Multiple Linear Regression

Dhruv Aggarwal

12/13/2020

# MULTIPLE LINEAR REGRESSION  
housing = read.table("http://www.jaredlander.com/data/housing.csv", sep =",", header = TRUE, stringsAsFactors=FALSE)  
  
# write.table(housing, "Shousing.csv", col.names = TRUE,row.names = FALSE, quote = FALSE, sep =",")  
# ?write.table # Help on write table!  
  
View(housing)  
  
str(housing)

## 'data.frame': 2626 obs. of 13 variables:  
## $ Neighborhood : chr "FINANCIAL" "FINANCIAL" "FINANCIAL" "FINANCIAL" ...  
## $ Building.Classification: chr "R9-CONDOMINIUM" "R4-CONDOMINIUM" "RR-CONDOMINIUM" "R4-CONDOMINIUM" ...  
## $ Total.Units : int 42 78 500 282 239 133 109 107 247 121 ...  
## $ Year.Built : int 1920 1985 NA 1930 1985 1986 1985 1986 1987 1985 ...  
## $ Gross.SqFt : int 36500 126420 554174 249076 219495 139719 105000 87479 255845 106129 ...  
## $ Estimated.Gross.Income : int 1332615 6633257 17310000 11776313 10004582 5127687 4365900 3637377 11246946 4115683 ...  
## $ Gross.Income.per.SqFt : num 36.5 52.5 31.2 47.3 45.6 ...  
## $ Estimated.Expense : int 342005 1762295 3543000 2784670 2783197 1497788 1273650 1061120 2440761 1231096 ...  
## $ Expense.per.SqFt : num 9.37 13.94 6.39 11.18 12.68 ...  
## $ Net.Operating.Income : int 990610 4870962 13767000 8991643 7221385 3629899 3092250 2576257 8806185 2884587 ...  
## $ Full.Market.Value : int 7300000 30690000 90970000 67556006 54320996 26737996 22210281 19449002 66316999 21821999 ...  
## $ Market.Value.per.SqFt : num 200 243 164 271 247 ...  
## $ Boro : chr "Manhattan" "Manhattan" "Manhattan" "Manhattan" ...

housing$Boro= as.factor(housing$Boro)# Boro - Borough - Town with its own Government.  
  
str(housing)

## 'data.frame': 2626 obs. of 13 variables:  
## $ Neighborhood : chr "FINANCIAL" "FINANCIAL" "FINANCIAL" "FINANCIAL" ...  
## $ Building.Classification: chr "R9-CONDOMINIUM" "R4-CONDOMINIUM" "RR-CONDOMINIUM" "R4-CONDOMINIUM" ...  
## $ Total.Units : int 42 78 500 282 239 133 109 107 247 121 ...  
## $ Year.Built : int 1920 1985 NA 1930 1985 1986 1985 1986 1987 1985 ...  
## $ Gross.SqFt : int 36500 126420 554174 249076 219495 139719 105000 87479 255845 106129 ...  
## $ Estimated.Gross.Income : int 1332615 6633257 17310000 11776313 10004582 5127687 4365900 3637377 11246946 4115683 ...  
## $ Gross.Income.per.SqFt : num 36.5 52.5 31.2 47.3 45.6 ...  
## $ Estimated.Expense : int 342005 1762295 3543000 2784670 2783197 1497788 1273650 1061120 2440761 1231096 ...  
## $ Expense.per.SqFt : num 9.37 13.94 6.39 11.18 12.68 ...  
## $ Net.Operating.Income : int 990610 4870962 13767000 8991643 7221385 3629899 3092250 2576257 8806185 2884587 ...  
## $ Full.Market.Value : int 7300000 30690000 90970000 67556006 54320996 26737996 22210281 19449002 66316999 21821999 ...  
## $ Market.Value.per.SqFt : num 200 243 164 271 247 ...  
## $ Boro : Factor w/ 5 levels "Bronx","Brooklyn",..: 3 3 3 3 3 3 3 3 3 3 ...

unique(housing$Boro)

## [1] Manhattan Brooklyn Queens Bronx Staten Island  
## Levels: Bronx Brooklyn Manhattan Queens Staten Island

head(housing)

## Neighborhood Building.Classification Total.Units Year.Built Gross.SqFt  
## 1 FINANCIAL R9-CONDOMINIUM 42 1920 36500  
## 2 FINANCIAL R4-CONDOMINIUM 78 1985 126420  
## 3 FINANCIAL RR-CONDOMINIUM 500 NA 554174  
## 4 FINANCIAL R4-CONDOMINIUM 282 1930 249076  
## 5 TRIBECA R4-CONDOMINIUM 239 1985 219495  
## 6 TRIBECA R4-CONDOMINIUM 133 1986 139719  
## Estimated.Gross.Income Gross.Income.per.SqFt Estimated.Expense  
## 1 1332615 36.51 342005  
## 2 6633257 52.47 1762295  
## 3 17310000 31.24 3543000  
## 4 11776313 47.28 2784670  
## 5 10004582 45.58 2783197  
## 6 5127687 36.70 1497788  
## Expense.per.SqFt Net.Operating.Income Full.Market.Value Market.Value.per.SqFt  
## 1 9.37 990610 7300000 200.00  
## 2 13.94 4870962 30690000 242.76  
## 3 6.39 13767000 90970000 164.15  
## 4 11.18 8991643 67556006 271.23  
## 5 12.68 7221385 54320996 247.48  
## 6 10.72 3629899 26737996 191.37  
## Boro  
## 1 Manhattan  
## 2 Manhattan  
## 3 Manhattan  
## 4 Manhattan  
## 5 Manhattan  
## 6 Manhattan

names(housing)=c("Neighborhood","Class", "Units", "YearBuilt", "SqFt", "Income", "IncomePerSqFt", "Expense", "ExpensePerSqFt", "NetIncome", "Value", "ValuePerSqFt", "Boro")  
  
tail(housing)

