Brooklyn Housing - Modeling Codebase

Kshitij Mittal

2022-12-18

Calling Relevant Packages

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(stringr)
library(readr)
library(relaimpo)
```

```
## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select

## Loading required package: boot

## Loading required package: survey

## Loading required package: grid

## Loading required package: Matrix

## Loading required package: survival
```

```
##
## Attaching package: 'survival'

## The following object is masked from 'package:boot':
##
##     aml

##
## Attaching package: 'survey'

## The following object is masked from 'package:graphics':
##
##     dotchart

## Loading required package: mitools

## This is the global version of package relaimpo.

## If you are a non-US user, a version with the interesting additional metric pmvd is available

## from Ulrike Groempings web site at prof.beuth-hochschule.de/groemping.
```

```
library(ggplot2)
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

1.1 Bring the data into R

```
dat_2016 = read.csv('/Users/kshitijmittal/Documents/UChicago Acad/01 Quarter 1/01 Stat Analysis/99 Final
dat_2017 = read.csv('/Users/kshitijmittal/Documents/UChicago Acad/01 Quarter 1/01 Stat Analysis/99 Final
dat_2018 = read.csv('/Users/kshitijmittal/Documents/UChicago Acad/01 Quarter 1/01 Stat Analysis/99 Final
dat_2019 = read.csv('/Users/kshitijmittal/Documents/UChicago Acad/01 Quarter 1/01 Stat Analysis/99 Final
dat_2020 = read.csv('/Users/kshitijmittal/Documents/UChicago Acad/01 Quarter 1/01 Stat Analysis/99 Final

colnames(dat_2016) = c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','l
colnames(dat_2017) = c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','l
colnames(dat_2018) = c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','l
colnames(dat_2019) = c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','l
colnames(dat_2020) = c('borough','neighborhood','bldclasscat','taxclasscurr','block','lot','easement','l
```

1.2 Join the data and make it usable for analysis

```
dat_2016 = dat_2016[5:dim(dat_2016)[1],]  #Removing top 4 blank/non-data rows
dat_2017 = dat_2017[5:dim(dat_2017)[1],]  #Removing top 4 blank/non-data rows
dat_2018 = dat_2018[5:dim(dat_2018)[1],]  #Removing top 4 blank/non-data rows
dat_2019 = dat_2019[5:dim(dat_2019)[1],]  #Removing top 4 blank/non-data rows
dat_2020 = dat_2020[8:dim(dat_2020)[1],]  #Removing top 7 blank/non-data rows

# Removing all blank rows from the datasets
dat_2016 = dat_2016[!apply(dat_2016 == "", 1, all),]
dat_2017 = dat_2017[!apply(dat_2017 == "", 1, all),]
dat_2018 = dat_2018[!apply(dat_2018 == "", 1, all),]
dat_2019 = dat_2019[!apply(dat_2019 == "", 1, all),]
dat_2020 = dat_2020[!apply(dat_2020 == "", 1, all),]

# Creating different vectors for columns as per data types
cols.num=c("borough","taxclasscurr","block","lot","zip","resunits","comunits","totunits","yrbuilt","tax
cols.num2=c("landsqft","grosssqft","price")
cols.str=c("neighborhood","bldclasscat","easement","bldclasscurr","address","bldclasssale")

#Starting off with cleaning the 2016 data
dat_2016[cols.num]=sapply(dat_2016[cols.num],as.numeric) #Converting char data type to numeric
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
dat_2016[cols.num2]=sapply(dat_2016[cols.num2],parse_number) #Extracting numbers out of character data
```

```
## Warning: 8467 parsing failures.
## row col expected actual
## 182  -- a number      -
## 215  -- a number      -
## 216  -- a number      -
## 217  -- a number      -
## 218  -- a number      -
## ... ... ........ ......
## See problems(...) for more details.

## Warning: 9644 parsing failures.
## row col expected actual
## 182  -- a number      -
## 215  -- a number      -
## 216  -- a number      -
## 217  -- a number      -
```

```
## 218  -- a number      -
## ... ... ........ ......
## See problems(...) for more details.
```

```
## Warning: 9881 parsing failures.
## row col expected actual
##   1  -- a number      -
##   3  -- a number      -
##   7  -- a number      -
##  10  -- a number      -
##  13  -- a number      -
## ... ... ........ ......
## See problems(...) for more details.
```

```
dat_2016[cols.str]=sapply(dat_2016[cols.str],str_trim) # Removing leading and trailing whitespaces from
dat_2016[cols.str]=sapply(dat_2016[cols.str],str_squish) # Removing middle whitespaces from characters
dat_2016$date=as.Date(dat_2016$date,"%m/%d/%Y") #Parsing date in an R friendly format

#Converting any blank values to 0 for this particular data set
dat_2016$price[is.na(dat_2016$price)]=0
dat_2016$comunits[is.na(dat_2016$comunits)]=0
dat_2016$resunits[is.na(dat_2016$resunits)]=0
dat_2016$totunits[is.na(dat_2016$totunits)]=0

#Repeating the above 5 types for 2017 data
dat_2017[cols.num]=sapply(dat_2017[cols.num],as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
dat_2017[cols.num2]=sapply(dat_2017[cols.num2],parse_number)
```

```
## Warning: 8221 parsing failures.
## row col expected actual
## 187  -- a number      -
## 188  -- a number      -
## 189  -- a number      -
## 190  -- a number      -
## 191  -- a number      -
## ... ... ........ ......
## See problems(...) for more details.
```

```
## Warning: 8888 parsing failures.
## row col expected actual
##  78  -- a number      -
## 172  -- a number      -
## 173  -- a number      -
## 187  -- a number      -
## 188  -- a number      -
## ... ... ........ ......
## See problems(...) for more details.
```

```
dat_2017[cols.str]=sapply(dat_2017[cols.str],str_trim)
dat_2017[cols.str]=sapply(dat_2017[cols.str],str_squish)
dat_2017$date=as.Date(dat_2017$date,"%m/%d/%y")

#Repeating the above 5 types for 2018 data
dat_2018[cols.num]=sapply(dat_2018[cols.num],as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
dat_2018[cols.num2]=sapply(dat_2018[cols.num2],parse_number)
dat_2018[cols.str]=sapply(dat_2018[cols.str],str_trim)
dat_2018[cols.str]=sapply(dat_2018[cols.str],str_squish)
dat_2018$date=as.Date(dat_2018$date,"%m/%d/%y")

#Repeating the above 5 types for 2019 data
dat_2019[cols.num]=sapply(dat_2019[cols.num],as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
dat_2019[cols.num2]=sapply(dat_2019[cols.num2],parse_number)
dat_2019[cols.str]=sapply(dat_2019[cols.str],str_trim)
dat_2019[cols.str]=sapply(dat_2019[cols.str],str_squish)
dat_2019$date=as.Date(dat_2019$date,"%m/%d/%y")

#Repeating the above 5 types for 2019 data
dat_2020[cols.num]=sapply(dat_2020[cols.num],as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion

## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
dat_2020[cols.num2]=sapply(dat_2020[cols.num2],parse_number)
dat_2020[cols.str]=sapply(dat_2020[cols.str],str_trim)
dat_2020[cols.str]=sapply(dat_2020[cols.str],str_squish)
dat_2020$date=as.Date(dat_2020$date,"%m/%d/%y")
```

```
data_comb = rbind(dat_2016, dat_2017, dat_2018, dat_2019, dat_2020)
head(data_comb)
```

```
##    borough neighborhood             bldclasscat taxclasscurr block lot easement
## 5        3   BATH BEACH 01 ONE FAMILY DWELLINGS            1  6360  13
## 6        3   BATH BEACH 01 ONE FAMILY DWELLINGS            1  6361  11
```

```
## 7        3   BATH BEACH 01 ONE FAMILY DWELLINGS            1  6364   2
## 8        3   BATH BEACH 01 ONE FAMILY DWELLINGS            1  6364  72
## 9        3   BATH BEACH 01 ONE FAMILY DWELLINGS            1  6371  79
## 10       3   BATH BEACH 01 ONE FAMILY DWELLINGS            1  6374  67
##    bldclasscurr            address aptnum   zip resunits comunits totunits
## 5            A5   8665 15TH AVENUE     NA 11228        1        0        1
## 6            A5 71 BAY 10TH STREET     NA 11228        1        0        1
## 7            A5 1649 BENSON AVENUE     NA 11214        1        0        1
## 8            A5 68 BAY 14TH STREET     NA 11214        1        0        1
## 9            A5   8668 19TH AVENUE     NA 11214        1        0        1
## 10           S1   8642 20TH AVENUE     NA 11214        1        1        2
##    landsqft grosssqft yrbuilt taxclasssale bldclasssale   price       date
## 5      1547      2224    1930            1           A5       0 2016-05-25
## 6      2900      1660    1930            1           A5  829000 2016-04-05
## 7      1638       972    1930            1           A5       0 2016-10-06
## 8      1950       972    1950            1           A5  790000 2016-06-21
## 9      2223      2520    1930            1           A5  788000 2016-03-31
## 10     1740      3240    1925            1           S1 1090000 2016-10-24
```

```
tail(data_comb)
```

```
##        borough    neighborhood            bldclasscat taxclasscurr block  lot
## 199017       3 WYCKOFF HEIGHTS 41 TAX CLASS 4 - OTHER            4  3407   26
## 199024       3 WYCKOFF HEIGHTS       44 CONDO PARKING            4  3290 1418
## 199034       3 WYCKOFF HEIGHTS       44 CONDO PARKING            4  3290 1422
## 199044       3 WYCKOFF HEIGHTS       44 CONDO PARKING            4  3310 1010
## 199054       3 WYCKOFF HEIGHTS       44 CONDO PARKING            4  3328 1086
## 199064       3 WYCKOFF HEIGHTS       44 CONDO PARKING            4  3328 1123
##        easement bldclasscurr                 address aptnum   zip resunits
## 199017                   Z9     378 WEIRFIELD STREET     NA 11237        0
## 199024                   RP   364 HARMAN STREET, PS2     NA 11237        0
## 199034                   RP   364 HARMAN STREET, PS6     NA 11237        0
## 199044                   RP 330 BLEECKER STREET, PK2     NA 11237        0
## 199054                   RP        358 GROVE ST, P25     NA 11237        0
## 199064                   RP   358 GROVE STREET, CB10     NA 11237        0
##        comunits totunits landsqft grosssqft yrbuilt taxclasssale bldclasssale
## 199017        0        0    12208         0      NA            4           Z9
## 199024        1        1        0         0    2017            4           RP
## 199034        1        1        0         0    2017            4           RP
## 199044        1        1        0         0    2013            4           RP
## 199054        1        1        0         0      NA            4           RP
## 199064        1        1        0         0      NA            4           RP
##          price       date
## 199017 3965000 2020-10-30
## 199024  470000 2020-12-01
## 199034  900000 2020-11-19
## 199044  960000 2020-12-21
## 199054       1 2020-03-09
## 199064  485000 2020-02-20
```

```
dim(data_comb)
```

```
## [1] 117151     21
```

```
dim(dat_2016)[1] + dim(dat_2017)[1] + dim(dat_2018)[1] + dim(dat_2019)[1] + dim(dat_2020)[1]
```

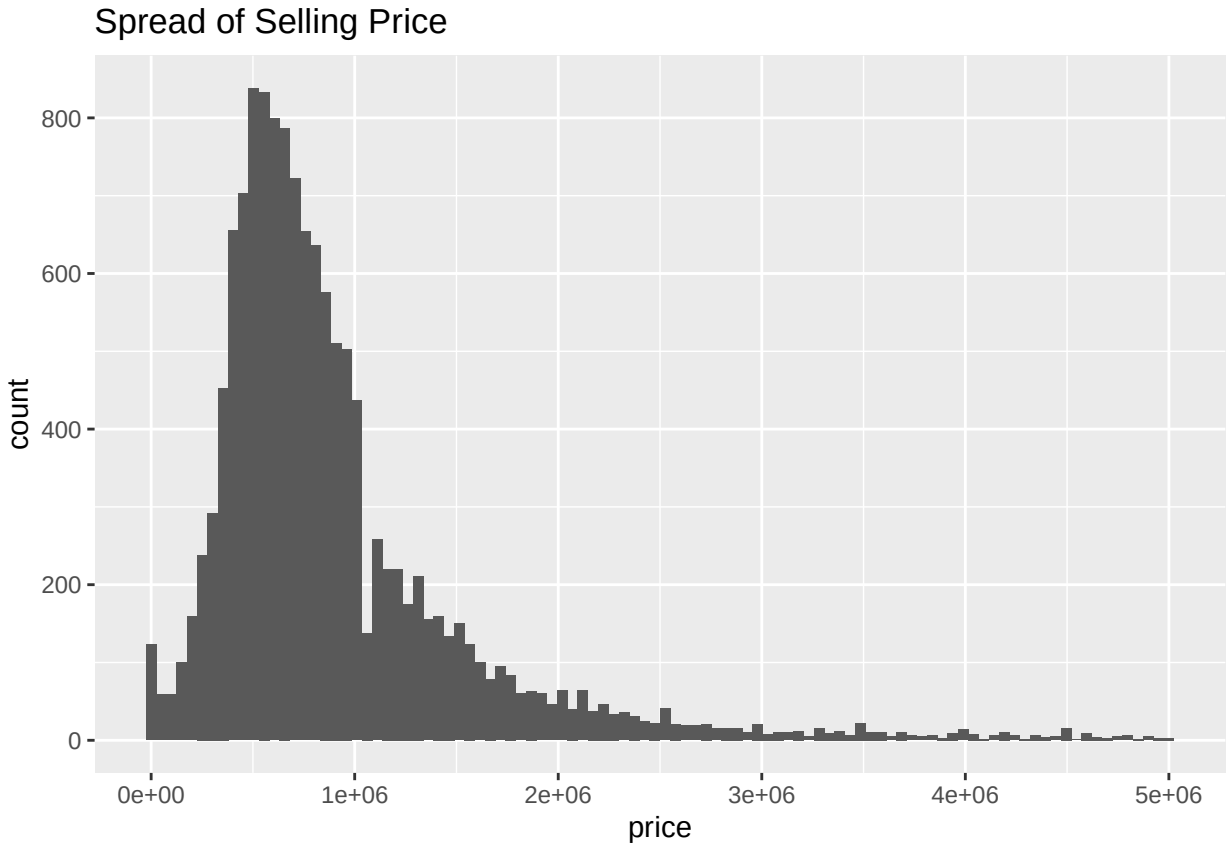
```
## [1] 117151
```

2.0 Exploratory Data analysis

```
# For the purposes of this analysis, we will only consider purchases
# of single-family residences and single-unit apartments or
# condos.

# Restrict the data to purchases where the building class at the time
# of sale starts with 'A' or 'R' and
index1 = with(data_comb, grepl("^A",bldclasssale))
index2 = with(data_comb, grepl("^R",bldclasssale))
data_comb1 = data_comb[index1|index2,]

# The number of total units and the number of residential units are both 1.
data_comb2 = data_comb1[data_comb1$totunits==1 & data_comb1$resunits==1,]
data_comb2 = data_comb2[!is.na(data_comb2$totunits),]
data_comb2 = data_comb2[!is.na(data_comb2$resunits),]

# Additionally restrict the data to observations where gross square footage is
# more than 0
data_comb3 = data_comb2[data_comb2$grosssqft>0,]
data_comb3 = data_comb3[!is.na(data_comb3$grosssqft),]

#Sale price is non-missing
data_comb4 = data_comb3[data_comb3$price>10,]
data_comb4 = data_comb4[!is.na(data_comb4$price),]
```

2.1 Feature Engineering

```
# Our goal will be to use linear regression to explain Brooklyn housing prices
# within the 2016-2020 window.  We will make predictions for the
# sale prices within this dataset, and extract most explanatory power out of our variables.

# Consider price as a potential response variable. Examining how it is
# distributed, and how it associates with the other variables in our data.

data_mod = data_comb4
ggplot(data_mod[data_mod$price<5000000,], aes(x=price)) + geom_histogram(bins = 100) + ggtitle("Spread
```

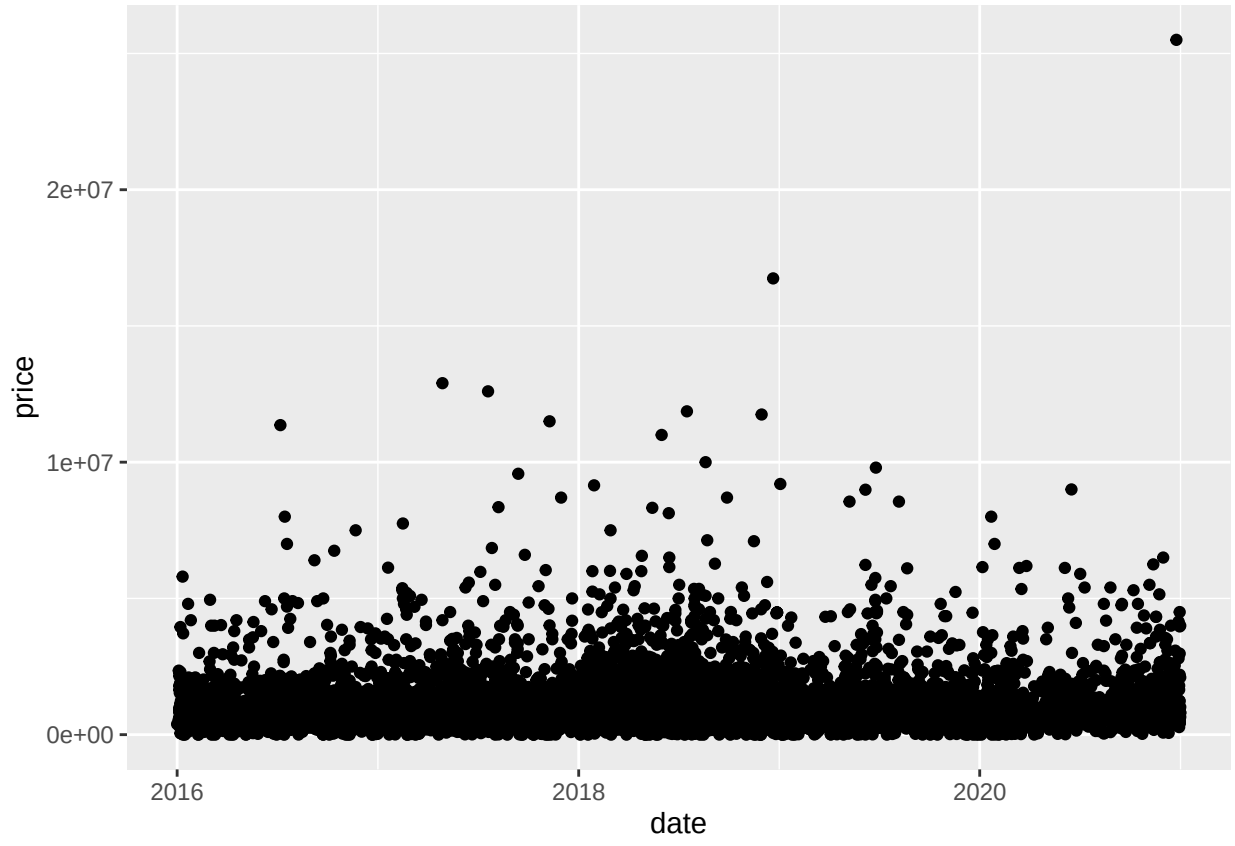



```
quantile(data_mod$price, probs = seq(.01, 1, by = .01))
```

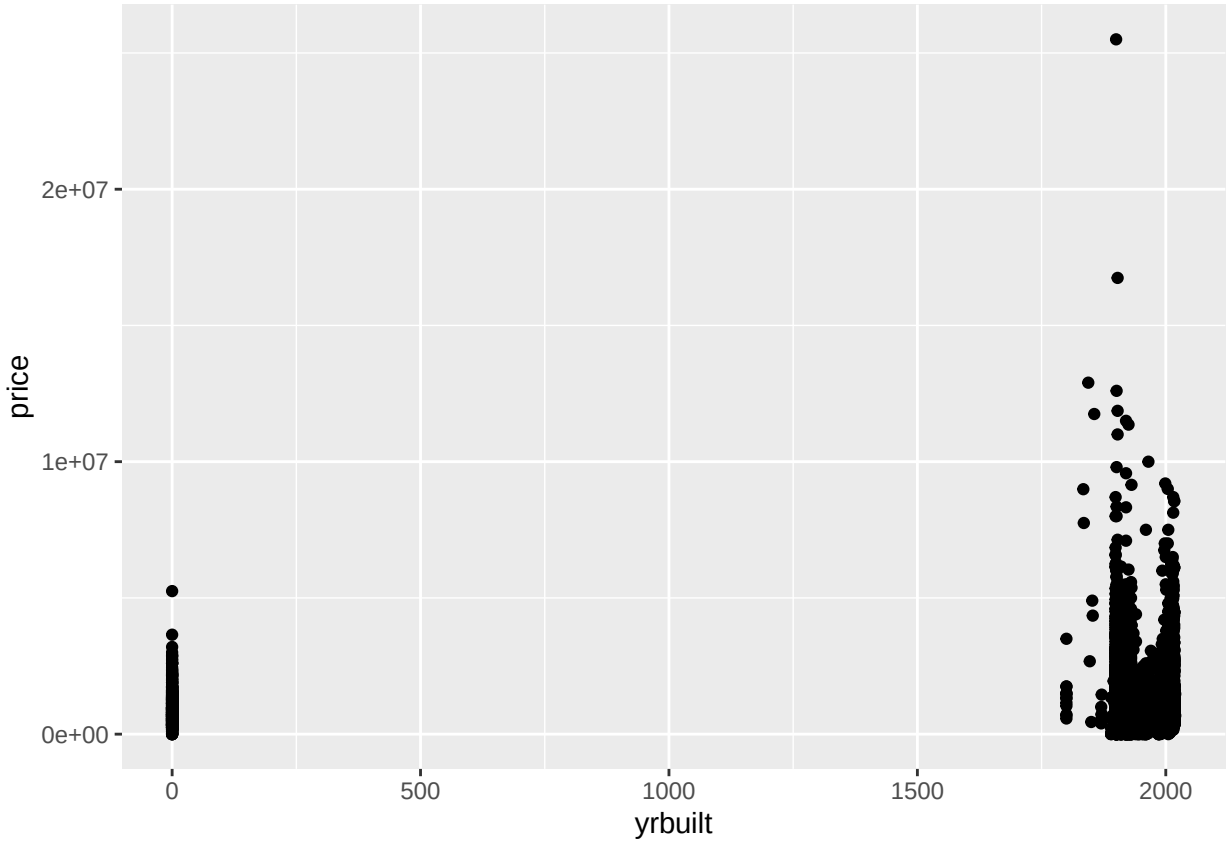
```
##         1%         2%         3%         4%         5%         6%         7%
##    31252.0   148612.5   200000.0   240000.0   265000.0   300000.0   319482.6
##         8%         9%        10%        11%        12%        13%        14%
##   335545.0   350000.0   365040.0   380000.0   393956.0   400548.6   415000.0
##        15%        16%        17%        18%        19%        20%        21%
##   421881.8   430857.1   441761.0   450000.0   460000.0   470000.0   480000.0
##        22%        23%        24%        25%        26%        27%        28%
##   491000.0   499925.5   505000.0   515000.0   525000.0   533010.0   540000.0
##        29%        30%        31%        32%        33%        34%        35%
##   550000.0   556676.9   565000.0   575000.0   580402.0   590585.0   600000.0
##        36%        37%        38%        39%        40%        41%        42%
##   606240.7   616041.0   625000.0   635000.0   644000.0   650000.0   660000.0
##        43%        44%        45%        46%        47%        48%        49%
##   670000.0   678000.0   686721.3   699000.0   700000.0   712775.0   725000.0
##        50%        51%        52%        53%        54%        55%        56%
##   735000.0   745000.0   750000.0   760000.0   775000.0   785650.0   799000.0
##        57%        58%        59%        60%        61%        62%        63%
##   804690.9   817000.0   830000.0   840000.0   850000.0   860000.0   875000.0
##        64%        65%        66%        67%        68%        69%        70%
##   890000.0   900000.0   915000.0   929000.0   945321.7   957735.0   973321.6
##        71%        72%        73%        74%        75%        76%        77%
##   985000.0   995000.0   999000.0  1020000.0  1075000.0  1100000.0  1135000.0
##        78%        79%        80%        81%        82%        83%        84%
```

```
##  1160089.9  1200000.0  1225000.0  1261534.4  1300000.0  1330000.0  1375000.0
##        85%        86%        87%        88%        89%        90%        91%
##  1418629.2  1465000.0  1505000.0  1567800.4  1625000.0  1700000.0  1788883.8
##        92%        93%        94%        95%        96%        97%        98%
##  1900000.0  2000000.0  2150000.0  2324398.1  2598199.2  2999999.1  3551369.7
##        99%       100%
##  4501332.5 25500000.0
```

