Examining Miami's Housing Market

Thomas Burrell (tmb9ccd)

February 23, 2022

1 Article

Link to Article (https://www.homemindset.com/why-is-miami-so-expensive/)

The article I chose tries to answer the question about why Miami's real estate is so expensive and what characteristics of a house lead to the highest prices. Some important characteristics when buying a house include location, amenities, and age. Urban areas tend to have lots of job opportunities which leads to higher average pay. However, because the cost of living is also very high in cities, it tends to cancel out or outpace the pay increase people get in cities.

Miami is the 2nd most expensive residential real estate market in the country

With cities like Los Angeles, New York, and San Francisco it make come as a surprise that Miami is so expensive, now the second most expensive residential real estate market in the Country. A household in Miami should expect to pay $2,653 a month toward home ownership costs, or roughly 81.6 percent of median income. I believe it would be beneficial to explore Miami's housing data to see if any trends occur that would help prospective home buyers make decisions before moving to an expensive area like Miami. The question I would like to answer is, are variables like structure quality, age, and floor area important to the sale price of a Miami house?

2 Data Summary

Link to Data (https://www.kaggle.com/deepcontractor/miami-housing-dataset)

My dataset consists of 17 variables that relate to the housing market in Miami and was pulled from Kaggle. The dataset consists of relevant attributes relating to the housing market in the Miami, FL. Here is a list of the variables used in the dataset along with their type.

```
## 'data.frame':   13932 obs. of  17 variables:
##  $ LATITUDE        : num  25.9 25.9 25.9 25.9 25.9 ...
##  $ LONGITUDE       : num  -80.2 -80.2 -80.2 -80.2 -80.2 ...
##  $ PARCELNO        : num  6.22e+11 6.22e+11 6.22e+11 6.22e+11 6.22e+11 ...
##  $ SALE_PRC        : num  440000 349000 800000 988000 755000 630000 1020000 850000 250
000 1220000 ...
##  $ LND_SQFOOT      : int  9375 9375 9375 12450 12800 9900 10387 10272 9375 13803 ...
##  $ TOT_LVG_AREA    : int  1753 1715 2276 2058 1684 1531 1753 1663 1493 3077 ...
##  $ SPEC_FEAT_VAL   : int  0 0 49206 10033 16681 2978 23116 34933 11668 34580 ...
##  $ RAIL_DIST       : num  2816 4359 4413 4585 4063 ...
##  $ OCEAN_DIST      : num  12811 10648 10574 10156 10837 ...
##  $ WATER_DIST      : num  348 338 297 0 327 ...
##  $ CNTR_DIST       : num  42815 43505 43530 43798 43600 ...
##  $ SUBCNTR_DI      : num  37742 37341 37329 37423 37551 ...
##  $ HWY_DIST        : num  15955 18125 18201 18514 17903 ...
##  $ age             : int  67 63 61 63 42 41 63 21 56 63 ...
##  $ avno60plus      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ month_sold      : int  8 9 2 9 7 2 2 9 3 11 ...
##  $ structure_quality: int  4 4 4 4 4 4 5 4 4 5 ...
```

There are almost 14,000 rows (which equates to 14,000 single-family homes sold in Miami).

Important variables to explore:

- SALE_PRC (sale price in dollars)
- TOT_LVG_AREA (floor area (square feet))
- Structure Quality (1-5)

```
summary(house_data[c(4, 6, 17)])
```

```
##     SALE_PRC         TOT_LVG_AREA   structure_quality
##  Min.   :  72000   Min.   : 854   Min.   :1.000
##  1st Qu.: 235000   1st Qu.:1470   1st Qu.:2.000
##  Median : 310000   Median :1878   Median :4.000
##  Mean   : 399942   Mean   :2058   Mean   :3.514
##  3rd Qu.: 428000   3rd Qu.:2471   3rd Qu.:4.000
##  Max.   :2650000   Max.   :6287   Max.   :5.000
```

- $310,000 house price (Median)
- 1878 Ft. Floor Area (Median)
- 3.514 Structure_quality rating (mean)

3 Data Validation

```
nrow(unique(house_data))
```

```
## [1] 13932
```

Each row in the dataset is unique

```
sum(is.na(house_data))
```

```
## [1] 0
```

There are 0 missing values from the dataset

```
nrow(na.omit(house_data))
```

```
## [1] 13932
```

