IMT 573: Problem Set 6 - Regression

Stephen Tucker

Novemebr 18th, 2022

Collaborators:

Instructions: Before beginning this assignment, please ensure you have access to R and RStudio.

1. Download the `problemset6.Rmd` file from Canvas. Open `problemset6.Rmd` in RStudio and supply your solutions to the assignment by editing `problemset6.Rmd`.
2. Replace the "Insert Your Name Here" text in the `author:` field with your own full name. Any collaborators must be listed on the top of your assignment.
3. All materials and resources that you use (with the exception of lecture slides) must be appropriately referenced within your assignment. In particular, note that Stack Overflow is licenses as Creative Commons (CC-BY-SA). This means you have to attribute any code you refer from SO.
4. Partial credit will be awarded for each question for which a serious attempt at finding an answer has been shown. But please **DO NOT** submit pages and pages of hard-to-read code and attempts that is impossible to grade. That is, avoid redundancy. Remember that one of the key goals of a data scientist is to produce coherent reports that others can easily follow. Students are *strongly* encouraged to attempt each question and to document their reasoning process even if they cannot find the correct answer. If you would like to include R code to show this process, but it does not run without errors you can do so with the `eval=FALSE` option as follows:

```
a + b # these object don't exist
# if you run this on its own it with give an error
```

6. When you have completed the assignment and have **checked** that your code both runs in the Console and knits correctly when you click `Knit PDF`, rename the knitted PDF file to `ps6_YourLastName_YourFirstName.pdf`, and submit the PDF file on Canvas.
7. Collaboration is often fun and useful, but each student must turn in an individual write-up in their own words as well as code/work that is their own. Regardless of whether you work with others, what you turn in must be your own work; this includes code and interpretation of results. The names of all collaborators must be listed on each assignment. Do not copy-and-paste from other students' responses or code.

Setup In this problem set you will need, at minimum, the following R packages.

```
# Load standard libraries
library(tidyverse)
library(AmesHousing)
```

Housing Values in Ames, Iowa

In this problem we will use the Ames Housing dataset that is available as part of the AmesHousing package. This dataset contains information about home sales in the town of Ames, Iowa. Information on variable names and other details can be found in the AmesHousing package documentation as well as here: http://jse.amstat.org/v19n3/decock/DataDocumentation.txt. Use this data to answer the following questions.

- Ames Housing dataset

Question 1: Load the package and use the `make_ames()` to store the dataset. Describe what this function does.

- This function loads data as an table data frame into a package.

```
ames_housing_data <- make_ames()
```

Question 2: Consider this data in context - what is the response variable of interest for a dataset on home sales? Filter the data to only contain observations where the Sale_Condition was "Normal." Select the following variables from the data and describe what each means: Lot_Frontage, Lot_Area, Bldg_Type, Overall_Qual, Overall_Cond, Year_Built, Gr_Liv_Area, TotRms_AbvGrd, Fireplaces, Garage_Cars, Garage_Area, Wood_Deck_SF, Total_Bsmt_SF, Full_Bath, Half_Bath, Year_Sold, and Sale_Price

- Response Variable: Sale_Price
- To get the other variable i used the filter function.

```
# str(ames_housing_data)  Checking my df  variables

# Steps:

# response variable for home sales : $ Sale_Price

# Filter the data to only contain observations where the Sale_Condition was "Normal." variable - $ Sale_

ames_housing_data_filter <- ames_housing_data %>%
  filter(Sale_Condition == "Normal")
```

