

Text as Data

Poli 176/DSC 161 Syllabus¹

Prof. Davi Moreira
Spring 2023

Course Information

Room: RWAC 115

Time: Mon/Wed/Fri 2:00-2:50pm

Professor

Email: dmoreira@ucsd.edu

Office Hour: Fri 11:00am-Noon, sign up [online](#),
Room: GPS 1214, or by appointment.

TA: Alison Sile Chen

Email: sic015@ucsd.edu

Office Hour: Thu 1-2 pm, Local: Ridgewalk Social nearby by Rimal

Overview and Course Goals

This course aims to introduce a novel approach for integrating text data into social science research design. We will demonstrate how text data can be utilized to describe the prevalence of a social behavior or phenomenon, as well as to make inferences about its origins. Thanks to the abundance of text and the emergence of new statistical methods, these inferences have become more accessible than ever before. Unlike the traditional focus on predictive tasks, the goal of this course is to emphasize inference in social science research. By doing so, we reconsider when and how some methods are useful, suggest new ways to evaluate them, and raise ethical considerations when using text as data.

The course will provide students with an overview of the techniques for text analysis. We will also provide a framework for incorporating text analysis into social science research design. Although a deep mathematical understanding of every approach may not be feasible within the given time-frame, students will acquire practical tools for quantitative analysis of texts and learn how these methods can contribute to social science inference.

By the end of the course the students will be able to:

1. Define and identify the social science tasks discovery, measurement, prediction, and causal inference in applied social science research.
2. Explain (but not necessarily derive) the foundations of statistical text analysis methods, including bag of words approaches, word embeddings, clustering, topic modeling, and classification.
3. Implement the above statistical text analysis methods in statistical software.
4. Produce a final project that applies statistical text analysis to a social science research task. Be able to describe not only what is learned from the analysis, but also its limitations.

¹This syllabus was developed based on the 2022 Poli 176/DSC 161 course edition. Thanks to Professor Molly Roberts for guidance and for sharing the material.

Course Format

This course will be held in person*. On Monday and Wednesday lectures will be held. Discussion and practice sections will be held on Fridays.

In-Person* Lecture and Discussion Section

Monday, Wednesday, and Friday, 2-2:50PM

*The Professor and/or the TA will invite students to watch asynchronous videos if necessary.

*Lectures will be podcasted, but students are strongly encouraged to attend lectures in person. Students are strongly encouraged to attend discussion sections on Friday in person and will receive participation points for doing so.

Asynchronous (Online) Course Elements

Videos, readings, quizzes, problem sets, and other resources available at: UC San Diego's Learning Management System: <https://canvas.ucsd.edu>. Readings are available through Perusall Login: UC San Diego Active Directory credentials.

A Typical Week in This Course

Monday

- Attend in-person lecture and complete readings. These lectures will introduce the foundational concepts in the course that will help you complete the weekly quizzes, midterm assignment, and final project.
- Complete assigned readings and make one annotation per reading/chapter on Perusall (linked in Canvas)

Wednesday

- Attend in-person lecture and complete readings. These lectures will introduce the foundational concepts in the course that will help you complete the weekly quizzes, midterm assignment, and final project.
- Complete assigned readings and make one annotation per reading/chapter on Perusall (linked in Canvas)

Friday

- Quizzes/problem sets due before section, at 2PM.
- In-person discussion section, 2-2:50PM. In this discussion section, students will apply concepts from the lecture in hands-on code examples.
- Students will code with the professor. Students receive participation points for attending discussion.

Who Are My Instructors?

Instructor



Davi Moreira
(he/him/his)

Visiting Assistant Professor
Political Science and GPS UCSD

www.davimoreira.com

Overall Course Expectations

What you can do to support your success in the course:	What I will do to support your success in the course:
Read the syllabus and stay current with course information	Be prepared and bring my enthusiasm for teaching to each session
Keep up with readings and assignments, as each one builds on the previous one.	Respond to emails within one working day, and provide timely feedback on assignments / submissions.
Contribute to the learning environment with fairness, cooperation, and professionalism	Establish a learning environment with fairness, cooperation and professionalism, and will take action if these principles are violated.
Treat your classmates, instructional assistants and myself honestly and ethically	Treat you honestly and ethically, and will address any concerns you might have
Commit to excel with integrity. Have the courage to act in ways that are honest, fair, responsible, respectful & trustworthy.	Uphold integrity standards and create an atmosphere that fosters active learning, creativity, critical thinking, and honest collaboration.
Manage your time, so you can stay on track with the course and complete tasks on time	Only assign work that is vital to the course, and will work to meet the standard credit hour allotment for the course.
Communicate with me if you determine that a deadline cannot be met due to extenuating circumstances	Consider requests for adjustments and will make reasonable exceptions available to all students when approved

