Airbnb in LA

Lily Damron, Brian Lin, Aida Ylanan, Ignat Kulinka, Liam Carrigan

Our Question

Based on house rules in Los Angeles Airbnb listings, which neighborhoods in LA are the most strict?

Data Clean-Up and Region Feature Creation

Data Clean-Up

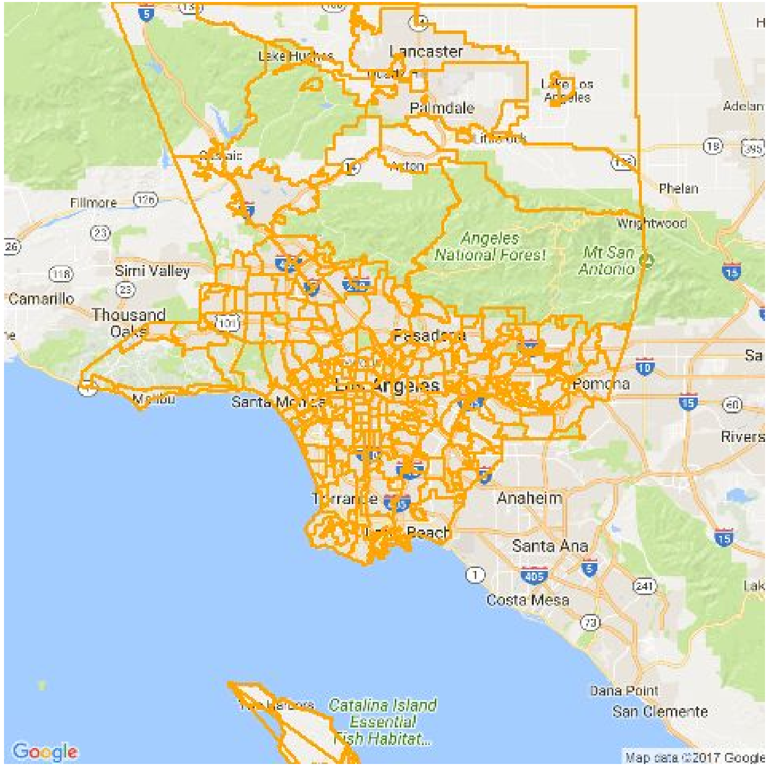
To clean the data, we removed commas and dollar signs from the price variables and empty spaces in the other variables

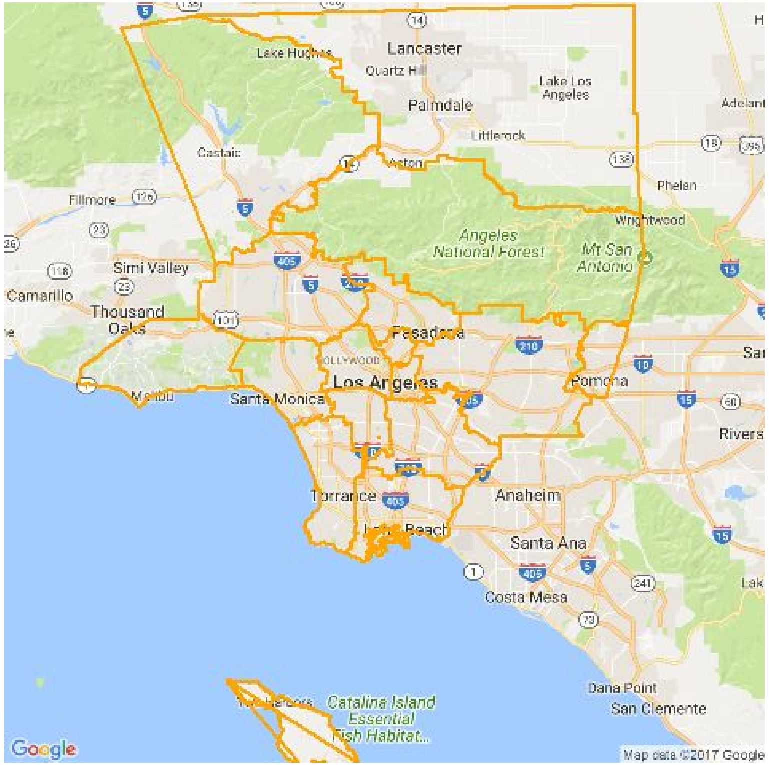
Region Feature Creation

To create the *region* variable, we used data from an LA Times article that mapped the different neighborhoods in Los Angeles.