## Neighborhood Class Units YearBuilt SqFt Income  
## 2621 NEW SPRINGVILLE R4-CONDOMINIUM 37 NA 47880 673193  
## 2622 ROSEBANK R4-CONDOMINIUM 52 NA 62391 831672  
## 2623 ARROCHAR-SHORE ACRES R4-CONDOMINIUM 102 1987 90618 1274089  
## 2624 GRANT CITY R4-CONDOMINIUM 100 1986 78903 1321625  
## 2625 GRANT CITY R4-CONDOMINIUM 159 1961 166712 2343971  
## 2626 GREAT KILLS R4-CONDOMINIUM 67 1965 108864 1298748  
## IncomePerSqFt Expense ExpensePerSqFt NetIncome Value ValuePerSqFt  
## 2621 14.06 336596 7.03 336597 2115260 44.18  
## 2622 13.33 326305 5.23 505367 3354003 53.76  
## 2623 14.06 637045 7.03 637044 5233000 57.75  
## 2624 16.75 673832 8.54 647793 4687000 59.40  
## 2625 14.06 1171985 7.03 1171986 5967531 35.80  
## 2626 11.93 722857 6.64 575891 3673011 33.74  
## Boro  
## 2621 Staten Island  
## 2622 Staten Island  
## 2623 Staten Island  
## 2624 Staten Island  
## 2625 Staten Island  
## 2626 Staten Island

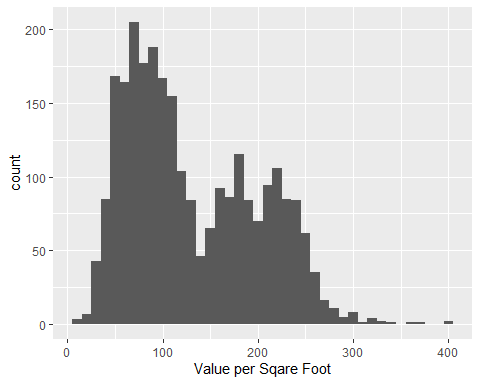
#Response Variable : ValuePerSqFt  
  
# Visulaize the Data  
require(ggplot2)

## Loading required package: ggplot2

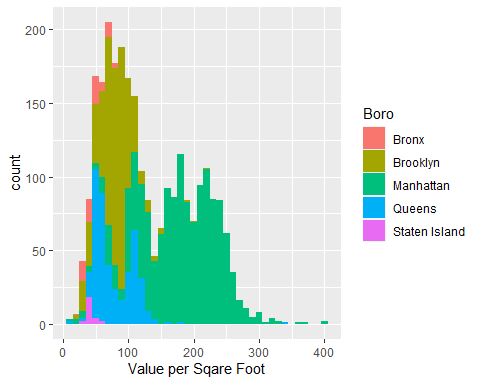
require(lattice)

## Loading required package: lattice

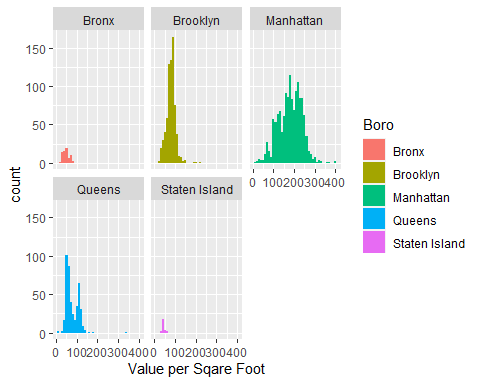
# Draw Histofgram of ValuePerSqFt  
ggplot(housing, aes(x=ValuePerSqFt))+geom\_histogram(binwidth=10)+labs(x="Value per Sqare Foot")



ggplot(housing, aes(x=ValuePerSqFt, fill=Boro))+geom\_histogram(binwidth=10)+labs(x="Value per Sqare Foot")

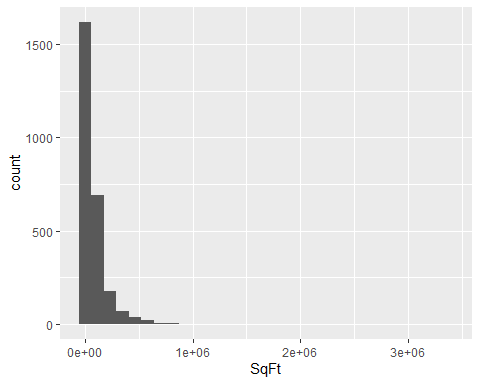


ggplot(housing, aes(x=ValuePerSqFt, fill=Boro))+geom\_histogram(binwidth=10)+labs(x="Value per Sqare Foot")+facet\_wrap("Boro")



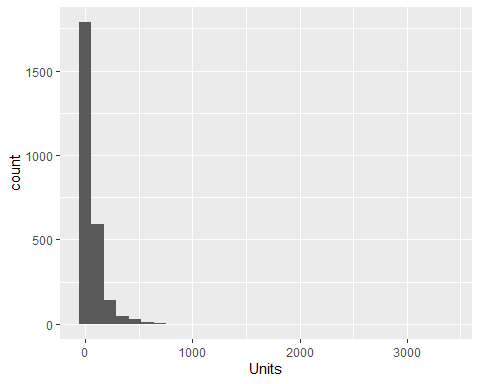
# Histogram for Square Footage and Units  
ggplot(housing, aes(x=SqFt))+geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



