testing_html

2023-04-19

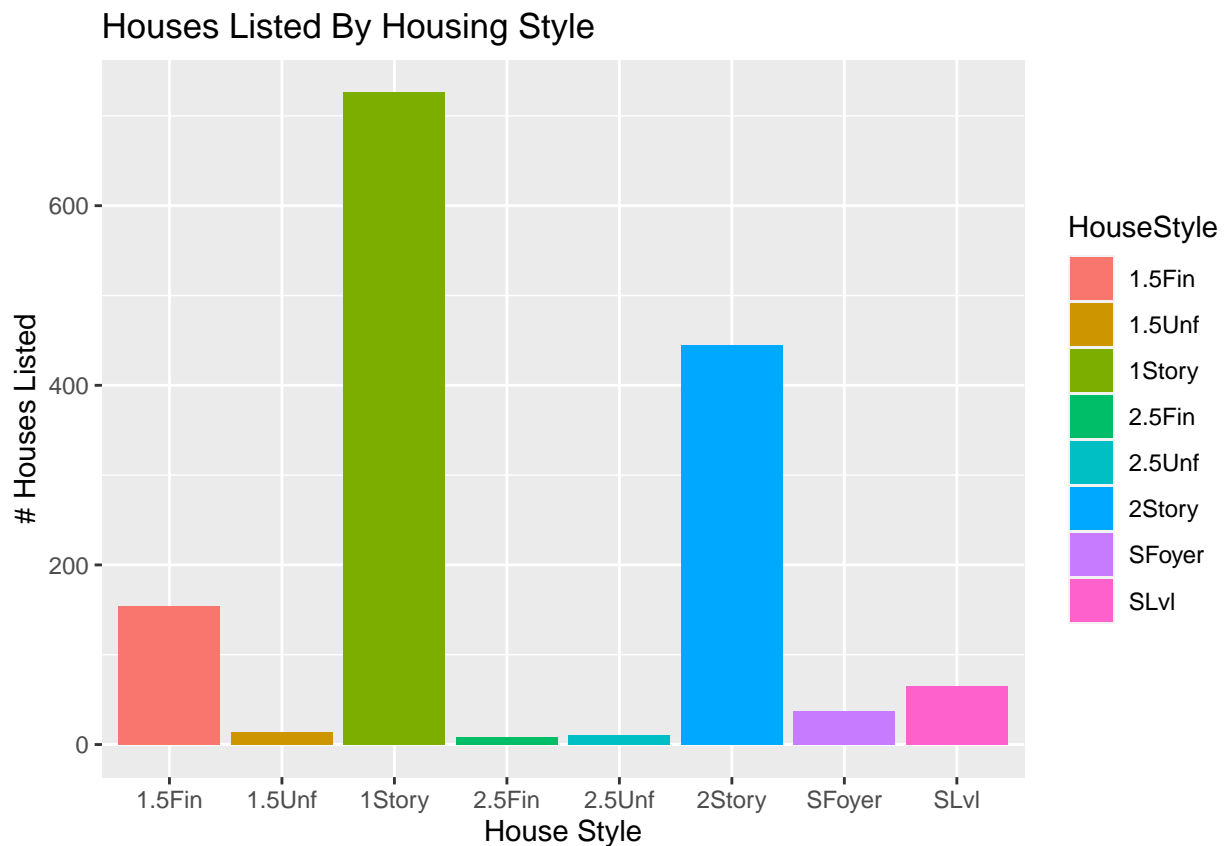```r
install.packages("readr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```r
library(readr)
library(ggplot2)

testingdataset <- read.csv("housing_train.csv")
```

```r
ggplot(data = testingdataset, mapping = aes(x = HouseStyle, fill = HouseStyle,)) +
  geom_bar()+
  labs(title = "Houses Listed By Housing Style",
       x = "House Style",
       y = "# Houses Listed",
       )
```


Bar Plot 1: In this bar plot, we chose to examine the relationship between the houses listed and the house style. We found that 1 story houses were the most prevalent of those listed, followed by 2 story houses, while 2.5 story houses with a finished second level were the least likely to be listed.

