Show me the Gratuity!

An analysis of restaurant wait staff demographics

4/10/2007

Introduction

In 2004, an estimated 2 million people were employed as waiters or waitresses (Lynn, M., 2004). Restaurants rely heavily upon this class of workers to provide the seamless connection between restaurant cuisine and an enjoyable eating experience for the customer. Often, the waiter/waitress has a large impact in determining a customer's opinion of the restaurant thus affecting the likelihood of a repeat customer. Likewise, the waiter/waitress has a vested interest in keeping the customer happy in that the waiter/waitress receives gratuities from patrons based on the quality of service the customers receive. Moreover, happy customers are more likely to give larger gratuities (Lynn, M., 2004).

Many factors such as wearing flair, squatting at the customer's table, smiling at the patrons, and telling jokes (to name a few) have been shown to affect the amount of gratuity that a server will receive from their customers. Many books and articles have been published to inform waiters and waitresses how to utilize these attributes in order to increase their gratuities. However, to the authors' knowledge, little information has been published on the demographics that give the largest gratuity percents. This information could be utilized by new restaurant waiters/waitresses that would like to maximize the gratuity potential. Therefore, the objectives of this study were to determine the demographics of region, race, gender, marital status, and type of restaurant that receive the largest gratuity percentage.

Questions

In identifying the underlying objective of our study, we set out to determine where a entrepreneur in the food service industry could obtain the highest yielding percent gratuity. To obtain our objective we set out to answer the following questions:

- What factors affect tipping percentage?
  - Region
  - State
  - Sex
  - Marital status
  - Type of restaurant
- Do certain restaurant chains perform differently in each region?
- Is there an explanation to why some of these issues may arise?
- Can we predict what the tipping percentage would be for a server in a type of restaurant based on its location in the U.S?

Data Collection and Restructuring

Data for this project was collected by Dr. Wm. Michael Lynn of Cornell University during the summer of 2006. Participants in the study used a web-based survey form to answer approximately 75 questions regarding past, current, and future employment in the food service industry. This data was then compiled into an Excel spreadsheet by Dr. Lynn for further analysis.

Like most other survey datasets where information is solicited from participants with no incentive to report information, this dataset contained many missing data points and improbable data

as a result of entry errors. A total of 2618 observations were reported using the web survey form, yet, all of this data is not probabilistic. For example, one participant reported 500% tips at Bruggers Bagels in Raleigh, NC, and two more participants reported 100% tips percentages at a Ponderosa Steakhouse and IHOP in Ohio and South Carolina, respectively. This data is unrealistic and, for this study, will be deleted from any further analysis. Conversely, participants that reported tipping percentage values of less than 5 % will be removed from further analysis. Dr. Lynn truncated his dataset at ~10%; however we felt that keeping the lower tail of tipping percentage might give light into relationships of tipping and gender, race, and server performance. To complete the objectives of this project of evaluating factors that affect tipping, some variables (remoteip, datercvd, submit_time, and more_mos) were removed from the dataset due to their unimportant and independence to the dependent variables. These variables are not expected to contribute to the understanding of variables that affect tipping and thus removed.

Aside from removing variables with no significant contribution to the study, new variables were formed to more broadly categorize the data. The variable 'State' was converted to its US postal service state abbreviation as listed on the USPS website (www.usps.com). Each state was then classified into the corresponding region of the US as determined by the US Census Bureau. The proportion of each race (asian, hispanic, black, and white) was classified into 10 categories (i.e. <10%, 10-20%, 20-30%, etc.) to evaluate trends of tipping as a particular proportion of a particular race increased. Additionally, a new variable of 'type of restaurant' was formed by classifying the restaurant name into the following categories: American, Asian, Bar, Café, Chain, Club, Deli, Family, Grill, Hotel, Pizza, Steak, Italian, Mexican, Seafood, Specialty, and Steak Chain. The following rules determined the classification of each restaurant into the 'type' variable. First, any duplicate restaurant names were found and were classified as a "chain" if they were located in a different state. Secondly, if the restaurant name contained any of the classifying variables, then those restaurants were automatically classified into those variables. Only after all other possibilities were exhausted were the remaining restaurants classified as 'specialty'.

Data Analysis

Statistical analysis for this dataset will be performed using two statistical software packages, R (http://cran.r-project.org) and SAS (SAS Institute, Inc., Cary, NC). The GLM procedure of SAS was used to estimate LS means. Region, gender, and marital status were used as fixed effects to predict tipping percentage. Means for each proportion of race were calculated by Excel and graphed using a PivotChart. Further graphical representations of the data were made using Excel and R for interactions and relationships of interest.

Results and Discussion

Region of Restaurant Location

After deleting and restructuring of the dataset, there were a total of 2181 observations used for the analysis. In the analysis of state effects, the dataset did not provide enough observations per state to achieve the desired resolution. Therefore to retrieve value from this particular aspect of the study, the states of which the restaurants were located were classified into their respective U.S. regional location. This classification was performed in order to gain more power in the analysis of some very general, but extremely interesting questions.

Figure 1 illustrates the average tipping percentage received for services by waiters and waitresses in a particular region of the US. The region of restaurant location significantly ($P<0.001$) affected the tipping percentage that a waiter received. Servers in the Northeast received, on average, 2.5% larger tips than servers in the Midwest region and approximately 1% larger tips than the South and Pacific regions. The West region did not statistically differ from the Northeast region on tipping percentage. This result could be explained by the type of people and the average wages of customers in each region. Even though the consumers in the Northeast region typically have higher annual salaries and higher costs of living, Northeasterners frequent restaurants more often. This could affect tipping percentage as customers are more likely to appreciate good service and express that appreciation in the form of larger gratuities. Contrarily, the frequency of which Midwesterners eat at restaurants is lower than other regions, and they could not appreciate quality service as much.

