Introduction to Python

Social Science Methods Workshops 2023, Lund University

Instructor: Nils Holmberg, Strategic Communication, Lund University

Contact: nils.holmberg@isk.lu.se

Dates: September 25—29, at 09—12 am (schedule)

Format: Campus-based

Workshop syllabus

A multitude of resources to learn Python exists online (see Matthes, 2019; also Géron, 2019). On this workshop, we will spend as much time as possible with hands-on practical usage of python programming. Datasets, code materials, and installation instructions will be provided at the start of the workshop. In order to prepare for the workshop you can explore the website!

Day 1: Setting up a python environment

There are many ways of setting up a **python environment**. Day 1 of the workshop we are going to have a look at some of the most common use cases, and discuss pros and cons with different setups depending on the specific research tasks you are trying to solve. We will also have a look at **basic python syntax** and write some simple python programs. The main format of instruction will be Jupyter notebooks.

Day 2: Getting started with data analysis

One of the most common use cases for python is **data analysis**. On day 2 of the workshop we are going to focus on a package called pandas. This package will allow us to import structured data such as excel files and other quantitative tabular data, and transform and summarize such data. We will also introduce scikit-learn for statistical analysis. Participants are invited to bring their own data files.

Day 3: Visualizing and reporting data

Analyzing data is only half of a data scientist's work. The other half is about using graphs and diagrams to **visualize data** in ways that will allow us to report our results as intuitive and compelling stories. On day 3 of the workshop we are going to focus on packages such as matplotlib and seaborn to produce various types of diagrams ranging from simple frequency diagrams to time series and heatmaps.

Day 4: Processing natural language texts

The opposite of structured data is **natural language texts**. However, as social scientists, we often encounter such unstructured text data. The challenge of managing and analyzing natural language data will be the main focus on day 4. To our help we will utilize both classical packages for data cleaning and tokenization such as NLTK, and more modern packages such as spaCy that relies on machine learning models to analyze language.

Day 5: Finishing with data collections

As researchers we not only need to analyze data, but the first step is usually to **collect empirical data**. Day 5 of the workshop we will turn our attention towards this challenge, and we will explore a couple of packages that can help with this. First off, there is Beautiful Soup which can be used to scrape and analyze web pages. Second, we will have a look at solutions for collecting survey data using python!

References

Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Matthes, E. (2019). Python crash course: A hands-on. *Project-Based Introduction to Programming*.