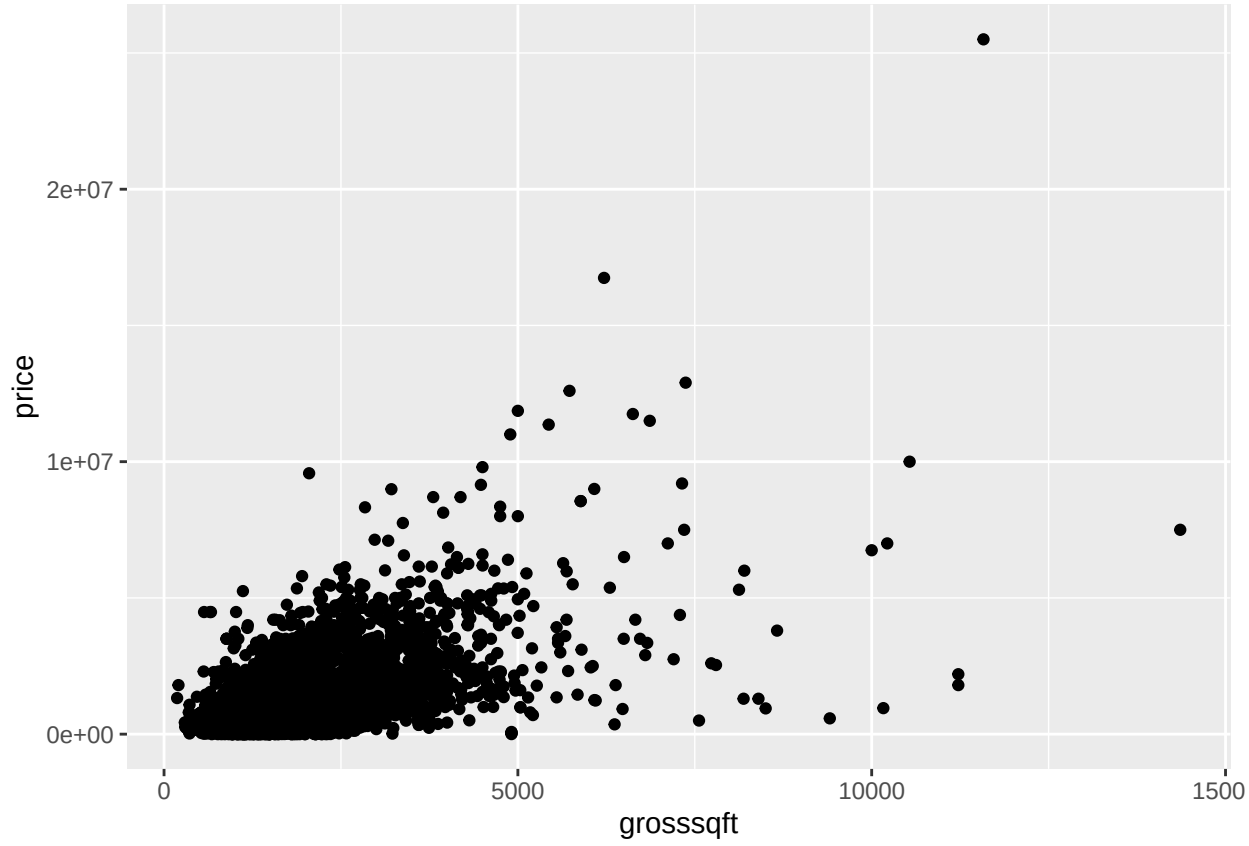
```
ggplot(data_mod, aes(date, price)) + geom_point()
```



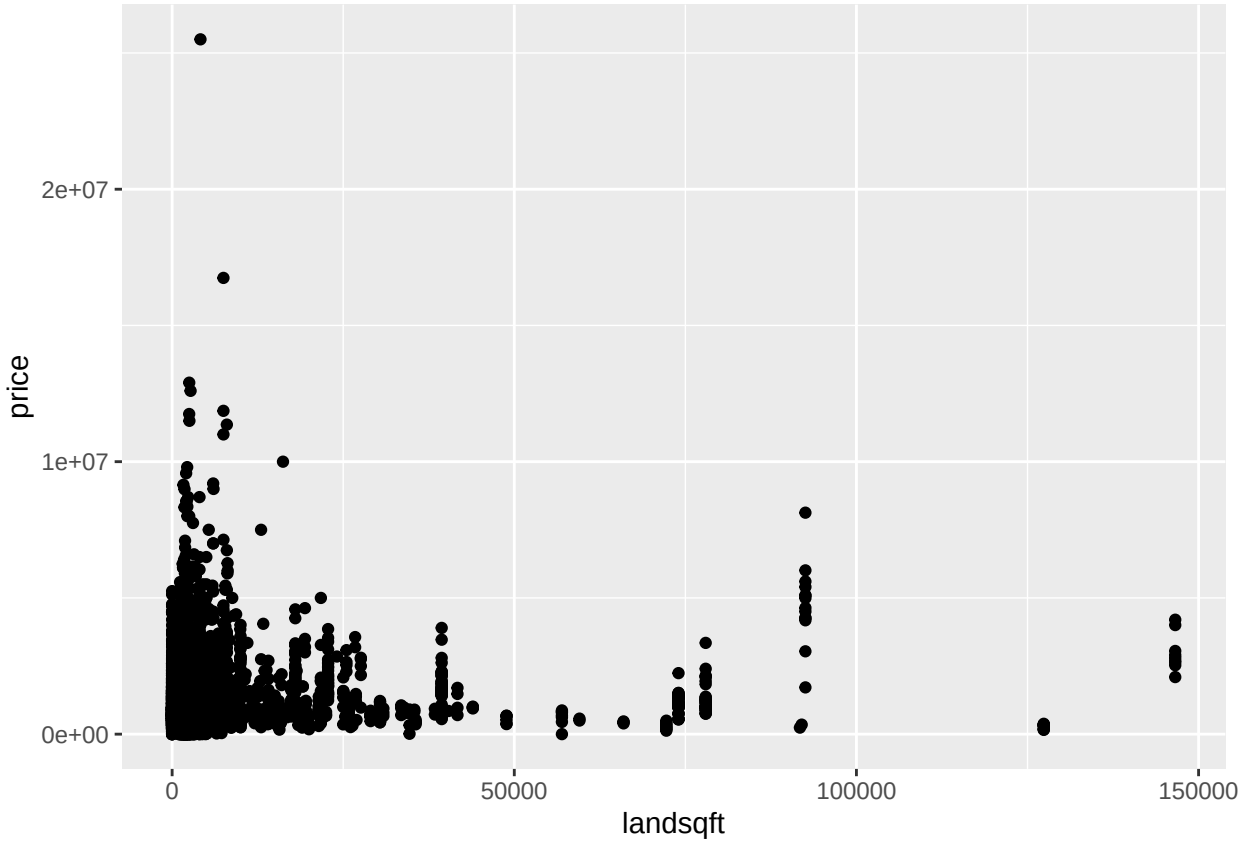

```
ggplot(data_mod, aes(yrbuilt, price)) + geom_point()
```



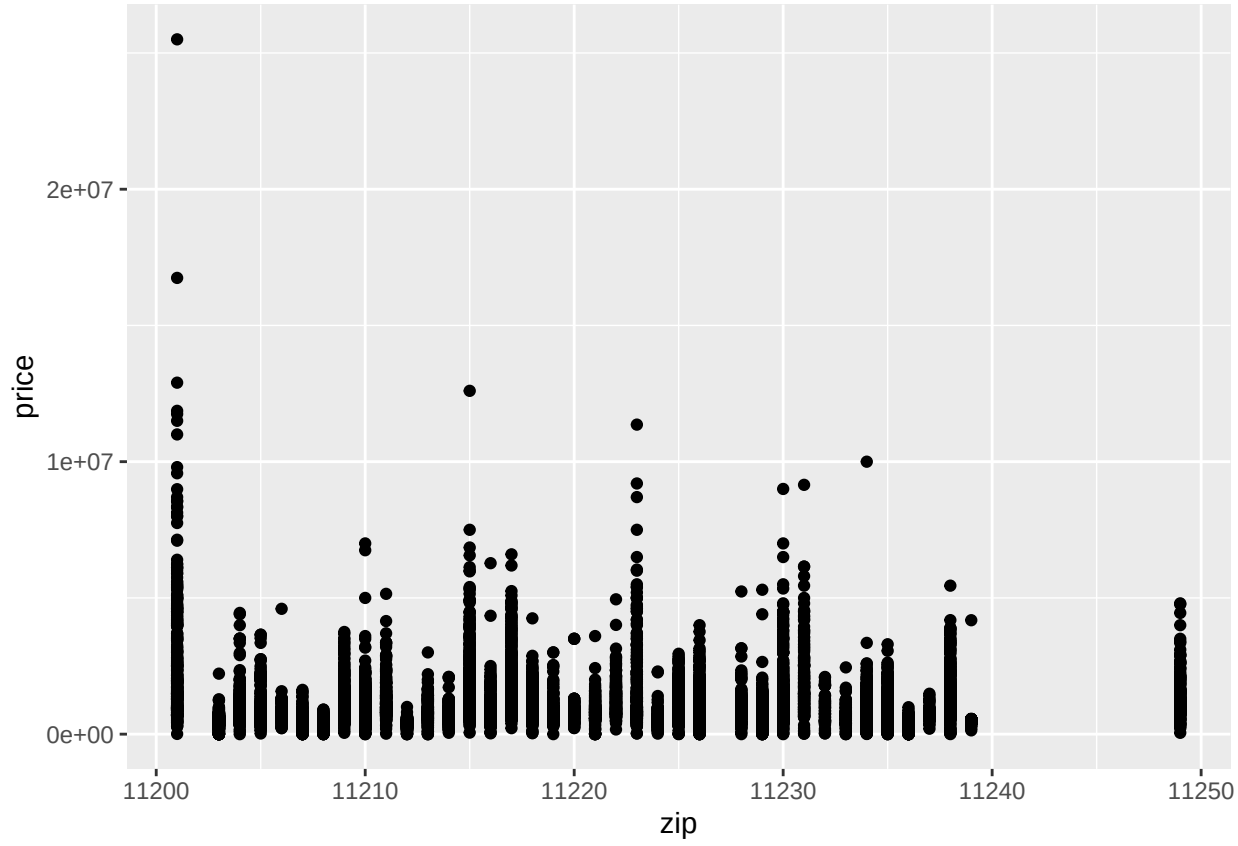

```
ggplot(data_mod, aes(grosssqft, price)) + geom_point()
```



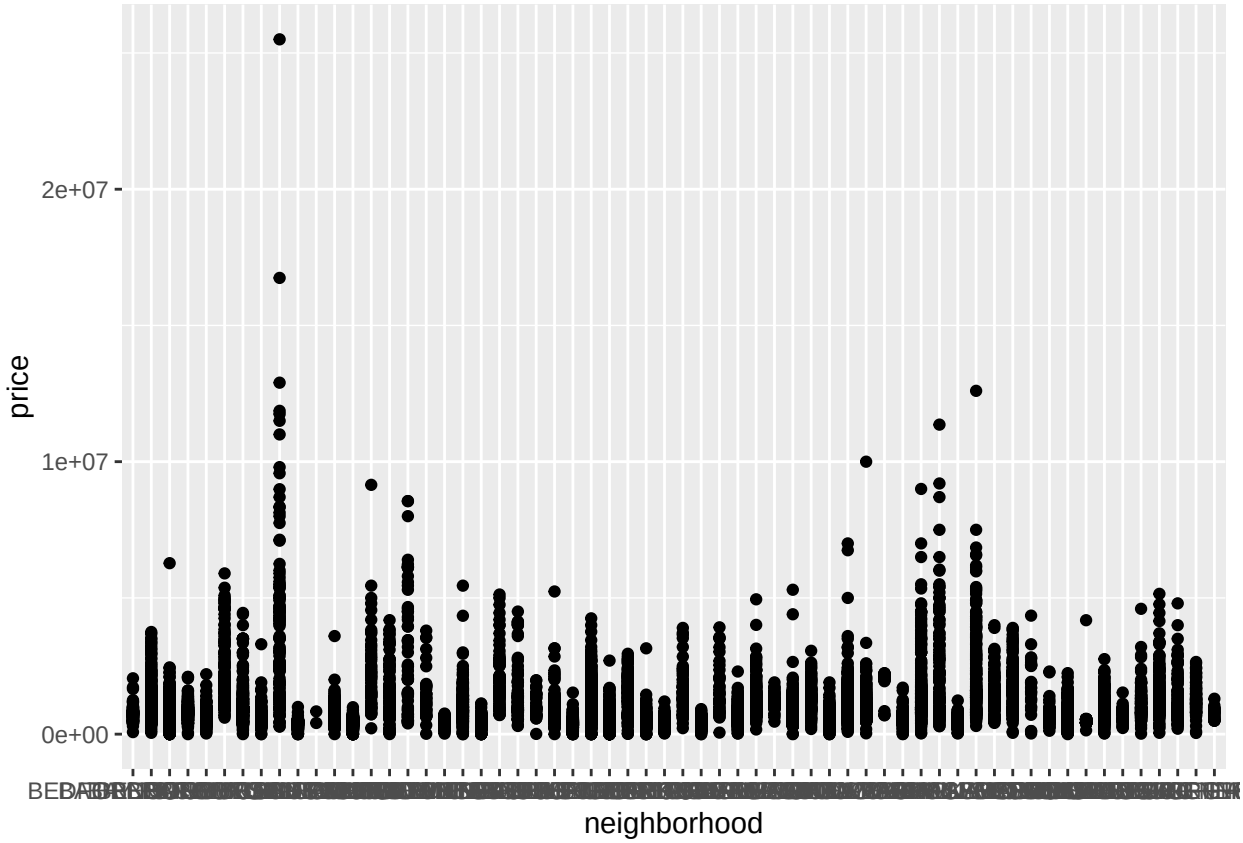

```
ggplot(data_mod, aes(landsqft, price)) + geom_point()
```



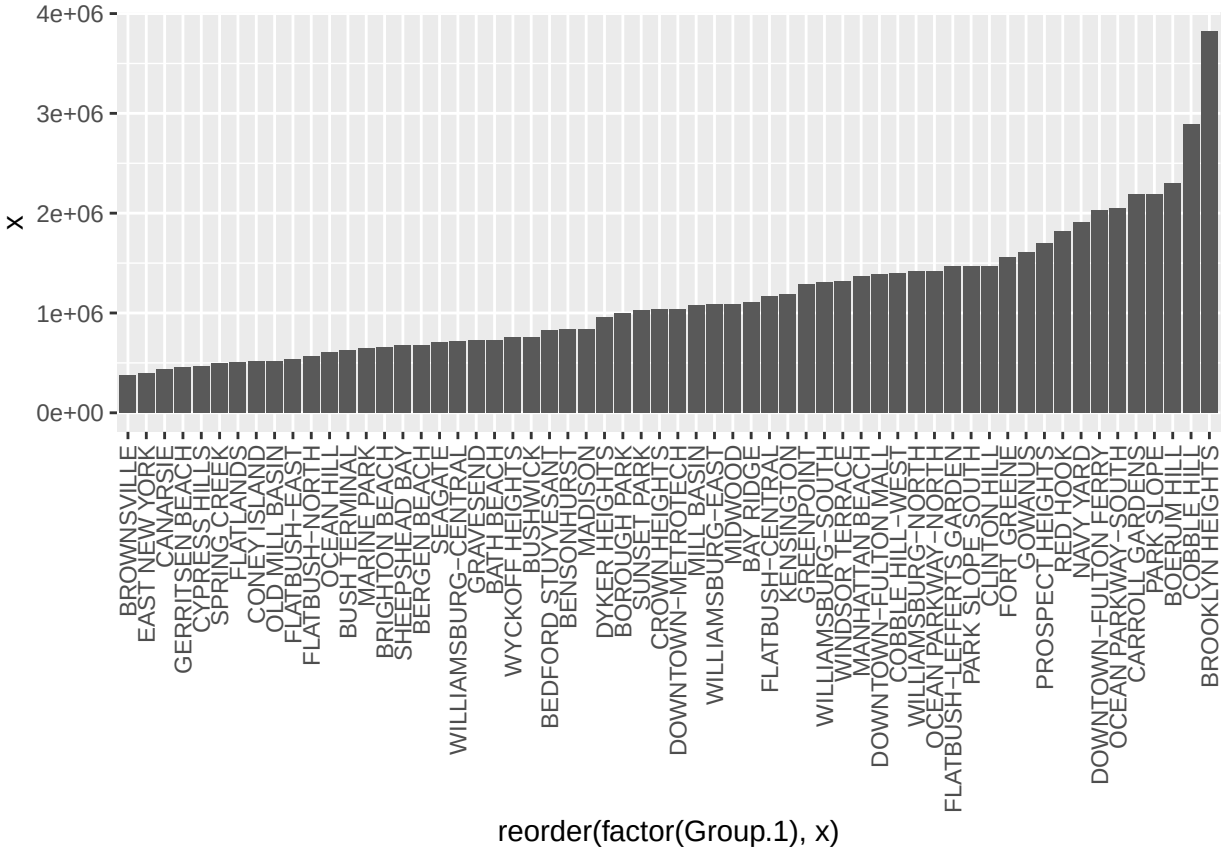

```
ggplot(data_mod, aes(zip, price)) + geom_point()
```



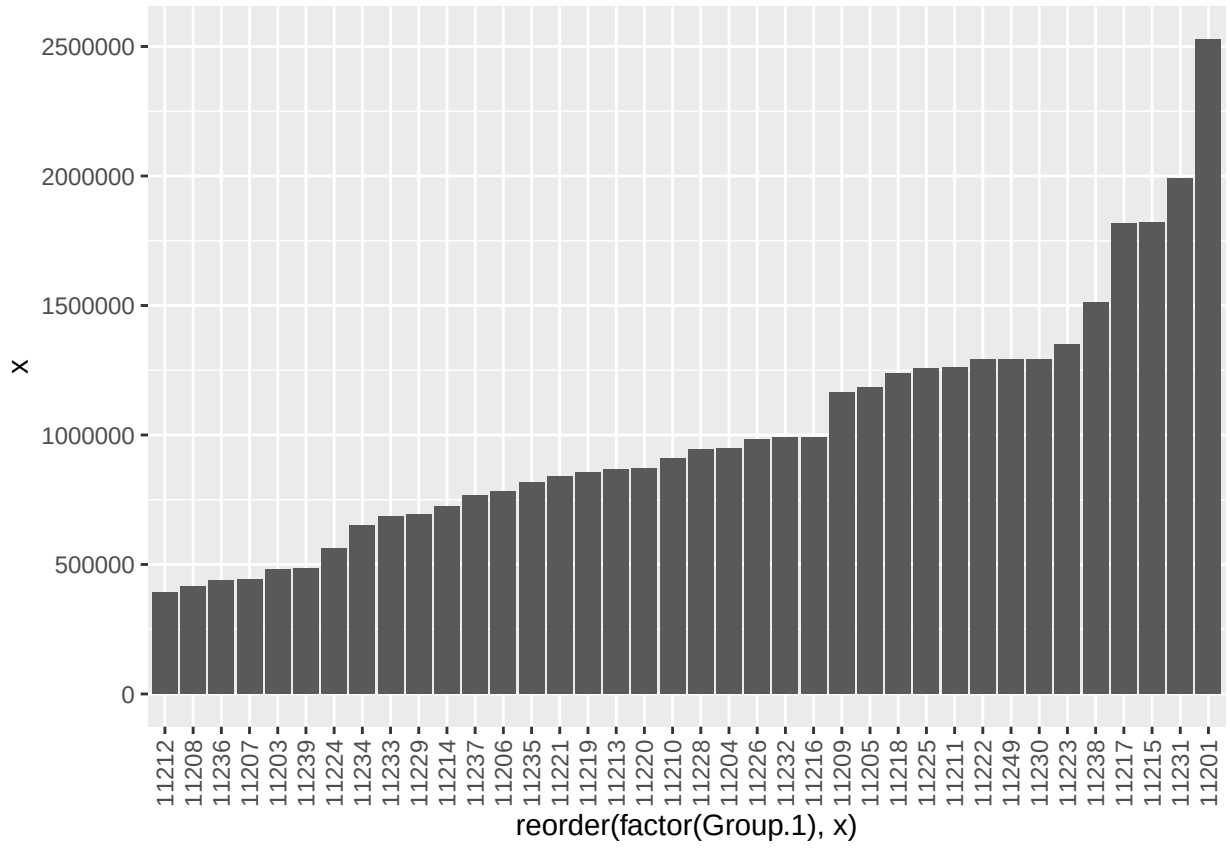

```
ggplot(data_mod, aes(neighborhood, price)) + geom_point()
```




```
#Analyzing prices by neighbourhoods
neigh_price=data.frame(aggregate(data_mod$price, list(data_mod$neighborhood), FUN=mean))
neigh_price = neigh_price %>% arrange(x)
ggplot(neigh_price, aes(x=reorder(factor(Group.1),x), y=x)) + geom_bar(stat = "identity") + theme(axis.
```

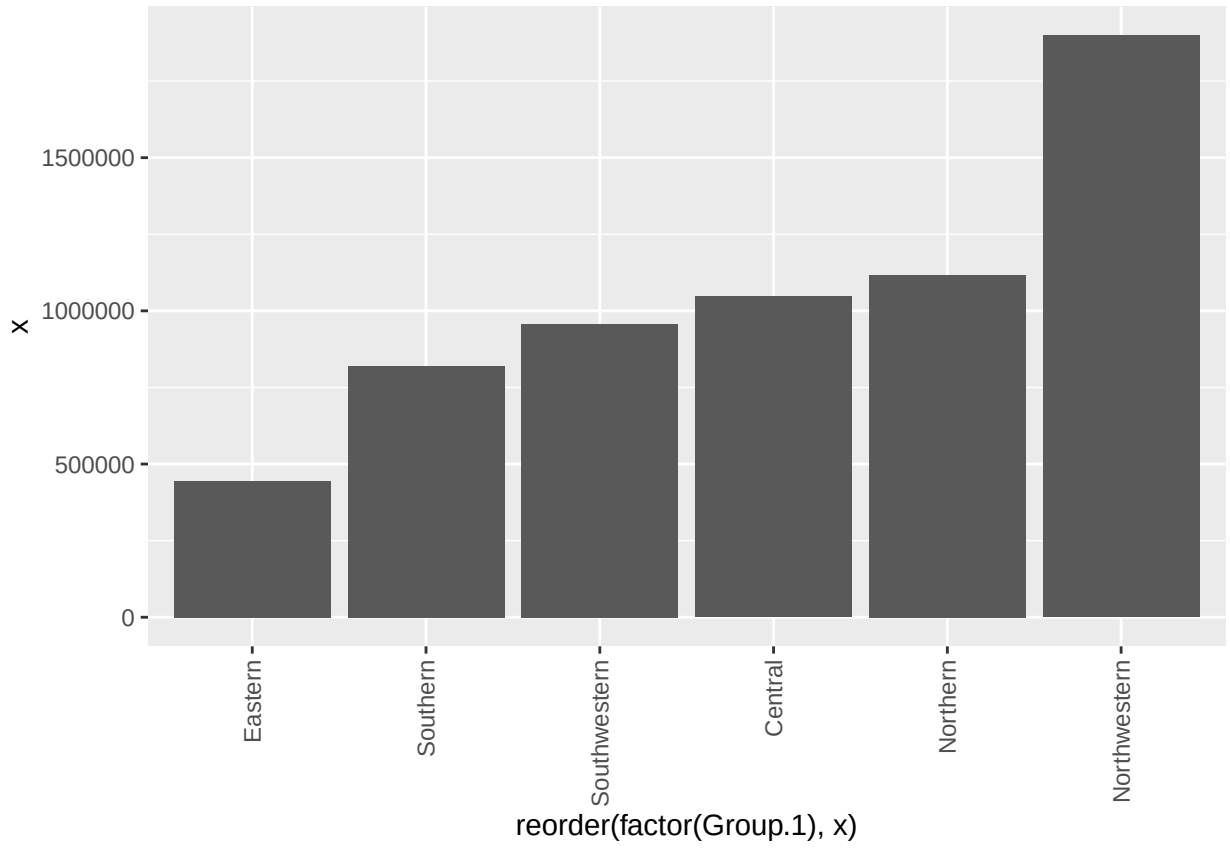



```
#Analyzing prices by zipcodes
zip_price=data.frame(aggregate(data_mod$price, list(data_mod$zip), FUN=mean))
zip_price = zip_price %>% arrange(x)
ggplot(zip_price, aes(x=reorder(factor(Group.1),x), y=x)) + geom_bar(stat = "identity") + theme(axis.te
```

