MA678_Report

Yirong Yuan

2022-12-11

Github Link:https://github.com/yiryuan/MA678_Final_Project.git (https://github.com/yiryuan/MA678_Final_Project.git)

<Abstract>

This project focuses on estimating the price of Boston Airbnb. In this project, I used the dataset which is downloaded from http://data.insideairbnb.com/united-states/ma/boston/2022-09-15/visualisations/listings.csv. (http://data.insideairbnb.com/united-states/ma/boston/2022-09-15/visualisations/listings.csv.) To estimate the price, I explored how factors, such as room type, minimum night, and so on, influenced the price of Boston Airbnb. The project has two parts: Exploratory Data Analysis and Multilevel Bayesian Analysis. Exploratory Data Analysis visually displayed the relationship between factor and price. In Multilevel Bayesian Analysis, I fitted a multilevel model based on the different neighborhoods to Airbnb prices for rentals in the Boston market to investigate the neighborhood heterogeneity.

<Introducation>

<Background>

Since 2007, Airbnb (ABNB) has been an online marketplace that connects people who want to rent out their homes with people looking for accommodations in specific locales. Today, Airbnb is a service operated and recognized around the world. Data analysis of the millions of listings offered through Airbnb is a critical element of the company. The massive amounts of data can be analyzed and used for security, business decision-making, understanding customer and provider (host) behavior and performance on the platform, guiding marketing programs, implementing innovative add-ons Serve, and more.

Location is key to valuable real estate. Homes in cities with little room for expansion are more beneficial than those with plenty of space. Consider the accessibility, appearance, and amenities of a neighborhood, as well as development plans. The community also influences the price of Airbnb.

<Data>

I selected the target variable ‘price’ and a subset of 9 predictor variables from the original data listings.csv. I deleted the rows in which the prices of Airbnb are negative because it did not make sense.

id: The Airbnb id number

price: the Airbnb price

latitude and longitude: The geographical information

neighbourhood: The name of neighbourhood

room_type: Type of room (Entire home/apt, Private room and Shared room)

minimum_nights: The minimum number of nights that a guest can book

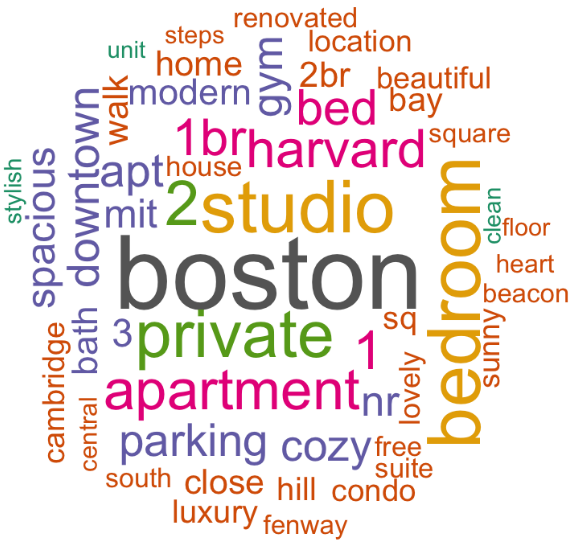
number_ of_reviews: Total number of Airbnb guest ratings

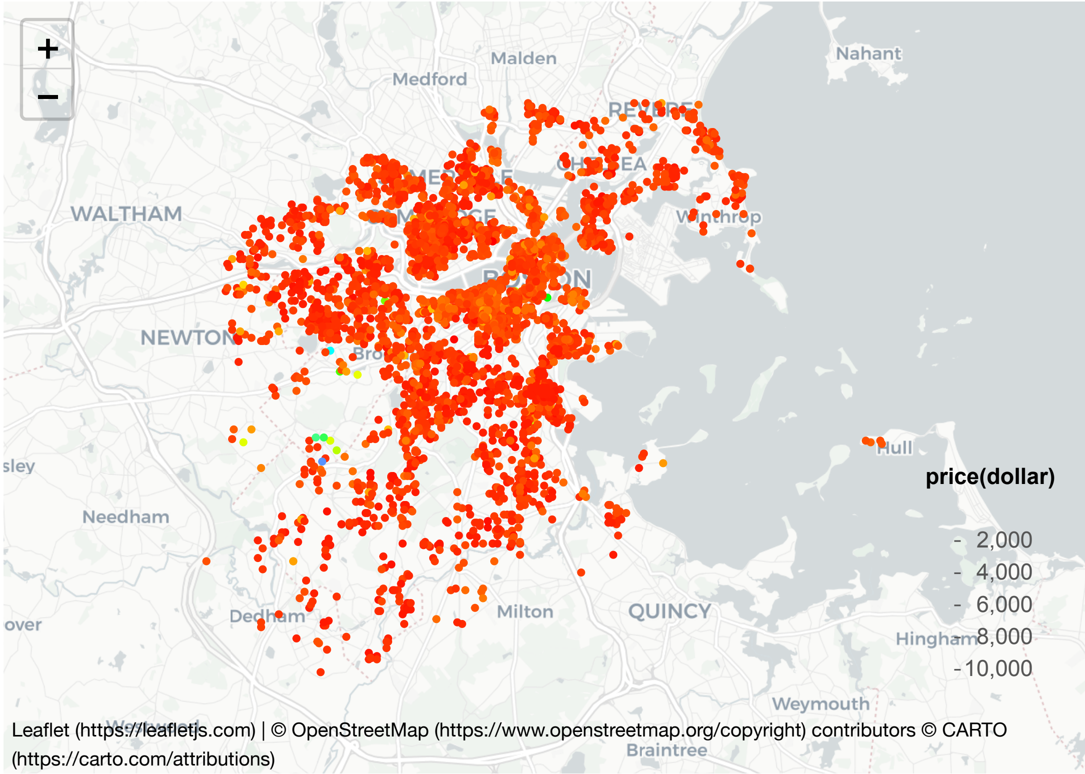
availability_365: The total number of days the listing is available for during the year

