Practical: Introduction to R

Laura Rodríguez Navas
rodrigueznavas@posgrado.uimp.es

January 2020

1. Generate the numbers 1, 2, ... , 12, and store the result in the vector x.

```
x <- c(1:12)
x
```

2. Generate four repetitions of the sequence of numbers (6, 2, 4).

```
s <- c(6, 2, 4)
rep(s, 4)
```

3. Generate the sequence consisting of six 9s, then five 2s, and finally four 5s. Store the numbers in a 5 by 3 matrix (populating it columnwise).

```
s <- c(rep(9, 6), rep(2, 5), rep(5, 4))
m <- matrix (s, 5, 3)
m
```

4. Generate a vector consisting of 20 numbers generated randomly from a normal distribution. Use the value 100 as seed (in order to be able to replicate the experiments). Setting the seed is done as follows:

   > set.seed(100)

   Then, calculate the following statistics about the generated vector: mean, median, variance and the standard deviation. Repeat the generation of the vector and the statistics with and without changing the seed and observe what happens.

```
calculation <- function(v) {
vmean <- mean(v)
vmedian <- median(v)
vvariance <- var(v)
vdeviation <- sd(v)
return_list <- list("mean" = vmean, "median" = vmedian,
"variance" = vvariance, "deviation" = vdeviation)
return(return_list)
}

# with seed
set.seed(100)
v <- rnorm(20)
values_list_with_seed <- calculation(v)
print(values_list_with_seed)

# without seed
v <- rnorm(20)
values_list_without_seed <- calculation(v)
print(values_list_without_seed)
```

5. From the resources provided with the course, download the file "data1.txt" that contains information about students.

   (a) Read the data into an R object named students (data is in a space-delimited text file and there is no header row).

```
students <- read.table("data1.txt", header = FALSE, sep = " ")
```

   (b) Add the following titles for columns (see section 9): height, shoesize, gender, population

```
names(students) <- c("height", "shoesize", "gender", "population")
```

   (c) Check that R reads the file correctly.

```
> students
   height shoesize gender population
1     181       44   male     kuopio
2     160       38 female     kuopio
3     174       42 female     kuopio
4     170       43   male     kuopio
5     172       43   male     kuopio
6     165       39 female     kuopio
7     161       38 female     kuopio
8     167       38 female    tampere
9     164       39 female    tampere
10    166       38 female    tampere
11    162       37 female    tampere
12    158       36 female    tampere
13    175       42   male    tampere
14    181       44   male    tampere
15    180       43   male    tampere
16    177       43   male    tampere
17    173       41   male    tampere
```

   (d) Print the header names only.

```
colnames(students)
```

   (e) Print the column height.

```
students[, 1, drop=FALSE]
```

   (f) What is the gender distribution (how many observations are in each groups) and the distribution of sampling sites (column population)?

```
summary(students$gender)
female  male
9       8
summary(students$population)
kuopio  tampere
7       10
```

   (g) Show the distributions in the above item at the same time by using a contingency table.

```
table(students$gender, students$population)
                kuopio tampere
    female      4      5
    male        3      5
```

(h) Make two subsets of your dataset by splitting it according to gender. Use data frame operations first and then do the same using the function subset. Use the help to understand how subset works.

First, using data frame operations:

```
male <- students[students$gender == "male", c("height", "shoesize", "population")]
female <- students[students$gender == "female", c("height", "shoesize", "population")]
male_s1
   height shoesize population
1     181       44     kuopio
4     170       43     kuopio
5     172       43     kuopio
13    175       42    tampere
14    181       44    tampere
15    180       43    tampere
16    177       43    tampere
17    173       41    tampere
female_s1
   height shoesize population
2     160       38     kuopio
3     174       42     kuopio
6     165       39     kuopio
7     161       38     kuopio
8     167       38    tampere
9     164       39    tampere
10    166       38    tampere
11    162       37    tampere
12    158       36    tampere
```

Secondly, using the function subset:

```
male_s2 <- subset(students, gender=="male", c("height", "shoesize", "population"))
female_s2 <- subset(students, gender=="female", c("height", "shoesize", "population"))
male_s2
   height shoesize population
1     181       44     kuopio
4     170       43     kuopio
5     172       43     kuopio
13    175       42    tampere
14    181       44    tampere
15    180       43    tampere
16    177       43    tampere
17    173       41    tampere
female_s2
   height shoesize population
2     160       38     kuopio
3     174       42     kuopio
6     165       39     kuopio
7     161       38     kuopio
8     167       38    tampere
9     164       39    tampere
10    166       38    tampere
11    162       37    tampere
12    158       36    tampere
```

(i) Make two subsets containing individuals below and above the median height. Use data frame operations first and then do the same using the function subset.

First, using data frame operations:

```
under_median_s1 <- students[students$height < median(students$height),
        c("height", "shoesize", "population")]
above_median_s1 <- students[students$height >= median(students$height),
        c("height", "shoesize", "population")]
under_median_s1
   height shoesize population
2     160       38     kuopio
6     165       39     kuopio
7     161       38     kuopio
8     167       38    tampere
9     164       39    tampere
10    166       38    tampere
11    162       37    tampere
12    158       36    tampere
above_median_s1
   height shoesize population
1     181       44     kuopio
3     174       42     kuopio
4     170       43     kuopio
5     172       43     kuopio
13    175       42    tampere
14    181       44    tampere
15    180       43    tampere
16    177       43    tampere
17    173       41    tampere
```

Secondly, using the function subset:

```
under_median_s2 <- subset(students, height < median(students$height),
        c("height", "shoesize", "population"))
above_median_s2 <- subset(students, height >= median(students$height),
        c("height", "shoesize", "population"))
under_median_s2
height shoesize population
2     160       38     kuopio
6     165       39     kuopio
7     161       38     kuopio
8     167       38    tampere
9     164       39    tampere
10    166       38    tampere
11    162       37    tampere
12    158       36    tampere
above_median_s2
height shoesize population
1     181       44     kuopio
3     174       42     kuopio
4     170       43     kuopio
5     172       43     kuopio
13    175       42    tampere
14    181       44    tampere
15    180       43    tampere
16    177       43    tampere
17    173       41    tampere
```

(j) Change height from centimetres to metres for all rows in the data frame. Do this using in three different ways: with basic primitives, a loop using for and the function apply.

```
students[1] <- students[1]*0.01

for(i in 1){
        students[1] <- students[1]*0.01
}

students[1] <- apply(students[1],2, function(x) x * 0.01)
```

(k) Plot height against shoesize, using blue circles for males and magenta crosses for females. Add a legend.

```
male_s3 <- subset(students, gender=="male", c("height", "shoesize"))
female_s3 <- subset(students, gender=="female", c("height", "shoesize"))
plot(male_s3, col="blue")
par(new = T)
plot(female_s3, pch=4, col="magenta", axes= F, xlab=NA, ylab=NA)
title(main = "height vs shoesize")
legend("bottomright", c("male","female"), col = c("blue","magenta"), lty = c(1,1))
```