

## SI 618 Homework 8 (100 points)

The purpose of this homework is to give you more experience with data aggregation in R, use of factors, SQL access, and visualization using ggplot() on a real-world dataset.

### Dataset

The dataset contains year-by-year fuel economy data for many different manufacturers and models of vehicles. The data used was originally downloaded from <http://www.fueleconomy.gov/feg/ws/index.shtml> and converted into a sqlite database. You don't need to download the data again, as it is provided in the attached zip file **si618\_hw8.zip**. The database has one table and that has the following columns:

- year - model year
- make - manufacturer (division)
- model - model name (carline)
- VClass - EPA vehicle size class
- cylinders - engine cylinders
- displ - engine displacement in liters
- trany - transmission
- city08 - city MPG for fuelType1
- highway08 - highway MPG for fuelType1
- comb08 - combined MPG for fuelType1

Your goal is to read data from this database, analyze it, and generate a report that looks like 'si618hw8\_sample\_solution.html' in the attached zip file. Your report doesn't need be pixel-for-pixel identical, but should capture the same set of data summaries with the same general graphic formatting as in the given plot forms.

### Data Aggregation and Visualization in R

You will write a RMarkdown document that will generate a report looking like the provided si618\_hw8\_sample\_solution.pdf. The steps involved and grading rubric are described there. In the provided RMarkdown template si618\_hw8\_template.rmd, you'll need to fill in the R code.

You should use the ggplot() function in the ggplot2 package and data.table functions (or functions in the plyr package) to make your plots. You'll generate quite a few plots here, and you don't want to generate them all manually by cutting and pasting code. One way could be to write a "for" loop in R to generate the two plots for each VClass in one iteration of the loop. (For an extra challenge, try to see if it's possible to do without the 'for' loop.)

Note that in your code, using the print statement with the ggplot object, like this:

```
ggobj = ggplot( *your settings here* ) *and possibly here*  
print(ggobj)  
should cause the plot to appear in the HTML file generated by RMarkdown.
```

## **What to submit:**

A zip file named 'si618\_hw8\_youruniqueusername.zip' containing:

- The R Markdown file named 'si618\_hw8\_report\_youruniqueusername.Rmd'
- The HTML or PDF report generated by your R Markdown file