

J298 Data Journalism, Spring 2018

This course is for students who want to make finding and reporting stories from data part of their toolkit. It will be useful for anyone interested in investigative journalism, which nowadays is often heavily data-driven, or those keen to use data to provide context and ground-truth for regular beat reporting. You should be comfortable with numbers, and thinking critically and quantitatively. You must be prepared to work with some simple code (in R and SQL), and to get your hands dirty with real-life, messy data!

Weekly schedule

We will meet in **108/Lower NG** on Thursdays from 6pm - 9pm. Your instructors, **Peter Aldhous** and **Amanda Hickman**, will maintain office hours over Skype. You are encouraged to arrange appointments to discuss your work.

- **Jan. 18: What is data?**

Categorical and continuous variables; basic operations for interviewing a dataset; sampling and margins of error; plotting and summarizing distributions; choosing bins for your data; basic newsroom math. (**Peter Aldhous**)

Assignment: File a preliminary pitch for a data-driven story. Also, read stories for discussion in week 2.

- **Jan. 25: Data journalism: Types of stories**

Through a tour of some great examples of data journalism, we'll get inspiration for our work in this course. (**Peter Aldhous**)

Assignment: Quiz on good practice in data analysis and data journalism.

- **Feb. 1: Working with spreadsheets**

Understanding spreadsheet functions and pivot tables is the foundation for the rest of your exploration of data. We'll learn how to troubleshoot and use spreadsheets. (**Amanda Hickman**)

Assignment Revise your pitch for a data-driven story.

- **Feb. 8: Acquiring, cleaning, and formatting data**

Where and how to find data online. Tips and tricks for downloading unruly data, including browser extensions to extract data from web tables and download from multiple files *en masse*. (**Peter Aldhous**)

Assignment Data cleaning exercise.

- **Feb. 15: R, R Studio, and the tidyverse**

Introduction to [R](#), [R Studio](#) and the [tidyverse](#) packages for data journalism. (**Peter Aldhous**)

Assignment Exercises to reinforce concepts learned in class.

- **Feb. 22: Data journalism in the tidyverse**

We'll use tidyverse packages to explore data on opioid prescription under Medicare in California, and related datasets on the doctors involved. (**Peter Aldhous**)

Assignment Further exercises with the data from class.

- **Mar. 1: Don't let data lie to you**

Data is as full of lies as people, but somehow we are inclined to believe numbers in ways we wouldn't believe sentences. We'll look at ways that numbers lie to us and people lie to each other with numbers. (**Amanda Hickman**)

Assignment Submit a reporting and data analysis plan, based on your earlier pitch.

- **Mar. 8: Introduction to databases and SQL**

We'll work [PostgreSQL](#) using SQL, to ask questions of data. (**Amanda Hickman**)

Assignment Skill reinforcement assignment in PostgreSQL.

- **Mar. 15: Finding stories using maps**

[QGIS](#) is a desktop geographic information system (or GIS) application that we'll use to view, edit, and analyze geographic data. (**Amanda Hickman**)

Assignment Skill reinforcement assignment in QGIS.

- **Mar. 22: Maps meet databases**

[PostGIS](#) adds support for geographic queries and objects to Postgres – it is a powerful tool for geographic analysis, and it plugs right into QGIS. We'll explore more advanced queries that take advantage of PostGIS's power. We'll also discuss the tool or strategy you'd like to tackle in week 12. (**Amanda Hickman**)

Assignment Skill reinforcement assignment in PostGIS.

- **Mar. 29: Spring Break – No class**

- **Apr. 5: More with fun with R**

Connecting to databases from R. Pulling data from the web. Correcting for inflation and seasonal trends. (**Peter Aldhous**)

Assignment Submit a progress report on your data analysis.

- **Apr. 12: Open class: To be determined**

As the semester progresses, we want to hear from you about tools you're interested in tackling. We're leaving week 12 open to new opportunities and inspiration that present themselves this semester. (**Amanda Hickman**)

Assignment Skill reinforcement assignment.

- **Apr. 19: Putting it all into practice: 1**

Drawing from the headlines, we'll attempt some original reporting from data. We may emerge with a story. We may crash and burn. The goal is for you to experience how this works in an actual newsroom, under deadline pressure. (**Peter Aldhous and Amanda Hickman**)

- **Apr. 26: Putting it all into practice: 2**

Continuation from last week's class. (**Peter Aldhous and Amanda Hickman**)

Assignment Submit your data journalism project.

- **May 4: Next steps: The cutting edge of data journalism**

Show and tell on advanced methods on data journalism, including machine learning and text analysis. (**Peter Aldhous and Amanda Hickman**)

Recommended reading

Sarah Cohen: *Numbers in the Newsroom: Using Math and Statistics in News*

Philip Meyer: *Precision Journalism: A Reporter's Introduction to Social Science Methods*

Attendance

Unexcused absence from two classes will drop you one letter grade; a third unexcused absence will result in an F. Excused absences will be permitted only in extraordinary circumstances. Regardless of the reason for an absence, students will be responsible for any assignments due and for learning material covered in class.

Grading

Class participation, weekly assignments: **90%** Attendance: **10%**

Good manners

Students must turn off the ringers on their cell phones before class begins. Students may not check e-mail, social media sites or other websites during lecture portions of class or while working on class exercises.

Academic dishonesty and plagiarism

The high academic standard at the University of California, Berkeley, is reflected in each degree that is awarded. As a result, it is up to every student to maintain this high standard by ensuring that all academic work reflects his/her own ideas or properly attributes the ideas to the original sources.

These are some basic expectations of students with regards to academic integrity:

- Any work submitted should be your own individual thoughts, and should not have been submitted for credit in another course unless you have prior written permission to re-use it in this course from this instructor.
- All assignments must use “proper attribution,” meaning that you have identified the original source of words or ideas that you reproduce or use in your assignment. This includes drafts and homework assignments!
- If you are unclear about expectations, ask your instructor.

Disability accommodations

If you need disability-related accommodations in this class, if you have emergency medical information you wish to share with the instructor, or if you need special arrangements in case the building must be evacuated, please inform the instructors as soon as possible by seeing one of us after class or making an appointment to visit during office hours. If you are not currently listed with DSP (Disabled Students' Program) but believe that you could benefit from their support, we encourage you to [apply online](#).