MA678 Project: Analyzing AirBnB Data in Boston

Shixin Liang

11/29/2019

Introduction

Background

Airbnb offers travellers someone's home as a place to stay instead of a hotel. And people can rent out extra space in their own home through Airbnb and make money for allowing a guest to stay the night. For travellers, they choose Airbnb because of many reasons: to shorten traveling time, to become part of local life or to experience individual room style. But for Airbnb hosts, their primary goal must be to earn money. Therefore, it is important to find a way to set a reasonable price for Airbnb properties and spare rooms.

Goal

This project aims to fit a suitable model which can provide a suggested price for an Airbnb property or spare room according to its attributes by using Airbnb Boston data. Then the model can be used as a basic pricing tool for Airbnb hosts.

Dataset

The Airbnb Boston dataset can be downloaded from the website: Airbnb Data Collection. (http://tomslee.net/airbnb-data-collection-get-the-data) The data is from July, 2016 to July, 2017. There are 12 csv files and I combined them into one data frame. The data contains 13 variables:

- room_id: A unique number identifying an Airbnb listing.
- host_id: A unique number identifying an Airbnb host.
- room_type: One of "Entire home/apt", "Private room", or "Shared room".
- neighborhood: A subregion of the city or search area for which the survey is carried out.
- reviews: The number of reviews that a listing has received. Airbnb has said that 70% of visits end up with a review, so the number of reviews can be used to estimate the number of visits.
- overall_satisfaction: The average rating (out of five) that the listing has received from those visitors who left a review.
- accommodates: The number of guests a listing can accommodate.
- bedrooms: The number of bedrooms a listing offers.
- price: The price (in $US) for a night stay.
- minstay: The minimum stay for a visit, as posted by the host.
- latitude and longitude: The latitude and longitude of the listing as posted on the Airbnb site
- last_modified: the date and time that the values were read from the Airbnb web site.

By looking through the data, we can see that there is no bathrooms, country, borough data for all tables. And for data from January, 2017 to July, 2017, there is no minstay data. Therefore, I ignored this four variables.

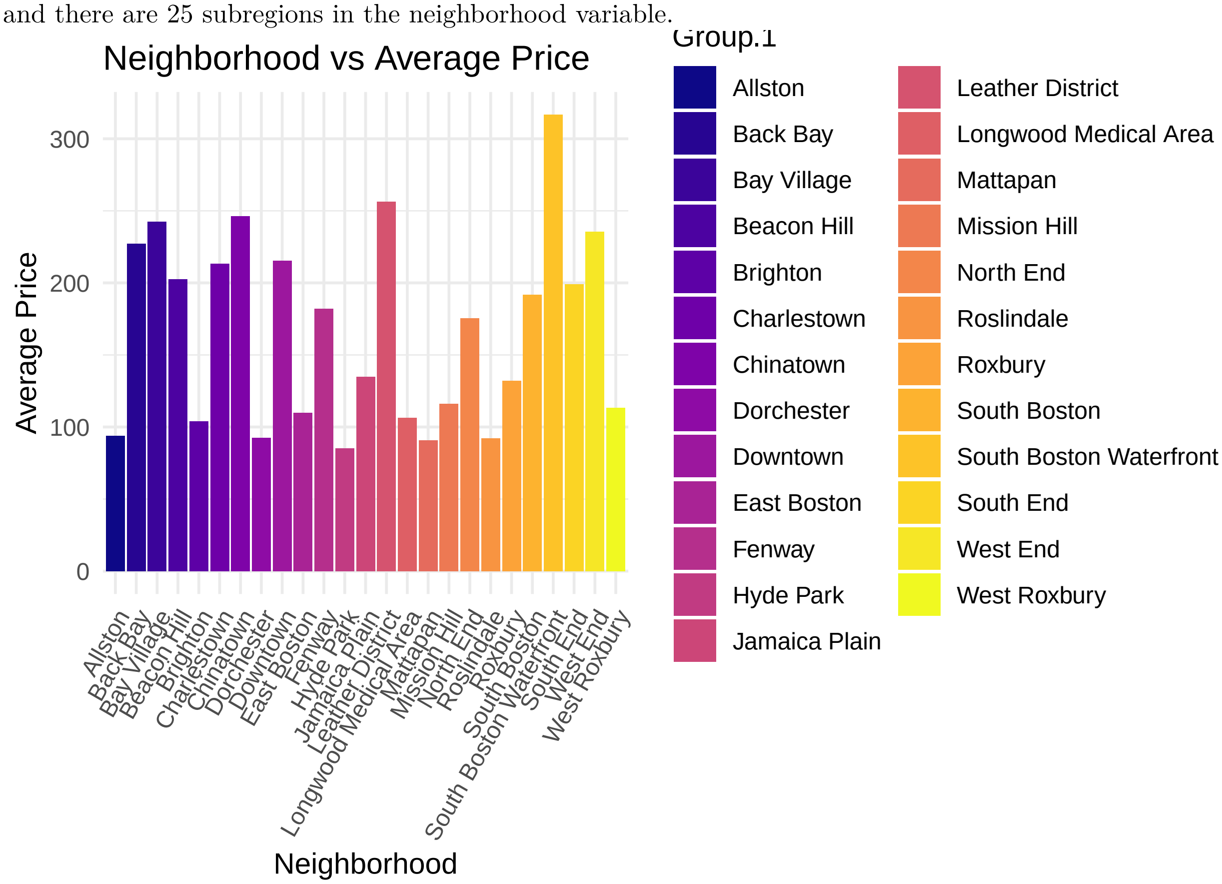
Exploratory Data Analysis

Prior to any analysis, the data was cleaned in the general following manner: 1) deleted redundant variables in the three latest tables 2) reorder columns in the tables to make all 12 tables consist with each other tables 3) merged 12 tables into 1 dataset 4) deleted NA data and outliers, and 5) converted string data in last_modified into datetime.

The following graphs are those that may potentially reveal patterns in data. Numerous other graphs were explored and have been included in the appendix.

Neighborhood

To begin, a bar graph of the average price as determined by the neighborhood. Overall, it shows that south boston waterfront is the neighborhood with highest average price. And the average price varies a lot from different neighborhood in Boston. It means that the data can be considered to be grouped in neighborhood, and there are 25 subregions in the neighborhood variable.

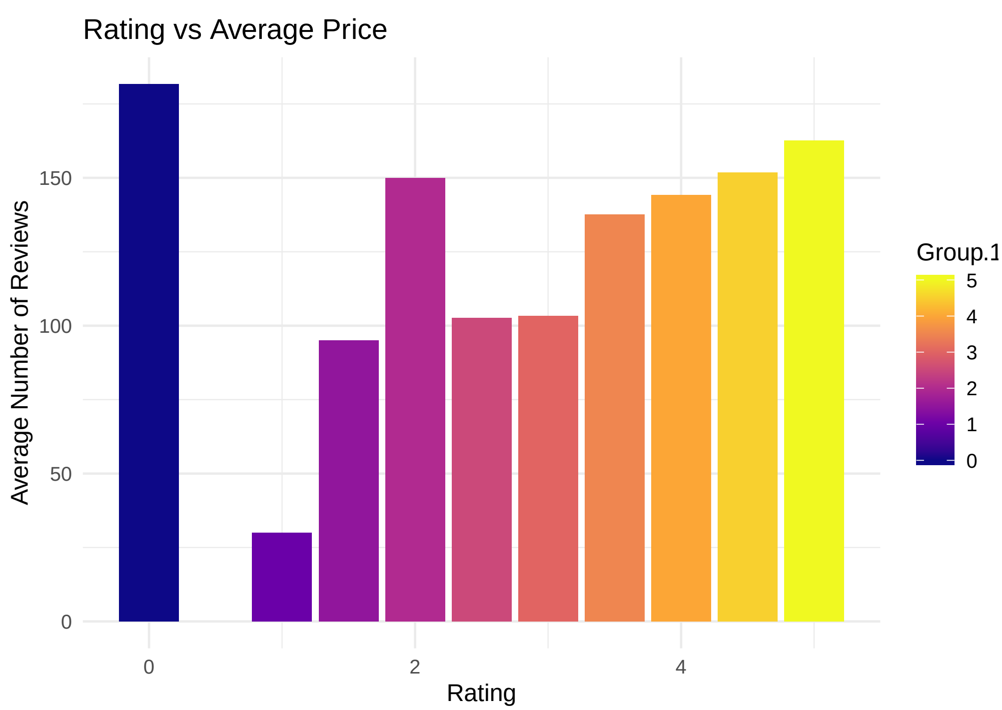


Accommodates vs Number of Bedrooms

This graph displays the number of guests a listing can accommodate arranged by the number of bedrooms a listing offers. This pair plot shows a positive relationships between variables accommodates and bedrooms. Overall, it shows that when the number of Bedrooms is larger, usually the accommodates is larger.

Overall Satisfaction vs Average Price

The following bar graph displays the average price as determined by the average rating of the listing. Overall, it shows that listings with rating 0 have the highest average price. It makes sense because no one will book a room with a too expensive price. Listings with rating 2 have relatively high average price. And for listings with rating higher than 2.5, there is a positive relationship between average price and average rating, which means higher rating higher price.



Model Building

After EDA, I chose these following variables to do regression. The dependent variable is price. The independent variables are room type, neighborhood, reviews, overall satisfaction, accommodates and bedrooms. For the variables accommodates and room type, an interaction is considered. I tried both linear regression models and multilevel linear regression models.

Linear Regression Model

The first model I used is a linear regression model:

```
##
## Call:
## lm(formula = price ~ room_type + neighborhood + reviews + overall_satisfaction +
##     accommodates * bedrooms, data = boston_airbnb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -523.75  -41.97   -4.42   27.57 1266.68
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          78.46235    2.71234  28.928  < 2e-16
## room_typePrivate room               -54.89179    1.27463 -43.065  < 2e-16
## room_typeShared room                -80.88714    3.63628 -22.244  < 2e-16
## neighborhoodBack Bay                101.18154    2.39462  42.254  < 2e-16
## neighborhoodBay Village             104.90576    6.16045  17.029  < 2e-16
## neighborhoodBeacon Hill              82.73431    2.64648  31.262  < 2e-16
## neighborhoodBrighton                  0.03862    2.59557   0.015  0.98813
## neighborhoodCharlestown              70.73468    3.21629  21.993  < 2e-16
## neighborhoodChinatown                90.80662    3.55883  25.516  < 2e-16
## neighborhoodDorchester               -5.51809    2.37143  -2.327  0.01998
## neighborhoodDowntown                 83.80115    2.67046  31.381  < 2e-16
## neighborhoodEast Boston               3.90846    2.72566   1.434  0.15159
## neighborhoodFenway                   63.58590    2.49399  25.496  < 2e-16
## neighborhoodHyde Park               -17.41747    5.38494  -3.234  0.00122
## neighborhoodJamaica Plain            16.34711    2.32206   7.040 1.96e-12
## neighborhoodLeather District        118.99326   11.60740  10.252  < 2e-16
## neighborhoodLongwood Medical Area    31.98765   11.92889   2.682  0.00733
## neighborhoodMattapan                  1.53270    5.63004   0.272  0.78544
## neighborhoodMission Hill             17.93331    3.26917   5.486 4.15e-08
## neighborhoodNorth End                36.76678    2.88282  12.754  < 2e-16
## neighborhoodRoslindale              -17.20241    3.76270  -4.572 4.85e-06
## neighborhoodRoxbury                  14.94185    2.78549   5.364 8.18e-08
## neighborhoodSouth Boston             52.27262    2.69139  19.422  < 2e-16
## neighborhoodSouth Boston Waterfront 164.40582    3.70529  44.371  < 2e-16
## neighborhoodSouth End                82.95361    2.41812  34.305  < 2e-16
## neighborhoodWest End                 79.24446    4.72958  16.755  < 2e-16
## neighborhoodWest Roxbury             -4.99432    4.35138  -1.148  0.25108
## reviews                              -0.08040    0.01231  -6.532 6.59e-11
## overall_satisfaction                 -4.00169    0.23173 -17.269  < 2e-16
## accommodates                          4.95933    0.57867   8.570  < 2e-16
## bedrooms                             39.91019    1.21148  32.943  < 2e-16
## accommodates:bedrooms                 1.84467    0.18065  10.211  < 2e-16
##
```

```
## (Intercept)                        ***
## room_typePrivate room              ***
## room_typeShared room               ***
## neighborhoodBack Bay               ***
## neighborhoodBay Village            ***
## neighborhoodBeacon Hill            ***
## neighborhoodBrighton
## neighborhoodCharlestown            ***
## neighborhoodChinatown              ***
## neighborhoodDorchester             *
## neighborhoodDowntown               ***
## neighborhoodEast Boston
## neighborhoodFenway                 ***
## neighborhoodHyde Park              **
## neighborhoodJamaica Plain          ***
## neighborhoodLeather District       ***
## neighborhoodLongwood Medical Area   **
## neighborhoodMattapan
## neighborhoodMission Hill           ***
## neighborhoodNorth End              ***
## neighborhoodRoslindale             ***
## neighborhoodRoxbury                ***
## neighborhoodSouth Boston           ***
## neighborhoodSouth Boston Waterfront ***
## neighborhoodSouth End              ***
## neighborhoodWest End               ***
## neighborhoodWest Roxbury
## reviews                            ***
## overall_satisfaction               ***
## accommodates                       ***
## bedrooms                           ***
## accommodates:bedrooms              ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.55 on 38694 degrees of freedom
## Multiple R-squared:  0.4876, Adjusted R-squared:  0.4871
## F-statistic:  1188 on 31 and 38694 DF,  p-value: < 2.2e-16
```

The p-value in the result shows that most of the coeffcients are statistically significant except some levels of the neighborhood variable. Therefore, I also checked a model without the neighborhood variable.

The second model I used is a linear regression model without the neighborhood variable:

```
##
## Call:
## lm(formula = price ~ room_type + reviews + overall_satisfaction +
##     accommodates * bedrooms, data = boston_airbnb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -500.91  -47.07   -9.93   30.14 1321.58
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)            152.19263    2.19536  69.325   <2e-16 ***
## room_typePrivate room  -96.91929    1.23004 -78.794   <2e-16 ***
## room_typeShared room  -109.57187    3.90779 -28.039   <2e-16 ***
## reviews                 -0.14037    0.01313 -10.687   <2e-16 ***
## overall_satisfaction    -4.08940    0.24736 -16.532   <2e-16 ***
## accommodates             5.23598    0.62115   8.430   <2e-16 ***
## bedrooms                31.44194    1.29373  24.303   <2e-16 ***
## accommodates:bedrooms    1.66723    0.19446   8.574   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 97.1 on 38718 degrees of freedom
## Multiple R-squared:  0.3971, Adjusted R-squared:  0.397
## F-statistic:  3644 on 7 and 38718 DF,  p-value: < 2.2e-16
```

The p-value looks better and all the coefficients are statistically significant. However, its R-squared turns smaller. Private room has 96.91929 lower weighted price than the Entire home/apt. Shared room has 109.57187 lower weighted price than the price of Entire home/apt. When one review increases, keeping other variables the same, price will decrease by 0.14037. When one unit increases in overall satisfaction, keeping other variables the same, price will decrease by 4.08940. When accommodates increases by one, keeping other variables the same, price will increase by 6.90321 ($5.23598 + 1.66723 = 6.90321$). When the number of bedrooms increases by one, keeping other variables the same, price will increase by 33.10917 ($31.44194 + 1.66723 = 33.10917$).

