# SI 618 Homework 8

## Getting Data from SQLite Database (10 points)

In the data preparation step, a SQLite database has been created and populated with vehicle data. Now the data in the database is retrieved using R package DBI and RSQLite and stored in a data frame named vehicles. Here are the first 10 rows of the data frame, and the summary.

##	yea	r m	ake		model		VClas	s cyli	ndars	dienl	
##		5 Alfa Ro		er Veloc		Тио	Seater	-	4	2.0	
	2 198		_		arossa		Seater		12	4.9	
	3 198		dge			Subcompa			4	2.2	
##			dge B150/E		_	Барсошре	Van		8	5.2	
	5 199		•	gacy AWD		Compa	act Car		4	2.2	
##			aru bog	-	Loyale	_	act Car		4	1.8	
	7 199		aru		Loyale	-	act Car		4	1.8	
##			rota		orolla	-	act Car		4	1.6	
##		J			orolla		act Car		4	1.6	
	10 199	•	rota		orolla		act Car		4	1.8	
##	10 100		any city08				100 041	Z	-	1.0	
##	1	Manual 5-			25	21					
		Manual 5-			14	11					
##		Manual 5-	-		33	27					
##		omatic 3-			12	11					
##		Manual 5-			23	19					
##		omatic 3-	-		24	22					
##		Manual 5-	-		29	25					
##		omatic 3-	_		26	24					
##		Manual 5-			31	26					
##		omatic 4-	-	3	30	25					
##		-			model			VClass			
##	Min.	year Min. :1984 Len							Length:35719		
##			•	ass :character		Length:35719 Class:character		Class : character			
##				aracter				Mode			
##	Mean									acter	
##		3rd Qu.:2008									
##	Max. :2016										
##		cylinders displ		las	trany			city08			
##	•	Min. : 2.000 Min.		_		th:35719		Min. : 6.00			
##				:2.200	_	:chara		1st Qu			
##		n : 6.000			Mode			Median			
##	Mean						Mean	:17.5	54		
##	3rd Q	3rd Qu.: 6.000 3rd Qu.:4.300					3rd Qu	.:20.0	00		
##	Max.	Max. :16.000 Max. :8.400					Max.	:53.0	00		
##	hig	highway08 com		800							
##			Min. :	7.00							
##		1st Qu.:20.00 1st		16.00							
##		n:23.00	Median :								
##	Mean	:23.68	Mean :	19.79							
##	3rd Q	3rd Qu.:27.00 3rd Qu.:22.0									
##	Max.	:61.00	Max. :	53.00							

### Converting to Factors (10 points)

To make downstream analysis easier, we convert the data in columns vehicles\$make, vehicles\$VClass, vehicles\$cylinders, and vehicles\$trany into factors. Here is the summary of the data frame after the conversion.

```
##
         vear
                            make
                                           model
##
                    Chevrolet: 3635
                                        Length: 35719
    Min.
            :1984
##
    1st Qu.:1990
                    Ford
                              : 2958
                                        Class : character
##
    Median:1999
                    Dodge
                                2465
                                        Mode
                                              :character
                    GMC
##
    Mean
            :1999
                              : 2306
##
    3rd Qu.:2008
                    Tovota
                              : 1821
##
    Max.
            :2016
                    BMW
                              : 1518
##
                              :21016
                     (Other)
##
                              VClass
                                              cylinders
                                                                 displ
    Compact Cars
                                                                     :0.600
##
                                  : 5160
                                           4
                                                   :13596
                                                             Min.
                                                   :12522
##
    Subcompact Cars
                                  : 4643
                                           6
                                                             1st Qu.:2.200
##
    Midsize Cars
                                  : 4035
                                           8
                                                   : 7938
                                                             Median :3.000
    Standard Pickup Trucks
                                                      759
                                                                     :3.328
##
                                  : 2354
                                           5
                                                             Mean
##
    Sport Utility Vehicle - 4WD: 2090
                                           12
                                                      505
                                                             3rd Qu.:4.300
##
    Two Seaters
                                           3
                                                      195
                                                                     :8.400
                                  : 1734
                                                             Max.
                                                      204
##
    (Other)
                                  :15703
                                           (Other):
##
                                  city08
                                                                      comb08
                                                  highway08
                 trany
##
    Automatic 4-spd:11035
                                      : 6.00
                                                       : 9.00
                                                                 Min.
                                                                         : 7.00
                              Min.
##
    Manual 5-spd
                    : 8252
                              1st Qu.:15.00
                                                1st Qu.:20.00
                                                                 1st Qu.:16.00
    Automatic 3-spd: 3151
                              Median :17.00
                                                Median :23.00
                                                                 Median :19.00
                    : 2206
##
    Manual 6-spd
                              Mean
                                      :17.54
                                                Mean
                                                       :23.68
                                                                 Mean
                                                                         :19.79
    Automatic (S6): 2201
##
                              3rd Qu.:20.00
                                                3rd Qu.:27.00
                                                                 3rd Qu.:22.00
##
    Automatic 5-spd: 2179
                              Max.
                                      :53.00
                                                Max.
                                                       :61.00
                                                                 Max.
                                                                         :53.00
                    : 6695
    (Other)
```

