

STAT 231 Data Science

FALL 2021

Instructor

Brittney E. Bailey
(Prof. Bailey)
DM via Campuswire

Class meetings

01: T/Th 8:30 – 9:50 a.m.
02: T/Th 11:30 a.m. – 12:50 p.m.
Webster 102

Office hours

T/Th 2:30 – 4 p.m. ET
[Zoom](#)

SDS Fellows hours

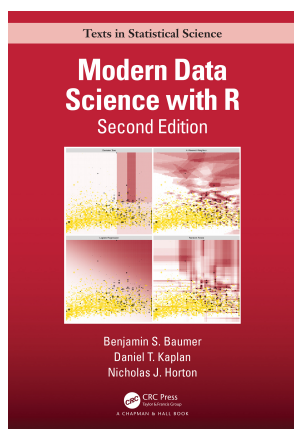
Sunday – Thursday, 7 – 9 p.m.
SCCE E208

Required* material

*See page 2 for details

Textbook

Modern Data Science with R
(2nd Edition) by
Baumer, Kaplan, Horton
[FREE web version](#)
ISBN: 9780367191498



Computer

Required for coding

Description

Computational data analysis is an essential part of modern statistics and data science. This course provides a practical foundation for students to think with data by participating in the entire data analysis cycle. Students will generate statistical questions and then address them through data acquisition, cleaning, transforming, modeling, and interpretation. This course will introduce students to tools for data management, wrangling, and databases that are common in data science and will apply those tools to real-world applications. Students will undertake practical analyses of large, complex, and messy data sets leveraging modern computing tools.

Requisite: STAT 111 or 135 (Intro Stat) and COSC 111 (Intro CS) or instructor consent.

I assume all students are familiar with [R/RStudio](#), [reproducible analysis](#), and [multiple regression](#).

Learning goals

By the end of this course, students will be able to

- ▶ demonstrate comprehensive knowledge of data wrangling using a novel dataset, including gathering, reshaping, and cleaning data, with a reproducible workflow
- ▶ demonstrate your ability to uncover patterns and narratives across a variety of data types including spatial data, textual data, and network data
- ▶ articulate a novel question that you can address with data and recognize when a question cannot be answered with data (or cannot be answered with the data at hand)
- ▶ communicate data narratives via effective visualizations, writing, and oral presentation, including acknowledging the limitations of the data and what the data cannot tell us
- ▶ support your peers, serve as resources for one another, and recognize the value of collaboration and teamwork
- ▶ identify some of the ethical considerations in data science, and contribute thoughtful opinions to current discussions in this area

Approach

This is an active learning course that requires wrestling with the material both in and out of class. Only 3 of the 12 hours that students are expected to dedicate to this course occur in person, so we will dedicate those precious few hours to active engagement with the material and timely feedback from me.

The typical flow will be:

- 1) complete assigned material, take notes on your own before synchronous sessions, and ask clarifying questions in office hours, on [Campuswire](#), or in class
- 2) actively problem-solve with peers in synchronous session
- 3) demonstrate individual comprehension through homework, reflections, and projects

“When learning is easy, it is often superficial and soon forgotten.”

Brown, et al. (2014).
Make it Stick.

Required material

Textbook We will use [Modern Data Science with R \(2nd Edition\)](#) by Baumer, Kaplan, and Horton as our primary source for learning outside of class time. You can access the text for [free online](#) or purchase a new or used copy from an [online book vendor](#).

Three copies of the textbook are available on reserve at the circulation desk of the Keefe Science Library (to your left when you enter the first floor of the Science Center) and can be checked out for 4 hours at a time.

Computer We will be working in RStudio all semester, and this work cannot be done on a tablet or cell phone (Chromebooks may also pose a challenge). Please bring your computer and charger to class each day so you are prepared to work in RStudio. If you are struggling with connectivity or other device issues (e.g., broken computer and need a loaner) please reach out to me as soon as possible and we will work with IT to get you what you need.

