Stat 480

10 April 2007

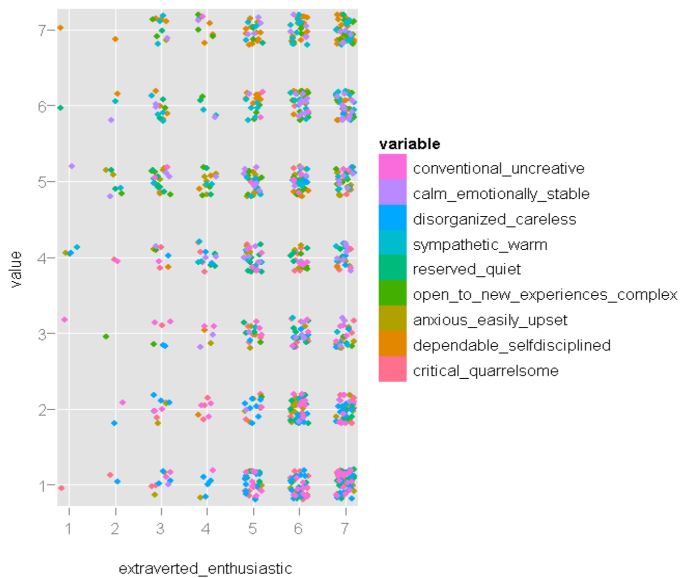
Server Survey Tip Analysis

Introduction

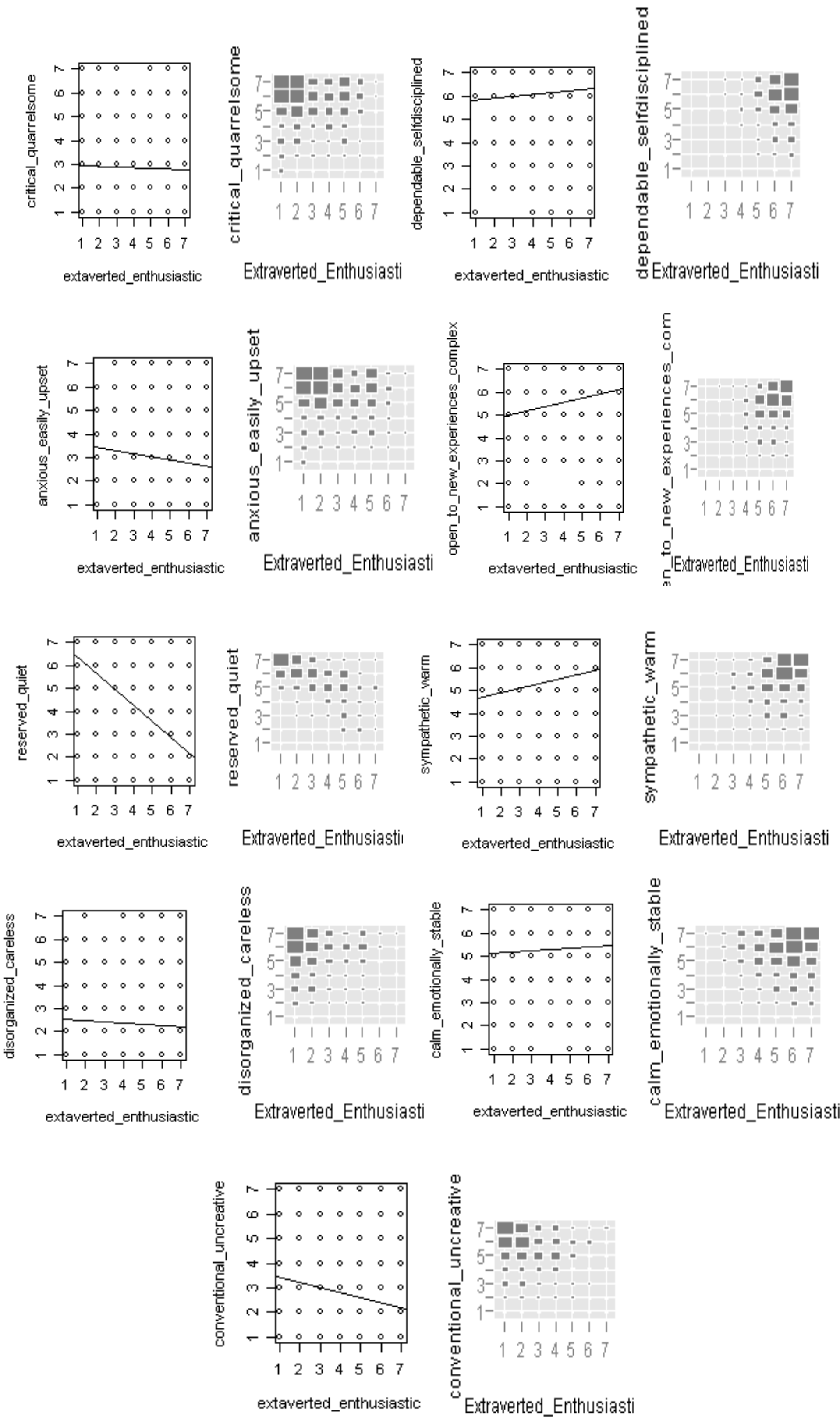
The United States Department of Agriculture maintains that on average, Americans spend just over 1.5 hours eating in restaurants each week. Multiplied by an estimate of about 301,575,683 citizens currently living in the continental U.S. that makes about 483 million hours Americans spend in restaurants as a whole. With this said, that means that there are 483 million hours possible for servers that work in those restaurants to accumulate tips as part of their weekly wages. Despite this abundance of individuals willing to tip for services that servers provide, the amount that each server will take home at weeks end varies greatly. The reasons for this variability is something that we are going to attempt to identify with this tip analysis. Given a dataset compiled by Prof William Michael Lynn of Cornell University, we will attempt to weed through a list of 77 variables in order to isolate as well as quantify the effect that a number of factors have on the success of a server as well as pinpoint some of the things that servers can do (or shouldn't do) in order to maximize the size of their pay checks. The origin of the data is very broad in that restaurants from all over the world have contributed to this database. Servers from a large bank of restaurants, identified through i.p. addresses, logged on to a website where they would not only surrender information about their daily activities but also their emotional condition and the type and numbers of people that visited their establishment. Using this data, the responses of the servers were classified and coded in a manner easy to import to a spreadsheet. If any observations had missing values, it seemed most convenient to exclude this observation from the dataset we used for analysis. From here we set goals which dealt with examining how a server's effort and attitude influences their tip size, as well as how age and demographics play their part. Though it would be nice to have complete control over the tip one receives, there are also a few factors that an individual cannot control which positively and negatively affect their daily revenue. By grouping these variables by nature, we are going to attempt to classify the servers "fixed" situation and then determine where variability lies that can be manipulated by the individual server to provide a formula for a successful food service employee.

