

# Airbnb Project

*Helen Zhou*

*11/26/2017*

```
## Loading required package: gridExtra  
## Loading required package: Matrix
```

## **Midterm Project\_Airbnb**

This project is designed to display data analysis based on Inside Airbnb data set from <http://insideairbnb.com/get-the-data.html>. Datas contain two cities: Boston and New York City. For each city, I use property listing data.

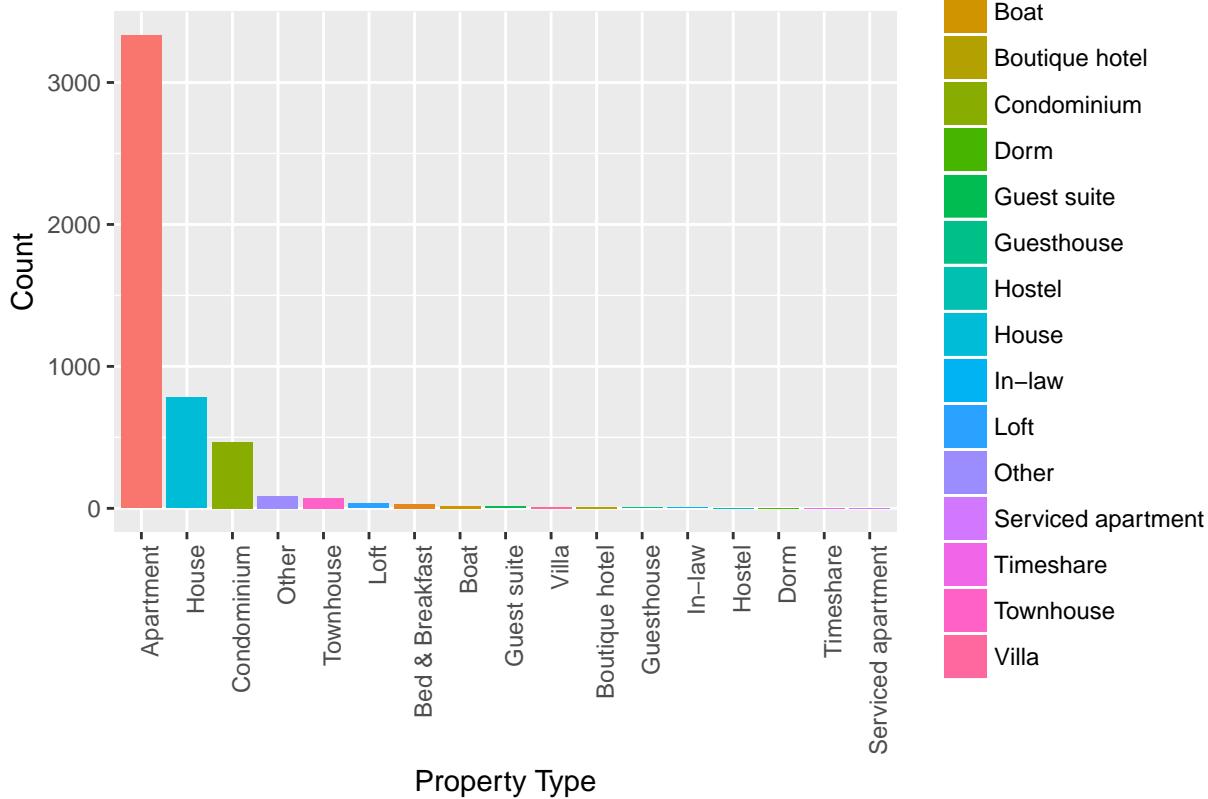
The model I built is mainly focused on the price prediction.

The major procedure can be summarized as below: 1. Data Reading and Preparation 2. EDA 4. Multi-level linear model 5. Reference and appendix