Multilevel Linear Regression Model

Random Intercept

I fitted a multilevel model to allow varying intercepts for neighborhood. The model is:

```
## lmer(formula = price ~ room_type + reviews + overall_satisfaction +
##     accommodates * bedrooms + (1 | neighborhood), data = boston_airbnb)
##                       coef.est coef.se
## (Intercept)           125.39    10.04
## room_typePrivate room -55.02     1.27
## room_typeShared room  -80.98     3.64
## reviews                -0.08     0.01
## overall_satisfaction   -4.00     0.23
## accommodates            4.96     0.58
## bedrooms               39.91     1.21
## accommodates:bedrooms   1.84     0.18
##
## Error terms:
##  Groups       Name        Std.Dev.
##  neighborhood (Intercept) 48.97
##  Residual                 89.55
## ---
## number of obs: 38726, groups: neighborhood, 25
## AIC = 458185, DIC = 458172.7
## deviance = 458168.8
```

All of the absolute t value of the coefficients are larger than 1.96, which means the coefficients are statistically significant at 95% level. Private room has 55.02 lower weighted price than the Entire home/apt. Shared room has 80.98 lower weighted price than the price of Entire home/apt. When one review increases, keeping other variables the same, price will decrease by 0.08. When one unit increases in overall satisfaction, keeping other variables the same, price will decrease by 4.00. When accommodates increases by one, keeping other variables the same, price will increase by 6.8 ($4.96 + 1.84 = 6.8$). When the number of bedrooms increases by one, keeping other variables the same, price will increase by 41.75 ($39.91 + 1.84 = 41.75$). The variance among different neighborhoods is 48.97. The within-neighborhood variance is 89.55. The variance among different neighborhoods is lower than the within-neighborhood variance. It means the pooling effect is strong.

It is strange that the coefficient of overall satisfaction is negative, which means higher rating lower price. It is against common sense. Therefore, I would like to expand this model to allow varying slopes for overall satisfaction.

Random Intercept And Random Slope

I fitted a multilevel model to allow varying intercepts for neighborhood and varying slopes for overall satisfaction. The model is:

```
## lmer(formula = price ~ room_type + reviews + accommodates * bedrooms +
##     (1 + overall_satisfaction | neighborhood), data = boston_airbnb)
##                        coef.est coef.se
## (Intercept)           100.60     8.96
## room_typePrivate room -54.59     1.27
## room_typeShared room  -79.66     3.61
## reviews                -0.10     0.01
## accommodates            4.77     0.57
## bedrooms               40.71     1.20
## accommodates:bedrooms   1.82     0.18
##
## Error terms:
##  Groups       Name                 Std.Dev. Corr
##  neighborhood (Intercept)          70.85
##               overall_satisfaction  9.61    -0.79
##  Residual                          88.72
## ---
## number of obs: 38726, groups: neighborhood, 25
## AIC = 457554, DIC = 457541
## deviance = 457536.2
```

All of the coefficients are statistically significant. The coefficients are similar to the previous model. Private room has 54.59 lower weighted price than the Entire home/apt. Shared room has 79.66 lower weighted price than the price of Entire home/apt. When one review increases, keeping other variables the same, price will decrease by 0.10. When accommodates increases by one, keeping other variables the same, price will increase by 6.59 ($4.77 + 1.82 = 6.59$). When the number of bedrooms increases by one, keeping other variables the same, price will increase by 42.53 ($40.71 + 1.82 = 42.53$). The variance among different neighborhoods are 70.85. The within-neighborhood variance is 88.72. The slope of overall satisfaction is 9.61, and there is correlation with intercept of -0.79.

Now, the slope of overall satisfaction is positive. It is much more reasonable than the last model.

Model Selection

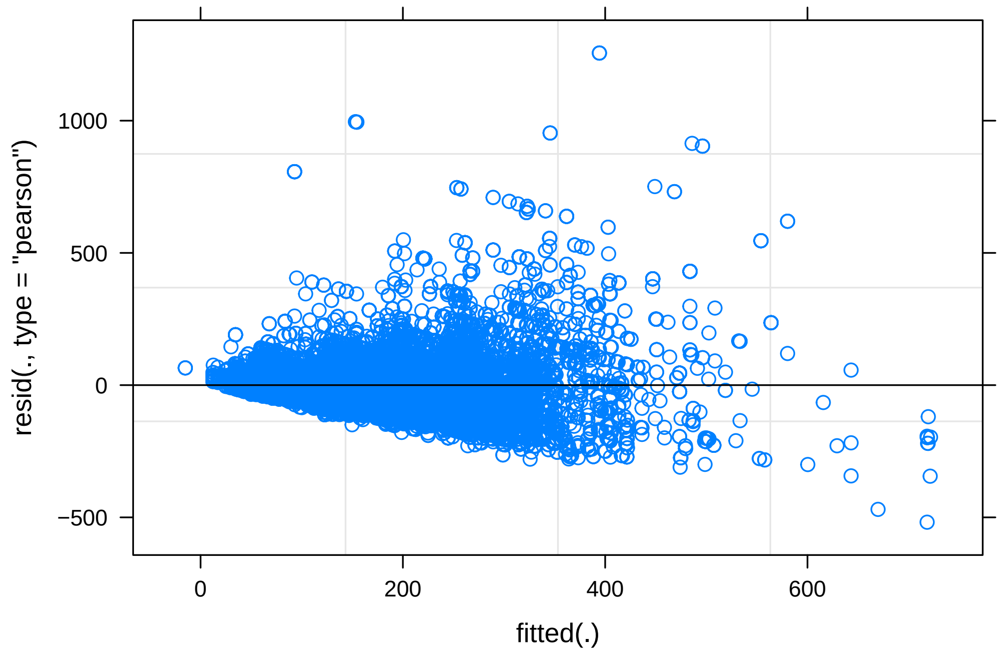
I used AIC and BIC to select the best models. Lower AIC and lower BIC means better performance.

Model	AIC	BIC
lm_fit1	458068.3	458351.0
lm_fit2	464312.6	464389.7
lmer_fit1	458185.0	458270.6
lmer_fit2	457553.5	457647.7

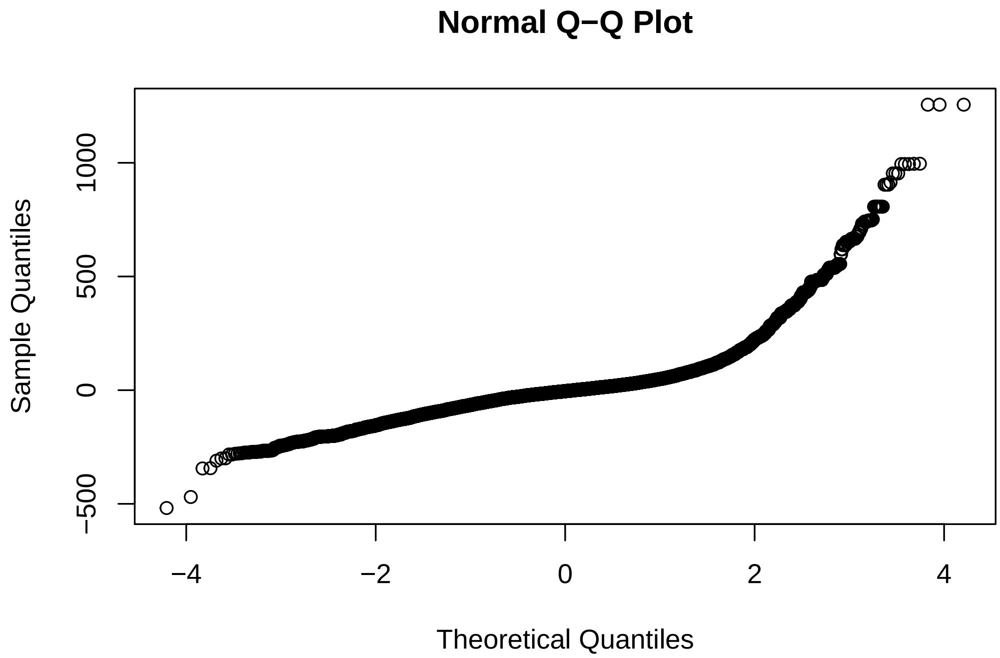
The model with random intercept and random slope (lmer_fit2) is a better model comparing to the other three models because it has the smallest AIC and BIC.

Model Checking

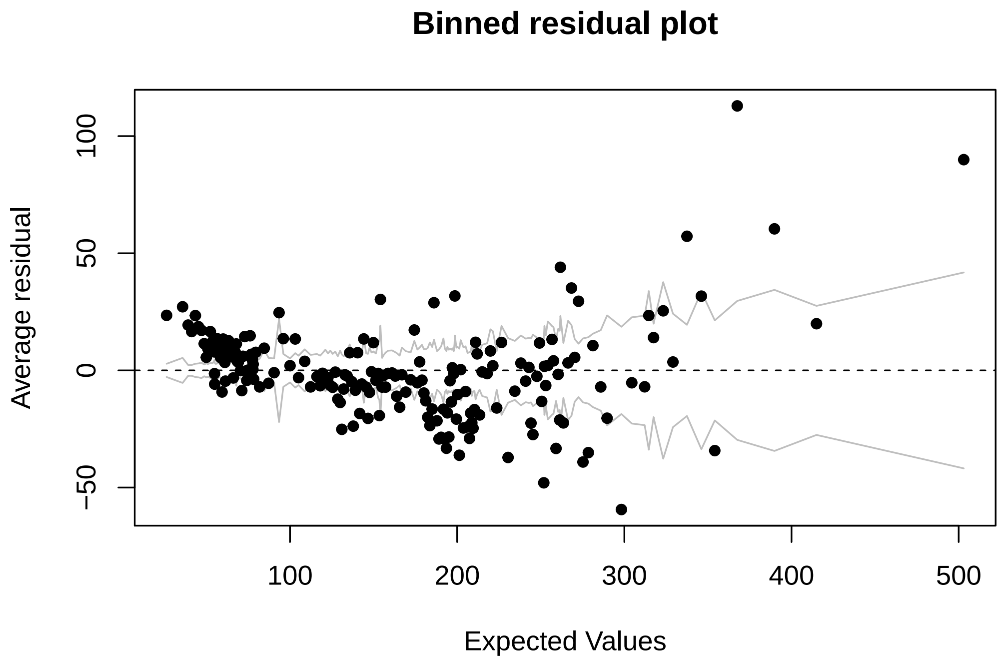
I plotted residual plot, normal Q-Q plot and binned residual plot to check the residual of the model I chose (lmer_fit2).



The residual plot shows a non-random pattern. It means the assumption that the relationship between price and other variables is linear is not reasonable. And the variances of the error terms may not be equal.



The points in the normal Q-Q plot for the residuals do not rest on a line. It means the model does not meet the normality assumptions.

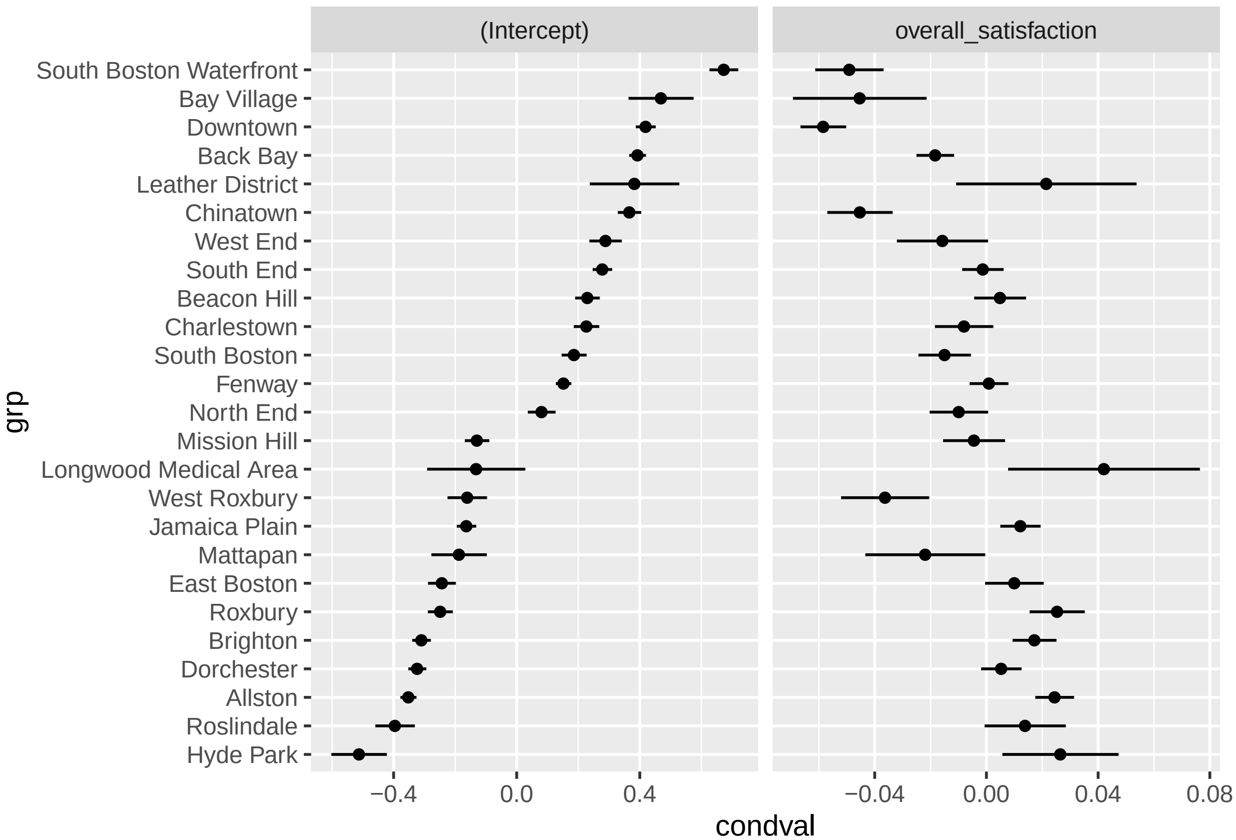


The binned residual plot shows that many points rest on the left. And there are some points falling outside of the 2 standard-error bounds.

