Introduction:

The purpose of this assignment is to determine 2 variables to use for initial regression modelling using the data from week 1. The regressions will be performed against the Sales Price of the training set of data to determine if we can see any predicting metrics to determine future sales prices of houses within the Ames, Iowa market. The same steps that were taken in Assignment 1 to subset and create the training set of data will be used in assignment 2.

File:ames_housing_data.csv

Code file: Assignment2.R

Task:

Section 1: Define eligible population for sample output and selection 2 predictors

We will continue with the eligible population set from week one where we used single family homes as our subset of data to perform our modelling efforts. The sample group will be defined only within the Ames, Iowa market. This is also considered as our drop conditions when filtering out the records that we don't need for our analysis. We first concluded a set of conditions that labeled the records for what was considered as "drop conditions" for an appropriate housing subset.

Condition	Count
Not SFR	505
Non-Normal Sale	423
Street Not Paved	6
Built Pre-1950	489
No Basement	28
LT 800 SqFt	9
Eligible Sample	1,470
Total	**2,930**

Here we can determine that we will start with 1,470 homes that are eligible for further analysis.

Section 2:Simple Linear Regression Modelling

The two variables that were selected were the total house square footage and total quality index

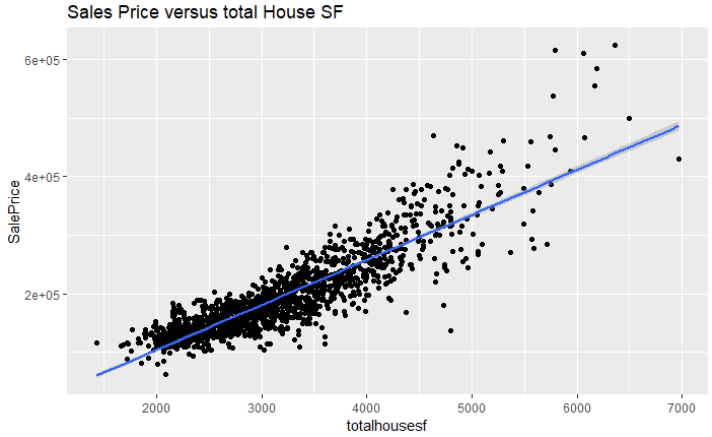
Total house square footage:

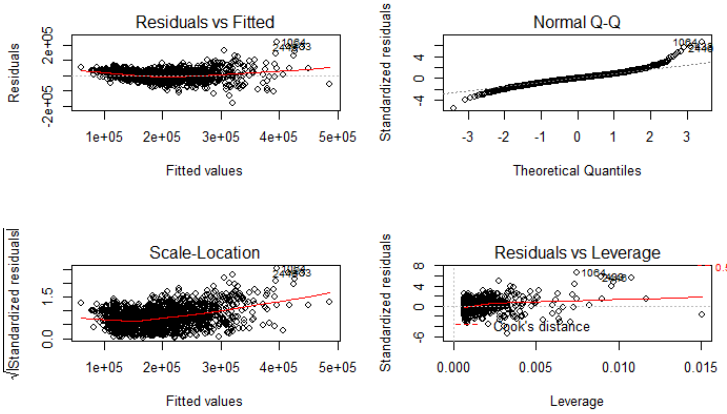
Simple linear regression of total house square footage against sales price.

```
Residuals:
    Min      1Q  Median      3Q     Max
-181631  -18098    -372   17624  219111

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -49200.584   3486.873  -14.11   <2e-16 ***
totalhousesf     76.819      1.058   72.62   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33420 on 1467 degrees of freedom
Multiple R-squared:  0.7824,    Adjusted R-squared:  0.7822
F-statistic:  5274 on 1 and 1467 DF,  p-value: < 2.2e-16
```





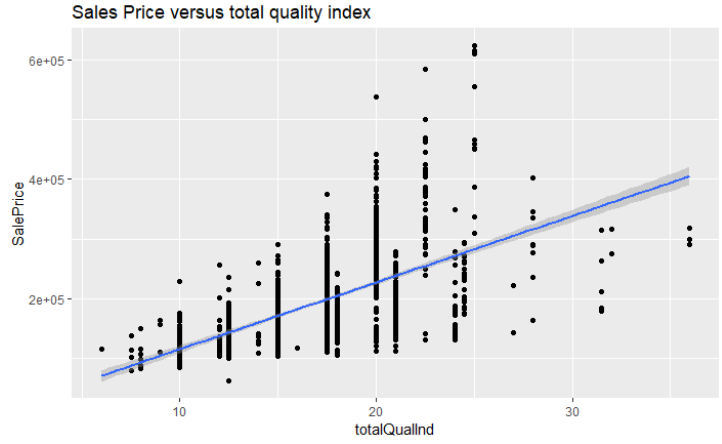
Total Quality Index:

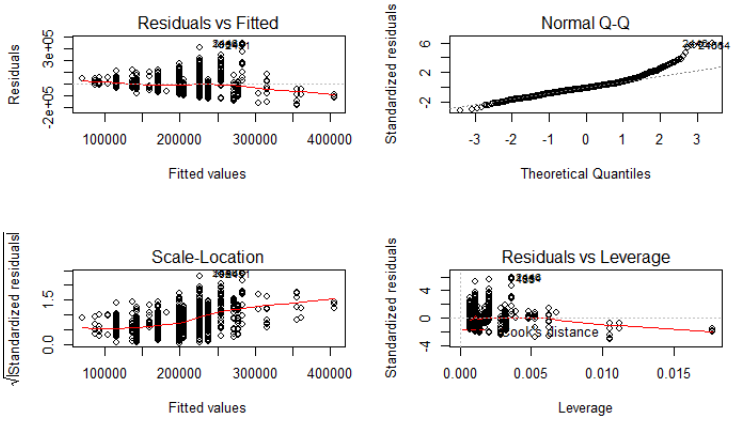
Simple linear regression of total quality index against sales price.

```
Residuals:
    Min      1Q  Median      3Q     Max
-176913  -33480   -6063   25920  342270

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3563.3     7122.7    0.50    0.617
totalQualInd  11166.7      403.9   27.65   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 58090 on 1467 degrees of freedom
Multiple R-squared:  0.3425,    Adjusted R-squared:  0.3421
F-statistic: 764.3 on 1 and 1467 DF,  p-value: < 2.2e-16
```





Multiple Linear Regression Model:

```
Residuals:
    Min      1Q  Median      3Q     Max
-126833  -17546    -371   15807  209921

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -92524.839   4040.436  -22.90   <2e-16 ***
totalQualInd   4161.499    239.793   17.36   <2e-16 ***
totalhousesf     67.926      1.092   62.23   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30450 on 1466 degrees of freedom
Multiple R-squared:  0.8195,    Adjusted R-squared:  0.8192
F-statistic:  3327 on 2 and 1466 DF,  p-value: < 2.2e-16
```