Statistics 133 Project Report:
Determinants of Restaurant Pricing and Regional Differentiations of Tastes in China

Group 13 Members: Yuqing Jiang, Zhengyun Sun, Jiangzhen Yu, Lisa Ann Yu

1. Introduction

Nowadays, people rely on websites such as Yelp.com to choose a restaurant. As a result of crowdsourcing of restaurant reviews and ratings, individuals' preferences accumulate to form a database on overall tastes for various cuisines across different regions. Therefore, it is revealing to explore such regional differentiations of tastes, where geographic location could serve as representatives of characteristics of the cities. With extreme diversity and regional variations, China serves as an exemplary case in exploring how preferred tastes differ across regions.

In this project, we analyze data collected from Yelp's Chinese counterpart, dianping.com, to examine the following questions: (1) What factors determine the pricing of a restaurant? (2) How do the preferences for major cuisines differ across Chinese provincial capitals? In order to answer these questions, we look at the ratings and pricing of restaurants in each provincial capital that serve Chuan, Yue, or local cuisine. Chuan and Yue cuisines are the two most widely recognized and served cuisines among all varieties of Chinese cooking. Chuan cuisine, which originated from Chengdu in the southwestern part of China, uses common ingredients and flavorful spices such as hot peppers. Yue cuisine, on the other hand, uses seafood and aims to preserve the original flavors of the ingredients. In common conception, Chuan cuisine represents everyday, common tastes, while Yue cuisine is associated with high-end connoisseur preferences. The local cuisines differ significantly across regions and in our analyses are mostly used as benchmarks for comparison. Although China exhibits great regional diversity, we could still use geographic locations as broad indicators for characteristics such as levels of economic development, size of immigrant populations and degree of openness.

In answering the questions proposed, we first analyzed the degree to which the flavor ratings and environment serve as predictors of restaurant prices. Second, we explored how the popularity of Chuan and Yue cuisines vary across Chinese cities. To quantify the geographic locations, we study the relationship between the distances of a capital city to the origin of a cuisine and the ratings of the restaurants serving the cuisine. That is, for each provincial capital, we compared the ratings of Chuan and Yue style restaurants, using local cuisine ratings as a benchmark. In doing so, we could answer questions such as: how do Beijingers rate Chuan, Yue and local cuisine? Beijing is further away from Chengdu (origin of Chuan cuisine) than Shanghai: do Beijingers like Chuan cuisine better than Shanghainese? To visualize the results, we use scatterplots with fitted regression lines, boxplots, and maps to characterize the relations.

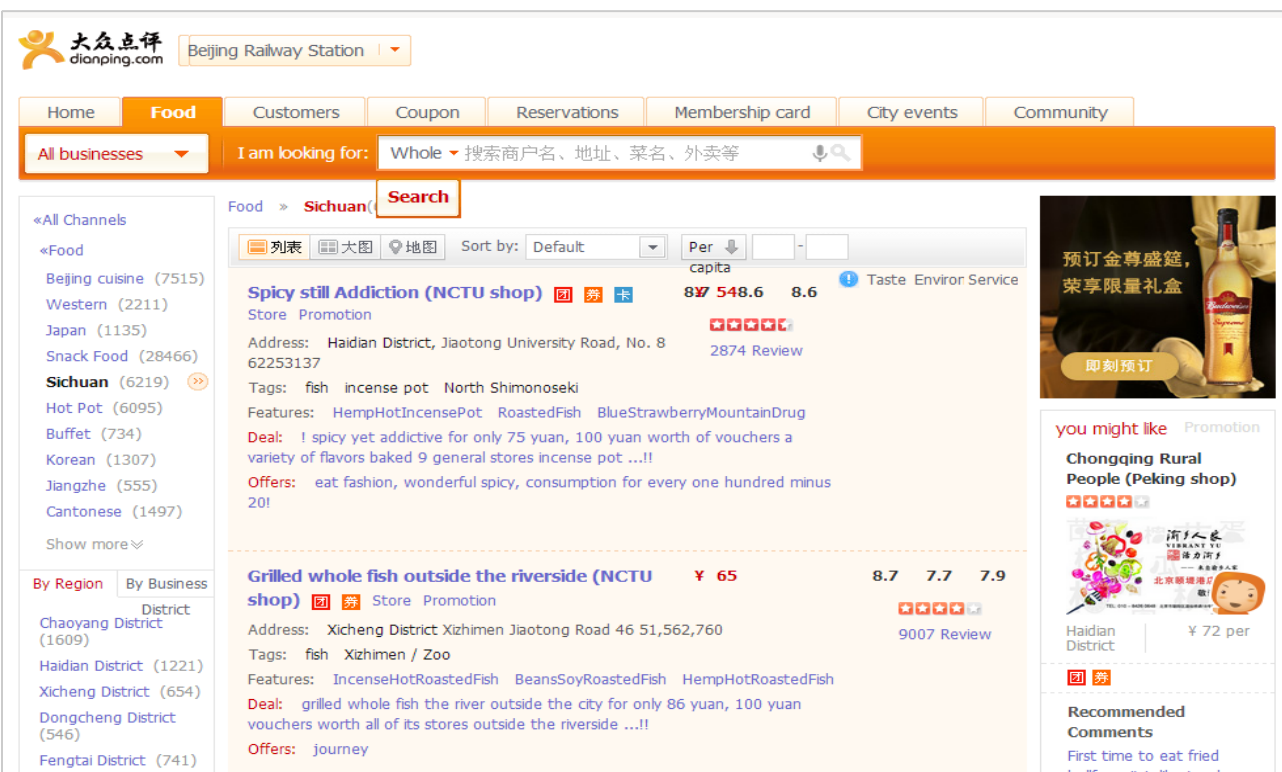
In addition to the distance variable, we also group the cities by regions, namely as southern, northern and western. The three regions have very different socioeconomic characteristics and local cultures. Southern cities have the most advanced economies and attract large numbers of immigrants from inland provinces. With regard to local cuisine taste, southern cuisines are very similar to Yue cuisine in terms of mild taste and preservations of the original ingredient flavors. Northern cuisines, on the other hand, have much stronger and more flavorful tastes that resemble Chuan style. The western cities are more culturally diverse, due to the high proportions of ethnic minorities, including a Muslim population. Such divergence from mainstream Han culture reflects in the western cities' food preferences. Given these distinctions, it is informative to consider how the popularity of Yue and Chuan cuisines vary across the three regions.

If we consider cuisine as a representative part of a culture, its level of acceptance reveals the popularity of the particular local culture associated with the cuisine. In addition, the degree to which a city accepts exotic cuisines may also hint at the openness of the city to outside influence, culture, and perhaps, immigrants. In this sense, the results we discovered on the regional preferences for cuisines might be revealing in terms of uncovering some cultural characteristics of the provincial capitals in China.

2. Data collection

In order to get data on restaurants reviews in the 31 provincial capitals in China, we did web scraping from dianping.com. The website dianping.com is the Chinese version of Yelp.com, on which there are listings of restaurants and many other businesses in major cities in China. The website has a large, booming user base in nearly every provincial capital. For each restaurant, the website provides information on its category, ratings, average per person prices, and users' reviews. Similar to Yelp, the ratings and price information are evaluated based on individual users' feedbacks. The major advantage of dianping.com is that each restaurant is rated by quality of the food (flavor), environment and service on a scale of 1 to 10. With such detailed ratings, we could get a more comprehensive picture of the restaurants that serve different cuisine. In addition, the average price is also quantified as a number rather than a price range. Hence, with these divisions, we could explore the relationship between ratings on food quality, environment and price. Figure 1 displays a Google-translated sample webpage of Chuan style cuisine restaurants in Beijing.

Specifically, we look at restaurants that serve Sichuan (Chuan), Guangdong (Yue) and local cuisine in each city. Chuan and Yue cuisine are the two most popular and widely served styles of cooking in China, yet the two have very different features that could effectively represent nearly



opposing preferences on food. Local cuisine differs significantly across the country and the information is collected mainly to serve as a benchmark for comparison. For each restaurant, we collected data on average price, flavor rating, environment rating, and number of reviews.

To make the sample representative, we chose sample sizes based on the total number of restaurants of each style in each city by criteria in Table 1 in the appendix. Using this standard, Table 2 in the appendix shows the number of pages scraped for each type of restaurants in each city (Part 1 of the R Code section in the Appendix). Second, we exclude all the restaurants with fewer than 5 reviews or with any of the variables missing to ensure accuracy. After the above procedures, we get a data set with 7822 observations. Table 3 displays a sample of our final restaurant data set. As shown, each observation corresponds to a restaurant and contains the information of a unique identifier (ID), city, category (Chuan, Yue or Local), rating of flavor, rating of environment, average price and number of reviews.

In addition to the information of the restaurants, we also collected the geographic coordinates of each provincial capital (Part 3 of the R Code section in the Appendix). Using this information, we calculated the distance from each city to the origin cities of Chuan and Yue cuisine, namely, Chengdu and Guangzhou respectively. Physical distance in some sense could serve as a proxy for the degree of differences in taste, preference and culture in general. We recognize there are many other characteristics of the cities that could potentially be used to categorize them. Therefore, we also include some categorical/dummy variables in the analyses.

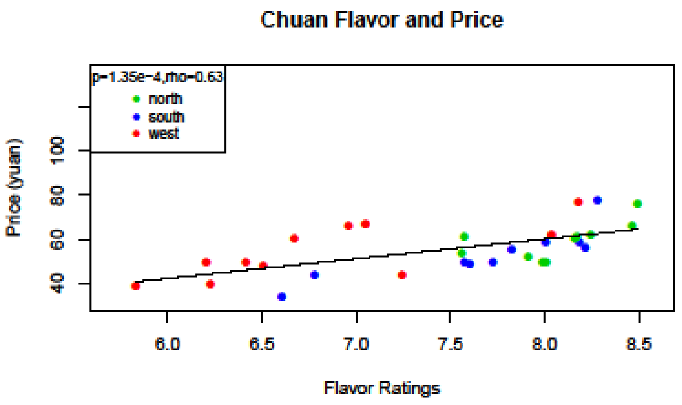
To produce the final dataset for analysis, we calculated the weighted average for ratings on flavor, environment and pricing for each city-cuisine pair, weighted by number of reviews (Part 2 of the R Code section in the Appendix). Since number of reviews represents the popularity of a restaurant in the city, we want to allocate more weight to the more popular restaurants that are better representatives of the city's preferences. In addition, we also calculated the distance between each city and the two origin cities of Chuan and Yue cuisine to get the distance variables. As for categorization, we visually grouped the cities into northern, western, or southern regions based on their geographic locations (Table 4 in appendix).

