PROBLEM STATEMENT: The most important question is how and why the square foot prices has been differing a lot, as it's charged a lot when it comes to city view are and when it comes to village area some prices has been dropped drastically.

1. Neighborhood to directly c0-related to the amenities and the more the amenities they have the more the price has been increased, so question is how so we can say that facilities like T-station, grocery shops etc.
2. The prices are depending on Number of rooms, floors, baths, and condition of house & Kitchen too.
3. As, prices has not been decreed during the home crisis which was happened in 2008 in US.
4. As, the factors that affecting housing prices are property size, how old the home are, the condition of house, number of rooms, amenities, condition of structure.
5. As, we can't conclude the ground area officially because in the US they don't conclude basement square footage but it's an important factor when you see it technique.
6. Age and location are inversely proportional to each other as it completed one year how pp has been reduce $400 but if location was premium the price has been increased .

```
#There's are 2 categorical variable i.e,. overall.cond and overall.qual .

Q1 <- quantile(ames$SalePrice, na.rm=F, probs = c(.20,.40,.60,.80))
Q2 <- quantile(ames$Gr.Liv.Area,na.rm=F,probs = c(.20,.40,.60,.80))
Q3<-quantile(ames$Overall.Cond,na.rm=F,probs = c(.20,.40,.60,.80))
Q4<-quantile(ames$Overall.Qual,na.rm=F,probs = c(.20,.40,.60,.80))
Q5<- quantile(ames$Total.Bsmt.SF,na.rm=T,probs = c(.20,.40,.60,.80))
```

In this data Set, we are having two categorical variables i.e, overall. cond and overall. qual. as it's in Integer Format we converted these in factor format. Also, according to the data set following factors are affecting the Sale prices of the data set as Sale Price, Ground Living area, Overall condition, overall quality, and Total square feet of the basement area. In overall quality and condition, variable data is overall divided into ratings. In the basement square foot variable, the data has been showing what was the per square foot area has been used/present. Here we uses quantile function where variable is divided into four parts. sale price of 40% is 230000 (Two Hundred and Thirty Thousand)and 20% is 124000 (one hundred twenty-four thousand). In ground living area 80% is 1838(one thousand eight hundred and thirty-eighth). In overall condition 5 means"Average" and 6 means"Above-Average" so 80% having only 6 and rest of three having 5. In overall quality variable having 5,6.7 7 means"GOOD", 20% having 5, 40 and 60% having 6 and 80% having 7. In Total square feet of basement area , 20% occupy 741 SPFA, 40% covers 912 SPFA, 60% occupy 1090 SPFA and 80% occupy 1392 SFPA.

```
#counts tablea
matrix(c(Q1,Q2,Q3,Q4,Q5),ncol=2)
```

```
##           [,1] [,2]
##  [1,] 124000.0    5
##  [2,] 146500.0    6
##  [3,] 178536.0    5
##  [4,] 230000.0    6
##  [5,]   1064.8    6
##  [6,]   1327.6    7
##  [7,]   1560.0  741
##  [8,]   1838.0  912
##  [9,]      5.0 1090
## [10,]      5.0 1392
```

```
chisq.test(matrix(c(Q1,Q2,Q3,Q4,Q5),ncol=2))
```

```
##
##  Pearson's Chi-squared test
##
## data:  matrix(c(Q1, Q2, Q3, Q4, Q5), ncol = 2)
## X-squared = 496704, df = 9, p-value < 2.2e-16
```

The p- value is very low so we reject the Null Hypothesis.

```
ROOF_STYLE <- as.numeric(factor(ames$Roof.Style))
cor(ames$SalePrice, ROOF_STYLE)
```

```
## [1] 0.2571701
```

```
FOUNDATION <- as.numeric(factor(ames$Foundation))
cor(ames$SalePrice, FOUNDATION)
```

```
## [1] 0.4066008
```

```
NEIBOURHOOD <- as.numeric(factor(ames$Neighborhood))
```

In roof style the numbers is assigned Flat (6),Gable(5),Gabrel,(Barn)(4),Hip(3),Mansard(2),Shed(1) as it's shown from the report people preferred the Mansard and Hip Roof style. In foundation variable it it shown that wood and PConc is mostly used .

