2017

Assignment 2

PREDICT 410 FALL 2017

ZEESHAN LATIFI

NORTHWESTERN UNIVERSITY

Introduction:

This is an exploratory data analysis of housing data for Ames, Iowa. In this analysis, we'll be using determining factors that can help predict the sales prices for a typical home in Ames, Iowa. The data has been provided by DeCock (2011). We will be looking for several predictor variables that will help determine our response variable: Sales Price.

To do this, we'll work through many aspects of data analysis. Initially, we'll evaluate our data, define the sample population, conduct a data quality check, perform analysis, and finally model our findings. From our model, we'll be able to assess whether our response variable can be predicted accurately using other variables provided in the data set.

Sample Population:

There are 82 variables and 2,930 observations in the Ames, Iowa data set. We begin by conducting a waterfall in R to clean our data set. By evaluation of each variable, we've identified what constitutes a 'typical' home in Ames.

We began with filtering on the 'SubClass' field. This field identifies the class of the home. The decision was to keep only homes that are:

- 1-STORY 1946 & NEWER ALL STYLES
- 2-STORY 1946 & NEWER
- SPLIT OR MULTI-LEVEL

We then removed all non-residential zoning, keeping only residential high, medium, and low density. Next removed all homes that were not on a paved street and did not have all public utilities included. A 'typical' home should have all standard utilities available.

To keep with our assumptions, the decision was made to only include homes that are in overall condition and quality of a 5 or higher. This means homes quality and condition ranked 'average' or higher. The same decision was made when filtering on the homes' exterior quality and condition. To eliminate homes that may skew our data set, only houses that were built in 1950 or later were included. Per our definition of the 'typical' home, we decided to only include homes that have square footage of 600 $ft.^2$ or higher for the first level. With that in mind we also eliminated homes without a paved driveway or central air. We've also eliminated townhomes and homes with lot areas above 200,000 ft^2. With these transformations, the observations were reduced down to 1,517. Figure 1 displays the count for each of the reductions.

Figure 1:

```
                                 [,1]
01: Not SFR                      1158
02: Non-Residential Zoning         82
03: Street Not Paved                2
04: Not All Utilities Included      2
05: Overall Quality Under 5        90
06: Overall Condition Under 5      32
07: Homes Built Pre-1950           17
08: Below Good Exterior Quality     2
09: Below Good Exterior Condition   5
10: First Floor Under 600 SqFt      1
11: No Central Air                  4
12: No Paved Driveway              16
13: Not a Single Family Home        1
14: Not a Normal Lot Area           1
99: Eligible Sample              1517
```

Exploratory Analysis:

To help predict the sales price for a home we should understand the response variable first. Figure 5 represents a summary for the 'SalesPrice' data in our sample population.

Figure 2:

