Lab 4: Model Assessment

We will use the `Auto` data set in the `ISLR` package.

```
library(ISLR)
library(tidyverse)
library(knitr)
library(tidymodels)

head(Auto) %>%
  kable()
```

mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin	name
18	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu
15	8	350	165	3693	11.5	70	1	buick skylark 320
18	8	318	150	3436	11.0	70	1	plymouth satellite
16	8	304	150	3433	12.0	70	1	amc rebel sst
17	8	302	140	3449	10.5	70	1	ford torino
15	8	429	198	4341	10.0	70	1	ford galaxie 500

Before we begin, be sure to set the seed for reproducibility.

```
set.seed(445)
```

0.1 Validation Set Approach

1. Split the data into 50% training and 50% test data.
2. Fit a linear model of `mpg` on `horsepower` using your training data.
3. Estimate the test error by using test MSE.
4. Repeat steps 2-3 for a cubic and quadratic model. Which model would you pick?
5. Repeat steps 1-4 after reseting the seed

```
set.seed(42)
```

6. Did you get the same results? Is this what you expected to happen?

0.2 LOOCV

1. Get the estimate of test MSE for the linear model using LOOCV.
2. Repeat steps 2-3 for a cubic and quadratic model. Which model would you pick?

0.3 k-Fold CV

1. Using $k = 10$-fold CV, compute the k-fold CV estimate of the test MSE for polynomial models of order $i = 1, \ldots, 10$. (Hint: you can use the `poly` function in your formula to specify a polynomial model.)
2. Plot the estimated test MSE vs. the polynomial order.
3. Which of these models would you choose?

0.4 Bonus

1. Write your own k-fold CV function that will calculate CV for the KNN Regression model. You function should take as parameters
   - CV k value
   - KNN K value
   - Data
   - A vector of names (character) of predictor columns
   - A character string of the response column

   And return the estimated test MSE.
2. Use your function to estimate the test MSE using 10-fold CV for KNN models with

$K = 1, 5, 10, 20, 100$ of a model predicting `mpg` using the `horsepower` predictor variable in the `Auto` data set.

3. Compare your results to the previous k-Fold CV method.