Use the dataset bwght.csv for this exercise.

Define a binary variable *smokes* if the woman smokes during pregnancy. (There are a lot of variables that should be factors and vice versa, you'll need to clean up the dataset a little first).

1) Estimate a logit model relating *smokes* to *motheduc*, *white*, *lfaminc.* Find the LOOCV and the 10-fold CV classification error rate for this model.

2) Augment this model by adding variables *fatheduc* and *cigprice*. Does this model perform better in terms of LOOCV and the 10-fold CV classification error rate?

3) Divide the data into a training set and a test set. Use the better fitting model of the two to generate a confusion matrix and find the Test Error Rate.

4) Find the bootstrapped coefficients and standard errors using the `boot()` function for the better fitting model.

5) Finally, use the `caret` package to generate a cross-validated k-nearest neighbors model using the predictors in the baseline model. Generate a confusion matrix and find the Test Error rate of this model.