Introduction

Tipping at restaurants where there is table service is expected in the United States. Many waiters and waitresses rely on these tips because restaurants are not required to pay them the regular minimum wage. The US federal minimum wage for a tipped server is only $2.13 an hour. So, servers are expected to earn most of their wages from tips. The normal tip that customers leave is 15% of the total price of the meal.

We are interested in what factors contribute to the size of the tip that servers receive. Are tips that a server receives influenced by the server's behavior, the time of day that they work, or the kinds of customers that they serve? Is tipping different in different parts of the country? We will use data collected by Dr. William Michael Lynn, a professor of consumer behavior and marketing at the Cornell University School of Hotel Administration. This data comes from an online survey that was filled out by servers from across the country. Some of the variables include the restaurant name and location that the server worked in, the size of the tips they received, their number of years of experience, how often they engage in certain behaviors, and their genetic qualities such as gender, age, and hair color. We will use this data to answer the above questions about tipping.

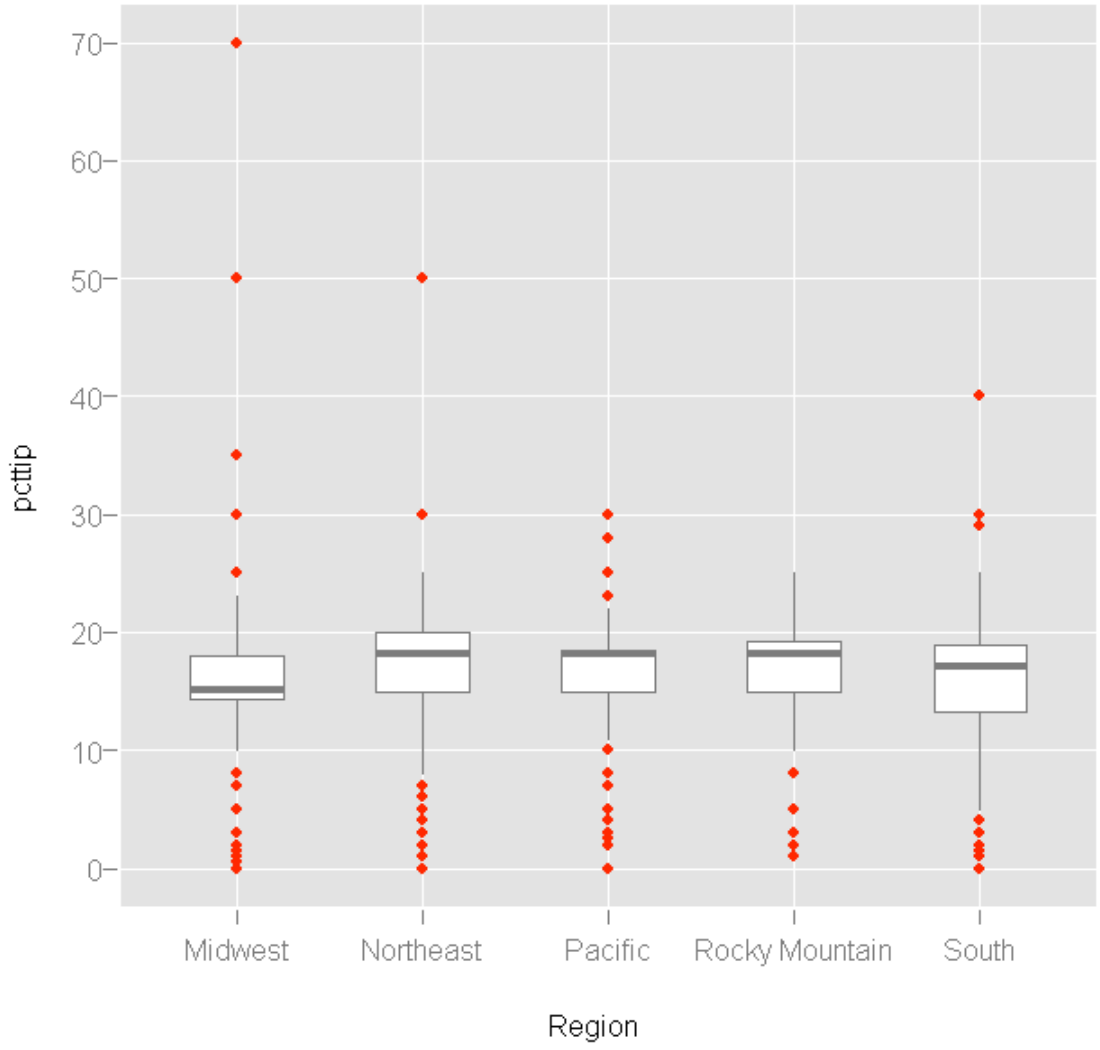
In working with the tips data set, we decided to narrow our focus to the United States only. We segmented the United States into 5 regions: the Northeast, South, Midwest, Rocky Mountain, and Pacific regions. Any observations that was not from the United States were eliminated. Then, because we wanted to focus on average tip percentages, we eliminated all the observations that did not have a value for this variable. There were a few locations such as fast food restaurants that listed a 0% tip. Since these locations do not have wait staff, they were eliminated as to not confuse the restaurant type in our analysis. Also, we eliminated those that had average tip percentages greater than 70% because those values did not seem plausible. Finally, we changed numbered values to their corresponding meanings as to ensure proper categorical usage as well as to simplify our analysis and scatter-plot creation techniques for the variables for group size, time of day, and behaviors.

