

Multiple Linear Regression Model: Brooklyn Home Prices 2016-2020

ANambiar

2023-01-24

Part 1: Analyze 2016-2020 Brooklyn Home Prices

Step 1: Import and prepare the data for analysis

#1.1 Bring the data into R

```
library(dplyr)

#load data
data2016 <- read.csv('/Users/ankitanambiar/Desktop/2022-2023 school/Fall 2022/Statistical Analysis/stats final/2016_brooklyn.csv', header = T, skip = 4, fill=TRUE)
head(data2016)
```

##	BOROUGH.	NEIGHBORHOOD.		
## 1	3	BATH BEACH		
## 2	3	BATH BEACH		
## 3	3	BATH BEACH		
## 4	3	BATH BEACH		
## 5	3	BATH BEACH		
## 6	3	BATH BEACH		
##	BUILDING.CLASS.CATEGORY.	TAX.CLASS.AT.PRESENT.	BLOCK.	
## 1	01	ONE FAMILY DWELLINGS	1	6360
## 2	01	ONE FAMILY DWELLINGS	1	6361
## 3	01	ONE FAMILY DWELLINGS	1	6364
## 4	01	ONE FAMILY DWELLINGS	1	6364
## 5	01	ONE FAMILY DWELLINGS	1	6371
## 6	01	ONE FAMILY DWELLINGS	1	6374
##	LOT.	EASE.MENT.	BUILDING.CLASS.AT.PRESENT.	
## 1	13	NA	A5	
## 2	11	NA	A5	
## 3	2	NA	A5	
## 4	72	NA	A5	
## 5	79	NA	A5	
## 6	67	NA	S1	
##	ADDRESS.	APARTMENT.NUMBER.	ZIP.CODE.	
## 1	8665 15TH AVENUE		11228	
## 2	71 BAY 10TH STREET		11228	
## 3	1649 BENSON AVENUE		11214	
## 4	68 BAY 14TH STREET		11214	
## 5	8668 19TH AVENUE		11214	
## 6	8642 20TH AVENUE		11214	
##	RESIDENTIAL.UNITS.	COMMERCIAL.UNITS.	TOTAL.UNITS.	LAND.SQUARE.FEET.
## 1	1	-	1	1,547
## 2	1	-	1	2,900
## 3	1	-	1	1,638
## 4	1	-	1	1,950
## 5	1	-	1	2,223
## 6	1	1	2	1,740
##	GROSS.SQUARE.FEET.	YEAR.BUILT.	TAX.CLASS.AT.TIME.OF.SALE.	
## 1	2,224	1930	1	
## 2	1,660	1930	1	
## 3	972	1930	1	
## 4	972	1950	1	
## 5	2,520	1930	1	
## 6	3,240	1925	1	
##	BUILDING.CLASS.AT.TIME.OF.SALE.	SALE.PRICE.	SALE.DATE.	
## 1	A5	-	5/25/2016	
## 2	A5	829,000	4/5/2016	
## 3	A5	-	10/6/2016	
## 4	A5	790,000	6/21/2016	
## 5	A5	788,000	3/31/2016	
## 6	S1	1,090,000	10/24/2016	

```
data2017 <- read.csv('/Users/ankitanambiar/Desktop/2022-2023 school/Fall 2022/Statistical Analysis/stats final/2017_brooklyn.csv', header = T, skip = 4, fill=TRUE)
head(data2017)
```

```

## BOROUGH. NEIGHBORHOOD. BUILDING.CLASS.CATEGORY.
## 1 3 BATH BEACH 01 ONE FAMILY DWELLINGS
## 2 3 BATH BEACH 01 ONE FAMILY DWELLINGS
## 3 3 BATH BEACH 01 ONE FAMILY DWELLINGS
## 4 3 BATH BEACH 01 ONE FAMILY DWELLINGS
## 5 3 BATH BEACH 01 ONE FAMILY DWELLINGS
## 6 3 BATH BEACH 01 ONE FAMILY DWELLINGS
## TAX.CLASS.AS.OF.FINAL.ROLL.17.18 BLOCK. LOT. EASE.MENT.
## 1 1 6360 55 NA
## 2 1 6360 157 NA
## 3 1 6361 20 NA
## 4 1 6367 42 NA
## 5 1 6393 16 NA
## 6 1 6405 49 NA
## BUILDING.CLASS.AS.OF.FINAL.ROLL.17.18 ADDRESS. APARTMENT.NUMBER.
## 1 A5 28 BAY 10TH STREET
## 2 A5 36 BAY 10TH STREET
## 3 A5 47 BAY 10TH STREET
## 4 S1 1730 86TH STREET
## 5 A5 145 BAY 8TH STREET
## 6 S1 1863 BATH AVENUE
## ZIP.CODE. RESIDENTIAL.UNITS. COMMERCIAL.UNITS. TOTAL.UNITS. LAND.SQUARE.FEET.
## 1 11228 1 0 1 1,547
## 2 11228 1 0 1 1,547
## 3 11228 1 0 1 1,933
## 4 11214 1 1 2 1,342
## 5 11228 1 0 1 1,716
## 6 11214 1 1 2 1,172
## GROSS.SQUARE.FEET. YEAR.BUILT. TAX.CLASS.AT.TIME.OF.SALE.
## 1 1,428 1930 1
## 2 1,428 1930 1
## 3 1,660 1930 1
## 4 1,920 1926 1
## 5 1,554 1920 1
## 6 1,800 1920 1
## BUILDING.CLASS.AT.TIME.OF.SALE. SALE.PRICE. SALE.DATE.
## 1 A5 725,000 6/27/17
## 2 A5 800,000 2/27/17
## 3 A5 0 3/4/17
## 4 S1 1,200,000 3/31/17
## 5 A5 10 9/15/17
## 6 S1 1,060,000 4/21/17

```

```

data2018 <- read.csv('/Users/ankitanambiar/Desktop/2022-2023 school/Fall 2022/Statistical Analysis/stats final/2018_brooklyn.csv', header = T, skip = 4, fill=TRUE)
head(data2018)

```

```

## BOROUGH. NEIGHBORHOOD. BUILDING.CLASS.CATEGORY.
## 1      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 2      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 3      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 4      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 5      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 6      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## TAX.CLASS.AS.OF.FINAL.ROLL.18.19 BLOCK. LOT. EASE.MENT.
## 1      1      6360      23      NA
## 2      1      6366      69      NA
## 3      1      6366      72      NA
## 4      1      6367      41      NA
## 5      1      6371      20      NA
## 6      1      6373      73      NA
## BUILDING.CLASS.AS.OF.FINAL.ROLL.18.19 ADDRESS.
## 1      A5      8645 15TH AVENUE
## 2      A1      8658 BAY 16TH STREET
## 3      A1      8664 BAY 16TH STREET
## 4      S1      1728 86TH STREET
## 5      A9      75 BAY 20TH STREET
## 6      A1      50 BAY 23RD STREET
## APARTMENT.NUMBER. ZIP.CODE. RESIDENTIAL.UNITS. COMMERCIAL.UNITS. TOTAL.UNITS.
## 1      11228      1      0      1
## 2      11214      1      0      1
## 3      11214      1      0      1
## 4      11214      1      1      2
## 5      11214      1      0      1
## 6      11214      1      0      1
## LAND.SQUARE.FEET. GROSS.SQUARE.FEET. YEAR.BUILT. TAX.CLASS.AT.TIME.OF.SALE.
## 1      1,547      1,428      1930      1
## 2      4,833      1,724      1930      1
## 3      4,833      2,300      1925      1
## 4      1,342      1,920      1931      1
## 5      2,417      1,742      1930      1
## 6      2,417      1,672      1930      1
## BUILDING.CLASS.AT.TIME.OF.SALE. SALE.PRICE. SALE.DATE.
## 1      A5      $750,000      5/18/18
## 2      A1      $0      10/25/18
## 3      A1      $1,720,000      12/12/18
## 4      S1      $1,380,000      7/26/18
## 5      A9      $710,000      2/21/18
## 6      A1      $0      1/8/18

```

```

data2019 <- read.csv('/Users/ankitanambiar/Desktop/2022-2023 school/Fall 2022/Statistical Analysis/stats final/2019_brooklyn.csv', header = T, skip = 4, fill=TRUE)
head(data2018)

```

```

## BOROUGH. NEIGHBORHOOD. BUILDING.CLASS.CATEGORY.
## 1      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 2      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 3      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 4      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 5      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## 6      3      BATH BEACH  01 ONE FAMILY DWELLINGS
## TAX.CLASS.AS.OF.FINAL.ROLL.18.19 BLOCK. LOT. EASE.MENT.
## 1      1      6360      23      NA
## 2      1      6366      69      NA
## 3      1      6366      72      NA
## 4      1      6367      41      NA
## 5      1      6371      20      NA
## 6      1      6373      73      NA
## BUILDING.CLASS.AS.OF.FINAL.ROLL.18.19 ADDRESS.
## 1      A5      8645 15TH AVENUE
## 2      A1      8658 BAY 16TH STREET
## 3      A1      8664 BAY 16TH STREET
## 4      S1      1728 86TH STREET
## 5      A9      75 BAY 20TH STREET
## 6      A1      50 BAY 23RD STREET
## APARTMENT.NUMBER. ZIP.CODE. RESIDENTIAL.UNITS. COMMERCIAL.UNITS. TOTAL.UNITS.
## 1      11228      1      0      1
## 2      11214      1      0      1
## 3      11214      1      0      1
## 4      11214      1      1      2
## 5      11214      1      0      1
## 6      11214      1      0      1
## LAND.SQUARE.FEET. GROSS.SQUARE.FEET. YEAR.BUILT. TAX.CLASS.AT.TIME.OF.SALE.
## 1      1,547      1,428      1930      1
## 2      4,833      1,724      1930      1
## 3      4,833      2,300      1925      1
## 4      1,342      1,920      1931      1
## 5      2,417      1,742      1930      1
## 6      2,417      1,672      1930      1
## BUILDING.CLASS.AT.TIME.OF.SALE. SALE.PRICE. SALE.DATE.
## 1      A5      $750,000      5/18/18
## 2      A1      $0      10/25/18
## 3      A1      $1,720,000      12/12/18
## 4      S1      $1,380,000      7/26/18
## 5      A9      $710,000      2/21/18
## 6      A1      $0      1/8/18

```

```

data2020 <- read.csv('/Users/ankitanambiar/Desktop/2022-2023 school/Fall 2022/Statistical Analysis/stats final/2020_brooklyn.csv', header = T, skip = 6, fill=TRUE)
data2020 <- data2020[,-c(1),]
head(data2020)

```

##	BOROUGH	NEIGHBORHOOD	BUILDING.CLASS.CATEGORY	TAX.CLASS.AT.PRESENT	BLOCK	LOT
## 2	3	BATH BEACH	01 ONE FAMILY DWELLINGS		1	6359 70
## 3	3	BATH BEACH	01 ONE FAMILY DWELLINGS		1	6360 48
## 4	3	BATH BEACH	01 ONE FAMILY DWELLINGS		1	6360 56
## 5	3	BATH BEACH	01 ONE FAMILY DWELLINGS		1	6362 23
## 6	3	BATH BEACH	01 ONE FAMILY DWELLINGS		1	6367 24
## 7	3	BATH BEACH	01 ONE FAMILY DWELLINGS		1	6371 19
##	EASE.MENT	BUILDING.CLASS.AT.PRESENT	ADDRESS	APARTMENT.NUMBER		
## 2	NA	S1	8684 15TH AVENUE			
## 3	NA	A5	14 BAY 10TH STREET			
## 4	NA	A5	30 BAY 10TH STREET			
## 5	NA	A9	43 BAY 11TH STREET			
## 6	NA	A9	8645 BAY 16TH STREET			
## 7	NA	A9	79 BAY 20TH STREET			
##	ZIP.CODE	RESIDENTIAL.UNITS	COMMERCIAL.UNITS	TOTAL..UNITS	LAND..SQUARE.FEET	
## 2	11228	1	1	2	1,933	
## 3	11228	1	0	1	2,513	
## 4	11228	1	0	1	1,547	
## 5	11228	1	0	1	2,280	
## 6	11214	1	0	1	1,571	
## 7	11214	1	0	1	2,320	
##	GROSS..SQUARE.FEET	YEAR.BUILT	TAX.CLASS.AT.TIME.OF.SALE			
## 2	4,080	1930	1			
## 3	1,428	1930	1			
## 4	1,428	1930	1			
## 5	1,052	1901	1			
## 6	1,456	1935	1			
## 7	1,566	1930	1			
##	BUILDING.CLASS.AT.TIME.OF.SALE	SALE.PRICE	SALE.DATE			
## 2	S1	1,300,000	4/28/20			
## 3	A5	849,000	3/18/20			
## 4	A5	75,000	11/30/20			
## 5	A9	0	11/4/20			
## 6	A9	830,000	6/26/20			
## 7	A9	0	2/21/20			

```

# change column names
# change data types for DATE

colnames(data2016)=c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','bldclasscurr','address','aptnum','zip','resunits','comunits','totunits','landsqft','grosssqft','yrbuilt','taxclasssale','bldclasssale','price','date')
data2016$date <- as.Date(data2016$date, "%m/%d/%Y")

colnames(data2017)=c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','bldclasscurr','address','aptnum','zip','resunits','comunits','totunits','landsqft','grosssqft','yrbuilt','taxclasssale','bldclasssale','price','date')
data2017$date <- as.Date(data2017$date, "%m/%d/%Y")

colnames(data2018)=c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','bldclasscurr','address','aptnum','zip','resunits','comunits','totunits','landsqft','grosssqft','yrbuilt','taxclasssale','bldclasssale','price','date')
data2018$date <- as.Date(data2018$date, "%m/%d/%Y")

colnames(data2019)=c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','bldclasscurr','address','aptnum','zip','resunits','comunits','totunits','landsqft','grosssqft','yrbuilt','taxclasssale','bldclasssale','price','date')
data2019$date <- as.Date(data2019$date, "%m/%d/%Y")

colnames(data2020)=c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','bldclasscurr','address','aptnum','zip','resunits','comunits','totunits','landsqft','grosssqft','yrbuilt','taxclasssale','bldclasssale','price','date')
data2020$date <- as.Date(data2020$date, "%m/%d/%Y")