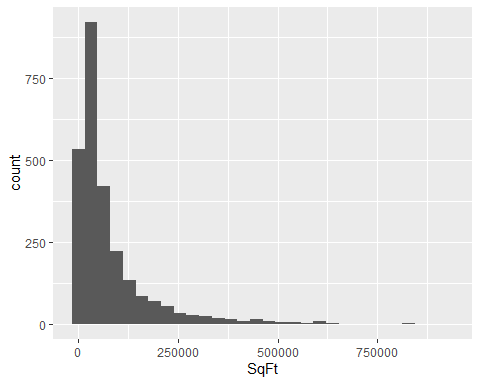
ggplot(housing, aes(x=Units))+geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ggplot(housing[housing$Units<1000,], aes(x=SqFt))+geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

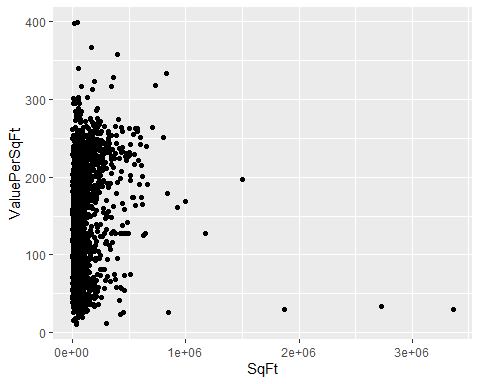


ggplot(housing[housing$Units<1000,], aes(x=Units))+geom\_histogram()

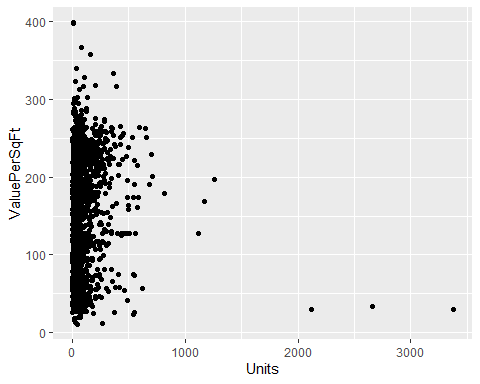
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



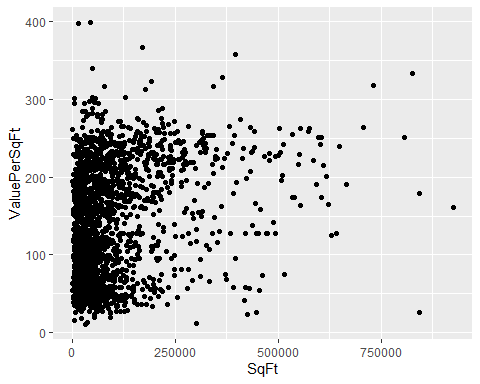
ggplot(housing, aes(x=SqFt, y=ValuePerSqFt))+geom\_point()



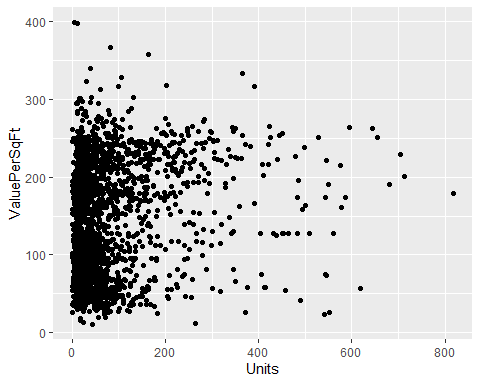
ggplot(housing, aes(x=Units, y=ValuePerSqFt))+geom\_point()



ggplot(housing[housing$Units<1000,], aes(x=SqFt, y=ValuePerSqFt))+geom\_point()



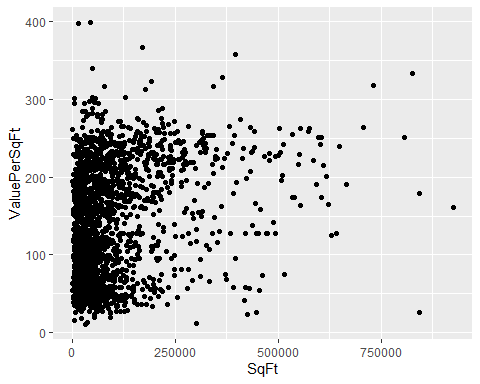
ggplot(housing[housing$Units<1000,], aes(x=Units, y=ValuePerSqFt))+geom\_point()



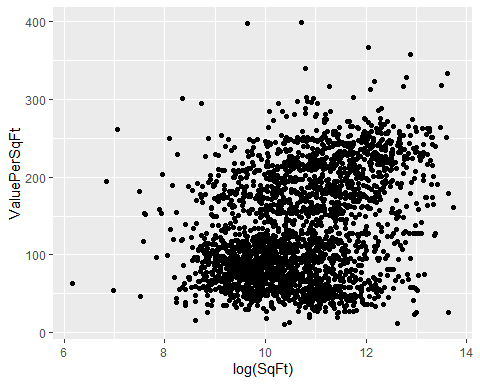
# How many Housing$Units greater than 1000  
sum(housing$Units>=1000)

