PPHA 35577: Big Data and Development

Winter 2025

Professor Austin Wright (austinlw@uchicago.edu)
Administrative TA Vitor Silveira (vitorsilveira@uchicago.edu)
Coding TA Akbar Saputra (akbarsaputra@uchicago.edu)
Section Times 9:30-10:50 AM Tuesday/Thursdays - Keller 0023
Professor Office Hours 8:00-9:20 AM Thursdays - Keller 2005

Description: *Big Data and Development* is a seminar course focused on the use of innovative data capture and analysis techniques to investigate topics related to economic and political development. Microlevel data is increasingly used to target and evaluate development interventions. In this course, students will engage with cutting-edge theoretical and quantitative research, drawing on readings in economics, political science, and data science. The course is organized around a set of core topics, including political and economic development, community-driven aid interventions, crime and corruption, and climate change. Course assessments will include two short research briefs, an econometric exercise in R, a partly in-class policy exercise, and a final research design.

Lectures: This course consists of live lectures of 1 hour and 20 minutes 2 times per week. All lectures will be in person. Tuesday's lectures will focus on a core and supplementary paper. Core readings are expected before class and supplementary papers are highly recommended. Thursday lectures will cover methods, a policy exercise, and eventually project presentations. A rough weekly schedule is visible in the course calendar. If a student requires remote accommodations, only Student Disability Services (not your instructor) can approve students to take their courses on a fully remote basis. Please contact Dean of Students Kate Biddle (kbiddle@uchicago.edu) for more information.

Assignments and Grading: The final grade for the course will be a function of participation, two exercises, and a final research design paired with an in-class presentation. The exercises (one econometric and the other policy focused) will resemble problem sets. Both exercises will be paired with in-class discussions and lectures. A participation grade will include attendance, participation in discussion, and check-in with Professor Wright on project status.

The research design will be approximately 4000 words (approximately 10 pages), including a statement of the research question and theoretical argument, a brief review of the relevant literature, and a research design, which includes an identification strategy, known and prospective data (including innovative measurement and collection protocols), and preliminary results to assess the

feasibility of the identification strategy. The projects may be done individually or in groups of two or three, but please note that group projects are expected to be more expansive (outlining more robustness checks and with a greater focus on results). Note that the data collection and cleaning process is usually the most time consuming part of any research project and being unable to produce full results before the end of the quarter is understandable, but it is expected that you think through this process early. Feel free to check in with Professor Wright and/or the TAs on any related concerns. The most compelling proposals from the last cycle have turned into larger projects, research assistant/associate opportunities, and post-graduation employment. Take your ideas seriously!

Grade Breakdown: The grade for the course will be broken down as follows:

- Participation: **15%**
- Exercises (2): **30%**
- Project Presentation: **15%**
- Final Research Design: **40%**

Late assignments/requests for extensions: Late assignments will be accepted within two business days with a 25% penalty; other assignments submitted later than two business days will not be accepted. Plan your commitments accordingly.

Stata and R Support: Harris offers free tutoring support to students in need of one-on-one help with their core courses as well as coding in Stata, R, and Python. Tutoring opens on Monday of Week 3 each quarter and students can utilize up to 10 hours total of tutoring per quarter. If you would like to learn more about the tutoring program or book an appointment here: `https://canvas.uchicago.edu/courses/36319/pages/core-tutoring-program?module_item_id=1415052`.

Recommended Textbooks: These textbooks are a recommended resource for understanding the identification strategies used in various papers we will study. You likely have these manuscripts from prior courses at Harris. Other course readings, made available via Canvas, will serve as the primary texts.

- *Mostly Harmless Econometrics* by Joshua D. Angrist and Jorn-Steffen Pischke
- *Introductory Econometrics: A Modern Approach* (7th Ed.) by Jeffrey M. Wooldridge

Ethical Academic Conduct: The University's Academic Policies and Procedures and guidance regarding Civil Conduct apply to all activities in our course. If you need to review the University's policies, please see: `https://studentmanual.uchicago.edu/Policies#Honesty` and

https://studentmanual.uchicago.edu/university. By taking this course, you explicitly pledge your honor that you will not cheat (or help others to cheat) in any way on the assignments. I adhere to the official Harris School protocol for ethical violations: *Harris Procedures for Allegations of Plagiarism, Cheating, and Academic Dishonesty.*

—First Violation

If a student is accused by an instructor or teaching assistant of plagiarism, cheating, or any other form of academic dishonesty, the student will be summoned to meet with the Dean of Students and the instructor. In the meeting, the student and instructor both present information about the situation. If it is determined by the instructor and the Dean of Students that the student has, in fact, plagiarized or cheated, the following sanctions will be imposed for the first violation:

- The student will generally receive a grade of 0 on the assignment or exam in question (subject to the discretion of the instructor). They may be penalized in other ways, up to and including failing the class.
- The student may be asked to re-do the assignment or retake the exam (without credit) to ensure that the student has learned how to properly cite sources or demonstrate that they have command of the material covered.
- A formal letter of finding is sent to the student stating that the student has been found in violation of the code of academic honesty and what the sanctions were. The letter, along with any evidence presented, is archived in Harris Student Affairs records until the student graduates if the student has no other violations.