Technology



Moodle: Our central hub for **course content** and will be updated regularly with course information, videos, pdf previews of assignments, and eventually grades. Moodle will link out to all external resources for the course. It is your responsibility to check Moodle frequently for updates.

Campuswire: Our primary tool for **course communication** (DMs, discussion, Q&A, chat, announcements). I will send you an invitation before the semester begins. See the [Communication FAQs](#) for more detail.

- ▶ Please use DMs, not email, for any personal communications with me. You can revert to emailing me *after the course ends* to keep in touch :)
- ▶ All non-personal questions/comments should be posted in the the Class Feed or Chat.
- ▶ All announcements will also be sent via email when I post them through Campuswire.
- ▶ *Why Campuswire?* There are a lot of options for communication within the course (Moodle, Slack, Google forums, etc.). Campuswire combines the best of the features from each of those and additionally integrates Markdown to make it easier to ask and answer questions with appropriate and clear formatting for code and equations. It is also intuitive to use and saves your inbox (with appropriate adjustment to notification settings)!

GitHub: Our primary tool for **assignment distribution, collaborative coding, and file version control**.

There are many GitHub interfaces, and this class will focus primarily on using GitHub within RStudio. Guidance will be provided for getting set up and using GitHub.

Gradescope: Our primary tool for **assignment submission** and feedback. Log in via SSO.

- ▶ All uploads must be pdfs (try [Scannable iOS](#) or [Adobe Scan](#) for mobile scan-to-pdf). Visit the [Gradescope Help Page](#) for additional guidance.

RStudio: We will use R software with the RStudio interface throughout the course for statistical analysis. While it is possible to use R through a version of RStudio accessible on the web at r.amherst.edu, **I expect you to install R and RStudio on your own machine**, if possible. R, RStudio, and the necessary TeX package are all freely available.

- ▶ There are numerous resources available for help getting set up on your own machine. We recommend these [detailed instructions for installation](#) of R, RStudio, and LaTeX ([TinyTeX](#)) for Mac and Windows, which have been created by [Kitty Girjau '21](#).

Course components

I usually include the grading scheme at this point in the syllabus. This semester, the grading scheme will be set up in conversation with the students in both sections.

*We will experiment at least initially with **feedback only** on your assignments (no letter or percentage grades). Although a grade will not be provided initially, **assignments turned in for feedback will factor into your final grade in the course**, and you will not have the opportunity to make up missed assignments after their due dates.*

Weekly course components

Engagement (2–3 hours) There are a number of ways to stay engaged with the course (some less optional than others), including maintaining an active presence in class and on Campuswire (engaging with peers by asking and answering questions, sharing content), regularly accessing material through Moodle and GitHub, completing assignments, and interacting with me or the SDS Fellows.

In-class time may involve active problem-solving, data analysis, project work, or discussion of assigned readings. There will be a mixture of small-group activities, whole-class discussions, and demo-based sessions lead by me.

Active reading (2–3 hours) Reading, note-taking, and constructive learning strategies are expected of each student. To assist with these efforts, guidelines for active reading, effective note-taking, and metacognitive learning strategies will be discussed and emphasized, and resources for success will be available on Moodle.

Problem sets (3–5 hours) After actively reading the assigned material, check your initial understanding of the material by completing the **Reading Set** for the week, a set of problems designed to promote active reading of the textbook. The Reading Sets should be done on your own as you initially wrestle with the material.

During the week, get further practice with the material by working through the **Practice Set**, a set of problems designed to give you practice beyond the examples produced in the text. You may work through these problems with peers, but all work must be completed by you (see [Honor Code](#)) and you must indicate who you worked with. Even then, the best approach here is to try the problems on your own before discussing them with peers, and then write your final solutions yourself.

Assigned readings and Reading Sets should be completed by 10 p.m. ET on Mondays. This will give you the foundation you need to work through our labs on Tuesdays and Thursdays and to complete your Practice Set by Friday at 10 p.m. ET.

