



THE OHIO STATE UNIVERSITY

JOHN GLENN COLLEGE OF PUBLIC AFFAIRS

PUB AFRS 4040– 3 Credits

Public Sector Data Sciences and Management
(Undergraduate)

Prerequisites:

GE Data Analysis Class

Instructor: Joshua Hawley
Office: Page Hall and CHRR
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 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 the following;

1. Data use and decision making in the public sector;
2. Legal and ethical framework for data use in the public sector;
3. Use of data sciences in evaluation, planning, and performance reporting;
4. Visualization techniques in public policy analysis;
5. Geographic analysis; and
6. Statistical tools to improve the use of data in decision making.

This course is designed to be an introduction for students interested in big data and public affairs. Data science methods will be described as utilized in the public sector. This requires both an understanding of the technical and methodological challenges of working with government data, but also the ethical and legal restrictions government places on data utilization. This class builds on some of the core content areas the Glenn College curriculum teaches. For example, the sequence of policy analysis and evaluation courses. Secondly, it deepens knowledge in the decision sciences courses.

The course assumes a familiarity with statistics as in the GE required in Data Sciences. Students will get the most of the class having already completed at least one course in statistical programming. The class is not a substitute for statistical training or programming skill.

The student project conducted as part of this class makes use of statistical programming skills. Students conduct a semester long project on an applied data topic that is relevant to state agency's needs.

Objectives

1. Upon completing this course, students will
 - a. ... gain an understanding of the current technologies and ways government are using data to inform policy decisions;
 - b. ... be able to understand the structure and use of both administrative data and survey data for policy decisions;
 - c. ... learn the ethical and legal framework for using public data to carry out data science;
 - d. ... be able to conduct analyses of administrative and survey data on applied policy problems; and,
 - e. ... 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.

Recommended

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.

John Fantuzzo and Dennis P. Culhane, eds., (2015). *Actionable Intelligence Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government* (Palgrave, NY)

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 traditional OSU grading system.

A 100% to 93%
A- < 93% to 90%
B+ < 90% to 87%
B < 87% to 83%
B- < 83% to 80%

C+ < 80% to 77%
C < 77% to 73%
C- < 73% to 70%
D+ < 70% to 67%
D < 67% to 60%
E < 60% to 0%

SAMPLE

Assignments

Class Memos - 20%

There are eight memos listed in the syllabus. Each one is worth 10 points, and will comprise 20% of the final grade.

Midterm grade – 15%

A midterm exam will be given.

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.

Final Project – 35%

The final project will be a group exercise with a real-life data science problem. Depending on the number of students in the class, individuals will be asked to work on a specific project and present the results in a state or local government agency. For example, if the Ohio Department of Education is evaluating their accountability data to determine how to improve the quality of the measures used to assess “Student performance” – the students would review data (publically available) and analyze the 1) data collection system, 2) the metrics used to develop accountability measures; and 3) perform statistical analyses to understand trends in the performance of student academic outcomes.

The final report will be 1) an approximately 20 page written document that summarizes the analyses and the results, and 2) a formal presentation for the government agency. Students will divide up work under the supervision of the instructor and will be responsible for a series of mini-deliverables over the course of the term:

- Deliverable 1: Proposal and data needs document. (10%) (Week X)
- Deliverable 2: Draft Final Report (10%) (Week X)
- Deliverable 3: Final Report and Presentation (15%) (Week X)

Additional guidelines will be distributed for the project will be distributed at the start of the class term.

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.

Attendance Policy

XXXX

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.
Overview of the class

Class	Topic	Reading
Class 1	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. Harvard Business Review. Retrieved from https://hbr.org/2014/11/with-big-data-comes-big-responsibility .
Class 2	Data Science and Public Policy; Reducing uncertainty with data; Going	Mergel, I., Rethemeyer, R. K. and Isett, K. (2016), Big Data in Public Affairs.

