

MATH/COSC 3570 Section 101: Introduction to Data Science

Instructor: Dr. Cheng-Han Yu

Spring 2024

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): No TA :(

Web: math3570-s24.github.io/website/

Class Hours: TuTh 2 - 3:15 PM

Class Room: Olin Engineering 198

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 required textbook.

Office Hours

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

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.

To: **Cheng-Han Yu** ▾

Cc:

Subject: **[math3570] Schedule a Teams meeting talking about my project**

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-s24.github.io/website/>.
- Course grades are saved in **D2L > Assessments > Grades**.

Grading Policy

- **40%** In-class lab exercises and participation
- **30%** Homework
- **30%** Final project competition
- *Extra credit opportunities*
- Every student has to participate (in-person) in the final presentation in order 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+	$[77, 80)$
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.
- **No make-up lab exercises** for any reason.

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 got [COVID](#) or [excused absence](#).
- If you miss a homework assignment due to COVID-19 symptoms, exposure, diagnosis, quarantine, and/or isolation, or you have an excused absence as defined in [Attendance in Academic Regulations](#), 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.

Group Projects

- 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. **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 **Monday, 5/6 10:30 AM - 12:30 PM**.

Sharing/Reusing Code Policy

- 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.*
- You know what I am talking about. Yes, **DO NOT CHEAT**.

Accommodation

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

Attendance and COVID-19

- It is **your** responsibility as a Marquette University student to protect the health and safety of our community in this course. The University has mandated vaccination for all students for the academic year 2022-2023. COVID Cheq will still be used, but those fully vaccinated that provide documentation will no longer receive the daily symptom screening survey. Instead, they will receive a daily email providing a green check. You may be required to show your automated green check to show clearance for entry into campus buildings. If you are experiencing symptoms related to COVID-19, you must follow the link in the automated email to report symptoms.
- Visit [What to do if you are exposed to COVID-19 or test positive](#) website for university guidelines on the best course of action.
- Visit guidance on Spring 2023 [Class attendance, withdrawal, and grading](#)
 - Students are responsible for contacting instructors **prior** to the missed class session to indicate absence and the need to make up classwork/assignments.
 - Students requesting make up classwork/assignments are required to provide the **COVID Cheq “stop sign”** to confirm inability to attend class.

Tentative Course Schedule

Week 1, 1/15 - 1/21: Syllabus, Introduction to Data Science, Posit Cloud

Week 2, 1/22 - 1/28: Git/GitHub, Quarto Documents

- **Drop deadline 1/24 11:59 PM**

Week 3, 1/29 - 2/4: Basic R/Python Syntax

Week 4, 2/5 - 2/11: R/Python packages for Data Science

Week 5, 2/12 - 2/18: Data Importing, Data Visualization

Week 6, 2/19 - 2/25: Data Visualization

Week 7, 2/26 - 3/3: Interactive Data Visualization, Data Wrangling

Week 8, 3/4 - 3/10: Data Wrangling

Week 9, 3/11 - 3/17:

- **No class this week (Spring Break)**
- **Midterm grade submission 3/12 by noon**

Week 10, 3/18 - 3/24: Probability and Statistics

Week 11, 3/25 - 3/31: Linear Regression

- **No class on 3/28 (Easter Break)**

Week 12, 4/1 - 4/7: Logistic Regression

Week 13, 4/8 - 4/14: K-Nearest Neighbors

- **Withdrawl Deadline 4/12**

Week 14, 4/15 - 4/21: Principal Component Analysis

Week 15, 4/22 - 4/28: K-Means Clustering

Week 16, 4/29 - 5/5: Dashboards and Websites

Week 17, 5/6 - 5/12:

- **Project Presentation: Monday, 5/6 10:30 AM - 12:30 PM**
- **Final grade submission 5/14 by noon**

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