It looks like more sales in summer. prices are higher in summer. Also,there was a HOME CRISIS IN the US (2008) as it doesn't look like home prices have been changed over the period. The ground living area is almost around 500 square to 3000 square feet. In total rooms above ground the prices is been increases. In Basement finished area it's shown from the graph if the Basement size of the Basement increases then prices also increase.

```
summary(lm(SalePrice ~ Bsmt.Unf.SF + BsmtFin.SF.1, data = ames))
```

```
##
## Call:
## lm(formula = SalePrice ~ Bsmt.Unf.SF + BsmtFin.SF.1, data = ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -627176  -41436   -9580   32175  468409
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  77164.169   2803.297   27.53   <2e-16 ***
## Bsmt.Unf.SF     91.798      3.001   30.58   <2e-16 ***
## BsmtFin.SF.1   118.220      2.895   40.83   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 62700 on 2926 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.3843, Adjusted R-squared:  0.3838
## F-statistic:   913 on 2 and 2926 DF,  p-value: < 2.2e-16
```

```
summary(lm(SalePrice ~ Bsmt.Unf.SF + ames$Total.Bsmt.SF, data = ames))
```

```
##
## Call:
## lm(formula = SalePrice ~ Bsmt.Unf.SF + ames$Total.Bsmt.SF, data = ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -637449  -40115  -13480   33516  412158
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        62449.497   2964.692  21.064  < 2e-16 ***
## Bsmt.Unf.SF          -16.954      2.839  -5.972 2.63e-09 ***
## ames$Total.Bsmt.SF   121.588      2.832  42.934  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61530 on 2926 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.407,  Adjusted R-squared:  0.4066
## F-statistic:  1004 on 2 and 2926 DF,  p-value: < 2.2e-16
```

```
summary(lm(SalePrice ~ Bsmt.Unf.SF + BsmtFin.SF.1 + Total.Bsmt.SF, data = ames))
```

```
##
## Call:
## lm(formula = SalePrice ~ Bsmt.Unf.SF + BsmtFin.SF.1 + Total.Bsmt.SF,
##     data = ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -657631  -39072  -12586   32613  407494
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   63643.177   2956.222  21.529  < 2e-16 ***
## Bsmt.Unf.SF      18.129      6.764   2.680  0.00739 **
## BsmtFin.SF.1     40.223      7.046   5.709 1.25e-08 ***
## Total.Bsmt.SF    84.864      7.023  12.084  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61200 on 2925 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.4135, Adjusted R-squared:  0.4129
## F-statistic: 687.5 on 3 and 2925 DF,  p-value: < 2.2e-16
```

