CONDITIONALS, LOOPS, AND APPLY

There are now three rules that apply to all projects:

a) Follow instructions *precisely*. If I do not tell you what to write on a particular line, leave it blank.
b) Do not use any functions or approaches to problems that we have not yet learned in this course.
c) All code must be *scalable by sample size* unless specifically noted otherwise. This means your code should work equally well on a dataset with N=10 as N=1000.

Part 1 – Set up a new R Studio Project and Install New Package

1. Create a new R Studio project with correct file and directory structure. Create an R script file called week3.R and save it appropriately. Save the week3.csv data file from Blackboard.
2. In the R console or via the menus, install the *rstudioapi* package if you don't have it already.

Part 2 – Data Import and Cleaning

3. **Line 1:** Write a comment that says: **Data Import and Cleaning**
4. **Line 2:** Write the following: **library(rstudioapi)**
5. **Line 3:** Write the following: **setwd(dirname(rstudioapi::getActiveDocumentContext()$path))** If you installed *rstudioapi* correctly, these two lines together will set the working directory to the location you currently have week3.R, and you will see no errors/output when executing them.
6. **Line 5:** Use *read.csv* to create a data frame called *raw_df* from the CSV file you downloaded earlier.
7. **Line 6:** Recast (i.e., convert and save over itself) timeStart into POSIX format. You should see the type change from *factor* in the Environment panel when you do this.
8. **Line 7:** Recast timeEnd into POSIX format.
9. **Line 8:** Everyone that participated in June 2017 in this file was a research assistant testing the code. Create a new data frame called *clean_df* that contains only real participants.
10. **Line 9:** Q6 of this survey read "If you're paying attention, answer Strongly Disagree (1) to this question." Update *clean_df* so that it *only* contains people who were paying attention.

Part 3 – Analysis

11. **Line 11:** Write a comment that says: **Analysis**
12. **Line 12:** Save the total time spent on the study by each participant in seconds into a new variable in *clean_df* called *timeSpent* using *difftime*(). You will probably want to look at the R documentation for *difftime*().
13. **Line 13:** Create a histogram of *timeSpent*. You will need to recast *timeSpent* as numeric first.
14. **Line 14:** Using *lapply*, iterate over the 5^{th} – 14^{th} columns of *clean_df* and run the function *table*. Save the result to a variable called *frequency_tables_list*.
15. **Line 15:** Using *lapply*, iterate over the *frequency_tables_list* variable and run the *barplot* function. If done correctly, you should see 10 bar charts appear in the Plots panel of R Studio.
16. **Line 16:** Participants should have answered q1 with a greater or equal value to what they answered q2 with, and they never should have answered q2 with the same value as what they answered q3 with. Display a count of the number of times in *clean_df* that this actually happened.

Part 4 – Submission (same as usual)