Syllabus

STAT 5731

**Introduction to R for Data Science I: Basic R**

* Autumn 2025
* 1 Credit Hour
* Online

## Course overview

### Instructor

* Vincent Q. Vu, Ph.D.
* Email: [vqv@stat.osu.edu](mailto:vqv@stat.osu.edu)
* Office hours on Zoom on Thursdays TBD, except holidays
  + Fridays, 9:00am — 10:00am. See CarmenZoom for schedule and meeting links.

|  |
| --- |
| Note |
| My preferred method of contact is the CarmenCanvas Inbox. When writing an email, 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 Description

[R](https://www.r-project.org) 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.

### 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](https://ggplot2.tidyverse.org) 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](https://quarto.org).

|  |
| --- |
| Generative AI tools and this course |
| The learning goals of this course are to help you learn how to use R for data science. This includes basic coding skills, data wrangling, and data visualization. Given that these skills are a fundamental part of becoming proficient in using R, it is important that you learn how to do them yourself.  Generative AI tools, such as 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. Importantly, 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 use it help you debug your code and decipher error messages. Basically, you can use AI tools like 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 used 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](https://oaa.osu.edu/academic-integrity-and-misconduct).  The assignment and academic policies sections of the syllabus has more details on the use of AI tools in specific assignments and policies. |

## Prerequisites

STAT 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.

## 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](go.osu.edu/canvas-notifications) 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.

### Credit hours and work expectations

This is a 7-week, 1-credit-hour course. According to [Ohio State policy](go.osu.edu/credithours), 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 also 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 materials and technologies

### Textbooks

#### Required

* [**R4DS2E**] Wickham, Çetinkaya-Rundel, and Grolemund (2023): *R for Data Science*, 2nd Edition. Electronic version: <r4ds.hadley.nz>. This web version of the book can be accessed freely from any web browser.

### 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** <https://cloud.r-project.org>
* **RStudio** Desktop IDE <https://posit.co/download/rstudio-desktop>
* **Quarto** <https://quarto.org/docs/get-started>

|  |
| --- |
| Important |
| Note that as of August 2025, 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. |

## Grading and instructor response

### How your grade is calculated

| Assignment Category | Points and/or Percentage |
| --- | --- |
| Participation | 10% |
| Homework | 60% |
| Final Exam | 30% |
| **Total** | **100%** |

### Description of major course assignments

#### Participation

* 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, like. 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 |
| Late homework assignments will automatically receive a 10 percentage point deduction each day that they are late. After the 5th day submissions will no longer be accepted. All other assignments (discussion and final exam…) will not be accepted after the due date. Please pay attention to the exact date and time that an assignment is due. For example, if an assignment is due at 11:59:00 PM, and you submit it at 10 seconds after at 11:59:10 PM, then it will be considered late.  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. |

### Grading Scale

* 93-100: A
* 90-92: A-
* 87-89: B+
* 83-86: B
* 80-82: B-
* 77-79: C+
* 73-76: C
* 70-72: C-
* 67-69: D+
* 60-66: D
* Under 60: E

|  |
| --- |
| 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.

## Academic policies

The following are important academic policies to be aware of.

* Academic Misconduct; Artificial Intelligence and Academic Integrity; Religious Accommodations; Disability Statement (with Accommodations for Illness); Intellectual Diversity; Grievances and Solving Problems; Creating an Environment Free from Harassment, Discrimination, and Sexual Misconduct.
  + [Policy statement](https://ugeducation.osu.edu/academics/standard-syllabus/standard-syllabus-statements)
* Copyright; Counseling and Consultation Services / Mental Health Statement; Content Warning Language
  + [Policy statement](https://ugeducation.osu.edu/academics/standard-syllabus/optional-syllabus-statements)

## Course Schedule

Refer to our Carmen course page for up-to-date assignment due dates.

| Week | Date | Topics/Readings/Assignments |
| --- | --- | --- |
| 1 | 8/26 | Module 1: Introduction to R, RStudio and very basic Quarto |
| 1 | 8/26 | Reading: *R4DS2E Introduction, 2, 28.1–.5.1* |
| 1 | 8/29 | **Discussion 1 due** |
| 2 | 9/2 | Module 2: Basic visualization with ggplot2 |
| 2 | 9/2 | Reading: *R4DS2E 1* |
| 2 | 9/5 | **Discussion 2 due** |
| 3 | 9/8 | **Homework 1 due** |
| 3 | 9/9 | Module 3: Data transformation and summarization |
| 3 | 9/9 | Reading: *R4DS2E 3* |
| 3 | 9/12 | **Discussion 3 due** |
| 4 | 9/15 | **Homework 2 due** |
| 4 | 9/16 | Module 4: Good coding style; Good data style — data tidying |
| 4 | 9/16 | Reading: *R4DS2E 4–5* |
| 4 | 9/19 | **Discussion 4 due** |
| 5 | 9/22 | **Homework 3 due** |
| 5 | 9/23 | Module 5: Scripts and projects; Data import |
| 5 | 9/23 | Reading: *R4DS2E 6–8* |
| 5 | 9/26 | **Discussion 5 due** |
| 6 | 9/29 | **Homework 4 due** |
| 6 | 9/30 | Module 6: From data to analysis to final product with Quarto |
| 6 | 9/30 | Reading: *R4DS2E 28–29* |
| 6 | 10/3 | **Discussion 6 due** |
| 7 | 10/6 | **Homework 5 due** |
| 7 | 10/6 | Final Exam opens |
| 7 | 10/10 | **Final Exam due** |