Model Improvement

The residual plots show that the model we chose (lmer_fit2) do not fit the data well. To improve the model, I transformed the outcome variable by logging the price variable. The improved model is:

```
## lmer(formula = log.price ~ room_type + reviews + accommodates *
##     bedrooms + (1 + overall_satisfaction | neighborhood), data = boston_airbnb_log)
##                        coef.est coef.se
## (Intercept)             4.62     0.05
## room_typePrivate room -0.51     0.01
## room_typeShared room  -0.80     0.02
## reviews                 0.00     0.00
## accommodates            0.06     0.00
## bedrooms                0.21     0.01
## accommodates:bedrooms -0.01     0.00
##
## Error terms:
##  Groups       Name                 Std.Dev. Corr
##  neighborhood (Intercept)          0.33
##               overall_satisfaction 0.03     -0.66
##  Residual                          0.41
## ---
## number of obs: 38725, groups: neighborhood, 25
## AIC = 41346.5, DIC = 41184.3
## deviance = 41254.4
```

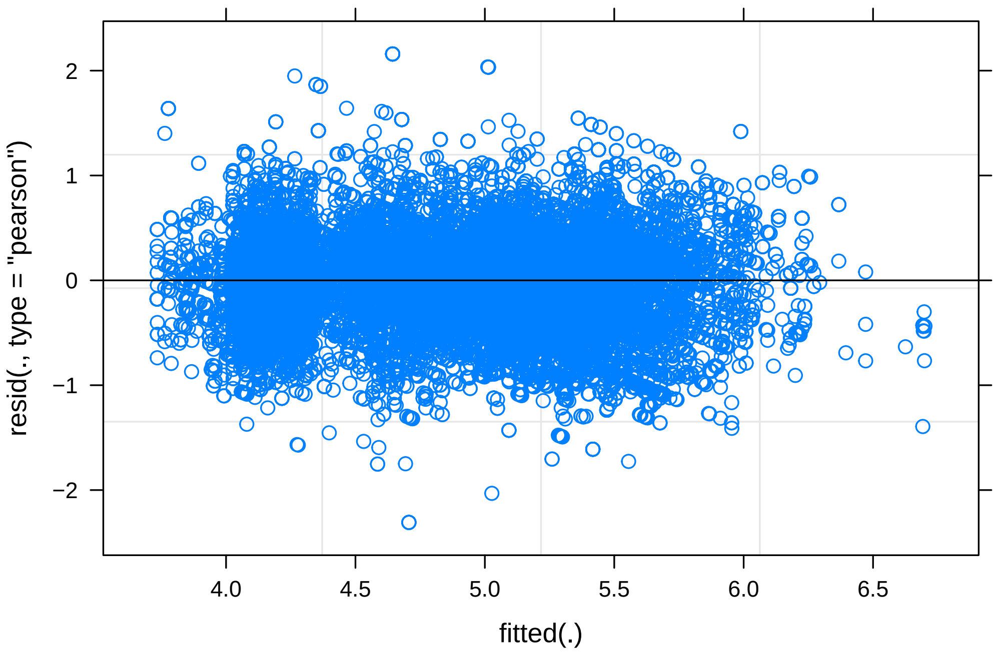


All of the coefficients are statistically significant. Private room has 40% ($exp(-0.51) = 0.6004956$) lower weighted price than the Entire home/apt. Shared room has 55% ($exp(-0.80) = 0.449329$) lower weighted price than the price of Entire home/apt. The coefficient of reviews is zero, this variable can be ignored. When accommodates increases by one, keeping other variables the same, price will increase by 5% ($exp(0.06 - 0.01) = 1.051271$). When the number of bedrooms increases by one, keeping other variables the same, price will increase by 22% ($exp(0.21 - 0.01) = 1.221403$). The variance among different neighborhoods are 0.33. The within-neighborhood variance is 0.41. The slope of overall satisfaction is 0.03, and there is correlation with intercept of -0.66. The random effects are shown on the plot above.

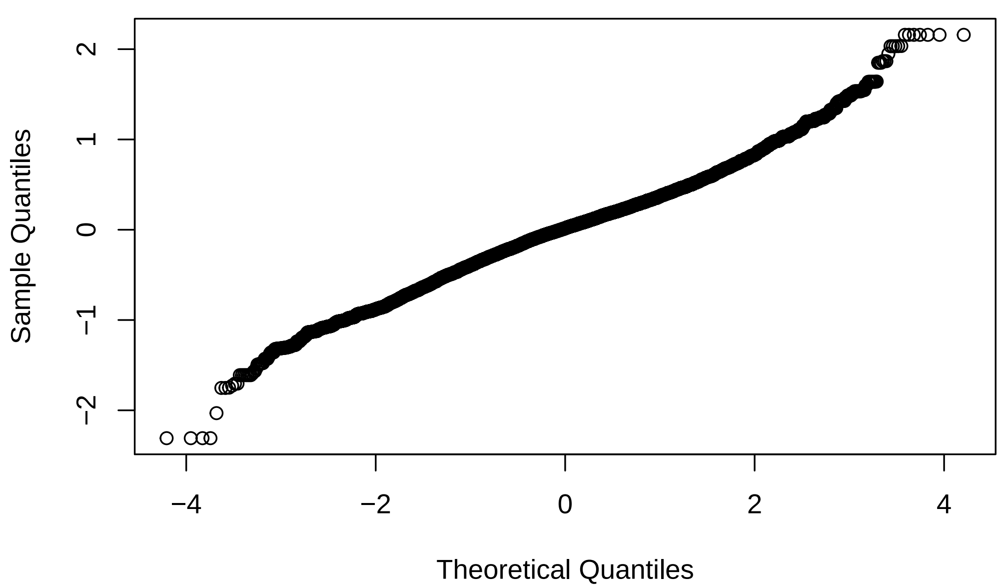
I also checked the AIC and BIC of the improved model. We can see that its AIC and BIC are much smaller than other models.

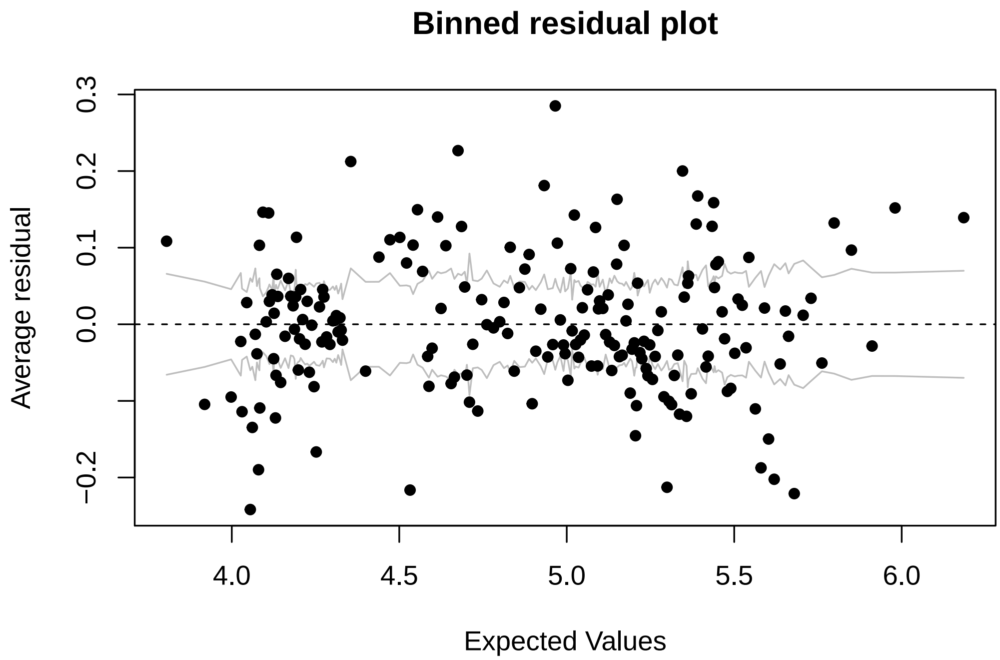
Model	AIC	BIC
lm_fit1	458068.34	458350.96
lm_fit2	464312.63	464389.71
lmer_fit1	458185.00	458270.65
lmer_fit2	457553.52	457647.72
lmer_fit3	41346.52	41440.72

I plotted residual plot, normal Q-Q plot and binned residual plot again to check the residual of the improved model (lmer_fit3).



Normal Q−Q Plot





The residual plot and normal Q-Q plot look much better. There is no pattern in the residual plot and most of the points in the normal Q-Q plot for the residuals rest on a line. For the binned residual plot, the residuals now form a "horizontal band" around the 0 line. However, more points fall outside of the 2 standard-error bounds.

Conclusion

Overall, the model with varying intercepts for neighborhood and varying slopes for overall satisfaction and transformed by logging the price variable (lmer_fit3) is my final choice. From the coefficients, we can see that the most important variable is room type. The second important variable is the number of bedroom. The neighborhood variable and the overall satisfaction variable also play important roles in the model.

The binned residual plot shows there are still problems in this models because there are many points falling outside of the 2 standard-error bounds. It may because there are still outliers in the data. A better way to do data cleaning is needed in the future. The number of reviews variable worths a deeper study. By now, the coefficient of the number of reviews variable is zero. There must be some relationships between the number of positive reviews and negative reviews. For future direction, I could collect the text data of the reviews and do sentiment analysis of the reviews text. Then I could use the number of positive reviews and negative reviews as new predictors.

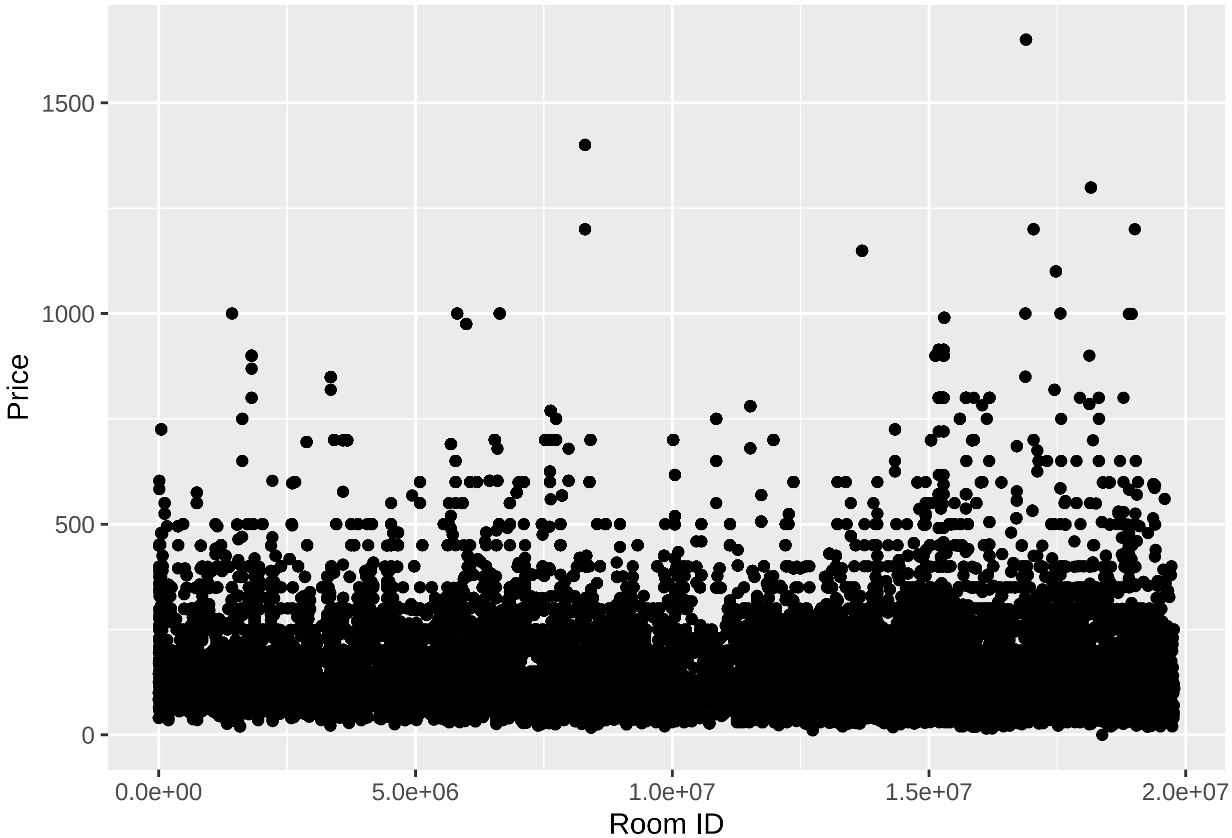
References

"Airbnb Data Collection: Get the Data." n.d. http://tomslee.net/airbnb-data-collection-get-the-data.

"Airbnb." 2019. https://en.wikipedia.org/w/index.php?title=Airbnb&action=edit.

Appendix A: Outliers Recognition

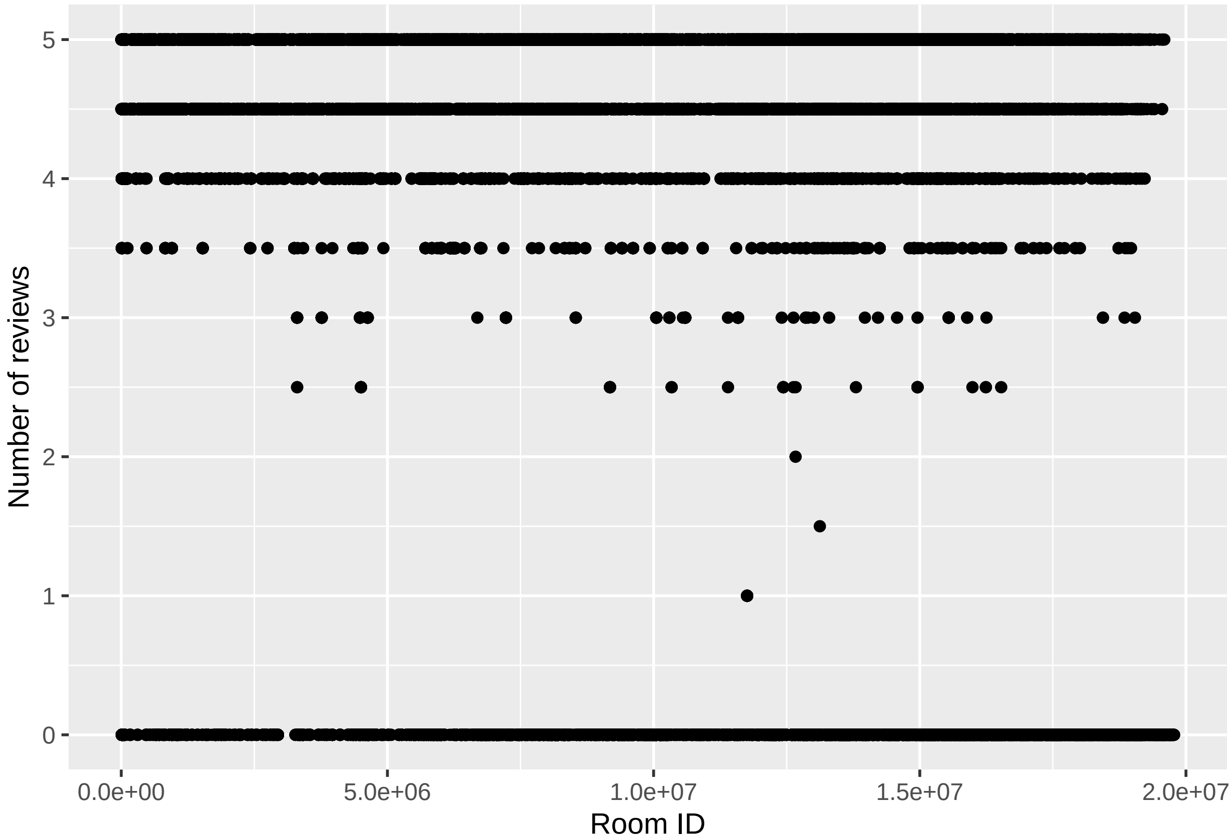
price




```
##  [1] room_id              host_id              room_type
##  [4] neighborhood         reviews              overall_satisfaction
##  [7] accommodates         bedrooms             price
## [10] latitude             longitude            date
## <0 rows> (or 0-length row.names)

##  [1] room_id              host_id              room_type
##  [4] neighborhood         reviews              overall_satisfaction
##  [7] accommodates         bedrooms             price
## [10] latitude             longitude            date
## <0 rows> (or 0-length row.names)

##  [1] room_id              host_id              room_type
##  [4] neighborhood         reviews              overall_satisfaction
##  [7] accommodates         bedrooms             price
## [10] latitude             longitude            date
## <0 rows> (or 0-length row.names)

##  [1] room_id              host_id              room_type
##  [4] neighborhood         reviews              overall_satisfaction
##  [7] accommodates         bedrooms             price
## [10] latitude             longitude            date
## <0 rows> (or 0-length row.names)
```

I would like to delete them because their price are too high. It may affect the regression results.

