STA 199 - Introduction to Data Science and Statistical Thinking

Fall 2025

Table of contents

TL;DR **2**
- Course Overview . . . 2
- Required Materials . . . 2
- Assessment & Grading . . . 3
- Key Policies . . . 3
  - Academic Honesty . . . 3
  - Deadlines & Late Work . . . 3
  - Attendance . . . 4
- Success Tips . . . 4
- Important Dates . . . 4
- Getting Help . . . 4

Long version **4**
- Course learning objectives . . . 5
- Course materials . . . 5
  - Textbooks . . . 5
  - Computing . . . 5
- Course community . . . 5
  - Inclusive community . . . 5
  - Personal pronouns . . . 6
  - Accessibility . . . 6
  - Communication . . . 6
  - Email . . . 7
- Five tips for success . . . 7
- Getting help . . . 8
- Course components . . . 8
  - Lectures . . . 8

Labs . . . 9
Teams . . . 9
Activities & Assessment . . . 9
Grading . . . 11
Course policies . . . 12
Duke Community Standard . . . 12
Academic honesty . . . 12
Late work & extensions . . . 14
One-time late penalty waiver . . . 15
Regrade requests . . . 15
Attendance policy . . . 15
Lecture recording request . . . 15
Accommodations . . . 16
Academic accommodations . . . 16
Religious accommodations . . . 16
Important dates . . . 16

TL;DR

for quick reference...[1] but really, read the long version!

Course Overview

Learn to explore, visualize, and analyze data using R and RStudio through real-world problems. Covers data wrangling, visualization, modeling, statistical inference, and ethical considerations in data science, including responsible AI tool usage.

Required Materials

- **Textbooks** (free online): *R for Data Science, 2e* and *Introduction to Modern Statistics*
- **Computing**: Laptop required for all classes; R accessed through Duke containers, no installation needed
- **Platform**: Course website at sta199-f25.github.io – everything you need is there or linked from there!

[1] The TL;DR short version was generated by Claude Sonnet 4 on 2025-08-23 with the prompt "Summarize the following syllabus to a 1-pager" and providing the long version content for context.

Assessment & Grading

Component	Weight	Details
Lectures	5%	Attendance/participation via Wooclap (18/26 lectures minimum for full score)
Labs	8%	Team-based, completed in class (lowest 2 dropped)
Homework	12%	Individual (lowest grade dropped)
Project	15%	Team-based data analysis with write-up and presentation
Exam 1	20%	In-class + take-home components
Exam 2	20%	In-class + take-home components
Final	20%	In-class only

Key Policies

Academic Honesty

- **Individual work**: Homework and exams must be completed alone
- **Team collaboration**: Expected for labs and project
- **AI tools**: Allowed for code assistance with proper citation; prohibited for direct narrative generation
- **Online resources**: Permitted with explicit citation

Deadlines & Late Work

- **Homework**: Up to 3 days late (-5% per day)
- **Labs**: No late work (completed in class)
- **Exams/Project**: No extensions or make-ups
- **One-time late penalty waiver**: One-time waiver for homework late penalty, request before deadline

Attendance

- **Lectures**: Participation tracked via Wooclap
- **Labs**: Mandatory attendance (no make-ups possible)
- **Lecture recordings**: Available for excused absences only

Success Tips

1. **Prepare**: Watch videos and read before class
2. **Engage**: Attend all lectures and labs actively
3. **Ask questions**: Use office hours and Ed Discussion forum
4. **Start early**: Don't procrastinate on assignments
5. **Stay current**: Content builds progressively

Important Dates

- **Classes begin**: Aug 25
- **Exam 1**: Oct 2 (in-class) / Oct 4 (take-home due)
- **Fall break**: Oct 13-14
- **Project**: Nov 10 (presentation in lab + write-up due)
- **Exam 2**: Nov 20 (in-class) / Nov 22 (take-home due)
- **Classes end**: Dec 5
- **Final exam**: Dec 12

Getting Help

- **Office hours**: Regular TA and instructor availability
- **Ed Discussion**: Online forum for course questions
- **Email**: Include "STA 199" in subject line
- **Accommodations**: Contact SDAO for academic accommodations

Note: You cannot pass without completing the project. The course emphasizes hands-on learning, ethical data science practices, and effective communication of results.

