Stat 850: Computing Tools for Statisticians

Instructor: Susan Vanderplas   Time: TR 9:30-10:45
E-mail: svanderplas2@unl.edu   Location: Keim 214
Office: Hardin 343D
Student Hours: By Appointment - use this link

1   Course Description

Introductions to statistical computing packages and document preparation software. Topics include: graphical techniques, data management, Monte Carlo simulation, dynamic document preparation, presentation software.

1.1   Goals

1. Learn how to use R and/or Python for data analysis, data processing, data visualization, and statistical simulation.
2. Become familiar with the process, techniques, and goals of exploratory data analysis.
3. Create, assess, and debug code effectively.
    i. Use online resources to find software to perform a task, comparing approaches taken by competing programs.
    ii. Read error messages, find related problems in online forums, and isolate the conditions necessary to generate the error.
    iii. Generate minimum working examples or reproducible examples of errors in order to ask for help effectively.
4. Communicate statistical results using reproducible, dynamic tools. Understand the importance of reproducibility in scientific computation.

1.2   Objectives

(what you should be able to do at the end of this course)

A. Clean and format the data appropriately for the intended analysis or visualization method. (Goals: 1)

B. Explore a data set using numerical and visual summaries, developing questions which can be answered using statistics. (Goals: 1, 2)

C. Evaluate methods or software to assess relevance to a problem. Compare similar options to determine which are more appropriate for a given application (Goals: 1, 3)

D. Test and debug software, using the following sequence: (Goals: 3, 4)

    1. Reproduce the error in a new environment,
    2. Create a minimal reproducible example,
    3. Research the error message and evaluate online resources for relevance,
    4. Ask for help, describing the error or problem appropriately.

E. Document the data, methods, and results of an analysis using reproducible methods. (Goals: 1, 2, 4)

F. Construct a reproducible statistical simulation for a given problem using methods such as MCMC, inverse probability sampling, and rejection sampling. (Goals: 1, 3, 4)

2   Materials and Resources

In this course, I will assume that you have access to:

- A computer (ideally a laptop) capable of running R, python, RStudio, and quarto.
- Access to a scanner or camera to scan or photograph handwritten sketches, etc. for homework assignments.

2.1   Textbook

In keeping with the principles of this course, any course materials are made available on GitHub, in the (continuously evolving) course textbook. The book is laid out with the same structure as the course. In order to avoid duplicating content available elsewhere, where it is appropriate, there are links provided to relevant material available on other sites. This makes the course easier to maintain, but it also ensures you get the most relevant and up to date instructions.

In addition, you may find it useful to reference some of the following resources that I have consulted while assembling the textbook. Most are available online for free, though some require an institutional email address.

- R for Data Science - Volunteer translations for this exist for Spanish, Portuguese, Turkish, and Italian.

- Advanced R

- Python for Everybody

- Python for Data Analysis *

- Python Data Science Handbook *

* Available online for free if you register with your UNL email address.

Additional readings will be selected from various other books and will be made available via Library Course Reserves.

