Researchers: *Sapir Dooley, Emily Gill, Austin Pigg, Sedrick Utt*

Presentation Title: TK Paddles LLC Customer Data Analysis

Research Focus: Profit data and customer demographic data related to Asian-American Greek organizations.

Abstract:

There are hundreds of individual Greek-Letter organizations in the United States, many of which have sub-branches or "chapters" numbering anywhere from two to fifty or more. Of these organizations, a little over one hundred are of Asian Pacific Islander Desi American (APIDA) backgrounds. It is in these organizations that the tradition of gifting decorative paddles - adorned with organizational letters and other customizations - is most common. Through our research, we have made several findings regarding the distributions of these organizations, the member and organization densities across the United States, and the temporal nature of the tradition.

Our research indicates that the distribution of these organizations reaches most of the continental United States, save for much of the Rocky Mountain regions, with the highest number and greatest diversity of operating organizations in the Northeast. By researching the temporal distribution of profit earned by TK Paddles, it becomes evident that "Paddle Gifting" is seasonal in nature, with the peaks centering around March and December. Our research reflects the current client base of TK Paddles and what entities should be pursued further in a cost-benefit analysis sense.

Introduction

In the United States, there are hundreds of Greek-Letter organizations. Many of these organizations are made up of chapters that are present on campuses throughout the nation. Each chapter of any given organization will have a population of members whose numbers will vary. In some cases, the number of active members in a chapter can reach into the hundreds. While it is not the case for all of the organizations present on college campuses, many are founded to represent individual demographics of the larger population. These demographics are usually communities of individuals that share cultural or religious ties.

One of these communities is the Asian Pacific Islander Desi American (APIDA) community. Of the Greek-Letter organizations represented in the United States, over 100 are made up of individuals of APIDA backgrounds. In APIDA organizations it is a common tradition for current members to welcome in new members by gifting them decorative paddles. The paddles often showcase the letters of the organization and are customized with other decorations. This could require a purchase of dozens of paddles for each chapter, depending on the number of new members.

TK Paddles is a small woodworking company that manufactures the decorative paddles used by these Greek-letter organizations. The owner of TK Paddles knows his biggest clients are the APIDA organizations that give paddles to their new members at the end of a semester. To grow the business and maximize profits, the owner wants to understand where the large APIDA communities are located throughout the country to help with focusing advertising efforts. Additionally, a large portion of the organizations of interest are licensed to purchase paddles through a competitor of TK Paddles. This being the case, the owner of TK Paddles would benefit from knowing which organizations are under this license.

Methods

In our research, we analyze four study groups: 2018 customers, 2019 customers, 2020 customers, and future prospective customers. Within the first three groups are the following attributes: organization name, locations (city, state, latitude, and longitude), and profit. The attributes within the final group are the same with the addition of total member count, whether or not the organization has a history of paddle gifting, whether or not the organization has their name licensed, as well as links to the organization's website and Instagram.

Data Gathering

Our research area includes the entirety of the contiguous United States, with specific data points collected at individual universities and institutions of higher learning. Data is gathered manually through several avenues. First, an extensive list of organizations is compiled using the list available on Wikipedia and the list found in the National Asian Pacific Islander Desi American Panhellenic Association (NAPA) resource guide. Next, geographic information - including latitude and longitude, city and university - is found from organizations' national websites and through the "Asian American Greek Survey" (Hayward) and confirmed through University websites. Paddle gifting information as well as some chapter size information is found in "History of Asian American Fraternal Organizations" and collected manually through organizations' social media accounts including Instagram and Facebook. A majority of the chapter size data is collected from chapter websites. Finally, licensing status is obtained from the largest greek licensing organization, "Affinity."

Data Cleaning

A variety of different methods were used to clean the data. The "date" attributes of the original data were formatted from mm/dd/yyyy to yyyy-mm-dd using Python 3's datetime package in Jupyter Notebook. In lines 45 through 102 of the R-code lies the rest of the data cleansing. Here the profit data is converted to integers from strings. The attributes "cumsum" and "meanProfit" are produced to represent the cumulative profit over the span of a year and the mean profit for each year. Each entry is given a unique ID, and the distance of each entry is calculated from College Station.

Figures

Figures 1 through 15 are made using the 2018, 2019, and 2020 customer demographics datasets, figures 16 through 31 are made using the future customer demographics dataset. The first set of 6 figures (fig. 1 - 6) depict the distribution of individual orders and their related information . The first 3 figures (fig. 1 -3) are maps that show the organization, location, and profit made per order in a static plot using the geom_polygon and geom_point functions of ggplot2. The second three figures (fig. 4 - 6) are maps depicting the same information in a slightly easier format using an interactive plot also made with geom_polygon and geom_point but then put through ggplotly. Both sets were made using each tables' "latitude" and "longitude" attributes. Figures 7 - 9 are box and whisker plots depicting the variation in profit per order for each organization, these tables were made using the "profit" and "organization" attributes in the geom_boxplot function of ggplot2. Figures 10 through 12 are dot plots showing the temporal distribution of orders and their related profit and organization using attributes of the same name and the "date" attributes in the geom_point function of ggplot2. Figures 13 - 15 are line plots showing the net profit made