Attitude

Attitude is a good measure of how successful a business and its employees are. As stated in a Charleston Regional Business Journal, "your attitude toward business directly affects what you achieve." In this case, the business is a restaurant and the achievement is the tip that a server receives as compensation for the services they provide. (http://www.charlestonbusiness.com/pub/12_25/news/8177-1.html) Furthermore, a server's attitude is something that can be changed in an effort to secure a higher value of financial return. Given the server survey dataset, complete with 77 different variables, it seemed that only a hand full of them actually related to the emotional aspect of the restaurant server. These variables dealt with the creativity, disposition, organization, and likelihood to engage in conflict. The scale that each of these qualities were rated on was 1 to 7 where 1 was "agree least" to 7 which was "strongly agree." As seen in the following graphic, it seems that in relation to a positive quality like having enthusiasm and an extroverted social position, those that ranked themselves high in terms of one positive quality also ranked themselves high in the other positive qualities measured. On the other hand, those that felt that they ranked low in relation to a positive quality, ranked themselves high in terms of a negative quality. Positive qualities are classified as those traits that would be favorable and most inviting whereas negative qualities are those that portray aggression or a withdrawn mind-set.



After observing the fact that one's self classification is indicative of how individuals rate themselves on similar and contrasting qualities, it was found that we could summarize their attitudes into variables called "posprop" and "negprop" (posprop = positive properties; negprop = negative properties). Which variables should be grouped was somewhat trivial until we used linear regression to observe the trends of the rankings. If a best fit line exhibits a positive correlation, that tells us that as one variable increases, so does the other thus indicating that the two considered variables could be grouped. If the correlation of the two variables is negative however, that means that as one variable's value increases, the other considered variable's value decreases suggesting that these variables be classified into opposing groups. Shown below are scatter-plots of data where each variable is compared to a positive quality. Observing the correlation allows us to be able to classify the variable in question as either positive or negative. To expand on how the data is distributed, a fluctuation table of the data is placed to the right of the fitted scatter-plot. Including these fluctuation tables better helps explain why the best fit lines follow the trends that they attempt to estimate.