Figure 1. Average tipping percentage received by servers in a specific region of the US.
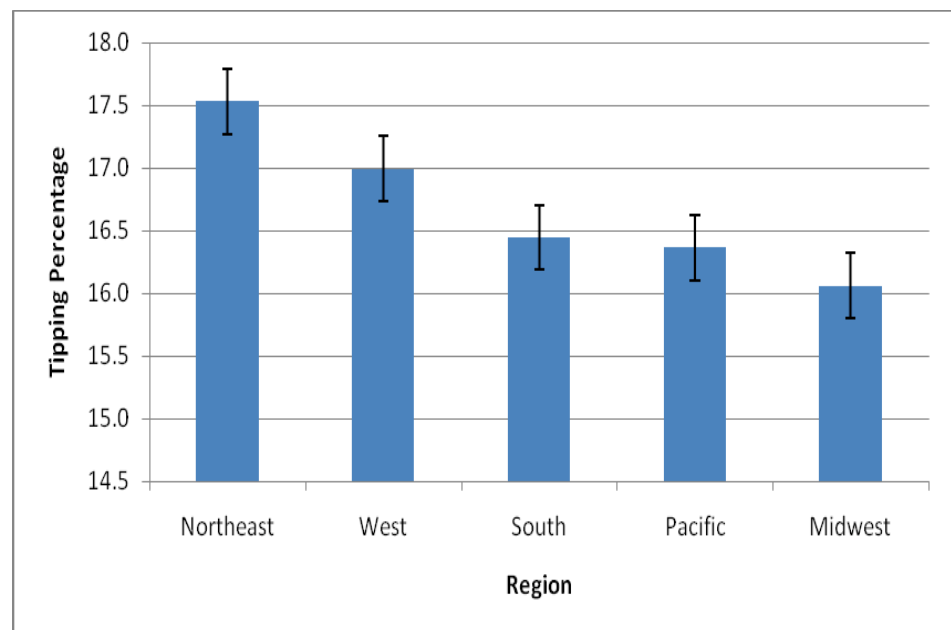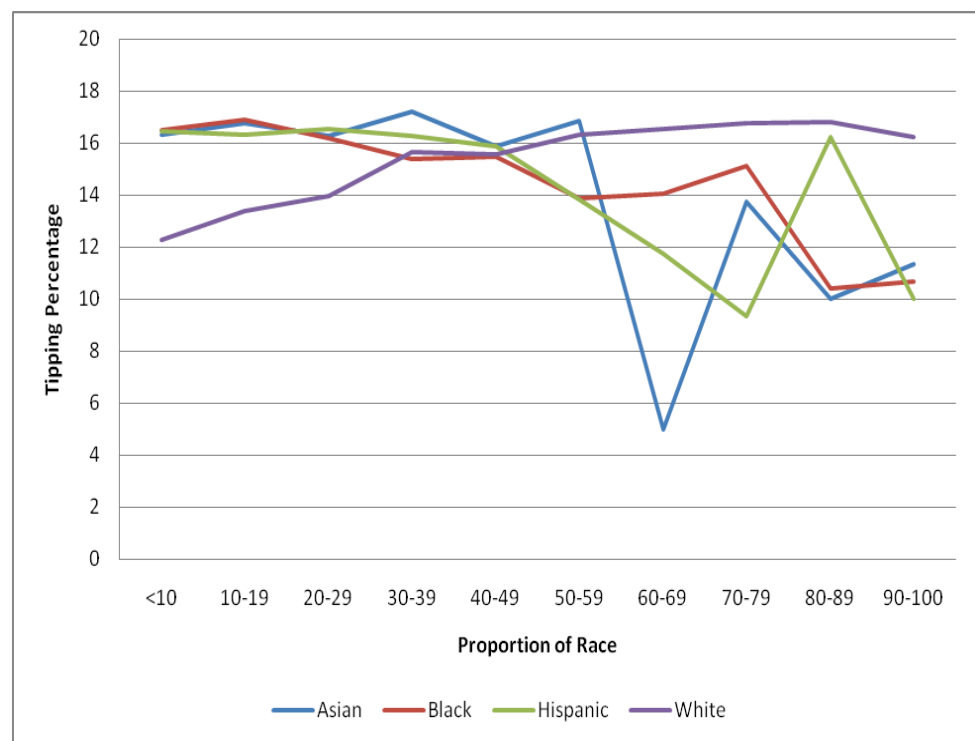

Proportion of Race

The average customer tipping percentage based on proportion of asian, black, hispanic, and white people that visit a restaurant is shown in Figure 2. A positive linear trend exists between the proportion of white customers that visit a restaurant and tipping percentage. Contrairly, negative linear trends were found between proportions of asian, black, and hispanic races and tipping

percentages received by the server. While the asian, black, and hispanic races follow general linear trends (Figures 3 and 4), a large amount of variantion exists in average tipping percentage, especially as the proportion of these races become the majority proportion served at a restaurant.   For example, the hispanic race aveages approximately 10% when the restaurant serves between 70-79% hispanic populations. The tipping percentage increases to 16% for a 80-89% hispanic populaiton and then decreases back to 10% for a 90-100% hispanic populations. This event could be due to the low amount of data collected

Figure 2. Average customer tipping percentage compared to the proportion of race served at a particular restaurant.