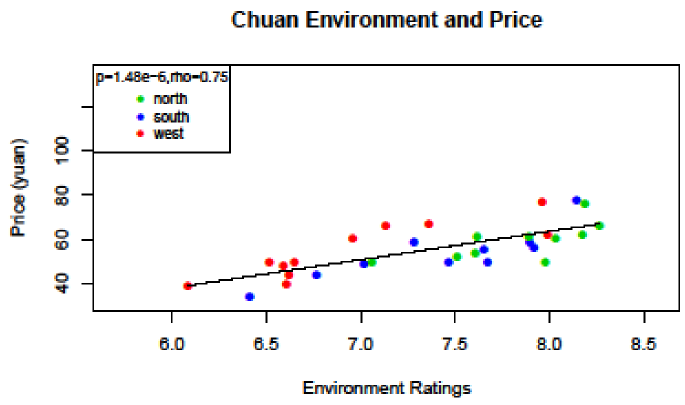
3. Analysis and Data visualization

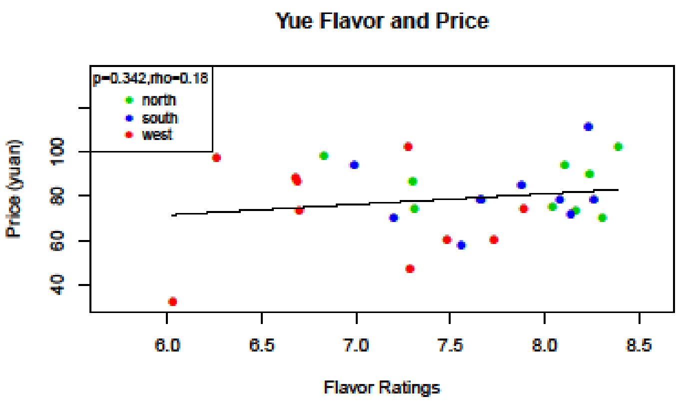
3.1 Determinants of Pricing

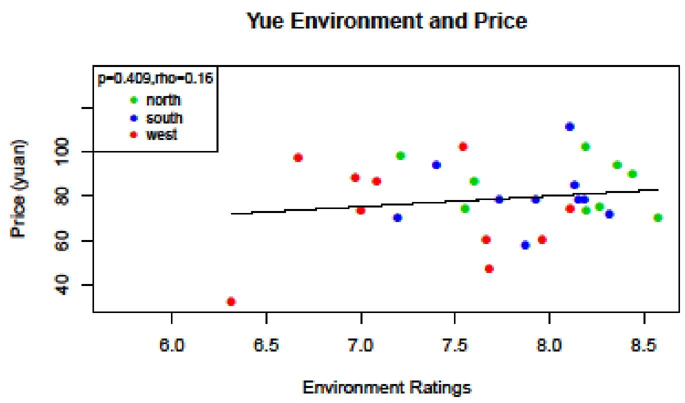
What goes into a restaurant's price? We have ratings for flavor and environment for each restaurant. Theoretically, as both ratings increase, the price should increase. However, which affects the price more? To test this, we compared the correlations between flavor ratings and the price of the restaurant and environment ratings and the price of the restaurant. Quantitatively, we looked at ρ, the correlation coefficient for the regression, how closely the data points followed the line; the p value, how likely they followed the line by random chance alone; and the slope of the regression line, how much the price increased as the flavor and environment increased.

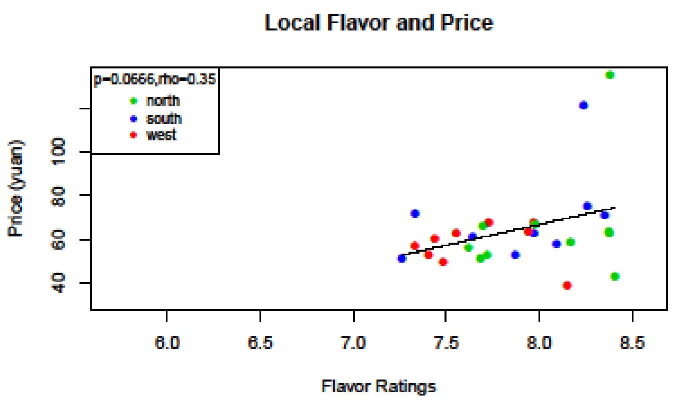
For every cuisine type, environment is more strongly correlated with price than flavor is, in terms of both ρ and the p value, except for ρ for Yue (see Table 5 in the Appendix). However, in that case, flavor has a ρ of 0.18 and environment has a ρ of 0.16, both of which are so weak that the difference of 0.02 hardly matters. Intuitively, it makes sense that environment is a better predictor of price than flavor is, for both cheap and expensive restaurants might taste good to its customers, but the environment and setting depends more on the price. One would expect to pay more at a sit-down restaurant with waiters and servers than at a fast-food place where customers stand in line to order food and then pick it up when their number is called. Customers really pay more for the experience of being served at fancy restaurants than the quality of food.

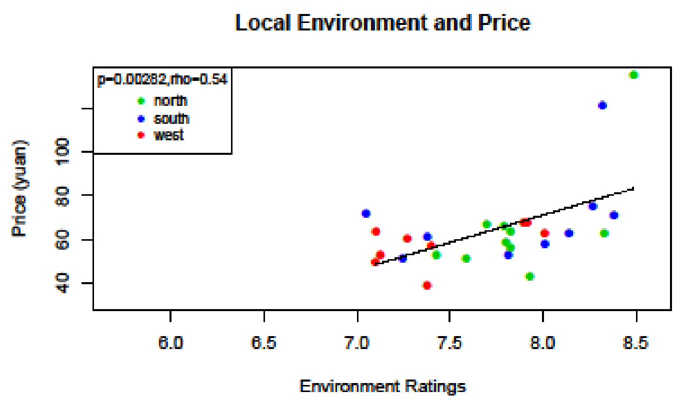

Chuan Flavor and Price
p=1.35e−4,rho=0.63
north
south
west
Price (yuan)
Flavor Ratings

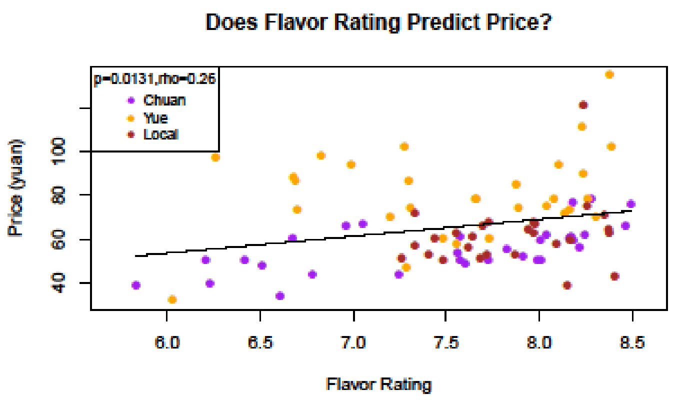

Chuan Environment and Price
p=1.48e−6,rho=0.75
north
south
west
Price (yuan)
Environment Ratings

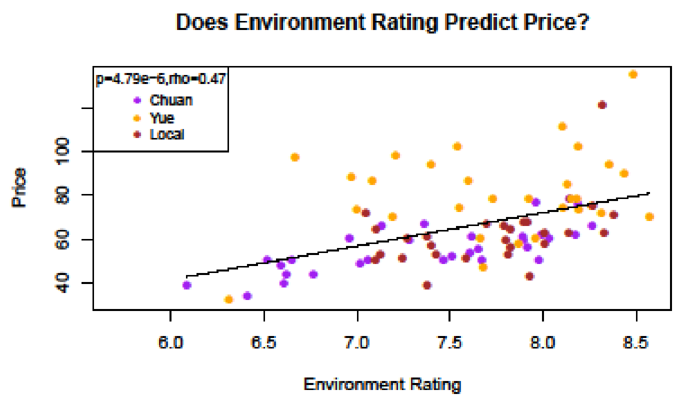

Yue Flavor and Price
p=0.342,rho=0.18
north
south
west
Price (yuan)
Flavor Ratings


Yue Environment and Price
p=0.409,rho=0.16
north
south
west
Price (yuan)
Environment Ratings


Local Flavor and Price
p=0.0666,rho=0.35
north
south
west
Price (yuan)
Flavor Ratings


Local Environment and Price
p=0.00282,rho=0.54
north
south
west
Price (yuan)
Environment Ratings


Does Flavor Rating Predict Price?
p=0.0131,rho=0.26
Chuan
Yue
Local
Price (yuan)
Flavor Rating


Does Environment Rating Predict Price?
p=4.79e−6,rho=0.47
Chuan
Yue
Local
Price
Environment Rating

We removed an outlier on the Yue graph in which a low flavor rating of 5.7 and low environment rating of 6.2 had a price of 167, well above any other restaurant. Including that data point in the data set for regression actually resulted in a negative regression line for both flavor and environment.

Even without that outlier, the regression between both flavor and environment and price was not significant for Yue cuisine, though it was for every other type of cuisine, except for local cuisine's flavor vs. price regression, which had a p of 0.0666. Compared to local and Chuan cuisine, Yue's prices seemed much more spread out. Both correlation coefficients were small, 0.18 and 0.16, which is close to there being no correlation at all. Yue cuisine is the Chinese equivalent of French cuisine in America, considered high-class and expected to have high quality of food and a nice setting. The prices are higher than Chuan and local cuisines as seen on the graph. Perhaps there is little variation in Yue cuisine restaurants since they are all expected to be high-class, so the resulting ratings for flavor and environment are left to the whim of the customer.