Percent Tip Per Region

First of all, we wanted to explore if the average percentage of the total bill that the server receives as a tip is different for different regions of the country. We thought that the coastal regions (Northeast and Pacific) would have higher average tips than non-coastal regions. This is because most of the densely populated areas are along the coasts and prices are often higher in these large cities (like New York City or Los Angeles).

Plot 1

Average Percentage Tips by Region



According to Plot 1, the highest average tip percentages are generated in the Northeast, Pacific, and Rocky Mountain regions. The lowest average tip percentage is in the Midwest. One thing to note, however, is that the spread is larger for the Midwest and Northeast regions than it is for the others. This affects the means for those regions, and,

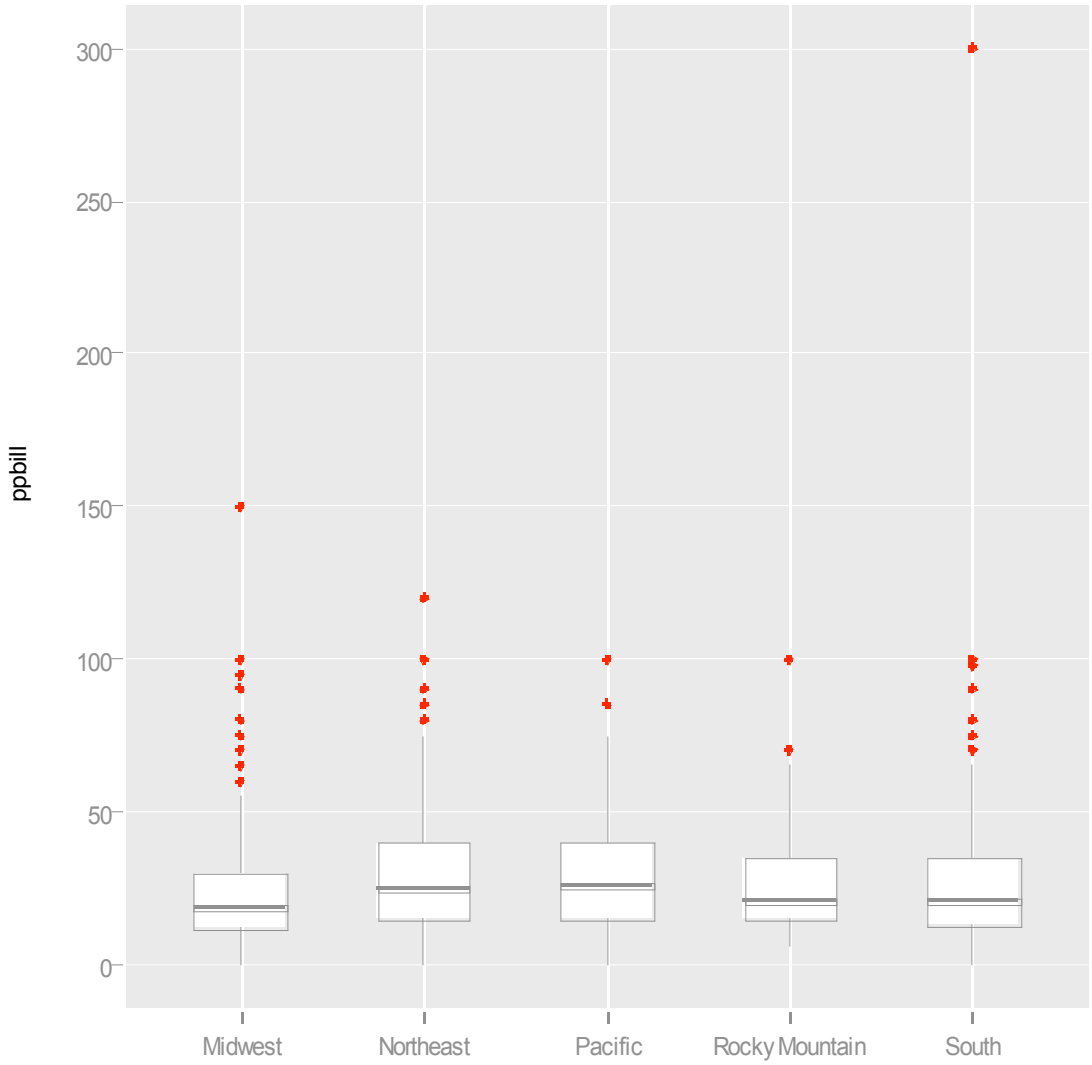
therefore, our inferences. Otherwise, our assumption is correct, the Northeast and Pacific regions, along with the Rocky Mountain region, have higher average tip percentages than the other regions.