Figures 3, 4, and 5. Average tipping percentage by proportion of Asians (Figure 3.), Blacks (Figure 4.), and Whites (Figure 5.) served at a particular restaurant.
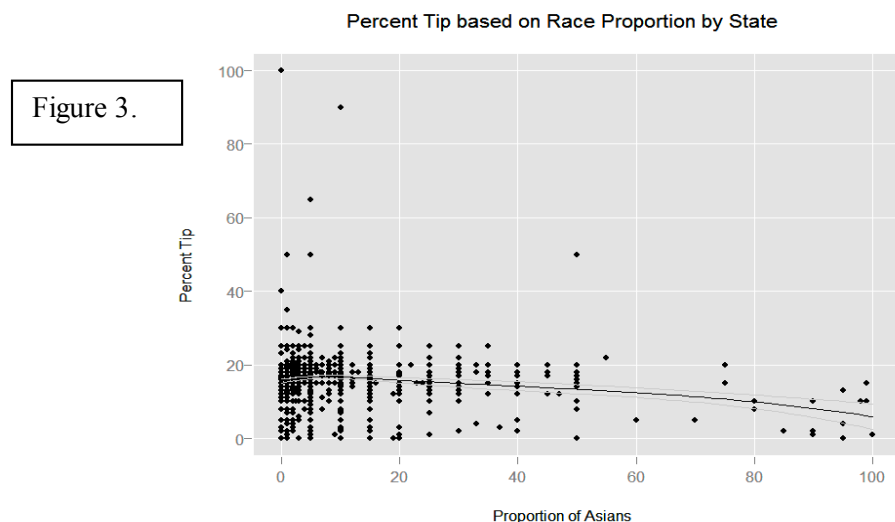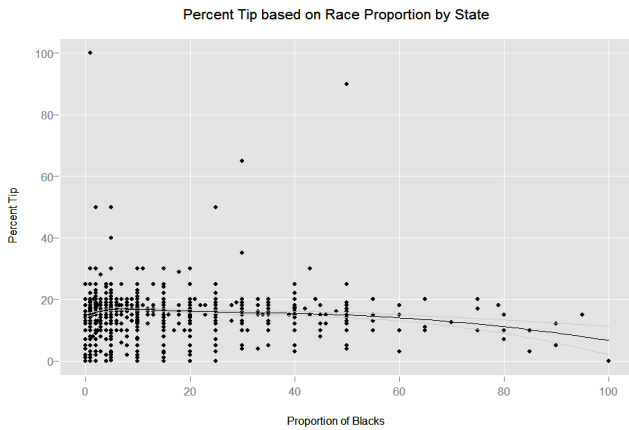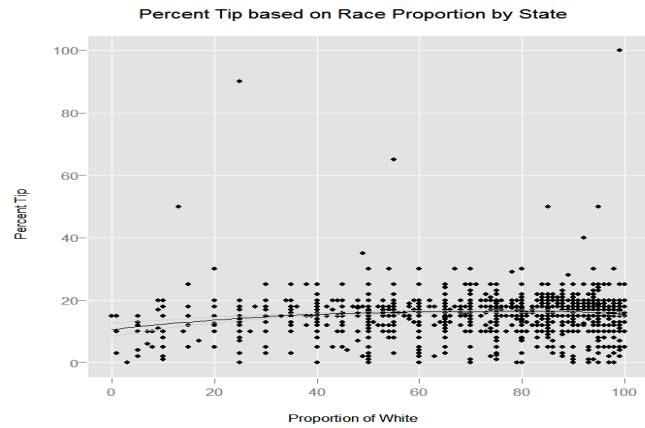

Figure 3.

Figure 4.

Figure 5.

Percent Tip based on Race Proportion by State

Percent Tip based on Race Proportion by State

Interaction of Region of Restaurant Location and Proportion of Race

As previously reported, the hightest tipping percentages have been a result of waiters and watiresses working in the Northeast region and serving primarily a white population. The least gratuity percentage was found in the Midwest region of the US. Figure 6 illustrates the proportion of asians, hispanics, blacks, and whites being served at restaurants in a specific region of the US. The proportion of each race is logical in terms of racial location in the regions of the U.S. For example, the Pacific region has the highest Asian population and the South is the greatest population of black citizens. Interestingly, restaurants in the Northeast and Midwest regions of the US serve approximately the same proportion of each race, yet have had significantly large differences in tipping percentages. From this conclusion, it is suggested that the cultural habits of a particular region plays a larger role in determing the estimated tipping percentage as compared to race. Yet, because of differences in traditions and dietary eating habits within each region, one could expect a large amount of variation in tipping percentage within each region.

Figure 6. Proportion of race served at restaurants in a particular region of the US.
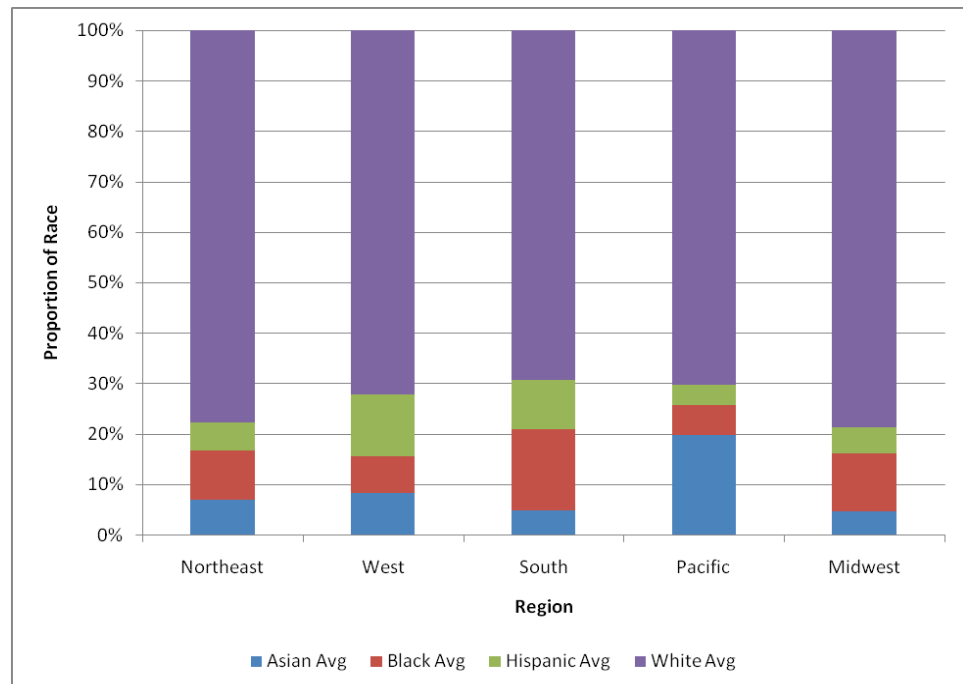
The finding that the difference between regions is

