

Midterm Project: Yelp Data Challenge

Xiang Zhao

12/14/2017

1. Introduction

Yelp is a website and also an app collecting the informations of business holders and users. With the increasing population from asia, the number of asian restaurants rising quickly and have been more popular than before. For asian restaurants' holders, in order to offer a better quality for customers and making more profits, analyzing the relationship between the ratings of restaurants and restaurants' informations is important. So, I collected the data from Yelp using SQL in R and filter four styles of restaurants, Chinese, Japanese, Korean and Southeast Asian restaurants with their attributes like price range and noise level and so on to fit several models in order to achieve my goal, finding the relationship between the ratings and restaurants' informations.

At the beginning, I collected the data with all the informations and attributes of restaurants and informations of users. I filter the Chinses, Japanese, Korean and Southeast Asian restaurants with all their informations first. Then I filter six attributes:whether a restaurant has free/paid/no WiFi, the price range of the restaurant, the choices of parking of a restaurant, the noise level of a restaurant, whether a restaurant has a TV or not and whether a restaurant has outdoor seating or not. Next I combined the 'attributes' dataset with 'business' dataset, which gives a whole dataset containing all the informations and attributes of four styles of restaurants. Later, I filter the stars rated by users from 'review' dataset and join it with the 'user' dataset containing users' informations. Finally I join the two datasets of users& reviews and restaurants' informations into one dataset and named it "yelp" which has 472741 rows and 22 variables after cleaning all the NAs.

After getting the clean dataset, I first did EDA especially on the relationships between predictors and response graphically. Then I designed several models to analyze and check these potential relationships statistically and numerically. Finally I summarized the conclusions.

2. Data & Method

2.1 Data source

The data is from yelp: <https://www.yelp.com/dataset/challenge>.

Due to the huge size of dataset after filtering, I saved the filtered dataset into RDS file and reread it in another Rmd file then analyzing. If you need to see the code of reading and cleaning the data, look at ‘Data collecting & cleaning.Rmd’.

Part of data shows below:

1st-5th columns:

business_id	restaurant_style	restaurant_name	restaurant_city	restaurant_state
iMoFE2g4kDG4FfKLJvk3Jw	Korean	Buk Chang Dong Soon Tofu	North York	ON
iMoFE2g4kDG4FfKLJvk3Jw	Korean	Buk Chang Dong Soon Tofu	North York	ON
iMoFE2g4kDG4FfKLJvk3Jw	Korean	Buk Chang Dong Soon Tofu	North York	ON
iMoFE2g4kDG4FfKLJvk3Jw	Korean	Buk Chang Dong Soon Tofu	North York	ON
iMoFE2g4kDG4FfKLJvk3Jw	Korean	Buk Chang Dong Soon Tofu	North York	ON
iMoFE2g4kDG4FfKLJvk3Jw	Korean	Buk Chang Dong Soon Tofu	North York	ON

6th-9th columns:

restaurant_stars	restaurant_review_count	restaurant_WiFi	restaurant_price_range
4	267	no	1
4	267	no	1
4	267	no	1
4	267	no	1
4	267	no	1
4	267	no	1

10th-13rd columns:

garage_parking	street_parking	validated_parking	lot_parking	valet_parking
false	true	false	false	false
false	true	false	false	false
false	true	false	false	false
false	true	false	false	false
false	true	false	false	false
false	true	false	false	false

15th-17th columns:

restaurant_noise_level	restaurant_TV	restaurant_outdoor_seating
average	0	0
average	0	0

restaurant_noise_level	restaurant_TV	restaurant_outdoor_seating
average	0	0
average	0	0
average	0	0
average	0	0

18th-22nd columns:

user_id	user_stars	user_name	user_review_count	user_average_stars
a4PU5fqFJynStdXmIxRflg	3	Fiona	37	3.55
zxnJHs9eEYfVq76LDTg9JA	5	Fiona	10	3.4
8P2LkzPGV4ID_fE7gJGKGg	3	Nadia	153	3.23
6LjTkzT-hFtf_YWjMMQffQ	4	Naseer	3	2
X2FKoMQOkGr17HOdqo6B0g	2	Christine	27	3.07
edEbC4fEPqq2BhbDaFj6Yw	5	Emily	4	4.25

2.2 Method

Tools: SQL, R, csv, RDS, online searching

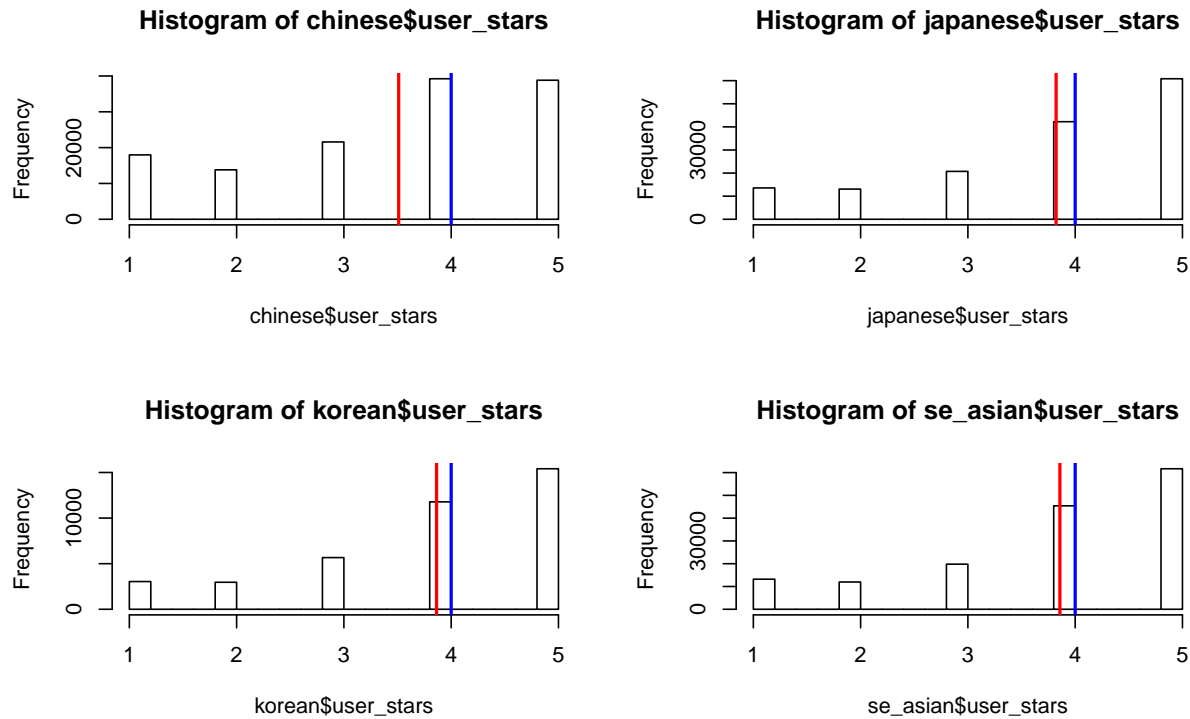
Packages & Functions: ggplot2::ggplot, lme4::lmer&glmer, VGAM::vglm, ordinal:clmm

Models: Multilevel Linear Model, Cumulative Logit Model, Multilevel Multinomial Logistic Regression

3. EDA

3.1 Histogram of stars rated by users toward each four styles of restaurants

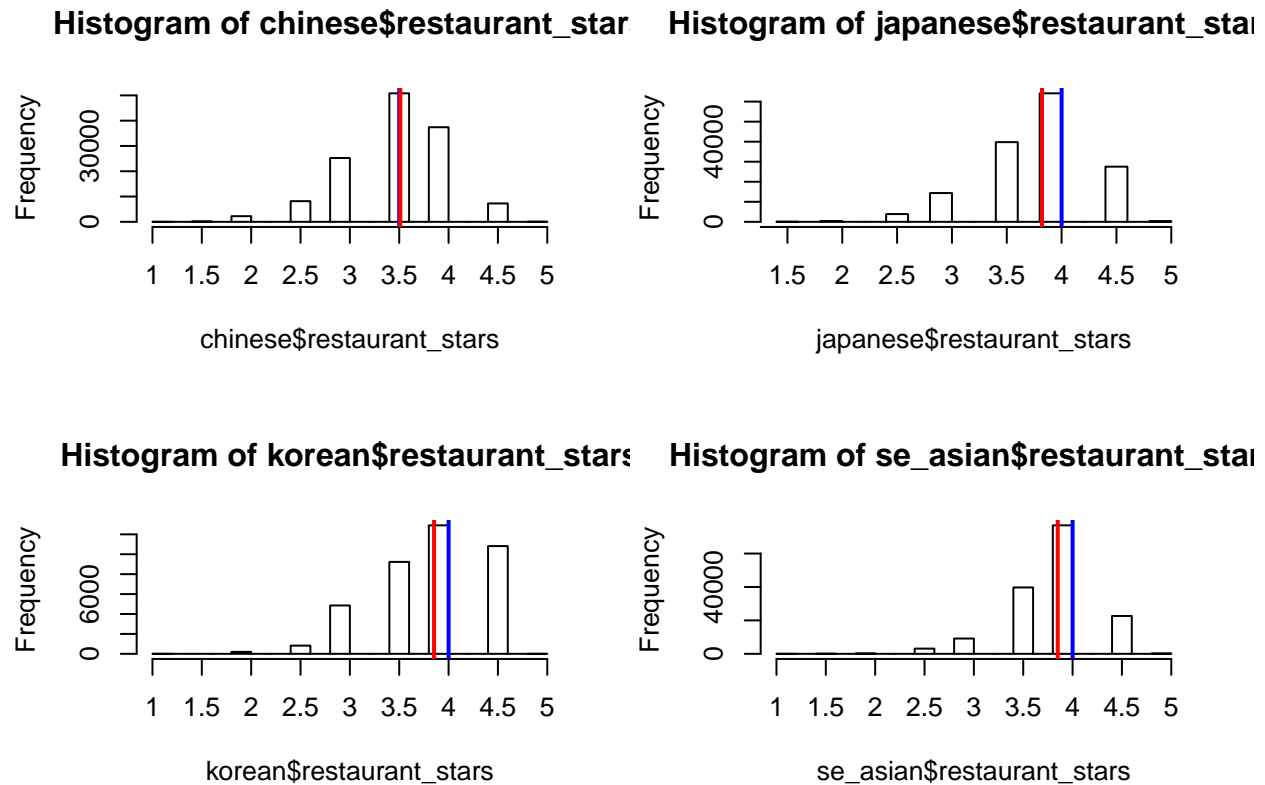
Figure 3.1: Histogram of stars rated by users toward each four styles of restaurants



In the figure 3.1, the red line represents the mean and the blue line stands for the median, we can see that all restaurants have a median of 4 stars rated by customers while Chinese restaurants have a lower mean than Japanese and Southeast Asian restaurants and Korean restaurants have a higher mean. We also can tell that for Japanese, Korean and Southeast Asian restaurants, the most frequent star rated by customers is 5 and the trend for rating from 2 stars to 5 stars is increasing. While for Chinese restaurants, the most frequent star of rating is 4 which is different from another three types of restaurants. Moreover, all these four style of restaurants have more 1-star ratings than 2-star ratings and it's more clear existed among Chinese restaurants.

