JOUR 460
Data Storytelling
SPRING 2021

Margaret Yee Man Ng, Ph.D.
Email: ymn at illinois dot edu
Lectures: Mondays 12:00PM - 01:50PM
Office hour: Mondays after class & Wednesdays 10:00AM-11:30AM

TA: Maggie Mengqing Zhang
Email: mz44 at illinois dot edu
Office hour: Wednesdays 12:00PM-1:50PM

"You want statistics to tell you the truth. You can find truth there if you know how to get at it, and romance, human interest, humor and fascinating revelations as well." — Joseph Pulitzer

Course Description:

Data has become an invaluable resource for journalists to expose stories hidden in the numbers and produce great stories. However, like any sources, data sets should be treated with skepticism and ethical considerations. This class introduces students to the development process and ethical guidelines for creating data stories: acquire, clean, analyze, and visualize data using various tools of data journalism.

Prerequisites:

Basic programming skills are preferred, as some course materials are built on Python and Jupyter Notebook. However, the most important prerequisite is a willingness to work hard on possibly unfamiliar materials. If you want to be prepared with Python, please visit Data Quest (this is a site I used when I first got started with coding) and go through some of the interactive exercises. I found the *Python for Data Science: Fundamentals class* looks pretty short but to the point. And it is free for most of the basic materials. If you go through most of the materials in that class, you can pretty much handle everything in the class.

Credit Hours: 4

Course Objectives:

At the completion of this course, students will be able to:

1. Understand the history and emerging trends of data journalism.
2. Recognize various sources of data and how to negotiate access to data with officials by using FOIA requests.
3. Spot errors and deal with missing values and messy data.
4. Acquire, clean, and analyze different types of data by using various tools of data journalism (Excel, Open Refine, SQL, Python).
5. Ask interesting and critical questions of data.
6. Design basics, effective data visualization.

7. Evaluate effectiveness of data-based storytelling projects.

Books & Resources:

- (REQUIRED) Houston, B. (2018). *Data for Journalists (5th Edition)* Cheshire, CT: Graphics Press. ISBN: 978-0815370406

- (Recommended) Healy, K. (2019), *Data Visualization: A Practical Introduction* Princeton, NJ: Princeton University Press. http://socviz.co/.

- You should scan the Web each week for stories involving data analysis. The IRE and NICAR website, along with Fivethirtyeight and ProPublica are some good places to start.

Grade Distribution:

Here is a very brief description of each of the grading component. Please refer to Compass for details.

Participation	10%
Festival of Data	5%
Checkpoints	25%
Assignment 1, 2, 3	25%
FOIA Request	15%
Final project	20%

- **Participation (10%)**: Students are expected to attend Monday class sessions, participate in discussions, and ask questions. Attendance during the weeks of guest speakers is strictly required. See below for my attendance policy.

- **Festival of Data (5%)**: Each student will choose a recent data driven story to present in class (5-10 minutes). Prepare to discuss the use of data and visualization, the strengths and weaknesses of the story. Post the story and discussion (250 words) in the "Festival of Data" Discussion Forum on Compass a day before the presentation.

- **Checkpoints (25%)**: There are six (6) checkpoints across the semester. Only the top five (5) checkpoints with higher scores will be included in the final grade. Each checkpoint will have around 10 - 15 multiple-choice questions. They are all "open book."

- **Assignment 1, 2, & 3: (25%)**: Data analysis with Excel, SQL, and Python.

- **FOIA Request (15%)**: Over the course of the first half of the semester, you will identify and negotiate to acquire a database from a local/state public agency. File a Freedom of Information Act (FOIA) or other public records request to a public-funded agency, then follow up with needed negotiations. Success will not depend on getting the data; instead, what counts is the evidence of your thoroughness and persistence. This assignment will include a two-page report on the process, plus documentation.

- **Proposal, Presentation & Final Project (20%)**: Using the database gathered from FOIA request, students will do a data project. You will write a proposal about what specific questions you are looking at. You will do your storytelling through Jupyter notebook with the help of data analysis/data visualization. Students will receive reviews from peer to further develop and improve their work before submitting their final work.

- **Graduate students are required to do additional readings and writings (20%).** Details will be announced later in the semester.

Letter Grade Distribution:

Grades in the **A** range represent work that is **excellent**; Grades in the **B** range represent performance that is **meets expectations**. This class does not offer "incomplete" as grades.

>= 97.00	A+		
93.00 - 96.99	A	73.00 - 76.99	C
90.00 - 92.99	A-	70.00 - 72.99	C-
87.00 - 89.99	B+	67.00 - 69.99	D+
83.00 - 86.99	B	63.00 - 66.99	D
80.00 - 82.99	B-	60.00 - 62.99	D-
77.00 - 79.99	C+	<= 59.99	F

Software & Platforms:

We'll be using a handful of free and open source software tools this semester:

- Excel
- Open Refine
- SQLite Manager
- Python with Jupyter Notebook/Google Colab (e.g., Pandas, Plotly, etc.)
- Text editor (e.g., Sublime Text or Atom)

Course Policy:

- **A Culture of Respect** We have students from different departments and have different comfort levels with data/coding. For students who are hesitant and know they need a lot of personal care, I am always happy to give you more resources to learn the topics; for students who are ahead in coding skills, the coding sessions may be pretty basic for you. But I see the benefit of a class that provides you space, time, and accountability to work on a data project.

