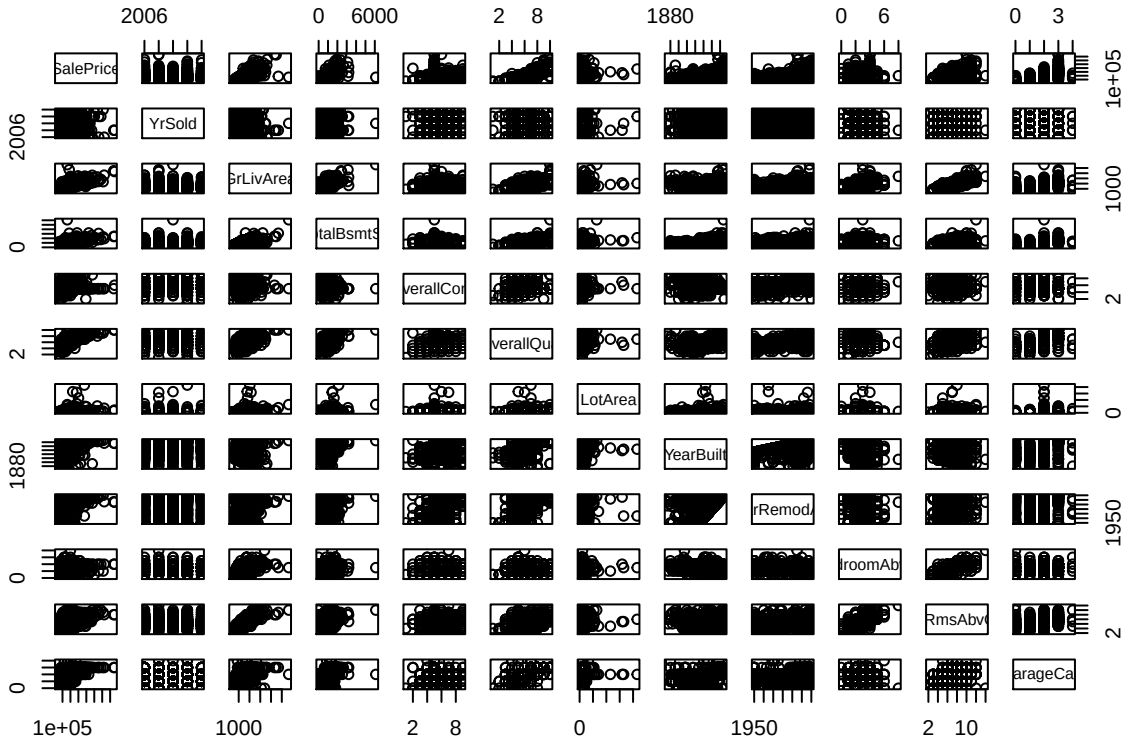
Lab 2

Group 17

1/29/2020

Exersize 1

Below is a visualization of our scatterplot matrix between 12 variables we believed to be associated with Sale Price. It is a little difficult to read in a pdf format, so if you are interested in getting a better view of each plot, run our rscript in github. The variables from left to right are sale price, year sold, total above ground living, total basement square footing, overall condition, overall quality, lot area, year built, year remodelled, bedrooms above ground, total rooms above ground, and garage size



