MATH/COSC 3570 Section 101: Introduction to Data Science

Instructor: Dr. Cheng-Han Yu

Spring 2025

E-mail: cheng-han.yu@marquette.edu
Office Hours: TuTh 4:50 - 5:50 PM, Wed 12 - 1 PM
Office: Cudahy Hall 353
Teaching Assistant (TA): Qishi Zhan
TA office hours: TBD

Web: math3570-s25.github.io/website/
Class Hours: TuTh 2 - 3:15 PM
Class Room: Lalumiere Language Hall 232
TA E-mail: qishi.zhan@marquette.edu

Course Objectives

MATH/COSC 3570 introduces main aspects of doing a practical data science project, from importing data to deploying what is learned from data. Topics include basic data science tools, R/Python programming, statistical machine learning methods, and data science documenting.

Prerequisites

In Bulletin, the prerequisites include COSC 1010 (Intro Programming) and MATH 4720 (Intro Stats). The course will also assume facility with using the internet and a personal computer/laptop. The course involves coding in R and Python using Posit Cloud, a cloud version of RStudio integrated development environment (IDE).

Textbook

No textbook is required for this course. Course materials are mainly Dr. Yu's slides. Below are some good references.

- *(**r4ds**) R for Data Science (2e)* by Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund.
- *(**tmwr**) Tidy Modeling with R* by Max Kuhn and Julia Silge.
- *(**py4da**) Python for Data Analysis* by Wes McKinney.
- (**IS**) *Introduction to Statistics* by Cheng-Han Yu. (Good resource for brushing up your basic probability, statistics and simple linear regression knowledge.)

Office Hours

- TuTh 4:50 - 5:50 PM and Wed 12 - 1 PM in CU 353.

TA Office Hours

- To be determined.

E-mail Policy

- I will attempt to reply your email quickly, at least **within 24 hours**.
- **Expect a reply on Monday if you send a question during weekends**. If you do not receive a response from me within two days, re-send your question/comment in case there was a "mix-up" with email communication (Hope this won't happen!).
- Please start your e-mail subject line with **[math3570]** or **[cosc3570]** followed by a clear description of your question. See an example in Figure 1.
- Email etiquette is important. Please read this article to learn more about email etiquette.

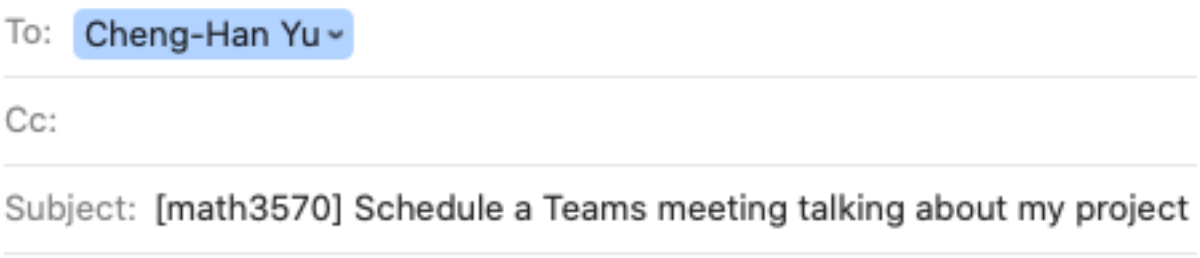

Figure 1: Email Subject Line Example

- I am more than happy to answer your questions about this course or statistics in general. However, due to time constraint, I may choose **NOT** to respond to students' e-mail if
  1. The student could answer his/her own inquiry by reading the syllabus or information on D2L.
  2. The student is asking for an extra credit opportunity. The answer is "no".
  3. The student is requesting an extension on homework. The answer is "no".
  4. The student is asking for a grade to be raised for no legitimate reason. The answer is "no".
  5. The student is sending an email with no etiquette.

Course Websites

- All course materials and news are posted on the course websites https://math3570-s25.github.io/website/.
- Course grades are saved in **D2L** > **Assessments** > **Grades**.

Grading Policy

- Your grade is from the following categories and distribution
  - **25%** In-class lab exercises and participation.
  - **30%** Homework
  - **20%** Midterm mini project
  - **25%** Final project competition
  - *Extra credit opportunities*
- Every student has to participate (in-person) in the final presentation to pass the course.
- You will **NOT** be allowed any extra credit projects/homework/exam to compensate for a poor grade. Everyone is given the same opportunity to do well in this class. I may use class participation to make grade adjustments at the end of the semester.
- The final grade is based on the grade-percentage conversion Table 1 on the next page. $[x, y)$ means greater than or equal to x and less than y. For example, 94.1 is in $[94, 100]$ and the grade is A and 93.8 is in $[90, 94)$ and the grade is A-.

