

ASSIGNMENT 6.1

Blaine Blasdel

2020-07-09

Assignment 6.1

Question 6.1a

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

Answer - At this time the data looks fairly clean. There does appear to be missing data in City name but that field seems redundant with postal city. Once we have looked at the outliers and the model, additional data might be relevant to be cleaned.

Question 6.1b

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.

Answer - see below: house_lm will contain sale price (as the outcome) and square feet of the lot (as the predictor). house_multi_lm will contain addition predictors including the year it was built, total square feet of the living area, number of bedrooms and the two common number of bathrooms.

```
house_lm <- lm(sale_price ~ sq_ft_lot, data = house_df )  
  
house_mult_lm <- lm(sale_price ~ sq_ft_lot + year_built + square_feet_total_living + bedrooms + bath_fu
```

Question 6.1c

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?

Answer: house_lm: R2 = 0.01435 & Adjusted R2 = 0.01428 house_multi_lm: R2 = 0.2208 & Adjusted R2 = 0.2204

With the single predictor of the lot size, we know it is an extremely weak prediction. We also know that the predictor of lot size only can explain 1.44% of the variance in sale price. The multi-predictors does a better job as the coefficient now shows a moderate prediction (vice weak). The inclusion of the additional predictors explains 22.1% of the variance in price, much better then 2%.

```
summary(house_lm)
```

```
##
## Call:
## lm(formula = sale_price ~ sq_ft_lot, data = house_df)
##
## 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 ***
## 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(house_mult_lm)
```

```
##
## Call:
## lm(formula = sale_price ~ sq_ft_lot + year_built + square_feet_total_living +
##      bedrooms + bath_full_count + bath_half_count, data = house_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2129073  -119468   -41494    46889   3763557
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -4.857e+06  4.332e+05 -11.212  < 2e-16 ***
## sq_ft_lot       2.742e-01  5.918e-02   4.633  3.64e-06 ***
## year_built     2.554e+03  2.186e+02  11.681  < 2e-16 ***
## square_feet_total_living 1.663e+02  4.747e+00  35.042  < 2e-16 ***
## bedrooms      -1.023e+04  4.576e+03  -2.236  0.02540 *
## bath_full_count  1.689e+04  6.092e+03   2.772  0.00558 **
## bath_half_count  9.098e+03  6.351e+03   1.433  0.15202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 357000 on 12858 degrees of freedom
## Multiple R-squared:  0.2208, Adjusted R-squared:  0.2204
## F-statistic: 607.3 on 6 and 12858 DF, p-value: < 2.2e-16
```

Question 6.1d

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(house_lm)
```

```
## sq_ft_lot  
## 0.1198122
```

```
lm.beta(house_mult_lm)
```

```
##          sq_ft_lot          year_built square_feet_total_living  
##          0.03860075          0.10873952          0.40715831  
##          bedrooms        bath_full_count        bath_half_count  
##          -0.02216417          0.02717757          0.01184097
```

Answer: The standardized beta tells the degree of importance of the variable to the outcome. In the single linear model we see that the square foot of the lot has some importance (.1198) but not great. In the multiple regression model, the total living space clearly has the highest importance the next closest is the year it was built. These make sense to me. What is counter-intuitive is that the total number of bedrooms is a negative relation, although small (-0.0222), meaning the more bathrooms typically the less the price.

Question 6.1e

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

```
confint(house_lm)
```

```
##          2.5 %          97.5 %  
## (Intercept) 6.343730e+05 6.492698e+05  
## sq_ft_lot   7.291208e-01 9.728641e-01
```

```
confint(house_mult_lm)
```

```
##          2.5 %          97.5 %  
## (Intercept) -5.706533e+06 -4.008138e+06  
## sq_ft_lot    1.581759e-01  3.901644e-01  
## year_built    2.125018e+03  2.982041e+03  
## square_feet_total_living 1.570362e+02 1.756455e+02  
## bedrooms     -1.919987e+04 -1.260103e+03  
## bath_full_count 4.945975e+03 2.882831e+04  
## bath_half_count -3.351059e+03 2.154782e+04
```

Answer: The square foot of the lot size and the total living space have the smallest confidence intervals (likely to be representative of the model) of my variables meaning they are the most significant when determining the price of house. The count of half bathrooms crosses zero so it is not a good representative. The number of bedrooms, number of full bathrooms, and the year built are all significant but less representative.

Question 6.1f

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(house_lm, house_mult_lm)
```

```
## Analysis of Variance Table
##
## Model 1: sale_price ~ sq_ft_lot
## Model 2: sale_price ~ sq_ft_lot + year_built + square_feet_total_living +
##          bedrooms + bath_full_count + bath_half_count
##   Res.Df      RSS Df Sum of Sq    F    Pr(>F)
## 1  12863 2.0734e+15
## 2  12858 1.6391e+15   5 4.3429e+14 681.37 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: In this case the $\Pr(>F)$ or the p-value associated with the F-Statistic is very small ($2.2e-16$) showing a significant improvement over the original model.

Question 6.1g

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

```
house_df$residuals <- resid(house_mult_lm)
house_df$standardized.residuals <- rstandard(house_mult_lm)
house_df$studentized.residuals <- rstudent(house_mult_lm)
house_df$cooks.distance <- cooks.distance(house_mult_lm)
house_df$dfbeta <- dfbeta(house_mult_lm)
house_df$dffit <- dffits(house_mult_lm)
house_df$leverage <- hatvalues(house_mult_lm)
house_df$covariance.ratios <- covratio(house_mult_lm)
```

Question 6.1h

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

Answer: 2.6% of the cases have a large residual, less than the 5% expected.

Question 6.1i

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

```
## [1] 12865
## [1] 334
## [1] "2.6"
## [1] 12865
## [1] 253
## [1] "2"
```

Question 6.1j

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

Answer: About 334 have large residuals. See table below.