	<p>from discovery to reform or innovation with data; Limitations of data systems</p>	<p>Public Admin Rev, 76: 928–937. doi:10.1111/puar.12625</p> <p>D. Weimer & A. Vining (2011), <i>Policy Analysis Chapter 11: Gathering Information for Policy Analysis</i></p> <p>Advanced</p> <p>D. Stone (2012) <i>Policy Paradox</i>, Chapter 5: Numbers</p>
--	--	--

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 Literacy in Decision Making

Class	Topic	Reading
Class 3	<p>Data typologies, structured and unstructured data; Sources of public data; Relationship to private sector; Data management systems; Levels of government and data analysis</p>	<p>Cukier, K. N. (2010, February 25). Data, data everywhere. <i>Economist</i> . Retrieved from http://www.economist.com/node/15557443.</p> <p>Llorente , Alejandro , Manuel Garcia-Herranz, Manuel Cebrian , and Esteban Moro. 2015. Social Media Fingerprints of Unemployment. <i>PLoS One</i> 10 (5)</p> <p>Kitchin , Rob , and Gavin McArdle . 2016 . What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets. <i>Big Data & Society</i>. http://bds.sagepub.com/content/3/1/2053951716631130</p>
Class 4	<p>Working with data in government, Common structure of administrative systems; Contractors and private vendors; Legacy technical systems; Contracting policy and effects on reform</p>	<p>Janssen , Marijn , and Jeroen van den Hoven. 2015. Big and Open Linked Data in Government: A Challenge to Transparency and Privacy? <i>Government Information Quarterly</i> 32 (4): 363 – 68</p> <p>Lavertu, Stéphane. 2015. We All Need Help: “Big Data” and the Mismeasure of Public Administration. <i>Public Administration Review</i>.</p>

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 and Ethics

Class	Topic	Reading
Class 5	Ethics in Big Data; Differences in ethics and data flows (e.g., Internet of things vs. PII); Major ethical limitations on data use in government;	<p>Boyd, d., & Crawford, K. (2012). Critical questions for big data. <i>Information, Communication and Society</i>, 15(5), 662–679.</p> <p>Zwitter, A. (2014, November). Big data ethics. <u>Big Data and Society</u></p> <p>*Foster et. al (2016), Chapter 11: Privacy and Confidentiality</p> <p>Paul G. Stiles and Roger A. Boothroyd (2015). Ethical Use of Administrative Data for Research Purposes; in John Fantuzzo and Dennis P. Culhane, eds., <u>Actionable Intelligence Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government</u> (Palgrave, NY)</p>
Class 6	Personal privacy and data utilization; discrimination; social media and internet	<p>Dannen, C. (2009, July 1). On Facebook? New algorithm can guess your SSN. <i>Fast Company</i>. Retrieved from <u>http://www.fastcompany.com/1305136/facebook-newalgorithm-can-guess-your-ssn</u></p> <p>NSA Domestic Surveillance Directorate. (n.d.). Your data: If you have nothing to hide, you have nothing to fear. Retrieved from <u>https://nsa.gov1.info/data</u>.</p> <p>McCallister, E., Grance, T., & Scarfone, K. (2010). Guide to protecting the confidentiality of personally identifiable information (PII): Recommendations of the National Institute of Standards and Technology (NIST Special Publication 800-122). Gaithersburg, MD: NIST.</p>

***Memo 3:** What can you find out about yourself online? Do you use google, facebook or twitter? Check out your you tube 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.

Week 4: Data and Governance

Class 7	Data governance; Managing data systems in local and state government; Collaborative Governance; Disaster planning for data releases	Erika Kitzmiller and TC Burnett (2015) The AISP Network: Three Organizational Models for Building, Using, and Sustaining Integrated Data Systems; in John Fantuzzo and Dennis P. Culhane, eds., <u>Actionable Intelligence Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government</u> (Palgrave, NY) SLDS Best Practice Brief, P-20W+ Data Governance (https://slds.grads360.org/#communities/pdc/documents/2717)
Class 8	Case Study: What happens when data get released?	Data Breach Scenario Activity (deals with unsecured passwords, and organizational responses)

Memo 4: Describe governance procedure of one data system at the state or local level. To answer this you will need to answer the following questions: 1) What is the coverage of the data system; 2) How are the data systems managed by the government and private organizations? 3) What is the purpose of the data systems?

