Ruben Pena
MA3740
Due:10/15/2018

Project 1

Mr. Chesterfield,

I'm glad to be of assistance! Here is my analysis and breakdown of your customer data that you provided. I will go in the same order as your queries.

1. How old are our customers?

Using SAS, I was able to determine that, on average, the age of your customers is 35 years old. The table below shows both the mean (average age) and median ages of your customers. Mean refers to the average of ages. In other words, if you were to add all the ages together and divide that sum by the total number of ages (in this instance 554) you would get a mean of 35. The median refers to the center of the ages. If you were to line all ages in order on a line the median will cut these in half (in the simplest of explanations).

Analysis Variable: age

N	Mean	Median	Std Dev
554	35.0920578	34.0000000	9.0183964

A great way to visualize distributions is a histogram. The chart below shows a graphical representation of the ages of your customers.
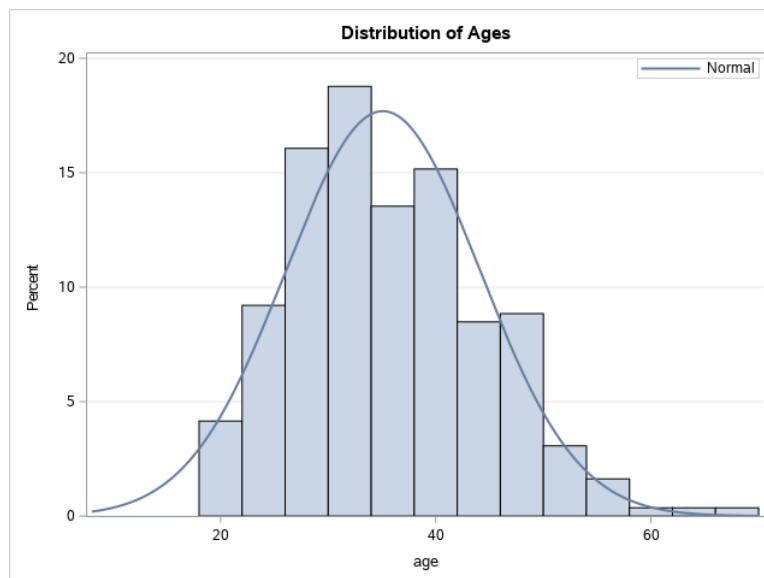
| Analysis Variable: age	
Lower 95% CL for Mean	Upper 95% CL for Mean
34.3394407	35.8446749

To check that the estimation of the average age of your customers is 35, I checked the confidence interval. This means that there is a 95% likelihood that the actual average age of your population of customers lies between 34 and 35.8 or 36.

2. How do people hear about our website?

According to the data provided in the "hear" column, there are four ways for your customers to hear about your website: Facebook, friend, Google, and magazine. A quick analysis shows that most of your customers hear about your website from Facebook.

```
faceb friend google  magaz
  250    118    121     65
```


How Customers Find Your Site
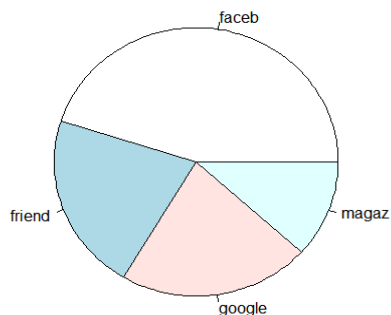
How Customers Find Your Site
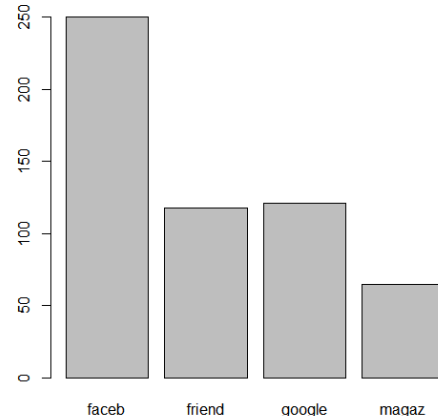
3. Is there a relationship between the amount of money people spend and how they heard about our website? If so, what group(s) are spending more or less money?

In order to answer this question, I will plot boxplots of the amount of money spent by each group and then compare the means and medians of the groups.

