STA 199: Intro to Data Science and Statistical Thinking

Summer 2025

Table of contents

Course learning objectives . . . 2
Course materials . . . 2
- Textbooks . . . 2
- Computing . . . 2

Course components . . . 2
- Lectures and labs . . . 2
- Activities & Assessment . . . 3

Grading . . . 4
Course community . . . 5
- Inclusive community . . . 5
- Personal pronouns . . . 5
- Accessibility . . . 6
- Communication . . . 6
- Email . . . 6

Five tips for success . . . 6
Getting help . . . 7
Course policies . . . 7
- Duke Community Standard . . . 7
- Academic honesty . . . 7
- Late work & extensions . . . 9
- Waiver for extenuating circumstances . . . 10
- Regrade requests . . . 10
- Attendance policy . . . 10

Accommodations . . . 11
- Academic accommodations . . . 11
- Religious accommodations . . . 11

Course learning objectives

By the end of the semester, you will…

- learn to explore, visualize, and analyze data in a reproducible and shareable manner using R and RStudio
- gain experience in data wrangling and munging, exploratory data analysis, predictive modeling, and data visualization
- work on problems and case studies inspired by and based on real-world questions and data
- learn to communicate results through written assignments and project presentation effectively

Course materials

Textbooks

All books are **freely available online**.

- R for Data Science, 2e, Wickham, Çetinkaya-Rundel, Grolemund. O'Reilly, 2nd edition, 2023. Hard copy available on Amazon.
- Introduction to Modern Statistics, Çetinkaya-Rundel, Hardin. OpenIntro Inc., 2nd Edition, 2023. Hard copy available on Amazon.

Computing

You will need a laptop you can bring to lecture and lab for this course. We will use the statistical software R. Students will be able to access R through Docker containers provided by Duke Office of Information Technology. See the computing page for more information.

Course components

Lectures and labs

Lectures and labs are designed to be interactive, so you gain experience applying new concepts and learning from each other. My role as instructor is to introduce you to new methods, tools, and techniques, but it is up to you to take them and use them. A lot of what you do in this course will involve writing code, and coding is a skill that is best learned by doing. Therefore, as much as possible, you will work on various tasks and activities throughout each lecture and lab. You are expected to prepare for class by completing assigned readings, attending lectures and lab sessions, and meaningfully contributing to in-class exercises and discussions.

Additionally, most lectures will feature application exercises that will be graded based on completing what we do in class.

You are expected to bring a laptop (or Chromebook) to each class so that you can participate in the in-class exercises. Please ensure your device is fully charged before you come to class, as the number of outlets in the classroom will not be sufficient to accommodate everyone. A tablet also works, but the user experience will be much smoother on a laptop.

Activities & Assessment

You will be assessed based on five components: application exercises, labs, exams, project, and teamwork.

Labs

In labs, you will apply what you've learned in the videos and during lectures to complete data analysis tasks. You may discuss lab assignments with other students; however, the lab should be completed and submitted individually. Lab assignments must be typed up using Quarto, all work must be pushed to your GitHub repository for the lab, and the lab's PDF output must be submitted on Gradescope by the deadline.

Labs are due at 11:59pm ET on the indicated due date. Generally, for Monday labs, this is the following Wednesday; for Thursday labs, this is the following Sunday.

The lowest lab grade will be dropped at the end of the semester.

Exams

This course will have two exams: a midterm and a final. The midterm will include an in-class component (with a cheat sheet) and an open-note take-home component, while the final will include only an in-class component (with a cheat sheet).

You can demonstrate what you've learned in the course thus far through these exams. The exams will focus on both conceptual understanding of the content and application through analysis and computational tasks. The exam's content will be related to the content in videos and reading assignments, lectures, application exercises, and labs.

More details about the exams will be given during the semester.

Project

The project aims to apply what you've learned throughout the semester to analyze an interesting data-driven research question. The project will be completed in teams, and each team will present their work in the last lab session of the semester. The write-up will be due on the same day.

You cannot pass this course if you have not completed the project.

More information about the project will be provided during the semester.

Application exercises

Parts of some lectures will be dedicated to working on Application Exercises (AEs). These exercises allow you to apply the statistical concepts and code introduced in the preparation materials. These AEs are due by 10:45 am ET on the day of the lecture they are covered. To submit the AEs, you only need to push your work to your GitHub repo.

Because these AEs are for practice, they will be graded based on attempt, i.e., a good-faith effort has been made in attempting all parts.

Successful on-time completion of at least 80% of AEs will result in full credit for AEs in the final course grade.

