SSC 442 Lab 2

Team Awesome (19) - Sayem Lincoln, Joshua Schwimmer, John Townshend.

1/27/2020

Lab 2

Excercise 1

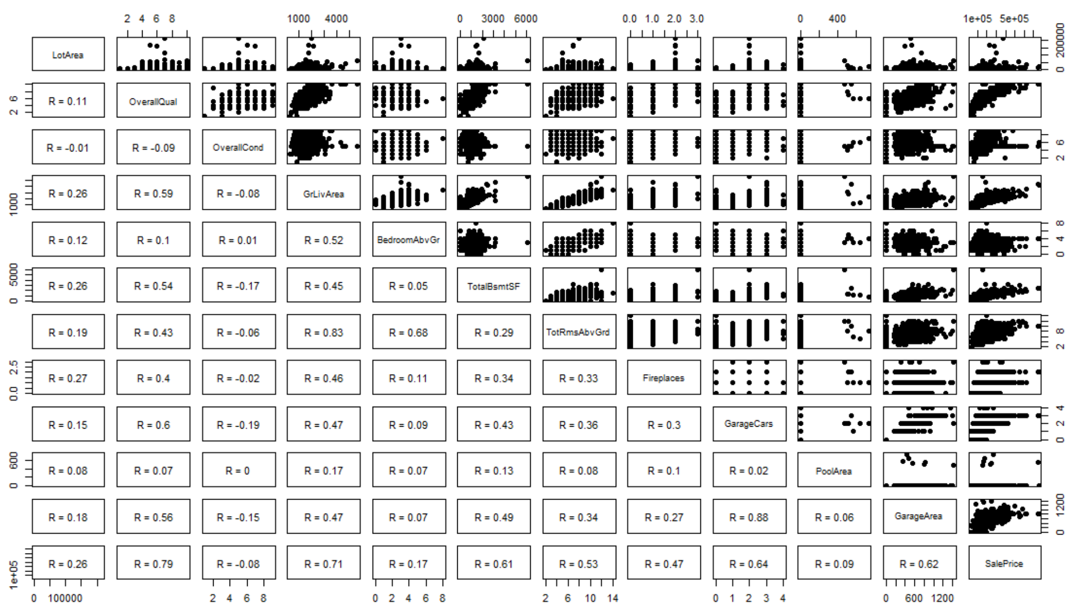
Part 1

The Ames dataframe is saved. And the text file for the Ames dataframe is also uploadecd to the repo.

Part 2

Sactter plot matrix

```
pairs(dfAmesCleaner[,1:12], upper.panel = upper.panel, lower.panel = panel.cor)
```



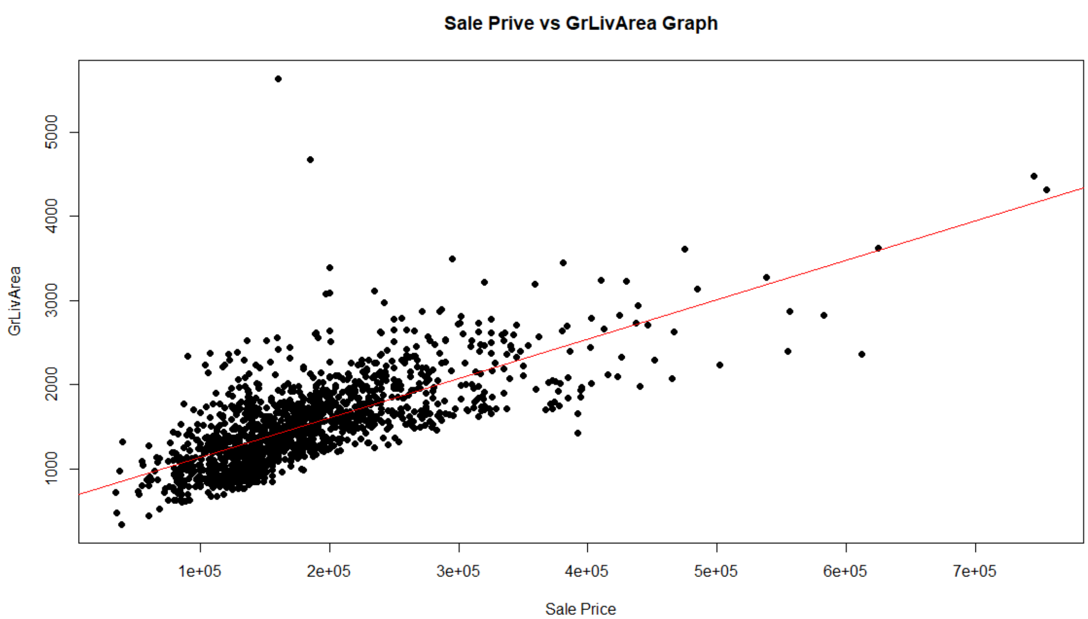
I believe there is correlation between the variables and SalePrice, as a house condition should depend on its price. I believe all of them should have some form of correlations with SalePrice, each of them influencing the price and vice versa.

Part 3

The ones that I believe to have a correlation to SalePrice are - “LotArea” “OverallQual” “OverallCond” “GrLivArea” “BedroomAbvGr” “TotalBsmtSF” “TotRmsAbvGrd” “GarageArea” As these graphs show a scatter plot compared to the others when plotted in a scatter plot matrix as from a scatter plot a coorelation causation trejectory can be formed,so I believe these Variables are correlated to SalePrice. So from Part 2 only these three variables does not show any correlation - “Fireplaces” “GarageCars” “PoolArea” whereas the rest do.

Part 4

```
plot(dfAmesCleaner$SalePrice, dfAmesCleaner$GrLivArea, main="Sale Prive vs
GrLivArea Graph",
     xlab="Sale Price", ylab="GrLivArea",
pch=19)+abline(lm(dfAmesCleaner$GrLivArea~ dfAmesCleaner$SalePrice,
data=dfAmesCleaner), col="red")# regression line (y~x)
```




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

The largest outlier that is above the regression line - it is the one with GrLiveArea of 5642 and SalePrice of 160000.

Produce the other information about this house? The other informations about the house - “LotArea” - 63887 “OverallQual” - 10 “OverallCond” - 5 “GrLivArea” - 5642 “BedroomAbvGr” - 3 “TotalBsmtSF”- 6110 “TotRmsAbvGrd”-12 “Fireplaces” - 3 “GarageCars” - 3 “PoolArea” - 480 “GarageArea” - 1418 “SalePrice” - 160000

Exercise 2

Part 1

Simple linear regression - running SalePrice against Garage Area

```
lm.fit1 = lm(dfAmesCleaner$SalePrice~dfAmesCleaner$GarageArea)
```

Part 2

```
summary(lm.fit2)

##
## Call:
## lm(formula = dfAmesCleaner$SalePrice ~ MSSubClass + LotFrontage +
##     LotArea + OverallQual + OverallCond + YearBuilt + YearRemodAdd +
##     MasVnrArea + BsmtFinSF1 + BsmtFinSF2 + BsmtUnfSF + TotalBsmtSF +
##     X1stFlrSF + X2ndFlrSF + LowQualFinSF + GrLivArea + BsmtFullBath +
##     BsmtHalfBath + FullBath + HalfBath + BedroomAbvGr + KitchenAbvGr +
##     TotRmsAbvGrd + Fireplaces + GarageYrBlt + GarageCars + GarageArea +
##     WoodDeckSF + OpenPorchSF + EnclosedPorch + X3SsnPorch + ScreenPorch +
##     PoolArea + MiscVal + MoSold + YrSold, data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -442865  -16873   -2581   14998  318042
##
## Coefficients: (2 not defined because of singularities)
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.232e+05  1.701e+06  -0.190 0.849317
## MSSubClass    -2.005e+02  3.449e+01  -5.814 8.03e-09 ***
## LotFrontage   -1.161e+02  6.124e+01  -1.896 0.058203 .
## LotArea        5.454e-01  1.573e-01   3.466 0.000548 ***
## OverallQual    1.870e+04  1.478e+03  12.646  < 2e-16 ***
## OverallCond    5.227e+03  1.367e+03   3.824 0.000139 ***
## YearBuilt      3.170e+02  8.762e+01   3.617 0.000311 ***
## YearRemodAdd   1.206e+02  8.661e+01   1.392 0.164174
## MasVnrArea     3.160e+01  7.006e+00   4.511 7.15e-06 ***
## BsmtFinSF1     1.739e+01  5.835e+00   2.980 0.002947 **
## BsmtFinSF2     8.362e+00  8.763e+00   0.954 0.340205
## BsmtUnfSF      5.006e+00  5.275e+00   0.949 0.342890
## TotalBsmtSF           NA         NA      NA       NA
## X1stFlrSF      4.591e+01  7.356e+00   6.241 6.21e-10 ***
## X2ndFlrSF      4.668e+01  6.099e+00   7.654 4.28e-14 ***
## LowQualFinSF   3.415e+01  2.788e+01   1.225 0.220788
## GrLivArea             NA         NA      NA       NA
## BsmtFullBath   8.980e+03  3.194e+03   2.812 0.005018 **
## BsmtHalfBath   2.490e+03  5.071e+03   0.491 0.623487
## FullBath       5.390e+03  3.529e+03   1.527 0.126941
## HalfBath      -1.119e+03  3.320e+03  -0.337 0.736244
## BedroomAbvGr  -1.023e+04  2.154e+03  -4.750 2.30e-06 ***
```

```
## KitchenAbvGr  -2.193e+04  6.704e+03  -3.271 0.001105 **
## TotRmsAbvGrd   5.440e+03  1.486e+03   3.661 0.000263 ***
## Fireplaces     4.375e+03  2.188e+03   2.000 0.045793 *
## GarageYrBlt   -4.914e+01  9.093e+01  -0.540 0.589011
## GarageCars     1.679e+04  3.487e+03   4.815 1.68e-06 ***
## GarageArea     6.488e+00  1.211e+01   0.536 0.592338
## WoodDeckSF     2.155e+01  1.002e+01   2.151 0.031713 *
## OpenPorchSF   -2.315e+00  1.948e+01  -0.119 0.905404
## EnclosedPorch  7.233e+00  2.061e+01   0.351 0.725733
## X3SsnPorch     3.458e+01  3.749e+01   0.922 0.356593
## ScreenPorch    5.797e+01  2.040e+01   2.842 0.004572 **
## PoolArea      -6.126e+01  2.984e+01  -2.053 0.040326 *
## MiscVal       -3.850e+00  6.955e+00  -0.554 0.579980
## MoSold        -2.240e+02  4.227e+02  -0.530 0.596213
## YrSold        -2.536e+02  8.454e+02  -0.300 0.764216
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36790 on 1086 degrees of freedom
##   (339 observations deleted due to missingness)
## Multiple R-squared:  0.8095, Adjusted R-squared:  0.8036
## F-statistic: 135.7 on 34 and 1086 DF,  p-value: < 2.2e-16
```

Is there a relationship between the predictors and the response?

Ans - Yes there is a relation ship between them, they are corraleted to each other. If there was no correaltion then residuals would not be possible.

Which predictors appear to have a statistically significant relationshipto the response?

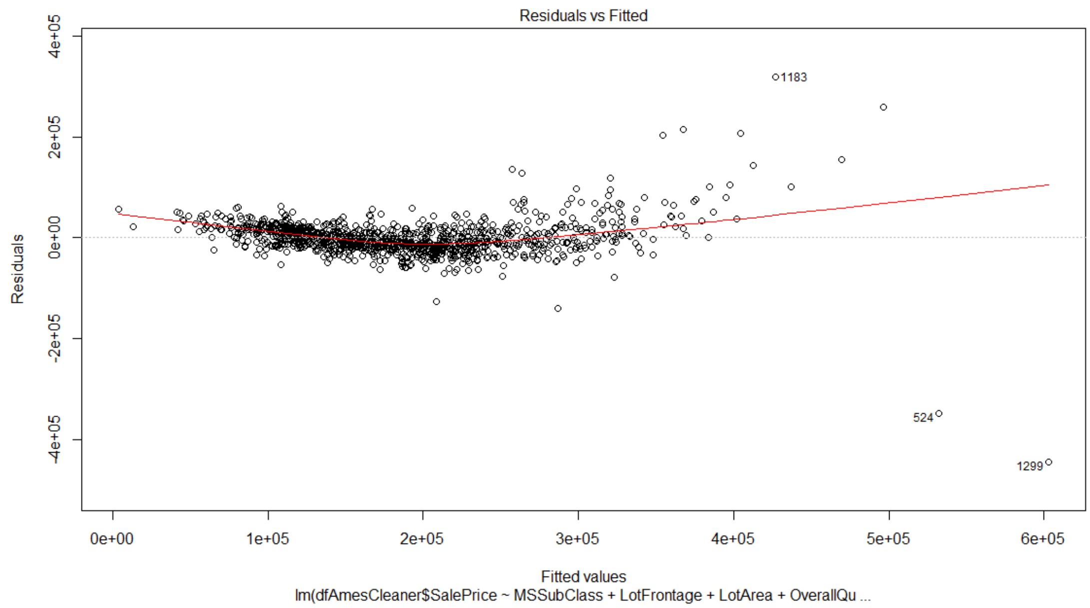
Ans - The followings processed a statistical value - Id, MSSubClass, LotFrontage, LotArea, OverallQual, OverallCond, YearBuilt, YearRemodAdd MasVnrArea, BsmtFinSF1, BsmtFinSF2, BsmtUnfSF, X1stFlrSF, X2ndFlrSF, LowQualFinSF BsmtFullBath, BsmtHalfBath, FullBath, HalfBath, BedroomAbvGr, KitchenAbvGr, TotRmsAbvGrd, Fireplaces, GarageYrBlt, GarageCars, GarageArea, WoodDeckSF, OpenPorchSF, EnclosedPorch, X3SsnPorch, ScreenPorch, PoolArea, MiscVal, MoSold, YrSold, SalePrice - so the following should have be statistically significant.