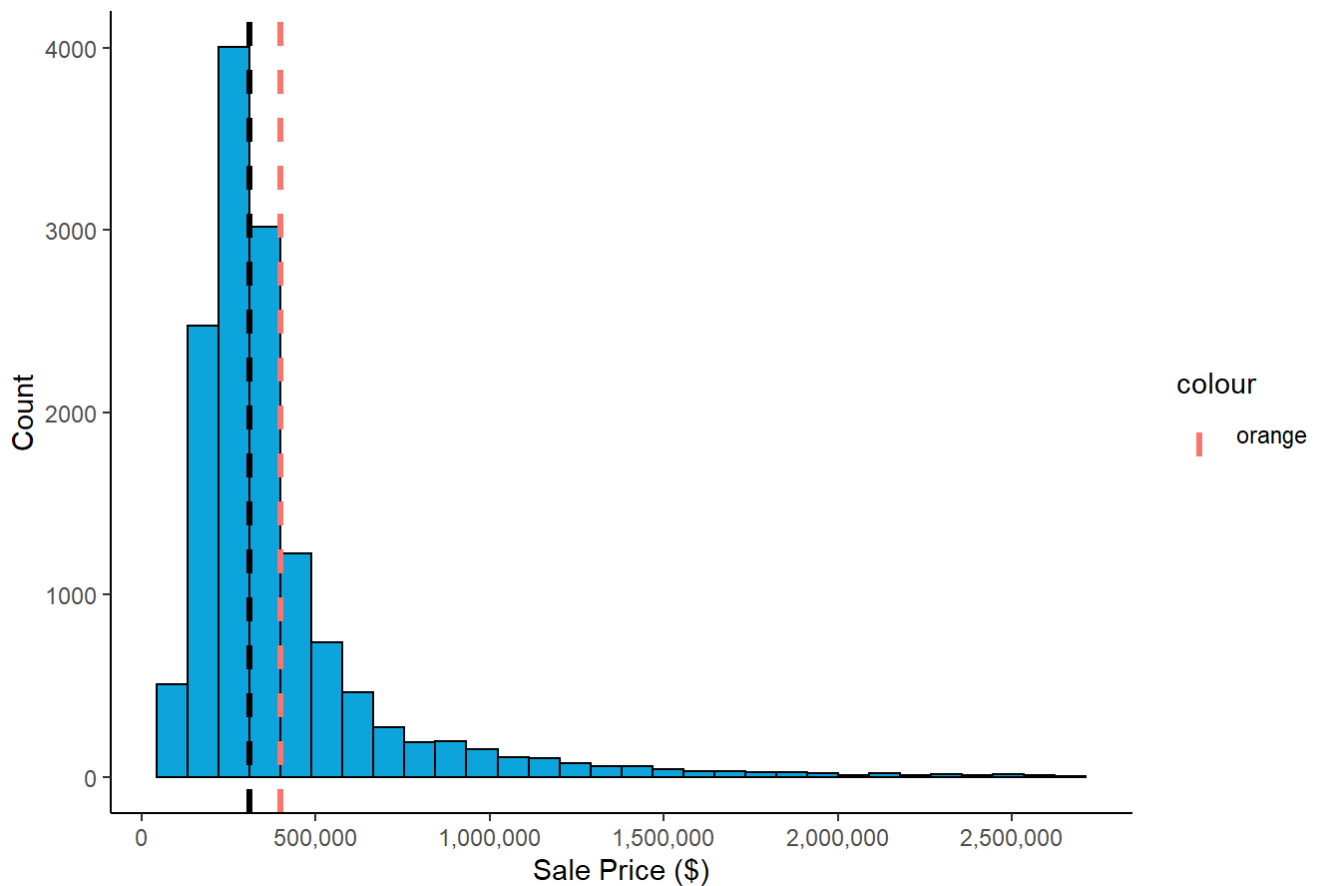Nothing to omit because there are 0 NA

The dataset is extremely clean with no rows or columns to edit.

4 Plots/Graphs

4.1 Home Price Graphic

First, a histogram will be plotted of Home Sale Price to see what kind of distribution and spread our data has.

Histogram of Home Sale Price in Miami, FL


As we can see from the histogram,

- The median home price is $310,000
- The mean home price is around $400,000.

