Is renting my house a good investment?

A spatial analysis of AirBnb's in New York.

Carol Wang Xiaojun Zheng Alessandro Zito

Case Study 2 - Stat 723

February 4, 2020

Overview

1. Introduction
2. Data
3. EDA
4. Methods
   - Voronoi Tassellation
   - Text Analysis
   - Conditional Autoregressive Model

Introduction

AirBnb is an on-line platform for "sharing economy" where house-owners rent their places for short-term stays. Having access to prices, locations and other factors allows to detect factors that are associated to success/high price of the house.

We have access to **48,895 listings** distributed across the 5 boroughs on New York (Manhattan, Queens, Staten Island, The Bronx and Brooklyn) and collected all in 2019.

We focus on answering the following questions

- What are the main determinants of price/popularity?
- Are boroughs and neighbourhoods heterogeneous in terms of price?
- Are boroughs and neighbourhoods heterogeneous in terms of type of house?
- What are the "most successful" words for a listing name?

Data

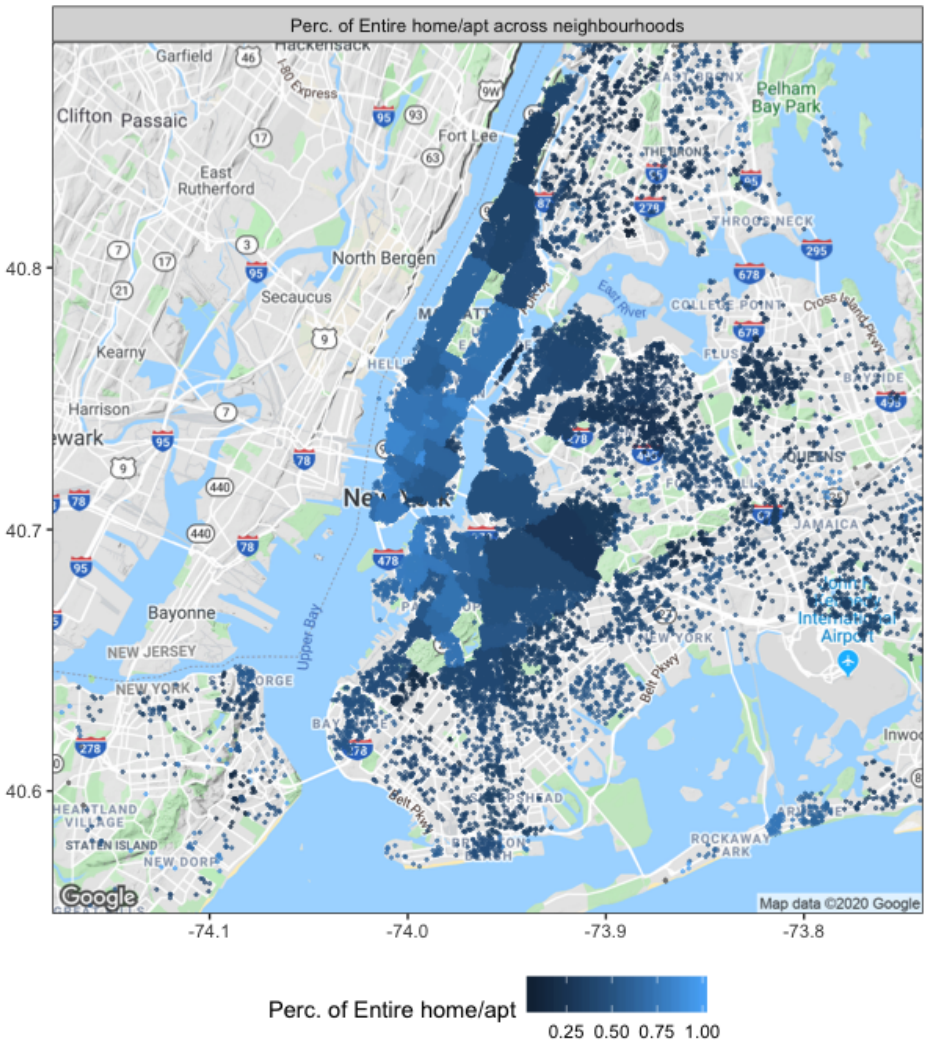
Our data include information on house price, location (latitude and longitude), boroughs and neighbourhoods, type of room, (total and average per month) number of reviews, date of last review, minimum nights of stay required and days available.