```
## # A tibble: 334 x 6
##   sq_ft_lot year_built square_feet_tot~ bedrooms bath_full_count
##   <dbl>    <dbl>         <dbl>    <dbl>         <dbl>
## 1      7280     2005         4160      4             2
## 2     112650    2007         4920      4             4
## 3     225640    1955          660      0             1
## 4     236966    2008         3840      0             0
## 5      63162    2008         5800      5             4
## 6       8752    2005         3360      2             2
## 7     14043    1918          900      2             1
## 8     18498    2014         4710      4             2
## 9     89734    2016         5060      4            23
## 10    288367    2008         6880      5             1
## # ... with 324 more rows, and 1 more variable: bath_half_count <dbl>
```

Question 6.1k

Investigate further by calculating the leverage, cooks distance, and covariance ratios. Comment on all cases that are problematic.

Answer: None of the cases that have a large residual have a Cook's distance greater than 1, meaning they do not have any undue influence on the model. Regarding, leverage, using the book's calculation for average leverage (.000544), of the 334 with large residual, I show 230 cases that are within two times the average and 264 cases that are within three times the average. That gives me 70 cases not within three times the average that could be problematic. Using the formula for upper and lower covariance values, I have 18 cases that do not fall outside the range and could be problematic.

```
house_df[house_df$large.residual, c("cooks.distance", "leverage", "covariance.ratios")]
```

```
## # A tibble: 334 x 3
##   cooks.distance leverage covariance.ratios
##   <dbl>    <dbl>         <dbl>
## 1    0.000257 0.000383         0.998
## 2    0.00111 0.00126         0.998
## 3    0.00343 0.00267         0.998
## 4    0.00309 0.00491         1.00
## 5    0.00116 0.00131         0.998
## 6    0.000457 0.000739         0.999
## 7    0.00360 0.00202         0.996
## 8    0.000613 0.000655         0.998
## 9    0.270    0.120         1.13
## 10   0.00733 0.00403         0.998
## # ... with 324 more rows
```

```
write.table(house_df[house_df$large.residual, c("cooks.distance", "leverage", "covariance.ratios")])
```