Long version

that you should read carefully…

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 tidying, exploratory data analysis, data visualization, predictive and descriptive modeling, and statistical inference;
- work on problems and case studies inspired by and based on real-world questions and data;
- explore ethical considerations in data science, including issues of misrepresentation, privacy, and bias;
- responsibly leverage AI tools within data science workflows while critically assessing the validity and potential biases in AI-generated insights;
- 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 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. 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

Using pronouns can help foster a respectful campus environment where all community members can thrive. Sharing pronouns is always optional for members of the Duke community. If you would like to share yours, you can update them in DukeHub. You can learn more at the DukeHub & Zoom Tutorials.

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-f25.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 mc301@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 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. **Watch the videos and do the readings before each class.**

Come prepared so you can deeply engage with the material during lectures and labs instead of using class time to learn the basics.

2. **Be present and engaged in every lecture and lab.**

In lectures do the application exercises, ask questions, and participate in discussions. In labs, work on the lab exercises, ask questions, and collaborate with your teammates. If you miss a class, make sure to catch up on the material before the next class.

3. **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.

4. **Do the homework.**

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 components

Lectures

Lectures 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, most lectures will feature application exercises that will serve as opportunities to practice what you're learning as you're learning it and be great preparation for the assignments and exams. You are expected to prepare for these by completing assigned readings and videos.

You are expected to bring a laptop (or Chromebook) to each class so that you can participate in the application activities. 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.

Labs

Labs are designed to be hands-on at all times, therefore you're similarly expected to bring a laptop to each lab session.

During labs you will get a brief introduction to the week's assignments from your TA and then work on the lab exercise with your teammates. You will often also have time to start working on your homework assignment during lab, once you complete the lab exercise.

Teams

You will start off the semester in randomly assigned teams for each lab session for the first few weeks. Then, you will be assigned to a project team that you will work with for the remainder of the semester.

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 seven components: lecture, lab, homework, project, two exams, a final, and attendance and participation.

Lectures

Attendance and participation will be tracked through in-class questions via Wooclap, which you can access on your laptop or phone. You will earn points for answering questions during lecture, and these points will contribute to your attendance and participation grade.

There are 26 lectures during the semester. You must participate in Wooclap questions in at least 18 lectures to get full credit on this component. Otherwise your grade on this component will be calculated as the percentage of lectures you attended and participated, e.g., if you attend and participate in 22/26 lectures, you get the full 5% but if you attend and participate in 15/26 lectures, you get (15/26)$5\% = 2.88\%$.

Application exercises won't be graded directly, but we will track activity on them to ensure you're staying engaged with the course and exams will feature topics and questions from these exercises.

Labs

In the first few weeks of the semester, you will be randomly assigned to a team and work with your teammates on an exercise that is due at the end of the lab session. Once project teams are formed, you will work with that team on the lab exercises. You must be present in lab to complete the lab assignment. There is no way to make up for missing a lab.

You will submit your lab assignments by pushing your work to your GitHub repository for the lab and submitting the PDF output on Gradescope *by the end of your lab session.*

There are eight graded labs in the semester. The lowest two lab grades will be dropped at the end of the semester, which means you can miss up to two lab sessions with no penalty.

Homework

You will complete homework assignments individually. You may start working on your homework assignment during your lab session, once you complete your lab exercise, but you will need to finish it outside of class. Your homework will be comprised of some practice exercises where you can get immediate feedback from AI tools designed for this course and some graded exercises.

You will submit your homework assignments by pushing your work to your GitHub repository for the homework and submitting the PDF output on Gradescope *by 11:59pm on Sundays.*

There are six graded homework assignments in the semester. The lowest homework grade will be dropped at the end of the semester, which means you can miss one homework assignment with no penalty.

Exams

This course will have three exams: two exams and a final. The exams 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.

You will submit your take-home exam by pushing your work to your GitHub repository for the exam and submitting the PDF output on Gradescope *by the stated deadline on the schedule.*

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 during a lab session in the latter part 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
Lectures	5%
Labs	8%
HW	12%
Project	15%
Exam 1	20%
Exam 2	20%
Final	20%

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
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 policies

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.

Academic honesty

TL;DR: Don't cheat!

What is allowed and what is not?