<Method>

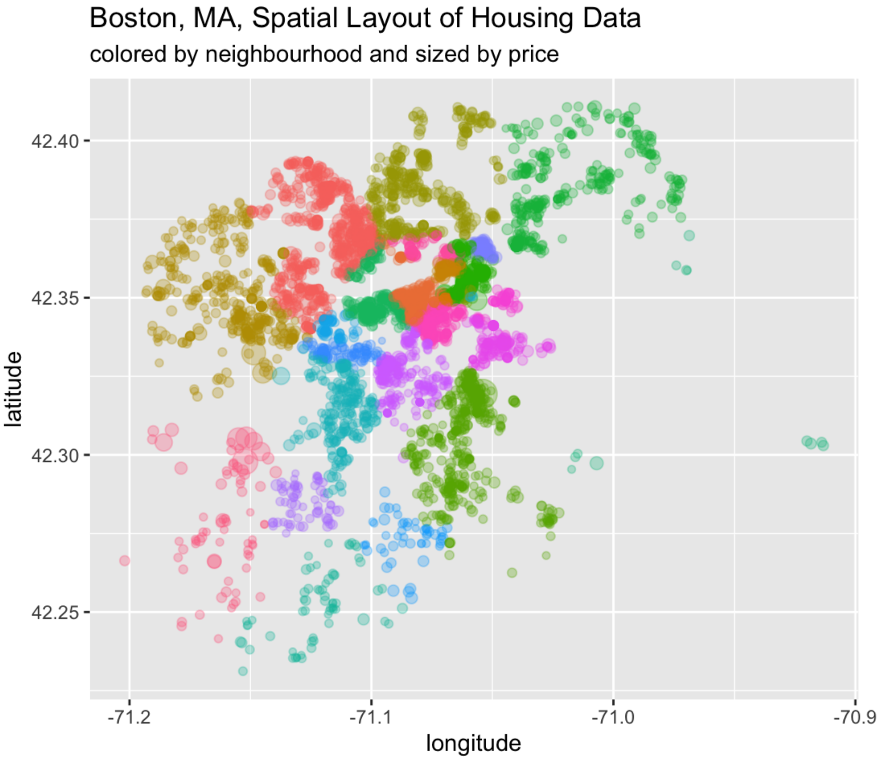
<Exploratory Data Analysis>

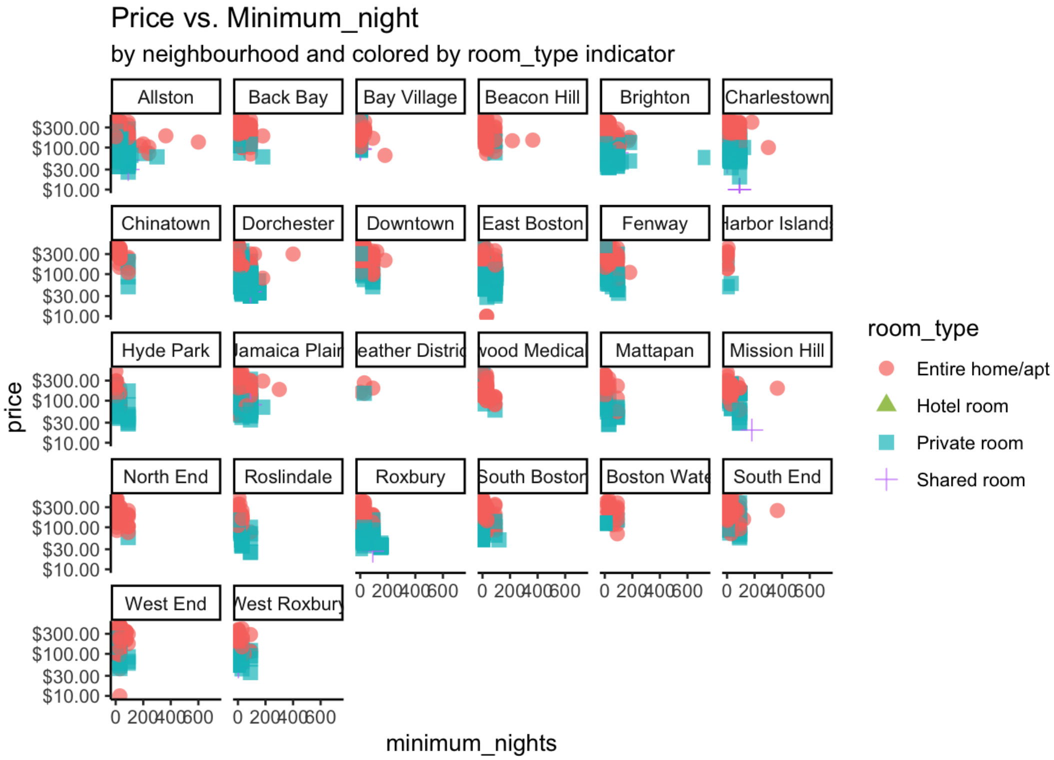
Figures 2 and 3 both display the relationship between price and geography. Figure 2 shows the detailed geographic location of the listings. Figure 3 shows more clearly the price situation in different neighborhoods. Figure 4 shows the relationship between listings’ prices and minimum nights in different neighborhoods and housing types. Figure 5 shows the relationship between listings’ prices and the number of reviews in different neighborhoods and housing types.

