16-9.pdf) (<http://www.sjsu.edu/senate/docs/S16-9.pdf>), relevant university policy concerning all courses, such as student responsibilities, academic integrity, accommodations, dropping and adding, consent for recording of class, etc. and available student services (e.g. learning assistance, counseling, and other resources) are listed on the [Syllabus Information](https://www.sjsu.edu/curriculum/courses/syllabus-info.php) (<https://www.sjsu.edu/curriculum/courses/syllabus-info.php>) web page. Make sure to visit this page to review and be aware of these university policies and resources.

Course Schedule

Date	Topic
W 1/24	Course Overview and Introduction to R
M 1/29	R fundamentals
W 1/31	R fundamentals
M 2/5	R fundamentals
W 2/7	Exploring data with R
M 2/12	Exploring data with R

W 2/14	Data visualization
M 2/19	Data visualization
W 2/21	Data visualization
M 2/26	Data manipulation and management
W 2/28	Data manipulation and management
M 3/4	Programming basics
W 3/6	Programming basics
M 3/11	Programming basics
W 3/13	Programming basics
M 3/18	Programming basics
W 3/20	Programming basics
M 3/25	Programming basics
W 3/27	Simulations
M 4/1	<i>Cesar Chavez Day - No class</i>
W 4/3	<i>Spring Break - No class</i>
M 4/8	Simulations
W 4/10	Simulations
M 4/15	Simulations

W 4/17	Statistical inference
M 4/22	Statistical inference
W 4/24	Statistical inference
M 4/29	Statistical inference
W 5/1	Prediction
M 5/6	Prediction
W 5/8	Prediction
M 5/13	Final presentations and review