Syllabus

Statistics 100

*Fall 2024*

Course Instructor: Julie Vu (julievu@g.harvard.edu)

Class Time: Mon/Wed 1:30 - 2:45pm

Course Site: https://canvas.harvard.edu/courses/138349

1. Learning Outcomes

In this course, you will learn how to think critically with data by engaging with the entire data analysis process, as shown in the diagram below. While most of our time will be spent learning techniques related to the *Exploration and Visualization* and *Modeling and Inference* stages, you will also practice the other important components of analyzing data. Since computation is an integral part of modern statistical work, you will learn to write R code – leveraging the RStudio user interface – to analyze data. Furthermore, you will practice effectively communicating the results of analyses.
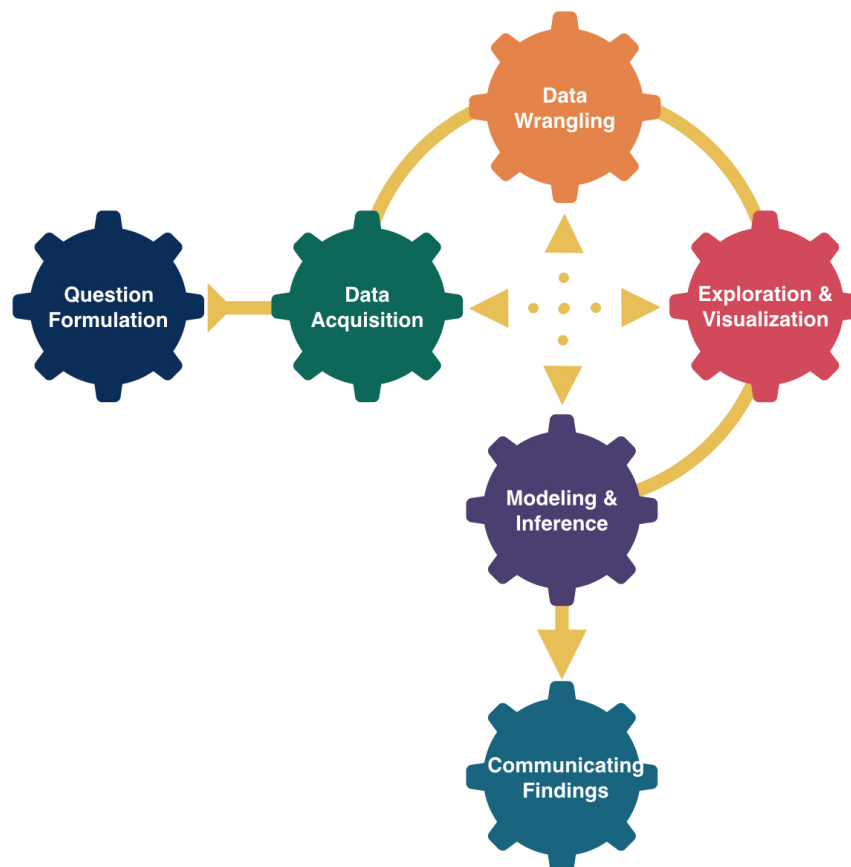
By the end of the course, you will be better prepared to accomplish the following tasks:

Question Formulation:

– Translate a research problem into a set of questions that can be answered with data.
– Formulate data questions as measurable statements about parameters in a model.

Data Acquisition:

– Determine the necessary data to conduct analyses.
– Reflect on how design structures and data collection impact potential conclusions.
– Identify potential ethical concerns surrounding data collection and data privacy.

Data Wrangling:

– Examine datasets to determine what wrangling may be required (e.g., removing missing values, filtering out variables or observations, collapsing categories of a categorical variable).
– Apply basic data wrangling operations.

Exploration and Visualization:

– Understand key principles of designing and creating effective data visualizations.
– Create graphs and draw sound conclusions from graphs.
– Compute and interpret summary statistics.

Modeling and Inference:

– Understand and be able to explain key probabilistic and inferential concepts such as sampling, variability, random variables, distributions, confidence, and statistical significance.
– Determine the correct model for a given problem and set of data.
– Appropriately apply and draw inferences from a statistical model, including quantifying and interpreting the uncertainty in model estimates.
– Consider the ethical implications of various modeling practices.

