STAT 4380

Data Science (3 credits)
Spring 2026
Section 001, CRN 35407
Mendel Hall G83
TTh 4 – 5:15 PM

Student hours – come talk to me!
Mondays 3 – 4:30PM (Zoom)
Tuesdays 1:15 – 2:15PM (SAC 370)
Thursdays 11AM – 12PM (SAC 370)

Instructor

Katie Fitzgerald, PhD Statistics
Assistant Professor of Statistics
kaitlyn.fitzgerald@villanova.edu
Office: SAC 370

Course Description

Data-savvy professionals are in high demand in businesses, public agencies, and nonprofits. The supply of professionals who can work effectively with data at scale is limited and is reflected by rapidly rising demand and salaries for data scientists. This course explores how real-world data are gathered, managed, and used for making decisions or predictions. Students will use R, one of the most popular software used in data science practice, to learn data wrangling and data visualization with applications from a variety of disciplines.

Prerequisite: Introductory statistics course or instructor permission

Course Objectives
After this course, you will be able to:
- Understand the role of data science and its role in society
- Read, explore, clean, wrangle, visualize and analyze data in a reproducible manner
- Discuss how choices about data collection, extraction and processing affect bias and generalizability
- Identify and describe ethical issues related to data access and use
- Ask good questions and produce meaningful analyses of real-world datasets
- Communicate data effectively

Tips for Success

- Dedicate yourself to being an engaged learner and contributing to a thoughtful learning environment for your peers.
- Collaborate! Get to know your classmates. This is a very hands-on class, and you will be working together extensively.
- Ask questions! Ask them in class, on Piazza, and/or in student hours.
- Google is a coders best friend! Answers & discussions on stackoverflow.com are often particularly helpful for R related questions. See end of syllabus for discussion of use of AI in this class.
- Embrace the struggle & don't shy away from confusion or uncertainty. After all, statistics is the "science of uncertainty," and being "good at math" is being good at being stuck…

How to access support for this class

- Our course website will be the central hub for accessing course materials, assignments, and announcements: https://nova-stat-4380.netlify.app. Blackboard will be used primarily for assignment submissions and grade posting.
- **Piazza** will be our preferred platform for questions about homework, code, and general course questions. The system is integrated with Blackboard and is highly catered to getting you help quickly and efficiently from classmates and the instructor. **Rather than emailing questions to the instructor, you should post your questions on Piazza**. Email should be reserved for personal questions such as those regarding absences or grades.
- Come to my student hours! I have over three hours *outside of class time* each week specifically dedicated to helping YOU succeed in this course. Even if you don't know what your specific questions or points of confusion are, we can figure that out together.:
    - Mondays, 3 – 4:30PM (Zoom)
    - Tuesdays, 1:15 – 2:15PM (SAC 370, Zoom by request)
    - Thursdays, 11AM – 12PM (SAC 370, Zoom by request)
- Contact me about any concerns. Best way to reach me is via email (kaitlyn.fitzgerald@villanova.edu). I do my best to respond within 24 hours Monday – Friday.

Required Course Materials

Books: This course utilizes three textbooks, all of which are available for free:

- Available as eBook through Villanova Library:
    1. Communicating with Data: The Art of Writing for Data Science by Deborah Nolan and Sara Stoudt. Oxford University Press, 2021. ISBN: 978-0-1988-6275-8
- Available as free online textbook:
    2. Modern Data Science with R by Benjamin S. Baumer, Daniel T. Kaplan, and Nicholas J. Horton
    3. R for Data Science by Hadley Wickham & Garrett Grolemund

Software: This course will utilize the statistical software R and RStudio. Students will receive instructions in the first week of class for how download it onto their personal computer.

Hardware: Students are expected to bring a laptop to all class sessions. If access to a laptop is an issue, then please contact the course instructor and an accommodation will be made. This requirement will not prevent students from taking this course.

Course design[1]

Modern data analysis is inherently tied to coding and computational tools, and coding is learned by doing. To facilitate this type of learning, this course utilizes a flipped learning design. Each week, you will watch a series of lecture videos *before* class on Tuesdays. Then class time will be dedicated to hands-on application exercises and lab assignments.

Lecture Recordings

At the beginning of each week, you will watch a series of lecture videos that introduce new data science techniques. You will watch the recordings in Perusall and contribute comments and/or questions to demonstrate engagement, due by class time on Tuesdays.

Application Exercises

The majority of class-time on Tuesdays will be dedicated to working on Application Exercises (AEs) in RStudio, designed to help you practice the new skills, code, and concepts introduced in that week's lecture videos. AEs are due at the end of class on Tuesdays and serve as your "exit ticket". AEs are graded on completion; demonstrating that a good faith effort has been made on the assignment will earn full credit. If you are unable to attend class on a Tuesday, you may still submit your AE for half credit. The two lowest AE scores will be dropped at the end of the semester to accommodate occasional absences.

