

Create a .R file. Copy the assignment prompt, below, and paste it into the .R file. Write your R code for each question beneath that question. Also write at least one sentence (in English, not code) answering the question in a comment beginning with #. If you aren't sure how to answer a question, write pseudocode and/or describe how you would approach it.

When done, upload this .R file directly to Canvas. Some questions have multiple parts - make sure to answer all of them.

BEFORE YOU START:

Download (to your class directory, meaning folder on your machine) and load the "my_data.csv" into R using the following code. (Don't forget to replace the file path in parentheses with where you saved the file on your machine):

```
# my_data <- read.csv("User/sts115_yourname/my_data.csv")
```

Use this dataset to answer the relevant homework questions, below.

Tip: If you run into issues loading in the dataset, try changing your working directory in R. Just like you can move around in command line, you can move directories in R using set.wd. For example:

```
# setwd("/path/to/my/directory")
```

If you need more help loading a file and setting your working directory, see 7.0 (https://ucdavisdatalab.github.io/adventures_in_data_science/data-forensics.html) in the course reader.

```
#####
```

1. Interview 5 people from outside of STS115 and add their data to the "my_data" data frame from class to create a new data frame called "our_data". You will use this for the subsequent questions. (Please do not interview someone who a classmate has already interviewed. You can call, Zoom, etc. but you need to actually talk to other human beings to complete this assignment!)

2. Come up with a yes/no or true/false question you'd have liked the survey to have asked. Make up these data for every subject in the survey, then create it as a logical vector and add it to the data frame.

3. Use a single function to return the class of each column in the data frame.

4. Look up how to use the function `max`. Use it to calculate the longest commute distance and travel time.

5. Calculate travel speed across subjects as miles per minute. Assign this to a new vector ``mi.per.min`` and add it to the data frame.

6. Use a function to return the total number of elements in `mi.per.min`.

7. Index/subset the vector ``major`` to get a new vector that contains the 3rd, 1st, and 9th elements (in that order).

8. R's ``[`` indexing operator accepts several different types of indexes, not only positive whole numbers. For example, the operator accepts negative numbers as indexes. Using the vector ``places``, try out three to five different negative indexes. Based on the results, what do you think the `[`` operator does with negative indexes?

9. Consider the R code ``c(3, 3.1, "4", -1, TRUE)``.

a. WITHOUT running the code, what data type you think the result will have and why?

b. Now run the code to check whether your guess was correct. If it was not correct, explain what the actual data type is and why. If your guess was correct, write a new, different line of code that yields the same resulting data type.

10. Run the code ``"four" < "five"``. Paste the output from R into a comment and explain why you think it provided that result.

11. There are several major mistakes in the data entered in class.

a. Describe in complete sentences what at least 2 of the errors are. Make some guesses as to how those errors may have happened, and how they might affect analyses and/or re-use of these data.

b. Pretend the data frame was too big for you to view it manually. List specific function calls you could use in R to help you find these mistakes programmatically.