# 6.1 Housing Data

Anna Harvey

7/12/2020

**a. Explain why you chose to remove data points from your 'clean' dataset.**

n/a

**b. Create two variables; one that will contain the variables Sale Price and Square Foot of Lot (same variables used from previous assignment on simple regression) and one that will contain Sale Price and several additional predictors of your choice. Explain the basis for your additional predictor selections.**

```
sale_sqft <- lm(survey$SalePrice~survey$sq_ft_lot)
sale_vars <- lm(survey$SalePrice~survey$square_feet_total_living + survey$year_built)
```

I chose the square feet of the living space and the year the house was built as additional predictors. I chose these because the amount of living space and how old the house is are likely to determine the sale price of the house.

**c. Execute a summary() function on two variables defined in the previous step to compare the model results. What are the R2 and Adjusted R2 statistics? Explain what these results tell you about the overall model. Did the inclusion of the additional predictors help explain any large variations found in Sale Price?**

```
summary(sale_sqft)
```

```
##
## Call:
## lm(formula = survey$SalePrice ~ survey$sq_ft_lot)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2016064  -194842   -63293    91565  3735109
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.418e+05  3.800e+03  168.90   <2e-16 ***
## survey$sq_ft_lot 8.510e-01  6.217e-02   13.69   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 401500 on 12863 degrees of freedom
## Multiple R-squared:  0.01435,    Adjusted R-squared:  0.01428
## F-statistic: 187.3 on 1 and 12863 DF,  p-value: < 2.2e-16
```

```
summary(sale_vars)
```

```
##
## Call:
## lm(formula = survey$SalePrice ~ survey$square_feet_total_living +
##     survey$year_built)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1719467  -121308   -42621    44230  3916857
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -5.114e+06  3.808e+05  -13.43   <2e-16 ***
## survey$square_feet_total_living  1.714e+02  3.346e+00   51.24   <2e-16 ***
## survey$year_built                2.679e+03  1.923e+02   13.93   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 357500 on 12862 degrees of freedom
## Multiple R-squared:  0.2184, Adjusted R-squared:  0.2183
## F-statistic:  1797 on 2 and 12862 DF,  p-value: < 2.2e-16
```

Sale Price vs Sq_ft_lot R2 = 0.01435 Adjusted R2 = 0.01428

Sale Price vs Sq Ft Living vs Year Built R2 = 0.2184 Adjusted R2 = 0.2183

R2 and adjusted R2 increases with the addition of more variables. This tells me that the addition of my choice of variables increased the variation of the output value. The R2 and adjusted R2 for each variable are very close to each other. This tells me that the variables are significant and do affect the output variable.

## d. Considering the parameters of the multiple regression model you have created. What are the standardized betas for each parameter and what do the values indicate?

```
lm.beta(sale_vars)
```

```
## survey$square_feet_total_living               survey$year_built
##                       0.4196286                       0.1140856
```

Standardized beta for Sq Ft Living Space = 0.4196286 Standardized beta for Year Built = 0.1140856

This tells me that the square footage of the living space plays a more important role in determining the sale price than the year the house was built.

### e. Calculate the confidence intervals for the parameters in your model and explain what the results indicate.

```
confint(sale_vars)
```

```
##                                       2.5 %        97.5 %
## (Intercept)                     -5860366.2970 -4367673.7301
## survey$square_feet_total_living      164.8777      177.9934
## survey$year_built                   2302.1248     3056.0179
```

The variation in the confidence intervals for both the Living Space and Year Built are fairly close to each other, indicating they are likely good representations of the general trend. Living Space is very close together which follows the information from the beta values indicating that it is a more important variable than Year Built for determining the sale price.

### f. Assess the improvement of the new model compared to your original model (simple regression model) by testing whether this change is significant by performing an analysis of variance.

```
anova(sale_sqft, sale_vars)
```

```
## Analysis of Variance Table
##
## Model 1: survey$SalePrice ~ survey$sq_ft_lot
## Model 2: survey$SalePrice ~ survey$square_feet_total_living + survey$year_built
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1  12863 2.0734e+15
## 2  12862 1.6441e+15  1 4.2931e+14 3358.7 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F = 3358.7 This does seem to indicate that the addition of the variables caused a significant improvement to the model.

### g. Perform casewise diagnostics to identify outliers and/or influential cases, storing each function's output in a dataframe assigned to a unique variable name.

```
sale_varsdf <- data.frame(survey$SalePrice, survey$square_feet_total_living, survey$year_built)
sale_varsdf$residuals <- resid(sale_vars)
sale_varsdf$rstandard <- rstandard(sale_vars)
sale_varsdf$cooks <- cooks.distance(sale_vars)
sale_varsdf$covratio <- covratio(sale_vars)
sale_varsdf$leverage <- hatvalues(sale_vars)
```

**h. Calculate the standardized residuals using the appropriate command, specifying those that are +-2, storing the results of large residuals in a variable you create.**

```
sale_varsdf$large_residuals <- sale_varsdf$rstandard > 2 | sale_varsdf$rstandard < -2
large_residuals <- sale_varsdf$rstandard > 2 | sale_varsdf$rstandard < -2
```

**i. Use the appropriate function to show the sum of large residuals.**

```
sum(large_residuals)
```

```
## [1] 327
```

**j. Which specific variables have large residuals (only cases that evaluate as TRUE)?**

```
sale_varsdf[sale_varsdf$large_residuals == TRUE, c("survey.SalePrice", "survey.square_feet_total_living
```

```
##       survey.SalePrice survey.square_feet_total_living survey.year_built
## 6               184667                            4160              2005
## 25              265000                            4920              2007
## 115            1390000                             660              1955
## 178             390000                            5800              2008
## 239            1588359                            3360              2005
## 246            1450000                             900              1918
## 287             163000                            4710              2014
## 295             270000                            5060              2016
## 300             200000                            6880              2008
## 341             300000                            4490              2008
## 359             187000                            5140              2008
## 385            2500000                            6310              2005
## 396            2169000                            5080              2005
## 475            1534000                            3320              1963
## 482             555000                            6380              2007
## 508              65000                            3700              2010
## 528             435000                            5830              2014
## 661            2569000                            8090              2006
## 670            2583000                            4710              1992
## 679             350000                            8490              2008
## 784             325000                            5150              2015
## 811            3000000                            5270              2001
## 853            1595000                            3470              1985
## 877             157000                            4640              2008
## 916             450000                            7640              2007
## 1009            175000                            4040              2005
## 1119            279150                            6340              2007
## 1142            279150                            5980              2007
```

| | | | |
|---|---|---|---|
| ## 1155 | 32000 | 4740 | 2007 |
| ## 1305 | 1316000 | 2800 | 1929 |
| ## 1345 | 233333 | 3990 | 2008 |
| ## 1380 | 80000 | 3660 | 2008 |
| ## 1442 | 1085000 | 340 | 1954 |
| ## 1492 | 285000 | 4610 | 2007 |
| ## 1504 | 452800 | 7780 | 2008 |
| ## 1550 | 276450 | 5380 | 2007 |
| ## 1633 | 5000 | 3130 | 2014 |
| ## 1650 | 349999 | 9360 | 2009 |
| ## 1716 | 1000 | 4610 | 2015 |
| ## 1745 | 2235000 | 5360 | 2006 |
| ## 1870 | 2988000 | 10630 | 2003 |
| ## 1962 | 2493000 | 5300 | 2006 |
| ## 1963 | 1600000 | 3090 | 1978 |
| ## 1964 | 1600000 | 550 | 1979 |
| ## 1976 | 2625000 | 3830 | 2009 |
| ## 1977 | 2625000 | 3330 | 2009 |
| ## 1978 | 2625000 | 3480 | 2008 |
| ## 1979 | 2625000 | 3370 | 2008 |
| ## 1980 | 2625000 | 3350 | 2008 |
| ## 1981 | 2625000 | 4080 | 2008 |
| ## 1982 | 2625000 | 3690 | 2009 |
| ## 2020 | 1862000 | 5310 | 1985 |
| ## 2022 | 1384950 | 900 | 1940 |
| ## 2099 | 1975000 | 5020 | 2004 |
| ## 2137 | 1787000 | 2820 | 1972 |
| ## 2157 | 1678000 | 3130 | 1965 |
| ## 2257 | 20000 | 2940 | 2015 |
| ## 2264 | 1651000 | 1540 | 1938 |
| ## 2302 | 450000 | 6340 | 2009 |
| ## 2360 | 1640000 | 3200 | 1968 |
| ## 2361 | 1640000 | 310 | 1964 |
| ## 2469 | 1710000 | 2980 | 1980 |
| ## 2604 | 40191 | 3060 | 2008 |
| ## 2684 | 2590000 | 3410 | 2008 |
| ## 2685 | 2590000 | 3660 | 2010 |
| ## 2686 | 2590000 | 4500 | 2008 |
| ## 2687 | 2590000 | 3480 | 2010 |
| ## 2688 | 2590000 | 3310 | 2008 |
| ## 2689 | 2590000 | 4340 | 2008 |
| ## 2690 | 2590000 | 4390 | 2009 |
| ## 2699 | 379950 | 5270 | 2015 |
| ## 2708 | 2300000 | 4710 | 2015 |
| ## 2709 | 2300000 | 5100 | 2016 |
| ## 2710 | 2300000 | 4910 | 2015 |
| ## 2717 | 2598000 | 5390 | 2007 |
| ## 2742 | 10570 | 3950 | 2013 |
| ## 2852 | 3995000 | 11810 | 2000 |
| ## 2934 | 1500 | 4610 | 2015 |
| ## 2937 | 20000 | 4610 | 2015 |
| ## 3097 | 2549000 | 5320 | 2007 |
| ## 3102 | 2080000 | 4790 | 2008 |
| ## 3110 | 1250000 | 1290 | 1961 |

