BKSales

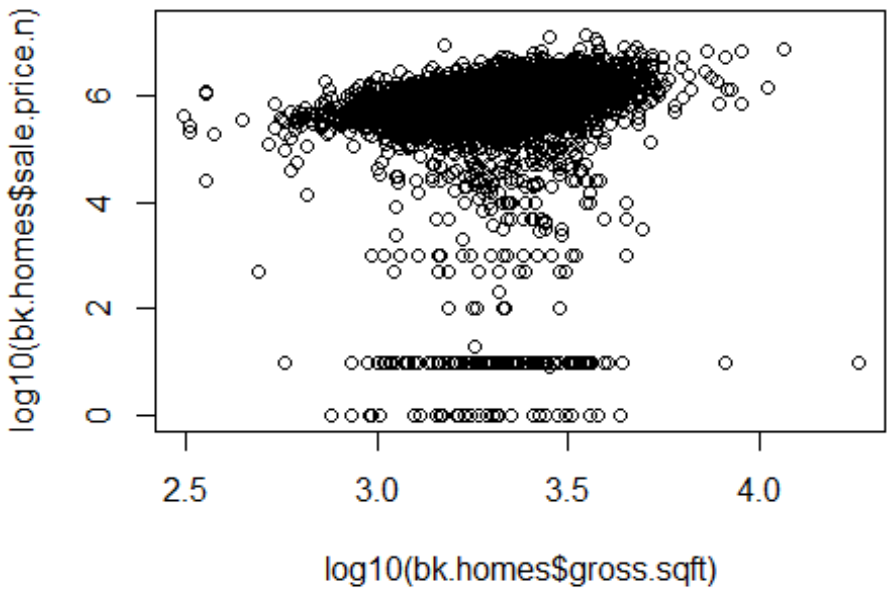
George Sturrock

June 10, 2017

The source data set is Rolling home sales data from Brooklyn, New York. This data is freely accessible on New York City's .gov website. The following code cleanses and prepares the data for analysis.

The following plot normalizes and illustrates the relationship between Gross Square Feet and Sales Price. The plot appears to show a positive correlation between Gross Square Feet and Sales price. As gross square footage increases, the sales price increases as well. However, there are several data point where the sales price is near or equal to zero. This muttles the picture and makes it hard to visually determine if a relationship between these two variables exists.

```
# TODO: complete plot() with log10 of
bk.homes$gross.sqft,bk.homes$sale.price.n
#   as above "bk.sale"
plot(log10(bk.homes$gross.sqft), log10(bk.homes$sale.price.n))
```



The following plot is similar to the prior plot. However, outliers with sales prices less that $5 have been removed in an attempt to provide a more accurate picture of the relationship

between Gross Square Feet and Sales Price. The plot with outliers removed clearly shows a positive relationship between gross square feet and sales price. A linear regression line has been added to the plot to verify this conclusion.

```
reg1 <- lm(log10(bk.homes$sale.price.n) ~ log10(bk.homes$gross.sqft))
plot(log10(bk.homes$gross.sqft),log10(bk.homes$sale.price.n))
abline(reg1)
```