

# Chicago Restaurant Pricing Analysis Project

Isha Singh

2024-10-27



(image taken by Isha Singh)

## SECTION 01: INTRODUCTION

The Chicago restaurant pricing dataset consists of approximately 1,500 observations of restaurants within the Chicago area. Each observation includes the restaurant's name, distance from a central location (Northwestern downtown campus) in miles, the price of a meal, the price of a drink, and the category or cuisine of the restaurant (American, Indian, Mexican, etc.). For the purpose of improved analysis, the collected data was also modified to incorporate the following: a 12 percent tax, the total cost of the meal including transportation, and a ratio of the price of the food to the total cost.

There are various types of parameters within the dataset, which further contribute to explaining the analysis. The dataset contains numerical variables for pricing and distance to the central location, as well as categorical variables (different cuisines). Since numerical variables contribute to understanding pricing trends across various cuisines and their distances, they are very important. The initial descriptive analysis provides information on the skewness and variability in the price distribution by cuisine, ranging from inexpensive to

expensive categories. In order to gain a better understanding of the factors influencing the Chicago area, the analysis aims to emphasize the variables.

## PRE - PROCESSING SECTION

### STRUCTURE OF THE DATA FRAME

```
## 'data.frame': 1500 obs. of 5 variables:  
## $ Name : chr "Chinese Buffet" "Burger Joint" "Thai Noodle House" "Korean BBQ" ...  
## $ Distance : num 1.245 2.085 0.567 0.916 3.674 ...  
## $ Price_of_Meal : num 11.79 14.4 8.16 8.23 6.08 ...  
## $ Price_of_Drink: num 5.08 13.08 8.68 6.9 4.46 ...  
## $ Category : chr "Chinese" "Mexican" "Thai" "Korean" ...
```

### PRE\_PROCESSING OF NEW FEATURES

```
#FoodPrice by adding the Meal Price and the Drink Price  
# and adding a 12% tax for the total.  
observations$FoodPrice <- (observations$Price_of_Meal +  
                           observations$Price_of_Drink) * 1.12  
  
# (e) TT_Price the total price of a meal, considering the  
# transportation as well, use 3.5 dollars per mile when the distance is longer than 1 mile.  
observations$TT_Price <- ifelse(observations$Distance > 1,  
                               observations$FoodPrice +  
                               (observations$Distance * 3.5),  
                               observations$FoodPrice)  
  
# Status, the status of the restaurant categorizing it as  
# inexpensive when the FoodPrice is less or equal to 10 dollars,  
# Normal for a price more than 10 but less than 20 dollars and  
# Expensive if the FoodPrice is higher than 20 dollars. Change this variable to a factor.  
observations>Status <- ifelse(observations$FoodPrice <= 10, "Inexpensive",  
                             ifelse(observations$FoodPrice  
                                   >= 20, "Expensive", "Normal"))  
  
# Convert Status to a factor  
observations>Status <- as.factor(observations>Status)  
  
# Calculate the Ratio of FoodPrice to TT_Price  
observations$Ratio <- observations$FoodPrice / observations$TT_Price
```

## SECTION 01A: SUMMARIZING THE DATA

(1)(a) **(1 point)** Use summary() to obtain and present descriptive statistics from mydata. Use table() to present a frequency table using Category and Status.

Table 1: Descriptive Statistics

Distance	Price_of_Meal	Price_of_Drink	FoodPrice	TT_Price	Ratio
Min. : 0.1068	Min. : 1.000	Min. : 1.270	Min. : 3.125	Min. : 4.11	Min. :0.1384
1st Qu.: 1.5084	1st Qu.: 4.450	1st Qu.: 5.178	1st Qu.:13.751	1st Qu.: 22.22	1st Qu.:0.5297
Median : 3.0351	Median : 7.965	Median : 8.400	Median :18.575	Median : 28.60	Median :0.6376
Mean : 3.0503	Mean : 8.051	Mean : 8.605	Mean :18.654	Mean : 29.03	Mean :0.6642
3rd Qu.: 4.5012	3rd Qu.:11.360	3rd Qu.:12.255	3rd Qu.:23.002	3rd Qu.: 35.79	3rd Qu.:0.7814
Max. :18.0000	Max. :56.000	Max. :16.000	Max. :80.315	Max. :123.89	Max. :1.0000

Table 2: Frequency Table: Category vs. Status

	Expensive	Inexpensive	Normal
<b>American</b>	114	37	151
<b>Brazilian</b>	27	11	27
<b>Chinese</b>	30	9	30
<b>French</b>	23	7	24
<b>German</b>	21	9	24
<b>Indian</b>	24	9	28
<b>Italian</b>	61	13	56
<b>Japanese</b>	49	11	69
<b>Korean</b>	32	6	30
<b>Mediterranean</b>	42	25	53
<b>Mexican</b>	76	10	75
<b>Peruvian</b>	31	1	24
<b>Spanish</b>	18	8	27
<b>Thai</b>	54	9	59
<b>Vietnamese</b>	29	7	20

**Question (1 point): Briefly discuss the variable types and distributional implications such as potential skewness and outliers.**

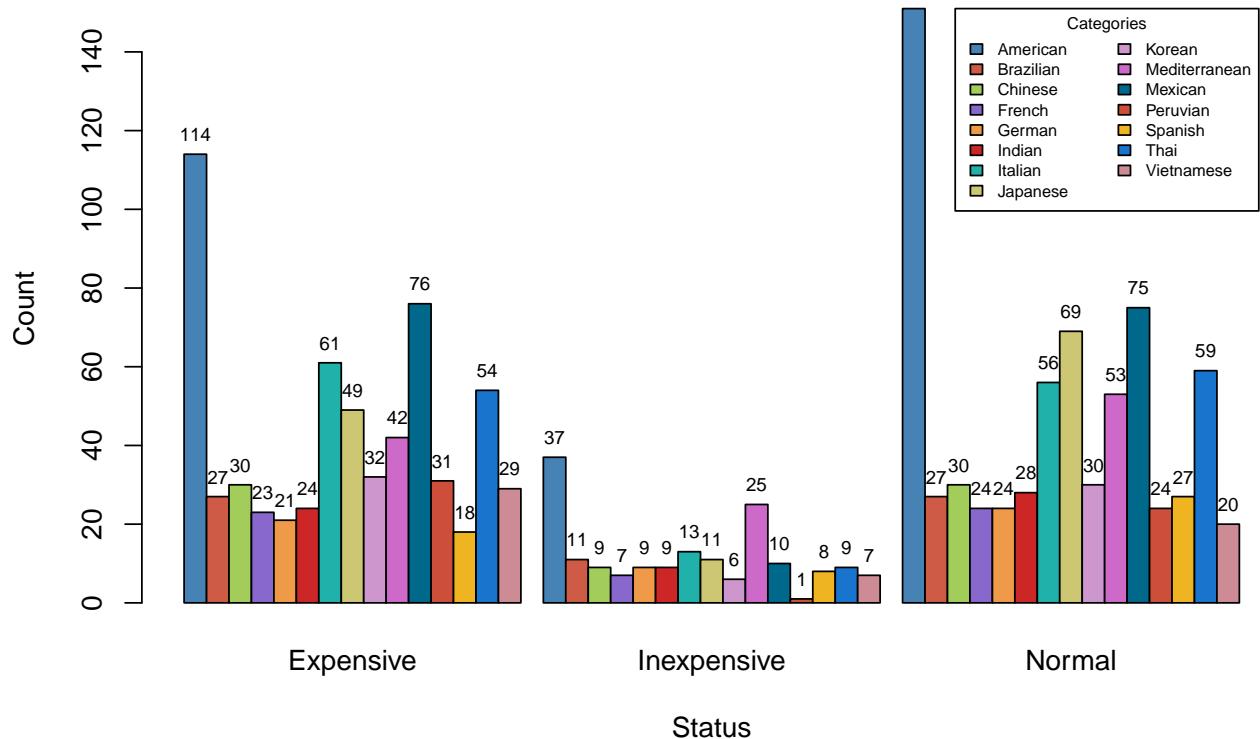
The categorical variables in the dataset are Category and Status for the pricing level. The numerical variables consist of Distance, Meal Price, Drink Price, Food Price, Total Transportation Price, and Ratio of Food Price to Total Cost. Among these numerical variables, the meal price and total transportation price show right skewness as the mean is shown to be higher than the median. In regard to the meal price in detail, the mean is \$8.051, while the median is \$7.96, which indicates a right skew due to high-priced meals. Similarly, the total transportation price has a mean of \$29.03 and a median of \$28.60, with values approaching a maximum of \$123.89, possibly due to restaurants located farther from the Northwestern downtown campus, causing high transportation costs. The high-cost outliers—meal price at \$56 and distance being 18 miles—pull the mean in an upward direction, which causes variability within the dataset. We can see the spread shows nearby restaurants that are inexpensive, expensive, and in the middle, with different types of cuisines.

(1)(b) **(2 point)** Generate a table of counts using Category and Status. Add margins to this table. Lastly, present a barplot of these data; ignoring the marginal totals.

