***Part 2***

TRY ITS

* GrLivArea is a variable that holds the size of the living areas for the houses
* Regression on SalePrice of GrLivArea:

A screenshot of a cell phone

Description automatically generated

* Plot of lm.fit:

A close up of a map

Description automatically generated

* Regression on SalePrice of GrLivArea controlling for LotArea:

A screenshot of text

Description automatically generated

A close up of a map

Description automatically generated

Controlling for Lot Area does not change the qualitative conclusions made by the previous regression but does change the quantitative results as the F statistic is lowered significantly but maintains the same p value.

EXERCISE 2

1. A screenshot of a cell phone

   Description automatically generated
   1. Having an outdoor garage is estimated to have a lower Sale Price by $72,859 than a house with an indoor garage.
2. A close up of a newspaper

   Description automatically generated
   1. By looking at the F statistic and the Multiple/adjusted R^2, there does appear to be a relationship between the predictors and the response. Adding all of the variables in the data set, it can be seen that not all of the predictors are statistically significant but a good amount of them are indicating a relationship with SalePrice.
   2. The predictors that have statistically significant relationships with SalePrice are LotArea, OverallQual, OverallCond, YearBuilt, X1stFlrSF, X2ndFlrSF, BedroomAbvGr, BsmtFullBath, and GarageCars.
   3. The coefficient for the year variable suggests that for every one year newer that the house is (one year increase from when it was built), the price fo the house increases by $425,800.
3. A close up of a map

   Description automatically generated A close up of a map

   Description automatically generated
   1. The fit seems to more or less be pretty good especially since we know that the R^2 is around 0.7. That being said, there do appear to be some large outliers towards the upper right hand corner of the graph, indicating higher residuals and fitted values than the majority than the rest of the data points. There are also outliers at the bottom right hand corner which display high fitted values but low residuals. The points centered around 2e+05 fitted value and a residual below zero seem to have larger leverage in pulling the fitted line below zero to the negative residual half. At larger fitted values, the points start to trend upward with residuals indicating high leverage in pulling up the fit line.
4. A close up of a newspaper

   Description automatically generated A close up of a newspaper

   Description automatically generated A close up of a newspaper

   Description automatically generated
   1. The only statistically significant interaction term in the three models above is that between LotArea (lot area), X1stFlrSF(Square footage of the first floor), and X2ndFlrSF(Square footage of the second floor), demonstrated through the ‘\*\*\*’ next to the predictor.
5. 1. LOG TRANSFORMATION

A screenshot of a cell phone

Description automatically generated

* 1. SQUARED TRANSFORMATION

A close up of a map

Description automatically generated

* 1. A close up of a map

     Description automatically generated

The log and the square root transformations both brought the residual and fitted values to smaller values but the log transformation linearized the graphs the most, so it would be noteworthy to include a log transformation over the other two transformations in a model of SalePrice.