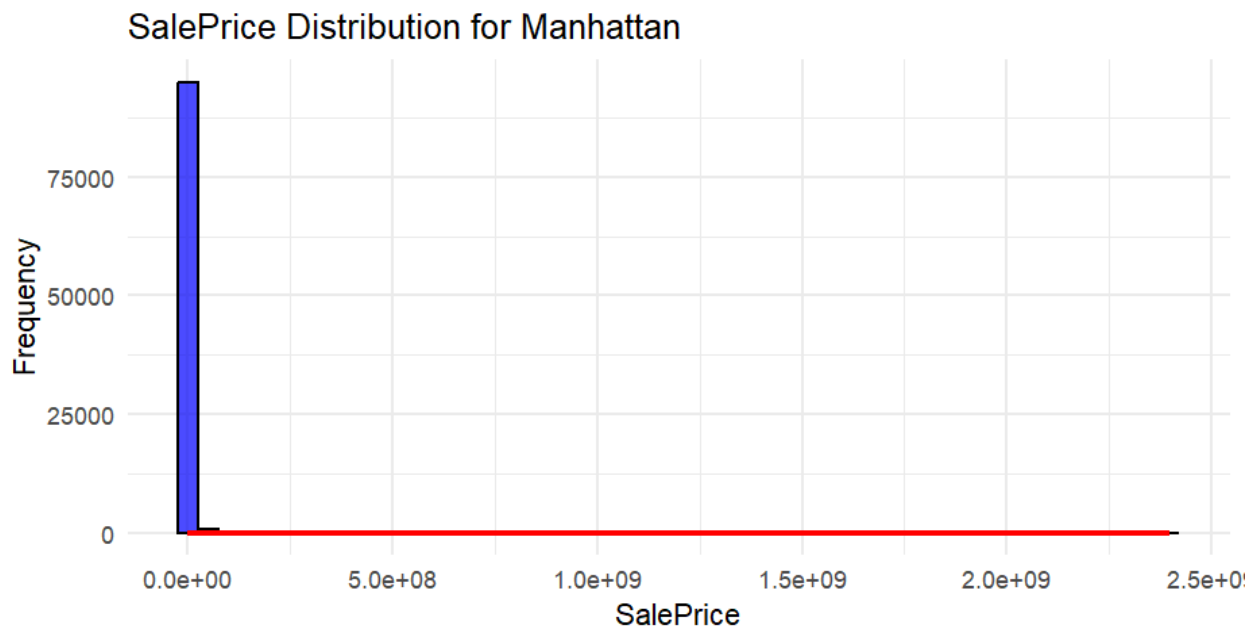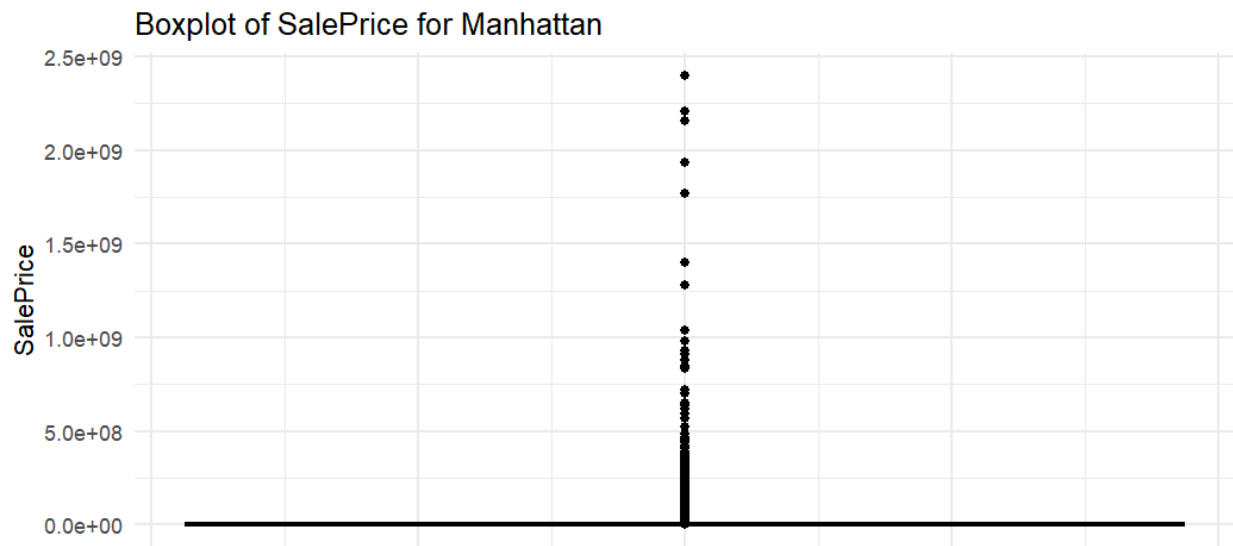## HI PROFESSOR 🙂

