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 1.5% (0.0155) 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

Questions 4-5

Question 6

Question 7