Empirical Reporting

(Otherwise known as MCO 510, Data Journalism, Spring 2024)

2023-12-22

Table of contents

CLASS DETAILS

M-W 11:30-1:20p Cronkite 404b (behind the elevators) **Canvas**: https://asu.instructure.com/courses/171974 **Slack Workspace**: https://asu-2241-mco510-20559.slack.com

Sarah Cohen sarah.h.cohen@asu.edu Slack: @shcohen1

I'm available to work with you almost any time Mondays through Thursdays except 9:00-1:30 on class days - Just let me know in advance when you'd like to meet. We can talk via Zoom or in person. I will also hold some study halls during the semester as you want them.

Cronkite policies, including the Diversity Principles, disability and religious accommodations, policies on sexual violence and harassment, and the Cronkite Academic Integrity Policy, are included on left-hand menu in the Canvas site for this and all Cronkite classes. These policies apply to all students in all classes.

Description

This is a class in data journalism focused the systematic use of public records, documents and other material in public affairs reporting. Previous generations might have called it "computer-assisted reporting" or "precision journalism". Whatever the name, the idea is the same: original reporting and analysis in the service of stories, particularly in the accountability and explanatory genres.

Figure 1: Chris Wormhoudt

Analyzing records can help you confirm or disprove a tip, isolate anecdotes, and explore stories or patterns that no one has thought to mention. When you use data-driven techniques, your street reporting will become deeper and more enlightening. Better yet, the techniques you'll learn in this class will often yield exclusive stories.

By the end of this semester, you will be able to:

- Identify how and when analysis of electronic records might strengthen your story, provide important insights, or suggest new lines of reporting.
- Apply the R programming language for reporting in efficient, effective and ethical ways that avoid errors.
- Integrate data journalism techniques with traditional reporting methods.
- Write effectively with and about data, whether for text or for the ear.

Prerequisites

This is a required course for the Masters' of Arts in Investiative Journalism at Cronkite. It is tightly coordinated with the rest of the curriculum, and depends on your having mastered the work done in Boot Camp in the fall semester. If you have not done this, you'll need to review on your own:

- Spreadsheet basics: Simple calculations, sorting, filtering and pivot tables. There are videos and exercises in the MAIJ textbook in case you need a refresher.
- Ethics Standard journalistic ethics as practiced in the United States.
- Basic numeracy: As graduate students, we expect you to come to school with basic numeracy skills such as computing percents or averages. There are materials provided to you if you need refreshers.

You are not expected to have acquired or used data for your stories in the past, or to have ever tried to write any computer code in any language.

Required materials and texts

All of the hardware and software required for class is available in our classroom, in the lab in Cronkite 320, and in the ASU lab in UCENT across from Cronkite.

If you choose to use your own laptop or desktop computer, it must be running Mac OS 11 (Big Sur) or newer[1], or Windows 10 or newer. Please contact me if you're not sure if your operating system needs updating. Macs that are more than 10 years old won't work. Neither will Chrome books or iPads.

I suggest working on a large screen with a regular mouse whenever you can, even if it means doing some of your work in a lab.

[1] Macs made in the past decade will generally work

Books and software

You do not need to buy any books or software.

The following free software will be used extensively in this class.

- The Slack app (available in Apple app stores, Android stores and Windows store). When possible, use the app rather than the version in your browser. The browser lacks some of the functions that are in the app.
- R and RStudio, both free community versions. You don't need to install them yet. We're going to do that together.

At the end of the semester, we'll be exploring some of the advanced features of Microsoft Excel and how they differ from Google sheets, but you should feel free to use lab computers for this.

There will be other free software that we'll use for specific projects. I'll provide links to those as we go along.

All of the readings and materials will be provided to you at no cost.

Recommended

Books you may already have

The Art of Access: Strategies for Acquiring Public Records, David Cullier and Charles N. Davis, 2nd edition (2018).

Investigative Reporters' Handbook: A Guide to Documents, Databases, and Techniques, 6th Edition , by Brant Houston and Mark Horvit, 2021. Available for rent or purchase from MacMillan Learning.