Ruben Pena
MA3740
Due:10/15/2018

First the visuals:
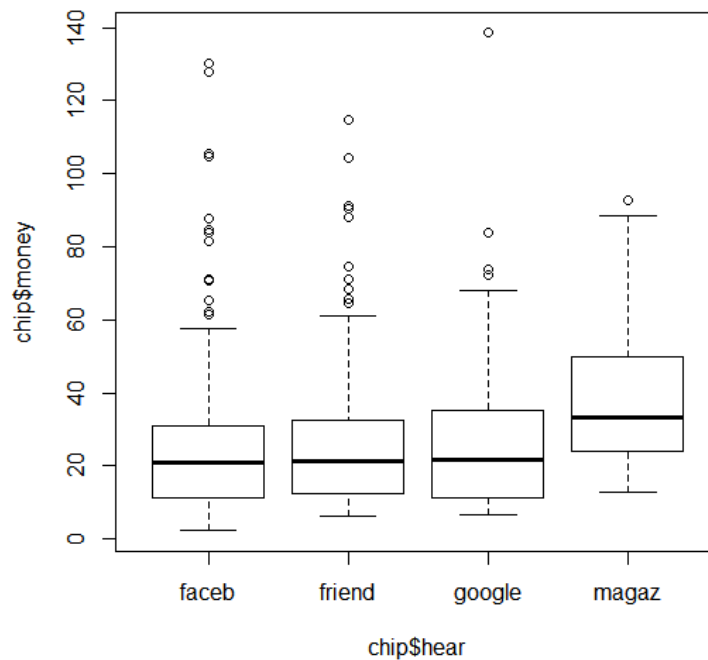
These plots are difficult to interpret without some statistical training. The values on the vertical axis show the amount of money, while the horizontal are how the customers hear about your site. It shows that the means (the bold solid bar inside each box) for each group are roughly between $20 and $40.

To get a clearer idea, I created a table that aggregates the values for each way that a customer hears about your site.

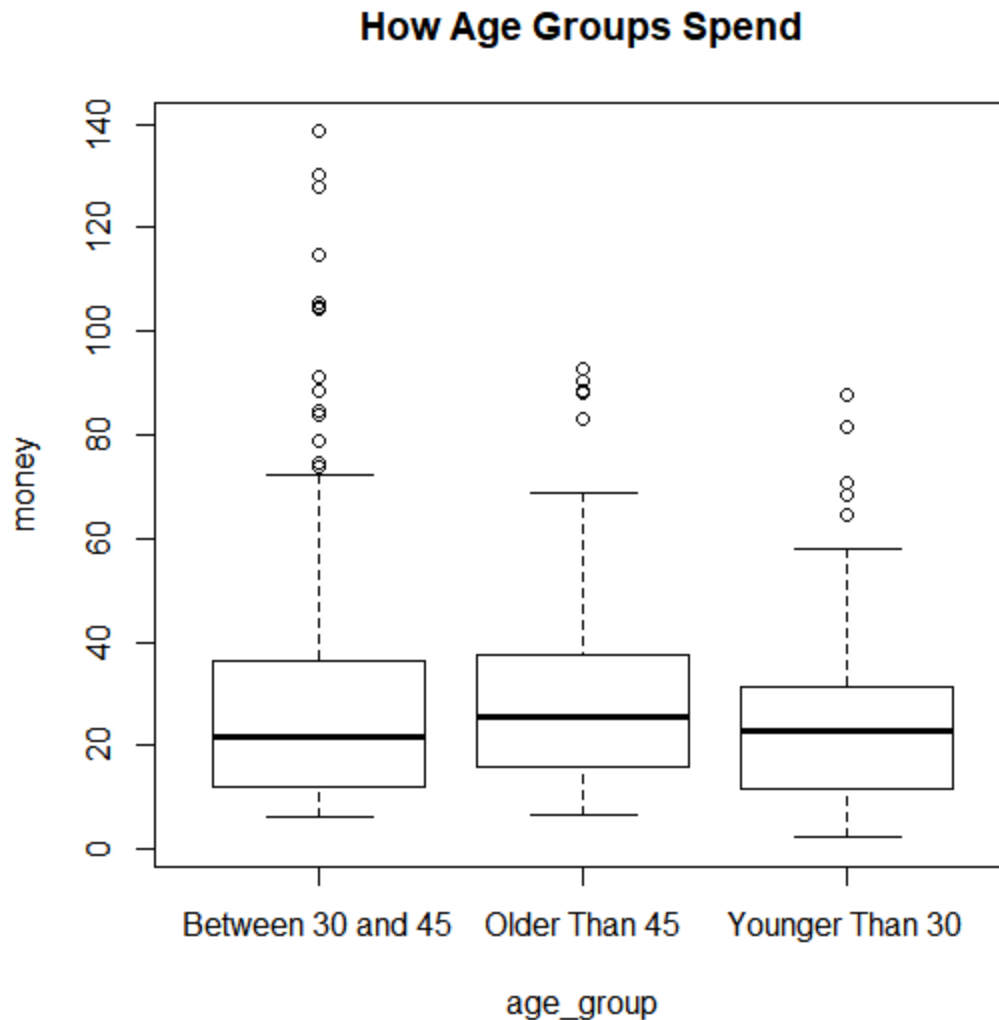| hear	money.Min.	money.1st Qu.	money.Median	money.Mean	money.3rd Qu.	money.Max.
faceb	2.22000	11.19000	20.78000	24.91180	31.00000	130.28000
friend	6.30000	12.42500	21.19000	27.26347	32.60500	114.92000
google	6.58000	11.45000	21.51000	25.93041	35.20000	138.57000
magaz	12.84000	23.84000	33.39000	39.60600	50.07000	92.81000

