INFO 3401: Information Exploration

M/W/F, 2:00-2:50, Humanities 1B90

Prof. Danielle Albers Szafir
danielle.szafir@colorado.edu
Office Hours: Monday 3-4 & by appointment, ENVD 201

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.

This course will build on skills you've developed in earlier Information Science courses. In this course, you will build on this foundational toolbox by:

- Gaining more experience in applying your skills to different datasets
- Developing deeper familiarity with existing skills, including exposure to new methods and techniques
- Fusing these skills together to form a complete exploratory pipeline by understanding the relationship between different data, approaches, user needs, tools, and techniques.

Textbooks:

Max Shron, "Thinking with Data"
Cathy O'Neil & Rachel Schutt, "Doing Data Science: Straight Talk from the Frontline"
Additional readings will be made available through D2L.

Note that these texts are available for free to all students through CU Library's Safari.

Prerequisites:

One of INFO 1111, 1121, 2131
One of INFO 2201 (or equivalent Data Structures course), 2301 (or equivalent intermediate statistics course)
Both of INFO 1201 (or equivalent Intro Programming course) and 1301 (or equivalent Intro Statistics course)

Topic Schedule: (Subject to Change):

08.28 - 09.01	Scaffolding & Overview: ● Syllabus Overview ● Introductions ● Tech Infrastructures ● The Exploratory Pipeline

09.04: No Class (Labor Day)

09.06 - 09.08	Data & People: ● Requirements Analysis ● Confirmation Bias
09.11 - 09.15	What Data Looks Like: ● Files & Databases Primer ● Data Collection & Scraping ● Data Wrangling & ETL ● Data Fusion
09.18 - 09.22	Exploratory Data Analysis: ● Common tools and techniques ● Sampling and approximation ● Working with experts through think-aloud studies
09.25 - 09.29	Visual Analytics & High-Dimensional Data: ● Basic visualizations for exploration and explanation ● Aggregation and other ways to handle large data ● "The Visual Statistician"
10.02 - 10.06	Qualitative Methods (*Guest Lecturers, Topics TBD*)
10.09 - 10.13	Statistical Foundations & Usability Data: ● Probability & Statistics Primer ● Hypothesis Testing & Validity ● Surveying the many ways of using and interpreting statistical results ● Preregistration
10.16 - 10.20	Core Statistical Algorithms: ● K-Means ● K-NN ● Regressions & kernels

10.20: Mid-Term Exam

10.23 - 10.27	Text Analytics: ● Text as Data ● Document Clustering ● Sentiment Analysis ● Topic Modeling
10.30 - 11.03	Causality & Prediction: ● Correlation versus Causation ● Methods and metrics for inferring causality ● Basics of data experimentation
11.06 - 11.10	Time Series: ● Modeling through Regression & Forecasting ● Data Smoothing
11.13 - 11.17	Data Engineering:

	● Feature Selection ● Data Leakage ● Entropy
Fall Break	
11.27 - 12.01	Network Data: ● Network Features ● Basic Network Algorithms
12.04 - 12.08	Catch-Up Time (or Final Project Work)
12.11 - 12.13	Final Project Presentations

Assignments:

- **Participation** (20% of your final grade): This course will require regular participation in-class. You are allowed three excused 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.

- **Lab Activities** (20% of your final grade): Every Friday will be devoted to a hands-on lab activity to be completed in-class. You will be able to drop your lowest score on two in-class labs. There will be no make-ups for lab activities.

- **Homework** (20% of your final grade): You will be expected to contribute to regular individual and small group assignments.

- **Mid-Term Exam** (20% of your final grade): The midterm is tentatively scheduled for 10.20 during our usual class time. It will cover all the materials through 10.16.

- **Final Project** (20% 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. Final projects will be presented during the end-of-term Information Science Showcase, with the specific date and time TBD.

- **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, 10.30. Also, there is a typo in this syllabus. Show me the typo by the start of the second lecture for 2 extra credit points on your first homework assignment.

Late Policy:

All assignments 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.

Grade Distribution:

- >= 93.0%: 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
- < 60.0%: F

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 1 business day. 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 ringers; 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:

The classroom is an environment where everyone should be presented with equal opportunities to succeed. As such, I am happy to make any reasonable accommodations necessary due to disability. If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to me in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environmnt. Information on requesting accommodations is located on the [Disability Services website](http://www.colorado.edu/disabilityservices/students) (www.colorado.edu/disabilityservices/students). 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 Conditions](#) under the Students tab on the Disability Services website and discuss your needs with your professor. I understand that emergency situations arise quickly and unexpectedly. Please let me know as soon as possible in the event that a medical situation arises that interferes with your ability to either attend or complete assigned work in this class.

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, I am happy to work with you to accommodate absences due to religious observance so long as I am notified of the impending absence by the end of Week 3 (9.15). Beyond this point, absences will count towards your . 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 maintaining a positive learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct, discrimination, harassment or related retaliation against or by any employee or student. CU's Sexual Misconduct Policy prohibits sexual assault, sexual exploitation, sexual harassment, intimate partner abuse (dating or domestic violence), stalking or related retaliation. CU Boulder's Discrimination and Harassment Policy prohibits discrimination, harassment or related retaliation based on race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. Individuals who believe they have been subject to misconduct under either policy should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127. Information about the OIEC, the above referenced policies, and the campus resources available to assist individuals regarding sexual misconduct, discrimination, harassment or related retaliation can be found at the [OIEC website](#).

Honor Code:

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to [the academic integrity policy](#). Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, resubmission, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code Council (honor@colorado.edu; 303-735-2273). Students who are found responsible for violating the academic integrity policy will be subject to non-academic sanctions from the Honor Code Council as well as academic sanctions from the faculty member. Additional information regarding the 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.