Introduction to R: R Miscellanea

Cristian Huse

University of Oldenburg

Overview

- The aim of this topic is to show you how **R** is much more than a fancy calculator
- We will discuss:
  - How to organize a project
  - Different data types, while focusing on **data.frame**
  - How to read, manage, subset, and process data
  - How to use fancier graphs through one **ggplot** example
  - Some useful operators, e.g., **%>%**, |>, **::**
  - ...
- The aim is to give you a broader view of what R can do, but this is still a limited and biased account

Lab Session: Misc R

Structure of the Session

- Project organization
- "Hello, World!" plus the pipe operator (**%>%**, from package magrittr)
  - Note: R 4.1 introduces a new pipe operator |> as part of base R (may discuss details later)
- Data structures and data.frame
- Loading data
  - **read.csv**, subsetting, data overview
- Bonus: A plot using **ggplot**
- Some data work: **data.table**
  - Reading large data sets plus the double colon operator (**::**)
  - Overview of data.table operations
- Appendix: Reading data from a webpage

Project Organization

- When starting a project, create a new folder/directory which bears the project name, e.g., **myproject1**
  - (avoid using empty spaces, " ", in naming your files and directories)

- In that folder, create the following sub-folders:
  - **data_raw** (or **Data_raw**): for the raw data
  - **data** (or **Data**): for the clean/processed data
  - **results** (or **Results**): for the output created, e.g., tables
  - **figs** (or **Figs**): for the plots/figures created

- A better/more advanced solution for seamless work: **RStudio Projects**
  - Using RStudio Projects
  - RStudio Projects

Instructions for the Lab Session Misc R

- Please load the file **IntroR_MiscR.R**
- Go through the file line-by-line, consulting the help whenever needed – this is your exercise for next week (book 2-3 hours)

Take-aways

- The aim of this slide deck is to provide you a broader view of what R can do in terms of reading, processing, and summarizing data;
- Is is fundamental that you run in detail the associated .R files and make sure you understand the commands used;
- **Just to make clear:** You are not expected to **memorize** the contents of the session, but to able to use them within a project – familiarity will come with use!

Take-aways (2)

Repeating myself:

- From my own experience, the best way to learn is to get your hands dirty with data:
  - Go through the files in detail
  - Take something you know and have done before and re-do the project using a new language
  - There are countless channels, tutorials, books, and communities, e.g., Stack Overflow

- As in everything, the contents and the approach pursued here are biased, incomplete, and reflect (my) personal taste

Selected References

- An Introduction to R
- R Data Import/Export
- Wickham & Grolemund's R for Data Science
- data.table official webpage
  - Check the nice vignettes
- data.table cheatsheet

Appendix

Setup: Package Management

- Installing packages one-by-one can become tedious, but one can use **pacman**, a package management tool to automatize the process:

```
## Load and install today's packages
if (!require("pacman")) install.packages("pacman")
pacman::p_load(mfx, tidyverse, hrbrthemes, estimatr,
               ivreg, fixest, sandwich, lmtest,
               margins, vtable, broom, modelsummary,
               data.table,fastverse)
## Make sure we have at least version 0.6.0 of ivreg
if (numeric_version(packageVersion("ivreg")) <
    numeric_version("0.6.0"))
  install.packages("ivreg")
## Optional -- ggplot2 plotting theme
theme_set(hrbrthemes::theme_ipsum())
```