```

## "cooks.distance" "leverage" "covariance.ratios"
## "1" 0.000257383930524949 0.000383329733282036 0.998370902454602
## "2" 0.00111367769184751 0.00125820174252634 0.998434961540868
## "3" 0.00342501158315409 0.00267189997066817 0.998347628558405
## "4" 0.00308546583020466 0.00491369317553282 1.00309340268102
## "5" 0.00116202032434918 0.0013122754062495 0.998487812574316
## "6" 0.000456916247696029 0.000739155673581695 0.998930058210765
## "7" 0.00360374087895116 0.00201837783539165 0.995780065616874
## "8" 0.000612947509283165 0.000655351141238256 0.997639933377151
## "9" 0.269539291306454 0.120364229167083 1.12894214109025
## "10" 0.00733395204764774 0.00403449865904648 0.997686899995633
## "11" 0.000293951011758788 0.000486210398608829 0.998728356792143
## "12" 0.00583038210873332 0.00884069343579145 1.0069570612335
## "13" 0.00159475846926246 0.00142362050028979 0.997707502421256
## "14" 0.00320370317684461 0.0020907848033813 0.996812909913258
## "15" 0.00151404665179541 0.00139145277022301 0.997797212904713
## "16" 0.000552318578839859 0.000699786856475117 0.998239688439155
## "17" 0.00304604551429804 0.00394282056584676 1.00156316093324
## "18" 0.00250538842345315 0.0026324320382144 0.999562099328376
## "19" 0.000806755114442958 0.00103852074873737 0.998626491485895
## "20" 0.000797016838675404 0.00130958911815408 0.999538347086014
## "21" 0.00331239438969213 0.00369528891039208 1.00084272530017
## "22" 0.00189385426463357 0.000689807812555491 0.990813694860428
## "23" 0.0379282875171551 0.0163169342754208 1.00831117204959
## "24" 0.00195742650524271 0.00232072079333373 0.999660338274135
## "25" 0.00104342631633625 0.00166103046711746 0.999816521299398
## "26" 0.00535276015425709 0.00135739639862703 0.986964994781591
## "27" 0.000518283704364107 0.000833764392311292 0.999011704253288
## "28" 0.0013541290340918 0.00129624136230475 0.997866778261491
## "29" 0.00519395966108859 0.0033601801099693 0.998038934757099
## "30" 0.000442657623514157 0.000650299758731963 0.998603081919238
## "31" 0.0027072813706656 0.00202067532258803 0.997473059054203
## "32" 0.00154327935466256 0.00132716541797209 0.997448805528421
## "33" 0.0160060353530371 0.00978541044834077 1.00421171548595
## "34" 0.00121836017568014 0.00179378460786451 0.999755640443345
## "35" 0.000698035752730612 0.00115951943069253 0.99941292631433
## "36" 0.000441430722721525 0.000605449107640067 0.998374034535163
## "37" 0.00164252719749003 0.00185193278013228 0.999024053861432
## "38" 0.000882444846627467 0.00118438067866849 0.998893562809692
## "39" 0.00538316760093428 0.00330190582746794 0.997659394184877
## "40" 0.0011777945011976 0.00117844271506162 0.997920425691045
## "41" 0.000495015201018361 0.000715113386657249 0.99862418744688
## "42" 0.0159311746689521 0.00614986662102166 0.996899937375352
## "43" 0.00123425835485701 0.000954727218462324 0.996581816175303
## "44" 0.00111146110704511 0.000903543315240914 0.9967684938796
## "45" 0.00631028576809476 0.00716822602440538 1.00441670175543
## "46" 0.00212920228205053 0.00115980435877814 0.994727398519516
## "47" 0.000493210417695706 0.000552138451071156 0.99769629430717
## "48" 0.00146851780599289 0.000773117971449395 0.99409855997143
## "49" 0.00238358308524758 0.000723706337514985 0.988778315554249
## "50" 0.00076019404383784 0.000210585975169376 0.987071891857353
## "51" 0.00198030874162637 0.000567618741827188 0.987885782028132
## "52" 0.000741357856026638 0.000206315212464459 0.98713006769146
## "53" 0.00129695042047597 0.000366157780887632 0.987483309298344

```

"54" 0.00271057887196207 0.000860102247278653 0.989451199578894
 ## "55" 0.00166914999638745 0.000498287142340821 0.988340548418354
 ## "56" 0.00114258058905006 0.00191885361232428 1.00019996077263
 ## "57" 0.0023476930577989 0.00164363121681254 0.996758248411504
 ## "58" 0.000574412249496688 0.000717113062297557 0.998212402087472
 ## "59" 0.000758197305496767 0.000520539442768093 0.995525091711362
 ## "60" 0.00057422728714872 0.00052087761339067 0.996870104216538
 ## "61" 0.000183103195757758 0.00027139296209997 0.998246517885125
 ## "62" 0.0018298872136692 0.000950405189272279 0.9941781001615
 ## "63" 0.00410735196085884 0.00524275323461544 1.00283445023719
 ## "64" 0.000825783686878585 0.000818343945746008 0.997522771332619
 ## "65" 0.00385929754903071 0.00168953422551744 0.993560620328082
 ## "66" 0.00066780160392535 0.000567411195935589 0.996633688500087
 ## "67" 0.000426576154716979 0.000640627901325627 0.998649798171561
 ## "68" 0.00254256624394798 0.000738453482420168 0.988229227549838
 ## "69" 0.00117347407511062 0.000368319273206033 0.988828145799571
 ## "70" 0.00156608527271913 0.000560459268306057 0.990499694286414
 ## "71" 0.00120257438483009 0.000364301809310667 0.988390204505272
 ## "72" 0.000708251393873085 0.000202616397142849 0.987494965923701
 ## "73" 0.00148788722858822 0.000514991441055669 0.990095801942355
 ## "74" 0.00141618131960244 0.000507655581177514 0.990464336737204
 ## "75" 0.0345732588300186 0.0255437229514183 1.02162248612249
 ## "76" 0.00109666539115025 0.000689929417485282 0.995190279165194
 ## "77" 0.00119045531806037 0.000866532005521396 0.996185795410542
 ## "78" 0.00125783899427389 0.000867998835038566 0.995901435914524
 ## "79" 0.00268682375712967 0.00128941450326911 0.993918097161537
 ## "80" 0.000462160352496803 0.000438053475329783 0.996967218694332
 ## "81" 0.0285098697362176 0.00828751807483054 0.995861904450998
 ## "82" 0.00123310836661982 0.000954727218462324 0.996586389596583
 ## "83" 0.00119093559971628 0.000954727218462324 0.996754120112777
 ## "84" 0.00230554712728193 0.00116246950383082 0.994171198231661
 ## "85" 0.000734601298119323 0.000665069680718185 0.997006470860524
 ## "86" 0.000910869597722001 0.00102060570359484 0.998168769772574
 ## "87" 0.000649395203264094 0.00092045478601524 0.998779393699225
 ## "88" 0.00466720150958387 0.000573551265637859 0.97050060449904
 ## "89" 0.00423181512629417 0.000526679872941354 0.970834236845555
 ## "90" 0.00289798835237338 0.00037580702681909 0.971885380131263
 ## "91" 0.00242437877059519 0.00031928139480603 0.972268118567294
 ## "92" 0.00250537457027 0.000328750184652342 0.972174679175554
 ## "93" 0.00267836456700831 0.000350330561158593 0.972105751106264
 ## "94" 0.00250481582816416 0.000328667519885534 0.972173790345692
 ## "95" 0.00331793391114126 0.000418806308971858 0.97114426332352
 ## "96" 0.00230164695755809 0.000306906205908092 0.972604375540884
 ## "97" 0.00219096277816158 0.000292335320061951 0.972607733376152
 ## "98" 0.00250971069645491 0.000332523924363174 0.972451643757861
 ## "99" 0.00281809324757346 0.000364669854149164 0.971813186877459
 ## "100" 0.00219384449338318 0.000292780599642761 0.972613967589566
 ## "101" 0.00281696545344435 0.000364508498649274 0.971811810357695
 ## "102" 0.00466739465590998 0.000573595855125611 0.97050174754402
 ## "103" 0.00466705457334501 0.000573516200035378 0.970499673395448
 ## "104" 0.00466705457334501 0.000573516200035378 0.970499673395448
 ## "105" 0.0046674182539061 0.000573601201670815 0.970501881751587
 ## "106" 0.00466720150958387 0.000573551265637859 0.97050060449904
 ## "107" 0.00466720150958387 0.000573551265637859 0.97050060449904

"108" 0.00423322852646142 0.000526997103639079 0.970842556992224
