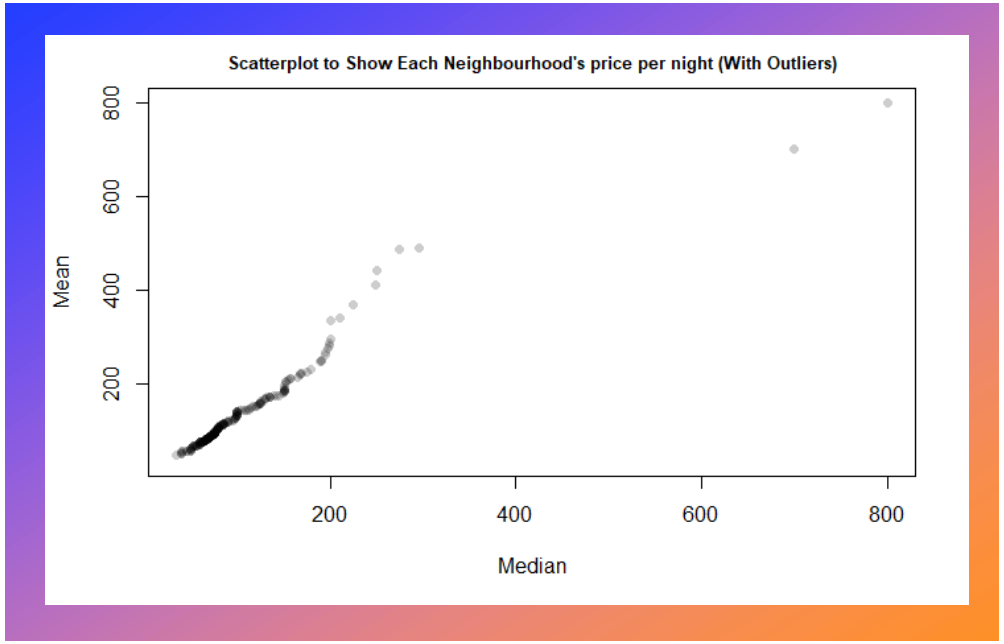
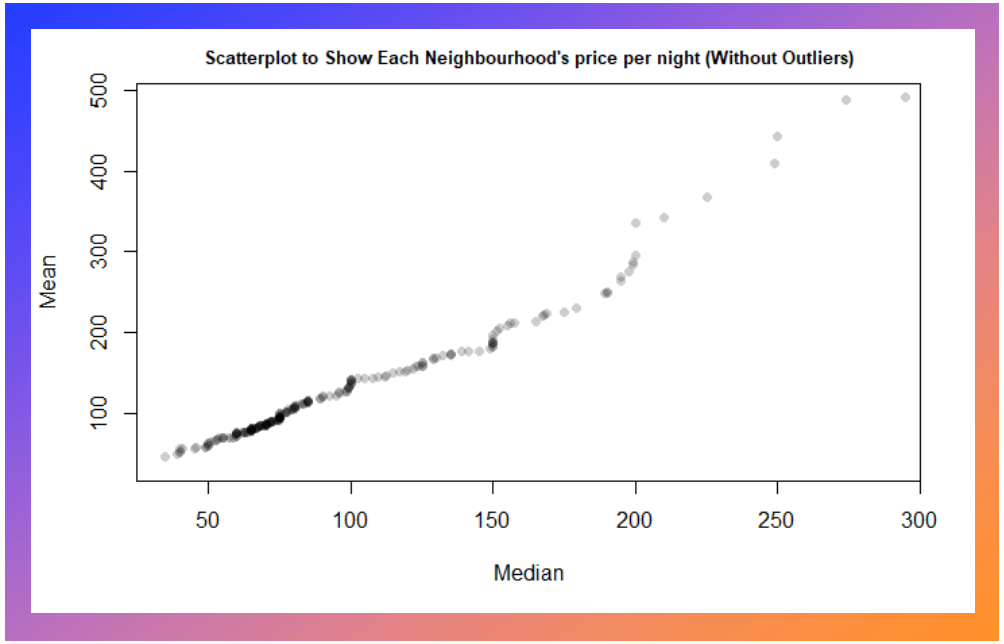
DATA ANALYSIS