3.2 Histogram of average stars of each four styles of restaurants

Figure 3.2. Histogram of average stars of each four styles of restaurants



In the figure 3.2, the red line represents the mean and the blue line stands for the median, we can see that Chinese restaurants have the lowest mean and median of average stars at 3.5 while other three types of restaurants have median stars at 4 and mean higher than 3.5 closing to 4. And from the histogram we can tell that a majority of Chinese, Japanese and Southeast Asian restaurants have average stars of 3.5 or 4, less restaurants have average 4.5 stars. Compared to these three types of restaurants, Korean restaurants seem to have better rating that most restaurants have average 4 or 4.5 stars, and then 3.5 even though the median and mean is relatively same as Southeast Asian and Japanese restaurants.

3.3 Stacked bar plot of count of stars within each predictor.

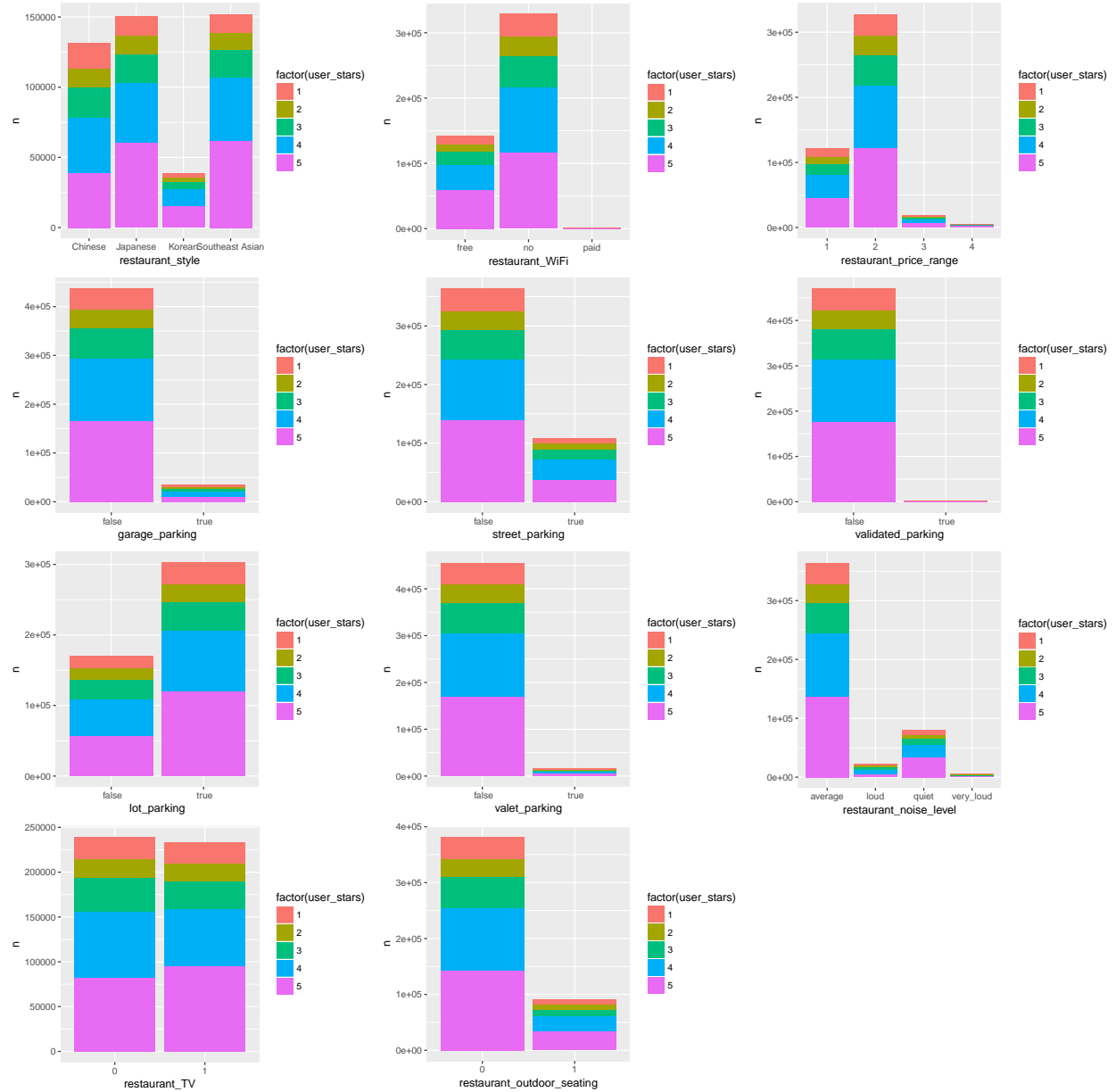


Figure3.3: Stacked bar plot of count of stars within each predictor

In figure3.3, we can see the count of each variable and stars rated by customers.

- In the first graph, the number of southeast asian restaurants is more than Japanese restaurants, than Chinese restaurants, Korean restaurants have the least number. In each category, the count of stars is increasing from 1 star to 5 stars.
- In the second graph, a vast majority of restaurants don't have WiFi, about 1/3 restaurants have free WiFi, just about a hundred restaurants have paid WiFi. In each category, the count of stars is increasing from 1 star to 5 stars.
- In the third graph, more than 300000 restaurants are in price range 2 which is the second cheap range, and then is the cheapest range 1 which has more than 100000 restaurants. The sum of restaurants in range 3 and range 4 are less than 30000. In each category, the count of stars is increasing from 1 star to 5 stars.
- In the fourth to eighth graphs, showing whether a restaurants has garage, street parking, parking lot, valet parking and validated parking. Besides parking lots, more restaurants don't have garage, valet parking, validated parking and street parking while more restaurants have parking lots. For validated parking, there is too few restaurants got.
- In the ninth graph, the noise of restaurants is grouped by 4 categories: quiet, average, loud and very loud. The most restaurants, more than 350000, have an average noise level, less than 100000 restaurants have a quiet level and the rest of restaurants have loud or very loue noise level.
- In the tenth graph, we can see that about 1/2 restaurants have TV and the other 1/2 don't have. Even the number of restaurants who have TV is a little bit smaller than the number of restaurants who don't have TV, the number of 5-star ratings of restaurants having TV is more and in each category, the count of stars is increasing from 1 star to 5 stars.
- In the eleventh graph, we can see that more than 350000 restaurants don't have outdoor seatings, less than 100000 restaurants have outdoor seating. In each category, the count of stars is increasing from 1 star to 5 stars.

3.4 Stacked bar plot of percentage of stars within each predictor.

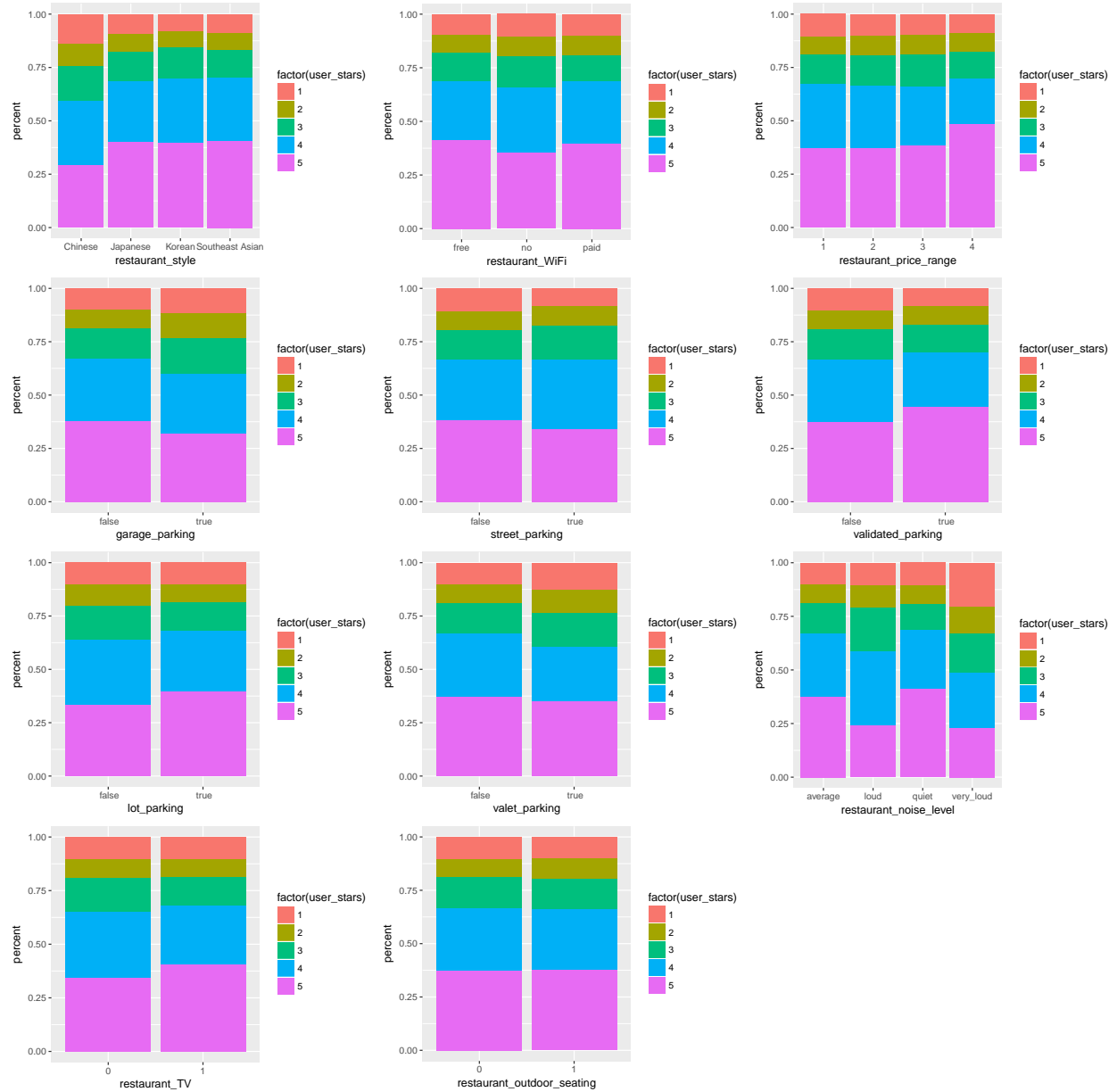


Figure3.4: Stacked bar plot of percentage of stars within each predictor

In figure 3.4, we can see the percent of stars rated by customers in each variable.