In the first Interpretation, we see clearly that the unfinished basement value is 108$ per edition square feet and the finished basement value is 125$ per square foot. As the standard error is 55190 so the sale price is increasing in dollars 1100. In the second Interpretation, it increases the number of square feet while holding the total basement, here if we are converting unfinished basement to finished basement it increases the prices by $17.62 cents per square. In the third model total basement, SF are having no values in our data set so I put NA values.

```
m1 <- lm(ames$SalePrice ~ ames$Gr.Liv.Area + ames$Bsmt.Unf.SF +ames$BsmtFin.SF.1+ ames$Full.Bath
+ames$Half.Bath)
m1
```

```
##
## Call:
## lm(formula = ames$SalePrice ~ ames$Gr.Liv.Area + ames$Bsmt.Unf.SF +
##     ames$BsmtFin.SF.1 + ames$Full.Bath + ames$Half.Bath)
##
## Coefficients:
##       (Intercept)   ames$Gr.Liv.Area   ames$Bsmt.Unf.SF  ames$BsmtFin.SF.1
##         -20920.86              62.40              48.95              81.78
##    ames$Full.Bath     ames$Half.Bath
##          24630.39           15799.49
```

```
summary(m1)
```

```
##
## Call:
## lm(formula = ames$SalePrice ~ ames$Gr.Liv.Area + ames$Bsmt.Unf.SF +
##     ames$BsmtFin.SF.1 + ames$Full.Bath + ames$Half.Bath)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -720552  -21561   -1602   19668  271106
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -20920.857   3040.305  -6.881 7.23e-12 ***
## ames$Gr.Liv.Area      62.399      2.677  23.305  < 2e-16 ***
## ames$Bsmt.Unf.SF      48.949      2.561  19.113  < 2e-16 ***
## ames$BsmtFin.SF.1     81.777      2.426  33.713  < 2e-16 ***
## ames$Full.Bath     24630.386   2081.264  11.834  < 2e-16 ***
## ames$Half.Bath     15799.487   2028.738   7.788 9.39e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47200 on 2923 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.6515, Adjusted R-squared:  0.6509
## F-statistic:  1093 on 5 and 2923 DF,  p-value: < 2.2e-16
```

```
old_town <- subset(ames, ames$Neighborhood=="OldTown")
summary(old_town)
```

```
##     ï..Order           PID             MS.SubClass      MS.Zoning
##  Min.   : 159.0   Min.   :535353130   Min.   : 20.00   Length:239
##  1st Qu.: 709.5   1st Qu.:902105085   1st Qu.: 30.00   Class :character
##  Median :1314.0   Median :902206020   Median : 50.00   Mode  :character
##  Mean   :1454.3   Mean   :842529958   Mean   : 63.95
##  3rd Qu.:2004.5   3rd Qu.:902400100   3rd Qu.: 70.00
##  Max.   :2690.0   Max.   :903476110   Max.   :190.00
##
##   Lot.Frontage       Lot.Area        Street             Alley
##  Min.   : 30.00   Min.   : 2500   Length:239         Length:239
##  1st Qu.: 52.00   1st Qu.: 5812   Class :character   Class :character
##  Median : 60.00   Median : 7800   Mode  :character   Mode  :character
##  Mean   : 61.78   Mean   : 8241
##  3rd Qu.: 63.00   3rd Qu.:10128
##  Max.   :153.00   Max.   :33120
##  NA's   :10
##   Lot.Shape         Land.Contour        Utilities          Lot.Config
##  Length:239         Length:239         Length:239         Length:239
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   Land.Slope        Neighborhood       Condition.1        Condition.2
##  Length:239         Length:239         Length:239         Length:239
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   Bldg.Type         House.Style         Overall.Qual     Overall.Cond
##  Length:239         Length:239         Min.   : 2.000   Min.   :1.000
##  Class :character   Class :character   1st Qu.: 5.000   1st Qu.:5.000
##  Mode  :character   Mode  :character   Median : 5.000   Median :6.000
##                                        Mean   : 5.205   Mean   :6.301
##                                        3rd Qu.: 6.000   3rd Qu.:7.000
##                                        Max.   :10.000   Max.   :9.000
##
##    Year.Built   Year.Remod.Add  Roof.Style         Roof.Matl
##  Min.   :1872   Min.   :1950   Length:239         Length:239
##  1st Qu.:1910   1st Qu.:1950   Class :character   Class :character
##  Median :1920   Median :1968   Mode  :character   Mode  :character
##  Mean   :1922   Mean   :1974
##  3rd Qu.:1938   3rd Qu.:2000
##  Max.   :2008   Max.   :2009
##
##  Exterior.1st       Exterior.2nd       Mas.Vnr.Type        Mas.Vnr.Area
##  Length:239         Length:239         Length:239         Min.   :  0.00
##  Class :character   Class :character   Class :character   1st Qu.:  0.00
##  Mode  :character   Mode  :character   Mode  :character   Median :  0.00
```

```
##                                                      Mean   : 14.29
##                                                      3rd Qu.:  0.00
##                                                      Max.   :771.00
##
##   Exter.Qual         Exter.Cond         Foundation         Bsmt.Qual
##  Length:239         Length:239         Length:239         Length:239
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   Bsmt.Cond         Bsmt.Exposure      BsmtFin.Type.1      BsmtFin.SF.1