Week 5: Data Transparency and Democracy

Class 9	Government uses of data; Voting systems and data; FOIA or State Open records requests; Whistle blowers and data systems;	Fung, A. (2015). Infotopia: Unleashing the power of transparency Schneier (2015), Data and Goliath, pp 1-103 How to Make a Open Records Request. http://www.dispatch.com/article/20130314/BLOGS/303149737 Forty-four states and DC have refused to give certain voter information to Trump commission
---------	--	---

		<p>http://www.cnn.com/2017/07/03/politics/kris-kobach-letter-voter-fraud-commission-information/index.html</p> <p>GOP Effort To Make Environmental Science 'Transparent' Worries Scientists https://www.npr.org/2017/07/20/537243392/gop-effort-to-make-environmental-science-transparent-worries-scientists</p>
Class 10	Case Study: Do data systems improve decisions and democratic access?	<p><u>Will Democracy Survive Big Data and Artificial Intelligence?</u> by Dirk Helbing; Bruno S Frey; Gerd Gigerenzer; Ernst Hafen; Michael Hagner; Yvonne Hofstetter; Jeroen van den Hoven; Roberto V Zicari; Andrej Zwitter, Scientific American, Feb 2017</p>

Memo 5: What do you think, can data systems improve decision making and democracy? Does it matter if data systems improve political decision making, or is it more important to improve efficiency?

Week 6: Laws and Data Utilization

Class 11	Federal laws governing data systems; State legal framework on data utilization; Acceptable uses of data by government; Acceptable Uses of data outside of government;	<p>Peralta, John (2015), "Legal Issues in the Use of Electronic Data Systems for Social Science Research," in John Fantuzzo and Dennis P. Culhane, eds., <u>Actionable Intelligence Using Integrated Data Systems to Achieve a More Effective, Efficient, and Ethical Government</u> (Palgrave, NY)</p>
Class 12	Case Study: FERPA, WIOA, and Performance Matches	<p>Review FERPA Language https://www2.ed.gov/policy/gen/reg/ferpa/index.html</p> <p>Integrated Data Systems and Student Privacy, https://studentprivacy.ed.gov/resources/integrated-data-systems-and-student-privacy</p> <p>Joint Guidance on Data Matching to Facilitate WIOA Performance Reporting and Evaluation https://studentprivacy.ed.gov/resources/joint-guidance-data-matching-facilitate-wioa-performance-reporting-and-evaluation</p>

Memo 6: Provide a one page summary of one of the following legal systems, and describe limitations on how the data can be used? (FERPA, HIPAA).

Week 7: Data, Performance, and Incentives

Class 13	Data and Performance Systems; Federal Program Standards; State Program Standards; Evidence Based Policy; Incentives and Standards; Content Standards in Education	<p>How's My Program Doing? The Question That Doesn't Always Have a Good Answer http://www.governing.com/columns/smart-mgmt/gov-program-evaluation-offices.html</p> <p>Did Performance Measurement Cause America's Police Problem?; http://www.governing.com/columns/smart-mgmt/gov-policing-performance-measurement.html</p> <p>Head Start Perf. Standards, https://www.acf.hhs.gov/ohs/policy</p> <p>Workforce Investment and Innovation Act Standards https://www.doleta.gov/performance/guidance/tools_commonmeasures.cfm</p>
Class 14	Case Study: Workforce Success Measures and Workforce Incentives	<p>Ohio Workforce Success Measures, https://workforcesuccess.chrr.ohio-state.edu/about</p> <p>(Heinrich, Mueser, Troske, Jeon, & Kahvecioglu, 2013) ; Do Public Employment and Training Programs Work</p>

Memo 7: How do performance standards improve the decision making for government? What are the best practices in performance standards?