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 62380  150500  185000  209400  242000  755000
```

As shown above, the mean pricing for a home in Ames according to the selected criteria is $209,400. We found no values that seemed out of the ordinary here. There is potential that the max value and min value can be outliers, but it's unable to be determined at this time.

Predictor Variables Analysis:

The two best predictor variables we've decided to conduct further analysis on is the "Year Built" variable and "First Floor Square Footage". Generally, 'year' would be considered a discrete variable. But since we are looking at a larger data set, in this instance we will use 'Year Built' as a continuous variable.

Figure 3:

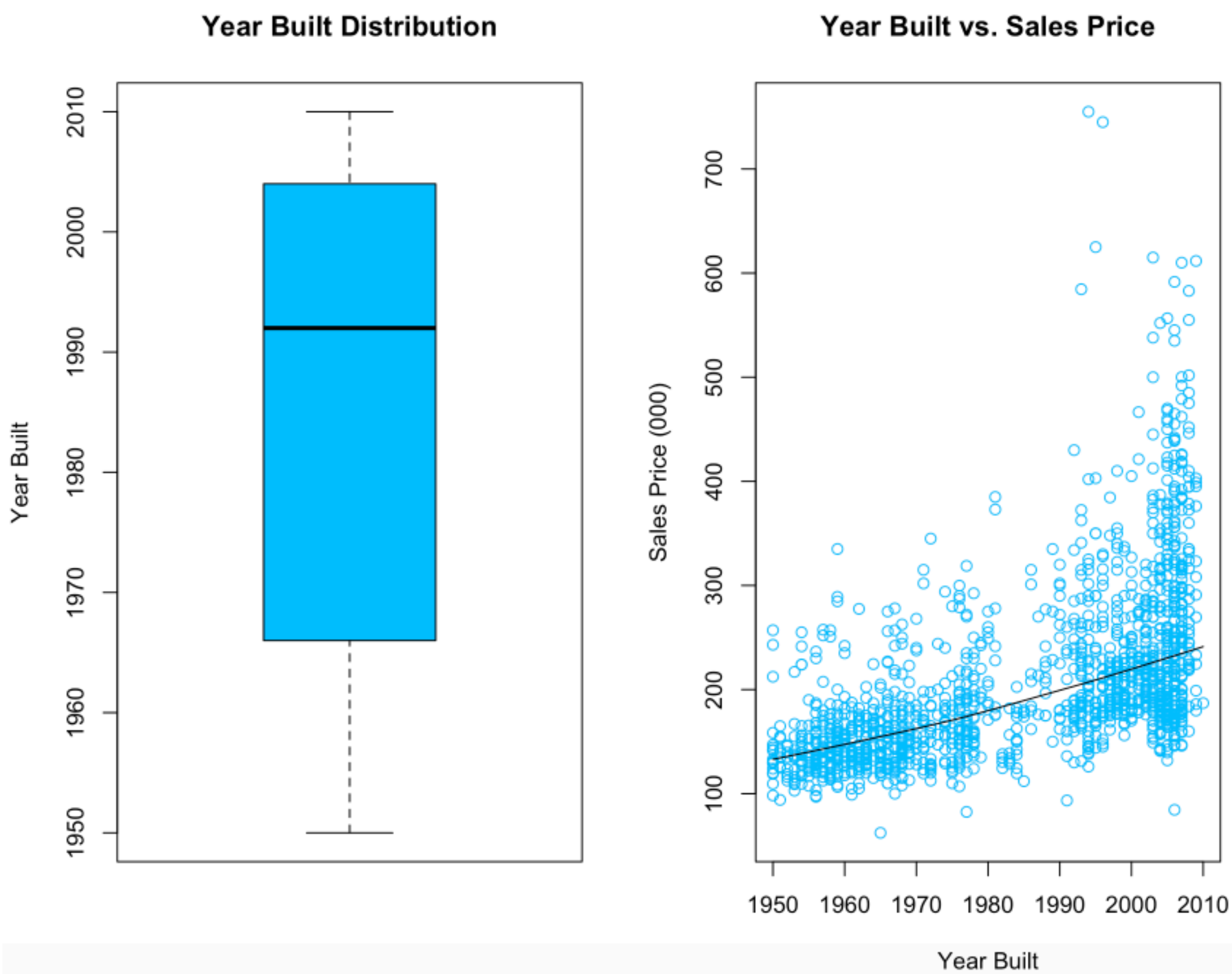
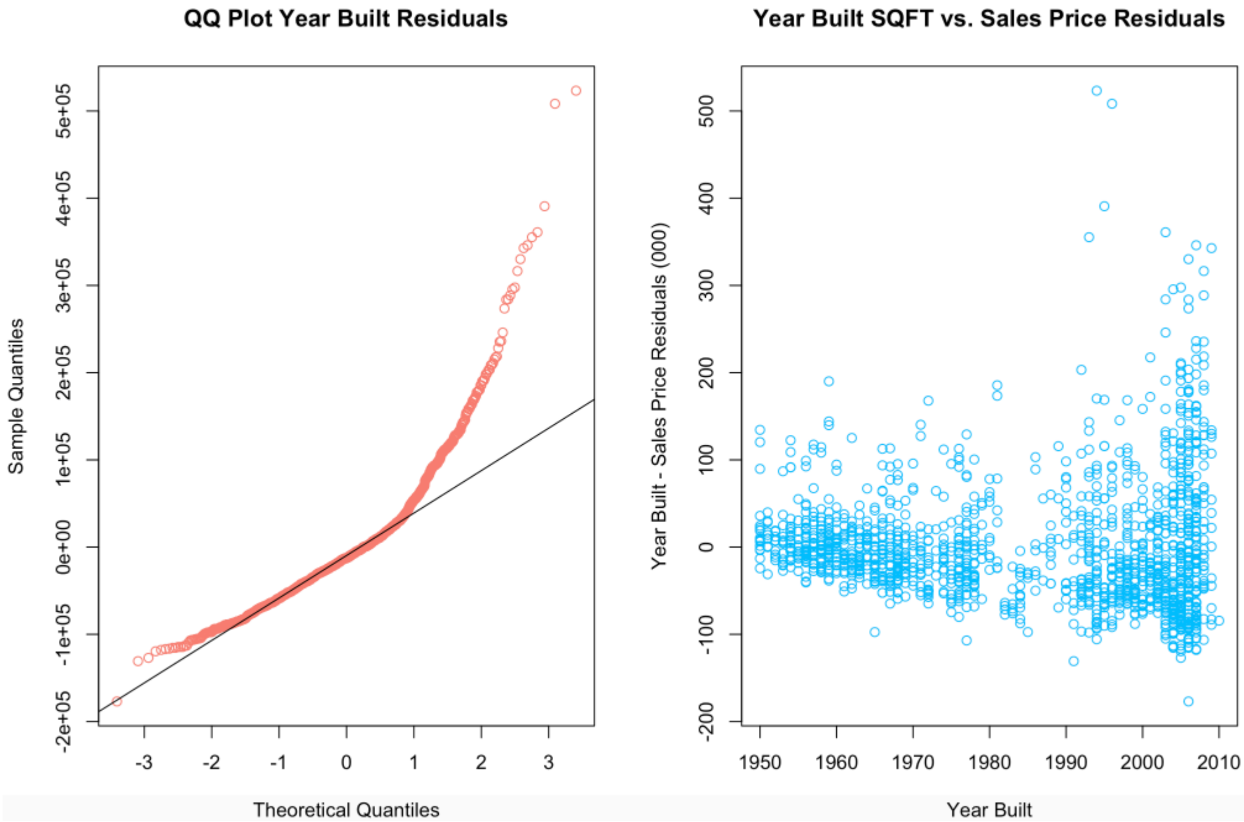


Figure 3 shows us the distribution of the homes built in the Ames, Iowa dataset after we've eliminated several factors. It also shows us the correlation with the sales price along with its LOESS line.

When evaluating the residuals, for the year built, we've included the residual plot, as well as the QQ plot of residuals. As you can see in Figure 4 below, the residual plot shows 'Year Built' not being an accurate standalone predictor of 'Sales Price'.

Figure 4:

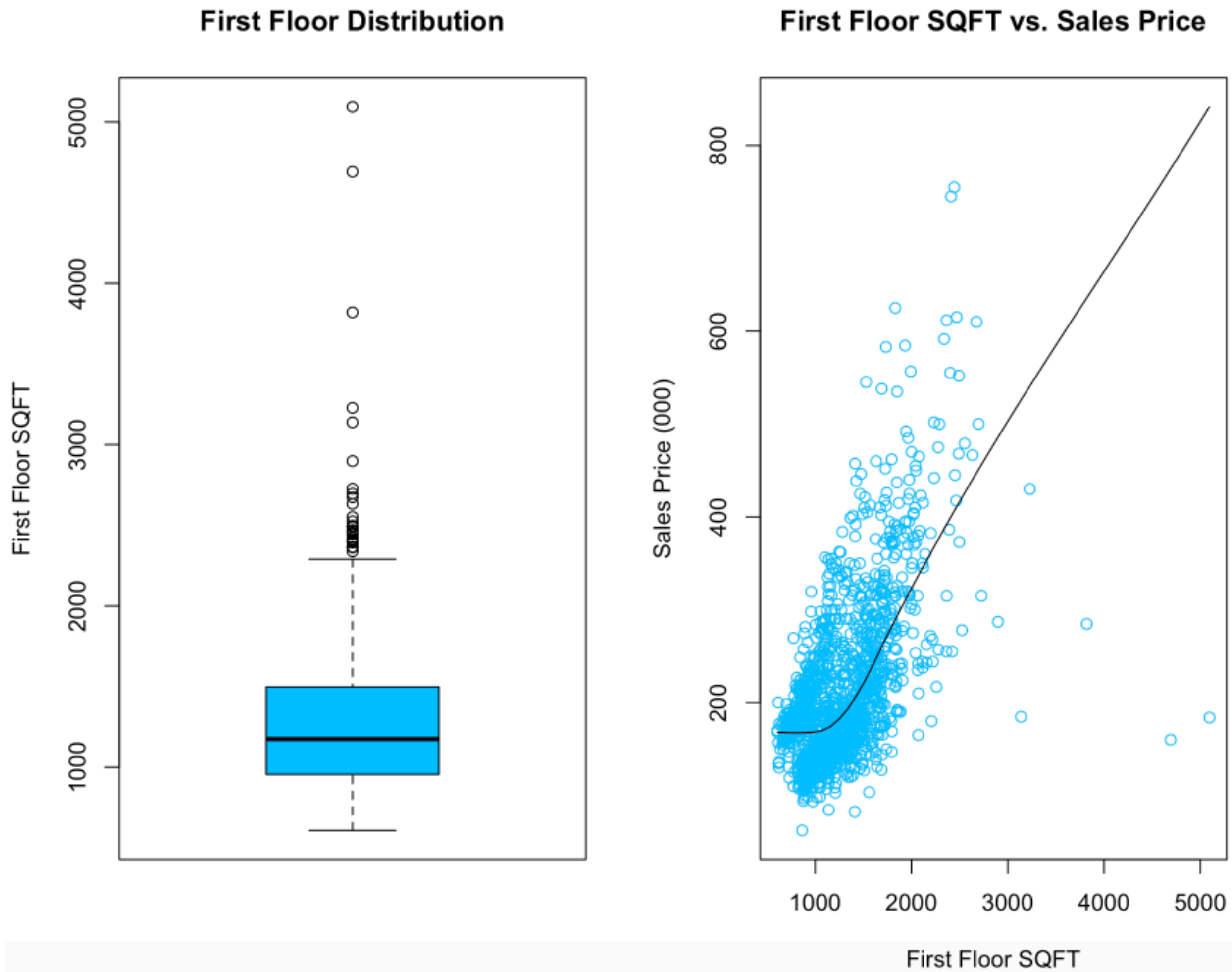