```

1.2 Join the data and make it usable for analysis

```

# bind
allyearsdata <- rbind(data2016, data2017, data2018, data2019, data2020)
#allyearsdata
# 119304 rows

# pay attention to white space and special characters
library(stringr)
allyearsdata$bldclasscat <- str_squish(allyearsdata$bldclasscat)
allyearsdata$bldclasssale <- str_squish(allyearsdata$bldclasssale)
allyearsdata$neighborhood <- str_squish(allyearsdata$neighborhood)
allyearsdata$address <- str_squish(allyearsdata$address)
allyearsdata$price <- str_squish(allyearsdata$price)
allyearsdata$grosssqft <- str_squish(allyearsdata$grosssqft)

# change data types
# pay attention to white space and special characters
class(allyearsdata$resunits) = "numeric"

```

```
## Warning in class(allyearsdata$resunits) = "numeric": NAs introduced by coercion
```

```
class(allyearsdata$comunits) = "numeric"
```

```
## Warning in class(allyearsdata$comunits) = "numeric": NAs introduced by coercion
```

```
class(allyearsdata$totunits) = "numeric"
```

```
## Warning in class(allyearsdata$totunits) = "numeric": NAs introduced by coercion
```

```
class(allyearsdata$zip) = "character"
```

```
allyearsdata$landsqft <- gsub(",", "", allyearsdata$landsqft)  
class(allyearsdata$landsqft) = "numeric"
```

```
## Warning in class(allyearsdata$landsqft) = "numeric": NAs introduced by coercion
```

```
allyearsdata$grosssqft <- gsub(",", "", allyearsdata$grosssqft)  
class(allyearsdata$grosssqft) = "numeric"
```

```
## Warning in class(allyearsdata$grosssqft) = "numeric": NAs introduced by coercion
```

```
allyearsdata$price <- gsub(",", "", allyearsdata$price)  
allyearsdata$price <- gsub("\\$", "", allyearsdata$price)  
allyearsdata$price <- str_squish(allyearsdata$price)  
#plot(allyearsdata$date, allyearsdata$price)  
class(allyearsdata$price) = "numeric"
```

```
## Warning in class(allyearsdata$price) = "numeric": NAs introduced by coercion
```

```
str(allyearsdata)
```



```
## 'data.frame':    119351 obs. of  21 variables:
## $ borough      : int  3 3 3 3 3 3 3 3 3 3 ...
## $ neighborhood: chr  "BATH BEACH" "BATH BEACH" "BATH BEACH" "BATH BEACH" ...
## $ bldclasscat  : chr  "01 ONE FAMILY DWELLINGS" "01 ONE FAMILY DWELLINGS" "01 ONE FAMILY DWELLINGS" "01 ONE FAMILY DWELLINGS" ...
## $ taxclasscurr: chr  "1" "1" "1" "1" ...
## $ block        : int  6360 6361 6364 6364 6371 6374 6376 6394 6401 6401 ...
## $ lot          : int  13 11 2 72 79 67 13 5 7 8 ...
## $ easement     : logi  NA NA NA NA NA NA ...
## $ bldclasscurr: chr  "A5" "A5" "A5" "A5" ...
## $ address      : chr  "8665 15TH AVENUE" "71 BAY 10TH STREET" "1649 BENSON AVENUE" "68 BAY 14TH STREET" ...
## $ aptnum       : chr  " " " " " " " " " ...
## $ zip          : chr  "11228" "11228" "11214" "11214" ...
## $ resunits     : num  1 1 1 1 1 1 1 1 1 1 ...
## $ comunits     : num  NA NA NA NA NA 1 NA NA 1 1 ...
## $ totunits     : num  1 1 1 1 1 2 1 1 2 2 ...
## $ landsqft     : num  1547 2900 1638 1950 2223 ...
## $ grosssqft    : num  2224 1660 972 972 2520 ...
## $ yrbuilt      : int  1930 1930 1930 1950 1930 1925 1915 1940 1920 1920 ...
## $ taxclasssale: num  1 1 1 1 1 1 1 1 1 1 ...
## $ bldclasssale: chr  "A5" "A5" "A5" "A5" ...
## $ price        : num  NA 829000 NA 790000 788000 1090000 NA 920000 40000 NA ...
## $ date         : Date, format: "2016-05-25" "2016-04-05" ...
```

```
#allyearsdata
#119304
```

1.3 Filter the data and make transformations specific to

this analysis

```
allyearsdata<- filter(allyearsdata, allyearsdata$bldclasssale %in% c("A0", "A1", "A2", "A3", "A4", "A5", "A6", "A7", "A8", "A9", "RA", "RB", "RG", "RH", "RK", "RP", "RR", "RS", "RT", "RW", "R0", "R1", "R2", "R3", "R4", "R5", "R6", "R7", "R8", "R9"))
#allyearsdata
#41910

allyearsdata<-allyearsdata[allyearsdata$totunits==1, ]
allyearsdata<-allyearsdata[allyearsdata$resunits==1, ]
#allyearsdata
#38850

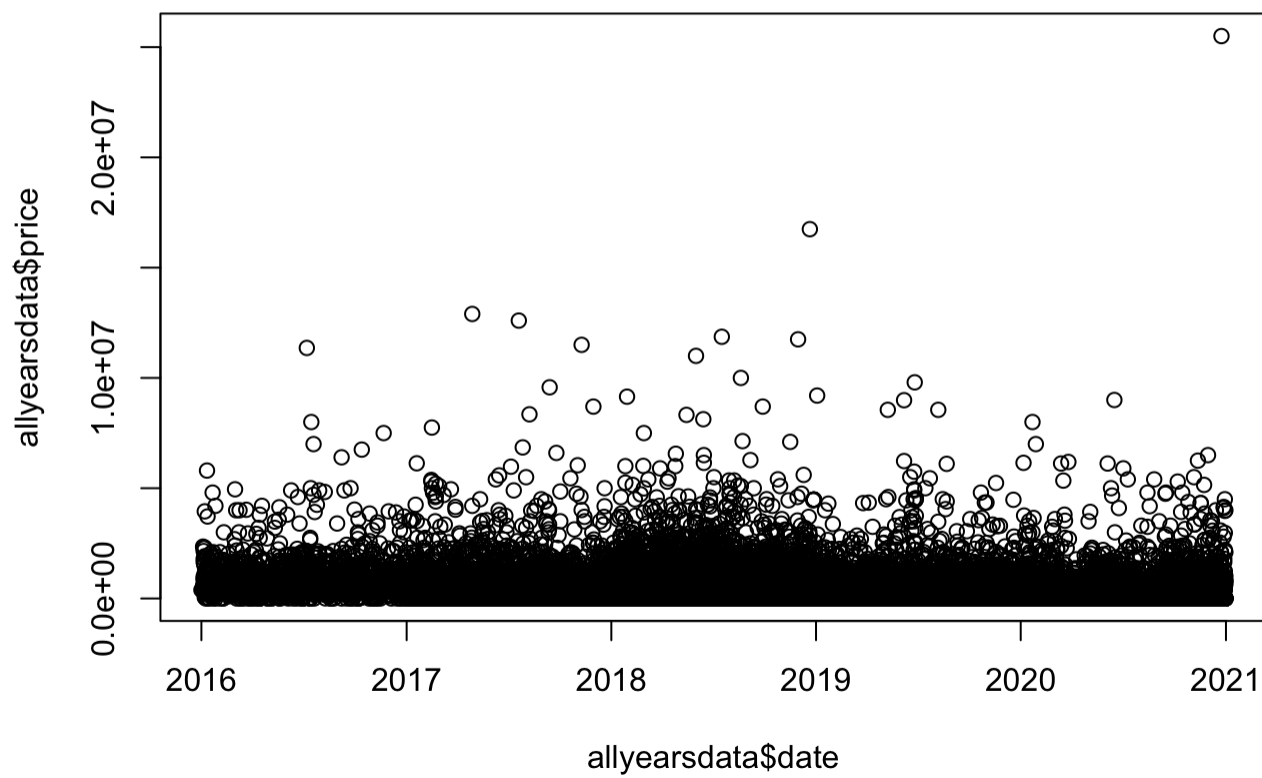
allyearsdata<-allyearsdata[!is.na(allyearsdata$grosssqft), ]
# allyearsdata %>% filter(!is.na(allyearsdata$grosssqft))
allyearsdata<-allyearsdata[allyearsdata$grosssqft>0, ]
#allyearsdata
#20774

allyearsdata<-allyearsdata[!is.na(allyearsdata$price), ]
# allyearsdata %>% filter(!is.na(allyearsdata$price))
#allyearsdata
#19593
```

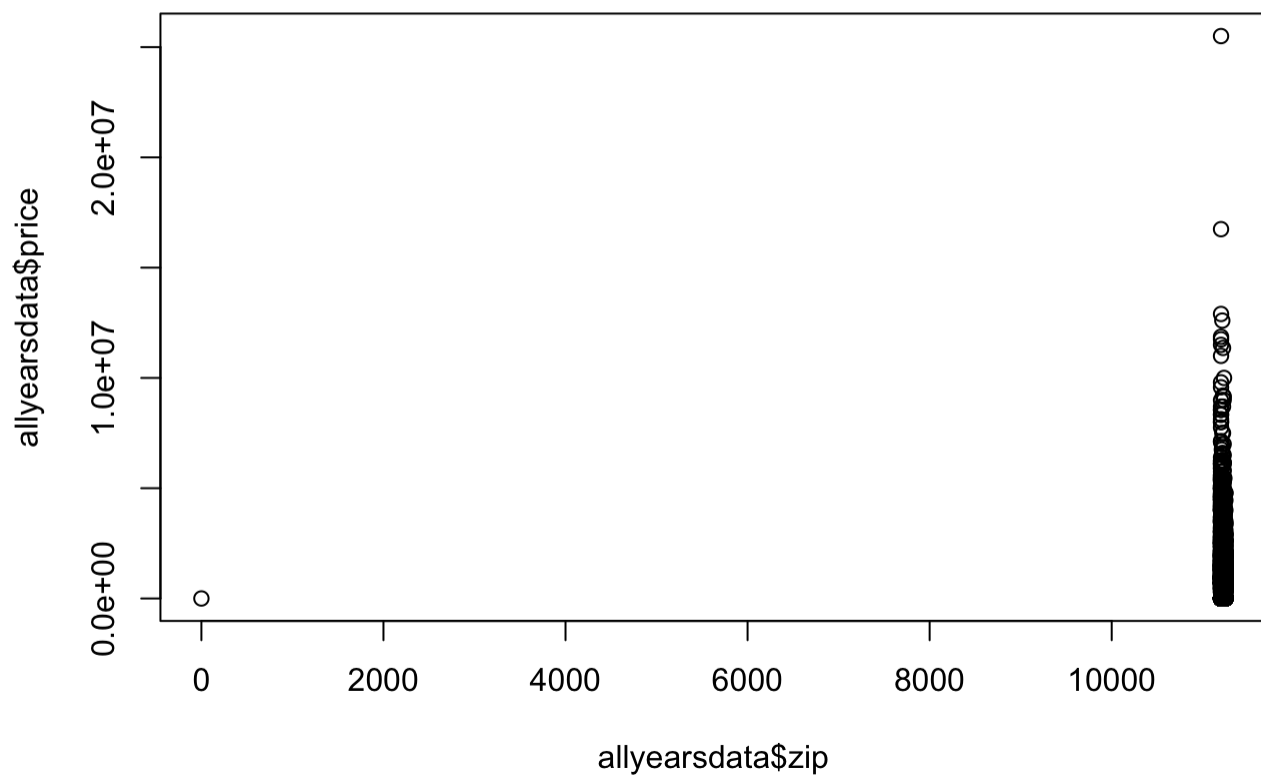
Step 2: EDA and feature engineering

2.1 Exploratory data analysis

```
#allyearsdata$price
plot(allyearsdata$date, allyearsdata$price)
```



```
plot(allyearsdata$zip, allyearsdata$price)
```



```
# revisited data cleaning steps  
# made zip code a character type  
# unique(allyearsdata$neighborhood)  
# many neighborhood categories, consider finding similarities
```

