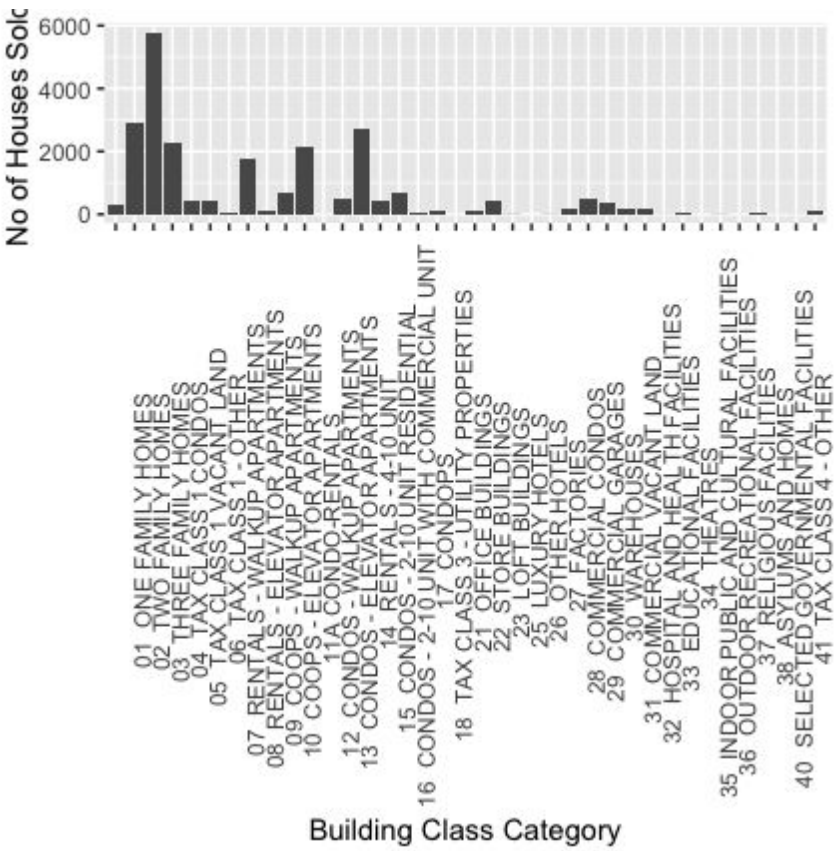
CSE 587 - Data Intensive Computing

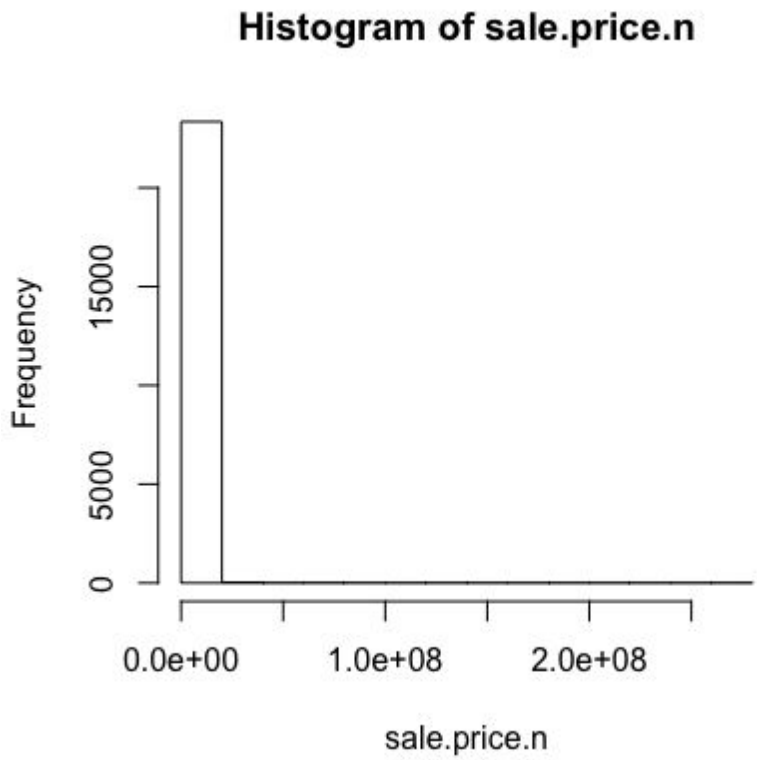
Problem 3 : Data economy: A real case study