Quizzes

Most weeks (10x per semester), one class period (usually Thursdays) will start with a brief quiz (closed notes). There are no make-up quizzes if you are late or absent, but the lowest two quiz grades will be dropped at the end of the semester.

Labs

The majority of class-time on Thursdays will be dedicated to lab assignments. The labs are a more in-depth application of the week's material that will have you complete scaffolded analyses of a real dataset using RStudio. You will collaborate in randomly assigned teams of 2-3, and teams will rotate approximately every 2 weeks. Labs are due by class-time the following Thursday. *The lowest lab grade will be dropped at the end of the semester.*

Project

The purpose of the final project is to apply what you've learned throughout the semester to investigate a real-world social good problem by analyzing a relevant dataset of your choosing. You will identify an issue you care about (e.g., mental health, food insecurity, animal adoption, mass incarceration) and find a dataset that allows you to better understand the problem and/or proposed solutions. The project will be completed in self-assigned teams of 2-3. You will be asked to present your findings in a written report and an oral presentation. See course website for details.

---

[1] Much of the course design, activities, and assessments are adapted from Mine Çetinkaya-Rundel's *Data Science in a Box* curriculum and Duke University's open source Introduction to Data Science course under the Creative Commons Attribution Share Alike 4.0 International.

Data Ethics Readings & Community Annotations

Each week, you will read an article or book excerpt that engages with issues related to data ethics and/or data communication. You will engage with the reading via the community annotation tool Perusall. Annotations are due by class-time on Thursdays.

Late/makeup work

Here is how deadlines work in the real world: they exist and they're important. However, there's a certain amount of flexibility with them. If you need a little longer on something, you communicate with whoever has set you the deadline and ask if you can have a few more days. This is usually not a big deal, but if it happens a lot, people will start asking you if everything is all right.

That is also how deadlines work in this class. You may communicate with me (via an extension request form, available on the course website) to ask for an extension on anything you need, and that's mostly fine. If you ask for lots of extensions, we'll work together to find ways to help you adjust and better keep up with the work in the course.

If there are life circumstances that are having a longer-term impact on your academic performance or well-being, come talk to me, and we can work towards a solution and connect you to the support you need.

Attendance & Engagement

Due to the hands-on and collaborative nature of this class, attendance is very important. Students are entitled to one excused absence for any reason that may contribute to their personal wellness. Students must advise the instructor by email before class of their intent to utilize a Personal Day as the reason for their absence. A Personal Day will not be approved retroactively. Students may, but are not required, to provide additional information regarding their absence. Additionally, a Personal Day may not:

- be used immediately preceding or following a University holiday or break period;
- be used on days when exams, presentations or other major assignments are scheduled.

A Personal Day does not grant an automatic extension for items due. Students remain responsible for all assignments, exams, presentations, etc. due on that date. It is in the instructor's discretion to determine whether any extension is appropriate given individual circumstances.

Grading

Lecture Recordings	5%
Application Exercises	5%
Readings & Annotations.	5%
Quizzes	25%
Labs	30%
Project	30%

Final letter grades will be assigned according to the scale below

A	93-100%	B+	87-89%	C+	77-79%	D+	67-69%	F	0-59%
A-	90-92%	B	83-86%	C	73-76%	D	63-66%		
		B-	80-82%	C-	70-72%	D-	60-62%		

Important Dates

January 13 (Tue)	First day our class meets
January 18 (Sun)	Add/Drop deadline
Feb 5 (Th)	Project Brainstorm Due
Feb 19 (Th)	NO CLASS (flex day)
Feb 21 (Sat)	Project Proposal Due
Mar 3, 5 (Tue, Th)	NO CLASS (Spring Break)
Mar 12 (Th)	Project Cleaning / EDA due
Mar 31 (Tue)	Project Preliminary Data Story due
April 2 (Th)	NO CLASS (Easter Break)
April 16 (Th)	Project presentation rough draft due
April 21 (Tu)	Project report rough draft due
April 28, 30 (Tue, Th)	Final Presentations
May 4 (Mon)	Final report due
May 11 (Mon)	Final Presentations (continued), 2:30 – 5pm

Course Community & Policies

Inclusive Community

Inclusive Community

It is my intent that this course models and fosters justice, equity, diversity, and inclusion. We will engage with these values both in content and in practice. Data and statistics can be tools to tell diverse stories and help us learn about the state of the world from a perspective beyond our own lived experience. When used responsibly and with integrity, they can amplify the experiences of vulnerable and historically excluded populations. For example, they can be used to shed light on disparities in our schools, healthcare system, and criminal justice system.