2.2 Pre-modeling and feature engineering

```
# outliers in price
allyearsdata <- allyearsdata[allyearsdata$price>6000, ]
allyearsdata <- allyearsdata[allyearsdata$price<6000000, ]
#allyearsdata
#13488 rows of data

allyearsdata$neighborhood[allyearsdata$neighborhood == "SHEEPSHEAD BAY"] <- "SHEEPSHEAD_BAY_AND_BE
RGEN_BEACH"
allyearsdata$neighborhood[allyearsdata$neighborhood == "BERGEN BEACH"] <- "SHEEPSHEAD_BAY_AND_BERG
EN_BEACH"

allyearsdata$neighborhood[allyearsdata$neighborhood == "WILLIAMSBURG-SOUTH"] <- "WILLIAMSBURG-SOUT
H_AND_WINDSOR_TERRACE"
allyearsdata$neighborhood[allyearsdata$neighborhood == "WINDSOR TERRACE"] <- "WILLIAMSBURG-SOUTH_A
ND_WINDSOR_TERRACE"

allyearsdata$neighborhood[allyearsdata$neighborhood == "BUSHWICK"] <- "BUSHWICK_AND_BOROUGH PARK_A
ND_SUNSET PARK"
allyearsdata$neighborhood[allyearsdata$neighborhood == "BOROUGH PARK"] <- "BUSHWICK_AND_BOROUGH PA
RK_AND_SUNSET PARK"
allyearsdata$neighborhood[allyearsdata$neighborhood == "SUNSET PARK"] <- "BUSHWICK_AND_BOROUGH PAR
K_AND_SUNSET PARK"

allyearsdata$neighborhood[allyearsdata$neighborhood == "DYKER HEIGHTS"] <- "DYKER _HEIGHTS_AND_WYC
KOFF_HEIGHTS"
allyearsdata$neighborhood[allyearsdata$neighborhood == "WYCKOFF HEIGHTS"] <- "DYKER _HEIGHTS_AND_W
YCKOFF_HEIGHTS"

allyearsdata$neighborhood[allyearsdata$neighborhood == "WILLIAMSBURG-CENTRAL"] <- "WILLIAMSBURG-CE
NTRAL_AND_FLATBUSH-NORTH"
allyearsdata$neighborhood[allyearsdata$neighborhood == "FLATBUSH-NORTH"] <- "WILLIAMSBURG-CENTRAL_
AND_FLATBUSH-NORTH"
#16017
#155981

allyearsdata$neighborhood[allyearsdata$neighborhood == "OLD MILL BASIN"] <- "OLD MILL BASIN_AND_CO
NEY ISLAND"
allyearsdata$neighborhood[allyearsdata$neighborhood == "CONEY ISLAND"] <- "OLD MILL BASIN_AND_CONE
Y ISLAND"
#-218274
#-218413

allyearsdata$neighborhood[allyearsdata$neighborhood == "KENSINGTON"] <- "KENSINGTON_AND_FLATBUSH-C
ENTRAL"
allyearsdata$neighborhood[allyearsdata$neighborhood == "FLATBUSH-CENTRAL"] <- "KENSINGTON_AND_FLAT
BUSH-CENTRAL"
#RMSE went up a little
#460595
#442170

allyearsdata$neighborhood[allyearsdata$neighborhood == "CYPRESS HILLS"] <- "CYPRESS HILLS_AND_GERR
```

```

ITSEN BACH"
allyearsdata$neighborhood[allyearsdata$neighborhood == "GERRITSEN BACH"] <- "CYPRESS HILLS_AND_GERRITSEN BACH"
#-254821
#-268848

allyearsdata$neighborhood[allyearsdata$neighborhood == "FLATLANDS"] <- "FLATLANDS_AND_SPRING_CREEK"
allyearsdata$neighborhood[allyearsdata$neighborhood == "SPRING CREEK"] <- "FLATLANDS_AND_SPRING_CREEK"
#-223755
#-238587

allyearsdata$neighborhood[allyearsdata$neighborhood == "DOWNTOWN-FULTON FERRY"] <- "DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE"
allyearsdata$neighborhood[allyearsdata$neighborhood == "PARK SLOPE"] <- "DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE"
#1298791
#1300480

# check neighborhood rows
# unique(allyearsdata$neighborhood)
#unique(data2019$neighborhood)

# model testing
#lm1 <- lm(price ~ grosssqft +neighborhood, data = allyearsdata)
#lm2 <- lm(price ~ grosssqft +landsqft + yrbuilt +block +neighborhood, data = allyearsdata)
#lm4 <- lm(price ~ grosssqft +landsqft + yrbuilt +block +zip, data = allyearsdata)
#lm5 <- lm(price ~ grosssqft +zip +block, data = allyearsdata)
#lm6 <- lm(price ~ grosssqft +landsqft + zip + yrbuilt +block +neighborhood, data = allyearsdata)

```

2.3 Reach a stopping point

```

# finalized model
lm0 <- lm(price ~ sqrt(grosssqft) + neighborhood, data = allyearsdata)
summary(lm0)

```

```
##
## Call:
## lm(formula = price ~ sqrt(grosssqft) + neighborhood, data = allyearsdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2974822  -192150    -914   160329  4014908
##
## Coefficients:
##                                     Estimate Std. Error
## (Intercept)                      -683973.1    41241.4
## sqrt(grosssqft)                   42983.1     494.7
## neighborhoodBAY RIDGE              62410.9    42332.1
## neighborhoodBEDFORD STUYVESANT    -39629.1    43576.9
## neighborhoodBENSONHURST          -69417.7    47591.8
## neighborhoodBOERUM HILL          1217711.4    54878.7
## neighborhoodBRIGHTON BEACH       -98770.8    55433.6
## neighborhoodBROOKLYN HEIGHTS     1624321.8    56719.6
## neighborhoodBROWNSVILLE        -439980.8    49068.8
## neighborhoodBUSH TERMINAL         20688.3   316999.6
## neighborhoodBUSHWICK_AND_BOROUGH PARK_AND_SUNSET PARK  57614.6    41347.4
## neighborhoodCANARSIE            -378511.0    42436.1
## neighborhoodCARROLL GARDENS       1039975.8    68916.2
## neighborhoodCLINTON HILL          512399.5    57706.4
## neighborhoodCOBBLE HILL          1285681.4    70998.3
## neighborhoodCOBBLE HILL-WEST      602965.4    76409.2
## neighborhoodCROWN HEIGHTS         34495.3    45893.6
## neighborhoodCYPRESS HILLS_AND_GERRITSEN BACH        -494297.2    50749.1
## neighborhoodDOWNTOWN-FULTON MALL    718209.7    56069.2
## neighborhoodDOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE    1042146.5    45118.6
## neighborhoodDOWNTOWN-METROTECH     415209.9    74026.2
## neighborhoodDYKER _HEIGHTS_AND_WYCKOFF_HEIGHTS      -9124.8    44903.0
## neighborhoodEAST NEW YORK        -406403.4    42900.7
## neighborhoodFLATBUSH-EAST        -405873.4    40669.8
## neighborhoodFLATBUSH-LEFFERTS GARDEN    186285.6    54346.9
## neighborhoodFLATLANDS_AND_SPRING_CREEK    -481395.0    45193.7
## neighborhoodFORT GREENE          630696.1    66996.2
## neighborhoodGERRITSEN BEACH       -337453.0    44781.7
## neighborhoodGOWANUS              691073.2    77764.4
## neighborhoodGRAVESEND            -90186.9    44860.4
## neighborhoodGREENPOINT           553905.1    50180.1
## neighborhoodKENSINGTON_AND_FLATBUSH-CENTRAL    52491.3    43375.7
## neighborhoodMADISON              -86669.9    42554.2
## neighborhoodMANHATTAN BEACH        115551.1    54833.6
## neighborhoodMARINE PARK          -246172.6    41331.9
## neighborhoodMIDWOOD              -42580.6    42765.8
## neighborhoodMILL BASIN           -357556.5    47620.7
## neighborhoodNAVY YARD             536307.5    93954.3
## neighborhoodOCEAN HILL           -263409.2    65649.1
## neighborhoodOCEAN PARKWAY-NORTH    295049.5    43941.1
## neighborhoodOCEAN PARKWAY-SOUTH    589023.6    53442.0
## neighborhoodOLD MILL BASIN_AND_CONEY ISLAND    -335886.9    41896.7
## neighborhoodPARK SLOPE SOUTH       643580.2    56899.8
## neighborhoodPROSPECT HEIGHTS      850126.9    53037.9
```

## neighborhoodRED HOOK	690324.8	89774.8
## neighborhoodSEAGATE	-499451.3	69862.0
## neighborhoodSHEEPSHEAD_BAY_AND_BERGEN_BEACH	-238155.1	41188.9
## neighborhoodWILLIAMSBURG-CENTRAL_AND_FLATBUSH-NORTH	-321568.3	47533.2
## neighborhoodWILLIAMSBURG-EAST	413456.7	51787.0
## neighborhoodWILLIAMSBURG-NORTH	745529.8	48509.8
## neighborhoodWILLIAMSBURG-SOUTH_AND_WINDSOR_TERRACE	518174.7	49570.2
##	t value	Pr(> t)
## (Intercept)	-16.585	< 2e-16 ***
## sqrt(grosssqft)	86.887	< 2e-16 ***
## neighborhoodBAY RIDGE	1.474	0.140420
## neighborhoodBEDFORD STUYVESANT	-0.909	0.363153
## neighborhoodBENSONHURST	-1.459	0.144697
## neighborhoodBOERUM HILL	22.189	< 2e-16 ***
## neighborhoodBRIGHTON BEACH	-1.782	0.074807 .
## neighborhoodBROOKLYN HEIGHTS	28.638	< 2e-16 ***
## neighborhoodBROWNSVILLE	-8.967	< 2e-16 ***
## neighborhoodBUSH TERMINAL	0.065	0.947966
## neighborhoodBUSHWICK_AND_BOROUGH PARK_AND_SUNSET PARK	1.393	0.163514
## neighborhoodCANARSIE	-8.920	< 2e-16 ***
## neighborhoodCARROLL GARDENS	15.090	< 2e-16 ***
## neighborhoodCLINTON HILL	8.879	< 2e-16 ***
## neighborhoodCOBBLE HILL	18.109	< 2e-16 ***
## neighborhoodCOBBLE HILL-WEST	7.891	3.22e-15 ***
## neighborhoodCROWN HEIGHTS	0.752	0.452283
## neighborhoodCYPRESS HILLS_AND_GERRITSEN BACH	-9.740	< 2e-16 ***
## neighborhoodDOWNTOWN-FULTON MALL	12.809	< 2e-16 ***
## neighborhoodDOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE	23.098	< 2e-16 ***
## neighborhoodDOWNTOWN-METROTECH	5.609	2.08e-08 ***
## neighborhoodDYKER _HEIGHTS_AND_WYCKOFF_HEIGHTS	-0.203	0.838973
## neighborhoodEAST NEW YORK	-9.473	< 2e-16 ***
## neighborhoodFLATBUSH-EAST	-9.980	< 2e-16 ***
## neighborhoodFLATBUSH-LEFFERTS GARDEN	3.428	0.000611 ***
## neighborhoodFLATLANDS_AND_SPRING_CREEK	-10.652	< 2e-16 ***
## neighborhoodFORT GREENE	9.414	< 2e-16 ***
## neighborhoodGERRITSEN BEACH	-7.536	5.17e-14 ***
## neighborhoodGOWANUS	8.887	< 2e-16 ***
## neighborhoodGRAVESEND	-2.010	0.044410 *
## neighborhoodGREENPOINT	11.038	< 2e-16 ***
## neighborhoodKENSINGTON_AND_FLATBUSH-CENTRAL	1.210	0.226241
## neighborhoodMADISON	-2.037	0.041700 *
## neighborhoodMANHATTAN BEACH	2.107	0.035109 *
## neighborhoodMARINE PARK	-5.956	2.65e-09 ***
## neighborhoodMIDWOOD	-0.996	0.319428
## neighborhoodMILL BASIN	-7.508	6.36e-14 ***
## neighborhoodNAVY YARD	5.708	1.17e-08 ***
## neighborhoodOCEAN HILL	-4.012	6.04e-05 ***
## neighborhoodOCEAN PARKWAY-NORTH	6.715	1.96e-11 ***
## neighborhoodOCEAN PARKWAY-SOUTH	11.022	< 2e-16 ***
## neighborhoodOLD MILL BASIN_AND_CONEY ISLAND	-8.017	1.17e-15 ***
## neighborhoodPARK SLOPE SOUTH	11.311	< 2e-16 ***
## neighborhoodPROSPECT HEIGHTS	16.029	< 2e-16 ***
## neighborhoodRED HOOK	7.690	1.58e-14 ***
## neighborhoodSEAGATE	-7.149	9.18e-13 ***


```
## neighborhoodSHEEPSHEAD_BAY_AND_BERGEN_BEACH -5.782 7.54e-09 ***
## neighborhoodWILLIAMSBURG-CENTRAL_AND_FLATBUSH-NORTH -6.765 1.39e-11 ***
## neighborhoodWILLIAMSBURG-EAST 7.984 1.53e-15 ***
## neighborhoodWILLIAMSBURG-NORTH 15.369 < 2e-16 ***
## neighborhoodWILLIAMSBURG-SOUTH_AND_WINDSOR_TERRACE 10.453 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 445100 on 13484 degrees of freedom
## Multiple R-squared:  0.6146, Adjusted R-squared:  0.6132
## F-statistic: 430 on 50 and 13484 DF, p-value: < 2.2e-16
```

```
#0.6146 = R^2
```

```
#445100 = RMSE
```

```
#50 = df
```

```
#meets conditions
```

```
saveRDS(list(model=lm0, data=allyearsdata), file='ankitanambiar.RDS')
```

