Assignment: ASSIGNMENT 8, Housing data analysis

Name: Shekhar , Manish

Date: 2021-05-07

```r
# load readxl library and set directory
# Read the housing data set
library(readxl)
setwd("/Users/mshekhar/Desktop/R Programming/DSC520/stats_for_data_science/stats_for_data_science")
mydata <- read_excel("week-6-housing.xlsx")

# check the structure of data and some basic stats
str(mydata)
```

```
## tibble[,24] [12,865 x 24] (S3: tbl_df/tbl/data.frame)
##  $ Sale Date               : POSIXct[1:12865], format: "2006-01-03" "2006-01-03" ...
##  $ Sale Price              : num [1:12865] 698000 649990 572500 420000 369900 ...
##  $ sale_reason             : num [1:12865] 1 1 1 1 1 1 1 1 1 1 ...
##  $ sale_instrument         : num [1:12865] 3 3 3 3 3 15 3 3 3 3 ...
##  $ sale_warning            : chr [1:12865] NA NA NA NA ...
##  $ sitetype                : chr [1:12865] "R1" "R1" "R1" "R1" ...
##  $ addr_full               : chr [1:12865] "17021 NE 113TH CT" "11927 178TH PL NE" "13315 174TH AVE N
##  $ zip5                    : num [1:12865] 98052 98052 98052 98052 98052 ...
##  $ ctyname                 : chr [1:12865] "REDMOND" "REDMOND" NA "REDMOND" ...
##  $ postalctyn              : chr [1:12865] "REDMOND" "REDMOND" "REDMOND" "REDMOND" ...
##  $ lon                     : num [1:12865] -122 -122 -122 -122 -122 ...
##  $ lat                     : num [1:12865] 47.7 47.7 47.7 47.6 47.7 ...
##  $ building_grade          : num [1:12865] 9 9 8 8 7 7 10 10 9 8 ...
##  $ square_feet_total_living: num [1:12865] 2810 2880 2770 1620 1440 4160 3960 3720 4160 2760 ...
##  $ bedrooms                : num [1:12865] 4 4 4 3 3 4 5 4 4 4 ...
##  $ bath_full_count         : num [1:12865] 2 2 1 1 1 2 3 2 2 1 ...
##  $ bath_half_count         : num [1:12865] 1 0 1 0 0 1 0 1 1 0 ...
##  $ bath_3qtr_count         : num [1:12865] 0 1 1 1 1 1 1 0 1 1 ...
##  $ year_built              : num [1:12865] 2003 2006 1987 1968 1980 ...
##  $ year_renovated          : num [1:12865] 0 0 0 0 0 0 0 0 0 0 ...
##  $ current_zoning          : chr [1:12865] "R4" "R4" "R6" "R4" ...
##  $ sq_ft_lot               : num [1:12865] 6635 5570 8444 9600 7526 ...
##  $ prop_type               : chr [1:12865] "R" "R" "R" "R" ...
##  $ present_use             : num [1:12865] 2 2 2 2 2 2 2 2 2 2 ...
```

```r
summary(mydata)
```

```
##    Sale Date                     Sale Price       sale_reason
##  Min.   :2006-01-03 00:00:00   Min.   :    698   Min.   : 0.00
##  1st Qu.:2008-07-07 00:00:00   1st Qu.: 460000   1st Qu.: 1.00
##  Median :2011-11-17 00:00:00   Median : 593000   Median : 1.00
##  Mean   :2011-07-28 15:07:32   Mean   : 660738   Mean   : 1.55
##  3rd Qu.:2014-06-05 00:00:00   3rd Qu.: 750000   3rd Qu.: 1.00
##  Max.   :2016-12-16 00:00:00   Max.   :4400000   Max.   :19.00
##  sale_instrument  sale_warning         sitetype           addr_full
##  Min.   : 0.000   Length:12865       Length:12865       Length:12865
##  1st Qu.: 3.000   Class :character   Class :character   Class :character
##  Median : 3.000   Mode  :character   Mode  :character   Mode  :character
```

```
##  Mean   : 3.678
##  3rd Qu.: 3.000
##  Max.   :27.000
##       zip5          ctyname           postalctyn             lon
##  Min.   :98052   Length:12865       Length:12865       Min.   :-122.2
##  1st Qu.:98052   Class :character   Class :character   1st Qu.:-122.1
##  Median :98052   Mode  :character   Mode  :character   Median :-122.1
##  Mean   :98053                                         Mean   :-122.1
##  3rd Qu.:98053                                         3rd Qu.:-122.0
##  Max.   :98074                                         Max.   :-121.9
##       lat        building_grade  square_feet_total_living    bedrooms
##  Min.   :47.46   Min.   : 2.00   Min.   :  240             Min.   : 0.000
##  1st Qu.:47.67   1st Qu.: 8.00   1st Qu.: 1820             1st Qu.: 3.000
##  Median :47.69   Median : 8.00   Median : 2420             Median : 4.000
##  Mean   :47.68   Mean   : 8.24   Mean   : 2540             Mean   : 3.479
##  3rd Qu.:47.70   3rd Qu.: 9.00   3rd Qu.: 3110             3rd Qu.: 4.000
##  Max.   :47.73   Max.   :13.00   Max.   :13540             Max.   :11.000
##  bath_full_count  bath_half_count  bath_3qtr_count   year_built
##  Min.   : 0.000   Min.   :0.0000   Min.   :0.000   Min.   :1900
##  1st Qu.: 1.000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:1979
##  Median : 2.000   Median :1.0000   Median :0.000   Median :1998
##  Mean   : 1.798   Mean   :0.6134   Mean   :0.494   Mean   :1993
##  3rd Qu.: 2.000   3rd Qu.:1.0000   3rd Qu.:1.000   3rd Qu.:2007
##  Max.   :23.000   Max.   :8.0000   Max.   :8.000   Max.   :2016
##  year_renovated    current_zoning       sq_ft_lot        prop_type
##  Min.   :   0.00   Length:12865       Min.   :    785   Length:12865
##  1st Qu.:   0.00   Class :character   1st Qu.:   5355   Class :character
##  Median :   0.00   Mode  :character   Median :   7965   Mode  :character
##  Mean   :  26.24                      Mean   :  22229
##  3rd Qu.:   0.00                      3rd Qu.:  12632
##  Max.   :2016.00                      Max.   :1631322
##   present_use
##  Min.   :  0.000
##  1st Qu.:  2.000
##  Median :  2.000
##  Mean   :  6.598
##  3rd Qu.:  2.000
##  Max.   :300.000
```

b.i. Explain any transformations or modifications you made to the dataset

```
## -------------- 1. Changed the column name to remove spaces -----------------
# Change the column names to remove spaces
colnames(mydata)[1] <- "Sale_Date"
colnames(mydata)[2] <- "Sale_Price"

## -------------- 2. Checking for NAs in the data ---------------
apply(mydata, 2, function(x) any(is.na(x) | is.infinite(x)))
```

```
##                Sale_Date                Sale_Price               sale_reason
##                    FALSE                     FALSE                     FALSE
##          sale_instrument              sale_warning                  sitetype
```

```
##                   FALSE                    TRUE                   FALSE
##               addr_full                    zip5                 ctyname
##                   FALSE                   FALSE                    TRUE
##              postalctyn                     lon                     lat
##                   FALSE                   FALSE                   FALSE
##          building_grade square_feet_total_living                bedrooms
##                   FALSE                   FALSE                   FALSE
##         bath_full_count         bath_half_count         bath_3qtr_count
##                   FALSE                   FALSE                   FALSE
##              year_built          year_renovated          current_zoning
##                   FALSE                   FALSE                   FALSE
##                sq_ft_lot               prop_type             present_use
##                   FALSE                   FALSE                   FALSE
```

```r
# check if there are any NAs in sale_warning
sum(is.na(mydata$sale_warning))
```

```
## [1] 10568
```

```r
# check if there are any NAs in ctyname
sum(is.na(mydata$ctyname))
```

```
## [1] 6078
```

```r
# We found that sale_warning = 10568 and ctyname = 6078 have NAs
# sale_warning sounds like a bad remark and if it is not present.
# Deriving sale_warning indicator and adding it to the data
mydata <- cbind(mydata, "Sale_Warning_in" = as.numeric(!(is.na(mydata$sale_warning))))
```

```r
# cityname
# Not worried about this field as we have other attributes like zip code,
# latitude, longitude to predict sale price.
# We can ignore this field while model creation

## ---------------- 3. changing categorical variables to factors ------------------
# changing sitetype to factor
# check distinct values
unique(mydata$sitetype)
```

```
## [1] "R1" "R2" "R3" "DV" "A1" "R4" "C1"
```

```r
mydata$sitetype <- factor(mydata$sitetype, labels = c(1,2,3,4,5,6,7))

# changing current zoning to factor
# check current zoning unique values and convert to factor
unique(mydata$current_zoning)
```

```
##  [1] "R4"      "R6"      "URPSO"   "RA5"     "R3"      "R5"      "RA2.5"
##  [8] "RA10"    "R12"     "RA5P"    "R1"      "RA2.5SO" "RA2.5P"  "R4/C"
## [15] "EH"      "R1P"     "BC"      "R8"      "A10"     "R6/C"    "R18"
## [22] "A10SO"   "RA10P"   "GC"
```

```r
mydata$current_zoning <- factor(mydata$current_zoning, labels = c(1:24))

# check unique values in prop_type
unique(mydata$prop_type)
```

```
## [1] "R"
```

```r
# As all the values are R, really this variable won't be a good predictor.
# Ignoring it

# don't need addr_full in the final data set as we have zip, latitude, longitude
# ignoring cityname and postalctyyn for the same reason
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
# picking only numerical attributes to check correlation to sale_price
hs_data_prcsd_num <- mydata %>% dplyr::select("Sale_Price","sale_reason","Sale_Warning_in",
"sale_instrument","zip5","lon","lat","building_grade",
"square_feet_total_living","bedrooms","bath_full_count",
"bath_half_count","bath_3qtr_count","year_built",
"year_renovated","sq_ft_lot","present_use")
```

```r
# create correlation matrix
cor(hs_data_prcsd_num, method = "pearson", use = "complete.obs")
```

```
##                           Sale_Price  sale_reason Sale_Warning_in
## Sale_Price                1.00000000 -0.116643537     0.083211139
## sale_reason              -0.11664354  1.000000000     0.355583265
## Sale_Warning_in           0.08321114  0.355583265     1.000000000
## sale_instrument          -0.04070601  0.399856649     0.343648065
## zip5                      0.06014866  0.001393069     0.029425761
## lon                       0.04684745 -0.013531540     0.036575404
## lat                       0.02119829 -0.031965818    -0.034812917
## building_grade            0.39122909 -0.078912289    -0.063020490
## square_feet_total_living  0.45458758 -0.065212764    -0.012080335
## bedrooms                  0.22546748 -0.056284214    -0.007526779
## bath_full_count           0.28484899 -0.073127789    -0.015603326
## bath_half_count           0.16582843 -0.051526350    -0.016935167
## bath_3qtr_count           0.03574175 -0.000687221    -0.021751887
## year_built                0.24267127 -0.120238360    -0.052099288
## year_renovated            0.03286429  0.024343886     0.040703002
## sq_ft_lot                 0.11981223  0.034005560     0.086937090
## present_use              -0.02542926 -0.002062742     0.003681498
##                          sale_instrument         zip5         lon         lat
## Sale_Price                  -0.040706012  0.060148664  0.04684745  0.02119829
## sale_reason                  0.399856649  0.001393069 -0.01353154 -0.03196582
## Sale_Warning_in              0.343648065  0.029425761  0.03657540 -0.03481292
## sale_instrument              1.000000000 -0.016103809 -0.02192079 -0.04966602
## zip5                        -0.016103809  1.000000000  0.35337066 -0.13431164
## lon                         -0.021920789  0.353370662  1.00000000 -0.04101493
## lat                         -0.049666024 -0.134311637 -0.04101493  1.00000000
## building_grade              -0.048442696  0.103989081  0.03312593  0.01217953
```

```
## square_feet_total_living    -0.007000713  0.095658773  0.06857118 -0.06383585
## bedrooms                    -0.001834812 -0.036622671 -0.24149623 -0.15481695
## bath_full_count             -0.025098358  0.098244669  0.13139051  0.07757143
## bath_half_count             -0.034397435  0.024996235  0.02574174 -0.04252565
## bath_3qtr_count              0.015641257 -0.074930308 -0.13644355 -0.10044764
## year_built                  -0.096159040  0.100989530  0.34235695  0.37787153
## year_renovated               0.011207272 -0.005762662 -0.02017440 -0.09174116
## sq_ft_lot                    0.124543302  0.089844639  0.20108459 -0.14545622
## present_use                  0.001793940  0.029112870  0.12844704  0.01760738
##                           building_grade square_feet_total_living     bedrooms
## Sale_Price                    0.39122909              0.454587585  0.225467478
## sale_reason                  -0.07891229             -0.065212764 -0.056284214
## Sale_Warning_in              -0.06302049             -0.012080335 -0.007526779
## sale_instrument              -0.04844270             -0.007000713 -0.001834812
## zip5                          0.10398908              0.095658773 -0.036622671
## lon                           0.03312593              0.068571180 -0.241496234
## lat                           0.01217953             -0.063835847 -0.154816948
## building_grade                1.00000000              0.745180518  0.342565614
## square_feet_total_living      0.74518052              1.000000000  0.575347048
## bedrooms                      0.34256561              0.575347048  1.000000000
## bath_full_count               0.45941631              0.517783777  0.304598944
## bath_half_count               0.28076582              0.316856642  0.162334411
## bath_3qtr_count               0.06951439              0.198673964  0.247340383
## year_built                    0.36192968              0.306427289  0.012441913
## year_renovated               -0.01902482              0.043396050  0.014923912
## sq_ft_lot                     0.13843064              0.234104503  0.047173368
## present_use                   0.04472257              0.032493682 -0.038895101
##                           bath_full_count bath_half_count bath_3qtr_count
## Sale_Price                     0.28484899      0.16582843     0.035741748
## sale_reason                   -0.07312779     -0.05152635    -0.000687221
## Sale_Warning_in               -0.01560333     -0.01693517    -0.021751887
## sale_instrument               -0.02509836     -0.03439744     0.015641257
## zip5                           0.09824467      0.02499624    -0.074930308
## lon                            0.13139051      0.02574174    -0.136443549
## lat                            0.07757143     -0.04252565    -0.100447636
## building_grade                 0.45941631      0.28076582     0.069514388
## square_feet_total_living       0.51778378      0.31685664     0.198673964
## bedrooms                       0.30459894      0.16233441     0.247340383
## bath_full_count                1.00000000      0.21731460    -0.379105667
## bath_half_count                0.21731460      1.00000000    -0.315842525
## bath_3qtr_count               -0.37910567     -0.31584253     1.000000000
## year_built                     0.45259966      0.19849009    -0.152354857
## year_renovated                 0.02218095     -0.02665661     0.018358504
## sq_ft_lot                      0.04699255      0.02810463     0.049156430
## present_use                    0.03364950      0.01931299     0.001521253
##                            year_built year_renovated   sq_ft_lot  present_use
## Sale_Price                 0.24267127    0.032864291  0.11981223 -0.025429262
## sale_reason               -0.12023836    0.024343886  0.03400556 -0.002062742
## Sale_Warning_in           -0.05209929    0.040703002  0.08693709  0.003681498
## sale_instrument           -0.09615904    0.011207272  0.12454330  0.001793940
## zip5                       0.10098953   -0.005762662  0.08984464  0.029112870
## lon                        0.34235695   -0.020174395  0.20108459  0.128447040
## lat                        0.37787153   -0.091741159 -0.14545622  0.017607381
## building_grade             0.36192968   -0.019024817  0.13843064  0.044722573
```

```
## square_feet_total_living  0.30642729    0.043396050  0.23410450  0.032493682
## bedrooms                  0.01244191    0.014923912  0.04717337 -0.038895101
## bath_full_count           0.45259966    0.022180945  0.04699255  0.033649501
## bath_half_count           0.19849009   -0.026656615  0.02810463  0.019312989
## bath_3qtr_count          -0.15235486    0.018358504  0.04915643  0.001521253
## year_built                1.00000000   -0.224586183 -0.13491395  0.130783379
## year_renovated           -0.22458618    1.000000000  0.06320824 -0.017558495
## sq_ft_lot                -0.13491395    0.063208244  1.00000000  0.059320852
## present_use               0.13078338   -0.017558495  0.05932085  1.000000000
```

```r
# We can see some of the lowest correlations to sale_price are -
# 1. sale price ~ sale_intrument = -0.040706012
# 2. sale_price ~ zip5 = 0.060148664
# 3. sale_price ~ log = 0.04684745
# 4. sale_price ~ lat = 0.02119829
# 5. sale_price ~ bath_3qtr_count = 0.035741748
# 6. sale_price ~ year_renovated = 0.032864291
# 7. sale_price ~ present_use = -0.025429262
# 8. sale_price ~ sale_warning_in = 0.083211139
```

b. ii. 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.

```r
# creating model with just sale price and square feet of the sq_ft_lot
hs_data_smpl_rgr_mdl <- lm(Sale_Price ~ sq_ft_lot, data = mydata)

# creating data set with several additional predictors of my choice
# My choice is purely based on their correlation
# with predicted variable Sale_Price
# Highest correlations are -
# 1. Sale_Price ~ sale_reason = -0.116643537
# 2. Sale_Price ~ building_grade = 0.39122909
# 3. Sale_Price ~ square_feet_total_living = 0.454587585
# 4. Sale_Price ~ bedrooms = 0.225467478
# 5. Sale_Price ~ bath_full_count = 0.28484899
# 6. Sale_Price ~ bath_half_count = 0.16582843
# 7. Sale_Price ~ year_built = 0.24267127
# 8. Sale_Price ~ sq_ft_lot = 0.11981223
```

```
# Checking correlation of Sale_Price with factor variables - sitetype, current_zoning
# applying Kruskal-Wallis chi-squared test as categorical variable is not
# dichotomous
kruskal.test(mydata$Sale_Price~mydata$sitetype)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  mydata$Sale_Price by mydata$sitetype
## Kruskal-Wallis chi-squared = 387.31, df = 6, p-value < 2.2e-16
```

```
kruskal.test(mydata$Sale_Price~mydata$current_zoning)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  mydata$Sale_Price by mydata$current_zoning
## Kruskal-Wallis chi-squared = 1636.5, df = 23, p-value < 2.2e-16
```

```
# Both the variables are highly correlated to Sale_Price and will include
# them as well as the predictor variables

# Run 1 : creating multiple regression model with sale price and several
# other predictors found relevant based on correlation exercise
hs_data_multi_rgr_mdl <- lm(Sale_Price ~ sale_reason+building_grade+
square_feet_total_living+bedrooms+bath_full_count+bath_half_count+
year_built+sq_ft_lot+sitetype+current_zoning, data = mydata)
```