attributable to cultural differences rather than differences in race populations is supported by evaluting the variances for each region. Average variance for tipping percentage by region of the US is shown in Figure 7. The aveage variance for the Northeast region is 9.57 whereas the aveage variance for the Midwest region is 13.04. The region with the largest variance in tipping percentage was the Southern region of the US of which those restaurants served the most diverse populations of race, whereas the white proportion comprised the largest percentage of people in restaurants in the Midwest and Northeast regions.

Figure 7. Average variance in tipping percentage in a particular region of the US.
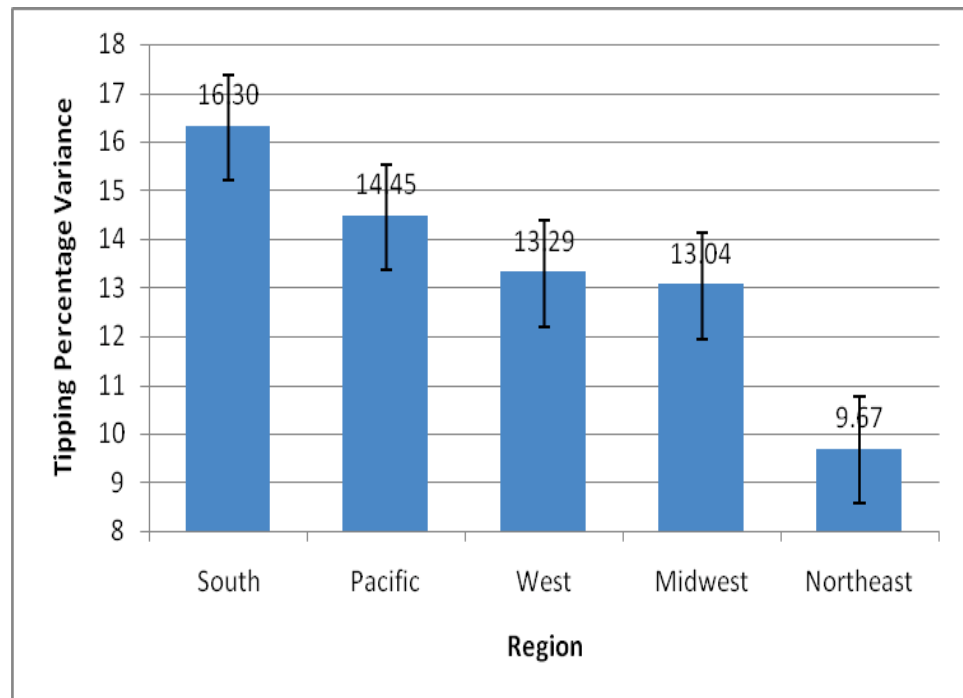
Interaction of Region of Restaurant Location and Proportion of Race

The proportion of black people being served in restaurants in a given state appears to affect the percent tip that a server may receive. Typically, the states are fairly evenly distributed based on their grouping with an approximate average between 15 to 20% gratuity. However, some of the outlying states such as Georgia (GA), Louisiana (LA), and Nebraska (NE) show that there is some variance in the data. Louisiana on average has the highest tipping percentage and the second highest proportion of black people in their population. In addition, Georgia has the lowest tipping percentage based the number of black people in their population.

Figure 8.
Average tipping
percentage based
on proportion
of the Black
race.
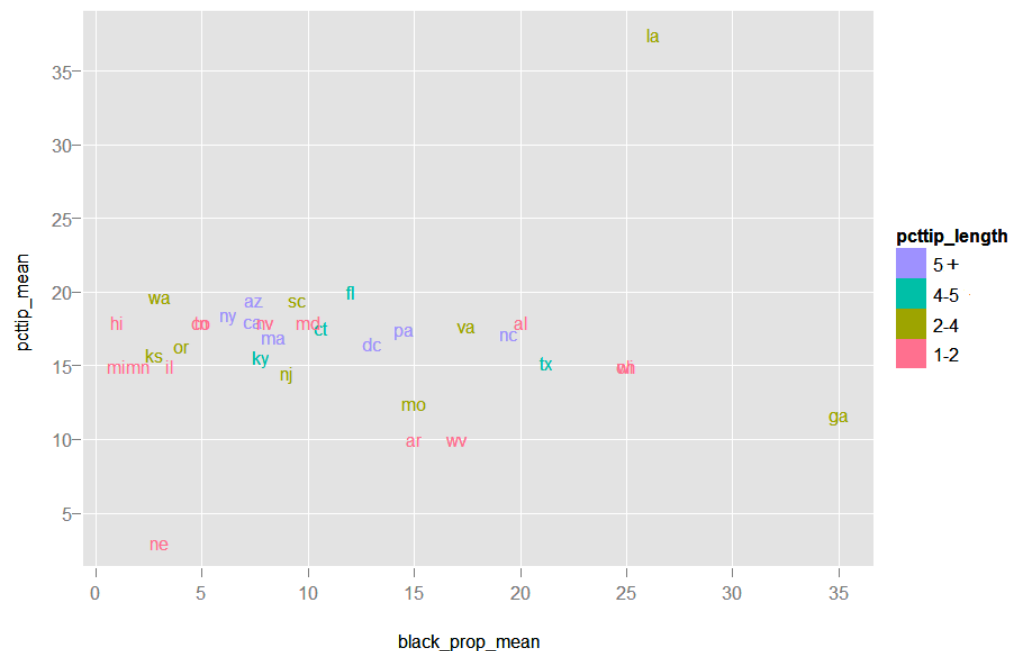
Further
analysis of the affect of race on percent tip reveals numerous trends in the data set. Similar to the
proportion of blacks in the population, the number of Asians and Hispanics also shows a trend toward
a reduction in percent tip as proportion increases. The proportion of Asians in Hawaii illustrates
that this population still maintains an average level of percent tip based the race of the population.