3   Class Schedule & Topic Outline

This schedule is tentative and subject to change. Students are expected to read the corresponding textbook chapter (linked in Canvas) prior to coming to class.
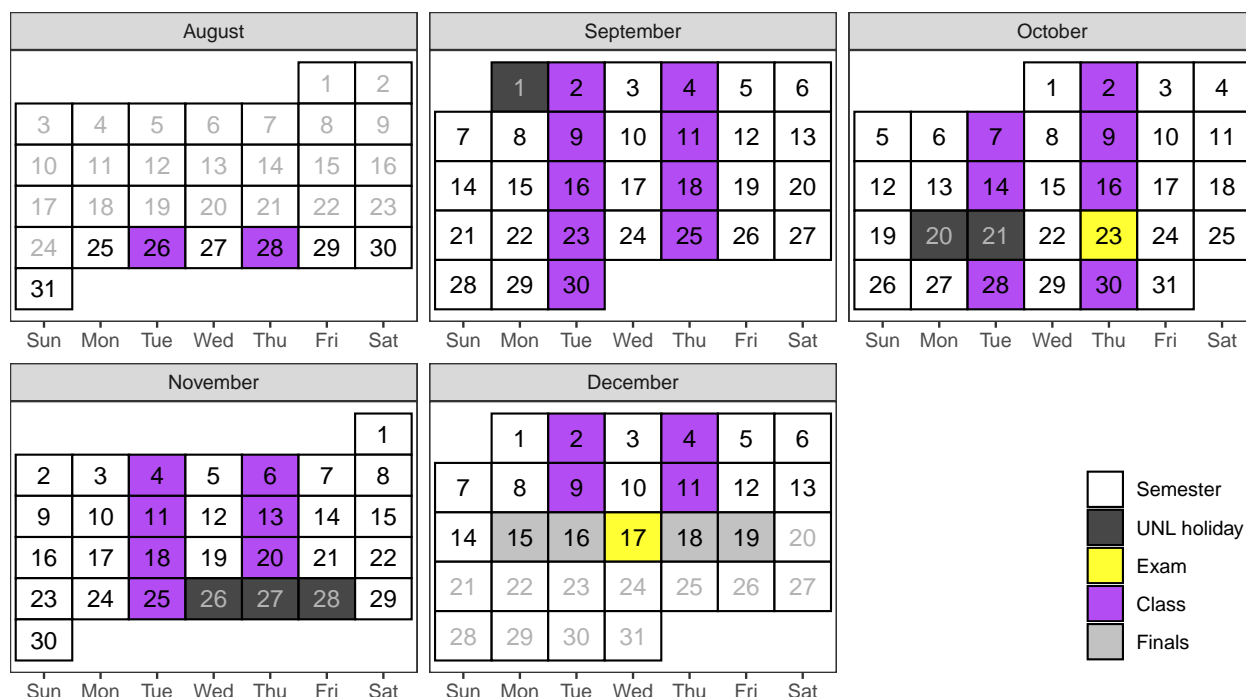
August						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

September						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

October						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

November						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

December						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

Legend:
- Semester
- UNL holiday
- Exam
- Class
- Finals

Table 1: Tentative schedule of class topics & project due dates

Week	Topic	Important Dates
1	Exploring the Toolbox	
2	Introduction to Programming and Variable Types	
3	Data and Control Structures	
4	Data Programming	
5	Reading and Exploring Data	
6	Data Cleaning and Manipulation	
7	Data Joins &Transformations	
8	Graphics	
9	Fall Break & Midterm Exam	Midterm: Oct 23
10	Reproducibility & Documents	
11	Functional Programming	
12	Simulation	
13	Simulation	
14	Thanksgiving	
15	Interactive Graphics	
16	Final Part 1	
17	Final Exam	Final: Dec 17

4   Course Essentials

4.1   Assessment/Grading

Assignments	Weight
Reading Quizzes & Participation	10%
Weekly Homework	50%
Midterm	20%
Final	20%

Lower bounds for grade cutoffs are shown in the following table. Do NOT ask for a grade bump or to have your grade rounded up - this is a statistics course, and I expect you to understand the mathematical rules of rounding.

Letter grade	X +	X	X -
A	96.5	93.5	89.5
B	86.5	83.5	79.5
C	76.5	73.5	69.5
D	66.5	63.5	60.5
F	<60.5		

Interpretation of this table:

- A grade of 86.4 will receive a B, and a grade of 86.6 will receive a B+.
- A grade of 73.2 will receive a C-, and a grade of 73.9 will receive a C+.
- Anything below a 59.5 will receive an F.

4.2 Course Components

Reading Quizzes 📖

This class uses a flipped classroom model, which assumes that you read the material and/or watch the videos before class so that you are prepared to apply the material during in-class activities. This model is most familiar in science laboratory settings - there may be a brief review of material at the start of the lab period, but most of your class time will be spent working interactively in pairs or small groups.

To ensure that you are prepared for class, there are low-stakes quizzes to accompany most textbook chapters; these quizzes will highlight some important items to remember from each chapter. You are encouraged to have the textbook chapter(s) open when you complete the quiz.

Quizzes will be due at 8am on class days (if there are multiple textbook chapters assigned for the week, one quiz may be due on Tuesday and one on Thursday, depending on whether the material is intended to be consumed sequentially or in parallel).

Homework 💻

Weekly formative assignments (assignments meant to help you practice skills) will be given throughout the semester.

- You will typically have one week to work on each of the assignments.
- Assignments must be submitted in the file format specified, and should run or compile as submitted for credit to be given.
- I will attempt to grade formative assignments within a week of the specified due date.
- The assignment due date in Canvas will be set to at least 48 hours before the hard deadline for assignment submission. This provides an implicit "grace period" for assignment submission. No additional extensions will be provided.

