

# DS Brush-up Courses - Introduction to Computing

Lecturer: *Hrvoje Stojic*, Universitat Pompeu Fabra

## Overview and Objectives

The brush up presents the computing infrastructure that will be used throughout the master - focusing on Unix, computing on servers and cloud (Amazon web services). We will provide a more thorough overview of R which will be the main pillar of the master program. Python is requirement for some of the courses and we will provide an introduction to Python as well, but on more basic level.

## Course Outline

### 1. Computing Infrastructure & Introduction to Unix

- Intro to the computing environment
- Linux and introduction to shell
- Cloud computing with Amazon Web Services

### 2. R basics & Programming in R

- Installing R and RStudio IDE
- Getting help
- Installing & loading packages
- Data types
- Control structures (if, for, while)
- Functions
- Debugging

### 3. Data Manipulation & Descriptive Statistics

- Data import/export
- Data display
- Subsetting, Merging, Reshaping data, Variable transformations
- Text manipulation
- Basic statistics
- Describing and simulating probability distributions
- Basic graphics

## DS Brush-up Courses - Introduction to Computing

### 4. Statistics and more advanced topics

- Linear regression
- Data wrangling with dplyr and tidyr
- Advanced graphics with ggplot2
- Reproducibility and R-markdown

### 5. Introduction to Python

- iPython interpreter
- Variable assignment
- Lists, dictionaries, and tuples
- Control statements: if, for, while loops; list comprehensions
- Functions
- String methods
- Basic reading from and writing to file
- Static HTML web scraping with Beautiful Soup

## Evaluation

The evaluation of the course is based on assignments in which students will be required to implement the procedures illustrated in class.

## Materials

The resources listed here are not strictly used to deliver the brush up course on programming or any other course in the master program. This is a list of recommended resources that you can use on your own before starting with the master program, to get familiar with the Unix environment and programming languages that will be used the most in the program. Of course, there are many more good resources on the Internet and feel free to use any of them to prepare yourself for this course and the master program.

## Unix

- Online command line/shell tutorial - <http://www.learnshell.org/>
- Another command line/shell tutorial - <http://cli.learncodethehardway.org/book/>

## DS Brush-up Courses -

# Introduction to Computing

### R resources

- Adler, Joseph (2010). \*R in a Nutshell\*. O'Reilly (Basics of R)
- Matloff, Norman (2011). \*The Art of R Programming: A Tour of Statistical Software Design\*. William Pollock. (Covers basics but also some advanced material, very useful for getting deeper insight into how R works)
- Wickham, Hadley (2014). [\*Advanced R\*](<http://adv-r.had.co.nz/>). Chapman & Hall/CRC The R Series. (Advanced material, suitable if you already have a decent working knowledge of R)
- Online R cookbook - <http://www.statmethods.net/> (Quick recipes for various types of analysis)

### Python resources

- Learn python the hard way - <http://learnpythonthehardway.org/book/> (Great online resource for learning Python through exercises)
- McKinney, Wes (2012). \*Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython\*. O'Reilly