Econ 370: Data Science for Economic Analysis

Department of Economics
Williams College
Fall 2026

1 Contact Information and Course Logistics

1.1 Professor: Dr. Pamela Jakiela

Email: `pj5@williams.edu`

1.2 Teaching Assistant: Christopher Hughes

Email: `cph4@williams.edu`

1.3 Course Meetings

Mondays, Wednesdays, Fridays from 11:00 AM to 12:15 PM in 129 Schapiro

1.4 Course Websites

1. Course materials are available on the public website for the course:

   `https://pjakiela.github.io/ECON370/`.

   That is where I will post lecture slides, information about assignments, class schedules, and other course information.

2. All assignments will be submitted through the course gradescope page:

   `https://www.gradescope.com/courses/1390275`

   All assignments will be submitted through gradescope. You'll receive an email inviting you to join the course, and that will allow you to set up a gradescope account linked to your Williams email address.

1.5 Office Hours

My drop-in office hours are on Mondays from 2:00 to 3:00 and Thursdays from 3:00 to 4:00 PM in 339 Schapiro. If you have a course that conflicts with my scheduled office hours, you can email me to set up an appointment.

TA Christopher Hughes will also be holding weekly office hours, time and location TBD.

1.6 Communication

Email is the best way to contact me (`pj5@williams.edu`). I will try to respond to course-related emails withing two working days of receipt.[1] I am sometimes overwhelmed by the volume of email I receive; if I have not responded to you within two working days, please feel free to send me an email reminder.

2 Course Description

2.1 From the Course Catalog

This course provides a hands-on introduction to data science tools most relevant for economic analysis including data visualization, machine learning, and text analysis. Economists and other social scientists use these data science tools differently than researchers in statistics and computer science – conducting empirical analysis that is explicitly grounded in economic theory and focusing on causal inference rather than prediction. Through a combination of lectures, hands-on labs, and group projects, students will develop the theoretical and practical skills needed to analyze economic data using modern data science techniques in both R and Python.

2.2 Learning Objectives

I have three main objectives in teaching ECON 370:

1. To help students develop advanced data skills, with a particular focus on identifying new sources of data, data cleaning/wrangling, and visualization;
2. To help students hone their ability to master new empirical tools (e.g. programming languages, statistical methods) independently to keep their skills current; and
3. To introduce students to new statistical tools (e.g. machine learning, text analysis) which are increasingly being incorporated into the empirical economist's toolbox.

2.3 Who Should Take ECON 370?

ECON 370 is an advanced elective intended for economics majors who have completed the intermediate level courses. ECON 255 (or its equivalent in the statistics department, STAT 346) is a required prerequisite.

3 Tentative Class Schedule and Important Dates

The updated course schedule is posted online at: `https://pjakiela.github.io/ECON370/schedule.html`. There is a final exam during the exam period which must be taken at the scheduled exam time and cannot be taken early under any circumstances.

[1] So, if you send me an email on Tuesday at noon, I will try to respond by the end of the day on Thursday. I do not check my Williams email on weekends. If you send me an email on Friday afternoon, I will respond to it by the end of the day on Tuesday.

Date(s)	Day	Description
09/11	F	Introduction
09/14	M	Exploratory data analysis
09/16	W	Lab: exploratory data analysis
09/18	F	Unsupervised learning
09/21	M	Lab: unsupervised learning
09/23	W	Data visualization, in-class worksheet, first EDA project group meetings
09/25	F	**Guest speaker and seminar: Kosuke Imai**
09/28	M	Numerical approaches to OLS
09/30	W	*No class: EDA project group meetings, posters due Thursday 10/1 at midnight*
10/2	F	Lab: numerical approaches to OLS*
10/5	M	Cross-validation*
10/7	W	**EDA project poster session**
10/9	F	Lab: cross-validation*
10/12	M	*No class: Reading Period*
10/14	W	Causal inference, potential outcomes, and RCTs*
10/16	F	*Mountain Day, no class**
10/19	M	Subset selection, regularization, lasso
10/21	W	Lab: lasso
10/23	F	Post-double-selection lasso, in-class worksheet
10/26	M	Random forests
10/28	W	Lab: random forests
10/30, 11/2	F, M	*No class: first RCT project group meetings*
11/4	W	Treatment effect heterogeneity, causal forests
11/6	F	Lab: treatment effect heterogeneity, causal forests
11/9	M	*No class: second RCT project group meetings*
11/11, 11/13	W, F	**RCT project presentations**
11/16	M	Lab: simple web-scraping, regular expressions
11/18	W	Word frequencies, tf-idf, in-class worksheet
11/20	F	Lab: working with text, word frequencies, tf-idf
11/23	M	*No class: first text as data project group meetings*
11/25, 11/27	W, F	*Thanksgiving break, no class*
11/30	M	Characterizing documents
12/2	W	Lab: characterizing documents
12/4, 12/7	F, M	*No class: second text as data project group meetings*
12/9, 12/11	W, F	**Text as data project presentations**

