POLS 260: Introduction to Social Data Analytics
North Dakota State University, Fall 2021, 3 Credits

Basic Information

Professor: Dan Pemstein
Email: daniel.pemstein@ndsu.edu
Slack: NDSUPOLS260.slack.com
Phone: 701-231-6563

Class Location: TBA
Class Time: TBA
Office Location: Putnam 12
Office Hours: TBA

Course Description

Bulletin Description

Collection, analysis, visualization, and presentation of digital social data, with a focus on political science and public policy applications, and ethical social data science.

Course Overview

POLS 260 is an introduction to the field of social data analytics, with a focus on applications to the study of politics and public policy. Social data analytics (often called social data science) is a fast emerging field that combines social scientific theory and methods with sophisticated application of statistical and computational tools to leverage the massive explosion of social data produced by our digitally connected society. This course offers a hands-on introduction to the field, providing students with a foundation in digital social data collection, analysis, visualization, and presentation, in the context of politics and public policy. It also asks students to critically examine the social implications of social data science, particularly equity and privacy concerns. Finally, students will enjoy multiple guest lectures from social data scientists working for social media companies, media outlets, governments, political campaigns, and other data-driven employers.

Course Objectives

After completing POLS 260, students should be familiar with basic data manipulation and analysis in the R language for statistical computing, be able to scrape unstructured online data and work with application programmer interfaces provided by social media firms, apply basic natural language processing and network analysis tools to social data, visualize spatial data, work with databases, be able to construct useful visualizations and clearly, engagingly, and interactively present findings from social data analyses. Most importantly, students should be able to *ethically* deploy these tools to *leverage domain expertise in social science* and *effectively communicate* their analyses to non-technical audiences including policy-makers, managers, and average citizens.

Required Student Resources

The main technical reference for this course is available online for free, although a print version is available for purchase.

- Hadley Wickham and Garrett Grolemund. 2017. *R for Data Science.* Sebastopol, CA: O'Reilly Media.

All substantive readings for the course are available online, or on blackboard. The following technical resources may also prove helpful:

Advanced R	R Cookbook	R Markdown	ggplot2
Mastering Shiny	Pro Git	Stack Overflow	

Schedule

1   Introduction, R and R Markdown

"I've come to the conclusion that rather than "fixing" R, it would be much more productive to simply start over and build something better."

— Ross Ihaka, R Co-Creator (2010)

Topics: installing R, basic computation, generating reports, data science and democracy
Media Pairing: Weapons of Math Destruction
Technical Reading: R4DS 1, 4, 6, 8, R Markdown Cheat Sheet, R Markdown Examples
Applied Reading: What is social data science and how is it done?

2   Data Structures and Manipulation

"Storage experts are hoarders."

— Marie Kondo (2019)

Topics: types of data structures, loading data, data wrangling in base R and tidy, cross-national trends in democratic institutions
Media Pairing: Swedish study finds that the Republican Party has moved toward illiberalism
Technical Reading: R4DS 5, 9–13
Applied Reading: V-Dem Democracy Report 2021 (pp. 6–9, 13–19)
Project: Varieties of Democracy

3   Getting a Feel for your Data

"It has been called the interocular traumatic test; you know what the data mean when the conclusion hits you between the eyes."

— Edwards, Lindman, and Savage (1963)

Topics: summarizing data numerically/visually, sanity checks, measuring internet politics
Media Pairing: On Average
Technical Reading: R4DS 7, Kaplan Ch. 3
Applied Reading: Measuring Internet Politics: Introducing the Digital Society Project

4 Traditional Social Science Datasets

> "*Thou shalt not sit*
> *With statisticians nor commit*
> *A social science.*"

> — W.H. Auden, Under Which Lyre (1946)

Topics: survey, cross-sectional, and panel data, digital politics
Media Pairing: The History that Defines our Technological Future
Technical Reading: R4DS 14–16
Applied Reading: 'Extremely aggressive' internet censorship spreads in the world's democracies
Project: Quantifying internet censorship worldwide

5 Version Control with Git

> "*It's no use, Mr. James—it's turtles all the way down.*"

> — J.R. Ross (1967)

Topics: version control, collaborative software design, inclusive open source communities
Media Pairing: How open are open source communities?
Technical Reading: Pro Git Ch. 1–2
Applied Reading: The Software that Builds Software

6 Basic Data Visualization

> "*Whenever I am infuriated, I revenge myself with a new diagram.*"

> — Florence Nightingale (1857)