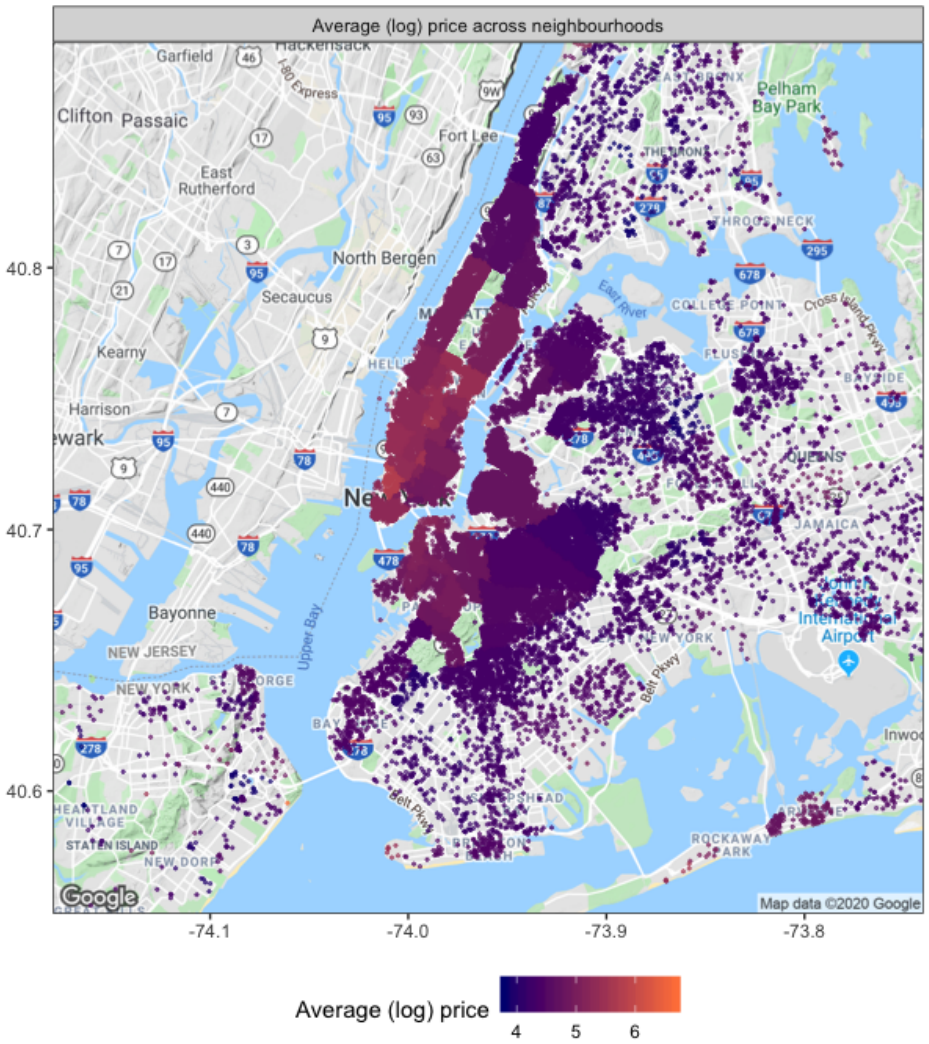
Data cleaning:

- Consider only listings with date of last review within one year of the latest last review (from June 2016 to June 2017).
- Select houses with price $\in [10, 2000]$ (close to true AirBnb distribution)
- Drop houses with availability less than 10
- Group availability into three groups: **Low** (less than 2 months), **Medium** (between 2 and 6 months), **High**(more than 6 months)
- Introduce months of activity = number of reviews / reviews per month

$\Longrightarrow$ **Reduced sample size = 24,165**

EDA - Neighbourhood heterogeneity





EDA - Neighbourhood heterogeneity

Heterogeneity across boroughs and neighbourhoods can be tested by via a (Monte Carlo simulated) χ^2-test.

