POLI3148 Data Science in Politics and Public Administration

Semester: 2023 Fall Credit: 6 Instructor: Dr. Chen Haohan

Big data play increasingly important roles in politics and public administration. This course aims to help future PPA practitioners and researchers make sense of big data. It guides students to ask and answer the following questions: What are big data? With big data, what questions can policymakers and researchers ask and answer? How do we collect and analyze big data? This course introduces students to state-of-the-art data science techniques. Topics include introductory programming for data science, data collection, data visualization, and machine learning. Upon completion, students are expected to get hands-on experience with data science and understand the promise, limitations, and pitfalls of big data in politics and public administration.

Instructor's Contact

- Email address: haohan@hku.hk
- Office address: Room 9.50, 9/F, The Jockey Club Tower
- Office hours: Tuesdays 14:00-16:00
  Please book meetings through Calendly*:* *https://calendly.com/haohanchen/office-hours*

Class Meetings

- Lecture + Lab: Thursdays 13:30 - 14:20 (CPD-3.41)
- DaSPPA Festival: Dec 7, 2023

Course Objectives

1. Introduce what big data and data science means and why they are relevant to politics and public administration.
2. Get hands-on experience with basic data science tools.
3. Critically evaluate the promise and limitations of data science in PPA.

Course Learning Outcomes

1. "Think like a data scientist" when asking and answering political and policy questions.
2. Master elementary skills for data collection, visualization, and analysis.
3. Effectively communicate data-driven political and policy analysis.

Schedule

Date	Week	Lecture (1st half)	Lecture (2nd half)	Due
09/07	2	Welcome	*R* you ready?	
09/14	3	*R* Basics (1)	*R* Basics (2)	
09/21	4	*R* Basics (3)	Data Wrangling (1)	
09/28	5	Data Wrangling (2)	Machine Learning Overview	
10/05	6	Data Visualization (1)	Linear Regression	
10/12	7	Data Visualization (2)	Classification	
10/19	8	*Reading week. No class.*		
10/26	9	Data Visualization (3)	Resampling Methods	A1
11/02	10	Data Visualization (4)	Model Selection and Regularization	B1
11/09	11	Text Mining (1)	Tree-Based Methods	
11/16	12	Text Mining (2)	Unsupervised Learning	
11/23	13	Text Mining (3)	Text Mining (4)	A2
11/30	14	Putting Everything Together	Debriefing and Q&A	
12/07	R	*DaSPPA Festival!*		G1
12/14	A	*Group Final Project Replication Dossier Due*		G2
12/19	A	*Personal DaSPPA Portfolio Submission Due*		A3

Output and Assessment

Type	Output	Weight	Due
Participation (30%)	[P1] Lecture Participation	15%	-
	[P2] In-Class Exercises	15%	-
Individual Assignments (40%)	[A1] Problem Set 1	15%	10/28
	[A2] Problem Set 2	15%	11/25
	[A3] Personal DaSPPA Portfolio	10%	12/19
Group Final Project (30%)	[G1] Poster Presentation at DaSPPA Festival	20%	12/07
	[G2] Replication Dossier	10%	12/14
Bonus (10% max.)	[B1] Research Proposal	5%	11/04
	[B2] Others	5%	-

Primary References

R Basics

[Douglas et al.] Douglas, Alex, Deon Roos, , Francesca Mancini, Ana Couto & David Lusseau. *An Introduction to R*. E-book: intro2r.com

[R4DS] Wickham, Hadley, Mine Cetinkaya-Rundel, & Garrett Grolemund. *R for Data Science (2e)*. E-book: r4ds.hadley.nz

Data Visualization

[Kabacoff] Kabacoff, Rob. *Data Visualization with R*. 2020. E-book: rkabacoff.github.io/datavis

[Healy] Healy, Kieran. *Data visualization: a practical introduction*. Princeton University Press, 2018. E-book: socviz.co

Machine Learning

[ISL] James, Gareth, Daniela Witten, Trevor Hastie, & Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R (2nd ed.)*. 2021. E-book: statlearning.com (click "Download ISL with R")

[HOML] Boehmke, Bradley, Brandon Greenwell. Hands-On Machine Learning with R. E-book: bradleyboehmke.github.io/HOML

Text Mining

[TMR] Silge, Julia, and David Robinson. *Text Mining with R*. 2021. E-book: tidytextmining.com

Reading Materials by Week

(09/07) Week 1: Welcome | *R* you ready?

