# Computational Social Science
# Rutgers University

# Syllabus

### Dr. Thomas Davidson

### Spring 2022

## CONTACT AND LOGISTICS

E-mail: thomas.davidson@rutgers.edu.

Website: https://github.com/t-davidson/css-spring-2022 and Canvas.

Class meetings: MW 2-3:20 p.m, Campbell Hall A5, College Avenue Campus.

Office hours: W 5:00-6:00 p.m, 109 Davison Hall, Douglass Campus or by appointment. Book a 15-minute meeting during regular hours using this link: https://thomasdavidson.youcanbook.me/

## COURSE DESCRIPTION

This course introduces students to the growing field of computational social science. Students will learn to collect and critically analyze social data using a range of techniques including natural language processing, machine learning, and agent-based modeling. We will discuss how these techniques are used by social scientists and consider the ethical implications of big data and artificial intelligence. Students will complete homework assignments involving coding in the R programming language to analyze several different datasets and will complete a group project to create a web-based application for data analysis and visualization.

## LEARNING GOALS

- Become competent at using R and RStudio
- Develop proficiency in data merging, cleaning, and basic analysis
- Understand and implement various methods for online data collection, natural language process, and machine learning
- Use RShiny to develop a web application for data analysis and visualization
- Identify important ethical issues related to the use of social data and computational methods

# PREREQUISITES

*Data 101* or equivalent. Enrolled students must have experience writing basic programs in a general purpose programming language, e.g. R, Python, Java, C. We will review the fundamentals for programming and data science in R in weeks 1-3.

# ASSESSMENT

- 10% Class participation
  - Students are expected to attend all class meetings and to actively participate in class discussions
- 60% Homework assignments (4 x 15%)
  - Students will complete a series of homework assignments to gain experience using R for data science
- 30% Final project
  - Students will complete a final project involving the use of RShiny to build an interactive web application for data analysis and visualization. Students may work individually or as part of a group.

## Rubric

Final grades will be determined according to the following rubric:

- A: 90-100%
- B+: 85-89%
- B: 80-84%
- C+: 75-79%
- C: 70-74%
- D: 60-69%
- F: <60%

# READINGS

Most of the readings will consist of chapters from the textbooks listed below. These readings are intended to build familiarity with key concepts and programming skills. Some weeks there will be an additional reading to highlight how data science techniques are used in empirical social scientific research. Links to each week's readings will be posted on Canvas.

## Textbooks

All textbooks are available for free online (hover over titles for links).

- Matthew Salganik. 2017. *Bit by Bit*. Princeton University Press. ISBN: 0691196109
- Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. (*R4DS*). O'Reilly Media, Inc. ISBN: 1491910399
- Silge, Julia, and David Robinson. 2017. *Text Mining with R: A Tidy Approach.* O'Reilly Media. ISBN: 1491981652

# COURSE RESOURCES

The course will be organized using two different tools, Github and Canvas. Canvas will be used for class communication, short quizzes, and to host readings. Github Classroom will be used for the submission of

assignments.

# TECHNOLOGY REQUIREMENTS

Students will be required to have access to a computer to complete assignments. Ideally, students should bring a laptop computer to class.

Please visit the Rutgers Student Tech Guide page for resources available to all students. If you do not have the appropriate technology for financial reasons, please email Dean of Students deanofstudents@echo.rutgers.edu for assistance. If you are facing other financial hardships, please visit the Office of Financial Aid.

# COURSE POLICIES

The Rutgers Sociology Department strives to create an environment that supports and affirms diversity in all manifestations, including race, ethnicity, gender, sexual orientation, religion, age, social class, disability status, region/country of origin, and political orientation. We also celebrate diversity of theoretical and methodological perspectives among our faculty and students and seek to create an atmosphere of respect and mutual dialogue. We have zero tolerance for violations of these principles and have instituted clear and respectful procedures for responding to such grievances.

Students must abide by the Code of Student Conduct at all times, including during lectures and in participation online.

