JOU 4306: Advanced Data Journalism

Spring 2026 | 1:55 p.m. to 4:55 p.m. Wednesdays | Weimer 2052

Lecturer Gavin Off
Email: gavin.off@ufl.edu
Office hours: Wednesdays and Thursdays in room 1200A no appointment. Any day via Zoom

MY EXPERIENCE:

I landed my first newspaper job in 2002. It was with a small paper in southwest Florida. I covered traffic accidents, local governments, police and everything in between. But by 2006, I was ready to move on and applied for a reporter position at the Louisville Courier-Journal in Kentucky. Two weeks later, I called and asked the editor if he received my resume. He said he got my resume and 200 others. Needless to say, I did not get the job.

Geez, I thought. What could I do to separate myself from other future job applicants?

Later that year, I enrolled at the University of Missouri's master's program. Missouri is home to Investigative Reporters and Editors and the National Association for Computer-Assisted Reporting – two nonprofits. That's where I learned data journalism.

I've worked in three newsrooms since graduating from Missouri in 2008. All three had mass layoffs. But my name was never called. Why not? I'm convinced it's because I know data journalism. I had a skill that no one else in the newsroom had and that brought value to the paper, which – at least as far as I'm concerned – saved my job.

COURSE DESCRIPTION

Some of the most impactful stories written are investigative projects rooted in data, thousands, sometimes millions of records. These in-depth series can root out corruption and protect those in need. But they often require advanced data journalism techniques that can't be done with Excel. Welcome R, a powerful, open-source database manager. In this class, students will use R to clean, query and make graphics out of raw data. They'll also learn how to harness the power of artificial intelligence. They'll use AI to search, categorize, summarize and structure unformatted text, allowing reporters to find stories previously left uncovered.

COURSE OBJECTIVES

You'll be asked to learn how to write code in R, a database manager. You'll also be asked to learn AI tools to manipulate unstructured data and turn it into files you can analyze. The ultimate goal: To become a better journalist.

LEARNING OBJECTIVES

In this class you are enterprise reporters. The objective is to become better ones. To do that, you'll be asked to:
- Understand what data are publicly available
- Learn how to request data by using public records laws
- Scrape data from PDFs and websites using R
- Master cleaning and querying data using R

- Create charts in using R
- Use AI tools to search, categorize and structure PDF files
- Produce an accurate, fair, engaging publishable story based on data

OTHER (IMPORTANT) NOTES

- It is up to you to stay on pace, learn the material and do the work. I'll help every way I can, but you need to put in the effort.
- Pay attention to Canvas. This is a must.
- These classes build on each other. If you do not understand a concept, let me know. I want to catch you up to speed before you get too far behind.
- The more you use the skills that you learn in this class, the better data journalists you'll become. Practice, practice, practice.

BOOKS/MATERIALS

- Excel. This is part of the Microsoft Office package available to students for free.
- R and Rstudio. We'll download it in class using these links: https://cran.r-project.org/ and https://posit.co/downloads/
- A Google Pinpoint account
- Recommended: An 8-gb flash drive to store files.

GRADING

Your grade will be determined on the following scale:

A	93 – 100	B-	80 – 82.9	D+	67 – 69
A-	90 – 92.9	C+	77 – 77.9	D	63 – 66.9
B+	87 – 89.9	C	73 – 76.9	D-	60 – 62.9
B	83 – 86.9	C-	70 – 72.9		

A – Superior. This work shows a superior understanding of the concepts, research, and analysis required by the assignment with few, if any, data-related errors. All answers must be supported by work that is clear, accurate and reproducible.
B – Above Average. This work shows a very good understanding of the concepts, research, and analysis required by the assignment with minimal data-related errors. Answers must be supported by work that is clear, accurate and reproducible.
C – Average. This work demonstrates a basic understanding of the concepts but is deficient in the research and analysis required by the assignment and includes many data-related errors. Some of the work needed to produce the answers is unclear, inaccurate or not reproducible.
D – Below Average. This work does not demonstrate an understanding of the concepts, is seriously deficient in the research and analysis required by the assignment, and includes excessive data-related errors. Little or no work is shown.
F – Deficient. This work is deficient on almost all counts.