Price Per Bill Per Region

In relation to our previous plot, we wanted to see if there was a relationship between the size of the average tip percentages and the price per bill. We thought that tips would be higher for larger bills.

Plot 2

Price per Bill by Region



From Plot 2, it is more evident that there is a relationship between price per bill and the average tip (from Plot 1). The means for price per bill follow a very similar pattern to those for average tip percentages. The coastal regions (the Northeast and the Pacific)

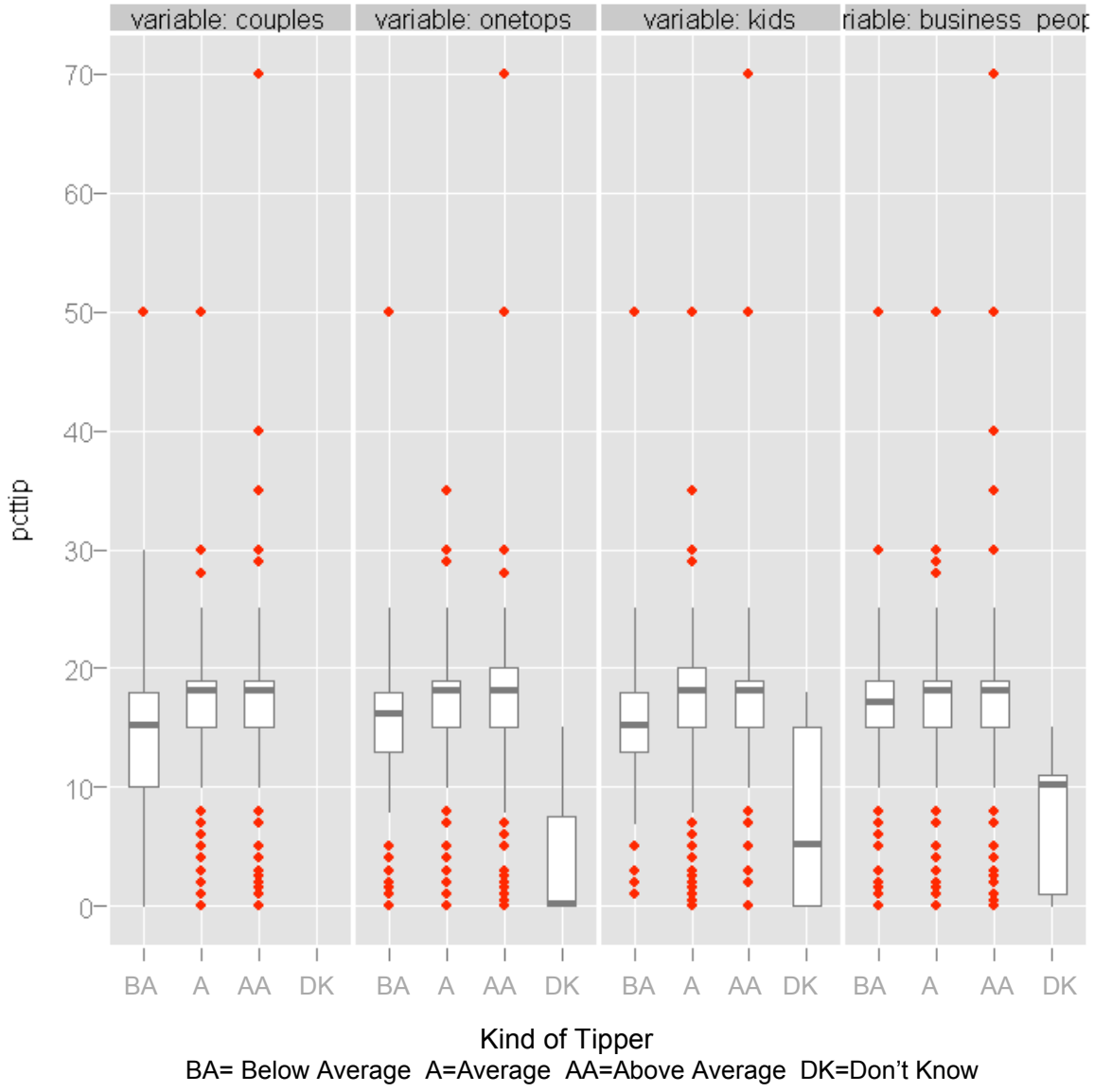
have the highest prices per bill while the Midwest has the lowest. So, our assumption is correct, those who pay more also leave a larger tip.

