**HW#7 (Due Apr 27th)**

Consider the cps_income dataset found in the 0.Data folder, which contains data from "current population survey". As in HW#6, we will be predicting whether an individual makes MORE or LESS than 50k per year.

**DOWNLOAD THE DATASET, IMPORT INTO Rstudio AND**
**RUN THIS CODE FIRST:**
YOUR_DATA_NAME = data.frame(YOUR_DATA_NAME)
**THEN IDENTIFY AND USE THE FIVE QUANTITATIVE (CONTINUOUS) VARIABLES.**

**Note:** You must include code and any output.

1) Generate an SVM model using the numeric variables you used in HW6 Q2 for all the observations to predict the value of the INCOME column. Answer this question with a SVM plot for two predictor variables of your choosing.

2) Using the same assignment of observations to a training and testing set as in HW6 Q2, generate the SVM model using the training set and obtain the misclassification rate on the testing set. Comment on which of your models in HW6 Q1(e), Q2, and here seems to perform the best.
**Note:** If you cannot obtain the same assignment of observations to a training and testing set as in HW6 Q2, randomly assign 30,000 observations to a training set and the rest to a testing set.

3) Generate two more SVM models (one by changing the cost parameter and another by changing the kernel type) for the TRAINING data. Obtain misclassification rates for both based on the TESTING data and decide which of the three SVM models (including the model from (2)) you prefer to use based on the three misclassification rates.
**Note:** If the model takes longer than 10 minutes to run, you should probably "stop" the process and start again with a different cost or kernel value.

4) Obtain an average 5-fold cross validation misclassification rate for your chosen model from question (3).

5) (a) Compute the False Positive Rate and False Negative Rate for the SVM model in (2) on the " <=50K" class.

   (b) Compute the False Positive Rate and False Negative Rate for the SVM model in (2) on the " >50K" class.