MATH/COSC 3570 Section 101: Introduction to Data Science

Spring 2026

Instructor: Dr. Cheng-Han Yu

Email: cheng-han.yu@marquette.edu
Website: math3570-s26.github.io/website/
Office Hours: TuTh 3:30–4:30PM, We 2-3PM Cudahy Hall 353
Class Hours: TuTh 2:00–3:15PM
TA: No TA :(
Class Room: Cudahy Hall 120

---

1 Course Objectives

MATH/COSC 3570 introduces the main aspects of doing a practical data science project, from importing and cleaning data to building models and communicating results. The course uses a modern workflow with R and Python, Git and GitHub, Quarto documents, and cloud based computing in Posit Cloud.

In this AI aware edition of the course, students learn to use generative AI tools such as ChatGPT as assistants for coding, debugging, documenting, and idea generation, while remaining responsible for all analytical decisions, validation, and interpretation.

2 Prerequisites

COSC 1010 Introduction to Programming and MATH 4720 Introduction to Statistics. The course assumes that students are comfortable with a personal computer or laptop and with using the internet. The course involves coding in R and Python using Posit Cloud, a cloud version of RStudio integrated development environment (IDE).

2.1 Technology Requirements

Students will need

- A laptop that can access the internet reliably.
- A Posit Cloud account for R and Python coding in the browser.
- A GitHub account to access course repositories and submit homework.
- Access to a generative AI tool such as ChatGPT, Gemini, or Copilot.

3 Textbooks

- There is **no required textbook** for this course. Course materials will consist of Dr. Yu's slides, lab notebooks, and online references.
- Below are useful references for deeper study.

- **(r4ds)** *R for Data Science (2e)* by Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Grolemund.
- **(tmr)** *Tidy Modeling with R* by Max Kuhn and Julia Silge.
- **(py4da)** *Python for Data Analysis (3e)* 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.)

4 Course Management

- All course materials are posted on our course website https://math3570-s26.github.io/website/.
- Course grades are saved and managed in **D2L** > **Assessments** > **Grades**.

5 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.

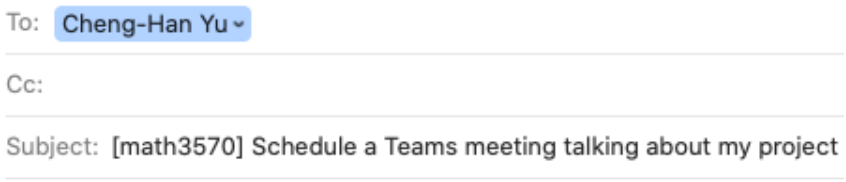


Figure 1: Email Subject Line Example

- Email etiquette is important. Please read this article to learn more about email etiquette.
- 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 the course website or 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.

6 Grading Policy

- Your grade is from the following categories and distribution
  - **25%** In-class lab activities
  - **10%** In-class AI activities
  - **45%** Mini projects
  - **20%** Final project competition
- You have to participate (in-person) in the final presentation to pass the course.
- There will be **no** individualized extra credit homework/project/exercise/exam to compensate for a poor grade. All students have the same opportunities to succeed. Class participation may be used for 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	[83, 87)
B-	[80, 83)
C+	[77, 80)
C	[73, 77)
C-	[70, 73)
D+	[65, 70)
D	[60, 65)
F	[0, 60)

6.1 Lab activities

- **In-class** labs are short activities that you complete during class. They are usually graded as **complete or incomplete**.
- In class labs are used as evidence of both attendance and participation.
- You may have up to two missing or incomplete labs without penalty.
- For each additional missing or incomplete lab, **2%** points will be deducted from your final course percentage.

6.2 AI activities

- AI activities are short presentations during class. They are usually graded as **complete or incomplete**.
- AI activities are used as evidence of both attendance and participation.

- Groups take turn to present what they learn from GenAI about data science.

6.3 Mini projects

- You will work as a team on **3** mini projects.
- This project will focus on a subset of the course content up to that point, for example data wrangling and visualization or a simple predictive model.
- The mini projects will include
  - A documented GitHub repository.
  - A Quarto report describing your question, data, methods, and findings.
  - A short in-class group presentation.
  - An AI usage documents that shows how you used generative AI and how you verified any AI generated content.
- More detailed instructions and rubric will be provided later in the semester.

6.4 Final project competition

- The final project is a team based competition. Each team will choose one of the following directions or a related idea approved by the instructor.
  1. A **data analysis** project using statistical models or machine learning algorithms.
  2. A **tutorial** style project that introduces an **R or Python package** not covered in class, including a **live demo**.
  3. A project that introduces and demonstrates a **(AI) data science tool** for visualization, computing, or workflow that was not used in class, including a **live demo**.
  4. A project that introduces another **programming language** for data science, for example Julia, SQL, MATLAB, or SAS, with a **live demo**.
  5. A **web development** project such as a Shiny app or dashboard with a **live demo**.
- You must complete the final project and be present for the presentation to pass this course.
- Each project must include an AI usage appendix that documents prompts, outputs, and team decisions about what to trust, modify, or reject.
- The final project presentation is on **Monday, 5/4 10:30 AM - 12:30 PM**.

