Homework 10 (Optional)

Instructions

Due: 5:00pm on Thursday, December 16th

1. Add your name between the quotation marks on the author line in the YAML above.
2. Compose your answer to each problem between the bars of red stars.
3. Commit your changes frequently.
4. Be sure to knit your .Rmd to a .pdf file.
5. Push both the .pdf and the .Rmd file to your repo on the class organization before the deadline.

Regression Competition Finale

We return one final time to the data set from Homework 3 to build the best possible regression model, using the full power of the tools developed in this course.

As before, you will construct your model using a training data set with information on 66 variables recorded for 1808 houses. I've held back the data on 600 other houses which will serve as the test data set for assessing the predictive accuracy of your model.

You should record your answers in this .Rmd file. However, you are encouraged to use a separate .Rmd file for scratchwork. The assignment is divided into several **Components** to help organize your work. Put all work you want graded between the bars of red stars in the corresponding section.

Grading

This assignment is entirely **optional**. Your model performance (measure in rMSE) will be compared against a few benchmark models that I've made on the same data, and you will be awarded extra credit points corresponding to how well your model does. The minimum extra credit awarded represents approximately 20% of a typical homework assignment, and can be achieved by building a model which exceeds the accuracy of the full linear model on test data.

The Data

The training data (predictors and response) as well as test data (predictors only) can be imported by running the following code:

```
house_train<-read_csv("house_train.csv")
house_test<-read_csv("house_test.csv")
```

Additionally, the `data_description.txt` file in the hw_10 repo gives a full description of the variables appearing in the data set.

There is one special column of note:

- `Sale_Price` is your response variable and should not be included as a predictor.

Your Model

Use the space below to construct your model.

Your Predictions

Create a data frame with a single column of your 597 predictions for `Sale_Price` on the test data. Label the column `preds` and save the data frame as `my_preds`. If you performed any variable transformations on the response, be sure you transform your predictions back to the appropriate units.

Export your Predictions

Run the following code chunk to create a .csv file in your repo containing a data frame of your predictions. To do so:

1. Remove the # symbol
2. Replace `FirstName_LastName` with your actual first and last name

```
# write_csv(my_preds, "FirstName_LastName_preds.csv")
```
