**Homework 2**

Use caret for all the models except where mentioned to use another package. Review examples provided on Canvas before attempting homework. Please turn in a Quarto (qmd) file or a word/pdf document generated from the qmd file or both.

1. Load the College.csv file on Canvas into a dataframe called college (or give it any name you want). A description of the dataset can be found at <https://www.rdocumentation.org/packages/ISLR2/versions/1.3-1/topics/College>
2. Check if there are any missing values. Replace any missing values with the mean value for the column rounded up to the nearest integer. Hint: Lookup the ceiling function and use it.
3. Remove the college name column from the dataframe as it is not useful in prediction
4. Split the dataset into 80% training and 20% test
5. Fit a linear multiple regression model to the training set to predict graduation rate using all the other features/variables. What variables are significant at the p = 0.001 level? Predict the graduation rate using the test dataset and report the root mean squared error (RMSE) for the test dataset.
6. Perform backward subset selection using the leaps library. Which model is best according to BIC? Are the variables chosen by the subset selection different from the list of statistically significant variables from fitting the linear regression model above? If so, which ones are different?
7. Fit a ridge regression model on the training set for predicting graduation rate using all the predictors, with *λ* chosen by 5-fold repeated cross-validation with 5 repeats. Predict graduation rate using the test dataset and report the test root mean squared error (RMSE). Report the value of *λ* used in the model.
8. Run the same ridge regression model above but this time use a grid of lamdas to search over. Use a grid of 50 values ranging from λ = 0.001 to λ = 10000. Report the λ that was chosen.
9. Fit a Generalized Additive local regression (LOESS) model on the training set for predicting graduation rate using all the predictors. Do not specify a span so caret can choose a span. Report the span chosen by caret. Predict graduation rate in the test dataset and report the test root mean squared error (RMSE).
10. Fit a Generalized Additive Spline regression model on the training set for predicting graduation rate using all the predictors. Do not specify a degree of freedom and let caret choose it. Report the degree of freedom chosen by caret. Predict graduation rate in the test dataset and report the test root mean squared error (RMSE).