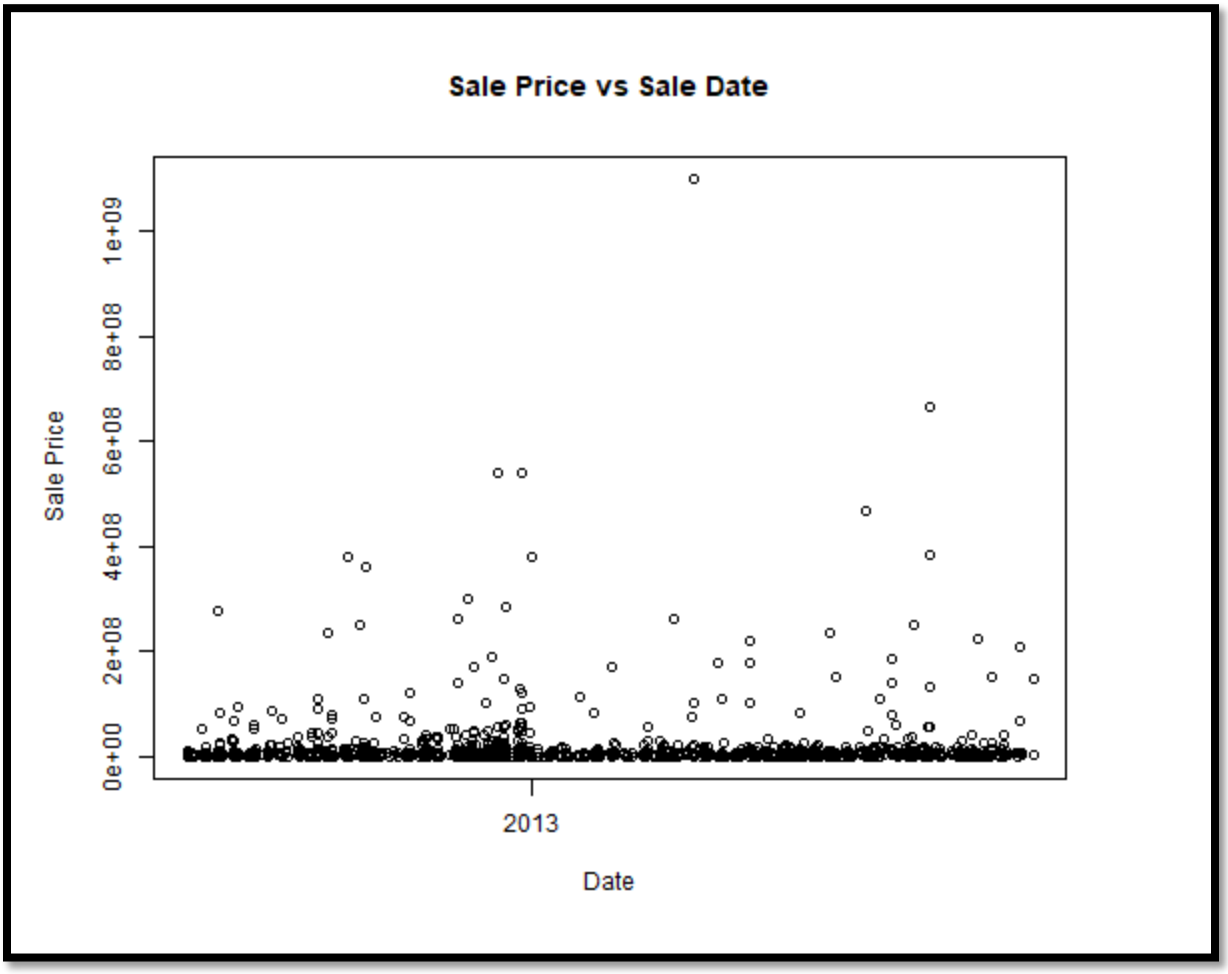
Connor Deakin
10/05/18
MAT 345

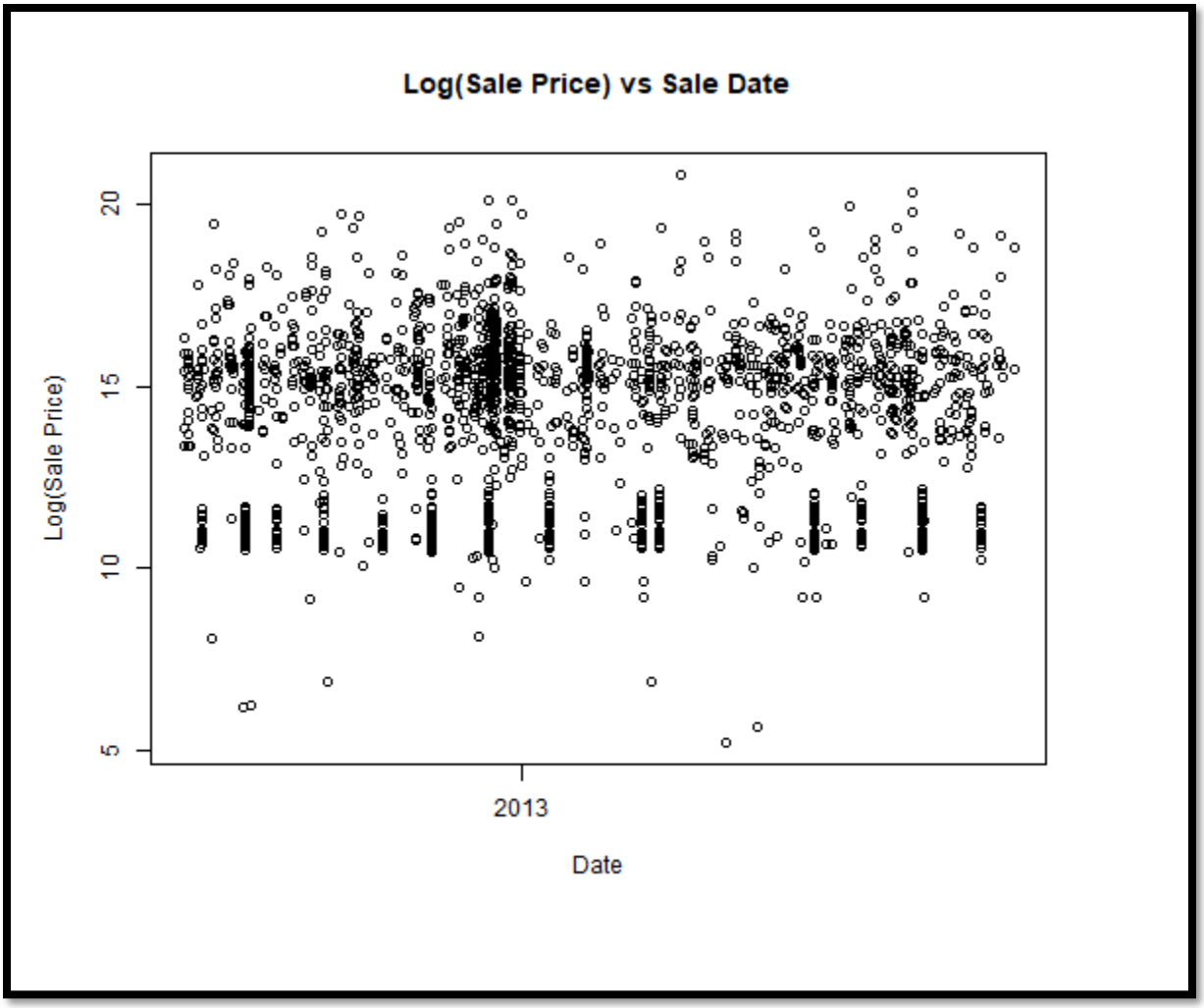
Project 1

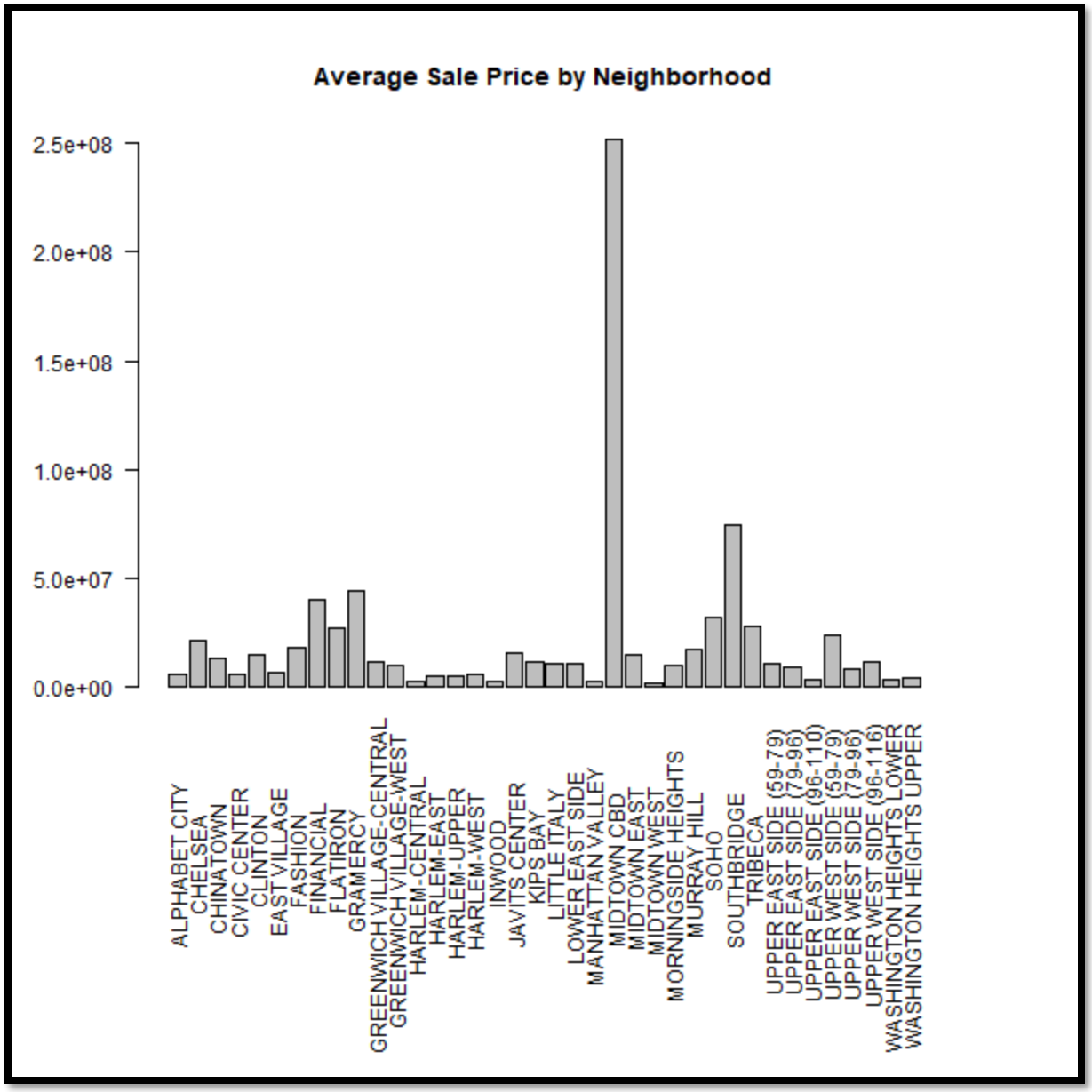