- In the first graph, the percent of 5-star ratings among Japanese, Korean and Southeast Asian restaurants are relatively same and they are higher than those of Chinese restaurants'.
- In the second graph, the restaurants with free WiFi have the bigger proportion of 5-star rating, and then is paid WiFi and the last is no WiFi. It seems that as long as there is a WiFi, higher chance to be rated of higher score.
- In the third graph, a clear rise of percent of 5-star rating between restaurants within price range 4 and restaurants within other three ranges. So it's kind of contradictory that people want to find a restaurant inexpensive. But we should be reminded that the absolute number of restaurants within range 4 is highly small which means this part of rating may be rated by higher income people who seek for high quality of the restaurants like decoration which may lead to higher cost.
- In the fourth to eighth graphs, showing whether a restaurant has garage, street parking, parking lot, valet parking and validated parking. It shows that restaurants without garage or without street parking or without valet parking tend to have higher proportion of 5-star ratings while restaurants having parking lots and validated parking tend to have higher proportion of 5-star ratings.
- In the ninth graph, we can see that restaurants are in quiet level and average noise level are going to have higher percent of 5-star ratings while loud and very loud restaurants result in lower proportion of 5-star ratings.
- In the tenth graph, restaurants having TV tend to have a 5% to 6% higher proportion of 5-star ratings than restaurants without TVs.
- In the eleventh graph, we can see that whether restaurants having or not having outdoor seatings rarely have influence on the proportion of 5-star ratings.

4. Model fitting & Analysis

4.1 Linear Regression with Mixed Effects

I fit 3 types of linear models in this section. First, I use four datasets to fit four models of each style individually. Second I use the sample dataset, treating four styles as fixed effect to fit a model with no pooling of style. Third, I use the sample dataset to fit a model treating styles as random effects. All these three models contain the variables I selected in section3 and besides, all these three models are including variable ‘city:state’ as random effect by varying intercepts.

4.1.1 Fit model individually with respect to the style of restaurants

First I fit the model only with data of Chinese restaurants.

Model 1:

$$Star_{[i]} = \beta_0 + \beta_1 WiFi_i + \beta_2 PriceRange_i + \beta_3 Garage_i + \beta_4 StreetParking_i + \beta_5 ValidatedParking_i + \beta_6 ParkingLot_i + \beta_7 ValetParking_i + \beta_8 NoiseLevel_i + \beta_9 TV_i + \beta_{10} OutdoorSeating_i + City : State_{j[i]} + \epsilon_i,$$

where $City : State_j \sim N(\mu_{City:State}, \sigma_{City:State}^2)$

Only with the data of Chinese restaurants.

Second I fit the model only with data of Japanese restaurants.

Model 2:

$$Star_{[i]} = \beta_0 + \beta_1 WiFi_i + \beta_2 PriceRange_i + \beta_3 Garage_i + \beta_4 StreetParking_i + \beta_5 ValidatedParking_i + \beta_6 ParkingLot_i + \beta_7 ValetParking_i + \beta_8 NoiseLevel_i + \beta_9 TV_i + \beta_{10} OutdoorSeating_i + City : State_{j[i]} + \epsilon_i,$$

where $City : State_j \sim N(\mu_{City:State}, \sigma_{City:State}^2)$

Only with the data of Japanese restaurants.

Third I fit the model only with data of Korean restaurants.

Model 3:

$$Star_{[i]} = \beta_0 + \beta_1 WiFi_i + \beta_2 PriceRange_i + \beta_3 Garage_i + \beta_4 StreetParking_i + \beta_5 ValidatedParking_i + \beta_6 ParkingLot_i + \beta_7 ValetParking_i + \beta_8 NoiseLevel_i + \beta_9 TV_i + \beta_{10} OutdoorSeating_i + City : State_{j[i]} + \epsilon_i,$$

where $City : State_j \sim N(\mu_{City:State}, \sigma_{City:State}^2)$

Only with the data of Korean restaurants.

Finally I fit the model only with data of Southeast Asian restaurants.

Model 4:

$$Star_{[i]} = \beta_0 + \beta_1 WiFi_i + \beta_2 PriceRange_i + \beta_3 Garage_i + \beta_4 StreetParking_i + \beta_5 ValidatedParking_i + \beta_6 ParkingLot_i + \beta_7 ValetParking_i + \beta_8 NoiseLevel_i + \beta_9 TV_i + \beta_{10} OutdoorSeating_i + City : State_{j[i]} + \epsilon_i,$$

where $City : State_j \sim N(\mu_{City:State}, \sigma_{City:State}^2)$

Only with the data of Southeast Asian restaurants.

Check the residual of the model:

Figure4.1: Binned residual plot(Chinese)

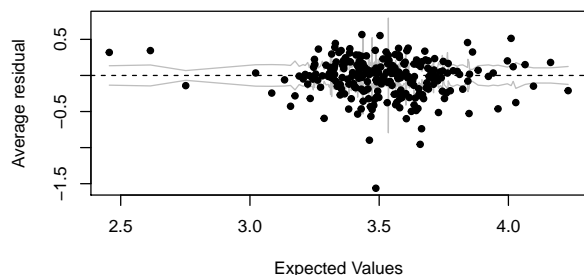


Figure4.2: Binned residual plot(Japanese)

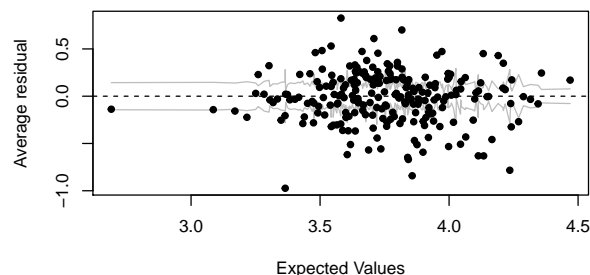


Figure4.3: Binned residual plot(Korean)

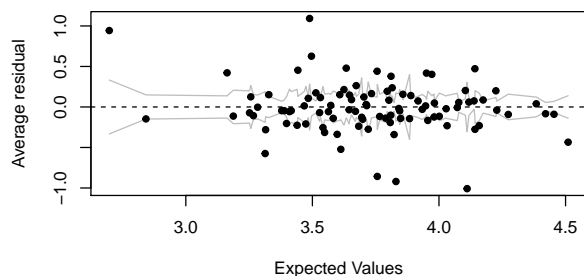
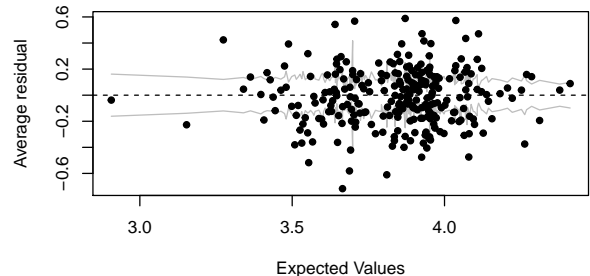


Figure4.4: Binned residual plot(Southeast Asian)



From figure4.1 to figure4.4, we can see the graphs of residual for each model. Basically, the residual plots look relatively good since most residuals are less than 0.5 the plots show symmetrical form. The model of Korean restaurants and Southeast Asian are better than another two because the residual of Korean restaurants are smoothly closer to zero rather than showing a kind of circle or round around axis. And the residual of Southeast Asian restaurants have a narrower range of $(-0.6, 0.6)$ which is better than $(-1.0, 1.0)$, it means the model for Southeast Asian restaurants is fited precisely.

Compare and interpret coefficients:

	Chinese	Japanese	Southeast Asian	Korean
Intercept	3.4604814	3.6375503	3.7802886	4.3633745
restaurant_WiFino	-0.0742453	-0.1252521	-0.0001770	-0.3159329
restaurant_WiFipaid	0.3910925	-0.3251938	-0.0840379	0.0178399
restaurant_price_range2	-0.0762948	-0.0172033	0.0168487	-0.2816079
restaurant_price_range3	0.2411818	0.1966469	0.2786113	-0.7230187
restaurant_price_range4	0.2925457	0.1249455	0.5790789	NA
garage_parking true	0.0624246	-0.1297917	-0.2807717	-0.6903903
street_parking true	0.2367694	0.1368207	0.1555946	-0.1428238
validated_parking true	0.3180269	0.1373965	0.2210321	-0.3301659
lot_parking true	0.1697010	0.0323071	0.1063941	-0.0034391
valet_parking true	-0.0963215	-0.2383541	-0.2055574	0.5747600
restaurant_noise_level	-0.1036612	-0.0814839	0.0140009	-0.1348616

	Chinese	Japanese	Southeast Asian	Korean
restaurant_noise_levelquiet	0.0704491	0.1824398	0.0119266	0.0229342
restaurant_noise_levelvery_loud	-0.8383866	0.1696895	-0.7398966	-0.5233938
restaurant_TV1	0.0810256	-0.0488169	-0.0256351	-0.1244898
restaurant_outdoor_seating1	-0.1163634	-0.0570491	-0.0037261	0.0342456

From the coefficients' table, we can see a lot of informations through comparing the coefficients across each style of restaurants.

- For intercept, Korean restaurants have the highest average stars of 4.36 rated by customers, 0.5 higher than the second restaurants of Southeast Asian. Chinese restaurants have a lowest average stars of 3.4.

- For WiFi variable, the customers rating Koreans restaurants are more sensitive to the unavailability of WiFi. It decreases 0.31 of stars compared to the Korean restaurants who have free WiFi. While for the paid WiFi, Chinese restaurants and Japanese restaurants who have paid WiFi tend to have a more than 0.3-stars ratings than restaurants who have free WiFi, which is counterintuitive except for the paid WiFi may be much faster than free WiFi.

- For price range variable, the 2-price-range Chinese restaurants and Japanese restaurants tend to have lower ratings than 1-price-range restaurants. The 3&4-price-range Chinese and Japanese restaurants are going to have higher ratings than 1-price-range restaurants. For Southeast Asian restaurants, the ratings tend to increase along with the rise of price range, the 4-price-range Southeast Asian restaurants are going to have 0.6 higher ratings than 1-price-range restaurants on average. Well for Korean restaurants, our sample data does not contain the 4-price-range restaurants, but with the rise from 1 to 3 price range, the ratings of restaurants are going to decrease and the average rating of 3-price-range Korean restaurants tend to have 0.7 stars lower than 1-price-range restaurants. By summary, 4-price-range Chinese restaurants, 3-price-range Japanese restaurants, 4-price-range Southeast Asian restaurants and 1-price-star Korean restaurants tend to have the highest ratings among each style with respect to price range.