b. iii. 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?

```
# check summary of simple regression model
summary(hs_data_smpl_rgr_mdl)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sq_ft_lot, data = mydata)
##
## 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
```

```
#Multiple R-squared:  0.01435,  Adjusted R-squared:  0.01428
#F-statistic: 187.3 on 1 and 12863 DF,  p-value: < 2.2e-16

# check summary of multiple regression model
summary(hs_data_multi_rgr_mdl)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sale_reason + building_grade + square_feet_total_living +
##     bedrooms + bath_full_count + bath_half_count + year_built +
##     sq_ft_lot + sitetype + current_zoning, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2045030  -121627   -41643    47525  3657041
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.848e+06  5.584e+05  -3.310 0.000935 ***
## sale_reason              -1.133e+04  1.165e+03  -9.724  < 2e-16 ***
## building_grade            2.596e+04  4.704e+03   5.520 3.46e-08 ***
## square_feet_total_living  1.456e+02  6.064e+00  24.003  < 2e-16 ***
## bedrooms                 -1.906e+04  4.867e+03  -3.917 9.03e-05 ***
## bath_full_count           1.533e+04  6.027e+03   2.544 0.010973 *
## bath_half_count           1.924e+04  6.348e+03   3.031 0.002445 **
## year_built                1.429e+03  2.725e+02   5.243 1.61e-07 ***
## sq_ft_lot                 2.695e-01  7.200e-02   3.743 0.000182 ***
## sitetype2                 8.307e+04  2.380e+05   0.349 0.727110
## sitetype3                -3.429e+05  1.901e+05  -1.804 0.071244 .
## sitetype4                -2.182e+05  1.244e+05  -1.754 0.079421 .
## sitetype5                -3.186e+05  1.257e+05  -2.536 0.011234 *
## sitetype6                -3.619e+05  1.558e+05  -2.323 0.020185 *
## sitetype7                -3.799e+05  3.712e+05  -1.023 0.306113
## current_zoning2           2.891e+06  3.806e+05   7.596 3.26e-14 ***
## current_zoning3          -3.569e+05  3.746e+05  -0.953 0.340767
## current_zoning4          -7.066e+05  1.436e+05  -4.920 8.77e-07 ***
## current_zoning5          -7.278e+05  3.747e+05  -1.943 0.052094 .
## current_zoning6          -6.597e+05  1.368e+05  -4.824 1.42e-06 ***
## current_zoning7          -7.086e+05  1.363e+05  -5.198 2.04e-07 ***
## current_zoning8          -8.296e+05  1.568e+05  -5.290 1.25e-07 ***
## current_zoning9          -7.010e+05  1.881e+05  -3.727 0.000194 ***
## current_zoning10         -6.526e+05  1.354e+05  -4.821 1.44e-06 ***
## current_zoning11         -6.094e+05  1.343e+05  -4.537 5.76e-06 ***
## current_zoning12         -6.255e+05  1.399e+05  -4.472 7.83e-06 ***
```

```
## current_zoning13         -6.927e+05  1.343e+05  -5.158 2.54e-07 ***
## current_zoning14         -7.019e+05  1.343e+05  -5.225 1.77e-07 ***
## current_zoning15         -1.705e+05  1.509e+05  -1.130 0.258483
## current_zoning16         -2.342e+05  1.430e+05  -1.638 0.101544
## current_zoning17         -7.102e+05  1.406e+05  -5.049 4.49e-07 ***
## current_zoning18         -2.589e+05  2.454e+05  -1.055 0.291434
## current_zoning19         -6.443e+05  1.354e+05  -4.758 1.98e-06 ***
## current_zoning20         -7.247e+05  1.669e+05  -4.343 1.42e-05 ***
## current_zoning21         -6.687e+05  1.462e+05  -4.574 4.83e-06 ***
## current_zoning22         -7.133e+05  1.337e+05  -5.336 9.66e-08 ***
## current_zoning23         -7.711e+05  1.399e+05  -5.512 3.61e-08 ***
## current_zoning24         -6.504e+05  1.346e+05  -4.832 1.37e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 349500 on 12827 degrees of freedom
## Multiple R-squared:  0.255,  Adjusted R-squared:  0.2528
## F-statistic: 118.7 on 37 and 12827 DF,  p-value: < 2.2e-16
```

```r
# Run 2 : Recreating multiple regression model by applying backward elimination
# of non significant variables. sitetype appears non-significant
# with p-values > 0.05
hs_data_multi_rgr_mdl_2 <- lm(Sale_Price ~ sale_reason+building_grade+
square_feet_total_living+bedrooms+bath_full_count+bath_half_count+
year_built+sq_ft_lot+current_zoning, data = mydata)
# check summary of multiple regression model
summary(hs_data_multi_rgr_mdl_2)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sale_reason + building_grade + square_feet_total_living +
##     bedrooms + bath_full_count + bath_half_count + year_built +
##     sq_ft_lot + current_zoning, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2058754  -122967   -42250    47356  3661440
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.991e+06  5.425e+05  -3.671 0.000243 ***
## sale_reason              -1.147e+04  1.166e+03  -9.835  < 2e-16 ***
## building_grade            2.609e+04  4.699e+03   5.552 2.87e-08 ***
## square_feet_total_living  1.474e+02  6.057e+00  24.328  < 2e-16 ***
## bedrooms                 -1.721e+04  4.853e+03  -3.545 0.000393 ***
## bath_full_count           1.441e+04  6.027e+03   2.390 0.016863 *
## bath_half_count           1.736e+04  6.341e+03   2.737 0.006205 **
## year_built                1.385e+03  2.721e+02   5.089 3.66e-07 ***
## sq_ft_lot                 2.721e-01  7.176e-02   3.792 0.000150 ***
## current_zoning2           2.888e+06  3.810e+05   7.580 3.68e-14 ***
## current_zoning3          -3.542e+05  3.751e+05  -0.945 0.344921
## current_zoning4          -7.344e+05  1.437e+05  -5.111 3.25e-07 ***
## current_zoning5          -7.237e+05  3.751e+05  -1.929 0.053711 .
## current_zoning6          -6.579e+05  1.369e+05  -4.806 1.56e-06 ***
## current_zoning7          -7.115e+05  1.364e+05  -5.215 1.87e-07 ***
```

```
## current_zoning8          -8.201e+05  1.570e+05  -5.223 1.79e-07 ***
## current_zoning9          -6.988e+05  1.883e+05  -3.711 0.000207 ***
## current_zoning10         -6.510e+05  1.355e+05  -4.804 1.58e-06 ***
## current_zoning11         -6.096e+05  1.345e+05  -4.533 5.86e-06 ***
## current_zoning12         -6.238e+05  1.400e+05  -4.455 8.47e-06 ***
## current_zoning13         -6.950e+05  1.344e+05  -5.170 2.38e-07 ***
## current_zoning14         -6.987e+05  1.345e+05  -5.195 2.08e-07 ***
## current_zoning15         -1.692e+05  1.510e+05  -1.120 0.262663
## current_zoning16         -2.335e+05  1.432e+05  -1.630 0.103021
## current_zoning17         -7.061e+05  1.408e+05  -5.015 5.37e-07 ***
## current_zoning18         -2.580e+05  2.457e+05  -1.050 0.293614
## current_zoning19         -6.412e+05  1.356e+05  -4.730 2.27e-06 ***
## current_zoning20         -7.213e+05  1.671e+05  -4.317 1.59e-05 ***
## current_zoning21         -6.596e+05  1.462e+05  -4.512 6.48e-06 ***
## current_zoning22         -7.120e+05  1.338e+05  -5.320 1.05e-07 ***
## current_zoning23         -7.741e+05  1.400e+05  -5.530 3.27e-08 ***
## current_zoning24         -6.495e+05  1.347e+05  -4.820 1.45e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 350000 on 12833 degrees of freedom
## Multiple R-squared:  0.2528, Adjusted R-squared:  0.251
## F-statistic: 140.1 on 31 and 12833 DF,  p-value: < 2.2e-16
```

```r
#Multiple R-squared:  0.2528,   Adjusted R-squared:  0.251
#F-statistic: 140.1 on 31 and 12833 DF,  p-value: < 2.2e-16

# Run 3 : Recreating multiple regression model by applying backward elimination
# of non significant variables. current_zoning appears non-significant with
# some created variables of it having p-values > 0.05
hs_data_multi_rgr_mdl_3 <- lm(Sale_Price ~ sale_reason+building_grade+
square_feet_total_living+bedrooms+bath_full_count+bath_half_count+year_built+sq_ft_lot, data = mydata)
# check summary of multiple regression model
summary(hs_data_multi_rgr_mdl_3)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sale_reason + building_grade + square_feet_total_living +
##     bedrooms + bath_full_count + bath_half_count + year_built +
##     sq_ft_lot, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2130758  -119409   -44728    41346  3722325
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -4.238e+06  4.339e+05  -9.767  < 2e-16 ***
## sale_reason              -1.163e+04  1.182e+03  -9.844  < 2e-16 ***
## building_grade            2.985e+04  4.446e+03   6.714 1.97e-11 ***
## square_feet_total_living  1.421e+02  5.957e+00  23.847  < 2e-16 ***
## bedrooms                 -8.170e+03  4.593e+03  -1.779   0.0753 .
## bath_full_count           1.384e+04  6.073e+03   2.279   0.0227 *
## bath_half_count           5.876e+03  6.324e+03   0.929   0.3528
## year_built                2.159e+03  2.202e+02   9.806  < 2e-16 ***
```

```
## sq_ft_lot                 2.972e-01  5.888e-02   5.046 4.56e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 355100 on 12856 degrees of freedom
## Multiple R-squared:  0.2295, Adjusted R-squared:  0.2291
## F-statistic: 478.8 on 8 and 12856 DF,  p-value: < 2.2e-16
```

```r
# Intercept has drastically changed. Also p-values for bedrooms and
# bath_half_count have changed and they no more significant and
# have p-values > 0.05

# Run 4 : Recreating multiple regression model by applying backward elimination
# of non significant variables. bedrooms, bath_half_count appears non-significant
# with p-values > 0.05
hs_data_multi_rgr_mdl_4 <- lm(Sale_Price ~ sale_reason+building_grade+
square_feet_total_living+bath_full_count+year_built+sq_ft_lot, data = mydata)

# check summary of multiple regression model
summary(hs_data_multi_rgr_mdl_4)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sale_reason + building_grade + square_feet_total_living +
##     bath_full_count + year_built + sq_ft_lot, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2136788  -119779   -44670    41079  3720748
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -4.477e+06  4.165e+05 -10.749  < 2e-16 ***
## sale_reason              -1.156e+04  1.180e+03  -9.792  < 2e-16 ***
## building_grade            3.106e+04  4.404e+03   7.053 1.84e-12 ***
## square_feet_total_living  1.374e+02  5.083e+00  27.038  < 2e-16 ***
## bath_full_count           1.288e+04  6.040e+03   2.132    0.033 *
## year_built                2.268e+03  2.132e+02  10.640  < 2e-16 ***
## sq_ft_lot                 3.132e-01  5.805e-02   5.395 6.97e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 355100 on 12858 degrees of freedom
## Multiple R-squared:  0.2293, Adjusted R-squared:  0.2289
## F-statistic: 637.6 on 6 and 12858 DF,  p-value: < 2.2e-16
```

```r
# Intercept seems to have stabilized with highly significant p-value
# All remaining predictor variables are highly significant with p-values < 0.05
# F-statistic has significantly improved with highly significant p-value
# Residual standard error: 355100 on 12858 degrees of freedom
# Multiple R-squared:  0.2293,  Adjusted R-squared:  0.2289
# F-statistic: 637.6 on 6 and 12858 DF,  p-value: < 2.2e-16
```

Comparing simple regression to multi regression model

1. Multiple R-squared (variation explained) - In simple regression sq_ft_lot explains only 1.4% of variation in Sale_Price while in final multi regression model 4 predictors together explains almost 23% variation in Sale_Price.
2. Adjusted R-squared - Difference between adjusted R-squared and R-squared in multi regression model is 0.0004 while same difference in simple linear regression is 0.001 which is higher. In ideal world we would like adjusted R-squared to be same as R-squared to say that model can be more genralized and represent population.
3. P-value of predictors - P-values for each predictor in both simple regression and multi regression are highly significant and almost all of them are well below critical value of 0.05.
4. F-statistic - Fscore of multi regression model is almost 4.5 times that of simple regression model. P-value of both models is highly significant. This means that R-squared of multi regression model is more significant than that of simple regression model.
Overall multiple regression model is a much better predictive model for Sale_Price

b. iv. 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?

Absolute values of beta coeffients of multi regression model are comparatively higher and showing higher strength of the effect of each individual independent variable to the dependent variable. Most of the coefficients are positive showing positive effect on Sale_price.
This mean that unit change in square_feet_total_living will cause Sale_Price to go up by 137.4 units, while keeping effect of all other predictors constant. We can say that effect of sq_feet_tital_living is more compared to sq_ft_lot on Sale_Price when then are seen individually keep other predictors constant. p-values associated with t values of precitors are very significant and well below cut-off value of 0.05 and in most case below 0.01 which would they will be significant in more than 99% cases. With t value of 27.038 and p-value < 0.01

square_feet_total_living is most significant precitor of the model.

```
summary(hs_data_multi_rgr_mdl_4)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sale_reason + building_grade + square_feet_total_living +
##     bath_full_count + year_built + sq_ft_lot, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2136788  -119779   -44670    41079  3720748
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -4.477e+06  4.165e+05 -10.749  < 2e-16 ***
## sale_reason              -1.156e+04  1.180e+03  -9.792  < 2e-16 ***
## building_grade            3.106e+04  4.404e+03   7.053 1.84e-12 ***
## square_feet_total_living  1.374e+02  5.083e+00  27.038  < 2e-16 ***
## bath_full_count           1.288e+04  6.040e+03   2.132    0.033 *
## year_built                2.268e+03  2.132e+02  10.640  < 2e-16 ***
## sq_ft_lot                 3.132e-01  5.805e-02   5.395 6.97e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 355100 on 12858 degrees of freedom
## Multiple R-squared:  0.2293, Adjusted R-squared:  0.2289
## F-statistic: 637.6 on 6 and 12858 DF,  p-value: < 2.2e-16
```

Absolute value of beta coeffient of simple regression model is comparatively lower i.e. 0.85 and thus showing weaker strength of the effect of predictor variable to predicted variable. Coefficient is positive showing positive relationship. One unit change in sq_ft_lot will cause 0.85 units of change in Sale_Price. p-value for sq_ft_lot is highly significant and well below 0.01 which means in 99%+ cases this effect will stant true.

```
summary(hs_data_smpl_rgr_mdl)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sq_ft_lot, data = mydata)
##
## 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
```

We can also get standardized beta estimates using lm.beta() of QuantPsyc package.

```
library(QuantPsyc)
```

```
## Loading required package: boot

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

##
## Attaching package: 'QuantPsyc'

## The following object is masked from 'package:base':
##
##     norm
```

```
lm.beta(hs_data_multi_rgr_mdl_4)
```

```
##              sale_reason           building_grade square_feet_total_living
##              -0.07645294               0.08392564               0.33641067
##          bath_full_count               year_built                sq_ft_lot
##               0.02072611               0.09658661               0.04409583
```

As standardized beta values are calculated in standard deviation units for all predictors they can be easily compared to each other. We can see that for square__feet__total__living standard beta value is 0.33 (one standard deviation unit increase in square__feet__total__living will cause 0.33 standard deviation unit increase in Sale Price) which is much higher than any other predictor's value and thus it indicates highest effect of square__feet__total__living on Sale__Price.

b v. Calculate the confidence intervals for the parameters in your model and

explain what the results indicate.

```
# calculating confidence interval for the parameters of simple regression model
confint(hs_data_smpl_rgr_mdl)
```

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

```r
# calculating confidence interval for the parameters of multiple regression model 4
confint(hs_data_multi_rgr_mdl_4)
```

```
##                                  2.5 %        97.5 %
## (Intercept)              -5.293246e+06 -3.660485e+06
## sale_reason              -1.386989e+04 -9.243201e+03
## building_grade            2.242855e+04  3.969337e+04
## square_feet_total_living  1.274738e+02  1.474013e+02
## bath_full_count           1.038854e+03  2.471804e+04
## year_built                1.850308e+03  2.685978e+03
## sq_ft_lot                 1.994069e-01  4.269936e-01
```

In simple linear regression model confidence interval is comparatively less wide and thus sample b value of predictor sq_ft_lot is more representative of being closer to beta value in population. In multi regression model, confidence intervals for few variables like bath_full_count, sale_reason, and even building_grade are wider and thus we are comparatively lesser confident about the mean of the future values. It is generally suggested to increase the sample size if CI is too wide, which doesn't seems to be the case in either model. In either model confidence interval does not cross zero for any of the predictors and thus it is indicative that relationships strength is good between predictors and predcited variables and thus model will not be predicting positive in some samples while negative in other samples because of particular predictor.

b. vi. 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.

```r
# Running ANOVA for simple regression model
aov(Sale_Price ~ sq_ft_lot, data = mydata)
```

```
## Call:
##    aov(formula = Sale_Price ~ sq_ft_lot, data = mydata)
##
## Terms:
##                    sq_ft_lot    Residuals
## Sum of Squares  3.019674e+13 2.073377e+15
## Deg. of Freedom            1        12863
##
## Residual standard error: 401483.8
## Estimated effects may be unbalanced
```

```r
# Running ANOVA for multiple regression model
aov(Sale_Price ~ sale_reason+building_grade+
square_feet_total_living+bath_full_count+year_built+sq_ft_lot, data = mydata)
```

```
## Call:
##    aov(formula = Sale_Price ~ sale_reason + building_grade + square_feet_total_living +
##     bath_full_count + year_built + sq_ft_lot, data = mydata)
##
## Terms:
##                  sale_reason building_grade square_feet_total_living
## Sum of Squares  2.862062e+13   3.089249e+14             1.249171e+14
## Deg. of Freedom            1              1                        1
##                 bath_full_count   year_built    sq_ft_lot    Residuals
## Sum of Squares     4.206324e+12 1.202280e+13 3.669902e+12 1.621212e+15
## Deg. of Freedom               1            1            1        12858
##
## Residual standard error: 355085.7
## Estimated effects may be unbalanced
```

We can observe that Residual standard error in case of simple linear regression is higher than that of multi regression model and thus there is more unexplained variation in data by simple regression model predictor that multi regression model predictors together. In other words multi regression model is better model to explain variation in the data.

b. vii. Perform casewise diagnostics to identify outliers and/or influential cases,

storing each function's output in a dataframe assigned to a unique variable name.

Statistics to identify outliers in the data by checking residuals using resid(), standard residuals using rstandard() and studentized residual using rstudent().

Identifying influencial cases -

1. Cook's distance - Access overall influence of a single case on the model.

Cook and Weisberg (1982) have suggested that values greater than 1 may be cause for concern. If a point is a significant outlier on Y, but its Cook's distance is $<$ 1, there is no real need to delete that point since it does not have a large effect on the regression analysis. Using cooks.distance()

2. Hat value (leverage) - Value of 0 would mean that case has no influence, while value of 1 would indicate that case has complete influence over prediction. If no cases exert undue influence over the model then we would expect all of the leverage values to be close to the average value ((k + 1)/ n). Using hatvalues().

3. DFFit & DFbeta - It is the difference between the predicted value for a

case when the model is fitted including that case and when the model is fitted excluding that case. Using dffits() and DFBeta using dfbeta().

