Information Exposition

INFO 3402; Spring 2022

Tuesdays and Thursdays; 11:00 am – 12:15 pm
Eaton Humanities 1B80

Brian Keegan, Ph.D.
Assistant Professor, Information Science
E-mail: brian.keegan@colorado.edu
Office: INFO 129
Office hours: Thursdays, 9:30 am – 10:45 am

Samantha Dalal
Teaching Assistant, Ph.D. Student, Information Science
E-mail: samantha.dalal@colorado.edu
Office hours: Fridays, 11:00 am – 1:00 pm

Course Description

This course teaches students how to communicate the findings from their data analyses and to understand the ethics and implications of data communication and storytelling. Students will develop their computational skills for reshaping, analyzing, summarizing, and visualizing quantitative data using Python's scientific libraries. Students will also develop their communication skills for interpreting, designing, and storytelling the findings of their analyses for general and professional audiences. Students will be evaluated on both their computational and communication skills through quizzes, weekly assignments, blog posts, and a final project. There is no midterm or final exam. Students will:

- Analyze and interpret relationships in quantitative data
- Create effective data visualizations through iterative design
- Communicate findings for general and professional audiences
- Think critically about and critique data narratives and visualizations

Course Design

Class will meet two times per week on Tuesdays and Thursdays from 11:00 am to 12:15 pm in Eaton Humanities 1B80. *Attendance is required*. During class expect to participate in coding exercises and discussions, critique visualizations and narratives, and assist your peers in projects. Tuesday's lecture will introduce a new concept with slides and a notebook and will have exercises to complete in class or take home. The Weekly Assignment will be also be introduced on Tuesday and will be due on Friday. Thursday's lecture will review concepts from Tuesday's class and exercises, have a session where we critique a data narrative or visualization, time to work on the Weekly Assignment, and will end with the Weekly Quiz. The class is split up into six modules for different data communication skills (shaping, distributions, comparisons, trends, relationships, and spatial). There will be a Module Assignment at the end of each Module where students will write a blog post for a general or professional audience communicating the findings of their analysis of a data set.

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

Course Website and Materials
There is no textbook required for class, but there will be required readings, quizzes, and assignments which will be made available and submitted through Canvas:

https://canvas.colorado.edu/courses/79135

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.* through announcements in class and on Canvas. Students should email the instructor if there are any discrepancies or questions.

Publishing
The class will use of the Medium blogging platform. Instruction on how to create accounts, read, write, and post to the class publication will be covered in class. There is extensive documentation in the Medium Help Center as well as multiple tutorials. Students will write their Module Assignments on Medium and submit links via Canvas with the expectation that their writing can be read by the general public. If students are unable or do not want to use the Medium platform, they should email the instructor before Friday, January 21 to work out an alternative arrangement.

Computing
Students will use programming languages for data analysis and visualization. Jupyter notebooks written in Python 3 will be used for all in-class examples and assignments. The Anaconda distribution of Python 3.8 (or above) is *strongly* recommended to provide all of these programs and other libraries. We will be using the Matplotlib and Seaborn libraries for data visualization. Lectures will include exercises and presentations with the expectation that students participate with their own laptop computers. If students cannot bring a laptop to class, they should email the instructor to work out an alternative arrangement.

If students wish to use an alternative programmatic data analysis software (R, Matlab, Julia, *etc.*) or other Python data visualization libraries (Plotly, Altair, Bokeh, *etc.*) they are welcome to do so, but instructional support will only be provided for Python and Matplotlib. Students are *not* permitted to use spreadsheet (Microsoft Excel, Apple Numbers, *etc.*) or business intelligence (Tableau, Microsoft PowerBI, *etc.*) software for assignments. The instructors reserve the right to conduct a code review on any assignment submitted by a student to ensure academic integrity. Students who are unable or unwilling to describe how their submitted code works will lose all credit on the assignment.

Evaluation
Students will be evaluated through four mechanisms. The class has no midterm or final exam.