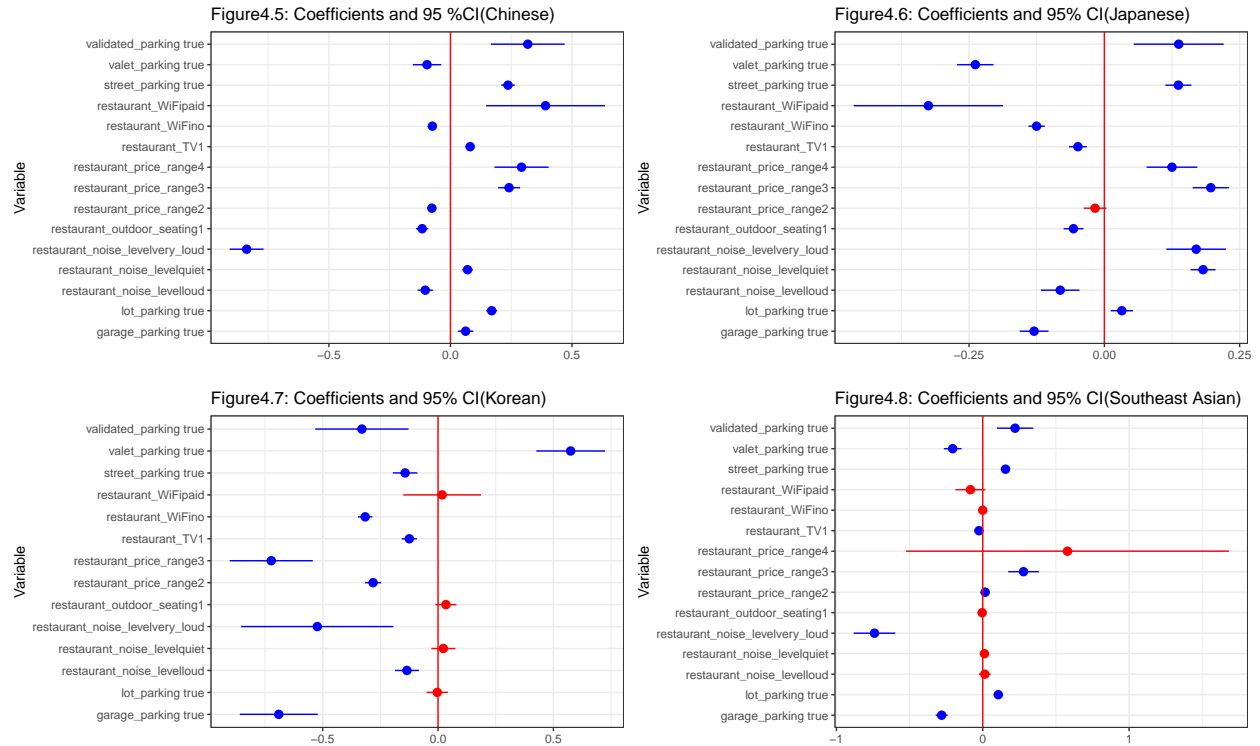
- For variables contain parking informations, Chinese restaurants who have garage, street parking, parking lots and validated parking tend to have higher ratings than without these conditions especially of a 0.3 higher rating of restaurants who have validated parking than who don't have, while the restaurants with valet parking tend to have 0.1 lower ratings than restaurants without valet parking. Japanese restaurants who have who have street parking, validated parking and parking lots tend to have 0.13, 0.13 and 0.03 higher ratings than without these conditions, while for Japanese restaurants who have garage and valet parking tend to have 0.13 and 0.23 lower ratings than without these conditions. Southeast Asian restaurants with street parking, validated parking and parking lots tend to have 0.16, 0.22 and 0.11 higher ratings than restaurants without these conditions, while Southeast Asian restaurants with garage and valet parking tend to have 0.28 and 0.21 lower ratings than restaurants with garage and valet parking. For Korean restaurants with garage, street parking, validated parking and parking lots tend to have 0.69, 0.14, 0.33 and 0.003 lower ratings than restaurants with these conditions. But for Korean restaurants with valet parking tend to have 0.57 higher ratings than restaurants who don't have valet parking.

- For noise level variable, the baseline is average noise level, the coefficients of noise level of models for Chinese and Korean restaurants kind of make sense since louder noise level tend to have lower average ratings and quieter noise level tend to have higher ratings. For example, Chinese restaurants with quiet noise level tend to have 0.70 higher stars than average-noise-level restaurants and restaurants with loud and very loud noise level tend to have 0.10 and 0.84 lower ratings than restaurants with average noise level, similar as Korean restaurants. While quiet and very loud Japanese restaurants seem to have around 0.17 higher ratings than restaurants with average noise level but loud-noise Japanese restaurants tend to have 0.08 lower ratings. Only very loud Southeast Asian restaurants tend to have 0.74 lower ratings while loud and quiet restaurants seem to have same ratings as average restaurants since the coefficients are around 0.01.

- For having or not having a TV as a variable, Chinese restaurants with TV tend to have 0.08 higher ratings and Japanese, Korean and Southeast Asian restaurants may have 0.05, 0.03 and 0.01 lower ratings than restaurants without TV.

- For having or not having outdoor seatings, except for Korean restaurants having outdoor seatings tend to have 0.03 higher ratings, Chinese, Japanese and Southeast Asian restaurants with outdoor seatings may have 0.12, 0.06 and 0.04 lower ratings compared to restaurants without outdoor seatings.

Check the coefficients' significance:



From figure 4.5 to figure 4.8 we can see the 95% confidence interval of each coefficients of each style of restaurants and whether it crosses the point zero, the vertical red line is the line of $y = 0$, if it does, the coefficient may not be significant at level 5%. For Chinese restaurants, all the coefficients seem to be significant at level 5%. For Japanese restaurants, the variable “price range” at level 2 may be not significant since its confidence interval crosses 0 but rest of the variables are significant at level 5%. For Korean restaurants, indicators paid WiFi, having outdoor seating, quiet noise level and having parking lot are not significant at level 5%. For Southeast Asian restaurants, indicators paid WiFi, no WiFi, price range 4, having outdoor seating, quiet noise level and loud noise level are not significant at level 5%.

4.1.2 Multilevel linear model (use sample data and treat 'restaurant_style' as fixed effect)

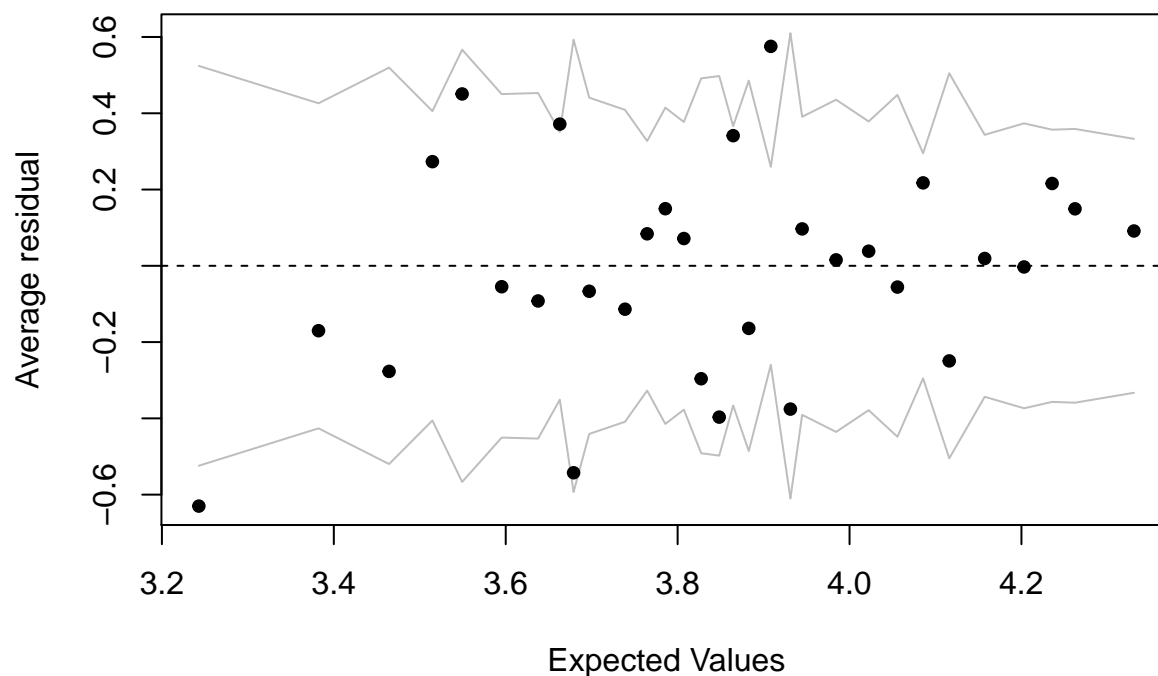
Model 5:

$$Star_{[i]} = \beta_0 + \beta_1 WiFi_i + \beta_2 PriceRange_i + \beta_3 Garage_i + \beta_4 StreetParking_i + \beta_5 ValidatedParking_i + \beta_6 ParkingLot_i + \beta_7 ValetParking_i + \beta_8 NoiseLevel_i + \beta_9 TV_i + \beta_{10} OutdoorSeating_i + \beta_{11} Style_i + City : State_{j[i]} + \epsilon_i,$$

where $City : State_j \sim N(\mu_{City:State}, \sigma_{City:State}^2)$

Check the residual of the model:

Figure4.9: Binned residual plot of fitting rated stars of sample data



The binned residual plot is kind of good because the residuals drops around 0 not too unbalanced and the range of residuals is 1.2 which is considered not too high for me.

Interpret coefficients:

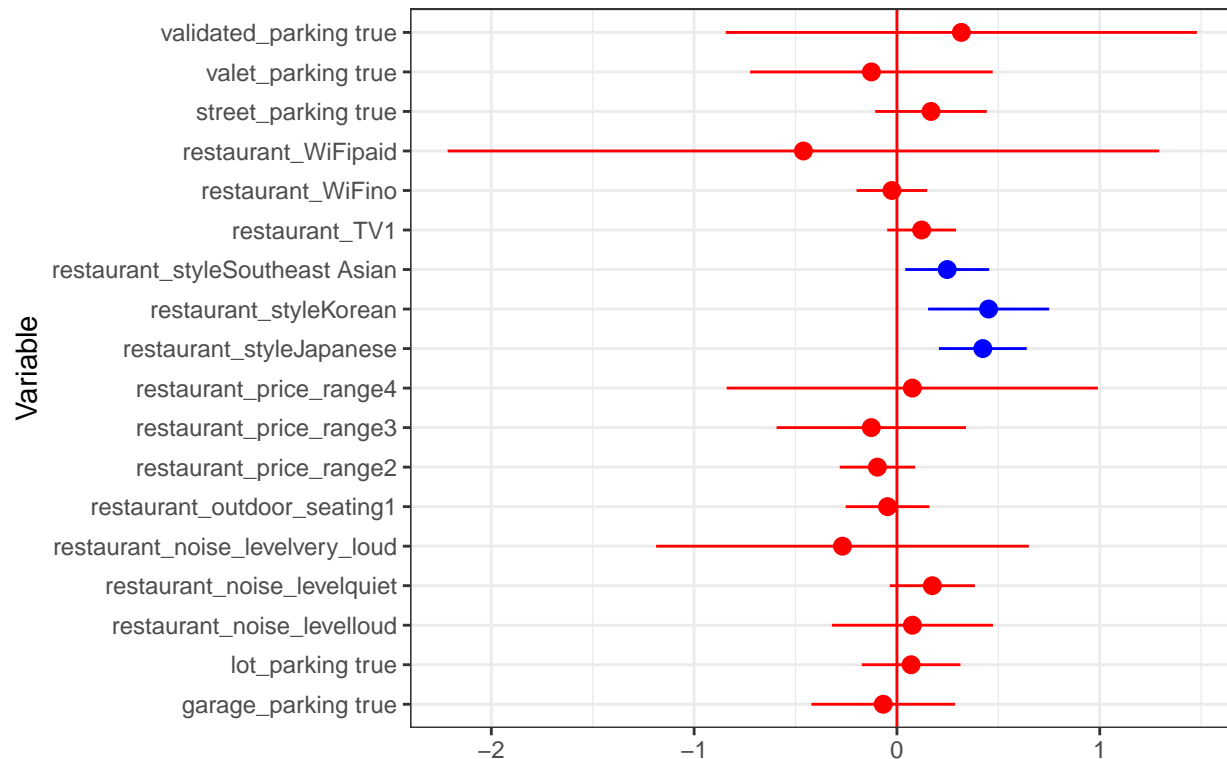
	Estimate	Std. Error
(Intercept)	3.4580916	0.1697024
restaurant_WiFino	-0.0248011	0.0887901
restaurant_WiFipaid	-0.4611872	0.8949567
restaurant_price_range2	-0.0962705	0.0946798
restaurant_price_range3	-0.1264599	0.2382296
restaurant_price_range4	0.0758808	0.4669444