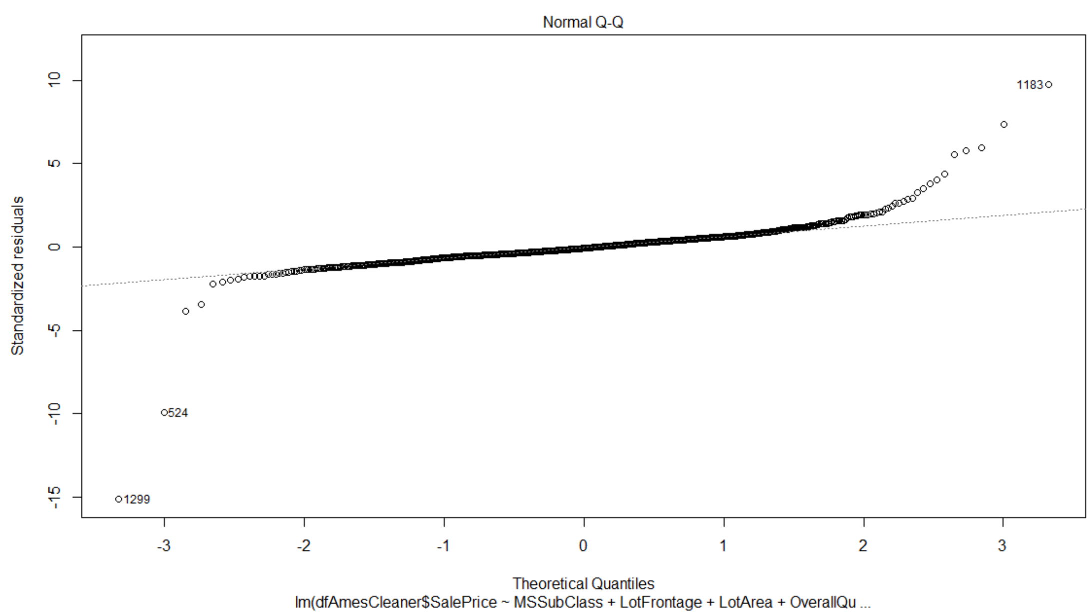
What does the coefficient for the year variable suggest?

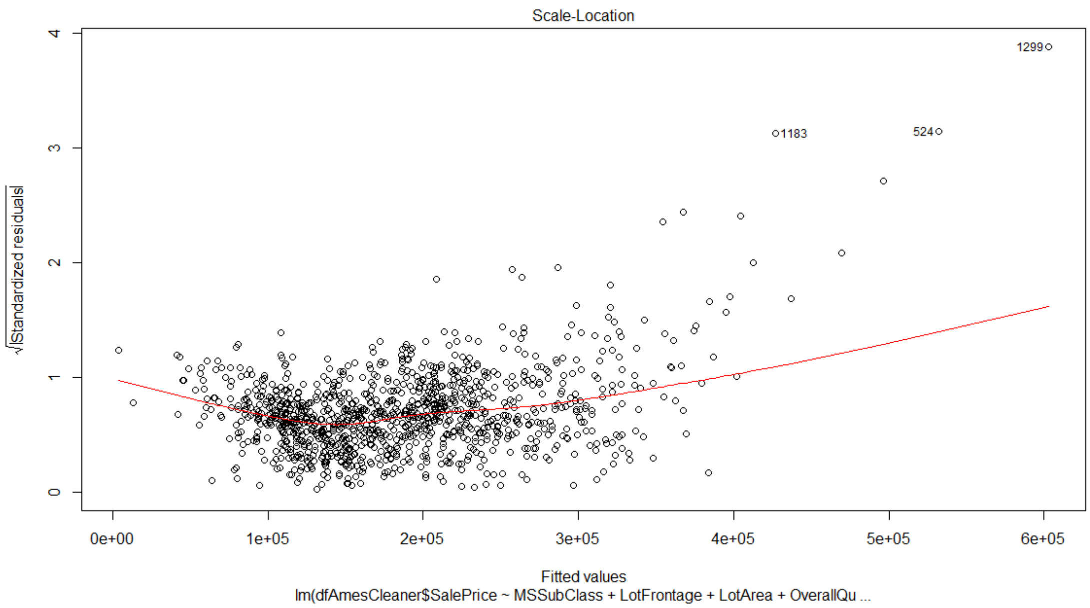
Ans - It suggests that Year is negatively correlated to SalePrice, meaning as the number of years passed some property values went down, which is a common notion for property prices, and that is relfected by the coeffecient that is processed out from the residuals and the model.

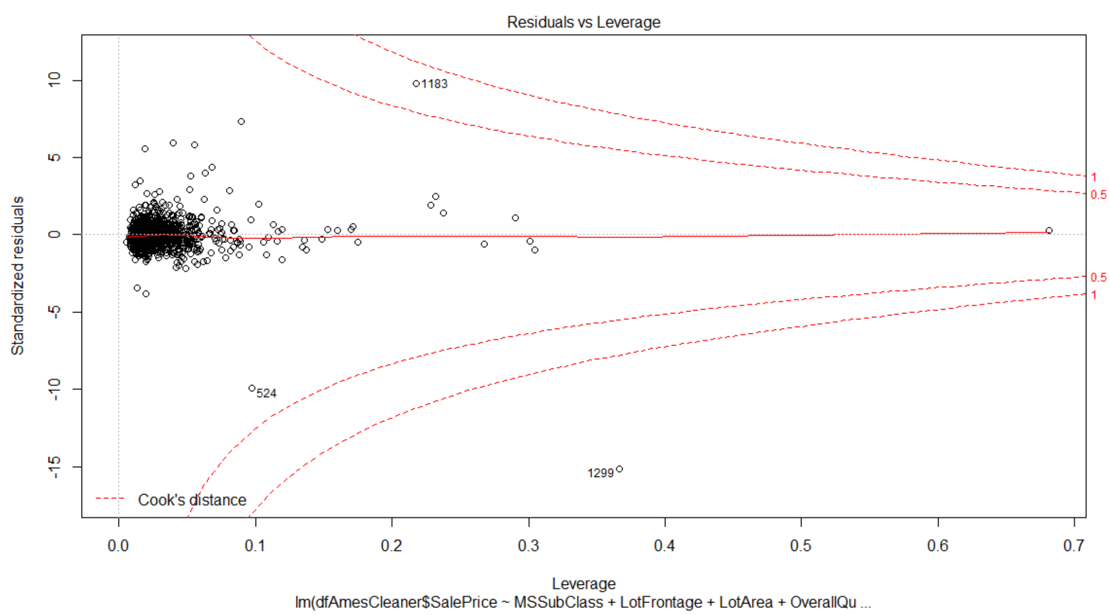
Part 3

```r
plot(lm.fit2)
```


Residuals vs Fitted
Residuals
4e+05
2e+05
0e+00
-2e+05
-4e+05
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
1183
524
1299
Fitted values
lm(dfAmesCleaner$SalePrice ~ MSSubClass + LotFrontage + LotArea + OverallQu ...


Normal Q-Q
Standardized residuals
10
5
0
-5
-10
-15
-3
-2
-1
0
1
2
3
1183
524
1299
Theoretical Quantiles
lm(dfAmesCleaner$SalePrice ~ MSSubClass + LotFrontage + LotArea + OverallQu ...





Comment on any problems you see with the fit.

Some of the variables have NA values so those variables do not form a residual output. So, forming a regression on all Variavbles does not better the model.

Do the residual plots suggest any unusually large outliers?

Ans- Yes there are some large outliers in all the graphs.

Does the leverage plot identify any observations with unusually high leverage?

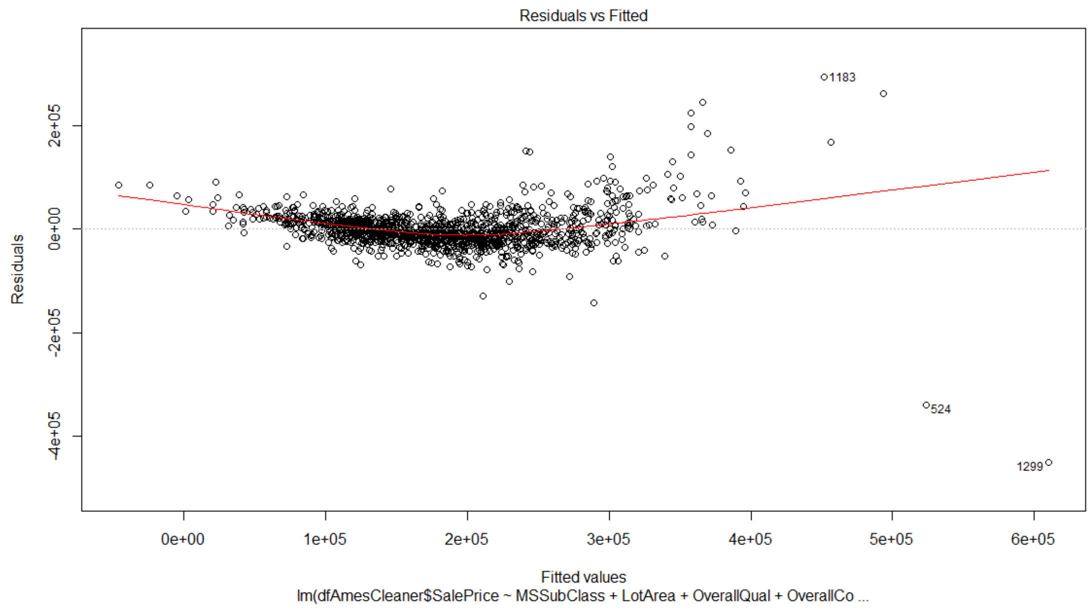
Ans - Yes it does, there is a levargae value of 0.7. which is a very high outlier.

