



THE OHIO STATE UNIVERSITY

JOHN GLENN COLLEGE OF PUBLIC AFFAIRS

**PUB AFRS 5800 – 3 Credits
(Spring 2018)**

Data Sciences for Public Policy Analysis
(Undergraduate)

Prerequisites:

- *PA 3000 or concurrent enrollment;**
- *Would suggest some programming experience;**
- *Senior or Junior Standing**
- *Preference Given to Glenn College B.S. or B.A. Students.**

Monday, Wednesday; 3:55-5:15; Hands on sessions (Fridays, 2 hours/week) – Distance Learning Lab Page Hall

Instructor: Joshua Hawley
Office: 250 Page Hall
Email: Hawley.32@osu.edu
Office Hours: TBD

Course goal and objectives:

Goals

This course provides an orientation to the use of data for decision-making in the areas of public sector. The emphasis in the course is how to use data in context – when organizations require the analysis of sophisticated data in order to achieve goals or priorities. Topics in the class include;

1. Data use in the public sector;
2. Development of tools for decision-making;
3. Use of visualization techniques;
4. Geographic analysis.

This course is the first required part of a program designed for students interested in exploring careers in policy research and evaluation using data analytics. Students will receive instruction on data based methods to investigate policy problems in this class. The class introduces software necessary for acquiring, analyzing and visualizing the results of policy analysis. Moreover, the class offers contextual information on data governance, legal frameworks, and institutional

review procedures required to analyze data from government. Finally, the course provides instruction on methodological tools (regression analysis, GIS, visualization) that are often used to conduct formal evaluations of public policies. This course can be taken as a stand-alone 3 credit elective. It can also be taken as part of the preparation for the Public-Sector Data Science Internship Program, which is offered in the Summer.

Taking the class does not mean you are accepted into the internship program. Application procedures for the internship program require a formal application submission and interview. In 2018 the Due Date for the Internship is February 1, 2018.

Objectives

1. Upon completing this course the student will gain an understanding of the current technologies and ways government are using data to inform policy decisions;
2. Upon completing this course the student will be able to understand the structure and use of both administrative data and survey data for policy decisions;
3. Upon completing this course the student will be able to conduct analyses of administrative and survey data on applied policy problems; and,
4. The student will be able to write a structured policy memo that informs policy decisions based on data analyses.

Course Materials

Required:

Students can access textbook information via the Barnes & Noble bookstore website: www.shopOhioState.com as well as from their BuckeyeLink Student Center. This information is disseminated by B&N to all area bookstores. You may buy from a store of your choice and/or shop for books (always use ISBN# for searches) on line.

Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (Eds.). (2016). *Big data and social science: a practical guide to methods and tools*. CRC Press.

Hadley Wickham and Garrett Grolemund. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* 1st Edition. O'Reilly

Grading:

The course will be graded using a satisfactory/unsatisfactory grade (S/U). In this case, satisfactory is defined as “mastery” of the material in the class. The material in the class are the technical assignments presented to learn the programming, and also writing assignments designed to ensure that students can present the results of technical programming in written forms that policy makers can understand. The cut-score for mastery on a specific assignment will

be available, but a general guideline is that we expect a 90% on weekly assignments and the final exam to be considered for mastery.

A word about this grading rule. Since the primary motivation for this class is preparation for an internship, we are teaching behavioral norms. Internship supervisors will require a high level of performance, and therefore we are running the class under those same expectations. This mastery expectation is based on the principles of competency-based learning. I would be happy to talk with you about this grading expectation if you have any concerns.

Assignments

Class Participation - 20%

As a seven-week course, the schedule is compressed for the amount of information required to prepare students for the internship experience. More than one absence will result in a reduction of the attendance grade.

Final Exam – 30%

A final exam will be given that provides a data set, a semi-structured problem, and a format for a written report. You will be given one week to produce a memo on a policy problem using the materials introduced in the class.

Course Exercises – 50% (7 points each)

In each of the seven weeks there will be task assigned that are designed to apply material from lectures. These tasks relate to the ultimate goal of producing research in a state agency over the summer as part of the internship experience

Course Policies

Writing Assistance

The Writing Center is a key resource for writing assistance. The following is from their website (<https://cstw.osu.edu/students>): The Writing Center works “one-one-one with undergraduate and graduate students, faculty, and staff at Ohio State on writing projects. Writing consultants are able to review writing at any stage, from brainstorming to a final draft, as well as help with non-paper assignments such as presentations, blogs, etc.” Consultants are available by appointment, on a walk-in basis or online.

We also suggest that students have taken 2367 in the Glenn College prior to enrolling in this class.

Attendance Policy

See “Class Participation” under assignments.”

Digital Etiquette

Lap-tops, tablets and phones are required for this class. However, limit personal lap-top and tablet use in class for class purposes only (ex. note taking). Phones are not to be out or used in class except for cases of emergency. Please let me know if you need to have your phone out in class. If you think you may distract other students, consider sitting towards the back. If a student is distracting you, feel free to change seats.