Italics indicates no regularly schedule class meeting.

Bold indicates student presentations during class meeting time.

* indicates subject to change depending on date of Mountain Day.

4 Readings

Data science is a new and rapidly evolving field, and there is no single textbook that is appropriate for the course. We will make use of several reference texts, all of which are freely available online. An up-to-date list of useful references is available at `https://pjakiela.github.io/ECON370/references.html`. Links to recommended readings associated with each course module will be posted on the course website.

You should download a pdf copy of *An Introduction to Statistical Learning* by James, Witten, Hastie, and Tibshirani. I recommend downloading the second edition of the R version, though students with prior Python experience are welcome to use the Python version instead (the two versions are pretty much identical except for the labs, which we will not be using). Both versions are available at `https://www.statlearning.com/`.

5 Assignments and Grading

Grades are calculated as follows:

Class Participation	10 points
In-Class Worksheets (3)	10 points
Lab Assignments	10 points
Exploratory Data Analysis Project	10 points
Treatment Effect Heterogeneity Project	20 points
Text as Data Project	20 points
Final Exam	20 points

5.1 Labs

Each course module will include a lab component that you will complete in R or Python (or, in one case, both). You will submit your code and data files via gradescope, and you may also be asked to answer some questions (in gradescope) about your results.

The labs account for ten percent of your final grade. The purpose of the labs is to give you the opportunity to gain hands-on experience using the empirical methods that we are learning. All scripts that you submit must run from start to finish and generate the correct results using only the methods and packages covered in class (unless otherwise noted).

5.2 Group Projects

Over the course of the semester, you will work in groups to complete three empirical projects and present your results to the class. The three projects are:

- An **Exploratory Data Analysis Project** where you seek to identify empirical patterns in cross-country data.

- A **Treatment Effect Heterogeneity Project** where you replicate and extend a randomized control trial, testing for differential impacts of a program or policy on subgroups defined by baseline characteristics.
- A **Text as Data Project** where you build and analyze a text data source.

Each of these projects will involve three graded components: initial meetings with me to discuss your work-in-progress, a presentation to the class, and the submission of your presentation materials (either a poster or a slide deck together with replication files that generate all of your tables and figures).

In your meetings with me, you will discuss your work on these projects. Members of your group will describe your data sources and the analytical approach you are taking. I will ask you questions about the work you have done, and we will discuss your plans for completing the project. During these meetings, we may discuss draft tables or visualizations that you have sent to me in advance. You will be allowed to bring hand-written note cards to our meetings, but you cannot read from your computer or from any printed material. In some cases, near the end of these meetings, I may invite you to open your laptop to show me code that you are struggling with – but this will only happen after I have evaluated your ability to discuss your analysis in your own words, without reference to code or other written materials.

Similar rules apply when you are presenting your work to the class: you cannot use any supporting material beyond your poster/slides and hand-written note cards (and reading directly from your slides will be heavily penalized). The goal of these projects is give you the opportunity to apply the tools we are learning in novel contexts, and the goal of the presentation is to share your knowledge with the class.

As discussed above, you will also submit replication files: a zipped folder containing both your raw data and the R and/or Python code that you used to process and analyze that data. You are responsible for explaining each piece of code to me upon request.

