Problem Set 7

name?!

date?!

For this problem set we will work with motor vehicle crash data from New York City. You can read more about this publicly available data set on their website.

The data is called "Motor_Vehicle_Collisions_Crashes". We want you to perform the following:

1. Rename the column names to lower-case and replace spaces with an underscore.
2. Select only:
    - crash_date
    - number_of_persons_injured
    - contributing_factor_vehicle_1
    - vehicle_type_code_1
3. Drop all rows with an NA value
4. Lower case the vehicle_type_code_1 variable and replace spaces with a dash.
5. Filter the data for vehicles that have a count/appear in the data set 500 times or more
    - Hints: group_by(), mutate(), n(), filter()
6. Calculate the percentage by vehicle
7. Which vehicle group accounted for 0.3% (0.00374) of the accidents?

We have grouped the questions below to push you to perform commands with less code. As you're building your code we recommend going line by line to test, then combining.

Questions 1-3

```
df_motor <- df_motor %>%
  # lower case and remove spaces
  rename_with(~ tolower(gsub(" ","_", .x, fixed=TRUE))) %>%
  # select certain columns
  select(crash_date, crash_time,
         number_of_persons_injured,
         contributing_factor_vehicle_1,
         vehicle_type_code_1) %>%
  # drop NA rows
  drop_na()
```

Questions 4-5

```
# lower case vehicles and add dash between spaces
df_motor <- df_motor %>%
  mutate(vehicle_type_code_1 =
           gsub(" ", "-", ignore.case=T, tolower(vehicle_type_code_1))) %>%
  # organize by vehicles
  group_by(vehicle_type_code_1) %>%
  # create a variable for counts
  mutate(count = n()) %>%
  # filter counts > 500
  filter(count > 500)
```

Question 6

```
# calculate percentage by vehicle
df_motor %>%
  group_by(vehicle_type_code_1) %>%
  summarize(count = n(),
            perc = count/nrow(df_motor))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 13 x 3
##    vehicle_type_code_1                count    perc
##    <chr>                              <int>   <dbl>
##  1 ambulance                            692 0.00375
##  2 bike                                1825 0.00989
##  3 box-truck                           3830 0.0208
##  4 bus                                 2862 0.0155
##  5 convertible                          577 0.00313
##  6 dump                                 543 0.00294
##  7 motorcycle                          1214 0.00658
##  8 pick-up-truck                       5411 0.0293
##  9 sedan                              85181 0.461
## 10 station-wagon/sport-utility-vehicle 71728 0.389
## 11 taxi                                8104 0.0439
## 12 tractor-truck-diesel                1434 0.00777
## 13 van                                 1177 0.00638
```

Question 7

ambulances * count: 692 * perc: 0.003