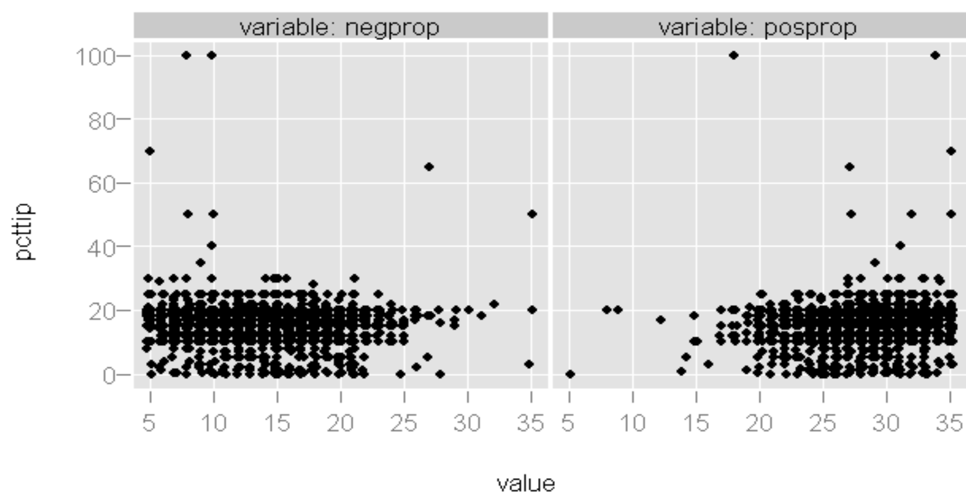


Now that we have determined how the variables relate to one another, the following hierarchy of variables was generated:

Positive Properties(Posprop) : 1.)dependable/selfdisciplined
2.)open to new things/complex
3.)sympathetic/warm
4.)calm/emotionally stable
5.)extroverted/enthusiastic.

Negative Properties(Negprop) : 1.)critical/quarrelsome
2.)anxious/easily upset
3.)reserved/quiet
4.)disorganized/careless
5.)conventional/uncreative

By summing across these groups, we were able to solve for which of these two groups increased the frequency of getting a larger tip given the bill amount. Seen below, as the strength of negative properties increases, the percentage of the bill that is the tip takes smaller and smaller values. To the contrary, as the intensity of positive properties increases, so does the frequency of ever increasing tip percentages.



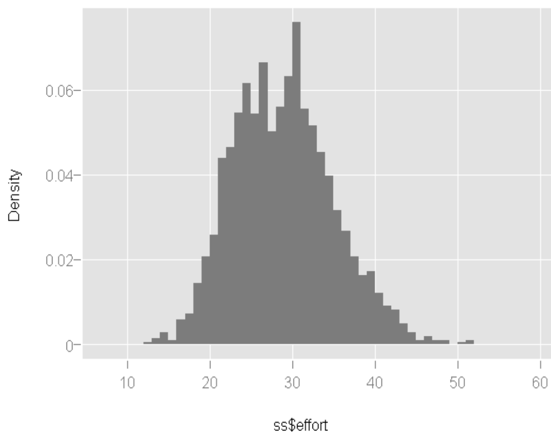
To generalize these results, having a strong positive attitude makes a server much more successful at securing a larger tip. This is further supported in Dr. Michael Lynn's guide titled "Mega Tips" which suggests that something as simple as smiling to display a positive attitude could mean a percentage increase in tip of up to 140%. (http://people.cornell.edu/pages/wml3/pdf/megatips.pdf) Don't be narrow-minded though, smiling is just one of the ways that one can portray a positive attitude. A server's success could be just as great if not better by simply exhibiting more of an up-beat attitude.

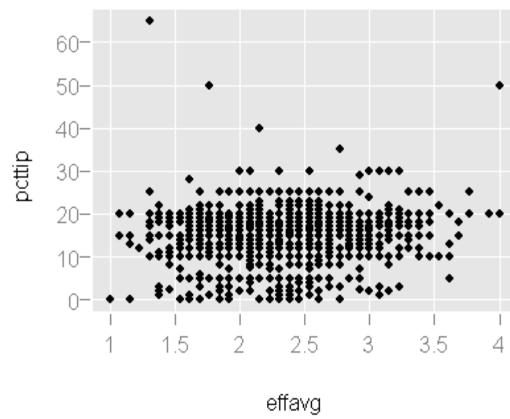
Effort

This survey information contains a group of variables, which effectively measure the total effort a waiter claims to put into serving their customers. Specifically, these are how often the waiter wore flair, introduced themselves by name, tried suggestive selling, squatted next to the table, touched customers, told jokes, repeated customers' orders back to them, called customers by name, drew pictures on checks, gave big smiles, wrote thanks on checks, forecasted good weather and complemented customers on their food choice.

These variables are all standardized on a scale from 1 to 4. Responses of 1 means they never use that method and a response of 4 means they always use that method. Since these variables are standardized on the same scale, we can summarize them all into a single variable "effort", which is a sum of each of these 13 variables, and "effavg", which is effort divided by 13.

