



STA 199: Intro to Data Science



Intro to data science and statistical thinking. Learn to explore, visualize, and analyze data to understand natural phenomena, investigate patterns, model outcomes, and make predictions, and do so in a reproducible and shareable manner. Gain experience in data wrangling and munging, exploratory data analysis, predictive modeling, data visualization, and effectively communicating results. Work on problems and case studies inspired by and based on real-world questions and data. The course will focus on the R statistical computing language.



Course info



Lectures

 Soc Sci 139  Tue and Thu 10:05a - 11:20a








Labs

Lab 01  Old Chem 003  Fri 10:05a - 11:20p

Lab 02  Soc Psych 127  Fri 11:45a - 1p

Lab 03  Old Chem 003  Fri 1:25p - 2:40p

Teaching team and office hours

Instructor	Prof. Maria Tackett (http://stat.duke.edu/~mt324/)	 (mailto:maria.tackett@duke.edu)  (https://github.com/matackett)	Thu 1p - 2:30p	Old Chem 118B
TAs	Salvador Arellano (https://www.linkedin.com/in/salvador-chavero-arellano-405969168/)	 (mailto:salvador.chavero.arellano@duke.edu)  (https://github.com/salvadorchavero)	Wed 5p - 7p	Old Chem 203B
	Max Bartlett (http://maxbartlett.com/)	 (mailto:maxwell.bartlett@duke.edu)  (https://github.com/MaxBartlett)	Tue 6:30p - 8:30p	Old Chem 203B
	Meredith Brown (https://www.linkedin.com/in/meredith-brown-807964172/)	 (mailto:meredith.brown@duke.edu)  (https://github.com/meredithb3)	Mon 12:30p - 2:30p	Old Chem 203B
	Steven Herrera (https://www.linkedin.com/in/rosvidstevenherrera/)	 (mailto:rosvid.herrera.tenorio@duke.edu)  (https://github.com/stevenherrera24)	Thu 5:30p - 7:30p	Old Chem 203B
	Malavi Ravindran (https://www.linkedin.com/in/malavi-ravindran-332035175)	 (mailto:malavi.ravindran@duke.edu)  (https://github.com/MalaviRavindran)	Mon 3p - 5p	Old Chem 203B
	Becky Tang (https://beckytang.rbind.io/)	 (mailto:becky.tang@duke.edu)  (https://github.com/beckytang)	Wed 3p - 5p	Old Chem 203B

Texts

All books are freely available online. Hardcopies are also available for purchase.

R for Data Science (http://r4ds.had.co.nz/)	Grolemund, Wickham	O'Reilly, 1st edition, 2016
OpenIntro Statistics (https://www.openintro.org/stat/textbook.php)	Diez, Barr, Çetinkaya-Rundel	CreateSpace, 4th Edition, 2019
Introductory Statistics with Randomization and Simulation (https://www.openintro.org/stat/textbook.php?stat_book=isrs)	Diez, Barr, Çetinkaya-Rundel	CreateSpace, 1st Edition, 2014

Materials

You should bring a fully-charged laptop, tablet with keyboard, or comparable device to every lecture and lab session.



This course has achieved Duke's Green Classroom Certification. The certification indicates that the faculty member teaching this course has taken significant steps to green the delivery of this course. Your faculty member has completed a checklist indicating their common practices in areas of this course that have an environmental impact, such as paper and energy consumption. Some common practices implemented by faculty to reduce the environmental impact of their course include allowing electronic submission of assignments, providing online readings and turning off lights and electronics in the classroom when they are not in use. The eco-friendly aspects of course delivery may vary by faculty, by course and throughout the semester. Learn more at <https://sustainability.duke.edu/action/certification/classroom> (<https://sustainability.duke.edu/action/certification/classroom>).

 DUKE STATSCI ([HTTP://STAT.DUKE.EDU/](http://stat.duke.edu/))

 RSTUDIO ([HTTPS://RSTUDIO.CLOUD/](https://rstudio.cloud/))

 GITHUB ([HTTPS://GITHUB.COM/STA199-FA19](https://github.com/STA199-FA19)) (./)



Policies

 (policies.pdf) Click the icon to download a PDF copy of the course policies.

Course Learning Objectives

- Learn to explore, visualize, and analyze data in a reproducible and shareable manner
- 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 effectively communicate results through written assignments and final project presentation

Activities & Assessments

The following activities and assessments will help you successfully achieve the course learning objectives. By experiencing the course content in different ways, you will not only gain a better understanding of data science, but you will also get experiences that can guide you as you apply what you've learned in future academic and professional projects.