Part 2: Analyze 2020 Q3 to Q4 shift

1.1 Prepare Data

```
library(lubridate)
library(dplyr)
data2020 <- allyearsdata
#data2020

data2020 <- data2020 %>%
  mutate(month = format(date,"%m"))
class(data2020$month) = "numeric"
str(data2020)
```

```
## 'data.frame':   13535 obs. of  22 variables:
## $ borough      : int  3 3 3 3 3 3 3 3 3 3 ...
## $ neighborhood: chr  "BATH BEACH" "BATH BEACH" "BATH BEACH" "BATH BEACH" ...
## $ bldclasscat  : chr  "01 ONE FAMILY DWELLINGS" "01 ONE FAMILY DWELLINGS" "01 ONE FAMILY DWELLI
NGS" "01 ONE FAMILY DWELLINGS" ...
## $ taxclasscurr: chr  "1" "1" "1" "1" ...
## $ block        : int  6361 6364 6371 6394 6401 6401 6402 6409 6414 6425 ...
## $ lot          : int  11 72 79 5 18 19 31 36 36 15 ...
## $ easement     : logi  NA NA NA NA NA NA ...
## $ bldclasscurr: chr  "A5" "A5" "A5" "A5" ...
## $ address      : chr  "71 BAY 10TH STREET" "68 BAY 14TH STREET" "8668 19TH AVENUE" "160 BAY 10T
H STREET" ...
## $ aptnum       : chr  " " " " " " " " " ...
## $ zip          : chr  "11228" "11214" "11214" "11228" ...
## $ resunits     : num  1 1 1 1 1 1 1 1 1 1 ...
## $ comunits     : num  NA NA NA NA NA NA NA NA NA NA ...
## $ totunits     : num  1 1 1 1 1 1 1 1 1 1 ...
## $ landsqft     : num  2900 1950 2223 2469 2417 ...
## $ grosssqft    : num  1660 972 2520 1836 1462 ...
## $ yrbuilt      : int  1930 1950 1930 1940 1925 1925 1924 1915 1910 1935 ...
## $ taxclasssale: num  1 1 1 1 1 1 1 1 1 1 ...
## $ bldclasssale: chr  "A5" "A5" "A5" "A5" ...
## $ price        : num  829000 790000 788000 920000 839000 854000 750000 699000 950000 699000 ...
## $ date         : Date, format: "2016-04-05" "2016-06-21" ...
## $ month        : num  4 6 3 2 7 2 7 2 9 5 ...
```

```
#data2020
```

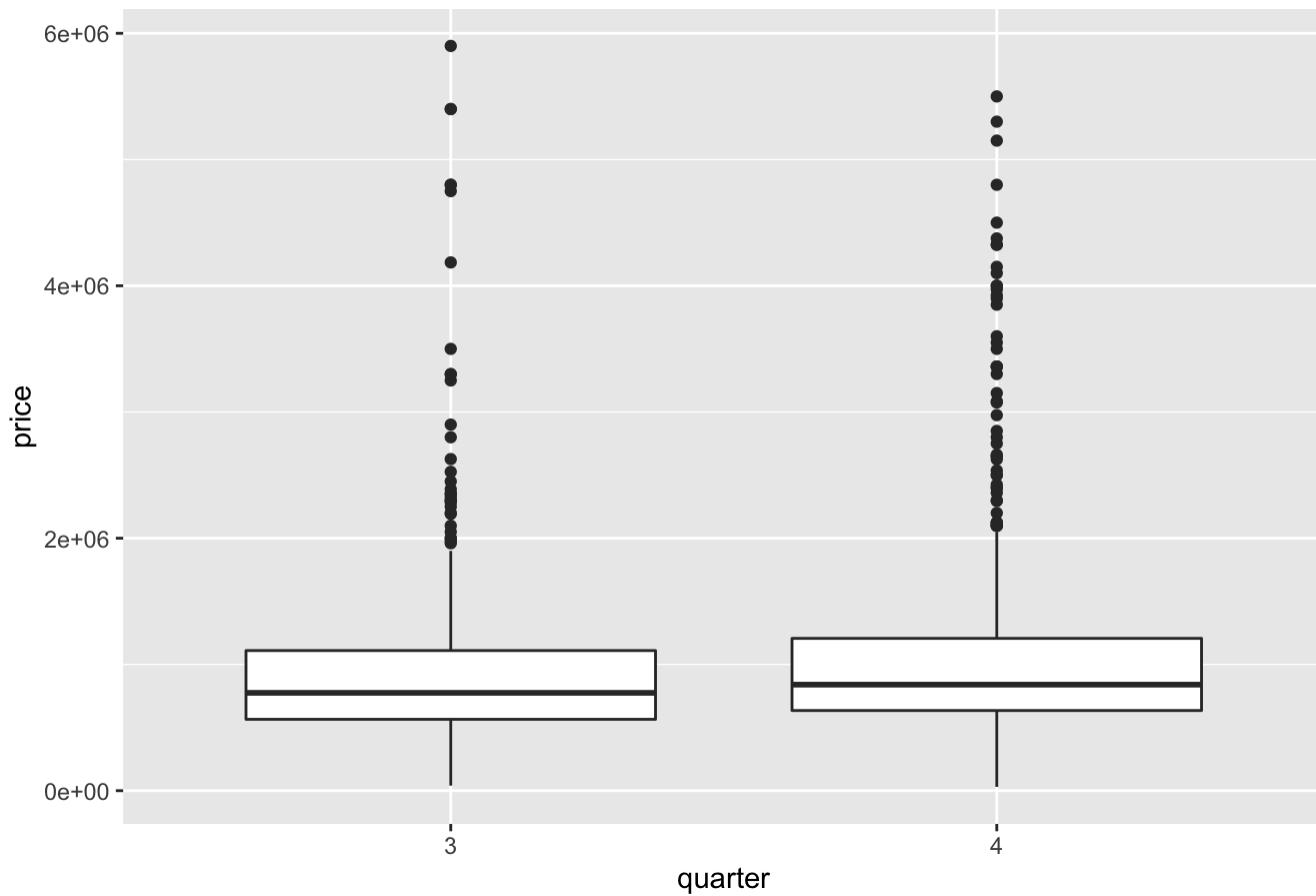
```
data2020 <- data2020 %>%
  mutate(quarter = case_when(month >6 & month <10 ~ "3",
                              month >9 & month <13 ~ "4")) %>%
  filter(year(ymd(date)) == "2020") %>%
  filter(quarter %in% c("3", "4"))

class(data2020$quarter) = "character"
```

1.2 Exploratory Analysis

```
#box plot
library(ggplot2)
library(gridExtra)
library(grid)
library(ggplot2)
library(lattice)
ggplot(data2020, aes(quarter, price)) +
  geom_boxplot() + ggtitle("Brooklyn Home Prices for 2020 Q3 and 2020 Q4")
```

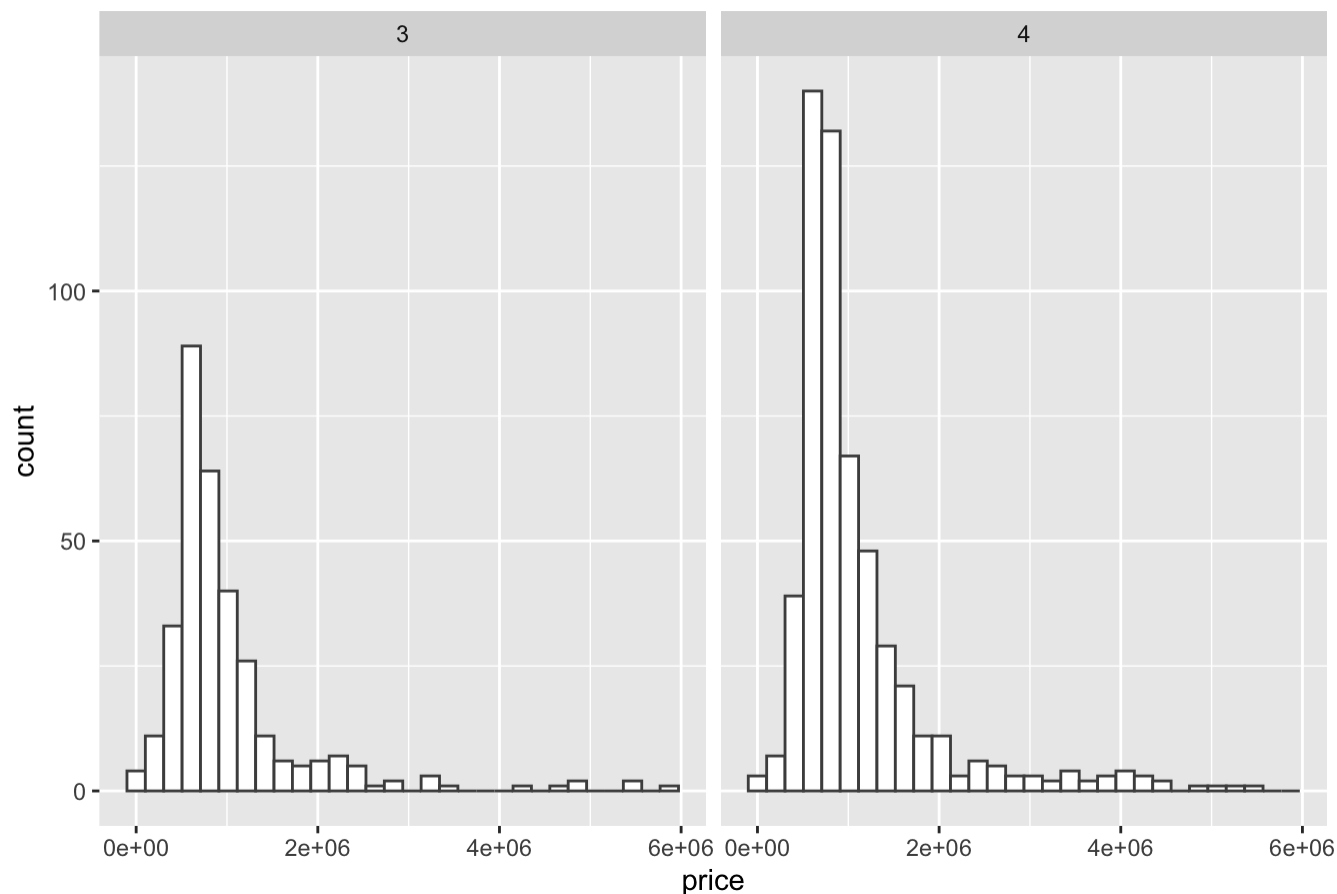
Brooklyn Home Prices for 2020 Q3 and 2020 Q4



```
#histogram
p1 <- ggplot(data2020, aes(price)) +
  geom_histogram(fill = "white", color = "grey30") +
  facet_wrap(~ quarter) + ggtitle("Brooklyn Home Price Distribution for Q3 and Q4 in 2020")
p1
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Brooklyn Home Price Distribution for Q3 and Q4 in 2020



```
#finding means of q3 and q4
library(ggplot2)
library(dplyr)
data2020 %>%
  group_by(quarter) %>%
  summarise(avg_price = mean(price))
```

```
## # A tibble: 2 × 2
##   quarter avg_price
##   <chr>      <dbl>
## 1 3         1011666.
## 2 4         1088537.
```

```
summary(data2020 %>% filter(quarter == 3) %>% .$quarter)
```

```
##   Length      Class      Mode
##      321 character character
```

```
summary(data2020 %>% filter(quarter == 4) %>% .$price)
```

```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 30000 635000   840000 1088537 1206500 5500000
```

```
#3 1011666
```

```
#4 1088537
```

```
#t test
```

```
t.test(data2020$price[data2020$quarter == '3'], data2020$price[data2020$quarter == '4'])
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: data2020$price[data2020$quarter == "3"] and data2020$price[data2020$quarter == "4"]
```

```
## t = -1.3255, df = 657.03, p-value = 0.1855
```

```
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -190743.32 37000.96
```

```
## sample estimates:
```

```
## mean of x mean of y
```

```
## 1011666 1088537
```

```
# fail to reject null that true difference in means between group 3 and group 4 is equal to 0
```

```
# cannot accept that there is statistical difference between the two means
```

```
#t.test(log(price) ~ quarter, data = data2020
```

```
# significant difference
```

```
# log for not normal data
```

1.3 Model Creation for Q3 to Q4 Shift

```
lm2 <- lm(price ~ quarter + sqrt(grosssqft) + neighborhood, data = data2020)
```