Average Tips by Size of Group

Next, we want to see if the average tips that a server receives is affected by the size of the group that they serve. We have data for four types of groups: couples, persons eating alone, families with kids, and business people. We thought that couples and business people would most likely tip more than the others. This is because when one is at a restaurant with co-workers or significant others, one feels more obligated to leave a proper tip.

Plot 3

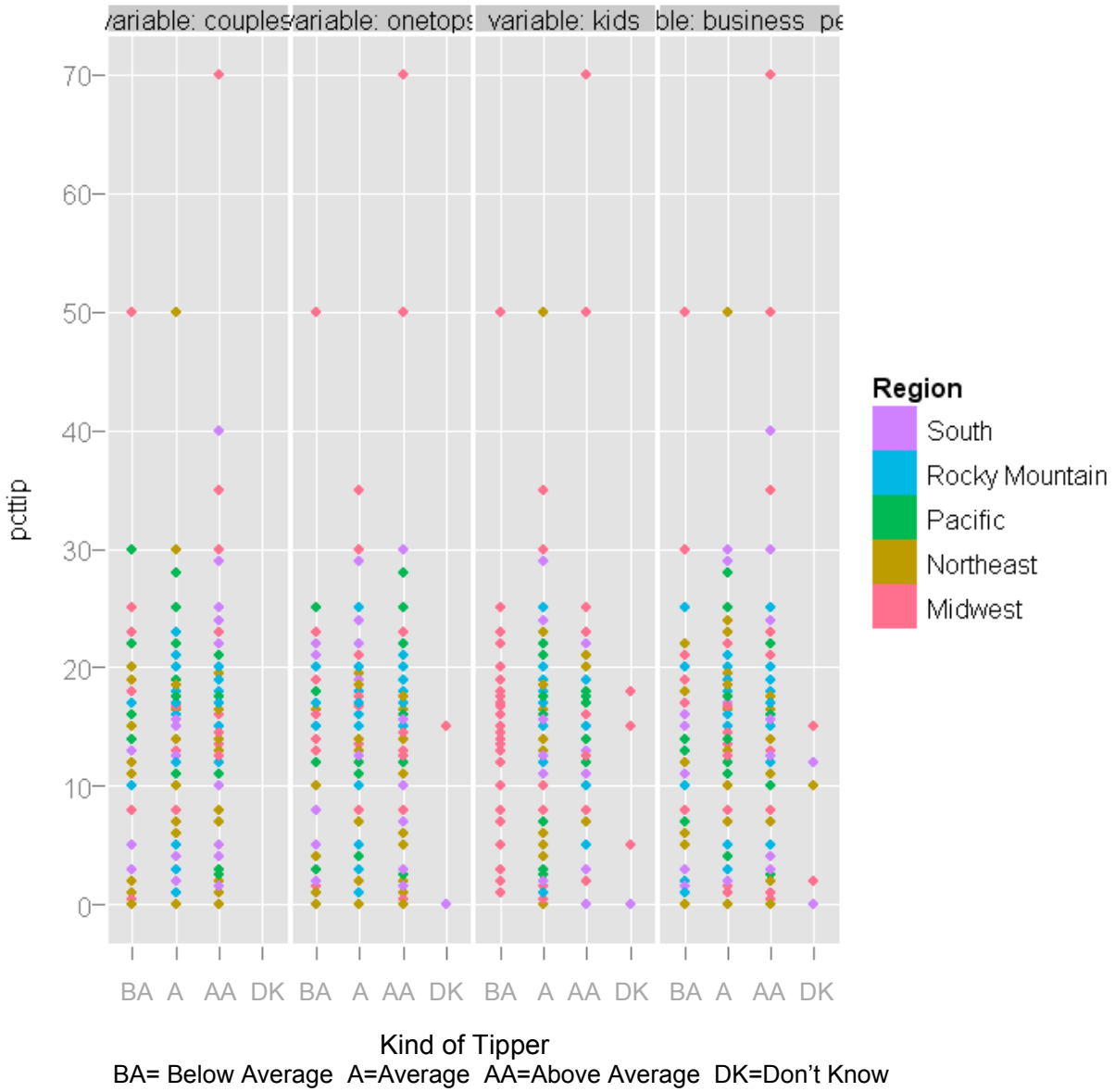
Average Tip Percentages by Group Size



From Plot 3, we discovered that there is no big difference in the average size of the tip given for the average and the above average tippers in each group. So, it looks like most customers, regardless of who they eat