

Math 23c Final Project

Derek Chang

05/14/2021

Overview of the data - House Listing and Assess Data in MA

```
## # A tibble: 6 x 13
##   city ls_year price assess ls_month lot_size res_area house_age style
##   <fct> <fct>   <dbl>  <dbl> <fct>      <dbl>   <dbl>    <int> <fct>
## 1 ABIN~ 2011   424250 410100 6          30039.    2136      19 colo~
## 2 ABIN~ 2011   180000 243900 6          17407.    1190      56 cape
## 3 ABIN~ 2011   223000 206000 6          25247.    1000      51 ranch
## 4 ABIN~ 2011   409000 407700 4          20020.    2642      29 colo~
## 5 ABIN~ 2011   174000 214400 9           7619.     980     120 conv~
## 6 ABIN~ 2011   260000 237800 7          12179.    1346     110 conv~
## # ... with 4 more variables: num_rooms <fct>, stories <dbl>,
## #   inc_medianhhd_blkgrp <dbl>, age_med_blk <dbl>

## # A tibble: 6 x 9
##   pop_blk black_share vacant_share latitude longitude distance_firest~
##   <dbl>   <dbl>      <dbl>   <dbl>   <dbl>      <dbl>
## 1     71     1.41        0     42.1    -71.0      1575.
## 2    469     0.426      0     42.1    -71.0      766.
## 3    405     0.988      2.01    42.1    -71.0     1651.
## 4    500      1        4.07    42.1    -71.0      984.
## 5     28      0        0     42.1    -70.9      887.
## 6     65     6.15        0     42.1    -70.9      268.
## # ... with 3 more variables: distance_hospital <dbl>, distance_police <dbl>,
## #   distance_prischool <dbl>

## # A tibble: 6 x 7
##   distance_library distance_pubsch~ distance_townha~ distance_train vacant high
##   <dbl>           <dbl>           <dbl>           <dbl> <lgl> <lgl>
## 1      782.         846.         511.         2660. FALSE TRUE
## 2     1107.        1034.        1060.        2957. FALSE FALSE
## 3     2107.         248.        1919.        2851. FALSE FALSE
## 4     1990.         964.        1893.        2315. FALSE TRUE
## 5     2015.        1769.        2224.        2264. FALSE FALSE
## 6     1291.         432.        1459.        1067. FALSE FALSE
## # ... with 1 more variable: near_school <chr>
```

