All,

Your goal for this assignment will be to apply the subset selection and dimension reduction techniques we covered in Chapter 6 to a new dataset to create the best model possible. We will also apply some of the nonlinear techniques we discussed at the beginning of Chapter 7. The datasets for this week:

Ames Iowa Housing Data: This is a large dataset containing information about houses sold in Ames, Iowa between 2006 and 2010. Your job is to build a model that will predict the selling price for the house based upon predictor variables contained in the data. Note that the data contains several predictors with many NA values – I would suggest first doing a summary on the entire dataset to see which predictors have a large number of NAs, and removing these predictors. Then use na.omit to eliminate observations that have NAs in other predictors. You should also read the data description file to see how to eliminate a few outlier observations. Apply the subset selection techniques we discussed as well as dimension reduction; I suggest using cross validation to select the best model.

Battery Life: This dataset contains lifetime and voltage readings for two types of batteries. Your task is to build a model for the Voltage as a function of the Time for each battery type (there will be two models). Use the subset() function in R to break the dataset up into two subsets (one for each battery type), and then use splines to fit a model to each dataset.

Please submits the results of your analysis by midnight on Sunday, March 6.

Best,
Dr Semper