adding variables to check each of above case level statistics to original data frames.

```
# adding case level residual/outlier and influencial stats for multi regression model
mydata$residuals <- resid(hs_data_multi_rgr_mdl_4)
mydata$standardized.residuals <- rstandard(hs_data_multi_rgr_mdl_4)
mydata$studentized.residuals <- rstudent(hs_data_multi_rgr_mdl_4)
mydata$cooks.distance <- cooks.distance(hs_data_multi_rgr_mdl_4)
mydata$dfbeta <- dfbeta(hs_data_multi_rgr_mdl_4)
mydata$dffit <- dffits(hs_data_multi_rgr_mdl_4)
mydata$leverage <- hatvalues(hs_data_multi_rgr_mdl_4)
mydata$covariance.ratios <- covratio(hs_data_multi_rgr_mdl_4)
# looking at the data
# mydata
# writing the saved stats for each case into a table
write.table(mydata, "House Sales With Diagnostics.dat", sep = "\t", row.names = FALSE)
# nrow(mydata) -- 12865
# check if about 5% of cases (<= 643 cases) have standardized residual within +-2.
sum(mydata$standardized.residuals > 2 | mydata$standardized.residuals < -2)
```

```
## [1] 329
```

```
# As only 329 cases are outside range or have large residuals, we are well within the range of 5% outli

# To exactly identify outliers we can add a variable called large.residual in the data frame to save th
mydata$large.residual <- mydata$standardized.residuals > 2 | mydata$standardized.residuals < -2
# we can now select the outlier cases by select rows with large.residual = TRUE
# model is fairly accurate

# check how many cases have standard residuals > 3 which may be we can investigate further
sum(mydata$standardized.residuals > 3 | mydata$standardized.residuals < -3)
```

```
## [1] 213
```

```
# create a variable to flag cases with very large residual
mydata$very.large.residual <- mydata$standardized.residuals > 3 | mydata$standardized.residuals < -3
# 213 cases out of 12865 -- about 1.65%

# Let's look at the leverage (hat value), cook's distance, and covariance ratio
# for cases with large.residual = TRUE
mydata[mydata$large.residual, c("cooks.distance","leverage","covariance.ratios")]
```

```
##       cooks.distance     leverage covariance.ratios
## 25      0.0010257923 0.0011869228         0.9984429
## 108     0.0033044011 0.0042136523         1.0017920
## 115     0.0024117590 0.0018063670         0.9972755
## 178     0.0011465389 0.0012854582         0.9984365
## 239     0.0001090031 0.0001668056         0.9982225
## 246     0.0031564325 0.0017464582         0.9954241
## 287     0.0006893490 0.0007933254         0.9980305
## 295     0.2396277578 0.1188300220         1.1278061
## 300     0.0075267399 0.0043854071         0.9984257
```

```
## 341      0.0004487567 0.0006708699         0.9986685
## 344      0.0057351762 0.0089305205         1.0071144
## 359      0.0017397504 0.0014731451         0.9975264
## 385      0.0025307845 0.0015599477         0.9959382
## 396      0.0014805940 0.0014632775         0.9981591
## 475      0.0006986847 0.0009934765         0.9988611
## 482      0.0031738517 0.0039360295         1.0014278
## 508      0.0033052073 0.0053950353         1.0036382
## 528      0.0010875245 0.0016248737         0.9996238
## 576      0.0008319101 0.0013155085         0.9994538
## 661      0.0027968265 0.0028437415         0.9996540
## 670      0.0016660191 0.0005997724         0.9906016
## 679      0.0070450303 0.0036509393         0.9968763
## 742      0.0004484261 0.0007724219         0.9991065
## 784      0.0019524886 0.0023320386         0.9996953
## 802      0.0010311108 0.0015000453         0.9994297
## 811      0.0035835198 0.0009199438         0.9867072
## 852      0.0009102800 0.0015231355         0.9997943
## 853      0.0004455858 0.0007137136         0.9988811
## 877      0.0016204219 0.0014373881         0.9976943
## 916      0.0045527934 0.0030400158         0.9978991
## 1009     0.0002877654 0.0004295533         0.9984230
## 1099     0.0009422438 0.0015341461         0.9997424
## 1119     0.0020112581 0.0015651144         0.9972234
## 1142     0.0011810063 0.0010314525         0.9972200
## 1155     0.0149292319 0.0095001048         1.0041643
## 1305     0.0011346071 0.0016966438         0.9996980
## 1345     0.0002542520 0.0004161686         0.9986340
## 1368     0.0007075141 0.0011555917         0.9993700
## 1380     0.0003101548 0.0004637808         0.9984617
## 1442     0.0016275507 0.0015502173         0.9981018
## 1492     0.0010091685 0.0012623284         0.9987649
## 1504     0.0048822856 0.0030902960         0.9976377
## 1543     0.0010641713 0.0017134567         0.9998962
## 1550     0.0008218085 0.0008187770         0.9975437
## 1633     0.0001621965 0.0002170445         0.9979159
## 1650     0.0135138750 0.0062910557         0.9987168
## 1716     0.0009970945 0.0007971874         0.9965835
## 1745     0.0009432943 0.0007656760         0.9966233
## 1870     0.0074126080 0.0066798036         1.0030493
## 1962     0.0024740915 0.0014260536         0.9953770
## 1963     0.0006106903 0.0007478570         0.9981840
## 1964     0.0008404656 0.0004465255         0.9938367
## 1976     0.0013242658 0.0003997737         0.9883823
## 1977     0.0006853792 0.0001884214         0.9869458
## 1978     0.0012447634 0.0003547031         0.9875953
## 1979     0.0006774109 0.0001869084         0.9869939
## 1980     0.0012740900 0.0003552299         0.9873042
## 1981     0.0014775095 0.0004630181         0.9889040
## 1982     0.0012863954 0.0003793806         0.9880669
## 2022     0.0013424594 0.0009537856         0.9961454
## 2099     0.0005363480 0.0006876495         0.9982629
## 2137     0.0005589791 0.0003740237         0.9952344
## 2157     0.0005565648 0.0005217858         0.9970069
```

```
## 2264    0.0031592360 0.0015138722          0.9941316
## 2302    0.0017626301 0.0024168498          1.0001911
## 2360    0.0005459138 0.0005608369          0.9974005
## 2361    0.0024410539 0.0009953756          0.9922280
## 2469    0.0004473775 0.0003714912          0.9963336
## 2684    0.0015598153 0.0004735320          0.9885261
## 2685    0.0012316856 0.0003774574          0.9885464
## 2686    0.0014329427 0.0005141746          0.9904813
## 2687    0.0012141040 0.0003613572          0.9881655
## 2688    0.0011725802 0.0003451790          0.9880091
## 2689    0.0013120914 0.0004579103          0.9901284
## 2690    0.0013092871 0.0004699203          0.9904396
## 2699    0.0340776069 0.0254130949          1.0215325
## 2708    0.0013991091 0.0008301737          0.9949675
## 2709    0.0012136735 0.0008456224          0.9959317
## 2710    0.0017277029 0.0010943362          0.9956365
## 2717    0.0030168291 0.0015203003          0.9945262
## 2742    0.0005096100 0.0004644780          0.9968336
## 2852    0.0323551464 0.0084297003          0.9945063
## 2934    0.0009961450 0.0007971874          0.9965880
## 2937    0.0009613303 0.0007971874          0.9967538
## 3097    0.0026676183 0.0014175848          0.9948117
## 3102    0.0007357039 0.0006850037          0.9971425
## 3110    0.0008237369 0.0008555478          0.9977359
## 3111    0.0003968700 0.0005785581          0.9985112
## 3168    0.0039875726 0.0004821071          0.9699125
## 3169    0.0033779971 0.0004123887          0.9701391
## 3170    0.0029576959 0.0003738115          0.9711367
## 3171    0.0027292290 0.0003492111          0.9714720
## 3172    0.0027537800 0.0003513317          0.9713897
## 3173    0.0027317318 0.0003496169          0.9714795
## 3174    0.0027533928 0.0003512709          0.9713887
## 3175    0.0030548353 0.0003796849          0.9706469
## 3176    0.0027749988 0.0003587433          0.9717792
## 3177    0.0026791022 0.0003464883          0.9717788
## 3178    0.0026799098 0.0003466282          0.9717818
## 3179    0.0028631251 0.0003611646          0.9710675
## 3180    0.0026810696 0.0003468243          0.9717859
## 3181    0.0028623150 0.0003610450          0.9710660
## 3182    0.0039884401 0.0004822319          0.9699139
## 3183    0.0039868755 0.0004820066          0.9699114
## 3184    0.0039868755 0.0004820066          0.9699114
## 3185    0.0039885428 0.0004822466          0.9699141
## 3186    0.0039875726 0.0004821071          0.9699125
## 3187    0.0039875726 0.0004821071          0.9699125
## 3188    0.0033853118 0.0004134073          0.9701493
## 3189    0.0030536527 0.0003795151          0.9706449
## 3190    0.0027516005 0.0003509865          0.9713838
## 3191    0.0030531822 0.0003794474          0.9706441
## 3192    0.0027294619 0.0003492496          0.9714728
## 3193    0.0027512095 0.0003509237          0.9713827
## 3194    0.0027509545 0.0003508825          0.9713819
## 3195    0.0026794947 0.0003465567          0.9717803
## 3196    0.0026803440 0.0003467022          0.9717834
```

```
## 3197     0.0027812390 0.0003597503          0.9717962
## 3198     0.0039902426 0.0004824902          0.9699168
## 3199     0.0039900914 0.0004824686          0.9699165
## 3200     0.0039868755 0.0004820066          0.9699114
## 3201     0.0039868755 0.0004820066          0.9699114
## 3202     0.0039868755 0.0004820066          0.9699114
## 3424     0.0024495202 0.0040292686          1.0022766
## 3464     0.0029958323 0.0003794015          0.9712020
## 3465     0.0028920607 0.0003723198          0.9716726
## 3466     0.0026276922 0.0003464232          0.9723251
## 3467     0.0026294974 0.0003467421          0.9723320
## 3468     0.0029970621 0.0003795821          0.9712041
## 3469     0.0028044725 0.0003604398          0.9716123
## 3470     0.0026272739 0.0003463465          0.9723233
## 3471     0.0028929496 0.0003724558          0.9716744
## 3472     0.0026776797 0.0003492414          0.9720233
## 3473     0.0027005098 0.0003511464          0.9719381
## 3474     0.0029999885 0.0003800094          0.9712089
## 3475     0.0026801737 0.0003496534          0.9720308
## 3476     0.0037772871 0.0004924125          0.9721523
## 3477     0.0026793022 0.0003495111          0.9720283
## 3478     0.0039134029 0.0004820066          0.9704772
## 3479     0.0039134029 0.0004820066          0.9704772
## 3480     0.0039134029 0.0004820066          0.9704772
## 3481     0.0039134029 0.0004820066          0.9704772
## 3482     0.0039134029 0.0004820066          0.9704772
## 3483     0.0039134076 0.0004820073          0.9704772
## 3484     0.0033146796 0.0004122789          0.9707011
## 3485     0.0028051885 0.0003710987          0.9724468
## 3486     0.0028040584 0.0003603760          0.9716114
## 3487     0.0027138345 0.0003583941          0.9723855
## 3488     0.0026989172 0.0003508860          0.9719335
## 3489     0.0026767519 0.0003490836          0.9720201
## 3490     0.0027005680 0.0003511558          0.9719383
## 3491     0.0037826809 0.0004935759          0.9721801
## 3492     0.0028922054 0.0003723420          0.9716729
## 3493     0.0026280542 0.0003464886          0.9723266
## 3494     0.0029973689 0.0003796270          0.9712046
## 3495     0.0039134076 0.0004820073          0.9704772
## 3496     0.0039134076 0.0004820073          0.9704772
## 3497     0.0039134076 0.0004820073          0.9704772
## 3523     0.0026507975 0.0015849983          0.9957735
## 3837     0.0006838252 0.0007383763          0.9977580
## 3918     0.0012625486 0.0015651144          0.9990412
## 3919     0.0007652759 0.0010314525          0.9987521
## 4055     0.0014804592 0.0011854653          0.9969800
## 4056     0.0045005009 0.0013619354          0.9893770
## 4248     0.0007189969 0.0012086223          0.9994892
## 4391     0.0011440636 0.0019973401          1.0003650
## 4435     0.0003501356 0.0004602009          0.9981078
## 4648     0.1052345743 0.0102109306          0.9722188
## 4649     0.6906373336 0.0441092700          0.9884534
## 4671     0.0033942707 0.0042189274          1.0017201
## 4695     0.0015209988 0.0004222043          0.9873116
```

```
## 4696    0.0012415938 0.0008215703           0.9956192
## 4740    0.0003991572 0.0006797623           0.9989884
## 4821    0.0018888427 0.0015127881           0.9973094
## 4834    0.0011454391 0.0009537856           0.9969297
## 4840    0.0156021509 0.0095001048           1.0038956
## 4934    0.0009536065 0.0009580187           0.9977148
## 5083    0.0901980470 0.0309037562           1.0213738
## 5491    0.0001204817 0.0001827619           0.9982166
## 5494    0.0001184603 0.0001828554           0.9982600
## 5495    0.0001210330 0.0001833950           0.9982144
## 5496    0.0001205418 0.0001828299           0.9982163
## 5497    0.0002495714 0.0003894789           0.9984937
## 5549    0.0007267978 0.0011075487           0.9991545
## 5935    0.0013552866 0.0010634637           0.9967606
## 6055    0.0009548636 0.0013018514           0.9990555
## 6230    0.0016927822 0.0003674489           0.9834783
## 6231    0.0016557276 0.0003716195           0.9840513
## 6232    0.0018284484 0.0004261427           0.9847252
## 6233    0.0018688703 0.0003973082           0.9831440
## 6234    0.0016795075 0.0003655666           0.9835237
## 6235    0.0016921418 0.0003673263           0.9834790
## 6236    0.0016791110 0.0003654903           0.9835241
## 6237    0.0017228537 0.0003748257           0.9835249
## 6238    0.0019715318 0.0004140669           0.9829447
## 6239    0.0016560923 0.0003716927           0.9840510
## 6429    0.0051922484 0.0003579187           0.9468824
## 6430    0.0028656656 0.0001849280           0.9431183
## 6431    0.0034505351 0.0002283094           0.9445494
## 6432    0.0040677930 0.0002746190           0.9456881
## 6433    0.0044944781 0.0003068287           0.9463174
## 6434    0.0048150696 0.0003295367           0.9464729
## 6435    0.0087269903 0.0006216813           0.9488530
## 6436    0.0035389281 0.0002355183           0.9448734
## 6437    0.0028910251 0.0001879459           0.9435341
## 6438    0.0030345974 0.0001959032           0.9431503
## 6439    0.0044687097 0.0003045868           0.9462308
## 6440    0.0050615201 0.0003607200           0.9486138
## 6441    0.0072029810 0.0004873128           0.9460152
## 6442    0.0044465560 0.0003022794           0.9460878
## 6443    0.0062548245 0.0004833333           0.9527154
## 6444    0.0065752820 0.0005214943           0.9539694
## 6445    0.0066645597 0.0004997116           0.9512863
## 6446    0.0063454767 0.0004937736           0.9530550
## 6447    0.0071497350 0.0005789957           0.9549787
## 6448    0.0026806822 0.0002002402           0.9508024
## 6449    0.0059395065 0.0004822235           0.9549975
## 6450    0.0077410779 0.0006442085           0.9562607
## 6451    0.0065804276 0.0004930996           0.9512497
## 6452    0.0063582774 0.0004955632           0.9531320
## 6453    0.0076861191 0.0006311932           0.9556587
## 6454    0.0067255502 0.0005369915           0.9542924
## 6455    0.0062439332 0.0004799401           0.9524606
## 6456    0.0029151760 0.0002198457           0.9512864
## 6457    0.0047098882 0.0003741964           0.9539020
```

```
## 6512      0.0027687042 0.0013074985           0.9938082
## 6527      0.0011439540 0.0014878729           0.9991076
## 6739      0.0010574400 0.0010579402           0.9977993
## 6796      0.0002679505 0.0004499813           0.9987267
## 6938      0.0008439725 0.0003966984           0.9928578
## 6939      0.0012674214 0.0006937685           0.9942942
## 6940      0.0005499315 0.0002296020           0.9916774
## 6941      0.0005484610 0.0002296106           0.9917020
## 6942      0.0011257892 0.0005881841           0.9938582
## 6943      0.0021717622 0.0011535140           0.9945433
## 6944      0.0015445694 0.0008612923           0.9945893
## 6945      0.0008439725 0.0003966984           0.9928578
## 6946      0.0005520202 0.0002309316           0.9916967
## 6947      0.0005282048 0.0002164612           0.9914939
## 6948      0.0005193341 0.0002109403           0.9914059
## 7039      0.0008794222 0.0013133248           0.9993099
## 7147      0.0008071568 0.0008465522           0.9977625
## 7167      0.0040556743 0.0007040698           0.9794912
## 7210      0.0034382281 0.0013317753           0.9920764
## 7211      0.0049454782 0.0015271078           0.9897918
## 7446      0.0019642756 0.0002300190           0.9686643
## 7447      0.0019171957 0.0002216273           0.9682447
## 7448      0.0019661194 0.0002303729           0.9686836
## 7449      0.0019179172 0.0002216866           0.9682413
## 7450      0.0019172881 0.0002215034           0.9682251
## 7451      0.0019167196 0.0002213281           0.9682090
## 7452      0.0019168567 0.0002216008           0.9682465
## 7453      0.0018489967 0.0002157551           0.9685378
## 7454      0.0018541697 0.0002165786           0.9685709
## 7455      0.0018138062 0.0002095564           0.9682144
## 7456      0.0019172478 0.0002216315           0.9682445
## 7457      0.0019590073 0.0002291654           0.9686307
## 7458      0.0019156108 0.0002212414           0.9682150
## 7459      0.0019175169 0.0002199960           0.9679999
## 7460      0.0019172881 0.0002215034           0.9682251
## 7461      0.0019152975 0.0002203961           0.9680963
## 7462      0.0019160039 0.0002201497           0.9680481
## 7463      0.0019152776 0.0002203946           0.9680964
## 7507      0.0023158182 0.0024123317           0.9993094
## 7649      0.0006917948 0.0011398730           0.9993746
## 7650      0.0006918818 0.0011402519           0.9993754
## 7683      0.0103781570 0.0159671542           1.0143040
## 7791      0.0010461800 0.0015224008           0.9994529
## 7871      0.0046273592 0.0030157643           0.9977352
## 8154      0.0002961755 0.0004439867           0.9984482
## 8232      0.0024292806 0.0023320386           0.9989182
## 8262      0.0253133456 0.0181771335           1.0137703
## 8320      0.0064410706 0.0050596919           1.0007889
## 8377      0.3807445906 0.0644221375           1.0471077
## 8457      0.0005906999 0.0007478570           0.9982856
## 8458      0.0008222984 0.0004465255           0.9939909
## 8535      0.0003569741 0.0005152238           0.9984215
## 8541      0.0001631968 0.0002170445           0.9978983
## 8698      0.0002796865 0.0003719032           0.9980529
```

```
## 8710     0.0022473554 0.0007041959         0.9891443
## 8763     0.0005216706 0.0006359458         0.9980577
## 8887     0.0129194672 0.0027836422         0.9857741
## 8911     0.0013009607 0.0019755686         1.0000169
## 8946     0.0013009607 0.0019755686         1.0000169
## 9215     0.0024188561 0.0026122144         0.9996397
## 9293     0.0011160547 0.0016248737         0.9995570
## 9369     0.0013616896 0.0021917458         1.0003758
## 9420     0.0060749032 0.0073714500         1.0048367
## 9453     0.0005106974 0.0007971874         0.9989027
## 9528     0.0056161190 0.0022635893         0.9933926
## 9546     0.0006538835 0.0007383763         0.9979122
## 9722     0.0007277701 0.0006797623         0.9971501
## 10125    0.0052324495 0.0048478089         1.0013104
## 10318    0.0133567782 0.0224122117         1.0212129
## 10371    0.0021064306 0.0028288247         1.0005472
## 10418    0.0015299893 0.0023617801         1.0004458
## 10478    0.0004579563 0.0007820097         0.9990968
## 10623    0.0004269388 0.0006254808         0.9985711
## 10707    0.0001299349 0.0002014687         0.9982897
## 10723    0.0004312082 0.0006878271         0.9988453
## 10741    0.0002559881 0.0004202566         0.9986450
## 10787    0.0027445767 0.0014033335         0.9945172
## 10844    0.0016949044 0.0021391104         0.9996722
## 10958    0.0004796424 0.0004712638         0.9971420
## 10995    0.0030409613 0.0051210076         1.0034327
## 11165    0.0013206888 0.0014774990         0.9986218
## 11289    0.0006334948 0.0002389100         0.9907170
## 11413    0.0027417091 0.0012059380         0.9931160
## 11558    0.0042468752 0.0009956524         0.9853908
## 11586    0.0016574128 0.0011102733         0.9959786
## 11728    0.0068990834 0.0047846693         0.9998698
## 11758    0.0022656737 0.0028288247         1.0003331
## 11772    0.0120241217 0.0018168689         0.9774022
## 11822    0.0025447045 0.0006655231         0.9867224
## 11898    0.0327527392 0.0348098650         1.0330472
## 11899    0.0767419624 0.0375479305         1.0318106
## 11982    0.0086379247 0.0030400158         0.9928114
## 11992    0.0732808662 0.0046287603         0.9463640
## 12212    0.0020701946 0.0019228612         0.9983743
## 12255    0.0041240793 0.0022774808         0.9959444
## 12256    0.0011712201 0.0011821940         0.9979574
## 12392    0.0008750702 0.0010714059         0.9985077
## 12472    0.0042021297 0.0013192567         0.9897842
## 12487    0.0110948252 0.0046009420         0.9960107
## 12577    0.0039696302 0.0014696028         0.9917631
## 12582    0.0186385515 0.0024805710         0.9747313
## 12643    0.0160063267 0.0019543570         0.9716895
## 12686    0.0001473875 0.0002401013         0.9984466
## 12764    0.0054974616 0.0021949561         0.9932346
```

```r
# check if any outlier cases have cook's distance > 1
sum(mydata[mydata$large.residual, c("cooks.distance")] > 1)
```

```
## [1] 0
```

```r
# None of the cases have cooks distance > 1, so none of the cases have undue
# influence on the model

# calculate average leverage using formula = (k+1/n)
avg_leverage <- (6+1)/12865
# three times average leverage
times_3_leverage <- avg_leverage*3
# check if there are outlier cases with levarage > 3 times the average leverage
sum(mydata[mydata$large.residual, c("levarage")] > times_3_leverage)
```

```
## [1] 0
```

```r
# There are none
```

b. viii. Calculate the standardized residuals using the appropriate command,

specifying those that are +-2, storing the results of large residuals in a

variable you create.

Answered this question as part of analyis in last question. Below is the

command used. Variable is already added to the data frame mydata

- mydata$large.residual < -mydata$**standardized.residuals > 2 | mydata$standardized.residuals < -2 -**

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

Answered in b. vii. as part of analysis of residual and influence.

```r
# nrow(mydata) -- 12865
# check if about 5% of cases (<= 643 cases) have standardized residual within +-2.
sum(mydata$standardized.residuals > 2 | mydata$standardized.residuals < -2)
```

```
## [1] 329
```

```r
# As only 329 cases are outside range or have large residuals, we are well
# within the range of 5% outliers.
```