- χ^2-test between boroughs and room type: p-value ≈ 0.000
- χ^2-test between neighbourhoods and room type: p-value $=$ 0.0004998 (based on Hope(1968) Monte Carlo test)

Both tests indicate that the distribution of room type is not independent across both neighbourhoods and boroughs

How should we account for this heterogeneity?

Methods (I) - Adjacency matrix construction

The heterogeneity across neighbourhoods requires to take into account for a spacial auto-correlation in our model.
$\Longrightarrow$ Need to construct an **adjacency matrix**.

$$W = (w_{ij})_{i,j=1}^{p}$$

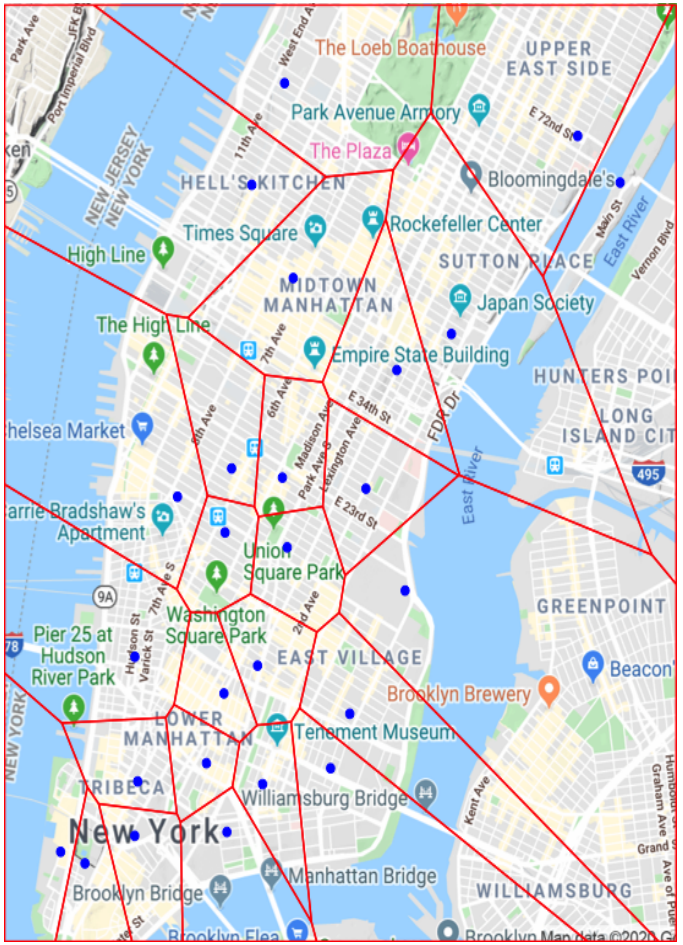
with p is n. of neighbourhoods, and $w_{i,j} = 1$ if i and j are contiguous.

Technique: **Voronoi Tassellation**

For $k = 1, \ldots, 1000$:

1. Sample 1 point for each neighbourhood
2. Get W_k with Voronoi Tassellation

If $\frac{w_{ij,k}}{1000} \geq 0.8$, then the neighbourhoods are contiguous.



Methods (II) - Text analysis

Methods (II) - Text Analysis

- Select the unique words from names of Airbnb
- Choose the Words whose frequency is higher than 8 as predictors
- log (Price) $\sim 1 +$ Words $+ \ldots$ (variables from the original dataset)
- log (Reviews per month) $\sim 1 +$ Words $+ \ldots$ (variables from the original dataset)

Important words:

- Price: luxury, loft, townhouse, triplex, stunning, events, spectacular, '5th' ...
- Popularity: private, LGA, JFK, parking, metro, min, superhost, hotellike, close ...

