

Beginner R exercises

Creating and removing objects in the environment

B 1) Run the following code in R:

```
a = b = 1  
a = 2
```

What happened? How do you know (which command)? How many objects did you create in your environment? Try assigning the values using the `<-` operator. Does the result differ from the result when you are using `=` to create the variables?

B 2) Remove the variables `a` and `b` from your environment using the `rm()` function.

B 3) Run the following code in R:

```
apple <- banana <- lemon <- 1
```

Remove all objects with an 'n' in the name.

B 4) Run the following code in R:

```
apples <- 5; pears <- 3; pineapples <- 6
```

Then try the following code:

```
apples + pineapples  
apples + Pears
```

Why can't you add `apples` and `Pears` ?

B 5) Use R to compute the square root of 81 and store the result in `t1` .

B 6) Use R to compute 81 to the power a half and store the result in `t2` .

B 7) Use the `==` operator to check whether the contents of `t1` and `t2` are the same.

Working directory and help functionality

- B 8) With the command `getwd()` you can find the current working directory of the R process. Explain what the working directory means and why it is important.
- B 9) Change the working directory of the R session to a folder of your preference. What function do you use? How can you check whether your change has worked?
- B 10) Run the following code in R:

```
demo(package = .packages(all.available = TRUE))
```

What does this code do? How do you run the demo `persp` (in package `graphics`)?

- B 11) Run the following code in R:

```
a <- c(1, 6, 7, 8, 9, NA)
mean(a)
```

This gives `NA`, why? Check `?mean` to find out what arguments the `mean()` function takes as input. Can you find a way to compute the mean with the missing value removed?

- B 12) According to Google there exists an R-function called `mvrnorm()`. Typing `mvrnorm` or `?mvrnorm`, however, gives an error. Why? Which library do you have to load first? How?
- B 13) What does `%>%` do? Can you ask help with `?%>%`, like in `?mean`? If not, how? What is `%>%` doing?
- B 14) The `cor()` function computes correlations. As an example, run the following code in R:

```
cor(c(1, 2, 3, 4), c(1, 4, 7, 15))
```

How can you find out (which function) whether this correlation is statistically significant? How can you find such a function? What is the confidence interval of the correlation? How can you find all functions that have 'cor' in their name?

Modes of objects

B 15) Run the following code in R:

```
a1 <- '1'; a2 <- 1; a3 <- TRUE
```

What are the modes of **a1** , **a2** , and **a3** ? Now run the following code in R:

```
b1 <- c(a1, a2); b2 <- c(a1, a3); b3 <- c(a2, a3); b4 <- c(a1, a2, a3)
```

What are the modes of **b1** , **b2** , **b3** , and **b4** ? Can you explain?

- B 16) Why does **TRUE/TRUE** yield 1, **FALSE/TRUE** yield 0, **FALSE/FALSE** yield **NaN** , and **TRUE/FALSE** yields **Inf** ?
- B 17) Convert the logical **TRUE** to a numerical. Convert this numerical to a character. Next, convert the logical **TRUE** to a character? Why do these results differ?
- B 18) How can you check whether a vector is numeric? Is the vector **c(1,0)** a numeric vector? And the vector **c(TRUE, FALSE)** ? And the vector **c(TRUE, FALSE, 1, 0)** ? Why?

Data types: Vectors

- B 19) Use the `c()` function (or `:`) to create the following vector:

-2 -1 0 1 2 3 4 5 6 7 8

What is the length of this vector?

- B 20) Make a vector that starts at -5 and goes to 5 in steps of 0.5. Can you find out more ways to construct this vector than by using the `c()` function?

- B 21) Create the following vector:

13 15 17 19 21 23 25 27 29 31 33

- B 22) Run the following code in R:

```
a <- 1:5
b <- -a
```

Combine the vectors `a` and `b` into one vector.

- B 23) What is the result of `rep(2, 10)`? What function argument receives the value 10?