##  Length:239         Length:239         Length:239         Min.   :   0.0
##  Class :character   Class :character   Class :character   1st Qu.:   0.0
##  Mode  :character   Mode  :character   Mode  :character   Median :   0.0
##                                                          Mean   : 140.1
##                                                          3rd Qu.: 240.5
##                                                          Max.   :1088.0
##
##  BsmtFin.Type.2      BsmtFin.SF.2      Bsmt.Unf.SF     Total.Bsmt.SF
##  Length:239         Min.   :   0.00   Min.   :   0.0   Min.   :   0.0
##  Class :character   1st Qu.:   0.00   1st Qu.: 422.5   1st Qu.: 662.0
##  Mode  :character   Median :   0.00   Median : 672.0   Median : 788.0
##                     Mean   :  21.33   Mean   : 636.2   Mean   : 797.7
##                     3rd Qu.:   0.00   3rd Qu.: 836.0   3rd Qu.: 945.0
##                     Max.   :1073.00   Max.   :1595.0   Max.   :2171.0
##
##    Heating           Heating.QC        Central.Air         Electrical
##  Length:239         Length:239         Length:239         Length:239
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   X1st.Flr.SF      X2nd.Flr.SF     Low.Qual.Fin.SF    Gr.Liv.Area
##  Min.   : 407.0   Min.   :   0.0   Min.   :   0.00   Min.   : 407
##  1st Qu.: 792.0   1st Qu.:   0.0   1st Qu.:   0.00   1st Qu.:1036
##  Median : 908.0   Median : 544.0   Median :   0.00   Median :1374
##  Mean   : 945.3   Mean   : 470.2   Mean   :  16.48   Mean   :1432
##  3rd Qu.:1074.0   3rd Qu.: 734.5   3rd Qu.:   0.00   3rd Qu.:1700
##  Max.   :1766.0   Max.   :1818.0   Max.   :1064.00   Max.   :3608
##
##  Bsmt.Full.Bath   Bsmt.Half.Bath      Full.Bath       Half.Bath
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:1.000   1st Qu.:0.0000
##  Median :0.0000   Median :0.00000   Median :1.000   Median :0.0000
##  Mean   :0.1506   Mean   :0.04603   Mean   :1.331   Mean   :0.1715
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:2.000   3rd Qu.:0.0000
##  Max.   :2.0000   Max.   :1.00000   Max.   :3.000   Max.   :1.0000
##
```

```
##  Bedroom.AbvGr   Kitchen.AbvGr  Kitchen.Qual       TotRms.AbvGrd
##  Min.   :0.000   Min.   :1.00   Length:239         Min.   : 3.000
##  1st Qu.:2.000   1st Qu.:1.00   Class :character   1st Qu.: 5.000
##  Median :3.000   Median :1.00   Mode  :character   Median : 6.000
##  Mean   :2.845   Mean   :1.13                      Mean   : 6.485
##  3rd Qu.:3.000   3rd Qu.:1.00                      3rd Qu.: 8.000
##  Max.   :6.000   Max.   :3.00                      Max.   :12.000
##
##   Functional          Fireplaces     Fireplace.Qu       Garage.Type
##  Length:239         Min.   :0.0000   Length:239         Length:239
##  Class :character   1st Qu.:0.0000   Class :character   Class :character
##  Mode  :character   Median :0.0000   Mode  :character   Mode  :character
##                     Mean   :0.3013
##                     3rd Qu.:1.0000
##                     Max.   :2.0000
##
##  Garage.Yr.Blt  Garage.Finish       Garage.Cars     Garage.Area
##  Min.   :1900   Length:239         Min.   :0.000   Min.   :   0.0
##  1st Qu.:1927   Class :character   1st Qu.:1.000   1st Qu.: 228.0
##  Median :1950   Mode  :character   Median :1.000   Median : 331.0
##  Mean   :1950                      Mean   :1.356   Mean   : 368.5
##  3rd Qu.:1968                      3rd Qu.:2.000   3rd Qu.: 520.0
##  Max.   :2008                      Max.   :4.000   Max.   :1488.0
##  NA's   :31
##  Garage.Qual        Garage.Cond        Paved.Drive         Wood.Deck.SF
##  Length:239         Length:239         Length:239         Min.   :  0.00
##  Class :character   Class :character   Class :character   1st Qu.:  0.00
##  Mode  :character   Mode  :character   Mode  :character   Median :  0.00
##                                                           Mean   : 43.02
##                                                           3rd Qu.: 24.00
##                                                           Max.   :631.00
##
##  Open.Porch.SF    Enclosed.Porch    X3Ssn.Porch        Screen.Porch
##  Min.   :  0.00   Min.   :  0.00   Min.   :  0.0000   Min.   :  0.000
##  1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:  0.0000   1st Qu.:  0.000
##  Median :  0.00   Median : 26.00   Median :  0.0000   Median :  0.000
##  Mean   : 43.57   Mean   : 75.27   Mean   :  0.5858   Mean   :  9.795
##  3rd Qu.: 49.00   3rd Qu.:129.00   3rd Qu.:  0.0000   3rd Qu.:  0.000
##  Max.   :547.00   Max.   :432.00   Max.   :140.0000   Max.   :480.000
##
##    Pool.Area   Pool.QC             Fence           Misc.Feature
##  Min.   :0   Length:239         Length:239         Length:239
##  1st Qu.:0   Class :character   Class :character   Class :character
##  Median :0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :0
##  3rd Qu.:0
##  Max.   :0
##
##     Misc.Val          Mo.Sold          Yr.Sold      Sale.Type
##  Min.   :   0.00   Min.   : 1.000   Min.   :2006   Length:239
##  1st Qu.:   0.00   1st Qu.: 5.000   1st Qu.:2007   Class :character
##  Median :   0.00   Median : 6.000   Median :2008   Mode  :character
```

```
##  Mean   :  42.68   Mean   : 6.335   Mean   :2008
##  3rd Qu.:   0.00   3rd Qu.: 8.000   3rd Qu.:2009
##  Max.   :4500.00   Max.   :12.000   Max.   :2010
##
##  Sale.Condition       SalePrice
##  Length:239         Min.   : 12789
##  Class :character   1st Qu.:103350
##  Mode  :character   Median :119900
##                     Mean   :123992
##                     3rd Qu.:140000
##                     Max.   :475000
##
```