```
summary(lm2)
```

```
##
## Call:
## lm(formula = price ~ quarter + sqrt(grosssqft) + neighborhood,
##     data = data2020)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2880033  -177140       99   157940  2563666
##
## Coefficients:
##                                     Estimate Std. Error
## (Intercept)                      -807947      159336
## quarter4                          76132       32122
## sqrt(grosssqft)                   43490       2274
## neighborhoodBAY RIDGE             180786      147549
## neighborhoodBEDFORD STUYVESANT   -118567      173411
## neighborhoodBENSONHURST           87355      194725
## neighborhoodBOERUM HILL          1221890     262716
## neighborhoodBROOKLYN HEIGHTS     1882459     190803
## neighborhoodBROWNSVILLE        -349132      179487
## neighborhoodBUSHWICK_AND_BOROUGH PARK_AND_SUNSET PARK  256623     151836
## neighborhoodCANARSIE             -298247      153871
## neighborhoodCARROLL GARDENS       2723131     470430
## neighborhoodCLINTON HILL          1767951     345145
## neighborhoodCOBBLE HILL          1983457     342395
## neighborhoodCOBBLE HILL-WEST       812774     465480
## neighborhoodCROWN HEIGHTS         214619     191932
## neighborhoodCYPRESS HILLS_AND_GERRITSEN BACH        -281984     182418
## neighborhoodDOWNTOWN-FULTON MALL    964679     346361
## neighborhoodDOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE    1936394     204466
## neighborhoodDYKER _HEIGHTS_AND_WYCKOFF_HEIGHTS        5449      155107
## neighborhoodEAST NEW YORK         -282242     156285
## neighborhoodFLATBUSH-EAST         -285946     144186
## neighborhoodFLATBUSH-LEFFERTS GARDEN    292624     227622
## neighborhoodFLATLANDS_AND_SPRING_CREEK    -363731     156428
## neighborhoodFORT GREENE           2025301     468162
## neighborhoodGERRITSEN BEACH        -152491     167202
## neighborhoodGOWANUS               824296     465038
## neighborhoodGRAVESEND              11295     158495
## neighborhoodGREENPOINT            485940     260034
## neighborhoodKENSINGTON_AND_FLATBUSH-CENTRAL    247871     157441
## neighborhoodMADISON                2654      150128
## neighborhoodMANHATTAN BEACH        399487     187286
## neighborhoodMARINE PARK           -134926     143962
## neighborhoodMIDWOOD               34597      153176
## neighborhoodMILL BASIN            -256486     173900
## neighborhoodOCEAN HILL            -312288     342467
## neighborhoodOCEAN PARKWAY-NORTH      473612     153777
## neighborhoodOCEAN PARKWAY-SOUTH     1546299     191595
## neighborhoodOLD MILL BASIN_AND_CONEY ISLAND    -196530     149080
## neighborhoodPARK SLOPE SOUTH        822716     215428
## neighborhoodPROSPECT HEIGHTS       1284518     343406
## neighborhoodRED HOOK               859319     465099
## neighborhoodSEAGATE               -322489     207628
```

## neighborhoodSHEEPSHEAD_BAY_AND_BERGEN_BEACH	-181057	148726	
## neighborhoodWILLIAMSBURG-CENTRAL_AND_FLATBUSH-NORTH	-159046	206853	
## neighborhoodWILLIAMSBURG-EAST	922445	465008	
## neighborhoodWILLIAMSBURG-NORTH	2141688	227184	
## neighborhoodWILLIAMSBURG-SOUTH_AND_WINDSOR_TERRACE	791847	194956	
##	t value	Pr(> t)	
## (Intercept)	-5.071	4.90e-07	***
## quarter4	2.370	0.018013	*
## sqrt(grosssqft)	19.129	< 2e-16	***
## neighborhoodBAY RIDGE	1.225	0.220825	
## neighborhoodBEDFORD STUYVESANT	-0.684	0.494333	
## neighborhoodBENSONHURST	0.449	0.653835	
## neighborhoodBOERUM HILL	4.651	3.85e-06	***
## neighborhoodBROOKLYN HEIGHTS	9.866	< 2e-16	***
## neighborhoodBROWNSVILLE	-1.945	0.052095	.
## neighborhoodBUSHWICK_AND_BOROUGH PARK_AND_SUNSET PARK	1.690	0.091381	.
## neighborhoodCANARSIE	-1.938	0.052928	.
## neighborhoodCARROLL GARDENS	5.789	1.01e-08	***
## neighborhoodCLINTON HILL	5.122	3.76e-07	***
## neighborhoodCOBBLE HILL	5.793	9.84e-09	***
## neighborhoodCOBBLE HILL-WEST	1.746	0.081166	.
## neighborhoodCROWN HEIGHTS	1.118	0.263807	
## neighborhoodCYPRESS HILLS_AND_GERRITSEN BACH	-1.546	0.122534	
## neighborhoodDOWNTOWN-FULTON MALL	2.785	0.005472	**
## neighborhoodDOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE	9.470	< 2e-16	***
## neighborhoodDYKER _HEIGHTS_AND_WYCKOFF_HEIGHTS	0.035	0.971985	
## neighborhoodEAST NEW YORK	-1.806	0.071292	.
## neighborhoodFLATBUSH-EAST	-1.983	0.047679	*
## neighborhoodFLATBUSH-LEFFERTS GARDEN	1.286	0.198954	
## neighborhoodFLATLANDS_AND_SPRING_CREEK	-2.325	0.020301	*
## neighborhoodFORT GREENE	4.326	1.70e-05	***
## neighborhoodGERRITSEN BEACH	-0.912	0.362028	
## neighborhoodGOWANUS	1.773	0.076675	.
## neighborhoodGRAVESEND	0.071	0.943205	
## neighborhoodGREENPOINT	1.869	0.062011	.
## neighborhoodKENSINGTON_AND_FLATBUSH-CENTRAL	1.574	0.115784	
## neighborhoodMADISON	0.018	0.985898	
## neighborhoodMANHATTAN BEACH	2.133	0.033217	*
## neighborhoodMARINE PARK	-0.937	0.348915	
## neighborhoodMIDWOOD	0.226	0.821363	
## neighborhoodMILL BASIN	-1.475	0.140620	
## neighborhoodOCEAN HILL	-0.912	0.362099	
## neighborhoodOCEAN PARKWAY-NORTH	3.080	0.002140	**
## neighborhoodOCEAN PARKWAY-SOUTH	8.071	2.46e-15	***
## neighborhoodOLD MILL BASIN_AND_CONEY ISLAND	-1.318	0.187774	
## neighborhoodPARK SLOPE SOUTH	3.819	0.000144	***
## neighborhoodPROSPECT HEIGHTS	3.741	0.000196	***
## neighborhoodRED HOOK	1.848	0.065017	.
## neighborhoodSEAGATE	-1.553	0.120758	
## neighborhoodSHEEPSHEAD_BAY_AND_BERGEN_BEACH	-1.217	0.223804	
## neighborhoodWILLIAMSBURG-CENTRAL_AND_FLATBUSH-NORTH	-0.769	0.442183	
## neighborhoodWILLIAMSBURG-EAST	1.984	0.047619	*
## neighborhoodWILLIAMSBURG-NORTH	9.427	< 2e-16	***
## neighborhoodWILLIAMSBURG-SOUTH_AND_WINDSOR_TERRACE	4.062	5.34e-05	***

```
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 445100 on 825 degrees of freedom  
## Multiple R-squared:  0.7224, Adjusted R-squared:  0.7066  
## F-statistic: 45.69 on 47 and 825 DF,  p-value: < 2.2e-16
```

Model Creation Explanation:

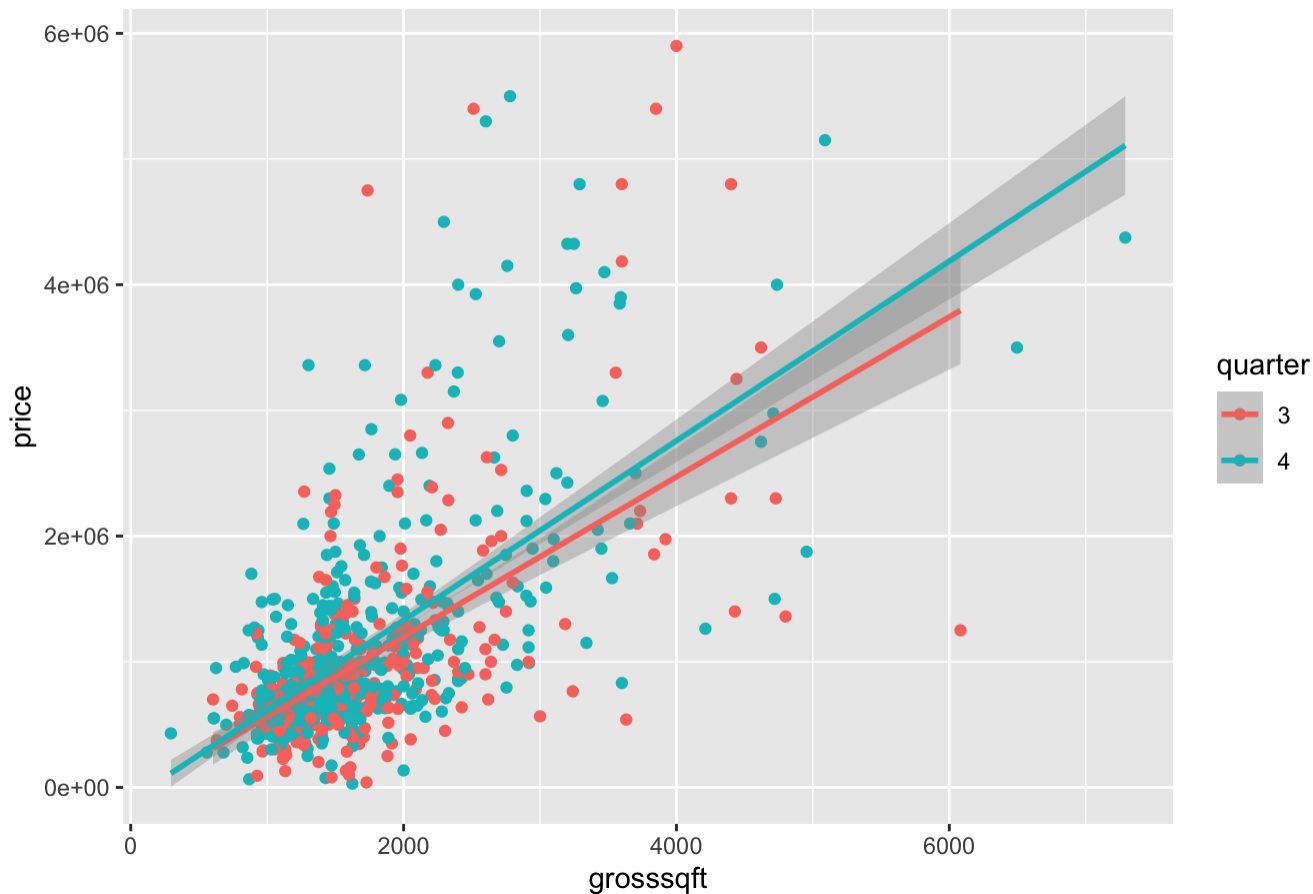
To answer if Brooklyn home purchase prices changed between Quarter 3 (Q3) and Quarter 4 (Q4) in 2020, I have created a price predictor model to analyze the Q3 to Q4 shift. My Brooklyn home price predictor model for this business case originated from a previous Brooklyn home price predictor model I created from data on real estate purchases within the borough of Brooklyn from 2016-2020. My previous model met all the conditions for predictive accuracy, so I decided to include the 2 predictor variables I had in my new model: gross square feet and neighborhood. Along with the predictor variables, I used the same filters that worked for my previous model, ex: only prices between \$6K to \$6M. I kept only 2020 data in Q3 and Q4 since those are the periods we are focusing on. I then created a new column for Quarter in the dataset, made it a 'character' type, and added it as a categorical variable into my price predictor model.

The final Brooklyn home price predictor model uses the following 3 predictors: gross square feet, neighborhood, and quarter. The regression output for this model has promising qualities. The residual standard error (RMSE) was 443400, the multiple R-squared is 0.7219, and has 39 degrees of freedom. These values show that the Brooklyn home price predictor model for Q3 and Q4 of 2020 has high predictive accuracy, so we can feel confident in the model's predictive power and proceed in using it to answer if Brooklyn home purchase prices changed between Q3 2020 and Q4 2020.

1.4 Q3 to Q4 Shift Analysis with Gross Square Feet

```
ggplot(data2020, aes(x=grosssqft, y=price, color=quarter)) + geom_point() +  
  stat_smooth(method = "lm",  
              formula = y ~ x,  
              geom = "smooth") + ggtitle("Brooklyn Home Price based on Gross Square Feet for 2020  
Q3 to 2020 Q4")
```


Brooklyn Home Price based on Gross Square Feet for 2020 Q3 to 2020 Q4



Gross Square Feet Analysis:

For the predictor variable gross square feet, I found gross square feet had a significant relation with price, with a p-value of $< 2e-16$.

I then created the “Brooklyn Home Price based on Gross Square Feet for 2020 Q3 to 2020 Q4” plot, below. The graph shows how price increases with greater square feet, which is expected. More importantly, when comparing the linear regression lines of Q3 and Q4, we see that Q4 homes were sold at higher prices than Q3 for the same square feet. The difference in price between the quarters is present for all square feet but gets larger as square feet increases, seen in the not parallel linear regression lines of Q4 and Q3. At 2000 square feet, there is an approximate \$100k increase from Q3 to Q4, while, at 6000 square feet, there is an approximate \$400k increase. There is an effect of quarter and square feet on price, signaling that square feet is important to consider for the price change between 2020 Q3 to 2020 Q4.

1.4 Q3 to Q4 Shift Analysis with Neighborhood