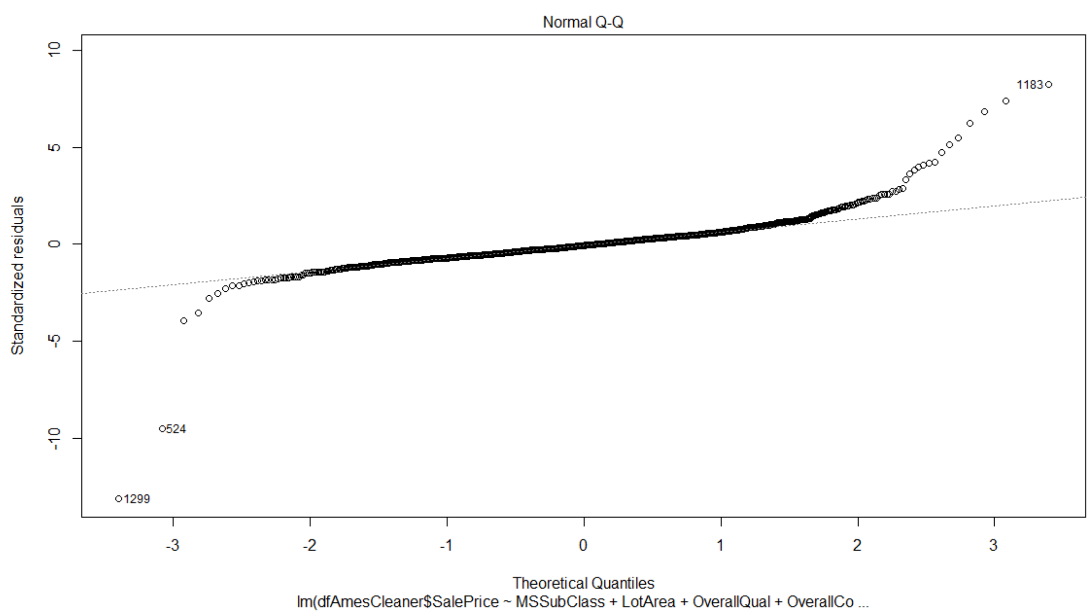
Part 4

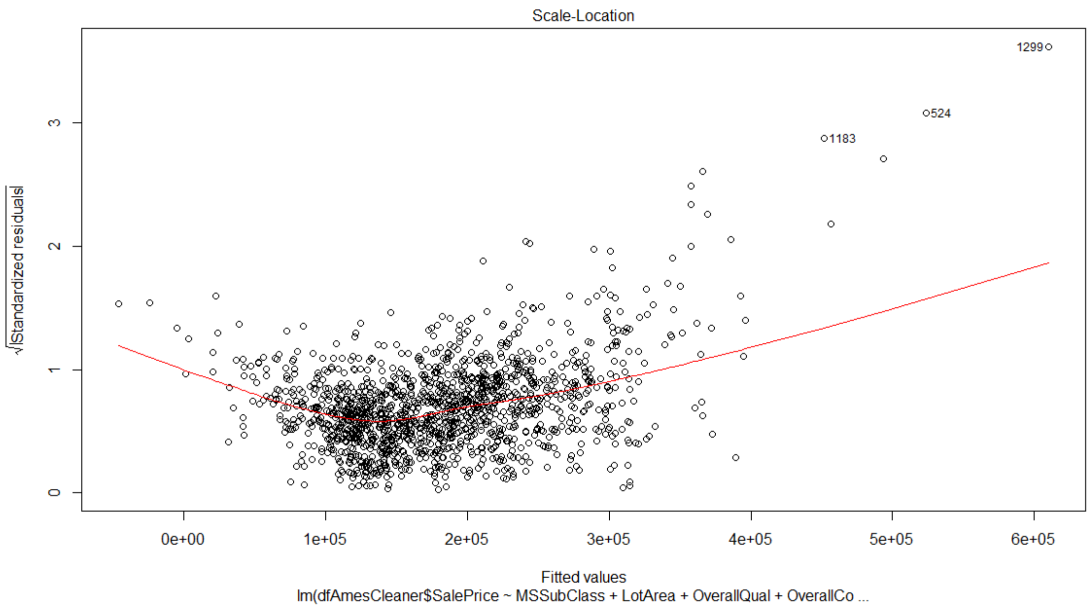
linear regression models with some well-chosen interaction effects

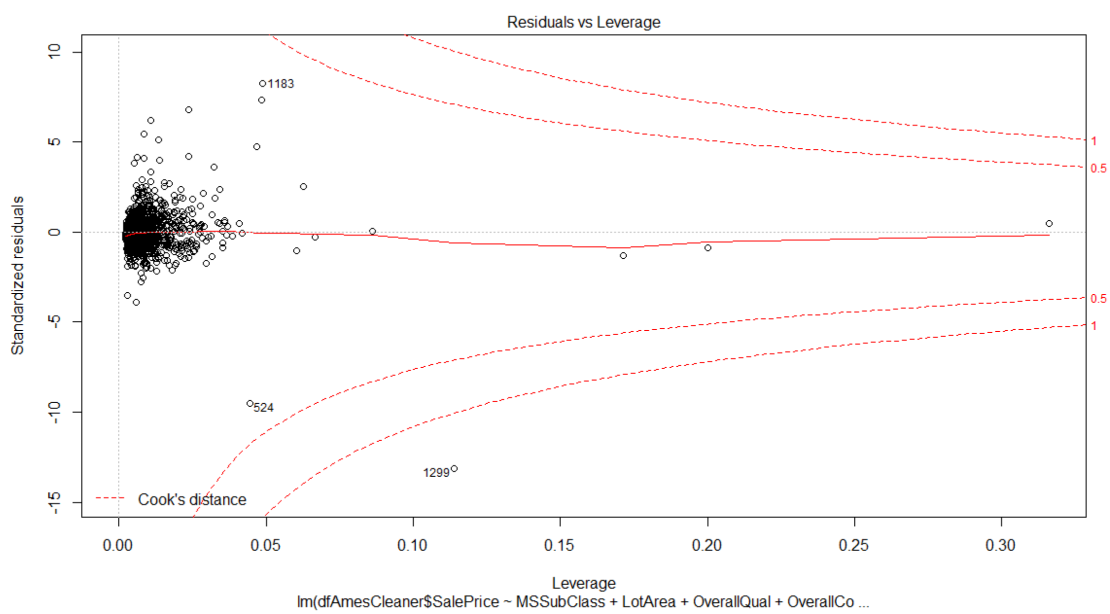
linear models for ***

```
summary(lm.fit3)

##
## Call:
## lm(formula = dfAmesCleaner$SalePrice ~ MSSubClass + LotArea +
##     OverallQual + OverallCond + YearBuilt + MasVnrArea + X1stFlrSF +
##     X2ndFlrSF + HalfBath + KitchenAbvGr + TotRmsAbvGrd + GarageCars,
##     data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -449973  -18596   -2222   14263  293482
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -9.260e+05  9.317e+04  -9.940  < 2e-16 ***
## MSSubClass   -1.152e+02  2.663e+01  -4.324 1.64e-05 ***
## LotArea       5.723e-01  1.026e-01   5.578 2.90e-08 ***
## OverallQual   1.954e+04  1.154e+03  16.928  < 2e-16 ***
## OverallCond   5.741e+03  9.582e+02   5.991 2.63e-09 ***
## YearBuilt     4.300e+02  4.735e+01   9.080  < 2e-16 ***
## MasVnrArea    3.450e+01  6.073e+00   5.682 1.61e-08 ***
## X1stFlrSF     7.084e+01  4.413e+00  16.051  < 2e-16 ***
## X2ndFlrSF     4.752e+01  4.537e+00  10.475  < 2e-16 ***
## HalfBath     -4.674e+02  2.544e+03  -0.184 0.854279
## KitchenAbvGr -1.883e+04  5.262e+03  -3.578 0.000358 ***
## TotRmsAbvGrd  7.751e+02  1.115e+03   0.695 0.487019
## GarageCars    1.193e+04  1.762e+03   6.774 1.82e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36460 on 1439 degrees of freedom
##   (8 observations deleted due to missingness)
## Multiple R-squared:  0.7903, Adjusted R-squared:  0.7885
## F-statistic: 451.9 on 12 and 1439 DF,  p-value: < 2.2e-16

plot(lm.fit3)
```


Residuals vs Fitted
Residuals
Fitted values
lm(dfAmesCleaner$SalePrice ~ MSSubClass + LotArea + OverallQual + OverallCo ...
1183
524
1299
2e+05
0e+00
-2e+05
-4e+05
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(dfAmesCleaner$SalePrice ~ MSSubClass + LotArea + OverallQual + OverallCo ...
1183
524
1299
10
5
0
-5
-10
-3
-2
-1
0
1
2
3



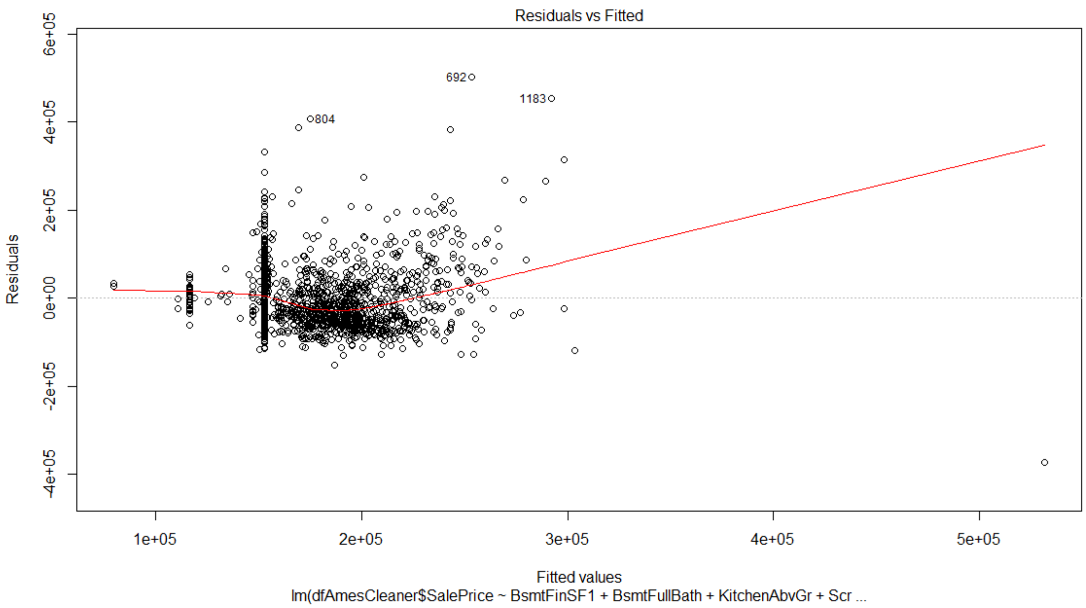


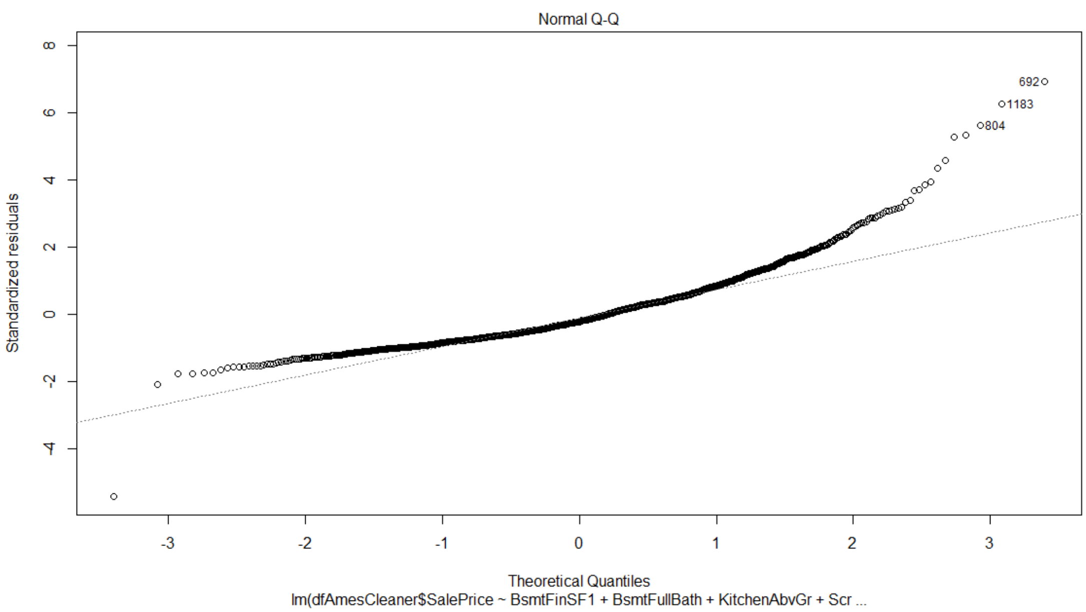
linear models for **

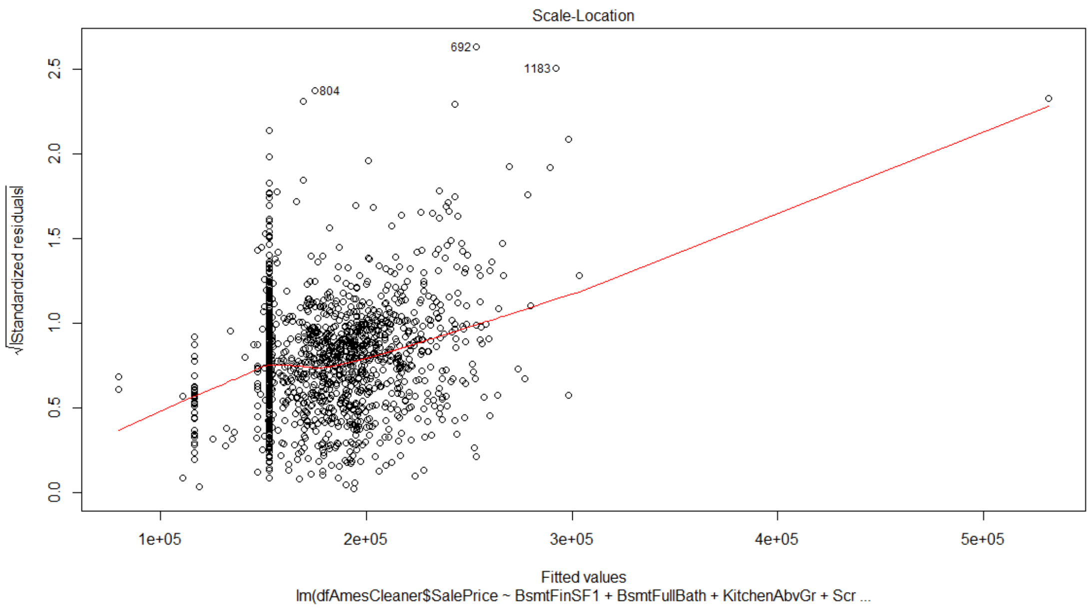
```
summary(lm.fit4)

##
## Call:
## lm(formula = dfAmesCleaner$SalePrice ~ BsmtFinSF1 + BsmtFullBath +
##     KitchenAbvGr + ScreenPorch, data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -371542  -49400  -15870   33209  501842
```

