Data Science for Economic Development
Pierre Biscaye, Chaire de Professeur Junior
Email : pierre.biscaye@uca.fr

Université Clermont Auvergne
Masters 2 AED / Magistère 3

Spring 2025

Course objectives

- Understand how to think about using data in applied economic research
- Build a foundation for using important tools in data science for economics: generative AI, big data, geospatial analysis, machine learning, and more
- Develop competency in working with different types of data using Python

Course logistics

- Language of instruction: English
- Class time: 20 hours of *cours magistraux* in 10 two-hour sessions
  - Salle 212 RO
  - **Bring your laptops** to class!
- Office hours: Thursdays from 9-11am, Office 425
- Assignments: weekly coding practice, not graded
- Evaluation:
  - 1-hour final exam
  - Mémoire: Preparation of short original research paper using tools from class to answer an economic question
    - In groups of 3-4
    - Paper: 10-15 pages including references and exhibits
      - 15-20 pages for groups of 4
    - Code: must submit the set of code files used to produce the paper exhibits from original data (not graded)
    - Presentation: 5-minute presentation of proposed research in last lecture

Course outline and schedule

1. Introduction and Python basic (9 January)
   a. Lecture: Course overview and introduction to data science
   b. Python
      i. Using Jupyter Notebook
      ii. Python basics: variables, types, indexing
      iii. Python errors/debugging
2. Reproducibility, more Python basics (13 January)
   a. Lecture: Reproducibility, organizing files and code
   b. Python
      i. Functions and methods
      ii. Loops, conditionals, list comprehension
      iii. Importing modules/packages
      iv. Working with Files

     v. Writing code well
3. Data wrangling and visualization (January 15)
  a. Lecture: Value of data visualization
  b. Python
    i. Pandas and data frames
    ii. Importing, subsetting, joining data
    iii. Summary statistics
    iv. Scatter plot, bar plot, histogram
    v. Using matplotlib
    vi. Simple linear regression
4. Generative AI and big data (January 16)
  a. Lecture
    i. Generative AI: What is generative AI; Large Language Models/Chatbots; Applications of generative AI in economics
    ii. Big Data: Data types, data sources, big data, tools for working with big data
5. Geospatial analysis 1 (January 21)
  a. Lecture: Geospatial data in economics
  b. Python
    i. Tabular spatial data
    ii. Shapely and geopandas
    iii. Rasters
    iv. Visualizing spatial data
    v. Map algebra
6. Geospatial analysis 2 (January 23)
  a. Lecture: Remote sensing in economics
  b. Python
    i. Satellite imagery and remote sensing
    ii. Google Earth Engine
    iii. Accessing spatial data
    iv. Preparing spatial data for analysis
    v. Point-level data analysis
7. Machine learning 1 (January 28)
  a. Lecture: Intro to ML, supervised ML
  b. Python
    i. Regression as machine learning foundation
    ii. Fitting and regularization
    iii. Validation
    iv. Pre-processing
8. Machine learning 2 (January 30)
  a. Lecture: Unsupervised ML, bias and issues with ML
  b. Python
    i. Classification
    ii. Machine learning pipeline
    iii. Unsupervised ML: clustering and dimensionality reduction
    iv. Other topics
9. Web scraping and text analysis (February 5)
  a. Lecture: Additional tools for accessing data for economics research
  b. Python
    i. Web APIs

      ii.   Web scraping
     iii.   Text preprocessing
     iv.   Bag-of-words
      v.   Word embeddings
10. Wrapping up, project discussion (February 6)
    a.   Cover spillover material from previous lectures
    b.   Lecture: Key lessons
    c.   Presentation of research proposals for mémoire

Python Resources

1. Guide to setting up Anaconda, Python, and Jupyter Notebooks for the course:
   https://github.com/pbiscaye/IntroDataScienceEcon/blob/main/Data_Science_Class_Setup_Guide.pdf
2. All links below as well as many others are included in my list of useful data science resources:
   https://github.com/pbiscaye/Teaching/blob/main/Data%20science%20resources%20index.pdf
3. Getting started with jupyter notebooks: https://www.dataquest.io/blog/jupyter-notebook-tutorial/
4. UC Berkeley D-Lab Python resources: https://github.com/dlab-berkeley/python-berkeley
5. UC Berkeley D-Lab Python fundamentals: https://github.com/dlab-berkeley/Python-Fundamentals
6. Arthur Turrell Python for data science:
   https://aeturrell.github.io/python4DS/welcome.html
7. Gabor Bekes coding for data analysis with Python: https://github.com/gabors-data-analysis/da-coding-python
8. Data camp data scientist with Python: https://app.datacamp.com/learn/career-tracks/data-scientist-with-python?version=5
9. NYU Python data bootcamp: https://nyudatabootcamp.gitbook.io/thebook/
10. NYU intro to Python for science:
    https://physics.nyu.edu/pine/pymanual/html/pymanMaster.html