Contextual Reading

Articles will be provided for each week in order to give you different perspectives on the week's material or provide context for the week's topic. These articles may relate to how the material is used in the real world, what statisticians were thinking when the concept was formalized, differences between implementations in R/Python/SAS, or ethical or societal challenges associated with the material. You will be expected to summarize and respond to the articles in free-response (essay/short-answer) questions on your midterm and final exam.

Exams (Midterm and Final)

Tentative dates: - Midterm: - October 23 (in-class portion) - October 23-24 (outside of class) - Final: - December 11-12 (outside of class), - December 17 (scheduled final exam), 10am - 12pm

Exams will be hybrid, with an in-class portion as well as a take-home portion. On exams, you may NOT use AI or external resources like StackOverflow unless indicated by the problem.

Exam problems which ask you to program will be conceptually similar to the homework assignments. Exams will also include questions which require you to write a paragraph or a short essay. It is difficult to write coherently about a topic if you don't have a good understanding of that topic and related concepts. Testing your ability to write and explain something is one way of measuring higher level understanding and integration of course material.

4.3 Course Policies

Late Work Policy ⏰

Late assignments will be accepted only under extenuating circumstances, and only if you have contacted me **prior** to the assignment due date and received permission to hand the assignment in late. I reserve the right not to grade (or to assign a 0 to) any assignments received after the assignment due date without prior approval. Most lab assignments build in a "grace period" - they are typically due on Friday and the assignment closes Sunday evening. This additional time serves as a general deadline extension to provide an extra few days to work on the assignment.

Exam Make-Up Policy

Students should make every possible attempt to ensure that they are available for both the midterm and final exam. The following policies will determine whether a student is eligible for a make-up exam or other accommodation:

- Only university-excused absences will be accommodated.

- If there is a scheduled conflict with an exam date, students should contact the instructor as soon as practicable. In most cases, conflicts must be identified within the first two weeks of the semester.

- In the case of emergent conflicts (illness, bereavement), students must notify the instructor as soon as possible. In most cases, this notice must be received before the scheduled exam date.

- If the midterm exam is missed due to a university-excused absence, the weight of the midterm will be transferred to the final exam.

- If the final exam is missed due to a university-excused absence, the instructor will work with the student to determine an appropriate accommodation or alternate assignment.

Attendance & Participation

You are expected to attend class and participate in online discussions. Consistent, repeated failure to attend class or actively participate in class will affect the participation portion of your grade.

Excused Absences

The university lists several reasons that a student may miss a class or significant component of the course such as an exam Faculty Senate Class Attendance Policy.

- illness of self or dependent,
- participation in UNL-sponsored activities,
- military service or jury duty,
- bereavement, or
- religious observance

In all cases, students are expected to provide notice to the instructor or the university as soon as possible.

Religious observance requests must be made by the 2nd week of class.

Conflicts with scheduled UNL-sponsored activities (including participation in conferences) should be communicated to the instructor by the 2nd week of class (if scheduled at the start of the semester) or with at least 2 weeks notice. Emergent conflicts will be handled on a case-by-case basis, but prior notice is expected.

To initiate bereavement leave, students should contact the Vice Chancellor for Student Affairs, and that office will coordinate with instructors.

In the case of extended or chronic illness, the student should work with Disability Services to obtain formal accommodations. Disability Services may require medical documentation as part of the accommodations process.

> Note that scheduling of vacations or international travel is NOT considered a university approved absence.

Illness 🥴 & Excused Absences

If you are feeling ill, please **do not come to class**. Instead, review the material and work on the homework/lab assignment, and then schedule an appointment with me to meet virtually within a week of your absence. In the appointment reason field on Calendly, indicate that this appointment is to substitute for your in-class participation on the date you missed.

If you need to miss more than 1-2 classes for illness, I reserve the right to require documentation from Disability Services or a medical provider.

Inclement Weather 🌨️ 🌀 🌧️

If in-person classes are canceled, you will be notified of the instructional continuity plan for this class by Canvas Announcement. When there is power in the Lincoln area, we may hold class via Zoom; however, I may assign an asynchronous activity instead. You will be held to the same participation and professionalism standards on Zoom as you are in person, and participation and attendance will be graded accordingly. That is – cameras should be on, you should be fully clothed, and you should be an active participant and not an empty black box. If there is a reason you don't want your camera to be on (e.g. if you're in an environment you can't fully control, or you're sick and don't want to be on camera), just let me know. Pets are 100% welcome to join the zoom call ❤️ 🐶 🐱.

