INFO 3401: Information Exploration

M/W/F, 10:00-10:50, Humanities 125

Prof. Danielle Albers Szafir
danielle.szafir@colorado.edu

Office Hours: Friday, 11am-12pm & by appointment, ATLAS 235C

Course Description

Information empowers people to build deeper understandings of the world and make more informed decisions. However, the increasing volume and variety of available information makes it hard for people to make sense of that data. This course will allow you to build the skills necessary to work with stakeholders to explore and build novel insights through data. You will gain hands-on experience with different tools and techniques for exploring information, including statistical methods, qualitative analyses, and visual analytics. You will learn how to generate and synthesize new findings from data, combine information from multiple sources, and identify questions and findings that are directly relevant to people.

Required Text

Nate Silver. The Signal and the Noise: Why So Many Predictions Fail--But Some Don't

Course Structure

The intention of this class is to help you combine the skills you've built in the Computational Thinking, Quantitative Thinking, and Foundations courses. By the end of this class, you will have gained experience applying the methods you've used to work with small datasets to larger and more complex problems. As such, the course will rely heavily on a combination of video lectures to be watched before class, in-class lectures and discussion, and in-class hands-on activities where we will work through data problems using the concepts from lecture.

While this course is not a programming course, many in-class activities and projects will require programming in-class. As a result, please plan on regularly bringing a laptop to class that you can use to develop and run your own programs. We will primarily use Python; however, time permitting, we will also likely work with other languages such as R. If you do not have access to a machine that will support your work in this class, please let Dr. Szafir know by the end of Week 1.

Due to the use of video lectures, there will be no required textbook for this class. Additional readings may be provided to support individual concepts, but will be made available through Canvas. The discussion board on Canvas will be actively monitored to support questions that arise from video lectures and we will additionally take time during each class period for discussions and clarifications. If you find yourself in a situation where you would like additional readings to support the lectures, please email Dr. Szafir for additional materials. Similarly, if you find resources online that are particularly useful, please feel free to share these with the class on the Canvas discussion board.

Topic Schedule: (Subject to Change)

Module One: January 15-17	History ● DS: Pipelines & History ● SE: Repositories ● Textbook Reading: Preface (Optional)
Module Two: January 22-24	Scraping & Collating Data ● DS: Scoping & Collection Practices ● SE: Terminals
Module Three: January 27 - February 3	Economic Data ● DS: Probability & Causality ● SE: Objects & Program Structure ● Textbook Reading: Ch 1, Ch 8
Module Four: February 10 - February 21	Political Data ● DS: Sampling, Uncertainty, & Inference ● SE: Databases ● Textbook Reading: Ch 2
Module Five February 24-March 6	Sports Data ● DS: Clustering ● DS/SE: Visualization Toolkits ● Textbook Reading: Ch 3
Module Six March 9-March 20	Climate Data ● DS: Bayesian Modeling, Evaluation Metrics ● DS/SE: SciKitLearn ● Textbook Reading: Ch 4, Ch 12
No Class. Spring Break	
Module Seven March 30-April 10	Disaster Data ● DS: Feature Engineering, Data Wrangling ● SE: Spatial Data Structures ● Textbook Reading: Ch 5
Module Eight April 13-April 24	Text Data ● DS: Text Processing, Topic Modeling ● SE: Regular Expressions ● Textbook Reading: Ch
April 27-May 1	Project Presentations

Course Objectives & Outcomes

The purpose of this course is to help you build foundational skills in exploratory analytics, with an emphasis on transitioning the skills you've learned in prior classes to larger, more realistic datasets and challenges. Upon completing this course, you should expect to:

- Understand foundational skills associated with computational data exploration and manipulation, including scoping exploratory projects, engaging in data collection and wrangling, and leveraging a broad set of techniques for data sensemaking.
- Have experience with common tools in data science, including basics version control, file manipulation, program structure, and translating algorithms to code.
- Be fluent in basic algorithms and approaches for data analysis, including basic machine learning, visual analytics, and statistical analysis methods.
- Be comfortable working with data in different forms to accomplish a goal.

Assignments

- Participation (10% of your final grade): This course will require regular participation in-class. You are allowed three absences before missing class will begin to affect your grade: 3% of your participation grade will be deducted per additional absence beyond the first three. There will be no exceptions to this rule.

- Quizzes (20% of your final grade): As this course focuses classroom time primarily on questions, discussion, and in-class activities, you will be asked to complete weekly quizzes based on materials covered in the video lectures. Quizzes will be due by 11:59pm on Fridays. Your lowest two quiz grades will be dropped from your grade.

- Homework (40% of your final grade): Each week (except for exam weeks), you will receive a set of problems that we will work on in small groups as part of our in-class activities. You are responsible for completing the full set of problems; however, some problem sets may require additional time to complete outside of class. The full problem set for each week is due by 11:59pm on Sunday evenings through the course GitHub. Note that each person must submit their own problem set even if working with others to complete the homework.

- Final Project (30% of your final grade): Over the course of the term, you will identify a stakeholder with a problem that could be addressed through data, collect and analyze data of relevance to their problem, and recommend a solution to that problem using both a formal write-up and presentation. More details about this project will be announced during Week Two.