Figure 9. Average tipping percentage based on proportion of the Asian race.
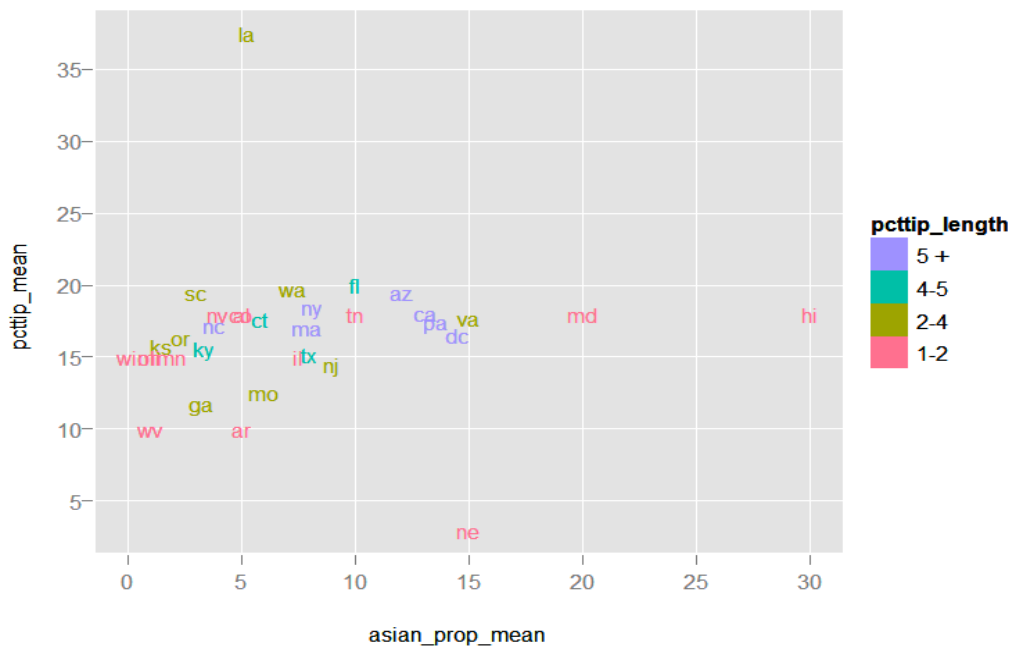
States with the largest percentage of white people are New Hampshire (NH), Michigan (MI), Minnesota (MN), and Kansas (KS). Similar to the Black and Asian races, Louisiana (LA) ranks the highest in tipping percentage with Florida (FL), Washington (WA), Arizona (AZ) and South Carolina (SC) following in second through fifth place. Consistently, Nebraska (NE) has the lowest tipping customers regardless of race. This possibly contributed to the regional differences found between the Northeast and Midwest regions in that all other states in each respective region are clustered relatively close together. When Nebraska observations are averaged into regional data, the consistently lower state of Nebraska decreased the overall mean. There must be some caution taken in interpreting these results since the data contained extreme values. This made determining trends in the data difficult because one state could affect the entire region.

