Syllabus

Course description

STA 313 - Advanced Data Visualization is all about the art and science of visualizing data. Three themes (what, why, and how) will run alongside each other as we cycle through the course topics. In "what" we focus on specific types of visualizations for a particular purpose (e.g. maps for spatial data, Sankey diagrams for proportions, etc.) as well as the tooling to produce them (e.g. specific R packages). In "how" we focus on the process – each visualization starts with a design (which we'll often ask you to do with a rough sketch accompanied by pseudo code), then often needs pre-processing of the data (wrangling, reshaping, joining, etc. to get it into a tidy, rectangular format for visualization), then attributes of the data are mapped to plot aesthetics, then the creator of the visualization needs to make a series of strategic decisions about visual encoding (e.g. accessibility concerns), and finally creating effective visualizations requires post-processing for visual appeal as well as annotation. In "why" we discuss the theory that ties the "how" and the "what" together, often focusing on the grammar of graphics. Like any data analysis, data visualization is also an iterative process. We don't expect you to land on the perfect visualization on the first try, so we promote the iterative process via critical and constructive review of one's own and each others' work. Independent modules will also touch on topics such as using statistical graphics for visual inference, creating data-based art, and a review of the literature on non-visual approaches to representing data.

The course will focus on the use of the R statistical programming language and introduce you to a variety of modern data visualization packages in R. In addition, you will continue to use hone their data science workflow skills that they acquired in pre-requisite courses by working with Git and GitHub for version control and collaboration.

Prerequisites

This course assumes that this is not your first interaction with working with data in R, using RStudio, and along with version control with Git, and collaboration on GitHub. Any of the following courses meet the prerequisite for the course: STA 198, STA 199, or STA 210. The course will start with a quick review of the relevant technologies.

Learning goals

- Understand the principles of designing and creating effective data visualizations.
- Evaluate, critique, and improve upon one's own and others' data visualizations based on how good a job the visualization does for communicating a message clearly and correctly.
- Post-process and refine plots for effective communication.
- Use visualizations for evaluating statistical models and for statistical inference.
- Master using R and a variety of modern data visualization packages to create data visualizations.
- Work reproducibly individually and collaboratively using Git and GitHub.

Textbooks

Readings for the course will come from the following textbooks. All of them are **freely available online** and you do not need to purchase a physical copy of either book to succeed in this class.

1. `[ggplot2-book]` Hadley Wickham, Danielle Navarro, and Thomas Lin Pedersen. ggplot2: Elegant Graphics for Data Analysis. (in progress) 3rd edition. Springer, 2023.
2. `[socviz]` Kieran Healy. Data Visualization: A Practical Introduction. Princeton University Press, 2018.
3. `[fdv]` Claus O. Wilke. Fundamentals of Data Visualization. O'Reilly Media, 2019.
4. `[r4ds]` Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund. R for Data Science. (in progress) 2nd edition. O'Reilly, 2022.

Course community

Duke Community Standard

All students must adhere to the Duke Community Standard (DCS): Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, and accountability. Citizens of this community commit to reflect upon these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity.

To uphold the Duke Community Standard, students agree:

- 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.

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.

Furthermore, I would like to create a learning environment for my students that supports a diversity of thoughts, perspectives and experiences, and honors your identities. To help accomplish this:

- If you have a name that differs from those that appear in your official Duke records, please let me know! You'll be able to note this in the Getting to know you survey.
- 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 something was said in class (by anyone) that made you feel uncomfortable, please let me or a member of the teaching team know.

Pronouns

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

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.

Communication

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

I will regularly send course announcements via email and Sakai, make sure to check one or the other of these regularly. If an announcement is sent Monday through Thursday, I will assume that you have read the announcement by the next day. If an announcement is sent on a Friday or over the weekend, I will assume that you have read it by Monday.

Where to get 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 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 course Slack. There is a chance another student has already asked a similar question, so please check the other posts on Slack before adding a new question. If you know the answer to a question posted on Slack, I encourage you to respond!

Check out the Support page for more resources.

I want to make sure that you learn everything you were hoping to learn from this class. If this requires flexibility, please don't hesitate to ask.

- You *never* owe me personal information about your health (mental or physical) but you're always welcome to talk to me. If I can't help, I likely know someone who can.
- I want you to learn lots of things from this class, but I primarily want you to stay healthy, balanced, and grounded during this crisis.