Goals:The goals of this project was to take a piece of real world data provided and do some exploratory data analysis.
Data Source: Real Estate data for 5 Boroughs
Data Format: Excel files

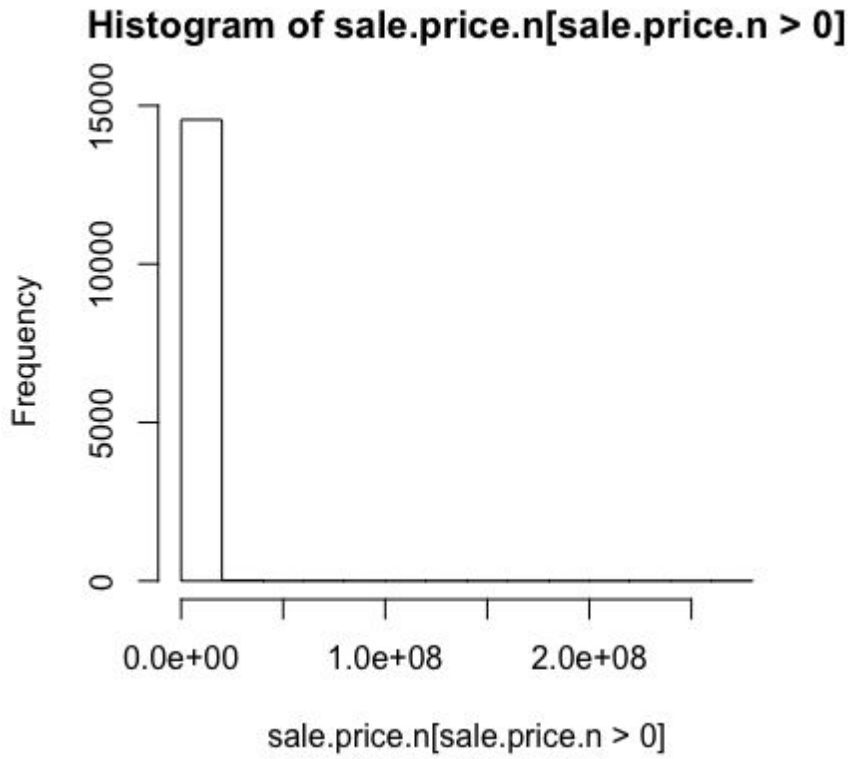
Analysis on Brooklyn Borough :

