

Exercises for the course
“An introduction to R”

Exercise session Reading and writing data: Friday - March 9, 2018

Exercise 1: *Reshaping data frames.*

- a) Download the two data files `Catfish.csv` and `Treatment.csv` from the web page and import them into R. `Catfish.csv` contains weight measurements (variables `March` and `April`) of two catfish species that are commonly used in aquaculture. Both species have been reared under different temperature and food treatments, which are specified in `Treatment.csv`. Get an overview of the structure of the data frames using `str()` and `head()`.
- b) Convert the data frame into a long format (`Catfish_long`), so that weight measurements for `March` and `April` are combined in one column named `Weight`. Use the `gather()` function from the `tidyr` package.

Try also to convert the data frame back into its original format with separate columns for `March` and `April`.

Exercise 2: *Adding new variables.*

- a) Add a new column `sqrtWeight` to the data frame, which displays the square root of the weight measurements.
- b) You would like to merge the two columns `Genus` and `Species` into one variable called `Species`. The result should look like this: **`Silirus.glanis`**. Fortunately, the `dplyr` package provides functions to split and combine columns.

Exercise 3: *Combining data frames.*

The treatment data has been saved in a different file. The `dplyr` package has many convenient functions to join the temperature and food treatment data with the weight measurements (`Catfish_Treatment`). Try the three functions `left_join()`, `right_join`, and `inner_join()`.

Exercise 4: *Subsetting data.*

- a) Use `subset()` and logical operators to make a new data frame that just contains females with a weight over 72g.
- b) Make another subset for males, which contains all males that have a weight under or equal to 20g or over 70g.
- c) Take a random subset that contains 65% of the data.

Exercise 5: *Summarising data with dplyr.*

- a) Use the data frame `Catfish_Treatment` and the `summarise()` function from `dplyr` to calculate the mean, the standard deviation, and the variance of weight for all individuals.
- b) Get the mean, the standard deviation, and the sample size for each food treatment separately.

HINT: You have to group your data first before you calculate summaries