A very noticeable behavior from this table is that customers that hear about your site from magazines tend to have larger minimum and average purchases but are the smallest proportion of your customer base. If you were to increase magazine marketing, you could potentially see a substantial increase in revenue.

It can also be noted that customers who hear about your site through Facebook have the smallest minimum and average, but also some of the largest maximum purchases. Coupled with them being the largest proportion of your base makes them the largest source of your revenue. A promotion that brings up the average spending of Facebook users would be very beneficial for revenues.

4. How many customers belong to each group of Ages? Do these different age groups typically spend different amounts of money on their orders?

As this question is similar to the previous question, I'll use the same format for explaining this one.

The graphic:

How Age Groups Spend

From this graphic we can see that, on average, customers older than 45 spend more. Let's take a look at a summarized table of these results.

age_group	money.Min.	money.1st Qu.	money.Median	money.Mean	money.3rd Qu.	money.Max.
Between 30 and 45	6.27000	12.05500	21.56000	28.19032	36.30500	138.57000
Older Than 45	6.62000	16.02000	25.46000	31.27457	37.50000	92.81000
Younger Than 30	2.22000	11.69000	22.68000	23.83294	31.24500	87.82000

The results of the table verify. It can be noted that customers under 30 tend to spend less on average and have smaller minimum and maximum purchases. Customers over 45 may have the most average spending, but customers between 30 and 45 aren't far behind in average spending but have far higher maximum spending. This is also verified in the visual by the open circles in the higher ranges on the vertical axes(money).

5. **What percent of our customers are buying gifts for people? Are gift-buyers typically older or younger than customers who are not gift-buyers?**
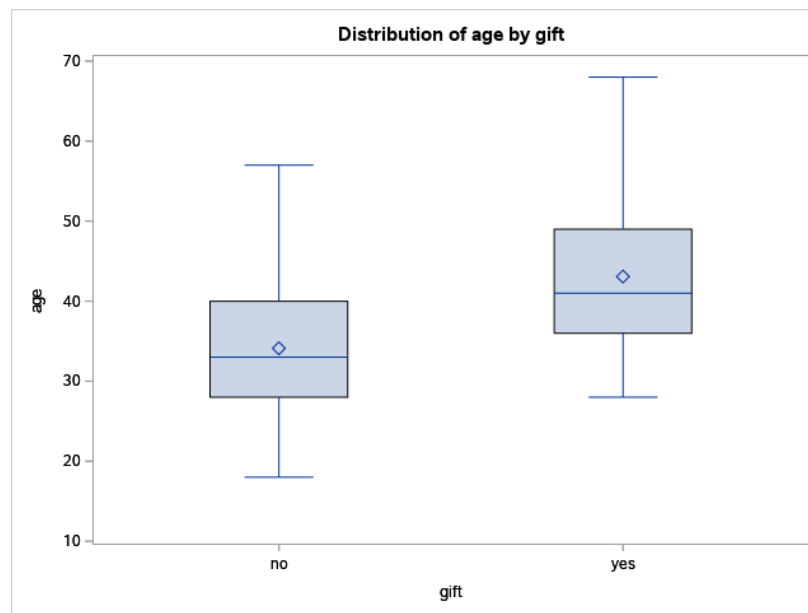
These boxplots show that your gift-buyers tend to be older than 30 years old with the average age (mean) being near 43(the diamond shape). The median age of your gift buyers is near 41 years of age. It should be noted that the number of non-gift-givers significantly outweighs the number of gift-givers. The tables below show that it is a 494:60 proportion. The actual percent of customers buying gifts would be 60/554 or 10.83%.

gift=no

Analysis Variable: age		
N	Mean	Median
494	34.1214575	33.0000000

gift=yes

Analysis Variable: age		
N	Mean	Median
60	43.0833333	41.0000000

6. Is it true that customers under the age of 26 who saw the advertisement on Facebook are spending less money per order than people who are not in this group?