Homework

In homework, you will apply what you've learned during lecture and lab to complete data analysis tasks. You may discuss homework assignments with other students; however, homework should be completed and submitted individually. Homework must be typed up using R Markdown and submitted in the appropriate GitHub repository.

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

Labs

In labs, you will apply the concepts discussed in lecture to various data analysis scenarios, with a focus on the computation. You will work on lab assignments in teams, and all team members are expected to contribute equally to the completion of each assignment. You are expected to use the team's Git repository on the course's GitHub page as the central platform for collaboration. Commits to this repository will be used as a metric of each team member's relative contribution for each lab. You will also be asked to evaluate your team members' performance periodically during the semester.

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

Exams

The exams are an opportunity to assess the knowledge and skills you've learned. Both exams will be take-home assignments that you are expected to complete individually. Each exam will include small analysis and computational tasks related to the content discussed in lectures, application exercises, homework assignments, and labs. More details about the content and structure of the exams will be discussed during the semester.

Final Project

The purpose of the project is to apply what you've learned throughout the semester to analyze an interesting data-based research question. The project will be completed in teams, and each team will present their results during the final exam period, Saturday, December 14, 7p - 10p. *You must complete the final project and present your work in class to pass the course.*

Participation & Application Exercises

Application exercises give you an opportunity to practice using the statistical concepts and/or code discussed in lecture on short data analyses. They will typically be started during class and may be assigned to be completed by the next class meeting. In general, these assignments are shorter than homework assignments and will be completed in teams.

Writing Exercises

There will be several writing exercises throughout the semester. Each exercise will include a writing prompt related to the content discussed in class along with peer review. More details will be provided before the first exercise.

Teamwork

To construct functional and diverse teams, you will be asked to complete a short survey to gauge your previous exposure to programming topics. After completing the survey, you will be assigned to teams of 3-4 students; these teams will stay consistent throughout the semester (barring extraordinary circumstances). You will work in these teams during class, on application exercises, on labs, and on the final project.

Grade Calculation

The final grade will be calculated as follows:

Homework	20%
Labs	15%
Exam 1	17.5%
Exam 2	17.5%
Final Project	15%
Participation & Application Exercises	5%
Writing Exercises	5%
Teamwork	5%

Class attendance in lecture and lab is a firm expectation; frequent absences or tardiness will be considered a legitimate cause for grade reduction.

If you have a cumulative numerical average of 90 - 100, you are guaranteed at least an A-, 80 - 89 at least a B-, and 70 - 79 at least a C-. The exact ranges for letter grades will be determined after Exam 2.

Inclusion

In this course, we will strive to create a learning environment that is welcoming to all students and that is in alignment with Duke's Commitment to Diversity and Inclusion (<https://provost.duke.edu/initiatives/commitment-to-diversity-and-inclusion>). If there is any aspect of the class that is not welcoming or accessible to you, please let me know immediately. Additionally, if you are experiencing something outside of class that is affecting your performance in the course, please feel free to talk with me and/or your academic dean.

Accessibility

Duke University is committed to providing equal access to students with documented disabilities. Students with disabilities may contact the Student Disability Access Office (SDAO) (<https://access.duke.edu>) to ensure your access to this course and to the program. There you can

engage in a confidential conversation about the process for requesting reasonable accommodations both in the classroom and in clinical settings. Students are encouraged to register with the SDAO as soon as they begin the program. Please note that accommodations are not provided retroactively. More information can be found online at access.duke.edu (<https://access.duke.edu>) or by contacting SDAO at 919-668-1267, SDAO@duke.edu (<mailto:sdao@duke.edu>).

Where to find help

- If you have a question during 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 any of the office hours posted on the home page to ask questions as you study the course content and work through assignments. A lot of questions are most effectively answered in-person, 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 Piazza since there are likely other students with the same questions. The questions you post will be visible to the entire class, so please email Professor Tackett directly with any specific questions about grades or personal matters.

Sometimes you may need help with the class that is beyond what can be provided by the teaching team. In that instance, I encourage you to visit the Academic Resource Center.