Table 3: Category and Status

	Expensive	Inexpensive	Normal	Sum
American	114	37	151	302
Brazilian	27	11	27	65
Chinese	30	9	30	69
French	23	7	24	54
German	21	9	24	54
Indian	24	9	28	61
Italian	61	13	56	130
Japanese	49	11	69	129
Korean	32	6	30	68
Mediterranean	42	25	53	120
Mexican	76	10	75	161
Peruvian	31	1	24	56
Spanish	18	8	27	53
Thai	54	9	59	122
Vietnamese	29	7	20	56
Sum	631	172	697	1500

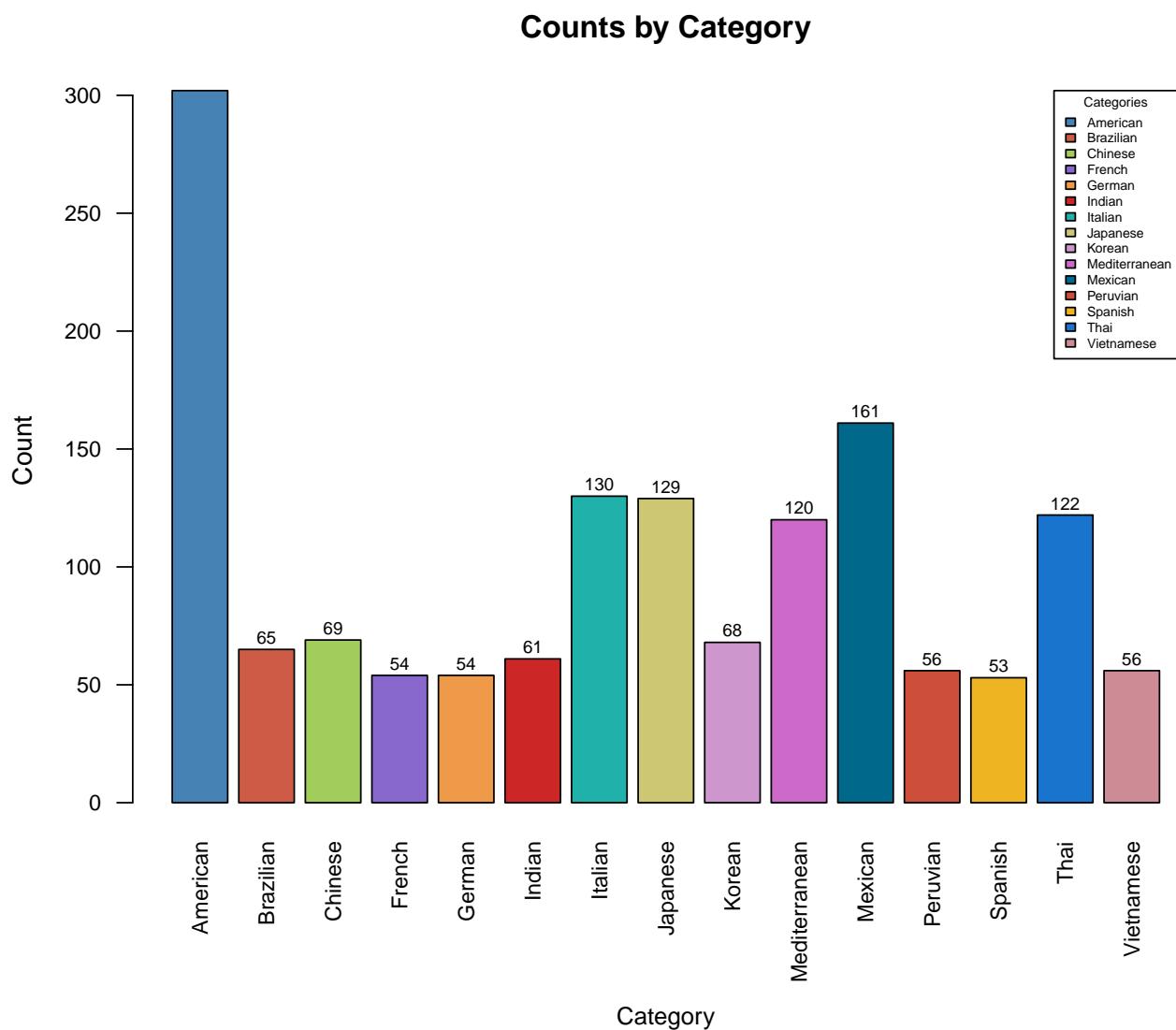
### Status of Cuisines



(1)(c) (1 point) Generate a table of counts using only Category. Present a barplot of these data, add a legend.

Table 4: Category Counts

Category	Count
American	302
Brazilian	65
Chinese	69
French	54
German	54
Indian	61
Italian	130
Japanese	129
Korean	68
Mediterranean	120
Mexican	161
Peruvian	56
Spanish	53
Thai	122
Vietnamese	56



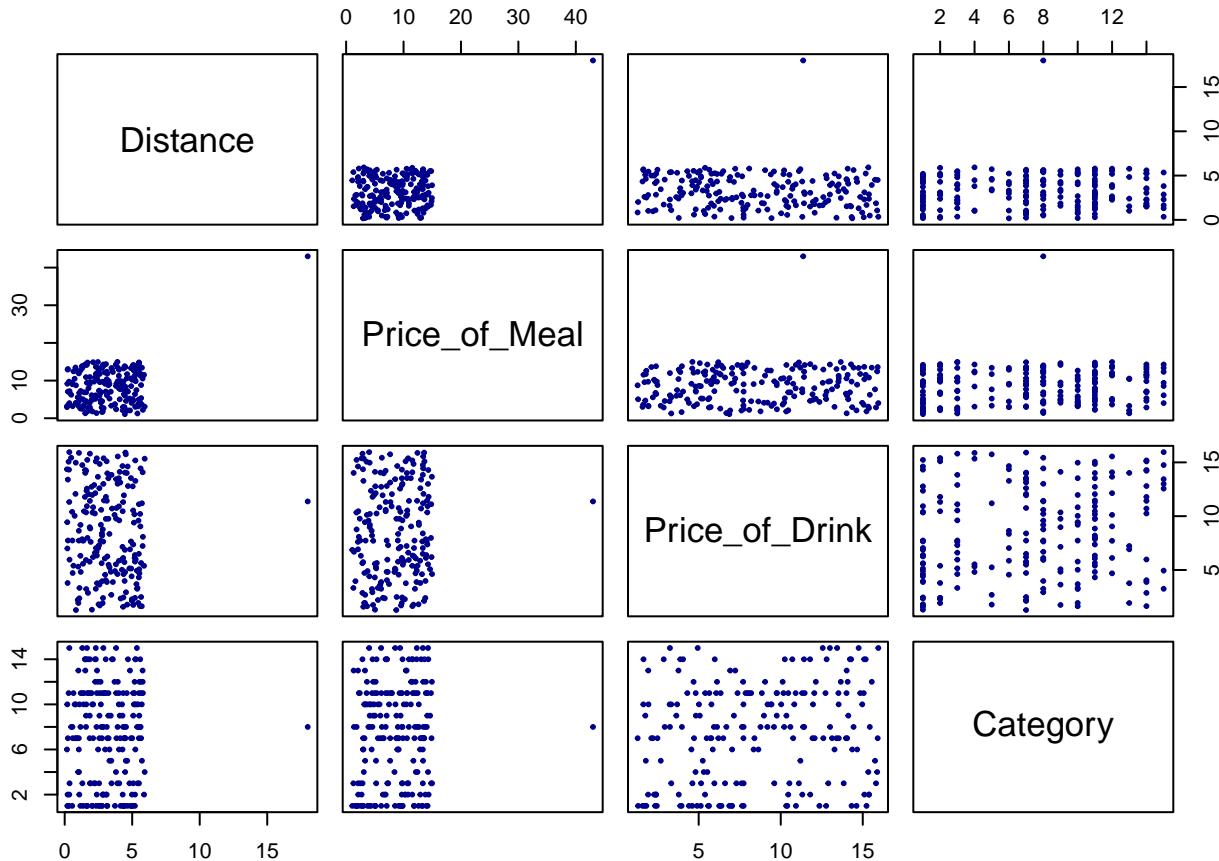
**Essay Question (2 points):** Discuss the Category distribution of restaurants first, what do you observe? Was that expected? What stands out about the distribution of restaurants by Status if any?

The category of the different cuisines of the restaurants shows a significant variation in the number of restaurants. There is the highest count of American restaurants at 302, followed by Mexican cuisine (161), Italian (130), and Thai (122). The trend overall shows Chicagoans' preferences or, in a way, the availability of different types of cuisines in a particular area according to distance. Spanish (53), Brazilian (65), and Peruvian (56) are among the cuisines with comparatively lower counts, suggesting that their availability is restricted. This indicates that Mexican food is very popular in the area around Northwestern Chicago.

One can gain a better sense of pricing levels by looking at the status distribution for each cuisine group. For instance, many restaurants serve American, Mexican, and Italian cuisines in a variety of pricing ranges (expensive, normal, and low), so people of all income levels can eat at the restaurant of their choice. Additionally, the more affordable options and limited price range of Peruvian and Brazilian cuisines suggest that their costs are more affordable. After consideration, the presence of multiple pricing points indicates a variety of restaurants catering to different income levels.

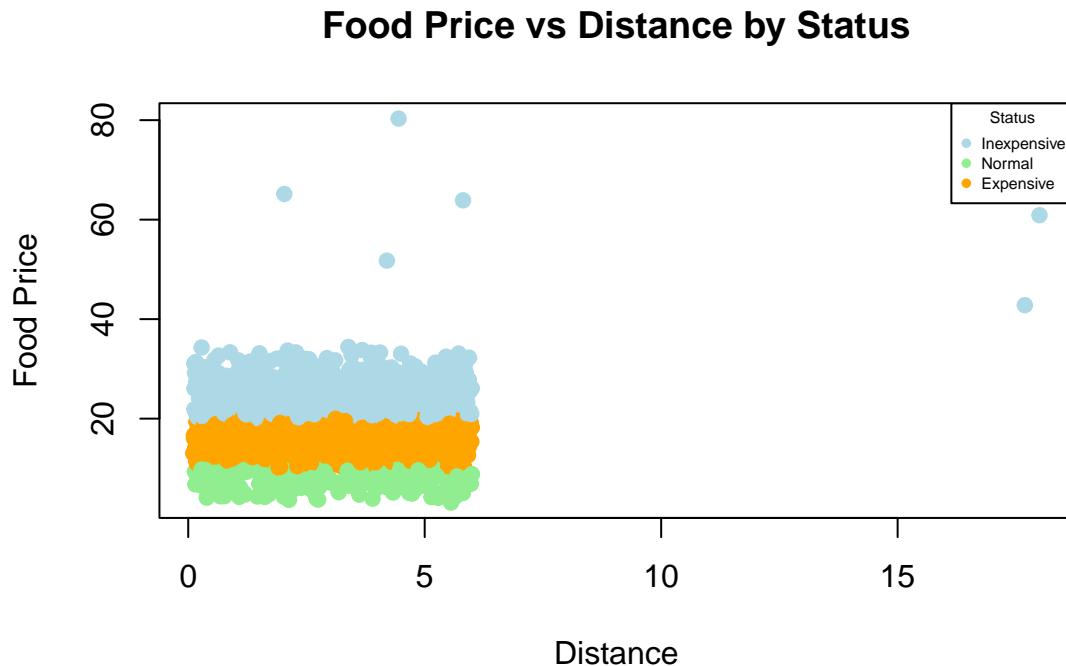
(1)(d) **(3 point)** Select a simple random sample of 200 observations from "mydata" and identify this sample as "work." Use `set.seed(123)` prior to drawing this sample. Do not change the number 123. Note that `sample()` "takes a sample of the specified size from the elements of x." Using "work", construct a scatterplot matrix of variables 2-5 with `plot(work[, 2:5])`