```
ggplot(data = testingdataset, mapping = aes(x = BldgType, fill = BldgType)) +
  geom_bar() +
  labs(title = "Houses Listed By Building Type",
       x = "Building Type",
       y = "# Houses listed",
       )
```
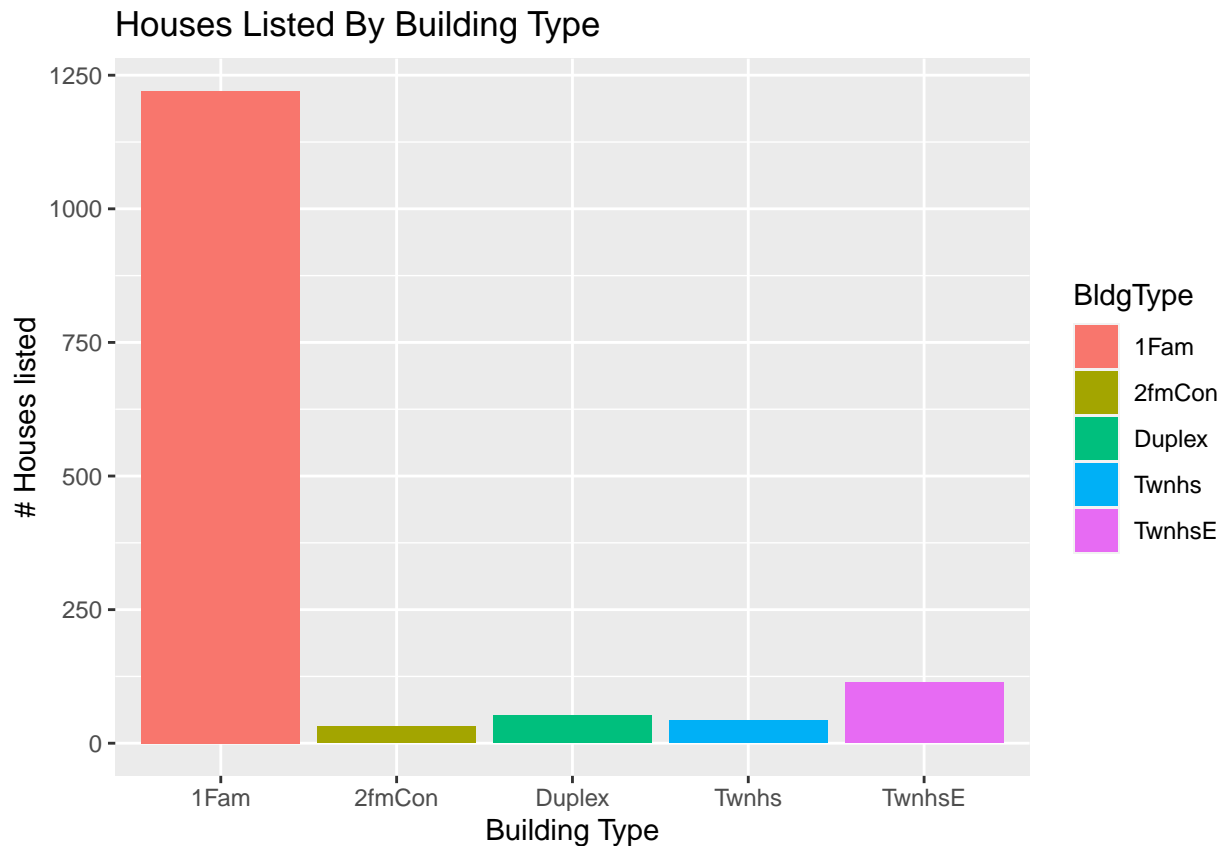

Bar Plot 2: In this bar plot we examine the relationship between the houses listed and the building type. According to the data, we found that single-family detached houses were by far the most common, with nearly 1250 listed. Townhouse end units come in second in terms of frequency among the listed houses, but are still far less common at about 125 listed. The least common building type listed was the two-family conversion (houses that were originally built as single-family), with approximately 35 listings.