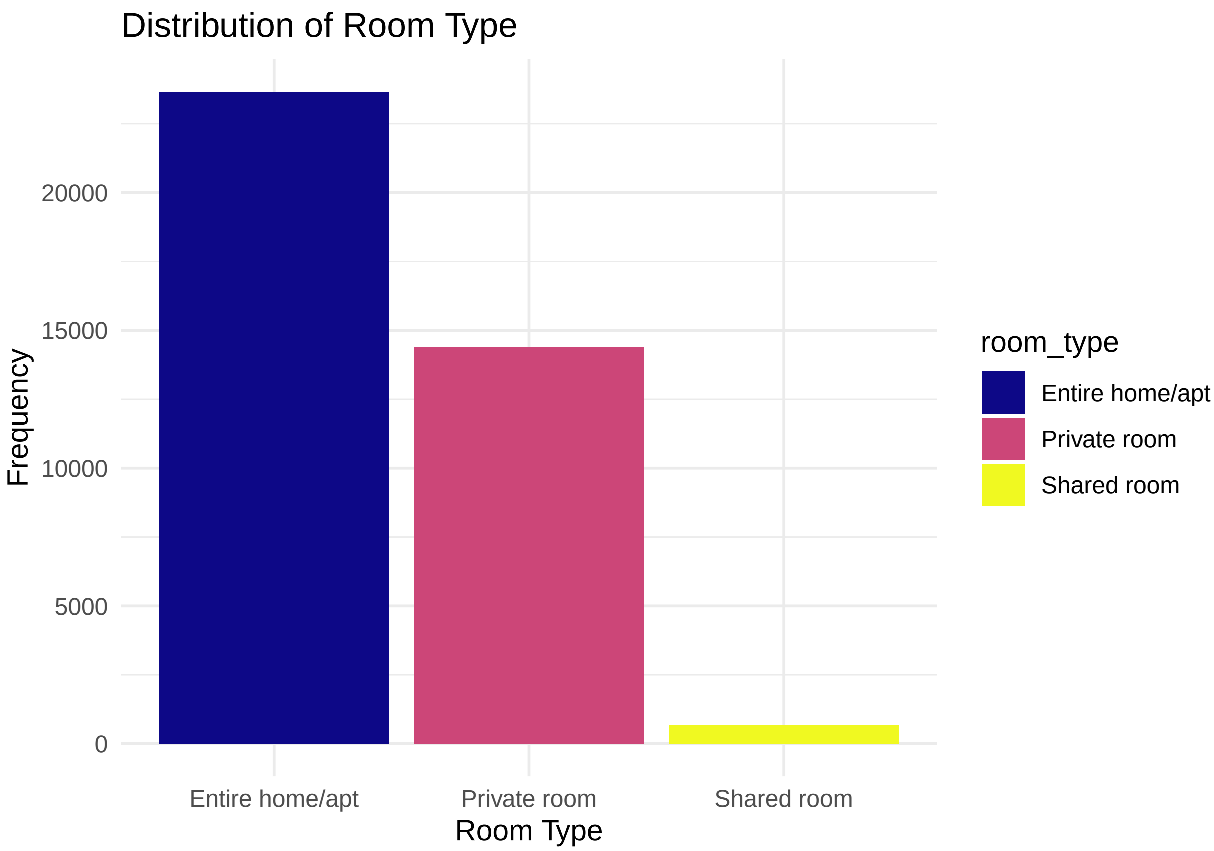
overall_satisfaction



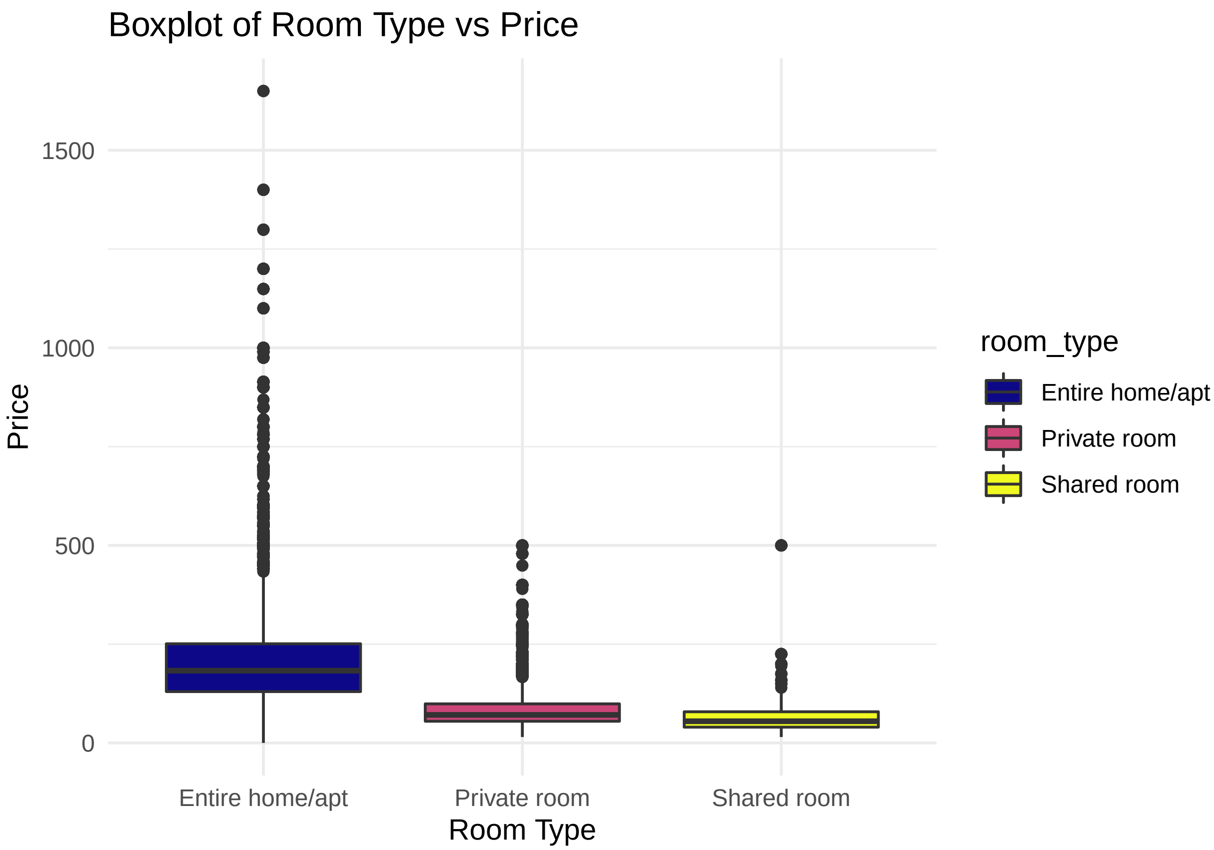

```
##    room_id  host_id    room_type neighborhood reviews overall_satisfaction
## 1 11757251 26873897 Private room     Brighton       3                    1
## 2 11757251 26873897 Private room     Brighton       3                    1
## 3 11757251 26873897 Private room     Brighton       3                    1
## 4 11757251 26873897 Private room     Brighton       3                    1
## 5 11757251 26873897 Private room     Brighton       3                    1
## 6 11757251 26873897 Private room     Brighton       3                    1
## 7 11757251 26873897 Private room     Brighton       3                    1
## 8 11757251 26873897 Private room     Brighton       3                    1
## 9 11757251 26873897 Private room     Brighton       3                    1
##   accommodates bedrooms price latitude longitude       date
## 1            1        1    30 42.34538 -71.13764 2016-07-16
## 2            1        1    30 42.34538 -71.13764 2016-08-19
## 3            1        1    30 42.34538 -71.13764 2016-09-16
## 4            1        1    30 42.34538 -71.13764 2016-10-18
## 5            1        1    30 42.34538 -71.13764 2017-02-16
## 6            1        1    30 42.34538 -71.13764 2017-03-12
## 7            1        1    30 42.34538 -71.13764 2017-04-08
## 8            1        1    30 42.34538 -71.13764 2017-05-05
## 9            1        1    30 42.34538 -71.13764 2017-06-10
```

I will keep these rating of 1.0 points because they are reasonable.

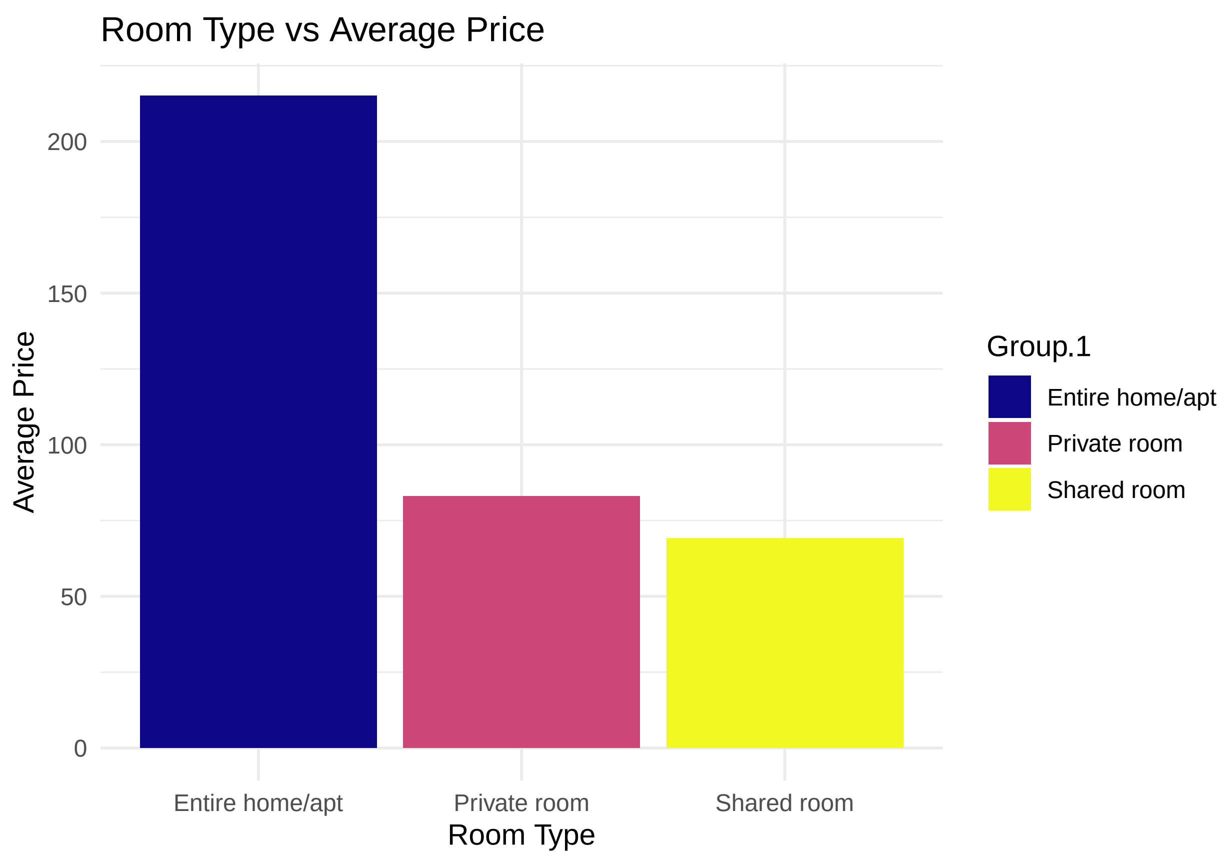
Appendix B: Additional Graphs



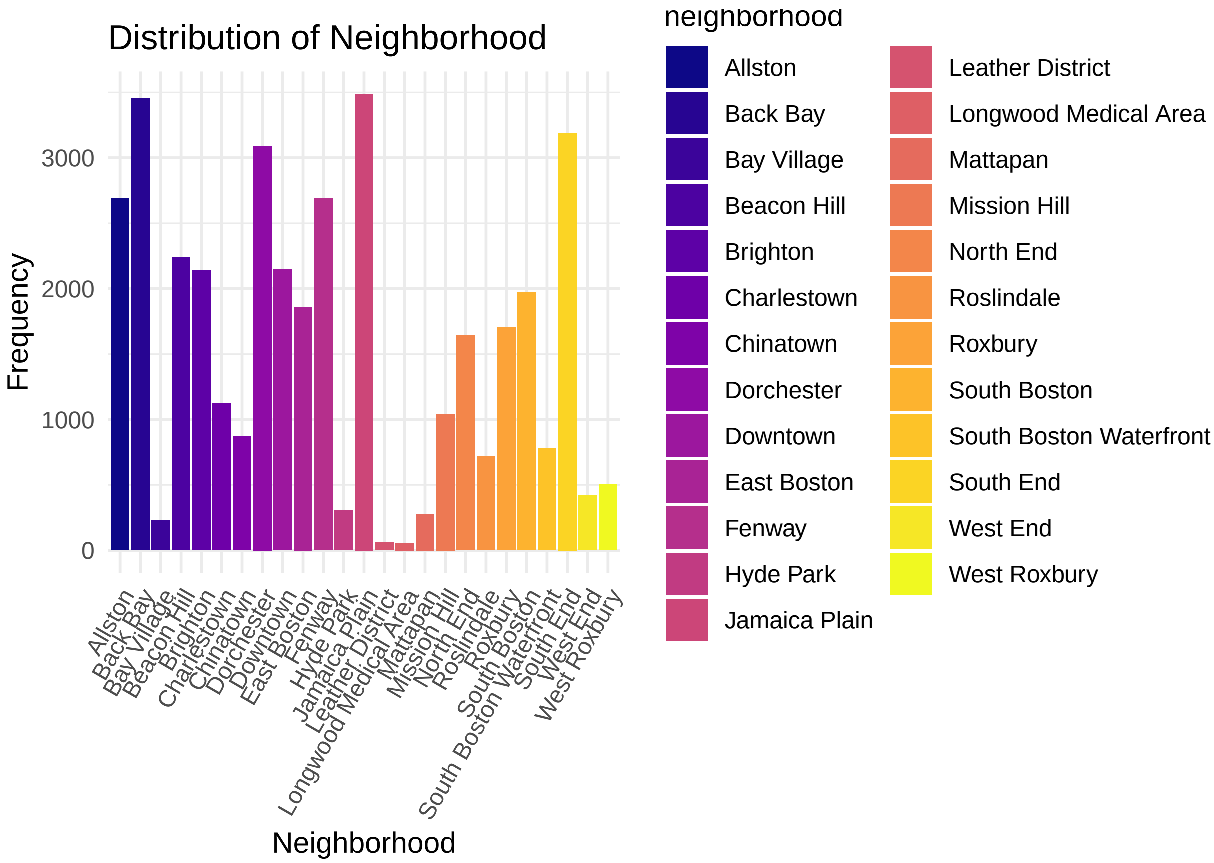
We can see that most rooms are entire home/apartment and private room. Only a few of the rooms are shared room.



