

Sport Analytics

Carlo A. Favero,¹

¹Bocconi University & CEPR

20630 Lecture 2

Going to the data with R

We will use R and run R through the user interface RStudio. When using R and studio Chat GPT can be used as an online assistant.

- What can R do for you ?
- A statistical package is needed to take all the modelling steps in practice. Going from loading the data, to transforming them, graphing them, analyzing them, modelling them and using them to estimate parameters and to generate model based simulations and, eventually, to producing a report.

R and RStudio

- We will use R and run R through the user interface RStudio.
- When using R and studio Chat GPT can be used as an online assistant.
- RStudio is an integrated development environment (IDE) for the R programming language. It provides a user-friendly and feature-rich environment for working with R, making it easier for data scientists, statisticians, and researchers to write, run, and manage R code.

RStudio

- **Script Editor:** RStudio includes a script editor with features like syntax highlighting, code completion, and code formatting to enhance your coding experience.
- **Console:** It has an interactive R console that allows you to run R code and see the results in real-time. You can also save your command history for future reference.
- **Data Viewer:** RStudio provides a data viewer that lets you examine and manipulate data frames and other data structures visually.
- **Plots and Visualization:** You can create and customize plots and charts using R's graphical capabilities.
- **Package Management:** It simplifies the management of R packages, allowing you to install, update, and load packages with ease.
- **Integrated Help:** RStudio integrates with R's documentation and help system.
- **Version Control:** It offers version control integration with Git and SVN, allowing you to manage your R projects and code revisions efficiently.
- **Project Management:** RStudio supports project organization, making it easy to keep your code, data, and files organized within a project folder.
- **R Markdown:** RStudio supports R Markdown, a format for creating dynamic documents that combine R code, text, and visualizations.
- **Shiny:** RStudio includes Shiny, a web application framework for building interactive web applications with R.

Learning R

The best approach is learning by doing, there are many good online tutorials. However, Chat GPT is probably the best assistant.

- <https://data-flair.training/blogs/r-list-tutorial/>
- <https://support.rstudio.com/hc/en-us/articles/200552336-Getting-Help-with-R>
- <https://www.r-bloggers.com/getting-help-in-r/>
- <https://bookdown.org/ndphillips/YaRrr/projects-in-rstudio.html>

Ask Chat GPT

Chat GPT can assist you in using R with RStudio in various ways, including:

- **Coding and Syntax Assistance:** I can provide explanations, examples, and code snippets to help you understand and write R code in RStudio. If you encounter errors or have questions about specific functions or packages, feel free to ask for guidance.
- **Data Manipulation and Analysis:** I can assist you in performing data manipulation, cleaning, and analysis tasks in RStudio.
- **Package Recommendations:** If you're looking for R packages to solve specific tasks or problems, I can suggest relevant packages and explain how to install and use them.
- **R Markdown and Reports:** I can assist you in creating dynamic reports using R Markdown in RStudio.
- **Plotting and Visualization:** I can help you create various types of plots and visualizations in RStudio using packages like ggplot2, base R graphics, and others. I can also explain how to customize and export your plots.
- **Debugging and Troubleshooting:** If you encounter errors or bugs in your R code within RStudio, I can help you understand and debug the issues.
- **Version Control with Git:** I can guide you through the process of using Git for version control within RStudio.
- **Project Organization:** I can provide advice on organizing your R projects within RStudio.
- **Customization and Configuration:** I can help you customize your RStudio environment.
- **Tips and Best Practices:** I can offer tips and best practices for efficient coding, data analysis, and project management.

Beyond coding with R

- Accessing Application Programme Interfaces (API) and running them to webscrape data with R
- creating web applications (Rshiny)
- incorporating programmes into documents (R Markdown)

Set-up

- Using Section 1-2 of Torfs and Bauer Install R and R studio on your computer and learn how to run them
- The R-studio layout
- Your first code upload **intro.R**
- clear the memory
- set-up working directory (the one where your code is)
- Learn what a package (library) is and how to install it
- read Torfs and Bauer and go through all the "ToDo" points (exercise 1).

Packages in R

R packages are a collection of R functions, compiled code and sample data.

They are stored under a directory called "library" in the R environment. By default, R installs a set of packages during installation.

```
# load packages  
listofpackages =  
c("ellipse", "reshape2", "ggplot2", "dygraphs", "plyr", "dplyr")  
for (j in listofpackages){  
  if(sum(installed.packages()[, 1] == j) == 0) {  
    install.packages(j)  
  }  
  library(j, character.only = T)  
}
```

Fundamentals in R

- data-objects
- data-handling
- getting data from the web with R
- programming and Control Flow
- modelling
- Dynamic Documents

Data-Objects in R

When working with data in R the first step is to know how to identify them and organize them.

- R makes available various common **datatypes**: double, integer, complex, logical, character, factor, date and time, and missing data.
- Data are then organized into datasets that are called **data structures** .
- The basic data structures we shall use are: vector, matrix, array, dataframes, tibble, time-series, and lists
- code **data_ob.R**

Data-Handling in R

- code **dH.R** (must be placed in the same directory with associated data files)
- Importing and Exporting Data from different formats
- setting-up dataframes, selecting data, transforming data
- subsetting data, omitting na
- data grouping (exercise 2)

Getting data from the web with R

- code **download.R**
- getting data from <https://finance.yahoo.com/> using `quantmod`
- getting data from <https://fred.stlouisfed.org/> using `quantmod` (in alternative you can use `fredR`)
- getting data from an URL, the case of COVID data for Italian Regions
- getting data using API's
 - <https://www.r-bloggers.com/2015/11/accessing-apis-from-r-and-a-little-r-programming/>
 - <https://git.nomics.world/dbnomics/rdbnomics>
 - webscraping

Programming and Control Flow in R

- code **prog.R**
- if-else statements
- using switch
- loops
- functions in R

Running a model with R

- **Problem** : What is the probability that a Great NBA Team (Win% = 85%) Looses Two Consecutive Games at Some Point in the Season ?
- **Solution**: Simulate the binomial distribution

Dynamic Documents in R

- an introduction to R Markdown to run R and create a Report with one file
- a self-explanatory illustrative code **example.rmd**