```
ggplot(data = testingdataset, mapping = aes(BldgType, HouseStyle, color = BldgType)) +
  geom_count()+
  labs(title = "Building Type & Style",
       x = "Building Type",
       y = "House Style",
       size = "Number of Listed Properties",
       color = "Building Type")
```

Building Type & Style
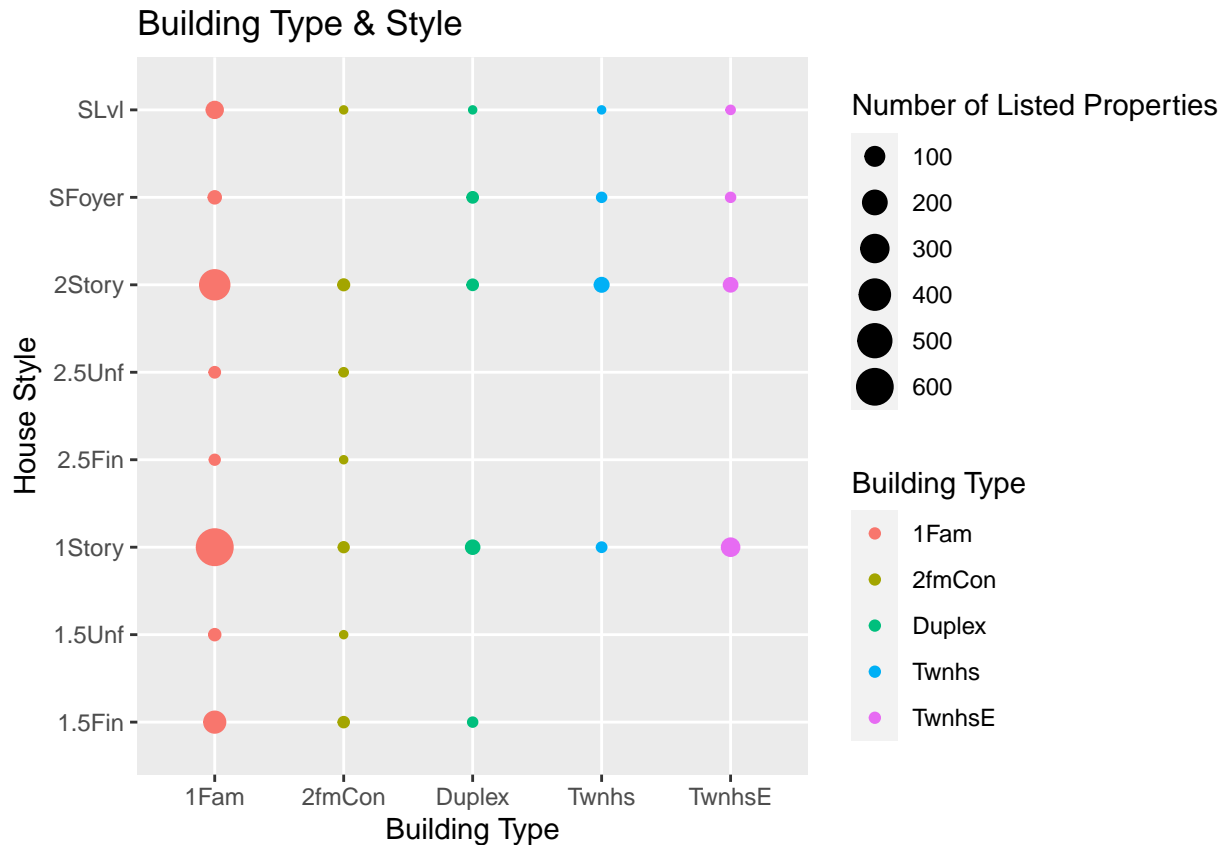
Plot 3: In this plot we examined the relationship between several variables: the house style, building type, and number of listings. The most common combination of house style and building type we found listed was the 1-story, single-family home, followed closely by the 2-story, single-family home. In general, we noticed that regardless of house style, single-family homes tended to outnumber all other building types. And, as established in the previous plot, two-family conversions were the least frequently listed houses, despite having a greater variety of house styles than the similarly uncommon townhomes. Altogether, this indicates that building type may play a more significant role than house style in terms of which house characteristics prospective homebuyers are prioritizing.

```
p <- ggplot(testingdataset, aes(YearBuilt, HouseStyle)) + geom_point()
# Use vars() to supply faceting variables:
p + facet_wrap(vars(BldgType)) +
