Data & Policy Summer School Program
Required Modules Sample Syllabus (2024)

Professors: Austin Wright (Program Lead and Academic Track Lead) and Jeff Levy (Professional Track Lead)
Head TAs: Jose Macias and Alex Sobczynski
Teaching Assistants: TBD

Course Description

This course is an introduction to research design and data analytics. We will focus on establishing a foundation for studying public policy problems through data. This course consists of two complementary sections – Data Analytics and Programming in R. These two sections compare to one full course at the University of Chicago Harris School of Public Policy.

Data Analytics [DA] provides an introduction to the statistical foundations, tools, and methods employed by public policy researchers. In this section we will explore the fundamental problem of causal inference, and learn how to use data, research design, and statistical modeling to navigate around this problem. The suite of tools taught for DA shall depend on your track.

Programming in R [R] introduces students to the powerful statistical modeling software known as R. This is a free and open source software that is constantly being expanded and upgraded. Knowledge of programming in this language is in high demand in policy job markets. In this section of the course, you will be introduced to many of the key uses of the software — including data wrangling, applications of causal inference methods, data visualizations, and more. The suite of tools taught for R shall depend on your track.

The Capstone is an opportunity to apply the coursework of Data Analytics and R Programming in a research project with open access data. Datasets have been selected in three areas and students may choose one based on interest. One or two of these datasets shall serve as the basis for a couple of assignments in the Programming in R assignments. You will work with a small group of peers ("capstone group") to collaborate and share ideas, but you will each submit an individual project. Each student must submit a project that reflects independent work. The final product is something you will be able to showcase in your resume, LinkedIn, job, or graduate school application. Professional track students will provide a *Policy Memo* while academic track students will deliver *Research Notes*. Capstone topics may include:

1. Economic Policy - Labor Market shifts due to automation in the early 20th century among telephone operators also known as "[Hello Girls](#)"

2. Urban and Social Policy - Chicago Array of Things; Analyzing the environmental, air quality, and computer vision data collected from sensors at Chicago bus stops.
3. International Development Policy - Identifying social and economic change at the household level in Ghana with survey data over nearly a decade from the Yale Economic Growth Center.

Virtual Format

Academic lectures for Data Analytics and R Programming are delivered via weekly video modules (pre-recorded lectures). These modules will provide students with the knowledge and materials to understand the basics of data analytics as applied to public policy. During the Capstone Weeks, faculty will host synchronous lectures for each research project. You are strongly encouraged to attend the synchronous lecture. Live office hours are held multiple times throughout each week, enabling students to connect with faculty and graduate teaching assistants. The schedule and how to access the virtual office hours will be available on the Canvas calendar.

Course Materials

All necessary course materials will be made accessible via Canvas. There are no required textbooks. If you would like a supplementary text that covers many of the same topics, then we recommend by category the following:

Data Analytics:

- *Thinking Clearly With Data: A Guide to Quantitative Reasoning and Analysis* by Bueno De Mesquita and Fowler.
- *Mastering 'Metrics: The Path from Cause to Effect* by Angrist and Pischke is recommended for Quantitative Analysis.
- *Mostly Harmless Econometrics* by Angrist and Pischke.

R Programming:

- *R for Data Science (R4DS)* by Hadley Wickham and Garret Grolemund
  - This online textbook is free, and a new version has been posted as well.

Assignments and Grading

The graded assignments for this course will consist of homework assignments. Each homework assignment will be released via Canvas, and you will submit your homework in Canvas by the required deadline. The assignments will be due on Mondays following Weeks 1, 2, 4 and 5 by 12:00 pm Central Daylight Time, and include separate evaluations for both R Programming and Data Analytics.

The concept quizzes will be available after each video lecture, and the questions will relate directly to content from the preceding lecture. These concept quizzes will be an excellent way to gauge your knowledge of the material covered in lecture, and they will prepare you for questions on your homework assignments. Concept quizzes are optional and not graded.

Communication

Communication from instructors and TAs to students will happen through posting of materials on Canvas, including postings to Announcements and on Ed Discussion.

The most effective way for the teaching team to support students is for student to submit questions in the following forms:

- Questions regarding course material should be asked in the live Office Hours or posted on Ed Discussion, a forum that is monitored by the teaching assistants and instructors.
- Questions regarding scheduling should be directed to the Head TA.

Ed Discussion Board: This course will also have a dedicated Ed Discussion board. Ed Discussion is a free online gathering place where students can engage with each other regarding course material, under the guidance of their instructors. Setting up an ed discussion account will be part of your first assignment, but participation is optional. On Ed Discussion, you can both ask and answer questions related to homework assignments and course materials. If you provide a strong answer, one of the teaching assistants might endorse your answer.