Another method of analyzing how flavor and price are related compared to how environment and price are related is by looking at the slopes of the regression lines. For all three cuisine types, environment had a steeper slope than flavor did, suggesting that the same increase in environment rating as flavor rating resulted in a larger increase in price for each environment plot compared to flavor plot. The local cuisine had the steepest regression slopes (19.06 yuan/rating unit for flavor and 25.19 for environment, followed by Chuan (8.899 for flavor and 12.81 for environment), then Yue (4.828 for flavor and 4.835 for environment) style cuisines for both flavor and environment ratings. The local cuisine had a smaller range of both flavor and environment ratings, but about the same price range as Chuan and Yue cuisines, with the exception of two high prices in Beijing and Shanghai. Therefore, the slope of the regression line for local cuisine had to be steep in order to compensate for the small range of the ratings.

As with any correlation plot, it is impossible to know whether increased tastiness of food leads to higher prices or higher prices lead to tastier food. Likewise, it is also impossible to know whether a better environment leads to higher prices or vice versa. These plots are intended to show only the relation between each pair of variables.

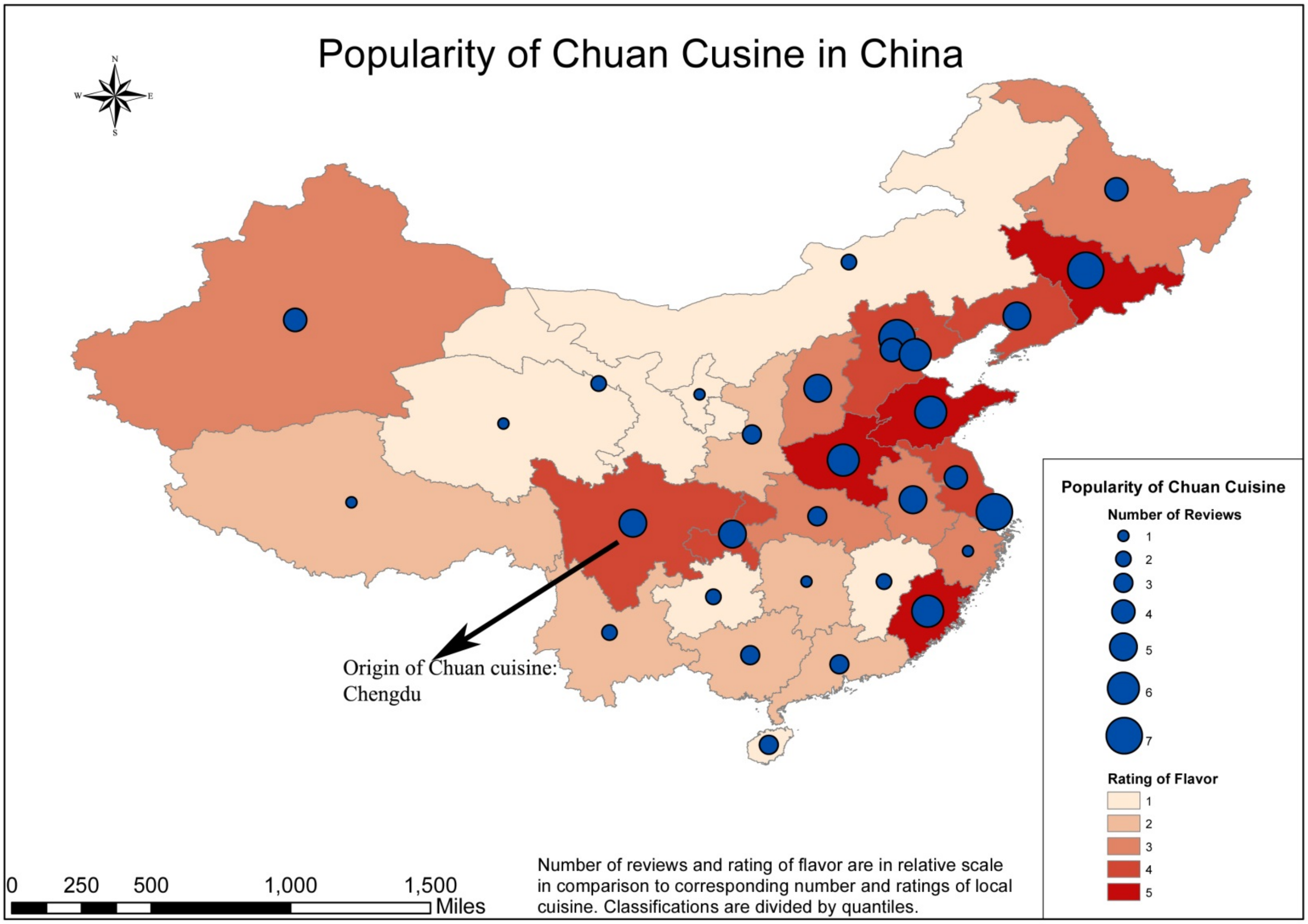
3.2 Variations of Popularity

1) Maps of popularity

The two maps illustrate the popularity of Sichuan and Yue cuisine in China based on average number of reviews and rating on flavor in relative scale. For average number of reviews and ratings, the relative measures are defined as follows:

$$comp.reviews = \frac{avg\ number\ of\ reviews_{Chuan/Yue} - avg\ number\ of\ reviews_{Local}}{avg\ number\ of\ reviews_{Local}}$$

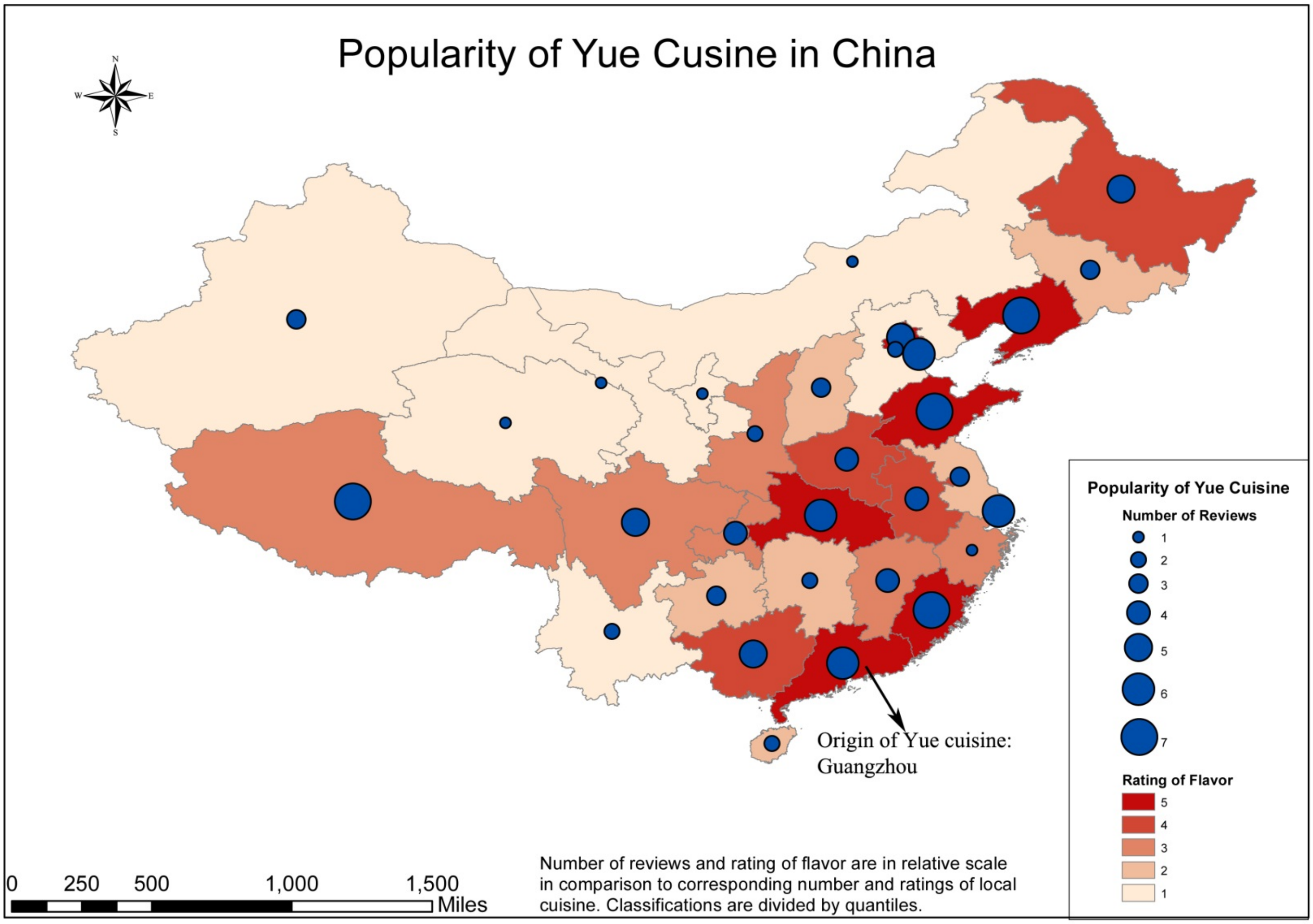
$$comp.flavor.rating = \frac{avg\ rating\ on\ flavor_{Chuan/Yue} - avg\ rating\ on\ flavor_{Local}}{avg\ rating\ on\ flavor_{Local}}$$



Notes: The scale shown here are in relative scare in comparison to the corresponding number of reviews and ratings of local cuisine, see equations follow for detailed. Specific scales 1 to 5 and 1 to 7 represent percentiles in the relative scale. e.g. number of reviews equal 1 if the city's relative number of reviews of Chuan cuisine is in the bottom 14 percent of all cities' relative number of reviews; rating of flavor equals 1 if the city's relative rating of Chuan restaurant is in the bottom 20 percent of all cities' relative ratings of the cuisine.

By converting the average number of reviews and flavor rating into a ratio between Chuan/Yue cuisine popularity measures and local cuisine popularity measures, we eliminate the differences among the cities. On the maps, color shade represents relative ratings on flavor and dot size shows relative number of reviews, to the nearest integer. Because the cities we choose are all provincial capitals and the provincial capitals are the largest cities in each province, preferences of the capitals could be reasonable indicators for the preferences of the provinces.

