Midterm

February 21, 2019

Exploratory data analysis

1. Import Condominium Comparable Rental Income data for Bronx, Brooklyn, Manhattan, Queens, and Staten Island. Create a data frame, call it CondosNYC. Based on the table below your columns in CondosNYC should be relabeled using column names in **New column name**. Report number of rows, columns and head of the table.

Current column name	New column name
Neighborhood	Neighborhood
Building.Class	Class
Total.Units	Units
Year.Built	YearBuilt
Gross.SqFt	SqFt
Estimated.Gross.Income	Income
Gross.Income.per.SqFt	IncomePerSqFt
Estimated.Expense	Expense
Expense.per.SqFt	ExpensePerSqFt
Net.Operating.Income	NetIncome
Full.Market.Value	Value
Market.Value.per.SqFt	ValuePerSqFt
Boro	Boro

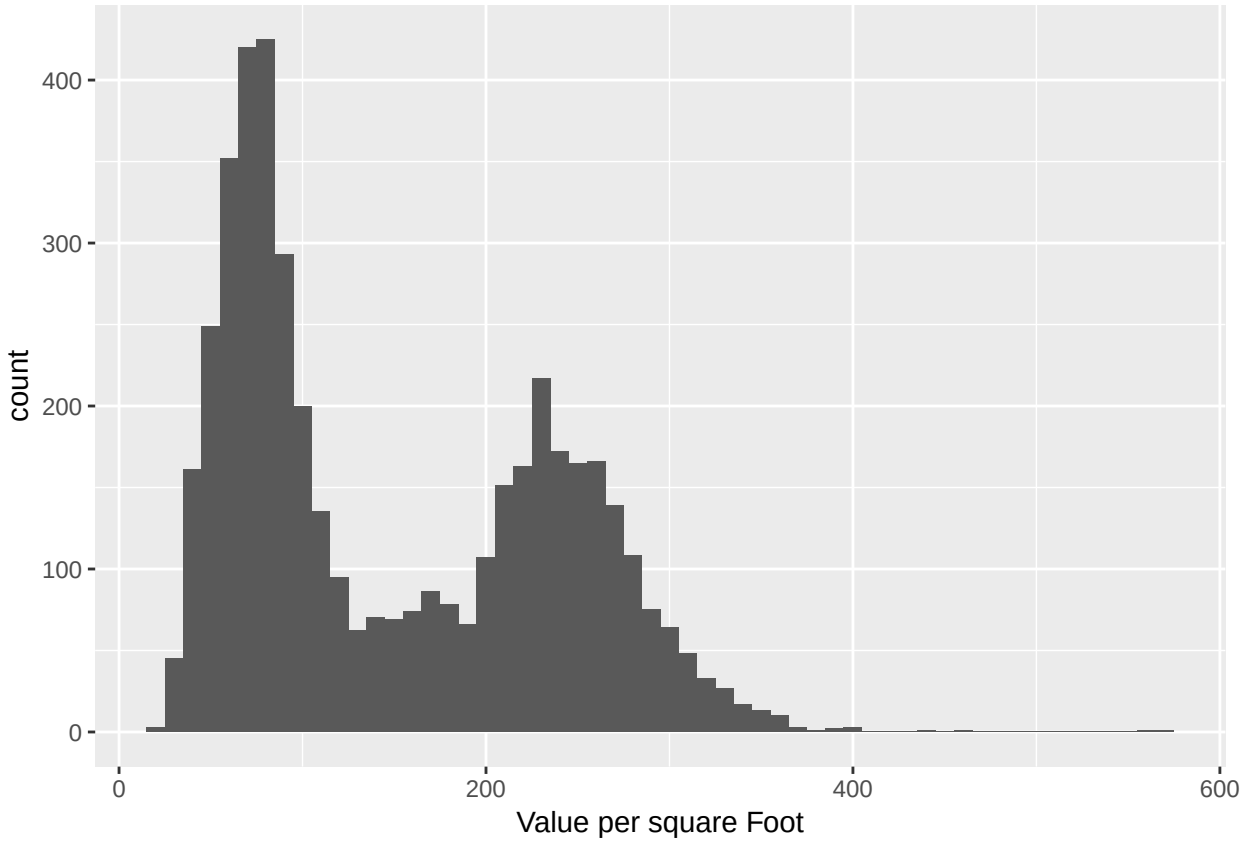
```
## [1] 22925    13
```

```
##            Neighborhood         Class Units YearBuilt    SqFt     Income
## 1    DOWNTOWN-METROTECH D4  -ELEVATOR  1026      1950 883,265 21,224,858
## 2  DOWNTOWN-FULTON MALL  C6  -WALK-UP    16      1905   9,784    325,318
## 3           BOERUM HILL D4  -ELEVATOR    61      1900 107,000  3,289,180
## 4           BOERUM HILL  C6  -WALK-UP    25      1914  18,530    530,514
## 5           BOERUM HILL  C6  -WALK-UP    25      1912  17,855    516,367
## 6 DOWNTOWN-FULTON FERRY D4  -ELEVATOR    11      1915  17,400    382,278
##   IncomePerSqFt    Expense ExpensePerSqFt  NetIncome      Value
## 1         24.03 11,129,139          12.60 10,095,719 81,260,000
## 2         33.25    121,909          12.46    203,409  1,637,000
## 3         30.74    959,790           8.97  2,329,390 18,749,000
## 4         28.63    231,625          12.50    298,889  2,406,000
## 5         28.92    163,909           9.18    352,458  2,837,000
## 6         21.97    164,952           9.48    217,326  1,724,000
##   ValuePerSqFt     Boro
## 1        92.00 Brooklyn
## 2       167.31 Brooklyn
## 3       175.22 Brooklyn
## 4       129.84 Brooklyn
## 5       158.89 Brooklyn
## 6        99.08 Brooklyn
```