EARNING THE GRADE YOU WANT

- Participation – 10%
- Homework – 25%
- Tests – 25%
- Final story – 40%

PARTICIPATION

Participation, what does that mean? It means to show up to class and show up to class on time. It means to ask questions and volunteer to answer them.

HOMEWORK ASSIGNMENTS

Most of the homework assignments will be short. Some will involve R and some will involve finding and dissecting good, data-centered stories. There will be some weeks in which you will file story memos that show your findings until that point. I will edit and grade these, and my edits are to be incorporated into your next memo. The point of these memos is to make sure you're on track to finish a final story that is worthy of being published.

TESTS

R test:

These will test your skills on cleaning and querying data using the computer language R. You'll be provided the raw data. You'll also be given a worksheet of possible R queries that you may be tested on. You may use the worksheet or other notes.

FINAL STORY

This is a data-centered story that you – or a group of you – complete specifically for this class. The goal is to write a publishable story by the end of the semester. I will likely choose the dataset for you but please offer suggestions if there is good, available data on a topic that you're passionate about.

Keep in mind that this is a news story. It is meant to inform, educate or shine light on an issue.

To make sure you're on pace to complete the story, we'll have periodic check-ins, where you'll update me – and sometimes the entire class – on where your data analysis and reporting stands. These check-ins are to make sure that you're working on the story throughout the semester. This is not an assignment that you can complete in the final few weeks.

To do well, your story needs to include all the necessary components of a strong data-centered story. These components include: findings from the data, interviews with experts and people effected, terrific writing, concise storytelling and possibly additional documents.

One last note…this is an advanced data journalism class. An advanced story with advanced reporting, writing and data skills is expected.

CLASS POLICIES

Attendance is imperative. These classes build on each other. What you learn today will help you understand the skills I'll teach in the next class.

We all are expected to attend each scheduled class and arrive on time prepared to participate. This includes coming to class with a computer and the necessary software loaded onto your machine.

Missing class does not change a deadline. All assignments are due when specified regardless of whether you expect to be in class, unless I've approved an alternate deadline in advance.

If there are special circumstances that prevent you from making arrangements for work missed because of an absence, please contact me or have a family member contact me as soon as possible, and I will do my best to work with you. I may require documentation.

This is a hands-on class that requires your participation. Part of that participation means asking questions, especially if you do not understand or cannot replicate a data query. Please speak up. If you have questions about a code that we're writing, chances are other students have the same questions. I need to know so I can help.

Behavior policies:
- Do not have conversation in class. This is disruptive.
- Do not arrive late or leave early. This is disruptive.
- Silence and keep your personal devices out of site. They are disruptive.
- Think twice about bringing food to class.

SUBMISSION OF ASSIGNMENTS

All written assignments must be submitted on the day they are due according to the guidelines outlined above and per the assignment description. Typically, assignments will be due at 11:59 p.m. on the day they are due. Please note that I will try to give you feedback as quickly as possible, but I too can get busy as the semester progresses. All grades will be posted on Canvas.

You may turn in an assignment late once without penalty. After that, I will not accept late assignments.

ACADEMIC INTEGRITY

UF students must adhere to The Honor Pledge which states, "We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code."

On all student work at UF, the following pledge is either required or implied: "On my honor, I have neither given nor received unauthorized aid in doing this assignment." The Honor Code (https://sccr.dso.ufl.edu/process/student-conduct-code) specifies a number of behaviors that are in violation of this code and the possible sanctions. Also, students are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have any questions or concerns, please consult with me. Violations can result in a failing grade for the course and referral to the dean of students.

STUDENTS WITH DISABILITIES

Students with disabilities who experience learning barriers and would like to request academic accommodation should connect with the Disability Resource Center. It is important to share any accommodation letter with me and discuss access needs as early as possible in the semester.