### Scatterplot Matrix: Distance, Meal Price, Drink Price, and Food Price



## SECTION 02: SUMMARIZING THE DATA USING GRAPHICS

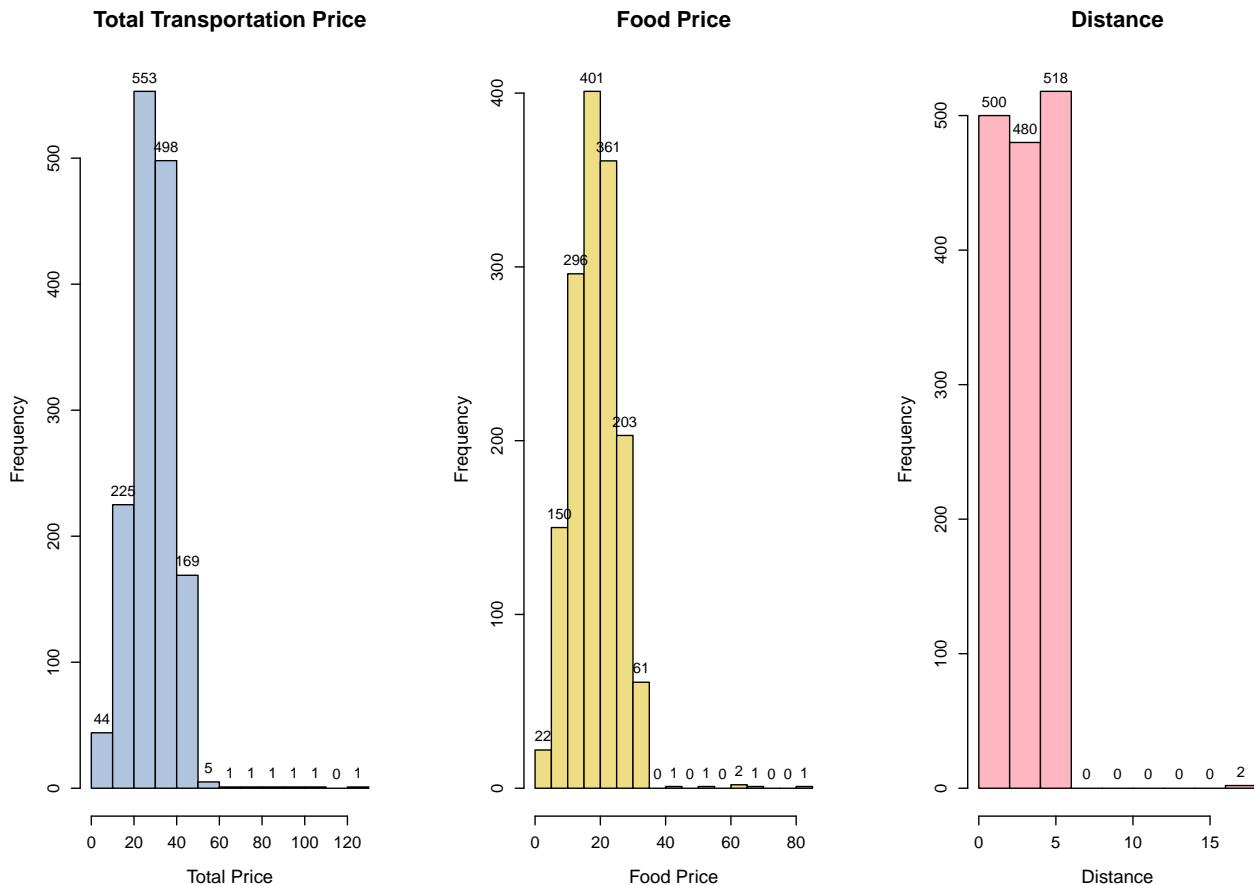
(2)(a) (1 point) Use “observations” to plot FoodPrice versus Distance. Color code data points by Status.



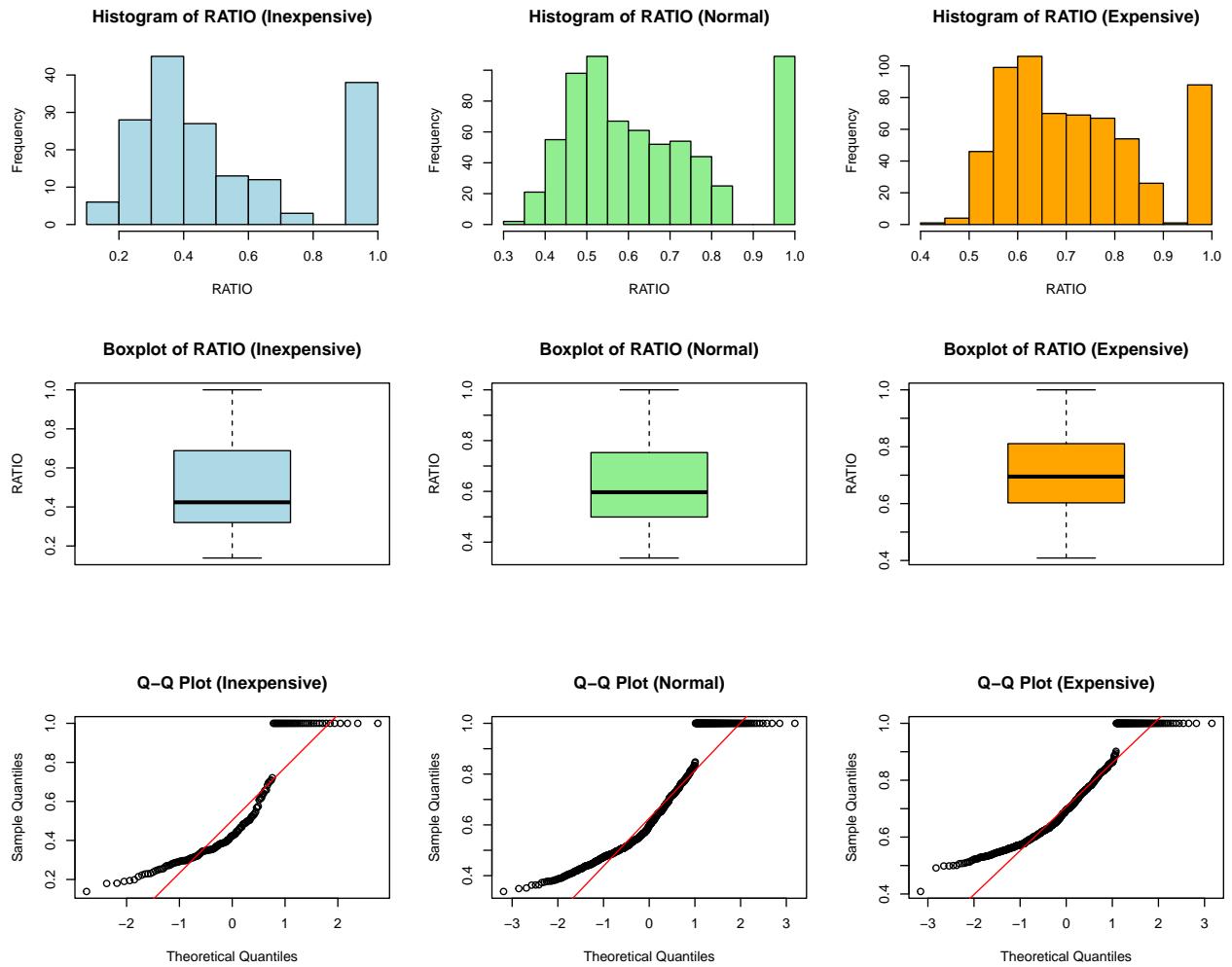
**Essay Question (3 points): Are there outliers? Are the outliers for all statuses or specific ones, can you explain the reason that these outliers are present?**

The scatterplot of the price of food vs. distance, specific to status, reveals distinct outliers, particularly within the “Normal” status category, with a few instances in the “Inexpensive” category as well. The outliers appear as data points above \$60, with some reaching just over \$80. The high food prices are due to restaurants located farther from Streeterville or the Northwestern downtown campus, which results in additional transportation costs, thereby increasing the total cost. Moreover, the outliers are not evenly distributed across all status categories. They appear mainly in the “Normal,” “Inexpensive,” and “Expensive” categories, which indicates a wide range in pricing even for restaurants considered “normal” and “inexpensive” or “expensive” based on their cuisine. The variability reflects the different types of cuisines and their categories, based on lower-priced to high-cost meals. The presence of these outliers highlights a range of restaurant options, from affordable to high-end dining experiences, within similar status categories.

