Exercises

Vectors

1) The pH measured in 10 sites of a lake were analyzed and the values obtained were: 6.6 – 7 – 6.5 – 7.2 – 8 – 5.9 – 7.6 – 6.6 – 8.1 – 7.5.

a) Create an object that contains the values above and call it pH. Using the functions that we learned, calculate the mean and standard deviation for these samples, and obtain the minimum and maximum values.

b) The value 5.9 is wrong due to a reading error of the equipment. Exclude this value from the vector that you created and re-calculate the mean, standard deviation, and the minimum and maximum values.

c) Sort the number from the smallest to the largest.

2) Two areas were selected for a vegetation study. The first area is called "Green Valley" and the second one "Cloud mountain". In each of these areas, 3 samples were collected.

a) Use the function "rep" to create an object called "areas" which contain the name "Green Valley" and "Cloud mountain" repeated three times each. Use the function "factor" to tell R that this vector is a category with two levels.

3. Calculate:

a) natural logarithm of 4

c) square root of 36;

d) create an object called "x" that contain the values: 3, 4, 2, 7, 5, 3, 9, 1. After that, raise the numbers to the power of 2.

4. Let's flip a coin 100 times and compute how many times each event (heads and teals) occur. This can be easily done in R! Follow the steps below.

a) create a coin (tip: create a vector with the elements "heads" and "tails")

b) create a sampling procedure (tip: don't forget to set replace=T and save the results in another object)

c) Count the number of each event (tip: you can use the function `sum` and logical symbol "=="). To confirm the results, use the function `table` (ex: `table (object name)`)

5. Create the objects

"a", which is composed of the numbers 2,3,4,5,1.

"b", which is composed of the numbers 0,1,5,3,6.

a) Test if any value of "a" is smaller or equal to 2 (tip: you can use the function "any")

b) Test if all values of "a" are larger than 3.

c) Test if the values of b are different from 5

6. Use the functions `union`, `intersect` e `setdiff` to find the union, intersect and the difference between the two vectors below.

x<-c(1,2,3,4,5)

y<-c(4,5,6,7)

Dataframes

1) Load the file "fish lakes" in R. This dataset contains information about the abundance of five fish species in 18 lakes. Lake area, the presence of human settlements, dissolved oxygen, conductivity, phosphate, and chlorophyll-a were also recorded.

a) Conduct a general summary in the dataset. What are the maximum values of dissolved oxygen and chlorophyll-a?

b) Calculate the average size of the lakes.

c) Calculate the average concentration of dissolved oxygen in lakes located in urban and non-urban areas (you can use either `tapply` or `aggregate`).

d) Select only the data collected in urban zones AND in lakes with an area larger than 100 ha.

e) Now, select the data collected in urban zones AND in lakes that have a fish abundance higher than 20 individuals. Use the function `rowSums` to calculate the abundance of fish in each lake.

f) Create a new objected based on the fish database that contain only the five species of fish and their correspondent abundance in each site. Call this object "spp". Create another object called "env" that contain the environmental variables (area, presence_of_humans, DO, conductivity, phosphate, and chlorophyll)

g) Calculate the total abundance and the average abundance of each species using the object "spp" that you created. Use the functions `colSums` e `colMeans`.

h) Transform the species abundance into presence-absence data (0-1) and save the object with the name "spp_pa".

i) Calculate the fish richness in each lake using the object "spp_pa". Use the function rowSums.

j) We saw in the last section (vectors) that `cbind` could be used to group two vectors together. Can we group the dataframes "env" and "spp" into a single database using the same function?

2) Load the file "biomass" in R. Rearrange this dataframe in order to create a site x species matrix for community data analysis. Create one matrix based on numerical abundance data and the other based on total biomass. (tip: use the function cast).