In these total rooms above ground price is also decreasing and full bath prices are decreasing and half bath prices are increasing.

```
M12 <- lm(ames$SalePrice ~ ames$Gr.Liv.Area + ames$Bsmt.Unf.SF +ames$BsmtFin.SF.1  + ames$Bedroo
m.AbvGr + ames$TotRms.AbvGrd, data = ames)
summary(M12)
```

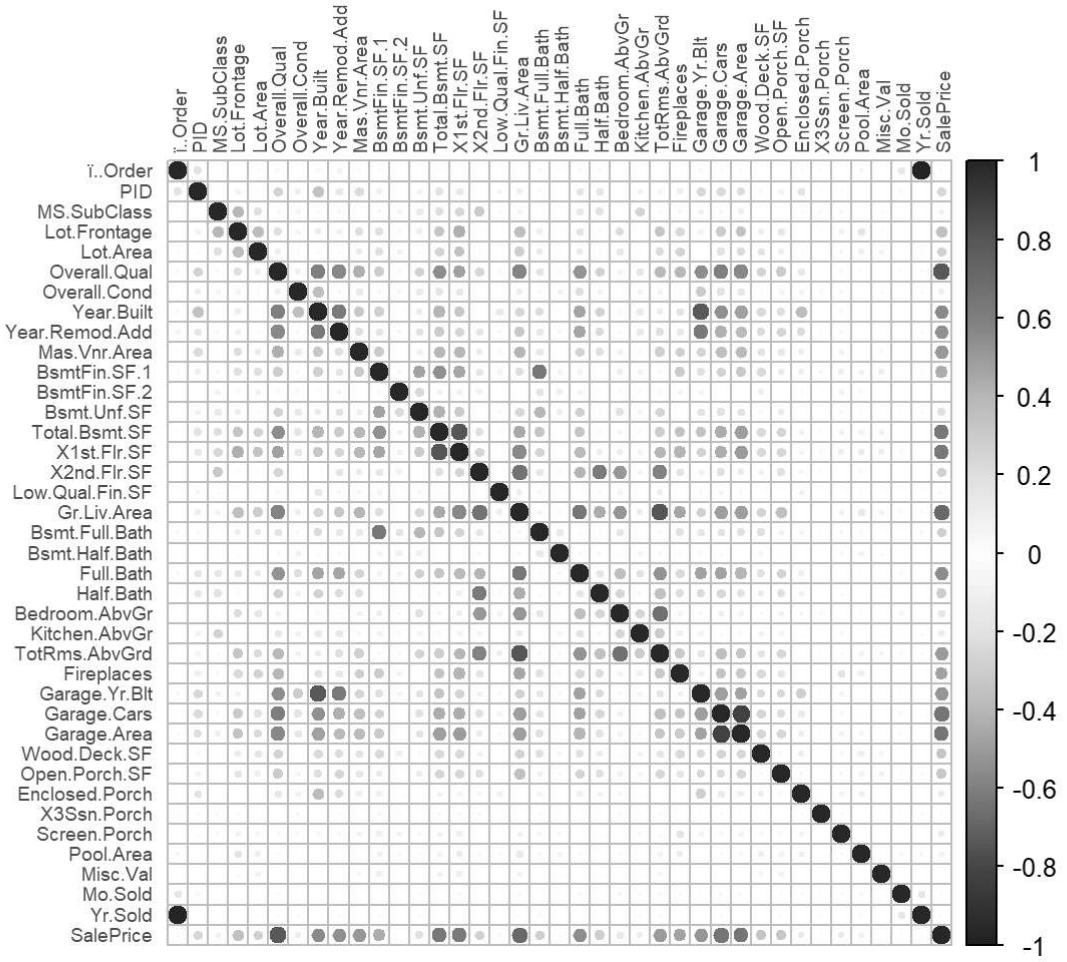
```
##
## Call:
## lm(formula = ames$SalePrice ~ ames$Gr.Liv.Area + ames$Bsmt.Unf.SF +
##     ames$BsmtFin.SF.1 + ames$Bedroom.AbvGr + ames$TotRms.AbvGrd,
##     data = ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -802532  -21831     360   20506  257685
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          19815.810   4006.647   4.946 8.01e-07 ***
## ames$Gr.Liv.Area       103.078      3.124  32.994  < 2e-16 ***
## ames$Bsmt.Unf.SF        44.754      2.439  18.348  < 2e-16 ***
## ames$BsmtFin.SF.1       67.766      2.425  27.947  < 2e-16 ***
## ames$Bedroom.AbvGr  -20880.711   1445.337 -14.447  < 2e-16 ***
## ames$TotRms.AbvGrd   1705.290   1083.565   1.574    0.116
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46700 on 2923 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.6588, Adjusted R-squared:  0.6583
## F-statistic:  1129 on 5 and 2923 DF,  p-value: < 2.2e-16
```

```
#converting data

num_cols <- unlist(lapply(ames, is.numeric))

numeric_data <- ames[ , num_cols]

#replaced empty spaces with NA Values
Replaced_values<- numeric_data %>%
  mutate_if(is.integer,function(n) ifelse(is.na(n),mean(n,na.rm=TRUE),n))

#Creating a co-relation plot for numerical data.
correlation_matrix<-cor(Replaced_values)
corrplot(correlation_matrix,tl.cex=0.6, fig.width=14)
```

```
## Warning in text.default(pos.xlabel[, 1], pos.xlabel[, 2], newcolnames, srt =
## tl.srt, : "fig.width" is not a graphical parameter
```

```
## Warning in text.default(pos.ylabel[, 1], pos.ylabel[, 2], newrownames, col =
## tl.col, : "fig.width" is not a graphical parameter
```

```
## Warning in title(title, ...): "fig.width" is not a graphical parameter
```



For the co-relation matrix we used, corplot(). The association between two variables is referred to as correlation. The degree of linear correlation between any two random variables is referred to as this. This relationship can be represented as a set of values in the interval [-1, 1]. The value -1 denotes a non-linear (negative) relationship, 1 denotes a positive linear relationship, and 0 denotes a relationship that is neither positive nor negative. However, a value of 0 does not imply that the variables are fully independent of one another. Correlation Matrices calculate the degree of linear association between a collection of random variables, one pair at a time, for each set of pairs in the data.

Since the skewness is negative(-0.01478586),this indicates that distribution is left-skewed and Kurtosis is positive (4.509277) , this indicates that distribution is right-skewed.