Topics: ggplot basics, patterns for social data viz, (mis)informing the public with graphs
Media Pairing: Florence Nightingale: Data Viz Pioneer
Technical Reading: R4DS 3
Applied Reading: 538's Weirdest (and Best) Charts of 2020
Project: Visualizing the Supreme Court using Martin-Quinn Scores

7 Less Basic Data Visualization

> "*Above all else, show the data.*"

> — Edward Tufte, The Visual Display of Quantitative Information (1983)

Topics: combining plots, avoiding chart-junk, communicating data to policy-makers
Media Pairing: Visualizing the spread of Coronavirus
Technical Reading: R4DS 17–18
Applied Reading: Guidelines for Good Visual Information Representations

8   Web Scraping

> "*And again, the Internet is not something that you just dump something on, it's not a big truck. It's a series of tubes.*"
>
> — Senator Ted Stevens (2006)

Topics: basic html, simple scraping in R, lobbying and campaign finance
Media Pairing: Meeting the pioneers of the world wide web
Technical Reading: R4DS 19, Tidy web scraping in R
Applied Reading: The solution to lobbying is more lobbying
Project: Visualizing campaign finance

9   Scraping Social Media

> "*What if we just went home and read books to each other?*"
>
> — Gary Shteyngart, Super Sad True Love Story (2010)

Topics: Twitter research API, JSON and XML data, regulating social media
Media Pairing: What you need to know about APIs
Technical Reading: R4DS 20–21, Twitter Research API
Applied Reading: Inside the Making of Facebook's Supreme Court

10   Textual Data

> "*A word after a word after a word is power.*"
>
> — Margaret Atwood, Spelling (2003)

Topics: parsing, sentiment analysis, document classification, political speech
Media Pairing: Using human language to make computers think more like us
Technical Reading: Text Mining with R 1–4
Applied Reading: Does Gender Still Matter for Politics? The case of the 2018 U.S. Elections . . .
Project: Candidate tweets and gender norms

11   Network Data

> "*Man, Farmville is so huge! Do you realize it's the second-biggest browser-based social-networking-centered farming game in the WORLD?.*"
>
> — Randall Munroe, Online Communities 2 (2010)

Topics: network visualization, centrality and other metrics, disinformation campaigns
Media Pairing: Country of Liars
Technical Reading: Networks Demystified Parts 1 & 2, Intro to Network Analysis with R
Applied Reading: Cross-platform disinformation campaigns, Five myths about misinformation

12   Interactive Data Presentation with Shiny

> *"Shiny, let's be bad-guys."*
>
> — Jayne Cobb, Serenity (2005)

Topics: data dashboards, open government data and accountability
Media Pairing: City Health Dashboard at NYU
Technical Reading: Mastering Shiny 1, 2, 3
Applied Reading: Open Government Data
Project: Visualizing, summarizing, and exploring disinformation networks

13   Spatial Data

> *"Every cubic inch of space is a miracle."*
>
> — Walt Whitman (1856)

Topics: maps, coordinate systems, spatial demography, social sorting
Media Pairing: Data as Protest: Data for Black Lives with Yeshi Milner
Technical Reading: Drawing beautiful maps programmatically with R, sf and ggplot2:
Basics, Layers, Layouts
Applied Reading: The Biggest Barriers to COVID Vaccination for Black and Latinx People

14   Reproducibility and Open Science

> *"Information is power. But like all power, there are those who want to keep it*
> *for themselves. The world's entire scientific and cultural heritage, published over*
> *centuries in books and journals, is increasingly being digitized and locked up by a*
> *handful of private corporations."*
>
> — Aaron Swartz, Guerilla Open Access Manifesto (2008)

Topics: replication, data sharing, and open access publishing
Media Pairing: Cosmic Habituation
Technical Reading: The garden of forking paths
Applied Reading: More social science studies just failed to replicate. Here's why this is good.
Project: The Racial Geography of Covid-19

15   Big Data

> *"I wish I was big."*
>
> — Josh Baskin (1988)

Topics: databases, cluster computing, social externalities of big data
Media Pairing: Saving Big Data From Itself
Technical Reading: CCAST User Guide, Using dplyr with databases, Database Queries with R
Applied Reading: Big Data, Big Waste? A Reflection on the Environmental Sustainability...

