Resultados

Cícero

2022-09-30

R Markdown

Variáveis testadas:

SalePrice x LotArea : Significativa

SalePrice x ExterQual : Siginificativa

SalePrice x ExterCond : significativa

SalePrice x OverallCond : Meio meh

SalePrice x Neighborhood : significativa para algumas vizinhaças

SalePrice x TotalBsmtSF : significativa

SalePrice x MiscVal : Poha nenhuma

SalePrice x MiscFeature : Poha nenhuma

```
dados <- read.csv('train.csv')
modeloLotArea <- lm(SalePrice ~ LotArea,dados)
summary(modeloLotArea)
```

```
##
## Call:
## lm(formula = SalePrice ~ LotArea, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -275668  -48169  -17725   31248  553356
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.588e+05  2.915e+03   54.49   <2e-16 ***
## LotArea     2.100e+00  2.011e-01   10.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 76650 on 1458 degrees of freedom
## Multiple R-squared:  0.06961,    Adjusted R-squared:  0.06898
## F-statistic: 109.1 on 1 and 1458 DF,  p-value: < 2.2e-16
```

```r
modeloExterQual <- lm(SalePrice ~ ExterQual,dados)
summary(modeloExterQual)
```

```
##
## Call:
## lm(formula = SalePrice ~ ExterQual, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -207361  -34341   -5841   25211  513366
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   367361       7972   46.08   <2e-16 ***
## ExterQualFa  -279376      17310  -16.14   <2e-16 ***
## ExterQualGd  -135728       8386  -16.18   <2e-16 ***
## ExterQualTA  -223020       8198  -27.20   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 57490 on 1456 degrees of freedom
## Multiple R-squared:  0.4774, Adjusted R-squared:  0.4763
## F-statistic: 443.3 on 3 and 1456 DF,  p-value: < 2.2e-16
```

```r
modeloExterCond <- lm(SalePrice ~ ExterCond,dados)
summary(modeloExterCond)
```

```
##
## Call:
## lm(formula = SalePrice ~ ExterCond, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -149135  -50660  -16898   32193  570965
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   201333      45384   4.436 9.84e-06 ***
## ExterCondFa   -98738      47753  -2.068   0.0388 *
## ExterCondGd   -32436      45847  -0.707   0.4794
## ExterCondPo  -124833      90767  -1.375   0.1692
## ExterCondTA   -17298      45437  -0.381   0.7035
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 78610 on 1455 degrees of freedom
## Multiple R-squared:  0.02362,    Adjusted R-squared:  0.02093
## F-statistic: 8.799 on 4 and 1455 DF,  p-value: 5.107e-07
```

```r
modeloOverallCond <- lm(SalePrice ~ OverallCond,dados)
summary(modeloOverallCond)
```

```
##
## Call:
## lm(formula = SalePrice ~ OverallCond, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -159924  -49459  -16590   30881  576439
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   211910      10597  19.997  < 2e-16 ***
## OverallCond    -5558       1864  -2.982  0.00291 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79230 on 1458 degrees of freedom
## Multiple R-squared:  0.006062,   Adjusted R-squared:  0.00538
## F-statistic: 8.892 on 1 and 1458 DF,  p-value: 0.002912
```

```
modeloNeighborhood <- lm(SalePrice ~ Neighborhood,dados)
summary(modeloNeighborhood)
```

```
##
## Call:
## lm(formula = SalePrice ~ Neighborhood, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -162271  -27552   -5324   19685  419705
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           194871      13097  14.879  < 2e-16 ***
## NeighborhoodBlueste   -57371      40367  -1.421 0.155463
## NeighborhoodBrDale    -90377      18809  -4.805 1.71e-06 ***
## NeighborhoodBrkSide   -70037      14893  -4.703 2.81e-06 ***
## NeighborhoodClearCr    17695      16603   1.066 0.286721
## NeighborhoodCollgCr     3095      13819   0.224 0.822820
## NeighborhoodCrawfor    15754      15123   1.042 0.297712
## NeighborhoodEdwards   -66651      14166  -4.705 2.78e-06 ***
## NeighborhoodGilbert    -2016      14437  -0.140 0.888944
## NeighborhoodIDOTRR    -94747      15822  -5.988 2.67e-09 ***
## NeighborhoodMeadowV   -96294      18522  -5.199 2.29e-07 ***
## NeighborhoodMitchel   -38601      15200  -2.540 0.011204 *
## NeighborhoodNAmes     -49024      13582  -3.609 0.000318 ***
## NeighborhoodNoRidge   140424      15577   9.015  < 2e-16 ***
## NeighborhoodNPkVill   -52176      22260  -2.344 0.019217 *
## NeighborhoodNridgHt   121400      14470   8.390  < 2e-16 ***
## NeighborhoodNWAmes     -5821      14542  -0.400 0.689011
## NeighborhoodOldTown   -66646      14047  -4.744 2.30e-06 ***
## NeighborhoodSawyer    -58078      14523  -3.999 6.69e-05 ***
## NeighborhoodSawyerW    -8315      14864  -0.559 0.575974
## NeighborhoodSomerst    30509      14333   2.129 0.033456 *
## NeighborhoodStoneBr   115628      16975   6.812 1.42e-11 ***
```

```
## NeighborhoodSWISU     -52280      16975  -3.080 0.002111 **
## NeighborhoodTimber     47377      15756   3.007 0.002686 **
## NeighborhoodVeenker    43902      20895   2.101 0.035810 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54000 on 1435 degrees of freedom
## Multiple R-squared:  0.5456, Adjusted R-squared:  0.538
## F-statistic: 71.78 on 24 and 1435 DF,  p-value: < 2.2e-16
```

```
modeloTotalBsmtSF <- lm(SalePrice ~ TotalBsmtSF,dados)
summary(modeloTotalBsmtSF)
```

```
##
## Call:
## lm(formula = SalePrice ~ TotalBsmtSF, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -582310  -39612  -14095   33315  420018
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 63430.629   4286.892   14.80   <2e-16 ***
## TotalBsmtSF   111.110      3.745   29.67   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 62750 on 1458 degrees of freedom
## Multiple R-squared:  0.3765, Adjusted R-squared:  0.3761
## F-statistic: 880.3 on 1 and 1458 DF,  p-value: < 2.2e-16
```

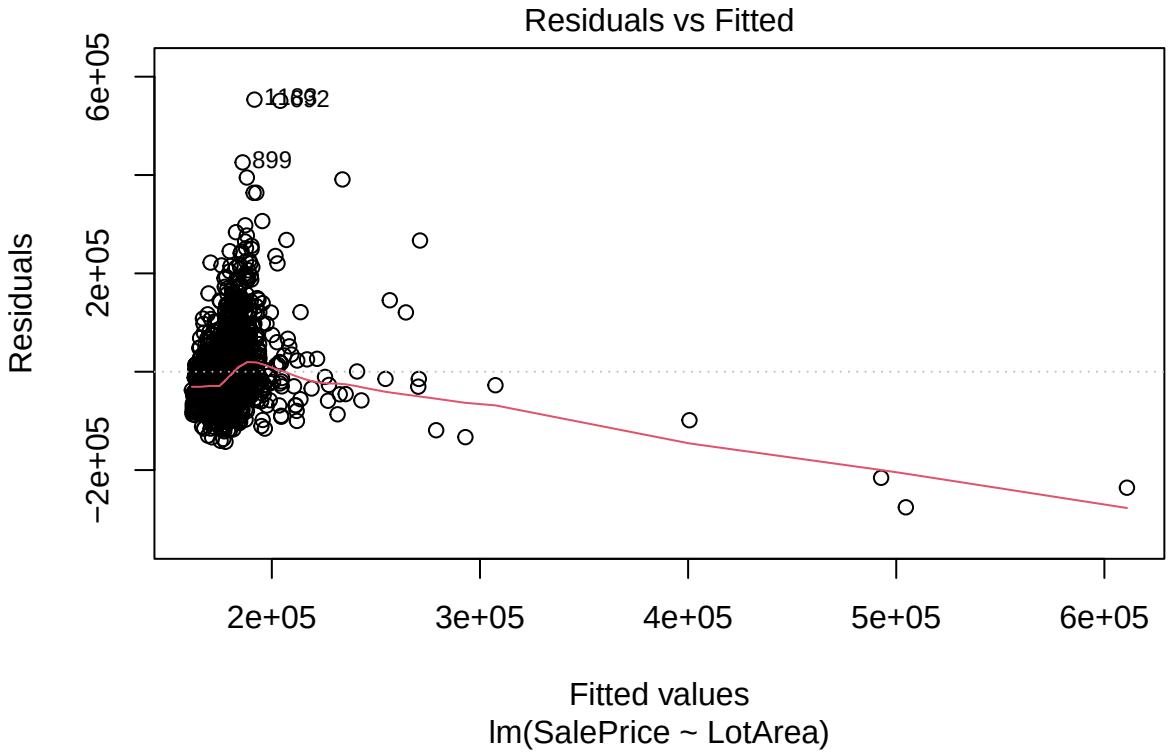
```
modeloMiscVal <- lm(SalePrice ~ MiscVal,dados)
summary(modeloMiscVal)
```