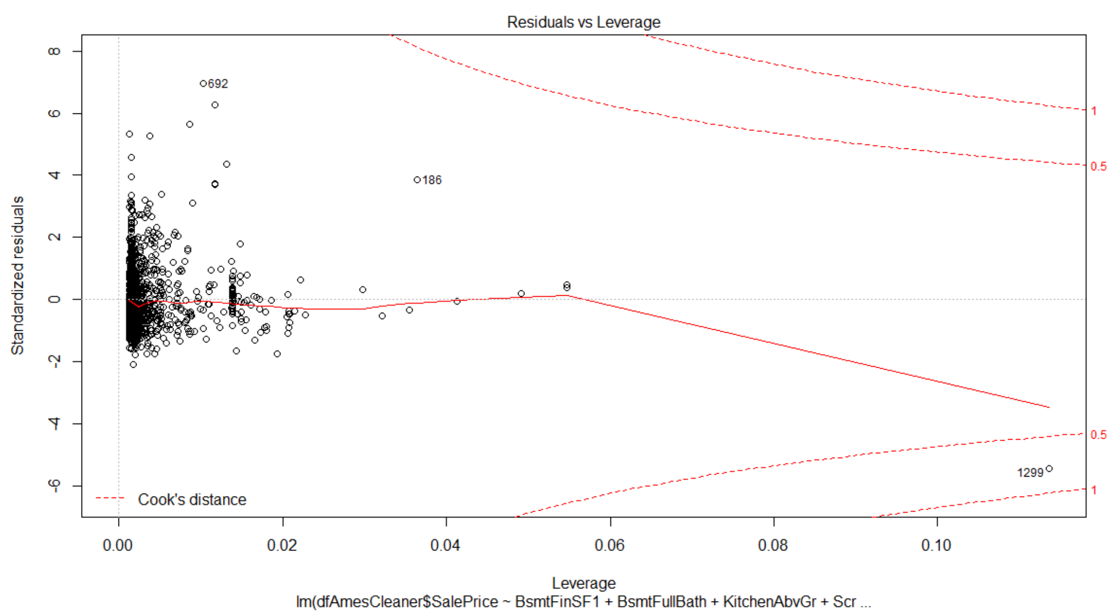
```
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  188986.430   9619.707  19.646  < 2e-16 ***
## BsmtFinSF1       69.136      5.499  12.572  < 2e-16 ***
## BsmtFullBath  -5613.972   4814.904  -1.166 0.243823
## KitchenAbvGr -36422.236   8660.583  -4.206 2.76e-05 ***
## ScreenPorch     117.493     34.183   3.437 0.000604 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 72560 on 1455 degrees of freedom
## Multiple R-squared:  0.168,  Adjusted R-squared:  0.1657
## F-statistic: 73.43 on 4 and 1455 DF,  p-value: < 2.2e-16
```

```r
plot(lm.fit4)
```


Residuals vs Fitted
Residuals
Fitted values
lm(dfAmesCleaner$SalePrice ~ BsmtFinSF1 + BsmtFullBath + KitchenAbvGr + Scr ...
692
1183
804
6e+05
4e+05
2e+05
0e+00
-2e+05
-4e+05
1e+05
2e+05
3e+05
4e+05
5e+05


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(dfAmesCleaner$SalePrice ~ BsmtFinSF1 + BsmtFullBath + KitchenAbvGr + Scr ...
692
1183
804
8
6
4
2
0
-2
-4
-3
-2
-1
0
1
2
3



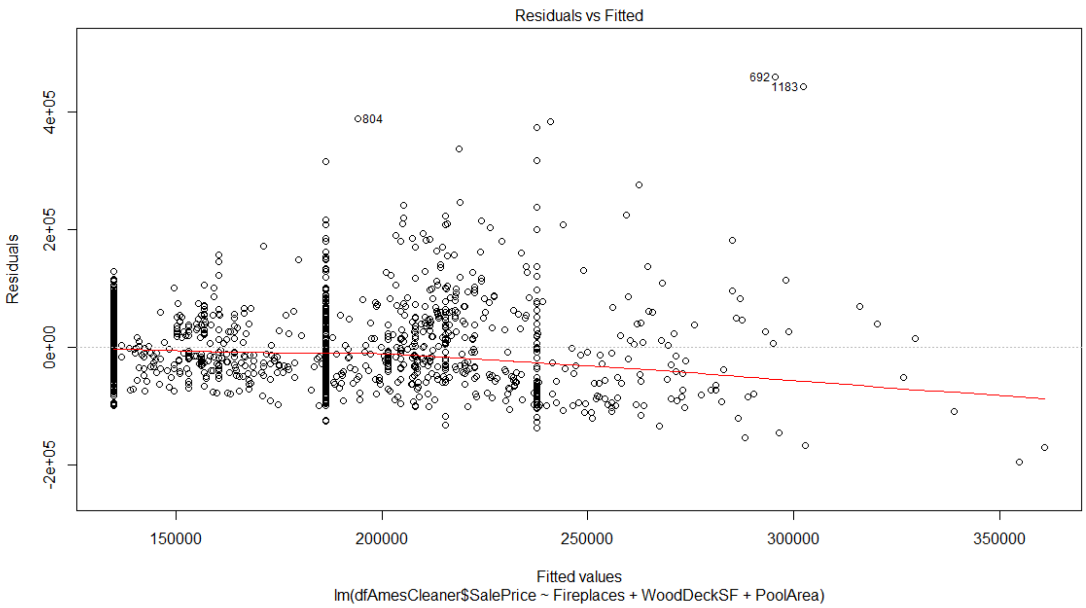


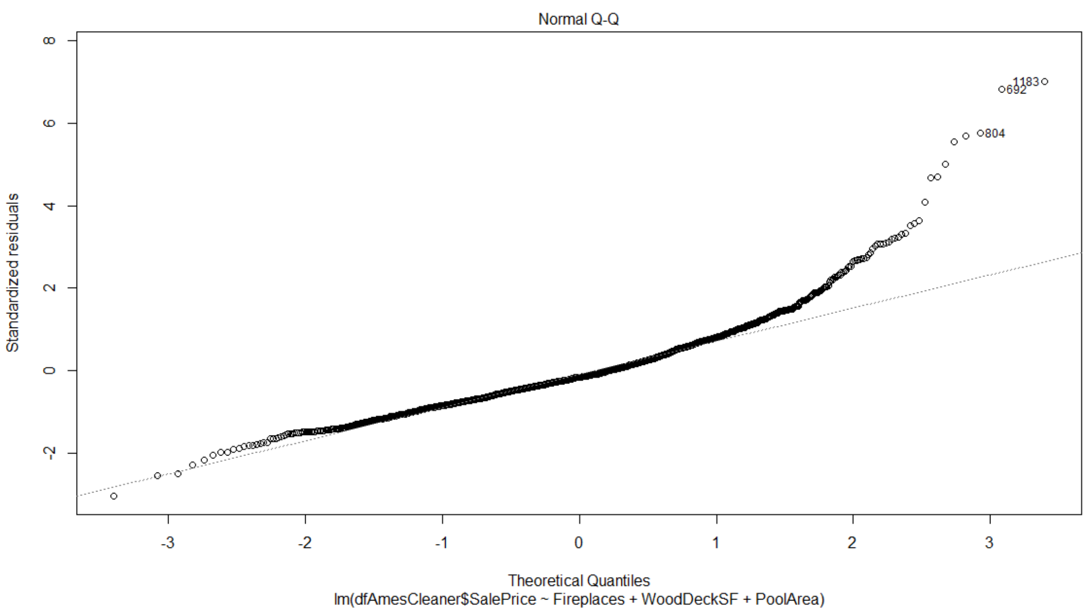
linear models for *

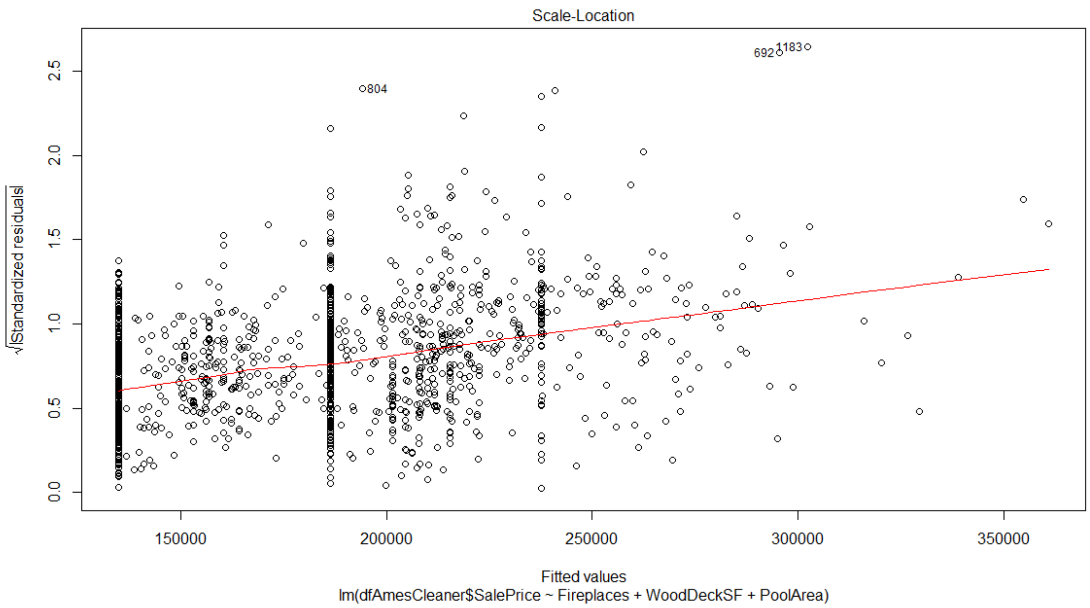
```
summary(lm.fit5)

##
## Call:
## lm(formula = dfAmesCleaner$SalePrice ~ Fireplaces + WoodDeckSF +
##     PoolArea, data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -194717  -42546  -10226   30977  459672
```

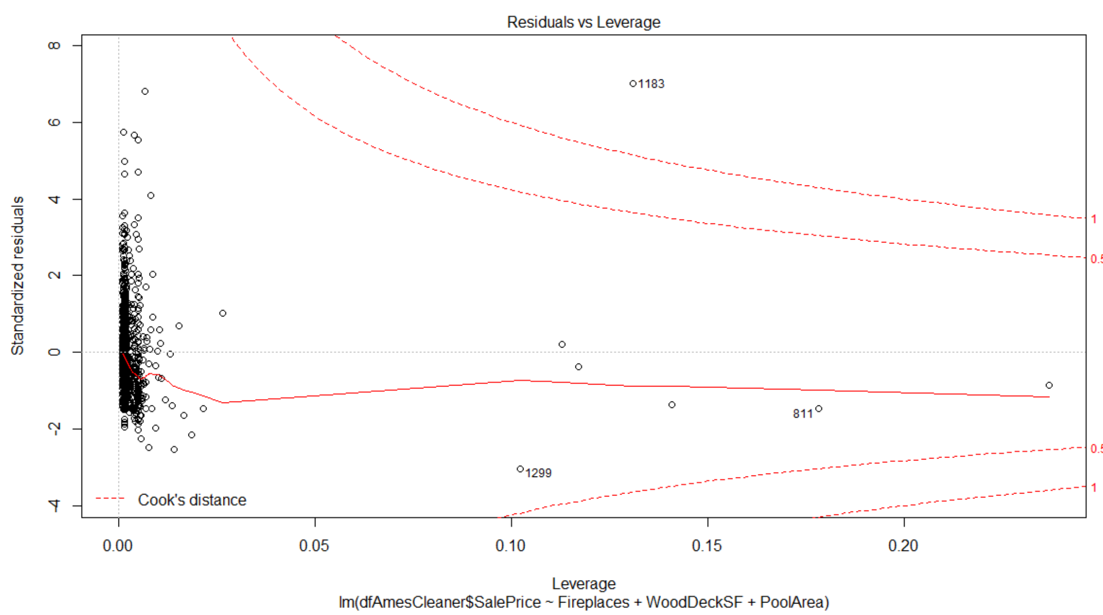
```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 135060.17    2650.54  50.956   <2e-16 ***
## Fireplaces   51243.38    2816.38  18.195   <2e-16 ***
## WoodDeckSF     151.26      14.46  10.461   <2e-16 ***
## PoolArea        69.91      44.40   1.575    0.116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67720 on 1456 degrees of freedom
## Multiple R-squared:  0.2749, Adjusted R-squared:  0.2734
## F-statistic:   184 on 3 and 1456 DF,  p-value: < 2.2e-16
```

```
plot(lm.fit5)
```