2. Recreate the following graph. Hint: binwidth=10

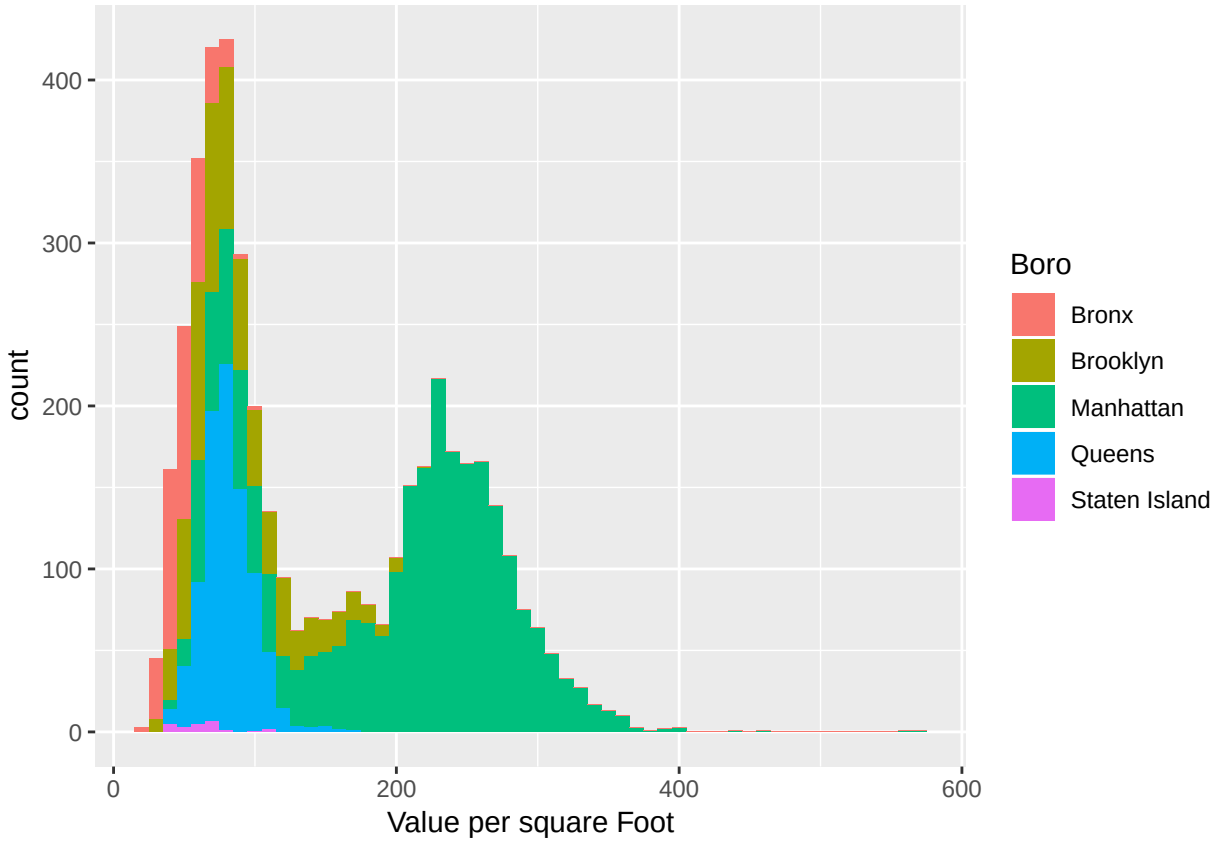
```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
## Warning: Removed 18354 rows containing non-finite values (stat_bin).
```



