NOTE: This syllabus is tentative and may be updated during the quarter if needed. Any changes will be communicated in class or on the class website.

Data Science for Social Good
Communication 188C | Spring 2025
Monday and Wednesday, 2:00 to 3:15
Kaplan Hall A65

Professor: Homa Hosseinmardi
Email: homahm@ucla.edu
Office Hours: by appointment

Course Description

In this class, we will explore how data science can be applied to real-world problems in **sociology, psychology, public health, education, journalism, political science,** and beyond. We will cover techniques for collecting and parsing data, **ranging from interpretable linear models for small samples to identifying patterns in large-scale datasets in longitudinal studies**. Additionally, we will discuss principles for effectively communicating results to the target audience.
We will also examine recent research findings across various fields to learn best practices for applying these techniques in real-world scenarios. An introductory course in applied statistics and experience coding in **R, Python,** or another high-level programming language is recommended to get the most out of this course.

Textbooks and Readings

[1] Big Data and Social Science (Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences), 2nd Edition, by Ian Foster, Rayid Ghani, Frauke Kreuter (Editors)
[2] Computational Analysis of Communication, by Carlos Arcila Calderon, Damian Trilling, and Wouter van Atteveldt
[3] Bit by Bit: Social Research in the Digital Age, by Matthew Salganik (2018)
[4] Thinking Clearly with Data, by Anthony Fowler and Ethan Bueno de Mesquita
[5] Data Visualization - A practical introduction, by Kieran Healy

** Additional papers and readings will be assigned throughout the quarter (TBD)*

Course Schedule (Updated for 10 Weeks)

Week 1: Introduction – What is Data Science for Social Good? (March 31, April 2)

- **Session 1:**
  - Overview of DSSG and its societal applications
- **Session 2:**
  - Characteristics of big data
  - Data collection methods

  References and readings:
  - Chaper 2, Ref [3]

- "Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking." Ohme, Jakob, Theo Araujo, Laura Boeschoten, Deen Freelon, Nilam Ram, Byron B. Reeves, and Thomas N. Robinson, Communication Methods and Measures 18, no. 2 (2024): 124-141.

Week 2: Unstructured and Semi-Structured Data (April 7 & 9)

• **Session 3:**
  • Unstructured data
  • Text representation
  • From bag-of-words to embedding techniques

• **Session 4:**
  • Networked Data
  • Network inference (network features, community detection, link prediction)

Week 3: Computational Basics (April 14 & 16)

• **Session 5: Correlation and Causation**
  • Does a relationship exist?
  • Correlation is not causation
  • Data Visualization
  • Best practices

  References and readings:
  - *Data Visualization*: https://socviz.co/

• **Session 6: Descriptive Analytics**
  • Unsupervised methods
  • Clustering
  • Model selection

Week 4: Machine Learning Basics (April 21 & 23)

• **Session 7:**
  • Weak-supervised
  • Supervised methods

• **Session 8:**
  • Large Language Models (LLMs)
  • Generative AI

Week 5: Interpretable Machine Learning (April 28 & 30)

• **Session 9:**
  • Introduction to Explainable AI (XAI) techniques: SHAP, LIME
  • Fairness in AI

• **Session 10:**
  • Algorithmic audits
  • Red teaming

Week 6: Algorithmic Bias and Fairness in AI (May 5 & 7)

• **Session 1:**
  • **Guest Lecture: Ninareh Mehrabi** – Red teaming LLMs

• **Session 2:**
  • Mid-term project presentations
    • **Lightning talks**

Week 7: Real-World Applications – Health, Environment, and Justice (May 12 & 14)

• **Session 1:**

    - Science of Science
    - AI applications in public health, environment, and social justice
- **Session 2:**
    - **Guest Lecture: Loes Olde Loohuis** – Precision Medicine

Week 8: Networks and Influence – Understanding Social Structures (May 19 & 21)

- **Session 1:**
    - How algorithms shape online discourse and misinformation
    - Social media's role in politics and elections
- **Session 2:**
    - **Guest Lecture: Amir Ghasemian** – Social implications of a networked world

Week 9: Real-World Applications – Health, Environment, and Justice (May 26 & 28)

- **Session 1:**
    - **Holiday – No Class**
- **Session 2:**
    - **Guest Lecture: Rachit Dubey** – Data-Driven Policymaking

Week 10: Data Ethics and Project Presentations (June 2 & 4)

- **Session 1:**
    - Final project presentations
- **Session 2:**
    - Data ethics and digital privacy regulations

Grading

Quizzes (5, random): 20%

Homework (4 assignments): 20%

Project-based components:

- **Midterm: 20%**
    - Lightning talk (3-4 min) **10%**
    - Project proposal (2-page write-up) **10%**
- **Final Project: 40%**
    - Project write-up (8-10 pages) – **30%**
    - Presentation – **10%**

Academic Integrity

All students are expected to adhere to the UCLA Student Conduct Code. Cheating, plagiarism, or any form of dishonest behavior will be reported to the Dean of Students and may result in disciplinary action. For an overview of UCLA's academic integrity policies, see: https://deanofstudents.ucla.edu/academic-integrity

Accommodations & Accessibility

If you need academic accommodations due to a disability, please contact the Center for Accessible Education (CAE) at www.cae.ucla.edu, visit their office at A255 Murphy Hall, or call (310) 825-1501. To request accommodations, you first need to register with CAE through the student portal. If you are seeking to begin the registration process, please submit your request via the CAE website.