

CMSE 830: Foundations of Data Science

Table of Contents

- [CMSE 830: Foundations of Data Science](#)
 - [Table of Contents](#)
 - [Who-When-Where](#)
 - [Office Hours and Locations:](#)
 - [Course Description](#)
 - [Topics Covered](#)
 - [Course Structure](#)
 - [Required Materials for class](#)
 - [Textbook](#)
 - [Computer](#)
 - [Course Activities](#)
 - [Grading information](#)
 - [Grading scale](#)
 - [Python setup](#)
 - [Other Important Information](#)

Who-When-Where

Instructor:

Dr. Luciano G. Silvestri

Fixed-term Assistant Professor

Computational Mathematics, Science and Engineering (CMSE)

Engineering Building 2502

silves28@msu.edu

Teaching Assistant:

Saraswathi Baskaran

MSDS Student

Computational Mathematics, Science and Engineering (CMSE)

baskar12@msu.edu

Days and time:

Tuesday and Thursday;

8:30 A.M. - 9:50 A.M.

Location:

Wonders Hall C211

Office Hours and Locations:

Instructor: Office hours are only in-person in the instructor's office (EB 2502). Office hours are only upon request and will start the week of Sep 2 (Labour Day) and can be arranged by emailing the instructor. An

email will be sent to you concerning the location and time of office hours. If you are not available in these hours, please contact the instructor.

Teaching Assistant:

Tue 12:00PM - 2:00PM in-person in ENG 2508

Thu 12:00PM - 2:00PM Zoom, link TBD

Information Technology: Any course personnel can help you with questions pertaining to the course material, including in-class and homework assignments. Technical questions, including issues with Python and/or Jupyter, should be first directed to your professor, but specific issues with [JupyterHub] (<https://jupyterhub.egr.msu.edu>) can be sent to support@egr.msu.edu.

If you have issues relating to class administration, including missed classes, illness, VISA issues, or school-sponsored activities please contact the instructor.

Credits:3

Prerequisites: At least three semesters of calculus, including multivariable calculus

Basic knowledge of statistics

Basic knowledge of linear algebra

Equivalent of CMSE 802 or very strong Python coding skills

Fully-operational, always-charged laptop with Python installed

Course Description

This course introduces the foundations of data science (DS) beginning with data types, cleaning, munging and visualization. Mathematical foundations are given with an emphasis on linear algebra. Specialized data types are covered from a wide array of applications, including time series, spatial, graphical and image data.

The course is project driven with two projects, one in the basics of DS and the capstone combining DS and basic machine learning. The projects involve designing and building web apps for DS applications. Your projects and operational web apps, built using Streamlit, will be placed at GitHub. There is another document that describes the projects in much more detail: see D2L and Teams for all relevant documents.

An emphasis is placed on visualization throughout, with a week devoted to this topic.

A second emphasis, distinct from learning basic computational skills, is on thinking in data science. We will use the textbook Thinking Clearly with Data to guide our thinking on thinking.

The flow of the class is based on a mixed lecture and flipped classroom format. The basic pattern will be lectures on Tuesday that lead into group in-class coding projects on Thursday. There may be pre-class work before the in-class to ensure that you are prepared, mainly in the form of homework assignments.

There are no exams in this course, only homeworks, in-class projects and projects. 30% of your grade is based on these two projects.

We will use Teams (for communication) and D2L (for assignments). Do not use email. You have received an invitation to Teams via an email sent from D2L.

Topics Covered

Here is a tentative list of topics, not necessarily in order, covered in the class. The list is tentative because it will change during the semester based on students' understanding and interests.

- Data Wrangling and Munging (e.g., File Types, Imputation, Feature Selection/Engineering)
- Visualization
- Advanced Matplotlib
- Libraries for DS and ML (e.g., Yellowbrick)
- Chart Types and Uses
- Effective Communication
- Exploratory Data Analysis
- Imputing and Missingness
- Linear Algebra
- Data Types: Graph, Spatiotemporal, Image, Text
- Ethics, Security and Privacy

Course Structure

The course has a very structured format:

- one topic per week
- a lecture on the first class of the week (usually Tuesday)
- group-based projects on the second class of the week (usually Thursday)
- one homework per week on that week's topic
- no exams
- two projects (midterm and final)

Unfortunately, this structure is disrupted two times during the fall semester: Fall Break (20-21 Oct) and Thanksgiving (28-29 Nov). In these shorter weeks, we will default to in-class projects (no lecture those weeks).