Graph

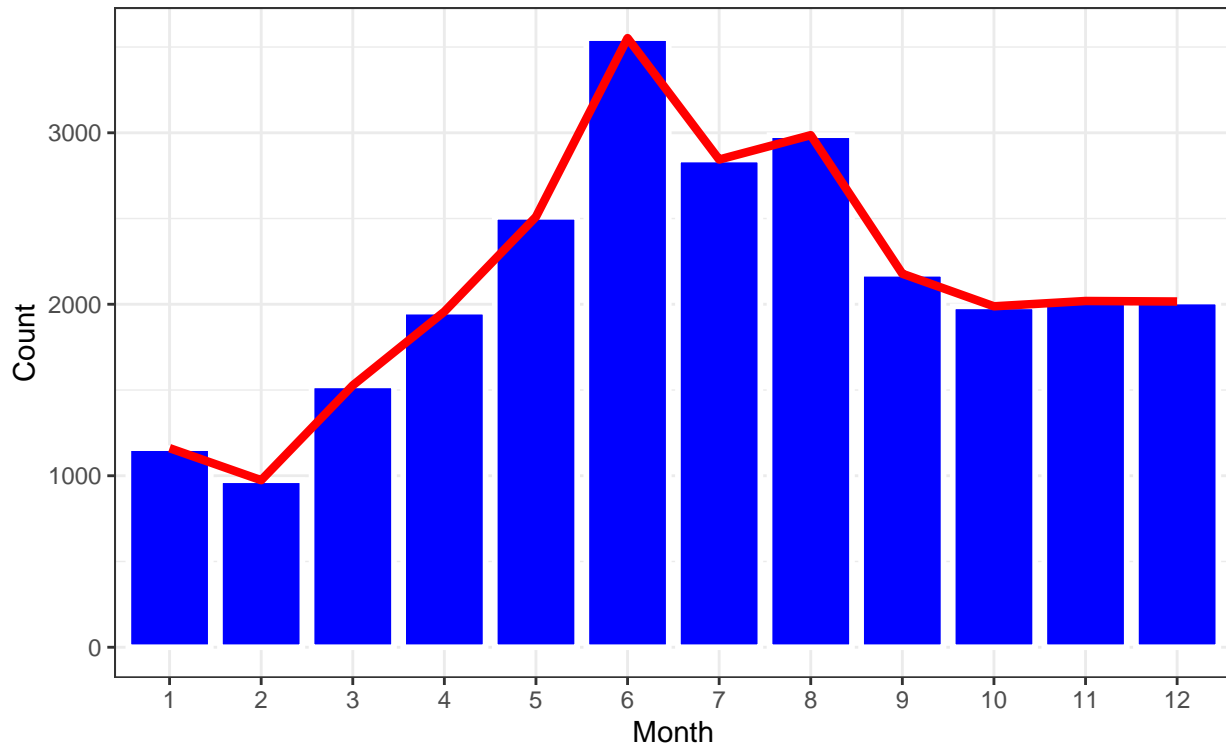
##

1 2 3 4 5 6 7 8 9 10 11 12

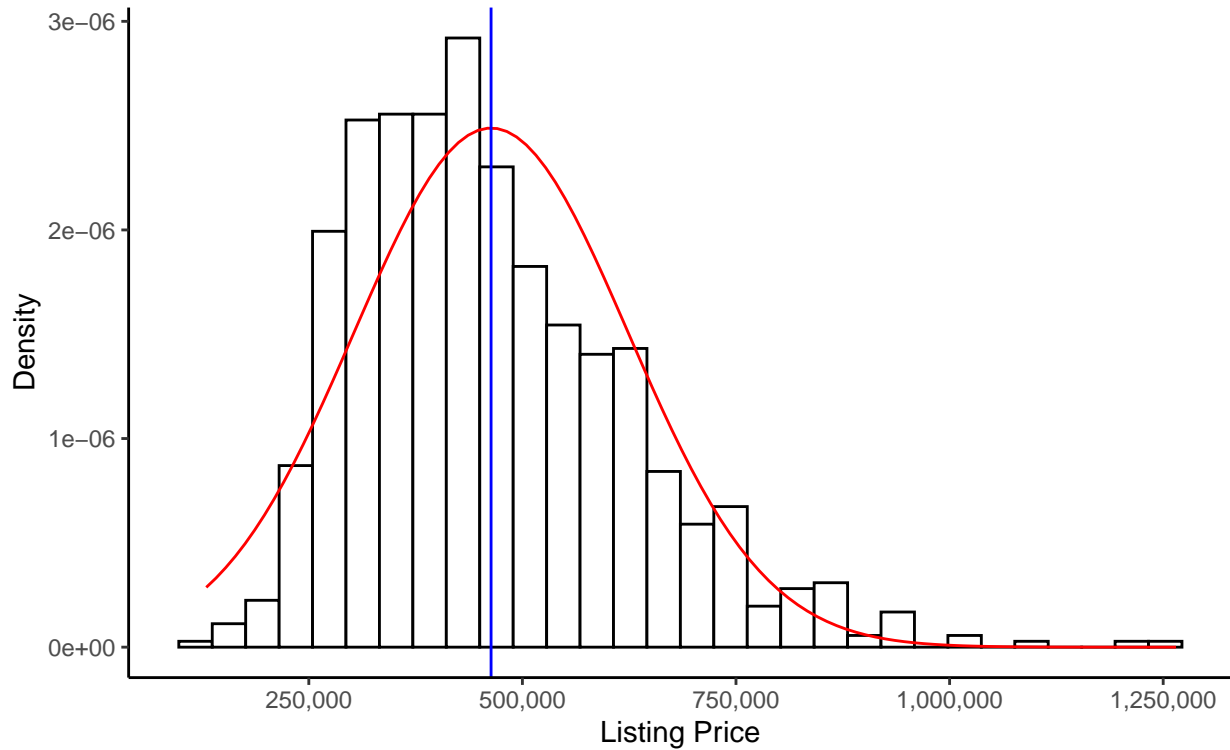
1161 974 1527 1957 2510 3552 2844 2986 2178 1988 2019 2016