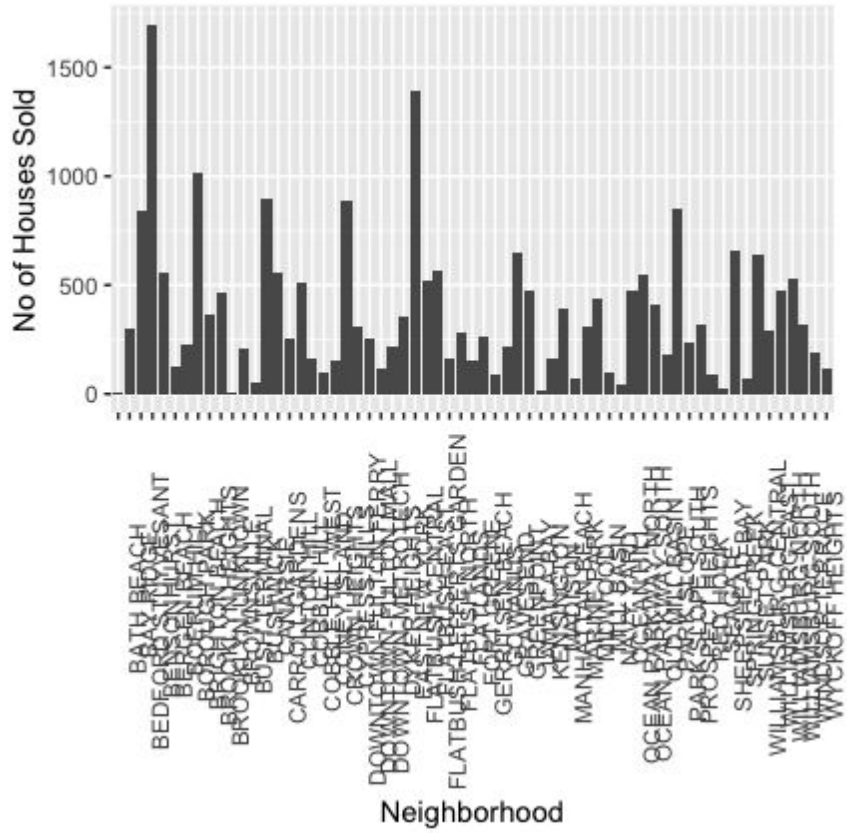


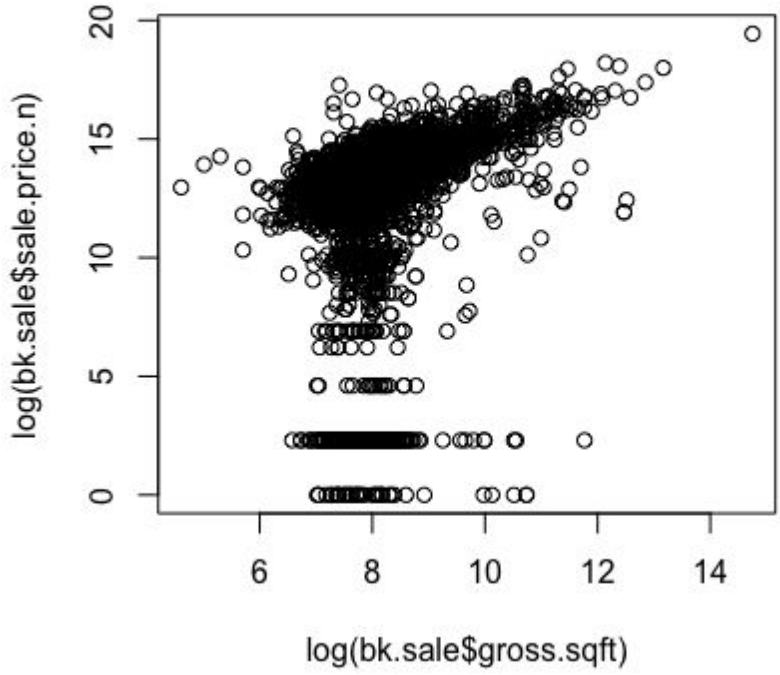
The below graph represents the sale prices of the houses that are within the 100,000$



The below graph depicts the sale price and the count of houses for which sale price > 0.

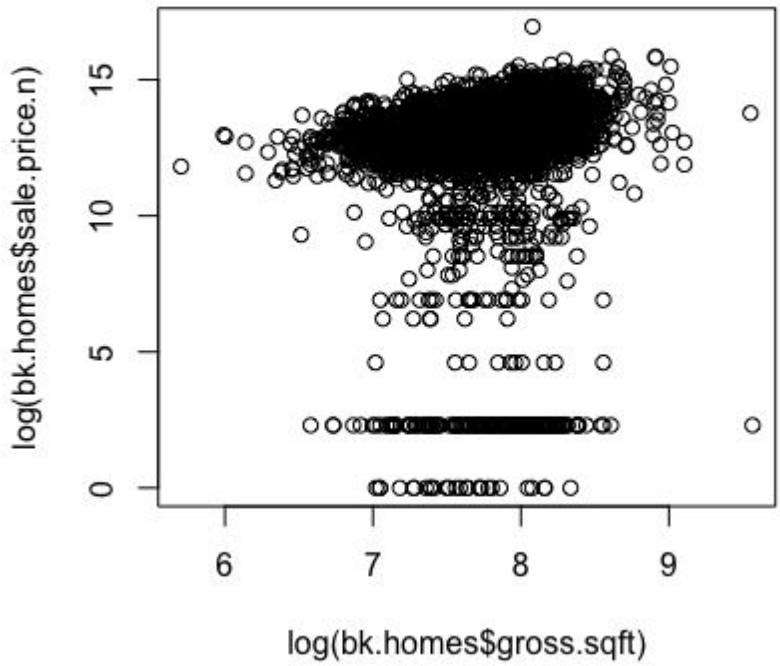






Above graph shows Sale price vs Gross square scatter plot for houses whose sale price is not 0.