question 2 continued

- Select the following variables from the data and describe what each means: Lot_Frontage, Lot_Area, Bldg_Type, Overall_Qual, Overall_Cond, Year_Built, Gr_Liv_Area, TotRms_AbvGrd, Fireplaces, Garage_Cars, Garage_Area, Wood_Deck_SF, Total_Bsmt_SF, Full_Bath, Half_Bath, Year_Sold, and Sale_Price
- Lot_Frontage :(Continuous), Linear feet of street connected to propert
- Lot_Area: (Continuous), Lot size in square feet
- Overall_Qual: (Ordinal): Rates the overall material and finish of the house
- Overall_Cond: (Ordinal): Rates the overall condition of the house
- Year_Built: Discrete): Original construction date
- Gr_Liv_Area, (Continuous): Above grade (ground) living area square feet
- TotRms_AbvGrd: (Discrete): Total rooms above grade (does not include bathrooms)
- Fireplaces: (Discrete): Number of fireplaces
- Garage_Cars: (Discrete): Size of garage in car capacity
- Garage_Area: (Continuous): Size of garage in square feet
- Wood_Deck_SF: (Continuous): Wood deck area in square feet

- Total_Bsmt_SF: (Continuous): Total square feet of basement area
- Full_Bath: (Discrete): Full bathrooms above grade
- Half_Bath: (Discrete): Half baths above grade
- Year_Sold: (Discrete): Year Sold (YYYY)
- Sale_Price: (Continuous): Sale price
- research : http://jse.amstat.org/v19n3/decock/DataDocumentation.txt

```
# code
ames_housing_data_select <- ames_housing_data_filter %>%
  select( Lot_Frontage, Lot_Area, Bldg_Type, Overall_Qual, Overall_Cond, Year_Built, Gr_Liv_Area, TotRms

# str(ames_housing_data_select)  Checking my data variables
```

Question 3: Provide a brief dive into the data and discuss any salient aspects of the variables: missingness, ranges, distributions, etc. Does each observation have complete data (Hint: you can use the complete.cases function in R)?

- after checking for true & false and creating frequency table we can see that the data is complete with not having any falses. This is pertains to the data filtered and then selected.

```
complete_data_beta <- complete.cases(ames_housing_data_select)

complete_data_beta <- table(complete_data_beta)
complete_data_beta <- data.frame(complete_data_beta)

print(complete_data_beta)
```

```
##   complete_data_beta Freq
## 1               TRUE 2413
```

Question 4: For each predictor, fit a simple (i.e. using only the one variable) linear regression model to predict the home sale price. Dummify variables as/when needed. In which of the models is there a statistically significant association between the predictor and the response? Describe your results.

- ideally if we are runnin each regression singulary it might be ideal to run some type of function to optimize time. For sake of practice it probably better if we drill it one by one.

model 1

```
model_1 <- lm(Sale_Price ~ Lot_Frontage, data = ames_housing_data_select)

summary(model_1)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Lot_Frontage, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -156174  -46841  -15925   31728  562958
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  156742.56    2757.04  56.852  < 2e-16 ***
## Lot_Frontage    339.42      42.54   7.979 2.26e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 70080 on 2411 degrees of freedom
## Multiple R-squared:  0.02573,    Adjusted R-squared:  0.02532
## F-statistic: 63.67 on 1 and 2411 DF,  p-value: 2.256e-15
```

model 2

```
model_2 <- lm(Sale_Price ~ Lot_Area, data = ames_housing_data_select)

summary(model_2)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Lot_Area, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -313698  -42704  -15584   28579  552187
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.517e+05  2.196e+03   69.08   <2e-16 ***
## Lot_Area    2.374e+00  1.690e-01   14.05   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 68260 on 2411 degrees of freedom
## Multiple R-squared:  0.07566,    Adjusted R-squared:  0.07528
## F-statistic: 197.4 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 3

```
model_3 <- lm(Sale_Price ~ Overall_Qual, data = ames_housing_data_select)

summary(model_3)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Overall_Qual, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -167813  -22719   -3026   19087  277187
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                    60600      23076   2.626  0.00869 **
## Overall_QualPoor               -2037      26311  -0.077  0.93829
## Overall_QualFair               25927      24134   1.074  0.28281
## Overall_QualBelow_Average      48313      23263   2.077  0.03792 *
## Overall_QualAverage            75426      23125   3.262  0.00112 **
## Overall_QualAbove_Average     102120      23131   4.415 1.06e-05 ***
## Overall_QualGood              144010      23147   6.222 5.79e-10 ***
## Overall_QualVery_Good         205367      23211   8.848  < 2e-16 ***
## Overall_QualExcellent         289922      23611  12.279  < 2e-16 ***
## Overall_QualVery_Excellent    417213      25279  16.504  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 39970 on 2403 degrees of freedom
## Multiple R-squared:  0.6841, Adjusted R-squared:  0.6829
## F-statistic: 578.2 on 9 and 2403 DF,  p-value: < 2.2e-16
```

model 4