It's shows's that there are a ouliers present in living area and garage area.

```
lm11 <- lm(SalePrice ~ ames$Neighborhood + ames$Exterior.1st, data = ames)
summary(lm11)
```

```
##
## Call:
## lm(formula = SalePrice ~ ames$Neighborhood + ames$Exterior.1st,
##     data = ames)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -183491  -26038   -4269   20588  436259
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                159533.7    12629.3  12.632  < 2e-16 ***
## ames$NeighborhoodBlueste   -35709.3    18958.0  -1.884 0.059719 .
## ames$NeighborhoodBrDale    -77671.0    13742.7  -5.652 1.74e-08 ***
## ames$NeighborhoodBrkSide   -59045.2    11074.0  -5.332 1.05e-07 ***
## ames$NeighborhoodClearCr    20474.6    12552.0   1.631 0.102960
## ames$NeighborhoodCollgCr     4948.4    10126.4   0.489 0.625122
## ames$NeighborhoodCrawfor    11661.4    11114.3   1.049 0.294161
## ames$NeighborhoodEdwards   -56786.2    10478.4  -5.419 6.47e-08 ***
## ames$NeighborhoodGilbert    -2815.1    10431.5  -0.270 0.787283
## ames$NeighborhoodGreens      7156.7    20709.3   0.346 0.729683
## ames$NeighborhoodGrnHill    82311.4    37459.5   2.197 0.028075 *
## ames$NeighborhoodIDOTRR    -80770.4    11235.6  -7.189 8.29e-13 ***
## ames$NeighborhoodLandmrk   -60716.5    51806.3  -1.172 0.241297
## ames$NeighborhoodMeadowV -153451.1    14030.1 -10.937  < 2e-16 ***
## ames$NeighborhoodMitchel   -26503.9    10915.3  -2.428 0.015237 *
## ames$NeighborhoodNAmes     -41374.7    10165.3  -4.070 4.82e-05 ***
## ames$NeighborhoodNoRidge   140170.6    11434.9  12.258  < 2e-16 ***
## ames$NeighborhoodNPkVill   -49056.7    14975.9  -3.276 0.001066 **
## ames$NeighborhoodNridgHt   122355.5    10432.6  11.728  < 2e-16 ***
## ames$NeighborhoodNWAmes      2015.1    10853.8   0.186 0.852723
## ames$NeighborhoodOldTown   -61199.5    10435.0  -5.865 5.01e-09 ***
## ames$NeighborhoodSawyer    -49043.5    10719.8  -4.575 4.96e-06 ***
## ames$NeighborhoodSawyerW    -6113.3    10746.0  -0.569 0.569474
## ames$NeighborhoodSomerst    30591.9    10392.7   2.944 0.003270 **
## ames$NeighborhoodStoneBr   116283.6    12169.7   9.555  < 2e-16 ***
## ames$NeighborhoodSWISU     -52692.7    12312.6  -4.280 1.93e-05 ***
## ames$NeighborhoodTimber     50332.4    11386.4   4.420 1.02e-05 ***
## ames$NeighborhoodVeenker    54573.5    14411.0   3.787 0.000156 ***
## ames$Exterior.1stAsphShn      432.1    36897.7   0.012 0.990656
## ames$Exterior.1stBrkComm    18583.9    22232.3   0.836 0.403283
## ames$Exterior.1stBrkFace    49469.0     9648.5   5.127 3.14e-07 ***
## ames$Exterior.1stCBlock    -22432.0    36915.6  -0.608 0.543464
## ames$Exterior.1stCemntBd    89673.9     9789.4   9.160  < 2e-16 ***
## ames$Exterior.1stHdBoard    19036.5     8404.2   2.265 0.023579 *
## ames$Exterior.1stImStucc   -37704.2    51892.5  -0.727 0.467540
## ames$Exterior.1stMetalSd    20078.1     8136.0   2.468 0.013652 *
## ames$Exterior.1stPlywood    30233.9     8841.3   3.420 0.000636 ***
## ames$Exterior.1stPreCast   157252.6    51600.5   3.048 0.002328 **
## ames$Exterior.1stStone      67969.4    37061.0   1.834 0.066759 .
## ames$Exterior.1stStucco     33212.2    10965.8   3.029 0.002478 **
## ames$Exterior.1stVinylSd    38182.8     8220.2   4.645 3.55e-06 ***
```

```
## ames$Exterior.1stWd Sdng   26840.8     8158.4   3.290 0.001014 **
## ames$Exterior.1stWdShing   23415.9    10375.6   2.257 0.024093 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50900 on 2887 degrees of freedom
## Multiple R-squared:  0.5998, Adjusted R-squared:  0.594
## F-statistic:   103 on 42 and 2887 DF,  p-value: < 2.2e-16
```

In if the pvalue is between 0 and 0.001 then it will have 3 stars,if it is between 0.001 and 0.01 it will have 2 stars,if it is between 0.01 and 0.05 it will have 1 star,if it is between 0.05 and 0.1 it will have a dot and if it is between 0.1 and 1 it will not have anything (just a space). as p value is very close to zero.2.2e-16 as the p value would indicate a significant result

