

---

[MATH 118] INTRODUCTION TO STATISTICAL LEARNING  
FALL 2022

---

<b>SECTION:</b>	MATH 0118A	<b>PROFESSOR:</b>	Becky Tang (she/her)
<b>CLASS HOURS:</b>	MWF 13:45-14:35	<b>OFFICE:</b>	Warner 214
<b>ROOM:</b>	Warner 100	<b>E-MAIL:</b>	btang@middlebury.edu
<b>PREREQS:</b>	NA	<b>OFFICE HOURS:</b>	M 3:30pm-5:30pm F 10:30am-12:00pm

---


### COURSE DESCRIPTION


In this course students will gain exposure to the entire data science pipeline: forming a statistical question, collecting and cleaning data sets, performing exploratory data analyses, identifying appropriate statistical techniques, and communicating the results, all the while leaning heavily on open source computational tools, in particular the R statistical software language. We will focus on analyzing real, messy, and large data sets, requiring the use of advanced data manipulation/wrangling and data visualization packages. Students will be required to bring a laptop (owned or college-loaned) to class as many lectures will involve in-class computational activities.

### KEY LEARNING OUTCOMES

- ☐ Learn to explore, visualize, and analyze data in a reproducible and shareable manner
- ☐ Think critically about the utility of a visualization
- ☐ Gain experience in data wrangling, exploratory data analysis, predictive modeling, and data visualization
- ☐ Learn to effectively communicate results through written assignments and final project presentation
- ☐ Have fun!







### TEXTBOOKS AND COURSE MATERIALS

 **ONLINE RESOURCES:** Content (slides, homework, schedule) will be housed mainly on the course website: <https://math118-fall2022.github.io/website/>. I encourage you to bookmark this website.




 **TEXTBOOK:** There is no required textbook for this course. However, the following may be useful:

- *R for Data Science* by Golemund, Wickham: <https://r4ds.had.co.nz/>
- *OpenIntro Statistics 4th Edition* by Diez, Barr, Çetinkaya-Rundel: <https://www.openintro.org/book/os/>

## CLASS EXPECTATIONS

-  **You are expected to physically show up to class and actively participate**, conditional on classes being in-person. You are an integral part of the class community! Exceptions include previously-communicated illness or planned absence.
-  **Please arrive on time.** I expect everyone, myself included, to arrive on time and dedicate full attention during the class. In turn, I will do my best to always end class at the designated time.
-  **Use of laptops will be often be necessary during the lectures and labs.** However, please be respectful of both the professor and your peers. Please ensure that your laptop has sufficient battery for the duration of the class.
-  **Cell phones should be turned to silent.** I don't mind cell phones in class, but please silence them so as to not disrupt the class.
-  **Please ask questions!**
-  **I encourage discussion amongst yourselves**, especially for clarification or help! However, please be mindful of volume so that the conversations will not be disruptive to the class.

## RESOURCES

-  **Office hours.** This time is meant for you! Please come by to ask questions, chat with me, or work on homework. You should never worry about disturbing me during this time.
-  **One-on-one meetings.** If you would like to meet with me one-on-one, please send me an e-mail or approach me after class so we can schedule a time.
-  **Your peers.** I encourage students to work together and discuss material! However, unless the assignment explicitly states that it is to be completed as group work, the submitted material must be your own.

# COLLEGE POLICIES AND RESOURCES

## ACADEMIC INTEGRITY

As an academic community devoted to the life of the mind, Middlebury requires every student to reflect complete intellectual honesty in the preparation and submission of all academic work. Details of our Academic Honesty, Honor Code, and Related Disciplinary Policies are available in Middlebury's handbook.

## HONOR CODE PLEDGE

The Honor Code pledge reads as follows: "I have neither given nor received unauthorized aid on this assignment." It is the responsibility of the student to write out in full, adhere to, and sign the Honor Code pledge on all examinations, research papers, and laboratory reports. Faculty members reserve the right to require the signed Honor Code pledge on other kinds of academic work.

## DISABILITY ACCESS AND ACCOMMODATION

Students who have Letters of Accommodation in this class are encouraged to contact me as early in the semester as possible to ensure that such accommodations are implemented in a timely fashion. For those without Letters of Accommodation, assistance is available to eligible students through the Disability Resource Center (DRC). <https://www.middlebury.edu/office/disability-resource-center>. The DRC provides support for students with disabilities and facilitates the accommodations process by helping students understand the resources and options available and by helping faculty understand how to increase access and full participation in courses. DRC services are free to all students. Please contact ADA Coordinators Jodi Litchfield and Peter Ploegman of the DRC at [ada@middlebury.edu](mailto:ada@middlebury.edu) for more information. All discussions will remain confidential.