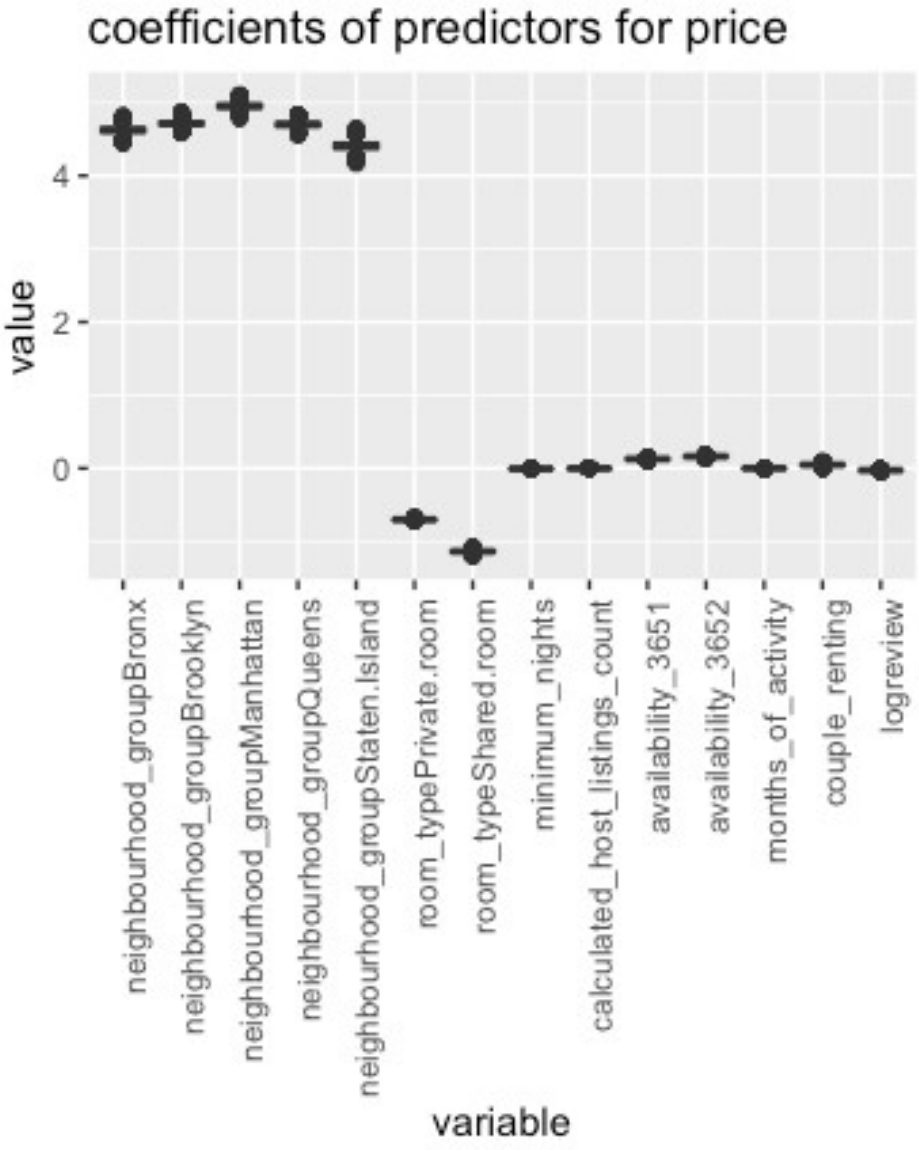
Methods (III) - CAR Model

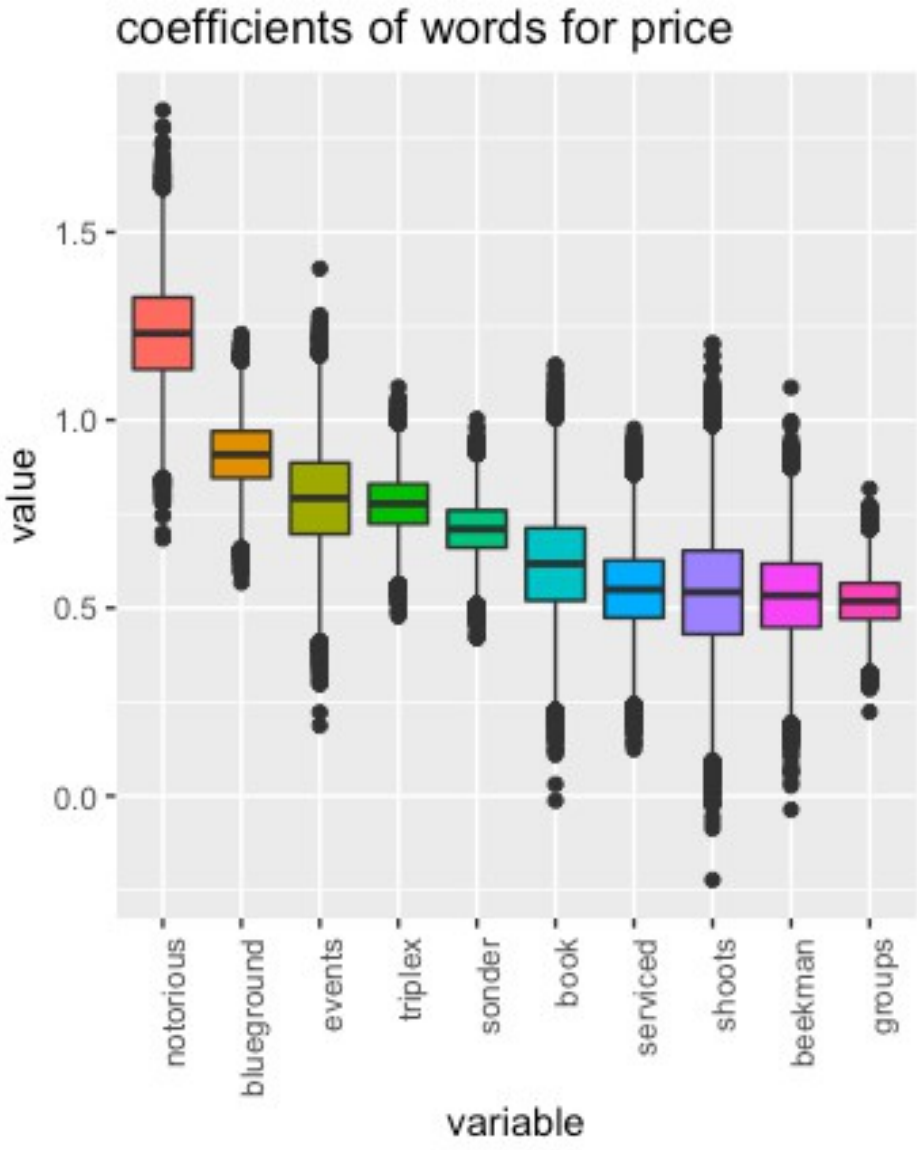
- Two-level areal data model with conditional autoregressive prior
- K non-overlapping areal units, each with m_k individuals

$$
\begin{aligned}
Y_{kj} &\sim N(x_{kj}^T\beta + \psi_{kj}, \nu^2), k = 1, \cdots, K, j = 1, \cdots, m_k, \\
\beta &\sim N(\mu_\beta, \Sigma_\beta), \\
\psi_{kj} &= \phi_k + \zeta_{\lambda(k,j)}, \\
\phi_k|\phi_{-k} &\sim N\left(\frac{\rho\Sigma_j w_{kj}\phi_j}{\rho\Sigma_j w_{kj}\phi_j + 1 - \rho}, \frac{\tau^2}{\rho\Sigma_j w_{kj}\phi_j + 1 - \rho}\right),
\end{aligned}
$$