Students must abide by the university's Academic Integrity Policy. Violations of academic integrity will result in disciplinary action. Please review this policy or contact Professor Davidson if there is something you are unsure about.

If you have a documented disability and require accommodations to obtain equal access in this course, please contact me during the *first week of classes*. Students with disabilities must be registered with the Office of Student Disability Services and must provide verification of their eligibility for such accommodations. See end of syllabus for further details.

*COVID-19*: Following university policy, students are required to wear masks at all times during in-person classes. I recommend students wear an N-95 or equivalent for maximum protection. While the science is continually evolving, current evidence suggests that cloth masks are ineffective at preventing infection during periods of sustained social interactions. Please do not attend class or office hours if you have symptoms or are required to quarantine. If you or your family are affected in any way that impedes your ability to participate in this class, please contact me as soon as you can so that we can make necessary arrangements.

# COURSE OUTLINE

## Week 1, 1/19 (Wednesday only)

**Introduction to Computational Social Science**

*Readings*

- Wednesday:
    - *Bit by Bit*, C1
    - *R4DS*: C1 & 27 [Note: Chapter numbers correspond to the online book; physical book numbers are different]

### Week 2, 1/24 & 1/26

**Data Structures in R**

*Readings*

- Monday:
    - *R4DS*: C2,4, skim 20.
- Wednesday:
    - *R4DS*: C17-19, 21.

### Week 3, 1/31 & 2/2

**Programming Fundamentals**

*In-Person Instruction Resumes*

*Readings*

- Monday:
    - *R4DS*: C5, 9,
- Wednesday:
    - *R4DS*: C10, 13.

*Assignment 1 released: Using R for Data Science.*

### Week 4, 2/7 & 2/9

**Data Collection I: Collecting Data Using Application Programming Interfaces**

*Readings*

- Monday:
    - *Bit by Bit*, C2
- Wednesday:
    - *R4DS*: C3

### Week 5, 2/14 & 2/16

**Data Collection II: Scraping Data From the Web**

*Readings*

- Monday:
    - *Bit by Bit*, C6
- Wednesday:
    - *R4DS*: C14, 16

*Recommended*

- Fiesler, Casey, Nate Beard, and Brian C Keegan. 2020. "No Robots, Spiders, or Scrapers: Legal and Ethical Regulation of Data Collection Methods in Social Media Terms of Service." In *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media*, 187–96.

*Assignment 2: Collecting and storing data released.*

## Week 6, 2/21 & 2/23

**Data Collection III: Online Experiments and Surveys**

*Readings*

- Monday:
    - *Bit by Bit*, C3-5
- Wednesday:
    - TBD
    - R Shiny tutorial: https://shiny.rstudio.com/tutorial/

## Week 7, 2/28 & 3/2

**Natural Language Processing I: The Vector-Space Model**

*Readings*

- Monday:
    - Evans, James, and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42 (1): 21–50.
- Wednesday:
    - *Text Mining with R*, C1 & 3

## Week 8, 3/7 & 3/9

**Natural Language Processing II: Word Embeddings**

*Readings*

- Monday:
    - *Text Mining with R*: C5.
- Wednesday:
    - Hvitfeldt, Emil and Julia Silge. 2020 *Supervised Machine Learning for Text Analysis in R.* Chapter 5.

*Recommended*

- Kozlowski, Austin, Matt Taddy, and James Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." *American Sociological Review*, September, 000312241987713.

*Spring Break*

## Week 9, 3/21 & 3/23

**Natural Language Processing III: Topic Models**

*Assignment 3: Natural language processing released.*

*Readings*

- Monday:
    - *Text Mining with R*: C6
    - Mohr, John, and Petko Bogdanov. 2013. "Introduction—Topic Models: What They Are and Why They Matter." *Poetics* 41 (6): 545–69.
- Wednesday:

– Roberts, Margaret, Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses: Structural Topic Models for Survey Responses." *American Journal of Political Science* 58 (4): 1064–82.

## Week 10, 3/28 & 3/30

**Machine Learning I: Prediction and Explanation**

*Readings*

- Monday:
    – Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45: 27–45.
- Wednesday:
    – Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106.