  labs(title = "Style of Houses Built grouped by Building Type (approx. 1860-2020)",
       x = "Year Built",
       y = "Style of House",
       )
```

# Style of Houses Built grouped by Building Type (approx. 1860–2020)
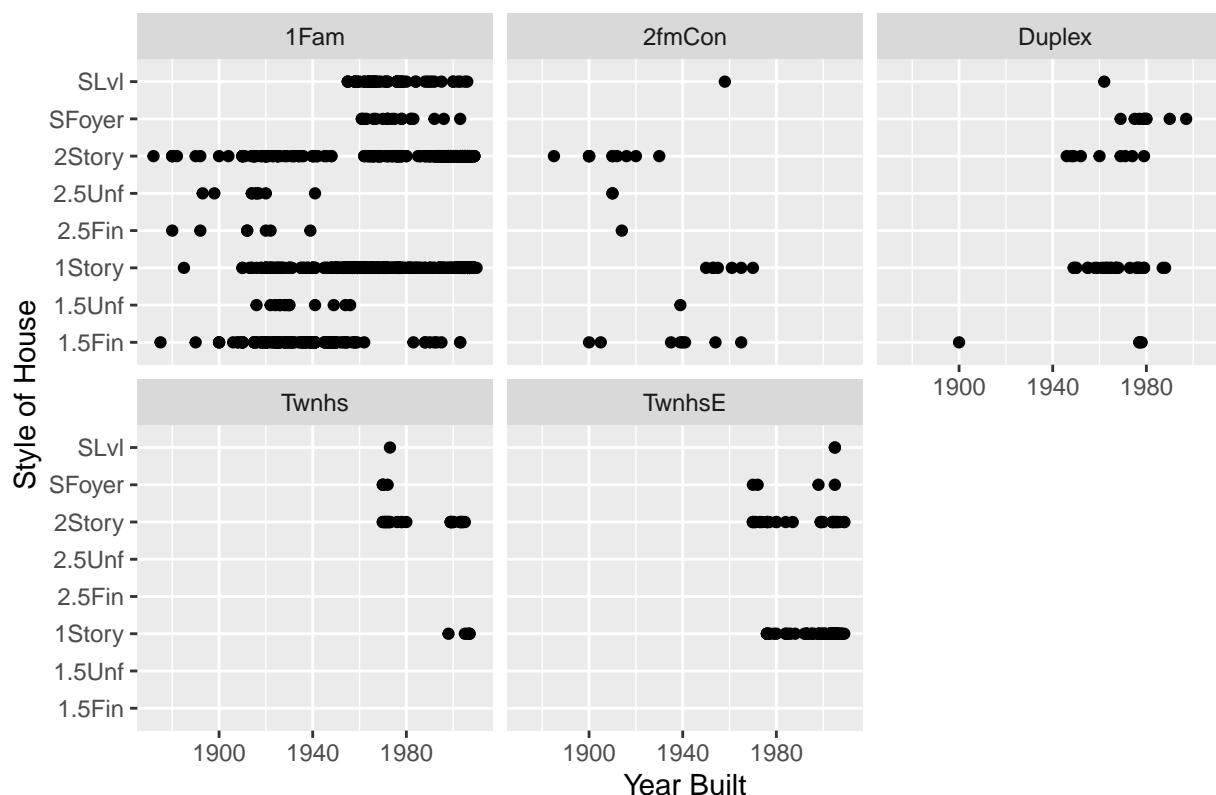
Plot 4: In these plots we examined the relationship between house style and year built for each building type. With regards to single-family homes, 2 story houses and 1.5 story houses with a finished second level seem to be the most consistently built homes since the late 1800s, with 1 story houses eventually exceeding their frequency after 1900. Two-family conversions, while never particularly common, were more prevalent prior to the 1970s, after which they stopped being built entirely. Duplexes, on the other hand, only started being built in (relatively) large numbers after 1940. Similarly, we found that townhomes, regardless of style or type, only started being built around the 70s. Additionally, across all building types, 1 and 2 story houses styles tended to be the most common.

```
ggplot(testingdataset, aes(YrSold, fill = HouseStyle)) +
  geom_histogram() +
    labs(title = "Types of Houses Sold (2006-2010)",
        x = "Year Sold",
        y = " ",
      )
```

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

Types of Houses Sold (2006–2010)
