Task 2. R in practice

Research Methods Data and statistics with R

Jüri Lillemets

2019-09-05

Get data

For the following exercises use a data set from `agridat` package. Use the following commands to install the package on your computer.

```
install.packages('agridat')
library('agridat')
```

When the package is loaded you can access all the data sets included in the package. Store a dataset named `australia.soybean` in an object with shorter name (e.g. `soyb`) for convenience. Type `?australia.soybean` for details on the data.

```
soyb <- australia.soybean
```

Get an overview of the data

Use functions `str()`, `summary()`, `head()` to get an understanding of the data frame.

Create a new variable

The data contains yield of soybeans (tonnes/hectare, `yield`) and percentage of oil in it (`oil`). Create a new variable containing *tonnes of oil per hectare* and name it whatever you want. You need to multiply `yield` by `oil`.

Convert data to wide format

Use variable `year` to store variable `protein` in new variables `1970` and `1971`. You need to use the command `spread()` from `tidyr` package.

Merge new rows with the data frame

Create a new data frame with some values for each location.

```
soybLoc <- data.frame(loc = unique(soyb$loc), price = c(21, 34, 55, 89))
```

Next, join this new data frame with data in `soyb`. Use `merge()` function or `left_join()` in `dplyr` package.

Aggregate the data

Calculate mean yield (`yield`) for each location (`loc`) using `aggregate()` function.

Remove duplicates

Remove all rows containing duplicated values of locations (`loc`).