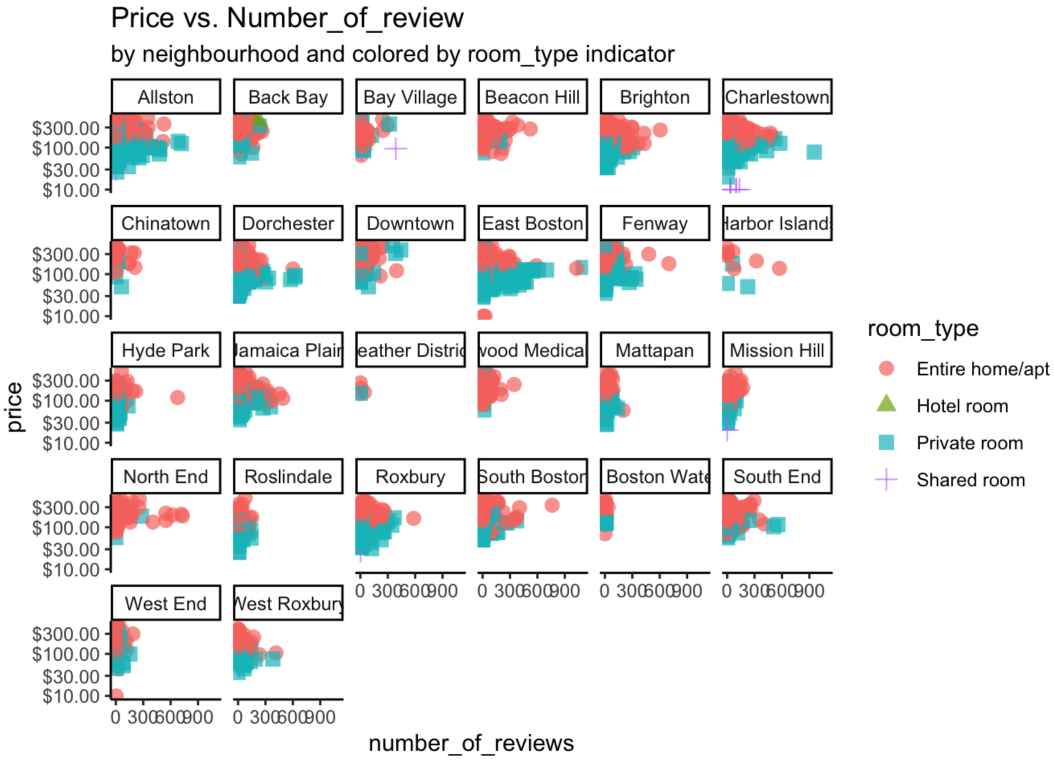




Leaflet (https://leafletjs.com) | © OpenStreetMap (https://www.openstreetmap.org/copyright) contributors © CARTO (https://carto.com/attributions)


Boston, MA, Spatial Layout of Housing Data
colored by neighbourhood and sized by price
latitude
42.40
42.35
42.30
42.25
-71.2
-71.1
-71.0
-70.9
longitude

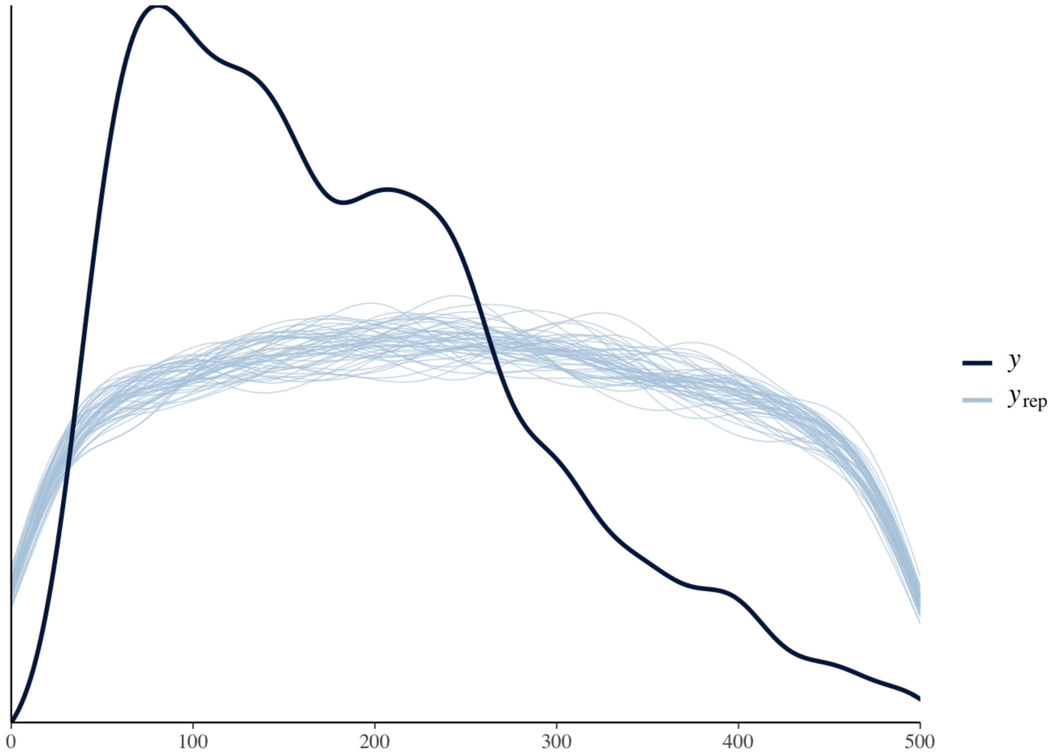

Price vs. Minimum_night
by neighbourhood and colored by room_type indicator
Allston
Back Bay
Bay Village
Beacon Hill
Brighton
Charlestown
Chinatown
Dorchester
Downtown
East Boston
Fenway
Harbor Islands
Hyde Park
Jamaica Plain
Leather District
Longwood Medical
Mattapan
Mission Hill
North End
Roslindale
Roxbury
South Boston
Boston Waterfront
South End
West End
West Roxbury
price
$300.00
$100.00
$30.00
$10.00
0 200 400 600
minimum_nights
room_type
Entire home/apt
Hotel room
Private room
Shared room