```
Residuals:
    Min      1Q  Median      3Q     Max
-176964  -42505  -12419   23164  523287

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.712e+06  1.889e+05  -24.95   <2e-16 ***
YearBuilt    2.479e+03  9.514e+01   26.06   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 70800 on 1515 degrees of freedom
Multiple R-squared:  0.3095,    Adjusted R-squared:  0.3091
F-statistic: 679.1 on 1 and 1515 DF,  p-value: < 2.2e-16
```

Our second predictor variable will be the first floor square footage of a home in Ames.

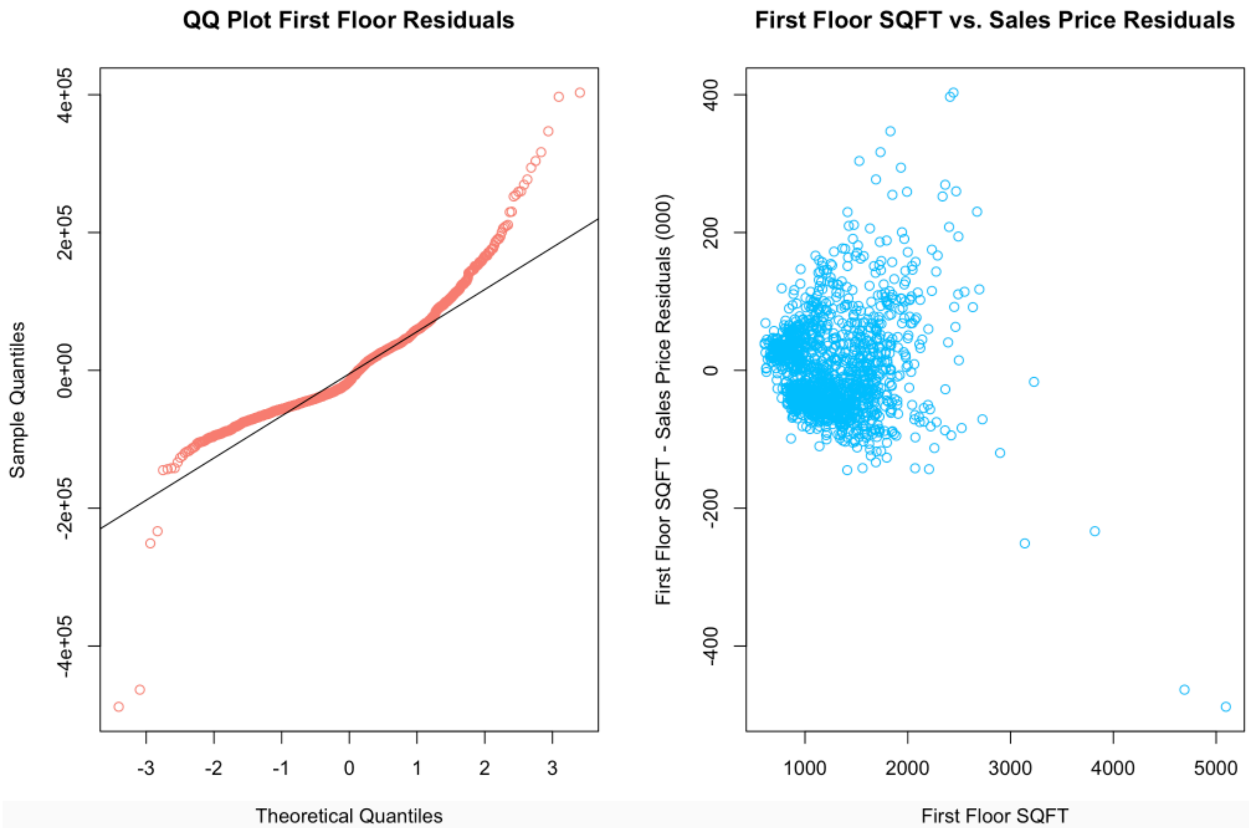
Figure 5:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
608	955	1175	1263	1498	5095



As shown in the figure above, the square footage for the first floor in Ames is about 1263 $ft.^2$. You can also view the distribution of the homes in Ames in terms of the first floor square footage. Some potential outliers exist, based on the square footage, we've decided the prices are representative of the population. We also evaluate the residuals for this variable as presented in Figure 6 below. As you can see the R^2 is low, therefore this may not be a good predictor of sales price as we first assumed. The outliers in this variable for our sample size are having a significant effect on the outcome.

Figure 6:

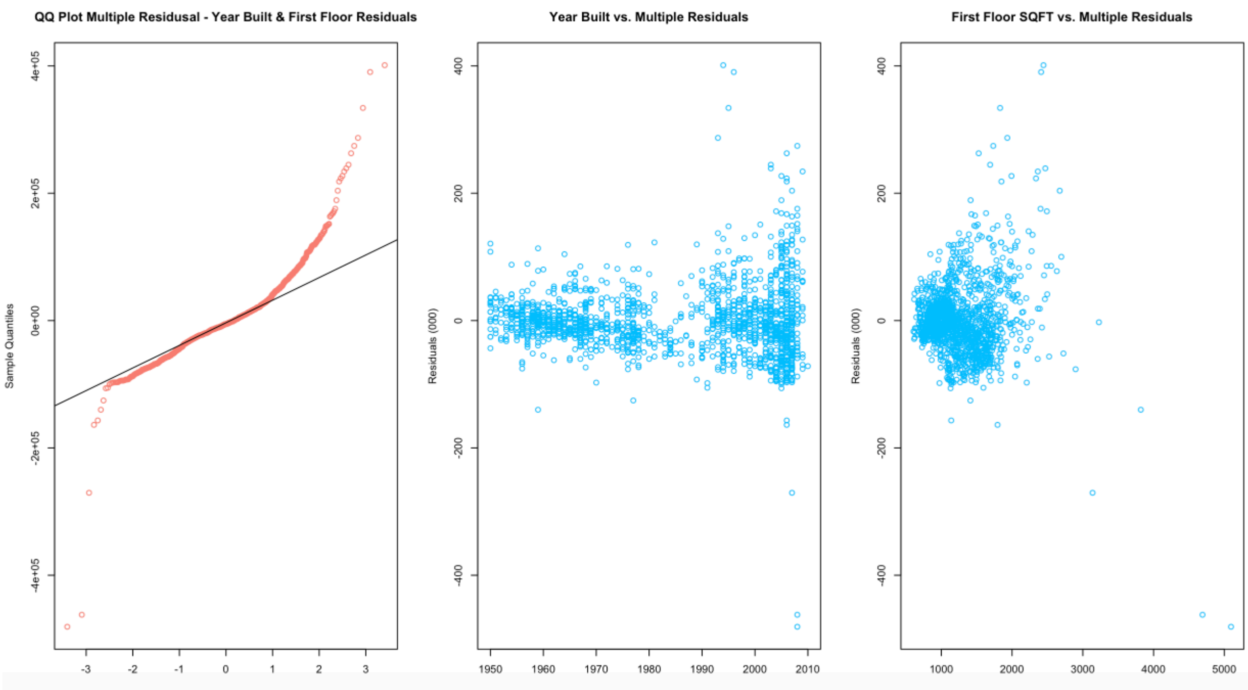



```
Residuals:
    Min      1Q  Median      3Q     Max
-488209  -46220  -16142   36253  402969