1. Please read UC San Diego's Policy on [Integrity of Scholarship](#) and take the integrity pledge!

Readings

You will find readings on the course website on Canvas through the Perusall application. Students are expected to make at least one comment on each reading/chapter as part of their participation grade.

Course Software

We will use R (<http://www.r-project.org>) and RStudio (<https://posit.co/download/rstudio-desktop/>) to introduce the computational tools within the course. Students more familiar with Python are welcome to use it for assignments and projects.

Assignments, Projects, and Grading

Summary of Grade Criteria

Assignment	Weight	Due Date
Participation	20%	Readings.
Quizzes	25%	Fridays, 2-2:50pm.
Midterm	25%	Friday, May 6, 2pm
Final Project Hackathon	30%	Due Wednesday, May 25, 11:59pm

Course Policies

Academic Accommodations

Students requesting accommodations for this course due to a disability must have a current Authorization for Accommodation (AFA) letter issued by the Office for Students with Disabilities (OSD: <http://disabilities.ucsd.edu>), which is now located on the 3rd Floor of Pepper Canyon Hall. AFA letters are now provided to Faculty electronically by OSD, per student request. Requests for accommodation must be made at least two weeks in advance of the midterm exam. You can find further information [here](#) and [here](#).

Grading Scale

Grade	Percentage
A+	97% - 100%
A	93% - 96.9%
A-	90% - 92.9%
B+	88% - 89.9%
B	83% - 87.9%
B-	80% - 82.9%
C+	78% - 79.9%
C	73% - 77.9%
C-	70% - 72.9%
D	60% - 69.9%
F	0% - 59.9%

Grading Procedure and Feedback

Students will be graded on an absolute scale over the course of the quarter. At the end of the quarter, the average of the scores in the class will become 100%.

Rubrics for Final Projects evaluation

1. Relevance of research question (20%)
2. Data quality and preprocessing (20%)
3. Analysis methodology (20%)
4. Interpretation of results (20%)
5. Discussion of limitations (20%)

Attendance and Participation

Students will receive attendance and participation points by

- 1) Commenting on readings through Perusall.
- 2) Attending and participating in the weekly discussion section.

Late Submissions & make up exam policy

Late submissions will not be graded (i.e., you will receive a zero for the assignment/project). There will be no make-up examinations without an approved and documented excuse. Acceptable excuses include illness, which must be documented by the UCSD health service or your physician, or a death or serious illness in the immediate family. I will accept no other excuses aside from those recognized by standing University policy.

Grade disputes

If you believe there has been an error or oversight in grading your work you may petition the instructor to have your grade changed. To do so you must submit a written memo of no more than one page explaining the error and stating the grade you believe that you deserve. We will not entertain grade complaints beyond three calendar days after the assignment has been returned to you.

Academic Integrity

Students in this course are expected to comply with UCSD's Policy on Integrity of Scholarship. In particular, plagiarism is considered a dishonest practice and a serious academic offense. Hence, there will be a zero tolerance policy with respect to these practices: any student violating the obligation of academic integrity during the term will automatically fail the class. You are not allowed to use course materials from previous years or to post solutions from this year's assignments anywhere online or to pass them onto next year's students. Failure to adhere to this policy will result in initiation of academic integrity proceedings, including after the conclusion of the quarter.

You may ask other students to read and comment on your work but all analytic work is expected to be your own. You will appropriately recognize and cite all sources of data or information you use. If you are unsure how to do this, please come to office hours or ask in class. If you are unsure what constitutes plagiarism, please see UCSD's academic honesty policy here: <http://academicintegrity.ucsd.edu/>. All cases of suspected plagiarism and cheating will be referred to the office of academic integrity.