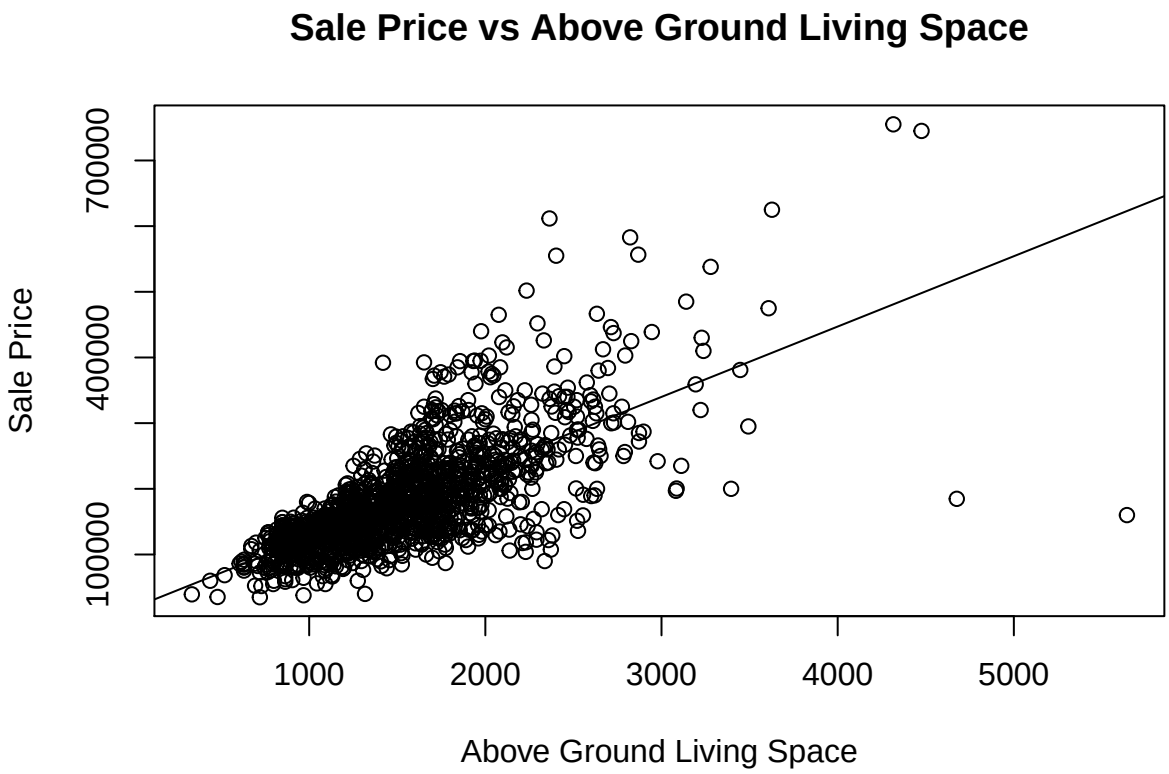
In addition to these plots, we've included a matrix of correlation between the variables. The results can be seen below

```
##                SalePrice      YrSold   GrLivArea TotalBsmtSF OverallCond
## SalePrice     1.00000000 -0.02892259  0.70862448  0.61358055 -0.07785589
## YrSold       -0.02892259  1.00000000 -0.03652582 -0.01496865  0.04394975
## GrLivArea     0.70862448 -0.03652582  1.00000000  0.45486820 -0.07968587
## TotalBsmtSF   0.61358055 -0.01496865  0.45486820  1.00000000 -0.17109751
## OverallCond  -0.07785589  0.04394975 -0.07968587 -0.17109751  1.00000000
## OverallQual   0.79098160 -0.02734671  0.59300743  0.53780850 -0.09193234
## LotArea       0.26384335 -0.01426141  0.26311617  0.26083313 -0.00563627
```

```
## YearBuilt     0.52289733 -0.01361768  0.19900971  0.39145200 -0.37598320
## YearRemodAdd  0.50710097  0.03574325  0.28738852  0.29106558  0.07374150
## BedroomAbvGr  0.16821315 -0.03601389  0.52126951  0.05044996  0.01298006
## TotRmsAbvGrd  0.53372316 -0.03451635  0.82548937  0.28557256 -0.05758317
## GarageCars    0.64040920 -0.03911690  0.46724742  0.43458483 -0.18575751
##               OverallQual     LotArea   YearBuilt YearRemodAdd BedroomAbvGr
## SalePrice     0.79098160  0.26384335  0.52289733   0.50710097   0.16821315
## YrSold       -0.02734671 -0.01426141 -0.01361768   0.03574325  -0.03601389
## GrLivArea     0.59300743  0.26311617  0.19900971   0.28738852   0.52126951
## TotalBsmtSF   0.53780850  0.26083313  0.39145200   0.29106558   0.05044996
## OverallCond  -0.09193234 -0.00563627 -0.37598320   0.07374150   0.01298006
## OverallQual   1.00000000  0.10580574  0.57232277   0.55068392   0.10167636
## LotArea       0.10580574  1.00000000  0.01422765   0.01378843   0.11968991
## YearBuilt     0.57232277  0.01422765  1.00000000   0.59285498  -0.07065122
## YearRemodAdd  0.55068392  0.01378843  0.59285498   1.00000000  -0.04058093
## BedroomAbvGr  0.10167636  0.11968991 -0.07065122  -0.04058093   1.00000000
## TotRmsAbvGrd  0.42745234  0.19001478  0.09558913   0.19173982   0.67661994
## GarageCars    0.60067072  0.15487074  0.53785009   0.42062215   0.08610644
##               TotRmsAbvGrd  GarageCars
## SalePrice      0.53372316  0.64040920
## YrSold        -0.03451635 -0.03911690
## GrLivArea      0.82548937  0.46724742
## TotalBsmtSF    0.28557256  0.43458483
## OverallCond   -0.05758317 -0.18575751
## OverallQual    0.42745234  0.60067072
## LotArea        0.19001478  0.15487074
## YearBuilt      0.09558913  0.53785009
## YearRemodAdd   0.19173982  0.42062215
## BedroomAbvGr   0.67661994  0.08610644
## TotRmsAbvGrd   1.00000000  0.36228857
## GarageCars     0.36228857  1.00000000
```

As we can see above the correlation factor relating each variable to one another is displayed, however, we are most concerned with the first column relating each variable to sale price. The variables with strong correlative relationships are above ground living area and overall quality. Variable with moderate correlative relationshios include total basement square footage, year built, year remodelled, total rooms above ground, garage size. The ones that didn't quite match our beliefs with weaker relationships were year sold, overall condition, lot area, and bedrooms above ground. Most of the correlations matched our earlier beliefs, all the variables that imply a bigger house are strongly correlated with an increase in sale price,newer and higher quality houses also sold for more, we expected condition to be positively correlated with sale price, we am unsure why this has a negative correlation, perhaps because condition is negatively correlated with basement size and YearBuilt. We expected Yrsold to be positively correlated with the saleprice but considering the timing of the dataset around the recession a negative correlation makes sense.

