Airbnb in Los Angeles: Regional Study of House Rules

Brian Lin[1], Lily Damron[1], Aida Ylanan[1], Liam Carrigan[1], and Ignat Kulinka[1]

[1]Department of Statistics, University of California Los Angeles,

December 15, 2017

Abstract

Airbnb listings provide new and unique ways to explore a city. The site, which allows homeowners in the city to open up and share their space with visitors, shares listing information such as the listing's location, price per night, and house rules. Using listing location and house rules, our team conducted a regional analysis of Los Angeles by comparing the average "strictness" level of regions in the city based on data from its Airbnb listings. We created two new variables, *region* and *strictness*, to conduct this analysis. We found that listings in West Los Angeles are stricter, on average, than listings in East Los Angeles. We also found No Smoking and No Partying to be the most common house rules in the city.

1 Introduction

Airbnb, an online marketplace that allows homeowners to advertise their property as a space for visitors to rent, makes its listing data publicly available online. Users who download the dataset get a record of all the information provided about an individual listing, including its location, room type, host name, house rules, listing description, and review count. A listing's location can be described using its coordinates or its neighborhood, as defined by Airbnb. The rich information provided by the geospatial data on every listing inspired our team to pursue a spatial analysis of the data, comparing different regions of Los Angeles based on sampled Airbnb listings from the area.

Los Angeles is the second largest city in the United States according to population size[1], and its many neighborhoods reflect the rich diversity of its citizens. Neighborhoods, both in the city proper and in its peripheries, vary widely in median income level[2] and racial demographics[3], with dramatic differences in the two statistics often observed in comparisons between the eastern and western halves of the city. Many other characteristics can be used to differentiate one neighborhood from another, though these can often be harder to quantify than race and income.

Airbnb listings provide a unique opportunity to compare neighborhoods by using listing data provided by homeowners themselves. Textual analysis of these listings reveals some insight about those homeowners in the area who are willing to share and advertise their living space. House rules, for example, provide information that can be used to explore the strictness of a neighborhood, a quality not easily quantifiable.

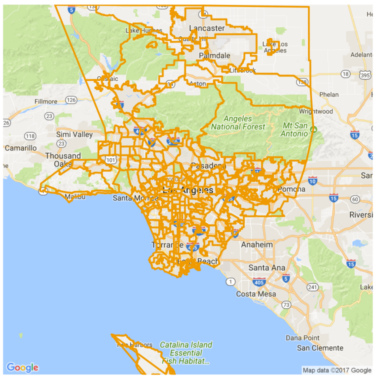
2 Methodology

2.1 Data Cleaning

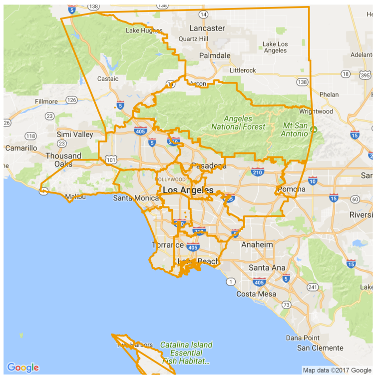
Before we could perform exploratory analysis, some basic data cleaning had to be performed. Unwanted symbols in the *price* variable, such as dollar signs and commas, had to be removed in order to calculate summary statistics.

2.2 Feature Creation

Performing a structural analysis of the dataset revealed that Airbnb had partitioned its listings into 254 unique neighborhoods, with listing density ranging from a single listing in one neighborhood to 2046 listings in another. The disparity in listing count per regions clearly rendered our current plan for neighborhood comparison invalid, and our group sought strategies to make listing distribution more equitable between areas. We decided to group each neighborhood into larger areas using region partitions defined by a Los Angeles Times neighborhood analysis[4]. The article provided a table that matched each neighborhood to a particular region; we scraped this table using R and joined this table with our existing listings dataset by *neighborhood*. Plots showing the outlines of the previous neighborhood listings and the new regional listings can be found below:



(a) Original Outlines



(b) Updated Outlines

The new *region* variable we created reduced area count from 254 neighborhoods to 16 larger regions. Though listings from areas north of Los Angeles City are much fewer in number than those in the city proper and to the south, the distribution of listings per region was much more equitable than the original distributions per neighborhood. The plot showing the distribution of listings per region can be found below:

