



THE OHIO STATE UNIVERSITY

STAT 5731

Introduction to R for Data Science I: Basic R

(28164)

Autumn 2026

Syllabus - Instructor Information

Vincent Q. Vu

Professional Qualifications

- Ph.D., Statistics, University of California, Berkeley
- M.A., Statistics, University of California, Berkeley
- B.A., Statistics, University of California, Berkeley
- Associate Professor, Department of Statistics, The Ohio State University

Syllabus - Required or Recommended Readings

Required

- [R4DS2E] Wickham, Çetinkaya-Rundel, and Grolemund (2023): R for Data Science, 2nd Edition. Electronic version: R for Data Science (2e). This web version of the book can be accessed freely from any web browser.
- Weekly modules with content created by the instructor for the exclusive use of students enrolled in the course, including lecture slides, lecture recordings, and accompanying course pages. See the course schedule for topics.

Syllabus - Course Schedule

WeekTopic

1	Introduction to R, RStudio and very basic Quarto
2	Basic visualization with ggplot2
3	Data transformation and summarization
4	Good coding style; Good data style — data tidying
5	Scripts and projects; Data import
6	From data to analysis to final product with Quarto
7	Final Exam

Course Details - Instructor Contact Details

- Email: vqv@stat.osu.edu

My preferred method of contact is the CarmenCanvas Inbox. If you email me, please include “STAT 5731” in the subject line so that I can identify it quickly. When you write an email, please be clear and specific about your question or issue. It may be helpful to include a screenshot, but do not use a screenshot as the only form of communication. If you have a question about a specific problem, please explain in words what you are trying to do and what you have tried so far. The more context you can provide, the better I can help you. If you do not provide any context, I will reply with a very brief response, but do not interpret my brevity as a sign of rudeness or lack of interest. I am happy to help you, but I need to know what you are trying to do and what you have tried so far.

Course Details - Office Hours

See CarmenZoom for schedule and meeting links.

Course Details - Class Meeting Details

Class Meeting Details: TBA - Online -

Course Details - Course Description

The first course in a sequence designed for teaching students how to use R effectively for doing data science. This course introduces the basic flow and focuses on basic usage of important tools in R for visualization, transformation, and organization of data. Prereq: 1350, 1350.01, 1350.02, 1430, 1430.01, 1430.02, 1450, 1450.01, 1450.02, 1550, 2450, 2450.01, 2450.02, 2480, 2480.01, 2480.02, 3201, 3202, 3450, 3450.01, 3450.02, 3460, 3470, 3470.01, 3470.02, 4202, 5301, or 5302, or equiv., or graduate standing, or permission of instructor. Not open to students with credit for Stat 5730.

Prerequisite(s) / Corequisites: See course description.

Course Details - More About This Course

Course Description

R is a freely available statistical computing environment and programming language. It has become a dominant workhorse for modern statistical research and data analysis and has been widely adopted in industrial data analytics as well. This course is part of a sequence whose goal is to teach students how to use R effectively for doing data science – importing raw data and transforming it into insights and knowledge that can be communicated with others. Throughout the sequence, there will be an emphasis on coding practices for ensuring reliability, reproducibility, and transparency of data analyses. Part I of the sequence introduces the basic workflow and focuses on basic usage of important tools in R for visualization, transformation, and organization of data.

How this online course works

Mode of delivery

This course is 100% online. There are no required sessions when you must be logged in to Carmen at a scheduled time. I will send all important class-wide messages through the Announcements tool in Carmen. Please check your notification preferences to ensure you receive these messages.

Pace of online activities

This course is divided into weekly modules, consisting of short video lectures and assignments, that are generally released at the beginning of each week. Students are expected to keep pace with weekly deadlines but may schedule their efforts freely within that time frame.

Because this is an accelerated course, there is an intentional overlap between modules. New modules release at the start of each week, but homework for each module isn't due until later, so you will often be finishing the homework for one module while starting the videos and discussion for the next.

Jumping between modules can feel a bit disjointed at times. Because the modules release weekly, it is important not to let the homework delay your start on the new material. My advice is to prioritize each week's discussion assignment, as it keeps you anchored in the current week's material, while homework acts as a delayed review of the previous week. Discussion assignments are graded complete/incomplete to help facilitate that transition.

Credit hours and work expectations

This is a 7-week, 1-credit-hour course. According to Ohio State policy, students should expect around 2 hours per week of time spent on direct instruction (instructor content and Carmen activities, for example) in addition to 4 hours of homework (reading and assignment preparation, for example) to receive a grade of at least (C) average.

Participation requirements

Because this is an online course, your attendance is based on your online activity and participation. The following is a summary of students' expected participation:

Participating in online activities

You are expected to log in to the course in Carmen every week. Online activities include watching prerecorded lecture videos and posting in Carmen discussion forums, including a weekly discussion assignment. During most weeks you will probably log in many times. If you have a situation that might cause you to miss an entire week of class, discuss it with me as soon as possible.