```
#plot, with 5 most significant neighborhoods
#filter significant neighborhoods
filtered_neighborhood_data <- filter(data2020, neighborhood=="BROOKLYN HEIGHTS" | neighborhood=="DO
WNTOWN-FULTON_FERRY_AND_PARK_SLOPE"
                                   | neighborhood=="OCEAN PARKWAY-SOUTH" | neighborhood=="WILLIAMS
BURG-NORTH")
filtered_neighborhood_data
```

##	borough	neighborhood
## 445710	3	BROOKLYN HEIGHTS
## 445810	3	BROOKLYN HEIGHTS
## 445910	3	BROOKLYN HEIGHTS
## 446110	3	BROOKLYN HEIGHTS
## 466110	3	BROOKLYN HEIGHTS
## 466210	3	BROOKLYN HEIGHTS
## 466310	3	BROOKLYN HEIGHTS
## 466410	3	BROOKLYN HEIGHTS
## 466510	3	BROOKLYN HEIGHTS
## 466610	3	BROOKLYN HEIGHTS
## 466810	3	BROOKLYN HEIGHTS
## 153941	3	OCEAN PARKWAY-SOUTH
## 154141	3	OCEAN PARKWAY-SOUTH
## 154231	3	OCEAN PARKWAY-SOUTH
## 154241	3	OCEAN PARKWAY-SOUTH
## 154251	3	OCEAN PARKWAY-SOUTH
## 154261	3	OCEAN PARKWAY-SOUTH
## 154281	3	OCEAN PARKWAY-SOUTH
## 154301	3	OCEAN PARKWAY-SOUTH
## 154311	3	OCEAN PARKWAY-SOUTH
## 154471	3	OCEAN PARKWAY-SOUTH
## 154561	3	OCEAN PARKWAY-SOUTH
## 158961	3	DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE
## 158971	3	DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE
## 159001	3	DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE
## 159021	3	DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE
## 159071	3	DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE
## 159081	3	DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE
## 159091	3	DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE
## 159131	3	DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE
## 159151	3	DOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE
## 191231	3	WILLIAMSBURG-NORTH
## 191241	3	WILLIAMSBURG-NORTH
## 191741	3	WILLIAMSBURG-NORTH
## 191751	3	WILLIAMSBURG-NORTH
## 191761	3	WILLIAMSBURG-NORTH
## 191771	3	WILLIAMSBURG-NORTH

##	bldclasscat	taxclasscurr	block	lot	easement
## 445710	01 ONE FAMILY DWELLINGS		1	260	43 NA
## 445810	01 ONE FAMILY DWELLINGS		1	260	57 NA
## 445910	01 ONE FAMILY DWELLINGS		1	263	60 NA
## 446110	01 ONE FAMILY DWELLINGS		1	268	26 NA
## 466110	11 SPECIAL CONDO BILLING LOTS		2	239	1106 NA
## 466210	11 SPECIAL CONDO BILLING LOTS		2	239	1120 NA
## 466310	11 SPECIAL CONDO BILLING LOTS		2	239	1134 NA
## 466410	11 SPECIAL CONDO BILLING LOTS		2	239	1143 NA
## 466510	11 SPECIAL CONDO BILLING LOTS		2	239	1154 NA
## 466610	11 SPECIAL CONDO BILLING LOTS		2	239	1166 NA
## 466810	11 SPECIAL CONDO BILLING LOTS		2	239	1199 NA
## 153941	01 ONE FAMILY DWELLINGS		1	6681	292 NA
## 154141	01 ONE FAMILY DWELLINGS		1	7104	516 NA
## 154231	01 ONE FAMILY DWELLINGS		1	7107	68 NA
## 154241	01 ONE FAMILY DWELLINGS		1	7107	89 NA

## 154251	01 ONE FAMILY DWELLINGS	1	7107	133	NA	
## 154261	01 ONE FAMILY DWELLINGS	1	7107	157	NA	
## 154281	01 ONE FAMILY DWELLINGS	1	7109	66	NA	
## 154301	01 ONE FAMILY DWELLINGS	1	7125	42	NA	
## 154311	01 ONE FAMILY DWELLINGS	1	7126	2	NA	
## 154471	01 ONE FAMILY DWELLINGS	1	7184	117	NA	
## 154561	01 ONE FAMILY DWELLINGS	1	7224	35	NA	
## 158961	01 ONE FAMILY DWELLINGS	1	942	7	NA	
## 158971	01 ONE FAMILY DWELLINGS	1	942	8	NA	
## 159001	01 ONE FAMILY DWELLINGS	1	952	26	NA	
## 159021	01 ONE FAMILY DWELLINGS	1	988	29	NA	
## 159071	01 ONE FAMILY DWELLINGS	1	1073	39	NA	
## 159081	01 ONE FAMILY DWELLINGS	1	1075	58	NA	
## 159091	01 ONE FAMILY DWELLINGS	1	1077	17	NA	
## 159131	01 ONE FAMILY DWELLINGS	1	1088	57	NA	
## 159151	01 ONE FAMILY DWELLINGS	1	1095	19	NA	
## 191231	01 ONE FAMILY DWELLINGS	1	2336	15	NA	
## 191241	01 ONE FAMILY DWELLINGS	1	2336	24	NA	
## 191741	11 SPECIAL CONDO BILLING LOTS	2	2331	1144	NA	
## 191751	11 SPECIAL CONDO BILLING LOTS	2	2337	1201	NA	
## 191761	11 SPECIAL CONDO BILLING LOTS	2	2337	1202	NA	
## 191771	11 SPECIAL CONDO BILLING LOTS	2	2337	1203	NA	
##	bldclasscurr	address	aptnum	zip	resunits	comunits
## 445710	A4	288 HICKS STREET		11201	1	0
## 445810	A4	314 HICKS STREET		11201	1	0
## 445910	A9	48 SIDNEY PLACE		11201	1	0
## 446110	A9	135 CLINTON STREET		11201	1	0
## 466110	RR	1 CLINTON STREET, 4C	4C	11201	1	0
## 466210	RR	1 CLINTON STREET, 5F	5F	11201	1	0
## 466310	RR	1 CLINTON STREET, 8C	8C	11201	1	0
## 466410	RR	1 CLINTON STREET, 10B	10B	11201	1	0
## 466510	RR	1 CLINTON STREET, 12C	12C	11201	1	0
## 466610	RR	1 CLINTON STREET, 16B	16B	11201	1	0
## 466810	RR	1 CLINTON STREET, 24C	24C	11201	1	0
## 153941	A3	1925 EAST 3RD STREET		11223	1	0
## 154141	A1	2071 EAST 1ST STREET		11223	1	0
## 154231	A1	2097 EAST 4TH STREET		11223	1	0
## 154241	A1	2047 EAST 4TH STREET		11223	1	0
## 154251	A1	1996 EAST 5TH STREET		11223	1	0
## 154261	A1	2056 EAST 5TH STREET		11223	1	0
## 154281	A1	2057 OCEAN PARKWAY		11223	1	0
## 154301	A5	81 GRAVESEND NECK ROAD		11223	1	0
## 154311	A5	2182 WEST STREET		11223	1	0
## 154471	A5	2401 EAST 7TH STREET		11223	1	0
## 154561	A9	729 GERALD COURT		11235	1	0
## 158961	A4	123 6 AVENUE		11217	1	0
## 158971	A4	121 6 AVENUE		11217	1	0
## 159001	A5	720 SACKETT STREET		11217	1	0
## 159021	A5	378 5TH STREET		11215	1	0
## 159071	A4	31 PROSPECT PARK WEST		11215	1	0
## 159081	A4	541 1 STREET		11215	1	0
## 159091	A4	544 1ST STREET		11215	1	0
## 159131	A4	469 8TH STREET		11215	1	0
## 159151	A4	658 10TH STREET, 4	4	11215	1	0

## 191231	A4	176 NORTH 6TH STREET	11211	1	0
## 191241	A9	580 DRIGGS AVE	11211	1	0
## 191741	RR 308 NORTH 7TH STREET, PHB	PHB 11211	1	0	
## 191751	RR 207 NORTH 5TH STREET, A	A 11211	1	0	
## 191761	RR 207 NORTH 5TH STREET, B	B 11211	1	0	
## 191771	RR 207 NORTH 5TH STREET, PH	PH 11211	1	0	

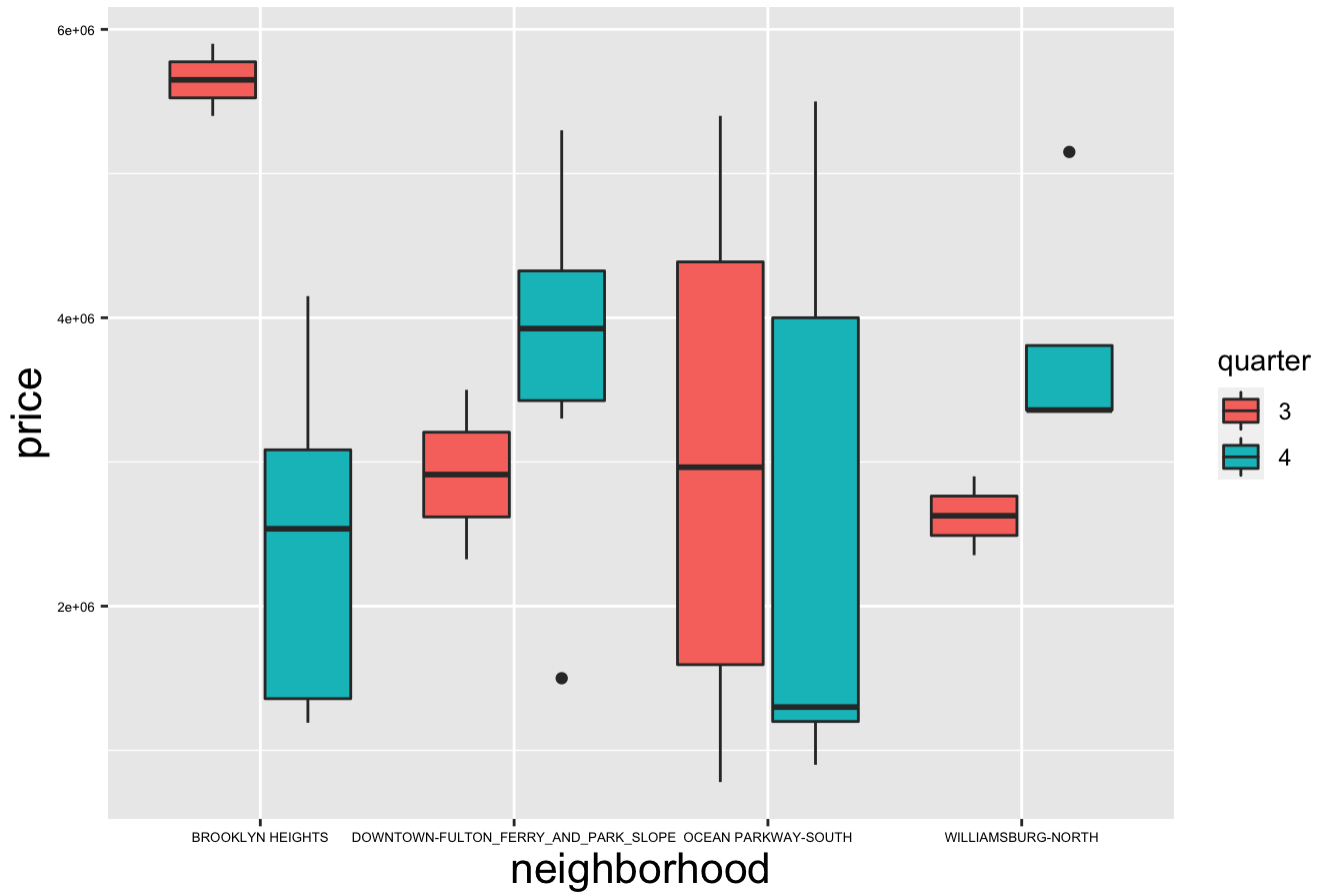
##	totunits	landsqft	grosssqft	yrbuilt	taxclass	sale	bldclass	sale	price
## 445710	1	1750	3264	1930		1	A4		3972000
## 445810	1	1873	4000	2012		1	A4		5900000
## 445910	1	2117	3850	1901		1	A9		5400000
## 446110	1	1163	2760	1901		1	A9		4150000
## 466110	1	25477	1672	2018		2	RR		2649000
## 466210	1	25477	1065	2018		2	RR		1358346
## 466310	1	25477	1456	2018		2	RR		2300000
## 466410	1	25477	932	2018		2	RR		1191353
## 466510	1	25477	1456	2018		2	RR		2536564
## 466610	1	25477	907	2018		2	RR		1271794
## 466810	1	25477	1981	2018		2	RR		3083993
## 153941	1	5000	2780	1920		1	A3		5500000
## 154141	1	2325	2610	1925		1	A1		2627000
## 154231	1	3600	2176	1992		1	A1		3300000
## 154241	1	4000	1736	1925		1	A1		4750000
## 154251	1	4000	2400	1931		1	A1		4000000
## 154261	1	4000	2513	1925		1	A1		5400000
## 154281	1	5000	6082	1899		1	A1		1250000
## 154301	1	1769	1024	1930		1	A5		780000
## 154311	1	3947	2100	1945		1	A5		1200000
## 154471	1	1743	1536	1920		1	A5		1300000
## 154561	1	2000	974	1930		1	A9		900000
## 158961	1	1992	3248	1899		1	A4		4325000
## 158971	1	1992	3200	1899		1	A4		4325000
## 159001	1	1667	1500	1899		1	A5		2325000
## 159021	1	1567	2700	1901		1	A5		3550000
## 159071	1	2350	2603	1901		1	A4		5300000
## 159081	1	2000	4620	1901		1	A4		3500000
## 159091	1	2100	4720	1910		1	A4		1500000
## 159131	1	1700	2397	1901		1	A4		3301000
## 159151	1	1875	2529	1920		1	A4		3925000
## 191231	1	2000	5089	1899		1	A4		5150000
## 191241	1	1280	2325	1899		1	A9		2900000
## 191741	1	13800	1271	2017		2	RR		2353500
## 191751	1	2500	1716	2017		2	RR		3360000
## 191761	1	2500	1303	2017		2	RR		3360000
## 191771	1	2500	2233	2017		2	RR		3360000

##	date	month	quarter
## 445710	2020-12-30	12	4
## 445810	2020-07-02	7	3
## 445910	2020-07-10	7	3
## 446110	2020-12-28	12	4
## 466110	2020-12-10	12	4
## 466210	2020-12-15	12	4
## 466310	2020-12-14	12	4
## 466410	2020-12-10	12	4
## 466510	2020-12-21	12	4

##	466610	2020-12-17	12	4
##	466810	2020-12-22	12	4
##	153941	2020-11-05	11	4
##	154141	2020-07-07	7	3
##	154231	2020-08-03	8	3
##	154241	2020-09-15	9	3
##	154251	2020-12-14	12	4
##	154261	2020-08-26	8	3
##	154281	2020-07-15	7	3
##	154301	2020-07-21	7	3
##	154311	2020-10-29	10	4
##	154471	2020-11-12	11	4
##	154561	2020-10-23	10	4
##	158961	2020-11-17	11	4
##	158971	2020-11-17	11	4
##	159001	2020-07-20	7	3
##	159021	2020-12-04	12	4
##	159071	2020-10-07	10	4
##	159081	2020-09-04	9	3
##	159091	2020-10-22	10	4
##	159131	2020-12-04	12	4
##	159151	2020-10-13	10	4
##	191231	2020-11-23	11	4
##	191241	2020-09-16	9	3
##	191741	2020-07-14	7	3
##	191751	2020-11-06	11	4
##	191761	2020-11-06	11	4
##	191771	2020-11-06	11	4

