R Statistical Software - An Introduction for Econ 19

UCLA - Econ 19 - Fall 2018

François Geerolf

The only thing you need to do before class is to download R statistical software (that is, point 1 and point 2 below). I will guide you through all the subsequent steps.

Getting you started with R Statistical Software

Downloading. You need to install `R` and `Rstudio`:

1. First you must get the **R statistical software**, which you may download on the UCLA website here. The latest release (2018-07-02, Feather Spray) is version 3.5.1. For Mac OSX: download here. For Windows: download here.
2. Second, I recommend you use a Graphical User Interface (GNU) for R such as **R Studio**. R Studio's latest release is 1.1.456: download here.

Introduction to R. Cheatsheets are a great way to get started on R. Many are available here, but the 2 main cheatsheets are:

- Base R Cheatsheet.
- Advanced R Cheatsheet

Packages

We are going to use `tidyverse`, from Hadley Wickham, for data manipulation as well as plotting data. This cheatsheet has a beginner's introduction to `tidyverse`. `tidyverse` is a powerful collection of R packages that are data tools for transforming and visualizing data. Datacamp has a free tutorial for `tidyverse`, which can get you started. The following packages are particularly useful:

- `dplyr` for data manipulation. Cheatsheet. Note, in particular, the use of pipes %>%:
  - x %>% f(y) is the same as f(x, y).
  - y %>% f(x, ., z) is the same as f(x, y, z).
  - "Piping" with %>% makes code more readable
- `ggplot2` for data visualization. Cheatsheet.
- `stringr` for string manipulation. Cheatsheet. Cheatsheet on Regular Expressions.

In addition to the `tidyverse` collection of R packages, we should also use the following packages:

- `rvest` for web scrapping, which has a tutorial here.

R-markdown

This document was created using `R-markdown`, which you can learn using this cheatsheet, as well as this reference guide.