Residuals vs Fitted
Residuals
Fitted values
lm(dfAmesCleaner$SalePrice ~ Fireplaces + WoodDeckSF + PoolArea)
692
1183
804
4e+05
2e+05
0e+00
-2e+05
150000
200000
250000
300000
350000


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(dfAmesCleaner$SalePrice ~ Fireplaces + WoodDeckSF + PoolArea)
1183
692
804



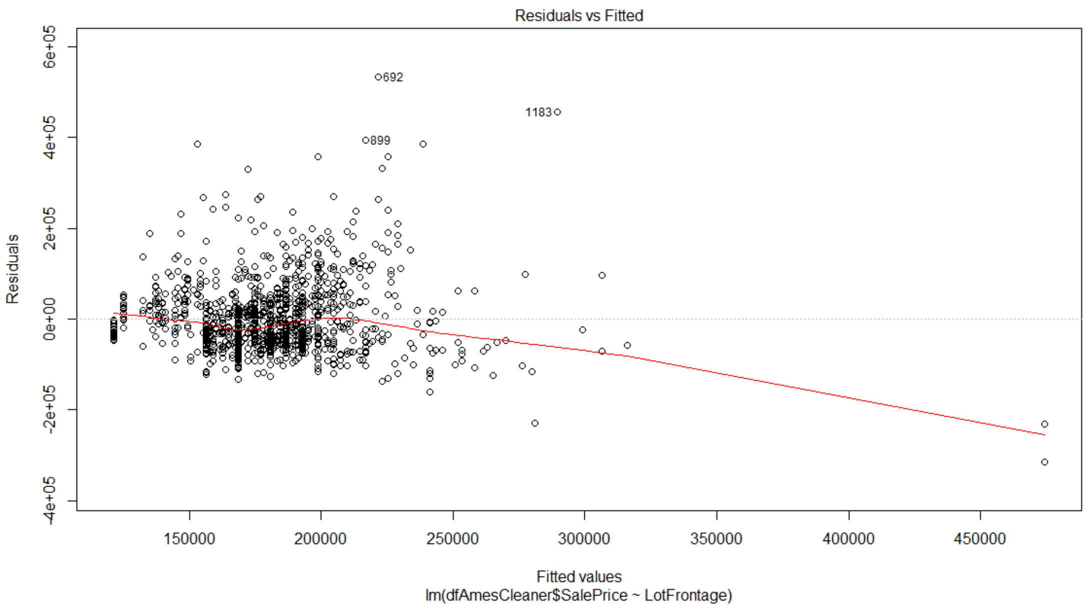


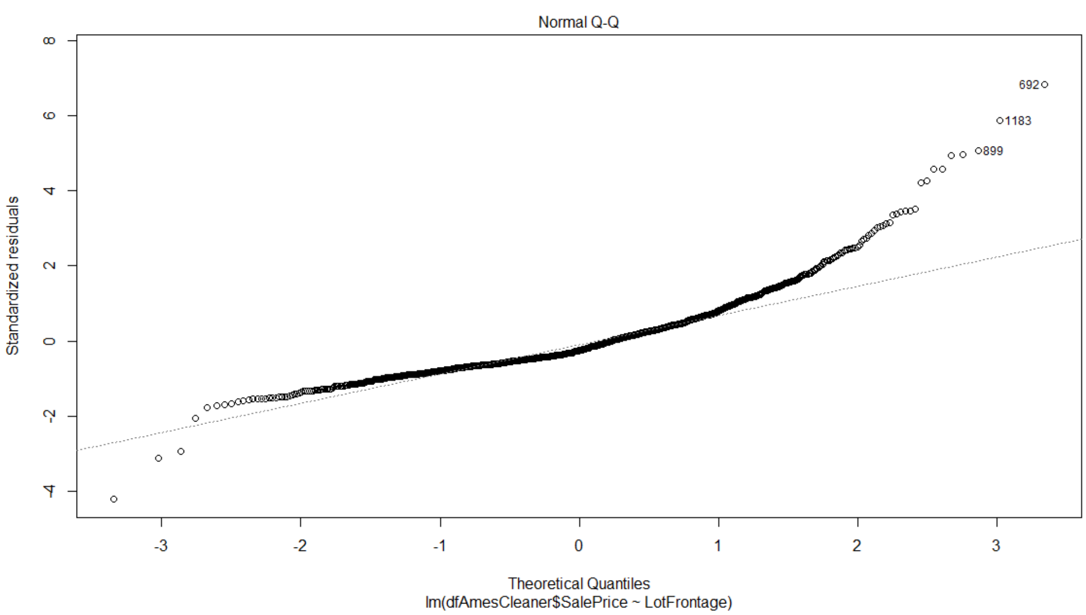
linear models for .

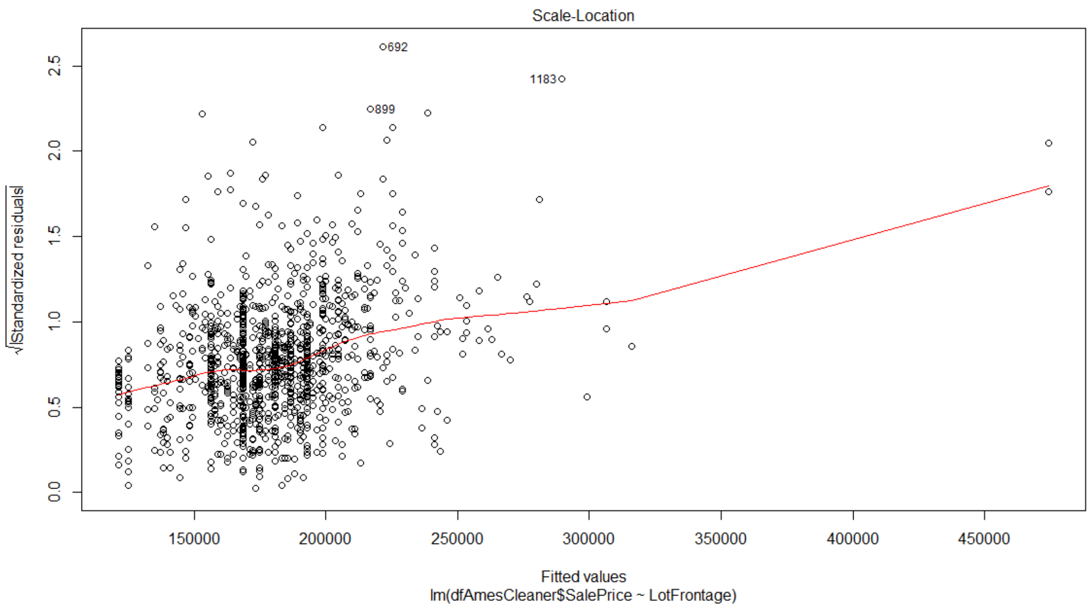
```
summary(lm.fit6)

##
## Call:
## lm(formula = dfAmesCleaner$SalePrice ~ LotFrontage, data = Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -314258  -48878  -19402   33290  533217
##
```

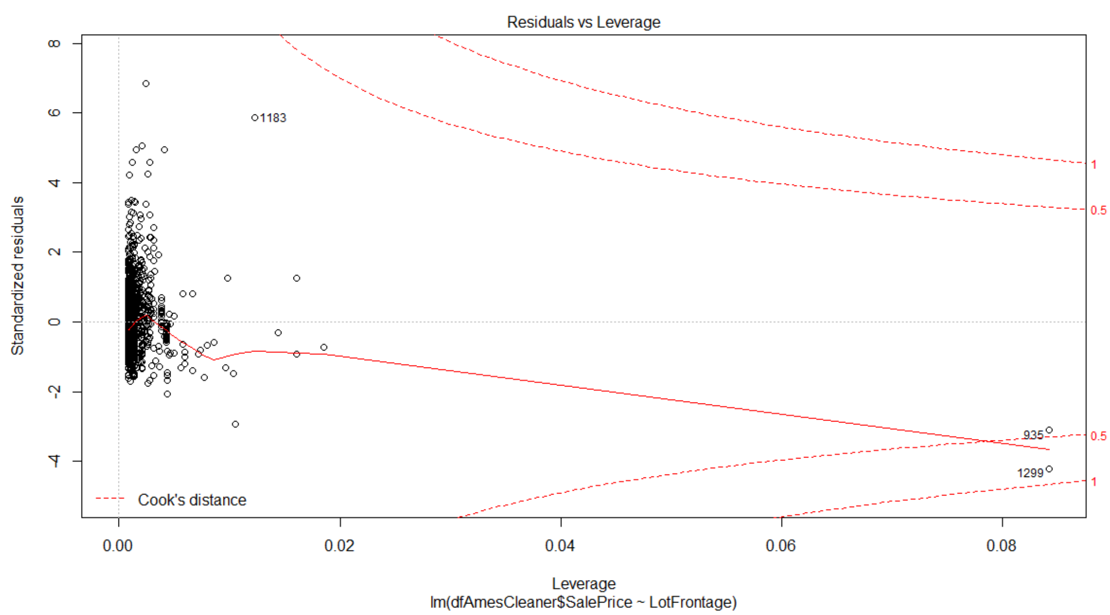
```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 96149.04    6881.97   13.97   <2e-16 ***
## LotFrontage  1208.02      92.83   13.01   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 78090 on 1199 degrees of freedom
##   (259 observations deleted due to missingness)
## Multiple R-squared:  0.1238, Adjusted R-squared:  0.123
## F-statistic: 169.4 on 1 and 1199 DF,  p-value: < 2.2e-16
```

```
plot(lm.fit6)
```


Residuals vs Fitted
692
1183
899
Residuals
6e+05
4e+05
2e+05
0e+00
-2e+05
-4e+05
150000
200000
250000
300000
350000
400000
450000
Fitted values
lm(dfAmesCleaner$SalePrice ~ LotFrontage)


Normal Q-Q
692
1183
899
Standardized residuals
8
6
4
2
0
-2
-4
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(dfAmesCleaner$SalePrice ~ LotFrontage)





Do any interactions appear to be statistically significant?

Ans - Yes, the linear regression model 3, seems to have a more of a statistical significance then that of the rest, as the data shows more correlation, then the rest of the models - as the other model's data reprersntation is all over the chart.