1[a-d]

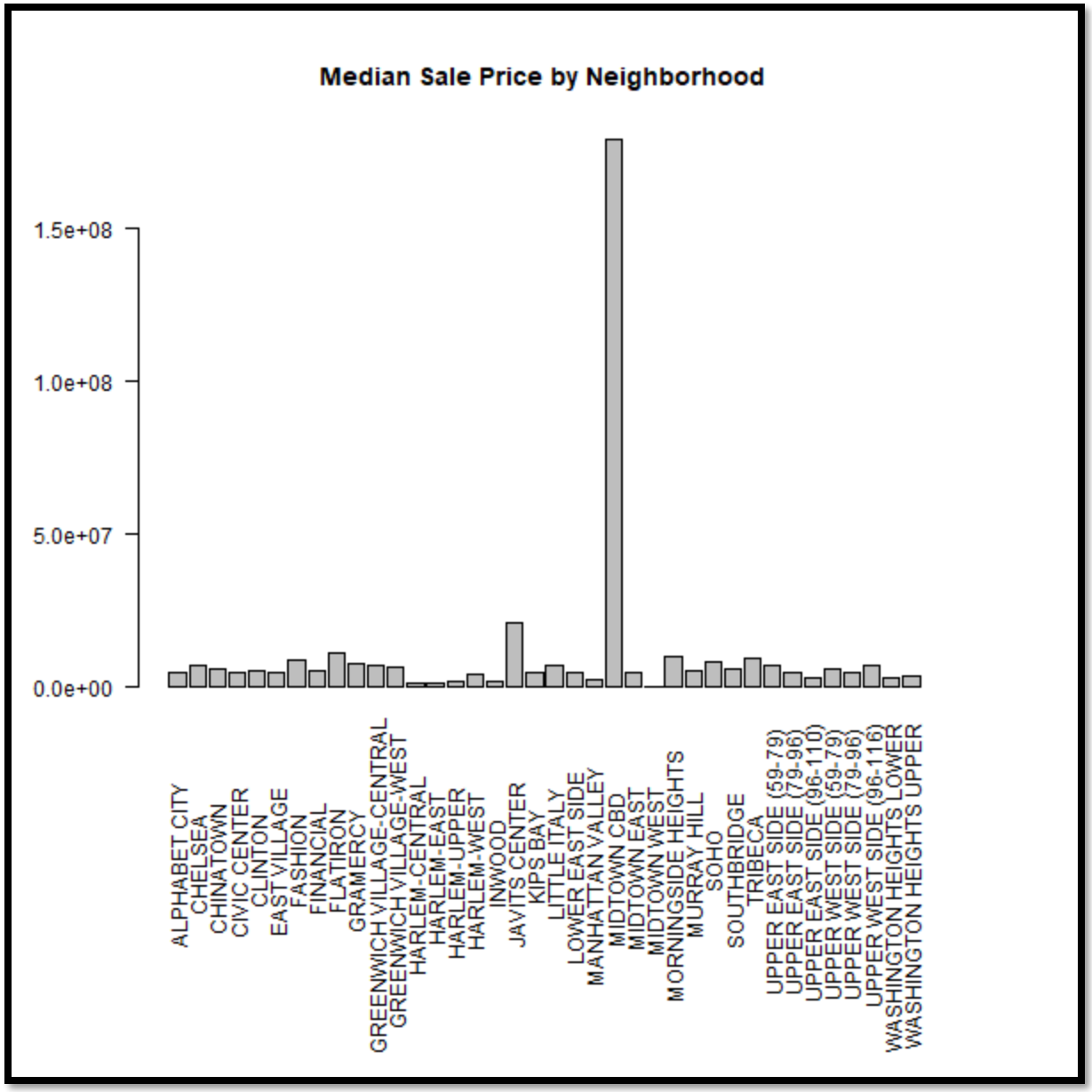
There is a file included with the zip that this document came in called main.R. It contains the R code that was used to load in and clean the data. Data was considered invalid if the gross square footage was equal to zero or the sale price was less than 100.