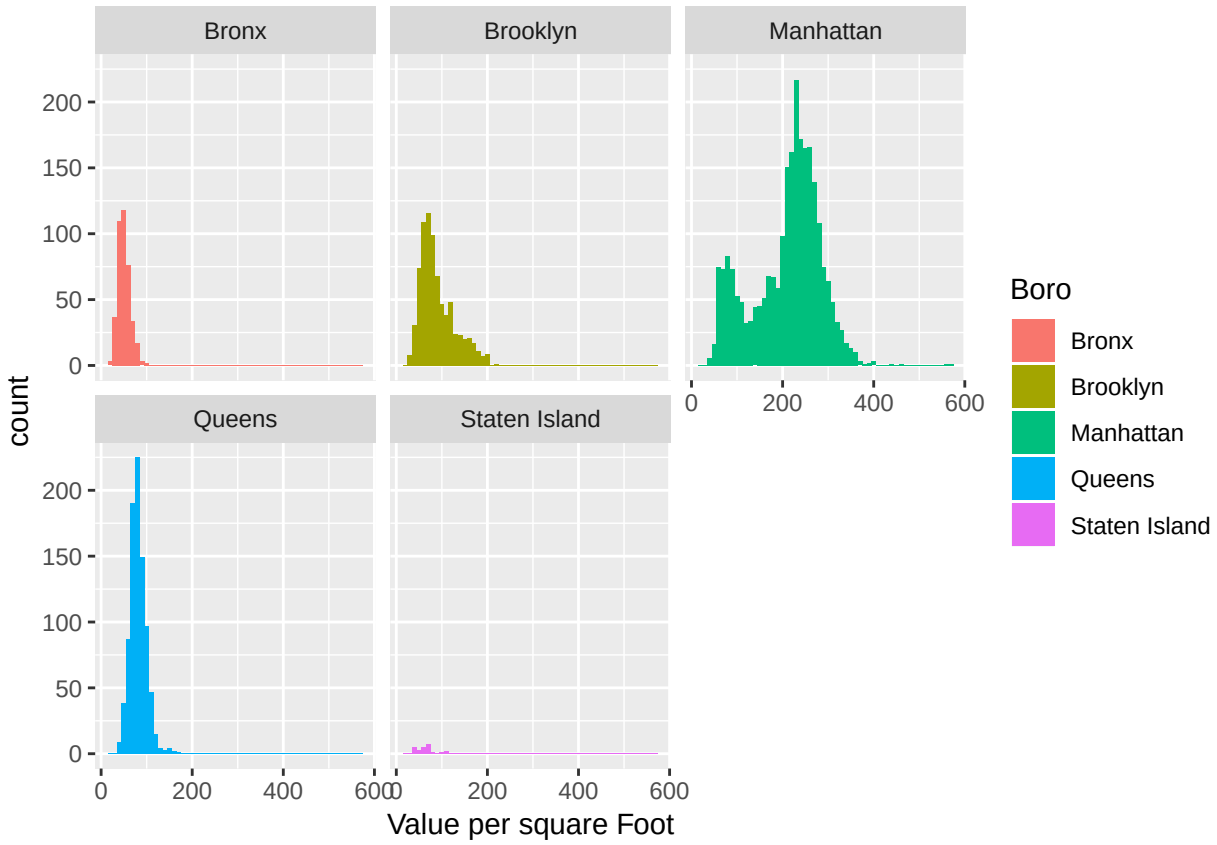
3. Recreate the following graph. Hint: binwidth=10

```
## Warning: Removed 18354 rows containing non-finite values (stat_bin).
```



