Waltham Real Estate Pricing Project

Hieu Nguyen

4/24/2021

Project Goals

1. Produce actionable visualization for a real estate developer on the housing market trends in Waltham, Massachusetts.
2. Produce a simple prediction model on a house based on simple set of variables to predict housing price before starting construction.

Packages Used

```
library(ggplot2); library(ggthemes); library(MASS); library(texreg); library(tidyr)
```

The Dataset

The data file contains data describing more than 2,400 recent sales of single-family houses in the City of Waltham, Massachusetts.

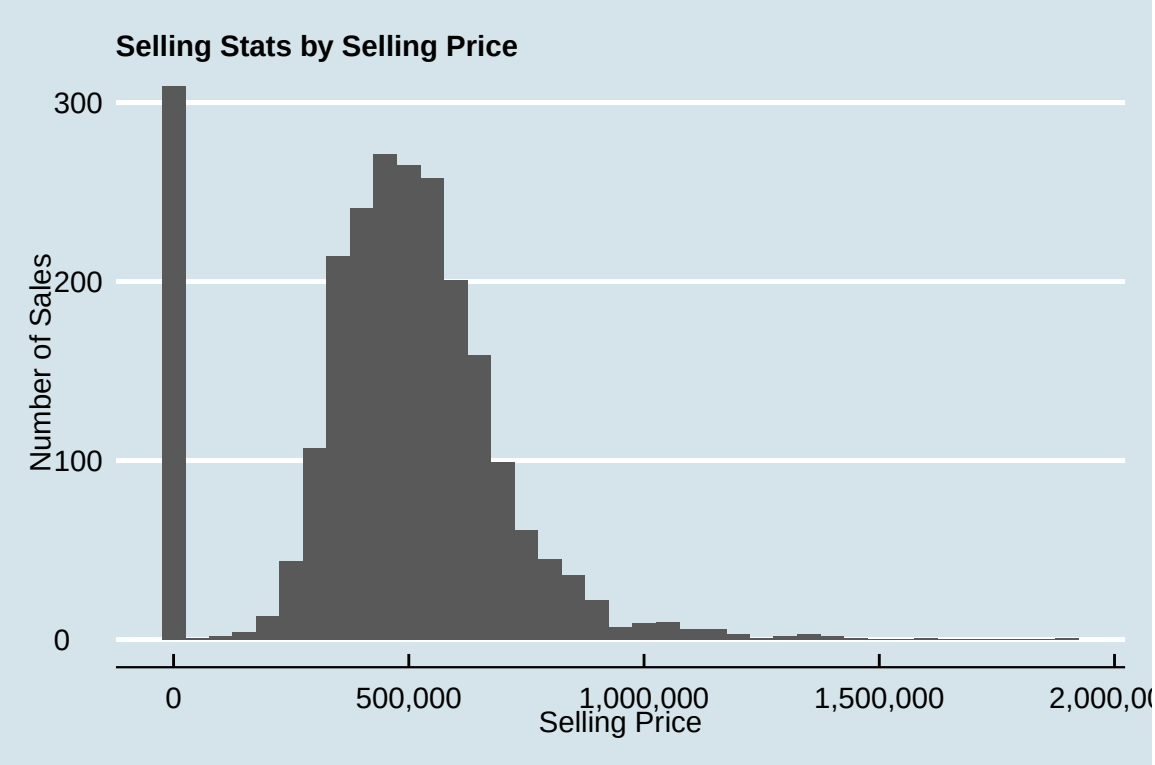
```
load("Waltham_Real_Estate_Project_2020.RData")
summary(Waltham_Real_Estate_Project_2020)
```

```
##    Address          Selling_Price          Beds           Baths
##  Length:2404        Min.   :      1   Min.   :1.000   Min.   :1.000
##  Class :character   1st Qu.: 356500   1st Qu.:3.000   1st Qu.:1.500
##  Mode  :character   Median : 475000   Median :3.000   Median :2.000
##                     Mean   : 458555   Mean   :3.212   Mean   :1.909
##                     3rd Qu.: 595000   3rd Qu.:4.000   3rd Qu.:2.500
##                     Max.   :1925000   Max.   :9.000   Max.   :6.000
##                                       NA's   :3
##  Building_Sq_Ft   Lot_Sq_Ft       Year_Built      Zip_Code
##  Min.   : 546   Min.   : 1306   Min.   :1755   Min.   :1.000
##  1st Qu.:1370   1st Qu.: 5000   1st Qu.:1932   1st Qu.:1.000
##  Median :1724   Median : 7000   Median :1950   Median :2.000
##  Mean   :1858   Mean   : 8643   Mean   :1950   Mean   :1.896
##  3rd Qu.:2138   3rd Qu.:10018   3rd Qu.:1963   3rd Qu.:3.000
##  Max.   :7818   Max.   :95396   Max.   :2020   Max.   :3.000
##
##    Date_Sold                   Elementary_School      Heat
##  Min.   :2012-04-30 00:00:00   Min.   :1.000     Min.   : 1.000
##  1st Qu.:2013-09-12 00:00:00   1st Qu.:2.000     1st Qu.: 5.000
##  Median :2015-02-26 00:00:00   Median :3.000     Median : 5.000
##  Mean   :2015-09-30 05:54:36   Mean   :2.858     Mean   : 5.738
##  3rd Qu.:2017-07-30 00:00:00   3rd Qu.:4.000     3rd Qu.: 7.000
##  Max.   :2020-07-28 00:00:00   Max.   :6.000     Max.   :13.000
##
```

```
##    Central_AC     Finished_Basement    Parking          Rooms
##  Min.   :0.0000   Min.   :0.0000    Min.   :1.000   Min.   : 3.000
##  1st Qu.:0.0000   1st Qu.:0.0000    1st Qu.:1.000   1st Qu.: 6.000
##  Median :0.0000   Median :0.0000    Median :2.000   Median : 6.000
##  Mean   :0.3814   Mean   :0.4497    Mean   :1.933   Mean   : 6.492
##  3rd Qu.:1.0000   3rd Qu.:1.0000    3rd Qu.:3.000   3rd Qu.: 7.000
##  Max.   :1.0000   Max.   :1.0000    Max.   :3.000   Max.   :20.000
##
##  Building_Grade Building_Condition    Street
##  Min.   :2.0    Min.   :1.000      Length:2404
##  1st Qu.:4.0    1st Qu.:5.000      Class :character
##  Median :4.0    Median :5.000      Mode  :character
##  Mean   :4.2    Mean   :5.152
##  3rd Qu.:4.0    3rd Qu.:5.000
##  Max.   :8.0    Max.   :9.000
##
```

