Syllabus - STA-209 (Spring 2026)

Instructor:

- Vinny Paris
- Noyce 2522
- parisvin at gri. . . . .

Class Meetings:

- SST-209-05
- HSSC N2170
- MWF 1:00-2:20 PM

Office Hours:

- Location: Noyce 2522 (Vinny's Office)
- Monday 4:10-5pm
- Tuesday 10am - noon
- Open door -> Feel free to pop your head in, best chances is afternoons on Tuesday/Thursday
- If these times do not work for you, *please* message me and we will find a time to meet.

Mentor Information:

- Evelyn Hao
- Email: haoliang @ gr. . .
- Mentor sessions: Wednesday, 7-8, HSSC N2170

Course Description:

This course covers the application of basic statistical methods such as univariate graphics and summary statistics, basic statistical inference for one and two samples, linear regression, one- and two-way ANOVA, and categorical data analysis. Students will use statistical software to analyze data and conduct simulations. (You will learn all these terms throughout the semester!)

A personal goal of mine is to help you learn how to make publication ready graphics that you can use in the rest of your academic careers.

Please note the software we use is a tool for this course and I do not view this course as a coding class (please think about STA 230 for that next semester).

Texts:

- Introduction to Modern Statistics, 1st Edition (Free online!)

All additional materials will be posted on the course website.

There will be required readings from other sources which will not be provided. Mostly these will be short published papers. They will not be provided in order for you to get comfortable finding papers online given suffcent information (title, author, year, journal)

Aims and Objectives

This course aims to introduce students to the field of statistics, including its vocabulary and fundamental principles. The course will prepare students to read, recognize, interpret, and discuss statistical concepts and their use in scientific applications. The course will provide students an understanding of the role of statistics within the scientific method, and provide students the tools to use data to make informed conclusions.

Learning Objectives   After completing this course, students should be able to:

- Apply methods of exploration, visualization, and statistical analysis to data in order to illustrate key findings and make justifiable inferences using statistical software (R)

- Communicate the methods and results of statistical analyses succinctly and accurately in both writing and speaking

- Read, identify, and critique the statistical concepts and choices of data presentation used in various media publications (newspaper articles, reports, blogs, etc.)

Grading

Engagement and Participation - 0%

Participation in a lab-heavy course is needed. During labs you are expected to help your partner(s) learn the material (which goes beyond simply answering the lab questions), and they are expected to help you. For this, labs are considered the longer activities we do inside the classroom. There is no formal attendance; I recognize you are adults and are capable of making your own decisions. I cannot stress enough though how important the act of showing up to the classroom is for your academic success, in this course and in later in life. I won't claim causality, but there is definitely a correlation.

Assignments - 10%

Assignnments will be assigned regularly. You may turn in exactly two late assignments and recieve full credit. All other late assignments will not be graded (although they will still be commented on). Your lowest score will be dropped (which can include a 0). These will be due on Canvas, generally due 9pm.

Some assignemnts will be done, at least partially, in-class which contain questions that you and your partner(s) will work on answering together. Some assignment questions will be scored for accuracy with feedback given, while others may be scored for effort/completion.

I encourage you to work with other students or visit the Math Lab for help on homework questions, but you should clearly understand all your answers and your assignment should be entirely in your own words. The homework is intended primarily as an exercise to practice and develop your understanding and to assist in identifying weaknesses: you do yourself a disservice if you submit work that you do not fully understand. I try to not assign "busy work".

If you engage in significant collaboration with classmates or tutors, you must explicitly acknowledge that person(s) on the top of your assignment (again, you are encouraged to collaborate, and there are times I will actively instruct you to work in groups). For example, Jack and Jill working together and producing similar answers (right or wrong) is fine. Jack turning in verbatim Jill's answers is not okay. Jack turning in code that is exactly the same as Jill's is also not okay.

Additionally, assignments are encouraged to be completed collaboratively, but if it becomes clear that you or your partner(s) are using a "divide and conquer" approach to answering questions your score on that assignment will be penalized.

Required Readings - 10%

At least three papers will be assigned during the course of the semester to read. They will not be overly long nor mathematically difficult to read. The goal of these papers is to help solidify the big picture understanding that I want you to walk out of here with. A single paper will be assigned one to two weeks before a small quiz on it will be done in class. These quizzes will be closed notes/papers.

These papers will be distinct from the suggested reading tab in the module summary.

Exams (3) - 20% each

There will be 2 midterms and a final exam. Each midterm will focus on a specific set of content, but because course topics are naturally cumulative there may be some questions involving earlier content. The 3 exams will contribute a total 60% towards your end-of-semester grade. Your lowest exam grade will be replaced by it's percentage's square root mulitplied by ten (eg if your lowest test score is 64% it will become 10*sqrt(64) = 80%).