Other data journalism texts

These generally use SQL (Structured Query Language) rather than R for more advanced data work, but the concepts are really good:

- Data for journalists: A practical guide ($$$), Brant Houston: The GOAT of data journalism texts, goes through a lot of the basics in a way that focuses on story rather than presentation.
- Data plus journalism ($$$), Mike Reilley and Samantha Sunne. Both of these authors take a kitchen-sink approach to data and journalism, and the book is sort of the same, but it has good exercises and a laundry list of resources in each chapter. Its organization is a bit muddled.
- The data journalism handbook 2" (Free), European Journalism Centre and Google News Initiative. This free book is a thoughtful collection of essays about the international movement.

I'll provide other resources for learning R for journalism when we get to that part of the course.

Grading

The class uses a point system that initially adds up to 1,000 points for the semester. It's possible this number of total points may fall if we find there are just too many assignments. If so, I'll provide an optional assignment during the exam period to make up the difference.

Daily work: 400

Class preparation and labs: 300

"What would you do?" posts and discussions : 100

Graded homework: 150

Projects : 450

Data replication project: 200

Final story memo : 250

All work will be submitted and feedback reviewed through Canvas based on the due date shown there in Mountain Standard Time.

Daily work

400 points

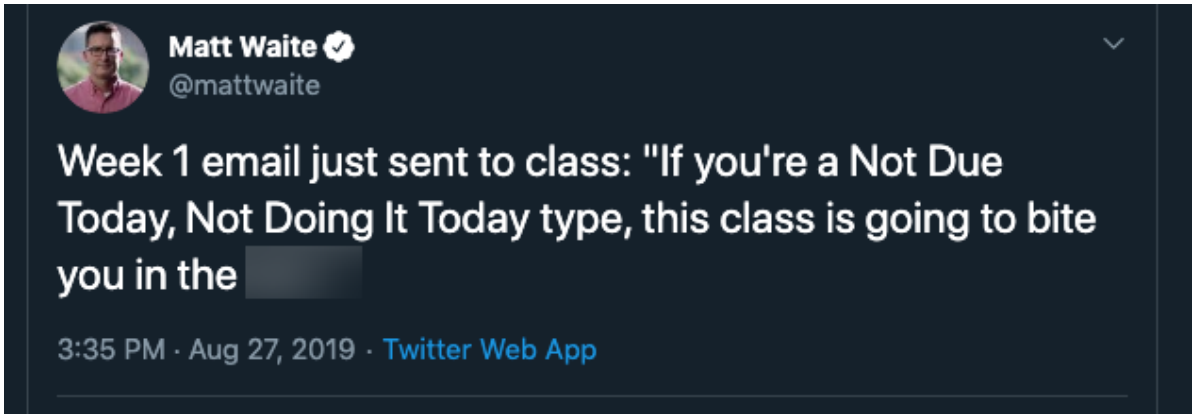
Labs, pre-labs and class preparation

Doing well in this class means working slowly, methodically and consistently throughout the semester.

Nearly half of your grade is dependent on this slow and steady work.

There will be at least one low-stakes assignment nearly every week. Sometimes, I'll ask you to reflect on what you did and provide feedback. Other times, you will show that you went through a tutorial on your own. We'll do a lot of labs in class.

Prelabs are exercises intended to force you to slow down and go through a tutorial on your own, providing feedback on areas that are muddy or difficult. I'll use them as diagnostics to decide how to focus our class time. *Labs* are exercises for you to practice and apply what you've learned.



These will be submitted in Canvas assignments, but I haven't yet produced all of them. This means the total points in Canvas won't be correct until partway through the semester.

What would you do?

You've had some background in journalism ethics, but there are specific issues that come up surrounding acquiring, analyzing and presenting data. You will be presented with four scenarios that are based on real-world situations. After reading the material, you will be asked to reflect on some specific questions on a discussion board, which will form the basis of our in-class discussion. Each of these is worth 25 points.

Replication project

200 points

There are three separate assignments for this project, which challenges you to reproduce some sentences published in real stories using the same data used by the reporters. I'll provide the stories and the data. Your first job is to identify paragraphs in the story that depended on the dataset that you've been given. The second gives you a chance to turn in a draft of your work so that I can help you. The last one is the final version.

Story memo

250 points

You'll develop a story pitch using a combination of story backgrounding and data analysis. I'll give you a choice of at least three datasets that I know will work for this assignment. You may propose your own, but it must be easy to obtain, relatively clean, and complicated or large enough that there is something for you to analyze beyond a basic statistic. Generally, the

dataset will be of individual records (accidents, inspections, deaths or events), not statistical compilations. You'll understand this better as we move along.