Figure10. Average tipping percentage based on proportion of the Black race.
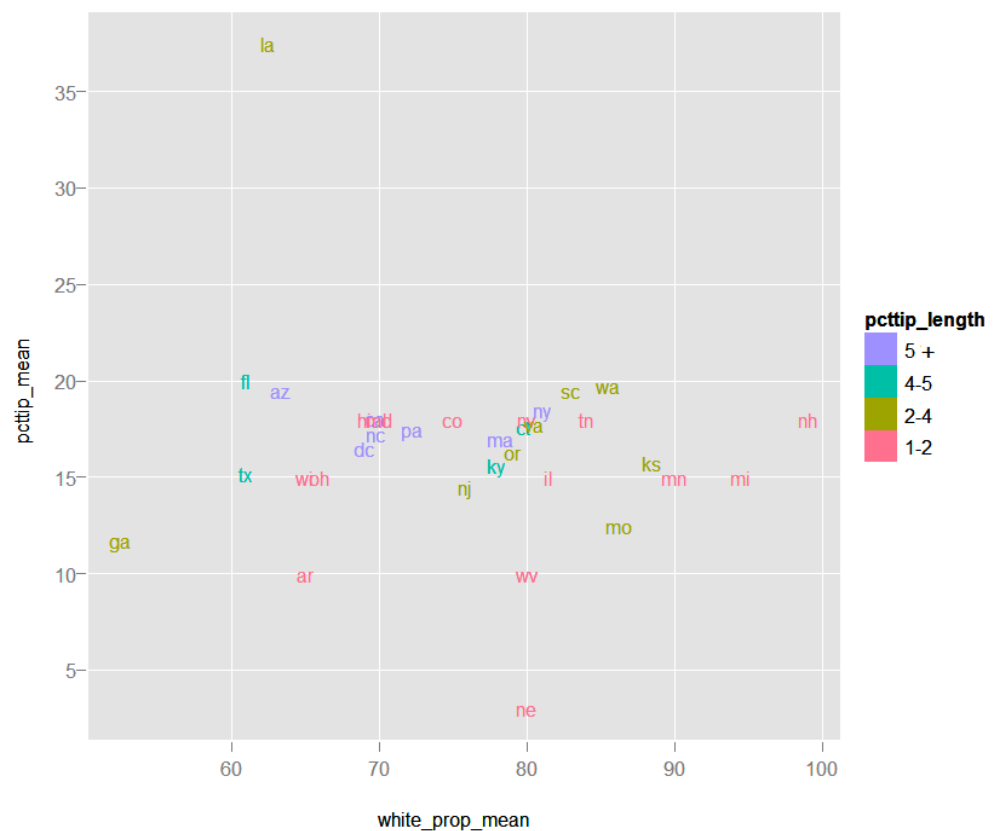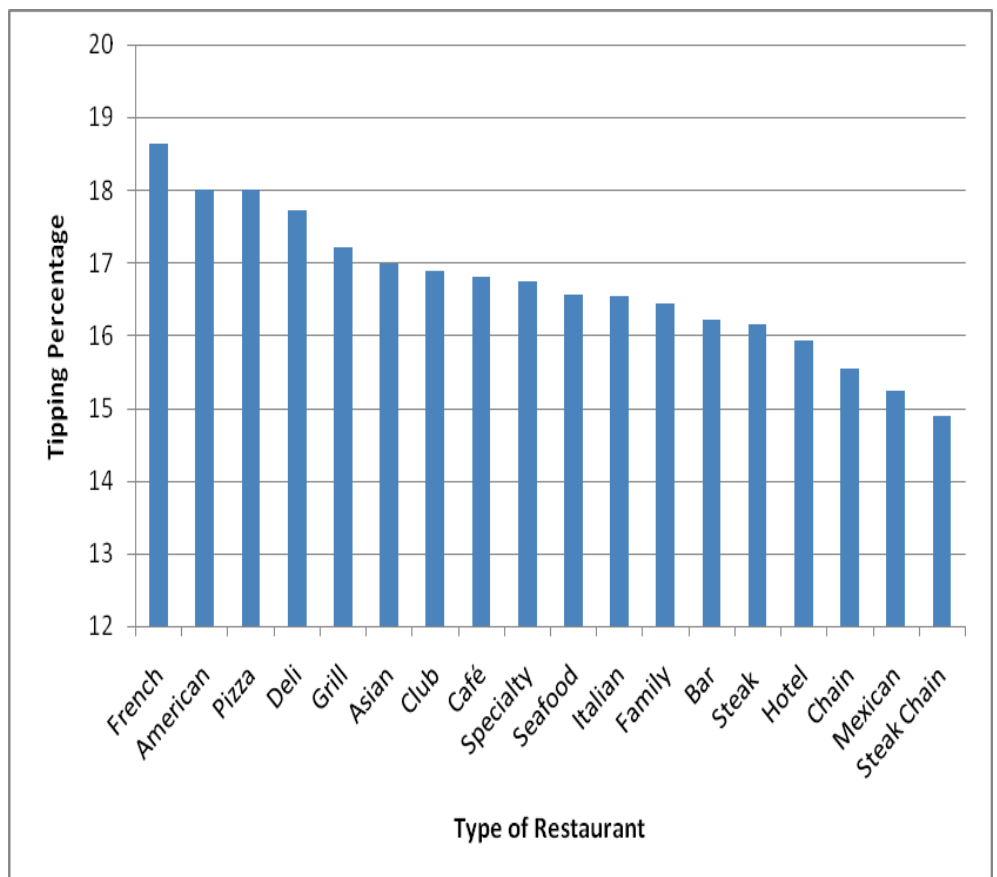
Restaurant Type

The type of restaurant in which servers work is also important in determing the expected potential gratuites one could receive. In this study, each restaurant was classified into 18 different types of restaurants. For example, Applebees Bar and Grill was classified as a restaurant chain whereas Outback Steakhouse was classified as a steak chain restaurant. Restaurant types whose servers receive the largest gratuity percent were French restaurnts followed by American and Pizza eating establishments. Restaurants receinving the lowest aveage server tipping percentages were Steak Chain and Mexican restaurants. The difference in tipping percentage between the lowest and highest type of restaurant equalled approximately 3.5%. This result could again be attribuatble to the low frequency of the observations in the French, American, Pizza, and Deli restaurant types. Those four restaurant types combined for a total of 57 observations as compared to 765 observations for the lowest four restaurant types of Hotel, Chain, Mexican, and SteakChain.

Table 1. Frequency and percent of type of restaurant reported in a web-based survey.

Figure 11. Average tipping percentage received by servers at a particular type of restaurant.

Type	N	%
American	8	0.37%
Asian	60	2.75%
Bar	74	3.39%
Café	126	5.78%
Chain	629	28.84%
Club	40	1.83%
Deli	7	0.32%
Family	38	1.74%
French	11	0.50%
Grill	100	4.59%
Hotel	50	2.29%
Italian	64	2.93%
Mexican	36	1.65%
Pizza	4	0.18%
Seafood	36	1.65%
Specialty	802	36.77%
Steak	46	2.11%
Steak Chain	50	2.29%
Grand Total	**2181**	**100.00%**

Interaction of Region of Restaurant Location and Age of Server

Based on the findings displayed in Figure 12, one could expect to have a higher tip based on their year's experience. However, caution must be expressed in this conclusion. Looking at the figure you notice that the percent tip is decreased from 45-50 years of experience. This could be due to the lack of data for this particular age group. The total observations for the last four age categories were 5, 1, 2, and 2 for the 35-40, 40-45, 45-50, and 50-55, respectively. Sufficient data in all age categories one could expect to see more of a positive linear trend in tipping percentage based on years of experience.