Required Materials for class

Textbook

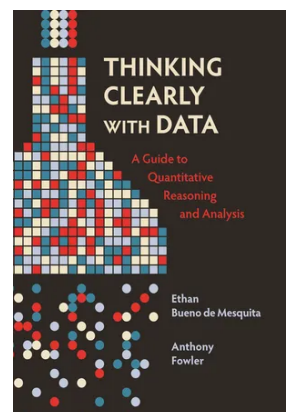
There are two textbooks for the course. The "technical" textbook is in the form of PDF notes that will be available in both D2L and Teams. These notes form the mathematical and computational core of the course.

Data science, however, is not about blindly calling Python libraries. It is easy to make quite serious conceptual mistakes. So that we all end up being the best data scientists in the world, we will also use the textbook *Thinking Clearly with Data*: A Guide to Quantitative Reasoning and Analysis, by Ethan Bueno de Mesquita and Anthony Fowler [ISBN 9780691214368]

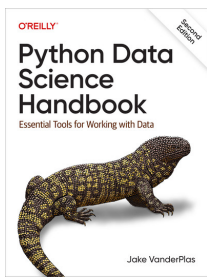
([https://press.princeton.edu/books/hardcover/9780691214368/thinking-clearly-with-data?](https://press.princeton.edu/books/hardcover/9780691214368/thinking-clearly-with-data?srsltid=AfmBOooQu3GaiqGTTq_ZjB92tNfH4xJsUjH2IR6_3IW6SIF1kC9bJZ38)

[srsltid=AfmBOooQu3GaiqGTTq_ZjB92tNfH4xJsUjH2IR6_3IW6SIF1kC9bJZ38](https://press.princeton.edu/books/hardcover/9780691214368/thinking-clearly-with-data?srsltid=AfmBOooQu3GaiqGTTq_ZjB92tNfH4xJsUjH2IR6_3IW6SIF1kC9bJZ38)).

This book will help us *think* about data science. It is worth skimming the entire textbook as soon as you can so that you know where we are going.

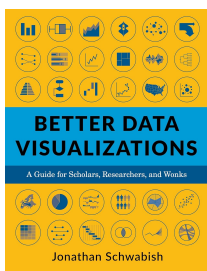
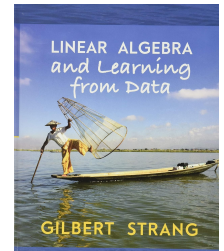


There are three *optional* textbooks that I recommend; which of them is useful to you depends on your background.



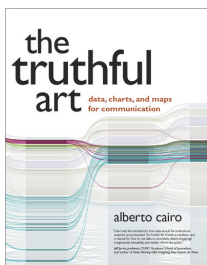
If your Python background is weak, you should consider this book. Geared toward data science, it might be useful to you even if you have proficient Python skills. We will cover some of this in class, but not nearly the entire content of this book. As the title suggests, it is a *handbook*, good for looking up many powerful data science techniques. (Be sure to get the latest edition.)

Linear algebra, combined with statistics, is at the core of data science and machine learning; having a solid background will greatly aid learning the concepts. We will spend a couple of weeks on linear algebra. However, if you want a more solid background with a data science slant, you should consider this text.



Finally, communicating your results is a key aspect of data science. We will spend a bit more than a week on visualization, but you may want a more complete reference. These are quite good and relatively inexpensive options.

Two suggestions are **Better Visualization** by Jonathan Schwabish and **The Truthful Art** by Laver Cairo.



Computer

In-class programming assignments are a critical part of the learning process in this course. To that end, **you are expected to bring your laptop and its power cord every day.**

You are expected to join the Microsoft Teams for CMSE 492 (see the front page of the course website for the appropriate link for this semester). Details for doing so are outlined in later sections.

Details regarding the software needed for this course are provided in the Software Setup Guide in D2L.

Course Activities

Class attendance and participation: Active class participation (led both by the instructor and by students) is critical to the success of this course. As such, you are expected to attend class every week, bring the required materials (most importantly, your computer and power cord) and to actively participate in the in-class discussion.