Lectures and lab

The goal of both the lectures and the labs is for them to be as interactive as possible. My role as instructor is to introduce you new tools and techniques, but it is up to you to take them and make use of 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 be working on a variety of tasks and activities throughout each lecture and lab. Attendance

will not be taken during class but you are expected to attend all lecture and lab sessions and meaningfully contribute to in-class exercises and discussion.

You are expected to bring a laptop to each class so that you can take part in the in-class exercises. Please make sure your laptop is fully charged before you come to class as the number of outlets in the classroom will not be sufficient to accommodate everyone. See [Technology accommodations] if you need a loaner laptop.

Assessment

Assessment for the course is comprised of four components: attendance and participation, reading quizzes, homework assignments, and projects.

- **Attendance and participation** is required throughout the semester. Students who attend at least 80% of the lectures and participate regularly during lecture will receive full credit for this portion of their grade. Participation in labs as well as on Slack will also count towards this component.
- **Reading quizzes**, due every other week (roughly), completed individually. Lowest quiz score is dropped.

  Reading quizzes will be linked from the course schedule. They always cover reading that is due since the previous quiz and up to and including the deadline for the given quiz. They're due by 10 am ET (beginning of class) on the indicated day on the course schedule.
- **Homework assignments**, due every other week (roughly), completed individually. Lowest homework assignment score is dropped.

  Homework assignments are due by 5 pm ET on the indicated day on the course schedule.
- **Projects** (2), mid-semester and end of semester, completed in teams.
  - Project 1: Teams will be given a dataset (or a set of datasets to choose from) to visualize. Project 1 is worth 15% of the course grade.
  - Project 2: Teams will choose the focus of their own project. One requirement is that you need to do something new. Project 2 is worth 25% of the course grade.

  The deliverables for each project will include a data visualization, a write up of the process and findings, and a presentation. For the second project, you will be encouraged to think beyond a traditional two-dimensional data visualization (e.g. interactive web apps/dashboards, data art, generative art, physical/tangible visualizations, **ggplot2** extensions, etc.).

  Each project will have a peer review component to provide at least one round of feedback during the process of development. Teams will provide periodic peer feedback to their

teammates while working on the projects as well as upon completion. The scores from the peer evaluations, along with individual contributions tracked by commits on GitHub, will be used to ensure that each student has contributed to the teamwork.

All team members must take part in the presentation. Presentations can be given in person in class, or via Zoom if the team prefers. My preference is that the team stick to one method of delivery (all presenters in person or all presenters on Zoom), but I realize a lot can change throughout this semester, and we'll adjust accordingly.

All work is expected to be submitted by the deadline and there are no make ups for any missed assessments. See the late work policy for more details.

Grading

The final course grade will be calculated as follows:

Category	Percentage
Attendance and participation	5%
Reading quizzes	10%
Homework assignments	45%
Project 1	15%
Project 2	25%

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

These are upper bounds for grade cutoffs, depending on the class performance the cutoffs may be lowered but they won't be increased.

Teams

You will be assigned to a different team for each of your two projects. You are encouraged to sit with your teammates in lecture and you will also work with them in the lab sessions. All team members are expected to contribute equally to the completion of each project and you will be asked to evaluate your team members after each assignment is due. Failure to adequately contribute to an assignment will result in a penalty to your mark relative to the team's overall mark.

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

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 team work should be completed collaboratively.
  - The homework assignments must also be completed individually and you are welcomed to discuss the assignment with classmates at a high level (e.g., discuss what's the best way for approaching a problem, what functions are useful for accomplishing a particular task, etc.). However you may not directly share answers to lab questions (including any code) with anyone other than myself and the teaching assistants.
  - The reading quizzes must be completed individually with absolutely no communication with classmates.
  - For the projects, collaboration within teams is not only allowed, but expected. Communication between teams at a high level is also allowed however you may not share code or components of the project across teams.
  - On individual assignments you may not directly share code with another student in this class, and on team assignments you may not directly share code with another team in this class.

- **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, the same as other online resources. There are two guiding principles that govern how you can use AI in this course:[1] (1) *Cognitive dimension:* Working with AI should not reduce your ability to think clearly. We will practice using AI to facilitate—rather than hinder—learning. (2) *Ethical dimension*: Students using AI should be transparent about their use and make sure it aligns with academic integrity.
  - **AI tools for code:** You may make use of 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 for citing AI-generated content.
  - **AI tools for narrative:** Unless instructed otherwise, you may not use generative AI to write 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.

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

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 result in a 0 for the 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 within a timely manner. We understand that things come up periodically that could make it difficult to submit an assignment by the deadline.