with, tip about the same. On the other hand, for those who are rated as tipping below average, business people give higher average tips than other groups of people. This may be because servers expect higher tips from business people, so below average for them is higher than below average for other types of groups.

Plot 4

Average Tips by Group Size—colored by Region



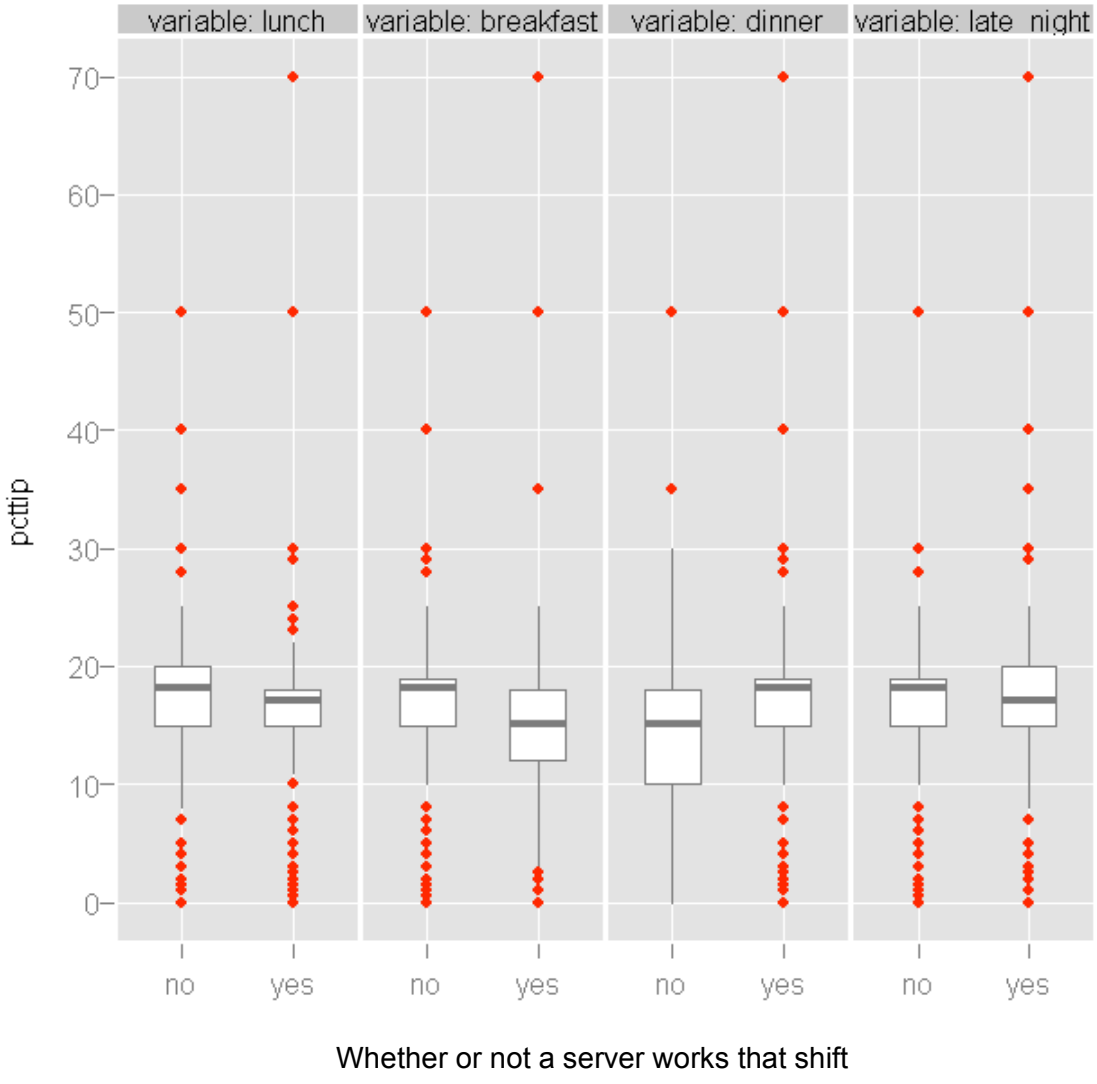
We also wanted to see if there was any noticeable differences between regions of the United States in the kind of tippers that different groups are. The only noticeable thing we found was that all of the families with kids that tip below average are from the Midwest, which can be seen in Plot 4. Other than that, there are no noticeable differences between the regions.

