# R Scripting

## Exercises for unit 1 - R basics and data structures (I)

Marcus Wurzer

Please solve the following problems!

1. Try to answer quiz questions 1. to 3. at the beginning of the "Data structures" chapter of the coursebook: http://adv-r.had.co.nz/Data-structures.html
2. Coursebook exercises:
   a. Why is `1 == "1"` true? Why is `-1 < FALSE` true? Why is `"one" < 2` false?
   b. What happens to a factor when you modify its levels?
   ```
   f1 <- factor(letters)
   levels(f1) <- rev(levels(f1))
   ```
   c. What does this code do? How do f2 and f3 differ from f1?
   ```
   f2 <- rev(factor(letters))
   f3 <- factor(letters, levels = rev(letters))
   ```
3. Create the following vectors using `rep()` and `seq()` (as needed):

   `0 -0.1 -0.2 -0.3 -0.4 -0.5 -0.6 -0.7 -0.8 -0.9 -1 -1.1 -1.2 -1.3 -1.4 -1.5`

   `1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5`

   `0 0 0 0 0 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4`

   `1 2 3 4 5 2 3 4 5 6 3 4 5 6 7 4 5 6 7 8 5 6 7 8 9`
4. Create a vector called `colors.6` consisting of the six elements `"red"`, `"brown"`, `"blue"`, `"green"`, `"orange"` and `"black"` (in this order), then use `colors.6`, `rep()` and `seq()` as needed to create the following vector:

   `"red" "brown" "blue" "brown" "blue" "green" "blue" "green" "orange" "green" "orange" "black"`
5. Construct a vector `x` that contains integers, real numbers, chains of characters, and several `NA` missing values. Test for the positions of the missing values using the `is.na()` function. Produce the subvector where all missing values have been eliminated.
6. Use R to produce a vector containing all integers from 1 to 100 that are not divisible by 7.
7. Suppose that `queue <- c("Steve", "John", "Alison", "Liam")` and that `queue` represents a supermarket queue with Steve first in line. Using R expressions, update the supermarket queue successively:
   a. Barry arrives.
   b. Steve is served and leaves.
   c. Pam arrives and talks her way to the front with one item.
   d. Barry gets impatient and leaves.

e. Alison gets impatient an leaves. (Do not assume that you know where in the queue Alison is standing.)
f. Use `which()` to find out the position of John in the queue.

8. Which of the following assignments will be successful? What will the vectors `x`, `y`, and `z` look like at each stage?

```
rm(list = ls())
x <- 1
x[3] <- 3
y <- c()
y[2] <- 2
y[3] <- y[1]
y[2] <- y[4]
z[1] <- 0
```

9. Show that, when `a` is a scalar (i.e., a vector with just one element) and `x` is a vector, `match(a, x)` is equivalent to `min(which(x == a))`. Compare this with the use of `%in%`.

10. Use `sample()` to simulate percentage grades of 50 students (0%-100%, only integer values possible) and save the results. Create a second vector containing the gender of these 50 students (again using `sample()`.

a. Five students' values should be missing because these students were ill at the date of the final exam. Randomly choose five of the 50 percentage grades and replace them by NA's.
b. Compute the overall mean percentage and the groupwise percentages for males and females, respectively.
c. Convert the percentage grades into a factor `grades` that contains the student's grades according to the Austrian grading system (from "Sehr gut" to "Nicht genügend"). Use the standard grading system of FH Technikum:
   - $< 50\%:$ *Nicht genügend*
   - $>= 50\% - < 63\%:$ *Genügend*
   - $>= 63\% - < 75\%:$ *Befriedigend*
   - $>= 75\% - < 88\%:$ *Gut*
   - $>= 88\%:$ *Sehr gut*
d. Compute the grade average for the whole class.

11. Why do we get a missing value after this sequence of operations?

```
grades <- c("Sehr Gut", "Gut", "Befriedigend", "Gut") # e.g., grades of four pupils
grades <- factor(grades, levels = c("Sehr gut", "Gut", "Befriedigend",
                                    "Genügend", "Nicht genügend"))
grades
```

```
## [1] <NA>         Gut          Befriedigend Gut
## Levels: Sehr gut Gut Befriedigend Genügend Nicht genügend
```