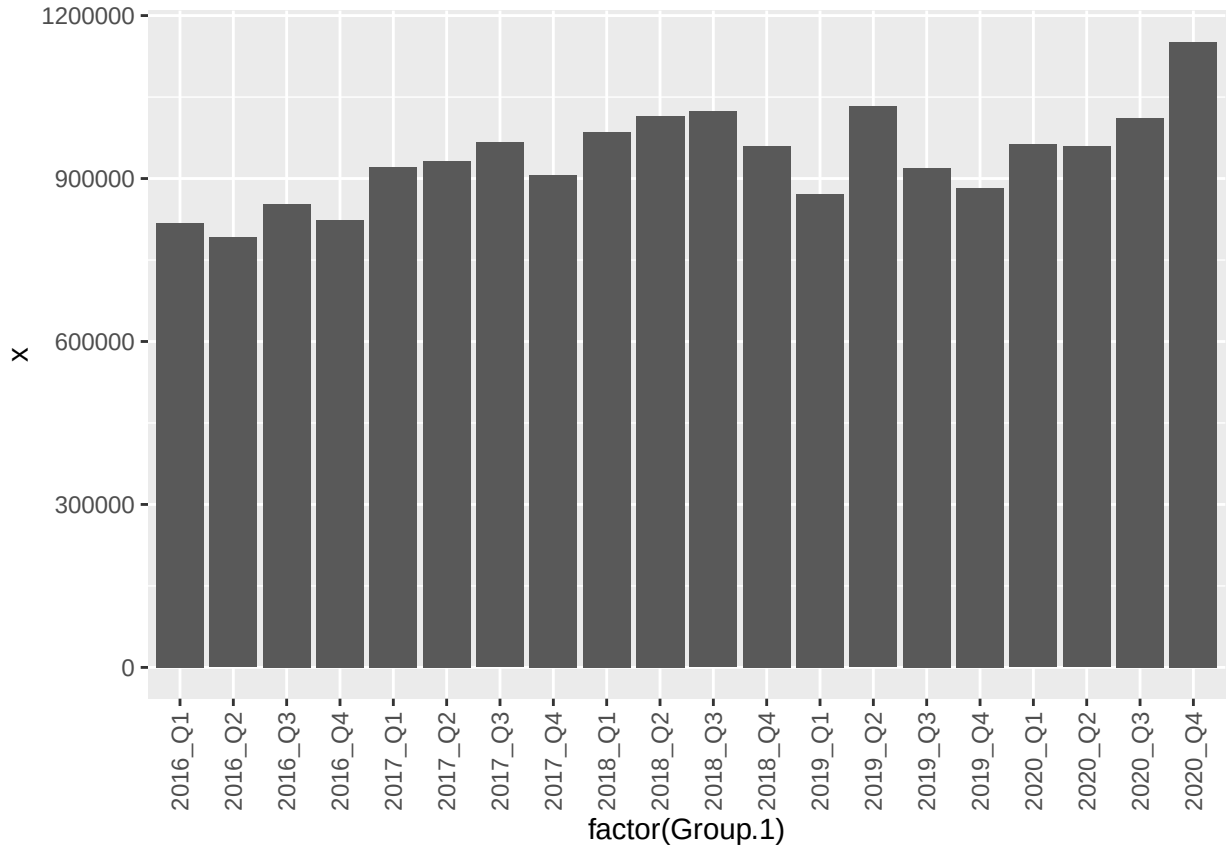


2.3 Pre-modeling and feature engineering

```
# Making a new categorical variable for zipcodes by location - obtained from
# New York City resources
Central = c(11213, 11216, 11225, 11233, 11226, 11218)
Eastern = c(11212, 11236, 11207, 11208, 11239, 11203)
Northern = c(11206, 11221, 11237, 11222, 11211, 11249)
Northwestern = c(11201, 11217, 11238, 11205, 11251, 11231, 11215)
Southern = c(11234, 11224, 11229, 11235, 11210, 11230, 11223)
Southwestern = c(11209, 11220, 11204, 11214, 11219, 11228, 11232)

data_mod$locality = as.factor(ifelse(data_mod$zip %in% Central, 'Central',
                                     ifelse(data_mod$zip %in% Eastern, 'Eastern',
                                            ifelse(data_mod$zip %in% Northern, 'Northern',
                                                   ifelse(data_mod$zip %in% Northwestern, 'Northwestern
                                                          ifelse(data_mod$zip %in% Southern, 'Southern'
                                                                 ifelse(data_mod$zip %in% Southwestern,

local_price=data.frame(aggregate(data_mod$price, list(data_mod$locality), FUN=mean))
ggplot(local_price, aes(x=reorder(factor(Group.1),x), y=x)) + geom_bar(stat = "identity") + theme(axis.
```




```
# Making a quarter variable for modelling

data_mod$quarter = as.factor(ifelse(data_mod$date >= "2016-01-01" & data_mod$date <= "2016-03-31","2016
                                    ifelse(data_mod$date >= "2016-04-01" & data_mod$date <= "2016-06-30
                                           ifelse(data_mod$date >= "2016-07-01" & data_mod$date <= "201
                                                  ifelse(data_mod$date >= "2016-10-01" & data_mod$date
                                                         ifelse(data_mod$date >= "2017-01-01" & data_mo
                                                                ifelse(data_mod$date >= "2017-04-01" &
                                                                       ifelse(data_mod$date >= "2017-07
                                                                              ifelse(data_mod$date >= "
                                                                                     ifelse(data_mod$da
                                                                                            ifelse(data_
                                                                                                   ifels

quarter_price=data.frame(aggregate(data_mod$price, list(data_mod$quarter), FUN=mean))
ggplot(quarter_price, aes(x=factor(Group.1), y=x)) + geom_bar(stat = "identity") + theme(axis.text.x =
```




```
# Removing static/non-predictive columns
delcols = c('easement','aptnum','address','borough','resunits','comunits','totunits')
data_mod2 = subset(data_mod, select = -c(easement,aptnum,address,borough,resunits,comunits,totunits))

# Factorizing categorical variables
factcols = c('neighborhood','bldclasscat','taxclasscurr','block','lot','bldclasscurr','zip','yrbuilt','
data_mod2[,factcols] = lapply(data_mod2[,factcols],factor)
str(data_mod2)
```

```
## 'data.frame':    13664 obs. of  16 variables:
##  $ neighborhood: Factor w/ 60 levels "BATH BEACH","BAY RIDGE",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ bldclasscat : Factor w/ 6 levels "01 ONE FAMILY DWELLINGS",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ taxclasscurr: Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ block       : Factor w/ 3421 levels "1","20","27",..: 2101 2102 2106 2113 2117 2117 2118 2121 212
##  $ lot         : Factor w/ 1107 levels "1","2","3","4",..: 11 72 79 5 18 19 31 36 36 15 ...
##  $ bldclasscurr: Factor w/ 15 levels "A0","A1","A2",..: 6 6 6 6 6 6 2 9 2 6 ...
##  $ zip         : Factor w/ 38 levels "11201","11203",..: 26 13 13 26 13 13 13 13 13 26 ...
##  $ landsqft    : num  2900 1950 2223 2469 2417 ...
##  $ grosssqft   : num  1660 972 2520 1836 1462 ...
##  $ yrbuilt     : Factor w/ 108 levels "0","1800","1834",..: 40 55 40 48 36 36 35 29 26 44 ...
##  $ taxclasssale: Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ bldclasssale: Factor w/ 15 levels "A0","A1","A2",..: 6 6 6 6 6 6 2 9 2 6 ...
##  $ price       : num  829000 790000 788000 920000 839000 854000 750000 699000 950000 699000 ...
##  $ date        : Date, format: "2016-04-05" "2016-06-21" ...
##  $ locality    : Factor w/ 6 levels "Central","Eastern",..: 6 6 6 6 6 6 6 6 6 6 ...
##  $ quarter     : Factor w/ 20 levels "2016_Q1","2016_Q2",..: 2 2 1 1 3 1 3 1 3 2 ...
```

3.0 Initial Modelling - without transformations

```
naive.lm2 = lm(price~neighborhood+bldclasscat+taxclasscurr+bldclasscurr+landsqft+grosssqft+taxclasssale+
summary(naive.lm2)
```

```
##
## Call:
## lm(formula = price ~ neighborhood + bldclasscat + taxclasscurr +
##     bldclasscurr + landsqft + grosssqft + taxclasssale + bldclasssale +
##     locality + quarter, data = data_mod2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6550915  -185403    14029   169648 13750503
##
## Coefficients: (10 not defined because of singularities)
##                                          Estimate Std. Error t value
## (Intercept)                             2.024e+04  3.210e+05   0.063
## neighborhoodBAY RIDGE                   3.496e+04  5.484e+04   0.638
## neighborhoodBEDFORD STUYVESANT         -8.790e+04  8.172e+04  -1.076
## neighborhoodBENSONHURST                -6.344e+04  6.160e+04  -1.030
## neighborhoodBERGEN BEACH               -3.488e+05  7.977e+04  -4.373
## neighborhoodBOERUM HILL                 8.046e+05  9.811e+04   8.200
## neighborhoodBOROUGH PARK                6.761e+04  5.872e+04   1.151
## neighborhoodBRIGHTON BEACH             -1.844e+04  8.914e+04  -0.207
## neighborhoodBROOKLYN HEIGHTS            2.178e+06  9.734e+04  22.372
## neighborhoodBROWNSVILLE                -1.824e+05  8.527e+04  -2.139
## neighborhoodBUSH TERMINAL               7.540e+03  3.835e+05   0.020
## neighborhoodBUSHWICK                    8.992e+04  9.302e+04   0.967
## neighborhoodCANARSIE                   -2.029e+05  7.989e+04  -2.539
## neighborhoodCARROLL GARDENS             9.780e+05  1.265e+05   7.733
## neighborhoodCLINTON HILL                6.990e+04  1.020e+05   0.685
## neighborhoodCOBBLE HILL                 1.342e+06  1.126e+05  11.920
## neighborhoodCOBBLE HILL-WEST            3.196e+05  1.298e+05   2.462
## neighborhoodCONEY ISLAND               -2.529e+05  8.218e+04  -3.077
## neighborhoodCROWN HEIGHTS              -4.708e+04  8.190e+04  -0.575
## neighborhoodCYPRESS HILLS              -3.040e+05  8.606e+04  -3.533
## neighborhoodDOWNTOWN-FULTON FERRY       7.102e+05  1.031e+05   6.891
## neighborhoodDOWNTOWN-FULTON MALL        3.141e+05  9.720e+04   3.231
## neighborhoodDOWNTOWN-METROTECH          6.842e+04  1.139e+05   0.601
## neighborhoodDYKER HEIGHTS              -6.927e+03  5.852e+04  -0.118
## neighborhoodEAST NEW YORK              -1.762e+05  8.006e+04  -2.200
## neighborhoodFLATBUSH-CENTRAL            5.535e+04  7.671e+04   0.722
## neighborhoodFLATBUSH-EAST              -2.645e+05  7.396e+04  -3.576
## neighborhoodFLATBUSH-LEFFERTS GARDEN    1.528e+05  8.902e+04   1.716
## neighborhoodFLATBUSH-NORTH             -1.449e+05  8.386e+04  -1.728
## neighborhoodFLATLANDS                  -3.516e+05  7.967e+04  -4.413
## neighborhoodFORT GREENE                 2.645e+05  1.128e+05   2.345
## neighborhoodGERRITSEN BEACH            -3.128e+05  7.680e+04  -4.073
## neighborhoodGOWANUS                     3.511e+05  1.212e+05   2.897
## neighborhoodGRAVESEND                  -4.871e+04  6.460e+04  -0.754
## neighborhoodGREENPOINT                  8.597e+05  1.033e+05   8.326
## neighborhoodKENSINGTON                  2.765e+05  1.130e+05   2.447
```

```
## neighborhoodMADISON                          -7.614e+03  7.467e+04  -0.102
## neighborhoodMANHATTAN BEACH                   1.250e+05  8.625e+04   1.449
## neighborhoodMARINE PARK                      -1.582e+05  7.351e+04  -2.153
## neighborhoodMIDWOOD                           1.337e+04  7.482e+04   0.179
## neighborhoodMILL BASIN                       -3.985e+05  7.974e+04  -4.997
## neighborhoodNAVY YARD                         2.478e+05  1.384e+05   1.791
## neighborhoodOCEAN HILL                       -1.898e+05  1.022e+05  -1.858
## neighborhoodOCEAN PARKWAY-NORTH               3.610e+05  7.507e+04   4.809
## neighborhoodOCEAN PARKWAY-SOUTH               8.134e+05  8.415e+04   9.667
## neighborhoodOLD MILL BASIN                   -2.392e+05  7.528e+04  -3.177
## neighborhoodPARK SLOPE                        8.322e+05  9.336e+04   8.914
## neighborhoodPARK SLOPE SOUTH                  3.063e+05  1.042e+05   2.939
## neighborhoodPROSPECT HEIGHTS                  4.696e+05  9.642e+04   4.870
## neighborhoodRED HOOK                          2.576e+05  1.287e+05   2.002
## neighborhoodSEAGATE                          -4.402e+05  1.015e+05  -4.337
## neighborhoodSHEEPSHEAD BAY                   -1.069e+05  7.515e+04  -1.422
## neighborhoodSPRING CREEK                     -2.984e+05  9.203e+04  -3.242
## neighborhoodSUNSET PARK                       1.303e+05  7.659e+04   1.701
## neighborhoodWILLIAMSBURG-CENTRAL             -8.778e+04  1.177e+05  -0.746
## neighborhoodWILLIAMSBURG-EAST                 6.463e+05  1.073e+05   6.021
## neighborhoodWILLIAMSBURG-NORTH                9.022e+05  9.798e+04   9.208
## neighborhoodWILLIAMSBURG-SOUTH                8.410e+05  1.072e+05   7.848
## neighborhoodWINDSOR TERRACE                   5.318e+05  9.500e+04   5.598
## neighborhoodWYCKOFF HEIGHTS                   2.472e+05  1.648e+05   1.500
## bldclasscat11 SPECIAL CONDO BILLING LOTS     -7.124e+05  3.535e+05  -2.015
## bldclasscat12 CONDOS - WALKUP APARTMENTS     -5.294e+05  6.232e+05  -0.849
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS   -2.358e+05  3.119e+05  -0.756
## taxclasscurr2                                       NA         NA      NA
## bldclasscurrA1                                9.835e+04  3.905e+05   0.252
## bldclasscurrA2                                2.402e+05  4.144e+05   0.580
## bldclasscurrA3                                2.628e+05  4.562e+05   0.576
## bldclasscurrA4                                4.855e+05  5.274e+05   0.920
## bldclasscurrA5                               -4.986e+04  3.662e+05  -0.136
## bldclasscurrA6                                1.826e+05  6.219e+05   0.294
## bldclasscurrA7                                2.655e+06  3.537e+05   7.509
## bldclasscurrA9                               -1.047e+05  3.112e+05  -0.336
## bldclasscurrR2                                3.032e+03  5.385e+05   0.006
## bldclasscurrR4                                      NA         NA      NA
## bldclasscurrRR                                      NA         NA      NA
## landsqft                                      5.036e+00  6.373e-01   7.901
## grosssqft                                     5.272e+02  8.041e+00  65.563
## taxclasssale2                                       NA         NA      NA
## bldclasssaleA1                               -1.537e+05  2.366e+05  -0.649
## bldclasssaleA2                               -2.967e+05  2.741e+05  -1.082
## bldclasssaleA3                               -1.764e+05  3.357e+05  -0.525
## bldclasssaleA4                               -3.756e+05  4.268e+05  -0.880
## bldclasssaleA5                               -1.169e+05  1.933e+05  -0.605
## bldclasssaleA6                                      NA         NA      NA
## bldclasssaleA7                                      NA         NA      NA
## bldclasssaleA9                                      NA         NA      NA
## bldclasssaleR2                                      NA         NA      NA
## bldclasssaleR4                                      NA         NA      NA
## bldclasssaleRR                                      NA         NA      NA
## localityEastern                              -1.221e+05  4.329e+04  -2.821
```

```
## localityNorthern                          -8.477e+04  5.631e+04  -1.505
## localityNorthwestern                       4.741e+05  5.489e+04   8.637
## localitySouthern                          -2.179e+03  3.730e+04  -0.058
## localitySouthwestern                       9.416e+04  5.520e+04   1.706
## quarter2016_Q2                             1.319e+04  3.343e+04   0.395
## quarter2016_Q3                             3.389e+04  3.241e+04   1.046
## quarter2016_Q4                             5.626e+04  3.354e+04   1.677
## quarter2017_Q1                             1.051e+05  3.321e+04   3.165
## quarter2017_Q2                             1.605e+05  3.276e+04   4.899
## quarter2017_Q3                             2.284e+05  3.287e+04   6.948
## quarter2017_Q4                             2.173e+05  3.322e+04   6.542
## quarter2018_Q1                             1.690e+05  3.024e+04   5.589
## quarter2018_Q2                             2.098e+05  3.006e+04   6.980
## quarter2018_Q3                             2.009e+05  3.010e+04   6.674
## quarter2018_Q4                             2.161e+05  3.021e+04   7.153
## quarter2019_Q1                             1.279e+05  3.517e+04   3.636
## quarter2019_Q2                             2.410e+05  3.336e+04   7.225
## quarter2019_Q3                             1.978e+05  3.340e+04   5.921
## quarter2019_Q4                             1.773e+05  3.380e+04   5.245
## quarter2020_Q1                             2.551e+05  3.503e+04   7.282
## quarter2020_Q2                             2.278e+05  3.958e+04   5.755
## quarter2020_Q3                             2.065e+05  3.836e+04   5.384
## quarter2020_Q4                             2.965e+05  3.315e+04   8.943
##                                           Pr(>|t|)
## (Intercept)                               0.949720
## neighborhoodBAY RIDGE                     0.523789
## neighborhoodBEDFORD STUYVESANT            0.282111
## neighborhoodBENSONHURST                   0.303152
## neighborhoodBERGEN BEACH                  1.24e-05 ***
## neighborhoodBOERUM HILL                   2.63e-16 ***
## neighborhoodBOROUGH PARK                  0.249568
## neighborhoodBRIGHTON BEACH                0.836109
## neighborhoodBROOKLYN HEIGHTS               < 2e-16 ***
## neighborhoodBROWNSVILLE                   0.032428 *
## neighborhoodBUSH TERMINAL                 0.984315
## neighborhoodBUSHWICK                      0.333684
## neighborhoodCANARSIE                      0.011117 *
## neighborhoodCARROLL GARDENS               1.13e-14 ***
## neighborhoodCLINTON HILL                  0.493400
## neighborhoodCOBBLE HILL                    < 2e-16 ***
## neighborhoodCOBBLE HILL-WEST              0.013843 *
## neighborhoodCONEY ISLAND                  0.002092 **
## neighborhoodCROWN HEIGHTS                 0.565372
## neighborhoodCYPRESS HILLS                 0.000413 ***
## neighborhoodDOWNTOWN-FULTON FERRY         5.83e-12 ***
## neighborhoodDOWNTOWN-FULTON MALL          0.001237 **
## neighborhoodDOWNTOWN-METROTECH            0.547920
## neighborhoodDYKER HEIGHTS                 0.905768
## neighborhoodEAST NEW YORK                 0.027795 *
## neighborhoodFLATBUSH-CENTRAL              0.470552
## neighborhoodFLATBUSH-EAST                 0.000350 ***
## neighborhoodFLATBUSH-LEFFERTS GARDEN      0.086098 .
## neighborhoodFLATBUSH-NORTH                0.083987 .
## neighborhoodFLATLANDS                     1.03e-05 ***
```

```
## neighborhoodFORT GREENE                    0.019025 *
## neighborhoodGERRITSEN BEACH                4.68e-05 ***
## neighborhoodGOWANUS                        0.003778 **
## neighborhoodGRAVESEND                      0.450783
## neighborhoodGREENPOINT                      < 2e-16 ***
## neighborhoodKENSINGTON                     0.014422 *
## neighborhoodMADISON                        0.918780
## neighborhoodMANHATTAN BEACH                0.147343
## neighborhoodMARINE PARK                    0.031369 *
## neighborhoodMIDWOOD                        0.858181
## neighborhoodMILL BASIN                     5.91e-07 ***
## neighborhoodNAVY YARD                      0.073265 .
## neighborhoodOCEAN HILL                     0.063247 .
## neighborhoodOCEAN PARKWAY-NORTH            1.53e-06 ***
## neighborhoodOCEAN PARKWAY-SOUTH             < 2e-16 ***
## neighborhoodOLD MILL BASIN                 0.001492 **
## neighborhoodPARK SLOPE                      < 2e-16 ***
## neighborhoodPARK SLOPE SOUTH               0.003302 **
## neighborhoodPROSPECT HEIGHTS               1.13e-06 ***
## neighborhoodRED HOOK                       0.045344 *
## neighborhoodSEAGATE                        1.46e-05 ***
## neighborhoodSHEEPSHEAD BAY                 0.154930
## neighborhoodSPRING CREEK                   0.001188 **
## neighborhoodSUNSET PARK                    0.088876 .
## neighborhoodWILLIAMSBURG-CENTRAL           0.455861
## neighborhoodWILLIAMSBURG-EAST              1.78e-09 ***
## neighborhoodWILLIAMSBURG-NORTH              < 2e-16 ***
## neighborhoodWILLIAMSBURG-SOUTH             4.56e-15 ***
## neighborhoodWINDSOR TERRACE                2.21e-08 ***
## neighborhoodWYCKOFF HEIGHTS                0.133639
## bldclasscat11 SPECIAL CONDO BILLING LOTS   0.043892 *
## bldclasscat12 CONDOS - WALKUP APARTMENTS   0.395622
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS 0.449576
## taxclasscurr2                                   NA
## bldclasscurrA1                             0.801164
## bldclasscurrA2                             0.562224
## bldclasscurrA3                             0.564613
## bldclasscurrA4                             0.357364
## bldclasscurrA5                             0.891706
## bldclasscurrA6                             0.769084
## bldclasscurrA7                             6.39e-14 ***
## bldclasscurrA9                             0.736634
## bldclasscurrR2                             0.995508
## bldclasscurrR4                                  NA
## bldclasscurrRR                                  NA
## landsqft                                   2.99e-15 ***
## grosssqft                                   < 2e-16 ***
## taxclasssale2                                   NA
## bldclasssaleA1                             0.516139
## bldclasssaleA2                             0.279078
## bldclasssaleA3                             0.599334
## bldclasssaleA4                             0.378868
## bldclasssaleA5                             0.545505
## bldclasssaleA6                                  NA
```

```
## bldclasssaleA7                                    NA
## bldclasssaleA9                                    NA
## bldclasssaleR2                                    NA
## bldclasssaleR4                                    NA
## bldclasssaleRR                                    NA
## localityEastern                                0.004792 **
## localityNorthern                               0.132269
## localityNorthwestern                            < 2e-16 ***
## localitySouthern                               0.953416
## localitySouthwestern                           0.088082 .
## quarter2016_Q2                                 0.693173
## quarter2016_Q3                                 0.295689
## quarter2016_Q4                                 0.093485 .
## quarter2017_Q1                                 0.001556 **
## quarter2017_Q2                                 9.74e-07 ***
## quarter2017_Q3                                 3.88e-12 ***
## quarter2017_Q4                                 6.31e-11 ***
## quarter2018_Q1                                 2.33e-08 ***
## quarter2018_Q2                                 3.10e-12 ***
## quarter2018_Q3                                 2.59e-11 ***
## quarter2018_Q4                                 8.95e-13 ***
## quarter2019_Q1                                 0.000278 ***
## quarter2019_Q2                                 5.32e-13 ***
## quarter2019_Q3                                 3.29e-09 ***
## quarter2019_Q4                                 1.58e-07 ***
## quarter2020_Q1                                 3.50e-13 ***
## quarter2020_Q2                                 8.86e-09 ***
## quarter2020_Q3                                 7.42e-08 ***
## quarter2020_Q4                                  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 537500 on 12376 degrees of freedom
##   (1185 observations deleted due to missingness)
## Multiple R-squared:  0.6458, Adjusted R-squared:  0.6429
## F-statistic: 221.3 on 102 and 12376 DF,  p-value: < 2.2e-16
```

```
naive.lm3 = lm(price~neighborhood+bldclasscat+taxclasscurr+bldclasscurr+(landsqft*grosssqft)+taxclasssa
summary(naive.lm3)
```

```
##
## Call:
## lm(formula = price ~ neighborhood + bldclasscat + taxclasscurr +
##     bldclasscurr + (landsqft * grosssqft) + taxclasssale + bldclasssale +
##     locality + quarter, data = data_mod2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6546845  -185179    12131   167717 13877684
##
## Coefficients: (10 not defined because of singularities)
##                                           Estimate Std. Error t value
## (Intercept)                              5.766e+04  3.208e+05   0.180
## neighborhoodBAY RIDGE                    2.834e+04  5.480e+04   0.517
```

```
## neighborhoodBEDFORD STUYVESANT            -9.154e+04  8.165e+04  -1.121
## neighborhoodBENSONHURST                   -7.194e+04  6.157e+04  -1.169
## neighborhoodBERGEN BEACH                  -3.572e+05  7.971e+04  -4.482
## neighborhoodBOERUM HILL                    8.056e+05  9.802e+04   8.219
## neighborhoodBOROUGH PARK                   6.134e+04  5.868e+04   1.045
## neighborhoodBRIGHTON BEACH                -2.179e+04  8.905e+04  -0.245
## neighborhoodBROOKLYN HEIGHTS               2.119e+06  9.793e+04  21.638
## neighborhoodBROWNSVILLE                   -1.923e+05  8.521e+04  -2.257
## neighborhoodBUSH TERMINAL                 -8.660e+03  3.831e+05  -0.023
## neighborhoodBUSHWICK                       9.065e+04  9.292e+04   0.975
## neighborhoodCANARSIE                      -2.106e+05  7.983e+04  -2.638
## neighborhoodCARROLL GARDENS                9.779e+05  1.264e+05   7.740
## neighborhoodCLINTON HILL                   7.266e+04  1.019e+05   0.713
## neighborhoodCOBBLE HILL                    1.355e+06  1.125e+05  12.044
## neighborhoodCOBBLE HILL-WEST               3.215e+05  1.297e+05   2.479
## neighborhoodCONEY ISLAND                  -2.614e+05  8.211e+04  -3.183
## neighborhoodCROWN HEIGHTS                 -5.304e+04  8.183e+04  -0.648
## neighborhoodCYPRESS HILLS                 -3.104e+05  8.599e+04  -3.610
## neighborhoodDOWNTOWN-FULTON FERRY          6.883e+05  1.031e+05   6.679
## neighborhoodDOWNTOWN-FULTON MALL           3.101e+05  9.711e+04   3.194
## neighborhoodDOWNTOWN-METROTECH             5.580e+04  1.138e+05   0.490
## neighborhoodDYKER HEIGHTS                 -1.542e+04  5.849e+04  -0.264
## neighborhoodEAST NEW YORK                 -1.864e+05  8.001e+04  -2.330
## neighborhoodFLATBUSH-CENTRAL               4.725e+04  7.665e+04   0.616
## neighborhoodFLATBUSH-EAST                 -2.711e+05  7.390e+04  -3.668
## neighborhoodFLATBUSH-LEFFERTS GARDEN       1.486e+05  8.893e+04   1.671
## neighborhoodFLATBUSH-NORTH                -1.526e+05  8.379e+04  -1.821
## neighborhoodFLATLANDS                     -3.585e+05  7.960e+04  -4.503
## neighborhoodFORT GREENE                    2.618e+05  1.126e+05   2.324
## neighborhoodGERRITSEN BEACH               -3.225e+05  7.675e+04  -4.202
## neighborhoodGOWANUS                        3.441e+05  1.211e+05   2.841
## neighborhoodGRAVESEND                     -5.661e+04  6.455e+04  -0.877
## neighborhoodGREENPOINT                     8.546e+05  1.032e+05   8.284
## neighborhoodKENSINGTON                     2.708e+05  1.129e+05   2.398
## neighborhoodMADISON                       -1.450e+04  7.460e+04  -0.194
## neighborhoodMANHATTAN BEACH                1.207e+05  8.617e+04   1.401
## neighborhoodMARINE PARK                   -1.664e+05  7.345e+04  -2.266
## neighborhoodMIDWOOD                        7.677e+03  7.475e+04   0.103
## neighborhoodMILL BASIN                    -4.046e+05  7.967e+04  -5.078
## neighborhoodNAVY YARD                      2.568e+05  1.382e+05   1.858
## neighborhoodOCEAN HILL                    -1.977e+05  1.021e+05  -1.937
## neighborhoodOCEAN PARKWAY-NORTH            3.552e+05  7.501e+04   4.736
## neighborhoodOCEAN PARKWAY-SOUTH            8.077e+05  8.407e+04   9.607
## neighborhoodOLD MILL BASIN                -2.483e+05  7.523e+04  -3.301
## neighborhoodPARK SLOPE                     8.313e+05  9.327e+04   8.913
## neighborhoodPARK SLOPE SOUTH               2.973e+05  1.041e+05   2.855
## neighborhoodPROSPECT HEIGHTS               4.915e+05  9.643e+04   5.097
## neighborhoodRED HOOK                       2.562e+05  1.286e+05   1.993
## neighborhoodSEAGATE                       -4.468e+05  1.014e+05  -4.405
## neighborhoodSHEEPSHEAD BAY                -1.079e+05  7.508e+04  -1.437
## neighborhoodSPRING CREEK                  -3.039e+05  9.195e+04  -3.305
## neighborhoodSUNSET PARK                    1.209e+05  7.654e+04   1.579
## neighborhoodWILLIAMSBURG-CENTRAL          -8.232e+04  1.176e+05  -0.700
## neighborhoodWILLIAMSBURG-EAST              6.399e+05  1.072e+05   5.967
```

```
## neighborhoodWILLIAMSBURG-NORTH              9.168e+05  9.793e+04   9.362
## neighborhoodWILLIAMSBURG-SOUTH              8.347e+05  1.071e+05   7.797
## neighborhoodWINDSOR TERRACE                 5.269e+05  9.491e+04   5.552
## neighborhoodWYCKOFF HEIGHTS                 2.413e+05  1.647e+05   1.465
## bldclasscat11 SPECIAL CONDO BILLING LOTS   -6.787e+05  3.532e+05  -1.921
## bldclasscat12 CONDOS - WALKUP APARTMENTS   -5.289e+05  6.226e+05  -0.850
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS -2.415e+05  3.116e+05  -0.775
## taxclasscurr2                                      NA         NA      NA
## bldclasscurrA1                              9.503e+04  3.901e+05   0.244
## bldclasscurrA2                              2.312e+05  4.140e+05   0.558
## bldclasscurrA3                              2.543e+05  4.558e+05   0.558
## bldclasscurrA4                              4.888e+05  5.269e+05   0.928
## bldclasscurrA5                             -5.328e+04  3.659e+05  -0.146
## bldclasscurrA6                              1.759e+05  6.213e+05   0.283
## bldclasscurrA7                              2.620e+06  3.534e+05   7.413
## bldclasscurrA9                             -1.084e+05  3.109e+05  -0.349
## bldclasscurrR2                             -3.702e+03  5.380e+05  -0.007
## bldclasscurrR4                                     NA         NA      NA
## bldclasscurrRR                                     NA         NA      NA
## landsqft                                   -2.377e-01  1.226e+00  -0.194
## grosssqft                                   5.090e+02  8.807e+00  57.796
## taxclasssale2                                      NA         NA      NA
## bldclasssaleA1                             -1.527e+05  2.364e+05  -0.646
## bldclasssaleA2                             -2.946e+05  2.739e+05  -1.076
## bldclasssaleA3                             -1.753e+05  3.354e+05  -0.523
## bldclasssaleA4                             -3.701e+05  4.264e+05  -0.868
## bldclasssaleA5                             -1.159e+05  1.931e+05  -0.600
## bldclasssaleA6                                     NA         NA      NA
## bldclasssaleA7                                     NA         NA      NA
## bldclasssaleA9                                     NA         NA      NA
## bldclasssaleR2                                     NA         NA      NA
## bldclasssaleR4                                     NA         NA      NA
## bldclasssaleRR                                     NA         NA      NA
## localityEastern                            -1.217e+05  4.325e+04  -2.815
## localityNorthern                           -8.630e+04  5.626e+04  -1.534
## localityNorthwestern                        4.737e+05  5.484e+04   8.638
## localitySouthern                           -1.495e+03  3.726e+04  -0.040
## localitySouthwestern                        9.589e+04  5.515e+04   1.739
## quarter2016_Q2                              1.297e+04  3.340e+04   0.388
## quarter2016_Q3                              3.412e+04  3.237e+04   1.054
## quarter2016_Q4                              5.573e+04  3.351e+04   1.663
## quarter2017_Q1                              1.044e+05  3.318e+04   3.146
## quarter2017_Q2                              1.593e+05  3.273e+04   4.869
## quarter2017_Q3                              2.278e+05  3.284e+04   6.938
## quarter2017_Q4                              2.162e+05  3.319e+04   6.515
## quarter2018_Q1                              1.689e+05  3.021e+04   5.592
## quarter2018_Q2                              2.090e+05  3.003e+04   6.960
## quarter2018_Q3                              1.999e+05  3.007e+04   6.647
## quarter2018_Q4                              2.130e+05  3.019e+04   7.056
## quarter2019_Q1                              1.261e+05  3.514e+04   3.588
## quarter2019_Q2                              2.404e+05  3.333e+04   7.212
## quarter2019_Q3                              1.968e+05  3.337e+04   5.897
## quarter2019_Q4                              1.765e+05  3.377e+04   5.228
## quarter2020_Q1                              2.536e+05  3.500e+04   7.248
```

