Stat 892/992: Special Topics in Data Visualization

Instructor: Susan Vanderplas	Time: TBD
E-mail: susan.vanderplas@unl.edu	Location: TBD
Office: Hardin 343D	
Student Hours: Schedule here	

Course Description

Design and use of data visualizations for statistical communication. Topics include the grammar of graphics, methods for evaluating graphics for utility, visual inference and visual statistics, high dimensional graphics, and exploratory data analysis methods. This course will be reading and writing intensive. Familiarity with R or python programming (pandas, numpy) is expected.

Course Goals

1. Assess the consequences of different graphical design decisions, identifying primary and secondary comparisons and likely areas of user focus.
2. Compare the principles of the grammar of graphics with implementations in different software packages and identify the consequences of these implementation decisions.
3. Evaluate existing visualization studies and suggest designs for new experiments to evaluate the effectiveness of different visualization design decisions.
4. Discuss the similarities and differences between visual inference and classical Frequentist or Bayesian statistical inference procedures.
5. Use appropriate methods to view high-dimensional data and discuss the strengths and weaknesses of different approaches to high dimensional data visualization, including nonlinear dimension reduction (t-SNE, UMAP) and interactive tours.

Course Objectives

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

A. Given a graphic, identify primary and secondary comparisons, likely areas of user focus, and describe the consequences of different design decisions. Suggest (and create) other alternative graphics that improve deficiencies in the original, and compare the strengths and weaknesses of the different versions of the chart before deciding on an optimal version. (Goals: 1)

B. Discuss or write a comparison between one or more software implementations of the grammar of graphics and the theoretical construct. Identify and critique the implementation decisions and assess the consequences of these decisions for usability and consistency within a syntactic framework. Examine defaults within each implementation and evaluate the pros and cons of these default design constructs when creating graphics. (Goals: 1, 2)

C. Develop a user study that evaluates a visualization design in a way that extends existing research in the field. Develop appropriate data sets and models, identify necessary experimen-

tal controls, and generate relevant stimuli for the experiment. In a mini-paper, motivate the experiment by connecting it to existing visualization literature and grounds the study conceptually. If applicable, examine the role the grammar of graphics plays in the implementation of visualization research. (Goals: 3; 2, 4 optional)

D. Using a specific example, develop a visual inference experiment that answers a research question. Write a paper comparing and contrasting the goals and implementation of a statistical inference procedure and a visual inference procedure, evaluating what can be learned from each and what factors may complicate each procedure. (Goals: 3, 4)

E. Describe the differences between two high-dimensional data visualization techniques, and what can be seen from each. For a given dataset or scenario, suggest a procedure for navigating high-dimensional data to explore for relationships between the variables. (Goals: 1, 5)

Textbooks

All textbooks which are required for this course are electronically reserved through the library - you should be able to access them for free.

- Getting (more out of) Graphics by Antony Unwin
- Exploratory Data Analysis by John Tukey
- Visualization Analysis and Design by Tamara Munzner
- Fundamentals of Data Visualization by Claus Wilke

Assessment/Grading

Assignments	Weight
Weekly Writing Assignments	30%
Participation and Discussion	20%
Papers	50%

Missing more than 4 classes without *prior* instructor permission will result in a reduced grade beyond the 20% allocated to class participation.

The 892 and 992 versions of this class will have different assignments and papers, but the overall grading scheme will be similar. In general, students taking the 992 course will be expected to discuss concepts at a higher level (this does not necessarily mean writing more text) and will write more papers.

Lower bounds for grade cutoffs are shown in the following table. I will not "round up" grades at the end of the semester beyond strict mathematical rules of rounding.

Letter grade	X +	X	X -
A	97	94	90
B	87	84	80
C	77	74	70
D	67	64	61
F	<61		

Interpretation of this table:

- A grade of 86.4 will receive a B.
- A grade of 76.8 will receive a C+.
- A grade of 73.1 will receive a C-.
- Anything below a 61 will receive an F.