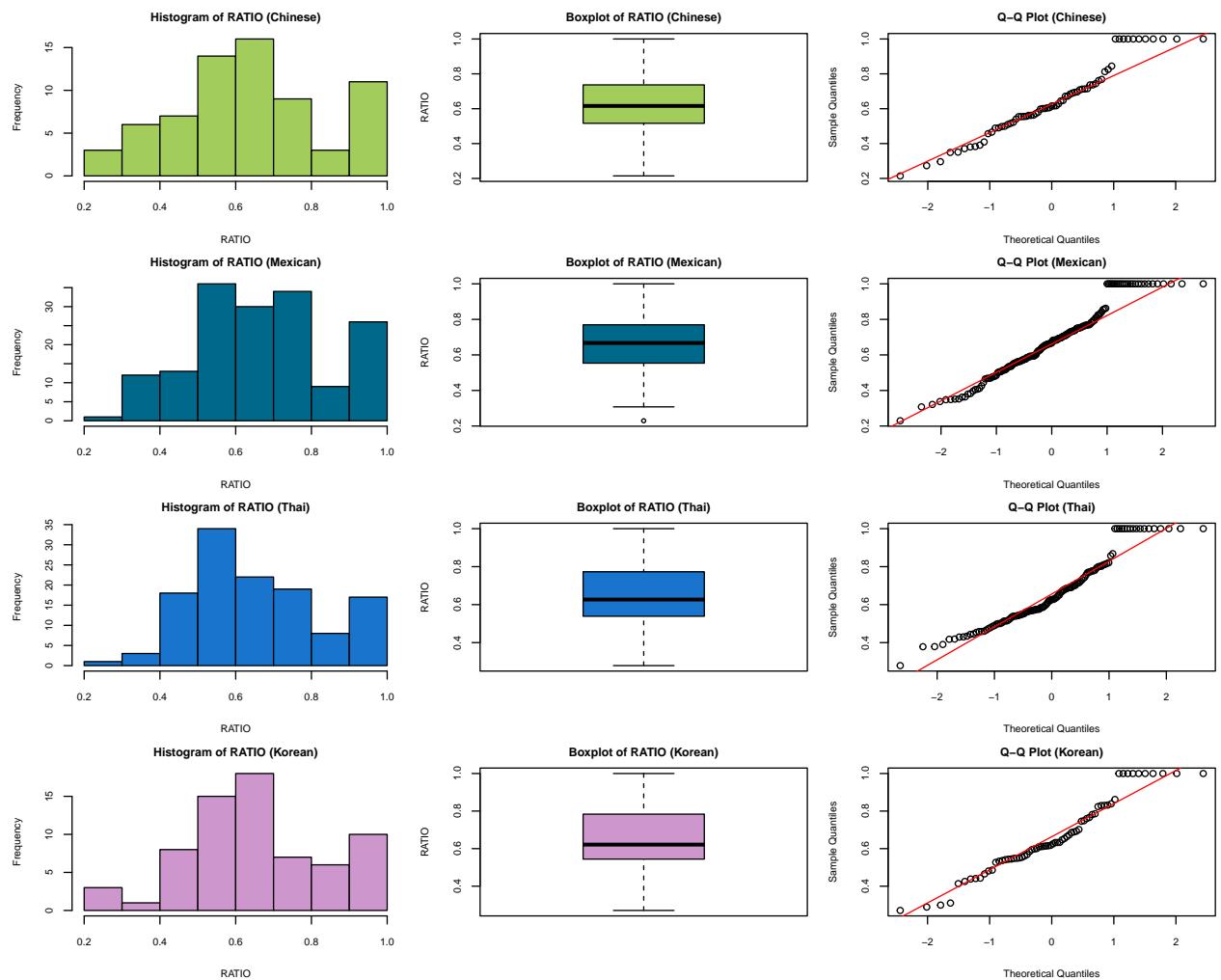
(2)(b) (3 points) Use “observations” to research the variables TT\_Price, FoodPrice, and Distance. Run histograms for all variables and identify the distribution they follow. Discuss Skewness for any of those. To further understand the FoodPrice, drill further in researching on the components of Meal and Drink Price. Use histograms and boxplots for this investigation. ( Free form analysis), that will help you to form conclusions.

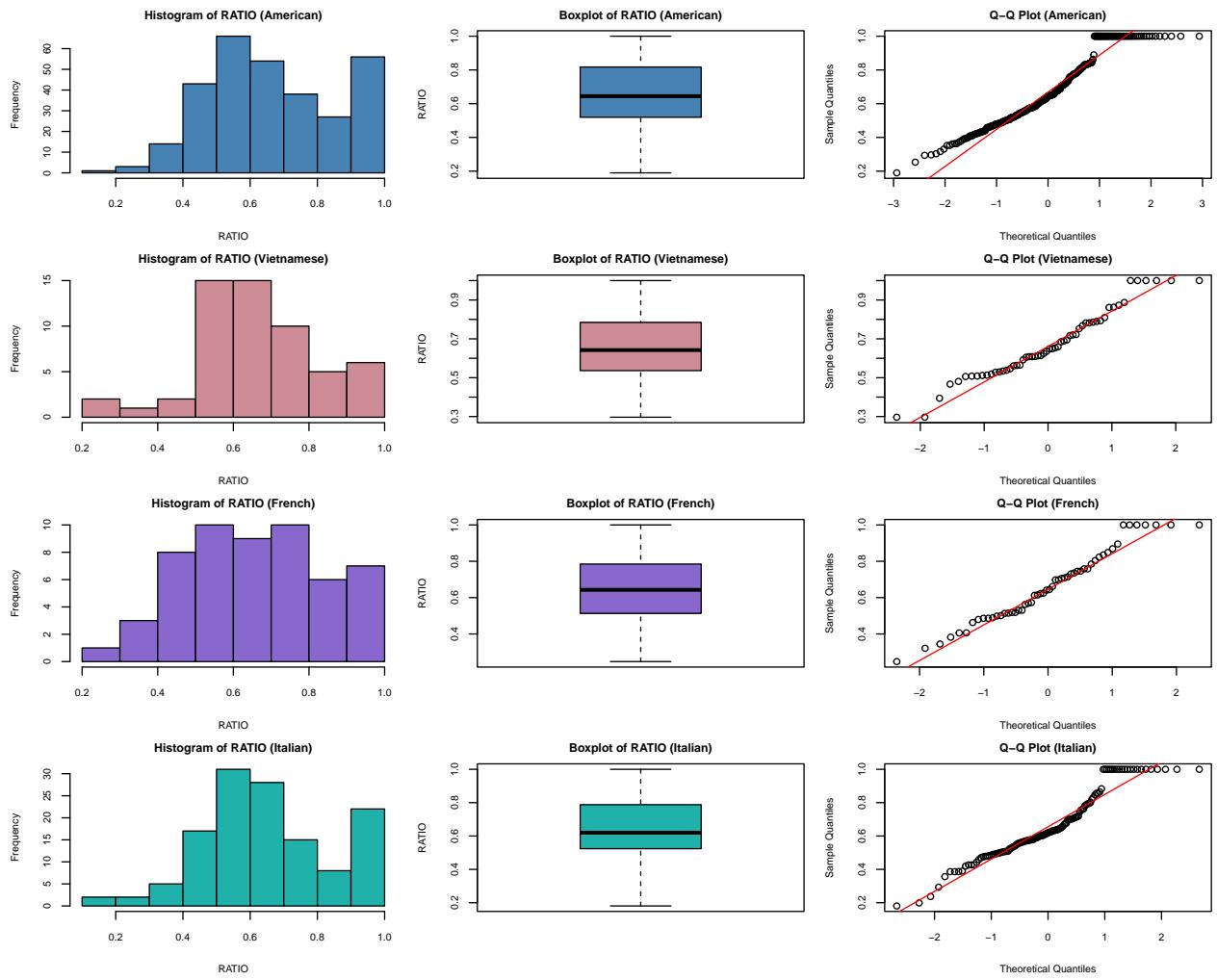


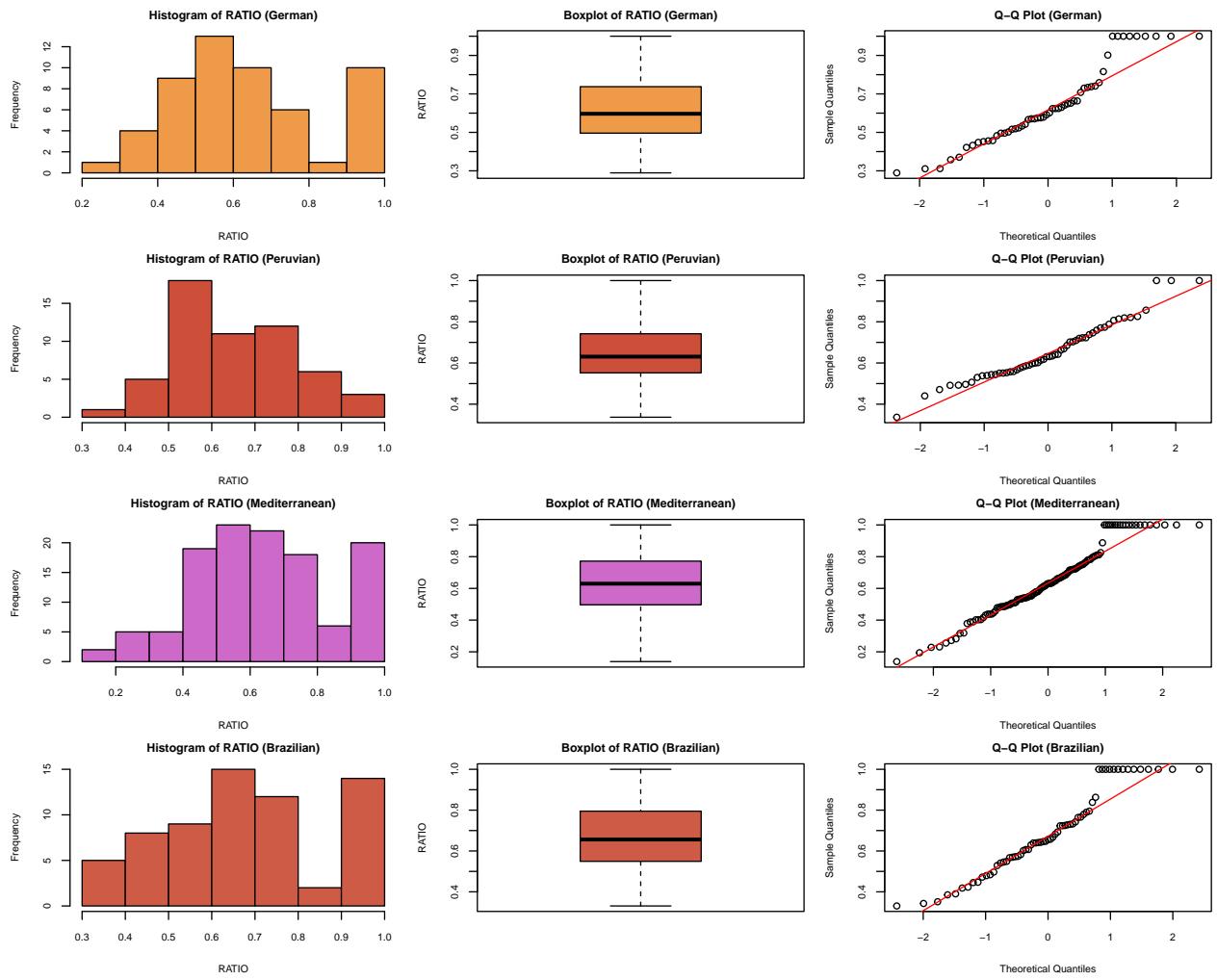
and ggplot2. The first row would show the histograms, the second row the boxplots and the third row the Q-Q plots. Be sure these displays are legible.