Project 2

Hypothesis – The most expensive neighborhoods are in the most expensive region

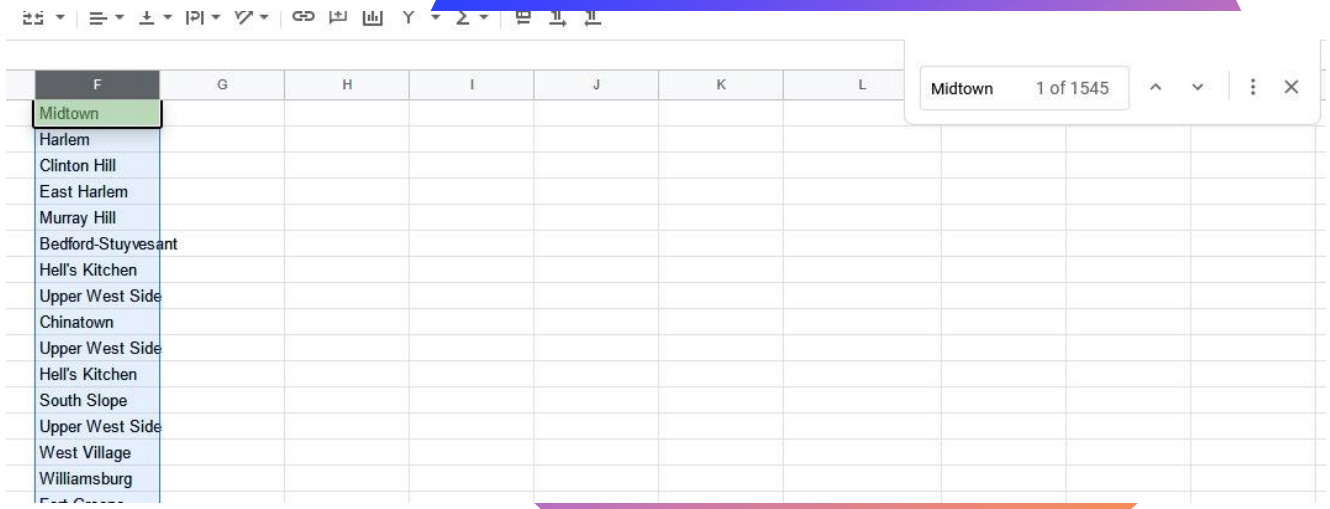
My aim for my part of the project is to delve deeper into my previous project which was about which region in New York was the most and least expensive on average.





Scatter Plot

- I started by generating a scatter plot to visualize the mean and median price of each Neiborhood's price. I quickly noticed it was hard to read due to some outliers far to the top right, I made a second scatter plot excluding those outliers to make the scatter plot more readable.
- From the scatter plot it is easy to see how the average price can vary.
- By making the dots on the scatter plot partly transparent you can see where overlap of the dots happens easier which exposed a lot of dots around the 50-100 area. Previously it was hard to see how many dots were in a busy area of the scatter plot



Comparing Top 10

neighbourhood <chr>	Mean <dbl>	Median <dbl>	Total <dbl>
Fort Wadsworth	800	800	1
Woodrow	700	700	1
Tribeca	491	295	177
Neponsit	275	274	3
NoHo	296	250	78
Willowbrook	249	249	1
Flatiron District	342	225	80
Midtown	283	210	1545
Financial District	225	200	744
West Village	268	200	768