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.
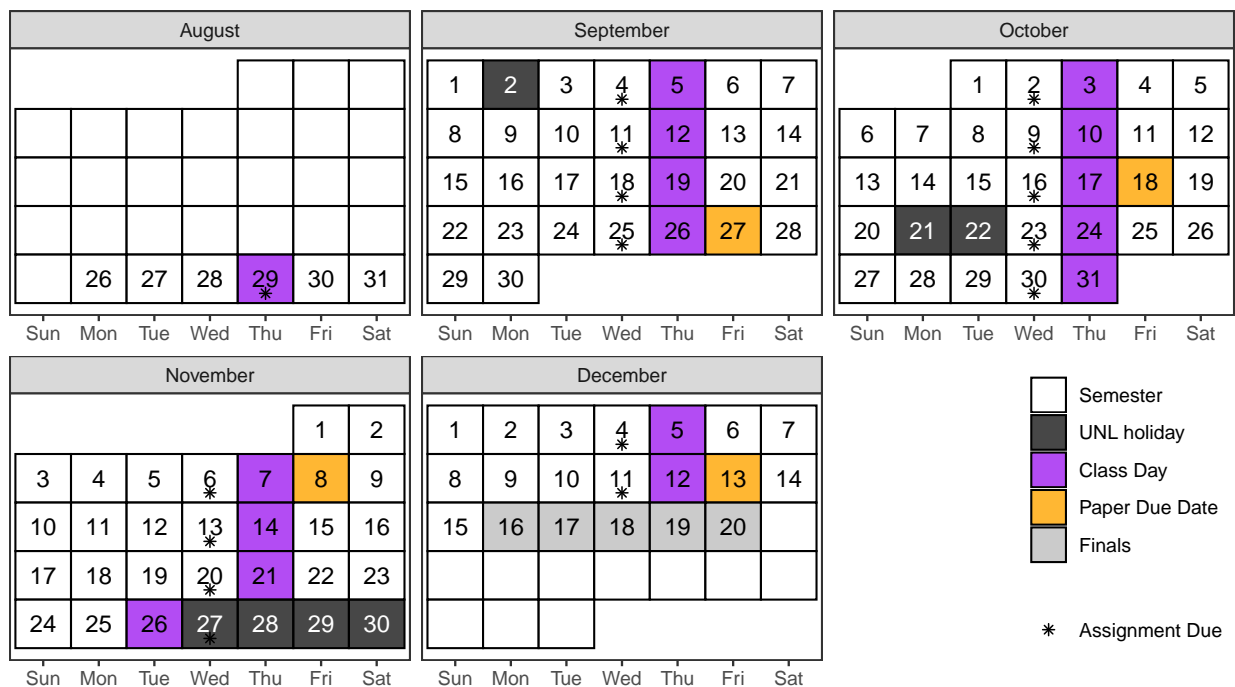

Table 3: Tentative schedule of class topics & project due dates

Date	Time	Topic
Aug 29		Syllabus, Introduction
Sep 5		Grammar of Graphics
Sep 12		Modifying the Grammar
Sep 19		Other Grammars

Date	Time	Topic
Sep 26		Evaluating Graphics
Sep 27	12pm	992 Paper #1 Due
Oct 3, Oct 10		Graphical Testing
Oct 17		Audience Considerations
Oct 18	12pm	892 Paper #1 Due
Oct 24		Visual Statistics
Oct 31, Nov 7		Visual Inference
Nov 8	12pm	992 Paper #2 Due
Nov 14, Dec 5		High Dimensional Vis
Nov 21		Tours
Nov 26		Dimension Reduction
Dec 12		Data Art/Infographics
Dec 13	12pm	Final Paper Due (892/992)

Course Policies

Attendance

You are expected to attend class. If you are feeling ill, please **do not come to class**. Contact me (you don't have to provide details) and let me know why you're not in class. If you are feeling well enough to participate virtually, contact me at least 2 hours prior to class and I will do my best to facilitate your participation.

If you miss more than 1 lecture, contact me and we will discuss how to proceed with the course. Missing more than 2 lectures will not only impact your participation grade, but may result in grade reduction or failing the course. Consistent participation is essential to get something useful out of a discussion-based course.

Feel free to contact me and let me know if you are having trouble or anticipate the possibility of trouble in the near future. It is much easier to communicate **early and often** than to fix problems that arise later in the semester.

Assignments

This class is very different from other advanced statistics classes: there will be more reading and writing, and more discussion, than you would typically encounter in a traditional statistics course. There will most likely be no proofs, and there should be relatively little computation (though you will be programming in order to create examples for writing and discussions).

Weekly Writing Assignments

There will be short (1-2 page) writing assignments that summarize the reading for each week. These assignments are intended to prompt you to reflect on the reading before we discuss items in class. Assignments will be graded primarily for effort and insight.

Participation and Discussion

This class will be primarily discussion based - you are expected to prepare for class by doing the reading and writing assignments before class starts. Many of these topics will require you to read and digest the material (you may have to read it several times!) in order to be prepared for the class discussions.

In a discussion class, it is important to both contribute regularly and to not dominate the discussion. The following principles (from https://tll.mit.edu/teaching-resources/inclusive-classroom/discussion-guidelines/) will guide discussions in this class, and adherence will be part of the participation grade.

- Listen respectfully, without interrupting.
- Listen actively and with an ear to understanding others' views. (Don't just think about what you are going to say while someone else is talking.)
- Criticize ideas, not individuals.
- Avoid blame, speculation, and inflammatory language.
- Allow everyone the chance to speak.
- Avoid assumptions about any member of the class or generalizations about social groups.
- We are accountable for our words and their impact.
- Personal information that comes up in the conversation should be kept confidential.

Papers

Two to three longer (~15-page) papers will be assigned during the course of the semester. Papers must be written using LaTeX, with BibTeX used for citations (it is also acceptable to use Quarto/RMarkdown to generate PDF content, using LaTeX/BibTeX under the hood). I recommend using Zotero + ZoteroConnector to keep track of papers and generate references; be aware that you will often have to fill in additional information that isn't provided by ZoteroConnector (but this is still faster than generating the BibTeX yourself).