Now let us turn to the details of the maps. First, the two measures of popularity often go hand in hand. That is, users from cities that rate the cuisine higher in terms of the flavor of the restaurants also tend to write more reviews for those restaurants. Second, cities that show limited interest in one cuisine compared to its local food tend to have rather limited interest in the other non-local cuisine as well. That is, if a city does not like Chuan cuisine, it also has limited passion for Yue cuisine. Such cities include Hohhot in Inner Mongolia, Xining in Qinghai, Lanzhou in Gansu and Yinchuan in Ningxia. These four cities have relatively high representation of ethnic minority population, in particular Muslim population, and are located in the northwestern inland

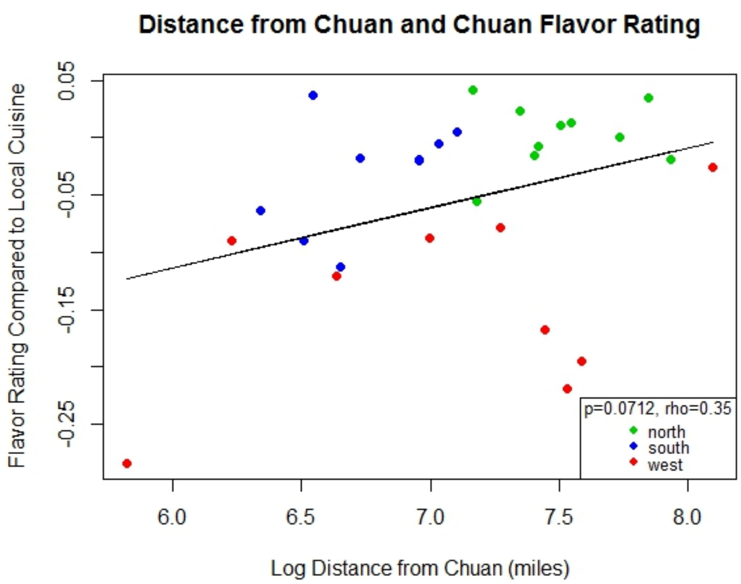


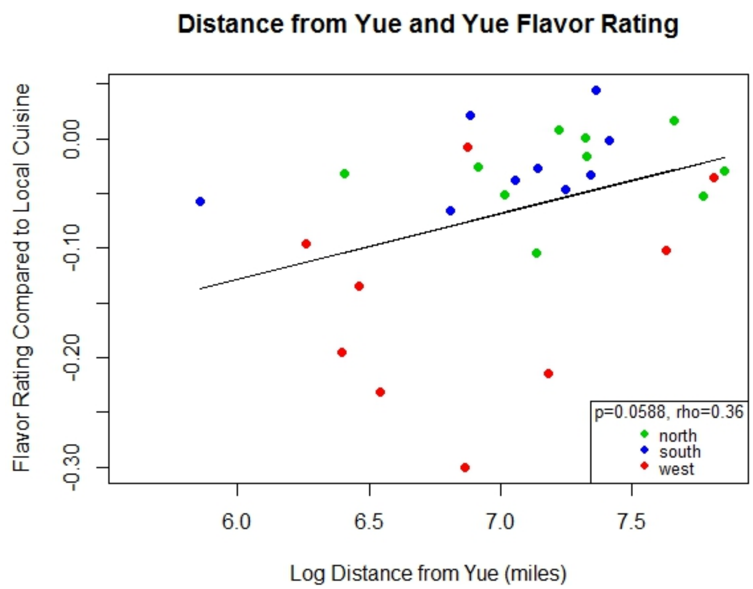
Notes: The scale shown here are in relative scare in comparison to the corresponding number of reviews and ratings of local cuisine, see equations follow for detailed. Specific scales 1 to 5 and 1 to 7 represent percentiles in the relative scale. e.g. number of reviews equal 1 if the city's relative number of reviews of Yue cuisine is in the bottom 14 percent of all cities' relative number of reviews; rating of flavor equals 1 if the city's relative rating of Yue restaurant is in the bottom 20 percent of all cities' relative ratings of the cuisine.

region. In comparison, the coastal cities and the metropolitans are much more welcoming to the non-local cuisine, as shown by both high ratings on flavor and large number of reviews. One possible reason is that these coastal cities absorb large numbers of immigrants from all parts of the country, making the taste of the cities much more diverse and open.

For Chuan cuisine, the coastal provinces in the northern part of the country tend to like the cuisine better than their inland counterparts, except for Xinjiang (Urumqi), to a certain degree. At the same time, the southern cities exhibit comparatively limited interests in Chuan cuisine except for, quite surprisingly, Fuzhou in Fujian province. This result is not completely unexpected as the preference of southern China for mildness and ingredients in its original taste is not really compatible with Sichuan cuisine that features heavy use of hot peppers. Similarly, the four southwestern provinces also exhibit limited interests in Sichuan cuisine, probably due to the similarity between their local cuisine and the Sichuan cuisine. The interests for Yue cuisine shift southward as comparing to Chuan cuisine. The cities that demonstrated truly burning interests for Yue cuisine are those that are relatively proximate to major fishing grounds and production sites of seafood as well as freshwater fish. These products are characteristic ingredients of Yue cuisine and Yue style of cooking tends to be quite picky on the quality of ingredients. Such relation may in part explain the interest in Yue cuisine in Shenyang and Jinan that are located along China's northern coastlines.

2) Popularity and distance to the origins





To quantify the geographic relations, we use distance to the origin as a way to measure the degree of separation and regional differences. Supposedly, the greater the distance to the origin, the more distinct the city is in comparison to the origin city of the cuisine in terms of taste, culture, and preferences. From a demographic perspective, it is also likely that the greater the distance between the origin and the comparison city, the fewer number of immigrants from the origin city in the comparison city.

The first plot shows the relationship between a city's distance to the origin of Chuan cuisine (Chengdu) in log scale and the relative flavor rating of Chuan cuisine (defined as the same in the map scale) restaurants in the city. Surprisingly, there is a relatively significant positive correlation between these two variables. That is, the further away a city is from Chengdu, the better the city likes Chuan cuisine. Also, northern cities have higher ratings for Chuan food than those of the southern cities, while both northern and southern cities have much higher ratings for the cuisine than cities in western part of the country. The fact that northern cities like Chuan cuisine better together with their greater distances to Chengdu in the southwest contributes to the positive association found here.

The second plot shows the relationship between a city's distance to the origin of Yue cuisine (Guangzhou) in log scale and the relative flavor ratings of restaurants serving Yue cuisine in that city. Again, there is a significant positive correlation between these two variables. In contrast to the Chuan cuisine case study, southern cities favor Yue cuisine more than northern cities, and western cities still have relatively low ratings. Since Yue cuisine shares many similarities with southern local cuisine, southern people might be more likely to accept Yue style cooking.

We removed the data point corresponding to Chengdu from the Chuan cuisine plot and the data point corresponding to Guangzhou from the Yue cuisine plot. Since those two cities are where Chuan and Yue cuisine originated and occur at the highest densities, their flavor ratings are affected by variables besides distance from Chengdu and Guangzhou, respectively. Besides, a log distance of 0 miles is undefined and a log distance of something approaching 0 is relatively far from the other data points along the log distance scale.

Overall, there is a positive correlation between distance and relative ratings for both of the cuisines. This result indicates that although greater distances may represent cultural barriers and fewer immigrants from the origin cities, such distances are not an obstacle for the acceptance of the cuisines.

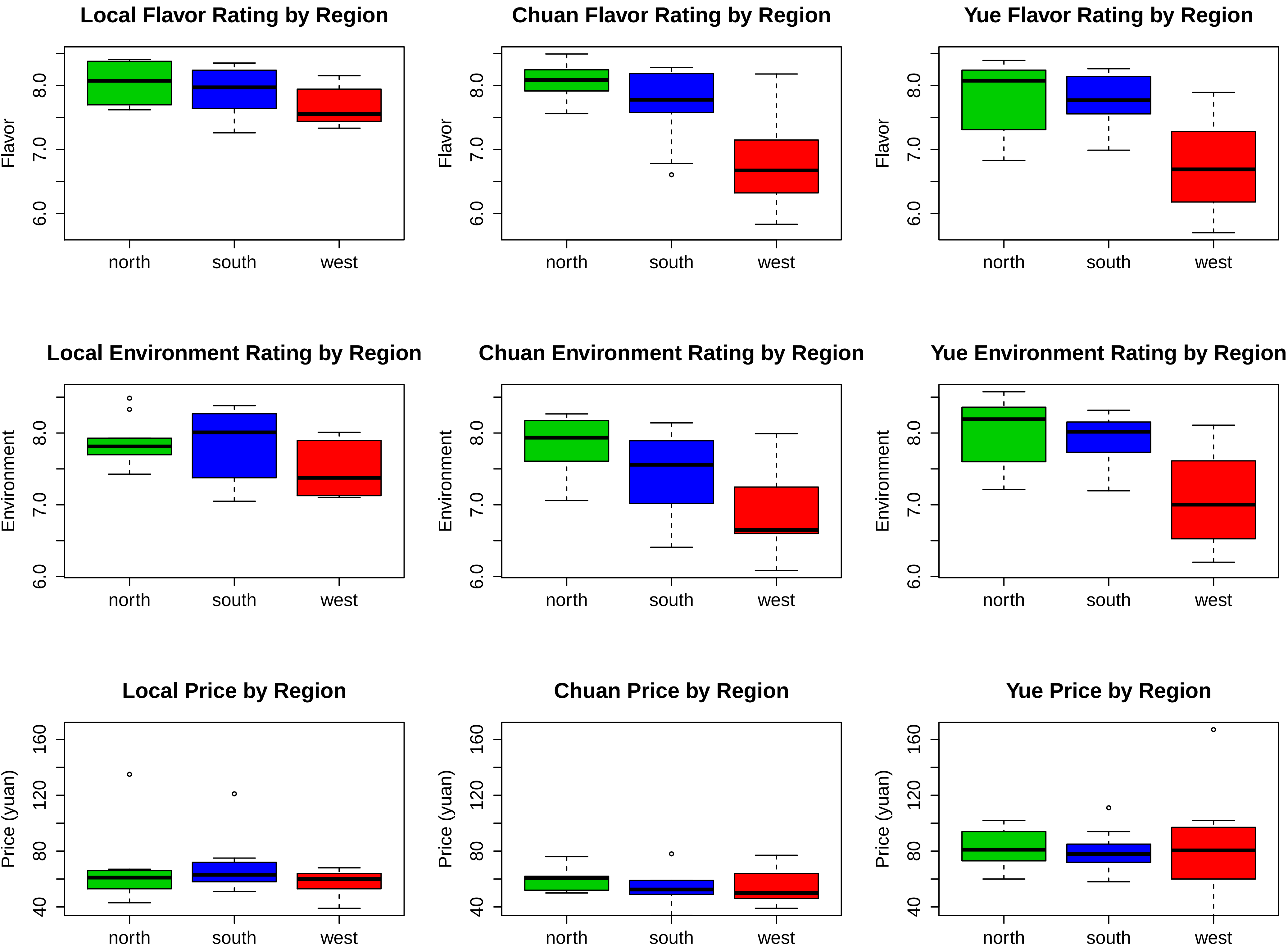
3) Decomposition by regions

To further explore the regional variations, we use box plots to visualize the average and variance of ratings and prices by region and by cuisines. Each plot corresponds to a specific cuisine's rating or prices and different colors represent different regions.

