Information Exposition

INFO 3402; Spring 2019

Monday, Wednesday, Friday 11:00–11:50; CASE E250

Assistant Professor Brian Keegan
E-mail: brian.keegan@colorado.edu
Office: TLC 203
Office hours: Friday, TLC 266, 09:00–11:00 or by appointment

Course Description

This course teaches students to communicate information to a wider audience and construct stories with data across a variety of domains. Students will learn to use data for rhetorical purposes, applying visual, statistical and interpretative methods. Students will learn to think critically about the ethical and social implications of using data in expository media, including identification of bias.

Learning objectives

- Master seven fundamental data analysis techniques
- Improve confidence in collecting and reshaping data
- Identify interesting questions within found data
- Use strategies like persuasion and storytelling to communicate findings
- Think critically about the opportunities and limitations of data

Course Design

Class will meet three times per week on Monday, Wednesday, and Friday from 11:00 to 11:50 in CASE E250. The format of each class will vary between lectures, analysis labs, and presentations. Student performance will be evaluated through a combination of Module Assignments, Weekly Presentations, and a Final Project. There is no final exam.

The class is split up into seven two-week modules. Each module corresponds to a fundamental data analysis technique: (1) *exploring* an existing data set, (2) *retrieving* new data from the web, (3) *cleaning* data into "tidier" formats, (4) *combining* different data sets together, (5) *aggregating* data into summaries, (6) *inferring* patterns in the data, and (7) *extrapolating* findings to new data.

The first week of each module will be a "show" week dedicated to reviewing or introducing the method. The second week of each module will be a "tell" week dedicated to connecting the technique to larger social and ethical implications. In each week, Mondays will be a lecture providing motivation and background, Wednesdays will be a combination of lecture and exercises, and Fridays will be Weekly Presentations.

Prerequisites

Students should have completed the sequences of INFO 2201, INFO 2301, and INFO 3401 or similar coursework covering intermediate computational reasoning, intermediate statistical reasoning, and exploratory data analysis before enrolling in INFO 3402. If you have questions about these prerequisites, please email the instructor.

Requirements
Students' regular and sustained participation in all class activities as well as punctual and thorough completion of assignments are essential. If you need to be excused from attending a class session or need an extension to an assignment, please email instructor at least 24 hours in advance.

Course Website and Materials
There is no textbook required for class, but there will be required readings, tutorials, and other material, which will be made available through Canvas:

https://canvas.colorado.edu/courses/26235

Once the semester begins, this PDF version of the syllabus will be revised infrequently and any revised requirements will be posted as announcements and updated course schedule to Canvas. The instructor reserves the right to make changes to the course's schedule, evaluation criteria, policies, *etc.* in this syllabus through announcements in class and on Canvas, so please check Canvas regularly. Students should email the instructor if there are any discrepancies or questions.

The class will make extensive use of the Medium blogging platform. Instruction on how to create accounts, read, write, and post to the class publication will be covered in the first week of class. There is extensive documentation in the Medium Help Center as well as multiple tutorials. Students will submit their Module Assignments on Medium—rather than Canvas—with the expectation that their writing will be read by the general public. If students are unable to or do not want to use the Medium platform, they should email the instructor immediately to work out an alternative arrangement.

Statistical Computing
Students will need to use statistical computing software. Jupyter notebooks written in Python 3 will be used for all in-class examples and assignments. The Anaconda distribution of Python 3.5 (or above) is strongly recommended to provide all of these programs and other libraries. If students cannot bring a laptop to class, they should email the instructor to work out an alternative arrangement.

Evaluation
Students will be evaluated through three different mechanisms.

Module Assignments (35%) Module Assignments are intended to develop students' skill and confidence in using oral and written communication to share the findings of quantitative data analyses to general audiences. There are seven Module Assignments in total, one per module. Each Module Assignment is worth 5% of the final grade (35% cumulative) and are due the Monday before the start of the next module. The format and evaluation criteria of each Module Assignment will vary, but will share two themes: "showing" how you did a data analysis in a tutorial-like format or "telling" what you found from a data analysis in a reporting-like format. Each Module Assignments will be published as blog posts via the class's Medium publication. In the absence of an approved excuse, late submissions will be docked 2% of their value for every hour elapsed since the deadline: *assignments submitted after Wednesday at 13:00 (50 hours after the deadline) will lose all credit.*