```
## 3111      1250000      1920      1961
## 3168      3175000      1290      2008
## 3169      3175000      1290      2008
## 3170      3175000      1600      2009
## 3171      3175000      1740      2008
## 3172      3175000      1710      2008
## 3173      3175000      1740      2008
## 3174      3175000      1710      2008
## 3175      3175000      1460      2008
## 3176      3175000      1820      2009
## 3177      3175000      1840      2008
## 3178      3175000      1840      2008
## 3179      3175000      1600      2008
## 3180      3175000      1840      2008
## 3181      3175000      1600      2008
## 3182      3175000      1290      2008
## 3183      3175000      1290      2008
## 3184      3175000      1290      2008
## 3185      3175000      1290      2008
## 3186      3175000      1290      2008
## 3187      3175000      1290      2008
## 3188      3175000      1290      2008
## 3189      3175000      1460      2008
## 3190      3175000      1710      2008
## 3191      3175000      1460      2008
## 3192      3175000      1740      2008
## 3193      3175000      1710      2008
## 3194      3175000      1710      2008
## 3195      3175000      1840      2008
## 3196      3175000      1840      2008
## 3197      3175000      1820      2009
## 3198      3175000      1290      2008
## 3199      3175000      1290      2008
## 3200      3175000      1290      2008
## 3201      3175000      1290      2008
## 3202      3175000      1290      2008
## 3260      1750000      4604      1999
## 3424        20146      4290      2008
## 3464      3150000      1460      2008
## 3465      3150000      1600      2009
## 3466      3150000      1840      2008
## 3467      3150000      1840      2008
## 3468      3150000      1460      2008
## 3469      3150000      1600      2008
## 3470      3150000      1840      2008
## 3471      3150000      1600      2009
## 3472      3150000      1740      2008
## 3473      3150000      1710      2008
## 3474      3150000      1460      2008
## 3475      3150000      1740      2008
## 3476      3150000      1830      2008
## 3477      3150000      1740      2008
## 3478      3150000      1290      2008
## 3479      3150000      1290      2008
```

| | | | |
|---|---|---|---|
| ## 3480 | 3150000 | 1290 | 2008 |
| ## 3481 | 3150000 | 1290 | 2008 |
| ## 3482 | 3150000 | 1290 | 2008 |
| ## 3483 | 3150000 | 1290 | 2008 |
| ## 3484 | 3150000 | 1290 | 2008 |
| ## 3485 | 3150000 | 1840 | 2010 |
| ## 3486 | 3150000 | 1600 | 2008 |
| ## 3487 | 3150000 | 1840 | 2009 |
| ## 3488 | 3150000 | 1710 | 2008 |
| ## 3489 | 3150000 | 1740 | 2008 |
| ## 3490 | 3150000 | 1710 | 2008 |
| ## 3491 | 3150000 | 1840 | 2008 |
| ## 3492 | 3150000 | 1600 | 2009 |
| ## 3493 | 3150000 | 1840 | 2008 |
| ## 3494 | 3150000 | 1460 | 2008 |
| ## 3495 | 3150000 | 1290 | 2008 |
| ## 3496 | 3150000 | 1290 | 2008 |
| ## 3497 | 3150000 | 1290 | 2008 |
| ## 3523 | 2000000 | 3980 | 1955 |
| ## 3837 | 130000 | 4480 | 2014 |
| ## 3918 | 500000 | 6340 | 2007 |
| ## 3919 | 475000 | 5980 | 2007 |
| ## 4055 | 2000000 | 4110 | 1991 |
| ## 4056 | 2000000 | 900 | 1937 |
| ## 4285 | 155026 | 4180 | 2003 |
| ## 4435 | 188750 | 4330 | 2011 |
| ## 4648 | 4400000 | 5790 | 1999 |
| ## 4649 | 4400000 | 2410 | 1935 |
| ## 4671 | 800000 | 7810 | 2011 |
| ## 4695 | 2300000 | 1430 | 2003 |
| ## 4696 | 2300000 | 5330 | 2005 |
| ## 4740 | 1448000 | 2900 | 1974 |
| ## 4750 | 2300000 | 7640 | 2007 |
| ## 4821 | 1430000 | 890 | 1999 |
| ## 4834 | 1300000 | 900 | 1940 |
| ## 4840 | 6000 | 4740 | 2007 |
| ## 4934 | 1780000 | 4210 | 1959 |
| ## 5083 | 698 | 5830 | 1969 |
| ## 5491 | 1550000 | 2960 | 2011 |
| ## 5494 | 1550000 | 3010 | 2011 |
| ## 5495 | 1550000 | 2960 | 2011 |
| ## 5496 | 1550000 | 2960 | 2011 |
| ## 5497 | 1550000 | 2950 | 2011 |
| ## 5549 | 241000 | 4130 | 2013 |
| ## 5935 | 180000 | 5640 | 2012 |
| ## 6055 | 130000 | 3500 | 2013 |
| ## 6230 | 2885000 | 3320 | 2012 |
| ## 6231 | 2885000 | 3530 | 2014 |
| ## 6232 | 2885000 | 3720 | 2013 |
| ## 6233 | 2885000 | 3150 | 2013 |
| ## 6234 | 2885000 | 3340 | 2012 |
| ## 6235 | 2885000 | 3320 | 2012 |
| ## 6236 | 2885000 | 3340 | 2012 |
| ## 6237 | 2885000 | 3320 | 2013 |

| | | | |
|---|---|---|---|
| ## 6238 | 2885000 | 3060 | 2013 |
| ## 6239 | 2885000 | 3530 | 2014 |
| ## 6429 | 4380542 | 3290 | 2012 |
| ## 6430 | 4380542 | 2450 | 2010 |
| ## 6431 | 4380542 | 2750 | 2012 |
| ## 6432 | 4380542 | 3010 | 2012 |
| ## 6433 | 4380542 | 3200 | 2010 |
| ## 6434 | 4380542 | 3200 | 2012 |
| ## 6435 | 4380542 | 3620 | 2012 |
| ## 6436 | 4380542 | 2810 | 2012 |
| ## 6437 | 4380542 | 2550 | 2010 |
| ## 6438 | 4380542 | 2440 | 2011 |
| ## 6439 | 4380542 | 3160 | 2011 |
| ## 6440 | 4380542 | 3400 | 2012 |
| ## 6441 | 4380542 | 2960 | 2012 |
| ## 6442 | 4380542 | 3110 | 2012 |
| ## 6443 | 4140203 | 2900 | 2012 |
| ## 6444 | 4140203 | 3220 | 2012 |
| ## 6445 | 4140203 | 2510 | 2013 |
| ## 6446 | 4140203 | 2970 | 2013 |
| ## 6447 | 4140203 | 3470 | 2012 |
| ## 6448 | 4140203 | 2580 | 2011 |
| ## 6449 | 4140203 | 3590 | 2012 |
| ## 6450 | 4140203 | 3890 | 2012 |
| ## 6451 | 4140203 | 2520 | 2012 |
| ## 6452 | 4140203 | 2990 | 2013 |
| ## 6453 | 4140203 | 3620 | 2013 |
| ## 6454 | 4140203 | 3300 | 2012 |
| ## 6455 | 4140203 | 2830 | 2012 |
| ## 6456 | 4140203 | 2680 | 2012 |
| ## 6457 | 4140203 | 3330 | 2012 |
| ## 6512 | 2500000 | 5040 | 1990 |
| ## 6527 | 275000 | 5050 | 2012 |
| ## 6739 | 50000 | 3850 | 2013 |
| ## 6796 | 226610 | 4000 | 2014 |
| ## 6821 | 20000 | 4200 | 1955 |
| ## 6931 | 149650 | 4000 | 2004 |
| ## 6938 | 2300000 | 3970 | 2013 |
| ## 6939 | 2300000 | 4610 | 2013 |
| ## 6940 | 2300000 | 3360 | 2013 |
| ## 6941 | 2300000 | 3370 | 2013 |
| ## 6942 | 2300000 | 4390 | 2013 |
| ## 6943 | 2300000 | 4380 | 2013 |
| ## 6944 | 2300000 | 4790 | 2013 |
| ## 6945 | 2300000 | 3970 | 2013 |
| ## 6946 | 2300000 | 3370 | 2013 |
| ## 6947 | 2300000 | 3260 | 2013 |
| ## 6948 | 2300000 | 3210 | 2013 |
| ## 7039 | 1920000 | 5320 | 1976 |
| ## 7147 | 254000 | 5370 | 2014 |
| ## 7167 | 3200000 | 4460 | 1975 |
| ## 7210 | 2500000 | 4460 | 1978 |
| ## 7211 | 2500000 | 3640 | 1982 |
| ## 7446 | 3462000 | 2650 | 2013 |

| | | | |
|---|---|---|---|
| ## 7447 | 3462000 | 2520 | 2013 |
| ## 7448 | 3462000 | 2655 | 2013 |
| ## 7449 | 3462000 | 2520 | 2013 |
| ## 7450 | 3462000 | 2515 | 2013 |
| ## 7451 | 3462000 | 2510 | 2013 |
| ## 7452 | 3462000 | 2520 | 2013 |
| ## 7453 | 3462000 | 2630 | 2012 |
| ## 7454 | 3462000 | 2640 | 2012 |
| ## 7455 | 3462000 | 2530 | 2012 |
| ## 7456 | 3462000 | 2520 | 2013 |
| ## 7457 | 3462000 | 2640 | 2013 |
| ## 7458 | 3462000 | 2510 | 2013 |
| ## 7459 | 3462000 | 2445 | 2013 |
| ## 7460 | 3462000 | 2515 | 2013 |
| ## 7461 | 3462000 | 2475 | 2013 |
| ## 7462 | 3462000 | 2460 | 2013 |
| ## 7463 | 3462000 | 2475 | 2013 |
| ## 7507 | 229687 | 5150 | 2015 |
| ## 7649 | 280000 | 4350 | 2014 |
| ## 7650 | 280000 | 4350 | 2014 |
| ## 7791 | 525400 | 5620 | 2015 |
| ## 7871 | 347000 | 6650 | 2015 |
| ## 8119 | 100000 | 4200 | 1955 |
| ## 8154 | 192000 | 3960 | 2014 |
| ## 8232 | 226000 | 5150 | 2015 |
| ## 8262 | 1300000 | 13540 | 1999 |
| ## 8320 | 302500 | 5820 | 2016 |
| ## 8377 | 14000 | 8750 | 1996 |
| ## 8457 | 1586000 | 3090 | 1978 |
| ## 8458 | 1586000 | 550 | 1979 |
| ## 8535 | 245000 | 4420 | 2014 |
| ## 8541 | 2500 | 3130 | 2014 |
| ## 8698 | 1700000 | 3380 | 2014 |
| ## 8710 | 2750000 | 4950 | 1996 |
| ## 8763 | 1650000 | 3460 | 1972 |
| ## 8887 | 3340000 | 6360 | 2008 |
| ## 8911 | 2160200 | 6280 | 2003 |
| ## 8946 | 2160200 | 6280 | 2003 |
| ## 9215 | 1401322 | 1650 | 1954 |
| ## 9293 | 425000 | 5830 | 2014 |
| ## 9420 | 320000 | 4800 | 2015 |
| ## 9453 | 299500 | 4610 | 2015 |
| ## 9528 | 8000 | 7110 | 2006 |
| ## 9546 | 150000 | 4480 | 2014 |
| ## 9722 | 1700000 | 2900 | 1974 |
| ## 10125 | 2200000 | 6680 | 1998 |
| ## 10318 | 2140000 | 5150 | 1997 |
| ## 10371 | 299000 | 4930 | 2016 |
| ## 10418 | 325000 | 4770 | 2003 |
| ## 10623 | 321000 | 4570 | 2016 |
| ## 10707 | 1442500 | 2630 | 1987 |
| ## 10723 | 1850000 | 4790 | 2008 |
| ## 10741 | 285000 | 4200 | 2015 |
| ## 10787 | 2200000 | 4640 | 1952 |