Monthly House Listings

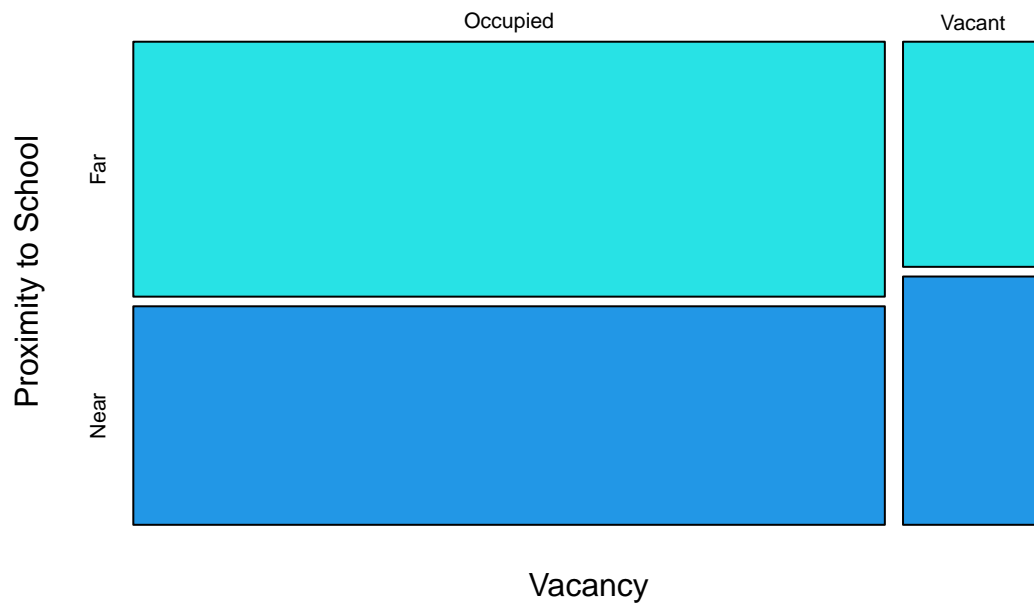
for Towns in Massachusetts



Listing Price Distribution for Towns around My Area



Vacancy vs School Proximity



Analysis

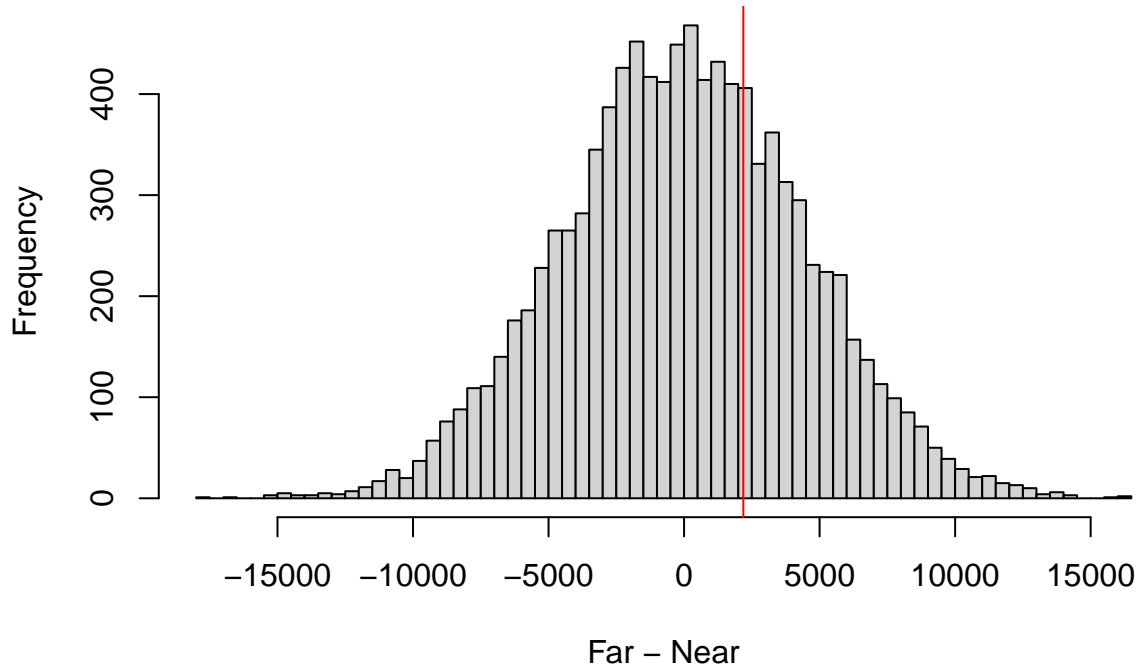
```
## [1] 437322.1
```

```
## [1] 435136.3
```

```
## [1] 2185.805
```

