Analytics for a Changing Climate: Introduction to Social Data Science

Stanford Summer Course 2023 | Instructor: Tyer McDaniel, Sociology, tylermc@stanford.edu

June 26-August 17 | Tuesday/Thursday, 1:30pm-3:30pm | Building: 320 Room 109

Syllabus

Office Hours: Fridays and Mondays, 11:00am-12:30pm https://calendly.com/tylermcdaniel/tyler-s-office-hours

Course Description: Data science has rapidly gained recognition within the social sciences because it offers powerful new ways to ask questions about social systems and problems. This course will examine how tools from data science can be used to analyze pressing issues relating to disaster, inequality, and scarcity in the Anthropocene (the current period in which humans are the primary driver of planetary changes). We will explore how a range of computational methods can be used to garner new meanings from sources such as weather monitors, press releases, websites, government programs, and more. This is a hands-on, interactive course culminating in a social data science project designed by the student or a team of up to four students. Most class sessions will be taught interactively using Jupyter Notebooks. Students will follow along with workshop-style lectures by using and modifying the provided R/Python code in real time to analyze data and visualize results. The course will cover such topics as the South African water crisis, Hurricane Katrina, the California Wildfires, and water rights along the Colorado River. Students will learn to explore text data with tools such as word embeddings, topic models, and sentiment analysis. Students will gain experience with R (and possibly Python) and will learn about a range of packages for cleaning data, linking and matching records, and mapping their results.

Who is this course for? This course is primarily designed for people with no or very little exposure to programming for social science. While this course may be useful for programmers and computer scientists seeking social applications for their skills, social science researchers interested in the topic of climate change, the materials will mostly be introductory. There are no prerequisites, other than a willingness to engage!

This course is also for students who seek to develop a project related to climate change and data science. Over the course of eight weeks, students will have opportunities to pose questions and ideas, receive feedback, and collaborate with others on building out a complete research project.

Learning Goals: At the end of the course, students will have experience using data science tools to address some major questions related to climate change and society. We will learn what questions can be addressed by data science and consider how we can effectively convey this information with the tools that we have. We will also think critically about power, inequality, and our role as data scientists in the twenty-first century.

Classroom policies: Please arrive on time and having completed the assigned materials. Being prepared to discuss the readings/videos/audio will enrich the learning experience of the entire class. Completing the assigned materials by Tuesday of each week will also allow students to identify areas of interest and consider possible research questions for problem sets or for the final project.

I ask that for 10 minutes of each class, we keep laptops away. We will be working on a series of hands-on activities to communicate some simple data during this time, as a warm-up activity to each class. By doing so, we can clear our minds, think creatively, and get in the right headspace to do some data science!

Class Schedule and Reading List

Readings should be done by class on Tuesday each week.

Week 1: Getting Started with Data Science in a Changing Climate (slides)

- Why ideas of 'planetary boundaries' must uphold environmental justice

- The Useful Discomfort of Critical Climate Social Science

- *Optional:* Perry, Keston. Realising Climate Reparations: Towards a Global Climate Stabilization Fund and Resilience Fund Programme for Loss and Damage in Marginalised and Former Colonised Societies

- *Optional:* Johan Rockström et al. 2023. Nature. Safe and Just Earth System Boundaries

- *Optional:* The Capitalinian: The First Geological Age of the Anthropocene

Week 2: Online Text as Data (slides)

- The Social Roots of Risk: Producing Disasters, Promoting Resillience. Chapter 1

- Twitter Bots are a Major Source of Climate Disinformation

- *Optional*: Marlow, Thomas, Sean Miller, and J. Timmons Roberts. 2021. Climate Policy. Bots and online climate discourses: Twitter discourse on President Trump's announcement of U.S. withdrawal from the Paris Agreement

- *Optional*: Locke et al., 2021. A Nature-Positive World: The Global Goal for Nature

Week 3: Text Analysis Tools, Part 1: Text Mining (slides)

- Floodlines, Episode 1: Antediluvian

- America's Deathtrap Schools

- Children of Katrina. Chapters 1-2

- *Optional*: There's No Such Thing as a Natural Disaster

- *Optional:* Baker, Dominique J., Karly Ford, Samantha Viano, and Marc Johnson-Guerrero. Racial Category Usage in Education Research: Examining the Publications from AERA Journals

- *Optional:* Which Towns are Worth Saving?