```
## 10844          1200000                    1800             1918
## 10958            18000                    4010             1997
## 10995          1050000                    9070             2005
## 11165          1890000                    4140             1995
## 11289          2300000                    3310             1982
## 11413          2025000                    2610             1948
## 11558          2150000                     730             1952
## 11586          1700000                    2770             1938
## 11728          1840000                    2280             1981
## 11758           269000                    4930             2016
## 11772          3750000                    6600             2001
## 11822          2850000                    4240             2005
## 11898          2165000                    3690             1999
## 11899          2165000                    1230             1999
## 11982          3175000                    7640             2007
## 11992          4311000                    1670             1964
## 12212          1200000                     820             1932
## 12255          1730000                    1650             1980
## 12256          1730000                    3260             1980
## 12392          1450000                    2140             1987
## 12472          2350000                    3150             1955
## 12487          2200000                    5000             1962
## 12577          2700000                    5030             2005
## 12582          3950000                    7070             1974
## 12643          3850000                    4690             2009
## 12686           148200                    3720             2006
## 12759          1700000                    4380             1990
## 12764          2988000                    6990             2006
## 12816          2050000                    5270             2015
```

**k. Investigate further by calculating the leverage, cooks distance, and covariance rations. Comment on all cases that are problematics.**

```
sale_varsdf[sale_varsdf$large_residuals == TRUE, c("cooks", "leverage", "covratio")]
```

```
##              cooks     leverage  covratio
## 6     0.0004664609 2.893507e-04 0.9993949
## 25    0.0009753878 5.278467e-04 0.9994688
## 115   0.0020299462 5.845638e-04 0.9983898
## 178   0.0018244399 9.228660e-04 0.9997742
## 239   0.0002198987 1.479589e-04 0.9993415
## 246   0.0066006892 1.561306e-03 0.9988414
## 287   0.0010683643 4.772308e-04 0.9991448
## 295   0.0012422792 6.082382e-04 0.9994132
## 300   0.0064548391 1.591739e-03 0.9989924
## 341   0.0005442432 3.857083e-04 0.9996320
## 359   0.0014782483 6.146711e-04 0.9991662
## 385   0.0043135187 1.224730e-03 0.9989968
## 396   0.0016694478 5.905161e-04 0.9988469
## 475   0.0008135770 4.637368e-04 0.9994699
## 482   0.0021170740 1.260140e-03 1.0003199
## 508   0.0004022547 2.183994e-04 0.9991632
```

```
## 528   0.0017586547 9.402696e-04 0.9998662
## 661   0.0057964869 2.601807e-03 1.0012840
## 670   0.0031300034 4.972622e-04 0.9963319
## 679   0.0146362888 2.968097e-03 0.9997635
## 784   0.0011803649 6.373256e-04 0.9995755
## 811   0.0060910722 6.817328e-04 0.9946753
## 853   0.0003225509 1.950450e-04 0.9992714
## 877   0.0009214220 4.319784e-04 0.9991734
## 916   0.0072367610 2.192226e-03 1.0001222
## 1009  0.0004091340 2.609966e-04 0.9993977
## 1119  0.0037010280 1.235103e-03 0.9993744
## 1142  0.0027194222 1.022383e-03 0.9993965
## 1155  0.0013170704 4.634135e-04 0.9987091
## 1305  0.0021113566 1.320236e-03 1.0004367
## 1345  0.0003479257 2.599336e-04 0.9995568
## 1380  0.0003464408 2.009067e-04 0.9992279
## 1442  0.0014502269 6.768441e-04 0.9994119
## 1492  0.0006475096 4.204073e-04 0.9995764
## 1504  0.0079385980 2.305177e-03 1.0001359
## 1550  0.0015491030 7.182887e-04 0.9994436
## 1633  0.0003408684 1.975727e-04 0.9992239
## 1650  0.0237019887 3.888704e-03 0.9998782
## 1716  0.0013617758 4.526234e-04 0.9985820
## 1745  0.0020676432 7.101284e-04 0.9989076
## 1870  0.0122987518 5.588886e-03 1.0043155
## 1962  0.0031283954 6.832116e-04 0.9977157
## 1963  0.0003982272 1.948616e-04 0.9989986
## 1964  0.0017909124 3.951219e-04 0.9974597
## 1976  0.0017640939 2.339192e-04 0.9951983
## 1977  0.0013934335 1.674806e-04 0.9945890
## 1978  0.0014328789 1.767488e-04 0.9947468
## 1979  0.0013645416 1.647790e-04 0.9946134
## 1980  0.0013535941 1.628304e-04 0.9945893
## 1981  0.0020084343 2.793419e-04 0.9954890
## 1982  0.0016356662 2.109031e-04 0.9950260
## 2020  0.0012221030 8.367764e-04 1.0000491
## 2022  0.0029495101 8.578213e-04 0.9986875
## 2099  0.0010943702 5.673553e-04 0.9994517
## 2137  0.0007736426 2.304123e-04 0.9981156
## 2157  0.0009897771 3.861360e-04 0.9988267
## 2257  0.0003150914 2.045951e-04 0.9993605
## 2264  0.0038980620 8.708916e-04 0.9979757
## 2302  0.0026322104 1.229012e-03 0.9999658
## 2360  0.0007899378 3.477326e-04 0.9989921
## 2361  0.0030113026 5.568932e-04 0.9970108
## 2469  0.0004280295 1.613032e-04 0.9985386
## 2604  0.0002089868 1.424471e-04 0.9993493
## 2684  0.0013336264 1.688865e-04 0.9948856
## 2685  0.0015864972 2.125071e-04 0.9952298
## 2686  0.0024463729 3.886704e-04 0.9962233
## 2687  0.0014664406 1.894597e-04 0.9950153
## 2688  0.0012815502 1.591432e-04 0.9947669
## 2689  0.0022381170 3.433787e-04 0.9960219
## 2690  0.0023152329 3.603098e-04 0.9961032
```

```
## 2699  0.0011678370 6.853058e-04 0.9997270
## 2708  0.0018403998 4.829746e-04 0.9980520
## 2709  0.0021098330 6.231975e-04 0.9984894
## 2710  0.0019749571 5.489309e-04 0.9982666
## 2717  0.0037587786 7.228403e-04 0.9973217
## 2742  0.0006546637 2.806310e-04 0.9988822
## 2852  0.0585035861 7.417413e-03 1.0021976
## 2934  0.0013605077 4.526234e-04 0.9985839
## 2937  0.0013140035 4.526234e-04 0.9986557
## 3097  0.0034087423 6.913469e-04 0.9974783
## 3102  0.0012417631 4.821890e-04 0.9989145
## 3110  0.0007994239 3.874035e-04 0.9991773
## 3111  0.0005518095 3.464887e-04 0.9994657
## 3168  0.0063615130 3.373224e-04 0.9874295
## 3169  0.0063615130 3.373224e-04 0.9874295
## 3170  0.0049800930 2.754245e-04 0.9879069
## 3171  0.0041938513 2.357635e-04 0.9880707
## 3172  0.0043116463 2.414307e-04 0.9880275
## 3173  0.0041938513 2.357635e-04 0.9880707
## 3174  0.0043116463 2.414307e-04 0.9880275
## 3175  0.0054390063 2.947875e-04 0.9876701
## 3176  0.0040859024 2.326081e-04 0.9882221
## 3177  0.0038271713 2.180110e-04 0.9882149
## 3178  0.0038271713 2.180110e-04 0.9882149
## 3179  0.0047751411 2.635592e-04 0.9878698
## 3180  0.0038271713 2.180110e-04 0.9882149
## 3181  0.0047751411 2.635592e-04 0.9878698
## 3182  0.0063615130 3.373224e-04 0.9874295
## 3183  0.0063615130 3.373224e-04 0.9874295
## 3184  0.0063615130 3.373224e-04 0.9874295
## 3185  0.0063615130 3.373224e-04 0.9874295
## 3186  0.0063615130 3.373224e-04 0.9874295
## 3187  0.0063615130 3.373224e-04 0.9874295
## 3188  0.0063615130 3.373224e-04 0.9874295
## 3189  0.0054390063 2.947875e-04 0.9876701
## 3190  0.0043116463 2.414307e-04 0.9880275
## 3191  0.0054390063 2.947875e-04 0.9876701
## 3192  0.0041938513 2.357635e-04 0.9880707
## 3193  0.0043116463 2.414307e-04 0.9880275
## 3194  0.0043116463 2.414307e-04 0.9880275
## 3195  0.0038271713 2.180110e-04 0.9882149
## 3196  0.0038271713 2.180110e-04 0.9882149
## 3197  0.0040859024 2.326081e-04 0.9882221
## 3198  0.0063615130 3.373224e-04 0.9874295
## 3199  0.0063615130 3.373224e-04 0.9874295
## 3200  0.0063615130 3.373224e-04 0.9874295
## 3201  0.0063615130 3.373224e-04 0.9874295
## 3202  0.0063615130 3.373224e-04 0.9874295
## 3260  0.0005713790 4.231622e-04 0.9997120
## 3424  0.0008288589 3.301445e-04 0.9988075
## 3464  0.0053372173 2.947875e-04 0.9879098
## 3465  0.0048859524 2.754245e-04 0.9881441
## 3466  0.0037537572 2.180110e-04 0.9884487
## 3467  0.0037537572 2.180110e-04 0.9884487
```