Grading

The final course grade will be calculated as follows:

Category	Percentage
Labs	30%
Project	20%
Midterm	20%
Final	20%
Application Exercises	5%
Lab Attendance	5%

The final letter grade will be determined based on the following thresholds:

Letter Grade	Final Course Grade
A	>= 93
A-	90 - 92.99
B+	87 - 89.99

Letter Grade	Final Course Grade
B	83 - 86.99
B-	80 - 82.99
C+	77 - 79.99
C	73 - 76.99
C-	70 - 72.99
D+	67 - 69.99
D	63 - 66.99
D-	60 - 62.99
F	< 60

Course community

Inclusive community

It is my intent that students from all diverse backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that the students bring to this class be viewed as a resource, strength, and benefit. I intend to present materials and activities that respect diversity and align with Duke's Commitment to Diversity and Inclusion. Furthermore, I would like to create a learning environment for my students that supports a diversity of thoughts, perspectives, and experiences and honors their identities. To help accomplish this:

- If you feel like your performance in the class is being impacted by your experiences outside of class, please don't hesitate to come and talk with me. If you prefer to speak with someone outside of the course, your academic dean is an excellent resource.
- I (like many people) am still in the process of learning about diverse perspectives and identities. If anything was said in class (by anyone) that made you feel uncomfortable, please let me or a teaching team member know.

Personal pronouns

Pronouns are meaningful tools to communicate identities and experiences, and using pronouns supports a campus environment where all community members can thrive. You can update your gender pronouns in Duke Hub and learn more about personal pronouns at the Center for Sexual and Gender Diversity's website.

Accessibility

If any portion of the course is not accessible to you due to challenges with technology or the course format, please let me know so we can make appropriate accommodations.

The Student Disability Access Office (SDAO) is available to ensure that students can engage with their courses and related assignments. Students should contact the SDAO to request or update accommodations under these circumstances.

Communication

All lecture notes, assignment instructions, an up-to-date schedule, and other course materials may be found on the course website: sta199-su25.github.io.

Announcements will periodically be emailed through Canvas Announcements. Please check your email regularly to ensure you have the latest announcements for the course.

Email

If you have questions about assignment extensions, accommodations, or any other matter not appropriate for the class discussion forum, please email me directly at marie.neubrander@duke.edu. **If you do so, please include "STA 199" in the subject line.** Barring extenuating circumstances, I will respond to STA 199 emails within 48 hours, Monday through Friday. Response time may be slower for emails sent Friday evening through Sunday.

Five tips for success

Your success in this course depends very much on you and the effort you put into it. The course team will help you by providing you with materials answering questions, and setting a pace, but for this to work, you must do the following:

1. Complete all the preparation work before class.
2. Ask questions. As often as you can. In class, out of class. Ask me, ask your friends, ask the person sitting next to you. This will help you more than anything else. If you get a question wrong on an assessment, ask us why. If you're not sure about the lab, ask. If you hear something on the news that sounds related to what we discussed, ask. If the reading is confusing, ask.
3. Do the readings!

4. Do the lab! The earlier you start, the better. It's not enough to just mechanically plow through the exercises. You should ask yourself how these exercises relate to earlier material and imagine how they might be changed (to make questions for an exam, for example).
5. Don't procrastinate. The content builds upon what was taught in previous weeks, so if something is confusing to you in Week 2, Week 3 will become more confusing, Week 4 even worse, etc. Don't let the week end with unanswered questions. But if you find yourself falling behind and not knowing where to begin asking, come to office hours and work with a member of the teaching team to help you identify a good (re)starting point.

Getting help

- If you have a question during the lecture or lab, feel free to ask it! There are likely other students with the same question, so by asking, you will create a learning opportunity for everyone.
- The teaching team is here to help you be successful in the course. You are encouraged to attend office hours to ask questions about the course content and assignments. Many questions are most effectively answered as you discuss them with others, so office hours are a valuable resource. Please use them!
- Outside of class and office hours, any general questions about course content or assignments should be posted on the class discussion forum, Ed Discussion. There is a chance another student has already asked a similar question, so please check the other posts on the forum before adding a new question. If you know the answer to a question, I encourage you to respond!

Check out the Support tab for more resources.

Course policies

Duke Community Standard

As a student in this course, you have agreed to uphold the Duke Community Standard (video here!) and the practices specific to this course.

Academic honesty

TL;DR: Don't cheat!

Please abide by the following as you work on assignments in this course:

- **Collaboration:** Only work that is clearly assigned as teamwork should be completed collaboratively.
  - You may discuss lab assignments with other students; however, you may not directly share (or copy) code or write-up with other students. For team assignments, you may collaborate freely within your team. You may discuss the assignment with other teams; however, you may not directly share (or copy) code or write-up with another team. Unauthorized sharing (or copying) of the code or write-up will be considered a violation for all students involved.
  - You may not discuss or otherwise work with others on the exams. Unauthorized collaboration or using unauthorized materials will be considered a violation for all students involved. More details will be given closer to the exam date.
  - Collaboration within teams is not only allowed but expected for the project. Communication between teams at a high level is also allowed; however, you may not share code or project components across teams.
  - On individual assignments, you may not directly share work (including code) with another student in this class; on team assignments, you may not directly share work (including code) with another team.
- **Online resources:** I know that a huge volume of code is available online to solve any number of problems. Unless I explicitly tell you not to use something, the course's policy is that you may make use of any online resources (e.g., StackOverflow), but **you must explicitly cite where you obtained any code you directly use (or use as inspiration)**. Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism.
- **Use of generative artificial intelligence (AI)**: You should treat generative AI, such as ChatGPT, like other online resources. Two guiding principles govern how to use AI in this course:
  1. *Cognitive dimension:* Working with AI should not reduce your thinking ability. We will practice using AI to facilitate—rather than hinder—learning.
  2. *Ethical dimension*: Students using AI should be transparent about their use and ensure it aligns with academic integrity.
  - ✔ **AI tools for code:** You may use the technology for coding examples on assignments; if you do so, you must explicitly cite where you obtained the code. Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism. You may use these guidelines to cite AI-generated content. **The bare minimum citation must include the AI tool you're using (e.g., ChatGPT) and your prompt.** The prompt you use cannot be copied and pasted directly from the assignment; you must create a prompt yourself.