	Estimate	Std. Error
garage_parking true	-0.0676114	0.1805795
street_parking true	0.1680177	0.1403383
validated_parking true	0.3175726	0.5926314
lot_parking true	0.0699468	0.1240592
valet_parking true	-0.1259401	0.3050991
restaurant_noise_level loud	0.0764199	0.2026818
restaurant_noise_level quiet	0.1745072	0.1071764
restaurant_noise_level very_loud	-0.2685261	0.4688006
restaurant_TV1	0.1218393	0.0865625
restaurant_outdoor_seating1	-0.0462853	0.1054390
restaurant_style Japanese	0.4234650	0.1103904
restaurant_style Korean	0.4519414	0.1522557
restaurant_style Southeast Asian	0.2478029	0.1055986

From the coefficients table, we see that on average ratings, Southeast Asian restaurants tend to have 0.24 higher ratings, Japanese restaurants tend to have 0.42 higher ratings and Korean restaurants tend to have 0.45 higher ratings, the average star of average-noise-level Chinese restaurants without WiFi, garage, street parking, parking lot, validated parking, valet parking, TV, outdoor seating and within price range 1 is 3.5. For each variable, free WiFi, price range 4, having street parking, having validated parking, having parking lots, not having garage, not having valet parking, quiet atmosphere, having TV, not having outdoor seating will bring the restaurants higher scored ratings.

Check the coefficients' significance:

Figure4.10: Coefficients and 95% CI



From the graph of confidence intervals, we can see that only the coefficients of style of restaurant is significant at level 5%, other variables are close to zero or have big standard errors.

4.1.3 Multilevel linear model (use sample data and treat ‘restaurant_style’ as random effect)

Model 6:

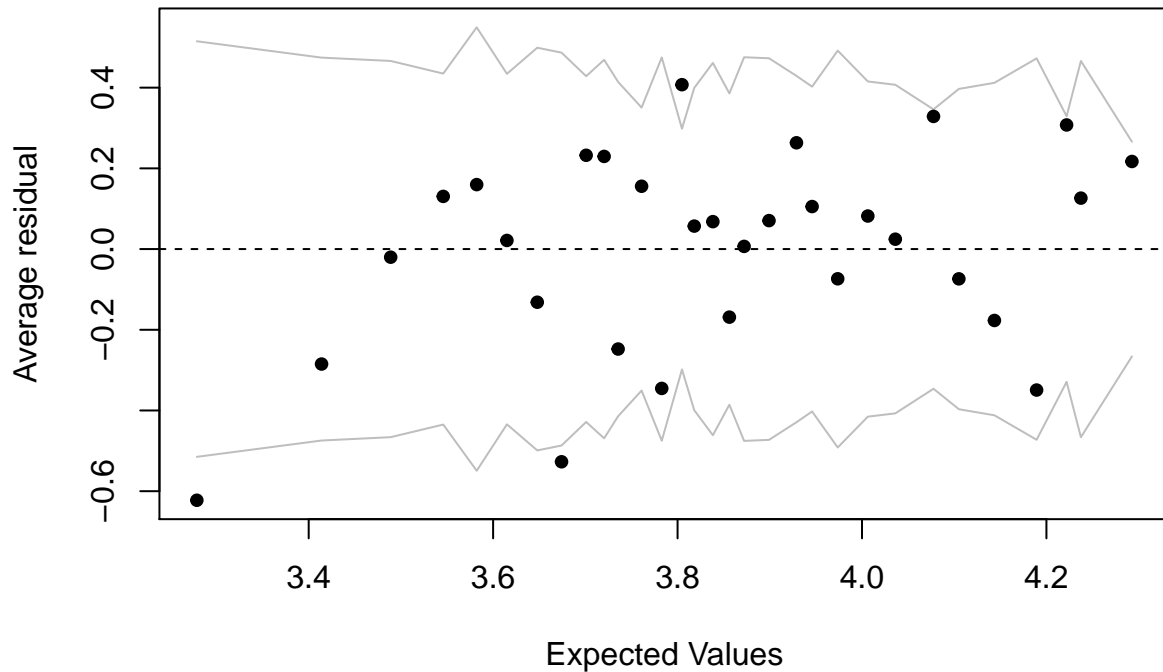
$$Star_{[i]} = \beta_0 + \beta_1 WiFi_i + \beta_2 PriceRange_i + \beta_3 Garage_i + \beta_4 StreetParking_i + \beta_5 ValidatedParking_i + \beta_6 ParkingLot_i + \beta_7 ValetParking_i + \beta_8 NoiseLevel_i + \beta_9 TV_i + \beta_{10} OutdoorSeating_i + Style_{k[i]} + City : State_{j[i]} + \epsilon_i,$$

where $City : State_j \sim N(\mu_{City:State}, \sigma_{City:State}^2)$,

and $Style_k \sim N(\mu_{Style}, \sigma_{Style}^2)$.

Check the residual of the model:

Figure4.11: Binned residual plot of fitting rated stars of sample data:



The residual plot is better than the former one since the range of residuals decreases and most residuals would be between 0 to 0.4 which can be somewhat precise when predicting the ratings.

Interpret coefficients:

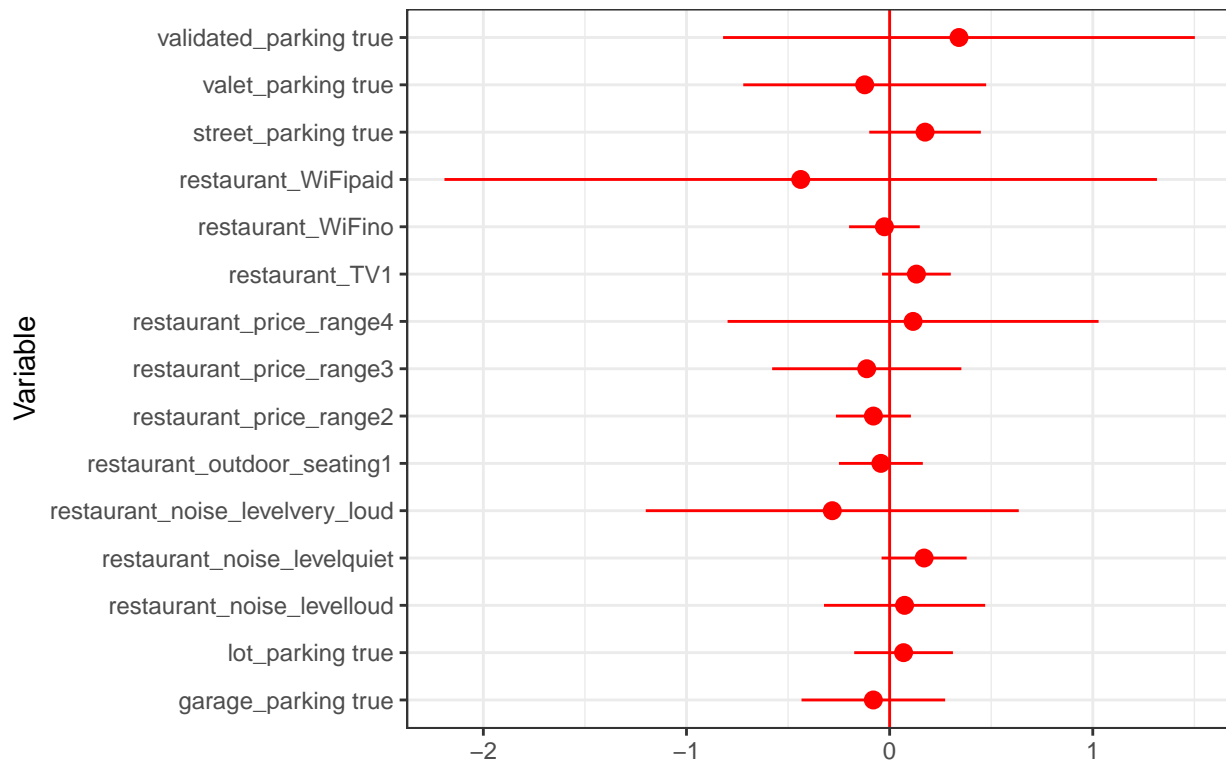
	Estimate	Std. Error
(Intercept)	3.7105977	0.1920120
restaurant_WiFino	-0.0256253	0.0887848
restaurant_WiFipaid	-0.4377416	0.8946775

	Estimate	Std. Error
restaurant_price_range2	-0.0799349	0.0941523
restaurant_price_range3	-0.1127681	0.2375745
restaurant_price_range4	0.1153597	0.4658396
garage_parking true	-0.0802873	0.1803365
street_parking true	0.1743039	0.1401028
validated_parking true	0.3412449	0.5924430
lot_parking true	0.0685476	0.1239322
valet_parking true	-0.1226750	0.3050268
restaurant_noise_levelquiet	0.0734944	0.2023519
restaurant_noise_levelquiet	0.1696887	0.1068212
restaurant_noise_levelvery_loud	-0.2824323	0.4683882
restaurant_TV1	0.1318700	0.0862675
restaurant_outdoor_seating1	-0.0431546	0.1053209

This time, the average stars of a restaurant with free WiFi, at price level 1, no garage, no valet parking, no street parking, no validated parking, no parking lot, at average noise level, not having TV and without outdoor seating should be 3.7. Free WiFi, price range 4, no valet parking, having street parking, validated parking and parking lot, at quiet noise level, having TV and without outdoor seating tend to increase the stars of rating.

Check the coefficients' significance:

Figure4.12: Coefficients and 95% CI



From graph, none of our variables is significant at level 5%.