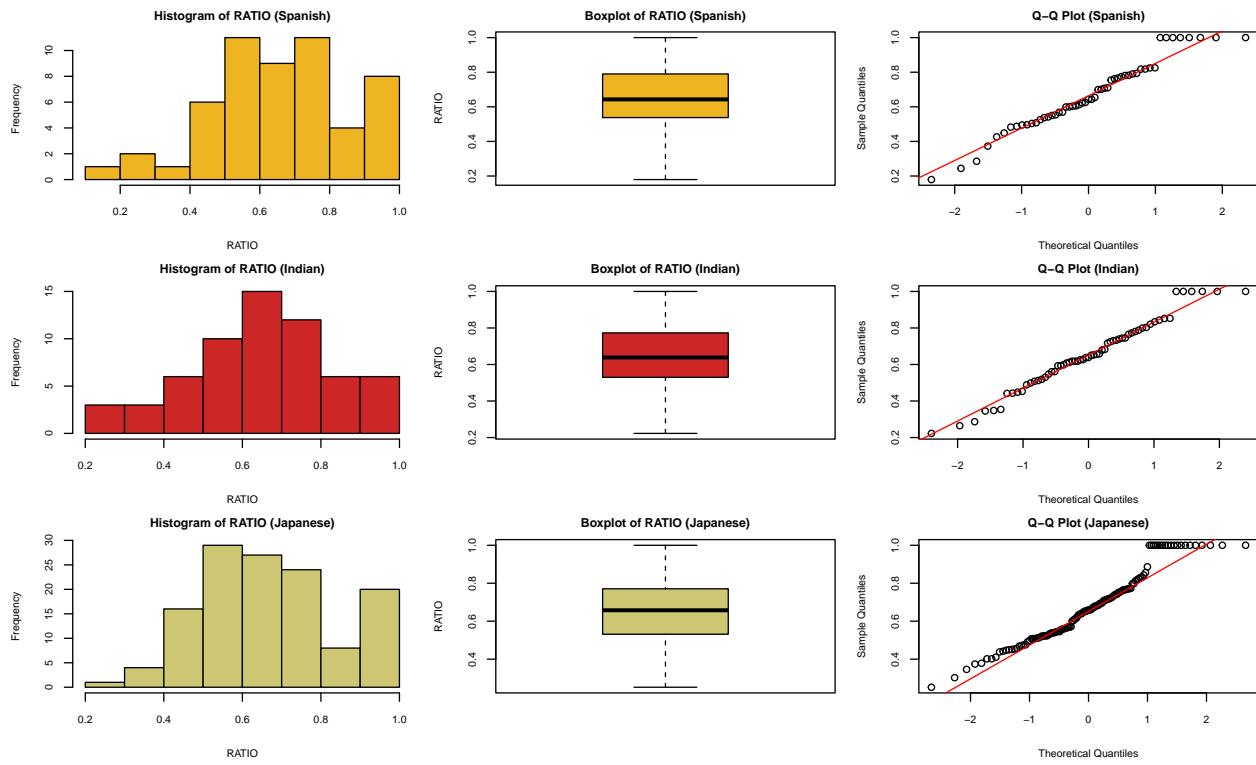


(3)(b) (5 points) Repeat the process for the step 3a but now replace the Status with Category.









**Essay Question (5 points): Compare the displays. How do the distributions compare to normality? Take into account the criteria discussed to evaluate non-normality.**

#### COMPARISON OF STATUS

The multi-figured plot created to understand the Status Ratio shows a distinctive pattern. Beginning with the Inexpensive category, the histogram is right-skewed, with most values clustering between 0.3 and 0.6. This indicates that inexpensive restaurants typically offer affordable meal prices relative to the total cost, which includes lower food prices and/or transportation costs. Near the 1.0 ratio, there is an outlier that points to the existence of more affordable restaurants. In the Normal category, the more uniform distribution covers a wider range of 0.3 to 1.0. The distribution is right-skewed, and there are further variations in the ratio of food to total transportation costs indicating major variations in meal costs.

The outliers that we see above the 0.9 threshold are caused by exceptionally high total prices, most likely from premium locations. Finally, the distribution falls between 0.5 and 0.9 in the Expensive category, with outliers at each extreme. There are a few outliers that indicate upscale restaurants with low ratios that are both near 1.0 and below 0.5.

This suggests that these restaurants are either very close to the university or follow a distinct pricing strategy, such as a fine dining experience. The boxplots tell a better and more emphasized story. The inexpensive boxplot shows a compact distribution with a range from approximately 0.3 to 0.6, which reinforces the histogram's understanding of inexpensive restaurants maintaining lower ratios. Outliers in the normal status exceed the 0.9 threshold, and the interquartile range is larger (~0.25). These outliers probably represent areas with high expenditures generally because of either premium costs or farther locations. Finally, the costly status shows a wide range of values between about 0.5 and 0.7, with some outliers at 0.9 and above and a few below 0.5. This distribution suggests that some expensive restaurants have incredibly high or low ratios, which may be due to their proximity to the institution or a combination of their fine dining experiences and/or unique pricing strategies.

#### COMPARISON OF CATEGORY

For the variable Ratio (Food Price / Total Transportation Price), each of the 15 distinct cuisine groups has

distinct distribution features, showing a range of normal and non-normal tendencies and highlighting the influence of outliers. A more symmetric distribution is seen in the Mexican and American cuisines, with central values ranging from 0.5 to 0.7. In contrast, the frequencies of Chinese and Mediterranean cuisines are greater, ranging from 0.3 to 0.6, and they exhibit a tilt towards the right. This implies lower ratios because Chinese and Mediterranean restaurants typically have cheaper meal prices or transportation expenses.

To further demonstrate the ratio of food price to total transportation cost, the boxplots above show different trends among cuisines. Restaurants based on cuisines of Mexico, Italy, and the United States had more outliers and a wider interquartile range. For example, outliers in the Mexican food category are close to 1.0, meaning that businesses with higher ratios either provide more expensive food or are farther away from Northwestern's central campus. Additionally, the boxplots for Chinese and Mediterranean cuisines are more compressed, suggesting less variability and a more consistent pricing pattern. However, even in these more stable categories, there are periodic outliers above 0.8, which may be the result of the expensive setting or unique ambiance of the restaurant.

The Q-Q figure illustrates how normalcy levels vary. The American and Mexican categories follow the line more closely and exhibit a more symmetric distribution, with occasional tail fluctuations. However, the Chinese, Thai, and Vietnamese groups show substantial deviations from the normal line for ratios below 0.6, indicating severe non-normality and right skewness. This is due to outliers that extends the higher tail in categories like Mexican and more affordable meals and transportation in these categories.

#### OVERALL COMPARISON OF STATUS AND CATEGORY

The data additionally shows how location and the wide variety of methods of pricing influence different cuisine groups, including Mexican and American. Overall, our analysis illustrates how pricing structures and location effects differ significantly across various types of cuisine groups, with some expensive outliers that exist with more affordable options like Chinese and Mediterranean and popular options like American.

In order to compare the ratio distributions with status and category, the study continues with a comparison of both non-normalcy and normality aspects. Understanding variation, outliers, and right skewness is vital for it. Approximately 1.0, 0.9, and 0.8 were the outliers for the status grouping for the normal, costly, and affordable groups, respectively.

Chinese and Mediterranean categories exhibit skewness with constant ratios, suggesting less price unpredictability, whereas American and Mexican categories have distributions that are closer to symmetry. The Mexican and Japanese categories, which displayed expensive locales, included a lot of outliers. Because ratio distributions are skewed across various cuisines and statuses, these patterns demonstrate non-normality. Therefore, both status and category show normality where they approach symmetry, whereas in the inexpensive category, Chinese and Mediterranean indicate non-normality due to lack of symmetry and skewness.