(Theses neighborhood having the outliers-BrkSide,CollgCr,Edwards,Gilbert,Mitchel,OldTown). Also different neighborhood having different prices.

```
NE <- table(ames$Neighborhood , ames$Exterior.1st)

aov2 <- aov(SalePrice ~ ames$Neighborhood + ames$Exterior.1st, data = ames)
summary(aov2)
```

```
##                      Df    Sum Sq   Mean Sq F value Pr(>F)
## ames$Neighborhood   27 1.072e+13 3.969e+11  153.18 <2e-16 ***
## ames$Exterior.1st   15 4.961e+11 3.307e+10   12.76 <2e-16 ***
## Residuals         2887 7.480e+12 2.591e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov2$coefficients
```

```
##              (Intercept) ames$NeighborhoodBlueste  ames$NeighborhoodBrDale
##              159533.6874              -35709.3320              -77671.0196
## ames$NeighborhoodBrkSide ames$NeighborhoodClearCr ames$NeighborhoodCollgCr
##              -59045.1535               20474.5766                4948.3510
## ames$NeighborhoodCrawfor ames$NeighborhoodEdwards ames$NeighborhoodGilbert
##               11661.3929              -56786.2450               -2815.0948
##  ames$NeighborhoodGreens ames$NeighborhoodGrnHill  ames$NeighborhoodIDOTRR
##                7156.7386               82311.3965              -80770.4270
## ames$NeighborhoodLandmrk ames$NeighborhoodMeadowV ames$NeighborhoodMitchel
##              -60716.4558             -153451.0995              -26503.9395
##   ames$NeighborhoodNAmes ames$NeighborhoodNoRidge ames$NeighborhoodNPkVill
##              -41374.6773              140170.5602              -49056.7277
## ames$NeighborhoodNridgHt  ames$NeighborhoodNWAmes ames$NeighborhoodOldTown
##              122355.4911                2015.1347              -61199.5238
##  ames$NeighborhoodSawyer ames$NeighborhoodSawyerW ames$NeighborhoodSomerst
##              -49043.4776               -6113.3221               30591.9381
## ames$NeighborhoodStoneBr   ames$NeighborhoodSWISU  ames$NeighborhoodTimber
##              116283.6346              -52692.6607               50332.3687
## ames$NeighborhoodVeenker ames$Exterior.1stAsphShn ames$Exterior.1stBrkComm
##               54573.4993                 432.1485               18583.9134
## ames$Exterior.1stBrkFace  ames$Exterior.1stCBlock ames$Exterior.1stCemntBd
##               49469.0081              -22431.9558               89673.8986
## ames$Exterior.1stHdBoard ames$Exterior.1stImStucc ames$Exterior.1stMetalSd
##               19036.5060              -37704.2477               20078.1325
## ames$Exterior.1stPlywood ames$Exterior.1stPreCast   ames$Exterior.1stStone
##               30233.9098              157252.5576               67969.4318
##  ames$Exterior.1stStucco ames$Exterior.1stVinylSd ames$Exterior.1stWd Sdng
##               33212.1695               38182.7684               26840.8239
## ames$Exterior.1stWdShing
##               23415.8870
```

```
# Creating contingency table from data using table()
conTable = table(NE)
print(conTable)
```

```
## NE
##   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  19  20
## 273  37  22  13  13   7   4   7   2   3   1   6   1   2   1   3   5   1   2   1
##  21  23  24  25  26  28  29  30  31  32  33  35  36  37  38  40  41  42  45  46
##   2   4   1   1   1   1   2   1   1   1   1   1   3   2   3   1   1   2   1   2
##  52  59  67  73  79  88  97 116 117 126 225
##   1   1   1   1   1   1   1   1   1   2   1
```

we are going to perform Analysis Of Variance (the two way ANOVA). ANOVA comes under the linear model so here we are using both lm and aov function to compare the results. The residuals and sum square having a magnitude of impact of matters as difference is wider i.e, 640,885,464.Degree of freedom means ,“how many variables has been added”.so difference is very big and on the bases of that we are rejecting NULL HYPOTHESIS.Both variables p values are smaller than 0.05 then it is conclude that there are significant difference among the mean of that two

