[MATH 118] INTRODUCTION TO STATISTICAL LEARNING
FALL 2022

SECTION:	MATH 0118A	**PROFESSOR:**	Becky Tang
CLASS HOURS:	MWF 13:45-14:35	**OFFICE:**	Warner 214
ROOM:	Warner 100	**E-MAIL:**	btang@middlebury.edu
PREREQS:	NA	**OFFICE HOURS:**	T 15:00-17:00
			F 10:30-12:00

COURSE DESCRIPTION

Intro to data science and statistical thinking. Learn to explore, visualize,and analyze data to understand natural phenomena, investigate patterns, model outcomes,and make predictions, and do so in a reproducible and shareable manner. Gain experience in data wrangling and munging, exploratory data analysis, predictive modeling, data visualization, and effectively communicating results. Work on problems and case studies inspired by and based on real-world questions and data. The course will introduce and focus on the R statistical computing language.

KEY LEARNING OUTCOMES

❏ Learn to explore, visualize, and analyze data in a reproducible and shareable manner

❏ 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:** Lecture notes and slides will be provided to students throughout the semester.

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

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

CLASS POLICY AND RULES OF CONDUCT

- ⚧ **Students are expected to physically show up to class and actively participate**, conditional on classes being in-person. Exceptions include previously-communicated illness or planned absence.
- ☉ **Please arrive on time.** I expect students 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 fellow students. 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 between students**, especially for clarification or help! However, please be mindful of volume so that the conversations will not be disruptive to the class.

GRADING

- ✳ Unless otherwise noted, assignments should be turned in on Canvas.
- ✳ Late work 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.
- ✳ 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.

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

Tentative Course Content

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

Week	Date	Topics
1	9/12	M - Welcome!
		W - Meet the toolkit + installation
		F - Data visualization
2	9/19	M - Visualizing different types of data
		W - Tidy data and data wrangling
		F - Tidy data and data wrangling (cont.)
3	9/26	M - Joining data from multiple sources
		W - Making pretty tables + Reshaping
		F - Introduction to probability
4	10/3	M - Conditional Probability
		W - **Lab:** activities covering discussed topics.or lubridate or spatial
		F - Confounding and Simpson's Paradox
5	10/10	M - Data Ethics
		W - Midterm I
		F - No class. Please complete mid-semester feedback form.
6	10/17	M - Estimation via bootstrapping
		W - Bootstrapping cont. + Discuss Midterm 1
		F - Simulation-based testing: Part 1
7	10/24	M - Simulation-based testing: Parts 1 and 2
		W - Simulation-based testing: Part 2
		F - The Central Limit Theorem
8	10/31	M -Inference with the CLT
		W - Two-sample inference
		F - More than two groups
9	11/7	M - Introducing linear models
		W - Linear models (cont.)
		F - Multiple linear regression (MLR)
10	11/14	M - MLR: Inference + conditions
		W - **Lab:** activities covering discussed topics.
		F - Categorical data
-	11/21	Thanksgiving break

11	11/28	M - Project work day
		W - Spatial data and visualization
		F - Strings
12	12/5	M + W - Work on final project
		F - Project presentations
	12/12	M - Project presentations (cont.)

NOTE: The above actual dates may be modified due to the requirements of the class. Also, the indicated dates may be moved backward or forward depending on class progress and my conference travel. **Exact dates and instructions will be announced on course webpage**.