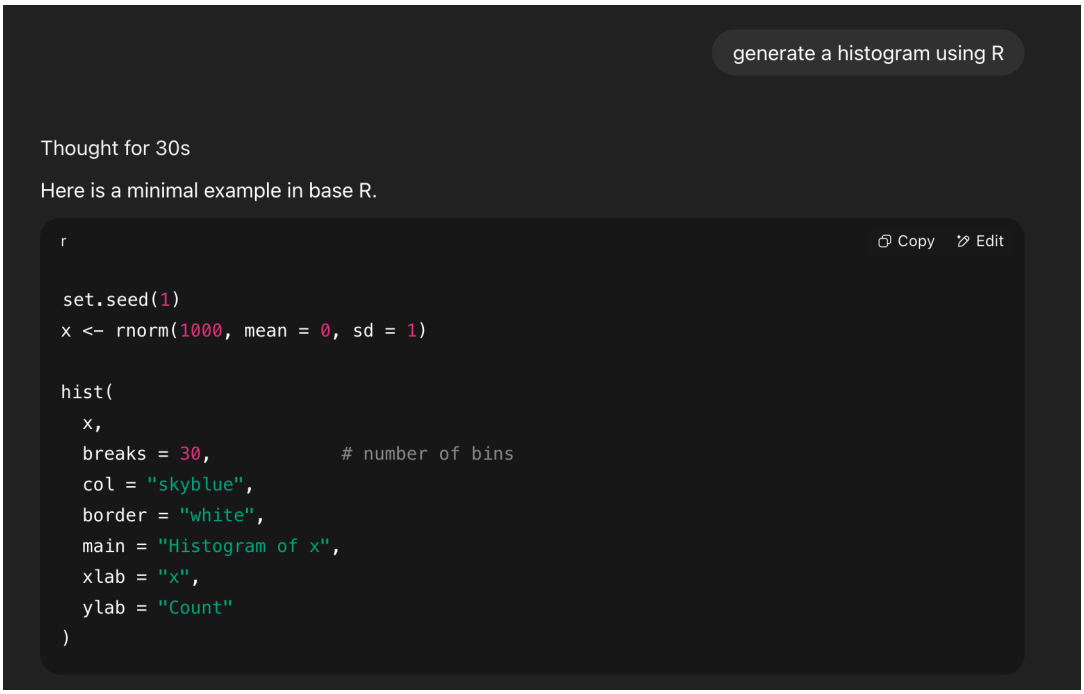
7 Generative Artificial Intelligence (GenAI) Policy

- *You are responsible for the content of all work submitted for this course.*
- For any of your work, you are allowed to use generative AI tools such as ChatGPT to generate a draft of your work.
- To avoid any academic integrity issue, you **must cite your AI usage, or screenshot your entire AI usage history**. Check the followings on how to cite it.
  - How to cite ChatGPT
  - How to Cite AI-Generated Content

- If you use GenAI, please include the followings in your submitted work:
  - **How I used AI (prompts or questions)**
  - **Generated output (screenshot or copy-paste excerpt)**
  - **How I used the output**

Here is an example.

- **How I used AI (prompts and questions)**
  - *I asked ChatGPT to generate a histogram using R.*
- **Generated output (screenshot or copy-paste excerpt)**



- **How I used the output**
  - *I reviewed the suggestions, but I did not use the exact code. Instead, I change the code format and breaks value to 50.*

8 Academic Integrity

- Watch this video about academic integrity issue of using GenAI.
- 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.*

9 Accommodation

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

10 Tentative Course Schedule

Week 1, 1/12 - 1/18: Syllabus, Introduction to data science, Posit Cloud, AI in data science workflows

Week 2, 1/19 - 1/25: Git, GitHub, and Quarto

- **Drop deadline 1/20 11:59 PM**

Week 3, 1/26 - 2/1: Basic R and Python syntax

Week 4, 2/2 - 2/8: R and Python packages for data science

Week 5, 2/9 - 2/15: Data importing, Data visualization

Week 6, 2/16 - 2/22: Data visualization

Week 7, 2/23 - 3/1: Data wrangling

- **Mini Project 1 Presentation on 2/26: Exploratory Data Storytelling**

Week 8, 3/2 - 3/8: Data wrangling

Week 9, 3/9 - 3/15:

- **No class this week (Spring Break)**
- **Midterm grade submission 3/10 by noon**

Week 10, 3/16 - 3/22: Probabilistic and statistical simulation

Week 11, 3/23 - 3/29: Linear regression

- **Mini Project 2 Presentation on 3/26: Data Wrangling and Pipelines**

Week 12, 3/30 - 4/5: Logistic regression

- **No class on 4/2 (Easter Break)**

Week 13, 4/6 - 4/12: K-Nearest neighbors

- **Withdrawl Deadline 4/10**

Week 14, 4/13 - 4/19: Decision trees

- **Mini Project 3 Presentation on 4/16: Supervised Learning**

Week 15, 4/20 - 4/26: Principal component analysis

Week 16, 4/27 - 5/3: K-Means clustering

Week 17, 5/4 - 5/10:

- **Final Project Presentation: Monday, 5/4 10:30 AM - 12:30 PM**
- **Final grade submission 5/12 by noon**

*** Dr. Yu reserves the right to make changes to the syllabus.**