—Second Violation

If a student who has already been found in violation academic dishonesty is again accused of academic dishonesty, the case will be sent to the Harris Area Disciplinary Committee. Details about the Area Disciplinary Committee procedures can be found in the University Student Manual (https://studentmanual.uchicago.edu/area). If the student is found in violation of academic honesty a second time, the Area Disciplinary Committee can assign sanctions including suspension or expulsion from the University.

Copyrights and Course Content (Use of Course Hero and similar websites): This course is a work of original authorship. All course materials (including, but not limited to, class lectures and discussions, handouts, examinations, study guides, and web materials) and the intellectual content of the course itself are protected by United States Federal Copyright Law. Students are permitted to make notes solely for their own private educational use. Students and all other persons are expressly forbidden from recording lectures or discussions and from distributing or selling lectures notes and all other course materials without the prior written permission of the instructors. Because the instructors own the copyright to the classroom presentations and all course materials, any notes

taken during those presentations and subsequently sold or distributed to others would constitute an unauthorized derivative work and expose the person or persons involved to individual copyright infringement actions by the instructors.

Course Calendar: The following calendar is meant as a **rough guide**. We will do our best to keep the assignments/project dates unchanged. In terms of lecture material, this is the order of the material, but we expect some content to take longer than one lecture, so the dates may change. Additional readings will be posted on Canvas.

1. **Week of** January 6

   Topic: Using Big Data to study political and economic development

   —**For Thursday:** A review of the research process, coming up with project ideas and incorporating big data

2. **Week of** January 13

   Topic: Big data and climate change: political economy of the environment

   Core paper: Income Opportunities and Sea Piracy in Indonesia: Evidence from Satellite Data (`https://www.ne.su.se/polopoly_fs/1.265182.1452856653!/menu/standard/file/jobmarket_paper.pdf`)

   Supplemental paper: No supplemental paper this week

   —**For Thursday:** Metrics Exercise Part 1: Brainstorming ideas for the exercise

3. **Week of** January 20

   Topic: Big data and the environment (Part II)

   Core paper: Bureaucrat Incentives Reduce Crop Burning and Child Mortality in South Asia (`https://www.nature.com/articles/s41586-024-08046-z`)

   Supplemental paper: No supplemental paper this week

   —**For Thursday:** Metrics Exercise: Following the guidelines, write a 2-page memo regarding your proposal

4. **Week of** January 27

   Topic: Big data and politics: fighting corruption

Core paper: Electoral Accountability and Corruption: Evidence from the Audits of Local Governments (https://eml.berkeley.edu/~ffinan/Finan_Termlimits.pdf)

Supplemental paper: No supplemental paper this week

—**For Thursday:** Democracy under Threat: Electoral Reform and Political Violence (PDF posted on Canvas)

5. Week of February 3

Topic: Big data and economics: targeting aid

Core Paper: Machine learning and phone data can improve targeting of humanitarian aid (https://www.nature.com/articles/s41586-022-04484-9)

Supplemental Paper: Aid under Fire: Development Projects and Civil Conflict (https://www.jstor.org/stable/42920868)

—**For Thursday:** Start of the Policy Exercise

6. Week of February 10

Note: Classes might be remote for this week

Topic: Big data and politics: political engagement

Core paper: Buying Informed Voters: New Effects of Information on Voters and Candidates (https://academic.oup.com/ej/article/131/635/1105/5904046)

Supplemental paper for Thursday: Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials (https://guygrossman.com/assets/pdf/2019_Science_PE_voter_information_campaigns.pdf)

—**For Thursday:** Continuation of the Policy Exercise

7. Week of February 17

Topic: Big data and economics: economic growth

Core paper: How Much Should We Trust the Dictator's GDP Growth Estimates? (https://www.journals.uchicago.edu/doi/10.1086/720458)

Supplemental paper: Behavioral Foundations of Microcredit: Experimental and Survey Evidence from Rural India (https://www.aeaweb.org/articles?id=10.1257/aer.102.2.1118)

—**For Thursday:** Student Presentations (part I)

8. Week of February 24

Topic: Big data and conflict: population displacement

Core paper: Public Response to Government Alerts Saves Lives during Russian Invasion of Ukraine (`https://bfi.uchicago.edu/insight/finding/civil-response-to-government-alerts-declin`

Supplemental paper: Public response to government alerts saves lives during Russian invasion of Ukraine (`https://www.pnas.org/doi/10.1073/pnas.2220160120l`)

—**For Thursday:** Student Presentations (part II)

9. Week of March 3

Topic: Big data and crime

Core paper: Gang Rule: Understanding and Countering Criminal Governance (`https://www.nber.org/system/files/working_papers/w28458/w28458.pdf`)

Supplemental paper: No supplemental paper this week

—**For Thursday:** Turning research into insights, insights into impact, impact into career opportunities.

10. Week of March 10

Final research design submission. Date TBD.