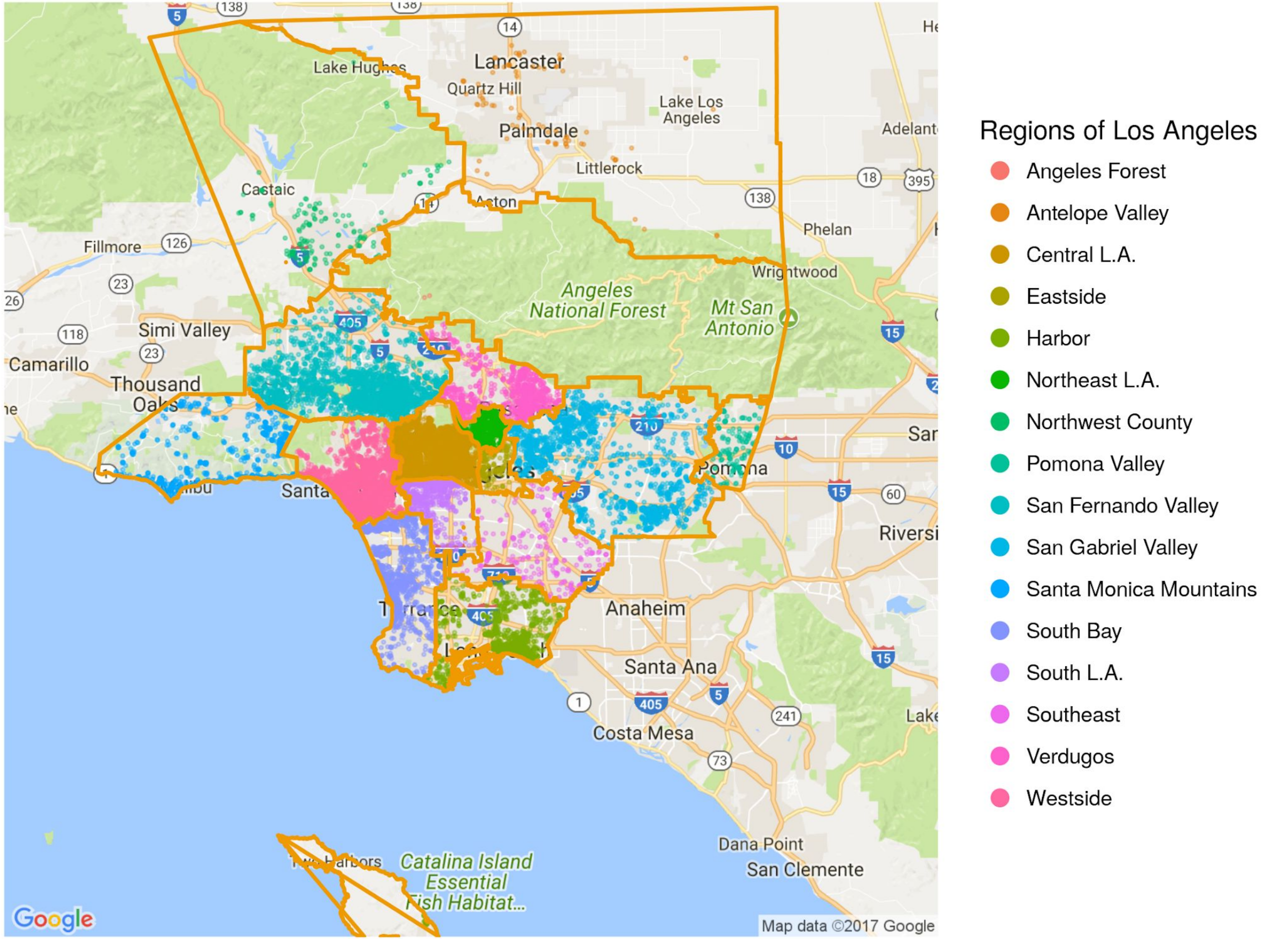
This variable allowed us to simplify the 273 neighborhoods into 16 larger regions in Los Angeles.

