COSC/MATH 3570: Introduction to Data Science

Fall 2018

Instructor:	Shion Guha	**Office Hours:**	Wed 2-5 pm
Email:	shion.guha@marquette.edu	**Office:**	318 Cudahy Hall

Description: This course is a first introduction to the new and rapidly evolving field of data science. We will understand the philosophy and basic concepts of doing data science through an active learning process. The student will learn how to use popular programming and data analysis platforms to clean, organize and analyze data from a modern machine learning perspective. Students are expected to read and practice specific lessons before coming to class. In class emphasis will be placed on practical interpretation, inference, visualization and navigating biases.

Prerequisites: math 1450 (calc 1) and cosc 1010 (intro programming); math 4710/4720 (intro statistics) or equivalent to be taken as pre or co-requisite. You will write code everyday in class so please plan accordingly.

Class: TR, 3:30-4:45 pm, 001 Cudahy Hall

Github: https://github.com/shionguha/cosc3570-introdatascience-fa18

Piazza: https://piazza.com/class/jl9ts6hpe6i5ss

Books: There are no **required** textbooks for this class. Let's all save some money! All recommended books, readings, datasets and code fragments will be posted on the course website as we progress through the class. All software used will be free and open source. We will use Python in this course.

Objectives: This course is a first introduction to modern data science for undergraduates. We will chiefly focus on an expansive, outward looking understanding of data science specifically featuring a rigorous introduction to data wrangling and a survey of different computational methods for analyzing complex datasets. At the end of the course, a successful student should be able to:

- import, wrangle and clean data using the Python stack,

- compute descriptive and summary statistics to understand data,

- understand and implement theories of data visualization,

- analyze, predict and make rigorous inferences about data from linear models.

Timeline:

- aug 28-30: introduction to the course; policies and expectations; python grokking

- sep 4-6: the data science process; asking research questions

- sep 11-13: data wrangling; data shaping and molding

- sep 18-20: data cleaning through tidy data concepts; visit from Kohl's data science team

- sep 25-27: dealing with missing and incomplete data

- oct 2-4: describing and summarizing data; **initial presentations!**

- oct 9-11: introduction to data visualization; visualization theories

- oct 16: constructing appropriate visualizations

- oct 17: **mid term project due!**

- oct 23-25: hypothesis testing and interpretations

- oct 30-nov 1: introduction to linear models; regression and classification

- nov 6-8: guest lecture on big data; least squares and residuals

- nov 13-15: regression inference

- nov 22: linear classifiers

- nov 27-29: k-nearest neighbors; visualizing and enhancing

- dec 4-6: data privacy; research ethics; algorithmic biases; **final presentations!**

- dec 15: **final project due!**

Grading Policy: class assignments (20%), initial presentations (10%), mid term project (20%), final presentations (10%), final project (20%), class performance (20%). This course will **not** be graded on a curve. The final grades will depend on the following scores:

A: 96 - 100; A-: 91 - 95; B+: 86 - 90; B: 81 - 85; B-: 76 - 80; C+: 71 - 75; C: 66 - 70; C-: 61 - 65; D+: 56 - 60; D-: 51 - 55; F: 0 - 50;

There are no regrade requests.

Course Policy:

- You are responsible for your own progress. Please check your marquette email, the course github repo and piazza about course announcements and news regularly.

- This is a course that utilizes active learning principles. As a result, there are no traditional lectures and you will be required to do readings for class everyday before arriving. These readings (for any week) will be posted the previous thursday by 12:00 pm except the first week, when it will be posted on saturday.

- As part of active learning, you are required to participate in class activities. There will be individual and group learning every day followed by immediate, in-class evaluation. You will be required to evaluate your peers.

- You are expected to ask, discuss and contribute to questions on piazza. Not only does this help you and enrich the course, but this will also count towards your class performance grade.

- Please bring a computing device everyday to class. We will be writing code together everyday. This is necessary.

- Each of you will create your own free github accounts to maintain your project and any code you write in class. This is part of a data science project's lifecycle and is expected to be shown to employers by data science job applicants.

- All submission of papers/project will be online via d2l. Any submission after the deadline will be considered late. I will post an exemplar and demonstrate in class, the right way to submit your mid term project.

- You must cite every reference in your papers and every library (if used) in your project in ACM style. Failure to do so will be regarded as a violation of academic integrity. Please refer to the section on academic integrity for more details.

- The mid term project paper should a maximum of 4 pages (excluding references), ACM extended abstract style adhering strictly to the current CSCW conference norms. It will be due at 11:59 PM local time. I will inform you about the expected content at an appropriate time.

- The final project paper should a maximum of 6 pages (excluding references), ACM extended abstract style adhering strictly to the current CSCW conference norms. It will be due at 11:59 PM local time. I will inform you about the expected content at an appropriate time.

- Regular attendance is essential to an active learning process. A student who incurs an excessive number of absences may be withdrawn from the class at the instructor's discretion.

- Please make sure your cell phone is turned fully off, or silent. No texting, reading emails, playing games, or anything else. I will reserve the right to ask you to leave the class if you are being distracting or disruptive.

Ott Memorial Writing Center: The Ott Memorial Writing Center offers free one-on-one consultations for all writers, working on any project, at any stage of the writing process. Marquette's writing center is a place for all writers who care about their writing, because every writer can benefit from conversation with an interested, knowledgeable peer. Writing center tutors can help you brainstorm ideas, revise a rough draft, or fine-tune a final draft. You can schedule a 30- or 60-minute appointment in advance (288-5542 or www.marquette.edu/writing-center), but walk-ins (in 240 Raynor or our other satellite locations) are also welcome. The Ott Memorial Writing Center also offers free workshops and hosts writing retreats.

Academic Integrity: Marquette University takes academic integrity very seriously. This is a core part of who we are as reflected by our Jesuit values. All students are required to take the Academic Integrity Tutorial. If you haven't, please go take it right now. All students are required to adhere to the Honor Pledge and follow the Honor Code. Please familiar yourself with the Academic Integrity website. There is a lot of useful information there.

I take a zero tolerance policy with violations of academic integrity. All papers and projects are run through a plagiarism detection software. If you are flagged, be assured that we will have a conversation. Let's not have that conversation shall we? If I determine that you have indeed violated academic integrity, you will receive a failing grade for that component of the course or for the entire course depending on the nature and severity of the violation. Please help me to make sure that there are no such incidents.

Accessibility Policy: If you have any accessibility needs, please contact Office of Disability Services (ODS) to register them as soon as possible. ODS works with students with documented disabilities to provide accommodations for their educational needs. The course has been designed for multiple different styles of learning. However, if you have any specific learning styles that you want me to know about which would not be addressed by AES, please reach out to me within the first week of class so I can try to accommodate.