Exercise 2

Data exploration

Explore the `airquality` dataset

i) Load the internal dataset `airquality`

ii) Assign a new name to the dataset (e.g. `airquality_1`)

iii) Get an overview 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?

iv) Calculate the mean temperature

v) What is the maximum measured ozone content?

For those who have time left...

vi) In which month and day was the maximum ozone content measured?

vii) Calculate the mean temperature for the month May

Hints

i) Use the command `data(airquality)` to load the dataset.

ii) Assign a new name by using an expression similar to: `new_name <- airquality`

iii) Get an overview of the dataset:

- Number of rows: Check the functions `dim` (`?dim`) or `nrow` (`?nrow`)
- Number of columns: Check the functions `dim` (`?dim`) or `ncol` (`?ncol`)
- Column classes: Check the function `class` (`?class`)

iv) Check the function `mean` (`?mean`). Select the column `Temp` (in the form of `df$Temp` or `df[, 4]`)

v) Check the function `max` (`?max`) and the argument `na.rm`

vi) Check the function `which.max`

vii) Use the function `subset` (`?subset`) to filter for the month May (`Month == 5`)