Neighborhood vs. Region Comparison



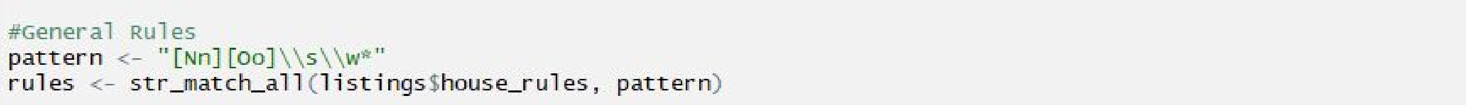


Distribution of Listings in Each Region



House Rule Feature: Creation Process

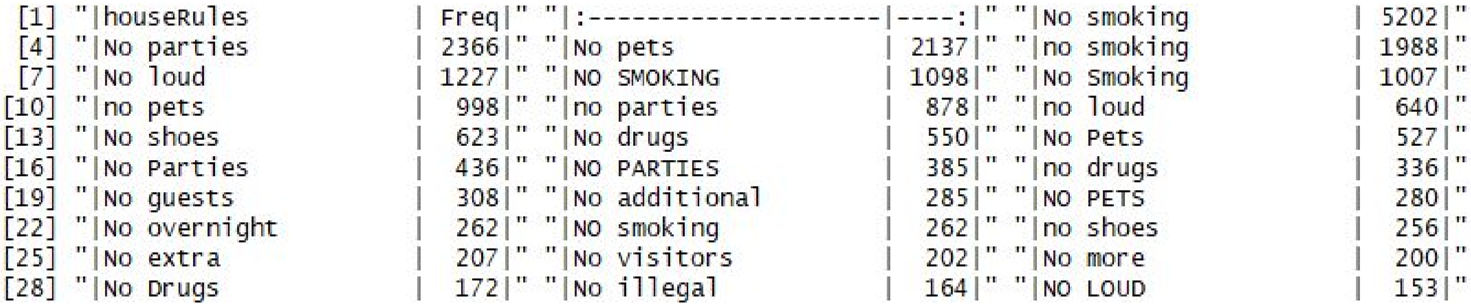
The general rules are identified by using regular expressions and are stored into a list:



```
#General Rules
pattern <- "[Nn][Oo]\\s\\w*"
rules <- str_match_all(listings$house_rules, pattern)
```

After that, we used nested for loops to store each rules into a vector and create a frequency table for all the possible rules.

Six rules with the highest frequency were picked as the *house rule* variables:



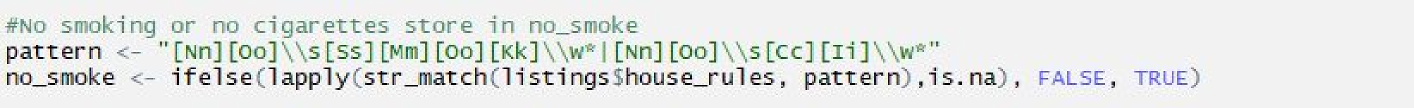
```
 [1] "|houseRules          | Freq|" "|:-------------------|----:|" "|No smoking          | 5202|"
 [4] "|No parties          | 2366|" "|No pets             | 2137|" "|no smoking          | 1988|"
 [7] "|No loud             | 1227|" "|NO SMOKING          | 1098|" "|No Smoking          | 1007|"
[10] "|no pets             |  998|" "|no parties          |  878|" "|no loud             |  640|"
[13] "|No shoes            |  623|" "|No drugs            |  550|" "|No Pets             |  527|"
[16] "|No Parties          |  436|" "|NO PARTIES          |  385|" "|no drugs            |  336|"
[19] "|No guests           |  308|" "|No additional       |  285|" "|NO PETS             |  280|"
[22] "|No overnight        |  262|" "|NO smoking          |  262|" "|no shoes            |  256|"
[25] "|No extra            |  207|" "|No visitors         |  202|" "|No more             |  200|"
[28] "|No Drugs            |  172|" "|No illegal          |  164|" "|NO LOUD             |  153|"
```