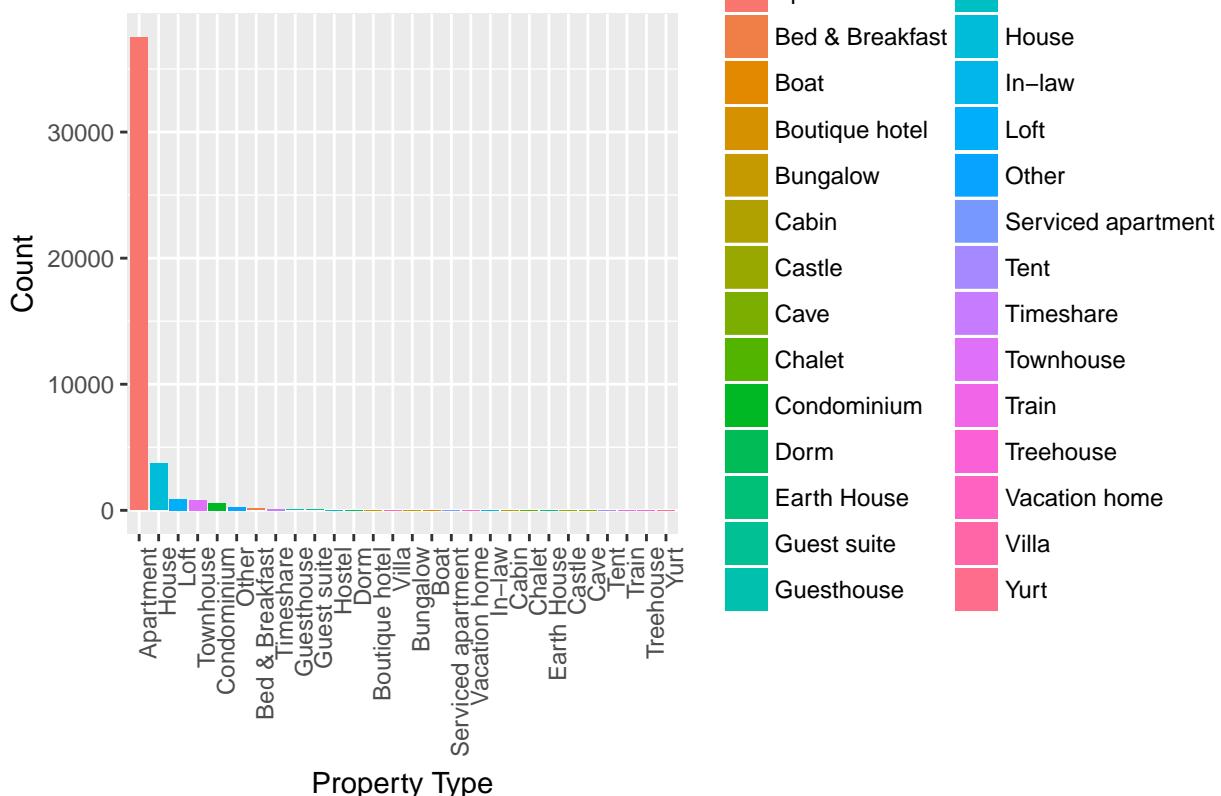
## **EDA and second round data cleaning**