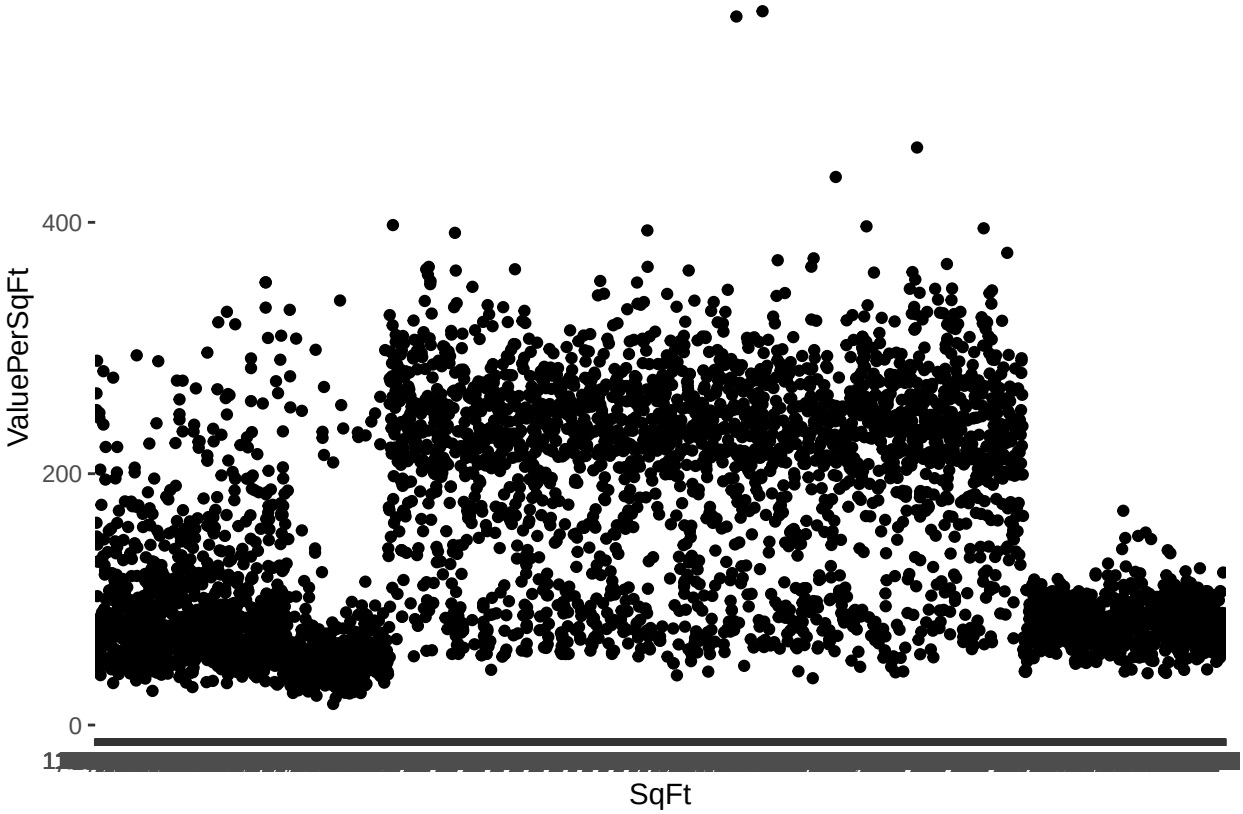
4. Recreate the following graph. Hint: binwidth=10

```
## Warning: Removed 18354 rows containing non-finite values (stat_bin).
```