R Resources

In HW1, you will download R and RStudio. Please use the following resources to do so.

Download R: here.

Download R Studio: here

Guide for Downloading R: here

Schedule

Week	Academic Track		Professional Track		Assignments
	Data Analytics	**R Programming**	**Data Analytics**	**R Programming**	
Week 1	1.1 - Foundations of Causal Inference for Public Policy	2.1 - Intro to R and RStudio (working dirs, projects, panes, R basics, etc)	1.1 - Foundations of Causal Inference for Public Policy	2.1 - Intro to R and RStudio (working dirs, projects, panes, R basics, etc)	HW1 - 10 pts (6/17)
	1.2 - Fundamentals of Working with Data	2.2 - Intro to tidyverse, fundamentals of data, basic visualization	1.2 - Fundamentals of Working with Data	2.2 - Intro to tidyverse, fundamentals of data, basic visualization	
Week 2	1.3 - Difference in means: RCTs (experimental ideal)	2.3 - Tidy data, data wrangling, and simple data cleaning	1.3 - Difference in means: RCTs (experimental ideal)	2.3 - Tidy data, data wrangling, and simple data cleaning	HW2 - 30 pts (6/24)
	1.4 - Bivariate regression: properties, testing, interpretation	2.4 - Data visualization and exploration (ggplot2, summarization)	1.4 - Bivariate regression: properties, testing, interpretation	2.4 - Data visualization and exploration (ggplot2, summarization)	
Week 3	1.5 - Multivariate regression: testing, interpretation, omitted variable bias	2.5 - APIs and policy applications (working with Census data)	1.5 - Multivariate regression: testing, interpretation, omitted variable bias	2.5 - APIs and policy applications (working with Census data)	

		2.6 - Programming concepts (for loops, functions, control flow)		2.6 - Programming concepts (for loops, functions, control flow)	
Week 4	1.6 - Functional Forms, Interactions, and Indicator Variables	2.7 - Descriptive stats in R (sample, distribution)	Catch up Week and Review Existing Materials	2.7 - Descriptive Statistics, Visuals, and mapping introduction in R (distributions, ggplot2)	HW3 - 30 pts (7/8)
	1.7 - Binary Dependent Variables	2.8 - Presenting a model; Causal inference in R (lm, stargazer)			
	1.8 - Differences in Differences				
Week 5	1.9 - Panel data designs: fixed effects, first differences	2.9 - Literate programming (RMarkdown, code syntax)	-	2.8 - More Advanced Data Visualization and Mapping 2.9 - QGIS Introduction (Required for Professional Track, Optional for Academic Track)	HW4 - 30 pts (7/15)
	1.10 - Regression discontinuity designs	2.10 - Presenting Models Graphically (RD, DiD)			
	1.11- Instrumental variables				
Week 6	**Capstone Project (Friday, 7/26)**				
Week 7					

Credential Award

In order to pass this course, you must successfully pass at least two of the three academic components: Data Analytics, R Programming, and Capstone research project. The passing score in each of the three components is 60%.

Upon successful completion of the coursework, you will be issued a Certification of Credential Completion by the University of Chicago Harris School of Public Policy. The Certification of Credential Completion contains a point grade for each academic component: Data Analytics, R Programming, and Capstone research project. You will receive an email from our office (harriscredential@uchicago.edu) with the electronic certification approximately four weeks after the program.

You will additionally receive an official UChicago Transcript with pass/fail marks (non-credit) for each academic component: Data Analytics, R Programming, and Capstone Research project. You will be able to access this transcript at the end of the summer (approximately early October). This is a non-credit program.

Sample Certification of Credential Completion

For your reference, we have enclosed a sample of a past Certification of Credential Completion, which contains a point grade for each academic component: Data Analytics, R Programming, and Capstone research project.

THE UNIVERSITY OF
CHICAGO

THE UNIVERSITY OF CHICAGO
HARRIS SCHOOL
OF PUBLIC POLICY
Social Impact, Down to a Science

The University of Chicago Harris School of Public Policy

certifies

Full Name

has successfully completed the non-degree, non-credit

Data and Policy Summer Scholar Program

an immersive seven-week experience with training in Data Analytics in Public Policy, Introduction to R Programming, a capstone research project with faculty, and rich connections with the UChicago community

June 12 – July 28, 2023
Course Dates

Austin Wright
Assistant Professor and Faculty Lead

This is the official report for the academic performance of

Full Name

with the completion of the following non-credit coursework for the

2023 Data and Policy Summer Scholar Program

Data Analytics in Public Policy

Hours of Instruction: 20

Grade: «D_point»

Introduction to Programming in R

Hours of Instruction: 20

Grade: «R_point»

Capstone Research Project

Hours of Instruction: 10

Grade: «Capstone_point»