We can now attempt to predict tip size with this effort variable, to see if greater effort results in larger tips. It is reasonable to assume there will be a point at which a customer may become annoyed with too much effort and the tip rate will either plateau or decrease.





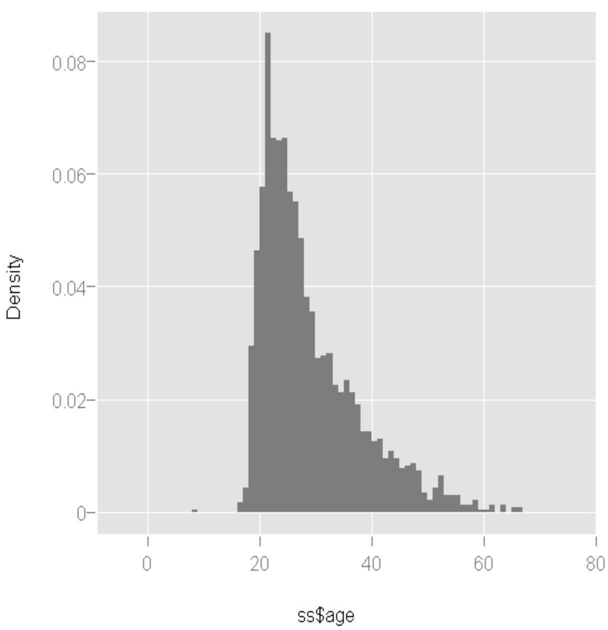
We can see that the distribution of effort is approximately normal. This is not too surprising, as it shows most people put forth an effort somewhere in the middle with few putting in "full" and few putting in no effort. It also appears that no relationship exists between the amount of effort a waiter puts forth to the percent tip they receive. We see that for most of the sample, most tipping rates are between 10 and 20% of the bill. This is consistent with data available from the bureau of labor statistics, which states "tips usually average between 10 and 20 percent of guests' checks" (http://www.bls.gov/oco/ocos162.htm). In support of the visually apparent lack of relationship, a linear model of fit illustrates the same conclusion. The R-square of the model is very small at less than 0.01, and the model fails to be significant on even a 0.05 level. Not only can we not predict pcttip with effort, but also the scatter plot suggests no other model would do better except a uniform distribution. This suggests the hypothesis that there would be a point of amount of effort to strive for is incorrect. It is worth noting, however, that for average effort levels above 3.5, there is only one point, which falls below a 10% tip rate. Though the lack of information that far into the end of the distribution is cause for concern, this potentially suggests that putting forth more effort will at least raise the minimum possible tip.

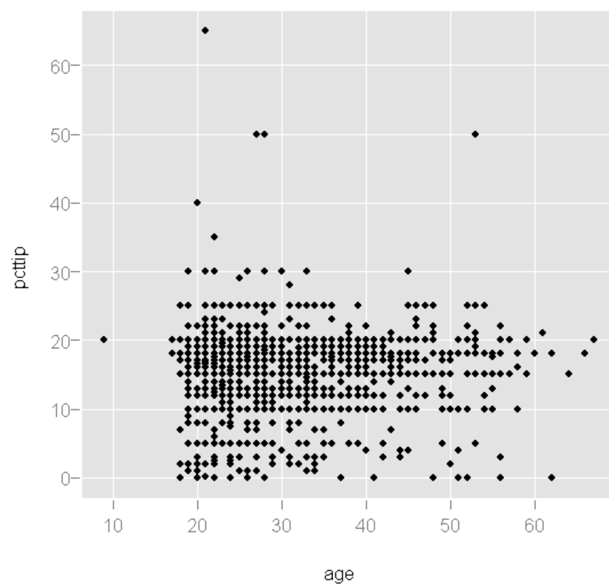
It is also interesting to consider which actions to take when considering effort as a whole – some actions may result in lower tips if done in excess. The functions available in the appendix automate comparing each action individually to the percentage tip in order to determine if they have a positive or negative effect, and if they significantly predict percent tip on their own. The actions of wearing flair, repeating an order, drawing on the check, and writing thanks on the check all decrease the percent tip as they're done more –

and therefore should be done less. The actions of touching a customer, telling jokes, and complementing do not significantly predict percent tips on a 0.01 level on their own and likely do not add significantly to a model with all of them.

