

IS590ODM: Open Data Mashups Syllabus

Instructor information

- Instructor: Elizabeth Wickes
- Office: 329 Information Sciences Building
- Email: wickes1@illinois.edu
- Cell phone: look in moodle, I'm not putting this on github (do not call after 9pm)

Course description

Description: Data sharing and modern open data standards have been creating large repositories of data that remain disconnected. Many data science and machine learning techniques are boosted by incorporating data representing a variety of domains and granularities. Topics on data curation, data cleaning, copyright, web scraping, storage, processing, and automation will be reviewed. This course seeks to explore techniques and perspectives of combining various data sources to create a dataset ready for analysis, but in a project oriented space so that each topic is synthesized with practice and experienced in context. Students will select a project area and explore the technical and conceptual requirements of that project space, eventually producing a proof of concept around it. All project domains and areas are open, with the only requirement being that they combine several data sources into a new dataset. This course is meant for students who have completed at least two semesters of coursework, are comfortable with programming in Python (the project can be completed in any language, but instruction will be in Python), and desire a space to explore and develop a capstone or independent study project. However, further work on the project is not a requirement.

Guest speakers and field experts from the University Library will be invited. Students will be encouraged to share and publish their datasets at the end of the semester.

Course prerequisites: IS452 or demonstrated programming experience, 20 hours of completed coursework

This course is worth 2 credit hours but is still significant work because of your project. You are expected to already have some exposure to each topic area around this project, so there will be very minimal outside readings, etc.

Learning Goals

This is a completely project oriented class. You will be working entirely on your project in this class. The first

half of the class will quickly get you started on your project and the second half will be you working on your project. Portions of the project will be due throughout the class, so no elements will take you by surprise.

- identify and assess relevant data sources for a research project
- process, clean, and quality check datasets
- combine multiple datasets into one for an analysis process
- produce quality dataset and project documentation
- effectively deliver various types of project pitches
- write effective conference talk proposals
- provide constructive peer feedback and peer reviews
- receive and interpret peer feedback

Class Time, Office Hours, and Lab Time

- Regular class times:
 - Thursdays from 2-4pm
 - In class attendance is absolutely required. Attendance will be taken, including the first day and final two presentation days.
 - Many assignments have elements that we will be working through together during class, and most class sessions will include some open lab time. This portion is also required.

Attendance

Attendance to class required and participation points will be awarded for each class.

Students are allowed one class absence either for illness or a scheduled event. Anyone needing to miss more than one class will need to speak to me to discuss absences and attendance points. Medical and accessibility considerations will be factored in, but you will need to meet with me before or ASAP to work out arrangements.

Required textbooks

None required.

Recommended reference books

Each of these are available as online ebooks from the library via Safari online access.

- Pandas for Everyone: Python Data Analysis (<https://www.amazon.com/Pandas-Everyone-Analysis-/>)

[Addison-Wesley-Analytics/dp/0134546938](https://www.amazon.com/Everyone-Advanced-Analytics-Graphics-Addison-Wesley/dp/0134546938))

- R for Everyone (<https://www.amazon.com/Everyone-Advanced-Analytics-Graphics-Addison-Wesley/dp/013454692X>)

Grading

Your final grade will be out of 500 points. Here are the big divisions for grades:

- Participation: 60 points (15% of your grade)
- Final project deliverables: 200 points (50% of your grades)
- Final project presentation: 50 points (12.5% of your grade)
- All other assignments and tasks: 90 points (22.5% of your grade). Each assignment is worth between 15 and 35 points, and are all listed in the course moodle.

These other assignments may be subject to change due to the experimental nature of this course, but the total value of their points will remain 90. They are designed to become parts of your final set of deliverables or get you started on the project.

You will earn participation points by actively participating in class. Based on absent, low participation, or strong participation.

A student may only receive a passing grade if they submit a final project. Your final grade may also be lowered if your project does not meet all requirements. This should not really be a problem, as the other assignments will force you to meet those requirements, but discretion may be used in edge cases where the completed project did not meet all the requirements.