Weekly Presentations (30%) Weekly Presentations are intended to develop students' communication skills by summarizing their findings and providing peer feedback. Each Friday, students should be prepared to present on their Module Assignment's progress, questions, and concerns. Approximately five students will be randomly selected to present their work-in-progress using a critical response process. Other students' participation as responders in the critical response process will also be evaluated. Each week's participation in the critical response process is worth 2% of the final grade (30% cumulative). If a student has a disability, anxiety, or another issue that limits their ability to participate in this format, please email the instructor. In the absence of an approved excuse, students who do not participate in a week's critical response process will not receive credit. There will be no make-up opportunities for missed Weekly Presentation credit.

Final Project (35%) The Final Project is intended to be a portfolio piece highlighting a student's analytical and communicative abilities. The project will likely—but does not have to—be an extension of a Module Assignment that goes deeper in data analysis and write-up with the goal of submitting for external publication as a guest blog post, op-ed, *etc*. A proposal will be presented in the middle of the term to go through the critical response process (10% of final grade), will be presented to the class in the final week (5% of final grade), and a written version will be posted to the class's Medium publication (20% of final grade). Further details about the Final Project will be collaboratively developed and detailed later in the course. In the absence of an approved excuse, late Final Project submissions will be docked 2% of their value for every hour elapsed since the deadline.

Course Policies

In-Class Confidentiality

The success of this class depends on students feeling comfortable sharing questions, ideas, concerns, and confusions about assignments, work-in-progress, and their personal experiences. Students may read, comment, and run on classmates' writing, code, and other class-related content for the sole purpose of use within this class. However, students may not use, run, copy, perform, display, distribute, modify, translate, or create derivative works of another student's work outside of this class without that student's expressed written consent or formal license. Furthermore, students may not create any audio, video, or other records during class time without the instructor's permission nor may students publicly share comments made in class attributable to another person's identity without that person's permission.

Critical Response Process

The class will make regular use of Liz Lerman's "Critical Response Process" for the Weekly Presentations and the Final Project. The details of this process will be covered in more detail in Week 1, but students will rotate through the roles of "artist" and "responder" regularly. Students are expected to participate in good faith when sharing statements of meaning, questions, neutral questions, and permissioned opinions.

Faculty Interaction

In addition to teaching this class, Professor Keegan also (1) manages a research program; (2) advises students; (3) performs service for the academic community; and (4) lives his life as a private citizen. He will check e-mail between 8:00 and 18:00 on non-holiday business days and try to respond to emails within 24 hours. He welcomes online or offline interactions outside of class, however these are not appropriate spaces for discussing class matters. E-mailing me or coming to my office hours are the best ways to get feedback outside of lecture.

Accommodations for Disabilities

I am committed to providing everyone the support and services needed to participate in this course. If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to the instructor 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 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 the instructor.

Religious Observance

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 assignments/attendance. If this applies to you, please email the instructor as soon as possible to make the appropriate accommodations.

Classroom Behavior
Students and instructors 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 differences of race, color, culture, religion, creed, politics, veteran's status, sexual orientation, gender, gender identity and gender expression, age, ability, and nationality. Class rosters are provided to the instructor with the student's legal name. The instructor will honor your request to address you by an alternate name or gender pronoun. Please advise the instructor of this preference early in the semester so that he may make appropriate changes. For more information, see the policies on class behavior and the student code.

Harassment and Discrimination
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 of the institution. 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 nonacademic sanctions from the Honor Code Council as well as academic sanctions from the faculty member. Additional information can be found at honorcode.colorado.edu.

Acknowledgements

The design and format of this course borrows from other courses.

- Bergstrom, Carl & West, Jevin (2017). *Calling Bullshit*. University of Washington.

This syllabus was typeset in LaTeX using Overleaf with the fbb/Bembo font and is derived from the `memoir` styles adapted by Kieran Healy and Benjamin 'Mako' Hill.

Course Outline

The schedule will evolve throughout the semester, so please consult the schedule online at Canvas for the most up-to-date information.

Week 1 – *Exploring*: Show

Monday, January 14; Wednesday, January 16; Friday, January 18

NOTE: *Professor Keegan will be out-of-town Friday, January 18. There will be a guest lecturer in his place.*

Course overview, reviewing fundamentals of exploratory data analysis, using Medium.