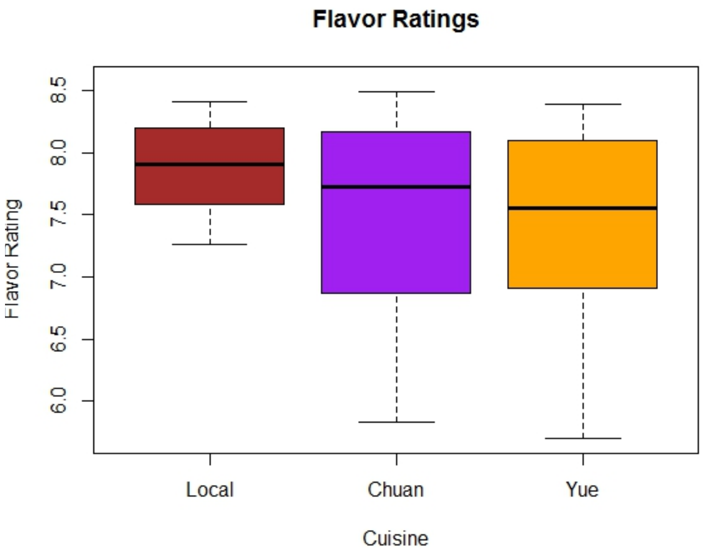
As indicated in the plots, among the ratings of flavor, environment and price, the variation of the price is the smallest for all regions. It makes sense intuitively since price is a relatively objective parameter compared to ratings of flavor and environment. Also, price clearly reflects the degree of economic development and the affluence of the cities, as the average prices for all three cuisines are highest in southern cities that are most economically advanced among the three regions. However, there is a clear price difference among the cuisine as the relative costs of Yue cuisine are much higher than the other two. Such results agree with our expectations that Yue restaurants tend to capture the higher end of the market.

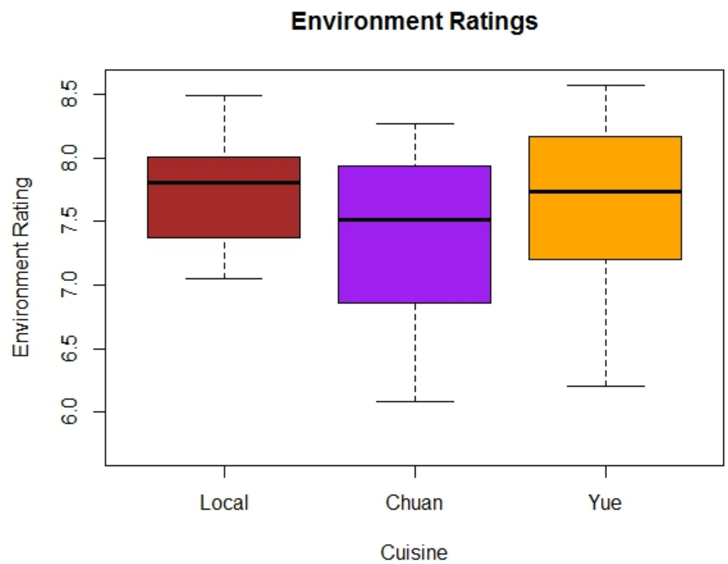
Confirming the previous conclusion, people from western region tend to rate both flavor and the environment low for all types of cuisine compared to those from south and north. There is also a considerable degree of within-region variation of ratings among the western cities as shown by the sizable interquartile ranges of these cities' ratings. This result confirms our expectation: since western China is relatively undeveloped compared to the north and the south, it attracts fewer immigrants from other cities living there. A large proportion of the population is composed of ethnic minorities. Therefore, a limited degree of openness as well as cultural differences between those ethnic minorities and Han leads to diverse preference on non-local cuisines, especially Yue cuisine, since it is far away from the west and its flavor is quite different from local western food.

When using the local ratings and pricing as benchmark for comparison, some interesting patterns emerge. First, northern people seem to like the two foreign cuisines better in terms of flavors compared to their local cuisine, while the other two regions clearly prefer their corresponding local cuisines. Similarly, restaurants in the northern cities that serve local cuisine receive lower ratings for their environment compared to those that serve either Chuan or Yue cuisine, while the opposite happens in the other two regions. Such phenomena could be due to the fact that local cuisines in the northern regions tend to be more rustic and homely, which is less attractive than the two foreign cuisines.


Local Flavor Rating by Region
Chuan Flavor Rating by Region
Yue Flavor Rating by Region
Flavor
6.0
7.0
8.0
north
south
west
Local Environment Rating by Region
Chuan Environment Rating by Region
Yue Environment Rating by Region
Environment
Local Price by Region
Chuan Price by Region
Yue Price by Region
Price (yuan)
40
80
120
160

4) Decomposition by cuisines

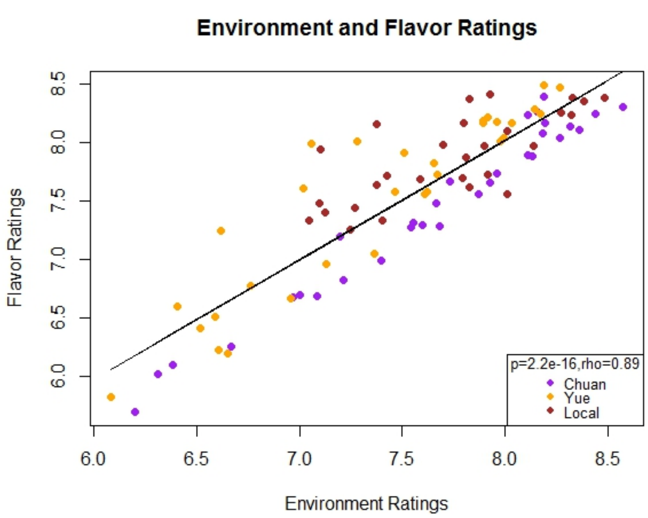




The following analyses are based on decomposition by cuisine. The first chart shows that local cuisine has the highest flavor rating and smallest variance compared to Chuan and Yue cuisines. This result is reasonable because people naturally prefer their local food or food from their origins. Chuan and Yue styles have similar averages and variances of flavor ratings across China. However, it is worth pointing out that the flavor ratings for Chuan cuisine are higher than that of the Yue cuisine, possibly because Chuan cuisine represents lower-end, family style of cooking that people have low expectations of.

The second chart shows that Yue style restaurants usually have better environments and Chuan style restaurants tend to feature more casual environments. This is as expected because Yue style is more likely to be high-end and delicate by nature, while Chuan cuisine tends to use common ingredients and is more compatible with an informal environment. Local cuisine restaurants have comparable, but more predictable (smaller variance) environment ratings.

5) Relationship between reviews and ratings

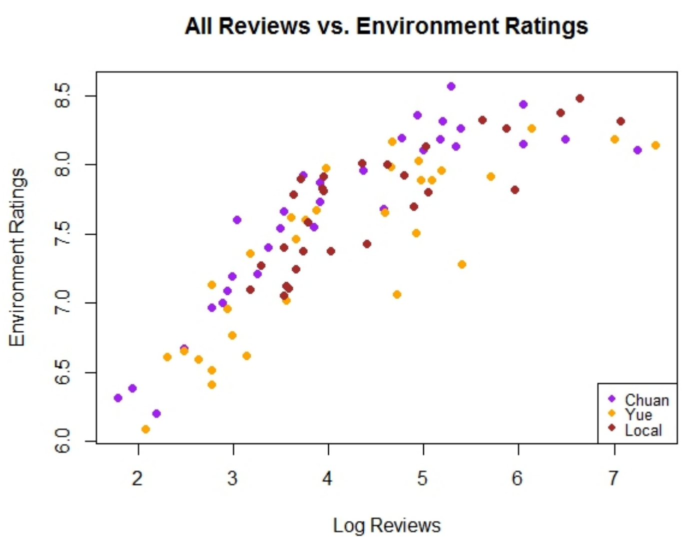


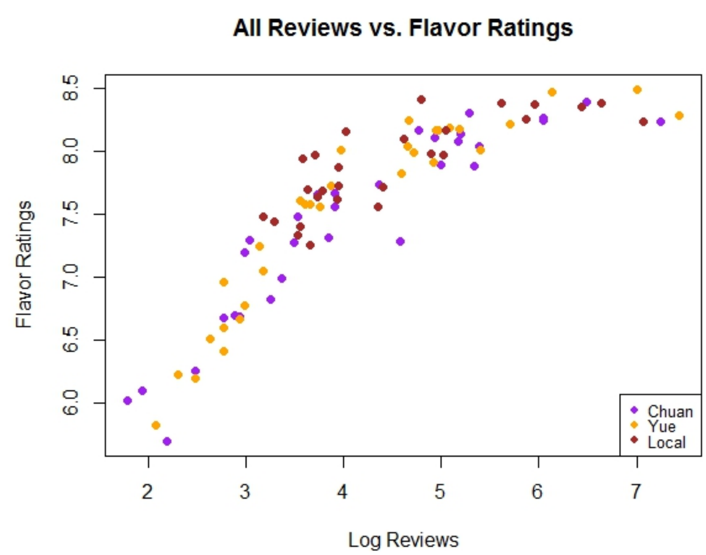
The following three plots explore the relationship between number of reviews, flavor ratings and environment ratings. The points are color-coded to represent different cuisine types. The first plot shows the relation between environment rating and flavor rating. As shown in the graph, flavor ratings and environment ratings are highly correlated, with p =2.2e-16 (highly statistically significant) and ρ=0.89 (strongly correlated). This makes sense because restaurants with better environments tend to have better flavors. Also, all data points for all three types of cuisines for all 31 cities is included, so the chance that this correlation arose out of random chance alone is very small, considering the number of data points. For a given environment rating, restaurants serving Yue cuisine tend to have higher flavor ratings than restaurants serving Chuan cuisine.