Communicating Findings:

– Develop a reproducible workflow using R Markdown documents.
– Interpret and communicate results of statistical analyses effectively for both a statistical and a non-statistical audience.
– Responsibly discuss the practical significance of findings and clearly communicate the limitations of analyses.
– Foster an open attitude towards other ways of examining and thinking about data.

---

2. Learning Environment

Class Meetings

This course will be conducted according to a flipped classroom model, in which class time is devoted to working through interactive labs rather than direct instruction. The success of a flipped classroom model depends on students taking responsibility for their learning and coming to class prepared to engage with the material. Since this model requires additional time commitment outside of class than a traditional lecture-based course, the Mon/Wed 1:30 - 2:45pm class sessions will be run as two equivalent sessions and students are only expected to attend one session per week (after Week 1), with the extra 75 minutes per week being used to prepare for class.

During class, students will work on problems in groups with classmates, with the teaching team available to provide support. This course format is designed to foster a collaborative learning environment. We understand that being asked to work on problems during class can feel intimidating, especially compared to simply listening to a traditional lecture. However, we believe that active practice and grappling with challenging material is how learning happens. Additionally, there is some evidence that students in classes with active learning perform better on exams than students in classes with traditional lecturing[1] and that active learning can decrease the achievement gap for underrepresented minorities and first generation college students.[2]

The in-class labs are not submitted for a grade and the solutions will be provided after class.

Attending class is an important component of the process of learning statistics as we will use class time to actively engage with the material. While class attendance is not factored into your engagement grade, attendance is still strongly recommended. If you miss a class session, you are responsible for learning the missed material; we recommend that you make a one-on-one office hours appointment to check in with any questions. If you will miss a significant number of classes, you should reach out to Julie with your plan for keeping up with the material.

Sections

Section is required and you should have enrolled for a section during registration.

- If you are not currently enrolled in a section, make sure to enroll as soon as possible. If none of the open section times work in your schedule, we recommend taking STAT 100 in a future semester.

Section attendance is factored into your engagement grade. You are allowed to miss up to two section meetings.

- If you need to attend a different section time in a particular week, please contact your section TF and the TF of the section you plan to attend to notify them.

- If you need to miss a meeting (i.e., won't be able to attend any section time that week), contact your section TF and specify the reason why you are missing section and your plan for keeping up with the material.

---

[1] Freeman, et al (2014)
[2] Theobald et al (2020)

Office Hours

The teaching team will host many drop-in office hours each week. Feel free to come to office hours to work on problem sets even if you do not yet have specific questions. All questions are welcome, including conceptual questions, specific questions about problems, questions about pursuing statistics and data science, etc.

One-on-one office hours by appointment (15 minutes per appointment) are also offered. These are great resources for asking about a concept in more detail, checking in about course progress, or chatting about study strategies. Note that these should **not** be used for problem set help unless you are catching up on assignments. To allow for everyone to have an opportunity to access one-on-one help, please sign up for at most 30 minutes per week.

Class Community

We will be using Slack as an enhanced discussion forum to build our sense of class community. Each section will have its own private channel and there will also be public channels for class announcements, asking questions, and sharing useful resources.

The #q-and-a channel on Slack will be a major resource for asking questions about material and problem set questions. Please post questions to this channel rather than privately messaging the teaching team so that other students can also view the question and answers, as well as potentially contribute answers.

The teaching team will moderate Slack regularly. Please be mindful of the honor code while using Slack; for example, avoid sharing written answers word-for-word. Detailed guidelines for using Slack will be provided on the course site.

Course Climate

We expect everyone in this class to strive to foster a learning environment that is equitable, inclusive, and welcoming. If you experience any barriers to learning, please bring your concerns to any member of the teaching team or a College administrator.

All members of the class will agree to abide by the following community norms:

> We pledge to make participation in our community a harassment-free experience for everyone, regardless of age, body size, visible or invisible disability, ethnicity, sex characteristics, gender identity and expression, level of experience, education, socio-economic status, nationality, personal appearance, race, religion, or sexual identity and orientation.

> We pledge to act and interact in ways that contribute to an open, welcoming, diverse, inclusive, and healthy community.