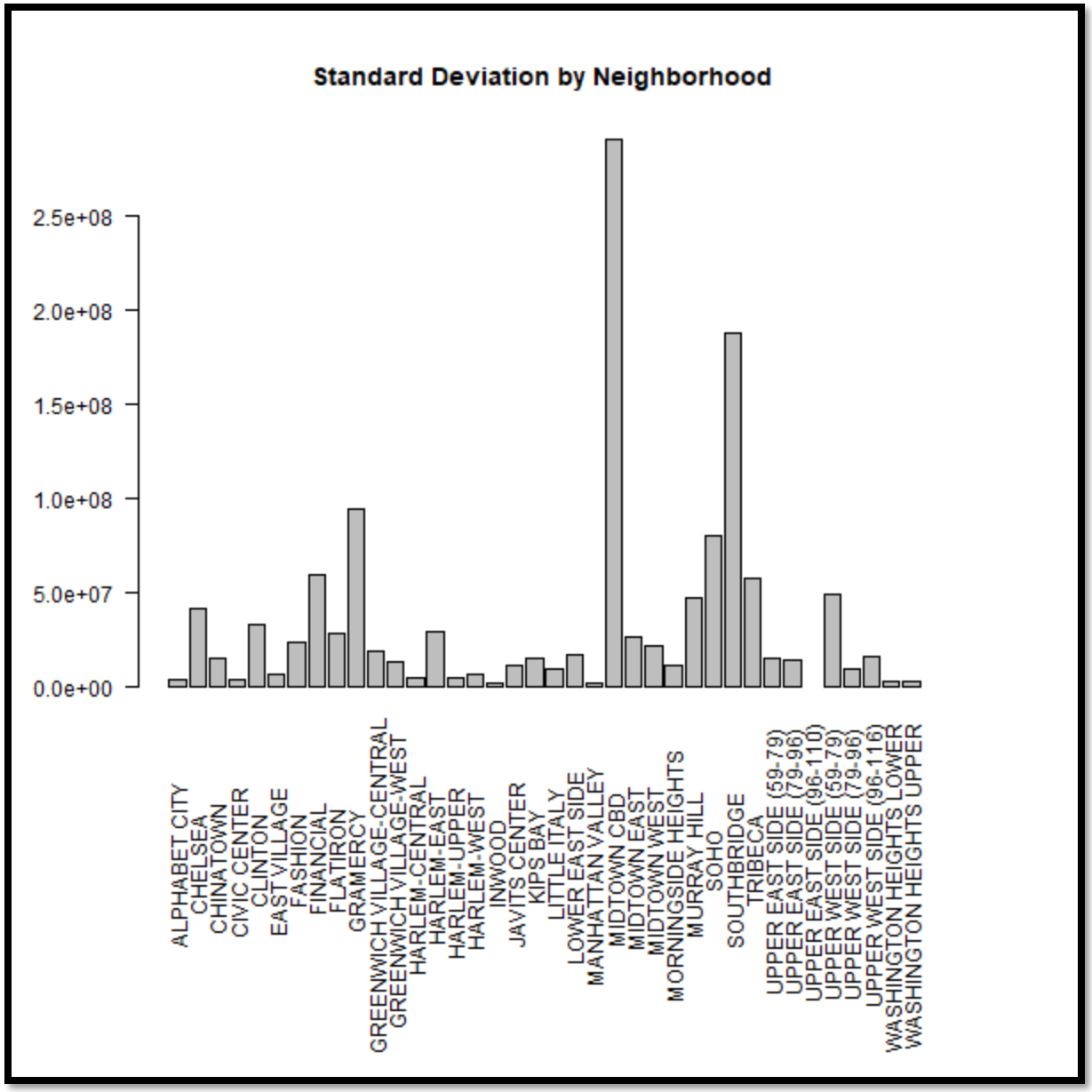
2[a-b]




