Introduction to R software
Session IV

Teaching Assitant: Manasa Shanta Yerramalla - manasa-shanta.yerramalla@inserm.fr
Coordination : Nolwenn Le Meur – nolwenn.lemeur@ehesp.fr

Objectives: The objectives of this lab are to merge dataset and start handling missing values.

A- Loading the transformed dataset

1. After launching RStudio and verify that you are in the right working directory using **getwd()** function call.

2. If you need to change the current working directory use either the drop-down menu or the **setwd()** function call with the appropriate path.

3. Open a new R script and save it as lab4.R using the drop-down menu.

4. As in lab 3, we will be using the SleepQualityData.Rdata object you created and saved in lab 2. This object should have been saved in your current working directory and you should be able to load the data using the **load()** function.

5. Load the data and verifying that it is consistent with what you expected

B- Merging dataset

In R, you can open more than one dataset at the same time. For instance, it allows to merge data from different sources that have a common identifier.
In this lab, we will add to the SleepQualityData information regarding the level of children's concentration at school.

1. Read a new dataset

1. Download from REAl the "SchoolConcentration.txt" _file and save in your current R working directory (where all your R lab scripts, objects and associated database are).
2. Read the dataset in R using one function of the read.csv() family functions (do not use the drop-down menu). Assign the name SchoolConcentration to that object.
3. Describe the new data using classical statistical parameter(s) and graphic(s)
4. Use the **merge()** function to append the SchoolConcentration data to the SleepQualityData dataset.

5. Save the data as a new R object

2. Handling missing values: an introduction

1. One variable has some missing values, which one?
2. Who are the individuals: retrieve their id, gender, and age? (hint have a look at the functions **which()** and **is.na()**)
3. Replace the missing values by the overall mean of the variable.
4. Save the new version of the SleepQualityData dataset in the SleepQualityData.Rdata file.

C- Is there an association between sleeping quality and school concentration?

1. We consider that a good sleeping quality is having a score equal to or above 6. Using the sleeping quality variable in its categorical form test whether sleeping quality influence school concentration. (remember to formulate the null and alternative hypothesis and verify the potential assumptions)
2. Compute the 95% confidence intervals of the school concentration levels for each sub-group of sleeping quality. Interpret the results.