## [1] 6

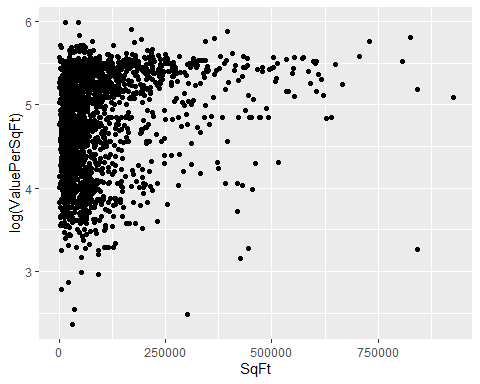
# Remove Housing$Units greater than 1000  
housing=housing[housing$Units<1000,]  
# Plot ValuePerSqFt against SqFt  
ggplot(housing, aes(x=SqFt, y=ValuePerSqFt))+geom\_point()



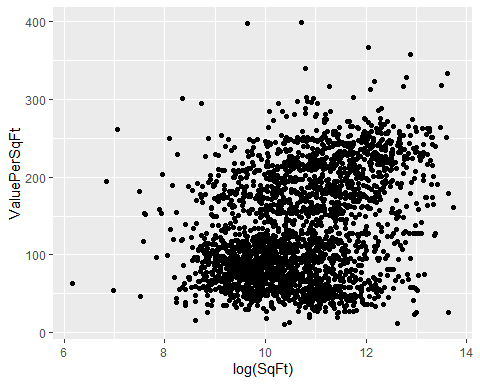
ggplot(housing, aes(x=log(SqFt), y=ValuePerSqFt))+geom\_point()



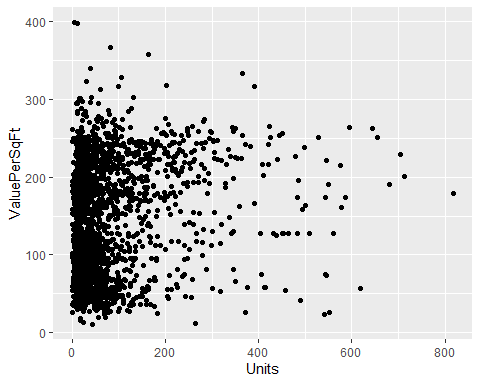
ggplot(housing, aes(x=SqFt, y=log(ValuePerSqFt)))+geom\_point()



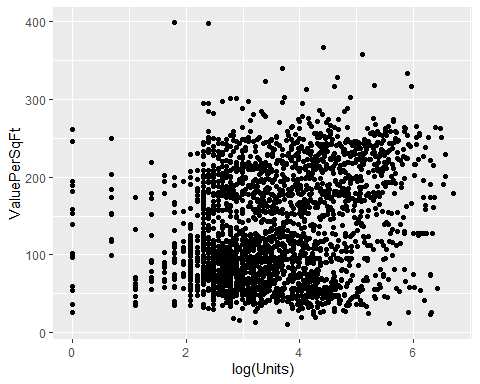
ggplot(housing, aes(x=log(SqFt), y=ValuePerSqFt))+geom\_point()



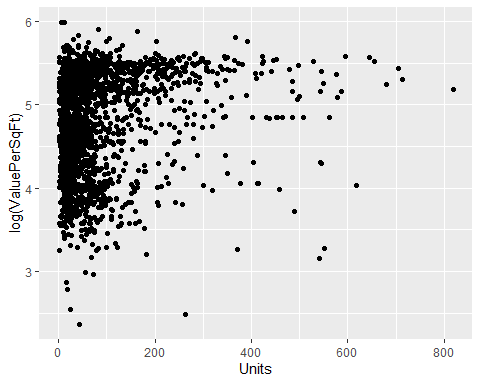
# Plot ValePerSqFt against Units  
ggplot(housing, aes(x=Units, y=ValuePerSqFt))+geom\_point()



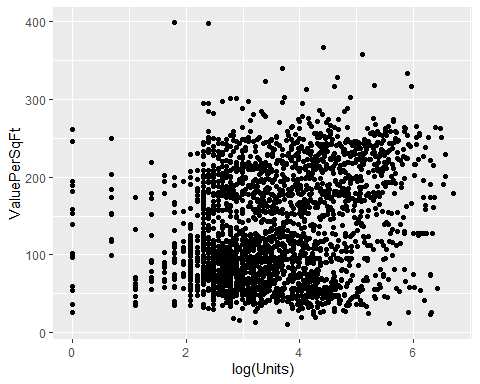
ggplot(housing, aes(x=log(Units), y=ValuePerSqFt))+geom\_point()



ggplot(housing, aes(x=Units, y=log(ValuePerSqFt)))+geom\_point()



ggplot(housing, aes(x=log(Units), y=ValuePerSqFt))+geom\_point()



# Fit the Model  
house1=lm(ValuePerSqFt~Units+SqFt+Boro, data=housing)  
summary(house1)

##   
## Call:  
## lm(formula = ValuePerSqFt ~ Units + SqFt + Boro, data = housing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -168.458 -22.680 1.493 26.290 261.761   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.430e+01 5.342e+00 8.293 < 2e-16 \*\*\*  
## Units -1.532e-01 2.421e-02 -6.330 2.88e-10 \*\*\*  
## SqFt 2.070e-04 2.129e-05 9.723 < 2e-16 \*\*\*  
## BoroBrooklyn 3.258e+01 5.561e+00 5.858 5.28e-09 \*\*\*  
## BoroManhattan 1.274e+02 5.459e+00 23.343 < 2e-16 \*\*\*  
## BoroQueens 3.011e+01 5.711e+00 5.272 1.46e-07 \*\*\*  
## BoroStaten Island -7.114e+00 1.001e+01 -0.711 0.477   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 43.2 on 2613 degrees of freedom  
## Multiple R-squared: 0.6034, Adjusted R-squared: 0.6025   
## F-statistic: 662.6 on 6 and 2613 DF, p-value: < 2.2e-16