Data cleaning

The price distribution indicates a large proportion of non-sales (sales at $1).



Note also that the Beds variable contains three unknowns. These data are removed from the analysis.

```
data <- drop_na(subset(Waltham_Real_Estate_Project_2020, Selling_Price > 1))
```

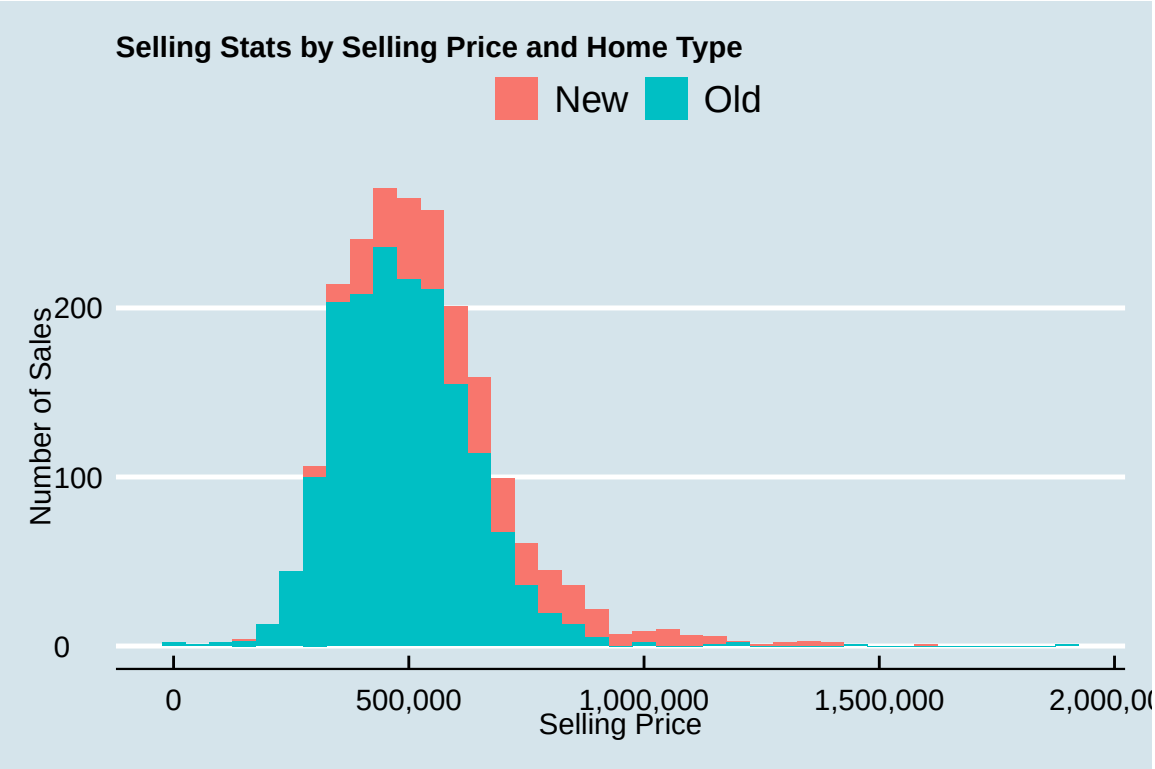
What should a New Housing Build Look Like?