independent variables.so, there's the significant difference these two variables F(25,17= 154 AS second F value is 13.) The model shows that the F statistic is very large and the p- value is very small, so it is essentially zero as this intercept to reject the Null Hypothesis and it's the evidence that there's a strong relation between Neighborhood and Exterior.1st.

```
# Two-way ANOVA with interaction effect
aov3 <- aov(SalePrice ~ Neighborhood * Exterior.1st, data = ames)
aov3 <- aov(SalePrice ~ Neighborhood + Exterior.1st + Neighborhood:Exterior.1st, data = ames)
summary(aov3)
```

```
##                              Df    Sum Sq   Mean Sq F value Pr(>F)
## Neighborhood                 27 1.072e+13 3.969e+11  170.40 <2e-16 ***
## Exterior.1st                 15 4.961e+11 3.307e+10   14.20 <2e-16 ***
## Neighborhood:Exterior.1st   132 1.064e+12 8.059e+09    3.46 <2e-16 ***
## Residuals                  2755 6.417e+12 2.329e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

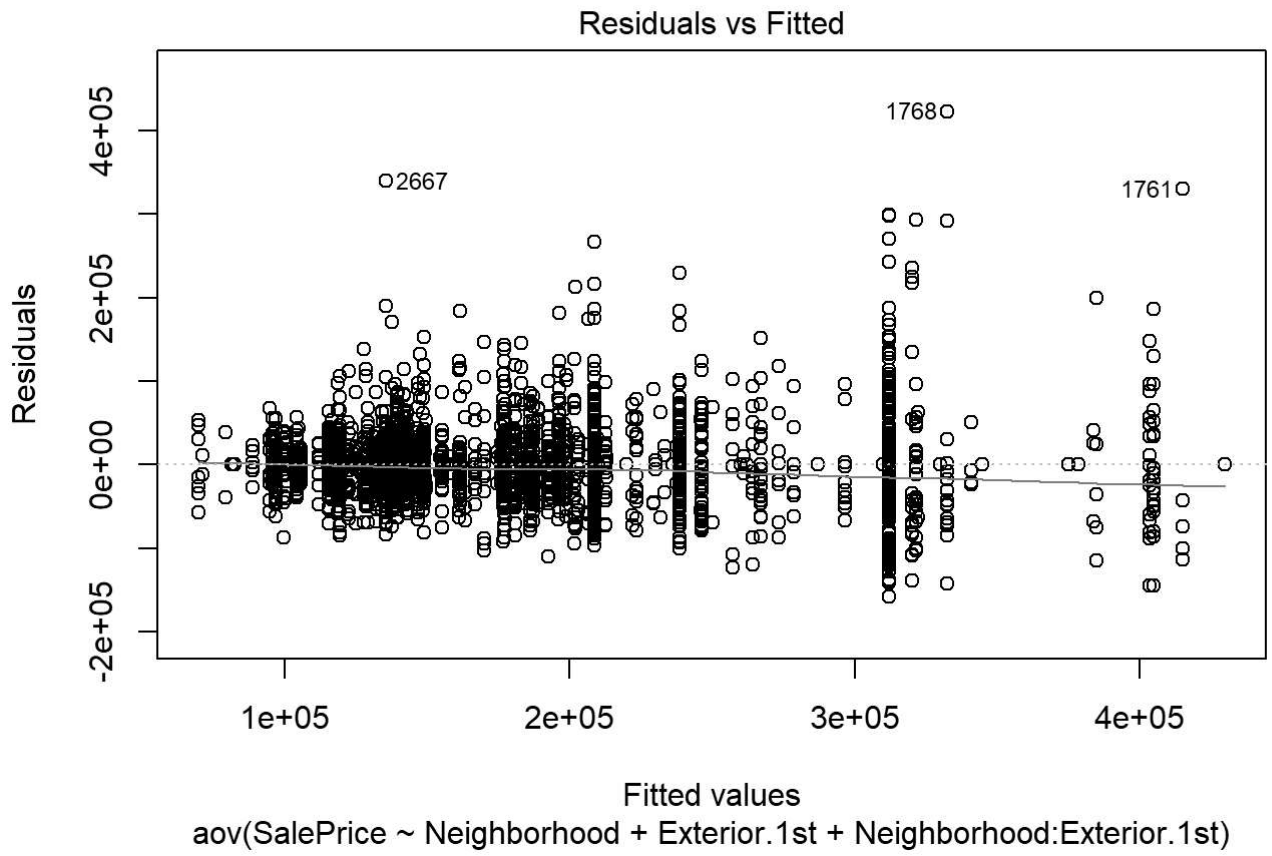
In this model, errors are greatly increased, our F-statistic larger, and our p-value much smaller so the model is not siginficant and 't match the default effects parameterization.Neighborhood having high sum of square and low p value it mean it's adding a lot of information. The p-value of the F-test is 170, and this probability is less than 0.05, your alpha, so you fail to reject the alternative hypothesis and can proceed as if the variances are equal between the groups.

```
group_by(ames, Neighborhood, Exterior.1st) %>%
  summarise(
    count = n(),
    mean = mean(SalePrice, na.rm = TRUE),
    sd = sd(SalePrice, na.rm = TRUE)
  )
```

```
## `summarise()` has grouped output by 'Neighborhood'. You can override using the `.groups` argu
ment.
```

```
## # A tibble: 175 x 5
## # Groups:   Neighborhood [28]
##    Neighborhood Exterior.1st count    mean     sd
##    <chr>        <chr>        <int>   <dbl>  <dbl>
##  1 Blmngtn      VinylSd         26 197936. 30000.
##  2 Blmngtn      WdShing          2 180094. 10739.
##  3 Blueste      HdBoard          3 182500  18875.
##  4 Blueste      MetalSd          7 126914. 12812.
##  5 BrDale       CemntBd          2 105500  19799.
##  6 BrDale       HdBoard         28 105616. 11996.
##  7 BrkSide      AsbShng          7 100500  28330.
##  8 BrkSide      BrkComm          1 202500     NA
##  9 BrkSide      BrkFace          1 210000     NA
## 10 BrkSide      CemntBd          1 159000     NA
## # ... with 165 more rows
```

```
plot(aov3, 1) #RESIDUALS FINDING
```



The residual V. fitted values plot for the final model shows a random scatter along the horizontal axis, suggesting that there is a linear association between the explanatory variables and price. We can also see that the points are equally scattered around the horizontal axis with no discernable pattern, indicating constant variance.

CONCLUSION:

It is shown from this dataset that the following factors are affecting the prices which are Finished and unfinished basement, Utilities, Building Type, Half bath, Full bath, Garage, ground living area, It is also shown that if we go for a full bath the price has been increased which is 21375$ per square feet and for the half bath, the price is 16485$. Utilities having only two variables only one value having NoSeWa and AllPub is having many values.

1. If we increase the area from one square foot the price increased to 0.038 % per square.
2. Properties are classified as Village Residential, Residential Low density & Residential medium density are more expensive than commercial properties.
3. The percentage prices prediction of these model is 95%.
4. The price and area are right skewed as it's ranges between 1000-2000 square per foot and price and area having the strong co-relation.so, as area increases, prices also increase.
5. Ames found to be distinctly numerous actual property market, market through sufficient variability in price, footage etc.
6. Also, there's home recission arise in US In year 2008 but as we can see there's no prices differences has been arrived, it means prices of houses has not been dropped that much.
7.Its shown that kitchen is the important factor if the kitchen was rated"Excellent" the price has been incrsesd by $4000 and also if we add fireplace then also prices has been incresead.