House Rule Feature: Creation Process

Once we decided our new variables, we can use regular expressions to match the rules.

When the patterns are matched, a logical value TRUE is then stored into the new *house rule* variable. On the other hand, when the patterns are not matched, FALSE is stored into the new *house rule* variable. In addition, rules with similar meanings are stored into the same rule.

Example



```
#No smoking or no cigarettes store in no_smoke
pattern <- "[Nn][Oo]\\s[Ss][Mm][Oo][Kk]\\w*|[Nn][Oo]\\s[Cc][Ii]\\w*"
no_smoke <- ifelse(lapply(str_match(listings$house_rules, pattern),is.na), FALSE, TRUE)
```

Limitations

Some observations in the original *house_rules* variable have repeated rules listed. These observations were not addressed, and some rules in the frequency table actually have a lower frequency.

However, such a problem is mitigated since only six rules were selected, and they have fairly high frequencies.

Defining “Strictness”

After creating new variables for house rules, we used these variables to determine a listing’s “strictness level.”

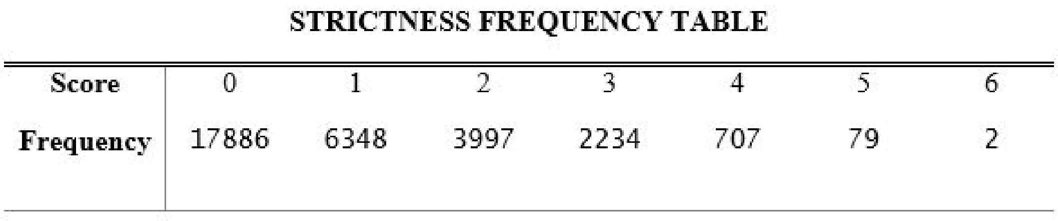
The new *house rule* variables we created are:

- No Smoking
- No Parties
- No Loud Noise
- No Drugs
- No Shoes
- No Pets

Defining “Strictness”

Our “strictness” score, valued 0 to 6, sums the house rule variables to determine the strictness of each listing.

A 0 represents the presence of none of the rules of interest in the listing description, while a 6 represents the presence of all of the rules of interest in the listing description, therefore making it a strict listing.