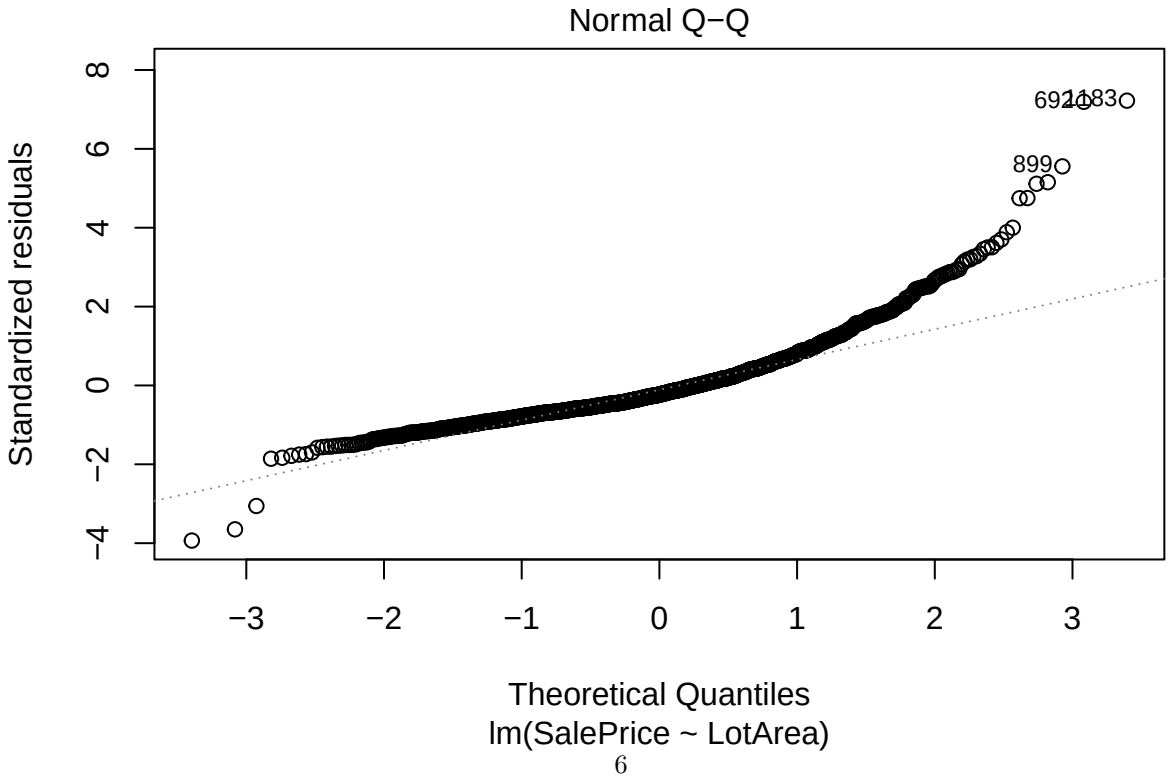
```
##
## Call:
## lm(formula = SalePrice ~ MiscVal, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -146169  -51069  -17569   32931  573931
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 181068.755   2087.330  86.747   <2e-16 ***
## MiscVal         -3.393      4.193  -0.809    0.418
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79450 on 1458 degrees of freedom
## Multiple R-squared:  0.000449,   Adjusted R-squared:  -0.0002366
## F-statistic: 0.6549 on 1 and 1458 DF,  p-value: 0.4185
```

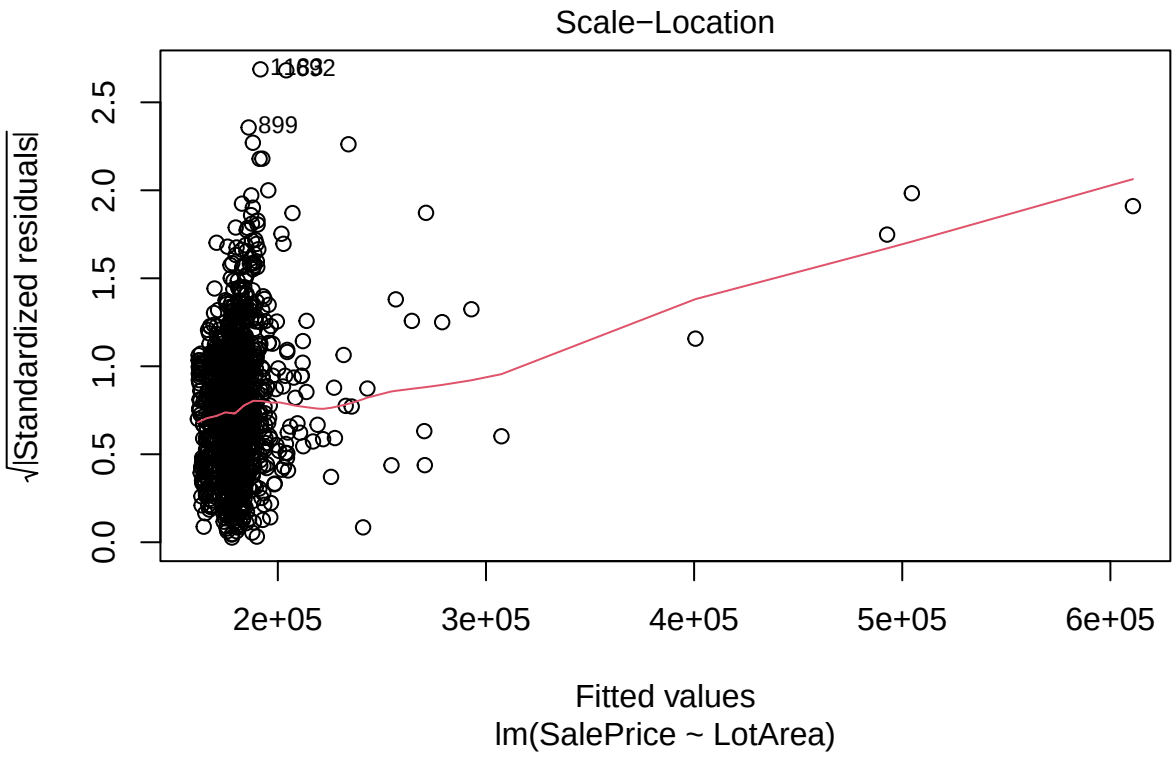
```r
modeloMiscFeature <- lm(SalePrice ~ MiscFeature,dados)
summary(modeloMiscFeature)
```

