We will build two models to predict the variable `cnt` from the `bike` dataset found on my website. We'll compare their ability to predict future data using k-fold cross validation.

1. Load the library `caret`.
2. Read in the `bike` dataset from my website. Read the help file associated with this dataset.
3. Write yourself a Mean Squared Error function. This function should have signature `MSE(y, yhat)` and should return a single number:$\text{MSE} = N^{-1}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2$.
4. Call the function `caret::createFolds` on the vector `cnt` and store the output in a variable named `folds`.
5. Recall how to extract a vector from a list that contains multiple vectors.
6. Create two vectors `mse_mean` and `mse_anova` using the function `rep`. These vector should be filled with K, the number of folds, `NA`s.
7. Write a for loop around each fold in your variable `folds`. Within each iteration of the for loop you should
   (a) Create two datasets: `training` and `testing`. The dataset `training` should contain all data but the current fold. The dataset `testing` should contain only the data from the current fold.
   (b) Calculate the mean of the variable `cnt` based on the `training` dataset. This is your first predictive model; a model that always predicts the mean, as calculated on the `training` dataset. Let's refer to this value as `yhat`.
   (c) Call your function `MSE` on `cnt` from the `testing` dataset, `y`, and your first model's predicted value (from the `training` dataset) `yhat`. Store this value into a vector `mse_mean`.
   (d) Calculate ANOVA on an appropriately recognized categorical variable of your choice from the `training` dataset.
   (e) Use the function `predict` to calculate `yhat` from this ANOVA model based on the `testing` dataset.
   (f) Call your function `MSE` on `cnt` from the `testing` dataset, `y`, and this ANOVA model's predicted value based on the `testing` dataset, `yhat`. Store this value into a vector `mse_anova`.
8. After the for loop, find the mean of the vectors `mse_mean` and `mse_anova`.
9. Which mean MSE is smaller?
10. Are these MSEs reasonable quantifications of your model's ability to predict future (not your current) data? Why or why not?
11. Why is ANOVA a better model?