### SECTION 04: GETTING INSIGHTS ABOUT THE POSSIBLE PREDICTORS

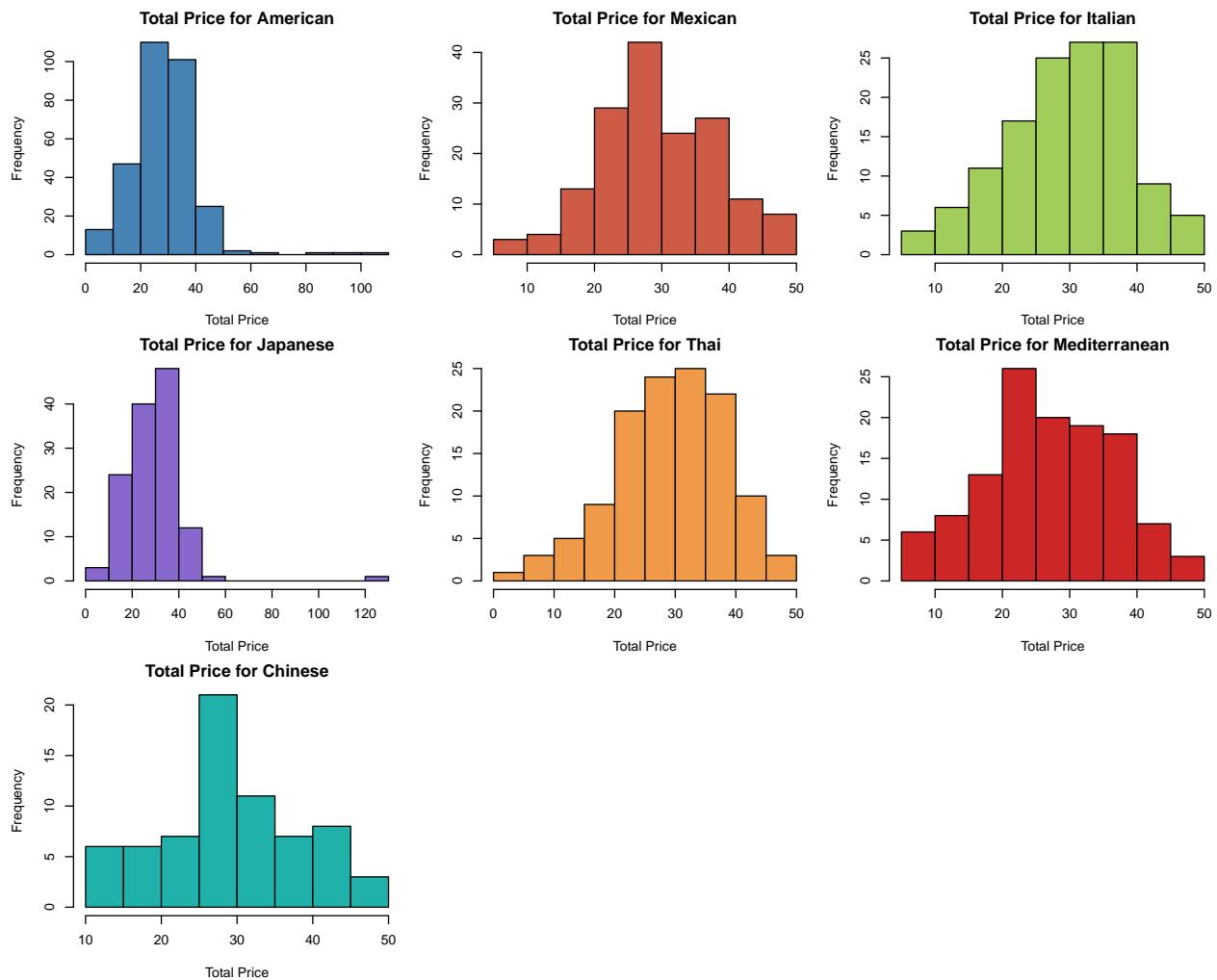
(4)(a) **(5 points)** Now we want to focus on the most common categories. Find the first 7 Categories in terms of the number of restaurants included to those. Run a summary() and histograms to identify and compare them. Create boxplots of TT\_Price by these 7 categories. To do this, you may want to identify those and then you can index the observations to get a working subset with only the observations of these categories. Also run boxplots for both TT\_Price and Distance by Common Category.

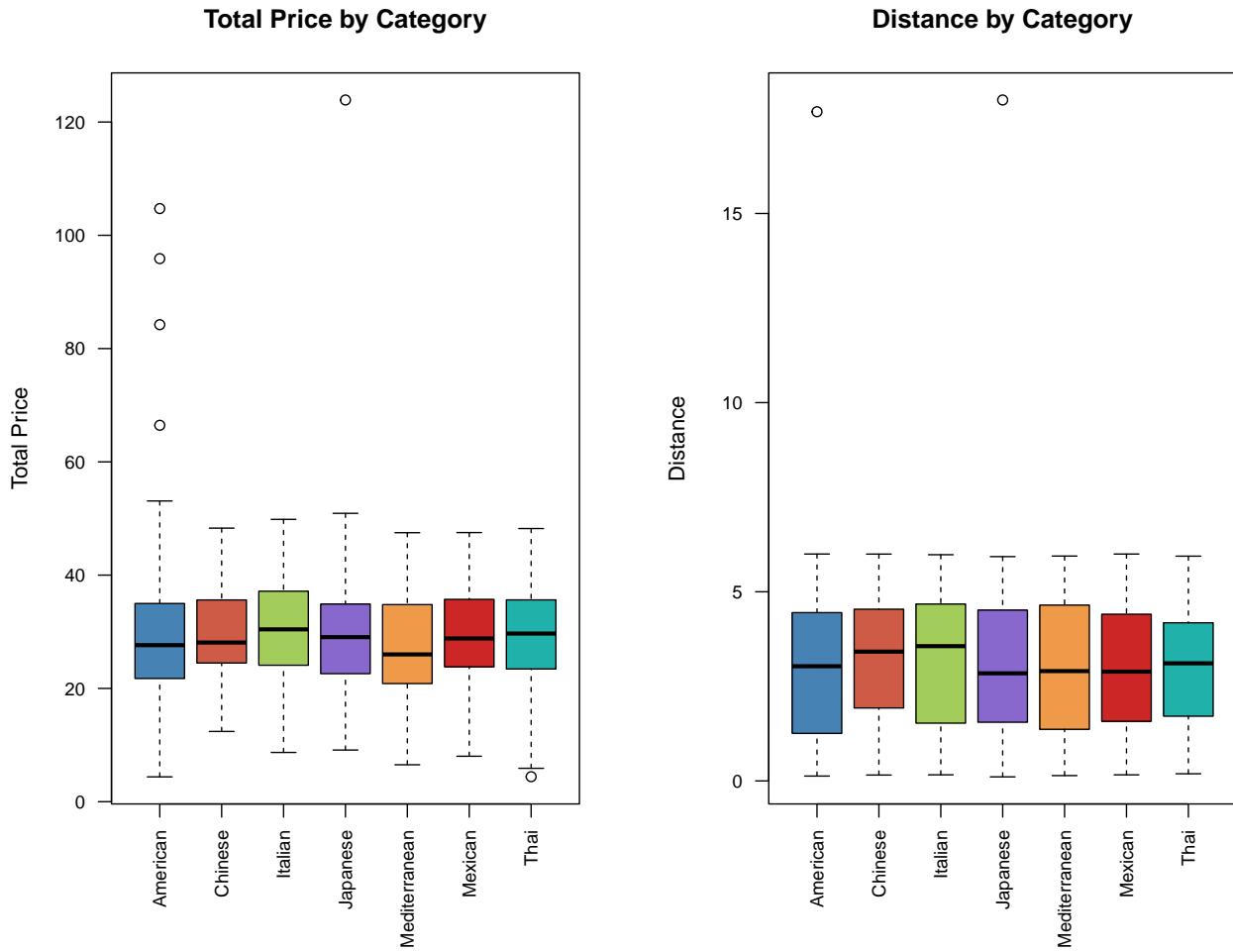
```
## [1] 0.2539486
##      Name          Distance      Price_of_Meal      Price_of_Drink
##  Length:1033     Min.   : 0.1068    Min.   : 1.000    Min.   : 1.280
##  Class :character 1st Qu.: 1.5243    1st Qu.: 4.580    1st Qu.: 5.040
##  Mode  :character  Median : 3.0248    Median : 8.000    Median : 8.310
##                  Mean    : 3.0487    Mean    : 8.142    Mean    : 8.464
```

```

##                               3rd Qu.: 4.4861   3rd Qu.:11.370   3rd Qu.:12.090
##                               Max.    :18.0000   Max.    :56.000   Max.    :16.000
## Category          FoodPrice      TT_Price      Status
## Length:1033       Min.    : 3.125   Min.    : 4.379   Expensive :426
## Class :character  1st Qu.:13.563   1st Qu.:22.531   Inexpensive:114
## Mode  :character  Median :18.368   Median :28.618   Normal    :493
##                           Mean    :18.598   Mean    :28.947
##                           3rd Qu.:22.982   3rd Qu.:35.427
##                           Max.    :80.315   Max.    :123.894
## Ratio
## Min.    :0.1384
## 1st Qu.:0.5241
## Median :0.6402
## Mean   :0.6655
## 3rd Qu.:0.7823
## Max.   :1.0000

```





**Essay Question (5 points)** Do you believe that these categories represent prices and distances of all restaurants well, or are there biases on their values?

I do not believe that these categories represent prices and distances of all restaurants well, as there are biases in their values. The seven most frequent categories analyzed—American, Mexican, Italian, Japanese, Thai, Chinese, and Mediterranean—showed a foundational overview of restaurant prices for the common ones. Price distribution is skewed, with most restaurants in the moderate price range and a few higher outliers initially focusing on mid-range costs. It is possible that this grouping will overlook other categories that provide less costly products or reflect unique price patterns. Additionally, the focus on these popular categories skews the representation, as it may exclude unique pricing and distance characteristics found in less frequent categories, such as Indian or French cuisine. The distance data, mainly within a 5-mile radius, is likely influenced by the proximity to Northwestern's campus, which might not reflect the broader Chicago restaurant landscape, particularly those farther from the center. This bias results in an incomplete view, as the concentration on common categories fails to capture the full range of prices and distances across all types of restaurants in the area.

## SECTION 05: GETTING INSIGHTS REGARDING DIFFERENT GROUPS IN THE DATA

(5)(a) (2 points) Use `aggregate()` with “observations” to compute the mean values of FoodPrice, Total Price (TT\_Price) and Ratio for each combination of Category and Status. Then, using `matrix()`, create matrices of the mean values.