```
## 3468  0.0053372173 2.947875e-04 0.9879098
## 3469  0.0046849661 2.635592e-04 0.9881072
## 3470  0.0037537572 2.180110e-04 0.9884487
## 3471  0.0048859524 2.754245e-04 0.9881441
## 3472  0.0041139291 2.357635e-04 0.9883060
## 3473  0.0042296400 2.414307e-04 0.9882633
## 3474  0.0053372173 2.947875e-04 0.9879098
## 3475  0.0041139291 2.357635e-04 0.9883060
## 3476  0.0037880303 2.197074e-04 0.9884344
## 3477  0.0041139291 2.357635e-04 0.9883060
## 3478  0.0062437444 3.373224e-04 0.9876717
## 3479  0.0062437444 3.373224e-04 0.9876717
## 3480  0.0062437444 3.373224e-04 0.9876717
## 3481  0.0062437444 3.373224e-04 0.9876717
## 3482  0.0062437444 3.373224e-04 0.9876717
## 3483  0.0062437444 3.373224e-04 0.9876717
## 3484  0.0062437444 3.373224e-04 0.9876717
## 3485  0.0041297344 2.408396e-04 0.9885208
## 3486  0.0046849661 2.635592e-04 0.9881072
## 3487  0.0039371732 2.291360e-04 0.9884845
## 3488  0.0042296400 2.414307e-04 0.9882633
## 3489  0.0041139291 2.357635e-04 0.9883060
## 3490  0.0042296400 2.414307e-04 0.9882633
## 3491  0.0037537572 2.180110e-04 0.9884487
## 3492  0.0048859524 2.754245e-04 0.9881441
## 3493  0.0037537572 2.180110e-04 0.9884487
## 3494  0.0053372173 2.947875e-04 0.9879098
## 3495  0.0062437444 3.373224e-04 0.9876717
## 3496  0.0062437444 3.373224e-04 0.9876717
## 3497  0.0062437444 3.373224e-04 0.9876717
## 3523  0.0031516419 8.461259e-04 0.9984756
## 3837  0.0009035567 4.093318e-04 0.9990987
## 3918  0.0023316105 1.235103e-03 1.0001495
## 3919  0.0017661073 1.022383e-03 1.0000484
## 4055  0.0009190774 3.045701e-04 0.9984275
## 4056  0.0076748548 9.372600e-04 0.9954514
## 4285  0.0005040196 2.917156e-04 0.9993164
## 4435  0.0006298517 3.527651e-04 0.9993371
## 4648  0.0249625344 9.532013e-04 0.9829702
## 4649  0.0412679063 1.029374e-03 0.9734661
## 4671  0.0039974176 2.311265e-03 1.0013404
## 4695  0.0021082688 2.486567e-04 0.9945596
## 4696  0.0023221566 6.979676e-04 0.9986051
## 4740  0.0003376386 2.147132e-04 0.9993479
## 4750  0.0030377724 2.192226e-03 1.0014613
## 4821  0.0009996267 3.569086e-04 0.9986314
## 4834  0.0025288899 8.578213e-04 0.9990301
## 4840  0.0013835219 4.634135e-04 0.9986089
## 4934  0.0018549558 8.318031e-04 0.9995060
## 5083  0.0050526138 1.436153e-03 0.9992115
## 5491  0.0002523170 1.635798e-04 0.9993177
## 5494  0.0002484208 1.647050e-04 0.9993428
## 5495  0.0002523170 1.635798e-04 0.9993177
## 5496  0.0002523170 1.635798e-04 0.9993177
```

```
## 5497  0.0002531760 1.634073e-04 0.9993128
## 5549  0.0004600685 3.168289e-04 0.9995343
## 5935  0.0024867718 8.422314e-04 0.9990113
## 6055  0.0003145941 2.149613e-04 0.9994244
## 6230  0.0020591500 1.897492e-04 0.9928470
## 6231  0.0023651015 2.270410e-04 0.9931872
## 6232  0.0024525684 2.426356e-04 0.9934181
## 6233  0.0020979429 1.884034e-04 0.9926483
## 6234  0.0020695093 1.913460e-04 0.9928740
## 6235  0.0020591500 1.897492e-04 0.9928470
## 6236  0.0020695093 1.913460e-04 0.9928740
## 6237  0.0021498317 1.986234e-04 0.9928756
## 6238  0.0020913876 1.850418e-04 0.9925288
## 6239  0.0023651015 2.270410e-04 0.9931872
## 6429  0.0061300865 1.874854e-04 0.9777108
## 6430  0.0059202411 1.667110e-04 0.9757505
## 6431  0.0059797139 1.736889e-04 0.9765041
## 6432  0.0058428794 1.739568e-04 0.9770827
## 6433  0.0054544428 1.648867e-04 0.9774238
## 6434  0.0059907580 1.816395e-04 0.9775083
## 6435  0.0069990643 2.210568e-04 0.9784576
## 6436  0.0059120325 1.726999e-04 0.9766372
## 6437  0.0056564746 1.607759e-04 0.9759715
## 6438  0.0063120189 1.778323e-04 0.9757739
## 6439  0.0056603235 1.707076e-04 0.9773758
## 6440  0.0063585188 1.965568e-04 0.9779589
## 6441  0.0058380478 1.729858e-04 0.9769711
## 6442  0.0058955217 1.772123e-04 0.9773064
## 6443  0.0050977569 1.723986e-04 0.9798524
## 6444  0.0052311711 1.828160e-04 0.9805223
## 6445  0.0059981754 1.953223e-04 0.9790890
## 6446  0.0053671634 1.830989e-04 0.9800403
## 6447  0.0056716563 2.034328e-04 0.9810504
## 6448  0.0051797909 1.693365e-04 0.9791460
## 6449  0.0059801345 2.172168e-04 0.9813053
## 6450  0.0070060323 2.627101e-04 0.9819468
## 6451  0.0056326441 1.833206e-04 0.9790655
## 6452  0.0053652371 1.834080e-04 0.9800820
## 6453  0.0062745117 2.290056e-04 0.9814082
## 6454  0.0053414175 1.882225e-04 0.9806909
## 6455  0.0051374917 1.725104e-04 0.9797068
## 6456  0.0053106364 1.756395e-04 0.9793957
## 6457  0.0053903273 1.905389e-04 0.9807542
## 6512  0.0034210414 6.510037e-04 0.9972114
## 6527  0.0011519050 5.869196e-04 0.9994477
## 6739  0.0005422379 2.629729e-04 0.9990539
## 6796  0.0004259225 2.974720e-04 0.9995291
## 6821  0.0016510477 9.316549e-04 0.9999262
## 6931  0.0004079721 2.499580e-04 0.9993414
## 6938  0.0013332318 2.843728e-04 0.9972400
## 6939  0.0017438684 4.410976e-04 0.9979102
## 6940  0.0010991039 2.017637e-04 0.9966273
## 6941  0.0011010038 2.025925e-04 0.9966371
## 6942  0.0015919676 3.791323e-04 0.9976767
```