No required reading

(09/14) Week 2: R Basics (1) (2)

Douglas et al. Chapters 1-3 (but **no need to read** 3.4)

(09/21) Week 3: R Basics (3) | Data Wrangling (1)

R Basics (3)

Douglas et al. Chapters 8, 9

Data Wrangling (1)

R4DS. Chapter 4. Data Transformation

(09/28) Week 4: Data Wrangling (2) | Machine Learning Overview

Data Wrangling (2)

R4DS. Chapter 4: Data Tidying

Machine Learning Overview

No required reading.

(10/05) Week 5: Data Visualization (1) | Linear Regression

Data Visualization (1)

Healy. Chapter 1: Look at data.

Kabacoff. Chapter 2: Introduction to ggplot2; Chapter 3: Univariate Graphs

Linear Regression

ISL. Chapter 3: Linear Regression

(10/12) Week 6: Data Visualization (2) | Classification

Data Visualization (2)

Kabacoff. Chapter 4: Bivariate Graphs; Chapter 5: Multivariate Graphs

Classification

ISL. Chapter 4: Classification
HOML. Chapter 5: Logistic Regression; Chapter 8: K-Nearest Neighbours

(10/19) Week 7: *Reading Week. No class.*

(10/26) Week 8: Data Visualization (3) | Resampling Methods

Data Visualization (3)

Kabacoff. Chapters 6: Maps; Chapter 7: Time-dependent Graphs

Resampling Methods

ISL. Chapter 5: Resampling Methods

(11/02) Week 9: Data Visualization (4) | Model Selection and Regularization

Data Visualization (4)

Kabacoff. Chapters 9: Other Graphs (focus on 9.4-9.8); Chapter 10: Customizing Graphs

Linear Model Selection and Regularization

ISL. Chapter 6: Linear Model Selection and Regularization
HOML. Chapter 6: Regularized Regression

(11/09) Week 10: Text Mining (1) | Tree-Based Methods

Text as Data (1)

TMR. Chapter 1: The tidy text format; Chapter 3: Analyzing word and document frequency.

Tree-based Methods

ISL. Chapter 8: Tree-Based Methods.
HOML. Chapter 9 Decision Trees; Chapter 10. Bagging; Chapter 11: Random Forest; Chapter 12 Gradient Boosting.

(11/16) Week 11: Text Mining (2) | Unsupervised Learning

Text as Data (2)

TMR. Chapter 2: Sentiment analysis with tidy data.

Unsupervised Learning

ISL. Chapter 12: Unsupervised Learning
HOML. Chapter 17: Principal Component Analysis; Chapter 20: K-means Clustering

(11/23) Week 12: Text Mining (3) (4)

TMR. Chapter 6: Topic Modeling

(11/30) Week 13: Putting Everything Together | Debriefing and Q&A

No required reading.

Output and Assessment Elaborated

Participation (30%)

Lecture participation (15%) recognizes attendance and active participation in discussions during lectures. You are encouraged to share your views on the reading materials and lectures. Grades will be awarded based on the frequency and the quality of your contribution.

In-class exercises (15%) are typically coding exercises that you complete during lectures under the instructor's guidance. You will have opportunities to polish your in-class exercises after class and submit your outputs. Grades will be awarded based on the quality of your submitted outputs.

Individual Assignments (40%)

Problem Sets (15% + 15%)

You will complete two Problem Sets, where you will practice your data science skills. The two assignments are individual tasks. You can discuss general concepts and techniques used in the assignment with your classmates. But your submission should be your own work.