b. x. Which specific variables have large residuals (only cases that evaluate

as TRUE)?

```r
mydata[mydata$large.residual == TRUE, c("Sale_Date","Sale_Price","sale_reason",
                                        "sale_instrument","sale_warning",
                                        "sitetype","addr_full","zip5",
```

```
                                          "ctyname","lon","lat","building_grade",
                                          "square_feet_total_living","bedrooms",
                                          "bath_full_count","year_built","year_renovated",
                                          "current_zoning","sq_ft_lot")]
```

```
##         Sale_Date Sale_Price sale_reason sale_instrument sale_warning sitetype
## 25     2006-01-11     265000           1               3         <NA>        4
## 108    2006-02-13    1520000          18               3           52        4
## 115    2006-02-15    1390000           1               3         <NA>        4
## 178    2006-03-03     390000           1               3         <NA>        4
## 239    2006-03-20    1588359           1               3         <NA>        4
## 246    2006-03-21    1450000           1               3         <NA>        4
## 287    2006-03-27     163000           1               3           49        6
## 295    2006-03-28     270000           1               3         <NA>        4
## 300    2006-03-29     200000           1               3         <NA>        4
## 341    2006-04-06     300000           1               3           56        4
## 344    2006-04-06      90000           1              26         <NA>        4
## 359    2006-04-10     187000           1               3         <NA>        4
## 385    2006-04-17    2500000           1               3         <NA>        4
## 396    2006-04-19    2169000           1               3         <NA>        4
## 475    2006-05-08    1534000           1               3         <NA>        4
## 482    2006-05-08     555000           1               3         <NA>        4
## 508    2006-05-15      65000          18              15        18 49        4
## 528    2006-05-18     435000           1               3         <NA>        4
## 576    2006-05-30     333000           1               3         <NA>        4
## 661    2006-06-19    2569000           1               3         <NA>        4
## 670    2006-06-20    2583000           1               3           45        4
## 679    2006-06-20     350000           1               3           49        4
## 742    2006-06-29     385000           1               3         <NA>        4
## 784    2006-07-12     325000           1               3           35        4
## 802    2006-07-14     340000           1               3           35        4
## 811    2006-07-19    3000000           1               3         <NA>        4
## 852    2006-07-26     250000           1               3           45        4
## 853    2006-07-27    1595000           1               3           45        4
## 877    2006-08-01     157000           1               3         <NA>        4
## 916    2006-08-10     450000           1               3           45        4
## 1009   2006-08-28     175000           1               3         <NA>        4
## 1099   2006-09-19     350000           1               3         <NA>        4
## 1119   2006-09-22     279150           1               3           49        4
## 1142   2006-09-27     279150           1               3           56        4
## 1155   2006-10-02      32000           1              26         <NA>        4
## 1305   2006-11-10    1316000           1               3         <NA>        4
## 1345   2006-11-17     233333           1               3         <NA>        4
## 1368   2006-11-27     375000           1               3         <NA>        4
## 1380   2006-12-01      80000           1               3         <NA>        4
## 1442   2006-12-20    1085000           1               3         <NA>        4
## 1492   2007-01-05     285000           1               3           56        4
## 1504   2007-01-09     452800           1               3         <NA>        4
## 1543   2007-01-30     375000           1               3           49        6
## 1550   2007-02-01     276450           1               3           56        4
## 1633   2007-02-27       5000           1               3           16        4
## 1650   2007-03-02     349999           1               3         <NA>        4
## 1716   2007-03-20       1000           1              15        18 51        4
## 1745   2007-03-29    2235000           1               3         <NA>        4
```

```
## 1870  2007-05-08  2988000       1       3    45    4
## 1962  2007-06-01  2493000       1       3  <NA>    4
## 1963  2007-06-01  1600000       1       3  <NA>    4
## 1964  2007-06-01  1600000       1       3  <NA>    4
## 1976  2007-06-05  2625000       1       3    45    4
## 1977  2007-06-05  2625000       1       3    45    4
## 1978  2007-06-05  2625000       1       3    45    4
## 1979  2007-06-05  2625000       1       3    45    4
## 1980  2007-06-05  2625000       1       3    45    4
## 1981  2007-06-05  2625000       1       3    45    4
## 1982  2007-06-05  2625000       1       3    45    4
## 2022  2007-06-12  1384950       1       3    45    4
## 2099  2007-06-22  1975000       1       3  <NA>    4
## 2137  2007-07-02  1787000       1       3  <NA>    4
## 2157  2007-07-05  1678000       1       3  <NA>    4
## 2264  2007-08-01  1651000       1       3  <NA>    4
## 2302  2007-08-08   450000       1       3  <NA>    4
## 2360  2007-08-22  1640000       1       3  <NA>    4
## 2361  2007-08-22  1640000       1       3  <NA>    4
## 2469  2007-09-25  1710000       1       3  <NA>    4
## 2684  2007-12-12  2590000       1       3    45    4
## 2685  2007-12-12  2590000       1       3    45    4
## 2686  2007-12-12  2590000       1       3    45    4
## 2687  2007-12-12  2590000       1       3    45    4
## 2688  2007-12-12  2590000       1       3    45    4
## 2689  2007-12-12  2590000       1       3    45    4
## 2690  2007-12-12  2590000       1       3    45    4
## 2699  2007-12-13   379950       1       3  <NA>    4
## 2708  2007-12-17  2300000       1       3    45    4
## 2709  2007-12-17  2300000       1       3    45    4
## 2710  2007-12-17  2300000       1       3    45    4
## 2717  2007-12-18  2598000       1       3  <NA>    4
## 2742  2007-12-27    10570       1       3  <NA>    4
## 2852  2008-02-26  3995000       1       3  <NA>    4
## 2934  2008-03-31     1500       1       3    51    4
## 2937  2008-04-01    20000       1       3    24    4
## 3097  2008-06-06  2549000       1       3  <NA>    4
## 3102  2008-06-09  2080000       1       3  <NA>    4
## 3110  2008-06-11  1250000       1       3    26    4
## 3111  2008-06-11  1250000       1       3    26    4
## 3168  2008-07-01  3175000       1       3    45    4
## 3169  2008-07-01  3175000       1       3    45    4
## 3170  2008-07-01  3175000       1       3    45    4
## 3171  2008-07-01  3175000       1       3    45    4
## 3172  2008-07-01  3175000       1       3    45    4
## 3173  2008-07-01  3175000       1       3    45    4
## 3174  2008-07-01  3175000       1       3    45    4
## 3175  2008-07-01  3175000       1       3    45    4
## 3176  2008-07-01  3175000       1       3    45    4
## 3177  2008-07-01  3175000       1       3    45    4
## 3178  2008-07-01  3175000       1       3    45    4
## 3179  2008-07-01  3175000       1       3    45    4
## 3180  2008-07-01  3175000       1       3    45    4
## 3181  2008-07-01  3175000       1       3    45    4
```

```
## 3182  2008-07-01  3175000   1  3  45  4
## 3183  2008-07-01  3175000   1  3  45  4
## 3184  2008-07-01  3175000   1  3  45  4
## 3185  2008-07-01  3175000   1  3  45  4
## 3186  2008-07-01  3175000   1  3  45  4
## 3187  2008-07-01  3175000   1  3  45  4
## 3188  2008-07-01  3175000   1  3  45  4
## 3189  2008-07-01  3175000   1  3  45  4
## 3190  2008-07-01  3175000   1  3  45  4
## 3191  2008-07-01  3175000   1  3  45  4
## 3192  2008-07-01  3175000   1  3  45  4
## 3193  2008-07-01  3175000   1  3  45  4
## 3194  2008-07-01  3175000   1  3  45  4
## 3195  2008-07-01  3175000   1  3  45  4
## 3196  2008-07-01  3175000   1  3  45  4
## 3197  2008-07-01  3175000   1  3  45  4
## 3198  2008-07-01  3175000   1  3  45  4
## 3199  2008-07-01  3175000   1  3  45  4
## 3200  2008-07-01  3175000   1  3  45  4
## 3201  2008-07-01  3175000   1  3  45  4
## 3202  2008-07-01  3175000   1  3  45  4
## 3424  2008-09-16    20146  18  3   5  4
## 3464  2008-10-01  3150000   1  3  45  4
## 3465  2008-10-01  3150000   1  3  45  4
## 3466  2008-10-01  3150000   1  3  45  4
## 3467  2008-10-01  3150000   1  3  45  4
## 3468  2008-10-01  3150000   1  3  45  4
## 3469  2008-10-01  3150000   1  3  45  4
## 3470  2008-10-01  3150000   1  3  45  4
## 3471  2008-10-01  3150000   1  3  45  4
## 3472  2008-10-01  3150000   1  3  45  4
## 3473  2008-10-01  3150000   1  3  45  4
## 3474  2008-10-01  3150000   1  3  45  4
## 3475  2008-10-01  3150000   1  3  45  4
## 3476  2008-10-01  3150000   1  3  45  4
## 3477  2008-10-01  3150000   1  3  45  4
## 3478  2008-10-01  3150000   1  3  45  4
## 3479  2008-10-01  3150000   1  3  45  4
## 3480  2008-10-01  3150000   1  3  45  4
## 3481  2008-10-01  3150000   1  3  45  4
## 3482  2008-10-01  3150000   1  3  45  4
## 3483  2008-10-01  3150000   1  3  45  4
## 3484  2008-10-01  3150000   1  3  45  4
## 3485  2008-10-01  3150000   1  3  45  4
## 3486  2008-10-01  3150000   1  3  45  4
## 3487  2008-10-01  3150000   1  3  45  4
## 3488  2008-10-01  3150000   1  3  45  4
## 3489  2008-10-01  3150000   1  3  45  4
## 3490  2008-10-01  3150000   1  3  45  4
## 3491  2008-10-01  3150000   1  3  45  4
## 3492  2008-10-01  3150000   1  3  45  4
## 3493  2008-10-01  3150000   1  3  45  4
## 3494  2008-10-01  3150000   1  3  45  4
## 3495  2008-10-01  3150000   1  3  45  4
```

```
## 3496  2008-10-01  3150000       1       3      45     4
## 3497  2008-10-01  3150000       1       3      45     4
## 3523  2008-10-09  2000000       1       3      45     4
## 3837  2009-04-21   130000       1       3    <NA>     3
## 3918  2009-05-27   500000       1       3      46     4
## 3919  2009-05-27   475000       1       3      46     4
## 4055  2009-07-21  2000000       1       3    <NA>     4
## 4056  2009-07-21  2000000       1       3    <NA>     4
## 4248  2009-09-24     1500       1      26      24     4
## 4391  2009-11-16   125000       4      18   23 31     4
## 4435  2009-12-01   188750       1       3      10     4
## 4648  2010-03-02  4400000       1       3   35 45     4
## 4649  2010-03-02  4400000       1       3   35 45     4
## 4671  2010-03-09   800000       1      22      15     4
## 4695  2010-03-16  2300000       1       3      60     4
## 4696  2010-03-16  2300000       1       3      60     4
## 4740  2010-04-05  1448000       1       3    <NA>     4
## 4821  2010-04-27  1430000       8       3   12 45     4
## 4834  2010-04-29  1300000       1       3   15 45     4
## 4840  2010-04-30     6000       1      26      24     4
## 4934  2010-05-25  1780000       1      22    <NA>     4
## 5083  2010-07-06      698       1      26      24     4
## 5491  2010-12-17  1550000       1       4      45     4
## 5494  2010-12-17  1550000       1       4      45     4
## 5495  2010-12-17  1550000       1       4      45     4
## 5496  2010-12-17  1550000       1       4      45     4
## 5497  2010-12-17  1550000       1       4      45     4
## 5549  2011-01-11   241000       1       3    <NA>     4
## 5935  2011-06-08   180000       1       3    <NA>     4
## 6055  2011-07-19   130000       1       3   26 49     4
## 6230  2011-09-13  2885000       1      22   15 45     4
## 6231  2011-09-13  2885000       1      22   15 45     4
## 6232  2011-09-13  2885000       1      22   15 45     4
## 6233  2011-09-13  2885000       1      22   15 45     4
## 6234  2011-09-13  2885000       1      22   15 45     4
## 6235  2011-09-13  2885000       1      22   15 45     4
## 6236  2011-09-13  2885000       1      22   15 45     4
## 6237  2011-09-13  2885000       1      22   15 45     4
## 6238  2011-09-13  2885000       1      22   15 45     4
## 6239  2011-09-13  2885000       1      22   15 45     4
## 6429  2011-11-17  4380542       1      22   11 45     4
## 6430  2011-11-17  4380542       1      22   11 45     4
## 6431  2011-11-17  4380542       1      22   11 45     4
## 6432  2011-11-17  4380542       1      22   11 45     4
## 6433  2011-11-17  4380542       1      22   11 45     4
## 6434  2011-11-17  4380542       1      22   11 45     4
## 6435  2011-11-17  4380542       1      22   11 45     4
## 6436  2011-11-17  4380542       1      22   11 45     4
## 6437  2011-11-17  4380542       1      22   11 45     4
## 6438  2011-11-17  4380542       1      22   11 45     4
## 6439  2011-11-17  4380542       1      22   11 45     4
## 6440  2011-11-17  4380542       1      22   11 45     4
## 6441  2011-11-17  4380542       1      22   11 45     4
## 6442  2011-11-17  4380542       1      22   11 45     4
```

```
## 6443  2011-11-17   4140203         1            3        11 45         4
## 6444  2011-11-17   4140203         1            3        11 45         4
## 6445  2011-11-17   4140203         1            3        11 45         4
## 6446  2011-11-17   4140203         1            3        11 45         4
## 6447  2011-11-17   4140203         1            3        11 45         4
## 6448  2011-11-17   4140203         1            3        11 45         4
## 6449  2011-11-17   4140203         1            3        11 45         4
## 6450  2011-11-17   4140203         1            3        11 45         4
## 6451  2011-11-17   4140203         1            3        11 45         4
## 6452  2011-11-17   4140203         1            3        11 45         4
## 6453  2011-11-17   4140203         1            3        11 45         4
## 6454  2011-11-17   4140203         1            3        11 45         4
## 6455  2011-11-17   4140203         1            3        11 45         4
## 6456  2011-11-17   4140203         1            3        11 45         4
## 6457  2011-11-17   4140203         1            3        11 45         4
## 6512  2011-12-08   2500000         1            3           60         4
## 6527  2011-12-09    275000         1            3           10         4
## 6739  2012-03-19     50000         1            3           15         4
## 6796  2012-04-06    226610         1            4         <NA>         4
## 6938  2012-05-21   2300000         1            3        45 56         4
## 6939  2012-05-21   2300000         1            3        45 56         4
## 6940  2012-05-21   2300000         1            3        45 56         4
## 6941  2012-05-21   2300000         1            3        45 56         4
## 6942  2012-05-21   2300000         1            3        45 56         4
## 6943  2012-05-21   2300000         1            3        45 56         4
## 6944  2012-05-21   2300000         1            3        45 56         4
## 6945  2012-05-21   2300000         1            3        45 56         4
## 6946  2012-05-21   2300000         1            3        45 56         4
## 6947  2012-05-21   2300000         1            3        45 56         4
## 6948  2012-05-21   2300000         1            3        45 56         4
## 7039  2012-06-18   1920000         1            3           45         4
## 7147  2012-07-12    254000         1            3         <NA>         4
## 7167  2012-07-20   3200000         1            3           15         4
## 7210  2012-07-30   2500000         1            3           45         4
## 7211  2012-07-30   2500000         1            3           45         4
## 7446  2012-10-02   3462000         1            3           45         4
## 7447  2012-10-02   3462000         1            3           45         4
## 7448  2012-10-02   3462000         1            3           45         4
## 7449  2012-10-02   3462000         1            3           45         4
## 7450  2012-10-02   3462000         1            3           45         4
## 7451  2012-10-02   3462000         1            3           45         4
## 7452  2012-10-02   3462000         1            3           45         4
## 7453  2012-10-02   3462000         1            3           45         4
## 7454  2012-10-02   3462000         1            3           45         4
## 7455  2012-10-02   3462000         1            3           45         4
## 7456  2012-10-02   3462000         1            3           45         4
## 7457  2012-10-02   3462000         1            3           45         4
## 7458  2012-10-02   3462000         1            3           45         4
## 7459  2012-10-02   3462000         1            3           45         4
## 7460  2012-10-02   3462000         1            3           45         4
## 7461  2012-10-02   3462000         1            3           45         4
## 7462  2012-10-02   3462000         1            3           45         4
## 7463  2012-10-02   3462000         1            3           45         4
## 7507  2012-10-12    229687         4           18     10 15 31         4
```

```
## 7649  2012-11-20      280000           1               3           45        4
## 7650  2012-11-20      280000           1               3           45        4
## 7683  2012-11-30     3000000           4              18        15 31        4
## 7791  2012-12-26      525400           1               3         <NA>        4
## 7871  2013-02-05      347000           1               3           10        4
## 8154  2013-04-29      192000           1               3         <NA>        4
## 8232  2013-05-14      226000           1               4        10 35        4
## 8262  2013-05-21     1300000           1              22         <NA>        4
## 8320  2013-06-03      302500           1               3         <NA>        4
## 8377  2013-06-13       14000           1              26         <NA>        4
## 8457  2013-07-02     1586000           1               3         <NA>        4
## 8458  2013-07-02     1586000           1               3         <NA>        4
## 8535  2013-07-16      245000           1               3           10        4
## 8541  2013-07-17        2500           1              26         <NA>        4
## 8698  2013-08-16     1700000           1               3           26        4
## 8710  2013-08-20     2750000           1               3         <NA>        4
## 8763  2013-08-29     1650000           1               3         <NA>        4
## 8887  2013-09-25     3340000           1               3         <NA>        4
## 8911  2013-10-01     2160200           1               3           40        4
## 8946  2013-10-10     2160200           1               4           41        4
## 9215  2013-12-31     1401322           1               3        35 45        4
## 9293  2014-02-13      425000           1               3         <NA>        4
## 9369  2014-03-24      150000           1               3         <NA>        4
## 9420  2014-04-11      320000           1               3         <NA>        4
## 9453  2014-04-18      299500           1               3         <NA>        4
## 9528  2014-05-07        8000           1              26           43        4
## 9546  2014-05-12      150000           1               3         <NA>        3
## 9722  2014-06-23     1700000           1               3         <NA>        4
## 10125 2014-09-30     2200000          18               3           52        4
## 10318 2014-11-25     2140000           1               3         <NA>        4
## 10371 2014-12-12      299000           1               3         <NA>        4
## 10418 2014-12-29      325000           1               3  15 22 46 51        4
## 10478 2015-01-28      365000           1               3         <NA>        4
## 10623 2015-03-25      321000           1               3         <NA>        4
## 10707 2015-04-23     1442500           1               3         <NA>        4
## 10723 2015-04-27     1850000           1               3         <NA>        4
## 10741 2015-04-28      285000           1               3         <NA>        4
## 10787 2015-05-11     2200000           1               3         <NA>        4
## 10844 2015-05-21     1200000           1               3           10        4
## 10958 2015-06-16       18000           1              26         <NA>        4
## 10995 2015-06-24     1050000           1               3        46 60        4
## 11165 2015-07-30     1890000           1               3         <NA>        4
## 11289 2015-08-26     2300000           1               3         <NA>        4
## 11413 2015-09-28     2025000           1               3        12 49        4
## 11558 2015-11-12     2150000           1               3           45        4
## 11586 2015-11-19     1700000           1               3         <NA>        4
## 11728 2016-01-19     1840000           1               3         <NA>        4
## 11758 2016-01-29      269000           1               3         <NA>        4
## 11772 2016-02-05     3750000           1               3         <NA>        4
## 11822 2016-02-25     2850000           1               3         <NA>        4
## 11898 2016-03-29     2165000           1               3         <NA>        4
## 11899 2016-03-29     2165000           1               3         <NA>        4
## 11982 2016-04-25     3175000           1               3           45        4
## 11992 2016-04-27     4311000           1               3         <NA>        4
```

```
## 12212 2016-06-22    1200000          1             3         <NA>        4
## 12255 2016-06-30    1730000          1             3         <NA>        4
## 12256 2016-06-30    1730000          1             3         <NA>        4
## 12392 2016-08-01    1450000          1             3         <NA>        4
## 12472 2016-08-23    2350000          1             3         <NA>        4
## 12487 2016-08-26    2200000         18             3           45        4
## 12577 2016-09-19    2700000          1             3         <NA>        4
## 12582 2016-09-20    3950000          1             3         <NA>        4
## 12643 2016-10-05    3850000          1             3         <NA>        4
## 12686 2016-10-14     148200          1            15         <NA>        4
## 12764 2016-11-08    2988000          1             3         <NA>        4
##                             addr_full  zip5   ctyname       lon      lat
## 25             25149 NE PATTERSON WAY 98053      <NA> -122.0032 47.65814
## 108     3230 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0972 47.63926
## 115                19656 NE REDMOND RD 98053      <NA> -122.0772 47.69595
## 178   13414 WOODINVILLE REDMOND RD NE 98052      <NA> -122.1343 47.72058
## 239                  24103 NE 122ND ST 98053      <NA> -122.0192 47.70719
## 246     3068 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0953 47.63820
## 287                   28218 NE 40TH ST 98053      <NA> -121.9611 47.64355
## 295                   5806 249TH CT NE 98053      <NA> -122.0053 47.65706
## 300                  8015 228TH AVE NE 98053      <NA> -122.0364 47.67546
## 341                  11650 154TH PL NE 98052   REDMOND -122.1328 47.70418
## 344                  1520 268TH AVE NE 98053      <NA> -121.9787 47.62110
## 359                  3131 269TH AVE NE 98053      <NA> -121.9790 47.63687
## 385                   20424 NE 64TH PL 98053      <NA> -122.0685 47.66419
## 396                  8024 255TH AVE NE 98053      <NA> -121.9978 47.67461
## 475     2454 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0879 47.63280
## 482                   3624 175TH CT NE 98052   REDMOND -122.1056 47.64296
## 508                  7717 252ND AVE NE 98053      <NA> -122.0033 47.67279
## 528     2829 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0929 47.63521
## 576                   25525 NE 67TH PL 98053      <NA> -121.9982 47.66396
## 661                  8027 255TH AVE NE 98053      <NA> -122.0000 47.67470
## 670     2612 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0889 47.63375
## 679            24300 NE UNION HILL RD 98053      <NA> -122.0140 47.66268
## 742                  6920 242ND AVE NE 98053      <NA> -122.0151 47.66757
## 784                  4224 244TH AVE NE 98053      <NA> -122.0130 47.64795
## 802                  4402 244TH AVE NE 98053      <NA> -122.0131 47.64924
## 811     2438 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0883 47.63227
## 852                   24805 NE 80TH ST 98053      <NA> -122.0080 47.67429
## 853                  24629 NE 109TH ST 98053      <NA> -122.0092 47.69466
## 877                  8303 250TH AVE NE 98053      <NA> -122.0064 47.67672
## 916                   7350 259TH PL NE 98053      <NA> -121.9927 47.66975
## 1009                 5235 279TH AVE NE 98053      <NA> -121.9675 47.65376
## 1099                  19736 NE 61ST PL 98053      <NA> -122.0757 47.66093
## 1119                 7515 238TH AVE NE 98053      <NA> -122.0238 47.67172
## 1142                 7533 238TH AVE NE 98053      <NA> -122.0236 47.67214
## 1155                 8700 196TH AVE NE 98053      <NA> -122.0782 47.67993
## 1305    2458 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0883 47.63285
## 1345                  16618 NE 47TH ST 98052   REDMOND -122.1184 47.65191
## 1368                   27935 E MAIN ST 98053      <NA> -121.9663 47.60549
## 1380                 3605 289TH AVE NE 98053      <NA> -121.9524 47.64088
## 1442            3122 W AMES LAKE DR NE 98053      <NA> -121.9623 47.63550
## 1492               11820 204TH AVE NE 98053      <NA> -122.0675 47.70476
## 1504                 25509 NE 100TH ST 98053      <NA> -121.9986 47.68701
```

```
## 1543                6616 247TH AVE NE 98053      <NA> -122.0087 47.66505
## 1550               11926 204TH AVE NE 98053      <NA> -122.0675 47.70566
## 1633                 8490 138TH LN NE 98052   REDMOND -122.1554 47.67857
## 1650  14338 WOODINVILLE REDMOND RD NE 98052      <NA> -122.1381 47.73097
## 1716                 22340 NE 65TH PL 98053      <NA> -122.0407 47.66467
## 1745                 23812 NE 61ST ST 98053      <NA> -122.0212 47.66215
## 1870                4130 244TH AVE NE 98053      <NA> -122.0108 47.64601
## 1962                8407 255TH AVE NE 98053      <NA> -122.0000 47.67678
## 1963    2030 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0877 47.62828
## 1964    2030 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0877 47.62828
## 1976                 14026 NE 85TH CT 98052   REDMOND -122.1526 47.67923
## 1977                 14009 NE 85TH CT 98052   REDMOND -122.1528 47.67900
## 1978                 14007 NE 85TH CT 98052   REDMOND -122.1532 47.67900
## 1979                 14005 NE 85TH CT 98052   REDMOND -122.1532 47.67926
## 1980                 14006 NE 85TH CT 98052   REDMOND -122.1532 47.67940
## 1981                 14018 NE 85TH CT 98052   REDMOND -122.1527 47.67959
## 1982                 14022 NE 85TH CT 98052   REDMOND -122.1527 47.67941
## 2022    2140 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0874 47.62941
## 2099                 5838 246TH PL NE 98053      <NA> -122.0097 47.65938
## 2137    3272 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0986 47.64015
## 2157    2240 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0875 47.63050
## 2264    2632 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0893 47.63440
## 2302                 18446 NE 95TH ST 98052   REDMOND -122.0926 47.68585
## 2360    2350 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0878 47.63100
## 2361    2350 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0878 47.63100
## 2469    2768 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0914 47.63484
## 2684                 5215 157TH CT NE 98052   REDMOND -122.1315 47.65483
## 2685                 5221 157TH CT NE 98052   REDMOND -122.1315 47.65503
## 2686                 5216 157TH CT NE 98052   REDMOND -122.1311 47.65488
## 2687                 5209 157TH CT NE 98052   REDMOND -122.1315 47.65469
## 2688                 5103 157TH CT NE 98052   REDMOND -122.1315 47.65451
## 2689                 5210 157TH CT NE 98052   REDMOND -122.1311 47.65473
## 2690                 5104 157TH CT NE 98052   REDMOND -122.1311 47.65456
## 2699                6415 196TH AVE NE 98053      <NA> -122.0798 47.66254
## 2708                9233 255TH AVE NE 98053      <NA> -122.0012 47.68269
## 2709                9113 258TH AVE NE 98053      <NA> -121.9957 47.68123
## 2710                9429 255TH AVE NE 98053      <NA> -122.0019 47.68349
## 2717                8212 255TH AVE NE 98053      <NA> -121.9980 47.67561
## 2742                 5826 249TH CT NE 98053      <NA> -122.0057 47.65873
## 2852               13707 160TH AVE NE 98052      <NA> -122.1272 47.72393
## 2934                 22340 NE 65TH PL 98053      <NA> -122.0407 47.66467
## 2937                 22340 NE 65TH PL 98053      <NA> -122.0407 47.66467
## 3097                8406 255TH AVE NE 98053      <NA> -121.9980 47.67671
## 3102                15335 NE 106TH CT 98052   REDMOND -122.1350 47.69462
## 3110    2524 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0886 47.63300
## 3111    2524 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0886 47.63300
## 3168                23541 NE 111TH ST 98053      <NA> -122.0252 47.69681
## 3169                23525 NE 111TH ST 98053      <NA> -122.0254 47.69680
## 3170             11070 ELLISTON WAY NE 98053      <NA> -122.0254 47.69661
## 3171             11030 ELLISTON WAY NE 98053      <NA> -122.0251 47.69601
## 3172             11019 ELLISTON WAY NE 98053      <NA> -122.0257 47.69573
## 3173             11027 ELLISTON WAY NE 98053      <NA> -122.0257 47.69581
## 3174             11031 ELLISTON WAY NE 98053      <NA> -122.0257 47.69590
## 3175             11035 ELLISTON WAY NE 98053      <NA> -122.0257 47.69601
```

```