```r
model_4 <- lm(Sale_Price ~ Overall_Cond, data = ames_housing_data_select)

summary(model_4)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Overall_Cond, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -143463  -38774  -11463   26537  604332
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   73000      37921   1.925 0.054338 .
## Overall_CondPoor              27233      46443   0.586 0.557675
## Overall_CondFair              20268      39469   0.514 0.607630
## Overall_CondBelow_Average     45231      38608   1.172 0.241501
## Overall_CondAverage          125463      37965   3.305 0.000965 ***
## Overall_CondAbove_Average     77668      38040   2.042 0.041287 *
## Overall_CondGood              81774      38082   2.147 0.031868 *
## Overall_CondVery_Good         82239      38328   2.146 0.031999 *
## Overall_CondExcellent        128703      39352   3.271 0.001089 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 65680 on 2404 degrees of freedom
## Multiple R-squared:  0.1466, Adjusted R-squared:  0.1437
## F-statistic: 51.61 on 8 and 2404 DF,  p-value: < 2.2e-16
```

model 5

```r
model_5 <- lm(Sale_Price ~ Year_Built, data = ames_housing_data_select)
summary(model_5)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Year_Built, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -108801  -37058  -12808   21710  547909
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.352e+06  8.169e+04  -28.79   <2e-16 ***
## Year_Built   1.283e+03  4.147e+01   30.94   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60060 on 2411 degrees of freedom
## Multiple R-squared:  0.2843, Adjusted R-squared:  0.284
## F-statistic: 957.5 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 6

```r
model_6 <- lm(Sale_Price ~ Gr_Liv_Area, data = ames_housing_data_select)
summary(model_6)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Gr_Liv_Area, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -178533  -27144    -653   21560  332083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15784.332   3148.708   5.013 5.75e-07 ***
## Gr_Liv_Area   108.151      2.026  53.388  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48060 on 2411 degrees of freedom
## Multiple R-squared:  0.5417, Adjusted R-squared:  0.5416
## F-statistic:  2850 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 7

```r
model_7 <- lm(Sale_Price ~ TotRms_AbvGrd, data = ames_housing_data_select)
summary(model_7)
```

```
##
```

```
## Call:
## lm(formula = Sale_Price ~ TotRms_AbvGrd, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -202752  -35059  -10059   25941  495146
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      27867       5406   5.154 2.75e-07 ***
## TotRms_AbvGrd    23199        826  28.086  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61630 on 2411 degrees of freedom
## Multiple R-squared:  0.2465, Adjusted R-squared:  0.2462
## F-statistic: 788.8 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 8

```
model_8 <- lm(Sale_Price ~ Fireplaces, data = ames_housing_data_select)
summary(model_8)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Fireplaces, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -140961  -37094   -9026   28006  504039
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   143026       1716   83.34   <2e-16 ***
## Fireplaces     53967       1938   27.85   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61750 on 2411 degrees of freedom
## Multiple R-squared:  0.2434, Adjusted R-squared:  0.2431
## F-statistic: 775.7 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 9

```
model_9 <- lm(Sale_Price ~ Garage_Cars, data = ames_housing_data_select)
summary(model_9)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Garage_Cars, data = ames_housing_data_select)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -250491  -33093   -5093   22907  501274
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    68828       2888   23.83   <2e-16 ***
## Garage_Cars    61633       1537   40.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54990 on 2411 degrees of freedom
## Multiple R-squared:    0.4,  Adjusted R-squared:  0.3997
## F-statistic:  1607 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 10

```r
model_10 <- lm(Sale_Price ~ Garage_Area, data = ames_housing_data_select)
summary(model_10)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Garage_Area, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -266768  -31604   -4646   24315  498285
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 74609.667   2825.596   26.41   <2e-16 ***
## Garage_Area   218.877      5.613   38.99   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55600 on 2411 degrees of freedom
## Multiple R-squared:  0.3868, Adjusted R-squared:  0.3865
## F-statistic:  1521 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 11