Communication expectations

- Respect and Professionalism: Communicate with courtesy and consideration at all times.
- Writing Standards: Use proper grammar, spelling, and punctuation.
- Tone and Civility: Foster a supportive environment and avoid sarcasm.
- Providing Context: When asking questions—whether by email or in discussion posts—provide enough detail for others to understand the issue (e.g., describe what you are trying to do, what you have already tried, and include supporting information if helpful). Screenshots may be included, but should not be your only explanation. Without context, responses will necessarily be brief.

Course Details - Course Learning Outcomes

Course expected learning outcomes

By the end of this course, students should successfully be able to:

1. Produce basic visualizations of data in R using the ggplot2 package.
2. Transform and summarize data in R for generating insights.
3. Organize (i.e. tidy) data in R in such a way that makes it easily amenable to visualization and analysis.

4. Import data in various tabular formats into R.
5. Organize code, data, and analyses into scripts and projects.
6. Author dynamic, reproducible documents that combine code, results, and prose with Quarto.

Course Details - Course Supplies

Course technology

Required software

You will need to install or have access to the latest versions of the following software. Even if you have installed these programs before, it is a good idea to check for updates. I will not provide support for outdated software.

- R: [Download R](#)
- RStudio Desktop IDE: [Download RStudio Desktop](#)
- Quarto: [Install Quarto](#)

Windows ARM-based PCs are incompatible with Quarto

As of August 2026, Windows ARM-based PCs, e.g. Microsoft Surface, are not fully compatible with Quarto, and users may encounter issues rendering Quarto documents when using these systems. If you are using a Windows ARM-based PC, please contact me as soon as possible so that I can provide an alternative, cloud-based solution for rendering Quarto documents.

Course Details - Detailed Course Schedule

Assignment due dates are in Carmen, which is authoritative — including any individually adjusted deadline.

Week	Date	Topics and readings
------	------	---------------------

1	8/25	Introduction to R, RStudio and very basic Quarto
		Reading: R4DS2E Introduction, 2, 28.1--.5.1
2	9/1	Basic visualization with ggplot2
		Reading: R4DS2E 1
	9/7	No class (Labor Day — university holiday)
3	9/8	Data transformation and summarization
		Reading: R4DS2E 3
4	9/15	Good coding style; Good data style — data tidying
		Reading: R4DS2E 4--5
5	9/22	Scripts and projects; Data import
		Reading: R4DS2E 6--8
6	9/29	From data to analysis to final product with Quarto
		Reading: R4DS2E 28--29
7	10/7	Final Exam

Course Details - Canvas Assignment Schedule

Date	Name	Assignment Type	Points
	Final Exam: Technical Q&A Only	Discussion	0
	General course Q&A	Discussion	0
	Module 1 Q&A	Discussion	0

Date	Name	Assignment Type	Points
	Module 2 Q&A	Discussion	0
	Module 3 Q&A	Discussion	0
	Module 4 Q&A	Discussion	0
	Module 5 Q&A	Discussion	0
	Module 6 Q&A	Discussion	0
8/28/26	Discussion for Module 1: Introduce yourself!	Discussion	1
9/4/26	Discussion for Module 2: Answer the quiz and share a plot	Discussion	1
9/8/26	Homework 1	Assignment	100
9/11/26	Discussion for Module 3: Find the quiz slide and brainstorm a way to compare the distribution of departure delays by month	Discussion	1
9/15/26	Homework 2	Assignment	100
9/18/26	Discussion for Module 4: What makes data tidy or untidy?	Discussion	1

Date	Name	Assignment Type	Points
9/22/26	Homework 3	Assignment	100
9/25/26	Discussion for Module 5: How do you organize the files and folders on your computer?	Discussion	1
9/29/26	Homework 4	Assignment	100
10/2/26	Discussion for Module 6: Which module was your favorite? What is your most used keyboard shortcut?	Discussion	1
10/6/26	Homework 5	Assignment	100
10/13/26	Final Exam	Assignment	100

Course Details - Grading Policy

Grading

How your grade is calculated

Assignment categoryWeight

Discussion (lowest score dropped; complete/incomplete)	10%
Homework	60%
Final Exam	30%
Total	100%

Description of major course assignments

Discussion

- Description

You are expected to watch all asynchronous lectures posted on the Carmen course site. Each module will include a discussion assignment with a varying prompt—for example, responding to a lecture quiz question, applying a technique (e.g., creating a plot), or modifying an example from the lecture.

These discussion assignments are designed to keep you engaged with the course material and to encourage learning from peers. You will not see other students' posts until you have submitted your own. Assignments are graded on a complete/incomplete basis (not correctness) and one lowest score will be dropped.