Week 2 – *Exploring*: Tell

Wednesday, January 23; Friday, January 25

Defining "bullshit" and types of empirical validity. **Assignment 1 due Monday, January 28.**

Week 3 – *Retrieving*: Show

Monday, January 28; Wednesday, January 30; Friday, February 1

Types of data and scraping from the web and APIs.

Week 4 – *Retrieving*: Tell

Monday, February 4; Wednesday, February 6; Friday, February 8

How to interview a dataset and ethics of data scraping. **Assignment 2 due Monday, February 11.**

Week 5 – *Cleaning*: Show

Monday, February 11; Wednesday, February 13; Friday, February 15

Handling missing data and creating tidy data.

Week 6 – *Cleaning*: Tell

Monday, February 18; Wednesday, February 20; Friday, February 22

NOTE: *Professor Keegan will be out-of-town Friday. There will be a guest lecturer in his place.*

Methods for imputation and viewpoint diversity. **Assignment 3 due Monday, February 25.**

Week 7 – *Combining*: Show

Monday, February 25; Wednesday, February 27; Friday, March 1

Using unique identifiers and comparing types of joins.

Week 8 – *Combining*: Tell

Monday, March 4; Wednesday, March 6; Friday, March 8

Visual design and de-anonymization attacks. **Assignment 4 due Monday, March 11.**

Week 9 – *Aggregating*: Show

Monday, March 11; Wednesday, March 13; Friday, March 15

Group-by aggregation, pivot tables, and summarization.

Week 10 – *Aggregating*: Tell

Monday, March 18; Wednesday, March 20; Friday, March 22

Cognitive biases, persuasion, and behavioral change. **Assignment 5 due Monday, April 1.**

Week 11 – Spring Break

Monday, March 25; Wednesday, March 27; Friday, March 29

No class.

Week 12 –*Inferring*: Show

Monday, April 1; Wednesday, April 3; Friday, April 5

Hypothesis testing, regression, and statistical power.

Week 13 – *Inferring*: Tell

Monday, April 8; Wednesday April 10; Friday, April 12

Correlation versus causation and counterfactuals. **Assignment 6 due Monday, April 15.**

Week 14 – *Extrapolating*: Show

Monday, April 15; Wednesday, April 17; Friday, April 19

Forecasting and prediction.

Week 15 – *Extrapolating*: Tell

Monday, April 22; Wednesday, April 24; Friday, April 26

Statistical paradoxes and logical fallacies. **Assignment 7 due Monday, April 29.**

Week 16 – Final presentations

Monday, April 29; Wednesday, May 1

Presenting final project assignments in class. **Final Project due Monday, May 7.**

Module	Week	Dates	Topics	Due Date
Exploring	1 2	Jan 14 – Jan 18 Jan 23 – Jan 25	*Show*: Administrivia, EDA *Tell*: Bullshit, Validity	Assignment 1: January 28
Retrieving	3 4	Jan 28 – Feb 1 Feb 4 – Feb 8	*Show*: Data types, Scraping *Tell*: Interviewing a dataset, Ethics	Assignment 2: February 11
Cleaning	5 6	Feb 11 – Feb 15 Feb 18 – Feb 22	*Show*: Missing data, Tidy data *Tell*: Imputation, Viewpoint diversity	Assignment 3: February 25
Combining	7 8	Feb 25 – Mar 1 Mar 4 – Mar 8	*Show*: Identifiers, Joins *Tell*: Visual design, Deanonymization	Assignment 4: March 11
Aggregating	9 10	Mar 11 – Mar 15 Mar 18 – Mar 22	*Show*: *Show*: Groupby, Summarization *Tell*: Cognitive biases, Persuasion	Assignment 5: April 1
	11	Mar 25 – Mar 29	🎉 🎉 🎉 Spring Break 🎉 🎉 🎉	
Inferring	12 13	Apr 1 – Apr 5 Apr 8 – Apr 12	*Show*: Hypothesis testing, Regression *Tell*: Causation, Counterfactuals	Assignment 6: April 15
Extrapolating	14 15	Apr 15 – Apr 19 Apr 22 – Apr 26	*Show*: Forecasting, Prediction *Tell*: Statistical paradoxes, Logical fallacies	Assignment 7: April 29
	16	Apr 29 – May 2	Final Project presentations	May 7