- *Optional*: Shaping the Coast with Permits: Making the State Regulatory Permitting Process Transparent with Text Mining

Week 4: Text Analysis Tools, Part 2: Sentiment (slides)

- Want to Survive Climate Change? You'll Need a Good Community

- Day Zero Cape Town

- Wetts, Rachel. 2022. PNAS. In climate news, statements from large businesses and opponents of climate action receive heightened visibility

Week 5: Text Analysis Tools, Part 3: Topic Modelling (slides)

- As the Colorado River Shrinks, Washington Prepares to Spread the Pain

- The Colorado River is Shrinking. See What's Using All the Water

- *Optional*: Casado-Pérez, Vanessa, Bruce E. Cain, Iris Hui, et al. 2015. All Over the Map: The Diversity of Western Water Plans

- *Optional*: The struggle for water : politics, rationality, and identity in the American Southwest. Chapter 1

Week 6: Maps and Spatial Data(slides)

- *Optional*: Maturo, Antonio and Veronica Moretti. 2018. Sociological Theories on Air Pollution: Between Environmental Justice and the Risk Society Approach

- *Optional*: Elliott, James and Jeremy Pais. 2008. Social Forces. Places as Recovery Machines: Vulnerability and Neighborhood Change After Major Hurricanes

- *Optional*: Elliott, Rebecca. 2020. Underwater: Loss, Flood Insurance, and the Moral Economy of Climate Change in the United States. Chapter 1

- *Optional*: Indigenous Land Return as Climate Justice

- *Optional*: Romero-Lankao, Patricia, Hua Qin, and Mercy Borbor-Cordova. 2013. Exploration of health risks related to air pollution and temperature in three Latin American cities

Week 7: Networks and Inequality

- Grasso, Marc and Richard Heede. 2023. Time to pay the piper: Fossil fuel companies' reparations for climate damages

- Howell, Junia and James Elliott. 2018. Socius. As Disaster Costs Rise, So Does Inequality

- Between the Devil, Debt and the Deep Blue Sea: Why the Global South Needs Reparations for the Climate Crisis

- *Optional*: Tackling Climate Change with Machine Learning

- *Optional*: Hauer, Mathew E. 2017. Migration induced by sea-level rise could reshape the US population landscape

- *Optional*: Farrell, Justin. 2016. Network structure and influence of the climate change counter-movement

Week 8: Final presentations

- No readings this week! Just work on your projects and come prepared to watch your classmates present!

Assignments

- Week 1 Problem Set   *Due: July 5th*
- Week 2 Problem Set   *Due: July10th*
- Week 3 Problem Set   *Due: July 17th*
- Final Project Proposal   *Due: July 17th*
- Week 4 Problem Set   *Due: July 24th*
- Week 5 Problem Set   *Due: July 31st*
- Week 6 Problem Set   *Due: August 7th*
- Week 7 Problem Set   *Due: August 14th*
- Final Presentation   *Due: August 21st*
- Final Project   *Due: August 24th*

Grading

Final grades will be determined as follows:

- Participation: 12.5%
- Problem Sets: 52.5%
- Final Project, Proposal, and Presentation: 35%

Participation

Students can participate in many ways. Engaging in class activities, working well with peers and in groups, and asking relevant questions during class are all valid forms of participation. Coming to class prepared (having done any readings or assignments) is vital!

Problem Sets

There will be seven problem sets over the course of the quarter. Students are expected to complete each of these on time and turn in assignments on Canvas. Most of the assignments should be turned in as .PDF files generated by R Markdown or Jupyter Notebook.

Part of each Thursday class will be devoted to answering questions about the problem set that is due the following Monday. Students with further questions, code bugs, and any other problems should come to office hours on Friday or schedule an appointment at another time.