Similar positive correlations appear between number of reviews (in log scale) and environment ratings as well as between number of reviews (in log scale) and flavor ratings. Ratings and number of reviews are mutually reinforcing. Restaurants with higher ratings for flavors and restaurants could attract more customers who then generate more reviews. On the other hand, a large number of reviews represents popularity of the restaurants, which then hints on high ratings.

Customers usually only write reviews if they are either extremely satisfied or extremely dissatisfied with the restaurant. Therefore, a restaurant with a large number of reviews might have high ratings or low ratings, or most commonly, ratings in the middle. In this case, it seems as though the majority of reviewers were satisfied, as there are no data points corresponding to a large number of reviews and low ratings.

Note that in both graphs, the slopes between the two variables become smaller as the number of reviews increases. That is, as average number of reviews gets large enough, the positive association with ratings vanishes. The relationship between both log reviews and environment and log reviews and flavor does not appear to be linear. More reviews does not necessarily correspond to a higher rating, as discussed above, and because the maximum rating a reviewer can give a restaurant is 10, after a certain number of reviews, the average rating cannot get any higher so it asymptotically approaches the high rating of about 8.3, as seen visually on the plots.





4. Concluding remarks

Using restaurant ratings and pricing data collected from dianping.com for restaurants in China's provincial capitals serving Chuan, Yue and local cuisine, we explored the determinants of price and popularity of the above cuisines in China. Our analyses suggest that while both flavor and environment ratings are strongly correlated with price, environment rating is a better determinant of prices among Chinese restaurants. In terms of popularity of Chuan and Yue cuisine, there is a considerable degree of regional variation. Overall, northern cities are most open to non-local cuisines while twestern cities are least acceptant of non-local cooking style. Chuan restaurants are relatively more popular compared to Yue restaurants, probably because the former are more casual and homely. In addition, coastal and metropolitan cities that attract large number of immigrants are generally more willing to welcome new cooking style. Specifically, large geographic distance does not function as a barrier for the acceptance of exotic cuisines. These brief analyses on cities' preferences for cuisines and determinants of price allude to China's regional cultural diversity and degree of openness. In particular, cities in the western regions with high representation of ethnic diversity and distinctive local culture show the least degree of acceptance to non-local cuisine, which reaffirm the existence of cultural differences.

This study of Chinese restaurants is far from complete and the conclusions presented here are limited in various aspects. For further studies, it would be revealing to explore the relationship between restaurant/cuisine popularity and the demographic composition of the cities. That is, how are cities that attract large number of immigrants or have diverse population profiles in terms of people's origins different from cities on the other end of the spectrum? How does ethnic composition factor in the picture? For example, to what degree do cities that are dominated by Han (non-minority) population accept cuisines of the ethnic minorities? Does economic development have an impact on a city's taste? These question raised here could serve as new starting points for continuous exploration of the topic.

Appendix

Table 1

Number of restaurants	< 500	500 -1000	1000 -5000	5000 -10000	>10000
Number of pages scraped[1]	3	5	10	15	20

[1] 15 restaurants per page

Table 2

	Chuan	Yue	Local cuisine
Shanghai	15	15	20
Beijing	15	15	15
Hangzhou	10	10	15
Guangzhou	10	20	N/A
Nanjing	10	10	15
Chengdu	15	5	N/A
Chongqing	15	5	N/A
Tianjin	10	5	10
Fuzhou	5	3	10
Wuhan	10	3	15
Xi’an	10	3	10
Shenyang	10	3	10
Jinan	5	3	10
Haikou	3	3	3
Shijiazhuang	10	3	3
Taiyuan	5	3	5
Hohhot	3	3	3
Changchun	5	3	10
Harbin	5	3	10
Hefei	5	3	10

Nanchang	5	3	10
Zhengzhou	10	5	10
Nanning	3	3	3
Guiyang	3	3	10
Kunming	5	5	10
Lahsa	5	1	10
Lanzhou	5	3	10
Xining	3	3	3
Yinchuan	3	3	3
Urumqi	10	3	10
Changsha	5	3	10

Table 3

	City	Category	ID	Price	Flavor	Environment	Review
16	Shanghai	Chuan	111	90	9	8.6	297
17	Shanghai	Chuan	112	84	9.2	9.1	1079
18	Shanghai	Chuan	113	80	9.1	8.9	6186
19	Shanghai	Chuan	114	73	9	8.7	10490

Table 4

North	Harbin, Changchun, Shenyang, Beijing, Tianjin, Zhengzhou, Shijiazhuang, Taiyuan, Xian, Jinan.
South	Wuhan, Changsha, Nanjing, Shanghai, Hangzhou, Fuzhou, Guangzhou, Hefei, Nanchang, Haikou.
West	Chengdu, Chongqing, Guiyang, Kunming, Nanning, Lhasa, Urumqi, Hohhot, Yinchuan, Xining, Lanzhou

Table 5

	p-value		ρ (corr. coefficient)		reg line slope	
	Flavor	Env	Flavor	Env	Flavor	Env
Chuan	0.0001346	1.481e-6	0.63	0.75	8.899	12.81
Yue	0.3418	0.4094	0.18	0.16	4.828	4.835

Local	0.06664	0.002828	0.35	0.54	19.06	25.19
All	0.01312	4.793e-6	0.26	0.47	7.722	15.33