Because our distribution is skewed right, it increases the mean of home price drastically because of the very expensive homes on the right side of the plot. This histogram's shape is typical when plotting home prices because there tends to be many middle and lower priced homes with a few very expensive homes in any given area.

4.2 SQ FT and Home Price Graphic

Next, I will plot a scatterplot of Sale Price ($) vs. Total floor area (SQ FT) to see if a positive relationship exists between the two variables.

Scatterplot of Sale Price ($) vs. Total floor area (SQ FT)
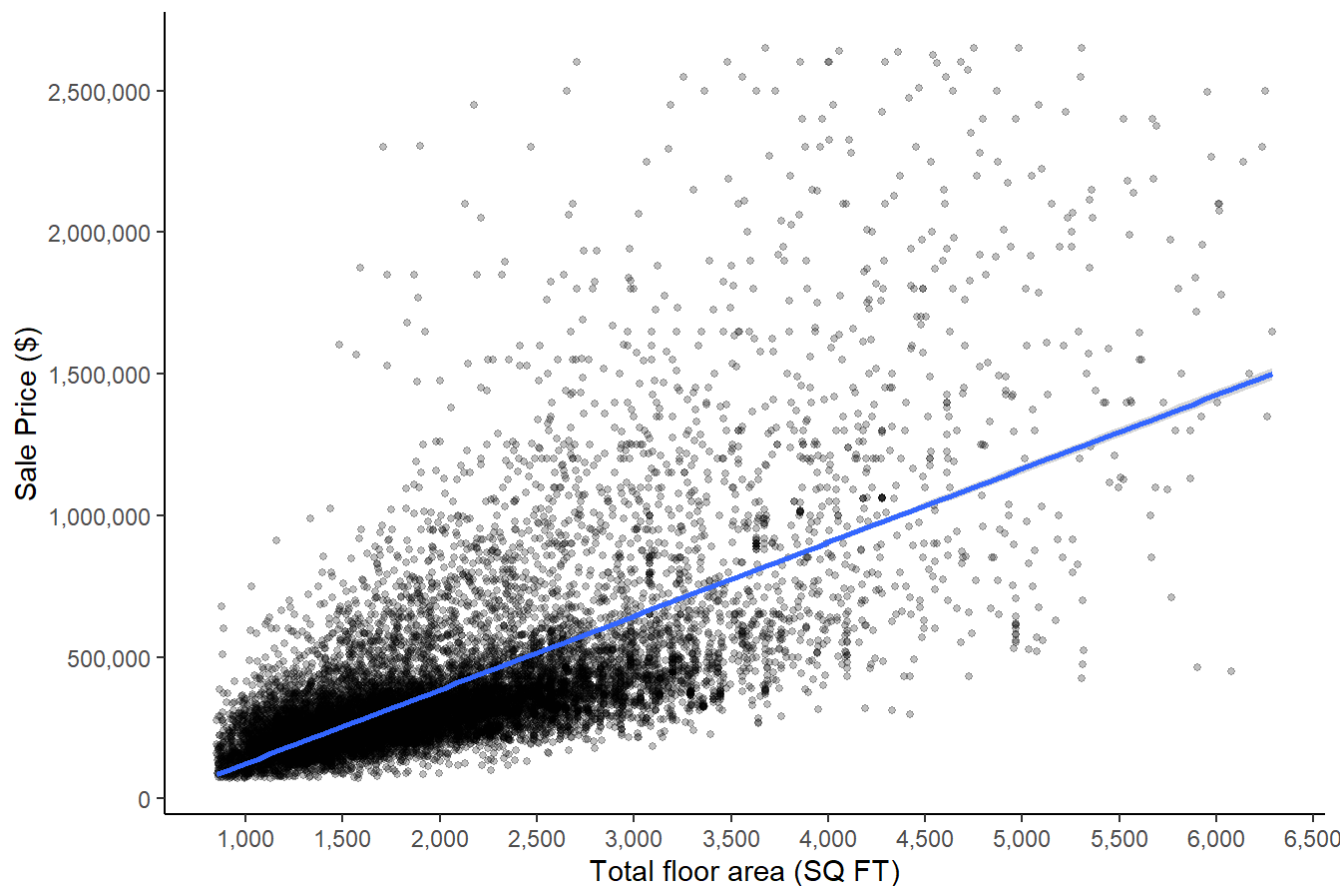
I decided to plot sale price against total floor area because the size of the house would seem like a very important factor when calculating price. We can see from the linear regression line that as floor area increases, the average sale price also increases.