Personal DaSPPA Portfolio (10%)

At the end of the semester, you will organize all your DaSPPA coursework, including notes and code you write for in-class exercises, problem sets, and the group project, into a folder/ document/ website and share it with the instructor. This portfolio is meant to comprehensively showcase your learning outcomes in this course, which you can re-use as part of your application portfolio for data science-related career opportunities in the future.

Group Research Project (30%)

In groups of **2-4**, you will design an original research project that uses data science to answer a question in politics and public administration of your choice. You are encouraged to use your creativity to decide the context and the methods of your study. A successful research project should (1) ask an interesting research question that is relevant to the substantive literature and/ or real-world events and (2) correctly apply the methods we learn in this course. The project will render two outputs: a poster presentation and a replication code dossier. See below for their requirements.

Poster Presentation at DaSPPA Festival (20%)

You will make a poster presentation of your final project at the DaSPPA Festival during the revision week (see details on the front page). The presentation should clearly communicate the motivation, research question, research design, findings, and conclusions of your project. **In-person participation in the poster presentation is mandatory**.

The poster may include the following sections:

- **Background:** Discuss why the topic is interesting and important with reference to the literature and/ or real-world events.
- **Research Question:** Discuss your puzzles with reference to the literature and/ or real-world events.

- **Data and Methods:** Discuss the data and methods you employ to answer your question.
- **Findings:** Describe and discuss the results you have obtained.
- **Conclusion:** Highlight the main takeaway from your project. You may also point out the limitations of your data and methods and suggest what can be done in the future to improve on them.
- (As footnotes) **Division of Labor**: Describe how the work has been allocated among group members

Replication Dossier (10%)

All the results in your poster presentation must be reproducible. In this task, you will organize your data, documented code, figures, tables, and the poster into a folder. Along with the above materials, you will provide a short *README* document that contains basic project information and an instruction on how to run your code to replicate the results.

Note: If you wish to do the "Group Final Project" as an individual project (that is, doing the project by yourself), you should discuss it with the instructor before the Reading Week. Permission may be granted on a case-by-case basis.

Bonus (10% max.)

Research Proposal (5% max.)

You may submit a 1-to-2-page research proposal to get feedback from the instructor. The proposal can include the following sections.

- Background: Briefly discuss why the topic is interesting and important with reference to the literature and your case of interest.
- Research Question: Discuss what puzzles your project tackles with reference to the literature and/or real-world political events.
- Data and Methods: Discuss the data and empirical method(s) you will employ to answer your question.
- Division of Labor: Explain how your group members have agreed to divide the work.

Others (5% max.)

Other bonus points may be awarded to recognize exceptional performance in participation, individual assignments, and the group final project.

Course Policy

Attendance and participation

Attendance is required. If you have medical reasons for your absence, please provide documented evidence (such as a certificate signed by a registered medical practitioner).

Late submission

One grade level will be deducted for one day of late submission. For example, A will be downgraded to A-, A- will be downgraded to B+ for one day of late submission. Works late for 5 days or above will be graded as F. If you have medical reasons for late submission, please provide documented evidence (such as a certificate signed by a registered medical practitioner). In principle, the instructor cannot waive the penalty for late submission due to non-medical reasons. **But you have 2 "free" late days (no question asked) across your individual coursework.**

Feedback

Department policy on provision of feedback for coursework assignments and written examinations: For written examinations: teachers should provide group feedback on overall student performance via Moodle or email within two weeks of publication of the final assessment results. Individual feedback on student performance may be provided at teachers' discretion. For coursework assignments: teachers should provide feedback to students within 30 days of the official date of submission or by the last day of teaching in each semester, whichever the teacher deems appropriate. In exceptional circumstances where this is not possible, teachers should : (i) notify students in advance of the expected date of feedback provision, and where possible, (ii) provide interim or generic feedback such as common errors and potential areas for improvement.