```r
model_11 <- lm(Sale_Price ~ Wood_Deck_SF, data = ames_housing_data_select)
summary(model_11)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Wood_Deck_SF, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -240983  -42770  -12651   28626  524838
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  157274.1     1681.1   93.55   <2e-16 ***
## Wood_Deck_SF    190.8       10.4   18.35   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 66500 on 2411 degrees of freedom
## Multiple R-squared:  0.1226, Adjusted R-squared:  0.1222
## F-statistic: 336.9 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 12

```
model_12 <- lm(Sale_Price ~ Total_Bsmt_SF, data = ames_housing_data_select)
summary(model_12)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Total_Bsmt_SF, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -161767  -37298  -12465   32574  420996
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   61539.792   2983.952   20.62   <2e-16 ***
## Total_Bsmt_SF   111.483      2.709   41.15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54410 on 2411 degrees of freedom
## Multiple R-squared:  0.4126, Adjusted R-squared:  0.4124
## F-statistic:  1694 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 13

```
model_13 <- lm(Sale_Price ~ Full_Bath, data = ames_housing_data_select)
summary(model_13)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Full_Bath, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -192509  -32019   -7516   19984  477981
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    65523       3623   18.09   <2e-16 ***
## Full_Bath      71497       2219   32.22   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59360 on 2411 degrees of freedom
## Multiple R-squared:  0.301,  Adjusted R-squared:  0.3007
## F-statistic:  1038 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 14

```r
model_14 <- lm(Sale_Price ~ Half_Bath, data = ames_housing_data_select)
summary(model_14)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Half_Bath, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -140460  -40836  -16460   25664  553540
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   159836       1735   92.14   <2e-16 ***
## Half_Bath      41624       2773   15.01   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 67890 on 2411 degrees of freedom
## Multiple R-squared:  0.08545,    Adjusted R-squared:  0.08507
## F-statistic: 225.3 on 1 and 2411 DF,  p-value: < 2.2e-16
```

model 15

```r
model_15 <- lm(Sale_Price ~ Year_Sold, data = ames_housing_data_select)
summary(model_15)
```

```
##
## Call:
## lm(formula = Sale_Price ~ Year_Sold, data = ames_housing_data_select)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -138274  -45957  -16957   30543  580498
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2289345    2217522  -1.032    0.302
## Year_Sold       1228       1104   1.112    0.266
##
## Residual standard error: 70980 on 2411 degrees of freedom
## Multiple R-squared:  0.0005122,  Adjusted R-squared:  9.766e-05
## F-statistic: 1.236 on 1 and 2411 DF,  p-value: 0.2664
```

Question 5: Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis $H_0 : \beta_j = 0$?

