CeDoSIA SS2020 - Exercise Sheet 1: Introduction to R

Vangelis Theodorakis, Xueqi Cao, Daniela Andrade Salazar, Julien Gagneur

08 June, 2020

Contents

1 Vectors 2
2 Factors 2
3 Computation on matrices 2
4 Data frame manipulation 3
5 Data frame operations 3
6 Looping and writing your own functions 3

1 Vectors

First, create three named numeric vectors of size 10, 11 and 12 respectively in the following manner:

- One vector with the "colon" approach: *from:to*
- One vector with the `seq()` function: *seq(from, to)*
- And one vector with the `seq()` function and the `by` argument: *seq(from, to, by)*

For easier naming you can use the vector `letters` or `LETTERS` which contain the latin alphabet in small and capital, respectively. In order to select specific letters just use e.g. `letters[1:4]` to get the first four letters. Check their types. What is the outcome? Where do you think the difference comes from?

Then combine all three vectors in a list. Check the attributes of the vectors and the list. What is the difference and why?

Hint: If list elements have no names, we can access them with the double brackets and an index, e.g. `my_list[[1]]`

2 Factors

```
f1 <- factor(letters)
levels(f1) <- rev(levels(f1))
f2 <- rev(factor(letters))
f3 <- factor(letters, levels = rev(letters))
```

The function `rev` reverses the order of an order-able object. What is the difference between f1, f2 and f3? Why?

3 Computation on matrices

Create a 10 by 5 matrix which contains the the numbers from 1 to 50 column-wise. Name the rows as 'row_n' and columns as 'col_n'. Compute the mean and sum of each row and column. Add vector seq(60,100,10) as another row to the matrix.

Generate another matrix with the same dimensions, containing random numbers between 1 and 100. Subtract this matrix from the first one.

Plot the covariance matrix of the columns of the resulting matrix with spearman correlation coefficients.

Hint: Check out functions paste0(), colMeans(), rowMeans(), colSums(), rowSums(), sample(), cor() and corrplot() (in package 'corrplot')

4 Data frame manipulation

Create a 3 by 4 matrix that contains the numbers 1 to 12 and then convert it into a data frame. Assign zero to the elements at row 2 which are greater than 4. Set the rownames to "row1", "row2", "row3" and column names to "col1", "col2", "col3" and "col4". Assign 0 to all elements in columns "col3" and "col4". Add a new column named "Letters" with values c("A", "B", "C"). Inspect the structure of the data frame.

5 Data frame operations

Compute the number of women who survived the Titanic. Start by loading the data into a data frame using the following command:

```
tab <- read.csv("extdata/titanic.csv")
```

6 Looping and writing your own functions

Write a function named "generateDataFrameSummary" which takes a data frame as input and outputs the medians of the rows and columns (NA values are discarded), and number of NA values in each row and column as a list.