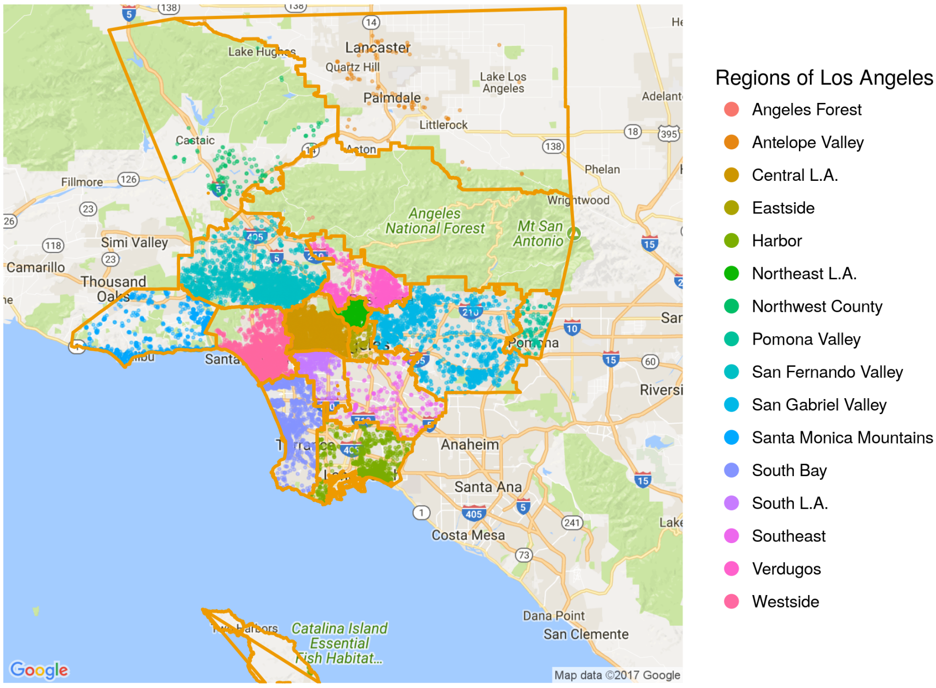


Figure 2: Distribution of Listings

We added another variable to the dataset that quantified more information about each listing. We called this the *strictness* variable, which counted the number of house rules listed in a particular listing. We generated regular expressions that would search the *house_rules* variable for certain phrases that indicated our house rules of interest. We extracted a number of house rules and created some linear models to determine which of these rules were significant. The rules are as follows: No Smoking, No Drugs, No Parties, No Drinking/Alcohol, No Loud Noise, and No Pets.

The *strictness* variable takes values from 0 to 6, each number a count of how many of the most popular rules are named in the listing. For example, a listing with No Smoking and No Drugs would get a *strictness* score of 2. Stricter listings have a *strictness* score closer to 6.

3 Conclusion

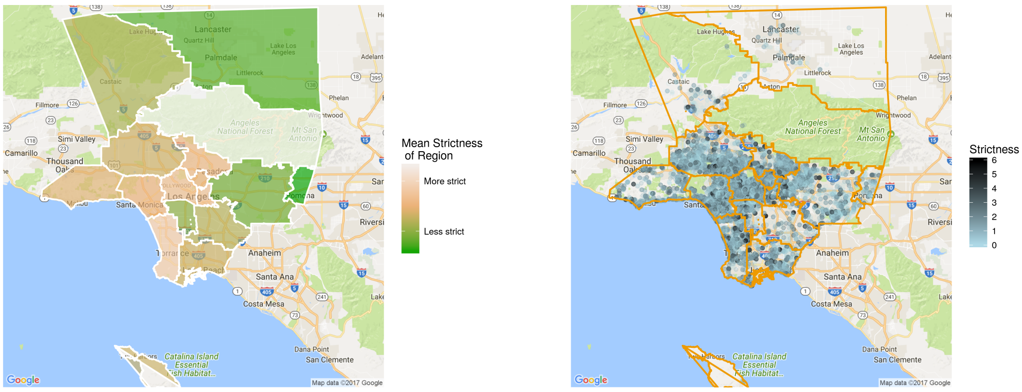
3.1 Results

The following comparative analyses exclude the Angeles Forest region, which only had one observation.

Comparative analysis of the different regions in Los Angeles reveals that South Bay is the strictest region in the area. The area with the most expensive listings, on average, is the Santa Monica Mountain region. The least expensive region, on average, is the Southeast region. A table of our results is displayed below:

Region	Mean Strictness	Mean Price	No Smoke	No Party	No Noise	No Drugs	No Shoes
Angeles Forest	1.20	449.80	0.60	0.20	0.20	0.00	0.20
South Bay	0.94	184.21	0.34	0.20	0.09	0.09	0.05
Eastside	0.86	120.35	0.31	0.22	0.07	0.05	0.05
Central L.A.	0.85	160.43	0.33	0.21	0.09	0.04	0.03
Verdugos	0.81	162.69	0.34	0.16	0.07	0.03	0.04
Westside	0.78	220.84	0.30	0.19	0.09	0.03	0.04
Northeast L.A.	0.76	115.53	0.35	0.17	0.06	0.02	0.02
Santa Monica Mountains	0.74	619.04	0.27	0.19	0.06	0.03	0.04
San Fernando Valley	0.70	182.35	0.27	0.16	0.08	0.04	0.04
Harbor	0.70	135.22	0.28	0.16	0.08	0.05	0.02
Southeast	0.65	89.92	0.21	0.16	0.08	0.07	0.05
Northwest County	0.63	122.02	0.25	0.11	0.07	0.05	0.06
South L.A.	0.58	100.04	0.24	0.15	0.05	0.05	0.03
San Gabriel Valley	0.54	106.12	0.21	0.11	0.03	0.05	0.03
Antelope Valley	0.51	144.22	0.24	0.07	0.04	0.04	0.04
Pomona Valley	0.47	89.09	0.22	0.09	0.05	0.05	0.02