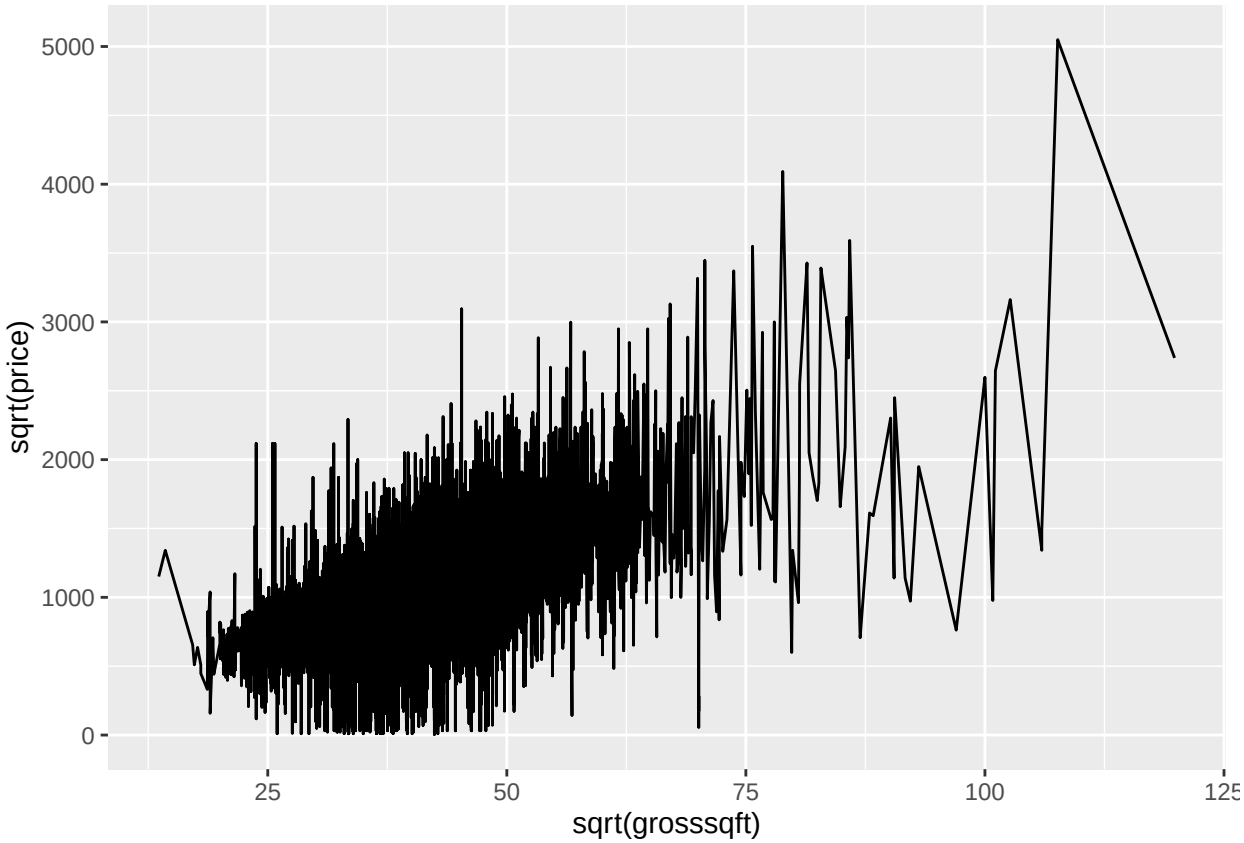
```
## quarter2020_Q2                          2.269e+05  3.954e+04   5.738
## quarter2020_Q3                          2.063e+05  3.832e+04   5.385
## quarter2020_Q4                          2.962e+05  3.312e+04   8.943
## landsqft:grosssqft                      3.397e-03  6.745e-04   5.036
##                                        Pr(>|t|)
## (Intercept)                            0.857344
## neighborhoodBAY RIDGE                  0.605002
## neighborhoodBEDFORD STUYVESANT         0.262210
## neighborhoodBENSONHURST                0.242612
## neighborhoodBERGEN BEACH               7.47e-06 ***
## neighborhoodBOERUM HILL                2.26e-16 ***
## neighborhoodBOROUGH PARK               0.295816
## neighborhoodBRIGHTON BEACH             0.806710
## neighborhoodBROOKLYN HEIGHTS            < 2e-16 ***
## neighborhoodBROWNSVILLE                0.024009 *
## neighborhoodBUSH TERMINAL              0.981969
## neighborhoodBUSHWICK                   0.329339
## neighborhoodCANARSIE                   0.008345 **
## neighborhoodCARROLL GARDENS            1.07e-14 ***
## neighborhoodCLINTON HILL               0.476045
## neighborhoodCOBBLE HILL                 < 2e-16 ***
## neighborhoodCOBBLE HILL-WEST           0.013195 *
## neighborhoodCONEY ISLAND               0.001461 **
## neighborhoodCROWN HEIGHTS              0.516889
## neighborhoodCYPRESS HILLS              0.000308 ***
## neighborhoodDOWNTOWN-FULTON FERRY      2.51e-11 ***
## neighborhoodDOWNTOWN-FULTON MALL       0.001408 **
## neighborhoodDOWNTOWN-METROTECH         0.623882
## neighborhoodDYKER HEIGHTS              0.792062
## neighborhoodEAST NEW YORK              0.019837 *
## neighborhoodFLATBUSH-CENTRAL           0.537587
## neighborhoodFLATBUSH-EAST              0.000245 ***
## neighborhoodFLATBUSH-LEFFERTS GARDEN   0.094683 .
## neighborhoodFLATBUSH-NORTH             0.068576 .
## neighborhoodFLATLANDS                  6.75e-06 ***
## neighborhoodFORT GREENE                0.020143 *
## neighborhoodGERRITSEN BEACH            2.67e-05 ***
## neighborhoodGOWANUS                    0.004503 **
## neighborhoodGRAVESEND                  0.380542
## neighborhoodGREENPOINT                  < 2e-16 ***
## neighborhoodKENSINGTON                 0.016486 *
## neighborhoodMADISON                    0.845854
## neighborhoodMANHATTAN BEACH            0.161170
## neighborhoodMARINE PARK                0.023489 *
## neighborhoodMIDWOOD                    0.918202
## neighborhoodMILL BASIN                 3.87e-07 ***
## neighborhoodNAVY YARD                  0.063183 .
## neighborhoodOCEAN HILL                 0.052773 .
## neighborhoodOCEAN PARKWAY-NORTH        2.21e-06 ***
## neighborhoodOCEAN PARKWAY-SOUTH         < 2e-16 ***
## neighborhoodOLD MILL BASIN             0.000966 ***
## neighborhoodPARK SLOPE                  < 2e-16 ***
## neighborhoodPARK SLOPE SOUTH           0.004310 **
## neighborhoodPROSPECT HEIGHTS           3.50e-07 ***
```

```
## neighborhoodRED HOOK                          0.046336 *
## neighborhoodSEAGATE                           1.07e-05 ***
## neighborhoodSHEEPSHEAD BAY                    0.150689
## neighborhoodSPRING CREEK                      0.000952 ***
## neighborhoodSUNSET PARK                       0.114325
## neighborhoodWILLIAMSBURG-CENTRAL              0.483965
## neighborhoodWILLIAMSBURG-EAST                 2.48e-09 ***
## neighborhoodWILLIAMSBURG-NORTH                 < 2e-16 ***
## neighborhoodWILLIAMSBURG-SOUTH                6.86e-15 ***
## neighborhoodWINDSOR TERRACE                   2.89e-08 ***
## neighborhoodWYCKOFF HEIGHTS                   0.142925
## bldclasscat11 SPECIAL CONDO BILLING LOTS      0.054720 .
## bldclasscat12 CONDOS - WALKUP APARTMENTS      0.395610
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS    0.438289
## taxclasscurr2                                       NA
## bldclasscurrA1                                0.807541
## bldclasscurrA2                                0.576545
## bldclasscurrA3                                0.576861
## bldclasscurrA4                                0.353652
## bldclasscurrA5                                0.884230
## bldclasscurrA6                                0.777132
## bldclasscurrA7                                1.32e-13 ***
## bldclasscurrA9                                0.727404
## bldclasscurrR2                                0.994510
## bldclasscurrR4                                      NA
## bldclasscurrRR                                      NA
## landsqft                                      0.846227
## grosssqft                                      < 2e-16 ***
## taxclasssale2                                       NA
## bldclasssaleA1                                0.518332
## bldclasssaleA2                                0.282116
## bldclasssaleA3                                0.601099
## bldclasssaleA4                                0.385452
## bldclasssaleA5                                0.548293
## bldclasssaleA6                                      NA
## bldclasssaleA7                                      NA
## bldclasssaleA9                                      NA
## bldclasssaleR2                                      NA
## bldclasssaleR4                                      NA
## bldclasssaleRR                                      NA
## localityEastern                               0.004888 **
## localityNorthern                              0.125051
## localityNorthwestern                           < 2e-16 ***
## localitySouthern                              0.967987
## localitySouthwestern                          0.082110 .
## quarter2016_Q2                                0.697759
## quarter2016_Q3                                0.291924
## quarter2016_Q4                                0.096260 .
## quarter2017_Q1                                0.001659 **
## quarter2017_Q2                                1.14e-06 ***
## quarter2017_Q3                                4.17e-12 ***
## quarter2017_Q4                                7.54e-11 ***
## quarter2018_Q1                                2.29e-08 ***
## quarter2018_Q2                                3.56e-12 ***
```

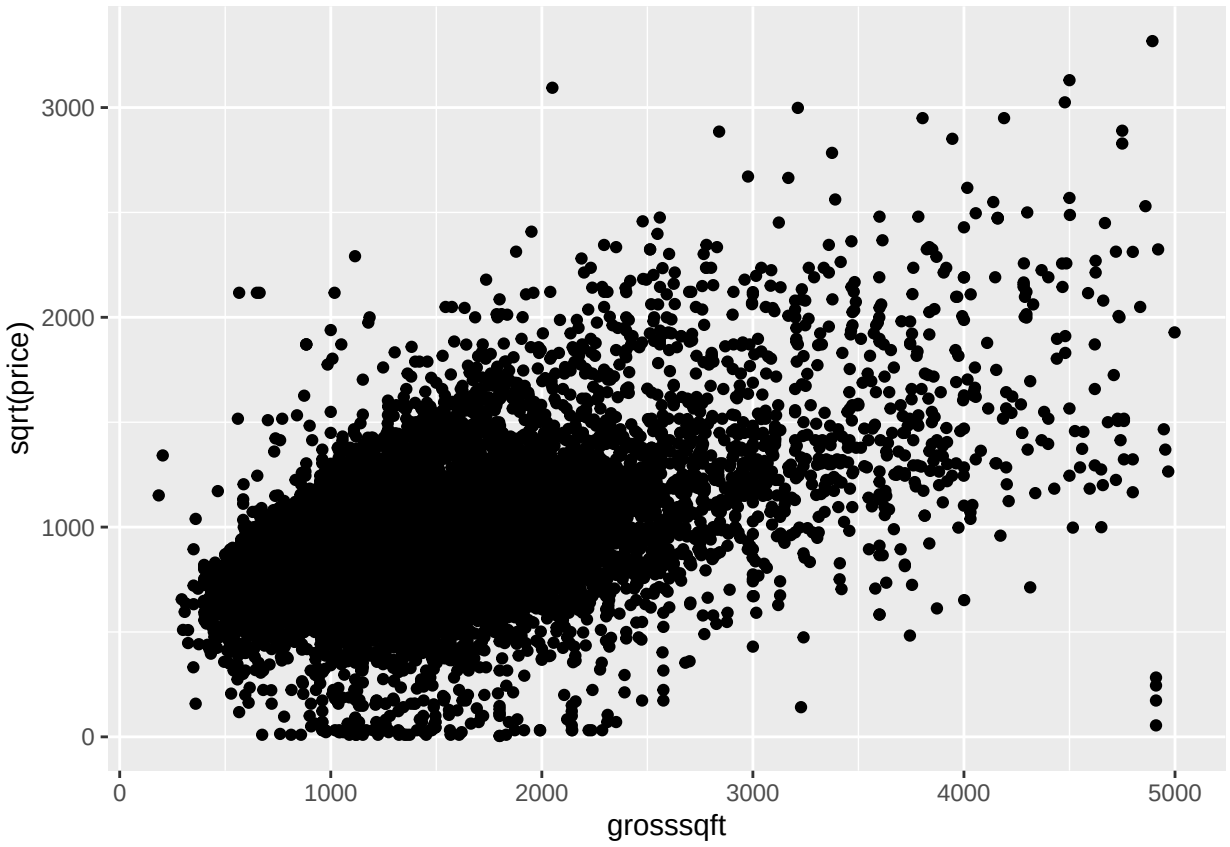
```
## quarter2018_Q3                              3.12e-11 ***
## quarter2018_Q4                              1.81e-12 ***
## quarter2019_Q1                              0.000334 ***
## quarter2019_Q2                              5.84e-13 ***
## quarter2019_Q3                              3.79e-09 ***
## quarter2019_Q4                              1.74e-07 ***
## quarter2020_Q1                              4.49e-13 ***
## quarter2020_Q2                              9.83e-09 ***
## quarter2020_Q3                              7.39e-08 ***
## quarter2020_Q4                               < 2e-16 ***
## landsqft:grosssqft                          4.82e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 536900 on 12375 degrees of freedom
##   (1185 observations deleted due to missingness)
## Multiple R-squared:  0.6466, Adjusted R-squared:  0.6436
## F-statistic: 219.8 on 103 and 12375 DF,  p-value: < 2.2e-16
```