Plagiarism and adapting the work of others

When the assignments require narrative or written (e.g. non-code) answers, appropriate attribution and citation is required. This is a graduate level course and all students are expected to know and understand the concept of plagiarism and academic integrity. Several writing centers offer services to answer questions about plagiarism. The instructor will follow one of the standard procedures when plagiarism is submitted: http://studentcode.illinois.edu/article1_part4_1-402.html. Students will be given 0 points for an assignment if it contains plagiarism in whole or in part. No make up work will be accepted to replace that grade. The instructor reserves the right to fail a student if more serious plagiarism is found.

The Code of Policies and Regulations Applying to All Students will be applied in all instances of academic misconduct committed by students. This applies to all exams, presentations, assignments and materials distributed or used in this course. You can review these policies at the following web site:

<http://admin.illinois.edu/policy/code/index.html>

Late policy and Homework extensions

Grace period

There is an 8 hour grace period for all assignment submissions. All assignments are due in moodle at 11:55pm, so any submissions turned in before 8am the next day will not receive a late penalty. The calendar in moodle can look confusing if you set a due date early in the morning, thus I've adopted this system.

Due dates

Most assignments are due on Sunday evenings. There are a few exceptions, so always check. There are also a few things you will need to have prepared for class, but won't be submitted until later. It will be noticeable that you don't have this work prepared. I will do my best to remind you, but this is largely your responsibility to review moodle and come to class prepared.

Extension

Each student may use up to two extension requests for regular homeworks. These will give you an extra four days to submit without penalty. Homework items meant to be used during class, the final project submission, and final presentation submissions are ineligible for any extensions aside from a personal crisis. You only need to email me to say that you are using an extension. You don't need to ask for permission (presuming that you're requesting it for an assignment that permits an extension).

Extension requests must be submitted by 11:55pm Monday (with an 8 hour grace period going into Tuesday).

Late penalties

Submission after the grace period and before 11:55pm Monday will receive a 10% penalty from the maximum value of the homework.

Submissions after 11:55pm Monday but before 11:55pm Tuesday will receive a 30% penalty from the maximum value of the homework.

Submissions after 11:55pm Tuesday will be marked as zero points.

Penalties will be taken off the maximum value only. Meaning that if you turn in a 20 point assignment one day late, the maximum points value for that assignment is 18 points. If you earn 15 points on what you've submitted, you will not take any penalties.

- Sundays 11:55pm, this is the usual due date

- Monday 11:55pm, last chance for late submission for the 10% late penalty from the max homework points. Last chance to request an extension.
- Tuesday 11:55pm, last chance for the late submission, with a 30% penalty from the max homework points.
- Any homeworks not submitted at this point without an extension will be given a 0.
- Thursday 11:55pm, due date for homeworks on an extension. Any homeworks on extension not submitted by this time will be given a 0.

What is a personal crisis?

Being behind on homework or just needing more time is not a personal crisis. This category involves house fires/floods, medical emergencies for yourself or a dependent, a death in your family, etc. Any student experiencing something like this is encouraged to reach out to their advisor and the student affairs team. You are welcome to contact me if you feel comfortable explaining the situation.

Request for Special Accommodations

We are committed to providing a learning environment where our students can succeed. If you require special accommodations, please contact us and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES, you may visit 1207 S. Oak Street, Champaign, call 217.333.4603, or email disability@illinois.edu. We will try to meet all accommodations once the process has started. Please note accommodations are not retroactive to the beginning of the semester, but begin the day you contact your professor with a current letter of accommodation from DRES.

Statement of Inclusion

<http://www.inclusiveillinois.illinois.edu/> As the state's premier public university, the University of Illinois at Urbana-Champaign's core mission is to serve the interests of the diverse people of the state of Illinois and beyond. The institution thus values inclusion and a pluralistic learning and research environment, one in which we respect the varied perspectives and lived experiences of a diverse community and global workforce. We support diversity of worldviews, histories, and cultural knowledge across a range of social groups including race, ethnicity, gender identity, sexual orientation, abilities, economic class, religion, and their intersections.