Weekly Assignments (30%). Weekly Assignments are intended to develop students' skill and confidence applying the technical and expository skills introduced during lecture. There will be a total of 15 Weekly Assignments. Each Weekly Assignment is worth 2% of the final grade (30% cumulative) and are due on Canvas by Sunday before midnight. In the absence of an approved excuse, late submissions will be docked 50% of their value for every day elapsed since the deadline: *assignments submitted after Tuesday before midnight will lose all credit*. The lowest Weekly Assignment grade will be automatically dropped and there are no opportunities for re-grades on assignments.

Weekly Quizzes (15%). Weekly Quizzes are intended to evaluate students' understanding of the concepts from the readings and lecture. There will be a total of 15 Weekly Quizzes. Each Weekly Quiz is worth 1% of the final grade (15% cumulative). The quizzes will be administered via Canvas and will take place

during class on Thursdays. The lowest Weekly Quiz grade will be automatically dropped. There are no opportunities to retake the quiz outside of class.

Module Assignments (30%). Module Assignments are intended to (1) develop students' confidence communicating with data to a general audience and (2) generate a public-facing portfolio. There are six Module Assignments in total, one per module. Each Module Assignment is worth 5% of the final grade (30% cumulative) and are due by 11:59pm the Wednesday after the end of a module. The format and evaluation criteria of each Module Assignment will vary, but will emphasize applying the module's concepts to a novel dataset. Each Module Assignment will be published as a blog post via the class's Medium publication. In the absence of an approved excuse, late submissions will be docked 50% of their value for every day elapsed since the deadline: *assignments submitted after Friday at 11:59pm will lose all credit*.

Final Project (25%). The Final Project is intended to be a portfolio piece highlighting a student's computational and communication skills. The project will be both a data analysis and write-up with the goal of submitting for external publication as an op-ed, guest column, *etc*. Further details about the Final Project will be collaboratively developed and detailed later in the course. In the absence of an approved excuse, late submissions will be docked 50% of their value for every day 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 classmates' writing, code, and other class-related content for 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. Students may not create any audio, video, or other records of lectures without the instructor's permission nor may students share comments made in class attributable to another person's identity without permission.

Classroom Behavior

Both students and faculty are responsible for maintaining an appropriate learning environment in all instructional settings, whether in person, remote or online. 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. For more information, see the policies on class behavior and the student code.

Requirements and Contingencies for COVID-19

As a matter of public health and safety, all members of the CU Boulder community and all visitors to campus must follow university, department and building requirements and all public health orders in place to reduce the risk of spreading infectious disease. Students who fail to adhere to these requirements will be asked to leave class, and students who do not leave class when asked or who refuse to comply with these requirements will be referred to Student Conduct and Conflict Resolution. For more information, see the policy on classroom behavior and the Student Code of Conduct. If you require accommodation because a disability prevents you from fulfilling these safety measures, please see the "Accommodation for Disabilities" statement.

CU Boulder currently requires masks in classrooms and laboratories regardless of vaccination status. This requirement is a precaution to supplement CU Boulder's COVID-19 vaccine requirement. Exemptions include individuals who cannot medically tolerate a face covering, as well as those who are hearing-impaired or otherwise disabled or who are communicating with someone who is hearing-impaired or otherwise disabled and where the ability to see the mouth is essential to communication. If you qualify for a mask-related accommodation, please follow the steps in the "Accommodation for Disabilities" statement on this syllabus. In addition, vaccinated instructional faculty who are engaged in an indoor instructional activity and are separated by at least 6 feet from the nearest person are exempt from wearing masks if they so choose.

If you feel ill and think you might have COVID-19, if you have tested positive for COVID-19, or if you are unvaccinated or partially vaccinated and have been in close contact with someone who has COVID-19, you should stay home and follow the further guidance of the Public Health Office (contacttracing@colorado.edu). If you are fully vaccinated and have been in close contact with someone who has COVID-19, you should self-monitor for symptoms and follow the further guidance of the Public Health Office.