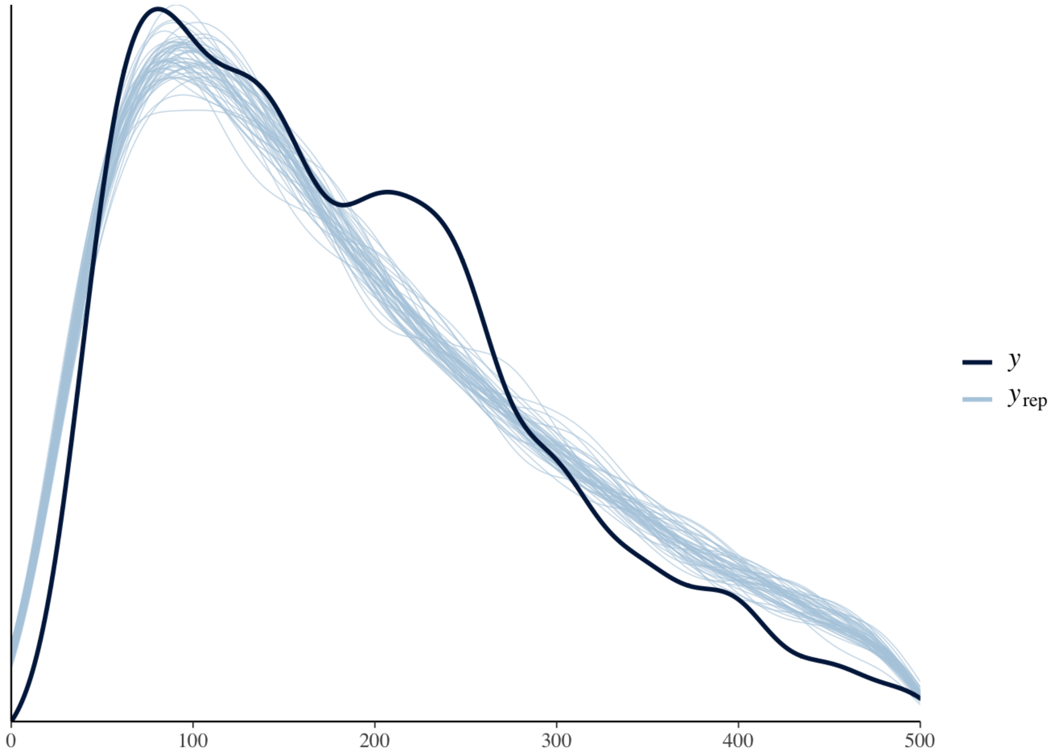


<Multilevel Bayesian Analysis>

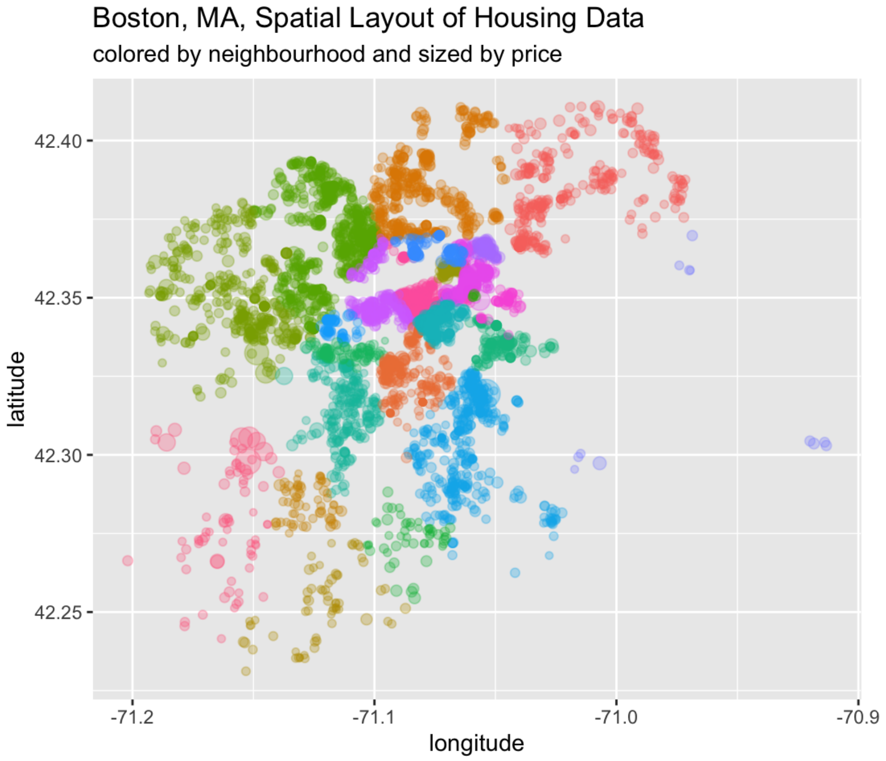
```
M1<-stan_glmer(price~latitude+longitude+factor(room_type)+minimum_nights+number_of_reviews+availability_365+(1|
neighbourhood),data=Boston_Airbnb_Selected)

M2<-stan_glmer(price~latitude+longitude+factor(room_type)+minimum_nights+number_of_reviews+availability_365+(1|
neighbourhood),family = neg_binomial_2,data=Boston_Airbnb_Selected)
```


