Looking at the ecological dataset

David Zelený

Before you start to do some analysis or visualization of a dataset, you usually need to explore it and to be able to access individual dataset parts. Here, we will upload a rather complex ecological dataset into R, observe it's structure, search within it certain values, and report them.

Upload the dataset from the Internet

The dataset is called `vltava` and is represented by a large *list* of components. To upload it into the R workspace, use the following code (you need to be connected to the Internet):

```
load (url ('https://github.com/zdealveindy/anadat-r/raw/master/data/vltava.RData'))
```

The `load` function loads an R object into R workspace; `url` function is used to wrap URL (website) link in case you want to download data directly from the internet. Note that the `vltava.RData` is an R object, not an R code (see *Importing data into R*, section Import .RData file for details).

Structure of the dataset

The object `vltava` appeared in your *R Global environment* (the workspace you are currently working in R), and you can use it for further work. The object is a list with five components (use `names (vltava)` to check their names). For this assignment, we will need to inspect only components `spe`, `eiv` and `env` (each of them is a `data.frame`):

- `spe` contains percentage covers of plant species (in columns) recorded in sampled sites (in rows); for example, species `AcerPla32` (second column, *Acer platanoides*, maple) is absent in the first five sites (value zero in the first five rows) and has percentage cover 49 percent in site 6;
- `eiv` contains so-called *Ellenberg indicator values* (in the columns) for individual species (in the rows); for example for `light` (second column, reflecting species ecological optima along light gradient), the species `AcerPla32` has the value 4 (on a scale from 1 to 9, i.e. it is not shade tolerant neither light demanding);
- `env` contains environmental descriptors (in columns) of individual sites (in rows); for example the `SLOPE` (second column) of site 3 (third row) is 30 degrees.

Within the dataset, search for the following information

1. What is the highest value of pH measured in the dataset? (`vltava` > `env` > `pH`)
2. What is the sum of species covers in the plot 33? (`vltava` > `spe` > row 33)
3. What is the mean soil depth? (`vltava` > `env` > `SOILDPT`)
4. How many missing values for species are there in Ellenberg indicator values for soil reaction? (`vltava` > `eiv` > `react`)
5. How many species are in the plot 64? (`vltava` > `spe` > sum the values in row 64 which are higher than 0)

Report the results

Please write exercise as an R code, save it as *.r file and upload into the NTU COOL. The code should be reproducible (see here what it means). Make sure that the first line of the code is the header line with your

name, student ID and homework number. The second line should load the data (copy the code above for it). From the third line of the code, you answer questions 1 to 5, with each answer represented by only one line of code. After the hashtag (#) at the same line, you write the correct value.

An example R code, at which the first question would be to find the minimum value of elevation (vltava > env > ELEVATION):

```
# REcol Homework 1, Author: David Zeleny, Student ID: p12345678
load (url (''https://github.com/zdealveindy/anadat-r/raw/master/data/vltava.RData''))
min (vltava$env$ELEVATION) # 405
... (here come answers to other questions) ...
```

Good luck!

Hints (try first to solve the exercise without using them)

1. `max`;
2. `sum`, subsetting by square brackets `[,]`;
3. `mean`;
4. `is.na` applied on the vector will return a vector of logical values, in which `TRUE` indicates missing value, while `FALSE` any existing value; if you apply `sum` on the vector of logical values, `TRUE` converts into `1` and `FALSE` into `0`;
5. Logical function *larger than zero* (`> 0`) applied on vector will turn the vector of species abundances into logical values (`TRUE` for values larger than 0). You can then `sum` the vector.