This graph plots the Log values of the above graph

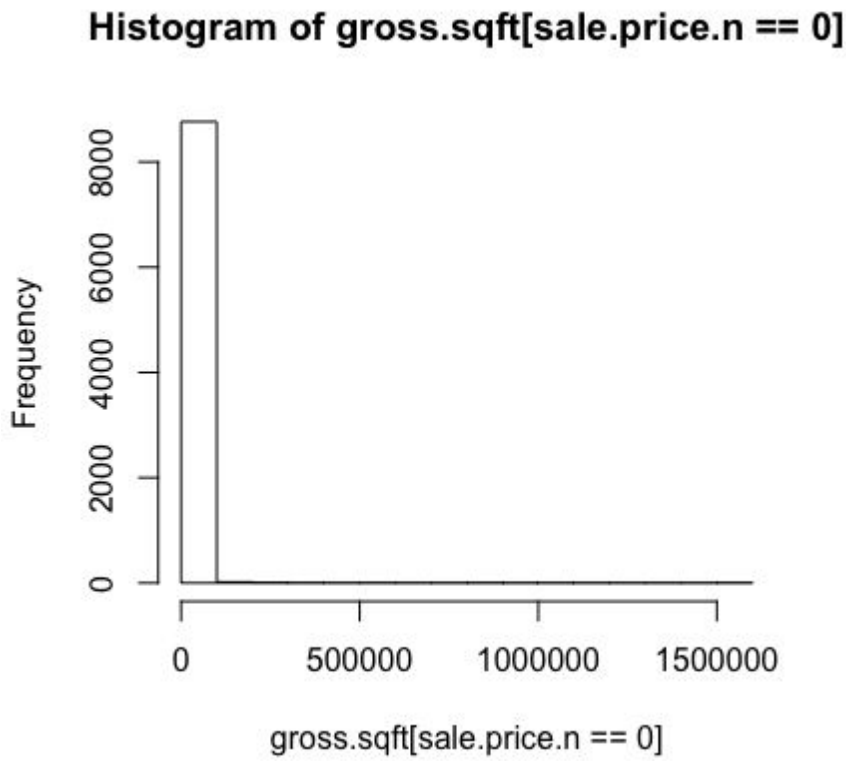


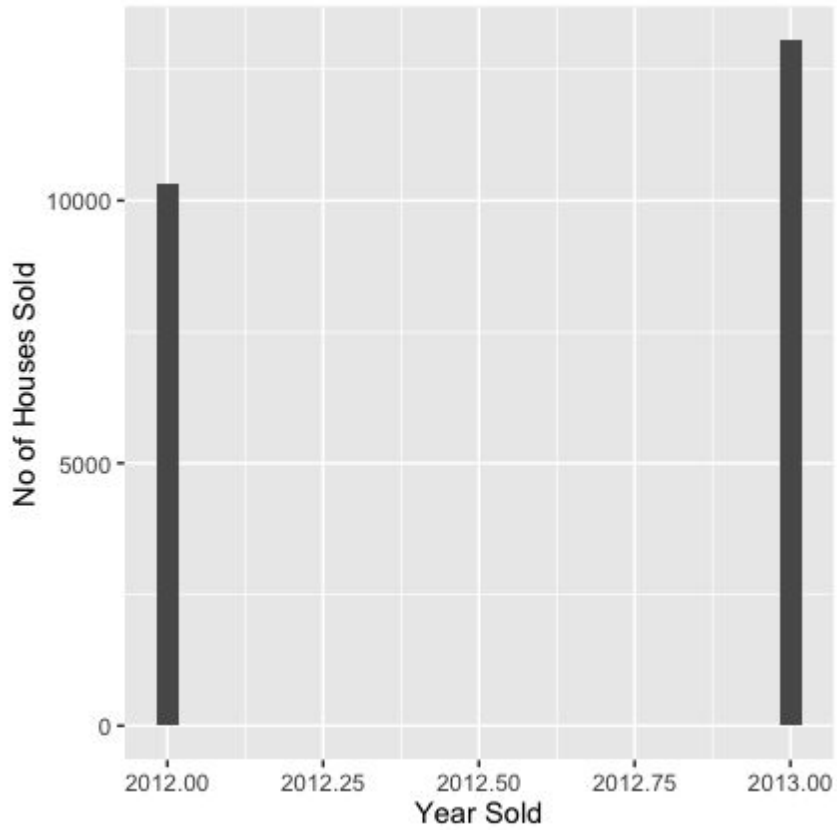
The below graph show Log of gross square feet vs sale price of the houses having Building class category of family.

