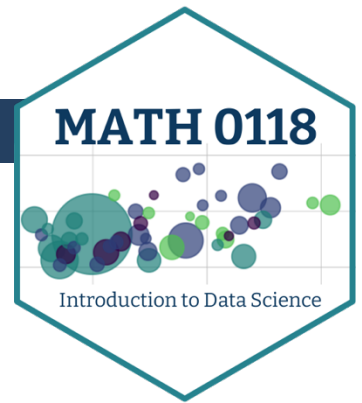


Fall 2023

STAT 0118: Introduction to Data Science



Instructor: Emily Malcolm-White (she/her)
Please address me as Professor Emily
emalcolmwhite@middlebury.edu

Office Hours: Mondays, Tuesdays, and Wednesdays
in WNS 215 (Professor Emily's Office)
Time TBD

CLASS MEETING FORMAT

Class: Section A: 8:15am-9:30am on Tuesday, Thursday in Warner 100

Class meetings will involve a combination of lecture, live coding, and active learning for the first 30-40 minutes, followed by group activities and homework for the last 30-40 minutes. The instructor and a teaching assistant will be available for support.

Students are required to bring a laptop to class each day. See more details below.

MATERIALS

- Students must have access to a laptop with R (<http://www.r-project.org>) and RStudio (<http://www.rstudio.com>) installed. Both are free and installation instructions are available on Canvas. We will walk through installation together on the first day of class.
 - Laptops with pre-installed R/RStudio can be borrowed from the Davis Family Library. This is a good option for those of you without access to a laptop or those of you who may be having a short-term issue with your laptop. See your instructor or the front desk of the Davis Library for more info.
- The website for this course is on Middlebury Canvas. Please check Canvas often for assignments, deadlines, resources, and announcements.
- There is no textbook requirement for this course.

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.

COURSE CALENDAR

This is a tentative course calendar and is subject to change throughout the semester.

	Topic	Date	To-Do
A	Introductions, Installation, Intro R, RMarkdown, File Paths	Tues, Sept 12	Homework A Due: Fri, Sept 15
B	Naming Objects, Wrangling data with dplyr: <code>filter</code> , <code>select</code> , <code>arrange</code> , Combining functions with the pipe <code>%>%</code>	Thurs, Sept 14	Homework B Due: Tues, Sept 19
C	Aggregating data with <code>summarize</code> , <code>group_by()</code>	Tues, Sept 19	Homework C Due: Fri, Sept 22
D	Making pretty tables with <code>kable</code>	Thurs, Sept 21	Homework D Due: Tues, Sept 26
E	Making plots with <code>ggplot2</code> : barplots and scatterplots	Tues, Sept 26	Homework E Due: Fri, Sept 29
F	Making plots with <code>ggplot2</code> : line graphs, histograms & boxplots	Thurs, Sept 28	Homework F Due: Tues, Oct 3
G		Tues, Oct 3	Homework G Due: Fri, Oct 6
H	Working with categorical data using <code>forcats</code>	Thurs, Oct 5	Homework H Due: Tues, Oct 10
			Midterm Project Due: Fri, Oct 13
I	Joining tables with <code>dplyr</code>	Tues, Oct 16	Homework I Due: Fri, Oct 18
J	Reshaping data with <code>tidyr</code> : <code>pivot_longer</code> , <code>pivot_wider</code> , <code>separate</code> , <code>unite</code>	Thurs, Oct 17	Homework J Due: Tues, Oct 24
K	Maps – basics and <code>leaflet</code> maps	Tues, Oct 24	Homework K Due: Fri, Oct 27
L	Maps – plotting points and <code>leaflet</code>	Thurs, Oct 26	Homework L Due: Tues, Oct 31
M	Simple Linear Regression, pretty output with <code>broom</code>	Tues, Oct 31	Homework M Due: Fri, Nov 3
N	Multiple Regression	Thurs, Nov 2	Homework N Due: Tues, Nov 7
O	Webscraping using <code>rvest</code> : tables	Tues, Nov 7	Homework O Due: Fri, Nov 10
P	Webscraping using <code>rvest</code> : text	Thurs, Nov 9	Homework P Due: Tues, Nov 14
Q	Working with dates and times using <code>lubridate</code>	Tues, Nov 14	Homework Q Due: Fri, Nov 17
R	Working with text as data using <code>stringr</code>	Thurs, Nov 16	Homework R Due: Tues, Nov 21
			Final Project Due: Fri, Dec 15

