HW 1

Enter your name and EID here:

Joseph Hendrix | jlh7459

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:

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. Have you heard about any (or all) of these cars?

```
# your code goes below (make sure to edit comment)
head(mtcars, n = 6L)
```

```
##                     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
```

Your answer goes here. Write sentences in bold.

I haven't heard of any of these cars, but I have heard of he Mazda RX7 and RX8, which are successors to the RX4.

Question 2: (2 pts)

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

```
# your code goes below (make sure to edit comment)
nrow(mtcars)
```

```
## [1] 32
```

```
ncol(mtcars)
```

```
## [1] 11
```

Your answer goes here. Write sentences in bold.

mtcars has 32 rows and 11 columns.

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.

```
# your code goes below (make sure to edit comment)
jlh7459 <- mtcars
```

Your answer goes here. Write sentences in bold.

Question 4: (2 pts)

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

```
# your code goes below (make sure to edit comment)
jlh7459$mpg[6]
```

```
## [1] 18.1
```

Your answer goes here. Write sentences in bold.

My birthday is on 9/6/2000.

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.

```
# your code goes below (make sure to edit comment)
length(jlh7459$mpg[jlh7459$mpg > 30])
```

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

Your answer goes here. Write sentences in bold. There are four rows where mpg is 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:

```
# Add a new variable to the dataset
jlh7459$kgl <- jlh7459$mpg * 0.425
jlh7459
```

```
##                     mpg cyl  disp  hp drat    wt  qsec vs am gear carb     kgl
## Mazda RX4          21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4  8.9250
## Mazda RX4 Wag      21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4  8.9250
## Datsun 710         22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1  9.6900
## Hornet 4 Drive     21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1  9.0950
## Hornet Sportabout  18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2  7.9475
## Valiant            18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1  7.6925
## Duster 360         14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4  6.0775
## Merc 240D          24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2 10.3700
##  [ reached 'max' / getOption("max.print") -- omitted 24 rows ]
```

```
max(jlh7459$kgl)
```

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

Your answer goes here. Write sentences in bold.

The maximum kilometers per liter in the dataset is 14.4075.

Part 2:

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

```
# your code goes below (make sure to edit comment)
mean(airquality$Ozone)
```

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

Your answer goes here. Write sentences in bold.

some of days don't have entries for the ozone. as such, some entries are NA, which you can't take the mean of

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?

Your answer goes here. Write sentences in bold.

na.rm takes boolean values to exclude NA entries

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!

```
# your code goes below (make sure to edit comment)
mean(airquality$Ozone, na.rm = TRUE)
```

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

Your answer goes here. Write sentences in bold.

https://www.statology.org/mean-function-in-r/ (https://www.statology.org/mean-function-in-r/)

The mean air quality for valid entries is 42.12931.

Formatting: (3 pts)

Knit your file! You can knit 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 (https://stackoverflow.com/) or crossvalidated (https://stats.stackexchange.com/).

Remember to select pages for each question when submitting your pdf to Gradescope.

```
##
sysname
##
"Darwin"
##
release
##
"20.6.0"
##
version
## "Darwin Kernel Version 20.6.0: Sun Nov  6 23:17:00 PST 2022; root:xnu-7195.141.46~1/R
ELEASE_X86_64"
##
nodename
##                                                    "wireless-10-145-13-161.publ
ic.utexas.edu"
##
machine
##
"x86_64"
##
login
##
"root"
##
user
##
"josephhendrix"
##
effective_user
##
"josephhendrix"
```