COURSE EVALUATIONS

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at **https://gatorevals.aa.ufl.edu/students/.** Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their

Canvas course menu under GatorEvals, or via **https://ufl.bluera.com/ufl/**. Summaries of course evaluation results are available to students at **https://gatorevals.aa.ufl.edu/public-results/**.

IN-CLASS RECORDING

The official UF policy regarding in-class recording, to comply with a 2021 Florida law can be found here.

HEALTH AND WELNESS

- U Matter, We Care: If you or someone you know is in distress, please contact umatter@ufl.edu, 352-392-1575, or visit U Matter, We Care website to refer or report a concern and a team member will reach out to the student in distress.
- Counseling and Wellness Center: Visit the Counseling and Wellness Center website or call 352-392-1575 for information on crisis services as well as non-crisis services.
- Student Health Care Center: Call 352-392-1161 for 24/7 information to help you find the care you need or visit the Student Health Care Center website.
- University Police Department: Visit UF Police Department website or call 352-392-1111 (or 911 for emergencies).
- UF Health Shands Emergency Room / Trauma Center: For immediate medical care call 352-733-0111 or go to the emergency room at 1515 SW Archer Road, Gainesville. Visit the UF Health Emergency Room and Trauma Center website.

ACADEMIC RESOURCES

- E-learning technical support: Contact the UF Computing Help Desk website, or phone 24/7 at 352-392-4357, or email helpdesk@ufl.edu.
- Career Connections Center: Career assistance and counseling services. Visit the website. Reitz Union Suite 1300, 352-392-1601.
- Library Support: Various ways to receive assistance with respect to using the libraries or finding resources. Visit the website.
- Teaching Center: General study skills and tutoring. Visit the website. Broward Hall, 352-392-2010 or to make an appointment 352- 392-6420.
- On-Campus Student Complaints: Details are available through the Student Honor Code and Student Conduct Code, also known as the Orange Book.

COURSE SCHEDULE

(Please note: The course schedule may change depending how well students are grasping the techniques described in class. ***Pay attention to Canvas for any changes.****)*

Week 1 – 1/14: Intro to data journalism and R

- What does it mean to be a project data reporter and what does it take to help produce large investigative series?
- Welcome to R and its import functions.

Week 2 – 1/21: Filtering in R

- How do you filter for direct matches, indirect matches, numbers and nulls.

Week 3 – 1/28: Grouping in R

- R can aggregate records and perform mathematical calculations, allowing reporters to identify trends, patterns and anomalies that make up the backbone of an investigation.

Week 4 – 2/4: Joining and binding in R

- One of the most powerful tools R possesses is the ability to make data frames talk to each other.

Week 5 – 2/11: Cleaning dirty data in R

- Before reporters can analyze data, they often have to create new fields. The mutate function is key, as is updating records with ifelse, parsing them and concatenating them.

Week 6 – 2/18: Cleaning dirty data in R (continued)

- Sometimes data is so dirty that we need two classes.

Week 7 – 2/25: Statistics in R

- Running significance tests, correlations and regressions can point you to a story.

Week 8 – 3/4: NICAR Conference (No class)

- R take-home test.

Week 9 – 3/11: Making charts in R

- Charts don't just complement data, they can help find stories, too.

Week 10 – 3/25: Scraping with R

- Often journalists can't download data from the web and requesting it would cost too much or take too long. How about scraping it?

Weel 11 – 4/1: Writing a data-centered story

- So you did all the data work, now how do you piece together an informative, thoughtful story with vivid descriptions and storytelling

Week 12 – 4/8: AI and spreadsheets

- Excel's Claude extension can extract, categorize and summarize your data, helping you uncover stories others can't.

Week 13 – 4/15: AI and PDFs

- AI tools can also help reporters make quick work of thousands of PDFs.

Week 14 – 4/22: AI wrap-up and final story discussion

- Possible R test. This will depend on how the final stories are going.