Load the NHANES dataset and use the view and slicing techniques to look at the dataset. How many column vectors do you see? How many rows are in the dataset? How many numerical column vectors do you see? For each potential predictive column vector find the: (a) minimum value, (b) maximum value, (c) median value, (d) mean value, (e) standard deviation (f) number of unique values. Why is this difficult give the data in its current state? Which column vectors do not work for (a)-(e)? What is wrong with them? Treating BMI as the Y value perform linear regression on each of the remaining column vectors and plot the result using the ggplot2 plot code in the CH3 JNB. Which variable has the best fit? Why is the fit so good do you think? If the method fails for some variables, why do you think that happens?

Load the NHANES dataset again and use the view or row index slicing as shown in examples to characterize the NHANES datset better. Now find all the potential predictor variables that are suitable for: (a) linear regression (b) logistic regression (c) Poisson regression. Don’t forget to explain why they are suitable (will need to show proof such as looking at the levels or running some summary statistics.

For the subset of variables that will work with Linear Regression select a 3-4 of interest (X1, X2, X3... ) to use as predictors for the Y variable of your choice. Do pair-wise correlations for the X variables and then build a multiple linear regression model with interaction terms for any correlations over 0.25. Next try to find a minimal form of X variables (1 or 2) and see if the predictive model is good.

For the subset of variables select a 3-4 of interest (X1, X2, X3... ) to use as predictors for a suitable Y variable of your choice using logistic regression. Do pair-wise correlations for the X variables and then build a multiple logistic regression model with interaction terms for any correlations over 0.25. Next try to find a minimal form of X variables (1 or 2) and see if the predictive model is good.

For the subset of variables select a 3-4 of interest (X1, X2, X3... ) to use as predictors for a suitable Y variable of your choice using Poisson regression. Do pair-wise correlations for the X variables and then build a multiple logistic regression model with interaction terms for any correlations over 0.25. Next try to find a minimal form of X variables (1 or 2) and see if the predictive model is good.