Average Tips by Shift Times

We decided to explore how the average tips that a server receives are affected by the shift(s) that a server works. We predicted that servers who work the dinner shift earn larger tips than those who do not.

Plot 5

Average Tips by Shift Times



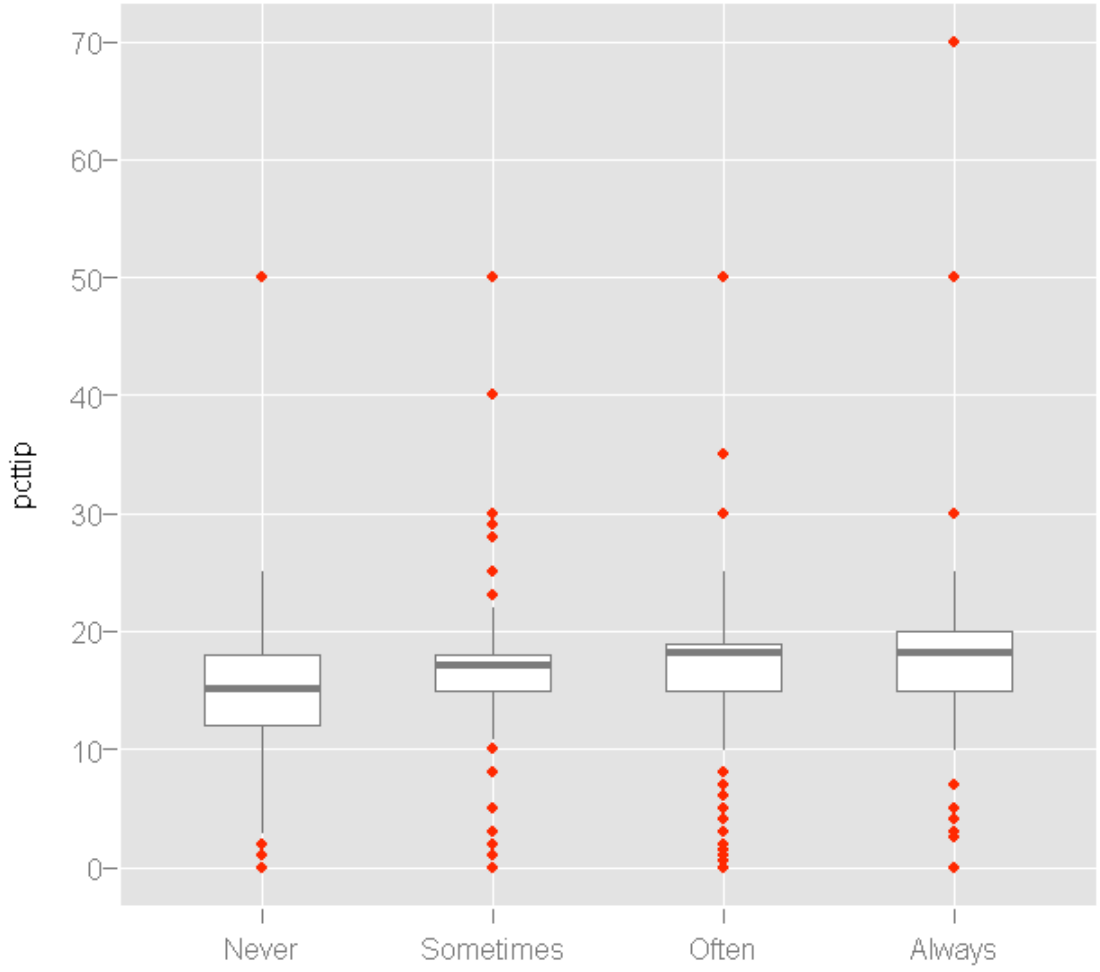
Plot 5 shows that our prediction is correct. We found that servers who do not work during the dinner shift make, on average, less in tips than those who do. One problem that should be noted with this plot is that, while many of these servers worked multiple shifts, we only have the average tip they made over all their shifts. So, our findings could be inaccurate. We also looked at this plot by region, but, because we did not find any differences between the regions, we discarded this plot.