3.1 Continuing Initial Modelling - Analyzing Transformations

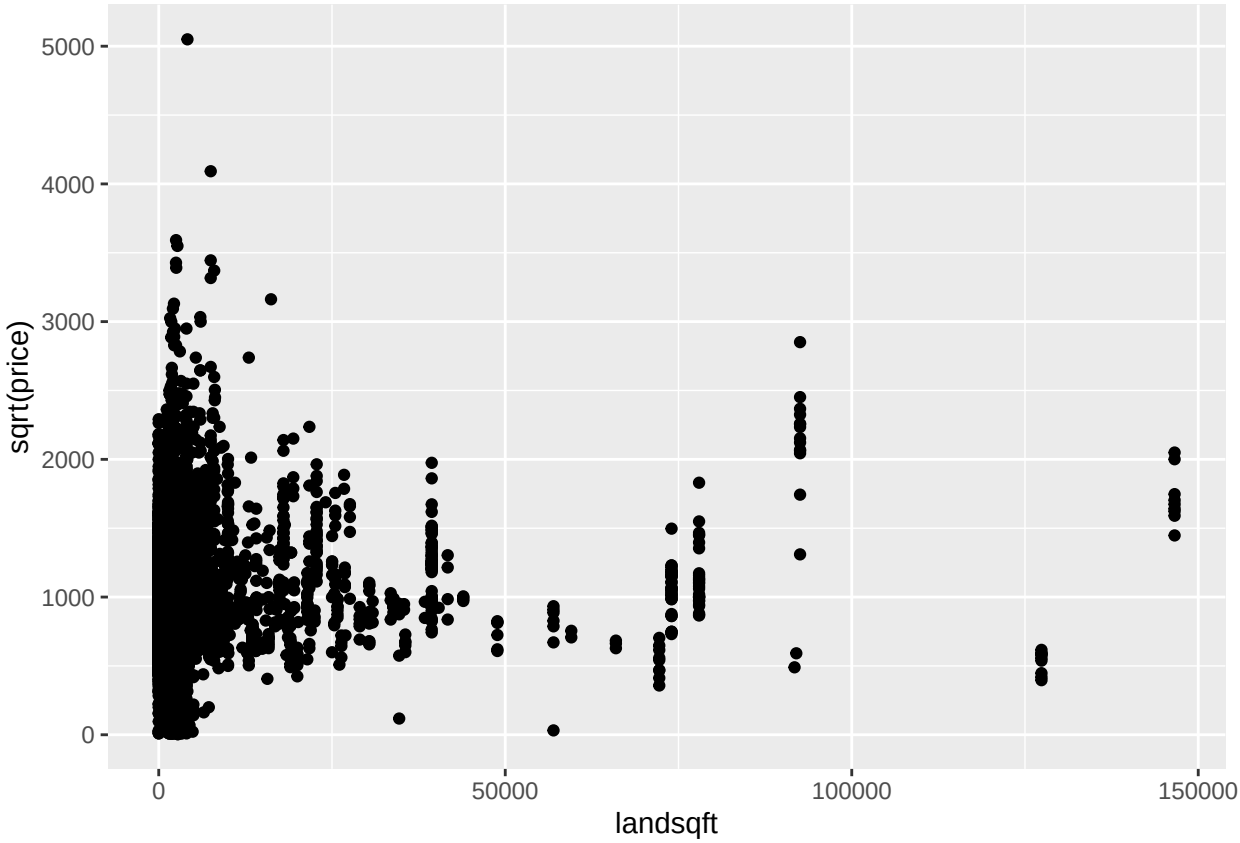
```
ggplot(data_mod2, aes(sqrt(grosssqft), sqrt(price))) + geom_line()
```



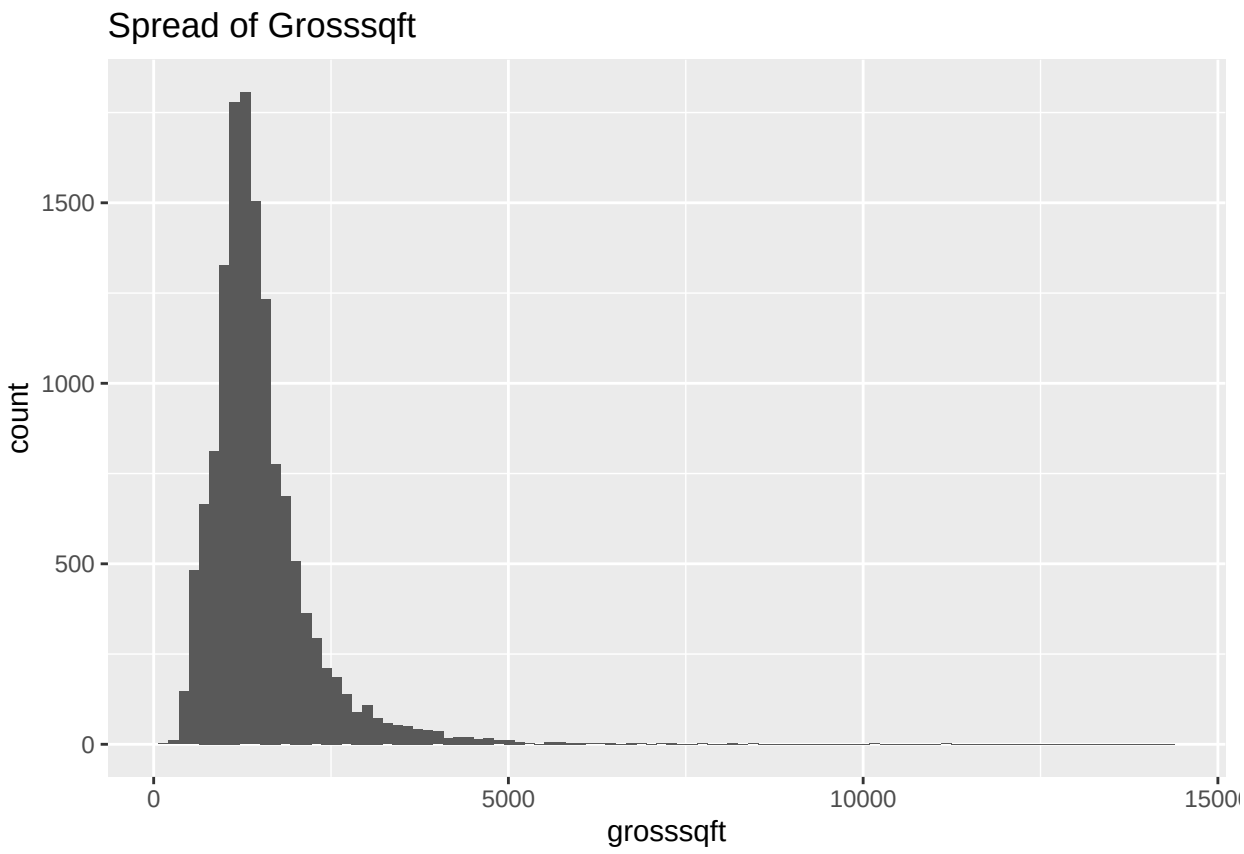
```
ggplot(data_mod2[data_mod2$grosssqft<5000,], aes(grosssqft, sqrt(price))) + geom_point()
```



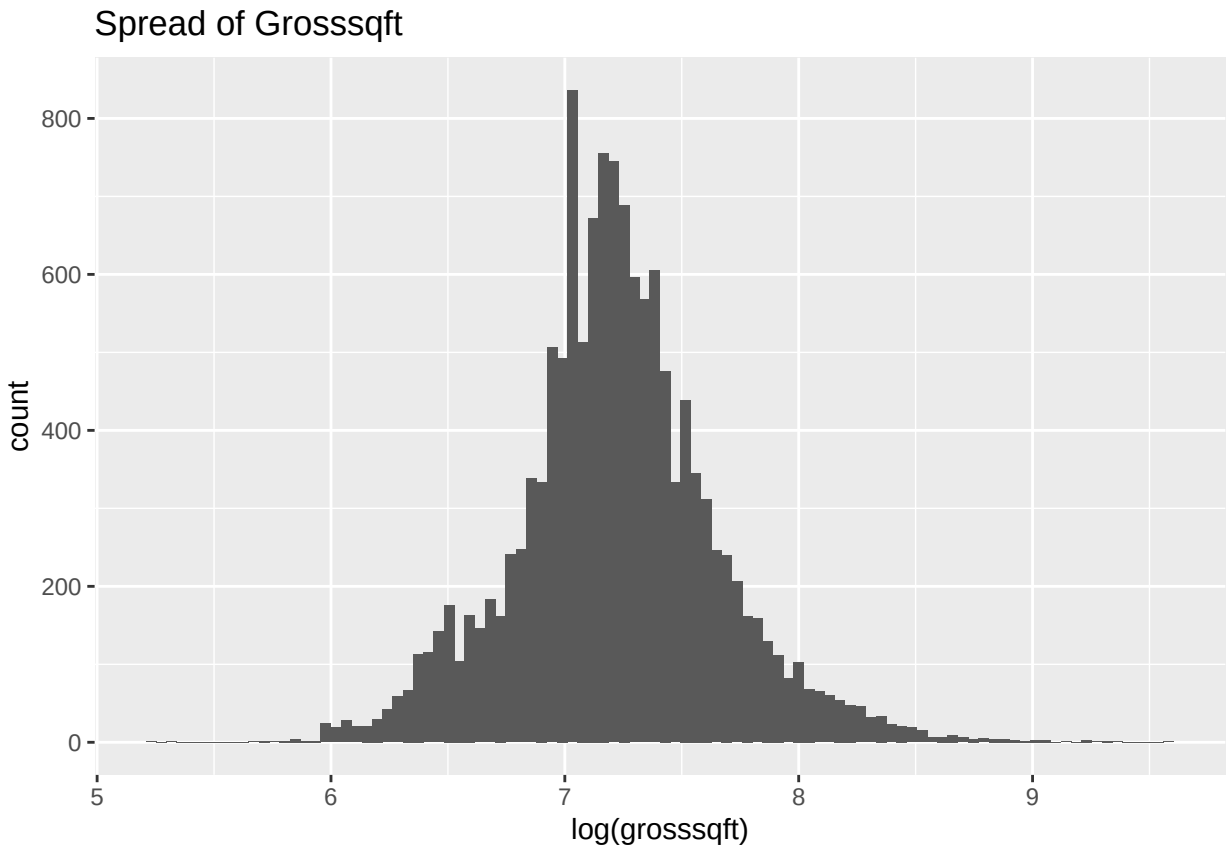

```
ggplot(data_mod2, aes(landsqft, sqrt(price))) + geom_point()
```



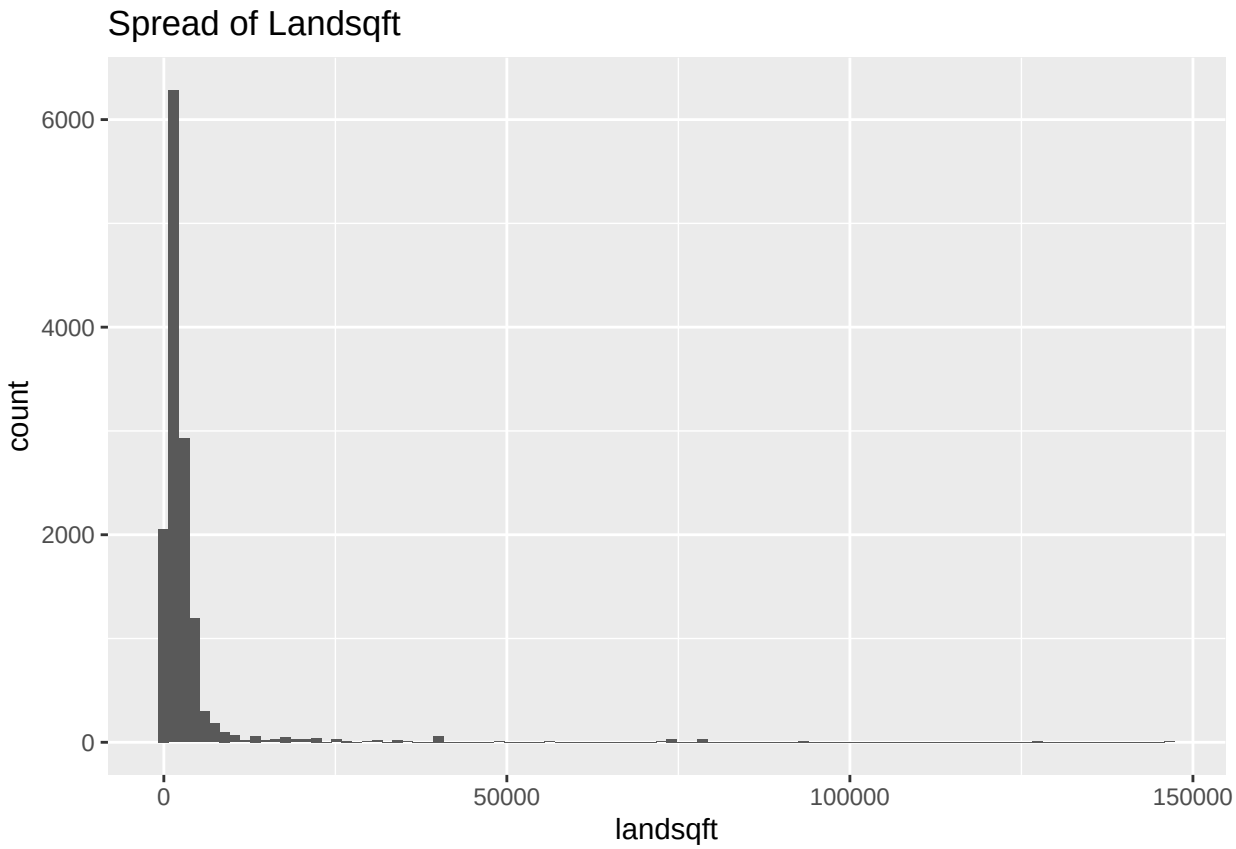

```
ggplot(data_mod, aes(x=grosssqft)) + geom_histogram(bins = 100) + ggtitle("Spread of Grosssqft")
```



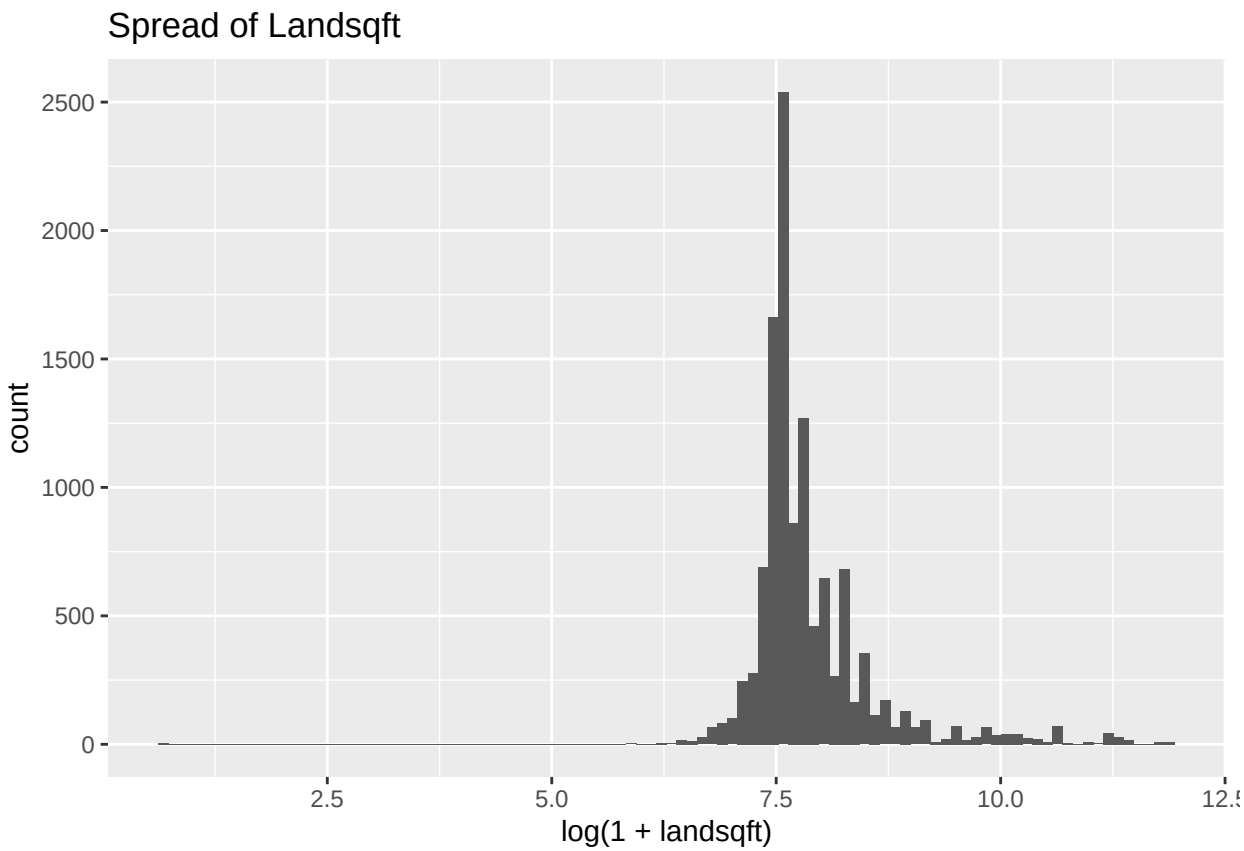

```
ggplot(data_mod, aes(x=log(grosssqft))) + geom_histogram(bins = 100) + ggtitle("Spread of Grosssqft")
```



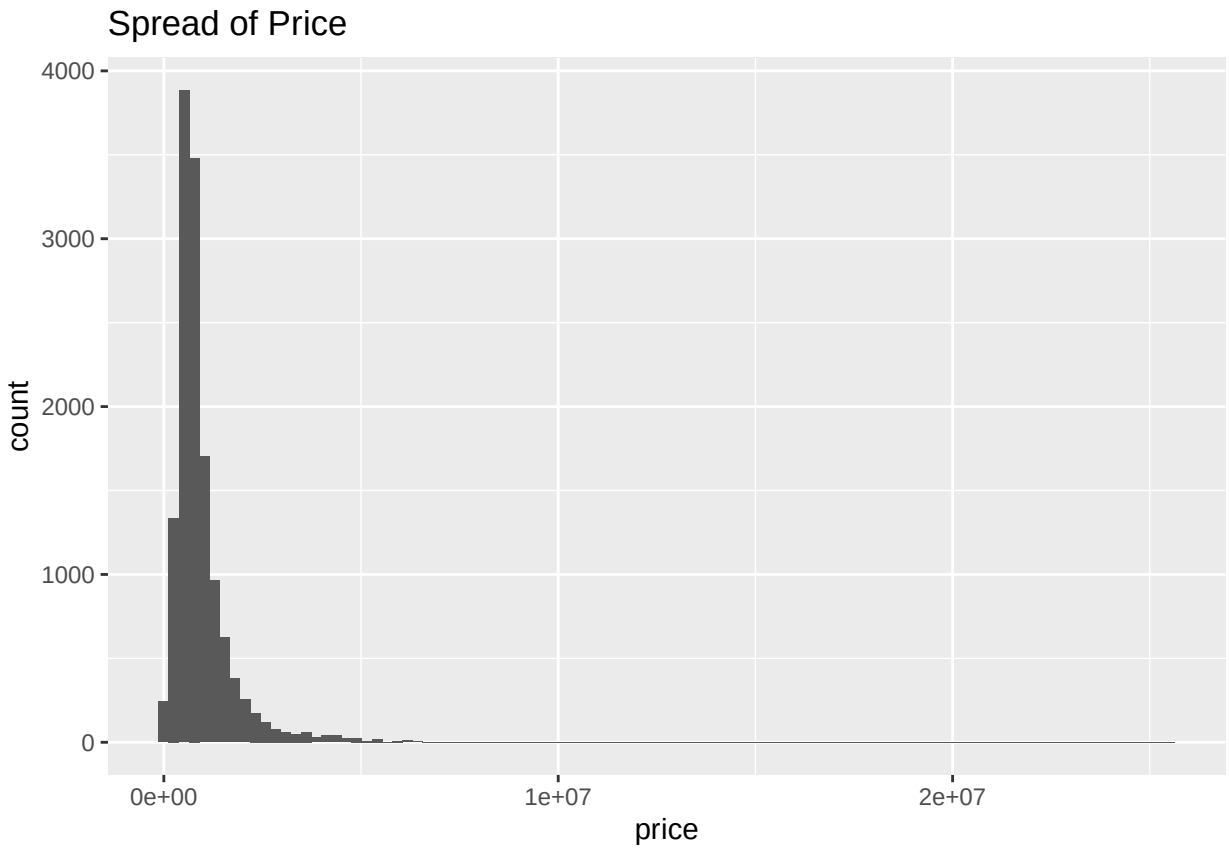

```
ggplot(data_mod, aes(x=landsqft)) + geom_histogram(bins = 100) + ggtitle("Spread of Landsqft")
```



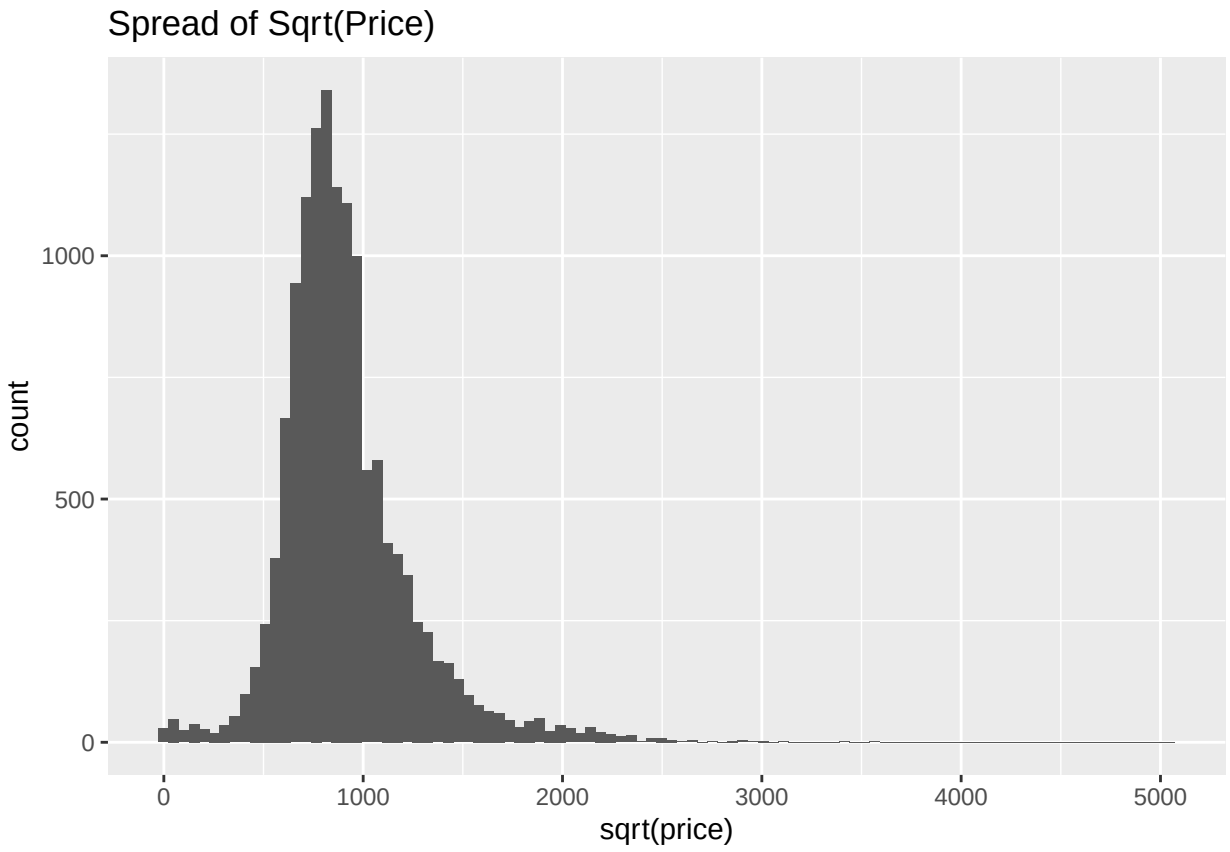

```
ggplot(data_mod[data_mod$landsqft>0,], aes(x=log(1+landsqft))) + geom_histogram(bins = 100) + ggtitle("S
```