Using this information, we created two plots that display strictness levels per listing and per region. The regional plot displays the average strictness level per region. The plot displaying individual listings also reveals the disparity in listing density per region, an important caveat to keep in mind when interpreting the results. The two plots can be viewed below:



(a) Mean Strictness per Region

(b) Strictness per Listing

We also looked within the *house_rules* and *strictness* variables to see which rules and which pairs of rules are most common. The most common rule is No Smoking, while the most common pair of rules is No Smoking and No Partying. The results are displayed on the stacked bar chart below:

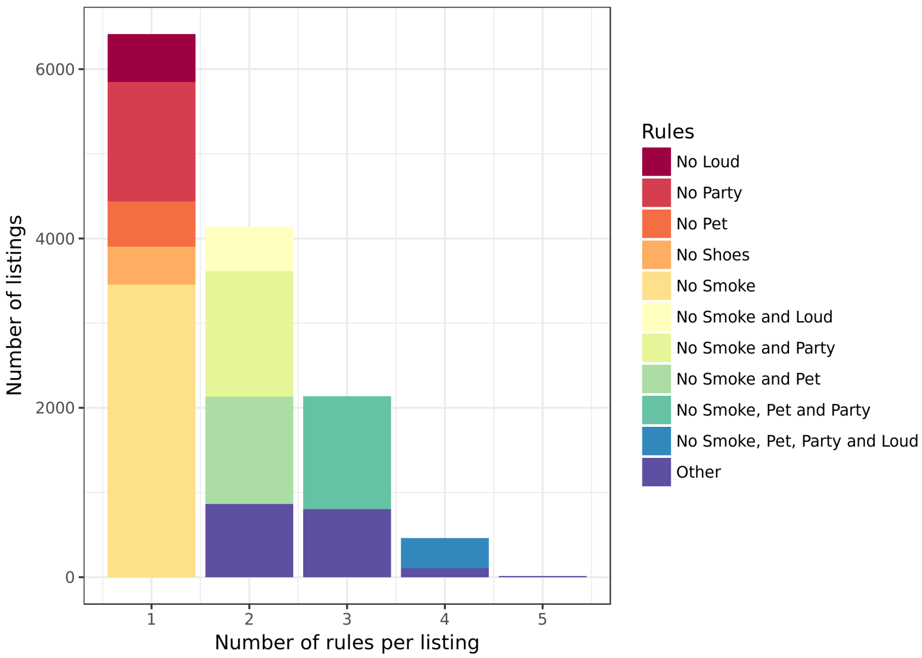


Figure 4: Distribution of Rules

3.2 Discussion

Two variables posed a potential problem for our analysis. We observed extreme values for the *minimum_nights* (365) and *maximum_nights* (9999) variables, suggesting the possibility of faulty Airbnb listings. Because these variables were not directly involved with our analysis of house rules, we chose to keep the data as is.

As discussed earlier, our grouping of Airbnb neighborhoods into larger regions still did not provide an equitable distribution of listings per region. The Angeles Forest region, for example, only had one listing. Further development of this analysis could involve a different regional breakdown – one that more evenly distributes listings per area. However, formal neighborhood partitions in Los Angeles are highly contested and change often, so a new regional breakdown would be difficult to find.

While it was fascinating to explore different regions in Los Angeles based on their average strictness levels from Airbnb, we did not investigate why these results may be so. One idea we discussed involved the different municipal codes per region; some neighborhoods in Los Angeles have laws that make renting out a space difficult for some homeowners. Further development of this analysis could compare these results to existing city laws. The relationship between strictness level and an area's socioeconomic makeup or its popularity with tourists could also be explored.

3.3 Contributions

Dataset selection and *region* variable research were done by Aida and Liam. She added the new variable to the dataset, with help from Ignat. Brian cleaned up the dataset and was instrumental in coding the *strictness* variable. He created linear models and ANOVA tables to explore which variables might be significant. Lily proposed the idea of measuring "strictness" and mapped the variable on some of the region plots, along with using geoJSON files to display neighborhood partitions. Ignat improved many of these plots by formatting their text and making them more readable. He also cleaned up our R code, created tables, and formatted this document in Latex. Liam recorded this process in a report write-up, which was later edited and expanded upon by Aida. All team members were involved in creating and editing the presentation.

References

[1] "The 50 Largest Cities in the United States." *Politifact.com*, www.politifact.com/largestcities/.

[2] "Mapping L.A. by Median Income" *Los Angeles Times*, maps.latimes.com/neighborhoods/income/median/neighborhood/list/.

[3] "Mapping L.A. by Ethnicity." *Los Angeles Times*, maps.latimes.com/neighborhoods/ethnicity/white/neighborhood/list/.

[4] "Neighborhoods List." *Los Angeles Times*, maps.latimes.com/neighborhoods/neighborhood/list/.