Basic data management in R

Exercises

Antonio Gasparrini

01 November 2023

Sections 1-2

1. Load the dataset AIRQUALITY stored in the DATASET package. Check the names of the variables. Change the names using the same labels but with lowercase letters (*hint*: use the function `tolower()`).
2. Load the dataset MTCARS stored in the DATASET package. Subset the dataset to the records of cars with manual transmission and weight higher than 2,500 pounds (see `help(mtcars)`), keeping the information on miles per gallon, number of cylinders, and displacement.
3. Extract the information on miles per gallon and gross horsepower in MTCARS ordered by the number of forward gears and the number of carburettors (see `help(mtcars)`), keeping just the relevant variables.
4. The text file DEPRIVATION (included in the folder FILES) contains some data about a deprivation score by age and alcohol consumption categories. Import the data into the session and merge this information with the data frame ESOPH in the DATASET package, creating a new data frame ESOPH2.

Sections 3-5

5. Aggregate the dataset BIRTHS in the package EPI (note: you might have to install and load it first) by sex of the baby and hypertension of the mother, computing for each category the total number of observations and the average birth weight (*hint*: produce two different aggregated data frames and merge them).
6. Repeat part of the previous exercises using the dataset MTCARS. Specifically, subset the records of cars with weight less than 3,000 pounds, aggregate computing the average of miles per gallon and number of cylinders by engine and transmission types, and finally extract the variable miles per gallon. Try using a single expression and then using a sequence of expressions with pipes.
7. Transform the following numeric variables of `births` in factors: low birth weight (yes/no), preterm (<37 weeks and >=37 weeks), hypertension (yes/no) and sex (male/female). Be careful about the original coding. Inspect the first 5 observations.
8. Create a new categorical variable `mpggp` in `mtcars` by categorising `mpg` with cut-offs placed at 20, and 30 miles per gallon. Then redefine `gear` as a categorical variable by transforming it into a factor, and then change the categories to '3' and '4 and more'. Summarize both variables.