Table 1: Grade-Percentage Conversion

Grade	Percentage
A	[94, 100]
A-	[90, 94)
B+	[87, 90)
B	[84, 87)
B-	[80, 84)
C+	[77, 80)

Grade	Percentage
C	[74, 77)
C-	[70, 74)
D+	[65, 70)
D	[60, 65)
F	[0, 60)

Lab exercises

- There are several **in-class** lab exercises, which are graded as **complete/incomplete** and used as evidence of attendance and class participation.
- You are allowed to have **two** incomplete lab exercises without penalty. Beyond that, **2%** grade percentage will be taken off for each missing/incomplete exercise.

Homework

- The homework assignments are *individual*. You should submit your own work.
- **You may not directly share or discuss answers/code with anyone other than the instructor**. But you are welcome to discuss the problems in general and ask for advice.
- Homework will be assigned through GitHub. You need to clone/pull the homework repo into Posit Cloud and work on the Quarto file in the repo. A step-by-step guide will be discussed in class before homework is assigned.
- You will have *at least one week* to complete your assignment.
- **No make-up homework** for any reason unless you have excused absences.
- If you miss a homework assignment due to excused absence, the homework percentage will be added to your final project. If you miss more than one assignment, only one assignment percentage can be added to the final project percentage. You get 0% for the other assignment.

Midterm mini project

- You will be team up to do the midterm mini project.
- More details about the mini project presentation will be released later.

Final project competition

- You will be team up to do the final project. Your project can be in either of the following categories:
  1. **Data analysis** using statistical models or machine learning algorithms
  2. Introduce a **R or Python package** not learned in class, including **live demo**
  3. Introduce a **data science tool (visualization, computing, etc)** not learned in class, including **live demo**
  4. Introduce a **programming language** not learned in class for doing data science, including **live demo**, Julia, SQL, MATLAB, SAS for example.
  5. **Web development**: Shiny website or dashboard, including **live demo**
- Details about the project will be provided as the course progresses. You must complete the final project and be in class to present it in order to pass this course.
- The final project presentation is on **Thursday, May 1, 2 PM** and **Monday, May 5, 10:30 AM - 12:30 PM**.

Generative AI and Sharing/Reusing Code Policy

Generative AI

- You are responsible for the content of all work submitted for this course. You may use generative AI tools such as ChatGPT or DALL-E to generate a first draft of text for your assignments, provided that this use is documented and cited.

Sharing/Reusing Code

- Unless explicitly stated otherwise, you may make use of any online resources, but you must **explicitly cite** where you obtained any code you directly use or use as inspiration in your solutions.
- Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism, regardless of source.

Academic Integrity

- This course expects all students to follow University and College statements on academic integrity.
- **Honor Pledge and Honor Code**: *I recognize the importance of personal integrity in all aspects of life and work. I commit myself to truthfulness, honor, and responsibility, by which I earn the respect of others. I support the development of good character, and commit myself to uphold the highest standards of academic integrity as an important aspect of personal integrity. My commitment obliges me to conduct myself according to the Marquette University Honor Code.*

Accommodation

If you need to request accommodations, or modify existing accommodations that address disability-related needs, please contact Disability Service.

Tentative Course Schedule

Week 1, 1/13(Mon) - 1/19(Sun): Syllabus, Introduction to Data Science, Posit Cloud

Week 2, 1/20(Mon) - 1/26(Sun): Git/GitHub, Quarto Documents

- **Drop deadline 1/21 11:59 PM**

Week 3, 1/27(Mon) - 2/2(Sun): Basic R/Python Syntax

Week 4, 2/3(Mon) - 2/9(Sun): R/Python packages for Data Science

Week 5, 2/10(Mon) - 2/16(Sun): Data Importing, Data Visualization

Week 6, 2/17(Mon) - 2/23(Sun): Data Visualization

Week 7, 2/24(Mon) - 3/2(Sun): Interactive Data Visualization, Data Wrangling

Week 8, 3/3(Mon) - 3/9(Sun): Data Wrangling

Week 9, 3/10(Mon) - 3/16(Sun):

- **No class this week (Spring Break)**
- **Midterm grade submission 3/11 by noon**

Week 10, 3/17(Mon) - 3/23(Sun): Probability and Statistics

Week 11, 3/24(Mon) - 3/30(Sun): Linear Regression

Week 12, 3/31(Mon) - 4/6(Sun): Logistic Regression

Week 13, 4/7(Mon) - 4/13(Sun): K-Nearest Neighbors

- **Withdrawl Deadline 4/11**

Week 14, 4/14(Mon) - 4/20(Sun): Principal Component Analysis

- **No class on 4/17 (Easter Break)**

Week 15, 4/21(Mon) - 4/27(Sun): K-Means Clustering

Week 16, 4/28(Mon) - 5/4(Sun): Dashboards and Websites

- **Project Presentation I: Thursday, 5/1, 2 - 3:15 PM**

Week 17, 5/5(Mon) - 5/11(Sun):

- **Project Presentation II: Monday, 5/5 10:30 AM - 12:30 PM**
- **Final grade submission 5/13 by noon**

*** I reserve the right to make changes to the syllabus.**