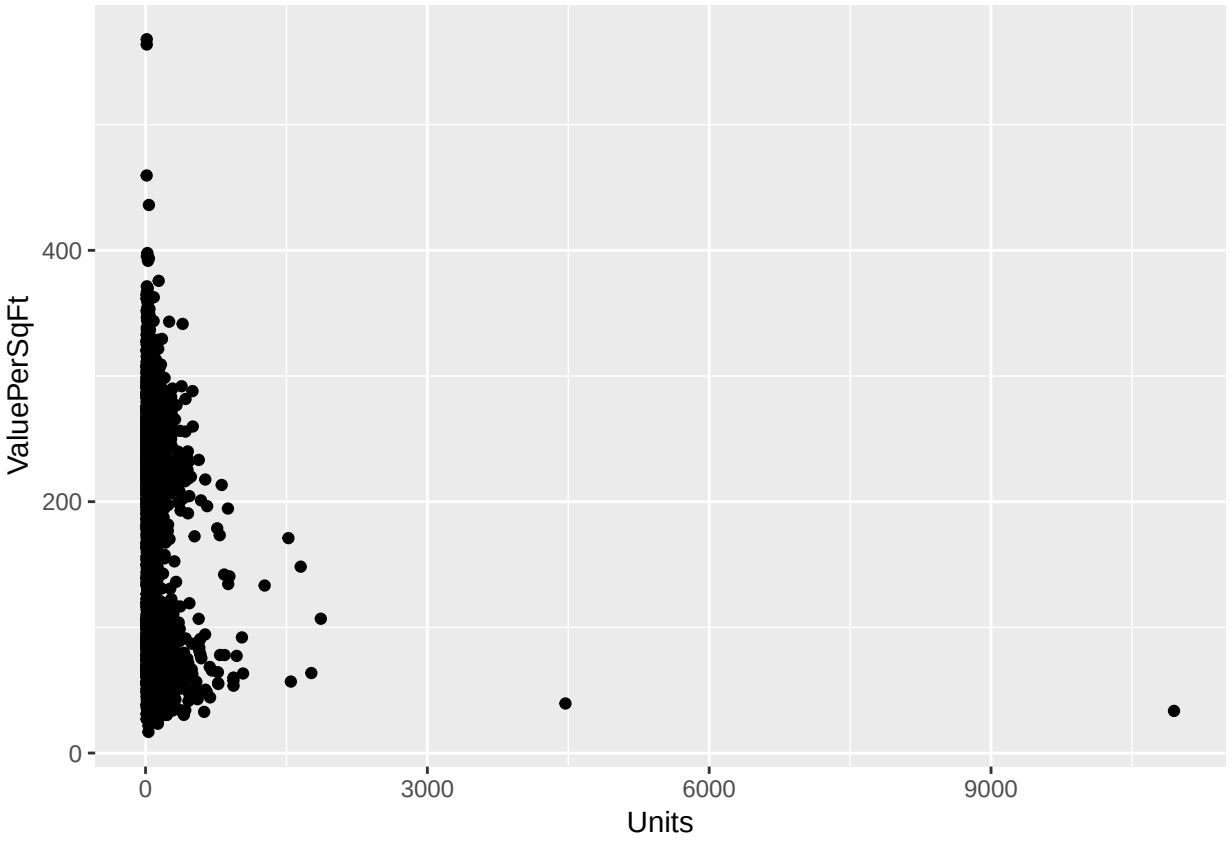


@.Write a code to recreate the following graphs:

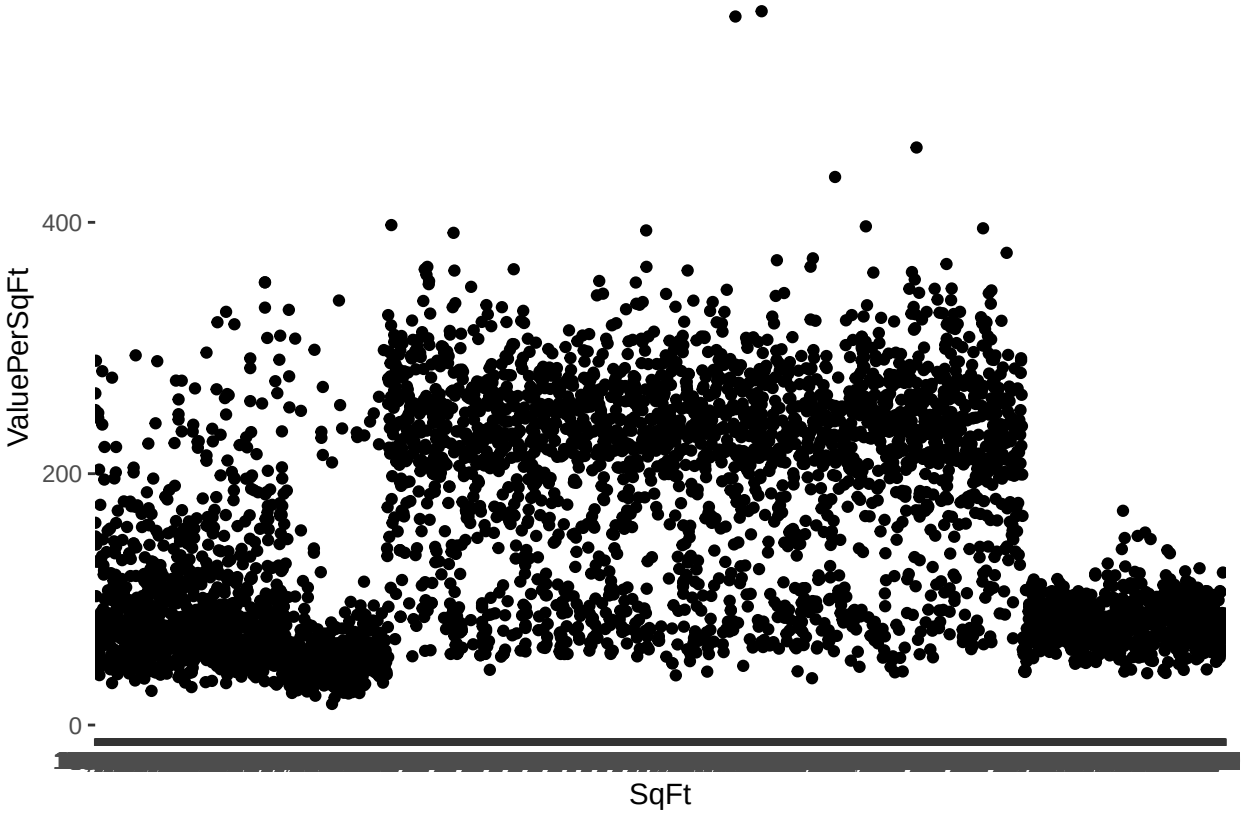
```
## Warning: Removed 18354 rows containing missing values (geom_point).
```



