SoDA 496

Spring 2023

Instructor:	Mike Burnham	**Time:**	Mo, We, Fr; 1:25-2:15
Email:	mlb6496@psu.edu	**Place:**	026 Burrowes
Course Page:	Github	**Office Hours:**	TBD

Course Overview

This seminar is designed to help you turn data into accessible and actionable knowledge. Whereas the fall term focused on data collection and research design, this term focuses on modeling, writing, and presenting findings. The course is broadly divided into two sections. The first section focuses on modeling social behavior both for inference and prediction while the second is on technical literacy. By the end of this course, students should be able to:

1. Communicate complex ideas in clear and concise writing
2. Understand how to model data for inference and prediction
3. Evaluate the validity of empirical claims
4. Employ multiple analytical tools including virtual environments, cloud computing, SQL, and Python
5. Analyze Data across a wide variety of mediums, including text, image, audio, and tabular data.

Course Materials

- Strunk, William, E.B. White (1918). The Elements of Style.

A copy of The Elements of Style is required for this course. You can purchase a copy for $5-$10. We will do some reading from the manual and your writing assignments will be graded according to its guidelines.

Course Requirements and Evaluation

Homework: 20% Students will complete four homework assignments that include writing and technical exercises. Details will be provided as we progress through the course.

Midterm: 20% A midterm exam that covers the first half of the course will be given. Questions will be in short answer format. Students are encouraged to use course slides as a study guide.

Attendance and Participation: 20% This is a small seminar. As such, attendance and participation are essential. To receive full grades students will:

1. Attend every class (except one day automatically forgiven).
2. Read the required materials before every class.
3. Participate in class discussion.
4. Present progress and provide feedback to peers.

Final Presentation: 10% During the final week of class, students will present their research to the class. Students are free to use whatever visual aids they deem appropriate (slides, posters, dashboards, etc.). Presentations should include theory, research methods, results, and a discussion about validity.

Final Paper: 30% Your final paper will include modeling and analysis of the data you collected during the Fall semester. Papers should test a clear hypothesis and communicate that the author understands their model and data. I will not be grading on if the hypothesis is supported, but on the quality of the research. Thus, papers should justify their choice of research design and call attention to the strengths and weaknesses in their data and model.

There is no minimum or maximum page limit. Rather, papers will be graded on if they are both complete and well edited. Students should not omit any important information, but they should not mistake a long paper for a good paper. At a minimum, each paper should include the following sections:

1. Introduction
2. Literature Review
3. Data and Methods
4. Analysis
5. Discussion

Some papers may want to include an additional theory section or multiple analysis sections for distinct hypotheses.

Letter Grade Conversion

Letter Grade	Numeric Grade
A	93%+
A-	90-93%
B+	87-90%
B	83-87%
B-	80-83%
C+	77-80%
C	70-77%
D	60-70%
F	60%-

Expectations/Procedures

RESPECT. In this course, we are all engaged in the endeavor of building a stronger understanding of Quantitative Methodology. Everyone comes to this course with a different background in the subject (particularly the statistics portion of the course). It is important that we all treat each other with the utmost respect.

OFFICE HOURS. Please come. I'm here to help. If my office hours conflict with another commitment, please e-mail me to find a time that works for both of us to meet. I'm really serious about this; coming to office hours with serious questions is the best predictor of success in this course.

SOFTWARE. Course assignments will focus on using Python and SQL as these are the most commonly used tools for data science positions in industry. You may use either R or Python to complete the analysis for your final paper.

WORKING TOGETHER. I encourage you to work together on homework. There is no better way to master this material than to work together on it. However, every keystroke of what you type and every mark you make with a pencil or pen must be your own work. You may not collaborate on the midterm and final exams, and you will complete your own final project (though you may—and should!—help each other with the project as we go).

LATE ASSIGNMENTS. Assignments not submitted by the assigned due date and time are late. Late submissions will be accepted; however, they will be subject to a one-half grade (5%) per day (including weekends) late penalty. All assignments must be completed in order to pass this course.

EXTENSIONS. Extensions will be granted in only the most severe circumstances. If you foresee the need for an extension, one needs to be requested and granted at least 24 hours before the due date. No one is entitled to an extension; they will be offered only at my discretion.