Academic Misconduct Warning

The Department of Politics & Public Administration expects that all students' work will conform to the highest standards of academic integrity. Student's work will be scrutinized for academic misconduct, which includes plagiarism of other's words and/or ideas, falsification, fabrication, and misuse of data.

Student's submitted work will be scrutinized for plagiarism through use of Turnitin via Moodle. In the event a student submits work that appears to be plagiarized—whether essays, presentations, or other course material—they will be asked to explain themselves to the lecturer. The Department does not tolerate plagiarism, whether direct, indirect, or self-plagiarism. Direct plagiarism is intentionally and completely lifting the words, equations, charts, graphs or artistic material of another author or authors. Indirect plagiarism is failing to cite completely or accurately, and/or copying themes, ideas, or sources the student has not read from another author or authors. Self-plagiarism is recycling papers, documents, equations, and so forth from a document previously submitted by the student without quotation, citation, or attribution of the previous work. Acts of plagiarism could result in heavy penalties, including disciplinary action. For more information about the policy on plagiarism at HKU, please visit http://www.hku.hk/plagiarism.

Appendix I. Programme Learning Outcomes

Bachelor of Social Sciences (Government and Laws) and LLB Curriculum

- *PLO1. Demonstrate a solid understanding of the key concepts in political science, political theory, public administration, international relations, legal analysis, legal theory, legal systems, and international law;*
- *PLO2. Possess effective skills in social scientific and legal research as applied to the study of public affairs and legal issues among different political, social, and cultural environments across nations;*
- *PLO3. Demonstrate a solid understanding of the body of legal knowledge and the capacity to critically analyse and evaluate legal principles at a level required to meet the standards and expectations of the legal profession and the community-at-large;*
- *PLO4. Show awareness of social issues and conditions, and utilize knowledge derived from political and legal studies and rhetorical advocacy techniques for the betterment of society;*
- *PLO5. Develop the skills and appreciation for teamwork through participating in group activities and internships;*
- *PLO6. Apply the knowledge, lawyering skills and legal reasoning to real situations in life, with a view to resolving issues, problems and disputes within the legal parameters.*

Bachelor of Social Sciences (Government and Laws) Curriculum (opt-out 4-year)

- *PLO1. Demonstrate a solid understanding of the key concepts in political science, political theory, public administration, international relations, legal analysis, legal theory, legal systems, and international law;*
- *PLO2. Possess effective skills in social scientific and legal research as applied to the study of public affairs and legal issues among different political, social, and cultural environments across nations;*
- *PLO3. Demonstrate a solid understanding of the body of legal knowledge and the capacity to critically analyse and evaluate legal principles at a level required to meet the standards and expectations of the legal profession and the community-at-large;*
- *PLO4. Show awareness of social issues and conditions, and utilize knowledge derived from political and legal studies and rhetorical advocacy techniques for the betterment of society;*
- *PLO5. Develop the skills and appreciation for teamwork through participating in group activities and internships.*

Bachelor of Social Sciences – Major/minor in PPA

- *PLO1. Demonstrate a solid understanding of the key concepts in political science, political theory, public administration, government operations, international relations and legal systems;*
- *PLO2. Understand critically and apply sensitively political science theories and research methods to analyze political complexities across national and international contexts;*
- *PLO3. Reflect critically on learning of political science and public administration theories in practice to meet the standards and expectations of the public sphere profession and the community-at-large;*

- *PLO4. Acquire specific knowledge and sensitive awareness of the formal, social and cultural aspects of political institutions and behavior;*
- *PLO5. Develop the skills and appreciation for teamwork through participating in group activities and internships;*
- *PLO6. Develop a comprehensive understanding of the complexities and intricacies of human society, and the ability to identify leverage points for policy and social actions.*

Appendix II: Course Grading Rubric

This appendix shows the standardized grading rubric from the Department of Politics and Public Administration, The University of Hong Kong.