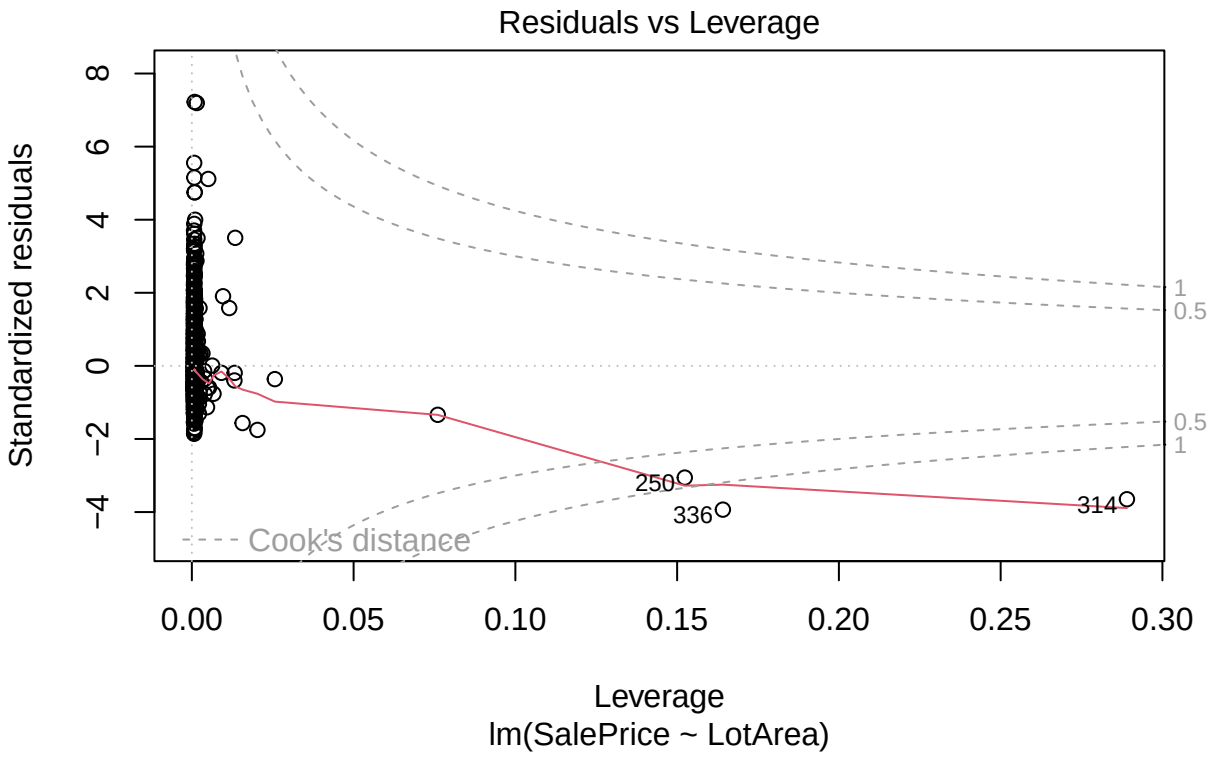
```
##
## Call:
## lm(formula = SalePrice ~ MiscFeature, data = dados)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -95195 -31063  -5188  20422 125812
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       170750      36304   4.703 2.05e-05 ***
## MiscFeatureOthr   -76750      51342  -1.495    0.141
## MiscFeatureShed   -19562      37037  -0.528    0.600
## MiscFeatureTenC    79250      62880   1.260    0.213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 51340 on 50 degrees of freedom
##   (1406 observations deleted due to missingness)
## Multiple R-squared:  0.1146, Adjusted R-squared:  0.06148
## F-statistic: 2.157 on 3 and 50 DF,  p-value: 0.1047
```