ACADEMIC DISHONESTY. I take violations of the University's academic dishonesty policy very seriously; it is printed on the next page. Please review the policy and let me know if you have any questions.

REGARDING GRADES. I do not give grades. You earn grades. It is essential that you are proactive regarding your performance in this course; do not wait until grades are posted and then ask how your grade could be improved. At that point, barring a mathematical error on my part, it cannot be. If, at any point, you are unsure of your current standing in the course, please come to my office hours. I may (or may not) offer extra credit assignments to the entire class during the semester. I am sometimes asked about extra-credit or additional assignments after the final grades have been tallied by students who are unhappy with their grades at the end of the semester. I will not offer such assignments to the class or individual students.

ACADEMIC DISHONESTY. The Department of Political Science, along with the College of the Liberal Arts and the University, takes violations of academic dishonesty seriously. Observing basic honesty in one's work, words, ideas, and actions is a principle to which all members of the community are required to subscribe.

All course work by students is to be done on an individual basis unless an instructor clearly states that an alternative is acceptable. Any reference materials used in the preparation of any assignment must be explicitly cited. Students uncertain about proper citation are responsible for checking with their instructor.

In an examination setting, unless the instructor gives explicit prior instructions to the contrary, whether the examination is in class or take home, violations of academic integrity shall consist but are not limited to any attempt to receive assistance from written or printed aids, or from any person or papers or electronic devices, or of any attempt to give assistance, whether the one so doing has completed his or her own work or not.

Lying to the instructor or purposely misleading any Penn State administrator shall also constitute a violation of academic integrity.

In cases of any violation of academic integrity it is the policy of the Department of Political Science to follow procedures established by the College of the Liberal Arts. More information on academic integrity and procedures followed for violation can be found at: http://www.la.psu.edu/current-students/student-services/academic-integrity/academic-integrity

NOTE TO STUDENTS WITH DISABILITIES. Penn State welcomes students with disabilities into the University's educational programs. Every Penn State campus has an office for students with disabilities.

The Student Disability Resources Web site provides contact information for every Penn State campus. For further information, please visit the Student Disability Resources Web site.

In order to receive consideration for reasonable accommodations, you must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation. If the documentation supports your request for reasonable accommodations, your campus's disability services office will provide you with an accommodation letter. Please share this letter with your instructors and discuss the accommodations with them as early in your courses as possible. You must follow this process for every semester that you request accommodations.

COUNSELING AND PSYCHOLOGICAL SERVICES. Many students at Penn State face personal challenges or have psychological needs that may interfere with interfere with their academic progress, social development, or emotional wellbeing. The university offers a variety of confidential services to help you through difficult times, including individual and group counseling, crisis intervention, consultations, online chats, and mental health screenings. These services are provided by staff who welcome all students and embrace a philosophy respectful of clients' cultural and religious backgrounds, and sensitive to differences in race, ability, gender identity and sexual orientation.

Counseling and Psychological Services at University Park (CAPS): 814-863-0395 Penn State Crisis Line (24 hours/7 days/week): 877-229-6400 Crisis Text Line (24 hours/7 days/week): Text LIONS to 741741

EDUCATIONAL EQUITY AND REPORTING BIAS INCIDENTS. State takes great pride to foster a diverse and inclusive environment for students, faculty, and staff. Acts of intolerance, discrimination, or harassment due to age, ancestry, color, disability, gender, gender identity, national origin, race, religious belief, sexual orientation, or veteran status are not tolerated and can be reported through Educational Equity via the Report Bias webpage. You may also contact one of the following offices:

University Police Services, University Park: 814-863-1111 Multicultural Resource Center, Diversity Advocate for Students: 814-865-1773 Office of the Vice Provost for Educational Equity: 814-865-5906 Office of the Vice President for Student Affairs: 814-865-0909 Affirmative Action Office: 814-863-0471

Call 911 in cases where physical injury has occurred or is imminent.

STUDENT CARE & ADVOCACY OFFICE. College presents a number of challenges for students, and Penn State maintains an office of Student Care & Advocacy that can point you in the right direction if you are facing any of the following issues:

- Death of an immediate family member
- Family crisis
- Mental health concern
- Self-injurious behavior
- Food insecurity
- Housing insecurity
- Medical emergency and/or hospitalization
- Local natural disaster
- Academic distress
- Unexpected events or challenges

If you have questions, concerns, or need more information, please do not hesitate to contact that office by phone at 814-863-2020 or by email at StudentCare@psu.edu. They encourage you to call or e-mail ahead.