Creating new variables: Age, Age_Type, Current_Tenure, Gas_Heat, Open_Space

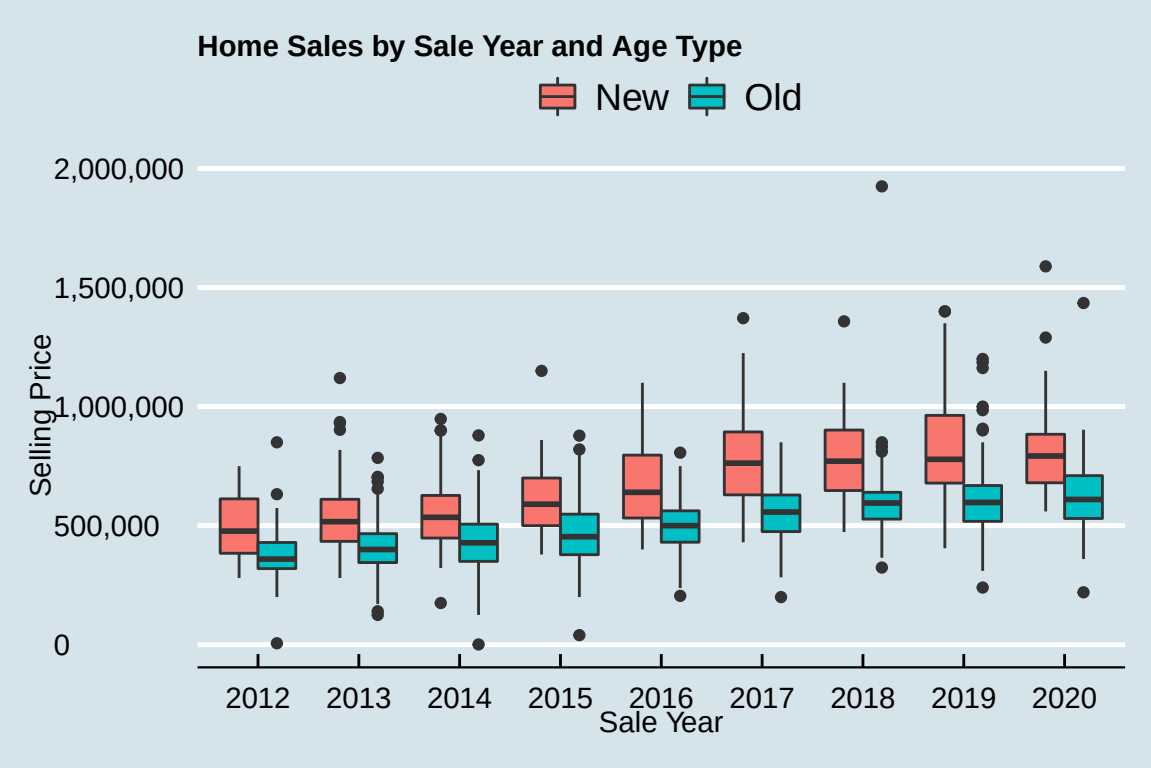
```
data$Age <- 2021 - data$Year_Built
data$Age_Type <- ifelse(data$Age < 50, "New", "Old")
data$Current_Tenure <- 2021 - as.numeric(format(data$Date_Sold,"%Y"))
data$Gas_Heat <- ifelse(data$Heat > 5 & data$Heat < 11, "Gas_Heat", "Other")
data$Open_Space <- data$Building_Sq_Ft/data$Lot_Sq_Ft
```

Newer homes are homes build after the year 1971 and older homes are build before. Gas_Heat variables capture home with gas as the fuel for their heating system. Open_Space captures the ratio between Building size and Lot size. Current_Tenure capture how long since the current owner bought the home.