Table 5: Mean Food Price by Category and Status

	Expensive	Inexpensive	Normal
American	25.92682	7.694400	15.68363
Brazilian	24.73499	8.538473	15.56966
Chinese	24.66949	7.257600	15.49221
French	24.87179	8.651200	15.61840
German	27.02880	7.165511	15.48447
Indian	24.29000	7.623467	15.97400
Italian	24.68994	8.042461	15.30820
Japanese	26.15749	8.597527	15.72756
Korean	24.44960	7.739200	15.30517
Mediterranean	24.25067	6.775104	15.37464
Mexican	24.83143	7.995680	15.19407
Peruvian	25.45579	8.164800	16.07060
Spanish	23.43973	7.120400	15.31413
Thai	24.67484	6.352889	15.33831
Vietnamese	25.83145	7.569600	15.57920

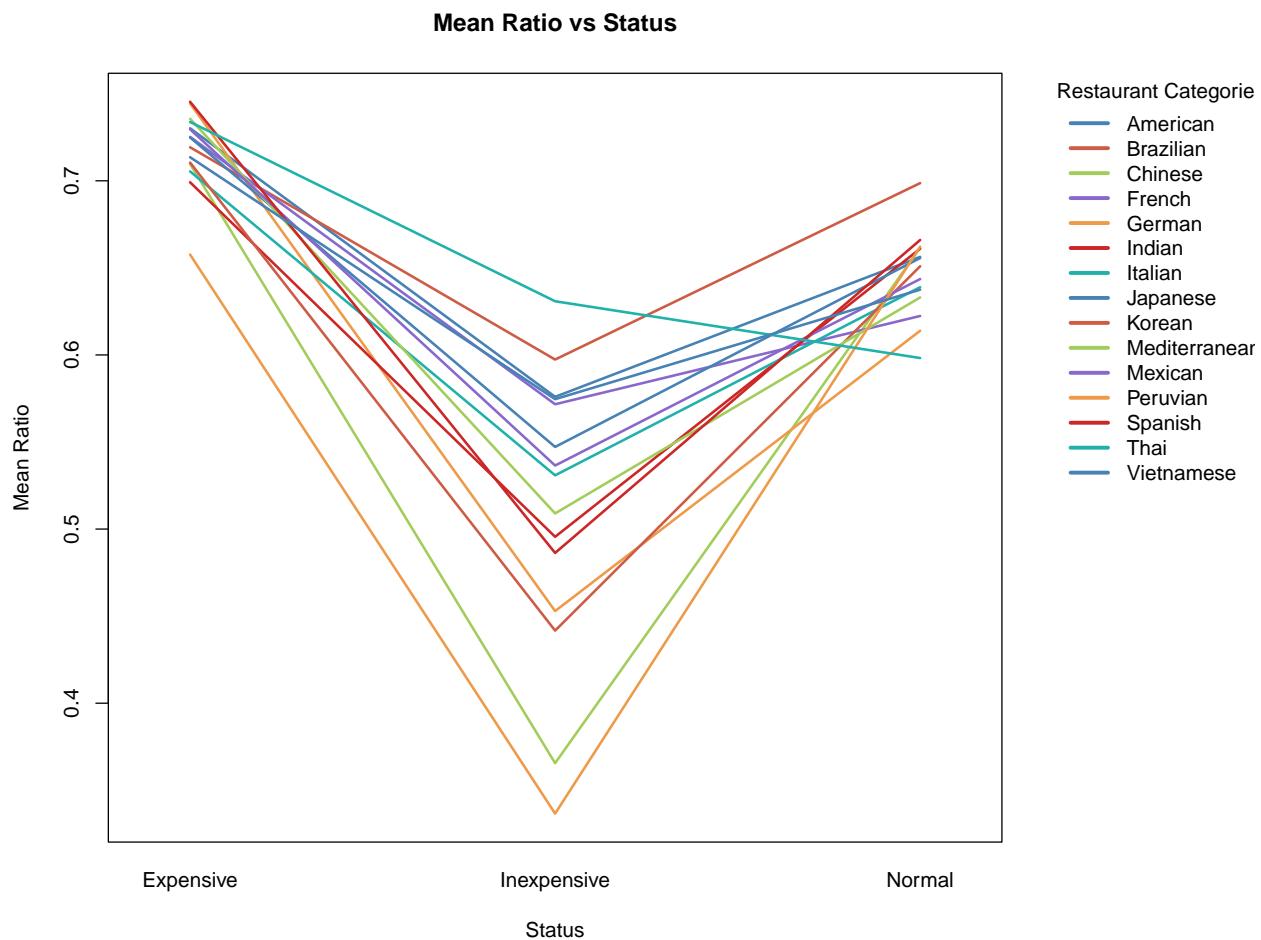
Table 6: Mean Total Price by Category and Status

	Expensive	Inexpensive	Normal
American	36.75438	16.80037	25.50683
Brazilian	35.69055	16.39421	24.21690
Chinese	35.98110	20.94730	24.98555
French	35.16581	19.33689	26.68316
German	37.46026	18.72008	26.46294
Indian	35.32786	19.43196	25.27885
Italian	36.03355	18.69537	25.84408
Japanese	37.55502	18.15672	25.73033
Korean	35.60245	21.18444	24.84102
Mediterranean	34.41559	16.41987	25.72178
Mexican	35.06276	18.76462	25.17744
Peruvian	39.14105	24.25578	25.38773
Spanish	32.55845	17.90086	24.31381
Thai	34.64510	11.53696	26.84235
Vietnamese	37.15587	14.04562	25.80418

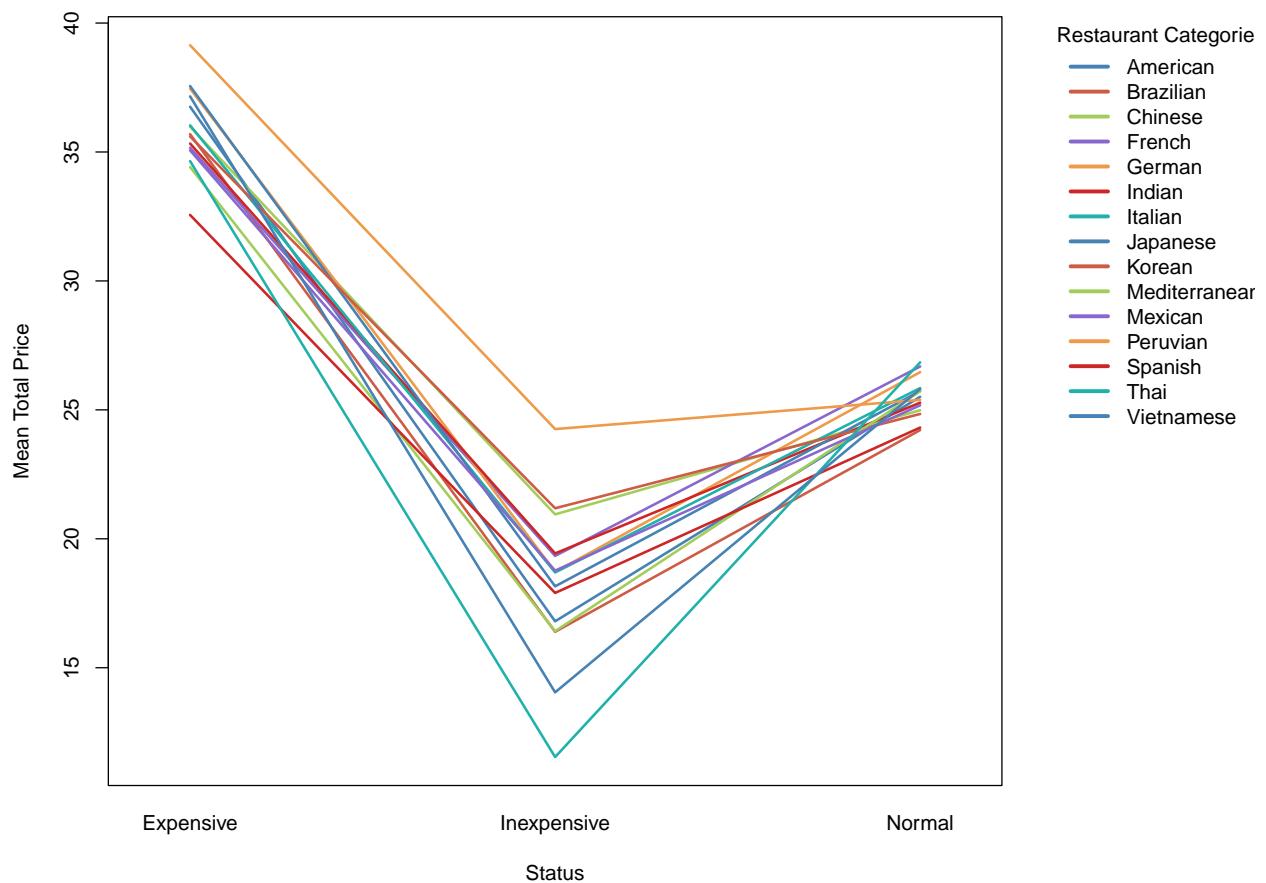
Table 7: Mean Ratio by Category and Status

	Expensive	Inexpensive	Normal
American	0.7301430	0.5759329	0.6563009
Brazilian	0.7192492	0.5973527	0.6986760
Chinese	0.7097093	0.3655694	0.6615340
French	0.7252118	0.5716840	0.6223786
German	0.7444220	0.4529178	0.6138806
Indian	0.6991462	0.4955675	0.6610121
Italian	0.7053896	0.5308661	0.6388957
Japanese	0.7249001	0.5471679	0.6556603
Korean	0.7104627	0.4416899	0.6509277
Mediterranean	0.7356227	0.5089925	0.6330252
Mexican	0.7295719	0.5364920	0.6435727
Peruvian	0.6576082	0.3366126	0.6621908
Spanish	0.7453898	0.4863019	0.6660011
Thai	0.7337822	0.6308264	0.5982128
Vietnamese	0.7135369	0.5746028	0.6374829