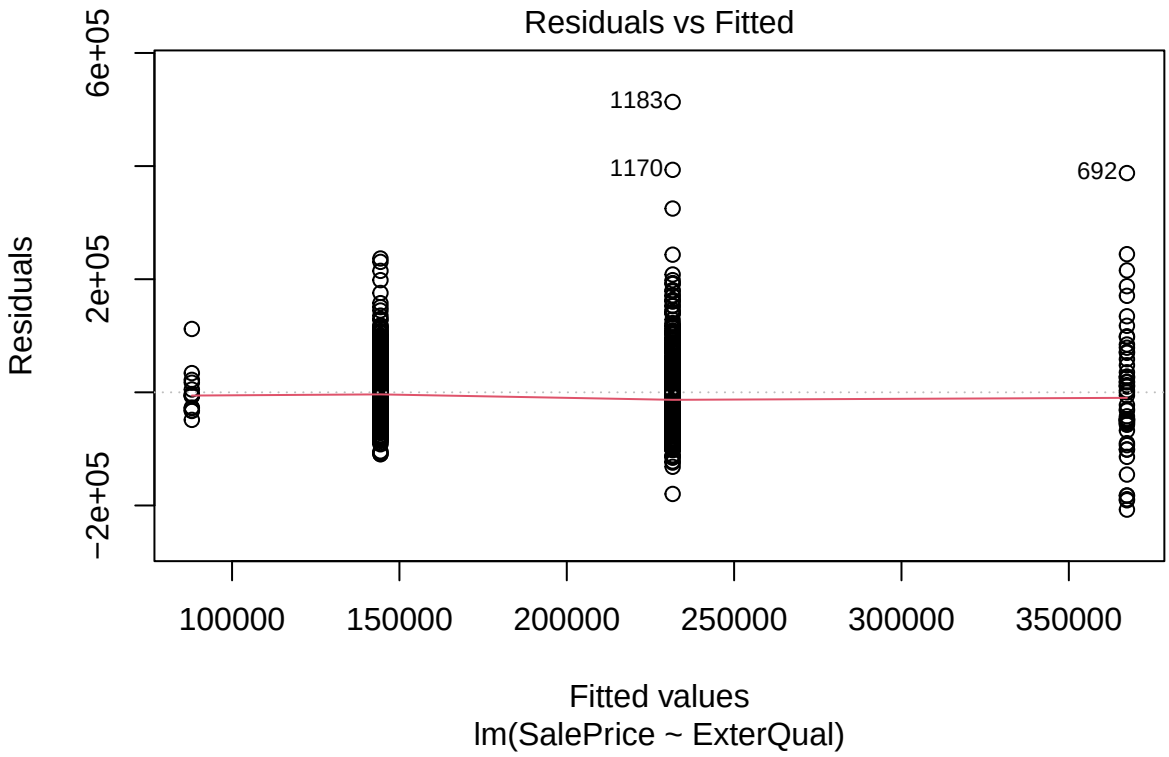
Lm Plots

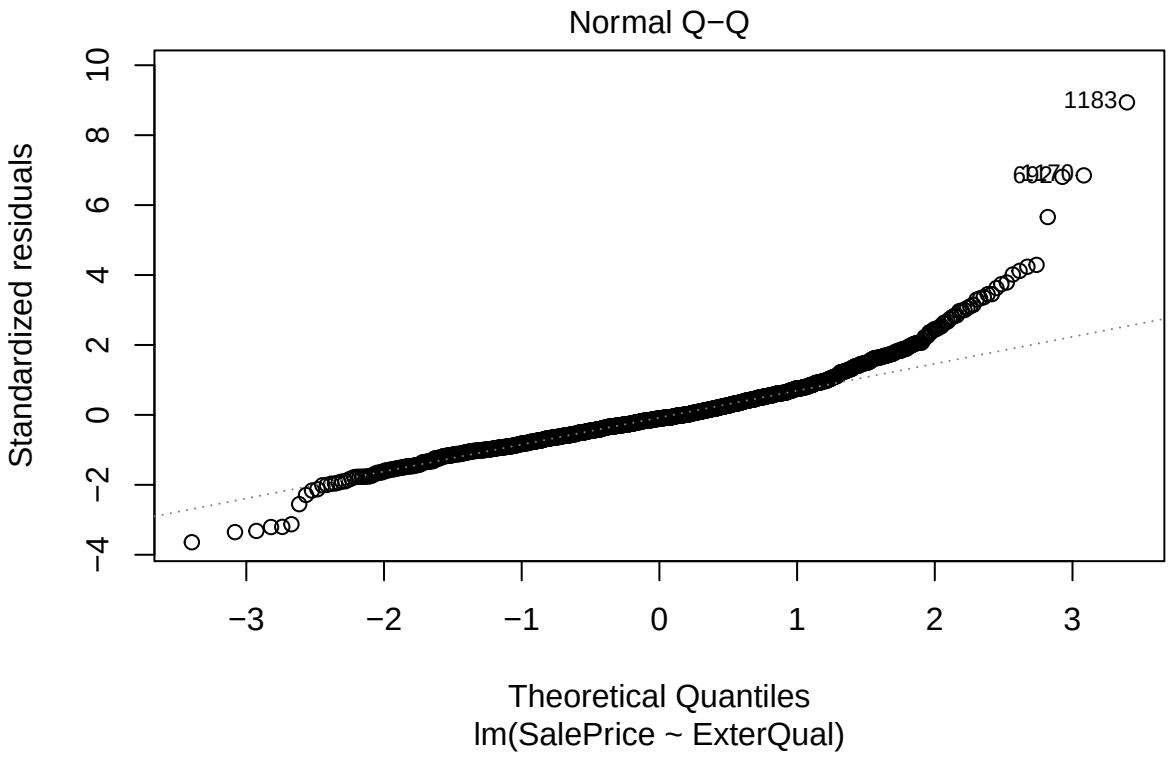

Residuals vs Fitted
Residuals
Fitted values
lm(SalePrice ~ LotArea)
899

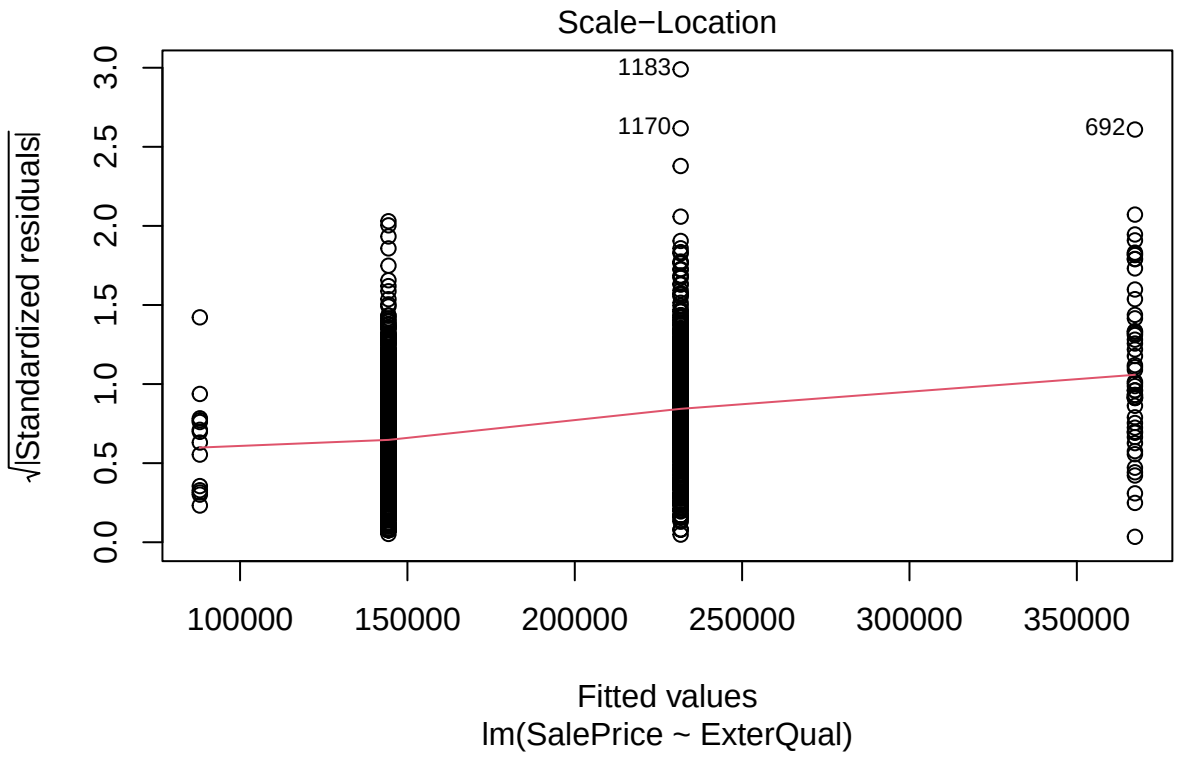

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(SalePrice ~ LotArea)
899

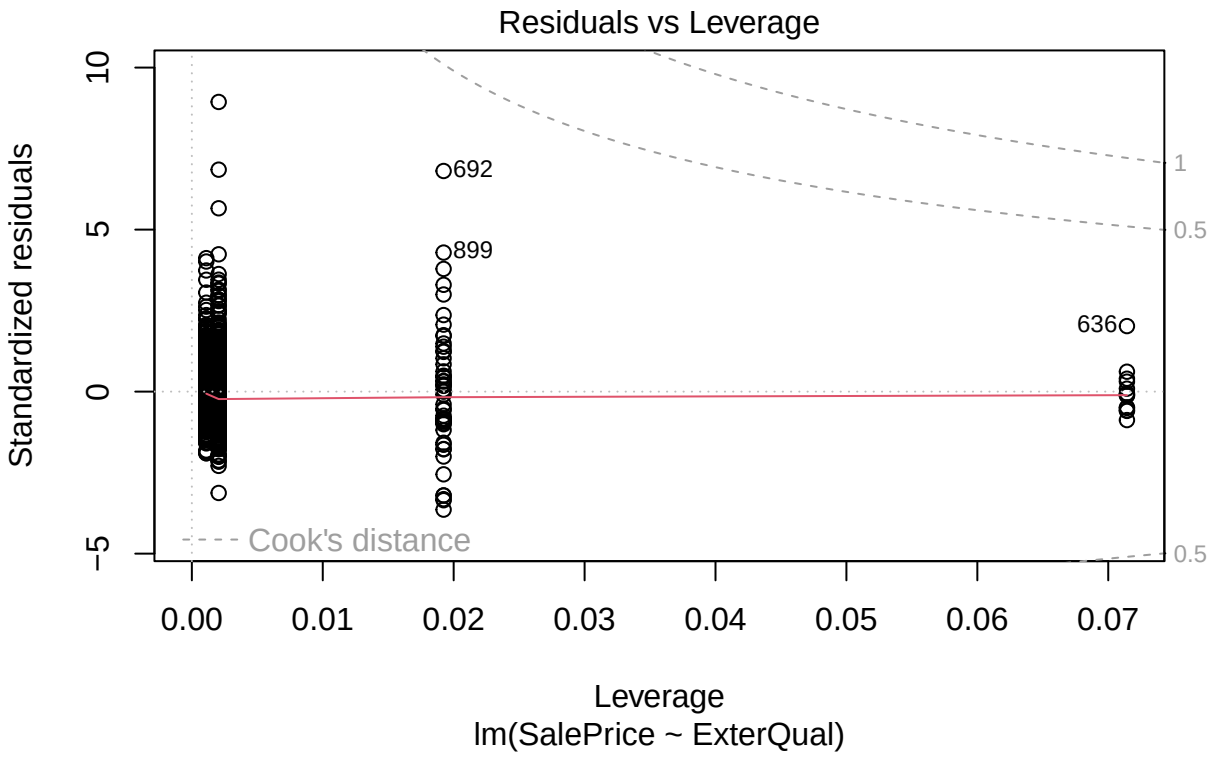

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
899
2e+05
3e+05
4e+05
5e+05
6e+05
Fitted values
lm(SalePrice ~ LotArea)

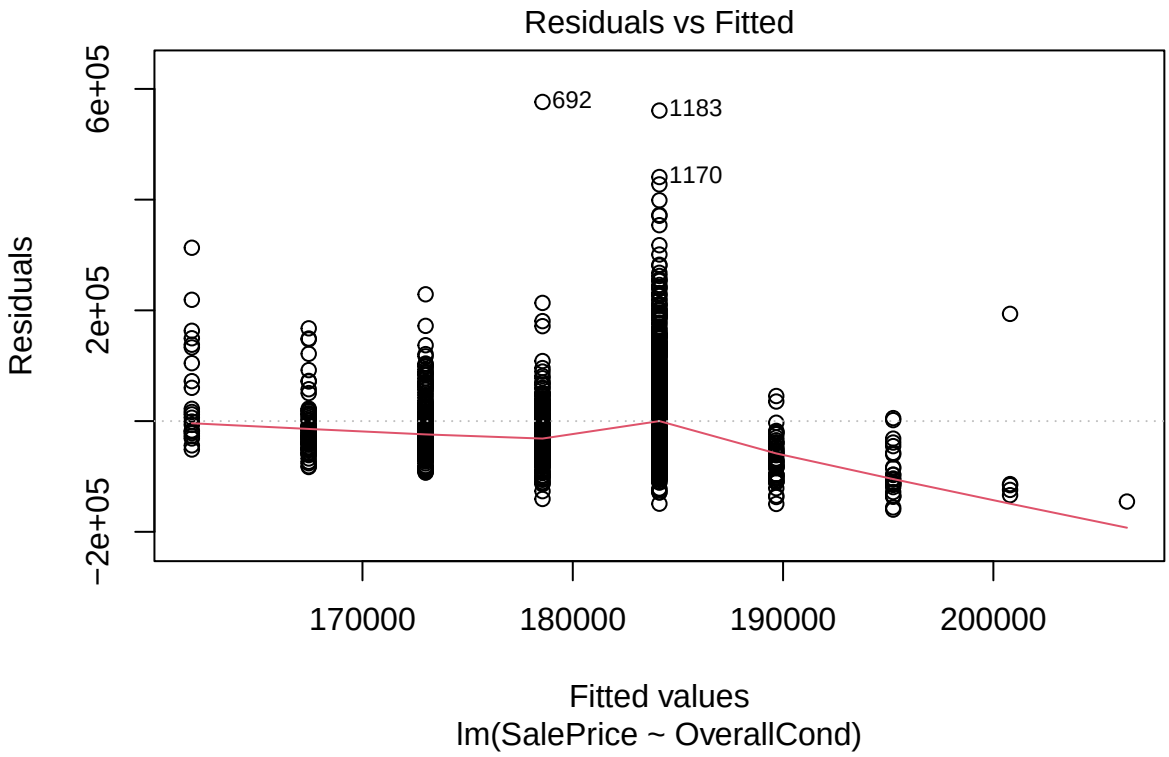

Residuals vs Leverage
Standardized residuals
8
6
4
2
0
−2
−4
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Leverage
lm(SalePrice ~ LotArea)
Cook's distance
1
0.5
0.5
1
250
336
314