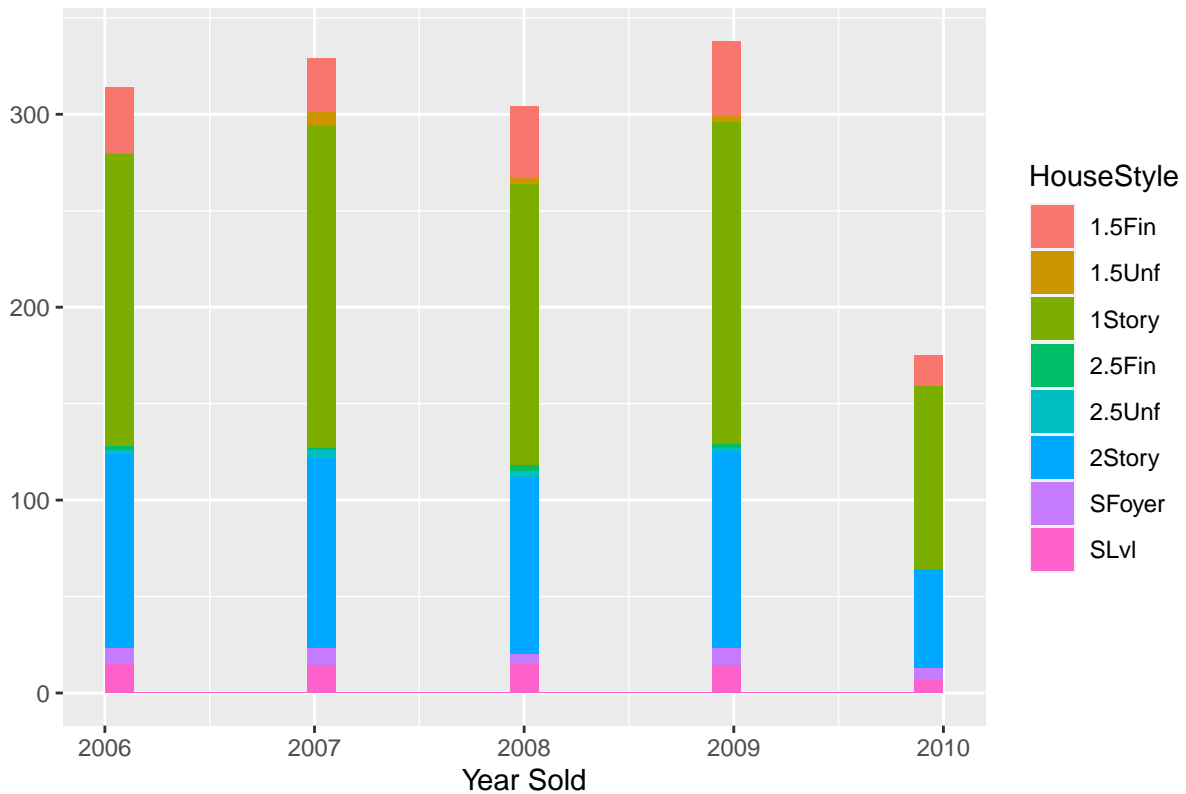
Plot 5: In this graph we examined the relationship between the house style, the year sold, and the number of listings. Of the houses sold in any given year, 1 story homes were the most prevalent – usually consisting approximately 50% of the listings. On the other hand, 2.5 story homes tended to be among the least frequently listed.

```
ggplot(testingdataset, aes(x=BldgType, y=SalePrice, fill=BldgType)) +
  geom_boxplot() +
    labs(title = "Central Tendencies of Sale's Price based on Building Type",
        x = "Building Type",
        y = "Sale's Price ($)",
      ) +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE, big.mark = ","))
```

## Central Tendencies of Sale's Price based on Building Type
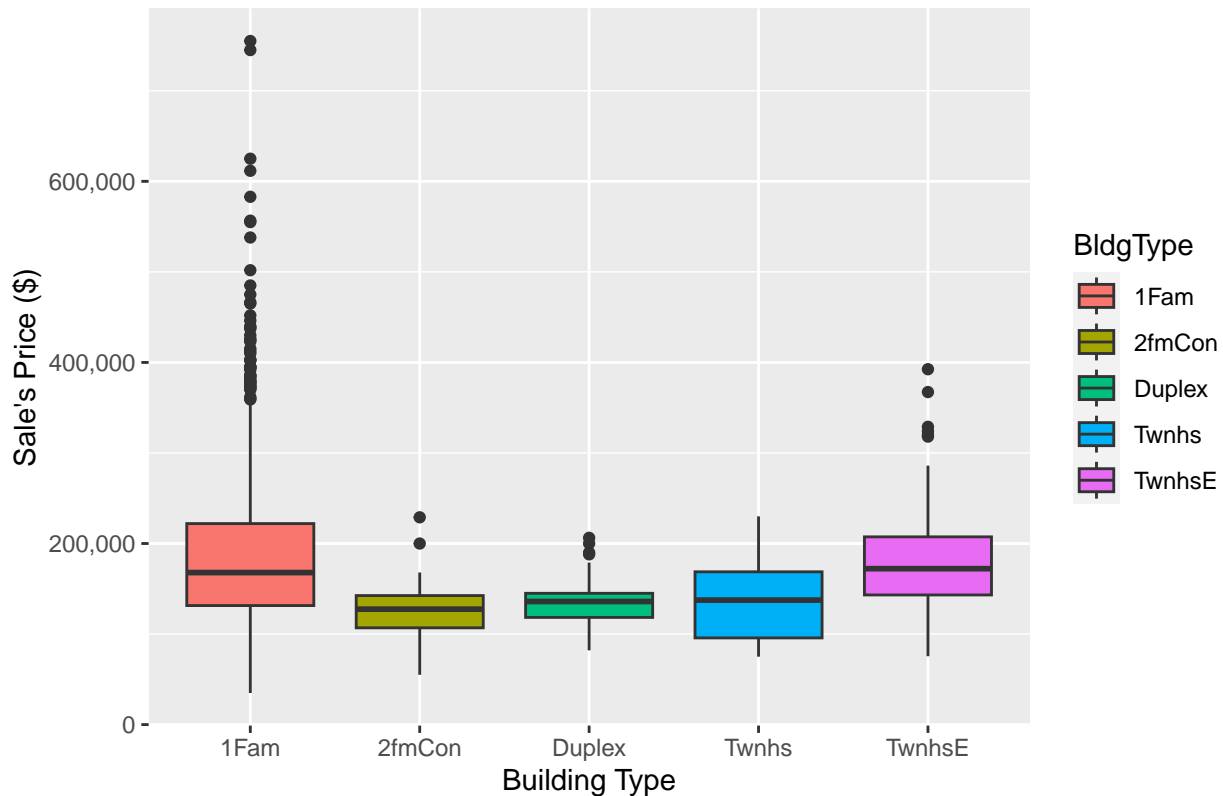
Plot 6: In this graph we examined the relationship between building type and sale price. Single-family homes and townhouses seemed to have comparable median sale prices, though the single-family homes did have a significantly higher number of more expensive outliers. The single-family homes also had a slightly lower minimum sale price, though this wider range is likely in part due to there being many more single family homes listed than townhomes and the greater variety of house styles that single-family homes have.