## 3176             11043 ELLISTON WAY NE 98053      <NA> -122.0258 47.69611
## 3177             11049 ELLISTON WAY NE 98053      <NA> -122.0258 47.69623
## 3178             11061 ELLISTON WAY NE 98053      <NA> -122.0259 47.69643
## 3179             11073 ELLISTON WAY NE 98053      <NA> -122.0260 47.69664
## 3180             11087 ELLISTON WAY NE 98053      <NA> -122.0261 47.69690
## 3181             11093 ELLISTON WAY NE 98053      <NA> -122.0260 47.69704
## 3182                 23518 NE 111TH ST 98053      <NA> -122.0255 47.69716
## 3183                 23526 NE 111TH ST 98053      <NA> -122.0254 47.69716
## 3184                 23558 NE 111TH ST 98053      <NA> -122.0249 47.69715
## 3185                 23553 NE 111TH ST 98053      <NA> -122.0249 47.69680
## 3186                 23549 NE 111TH ST 98053      <NA> -122.0250 47.69680
## 3187                 23533 NE 111TH ST 98053      <NA> -122.0253 47.69680
## 3188                 23517 NE 111TH ST 98053      <NA> -122.0255 47.69681
## 3189             11062 ELLISTON WAY NE 98053      <NA> -122.0253 47.69648
## 3190             11054 ELLISTON WAY NE 98053      <NA> -122.0252 47.69639
## 3191             11048 ELLISTON WAY NE 98053      <NA> -122.0252 47.69630
## 3192             11042 ELLISTON WAY NE 98053      <NA> -122.0252 47.69619
## 3193             11038 ELLISTON WAY NE 98053      <NA> -122.0252 47.69610
## 3194             11022 ELLISTON WAY NE 98053      <NA> -122.0251 47.69591
## 3195             11057 ELLISTON WAY NE 98053      <NA> -122.0259 47.69633
## 3196             11067 ELLISTON WAY NE 98053      <NA> -122.0260 47.69654
## 3197             11079 ELLISTON WAY NE 98053      <NA> -122.0261 47.69675
## 3198                 23502 NE 111TH ST 98053      <NA> -122.0258 47.69710
## 3199                 23510 NE 111TH ST 98053      <NA> -122.0256 47.69714
## 3200                 23534 NE 111TH ST 98053      <NA> -122.0253 47.69715
## 3201                 23542 NE 111TH ST 98053      <NA> -122.0252 47.69715
## 3202                 23550 NE 111TH ST 98053      <NA> -122.0250 47.69716
## 3424                  16771 NE 86TH CT 98052   REDMOND -122.1172 47.67879
## 3464                 11044 236TH PL NE 98053      <NA> -122.0241 47.69626
## 3465                 11020 236TH PL NE 98053      <NA> -122.0240 47.69591
## 3466                 11004 236TH PL NE 98053      <NA> -122.0240 47.69569
## 3467                 11005 236TH PL NE 98053      <NA> -122.0246 47.69567
## 3468                 11013 236TH PL NE 98053      <NA> -122.0246 47.69579
## 3469                 11021 236TH PL NE 98053      <NA> -122.0246 47.69590
## 3470                 11029 236TH PL NE 98053      <NA> -122.0247 47.69600
## 3471                 11037 236TH PL NE 98053      <NA> -122.0247 47.69609
## 3472                 11045 236TH PL NE 98053      <NA> -122.0247 47.69621
## 3473                 11053 236TH PL NE 98053      <NA> -122.0248 47.69631
## 3474                 11061 236TH PL NE 98053      <NA> -122.0248 47.69642
## 3475                 11077 236TH PL NE 98053      <NA> -122.0248 47.69664
## 3476             11001 ELLISTON WAY NE 98053      <NA> -122.0255 47.69545
## 3477             11011 ELLISTON WAY NE 98053      <NA> -122.0256 47.69562
## 3478                 23566 NE 111TH ST 98053      <NA> -122.0248 47.69715
## 3479                 23574 NE 111TH ST 98053      <NA> -122.0247 47.69716
## 3480                 23600 NE 111TH ST 98053      <NA> -122.0245 47.69715
## 3481                 23608 NE 111TH ST 98053      <NA> -122.0244 47.69715
## 3482                 23616 NE 111TH ST 98053      <NA> -122.0243 47.69716
## 3483                 23632 NE 111TH ST 98053      <NA> -122.0241 47.69716
## 3484                 23656 NE 111TH ST 98053      <NA> -122.0237 47.69716
## 3485                 11092 236TH PL NE 98053      <NA> -122.0242 47.69686
## 3486                 11052 236TH PL NE 98053      <NA> -122.0242 47.69633
## 3487                 11036 236TH PL NE 98053      <NA> -122.0241 47.69612
## 3488                 11028 236TH PL NE 98053      <NA> -122.0241 47.69601
## 3489                 11012 236TH PL NE 98053      <NA> -122.0240 47.69582
```

```
## 3490                 11069 236TH PL NE 98053      <NA> -122.0248 47.69653
## 3491             11018 ELLISTON WAY NE 98053      <NA> -122.0251 47.69581
## 3492             11014 ELLISTON WAY NE 98053      <NA> -122.0250 47.69571
## 3493             11006 ELLISTON WAY NE 98053      <NA> -122.0250 47.69558
## 3494             11007 ELLISTON WAY NE 98053      <NA> -122.0256 47.69554
## 3495                 23624 NE 111TH ST 98053      <NA> -122.0242 47.69716
## 3496                 23640 NE 111TH ST 98053      <NA> -122.0240 47.69715
## 3497                 23648 NE 111TH ST 98053      <NA> -122.0238 47.69716
## 3523    2400 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0879 47.63122
## 3837                 22465 NE 111TH LN 98053      <NA> -122.0384 47.69732
## 3918                 7515 238TH AVE NE 98053      <NA> -122.0238 47.67172
## 3919                 7533 238TH AVE NE 98053      <NA> -122.0236 47.67214
## 4055    2844 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0925 47.63591
## 4056    2844 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0925 47.63591
## 4248                 5106 272ND AVE NE 98053      <NA> -121.9751 47.65169
## 4391                  5920 216TH PL NE 98053      <NA> -122.0503 47.66073
## 4435                  15808 NE 51ST ST 98052   REDMOND -122.1297 47.65474
## 4648                 12025 154TH PL NE 98052      <NA> -122.1350 47.70801
## 4649                 12053 154TH PL NE 98052      <NA> -122.1345 47.70950
## 4671                 6910 264TH AVE NE 98053      <NA> -121.9852 47.66578
## 4695    2424 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0880 47.63192
## 4696    2424 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0880 47.63192
## 4740    2850 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0927 47.63610
## 4821                  8436 143RD CT NE 98052   REDMOND -122.1489 47.67850
## 4834    2140 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0874 47.62941
## 4840                 8700 196TH AVE NE 98053      <NA> -122.0782 47.67993
## 4934    3402 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0994 47.64049
## 5083          19805 NE NOVELTY HILL RD 98053      <NA> -122.0741 47.68643
## 5491                  16644 NE 43RD CT 98052   REDMOND -122.1188 47.64854
## 5494                  16525 NE 43RD CT 98052   REDMOND -122.1193 47.64811
## 5495                  16631 NE 43RD CT 98052   REDMOND -122.1191 47.64824
## 5496                  16639 NE 43RD CT 98052   REDMOND -122.1187 47.64828
## 5497                  16501 NE 43RD CT 98052   REDMOND -122.1199 47.64817
## 5549                 8036 196TH AVE NE 98053      <NA> -122.0770 47.67611
## 5935                10100 203RD AVE NE 98053      <NA> -122.0682 47.68972
## 6055                  19821 NE 61ST PL 98053      <NA> -122.0757 47.66146
## 6230                  17144 NE 46TH CT 98052   REDMOND -122.1106 47.65080
## 6231                  17372 NE 46TH CT 98052   REDMOND -122.1092 47.65088
## 6232                  17376 NE 46TH CT 98052   REDMOND -122.1088 47.65092
## 6233                  17380 NE 46TH CT 98052   REDMOND -122.1088 47.65078
## 6234                  17252 NE 46TH CT 98052   REDMOND -122.1102 47.65079
## 6235                  17256 NE 46TH CT 98052   REDMOND -122.1100 47.65079
## 6236                  17260 NE 46TH CT 98052   REDMOND -122.1098 47.65079
## 6237                  17364 NE 46TH CT 98052   REDMOND -122.1096 47.65078
## 6238                  17368 NE 46TH CT 98052   REDMOND -122.1093 47.65083
## 6239                  17384 NE 46TH CT 98052   REDMOND -122.1088 47.65062
## 6429                 17137 NE 120TH ST 98052   REDMOND -122.1113 47.70674
## 6430                 11818 171ST PL NE 98052   REDMOND -122.1119 47.70639
## 6431                17011 NE 118TH WAY 98052   REDMOND -122.1134 47.70580
## 6432                16943 NE 118TH WAY 98052   REDMOND -122.1138 47.70579
## 6433                16944 NE 118TH WAY 98052   REDMOND -122.1138 47.70624
## 6434                 16909 NE 120TH ST 98052   REDMOND -122.1145 47.70694
## 6435                 17128 NE 120TH ST 98052   REDMOND -122.1115 47.70718
## 6436                 17136 NE 120TH ST 98052   REDMOND -122.1112 47.70716
```

```
## 6437               11902 171ST PL NE 98052   REDMOND -122.1120 47.70651
## 6438               11719 171ST PL NE 98052   REDMOND -122.1125 47.70568
## 6439              16955 NE 118TH WAY 98052   REDMOND -122.1135 47.70579
## 6440               11703 169TH PL NE 98052   REDMOND -122.1146 47.70573
## 6441              16906 NE 118TH WAY 98052   REDMOND -122.1146 47.70631
## 6442              17020 NE 118TH WAY 98052   REDMOND -122.1132 47.70624
## 6443               12212 164TH CT NE 98052   REDMOND -122.1203 47.70956
## 6444               12238 164TH CT NE 98052   REDMOND -122.1203 47.70980
## 6445               12300 164TH CT NE 98052   REDMOND -122.1203 47.71005
## 6446               12308 164TH CT NE 98052   REDMOND -122.1202 47.71018
## 6447               12351 164TH CT NE 98052   REDMOND -122.1212 47.71034
## 6448               12261 164TH CT NE 98052   REDMOND -122.1213 47.70954
## 6449               12249 164TH CT NE 98052   REDMOND -122.1208 47.70993
## 6450               12237 164TH CT NE 98052   REDMOND -122.1208 47.70976
## 6451               12228 164TH CT NE 98052   REDMOND -122.1203 47.70968
## 6452               12250 164TH CT NE 98052   REDMOND -122.1202 47.70993
## 6453               12393 164TH CT NE 98052   REDMOND -122.1212 47.71090
## 6454               12375 164TH CT NE 98052   REDMOND -122.1212 47.71075
## 6455               12285 164TH CT NE 98052   REDMOND -122.1214 47.70982
## 6456               12273 164TH CT NE 98052   REDMOND -122.1214 47.70968
## 6457               12225 164TH CT NE 98052   REDMOND -122.1208 47.70963
## 6512    3018 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0937 47.63761
## 6527               5120 159TH AVE NE 98052   REDMOND -122.1281 47.65495
## 6739                21117 NE 78TH ST 98053      <NA> -122.0580 47.67346
## 6796                17340 NE 36TH ST 98052   REDMOND -122.1078 47.64256
## 6938               11516 174TH CT NE 98052   REDMOND -122.1068 47.70294
## 6939               11527 174TH CT NE 98052   REDMOND -122.1078 47.70295
## 6940               17317 NE 116TH CT 98052   REDMOND -122.1079 47.70435
## 6941               17321 NE 116TH CT 98052   REDMOND -122.1080 47.70417
## 6942               11586 174TH CT NE 98052   REDMOND -122.1070 47.70340
## 6943               11504 174TH CT NE 98052   REDMOND -122.1070 47.70294
## 6944               11503 174TH CT NE 98052   REDMOND -122.1073 47.70294
## 6945               11519 174TH CT NE 98052   REDMOND -122.1075 47.70294
## 6946               11571 174TH CT NE 98052   REDMOND -122.1076 47.70324
## 6947               17327 NE 116TH CT 98052   REDMOND -122.1077 47.70417
## 6948               17333 NE 116TH CT 98052   REDMOND -122.1074 47.70418
## 7039               18852 NE 116TH ST 98052      <NA> -122.0875 47.70434
## 7147               6226 148TH AVE NE 98052   REDMOND -122.1428 47.66317
## 7167    3408 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.1001 47.64084
## 7210               16224 NE 130TH ST 98052      <NA> -122.1227 47.71690
## 7211              13031 164TH AVE NE 98052      <NA> -122.1219 47.71798
## 7446              10814 240TH AVE NE 98053      <NA> -122.0181 47.69454
## 7447              10818 240TH AVE NE 98053      <NA> -122.0182 47.69465
## 7448              10932 240TH AVE NE 98053      <NA> -122.0185 47.69507
## 7449           10921 EASTRIDGE DR NE 98053      <NA> -122.0179 47.69513
## 7450           10915 EASTRIDGE DR NE 98053      <NA> -122.0179 47.69503
## 7451           10821 EASTRIDGE DR NE 98053      <NA> -122.0177 47.69484
## 7452           10811 EASTRIDGE DR NE 98053      <NA> -122.0176 47.69465
## 7453              10888 240TH AVE NE 98053      <NA> -122.0183 47.69475
## 7454              10910 240TH AVE NE 98053      <NA> -122.0183 47.69486
## 7455              10918 240TH AVE NE 98053      <NA> -122.0184 47.69498
## 7456              10940 240TH AVE NE 98053      <NA> -122.0186 47.69520
## 7457              10956 240TH AVE NE 98053      <NA> -122.0187 47.69530
## 7458           10989 EASTRIDGE DR NE 98053      <NA> -122.0182 47.69554
```

```
## 7459           10945 EASTRIDGE DR NE 98053      <NA> -122.0182 47.69544
## 7460           10937 EASTRIDGE DR NE 98053      <NA> -122.0181 47.69533
## 7461           10929 EASTRIDGE DR NE 98053      <NA> -122.0180 47.69523
## 7462           10909 EASTRIDGE DR NE 98053      <NA> -122.0178 47.69493
## 7463           10815 EASTRIDGE DR NE 98053      <NA> -122.0177 47.69474
## 7507               4224 244TH AVE NE 98053      <NA> -122.0130 47.64795
## 7649                13521 NE 70TH ST 98052   REDMOND -122.1588 47.66749
## 7650                13524 NE 70TH ST 98052   REDMOND -122.1588 47.66770
## 7683    2222 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0874 47.62981
## 7791                26510 NE 70TH ST 98053      <NA> -121.9864 47.66943
## 7871                25025 NE 80TH ST 98053      <NA> -122.0051 47.67247
## 8154                21817 NE 97TH PL 98053      <NA> -122.0484 47.68620
## 8232               4224 244TH AVE NE 98053      <NA> -122.0130 47.64795
## 8262                26408 NE 70TH ST 98053      <NA> -121.9857 47.66751
## 8320              11320 244TH AVE NE 98053      <NA> -122.0129 47.69933
## 8377                20210 NE 85TH ST 98053      <NA> -122.0724 47.68034
## 8457    2030 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0877 47.62828
## 8458    2030 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0877 47.62828
## 8535                13304 NE 75TH ST 98052   REDMOND -122.1624 47.67189
## 8541                8490 138TH LN NE 98052   REDMOND -122.1554 47.67857
## 8698    3040 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0950 47.63783
## 8710    3244 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0973 47.63979
## 8763    2608 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0886 47.63358
## 8887    3026 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0947 47.63759
## 8911                2006 250TH PL NE 98074 SAMMAMISH -122.0045 47.62730
## 8946                2006 250TH PL NE 98074 SAMMAMISH -122.0045 47.62730
## 9215             12210 WILLOWS RD NE 98052      <NA> -122.1536 47.70956
## 9293    2829 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0929 47.63521
## 9369               2722 288TH AVE NE 98053      <NA> -121.9538 47.63305
## 9420          22614 NE UNION HILL RD 98053      <NA> -122.0379 47.66734
## 9453                22340 NE 65TH PL 98053      <NA> -122.0407 47.66467
## 9528                20338 NE 85TH ST 98053      <NA> -122.0692 47.68341
## 9546               22465 NE 111TH LN 98053      <NA> -122.0384 47.69732
## 9722    2850 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0927 47.63610
## 10125               20015 NE 85TH ST 98053      <NA> -122.0729 47.67825
## 10318             8565 261ST AVE NE 98053      <NA> -121.9914 47.67951
## 10371   22016 NE REDMOND FALL CITY RD 98053      <NA> -122.0448 47.65178
## 10418            19841 NE REDMOND RD 98053      <NA> -122.0716 47.69623
## 10478               5715 251ST CT NE 98053      <NA> -122.0034 47.65602
## 10623               25621 NE 42ND PL 98053      <NA> -121.9967 47.64604
## 10707               16410 NE 43RD CT 98052   REDMOND -122.1208 47.64847
## 10723              15339 NE 106TH CT 98052   REDMOND -122.1350 47.69446
## 10741              6228 224TH AVE NE 98053      <NA> -122.0407 47.66293
## 10787   3230 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0972 47.63926
## 10844               15430 NE 51ST ST 98052   REDMOND -122.1341 47.65443
## 10958               24143 NE 58TH PL 98053      <NA> -122.0174 47.65712
## 10995              8513 255TH AVE NE 98053      <NA> -121.9994 47.67765
## 11165   2410 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0878 47.63155
## 11289   3260 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0978 47.64009
## 11413              8071 208TH AVE NE 98053      <NA> -122.0640 47.67567
## 11558             11004 132ND AVE NE 98052   REDMOND -122.1642 47.69914
## 11586   2864 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0929 47.63633
## 11728              21031 NE 122ND ST 98053      <NA> -122.0590 47.70618
## 11758   22016 NE REDMOND FALL CITY RD 98053      <NA> -122.0448 47.65178
```

```
## 11772   2808 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0915 47.63512
## 11822   2420 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0879 47.63179
## 11898              11207 248TH AVE NE 98053      <NA> -122.0090 47.69878
## 11899              11207 248TH AVE NE 98053      <NA> -122.0090 47.69878
## 11982                7350 259TH PL NE 98053      <NA> -121.9927 47.66975
## 11992               17656 NE 116TH ST 98052   REDMOND -122.1036 47.70440
## 12212                25205 NE 80TH ST 98053      <NA> -122.0020 47.67283
## 12255               19610 NE 116TH ST 98053      <NA> -122.0784 47.70355
## 12256               19610 NE 116TH ST 98053      <NA> -122.0784 47.70355
## 12392              11220 196TH AVE NE 98053      <NA> -122.0778 47.69907
## 12472   3020 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0945 47.63747
## 12487   2656 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0911 47.63458
## 12577               7702 196TH AVE NE 98053      <NA> -122.0751 47.67266
## 12582   2942 W LAKE SAMMAMISH PKWY NE 98052   REDMOND -122.0932 47.63718
## 12643               7917 219TH AVE NE 98053      <NA> -122.0480 47.67454
## 12686          3851 E AMES LAKE DR NE 98053      <NA> -121.9549 47.64393
## 12764               18025 NE 136TH ST 98052      <NA> -122.0988 47.72113
##       building_grade square_feet_total_living bedrooms bath_full_count
## 25                10                     4920        4               4
## 108                9                     4640        5               2
## 115                6                      660        0               1
## 178               11                     5800        5               4
## 239                9                     3360        2               2
## 246                6                      900        2               1
## 287                9                     4710        4               2
## 295               11                     5060        4              23
## 300               10                     6880        5               1
## 341               11                     4490        4               2
## 344                7                     2700        3               1
## 359               11                     5140        4               2
## 385               11                     6310        4               2
## 396               12                     5080        4               3
## 475               10                     3320        4               1
## 482               12                     6380        6               6
## 508                9                     3700        4               3
## 528                9                     5830        5               3
## 576               10                     4320        4               3
## 661               12                     8090        4               3
## 670               10                     4710        4               2
## 679               11                     8490        7               3
## 742               11                     4577        4               2
## 784               10                     5150        4               1
## 802               11                     4380        4               3
## 811               11                     5270        4               2
## 852               10                     3630        4               2
## 853                9                     3470        3               2
## 877               11                     4640        4               3
## 916               12                     7640        5               5
## 1009               9                     4040        4               3
## 1099              11                     4250        4               2
## 1119              11                     6340        4               4
## 1142              11                     5980        4               3
## 1155               8                     4740        6               5
## 1305               8                     2800        4               2
```

```
## 1345                10                  3990         5            3
## 1368                10                  4550         4            3
## 1380                 8                  3660         3            2
## 1442                 4                   340         0            0
## 1492                11                  4610         4            4
## 1504                12                  7780         5            3
## 1543                11                  4310         4            3
## 1550                11                  5380         4            3
## 1633                 9                  3130         5            2
## 1650                10                  9360         4            3
## 1716                 9                  4610         4            2
## 1745                11                  5360         4            3
## 1870                12                 10630         5            4
## 1962                12                  5300         4            3
## 1963                10                  3090         3            1
## 1964                 7                   550         1            1
## 1976                 9                  3830         5            3
## 1977                 9                  3330         4            2
## 1978                 9                  3480         4            3
## 1979                 9                  3370         4            2
## 1980                 9                  3350         4            3
## 1981                 9                  4080         6            3
## 1982                 9                  3690         5            3
## 2022                 7                   900         1            1
## 2099                11                  5020         4            3
## 2137                 8                  2820         3            1
## 2157                 9                  3130         4            2
## 2264                 5                  1540         3            1
## 2302                 9                  6340        10            2
## 2360                 9                  3200         3            1
## 2361                 5                   310         0            0
## 2469                 8                  2980         3            1
## 2684                10                  3410         5            3
## 2685                 9                  3660         4            3
## 2686                10                  4500         5            2
## 2687                 9                  3480         4            3
## 2688                10                  3310         4            2
## 2689                10                  4340         5            2
## 2690                10                  4390         5            3
## 2699                 9                  5270         4            3
## 2708                 9                  4710         5            3
## 2709                10                  5100         4            3
## 2710                 9                  4910         4            3
## 2717                12                  5390         4            3
## 2742                10                  3950         4            3
## 2852                13                 11810         7            4
## 2934                 9                  4610         4            2
## 2937                 9                  4610         4            2
## 3097                12                  5320         4            3
## 3102                11                  4790         4            3
## 3110                 6                  1290         2            0
## 3111                 7                  1920         5            2
## 3168                 7                  1290         2            1
## 3169                 7                  1290         2            2
```

```
## 3170      7      1600   3      2
## 3171      7      1740   3      2
## 3172      7      1710   3      2
## 3173      7      1740   3      2
## 3174      7      1710   3      2
## 3175      7      1460   3      2
## 3176      7      1820   3      2
## 3177      7      1840   3      2
## 3178      7      1840   3      2
## 3179      7      1600   3      2
## 3180      7      1840   3      2
## 3181      7      1600   3      2
## 3182      7      1290   2      1
## 3183      7      1290   2      1
## 3184      7      1290   2      1
## 3185      7      1290   2      1
## 3186      7      1290   2      1
## 3187      7      1290   2      1
## 3188      7      1290   2      2
## 3189      7      1460   3      2
## 3190      7      1710   3      2
## 3191      7      1460   3      2
## 3192      7      1740   3      2
## 3193      7      1710   3      2
## 3194      7      1710   3      2
## 3195      7      1840   3      2
## 3196      7      1840   3      2
## 3197      7      1820   3      2
## 3198      7      1290   2      1
## 3199      7      1290   2      1
## 3200      7      1290   2      1
## 3201      7      1290   2      1
## 3202      7      1290   2      1
## 3424      8      4290   4      3
## 3464      7      1460   3      2
## 3465      7      1600   3      2
## 3466      7      1840   3      2
## 3467      7      1840   3      2
## 3468      7      1460   3      2
## 3469      7      1600   3      2
## 3470      7      1840   3      2
## 3471      7      1600   3      2
## 3472      7      1740   3      2
## 3473      7      1710   3      2
## 3474      7      1460   3      2
## 3475      7      1740   3      2
## 3476      7      1830   3      1
## 3477      7      1740   3      2
## 3478      7      1290   2      1
## 3479      7      1290   2      1
## 3480      7      1290   2      1
## 3481      7      1290   2      1
## 3482      7      1290   2      1
## 3483      7      1290   2      1
```

```
## 3484                    7                                        1290          2                  2
## 3485                    7                                        1840          3                  2
## 3486                    7                                        1600          3                  2
## 3487                    7                                        1840          3                  2
## 3488                    7                                        1710          3                  2
## 3489                    7                                        1740          3                  2
## 3490                    7                                        1710          3                  2
## 3491                    7                                        1840          4                  1
## 3492                    7                                        1600          3                  2
## 3493                    7                                        1840          3                  2
## 3494                    7                                        1460          3                  2
## 3495                    7                                        1290          2                  1
## 3496                    7                                        1290          2                  1
## 3497                    7                                        1290          2                  1
## 3523                    8                                        3980          4                  3
## 3837                    9                                        4480          4                  2
## 3918                   11                                        6340          4                  4
## 3919                   11                                        5980          4                  3
## 4055                   11                                        4110          4                  1
## 4056                    8                                         900          1                  1
## 4248                    8                                        2610          3                  2
## 4391                    8                                        3140          2                  2
## 4435                   10                                        4330          4                  2
## 4648                   11                                        5790          3                  2
## 4649                    6                                        2410          3                  1
## 4671                   11                                        7810          5                  5
## 4695                    8                                        1430          2                  1
## 4696                   10                                        5330          4                  3
## 4740                   10                                        2900          3                  2
## 4821                    9                                         890          1                  1
## 4834                    7                                         900          1                  1
## 4840                    8                                        4740          6                  5
## 4934                    9                                        4210          4                  2
## 5083                   12                                        5830          4                  4
## 5491                    9                                        2960          4                  2
## 5494                    9                                        3010          4                  2
## 5495                    9                                        2960          4                  2
## 5496                    9                                        2960          4                  2
## 5497                    9                                        2950          5                  3
## 5549                   10                                        4130          3                  1
## 5935                   10                                        5640          3                  3
## 6055                    9                                        3500          4                  2
## 6230                   10                                        3320          4                  2
## 6231                   10                                        3530          4                  2
## 6232                   10                                        3720          4                  3
## 6233                   10                                        3150          4                  2
## 6234                   10                                        3340          5                  2
## 6235                   10                                        3320          5                  2
## 6236                   10                                        3340          5                  2
## 6237                   10                                        3320          5                  2
## 6238                   10                                        3060          4                  2
## 6239                   10                                        3530          4                  2
## 6429                    8                                        3290          4                  2
## 6430                    8                                        2450          4                  2
```