If you use graphics generated from code in your paper, you must include the code in a reproducible fashion (e.g. using Quarto/Sweave/Rmarkdown) and it must compile on my machine without modification.

Students taking the class at the 900 level will submit 3 papers (due in late Sept, early Nov, and mid Dec), while students taking the class at the 800 level will submit 2 papers (due mid-Oct and mid Dec).

Expectations

You can expect me to:

- reply to emails within 48 hours during the week (72 hours on weekends)
- be available in class to lead and participate in the discussion
- be available by appointment for additional help

I expect you to:

- Read the assigned material and complete the reflective writing 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

General 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 unwieldly, complex solutions that accomplish the same task.

While this is not an English class, grammar and spelling are important, as is your ability to communicate technical information in writing; both of these criteria will be used in addition to assignment-specific rubrics to evaluate your work.

Version Control

This class will require you to use Git and GitHub for assignments. If you are unfamiliar with Git/GitHub, talk to me before class starts so that I can arrange for you to obtain help from a peer.

I will use your version control history to verify that you are completing assignments before class (so you will be responsible for ensuring your work is pushed/submitted before class starts) and to assess the pattern of work for longer projects. Commit your changes frequently, and push any work you've done to github any time you take a break, even if the assignment is not complete. This will ensure that I have accurate information, and also will ensure that your work is backed up to the cloud regularly.

Late 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. Weekly writing assignments are intended to prepare you for class; if you do not complete these before the class period, they are not serving their designed purpose. There are very few circumstances under which I will allow you to submit weekly writing assignments late, and almost all of them involve hospitalization and missing the discussion portion of the class.

I reserve the right not to grade any assignments received after the assignment due date.

Make Mistakes, Have Opinions!

While this is not a programming class, many of the the same sentiments apply to class discussions.

Don't be afraid to state an opinion, even if no one else agrees. Different opinions make for great discussions. Don't be afraid to say "I don't understand XXX - it sounds like YYY, but what about ZZZ?". It is fine to not understand something, and often, it indicates a lack of clarity in how something is presented, or that your experience with a topic is different from the author's.

Similarly, if someone doesn't understand something that seems trivial to you, rather than saying "it's trivial", ask questions to determine how your understanding differs from their understanding - everyone's brain and perceptual systems are different! You may discover during this class that you have a color vision deficiency or that you don't process depth cues the same way as everyone else – both considerations that impact how graphics "work". Everyone in the class can learn something from these discoveries - while we can't experience how someone else's brain works, it's a good thing to discover just how much different things vary from individual to individual.

For instance, I have color vision deficiency (colorblindness) and prosopagnosia, which means I don't process faces holistically. As a result, I *hate* Chernoff faces because they just don't do anything for me. Perhaps they're useful for people with normal face perception – but that doesn't make my experience wrong or invalid - it just means I need to allow for the fact that maybe Chernoff faces are useful for some people (though evidence suggests they're probably not). Similarly, something might be objectively green, but I may say it's brown during a discussion. It's totally fine to (kindly) say in a class discussion that you perceive the object to be green, and then pivot to a discussion of whether the design is robust to color vision deficiency - a topic that is useful to consider when designing graphics. No one can say that my perception is wrong (because no one else can see inside my brain), but it is fine to say that your perception differs.

Inclement Weather

If in-person classes are canceled, you will be notified of the instructional continuity plan for this class by Canvas Announcement. In most circumstances where there is power in the Lincoln area, we will continue to hold class via Zoom.

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.

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.

Your written work will be graded on both your writing and argumentation skills and demonstration of course learning goals and objectives. Each written assignment will be accompanied by a rubric outlining basic standards and goals, but this should not be considered to be all inclusive - your primary goal in any assignment is to show me that you are engaging with the material.

Any sources which contribute to your writing must be cited, including AI engines and web pages that are not typical scholarly sources. I do not care which format you use for source citations (APA, Chicago, IEEE numeric are all fine) so long as you provide information which is as complete as possible AND which is sufficient to find the source material.

AI Policy

AI tools do not have the ability to provide insight into specific problems - they recombine text from other sources, but this often results in vague, bland, and muddled writing. It is unlikely that using AI to provide more than a basic structure/outline to get you started writing will be useful for this class.

If you use AI tools on any assignment in this course, you must provide the following at the time the assignment is submitted:

- an attachment or appendix to your assignment that includes:
    - the prompt you provided, including any back-and-forth prompt engineering
    - the AI response(s)
    - a document showing the difference between the AI response(s) and your final submission (and there should be many, many differences). The `diff` command (or `latexdiff`) may be helpful to generate this document.
- citations for any claims the AI made which remain in your final submission. As most AI engines do not cite contributing sources, you will need to find sources to support each claim made in your submission. **Sources which cannot be verified and are suspected to be AI hallucinations are cause for an automatic 0% on the assignment.**
- a citation to the AI engine used and its specific version, for reproducibility purposes.

I reserve the right to replace the grade on any assignment with an oral exam over that assignment's content and key concepts.

Required University Information

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