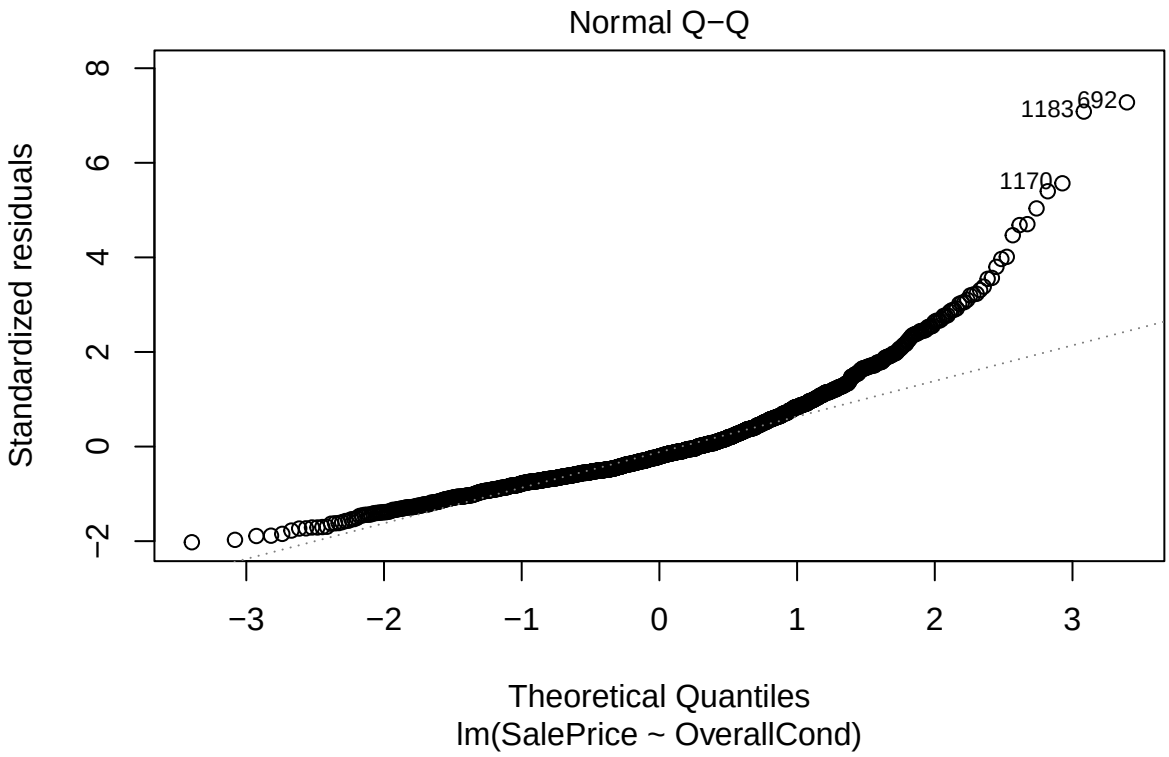

Residuals vs Fitted
Residuals
6e+05
2e+05
−2e+05
1183
1170
692
100000
150000
200000
250000
300000
350000
Fitted values
lm(SalePrice ~ ExterQual)

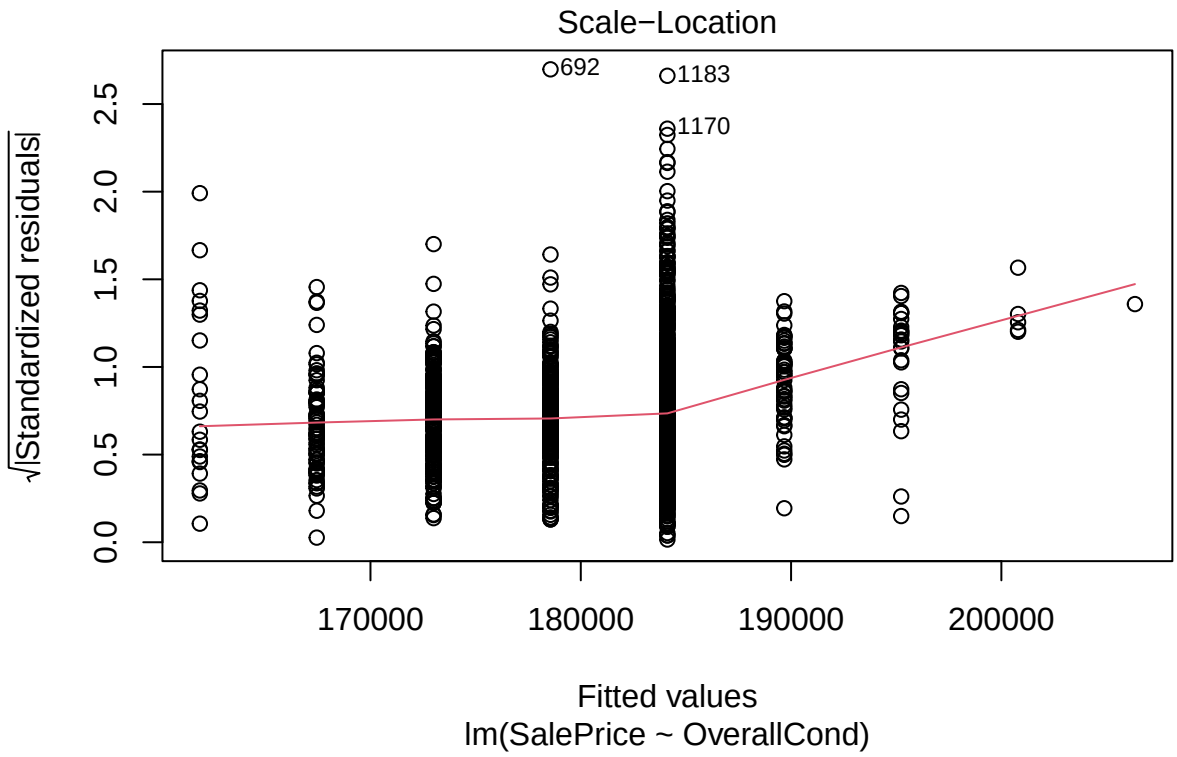

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(SalePrice ~ ExterQual)
1183
−4
−2
0
2
4
6
8
10
−3
−2
−1
0
1
2
3

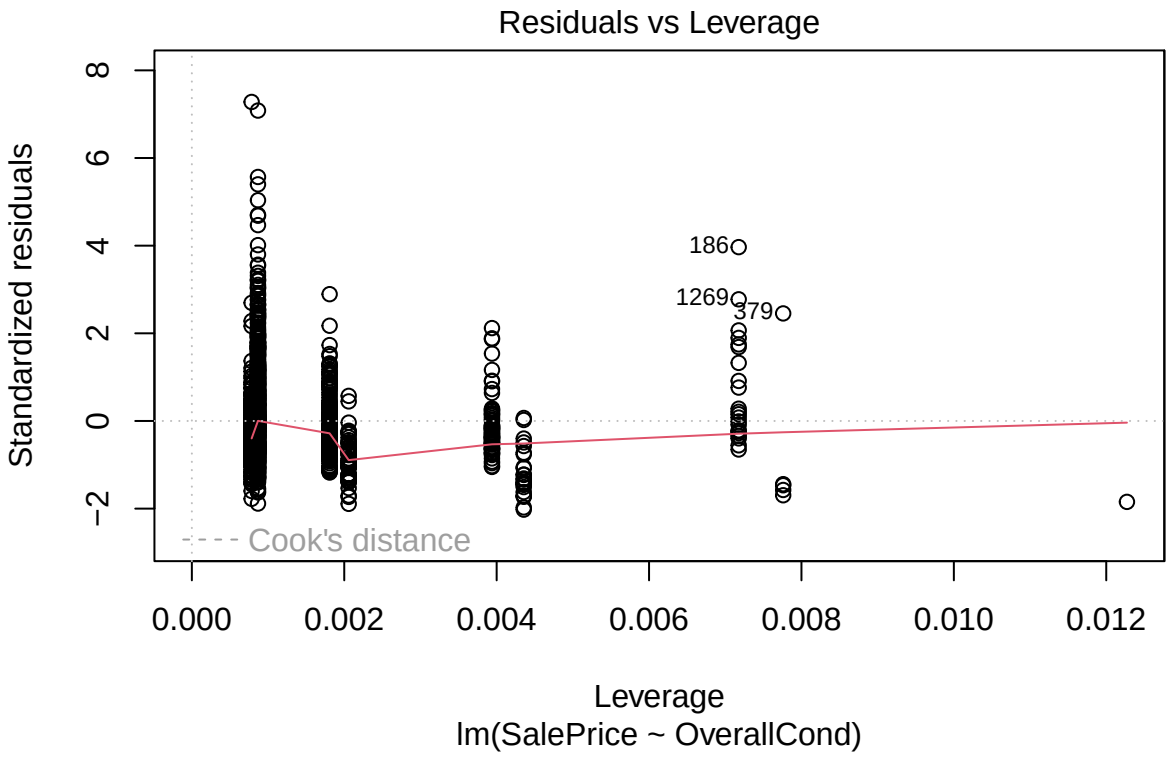

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1183
1170
692
100000
150000
200000
250000
300000
350000
Fitted values
lm(SalePrice ~ ExterQual)

