Homework 2: Make It Rain

CS 201-A01, Spring 2021

8 February 2021

This Homework was adapted from Homework 1 of Carnegie Mellon University's 36-350 course[1].

General instructions for homeworks: Upload both (i) the R Markdown file and (ii) the `nb.html` file to eCampus. You should give the commands to answer each question in its own code block.

Note: Your responses must be supported by both textual explanations and code you use to produce your result. Just examining your various objects in the "Environment" section of RStudio is insufficient – you must use scripted commands. If you are unsure what I mean by this, **ask me**.

1. **Problem 1:** The data set at http://www.statprogr.science/data/rnf6080.dat records hourly rainfall at a certain location in Canada, every day from 1960 to 1980.

   a. We need to load the data set into R using the command `read.table()`. Use `?read.table` to learn what arguments this function takes. Once you have the necessary input, load the data set into R and make it a data frame called `rain.df`.
   b. How many rows and columns does `rain.df` have? (If there are not 5070 rows and 27 columns, something is wrong; check the previous part to see what might have gone wrong in the previous part.)
   c. What are the names of the columns of `rain.df`?
   d. What is the value of row 5, column 7 of `rain.df`?
   e. Display the second row of `rain.df` in its entirety.
   f. Explain what this command does:

   ```
   names(rain.df) <- c("year","month","day",seq(0,23))
   ```

   by running it on your data and examining the object. (You may find the display functions `head()` and `tail()` useful here.) Is it clear now what the last 24 columns represent?

   g. Create a new column in the data frame called `daily`, which is the sum of the rightmost 24 columns. With this column, create a histogram of the values in this column, which are supposed to be daily rainfall values. What is wrong with this picture? **Hint:** This data file codes missing rainfall measurements as `-999`.
   h. Create a new data frame `rain.df.fixed` that takes the original and fixes it for the apparent flaw you have discovered. Having done this, produce a new histogram with the corrected data and explain why this is more reasonable.

2. **Problem 2:** Syntax and class-typing.

   a. For each of the following commands, either explain why they should be errors, or explain the non-erroneous result.

   ```
   vector1 <- c("5", "12", "7", "32")
   max(vector1)
   sort(vector1)
   sum(vector1)
   ```

[1]Shalizi, C. R. and Thomas, A. C. (2014), "Statistical Computing 36-350: Beginning to Advanced Techniques in R", http://www.stat.cmu.edu/ cshalizi/statcomp/14

b. For the next series of commands, either explain their results, or why they should produce errors.

```
vector2 <- c("5",7,12)
vector2[2] + vector2[3]

dataframe3 <- data.frame(z1="5",z2=7,z3=12)
dataframe3[1,2] + dataframe3[1,3]

list4 <- list(z1="6", z2=42, z3="49", z4=126)
list4[[2]]+list4[[4]]
list4[2]+list4[4]
```

3. **Problem 3:** Working with functions and operators.
   a. The colon operator will create a sequence of integers in order. It is a special case of the function `seq()` which you saw earlier in this assignment. Using the help command `?seq` to learn about the function, design an expression that will give you the sequence of numbers from 1 to 9673 in increments of 372. Design another that will give you a sequence between 1 and 10000 that is exactly 50 numbers in length.
   b. The function `rep()` repeats a vector some number of times. Explain the difference between `rep(1:3, times=3)` and `rep(1:3, each=3)`.