

# APEC 8222: Big Data Methods in Economics

## Fall 2023 Syllabus

### Instructor:

Justin Johnson ([jajohns@umn.edu](mailto:jajohns@umn.edu))

337f Ruttan Hall

### Office Hours

Tuesdays 1:00-2:00 P.M. in person (classroom or office 337F). If virtual is needed, go to <https://umn.zoom.us/my/jajohns> Appointments also welcome.

### Class time & location

Tuesday and Thursday, 11:45 A.M.–1:00 P.M.

October 24 – December 12

Ruttan Hall 135B

### Readings

#### Required texts:

Hastie, Tibshirani, and Friedman. 2009. *The Elements of Statistical Learning*, 2nd Edition. Springer. This book is available on Robert Tibshirani's website: <http://statweb.stanford.edu/~tibs/ElemStatLearn/>.

Turrell, Arthur. 2023. *Coding for Economists*. Freely available at <https://aeturrell.github.io/coding-for-economists/intro.html>

## Optional Texts:

Adhikari, Ani, John DeNero, David Wagner. *Computational and Inferential Thinking: The Foundations of Data Science*. 2nd Edition. Freely available at <https://inferentialthinking.com/chapters/intro.html>

Mueller and Guido (2017). Introduction to Machine Learning with Python. This book is a purchasable ebook, though the authors have provided free and legal copies in years past on their github page [https://github.com/amueller/introduction\\_to\\_ml\\_with\\_python](https://github.com/amueller/introduction_to_ml_with_python) and elsewhere.

Additional readings and materials will be on the course Canvas site

## Course Description

Challenges, techniques, and opportunities presented by data that has one or more of the following characteristics: large, unstructured, high frequency, variable quality. The course will consist of three parts: 1) computational tools for applying standard econometric techniques on large datasets, 2) working with new types of data, such as unstructured data (e.g. images, text) and web-scraped data, and 3) application of machine learning, AI and other new statistical learning techniques (e.g. classifiers, regression trees, machine learning, neural nets). This course is intended to be a broad introduction to many domains of knowledge with the goal of orienting students towards where they could learn more detailed information relevant to their research.

## Prerequisites

APEC 8221 or equivalent programming course: APEC 5031 or equivalent econometrics course.

## Objectives

You should leave this course with an understanding of and ability to estimate standard econometric models on large datasets, produce datasets for analysis from unstructured data, and apply several statistical learning techniques to answer economic questions. In particular, you will be able to

- Estimate standard econometric models on datasets containing millions of records
- Understand cross-validation approaches
- Assess the strengths and weaknesses of statistical learning techniques and evaluate their applicability to questions of causal inference.

- Gain competency on new machine learning models
- Achieve moderate skill level of Python programming

## Evaluation

Assignments: 70%

Independent Project: 20%

Class Participation: 10%

There will be six assignments during the course, which will consist of programming tasks designed to give you experience working with big and otherwise challenging data in the context of econometric analysis. We plan that you will have at least a week to complete them. You will submit Python-based Jupyter notebooks with your results and analysis, structured so that it can be run without modification. Assignments will be evaluated based on both functionality and the readability/organization of the code that you write. Please read the course policies below on collaboration and working together.

The independent project will be introduced in the middle of the semester and due near the end of the semester. You will submit a short (3-7 page) writeup, along with the code used. You will also give a short (5 minute) presentation on your results.

If you show up to class every day ready to learn with your computer setup with the day's tools, you will likely get full class participation points.

Late Policy: All assignments must be submitted on Canvas. Late assignments will be accepted for up to three days afterward. For each day late your grade will automatically reduce by 10% (for a maximum reduction of 30%).

There will be no opportunities for extra credit.

If you think that any grading was done incorrectly or unfairly, please come to my office hours. I am happy to revisit the grading, but will re-grade the entire assignment. The purpose of that policy is to be as fair as possible: if one problem needs re-grading, they probably all should have a second look.

