

Case Study 2-Group 4

Frances Hung, Yunran Chen, Keru Wu

Department of Statistical Science, Duke University

01/30/2019

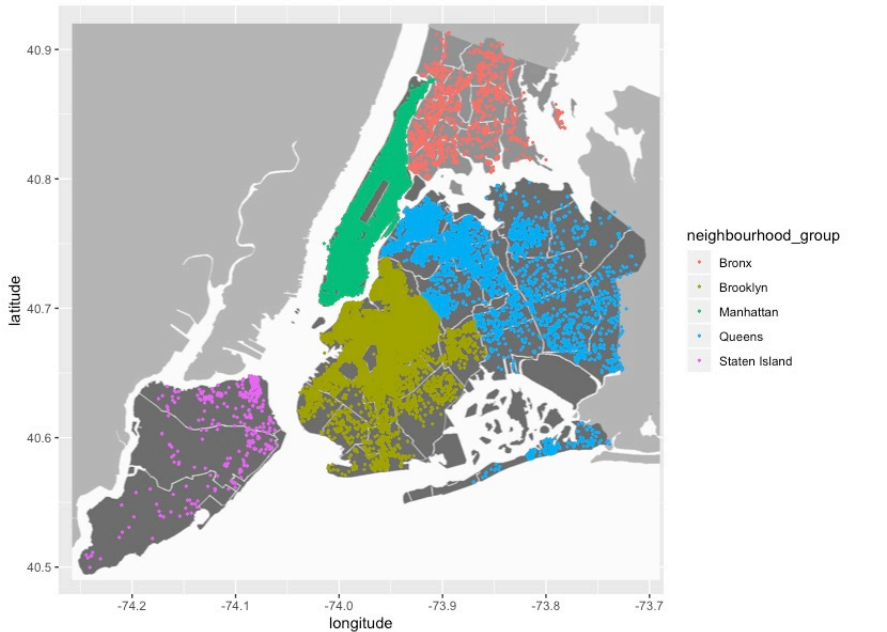
Introduction

- ▶ Data: 2019 Airbnb listings in NYC, 48895 observations.
- ▶ Goal: Identify discernible and interesting patterns among the listings in NYC.

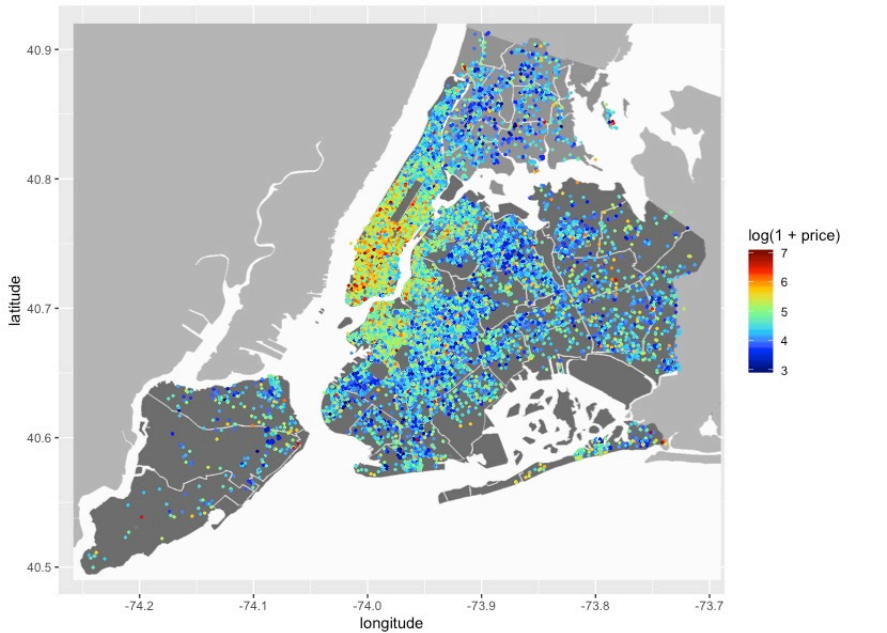
Data Preprocessing

- ▶ Delete id, host_name and last_review.
- ▶ Delete 11 listings with price 0.
- ▶ Missing data: 10052 in reviews_per_month, impute with 0.