Part 5

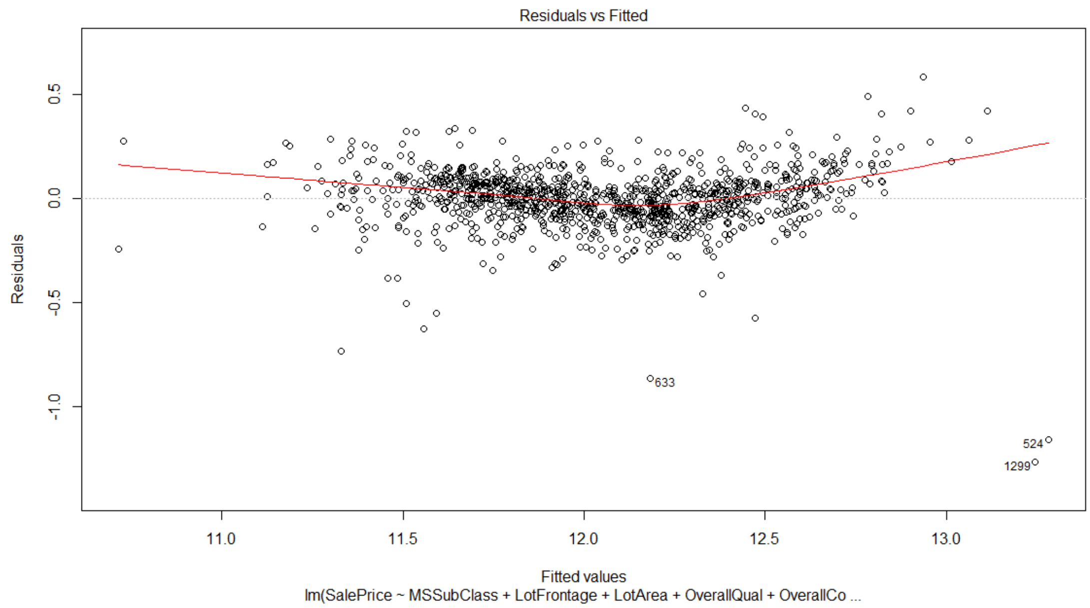
trying ln(x)

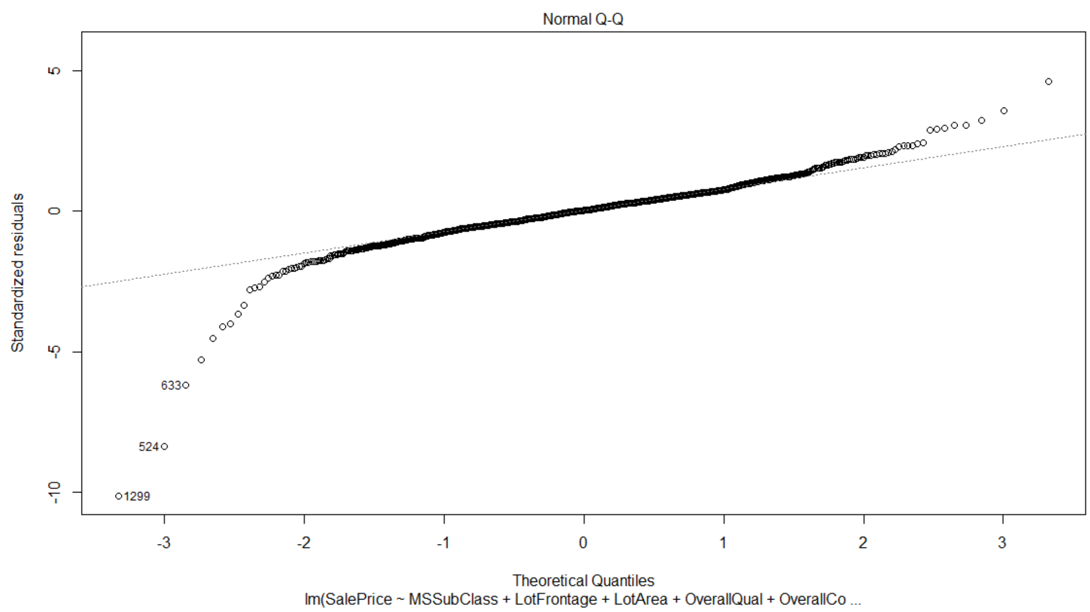
linear model for ln(x)

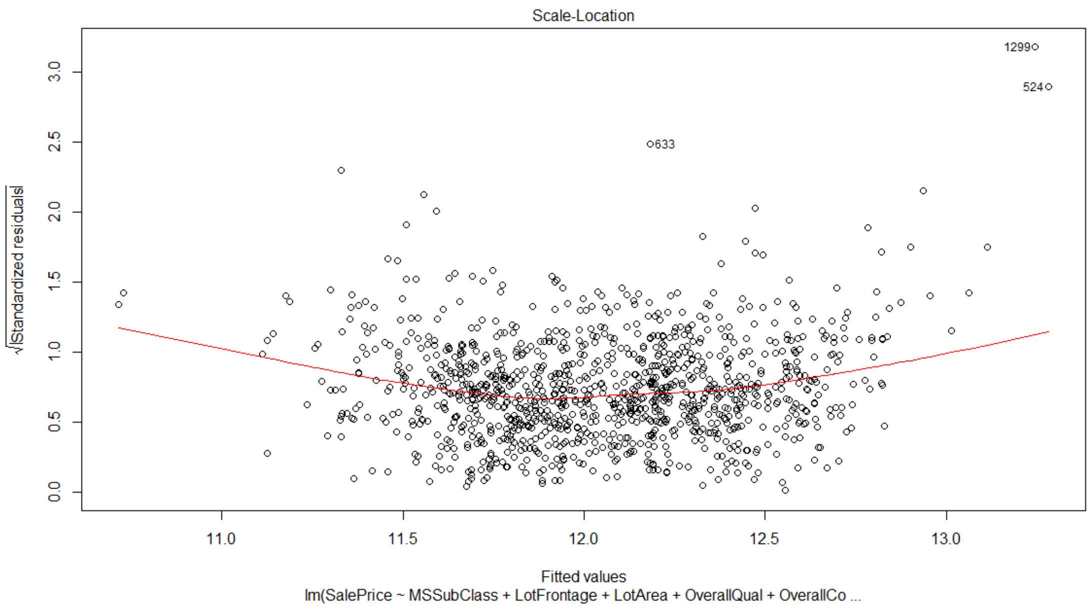
```
summary(lm.fit7)

##
## Call:
## lm(formula = SalePrice ~ MSSubClass + LotFrontage + LotArea +
##     OverallQual + OverallCond + YearBuilt + YearRemodAdd + MasVnrArea +
##     BsmtFinSF1 + BsmtFinSF2 + BsmtUnfSF + TotalBsmtSF + X1stFlrSF +
##     X2ndFlrSF + LowQualFinSF + GrLivArea + BsmtFullBath + BsmtHalfBath +
##     FullBath + HalfBath + BedroomAbvGr + KitchenAbvGr + TotRmsAbvGrd +
##     Fireplaces + GarageYrBlt + GarageCars + GarageArea + WoodDeckSF +
##     OpenPorchSF + EnclosedPorch + X3SsnPorch + ScreenPorch +
##     PoolArea + MiscVal + MoSold + YrSold, data = log_Ames)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26329 -0.06816  0.00364  0.07345  0.58461
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.6193416 49.6373326   0.053 0.957925
## MSSubClass    -0.0119718  0.0095469  -1.254 0.210112
## LotFrontage    0.0090561  0.0200622   0.451 0.651792
## LotArea        0.0914991  0.0141655   6.459 1.59e-10 ***
## OverallQual    0.5307110  0.0390010  13.608  < 2e-16 ***
## OverallCond    0.3427799  0.0341021  10.052  < 2e-16 ***
## YearBuilt      4.3555289  0.6563978   6.636 5.10e-11 ***
## YearRemodAdd   1.9589227  0.6488635   3.019 0.002595 **
## MasVnrArea    -0.0004934  0.0019408  -0.254 0.799370
## BsmtFinSF1     0.0107095  0.0020485   5.228 2.05e-07 ***
## BsmtFinSF2    -0.0039379  0.0026190  -1.504 0.132979
## BsmtUnfSF     -0.0009542  0.0037356  -0.255 0.798443
## TotalBsmtSF    0.0121522  0.0060066   2.023 0.043305 *
## X1stFlrSF     -0.0040418  0.0550967  -0.073 0.941535
## X2ndFlrSF     -0.0128756  0.0055002  -2.341 0.019417 *
## LowQualFinSF  -0.0063520  0.0069431  -0.915 0.360462
## GrLivArea      0.4704074  0.0695219   6.766 2.16e-11 ***
## BsmtFullBath   0.0546865  0.0170619   3.205 0.001389 **
## BsmtHalfBath  -0.0006978  0.0283299  -0.025 0.980353
## FullBath       0.1014748  0.0334067   3.038 0.002442 **
## HalfBath       0.0460988  0.0190604   2.419 0.015746 *
## BedroomAbvGr  -0.1056927  0.0284035  -3.721 0.000209 ***
## KitchenAbvGr  -0.3009993  0.0652447  -4.613 4.43e-06 ***
## TotRmsAbvGrd   0.0964442  0.0430869   2.238 0.025400 *
## Fireplaces     0.0485747  0.0136863   3.549 0.000403 ***
```

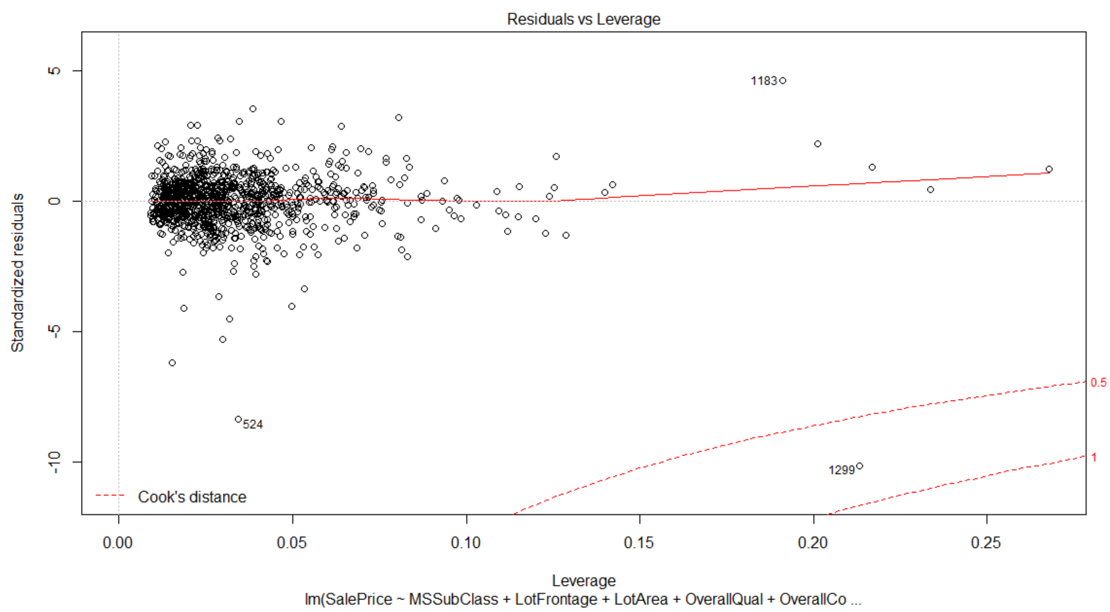
```
## GarageYrBlt    1.1124269  0.7118456   1.563 0.118406    
## GarageCars     0.1665843  0.0394182   4.226 2.58e-05 ***
## GarageArea     0.0037657  0.0246306   0.153 0.878516    
## WoodDeckSF     0.0029815  0.0018508   1.611 0.107488    
## OpenPorchSF    0.0028206  0.0024338   1.159 0.246743    
## EnclosedPorch  0.0054430  0.0028682   1.898 0.058000 .  
## X3SsnPorch     0.0044624  0.0065195   0.684 0.493828    
## ScreenPorch    0.0089372  0.0030815   2.900 0.003804 ** 
## PoolArea      -0.0279696  0.0095702  -2.923 0.003544 ** 
## MiscVal       -0.0081570  0.0039091  -2.087 0.037152 *  
## MoSold         0.0071512  0.0101640   0.704 0.481845    
## YrSold        -6.9854977  6.5114977  -1.073 0.283602    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1405 on 1084 degrees of freedom
##   (339 observations deleted due to missingness)
## Multiple R-squared:  0.8786, Adjusted R-squared:  0.8746 
## F-statistic:   218 on 36 and 1084 DF,  p-value: < 2.2e-16
```

```r
plot(lm.fit7)
```


Residuals vs Fitted
Residuals
Fitted values
lm(SalePrice ~ MSSubClass + LotFrontage + LotArea + OverallQual + OverallCo ...
633
524
1299


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(SalePrice ~ MSSubClass + LotFrontage + LotArea + OverallQual + OverallCo ...
633
524
1299





trying x^(2)

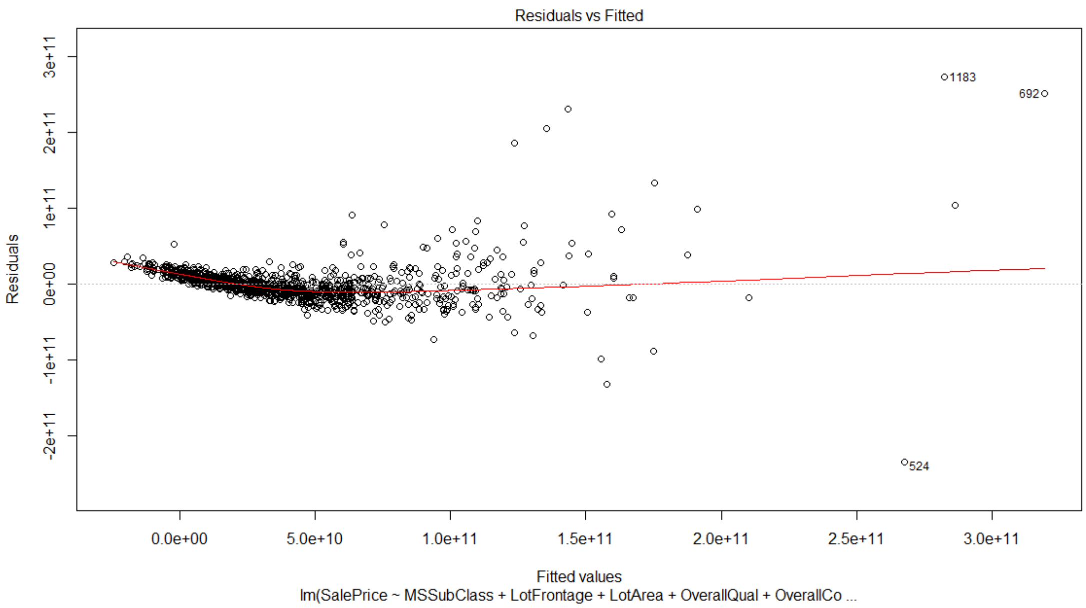
linear model for x^(2)