R code

```
#Part 1: scrape restaurant data from dianping.com
url = "http://www.dianping.com/search/category/1/10/g102o3p1"
temp = readLines(url)
indexscore1 = grep(".*score1.*", temp)
linesscore1 = temp[indexscore1[-1]]
flavor = as.numeric(gsub(pattern =".*([[:digit:]][.][[:digit:]]).*", replacement = "\\1", linesscore1))
indexscore2 = grep(".*score2.*", temp)
linesscore2 = temp[indexscore2[-1]]
environment = as.numeric(gsub(pattern =".*([[:digit:]][.][[:digit:]]).*", replacement = "\\1",
linesscore2))
indexprice = grep("Price", temp)
linesprice = temp[indexprice[-(1:2)]]
price = as.numeric(gsub(pattern = ".*>([[:digit:]]+)<.*", replacement = "\\1", linesprice))
indexreview = indexprice - 4
linesreview = temp[indexreview[-(1:2)]]
review = as.numeric(gsub(pattern = ".*>([[:digit:]]+)\345.*", replacement = "\\1", linesreview))
city = paste(rep("Shanghai", times = length(flavor)))
citycode = rep(1, times = length(flavor))
category = paste(rep("Chuan", times = length(flavor)))
categorycode = rep(1, times = length(flavor))
rescode = 1:length(flavor)
id = as.numeric(paste(as.character(citycode), as.character(categorycode), as.character(rescode),
sep = ""))
restaurant = data.frame(city, category, id, price, flavor, environment, review)
#loop through cities and cuisine
#city code and category code collected from dianping.com,
#need to vary and manually input for each combination
#based on table 2, vary the number of pages
capital = c(Shanghai = 1)
searchpage = as.character(1:15)
type = c(Yue = 103, Chuan = 102)
for (i in 1:length(capital)){
 for (j in 1:length(searchpage)){
        for (k in 1:length(type)){
        print(names(capital[i]))
        print(names(type[k]))
        print(searchpage[j])
        url = paste("http://www.dianping.com/search/category/", as.character(capital[i]), "/10/g",
type[k], "o3p", searchpage[j], sep = "")
        print(url)
        page = readLines(url)
```

```
        indexscore1 = grep(".*score1[^[:alpha:]].*", page)
        linesscore1 = page[indexscore1]
        flavor = as.numeric(gsub(pattern =".*(([[:digit:]][.][[:digit:]])|(-)).*", replacement = "\\1",
linesscore1))
        indexscore2 = grep(".*score2[^[:alpha:]].*", page)
        linesscore2 = page[indexscore2]
        environment = as.numeric(gsub(pattern =".*(([[:digit:]][.][[:digit:]])|(-)).*", replacement
= "\\1", linesscore2))
        indexprice = grep("average", page)
        linesprice = page[indexprice]
        price = as.numeric(gsub(pattern = ".*>([[:digit:]]+)<.*", replacement = "\\1", linesprice))
        indexreview = indexprice - 4
        linesreview = page[indexreview]
        review = as.numeric(gsub(pattern = ".*>([[:digit:]]+)\345.*", replacement = "\\1",
linesreview))
        city = paste(rep(names(capital[i]), times = length(flavor)))
        citycode = rep(capital[i], times = length(flavor))
        category = paste(rep(names(type[k]), times = length(flavor)))
        categorycode = rep(type[k], times = length(flavor))
        rescode = 1:length(flavor) + 15*(as.numeric(searchpage[j])-1)
        id = as.numeric(paste(as.character(citycode), as.character(categorycode),
as.character(rescode), sep = ""))
        temp = data.frame(city, category, id, price, flavor, environment, review)
        restaurant = rbind(restaurant, temp)
 }}}
#cleaning up data
restaurant = restaurant[restaurant$review >= 5,]
judge = is.na(restaurant$id)
restaurant = restaurant[!judge,]
restaurant = restaurant[-(1:15),]
#clear rows with NA
judge = sapply(restaurant[,5:7], is.na)
missing = apply(judge, 1, function(x){
 ifelse(sum(x)>0, TRUE, FALSE)
})
cleandata = restaurant[!missing,]
write.csv(restaurant, "cleandata.csv")

#Part 2: calculate aggregate statistics
#get weighted average
library(sqldf)
sumreview = sqldf("SELECT city, category, SUM(review) AS sumreview FROM cleandata
                GROUP BY city, category")
data = merge(cleandata, sumreview, by = c("city", "category"), all = TRUE)
data$weight = data$review/data$sumreview
data$wprice = data$weight*data$price
```

```
data$wflavor = data$weight*data$flavor
data$wenvironment = data$weight*data$environment
prov.category = sqldf("SELECT city, category, SUM(wprice) AS price, SUM(wflavor) AS
flavor,
                    SUM(wenvironment) AS environment, AVG(review) AS review FROM data
                    GROUP BY city, category")
write.csv(prov.category, "prov.category.csv")
#restructure the data
Chuan = prov.category[prov.category$category == "Chuan",][,-2]
colnames(Chuan) = c("city", "price.Chuan", "flavor.Chuan", "environment.Chuan",
"review.Chuan")
Yue = prov.category[prov.category$category == "Yue",][,-2]
colnames(Yue) = c("city", "price.Yue", "flavor.Yue", "environment.Yue", "review.Yue")
Local = prov.category[prov.category$category == "Local",][,-2]
colnames(Local) = c("city", "price.Local", "flavor.Local", "environment.Local", "review.Local")
temp = merge(Chuan, Yue, by = "city", all = TRUE)
province = merge(temp, Local, by = "city", all = TRUE)
rm(temp)
write.csv(province, "province.csv")

#Part 3: collect geographic coordinates, calculate distance
url = "http://www.tageo.com/index-e-ch-cities-CN.html"
page = readLines(url)
all = page[grep("Shanghai",page)]
#extract city name
city_pattern = "<b>([A-Z][a-z]+)</b>"
city_match = gregexpr(city_pattern, all)
temp_cities = substring(all, city_match[[1]], city_match[[1]]+attr(city_match[[1]],
"match.length")-1)
temp_cities = temp_cities[3:length(temp_cities)]
cities = gsub(city_pattern,"\\1",temp_cities)
#exctract coordinates info
coor_pattern = "<td width=[^>]+>[0-9]+</td><td>([0-9.]+)</td><td>([0-9.]+)</td>"
coor_match = gregexpr(coor_pattern,all)
temp_coor = substring(all,coor_match[[1]], coor_match[[1]]+attr(coor_match[[1]],
"match.length")-1)
lati = as.numeric(gsub(coor_pattern, "\\1", temp_coor))
long = as.numeric(gsub(coor_pattern, "\\2", temp_coor))
#construct data frame
geography = data.frame(city = cities, Latitude = lati, Longitude = long)
haikou = c(Longitude = 110.320, Latitude = 20.050)
yinchuan = c(Longitude = 106.320, Latitude = 38.470)
lhasa = c(Longitude = 91.12, Latitude = 29.65)
chengdu = c(LongChuan = 104.07, LatChuan = 30.67)
guangzhou = c(LongYue = 113.25, LatYue = 23.15)
fix = rbind(haikou, yinchuan, lhasa)
```

```
fix = cbind(fix, chengdu[1], chengdu[2], guangzhou[1], guangzhou[2])
write.csv(geography, "geography.csv")
#calculate distance
#convert degree to radius
deg2rad = function(deg) return(deg*pi/180)
#calculate great circle distance
gcd = function(long1, lat1, long2, lat2){
 R = 6371 #earth radius
 d = acos(sin(lat1)*sin(lat2) + cos(lat1)*cos(lat2)*cos(long2-long1))*R
 return(d)
}
geography$latYue = geography$Latitude[geography$city == "Guangzhou"]
geography$longYue = geography$Longitude[geography$city == "Guangzhou"]
geography$latChuan = geography$Latitude[geography$city == "Chengdu"]
geography$longChuan = geography$Longitude[geography$city == "Chengdu"]
#convert to radius
geo = apply(geography[,2:7], 2, deg2rad)
geofix = apply(fix, 2, deg2rad)
distChuan = rep(NA, length = length(geo[,1]))
for (i in 1:length(geo[,1])){
 distChuan[i] = gcd(geo[i,2], geo[i,1], geo[i,4], geo[i,3])
}
distYue = rep(NA, length = length(geo[,1]))
for (i in 1:length(geo[,1])){
 distYue[i] = gcd(geo[i,2], geo[i,1], geo[i,6], geo[i,5])
}
geography$distChuan = distChuan
geography$distYue = distYue
#add the three missing ones
distChuanfix = rep(NA, length = length(geofix[,1]))
for (i in 1:length(geofix[,1])){
 distChuanfix[i] = gcd(geofix[i,2], geofix[i,1], geofix[i,4], geofix[i,3])
}
distYuefix = rep(NA, length = length(geofix[,1]))
for (i in 1:length(geo[,1])){
 distYuefix[i] = gcd(geofix[i,2], geofix[i,1], geofix[i,6], geofix[i,5])
}
write.csv(geography, "geography.csv")
province = read.csv("province.csv")
province = merge(province, geography, by = "city", all = TRUE)
province = province[1:31,-(15:20)]
province$distChuan[province$city == "Haikou"] = distChuanfix[1]
province$distYue[province$city == "Haikou"] = distYuefix[1]
province$distChuan[province$city == "Yinchuan"] = distChuanfix[2]
province$distYue[province$city == "Yinchuan"] = distYuefix[2]
province$distChuan[province$city == "Lhasa"] = distChuanfix[3]
```

```
province$distYue[province$city == "Lhasa"] = distYuefix[3]
write.csv(province, "provincegeo.csv")

#Part 4: making plots
data<-read.csv("provincegeo.csv")
#add geographic region info
north<-c("Harbin","Changchun","Shenyang","Beijing","Tianjin","Zhengzhou",
      "Shijiazhuang","Taiyuan","Xian","Jinan")
south<-c("Wuhan","Changsha","Nanjing","Shanghai","Hangzhou","Fuzhou",
      "Guangzhou","Hefei","Nanchang","Haikou")
west<-
c("Chengdu","Chongqing","Guiyang","Kunming","Nanning","Lhasa","Urumqi","Hohhot",
     "Yinchuan","Xining","Lanzhou")
city<-c(north,south,west)
region<-c(rep("north",10),rep("south",10),rep("west",11))
regionType<-data.frame(city,region)
regionType<-regionType[order(regionType[,1]),]
col<-rep(NA,31)
for(i in 1:length(col)){
 if(regionType$region[i]=="north")
         col[i]<-"green3"
 if(regionType$region[i]=="south")
         col[i]<-"blue"
 if(regionType$region[i]=="west")
         col[i]<-"red"
}
##Determinants of Pricing (Section 3.1)
#Is flavor or environment a better predictor of price?
x11()
par(mfrow=c(2,4))
xscale<-c(min(data$flavor.Yue,na.rm=T),max(data$environment.Yue,na.rm=T))
yscale<-c(min(data$price.Yue,na.rm=T),max(data$price.Local,na.rm=T))
#Chuan
#Flavor
plot(data$flavor.Chuan,data$price.Chuan,xlab="Flavor Ratings",ylab="Price (yuan)",
         main="Chuan Flavor and Price",col=col,pch=16,xlim=xscale,ylim=yscale)
legend("topleft", legend = c("north","south","west"), pch=16,cex=0.8,
         col = c("green3", "blue","red"),title="p=1.35e-4,rho=0.63")
coefPrFl<-lm(data$price.Chuan~data$flavor.Chuan)
lines(data$flavor.Chuan,fitted(coefPrFl))
cor.test(data$flavor.Chuan,data$price.Chuan)
#Environment
plot(data$environment.Chuan,data$price.Chuan,xlab="Environment Ratings",
         ylab="Price (yuan)",main="Chuan Environment and Price",col=col,pch=16,
         xlim=xscale,ylim=yscale)
legend("topleft", legend = c("north","south","west"), pch=16,cex=0.8,
```

```
	col = c("green3", "blue","red"),title="p=1.48e-6,rho=0.75")
coefPrEnv<-lm(data$price.Chuan~data$environment.Chuan)
lines(data$environment.Chuan,fitted(coefPrEnv))
cor.test(data$environment.Chuan,data$price.Chuan)