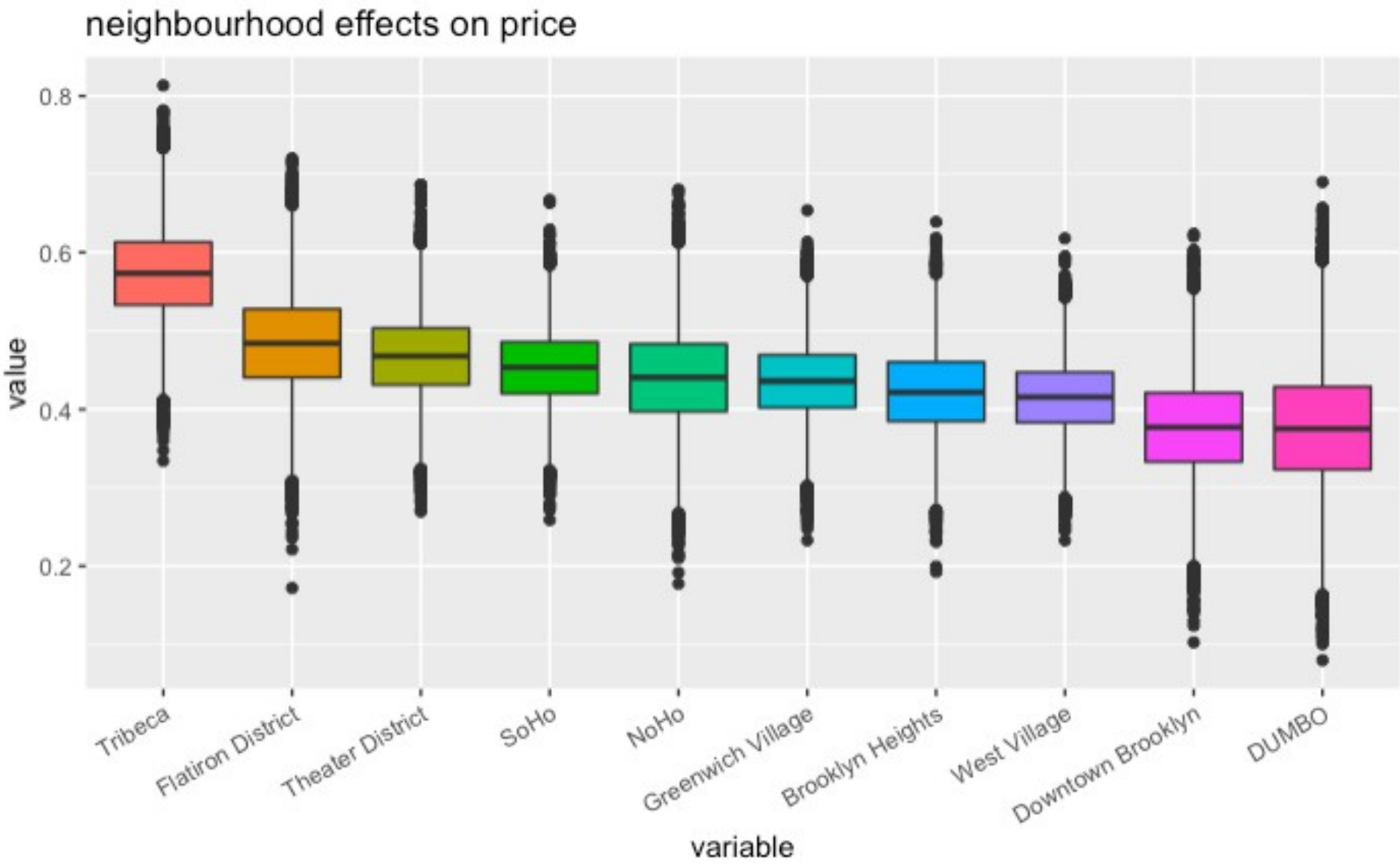
where ϕ_k is the effect of neighbourhood k.