Should a student contract any illness (COVID-19 or otherwise) that requires mandatory sequestration, intensive medical treatment, or extended convalescence and disrupts their ability to participate in class and complete assignments, the instructor will try to accommodate their condition without penalty with extensions and incompletes. This also applies if the student has a family member whose diagnosis, treatment, and recovery will affect their ability to participate. *Please do not ghost us*: students should notify the instructor as soon as possible of events that will impact their engagement with the class so that we can triage and develop an accommodation plan rather than scrambling to solve problems near deadlines at the end of the semester.

Accommodations 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, see Temporary Medical Conditions on the Disability Services website.

Preferred Student Names and Pronouns
CU Boulder recognizes that students' legal information does not always align with how they identify. Students may update their preferred names and pronouns via the student portal; those preferred names and pronouns are listed on instructor's class rosters. In the absence of such updates, the name that appears on the class roster is the student's legal name.

Honor Code
All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code academic integrity policy. Violations of the Honor Code may include, but are not limited to: 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 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 on the Honor Code website.

Harassment and Discrimination
CU Boulder is committed to fostering an inclusive and welcoming learning, working, and living environment. CU's discrimination and harassment policies prohibit acts of sexual misconduct (harassment, exploitation, and assault), intimate partner violence (dating or domestic violence), stalking, or protected-class discrimination or harassment by or against 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 email cureport@colorado.edu. Information about university policies, reporting options, and the support resources can be found on the OIEC website.

Please know that faculty and graduate instructors have a responsibility to inform OIEC when they are made aware of incidents of sexual misconduct, dating and domestic violence, stalking, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about their rights, support resources, and reporting options. To learn more about reporting and support options for a variety of concerns, visit Don't Ignore It.

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.

Acknowledgements

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.

Resources

These are books, papers, documentation, and blog posts that are helpful resources for understanding data cleaning, data visualization, and data communication. Some parts of these resources will be used as readings in class. For books, the links will take you to the CU Boulder Libraries page: click the "Full Text Online" button there to access the book. If you use these resources or other resources you find on the web (StackOverflow, blog posts, code repositories, library documentation, *etc*.) please make sure to reference them in your code, notebook markdown, or other documentation.

There is at least one error below: the link does not match the title of the resource. The *first* student send an email to brian.keegan@colorado.edu with (1) the subject "INFO 3402 Syllabus easter egg", (2) the title of the resource, and (3) the correct permanent link (click the blue chain icon to the right of an item title in CU Libraries OneSearch) will automatically receive full credit on a module assignment (*i.e*., you can skip it). The instructor will email the class when this opportunity closes.

- Data Cleaning and Analysis
  - Chen, D. (2018). *Pandas for Everyone: Python Data Analysis*.
  - Elgabry, O. (2019). "The Ultimate Guide to Data Cleaning." *Towards Data Science*.
  - Harrison, M. & Petrou, T. (2020). *Pandas 1.x Cookbook: Practical Recipes.*
  - Heydt, M. (2017). *Learning Pandas: High-Performance Data Manipulation and Analysis in Python*.
  - Klosterman, S. (2019). *Data Science Projects with Python*.
  - McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, Numpy, and IPython*.
  - Mertz, D. (2021). *Cleaning Data for Effective Data Science*.
  - Molin, S. & Jee, K. (2021). *Hands-On Data Analysis with Pandas*.
  - Nelli, F. (2018). *Python Data Analytics: With Pandas, NumPy, and Matplotlib*.
  - Osborne, J. (2013). *Best Practices in Data Cleaning*.
  - Stepanek, H. (2020). *Thinking in Pandas: How to use the Python Data Analysis Library the Right Way*.
  - Vanderplas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*.
  - Walker, M. (2020). *Python Data Cleaning Cookbook*.
  - Wickham, H. (2014). "Tidy Data." *J. Statistical Software*.
- Data Visualization
  - Berinato, S. (2016). *Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations*.
  - Döbler, M. & Grössmann, T. (2019). *Data Visualization with Python*.
  - Engebretsen, M. (2020). *Data Visualization in Society*.
  - Holtz, Y. (2021). Python Graph Gallery.
  - Jones, B. (2014). *Communicating Data with Tableau: Designing, Developing, & Delivering Data Visualizations*.
  - Keller, B. (2018). *Mastering Matplotlib 2.x: Effective Data Visualization Techniques with Python*.
  - Kirk, A. (2012). *Data Visualization: A Successful Design Process*.
  - Lima, M. (2011). *Visual Complexity: Mapping Patterns of Information*.
  - Matplotlib. (2021). Tutorials, Examples, and Twitter.
  - Meyer, M. & Fisher, D. (2018). *Making Data Visual: A Practical Guide to Using Visualization for Insight*.
  - Milovanović, I., Vettigli, G., & Foures, D. (2015). *Python Data Visualization Cookbook*.
  - Munzner, T. & Maguire, E. (2015). *Visualization Analysis & Design*.
  - Pajankar, A. (2021). *Hands-On Matplotlib: Learning Plotting and Visualizations with Python 3*.