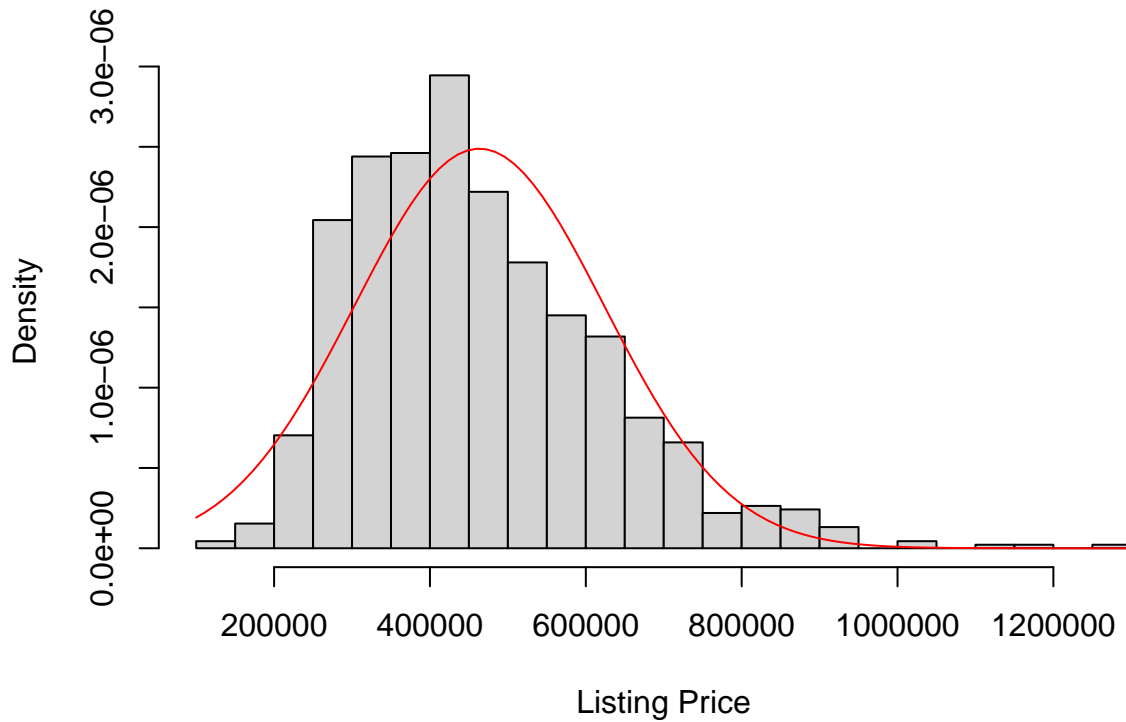
```
## [1] -6.87601
```

Price Difference based on School Proximity



```
## [1] 0.3102
```

Distribution of House Listing Price



```
## [1] -Inf 257879.4 328407.2 379262.8 422717.0 463332.6 503948.2 547402.4
## [9] 598257.9 668785.8 Inf

## [1] 4 4 5 9 6 7 10 10 9 9 5 7 3 3 2 7 4 9 3 4 8 8 5 8 7
## [26] 9 10 10 5 5 10 6 9 9 3 2 9 5 2 10 6 4 4 4 10 7 9 9 7 3
## [51] 10 10 10 8 10 4 7 7 9 9 5 6 10 3 10 4 5 10 7 10 2 7 5 10 10
## [76] 5 7 4 6 7 3 7 10 6 5 3 8 9 7 10 7 4 10 10 5 3 7 5 3 3
## [101] 9 5 4 4 6 6 4 9 9 10 10 6 5 3 6 10 5 10 10 10 7 5 9 4 7
## [126] 4 2 6 6 6 6 6 5 9 7 7 6 6 10 9 4 8 9 4 10 7 5 2 5 5
## [151] 10 3 8 4 5 6 5 3 8 3 10 5 5 10 8 10 10 4 10 6 6 7 7 5 5
## [176] 4 5 10 6 7 10 10 7 10 4 10 10 10 4 10 7 10 9 8 10 4 3 9 7 8
## [201] 8 10 10 10 8 5 7 5 10 4 5 8 6 10 4 5 6 5 9 4 4 10 8 5 10
## [226] 8 8 9 8 5 5 9 8 10 7 9 5 10 8 5 5 2 5 3 6 2 9 4 8 6
## [251] 4 3 3 5 5 5 6 3 10 3 9 2 5 1 4 5 4 6 3 4 4 6 9 2 2
## [276] 8 9 4 5 1 3 2 1 7 3 5 6 5 8 3 5 4 2 3 3 10 4 3 1 1
## [301] 3 4 3 2 4 4 6 