The specific list of datasets may be coordinated with your investigative reporting class, News21 and/or the Howard Center, so they're subject to change.

Although this is one-quarter of your grade, it's split among four assignments over more than a month. You'll build the memo a little at a time, so by the end, it won't seem like a big lift. I promise.

Graded homework

150 points

There will be three smaller graded assignments that are more important than daily work but don't rise to the level of a project. They are shown in Canvas each with 50 points.

Grading policies

50 points or more

Any assignment worth 50 points or more will be graded with letters, ranging from E to A+. It follows ASU's policy on grades. They will not be rounded up.

ASU's policy is:

- "A" is reserved for "excellent" or "outstanding" work, which goes beyond the explicit requirements in quality, depth, effort, or presentation.
- "B" means "successful". That means it's on time, accurate, written or presented in a professional way, and answers the requirements of the assignment.
- "C" is considered "Acceptable" work, in that it has met some, but not all, of the requirements.

Under 50 points

Smaller assignments and labs will be graded as follows:

Percentage	Definition
115%	I will award extra points when the work shows extraordinary effort, thought and engagement. You can do everything right, receive an A+ in the course, and never see this. Instead, it's a way for me to acknowledge unusually ambitious work that goes well beyond the original assignment.

Percentage	Definition
100%	You have completed all of the work assigned at an acceptable level of quality, and submitted it according to the instructions by the deadline. This might include running code, writing code, providing questions or feedback, or summarizing your results as a journalist. The work shows no fundamental factual errors and reflects an attempt to apply the material at hand. Sometimes, you'll end up explaining in your homework what you tried to accomplish but couldn't quite succeed — that's ok! I hope all of your assignments can acheive this level but understand that it's not always possible.
50%	You turned in a late or partial assignment or one that shows cursory review of the material. I hope these will be rare, but I expect most people will have one or two of these throughout the semester.
0%	You didn't turn anything in, or what was turned had no meaningful work done.

Course requirements

Showing up is 80 percent of life.

Woody Allen

This is an immersion, in-person class and is not offered via Zoom.

The minimal requirement is to arrive on time, prepared, and ready to work. Be sure to refresh your memory before class begins if you can't recall what you wrote or read in preparation.

Please refrain from eating, napping, using social media, shopping, or working on something else. I expect you to treat your colleagues with respect and be polite to any guests. The classroom is a strict no-eating zone.

Devices must be closed and turned off for guests and used only for class work the rest of the time. Research shows that you are much worse at multi-tasking than you think you are, and the class won't stop because you got sidetracked.

> **!** Devices in class
>
> I am easily distracted when someone is disengaged or doing something else during class. I can't stop staring at someone who nods off, is looking at their phone, or is eating lunch. I know this is a character flaw. I will try to overcome it, but if I start losing my place or stumbling over simple questions, I may ask you to leave the classroom or close your computer until you are ready to engage with the material we're working on.

I will not monitor messages in Slack or email during or just before class time. The only reliable way to get me a message between 11am and 1:20p Mondays and Wednesdays is to do so in

person in the classroom. If you need to get between those hours, please ask a classmate to get my attention.

Attendance and deadlines

There are no excused absences in this course, except for true medical or other emergencies or religious observances.

You might decide it's more important to go to a job interview, attend a family event, interview a source, or schedule a tight connection on a flight than to attend class or complete your work. If I were you, I might decide the same thing, but these are choices that reflect your priorities. They're not emergencies, nor are they excused absences.

As in all journalism work, you are expected to meet deadlines. Late assignments will be heavily penalized and those with small point counts may not be accepted at all.

That said, **give yourself a break**. There are **two no-questions-asked 48-hour extensions** on daily work during the semester. This includes only the labs/prelabs and graded homework, not any part of the project or "What would you do?" assignments. You don't a reason; just let me know before class or deadline.

> **i** May I be excused?
>
> The short answer is "yes". I don't take attendance and there is no class participation grade. I get it — sometimes you don't feel 100% or you have something more important to do or the bus broke down.
> But when you make that choice, you have to catch up on your own. I can't offer a Zoom option, and you'll receive a zero for any missed deadline.

Please advise me the first week of class of any ASU-endorsed absences such as a religious holiday or participation in an ASU team sport that will cause you to miss class so that I can plan accommodations.