EDA-borough

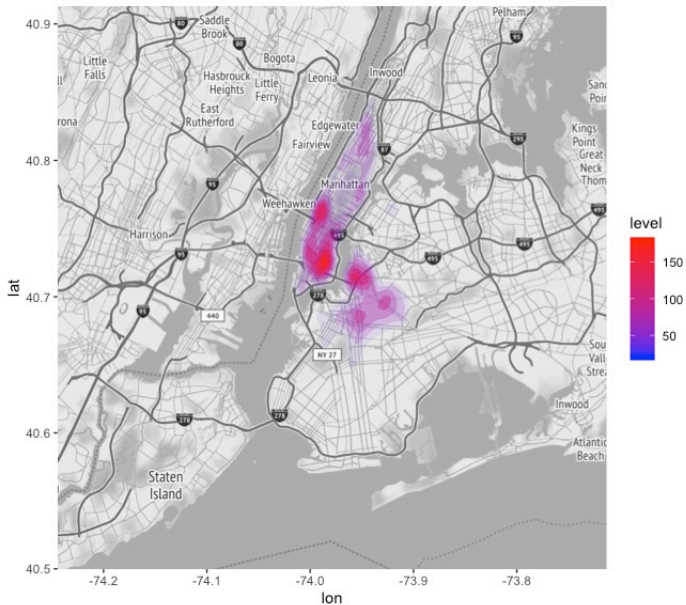


EDA-price



EDA-Traffic

Number of listings: KDE



Model

► Multilevel Conditional Autoregressive (CAR) Model

$$Y_{kj}|\mu_{kj} \sim f(y_{kj}|\mu_{kj}, \nu^2), \quad k = \text{neighbourhood} = 1, \dots, K \\ j = \text{listings} = 1, \dots, m_k$$

$$g(\mu_{kj}) = \mathbf{x}_{kj}^T \beta + \psi_{kj}$$

$$\psi_{kj} = \phi_k + \zeta_{kj}$$

► Priors

$$\beta \sim N(\mu_\beta, \Sigma_\beta)$$

$$\phi_k | \phi_{-k} \sim N\left(\frac{\rho \sum_{l=1}^K w_{kl} \phi_l}{\rho \sum_{l=1}^K w_{kl} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{l=1}^K w_{kl} + 1 - \rho}\right)$$

- w_{kl} denotes whether neighborhood k and l are adjacent.
- ρ denotes spatial dependence.

Model

- Priors (Cont'd)

$$\zeta_{kj} \sim N(0, \sigma^2)$$

$$\tau^2, \sigma^2 \sim \text{Inv-Gamma}(a, b)$$

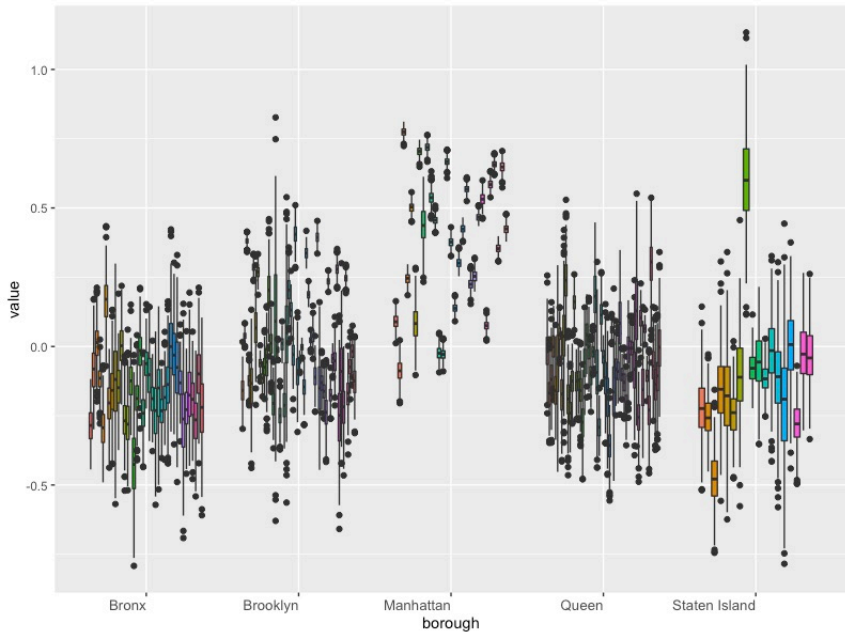
$$\rho \sim \text{Uniform}(0,1)$$

- x_{kj} include room_type, neighbourhood_group, availability_365, $\log(1+\text{reviews_per_month})$, minimum_nights, etc.
- $\psi_{kj} = \phi_k + \zeta_{kj}$ includes both spatial information and individual random effect.

Further process the data

- ▶ No data for exactly 217 neighbourhoods.
- ▶ Relocate neighbourhoods according to formal NYC shapefile data (195 neighbourhoods).
- ▶ 191 neighbourhoods have airbnb listings.
- ▶ Obtain adjacency matrix $W = (w_{kl})$

Neighbourhood Effect on log(price)



Discussion

- ▶ Include last_review: spatial temporal model.
- ▶ Nonlinear model: spline regression for x_{kj} .
- ▶ More spatial information: longitude & latitude