HW 1

Enter your name and EID here: Harini Shanmugam hs28663

You will submit this homework assignment as a pdf file on Gradescope.

For all questions, include the R commands/functions that you used to find your answer (show R chunk). Answers without supporting code will not receive credit. Write full sentences to describe your findings.

Part 1: (11 pts)

The dataset `mtcars` was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and other aspects of automobile design and performance for different cars (1973-74 models). Look up the documentation for this data frame with a description of the variables by typing `?mtcars` **in the console** pane.

Question 1: (2 pt)

Take a look at the first 6 rows of the dataset by using an `R` function in the code chunk below. Do you know about any (or all) of these cars?

```
# Displays the first 6 rows of the dataset
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

I know of the Mazda car brand, but apart from that, I do not know anything about the listed cars.

Question 2: (2 pts)

How many rows and columns are there in this data frame in total?

```
# Shows details of the  mtcars data set in the output pane
?mtcars
```

There are 32 rows and 11 columns in this data frame.

Question 3: (1 pt)

Save `mtcars` in your environment and name it as your `eid`. From now on, use this new object instead of the built-in dataset.

```
# Setting the mtcars data set to "hs28663" as an object
hs28663 <- mtcars
```

Question 4: (2 pts)

When is your birthday? Using indexing, grab the value of `mpg` that corresponds to the day of your birthday (should be a number between 1 and 31).

```
# Finding the mpg listed in the 16th row of the dataset
hs28663[16,1]
```

```
## [1] 10.4
```

The mpg that corresponds to the day of my birthday (the 16th) is 10.4 mpg.

Question 5: (2 pts)

Using logical indexing, count the number of rows in the dataset where the variable `mpg` takes on values greater than 30.

```
# Counting the number of rows where mpg is greater than 30
nrow(hs28663[hs28663$mpg > 30, ])
```

```
## [1] 4
```

There are 4 rows in the dataset where the variable mpg takes on values greater than 30.

Question 6: (2 pts)

Let's create a new variable called `kpl` which converts the fuel efficiency `mpg` in kilometers per liter. Knowing that 1 mpg corresponds to 0.425 kpl, complete the following code and calculate the max kpl:

```
# Converting mpg to kpl by a scalar mulitple and assigning it to the new object "kpl"
kpl <- hs28663$mpg*0.425
# Finding the max kpl from the dataset
max(kpl)
```

```
## [1] 14.4075
```

The max kpl is 14.4075 kpl.

Part 2: (6 pts)

Let's quickly explore another built-in dataset: `airquality` which contains information about daily air quality measurements in New York, May to September 1973.

Question 7: (2 pts)

Calculate the mean `Ozone` (in ppb). Why does it make sense to get this answer? *Hint: take a look at the column* `Ozone` *in the dataset.*

```
# Calculating mean of Ozone variable in airquality dataset
mean(airquality$Ozone)
```

```
## [1] NA
```

It makes sense to an answer of NA because there are some cells that are not numbers have "NA" instead.

Question 8: (2 pts)

Look at the documentation for the function `mean()` by running `?mean` **in the console**. What argument should be used to find the mean value that we were not able to get in the previous question? What type of values does that argument take?

An R object argument is needed to find the mean value. The argument takes numeric/logical vectors, date, date-time, and time interval objects. Since some of the cells had "NA", which are not numeric/logical vectors, date, date-time, or time interval objects, the mean() function could not complete the computation of the mean.

Question 9: (2 pts)

Sometimes the R documentation does not feel complete. We wish we had more information or more examples. Find a post online (include the link) that can help you use that argument in the `mean()` function. Then finally find the mean ozone!

```
# Calculating mean of Ozone variable in airquality dataset while accommodating
# for NA data cells
mean(airquality$Ozone, na.rm = TRUE)
```

```
## [1] 42.12931
```

The mean ozone is 42.12931 ppb.
I used the following website https://www.statology.org/mean-function-in-r/#:~:text=You%20can%20use%20the%20mean%20%28%29%20function%20in,in%20practice.%20Example%201%3A%20Calculate%20Mean%20of%20Vector

---

Formatting: (3 pts)

Knit your file! Into pdf directly or into html.

Is it working? If not, try to decipher the error message (look up the error message, consult websites such as stackoverflow or crossvalidated.

Once it knits in html, click on `Open in Browser` at the top left of the window pops out. Print your html file into pdf from your browser. Any issue? Ask your classmates or TA!

```
##                                       sysname
##                                       "Linux"
##                                       release
##                             "4.15.0-191-generic"
##                                       version
## "#202-Ubuntu SMP Thu Aug 4 01:49:29 UTC 2022"
##                                      nodename
##                     "educcomp02.ccbb.utexas.edu"
##                                       machine
##                                      "x86_64"
##                                         login
##                                     "unknown"
##                                          user
##                                     "hs28663"
##                                 effective_user
##                                     "hs28663"
```