COSC 4500/5500: Advanced Data Science

Spring 2019

Instructor:	Shion Guha	**Office Hours:**	W 10-1 pm
Email:	shion.guha@marquette.edu	**Office:**	318 Cudahy Hall

Description: Focuses on developing data products using the Javascript/D3 framework by combining concepts from human-computer interaction, visualization and design. Focus is also placed on model visualization, interpretation, A/B testing and design thinking.

Prerequisites: cosc 3570 (intro data science) or cosc 2100(data structures). I will assume that you either know or will pick up the basics of git and latex.

Is this class right for me/what do I expect from this course? This class is generally meant for upper level undergraduates and graduate students in data science, computer science, psychology, media studies, communication or any other students who are interested in developing design research thinking. Some frequently addressed points:

- This is not a programming course. I will not be "teaching" you Javascript/D3. You are expected to pick it up as we go along. All of you should have an object oriented programming background. If you do not, you will struggle in this class. if you need a refresher, I strongly encourage you to prepare for this by starting to read d3.js in action by elijah meeks and looking at Javascript/D3 online tutorials.
- You do not need a strong statistics or machine learning background for this class. While we will be visualizing models, any basic knowledge of statistical models (at the level of math 1700 or psych 2100) will do. All of you should already have such basic competency.
- This is a research seminar and not a lecture based course. There is significant amounts of reading, writing, discussing and critiquing each other's work. If you feel uncomfortable in such a setting, this class is not appropriate for you.
- This course develops crossover skills from many different areas. You do not need to have any type of design background to take this course. Of course, such a background is useful. If you can, I suggest taking this "10 hour" free MOOC from University of Virginia on Coursera.
- We will make frequent use of Github and Overleaf in this course. You are expected to already have (or pickup) these skills in your academic career. I will not be "teaching" them to you but will expect that you already have said skills.
- We meet one day a week in the evening for about 2.5 hours. This is usually how a research seminar is conducted. Given the depth of material to be covered, we will be attempting to do a lot in said time. There will be breaks but if you feel like this course format is not suitable for you, I would strongly encourage you to drop the course as a non-trivial amount of your grade depends on class performance and participation.
- Finally, to assess your readiness, I will be developing a "intro-tech-skills-demo" project that will be due within the first 3 days (Jan 20) of this course starting up. This will test all the technology and platform skills required for this course and will be worth 10 points. If you score 5 or below, I will strongly encourage you to drop the course. This is an implementation of the feedback that I received from the first iteration of this course that I ran last year. This is very quick but done to make sure that you can safely drop the course by the drop deadline which is on Jan 22.

Class: Th, 6:30-9:10 pm, 108 Cudahy Hall

Github: https://github.com/shionguha/cosc4500-advdatascience-sp19

Piazza: https://piazza.com/class/jqfn5cs2agq47a

Synchronous Learning: We will use Skype. I will provide more details on piazza. Undergraduates are not permitted to do this without prior permission on a case-by-case basis. In general, I wouldn't recommend this class for synchronous learning for anyone as you miss out on valuable class discussions as well as design activities that are relevant. Please come talk to me if you have any questions.

Books: This course is a research seminar. There are plenty of articles and books for this course. All article readings are available as online pdfs and you don't have to "buy" all the books per se though many are quite cheap; I will provide some and many are available online. however, think about these books as investments in your data science careers. They will be extremely useful beyond the classroom. Regardless, following my usual academic philosophy, I will not make any of these books "mandatory" for purchase but will instead **strongly recommend** them. They are relatively cheap on Amazon and the Marquette Librarians are pretty amazing at getting books via the Interlibrary Loan system. Make judicious use of these resources!

- d3.js in action by elijah meeks
- ways of knowing in hci by olson and kellogg.
- design of everyday things by don norman
- user interface design for programmers by joel spolsky

Objectives: This course aims to develop skills that can translate to building ethical, human centered, data products and services within organizations. At the end of the course, a successful student should be able to:

- design and develop data-driven applications using the Javascript/D3 stack,
- understand the various aspects of design research,
- develop a human-centered, design thinking research mindset,
- evaluate human interaction with data-driven models.

Timeline:

- jan 17: introduction to the course; very quick introduction to javascript/d3: scales and axes.
- **jan 20: "intro tech skills" assignment submission at 11:59 pm.**
- **jan 22: drop deadline at 11:59 pm.**
- jan 24: a crash course in design thinking; stanford d-school exercise
- jan 31: what is human computer interaction? **group formation deadline**
- feb 07: design research methods: sketching, brainstorming, storyboarding
- feb 14: design research methods: interviews
- feb 21: design research methods: data and tracking
- **feb 24: project proposal and plan due at 11:59 pm**

- feb 28: usability and user experience
- mar 07: introduction to visualization
- **mar 08: initial design prototype and evaluation plan at 11:59 pm!**
- mar 21: affordances in design
- mar 28: evaluation: in the lab
- apr 04: evaluation: in the wild
- apr 11: evaluation: A/B testing fundamentals
- **apr 18: interim evaluation results and feedback due at 11:59 pm**
- apr 25: evaluation: more A/B testing
- **may 02: final project presentations**
- **may 10: final paper submission at 11:59 pm**

Grading Policy: Both individual as well as group skills will be tested as part of this course. The following are the items that will be assessed as part of this course:

- **jan 20:** individual intro-tech-skills-demo submission (10%)
- individual weekly reading responses and discussion on piazza (10%)
- individual class activity performance and participation (10%)
- individual in-class research paper leadership and discussion (20%)
- **feb 24:** group project proposal plan (10%)
- **mar 08:** group design prototype and evaluation plan (10%)
- **apr 18:** group interim evaluation results (10%)
- **may 10:** group final project report (10%)
- **may 02:** group final project presentations (10%)

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; D-: 46 - 50; F: 0 - 45;
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. You will be required to post reading responses on piazza as part of class participation the day before (wednesday) by **9 pm**.

- 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 grade. Please monitor piazza everyday as I will be posting readings, questions and polls there regularly.
- Please bring a computing device as well as a pens/pencils/crayons everyday to class. If we won't be writing or experimenting with code, we will be doing design exercises.
- If you don't already, each of you will create your own free github accounts to maintain your project and any code you write or design prototypes that you create for this course. This is part of a data science project's lifecycle and is expected to be shown to employers by data science job applicants.
- If you don't already, you will be expected pick up LaTeX, specifically overleaf. This is the lingua franca in computer science and data science and is often used to write academic or white papers in serious data science teams. I recommend Overleaf as it is the "Google Docs" equivalent of LaTeX and allows you to write
- All submission of project reports will be online via d2l. Any submission after the deadline will be considered late. In addition, if there is any code as part of your submission, you will be required to send a pull request to the class github repository.
- 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.
- We will follow CHI Student Design Competition rules and format for this course. CHI is generally regarded as the top conference in HCI/UX in the world. CHI Proceedings Formats are here. Specifically, we will be using the Overleaf template. Submissions in any other format shall not be accepted.
- Following the norms of the CHI 2019 Student Design Competition, your group project topic should generally follow the "CHI 2019 Design Brief: Weaving the Threads within the Social Fabric". Please read the relevant section of the website. Throughout the semester, you will be tested via the criteria mentioned in that design brief.
- 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.