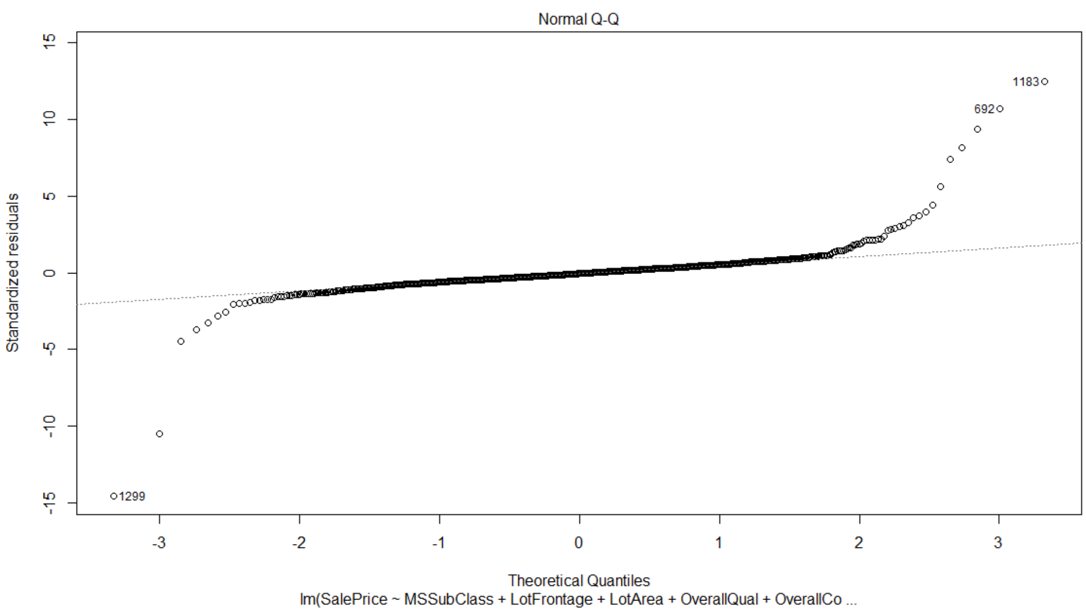
```
summary(lm.fit8)

##
## Call:
## lm(formula = SalePrice ~ MSSubClass + LotFrontage + LotArea +
##     OverallQual + OverallCond + YearBuilt + YearRemodAdd + MasVnrArea +
##     BsmtFinSF1 + BsmtFinSF2 + BsmtUnfSF + TotalBsmtSF + X1stFlrSF +
##     X2ndFlrSF + LowQualFinSF + GrLivArea + BsmtFullBath + BsmtHalfBath +
```

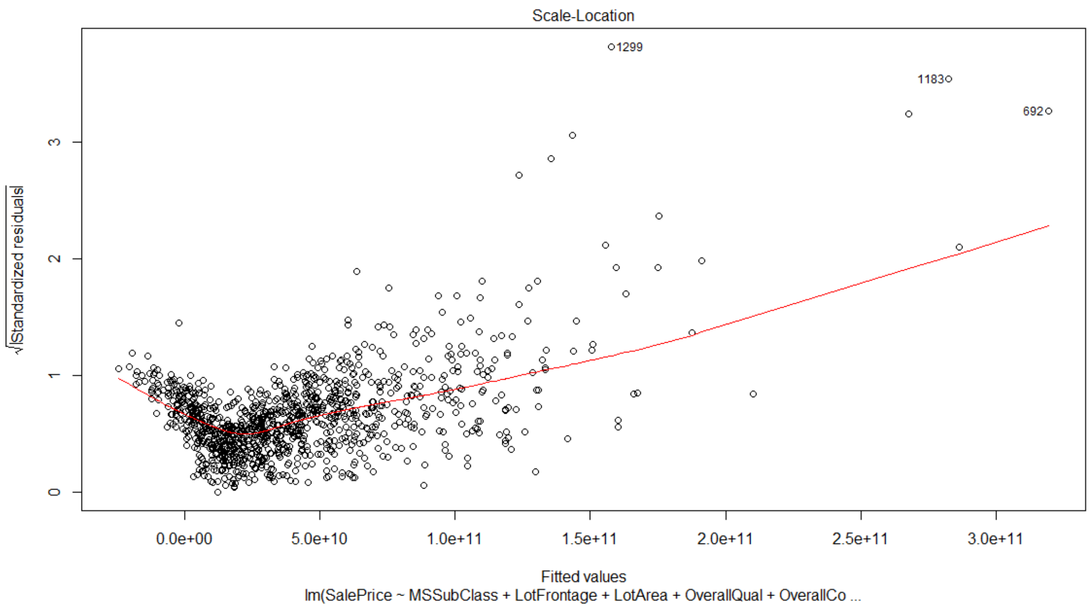
```
##     FullBath + HalfBath + BedroomAbvGr + KitchenAbvGr + TotRmsAbvGrd +
##     Fireplaces + GarageYrBlt + GarageCars + GarageArea + WoodDeckSF +
##     OpenPorchSF + EnclosedPorch + X3SsnPorch + ScreenPorch +
##     PoolArea + MiscVal + MoSold + YrSold, data = xsquared_Ames)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -2.336e+11 -1.112e+10 -1.397e+09  8.256e+09  2.724e+11
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.685e+11  5.978e+11  -0.282  0.77814
## MSSubClass    -4.670e+05  1.171e+05  -3.988 7.10e-05 ***
## LotFrontage   -5.087e+05  1.895e+05  -2.685  0.00737 **
## LotArea        1.216e+00  5.746e-01   2.116  0.03457 *
## OverallQual    7.601e+08  6.852e+07  11.093  < 2e-16 ***
## OverallCond    9.774e+07  6.934e+07   1.410  0.15894
## YearBuilt      2.956e+04  1.536e+04   1.925  0.05454 .
## YearRemodAdd  -1.341e+04  1.542e+04  -0.869  0.38486
## MasVnrArea     3.642e+04  5.490e+03   6.634 5.14e-11 ***
## BsmtFinSF1    -1.513e+04  2.619e+03  -5.777 9.95e-09 ***
## BsmtFinSF2    -1.832e+04  6.905e+03  -2.653  0.00808 **
## BsmtUnfSF     -1.024e+04  2.272e+03  -4.505 7.36e-06 ***
## TotalBsmtSF    4.789e+03  2.335e+03   2.051  0.04052 *
## X1stFlrSF      1.485e+04  3.287e+03   4.517 6.96e-06 ***
## X2ndFlrSF      3.440e+04  7.730e+03   4.450 9.49e-06 ***
## LowQualFinSF   2.380e+04  5.001e+04   0.476  0.63422
## GrLivArea     -3.150e+02  2.405e+03  -0.131  0.89583
## BsmtFullBath   2.533e+09  6.390e+08   3.965 7.83e-05 ***
## BsmtHalfBath   1.942e+09  1.139e+09   1.705  0.08856 .
## FullBath      -3.868e+08  4.603e+08  -0.840  0.40089
## HalfBath      -1.612e+09  6.840e+08  -2.357  0.01858 *
## BedroomAbvGr  -1.236e+09  1.936e+08  -6.387 2.51e-10 ***
## KitchenAbvGr  -2.174e+09  9.152e+08  -2.376  0.01770 *
## TotRmsAbvGrd   1.845e+08  6.186e+07   2.982  0.00292 **
## Fireplaces     3.035e+08  4.073e+08   0.745  0.45631
## GarageYrBlt   -1.881e+04  1.569e+04  -1.199  0.23080
## GarageCars     1.126e+09  3.836e+08   2.934  0.00341 **
## GarageArea     1.142e+04  6.340e+03   1.801  0.07195 .
## WoodDeckSF     2.724e+04  1.558e+04   1.748  0.08074 .
## OpenPorchSF   -1.414e+05  4.650e+04  -3.040  0.00242 **
## EnclosedPorch -1.536e+05  5.425e+04  -2.830  0.00473 **
## X3SsnPorch     3.300e+04  7.318e+04   0.451  0.65208
## ScreenPorch    7.206e+04  5.211e+04   1.383  0.16696
## PoolArea       9.492e+04  3.850e+04   2.465  0.01384 *
## MiscVal       -3.486e+03  2.762e+03  -1.262  0.20718
## MoSold        -2.880e+07  1.919e+07  -1.501  0.13363
## YrSold         3.869e+04  1.480e+05   0.261  0.79381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.588e+10 on 1084 degrees of freedom
##   (339 observations deleted due to missingness)
## Multiple R-squared:  0.7006, Adjusted R-squared:  0.6907
## F-statistic: 70.46 on 36 and 1084 DF,  p-value: < 2.2e-16
```

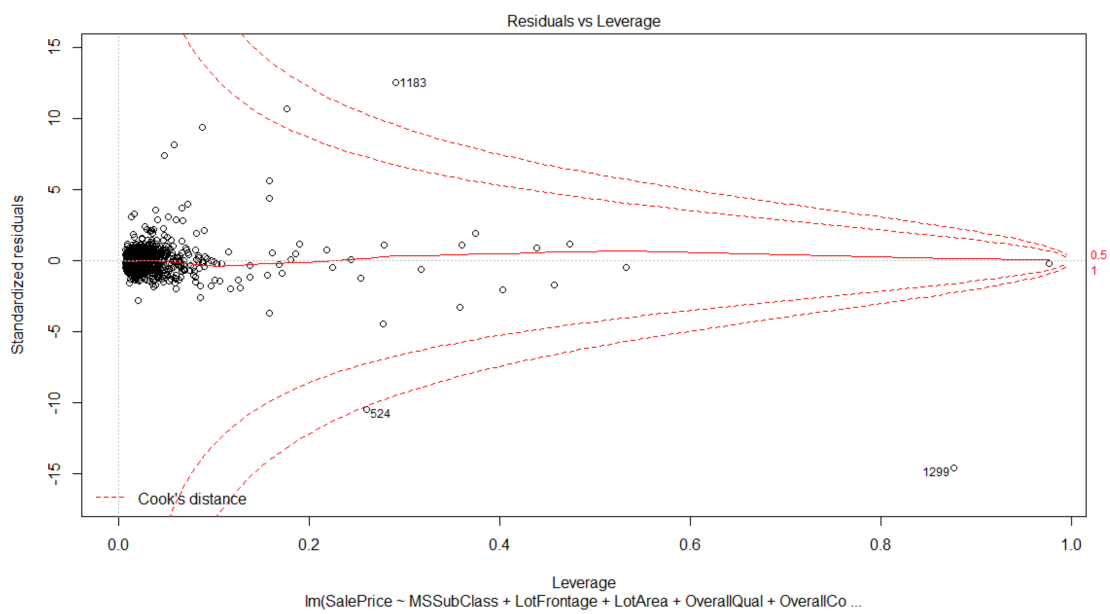
```r
plot(lm.fit8)
```


Residuals vs Fitted
Residuals
3e+11
2e+11
1e+11
0e+00
-1e+11
-2e+11
1183
692
524
0.0e+00
5.0e+10
1.0e+11
1.5e+11
2.0e+11
2.5e+11
3.0e+11
Fitted values
lm(SalePrice ~ MSSubClass + LotFrontage + LotArea + OverallQual + OverallCo ...