Figure 12. Average tipping percentage based on years experience.
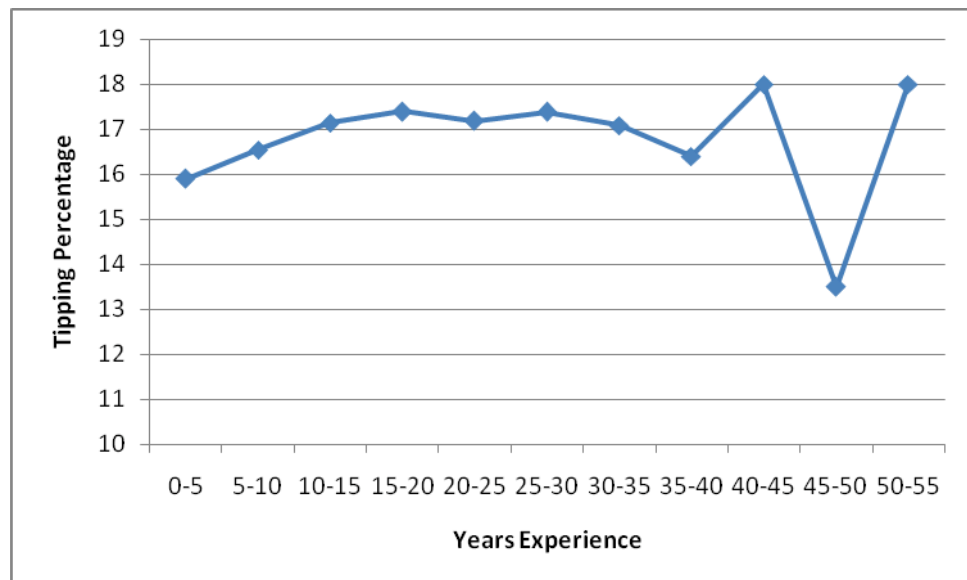
Further evaluation of server age reveals a significant difference in percent tip based on geographical region. Although certain regions have missing data for certain age groups, some conclusions can be drawn from the trends that are displayed in the Table 2. An ideal location for a server wanting to earn the highest gratuities would locate themselves in the Northeast region. However, looking at the interaction between age and region, the server would locate themselves in the South if they were between the age of 56-65. Contarily, the elderly servers could expect the highest percent tip in the Western region. Age is no doubt an important factor to consider when developing a model to predict the percent tip of a server, but more importantly, the interaction of region by age must not be overlooked.

Table 2. Tipping percentage by region and age group.

Interaction

Region	Age 16-25	26-35	36-45	46-55	56-65	66-75	Grand Total
Northeast	17.40	17.79	17.33	18.10	15.63		17.52
West	16.47	17.27	16.68	17.81	17.36	20.00	16.98
South	16.02	16.85	16.44	16.55	18.33	15.75	16.44
Pacific	15.33	17.50	18.00				16.36
Midwest	16.06	16.12	16.41	14.62	16.78		16.06
Grand Total	16.54	16.96	16.68	16.76	16.97	16.60	16.73

of Marital Status and Gender of Server

Another important factor to consider when looking at tipping percentage would be the marital status and sex of the server. Based on Figure 13, females that are married tend to receive the least tips when compared to single females. Additionally, males that are married receive a lower tipping percentage when compared to single men. Looking at the big picture, you notice that single men receive the highest tipping percentage followed by single men, single women, and them married women. Marital status and sex undoubtedly play a significant role in the amount of gratuity a server could expect during a particular eating experience. More data gathered on making the sex and marital of the server would increase the accuracy of this prediction. However, trends in the data can still be evaluated based on these assumptions.

Figure 13. Average Tipping Percentage by Gender and Marital status.
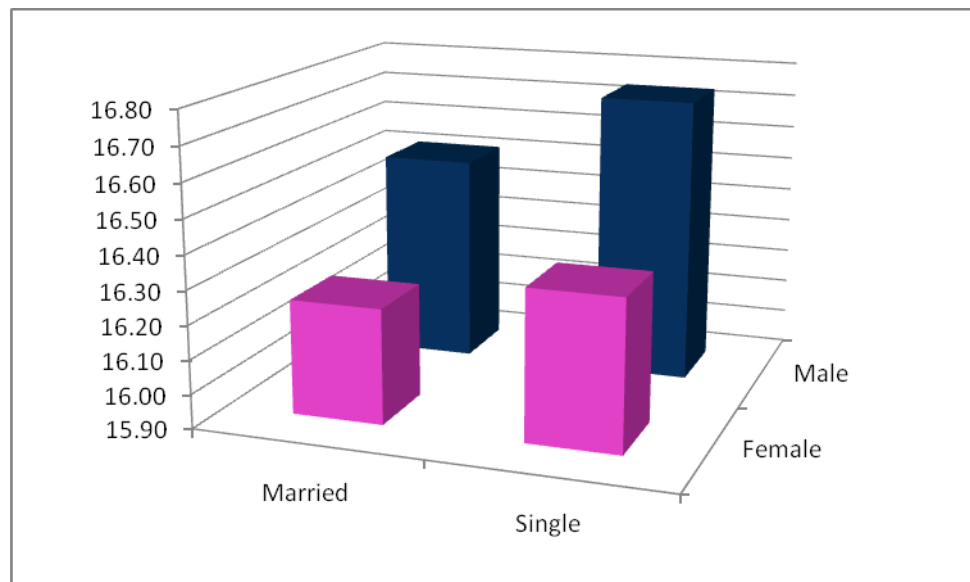
Conclusion

Many linear trends were found in this dataset for the variables of region, age, and proportion of race. The northeast region averaged the highest percent gratuity per meal whereas servers in the Midwest region averaged the lowest percent gratuities. An evaluation of the variances of each region and combining that information into the regional differences yielded the conclusion that the Northeast region not only had the highest server tips but also the lowest variance in tips. Therefore, if a server is beginning his/her career in the food service industry and desires to make the largest percent gratuity per meal, they should relocate to the Northeast region. The region with the highest variance was the South, possibly due to cultural and racial differences within the region, however the South ranked third of five regions in tipping percentage.