- Academic integrity and collaboration guidelines

Discussion posts should be your own work. They will generally require you to repeat an example or answer a question posed in the lecture videos. These assignments will be graded based on completion rather than correctness.

- Permitted uses of AI tools

You may not use AI tools to write these discussion posts.

Homework

- Description

Homework will be assigned (approximately) weekly, will be due on announced dates and will be graded. Learning to compute and program requires practice. Homework assignments will mainly consist of exercises designed to reinforce the concepts covered in class during the previous week.

- Academic integrity and collaboration guidelines

You may collaborate with classmates on your homework, but ultimately the code that you write and submission that you make must be your own work. For example, I encourage you to discuss strategies for solving problems, but the actual code and explanations that you write must be your own. Moreover, keep in mind the university policies on plagiarism. Do not copy or plagiarize anything you may find on the Internet or anything produced by an AI tool.

- Permitted uses of AI tools

You may use generative AI tools, e.g. ChatGPT, Claude, Gemini, etc... to help you understand concepts. For example, you can ask an AI tool about a specific function and its usage. However, you are not permitted to use of AI tools to write code or produce text for you. In particular, you should not use AI tools to write code for your homework assignments.

- Required disclosures

Any collaboration or use of AI tools must be acknowledged in the disclosures section of your homework submission.

Final Exam

- Description

There will be a “take home” final exam during the last week of the course. It will involve R programming and data analysis/visualization. The exam will be designed to take about 3 hours to complete, but it will be untimed so that you will have several days to complete it. Additional information about the exam will be announced on Carmen.

- Academic integrity and collaboration guidelines

The exam is an individual assignment. You should complete the exam on your own and your submission should be your own original work. You should not discuss it with anyone else.

- Permitted uses of AI tools

The use of AI tools is not allowed at all.

Late assignments

- Discussion: not accepted after the due date.
- Final exam: not accepted after the due date.
- Homework: accepted up to 5 days late, with a 10 percentage point deduction per day.

If you absolutely need to turn in an assignment late and have a valid excuse, please contact me for the necessary arrangements. However, you must notify me in advance in such a situation. Exceptions to this policy will be permitted only in extreme situations such as serious injury immediately prior to an assignment being due or severe illness requiring hospitalization.

Instructor feedback and response time

Please use the discussion board in Carmen for questions about the course material and assignments. If you have a question that is personal or that you would like to discuss privately, please email me. I will respond to questions posted on the discussion board or by email within 24 hours (except for weekends and university holidays). If you would like to meet with me over Zoom, please message me to set up an appointment.

Grading and feedback

For weekly assignments, you can generally expect feedback and grades within 10 days.

Course Details - Grading Scale

Letter Grade	Percentage Range
A	100% to 93%
A-	< 93% to 90%
B+	< 90% to 87%
B	< 87% to 83%

Letter Grade	Percentage Range
B-	< 83% to 80%
C+	< 80% to 77%
C	< 77% to 73%
C-	< 73% to 70%
D+	< 70% to 67%
D	< 67% to 60%
E	< 60% to 0%

Course Details - AI Use Policy

Generative AI tools and this course

The goal of this course is to help you learn how to use R for data science. This includes basic coding skills, data wrangling, and data visualization. These skills are a fundamental part of becoming proficient in using R, so it is important that you learn how to do them yourself.

Generative AI tools (e.g., ChatGPT, Gemini, and Claude) can be very powerful aids, but they are not a substitute for learning, and relying on them from the beginning can hinder your development. These tools often produce incorrect or misleading information, and they can be overly confident in their answers. This is easy for skilled R coders and data scientists to spot, but it can be very difficult for beginners like you. So it is important that you learn how to do things yourself first, and then, after completing this course, you can use AI tools to help with tedious/non-creative aspects of your work once you become more proficient.

I do permit students to use AI tools to help with understanding concepts and debugging errors. For example, you can ask an AI tool about a specific function and its usage, or you can use it to help you debug your code and decipher error messages. In this way, you may only use AI tools as an enhanced search engine.

However, you are not permitted to use AI tools to write code or produce text for your assignments.

If I suspect that you have misused an AI tool on an assignment for this course, I

will ask you to explain your process for completing the assignment in question. The unauthorized use of AI tools will result in referral to the Committee on Academic Misconduct.

The assignment and academic policies sections of the syllabus has more details on the use of AI tools in specific assignments and policies.

Syllabus - University Policies

Office of Undergraduate Education - Standard Syllabus Statements

- [Academic Misconduct](#)
- [Religious Accommodations](#)
- [Accommodations for Disability and Illness](#)
- [Intellectual Diversity](#)
- [Grievances and Solving Problems](#)
- [Creating an Environment Free from Harassment, Discrimination, and Sexual Misconduct](#)
- [Copyrights](#)
- [Content Warnings](#)
- [Mental Health Support](#)
- [Resources for Military-Connected Students](#)
- [Campus-wide Class Cancellations](#)