Due on November 9th, 2018 (No Quiz)

Submit wordprocessed report with font size 11. No screenshots. For last two questions a colored output is required.

1. The `hsb` dataframe from library `faraway` was collected as a subset of the High School and Beyond study conducted by the National Education Longitudinal Studies program of the National Center for Education Statistics. The variables are gender; race; socioeconomic status (SES); school type; chosen high school program type; scores on reading, writing, math, science, and social studies. We want to determine which factors are related to the choice of the type of program (academic, vocational or general) that the students pursue in high school. The response `prog` is multinomial with three levels.

   a) Make a table showing the proportion of males and females choosing the three different programs. Comment on the difference. Repeat this comparison but for SES rather than gender.

   b) Fit a multinomial response model for the program choice and examine the fitted coefficients. Of the five subjects, one gives unexpected coefficients. Identify this subject and suggest an explanation for this behavior.

   c) Use the BIC method to reduce the model. Which variables are in your selected model?

   d) Fit a multinomial model with two predictors `math` and `science`. Make a 2-in-1 plot showing the observed and predicted responses with different colors. Find the overall error rate.

   e) Fit a KNN model to predict the program of choice `prog` with two predictors `math` and `science`. Make a 2-in-1 plot showing the observed and predicted responses with different colors. Find the overall error rate. Report results for $k = 3$ and 5.