Proportion of race served in restaurants contained both positive (White) and negative (Black, Asian, and Hispanic) trends in gratuity percentage. Within each region, the South contained the largest diversity of race of all regions. The Midwest and Northeast regions contained similar proportions of race, yet all regions were primarily of the white race (approximately 70-80%). Therefore, since the white population has a positive linear trend in gratuity percentage as the proportion of whites increase, the regions of Northeast and Midwest should have the predicted

highest percent tips. Other factors must be contributable to tipping percentage since the Midwest region received the lowest percent tips.

Type of restaurant also played an important role when determining potential earnings in the food service industry. French, American, and Pizza restaurants yielded the highest tipping percentage whereas servers at Steak Chain, Mexican, and Chain restaurants received the lowest average tipping percentage. This was attributable to the low frequency of observations in the restaurant types with the highest earnings. In order to accurately determine the true mean and distribution of earnings, the dataset needs more total unambiguous observations to identify these parameters.

Marital status, gender, and years of experience are parameters that must be measured to accurately predict tipping percentage. The highest tipping percentage could be expected for single men followed by married men, single women, and married women. Although one would assume single women to receive the highest tips, analysis revealed that this was not the case and that the interaction of marital status and gender was non-significant ($P > 0.05$). Years of experience showed a positive linear trend from 0 to 30 year of experience. Caution must be taken when predicting tipping percentage beyond 30 years of experience in this dataset due to the large variation and low frequency of observations. However, one would expect a positive linear trend in tipping percentage by years of experience to continue if sufficient data were present in this study.

In order to accurately predict tipping percentage, a complex model of gender, marital status, region, years of experience, age, and type of restaurant must be utilized. Extending the data collection period on the web-based survey would have drastically increased the resolution at which one could examine relationships between these variables.

Appendix A.

Differences of Least Squares Means

Effect	Region	Region	Estimate	Standard Error	DF	t Value	Pr > |t|
Region	Midwest	Northeast	-1.4677	0.2221	2078	-6.61	<.0001
Region	Midwest	Pacific	-0.3029	1.1056	2078	-0.27	0.7841
Region	Midwest	South	-0.3851	0.2161	2078	-1.78	0.0748
Region	Midwest	Territory	-1.9393	2.1006	2078	-0.92	0.3560
Region	Midwest	West	-0.9335	0.2490	2078	-3.75	0.0002
Region	Northeast	Pacific	1.1648	1.1044	2078	1.05	0.2917
Region	Northeast	South	1.0826	0.2100	2078	5.16	<.0001
Region	Northeast	Territory	-0.4716	2.1000	2078	-0.22	0.8223
Region	Northeast	West	0.5342	0.2438	2078	2.19	0.0286
Region	Pacific	South	-0.08217	1.1032	2078	-0.07	0.9406
Region	Pacific	Territory	-1.6364	2.3629	2078	-0.69	0.4887
Region	Pacific	West	-0.6306	1.1102	2078	-0.57	0.5701
Region	South	Territory	-1.5542	2.0994	2078	-0.74	0.4592
Region	South	West	-0.5484	0.2383	2078	-2.30	0.0215
Region	Territory	West	1.0058	2.1031	2078	0.48	0.6325

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Region	5	1955	9.42	<.0001
Sex	1	1955	4.63	0.0315
Married	1	1955	6.24	0.0125
Sex*Married	1	1955	0.00	0.9642
Age_Group	6	1955	1.42	0.2014

Solution for Fixed Effects

Effect	Region	Age_Group	Sex	Married	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept					17.5984	1.1327	1956	15.54	<.0001
Region	Midwest				-0.7951	0.2531	1956	-3.14	0.0017
Region	Northeast				0.6593	0.2491	1956	2.65	0.0082
Region	Pacific				-0.6794	1.0866	1956	-0.63	0.5319
Region	South				-0.3963	0.2416	1956	-1.64	0.1011
Region	Territory				0.6796	2.0570	1956	0.33	0.7411
Region	West				0	.	.	.	.
Sex			0		0.4415	0.1747	1956	2.53	0.0116
Sex			1		0	.	.	.	.
Married				0	0.5564	0.2024	1956	2.75	0.0060
Married				1	0	.	.	.	.
Age_Group		.			-0.3349	1.8294	1956	-0.18	0.8548
Age_Group		16-19			-1.6925	1.2158	1956	-1.39	0.1640
Age_Group		20-29			-1.3332	1.1316	1956	-1.18	0.2389
Age_Group		30-39			-1.0265	1.1350	1956	-0.90	0.3659
Age_Group		40-49			-0.6643	1.1509	1956	-0.58	0.5638
Age_Group		50-59			-1.1165	1.2230	1956	-0.91	0.3614
Age_Group		60-69			0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Region	5	1956	9.43	<.0001

Sex	1	1956	6.39	0.0116
Married	1	1956	7.56	0.0060
Age_Group	6	1956	1.44	0.1968