Results

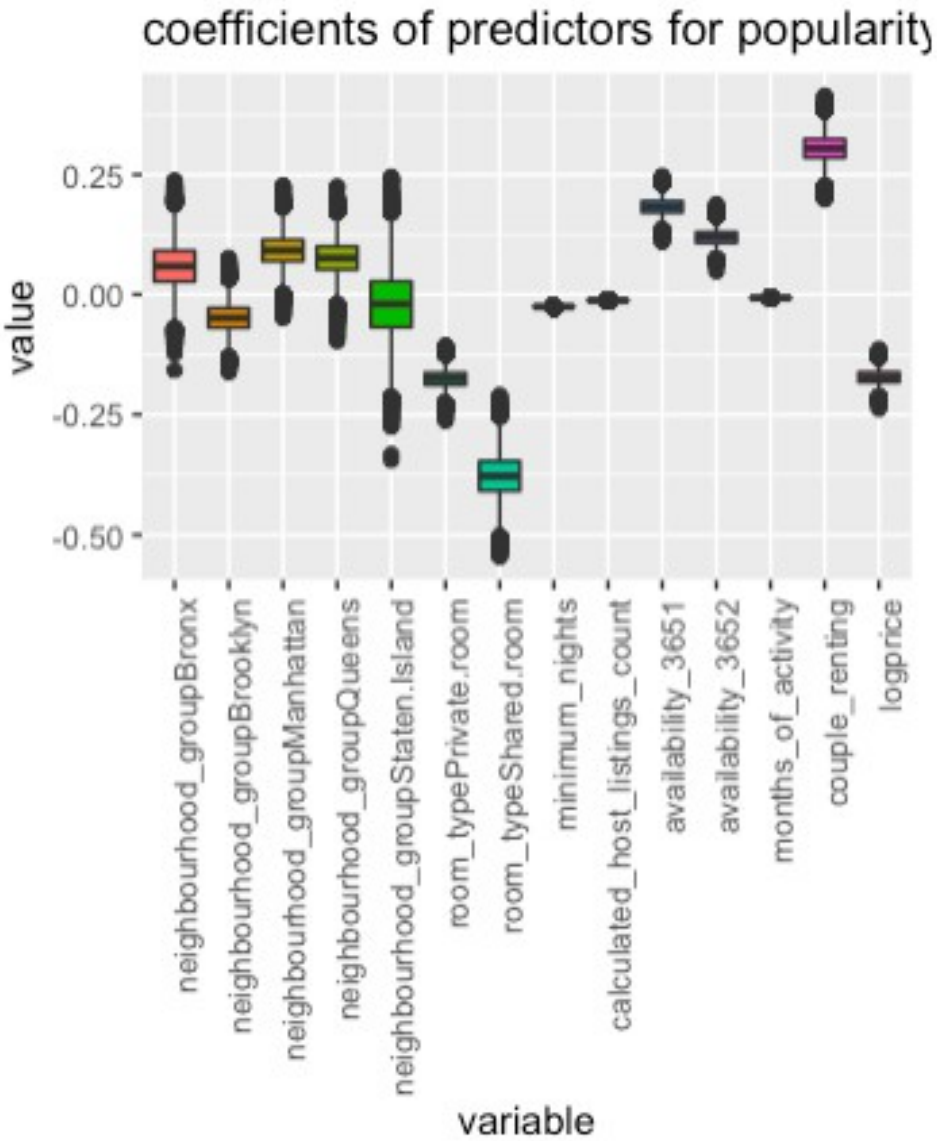

coefficients of predictors for price
value
4
2
0
neighbourhood_groupBronx
neighbourhood_groupBrooklyn
neighbourhood_groupManhattan
neighbourhood_groupQueens
neighbourhood_groupStaten.Island
room_typePrivate.room
room_typeShared.room
minimum_nights
calculated_host_listings_count
availability_3651
availability_3652
months_of_activity
couple_renting
logreview
variable