- ▶ I encourage you to start the assignments the day they are available. You can resubmit your work on Gradescope as many times as you want before the deadline, so you should not wait until the last minute to submit some version of your work.
- ▶ When you submit your work to Gradescope, make sure you have selected all pages that correspond to a particular problem when you upload your work to Gradescope. You will not get credit for work that is not assigned to a particular problem.
- ▶ After feedback has been provided, you are expected to review and reflect on the feedback, perhaps in consultation with any provided solutions, me, and/or the SDS Fellows to help you figure that out how to improve.

Microlectures (0.5–1 hour) Short lecture videos (*microlectures*, ~10 minutes or less) will be available roughly twice a week and should be viewed prior to the corresponding class day (e.g., you might plan to view the microlectures on Mondays and Wednesdays).

Reflections and connections

One of the most powerful tools for long-term learning is reflection, and this will be built into the course in several ways:

Learning reflections Reflect on your learning in the course (big ideas, things you are still struggling with) and/or your learning process.

Satistics connections Connect your knowledge in the course to your interests and real-world statistical encounters.

Project reflections Think about how you work in a group, how your group is functioning, and what steps can be taken to by you or your group members to improve how the group is working together.

Projects

Calendar Query (individual) This individual project will give you the opportunity to go through the entire data science cycle (asking a question, collecting data, analyzing the data, and communicating conclusions) while providing insight on a question of interest to you about how you spend your time.

Shiny App (group) You will work in small groups to create an interactive Shiny application that explores a topic of interest. The focus of this project is data wrangling, interactive data visualization, and effective communication.

GitHub Blog (group) This project will take the data analysis a step further and incorporate some of the exploratory data analysis techniques introduced in this class. This project is deliberately open-ended to allow you to explore your creativity and interests.

Weekly planning for STAT 231

Here is how you might plan your time on this course in a typical week.

Friday through Monday

- ▶ 2–5 hours: Actively read the textbook chapter(s) and work on the Reading Set for the week
- ▶ Turn in the Reading Set by 10 p.m. ET on Monday
- ▶ 15–20 minutes: Watch the microlecture before Tuesday's class time

Tuesday

- ▶ 80 minutes: Attend the class session
- ▶ Begin working on the Practice Set for the week

Wednesday

- ▶ 15–20 minutes: Watch the microlecture before Thursday's class time
- ▶ Continue working on the Practice Set for the week

Thursday

- ▶ 80 minutes: Attend the class session
- ▶ Continue working on the Practice Set for the week

Friday

- ▶ Turn in the Practice Set by 10 p.m. ET on Friday
- ▶ Reflect on what you have learned so far and what you are still struggling with

Support and resources

My pledge to support you

I strive to make this course a place where all students are welcome and all students can thrive. I look forward to working with you to understand your needs and support your academic success. If you would like to discuss your learning needs with me, please schedule a meeting.

Your mental and physical health are foundational to your overall success and take priority over academic performance. We can work together to make adjustments to the course so that you can take care of personal needs while also demonstrating your knowledge of the material. Please let me know as soon as something arises so we can work together to ensure your success in the course. I do not expect you to disclose personal details—just make me aware when flexibility might be needed so we can work something out. *I cannot and will not make adjustments after-the-fact for issues that were not discussed earlier.*

Additional support services are available on campus, including your [Class Dean](#), the [Counseling Center](#), and [student resource centers](#).

College accommodations

The college will provide accommodations and services after a student has completed the interactive process with [Accessibility Services](#) and is deemed eligible. Accommodations include but are not limited to flexibility with exams (e.g., extra time, reduced distraction, breaks, staggering), assistive devices and services (e.g., note takers), alternative formats for course materials (e.g., large print, braille), and flexibility with attendance. You can reach Accessibility Services via email at accessibility@amherst.edu or via phone at 413-542-2337. Please meet with me to discuss the best implementation of your accommodations once you have them in place—college accommodations can only be implemented after electing them for the course and notifying the instructor.

