

Introduction to Python

Week 1 (Introductory course)

Duration: 8 hours/2 days

Course description

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.

Schedule

Introduction to Python objects and data types

Pandas, data input/output

Exploratory data analysis, data visualization

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

Additional online resources:

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

Short biography



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.