Final Project, Proposal, and Presentation

The final project is intended to be an opportunity for students to apply some of the skills that we learn in this class toward a pressing and important issue. Much of this class is very introductory - we will go broad, but not deep - however, in the final project you can go deeper into one specific method or topic area.

Students can work in groups of up to *four* for the final projects. Individual projects are also fine - but I encourage you to collaborate if you have the capacity! Much of the best science is now done in teams, so this type of work is an important skill to practice. As is the norm in science, teams should write details on specifically who contributed what to the project.

Because the quarter is short, students should begin thinking about potential final projects early. The final project proposal is due at the same time as Problem Set 3, leaving roughly four more weeks for students to complete the project.

Students with Documented Disabilities

Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Professional staff will evaluate the request, review appropriate medical documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty. The letter will indicate how long it is to be in effect. Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations. Students should also send your accommodation letter to instructors as soon as possible.

From the OAE: "Documentation is not required to meet with us. To request a meeting with a Disability Adviser without medical documentation, email oae-contactus@stanford.edu."

A note on the syllabus: The planned readings and notebooks may evolve during the quarter to better reflect the class's collective interests, strengths, and progress. Any changes will be made in consultation with the class.

A note on late assignments: I recognize that unexpected events come up, and despite our best efforts, we sometimes miss deadlines. While it is important to keep up with the class assignments, it is possible to catch up with some additional effort.

Therefore, everyone is granted *2 late assignments, no questions asked*. These can be turned in up to one week after the initial deadline with no penalty. Afterwards, or for the third late assignment, students will lose 30% of the maximum grade for each week that they are late.

The purpose of this policy is to respect your privacy. You don't need to disclose to me *why* an assignment is late, if you don't want to. But in order for this policy to work well, it is vital that you only use these late assignments when you really need them. I highly, highly, encourage you to submit everything on time if you are able to.

If you have any questions, or if a situation comes up where you foresee needing more than 2 late assignments, please come to office hours or reach out!

A note on Chat GPT: As you might be aware, the latest versions of Chat GPT have the capability to provide human-like intuition and reasoning, and to answer novel problems in interesting ways. These are incredibly powerful tools. These tools might *supplement* our learning, but they should never *replace* our learning.

This class requires a lot of coding, problem-posing, and data interpreting. We might want to ask ChatGPT a question like, "how do you read a csv into R?" or "how can I add spaces to my R markdown document?" from time to time. I recommend going to sources like stackoverflow first, but sometimes these are not super helpful. In these cases, using ChatGPT could quickly provide a useful answer. I consider this acceptable as long as you do two things:

1. You are responsible for knowing what the code that ChatGPT gives you is doing. If you submit this code as a homework, you should be able to explain it to the class.

2. You will note at the beginning of your problem set or project whether ChatGPT was used or not.

There are cases where usage of ChatGPT is not acceptable. Importantly, *it is never acceptable to simply plug in a homework question from this class and let ChatGPT come up with an answer*. I consider this a violation of Stanford's Fundamental Standard. In general, if you would like to use ChatGPT make sure you are using it for the most specific aspect of a technical question possible (e.g. an edit to a line of code). Your written work should always be your own.

Resources

R cheat sheets These sheets are extremely helpful in documenting most or all of the functionalities of different R packages, including `dplyr`, `ggplot2`, and more.

R for Data Science, Hadley Wickham & Garrett Grolemund This book is an excellent guide to using R for data science, written by the same person (Hadley Wickham) who wrote `dplyr` and many of the R functions that you will use in this class.

Bit By Bit, Matthew J. Salganik This is probably the most complete book out there on doing data science for sociology (and other social sciences). The sections on research ethics are especially insightful, in my opinion.

SICSS Learning Materials This is a wonderful trove of materials geared toward graduate students in the social sciences seeking to learn (or improve upon) their data skills. It is organized by Matthew Salganik (from Bit By Bit) and Chris Bail.

Data Visualization: A Practical Introduction This is a great book if you are interested in visualizing data (which we all should be!). It includes lots of great examples, and also assumes no prior knowledge of R or Rstudio (but includes code so you can build your knowledge here too).