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 56993.024   5730.282   9.946   <2e-16 ***
FirstFlrSF    120.719      4.315  27.976   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 69190 on 1515 degrees of freedom
Multiple R-squared:  0.3406,    Adjusted R-squared:  0.3402
F-statistic: 782.6 on 1 and 1515 DF,  p-value: < 2.2e-16
```

Multiple Linear Regression:

Figure 7, below shows our analysis of multiple regression with both the year built and the square footage of the first floor of a home in Ames. There is a middle tier of homes that can be predicted but the outliers in our sample are too broad in order to use these as accurate predictor variables. Using multiple variables does not necessarily mean a better model. After utilizing both variables 'Year Built' and "First Floor SQFT' we've concluded that these may not be the most accurate predictor of 'Sales Price'.

Figure 7:



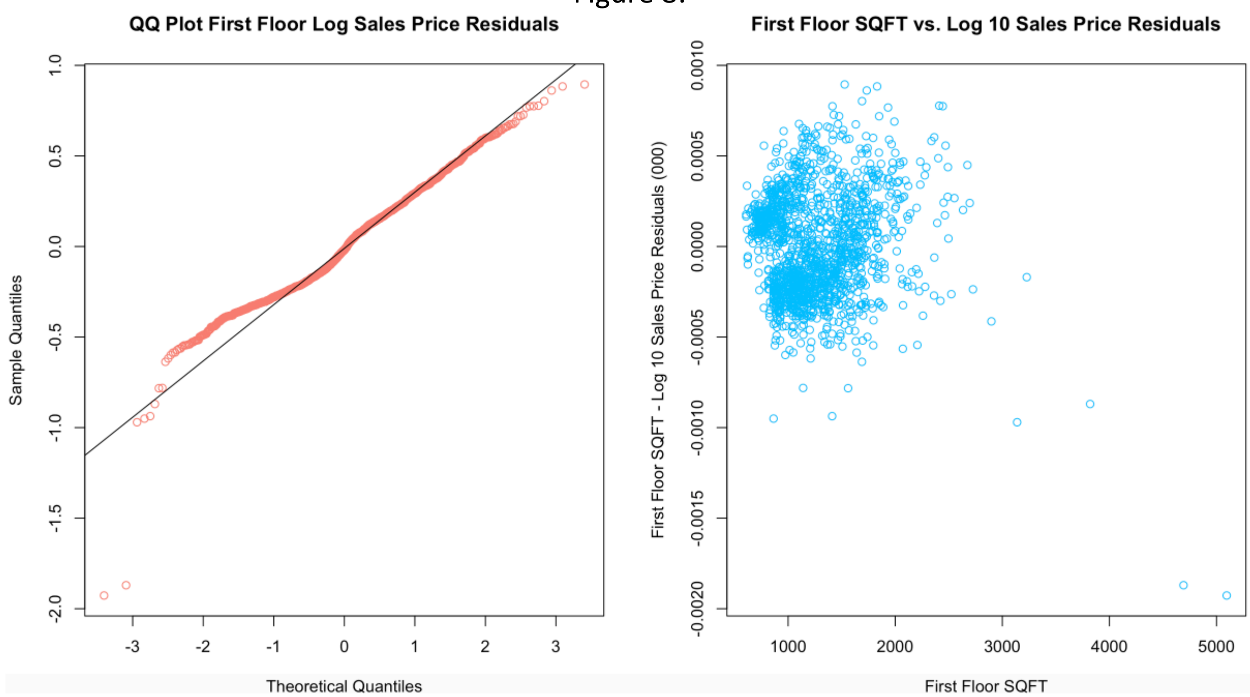

```
Residuals:
    Min      1Q  Median      3Q     Max
-480884  -27371   -5439   20514  401143

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.173e+06  1.508e+05  -27.67   <2e-16 ***
YearBuilt    2.140e+03  7.628e+01   28.06   <2e-16 ***
FirstFlrSF   1.060e+02  3.541e+00   29.93   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56140 on 1514 degrees of freedom
Multiple R-squared:  0.5662,    Adjusted R-squared:  0.5656
F-statistic:   988 on 2 and 1514 DF,  p-value: < 2.2e-16
```

Transformed Variable Analysis:

Based on our previous assessment, we've now decided to take a looked at a transformed regression of 'Sales Price'. We'll be doing this for 'Year Built', 'First Floor SQFT', and both combined in a multivariate analysis.

Right off the bat, for the first floor square footage, you can see the model fits much better than before taking the transformed data of sales price as shown in Figure 8.

Figure 8:




```
Residuals:
    Min      1Q  Median      3Q     Max
-1.9271 -0.2203 -0.0193  0.1988  0.8946

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.157e+01  2.396e-02  482.96   <2e-16 ***
FirstFlrSF  4.863e-04  1.804e-05   26.95   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2893 on 1515 degrees of freedom
Multiple R-squared:  0.3241,    Adjusted R-squared:  0.3237
F-statistic: 726.5 on 1 and 1515 DF,  p-value: < 2.2e-16
```

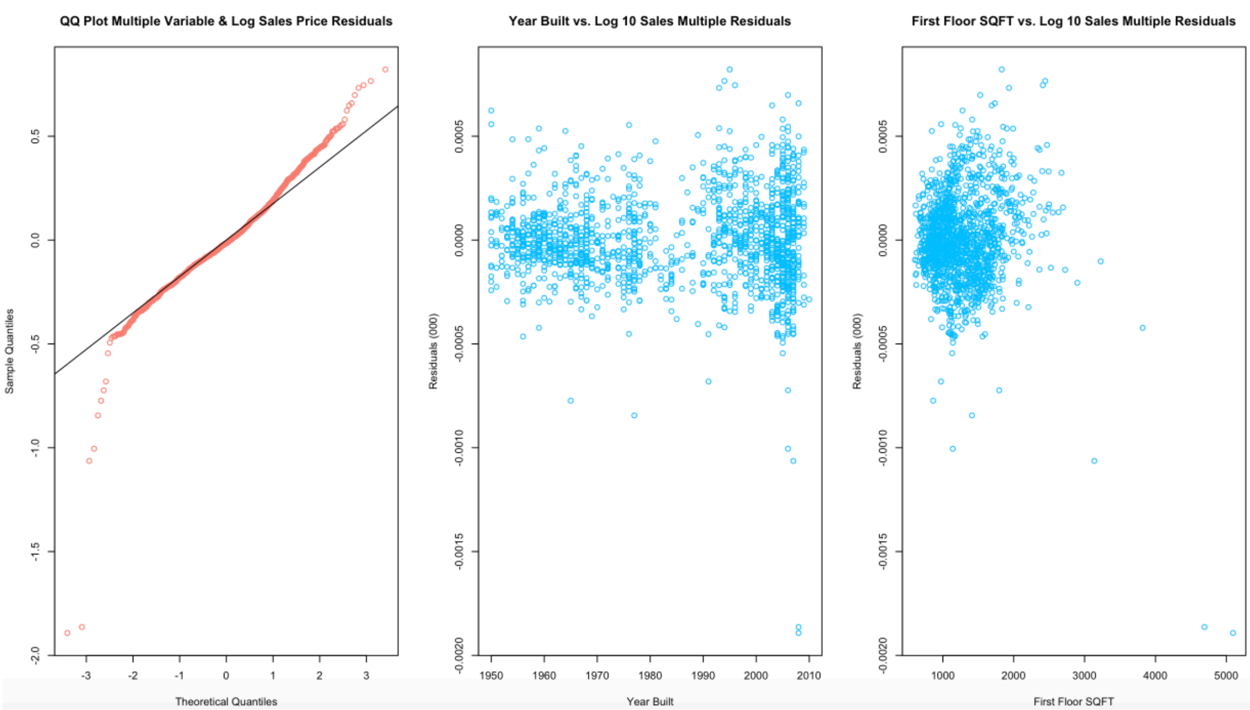
Now we'll take a look at the same, but for 'Year Built. The same can also be said for 'Year Built' as above. The transformed model fits much better along the line in comparison to the original.

Figure 9:


