

# Exercices on session *Statistics with R session*

## 3/3

Load the packages *vegan* and *tidyverse*.

### Climate of the cities

In this exercise, you will study the climate of 49 big cities in the world. The data are in the *cities\_climate.csv* file available [here](#).

- Load the data as a tibble and look at their structure. These data have been obtained from WorldClim. They contain:
  - long and lat: the longitude and latitude
  - t\_mean: Mean annual temperature (°C)
  - t\_diu: Daily temperature variation, average difference between the maximum and minimum temperature in the same month (°C)
  - t\_sd: Seasonal temperature variation, standard deviation of the mean temperature between months (°C)
  - t\_max: Maximum temperature of the warmest month (°C)
  - t\_min: Minimum temperature of the coldest month (°C)
  - p\_ann: Annual precipitation (mm)
  - p\_max: Precipitation in the wettest month (mm)
  - p\_min: Precipitation of the driest month (mm)
  - p\_cv: Coefficient of variation of precipitation between months

We are going to do a PCA on these data.

- Remove the latitude and longitude

- Put the variable *city* as a row names and remove the corresponding column (to be able to see the name of the cities in the output of the analysis).
- Do the PCA (don't forget to standardise the variables)
- Look at the output: what is the cumulative proportion of variance explained by the two first axes?
- Do a screeplot with a broken stick distribution: how many axes do you want to interpret?
- Let's first look at similarity between cities: which scaling do we need for the biplot? Do the biplot and interpret it.
- Let's now look at correlations between climatic variables: do the corresponding biplot and interpret the axes.
- Let's now prepare the data for a clustering: standardise them and compute the distance matrix.
- Do the hierarchical clustering using the method of Ward.
- Plot the dendrogram.
- Make 3 groups and extract them.
- Represent the groups on the PCA biplot.

## For next time

Review what we have seen about logical operators in the session *Getting started with R*.