Systems Biology course:
R programming exercises

September 24, 2014

1 Exercises: R programming

Where to find and how to install R If you do not have R installed on your machine, you would need to download it from `http://www.r-project.org/`. From the homepage, simply click on CRAN, Packages and select your favorite mirror (e.g. Austria). Download the precompiled binary distribution by clicking on your operating system. R comes with a simple editor, feel free to use it or use your favorite editor (e.g. Eclipse, Xcode, ...) to write your code.

Notes The exercises below cover distinct basic topics of R programming. Each of the following exercise requires both the concept presented in the tutorial as well as independent reading of the R manual (`http://cran.r-project.org/doc/manuals/R-intro.pdf`) and of the R documentation. Recall that each built-in function in R has an associated help page that can accessed as `?name` (e.g. `?paste`).
We strongly suggest to get familiar with the R data types (e.g. vectors, data frames, lists, ...) and the methods used to construct and manipulate them.

2 Beginner level

Exercise 1 Create a vector containing all odd numbers smaller than or equal to 75 i) using a for loop and logical operators; ii) using the built-in `seq` function.

Exercise 2 Create a function called `isEven` that takes an integer as argument and returns TRUE (boolean) if the number is even.

Exercise 3 Create a function called `sum10` that given a vector of 10 numbers, return their sum. Try to implement it using a for loop.

Exercise 4 Same as ex3, but use R basic functions.

Exercise 5 Given a string longer than 8 characters, extract the 4-prefix and the 4-suffix (e.g. SystemsBiology will give you *Syst* and *logy*)

Exercise 6 Given a string with your name, compute its reverse

Exercise 7 Given a string composed of two words separated by a blank space, return a string obtained by extracting and pasting (with no space) the first 3 characters from each of the words (e.g. from 'Systems Biology' you would obtain SysBio). Can you extend it to any number of sequences?

Exercise 8 We provide a dataset (in CSV format, data.csv). You can take a look at the first lines of the file using the Linux command `head` as you learned in the previous lecture. First, read in the file into R (hint: check out the function `read.csv2`). Then, plot the data as scatter plot and read the documentation on plot (e.g. check out `plot` and par for specifying plotting parameters).

Exercise 9 Use the dataframe you loaded in the previous exercise to create a new dataframe that contains only the first 10 rows of the original one and that has an additional column containing the element-wise product of the extracted rows. The resulting dataframe should look like:

```
   speed dist prod
1      4    2    8
2      4   10   40
3      7    4   28
4      7   22  154
5      8   16  128
6      9   10   90
7     10   18  180
8     10   26  260
9     10   34  340
10    11   17  187
```

hint: To solve this problem, read a bit on `cbind`.

Exercise 10 Write the dataframe you obtained in ex9 to a file using your favorite function (hint: check out either `write.table` or `write.csv`, `write.delim`, etc). Make sure that you can load the file back into the R session.

3 Intermediate level

Exercise 11 Produce a scatter plot of `wind` vs. `temperature` for the `airquality` dataset. What do we learn, i.e. what relationship do we see? Can you customize your plot? (hint: see `par`) Can we fit a simple linear regression model (hint: see `lm`) to the data and add it to the plot?

Exercise 12 Have a look at the ChickWeight data set. Plot the weight as a function of time for chick 7. Superimpose a line representing the same information for chick 9. Add a legend to the plot. Find out how to plot in a pdf device (see, e.g. ?pdf) and plot the previous graph to a pdf file. Extract all weights at time point $t = 21$ and arrange your result in a list where each slot corresponds to a Diet (there are 4 regimes, so it will be a list of 4 vectors). Hint: see ?split. Boxplot the weights in this 4 diets. Any difference in mean weight between diet 1 and 3? How could you go about testing this?

Exercise 13 Have a look at the eurodist data set. What is the class of this object? Do you know it? Sort the cities by increasing distance from Geneva. Hint: you may want to look at functions that allow you to coerce an object to a different class. Specifically, see ?as.matrix here.

Exercise 14 Write your own functions to compute mean and standard deviation of a numeric vector. Apply them to the precip data sets and cross check your results using the built-in mean and sdfunctions.

Exercise 15 Write your own function to transpose a matrix. Use of the built-in t function is considered cheating!

Exercise 16 Write the swiss dataset as csv file. Then, read the file into the console. How long does it take to execute the command?

Exercise 17 Write a function that takes a numeric vector as input and returns a vector of differences between consecutive elements. Is there a way to do the same using a built-in function? Apply both functions to the following input

```
v <- 1 : 10
v <- c(3,4,5,1,2,9,11,8)
v <- seq(1,100,2)
```

Exercise 18 Create three vectors containing the first 10 numbers divisible by 2, by 3 and by 4, respectively. Use these vectors to create a list, and give suitable names to the slots. Now that you have a list, can you compute the sum for each slot? If you sum the three slots, you get a vector of length 10. What is it?

Exercise 19 Write a function to convert Fahrenheit to Celsius.

Exercise 20 Write a function which takes a single numeric value x as argument and returns a logical value TRUE if x is larger than 5, FALSE otherwise. Then, extend the previous function by including the threshold (t) as an argument. Finally, given l <- list(seq(1,75, by = 3)) use the function you created in a to test each single entry in l. Set $t = 15$.