# To get only Coefficients  
house1$coefficients

## (Intercept) Units SqFt BoroBrooklyn   
## 4.430325e+01 -1.532405e-01 2.069727e-04 3.257554e+01   
## BoroManhattan BoroQueens BoroStaten Island   
## 1.274259e+02 3.011000e+01 -7.113688e+00

coef(house1)

## (Intercept) Units SqFt BoroBrooklyn   
## 4.430325e+01 -1.532405e-01 2.069727e-04 3.257554e+01   
## BoroManhattan BoroQueens BoroStaten Island   
## 1.274259e+02 3.011000e+01 -7.113688e+00

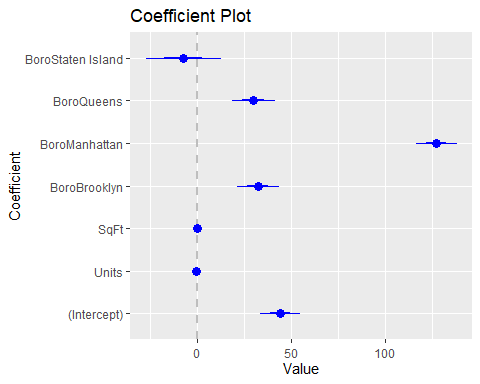
coefficients(house1)

## (Intercept) Units SqFt BoroBrooklyn   
## 4.430325e+01 -1.532405e-01 2.069727e-04 3.257554e+01   
## BoroManhattan BoroQueens BoroStaten Island   
## 1.274259e+02 3.011000e+01 -7.113688e+00

# Draw coefficients Plot  
#install.packages("coefplot")  
require(coefplot)

## Loading required package: coefplot

coefplot(house1)



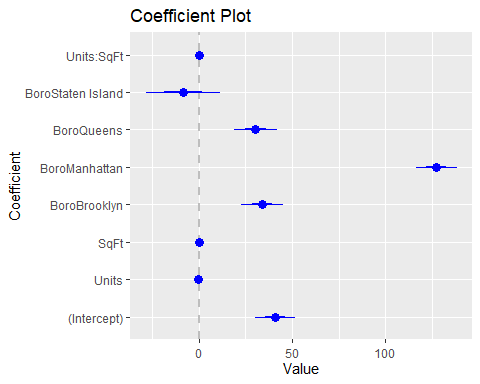
house2=lm(ValuePerSqFt~Units\*SqFt+Boro, data = housing)  
house3=lm(ValuePerSqFt~Units:SqFt+Boro, data = housing)  
house2$coefficients

## (Intercept) Units SqFt BoroBrooklyn   
## 4.093685e+01 -1.024579e-01 2.362293e-04 3.394544e+01   
## BoroManhattan BoroQueens BoroStaten Island Units:SqFt   
## 1.272102e+02 3.040115e+01 -8.419682e+00 -1.809587e-07

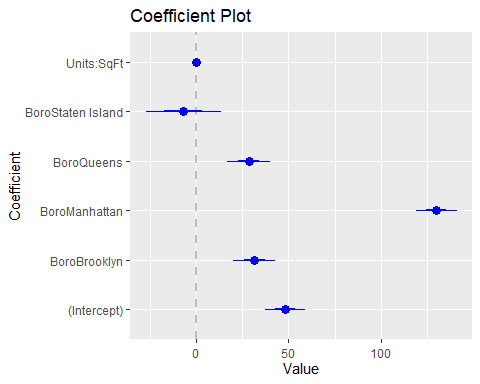
house3$coefficients

## (Intercept) BoroBrooklyn BoroManhattan BoroQueens   
## 4.804972e+01 3.141208e+01 1.302084e+02 2.841669e+01   
## BoroStaten Island Units:SqFt   
## -7.199902e+00 1.088059e-07

# draw Coef Plot  
coefplot(house2)



coefplot(house3)



house4=lm(ValuePerSqFt~ SqFt\*Units\*Income, data = housing)  
house4$coefficients

## (Intercept) SqFt Units Income   
## 1.116433e+02 -1.694688e-03 7.142611e-03 7.250830e-05   
## SqFt:Units SqFt:Income Units:Income SqFt:Units:Income   
## 3.158094e-06 -5.129522e-11 -1.279236e-07 9.107312e-14

# Interaction between TWO Categorical variables  
house5=lm(ValuePerSqFt~Units:Class\*Boro, data = housing)  
house5$coefficients