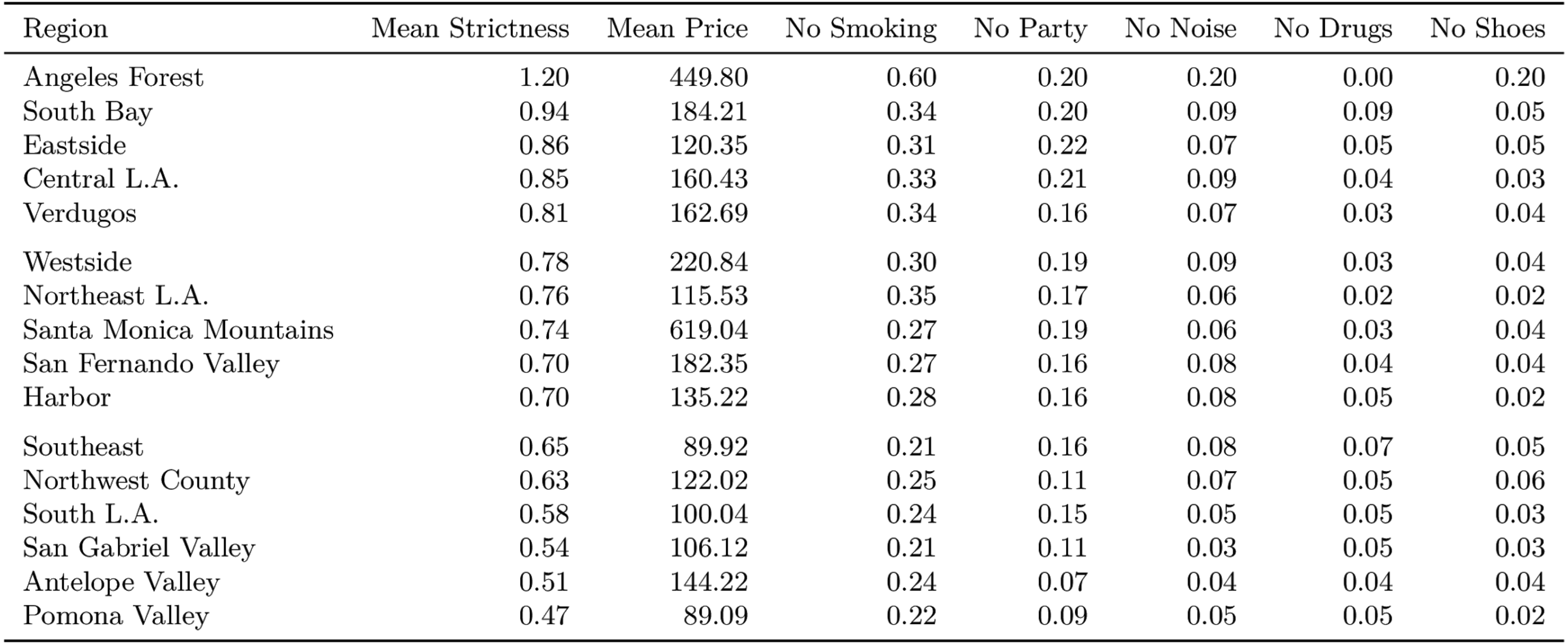


STRICTNESS FREQUENCY TABLE

Score	0	1	2	3	4	5	6
Frequency	17886	6348	3997	2234	707	79	2

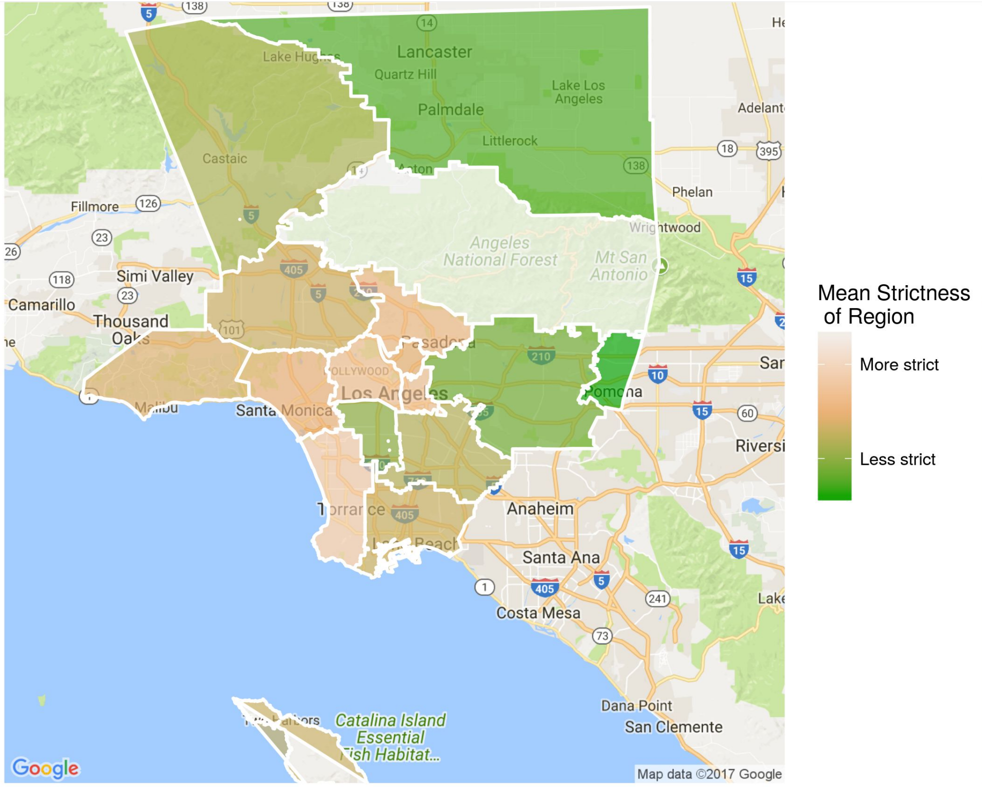
Summary of Analysis