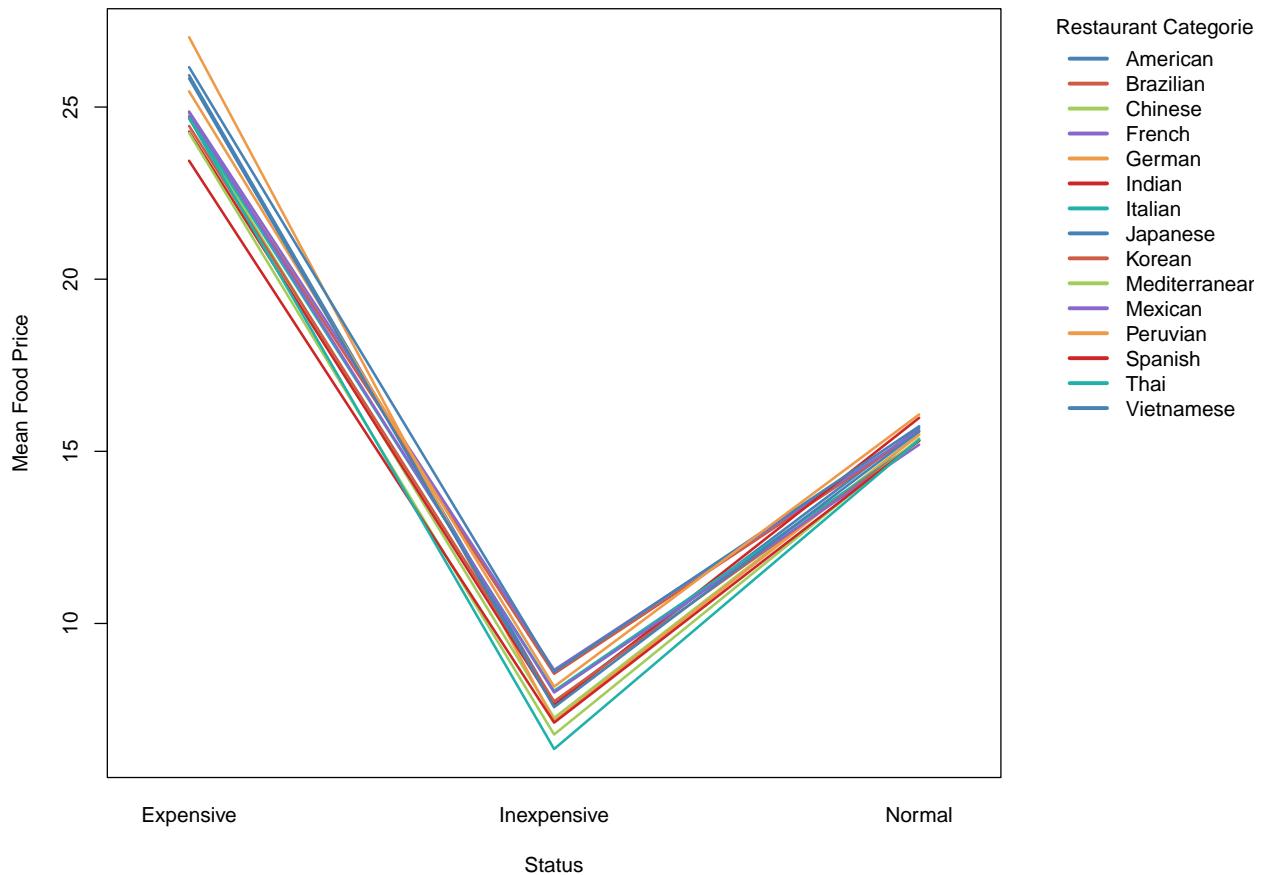
(5)(b) (4 points) Present three graphs. Each graph should include three lines, one for each Category. The first should show mean Ratio versus Status; the second, mean Total Price versus Status; the third, mean Food Price versus Status.



Mean Total Price vs Status



### Mean Food Price vs Status



**Essay Question (2 points):** What questions do these plots raise? Consider distance and category of restaurant.

A few questions arise during this research that focuses on features like food price, total transportation price, and ratio by category and status, specifically:

1. Does distance from the Northwestern downtown campus influence restaurant prices?
2. Are certain restaurant categories more inclined to offer expensive to inexpensive options based on location?
3. Do certain restaurant categories maintain similar price levels overall, regardless of distance, supporting consistent pricing?

**Essay Question (2 points):** What do these displays suggest about eating out expenses? Also, compare the ratio for each of the different categories, do they give an insight about locations?

The analysis in Tables 05, 06, and 07, along with the Mean Ratio, Total Price, and Food Price vs. Status, provides important insights into the expenses associated with different restaurant categories and their status. Table 05 illustrates the mean food prices by status for each cuisine. From this, we observe that German and Vietnamese cuisines have higher food prices even in the Normal status, showing that these cuisines carry a premium regardless of the status. Expensive restaurants display even higher average food prices; for instance, German cuisine is priced at \$27.03, while Thai in the Inexpensive category has a minimum of \$6.35. Given the impact of food prices alone, this implies that German and Japanese cuisines do not necessarily require the classification of "Expensive," as they are already pricy.

The overall cost, including transportation, is displayed in Table 06, which also provides insight into how distance may affect the price of food. Peruvian restaurants have the highest cost, at \$39.14, mainly because of their distance from the institution. For affordable meals, though, there's a lot of variation: Thai spots are the cheapest at \$11.54, while some Peruvian and Chinese places can go up to \$25.55. As inexpensive restaurants are far from Northwestern Downtown Campus, the transportation cost component is very important, contributing significantly to the total cost.

Table 07 and the Ratio vs. Status graph provide insights into how much of the total price is driven by food cost versus transportation. Higher ratios, like Spanish at 0.75, suggest that these restaurants are close to Northwestern University's downtown Chicago campus. In contrast, lower ratios at more affordable spots, like Peruvian at 0.34, indicate that these places are farther from campus, which increases transportation costs. The Ratio vs. Status graph highlights this pattern, showing how distance can impact dining costs, especially for budget-friendly choices. This pattern reinforces the idea that distance is a big factor in affordable dining, providing useful insights into how location can significantly affect the overall cost.

## Section 06: CONCLUSION FROM THE EXPLORATORY DATA ANALYSIS

*Essay Question (5 points) Based solely on these data, what are plausible statistical reasons that explain the expenditure of Chicagoans who reside in downtown on restaurant eating?*

The data clearly shows several factors that affect downtown Chicago residents' dining expenses, particularly for those who live close to Northwestern's campus. The type of cuisine is a crucial factor as some cuisines are always more expensive than others. For example, Table 05 shows that even when categorized as "Normal," German and Vietnamese restaurants typically charge higher prices. This indicates that these cuisines come with a natural price premium, so residents dining at these establishments pay extra, regardless of other factors. This price premium suggests that the choice of cuisine alone can elevate costs, particularly for those who favor distinctive, higher-cost dining options.

Transportation costs are another significant component. Table 06 reveals that as the distance from Northwestern's downtown campus increases, so does the total cost, especially for restaurants located farther away. For instance, Thai restaurants, which are usually closer, tend to be more reasonably priced since less money is spent on transportation, whereas Peruvian eateries, which are usually located farther from the university, display some of the highest total expenses. This means that people who live far away from Northwestern will most likely spend more time commuting, making them less inclined to dine at places farther away. Finally, Table 07 and the Ratio vs. Status graphs demonstrate how the restaurants' proximity to Northwestern's campus influences the food-to-transportation cost ratio. For example, cheaper and more convenient transportation lowers expenses for local establishments such as Spanish restaurants, which have a high food-to-total-cost ratio (about 0.75). On the other hand, areas further away, such as Peruvian destinations, incur more transportation costs, increasing the overall cost.

Residents of Chicago, particularly those who live closer to Northwestern's downtown campus, prefer to eat nearby to keep costs low, ensuring that the majority of the bill is spent on the meal itself. Overall, we can see that downtown Chicago residents' dining habits are influenced by the following factors: distance, cost, and preference of cuisine.

***Essay Question (15 points) Summarize your findings in a half a page paragraph***

Looking at the restaurant prices in Chicago, we can see some interesting trends, especially for people living near Northwestern's downtown campus. Table 05 shows that some cuisines, like German and Vietnamese, come with a noticeable price jump, even in the "Normal" range. For example, German restaurants charge around \$27 per meal, putting them on the expensive side, right up there with Vietnamese restaurants. Popular choices, such as American, Mexican, and Italian, are generally less expensive, whereas less common options, such as Peruvian and Brazilian, are more expensive. This pattern tells us that people who go for these unique, more expensive cuisines, like German or Vietnamese, are likely to end up spending more based just on their preferences.

Travel costs add another layer here, as shown in Table 06. We see a clear trend where the total dining cost rises with distance from Northwestern, with travel adding roughly \$3.50 per mile. Thai restaurants, which are often closer to campus, show a lower average total of \$11.54, while Peruvian spots, typically farther out, can hit total costs up to \$39.14. Histograms of these travel costs show a right skew, indicating that diners closer to campus generally pay less for travel. The scatterplot matrix backs this up, showing a positive link between distance and dining expenses, suggesting that folks may prefer closer spots to avoid extra travel costs.

Finally, Table 07 and the Ratio vs. Status graphs highlight how location affects the split between food and travel costs. For example, Spanish places have a high food-to-total cost ratio, around 0.75, which suggests that lower travel expenses mean more of the bill goes toward the actual meal. Meanwhile, spots farther out, like Peruvian places with a ratio near 0.34, show that travel costs are eating up a bigger part of the total. Boxplots, histograms, and Q-Q plots point out that distant spots have more cost variability and right-skewed distributions, showing the heavier impact of transportation. Overall, it appears that downtown Chicago residents' dining preferences are influenced by cuisine, travel costs, and distance. Those who live closer to Northwestern may pick nearby restaurants to make the most of their meal budget, but those who choose premium or farther restaurants must balance taste with their travel expenses.

## SECTION 07: APPENDIX