## (Intercept)   
## 5.494287e+01   
## BoroBrooklyn   
## 2.653390e+01   
## BoroManhattan   
## 1.191513e+02   
## BoroQueens   
## 2.671272e+01   
## BoroStaten Island   
## -1.297158e+01   
## Units:ClassR2-CONDOMINIUM   
## -3.041300e-01   
## Units:ClassR4-CONDOMINIUM   
## -1.062423e-01   
## Units:ClassR9-CONDOMINIUM   
## -4.872488e-02   
## Units:ClassRR-CONDOMINIUM   
## 5.268445e-03   
## Units:ClassR2-CONDOMINIUM:BoroBrooklyn   
## -2.656421e-05   
## Units:ClassR4-CONDOMINIUM:BoroBrooklyn   
## 1.255519e-01   
## Units:ClassR9-CONDOMINIUM:BoroBrooklyn   
## -1.039370e-01   
## Units:ClassRR-CONDOMINIUM:BoroBrooklyn   
## -6.328234e-01   
## Units:ClassR2-CONDOMINIUM:BoroManhattan   
## -1.156371e+00   
## Units:ClassR4-CONDOMINIUM:BoroManhattan   
## 2.433943e-01   
## Units:ClassR9-CONDOMINIUM:BoroManhattan   
## 1.276151e-02   
## Units:ClassRR-CONDOMINIUM:BoroManhattan   
## 1.430965e-02   
## Units:ClassR2-CONDOMINIUM:BoroQueens   
## 9.655360e-02   
## Units:ClassR4-CONDOMINIUM:BoroQueens   
## 6.848450e-02   
## Units:ClassR9-CONDOMINIUM:BoroQueens   
## -5.456138e-02   
## Units:ClassRR-CONDOMINIUM:BoroQueens   
## 6.057630e-01   
## Units:ClassR2-CONDOMINIUM:BoroStaten Island   
## NA   
## Units:ClassR4-CONDOMINIUM:BoroStaten Island   
## 1.048513e-01   
## Units:ClassR9-CONDOMINIUM:BoroStaten Island   
## NA   
## Units:ClassRR-CONDOMINIUM:BoroStaten Island   
## NA

# For ratio, need to use I() function  
house6=lm(ValuePerSqFt~ I(SqFt/Units)+Boro, data = housing)  
house6$coefficients

## (Intercept) I(SqFt/Units) BoroBrooklyn BoroManhattan   
## 43.754838763 0.004017039 30.774343209 130.769502685   
## BoroQueens BoroStaten Island   
## 29.767922792 -6.134446417

house7=lm(ValuePerSqFt~ (Units+SqFt)^2,housing)  
house7$coefficients

## (Intercept) Units SqFt Units:SqFt   
## 1.070301e+02 -1.125194e-01 4.964623e-04 -5.159669e-07

house8=lm(ValuePerSqFt~Units\*SqFt, housing)  
house8$coefficients

## (Intercept) Units SqFt Units:SqFt   
## 1.070301e+02 -1.125194e-01 4.964623e-04 -5.159669e-07

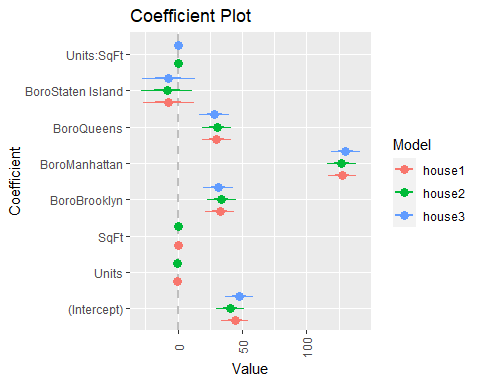
identical(house7$coefficients, house8$coefficients)

## [1] TRUE

house9=lm(ValuePerSqFt~ I(Units+SqFt)^2,housing)  
house9$coefficients

## (Intercept) I(Units + SqFt)   
## 1.147034e+02 2.107231e-04

# MULTIPLOT - coefplot  
multiplot(house1, house2, house3)



summary(house1)

##   
## Call:  
## lm(formula = ValuePerSqFt ~ Units + SqFt + Boro, data = housing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -168.458 -22.680 1.493 26.290 261.761   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.430e+01 5.342e+00 8.293 < 2e-16 \*\*\*  
## Units -1.532e-01 2.421e-02 -6.330 2.88e-10 \*\*\*  
## SqFt 2.070e-04 2.129e-05 9.723 < 2e-16 \*\*\*  
## BoroBrooklyn 3.258e+01 5.561e+00 5.858 5.28e-09 \*\*\*  
## BoroManhattan 1.274e+02 5.459e+00 23.343 < 2e-16 \*\*\*  
## BoroQueens 3.011e+01 5.711e+00 5.272 1.46e-07 \*\*\*  
## BoroStaten Island -7.114e+00 1.001e+01 -0.711 0.477   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 43.2 on 2613 degrees of freedom  
## Multiple R-squared: 0.6034, Adjusted R-squared: 0.6025   
## F-statistic: 662.6 on 6 and 2613 DF, p-value: < 2.2e-16

summary(house2)

##   
## Call:  
## lm(formula = ValuePerSqFt ~ Units \* SqFt + Boro, data = housing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -161.888 -22.867 1.802 26.431 261.733   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.094e+01 5.393e+00 7.590 4.41e-14 \*\*\*  
## Units -1.025e-01 2.728e-02 -3.755 0.000177 \*\*\*  
## SqFt 2.362e-04 2.245e-05 10.521 < 2e-16 \*\*\*  
## BoroBrooklyn 3.395e+01 5.556e+00 6.110 1.15e-09 \*\*\*  
## BoroManhattan 1.272e+02 5.444e+00 23.369 < 2e-16 \*\*\*  
## BoroQueens 3.040e+01 5.696e+00 5.338 1.02e-07 \*\*\*  
## BoroStaten Island -8.420e+00 9.985e+00 -0.843 0.399160   
## Units:SqFt -1.810e-07 4.530e-08 -3.995 6.65e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 43.08 on 2612 degrees of freedom  
## Multiple R-squared: 0.6058, Adjusted R-squared: 0.6047   
## F-statistic: 573.4 on 7 and 2612 DF, p-value: < 2.2e-16

summary(house3)