## Week 11, 4/4 & 4/6

**Machine learning II: Text Classification**

*Assignment 4: Machine learning released.*

*Readings*

- Monday:
    – TBD
- Wednesday:
    – Barberá, Pablo, Amber E. Boydstun, Suzanna Linn, Ryan McMahon, and Jonathan Nagler. 2020. "Automated Text Classification of News Articles: A Practical Guide." *Political Analysis*, June, 1–24.

*Recommended*

- Hanna, Alex. 2013. "Computer-Aided Content Analysis of Digitally Enabled Movements." *Mobilization: An International Quarterly* 18 (4): 367–388.

## Week 12, 4/11 & 4/13

**Machine learning III: Challenges**

*Readings*

- Monday:
    – Salganik, Matthew, Ian Lundberg, Alexander Kindel, et al. 2020. "Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration." *Proceedings of the National Academy of Sciences*.
    – Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of Machine Learning Research*, 81:1–15.
- Wednesday:
    – Project workshop

## Week 13, 4/18 & 4/20

**Machine learning IV: Image Classification**

*Readings*

- Monday:
  - Torres, Michelle, and Francisco Cantú. 2021. "Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data." *Political Analysis*, April, 1–19.
  - Gebru, Timnit, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. 2017. "Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighborhoods across the United States." *Proceedings of the National Academy of Sciences* 114 (50): 13108–13.
- Wednesday:
  - Project workshop

## Week 14, 4/25 & 4/27

**Simulation and Agent-based Models**

*Readings*

- Monday:
  - Macy, Michael, and Robert Willer. 2002. "From Factors to Factors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28 (1): 143–66.
- Wednesday:
  - TBD

## Week 15, 5/2 (Monday only)

**Project presentations**

*Final projects due TBD*

# Additional information

*The Rutgers University Student Assembly urges that the following information be included at the end of every syllabus.*

## Report a Bias Incident

Bias is defined by the University as an act, verbal, written, physical, psychological, that threatens, or harms a person or group on the basis of race, religion, color, sex, age, sexual orientation, gender identity or expression, national origin, ancestry, disability, marital status, civil union status, domestic partnership status, atypical heredity or cellular blood trait, military service or veteran status.

If you experience or witness an act of bias or hate, report it to someone in authority. You may file a report online and you will be contacted within 24 hours. The bias reporting page is here.

## Counseling, ADAP & Psychiatric Services (CAPS)

(848) 932-7884 / 17 Senior Street, New Brunswick, NJ 08901 / Link to website.

CAPS is a University mental health support service that includes counseling, alcohol and other drug assistance, and psychiatric services staffed by a team of professionals within Rutgers Health services to support students' efforts to succeed at Rutgers University. CAPS offers a variety of services that include: individual therapy, group therapy and workshops, crisis intervention, referral to specialists in the community, and consultation and collaboration with campus partners.

## Crisis Intervention

Link to website.

## Report a Concern:

Link to website.

## Violence Prevention & Victim Assistance (VPVA)

(848) 932-1181 / 3 Bartlett Street, New Brunswick, NJ 08901 / Link to website. The Office for Violence Prevention and Victim Assistance provides confidential crisis intervention, counseling and advocacy for victims of sexual and relationship violence and stalking to students, staff and faculty. To reach staff during office hours when the university is open or to reach an advocate after hours, call 848-932-1181.

## Disability Services

(848) 445-6800 / Lucy Stone Hall, Suite A145, Livingston Campus, 54 Joyce Kilmer Avenue, Piscataway, NJ 08854 / Link to website

Rutgers University welcomes students with disabilities into all of the University's educational programs. In order to receive consideration for reasonable accommodations, a student with a disability must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation: see guidelines. If the documentation supports your request for reasonable accommodations, your campus's disability services office will provide you with a Letter of Accommodations. Please share this letter with your instructors and discuss the accommodations with them as early in your courses as possible. To begin this process, please complete the Registration form on the ODS web site.