1-10 of 221 rows Previous 1 2

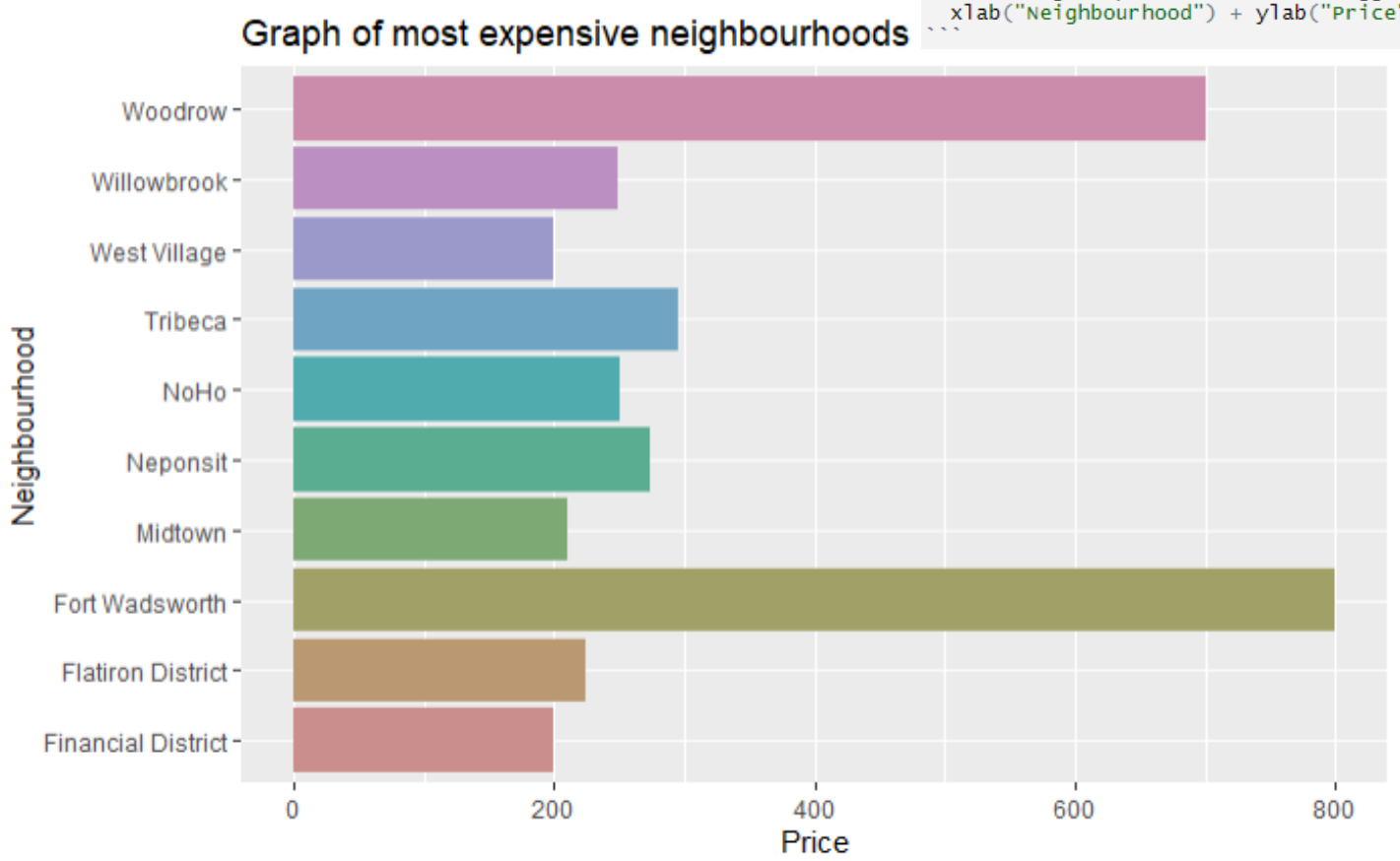
Firstly, I got the mean and medians of each Neiborhood. This allowed me to see what the average prices were much easier. However, the list was sorted alphabetically, not by price. This was a tricky fix but eventually got the list to sort by most expensive to least.

Now we can quite clearly see that Fort Wadsworth is the most expensive. Then I added the total of how many properties where in that neighborhood. This unveiled that Fort Wadsworth and some other of the Neiborhood's only had a few properties in that area. With this information I can now see that those could be outliers. To check that my calculations for the total were correct I went into the CVS and ran the find function to see how many of each Neiborhood were in the CVS file. This was good proof that my calculations were correct.

Table Visualized – Comparing Top 10

```r
# Load ggplot2
library(ggplot2)


# Barplot
ggplot(headofarranged, aes(x=as.factor(neighbourhood), y=Median, fill=as.factor(neighbourhood) )) +
  geom_bar(stat = "identity") +
  coord_flip() +
  scale_fill_hue(c = 40) +
  theme(legend.position="none") + ggtitle("Graph of most expensive neighbourhoods" ) +
  xlab("Neighbourhood") + ylab("Price")
```