Now lets look at the relationship between sale price and above ground living space. Below us a scatterplot and simple linear regression model plotted alongside it.




```
## integer(0)
```

As we can see there are a few outliers to this scatterplot. After calculating the absolute value of the residuals, there is one point in particular (the one on the bottom right corner of the plot) that is the other outlier. This point has an above ground living space of 5642 and a sale price of 16000. According to the model we would predict that with this GrLivArea of about 60000 for a residual of -462998.5. To see why this sale price was so low, we compared the other characteristic to the mean of each column. Factors that could have pulled the sale price down was the small lot area, however all the other characteristics of the line are in line or even better than the mean. For this reason, we can assume that this was some kind of favor for a family member or friend, or there could be some other type of payment not related to sale price associated with the house.

Exersize 2

After creating a regression model with sale price as the response and Garage Outside being the predictor we get the following:

```
ameslist <- fastDummies::dummy_cols(ameslist, select_columns = c("GarageType"))
ameslist$GarageOutside <- ifelse(ameslist$GarageType_Detchd == 1 | ameslist$GarageType_CarPort == 1, 1,
lmGarage.fit = lm(SalePrice ~ ameslist$GarageOutside)
summary(lmGarage.fit)
```

```
##
## Call:
```

```
## lm(formula = SalePrice ~ ameslist$GarageOutside)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -150409  -44237  -13043   25098  548598
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)               206402       2291   90.08   <2e-16 ***
## ameslist$GarageOutside    -72859       4276  -17.04   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 71840 on 1377 degrees of freedom
##   (81 observations deleted due to missingness)
## Multiple R-squared:  0.1741, Adjusted R-squared:  0.1735
## F-statistic: 290.3 on 1 and 1377 DF,  p-value: < 2.2e-16
```

After running this we see GarageOutside has a coeficient of -72859. This means that the model predicts if a house has a garage that is not connected to the house in some way, it will devalue the sale price by 72859.

Lets now explore a regression model with more predictors.The predictors contain all variables defined buy our Ames text file.

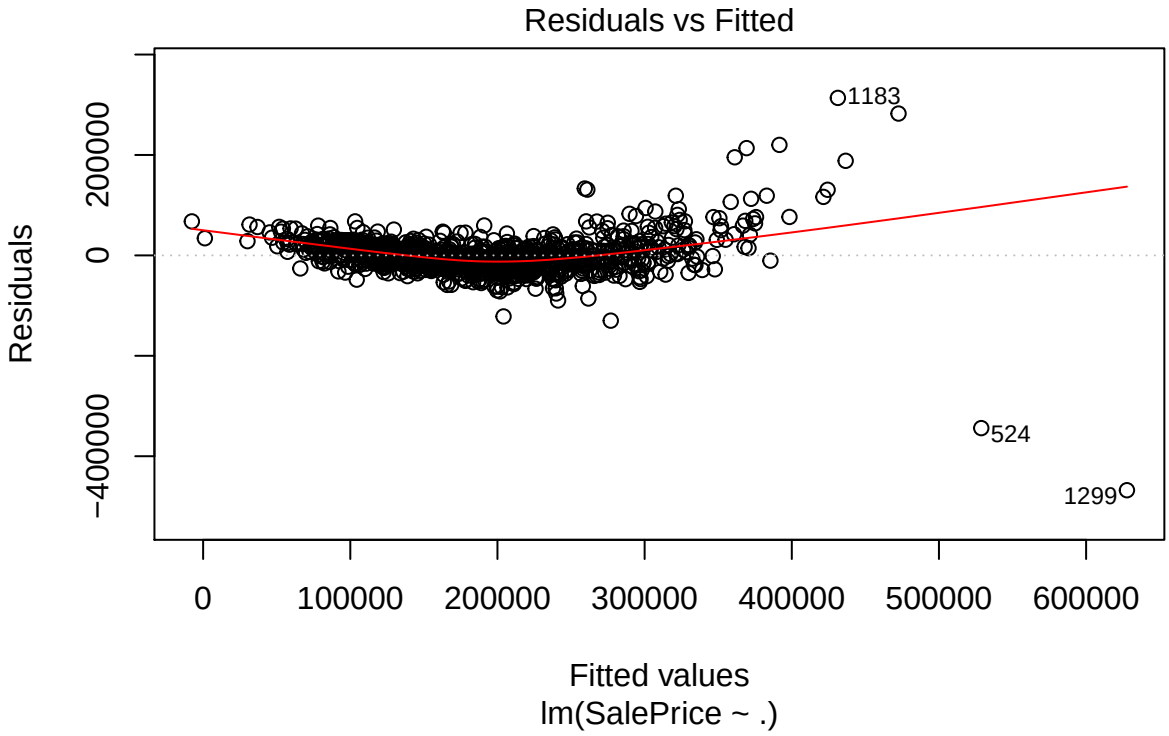
```
sp_model <- lm(SalePrice ~ ., data = Ames)
summary(sp_model)
```