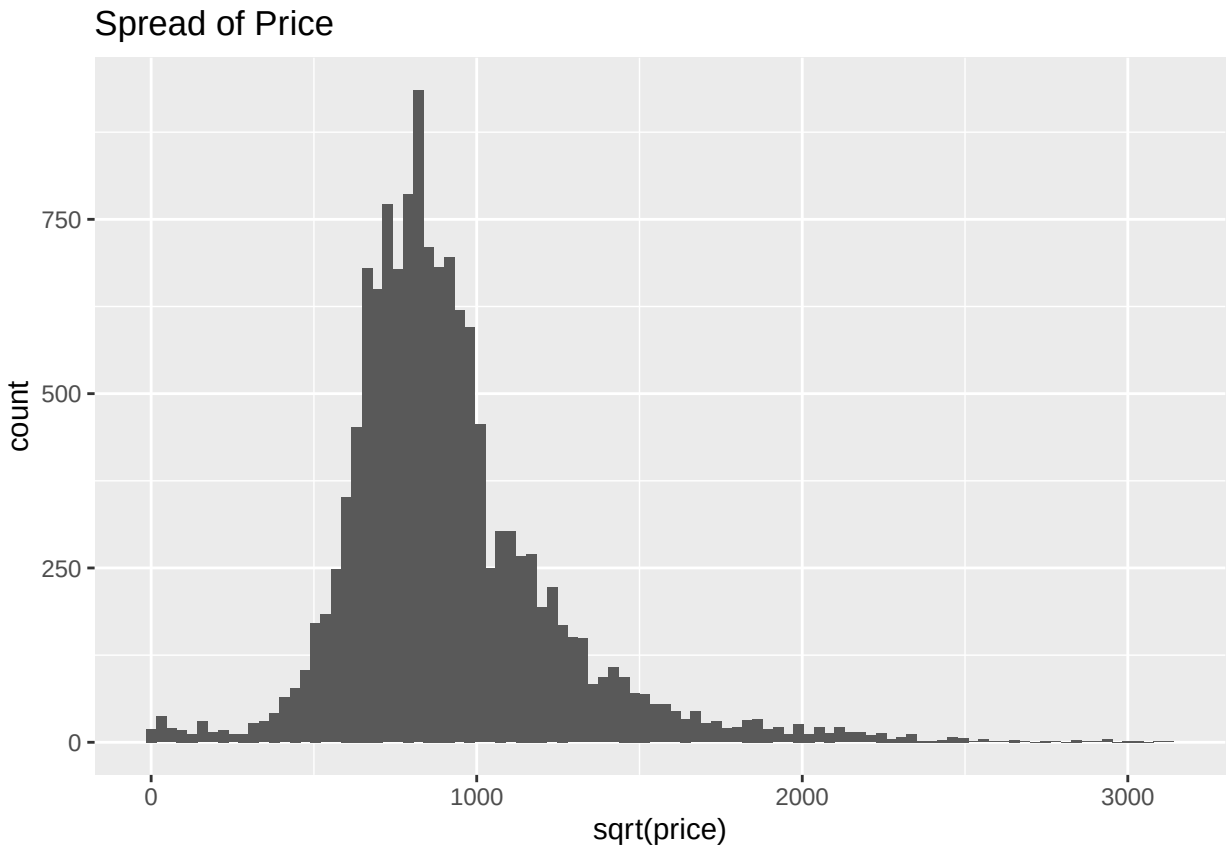

```
ggplot(data_mod, aes(x=price)) + geom_histogram(bins = 100) + ggtitle("Spread of Price")
```




```
ggplot(data_mod, aes(x=sqrt(price))) + geom_histogram(bins = 100) + ggtitle("Spread of Sqrt(Price)")
```




```
ggplot(data_mod[data_mod$price<10000000,], aes(x=sqrt(price))) + geom_histogram(bins = 100) + ggtitle("S
```



3.2 Continuing with feature engineering

```
# For experimenting - trying year and quarter separately also
data_mod2$quartf = as.factor(substr(data_mod2$quarter,6,7))
data_mod2$yearsl = as.factor(substr(data_mod2$quarter,1,4))

# Adding neighbourhood levels with price
new_neigh_level10 = as.vector(neigh_price$Group.1[55:60])
new_neigh_level9 = as.vector(neigh_price$Group.1[49:54])
new_neigh_level8 = as.vector(neigh_price$Group.1[43:48])
new_neigh_level7 = as.vector(neigh_price$Group.1[37:42])
new_neigh_level6 = as.vector(neigh_price$Group.1[31:36])
new_neigh_level5 = as.vector(neigh_price$Group.1[25:30])
new_neigh_level4 = as.vector(neigh_price$Group.1[19:24])
new_neigh_level3 = as.vector(neigh_price$Group.1[13:18])
new_neigh_level2 = as.vector(neigh_price$Group.1[7:12])
new_neigh_level1 = as.vector(neigh_price$Group.1[1:6])

data_mod2$new_neigh_level = as.factor(ifelse(data_mod2$neighborhood %in% new_neigh_level1, 'new_neigh_l
                                      ifelse(data_mod2$neighborhood %in% new_neigh_level2, 'new_neigh
                                             ifelse(data_mod2$neighborhood %in% new_neigh_level3, 'n
                                                    ifelse(data_mod2$neighborhood %in% new_neigh_leve
                                                           ifelse(data_mod2$neighborhood %in% new_ne
                                                                  ifelse(data_mod2$neighborhood %in%
```

```
                                                                                ifelse(data_mod2$neighborhoo
                                                                                       ifelse(data_mod2$neig
                                                                                              ifelse(data_mo
                                                                                                     ifelse
```

3.3 Intermediate Models - Deprioritized

```
trans.lm6 = lm(I(sqrt(price))~new_neigh_level+bldclasscat+(landsqft*I(log(grosssqft)))+locality+yearsl+q
summary(trans.lm6)
```

```
##
## Call:
## lm(formula = I(sqrt(price)) ~ new_neigh_level + bldclasscat +
##     (landsqft * I(log(grosssqft))) + locality + yearsl + quartf,
##     data = data_mod2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1582.39   -82.22     9.67    91.18  2776.67
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                -1.764e+03  3.973e+01 -44.413
## new_neigh_levelnew_neigh_level10            5.361e+02  1.278e+01  41.947
## new_neigh_levelnew_neigh_level2             1.031e+01  7.525e+00   1.370
## new_neigh_levelnew_neigh_level3             8.620e+01  9.170e+00   9.400
## new_neigh_levelnew_neigh_level4             1.125e+02  1.032e+01  10.898
## new_neigh_levelnew_neigh_level5             1.725e+02  9.843e+00  17.529
## new_neigh_levelnew_neigh_level6             1.969e+02  9.437e+00  20.866
## new_neigh_levelnew_neigh_level7             3.310e+02  1.208e+01  27.402
## new_neigh_levelnew_neigh_level8             3.315e+02  1.068e+01  31.030
## new_neigh_levelnew_neigh_level9             3.447e+02  1.563e+01  22.045
## bldclasscat04 TAX CLASS 1 CONDOS           -5.829e+01  1.163e+01  -5.014
## bldclasscat11 SPECIAL CONDO BILLING LOTS    6.992e+01  6.262e+01   1.117
## bldclasscat12 CONDOS - WALKUP APARTMENTS   -1.363e+02  1.391e+01  -9.797
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS -5.063e+00  7.782e+00  -0.651
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL -4.070e+01  9.894e+00  -4.114
## landsqft                                   -2.592e-02  2.655e-03  -9.763
## I(log(grosssqft))                           3.381e+02  5.297e+00  63.819
## localityEastern                            -1.006e+02  9.837e+00 -10.222
## localityNorthern                            1.058e+02  9.510e+00  11.128
## localityNorthwestern                        1.541e+02  9.850e+00  15.648
## localitySouthern                           -4.196e+01  7.162e+00  -5.858
## localitySouthwestern                        1.654e+01  7.954e+00   2.080
## yearsl2017                                  6.305e+01  6.191e+00  10.184
## yearsl2018                                  8.253e+01  6.336e+00  13.025
## yearsl2019                                  8.290e+01  6.373e+00  13.010
## yearsl2020                                  1.141e+02  6.795e+00  16.787
## quartfQ2                                    1.344e+01  5.013e+00   2.680
## quartfQ3                                    1.943e+01  5.020e+00   3.871
## quartfQ4                                    2.307e+01  5.008e+00   4.606
```

```
## landsqft:I(log(grosssqft))                    3.979e-03  3.732e-04  10.661
##                                                Pr(>|t|)
## (Intercept)                                     < 2e-16 ***
## new_neigh_levelnew_neigh_level10                < 2e-16 ***
## new_neigh_levelnew_neigh_level2                0.170668
## new_neigh_levelnew_neigh_level3                 < 2e-16 ***
## new_neigh_levelnew_neigh_level4                 < 2e-16 ***
## new_neigh_levelnew_neigh_level5                 < 2e-16 ***
## new_neigh_levelnew_neigh_level6                 < 2e-16 ***
## new_neigh_levelnew_neigh_level7                 < 2e-16 ***
## new_neigh_levelnew_neigh_level8                 < 2e-16 ***
## new_neigh_levelnew_neigh_level9                 < 2e-16 ***
## bldclasscat04 TAX CLASS 1 CONDOS               5.40e-07 ***
## bldclasscat11 SPECIAL CONDO BILLING LOTS       0.264211
## bldclasscat12 CONDOS - WALKUP APARTMENTS        < 2e-16 ***
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS     0.515320
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL 3.91e-05 ***
## landsqft                                        < 2e-16 ***
## I(log(grosssqft))                               < 2e-16 ***
## localityEastern                                 < 2e-16 ***
## localityNorthern                                < 2e-16 ***
## localityNorthwestern                            < 2e-16 ***
## localitySouthern                               4.79e-09 ***
## localitySouthwestern                           0.037575 *
## yearsl2017                                      < 2e-16 ***
## yearsl2018                                      < 2e-16 ***
## yearsl2019                                      < 2e-16 ***
## yearsl2020                                      < 2e-16 ***
## quartfQ2                                       0.007360 **
## quartfQ3                                       0.000109 ***
## quartfQ4                                       4.15e-06 ***
## landsqft:I(log(grosssqft))                      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 205.3 on 13634 degrees of freedom
## Multiple R-squared:  0.6356, Adjusted R-squared:  0.6349
## F-statistic: 820.2 on 29 and 13634 DF,  p-value: < 2.2e-16
```

```
sqrt(sum(((trans.lm6$fitted.values)^2 - data_mod2$price)^2)/length(trans.lm6$fitted.values))
```

```
## [1] 557243.9
```

```
#TRANSFORMING LANDSQFT
trans.lm7 = lm(I(sqrt(price))~new_neigh_level+bldclasscat+I(log(1+landsqft))+I(log(grosssqft))+locality+
summary(trans.lm7)
```

```
##
## Call:
## lm(formula = I(sqrt(price)) ~ new_neigh_level + bldclasscat +
##     I(log(1 + landsqft)) + I(log(grosssqft)) + locality + yearsl +
##     quartf, data = data_mod2)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1598.38   -83.88     9.51    93.05  2762.92
##
## Coefficients:
##                                           Estimate Std. Error t value
## (Intercept)                              -1974.5788    37.4991 -52.657
## new_neigh_levelnew_neigh_level10           531.0076    12.8661  41.272
## new_neigh_levelnew_neigh_level2              7.0977     7.5625   0.939
## new_neigh_levelnew_neigh_level3             84.4572     9.2149   9.165
## new_neigh_levelnew_neigh_level4            102.3138    10.3820   9.855
## new_neigh_levelnew_neigh_level5            163.7256     9.8892  16.556
## new_neigh_levelnew_neigh_level6            191.6039     9.4869  20.197
## new_neigh_levelnew_neigh_level7            323.6033    12.1328  26.672
## new_neigh_levelnew_neigh_level8            325.0899    10.7280  30.303
## new_neigh_levelnew_neigh_level9            346.3168    15.6969  22.063
## bldclasscat04 TAX CLASS 1 CONDOS           -26.2116    11.6725  -2.246
## bldclasscat11 SPECIAL CONDO BILLING LOTS    97.1843    62.8850   1.545
## bldclasscat12 CONDOS - WALKUP APARTMENTS   -99.8707    14.2009  -7.033
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS  45.3106     8.1888   5.533
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL -9.8950   10.2195  -0.968
## I(log(1 + landsqft))                         7.4560     0.7645   9.752
## I(log(grosssqft))                          360.6365     4.9481  72.883
## localityEastern                           -103.4349     9.8935 -10.455
## localityNorthern                           107.0475     9.5636  11.193
## localityNorthwestern                       155.2977     9.9030  15.682
## localitySouthern                           -42.2844     7.2026  -5.871
## localitySouthwestern                        16.4995     7.9967   2.063
## yearsl2017                                  63.7264     6.2240  10.239
## yearsl2018                                  83.3841     6.3697  13.091
## yearsl2019                                  83.5748     6.4062  13.046
## yearsl2020                                 114.5303     6.8306  16.767
## quartfQ2                                    13.3340     5.0409   2.645
## quartfQ3                                    19.6825     5.0468   3.900
## quartfQ4                                    23.9924     5.0353   4.765
##                                          Pr(>|t|)
## (Intercept)                               < 2e-16 ***
## new_neigh_levelnew_neigh_level10          < 2e-16 ***
## new_neigh_levelnew_neigh_level2           0.34798
## new_neigh_levelnew_neigh_level3           < 2e-16 ***
## new_neigh_levelnew_neigh_level4           < 2e-16 ***
## new_neigh_levelnew_neigh_level5           < 2e-16 ***
## new_neigh_levelnew_neigh_level6           < 2e-16 ***
## new_neigh_levelnew_neigh_level7           < 2e-16 ***
## new_neigh_levelnew_neigh_level8           < 2e-16 ***
## new_neigh_levelnew_neigh_level9           < 2e-16 ***
## bldclasscat04 TAX CLASS 1 CONDOS          0.02475 *
## bldclasscat11 SPECIAL CONDO BILLING LOTS  0.12227
## bldclasscat12 CONDOS - WALKUP APARTMENTS  2.12e-12 ***
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS 3.20e-08 ***
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL 0.33294
## I(log(1 + landsqft))                      < 2e-16 ***
## I(log(grosssqft))                         < 2e-16 ***
## localityEastern                           < 2e-16 ***
```

```
## localityNorthern                                < 2e-16 ***
## localityNorthwestern                            < 2e-16 ***
## localitySouthern                               4.44e-09 ***
## localitySouthwestern                            0.03911 *
## yearsl2017                                      < 2e-16 ***
## yearsl2018                                      < 2e-16 ***
## yearsl2019                                      < 2e-16 ***
## yearsl2020                                      < 2e-16 ***
## quartfQ2                                        0.00817 **
## quartfQ3                                       9.66e-05 ***
## quartfQ4                                       1.91e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 206.4 on 13635 degrees of freedom
## Multiple R-squared:  0.6317, Adjusted R-squared:  0.631
## F-statistic: 835.4 on 28 and 13635 DF,  p-value: < 2.2e-16
```

```
summary(log(1+data_mod2$landsqft))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   7.385   7.601   6.723   7.940  11.895
```

```
data_mod3 = data_mod2[data_mod2$landsqft>10,]
trans.lm7new = lm(I(sqrt(price))~new_neigh_level+bldclasscat+(I(log(landsqft))*I(log(grosssqft)))+locali
summary(trans.lm7new)
```

```
##
## Call:
## lm(formula = I(sqrt(price)) ~ new_neigh_level + bldclasscat +
##     (I(log(landsqft)) * I(log(grosssqft))) + locality + yearsl +
##     quartf, data = data_mod3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1591.24   -75.54    11.53    90.17  2622.91
##
## Coefficients:
##                                             Estimate Std. Error t value
## (Intercept)                                 4051.627    340.078  11.914
## new_neigh_levelnew_neigh_level10             593.695     14.403  41.220
## new_neigh_levelnew_neigh_level2               19.553      7.605   2.571
## new_neigh_levelnew_neigh_level3               93.097      9.339   9.968
## new_neigh_levelnew_neigh_level4              121.462     10.945  11.098
## new_neigh_levelnew_neigh_level5              181.249     10.240  17.701
## new_neigh_levelnew_neigh_level6              182.481      9.754  18.708
## new_neigh_levelnew_neigh_level7              341.898     13.649  25.048
## new_neigh_levelnew_neigh_level8              329.015     11.404  28.851
## new_neigh_levelnew_neigh_level9              304.447     18.753  16.235
## bldclasscat04 TAX CLASS 1 CONDOS            -109.225     16.655  -6.558
## bldclasscat11 SPECIAL CONDO BILLING LOTS     -92.970     63.185  -1.471
## bldclasscat12 CONDOS - WALKUP APARTMENTS    -208.792     21.100  -9.896
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS    -88.733     13.263  -6.690
```

```
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL -114.222     12.811  -8.916
## I(log(landsqft))                             -701.651     41.669 -16.839
## I(log(grosssqft))                            -542.274     47.376 -11.446
## localityEastern                              -104.156     10.343 -10.071
## localityNorthern                              135.017     11.943  11.305
## localityNorthwestern                          224.331     12.396  18.098
## localitySouthern                              -55.386      7.813  -7.089
## localitySouthwestern                           25.394      8.693   2.921
## yearsl2017                                     59.319      6.182   9.596
## yearsl2018                                     78.932      6.328  12.474
## yearsl2019                                     79.230      6.362  12.453
## yearsl2020                                    110.317      6.786  16.256
## quartfQ2                                       14.461      5.414   2.671
## quartfQ3                                       24.628      5.426   4.539
## quartfQ4                                       24.130      5.413   4.458
## I(log(landsqft)):I(log(grosssqft))            107.026      5.784  18.504
##                                              Pr(>|t|)
## (Intercept)                                   < 2e-16 ***
## new_neigh_levelnew_neigh_level10              < 2e-16 ***
## new_neigh_levelnew_neigh_level2               0.01015 *
## new_neigh_levelnew_neigh_level3               < 2e-16 ***
## new_neigh_levelnew_neigh_level4               < 2e-16 ***
## new_neigh_levelnew_neigh_level5               < 2e-16 ***
## new_neigh_levelnew_neigh_level6               < 2e-16 ***
## new_neigh_levelnew_neigh_level7               < 2e-16 ***
## new_neigh_levelnew_neigh_level8               < 2e-16 ***
## new_neigh_levelnew_neigh_level9               < 2e-16 ***
## bldclasscat04 TAX CLASS 1 CONDOS             5.68e-11 ***
## bldclasscat11 SPECIAL CONDO BILLING LOTS      0.14121
## bldclasscat12 CONDOS - WALKUP APARTMENTS      < 2e-16 ***
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS   2.33e-11 ***
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL  < 2e-16 ***
## I(log(landsqft))                              < 2e-16 ***
## I(log(grosssqft))                             < 2e-16 ***
## localityEastern                               < 2e-16 ***
## localityNorthern                              < 2e-16 ***
## localityNorthwestern                          < 2e-16 ***
## localitySouthern                             1.43e-12 ***
## localitySouthwestern                          0.00349 **
## yearsl2017                                    < 2e-16 ***
## yearsl2018                                    < 2e-16 ***
## yearsl2019                                    < 2e-16 ***
## yearsl2020                                    < 2e-16 ***
## quartfQ2                                      0.00757 **
## quartfQ3                                     5.71e-06 ***
## quartfQ4                                     8.35e-06 ***
## I(log(landsqft)):I(log(grosssqft))            < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 204.9 on 11619 degrees of freedom
## Multiple R-squared:  0.6576, Adjusted R-squared:  0.6568
## F-statistic: 769.5 on 29 and 11619 DF,  p-value: < 2.2e-16
```

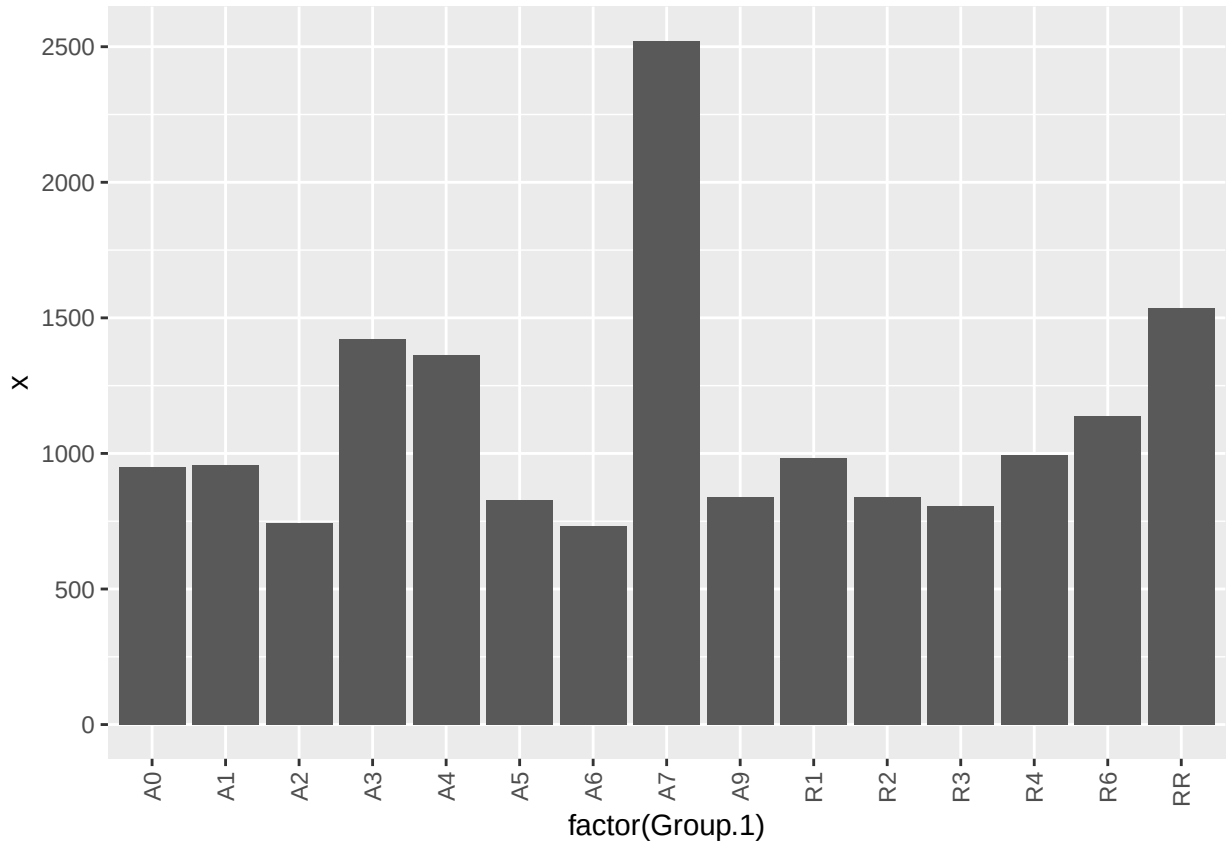
```r
cor(data_mod2$landsqft,data_mod2$grosssqft)
```

```
## [1] 0.01016188
```