4.3 Age Graphics

To see how influential age is on sale price of a house, I plotted a boxplot of sale price vs. age. The age variable is split up between 10 year increments with the age being rounded to the nearest 10.

Boxplot of Sale Price vs. Age
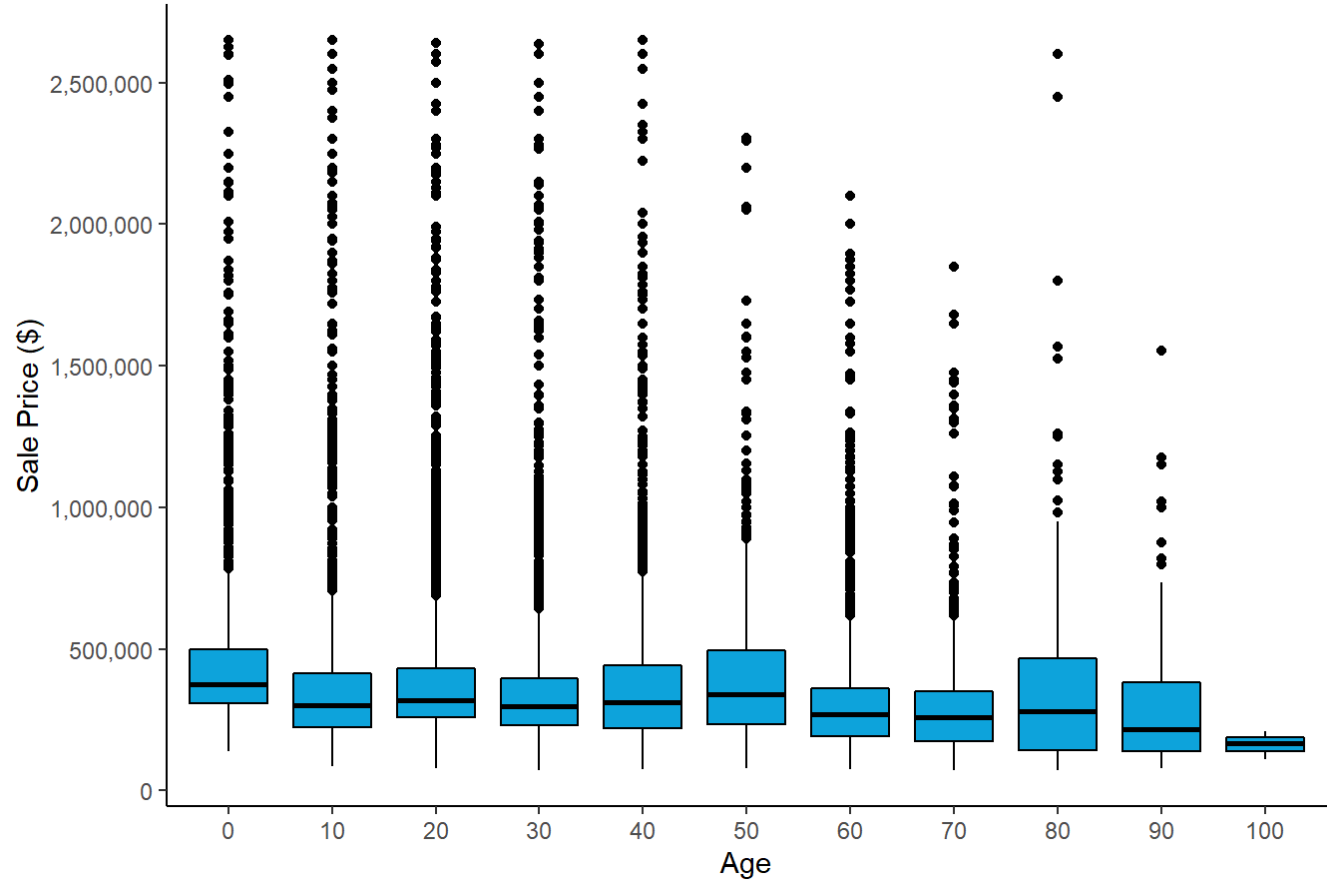
This plot does not provide too much value with all of the outliers included because it makes the y axis very large and hard to read the averages of each boxplot. To account for this, I will remove the outliers and zoom in on the boxplots.

