**2. Data**

The dataset we use in this report contains price and relative features of 36,000 Airbnb listings in 6 major U.S. cities. This dataset contains all the information that a guest can find on the Airbnb platform, including the geographical information, properties, amenities, booking policy, reviews and host profile. Details are listed in the Table 1.

Table 1 Variables in the Airbnb dataset

|  |  |
| --- | --- |
| Geographical information | City  Neighborhood  Longitude and Latitude  Zip code |
| Properties | Accommodates  Number of beds  Number of bedrooms  Number of bathrooms  Property type  Bed type |
| Amenities | Amenities listed by host |
| Booking policy | Cancellation policy  Cleaning fee (logical)  Instant bookable |
| Reviews | First review date  Last review date  Number of reviews  Review score rating |
| Host profile | Profile picture (logical)  Identity verified (logical)  Host since date  Response rate |
| Data Source: AAE 722 |  |

**2.1 Data Exploration**

The six cites in the Airbnb data are Boston, Chicago, Washington DC, Los Angeles, New York City and San Francisco. Figure 1 shows that the price distribution are similar in those cities. There is no significant difference or pattern in prices between the West Coast, the East Coast and the Central Region.

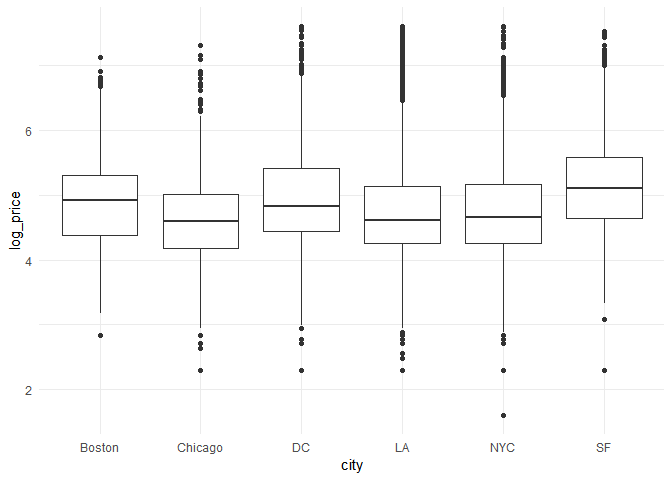
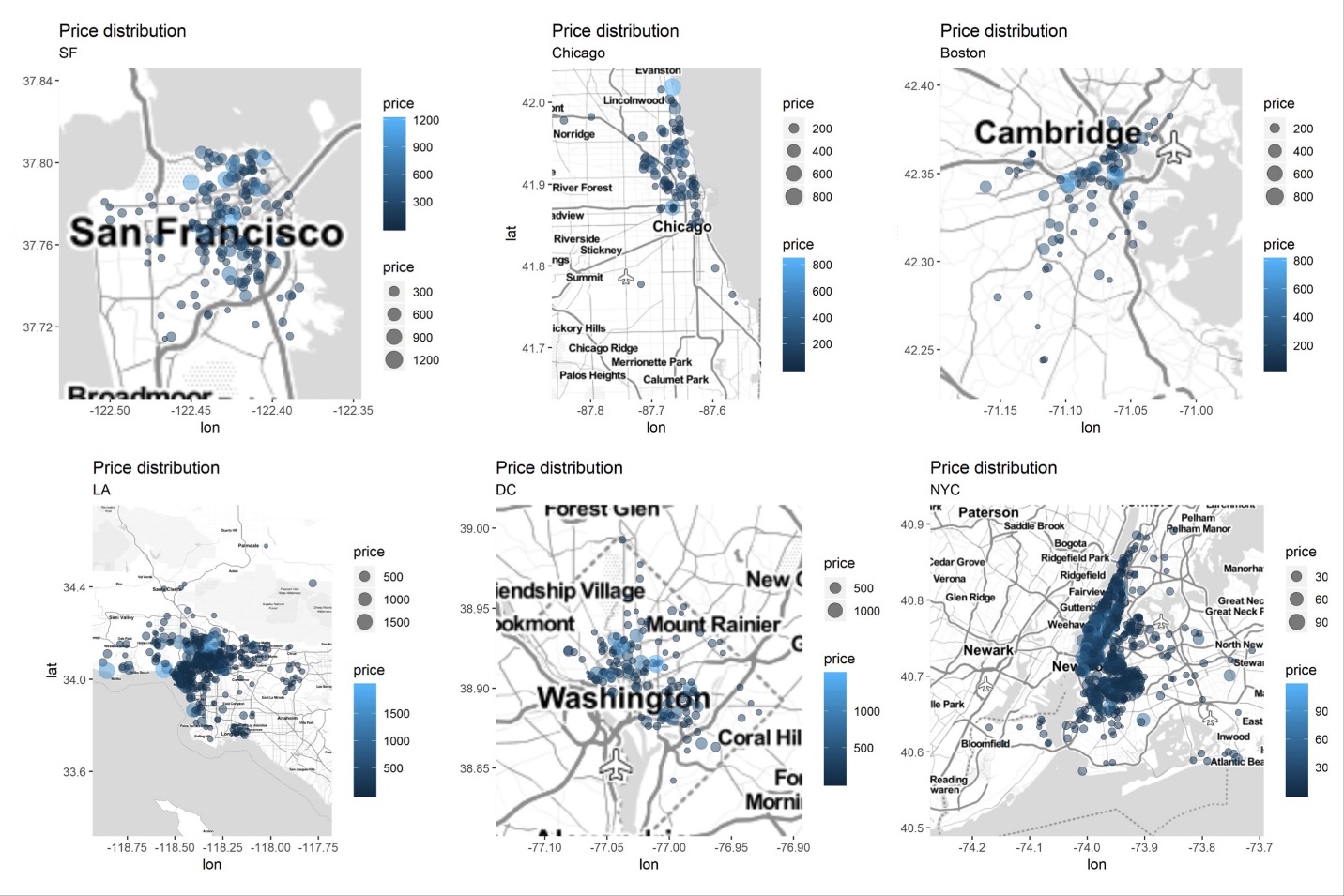


Figure 1 Price distributions across cities

Note: No significant difference or distribution pattern across the West Coast, the East Coast and the Central Region. Take the logarithm of the price to curve down the scale.

The six cites in the Airbnb data are Boston, Chicago, Washington DC, Los Angeles, New York City and San Francisco. Figure 1 shows that the price distribution are similar in those cities. There is no significant difference or pattern in prices between the West Coast, the East Coast and the Central Region.



**2.2 Data Processing**