PLPTH813, Homework 2

Due on 2/21/2019

The purpose of this homework is to practice R commands. Note that each question could need multiple lines of commands. Please do not hesitate to ask questions.

Problem 1:

x is a vector with 100 random numbers. Please run the following two commands in R first.

set.seed(100)

x <- round(runif(100)\*100)

Please write down R commands to tackle each following task :

1. extract the 2nd and 51th elements out of x
2. extract numbers larger than 20 and determine the total number of the extracted elements (two R commands)
3. sum and mean of all values
4. add 105 as one more value to x
5. only keep first 10 values

Problem 2:

*fruits* is a vector with 10 character elements.

fruits <- c("apple", "orange", "banana", "pear", "pineapple",

"kiwi", "blueberry", "lemon", "peach", "cherry")

Please write down R commands to tackle each following task :

1. determine character/letter number of each element
2. remove "orange" from the vector
3. check which element(s) containing "apple"
4. replace "apple" with "APPLE" for those fruit names containing "apple"
5. create a logical vector with 10 elements to indicate whether the fruit is your favorite or not  (e.g., fruits.like <- c(TURE, FALSE, …)); TRUE means favorite and FALSE means not ; then use this logical vector to extract your favorite fruit(s)

Problem 3:

The data of annual precipitation of many cities were deposited at:

<https://github.com/liu3zhenlab/teaching/raw/master/PLPTH813Bioinformatis/2019/data/HW02_annual.precipitation.txt>

1. please read this online file to a data frame variable of “ap” using read.delim and report the command.
2. report the command line to extract the annual precipitation data of “Boston”
3. report the command line to calculate the average precipitation across all the cities
4. plot a barplot of all the annual precipitation data; the height of each bar represents the annual precipitation; please report both the commands and the plot
5. plot a histogram to display the distribution of all the annual precipitations; please report both the commands and the plot
6. report the command line to extract the subset of all the cities with annual precipitation larger than 50 and output this subset, including annual precipitation, to a new file.

Problem 4 (from in-class problem):

Provide R scripts for the following steps:

1. create a data frame

The data frame should include three columns (1. Name 2. Major 3. Gender) and two or three rows (entries) related to the information about your neighbors and you

1. write an output file
2. read the file back to R and add one more column (e.g., favorite color)