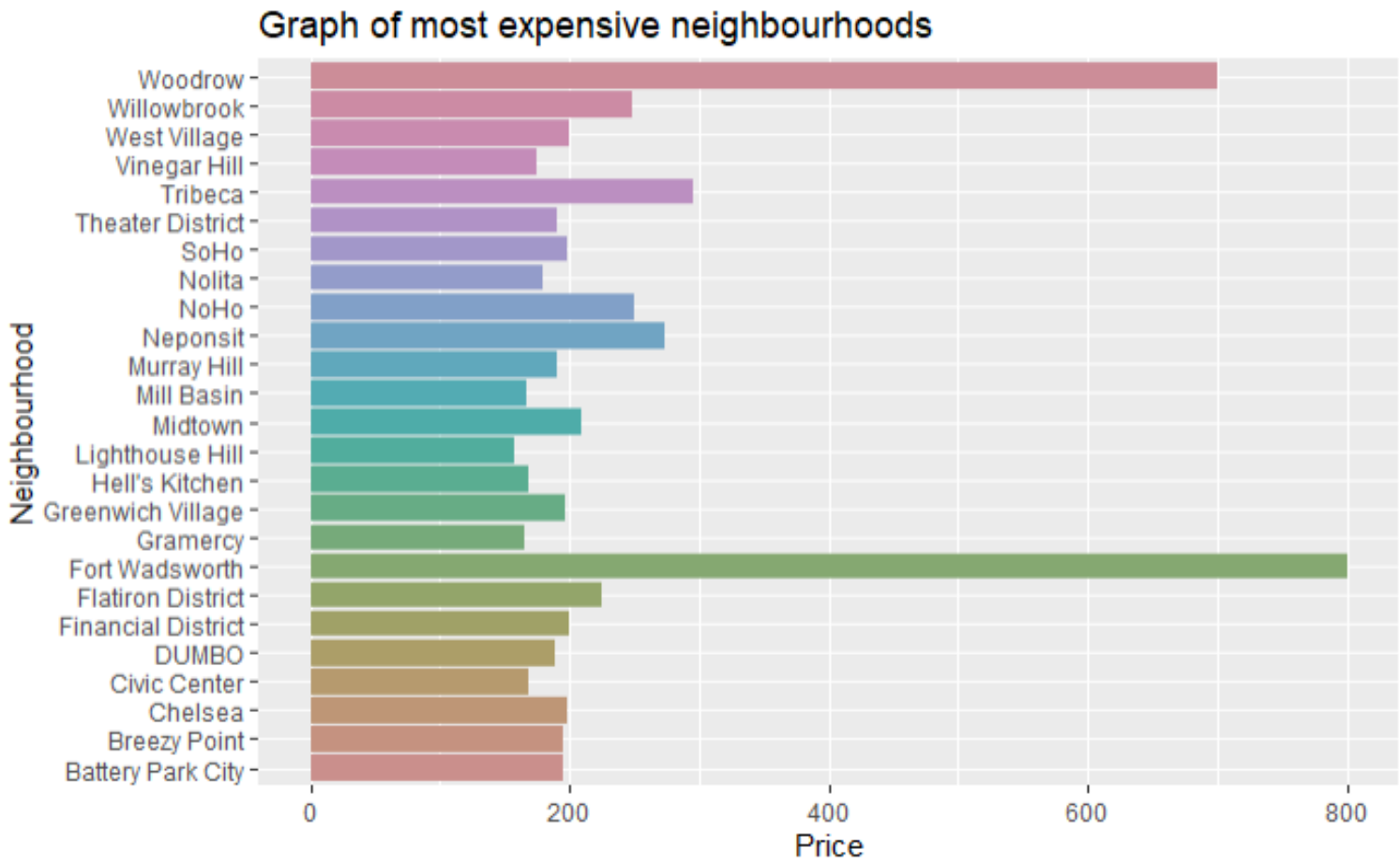


Woodrow - Staten Island
Willowbrook - Staten Island
West Village - Manhattan
Tribeca - Manhattan
NoHo - Manhattan
Neponsit - Queens
Midtown - Manhattan
Fort Wadsworth - Staten Island
Flatiron District - Manhattan
Financial District - Manhattan

Manhattan - 60%
Staten Island - 30%
Queens – 10%

Comparing Top 25

I decided to try comparing the top 25 to see if the trend would continue



64% 16% 8% 12%

Woodrow - Staten Island
Willowbrook - Staten Island
West Village - Manhattan
Tribeca - Manhattan
NoHo - Manhattan
Neponsit - Queens
Midtown - Manhattan
Fort Wadsworth - Staten Island
Flatiron District - Manhattan
Financial District - Manhattan
Chelsea – Manhattan
SoHo – Manhattan
Greenwich Village – Manhattan

Battery Park – Manhattan
Breezy Point – Queens
Murray Hills – Manhattan
Theater District – Manhattan
DUMBO – Brooklyn
Nolita – Manhattan
Vinegar Hills – Brooklyn
Civic Center – Manhattan
Hell's Kitchen – Manhattan
Mill Basin – Brooklyn
Gramercy – Manhattan
Lighthouse Hill – Staten Island

Comparing top 25

When comparing top 25 the percentages hold up to be similar to the top 10. The only real noticeable difference is Staten Island drops in percentage a bit and is replaced with Brooklyn

Conclusion - Was My Hypothesis Correct?

I conclude:

- Manhattan is the most expensive region of New York.
- It has 60% of the top 10 most expensive properties.
- It has 64% of the top 25 most expensive properties.
- If you are going to stay in New York beware that if you want to stay in Manhattan, it will likely cost more than any other region. This has its upsides though as it is in the middle of the city.
- Brooklyn could be a good middle ground as it is reasonably priced and is a nice area of New York.

neighbourhood_group <chr>	Mean <dbl>	Median <dbl>
Bronx	87	65
Brooklyn	124	90
Manhattan	197	150
Queens	100	75
Staten Island	115	75

5 rows