```
#plot
ggplot(filtered_neighborhood_data, aes(x=neighborhood, y=price, fill=quarter)) + geom_boxplot() +
ggtitle("Home Prices for Neighborhoods, 2020 Q3 to Q4")+ theme(axis.title = element_text(size = 1
6),axis.text = element_text(size= 5, color = "black"))
```

Home Prices for Neighborhoods, 2020 Q3 to Q4



```
lmnew <- lm(price ~ quarter + grosssqft + neighborhood, data = data2020)
summary(lmnew)
```

```
##
## Call:
## lm(formula = price ~ quarter + grosssqft + neighborhood, data = data2020)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -3252865  -164968     2451   140098   2577184
##
## Coefficients:
##                                     Estimate Std. Error
## (Intercept)                      150113.86   139157.26
## quarter4                          75382.74    31935.80
## grosssqft                          461.34       23.67
## neighborhoodBAY RIDGE             213056.79   146557.83
## neighborhoodBEDFORD STUYVESANT    -78669.73   172167.76
## neighborhoodBENSONHURST           108240.39   193569.39
## neighborhoodBOERUM HILL           1253855.59   260912.11
## neighborhoodBROOKLYN HEIGHTS      1883124.27   189677.44
## neighborhoodBROWNSVILLE         -349238.12   178454.65
## neighborhoodBUSHWICK_AND_BOROUGH PARK_AND_SUNSET PARK 258589.35   150949.38
## neighborhoodCANARSIE             -307069.35   152990.00
## neighborhoodCARROLL GARDENS        2620004.03   468386.83
## neighborhoodCLINTON HILL           1796663.95   342944.12
## neighborhoodCOBBLE HILL           2001533.52   340397.51
## neighborhoodCOBBLE HILL-WEST       839713.12   462794.72
## neighborhoodCROWN HEIGHTS          254392.52   190487.53
## neighborhoodCYPRESS HILLS_AND_GERRITSEN BACH    -262312.81   181349.59
## neighborhoodDOWNTOWN-FULTON MALL      955951.60   344334.44
## neighborhoodDOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE 1941965.58   203122.86
## neighborhoodDYKER _HEIGHTS_AND_WYCKOFF_HEIGHTS    31344.89   154160.54
## neighborhoodEAST NEW YORK          -283368.04   155385.53
## neighborhoodFLATBUSH-EAST          -263025.60   143321.87
## neighborhoodFLATBUSH-LEFFERTS GARDEN    347907.81   225943.22
## neighborhoodFLATLANDS_AND_SPRING_CREEK    -330892.77   155460.27
## neighborhoodFORT GREENE            2017381.47   465436.76
## neighborhoodGERRITSEN BEACH        -199028.39   166170.96
## neighborhoodGOWANUS                789713.09   462334.29
## neighborhoodGRAVESEND               982.46    157594.68
## neighborhoodGREENPOINT             499953.25   258533.15
## neighborhoodKENSINGTON_AND_FLATBUSH-CENTRAL    290335.17   156110.80
## neighborhoodMADISON                36533.77   149184.13
## neighborhoodMANHATTAN BEACH         449318.79   185872.29
## neighborhoodMARINE PARK            -127366.20   143132.11
## neighborhoodMIDWOOD                60539.00   152149.63
## neighborhoodMILL BASIN             -242737.74   172674.57
## neighborhoodOCEAN HILL             -373753.14   340387.99
## neighborhoodOCEAN PARKWAY-NORTH      479520.87   152859.97
## neighborhoodOCEAN PARKWAY-SOUTH     1546900.75   190449.20
## neighborhoodOLD MILL BASIN_AND_CONEY ISLAND    -193732.12   148222.72
## neighborhoodPARK SLOPE SOUTH        849797.35   214134.91
## neighborhoodPROSPECT HEIGHTS       1347511.28   341268.10
## neighborhoodRED HOOK               806620.04   462364.34
## neighborhoodSEAGATE               -296672.44   206315.43
## neighborhoodSHEEPSHEAD_BAY_AND_BERGEN_BEACH    -164229.90   147820.81
```

## neighborhoodWILLIAMSBURG-CENTRAL_AND_FLATBUSH-NORTH	-136069.39	205644.27	
## neighborhoodWILLIAMSBURG-EAST	957498.14	462312.09	
## neighborhoodWILLIAMSBURG-NORTH	2141939.05	225843.78	
## neighborhoodWILLIAMSBURG-SOUTH_AND_WINDSOR_TERRACE	811989.80	193786.70	
##	t value	Pr(> t)	
## (Intercept)	1.079	0.28102	
## quarter4	2.360	0.01848	*
## grosssqft	19.487	< 2e-16	***
## neighborhoodBAY RIDGE	1.454	0.14640	
## neighborhoodBEDFORD STUYVESANT	-0.457	0.64784	
## neighborhoodBENSONHURST	0.559	0.57619	
## neighborhoodBOERUM HILL	4.806	1.83e-06	***
## neighborhoodBROOKLYN HEIGHTS	9.928	< 2e-16	***
## neighborhoodBROWNSVILLE	-1.957	0.05068	.
## neighborhoodBUSHWICK_AND_BOROUGH PARK_AND_SUNSET PARK	1.713	0.08707	.
## neighborhoodCANARSIE	-2.007	0.04506	*
## neighborhoodCARROLL GARDENS	5.594	3.02e-08	***
## neighborhoodCLINTON HILL	5.239	2.05e-07	***
## neighborhoodCOBBLE HILL	5.880	5.96e-09	***
## neighborhoodCOBBLE HILL-WEST	1.814	0.06997	.
## neighborhoodCROWN HEIGHTS	1.335	0.18209	
## neighborhoodCYPRESS HILLS_AND_GERRITSEN BACH	-1.446	0.14843	
## neighborhoodDOWNTOWN-FULTON MALL	2.776	0.00562	**
## neighborhoodDOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE	9.561	< 2e-16	***
## neighborhoodDYKER _HEIGHTS_AND_WYCKOFF_HEIGHTS	0.203	0.83893	
## neighborhoodEAST NEW YORK	-1.824	0.06857	.
## neighborhoodFLATBUSH-EAST	-1.835	0.06683	.
## neighborhoodFLATBUSH-LEFFERTS GARDEN	1.540	0.12399	
## neighborhoodFLATLANDS_AND_SPRING_CREEK	-2.128	0.03359	*
## neighborhoodFORT GREENE	4.334	1.64e-05	***
## neighborhoodGERRITSEN BEACH	-1.198	0.23137	
## neighborhoodGOWANUS	1.708	0.08799	.
## neighborhoodGRAVESEND	0.006	0.99503	
## neighborhoodGREENPOINT	1.934	0.05348	.
## neighborhoodKENSINGTON_AND_FLATBUSH-CENTRAL	1.860	0.06327	.
## neighborhoodMADISON	0.245	0.80660	
## neighborhoodMANHATTAN BEACH	2.417	0.01585	*
## neighborhoodMARINE PARK	-0.890	0.37381	
## neighborhoodMIDWOOD	0.398	0.69081	
## neighborhoodMILL BASIN	-1.406	0.16017	
## neighborhoodOCEAN HILL	-1.098	0.27252	
## neighborhoodOCEAN PARKWAY-NORTH	3.137	0.00177	**
## neighborhoodOCEAN PARKWAY-SOUTH	8.122	1.66e-15	***
## neighborhoodOLD MILL BASIN_AND_CONEY ISLAND	-1.307	0.19157	
## neighborhoodPARK SLOPE SOUTH	3.969	7.86e-05	***
## neighborhoodPROSPECT HEIGHTS	3.949	8.53e-05	***
## neighborhoodRED HOOK	1.745	0.08143	.
## neighborhoodSEAGATE	-1.438	0.15083	
## neighborhoodSHEEPSHEAD_BAY_AND_BERGEN_BEACH	-1.111	0.26689	
## neighborhoodWILLIAMSBURG-CENTRAL_AND_FLATBUSH-NORTH	-0.662	0.50837	
## neighborhoodWILLIAMSBURG-EAST	2.071	0.03866	*
## neighborhoodWILLIAMSBURG-NORTH	9.484	< 2e-16	***
## neighborhoodWILLIAMSBURG-SOUTH_AND_WINDSOR_TERRACE	4.190	3.09e-05	***
## ---			


```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 442500 on 825 degrees of freedom
## Multiple R-squared:  0.7256, Adjusted R-squared:  0.71
## F-statistic: 46.42 on 47 and 825 DF,  p-value: < 2.2e-16
```