5 8 8 5 7 3 2 10 3 2 6 5 5 6 3 7 4 10
## [326] 2 2 2 4 8 3 4 6 10 7 2 2 5 7 5 6 5 7 2 3 2 1 9 4 10
## [351] 6 5 3 3 6 2 3 3 3 3 2 8 7 4 3 3 1 4 9 3 4 1 6 2 3
## [376] 5 10 9 2 4 3 4 5 7 1 3 10 3 3 6 2 4 5 1 9 2 3 2 9 3
## [401] 3 5 7 2 2 6 1 7 9 2 8 1 7 3 2 5 2 7 2 3 6 3 3 2 1
## [426] 3 3 3 4 7 2 3 6 4 1 3 2 7 4 4 7 5 3 1 3 3 3 7 3 4
## [451] 4 1 4 4 2 8 2 5 2 2 3 3 3 1 6 2 3 5 5 2 2 2 1 2 3
## [476] 2 3 3 6 2 2 2 1 1 2 2 3 2 4 2 2 2 3 2 3 2 4 2 1 3
## [501] 2 4 2 2 1 2 1 3 2 2 3 2 2 2 2 3 3 8 4 5 5 2 2 4 2
## [526] 2 2 1 2 2 1 2 6 1 2 3 2 2 4 2 1 1 2 3 1 2 3 3 1 4
## [551] 2 1 1 2 2 1 2 2 1 4 3 4 2 4 3 2 2 1 1 2 2 2 4 3 2
## [576] 2 3 1 1 2 1 2 2 1 6 3 5 5 8 4 6 8 10 8 5 4 10 5 10 3
## [601] 2 5 3 6 3 3 6 2 10 2 9 8 8 6 3 5 7 2 9 10 6 4 5 7 2
## [626] 8 4 9 8 10 4 3 7 4 4 6 2 7 9 2 3 9 4 2 4 4 3 4 2 5
## [651] 2 9 7 10 10 3 6 2 7 9 6 8 7 6 2 5 7 7 7 2 2 7 2 5 9
```

```

## [676] 2 2 9 6 4 6 6 2 2 4 6 10 5 8 4 6 5 3 6 7 5 4 2 3 7
## [701] 8 2 3 7 10 2 4 4 3 8 4 9 9 3 4 2 4 6 7 5 10 4 10 1 7
## [726] 5 5 8 3 4 5 5 2 4 7 4 2 7 6 9 6 10 10 7 6 2 1 9 9 