Log(Sale Price) vs Sale Date
Log(Sale Price)
5
10
15
20
2013
Date


Average Sale Price by Neighborhood
2.5e+08
2.0e+08
1.5e+08
1.0e+08
5.0e+07
0.0e+00
ALPHABET CITY
CHELSEA
CHINATOWN
CIVIC CENTER
CLINTON
EAST VILLAGE
FASHION
FINANCIAL
FLATIRON
GRAMERCY
GREENWICH VILLAGE-CENTRAL
GREENWICH VILLAGE-WEST
HARLEM-CENTRAL
HARLEM-EAST
HARLEM-UPPER
HARLEM-WEST
INWOOD
JAVITS CENTER
KIPS BAY
LITTLE ITALY
LOWER EAST SIDE
MANHATTAN VALLEY
MIDTOWN CBD
MIDTOWN EAST
MIDTOWN WEST
MORNINGSIDE HEIGHTS
MURRAY HILL
SOHO
SOUTHBRIDGE
TRIBECA
UPPER EAST SIDE (59-79)
UPPER EAST SIDE (79-96)
UPPER EAST SIDE (96-110)
UPPER WEST SIDE (59-79)
UPPER WEST SIDE (79-96)
UPPER WEST SIDE (96-116)
WASHINGTON HEIGHTS LOWER
WASHINGTON HEIGHTS UPPER


Median Sale Price by Neighborhood
1.5e+08
1.0e+08
5.0e+07
0.0e+00
ALPHABET CITY
CHELSEA
CHINATOWN
CIVIC CENTER
CLINTON
EAST VILLAGE
FASHION
FINANCIAL
FLATIRON
GRAMERCY
GREENWICH VILLAGE-CENTRAL
GREENWICH VILLAGE-WEST
HARLEM-CENTRAL
HARLEM-EAST
HARLEM-UPPER
HARLEM-WEST
INWOOD
JAVITS CENTER
KIPS BAY
LITTLE ITALY
LOWER EAST SIDE
MANHATTAN VALLEY
MIDTOWN CBD
MIDTOWN EAST
MIDTOWN WEST
MORNINGSIDE HEIGHTS
MURRAY HILL
SOHO
SOUTHBRIDGE
TRIBECA
UPPER EAST SIDE (59-79)
UPPER EAST SIDE (79-96)
UPPER EAST SIDE (96-110)
UPPER WEST SIDE (59-79)
UPPER WEST SIDE (79-96)
UPPER WEST SIDE (96-116)
WASHINGTON HEIGHTS LOWER
WASHINGTON HEIGHTS UPPER


Standard Deviation by Neighborhood
2.5e+08
2.0e+08
1.5e+08
1.0e+08
5.0e+07
0.0e+00
ALPHABET CITY
CHELSEA
CHINATOWN
CIVIC CENTER
CLINTON
EAST VILLAGE
FASHION
FINANCIAL
FLATIRON
GRAMERCY
GREENWICH VILLAGE-CENTRAL
GREENWICH VILLAGE-WEST
HARLEM-CENTRAL
HARLEM-EAST
HARLEM-UPPER
HARLEM-WEST
INWOOD
JAVITS CENTER
KIPS BAY
LITTLE ITALY
LOWER EAST SIDE
MANHATTAN VALLEY
MIDTOWN CBD
MIDTOWN EAST
MIDTOWN WEST
MORNINGSIDE HEIGHTS
MURRAY HILL
SOHO
SOUTHBRIDGE
TRIBECA
UPPER EAST SIDE (59-79)
UPPER EAST SIDE (79-96)
UPPER EAST SIDE (96-110)
UPPER WEST SIDE (59-79)
UPPER WEST SIDE (79-96)
UPPER WEST SIDE (96-116)
WASHINGTON HEIGHTS LOWER
WASHINGTON HEIGHTS UPPER