```
ggplot(testingdataset, aes(x=BldgType, y=GarageArea, fill=BldgType)) +
  geom_boxplot() +
  stat_summary(position = "dodge",
          fun=mean, geom='point', shape=20) +
  labs(title = "Central Tendencies of Garage Area based on Building Type",
      x = "Building Type",
      y = "Garage Area (Feet)",
      )

## Warning: Width not defined
## i Set with `position_dodge(width = ...)`
```

Central Tendencies of Garage Area based on Building Type
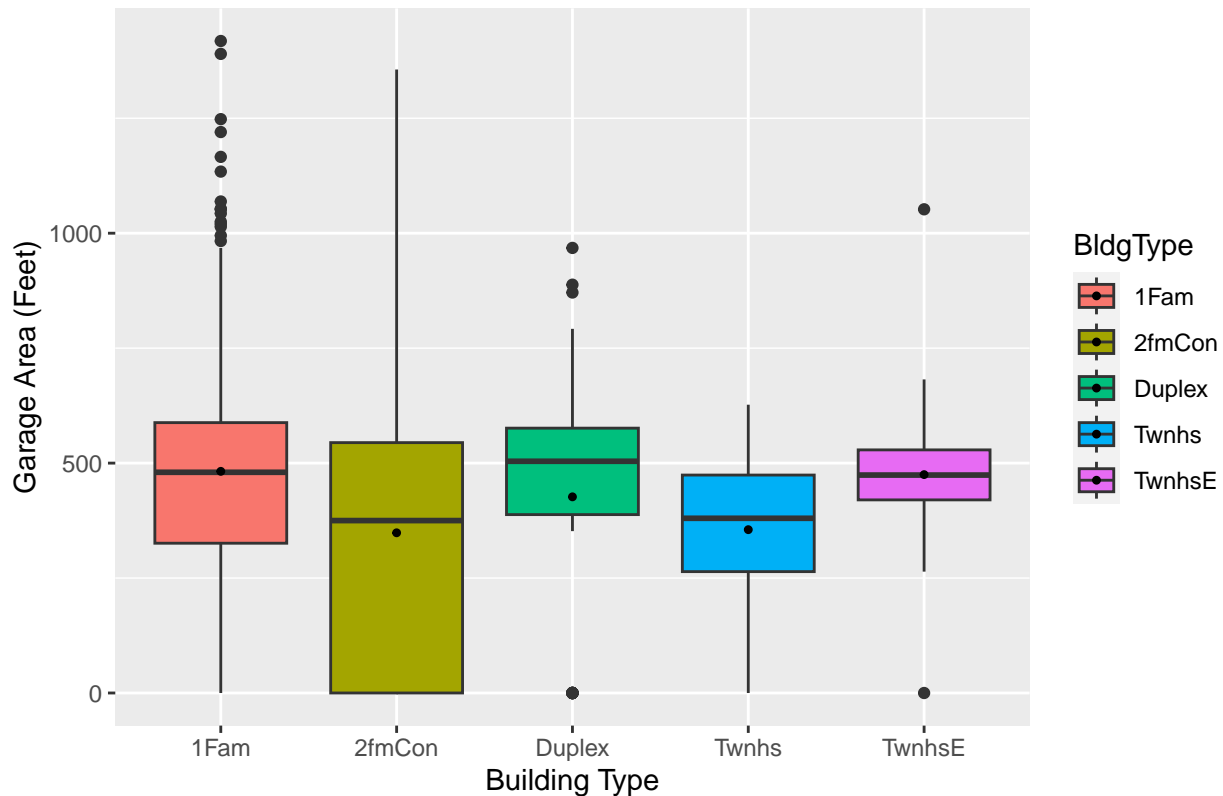
Plot 7: In this graph we examined the relationship between building type and garage area. Interestingly, the two-family conversions seemed to have the greatest spread/interquartile range in terms of garage area, with all the other building types having, in comparison, relatively consistently sized garages. However, as in the previous graph, the single-family homes still have the greatest number of outliers on both ends.