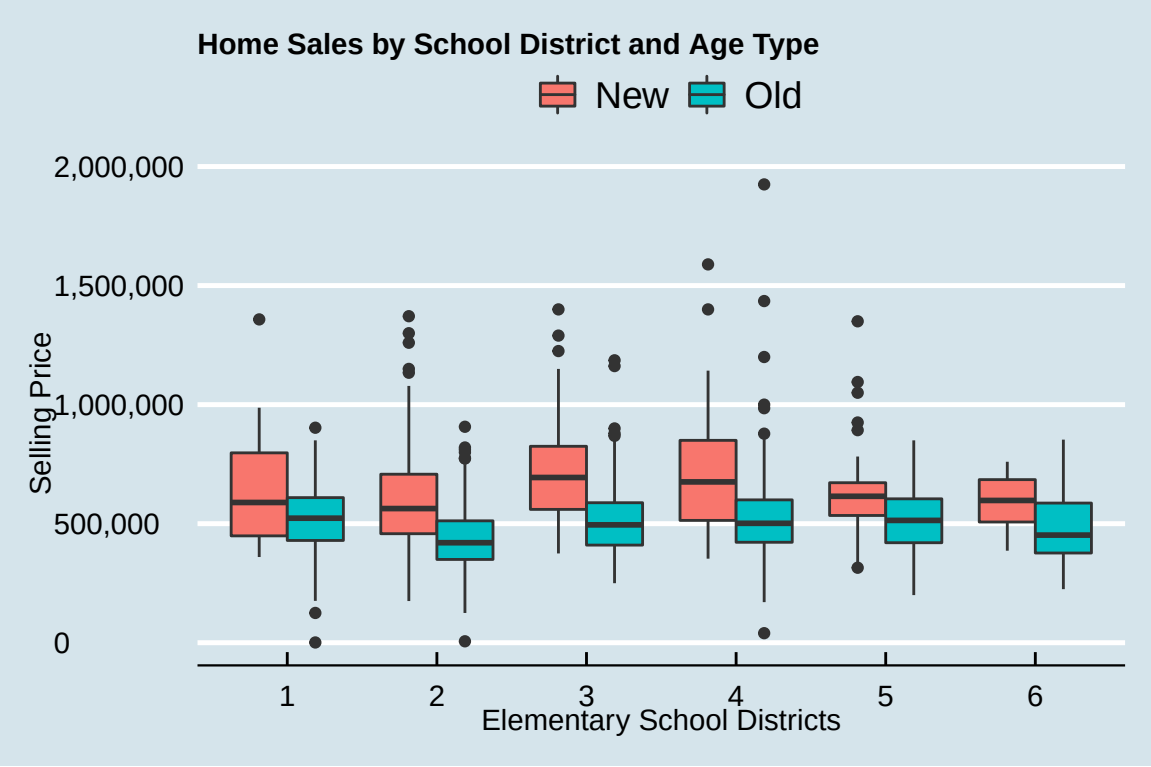
Distribution of Home Sales by Home Type



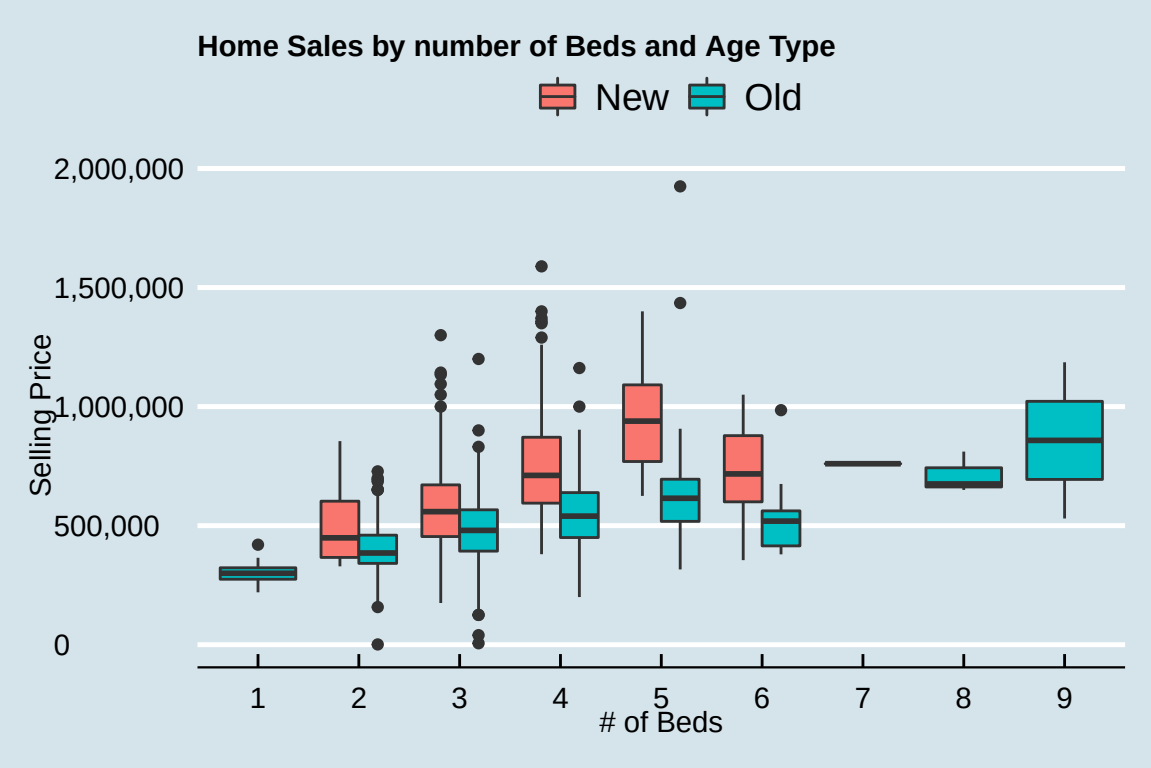
Newer homes sale prices are