Normal Q-Q
Standardized residuals
15
10
5
0
-5
-10
-15
1183
692
1299
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(SalePrice ~ MSSubClass + LotFrontage + LotArea + OverallQual + OverallCo ...




```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



trying square root

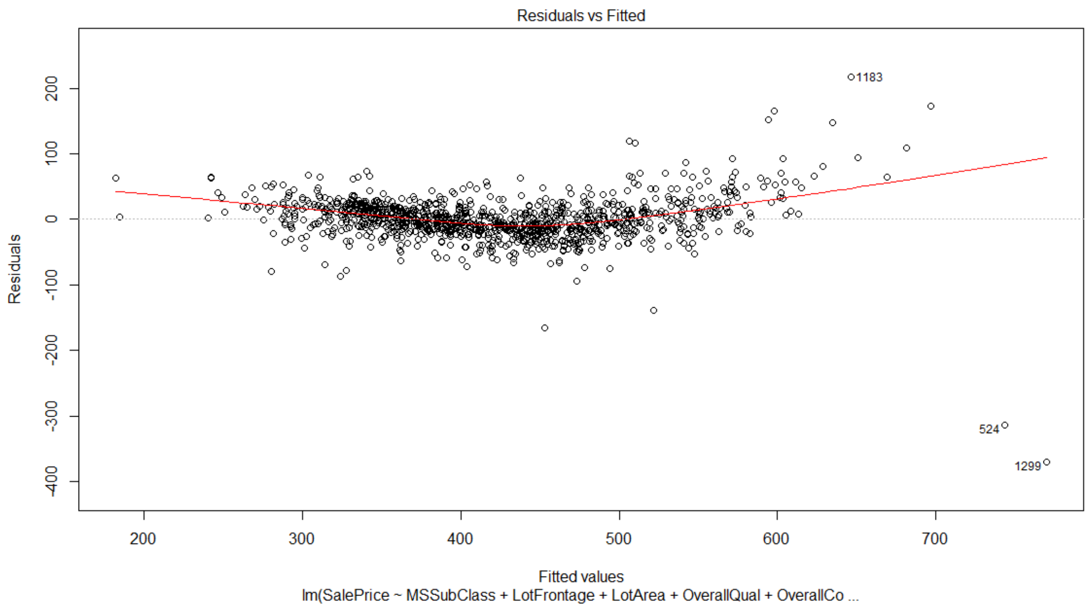
linear model for square root

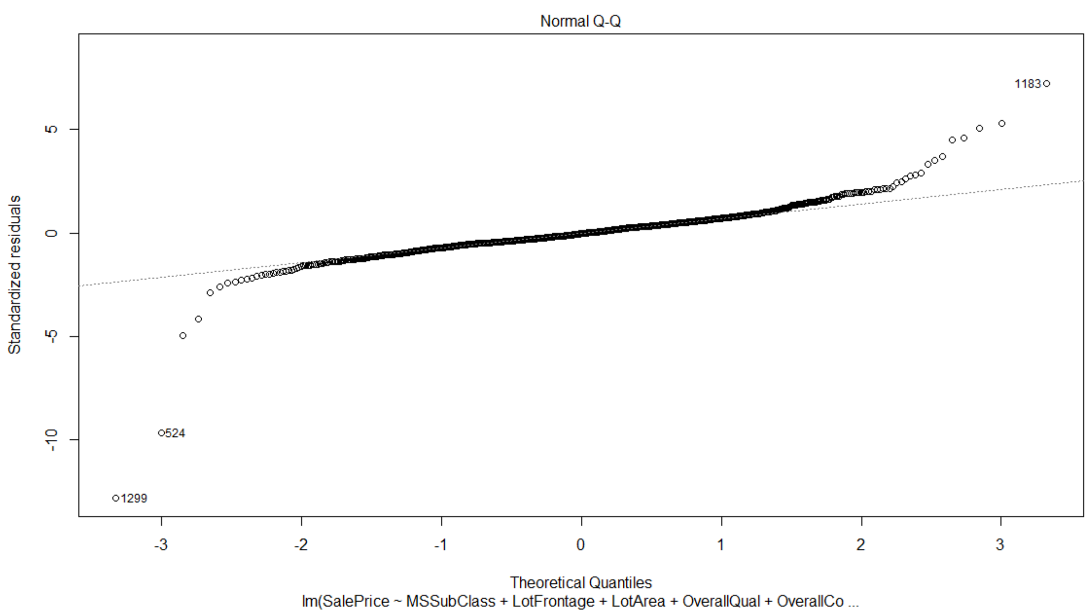
```r
summary(lm.fit9)
```

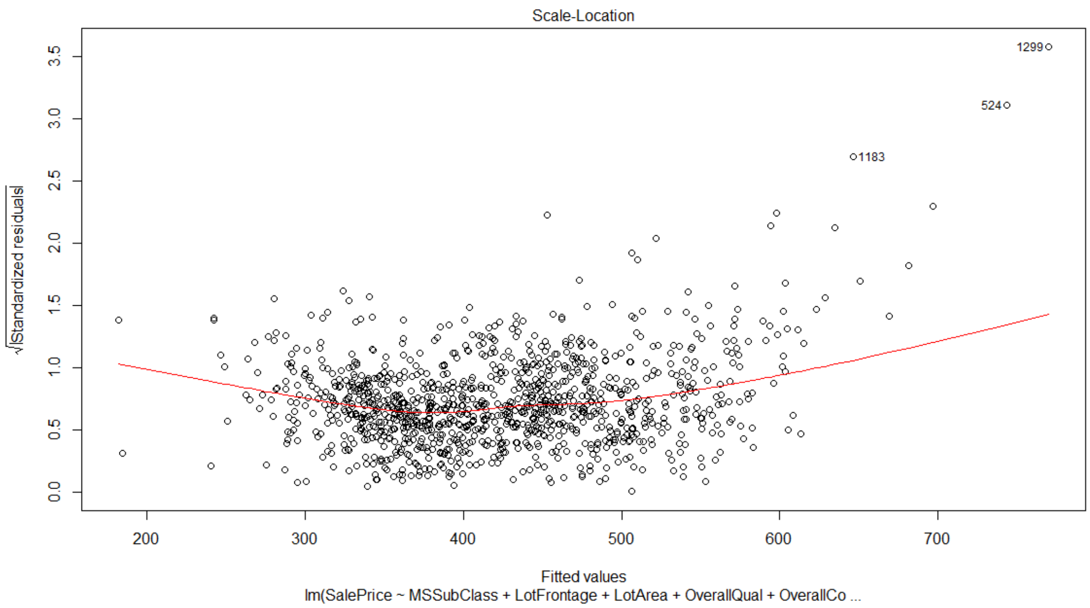
```
##
## Call:
## lm(formula = SalePrice ~ MSSubClass + LotFrontage + LotArea +
##     OverallQual + OverallCond + YearBuilt + YearRemodAdd + MasVnrArea +
##     BsmtFinSF1 + BsmtFinSF2 + BsmtUnfSF + TotalBsmtSF + X1stFlrSF +
##     X2ndFlrSF + LowQualFinSF + GrLivArea + BsmtFullBath + BsmtHalfBath +
##     FullBath + HalfBath + BedroomAbvGr + KitchenAbvGr + TotRmsAbvGrd +
##     Fireplaces + GarageYrBlt + GarageCars + GarageArea + WoodDeckSF +
##     OpenPorchSF + EnclosedPorch + X3SsnPorch + ScreenPorch +
##     PoolArea + MiscVal + MoSold + YrSold, data = root_Ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -369.82  -16.04   -1.04   15.20  216.88
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -913.41714 3117.01379  -0.293 0.769546
## MSSubClass      -2.05016    0.57873  -3.543 0.000413 ***
## LotFrontage     -0.39591    1.06997  -0.370 0.711443
## LotArea          0.29012    0.05545   5.232 2.01e-07 ***
## OverallQual     94.76502    7.21574  13.133  < 2e-16 ***
## OverallCond     42.34524    6.45701   6.558 8.42e-11 ***
## YearBuilt       33.71148    7.18527   4.692 3.05e-06 ***
## YearRemodAdd    17.48548    7.05054   2.480 0.013289 *
## MasVnrArea       0.28029    0.15593   1.798 0.072530 .
## BsmtFinSF1       0.55176    0.21819   2.529 0.011584 *
## BsmtFinSF2      -0.14610    0.23334  -0.626 0.531370
## BsmtUnfSF       -0.10421    0.30833  -0.338 0.735435
## TotalBsmtSF      0.67166    0.39400   1.705 0.088530 .
## X1stFlrSF       -3.23046    1.37305  -2.353 0.018812 *
## X2ndFlrSF       -1.62357    0.58761  -2.763 0.005824 **
## LowQualFinSF    -1.21239    0.65044  -1.864 0.062600 .
## GrLivArea        7.94421    1.52105   5.223 2.11e-07 ***
## BsmtFullBath    14.42487    7.24475   1.991 0.046723 *
## BsmtHalfBath    -1.69028   11.34438  -0.149 0.881584
## FullBath        21.55222   10.31754   2.089 0.036950 *
## HalfBath         7.39389    7.59571   0.973 0.330557
## BedroomAbvGr   -34.42419    7.48606  -4.598 4.76e-06 ***
## KitchenAbvGr   -80.48205   19.74450  -4.076 4.91e-05 ***
## TotRmsAbvGrd    24.99656    7.59550   3.291 0.001031 **
## Fireplaces      14.22018    5.20864   2.730 0.006434 **
## GarageYrBlt      4.06708    7.54632   0.539 0.590033
## GarageCars      50.00360   11.15964   4.481 8.23e-06 ***
## GarageArea       0.20275    0.53077   0.382 0.702546
## WoodDeckSF       0.34889    0.17087   2.042 0.041413 *
## OpenPorchSF      0.31577    0.27117   1.164 0.244490
## EnclosedPorch    0.39221    0.29077   1.349 0.177658
## X3SsnPorch       0.54796    0.60224   0.910 0.363091
## ScreenPorch      0.88403    0.29846   2.962 0.003123 **
```

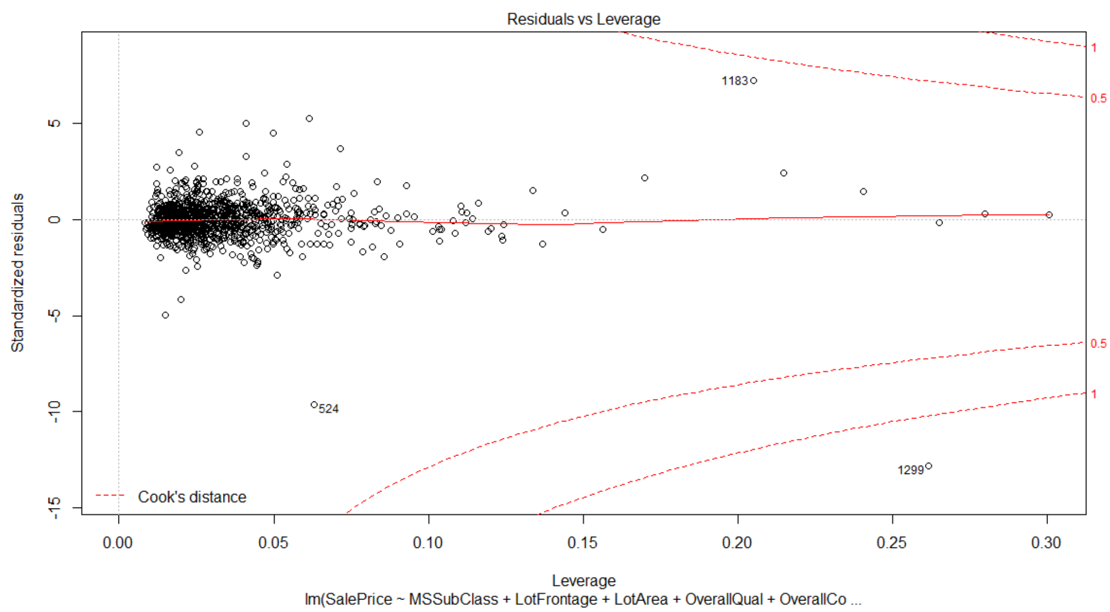
```
## PoolArea        -1.97972    0.65345  -3.030 0.002507 **
## MiscVal         -0.24343    0.22568  -1.079 0.280971
## MoSold           0.42651    2.00031   0.213 0.831196
## YrSold         -39.18403   69.43457  -0.564 0.572646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33.61 on 1084 degrees of freedom
##   (339 observations deleted due to missingness)
## Multiple R-squared:  0.8575, Adjusted R-squared:  0.8528
## F-statistic: 181.3 on 36 and 1084 DF,  p-value: < 2.2e-16
```

```r
plot(lm.fit9)
```


Residuals vs Fitted
Residuals
Fitted values
lm(SalePrice ~ MSSubClass + LotFrontage + LotArea + OverallQual + OverallCo ...
1183
524
1299
200
100
0
-100
-200
-300
-400
200
300
400
500
600
700


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(SalePrice ~ MSSubClass + LotFrontage + LotArea + OverallQual + OverallCo ...
1183
524
1299
5
0
-5
-10
-3
-2
-1
0
1
2
3





Do any of these make sense to include in a model of SalePrice?

Ans - Yes, it does. The log function can be used to plot prices for market inflation - as the log function is used for calculating elastic demand thus making adjustments within the log function depending on the inflation rate can help us formulate a regression model for SalePrice of a house during and after inflation.So, much like the above example other regression models for economic situations that impact the price of a house can also be plotted out - by using the log, square and square root functions.

Comment on your findings.

The graphs are a representation of the data being processed through all three individual functions - they all show correlation and none of them present a graph with large deviation, some of the plots are dispersed compared to the other ones but a regression line can still be plotted out.