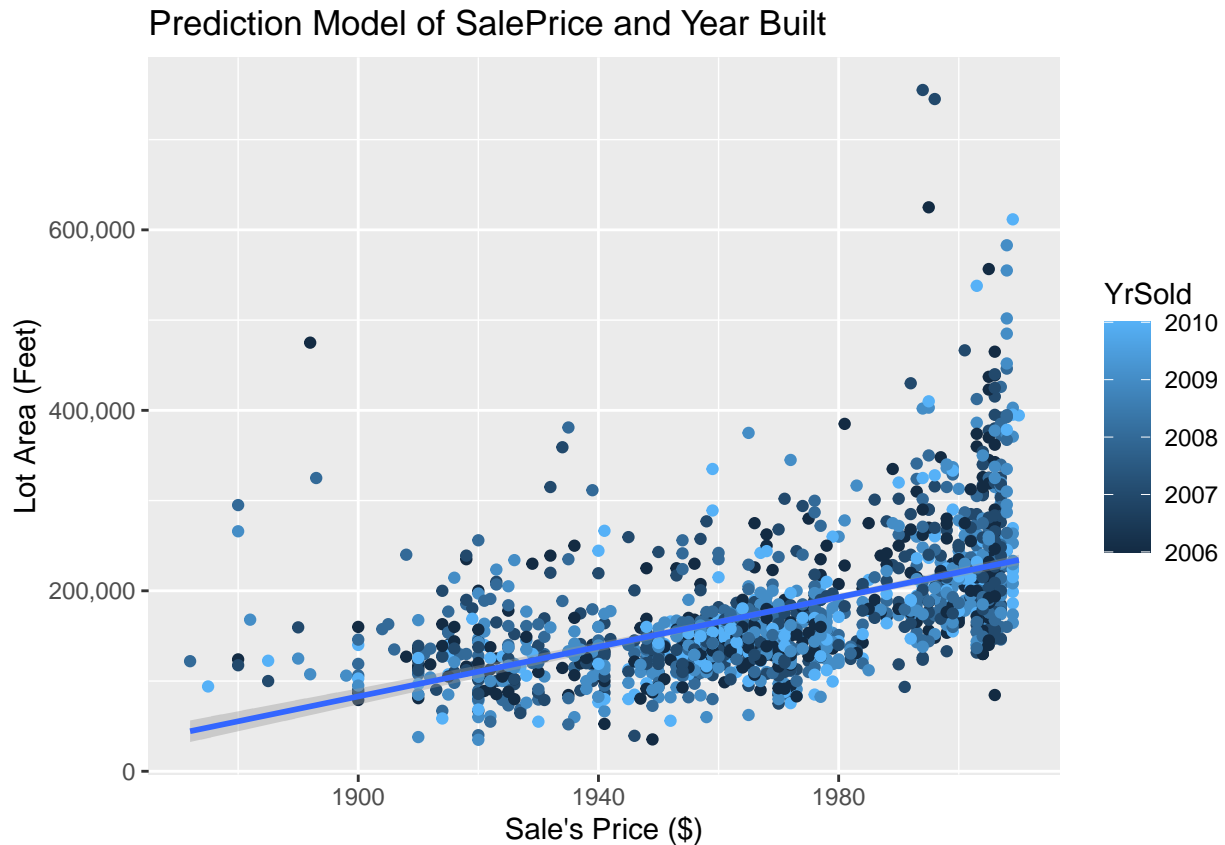
```r
ggplot(testingdataset,aes(YearBuilt, SalePrice, fill = YrSold, color = YrSold)) +
  geom_point() +
  geom_smooth(method='lm') +
  scale_y_continuous(labels = function(x) format(x, scientific = FALSE, big.mark = ",")) +
    labs(title = "Prediction Model of SalePrice and Year Built",
      x = "Sale's Price ($)",
      y = "Lot Area (Feet)",
      )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: The following aesthetics were dropped during statistical transformation: fill,
## colour
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```

Prediction Model of SalePrice and Year Built

Plot 8: In this graph, we examined the relationship between the year the houses were built, the sales price, and the year they were sold. In general, we found that the later the houses were built, the more they were sold for – likely a function of inflation. The year sold, on the other hand, did not seem to have as much of a discernible relationship with the sales price.