increasing faster than that of older homes



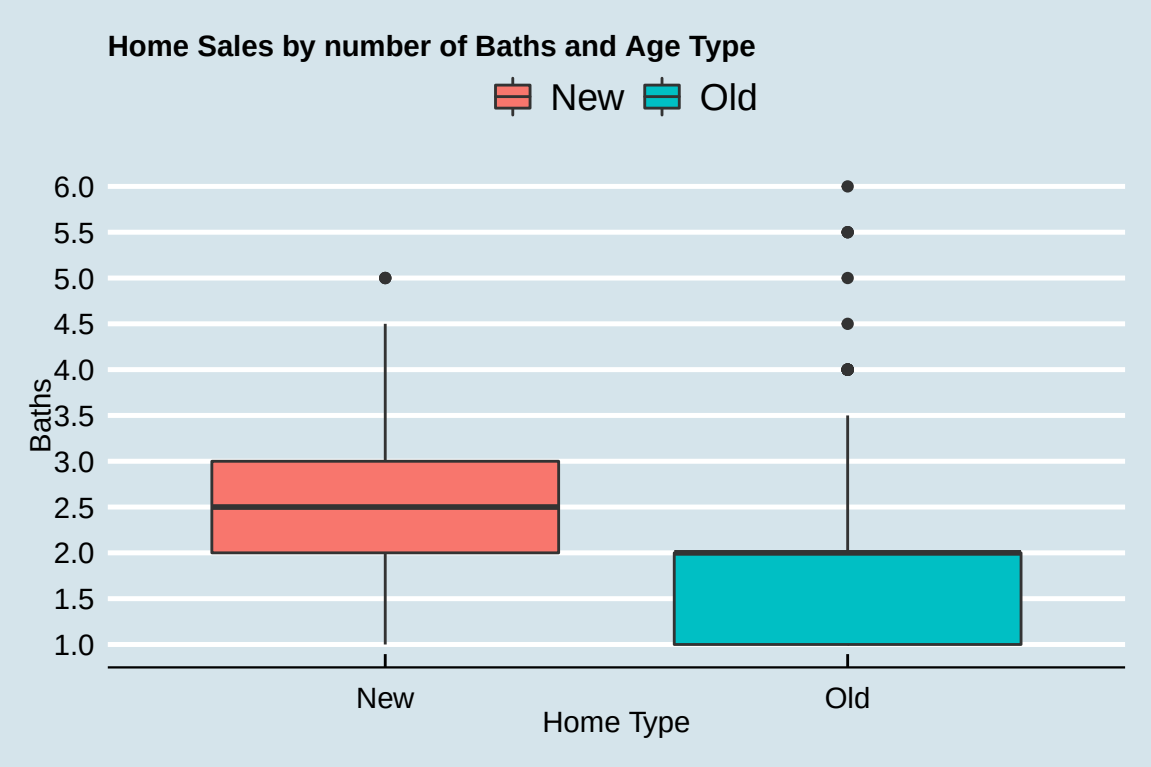
Newer homes sold for a much higher price in school district 3 and 4