 ## "109" 0.00331655775034007 0.000418610570152204 0.971142518158457
 ## "110" 0.00250217964051791 0.000328275676291926 0.972169431898287
 ## "111" 0.0033160089934592 0.000418532392022552 0.971141812904651
 ## "112" 0.00242473481916923 0.000319335248898731 0.972268788349605
 ## "113" 0.00250159177393454 0.000328187840149553 0.972168418879244
 ## "114" 0.00250120550514671 0.000328130025524587 0.972167744226652
 ## "115" 0.00219155352713265 0.000292427299411589 0.972609080018447
 ## "116" 0.00219280288686717 0.000292620580750979 0.97261180568873
 ## "117" 0.00231031334813488 0.000308213244336824 0.972619393722189
 ## "118" 0.00466783029799415 0.000573691684335527 0.970504070589849
 ## "119" 0.00466779206291959 0.000573683490135462 0.970503878347325
 ## "120" 0.00466705457334501 0.000573516200035378 0.970499673395448
 ## "121" 0.00466705457334501 0.000573516200035378 0.970499673395448
 ## "122" 0.00466705457334501 0.000573516200035378 0.970499673395448
 ## "123" 0.000483488490635384 0.000837428720132822 0.99918384549723
 ## "124" 0.000513256249282224 0.000471181945044284 0.996870207667327
 ## "125" 0.00325355557723867 0.000418479361736214 0.971692638720827
 ## "126" 0.0028292358208826 0.000373799390493557 0.972413343600491
 ## "127" 0.00214824196775484 0.000292247076403788 0.973143225121057
 ## "128" 0.00215093004157187 0.000292673058853598 0.97314928856666
 ## "129" 0.00325498986234318 0.000418687822067179 0.971694502571082
 ## "130" 0.002757913227453 0.000363679298167663 0.97234926353627
 ## "131" 0.00214758763288706 0.000292141979668965 0.973141613805851
 ## "132" 0.00283051094832594 0.000373986902266392 0.972414964735405
 ## "133" 0.00237829356423694 0.000319323807129967 0.972808823832107
 ## "134" 0.00245587827800702 0.000328496996217451 0.972713073072008
 ## "135" 0.00325838441175269 0.000419179224243153 0.971698769829972
 ## "136" 0.00238198764685643 0.000319889739392416 0.972815449974854
 ## "137" 0.00352109559484517 0.000471990029567121 0.972916243878789
 ## "138" 0.00238072010060209 0.000319696439204742 0.972813256582611
 ## "139" 0.00458081033558097 0.000573516200035378 0.971058127583851
 ## "140" 0.00458081033558097 0.000573516200035378 0.971058127583851
 ## "141" 0.00458081033558097 0.000573516200035378 0.971058127583851
 ## "142" 0.00458081033558097 0.000573516200035378 0.971058127583851
 ## "143" 0.00458081033558097 0.000573516200035378 0.971058127583851
 ## "144" 0.00458081128285495 0.00057351643622751 0.971058133903732
 ## "145" 0.00415301233417049 0.000526649826012915 0.971388366716165
 ## "146" 0.00228150339176991 0.00031160052976678 0.973270128711881
 ## "147" 0.00275731113062445 0.000363590535196568 0.972348448112135
 ## "148" 0.00221220371401542 0.000301539949097626 0.973206200703637
 ## "149" 0.0024535005300549 0.000328135005225978 0.97270898544941
 ## "150" 0.00237685543698298 0.000319101051294616 0.97280602447558
 ## "151" 0.00268192295243972 0.000356199655880551 0.972543008751056
 ## "152" 0.00517852896852539 0.000689200450639741 0.972933892673064
 ## "153" 0.00299455785379112 0.00039287330467474 0.972234061749773
 ## "154" 0.00214879703481136 0.00029233574606613 0.973144545444851
 ## "155" 0.00325534694795722 0.00041873964090322 0.971694960723864
 ## "156" 0.00458081128285495 0.00057351643622751 0.971058133903732
 ## "157" 0.00458081128285495 0.00057351643622751 0.971058133903732
 ## "158" 0.00458081128285495 0.00057351643622751 0.971058133903732
 ## "159" 0.00240146243271876 0.00153914209586646 0.996152879718985
 ## "160" 0.000606554384788247 0.000623221795580564 0.997463457961469
 ## "161" 0.00172231192062287 0.00202067532258803 0.999325025695766

"162" 0.00100625309622715 0.00132716541797209 0.99898684844448
 ## "163" 0.0010706111004342 0.000787452326144973 0.996160548701612
 ## "164" 0.00608908845227897 0.00176623876019583 0.98924351560713
 ## "165" 0.00074375156108517 0.00125388462747139 0.999541236248449
 ## "166" 0.000538179063296849 0.000750137630060078 0.998563336165926
 ## "167" 0.00144596154653493 0.00216640643930252 1.00017470597173
 ## "168" 0.000348060468199054 0.000467869401400809 0.998179821720803
 ## "169" 0.000644615437825109 0.00108074811209351 0.999355061077324
 ## "170" 0.10605341155971 0.0103184768940817 0.972432035810669
 ## "171" 0.68860768011274 0.0441637034796997 0.988749827687195
 ## "172" 0.00370629547883743 0.00409191582471202 1.00120702405528
 ## "173" 0.00184624846864659 0.000505764863742478 0.987214779145576
 ## "174" 0.00137738389019818 0.000964691412746963 0.996079297948872
 ## "175" 0.000364026938760907 0.000540195563761684 0.998518725582263
 ## "176" 0.000932196047990724 0.000788884614714331 0.996837591757888
 ## "177" 0.00200713979443639 0.00164363121681254 0.99754486460265
 ## "178" 0.0167093285350971 0.00978541044834077 1.00393915228598
 ## "179" 0.00123942296430265 0.00125489545190656 0.998041875657598
 ## "180" 0.0799703587777597 0.0307977237228173 1.02247806509614
 ## "181" 0.000152894832305823 0.00022661333305427 0.998201632539936
 ## "182" 0.000148087668675483 0.00022414281605117 0.998252463699951
 ## "183" 0.000259203725951328 0.000374859413454746 0.99828623110305
 ## "184" 0.000152921585980952 0.000226628148610211 0.998201366049048
 ## "185" 0.00058241140810976 0.000851935569721729 0.998793962972809
 ## "186" 0.00104086596378478 0.00159465270896969 0.999656507026796
 ## "187" 0.00214123120905667 0.00156677889356092 0.996915341490914
 ## "188" 0.00108457580772343 0.00154938669437898 0.999431193165666
 ## "189" 0.00112622264368175 0.000239239675897173 0.982971977762556
 ## "190" 0.00127342942585954 0.000281379112242391 0.983697099026176
 ## "191" 0.00161963489393509 0.000374878338815011 0.984562043679898
 ## "192" 0.00201846470884243 0.000414437723775165 0.982536029386915
 ## "193" 0.00319834915610788 0.000668354347262999 0.983109196536433
 ## "194" 0.00321038950302089 0.000668736597539989 0.98305226373065
 ## "195" 0.00319820600817708 0.000668339892647964 0.983109597549666
 ## "196" 0.00328185560946062 0.000685283846208607 0.983112506251906
 ## "197" 0.00201149893097494 0.000407136149088233 0.982265174951337
 ## "198" 0.00127374884141059 0.000281443969801345 0.983696818327415
 ## "199" 0.00333858794302623 0.000236396701373414 0.948157797652886
 ## "200" 0.00443401622125822 0.000289708724879855 0.943915514298292
 ## "201" 0.00384838869018069 0.000259161048198605 0.945538618636495
 ## "202" 0.00567357105133668 0.00038942849451699 0.946687836465919
 ## "203" 0.00685849408923828 0.000477057647006608 0.947474770380707
 ## "204" 0.00724230360995702 0.000505173797744353 0.947649839637588
 ## "205" 0.0102220845821762 0.000745290213326859 0.950150419275194
 ## "206" 0.00374011005488371 0.000253545513391485 0.945889599099
 ## "207" 0.00406536249914517 0.000268026236221252 0.944391859703588
 ## "208" 0.00343111931399779 0.000225549324567847 0.944187963363828
 ## "209" 0.00318230457711934 0.000222266138319377 0.947435028515076
 ## "210" 0.00802867033896048 0.00057654909605402 0.949216976274651
 ## "211" 0.00853695106745841 0.000588130910961099 0.947081355113196
 ## "212" 0.00574014641562243 0.000397435295007136 0.94715394956669
 ## "213" 0.00915222901978592 0.00071911092042013 0.953733289000138
 ## "214" 0.00750435794859625 0.000608802755626184 0.955089396393284
 ## "215" 0.00723015812753357 0.000551133920654844 0.952134105902394

"216" 0.00531322787402158 0.000423827109821166 0.954135940769749