  - Schwabish, J. (2014). "An Economist's Guide to Visualizing Data." *J. Economic Perspectives*.
  - Schwabish, J. (2021). *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks*.
  - Sosulski, K. (2019). *Data Visualization Made Simple: Insights into Becoming Visual*.
  - Tufte, E. (2001). *The Visual Display of Quantitative Information*.
  - Wilke, C. (2019). *Fundamentals of Data Visualization*.
  - Wilkinson, L. & Wills, G. (2005). *The Grammar of Graphics*.
  - Yau, N. (2011). *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*.
  - Yim, A., Chung, C., & Yu. A. (2018). *Matplotlib for Python Developers*.
- Data Communication and Storytelling
  - Abela, A. (2013). *Advanced Presentations by Design: Creating Communication That Drives Action*.
  - Allchin, C. (2021). *Communicating with Data: Making Your Case with Data*.
  - Andrews, R. (2019). *Info We Trust: How to Inspire the World with Data*.
  - Bach, B., *et al*. (2018). "Design Patterns for Data Comics." *CHI*.
  - Blount, T., *et al*. (2020). "Use of Narrative Patterns by Novice Data Storytellers." *CHIRA*.
  - Cairo, A. (2016). *The Truthful Art: Data, Charts, and Maps for Communication*.
  - Cairo, A. (2019). *How Charts Lie: Getting Smarter about Visual Information*.
  - Data Comics for Visual Storytelling.
  - Dykes, B. (2019). *Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals*.
  - Gemignani, Z. (2014). *Data Fluency: Empowering Your Organization with Effective Data Communication*.
  - Gershon, N. & Page, W. (2001). "What Storytelling Can Do for Information Visualization." *CACM*.
  - Jones, B. (2020). *Avoiding Data Pitfalls*.
  - Kahan, D., Jamieson, K. & Scheufele, D. (2017). *Handbook on the Science of Science Communication*.
  - Knaflic, C. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*.
  - Kriebel, A. & Murray, E. (2018). *#MakeoverMonday: Improving How We Visualize and Analyze Data*.
  - Matthews, R. & Wacker, W. (2007). *What's your story?: Storytelling to move markets*.
  - Narrative Patterns for Data-Driven Storytelling.
  - Nolan, D. & Stoudt, S. (2021). *Communicating with Data: The Art of Writing for Data Science*.
  - Ojo, A. & Heravi, B. (2017). "Patterns in Award Winning Data Storytelling." *Digital Journalism*.
  - Riche, N., Hunter, C., Diakopoulos, N., & Carpendale, S. (2018). *Data-Driven Storytelling*.
  - Segel, E. & Heer, J. (2010). "Narrative Visualization: Telling Stories with Data." *TVCG*.
  - Steele, J. & Iliinsky, N. (2010). *Beautiful Visualization: Looking at Data Through the Eyes of Experts*.
  - Swires-Hennessy, E. (2014). *Presenting Data: How to Communicate Your Message Effectively*.
  - Vora, S. (2019). *The Power of Data Storytelling*.
  - Wexler, S. (2021). *The Big Picture: How to Use Data Visualization to Make Better Decisions—Faster*.
  - Yang, L., *et al*. (2021). "Applying the Freytag's Pyramid Structure to Data Stories." *TVCG*.
  - Yau, N. (2013). *Data Points: Visualization that Means Something*.
- Data Journalism and Newsletters
  - Data & Eggs. *The DataFace*.
  - Data Journalism. *Global Investigative Journalism Network*.
  - DataIsBeautiful Subreddit & Twitter.
  - Data Journalism Mailbrew.
  - Freeguard, G. *Warning: Graphic Content*.
  - Gray, J. & Bounegru, L. (2019). *The Data Journalism Handbook*.
  - Hickey, W. Numlock News.
  - Ingraham, C. *The Why Axis*.
  - de Jong, W. ddj.news.
  - Off the Charts. *The Economist*.
  - Singer-Vine, J. *Data Is Plural*.
  - Sollazzo, G. *Quantum of Sollazzo*.
  - Warnes, S. *Fair Warning*.
  - The Week in Data. *Open Data Institute*.