6
## [751] 3 3 8 4 5 8 9 2 10 1 6 10 10 6 7 2 9 9 8 8 7 9 3 1 2
## [776] 3 8 9 10 9 9 6 8 9 7 4 2 10 4 10 3 7 9 8 9 4 6 9 9 7
## [801] 10 7 8 5 6 10 4 10 6 3 8 3 2 6 2 7 9 2 8 4 4 10 4 7 7
## [826] 10 9 3 3 5 2 8 9 8 7 9 2 10 8 8 9 10 4 9 9 9 2 1 3 2
## [851] 10 2 2 10 10 5 8 5 10 2 8 2 5 10 10 5 1 10 2 10 9 5 5 9 9
## [876] 8 6 6 3 10 4 8 5 9 6 4 6 2 7 8 4 10 6 10 4 5 4 8 8 2
## [901] 4 5 9 9 2 8 4 7 4 3

## [1] 46 146 117 110 98 77 76 64 76 100

## [1] 91 91 91 91 91 91 91 91 91 91

## [1] TRUE

## [1] 83.42857

## [1] 2.748034e-15

##
##          FAR NEAR
## FALSE 413 354
## TRUE 68 75

## [1] 0.1571429

## [1] 0.4714286

## [1] 910

##          [,1]      [,2]
## [1,] 361.58571 405.41429
## [2,] 67.41429 75.58571

## [1] 13.8406

## [1] 0.0001989898

##
##          FALSE TRUE
## FALSE 8825 4026
## TRUE 9652 3209

## [1] 0.743956

## [1] 0.1571429

## [1] 910

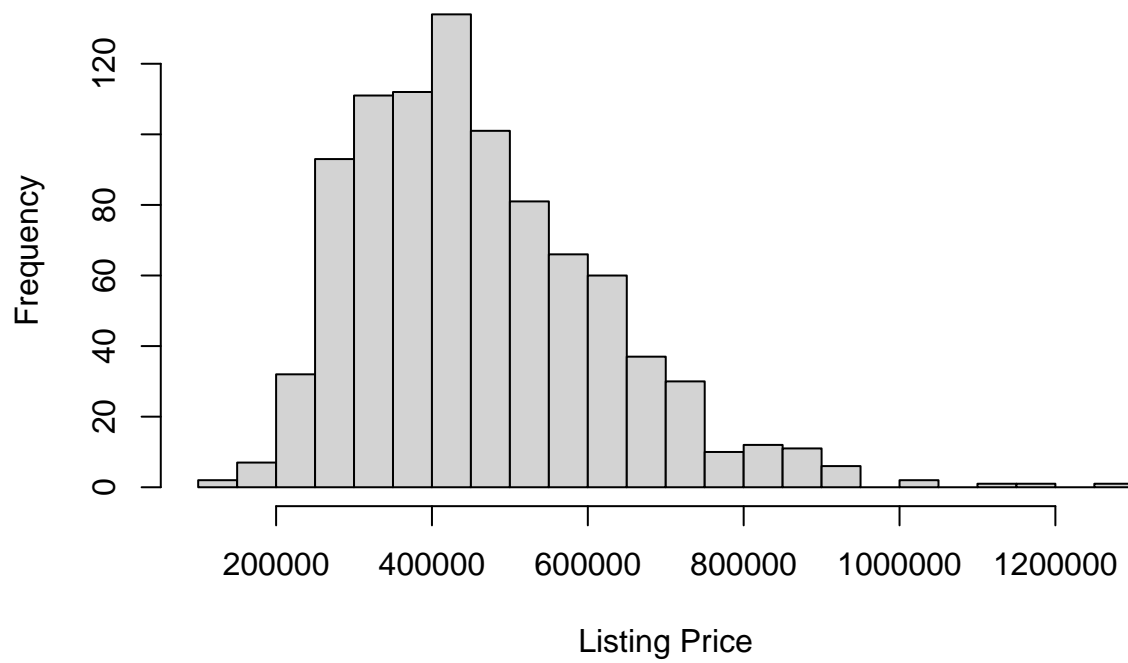
##          [,1]      [,2]
## [1,] 36.61429 196.3857
## [2,] 106.38571 570.6143

## [1] 3260.894

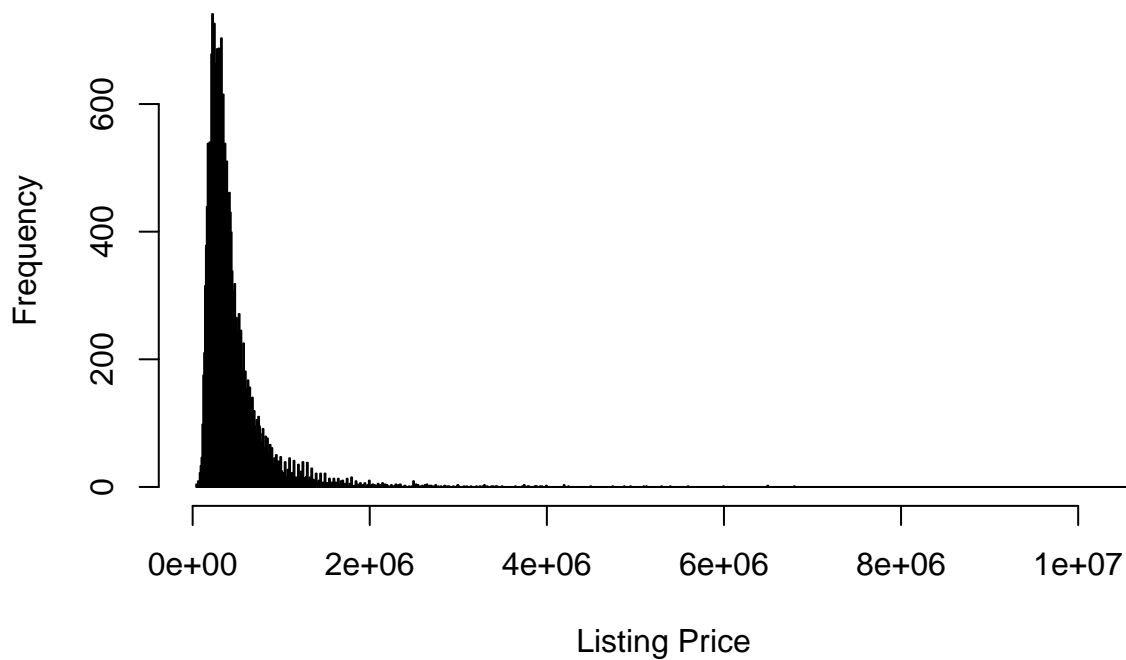
## [1] 0