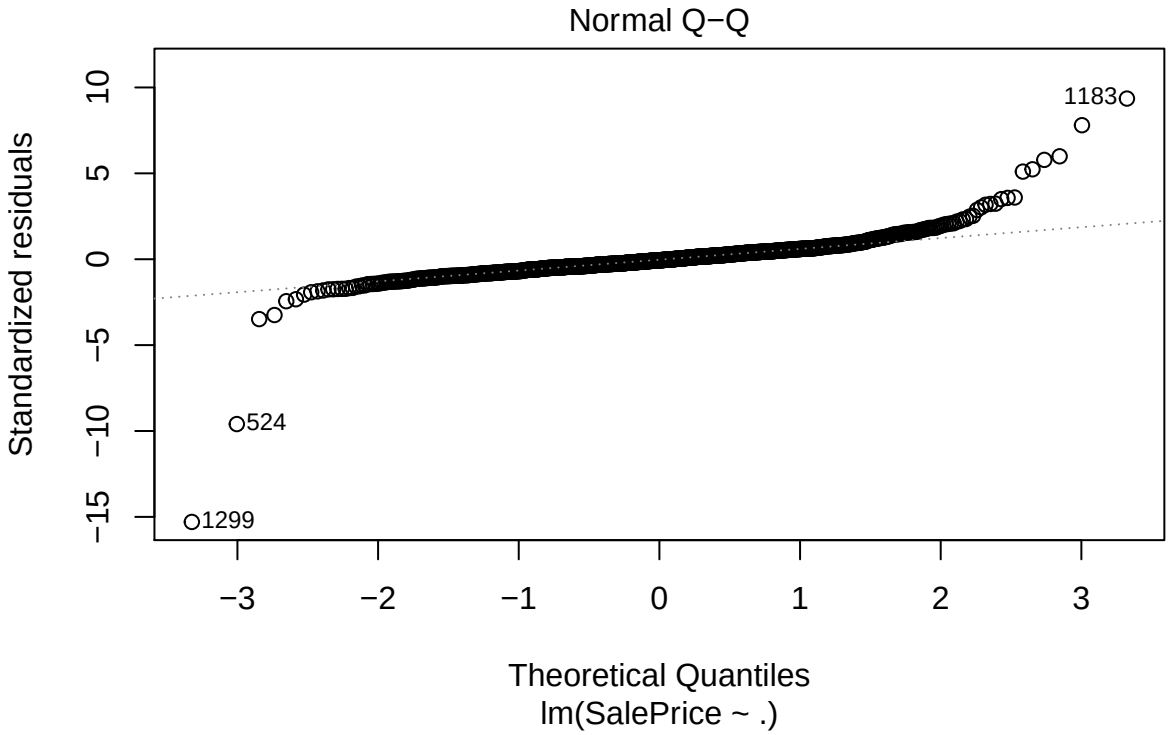
```
##
## Call:
## lm(formula = SalePrice ~ ., data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -467752  -16792   -2180   14737  313676
##
## Coefficients:
##                    Estimate   Std. Error t value    Pr(>|t|)
## (Intercept)   -165164.5301 1735624.0125  -0.095    0.924204
## Id                 -2.2838       2.7023  -0.845    0.398217
## LotFrontage         4.2922      59.1989   0.073    0.942213
## LotArea             0.5270       0.1598   3.297    0.001007 **
## OverallQual     18663.9699    1498.8854  12.452     < 2e-16 ***
## OverallCond      5610.1322    1393.0492   4.027 0.000060325 ***
## YearBuilt         356.2908      88.8156   4.012 0.000064406 ***
## YearRemodAdd      102.1997      88.2394   1.158    0.247031
## BsmtUnfSF         -12.2739       3.9076  -3.141    0.001729 **
## TotalBsmtSF        21.0194       5.8508   3.593    0.000342 ***
## X1stFlrSF          26.9960      29.2112   0.924    0.355604
## X2ndFlrSF          20.2645      28.6139   0.708    0.478968
## GrLivArea          21.9289      28.5270   0.769    0.442233
## BsmtFullBath     6911.3490    3250.1017   2.127    0.033684 *
## BsmtHalfBath      508.1952    5168.8700   0.098    0.921697
## FullBath         3371.4809    3586.1907   0.940    0.347359
## HalfBath         -823.5253    3367.7622  -0.245    0.806865
```

