

Exercices on session *Tidyverse 1/2*

For all the following exercises, use functions from the tidyverse whenever possible, and try to make one pipe per question.

Karnataka forest dataset

You are going to work on data from forests of the central Western Ghats of India. These data were available in a previous version of the package BIOMASS (v 2.1.11), and are called KarnatakaForest. You can access these data in a csv file [here](#).

- Read the data using the function `read.csv` (don't forget to look at the metadata in the help available here [here](#)). Rename the data to facilitate the writing of the script.
- Get an overview of the data and check their class.
- Transform the data to a tibble.
- Transform all non-numeric variables to a factor.

Diameters

- Rename the column *D* to *diam*
- What is the smallest diameter of the whole dataset (make sure you extract the value, not only read it in a output)?
- What are the full species names (genus and species) of the 5 biggest trees?
- How many different species are amongst these five biggest trees?
- Create a table (tibble) with the median diameter per plot.
- What is the most common family in the whole dataset?

- Create a tibble with the number of trees per family and per plot for plots BSP1 and BSP2, and order it by decreasing number of individuals per family and per plot.
- To how many families do trees that have a diameter in the fourth quartile (*i.e.* in the interval between the 3rd and the 4th quartile, *i.e.* bigger than the 3rd quartile) of the whole dataset belong?

NB: quartile can refer both to the value (most common use), or to an interval.

For next time

- Install the following package from CRAN: *janitor*
- Make sure you remember what we saw on factors and logical operators during the session [Getting started with R](#)
- Download the *RData* file available [here](#) and store them in your raw data folder.