Meetings, presentations, and replication files are part of a data analysis pipeline, and each step feeds into the next. For this reason, you cannot deviate from the analysis plan that we discuss in a meeting without written permission from me.

5.3 In-Class Worksheets

You will complete three in-class, one-question worksheets over the course of the semester. On each worksheet, you will need to answer a question based on either course material or the readings.

5.4 Final Exam

There is a (closed-book) final exam, worth twenty points, during the exam period. It must be taken at the scheduled exam time and cannot be taken early under any circumstances.

5.5 Class Participation

Active, constructive participation in class meetings is a critical part of the course. During each class, I will be assigning you a participation score on a scale of 0 to 4, as follows:

- 4: Present in class, paying attention, fully participating, and making constructive comments and/or asking thoughtful questions, working with a group in lab, not looking at materials that are unrelated to the course, not looking at your phone (much)
- 3: Present in class, fully engaged, limited contributions to class/lab discussion
- 1: Physically present but not fully engaged (e.g. looking at your phone or outside websites), or only present for part of the scheduled class period
- 0: Not present

5.6 Late Assignments

Unless otherwise stated, all (unexcused) late assignments will be penalized: the maximum grade will be lowered by 10 percent for every day late for the first five days (including weekends). Assignments can only be submitted more than five days late with permission from the instructor.

6 Honor Code

The English language content of your labs and presentation slides should reflect your own work (or the work of the members of your group, in the case of group projects). To avoid violating the honor code, any English text written or prepared by someone else must be identified as such and cited appropriately. Any assignment containing either an image/figure/graph or more than five consecutive words taken from another source (e.g. a published paper or a website) without attribution will automatically receive a zero.

You will also receive no credit for graphs or tables that are identical or nearly identical to existing work (e.g. graphs from published papers or those posted on blogs), even if you submit code that generates the table or figure from the relevant raw data files. This means that you will not receive credit if you submit someone else's replication file(s) as part of an empirical project or lab. If I find a table or figure that you submit posted elsewhere, you will automatically receive a zero for the assignment with no opportunity for renegotiation.

6.1 ChatGPT, etc.

Generative AI (e.g. ChatGPT, Claude, etc.) is an invaluable tool that is changing the way we write code. You are allowed to use AI to help you write and debug code – in fact, I encourage you to do so. My hope is to help you learn how to use AI responsibly and effectively. You will find that AI makes mistakes, and you are responsible for correcting these mistakes. This involves learning to improve the engineering of your prompts and iterate effectively with AI agents. The goal is to learn how to leverage AI to help you code faster while also avoiding the pitfalls of AI slop.

As a starting point, I suggest that you ask chat or claude to translate or debug small sections of code (based on your own pseudocode). You are always responsible for submitting code that is correct and for making sure that your work (i.e. tables and figures) is original. In addition, I reserve the right to ask you to explain any code that you submit; if you are unable to do so, you will not receive credit for the relevant assignment.

7 Health and Accessibility Resources

Students with disabilities or disabling conditions who experience barriers in this course are encouraged to contact me to discuss options for access and full course participation. The Office of Accessible Education is also available to facilitate the removal of barriers and to ensure access and reasonable accommodations. Students with documented disabilities or disabling conditions of any kind who may need accommodations for this course or who have questions about appropriate resources are encouraged to contact the Office of Accessible Education at `oaestaff@williams.edu`.

8 Classroom Culture

The Williams community embraces diversity of age, background, beliefs, ethnicity, gender, gender identity, gender expression, national origin, religious affiliation, sexual orientation, and other visible and non visible categories. I welcome all students in this course and expect that all students contribute to a respectful, welcoming and inclusive environment. If you have any concerns about classroom climate, please come to me to share your concern.

9 Concluding Remarks

Congratulations on making it to the end of the syllabus. Do professors enjoy writing syllabi? Not really. Do students enjoy reading them? Probably not. Hopefully, this document will provide us with a shared set of expectations for the semester, making the course more constructive and enjoyable for everyone. Also, if you send me an email containing a "bad" data visualization together with a sentence explaining why it is bad before midnight on September 13 (subject line: ECON 370 bad data viz), you will earn a point of extra credit.