Note -

It took super super long to retrieve most plots because it was loading for foreverrrr so for most of them i screenshotted the scripts I received

SalePrice Distribution for Manhattan

Boxplot of SalePrice for Manhattan

```
u poincs
> cat("Number of outliers in SalePrice:", nrow(outliers), "\n")
Number of outliers in SalePrice: 9617
> print(outliers[, c("SALE.PRICE")])  # Corrected column name
  [1]    43300000  148254147   11000000  591800000   99350000
  [6]   115000000  148254147   14500000    4750000    6900000
 [11]    21350000   23200000   74159182  175000000    4800000
 [16]   175000000    4980000    4820000    5500000    9000000
 [21]     5500000   20150000    4800000    5224735    4800000
 [26]    10100295   10100295   10100295   10100295   10100295
 [31]     7250000    5880000   18750000    6950000    4625000
 [36]     7300000    7100000    8000000    6750000   16200000
 [41]    18000000    9300000    5945000    6150000    4750000
 [46]     5050000   11100000   11345000    7450000   13750000
 [51]     5300000    4850000    8750000    5170000    5575000
 [56]     7079588    7900000    5298158    4851831   10003200
 [61]     5250000    5450000    7450000    4635000    6330000
 [66]     4775000   10950000   25000000   19600000    4700000
 [71]   205500000    5298158    4851831   10003200    6600000
 [76]     8400000    6670000    5900000    7250000   82664485
 [81]     5500000    7400000   15750000   14000000    7650000
 [86]     7350000    8470000    7750000    4600000   19650000
 [91]    19650000   10700000   12370800   12370800   14900000
 [96]    21250000    5793511    5956763    6000000    6600000
[101]     8000000   14000000    7200000   18000000    4760000
[106]    31725000    4600000    6405206    5900000    5203258
[111]     4850000    5508733    8500000    5995000   35500000
[116]    81500000    6255000   23500000   60500000   30000000
[121]    35000000    6850000  850000000   56500000    6200000
```