> *This Code of Conduct is adapted from the Contributor Covenant, version 2.0.*

3. Course Materials

Readings

All readings are from free, open-source texts. The links are posted on the course site. Readings will come from the following books:

– *Introductory Statistics for the Life and Biomedical Sciences*, by Vu and Harrington
– *ModernDive*, by Ismay and Kim
– *Introduction to Modern Statistics*, by Çetinkaya-Rundel and Hardin

In addition to textbook readings, students will also be introduced to news articles, published literature, journal editorials, etc. on relevant statistical topics.

Computing

The course will use the statistical language R via R Studio, which is freely accessible with the cloud-based interface Posit Cloud (https://posit.cloud/). This allows all computing to be done within a web browser with internet access.

R and *RStudio* are also freely available for all common operating systems and instructions for downloading R, *RStudio*, and LaTeX are available on the course site.

We will use computers extensively. Plan on bringing a computer to class meetings and section meetings. If this is an issue, please reach out to Julie.

Work done in any programming language other than R will not receive credit.

4. Course Grading

Assessments

Your course grade will be based on the following key components:

– Problem Sets (30%)
– Midterm Exam (25%) and Final Exam (30%)
– Quizzes (10%)
– Engagement (5%)

Problem Sets (30%)

Generally, there will be a problem set due Tuesday by 11:59pm. Check the course schedule for the specific deadlines, as the regular schedule may need to be adjusted to account for holidays and the midterm exam.

Solutions to the problem sets will not be posted; be sure to review the individualized feedback and ask clarification questions about any incorrect answers.

In order to be accommodating of everyone's personal situation, we will adhere to the following polices for deadline flexibility:

– The lowest problem set score will be dropped from the grading. This includes scores of 0, such as for a problem set that was not submitted.

– Each student has four extension days that can be used as needed, no questions asked. Using 1 extension day means that the submission deadline is extended by 24 hours. To use an extension day, submit the Problem Set Extension Form. If it turns out that you did not need to use an extension (or used fewer extension days than requested), let your section TF know so they can update the record to accurately reflect how many extension days you have used.

– We recommend that you use no more than one extension day on a problem set in order to avoid falling behind, but recognize that there may be circumstances in which using more than one extension day may be necessary.

– Once the extension days are used, no further extensions will be granted. In the event of serious illness or unexpected circumstances that may require additional flexibility, please contact Julie.

– Problem sets submitted late without an extension are subject to a late penalty.

Tips for success:

– Don't wait until the day the problem set is due to work on it! They are not structured to be completed in a single day, but rather to encourage consistent practice. The experience of working through a problem set will go more smoothly if you start early and give yourself enough time to take breaks, ask questions, discuss with classmates, and return to problems that you were stuck on.

– Ask for help. You are not expected to complete the problem set "in isolation". Instead, take advantage of the course support structure to ask questions and collaborate with peers.

– Develop a consistent routine for working on the problem set. Most weeks, there will be time during section to start the problem set. Identify specific times in your schedule that you can regularly spend working on the problem set; if these times coincide with office hours, then that is even better.

Exams (55%)

There will be a midterm exam (25%) and a final exam (30%).

Both the midterm and final exams consist of a 3-hour written component (completed in-person) and a 10-minute oral component (completed over Zoom).

– For the midterm exam, the written component will take place on October 16, tentatively from 6:00 - 9:00 pm (there will be no class on the day of the exam) and the oral component will take place virtually on October 17 or 18.

– The final exam will take place during final exam period at the time scheduled by the Registrar's Office. The oral component will take place virtually; exact dates still TBD.

– The midterm exam should only be missed due to extenuating circumstances. There will be no makeup midterm exam; the final exam will comprise 55% of the course grade for students who miss the midterm. Contact Julie if you anticipate needing to miss the midterm exam.

The written component is open-notes, open-book. The oral component is closed-notes, closed-book; students may only have their written component open during the oral exam.

No collaboration is allowed on exams.

Quizzes (10%)

Each week, you will take two short quizzes on Gradescope related to that week's material.

- The first quiz is due by noon on Mondays. This is a 'pre-quiz' meant to ensure that you have prepared for class.

- The second quiz is due by noon on Fridays. This is a 'post-quiz' meant to assess your understanding of that week's material after attending class.

- The lowest quiz grade will be dropped.

- No late quizzes will be accepted.

- The quizzes are open-note, open-book.

- Collaboration on the quizzes is allowed, in the sense that you may discuss the questions and answers with classmates and the teaching team. However, please be responsible when collaborating; i.e., we ask that you do not, for example, simply give a friend all the answers because they did not have time to prepare for class.

