

# Introduction to Python

Research and Expertise Centre for Survey Methodology (RECSM)  
Summer Methods School 2022

Tom Paskhalis

[tom@paskhal.is](mailto:tom@paskhal.is)

- **Duration:** 2 days (8 hours overall)
- **Dates:** 27–28 June 2022
- **Location:** RECSM, University Pompeu Fabra, Barcelona
- **Instructor:** Tom Paskhalis ([tom@paskhal.is](mailto:tom@paskhal.is))
- **Course Website:** [bit.ly/RECSM\\_Python](https://bit.ly/RECSM_Python)
- **Course Registration:**  
<https://eventum.upf.edu/80340/detail/recsm-summer-methods-school-2022.html>

## Overview

[Python](#) is a versatile, multi-purpose, open-source programming language which has become one of the [top choices](#) for data engineering and advanced analytics. Due to its interactive nature and easy syntax Python is increasingly used in academia, business, and even [journalism](#). The purpose of this course is to provide a gentle introduction to Python with a focus on its data analysis capabilities.

This course is aimed at participants who would like an overview of/introduction to data manipulation and statistical analysis in Python. It can also serve as an overview/refresher of the basics of Python programming. In the class we will focus on understanding the core Python objects and getting hands-on experience with the main aspects of data analysis for social scientists.

## Prerequisites

This is an introductory class and no prior experience with programming is required. Familiarity with some statistical analysis techniques (e.g. linear regression) and R/Stata would be an advantage, but is not essential.

## Software

In this course we will use Python 3 and Jupyter Lab/Notebook. The easiest way to install all of the required software is by using [Anaconda distribution](#).

Alternatively, you may choose to use [Google Colab](#), a cloud platform for hosting Jupyter Notebooks. Its interface is slightly different and you need to have a Google account, but it does not require any local installation.

## Class Schedule

Classes will take place on 27 and 28 June.

Date	Time (CEST)	Topic
27 June	09:00-10:45	Introduction to Python objects and data types
	10:45-11:15	Break
	11:15-13:00	Pandas, data input/output
28 June	09:00-10:45	Exploratory data analysis, data visualization
	10:45-11:15	Break
	11:15-13:00	Regression analysis, communicating results

## Materials

The course is designed to be self-contained. All the materials are available in this [GitHub repository](#). However, some students might find additional materials helpful. The following texts provide a good introduction to Python programming with a focus on data analysis applications:

- John Guttag. 2016. *Introduction to Computation and Programming Using Python: With Application to Understanding Data*. 2nd ed. Cambridge, Massachusetts: The MIT Press
- Wes McKinney. 2017. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 2nd ed. Sebastopol, CA: O'Reilly Media
- Al Sweigart. 2019. *Automate the Boring Stuff with Python*. 2nd ed. San Francisco, CA: No Starch Press

Additional online resources:

- [Python For You and Me](#)
- [Python Wikibook](#)
- [Python 3 documentation](#) (intermediate and advanced)

## Short bio

Tom Paskhalis is an Assistant Professor in Political Science and Data Science at [Trinity College Dublin, Department of Political Science](#). Previously, he was a Postdoctoral Fellow at [New York University, Center for Social Media and Politics](#). He received his PhD from [London School of Economics and Political Science, Department of Methodology](#). His research focuses on political communication, comparative politics, and computational analysis.

More information:

 [tom.paskhal.is](https://tom.paskhal.is)

 [tpaskhalis](#)

 [tpaskhalis](#)