```

Histogram of Listing Price



```
## [1] 910
## [1] 344560.9
## [1] 435927.9
## [1] 463332.6
## [1] 2.399275
## [1] 0.008213796
## [1] 2201.672
## [1] 2202.157
```



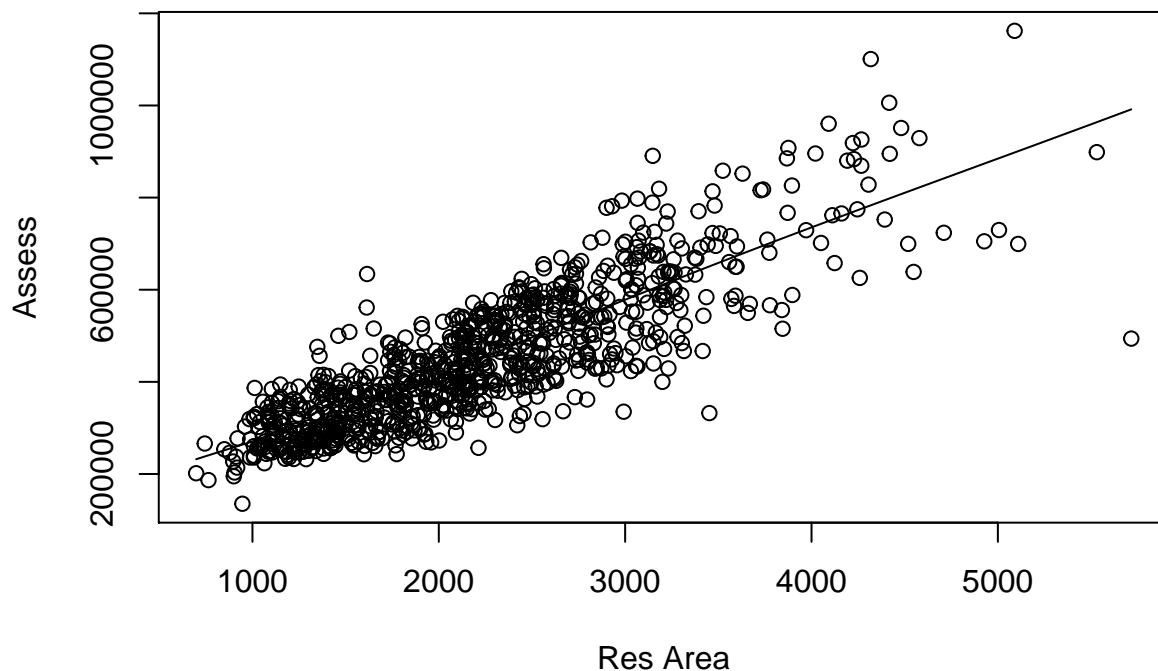
[1] 10600000

[1] 45000


```
## [1] 5.547497
## [1] 435927.9
## [1] 360881.6

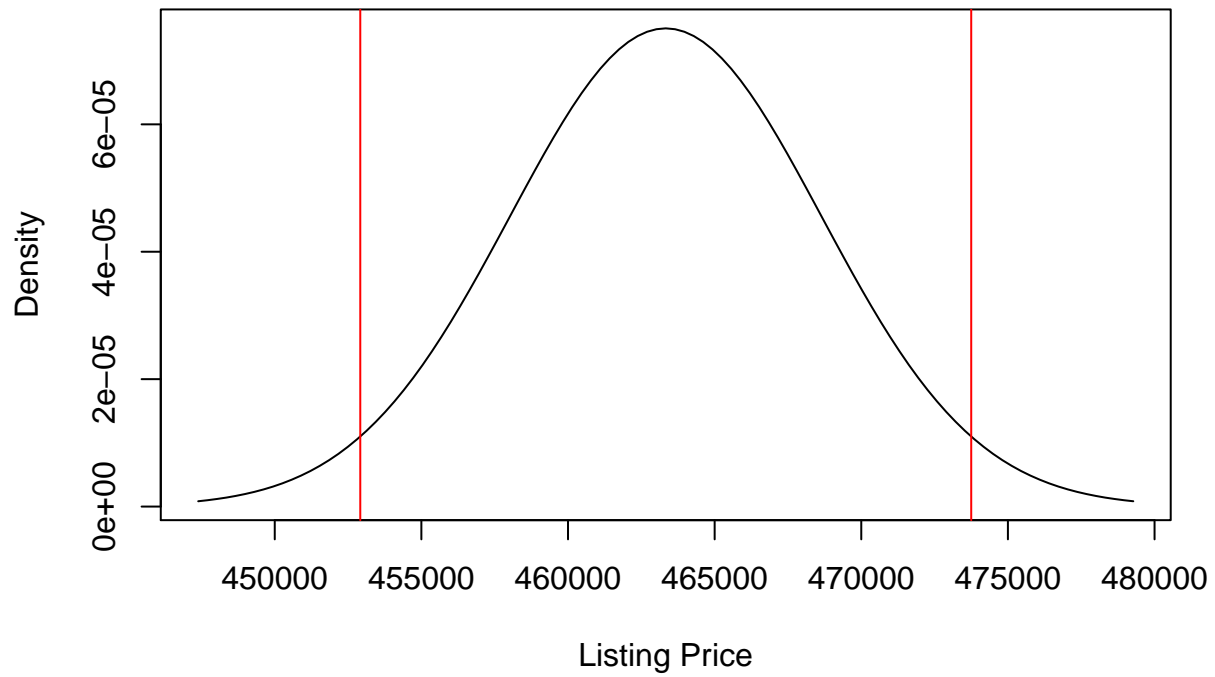
##
## Call:
## lm(formula = local$assess ~ local$res_area, data = local)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -505489  -52513   -3076    54325   320321
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.039e+05  7.744e+03  13.42  <2e-16 ***
## local$res_area 1.567e+02  3.301e+00  47.46  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 80550 on 908 degrees of freedom
## Multiple R-squared:  0.7127, Adjusted R-squared:  0.7124
## F-statistic: 2252 on 1 and 908 DF, p-value: < 2.2e-16
```

Relationship between Assess and Res Area



```
## [1] 0.8441997
## [1] 452916.5
## [1] 473748.7
```

Listing Price Distribution for Local Towns



25%

247500

75%

509000

25%

343125

75%

557875