#### Filter Down Data (30 points)

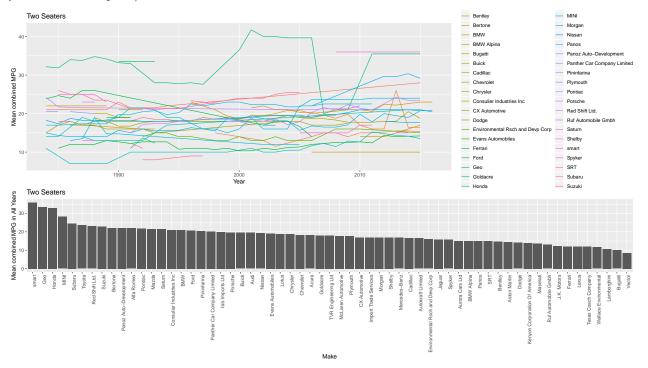
We will filter down the data such that only 'VClass' with more than 40 vehicles are kept. Here is the summary of the data frame after this subsetting step.

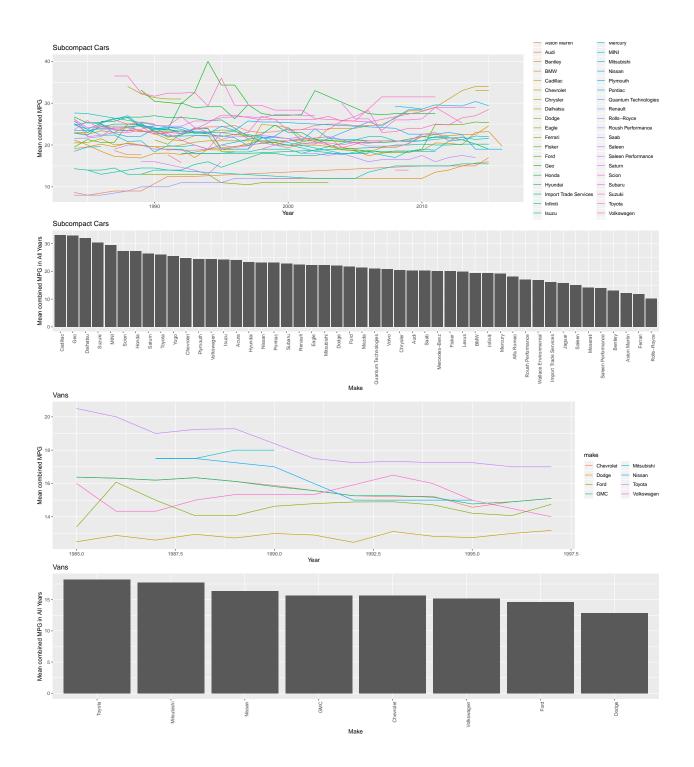
```
##
                            make
                                            model
         year
##
    Min.
            :1984
                     Chevrolet: 3633
                                        Length: 35708
##
    1st Qu.:1990
                     Ford
                              : 2958
                                        Class : character
                                2465
##
    Median:1999
                     Dodge
                                        Mode : character
##
    Mean
            :1999
                     GMC
                               : 2302
##
    3rd Qu.:2008
                     Toyota
                               : 1821
##
    Max.
            :2016
                     BMW
                               : 1518
##
                     (Other)
                              :21011
##
                              VClass
                                              cylinders
                                                                  displ
##
    Compact Cars
                                            4
                                                                     :0.600
                                  : 5160
                                                   :13594
                                                             Min.
    Subcompact Cars
                                                             1st Qu.:2.200
##
                                  : 4643
                                            6
                                                   :12518
##
    Midsize Cars
                                  : 4035
                                            8
                                                   : 7933
                                                             Median :3.000
##
    Standard Pickup Trucks
                                  : 2354
                                            5
                                                       759
                                                             Mean
                                                                     :3.328
    Sport Utility Vehicle - 4WD: 2090
                                                      505
##
                                            12
                                                             3rd Qu.:4.300
##
    Two Seaters
                                  : 1734
                                                       195
                                                             Max.
                                                                     :8.400
                                            (Other):
                                                      204
##
    (Other)
                                  :15692
##
                 trany
                                   city08
                                                  highway08
                                                                      comb08
                                                        : 9.00
##
    Automatic 4-spd:11026
                              Min.
                                      : 6.00
                                                Min.
                                                                  Min.
                                                                         : 7.00
```