```
## 6431          8        2750   4      2
## 6432          8        3010   4      2
## 6433          8        3200   5      2
## 6434          8        3200   5      2
## 6435          8        3620   5      3
## 6436          8        2810   4      2
## 6437          8        2550   4      2
## 6438          8        2440   3      2
## 6439          8        3160   4      2
## 6440          9        3400   4      3
## 6441          8        2960   4      3
## 6442          8        3110   4      2
## 6443          8        2900   5      3
## 6444          8        3220   5      3
## 6445          8        2510   4      3
## 6446          8        2970   4      3
## 6447          8        3470   5      3
## 6448          8        2580   4      2
## 6449          8        3590   5      2
## 6450          8        3890   5      2
## 6451          8        2520   4      3
## 6452          8        2990   4      3
## 6453          8        3620   5      3
## 6454          8        3300   5      3
## 6455          8        2830   5      3
## 6456          8        2680   4      2
## 6457          8        3330   5      2
## 6512         12        5040   4      3
## 6527         10        5050   5      1
## 6739         10        3850   3      4
## 6796          9        4000   4      3
## 6938          9        3970   4      2
## 6939          9        4610   4      3
## 6940          9        3360   4      2
## 6941          9        3370   4      2
## 6942          9        4390   5      3
## 6943          9        4380   5      4
## 6944          9        4790   5      2
## 6945          9        3970   4      2
## 6946          9        3370   4      2
## 6947          9        3260   4      2
## 6948          9        3210   4      2
## 7039         11        5320   4      3
## 7147         10        5370   5      3
## 7167         10        4460   4      2
## 7210          9        4460   4      3
## 7211          7        3640   5      3
## 7446          8        2650   4      2
## 7447          8        2520   4      2
## 7448          8        2655   4      2
## 7449          8        2520   3      2
## 7450          8        2515   3      2
## 7451          8        2510   3      2
## 7452          8        2520   4      2
```

```
## 7453               8                      2630       4           2
## 7454               8                      2640       4           2
## 7455               8                      2530       4           2
## 7456               8                      2520       4           2
## 7457               8                      2640       4           2
## 7458               8                      2510       3           2
## 7459               8                      2445       4           2
## 7460               8                      2515       3           2
## 7461               8                      2475       4           2
## 7462               8                      2460       4           2
## 7463               8                      2475       4           2
## 7507              10                      5150       4           1
## 7649               9                      4350       4           4
## 7650               9                      4350       4           4
## 7683              12                     13210       4           2
## 7791              11                      5620       3           3
## 7871              12                      6650       4           3
## 8154               9                      3960       4           3
## 8232              10                      5150       4           1
## 8262              12                     13540       7           1
## 8320              10                      5820       3           2
## 8377              12                      8750       5           2
## 8457              10                      3090       3           1
## 8458               7                       550       1           1
## 8535              10                      4420       5           2
## 8541               9                      3130       5           2
## 8698              10                      3380       4           2
## 8710              10                      4950       5           3
## 8763              10                      3460       5           2
## 8887              10                      6360       4           5
## 8911              13                      6280       5           4
## 8946              13                      6280       5           4
## 9215               7                      1650       4           1
## 9293               9                      5830       5           3
## 9369               8                      3240       3           2
## 9420               9                      4800       4           0
## 9453               9                      4610       4           2
## 9528              11                      7110       4           3
## 9546               9                      4480       4           2
## 9722              10                      2900       3           2
## 10125             12                      6680       4           3
## 10318             11                      5150       4           3
## 10371             10                      4930       4           0
## 10418              8                      4770       3           3
## 10478             11                      4400       4           2
## 10623             10                      4570       4           3
## 10707              9                      2630       4           2
## 10723             11                      4790       4           3
## 10741             10                      4200       4           3
## 10787              9                      4640       5           2
## 10844              7                      1800       4           2
## 10958             10                      4010       5           2
## 10995             11                      9070       6           5
## 11165             11                      4140       4           4
```

```
## 11289          9              3310        4           2
## 11413          8              2610        3           1
## 11558          5               730        2           1
## 11586          8              2770        4           1
## 11728          8              2280        2           1
## 11758         10              4930        4           0
## 11772         11              6600        4           2
## 11822         11              4240        3           2
## 11898         11              3690        4           2
## 11899          9              1230        1           1
## 11982         12              7640        5           5
## 11992          8              1670        3           1
## 12212          5               820        2           1
## 12255          5              1650        2           1
## 12256          8              3260        4           2
## 12392          7              2140        3           1
## 12472          8              3150        3           3
## 12487          9              5000        4           1
## 12577         10              5030        3           2
## 12582         12              7070        5           3
## 12643         13              4690        3           2
## 12686          9              3720        3           2
## 12764         12              6990        5           3
##       year_built year_renovated current_zoning sq_ft_lot
## 25          2007              0             22    112650
## 108         1952           2004             11     19173
## 115         1955              0             22    225640
## 178         2008              0             21     63162
## 239         2005              0             24      8752
## 246         1918              0             11     14043
## 287         2014              0             22     18498
## 295         2016              0             22     89734
## 300         2008              0             22    288367
## 341         2008              0             11     55303
## 344         2003              0             22    574992
## 359         2008              0             22    212572
## 385         2005              0             22     36362
## 396         2005              0             22    180774
## 475         1963           1989             11     10247
## 482         2007              0             11     15021
## 508         2010              0             22    305173
## 528         2014              0             11     10454
## 576         2008              0             22    223898
## 661         2006              0             22    176418
## 670         1992              0             11     15167
## 679         2008              0             22    118483
## 742         2011              0             22     98881
## 784         2015              0             22    221720
## 802         2008              0             22    221720
## 811         2001              0             11     18045
## 852         2007              0             22    243176
## 853         1985           2001             22    182516
## 877         2008              0             22    220413
## 916         2007              0             22    144683
```

```
## 1009          2005         0          22    35042
## 1099          2007         0          22   223027
## 1119          2007         0          22    45302
## 1142          2007         0          22    55756
## 1155          2007         0          22   544199
## 1305          1929      1993          11    10046
## 1345          2008         0          11    11842
## 1368          2010         0          22   204732
## 1380          2008         0          22    14820
## 1442          1954         0          22    29933
## 1492          2007         0          22   106722
## 1504          2008         0          17   266152
## 1543          2008         0          22   240855
## 1550          2007         0          22   106722
## 1633          2014         0          13     5930
## 1650          2009         0          21    45738
## 1716          2015         0          22    95989
## 1745          2006         0          19    46609
## 1870          2003         0          22   207781
## 1962          2006         0          22   179031
## 1963          1978         0          11    17715
## 1964          1979         0          11    17715
## 1976          2009         0          13     9480
## 1977          2009         0          13     7399
## 1978          2008         0          13    10309
## 1979          2008         0          13     7457
## 1980          2008         0          13     5788
## 1981          2008         0          13     8371
## 1982          2009         0          13     6731
## 2022          1940         0          11    18102
## 2099          2004         0          22    59677
## 2137          1972         0          11    17995
## 2157          1965      1997          11    13634
## 2264          1938         0          11    28183
## 2302          2009         0          11    17328
## 2360          1968         0          11    19556
## 2361          1964      1992          11    19556
## 2469          1980         0          11    19110
## 2684          2008         0          15     5794
## 2685          2010         0          15     5833
## 2686          2008         0          15     5169
## 2687          2010         0          15     5035
## 2688          2008         0          15     5650
## 2689          2008         0          15     5086
## 2690          2009         0          15     5476
## 2699          2015         0          23  1008414
## 2708          2015         0          22   108542
## 2709          2016         0          22   131301
## 2710          2015         0          22   146779
## 2717          2007         0          22   193842
## 2742          2013         0          22    78844
## 2852          2000         0          21   139392
## 2934          2015         0          22    95989
## 2937          2015         0          22    95989
```

```
## 3097              2007                    0                    22    178160
## 3102              2008                    0                    12      7701
## 3110              1961                    0                    11     18792
## 3111              1961                    0                    11     18792
## 3168              2008                    0                    24      2628
## 3169              2008                    0                    24      2628
## 3170              2009                    0                    24      6451
## 3171              2008                    0                    24      3833
## 3172              2008                    0                    24      4846
## 3173              2008                    0                    24      4862
## 3174              2008                    0                    24      4710
## 3175              2008                    0                    24      4350
## 3176              2009                    0                    24      4622
## 3177              2008                    0                    24      4697
## 3178              2008                    0                    24      5125
## 3179              2008                    0                    24      5481
## 3180              2008                    0                    24      5680
## 3181              2008                    0                    24      5277
## 3182              2008                    0                    24      2807
## 3183              2008                    0                    24      2482
## 3184              2008                    0                    24      2482
## 3185              2008                    0                    24      2828
## 3186              2008                    0                    24      2628
## 3187              2008                    0                    24      2628
## 3188              2008                    0                    24      3940
## 3189              2008                    0                    24      4093
## 3190              2008                    0                    24      4041
## 3191              2008                    0                    24      3989
## 3192              2008                    0                    24      3937
## 3193              2008                    0                    24      3885
## 3194              2008                    0                    24      3781
## 3195              2008                    0                    24      4910
## 3196              2008                    0                    24      5340
## 3197              2009                    0                    24      6952
## 3198              2008                    0                    24      3170
## 3199              2008                    0                    24      3140
## 3200              2008                    0                    24      2482
## 3201              2008                    0                    24      2482
## 3202              2008                    0                    24      2482
## 3424              2008                    0                    13     10922
## 3464              2008                    0                    24      3918
## 3465              2009                    0                    24      4036
## 3466              2008                    0                    24      4487
## 3467              2008                    0                    24      5453
## 3468              2008                    0                    24      4195
## 3469              2008                    0                    24      4179
## 3470              2008                    0                    24      4229
## 3471              2009                    0                    24      4279
## 3472              2008                    0                    24      3915
## 3473              2008                    0                    24      4424
## 3474              2008                    0                    24      4827
## 3475              2008                    0                    24      4948
## 3476              2008                    0                    24      4646
## 3477              2008                    0                    24      4607
```

```
## 3478          2008            0          24       2482
## 3479          2008            0          24       2482
## 3480          2008            0          24       2482
## 3481          2008            0          24       2482
## 3482          2008            0          24       2482
## 3483          2008            0          24       2483
## 3484          2008            0          24       2479
## 3485          2010            0          24       4408
## 3486          2008            0          24       4056
## 3487          2009            0          24       4424
## 3488          2008            0          24       3790
## 3489          2008            0          24       3476
## 3490          2008            0          24       4446
## 3491          2008            0          24       4140
## 3492          2009            0          24       4076
## 3493          2008            0          24       4698
## 3494          2008            0          24       4263
## 3495          2008            0          24       2483
## 3496          2008            0          24       2483
## 3497          2008            0          24       2483
## 3523          1955         1991          11      22545
## 3837          2014            0          23     104688
## 3918          2007            0          22      45302
## 3919          2007            0          22      55756
## 4055          1991            0          11      14388
## 4056          1937         1991          11      14388
## 4248          1988            0          22     228690
## 4391          2015            0          22     260488
## 4435          2011            0          13      11325
## 4648          1999            0           1     657816
## 4649          1935            0           2    1327090
## 4671          2011            0          22     277286
## 4695          2003            0          11      16105
## 4696          2005            0          11      16105
## 4740          1974            0          11      15188
## 4821          1999            0          13       7599
## 4834          1940            0          11      18102
## 4840          2007            0          22     544199
## 4934          1959         1989          11      15353
## 5083          1969            0          23    1127205
## 5491          2011            0          11       7743
## 5494          2011            0          11       6932
## 5495          2011            0          11       6396
## 5496          2011            0          11       7588
## 5497          2011            0          11       5815
## 5549          2013            0          22     104544
## 5935          2012            0          23     127591
## 6055          2013            0          22     223027
## 6230          2012            0          16       5000
## 6231          2014            0          16       5037
## 6232          2013            0          16       6237
## 6233          2013            0          16       5163
## 6234          2012            0          16       5013
## 6235          2012            0          16       5143
```

```
## 6236          2012            0             16      5100
## 6237          2013            0             16      5100
## 6238          2013            0             16      5103
## 6239          2014            0             16      4963
## 6429          2012            0             11      6712
## 6430          2010            0             11      4749
## 6431          2012            0             11      5816
## 6432          2012            0             11      8908
## 6433          2010            0             11      4584
## 6434          2012            0             11      4681
## 6435          2012            0             11      9901
## 6436          2012            0             11     13289
## 6437          2010            0             11      4368
## 6438          2011            0             11      4244
## 6439          2011            0             11      5778
## 6440          2012            0             11      6740
## 6441          2012            0             11      4451
## 6442          2012            0             11      5310
## 6443          2012            0             11      4647
## 6444          2012            0             11      4080
## 6445          2013            0             11      5032
## 6446          2013            0             11      4383
## 6447          2012            0             11      5326
## 6448          2011            0             11      5913
## 6449          2012            0             11      6254
## 6450          2012            0             11      5441
## 6451          2012            0             11      4442
## 6452          2013            0             11      4234
## 6453          2013            0             11      6168
## 6454          2012            0             11      4676
## 6455          2012            0             11      6695
## 6456          2012            0             11      7271
## 6457          2012            0             11      4596
## 6512          1990            0             11     17417
## 6527          2012            0             13     15681
## 6739          2013            0             22     62726
## 6796          2014            0             11     11780
## 6938          2013            0             11      6260
## 6939          2013            0             11      6260
## 6940          2013            0             11      6200
## 6941          2013            0             11      8108
## 6942          2013            0             11      6858
## 6943          2013            0             11      6475
## 6944          2013            0             11      6438
## 6945          2013            0             11      6260
## 6946          2013            0             11      6418
## 6947          2013            0             11      6594
## 6948          2013            0             11      7617
## 7039          1976         1998             19    151153
## 7147          2014            0             13     11718
## 7167          1975         2000             11     20119
## 7210          1978         2005             19    202387
## 7211          1982            0             19    155074
## 7446          2013            0             24      4644
```

```
## 7447           2013                 0                24        4887
## 7448           2013                 0                24        5078
## 7449           2013                 0                24        4409
## 7450           2013                 0                24        4400
## 7451           2013                 0                24        4408
## 7452           2013                 0                24        5140
## 7453           2012                 0                24        5021
## 7454           2012                 0                24        5099
## 7455           2012                 0                24        5127
## 7456           2013                 0                24        4850
## 7457           2013                 0                24        4499
## 7458           2013                 0                24        5232
## 7459           2013                 0                24        4400
## 7460           2013                 0                24        4400
## 7461           2013                 0                24        4400
## 7462           2013                 0                24        4403
## 7463           2013                 0                24        4416
## 7507           2015                 0                22      221720
## 7649           2014                 0                14        8038
## 7650           2014                 0                14        7796
## 7683           2008                 0                11       29728
## 7791           2015                 0                17      218439
## 7871           2015                 0                22      319294
## 8154           2014                 0                22       54014
## 8232           2015                 0                22      221720
## 8262           1999                 0                17      307752
## 8320           2016                 0                18      436507
## 8377           1996                 0                22     1631322
## 8457           1978                 0                11       17715
## 8458           1979                 0                11       17715
## 8535           2014                 0                14       10000
## 8541           2014                 0                13        5930
## 8698           2014                 0                11       16190
## 8710           1996                 0                11       14444
## 8763           1972              2007                11       10861
## 8887           2008                 0                11       21010
## 8911           2003                 0                 6       83813
## 8946           2003                 0                 6       83813
## 9215           1954                 0                 1      316245
## 9293           2014                 0                11       10454
## 9369           2015                 0                22      279219
## 9420           2015                 0                22      464350
## 9453           2015                 0                22       95989
## 9528           2006                 0                22      209088
## 9546           2014                 0                23      104688
## 9722           1974                 0                11       15188
## 10125          1998                 0                22      167270
## 10318          1997                 0                17      963702
## 10371          2016                 0                22      111950
## 10418          2003                 0                22      261795
## 10478          2015                 0                22       93654
## 10623          2016                 0                22      112384
## 10707          1987                 0                11        9705
## 10723          2008                 0                12        6451
```

```
## 10741        2015             0              22     30475
## 10787        1952          2004              11     19173
## 10844        1918             0              13     41789
## 10958        1997             0              19    117454
## 10995        2005             0              22    186525
## 11165        1995             0              11     14041
## 11289        1982             0              11     10532
## 11413        1948             0              22    192099
## 11558        1952             0              11     13260
## 11586        1938             0              11     13170
## 11728        1981             0              22    439956
## 11758        2016             0              22    111950
## 11772        2001             0              11     41217
## 11822        2005             0              11     15368
## 11898        1999             0              18   1166246
## 11899        1999             0              18   1166246
## 11982        2007             0              22    144683
## 11992        1964             0              11    425145
## 12212        1932             0              22    167838
## 12255        1980             0              22    227383
## 12256        1980             0              22    227383
## 12392        1987             0              22    188614
## 12472        1955             0              11     30894
## 12487        1962          1980              11      9600
## 12577        2005             0              22    227818
## 12582        1974          2003              11     29494
## 12643        2009             0              22     77418
## 12686        2006             0              22     19290
## 12764        2006             0              22    226512
```