Exams will be announced *at least* one week in advance. Exams will be closed notes, but you will be permitted to bring a note card (3in x 5in, both sides can be written on) and a calculator (no cellphone use allowed). You will not be responsible for writing your own R code on exams, but you should expect to encounter output from R as well as the code that produced it during each exam.

Alternative exam arrangements need to be made at least one week in advance of the time you plan to take the exam; this includes taking the exam in a different location, or times going beyond the given class time. Alternative arrangements are not guaranteed unless proper notification is given. There will be sufficent time after the announcement of an exam to seek an accomodation.

For those of you with accommodations, those accommodations are your right and I intend to respect those rights fully. Even if the semester has already started and you want/need an accomodation we (the college) can do that for you. Please just reach out. Also note that I am never given your diagnosis/background for why; I'm just informed there is an accommodation which I will meet.

Midterm and Final Project - 20%

- Data visualization at midterms - 10%

- Write up - 10%

There will be an ongoing group project throughout the semester. The project will include a few short check-in's with me during a lab periord in class. Near the midpoint of the semester lightening talks (<5 minutes) from each group giving a graphical analysis and exploration of the data will occur. The project will culminate at the end of the semester in a three to five page written report in the last week of the semester. More details will be announced later.

Policies

Extensions for Assignments

Twice, and only twice, a semester you may submit an assignment late for full credit. The late assignment must be handed in before the exam over that material. Assignments are not meant to be busy work but to give you practice with the material.

Please note that the deadline is the deadline and will be treated as a bright line cut off. Fairness is something that is very important to me and the only way I can find to be absolutely fair and not show favoritism is by making a bright line rule. The due date and time is the due date and time. If you are late you have the two late submission policy. Beyond that I will not give credit for late work. One of my favorite students I had to give a 0 to for a 9:01pm submission; I intend to be fair and that means a rigid application of the rules.

Class Sessions

A core component of our class meetings will be working through hands-on labs in a paired programming environment. After the first project, you will have the freedom to choose your partner, or work independently near someone that you can occasionally consult with. During labs it is essential that you and your partners work together, and that you assist your neighbors making certain that each of you understand your work equally well.

Most labs will begin with a brief "preamble" section that we will go through together as class. The purpose of this section is to introduce the topic of the lab and ensure a smooth start to each class meeting.

Attendance

Because this course involves some amount of group work, absences impact not only yourself but also your classmates. That said, I understand that missing class is sometimes necessary. I will not take attendance directly; be careful with this policy.

Please note that if you are sick I do NOT want you in class. Please do not come to class. If you show up and are clearly demostrating sympotoms of being sick you will be asked to leave.

Software

Software is increasingly an essential component of statistics and will play a role in this course. We will primarily use `R`, an open-source statistical software program.

You are welcome to use your own personal laptop, or a Grinnell College laptop, during the course. `R` is freely available and you can download it and it's UI companion, `R Studio`, here (note: `R` must be downloaded and installed before `R Studio`):

1. Download `R` from http://www.r-project.org/

2. Download `R Studio` from http://www.rstudio.com/

You may also work on a classroom computer, all of which will have `R` and `R Studio` pre-installed.

Alternatives to `R Studio` do exist and you are invited to use them if you are already familiar with them. Please understand though that my knowledge of some of them is more limited. The UI (user-interface) companion chosen will not affect the final code.

Grinnell hosts an online version of R Studio that you may use while on campus internet: https://rstudio.grinnell.edu/

Academic Honesty

At Grinnell College you are part of a conversation among scholars, professors, and students, one that helps sustain both the intellectual community here and the larger world of thinkers, researchers, and writers. The tests you take, the research you do, the writing you submit-all these are ways you participate in this conversation.

The College presumes that your work for any course is your own contribution to that scholarly conversation, and it expects you to take responsibility for that contribution. That is, you should strive to present ideas and data fairly and accurately, indicate what is your own work, and acknowledge what you have derived from others. This care permits other members of the community to trace the evolution of ideas and check claims for accuracy.

Failure to live up to this expectation constitutes academic dishonesty. Academic dishonesty is misrepresenting someone else's intellectual effort as your own. Within the context of a course, it also can include misrepresenting your own work as produced for that class when in fact it was produced for some other purpose. A complete list of dishonest behaviors, as defined by Grinnell College, can be found here.