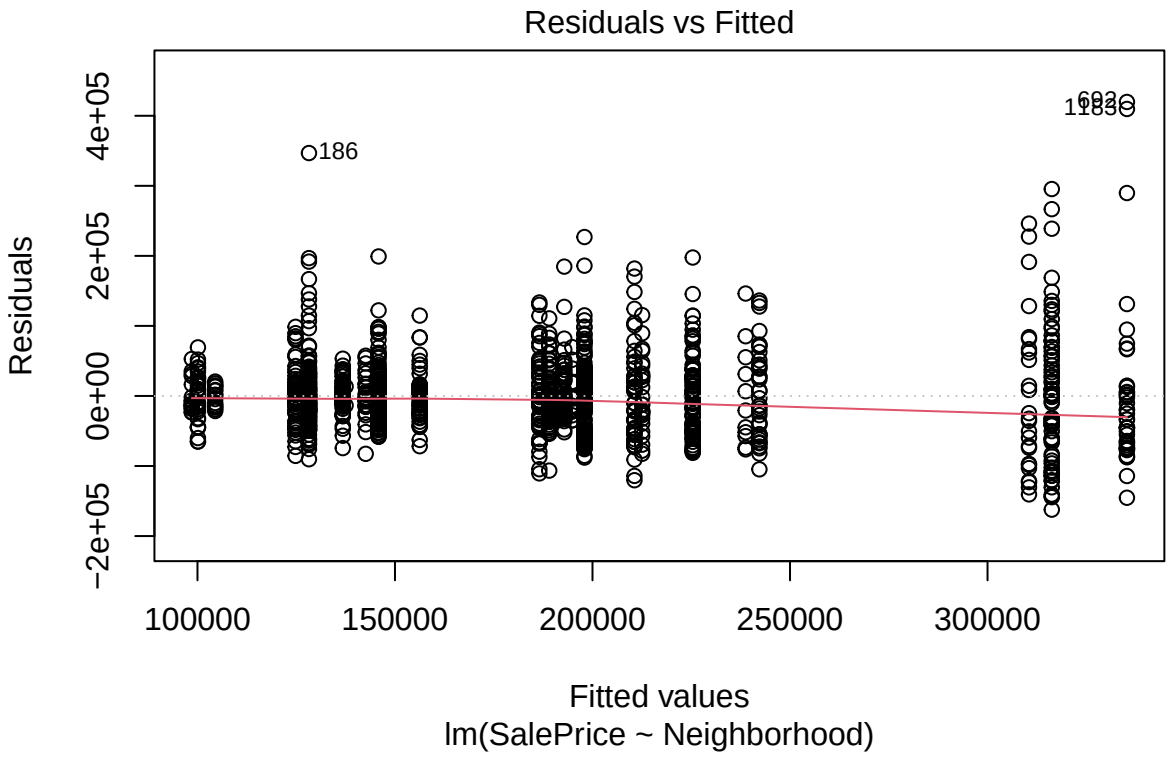

Residuals vs Leverage
Standardized residuals
10
5
0
−5
692
899
636
1
0.5
0.5
Cook's distance
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
Leverage
lm(SalePrice ~ ExterQual)

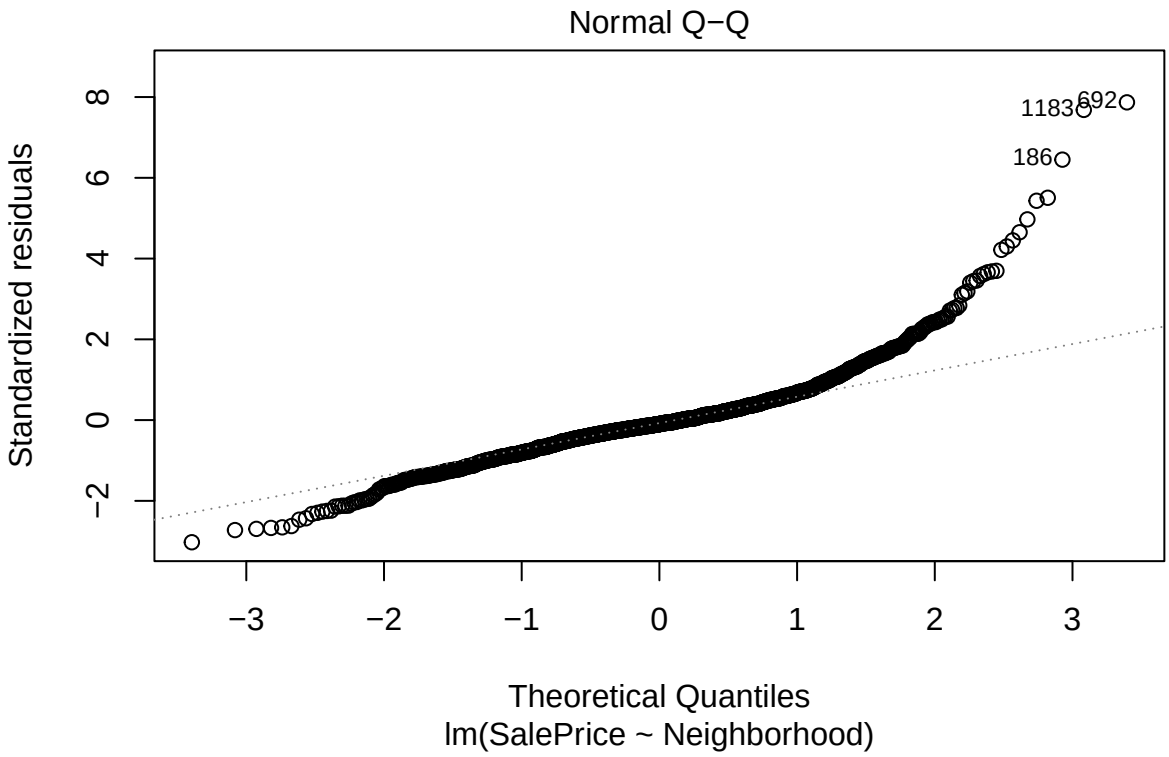

Residuals vs Fitted
Residuals
6e+05
2e+05
−2e+05
692
1183
1170
170000
180000
190000
200000
Fitted values
lm(SalePrice ~ OverallCond)

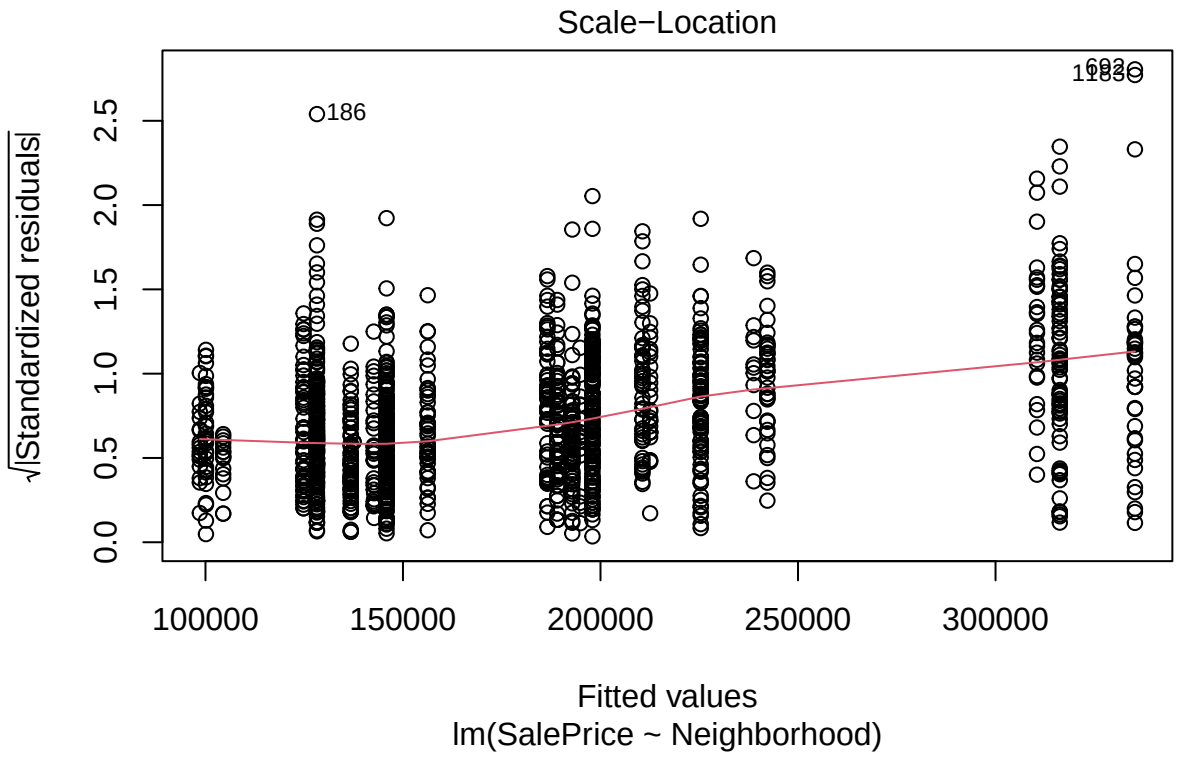

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(SalePrice ~ OverallCond)
1183
692
1170
−2
0
2
4
6
8
−3
−2
−1
0
1
2
3