```
confint(lmnew)
```

##	2.5 %
## (Intercept)	-123030.0766
## quarter4	12697.7575
## grosssqft	414.8679
## neighborhoodBAY RIDGE	-74613.3013
## neighborhoodBEDFORD STUYVESANT	-416608.1232
## neighborhoodBENSONHURST	-271706.0572
## neighborhoodBOERUM HILL	741725.9261
## neighborhoodBROOKLYN HEIGHTS	1510817.1283
## neighborhoodBROWNSVILLE	-699516.6971
## neighborhoodBUSHWICK_AND_BOROUGH PARK_AND_SUNSET PARK	-37700.6755
## neighborhoodCANARSIE	-607364.8004
## neighborhoodCARROLL GARDENS	1700633.9297
## neighborhoodCLINTON HILL	1123518.2668
## neighborhoodCOBBLE HILL	1333386.4444
## neighborhoodCOBBLE HILL-WEST	-68680.5497
## neighborhoodCROWN HEIGHTS	-119504.7203
## neighborhoodCYPRESS HILLS_AND_GERRITSEN BACH	-618273.6901
## neighborhoodDOWNTOWN-FULTON MALL	280076.9493
## neighborhoodDOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE	1543267.1779
## neighborhoodDYKER _HEIGHTS_AND_WYCKOFF_HEIGHTS	-271248.1420
## neighborhoodEAST NEW YORK	-588365.5292
## neighborhoodFLATBUSH-EAST	-544344.0130
## neighborhoodFLATBUSH-LEFFERTS GARDEN	-95583.4087
## neighborhoodFLATLANDS_AND_SPRING_CREEK	-636036.9631
## neighborhoodFORT GREENE	1103801.8960
## neighborhoodGERRITSEN BEACH	-525195.9887
## neighborhoodGOWANUS	-117776.8243
## neighborhoodGRAVESEND	-308351.2493
## neighborhoodGREENPOINT	-7506.8833
## neighborhoodKENSINGTON_AND_FLATBUSH-CENTRAL	-16085.9221
## neighborhoodMADISON	-256291.3514
## neighborhoodMANHATTAN BEACH	84480.5467
## neighborhoodMARINE PARK	-408312.1446
## neighborhoodMIDWOOD	-238106.9177
## neighborhoodMILL BASIN	-581670.9199
## neighborhoodOCEAN HILL	-1041881.5269
## neighborhoodOCEAN PARKWAY-NORTH	179480.6473
## neighborhoodOCEAN PARKWAY-SOUTH	1173078.7559
## neighborhoodOLD MILL BASIN_AND_CONEY ISLAND	-484670.1296
## neighborhoodPARK SLOPE SOUTH	429484.0053
## neighborhoodPROSPECT HEIGHTS	677655.3725
## neighborhoodRED HOOK	-100928.8489
## neighborhoodSEAGATE	-701637.3640
## neighborhoodSHEEPSHEAD_BAY_AND_BERGEN_BEACH	-454379.0305
## neighborhoodWILLIAMSBURG-CENTRAL_AND_FLATBUSH-NORTH	-539716.9395
## neighborhoodWILLIAMSBURG-EAST	50051.7968
## neighborhoodWILLIAMSBURG-NORTH	1698643.0389
## neighborhoodWILLIAMSBURG-SOUTH_AND_WINDSOR_TERRACE	431616.8257
##	97.5 %
## (Intercept)	423257.7911
## quarter4	138067.7198
## grosssqft	507.8058
## neighborhoodBAY RIDGE	500726.8880

## neighborhoodBEDFORD STUYVESANT	259268.6667
## neighborhoodBENSONHURST	488186.8401
## neighborhoodBOERUM HILL	1765985.2579
## neighborhoodBROOKLYN HEIGHTS	2255431.4178
## neighborhoodBROWNSVILLE	1040.4515
## neighborhoodBUSHWICK_AND_BOROUGH PARK_AND_SUNSET PARK	554879.3755
## neighborhoodCANARSIE	-6773.9076
## neighborhoodCARROLL GARDENS	3539374.1401
## neighborhoodCLINTON HILL	2469809.6363
## neighborhoodCOBBLE HILL	2669680.6040
## neighborhoodCOBBLE HILL-WEST	1748106.7914
## neighborhoodCROWN HEIGHTS	628289.7548
## neighborhoodCYPRESS HILLS_AND_GERRITSEN BACH	93648.0640
## neighborhoodDOWNTOWN-FULTON MALL	1631826.2566
## neighborhoodDOWNTOWN-FULTON_FERRY_AND_PARK_SLOPE	2340663.9808
## neighborhoodDYKER _HEIGHTS_AND_WYCKOFF_HEIGHTS	333937.9145
## neighborhoodEAST NEW YORK	21629.4465
## neighborhoodFLATBUSH-EAST	18292.8180
## neighborhoodFLATBUSH-LEFFERTS GARDEN	791399.0189
## neighborhoodFLATLANDS_AND_SPRING_CREEK	-25748.5781
## neighborhoodFORT GREENE	2930961.0343
## neighborhoodGERRITSEN BEACH	127139.2136
## neighborhoodGOWANUS	1697202.9997
## neighborhoodGRAVESEND	310316.1595
## neighborhoodGREENPOINT	1007413.3900
## neighborhoodKENSINGTON_AND_FLATBUSH-CENTRAL	596756.2561
## neighborhoodMADISON	329358.8914
## neighborhoodMANHATTAN BEACH	814157.0391
## neighborhoodMARINE PARK	153579.7506
## neighborhoodMIDWOOD	359184.9266
## neighborhoodMILL BASIN	96195.4456
## neighborhoodOCEAN HILL	294375.2411
## neighborhoodOCEAN PARKWAY-NORTH	779561.0851
## neighborhoodOCEAN PARKWAY-SOUTH	1920722.7367
## neighborhoodOLD MILL BASIN_AND_CONEY ISLAND	97205.8976
## neighborhoodPARK SLOPE SOUTH	1270110.6942
## neighborhoodPROSPECT HEIGHTS	2017367.1843
## neighborhoodRED HOOK	1714168.9192
## neighborhoodSEAGATE	108292.4822
## neighborhoodSHEEPSHEAD_BAY_AND_BERGEN_BEACH	125919.2351
## neighborhoodWILLIAMSBURG-CENTRAL_AND_FLATBUSH-NORTH	267578.1601
## neighborhoodWILLIAMSBURG-EAST	1864944.4834
## neighborhoodWILLIAMSBURG-NORTH	2585235.0695
## neighborhoodWILLIAMSBURG-SOUTH_AND_WINDSOR_TERRACE	1192362.7827

Neighborhood Analysis:

For the predictor variable neighborhood, I determined the neighborhoods that have the most significant relation with price, as those with the smallest p value which were < 2e-16. Overall, 27 neighborhoods had a significant relation with price, however, I decided to stick with the 5 most significant for simplicity: Brooklyn Heights, Downtown Fulton Ferry, Park Slope, South Ocean Parkway, and North Williamsburg.

For the 5 most significant neighborhoods, I plotted the box plot “Brooklyn Home Prices for Neighborhoods with Significant Price Relations from 2020 Q3 to 2020 Q4,” below. As seen in the graph, from Q3 to Q4, home prices in Brooklyn Heights and South Ocean Parkway dropped. However, home prices in Downtown Fulton Ferry, Park Slope, and North Williamsburg

increased. For Brooklyn Height, there was an approximate \$3.1M price drop while, in North Williamsburg, there was an approximate \$800k increase. This shows an effect of quarter and neighborhood on price, meaning neighborhood is important to consider for the price change between 2020 Q3 to 2020 Q4.

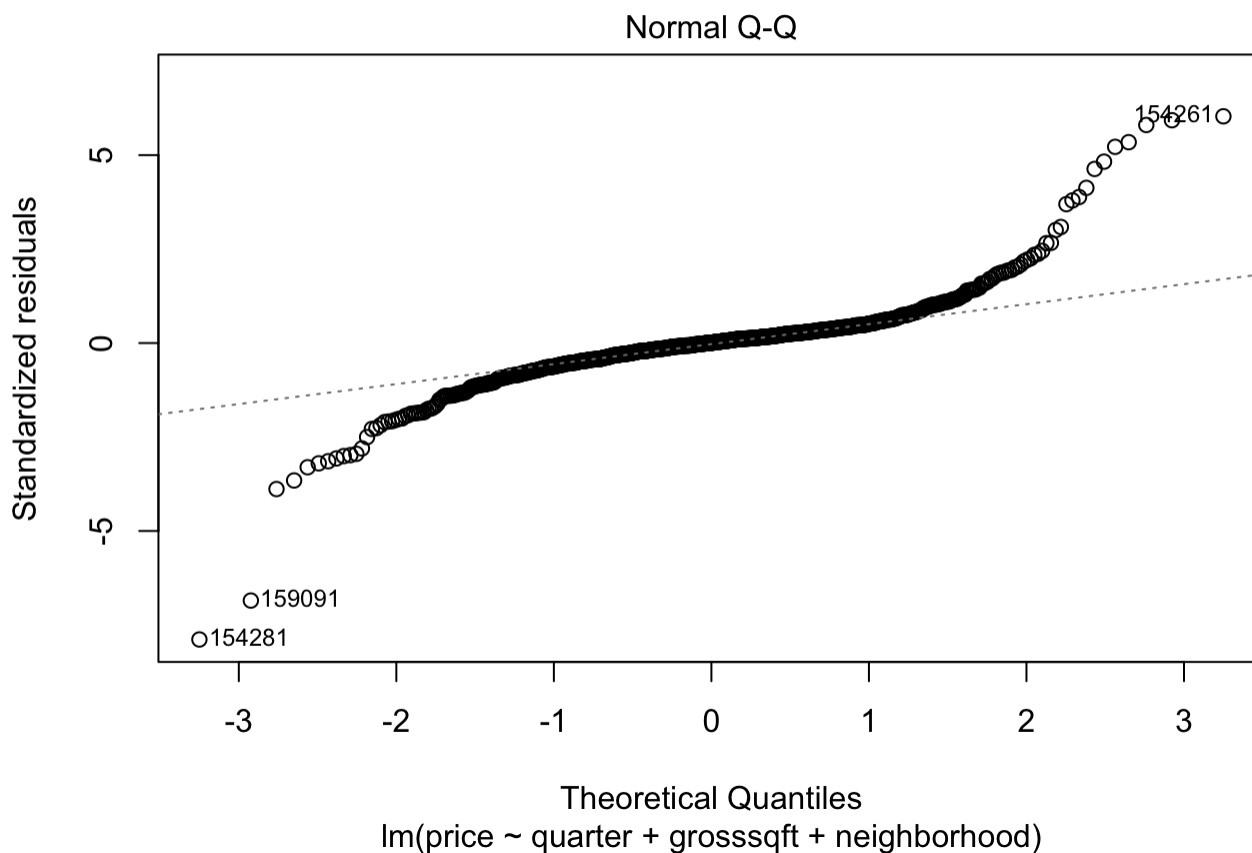
1.5 Check Assumptions for Linear Regression

```
library(car)
durbinWatsonTest(lmnew)
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 0.1264452 1.74299 0
## Alternative hypothesis: rho != 0
```

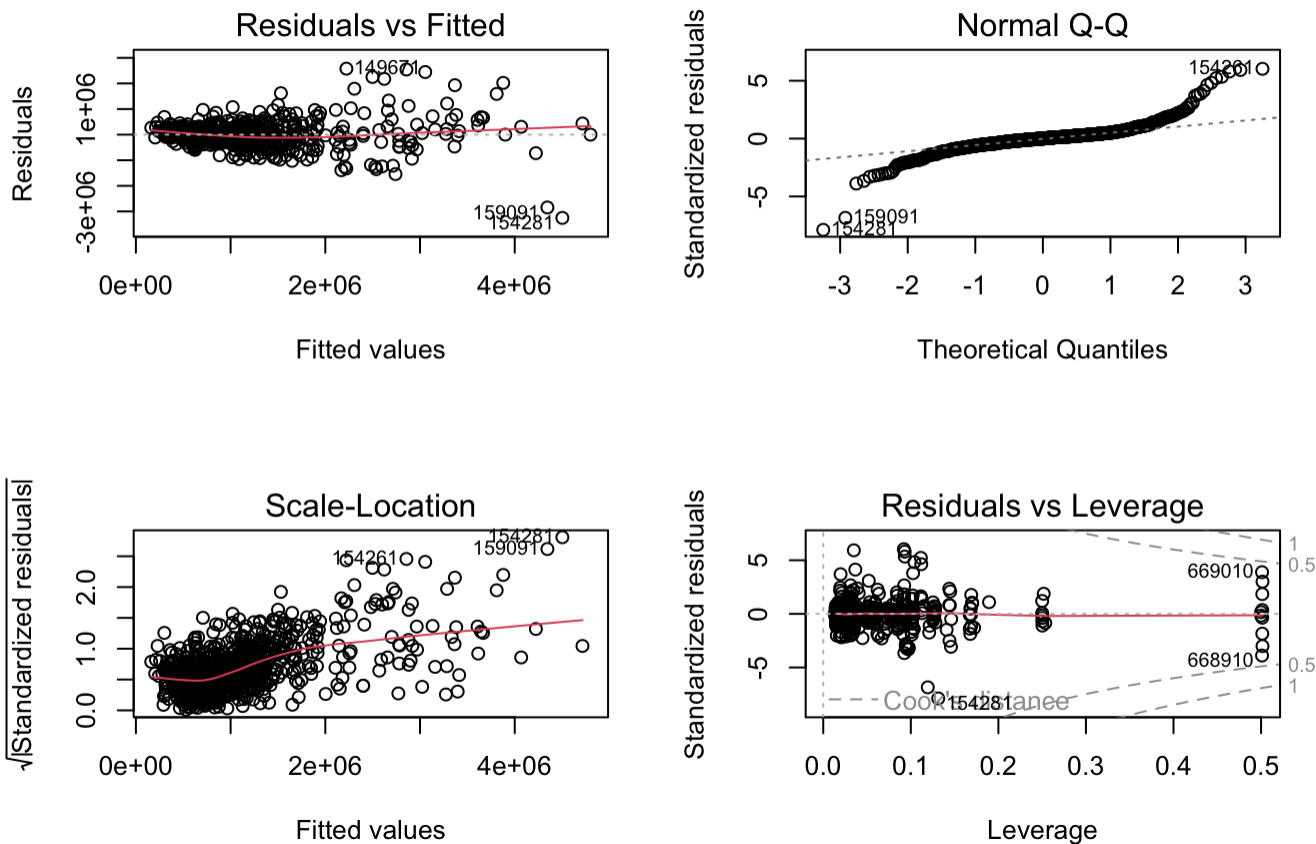
```
library(ggplot2)
#plot(lmnew, which = 1)
#not linear -->
#plot(lmnew$fitted.values, lmnew$model$BMI)
#hist(lmnew$residuals)
plot(lmnew, which=2)
```

```
## Warning: not plotting observations with leverage one:
## 209, 214, 447, 468, 793, 854
```



```
par(mfrow=c(2,2))
plot(lmnew)
```

```
## Warning: not plotting observations with leverage one:
##      209, 214, 447, 468, 793, 854
```



Assumption Analysis:

My price predictor model, as stated, uses linear regression to explain Brooklyn housing prices, so I checked whether the assumptions made by the linear regression model are met or not. First, I checked the linearity of the data with a Residual vs Fitted Plot, below. There is a clear pattern in the points, so the assumption that there is a linear relationship between price and the predictor variables is not met. Second, I tested the homoscedasticity with a scale-location plot, below. In the scale-location plot, the points are not equally spread, so there is a heteroscedasticity problem and we do not meet the homoscedasticity assumption.

Third, I checked the normality of the residuals with a QQ plot of the residuals, above. In the QQ plot, the majority of the points fall approximately along the reference line, so the normality assumption is met. However, there are deviating endpoints, which suggest a skewed distribution. Fourth, I checked that the predictors are independent using the Durbin-Watson test. I found a p-value of 0, so we conclude that the residuals in this regression model are autocorrelated and fail to meet our independence assumption. Fifth, I tested the independence of errors with a Residual vs Fitted Plot, above. In the plot, there is a slight correlation, so the independence of errors is not met. Overall, the majority of the Ordinary Least Squares (OLS) assumptions are not met, meaning that remodeling is necessary. A possible remodeling step is using a logistic transformation on price.

1.6 Conclusion for Q3 to Q4 Shift

With my price predictor model, I found that Brooklyn home purchase prices significantly changed between Q3 2020 and Q4 2020 with an additive increase of \$80,877.71 from Q3 2020 and Q4 2020. This was supported by the increase in price for all square footage from Q3 to Q4. However, we cannot apply this insight to all of Brooklyn because while overall Brooklyn home purchase prices increased, in certain neighborhoods there was a price decrease. Neighborhood should be accounted for to determine the price change between Q3 2020 and Q4 2020. Furthermore, the model did not meet all the OLS assumptions, so the results found should not be fully trusted and used for business decisions. The next steps include remodeling and considering a non-linear regression model to determine if Brooklyn home purchase prices significantly changed between Q3 2020 and Q4 2020.