Grade/Competency	A+, A, A-	B+, B, B-	C+, C, C-	D+, D, D-	F
Use of vocabulary and concepts	Student accurately and creatively uses concepts and key course vocabulary throughout the assignment, demonstrating a sophisticated understanding of each.	Student accurately uses concepts and key course vocabulary throughout the assignment, but does not demonstrate creativity in use or fluency.	Student uses concepts and key vocabulary from the course, but in a manner that does not demonstrate understanding or proficiency; use of concepts and vocabulary is perfunctory.	Student rehearses concepts or key course vocabulary but not in a way suggesting understanding at a university level.	Student fails to use concepts or key vocabulary correctly or at all.
Deployment of theories and argumentation	Student deploys theoretical arguments well using their own voice and substantive arguments in a sophisticated way.	Student deploys theoretical arguments well although voice, style and substantive critiques are similar to the source.	Student deploys theoretical vocabulary in a way commensurate with rules for argumentation, but does not show creativity or sophistication in substance or style.	Student rehearses theories and bits of argumentation from others and not in a way suggesting understanding at a university level.	Student fails to attempt argumentation or use of theoretical tools from the course.
Creativity	Students choice of topic, sources, assignment completion modality, arguments, and solutions show sophistication and critical thinking at a high level.	Students choice of topic, sources, assignment completion modality, arguments, and solutions show critical thinking skills.	Students choice of topic, sources, assignment completion modality, arguments, and solutions are average and "modal".	Student shows no more creativity than what is required to complete the task.	Student misunderstands creativity or fails to complete the assigned task.
Persuasiveness	Student makes an argument using appropriate language and rhetorical style necessary to persuade the reader to accept or accommodate their viewpoint.	Student makes an argument using appropriate language and rhetorical style necessary to complete the assignment.	Student makes an argument using either inappropriate language and/or rhetorical style. Ranting or editorializing.	Student rants or editorializes considerably, but stays largely on message.	Student rants incoherently.

(continued)

Grade/Competency	A+, A, A-	B+, B, B-	C+, C, C-	D+, D, D-	F
Use of fact and empirical evidence	Student brings factual evidence to bear upon the arguments and supports factual claims with adequate support from reputable sources.	Student brings factual evidence to bear upon some arguments and supports factual claims with support from limited or questionable sources.	Student brings some facts into their arguments but fails to provide support consistently for factual claims and uses trite or prohibited sources as support (e.g., Wikipedia).	Students factual claims are questionable or unsupported. Student rehearses facts from unacceptable sources (e.g., Yahoo answers).	Factual claims, if any are incorrect, ill supported, or incoherent within the argument.
Grammar and spelling	Students writing is grammatically correct and there are no spelling errors.	Students writing is grammatically correct in most instances and there are few spelling errors.	Students writing is grammatically correct in many instances but spelling errors are found throughout the document, consonant with ESL students.	Grammatical infelicities and spelling errors appear frequently in the document, but these are errors common to ESL students.	Grammar and spelling are unacceptable for university level writing for any student.
Mechanics and style	Students writing is fluid, fluent, and in an appropriate style for the task.	Students writing is fluent but stilted and/ or is an odd style for the task.	Students writing is halting and imbalanced and may be inappropriate for the task.	Students writing is only marginally acceptable for university level courses.	Students writing needs significant remediation by outside sources.
Citations	Student accurately and completely cites all sources, whether factual, argumentative, or theoretical claims according to the appropriate citation scheme.	Student accurately cites all sources, whether factual, argumentative, or theoretical claims but does not provide complete citations or uses an inappropriate citation scheme.	Student cites most expected sources, but does not provide accurate or complete citations.	Students citations are incomplete and inconsistent throughout the paper.	Student fails to cite at all.
Sources	Students choice of sources demonstrate sophisticated use of research resources. Sources are from reputable, academic sources.	Students choice of sources indicate a notable level of use of research resources. Sources are from reputable, academic sources.	Students choice of sources show minimal use of research resources. Sources are from a mix of academic and non-academic sources, some of questionable provenance.	Students include only minimal outside sources from sophomoric or prohibited sources (e.g., Encyclopedia Britannica online).	Student fails to demonstrate appropriate outside research.