The research questions we ask, the data we collect, and the way we use that data are infused with (often hidden) values about who and what matters in the world. For example, we should examine if and when marginalized people and their experiences are being excluded from our data, particularly when that data is used in countless ways to drive decision-making and inform society about the state of the world.

You will be asked to continually and critically engage with these ideas with each dataset and analysis you encounter. You are expected to engage your peers and new perspectives with curiosity, empathy, and intellectual humility. It is my intent that all students be well-served by this course, that your learning needs are met inside and outside the classroom, and that the diversity that you bring to this class be valued and utilized as a resource and strength.

I (like many people) am continually learning how to honor diverse perspectives and identities. If something was said in class (by me or a peer) that made you feel uncomfortable, please let me know. You will also have the opportunity to express concerns anonymously via check-in surveys. Villanova also encourages community members to submit any campus climate concerns at the following website: https://www1.villanova.edu/university/diversity-inclusion/report-climate-concern.html

Villanova affirms that diversity, equity and inclusion are integral components of the teaching and learning experience and an essential element of the ongoing intellectual, social and spiritual development of every member of the Villanova community. We believe that an inclusive community fosters an understanding and appreciation for diversity among our students, faculty, staff and administrators. We are committed to cultivating an academic environment that is marked by genuine curiosity about different perspectives, ardent receptivity to knowledge generated through intercultural connections, and a genuine sensitivity to the variety of human experiences.

Academic Integrity Policy

TL;DR: Don't cheat! And use AI responsibly and within guidelines provided

See full AI / LLM policy at the end of this document. Additionally, 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-ups with other students.
- Citing code/solutions: Unless explicitly stated otherwise, you may make use of online resources for coding on assignments. **However, the work must be primarily your own and may not be completed, in whole or in substantial part, by other humans or chatbots, AI, etc.** 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.

All students are expected to uphold Villanova's Academic Integrity Policy and Code. Any incident of academic dishonesty will be reported to the Dean of the College of Liberal Arts and Sciences for disciplinary action. For the College's statement on Academic Integrity, you should consult the Student Guide to Policies and Procedures. You may view the University's Academic Integrity Policy and Code, as well as other useful information related to writing papers, at the Academic Integrity Gateway web site: https://library.villanova.edu/research/subject-guides/academicintegrity

Office of Disabilities (ODS) and Learning Support Services (LSS)

If there is any portion of this class that is not accessible to you due to course format or challenges with technology, please let me know so we can make appropriate accommodations.

It is the policy of Villanova to make reasonable academic accommodations for qualified individuals with disabilities. All students who need accommodations should go to Clockwork for Students via myNOVA to complete the Online Intake or to send accommodation letters to professors. Go to the LSS website http://learningsupportservices.villanova.edu or the ODS website https://www1.villanova.edu/university/student-life/ods.html for registration guidelines and instructions. If you have any questions please contact LSS at 610- 519-5176 or learning.support.services@villanova.edu, or ODS at 610-519-3209 or ods@villanova.edu.

Absences for Religious Holidays

Villanova University makes every reasonable effort to allow members of the community to observe their religious holidays, consistent with the University's obligations, responsibilities, and policies. Students who expect to miss a class or assignment due to the observance of a religious holiday should discuss the matter with their professors as soon as possible, normally at least two weeks in advance. Absence from classes or examinations for religious reasons does not relieve students from responsibility for any part of the course work required during the absence.
https://www1.villanova.edu/villanova/provost/resources/student/policies/religiousholidays.html

AI /LLM policy (i.e. usage of ChatGPT, Gemini, etc.)[2]

TLDR: you're responsible for understanding how to solve problems, cite any use of AI

In general, we treat AI-based assistance, such as ChatGPT, the same way we treat collaboration with other people; you are welcome to talk about your ideas with other people, both inside and outside the classroom, as well as with AI-based assistants.

However, **all work you submit must be primarily your own, and may not be completed, in whole or in substantial part, by other humans or chatbots, AI, etc.** You also **must properly acknowledge (cite) any ideas / code / solutions that did not originate from you.** In all cases, you are responsible for understanding all work that is turned in and may be periodically asked to orally explain your answers.

I expect you *will* use AI / LLMs periodically to assist you in this course. Responsible use of AI is not "against the rules" and you should not feel the need to hide it. **If/when you use AI while working on an assignment, you are expected to provide the following with your submission:**
- A statement acknowledging your use of AI and which tool you used
- A precise description of the prompt(s) you used on which problem(s)
- A brief reflection of your takeaway / analysis of the output provided by the tool, including your level of confidence in it. A couple of sentences will suffice.