- B 24) Use the `rep()` function to create the following vector:

3 3 3 4 4 4 5 5 5 6 6 6 7 7 7

- B 25) Suppose you want to make the vector `"a" "a" "b" "b" "c" "c" "d" "d" "e" "e"` with the `rep()` function but `rep(letters[1:5], 2)` does not work. What do you have to change in this command?

- B 26) The following code gives the first 10 uneven numbers. What is the sum of these numbers?

```
(1:10) * 2 - 1
```

- B 27) Make a vector with the first 10 even numbers. Next, make a vector with the first 10 numbers divided by 5. Finally, create the vector `5 8 11 14 17 20 23 26 29 32`.

- B 28) `logical(5)` makes a logical vector of length 5. How can you make the same vector with the `vector()` function?

- B 29) Run the following code in R:

```
paste(rep(c('a', 'b'), each = 5), 1:5, sep = '.')
```

Next, create the following vector:

"x1m" "x1f" "x2m" "x2f" "y1m" "y1f" "y2m" "y2f"

B 30) Run the following code in R:

```
s <- c(5, 7, 2, 8)
```

First sort `s` from highest to lowest. What option in the `sort()` function do you use? Next, sort `s` from lowest to highest. Why don't you have to use the same option here?

B 31) Run the following code in R and explain why the second line gives a warning:

```
1:10 + 1:2 - 1  
1:10 + 1:3 - 1
```

Data types: Matrices

B 32) Use the `matrix()` function to create the following matrix:

25	24	23	22	21
20	19	18	17	16
15	14	13	12	11
10	9	8	7	6
5	4	3	2	1

B 33) Create the following matrix:

0	0	0	0
1	1	1	1
0	0	0	0
1	1	1	1

B 34) Run the following code in R:

```
m1 <- matrix(1:20, , 4)
```

Why don't you have to give R the number of rows for the matrix?

B 35) With what function can you transpose a matrix?

B 36) `mean(m1)` returns one number. How can you get the mean of each column in `m1` without calculating them each in turn?

B 37) Find out the values of the diagonal in matrix `m1`. Can you make a new matrix of 0's with on the diagonal the diagonal of `m1`?

B 38) Add a new row to matrix `m1` with the sum of each row of the matrix. Use the `rowSums()` function.

B 39) Run the following code in R:

```
m2 <- scale(m1)
```

What are the means of the columns in matrix `m2`? And the standard deviations? Given these values, can you find out what the `scale()` function does?

Data types: Data frames

B 40) Create the following data frame without typing out the first two columns:

	subject	time	score
1	1	t1	7
2	1	t2	8
3	2	t1	8
4	2	t2	8
5	3	t1	9
6	3	t2	8
7	4	t1	9
8	4	t2	7
9	5	t1	7
10	5	t2	6

B 41) Run the following code in R:

```
a <- matrix(1:10, , 2, TRUE)
```

Convert the matrix `a` to a data frame.

Data types: Lists

B 42) Create the following list:

```
$subject
[1] 1 2 3 4 5

$time
[1] "t1" "t2"

$score1
[1] 7 8 9 7 8

$score2
[1] 6 8 8 7 8
```

Data types: Factors

B 43) Run the following code in R:

```
g <- gl(4, 3)
```

Is `g` a vector? (If not, what is it?) How can you check?

B 44) Run the following code in R:

```
v <- rep(c('m', 'f'), 3)
```

Encode the variable `v` as factor levels. How do you check whether you succeeded?

B 45) Run the following code in R:

```
x <- 0:10%%3
```

Convert `x` to a factor. How many factor levels has `x`?

Indexing

46) Suppose you have `x <- 3`. Now type `x[3] <- 5`. What is the value of `x[2]`?

47) Run the following code in R:

```
a <- c(9, 2, 4, 5, 2, 7, 5)
```

Change the first element into an 8. Next, change the 2's in the vector into zero's.

48) Run the following code in R:

```
l <- matrix(c(1, 4, 6, 7, 4, 7), ncol = 3)
```

Use the square brackets to find out the second value of the third column of the matrix `l`.