Please see the UMN policy below on make-up work and policies concerning legitimate absences. In short, if a legitimate reason prohibits you from finishing an assignment on time, please let me know and we will make accommodations for you.

## Software

We will use the Python programming language (though there will be frequent reference to equivalent R functions). All software in the class is open source and freely available.

For Python, we will work through installation of the programming language and several supporting tools. This course strongly recommends that you bring your own laptop because having a well-setup laptop will serve you well into your academic career. It is not strictly required, but you will likely want to have administrator rights to your computer in order to install all of the supporting software. If you are using a personal computer, this should not be an issue. If you are using a university-owned computer, it might be. Reach out to me ASAP if you have questions about this. If you are not able to bring a suitable computer, please contact the instructors within the first week of class and we can discuss alternatives. It is possible to use a PC, Mac or Linux for this course, though all examples will be given on a PC.

Becoming skilled in Big Data is partly about mastering the tools and it will be your responsibility to come to class with your computer setup in a way for you to succeed. We will discuss any setup steps necessary in the lecture before it is to be used.

## Approximate Course Schedule

I will take more or less time on the topics as needed and will react to student feedback and requests.

Date	Topic	Readings	Assignments
24-Oct	Introduction & Course Overview	None	Assignment 1 Assigned (very short assignment)
26-Oct	Expanding your toolset, Introduction to Python [part 1]	Turrell, "Preliminaries" section: <a href="https://aeturrell.github.io/coding-for-economists/code-preliminaries.html">https://aeturrell.github.io/coding-for-economists/code-preliminaries.html</a>	
31-Oct	Introduction to Python [part 2]	Turrell, "Coding Basics" section: <a href="https://aeturrell.github.io/coding-for-economists/code-basics.html">https://aeturrell.github.io/coding-for-economists/code-basics.html</a>	Assignment 1 Due, Assignment 2 Assigned
2-Nov	Python on Big and/or Spatial Data	Optional: "Workflow Basics" and "Writing Code" sections	
7-Nov	Huge matrices with Numpy	<a href="https://nature.com/articles/s41586-020-2649-2">https://nature.com/articles/s41586-020-2649-2</a>	Assignment 2 Due, Assignment 3 Assigned

---

9-Nov	Huge tables with Pandas/Geopandas	Turrell, “Intro to Geo-Spatial Analysis” section: <a href="https://aeturrell.github.io/coding-for-economists/geo-intro.html">https://aeturrell.github.io/coding-for-economists/geo-intro.html</a>	
14-Nov	Machine Learning and Cross-Validation	“Machine Learning Methods Economists Should Know About”, Hastie et al. (2009) Chp 2	Assignment 3 Due, Assignment 4 Assigned
16-Nov	Regularization and Shrinkage	“Big Data: New Tricks for Econometrics”, Hastie et al (2009) Chp 3	Independent Project Assigned
21-Nov	Regression Trees, Random Forest and LULC classification	Hastie et al (2009) Chp 15	Assignment 4 Due, Assignment 5 Assigned
23-Nov	University Closed (Thanksgiving Holiday)		
Nov 28 (guest lecture/remote)	KGML (Knowledge Guided Machine Learning)	To be assigned.	List of 3 project ideas/datasets due
Nov 30 (guest lecture/remote)	R and Python Integration	<a href="https://aeturrell.github.io/coding-for-economists/coming-from-r.html">https://aeturrell.github.io/coding-for-economists/coming-from-r.html</a>	
5-Dec	Neural Nets	Hastie et al (2009) Chapter 11	Research proposal due
7-Dec	Convolutional Neural Net & Transformers	“Combining satellite imagery and machine learning to predict poverty”	Assignment 5 Due
12-Dec	Student Presentations!		Last Day of Class
15-Dec	No class		Research project due
21-Dec	No class		Final day of UMN semester

