Syllabus

About the Course

Instructors

- Naila Moreira
- Ben Baumer

Description

Data journalism is the practice of telling stories with data. This course will focus on journalistic practices, interviewing data as a source, and interpreting results in context. We will discuss the importance of audience in a journalistic context, and will focus on statistical ideas of variation and bias. The course will include hands-on work with data, using appropriate computational tools such as R, Python, and data APIs. In addition, we will explore the use of visualization and storytelling tools such as Tableau, `plot.ly`, and D3. No prior experience with programming or journalism is required. **{M}{WI}**

> **!** Important
>
> This course satisfies the **data in context** requirement for the SDS major.

Prerequisites

- an introductory statistics course (including SDS 201/210/220, SOC 201, GOV 203, ECO 220, PSY 201)

Learning goals

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

- write in a journalistic style
- write about numbers effectively in context
- use original sources and interviews to support a piece of investigative journalism
- create an original data graphic that draws the reader's attention to particular points of emphasis
- augment a piece of investigative journalism with data graphics, images, videos, or other multimedia elements

Textbooks

Required

None

Suggested as supplementary references

- *Numbers in the Newsroom*, 2nd edition, Sarah Cohen, IRE, 2014 ($10 for e-book).
- *Communicating with Data*, Nolan, Stoudt, 2021. ~$50. (Amazon)
- *Happy Git and GitHub for the UseR*, Bryan, Hester. Available for free online.
- *Modern Data Science with R*, Baumer, Kaplan, and Horton, CRC Press, 2021. ~$100 (CRC | Amazon) | Available for free online
- *R for Data Science*, Garrett Grolemund and Hadley Wickham, O'Reilly, 2017. Available free online.
- Shiny is an interactive web application framework for R. Available for free via our Posit Connect Server.

Evaluation

Time

This is a 4 credit course, meaning that by federal guidelines, it should consume about 12 hours per week of your time. We meet for 3 hours per week. That means **you should be spending about 9 hours per week**, or nearly 90 minutes per day, on this course **outside of class.**

> ⚠ Warning
>
> You should be spending **about 9 hours per week** on this course outside of class.

Grading

- Homework [15%]: Assignments will include reading responses and other short assignments
- News stories [75%]:
    1. Major assignment 1 [20%]: A short piece during the first quarter of the class.
    2. Major assignment 2: An investigative project that will spawn two deliverables:
        1. A longer piece of investigative journalism [30%]
        2. A `closeread` project [25%]
- Engagement [10%]: Active participation in class, engagement with group work, activity on GitHub, helpfulness on Slack, and regular attendance will comprise the remainder of your grade.

Extensions

Extensions up to 48 hours will typically be granted *when requested at least 48 hours in advance*. Longer extensions, or those requested within 48 hours of a deadline will typically not be granted. Please plan accordingly. Please note that because many of the assignments in this class are collaborative, individual extensions for group assignments will be problematic. **All extended deadlines will appear on Moodle.**

Late assignments will be penalized at the rate of 20% per day, up to a minimum grade of 20% of the assigned value.

Academic Integrity

We expect that you will maintain your academic integrity in this class. Please read Smith's Academic Integrity Statement. Please pay particular attention to this sentence in the "Examples of Violations" section:

> If you are finding ways to avoid the "thinking" component of your coursework, you should stop to ask yourself whether you are compromising your academic integrity.

Policies

Inclusion

Inclusion policy

We are committed to fostering a classroom environment where all students thrive. We are committed to affirming the identities, realities and voices of all students, especially those from historically marginalized or underrepresented backgrounds. We are dedicated to creating a space where everyone in the class is respected, is free from discrimination based on race, ethnicity, sexual orientation, religion, gender identity, disability status, and other identities, and feel welcome and ready to learn at your highest potential.
If you have any concerns or suggestions for how to make this class more inclusive, please reach out to your instructor.
We are here to support your learning and growth as data journalists and people!

Accommodations

Smith is committed to providing support services and reasonable accommodations to all students with disabilities. To request an accommodation, please register with the Accessibility Resource Center (ARC) at the beginning of the semester. To contact ARC, please email arc@smith.edu.

Attendance

We expect you attend class in person. When you come to class, we expect your full attention. Please put your phone away during class unless otherwise directed.

If you are unable to attend class for any reason, please follow the materials on the course website and check with another student about what happened in class.

Collaboration

Warning

Much of this course will operate on a collaborative basis, and you are expected and encouraged to work together with a partner or in small groups to study, complete labs, and prepare for exams. However, **all work that you submit for credit must be your own**. Copying and pasting sentences, paragraphs, or *blocks of code* from another student

or from online sources is not acceptable and will receive no credit. No interaction with anyone but the instructors is allowed on any exams or quizzes.

Generative AI

Generative AI and Academic Integrity

Please read the Smith Academic Integrity Board's statement on Generative Artificial Intelligence & Your Academic Integrity

We draw your attention to the following excerpt:

> **Tip**
>
> Any time you are using AI in a way that is substituting for the "thinking work" that you should be doing for a course, you should stop.