b. xi. Investigate further by calculating the leverage, cooks distance, and

covariance rations. Comment on all cases that are problematics.

```
# nrow(mydata) -- 12865

# Let's look at the leverage (hat value), cook's distance, and covariance ratio
# for cases with large.residual = TRUE
mydata[mydata$large.residual, c("cooks.distance","leverage","covariance.ratios")]
```

```
##       cooks.distance     leverage covariance.ratios
## 25      0.0010257923 0.0011869228         0.9984429
## 108     0.0033044011 0.0042136523         1.0017920
## 115     0.0024117590 0.0018063670         0.9972755
## 178     0.0011465389 0.0012854582         0.9984365
## 239     0.0001090031 0.0001668056         0.9982225
## 246     0.0031564325 0.0017464582         0.9954241
## 287     0.0006893490 0.0007933254         0.9980305
## 295     0.2396277578 0.1188300220         1.1278061
## 300     0.0075267399 0.0043854071         0.9984257
## 341     0.0004487567 0.0006708699         0.9986685
```

```
## 344      0.0057351762 0.0089305205          1.0071144
## 359      0.0017397504 0.0014731451          0.9975264
## 385      0.0025307845 0.0015599477          0.9959382
## 396      0.0014805940 0.0014632775          0.9981591
## 475      0.0006986847 0.0009934765          0.9988611
## 482      0.0031738517 0.0039360295          1.0014278
## 508      0.0033052073 0.0053950353          1.0036382
## 528      0.0010875245 0.0016248737          0.9996238
## 576      0.0008319101 0.0013155085          0.9994538
## 661      0.0027968265 0.0028437415          0.9996540
## 670      0.0016660191 0.0005997724          0.9906016
## 679      0.0070450303 0.0036509393          0.9968763
## 742      0.0004484261 0.0007724219          0.9991065
## 784      0.0019524886 0.0023320386          0.9996953
## 802      0.0010311108 0.0015000453          0.9994297
## 811      0.0035835198 0.0009199438          0.9867072
## 852      0.0009102800 0.0015231355          0.9997943
## 853      0.0004455858 0.0007137136          0.9988811
## 877      0.0016204219 0.0014373881          0.9976943
## 916      0.0045527934 0.0030400158          0.9978991
## 1009     0.0002877654 0.0004295533          0.9984230
## 1099     0.0009422438 0.0015341461          0.9997424
## 1119     0.0020112581 0.0015651144          0.9972234
## 1142     0.0011810063 0.0010314525          0.9972200
## 1155     0.0149292319 0.0095001048          1.0041643
## 1305     0.0011346071 0.0016966438          0.9996980
## 1345     0.0002542520 0.0004161686          0.9986340
## 1368     0.0007075141 0.0011555917          0.9993700
## 1380     0.0003101548 0.0004637808          0.9984617
## 1442     0.0016275507 0.0015502173          0.9981018
## 1492     0.0010091685 0.0012623284          0.9987649
## 1504     0.0048822856 0.0030902960          0.9976377
## 1543     0.0010641713 0.0017134567          0.9998962
## 1550     0.0008218085 0.0008187770          0.9975437
## 1633     0.0001621965 0.0002170445          0.9979159
## 1650     0.0135138750 0.0062910557          0.9987168
## 1716     0.0009970945 0.0007971874          0.9965835
## 1745     0.0009432943 0.0007656760          0.9966233
## 1870     0.0074126080 0.0066798036          1.0030493
## 1962     0.0024740915 0.0014260536          0.9953770
## 1963     0.0006106903 0.0007478570          0.9981840
## 1964     0.0008404656 0.0004465255          0.9938367
## 1976     0.0013242658 0.0003997737          0.9883823
## 1977     0.0006853792 0.0001884214          0.9869458
## 1978     0.0012447634 0.0003547031          0.9875953
## 1979     0.0006774109 0.0001869084          0.9869939
## 1980     0.0012740900 0.0003552299          0.9873042
## 1981     0.0014775095 0.0004630181          0.9889040
## 1982     0.0012863954 0.0003793806          0.9880669
## 2022     0.0013424594 0.0009537856          0.9961454
## 2099     0.0005363480 0.0006876495          0.9982629
## 2137     0.0005589791 0.0003740237          0.9952344
## 2157     0.0005565648 0.0005217858          0.9970069
## 2264     0.0031592360 0.0015138722          0.9941316
```

```
## 2302    0.0017626301 0.0024168498          1.0001911
## 2360    0.0005459138 0.0005608369          0.9974005
## 2361    0.0024410539 0.0009953756          0.9922280
## 2469    0.0004473775 0.0003714912          0.9963336
## 2684    0.0015598153 0.0004735320          0.9885261
## 2685    0.0012316856 0.0003774574          0.9885464
## 2686    0.0014329427 0.0005141746          0.9904813
## 2687    0.0012141040 0.0003613572          0.9881655
## 2688    0.0011725802 0.0003451790          0.9880091
## 2689    0.0013120914 0.0004579103          0.9901284
## 2690    0.0013092871 0.0004699203          0.9904396
## 2699    0.0340776069 0.0254130949          1.0215325
## 2708    0.0013991091 0.0008301737          0.9949675
## 2709    0.0012136735 0.0008456224          0.9959317
## 2710    0.0017277029 0.0010943362          0.9956365
## 2717    0.0030168291 0.0015203003          0.9945262
## 2742    0.0005096100 0.0004644780          0.9968336
## 2852    0.0323551464 0.0084297003          0.9945063
## 2934    0.0009961450 0.0007971874          0.9965880
## 2937    0.0009613303 0.0007971874          0.9967538
## 3097    0.0026676183 0.0014175848          0.9948117
## 3102    0.0007357039 0.0006850037          0.9971425
## 3110    0.0008237369 0.0008555478          0.9977359
## 3111    0.0003968700 0.0005785581          0.9985112
## 3168    0.0039875726 0.0004821071          0.9699125
## 3169    0.0033779971 0.0004123887          0.9701391
## 3170    0.0029576959 0.0003738115          0.9711367
## 3171    0.0027292290 0.0003492111          0.9714720
## 3172    0.0027537800 0.0003513317          0.9713897
## 3173    0.0027317318 0.0003496169          0.9714795
## 3174    0.0027533928 0.0003512709          0.9713887
## 3175    0.0030548353 0.0003796849          0.9706469
## 3176    0.0027749988 0.0003587433          0.9717792
## 3177    0.0026791022 0.0003464883          0.9717788
## 3178    0.0026799098 0.0003466282          0.9717818
## 3179    0.0028631251 0.0003611646          0.9710675
## 3180    0.0026810696 0.0003468243          0.9717859
## 3181    0.0028623150 0.0003610450          0.9710660
## 3182    0.0039884401 0.0004822319          0.9699139
## 3183    0.0039868755 0.0004820066          0.9699114
## 3184    0.0039868755 0.0004820066          0.9699114
## 3185    0.0039885428 0.0004822466          0.9699141
## 3186    0.0039875726 0.0004821071          0.9699125
## 3187    0.0039875726 0.0004821071          0.9699125
## 3188    0.0033853118 0.0004134073          0.9701493
## 3189    0.0030536527 0.0003795151          0.9706449
## 3190    0.0027516005 0.0003509865          0.9713838
## 3191    0.0030531822 0.0003794474          0.9706441
## 3192    0.0027294619 0.0003492496          0.9714728
## 3193    0.0027512095 0.0003509237          0.9713827
## 3194    0.0027509545 0.0003508825          0.9713819
## 3195    0.0026794947 0.0003465567          0.9717803
## 3196    0.0026803440 0.0003467022          0.9717834
## 3197    0.0027812390 0.0003597503          0.9717962
```

```
## 3198     0.0039902426 0.0004824902          0.9699168
## 3199     0.0039900914 0.0004824686          0.9699165
## 3200     0.0039868755 0.0004820066          0.9699114
## 3201     0.0039868755 0.0004820066          0.9699114
## 3202     0.0039868755 0.0004820066          0.9699114
## 3424     0.0024495202 0.0040292686          1.0022766
## 3464     0.0029958323 0.0003794015          0.9712020
## 3465     0.0028920607 0.0003723198          0.9716726
## 3466     0.0026276922 0.0003464232          0.9723251
## 3467     0.0026294974 0.0003467421          0.9723320
## 3468     0.0029970621 0.0003795821          0.9712041
## 3469     0.0028044725 0.0003604398          0.9716123
## 3470     0.0026272739 0.0003463465          0.9723233
## 3471     0.0028929496 0.0003724558          0.9716744
## 3472     0.0026776797 0.0003492414          0.9720233
## 3473     0.0027005098 0.0003511464          0.9719381
## 3474     0.0029999885 0.0003800094          0.9712089
## 3475     0.0026801737 0.0003496534          0.9720308
## 3476     0.0037772871 0.0004924125          0.9721523
## 3477     0.0026793022 0.0003495111          0.9720283
## 3478     0.0039134029 0.0004820066          0.9704772
## 3479     0.0039134029 0.0004820066          0.9704772
## 3480     0.0039134029 0.0004820066          0.9704772
## 3481     0.0039134029 0.0004820066          0.9704772
## 3482     0.0039134029 0.0004820066          0.9704772
## 3483     0.0039134076 0.0004820073          0.9704772
## 3484     0.0033146796 0.0004122789          0.9707011
## 3485     0.0028051885 0.0003710987          0.9724468
## 3486     0.0028040584 0.0003603760          0.9716114
## 3487     0.0027138345 0.0003583941          0.9723855
## 3488     0.0026989172 0.0003508860          0.9719335
## 3489     0.0026767519 0.0003490836          0.9720201
## 3490     0.0027005680 0.0003511558          0.9719383
## 3491     0.0037826809 0.0004935759          0.9721801
## 3492     0.0028922054 0.0003723420          0.9716729
## 3493     0.0026280542 0.0003464886          0.9723266
## 3494     0.0029973689 0.0003796270          0.9712046
## 3495     0.0039134076 0.0004820073          0.9704772
## 3496     0.0039134076 0.0004820073          0.9704772
## 3497     0.0039134076 0.0004820073          0.9704772
## 3523     0.0026507975 0.0015849983          0.9957735
## 3837     0.0006838252 0.0007383763          0.9977580
## 3918     0.0012625486 0.0015651144          0.9990412
## 3919     0.0007652759 0.0010314525          0.9987521
## 4055     0.0014804592 0.0011854653          0.9969800
## 4056     0.0045005009 0.0013619354          0.9893770
## 4248     0.0007189969 0.0012086223          0.9994892
## 4391     0.0011440636 0.0019973401          1.0003650
## 4435     0.0003501356 0.0004602009          0.9981078
## 4648     0.1052345743 0.0102109306          0.9722188
## 4649     0.6906373336 0.0441092700          0.9884534
## 4671     0.0033942707 0.0042189274          1.0017201
## 4695     0.0015209988 0.0004222043          0.9873116
## 4696     0.0012415938 0.0008215703          0.9956192
```

```
## 4740    0.0003991572 0.0006797623          0.9989884
## 4821    0.0018888427 0.0015127881          0.9973094
## 4834    0.0011454391 0.0009537856          0.9969297
## 4840    0.0156021509 0.0095001048          1.0038956
## 4934    0.0009536065 0.0009580187          0.9977148
## 5083    0.0901980470 0.0309037562          1.0213738
## 5491    0.0001204817 0.0001827619          0.9982166
## 5494    0.0001184603 0.0001828554          0.9982600
## 5495    0.0001210330 0.0001833950          0.9982144
## 5496    0.0001205418 0.0001828299          0.9982163
## 5497    0.0002495714 0.0003894789          0.9984937
## 5549    0.0007267978 0.0011075487          0.9991545
## 5935    0.0013552866 0.0010634637          0.9967606
## 6055    0.0009548636 0.0013018514          0.9990555
## 6230    0.0016927822 0.0003674489          0.9834783
## 6231    0.0016557276 0.0003716195          0.9840513
## 6232    0.0018284484 0.0004261427          0.9847252
## 6233    0.0018688703 0.0003973082          0.9831440
## 6234    0.0016795075 0.0003655666          0.9835237
## 6235    0.0016921418 0.0003673263          0.9834790
## 6236    0.0016791110 0.0003654903          0.9835241
## 6237    0.0017228537 0.0003748257          0.9835249
## 6238    0.0019715318 0.0004140669          0.9829447
## 6239    0.0016560923 0.0003716927          0.9840510
## 6429    0.0051922484 0.0003579187          0.9468824
## 6430    0.0028656656 0.0001849280          0.9431183
## 6431    0.0034505351 0.0002283094          0.9445494
## 6432    0.0040677930 0.0002746190          0.9456881
## 6433    0.0044944781 0.0003068287          0.9463174
## 6434    0.0048150696 0.0003295367          0.9464729
## 6435    0.0087269903 0.0006216813          0.9488530
## 6436    0.0035389281 0.0002355183          0.9448734
## 6437    0.0028910251 0.0001879459          0.9435341
## 6438    0.0030345974 0.0001959032          0.9431503
## 6439    0.0044687097 0.0003045868          0.9462308
## 6440    0.0050615201 0.0003607200          0.9486138
## 6441    0.0072029810 0.0004873128          0.9460152
## 6442    0.0044465560 0.0003022794          0.9460878
## 6443    0.0062548245 0.0004833333          0.9527154
## 6444    0.0065752820 0.0005214943          0.9539694
## 6445    0.0066645597 0.0004997116          0.9512863
## 6446    0.0063454767 0.0004937736          0.9530550
## 6447    0.0071497350 0.0005789957          0.9549787
## 6448    0.0026806822 0.0002002402          0.9508024
## 6449    0.0059395065 0.0004822235          0.9549975
## 6450    0.0077410779 0.0006442085          0.9562607
## 6451    0.0065804276 0.0004930996          0.9512497
## 6452    0.0063582774 0.0004955632          0.9531320
## 6453    0.0076861191 0.0006311932          0.9556587
## 6454    0.0067255502 0.0005369915          0.9542924
## 6455    0.0062439332 0.0004799401          0.9524606
## 6456    0.0029151760 0.0002198457          0.9512864
## 6457    0.0047098882 0.0003741964          0.9539020
## 6512    0.0027687042 0.0013074985          0.9938082
```

```
## 6527     0.0011439540 0.0014878729          0.9991076
## 6739     0.0010574400 0.0010579402          0.9977993
## 6796     0.0002679505 0.0004499813          0.9987267
## 6938     0.0008439725 0.0003966984          0.9928578
## 6939     0.0012674214 0.0006937685          0.9942942
## 6940     0.0005499315 0.0002296020          0.9916774
## 6941     0.0005484610 0.0002296106          0.9917020
## 6942     0.0011257892 0.0005881841          0.9938582
## 6943     0.0021717622 0.0011535140          0.9945433
## 6944     0.0015445694 0.0008612923          0.9945893
## 6945     0.0008439725 0.0003966984          0.9928578
## 6946     0.0005520202 0.0002309316          0.9916967
## 6947     0.0005282048 0.0002164612          0.9914939
## 6948     0.0005193341 0.0002109403          0.9914059
## 7039     0.0008794222 0.0013133248          0.9993099
## 7147     0.0008071568 0.0008465522          0.9977625
## 7167     0.0040556743 0.0007040698          0.9794912
## 7210     0.0034382281 0.0013317753          0.9920764
## 7211     0.0049454782 0.0015271078          0.9897918
## 7446     0.0019642756 0.0002300190          0.9686643
## 7447     0.0019171957 0.0002216273          0.9682447
## 7448     0.0019661194 0.0002303729          0.9686836
## 7449     0.0019179172 0.0002216866          0.9682413
## 7450     0.0019172881 0.0002215034          0.9682251
## 7451     0.0019167196 0.0002213281          0.9682090
## 7452     0.0019168567 0.0002216008          0.9682465
## 7453     0.0018489967 0.0002157551          0.9685378
## 7454     0.0018541697 0.0002165786          0.9685709
## 7455     0.0018138062 0.0002095564          0.9682144
## 7456     0.0019172478 0.0002216315          0.9682445
## 7457     0.0019590073 0.0002291654          0.9686307
## 7458     0.0019156108 0.0002212414          0.9682150
## 7459     0.0019175169 0.0002199960          0.9679999
## 7460     0.0019172881 0.0002215034          0.9682251
## 7461     0.0019152975 0.0002203961          0.9680963
## 7462     0.0019160039 0.0002201497          0.9680481
## 7463     0.0019152776 0.0002203946          0.9680964
## 7507     0.0023158182 0.0024123317          0.9993094
## 7649     0.0006917948 0.0011398730          0.9993746
## 7650     0.0006918818 0.0011402519          0.9993754
## 7683     0.0103781570 0.0159671542          1.0143040
## 7791     0.0010461800 0.0015224008          0.9994529
## 7871     0.0046273592 0.0030157643          0.9977352
## 8154     0.0002961755 0.0004439867          0.9984482
## 8232     0.0024292806 0.0023320386          0.9989182
## 8262     0.0253133456 0.0181771335          1.0137703
## 8320     0.0064410706 0.0050596919          1.0007889
## 8377     0.3807445906 0.0644221375          1.0471077
## 8457     0.0005906999 0.0007478570          0.9982856
## 8458     0.0008222984 0.0004465255          0.9939909
## 8535     0.0003569741 0.0005152238          0.9984215
## 8541     0.0001631968 0.0002170445          0.9978983
## 8698     0.0002796865 0.0003719032          0.9980529
## 8710     0.0022473554 0.0007041959          0.9891443
```

```
## 8763     0.0005216706 0.0006359458         0.9980577
## 8887     0.0129194672 0.0027836422         0.9857741
## 8911     0.0013009607 0.0019755686         1.0000169
## 8946     0.0013009607 0.0019755686         1.0000169
## 9215     0.0024188561 0.0026122144         0.9996397
## 9293     0.0011160547 0.0016248737         0.9995570
## 9369     0.0013616896 0.0021917458         1.0003758
## 9420     0.0060749032 0.0073714500         1.0048367
## 9453     0.0005106974 0.0007971874         0.9989027
## 9528     0.0056161190 0.0022635893         0.9933926
## 9546     0.0006538835 0.0007383763         0.9979122
## 9722     0.0007277701 0.0006797623         0.9971501
## 10125    0.0052324495 0.0048478089         1.0013104
## 10318    0.0133567782 0.0224122117         1.0212129
## 10371    0.0021064306 0.0028288247         1.0005472
## 10418    0.0015299893 0.0023617801         1.0004458
## 10478    0.0004579563 0.0007820097         0.9990968
## 10623    0.0004269388 0.0006254808         0.9985711
## 10707    0.0001299349 0.0002014687         0.9982897
## 10723    0.0004312082 0.0006878271         0.9988453
## 10741    0.0002559881 0.0004202566         0.9986450
## 10787    0.0027445767 0.0014033335         0.9945172
## 10844    0.0016949044 0.0021391104         0.9996722
## 10958    0.0004796424 0.0004712638         0.9971420
## 10995    0.0030409613 0.0051210076         1.0034327
## 11165    0.0013206888 0.0014774990         0.9986218
## 11289    0.0006334948 0.0002389100         0.9907170
## 11413    0.0027417091 0.0012059380         0.9931160
## 11558    0.0042468752 0.0009956524         0.9853908
## 11586    0.0016574128 0.0011102733         0.9959786
## 11728    0.0068990834 0.0047846693         0.9998698
## 11758    0.0022656737 0.0028288247         1.0003331
## 11772    0.0120241217 0.0018168689         0.9774022
## 11822    0.0025447045 0.0006655231         0.9867224
## 11898    0.0327527392 0.0348098650         1.0330472
## 11899    0.0767419624 0.0375479305         1.0318106
## 11982    0.0086379247 0.0030400158         0.9928114
## 11992    0.0732808662 0.0046287603         0.9463640
## 12212    0.0020701946 0.0019228612         0.9983743
## 12255    0.0041240793 0.0022774808         0.9959444
## 12256    0.0011712201 0.0011821940         0.9979574
## 12392    0.0008750702 0.0010714059         0.9985077
## 12472    0.0042021297 0.0013192567         0.9897842
## 12487    0.0110948252 0.0046009420         0.9960107
## 12577    0.0039696302 0.0014696028         0.9917631
## 12582    0.0186385515 0.0024805710         0.9747313
## 12643    0.0160063267 0.0019543570         0.9716895
## 12686    0.0001473875 0.0002401013         0.9984466
## 12764    0.0054974616 0.0021949561         0.9932346
```

```r
# check if any outlier cases have cook's distance > 1
sum(mydata[mydata$large.residual, c("cooks.distance")] > 1)
```

```
## [1] 0
```

```r
sum(mydata[, c("cooks.distance")] > 1)
```

```
## [1] 0
```

```r
# None of the cases have cooks distance > 1, so none of the cases have undue
# influence on the model

# calculate average leverage using formula = (k+1/n)
avg_leverage <- (6+1)/12865
# twice the average leverage
times_2_leverage <- avg_leverage*2
# three times average leverage
times_3_leverage <- avg_leverage*3
# check if there are outlier cases with levarage > 3 times the average leverage
sum(mydata[mydata$large.residual, c("levarage")] >= times_3_leverage)
```

```
## [1] 0
```

```r
# There are none. All the cases are within the range of three times the average.
sum(mydata[mydata$large.residual, c("levarage")] >= times_2_leverage)
```

```
## [1] 0
```

```r
# There are none. All the cases are within the range of two times the average.

# check the minimum and maximum covariance ratios
# calculate the covariance minimum and maximum range
cov_ratio_min <- 1-(3*7/12865) # 1-[3(k+1)/n]
cov_ratio_max <- 1+(3*7/12865) # 1+[3(k+1)/n]
# checking covariance ratios
min(mydata[mydata$large.residual, c("covariance.ratios")])
```

```
## [1] 0.9431183
```

```r
# 0.94 which is less that calculated minimum
max(mydata[mydata$large.residual, c("covariance.ratios")])
```

```
## [1] 1.127806
```

```r
# 1.127 which is greater than calculated maximum
# check cook's distance for out of range covariance ratio cases
mydata$outside.covariance.ratios <- mydata$covariance.ratios > cov_ratio_max | mydata$covariance.ratios
# check cook's distance for out of range covariance ratios
sum(mydata[mydata$outside.covariance.ratios, c("cooks.distance")] > 1)
```

```
## [1] 0
```

```r
# As there are no cases with cook's distance > 1 they are not alarming
```

b. xii. Perform the necessary calculations to assess the assumption of

independence and state if the condition is met or not.

```r
# execute Durbin-Watson test to assess the assumption of independent error
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:boot':
##
##     logit

## The following object is masked from 'package:dplyr':
##
##     recode
```

```
durbinWatsonTest(hs_data_multi_rgr_mdl_4)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.7302629     0.5394659       0
##  Alternative hypothesis: rho != 0
```

```
# Value is < 1 i.e. 0.539 and p-value is 0 which is < 0.05
# Result of the test conclude the residuals in the regression are positively
# correlated and thus fails the assumption of independence or errors.
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
dwtest(hs_data_multi_rgr_mdl_4)
```

```
##
##  Durbin-Watson test
##
## data:  hs_data_multi_rgr_mdl_4
## DW = 0.53947, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

```
# Alternative hypothesis is true and thus error terms or residuals are
# auto correlated.
```

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

```
# calculating mean vif
mean(vif(hs_data_multi_rgr_mdl_4))
```

```
## [1] 1.671316
```

```
# calculating vif
vif(hs_data_multi_rgr_mdl_4)
```

```
##              sale_reason           building_grade square_feet_total_living
##                 1.017011                 2.362311                 2.582797
##          bath_full_count               year_built                sq_ft_lot
```

```
##                1.576506                1.374726                1.114547
```

```
# calculating tolerance
1/vif(hs_data_multi_rgr_mdl_4)
```

```
##              sale_reason          building_grade square_feet_total_living
##                0.9832737               0.4233143                0.3871772
##          bath_full_count              year_built                sq_ft_lot
##                0.6343139               0.7274178                0.8972257
```

```
# Observations
# 1. There is no vif greater than 10
# 2. Average vif is > 1, though not substantially and thus there may be some
# bias in the model (not significant). Value of vif is way below 5 and near 1.
# 3. None of the tolerance is below 0.1, no we do not have any serious problem
# with the model
# 4. None of the tolerance is below 0.2
# looking at above observation we can say that no multi-collinearity assumption
# is met.
```

b. xiv. Visually check the assumptions related to the residuals using the plot()

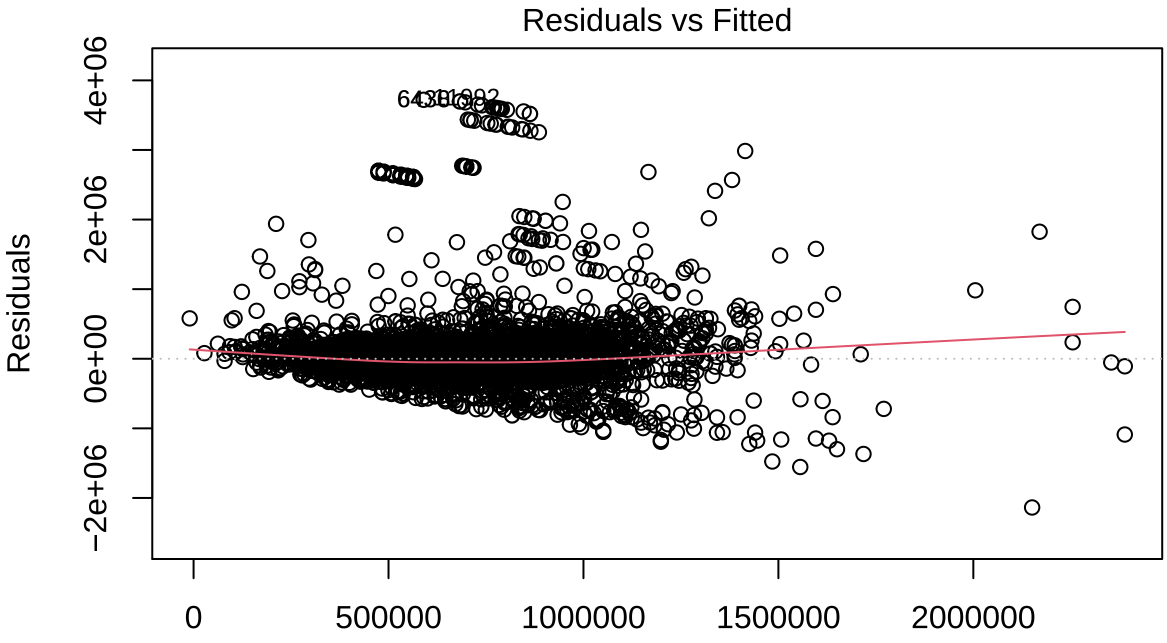
and hist() functions. Summarize what each graph is informing you of and if any

anomalies are present.

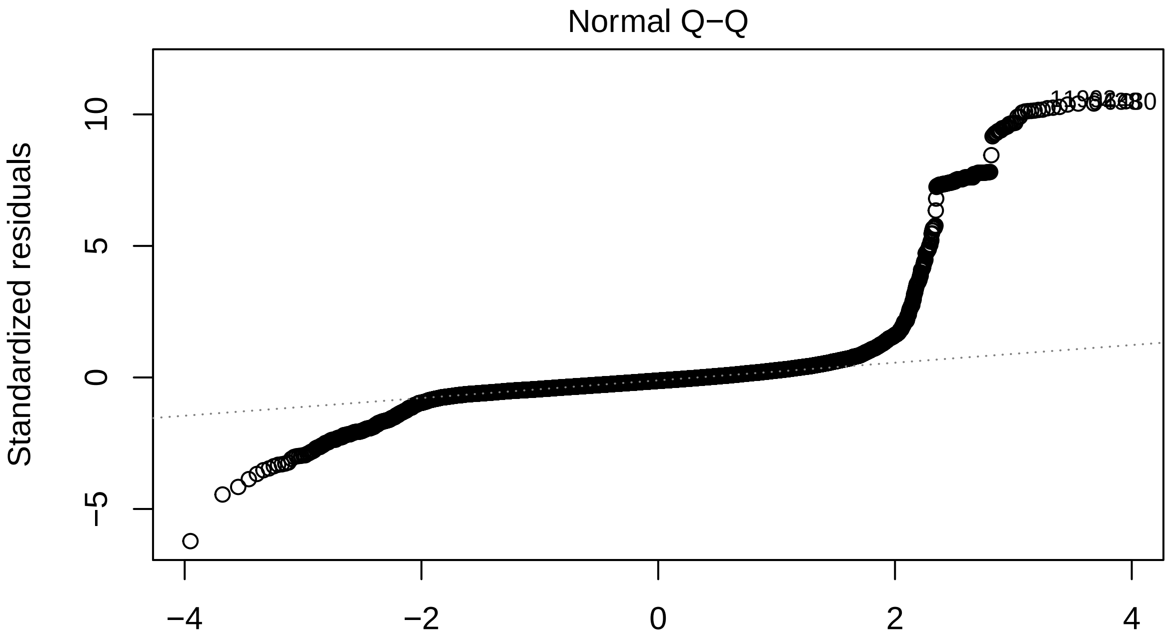
```
# Plotting standardized residual on y-axis and predicted values on x-axis.
# Plot is useful to determine if assumptions of random errors or homoscedasticity
# have been met.
# save fitted.values or predicted values as a variable in mydata data frame
mydata$fitted <- hs_data_multi_rgr_mdl_4$fitted.values

library(ggplot2)
library(qqplotr)
```

```
##
## Attaching package: 'qqplotr'