3[a]

Range of sale prices for all data: [1.83e+02, 1.10e+09]

Range of sale prices for each neighborhood

	NEIGHBORHOOD	MINUMUM	MAXIMUM
1	ALPHABET CITY	283	15500000
2	CHELSEA	242000	284429879
3	CHINATOWN	1950000	53183682
4	CIVIC CENTER	1200000	13650000
5	CLINTON	600000	170000000
6	EAST VILLAGE	635000	41000000
7	FASHION	1300000	103000000
8	FINANCIAL	2000000	151000000
9	FLATIRON	2769455	111500000
10	GRAMERCY	318000	380000000
11	GREENWICH VILLAGE-CENTRAL	295162	101000000
12	GREENWICH VILLAGE-WEST	560000	96436875
13	HARLEM-CENTRAL	500	55000000
14	HARLEM-EAST	183	252000000
15	HARLEM-UPPER	110000	19600000
16	HARLEM-WEST	500000	24000000
17	INWOOD	440000	9498200
18	JAVITS CENTER	1150000	27500000
19	KIPS BAY	725000	39000000
20	LITTLE ITALY	325000	33500000
21	LOWER EAST SIDE	1400000	103466615
22	MANHATTAN VALLEY	50246	7380000
23	MIDTOWN CBD	6192000	1100000000
24	MIDTOWN EAST	490	115000000
25	MIDTOWN WEST	10000	470000000
26	MORNINGSIDE HEIGHTS	2365764	18386667
27	MURRAY HILL	133000	300000000
28	SOHO	550000	540829842
29	SOUTHBRIDGE	1400000	539000000
30	TRIBECA	3300000	223000000
31	UPPER EAST SIDE (59-79)	28000	141510000
32	UPPER EAST SIDE (79-96)	729500	95000000
33	UPPER EAST SIDE (96-110)	3350000	3350000
34	UPPER WEST SIDE (59-79)	15220	221000000
35	UPPER WEST SIDE (79-96)	360547	52500000
36	UPPER WEST SIDE (96-116)	512750	60000000
37	WASHINGTON HEIGHTS LOWER	265000	14450000
38	WASHINGTON HEIGHTS UPPER	9586	12000000

3[b]

Median of sale prices for all data: 875000

Median of sale prices for each neighborhood

	NEIGHBORHOOD	MEDIAN
1	ALPHABET CITY	4653771
2	CHELSEA	6990000
3	CHINATOWN	6140000
4	CIVIC CENTER	4931000
5	CLINTON	5243000
6	EAST VILLAGE	5116961
7	FASHION	9000000
8	FINANCIAL	5700000
9	FLATIRON	11250000
10	GRAMERCY	7523375
11	GREENWICH VILLAGE-CENTRAL	7250000
12	GREENWICH VILLAGE-WEST	6575000
13	HARLEM-CENTRAL	1300000
14	HARLEM-EAST	1250000
15	HARLEM-UPPER	1900000
16	HARLEM-WEST	4250000
17	INWOOD	2258707
18	JAVITS CENTER	21250000
19	KIPS BAY	4750000
20	LITTLE ITALY	6950000
21	LOWER EAST SIDE	4825000
22	MANHATTAN VALLEY	2770780
23	MIDTOWN CBD	179625000
24	MIDTOWN EAST	4750000
25	MIDTOWN WEST	57900
26	MORNINGSIDE HEIGHTS	10376216
27	MURRAY HILL	5750000
28	SOHO	8500000
29	SOUTHBRIDGE	6225625
30	TRIBECA	9500000
31	UPPER EAST SIDE (59-79)	7000000
32	UPPER EAST SIDE (79-96)	4900000
33	UPPER EAST SIDE (96-110)	3350000
34	UPPER WEST SIDE (59-79)	6110000
35	UPPER WEST SIDE (79-96)	5025000
36	UPPER WEST SIDE (96-116)	6950000
37	WASHINGTON HEIGHTS LOWER	3300000
38	WASHINGTON HEIGHTS UPPER	3860126

3[c]