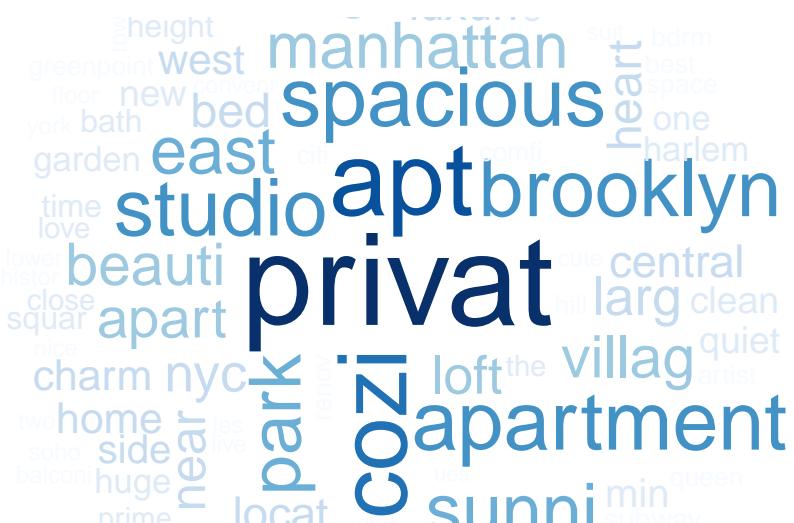
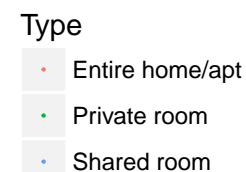
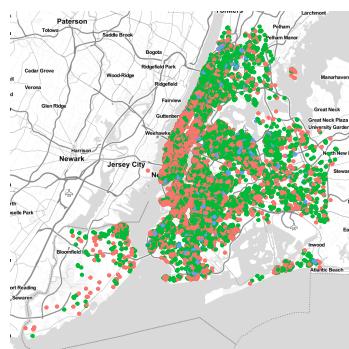
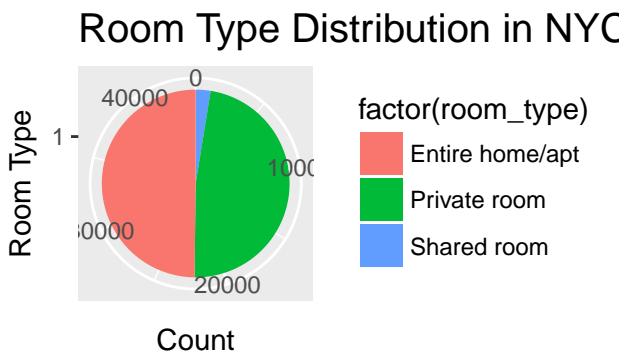
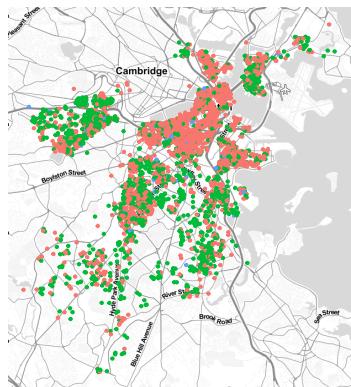
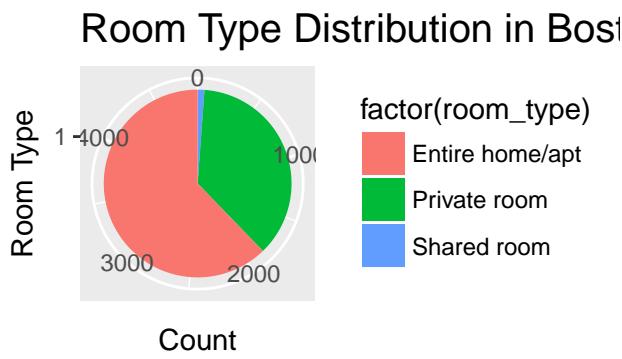
To find out people's experience with Airbnb in one city, we first need to know what are the options there

## Property Type Distribution in Boston



## Property Type Distribution in NYC





```
# distribution of price - BOS
qplot(bos_list$price,geom="histogram",binwidth = 0.5, main = "Histogram for Price in Boston", xlab = "Price")
p + coord_polar(theta = "y")
# distribution of price - NYC
```

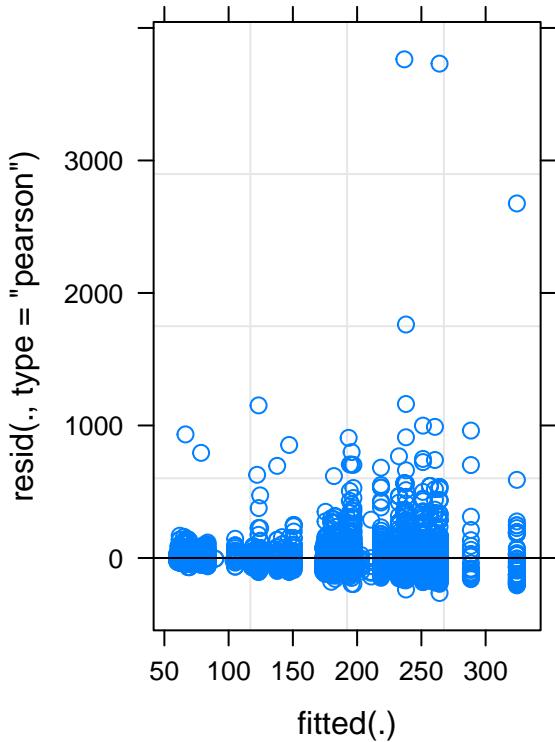
## Model and model visualization

I fit a multi-level model for each city.

$$y_i^{\text{price}} = \mu + \gamma_1^{\text{roomtype}} + \delta^{\text{neighbourhood}} + e_i$$

```
# Multi-level model
# Boston
m1 <- lmer(formula = price ~ 1 + (1 | room_type) + (1 | neighbourhood), data = bos_list_sum)
# New York
m2 <- lmer(formula = price ~ 1 + (1 | room_type) + (1 | neighbourhood), data = nyc_list_sum)
```

**residual plot Boston**



**residual plot NYC**