Average Tips by Server Behavior

We also looked into how servers' behaviors affect the size of tips that they receive. After looking at plots of all the behaviors that we had data for, we picked the two most influential behaviors. One was how often a server complements their customers' food choice. Of course, we assume that the more servers complement the customer, the more the customer will be willing to tip the server. As you can see in Plot 6, this is correct.

Plot 6

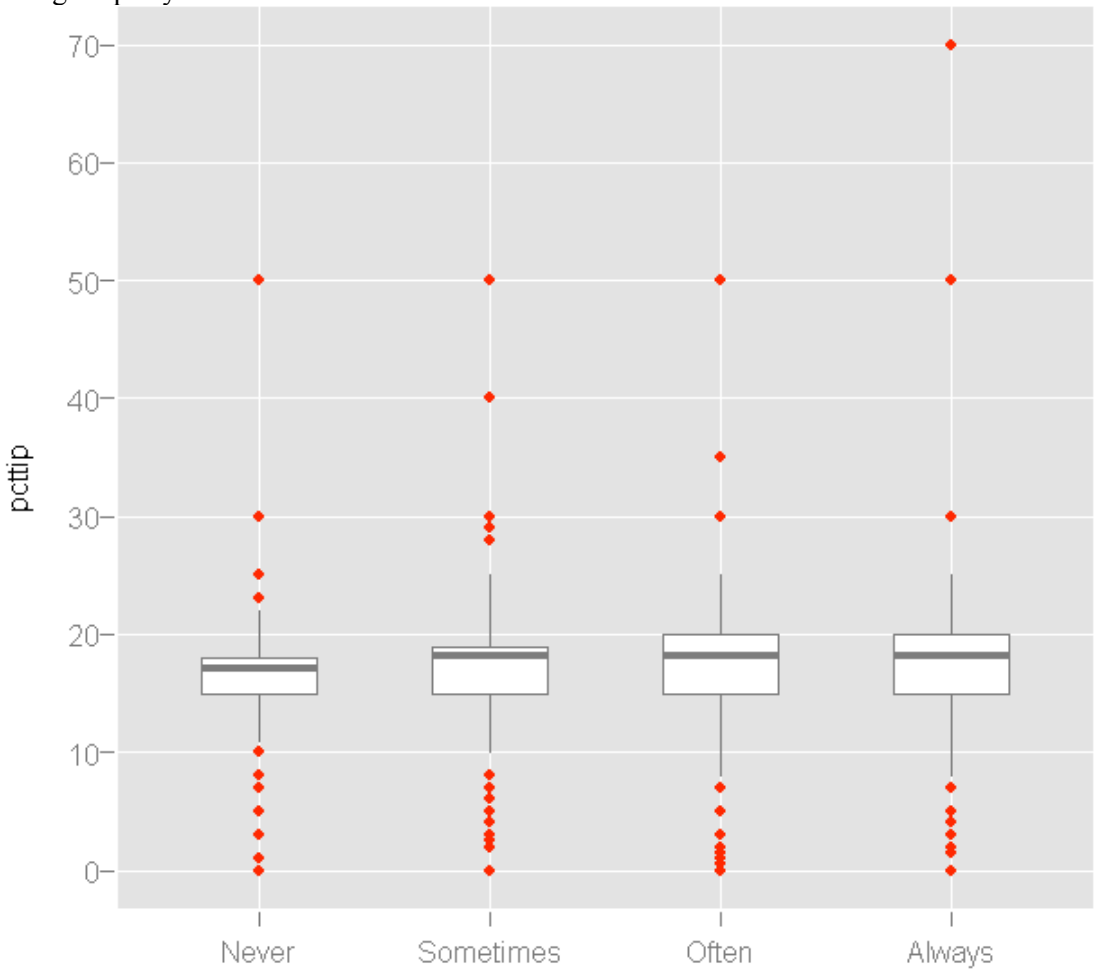
Average Tips by Complement



We also looked at how using the customers' name when speaking to them affects the average size of the tip that they leave. Again, we think that the more servers use their customers' names, the more tips they'll receive. As you can see in Plot 7, this is again correct. In the case of these two behaviors, the more positive behavior the servers directed toward the customers, the more likely they will receive a higher tip.

