Data Handling: Import, Cleaning and Visualisation

Exercise to lecture 4: csv and arrays

Dr. Aurélien Sallin

Working with a data frame

Import data

Have a look at the file `financial_data.txt` using your favorite text editor. What do you notice?

Import the table using the `read.csv()` function in your environment. Make sure you have the right path to access the .txt document. What does this parser do? Explore the data.frame. What is its structure? What are its dimensions?

Compute the summary statistics for each variable using the `summary()` command. What does this command give you? What do you notice? Make the necessary changes.

Variable creation

Create a new variable "costs", which is the revenue - profit. [There are many ways to create a variable in a data frame. Here, use the `$` index.]

Factor variable

Which variable is (should be) a factor? Recode this variable as a factor. What are the levels? Should we have the variable `Firm` as a factor?

Nests

Split your data using the factor variable into three data frames that are contained in a list. Compute the mean profit for each data frame. Hint: use the function `split`.

Advanced: map (not exam relevant)

Do the same as the exercise above using the `map` function. Install the packages `tidyr`, `dplyr`, and `purrr`.