EXTENDED ABSENCES. During your enrollment at Penn State, unforeseen challenges may arise. If you ever need to miss an extended amount of class in such a circumstance, please notify your professor so you can determine the best course of action to make up missed work. If your situation rises to a level of difficulty you cannot manage on your own with faculty support, reach out to the Student Care & Advocacy office by phone at (814-863-2020) or email them at StudentCare@psu.edu. Office hours are Monday-Friday, 8 a.m. to 5 p.m.

Tentative Course Outline:

Week 1 (Jan. 9, 11, 13): Course introduction

Week 2 (Jan. 18, 20): Writing

- Strunk and White Ch. 2
- Strunk and White Ch. 5

Week 3 (Jan. 23, 25, 27): Hardware, software, and data management

- Jupyter Lab
- Virtual Environments
- SQL

Week 4 (Jan. 30, Feb. 1, 3): Modeling for inference

- Haidt, Jonathan. "The emotional dog and its rational tail: a social intuitionist approach to moral judgment." Psychological review 108, no. 4 (2001): 814.
- Boussalis, Constantine, Travis G. Coan, Mirya R. Holman, and Stefan Müller. "Gender, candidate emotional expression, and voter reactions during televised debates." American Political Science Review 115, no. 4 (2021): 1242-1257.

Week 5 (Feb. 6, 8, 10): Validity

- What's wrong with social science and how to fix it
- Psychology might be a big stinkin' load of hogwash and that's just fine
- Barrett, Lisa Feldman. "The theory of constructed emotion: an active inference account of interoception and categorization." Social cognitive and affective neuroscience 12, no. 1 (2017): 1-23.

Week 6 (Feb. 13, 15, 17): Modeling for prediction

- Cranmer, Skyler J., and Bruce A. Desmarais. "What can we learn from predictive modeling?." Political Analysis 25, no. 2 (2017): 145-166.
- Dietrich, Bryce J., Ryan D. Enos, and Maya Sen. "Emotional arousal predicts voting on the US supreme court." Political Analysis 27, no. 2 (2019): 237-243.

Week 7 (Feb. 20, 22, 24): AI and foundation models

- Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. "On the opportunities and risks of foundation models." arXiv preprint arXiv:2108.07258 (2021).

Week 8 (Feb. 27, Mar. 1, 3): Midterm and one-on-one meetings

Week 9 (Mar. 6, 8 10): Spring break

Week 10 (Mar. 13, 15, 17): Time Series Week 1: Stationarity

- Linn, Suzanna, Lebo, Matthew and Webb, Clayton. "A Practical Guide to Time Series Analysis." Chapters 2 and 3.

Week 11 (Mar. 20, 22, 24): Time Series Week 2: Single Equation Regression

- Linn, Suzanna, Lebo, Matthew and Webb, Clayton. "A Practical Guide to Time Series Analysis." Chapter 5.

Week 12 (Mar. 27, 29, 31): Network analysis

- Lazer, David. "Networks in political science: Back to the future." PS: Political Science Politics 44, no. 1 (2011): 61-68.

Week 13 (Apr. 3, 5, 7): Text Analysis Week 1: Distributional Semantics

- O'Connor, Brendan, David Bamman, and Noah A. Smith. "Computational text analysis for social science: Model assumptions and complexity." In Second workshop on comptuational social science and the wisdom of crowds (NIPS 2011). 2011.

- Goldberg, Yoav, and Omer Levy. "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method." arXiv preprint arXiv:1402.3722 (2014).

Week 14 (Apr. 10, 12, 14): Text Analysis Week 2: Transfer Learning and Foundation Models

- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

- Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D. Kaplan, Prafulla Dhariwal, Arvind Neelakantan et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877-1901.

Week 15 (Apr. 17, 19, 21): Visualization

- Börner, Katy, Andreas Bueckle, and Michael Ginda. "Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments." Proceedings of the National Academy of Sciences 116, no. 6 (2019): 1857-1864.

Week 16 (Apr. 24, 26, 28): Presentations