Inclusive Classroom

Grinnell College makes reasonable accommodations for students with documented disabilities. To receive accommodations, students must provide documentation to the Coordinator for Disability Resources, information can be found here. If you plan on using accommodations in this course, you should speak with me as early as possible in the semester so that we can discuss ways to ensure your full participation in the course.

Religious Holidays

Grinnell College encourages students who plan to observe holy days that coincide with class meetings or assignment due dates to consult with your instructor in the first three weeks of classes so that you may reach a mutual understanding of how you can meet the terms of your religious observance, and the requirements of the course.

Pregnancy Related Conditions, Title IX

Grinnell College is committed to compliance with Title IX and to supporting the academic success of pregnant and parenting students and students with pregnancy related conditions. If you are a pregnant student, have pregnancy related conditions, or are a parenting student (child under 1-year needs documented medical care) who wishes to request reasonable related supportive measures from the College under Title IX, please email the Title IX Coordinator at titleix@grinnel.edu. The Title IX Coordinator will work with Disability Resources and your professors to provide reasonable supportive measures in support of your education while pregnant or as a parent under Title IX.

Getting Help

In addition to visiting office hours and completing the recommended readings, there are many other ways in which you can find help on assignments and projects.

The Data Science and Social Inquiry Lab (DASIL) is staffed by mentors who are experienced in R programming and may be able to troubleshoot coding problems you are having.

The Grinnell Math Lab is located on the 2nd floor of Noyce Science Center in Room 2012 and offers drop-in statistics tutoring.

The online platform Stack Overflow is a useful resource to find user-generated coding solutions to common R problems. Nearly all professionals have needed to "look up" a coding strategy on a site like Stack Overflow at some point in their career, and I have no problem with you doing the same on assignments or projects. However, if you make substantial use of a Stack Overflow answer (ie: actually integrating lines of code written by someone else into your work, not just getting help identifying the right functions/arguments) the expectation is that you cite or acknowledge doing so. I admit the line can be somewhat blurred but generally if you copy and paste 20+ lines of code and don't know what it's doing that is a problem.

Large Language Models

Large language models, such as ChatGPT, Bing Chat, or Bard, can be a useful tool for explaining and fixing errors in your R code, or helping you understand example code. I myself am optimistic about LLM's abilities to cut down on ''data cleaning'' and am looking forward to folding that into future statistical consulting work I might do. You are welcomed to use these tools in only in a single, very limited case: You may use LLM to help generate additional examples to help you study for your exams. I do not want you to use LLM for coding.

However, you are ultimately need to learn the material and LLM's can be incorrect espeically given I have quirks in how I want things interpretted/understood. The model may provide inaccurate information or generate text that is superficial and lacking sufficient detail. I encourage you to read Professor Erik Simpson's write-up on writing with LLMs to see some reasons why you shouldn't lean too heavily on these technologies. Nevertheless, you're welcome to use large language models in this course in the above manner. Overly relying on these tools may set you up to do poorly on exams.

To be abundantly clear I do not want you to use LLM's for coding in this course. Help pages, peers, stack overflow, DASIL are sources I would prefer you to learn to use.

Note: My stance on LLM usage through the course may change, but you will be given notice and, if applied retroactively, won't lower your grade

ROUGH Sketch of Schedule

Date	Topic
Jan 21	Intro, Install R, Fundementals
23	Visualization 1 + 2
26	Visualization accessibility + ggplot2
28	Numeric Summaries + Boxplots
30	Correlation + Catch Up Day (Code Along)
Feb 2	Simple Linear Regression
4	SLR: Categorical (+ Midterm Project Proposals)
6	SLR: Transformations
9	SLR lab/catch up
13	MLR lab
16	MLR
	END OF UNIT 1 MATERIAL
18	Probability 1 (Basics + Review) + basic sampling
20	Exam 1
23	Permutation Tests
25	Probability 2 (Normal) + Sampling Distribution
27	Z/t - tests
Mar 2	Z/t - tests Lab
4	Midterm Projects
6	Z/t -tests lab / Catch Up day
23	Errors
25	We Inferences 3
27	We Inferences 3
Apr 1	We Inference 3 Lab
	END OF UNIT 2 MATERIAL
3	Sampling Methods (and Final Project Intro)
6	Exam 2
8	Coefficient Testing
10	Coefficient Testing Lab
13	ANOVA
15	ANOVA Lab
17	Contingency Tables
20	Chi Squared Test
22	Tables and Chi Squared Test Lab
24	Logistic Regression
27	Logistic Regression Lab
29	Mixed Models
1	Mixed Models Lab
4	Bootstrapping
6	Bootstrapping Lab
8	Fun Day (Advanced Regression?)