##   
## Call:  
## lm(formula = ValuePerSqFt ~ Units:SqFt + Boro, data = housing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -165.666 -22.840 0.161 27.351 263.442   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.805e+01 5.425e+00 8.857 < 2e-16 \*\*\*  
## BoroBrooklyn 3.141e+01 5.668e+00 5.542 3.30e-08 \*\*\*  
## BoroManhattan 1.302e+02 5.561e+00 23.414 < 2e-16 \*\*\*  
## BoroQueens 2.842e+01 5.822e+00 4.881 1.12e-06 \*\*\*  
## BoroStaten Island -7.200e+00 1.020e+01 -0.706 0.48   
## Units:SqFt 1.088e-07 1.926e-08 5.649 1.79e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 44.07 on 2614 degrees of freedom  
## Multiple R-squared: 0.5873, Adjusted R-squared: 0.5865   
## F-statistic: 743.8 on 5 and 2614 DF, p-value: < 2.2e-16

summary(house4)

##   
## Call:  
## lm(formula = ValuePerSqFt ~ SqFt \* Units \* Income, data = housing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -226.020 -26.409 -7.011 20.752 274.520   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.116e+02 1.382e+00 80.791 <2e-16 \*\*\*  
## SqFt -1.695e-03 5.324e-05 -31.834 <2e-16 \*\*\*  
## Units 7.143e-03 3.886e-02 0.184 0.854   
## Income 7.251e-05 1.320e-06 54.919 <2e-16 \*\*\*  
## SqFt:Units 3.158e-06 1.328e-07 23.775 <2e-16 \*\*\*  
## SqFt:Income -5.130e-11 1.835e-12 -27.959 <2e-16 \*\*\*  
## Units:Income -1.279e-07 4.785e-09 -26.733 <2e-16 \*\*\*  
## SqFt:Units:Income 9.107e-14 4.519e-15 20.152 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 40.47 on 2612 degrees of freedom  
## Multiple R-squared: 0.6522, Adjusted R-squared: 0.6513   
## F-statistic: 699.8 on 7 and 2612 DF, p-value: < 2.2e-16

summary(house5)

##   
## Call:  
## lm(formula = ValuePerSqFt ~ Units:Class \* Boro, data = housing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -173.145 -22.686 0.031 26.007 259.975   
##   
## Coefficients: (3 not defined because of singularities)  
## Estimate Std. Error t value  
## (Intercept) 5.494e+01 9.267e+00 5.929  
## BoroBrooklyn 2.653e+01 9.500e+00 2.793  
## BoroManhattan 1.192e+02 9.389e+00 12.691  
## BoroQueens 2.671e+01 9.636e+00 2.772  
## BoroStaten Island -1.297e+01 1.706e+01 -0.760  
## Units:ClassR2-CONDOMINIUM -3.041e-01 4.244e-01 -0.717  
## Units:ClassR4-CONDOMINIUM -1.062e-01 1.430e-01 -0.743  
## Units:ClassR9-CONDOMINIUM -4.872e-02 1.236e-01 -0.394  
## Units:ClassRR-CONDOMINIUM 5.268e-03 3.715e-01 0.014  
## Units:ClassR2-CONDOMINIUM:BoroBrooklyn -2.656e-05 4.524e-01 0.000  
## Units:ClassR4-CONDOMINIUM:BoroBrooklyn 1.256e-01 1.466e-01 0.857  
## Units:ClassR9-CONDOMINIUM:BoroBrooklyn -1.039e-01 1.770e-01 -0.587  
## Units:ClassRR-CONDOMINIUM:BoroBrooklyn -6.328e-01 7.391e-01 -0.856  
## Units:ClassR2-CONDOMINIUM:BoroManhattan -1.156e+00 4.757e-01 -2.431  
## Units:ClassR4-CONDOMINIUM:BoroManhattan 2.434e-01 1.435e-01 1.696  
## Units:ClassR9-CONDOMINIUM:BoroManhattan 1.276e-02 1.250e-01 0.102  
## Units:ClassRR-CONDOMINIUM:BoroManhattan 1.431e-02 3.725e-01 0.038  
## Units:ClassR2-CONDOMINIUM:BoroQueens 9.655e-02 4.309e-01 0.224  
## Units:ClassR4-CONDOMINIUM:BoroQueens 6.848e-02 1.465e-01 0.467  
## Units:ClassR9-CONDOMINIUM:BoroQueens -5.456e-02 1.300e-01 -0.420  
## Units:ClassRR-CONDOMINIUM:BoroQueens 6.058e-01 9.633e-01 0.629  
## Units:ClassR2-CONDOMINIUM:BoroStaten Island NA NA NA  
## Units:ClassR4-CONDOMINIUM:BoroStaten Island 1.049e-01 2.058e-01 0.510  
## Units:ClassR9-CONDOMINIUM:BoroStaten Island NA NA NA  
## Units:ClassRR-CONDOMINIUM:BoroStaten Island NA NA NA  
## Pr(>|t|)   
## (Intercept) 3.45e-09 \*\*\*  
## BoroBrooklyn 0.00526 \*\*   
## BoroManhattan < 2e-16 \*\*\*  
## BoroQueens 0.00561 \*\*   
## BoroStaten Island 0.44711   
## Units:ClassR2-CONDOMINIUM 0.47371   
## Units:ClassR4-CONDOMINIUM 0.45750   
## Units:ClassR9-CONDOMINIUM 0.69346   
## Units:ClassRR-CONDOMINIUM 0.98869   
## Units:ClassR2-CONDOMINIUM:BoroBrooklyn 0.99995   
## Units:ClassR4-CONDOMINIUM:BoroBrooklyn 0.39180   
## Units:ClassR9-CONDOMINIUM:BoroBrooklyn 0.55715   
## Units:ClassRR-CONDOMINIUM:BoroBrooklyn 0.39197   
## Units:ClassR2-CONDOMINIUM:BoroManhattan 0.01513 \*   
## Units:ClassR4-CONDOMINIUM:BoroManhattan 0.08999 .   
## Units:ClassR9-CONDOMINIUM:BoroManhattan 0.91868   
## Units:ClassRR-CONDOMINIUM:BoroManhattan 0.96936   
## Units:ClassR2-CONDOMINIUM:BoroQueens 0.82271   
## Units:ClassR4-CONDOMINIUM:BoroQueens 0.64028   
## Units:ClassR9-CONDOMINIUM:BoroQueens 0.67473   
## Units:ClassRR-CONDOMINIUM:BoroQueens 0.52952   
## Units:ClassR2-CONDOMINIUM:BoroStaten Island NA   
## Units:ClassR4-CONDOMINIUM:BoroStaten Island 0.61041   
## Units:ClassR9-CONDOMINIUM:BoroStaten Island NA   
## Units:ClassRR-CONDOMINIUM:BoroStaten Island NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 42.61 on 2598 degrees of freedom  
## Multiple R-squared: 0.6165, Adjusted R-squared: 0.6134   
## F-statistic: 198.8 on 21 and 2598 DF, p-value: < 2.2e-16