Boxplot of Sale Price vs. Age (Outliers Removed)
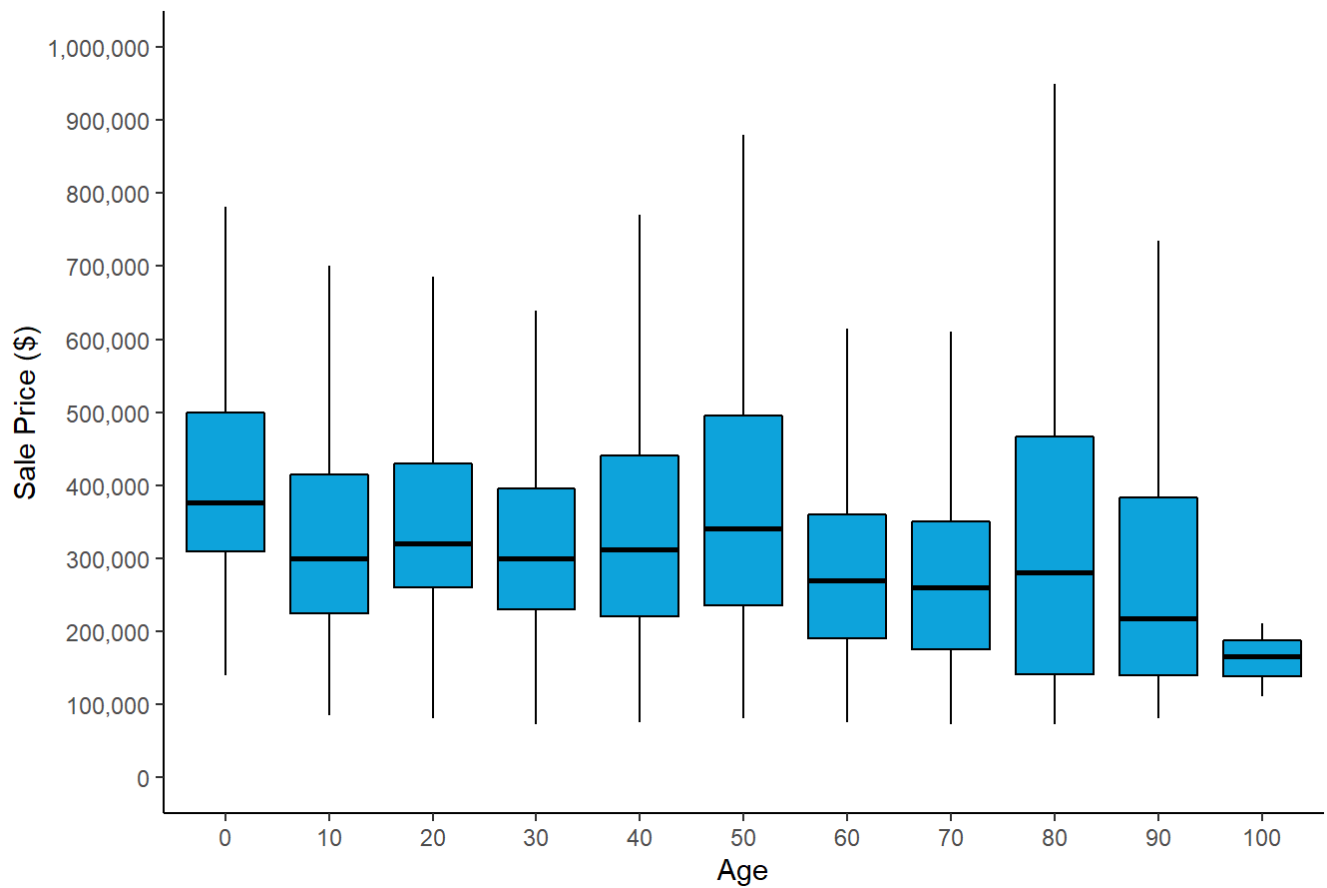
After making the necessary changes from the first boxplot, we can see that there is not much impact of age on how much a house cost in Miami.