Considerations for responsible AI / LLM usage

AI / LLMs are likely to be used in your future workplace and can be an effective tool for the modern statistician / data scientist. However, there are both effective and detrimental ways that LLMs can be used in a learning context. Here are a few things to consider when choosing whether/how to use AI in your coursework:
- AI / LLMs can hallucinate and provide incorrect answers. You must develop your own foundational knowledge of a subject in order to effectively judge and verify whether an LLM's output is trustworthy.
- AI / LLMs use an enormous amount of energy. In order to be climate-conscious, we need discerning use of AI and should be careful not to over-rely on it when other methods (e.g. Google search, non-AI computational tools, human effort) will suffice.
- Employers are interested in people who can, among other things, achieve Task X (with or without LLMs) correctly and efficiently and who can effectively document and communicate *how* they achieved Task X so that it can be verified and reproduced by someone else. This motivates the policy described above.
- One of the broader / more existential threats posed by AI is its potential to diminish human connection. Be mindful of how often you are turning to AI for help when you otherwise would be turning to a human. There is value in day-to-day interactions with classmates, tutors, and professors that go beyond efficiently completing an assignment.

Tips for when AI assistance can be especially useful and appropriate
- Debugging code or interpreting error messages
- Clarifying course concepts. For example:

---

[2] This policy has been adapted from language provided by Drs. Keegan Kang (Bucknell), Yimin Zhang, Michael Posner, and Villanova University

- o "provide me with an intuitive explanation of XYZ"
- o "help me understand why ABC happens in this context"
- o "I'm confused about the difference between ABC and XYZ"
- Deepening understanding of posted solutions
- Generating additional practice questions and step-by-step explanations when studying

(Non-exhaustive) scenarios when AI assistance would be inappropriate
- Copying / typing a homework problem into ChatGPT and having it generate a full solution from start to finish. Homework is intended to build your general problem solving intuition, and you are responsible for coming up with the steps to solve the problem.
- Copying output from ChatGPT directly into your submission. Just as you should not let a classmate write content directly into your submission, so also should you avoid using AI assistance in such a way that directly adds content to your submission.
- Using AI on an open-ended problem or writing assignment that asks for your reflection, opinion, or meta-cognitive thought-processes. It is considered academic irresponsibility to use AI to generate an answer that does not reflect what *you* truly think and believe. I am interested in what you think, not what an LLM thinks.
- Submitting an assignment that has ideas, code, or solutions that originated from AI but is not properly cited. This is plagiarism.

If at any point you are unsure whether a particular use-case of AI is appropriate, please ask!

Notice: Students should be aware that Mediasite Lecture Capture is being used to automatically record each class session. Some recordings may be available to students upon request.

Course Calendar (Tentative)

Week	Dates (M-F)	Topics	Tue	Th	Annotations / Other due	Project component due
1	Jan 12 - 16	Welcome to STAT 4380! Meet the toolkit	AE 01	Lab 01	Ann. 01	
2	Jan 19 - 23	Data Visualization	AE 02	Quiz 01, Lab 02	Ann. 02	
3	Jan 26 - 30	Data Wrangling I	AE 03	Quiz 02, Lab 03	Ann. 03	
4	Feb 2 - 6	Data Wrangling II	AE 04	Quiz 03, Lab 04	Ann. 04	Brainstorm (Th)
5	Feb 9 - 13	Data Wrangling III	AE 05	Quiz 04, Lab 05	Ann. 05	Choose teams (Tue)
6	Feb 16 - 20	Data Importing & Cleaning	Lab 06	NO CLASS (flex day)	Ann. 06	Proposal (Sat)
7	Feb 23 - 27	Review / Project	Quiz 05, Lab 07	Project work session	Ann. 07	
	Mar 2 - 6	SPRING BREAK				
8	Mar 9 - 13	Data Ethics	AE 06	Quiz 06 Project work session	Ann. 08	Cleaning & EDA (Th)
9	Mar 16 - 20	Working with strings	AE 07	Quiz 07, Lab 08	Ann. 09	
10	Mar 23 - 27	Web scraping	AE 08	Quiz 08, Lab 09	Ann. 10	
11	Mar 30 – Apr 3	Communicating results I	AE 09	NO CLASS (Easter)		Preliminary data story (Tue)
12	Apr 6 - 10	Misc. topics	AE 10	Quiz 09, Lab 10		
13	Apr 13 - 17	Communicating results II	AE 11	Quiz 10		Slides rough draft (Th)
14	Apr 20 - 24	Project peer review	Peer review	Peer review	End of course reflection	Report rough draft (Tue) Peer review due (Sat)
15	Apr 27 – May 1	Project presentations	Presentations I	Presentations II		Presentations Final report due (May 4)
	May 11 (Mon)	Presentations III				Presentations