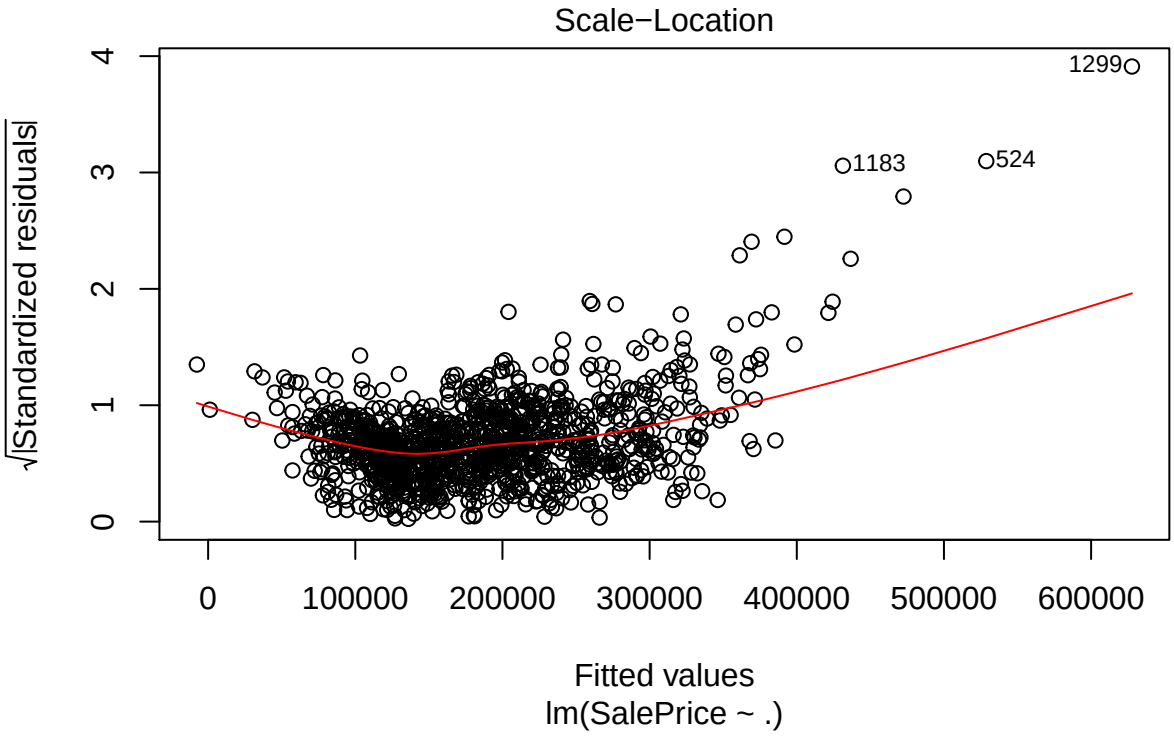
```
## BedroomAbvGr    -9382.9562    2177.9509  -4.308 0.000017939 ***
## KitchenAbvGr   -34641.1444    6473.0758  -5.352 0.000000106 ***
## TotRmsAbvGrd     6271.2513    1509.5892   4.154 0.000035169 ***
## Fireplaces       3839.9791    2227.2889   1.724    0.084979 .
## GarageYrBlt       -98.2759      92.7365  -1.060    0.289500
## GarageCars      17600.2393    3567.7720   4.933 0.000000935 ***
## GarageArea         14.1265      12.3418   1.145    0.252623
## WoodDeckSF         23.0022      10.1956   2.256    0.024261 *
## OpenPorchSF        -9.1742      19.7157  -0.465    0.641793
## EnclosedPorch       5.5595      21.0431   0.264    0.791678
## ScreenPorch        55.8002      20.5785   2.712    0.006801 **
## PoolArea          -84.3404      30.1707  -2.795    0.005273 **
## MoSold            -67.9712     429.5609  -0.158    0.874301
## YrSold           -313.8101     863.0879  -0.364    0.716234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37670 on 1096 degrees of freedom
##   (333 observations deleted due to missingness)
## Multiple R-squared:  0.8005, Adjusted R-squared:  0.7951
## F-statistic: 146.6 on 30 and 1096 DF,  p-value: < 2.2e-16
```