- 🚫 **AI tools for narrative:** Unless instructed otherwise, you may not use generative AI to generate a narrative that you then copy-paste verbatim into an assignment or edit and then insert into your assignment.
- ✔ **AI tools for learning:** You're welcomed to ask AI tools questions that might help your learning and understanding in this course.

In general, you may use generative AI as a resource as you complete assignments but not to answer the exercises for you. You are ultimately responsible for the work you turn in; it should reflect *your* understanding of the course content. Identifying AI-generated content is fairly straightforward. Any code identified as AI-generated but not cited as such and any narrative identified as AI-generated will be considered plagiarism and treated as such.

If you are unsure if using a particular resource complies with the academic honesty policy, please ask a teaching team member.

Regardless of course delivery format, it is the responsibility of all students to understand and follow all Duke policies, including academic integrity (e.g., completing one's own work, following proper citation of sources, adhering to guidance around group work projects, and more). Ignoring these requirements is a violation of the Duke Community Standard. Any questions and/or concerns regarding academic integrity can be directed to the Office of Student Conduct and Community Standards at conduct@duke.edu.

Any violations in academic honesty standards as outlined in the Duke Community Standard and those specific to this course will

- automatically results in a 0 for the relevant portion or the entire assignment,
- can further impact your overall course grade, and
- will be reported to the Office of Student Conduct & Community Standards for further action.

Late work & extensions

The due dates for assignments are there to help you keep up with the course material and to ensure the teaching team can provide feedback in a timely manner. We understand that things come up periodically that could make it difficult to submit an assignment by the deadline. Note that the lowest lab assignment will be dropped to accommodate such circumstances.

- Labs may be submitted up to 3 days late. A 5% deduction will be applied for each 24-hour period during which the assignment is late.
- No late work is accepted for application exercises since these are designed to help you prepare for other assessments in the course.
- No late work is accepted for exams.

- No late work is accepted for projects.

Waiver for extenuating circumstances

If circumstances prevent you from completing a lab by the stated due date, you may email the course coordinator, Dr. Mary Knox, before the deadline to waive the late penalty. In your email, you only need to request the waiver; you do not need to provide an explanation. This waiver may only be used **once** a semester, so only use it for a truly extenuating circumstance.

If circumstances have a longer-term impact on your academic performance, please let your academic dean know. They can be a resource. Please let me know if you need help contacting your academic dean.

Regrade requests

Regrade requests must be submitted on Gradescope within a week after an assignment is returned. Regrade requests will be considered if there was an error in the grade calculation or if a correct answer was mistakenly marked as incorrect. Requests to dispute the number of points deducted for an incorrect response will not be considered. Regrade requests are also not a mechanism for asking for clarification on feedback, those questions should be brought to office hours. Note that by submitting a regrade request, the **entire assignment may be regraded**, which could potentially result in losing points.

No grades will be changed after the final exam has been administered.

Attendance policy

Every student is expected to attend and participate in lecture and labs. There may be times, however, when you cannot attend class. If you miss a lecture, make sure to review the material and complete the application exercise, if applicable, before the next lecture. Labs are dedicated to completing the lab assignment and collaborating with your classmates and lab instructor. If you miss a lab session, make sure to communicate with your lab instructor about how you can make up your contribution. If you know you're going to miss a lab session and you're feeling well enough to do so, notify your lab TA and teammates ahead of time.

More details on Trinity attendance policies are available here.

Accommodations

Academic accommodations

If you need accommodations for this class, you will need to register with the Student Disability Access Office (SDAO) and provide them with documentation related to your needs. SDAO will work with you to determine what accommodations are appropriate for your situation. Please note that accommodations are not retroactive and disability accommodations cannot be provided until a Faculty Accommodation Letter has been given to me. Please contact SDAO for more information: sdao@duke.edu or access.duke.edu.

Religious accommodations

Students are permitted by university policy to be absent from class to observe a religious holiday. Accordingly, Trinity College of Arts & Sciences and the Pratt School of Engineering have established procedures to be followed by students for notifying their instructors of an absence necessitated by the observance of a religious holiday. Please submit requests for religious accommodations at the beginning of the semester so that we can work to make suitable arrangements well ahead of time. You can find the policy and relevant notification form here: trinity.duke.edu/undergraduate/academic-policies/religious-holidays