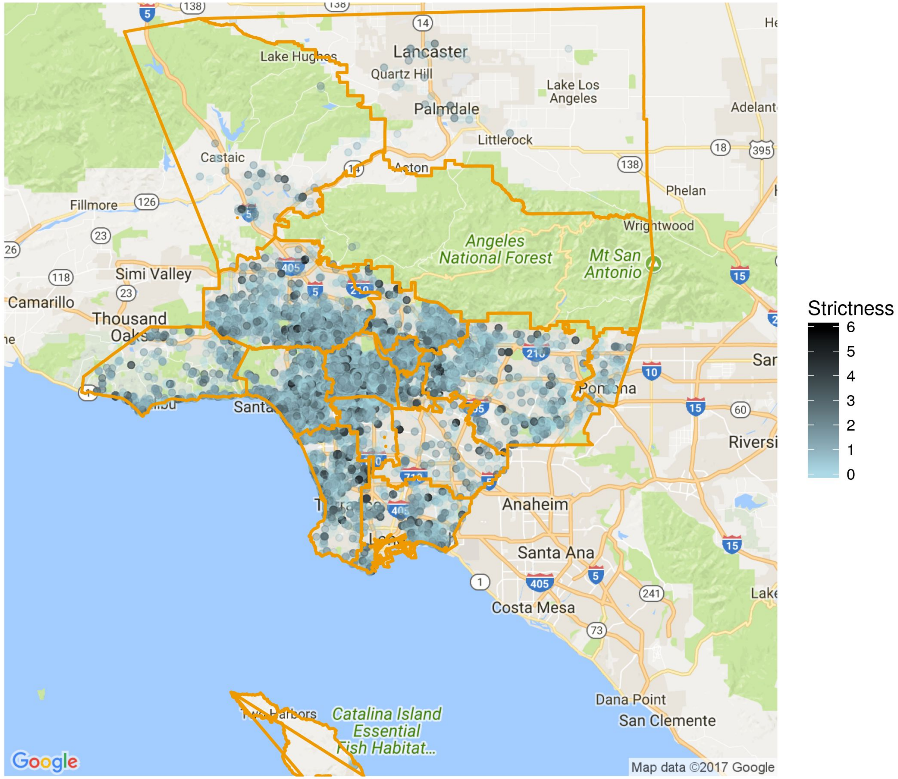
South Bay is the strictest region and Santa Monica Mountains is the most expensive region



Region	Mean Strictness	Mean Price	No Smoking	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