The Academic Resource Center (ARC) (<https://arc.duke.edu>) offers free services to all students during their undergraduate careers at Duke. Services include Learning Consultations, Peer Tutoring and Study Groups, ADHD/LD Coaching, Outreach Workshops, and more. Because learning is a process unique to every individual, they work with each student to discover and develop their own academic strategy for success at Duke. Contact the ARC to schedule an appointment. Undergraduates in any year, studying any discipline can benefit! Contact ARC@duke.edu (<mailto:arc@duke.edu>), 919-684-5917, 211 Academic Advising Center Building, East Campus – behind Marketplace.

Where to find course materials

All assignments and course materials may be found on the course website and GitHub (www.github.com/sta199-fa19) page. There is also an up-to-date course schedule ([./schedule.html](https://www.github.com/sta199-fa19/schedule.html)) where you can find the lecture notes discussed in each class meeting, assignment deadlines, and reading assignments to help you prepare for each class.

Announcements may also be sent to the class by email, so please check your email regularly.

Make-up Policy

Students who miss a class due to a scheduled varsity trip, religious holiday, or short-term illness should fill out an online NOVAP (<https://trinity.duke.edu/undergraduate/academic-policies/athletic-varsity-participation>), Religious Observance Notification (<https://trinity.duke.edu/undergraduate/academic-policies/religious-holidays>), or Incapacitation Form (<https://trinity.duke.edu/undergraduate/academic-policies/illness>), respectively. These excused absences do not excuse you from assigned homework. It will still be your responsibility to submit relevant assignments in accordance with the deadline.

If you have a personal or family emergency or health condition that affects your ability to participate in class, you should contact your academic dean's office. More information about this procedure may be found on the Personal Emergencies page (<https://trinity.duke.edu/undergraduate/academic-policies/personal-emergencies>) or provided by your academic dean.

Exam dates cannot be changed and no make-up exams will be given. If you must miss an exam, your absence must be officially excused before the exam due date. If your absence is excused, the missing exam grade will be imputed at the end of the semester based on your performance on other relevant course assignments.

The final project presentations will be during the university scheduled exam period, Saturday, December 14, 7p - 10p. **You must complete the final project and present your work during the exam period in order to pass the course.**

Late Work

Homework or lab assignments submitted late but within 24 hours of the deadline may be accepted with a 20% penalty. Homework or lab assignments submitted any later will not be accepted.

Late work will not be accepted for the exams or the final project.

Academic Honesty

By enrolling in this course, you have agreed to abide by and uphold the provisions of the Duke Community Standard (<https://studentaffairs.duke.edu/conduct/about-us/duke-community-standard>) as well as the policies specific to this course. Any violations will automatically result in a grade of 0 on the assignment and will be reported to Office of Student Conduct (<https://studentaffairs.duke.edu/conduct>) for further action.

- 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.
- **Reusing code:** Unless explicitly stated otherwise, you may make use of online resources (e.g. StackOverflow) for coding examples on assignments. If you directly use code from an outside source (or use it as inspiration), you must or explicitly cite where you obtained the code. Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism.

On individual assignments, you may not directly share code or write up with other students. On team assignments, you may not directly share code or write up with another team. Unauthorized sharing of the code or write up will be considered a violation for all students involved.

Regrade Requests

Regrade requests should be submitted through the regrade request form on Gradescope. **Requests for a regrade must be made within a week of when the assignment is returned.** Due to the time consuming nature of regrades, requests submitted later will not be regraded. Requests will be honored if there is an error in the grade calculation or a correct answer was mistakenly marked as incorrect. Please note that by submitting a regrade request, your entire assignment may be regraded and you may potentially lose points. Therefore, you should attend office hours to ask a member of the teaching team about your grading feedback before submitting a regrade request.

No grades will be changed after the final project presentations.

Note: Grades can only be changed by the Professor Tackett. Teaching Assistants cannot change grades on returned assignments.

Technology

Cell phones and other electronic devices should be turned off or put on silent during class. When you use a laptop (or if you use a tablet), please ensure that the volume is set to mute and the device is only used for class purposes. In general, you should focus on the class discussion/activity at hand and refrain from engaging in other work or outside activities.



DUKE STATSCI ([HTTP://STAT.DUKE.EDU/](http://stat.duke.edu/))



RSTUDIO ([HTTPS://RSTUDIO.CLOUD/](https://rstudio.cloud/))



GITHUB ([HTTPS://GITHUB.COM/STA199-FA19](https://github.com/STA199-FA19))

(./)