Age

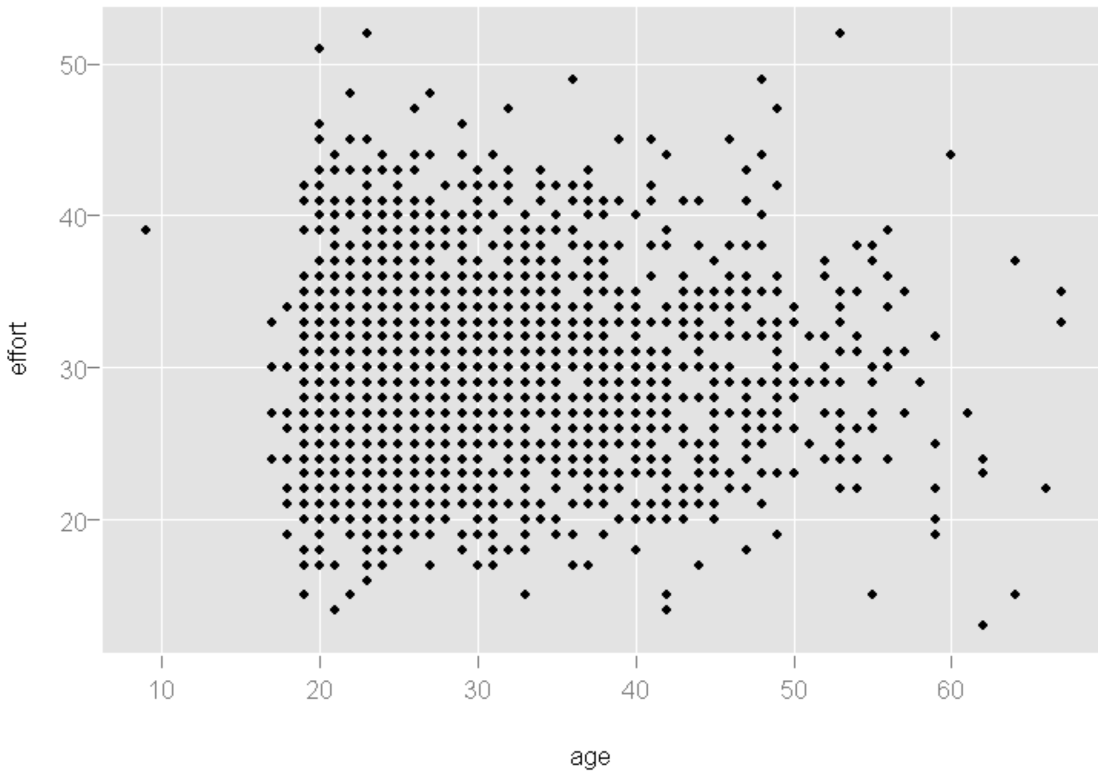
Let's now consider what relationship exists between a server's age and how well they are tipped. This is an attribute a server obviously cannot specifically attempt to control in an effort to improve their tipping rate. However, older servers likely are more knowledgeable on what effort to put forth, from their larger experience base of social situations. To analyze this relationship, it is convenient to clean the server's birth year variable and convert it directly into age. The method of standardizing these responses and removing ambiguous responses is available in the appendix.





We can see that the distribution of age is not normal; it is skewed to the right. This distribution is intuitive as it shows most waiters in this sample are aged in their late teens to early twenties. This quick preliminary visualization of how age predicts the percent tip received seems to suggest little noticeable relationship. Once again, most of the tipping rates fall heavily between the 10 to 20% range.