4.4   Expectations

You can expect me to:

- reply to emails within 48 hours during the week (72 hours on weekends)
- be available in class to assist with assignments
- be available by appointment for additional help or discussion

I expect you to:

- Read the module material and homework assignment before coming to class
- Engage with the material and your classmates during class
- Seek help when you do not understand the material
- Communicate promptly if you anticipate that you will have trouble meeting deadlines or participating in a portion of the course.
- Do your own troubleshooting before contacting me for help (and mention things you've already tried when you do ask for help!)
- Be respectful and considerate of everyone in the class

Communication

If you are lost or confused or have a problem related to the course, make contact early! It is much easier to address a small problem early than to fix a cascade of issues later in the semester.

Make Mistakes! 🚗 💥

Programming is the process of making a series of silly or stupid mistakes, and then slowly fixing each mistake (while potentially adding a few more). The only way to know how to fix these mistakes (and avoid them in the future) is to make them. (Sometimes, you have to make the same mistake a few dozen times before you can avoid it in the future). At some point during the class, you will find that you've spent 30 minutes staring at an error caused by a typo, a space, or a parenthesis in the wrong place. You may ask for help debugging this weird error, only to have someone immediately point out the problem... it is always easier to see these things in someone else's code. This is part of programming, it is normal, and you shouldn't feel embarrassed or sorry (unless you put no effort into troubleshooting the problem before you asked for help).

If you manage to produce an error I haven't seen before, then congratulations. You have achieved something special, and that achievement should be celebrated. Each new and bizarre error is an opportunity to learn a bit more about the programming language, the operating system, or the interaction between the two.

Evaluation Criteria

In every assignment, discussion, and written component of this class, you are expected to **demonstrate that you are intellectually engaging with the material**. I will evaluate you based on this engagement, which means that technically correct but low effort answers which do not demonstrate engagement or understanding will receive no credit.

When you answer questions in this class, your goal is to show that you either understand the material or are actively engaging with it. If you did not achieve this goal, then your answer is incomplete, regardless of whether or not it is technically correct. This is not to encourage you to add unnecessary complexity to your answer - simple, elegant solutions are always preferable to unwieldy, complex solutions that accomplish the same task.

Grammar and spelling are important, as is your ability to communicate technical information clearly in a written format; both of these criteria will be used in addition to assignment-specific rubrics to evaluate your work.

Academic Integrity and Class Conduct

You will be engaging with your classmates and me through in-person discussions and collaborative activities. It is expected that everyone will engage in these interactions civilly and in good faith. Discussion and disagreement are important parts of the learning process, but it is important that mutual respect prevail. Individuals who detract from an atmosphere of civility and respect will be removed from the conversation or the classroom.

Students are expected to adhere to guidelines concerning academic dishonesty outlined in Article III B.1 of the University's Student Code of Conduct. The Statistics Department academic integrity and grade appeal policy is available here.

AI and Explainability Policy

Any use of generative AI must be disclosed in an appendix to your submission - this includes brainstorming, editing, using AI as spell-check/grammar-check, and so on. You must document the following:

- the version of the generative AI used
- the full sequence of prompts and responses
- any additional inputs you provided to the AI system
- a "diff" between the AI responses and your submission, showing exactly what was generated by the AI system and what you changed.

It may be useful to leverage AI tools to ensure that your work conforms to grammar and style guidelines, but I very highly discourage the use of generative AI for content or code.

Oral Exams

I reserve the right to replace the grade on any written assignment with a graded oral exam on your submission. This policy is not intended to be punitive – it may be used to resolve issues where question wording was not clear or was misinterpreted. Rather, it is intended to be used to assess what you actually know when I suspect that the written assignment is not a good measure of your knowledge or skills.

Code

You must be able to explain any work you turn in. - If you cannot explain the logic behind your approach as well as how it works in practice, then you will not receive credit for your submission. This is in line with what is generally considered acceptable behavior in programming - reuse is fine (subject to the code's license), but you must still be able to **understand** and **modify** any code you did not write yourself.

- If you copy and modify code from another source (StackOverflow, Rbloggers, etc.), provide that source in a comment above the code. This provides you with documentation as to how you got to the solution, and also provides credit to the original source. (This practice has been helpful to me on many occasions when I come back to code I wrote long ago)

Exams

No external resources are allowed on exams (including, but not limited to: the internet, AI, StackOverflow, a friend) for assistance unless specifically stated in the problem (and then, only on that specific problem). Use of any external resources will be grounds for an academic integrity violation under Section 2.A.5, "Failing to follow the rules".

5   Required University Information

See https://executivevc.unl.edu/academic-excellence/teaching-resources/course-policies.