49) Run the following code in R:

```
m <- matrix(sample(1:100, 100, replace = T), 10, 10)
```

Select the third and fifth row of the matrix `m`.

50) Select those rows from matrix `m` for which the value in the first column is less than 5.

51) Select all but the fourth column of matrix `m`.

- B 52) Change the matrix `m` to a data frame. Give the columns of the data frame the names `'trial.1'`, `'trial.2'`, etc.
- B 53) Select the second to the fifth element of column `'trial.1'`. Select the first 2 elements of column `'trial.4'`.
- B 54) Run the following code in R:

```
b <- c(1, 2, 2, 1, 2, 1, 1, 2, 2, 1)
```

Suppose you want to change all ones into twos and all twos to ones. To try out, run the following code in R. It doesn't work. Can you find a method so that this is done correctly?

```
b[b==1] <- 2  
b[b==2] <- 1
```

- B 55) Run the following code in R:

```
n <- c('bananas', 'apples')
```

Use the `gsub()` function to change all a's in `n` to dots (e.g., bananas becomes b.n.n.s).

- B 56) Run the following code in R:

```
set.seed(1)  
grades <- data.frame(1:30, matrix(sample(4:10, 60, TRUE), , 2))  
names(grades) <- c('student', 'exer', 'exam')
```

Students pass a course when the average grade is at least 5.5 and both grades are larger than 5. Select (index) which students passed the course. Also find out how can you select which students did not pass the course.

Conditions

- 57) What does `(a < 0 | b < 0) (a * b < 0)` mean? If this condition is **TRUE** and **a** is negative, what do you know about **b**?
- 58) Why does `!sum(is.na(c(1, 2, 3, 4, NA, 7))) <= 2` return **FALSE**?
- B 59) Run the following code in R and explain the result:

```
x <- 1:24
which(24%x == 0)
```

- B 60) Run the following code in R:

```
x <- y <- 1:10
sum(x==y) == length(x==y)

x <- y <- 1:10
y[1] <- 0
sum(x==y) == length(x==y)
```

Explain the result of this code. Can you make a test whether **x** and **y** are identical using the `min()` function? And with the `mean()` or `prod()` function? Is there a special function to test whether two vectors are identical?

Sampling and simulating data

B 61) You can throw a die 100 times by typing the following code in R:

```
throws <- sample(1:6, 100, TRUE)
```

How can you see how often each number shows up? How often did each number show up?

B 62) Sample, with replacement, 4 cards from the set 'jack' , 'queen' , 'king' and 'ace' with equal probabilities.

B 63) Create 20 uniform random numbers between 0 and 100. Why is `sample(1:100, 20)` not correct?

B 64) Put 21 uniform random numbers in a vector. What is the median of this vector? Sort the vector from highest to lowest. What is the 11th element of the sorted vector?

B 65) Create 100 normally distributed numbers with `mean = 100` and `sd = 15` . What is the variance of these numbers? Run the code that you used for this again. Why is the variance not exactly the same the second time? What can you use so that, if you run your code again, the results are the same?

B 66) Run the following code in R:

```
x <- matrix(rnorm(100), 25, 4)
```

Find out the covariance matrix of matrix `x` ? Also find out the correlation matrix of matrix `x` ?

Reading and writing data

- B 67) What does the `read.csv()` function do? How would you read in Excel files? What package contains the function `read.spss()` ? Can you find other file types that R has read functions for?
- B 68) Read in the file `example.csv` from the online resources using the `read.csv()` function. How many observations does this data contain?
- B 69) Run the following code in R:

```
d <- data.frame(sex = c('m', 'f'),  
                Age = c(6.7, 6.5, 5.6, 5.4),  
                var1 = c(9, 5, 4, 4),  
                var2 = c(10, 5, 8, 4))
```

Write the data frame `d` to a file named `example.xlsx` so that Excel is able to open it. What options did you use?