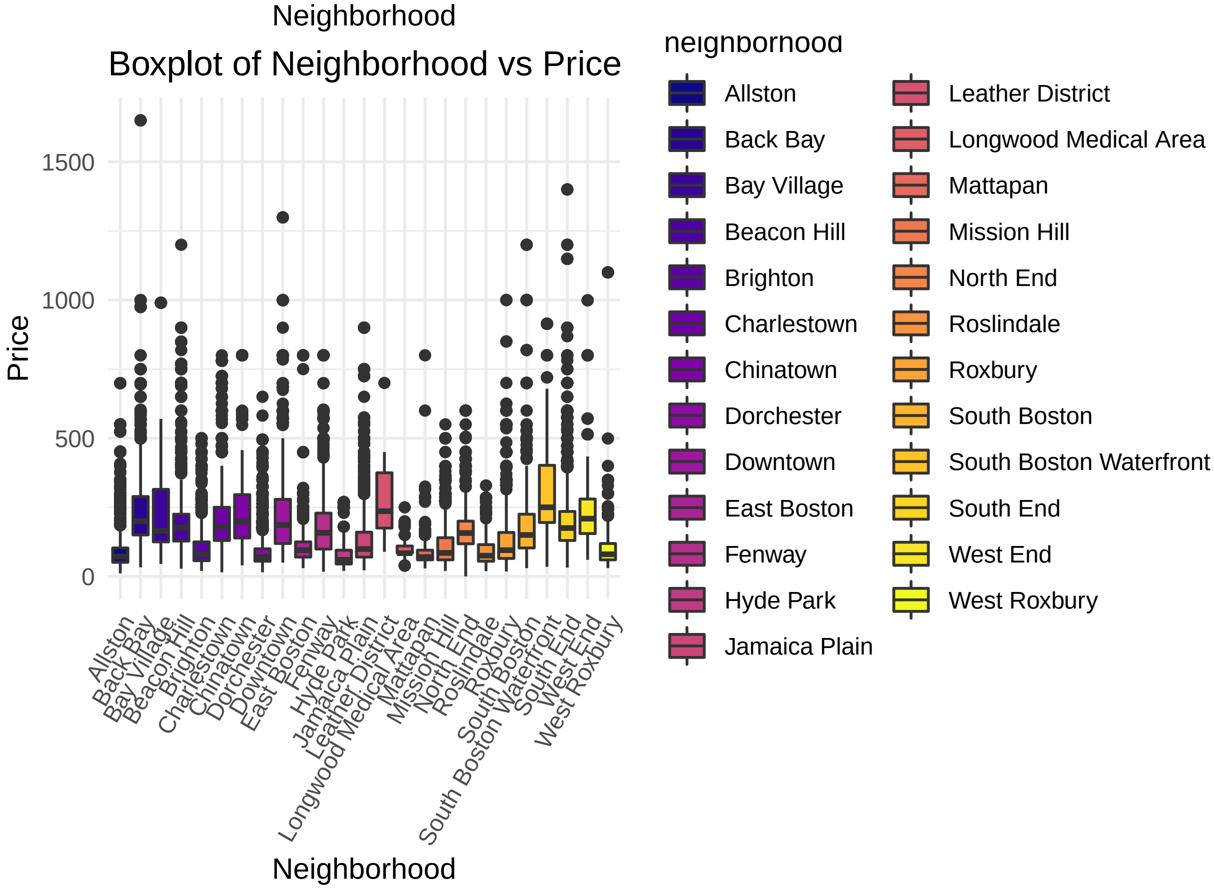
The average price of the three room types are close, but for the entire home/apartment type, there exis points with relatively high price.



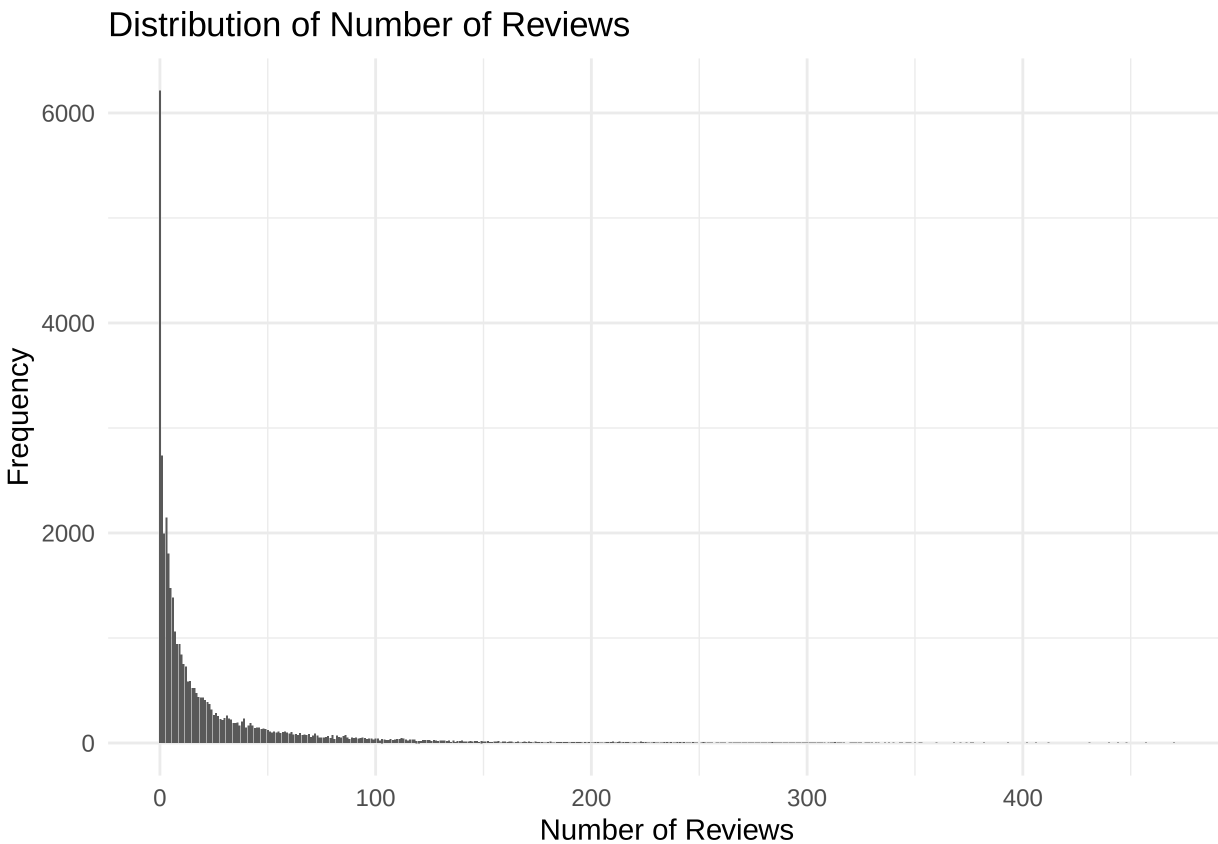
Entire room/apartment' average price is much higher than the other two types.



Distribution of Neighborhood
Frequency
3000
2000
1000
0
Neighborhood
Allston
Back Bay
Bay Village
Beacon Hill
Brighton
Charlestown
Chinatown
Dorchester
Downtown
East Boston
Fenway
Hyde Park
Jamaica Plain
Leather District
Longwood Medical Area
Mattapan
Mission Hill
North End
Roslindale
Roxbury
South Boston
South Boston Waterfront
South End
West End
West Roxbury
neighborhood



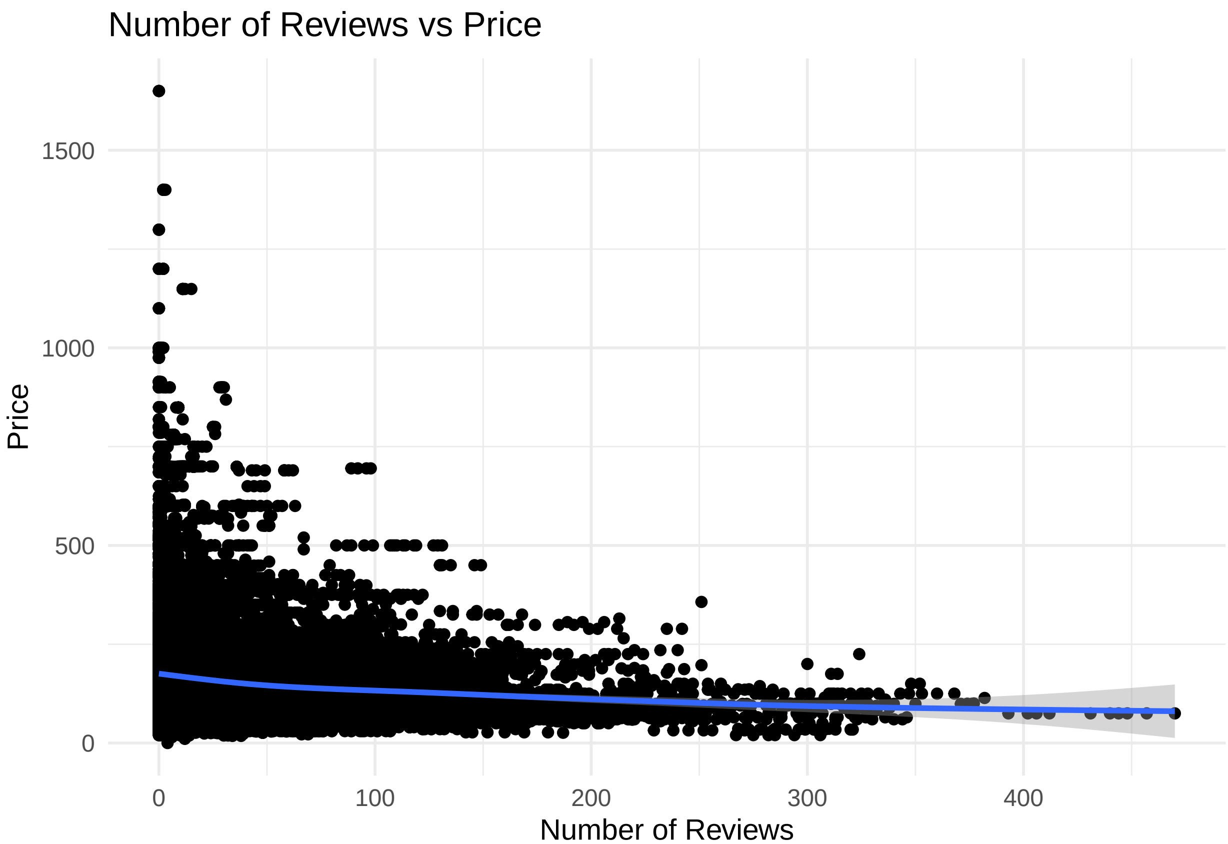
Boxplot of Neighborhood vs Price
Price
1500
1000
500
0
Neighborhood
Allston
Back Bay
Bay Village
Beacon Hill
Brighton
Charlestown
Chinatown
Dorchester
Downtown
East Boston
Fenway
Hyde Park
Jamaica Plain
Leather District
Longwood Medical Area
Mattapan
Mission Hill
North End
Roslindale
Roxbury
South Boston
South Boston Waterfront
South End
West End
West Roxbury
neighborhood