For this question, I subset the data to include only customers from Facebook and added the group for Under 26. On the bottom of the boxplot the x axis 0 represents customers who are Over 26, while the 1 represents customers under 26.

Those Facebook customers under 26 did indeed spend less money on average than customers over 26.
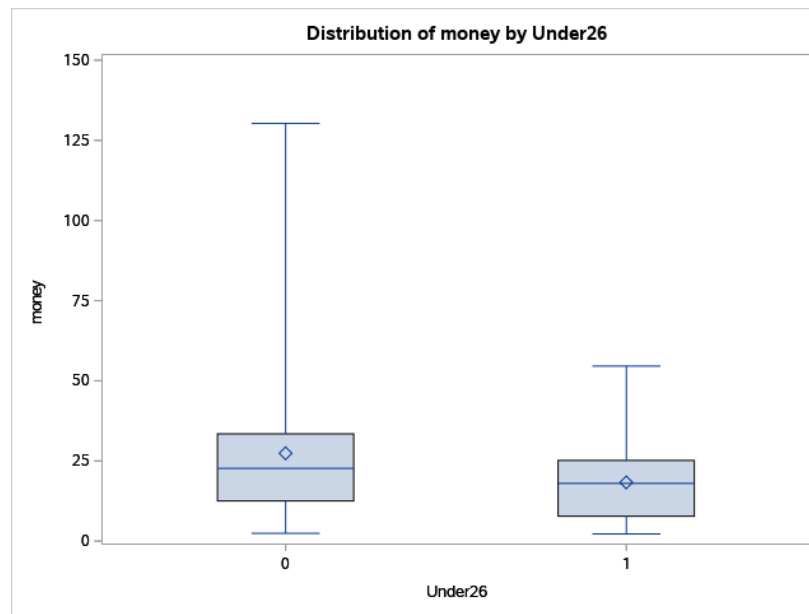
This can also be seen in a table format of Money by the two Age Groups. The 250 Facebook Customers over the age of 26 made purchases with higher average and median values than the 69 Facebook Customers under 26.

Under 26   **Over 26**

Analysis Variable: money		
N	Mean	Median
69	18.3215942	18.0000000

Analysis Variable: money		
N	Mean	Median
250	24.9118000	20.7800000

APPENDIX

Appendix 1

#SAS University Edition – SAS Studio

```
/*Question 1*/
FILENAME CSV "/folders/myfolders/chip.csv" TERMSTR=CRLF;
/** Import the CSV file.  **/

PROC IMPORT DATAFILE=CSV
                 OUT=chip
                 DBMS=CSV
                 REPLACE;
RUN;

/** Print the results. **/

PROC PRINT DATA=chip; RUN;

/** Unassign the file reference.  **/
FILENAME CSV;

/*--Analysis of Age--*/
proc means data=chip n mean median std; var age;run;

/*--Histogram of Age--*/
/*--From SAS University Snippets--*/
title 'Distribution of Age';
proc sgplot data=chip;
  histogram age;
  density age / lineattrs=(pattern=solid);
  keylegend / location=inside position=topright across=1;
  yaxis offsetmin=0 grid;
run;

/*--95% Confidence Interval--*/
proc means data=chip alpha=.05 clm; var age;run;

/*--Question 5 */
proc sort data=chip; by gift;
proc means data=chip n mean median;var age;by gift;
proc boxplot data=chip; plot age*gift;run;
```

Appendix 2

```
/* --Question 6-- */
data sub_chip; set chip; if hear='faceb';
proc print data=sub_chip; run;
data faceb; set sub_chip; if (age < 26) then Under26=1;else Under26=0;
proc print data=faceb;run;

/*--Box Plot--*/
proc sort data=faceb; by Under26;
title 'Money for Faceb Customersby Under/Over26';
proc boxplot data=faceb; plot money*Under26;run;

data under; set faceb; if Under26=1;
proc means data=under n mean median; var money; run;
data over; set faceb; if Under26=0;
proc means data=faceb n mean median; var money;run;
```

Appendix 3

```
#Using R Studio, attached using file viewer

chip <- read.csv("~/GitHub/MA3740.Lab/Project1/chip.csv")

require(dplyr)

groups = function(x)

 {

 if (x <30)  {result = "Younger Than 30"} else

 if (between(x,30,45)) {result = "Between 30 and 45"} else

 if (x >45) {result = "Older Than 45"}

 return(result)

}

age_group = sapply(chip$age,groups)

chip=cbind(chip,age_group)

plot(money~age_group,data=chip,main="How Age Groups Spend")

aggregate(money~age_group,data=chip,summary)
```