4.4 Structure Quality Graphics

Lastly, a boxplot of structure quality vs sale price to group the sale price by each structure quality group (1-5)

Boxplots of Sale Price vs. Structure Quality
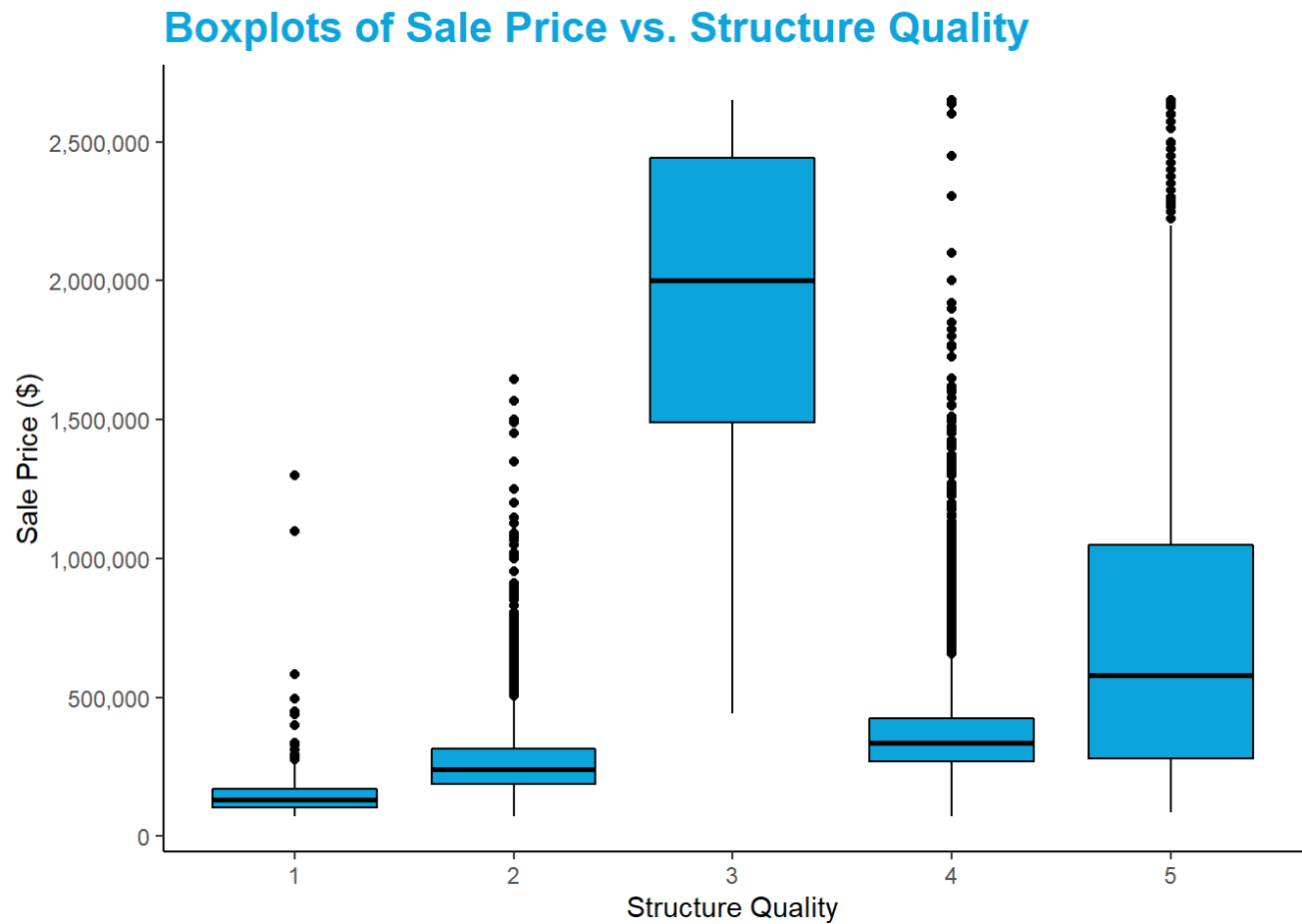
As we can see, as the structure quality increases, the sale price also increase, besides the 3 category. We should take a closer look at the number of values in each structure quality rating using a histogram.