Policy on late work depends on the particular course component:

- Reading quizzes: Late quizzes are not accepted and there are no make ups for missed quizzes.
- Homework assignments: GitHub repositories will be closed to contributions at the deadline. If you need to submit your work late, send a DM on Slack to the Head TA to reopen your repository.
  - Late, within 24 hours of deadline: -10% of available points
  - Late, within 24-48 hours of deadline: -20% of available points
  - More than 48 hours later: No credit, and we will not provide written feedback
- Projects: The following three components contribute to your project score.
  - Presentation: Late presentations are not accepted and there are no make ups for missed presentations.
  - Write up: GitHub repositories will be closed to contributions at the deadline. If you need to submit your work late, Slack/email me to reopen your repository.
    - Late, within 24 hours of deadline: -10% of available points
    - Late, within 24-48 hours of deadline: -20% of available points
    - More than 48 hours later: No credit, and we will not provide written feedback
  - Peer evaluation: Late peer evaluations are not accepted and there are no make ups for missed presentations. If you do not turn in your peer evaluation, you get 0 points for your own peer score as well, regardless of how your teammates have evaluated you.

Waiver for extenuating circumstances

If there are circumstances that prevent you from completing a lab or homework assignment by the stated due date, you may email or DM the Head TA before the deadline to waive the late penalty. In your email, you only need to request the waiver; you do not need to provide explanation. This waiver may only be used for once in the semester, so only use it for a truly extenuating circumstance.

If there are circumstances that are having a longer-term impact on your academic performance, please let your academic dean know, as they can be a resource. Please let me know if you need help contacting your academic dean.

Regrade requests

Regrade requests must be made within one week of when the assignment is returned, and must be typed up and submitted via email to the course Head TA. These will be considered if points were tallied incorrectly or if you feel your answer is correct but it was marked wrong. No regrade will be made to alter the number of points deducted for a mistake. There will be no grade changes after the second project presentations. Note that during the regrade process your score could go up or go down or not change.

No grades will be changed after the project presentations.

Attendance policy

Responsibility for class attendance rests with individual students. Since regular and punctual class attendance is expected, students must accept the consequences of failure to attend. More details on Trinity attendance policies are available here.

However, there may be many reasons why you cannot be in class on a given day, particularly with possible extra personal and academic stress and health concerns this semester. All course lectures will be recorded and available to enrolled students after class. If you miss a lecture, make sure to to review the lecture material before the next class session. Lab time is dedicated to working on your homework assignments and collaborating with your teammates on your project. If you miss a lab session, make sure to communicate with your team about how you can make up your contribution. Given the technologies we use in the course, this is straightforward to do asynchronously. If you know you're going to miss a lab session and you're feeling well enough to do so, notify your teammates ahead of time. Overall these policies are put in place to ensure communication between team members, respect for each others' time, and also to give you a safety net in the case of illness or other reasons that keep you away from attending class.

Note that attendance and participation is part of your grade as well.

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.

Inclement weather policy

In the event of inclement weather or other connectivity-related events that prohibit class attendance, I will notify you how we will make up missed course content and work. This might entail holding the class on Zoom synchronously or watching a recording of the class.

Policy on video recording course content

If you feel that you need record the lectures yourself, you must get permission from me ahead of time and these recordings should be used for personal study only, no for distribution. The full policy on recording of lectures falls under the Duke University Policy on Intellectual Property Rights, available at provost.duke.edu/sites/default/files/FHB_App_P.pdf. Unauthorized distribution is a cause for disciplinary action by the Judicial Board.

Accommodations

Academic accommodations

If you are a student with a disability and need accommodations for this class, it is your responsibility to register with the Student Disability Access Office (SDAO) and provide them with documentation of your disability. 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: https://trinity.duke.edu/undergraduate/academic-policies/religious-holidays.

Note: If you've read this far in the syllabus, post a picture of your pet and their name if you have one or your favorite meme to the #random channel on Slack!

Important dates

- **Wednesday, January 10:** Classes begin, Monday schedule (no class for us)
- **Wednesday, January 24:** Drop/add ends
- **Wednesday, February 28:** Project 1 presentations
- **Friday, February 23:** Mid-semester grades reported
- **Saturday - Sunday, March 9 - 17**: Spring Break
- **Wednesday, March 27:** Last day to withdraw with W
- **Wednesday, April 24**: Classes end
- **Thursday - Sunday, April 25 – 28**: Reading period
- **Wednesday, May 1, 9 am - 12 pm**: Project 2 presentations

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