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    2.00    8.00   23.55   27.00  470.00
```

Many listings have zero review. And the number of reviews varies a lot.

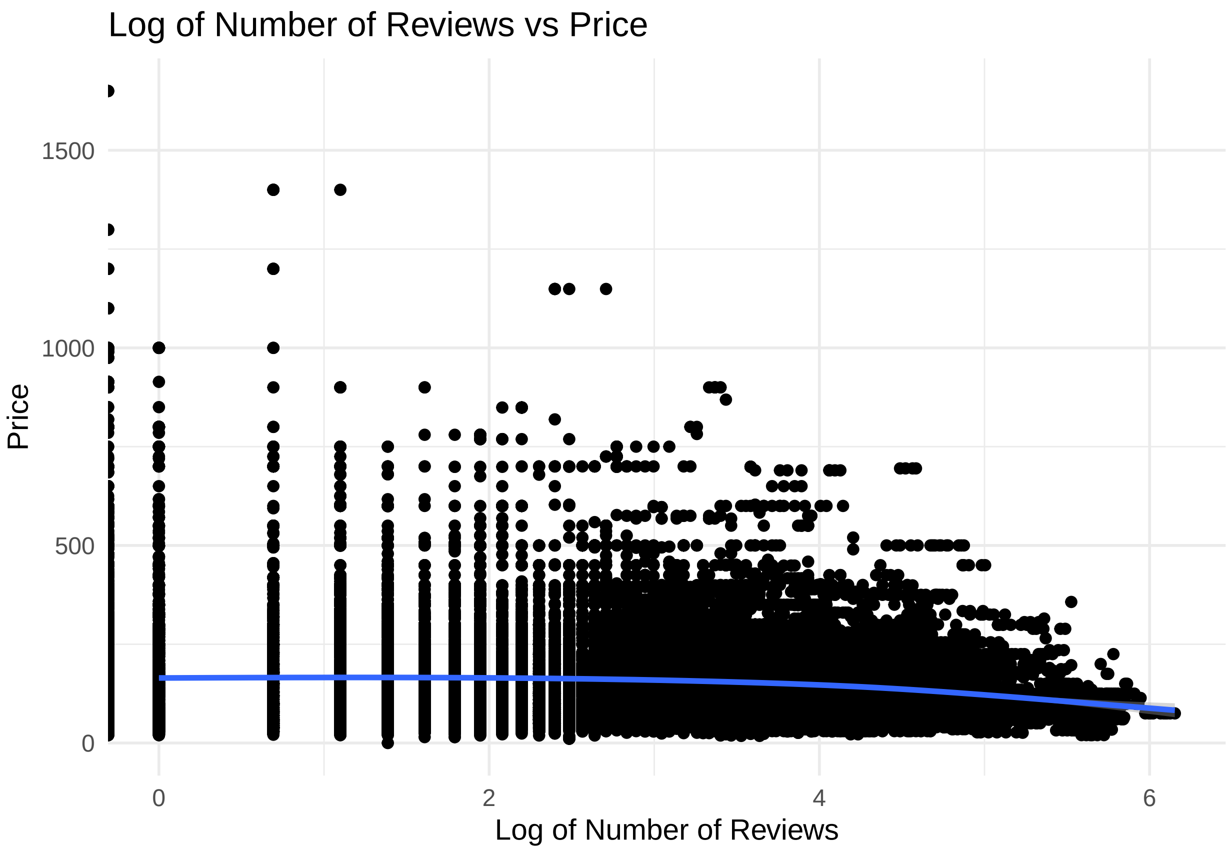
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



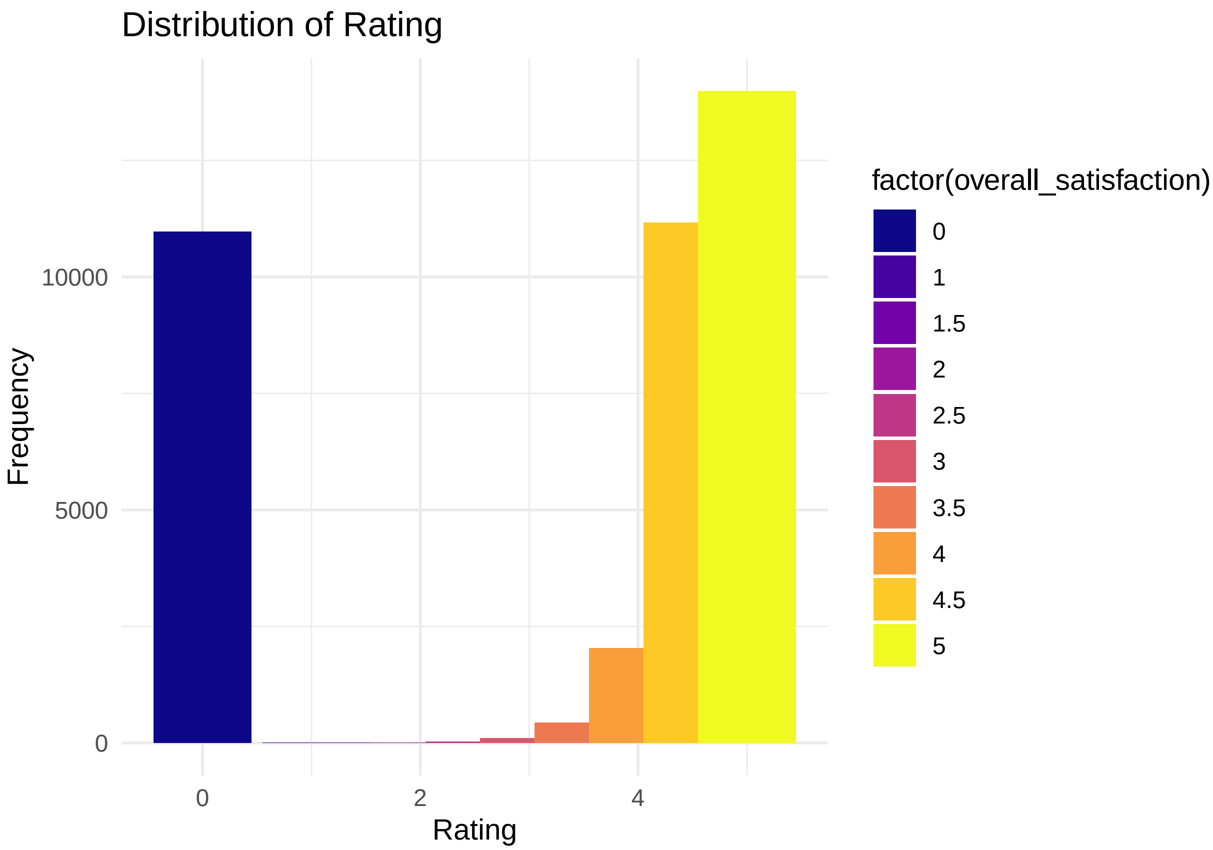
It shows more revies appear when the price is lower.

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 6209 rows containing non-finite values (stat_smooth).
```




```
## Warning: position_stack requires non-overlapping x intervals
```

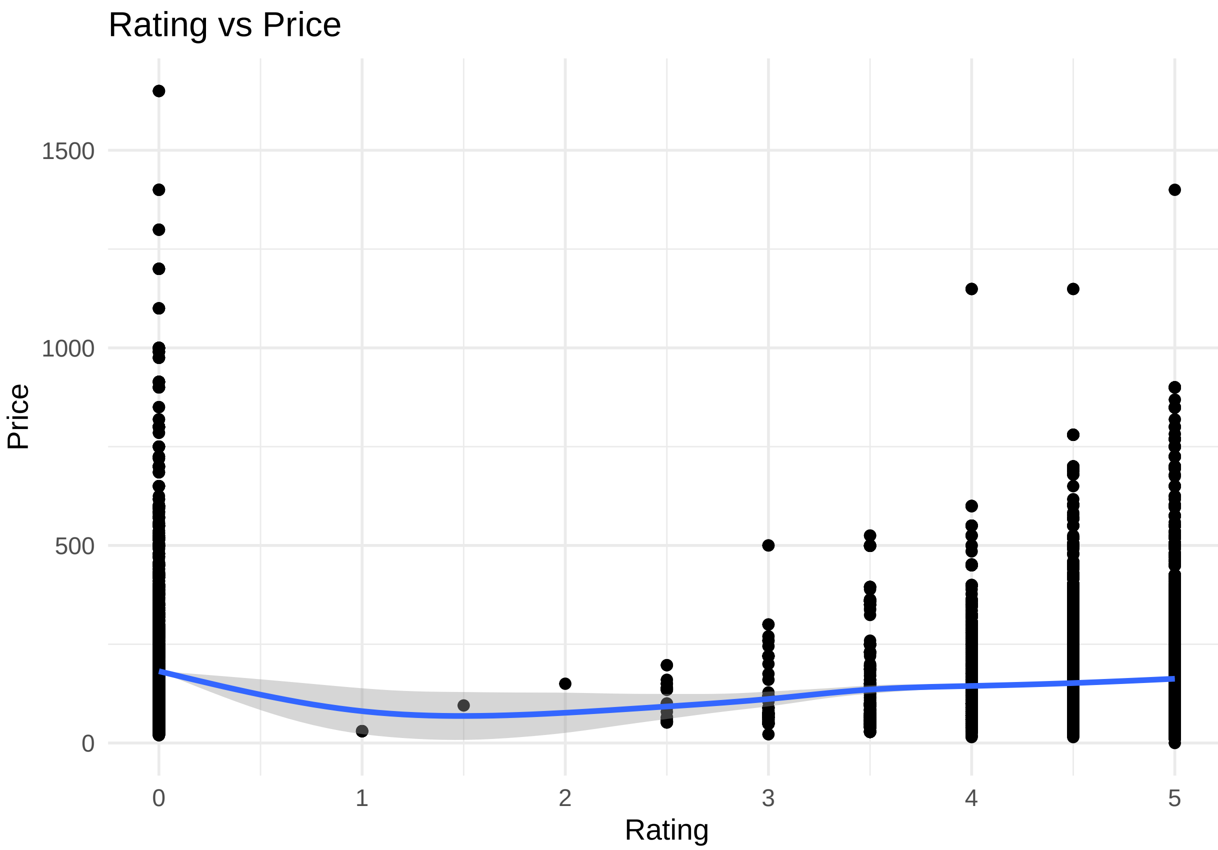



```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   4.500   3.363   5.000   5.000
```

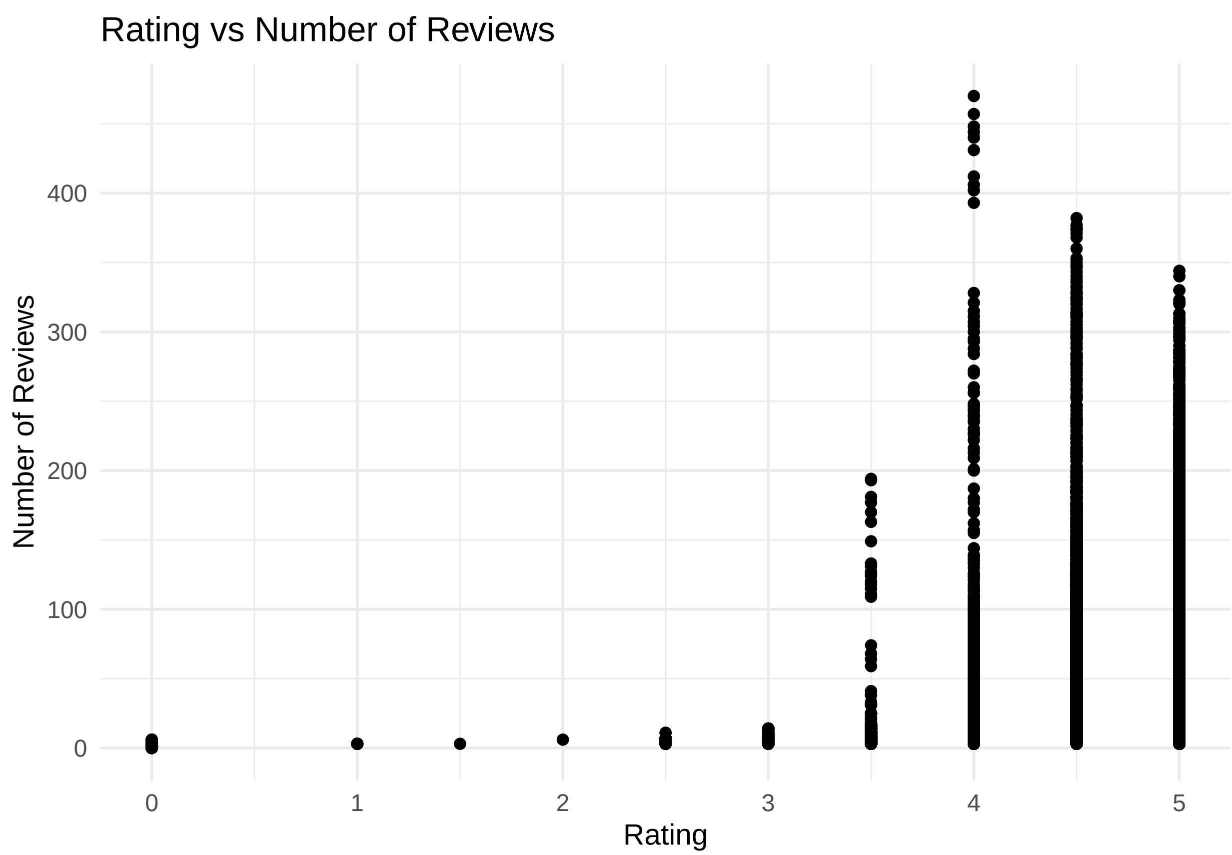
The rating is polarize. Mosting are 0 or 4.5~5.

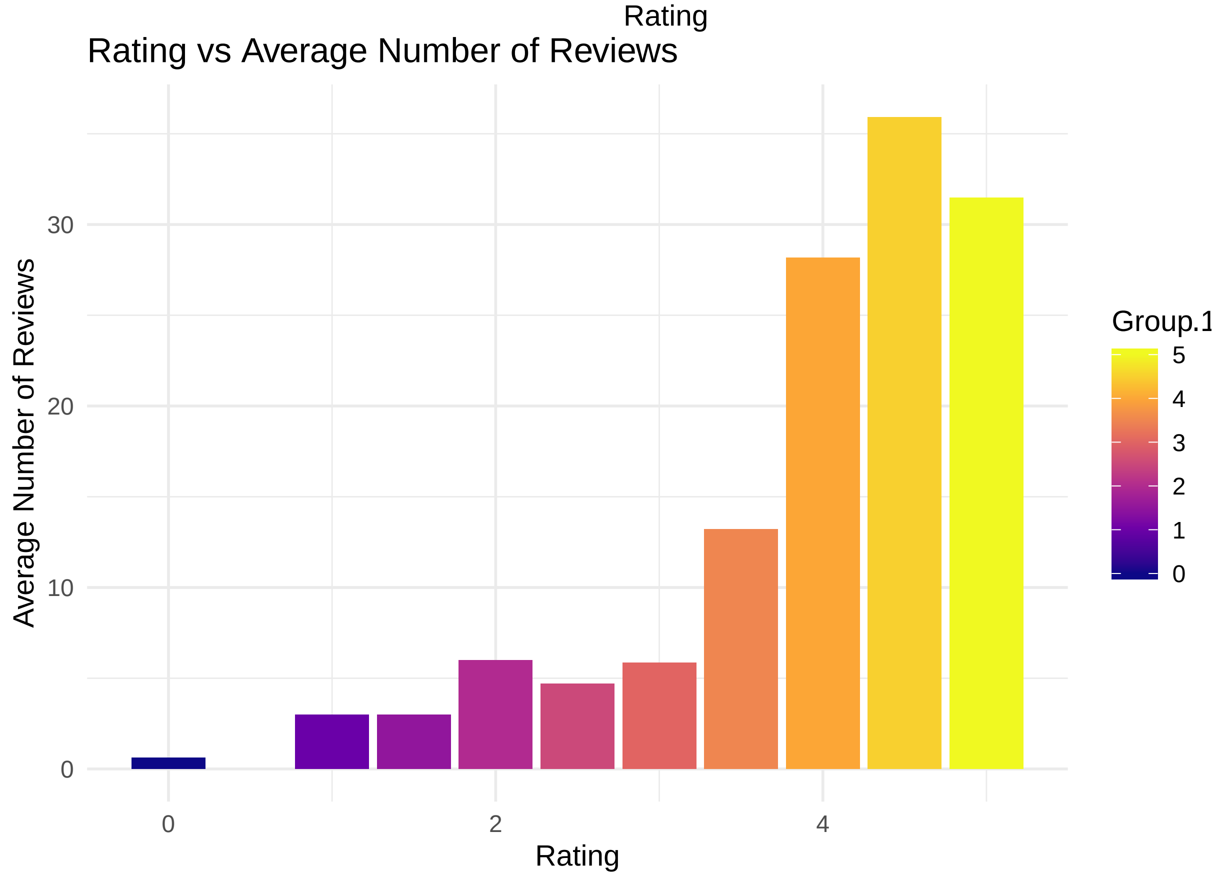
For those listings with 0 rating, most of them have 0 reviews, which means very likely that no one have been there. It make sense the rating is low.

