Instructor: Jon Williams   UNC – Chapel Hill   Spring Semester 2023
Office: Gardner Hall, 300A   Department of Economics
Email: jonwms@unc.edu

ECON 370: Economic Applications of Data Science

Schedule Information:

Class: T/TR 9:30-10:45 am
Class Location: Gardner 307

Instructor Office Hours: T/TR 11:00-12:30 pm
Instructor Location: Gardner 300A

TA (Haoran Zhang) Office Hours: F 1:00-2:00pm
TA Location: Gardner 409

Course Description:

Econ 370 is intended to introduce you to a programming language commonly used for economic applications of numerical methods and data science. The topics to be covered include an introduction and overview of the Python programming language, basic numerical methods, and data analysis and visualization. There will be lectures on each of these topics, but the core of the course will be a group project that directly applies these methods to study a topic of each group's choosing. This permits students to pursue more-advanced skills most relevant for their goals while establishing a strong foundation to build upon in future course work and employment.

Method of Evaluation:

Class Participation (20%): Attendance and Involvement in Discussions
Project Preliminary Slides, Programs, and Presentations (50% = 2 x 25%)
Project Final Slides, Programs, and Presentation (30%)

At the instructor's discretion, grades will be curved to have a distribution similar to previous sections of 500-level courses in the economics department at UNC.

Examination Policies and Dates (subject to change if needed):

Project Final Slides and Programs Due: May 9th, 8am
Project Final Presentations: May 9th, 8am-11am

Important Dates:

Classes begin (end) 1/9 (4/28)
Last day for late registration 1/13
Holidays 1/16, 4/28
Well-being days 2/13, 2/14, 4/6
Spring break 3/13-3/17

Assessments:

Competency as a programmer requires establishing a strong foundational understanding of best general (i.e., any language or task) and language-specific (e.g., Python) practices, so as to optimize your code for whatever task is at hand. Yet, each task is different and may require specialized knowledge and syntax that is most easily learned through doing.

The lectures throughout the course are intended to help establish a foundational understanding of the Python language. This is the reason for making 20% of your grade based on attendance and participation in discussions.

The project is intended to develop your ability to clearly define a research question, collect the data to pursue it, and then develop the necessary programming expertise to conduct the analysis and present your findings. The project should be completed with groups of 3-5 that share an interest in a topic. I have high expectations for the projects, so each member of the group should be continually contributing to all aspects to be successful (grades will differ across group members if inputs or outputs are substantially different). However, the group structure also allows for some specialization and expertise to be developed among the members. You will be evaluated on your project at three different times throughout the semester:
- Develop a slide deck that motivates and clearly defines a research question, and identifies and discusses the data that will be used to answer it. You will submit the slide deck and data resources for review. Your grade will be based on the slide deck, presentation, and usefulness and validity of data resource. Timing TBD. (25%)
- Your updated slide deck will incorporate feedback from the first presentation and build upon it to describe in detail the method or research design by which you'll evaluate your research question. In addition, you'll develop an initial set of Python programs to implement it and this code base will be described in the slide deck. You will submit the updated slide deck and code. Your grade will be based on the slide deck, presentation, and progress on programming. Timing TBD. (25%)
- Your final will involve presenting the entire updated slide deck that incorporates all feedback from first two presentations and includes your analysis and findings. In addition to submitting the slide deck, you will present a final version of the data and programs. Your grade will be based on the slide deck, presentation, and quality of the analysis (i.e., programming, research design, data resources, etc). During final exam time. (30%)

Regarding the type of topics that can be chosen, I am completely flexible. I want the topic to be of direct interest to members of the group and be useful in developing the desired skillset. Examples might include measuring effects of policies (e.g., success of regulatory interventions), monte-carlo exercises (e.g., comparing alternative ML algorithms for particular objectives), or economic modeling and simulation (e.g., develop behavioral economic model and calibrate to match real-world data).

Readings:

There will be no mandatory texts for the course. I will use Sakai (or other reasonable electronic means like email) to distribute notes, readings, homework assignments, etc. Please verify that you can log on to Sakai and access the information for this class. Attendance is still extremely important as materials will periodically be distributed in the classroom without prior notification. For those interested in further advancing their knowledge beyond notes from class, there are some great resources. Here are a few of my favorties:
- *Python for Data Analysis*: Wes McKinney (MK)
- *Introduction to Machine Learning with Python*: Andreas Muller and Sarah Guido (MG)
- *An Introduction to Statistical Learning (with Applications in R):* Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani

Tentative Course Outline (order/content subject to change at instructor's discretion):

Python Introduction and Overview:
- What is Python and how is it useful to economists?
- Anaconda installation
- Spyder Integrated Development Environment (IDE)

Basic Operations:
- Common object types: float, integers, strings
- Booleans, if-then statements, loops
- Numerical algebraic calculations
- Lists, list comprehensions, tuples, dictionaries
- Slicing

User-defined Functions:
- Structure, visibility, returns
- Objects, attributes, and methods

Familiarization w/ Capabilities of Common Packages
- OS: loading and saving data (input/output)
- Pandas: DataFrames calculations w/ Pandas
- Matplotlib: visualizing data

Applications of Common Packages
- Cleaning data: replace, map, apply, applymap, unique, strip, cut, qcut (MK 7)
- Reshaping and merging data: multiindex, stack, unstack, pivot, melt (MK 8.1, 8.3)
- Collapsing data: groupby, transform (MK 10.1, 10.2)

Other Useful and Advanced Topics
- Regular Expressions (Re): searching and matching strings (MK 7.3)
- Geospatial Data (Geopandas): plotting geospatial data and Chloropleths
- Webscraping (Requests, BeautifulSoup): extracting data from webpages
- Application Programming Interfaces (Pandas Datareader): APIs (MK 6.3)
- Basic Econometrics (Patsy, Statsmodels): linear regression, logit, etc (MK 13.3)
- Supervised Learning (Scikit-learn): regularization via ridge, LASSO, elastic-net (MG 2.2, 2.3, 2.4)

- Unsupervised Learning (Scikit-learn): clustering via k-means and dimension reduction via principal component analysis (MG 3.4.1, 3.5.1)
- ML with text data (nltk, wordcloud, autocorrect): Natural Language Processing, Bag of Words, Term Frequency-Inverse Document Frequency (MG 7)

Classroom etiquette:

My goal is to maintain a classroom environment that provides a good learning environment for everyone. To minimize distraction, you must turn off all cell phones, laptops, and other electronic devices during class, unless explicitly instructed otherwise, as there may be times during the course that an open laptop is appropriate and desirable. I expect you to arrive on time and prepared for the day's class.

Academic integrity:
You are encouraged to work together with other students on the homework. However, each student must write up her or his own version of the homework assignment. Exams are taken individually, without any help from other students or unapproved resources. Each student is expected to maintain academic integrity and follow the UNC Honor System.