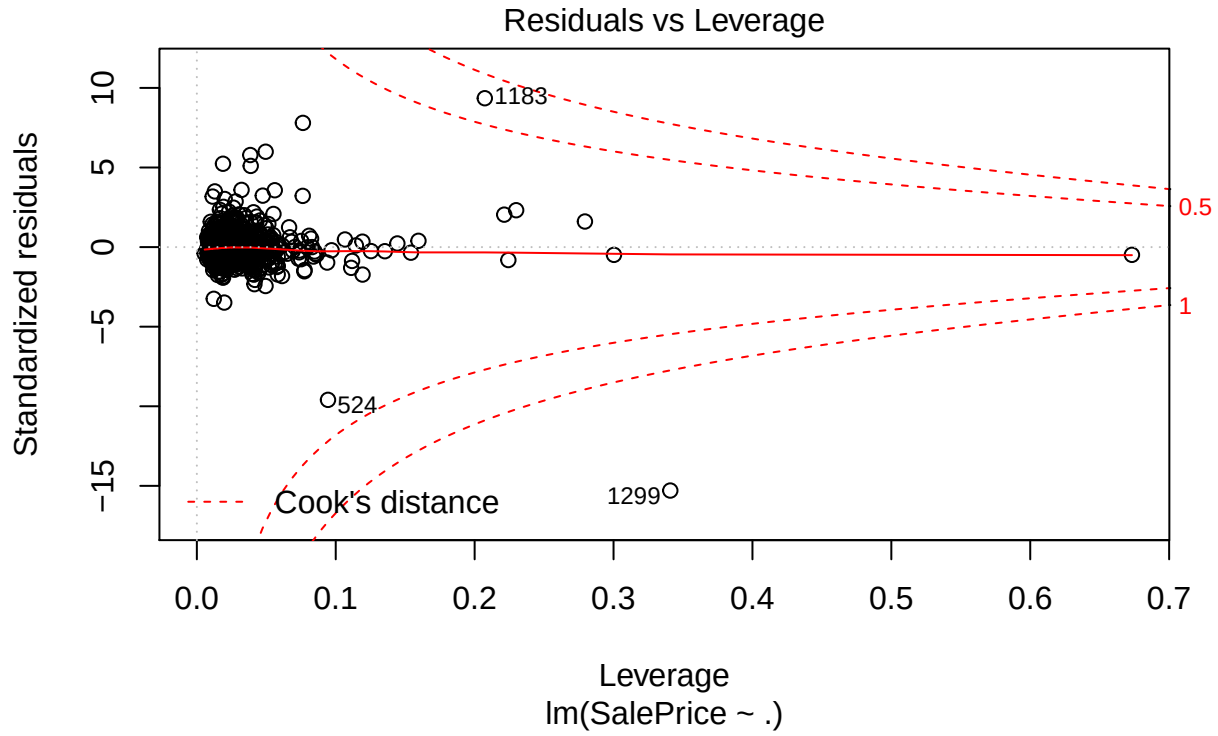
If you look at the output above, there are a few insights we can make. The first is that under the column each predictor has a linear relationship given the others stay constant. For example for Year Built we can interpret that given every other predictor stays constant, a one year increase in the year it was built will result in a 356 increase in the value of its sale price. The predictors that appear to have a statistically significant relationship to the response are those with low p-values and stars next to it. For example lot area, overall quality, bedrooms above ground, and many others are all statistically significant.

Lets now take a look at the diagnostic plots of the regression


Residuals vs Fitted
Residuals
200000
0
−400000
1183
524
1299
0
100000
200000
300000
400000
500000
600000
Fitted values
lm(SalePrice ~ .)


Normal Q−Q
Standardized residuals
10
5
0
−5
−10
−15
1183
524
1299
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(SalePrice ~ .)


Scale−Location
√|Standardized residuals|
1299
1183
524
0
1
2
3
4
0
100000
200000
300000
400000
500000
600000
Fitted values
lm(SalePrice ~ .)


Residuals vs Leverage
Standardized residuals
1183
524
1299
Cook's distance
0.5
1
10
5
0
−5
−15
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Leverage
lm(SalePrice ~ .)

Lets take a look at the residual vs fitted graph first. As we can see, it is a relatively straight line with equally spread out residuals. This is a good thing as it reflects that there are no non-linear relationships. However, if we take a closer look, there is a slight parabolic shape which could reflect a slight non-linear relationship. In addition, as sale price gets larger, we notice that the data begins to have larger residuals and more outliers.

Skipping the next two graphs we come to the Residual vs Leverage graph. This graph tells us if there are any outliers that are influencial to the regression line. The dotted lines represent Cook's distance and if dotted lines lay inside of them, it means that they are influencial in the regression model. As we can see, point 1183 and 1299 do just this.