Academic Misconduct Policy

ACADEMIC INTEGRITY (ACADEMIC MISCONDUCT)

From: <http://oaa.osu.edu/coamfaqs.html#academicmisconductstatement>

The Ohio State University and the Committee on Academic Misconduct (COAM) expect that all students have read and understand the University's [Code of Student Conduct](#), and that all students will complete all academic and scholarly assignments with fairness and honesty. Failure to follow the rules and guidelines established in the University's Code of Student Conduct may constitute "Academic Misconduct." Sanctions for the misconduct could include a failing grade in this course and suspension or dismissal from the University.

In the Ohio State University's [Code of Student Conduct](#), Section 3335-23-04 defines academic misconduct as: "Any activity that tends to compromise the academic integrity of the University, or subvert the educational process." Examples of academic misconduct include (but are not limited to) plagiarism, collusion (unauthorized collaboration), copying the work of another student, and possession of unauthorized materials during an examination. Ignorance of the University's [Code of Student Conduct](#) is never considered an "excuse" for academic misconduct.

Other sources of information on academic misconduct (integrity) to which you can refer include The Committee on Academic Misconduct web page: <http://oaa.osu.edu/coam.html>
Ten Suggestions for Preserving Academic Integrity: <http://oaa.osu.edu/coamtensuggestions.html>
Eight Cardinal Rules of Academic Integrity: www.northwestern.edu/uacc/8cards.html
If you have any questions about the above policy or what constitutes academic misconduct in this course, please contact me.

Accommodation Policy

The University strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability (including mental health, chronic or temporary medical conditions), please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with Student Life Disability Services. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. **SLDS contact information:** slds@osu.edu; 614-292-3307; slds.osu.edu; 098 Baker Hall, 113 W. 12th Avenue.

Mental Health Statement

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty

concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student’s ability to participate in daily activities. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know is suffering from any of the aforementioned conditions, you can learn more about the broad range of confidential mental health services available on campus via the

Office of Student Life Counseling and Consultation Services (CCS) by visiting ccs.osu.edu or calling 614-292- 5766. CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at 614--292--5766 and 24 hour emergency help is also available through the 24/7 National Suicide Prevention Hotline at 1--800--273--TALK or at suicidepreventionlifeline.org. Also, the OSU Student Advocacy Center is a resource to help students navigate OSU and to resolve issues that they encounter at OSU – visit <http://advocacy.osu.edu/> .

Course Schedule

Week 1: Overview of Big Data,

Data science, technical capacity in government, decision making and data.

Class	Topic	Reading
March 19	Data Science Overview, Administrative Data	Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (Eds.). (2016). <i>Big data and social science: a practical guide to methods and tools</i> . CRC Press.; Chapter 1 Pentland, A. S., & Berinato, S. (2014, November). With big data comes big responsibility. <i>Harvard Business Review</i> . Retrieved from https://hbr.org/2014/11/with-big-data-comes-big-responsibility .
March 21	Data Science and Public Policy	Mergel, I., Rethemeyer, R. K. and Isett, K. (2016), Big Data in Public Affairs. <i>Public Admin Rev</i> , 76: 928–937. doi:10.1111/puar.12625 D. Weimer & A. Vining (2011), <i>Policy Analysis Chapter 11: Gathering Information for Policy Analysis</i> Advanced D. Stone (2012) <i>Policy Paradox</i> , Chapter 5: Numbers

Memo 1: Respond to prompt on Canvas; “How does government currently use data to make decisions? Provide at least 2 examples of a situation where a local or state government makes use of data.

Week 2: Data Science Fundamentals - 1

Class	Topic	Reading
March 26	Overview of the R Interface; Inputting Public Data	Foster et. al, Chapter 2; R for Data Science Advanced Kuhn & Johnson (2013), Applied Predictive Modeling; “Chapter 3: Data Preprocessing”
March 28	Continue R Overview, Basic merged data; Simple descriptive output	R for Data Science; Advanced R. Nisbet, J. Elder, G. Miner (2015), Handbook of Statistical Analysis and Data Mining Applications; Chapter 4: Data Understanding and Preparation

Homework, CITI IRB Training; Read Basic HHS policy for protection of human research subjects, 45 C.F.R. 46 (2009).

Memo 2 – Respond to prompt on Canvas; “Answer the following questions based on the data we provide in class; 1) How has the demographic composition of Ohio’s schools changed in the past 10 years?; 2) What characteristic (test scores, poverty, demographic) seems to change the most over time?” Show at least four measures (summary statistics) that will help answer this question.

Week 3: Data Fundamentals 2 and Ethics

Class	Topic	Reading
April 2	Ethics in Big Data	Boyd, d., & Crawford, K. (2012). Critical questions for big data. Information, Communication and Society, 15(5), 662–679. Foster et. al (2016), Chapter 11: Privacy and Confidentiality Dannen, C. (2009, July 1). On Facebook? New algorithm can guess your SSN. Fast Company.