0
100
200
300
400
500
y
y_{rep}


0
100
200
300
400
500
y
y_{rep}

```
## stan_glmer
##  family:       neg_binomial_2 [log]
##  formula:      price ~ latitude + longitude + factor(room_type) + minimum_nights +
##     number_of_reviews + availability_365 + (1 | neighbourhood)
##  observations: 5182
## ------
##                               Median MAD_SD
## (Intercept)                   -95.5   60.9
## latitude                        2.8    0.7
## longitude                       0.3    0.7
## factor(room_type)Hotel room     0.6    0.1
## factor(room_type)Private room  -0.8    0.0
## factor(room_type)Shared room   -1.1    0.1
## minimum_nights                  0.0    0.0
## number_of_reviews               0.0    0.0
## availability_365                0.0    0.0
##
## Auxiliary parameter(s):
##                       Median MAD_SD
## reciprocal_dispersion 2.9    0.1
##
## Error terms:
##  Groups        Name        Std.Dev.
##  neighbourhood (Intercept) 0.27
## Num. levels: neighbourhood 26
##
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```



<Result>

I first tried to fit the multilevel models and then used the posterior predictive checking model to check these two models. The plots show the model with a negative binomial argument worked better(plot 2).

I arrived at the following formula of fixed effect:

Price = -93.7+ 2.9 * latitude + 0.3* longitude + 0.6* Hotel room -0.8 *Private room -1.1* shared room

The model shows that latitude, longitude, and price are positively correlated. Hotel rooms will make room prices more expensive, while private and shared rooms will make room prices cheaper. Some neighborhoods positively correlate with the price(such as Back bay, Chinatown, downtown, Fenway, Harbor Islands, North End, South Boston Waterfront, and West Roxbury.

