

Exercises for the course
“An introduction to R”

Exercise session Reading and writing data: Thursday - March 5, 2020

In this exercise session, you will analyze data on different measurements for two sparrow species. The individual tasks are split into six exercises. You should save all your analysis in a single script. Make your life easier and document the code in your script properly.

Exercise 1: *Import sparrow morphology data*

- Open RStudio, create a new script (**File** → **New File** → **R script**) and save your script, for example, as `sheet_6.R`. If you do not use RStudio, you can write and save the script with any editor as discussed in the lecture.
- Download the data file `sparrows.csv` from the webpage. This file contains wing, tarsus, head and bill measurements of two sparrow species. Use `read.table(file = "sparrows.csv", header = TRUE)` to import the data into an object named `sparrows`. Afterwards, use the `View()` function to display the data.
- Something must have gone wrong! All variables are combined in one big column. Can you figure out why R does not import the data set correctly? Write a short answer. Hint: Consult the help page `?read.table` and find out the default for the argument `'sep'`. You have to specify this argument correctly.
- Solve this issue by specifying the argument correctly and import the sparrow data again. Use `View()` to check if the import worked correctly.

Exercise 2: *Get an overview of the data*

- Get an overview of the sparrow data with the command `str()`.
- Return the minimum, median, mean and maximum for tarsus and bill measurements. Hint: You may use a single function to perform this task.

Exercise 3: *Checking and cleaning data frames*

- During the data entry, three rows have been entered twice. Which are these duplicate rows? Remove the duplicate rows from the data frame. Hint: It might be faster if you incorporate the function `which()`.
- Display the levels of the factor `sex`. Correct the typos by using `which()` and logical operators, such that `sex` contains only the levels `'Male'` and `'Female'`. Remove all other extra levels.

Exercise 4: *Missing values*

- Find out which rows in the variable `wing` contains NAs.
- Replace all NAs with the values 59, 56.5, and 57 (in this order). Use `which(is.na())` to check if your replacement worked.

Exercise 5: *Sorting data frames*

- Sort the data frame by the `wing` and `head` column, and create a new object `sparrows_sorted`.
- Show the individuals with the six longest bills. Hint: You may use the functions `head()` and `order()` in the same command line.

Exercise 6: *Export data*

- Save the ordered table `sparrows_sorted` into a comma separated file named `sparrows_sorted.csv`.
- Save your script and remove all objects from the working environment.