

## Professor

Prof. Paul Rognon

## Overview and Objectives

The "Coding Bootcamp in Python and R" provides basic training in programming with Python and R for data analysis and machine learning. This is an intensive 8-hour course based on a hands-on approach using Jupyter notebooks.

This course has been specially designed to provide you with essential coding skills that come in handy when initiating or continuing your training in Data Science. If you are interested in mastering this area and learning the tools to apply modern machine learning techniques, from data exploration to building predictive models and extracting insights, this intensive 8-hour course is the ideal preparation to get started.

The bootcamp will also give you the required preparation for other courses of the BSE Summer School in Data Science.

## Prerequisites

Although not mandatory, some knowledge of Python, Jupyter notebooks, and matrix algebra is recommended. Students are encouraged to install R and Python on their own laptop or desktop computer. This is also non-mandatory since the course is based on Jupyter notebooks that can be run with Google Collab.

## Course Outline

The course evolves along the following thematic units:

### 1. Programming with Python

- Running Python with Jupyter
- Basic variables: Numbers and strings
- Basic operations
- Loops, control flow
- Data structures: Lists, maps, reductions
- Functions and classes

- Inputs and outputs
- String manipulation
- Date/time manipulation
- Special packages: panda and numpy

## 2. Programming with R

- Running R with Jupyter
- Basic variables: Numbers and strings
- Basic Operations
- Loops, control flow
- Data structures: Lists, vectors, matrices, data frames
- Functions
- Inputs and outputs
- String manipulation
- Date/time manipulation
- Basic data visualization
- Special packages: ggplot2, dplyr

## Required Activities

There is no required activity.

## About the Instructor

Paul Rognon is a PhD Student in Statistics and Machine Learning at Universitat Pompeu Fabra and Universitat Politècnica de Catalunya. His background is in applied mathematics and mathematical statistics. His research currently focuses on graphical and high-dimensional statistical theory. He has previously worked on applications ranging from financial risk, functional genomics, text mining, and extreme weather risk.

## List of References

There is a bewildering number of online Python and R tutorials and books. Here are some recommendations:

### 1. Python

Online resources:

- [codecademy Python course](#): interactive tutorials for Python.
- Software Carpentry Python lessons: [Programming with Python](#) and [Plotting with Python](#). They are a bit more advanced than the code academy pages, and not interactive.

Book:

- [Python Data Science Handbook](#) by Jake Van der Plas: Python programming oriented to data science with an online version and Colab notebooks.

### 2. R

Online resources:

- [codecademy R course](#): interactive tutorials for R.
- Software Carpentry R lessons: [Programming with R](#) and [R for Reproducible Scientific Analysis](#).

Book:

- [R for Data Science](#) by R guru Hadley Wickham and Garrett Grolemund: seen by many as the R bible. It has an online version. Second edition to be published in 2023.