- Extra Credit (5% boost to your final grade): Dr. Szafir will release an extra credit assignment during Week Three. If you wish to complete this assignment, please submit it by 11:59pm on Monday, 3.30.

Grading

This course will use a standard, 100-point grading scale:

93.0% and above: A
90.0%–92.9%: A-
87.0%–89.9%: B+
83.0%–86.9%: B
80.0%–82.9%: B-
77.0%–79.9%: C+
73.0%–76.9%: C
70.0%–72.9%: C-
67.0%–69.9%: D+
63.0%–66.9%: D
60.0%–62.9%: D
Below 60.0%: F

Late Policy

All assignments and projects will be due by 11:59pm on the assigned due date. Extensions to assignment deadlines will only be granted in extreme circumstances. If you need an exception or extension to an assignment deadline for any reason, please let Dr. Szafir know as soon as possible and provide reasonable documentation as to the reason for your request. Extensions cannot be granted after the due date.

Policies & Commitments:

Information Science Teaching Pledge:

I pledge to give feedback to students constructively and quickly, specifically within 7 days of an assignment. I pledge to treat each student with respect. I invite constructive feedback if a student feels that I could improve my instruction or conduct in the classroom. I will do my best to respond to your emails within 2 business days. If you have not heard from me by then, I welcome follow-ups either in-class or over email.

Technology Requirements:

Students in this course will benefit from having a laptop or tablet available for notetaking, in-class work, homework, and presentations. If you do not have access to a laptop or tablet, please consult with me. However, students and the instructor alike are encouraged to quit mail and other applications that may be distracting; to turn off notifications and silence phones; and to put nonessential equipment away.

Open Discussion & Debate:

(adapted from Prof. Casey Fiesler's statement, with permission and thanks). In the classroom,

students and instructors need to feel comfortable sharing their opinions and questions openly, even when we disagree. Disagreement is expected, but must be respectful and civil at all times. Students should feel welcome to share thoughts during class discussion without any fear of being disparaged for their opinions. Like yourselves, I also have opinions, and I will attempt to surface my own biases when appropriate. These disagreements or differences of opinion will not impact grades, as long as students are respectful. I invite students to meet with me to discuss concerns and ideas about how to make our learning community a positive experience for all.

This policy extends to topics in the course. Please feel free to reach out to me should there be any topics you'd like to discuss more deeply in class or that are not currently part of the course but you would like to see integrated. While I cannot guarantee that every request will be accommodated, I will do my best to make sure the curriculum and topics align with the needs of all of the students in the course.

Accommodation for Disabilities:

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the Disability Services website. Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition or injury, see Temporary Medical Conditionsunder the Students tab on the Disability Services website.

Family Support Statement:

(Adapted from Dr. Melissa Cheyney, OSU) CU does not have a formal policy on children in the classroom. However, I firmly believe that parents should not be forced to choose between family and education. All exclusively breastfeeding babies are welcome in class as often as is necessary to support the breastfeeding relationship. For older children and babies, minor illnesses and unforeseen disruptions in childcare often put parents in the position of having to choose between missing class to stay home with a child and leaving them with someone you or the child does not feel comfortable with. While this is not meant to be a long-term childcare solution, occasionally bringing a child to class in order to cover gaps in care is perfectly acceptable. In all cases where babies and children come to class, I ask that you sit close to the door so that if your little one needs special attention and is disrupting learning for other students, you may step outside until their need has been met. Non-parents in the class, please reserve seats near the door for your parenting classmates.

I ask that all students work with me to create a welcoming environment that is respectful of all forms of diversity, including diversity in parenting status.

Mental Health Statement:

Success in this course depends heavily on your personal health and wellbeing. **Recognize** that stress is an expected part of the college experience, and it often can be compounded by unexpected setbacks or life changes outside the classroom. Your other instructors and I strongly encourage you to **reframe** challenges as an unavoidable pathway to success. **Reflect** on your role in taking care of yourself throughout the term, before the demands of exams and projects reach their peak. Please feel free to **reach out** to me about any difficulty you may be having that may impact your performance in this course or campus life as soon as it occurs and before it becomes too overwhelming. In addition to your academic advisor, I strongly encourage you to contact the many other support services on campus that stand ready to assist you.

Religious Holidays:

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, please alert me of any such situations at least one week before any potentially conflicting deadlines. See the campus policy regarding religious observances for full details.

Classroom Behavior:

Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records. For more information, see the policies on classroom behavior and the Student Code of Conduct.

Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation:

The University of Colorado Boulder (CU Boulder) is committed to fostering a positive and welcoming learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct intimate partner abuse (including dating or domestic violence), stalking, protected-class discrimination or harassment by members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or cureport@colorado.edu. Information about the OIEC, university policies, anonymous reporting, and the campus resources can be found on the OIEC website.

Please know that faculty and instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, discrimination, harassment and/or related retaliation, to

ensure that individuals impacted receive information about options for reporting and support resources.

Honor Code:

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code (honor@colorado.edu); 303-492-5550). Students who are found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code as well as academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found at the Honor Code Office website.

The first instance of academic dishonesty will result in a grade of 0 on the assignment in question. Subsequent violations will result in a failing grade for the course.