```
Residuals:
     Min       1Q   Median       3Q      Max
-1.08380 -0.18527 -0.03942  0.14831  1.24506

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.079e+01  7.297e-01  -14.79   <2e-16 ***
YearBuilt    1.157e-02  3.676e-04   31.49   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2736 on 1515 degrees of freedom
Multiple R-squared:  0.3956,    Adjusted R-squared:  0.3952
F-statistic: 991.5 on 1 and 1515 DF,  p-value: < 2.2e-16
```

Finally, we'll take a look at the multivariate model. In contrast to the untransformed model, this model fits much better. The goodness of fit in comparison to the other models is much higher. The R^2 value is approximately 0.63 when compared to the R2 value for the other models at 0.39 and 0.32 respectively

Figure 10:




```
Residuals:
     Min       1Q   Median       3Q      Max
-1.89209 -0.11830 -0.01497  0.11839  0.82152

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.675e+00  5.774e-01  -15.02   <2e-16 ***
YearBuilt    1.024e-02  2.920e-04   35.08   <2e-16 ***
FirstFlrSF   4.157e-04  1.356e-05   30.67   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2149 on 1514 degrees of freedom
Multiple R-squared:  0.6272,    Adjusted R-squared:  0.6267
F-statistic:  1273 on 2 and 1514 DF,  p-value: < 2.2e-16
```

Conclusion:

Based on our current analysis, the best model for predicting 'Sales Price' is the final model in which we analyze multiple variables against the log10(Sales Price). There is somewhat of a correlation between the variables and the Sales Price. R^2 is listed at footage ~0.63. When compared to our univariate analysis of year built and first floor square footage, this is the best model compiled. From our assessment above, we can infer the model with log10(Sales Price) and multivariate analysis may be the best possible predictor of the Sales Price per our sample population.

Code Appendix:

```
# Zeeshan Latifi
# 9.30.2017
# ames_waterfall.R

# Read in csv file for Ames housing data;

# Note that back slash is an escape character in R so we use \\ when we want \;
path.name <- '/Users/Zeeshan/Desktop/PREDICT 410/Week 1/';
file.name <- paste(path.name,'ames_housing_data.csv',sep='');

# Read in the csv file into an R data frame;
amesiowa.df <- read.csv(file.name,header=TRUE,stringsAsFactors=FALSE);

# Single ifelse() statement
# ifelse(condition, value if condition is TRUE, value if the condition is FALSE)

# Nested ifelse() statement
# ifelse(condition1, value if condition1 is TRUE,
#             ifelse(condition2, value if condition2 is TRUE,
#             value if neither condition1 nor condition2 is TRUE
#             )
# )


# Create a waterfall of drop conditions;
# Work the data frame as a 'table' like you would in SAS or SQL;
amesiowa.df$dropConditions <- ifelse(amesiowa.df$SubClass!= 020 & amesiowa.df$SubClass != 060 & amesiowa.df$SubClass
!= 080,'01: Not SFR',
  ifelse(amesiowa.df$Zoning!='RH' & amesiowa.df$Zoning!='RL' & amesiowa.df$Zoning!='RM','02: Non-Residential Zoning',
  ifelse(amesiowa.df$Street!='Pave','03: Street Not Paved',
  ifelse(amesiowa.df$Utilities!='AllPub', '04: Not All Utilities Included',
  ifelse(amesiowa.df$OverallQual<5, '05: Overall Quality Under 5',
  ifelse(amesiowa.df$OverallCond<5, '06: Overall Condition Under 5',
  ifelse(amesiowa.df$YearBuilt<1950, '07: Homes Built Pre-1950',
  ifelse(amesiowa.df$ExterQual!='TA' & amesiowa.df$ExterQual!='Gd'& amesiowa.df$ExterQual!='Ex', '08: Below Good Exterior
Quality',
  ifelse(amesiowa.df$ExterCond!='TA' & amesiowa.df$ExterCond!='Gd'& amesiowa.df$ExterCond!='Ex', '09: Below Good
Exterior Condition',
  ifelse(amesiowa.df$FirstFlrSF<600, '10: First Floor Under 600 SqFt',
  ifelse(amesiowa.df$CentralAir!='Y', '11: No Central Air',
  ifelse(amesiowa.df$PavedDrive!='Y', '12: No Paved Driveway',
  ifelse(amesiowa.df$BldgType!='1Fam', '13: Not a Single Family Home',
  ifelse(amesiowa.df$LotArea>200000, '14: Not a Normal Lot Area',
  '99: Eligible Sample')
  )))))))))))));


table(amesiowa.df$dropConditions)

# Save the table
waterfalls <- table(amesiowa.df$dropConditions);
```

```
# Format the table as a column matrix for presentation;
as.matrix(waterfalls,13,1)


# Eliminate all observations that are not part of the eligible sample population;
myeligible.population <- subset(amesiowa.df,dropConditions=='99: Eligible Sample');

# Check that all remaining observations are eligible;
table(myeligible.population$dropConditions);

head(myeligible.population)

