For your third assignment, you will continue to multiple Linear Regression models and also begin to look at Classification Models. The data sets:

**Predicting Car Trade In Value:** Returning to this model you worked on last week, look for collinearity (you may use the VIF function in the “car” package), high leverage observations, and nonlinear relationships. As before, train you model using the training data, and then try to estimate the test MSE using the test data.

Recommendations:

* You probability eliminated most of the collinearity last week, but if you can get the VIF function to work, you can check on the predictors you have left.
* Use residual plots to look for nonlinear influences from the predictors. Try cross terms as well.
* You may also want to use studentized residuals to check for outlier observations in the training data that should not be included in the model.
* You can check leverage with the hatvalues() function  see lab 3.6.2 on page 113 of the book.

[Car Worth - Train(1).csv](https://elearning.utdallas.edu/bbcswebdav/pid-1054384-dt-content-rid-10269355_1/xid-10269355_1) , [Car Worth - Test(1).csv](https://elearning.utdallas.edu/bbcswebdav/pid-1054384-dt-content-rid-10269356_1/xid-10269356_1)

**Predicting Carbon Monoxide Levels in Cigarettes:** The data for this study is given in cigarttes-data.csv. The columns are Brand, Tar (mg), Nicotine (mg), Weight (g), CO (mg). The response variable is CO (Carbon Monoxide). Your goal is to build  linear regression model that predicts the CO level using the other predictor variables.

Recommendations:

* Think about the predictors. Is brand a predictor – or just an identifier?
* Look for outliers
* Look for collinearity

[cigarettes-data.csv](https://elearning.utdallas.edu/bbcswebdav/pid-1054384-dt-content-rid-10269357_1/xid-10269357_1)

**Predicting Low Birth Weight:** The data for this study is included in the file lowbwt.csv. The goal of this study was to identify risk factors associated with giving birth to a low birth weight baby (weighing less than 2500 grams). Data were collected on 189 women, 59 of which had low birth weight babies and 130 of which had normal birth weight babies.  The variables are contained in the file lowbwt data dictionary.txt.

Recommendations:

* Create a logistic model that predicts the LOW variable in terms of the other variables. You probably want to leave BWT out, since that is not known before the birth.
* Try to identify the key predictor variables that are important

[lowbwt.csv](https://elearning.utdallas.edu/bbcswebdav/pid-1054384-dt-content-rid-10269358_1/xid-10269358_1) , [lowbwt-data dictionary.txt](https://elearning.utdallas.edu/bbcswebdav/pid-1054384-dt-content-rid-10269359_1/xid-10269359_1)