

Take-home exam

Introduction to R

Mattia Guerini

Academic Year: 2020/2021 (Fall Semester)

Details

Aim: the aim of the homework is that of gaining independence in problem solving (in this case: presenting a simple report with basic statistical evidence).

Output: you should prepare the R script (SURNAMES.R) with the complete sequence of R commands and a pdf (SURNAMES.PDF) reporting the relevant graph(s), the table(s) and your comments.

Deadline: 30 November 2020.

Groups: the exam shall be executed in groups of two students. If the total number of students is odd, one group will be formed by three members.

Delivery method: email me the two output files at mattia.guerini@sciencespo.fr.

Tasks:

- Look at the available datasets on the Eurostat webpage: <https://ec.europa.eu/eurostat/data/database>.
- Select one or two (at most) variables that you would like to study (look at the variable code as done in class).
- Use the library `eurostat` to load the chosen dataset(s) into R.
- Clean the data: remove the NA, select few specific countries, focus on some specific years of interest, etc. (use the library `tidyverse`).
- Plot a couple of graphs (use the library `ggplot2`) about the variable of interest and include at least one table.¹
- Write a short report (max. 4 pages) about the statistical evidence you find. The report should be composed of:
 1. *introduction* where you describe why you think the analysis can be of potential interest;
 2. *descriptive statistics* where you present and discuss the graphs and tables you have produced;
 3. *conclusions* where you summarize the main results, and discuss the limitations of your work and how you can possibly solve them.

¹To export the graphs you can click on 'export' after a graph appears on 'R Studio'.