Week 8: Data Standards

Class 15	Data Standards, Federal Standards; Comparable Data Systems; Interaction with Information Technology Systems	<p>(Additional readings to be determined)</p> <p>EPA Data Standards - https://www.epa.gov/data-standards/federal-national-and-international-data-standards</p>
Class 16	Case Study: CEDS (Common Education Data Standards) and EdFI	<p>Common Ed. Data Standards, https://ceds.ed.gov/</p> <p>CEDS and SLDS, https://ceds.ed.gov/pdf/ceds-and-slids-aligning-efforts.pdf</p>

		EdFI, https://techdocs.ed-fi.org/display/EFDS21/Ed-Fi+Data+Standard+v2.1
--	--	---

Week 9: Technology and Data Systems

Class 17	Databases (data management; data structures; data systems; filtering; predictive analytics)	<p>Kuhn and Johnson (2013) <u>Applied Predictive Analytics</u>, Springer, “Chapter 1: Introduction”</p> <p>Hadley Wickham and Garrett Grolemund. (2016). <i>R for Data Science: Import, Tidy, Transform, Visualize, and Model Data</i> 1st Edition. O’Reilly; “Chapter 8: Data Import with readr” and “Chapter 9: Tidy data with tidy”</p> <p><u>Zagorsky, J (2017) “Predictive Model Documentation for Student Success Dashboard” (CHRR, Columbus, OH)</u></p>
Class 18	Data engineering principles (distributed and cloud computing; ETL processes; basic querying across platforms)	<p>Kim, Trimi and Chung (2014) “Big Data Applications in the Government Sector” in <u>Communications of the ACM</u> (March, vol 57, 3, pp. 78-85</p> <p><u>Bahrani, M and Singal, M (2014), “The Role of Cloud Computing Architecture in Big Data” in Information Granularity, Big Data, and Computational Intelligence , pp. 275-295</u></p>

Week 10: Using Data to Produce Metrics, Dashboards, Scorecards

Class	Topic	Reading
Class 19	Data Uses, Subsetting Data; Creating Metrics	<p>Weiss, C. (1998). Improving the Use of Evaluations; Whose Job is it anyway. <i>Advances in Educational Productivity</i>, 7, 263-276.</p> <p>Jevin West (2014) How to improve the use of metrics: learn from game theory. <i>Nature</i> 465:871-872</p> <p>Hadley Wickham and Garrett Grolemund. (2016). <i>R for Data Science: Import, Tidy, Transform, Visualize, and Model Data</i> 1st Edition. O’Reilly; “Chapter 10: Relational Data with dplyr”</p>
Class 20	Reproducibility; Avoiding BS	Robert Matthews (2000) Storks deliver babies (p=0.008). <i>Teaching Statistics</i> 22:36-38

	Dashboards	<p>Food Stamp Fraud Case http://callingbullshit.org/case_studies/case_study_foodstamp_fraud.html</p> <p>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.</p> <p>Advanced Foster et al (2016), Chapter 10: Errors and Inference</p>
--	------------	---

Memo 8: 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 11: Data Visualization and Causality

Class	Topic	Reading
Class 21	Data Visualization	<p>Edward Tufte (1983) <i>The Visual Display of Quantitative Information</i> Chapters 2 (Graphical integrity) and 5 (Chartjunk: vibrations, grids, and ducks).</p> <p>Advanced Foster et al (2016), Chapter 9: Information Visualization</p>
Class 22	Correlation and Causation; Data relevance	<p>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)</p> <p>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</p> <p>Advanced Kuhn & Johnson (2013), Chapter 18: Measuring Predictor Importance</p>

Week 12: Prediction and Geography

Class	Topic	Reading
-------	-------	---------

Class 23	Prediction	<p>Edward Tufte (1974). Data Analysis for Politics and Policy (Chapter 2: Predictions and Projections)</p> <p>Martin O'Malley (2014). Doing What Works: Governing in 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>
Class 24	Geographical Analysis	<p>R-Intro to Spatial Data https://cran.r-project.org/doc/contrib/intro-spatial-rl.pdf</p> <p>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>

Week 13: Final Projects

Class 25	Presentations	
Class 26	Presentations	

SAMPLE