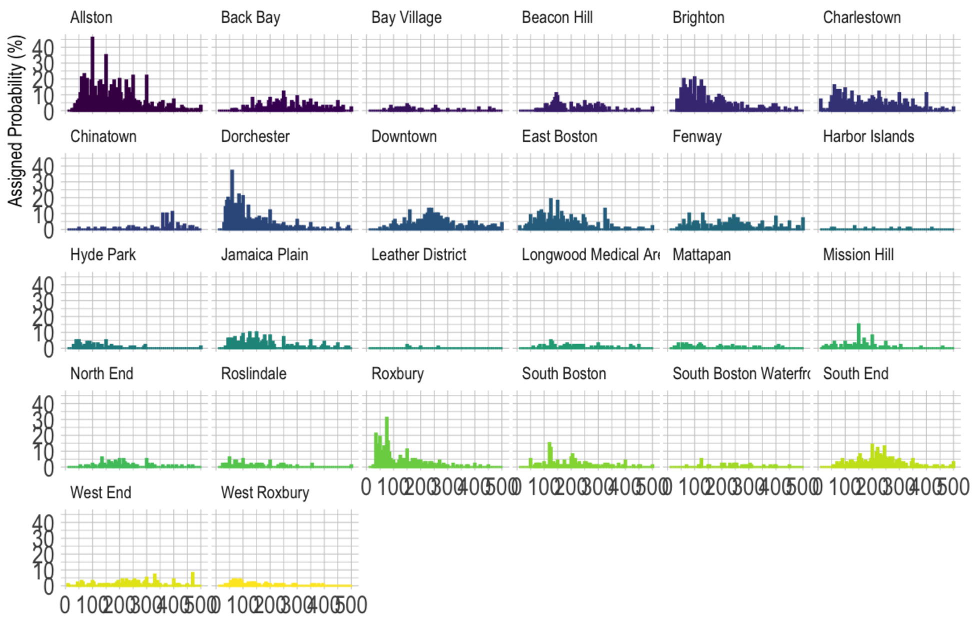
In the last plot, I color the neighborhood according to the ordering dictated by the estimated price intercepts (base_price). This way, the higher-priced areas will be more toward the red end of the color spectrum.

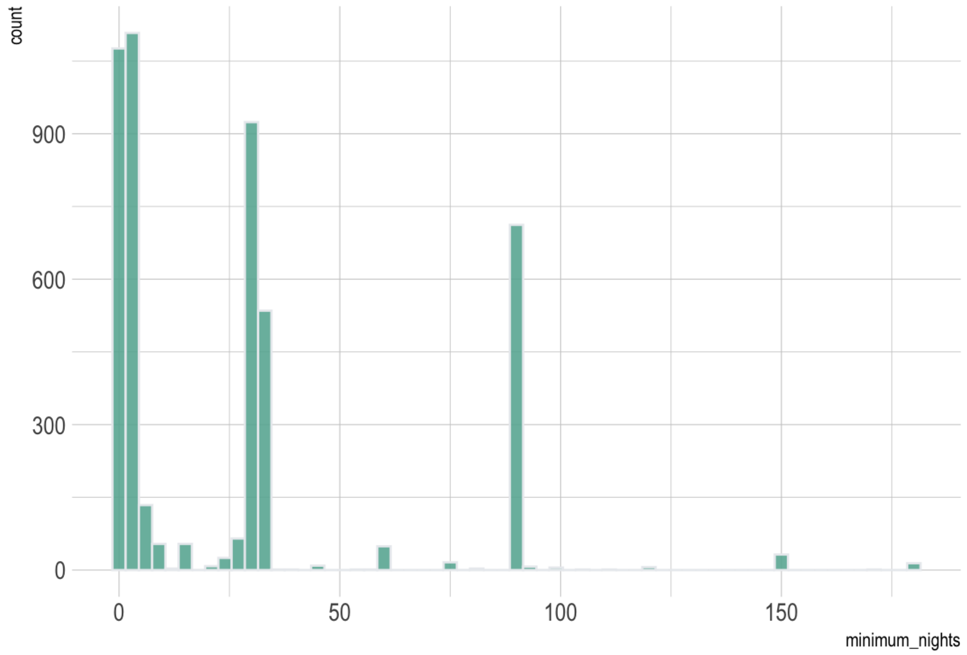
<Discussion>

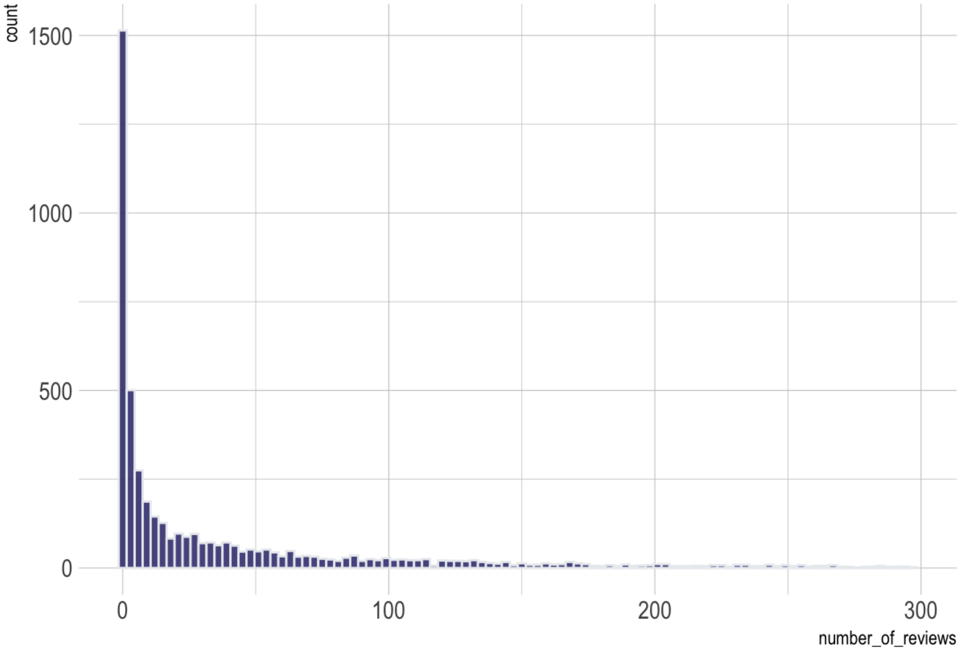
I found the relationship between price and geography, and room type through the multilevel model. But I didn’t find an association between price, minimum nights, and the number of reviews by the multilevel model. The previous EDA shows that the minimum nights and number of reviews are related to the price. Further research is needed on them.

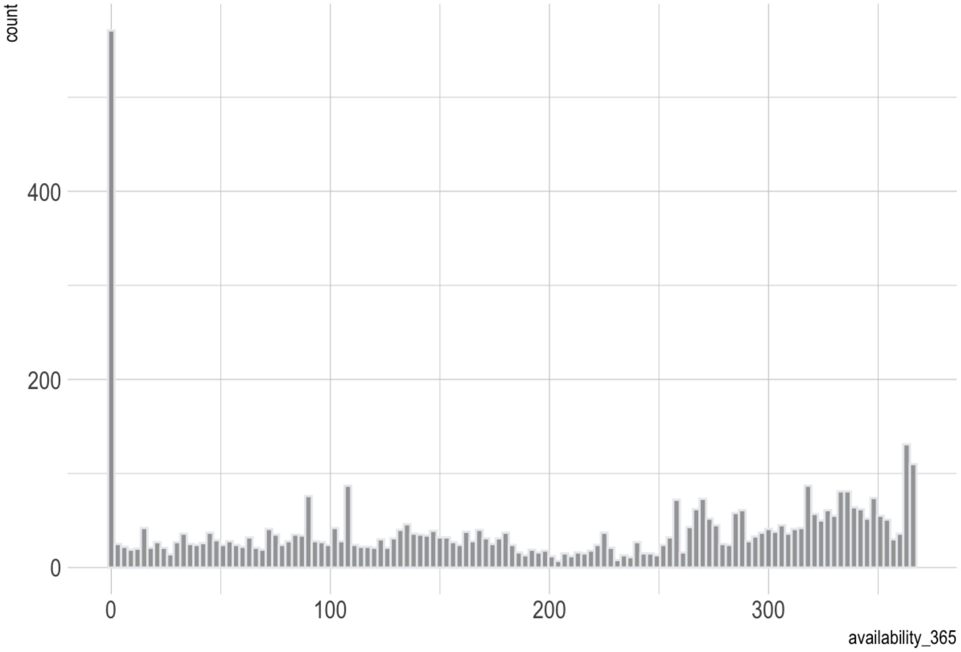
As seen from the posterior predictive checking model, my multilevel model does not fit perfectly, probably because of variables and data limitations. The fit of the price prediction model can be improved in the future.

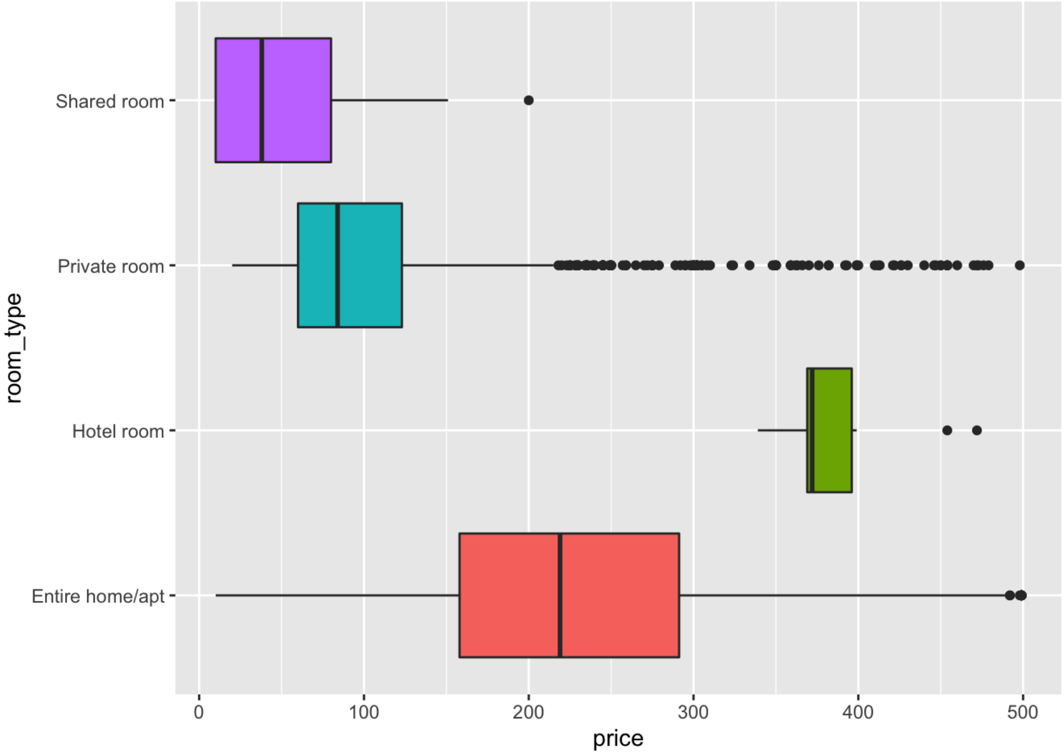
<Appendix>


Assigned Probability (%)
Allston
Back Bay
Bay Village
Beacon Hill
Brighton
Charlestown
Chinatown
Dorchester
Downtown
East Boston
Fenway
Harbor Islands
Hyde Park
Jamaica Plain
Leather District
Longwood Medical Ar
Mattapan
Mission Hill
North End
Roslindale
Roxbury
South Boston
South Boston Waterfr
South End
West End
West Roxbury
0
10
20
30
40
0
100
200
300
400
500


count
0
300
600
900
0
50
100
150
minimum_nights


count
1500
1000
500
0
0
100
200
300
number_of_reviews


count
400
200
0
0
100
200
300
availability_365


Shared room
Private room
Hotel room
Entire home/apt
room_type
0
100
200
300
400
500
price