Submitting written work:

- Students agree that by taking this course all required papers will be submitted for textual similarity and plagiarism review via Turnitin.com. All submitted papers will be included as source documents in the Turnitin.com reference database solely for the purpose of detecting plagiarism of such papers. Use of the Turnitin.com service is subject to the terms of use agreement posted on Canvas and the Turnitin.com site.

Collaboration Policy

- You are welcome to discuss ideas with classmates, but should submit your own original solutions.

AI Policy

I encourage you to use AI tools you believe will enhance your individual or group performance. In fact, some course assignments can require it. Learning to use AI is a valuable and emerging skill, and I am available to provide support and assistance with these tools during office hours or by appointment.

Be aware of the following guidelines:

- Providing low-effort prompts will result in low-quality outputs. You must refine your prompts to achieve desirable outcomes.
- Do not blindly trust the information provided by the output. If the output contains a number, index, or fact, assume it is incorrect and check its veracity in another source. Any errors or omissions resulting from using the AI tool will be your responsibility. Remember, the AI tool works better for topics that you already understand.
- While AI is a tool, you must acknowledge its use. Always cite! Include a paragraph or note at the end of any document that utilizes AI, explaining what prompts were used and how they were utilized to obtain the results. Failure to do so would violate academic honesty policies.

Subject to Change Policy

While I will try to adhere to the course schedule as much as possible, I also want to adapt to your learning pace and style. Therefore, the syllabus and course plan may change in the quarter. I always welcome feedback from you about what is working and not working for your learning in the course.

Etiquette

Here are some general rules that may seem fairly obvious, but we are putting them here to be clear on what our expectations are of you as professionals.

- Do not be disrespectful towards anyone, TAs, assistants, professors, or classmates.
- Do not spam TAs, assistants, or professors about anything. We try our best to be responsive, so we will get you as soon as we can.
- Name calling, harassment, or threats are not tolerated.
- No discriminatory language (ex: racist, homophobic) or hate speech will be tolerated.

Nondiscrimination and Harassment

UC San Diego is committed to creating an environment in which all students are able to learn and openly express themselves in an environment free from all forms of discrimination and harassment. I encourage all students to read campus policy [here](#).

Resources for Support and Learning

Learning and Academic Support

- **Ask a Librarian:** Library Support Chat or make an appointment with a librarian to focus on your research needs
- **Course Reserves, Connecting from Off-Campus and Research Support** Find supplemental course materials
- **First Gen Student Success Coaching Program:** Peer mentor program that provides students with information, resources, and support in meeting their goals.
- **Office of Academic Support & Instructional Services (OASIS):** Intellectual and personal development support
- **Writing Hub Services in the Teaching + Learning Commons:** One-on-one online writing tutoring and workshops on key writing topics
- **Supplemental Instruction:** Peer-assisted study sessions through the Academic Achievement Hub to improve success in historically challenging courses
- **Tutoring – Content:** Drop-in and online tutoring through the Academic Achievement Hub
- **Tutoring – Learning Strategies:** Address learning challenges with a metacognitive approach

Developing Writing Skills

You are encouraged to seek writing support in UCSD's Writing Center (<http://writingcenter.ucsd.edu>). Please consider this resource, especially if English is your second language.

Support for Well-being and Inclusion

- **Basic Needs at UCSD:** Any student who has difficulty accessing sufficient food to eat every day, or who lacks a safe and stable place to live is encouraged to contact: foodpantry@ucsd.edu — basicneeds@ucsd.edu — (858) 246-2632
- **Counseling and Psychological Services:** Confidential counseling and consultations for psychiatric service and mental health programming
- **Triton Concern Line:** Report students of concern: (858) 246-1111
- **Office for Students with Disabilities (OSD):** Supports students with disabilities and accessibility across campus
- **Community and Resource Centers Office of Equity, Diversity, and Inclusion:** As part of the Office of Equity, Diversity, and Inclusion the campus community centers provide programs and resources for students and contribute toward the evolution of a socially just campus (858).822-.3542 — diversity@ucsd.edu
- **Get Involved:** Student organizations, clubs, service opportunities, and many other ways to connect with others on campus
- **Undocumented Student Services:** Programs and services are designed to help students overcome obstacles that arise from their immigration status and support them through personal and academic excellence

Health and Well-Being

Health and Well-Being Statement Throughout your time at UC San Diego, you may experience a range of issues that can negatively impact your learning. These may include physical illness, housing or food insecurity, strained relationships, loss of motivation, depression, anxiety, high levels of stress, alcohol and drug problems, feeling down, interpersonal or sexual violence, or grief.

