Introduction to Data Science

Fall 2021

Course:	STA199-001
Professor:	Alexander Fisher
Contact:	alexander.fisher@duke.edu
Lecture:	TuTh 3:30-4:45pm, Social Sciences 136
Lab:	01L - Fr 10:15-11:30am, Perkins Link 087
	02L - Fr 12:00-1:15pm, Languages 211
	03L - Fr 1:45-3:00pm, Perkins Link 087
Websites:	sakai.duke.edu – central site
	sta199-f21-001.netlify.app/ – hub for course info
	introds.org – video lectures and resources
Textbooks:	R for Data Science by Grolemund & Wickham (1st ed. O'Reilly)
	Introduction to Modern Statistics by Çetinkaya-Rundel & Hardin (1st ed. OpenIntro)

Course objectives: By the end of this semester students will be able to:

1. explore, visualize and analyze data in a reproducible manner

2. gain experience in data wrangling, munging, exploratory data analysis, predictive modeling and data visualization

3. work on problems and case studies based on real-world questions and data

4. effectively communicate and present results

Grading policy:

Assignment	Brief description
Homeworks (25%)	Biweekly homeworks typed up in R Markdown on GitHub with final PDF submitted to Gradescope.
Labs (15%)	Weekly team assignments typed up in R Markdown on Github with final PDF submitted to gradescope.
Final project (15%)	Video and written presentation due during final exam period. Students will work with lab teams.
Exam 1 (17.5%)	Take home, open-notes
Exam 2 (17.5%)	Take home, open notes
Quizzes (5%)	Weekly quizzes on sakai due online, before class starts
Application excercises (2.5%)	In-class interactive notes, graded for completion
Teamwork (2.5%)	Consists of peer-feedback and meaningful commits

Letter grade	Numeric 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

Assignments (long description)

1. Homeworks:

For 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 GitHub and submitted as a PDF in Gradescope.

One homework assignment will be dedicated to a statistics experience. The statistics experience is an opportunity to engage with statistics and data science outside of the classroom through podcasts, books, seminars, data analysis competitions, and other activities. As you complete these experiences, the goal is to consider how the material you're learning in the course connects with society more broadly.

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

2. Labs:

In labs, you will apply the concepts discussed in lecture to various data analysis scenarios, with a focus on the computation. Most lab assignments will be completed 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, and there will be periodic peer

evaluation on the team collaboration. Lab assignments will be completed using R Markdown, correspond to an appropriate GitHub repository, and submitted for grading in Gradescope.

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

3. Final Project:

The purpose of the final project is to apply what you've learned throughout the semester to analyze an interesting data-driven research question. The project will be completed with your lab teams, and each team will present their work in video and in writing during the final exam period. More information about the project will be provided during the semester.

4. Exams:

There will be two, take-home, open-note exams. Through these exams you have the opportunity to demonstrate what you've learned in the course thus far. Each exam will include small analysis and computational tasks related to the content in the prepare, practice, and perform assignments. More details about the content and structure of the exams will be discussed during the semester.

5. Quizzes:

These are low-stakes quizzes that will be completed at the beginning of each unit (about once per week) as part of the preparation for lecture. These quizzes include multiple choice and fill-in-the-blank questions. Each quiz will be administered in Sakai and may be taken up to three times. The quiz is due by the beginning of lecture on the specified due date. An overall score of 80% on all quizzes will result in full credit for prep quizzes in the final course grade.

5. Application exercises:

A majority of the in-class lectures will be dedicated to working on Application Exercises (AEs). These exercises which give you an opportunity to practice apply the statistical concepts and code introduced in the prepare assignment. These AEs are due within three days of the corresponding lecture period. Specifically, AEs from Tuesday lectures are due Friday by 11:59pm ET, and AEs from Thursday lectures are due Sunday by 11:59pm ET.

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 80% of AEs will result in full credit for AEs in the final course grade.

In addition to AEs, there will be periodic activities to help build a learning community. These will be short, fun activities that will help everyone in the class connect throughout the semester.

Course policies

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. It is my intent to present materials and activities that are respectful of diversity and in alignment 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.

Attendance policy related to COVID 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 symptoms related to COVID-19, have had a known exposure to COVID-19, or have tested positive for COVID-19. 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 at 919-681-9355. To keep the university community as safe and healthy as possible, you will be expected to follow these guidelines. Please reach out to 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.

Academic honesty: You should be familiar with Duke's community standard:

https://studentaffairs.duke.edu/conduct/about-us/duke-community-standard

By enrolling in this course, you commit to upholding Duke's community standard reproduced as follows:

- I will not lie, cheat, or steal in my academic endeavors;

- I will conduct myself honorably in all my endeavors; and

- I will act if the Standard is compromised.

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 assignment and will be reported to the Office of Student Conduct for further action.

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

- You may discuss individual homework and 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.

- Reusing code: Unless explicitly stated otherwise, you may make use of online resources (e.g. Stack-Overflow) for coding examples on assignments. If you directly use code from an outside source (or use it as inspiration), 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.

Late policy:

- Homework and labs may be submitted up to 3 days late. There will be a 5% deduction for each 24-hour period the assignment is late. After 3 days, work will no longer be accepted. Note, your lowest homework and lab will be dropped.

- There is no late work accepted for prep quizzes or application exercises, since these are designed to help you prepare for labs and homework.

- The late work policy for exams will be provided with the exam instructions.

- The late work policy for the project will be provided with the project instructions.

Logistics

Communication: All lecture notes, assignment instructions, an up-to-date schedule, and other course materials may be found on the course website, sta199-f21-001.netlify.app

Announcements will be emailed to the class through sakai. Please check your email regularly to ensure you have the latest announcements for the course. For quick communication with your peers and the teaching team, see the course slack for general questions and discussion.

Accessibility: If there is any portion of the course that 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 are able to engage with their courses and related assignments. Students should be in touch with the Student Disability Access Office to request or update accommodations under these circumstances.

Please note: This syllabus is a dynamic document that is subject to change. Any changes made will be in the interest of creating a manageable workload and being compassionate towards students.

Additional resources

Academic Resource Center: There are times when you may need help with the class beyond what can be provided by the teaching team. In those instances, I encourage you to visit the Academic Resource Center. The Academic Resource Center (ARC) 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, 919-684-5917.

CAPS: Duke Counseling & Pyschological Services (CAPS) helps Duke Students enhance strengths and develop abilities to successfully live, grow and learn in their personal and academic lives. CAPS recognizes that we are living in unprecedented times and that the changes, challenges and stressors brought on by the COVID-19 pandemic have impacted everyone, often in ways that are tax our well-being. CAPS offers many services to Duke undergraduate students, including brief individual and group counseling, couples counseling and more. CAPS staff also provides outreach to student groups, particularly programs supportive of at-risk populations, on a wide range of issues impacting them in various aspects of campus life. CAPS provides services to students via Telehealth. To initiate services, you can contact their front desk at 919-660-1000.

Important dates:
Aug 23: Classes begin
Sep 3: Drop/add ends
Oct 4-5: Fall break
Nov 5: Last day to withdraw with W
Nov 24-26: Thanksgiving recess
Dec 3: Classes end
Dec 4-7: Reading period
Dec 8-13: Final exams

See https://registrar.duke.edu/fall-2021-academic-calendar for the full Duke academic calendar.