Here's what to do in the event of an emergency or illness:

- If there is an emergency, such as a car accident or a death in your close circle or family, please let me know as soon as possible and we'll work out what to do next. (If you've already worked with the dean of students, I can coordinate with them and you don't have to talk to me individually.)
- If you are sick, you should contact ASU Health Services or your doctor. They can provide documentation if you are too ill or contagious to leave home.

Accommodations in these circumstances can include extending deadlines and helping you find assistance such as ASU mental health resources and tutoring. I may be able to let you observe (without participating) via Zoom, but you shouldn't count on it.

Accuracy

If you pursue reporting with data-driven techniques, you'll often create original information that no one else can confirm. In fact, your sources are often happy to explain a result you present to them, even if it's entirely wrong. This puts an even bigger burden on you to understand your sources, to question your assumptions and find holes in your work. If an answer seems too newsworthy[2], walk through everything you've done step by step. Seek out errors; ask experts or colleagues to challenge your results. We'll spend a lot of time on practices that will help protect you from errors but your own skepticism (and sometimes even fear) is the best defense.

I don't expect your work to be polished and exact. However, work that shows a fundamental misunderstanding of the source, contains a severe miscalculation or misinterpretation, or ignores clear warning signs will receive an E. This is the equivalent of having to retract an entire story because it was based on an obvious error.

This requires you to gut-check all of your data work before you turn it in or include it in a sentence. Actually look at your results and decide "could this possibly be true?"

Having trouble?

Figure 2: by Fabian Moller

Learning data skills requires a lot of trial and error – a LOT of error. It's frustrating and maddening at times and you won't be alone. But there are some strategies that you can use to help get you past roadblocks.

[2]In the St. Pete Times story, the Labrador was nowhere in the top 10 list, and at the time small dogs were pretty rare.

- Slow down! Unlike the rest of journalism, going faster doesn't help. Separate the problem into pieces and solve it one step at a time.
- Don't get upset over errors. They don't reflect some kind of flaw in your work – they're supposed to be helpful. Try to read the error and piece together what's wrong. A story goes through many drafts before it's finished, and the same thing will happen with your programs.
- If you've been working on something for more than 20 minutes **without making any progress**, don't waste more time on it. If you're making just a little progress, keep going. But take a break and come back to it if you're stuck. Get help if you're really stuck.
- Remember the goal: We're not trying to become social scientists, computer programmers or data analysts. We're journalists. Simplify your questions, focusing on newsworthiness rather than academic research.
- Practice. Look for opportunities to hone your skills by applying them outside of class on a personal project or for a story or another course. If you can't think of any practice, try a #tidytuesday challenge.[3]

Cronkite school policies

Collaborations

All work submitted under your name must be your work and your work alone. When you are assigned work in teams, contributors should be credited when an individual was primarily responsible for a portion of the work. For example, if one student takes the lead on analyzing data and another on writing the results, you should give credit where it is due.

I will encourage you to work in pairs during some of the labs — it's much easier to solve problems when two people are looking at the same thing. But you must turn in your work separately and answer question prompts on your own.

Extra Credit

Simple. There is none. Why? This course is made up of a lot of little pieces. They can't be replaced by attending a lecture.

[3] In principle, I object to these "play with a dataset" exercises. As you'll see, understanding the data takes time and research, and in these cases the experts just assume they know. But they're useful for honing your skills.

Cronkite School Diversity Principles

The Walter Cronkite School of Journalism and Mass Communication embraces inclusivity in student, staff and faculty populations in order to create an academic environment that fosters diversity of thought and acceptance of all people regardless of race, gender, age, sexual orientation, or societal, political, cultural, economic, spiritual, or physical differences. To this end, the school directs efforts to the following four principles:

- Actively seek out and encourage diverse populations to become productive members of the faculty, staff and the student body;
- Create and maintain a work, learning and social environment that is cognizant and supportive of a diversity of human differences and beliefs;
- Incorporate within the formal content of the curriculum and in each course an affirmation of the core journalistic values of accuracy, fairness, ethical behavior and sensitivity when reflecting an increasingly multicultural world;
- Foster and support a climate in which events and activities of the school reflect diversity of awareness, sensitivity to and support for people of different origins, orientations and abilities.

Student Accessibility and Inclusive Learning Services (SAILS)

Student Accessibility and Inclusive Learning Services facilitates accommodations for students who have registered with that office. Accommodations are determined on a case-by-case, course-by-course basis through an interactive process with people who can best support student success. Prior to receiving accommodations from their instructor, students must receive verification of eligibility by registering with SAILS. Student SAILS information is confidential.