```
##
## Model Info:
##  function:     stan_glmer
##  family:       neg_binomial_2 [log]
##  formula:      price ~ latitude + longitude + factor(room_type) + minimum_nights +
##     number_of_reviews + availability_365 + (1 | neighbourhood)
##  algorithm:    sampling
##  sample:       4000 (posterior sample size)
##  priors:       see help('prior_summary')
##  observations: 5182
##  groups:       neighbourhood (26)
##
## Estimates:
##                                                         mean   sd     10%
## (Intercept)                                           -93.2   60.9 -168.5
## latitude                                                2.8    0.7    1.9
## longitude                                               0.3    0.7   -0.6
## factor(room_type)Hotel room                             0.6    0.1    0.5
## factor(room_type)Private room                          -0.8    0.0   -0.8
## factor(room_type)Shared room                           -1.1    0.1   -1.2
## minimum_nights                                          0.0    0.0    0.0
## number_of_reviews                                       0.0    0.0    0.0
## availability_365                                        0.0    0.0    0.0
## b[(Intercept) neighbourhood:Allston]                   -0.1    0.1   -0.2
## b[(Intercept) neighbourhood:Back_Bay]                   0.3    0.1    0.2
## b[(Intercept) neighbourhood:Bay_Village]                0.0    0.1   -0.1
## b[(Intercept) neighbourhood:Beacon_Hill]               -0.2    0.1   -0.3
## b[(Intercept) neighbourhood:Brighton]                  -0.1    0.1   -0.2
## b[(Intercept) neighbourhood:Charlestown]               -0.2    0.1   -0.3
## b[(Intercept) neighbourhood:Chinatown]                  0.3    0.1    0.2
## b[(Intercept) neighbourhood:Dorchester]                 0.0    0.1   -0.1
## b[(Intercept) neighbourhood:Downtown]                   0.2    0.1    0.1
## b[(Intercept) neighbourhood:East_Boston]               -0.4    0.1   -0.5
## b[(Intercept) neighbourhood:Fenway]                     0.1    0.1    0.0
## b[(Intercept) neighbourhood:Harbor_Islands]             0.1    0.2   -0.1
## b[(Intercept) neighbourhood:Hyde_Park]                 -0.2    0.1   -0.3
## b[(Intercept) neighbourhood:Jamaica_Plain]              0.0    0.1   -0.1
## b[(Intercept) neighbourhood:Leather_District]          -0.1    0.2   -0.4
## b[(Intercept) neighbourhood:Longwood_Medical_Area]      0.0    0.1   -0.1
## b[(Intercept) neighbourhood:Mattapan]                  -0.1    0.1   -0.2
## b[(Intercept) neighbourhood:Mission_Hill]               0.0    0.1   -0.1
## b[(Intercept) neighbourhood:North_End]                  0.1    0.1    0.0
## b[(Intercept) neighbourhood:Roslindale]                -0.2    0.1   -0.4
## b[(Intercept) neighbourhood:Roxbury]                   -0.3    0.1   -0.4
## b[(Intercept) neighbourhood:South_Boston]               0.0    0.1   -0.1
## b[(Intercept) neighbourhood:South_Boston_Waterfront]    0.2    0.1    0.0
## b[(Intercept) neighbourhood:South_End]                  0.0    0.1   -0.1
## b[(Intercept) neighbourhood:West_End]                   0.0    0.1   -0.1
## b[(Intercept) neighbourhood:West_Roxbury]               0.9    0.1    0.7
## reciprocal_dispersion                                   2.9    0.1    2.9
## Sigma[neighbourhood:(Intercept),(Intercept)]            0.1    0.0    0.0
##                                                         50%    90%
## (Intercept)                                           -95.5  -13.1
## latitude                                                2.8    3.7
## longitude                                               0.3    1.2
## factor(room_type)Hotel room                             0.6    0.8
## factor(room_type)Private room                          -0.8   -0.8
## factor(room_type)Shared room                           -1.1   -0.9
## minimum_nights                                          0.0    0.0
## number_of_reviews                                       0.0    0.0
## availability_365                                        0.0    0.0
## b[(Intercept) neighbourhood:Allston]                   -0.1    0.0
## b[(Intercept) neighbourhood:Back_Bay]                   0.3    0.4
## b[(Intercept) neighbourhood:Bay_Village]                0.0    0.1
## b[(Intercept) neighbourhood:Beacon_Hill]               -0.2   -0.1
## b[(Intercept) neighbourhood:Brighton]                  -0.1    0.0
## b[(Intercept) neighbourhood:Charlestown]               -0.2   -0.1
## b[(Intercept) neighbourhood:Chinatown]                  0.3    0.4
## b[(Intercept) neighbourhood:Dorchester]                 0.0    0.1
## b[(Intercept) neighbourhood:Downtown]                   