Pre-class assignments (PCAs): We will assign short assignments that are due prior to class. The purpose of these assignments is to practice material introduced in the lecture and prepare so that we can focus on experimentation and implementation in class. These assignments will typically consist of one or more short videos or reading assignments and related questions or problems and will be due before class via the course's Desire2Learn page. The deadline for each pre-class assignment is indicated on the course's Desire2Learn page. Completing these assignments will be critical for your success in this course. These assignments will not be graded.

In-class programming assignments (ICAs): The in-class assignments (ICAs) are done in small groups, so they are turned in with your group at the end of the class, *unless specific instructions are given*. Occasionally an ICA is long and we don't want you to stress; we want you to learn, and we might add a few hours to complete the ICA outside of class. In-class assignments are mainly graded based on effort and participation. This means that active participation and attendance are part of the grade. You, and your entire group, will receive zero credit for any project turned in after that time. If you do not show up for the in-class project, you will receive partial credit (1 point for submission) but cannot earn the full 2 points without attending class to participate in the group work. ICAs will be turned in at the end of the class session (or otherwise specified on D2L) via the course's Desire2Learn page. The assignments might be graded by a computer system or AI as such it is very important to follow the instructions in the assignments. Not following the instructions will result in a reduction of the grade. The percentage reduction at the instructor's discretion. **No late assignments will be accepted.**

Homework: You will have periodic homework assignments that are meant to help you work towards your learning. You can expect that they will require a week's worth of out-of-class effort, so you are encouraged to start your assignments as early as possible. The assignments might be graded by a computer system or AI as such it is very important to follow the instructions in the assignments. Not following the instructions will result in a reduction of the grade. The percentage reduction at the instructor's discretion.

Homework assignments that are submitted late will be accepted **for up to two days beyond the due date** (*i.e.*, 48 hours past the original deadline). **If the assignment is submitted within 24 hours of the original deadline, there is a 10% penalty.** This applies even if the assignment is 1 minute late. Similarly, **if it is submitted in the 24-48 hour window, a 20% deduction is applied.** Again, **after the 48th hour, the assignment will no longer be accepted.**

Semester Project: In place of a final exam, this course will have two semester projects. The learning you would get from exams is included through comprehensive homeworks that continuously reinforce the content. A lengthy description of the two projects is given as a separate document in D2L and Teams.

Excused Absences: Up to 2 excused absences are allowed during the semester. Any absences beyond that will impact your grade. If you have a legitimate reason to miss class (such as your job, graduate school visits, conferences or medical school interviews) you must arrange this ahead of time to be excused from class. Look at your calendar for the fall semester and think through dates you have obligations and let us know immediately.

Grading information

Attendance is mandatory. Because of the flipped nature of a project-based coding course, it is not possible to be successful if you are not present. See details below.

There are a variety of course activities, with a percentage of total grade listed. More detailed descriptions of each activity can be found elsewhere in the syllabus.

Activity	Grade Percentage
Lecture attendance	10%
In-class assignments (ICAs)	30%
Homework assignments	30%
Semester project	30%
Total	100%

Attendance and Participation Scoring:

- **In-class assignments (ICAs):** Each ICA day is worth 2 points - 1 point for attendance and 1 point for submitting a non-empty notebook with your group
- **Lecture attendance:** Each lecture day is worth 2 points - 1 point for attendance and 1 point for participation in class quiz questions

Grading scale

Grade	Percentage
4.0	≥ 90%
3.5	≥ 85%
3.0	≥ 80%
2.5	≥ 75%
2.0	≥ 70%
1.5	≥ 65%
1.0	≥ 60%
0.0	< 60%

Grades will not be curved - your grade is based on *your own effort and progress*, not on competition with your classmates.

The assignments associated with your project throughout the semester are graded in the same way as homework, with the exception of the final product, which obviously cannot be done late.

Excused Absences: Up to 2 excused absences are allowed during the semester. Any absences beyond that could impact your grade. If you have a legitimate reason to miss class (such as your job, graduate school visits, conferences or medical school interviews) you must arrange this *ahead of time* to be excused from class. Look at your calendar for the fall semester and think through dates you have obligations and let us know immediately.