Finally, we played around with the model to try to make better fit to the data by using interactions and transformation of predictor variables. The code exploring this is put below.

```
#4
inter_model1 <- lm(SalePrice ~ OverallQual+ GrLivArea + OverallQual*GrLivArea, data=Ames)
summary(inter_model1)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallQual + GrLivArea + OverallQual *
##     GrLivArea, data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -527522  -20527      30   17832  271107
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            10031.529  12099.368   0.829  0.40719
## OverallQual            15409.364   1948.165   7.910 5.06e-15 ***
## GrLivArea                -23.428      8.103  -2.891  0.00389 **
## OverallQual:GrLivArea     11.620      1.128  10.305  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41050 on 1456 degrees of freedom
## Multiple R-squared:  0.7336, Adjusted R-squared:  0.7331
## F-statistic:  1337 on 3 and 1456 DF,  p-value: < 2.2e-16
```

```
#statistically significant interaction between quality and above ground living area
inter_model2 <- lm(SalePrice ~ GrLivArea + OverallQual + YearBuilt+ YearRemodAdd + YearBuilt*YearRemodA
summary(inter_model2)
```

```
##
## Call:
## lm(formula = SalePrice ~ GrLivArea + OverallQual + YearBuilt +
##     YearRemodAdd + YearBuilt * YearRemodAdd, data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -391834  -22371   -2094   18004  294797
##
## Coefficients:
##                             Estimate   Std. Error t value Pr(>|t|)
```

```
## (Intercept)            1421174.2449 8157974.7104   0.174    0.862
## GrLivArea                   61.8285       2.5738  24.023   <2e-16 ***
## OverallQual              24092.1035    1237.7762  19.464   <2e-16 ***
## YearBuilt                 -997.8718    4180.2756  -0.239    0.811
## YearRemodAdd             -1173.1719    4097.5730  -0.286    0.775
## YearBuilt:YearRemodAdd       0.7193       2.0998   0.343    0.732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40620 on 1454 degrees of freedom
## Multiple R-squared:  0.7395, Adjusted R-squared:  0.7386
## F-statistic: 825.5 on 5 and 1454 DF,  p-value: < 2.2e-16
```

```
#no statistically significant interaction between yearbuilt and year remodeled
inter_model3 <- lm(SalePrice ~ GrLivArea + OverallQual + LotArea + OverallQual*LotArea, data=Ames)
summary(inter_model3)
```

```
##
## Call:
## lm(formula = SalePrice ~ GrLivArea + OverallQual + LotArea +
##     OverallQual * LotArea, data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -409997  -21465    -408   19042  292157
##
## Coefficients:
##                          Estimate    Std. Error t value Pr(>|t|)
## (Intercept)         -107938.67614    8137.11693 -13.265   <2e-16 ***
## GrLivArea                50.46805       2.68861  18.771   <2e-16 ***
## OverallQual           33250.30710    1361.20150  24.427   <2e-16 ***
## LotArea                   0.84443       0.59382   1.422    0.155
## OverallQual:LotArea       0.01056       0.09215   0.115    0.909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41620 on 1455 degrees of freedom
## Multiple R-squared:  0.7263, Adjusted R-squared:  0.7256
## F-statistic: 965.4 on 4 and 1455 DF,  p-value: < 2.2e-16
```

```
#interestingly not a significant interaction between lot area and quality
inter_model4 <- lm(SalePrice ~ GrLivArea + OverallQual + LotArea + LotFrontage +  LotFrontage*LotArea, d
summary(inter_model4)
```

```
##
## Call:
## lm(formula = SalePrice ~ GrLivArea + OverallQual + LotArea +
##     LotFrontage + LotFrontage * LotArea, data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -311278  -22063   -1696   19773  285743
##
```

```
## Coefficients:
##                            Estimate     Std. Error t value Pr(>|t|)
## (Intercept)         -156748.264960    6375.288509 -24.587  < 2e-16 ***
## GrLivArea               45.587153       3.021579  15.087  < 2e-16 ***
## OverallQual          34417.052784    1055.702035  32.601  < 2e-16 ***
## LotArea                  4.721621       0.359998  13.116  < 2e-16 ***
## LotFrontage            493.042287      63.874263   7.719 2.47e-14 ***
## LotArea:LotFrontage     -0.030306       0.002567 -11.808  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 41160 on 1195 degrees of freedom
##   (259 observations deleted due to missingness)
## Multiple R-squared:  0.7574, Adjusted R-squared:  0.7564
## F-statistic: 746.2 on 5 and 1195 DF,  p-value: < 2.2e-16
```

```
#statistically significant but small interaction between lot area and lot frontage.
#it may make sense to include the 2 significant interactions in the final model

#5
transformer_model1 <- lm(log(SalePrice) ~ ., data = Ames)
summary(transformer_model1)
```

```
