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 2026

Course objectives

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

Course logistics

- Language of instruction: English
- Course website: https://github.com/pbiscaye/IntroDataScienceEcon
- Class time: 20 hours of *cours magistraux* in 1 two-hour and 6 three-hour sessions
    o Salle 212 RO
    o **Bring your laptops** to class!
- Office hours: Office 425, Tuesdays from 15h30-16h30 et Thursdays from 11-12h
    o Sign-up: https://calendar.app.google/Tzo7ZMJo6U69kNE87
- Assignments: weekly coding practice, not graded
- Evaluation:
    o 1-hour final written **exam**
    o **Mémoire**: In groups of 3-4, apply tools from class and different data sources to analyze incidence of flooding in Nigeria's Jigawa State.
        ▪ Due date: March 2 by 17h
            • Intermediate outputs due periodically
        ▪ Final outputs
            • ~15 page report including exhibits
            • Zip file of all code files used to produce paper exhibits, with readme

Course outline and schedule (subject to change)

1. Introduction and Python basics (January 8, 14h-17h)
    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
        iv. Functions and methods
        v. Loops, conditionals, list comprehension
2. Reproducibility, data wrangling (January 15, 15h30-17h30)
    a. Lecture: Reproducibility, organizing files and code

    b. Python
- i. Writing code well
- ii. Importing modules/packages
- iii. Working with files
- iv. Pandas and data frames
- v. Importing, subsetting, joining data

3. Generative AI, data visualization (January 19, 14h-17h)
   a. Lecture: Generative AI: What is generative AI; Large Language Models/Chatbots; Applications of generative AI in economics
   b. Python
      - i. Summary statistics
      - ii. Scatter plot, bar plot, histogram
      - iii. Using matplotlib
      - iv. Exploring tabular data

4. Accessing and working with different data sources (January 26, 14h-17h)
   a. Lectures
      - i. Big Data in economics
      - ii. Web scraping
      - iii. Web APIs
   b. Python
      - i. Web scraping
      - ii. Web APIs

5. Spatial data and remote sensing (February 2, 14h-17h)
   a. Lecture: Introduction to spatial data and remote sensing
   b. Python
      - i. Tabular spatial data
      - ii. Shapely and geopandas
      - iii. Rasters
      - iv. Visualizing spatial data
      - v. Map algebra
      - vi. Satellite imagery and remote sensing
      - vii. Accessing spatial data
      - viii. Preparing spatial data for analysis
      - ix. Point-level data analysis

6. Machine learning (February 12, 14h-17h)
   a. Lecture: Introduction to machine learning
   b. Python
      - i. Regression as machine learning foundation
      - ii. Fitting and regularization
      - iii. Validation
      - iv. Pre-processing
      - v. Classification
      - vi. Machine learning pipeline
      - vii. Unsupervised ML: clustering and dimensionality reduction

7. Text analysis (February 23, 14h-17h)
   a. Lecture: Introduction to text analysis
   b. Python
      - i. Text preprocessing
      - ii. Bag-of-words
      - iii. Word embeddings

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