Honor code and academic integrity

"Every person's education is the product of their intellectual effort and participation in a process of critical exchange. Amherst College cannot educate those who are unwilling to submit their own work and ideas to critical assessment. Nor can it tolerate those who interfere with the participation of others in the critical process. Therefore, the College considers it a violation of the requirements of intellectual responsibility to submit work that is not one's own or otherwise to subvert the conditions under which academic work is performed by oneself or by others."

This course is filled with collaborative work, and you are expected and encouraged to work together with a partner or in small groups to study, complete assignments, and prepare for exams. However, every word that you write must be your own. Copying or "slightly rewording" sentences, paragraphs, or blocks of R code from another student is not acceptable and will receive a penalty. No interaction with anyone but the instructor is allowed on any exams or quizzes. Cases of dishonesty, plagiarism, etc., will be reported, per the full statement of the [Amherst College Honor Code](#). If you are not sure if you have plagiarized any material, a [handy diagram](#) is available to help you out.

Note on recording

Audio and/or video recording of classes without advance approval from the instructor or an approved disability accommodation is prohibited under the Student Code of Conduct. Any other audio and/or video recording of any individual without that individual's knowledge or permission (see Massachusetts General Law Part 4, Title I, Chapter 272, Section 99) is also not allowed under the code.

FAQs

ASSIGNMENTS

May I turn it in late? May I have an extension?

- ▶ Yes, at your own risk! Deadlines are set to ensure you keep up with the material, but I will accept homework until solutions are posted (usually within 12 hours of the deadline). An important part of your learning is timely feedback, and I cannot provide that if I grant individual extensions. Please manage your time accordingly. Unexpected delays/crises that occur on the day the assignment is due do not warrant extensions (submit what you have for credit).

The RStudio Server is down/frozen/broken. What do I do?

- ▶ Have you tried turning it off and back on again? Have you tried a different browser? Did you wait a few minutes and try again? If none of those worked, email askit@amherst.edu and Cc me.

Can I share your material with a friend not in this class or upload it to this site on the interwebs?

- ▶ Please do not. Hours and hours and hours and HOURS of work went into designing and creating this material for your personal success in this course. Please do not share material with a broader audience without my specific permission. Also, see [note about recording audio/video](#).

ATTENDANCE

I'm running late!

- ▶ We have limited time together so class will start promptly to make the most of that time. Out of courtesy to me and your peers, I expect you to arrive on time.

What if I have to miss a lab?

- ▶ It is your responsibility to stay engaged with the material regardless of your ability to attend class sessions. Let me know ahead of time, if possible, to ensure there are no consequences for a necessary absence. Regardless, make arrangements with a classmate to catch up on any missed material (I am unlikely to respond to "what did I miss?" or "did I miss anything important?" emails).

COMMUNICATION

I sent you a message forrrrrverrrrr ago--why haven't you responded yet?

- ▶ While I appreciate that students work at all hours, I must make time for myself and my family. As such, I will (generally) only respond to messages between 8 a.m. – 5 p.m. on weekdays, within 24 hours of receiving them (or by Monday if you email me on Friday or over the weekend) with two notable exceptions:
 - I am unlikely to respond to questions that can be answered by the syllabus or other content on Moodle, GitHub, or Campuswire. If you don't hear from me, verify the answer can't already be found elsewhere.
 - I am unlikely to respond to "what did I miss?" or "did I miss anything important?" emails.
- ▶ Campuswire sometimes glitches, so you may not get an email notification when I respond to a DM, regardless of your notification settings. You will need to check Campuswire directly.

Can I just DM you all my questions?

- ▶ Students often ask clarifying questions they think are specific to them or message me privately when they find I might have made a mistake, and I will have at least three other students DMing me the same thing. The responses are usually relevant to the entire class, and sometimes can get answered more quickly by a peer who sees it before me. Unless it's regarding a private matter, please post questions and comments to the Class Feed. You have the option to post anonymously to the class (I will still know who you are) if that makes you more comfortable.

I need to meet with you but I can't make it to office hours!

- ▶ Catch me in/before/after class or send me a DM to arrange a time to meet.