Tasks for day 5:

- Set up a new R project: day5

- load in the Boston housing data

- divide the data into training and test set (90% training, 10% testing)

**Read the info about the data set and explore the contents (and the relationships between the variables) by using methods we covered last week.**

Work in groups of 3-4 and put together a report (R Notebook) to carry out EDA (exploratory data analysis) and the results of a multivariate fit analysis. Your response variable in the set is:

**medv**

come up with the following models using the train data:

1. Bivariate linear model (1 response variable, 1 quantitative predictor)

2. Multivariate linear model (1 response variable, 1 quantitative predictor, 1 qualitative predictor)

3. Multivariate linear model that allowed interactions between predictors

4. All possible predictors (1 response, all predictors)

5. Amodel with an optimal set of predictors (1 response, X predictors) (explore stepwise/exhaustive search methods to select predictor variables)

- Use your models to predict medv on the test data and compare the performance of your models train vs test (R², Adjusted R²)

- Choose two models and compare MSE, RMSE