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

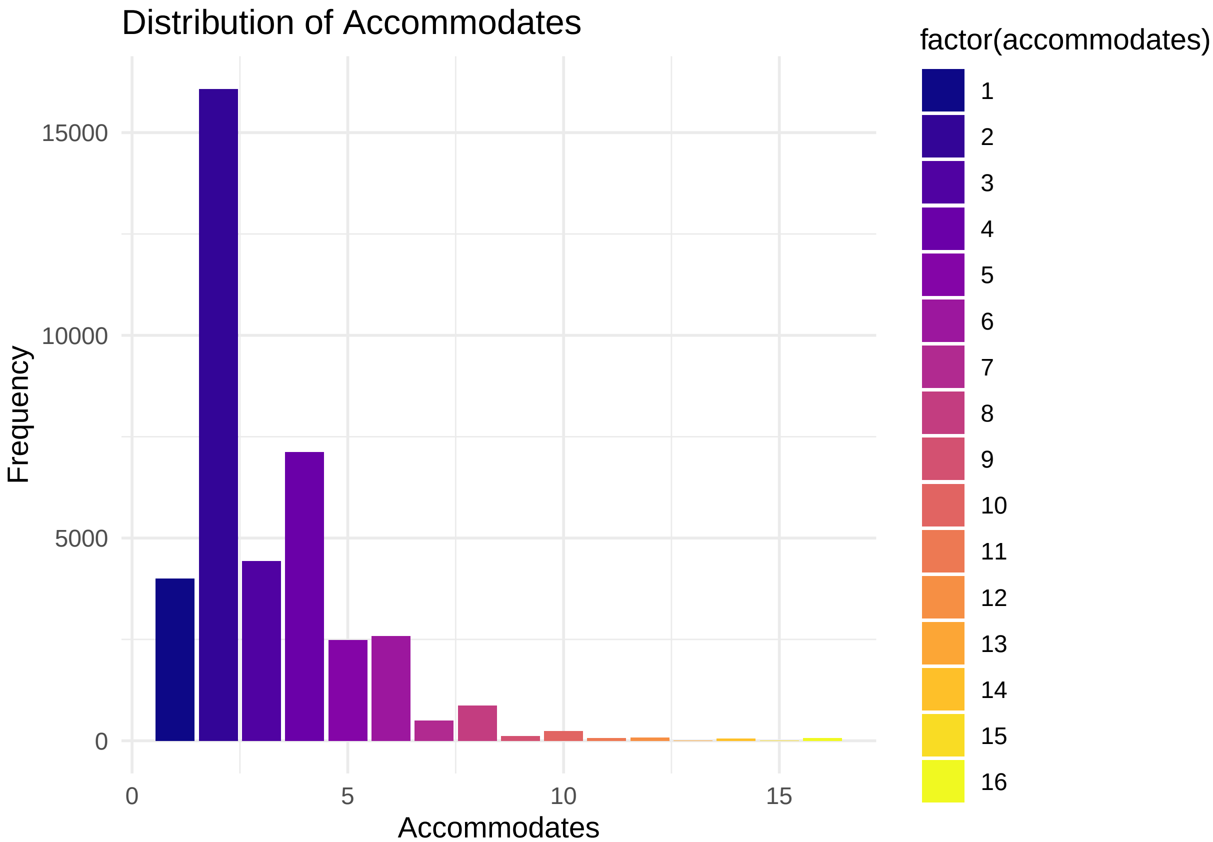


If we ignore the zero rating, the rating gets higher when the price turns higher.



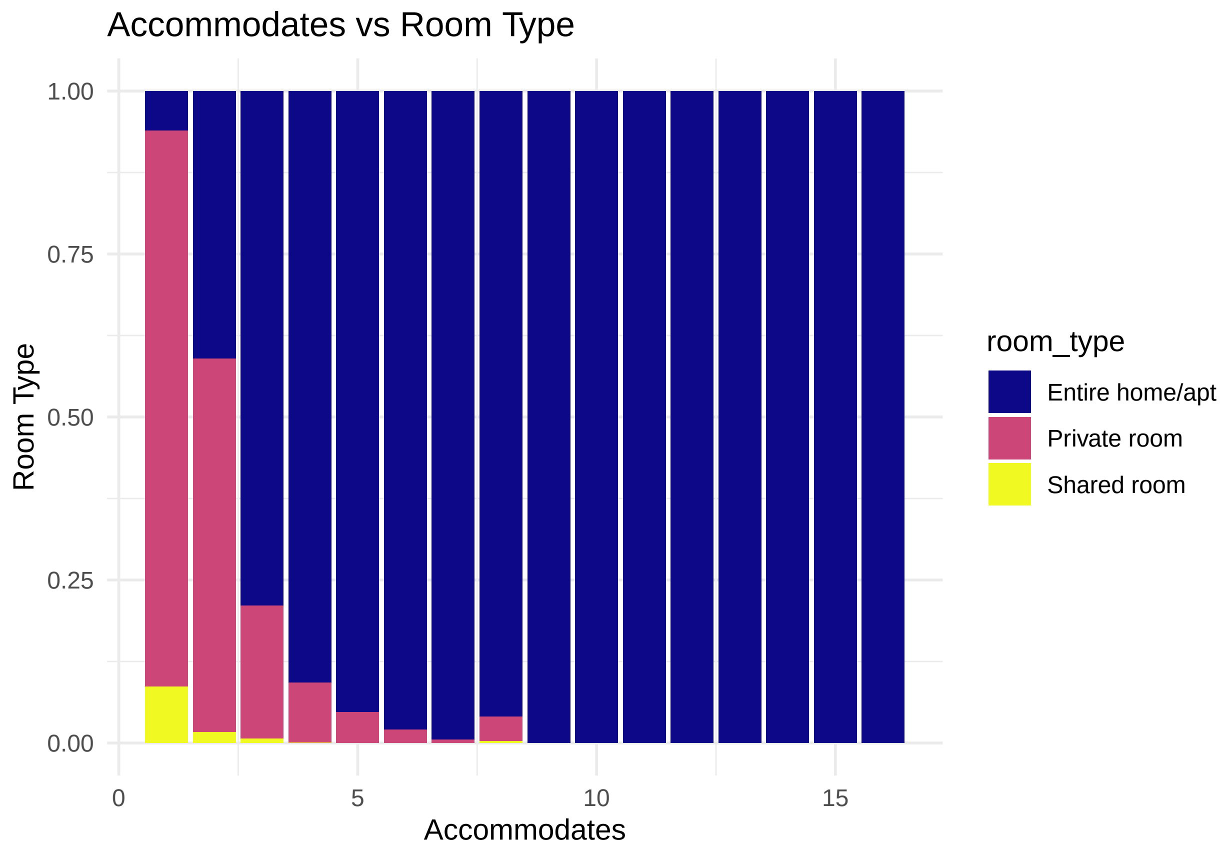


From the result, we can see that listings with rating 4.5 have highest number of reviews. It makes sense because listings with rating 5 many be expensive then fewer people will choose them.



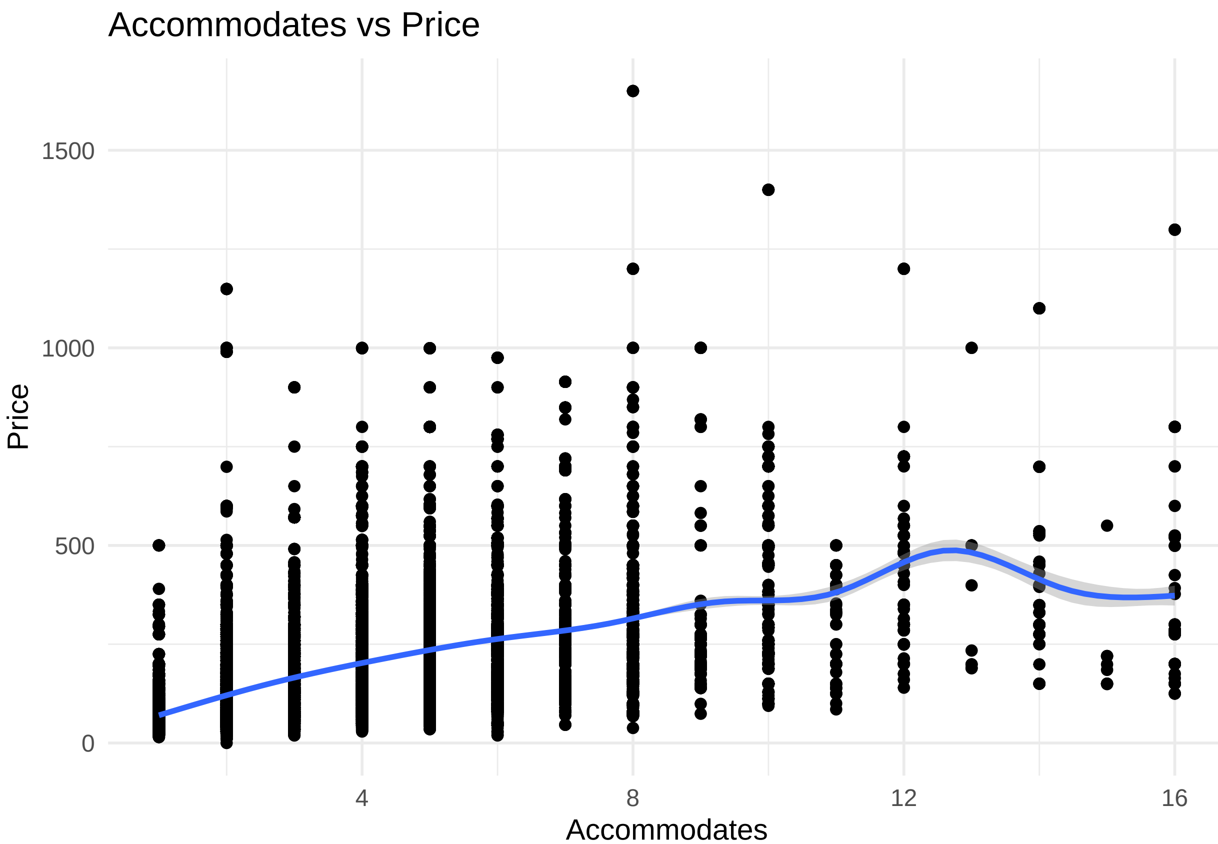

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   2.000   3.191   4.000  16.000
```

Listings with two accommodates are most common in Boston.

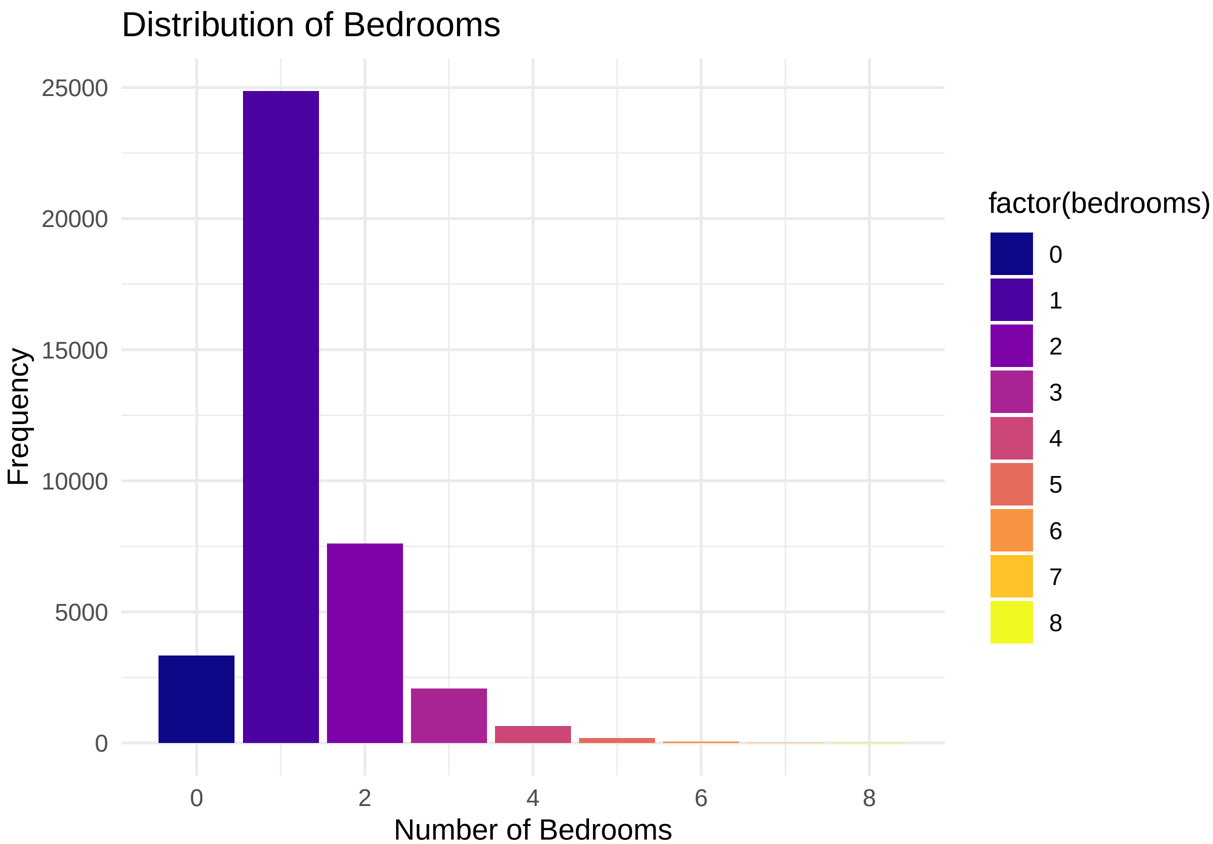


There is relationship between accommodates and room type. Most private rooms and shared rooms accept 1 to 2 accommodates. A listing with accommodates larger than five is most possibly an entire home or apartment.

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

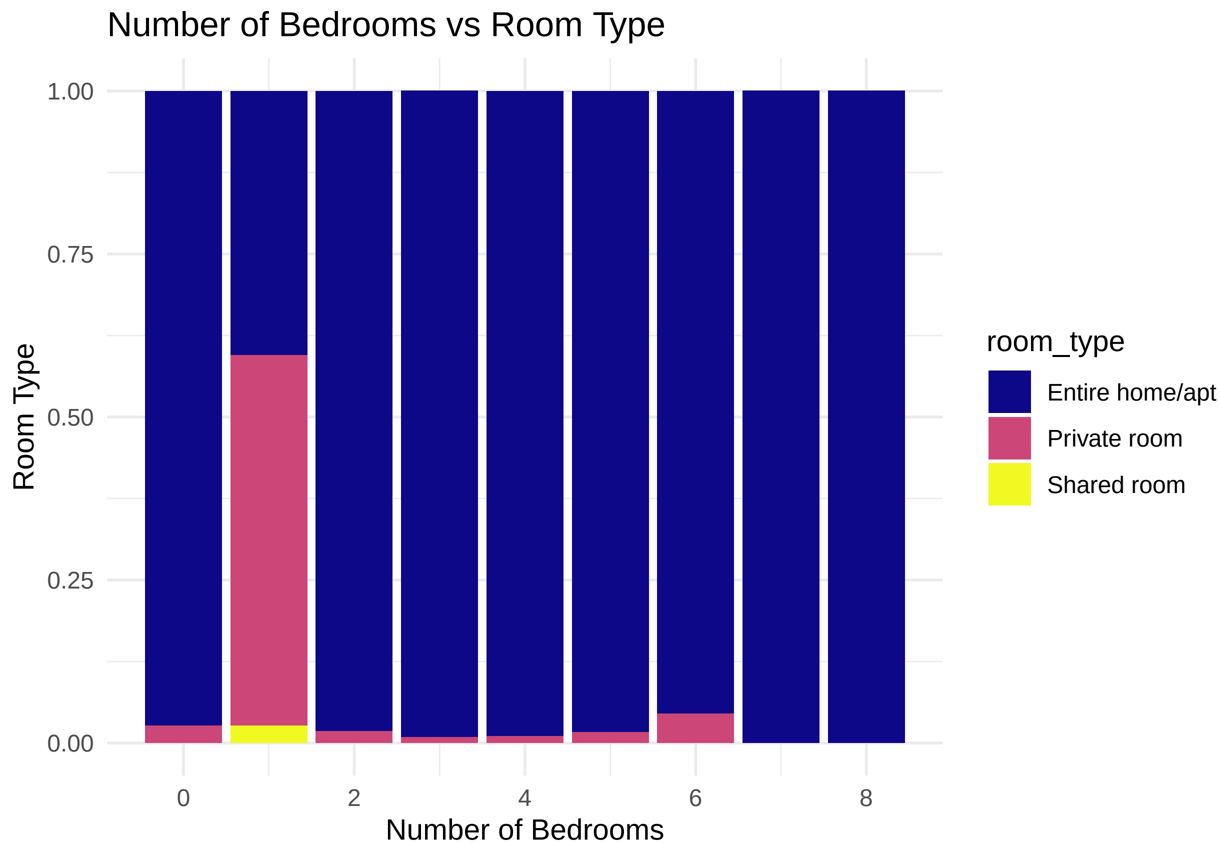


It seems when the number of accommodates grows, the price grows a little.

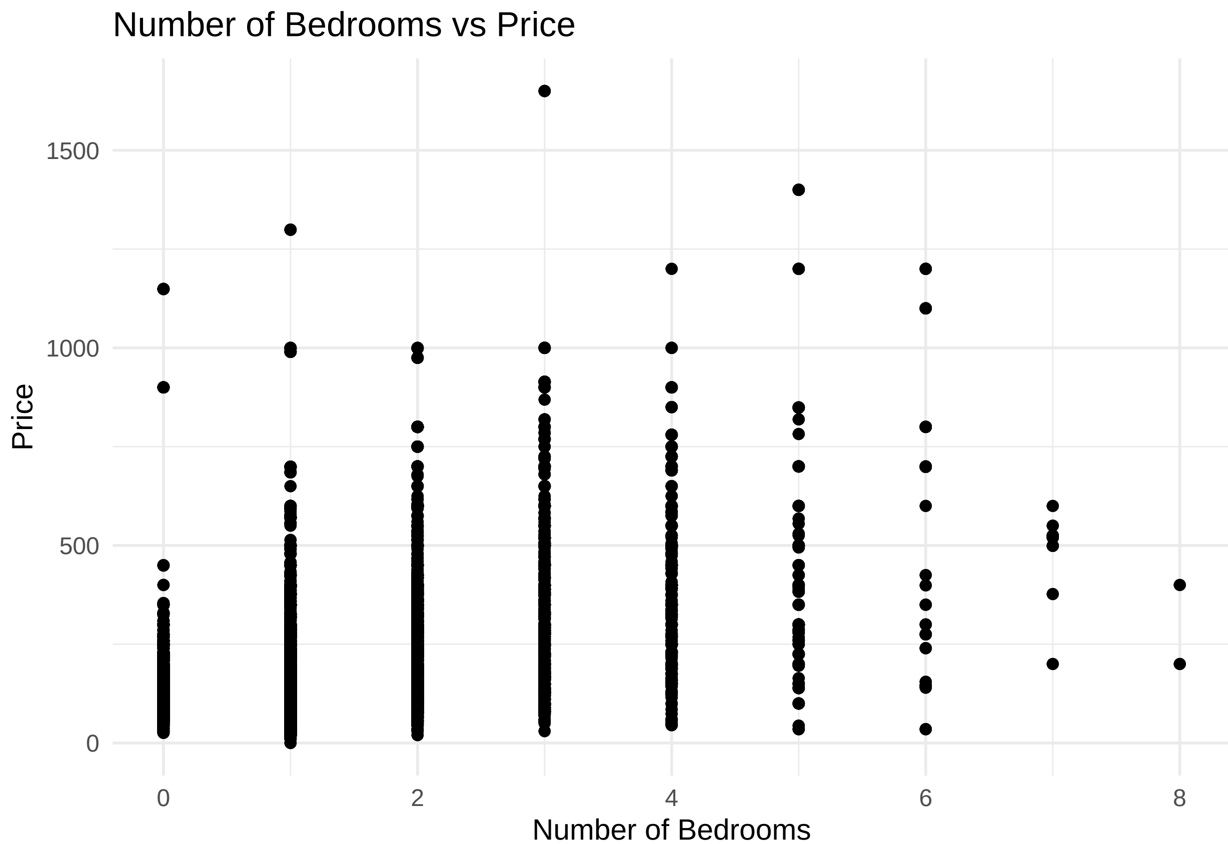