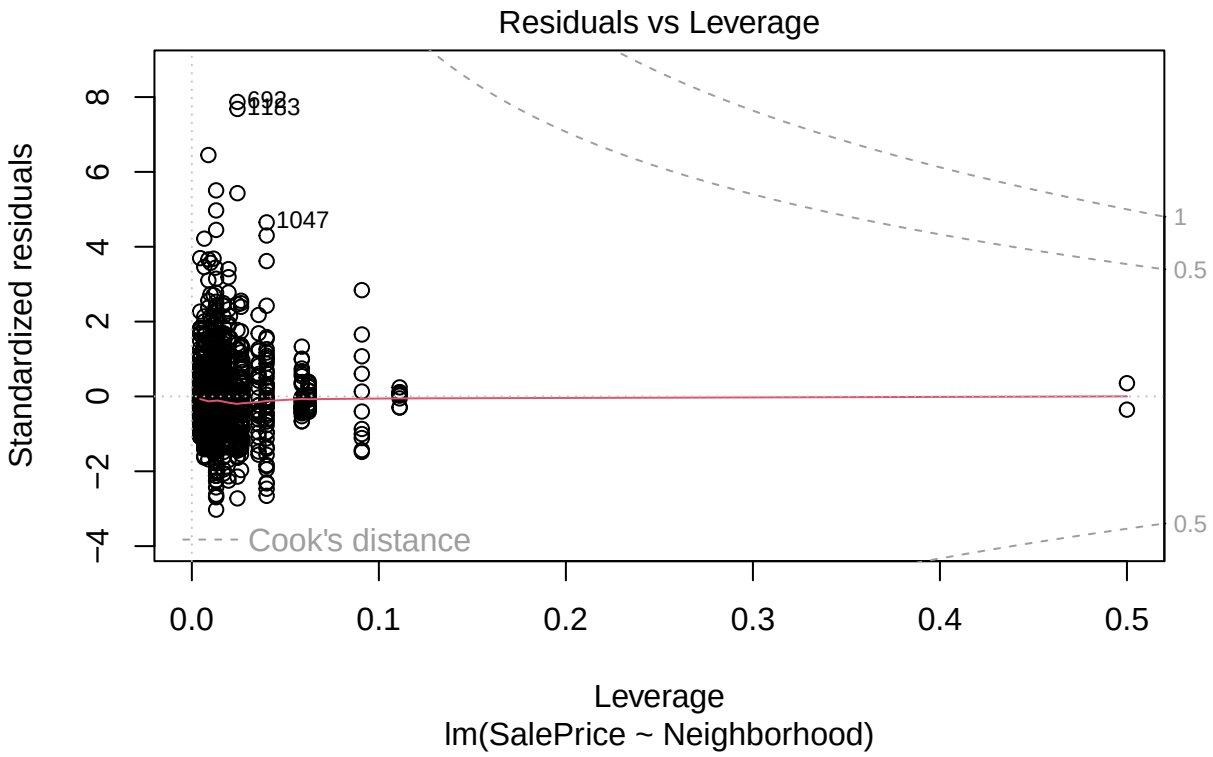

Scale−Location
692
1183
1170
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
170000
180000
190000
200000
Fitted values
lm(SalePrice ~ OverallCond)

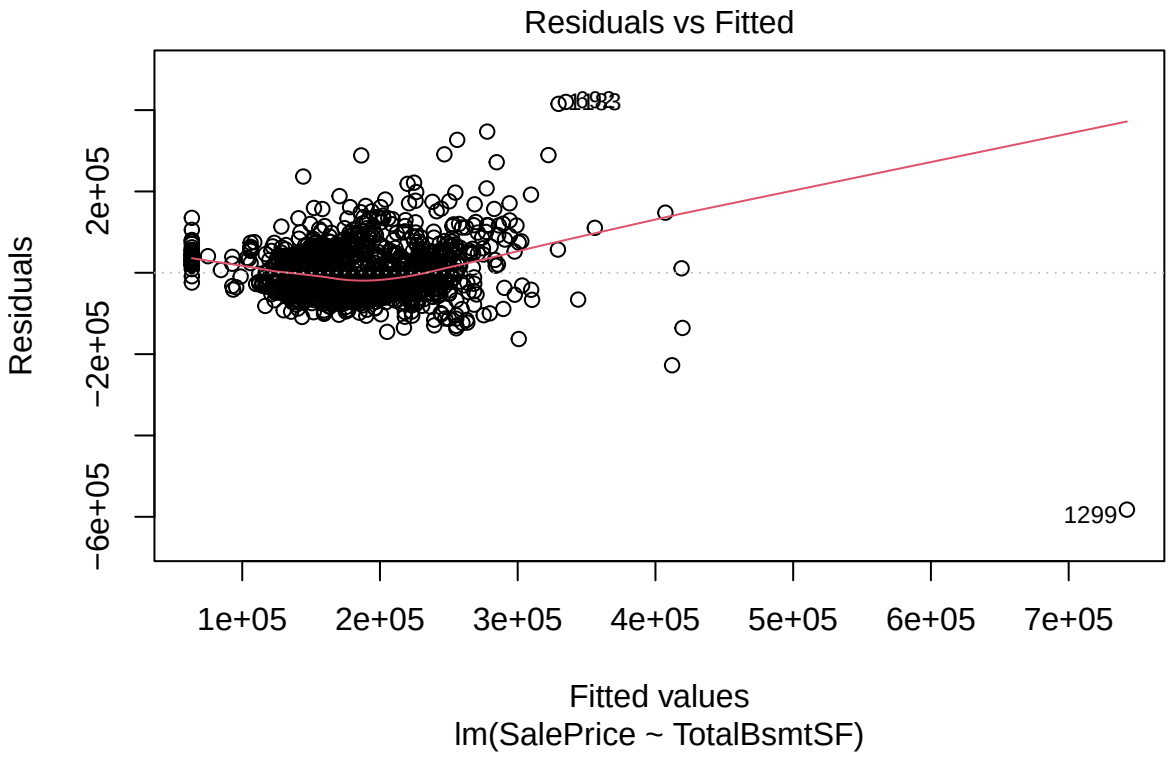

Residuals vs Leverage
Standardized residuals
8
6
4
2
0
−2
186
1269
379
Cook's distance
0.000
0.002
0.004
0.006
0.008
0.010
0.012
Leverage
lm(SalePrice ~ OverallCond)