 ## "217" 0.00658109591937211 0.000547736530524056 0.956169019319871
 ## "218" 0.00362308462753786 0.000274703448628208 0.951600553067857
 ## "219" 0.00561960491395035 0.000468739683879192 0.956186461379746
 ## "220" 0.00555368327558869 0.000477638931770964 0.957521034167923
 ## "221" 0.00710242310556529 0.000541075019972622 0.952095523099271
 ## "222" 0.00525536469227199 0.000420039929122831 0.954222640441984
 ## "223" 0.00896994429410069 0.000754820981169192 0.956858600588577
 ## "224" 0.00717582892763177 0.000586936934771532 0.955435700527616
 ## "225" 0.0115895690302616 0.000899476578920052 0.953337456796364
 ## "226" 0.00352038093841305 0.000269989754812785 0.952144205339868
 ## "227" 0.00599869460753722 0.000487228002540594 0.955020964314504
 ## "228" 0.00223814903917653 0.000997122026200959 0.993015976422772
 ## "229" 0.00115451394182786 0.00151080236696446 0.999148379556282
 ## "230" 0.000371856102666462 0.000650299758731963 0.999017367778786
 ## "231" 0.00123161021530869 0.00125000073941826 0.998046021924449
 ## "232" 0.00441783021494932 0.00766906757817268 1.00608270111876
 ## "233" 0.00025092880158679 0.00040170359072756 0.998567296629676
 ## "234" 0.000963920619648686 0.00139700342132114 0.999316364666412
 ## "235" 0.000220142336145986 0.000322893351688194 0.998271060718411
 ## "236" 0.00127262610684186 0.000611299239503028 0.993245597967382
 ## "237" 0.00155264156726897 0.00090119182846815 0.994896303929536
 ## "238" 0.000597558300111317 0.000251910346755071 0.9917869616482
 ## "239" 0.000594918041965514 0.000251553213447919 0.991813573618228
 ## "240" 0.00151108718458727 0.000814695117112127 0.994309567885349
 ## "241" 0.00245868955501958 0.0013558607878952 0.995009143295093
 ## "242" 0.00175659470265911 0.00102406509598435 0.995048144467275
 ## "243" 0.00127262610684186 0.000611299239503028 0.993245597967382
 ## "244" 0.00059762485367201 0.000252536094156221 0.991808829472105
 ## "245" 0.000596930850531484 0.000246026410410062 0.991576097850822
 ## "246" 0.000598868478080355 0.000244150978639601 0.991473701845694
 ## "247" 0.000997818211216286 0.00143582358344349 0.999336472295887
 ## "248" 0.000775909338208097 0.000788123385069345 0.997585936874706
 ## "249" 0.00585100552147275 0.00100774678723319 0.979624513030527
 ## "250" 0.00386437160191879 0.00152922949617662 0.992481296463763
 ## "251" 0.00357718086228746 0.00117151724515523 0.990133305156593
 ## "252" 0.00241207299273105 0.000287564566214918 0.969284450046439
 ## "253" 0.00266977022226755 0.000313312020429124 0.968816240759001
 ## "254" 0.0024040804811959 0.000286811494638808 0.969305371744075
 ## "255" 0.00203704271964238 0.000240854192875325 0.968979393997572
 ## "256" 0.00203981477251676 0.000241035428059924 0.968960507208668
 ## "257" 0.00204261775282671 0.000241220867133328 0.968941725407218
 ## "258" 0.00267061682445672 0.00031342707509272 0.968817938759232
 ## "259" 0.00234067722815932 0.000277883119305243 0.969143787163086
 ## "260" 0.00232312798160977 0.000276140033649453 0.969180551973776
 ## "261" 0.00253640288192749 0.00029749211333593 0.968782510487995
 ## "262" 0.00266964891619443 0.000313295489011872 0.968815992708094
 ## "263" 0.00242954449557062 0.000289286102151895 0.969247300475465
 ## "264" 0.00204058906671234 0.000241021012942149 0.968946701086668
 ## "265" 0.00284288870003096 0.00033058568394292 0.968542745537367
 ## "266" 0.00203981477251676 0.000241035428059924 0.968960507208668
 ## "267" 0.00277063765020979 0.000323352959333496 0.968650765142414
 ## "268" 0.0028063743083378 0.000326931058752611 0.968596763227615
 ## "269" 0.00277069510284787 0.000323360685784099 0.968650873210865

"270" 0.0024122675343462 0.00232072079333373 0.998915326628151
 ## "271" 0.000692202076239447 0.00104778765641085 0.999077803577652
 ## "272" 0.000692388596923858 0.0010482503750547 0.999078700043153
 ## "273" 0.00157806737771878 0.00221525435402339 1.000052914713
 ## "274" 0.0047775620265194 0.00318326108952513 0.998031109713808
 ## "275" 0.000275016393431277 0.000397082680392985 0.998304210942694
 ## "276" 0.00243081590999455 0.00232072079333373 0.998884955223578
 ## "277" 0.0282198764421084 0.0156828986228449 1.00964483290271
 ## "278" 0.00693165124941345 0.00517091207539152 1.00064504858322
 ## "279" 0.376761951507559 0.0648584650680358 1.04798546195204
 ## "280" 0.000477855245754919 0.000552138451071156 0.997802029292922
 ## "281" 0.0014370801406443 0.000773117971449395 0.994252658525883
 ## "282" 0.000390529798379769 0.000586819144639134 0.998597298538399
 ## "283" 0.000498171696392688 0.000715113386657249 0.998607395356791
 ## "284" 0.000209751172729719 0.000261433424201503 0.997751064995333
 ## "285" 0.00220008101852355 0.000699148159541966 0.989307498734713
 ## "286" 0.000465850016555904 0.000510930742446826 0.997584710289629
 ## "287" 0.0130226637561288 0.00302653329852622 0.987290045629808
 ## "288" 0.00128083641883174 0.00177672680702853 0.999579934280273
 ## "289" 0.00128083641883174 0.00177672680702853 0.999579934280273
 ## "290" 0.00280236021502446 0.00286283338985061 0.999690748778811
 ## "291" 0.000826271518372155 0.00103852074873737 0.998555020577982
 ## "292" 0.00137087323722742 0.00207774266443026 1.00011475018314
 ## "293" 0.00618151188173618 0.00731605357836287 1.00470133415624
 ## "294" 0.000643070430316837 0.000954727218462324 0.998935302282165
 ## "295" 0.0071626594201534 0.00273330310252101 0.993337567593871
 ## "296" 0.000580783125279023 0.000623221795580564 0.997620639405214
 ## "297" 0.00063923752996982 0.000540195563761684 0.99658205247576
 ## "298" 0.00144179501447737 0.00210830876250485 1.00005387108872
 ## "299" 0.0156529076866929 0.0226598391587872 1.02111143047112
 ## "300" 0.00226384366669782 0.0031256469669161 1.00092335598499
 ## "301" 0.0016863456700547 0.00224112387569768 0.999926412736773
 ## "302" 0.000435696217749467 0.000639179589614669 0.99858832207685
 ## "303" 0.000126801066223671 0.000184071507046077 0.998105126889001
 ## "304" 0.000436164272173802 0.000668578558216554 0.998729151073909
 ## "305" 0.000253078780301103 0.000424749505006028 0.998700085597907
 ## "306" 0.0030386018180864 0.00156751798389547 0.994747442394213
 ## "307" 0.00175425921066792 0.00212756009128179 0.999538139815152
 ## "308" 0.000527636334229567 0.000561715743986196 0.997530800997988
 ## "309" 0.00327906883974016 0.00438173421123678 1.00209805619081
 ## "310" 0.00121299626333662 0.00126234123809399 0.99815105572938
 ## "311" 0.000677386236397154 0.000254473107774603 0.990693553740403
 ## "312" 0.00306011532594868 0.00133934508928414 0.993207117005558
 ## "313" 0.00316346214853504 0.000779623081112377 0.985955645064465
 ## "314" 0.00178385127647314 0.00116358737804677 0.995879292905973
 ## "315" 0.00717440118533044 0.00485713256265891 0.999809549785276
 ## "316" 0.00243650480659752 0.0031256469669161 1.00071324427739
 ## "317" 0.0144784502407767 0.00225796777617427 0.978613743536127
 ## "318" 0.00261909886803433 0.000670427888585934 0.98641482914643
 ## "319" 0.0408630051387259 0.0351753222879337 1.03260080123683
 ## "320" 0.0862056639778361 0.0370847192745079 1.03024745800608
 ## "321" 0.00916104618430972 0.0033601801099693 0.99356853178604
 ## "322" 0.0758732644838351 0.00473473085099684 0.945771579601073
 ## "323" 0.0017267514855945 0.0017192117860799 0.998444299395309