log(bk.homes$sale.price.n)
6
8
10
12
14
16
6
7
8
9
log(bk.homes$gross.sqft)

bk.sale$sale.price.n
0.0e+00
1.0e+08
2.0e+08
0
500000
1500000
2500000
bk.sale$gross.sqft

Above depicts the gross square feet of houses whose sale price 0.

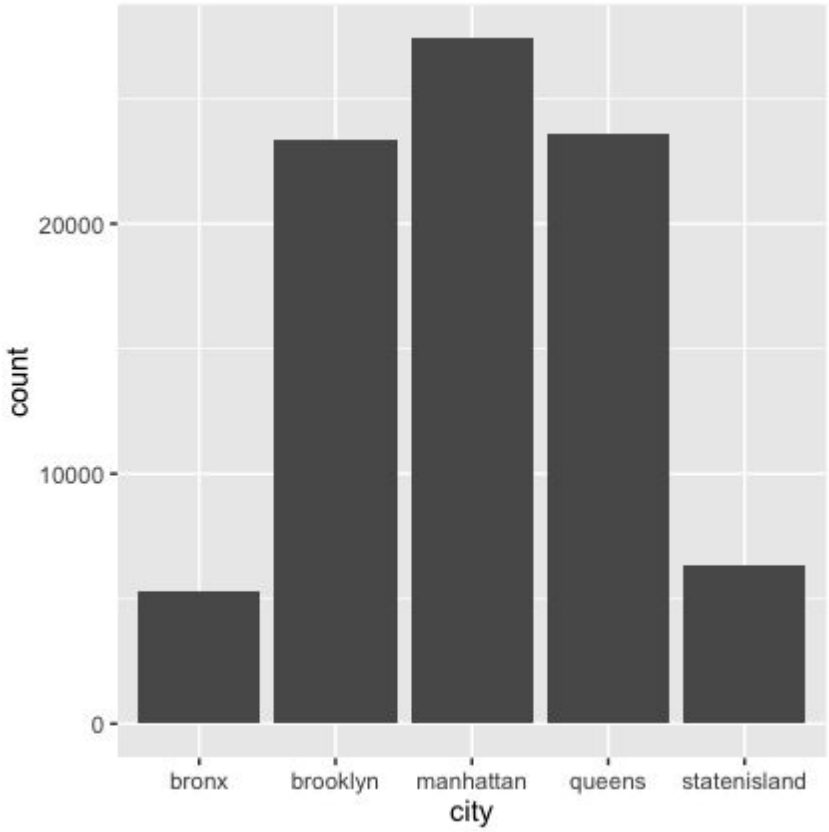


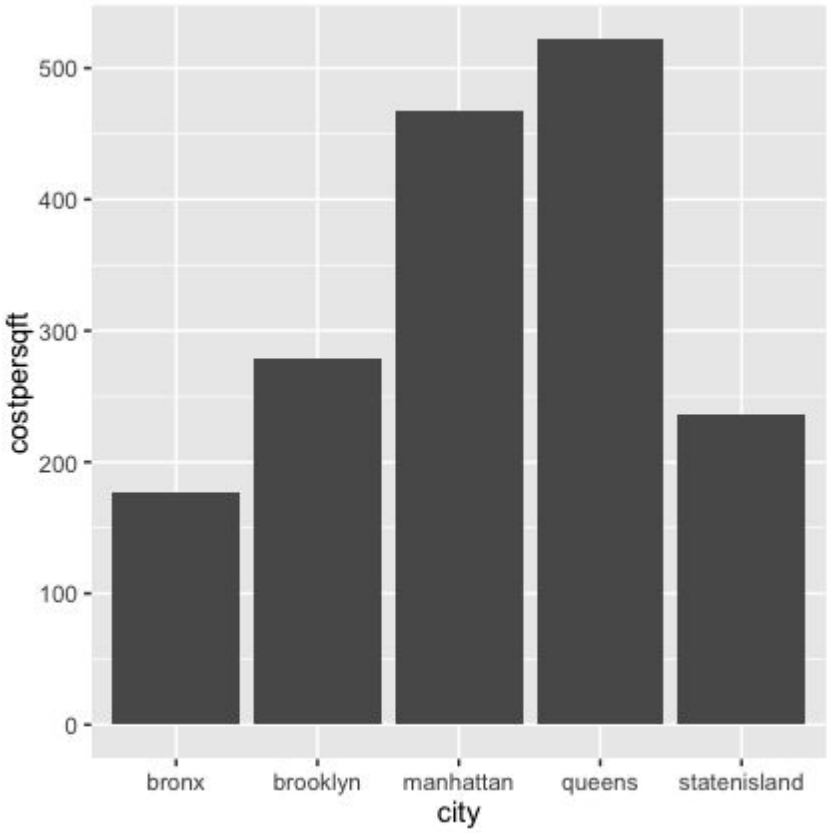


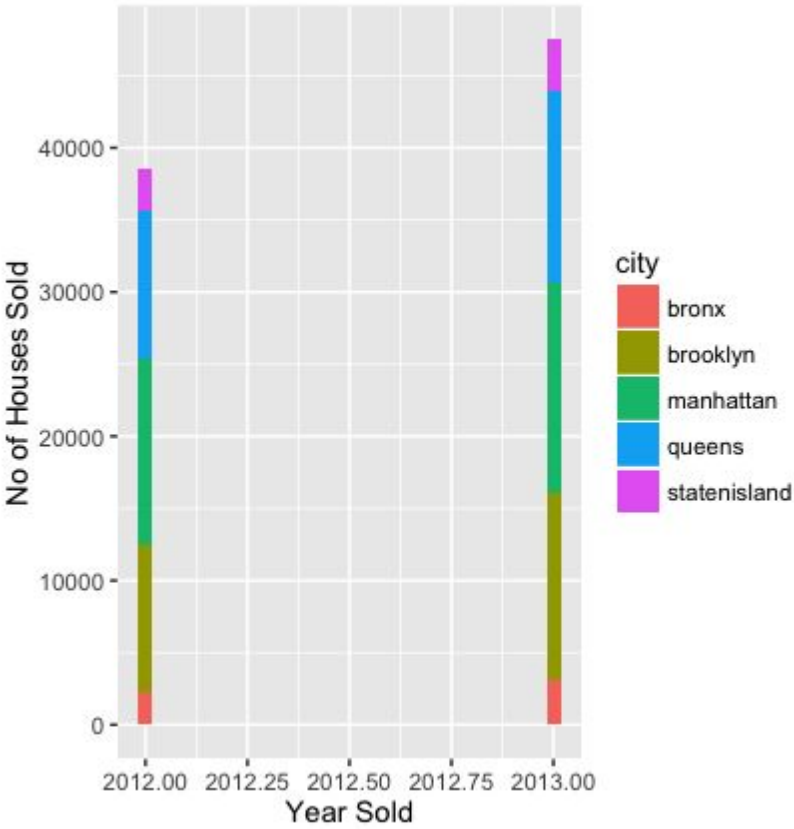
Observations :

1. The Number of houses sold in the year 2013 are more than the number of houses sold in the year 2012.
2. The Number of houses sold in the category “Three Class Family” are more.
3. The Number of houses sold in the neighbourhood Bedford stuyvesant are more.
4. We can see that most of the houses listings are in the size of (0,125000)

For Different Cities in NY (5 Boroughs)







Observations:

1. Cost per Square feet is highest in queens so queens real estate is expensive and Cost per Square feet is lowest in bronx so bronx is relatively cheap to live.
2. Number of houses sold are more in the year 2013 compared to the year 2012.
3. More Number of houses are sold in Manhattan and less number of houses are sold in bronx.