```
## 6943  0.0015852546 3.765171e-04 0.9976661
## 6944  0.0018719324 4.981012e-04 0.9981037
## 6945  0.0013332318 2.843728e-04 0.9972400
## 6946  0.0011010038 2.025925e-04 0.9966371
## 6947  0.0010844651 1.944384e-04 0.9965292
## 6948  0.0010802487 1.914325e-04 0.9964805
## 7039  0.0017636293 9.841865e-04 0.9999650
## 7147  0.0016989320 7.235079e-04 0.9993149
## 7167  0.0080037666 6.011161e-04 0.9915447
## 7210  0.0034800889 5.546969e-04 0.9964038
## 7211  0.0018864790 2.561550e-04 0.9953439
## 7446  0.0036456431 1.876792e-04 0.9868871
## 7447  0.0038433663 1.946625e-04 0.9866729
## 7448  0.0036392849 1.874697e-04 0.9868953
## 7449  0.0038433663 1.946625e-04 0.9866729
## 7450  0.0038522383 1.949902e-04 0.9866647
## 7451  0.0038612057 1.953223e-04 0.9866565
## 7452  0.0038433663 1.946625e-04 0.9866729
## 7453  0.0034644328 1.775582e-04 0.9868169
## 7454  0.0034519273 1.771395e-04 0.9868334
## 7455  0.0036098063 1.827092e-04 0.9866517
## 7456  0.0038433663 1.946625e-04 0.9866729
## 7457  0.0036586326 1.881113e-04 0.9868706
## 7458  0.0038612057 1.953223e-04 0.9866565
## 7459  0.0039865399 2.000377e-04 0.9865499
## 7460  0.0038522383 1.949902e-04 0.9866647
## 7461  0.0039266629 1.977694e-04 0.9865991
## 7462  0.0039561638 1.988839e-04 0.9865745
## 7463  0.0039266629 1.977694e-04 0.9865991
## 7507  0.0014626494 6.373256e-04 0.9992658
## 7649  0.0005467171 3.750522e-04 0.9995887
## 7650  0.0005467171 3.750522e-04 0.9995887
## 7791  0.0011445250 8.396526e-04 1.0001201
## 7871  0.0043052636 1.418340e-03 0.9995310
## 8119  0.0013458758 9.316549e-04 1.0001553
## 8154  0.0004468693 2.899718e-04 0.9994452
## 8232  0.0014741764 6.373256e-04 0.9992531
## 8262  0.0445055400 1.048109e-02 1.0078588
## 8320  0.0023704331 9.403814e-04 0.9994116
## 8377  0.0263960969 3.400356e-03 0.9982227
## 8457  0.0003857287 1.948616e-04 0.9990434
## 8458  0.0017530634 3.951219e-04 0.9975266
## 8535  0.0006474578 3.931427e-04 0.9994745
## 8541  0.0003429674 1.975727e-04 0.9992165
## 8698  0.0003905479 2.127067e-04 0.9991616
## 8710  0.0041172837 5.668439e-04 0.9957255
## 8763  0.0006975138 3.391819e-04 0.9991340
## 8887  0.0128047407 1.244190e-03 0.9942935
## 8911  0.0021982138 1.216459e-03 1.0001873
## 8946  0.0021982138 1.216459e-03 1.0001873
## 9215  0.0012471199 4.801058e-04 0.9988968
## 9293  0.0018004712 9.402696e-04 0.9998351
## 9420  0.0008278908 5.117880e-04 0.9996137
## 9453  0.0007098136 4.526234e-04 0.9995890
```

```
## 9528  0.0100392020 1.772567e-03 0.9980507
## 9546  0.0008646845 4.093318e-04 0.9991651
## 9722  0.0005923976 2.147132e-04 0.9985183
## 10125 0.0026517491 1.522336e-03 1.0005396
## 10318 0.0017609131 6.468996e-04 0.9989768
## 10371 0.0010176808 5.615556e-04 0.9995274
## 10418 0.0006858285 4.735117e-04 0.9996937
## 10623 0.0006563793 4.477275e-04 0.9996556
## 10707 0.0001445453 9.054752e-05 0.9992070
## 10723 0.0007332636 4.821890e-04 0.9996519
## 10741 0.0004678762 3.464935e-04 0.9996352
## 10787 0.0052823602 1.216153e-03 0.9984143
## 10844 0.0031108380 1.582041e-03 1.0004426
## 10958 0.0005425057 2.535700e-04 0.9989903
## 10995 0.0055155406 3.612158e-03 1.0027911
## 11165 0.0006900799 2.933313e-04 0.9988811
## 11289 0.0011758876 1.908893e-04 0.9961190
## 11413 0.0038167484 6.738844e-04 0.9969489
## 11558 0.0059203676 6.219845e-04 0.9942086
## 11586 0.0034271108 9.967534e-04 0.9988267
## 11728 0.0004760078 1.156990e-04 0.9974723
## 11758 0.0010922881 5.615556e-04 0.9994345
## 11772 0.0211815375 1.439821e-03 0.9914144
## 11822 0.0028122280 3.096546e-04 0.9942002
## 11898 0.0007945653 1.827581e-04 0.9973763
## 11899 0.0020091839 2.625214e-04 0.9951489
## 11982 0.0147427616 2.192226e-03 0.9977314
## 11992 0.0121394901 3.095031e-04 0.9733415
## 12212 0.0028335562 1.089707e-03 0.9995056
## 12255 0.0006600813 1.602514e-04 0.9975135
## 12256 0.0005042011 2.009995e-04 0.9986797
## 12392 0.0001886739 9.473346e-05 0.9989348
## 12472 0.0044534605 5.997623e-04 0.9956455
## 12487 0.0042234880 1.121244e-03 0.9987220
## 12577 0.0037179263 5.703393e-04 0.9962485
## 12582 0.0386471549 2.245096e-03 0.9904841
## 12643 0.0090749332 4.506111e-04 0.9866549
## 12686 0.0002951996 2.013046e-04 0.9994087
## 12759 0.0005506545 3.940090e-04 0.9996497
## 12764 0.0102988857 1.682586e-03 0.9976412
## 12816 0.0013304277 6.853058e-04 0.9995611
```

There are no problematic Cook's values since there are none greater than 1. The leverage values all lie between 0 and 1 so there are no problematic indicators there either. The covariance ratios also seem to be more or less within the approximate boundaries (between 1.0007 and 0.9993), however, there are a few that fall around the 0.98 range.

## l. Perform the necessary calculations to assess the assumption of independence and state if the condition is met or not.