Interpret the varying intercepts of random effect:

	city:state
Chinese	0.10330
Japanese	0.14650
Korean	0.12030
Southeast Asian	0.09686

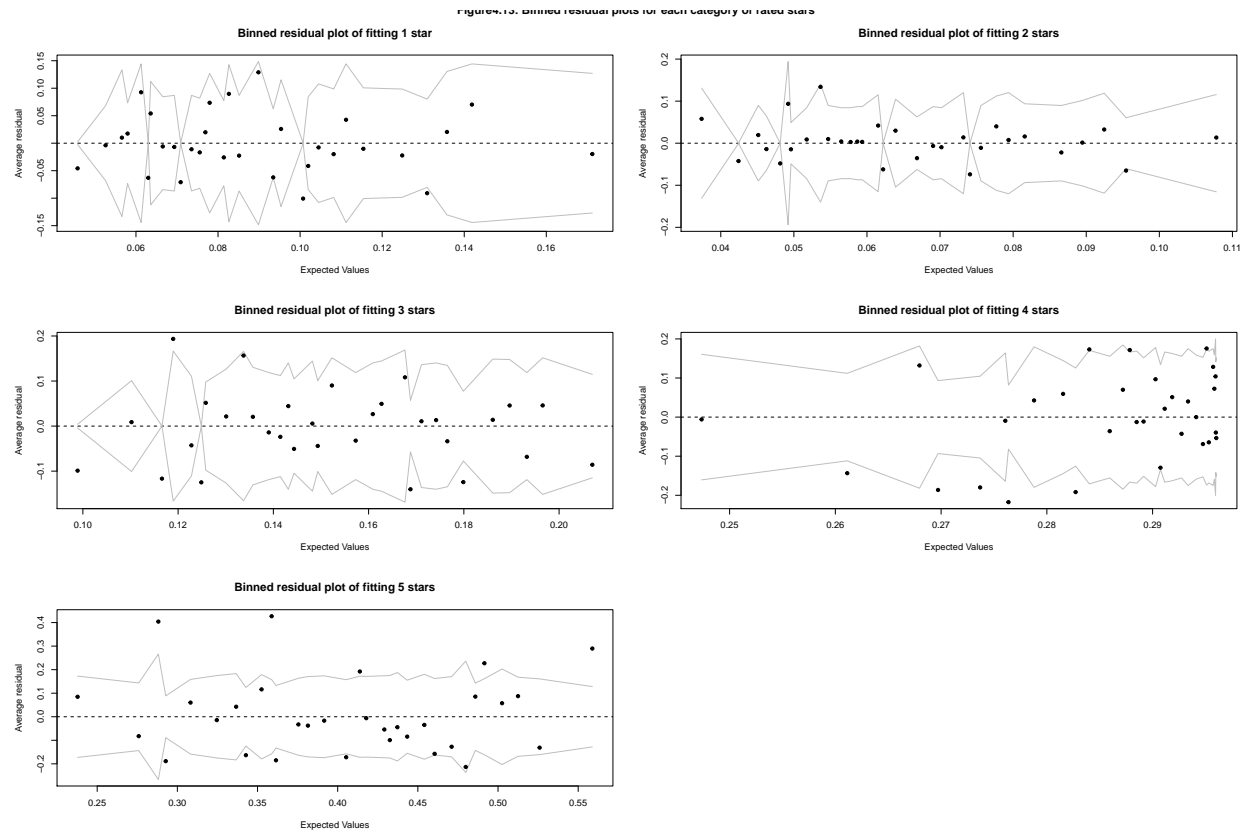
The random effects stand for the varying intercept of four styles of restaurants. Here is the table of four different intercepts. We can see that as we treat style as random effect, different styles of restaurants tend to have various intercept means different average ratings, Japanese tend to have the highest ratings since it has the highest random intercept 0.14, and 3.71 plus 0.14 which is the average stars of a Japanese restaurant to be 3.85 or 4 stars ratings.

4.2 Multinomial Logistic Regression with no random effects

I fit a cumulative logit model with the star of a user rating for a certain restaurant as response and the style, price range, noise level, parking conditions and wifi, TV and outdoor seatings of a restaurant as predictors. This model does not have random effects.

Model 7: Cumulative Logit Model

Check the residual of the model:



The residual plot of 1 star is relatively good since it has the smallest range and balanced while the 4 stars and 5 stars are kind of unbalanced and having bigger range which indicating the poor precision of model fitting.

Interpret coefficients:

Estimated Coefficients	
(Intercept):1	-1.8734112
(Intercept):2	-1.2409576
(Intercept):3	-0.3394474
(Intercept):4	0.8812186
restaurant_styleJapanese	-0.6233304
restaurant_styleKorean	-0.7409488
restaurant_styleSoutheast Asian	-0.3276963
restaurant_WiFino	0.1058993

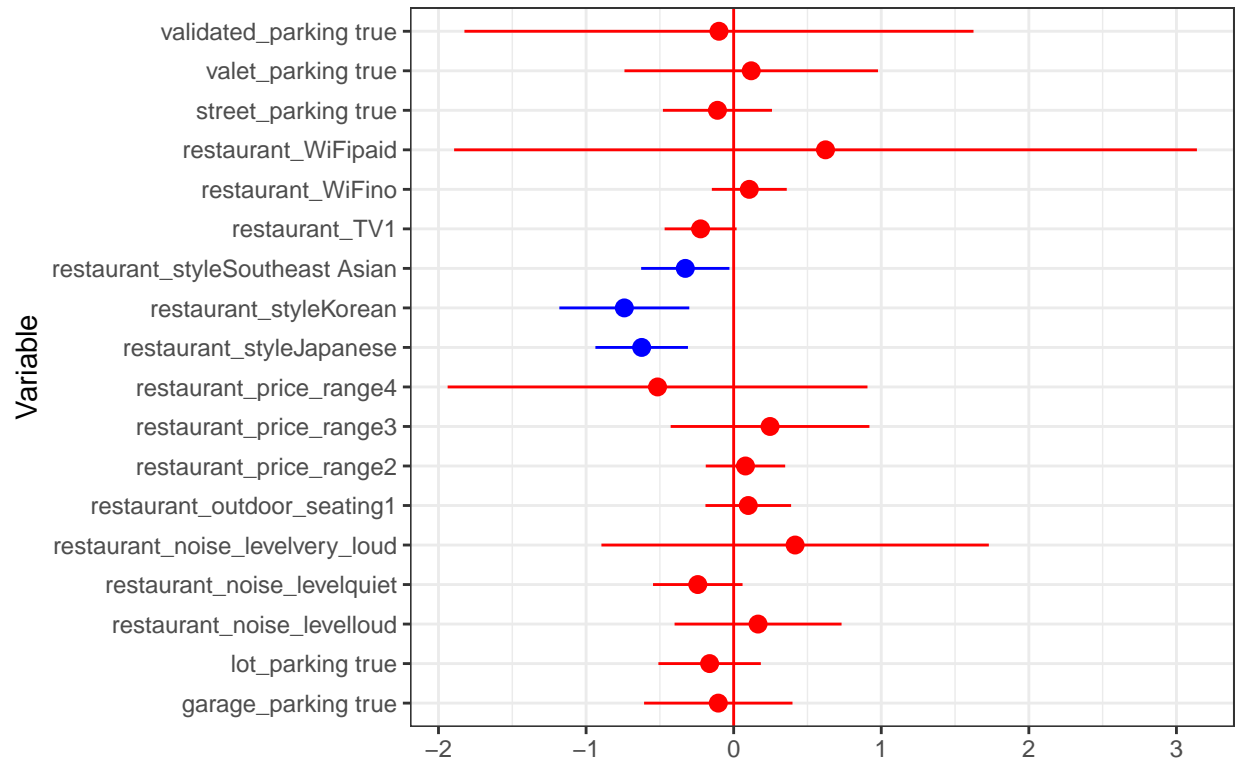
	Estimated Coefficients
restaurant_WiFipaid	0.6223533
restaurant_price_range2	0.0798676
restaurant_price_range3	0.2464221
restaurant_price_range4	-0.5157504
garage_parking true	-0.1042294
street_parking true	-0.1098878
validated_parking true	-0.0995466
lot_parking true	-0.1629981
valet_parking true	0.1188710
restaurant_noise_level loud	0.1652361
restaurant_noise_level quiet	-0.2432027
restaurant_noise_level very_loud	0.4163916
restaurant_TV1	-0.2240422
restaurant_outdoor_seating1	0.0989359

For Chinese restaurants has free WiFi, at price range1, no garage, street parking, parking lots, validated parking and valet parking, at average noise level, no TV and no outdoor seatings, the probability of getting 5 stars is:

$1 - \text{logit}^{-1}(0.88) = 0.29$, ###same calculation, we can calculate that the probabilities of Japanese, Korean and Southeast Asian restaurants under same circumstances getting 5 stars are: 0.44, 0.47 and 0.36. So the Korean restaurants tend to have higher probability to get 5 stars.

Check the coefficients' significance:

Figure4.14: Coefficients and 95% CI



Also only the coefficients of styles of restaurants are significant at level 5%, the rest of variables are not.

5. General Findings

As I've stated before, my goal is to find the relationship between the ratings and restaurants' informations, and I do have found something interesting.

- 1. Korean and Japanese restaurants have relatively same best ratings while Chinese restaurants have the worst ratings.
- 2. No WiFi may not lower the rating too much but a paid WiFi is going to lower the rating by almost 0.5 star.
- 3. Among the five parking informations, the validated parking and street parking tend to do better for the rating of restaurants, having both may help restaurants get 0.5 star higher rating.
- 4. It seems that the restaurants within price range 4 tend to have the best rating among 4 ranges. At beginning, I thought that inexpensiveness may lead to higher rating, but something beyond the cost of food may also have impact on ratings like decorations, giving customers first impressions.
- 5. For noise level, it seems that people are okay with loud noise but can't stand with very loud noise. Very loud noise level may lower the rating 0.27 star compared to the average-noise-level restaurants.
- 6. Having TV will only increase 0.12 or 0.13 star of rating which may not be an important effect. Personal guess, I think people may not have too much time watching TV via having dinner and sometimes a restaurant may only have one TV which can not be watched from all the directions for all customers, so it may not affect too much.
- 7. Whether a restaurant having or not having outdoor seating has tiny negative influence on the rating.

6. Further Discussion

6.1 Merits and demerits

The merit of linear regression is that it's easy to interpret. But for this topic, the star of rating is ordinal and has 5 levels, the best way is to fit the multilevel multinomial logit regression, so the demerit is the linear model may be not suitable to analyze since we only have 1,2,3,4,5 but we can calculate the 3.85 stars which will be confused. While the “MLML” is more advanced and difficult and it's harder to interpret and understand.