```
## "324" 0.00241078605127485 0.00149412799871436 0.995905727479779
## "325" 0.00112043488601784 0.00116632647009995 0.99805573261688
## "326" 0.000812436644438571 0.00104888836369852 0.998645488603813
## "327" 0.00537313451062185 0.00172817073442998 0.990481646453346
## "328" 0.00362572389564412 0.00210545140181011 0.996107903094211
## "329" 0.0053720123552743 0.00207418058186529 0.992790506227141
## "330" 0.0218207803449855 0.00294248730564416 0.975560739426715
## "331" 0.00765362371695188 0.000892525148893719 0.96919552951682
## "332" 0.00038509745895089 0.000620020026869079 0.998799488736521
## "333" 0.000280735363481233 0.000466589134991383 0.998719741915275
## "334" 0.00770188363366852 0.00315472787947162 0.994439121506902
```

Question 6.1l

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

Answer: With a p-value near zero, we can reject the null hypothesis. As the value of DW Statistic is less than one, our condition is not met.

```
durbinWatsonTest(house_mult_lm)
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 0.7208184 0.558353 0
## Alternative hypothesis: rho != 0
```

Question 6.1m

Perform the necessary calculations to assess the assumption of no multicollinearity and state if the condition is met or not.

Answer: The largest VIF is nowhere near 10 (being below 10) and our tolerance is above 0.2. The average which shows some correlation but not enough to cause concern. Based on what I see, the multi-regression model so no multicollinearity.

```
vif(house_mult_lm)
```

```
##          sq_ft_lot          year_built square_feet_total_living
##          1.145449          1.430110          2.227822
##          bedrooms          bath_full_count          bath_half_count
##          1.622102          1.586187          1.127451
```

```
1/vif(house_mult_lm)
```

```
##          sq_ft_lot          year_built square_feet_total_living
##          0.8730201          0.6992467          0.4488689
##          bedrooms          bath_full_count          bath_half_count
##          0.6164840          0.6304426          0.8869567
```

```
mean(vif(house_mult_lm))
```

```
## [1] 1.523187
```

Question 6.1n

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.

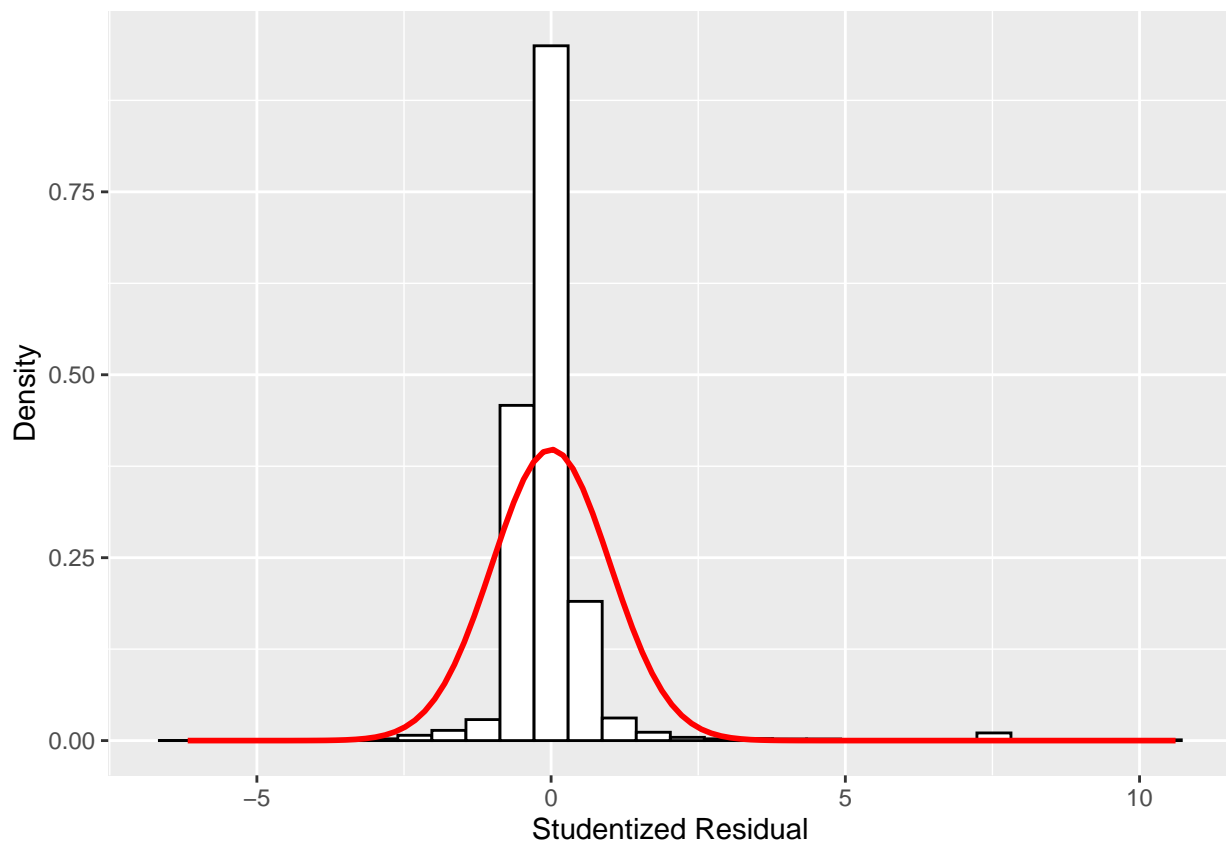
Answer: My histogram of Studentized Residuals is fairly normal. In the fitted values against residuals, my graph is clustered along the line but does not closely resemble any of the examples that would violate, randomness, heteroscedasticity or linearity. In the middle of my plot on observed values verse theoretical values you can see that there is a section in the middle that does deviate from normality and a skew. If the data was cleaned and additional outliers removed, it would most likely be closer to normal. Also, trying different variables and parameters may include additional findings.