- **Deadlines Matter:** Deadlines are of vital importance, not only for the class but also in the professional career of any journalists. Deadlines are sacred. Please carefully note these rule: Deadlines for Checkpoints (1-6) and Assignment 1-3 are strict as answers will be discussed in class the next day right away. For other assignments, students may submit assigned work up to five (5) days after the stated deadline, but there will be a 10 percent deduction of your assignment grade for every 24 hours that passes after a deadline in which you have not turned it in. Make-up work will not be offered except in extenuating circumstances (medical emergency or family emergency).

- **Attendance:** Students are expected to attend all Monday class sessions, participate in discussions, and ask questions. I have found that the best classes are where students are

engaged in their learning, even virtual. I realize that you will sometimes miss class, so you get three (3) unexcused absences. If you have four (4) unexcused absences, your final grade will drop one full letter grade. If you have five (5) unexcused absences you will either be dropped from the class or receive an F for your final grade. **However, attendance during the weeks of guest speakers is strictly required and does not counted towards those three (3) unexcused absences.**

- **Academic Integrity:** Discussion amongst students is encouraged, but **offering** and **accepting** solutions from others is an act of **cheating**. You are expected to be familiar with and to follow the UIUC Student Code in all matters related to this course. Specifically I call your attention to Part 4, "Academic Integrity Policy and Procedure"and to Section1-401(b), which states: "Students have been given notice of this rule by virtue of its publication. Regardless of whether a student has actually read this rule, a student is charged with knowledge of it..." I do not tolerate plagiarism.

JOUR 460 Data Storytelling Spring 2021

This class provides synchronous lectures (every Monday) and asynchronous content. Schedule is subjected to change, depending on how well students are progressing.

	Readings	Lectures	Self-guided materials	Homework
Orientation Week 1 introduces practices of data journalism.				
Week 1 (January 24 - 30)		What is Data? What is Data Journalism?	Course Overview	Syllabus Quiz Due by Sunday Jan 26, 5 PM Getting to Know Your Activity Throughout Week 1
Module 1: Finding & Cleaning data Module 1 shows students how to find reliable data, particularly through FOIA requests. We will learn the basic of spreadsheets and the workflow of cleaning messy data.				
Week 2 (January 31 - February 6)	*Data for Journalists* Chapter 1, 2, 3	Wrestling with Dirty Data + Open Refine	Freedom of Information Act	Checkpoint 1: FOIA Due by Sunday Feb 7, 5 PM
Week 3 (February 7 - 13)	*Data for Journalists* Chapter 8, 10	Guest Speaker: Pamela Dempsey on FOIA requests	Basic Excel (Sort, Filter, & Formulas)	Checkpoint 2: Excel Due by Sunday Feb 14, 5 PM
Module 2: Analyzing Data Like interviewing people, we interview data to look for newsworthy trends or events. Besides Excel, this module discusses the technical details of SQL				
Week 4 (February 14 - 20)		Advanced Excel (Pivot Table)	Interview Data + Basic Statistics	Assign. 1 on Excel Due by Sunday Feb 21, 5 PM
Week 5 (February 21 - 27)	*Data for Journalists* Chapter 4, 5, 9	SQL Queries I (Syntax, WHERE, GROUP BY, ORDER BY)	SQL Queries II (Wildcard Characters)	Checkpoint 3: SQL — Deer Hunting Accidents Due by Sunday Feb 28, 5 PM
Week 6 (February 28 - March 6)	*Data for Journalists* Chapter 6, 7	SQL Queries IV (JOIN)		Assign. 2 on SQL Due by Sunday Mar 7, 5 PM

Module 3: Visualizing Data
Module 3 reviews the history and emerging trends of data visualization in journalism. It also introduces types of charts, Tufte's Criteria for Good Visual Information Representation, and theories of colors and visual perception.

Week 7 (March 7 - 13)		Intro to Python	Data Visualization in Journalism	FOIA Request Soft Deadline Due by Sunday Mar 14, 5 PM
Week 8 (March 14 - 20)	Break		Types of Charts and Graphs & Visual Perception	Checkpoint 4: Chart Type & Visual Perception Due by Sunday Mar 21, 5 PM
Week 9 (March 21 - 27)		Data Wrangling with Pandas I (Data Type)		
Week 10 (March 28 - April 3)		Data Wrangling with Pandas II (JOIN & MERGE)	What Makes a Good Data Visualization?	Assign. 3 on Python Due by Wednesday Mar 31, 5 PM FOIA Request and Proposal Due by Sunday Apr 4, 5 PM

Module 4: Ethics and the Future
Module 4 explores key concepts in ethics for data journalism and talks about other burgeoning computational technology used in the field.

Week 11 (April 4 - 10)		Individual meeting	How Color Communicates Meaning	Checkpoint 5: Visualization Design Due by Sunday Apr 11, 5 PM
Week 12 (April 11 - 17)		Guest speaker - Jeremy CF Lin (Bloomberg) + Basic Visualization with Python	Ethical Dilemmas in Data Journalism	Checkpoint 6: Ethics Due by Sunday Apr 18, 5 PM
Week 13 (April 18 - 24)		Computational Journalism		
Wrap up & Presentation				
Week 14 (April 26)		Final Project Presentation		
Week 15	Peer Review due by May 2 Sunday at 5 p.m.			
Week 16	Revised Final Project due by May 9 Sunday at 5 p.m.			