---

## Course Policies

1. Class attendance is expected. You will not be graded on attendance, but we expect you to come: any material covered in class is fair game on assignments and exams. Equally importantly, the questions you and your peers ask and the comments you make during class will make for a better experience for everyone, so please make every effort to attend. That said, unexpected conflicts come up for all of us, so any lecture notes and slides used during class will be available online.
2. You may discuss and work on assignments with other students, but you must write up and turn in your own assignment. You all have different strengths and can learn from one another by working together, so we encourage you to do so. The point of this policy is so that you can learn more, not less, so please don't abuse the privilege and let someone else do the work and you simply copy their answer. That won't help you learn the material.
3. Do not violate the Student Conduct Code. The assignments you turn in and your answers on exams should reflect your own work. Simply copying someone else's work, or otherwise violating the Student Conduct Code ([http://regents.umn.edu/sites/default/files/policies/Student\\_Conduct\\_Code.pdf](http://regents.umn.edu/sites/default/files/policies/Student_Conduct_Code.pdf)) may result in a failing grade and/or additional University action. That may sound scary, and it's supposed to be, but know that if you turn in work that is your own, you have nothing to worry about. If you have any questions or concerns about this policy, please don't hesitate to ask.
4. Treat your classmates with respect. Everyone in the class is here to learn and has an equal right to be treated with respect. This means many things, but in summary, We will not tolerate any form of discrimination or sexual harassment. The University has official policies that you can read if you aren't sure what constitutes either of those things, but your common sense will take you a long way.

### Other notes:

If you require disability accommodations, please let me know at your earliest convenience and we will work together to arrange accommodations. If you have or think you may have a disability and have not yet contacted Disability Services (DS), please do so at 612-626-1333 to arrange for a confidential consultation.

### Chat GPT Notice

We will literally be learning about how ChatGPT, Large Language Models and Transformers work in this course. Not surprisingly, I am fine with any usage of ChatGPT in your work.

## University of Minnesota Policies

### Student Conduct Code:

The University seeks an environment that promotes academic achievement and integrity, that is protective of free inquiry, and that serves the educational mission of the University. Similarly, the University seeks a community that is free from violence, threats, and intimidation; that is respectful of the rights, opportunities, and welfare of students, faculty, staff, and guests of the University; and that does not threaten the physical or mental health or safety of members of the University community.

As a student at the University you are expected adhere to Board of Regents Policy: *Student Conduct Code*. To review the Student Conduct Code, please see: <http://regents.umn.edu/sites/default/files/policies>

Note that the conduct code specifically addresses disruptive classroom conduct, which means “engaging in behavior that substantially or repeatedly interrupts either the instructor’s ability to teach or student learning. The classroom extends to any setting where a student is engaged in work toward academic credit or satisfaction of program-based requirements or related activities.”

### Use of Personal Electronic Devices in the Classroom:

Using personal electronic devices in the classroom setting can hinder instruction and learning, not only for the student using the device but also for other students in the class. To this end, the University establishes the right of each faculty member to determine if and how personal electronic devices are allowed to be used in the classroom. For complete information, please reference: <http://policy.umn.edu/Policies/Education/Education/STUDENTRESP.html>.

### Scholastic Dishonesty:

You are expected to do your own academic work and cite sources as necessary. Failing to do so is scholastic dishonesty. Scholastic dishonesty means plagiarizing; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering, forging, or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis. (Student Conduct Code: [http://regents.umn.edu/sites/default/files/policies/Student\\_Conduct\\_Code.pdf](http://regents.umn.edu/sites/default/files/policies/Student_Conduct_Code.pdf)) If it is determined that a student has cheated, he or she may be given an “F” or an “N” for the course, and may face additional sanctions from the University. For additional information, please see: <http://policy.umn.edu/Policies/Education/Education/INSTRUCTORRESP.html>.

The Office for Student Conduct and Academic Integrity has compiled a useful list of Frequently Asked Questions pertaining to scholastic dishonesty: <http://www1.umn.edu/oscai/integrity/student/index.html>. If you have additional questions, please clarify with your instructor for the course. Your instructor can respond to your specific questions regarding what would constitute scholastic dishonesty in the context of a particular class-e.g., whether collaboration on assignments is permitted, requirements and methods for citing sources, if electronic aids are permitted or prohibited during an exam.