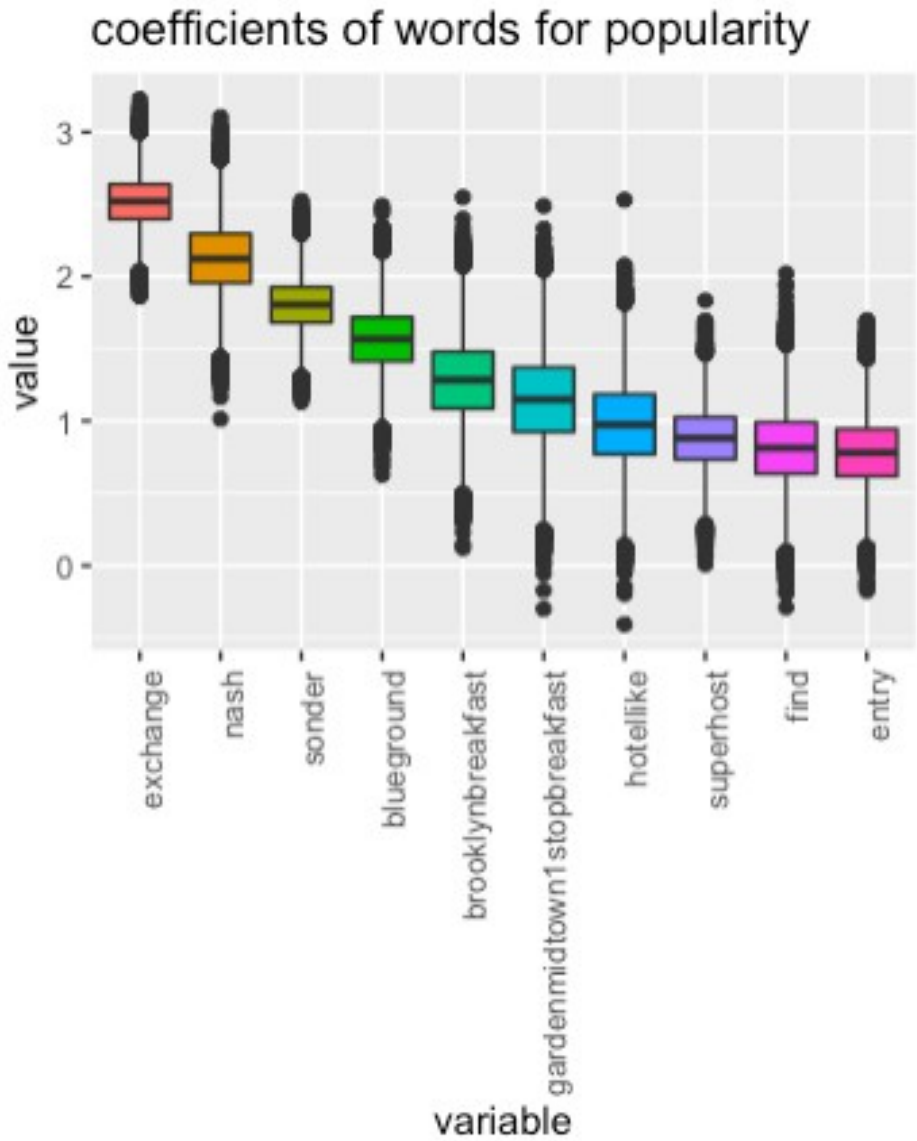

coefficients of words for price
value
1.5
1.0
0.5
0.0
notorious
blueground
events
triplex
sonder
book
serviced
shoots
beekman
groups
variable

Results


neighbourhood effects on price
value
0.8
0.6
0.4
0.2
Tribeca
Flatiron District
Theater District
SoHo
NoHo
Greenwich Village
Brooklyn Heights
West Village
Downtown Brooklyn
DUMBO
variable

Results


coefficients of predictors for popularity
value
0.25
0.00
-0.25
-0.50
neighbourhood_groupBronx
neighbourhood_groupBrooklyn
neighbourhood_groupManhattan
neighbourhood_groupQueens
neighbourhood_groupStaten.Island
room_typePrivate.room
room_typeShared.room
minimum_nights
calculated_host_listings_count
availability_3651
availability_3652
months_of_activity
couple_renting
logprice
variable


coefficients of words for popularity
value
3
2
1
0
exchange
nash
sonder
blueground
brooklynbreakfast
gardenmidtown1stopbreakfast
hotellike
superhost
find
entry
variable

Results

- Both Price & Popularity:
  - **Rental Agency**: notorious, sonder, blueground, exchange
- More Price
  - **Type of house**: triplex, bth, townhouse, floors, villa, townhome, hipster
  - **House Rules**: events, shoots, (no) service (fee)
- More popularity
  - **Demand**: (private) entry, (easy/self) check-in, laguardia (airport), wall (street), solo (traveler), overnight (bed/place)

Our AirBnb name: The Notorious B.N.B. {Triplex townhome in Manhattan for Shoots and Events with private entry and self check-in}