Mean of sale prices for all data: 8662951

Mean of sale prices for each neighborhood

	NEIGHBORHOOD	MEAN
1	ALPHABET CITY	5897204
2	CHELSEA	21540533
3	CHINATOWN	12985063
4	CIVIC CENTER	6030542
5	CLINTON	14856886
6	EAST VILLAGE	7159309
7	FASHION	18571245
8	FINANCIAL	39986667
9	FLATIRON	27591907
10	GRAMERCY	44527171
11	GREENWICH VILLAGE-CENTRAL	12026421
12	GREENWICH VILLAGE-WEST	10269727
13	HARLEM-CENTRAL	2563649
14	HARLEM-EAST	5503330
15	HARLEM-UPPER	4866437
16	HARLEM-WEST	6196898
17	INWOOD	3094446
18	JAVITS CENTER	15492857
19	KIPS BAY	11606250
20	LITTLE ITALY	10764088
21	LOWER EAST SIDE	10617596
22	MANHATTAN VALLEY	3054686
23	MIDTOWN CBD	252020355
24	MIDTOWN EAST	14787857
25	MIDTOWN WEST	2179821
26	MORNINGSIDE HEIGHTS	10376216
27	MURRAY HILL	17648795
28	SOHO	32515954
29	SOUTHBRIDGE	74206406
30	TRIBECA	28210357
31	UPPER EAST SIDE (59-79)	11171130
32	UPPER EAST SIDE (79-96)	9429458
33	UPPER EAST SIDE (96-110)	3350000
34	UPPER WEST SIDE (59-79)	23711921
35	UPPER WEST SIDE (79-96)	8248005
36	UPPER WEST SIDE (96-116)	12112997
37	WASHINGTON HEIGHTS LOWER	3937385
38	WASHINGTON HEIGHTS UPPER	4609474

3[d]

Standard deviation of sale prices for all data: 40249762

Standard deviation of sale prices for each neighborhood

	NEIGHBORHOOD	STANDARD DEVIATION
1	ALPHABET CITY	3877376
2	CHELSEA	41912252
3	CHINATOWN	15668884
4	CIVIC CENTER	4214875
5	CLINTON	33333589
6	EAST VILLAGE	6751300
7	FASHION	23765787
8	FINANCIAL	59350242
9	FLATIRON	29038627
10	GRAMERCY	94421252
11	GREENWICH VILLAGE-CENTRAL	19471467
12	GREENWICH VILLAGE-WEST	13923195
13	HARLEM-CENTRAL	4816716
14	HARLEM-EAST	29536932
15	HARLEM-UPPER	5411304
16	HARLEM-WEST	6505611
17	INWOOD	2429118
18	JAVITS CENTER	11288247
19	KIPS BAY	15011429
20	LITTLE ITALY	9992874
21	LOWER EAST SIDE	17168704
22	MANHATTAN VALLEY	1828899
23	MIDTOWN CBD	290619315
24	MIDTOWN EAST	27177034
25	MIDTOWN WEST	21850195
26	MORNINGSIDE HEIGHTS	11328489
27	MURRAY HILL	46983884
28	SOHO	80093116
29	SOUTHBRIDGE	187963326
30	TRIBECA	57698249
31	UPPER EAST SIDE (59-79)	15103714
32	UPPER EAST SIDE (79-96)	14074042
33	UPPER EAST SIDE (96-110)	NA
34	UPPER WEST SIDE (59-79)	49307740
35	UPPER WEST SIDE (79-96)	10076774
36	UPPER WEST SIDE (96-116)	16543803
37	WASHINGTON HEIGHTS LOWER	3170301
38	WASHINGTON HEIGHTS UPPER	3266536