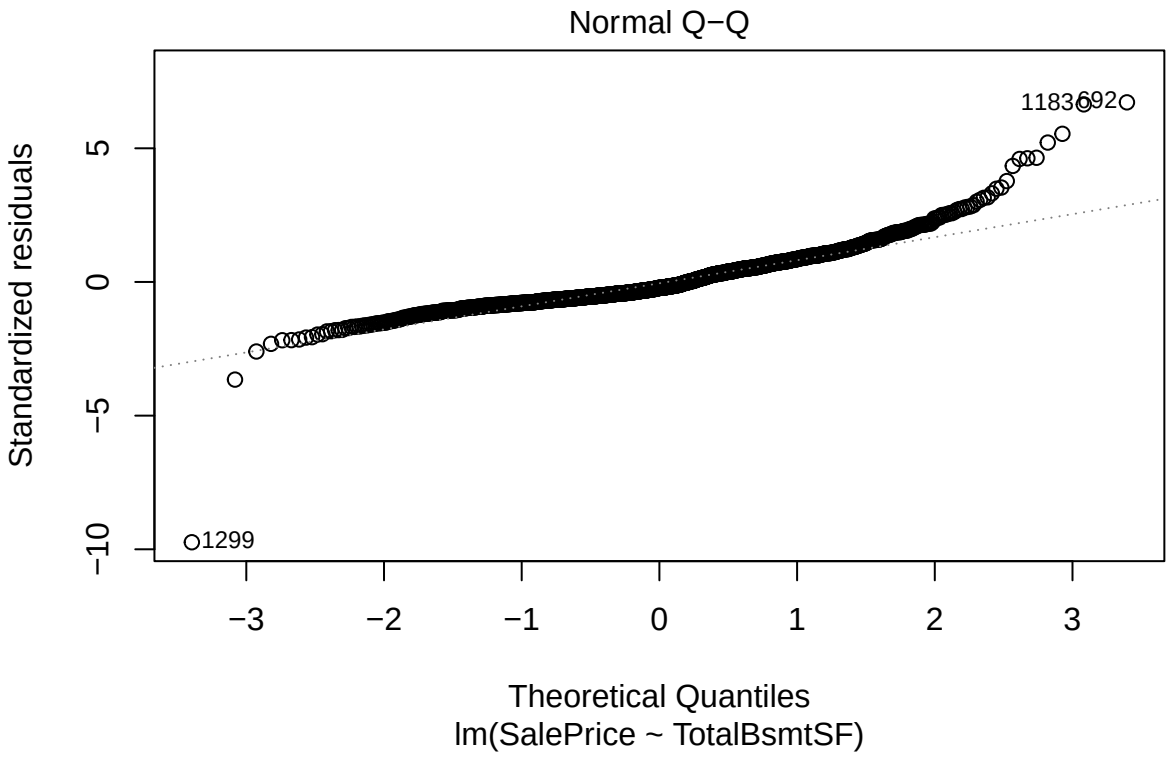

Residuals vs Fitted
Residuals
4e+05
2e+05
0e+00
−2e+05
186
100000
150000
200000
250000
300000
Fitted values
lm(SalePrice ~ Neighborhood)

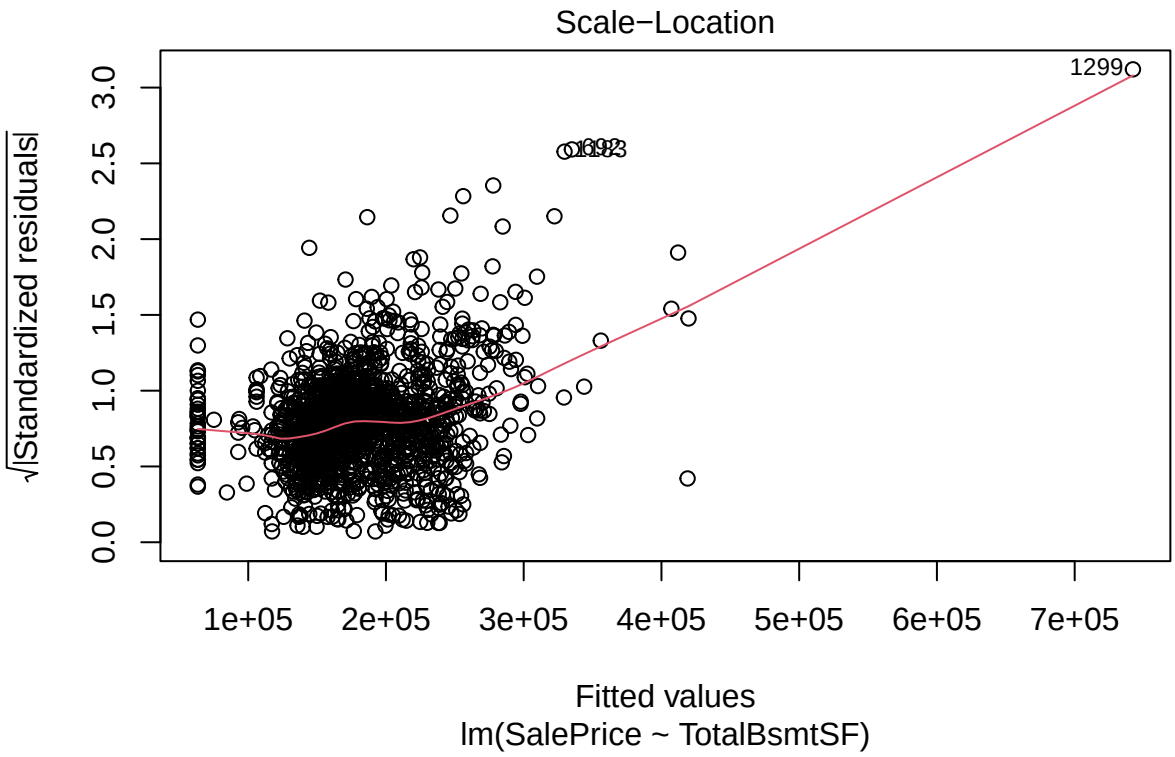

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(SalePrice ~ Neighborhood)
1183
692
186

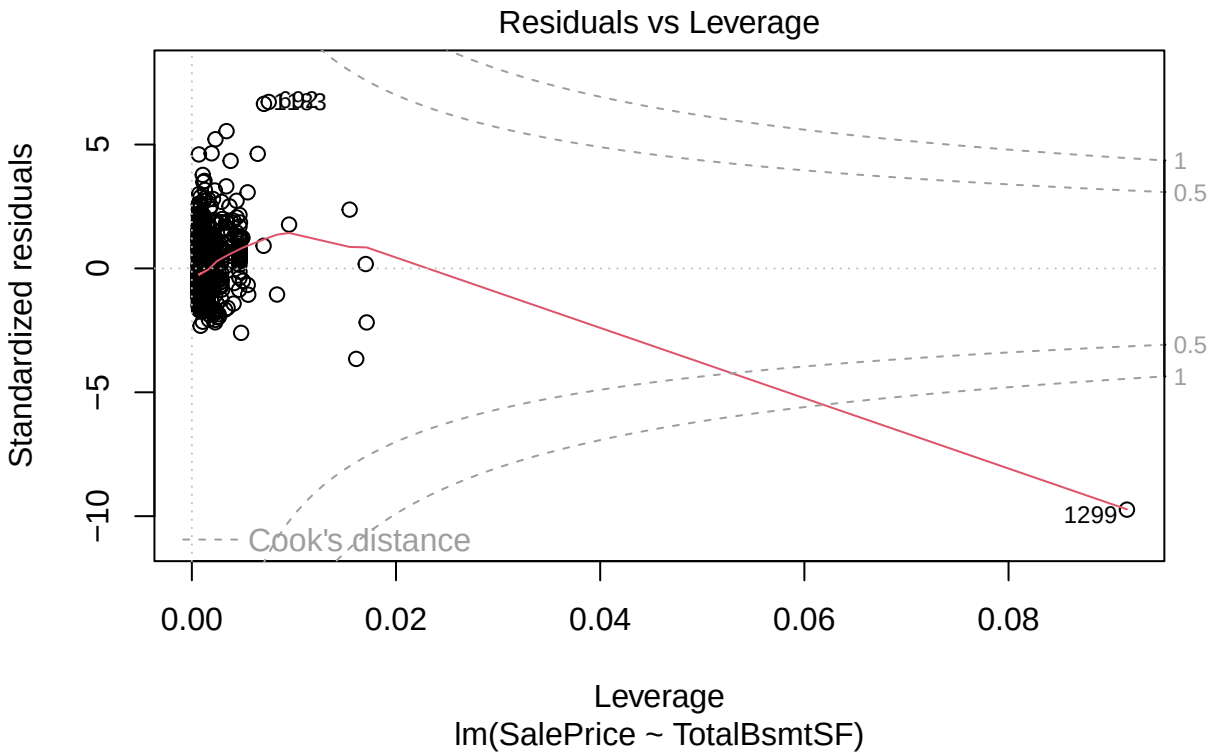

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
186
100000
150000
200000
250000
300000
Fitted values
lm(SalePrice ~ Neighborhood)


Residuals vs Leverage
Standardized residuals
8
6
4
2
0
−2
−4
692
1183
1047
1
0.5
0.5
Cook's distance
0.0
0.1
0.2
0.3
0.4
0.5
Leverage
lm(SalePrice ~ Neighborhood)


Residuals vs Fitted
Residuals
2e+05
−2e+05
−6e+05
1299
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
7e+05
Fitted values
lm(SalePrice ~ TotalBsmtSF)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(SalePrice ~ TotalBsmtSF)
1299
692
5
0
−5
−10
−3
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
1299
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
7e+05
Fitted values
lm(SalePrice ~ TotalBsmtSF)



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.