Histogram of Structure Quality of homes in Miami, FL
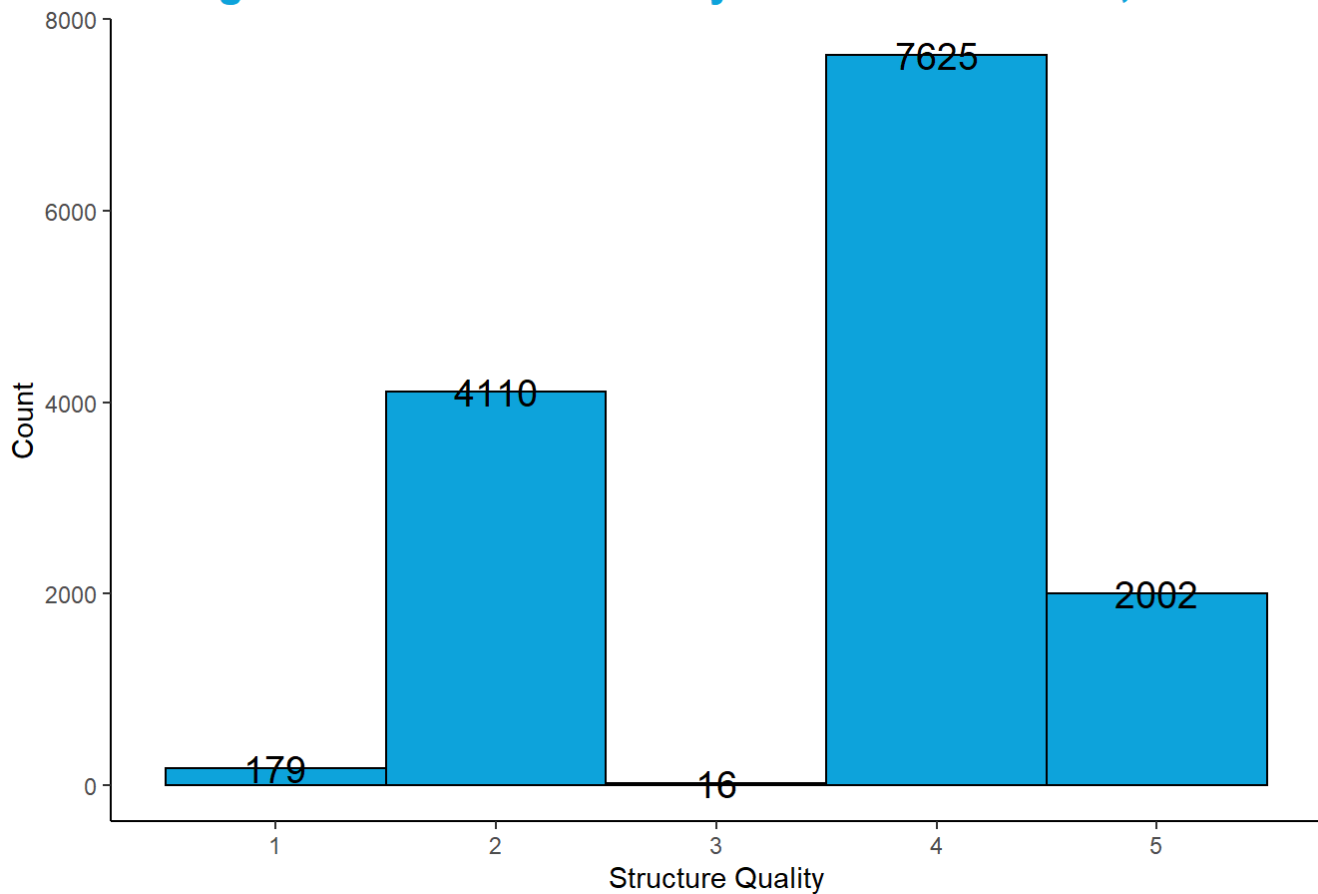
Although the structure quality labeled 3 has the highest average sale price at $2 million, that group only contains 16 homes. This means the mean is probably not reliable because of such a small sample size.

5 Conclusion/Limitations

In conclusion, we found that home price is moderately to strongly correlated to the square footage of a house. This can be seen by the construction of the scatter plot. Also, we can see that structure quality is an important factor in home price, as an increase in quality leads to a higher priced home. Some limitations to this data is the inability to compare Miami's home prices with other large cities or nearby towns. Also, I would like to get more info about the structure quality rating to get a better idea of what it means in more concrete terms. This data could also be used create a house price prediction model that could help people decide whether or not a house is overpriced.