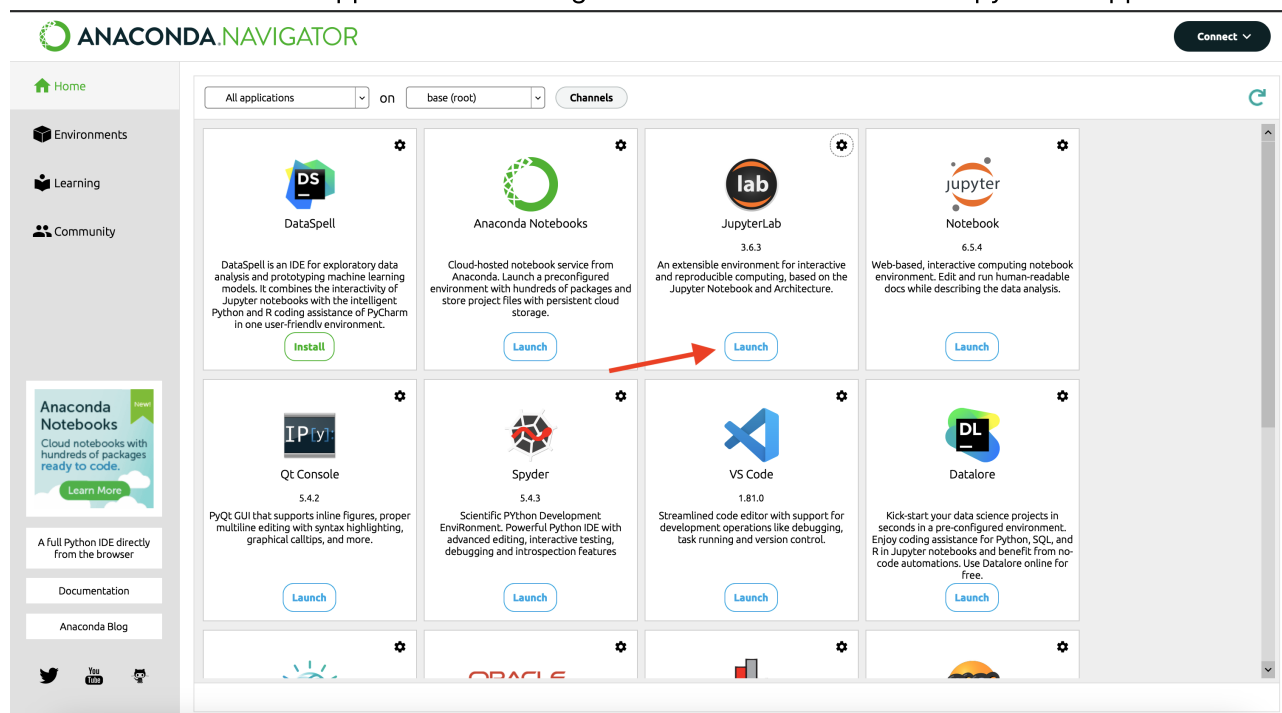
Python setup

In this class, you will do a lot of coding, and all of it in Python. We will exclusively use Python 3.13.x. If you have used Python before and you already have it on your laptop, upgrade to this version before class starts. You will be at a serious disadvantage if you are not using the same version as your classmates because of the team coding projects every week. We will use Jupyter notebooks (in Jupyter lab) for all assignments, so be sure that you are able to create and save notebooks.