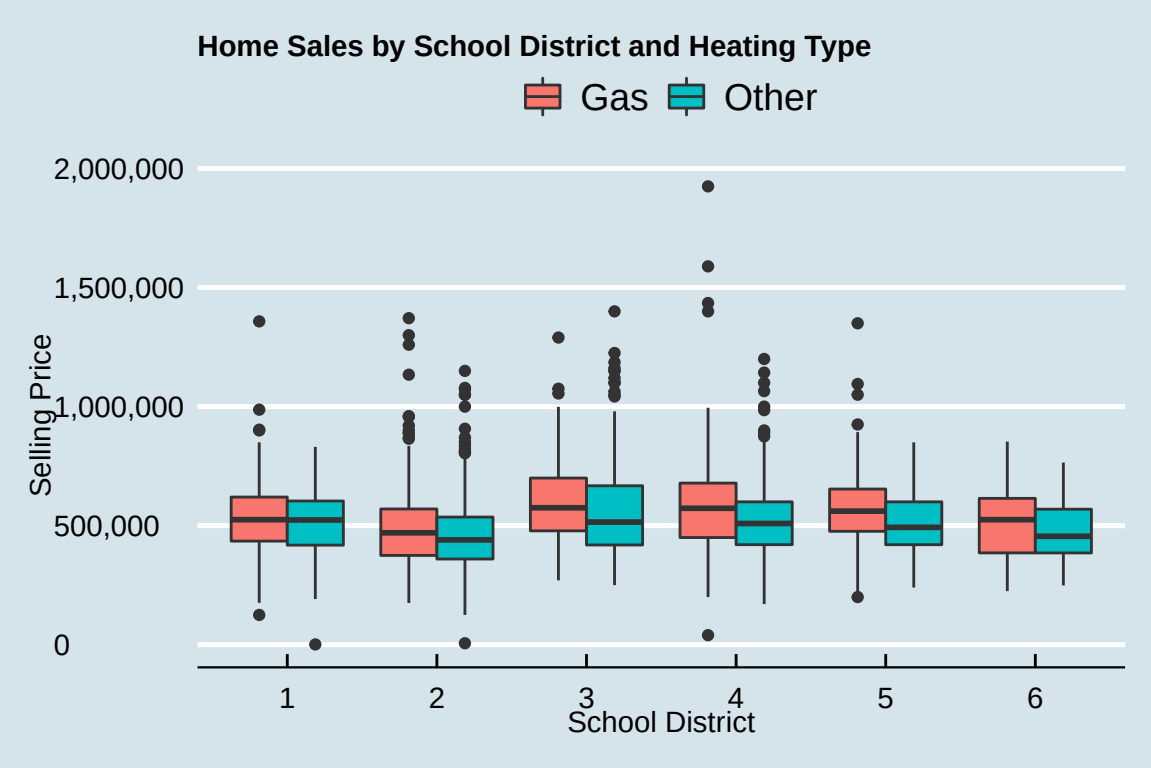
Newer homes with 5 beds sold at a highest average price among all categories