```
ggplot(testingdataset,aes(SalePrice, LotArea, fill = YrSold, color = YrSold)) +
  geom_point() +
  geom_smooth(method='lm') +
  scale_y_continuous(labels = function(x) format(x, big.mark = ",")) +
  scale_x_continuous(labels = function(x) format(x, scientific = FALSE, big.mark = ",")) +
    labs(title = "Prediction Model of SalePrice and Lot Area",
      x = "Sale's Price ($)",
      y = "Lot Area (Feet)",
      )
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: The following aesthetics were dropped during statistical transformation: fill,
## colour
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```

Prediction Model of SalePrice and Lot Area
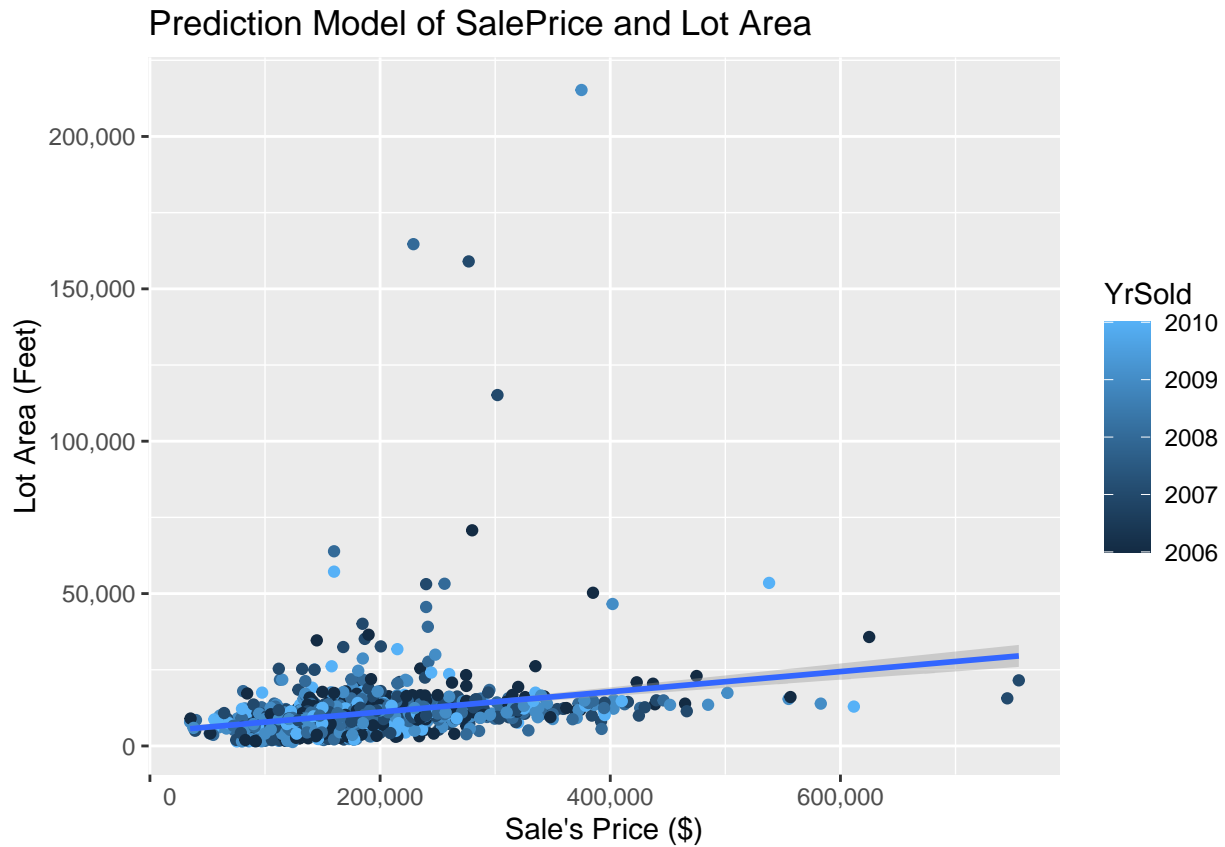
Plot 9: In this graph, we examined the relationship between the sale price, lot area, and year sold. We found that as the sale price increased, the lot area would also increase, albeit at a relatively low rate. We also noticed that there were a fair number of outliers that had relatively average or slightly above-average sales prices, but significantly higher lot areas than comparably priced homes. These outliers likely represent farmland or rural lots in which the price per acre is significantly lower than the price in more suburban or urban areas.