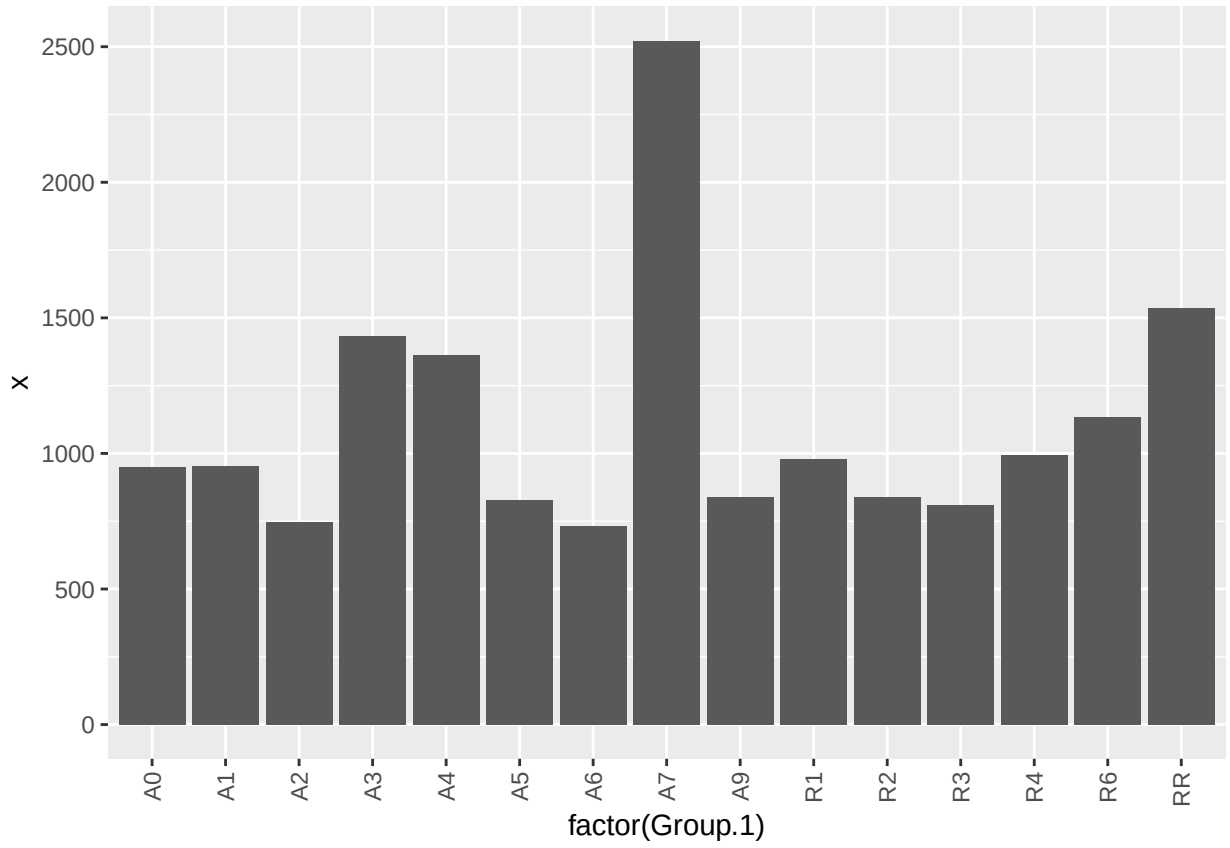
```r
# LANDSQFT Works with a log(x) transformation for predicting sqrt(price), but this is significant only
# That will reduce the rows to 11K, so not making that change
# Sticking to log(1+landsqft) only for now

# ADDING TAX and Building CLASSES TO THE MODEL
summary(lm(I(sqrt(price))~(bldclasscurr), data = data_mod2))
```

```
##
## Call:
## lm(formula = I(sqrt(price)) ~ (bldclasscurr), data = data_mod2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1381.7  -174.5   -37.8   126.6  3099.1
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     949.473    179.900   5.278 1.33e-07 ***
## bldclasscurrA1    5.773    179.994   0.032  0.97442
## bldclasscurrA2 -205.397    180.368  -1.139  0.25482
## bldclasscurrA3  473.492    181.702   2.606  0.00917 **
## bldclasscurrA4  413.254    180.482   2.290  0.02205 *
## bldclasscurrA5 -123.006    179.979  -0.683  0.49433
## bldclasscurrA6 -218.036    359.799  -0.606  0.54453
## bldclasscurrA7 1572.395    202.954   7.748 1.00e-14 ***
## bldclasscurrA9 -109.010    180.004  -0.606  0.54479
## bldclasscurrR1   32.586    180.246   0.181  0.85654
## bldclasscurrR2 -110.105    180.822  -0.609  0.54259
## bldclasscurrR3 -143.538    180.721  -0.794  0.42706
## bldclasscurrR4   43.470    180.027   0.241  0.80920
## bldclasscurrR6  188.583    183.284   1.029  0.30354
## bldclasscurrRR  587.877    202.954   2.897  0.00378 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 311.6 on 13649 degrees of freedom
## Multiple R-squared:   0.16,  Adjusted R-squared:  0.1591
## F-statistic: 185.7 on 14 and 13649 DF,  p-value: < 2.2e-16
```

```r
#Both Bldclasscurr and Bldclasssale explain close to 15% variance in sales prices, including one in the
#Bucketing Bldclasssale as per mean prices to reduce model degree of freedom

bldcurr_price=data.frame(aggregate(sqrt(data_mod2$price), list(data_mod2$bldclasscurr), FUN=mean))
ggplot(bldcurr_price, aes(x=factor(Group.1), y=x)) + geom_bar(stat = "identity") + theme(axis.text.x =
```




```
bldsale_price=data.frame(aggregate(sqrt(data_mod2$price), list(data_mod2$bldclasssale), FUN=mean))
ggplot(bldsale_price, aes(x=factor(Group.1), y=x)) + geom_bar(stat = "identity") + theme(axis.text.x = e
```




```
trans.lm8 = lm(I(sqrt(price))~new_neigh_level+bldclasscat+I(log(1+landsqft))+I(log(grosssqft))+locality+
summary(trans.lm8)
```

```
##
## Call:
## lm(formula = I(sqrt(price)) ~ new_neigh_level + bldclasscat +
##     I(log(1 + landsqft)) + I(log(grosssqft)) + locality + yearsl +
##     quartf + bldclasssale, data = data_mod2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1626.05   -79.77    11.60    91.39  2066.55
##
## Coefficients: (5 not defined because of singularities)
##                                          Estimate Std. Error t value
## (Intercept)                           -1737.4876   123.8992 -14.023
## new_neigh_levelnew_neigh_level10        520.8483    12.7377  40.890
## new_neigh_levelnew_neigh_level2          14.3839     7.6317   1.885
## new_neigh_levelnew_neigh_level3          94.9814     9.1579  10.371
## new_neigh_levelnew_neigh_level4         101.8441    10.2972   9.890
## new_neigh_levelnew_neigh_level5         162.9159     9.7988  16.626
## new_neigh_levelnew_neigh_level6         184.5881     9.3975  19.642
## new_neigh_levelnew_neigh_level7         323.5240    12.0017  26.957
## new_neigh_levelnew_neigh_level8         319.7660    10.6523  30.019
## new_neigh_levelnew_neigh_level9         350.3962    15.5041  22.600
```

```
## bldclasscat04 TAX CLASS 1 CONDOS               -63.4751   119.7854  -0.530
## bldclasscat11 SPECIAL CONDO BILLING LOTS        47.1968   132.7620   0.355
## bldclasscat12 CONDOS - WALKUP APARTMENTS      -167.6309   118.3019  -1.417
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS     -20.5604   117.7251  -0.175
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL   -74.5283   117.8888  -0.632
## I(log(1 + landsqft))                             6.8875     0.7543   9.130
## I(log(grosssqft))                              335.4705     5.1951  64.575
## localityEastern                                -91.2763     9.9741  -9.151
## localityNorthern                               113.4196     9.4439  12.010
## localityNorthwestern                           164.5975     9.7756  16.838
## localitySouthern                               -43.8088     7.4722  -5.863
## localitySouthwestern                            31.3959     8.1350   3.859
## yearsl2017                                      65.1852     6.1291  10.635
## yearsl2018                                      82.8249     6.2720  13.206
## yearsl2019                                      83.2264     6.3085  13.193
## yearsl2020                                     115.7083     6.7249  17.206
## quartfQ2                                        12.7435     4.9636   2.567
## quartfQ3                                        20.1079     4.9687   4.047
## quartfQ4                                        23.4506     4.9578   4.730
## bldclasssaleA1                                 -24.2473   117.4150  -0.207
## bldclasssaleA2                                 -49.7690   117.6709  -0.423
## bldclasssaleA3                                 111.1938   118.5969   0.938
## bldclasssaleA4                                   5.4781   117.8714   0.046
## bldclasssaleA5                                 -89.3512   117.4290  -0.761
## bldclasssaleA6                                  19.9780   234.7457   0.085
## bldclasssaleA7                                 811.5303   132.5898   6.121
## bldclasssaleA9                                 -64.4687   117.4586  -0.549
## bldclasssaleR1                                       NA         NA      NA
## bldclasssaleR2                                       NA         NA      NA
## bldclasssaleR3                                 -33.5530    25.5075  -1.315
## bldclasssaleR4                                       NA         NA      NA
## bldclasssaleR6                                       NA         NA      NA
## bldclasssaleRR                                       NA         NA      NA
##                                              Pr(>|t|)
## (Intercept)                                   < 2e-16 ***
## new_neigh_levelnew_neigh_level10              < 2e-16 ***
## new_neigh_levelnew_neigh_level2              0.059484 .
## new_neigh_levelnew_neigh_level3               < 2e-16 ***
## new_neigh_levelnew_neigh_level4               < 2e-16 ***
## new_neigh_levelnew_neigh_level5               < 2e-16 ***
## new_neigh_levelnew_neigh_level6               < 2e-16 ***
## new_neigh_levelnew_neigh_level7               < 2e-16 ***
## new_neigh_levelnew_neigh_level8               < 2e-16 ***
## new_neigh_levelnew_neigh_level9               < 2e-16 ***
## bldclasscat04 TAX CLASS 1 CONDOS             0.596185
## bldclasscat11 SPECIAL CONDO BILLING LOTS     0.722221
## bldclasscat12 CONDOS - WALKUP APARTMENTS     0.156513
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS   0.861360
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL 0.527272
## I(log(1 + landsqft))                          < 2e-16 ***
## I(log(grosssqft))                             < 2e-16 ***
## localityEastern                               < 2e-16 ***
## localityNorthern                              < 2e-16 ***
## localityNorthwestern                          < 2e-16 ***
```

```
## localitySouthern                              4.65e-09 ***
## localitySouthwestern                          0.000114 ***
## yearsl2017                                     < 2e-16 ***
## yearsl2018                                     < 2e-16 ***
## yearsl2019                                     < 2e-16 ***
## yearsl2020                                     < 2e-16 ***
## quartfQ2                                      0.010257 *
## quartfQ3                                      5.22e-05 ***
## quartfQ4                                      2.27e-06 ***
## bldclasssaleA1                                0.836396
## bldclasssaleA2                                0.672338
## bldclasssaleA3                                0.348478
## bldclasssaleA4                                0.962932
## bldclasssaleA5                                0.446733
## bldclasssaleA6                                0.932179
## bldclasssaleA7                                9.58e-10 ***
## bldclasssaleA9                                0.583108
## bldclasssaleR1                                      NA
## bldclasssaleR2                                      NA
## bldclasssaleR3                                0.188392
## bldclasssaleR4                                      NA
## bldclasssaleR6                                      NA
## bldclasssaleRR                                      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 203.2 on 13626 degrees of freedom
## Multiple R-squared:  0.6434, Adjusted R-squared:  0.6424
## F-statistic: 664.5 on 37 and 13626 DF,  p-value: < 2.2e-16
```

```
bld_sale_Alow = c('A0','A1','A2','A5','A6','A9')
bld_sale_Amed = c('A3','A4','A7')
bld_sale_Rlow = c('R1','R2','R3','R4','R5','R6','RR')
data_mod2$new_bld_sale = as.factor(ifelse(data_mod2$bldclasssale %in% bld_sale_Alow, 'bld_sale_Alow',
                                          ifelse(data_mod2$bldclasssale %in% bld_sale_Amed, 'bld_sal
                                                 ifelse(data_mod2$bldclasssale %in% bld_sale_Rlow, '

trans.lm9 = lm(I(sqrt(price))~new_neigh_level+bldclasscat+I(log(1+landsqft))+I(log(grosssqft))+locality
summary(trans.lm9)
```

```
##
## Call:
## lm(formula = I(sqrt(price)) ~ new_neigh_level + bldclasscat +
##     I(log(1 + landsqft)) + I(log(grosssqft)) + locality + yearsl +
##     quartf + new_bld_sale, data = data_mod2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1668.29   -82.97    10.01    92.16  2711.68
##
## Coefficients: (1 not defined because of singularities)
##                                             Estimate Std. Error t value
## (Intercept)                                -1895.3837    38.0260 -49.844
## new_neigh_levelnew_neigh_level10             521.9183    12.8367  40.658
```

```
## new_neigh_levelnew_neigh_level2                 9.0423     7.5316   1.201
## new_neigh_levelnew_neigh_level3                83.9562     9.1749   9.151
## new_neigh_levelnew_neigh_level4                99.8732    10.3392   9.660
## new_neigh_levelnew_neigh_level5               161.0735     9.8491  16.354
## new_neigh_levelnew_neigh_level6               192.8186     9.4462  20.412
## new_neigh_levelnew_neigh_level7               329.0375    12.0900  27.216
## new_neigh_levelnew_neigh_level8               322.8584    10.6832  30.221
## new_neigh_levelnew_neigh_level9               351.4788    15.6356  22.479
## bldclasscat04 TAX CLASS 1 CONDOS              -19.2063    11.6391  -1.650
## bldclasscat11 SPECIAL CONDO BILLING LOTS      118.4346    62.6410   1.891
## bldclasscat12 CONDOS - WALKUP APARTMENTS      -89.2760    14.1719  -6.300
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS     52.4764     8.1792   6.416
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL    1.1697    10.2247   0.114
## I(log(1 + landsqft))                            7.5532     0.7613   9.922
## I(log(grosssqft))                             346.8160     5.0848  68.206
## localityEastern                               -85.6525     9.9827  -8.580
## localityNorthern                              112.0648     9.5329  11.756
## localityNorthwestern                          161.0006     9.8735  16.306
## localitySouthern                              -24.3301     7.3552  -3.308
## localitySouthwestern                           33.3043     8.1077   4.108
## yearsl2017                                     63.7676     6.1968  10.290
## yearsl2018                                     83.3376     6.3419  13.141
## yearsl2019                                     83.1194     6.3784  13.031
## yearsl2020                                    114.5241     6.8008  16.840
## quartfQ2                                       12.8013     5.0192   2.550
## quartfQ3                                       19.8448     5.0248   3.949
## quartfQ4                                       23.8550     5.0134   4.758
## new_bld_salebld_sale_Amed                     104.0851     9.4791  10.981
## new_bld_salebld_sale_Rlow                           NA         NA      NA
##                                              Pr(>|t|)
## (Intercept)                                   < 2e-16 ***
## new_neigh_levelnew_neigh_level10              < 2e-16 ***
## new_neigh_levelnew_neigh_level2              0.229936
## new_neigh_levelnew_neigh_level3               < 2e-16 ***
## new_neigh_levelnew_neigh_level4               < 2e-16 ***
## new_neigh_levelnew_neigh_level5               < 2e-16 ***
## new_neigh_levelnew_neigh_level6               < 2e-16 ***
## new_neigh_levelnew_neigh_level7               < 2e-16 ***
## new_neigh_levelnew_neigh_level8               < 2e-16 ***
## new_neigh_levelnew_neigh_level9               < 2e-16 ***
## bldclasscat04 TAX CLASS 1 CONDOS             0.098936 .
## bldclasscat11 SPECIAL CONDO BILLING LOTS     0.058687 .
## bldclasscat12 CONDOS - WALKUP APARTMENTS     3.08e-10 ***
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS   1.45e-10 ***
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL 0.908926
## I(log(1 + landsqft))                          < 2e-16 ***
## I(log(grosssqft))                             < 2e-16 ***
## localityEastern                               < 2e-16 ***
## localityNorthern                              < 2e-16 ***
## localityNorthwestern                          < 2e-16 ***
## localitySouthern                             0.000943 ***
## localitySouthwestern                         4.02e-05 ***
## yearsl2017                                    < 2e-16 ***
## yearsl2018                                    < 2e-16 ***
```

```
## yearsl2019                                    < 2e-16 ***
## yearsl2020                                    < 2e-16 ***
## quartfQ2                                     0.010768 *
## quartfQ3                                     7.88e-05 ***
## quartfQ4                                     1.97e-06 ***
## new_bld_salebld_sale_Amed                     < 2e-16 ***
## new_bld_salebld_sale_Rlow                          NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 205.5 on 13634 degrees of freedom
## Multiple R-squared:  0.635,  Adjusted R-squared:  0.6342
## F-statistic: 817.8 on 29 and 13634 DF,  p-value: < 2.2e-16
```

```
calc.relimp(trans.lm9)
```

```
## Response variable: I(sqrt(price))
## Total response variance: 115465.3
## Analysis based on 13664 observations
##
## 30 Regressors:
## Some regressors combined in groups:
##         Group  new_neigh_level : new_neigh_levelnew_neigh_level10 new_neigh_levelnew_neigh_level2 ne
##         Group  bldclasscat : bldclasscat04 TAX CLASS 1 CONDOS bldclasscat11 SPECIAL CONDO BILLING LO
##         Group  locality : localityEastern localityNorthern localityNorthwestern localitySouthern loc
##         Group  yearsl : yearsl2017 yearsl2018 yearsl2019 yearsl2020
##         Group  quartf : quartfQ2 quartfQ3 quartfQ4
##         Group  new_bld_sale : new_bld_salebld_sale_Amed new_bld_salebld_sale_Rlow
##
##  Relative importance of 8 (groups of) regressors assessed:
##  new_neigh_level bldclasscat locality yearsl quartf new_bld_sale I(log(1 + landsqft)) I(log(grosssq
##
## Proportion of variance explained by model: 63.5%
## Metrics are not normalized (rela=FALSE).
##
## Relative importance metrics:
##
##                               lmg
## new_neigh_level      0.2295195115
## bldclasscat          0.0165758675
## locality             0.1330783955
## yearsl               0.0115736870
## quartf               0.0005955513
## new_bld_sale         0.0468671164
## I(log(1 + landsqft)) 0.0055025062
## I(log(grosssqft))    0.1912532160
##
## Average coefficients for different model sizes:
##
##                                              1group    2groups     3groups
## new_neigh_levelnew_neigh_level10           837.024537  832.03864  790.053707
## new_neigh_levelnew_neigh_level2             74.787550   66.33138   57.476546
## new_neigh_levelnew_neigh_level3            159.702316  155.98692  148.369610
## new_neigh_levelnew_neigh_level4            217.857133  233.83816  223.793524
```

```
## new_neigh_levelnew_neigh_level5               303.797803  302.25419  287.479783
## new_neigh_levelnew_neigh_level6               380.713470  365.75719  341.477934
## new_neigh_levelnew_neigh_level7               478.587730  502.42153  489.133913
## new_neigh_levelnew_neigh_level8               510.085166  513.02042  489.965399
## new_neigh_levelnew_neigh_level9               645.746257  658.92460  624.825571
## bldclasscat04 TAX CLASS 1 CONDOS              -24.112962  -54.57369  -90.338247
## bldclasscat11 SPECIAL CONDO BILLING LOTS      640.085121  472.81492  328.496494
## bldclasscat12 CONDOS - WALKUP APARTMENTS      -57.441759 -113.74907 -167.332713
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS     95.682846   44.01306   -7.697419
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL   83.513177   20.40009  -39.709117
## I(log(1 + landsqft))                            2.934897    8.78690   10.203621
## I(log(grosssqft))                             384.078228  426.36172  423.720285
## localityEastern                              -329.225128 -295.51312 -251.192457
## localityNorthern                              41.470690   92.26818  115.459089
## localityNorthwestern                         316.823622  319.25818  302.857523
## localitySouthern                             -111.537925 -106.40504  -91.661140
## localitySouthwestern                          -22.783328  -17.87358   -9.616491
## yearsl2017                                    45.712584   49.92126   53.685045
## yearsl2018                                    94.771539   70.39261   65.786543
## yearsl2019                                    57.246769   63.44331   68.950291
## yearsl2020                                   107.120691  108.17693  109.375238
## quartfQ2                                      14.737042   12.64517   11.915483
## quartfQ3                                      19.215603   18.53365   18.253552
## quartfQ4                                      11.866879   14.77078   17.601568
## new_bld_salebld_sale_Amed                    535.157008  430.41275  343.551515
## new_bld_salebld_sale_Rlow                    101.584360   79.42756   80.917009
##                                                   4groups     5groups
## new_neigh_levelnew_neigh_level10             736.24818104  680.675356
## new_neigh_levelnew_neigh_level2               48.42285265   39.129733
## new_neigh_levelnew_neigh_level3              138.04718053  125.865657
## new_neigh_levelnew_neigh_level4              203.20053480  178.789262
## new_neigh_levelnew_neigh_level5              266.34416216  242.055815
## new_neigh_levelnew_neigh_level6              313.63672507  284.453830
## new_neigh_levelnew_neigh_level7              461.78278098  430.041390
## new_neigh_levelnew_neigh_level8              457.97608265  423.969646
## new_neigh_levelnew_neigh_level9              574.02541692  518.872605
## bldclasscat04 TAX CLASS 1 CONDOS            -126.89441317 -162.166674
## bldclasscat11 SPECIAL CONDO BILLING LOTS     206.12182010  103.169348
## bldclasscat12 CONDOS - WALKUP APARTMENTS    -214.44614138 -253.719051
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS   -55.25719613  -96.892516
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL -93.38230664 -139.495299
## I(log(1 + landsqft))                           9.80877474    9.004759
## I(log(grosssqft))                            406.00025287  386.307001
## localityEastern                             -206.40552387 -166.255567
## localityNorthern                             123.67088008  124.312049
## localityNorthwestern                         276.53503154  246.413574
## localitySouthern                             -74.09899035  -57.410670
## localitySouthwestern                          -0.06057071    9.540165
## yearsl2017                                    56.65293798   58.912570
## yearsl2018                                    69.69316325   75.257905
## yearsl2019                                    73.29447238   76.586036
## yearsl2020                                   110.46050010  111.443852
## quartfQ2                                      11.79893814   11.933662
## quartfQ3                                      18.20468613   18.351575
```

```
## quartfQ4                                          19.65534944   21.144402
## new_bld_salebld_sale_Amed                        272.57692667  215.389472
## new_bld_salebld_sale_Rlow                         99.15472228  128.638543
##                                                     6groups    7groups     8groups
## new_neigh_levelnew_neigh_level10                 626.220896  573.39808  521.918340
## new_neigh_levelnew_neigh_level2                   29.498249   19.47053    9.042264
## new_neigh_levelnew_neigh_level3                  112.456041   98.33279   83.956167
## new_neigh_levelnew_neigh_level4                  152.976367  126.59344   99.873155
## new_neigh_levelnew_neigh_level5                  215.993566  188.85041  161.073537
## new_neigh_levelnew_neigh_level6                  254.527363  223.97980  192.818612
## new_neigh_levelnew_neigh_level7                  397.008583  363.39378  329.037534
## new_neigh_levelnew_neigh_level8                  389.978795  356.35100  322.858441
## new_neigh_levelnew_neigh_level9                  463.048940  407.33021  351.478767
## bldclasscat04 TAX CLASS 1 CONDOS                -196.044202 -229.90489 -265.969370
## bldclasscat11 SPECIAL CONDO BILLING LOTS          16.019069  -59.65354 -128.328500
## bldclasscat12 CONDOS - WALKUP APARTMENTS        -285.671598 -312.22448 -336.039066
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS      -132.756943 -164.43681 -194.286636
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL    -178.752837 -213.20505 -245.593416
## I(log(1 + landsqft))                              8.350541    7.90743    7.553226
## I(log(grosssqft))                               369.240685  356.07552  346.816021
## localityEastern                                 -132.720338 -106.05332  -85.652495
## localityNorthern                                 121.396128  116.96052  112.064842
## localityNorthwestern                             216.141969  187.45906  161.000618
## localitySouthern                                 -43.289431  -32.27469  -24.330143
## localitySouthwestern                              18.497991   26.47857   33.304278
## yearsl2017                                        60.713267   62.28602   63.767586
## yearsl2018                                        79.268001   81.38185   83.337592
## yearsl2019                                        79.144914   81.26518   83.119351
## yearsl2020                                       112.397321  113.39259  114.524068
## quartfQ2                                          12.144090   12.40992   12.801322
## quartfQ3                                          18.686690   19.19132   19.844795
## quartfQ4                                          22.279310   23.16665   23.855013
## new_bld_salebld_sale_Amed                        169.790072  133.48392  104.085053
## new_bld_salebld_sale_Rlow                        165.091755  205.29723  246.763081
```

```
sqrt(sum(((trans.lm9$fitted.values)^2 - data_mod2$price)^2)/length(trans.lm9$fitted.values))
```