```
durbinWatsonTest(sale_vars)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.7195885     0.5608136       0
##  Alternative hypothesis: rho != 0
```

The p-value is 0 which suggests the variables do significantly impact the outcome.

### m. Perform the necessary calculations to assess the assumption of no multi-collinearity and state if the condition is met or not.

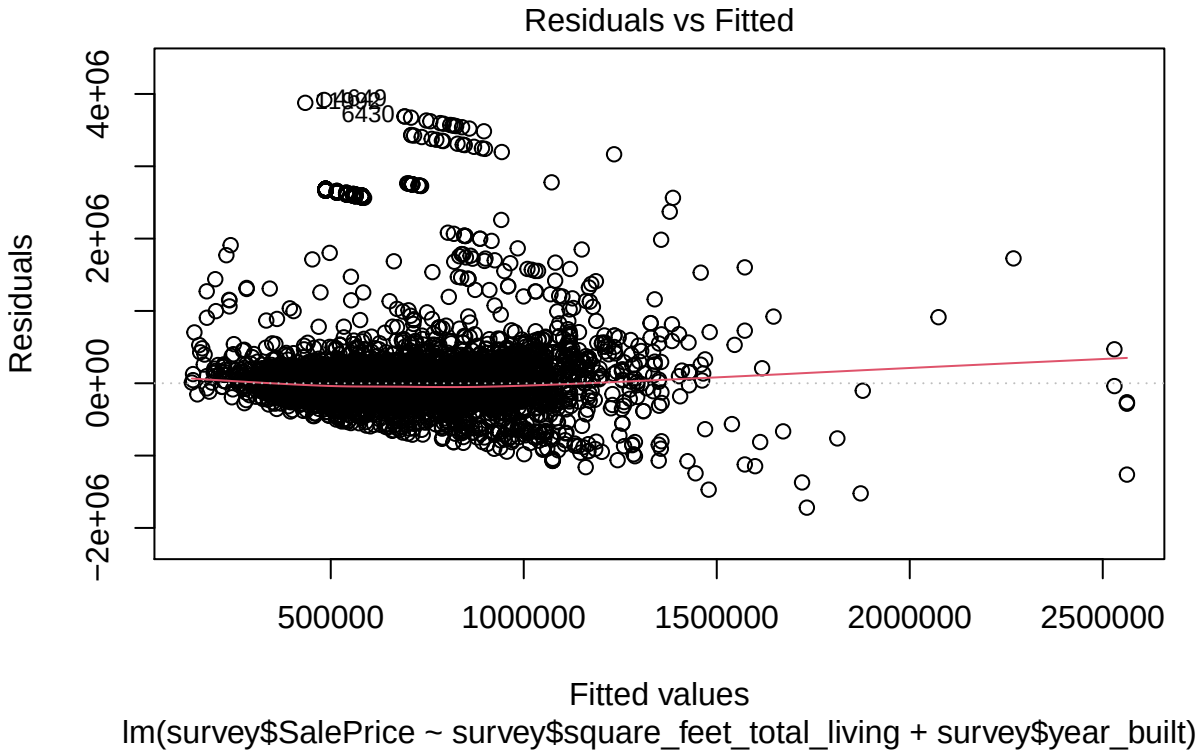
```
vif(sale_vars)
```

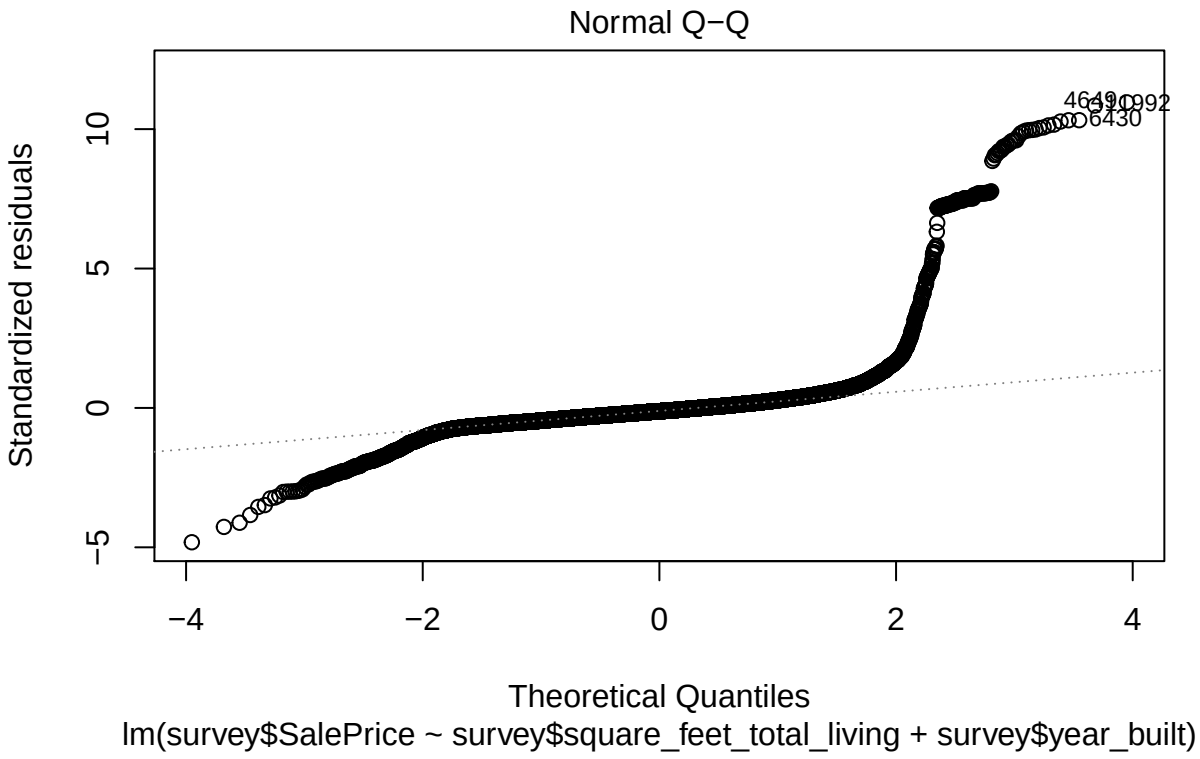
```