###############################################################################################
#Final Table

final.pop <- data.frame(myeligible.population$SubClass, myeligible.population$Zoning, myeligible.population$LotArea,
            myeligible.population$Street, myeligible.population$Utilities, myeligible.population$BldgType,
            myeligible.population$HouseStyle, myeligible.population$OverallQual, myeligible.population$OverallCond,
            myeligible.population$YearBuilt, myeligible.population$YearRemodel,myeligible.population$ExterQual,
            myeligible.population$ExterCond, myeligible.population$BsmtFinType1, myeligible.population$FirstFlrSF,
            myeligible.population$GarageCars, myeligible.population$PavedDrive, myeligible.population$PoolArea,
            myeligible.population$CentralAir, myeligible.population$SalePrice)

head(final.pop)

###############################################################################################
#Data Quality Check
as.data.frame(table(myeligible.population$SubClass))

as.data.frame(table(myeligible.population$Zoning))

summary(myeligible.population$LotArea)
myeligible.population[is.element(myeligible.population$LotArea, max(myeligible.population$LotArea)),]

as.data.frame(table(myeligible.population$Street))

as.data.frame(table(myeligible.population$Utilities))

as.data.frame(table(myeligible.population$BldgType))

as.data.frame(table(myeligible.population$HouseStyle))

as.data.frame(table(myeligible.population$OverallQual))

as.data.frame(table(myeligible.population$OverallCond))

as.data.frame(table(myeligible.population$YearBuilt))
summary(myeligible.population$YearBuilt)

as.data.frame(table(myeligible.population$YearRemodel))
summary(myeligible.population$YearRemodel)

as.data.frame(table(myeligible.population$ExterQual))

as.data.frame(table(myeligible.population$ExterCond))

as.data.frame(table(myeligible.population$BsmtFinType1))
```

```
as.data.frame(table(myeligible.population$OverallQual))

summary(myeligible.population$FirstFlrSF)
sd(myeligible.population$FirstFlrSF)

as.data.frame(table(myeligible.population$GarageCars))

as.data.frame(table(myeligible.population$PavedDrive))

summary(myeligible.population$PoolArea)
as.data.frame(table(myeligible.population$PoolArea))

as.data.frame(table(myeligible.population$CentralAir))

summary(final.pop$myeligible.population.SalePrice)
sd(final.pop$myeligible.population.SalePrice)


########################################################################################
#Exploratory Data Analysis
par(mfrow = c(1,1))

boxplot(myeligible.population$SalePrice)

qqplot(myeligible.population$YearBuilt, myeligible.population$SalePrice)
plot(myeligible.population$YearBuilt, myeligible.population$SalePrice)

######**************************************************************
par(mfrow = c(1,2))
boxplot(myeligible.population$YearBuilt, main = 'Year Built Distribution', col = 'deepskyblue', ylab = 'Year Built')
scatter.smooth(myeligible.population$YearBuilt, myeligible.population$SalePrice/1000, main = 'Year Built vs. Sales Price',
        col = 'deepskyblue', ylab = 'Sales Price (000)', xlab = 'Year Built')

par(mfrow = c(1,1))
boxplot(myeligible.population$FirstFlrSF)
qqplot(myeligible.population$FirstFlrSF, myeligible.population$SalePrice)
plot(myeligible.population$FirstFlrSF, myeligible.population$SalePrice)
summary(myeligible.population$FirstFlrSF)

######**************************************************************
par(mfrow = c(1,2))
boxplot(myeligible.population$FirstFlrSF, main = 'First Floor Distribution', col = 'deepskyblue', ylab = 'First Floor SQFT')
scatter.smooth(myeligible.population$FirstFlrSF, myeligible.population$SalePrice/1000, main = 'First Floor SQFT vs. Sales Price',
        col = 'deepskyblue', ylab = 'Sales Price (000)', xlab = 'First Floor SQFT')

par(mfrow = c(1,1))

plot(myeligible.population$OverallQual, myeligible.population$SalePrice)

plot(myeligible.population$OverallCond, myeligible.population$SalePrice)

plot(myeligible.population$LotArea, myeligible.population$SalePrice)
boxplot(myeligible.population$LotArea)


style_table <- table(myeligible.population$HouseStyle)
barplot(style_table)
```

```
bldg_table <- table(myeligible.population$BldgType)
barplot(bldg_table)

par(mfrow = c(1,2))
extqual_table <- table(myeligible.population$ExterQual)
barplot(extqual_table, col = 'deepskyblue', main = 'Exterior Quality')

extcond_table <- table(myeligible.population$ExterCond)
barplot(extcond_table,col = 'salmon', main = 'Exterior Condition')

par(mfrow = c(1,1))
plot(myeligible.population$YearRemodel, myeligible.population$SalePrice)
boxplot(myeligible.population$YearRemodel)


################################################################################################
#Regression Analysis on 3 variables
par(mfrow = c(1,2))
boxplot(myeligible.population$SalePrice, ylim = c(60000,760000), col = 'deepskyblue', main = 'Sales Price Box Plot', ylab = 'Sales Price')
boxplot(log10(myeligible.population$SalePrice), col = 'deepskyblue', main = 'Log10 Sales Price Box Plot', ylab = 'Sales Price')




qqplot(myeligible.population$YearBuilt, myeligible.population$SalePrice, ylim = c(60000,760000), col = 'deepskyblue',
      main = 'Year Built vs. Sales Price - QQ', ylab = 'Sales Price', xlab = 'Year Built')
plot(myeligible.population$YearBuilt, myeligible.population$SalePrice, ylim = c(60000,760000),
    col = 'deepskyblue', main = 'Year Built vs. Sales Price', ylab = 'Sales Price', xlab = 'Year Built')
cor(myeligible.population$YearBuilt, myeligible.population$SalePrice)