COURSE ASSESSMENT

The following components will determine your final grade. They will be weighted as follows:

- | | | |
|------------|------------------------|---|
| 50% | Homework | <i>The lowest two homework grades will be dropped.</i>
You are encouraged to discuss the homework problems with your peers, the teaching assistant, and the instructor. Check the course website regularly for homework assignments and deadlines. Typically, there will be a homework assignment for each class and homework will be due on Fridays and Tuesdays by 11:59pm. Assignments should be submitted on Canvas. |
| 20% | Midterm Project | You will analyze a provided dataset. You are encouraged to work in pairs, but you can also choose to work independently. More |
| 30% | Final Project | You will analyze a dataset of your choice. You are encouraged to work in pairs, but you can also choose to work independently. |

You will be given a numeric grade (out of 100) at the end of the course. It will be translated into a letter grade using the following scale.

F	D-	D	D+	C-	C	C+	B-	B	B+	A-	A
0 -59.9	60 -62.9	63 -66.9	67 -69.9	70 -72.9	73 -76.9	77 -79.9	80 -82.9	83 -86.9	87 -89.9	90 -92.9	93 - 100

LATE POLICY

Consistent engagement with the course material is essential for your learning and academic growth. However, I understand that unforeseen circumstances may occasionally arise.

- When you become aware that you won't be able to make a deadline, please email Professor Emily to let her know which homework you won't be submitting on time and what date you anticipate the homework will be done. You do not need to disclose *why* you are missing the deadline. **So long as you communicate to me before the deadline, no late penalty will be applied.**
- If you do not communicate with me before the deadline, late submissions will be subject to a penalty of 20% per day.

ACADEMIC INTEGRITY

You are bound by Middlebury College's honor code, including its policies on plagiarism and cheating. Violation of these rules is ground for failure.

I encourage you to seek help in understanding the concepts and problems in your assignments from various sources, including peers, instructors, peer tutors, class notes, textbooks, and online sources. To maintain academic integrity and properly credit your resources, it is important to:

1. Write up your own solutions to assignments independently and in your own words
2. Include an Acknowledgements section at the beginning or end of your assignment. In this section, give credit to people who have helped you, acknowledge useful websites, and list any other resources you used.

DIVERSITY & INCLUSION STATEMENT

It is my intent that students from all backgrounds and perspectives be well-served by this course, and that students' learning needs be addressed both inside and outside of class. I value the diversity that students bring to this class and view it as a resource and strength. I strive to present materials and activities that are inclusive and respectful of all backgrounds, including but not limited to: gender identity, sexual orientation, disability, age, socioeconomic status, ethnicity, race, religion, culture, and perspective. Your feedback about how to improve this course for yourself and other students is encouraged and appreciated. Feel free to reach out to me with suggestions.

NAME/PRONOUN POLICY

I respect your right to be addressed by your preferred name or gender pronoun. Please advise me of this preference early in the semester so that I can update my records accordingly. If anything changes throughout the semester, please let me know.

STATEMENT ON RELIGIOUS HOLIDAYS

Campus policies regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students, who because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. Please have a look through the course calendar and let me know as soon as possible if you anticipate such conflicts so that we can agree upon alternative arrangements.

ACCOMMODATIONS FOR DISABILITIES

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). Please contact ADA Coordinators Jodi Litchfield, Peter Ploegman or Dierdre Kelly of the DRC at ada@middlebury.edu for more information. All discussions will remain confidential.

STUDENT ATHLETES

Please have a look through the course calendar and let me know as soon as possible if you anticipate any conflicts with your athletic schedule so that we can agree upon alternative arrangements. Whenever possible, I would appreciate notice at least a week ahead of time of any explained absences.