## The following objects are masked from 'package:ggplot2':
##
##     stat_qq_line, StatQqLine
```

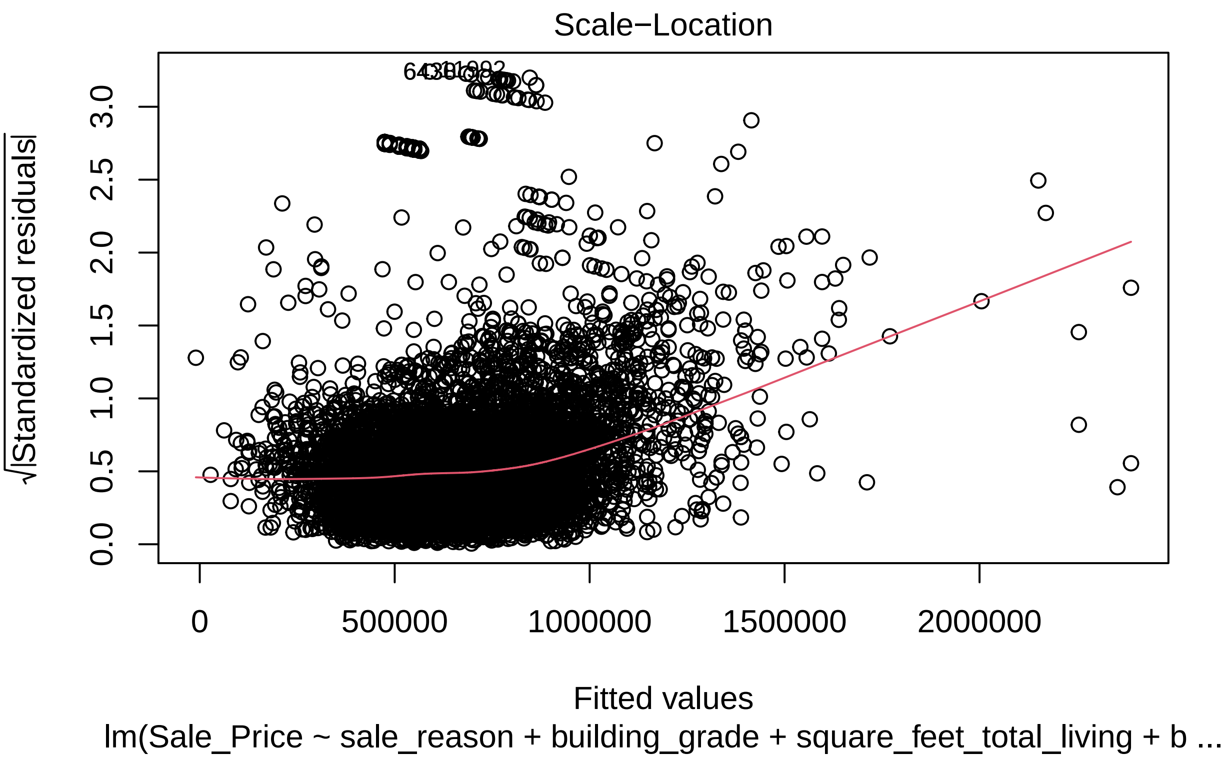
```
plot(hs_data_multi_rgr_mdl_4)
```

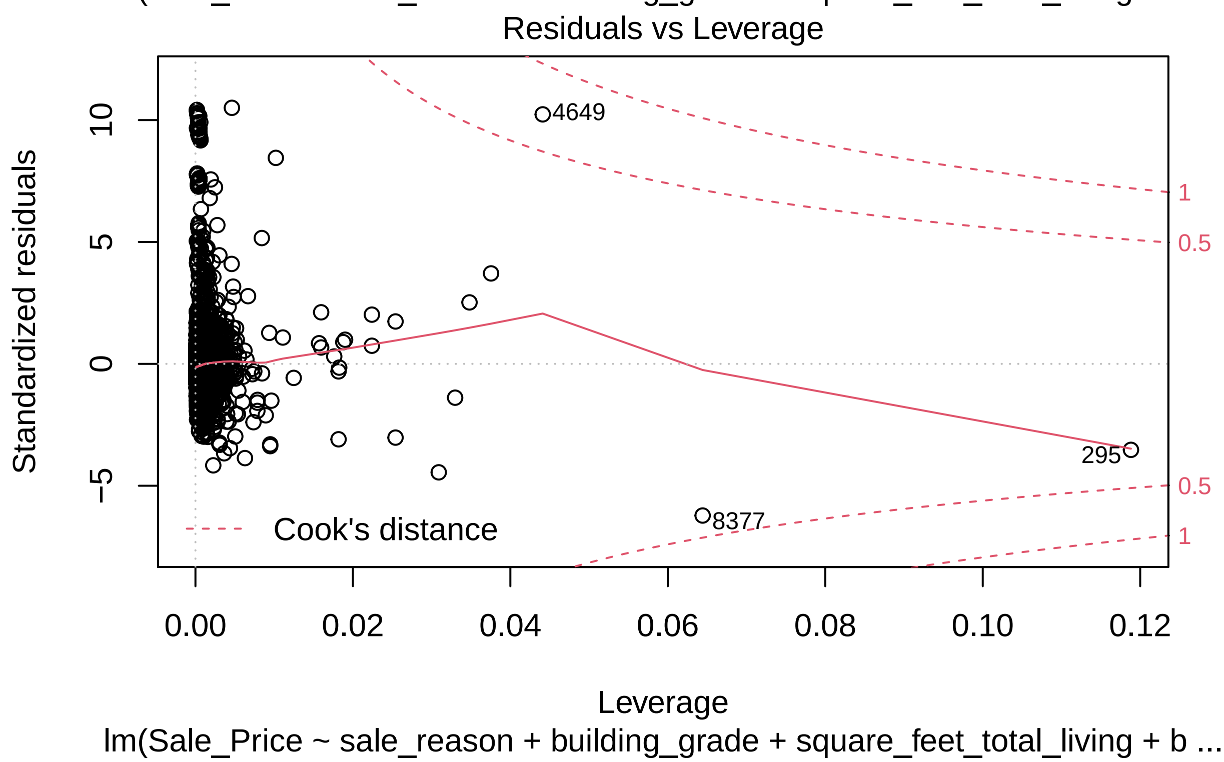


lm(Sale_Price ~ sale_reason + building_grade + square_feet_total_living + b ...



lm(Sale_Price ~ sale_reason + building_grade + square_feet_total_living + b ...

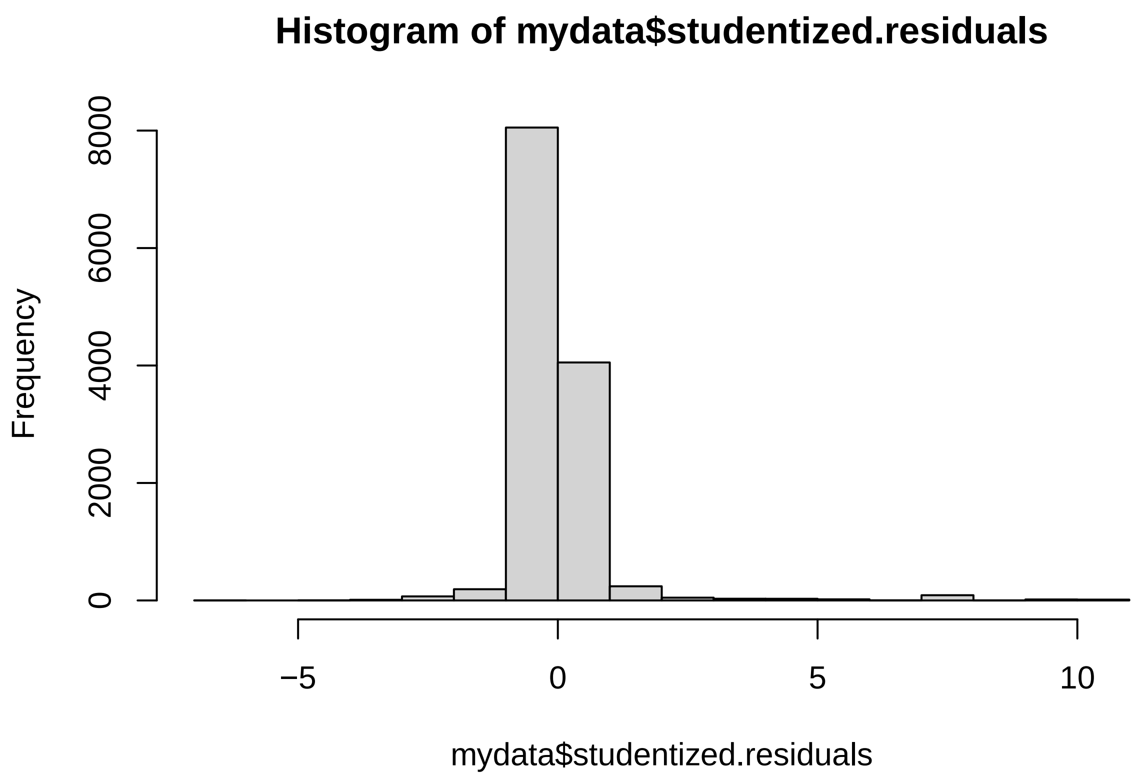




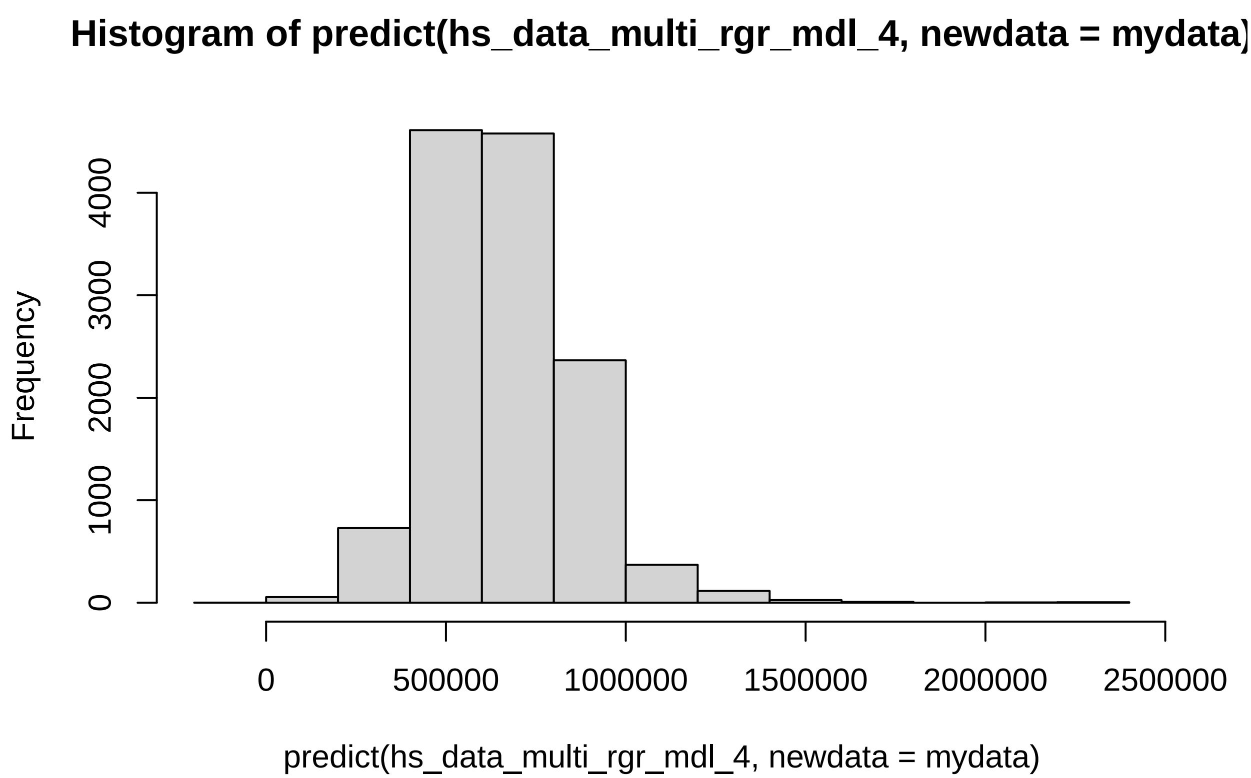

```
# Plot 1 - Scatterplot (Residual vs Fitted) - Access linearity of data
# Ideally, the residual plot will show no fitted pattern. That is, the red
# line should be approximately horizontal at zero. The presence of a pattern may
# indicate a problem with some aspect of the linear model. In our example,
# there is no pattern in the residual plot. This suggests that we can assume
# linear relationship between the predictors and the outcome variables.
```

```
# Plot 2 = QQ plot of residuals - Access normality of residuals
# It can be used to visually check the normality assumption. The normal
# probability plot of residuals should approximately follow a straight line.
# In our example, half of the points fall exactly on the reference line which
# about half does not. We cannot assume normality of residuals in our case.

# Plot 3 - It's also called Spread-Location or scale-location plot.
# It is to evaluate Homogeneity of variance. This plot shows if residuals
# are spread equally along the ranges of predictors. It's good if you see a
# horizontal line with equally spread points. In our example, while majority
# of points are concentrated together and maintains the straight horizontal
# red line, few observations on right exist with higher residuals that are
# changes the red line from being horizontal to slant. I will still say overall
# data satisfies Homogeneity of variance because majority of points are randomly
# distributed and keeping the red line almost horizontal. If assumption is
# failed, a possible solution to reduce the heteroscedasticity problem is to
# use a log or square root transformation of the outcome variable (y).
# For e.g. model2 <- lm(log(sales) ~ youtube, data = marketing)

# Plot 4 - Residuals vs Leverage - outliers and high leverage points.
# Outliers - An outlier is a point that has an extreme outcome variable value.
# The presence of outliers may affect the interpretation of the model,
# because it increases the RSE. Outliers can be identified by examining the
# standardized residual (or studentized residual), which is the residual
# divided by its estimated standard error. Standardized residuals can be
# interpreted as the number of standard errors away from the regression line. Observations whose standa
# High leverage points - A data point has high leverage, if it has extreme
# predictor x values. This can be detected by examining the leverage statistic
# or the hat-value. A value of this statistic above 2(k + 1)/n indicates an
# observation with high leverage, where k is the number of predictors and n
# is the number of observations.
# Influencial values - An influential value is a value, which inclusion or
# exclusion can alter the results of the regression analysis. Such a value is
# associated with a large residual. Not all outliers (or extreme data points)
# are influential in linear regression analysis. Statisticians have developed
# a metric called Cook's distance to determine the influence of a value.
# This metric defines influence as a combination of leverage and residual size.
# Outliers and high leverage points can be identified by inspecting the
# Residuals vs Leverage plot.
# In our example, the plot highlights the top 3 most extreme points
# (#295, #4649 and #8377) with (-4,10,-5) approx standard residuals respectively.
# These are > 3 in absolute values and thus should be considered possible outliers.
# There are no points with higher leverage i.e. with leverage > than 2 times avg(leverage).
# The Residuals vs Leverage plot can help us to find influential observations if any.
# On this plot, outlying values are generally located at the upper right corner
# or at the lower right corner. Those spots are the places where data points
# can be influential against a regression line. We do not really have any
# points on the other side red dotted lines with Cook's distance >= 1.
# None on top right corner or bottom right corner. Thus, we do not see any
# outliers that are influential enough to change the regression line.


# plotting histograms of studentized.residuals and predicted values using hist()
```

```r
hist(mydata$studentized.residuals)
```




```r
hist(predict(hs_data_multi_rgr_mdl_4, newdata = mydata))
```



```
# as the histogram of residuals and predicted values have almost same shape,
# we can conclude that the residuals are normally distributed.
```

b. xv. Overall, is this regression model unbiased? If an unbiased regression model,

what does this tell us about the sample vs. the entire population model?

```
# 1. Utilizing vif values to gauge model bias
# When we check multi collinearity we check for vif score
vif(hs_data_multi_rgr_mdl_4)
```

```
##              sale_reason           building_grade square_feet_total_living
##                 1.017011                 2.362311                 2.582797
##          bath_full_count               year_built                sq_ft_lot
##                 1.576506                 1.374726                 1.114547
```

```
# None of the vif scores are near 5 or greater and thus predictors does not
# have any significant multi collinearity. Multi collinearity problems consist of
# including, in the model, different variables that have a similar predictive
# relationship with the outcome.

# Average vif
mean(vif(hs_data_multi_rgr_mdl_4))
```

```
## [1] 1.671316
```

```
# Average vif is >1 but nowhere close to 5 or greater. Model does not appear to have significant proof

# 2. We can also check bias by checking the impact of adding predictors on
# coefficient of first chosen predictor (in case of housing data - sq_ft_lot).
# If the coefficient of sq_ft_lot gets impacted much by addition of new predictors
# that means addition may cause bias.

# Coefficient of sq_ft_lot in simple regression model as base to compare
summary(hs_data_smpl_rgr_mdl)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sq_ft_lot, data = mydata)
##
## 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
```

```
#             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 ***

# Trying to create models with one chosen variable at a time and observe change in coefficient of sq_ft
hs_data_multi_rgr_mdl_5 <- lm(Sale_Price ~ sq_ft_lot+square_feet_total_living, data = mydata)

summary(hs_data_multi_rgr_mdl_5)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sq_ft_lot + square_feet_total_living,
##     data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1953643  -119763   -41295    44410  3770022
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1.903e+05  8.771e+03  21.697   <2e-16 ***
## sq_ft_lot                1.006e-01  5.737e-02   1.754   0.0795 .
## square_feet_total_living 1.844e+02  3.300e+00  55.869   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 360200 on 12862 degrees of freedom
## Multiple R-squared:  0.2068, Adjusted R-squared:  0.2067
## F-statistic:  1677 on 2 and 12862 DF,  p-value: < 2.2e-16
```

```
#(Intercept)              1.903e+05  8.771e+03  21.697   <2e-16 ***
#sq_ft_lot                1.006e-01  5.737e-02   1.754   0.0795 .
#square_feet_total_living 1.844e+02  3.300e+00  55.869   <2e-16 ***
## difference caused in coefficient of sq_ft_lot (from simple regression) = 0.75

hs_data_multi_rgr_mdl_6 <- lm(Sale_Price ~ sq_ft_lot+building_grade, data = mydata)

summary(hs_data_multi_rgr_mdl_6)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sq_ft_lot + building_grade, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1943227  -135853   -48001    53756  3762342
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -5.147e+05  2.499e+04 -20.595  < 2e-16 ***
## sq_ft_lot       4.754e-01  5.805e-02   8.191 2.84e-16 ***
```

```
## building_grade  1.414e+05  3.025e+03  46.739  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 371200 on 12862 degrees of freedom
## Multiple R-squared:  0.1575, Adjusted R-squared:  0.1573
## F-statistic:  1202 on 2 and 12862 DF,  p-value: < 2.2e-16
```

```
# (Intercept)    -5.147e+05  2.499e+04 -20.595  < 2e-16 ***
# sq_ft_lot       4.754e-01  5.805e-02   8.191 2.84e-16 ***
# building_grade  1.414e+05  3.025e+03  46.739  < 2e-16 ***
## difference caused in coefficient of sq_ft_lot (from simple regression) = 0.375

hs_data_multi_rgr_mdl_7 <- lm(Sale_Price ~ sq_ft_lot+bath_full_count, data = mydata)

summary(hs_data_multi_rgr_mdl_7)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sq_ft_lot + bath_full_count, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4128410  -161910   -49329    72518  3698383
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     3.312e+05  1.002e+04   33.05   <2e-16 ***
## sq_ft_lot       7.576e-01  5.973e-02   12.68   <2e-16 ***
## bath_full_count 1.739e+05  5.225e+03   33.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 385300 on 12862 degrees of freedom
## Multiple R-squared:  0.09249,    Adjusted R-squared:  0.09235
## F-statistic: 655.4 on 2 and 12862 DF,  p-value: < 2.2e-16
```

```
# (Intercept)     3.312e+05  1.002e+04   33.05   <2e-16 ***
# sq_ft_lot       7.576e-01  5.973e-02   12.68   <2e-16 ***
# bath_full_count 1.739e+05  5.225e+03   33.28   <2e-16 ***
# difference caused in coefficient of sq_ft_lot (from simple regression) = 0.1

hs_data_multi_rgr_mdl_8 <- lm(Sale_Price ~ sq_ft_lot+year_built, data = mydata)

summary(hs_data_multi_rgr_mdl_8)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sq_ft_lot + year_built, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2441124  -166193   -48805    74921  3634286
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.170e+07  3.991e+05  -29.32   <2e-16 ***
## sq_ft_lot    1.104e+00  6.054e-02   18.23   <2e-16 ***
## year_built   6.191e+03  2.002e+02   30.93   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 387400 on 12862 degrees of freedom
## Multiple R-squared:  0.08259,    Adjusted R-squared:  0.08245
## F-statistic:   579 on 2 and 12862 DF,  p-value: < 2.2e-16
```

```
# (Intercept) -1.170e+07  3.991e+05  -29.32   <2e-16 ***
# sq_ft_lot    1.104e+00  6.054e-02   18.23   <2e-16 ***
# year_built   6.191e+03  2.002e+02   30.93   <2e-16 ***
# difference caused in coefficient of sq_ft_lot (from simple regression) = -.25

hs_data_multi_rgr_mdl_9 <- lm(Sale_Price ~ sq_ft_lot+sale_reason, data = mydata)

summary(hs_data_multi_rgr_mdl_9)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sq_ft_lot + sale_reason, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2073083  -194249   -65029    86963  3725586
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.695e+05  4.265e+03  156.98   <2e-16 ***
## sq_ft_lot    8.802e-01  6.175e-02   14.25   <2e-16 ***
## sale_reason -1.827e+04  1.314e+03  -13.90   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 398500 on 12862 degrees of freedom
## Multiple R-squared:  0.02894,    Adjusted R-squared:  0.02879
## F-statistic: 191.7 on 2 and 12862 DF,  p-value: < 2.2e-16
```

```
# (Intercept)  6.695e+05  4.265e+03  156.98   <2e-16 ***
# sq_ft_lot    8.802e-01  6.175e-02   14.25   <2e-16 ***
# sale_reason -1.827e+04  1.314e+03  -13.90   <2e-16 ***
# difference caused in coefficient of sq_ft_lot (from simple regression) = -0.03

summary(hs_data_multi_rgr_mdl_4)
```

```
##
## Call:
## lm(formula = Sale_Price ~ sale_reason + building_grade + square_feet_total_living +
##     bath_full_count + year_built + sq_ft_lot, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2136788  -119779   -44670    41079  3720748
```

```
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -4.477e+06  4.165e+05 -10.749  < 2e-16 ***
## sale_reason             -1.156e+04  1.180e+03  -9.792  < 2e-16 ***
## building_grade           3.106e+04  4.404e+03   7.053 1.84e-12 ***
## square_feet_total_living  1.374e+02  5.083e+00  27.038  < 2e-16 ***
## bath_full_count          1.288e+04  6.040e+03   2.132    0.033 *
## year_built               2.268e+03  2.132e+02  10.640  < 2e-16 ***
## sq_ft_lot                3.132e-01  5.805e-02   5.395 6.97e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 355100 on 12858 degrees of freedom
## Multiple R-squared:  0.2293, Adjusted R-squared:  0.2289
## F-statistic: 637.6 on 6 and 12858 DF,  p-value: < 2.2e-16
```

```
#(Intercept)             -4.477e+06  4.165e+05 -10.749  < 2e-16 ***
#sale_reason             -1.156e+04  1.180e+03  -9.792  < 2e-16 ***
#building_grade           3.106e+04  4.404e+03   7.053 1.84e-12 ***
#square_feet_total_living  1.374e+02  5.083e+00  27.038  < 2e-16 ***
#bath_full_count          1.288e+04  6.040e+03   2.132    0.033 *
#year_built               2.268e+03  2.132e+02  10.640  < 2e-16 ***
#sq_ft_lot                3.132e-01  5.805e-02   5.395 6.97e-08 ***
# difference caused in coefficient of sq_ft_lot (from simple regression) = 0.54

# As we can see that most difference caused to the coefficient of sq_ft_lot
# was when we added square_feet_total_living. Difference was 0.75. There may
# be some bias in the model because of this observation but it needs further
# analysis.

# 3. We can also observe correlation between variables of the model,
# specially checking correlation between sq_ft_lot and every other
# variable individually
cor(mydata[,c("sale_reason","building_grade","square_feet_total_living",
              "bath_full_count","year_built","sq_ft_lot")], method = "pearson")
```

```
##                          sale_reason building_grade square_feet_total_living
## sale_reason               1.00000000    -0.07891229              -0.06521276
## building_grade           -0.07891229     1.00000000               0.74518052
## square_feet_total_living -0.06521276     0.74518052               1.00000000
## bath_full_count          -0.07312779     0.45941631               0.51778378
## year_built               -0.12023836     0.36192968               0.30642729
## sq_ft_lot                 0.03400556     0.13843064               0.23410450
##                          bath_full_count year_built   sq_ft_lot
## sale_reason                  -0.07312779 -0.1202384  0.03400556
## building_grade                0.45941631  0.3619297  0.13843064
## square_feet_total_living      0.51778378  0.3064273  0.23410450
## bath_full_count               1.00000000  0.4525997  0.04699255
## year_built                    0.45259966  1.0000000 -0.13491395
## sq_ft_lot                     0.04699255 -0.1349139  1.00000000
```

```
# We see that there is some correlation positive correlation between
# sq_ft_lot and square_feet_total_living. We observed that addition of
# square_feet_total_living caused most change in the coefficient of sq_ft_lot as well.
```

```
# To conclude there may be little bias in the model because of correlation
# between sq_ft_lot and square_feet_total_living. But as we see the results of
# vif stats it is not significant enough.
```