		Retrieved from http://www.fastcompany.com/1305136/facebook-newalgorithm-can-guess-your-ssn NSA Domestic Surveillance Directorate. (n.d.). Your data: If you have nothing to hide, you have nothing to fear. Retrieved from https://nsa.gov1.info/data .
April 4	Complex Merging in Data	R for Data Science, chapter, 3* Foster et. al (2016). Chapter 3: Record Linkage

*Homework: What can you find out about yourself online? Do you use google, facebook or twitter? Check out your youtube viewing history for example. How much time do you spend on cat videos? I find it useful to also check the NSA document from the readings to determine how the government might be tracking me.

Carry out complex merge using the R application as directed in class. Provide summary statistics as discussed in class.

Week 4: Using Data to Produce Metrics, Dashboards, Scorecards

Class	Topic	Reading
April 9	Data Uses, Subsetting Data; Creating Metrics	Weiss, C. (1998). Improving the Use of Evaluations; Whose Job is it anyway. <i>Advances in Educational Productivity</i> , 7, 263-276. Jevin West (2014) How to improve the use of metrics: learn from game theory. <i>Nature</i> 465:871-872
April 11	Reproducibility; Avoiding Bullshit Dashboards	Robert Matthews (2000) Storks deliver babies (p=0.008). <i>Teaching Statistics</i> 22:36-38 Food Stamp Fraud Case (http://callingbullshit.org/case_studies/case_study_foodstamp_fraud.html) Ganapati, S. (2011). Using dashboards in government. Retrieved from IBM Center for the Business of Government website: http://www.businessofgovernment.org/report/use-dashboards-government . Advanced

		Foster et al (2016), Chapter 10: Errors and Inference
--	--	---

Homework: Find one example of bullshit using data analysis in a policy area such as education or health care. Write a one-page summary of what you think is wrong with the data analysis or data presentation.

Week 5: Data Visualization and Causality

Class	Topic	Reading
April 16	Data Visualization	Edward Tufte (1983) <i>The Visual Display of Quantitative Information</i> Chapters 2 (Graphical integrity) and 5 (Chartjunk: vibrations, grids, and ducks). Advanced Foster et al (2016), Chapter 9: Information Visualization
April 18	Correlation and Causation; Data relevance	Richard Murnane & John Willett (2011). <i>Methods Matter</i> , Chapters 1-3 (The challenge of educational research, the importance of theory, designing research to answer causal questions) LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011, Winter). Big data, analytics and the path from insights to value. MIT Sloan Management Review, 21–32 Advanced Kuhn & Johnson (2013), Chapter 18: Measuring Predictor Importance

Homework: Memo 5, Structured memo answering the importance of variables in the test data set. TBD.

Week 6: Prediction and Geography

Class	Topic	Reading
April 23	Prediction	Edward Tufte (1974). <i>Data Analysis for Politics and Policy</i> (Chapter 2: Predictions and Projections) Martin O'Malley (2014). <i>Doing What Works: Governing i</i> http://www.arcgis.com/home/ggge/webmap/viewer.html?webmap=7d987ba67f4640f0869acb82ba06422

		<p>8n the Age of Big Data. <i>Public Admin. Review</i>, 74,5,555-556</p> <p>Advanced</p> <p>Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland. A. (2014). Once upon a crime: Towards crime prediction from demographics and mobile data. In Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey, November 12–16, 2014 (pp. 427–434). New York, NY: Association for Computing Machinery.</p> <p>Kuhn, Max, and Kjell Johnson. <i>Applied predictive modeling</i>. New York: Springer, 2013; Section on Logistic Regression and Chapter 6: Linear Regression and Its Cousins</p>
April 25	Geographical Analysis	<p>R-Intro to Spatial Data https://cran.r-project.org/doc/contrib/intro-spatial-rl.pdf Example Maps - (Earthquakes) http://www.arcgis.com/home/group.html?id=d55ba7a328924596b9fa9d5d1876a970#overview (Wildfires) http://www.arcgis.com/home/item.html?id=df8bcc10430f48878b01c96e907a1fc3 (Education)</p> <p>Advanced</p> <p>Sirer, Maroulis et. al (2015), The Currents Beneath the Rising Tide of School Choice, <i>JPAM</i></p>

Memo: NONE For Week 6

Week 7: Demonstration

Class	Topic	Reading
April 30	Final Test Office Hours	Open house for help
May 2	Final Test Office Hours	Turn in Final Exam

Other tasks to do for people in internship

1. Finish your IRB Clearance (CITI training)
2. Meet with staff from agency
3. Write short summary of data tasks that agency is interested in and how it relates to theory presented in class.
4. Summary of programming needs from agencies
5. Project Work: Develop formal work order and plan for work with agency liaison.

SAMPLE