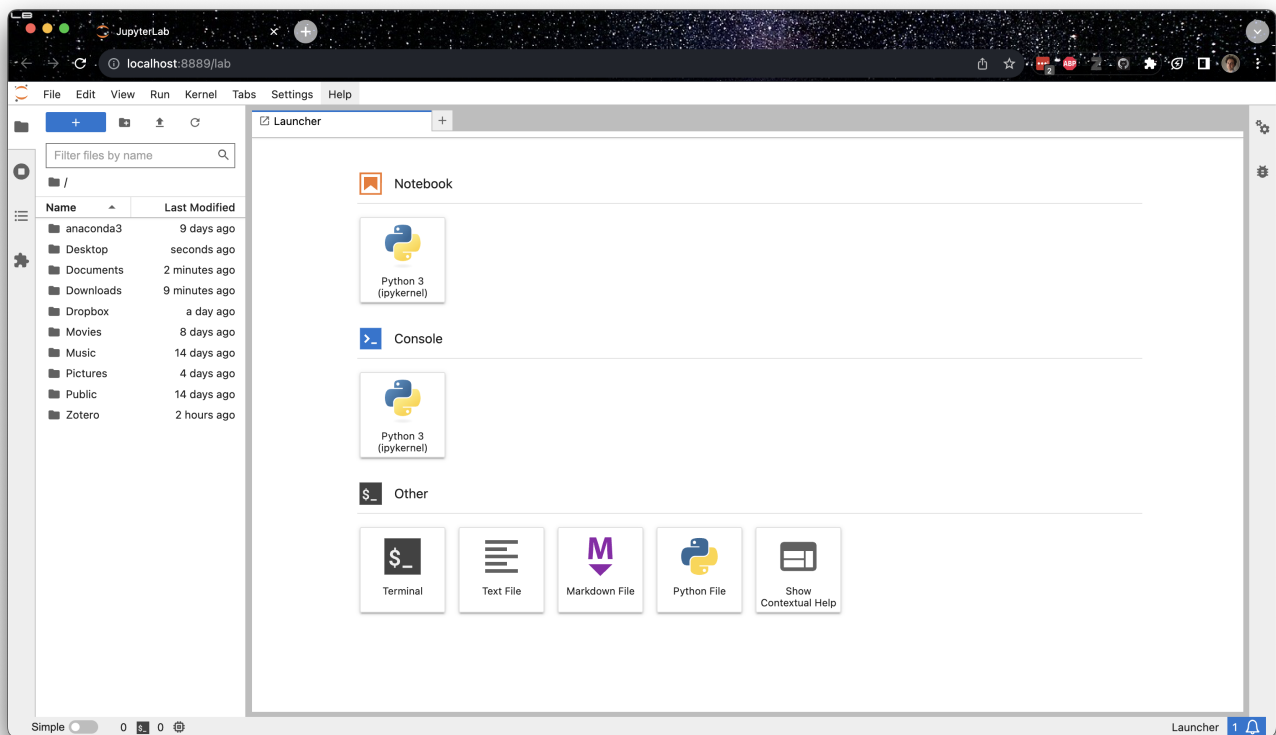
If this is your first time working with python, it is recommended that you use the Anaconda distribution. **If you don't have Anaconda installed already**, follow the instructions below to get the Anaconda distribution of Python installed on your computer. Even if you already have a version of Python installed on your machine, we encourage you to go through this installation process as the assignments will assume that you are working with the same versions of the Anaconda Python packages that the instructors are using. If you already specifically have Anaconda installed, we encourage you to update all of the Python packages (you may need to look up how to do this).

Instructions for downloading Anaconda:

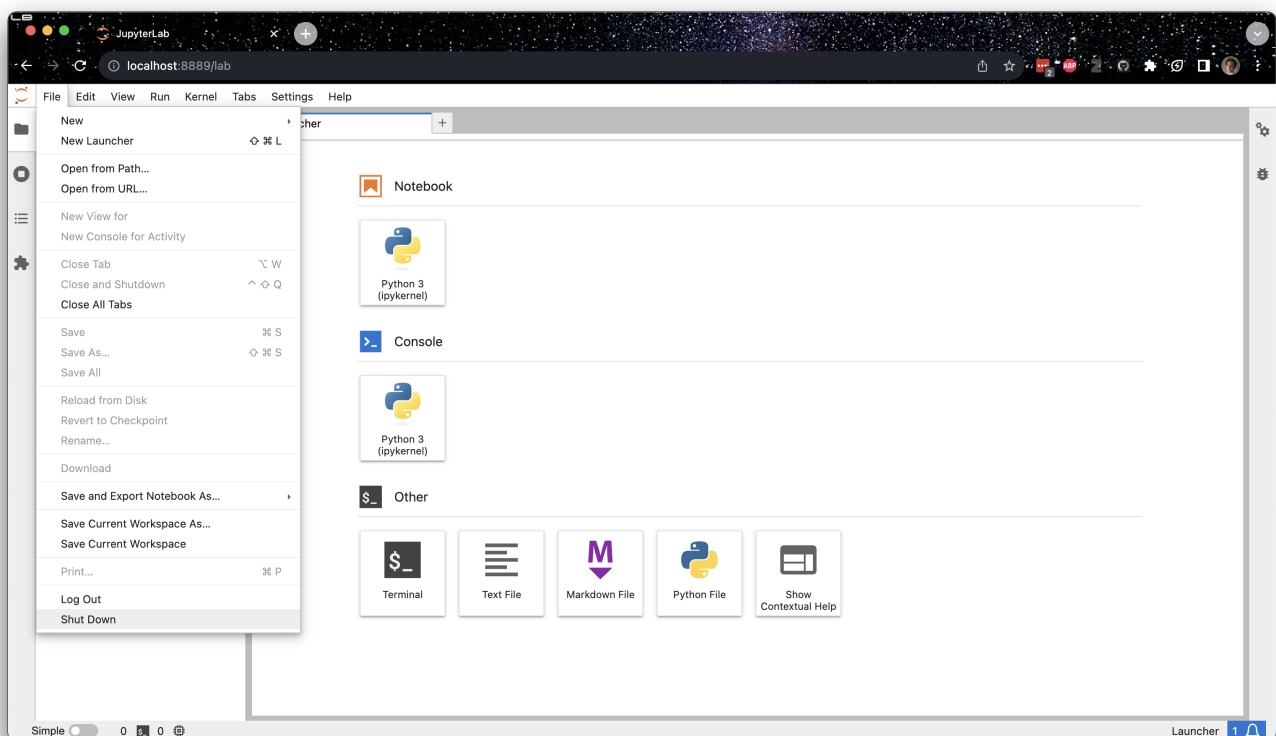
1. Go to the Anaconda Download webpage: <https://www.anaconda.com/download>
2. In the box "Free Download" click "Get Started" or "Skip registration" (at the bottom in small font).
3. Under "Distribution Installers" choose your system and download the files.
Mac Users: make sure to download the correct file. If you have a M1, M2, M* laptop choose the (Apple Silicon) file.
4. Download the Python 3.13 version (64 bit recommended).
5. Follow the directions at the bottom of the page to install Python on your specific operating system
6. Once installed launch the app Anaconda-Navigator and within it launch the JupyterLab app.



7. If everything goes correctly, a browser window should open up with the Jupyter interface running. If things don't work, don't worry, we will help you get started.



8. You can close this window by clicking on **File** in the top left corner of the window and then select **Shut Down** as shown in the figure.



If everything goes correctly, a browser window should open up with the Jupyter interface running. If things don't work, don't worry, we will help you get started.

Other Important Information

Course website and calendar: This course uses a Desire2Learn page for course organization, which can be found at <http://d2l.msu.edu>. Accompanying course information, including the official course calendar, can be found at this website. All assignments will be handed in via Desire2Learn. Consult the class website for instructions, but use Teams to find the latest information.

Email: At times, we will send out important course information via email. This email is sent to your MSU email address (the one that ends in "@msu.edu"). **You are responsible for all information sent out to your University email account, and for checking this account on a regular (daily) basis.**

Class discussion: We will be using [Teams](#) as our means of communicating about course content as the semester progresses. We believe that this will provide an excellent avenue to have discussions not only with course instructors, TAs, but also between you and your fellow classmates.

The Teams channels will be the place to go for any questions about assignments in the course. We encourage you to help out other classmates when you can!

In order to ensure that Teams is an appropriately used tool that does not become overly time-consuming for the course instructors, TA, we have a list of rules for how we expect you to use Teams. They are:

1. Before you ask a question, be sure to check the main help channel and other section channels to see if the question has already been answered.
2. The Teams group is primarily for you, the students, so help each other.
3. The Instructor will monitor the channels, but will defer to the students to work through things. They will only enter a conversation if students are going down the wrong path and/or there are too few other students involved. However, **you should not expect that the Instructor will always be available**. The Instructor will spend a limited amount of time "logged in" to Teams and we ask that you be respectful of their time.
4. Teams is meant to be used to help you when you are stuck with a minor issue. If you are having major issues or trouble understanding the concept, go to office hours or help room hours as they are meant for more in-depth discussions of course content.
5. Course instructors will rarely check Teams, only to examine progress. While they may offer help, do not rely on it. Instructors will not respond to the same student twice within a 30 minute time interval.
6. Only in rare cases should you contact an instructor through a private channel. But, if you are struggling, feel free to use this option.
7. Do not post your solutions to out-of-class assignments directly into Teams unless prompted by an instructor.
8. Be courteous to everyone on Teams. Students who are being rude or who are excessively posting might be banned from posting on the course Teams channel.