```
house_df$fitted <- house_mult_lm$fitted.values
```

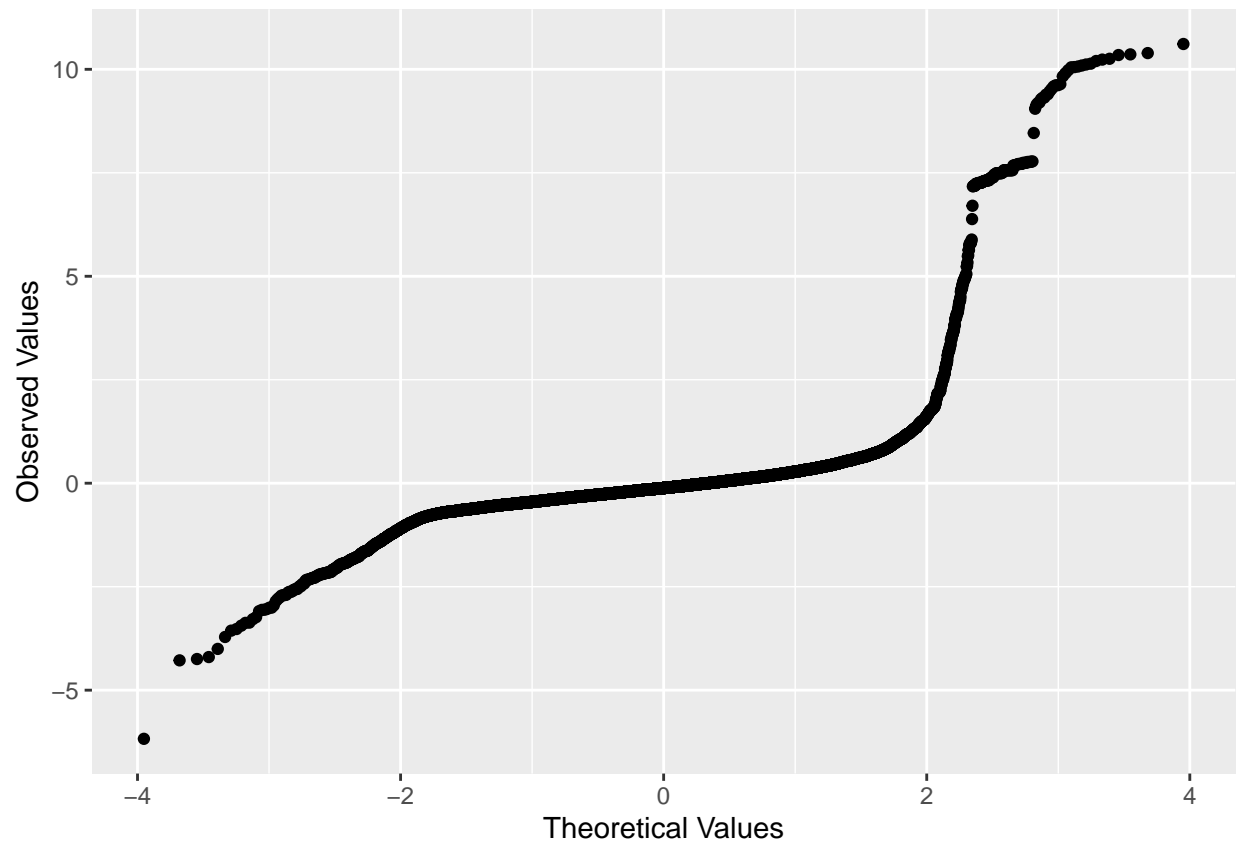
```
histogram <- ggplot(house_df, aes(studentized.residuals)) + geom_histogram(aes(y = ..density..), colour = 'red', fill = 'white', border = 'black') +
```

```
histogram + stat_function(fun = dnorm, args = list(mean = mean(house_df$studentized.residuals), na.rm = TRUE))
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