## survey$square_feet_total_living               survey$year_built
##                        1.103628                        1.103628
```

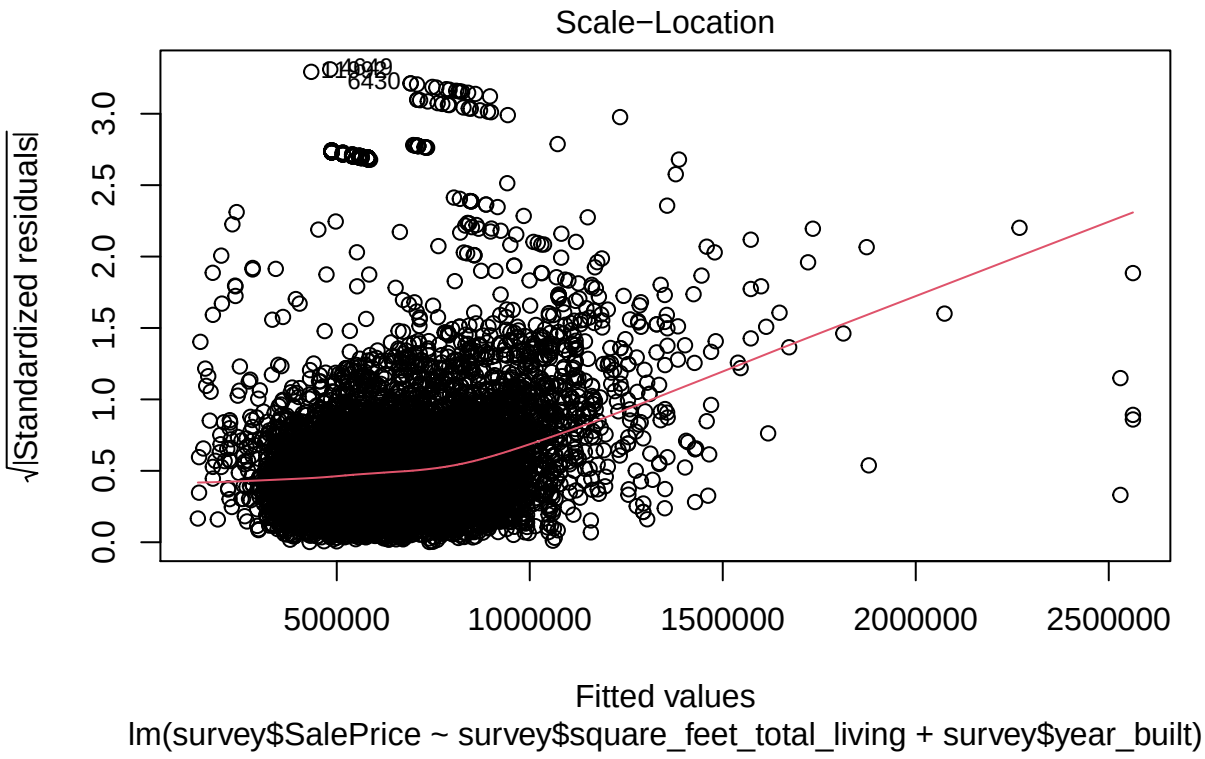
Both VIF figures are very close to 1 so we can conclude that there is no collinearity in the data.

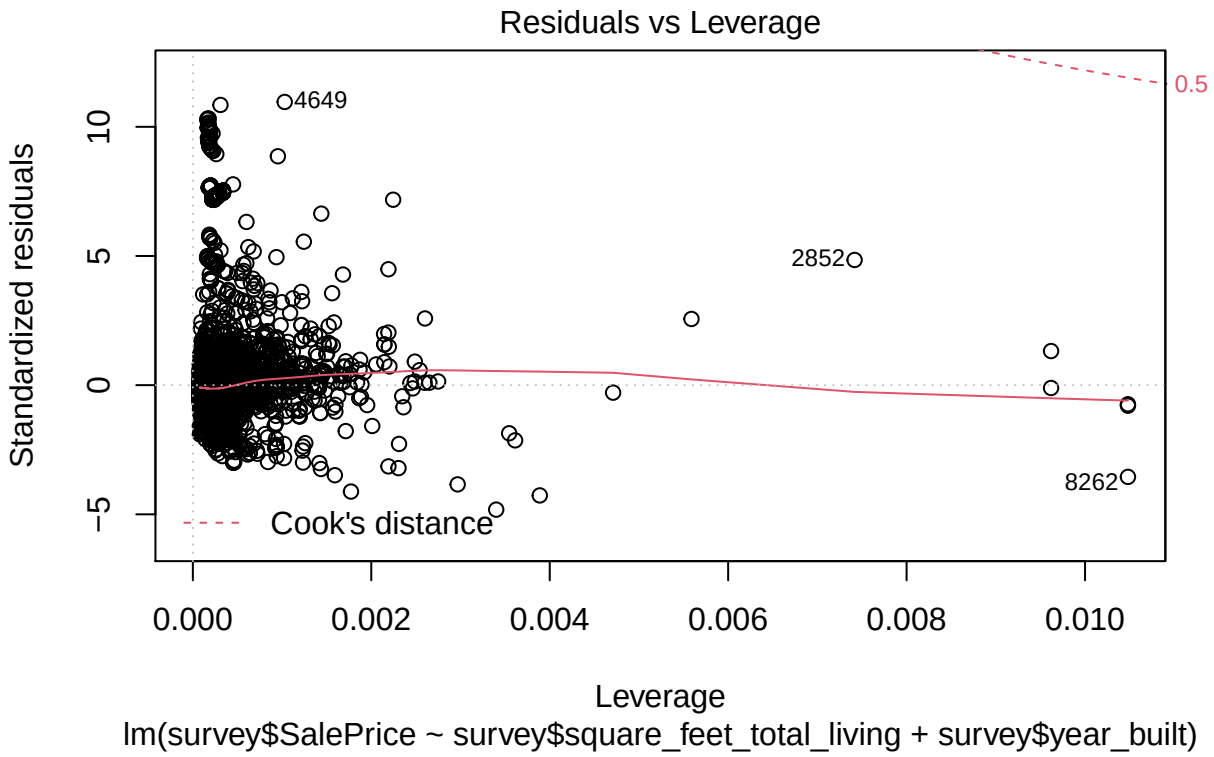
### n. Visually check the assumptions related to the residuals using the plot() and hist() functions. Summarize what each graph is informing you of and if any anomalies are present.

