

# STA 199 - Introduction to Data Science and Statistical Thinking

Fall 2024

## 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](#).

## 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 community

### Duke Community Standard

As a student in this course, you have agreed to uphold the [Duke Community Standard](#) and the practices specific to this course.

## **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](#). Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups.

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-f24.github.io](https://sta199-f24.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.

## Where to get 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 [ed, the course discussion forum](#). 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 [Help](#) tab for more resources.

## Email

If you have a question that is not appropriate for the public forum, you are welcome to email me directly. 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.

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

You are expected to bring a laptop, tablet, 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.

## **Teams**

You will be assigned to a team early on in the semester. You are encouraged to sit with your teammates in lectures, and you will also work with them in some of the lab sessions. All team members are expected to contribute equally to the completion of the project, and you will be asked to evaluate your team members throughout the semester. Failure to adequately contribute to any project component will result in a penalty to your mark relative to the team's overall mark.

You are expected to use the provided GitHub repository as the central collaborative platform. Commits to this repository will be used as a metric (one of several) of each team member's relative contribution to each project.

## **Activities & Assessment**

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

### **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 2 pm 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 completion, i.e., a good-faith effort has been made in attempting all parts.

*Successful on-time completion of at least 70% of AEs will result in full credit for AEs in the final course grade.*

## 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 8:30 am ET on the indicated due date (generally the Monday after the lab).

*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.

## Grading

The final course grade will be calculated as follows:

Category	Percentage
Labs	35%
Project	20%

Category	Percentage
Midterm	20%
Final	20%
Application Exercises	5%

While there are no specific points allocated to attendance, we will record your attendance periodically throughout the semester. This information will be used as “extra credit” if you’re in between two grades and a minor bump would help.

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

Letter Grade	Final Course Grade
A	$\geq 93$
A-	90 - 92.99
B+	87 - 89.99
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	$< 60$

### Five tips for success

Your success in this course depends very much on you and the effort you put into it. The course has been organized so that the burden of learning is on you. Your TAs and I 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 the TAs, 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 weekend 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.

## Course policies

### 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 am well aware that a huge volume of code is available on the web 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.

- < fa-check > **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.
- < fa-ban > **AI tools for narrative:** Unless instructed otherwise, you may not use generative AI to write a narrative on assignments.

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](mailto: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](#) 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.
- The late work policy for the project will be provided with the project instructions.

### **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 you feel a correct answer was mistakenly marked as incorrect. Requests to dispute the number of points deducted for an incorrect response will not be considered. Note that by submitting a regrade request, the entire question will be graded, which could potentially result in losing points.

*No grades will be changed after the project presentations.*


## Attendance policy

- **COVID-19 Symptoms, Exposure, or Infection:** Student health, safety, and well-being are the university's top priorities. To help ensure your well-being and the well-being of those around you, **please do not come to class if you have tested positive for COVID-19 or have possible symptoms and have not yet been tested.** If any of these situations apply to you, you must follow university guidance related to the ongoing COVID-19 pandemic and current health and safety protocols. If you are experiencing any COVID-19 symptoms, [contact student health \(dshcheckin@duke.edu, 919-681-9355\)](mailto:dshcheckin@duke.edu). Learn more about current university policy related to COVID-19 at <https://coronavirus.duke.edu>. You will be expected to follow these guidelines to keep the university community as safe and healthy as possible. Please contact me and your academic dean as soon as possible if you need to quarantine or isolate so that we can discuss arrangements for your continued participation in class.
- **Inclement weather:** If inclement weather or other connectivity-related events prohibit class attendance, I will notify you how we will make up missed course content and work. Asynchronous catch-up methods may apply.
- **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 for students to notify 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: <https://trinity.duke.edu/undergraduate/academic-policies/religious-holidays>.

 Click here for syllabus bounty!

If you've read this far in the syllabus, post a picture of your pet if you have one or your favorite meme on ed in [this thread](#)! If you're willing, share the name of your pet too!

## Important dates

 Important

TO DO: Update dates

- **Jan 10:** Classes begin
- **Jan 24:** Drop/add ends
- **Feb 15:** Exam 1 in-class + take-home released

- **Feb 19:** Exam 1 take-home due
- **Feb 26:** Project milestone 1 – working collaboratively due
- **Mar 11-15:** Spring break
- **Mar 25:** Project milestone 2 – proposals due
- **Mar 27:** Last day to withdraw with W
- **Apr 11:** Exam 2 in-class + take-home released
- **Apr 15:** Exam 2 take-home due
- **Apr 15:** Project milestone 3 – peer review due
- **Apr 22:** Project milestone 4 – presentations and write-up due (in lab)
- **Apr 24:** Classes end

HW deadlines are listed on the course schedule. Class ends on April 24, there is no final exam.

For more important dates, see the full [Duke Academic Calendar](#).