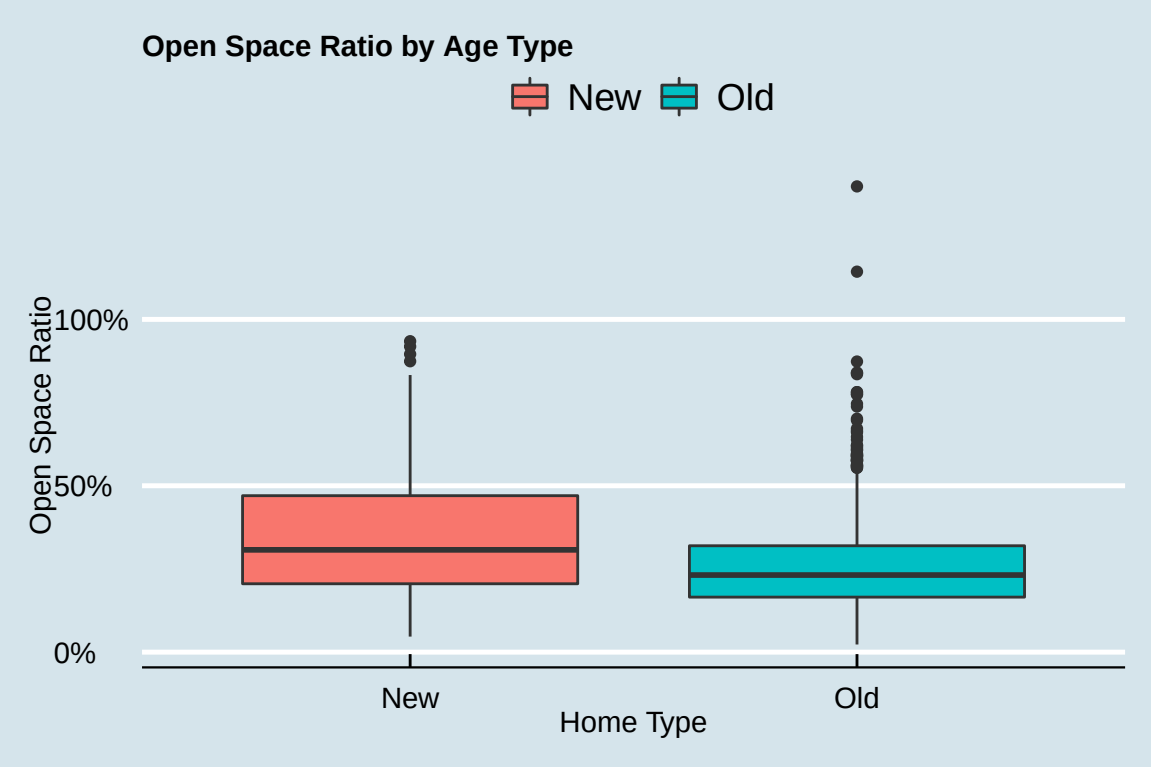
Newer homes have on average 2.5 baths



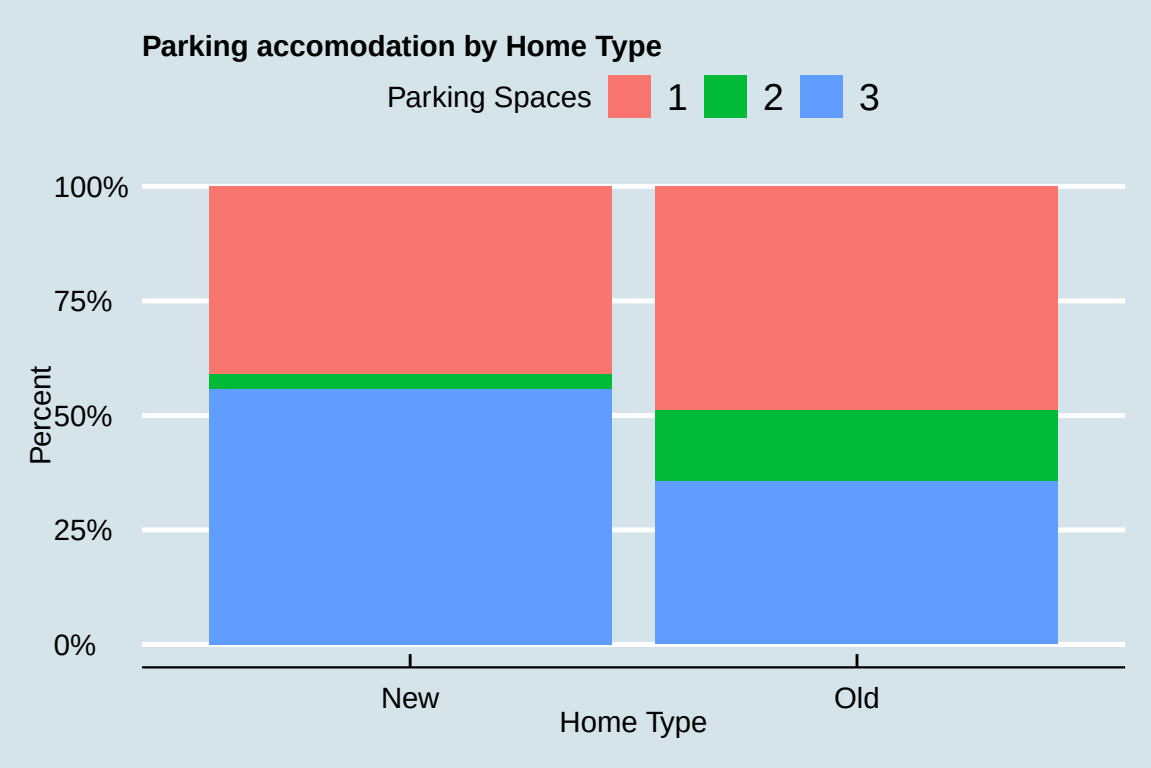
Homes with gas based heating system sold at a slightly higher average price than other types