```
plot(sale_vars)
```

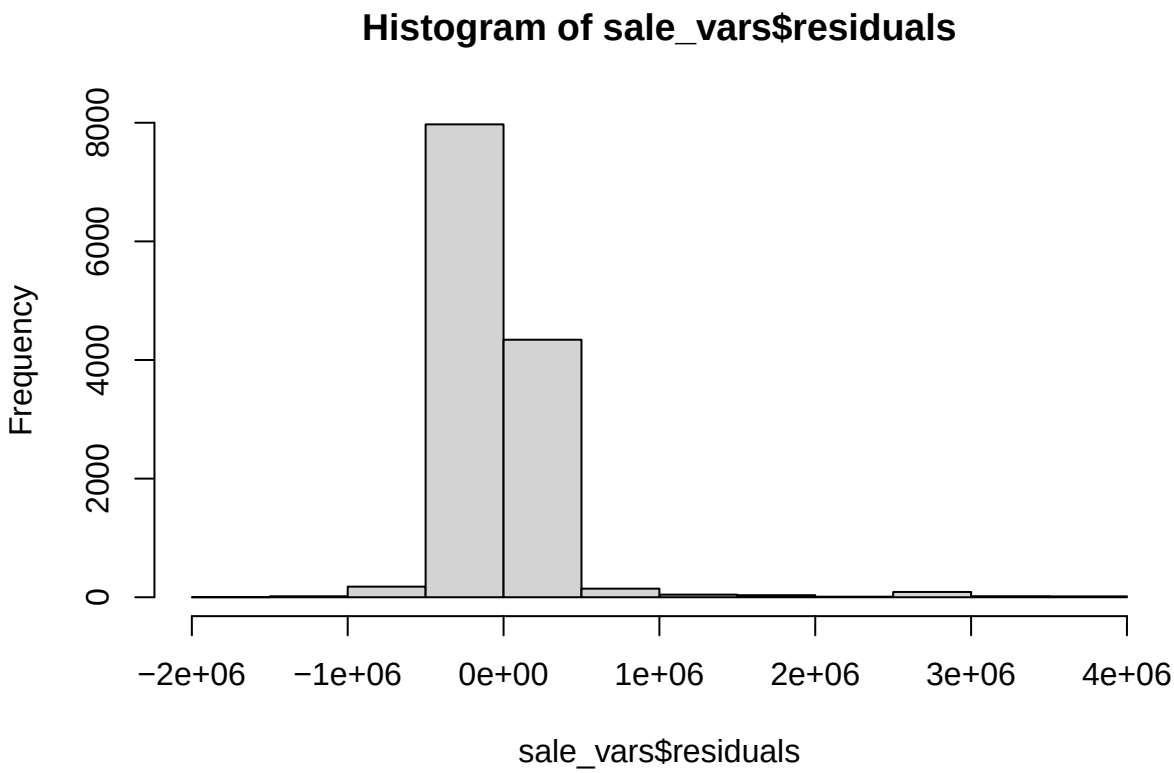



Normal Q−Q
Standardized residuals
10
5
0
−5
−4
−2
0
2
4
Theoretical Quantiles
lm(survey$SalePrice ~ survey$square_feet_total_living + survey$year_built)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
500000
1000000
1500000
2000000
2500000
Fitted values
lm(survey$SalePrice ~ survey$square_feet_total_living + survey$year_built)




```
hist(sale_vars$residuals)
```




By looking at the graph, it seems like the data may not be linear. There seems to be some outliers far out from the majority of the data in both directions.The normal Q-Q graph in particular shows a dip to the right in the mid-range of the standardized residuals.

### o. Overall, is this regression model unbiased? If an unbiased regression model, what does this tell us about the sample vs. the entire population model?

With the calculations done with the data, it seems that the data is largely unbiased. There may be some outliers but they may not be statistically significant enough to claim that the observations would be a poor representation of the entire population.