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. 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. "Sharing" includes, but is not limited to, messaging, emailing, or otherwise providing your work to another student or team.
  - Labs: Collaboration in teams for lab assignments is not only allowed but expected. You will work together with your lab team to complete the lab exercise. However, each student must submit their own write-up of the lab exercise, which should reflect their understanding and ideas. It's expected that lab submission will be similar across team members.
  - Project: Similarly, collaboration within teams is not only allowed but expected for the project. The difference is that the project is a single, collaborative work developed by the entire team. It is the team's responsibility to make sure all components of the project are vetted and revised by all team members, even if some division of labor takes place for first drafts of components. Communication between teams at a high level is also allowed; however, you may not share code or project components across teams.
  - Homework: You may discuss homework assignments with other students; however, you may not directly share (or copy) code or write-up with other students. For homework assignments, sharing (or copying) of the code or write-up will be considered a violation for all students involved, regardless of who initiated the sharing.
  - Exams: You may not discuss or otherwise work with others on the exams (in class and take home). On exams, collaboration, sharing (or copying) of the code, or using unauthorized materials will be considered a violation for all students involved, regardless of who initiated the sharing.

- **Use of 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, resulting in an automatic 0 for the relevant portion of the assignment.**
- **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 generative AI tools when you need help with assignments. However you should first attempt to solve the problem yourself. Your submission should not be a copy-paste of AI-generated content – you must edit the content to ensure it reflects your understanding, has your voice and intellectual input, and conforms with course materials, syntax, terminology, and style. Additionally, **the prompt you use cannot be copied and pasted directly from the assignment; you must create a prompt yourself.**

  You must also explicitly cite work submitted that is based on AI-generated content. 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), the model the tool uses, the date when you ran the prompt, and a link to the full transcript of the session starting with your prompt.** A new citation must be included for each exercise where you used AI tools, with a new link to the full transceript of the interaction.

  – ⊘ **AI tools for narrative:** Unless instructed otherwise, you may not use AI tools 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. However you should be critical of the answers you receive, as AI-generated content may not always be accurate or reliable. Use it to supplement your understanding, not as a substitute for learning.

  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 treated as plagiarism, resulting in an automatic 0 for the relevant portion of the assignment.**

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

What happens if you violate the academic honesty policy?

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

- will automatically result in a 0 for the relevant portion or the entire assignment or assessment,
- can result in further deductions to your overall course grade (e.g., drop down to the next letter grade or drop down to an F), and
- can be reported to the Office of Student Conduct & Community Standards for further action.

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.

Late work & extensions

The due dates for homework 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 homework assignment and lowest 2 lab assignments will be dropped to accommodate such circumstances.

- Homework assignments 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 labs since these are designed to be completed in class.
- No late work is accepted for exams.
- No late work is accepted for projects.

One-time late penalty waiver

If circumstances prevent you from completing a homework by the stated due date, you may email the course coordinator, Dr. Mary Knox, **before the deadline** to request 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** in the semester, so only use it wisely. The waiver may only be used for homework assignments, not for attendance/participation, labs, exams, or the project.

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

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 and a portion of your grade depends on this. There may be times, however, when you cannot attend class. Lecture recordings are available upon request for students who have an excused absence. See the Lecture recording request policy for more detail. 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 lab team and there is no way to make up for missing a lab.

More details on Trinity attendance policies are available here.

Lecture recording request

Lectures will be recorded on Panopto and will be made available to students with an excused absence upon request. Videos shared with such students will be available for a week after the lecture date. To request a particular lecture's video, please fill out the form at the link

below. Please submit the form within 24 hours of missing lecture to ensure you have sufficient time to watch the recording. Please also make sure that any official documentation, such as incapacitation forms, Dean's excuses, NOVAPs, and quarantine/removal from class notices from student health are also uploaded to the form.

https://forms.cloud.microsoft/r/jwFuhygVw5

About one week before each exam, the class recordings will be available to all students. These recordings will be available until the start of the exam.

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

Important dates

- **Aug 25:** Classes begin
- **Sep 1:** Labor Day
- **Sep 5:** Drop/add ends
- **Oct 2:** Exam 1 in-class + take-home released
- **Oct 4:** Exam 1 take-home due
- **Oct 13-14**: Fall break

- **Nov 7**: Last day to withdraw with W
- **Nov 10**: Project presentation in lab + write-up due
- **Nov 20:** Exam 2 in-class + take-home released
- **Nov 22:** Exam 2 take-home due
- **Nov 26-28**: Thanksgiving break
- **Dec 5**: Classes end
- **Dec 12**: Final exam

Lab and project deadlines are listed on the course schedule.

For more important dates, see the full Duke Academic Calendar.