Class attendance: This class is heavily based on material presented and worked on in class, and it is critical that you attend and participate fully every week! Therefore, class attendance is absolutely mandatory. An unexcused absence will result in zero points for that day.

Excused Absences: Up to 2 excused absences are allowed during the semester. Any absences beyond that *will impact* your grade. If you have a legitimate reason to miss class (such as your job, graduate school visits, conferences or medical school interviews) you must arrange this *ahead of time* to be excused from class. Look at your calendar for the fall semester and think through dates you have obligations and let us know immediately.

Use of Artificial Intelligence Tools: Please see the separate document in Teams and D2L on this topic.

Classroom behavior: Respectful and responsible behavior is expected at all times, which includes not interrupting other students, turning your cell phone off, refraining from non-course-related use of electronic devices, and not using offensive or demeaning language in our discussions. Flagrant or repeated violations of this expectation may result in ejection from the classroom, grade-related penalties, and/or involvement of the university Ombudsperson. In particular, behaviors that could be considered discriminatory or harassing, or unwanted sexual attention, will not be tolerated and will be immediately reported to the appropriate MSU office (which may include the MSU Police Department).

MSU welcomes a full spectrum of experiences, viewpoints, and intellectual approaches because they enrich the conversation, even as they challenge us to think differently and grow. However, we believe that expressions and actions that demean individuals or groups comprise the environment for intellectual growth and undermine the social fabric on which the community is based.

Importantly, keep in mind that we are building an international data science community. You may stay connected to your classmates for the rest of your career. Meet them, learn them, befriend them. If you don't know someone in the class, introduce yourself and find out some interesting facts about them.

Academic honesty: Intellectual integrity is the foundation of the scientific enterprise. In all instances, you must do your own work and give proper credit to all sources that you use in your papers and oral presentations - any instance of submitting another person's work, ideas, or wording as your own counts as plagiarism. This includes failing to cite any direct quotations in your essays, research paper, class debate, or written presentation. The MSU College of Natural Science adheres to the policies of academic honesty as specified in the General Student Regulations 1.0, Protection of Scholarship and Grades, and in the all-University statement on Integrity of Scholarship and Grades, which are included in Spartan Life: Student Handbook and Resource Guide. Students who plagiarize will receive a 0.0 in the course. In addition, University policy requires that any cheating offense, regardless of the magnitude of the infraction or punishment decided upon by the professor, be reported immediately to the dean of the student's college.

It is important to note that **plagiarism in the context of this course includes, but is not limited to, directly copying another student's solutions to in-class or homework problems; copying materials from online sources, textbooks, or other reference materials *without citing those references in your source code or documentation*, or having somebody else do your in-class work or homework on your behalf.** Any work that is done in collaboration with other students should state this explicitly, and their names, as well as yours, should be listed clearly. When collaborating with other students, you should still be coding/writing your own solutions to the assignments and should limit your collaboration to conceptual discussions about how one might go about solving the problems, not sharing exact solutions.

More broadly, we ask that students adhere to the Spartan Code of Honor academic pledge, as written by the Associated Students of Michigan State University (ASMSU):

"As a Spartan, I will strive to uphold values of the highest ethical standard. I will practice honesty in my work, foster honesty in my peers, and take pride in knowing that honor is worth more than grades. I will carry these values beyond my time as a student at Michigan State University, continuing the endeavor to build personal integrity in all that I do."

Accommodations: If you have a university-documented learning difficulty or require other accommodations, please provide the instructor with your VISA as soon as possible and speak with the instructor about how they can assist you in your learning. If you do not have a VISA but have been documented with a learning difficulty or other problems for which you may still require an accommodation, please contact MSU's Resource Center for People with Disabilities (355-9642) in order to acquire current documentation.