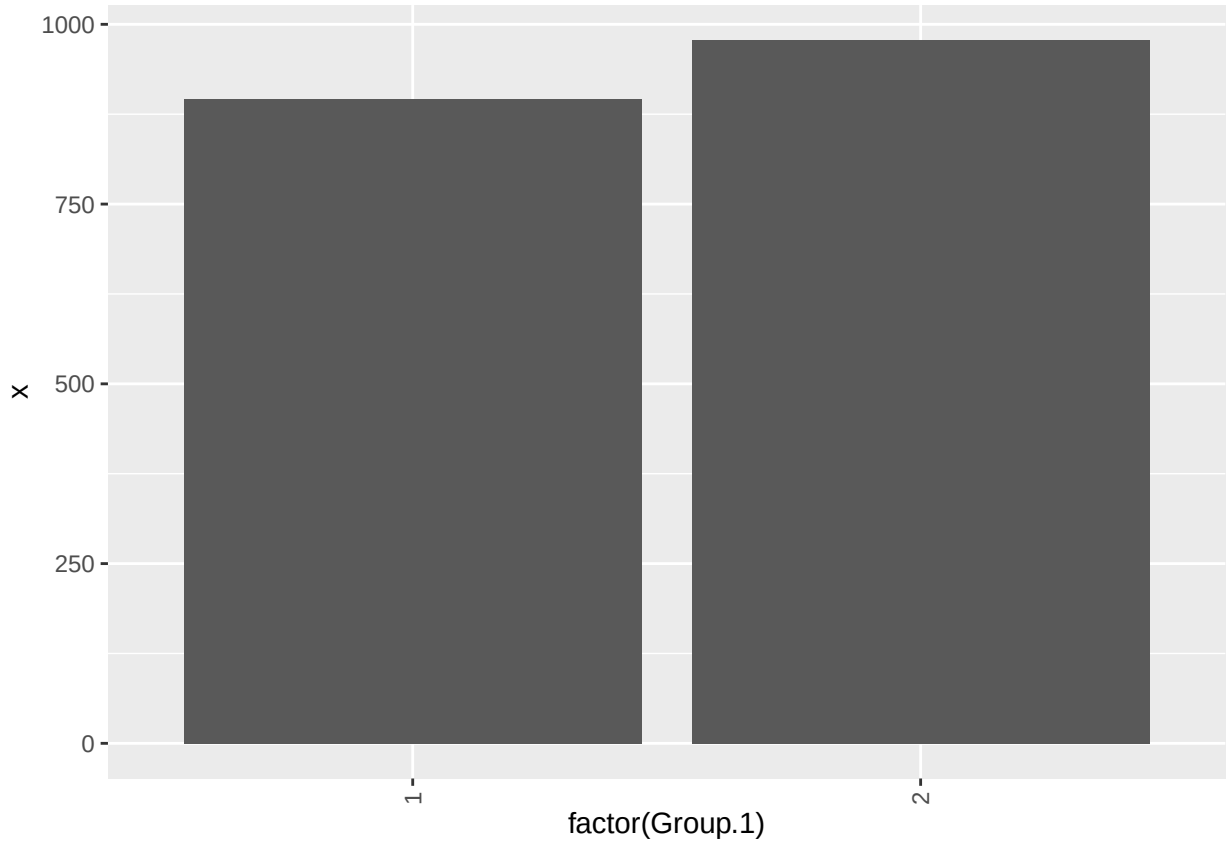
```
## [1] 554942
```

```
## No prediction power coming from tax sale category, not included in the model
summary(lm(I(sqrt(price))~(taxclasscurr*taxclasssale), data = data_mod2))
```

```
##
## Call:
## lm(formula = I(sqrt(price)) ~ (taxclasscurr * taxclasssale),
##     data = data_mod2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -966.9 -193.7  -60.6  102.2 4152.5
##
## Coefficients: (2 not defined because of singularities)
##                             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                  897.265      3.422  262.18   <2e-16 ***
## taxclasscurr2                 79.616      7.773   10.24   <2e-16 ***
## taxclasssale2                     NA         NA      NA       NA
## taxclasscurr2:taxclasssale2       NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 343.3 on 12477 degrees of freedom
##   (1185 observations deleted due to missingness)
## Multiple R-squared:  0.008338,   Adjusted R-squared:  0.008258
## F-statistic: 104.9 on 1 and 12477 DF,  p-value: < 2.2e-16
```

```
tax_sale_price=data.frame(aggregate(sqrt(data_mod2$price), list(data_mod2$taxclasssale), FUN=mean))
ggplot(tax_sale_price, aes(x=factor(Group.1), y=x)) + geom_bar(stat = "identity") + theme(axis.text.x =
```




```
trans.lm10 = lm(I(sqrt(price))~new_neigh_level+bldclasscat+I(log(1+landsqft))+I(log(grosssqft))+locality
summary(trans.lm10)
```

```
##
## Call:
## lm(formula = I(sqrt(price)) ~ new_neigh_level + bldclasscat +
##     I(log(1 + landsqft)) + I(log(grosssqft)) + locality + yearsl +
##     quartf + new_bld_sale + taxclasssale, data = data_mod2)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -1668.29   -82.97    10.01    92.16  2711.68
##
## Coefficients: (2 not defined because of singularities)
##                                            Estimate Std. Error t value
## (Intercept)                               -1895.3837    38.0260 -49.844
## new_neigh_levelnew_neigh_level10            521.9183    12.8367  40.658
## new_neigh_levelnew_neigh_level2               9.0423     7.5316   1.201
## new_neigh_levelnew_neigh_level3              83.9562     9.1749   9.151
## new_neigh_levelnew_neigh_level4              99.8732    10.3392   9.660
## new_neigh_levelnew_neigh_level5             161.0735     9.8491  16.354
## new_neigh_levelnew_neigh_level6             192.8186     9.4462  20.412
## new_neigh_levelnew_neigh_level7             329.0375    12.0900  27.216
## new_neigh_levelnew_neigh_level8             322.8584    10.6832  30.221
## new_neigh_levelnew_neigh_level9             351.4788    15.6356  22.479
## bldclasscat04 TAX CLASS 1 CONDOS            -19.2063    11.6391  -1.650
## bldclasscat11 SPECIAL CONDO BILLING LOTS    118.4346    62.6410   1.891
## bldclasscat12 CONDOS - WALKUP APARTMENTS    -89.2760    14.1719  -6.300
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS   52.4764     8.1792   6.416
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL  1.1697    10.2247   0.114
## I(log(1 + landsqft))                          7.5532     0.7613   9.922
## I(log(grosssqft))                           346.8160     5.0848  68.206
## localityEastern                             -85.6525     9.9827  -8.580
## localityNorthern                            112.0648     9.5329  11.756
## localityNorthwestern                        161.0006     9.8735  16.306
## localitySouthern                            -24.3301     7.3552  -3.308
## localitySouthwestern                         33.3043     8.1077   4.108
## yearsl2017                                   63.7676     6.1968  10.290
## yearsl2018                                   83.3376     6.3419  13.141
## yearsl2019                                   83.1194     6.3784  13.031
## yearsl2020                                  114.5241     6.8008  16.840
## quartfQ2                                     12.8013     5.0192   2.550
## quartfQ3                                     19.8448     5.0248   3.949
## quartfQ4                                     23.8550     5.0134   4.758
## new_bld_salebld_sale_Amed                   104.0851     9.4791  10.981
## new_bld_salebld_sale_Rlow                         NA         NA      NA
## taxclasssale2                                     NA         NA      NA
##                                           Pr(>|t|)
## (Intercept)                                < 2e-16 ***
## new_neigh_levelnew_neigh_level10           < 2e-16 ***
## new_neigh_levelnew_neigh_level2           0.229936
## new_neigh_levelnew_neigh_level3            < 2e-16 ***
## new_neigh_levelnew_neigh_level4            < 2e-16 ***
## new_neigh_levelnew_neigh_level5            < 2e-16 ***
## new_neigh_levelnew_neigh_level6            < 2e-16 ***
## new_neigh_levelnew_neigh_level7            < 2e-16 ***
## new_neigh_levelnew_neigh_level8            < 2e-16 ***
## new_neigh_levelnew_neigh_level9            < 2e-16 ***
## bldclasscat04 TAX CLASS 1 CONDOS          0.098936 .
## bldclasscat11 SPECIAL CONDO BILLING LOTS  0.058687 .
## bldclasscat12 CONDOS - WALKUP APARTMENTS  3.08e-10 ***
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS 1.45e-10 ***
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL 0.908926
## I(log(1 + landsqft))                       < 2e-16 ***
```

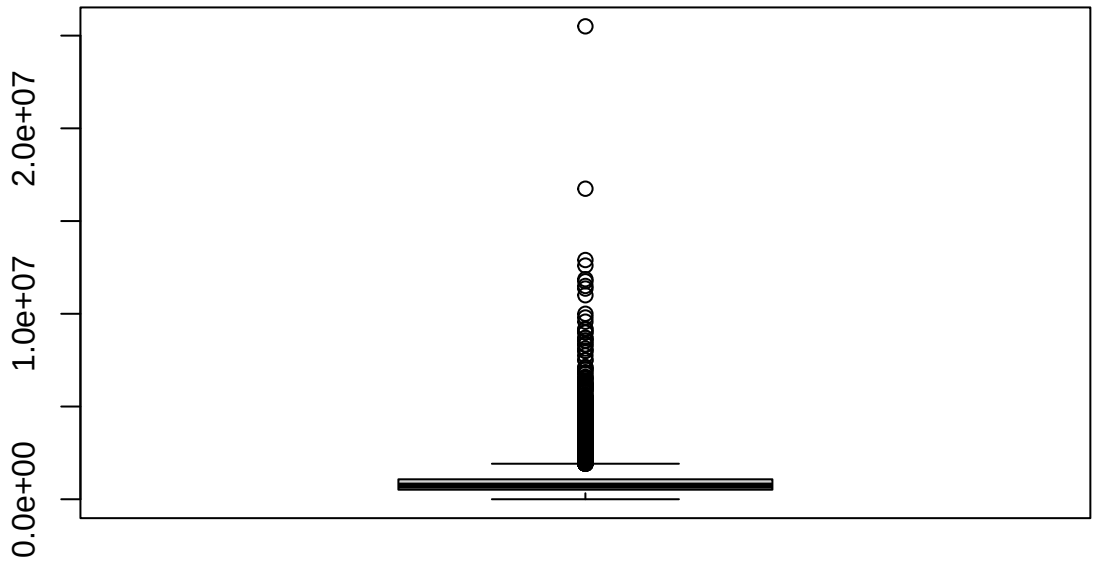
```
## I(log(grosssqft))                              < 2e-16 ***
## localityEastern                                < 2e-16 ***
## localityNorthern                               < 2e-16 ***
## localityNorthwestern                           < 2e-16 ***
## localitySouthern                              0.000943 ***
## localitySouthwestern                          4.02e-05 ***
## yearsl2017                                     < 2e-16 ***
## yearsl2018                                     < 2e-16 ***
## yearsl2019                                     < 2e-16 ***
## yearsl2020                                     < 2e-16 ***
## quartfQ2                                      0.010768 *
## quartfQ3                                      7.88e-05 ***
## quartfQ4                                      1.97e-06 ***
## new_bld_salebld_sale_Amed                      < 2e-16 ***
## new_bld_salebld_sale_Rlow                           NA
## taxclasssale2                                       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 205.5 on 13634 degrees of freedom
## Multiple R-squared:  0.635,  Adjusted R-squared:  0.6342
## F-statistic: 817.8 on 29 and 13634 DF,  p-value: < 2.2e-16
```

4.0 Final Model

```
##  Reducing Modelling data to exclude price outliers
summary(data_mod2$price)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##       20   515000   735000   953392  1075000 25500000
```

```
boxplot(data_mod2$price, data=data_mod2)
```




```
data_mod4 = data_mod2[data_mod2$price>=100000 & data_mod2$price<=7000000,]
dim(data_mod4)
```

```
## [1] 13437    20
```

```
trans.lm12 = lm(I(sqrt(price))~new_neigh_level+bldclasscat+I(log(1+landsqft))+I(log(grosssqft))+locality
summary(trans.lm12)
```

```
##
## Call:
## lm(formula = I(sqrt(price)) ~ new_neigh_level + bldclasscat +
##     I(log(1 + landsqft)) + I(log(grosssqft)) + locality + quarter,
##     data = data_mod4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1150.24   -86.32     2.06    81.56  1364.40
##
## Coefficients:
##                                          Estimate Std. Error t value
## (Intercept)                             -1812.8608    33.6540 -53.868
## new_neigh_levelnew_neigh_level10          485.1383    11.2338  43.186
## new_neigh_levelnew_neigh_level2            12.6889     6.6428   1.910
## new_neigh_levelnew_neigh_level3            88.8610     8.0540  11.033
## new_neigh_levelnew_neigh_level4           110.9818     9.0670  12.240
```

```
## new_neigh_levelnew_neigh_level5                170.2073     8.6285  19.726
## new_neigh_levelnew_neigh_level6                198.8590     8.2856  24.000
## new_neigh_levelnew_neigh_level7                317.8493    10.5647  30.086
## new_neigh_levelnew_neigh_level8                330.0290     9.3619  35.252
## new_neigh_levelnew_neigh_level9                350.9053    13.6734  25.663
## bldclasscat04 TAX CLASS 1 CONDOS               -30.4129    10.1441  -2.998
## bldclasscat11 SPECIAL CONDO BILLING LOTS       110.5011    54.5558   2.025
## bldclasscat12 CONDOS - WALKUP APARTMENTS      -110.8508    12.3351  -8.987
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS      29.5992     7.1404   4.145
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL   -22.9537     8.8796  -2.585
## I(log(1 + landsqft))                             6.8646     0.6642  10.335
## I(log(grosssqft))                              341.3813     4.3550  78.388
## localityEastern                               -104.3058     8.6683 -12.033
## localityNorthern                               103.2527     8.3350  12.388
## localityNorthwestern                           149.4454     8.6339  17.309
## localitySouthern                               -58.4784     6.2914  -9.295
## localitySouthwestern                            -3.2216     6.9676  -0.462
## quarter2016_Q2                                  11.4617    11.1572   1.027
## quarter2016_Q3                                  17.3430    10.8051   1.605
## quarter2016_Q4                                  41.1893    11.1937   3.680
## quarter2017_Q1                                  58.0006    11.0769   5.236
## quarter2017_Q2                                  66.4378    10.8761   6.109
## quarter2017_Q3                                  89.3882    10.9549   8.160
## quarter2017_Q4                                  95.9340    11.1067   8.638
## quarter2018_Q1                                  87.7443     9.8349   8.922
## quarter2018_Q2                                  97.6932     9.7747   9.995
## quarter2018_Q3                                  97.6467     9.7912   9.973
## quarter2018_Q4                                 100.4076     9.8467  10.197
## quarter2019_Q1                                  90.8988    11.7884   7.711
## quarter2019_Q2                                 119.3158    11.1606  10.691
## quarter2019_Q3                                 103.7956    11.1394   9.318
## quarter2019_Q4                                 101.2500    11.3114   8.951
## quarter2020_Q1                                 133.5932    11.7331  11.386
## quarter2020_Q2                                 115.2473    13.2202   8.718
## quarter2020_Q3                                 106.9237    12.7731   8.371
## quarter2020_Q4                                 144.0320    11.0344  13.053
##                                               Pr(>|t|)
## (Intercept)                                    < 2e-16 ***
## new_neigh_levelnew_neigh_level10               < 2e-16 ***
## new_neigh_levelnew_neigh_level2               0.056134 .
## new_neigh_levelnew_neigh_level3                < 2e-16 ***
## new_neigh_levelnew_neigh_level4                < 2e-16 ***
## new_neigh_levelnew_neigh_level5                < 2e-16 ***
## new_neigh_levelnew_neigh_level6                < 2e-16 ***
## new_neigh_levelnew_neigh_level7                < 2e-16 ***
## new_neigh_levelnew_neigh_level8                < 2e-16 ***
## new_neigh_levelnew_neigh_level9                < 2e-16 ***
## bldclasscat04 TAX CLASS 1 CONDOS              0.002722 **
## bldclasscat11 SPECIAL CONDO BILLING LOTS      0.042839 *
## bldclasscat12 CONDOS - WALKUP APARTMENTS       < 2e-16 ***
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS    3.41e-05 ***
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL  0.009748 **
## I(log(1 + landsqft))                           < 2e-16 ***
## I(log(grosssqft))                              < 2e-16 ***
```

```
## localityEastern                                < 2e-16 ***
## localityNorthern                               < 2e-16 ***
## localityNorthwestern                           < 2e-16 ***
## localitySouthern                               < 2e-16 ***
## localitySouthwestern                          0.643823
## quarter2016_Q2                                0.304298
## quarter2016_Q3                                0.108501
## quarter2016_Q4                                0.000234 ***
## quarter2017_Q1                                1.66e-07 ***
## quarter2017_Q2                                1.03e-09 ***
## quarter2017_Q3                                3.66e-16 ***
## quarter2017_Q4                                 < 2e-16 ***
## quarter2018_Q1                                 < 2e-16 ***
## quarter2018_Q2                                 < 2e-16 ***
## quarter2018_Q3                                 < 2e-16 ***
## quarter2018_Q4                                 < 2e-16 ***
## quarter2019_Q1                                1.34e-14 ***
## quarter2019_Q2                                 < 2e-16 ***
## quarter2019_Q3                                 < 2e-16 ***
## quarter2019_Q4                                 < 2e-16 ***
## quarter2020_Q1                                 < 2e-16 ***
## quarter2020_Q2                                 < 2e-16 ***
## quarter2020_Q3                                 < 2e-16 ***
## quarter2020_Q4                                 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 178.3 on 13396 degrees of freedom
## Multiple R-squared:  0.6718, Adjusted R-squared:  0.6708
## F-statistic: 685.6 on 40 and 13396 DF,  p-value: < 2.2e-16
```

```
sqrt(sum(((trans.lm12$fitted.values)^2 - data_mod4$price)^2)/length(trans.lm12$fitted.values))
```

```
## [1] 441420.2
```

```
calc.relimp(trans.lm12)
```

```
## Response variable: I(sqrt(price))
## Total response variance: 96602.56
## Analysis based on 13437 observations
##
## 40 Regressors:
## Some regressors combined in groups:
##         Group  new_neigh_level : new_neigh_levelnew_neigh_level10 new_neigh_levelnew_neigh_level2 new
##         Group  bldclasscat : bldclasscat04 TAX CLASS 1 CONDOS bldclasscat11 SPECIAL CONDO BILLING LOT
##         Group  locality : localityEastern localityNorthern localityNorthwestern localitySouthern loca
##         Group  quarter : quarter2016_Q2 quarter2016_Q3 quarter2016_Q4 quarter2017_Q1 quarter2017_Q2 q
##
##  Relative importance of 6 (groups of) regressors assessed:
##  new_neigh_level bldclasscat locality quarter I(log(1 + landsqft)) I(log(grosssqft))
##
## Proportion of variance explained by model: 67.18%
## Metrics are not normalized (rela=FALSE).
```

```
##
## Relative importance metrics:
##
##                             lmg
## new_neigh_level      0.254535733
## bldclasscat          0.024796734
## locality             0.154925444
## quarter              0.015938980
## I(log(1 + landsqft)) 0.006560037
## I(log(grosssqft))    0.215057577
##
## Average coefficients for different model sizes:
##
##                                                 1group     2groups
## new_neigh_levelnew_neigh_level10             758.515995  753.735502
## new_neigh_levelnew_neigh_level2               72.022421   61.954610
## new_neigh_levelnew_neigh_level3              149.647599  144.199062
## new_neigh_levelnew_neigh_level4              212.955884  229.091537
## new_neigh_levelnew_neigh_level5              292.566625  289.235276
## new_neigh_levelnew_neigh_level6              369.844224  351.202608
## new_neigh_levelnew_neigh_level7              462.461749  483.210684
## new_neigh_levelnew_neigh_level8              500.384851  502.412641
## new_neigh_levelnew_neigh_level9              636.600443  641.320444
## bldclasscat04 TAX CLASS 1 CONDOS             -24.807089  -37.667160
## bldclasscat11 SPECIAL CONDO BILLING LOTS     631.753293  433.421220
## bldclasscat12 CONDOS - WALKUP APARTMENTS     -60.688304 -108.896274
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS    86.323789   44.255767
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL  76.162142   18.513157
## I(log(1 + landsqft))                           2.644181    8.446142
## I(log(grosssqft))                            360.023207  402.796062
## localityEastern                             -327.711556 -285.684706
## localityNorthern                              34.173036   83.047652
## localityNorthwestern                         277.760563  270.987293
## localitySouthern                            -126.617024 -123.996102
## localitySouthwestern                         -38.480240  -39.511401
## quarter2016_Q2                               -10.274982   -3.429257
## quarter2016_Q3                                 4.771767    7.924844
## quarter2016_Q4                                 8.522494   17.558575
## quarter2017_Q1                                39.469044   45.798027
## quarter2017_Q2                                42.133787   48.221381
## quarter2017_Q3                                46.224444   58.916897
## quarter2017_Q4                                35.673869   53.954257
## quarter2018_Q1                                83.820130   61.783355
## quarter2018_Q2                                95.853371   72.522246
## quarter2018_Q3                                96.363186   72.325907
## quarter2018_Q4                                71.118834   57.321930
## quarter2019_Q1                                38.200667   54.492835
## quarter2019_Q2                                85.488634   95.325449
## quarter2019_Q3                                58.413384   73.125690
## quarter2019_Q4                                56.880650   70.068275
## quarter2020_Q1                                72.640127   90.342144
## quarter2020_Q2                                72.418263   85.330672
## quarter2020_Q3                                98.521319  100.256223
## quarter2020_Q4                               144.914052  141.380433
```

```
##                                              3groups     4groups
## new_neigh_levelnew_neigh_level10             698.321040  629.797461
## new_neigh_levelnew_neigh_level2               50.727669   38.859345
## new_neigh_levelnew_neigh_level3              134.816329  122.278262
## new_neigh_levelnew_neigh_level4              211.900236  183.106794
## new_neigh_levelnew_neigh_level5              269.269092  241.489212
## new_neigh_levelnew_neigh_level6              319.114297  282.083196
## new_neigh_levelnew_neigh_level7              455.940146  414.061751
## new_neigh_levelnew_neigh_level8              469.019267  425.449564
## new_neigh_levelnew_neigh_level9              584.084950  510.169390
## bldclasscat04 TAX CLASS 1 CONDOS             -46.500395  -48.986198
## bldclasscat11 SPECIAL CONDO BILLING LOTS     292.101196  197.013978
## bldclasscat12 CONDOS - WALKUP APARTMENTS    -139.756287 -150.948157
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS    15.497160    2.720474
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL -20.906145  -40.869762
## I(log(1 + landsqft))                           9.255983    8.277730
## I(log(grosssqft))                            392.160123  370.900183
## localityEastern                             -232.319326 -180.590339
## localityNorthern                             103.490897  109.460597
## localityNorthwestern                         247.466250  215.346537
## localitySouthern                            -109.298733  -90.896232
## localitySouthwestern                         -34.539915  -25.972548
## quarter2016_Q2                                 2.476128    6.753765
## quarter2016_Q3                                10.890183   13.402421
## quarter2016_Q4                                25.938380   32.437689
## quarter2017_Q1                                50.714223   54.045230
## quarter2017_Q2                                53.935005   58.493723
## quarter2017_Q3                                70.165519   78.455029
## quarter2017_Q4                                70.123825   81.792127
## quarter2018_Q1                                62.532912   72.934893
## quarter2018_Q2                                72.539618   82.868333
## quarter2018_Q3                                72.005843   81.886303
## quarter2018_Q4                                65.425259   80.819884
## quarter2019_Q1                                68.425562   78.396254
## quarter2019_Q2                               103.743887  109.975473
## quarter2019_Q3                                85.574650   94.231898
## quarter2019_Q4                                81.796075   90.395326
## quarter2020_Q1                               106.162798  118.060579
## quarter2020_Q2                                96.572585  104.713961
## quarter2020_Q3                               102.072784  103.684632
## quarter2020_Q4                               140.226211  140.578042
##                                              5groups     6groups
## new_neigh_levelnew_neigh_level10             557.711570  485.138280
## new_neigh_levelnew_neigh_level2               26.219203   12.688873
## new_neigh_levelnew_neigh_level3              106.886273   88.861001
## new_neigh_levelnew_neigh_level4              148.754387  110.981770
## new_neigh_levelnew_neigh_level5              208.222765  170.207342
## new_neigh_levelnew_neigh_level6              241.858125  198.858959
## new_neigh_levelnew_neigh_level7              366.934795  317.849343
## new_neigh_levelnew_neigh_level8              378.383538  330.028996
## new_neigh_levelnew_neigh_level9              431.171126  350.905296
## bldclasscat04 TAX CLASS 1 CONDOS             -43.750683  -30.412949
## bldclasscat11 SPECIAL CONDO BILLING LOTS     139.011203  110.501072
## bldclasscat12 CONDOS - WALKUP APARTMENTS    -141.292833 -110.850840
```

```
## bldclasscat13 CONDOS - ELEVATOR APARTMENTS      7.383159   29.599189
## bldclasscat15 CONDOS - 2-10 UNIT RESIDENTIAL  -41.211316  -22.953710
## I(log(1 + landsqft))                            7.293073    6.864598
## I(log(grosssqft))                             352.389083  341.381328
## localityEastern                              -136.942361 -104.305822
## localityNorthern                              108.275854  103.252730
## localityNorthwestern                          181.423114  149.445447
## localitySouthern                              -73.243178  -58.478434
## localitySouthwestern                          -15.308019   -3.221614
## quarter2016_Q2                                  9.578756   11.461747
## quarter2016_Q3                                 15.497584   17.342958
## quarter2016_Q4                                 37.280736   41.189345
## quarter2017_Q1                                 56.224844   58.000565
## quarter2017_Q2                                 62.379359   66.437801
## quarter2017_Q3                                 84.411844   89.388239
## quarter2017_Q4                                 89.776688   95.934017
## quarter2018_Q1                                 83.451315   87.744317
## quarter2018_Q2                                 93.591783   97.693161
## quarter2018_Q3                                 92.471520   97.646735
## quarter2018_Q4                                 94.222148  100.407602
## quarter2019_Q1                                 85.312976   90.898842
## quarter2019_Q2                                114.753265  119.315797
## quarter2019_Q3                                 99.798238  103.795557
## quarter2019_Q4                                 96.460493  101.249994
## quarter2020_Q1                                126.719638  133.593246
## quarter2020_Q2                                110.457221  115.247287
## quarter2020_Q3                                105.203106  106.923727
## quarter2020_Q4                                141.963288  144.032031
```