0.2    0.2
## b[(Intercept) neighbourhood:East_Boston]               -0.4   -0.3
## b[(Intercept) neighbourhood:Fenway]                     0.1    0.2
## b[(Intercept) neighbourhood:Harbor_Islands]             0.1    0.3
## b[(Intercept) neighbourhood:Hyde_Park]                 -0.2   -0.1
## b[(Intercept) neighbourhood:Jamaica_Plain]              0.0    0.1
## b[(Intercept) neighbourhood:Leather_District]          -0.1    0.1
## b[(Intercept) neighbourhood:Longwood_Medical_Area]      0.0    0.1
## b[(Intercept) neighbourhood:Mattapan]                  -0.1    0.1
## b[(Intercept) neighbourhood:Mission_Hill]               0.0    0.1
## b[(Intercept) neighbourhood:North_End]                  0.1    0.2
## b[(Intercept) neighbourhood:Roslindale]                -0.2   -0.1
## b[(Intercept) neighbourhood:Roxbury]                   -0.3   -0.3
## b[(Intercept) neighbourhood:South_Boston]               0.0    0.1
## b[(Intercept) neighbourhood:South_Boston_Waterfront]    0.2    0.3
## b[(Intercept) neighbourhood:South_End]                  0.0    0.1
## b[(Intercept) neighbourhood:West_End]                   0.0    0.1
## b[(Intercept) neighbourhood:West_Roxbury]               0.9    1.0
## reciprocal_dispersion                                   2.9    3.0
## Sigma[neighbourhood:(Intercept),(Intercept)]            0.1    0.1
## 
## Fit Diagnostics:
##            mean   sd    10%   50%   90%
## mean_PPD 230.9    3.0 227.0 230.9 234.6
## 
## The mean_ppd is the sample average posterior predictive distribution of the outcome variable (for details se
e help('summary.stanreg')).
## 
## MCMC diagnostics
##                                                      mcse Rhat n_eff
## (Intercept)                                          1.4  1.0   1859
## latitude                                             0.0  1.0   1899
## longitude                                            0.0  1.0   1765
## factor(room_type)Hotel room                          0.0  1.0  11373
## factor(room_type)Private room                        0.0  1.0   7871
## factor(room_type)Shared room                         0.0  1.0   8413
## minimum_nights                                       0.0  1.0   9503
## number_of_reviews                                    0.0  1.0   7588
## availability_365                                     0.0  1.0   8057
## b[(Intercept) neighbourhood:Allston]                 0.0  1.0    696
## b[(Intercept) neighbourhood:Back_Bay]                0.0  1.0    617
## b[(Intercept) neighbourhood:Bay_Village]             0.0  1.0   1237
## b[(Intercept) neighbourhood:Beacon_Hill]             0.0  1.0    795
## b[(Intercept) neighbourhood:Brighton]                0.0  1.0    899
## b[(Intercept) neighbourhood:Charlestown]             0.0  1.0    702
## b[(Intercept) neighbourhood:Chinatown]               0.0  1.0    987
## b[(Intercept) neighbourhood:Dorchester]              0.0  1.0    669
## b[(Intercept) neighbourhood:Downtown]                0.0  1.0    655
## b[(Intercept) neighbourhood:East_Boston]             0.0  1.0    829
## b[(Intercept) neighbourhood:Fenway]                  0.0  1.0    662
## b[(Intercept) neighbourhood:Harbor_Islands]          0.0  1.0   2639
## b[(Intercept) neighbourhood:Hyde_Park]               0.0  1.0   1513
## b[(Intercept) neighbourhood:Jamaica_Plain]           0.0  1.0    764
## b[(Intercept) neighbourhood:Leather_District]        0.0  1.0   3835
## b[(Intercept) neighbourhood:Longwood_Medical_Area]   0.0  1.0   1207
## b[(Intercept) neighbourhood:Mattapan]                0.0  1.0   1457
## b[(Intercept) neighbourhood:Mission_Hill]            0.0  1.0   1083
## b[(Intercept) neighbourhood:North_End]               0.0  1.0   1042
## b[(Intercept) neighbourhood:Roslindale]              0.0  1.0   1453
## b[(Intercept) neighbourhood:Roxbury]                 0.0  1.0    646
## b[(Intercept) neighbourhood:South_Boston]            0.0  1.0    839
## b[(Intercept) neighbourhood:South_Boston_Waterfront] 0.0  1.0   1340
## b[(Intercept) neighbourhood:South_End]               0.0  1.0    566
## b[(Intercept) neighbourhood:West_End]                0.0  1.0    906
## b[(Intercept) neighbourhood:West_Roxbury]            0.0  1.0   1322
## reciprocal_dispersion                                0.0  1.0   4978
## Sigma[neighbourhood:(Intercept),(Intercept)]         0.0  1.0    779
## mean_PPD                                             0.0  1.0   3768
## log-posterior                                        0.2  1.0    870
## 
## For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample size, a
nd Rhat is the potential scale reduction factor on split chains (at convergence Rhat=1).
```