Our perspective on AI and your learning

Our goal is to help you achieve the learning goals for this course using only the mental model you have built of the material we have covered, and *without* the aid of generative AI. While we accept the ubiquity of generative AI, we believe that helping you build your mental model of this material is where we can best contribute to your education. To that end, much of our time in class and many of our assessments will take place in AI-free environments. Other learning will take place outside of class, wherein you are free to use AI in whatever fashion you want (unless otherwise noted), provided that it is in compliance with Smith's Academic Integrity policies.

Please understand that while AI may be helpful to you in building your mental model of the material, it will not be available to you during many of our assessments, and we are comparatively less interested in your ability to complete tasks while using AI than we are in your ability to demonstrate knowledge using only your brain (and body).

This perspective applies to all course content, including mathematical equations and R code.

Usage

> **Important**
>
> - Use of generative AI is expressly prohibited on exams and oral presentations.
> - Use of generative AI is generally prohibited **inside the classroom**, although there may be exceptions.

- Generative AI may not be used for **writing**, including first drafts. The words need to come from you!
- Unless otherwise noted (such as the above), generative AI can be used whenever you are **outside of class**.

Please read our perspective on learning and generative AI. Please understand that careless and excessive use of generative AI will likely impede your ability to achieve the course learning goals.

> ⚠ Warning
>
> Remember that generative AI is not *intelligent*, doesn't *think*, and has no idea what is *true* or *false*. You are solely responsible for the veracity of anything (e.g., code or text) you submit.

Code of Conduct

As the instructor and assistants for this course, we are committed to making participation in this course a harassment-free experience for everyone, regardless of level of experience, gender, gender identity and expression, sexual orientation, disability, personal appearance, body size, race, ethnicity, age, or religion. Examples of unacceptable behavior by participants in this course include the use of sexual language or imagery, derogatory comments or personal attacks, deliberate misgendering or use of "dead" names, trolling, public or private harassment, insults, or other unprofessional conduct.

As the instructor and assistants we have the right and responsibility to point out and stop behavior that is not aligned to this Code of Conduct. Participants who do not follow the Code of Conduct may be reprimanded for such behavior. Instances of abusive, harassing, or otherwise unacceptable behavior may be reported by contacting the instructor.

> ❗ Important
>
> All students, the instructor, the lab instructor, and all assistants are expected to adhere to this Code of Conduct in all settings for this course: lectures, labs, office hours, tutoring hours, and over Slack.

This Code of Conduct is adapted from the Contributor Covenant, version 1.0.0, available here.

Resources

Moodle and course website

The course website and Moodle will be updated regularly with lecture handouts, project information, assignments, and other course resources. Homework and grades will be submitted to Moodle. Please check both regularly.

Computing

The use of the `R` statistical computing environment with the RStudio interface is thoroughly integrated into the course. You have two options for using `RStudio`:

- The **server** version of Posit Workbench on the web. The advantage of using the server version is that all of your work will be stored in the cloud, where it is automatically saved and backed up. This means that you can access your work from any computer with a web browser (Firefox is recommended) and an Internet connection.
- A **desktop** version of RStudio IDE installed on your machine. The downside to this approach is that your work is only stored locally, and you will have to manage your own installation.

Note that you do not have to choose one or the other – you may use both. However, it is important that you understand the distinction so that you can keep track of your work. Both `R` and `RStudio` are free and open-source, and are installed on most computer labs on campus.

Unless otherwise noted, you should assume that it will be helpful to bring a laptop to class. It is not *required*, but since there are no workstations in the classroom, we will need a critical mass (i.e. at least 12) computers in the classroom pretty much everyday.

Communication

- Slack is the primary forum for course-related discussions of all kinds. *Please do not email me with course-related questions!* Instead, post those `#questions` on Slack. If discretion is absolutely necessary, private message us on Slack.
- GitHub will host all of the code for projects associated with this course. The repository is private, but the website it serves is **public**.

Writing

Your ability to communicate results—which may be technical in nature—to your audience—which is likely to be non-technical—is critical to your success as a data analyst. The assignments in this class will place an emphasis on the clarity of your writing.

Writing Enriched Curriculum

This course is part of Smith College's Writing Enriched Curriculum. As such, the course supports the Writing Plan of the Program in Statistical & Data Sciences.

Please read the SDS Writing Plan for more information.

The Spinelli Center

The Spinelli Center (now in Seelye 207) supports students doing quantitative work across the curriculum. In particular, they employ:

- Data assistants
- Statistics TAs available from 7:00–9:00pm on Sunday–Thursday evenings in Burton 301. These students are trained to help you with your statistics questions, but may or may not be able to help you with your R questions.
- A Data Research and Statistics Counselor who keeps both drop-in hours and appointments. Students are welcome to email qlctutor@smith.edu to make an appointment with either the Data Counselor or one of the Data Assistants.

Your fellow students are also an excellent source for explanations, tips, etc.

Tentative Schedule

Please see the schedule for more specific information about readings and assignments.