
BASc0005: Quantitative Methods 2 - Data Science and Visualisation

Convenor: Dr. Ollie Ballinger

Module Overview

The module teaches quantitative skills, with an emphasis on the context and use of data. Students learn to focus on datasets that allow exploration of questions in society – in arts, humanities, sports, criminal justice, economics, inequality, or policy. The student will work with Python for:

- Data manipulation (cleaning and segmentation),
- Analysis (e.g., deriving descriptive statistics),
- Visualisation (graphing, mapping, and other forms of visualisation).

Students will engage with literature around their topics and connect their datasets and analyses to explore wider arguments, linking their results to these contextual considerations.

The module is assessed by a group research project that uses data analysis and visualisation to explore a real-world question. This model follows typical data-driven research projects at a postgraduate and postdoctoral level.

Module Overview Contents

1. Expectations
2. Themes and Focus
3. Module Strands
 - Data Contexts
 - Data Analysis
 - Data Design
4. Delivery: Data Workshops
5. Interdisciplinarity and Research-Based Teaching
6. Connection to Advanced Courses
7. Practicalities
 - Teaching Timetable
 - Course Deadlines
8. Assessment
 - Group Presentation (10%)
 - Problem Sets (30%)
 - Group Website (60%)
9. Recommended Prerequisites

Expectations

BASc0005 is a challenging module where students are expected to cover a large amount of material and work as a team on a data-driven research project with real-world applications. Expectations include:

- Attend all teaching sessions on time,
- Complete homework and participate in classes,
- Fully contribute to assessed group work and presentations,
- Be respectful of others' views and support their learning,
- Adhere to UCL's standards of transparency, scholarship, and ethics.

Students can expect support from the teaching team on module-related issues, helpful resources, and assistance with additional personal or academic needs. The team will respect students' views in the classroom and help build skills to meet the goals of the module, degree, and education.

Themes and Focus

Module Strands

- **Data Contexts:**
Understanding how data is accessed, analysed, communicated, and its impacts. This includes principles of design for mapping, data visualization, the Open Data movement, Government and Census data, private data and ethics, aggregation, and uncertainty. These strands will be covered in weekly lectures.
 - **Data Analysis:**
Explore how data is structured and how this affects how we work with it. Topics include numerical data, text analysis, geographical data, linked data (semantic web), normalization, scaling, and social media data.
 - **Data Design:**
Discuss the communication of data after analysis. Students will learn graphing, histograms, summarization, mapping, and explore broader ideas of data visualisation.
-

Delivery: Data Workshops

Students will be taught to use Python libraries (Matplotlib, Pandas, Basemap, NLTK) for importing and working with raw data, producing tables, scatter graphs, histograms, summary statistics (mean, median, mode, standard deviation,

etc.), and maps. iPython notebooks will be used to create a framework for combining narrative, explanation, images, and equations.

At the end of the module, students will have gained technical skills paired with the context of data: practical, political, communicative, ethical, and transformative. Weekly two-hour Python workshops will complement independent work.

Interdisciplinarity and Research-Based Teaching

Students will focus on datasets that allow exploration of societal issues and will use Python for data manipulation, analysis, and visualisation. The module aims to help students critically examine data, and representations of data, in visual form and journalism. The module is assessed through a group research project.

Connection to Advanced Courses

This module serves as a starting point for students interested in advanced study in data science, mapping, visualisation, or digital humanities. Examples include:

- CASA: Spatial Data Science and Visualisation,
 - Digital Humanities: MA/MSc, Digital Humanities,
 - Computer Science: MSc, Web Science and Big Data Analytics,
 - Geography/CEGE: MSc, Geographic Information Science,
 - CHIME: MSc, Data Science for Research in Health and Biomedicine.
-

Practicalities

Teaching Timetable

Lectures and workshops are based on set readings and are led by Ollie Ballinger and supported by technical workshops led by teaching assistants. The timetable, including locations, is available on the UCL common timetable system and on the “overview” tab of the module Moodle page.

Course Deadlines

Please see Moodle for the latest assessment deadlines.

Assessment

The module will be assessed through a group project that integrates context, analysis, design, communication, and contextualization of outputs. The main

assessment task is to produce a **website** where the group analyses, visualizes, and summarises a dataset, placing it in a wider context. The website will consist of several pages discussing:

- The nature and sources of the data,
- Problems encountered,
- Methods used,
- Visualisations created,
- Patterns found,
- Conclusions drawn.

Groups will be formed in the first week of the module, and each group will collectively choose the dataset(s) to work with. Assessment is divided into three components:

1. **Group Presentation (10%)**

This takes place the week after reading week and showcases the group's proposal and initial plans. The group presentation will demonstrate progress toward the website, and all students in the group should contribute. Each group will have around 10 minutes to present, including time for questions.

- **Assessment Criteria:**

Clear presentation of project goals, appropriate use of visual aids, prioritization, and pacing.

Demonstration of project objectives and planning.

Contributions from all team members.

- **Feedback:**

Written feedback will be provided by the beginning of the final week of term.

2. **Problem Sets (30%)**

Weekly coding workshops develop students' coding skills and data science knowledge. Collaboration is encouraged, but each problem set will contain an Assessed Question, which students must answer independently. At the end of the term, these answers will be entered into a Moodle quiz, and students will provide a link to a GitHub repository to demonstrate their engagement throughout the term.

- **Assessment Criteria:**

Completeness of problem sets, and correct answers to Assessed Questions.

3. **Group Website (60%)**

The website should analyze, visualize, and summarize a dataset or datasets, addressing the following elements:

- **Content:**

At least one context-centric entry, one visualisation-centric entry, and one methodological-centric entry.

- **Technical Sophistication:**
Use of Python (or other software) to analyze and visualize data.
 - **Explanation of Methodology:**
Explanation of techniques used in a clear and understandable manner.
 - **Reflection on Context:**
Discussion of data collection methodology, quality, and the potential impact of the analysis.
 - **Demonstration of Scholarship:**
Proper citation of sources, including data and code libraries.
 - **Contribution:**
Individual contributions will be assessed based on the project quality and their demonstrated input via the website. Students will receive and individualized mark using Individual Peer Assessed Contribution (IPAC) weighting.
 - **Feedback:**
Written feedback will be provided within four weeks of submission.
-

Recommended Prerequisites

Students are expected to have a basic familiarity with Python and iPython. Although the course does not focus on programming, Python will be used to manipulate, analyse, and display data.

Suggested Resources for Preparation:

1. **Codecademy:**
A beginner-friendly Python course (focusing on variables, lists, and dictionaries).
<https://www.codecademy.com/catalog/language/python>
2. **Think Python:**
A more traditional book available for free:
<https://www.greenteapress.com/thinkpython/>
3. **Software Carpentry Tutorials:**
Useful for data manipulation:
<https://swcarpentry.github.io/python-novice-inflammation/>
4. **Learn Python:**
An online interface for Python tutorials:
<https://www.learnpython.org>
5. **Learn Python the Hard Way:**
A structured course designed to create good habits, with a free online

version available here:
<https://learnpythonthehardway.org>
