

DS Brush-up Courses - Introduction to Computing

Lecturer: *Hrvoje Stojic*, University College London

Overview and Objectives

The course introduces the computing environment that will be used throughout the master - focusing on Unix, programming with Bash, computing on servers, version control with Git(hub) and R. We will provide particularly thorough overview of R which will be the main pillar of the master program.

Course Outline

1. Introduction to Unix & Programming with Bash

- Introduction to Unix
- Introduction to Bash (shell/terminal/command line)
- SSH and connecting to servers
- Version control with Git

2. R basics & Programming in R

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

3. Data Manipulation & Descriptive Statistics

- Data import/export
- Date and time
- Data display
- Basic statistics
- Graphics with ggplot2
- Text manipulation

4. Statistics and more advanced topics

- Data manipulation
- The apply family of functions

DS Brush-up Courses - Introduction to Computing

- Data wrangling with dplyr and tidyr
- Reproducibility and R-markdown
- Debugging
- Accessing R from terminal

This is the list of topics that we will cover, roughly in that order. You will receive handouts covering basic commands explicitly within these topics and will be expected to go through them before the classes. I believe that programming is best learnt by doing. Hence, in the classes, instead of showing you all the commands (for example in R) you will learn by doing, by solving practical problems using any method you want. I will be there to guide you through the process, answer any question and speed up the learning process.

Evaluation

The evaluation for the course is based on a final exam.

Hardware & Software

I expect you to come with a laptop on which you will implement the computing assignments that will be interspersed throughout the lectures. You should come with Unix based operating system (OS). A large portion of Data science tools are designed for Unix/Linux systems and working in such an environment is significantly easier. Moreover, connecting to servers and cloud computing platforms, as well as working with them, is much easier from a Unix based OS. Mac OS or any Linux OS would do, but I recommend Ubuntu based Linux distribution, since in that case I could directly help you if you encounter problems. Note that you can install Linux OS as dual-boot, alongside Windows (you can find many tutorials on the Internet on how to do this). If you have particularly strong machine, a virtual machine with a Linux operating system would work as well.

You should also install the following programs before the course starts. They are very common and it should be easy to find information on Internet.

- R (<https://www.r-project.org/>)
- Rstudio - IDE for R (<https://www.rstudio.com/>)
- Git (<https://git-scm.com/>)

Materials

The resources listed here are not strictly used to deliver the course. 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

DS Brush-up Courses - Introduction to Computing

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/>

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)