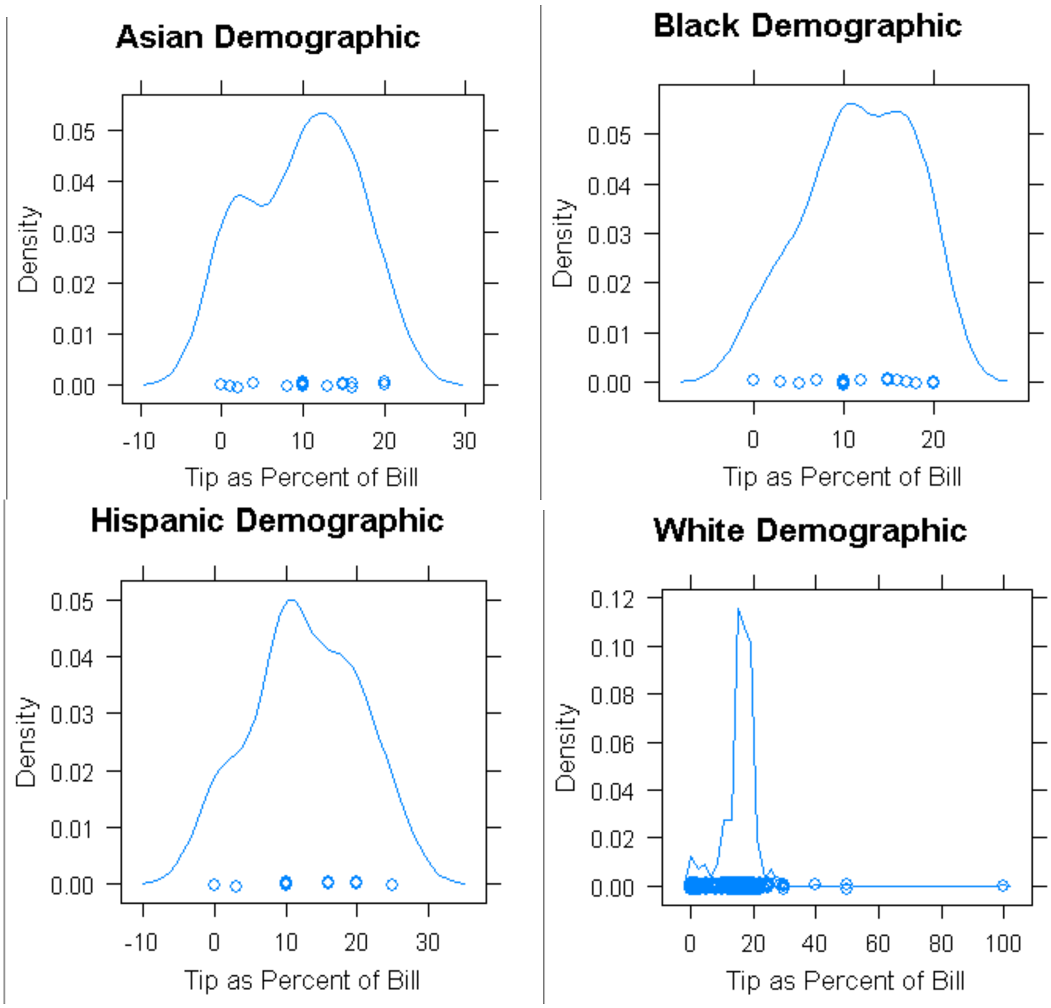
The model which predicts percent tip using the variable age gives an equation of "pcttip = 15.2 + 0.03 * age". We can see from this that the relationship is barely a positive one, which suggests higher tips for older waiters. Despite the seemingly weak visual relationship, the linear model significantly predicts percent tips on a 0.01 level. A scatter plot should show that as age increases, effort level either increases or decreases in order to produce higher percent tip.



This plot seems to show that an increase in age narrows the range of effort. This could suggest that the older waiters level out into an average amount of effort – not too much and not too little. This is consistent with earlier conclusions. If effort level does not have an effect on increasing the amount of tipping, it would be reasonable to conclude an individual would not find it necessary to put forth an extravagant amount of effort. It is interesting to note the response of a very young server. This response had seemingly meaningful responses except for age, so perhaps, if anything, the establishment needs a visit to ensure the adherence to standard labor laws.

Demographics

There are a multitude of factors that servers can change about themselves that impact the size of the compensation that they receive for their services. Among factors that cannot be changed, however, is the demographic of those that the server deals with. Below, we took a look at the days that one of four cultural demographics was in a clear majority within a restaurant establishment. Since a large aggregate of the tips came from that cultural demographic, we assume that this is the likely tip distribution with respect to that background.



As seen above, both the Asian and the Hispanic populations generally leave between –10 (implies they didn't leave enough to cover the bill) to 30 percent with the average being around 10%. The black demographic however pays closer to a minimum of 0 to 30 percent. The aggregate of the population that are classified as "black" however, tip much more consistently between 10 to 15 percent. Unlike any other demographic, white restaurant visitors tip a relatively consistent 15 to 20 percent with a small subset of the population paying between 0 and 10 percent. The type of demographic that visits a particular restaurant is determined by the location of the restaurant, the type of food it serves, and the type of customer who's needs it fulfills most. Zachary Brewster of North Carolina State University states that non-white groups tend to tip below the 15 to 20 percent quota due to a number of reasons. Some of these reasons are: unfamiliarity with tipping norms, household size, education levels, as well as income and other dominant demographic features. (http://userpages.umbc.edu/~clmallin/brew&mallFC.pdf) These as well as a multitude of other reasons explain the distributions of tips by demographic we see in the above density plots. Targeting these dynamics however would require information about the customer which is not only difficult to come by (even if you google it!) but also very difficult to collect in a non-privacy-violating way.