```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   1.000   1.291   2.000   8.000
```

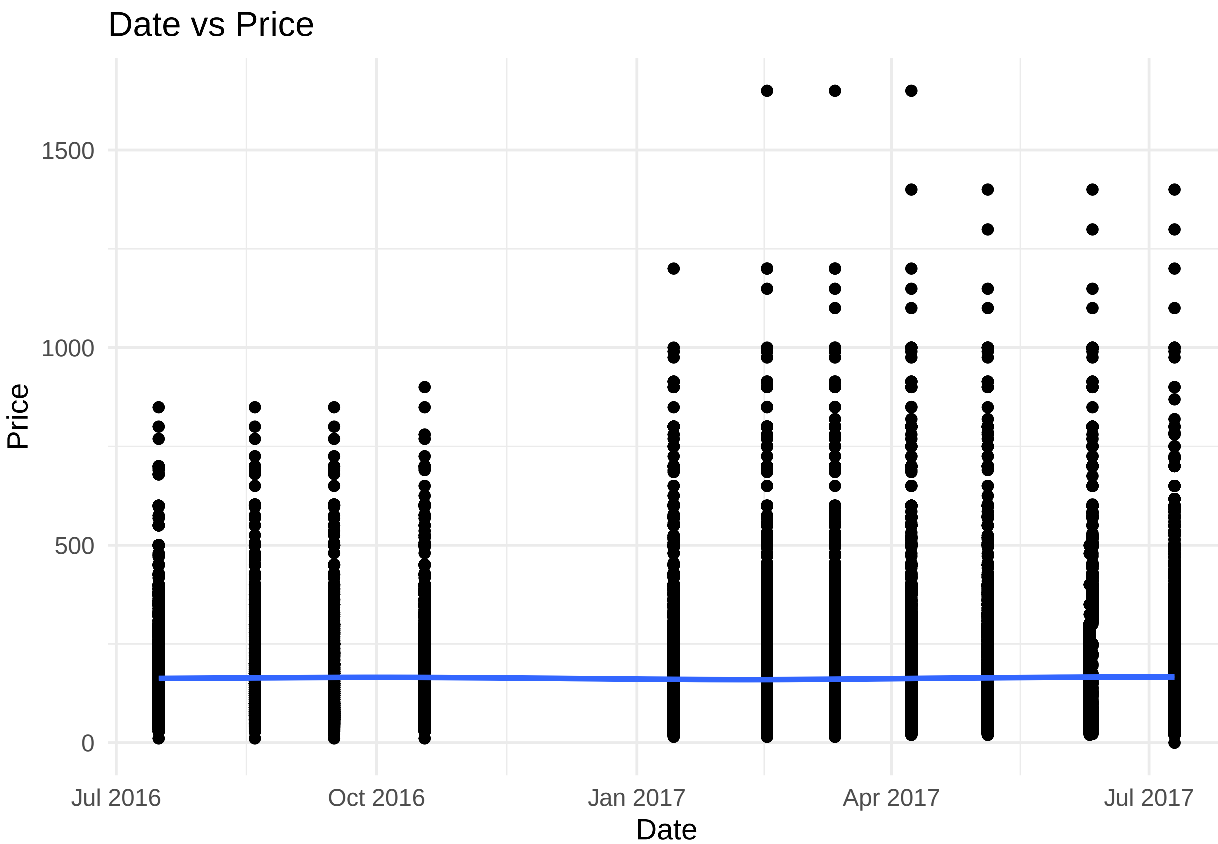
Listings with 1 rooms are most common in Boston.



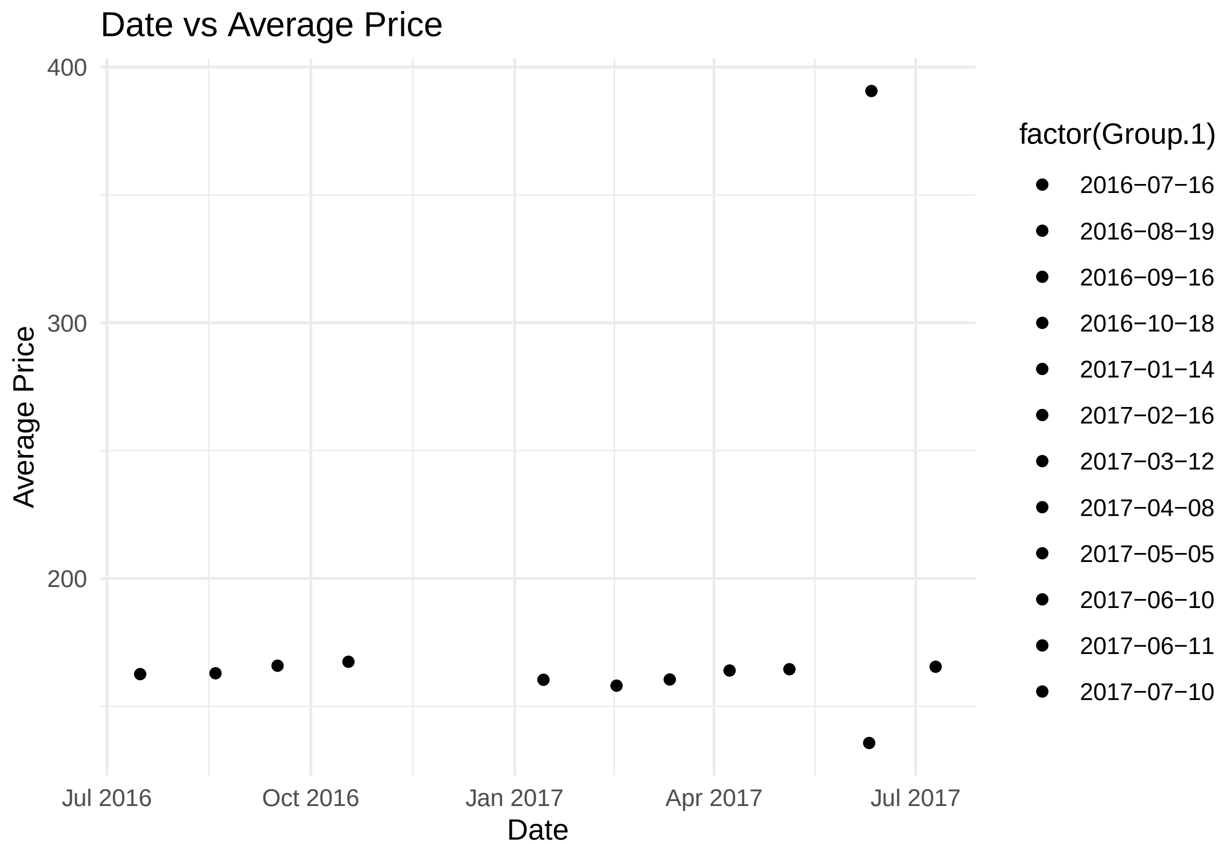
There is relationship between number of Bedrooms and room type.




```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



After Jan, 2017, some listings with higher price appeared. And the price is a little bit rising.



	Group.1	x
1	2016-07-16	162.5953
2	2016-08-19	162.9219
3	2016-09-16	165.8719
4	2016-10-18	167.4147
5	2017-01-14	160.3747
6	2017-02-16	158.0746
7	2017-03-12	160.4793
8	2017-04-08	163.9974
9	2017-05-05	164.5051
10	2017-06-10	167.1857
12	2017-07-10	165.4677

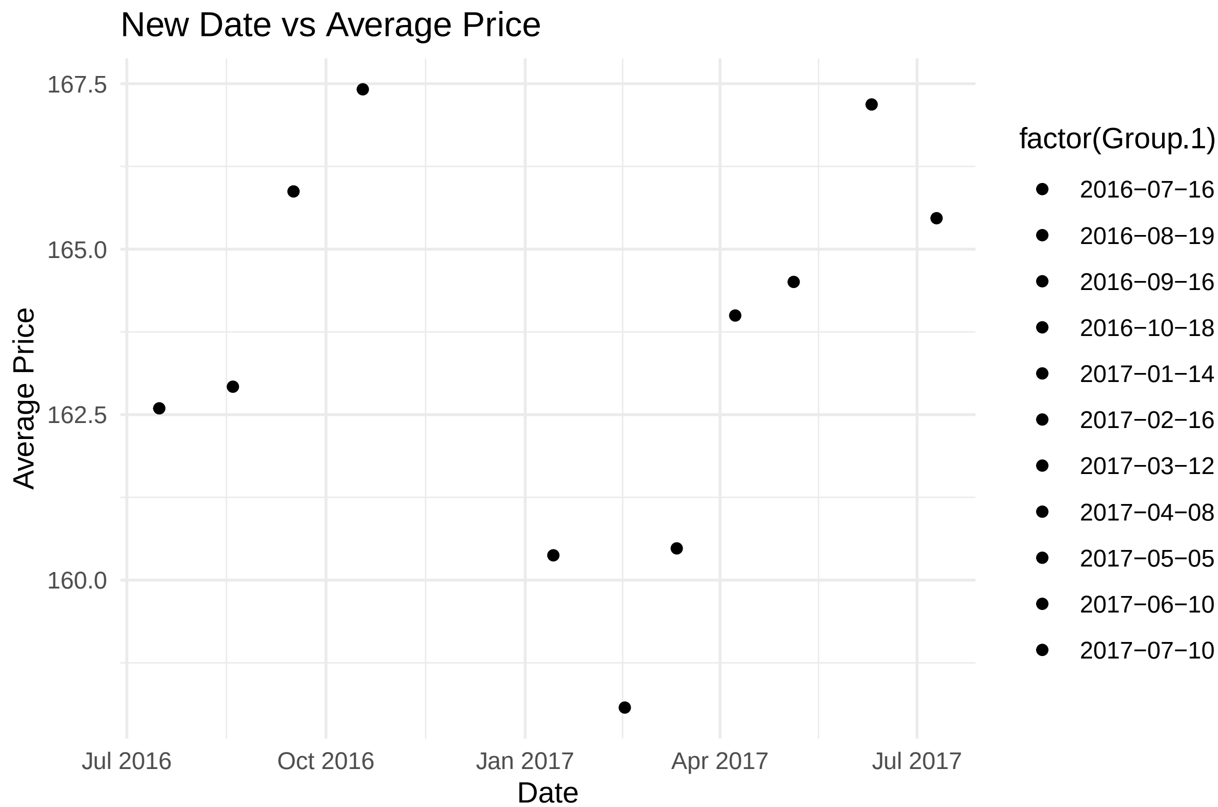


Figure 20

The price keep relatively stable over the time period between July, 2016 to July, 2017.