Academic Integrity

The school has zero tolerance for academic dishonesty; instructors enforce academic integrity in every course and educational activity offered or sanctioned by the school. Any allegation of academic dishonesty will be referred to an appointed academic integrity officer for investigation. They elevate cases to the school's standards committee for review and recommendation to the dean of the school. If the committee finds that a student has engaged in academic dishonesty in any form — including but not limited to cheating, plagiarizing, and fabricating — that student shall receive a grade of XE for the class and may face suspension or expulsion from ASU.

You must complete the Academic Integrity pledge within the first week of class, included in Canvas as a quiz.

::: {style="page-break-after: always;"} ::: {.content-visible when-format="html"}

Brodeur via Unsplash

:::

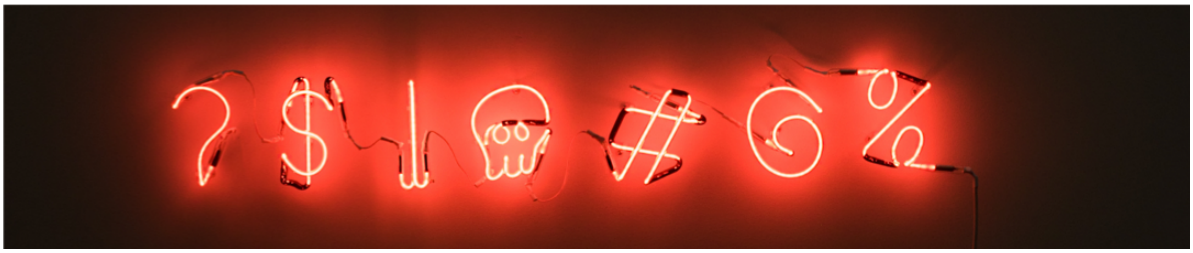


Figure 3: by Brodeur via Unsplash

FAQ

Q: I don't have any experience. What do I need to succeed?

Everyone in this class has gone through the boot camp required of all masters students. You covered spreadsheet basics and were expected to have basic numeracy skills such as computing percentages and averages. No other coding or data analysis background is expected.

Q: This looks like a lot of work. How long do I have to spend on it every week?

I try to follow the ASU standard for a 3-credit course — approximately 9 hours per week, including class and lab time. That suggests that, on average, you should expect to spend about 5 or 6 hours per week outside of class on preparation, homework and projects.

If you're spending considerably more time on assigned work and preparation, please let me know – that's not the intention.

Q: How should I communicate with you? A: Slack

Slack or in person is best. All assignments must be turned in via Canvas assignments unless I direct you elsewhere. Do NOT email assignments.

Q: How will I know how I'm doing?

Pay attention to the feedback on your routine homework and labs. I'll give you a sense of where you're excelling and where you might improve. I'm always happy to meet and discuss strategies that will help you get to the next level in your journalism or your data analysis skills.

I've found that Canvas isn't very accurate in estimating the final grade as we work through the semester, so I've turned that feature off. It will just mislead you.

Q: How do I ask good questions?

In technical work, asking a good question is an art. Try to:

- Describe what you are trying to do – what does success look like? What question are you trying to answer?
- Provide a snippet of the code or a screenshot of what you are trying. Include any context needed.
- Take a screen shot or quote the error message if there is one

Q: I already know how to code and analyze data. Why do I have to take this class?

Even if you know how to code, I presume you are at Cronkite because you want to apply all of your skills to a journalism career. You probably knew how to use a phone or take a picture before you came to j-school, but you may not have used them the way you do now. That said, talk with me if you're concerned that you won't be challenged enough or want to go further. We'll find ways to make the time productive and take you to the next level in data reporting.

Q: Can I use AI like ChatGPT to help me with my coding?

Sure — for help, not for answers. In fact, one of the things it's really good at is answering the question, "Why doesn't this code work?" You will have to determine if its answer is right. But for this purpose it's pretty similar to searching Google and Substack or asking colleague for help.

You are responsible for learning the methods taught in the class, not some other way to accomplish the same thing. (You don't have to memorize anything, but we're working in very limited version of the world of programming and R, and I won't go outside it to help you or accept your work.)

The free version of ChatGPT is a little too old for some of the things you'll be learning. Consider trying Perplexity.ai or Github Copilot instead.