### **Makeup Work for Legitimate Absences:**

Students will not be penalized for absence during the semester due to unavoidable or legitimate circumstances. Such circumstances include verified illness, participation in intercollegiate athletic events, subpoenas, jury duty, military service, bereavement, and religious observances. Such circumstances do not include voting in local, state, or national elections. For complete information, please see: <http://policy.umn.edu/Policies/Education/Education/MAKEUPWORK.html>.

### **Appropriate Student Use of Class Notes and Course Materials:**

Taking notes is a means of recording information but more importantly of personally absorbing and integrating the educational experience. However, broadly disseminating class notes beyond the classroom community or accepting compensation for taking and distributing classroom notes undermines instructor interests in their intellectual work product while not substantially furthering instructor and student interests in effective learning. Such actions violate shared norms and standards of the academic community. For additional information, please see: <http://policy.umn.edu/Policies/Education/Education/STUDENTRESP.html>.

### **Grading and Transcripts:**

The University utilizes plus and minus grading on a 4.000 cumulative grade point scale. Please see <http://policy.umn.edu/education/gradingtranscripts> for details..

### **Sexual Harassment**

“Sexual harassment” means unwelcome sexual advances, requests for sexual favors, and/or other verbal or physical conduct of a sexual nature. Such conduct has the purpose or effect of unreasonably interfering with an individual’s work or academic performance or creating an intimidating, hostile, or offensive working or academic environment in any University activity or program. Such behavior is not acceptable in the University setting. For additional information, please consult Board of Regents Policy: <http://regents.umn.edu/sites/default/files/policies/SexHarassment.pdf>



### **Equity, Diversity, Equal Opportunity, and Affirmative Action:**

The University provides equal access to and opportunity in its programs and facilities, without regard to race, color, creed, religion, national origin, gender, age, marital status, disability, public assistance status, veteran status, sexual orientation, gender identity, or gender expression. For more information, please consult Board of Regents Policy: [http://regents.umn.edu/sites/default/files/policies/Equity\\_Diversity\\_EO\\_AA.pdf](http://regents.umn.edu/sites/default/files/policies/Equity_Diversity_EO_AA.pdf).

### **Disability Accommodations:**

The University of Minnesota is committed to providing equitable access to learning opportunities for all students. The Disability Resource Center is the campus office that collaborates with students who have disabilities to provide and/or arrange reasonable accommodations.

If you have, or think you may have, a disability (e.g., mental health, attentional, learning, chronic health, sensory, or physical), please contact DS at 612-626-1333 to arrange a confidential discussion regarding equitable access and reasonable accommodations.

If you are registered with DS and have a current letter requesting reasonable accommodations, please contact your instructor as early in the semester as possible to discuss how the accommodations will be applied in the course.

For more information, please see the DS website, <https://diversity.umn.edu/disability/>.

### **Mental Health and Stress Management:**

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 and may reduce your ability to participate in daily activities. University of Minnesota services are available to assist you. You can learn more about the broad range of confidential mental health services available on campus via the Student Mental Health Website: <http://www.mentalhealth.umn.edu>.

### **Academic Freedom and Responsibility**

Academic freedom is a cornerstone of the University. Within the scope and content of the course as defined by the instructor, it includes the freedom to discuss relevant matters in the classroom. Along with this freedom comes responsibility. Students are encouraged to develop the capacity for critical judgment and to engage in a sustained and independent search for truth. Students are free to take reasoned exception to the views offered in any course of study

and to reserve judgment about matters of opinion, but they are responsible for learning the content of any course of study for which they are enrolled.\*

Reports of concerns about academic freedom are taken seriously, and there are individuals and offices available for help. Contact the instructor, the Department Chair, your adviser, the associate dean of the college, or the Vice Provost for Faculty and Academic Affairs in the Office of the Provost.

*\* Language adapted from the American Association of University Professors “Joint Statement on Rights and Freedoms of Students”.*