Exercices on session *Working with data*

Working with data from the *Inventário Florestal Nacional*

- Go to the website of the IFN (here) and download the data of the Rondônia for trees with a DBH 10cm (don't forget to look at the metadata!)
- Import these data into R (make sure you look at the data structure before)
- How many variables are there? how many observations?
- Look at the structure, everything is fine?
- Transform the following variable into factors: *Unidade amostral*, *estado de sanidade do fuste*, *qualidade do fuste*, *posição sociológica*, *hábito da planta* and check the structure again
- Transform the following variable into boleans: *árvore fora da floresta*, *presença ou ausência de lianas*, *indivíduo com ou sem coleta botânica* and check the structure again
- How many *Unidade amostral* is there?
- Add a colum with the DBH in meter
- What are the minimum, maximum and median height (*altura total*)
- Represent the distribution of DBH (in cm) using an histogram (*Bonus question for the foresters and forest ecologists: does this distribution make sense to you?*)
- How many trees are out of the forest?
- Which percentage of trees have liana?
- Make a graphical representation (barplot or Cleveland plot) of the occurrences of health levels (*estado de sanidade do fuste*)

Find the problems

We are working with a fake data set of vegetation survey in the DF, that contain the following variables:

- *site*: name of the site
- *plot*: plot ID (unique identifier)
- *height_max*: height of the tallest tree on the plot (in m)
- *DBH_max*: DBH (diameter at breast height) of the largest tree on the plot (in cm)

You can find this data set here.

There are 4 problems in this data frame, what are they?

For next time

- Install the following package from CRAN: *tidyverse*
- Download the following data sets here and here and store them in your raw data folder.
- Make sure you remember what we saw on factors and logical operators in the session Getting started with R