

Coding assignment

06.04.2010 SGN-6166 Introduction to R programming

Guidelines

The tasks are given in this document along with the file "Assignments_2010-04-06.r". You should write your codes into this file, too. Download the file from http://www.cs.tut.fi/~aho2/Assignments_2010-04-06.r.

In the file, the tasks are indicated by the notation:

```
##### TASK BEGINS #####
```

```
# Task x (y points)
```

```
#
```

```
# Possible task description...
```

```
#
```

```
##### TASK ENDS #####
```

After you have completed the tasks (11:45 at latest):

- 1) Save the file.
- 2) Close R and your R editor.
- 3) Then email your result file "Assignments_2010-04-06.r" to tommi.aho@tut.fi.

Task 1

Load the data from <http://www.cs.tut.fi/~aho2/datadirectory.zip> and extract it into a directory called 'datadirectory'. The directory contains 1000 data files named as 'datafile1', 'datafile2', ..., 'datafile999', 'datafile1000'. The files are stored in the binary format of R. In your code, access this directory programmatically. That is, your code should be located in a different directory than the data files. Each of the files contains 100 data points. Calculate the mean of data points in each file and plot all the 1000 means in decreasing order.

Task 2

Write a function that may have an arbitrary number of logical-valued arguments (TRUE, FALSE, or NA). The function should determine its return value according to the following logical rules:

- If none of the arguments has value FALSE, then return TRUE.
- Otherwise return FALSE.

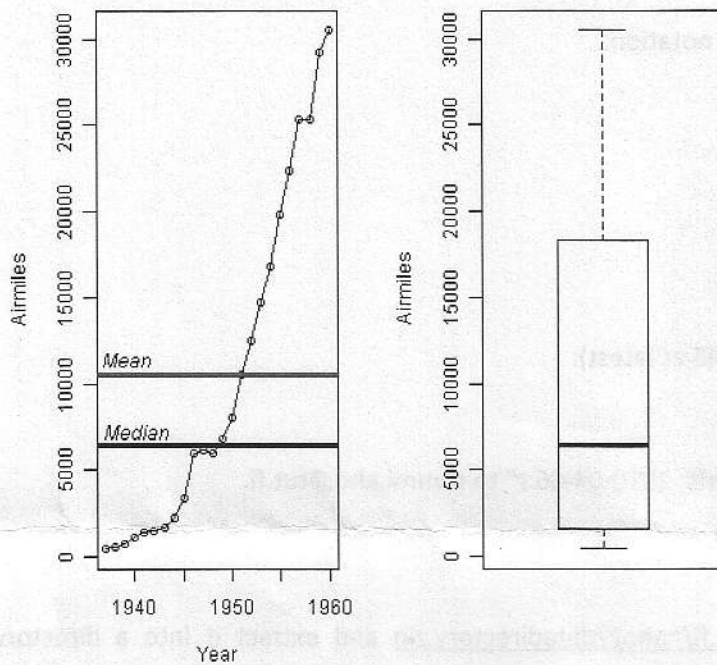
Test your function using the following inputs (that contain different number of arguments):

- TRUE
- NA

- FALSE, TRUE
- FALSE, TRUE, NA

Task 3

Load the data set called "airmiles" and implement a code that generates the following figure. Your code should also save the figure to a file. The file type should be Portable Network Graphics (".png").



Tasks 4

Implement a code that

- (1) Counts the total number of words in the given text. A word is now defined as a character string separated by a space. Therefore, the name "R" and the abbreviation "CRAN" are counted as words. Also, punctuation affects because "statistics" and "statistics." are now considered as two separate words.
- (2) Counts the number of unique words in the text.
- (3) Counts the number of unique words starting with lower-case 'a'.

Tasks 5

Implement the quicksort algorithm according to the following pseudocode (the algorithm and the pseudocode are explained at <http://en.wikipedia.org/wiki/Quicksort>):

```
function quicksort(array)
  var list less, greater
  if length(array) ≤ 1
    return array
  select and remove a pivot value pivot from array
  for each x in array
    if x ≤ pivot then append x to less
    else append x to greater
  return concatenate(quicksort(less), pivot, quicksort(greater))
```

The input of the function should be a vector of real numbers.