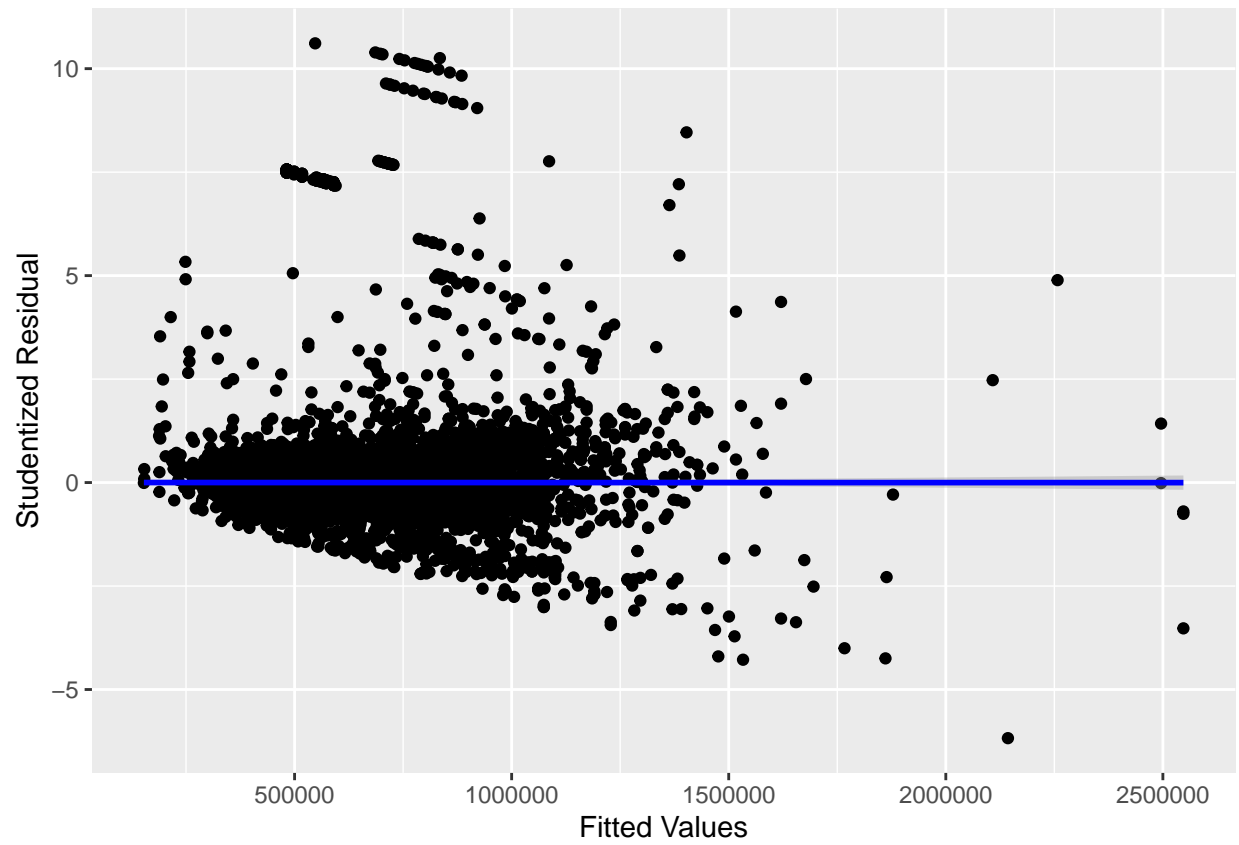


```
qqPlot.resid <- qplot(sample = house_df$studentized.residuals) + labs(x = "Theoretical Values", y = "Observed Values")
qqPlot.resid
```

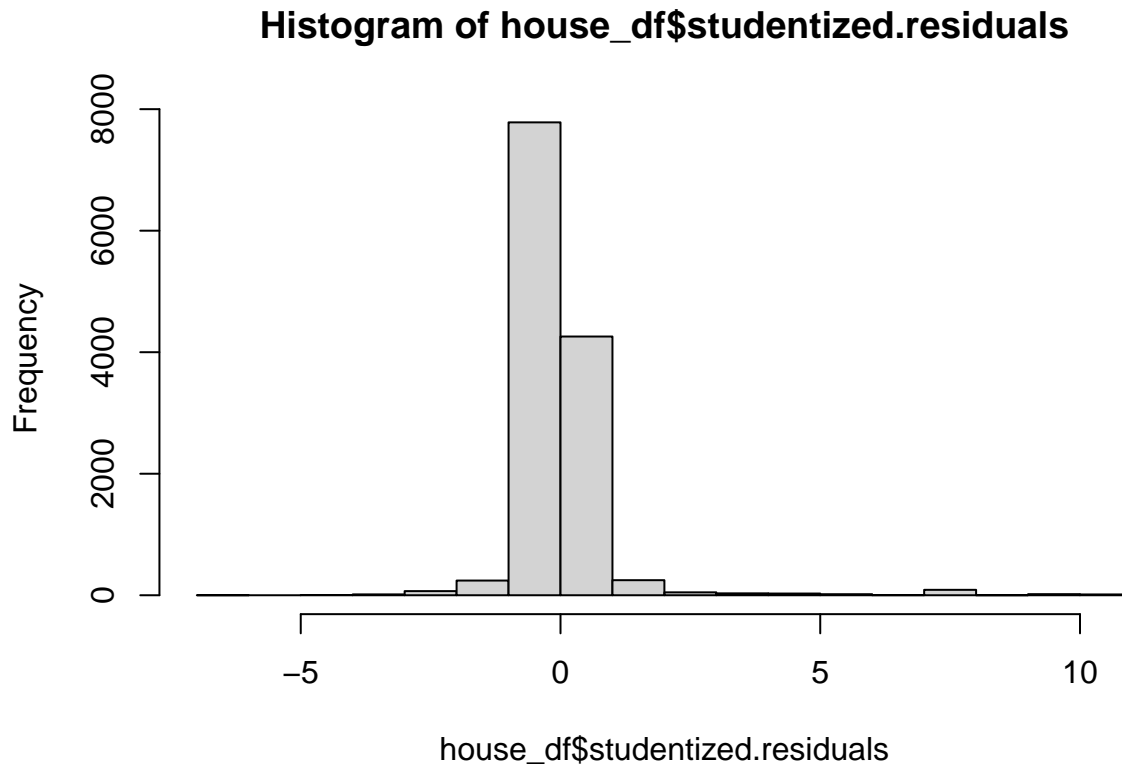


```
scatter <- ggplot(house_df, aes(fitted, studentized.residuals))
scatter + geom_point() + geom_smooth(method = "lm", colour = "Blue") + labs(x = "Fitted Values", y = "Studentized Residuals")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



```
hist(house_df$studentized.residuals)
```



Question 6.1o

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

Answer: There are clearly some outliers and influential cases in this model that may be skewing or at least, slightly deviating it from norm. The model does not violate randomness, heteroscedasticity, linearity, and is has no multicollinearity. Most likely with the removal of some of the outlier and influential cases this model would be fairly unbiased and a good representation of the for the population.

References