#for Chuan, environment is a better predictor
#Yue
#Flavor
plot(data$flavor.Yue[data$price.Yue!=167],data$price.Yue[data$price.Yue!=167],
	xlim=xscale,ylim=yscale,xlab="Flavor Ratings",ylab="Price (yuan)",
	main="Yue Flavor and Price",col=col,pch=16)
legend("topleft", legend = c("north","south","west"), pch=16,cex=0.8,
	col = c("green3", "blue","red"),title="p=0.342,rho=0.18")
coefPrFl<-lm(data$price.Yue[data$price.Yue!=167]~data$flavor.Yue[data$price.Yue!=167]
	,na.action=na.exclude)
lines(data$flavor.Yue[data$price.Yue!=167],fitted(coefPrFl))
cor.test(data$flavor.Yue[data$price.Yue!=167],data$price.Yue[data$price.Yue!=167])
#Environment
plot(data$environment.Yue[data$price.Yue!=167],data$price.Yue[data$price.Yue!=167],
	xlab="Environment Ratings",ylab="Price (yuan)",
	main="Yue Environment and Price",col=col,pch=16,xlim=xscale,ylim=yscale)
legend("topleft", legend = c("north","south","west"), pch=16,cex=0.8,
	col = c("green3", "blue","red"),title="p=0.409,rho=0.16")
coefPrEnv<-
lm(data$price.Yue[data$price.Yue!=167]~data$environment.Yue[data$price.Yue!=167],
	na.action=na.exclude)
lines(data$environment.Yue[data$price.Yue!=167],fitted(coefPrEnv))
cor.test(data$environment.Yue[data$price.Yue!=167],data$price.Yue[data$price.Yue!=167])
#neither is significant, environment is a slightly better predictor
#Local
#Flavor
plot(data$flavor.Local,data$price.Local,xlab="Flavor Ratings",ylab="Price (yuan)",
	main="Local Flavor and Price",xlim=xscale,ylim=yscale,col=col,pch=16)
legend("topleft", legend = c("north","south","west"), pch=16,cex=0.8,
	col = c("green3", "blue","red"),title="p=0.0666,rho=0.35")
coefPrFl<-lm(data$price.Local~data$flavor.Local,na.action=na.exclude)
lines(data$flavor.Local,fitted(coefPrFl))
cor.test(data$flavor.Local,data$price.Local)
#Environment
plot(data$environment.Local,data$price.Local,xlab="Environment Ratings",ylab=
	"Price (yuan)",main="Local Environment and Price",col=col,pch=16,
	xlim=xscale,ylim=yscale)
legend("topleft", legend = c("north","south","west"), pch=16,cex=0.8,
	col = c("green3", "blue","red"),title="p=0.00282,rho=0.54")
coefPrEnv<-lm(data$price.Local~data$environment.Local,na.action=na.exclude)
lines(data$environment.Local,fitted(coefPrEnv))
cor.test(data$environment.Local,data$price.Local)
```

```
#Environment is significant, p=0.002828, flavor is not
#all 3 grouped together
#Flavor
plot(c(data$flavor.Chuan,data$flavor.Yue[data$price.Yue!=167],data$flavor.Local),c(data$price
.Chuan,
        data$price.Yue[data$price.Yue!=167],data$price.Local),
  col=c(rep("purple",31),rep("orange",31),rep("brown",31)),pch=16,
        xlim=xscale,ylim=yscale,
        xlab="Flavor Rating",
        ylab="Price (yuan)",main="Does Flavor Rating Predict Price?")
legend("topleft", legend = c("Chuan","Yue","Local"), pch=16,cex=0.8,
        col = c("purple", "orange","brown"),title="p=0.0131,rho=0.26")
coefFl<-lm(c(data$price.Chuan,data$price.Yue[data$price.Yue!=167],data$price.Local)
   ~c(data$flavor.Chuan,data$flavor.Yue[data$price.Yue!=167],data$flavor.Local),na.action=na.
exclude)
lines(c(data$flavor.Chuan,data$flavor.Yue[data$price.Yue!=167],data$flavor.Local),fitted(coefF
l))
cor.test(c(data$flavor.Chuan,data$flavor.Yue[data$price.Yue!=167],data$flavor.Local),
        c(data$price.Chuan,data$price.Yue[data$price.Yue!=167],data$price.Local))
#Environment
plot(c(data$environment.Chuan,data$environment.Yue[data$price.Yue!=167],data$environment.
Local),
  c(data$price.Chuan,data$price.Yue[data$price.Yue!=167],data$price.Local),
        col=c(rep("purple",31),rep("orange",31),rep("brown",31)),pch=16,xlim=xscale,
        ylim=yscale,xlab="Environment Rating",ylab="Price",
        main="Does Environment Rating Predict Price?")
legend("topleft", legend = c("Chuan","Yue","Local"), pch=16,cex=0.8,
        col = c("purple", "orange","brown"),title="p=4.79e-6,rho=0.47")
coefEnv<-lm(c(data$price.Chuan,data$price.Yue[data$price.Yue!=167],data$price.Local)
       ~c(data$environment.Chuan,data$environment.Yue[data$price.Yue!=167],
        data$environment.Local),na.action=na.exclude)
lines(c(data$environment.Chuan,data$environment.Yue[data$price.Yue!=167],
     data$environment.Local),fitted(coefEnv))
cor.test(c(data$environment.Chuan,data$environment.Yue[data$price.Yue!=167],
        data$environment.Local),
    c(data$price.Chuan,data$price.Yue[data$price.Yue!=167],data$price.Local))

##Distance and Flavor Ratings (Section 3.2.2)
#just look at flavor ratings b/c flavor comes from 2 places
#Chuan

plot(log(data$distChuan),(data$flavor.Chuan-data$flavor.Local)/data$flavor.Local,
   xlab="Log Distance from Chuan (miles)",
   ylab="Flavor Rating Compared to Local Cuisine",main="Distance from Chuan and Chuan
Flavor Rating",
   col=col,pch=16)
```

```
legend("bottomright", legend = c("north","south","west"), pch=16,cex=0.8,
      col = c("green3", "blue","red"),title="p=0.0712, rho=0.35")

#correct by removing Chuan's data
coefChuan<-lm((data$flavor.Chuan[data$distChuan!=0]-data$flavor.Local[data$distChuan!=0])/
     data$flavor.Local[data$distChuan!=0]~log(data$distChuan[data$distChuan!=0]),na.action=na
.exclude)
lines(log(data$distChuan[data$distChuan!=0]),fitted(coefChuan))
cor.test((data$flavor.Chuan[data$distChuan!=0]-data$flavor.Local[data$distChuan!=0])/
     data$flavor.Local[data$distChuan!=0],log(data$distChuan[data$distChuan!=0]))#p=0.0712,r
ho=0.35

#Yue
plot(log(data$distYue),(data$flavor.Yue-data$flavor.Local)/data$flavor.Local,pch=16,
    col=col,xlab="Log Distance from Yue (miles)",
    ylab="Flavor Rating Compared to Local Cuisine",main="Distance from Yue and Yue Flavor
Rating")
legend("bottomright", legend = c("north","south","west"), pch=16,cex=0.8,
      col = c("green3", "blue","red"),title="p=0.0588, rho=0.36")

#correct by removing Yue's data
coefYue<-lm((data$flavor.Yue[data$distYue!=0]-data$flavor.Local[data$distYue!=0])/
     data$flavor.Local[data$distYue!=0]~log(data$distYue[data$distYue!=0]),
     na.action=na.exclude)
lines(log(data$distYue[data$distYue!=0]),fitted(coefYue))
cor.test(log(data$distYue[data$distYue!=0]),(data$flavor.Yue[data$distYue!=0]-
     data$flavor.Local[data$distYue!=0])/
     data$flavor.Local[data$distYue!=0])#p=0.0315,rho=0.41

##Decomposition by Region (Section 3.2.3)
x11()
par(mfrow=c(3,3))
yFl<-c(min(data$flavor.Yue,na.rm=T),max(data$flavor.Chuan,na.rm=T))
yEnv<-
c(min(data$environment.Chuan,na.rm=TRUE),max(data$environment.Yue,na.rm=TRUE))
yP = c(min(data$price.Local, na.rm = T), max(data$price.Yue, na.rm = T))
#Flavor
boxplot(split(data$flavor.Local,regionType$region),main="Local Flavor Rating by Region",
            ylab="Flavor",col=c("green3","blue","red"),ylim=yFl)
boxplot(split(data$flavor.Chuan,regionType$region),main="Chuan Flavor Rating by Region",
       ylab="Flavor",col=c("green3","blue","red"),ylim=yFl)
boxplot(split(data$flavor.Yue,regionType$region),main="Yue Flavor Rating by Region",
       ylab="Flavor",col=c("green3","blue","red"),ylim=yFl)
#Environment
boxplot(split(data$environment.Local,regionType$region),main="Local Environment Rating by
Region",
```

```
      ylab="Environment",col=c("green3","blue","red"),ylim=yEnv)
boxplot(split(data$environment.Chuan,regionType$region),main="Chuan Environment Rating
by Region",
      ylab="Environment",col=c("green3","blue","red"),ylim=yEnv)
boxplot(split(data$environment.Yue,regionType$region),main="Yue Environment Rating by
Region",
      ylab="Environment",col=c("green3","blue","red"),ylim=yEnv)
#Price
boxplot(split(data$price.Local,regionType$region),main="Local Price by Region",
         ylab="Price (yuan)",col=c("green3","blue","red"), ylim = yP)
boxplot(split(data$price.Chuan,regionType$region),main="Chuan Price by Region",
         ylab="Price (yuan)",col=c("green3","blue","red"), ylim = yP)
boxplot(split(data$price.Yue,regionType$region),main="Yue Price by Region",
         ylab="Price (yuan)",col=c("green3","blue","red"), ylim = yP)

##Decomposition by Cuisine (Section 3.2.4)
boxplot(data$flavor.Local,data$flavor.Chuan,data$flavor.Yue,names=c("Local",
  "Chuan","Yue"),xlab="Cuisine",ylab="Flavor Rating",main="Flavor Ratings",
  col=c("brown","purple","orange"),ylim=c(min(data$flavor.Yue),
  max(data$environment.Yue)))
boxplot(data$environment.Local,data$environment.Chuan,data$environment.Yue,
  names=c("Local","Chuan","Yue"),xlab="Cuisine",ylab="Environment Rating",
  main="Environment Ratings",ylim=c(min(data$flavor.Yue),
  max(data$environment.Yue)),col=c("brown","purple","orange"))

##Relationship between reviews and ratings (Section 3.2.5)
#Log Flavor
plot(log(c(data$review.Yue,data$review.Chuan,data$review.Local)),
    c(data$flavor.Yue,data$flavor.Chuan,data$flavor.Local),xlab="Log Reviews",
    ylab="Flavor Ratings",main="All Reviews vs. Flavor Ratings",
    col=c(rep("purple",31),rep("orange",31),rep("brown",31)),pch=16)
legend("bottomright", legend = c("Chuan","Yue","Local"), pch=16,cex=0.8,
      col = c("purple", "orange","brown"))

#Log Environment
plot(log(c(data$review.Yue,data$review.Chuan,data$review.Local)),
    c(data$environment.Yue,data$environment.Chuan,data$environment.Local),
    xlab="Log Reviews",
    ylab="Environment Ratings",main="All Reviews vs. Environment Ratings",
    col=c(rep("purple",31),rep("orange",31),rep("brown",31)),pch=16)
legend("bottomright", legend = c("Chuan","Yue","Local"), pch=16,cex=0.8,
      col = c("purple", "orange","brown"))
#Relation between flavor and environment
plot(c(data$environment.Yue,data$environment.Chuan,data$environment.Local),
    c(data$flavor.Yue,data$flavor.Chuan,data$flavor.Local),
    xlab="Environment Ratings",ylab="Flavor Ratings",
```

```
     main="Environment and Flavor Ratings",
     col=c(rep("purple",31),rep("orange",31),rep("brown",31)),pch=16)
legend("bottomright", legend = c("Chuan","Yue","Local"), pch=16,cex=0.8,
       col = c("purple", "orange","brown"),title="p=2.2e-16,rho=0.89")
coefFlEnv<-lm(c(data$flavor.Yue,data$flavor.Chuan,data$flavor.Local)~
    c(data$environment.Yue,data$environment.Chuan,data$environment.Local),
    na.action=na.exclude)
lines(c(data$environment.Yue,data$environment.Chuan,data$environment.Local),
    fitted(coefFlEnv))
cor.test(c(data$environment.Yue,data$environment.Chuan,data$environment.Local),
         c(data$flavor.Yue,data$flavor.Chuan,data$flavor.Local))
```