These concerns or stressful events may lead to diminished academic performance and affect your ability to participate in day-to-day activities. If there are issues related to coursework that are a source of particular stress or challenge, please speak with me, Professor Roberts, so that I am able to support you. UC San Diego provides a number of resources to all enrolled students, including:

- Counseling and Psychological Services (858-534-3755 — caps.ucsd.edu)
- Student Health Services (858-534-3300 — studenthealth.ucsd.edu)
- CARE at the Sexual Assault Resource Center (858-534-5793 — care.ucsd.edu)
- The Hub Basic Needs Center (858-246-2632 — basicneeds.ucsd.edu) We care about you at UC San Diego, and there is always help available.

Campus Policies

Below are links to useful campus policies – principles of community, integrity, and code of conduct that are also policies within this course:

- UC San Diego Principles of Community
- UC San Diego Policy on Integrity of Scholarship
- Religious Accommodation
- Nondiscrimination and Harassment
- UC San Diego Student Conduct Code

Books

We will read chapters from the following books:

- Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton University Press.

Course Schedule

Below is a tentative course schedule. Please refer to the Modules section on canvas for the most up to date schedule.

Legend: Required readings and tasks are marked with a (★).

I. Week 1

Read:

- ★ DiMaggio, Paul. "Adapting computational text analysis to social science (and vice versa)." Big Data & Society 2.2 (2015).
- ★ Catalinac, Amy. "From Pork to Policy: The Rise of Programmatic Campaigning in Japanese Elections." The Journal of Politics. 2016; 78 (1) :1-18.

Lectures:

- Introduction to the course
- Overview of the syllabus

Do:

- ★ Introduce yourself on the Canvas discussion board
- ★ Reading Annotations on Perusall

II. Week 2

Read:

- ★ Grimmer, Roberts and Stewart, Chapters 3-7

Lectures:

- Multinomial Language Model
- Vector Space Model and Similarity Metrics

Do:

- ★ Quiz 1

III. Week 3

Read:

- ★ Grimmer, Roberts and Stewart, Chapters 8 and 11
- ★ Nelson, Laura K. "Computational grounded theory: A methodological framework." Sociological Methods & Research 49.1 (2020): 3-42.

Lectures:

- Introduction to Discovery
- Discriminating Words

Do:

- ★ Quiz 2
- ★ Reading Annotations on Perusall

IV. Week 4

Read:

- ★ Grimmer, Roberts and Stewart, Chapters 12 and 13

Lectures:

- Topic Models

Do:

- ★ Quiz 3
- ★ Reading Annotations on Perusall

V. Week 5

Read:

- ★ Grimmer, Roberts and Stewart, Chapters 12 and 13

Lectures:

- Interpreting Topic Models

Do:

- ★ Quiz 4
- ★ Midterm Homework Assignment

VI. Week 6

Read:

- ★ Grimmer, Roberts and Stewart, Chapters 15-17
- ★ Gillion (2016), Chapter 2.

Lectures:

- Supervised Learning: An Overview
- Human Coding

Do:

- ★ Reading Annotations on Perusall
- ★ Final Project Hackathon: Submit a project proposal, including research question, dataset, analysis plan, and group members.

VII. Week 7

Read:

- ★ Grimmer, Roberts and Stewart, Chapters 19-20

Lectures:

- Supervised Learning Fundamentals
- Cross-Validation

Do:

- ★ Quiz 5
- ★ Final Project Hackathon: submit preliminary text analysis and results.
- ★ Reading Annotations on Perusall

VIII. Week 8

Read:

- ★ Grimmer, Roberts and Stewart, Chapters 22-23

Lectures:

- Overview of Prediction
- Validation of Prediction

Do:

- ★ Quiz 6
- ★ Reading Annotations on Perusall

IX. Week 9

Read:

- ★ Grimmer, Roberts and Stewart, Chapters 24-25

Lectures:

- Latent Variables and Causal Inference
- Text as Outcome

Do:

- ★ Final Project Hackathon: submit a draft of the final report, including research question, findings, limitations, and implications for political science research.
- ★ Reading Annotations on Perusall

X. Week 10

Do:

- ★ Presentations of Hackathon's Final Projects.