library(ggplot2)

library(tidyverse)

library(glmnet)

library(AppliedPredictiveModeling)

# Read in the cars datasets

data(FuelEconomy)

# Build the predictor and target elements for modeling

train\_x <- model.matrix(FE ~ EngDispl +

factor(NumCyl) +

Transmission +

AirAspirationMethod +

factor(NumGears) +

factor(TransLockup) +

factor(TransCreeperGear) +

DriveDesc +

factor(IntakeValvePerCyl) +

factor(ExhaustValvesPerCyl) +

CarlineClassDesc +

factor(VarValveTiming) +

factor(VarValveLift),

data = cars2010)[, -4]

train\_y <- cars2010$FE

# Build a LASSO regression

cars\_lasso <- glmnet(x = train\_x, y = train\_y, alpha = 1)

plot(cars\_lasso, xvar = "lambda")

# Perform a CV to select the optimal LASSO regression penalty

cars\_lasso\_cv <- cv.glmnet(x = train\_x, y = train\_y, alpha = 1)

plot(cars\_lasso\_cv)

cars\_lasso\_cv$lambda.min

cars\_lasso\_cv$lambda.1se

plot(cars\_lasso, xvar = "lambda")

abline(v = log(cars\_lasso\_cv$lambda.1se), col = "red", lty = "dashed")

abline(v = log(cars\_lasso\_cv$lambda.min), col = "black", lty = "dashed")

# Plot the important variables' coefficients at 1 SE above minimum penalty

coef(cars\_lasso, s = cars\_lasso\_cv$lambda.1se) %>%

broom::tidy() %>%

filter(row != "(Intercept)") %>%

ggplot(aes(value, reorder(row, value))) +

geom\_point() +

ggtitle("Influential Variables") +

xlab("Coefficient") +

ylab(NULL)