6.2 Further Direction: Try using `ordinal::clmm/clmm2` to fit a multinomial logistic regression with mixed effects.

(cont'd) **4.3 Multinomial Logistic Regression with mixed effects**

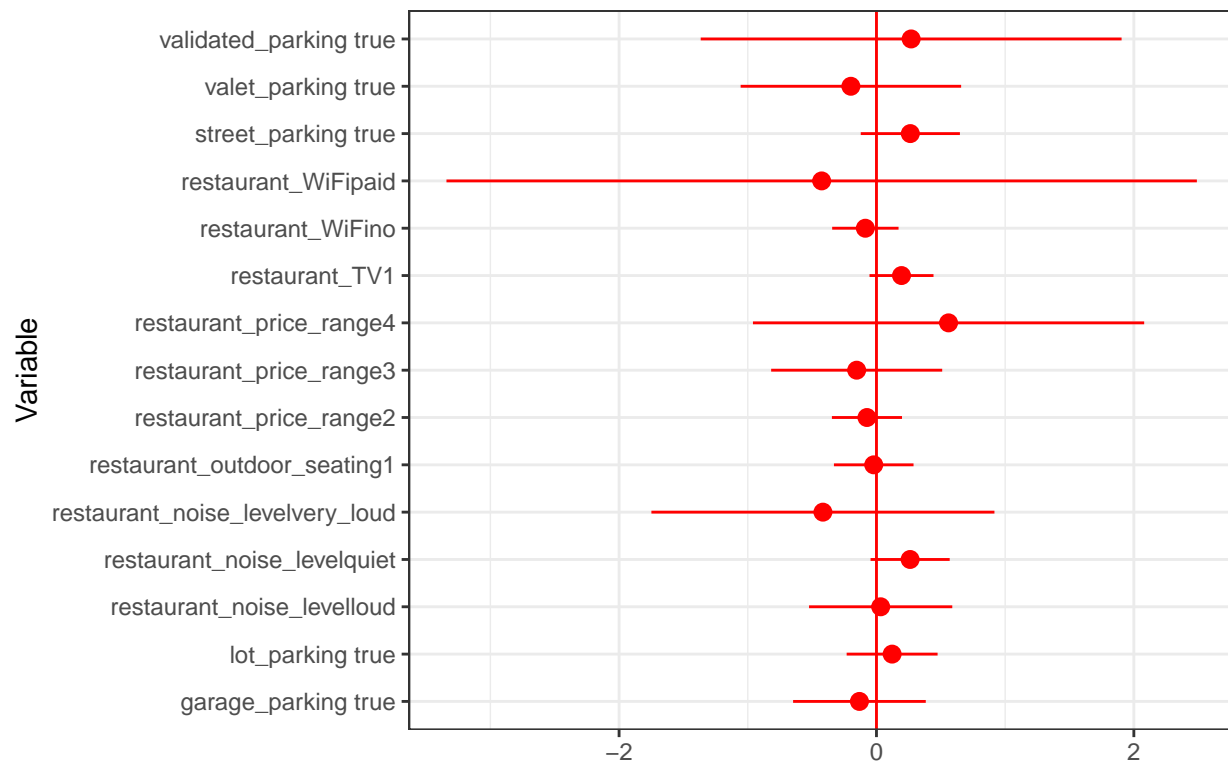
Model 8: Multilevel Multinomial Logistic Regression

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
1 2	-2.2614957	0.3040249	-7.4385221	0.0000000
2 3	-1.6260500	0.2963288	-5.4873169	0.0000000
3 4	-0.7167121	0.2912863	-2.4605078	0.0138741
4 5	0.5233803	0.2904777	1.8017914	0.0715782
restaurant_WiFino	-0.0863711	0.1315676	-0.6564772	0.5115171
restaurant_WiFipaid	-0.4257817	1.4874171	-0.2862557	0.7746822
restaurant_price_range2	-0.0739710	0.1389738	-0.5322660	0.5945418
restaurant_price_range3	-0.1537527	0.3390227	-0.4535173	0.6501763
restaurant_price_range4	0.5601749	0.7755068	0.7223340	0.4700891
garage_parking true	-0.1323355	0.2629993	-0.5031781	0.6148391
street_parking true	0.2630663	0.1965230	1.3386027	0.1807001
validated_parking true	0.2694941	0.8342584	0.3230343	0.7466693
lot_parking true	0.1216562	0.1802483	0.6749365	0.4997161
valet_parking true	-0.1988441	0.4371211	-0.4548947	0.6491850
restaurant_noise_level loud	0.0326908	0.2838893	0.1151533	0.9083236
restaurant_noise_level quiet	0.2615202	0.1568902	1.6668994	0.0955344
restaurant_noise_level very loud	-0.4160873	0.6800016	-0.6118917	0.5406094
restaurant_TV1	0.1946143	0.1268759	1.5338951	0.1250555
restaurant_outdoor_seating1	-0.0212759	0.1577436	-0.1348766	0.8927095

Check the coefficients' significance:

Figure4.15: Coefficients and 95% CI



The result also shows no variable is significant at level 5%.

Well, this part needs more time to learn, research and fit more types of model coping with 'Hess' and 'nAGQ' and understand the interpretation of the coefficients and varying intercepts of random effects.

Appendix:

Result of Model1:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## user_stars ~ restaurant_WiFi + restaurant_price_range + garage_parking +
##   street_parking + validated_parking + lot_parking + valet_parking +
##   restaurant_noise_level + restaurant_TV + restaurant_outdoor_seating +
##   (1 | restaurant_city:restaurant_state)
## Data: chinese
##
## REML criterion at convergence: 452290.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.4646 -0.5538  0.3092  0.9359  1.9417
##
## Random effects:
##      Groups                                Name          Variance Std.Dev.
## restaurant_city:restaurant_state (Intercept) 0.1033     0.3214
## Residual                                     1.8244     1.3507
## Number of obs: 131394, groups:  restaurant_city:restaurant_state, 190
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      3.460481   0.031447  110.04
## restaurant_WiFi    -0.074245   0.009681   -7.67
## restaurant_WiFipaid  0.391093   0.124722    3.14
## restaurant_price_range2 -0.076295   0.008556   -8.92
## restaurant_price_range3  0.241182   0.023064   10.46
## restaurant_price_range4  0.292546   0.056898    5.14
## garage_parking true    0.062425   0.016080    3.88
## street_parking true    0.236769   0.014132   16.75
## validated_parking true  0.318027   0.077457    4.11
## lot_parking true      0.169701   0.011287   15.03
## valet_parking true    -0.096321   0.029909   -3.22
## restaurant_noise_level loud -0.103661   0.016073   -6.45
## restaurant_noise_level quiet  0.070449   0.010912    6.46
## restaurant_noise_level very_loud -0.838387   0.035536  -23.59
## restaurant_TV1        0.081026   0.008341    9.71
## restaurant_outdoor_seating1 -0.116363   0.013031   -8.93
##
## Correlation matrix not shown by default, as p = 16 > 12.
## Use print(x, correlation=TRUE) or
##   vcov(x)      if you need it
```

Result of Model2:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## user_stars ~ restaurant_WiFi + restaurant_price_range + garage_parking +
```

```

##      street_parking + validated_parking + lot_parking + valet_parking +
##      restaurant_noise_level + restaurant_TV + restaurant_outdoor_seating +
##      (1 | restaurant_city:restaurant_state)
## Data: japanese
##
## REML criterion at convergence: 498840.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7861 -0.5712  0.2626  0.8048  2.0299
##
## Random effects:
##      Groups                                Name          Variance Std.Dev.
## restaurant_city:restaurant_state (Intercept) 0.1465     0.3828
## Residual                                     1.6089     1.2684
## Number of obs: 150429, groups:  restaurant_city:restaurant_state, 111
##
## Fixed effects:
##
##              Estimate Std. Error t value
## (Intercept)      3.637550   0.041999   86.61
## restaurant_WiFino      -0.125252   0.007699  -16.27
## restaurant_WiFipaid     -0.325194   0.070338   -4.62
## restaurant_price_range2  -0.017203   0.010670   -1.61
## restaurant_price_range3    0.196647   0.017109   11.49
## restaurant_price_range4    0.124945   0.023910    5.23
## garage_parking true      -0.129792   0.013559   -9.57
## street_parking true       0.136821   0.012295   11.13
## validated_parking true     0.137396   0.042405    3.24
## lot_parking true          0.032307   0.010401    3.11
## valet_parking true        -0.238354   0.017193  -13.86
## restaurant_noise_level loud -0.081484   0.018166   -4.49
## restaurant_noise_level quiet 0.182440   0.011740   15.54
## restaurant_noise_level very_loud 0.169690   0.028072    6.04
## restaurant_TV1           -0.048817   0.008459   -5.77
## restaurant_outdoor_seating1 -0.057049   0.009343   -6.11
##
## Correlation matrix not shown by default, as p = 16 > 12.
## Use print(x, correlation=TRUE) or
##      vcov(x)      if you need it

```

Result of Model3:

```

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## user_stars ~ restaurant_WiFi + restaurant_price_range + garage_parking +
##      street_parking + validated_parking + lot_parking + valet_parking +
##      restaurant_noise_level + restaurant_TV + restaurant_outdoor_seating +
##      (1 | restaurant_city:restaurant_state)
## Data: korean
##
## REML criterion at convergence: 124142.3
##
## Scaled residuals:

```

```

##      Min      1Q  Median      3Q      Max
## -3.0001 -0.4853  0.2527  0.7220  1.9299
##
## Random effects:
##      Groups                                Name      Variance Std.Dev.
## restaurant_city:restaurant_state (Intercept) 0.1203   0.3469
## Residual                                     1.4232   1.1930
## Number of obs: 38850, groups:  restaurant_city:restaurant_state, 41
##
## Fixed effects:
##                                     Estimate Std. Error t value
## (Intercept)                       4.363375   0.065494   66.62
## restaurant_WiFino                  -0.315933   0.015895  -19.88
## restaurant_WiFipaid                 0.017840   0.086046    0.21
## restaurant_price_range2            -0.281608   0.018035  -15.61
## restaurant_price_range3            -0.723019   0.091844   -7.87
## garage_parking true                 -0.690390   0.086335   -8.00
## street_parking true                 -0.142824   0.027089   -5.27
## validated_parking true              -0.330166   0.103130   -3.20
## lot_parking true                   -0.003439   0.023524   -0.15
## valet_parking true                  0.574760   0.075823    7.58
## restaurant_noise_level loud         -0.134862   0.026442   -5.10
## restaurant_noise_level quiet        0.022934   0.026623    0.86
## restaurant_noise_level very_loud   -0.523394   0.168415   -3.11
## restaurant_TV1                     -0.124490   0.016837   -7.39
## restaurant_outdoor_seating1         0.034246   0.023349    1.47
##
## Correlation matrix not shown by default, as p = 15 > 12.
## Use print(x, correlation=TRUE) or
## vcov(x) if you need it

```

Result of Model4:

```

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## user_stars ~ restaurant_WiFi + restaurant_price_range + garage_parking +
## street_parking + validated_parking + lot_parking + valet_parking +
## restaurant_noise_level + restaurant_TV + restaurant_outdoor_seating +
## (1 | restaurant_city:restaurant_state)
## Data: se_asian
##
## REML criterion at convergence: 501276.6
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -2.7109 -0.5992  0.1649  0.8425  1.8204
##
## Random effects:
##      Groups                                Name      Variance Std.Dev.
## restaurant_city:restaurant_state (Intercept) 0.09686   0.3112
## Residual                                     1.57755   1.2560
## Number of obs: 152068, groups:  restaurant_city:restaurant_state, 140
##

```

```
## Fixed effects:
##
##              Estimate Std. Error t value
## (Intercept)      3.780289   0.032831  115.14
## restaurant_WiFino -0.000177   0.007656   -0.02
## restaurant_WiFipaid -0.084038   0.052226   -1.61
## restaurant_price_range2  0.016849   0.007487    2.25
## restaurant_price_range3  0.278611   0.053409    5.22
## restaurant_price_range4  0.579079   0.563010    1.03
## garage_parking true -0.280772   0.020810  -13.49
## street_parking true  0.155595   0.011610   13.40
## validated_parking true  0.221032   0.063286    3.49
## lot_parking true    0.106394   0.010640   10.00
## valet_parking true   -0.205557   0.030688   -6.70
## restaurant_noise_level loud  0.014001   0.020605    0.68
## restaurant_noise_level quiet 0.011927   0.008027    1.49
## restaurant_noise_level very_loud -0.739897   0.072382  -10.22
## restaurant_TV1      -0.025635   0.007364   -3.48
## restaurant_outdoor_seating1 -0.003726   0.008817   -0.42
##
## Correlation matrix not shown by default, as p = 16 > 12.
## Use print(x, correlation=TRUE) or
##   vcov(x)      if you need it
```