qqplot(myeligible.population$FirstFlrSF, myeligible.population$SalePrice, ylim = c(60000,760000), col = 'deepskyblue',
      main = 'First Floor SQFT vs. Sales Price - QQ', ylab = 'Sales Price', xlab = 'First Floor SQFT')
plot(myeligible.population$FirstFlrSF, myeligible.population$SalePrice, ylim = c(60000,760000), col = 'deepskyblue',
    main = 'First Floor SQFT vs. Sales Price', ylab = 'Sales Price', xlab = 'First Floor SQFT')
cor(myeligible.population$FirstFlrSF, myeligible.population$SalePrice)




qqplot(myeligible.population$LotArea, myeligible.population$SalePrice, ylim = c(60000,760000), col = 'deepskyblue',
      main = 'Lot Area vs. Sales Price - QQ', ylab = 'Sales Price', xlab = 'Lot Area')
plot(myeligible.population$LotArea, myeligible.population$SalePrice, ylim = c(60000,760000), col = 'deepskyblue',
    main = 'Lot Area vs. Sales Price', ylab = 'Sales Price', xlab = 'Lot Area')
cor(myeligible.population$LotArea, myeligible.population$SalePrice)

par(mfrow = c(1,1))
boxplot(myeligible.population$YearBuilt, col = 'deepskyblue', ylab = 'Year Built')
```

```
#########################################################################################################
##################
#Assignment 2
# First floor predictor variable
par(mfrow = c(1,1))

model.1 <- lm(SalePrice ~ FirstFlrSF, data=myeligible.population)

# Display model summary
summary(model.1)

# List out components of lm object
names(model.1)

model.1$coefficients

par(mfrow = c(1,2))

qqnorm(model.1$residuals, main = 'QQ Plot First Floor Residuals', col = 'salmon')
qqline(model.1$residuals)

# Make a scatterplot
plot(myeligible.population$FirstFlrSF,model.1$residuals/1000, main = 'First Floor SQFT vs. Sales Price Residuals',
    col = 'deepskyblue', xlab = 'First Floor SQFT', ylab = 'First Floor SQFT - Sales Price Residuals (000)')


###########################################################################################
# Year Built predictor variable

model.2 <- lm(SalePrice ~ YearBuilt, data=myeligible.population)

# Display model summary
summary(model.2)

# List out components of lm object
names(model.2)

model.2$coefficients

par(mfrow = c(1,2))
qqnorm(model.2$residuals, main = 'QQ Plot Year Built Residuals', col = 'salmon')
qqline(model.2$residuals)

# Make a scatterplot
plot(myeligible.population$YearBuilt,model.2$residuals/1000, main = 'Year Built SQFT vs. Sales Price Residuals',
    col = 'deepskyblue', xlab = 'Year Built', ylab = 'Year Built - Sales Price Residuals (000)')

###########################################################################################
# Multiple regression plot with both variables

model.3 <- lm(SalePrice ~ YearBuilt + FirstFlrSF, data=myeligible.population)

# Display model summary
```

```
summary(model.3)

# List out components of lm object
names(model.3)

model.3$coefficients

par(mfrow = c(1,3))
qqnorm(model.3$residuals, main = 'QQ Plot Multiple Residusal - Year Built & First Floor Residuals', col = 'salmon')
qqline(model.3$residuals)

# Make a scatterplot
#par(mfrow = c(1,2))
plot(myeligible.population$YearBuilt,model.3$residuals/1000, main = 'Year Built vs. Multiple Residuals',
    col = 'deepskyblue', xlab = 'Year Built', ylab = 'Residuals (000)')
plot(myeligible.population$FirstFlrSF,model.3$residuals/1000, main = 'First Floor SQFT vs. Multiple Residuals',
    col = 'deepskyblue', xlab = 'First Floor SQFT', ylab = 'Residuals (000)')




##############################################################################################
#reproduce models 1,2,3 with log(salesprice)

model.4 <- lm(log(SalePrice) ~ FirstFlrSF, data=myeligible.population)

# Display model summary
summary(model.4)
model.4$coefficients

par(mfrow = c(1,2))
qqnorm(model.4$residuals, main = 'QQ Plot First Floor Log Sales Price Residuals', col = 'salmon')
qqline(model.4$residuals)

# Make a scatterplot
plot(myeligible.population$FirstFlrSF,model.4$residuals/1000, main = 'First Floor SQFT vs. Log 10 Sales Price Residuals',
    col = 'deepskyblue', xlab = 'First Floor SQFT', ylab = 'First Floor SQFT - Log 10 Sales Price Residuals (000)')

###########################################
#Model 5
model.5 <- lm(log(SalePrice) ~ YearBuilt, data=myeligible.population)

# Display model summary
summary(model.5)
model.5$coefficients

par(mfrow = c(1,2))
qqnorm(model.5$residuals, main = 'QQ Plot Year Built Log Sales Price Residuals', col = 'salmon')
qqline(model.5$residuals)

# Make a scatterplot
plot(myeligible.population$YearBuilt,model.5$residuals/1000, main = 'Year Built SQFT vs. Log 10 Sales Price Residuals',
    col = 'deepskyblue', xlab = 'Year Built', ylab = 'Year Built - Log 10 Sales Price Residuals (000)')

###########################################
#Model 6
```

```
model.6 <- lm(log(SalePrice) ~ YearBuilt + FirstFlrSF, data=myeligible.population)

# Display model summary
summary(model.6)
model.6$coefficients

par(mfrow = c(1,3))
qqnorm(model.6$residuals, main = 'QQ Plot Multiple Variable & Log Sales Price Residuals', col = 'salmon')
qqline(model.6$residuals)

# Make a scatterplot
plot(myeligible.population$YearBuilt,model.6$residuals/1000, main = 'Year Built vs. Log 10 Sales Multiple Residuals',
    col = 'deepskyblue', xlab = 'Year Built', ylab = 'Residuals (000)')
plot(myeligible.population$FirstFlrSF,model.6$residuals/1000, main = 'First Floor SQFT vs. Log 10 Sales Multiple Residuals',
    col = 'deepskyblue', xlab = 'First Floor SQFT', ylab = 'Residuals (000)')
```