

## Exercise 3

### Data exploration

#### Explore the temperature dataset

- i) Load the dataset `temperature.csv` from the `01_Data` folder and assign it to an object with a meaningful name (e.g. `temperature`)
- ii) Get an overview of the dataset:
  - View the first six rows of the dataset
  - How many rows does the dataset have?
  - How many columns does the dataset have?
  - What class do the columns have? Can you guess?
- iii) Calculate the mean temperature

#### *For those who have time left...*

- iv) In which months were the measurements taken?
- v) What month and day was the maximum temperature measured?
- vi) Load the internal dataset `airquality` and calculate the Pearson correlation between `Wind` and `Temp`. Do you expect a positive or negative correlation?

## Hints

- i) Load the dataset `temperature.csv` from the `01_Data` folder and assign it to an object with a meaningful name (e.g. `temp`)
  - Set the working directory to the `01_Data` folder in the course material (`?setwd` or look at the video tutorial in the `02_Slides_tutorials` folder)
  - Load the dataset with the function `read.csv` (`?read.csv`)
- ii) Get an overview of the dataset:
  - First six rows: Check the function `head` (`?head`)
  - Number of rows: Check the functions `nrow` (`?nrow`) or `dim` (`?dim`)
  - Number of columns: Check the functions `ncol` (`?ncol`) or `dim` (`?dim`)
  - Column classes: Check the function `class` (`?class`)
  - *All in one*: Check the function `str` (`?str`)
- iii) Check the function `mean` (`?mean`).  
 Select the column `temp` (in the form of `df$temp` or `df[, 2]`)
- iv) Check the function `unique` (`?unique`)
- v) Check the function `which.max` (`?which.max`)
- vi) Load the dataset with `data(airquality)`  
 Check the function `cor` (`?cor`) and the argument `method`