Result of Model5:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## user_stars ~ restaurant_WiFi + restaurant_price_range + garage_parking +
##   street_parking + validated_parking + lot_parking + valet_parking +
##   restaurant_noise_level + restaurant_TV + restaurant_outdoor_seating +
##   restaurant_style + (1 | restaurant_city:restaurant_state)
## Data: mysample
##
## REML criterion at convergence: 3307.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.6753 -0.5996  0.2363  0.7812  1.5070
##
## Random effects:
##   Groups                                Name      Variance Std.Dev.
## restaurant_city:restaurant_state (Intercept) 0.05574  0.2361
## Residual                                     1.54752  1.2440
## Number of obs: 1000, groups:  restaurant_city:restaurant_state, 79
##
## Fixed effects:
##
##              Estimate Std. Error t value
## (Intercept)      3.45809   0.16970  20.377
## restaurant_WiFino -0.02480   0.08879  -0.279
## restaurant_WiFipaid -0.46119   0.89496  -0.515
## restaurant_price_range2 -0.09627   0.09468  -1.017
## restaurant_price_range3 -0.12646   0.23823  -0.531
## restaurant_price_range4  0.07588   0.46694   0.163
```

```
## garage_parking true          -0.06761    0.18058   -0.374
## street_parking true          0.16802    0.14034    1.197
## validated_parking true       0.31757    0.59263    0.536
## lot_parking true             0.06995    0.12406    0.564
## valet_parking true           -0.12594    0.30510   -0.413
## restaurant_noise_levelcloud  0.07642    0.20268    0.377
## restaurant_noise_levelquiet  0.17451    0.10718    1.628
## restaurant_noise_levelvery_loud -0.26853    0.46880   -0.573
## restaurant_TV1               0.12184    0.08656    1.408
## restaurant_outdoor_seating1 -0.04629    0.10544   -0.439
## restaurant_styleJapanese     0.42347    0.11039    3.836
## restaurant_styleKorean       0.45194    0.15226    2.968
## restaurant_styleSoutheast Asian 0.24780    0.10560    2.347

##
## Correlation matrix not shown by default, as p = 19 > 12.
## Use print(x, correlation=TRUE) or
##   vcov(x)       if you need it
```

Result of Model6:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## user_stars ~ restaurant_WiFi + restaurant_price_range + garage_parking +
##   street_parking + validated_parking + lot_parking + valet_parking +
##   restaurant_noise_level + restaurant_TV + restaurant_outdoor_seating +
##   (1 | restaurant_style) + (1 | restaurant_city:restaurant_state)
##   Data: mysample
##
## REML criterion at convergence: 3307.7
##
## Scaled residuals:
##   Min      1Q  Median      3Q      Max
## -2.6710 -0.5987  0.2190  0.7918  1.4937
##
## Random effects:
##   Groups                                Name          Variance Std.Dev.
## restaurant_city:restaurant_state (Intercept) 0.05452  0.2335
## restaurant_style                  (Intercept) 0.03486  0.1867
## Residual                          1.54794  1.2442
## Number of obs: 1000, groups:
## restaurant_city:restaurant_state, 79; restaurant_style, 4
##
## Fixed effects:
##                                Estimate Std. Error t value
## (Intercept)                   3.71060    0.19201  19.325
## restaurant_WiFi                -0.02563    0.08878  -0.289
## restaurant_WiFipaid            -0.43774    0.89468  -0.489
## restaurant_price_range2        -0.07993    0.09415  -0.849
## restaurant_price_range3        -0.11277    0.23757  -0.475
## restaurant_price_range4         0.11536    0.46584   0.248
## garage_parking true            -0.08029    0.18034  -0.445
## street_parking true             0.17430    0.14010   1.244
## validated_parking true          0.34124    0.59244   0.576
```

```
## lot_parking true          0.06855    0.12393    0.553
## valet_parking true        -0.12267    0.30503   -0.402
## restaurant_noise_level loud 0.07349    0.20235    0.363
## restaurant_noise_level quiet 0.16969    0.10682    1.589
## restaurant_noise_level very_loud -0.28243    0.46839   -0.603
## restaurant_TV1           0.13187    0.08627    1.529
## restaurant_outdoor_seating1 -0.04315    0.10532   -0.410

##
## Correlation matrix not shown by default, as p = 16 > 12.
## Use print(x, correlation=TRUE) or
##   vcov(x)       if you need it
```

Result of Model7:

```
##
## Call:
## vglm(formula = ordered(user_stars) ~ restaurant_style + restaurant_WiFi +
##   restaurant_price_range + garage_parking + street_parking +
##   validated_parking + lot_parking + valet_parking + restaurant_noise_level +
##   restaurant_TV + restaurant_outdoor_seating, family = cumulative(parallel = T),
##   data = mysample)
##
##
## Pearson residuals:
##           Min       1Q   Median       3Q      Max
## logit(P[Y<=1]) -1.994 -0.2111 -0.1547 -0.1266  4.019
## logit(P[Y<=2]) -1.232 -0.2929 -0.1982 -0.1594  4.461
## logit(P[Y<=3]) -1.217 -0.8593 -0.3055  0.4764  2.602
## logit(P[Y<=4]) -2.214 -1.0235  0.2831  1.0341  1.606
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1      -1.87341    0.25000  -7.494 6.70e-14 ***
## (Intercept):2      -1.24096    0.24117  -5.146 2.67e-07 ***
## (Intercept):3      -0.33945    0.23644  -1.436 0.151095
## (Intercept):4       0.88122    0.23796   3.703 0.000213 ***
## restaurant_styleJapanese -0.62333    0.16002  -3.895 9.81e-05 ***
## restaurant_styleKorean  -0.74095    0.22457  -3.299 0.000969 ***
## restaurant_styleSoutheast Asian -0.32770    0.15260  -2.147 0.031758 *
## restaurant_WiFino      0.10590    0.12951   0.818 0.413522
## restaurant_WiFipaid    0.62235    1.28393   0.485 0.627871
## restaurant_price_range2  0.07987    0.13729   0.582 0.560744
## restaurant_price_range3  0.24642    0.34355   0.717 0.473194
## restaurant_price_range4 -0.51575    0.72567  -0.711 0.477255
## garage_parking true    -0.10423    0.25650  -0.406 0.684484
## street_parking true    -0.10989    0.18850  -0.583 0.559923
## validated_parking true  -0.09955    0.87993  -0.113 0.909927
## lot_parking true      -0.16300    0.17704  -0.921 0.357224
## valet_parking true     0.11887    0.43820   0.271 0.786182
## restaurant_noise_level loud 0.16524    0.28865   0.572 0.567022
## restaurant_noise_level quiet -0.24320    0.15473  -1.572 0.116001
## restaurant_noise_level very_loud 0.41639    0.66966   0.622 0.534077
## restaurant_TV1        -0.22404    0.12438  -1.801 0.071667 .
```



```

## restaurant_outdoor_seating1      0.09894      0.14806      0.668 0.504008
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of linear predictors: 4
##
## Names of linear predictors:
## logit(P[Y<=1]), logit(P[Y<=2]), logit(P[Y<=3]), logit(P[Y<=4])
##
## Residual deviance: 2782.511 on 3978 degrees of freedom
##
## Log-likelihood: -1391.256 on 3978 degrees of freedom
##
## Number of iterations: 5
##
## No Hauck-Donner effect found in any of the estimates
##
## Exponentiated coefficients:
##      restaurant_styleJapanese      restaurant_styleKorean
##      0.5361559                      0.4766614
## restaurant_styleSoutheast Asian      restaurant_WiFino
##      0.7205819                      1.1117099
##      restaurant_WiFipaid      restaurant_price_range2
##      1.8633078                      1.0831436
##      restaurant_price_range3      restaurant_price_range4
##      1.2794396                      0.5970524
##      garage_parking true      street_parking true
##      0.9010186                      0.8959346
##      validated_parking true      lot_parking true
##      0.9052477                      0.8495928
##      valet_parking true      restaurant_noise_levelquiet
##      1.1262246                      1.1796716
##      restaurant_noise_levelquiet restaurant_noise_levelvery_loud
##      0.7841125                      1.5164795
##      restaurant_TV1      restaurant_outdoor_seating1
##      0.7992814                      1.1039955

```

Result of Model8:

```

## Cumulative Link Mixed Model fitted with the Laplace approximation
##
## formula: ordered(user_stars) ~ restaurant_WiFi + restaurant_price_range +
##      garage_parking + street_parking + validated_parking + lot_parking +
##      valet_parking + restaurant_noise_level + restaurant_TV +
##      restaurant_outdoor_seating + (1 | restaurant_style) + (1 |
##      restaurant_state)
## data:      mysample
##
## link threshold nobs logLik  AIC      niter      max.grad cond.H
## logit flexible 1000 -1389.18 2820.36 2892(6105) 2.26e-03 1.4e+03
##
## Random effects:
## Groups          Name          Variance Std.Dev.

```

```

## restaurant_state (Intercept) 0.07924 0.2815
## restaurant_style (Intercept) 0.06517 0.2553
## Number of groups: restaurant_state 12, restaurant_style 4
##
## Coefficients:
##
## Estimate Std. Error z value Pr(>|z|)
## restaurant_WiFino -0.08637 0.13157 -0.656 0.5115
## restaurant_WiFipaid -0.42578 1.48742 -0.286 0.7747
## restaurant_price_range2 -0.07397 0.13897 -0.532 0.5945
## restaurant_price_range3 -0.15375 0.33902 -0.454 0.6502
## restaurant_price_range4 0.56017 0.77551 0.722 0.4701
## garage_parking true -0.13234 0.26300 -0.503 0.6148
## street_parking true 0.26307 0.19652 1.339 0.1807
## validated_parking true 0.26949 0.83426 0.323 0.7467
## lot_parking true 0.12166 0.18025 0.675 0.4997
## valet_parking true -0.19884 0.43712 -0.455 0.6492
## restaurant_noise_level loud 0.03269 0.28389 0.115 0.9083
## restaurant_noise_level quiet 0.26152 0.15689 1.667 0.0955
## restaurant_noise_level very_loud -0.41609 0.68000 -0.612 0.5406
## restaurant_TV1 0.19461 0.12688 1.534 0.1251
## restaurant_outdoor_seating1 -0.02128 0.15774 -0.135 0.8927
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
## Estimate Std. Error z value
## 1|2 -2.2615 0.3040 -7.439
## 2|3 -1.6260 0.2963 -5.487
## 3|4 -0.7167 0.2913 -2.461
## 4|5 0.5234 0.2905 1.802

```