Conclusion

In conclusion, we have shown several things a server can control to improve their tip rate as well as other factors they may not be able to control that could affect their tip rate. Attitude does play a significant role in the success of a waiter and as we saw, a large amount of qualities one would classify as positive and negative actually do assist. Generally, servers could heed the wisdom of their elders and realize putting forth excessive effort does not necessarily do much more than raise the “floor” (minimum) level of tip rates. It would be wisest to consider the activity individually, so as to be most efficient and not over-exert unnecessarily. The data suggests touching, joking, and complementing do not affect the tip they will receive. It also suggests that flair, repeating orders, and drawing or writing thanks on checks will negatively impact their tip rate. Though a server would prefer to select a restaurant to work at based on a demographic, the specific reasons for the differences by demographic become difficult to classify due to the sheer numbers of variables that could be taken into account. A possible extension of this paper could highlight the demographic effects and actually expand on major contributing factors. All things considered, the data supports government statistics that servers on a whole receive 10% to 20% of the bill in gratuity.

Appendix

```
########## PREPARE DATA ##########

ss<-read.csv("http://had.co.nz/stat480/data/server-survey.csv")
ss$id<-1:nrow(ss)
library(ggplot)
library(reshape)
options(scipen=9)
ss$sex[ss$sex == 2] <- NA
props <- !is.na(ss$pcttip) & na.omit(ss$pcttip < 1)
ss$pcttip[props] <- ss$pcttip[props] * 100
ss <- ss[!(!is.na(ss$pcttip) &  ss$pcttip > 70), ]


########## CREATE AGE VARIABLE ##########

ss <- ss[!is.na(ss$birth_yr), ]
birthyear <- ss$birth_yr
birthyear[birthyear==1802] <- NA
birthyear[birthyear==1894] <- 1984
birthyear[birthyear==982] <- 1982
birthyear[birthyear==192] <- NA
birthyear[birthyear<1900]+1900
birthyear[!is.na(birthyear)&birthyear<1900]<-
birthyear[!is.na(birthyear)&birthyear<1900]+1900
```

```
birthyear[birthyear==8197]<-1987
ss$birth_yr<-birthyear
ss <- ss[!is.na(ss$birth_yr), ]
ss$age<-2007-ss$birth_yr


########## CREATE EFFORT VARIABLE ##########

ss$effort<-( ss$flair + ss$intro + ss$selling + ss$squat + ss$touch + ss$jokes +
ss$repeat. + ss$customer_name + ss$draw
+ ss$smile + ss$thanks + ss$weather + ss$complement )
ss <- ss[!is.na(ss$effort), ]
ss$effavg <- ( ss$effort / 13)


########## ANALYZE EFFORT TO PCTTIP ##########

qplot(ss$effort,type="histogram",breaks=seq(10,55,by=1))
qplot(effort,pcttip,data=ss)
qplot(effavg,pcttip,data=ss)
summary(lm(ss$effavg~ss$pcttip))
summary(lm(ss$effort~ss$pcttip))

effc <- function(a){
lm(ss$pcttip~a)$coefficients
}
sapply(ss[,25:37],effc)

effr <- function(a){
summary(lm(ss$pcttip~a))$r.squared
}
sapply(ss[,25:37],effr)

########## ANALYZE AGE TO PCTTIP ##########

qplot(ss$age,type="histogram",breaks=seq(0,70,by=1))
qplot(age,pcttip,data=ss)
lm(ss$pcttip~ss$age)
aov(ss$pcttip~ss$age)
summary(lm(ss$pcttip~ss$age))
qplot(age,effort,data=ss)

########## PREPARE DATA ##########

ss<-read.csv("http://had.co.nz/stat480/data/server-survey.csv")
```