16   Ethical Data Science

> *"Our own values and desires influence our choices, from the data we choose to collect to the questions we ask. Models are opinions embedded in mathematics."*

> — Cathy O'Neil, Weapons of Math Destruction (2016)

Topics: AI/ML bias, data privacy, equity in tech, politics and sociology of data science
Media Pairing: Building a Movement to Dismantle the New Jim Code
Applied Reading: How the Cambridge Analytica scandal unravelled
Project: Anybody seen the popo's? Exploring police cruiser transponder data.

Evaluation

Summary

Projects	80%
Participation	20%

Fortnightly Projects

Working in groups of 3–4, students will demonstrate mastery of the course material by completing eight hands-on assignments, due every other week. Groups will be assigned randomly for every assignment.

Grades

I use a flat grade scale: $A \geq 90 > B \geq 80 > C \geq 70 > D \geq 60 > F$. If you have a complaint about an assignment grade you must type a formal appeal describing the problem. Your appeal should clearly explain why you believe your grade is incorrect and make a cogent argument for improving your grade. Attach a copy of the original assignment to your appeal.

Participation

Participation is worth 20 percent of your final grade and will reflect your engagement in and contribution to class. Participation can take many forms, including—but not limited to— asking questions, answering my queries, engaging in class debate, organizing or participating in study groups, asking or answering questions in public Slack channels, and taking an active role in class activities. In this course, working effectively with your project group members is also a fundamental aspect of participation. Towards the end of the semester, each student will write a short (1-2 page) paper making an evidence-based case for the participation grade that she feels she deserves. Students should also note any issues they experienced with group dynamics. These papers, and the persuasiveness of their arguments, will form the basis for students' participation grades. Although you have substantial leeway in how you make your case for your participation grade, your self-evaluation should follow this rough rubric:

A   Strong attendance, frequent and thoughtful verbal participation,
     active participation in group work
B   Strong attendance, regular and thoughtful verbal participation,
     active participation in group work
C   Strong attendance, some verbal participation, satisfactory participation in group work
D   Missed more than 3-4 classes, little to no verbal participation in class,
     lack of engagement in group work
F   Frequently miss class, no verbal participation, leave your group members hanging

Class Policies

Make-Up Assignments

Each student may drop one weekly project without penalty, so make-up assignments should only be an issue in exceptional circumstances. Please get in touch with the instructor as early as possible if you find yourself in this situation.

Office Hours

Office hours are a time for students to discuss any aspect of the course with the professor. I will hold "drop-in" office hours at the scheduled times, but you may also arrange to meet at another time if those times do not work for you.

Attendance

According to NDSU Policy 333 (http://www.ndsu.edu/fileadmin/policy/333.pdf), attendance in classes is expected. In this course attendance is mandatory unless you have a valid reason to miss a session.

   Veterans and student service members with special circumstances or who are activated are encouraged to notify the instructor as soon as possible and are encouraged to provide Activation Orders.

Academic Honesty

The academic community is operated on the basis of honesty, integrity, and fair play. NDSU Policy 335: Code of Academic Responsibility and Conduct applies to cases in which cheating, plagiarism, or other academic misconduct have occurred in an instructional context. Students found guilty of academic misconduct are subject to penalties, up to and possibly including suspension and/or expulsion. Student academic misconduct records are maintained by the Office of Registration and Records. Informational resources about academic honesty for students and instructional staff members can be found at http://www.ndsu.edu/academichonesty.

   Please make sure that you understand common standards of academic integrity and plagiarism and consult the instructor if you are ever in doubt. I have a no tolerance policy for academic misconduct and students who commit such misconduct should expect, at minimum, to receive a failing grade for this class.

Students with Special Needs

Any students with disabilities or other special needs, who need special accommodations in this course, are invited to share these concerns or requests with the instructor and contact the Disability Services Office (`http://www.ndsu.edu/disabilityservices`) as soon as possible.

Discrimination and Harassment

NDSU is committed to providing a safe and non-discriminatory learning, living, and working environment for all members of its university community. NDSU's policy on discrimination and harassment is available here and the equity office provides information about filing complaints here. Any form of violence or harassment, including sexual assault, relationship violence, and stalking is unwelcome at the University. NDSU provides a Sexual and Gender-Based Harassment and Sexual Assault Resource Guide.

Low-grade discrimination and harassment can be especially pernicious in a classroom setting. Please read this blog post and work to avoid the behaviors that the post describes.

Please note that the instructor has a mandatory responsibility to report instances of discrimination, harassment, sexual assault, and retaliation, as described here. What this means is that as your professor, I am required to report any incidents of such misconduct that I observe, or that students or others report to me.