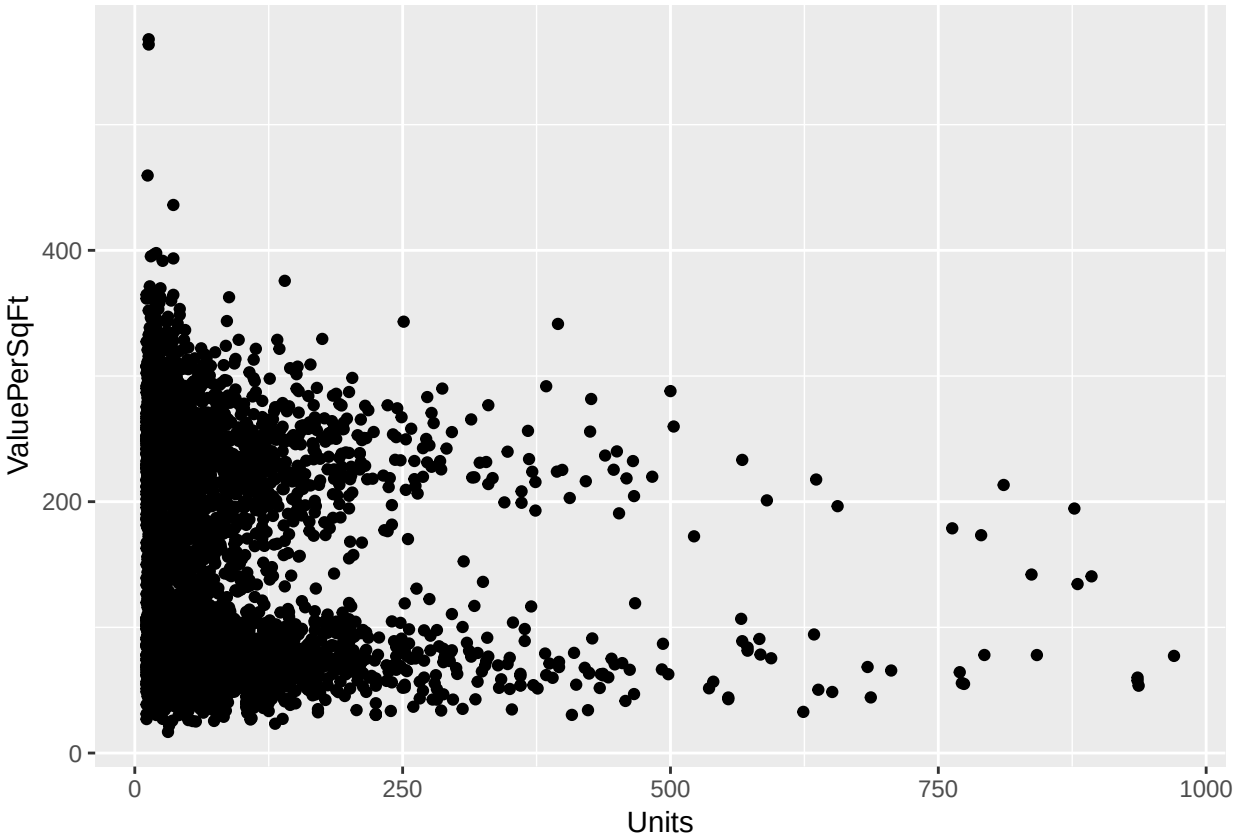

```
## Warning: Removed 18354 rows containing missing values (geom_point).
```




```
## Warning: Removed 18354 rows containing missing values (geom_point).
```




```
## Warning: Removed 18354 rows containing missing values (geom_point).
```



5. What insights do we gain from these graphs? How will they effect further analysis and modeling of the data?
6. For each Boro create a table that contans comlumns that represent smallest, average, median, and largest value of a condo.
7. For each condo classification in each Boro create a table that contans comlumns that represent smallest, average, median, and largest value of condos that have 100 or more units.
8. Which Boro hacs largest number neighborhoods? What is the largest number of neighborhoods?