########## ANALYZE ATTITUDES TO ONE ANOTHER ##########

```
ss4<-ss[,61:70]
dim(ss4)
ss4$hasna<-0
i<-0
j<-0
for(i in 1:2618){
for(j in 1:10){
if(is.na(ss4[i,j])){
ss4[i,11]<-1
}
}
}

ss4<-ss4[ss4$hasna==0,]
ss4<-ss4[,1:10]
dim(ss4)
ss4<-ss4[runif(100,0,2346),]
mlt<-melt(ss4,id=('extraverted_enthusiastic'))
qplot(extraverted_enthusiastic,value,data=mlt,colour=variable,type="jitt
er")

ss2<-ss[,61:70]
i<-0
j<-0
for(i in 1:2618){
	for(j in 1:10){
		if(is.na(ss2[i,j])){
			ss2[i,11]<-1
		}
	}
}
ss2<-ss2[ss2$hasna==0,]


ys<-names(ss2)
x11()
for(i in 1:(length(ys)-1)){
plot(ss2$extraverted_enthusiastic,ss2[,i],ylab=ys[i],xlab="extaverted_en
thusiastic")
abline(lm(ss2[,i]~extraverted_enthusiastic,ss2))
}


editedggfluctuation<-function (table, type = "size", floor = 0, ceiling
= max(table$freq,na.rm = TRUE),xlab,ylab)
{
    if (is.table(table))
        table <- as.data.frame(t(table))
    oldnames <- names(table)
    names(table) <- c("x", "y", "freq")
    table <- add.all.combinations(table, list("x", "y"))
    table <- transform(table, x = as.factor(x), y = as.factor(y))
    if (type == "size") {
        table <- transform(table, freq = sqrt(pmin(freq,
ceiling)/ceiling),
```

```
            border = ifelse(is.na(freq), "grey90", ifelse(freq >
                ceiling, "grey30", "grey50")))
        table[is.na(table$freq), "freq"] <- 1
    }
    table <- subset(table, freq * ceiling >= floor)
    if (type == "size") {
        p <- ggtile(ggplot(table, aesthetics = list(x = x, y = y,
            height = freq, width = freq, fill = border)), colour =
"white")
        p <- scmanual(p, "fill")
    }
    else {
        p <- ggtile(ggplot(table, aesthetics = list(x = x, y = y,
            fill = freq)), colour = "grey50")
        p <- scfillgradient(p, low = "white", high = "darkgreen")
    }
    p$xlabel <- xlab
    p$ylabel <- ylab
    p
}

editedggfluctuation(table(ss2$extraverted_enthusiastic,ss2[,2]),ylab=ys[
2],xlab="Extraverted_Enthusiastic")
editedggfluctuation(table(ss2$extraverted_enthusiastic,ss2[,3]),ylab=ys[
3],xlab="Extraverted_Enthusiastic")
editedggfluctuation(table(ss2$extraverted_enthusiastic,ss2[,4]),ylab=ys[
4],xlab="Extraverted_Enthusiastic")
editedggfluctuation(table(ss2$extraverted_enthusiastic,ss2[,5]),ylab=ys[
5],xlab="Extraverted_Enthusiastic")
editedggfluctuation(table(ss2$extraverted_enthusiastic,ss2[,6]),ylab=ys[
6],xlab="Extraverted_Enthusiastic")
editedggfluctuation(table(ss2$extraverted_enthusiastic,ss2[,7]),ylab=ys[
7],xlab="Extraverted_Enthusiastic")
editedggfluctuation(table(ss2$extraverted_enthusiastic,ss2[,8]),ylab=ys[
8],xlab="Extraverted_Enthusiastic")
editedggfluctuation(table(ss2$extraverted_enthusiastic,ss2[,9]),ylab=ys[
9],xlab="Extraverted_Enthusiastic")
editedggfluctuation(table(ss2$extraverted_enthusiastic,ss2[,10]),ylab=ys
[10],xlab="Extraverted_Enthusiastic")
```

########## ANALYZE ATTITUDES TO PCTTIP ##########

```
ss5<-ss[,c(22,61:70)]
i<-0
j<-0
ss5$hasna<-0
for(i in 1:2618){
      for(j in 1:11){
            if(is.na(ss5[i,j])){
                  ss5[i,12]<-1
            }
      }
}
ss5<-ss5[ss5$hasna==0,]
dim(ss5)
```

```
i<-0
for(i in 1:2327){
ss5$posprop[i]<-sum(ss5[i,c(2,4,6,8,10)])
ss5$negprop[i]<-sum(ss5[i,c(3,5,7,9,11)])
}
ss5<-ss5[!(ss5$pcttip==500),]
meltss5<-melt(ss5,id=names(ss5[,1:12]))
qplot(value,pcttip,data=meltss5,.~variable,type="jitter")
```

########## ANALYZE DEMOGRAPHICS TO PCTTIP ##########

```
summary(ss$asian_prop)
summary(ss$black_prop)
summary(ss$hispanic_prop)
summary(ss$white_prop)
asntip<-ss[(ss$asian_prop>=75),22]
blktip<-ss[(ss$black_prop>=75),22]
hisptip<-ss[(ss$hispanic_prop>=75),22]
wittip<-ss[(ss$white_prop>=75),22]
asntip<-na.omit(asntip)
blktip<-na.omit(blktip)
hisptip<-na.omit(hisptip)
wittip<-na.omit(wittip)
densityplot(asntip,xlab="Tip as Percent of Bill",main="Asian
Demographic",breaks=c(seq(-10,100,by=2)))
densityplot(blktip,xlab="Tip as Percent of Bill",main="Black
Demographic",breaks=c(seq(-10,100,by=2)))
densityplot(hisptip,xlab="Tip as Percent of Bill",main="Hispanic
Demographic",breaks=c(seq(-10,100,by=2)))
densityplot(wittip,xlab="Tip as Percent of Bill",main="White
Demographic",breaks=c(seq(-10,100,by=2)))
```