SalePrice vs Gross Square Feet with Outliers Highlighted (Manhattan)
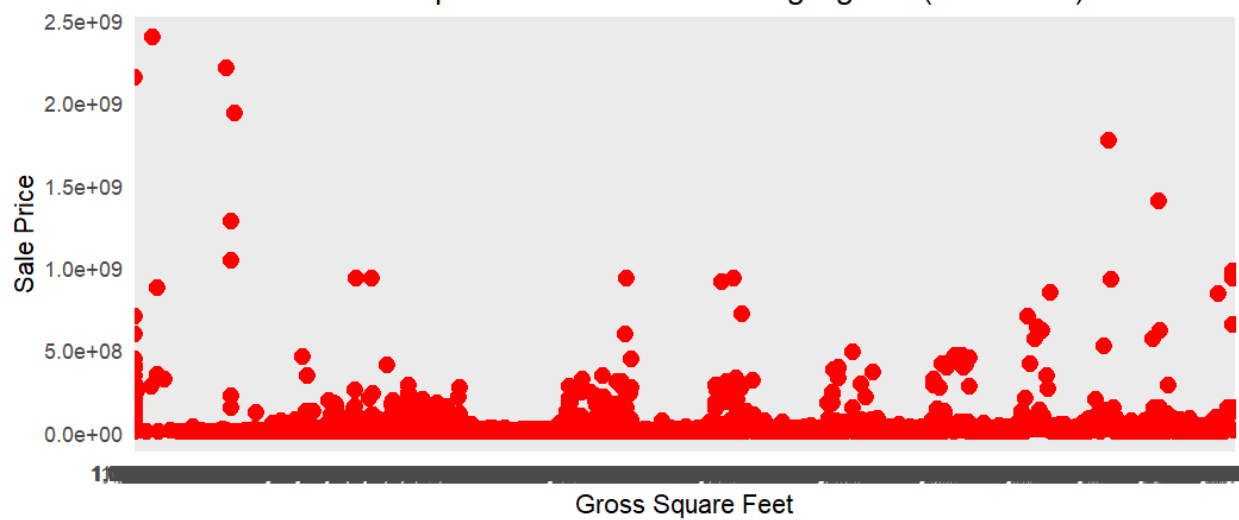
```
Call:
lm(formula = SALE.PRICE ~ LAND.SQUARE.FEET + GROSS.SQUARE.FEET +
    YEAR.BUILT + NEIGHBORHOOD + RESIDENTIAL.UNITS, data = nyc_dat
_small)

Residuals:
      Min        1Q    Median        3Q       Max
 -17689703  -1582497   -631624    317558 2140113753

Coefficients:
                                            Estimate Std. Error
(Intercept)                                4.598e+05  1.305e+06
LAND.SQUARE.FEET                           9.312e+00  1.126e+01
GROSS.SQUARE.FEET                          7.263e+02  1.030e+02
YEAR.BUILT                                 4.479e+02  1.852e+02
NEIGHBORHOODCHELSEA                        1.031e+06  1.360e+06
NEIGHBORHOODCHINATOWN                     -5.936e+04  1.681e+06
NEIGHBORHOODCIVIC CENTER                   1.494e+06  1.504e+06
NEIGHBORHOODCLINTON                       -1.496e+05  1.466e+06
NEIGHBORHOODEAST VILLAGE                  -9.995e+04  1.564e+06
NEIGHBORHOODFASHION                        2.598e+06  1.607e+06
NEIGHBORHOODFINANCIAL                     -3.467e+05  1.440e+06
NEIGHBORHOODFLATIRON                       3.355e+06  1.456e+06
NEIGHBORHOODGRAMERCY                      -3.034e+05  1.378e+06
NEIGHBORHOODGREENWICH VILLAGE-CENTRAL      8.245e+05  1.393e+06
NEIGHBORHOODGREENWICH VILLAGE-WEST         1.365e+06  1.360e+06
NEIGHBORHOODHARLEM-CENTRAL                -1.076e+06  1.334e+06
NEIGHBORHOODHARLEM-EAST                   -8.742e+05  1.481e+06
NEIGHBORHOODHARLEM-UPPER                  -1.324e+06  1.537e+06
NEIGHBORHOODHARLEM-WEST                   -3.203e+05  2.756e+06
NEIGHBORHOODINWOOD                        -1.341e+06  1.772e+06
```

```
> # Print the nearZeroVar metrics to see which columns are being
emoved
> print(nzv)
                              freqRatio percentUnique zeroVar
GROSS.SQUARE.FEET             38.275833    1.11725905   FALSE
RESIDENTIAL.UNITS             2.074959     0.06731899   FALSE
YEAR.BUILT                    1.220965     0.03995835   FALSE
SALE.PRICE                    29.606608    7.30273272   FALSE
BUILDING.CLASS.AT.TIME.OF.SALE 1.612055    0.02184915   FALSE
                                 nzv
GROSS.SQUARE.FEET                TRUE
RESIDENTIAL.UNITS               FALSE
YEAR.BUILT                      FALSE
SALE.PRICE                       TRUE
BUILDING.CLASS.AT.TIME.OF.SALE  FALSE
```


Sale Price Distribution by Gross Square Feet in Manhattan