summary(house6)

##   
## Call:  
## lm(formula = ValuePerSqFt ~ I(SqFt/Units) + Boro, data = housing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -167.585 -23.001 0.282 27.558 261.709   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.375e+01 5.560e+00 7.869 5.18e-15 \*\*\*  
## I(SqFt/Units) 4.017e-03 9.438e-04 4.256 2.15e-05 \*\*\*  
## BoroBrooklyn 3.077e+01 5.684e+00 5.414 6.72e-08 \*\*\*  
## BoroManhattan 1.308e+02 5.574e+00 23.461 < 2e-16 \*\*\*  
## BoroQueens 2.977e+01 5.842e+00 5.096 3.72e-07 \*\*\*  
## BoroStaten Island -6.134e+00 1.023e+01 -0.600 0.549   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 44.18 on 2614 degrees of freedom  
## Multiple R-squared: 0.5851, Adjusted R-squared: 0.5843   
## F-statistic: 737.2 on 5 and 2614 DF, p-value: < 2.2e-16

summary(house7)

##   
## Call:  
## lm(formula = ValuePerSqFt ~ (Units + SqFt)^2, data = housing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -296.38 -46.65 -12.21 48.52 284.15   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.070e+02 1.874e+00 57.121 < 2e-16 \*\*\*  
## Units -1.125e-01 3.974e-02 -2.831 0.00467 \*\*   
## SqFt 4.965e-04 3.225e-05 15.393 < 2e-16 \*\*\*  
## Units:SqFt -5.160e-07 6.541e-08 -7.889 4.45e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 63.42 on 2616 degrees of freedom  
## Multiple R-squared: 0.1443, Adjusted R-squared: 0.1433   
## F-statistic: 147.1 on 3 and 2616 DF, p-value: < 2.2e-16

summary(house9)

##   
## Call:  
## lm(formula = ValuePerSqFt ~ I(Units + SqFt)^2, data = housing)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -266.37 -48.80 -14.04 52.70 279.58   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.147e+02 1.575e+00 72.83 <2e-16 \*\*\*  
## I(Units + SqFt) 2.107e-04 1.193e-05 17.67 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 64.79 on 2618 degrees of freedom  
## Multiple R-squared: 0.1065, Adjusted R-squared: 0.1062   
## F-statistic: 312.1 on 1 and 2618 DF, p-value: < 2.2e-16

housingNew=read.table("http://www.jaredlander.com/data/housingNew.csv", sep =",", header = TRUE, stringsAsFactors=FALSE)  
# write.table(housingNew, "Testhousing.csv", col.names = NA,row.names = TRUE, quote = FALSE, sep =",")  
# To predict with 95 % confidence bounds  
housePredict=predict(house1, newdata=housingNew, se.fit=TRUE, interval="prediction", level=0.95)  
head(housePredict$fit)

## fit lwr upr  
## 1 74.00645 -10.813887 158.8268  
## 2 82.04988 -2.728506 166.8283  
## 3 166.65975 81.808078 251.5114  
## 4 169.00970 84.222648 253.7968  
## 5 80.00129 -4.777303 164.7799  
## 6 47.87795 -37.480170 133.2361

View(housePredict$fit)  
  
house4Predict=predict(house4, newdata=housingNew, se.fit=TRUE, interval="prediction", level=0.95)  
head(house4Predict$fit)

## fit lwr upr  
## 1 110.87020 31.488812 190.2516  
## 2 70.55054 -8.841812 149.9429  
## 3 124.94771 45.360975 204.5344  
## 4 150.81847 71.369611 230.2673  
## 5 101.93912 22.563391 181.3149  
## 6 100.59737 21.223953 179.9708

View(house4Predict$fit)