Newer homes took up a larger percentage of the lot than older homes do



More Newer homes have 3 parking spaces than those with a single parking space



Prediction Models

Both model perform really well, both explaining 78% of the total variations in the selling prices of homes in Waltham, Massachusetts in the past 9 years. Using the Step-wise model, we can provide a few rules of thumb about the expected selling prices of home in the Waltham real estate market:

1. Each additional bathroom increases the average home sale by $34,606, All-Else-Equals (AEE).
2. Each additional room increases the average home sale by $9,553, AEE.
3. Having central AC increases the average home sale by $25,890, AEE.
4. Having a Gas heating system increases the average home sale by $9,954, AEE.
5. Having a finished basement increases the average home sale by $18.039, AEE.
6. Each additional parking spot increases the average home sale by $12,931, AEE.
7. Each additional square feet of the building increases the average home sale by $93, AEE.
8. The average selling price of home increases by $32,948 each year in the past 9 years.
9. Newer homes on average sold for $20,099 higher, AEE.

```
##
## ================================================
##                      Full Model      Step Wise
## ------------------------------------------------
## (Intercept)         -778384.64 ***  -746670.66 ***
##                     (176248.25)     (173775.45)
## Beds                  -1705.21
##                       (3122.72)
## Baths                 34804.65 ***    34606.60 ***
##                       (3606.97)       (3591.24)
## Building_Sq_Ft           92.89 ***       92.77 ***
##                          (4.75)          (4.67)
## Lot_Sq_Ft                -0.88           -0.90
##                          (0.50)          (0.50)
## Year_Built              491.07 ***      475.22 ***
##                         (86.82)         (85.70)
## Zip_Code              16072.80 ***    16459.12 ***
##                       (2229.40)       (2202.67)
## Elementary_School      1571.16
##                       (1359.23)
## Central_AC            26145.37 ***    25889.92 ***
##                       (4273.37)       (4264.12)
## Finished_Basement    -18144.22 ***   -18038.11 ***
##                       (3933.37)       (3924.56)
## Parking               12889.14 ***    12931.25 ***
##                       (2170.57)       (2170.05)
## Rooms                  9957.38 ***     9553.39 ***
##                       (1816.49)       (1662.00)
## Building_Grade        47138.40 ***    47397.44 ***
##                       (4196.73)       (4188.00)
## Building_Condition     5538.96         5655.43
##                       (3167.57)       (3164.00)
## Gas_HeatOther        -10144.05 **     -9953.69 **
##                       (3823.88)       (3817.79)
## Age_TypeOld          -19048.88 **    -20099.73 **
##                       (7149.30)       (7099.30)
## Open_Space          -126058.59 ***  -126289.05 ***
##                      (19455.24)      (19428.48)
## Current_Tenure       -32931.83 ***   -32948.35 ***
##                        (793.18)        (792.87)
## ------------------------------------------------
## R^2                       0.78            0.78
## Adj. R^2                  0.77            0.77
## Num. obs.                2096            2096
## ================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```