

# Introduction to Python

Social Science Methods Workshops 2023, Lund University

**Instructor:** Nils Holmberg, Strategic Communication, Lund University

**Contact:** [nils.holmberg@isk.lu.se](mailto:nils.holmberg@isk.lu.se)

**Dates:** September 25—29, at 09—12 am ([schedule](#))

**Format:** Campus-based

## 1 Workshop syllabus

Install

### 1.1 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. In order to prepare for the workshop you can explore workshop [website](#)!

### 1.2 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 perform transformations and summary statistics on them. We will also introduce the package [scikit-learn](#) for inferential statistical analysis. Datasets will be provided at the workshop, but participants are also invited to bring their own data files.

### 1.3 Day 3: Visualizing and reporting data

Analyzing data is often 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.

### 1.4 Day 4: Processing natural language texts

The opposite of structured data is **natural language texts**. However, as social scientists, we often encounter data in this unstructured format. The challenge of managing and analyzing natural language data will be the main focus on day 3. 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.

### 1.5 Day 5: Getting started 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!

localization ([Géron, 2019](#); [Matthes, 2019](#)).

## 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*.