model 16

```
model_16 <- lm(Sale_Price ~., data = ames_housing_data_select)
summary(model_16)
```

```
##
## Call:
## lm(formula = Sale_Price ~ ., data = ames_housing_data_select)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -98710 -11944   -901  10569 154687
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -5.941e+05  6.834e+05  -0.869  0.38474
## Lot_Frontage                3.844e+01  1.415e+01   2.716  0.00665 **
## Lot_Area                    6.651e-01  5.944e-02  11.190  < 2e-16 ***
## Bldg_TypeTwoFmCon          -7.688e+03  3.175e+03  -2.421  0.01555 *
## Bldg_TypeDuplex            -2.001e+04  2.723e+03  -7.347 2.77e-13 ***
## Bldg_TypeTwnhs             -1.616e+04  2.494e+03  -6.480 1.11e-10 ***
## Bldg_TypeTwnhsE            -7.889e+03  1.890e+03  -4.174 3.11e-05 ***
## Overall_QualPoor            7.418e+03  1.484e+04   0.500  0.61734
## Overall_QualFair            2.489e+03  1.409e+04   0.177  0.85978
## Overall_QualBelow_Average  -6.596e+03  1.378e+04  -0.479  0.63210
## Overall_QualAverage        -4.781e+03  1.375e+04  -0.348  0.72799
## Overall_QualAbove_Average  -1.468e+02  1.379e+04  -0.011  0.99151
## Overall_QualGood            1.292e+04  1.385e+04   0.933  0.35069
## Overall_QualVery_Good       4.471e+04  1.396e+04   3.203  0.00138 **
## Overall_QualExcellent       1.056e+05  1.424e+04   7.415 1.69e-13 ***
## Overall_QualVery_Excellent  1.848e+05  1.516e+04  12.191  < 2e-16 ***
## Overall_CondPoor            2.069e+03  1.613e+04   0.128  0.89796
## Overall_CondFair            1.737e+04  1.386e+04   1.253  0.21034
## Overall_CondBelow_Average   2.755e+04  1.385e+04   1.989  0.04683 *
## Overall_CondAverage         4.046e+04  1.375e+04   2.944  0.00328 **
## Overall_CondAbove_Average   4.511e+04  1.374e+04   3.282  0.00104 **
## Overall_CondGood            5.748e+04  1.375e+04   4.181 3.00e-05 ***
## Overall_CondVery_Good       6.354e+04  1.383e+04   4.596 4.53e-06 ***
## Overall_CondExcellent       6.997e+04  1.413e+04   4.951 7.91e-07 ***
## Year_Built                  5.793e+02  2.605e+01  22.237  < 2e-16 ***
## Gr_Liv_Area                 6.452e+01  2.265e+00  28.482  < 2e-16 ***
## TotRms_AbvGrd              -3.699e+03  5.457e+02  -6.779 1.52e-11 ***
## Fireplaces                  7.577e+03  8.216e+02   9.223  < 2e-16 ***
## Garage_Cars                 3.346e+03  1.435e+03   2.332  0.01980 *
## Garage_Area                 1.170e+01  5.002e+00   2.338  0.01946 *
## Wood_Deck_SF                1.102e+01  3.696e+00   2.981  0.00291 **
## Total_Bsmt_SF               2.870e+01  1.507e+00  19.043  < 2e-16 ***
## Full_Bath                  -1.983e+03  1.321e+03  -1.501  0.13346
## Half_Bath                   9.286e+02  1.210e+03   0.768  0.44271
## Year_Sold                  -2.726e+02  3.398e+02  -0.802  0.42251
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21680 on 2378 degrees of freedom
## Multiple R-squared:  0.908,  Adjusted R-squared:  0.9067
## F-statistic: 690.5 on 34 and 2378 DF,  p-value: < 2.2e-16
```

Question 6: How do your results from (4) compare to your results from (5)? You need to compare the coefficients across the two models and report on the changes you observe and reasons why. What happened to the coefficients? What happened to the p-values? Why?

- It appears that where some of my coefficients were strongly statistically significant when tested in isolation, when tested in the multiple linear regression some variables co-coefficients become less significant. Another caveat of the co-coefficients are that when done in singularity there unit increase has a much larger impact compared to when done in the Multiple Linear Regression. The reason why it changes is possibly due to how much data it captures. I kept showed an example of a co-efficient to show the overall trend of what happens with the creation of S.R models and Multiple regression models.

Regarding the P value stays relatively the same in comparison to what it was in the singular models but it varied depending on the variable input.

Lot_Frontage 339.42 -S.R | 3.844e+01 - M.R

Lot_Area 2.374e+00 S.R | 6.651e-01 M.R SR - Singular regression | MR- Multiple Linear Regression

- Overall_QualPoor | S.R -2037 S.R | M.R 7.418e+03
- Overall_QualFair | S.R 25927 | M.R 2.489e+03
- Overall_QualBelow_Average | S.R 48313 | M.R -6.596e+03