Exercises R basics 1

1. Create a folder on your hard drive, which you will be using as the R working directory (if not already done so).
2. In RStudio, click Tools -> Global options -> tab General. Make sure R is pointing to the right folder to be used as working directory. If not, change the folder.
3. Have a look at the other options on the general tab, which are mainly about automatic saving and restoring files when opening or closing R.
   1. ‘Restore most recently opened project at startup’: opens the project at start up you were working in when you closed RStudio
   2. ‘Restore previously open source documents at startup’: automatically opens source documents, such as scripts and other files, at start up
   3. ‘Restore .RData into workspace at startup’: automatically loads your data into the environment at start up. RStudio prevents overwriting two different .RData files
   4. ‘Save workspace to .RData on exit’: if your data needs to be saved automatically to the .RData file on your hard drive
   5. ‘Always save history’: describes itself
4. Create your first project, called R seminars
5. Within the new project, create an R script called script
6. Make a vector called age, containing 5 decimal point numbers (choose the numbers yourself)
7. Modify the vector age in such a way it contains just the 1st, 2nd, 4th and 5th element. Store the result in a variable called age2
8. Calculate the minimum, maximum, mean and median of age
9. Coerce age to a character vector (but do not store the result)
10. Dichotomise age, by assigning observations below the mean the value 0, and observations above the mean the value 1. Note, you can do this manually since there are only 5 observations, but try to let R do this for you. Have a look at the documentation for the function ifelse() and make use of conditional operators < and >. Store the results in a variable called agedich.
11. Make a data frame called df, consisting of the following variables: variable id, containing a consecutive number from 1 to 5; variable age, containing the vector created in exercise 6; variable agedich containing the dichotomised sequence of exercise 10.
12. Coerce agedich in your data frame to a factor
13. Explore your data frame (column names, data classes, minimum values, means, etc.).
14. Set the following values in your data frame to missing (a missing value is denoted as NA in R):
    1. The observation in the 3rd row and 2nd column
    2. The entire 5th row
15. Calculate the mean of df$age. Why do you get an error? Look into the documentation of mean() how to solve this.
16. Return to the vector age, created in exercise 6. Add the first and second element. Multiply the third and fourth element. Do not store the results.
17. Once created, an object can be subsetted beyond its current dimensions. This is useful to insert items in already existing structures. For example, insert a random number of your liking in the 10th position of age and have a look at the result. What happened?
18. Make a named list, called my\_list containing the vector age and an item called data, containing the data frame df.
19. Display the value stored in the 2nd row, 3rd column of item data in list my\_list.
20. Find out what the function rnorm() does.
21. Create a vector x, containing 50 draws of a normal distribution with mean 5 and standard deviation 2.
22. Remove all objects created as part of these exercises. Use function rm().

**Turn to the dataset simulated data**

1. Save the dataset as a .csv file on your computer
2. Import the data file using read.csv() or read.csv2() (based on your computer settings with respect to decimal point and value separators. Read the R documentation of the abovementioned functions to familiarize yourself with the function you should use). Note that the full path towards the .csv file needs to be specified, using forward slashes (e.g. C:/My Documents/R working directory/simulated\_data.csv).
3. Determine the number of columns and rows in the data set, determine the column names.
4. Coerce the variables sex and subtype to factors.
5. Calculate the mean age.
6. Tough question: Normalize the ESR to calculate the JADAS10 in the next exercise. Remember: ESR <20 becomes 0; ESR between 20 and 120 becomes (ESR – 20) / 10; ESR above 120 becomes 10. Make use of the function ifelse(), already used in exercise 10. Because there are now three possible conditions, you have to use it twice, nested inside itself.
7. Calculate the JADAS10 at baseline. Remember, the JADAS10 is the sum of the active joints, the normalized ESR, patient VAS well-being and physician’s global assessment. Make use of vector based calculations.
8. Calculate the mean JADAS10 per JIA subtype.