Plot 7

Average Tips by Customer Name



Again, we looked at plots 6 and 7 by region, but didn't find any patterns or differences between regions of the United States.

Conclusion

From our analysis we can conclude if one wants to be a successful server, she or he can take various steps to better the odds of receiving high-percent tips. First, the highest tips

can be made either in the Northeast or Pacific regions, so working at a restaurant there is best for large tips. In addition, we have discovered that, in general, working in restaurants that do not serve families with kids in the Midwest will yield a higher tip overall. We also discovered that working a dinner shift significantly increases the size of the tips that can be gathered. While it is difficult to determine that the dinner tips are the highest, it is clear to us that not working that shift has a negative impact on amount of tips that they receive. Finally, complementing customers and calling them by name increases sizes of tips. In fact the more customers that a waiter or waitress complements, the higher a tip he can expect. Always complementing and always using a customer's first name works best. Creating a warm and friendly atmosphere makes the customer happy, which encourages larger tips.

Appendix
Reproducibility

In order for the reader to be able to reproduce our findings, we will describe what we did with the data to produce our findings. Before using R, we made several changes to the original Excel data file. One of these changes was changing all the values for the state variable into their two letter postal codes. We eliminated all the observations that weren't from the United States. Once we had the state variable fixed, we created the Region variable by dividing the United States into 5 regions: the Northeast, South, Midwest, Rocky Mountain, and Pacific regions. Also, for the categorical variables for group size, time of day, and server behavior, we changed the numerical values provided into their correct meanings. For example, for the shift time variables, a 0 meant the server did not work that shift and a 1 meant that he or she did. To make the plots easier to read, we changed the 0 and 1 to "no" and "yes." Then, we used R to modify the pcttip variable and create our plots, using our modified Excel data file.

Code:

```
#Load ggplot, which enables us to plot the data
library(ggplot)
#Read in the data file
ss<-read.csv(file.choose())
#Turn all the values for average tips from proportions into percentages
props <- !is.na(ss$pcttip) & (ss$pcttip < 1)
ss$pcttip[props] <- ss$pcttip[props] * 100
#Eliminate all values for average tip percentages greater than 70%
table(round_any(ss$pcttip, 1), exclude=NULL)
ss <- ss[!(!is.na(ss$pcttip) &  ss$pcttip > 70), ]
#The following 6 lines of codes put the levels of each factor in the correct order (otherwise, R
#would put them in alphabetical order)
ss$couples<-factor(ss$couples, levels=c("Below Average", "Average", "Above Average", "Don't
        Know"))
ss$onetops<-factor(ss$onetops, levels=c("Below Average", "Average", "Above Average", "Don't
        Know"))
ss$kids<-factor(ss$kids, levels=c("Below Average", "Average", "Above Average", "Don't
        Know"))
ss$business_people<-factor(ss$business_people, levels=c("Below Average", "Average", "Above
        Average", "Don't Know"))
ss$customer_name<-factor(ss$customer_name, levels=c("Never", "Sometimes", "Often",
        "Always"))
ss$complement<-factor(ss$complement, levels=c("Never", "Sometimes", "Often", "Always"))
#In order to plot more than one variable in a plot, we will use the next three lines of code to melt
#the data.
ss$id <- 1:nrow(ss)
```

```
#The following combines the shift time variables together.
ssm <- melt(ss, id=c("id","pcttip", "Region"), m=c("lunch","breakfast", "dinner", "late_night"))
#The following combines the group size variables together.
ssg<-melt(ss, id=c("id", "pcttip", "Region"), m=c("couples", "onetops", "kids",
          "business_people"))
#The next seven lines of code create each of the plots seen in the report:
#Plot 1: Average tips by Region
qplot(Region,pcttip,data=ss, type="boxplot")
#Plot 2: Price Per Bill by Region
qplot(Region,ppbill,data=ss, type="boxplot")
#Plot 3: Average tips by Group Size
qplot( value, pcttip,data=ssg, facet = . ~ variable, type="boxplot")
#Plot 4: Average tips by Group Size, colored by Region
qplot( value, pcttip,data=ssg, facet = . ~ variable, colour=Region)
#Plot 5: Averages tips by Shift Times
qplot( value, pcttip,data=ssm, facet = . ~ variable, type=”boxplot”)
#Plot 6: Average tips by Complement
qplot(complement, pcttip, data=ss, type="boxplot")
#Plot 7: Average tips by Customer Name
qplot(customer_name, pcttip, data=ss, type="boxplot")
```