Analytics Foundations: Lab 11

1. Use the same cars2010 dataset you have used in previous labs. To obtain this data, submit the following code:

```
library(AppliedPredictiveModeling)

data(FuelEconomy)
```

   This dataset has variables pertaining to fuel economy of various cars. **Do not** create a training and test set. Just use the whole cars2010 dataset for the following analysis. The cars2011 and cars2012 datasets will be used at later time periods..

   Perform the following analysis:

   a. Run a LASSO regression predicting the **FE** variable using all the remaining variables. Some of these predictor variables are coded as numeric, but should be treated as categorical. The only numeric variables in your dataset should be **EngDispl**. All remaining variables are categorical.
      a. Plot the coefficients and how they change across different levels of lambda.

   b. Perform a CV LASSO to optimize the lambda value.
      a. What is the value of lambda that minimizes the MSE?
      b. What is the value of lambda one standard error above the minimum MSE value?
      c. How many variables are left at the penalty value that is one standard error above the minimum MSE value (think of variables as a whole, not per category)? (HINT: Look at the coefficients from the model with `coef` function.)

   c. Obtain the variables from the LASSO regression at the penalty value that is one standard error above the minimum MSE value. The multiple linear regression with p-value selection (Lab 6) left the variables **EngDispl**, **NumCyl**, **Transmission**, **AirAspirationMethod**, **NumGears**, **TransLockup**, **DriveDesc**, **IntakeValvePerCyl**, **CarlineClassDesc**, and **VarValveLift**.
      a. What variables were left in your LASSO at the 1SE above minimum MSE penalty value?
      b. How does this compare against the variables from the multiple linear regression?