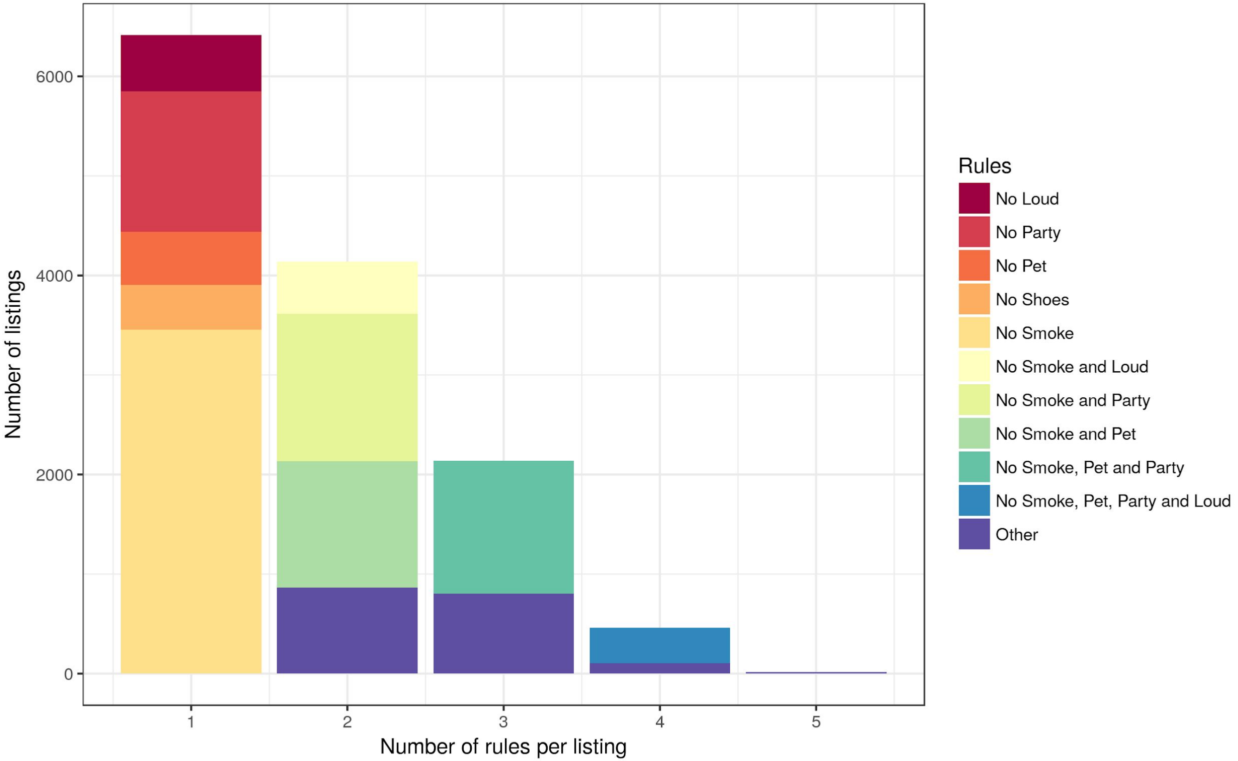
Graphing Strictness





Distribution of Rules for Each Strictness Level

Rule combinations differ per strictness level



Difficulties and Potential Problems

- The Angeles Forest region only has one observation, skewing the results and “strictness” score for this region.
- Similarly, the regions we created are not of equal size. Though the *region* variable allows for a more equitable comparison between listings than using the dataset’s original *neighbourhood* variable, some regions had a lot more observations than others.
- Extreme values for minimum_nights (365) and maximum nights (9999) we observed potentially skewed our analysis of the data.

Future Research

- Strictness score could be adjusted to account for less common rules that we did not include in our original analysis.
- Listings could be partitioned differently from the way they were separated in our *region* variable from the LA Times. New regions could allow for a more equitable distribution of observations per region.

Conclusions and Applications

Strictest Region: South Bay

Most Expensive Area: Santa Monica Mountains

Most Common Rule Combinations: no smoking & no parties (2 rules); no smoking, no pets, no parties (3 rules)

Application: Airbnb could implement a strictness score to their listings as one more piece of info that customers can use to find a listing that fits their needs.