## CENTER FOR TEACHING, LEARNING, AND RESEARCH (CTRL)

The CTRL provides academic support for students in many specific content areas and in writing across the curriculum through both professional and peer tutors. The Center is also the place where students can find assistance in time management and study skills. These services are free to all students. [go.middlebury.edu/connect](https://go.middlebury.edu/connect)

## TYPES OF ASSIGNMENTS

- \* **Homework.** Assigned weekly and typically completed individually. In homework, you will apply what you've learned during lecture and lab to complete data analysis tasks. *Lowest score will be dropped.*
- \* **Application exercises.** Small assignments designed to practice using the statistical concepts and/or code discussed in lecture on short data analyses. Every lecture period will usually have an associated application exercise, with the intention that students will be able to start the exercise in class. Graded based on a good-faith effort has been made in attempting all parts. Successful on-time completion of at least 90% will result in full points. Deadlines for application exercises will be 11:59pm before the next lecture. For example, if the exercise is assigned on Friday, you have until Sunday 11:59pm to complete it. *Lowest two scores will be dropped.*
- \* **Midterm exam.** One take-home midterm exam designed as an opportunity to assess the knowledge and skills you've learned. The midterm exam will include small analysis and computational tasks related to the content discussed in lectures, application exercises, homework assignments
- \* **Final project.** The purpose of the project is to apply what you have learned throughout the semester to analyze an interesting data-based research question of your choosing. The final project will include a written component as well as an in-person class presentation. The presentations are scheduled to take place during the last week of classes, though there will be opportunity to revise the written portion up until the end of finals. **You must present the final project to pass the course.**
- \* **Sharing a visualization.** Each one of you will be asked to sign-up to briefly share with the class a data visualization that you find interesting, meaningful, and worth discussing. The visualization can come from any source, including the web, a magazine, a book, etc. You should upload the visualization onto Canvas *at least 48 hours* before the class in which you are presenting.
- \* **Participation:** I expect you to be present and active learners in this course! Throughout the semester, there will be various small assignments that will count towards participation (e.g. reflection exercises, giving constructive feedback to each other).

## GRADING

- \* Unless otherwise noted, assignments should be turned in on Canvas.
- \* Late work for homework will always be considered within one week of the original due date. Unless otherwise stated, the late policy is as follows: for every 24-hour period the assignment is late, 10% from the maximum possible grade will be deducted.
- \* I will do my best to return assignments within one week of submission.
- \* Regrade requests: I do allow regrade requests, which must be submitted within one week of when the assignment is returned. These should be entered as a "Comment" on the corresponding assignment in Canvas, after which you must notify me via e-mail that you have a regrade request. Keep in mind that regrade requests do not guarantee points back, and may result in a lower score if I see that I incorrectly marked something as correct.

Component	Percentage
Homework	30%
Application Exercises	15%
Midterm 1	20%
Final Project	25%
Sharing visualization	5%
Participation	5%

## TENTATIVE COURSE CONTENT

\*\* Homework will generally be released on Wednesday after class, and due the following Tuesday at 11:59pm \*\*

Week	Date	Topics
1	9/12	M - Welcome! - HW0 assigned W - Meet the toolkit + installation F - Data visualization + HW 01 assigned
2	9/19	M - Visualizing different types of data T - HW 01 due at 11:59pm W - Tidy data and data wrangling + HW 02 assigned F - Tidy data and data wrangling (cont.)
3	9/26	M - Joining data from multiple sources T - HW 02 due at 11:59pm W - Making pretty tables + Reshaping + HW 03 assigned F - Introduction to probability
4	10/3	M - Conditional Probability T - HW 03 due at 11:59pm W - Dates and strings + HW 04 assigned F - Confounding and Simpson's Paradox
5	10/10	M - Data Ethics - Introduce final project T - HW 04 due at 11:59pm W - Estimation via bootstrapping F - <b>No class</b> . Please complete mid-semester feedback form.
6	10/17	M - <b>Midterm</b> W - Bootstrapping cont. + HW 05 assigned F - Simulation-based testing: Part 1
7	10/24	M - Simulation-based testing: Part 1 (cont.) Discuss Midterm T - HW 05 due at 11:59pm W - Simulation-based testing: Part 2 + HW 06 assigned F - Project work day
8	10/31	M - The Central Limit Theorem T - HW 06 due at 11:59pm

		W - Inference with the CLT + HW 07 assigned F - Two-sample inference
<b>9</b>	11/7	M - Introducing linear models T - HW 07 due at 11:59pm W - Linear models (cont.) + HW 08 assigned F - Multiple linear regression (MLR)
<b>10</b>	11/14	M - MLR: Inference + conditions T - HW 08 due at 11:59pm W - Categorical data + HW 09 assigned F - ANOVA and/or Project work day - Project proposal due
<b>-</b>	11/21	<b>Thanksgiving break</b>
<b>11</b>	11/28	M - Project work day T - HW 09 due at 11:59pm W - Spatial data and visualization F - Text analysis - Project draft due
<b>12</b>	12/5	M + W - Case studies R - Final project due 12:00pm F - <b>Project presentations</b>
	12/12	M - <b>Project presentations (cont.)</b>

**NOTE:** The above dates and content may be modified due to the requirements of the class. Also, the indicated dates of course content and homework may be moved backward or forward depending on class progress and my conference travel.

**Midterm and final project dates will not change.**