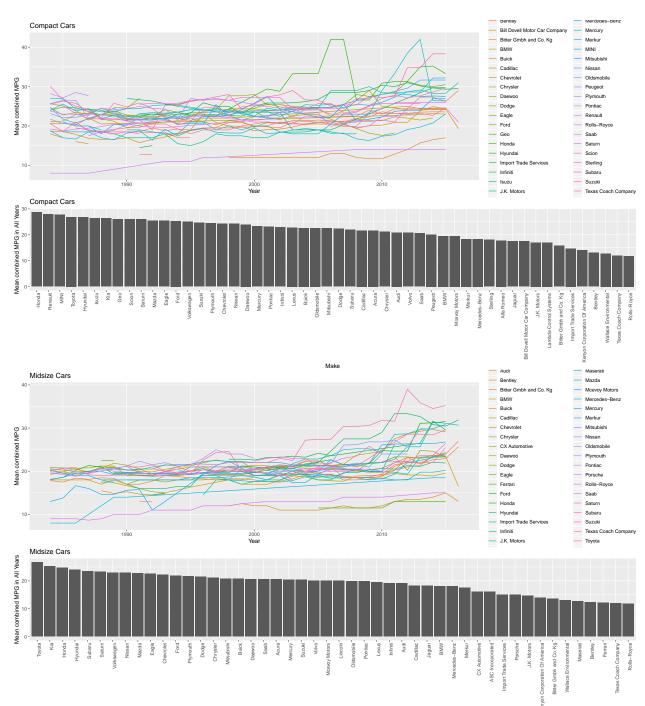
```
##
    Manual 5-spd
                    : 8250
                              1st Qu.:15.00
                                               1st Qu.:20.00
                                                                 1st Qu.:16.00
    Automatic 3-spd: 3151
##
                              Median :17.00
                                               Median :23.00
                                                                 Median :19.00
    Manual 6-spd
                    : 2206
                                      :17.54
                                                                         :19.79
##
                              Mean
                                               Mean
                                                       :23.68
                                                                 Mean
##
    Automatic (S6)
                   : 2201
                              3rd Qu.:20.00
                                               3rd Qu.:27.00
                                                                 3rd Qu.:22.00
##
    Automatic 5-spd: 2179
                              Max.
                                      :53.00
                                               Max.
                                                       :61.00
                                                                 Max.
                                                                         :53.00
    (Other)
                    : 6695
##
##
      vclass.ct
##
    Min.
           : 45
##
    1st Qu.:1143
##
    Median:2090
##
    Mean
            :2629
##
    3rd Qu.:4643
##
    Max.
            :5160
##
```

#### Fuel Economy of Vehicles of Different Makes (50 points)

For each vehicle class in filtered down data, we plot the mean combined MPG (average of data in vehicles\$comb08) for each vehicle maker every year. And then, we compute the mean combined MPG in all years for each vehicle maker, and plot it. Both charts are created with ggplot(). Note how the vehicle makers are ranked in the second plot. Use fig.width=16. To suppress messages from ggplot regarding groups with only one observation, set warning=FALSE, message=FALSE (we recommend setting this option only once your code is complete).







Make