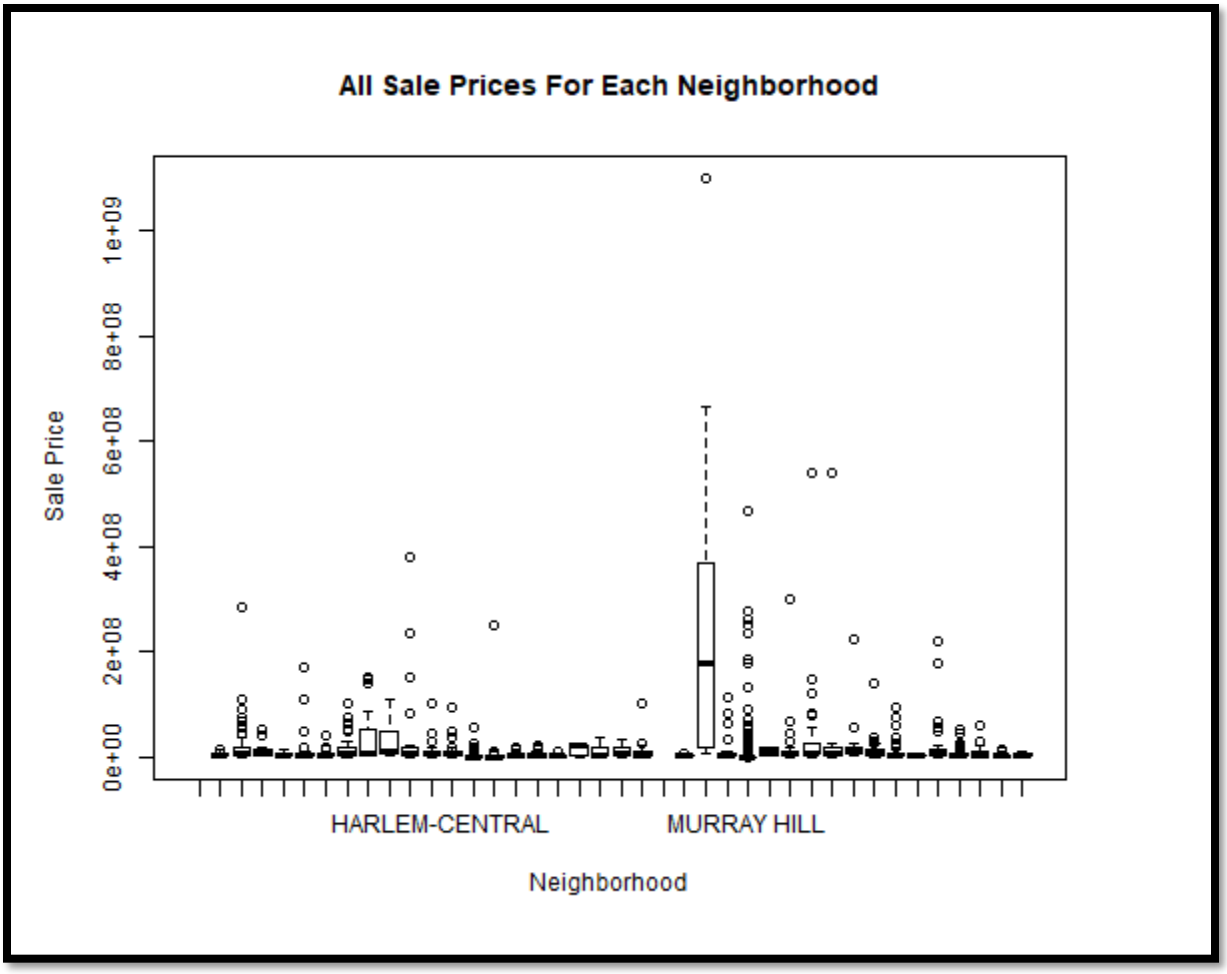
Conclusions

The first data to analyze is the graph of sale prices versus time. Though there are spikes here and there, this graph lacks a recognizable pattern. To see if there was any trend over the short timespan covered by the data, I plotted the sale prices on a logarithmic scale. This is the second graph in part 2. Even with this change in scale, no pattern is clear.

On the other hand, the bar plots of the sale prices for each neighborhood are easy to draw conclusions from. As shown by the graph of average sale prices, Midtown CBD and Southbridge have the highest prices of any residential area. Though this might be useful if one is looking for housing in the area, it isn't very interesting.

When combined with the graph of standard deviations for each neighborhood, some intriguing patterns show up. In general, it appears that neighborhoods with higher average prices also have a higher standard deviation in prices. When viewed side by side, it is easy to see this relationship as the graphs look very similar. This means that a neighborhood with a higher average price will likely contain a large range of prices. Take Southbridge versus Midtown West as an example. The minimum sale price for a property in Southbridge is lower than highest sale price in Midtown West. This would not be obvious if we only considered the graph of average sale prices.

Another interesting property shows up when comparing the graph of medians to averages. The medians are lower than the averages for almost all neighborhoods. This indicates that cheaper housing is more common than expensive housing. To confirm this suspicion, I created one last graph to show all prices for each neighborhood. This graph is on the next page.



Additionally

Apologies for the poor formatting and the lack of more luxurious graphs. This was my first time using R. I used R for cleaning the data, creating graphs and calculating statistical data. I could have finished this project more quickly but using R for the first time slowed me down. I do not believe that R was very effective for this task. It would have been faster and more hassle free to use Python for calculating statistical data and Excel for producing graphs. However, because an R example was provided, I took the opportunity to pick up the language and complete the project using it.