Engagement (5%)

- Engagement on Slack: Make at least two posts before October 17 and then at least two additional posts before the final exam date. Posts that count include asking a question about course content, answering someone else's question, posting a useful resource and why you found it helpful, and creating an example that illustrates a recent concept. Note that the Slack prompt-based discussion posts do not count toward this engagement requirement.

- Attend at least one office hour before October 17 and attend at least one additional office hour before the final exam. Students are not required to stay for the entire duration of the office hour to receive engagement credit. Come introduce yourself, ask a question about course content, talk about statistics outside of Stat 100, etc. To receive credit for your office hour visits, refer to the Google Forms linked on Canvas.

- You are expected to attend class and section. While class attendance is not factored into your engagement grade, attendance is still strongly recommended. Attendance at section is recorded.

Regrade Request Policy

If you believe a grader misapplied the rubric on one of your assessments, you can submit a regrade request within Gradescope. Here are a few important notes about our regrade policy:

- These must be submitted within a week of when the grade is released on Gradescope.

- In your request, make sure to clearly and succinctly state what error you believe occurred.

- We will regrade the whole problem, possibly the whole assessment.

- We will not accept arguments related to effort (i.e., that the grade should be increased because you worked hard on the assignment) or related to a disagreement with the distribution of points on the rubric (i.e., no regrade will be made to alter the number of points deducted for a mistake).

- Your grade may **increase, stay the same, or decrease**. If your grade decreases, we will not accept requests to revert back to the original grade.

- Please remember to be respectful and follow the STAT 100 Code of Conduct in your request.

---

5. Academic Integrity

We encourage you to collaborate with classmates while working on problem set questions, but you must be sure to understand a solution well enough to be able to work a similar problem on your own. Solutions must always be written in your own words; this also applies to any program code. Copying or paraphrasing someone else's solution is a violation of the Honor Code. You are allowed to use R functions not covered in STAT 100 but if your code deviates significantly from that taught in STAT 100, we may schedule a meeting to discuss your work and affirm the work is your own.

Solutions to problem sets from previous versions of the course may be available in various places online. Copying answers from solution sets is a violation of the Honor Code and any instances of doing so that we detect will be reported to the Honor Council.

The Harvard College Honor Code states:

> Members of the Harvard College community commit themselves to producing academic work of integrity – that is, work that adheres to the scholarly and intellectual standards of accurate attribution of sources, appropriate collection and use of data, and transparent acknowledgement of the contribution of others to their ideas, discoveries, interpretations, and conclusions. Cheating on exams or problem sets, plagiarizing or misrepresenting the ideas or language of someone else as one's own, falsifying data, or any other instance of academic dishonesty violates the standards of our community, as well as the standards of the wider world of learning and affairs.

More information about the Honor Code as well as resources for students may be found at `https://honor.fas.harvard.edu`.

Students who sell, post, publish, or distribute course materials without written permission, whether for the purpose of soliciting answers or otherwise, may be subject to disciplinary action, up to and including requirement to withdraw from the college.

---

6. Generative AI Policy

While generative artificial intelligence (GAI) tools such as ChatGPT are capable of generating code, analyzing data, and producing written summaries, this course is intended to help students develop their **own** abilities to write code, analyze data, and thoughtfully communicate results. Therefore, we expect that all work (including code, written work, oral assessments) that students submit will be their own. We specifically forbid the use of generative AI tools to answer assessment questions, unless the assignment specifically states that it is allowed. Violations of this policy will be considered academic misconduct.

The purpose of this policy is not to lessen student access for support but to ensure that students gain important skills. While we recognize that GAI tools can be powerful assistants, existing software has been shown to be error-prone in many cases, including misstating facts and even completely fabricating research results. We believe that responsibly and effectively using AI tools requires some base knowledge (such as applied domain knowledge or coding experience) in addition to informed skepticism and critical thinking.

Note that different classes at Harvard may implement different AI policies and students are responsible for conforming to course-specific expectations.

---

7. Accessibility

Harvard College is committed to working with all students. If you have a disability and would like to request accommodation for this reason, please contact the Accessible Education Office (https://aeo.fas.harvard.edu/). Advance notice and appropriate documentation are required for any accommodations.