

Exercise 4

Data manipulation

Manipulate the `temperature` dataset

- i) Load the dataset `temperature.csv` from the `01_Data` folder and assign it to an object with a meaningful name
- ii) Filter the dataset for the site `Bern`
- iii) Add a column with the variable `year` (the year is 2013)

For those who have time left...

- iv) Create a new `date` column
 - Create a new column that is a combination of the variables `year`, `month` and `day` (in the form of "2013-01-25")
 - Convert the `class` of the column from `"character"` to `"Date"`
 - Calculate the number of days between the first and last measurement
- v) Calculate the average temperature for periods without frost (i.e. the temperature is above 0 °C) for the site `Zurich`
- vi) Load the internal dataset `airquality` and change the column names to lower case

Hints

- i) Load the dataset `temperature.csv` from the `01_Data` folder and assign it to an object with a meaningful name
 - Set the working directory to the `01_Data` folder in the course material (`?setwd` or look at the slides in the `02_Slides` folder)
 - Load the dataset with the function `read.csv` (`?read.csv`)
- ii) Check the function `subset` (`?subset`)
- iii) Use a command in the form of `df$year` and assign the value 2013 to this column.
- iv) Create a new `date` column
 - Combine the columns `year`, `month` and `day`:
Check the function `paste` (`?paste`) and the argument `sep` inside the function `paste`
 - Add this vector to the dataset and name the column `date`
 - Check the function `as.Date` (`?as.Date`) to change the class of the column from `"character"` to `"Date"`
 - Calculate the duration of the measurements:
Calculations are possible with variables of class `"Date"`
- v) Calculate the average temperature for periods without frost (i.e. the temperature is above 0 °C) for the site `Zurich`
 - Filter the dataset with the function `subset` (`?subset`)
 - Calculate the average temperature with the function `mean` (`?mean`)
- vi) Change the column names to lower case
 - Use the function `colnames` (`?colnames`) to change column names
→ See file `02_Vector_data_frame_list.R` in the `03_Scripts` folder for an example
 - Try to do it manually and also try to find a function that does this task (e.g. by asking Google)