##
## Call:
## lm(formula = log(SalePrice) ~ ., data = Ames)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.86251 -0.06914  0.00737  0.07665  0.51255
##
## Coefficients:
##                    Estimate    Std. Error t value Pr(>|t|)
## (Intercept)   15.4526582288  6.9111821507   2.236 0.025560 *
## Id            -0.0000079960  0.0000107603  -0.743 0.457575
## LotFrontage    0.0003898468  0.0002357275   1.654 0.098454 .
## LotArea        0.0000018538  0.0000006364   2.913 0.003655 **
## OverallQual    0.0862135878  0.0059684988  14.445  < 2e-16 ***
## OverallCond    0.0509021168  0.0055470635   9.176  < 2e-16 ***
## YearBuilt      0.0025810879  0.0003536601   7.298 5.59e-13 ***
## YearRemodAdd   0.0011489936  0.0003513655   3.270 0.001109 **
## BsmtUnfSF     -0.0000287677  0.0000155599  -1.849 0.064751 .
## TotalBsmtSF    0.0000719466  0.0000232975   3.088 0.002065 **
## X1stFlrSF      0.0000771173  0.0001163177   0.663 0.507477
## X2ndFlrSF      0.0000138079  0.0001139395   0.121 0.903565
## GrLivArea      0.0001041439  0.0001135933   0.917 0.359442
## BsmtFullBath   0.0574278074  0.0129417689   4.437 1.00e-05 ***
## BsmtHalfBath   0.0107738182  0.0205822238   0.523 0.600765
## FullBath       0.0509226333  0.0142800614   3.566 0.000378 ***
## HalfBath       0.0308625215  0.0134102881   2.301 0.021556 *
## BedroomAbvGr  -0.0012403774  0.0086725095  -0.143 0.886297
## KitchenAbvGr  -0.1347950533  0.0257755168  -5.230 2.03e-07 ***
## TotRmsAbvGrd   0.0207205616  0.0060111210   3.447 0.000588 ***
## Fireplaces     0.0446855356  0.0088689711   5.038 5.49e-07 ***
```

```
## GarageYrBlt   -0.0000096161  0.0003692731  -0.026 0.979230
## GarageCars     0.0703342813  0.0142067189   4.951 8.55e-07 ***
## GarageArea     0.0000402521  0.0000491445   0.819 0.412932
## WoodDeckSF     0.0000961153  0.0000405983   2.367 0.018083 *
## OpenPorchSF    0.0000474120  0.0000785073   0.604 0.546023
## EnclosedPorch  0.0001738924  0.0000837928   2.075 0.038195 *
## ScreenPorch    0.0003467082  0.0000819428   4.231 2.52e-05 ***
## PoolArea      -0.0006151074  0.0001201387  -5.120 3.61e-07 ***
## MoSold         0.0004285207  0.0017104934   0.251 0.802229
## YrSold        -0.0060968302  0.0034367798  -1.774 0.076342 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.15 on 1096 degrees of freedom
##   (333 observations deleted due to missingness)
## Multiple R-squared:  0.8611, Adjusted R-squared:  0.8573
## F-statistic: 226.4 on 30 and 1096 DF,  p-value: < 2.2e-16
```

```
#using log of sale price allows us to see percent change effects of the variables

transformer_model2 <- lm(SalePrice ~ log(GrLivArea), data = Ames)
summary(transformer_model2)
```

```
##
## Call:
## lm(formula = SalePrice ~ log(GrLivArea), data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -247772  -31767   -1680   24759  391583
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1022316      32624  -31.34   <2e-16 ***
## log(GrLivArea)   165558       4484   36.92   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 57130 on 1458 degrees of freedom
## Multiple R-squared:  0.4832, Adjusted R-squared:  0.4828
## F-statistic:  1363 on 1 and 1458 DF,  p-value: < 2.2e-16
```

```
#allows you to see the effects of a percent change in GrLivArea

transformer_model3 <- lm(SalePrice ~ GrLivArea + I(GrLivArea^2), data = Ames)
summary(transformer_model3)
```

```
##
## Call:
## lm(formula = SalePrice ~ GrLivArea + I(GrLivArea^2), data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -321613  -30369    -876   22954  338146
##
## Coefficients:
##                     Estimate    Std. Error t value  Pr(>|t|)
## (Intercept)    -13283.688506   8583.746149  -1.548     0.122
## GrLivArea         145.549389      9.276039  15.691   < 2e-16 ***
## I(GrLivArea^2)     -0.010250      0.002361  -4.341 0.0000152 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55730 on 1457 degrees of freedom
## Multiple R-squared:  0.5085, Adjusted R-squared:  0.5078
## F-statistic: 753.7 on 2 and 1457 DF,  p-value: < 2.2e-16
```

```
#as above ground living area increases it becomes less significant

transformer_model4 <- lm(SalePrice ~ LotArea + I(LotArea^2), data = Ames)
summary(transformer_model4)
```

```
##
## Call:
## lm(formula = SalePrice ~ LotArea + I(LotArea^2), data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -236109  -46223  -15645   32590  534227
##
## Coefficients:
##                     Estimate      Std. Error t value Pr(>|t|)
## (Intercept)  123300.85821909   4376.49958756   28.17   <2e-16 ***
## LotArea           6.02902692      0.41858791   14.40   <2e-16 ***
## I(LotArea^2)     -0.00002753      0.00000260  -10.59   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 73890 on 1457 degrees of freedom
## Multiple R-squared:  0.1361, Adjusted R-squared:  0.1349
## F-statistic: 114.8 on 2 and 1457 DF,  p-value: < 2.2e-16
```

```
#Same for lot area

transformer_model5 <- lm(SalePrice ~ sqrt(LotArea), data = Ames)
summary(transformer_model5)
```

```
##
## Call:
## lm(formula = SalePrice ~ sqrt(LotArea), data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -263407  -46335  -16276   32787  537176
##
## Coefficients:
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   81166.31    6905.23   11.75   <2e-16 ***
## sqrt(LotArea)  1013.33      67.33   15.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 73930 on 1458 degrees of freedom
## Multiple R-squared:  0.1345, Adjusted R-squared:  0.1339
## F-statistic: 226.5 on 1 and 1458 DF,  p-value: < 2.2e-16
```

```
#Taking the sqrt of lot area increases the statistical significance but I am still not sure why you wou
```