Course Outline

Please consult the schedule online at Canvas for the most up-to-date information.

Week 1 – Shaping: Loading

Tuesday, January 11; Thursday, January 13

Introductions and loading data.

Week 2 – Shaping: Aggregating

Tuesday, January 18; Thursday, January 20

Pivot tables and groupby-aggregation.

Week 3 – Shaping: Combining

Tuesday, January 25; Thursday, January 27

Types of joins and evaluating results of a join.

Week 4 – Shaping: Tidying

Tuesday, February 1; Thursday, February 3

Wide versus tidy data and melting.

Week 5 – Distribution: Histograms

Tuesday, February 8; Thursday, February 10

Counts, cuts, and transformations.

Week 6 – Distribution: Audience

Tuesday, February 15; Thursday, February 17

Understanding and writing for an audience.

Week 7 – Comparison: Cat plots

Tuesday, February 22; Thursday, February 24

Cat plots, hues, and facets.

Week 8 – Comparison: Persuasion

Tuesday, March 1; Thursday, March 3

Strategies for persuading an audience.

Week 9 – Trend: Time series

Tuesday, March 8; Thursday, March 10

Temporal data, time series plots, and forecasting.

Week 10 – Trend: Uncertainty

Tuesday, March 15; Thursday, March 17

Communicating uncertainty about predictions.

Week 11 – Spring Break

Tuesday, March 22; Thursday, March 24

No class!

Week 12 – Relationship: Scatter plots
Tuesday, March 29; Thursday, March 31
Scatter plots, trends, and correlations.

Week 13 – Relationship: Validity
Tuesday, April 5; Thursday, April 7
Types of research validity and counter-factual thinking.

Week 14 – Spatial: Choropleths
Tuesday, April 12; Thursday, April 14
Spatial data, choropleths, and spatial joins.

Week 15 – Spatial: Conventions
Tuesday, April 19; Thursday, April 21
Designing for socio-cultural conventions and highlighting anomalies.

Week 16 – Projects
Tuesday, April 26; Thursday, April 28
Working on final projects.

Module	Week	Dates	Type	Skill
Shaping	1	Jan 11, Jan 13	Computation	Loading
	2	Jan 18, Jan 20	Computation	Aggregating
	3	Jan 25, Jan 27	Computation	Combining
	4	Feb 1, Feb 3	Computation	Tidying
Distribution	5	Feb 8, Feb 10	Computation	Histograms
	6	Feb 15, Feb 17	Communication	Audience
Comparison	7	Feb 22, Feb 24	Computation	Cat plots
	8	Mar 1, Mar 3	Communication	Persuasion
Trend	9	Mar 8, Mar 10	Computation	Time series
	10	Mar 15, Mar 17	Communication	Uncertainty
	11	Mar 22, Mar 24	Spring Break	
Relationship	12	Mar 29, Mar 31	Computation	Scatter plots
	13	Apr 5, Apr 7	Communication	Validity
Spatial	14	Apr 12, Apr 14	Computation	Choropleths
	15	Apr 19, Apr 21	Communication	Conventions
Projects	16	Apr 26, Apr 28	Projects	