for each year from 2018 to 2020, made using the “cumsum” attribute mentioned earlier in the geom_point and geom_line functions of ggplot2. The next 10 figures were made in the same fashion as the first 6, the first half made in ggplot2 and the second half made using ggplotly. Figures 16 and 22 show the distribution of all organizations related in some way to previous customers and their related data including organization, university, and total members. Figures 17 and 23 show a filter of figures 16 and 22, showing only organizations with a number of members greater than 200 by creating a subset of the ‘future’ dataset where ‘totalMembers” > 200. Figures 18 and 24 show the same data as figures 16 and 22, however this time the colors represent organizations that do or do not have a history of paddle gifting. Figures 19 and 25 show the distribution of organizations with or without licensing agreements. Figures 20 and 26 filter the distribution by organization type: sorority, fraternity, or co-ed. Figure 27 shows the distance of each organization to College Station. Distances were calculated by first creating entry IDs. Then a dummy variable ‘future$ZZZ=1’ is created so that data can be joined to itself. Then the line ‘temp <- future %>% full_join(future,c("ZZZ"="ZZZ")) %>% ’ creates a data frame of all combinations of rows from ‘future’. The second line filters out the rows with the same ID. The last line computes the distance. Figure 28 shows the same information as figures 19 and 25 in the form of a pie chart using the geom_bar function modified using coord_polar. Figure 28 does the same thing with the information from figures 18 and 24. Figure 29 shows the different states represented in the future dataset in the form of a pie chart. Figure 30 shows the same data as figures 20 and 26 in the form of a pie chart as well.

Results

Figure 1.

Figure 2.

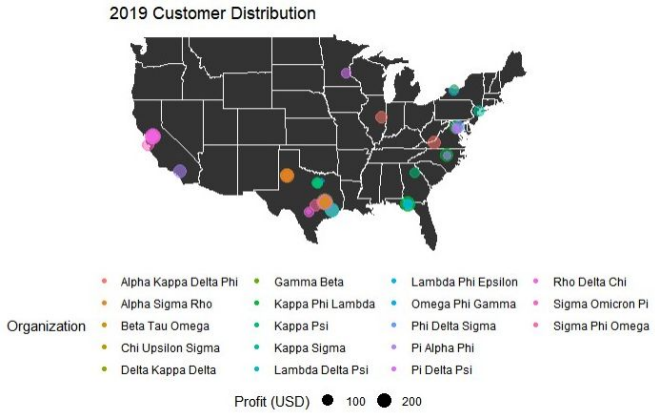


Figure 3.

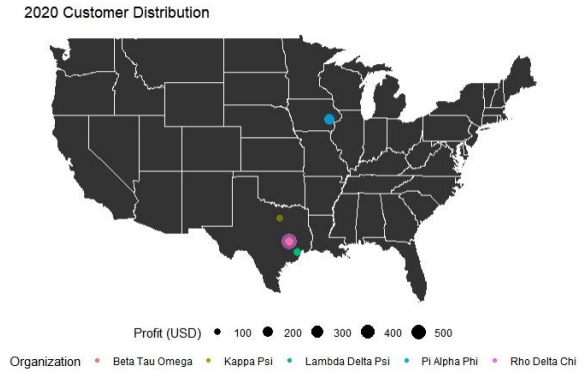


Figure 4.

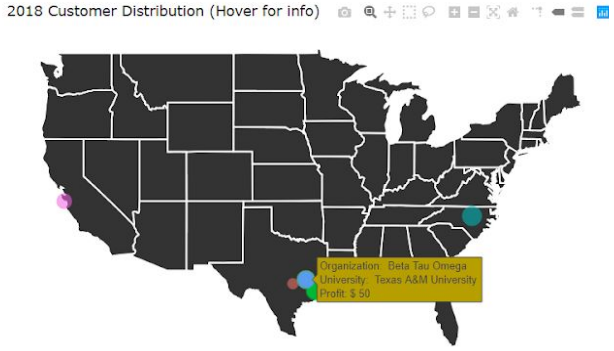


Figure 5.

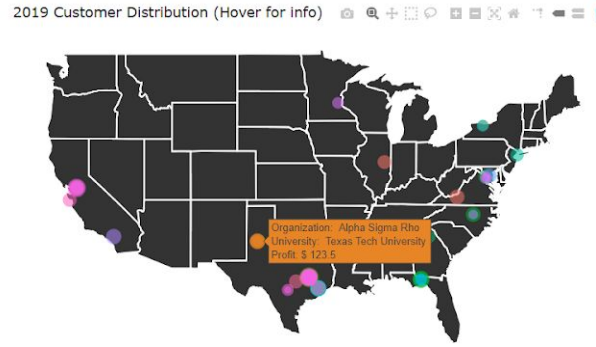


Figure 6.

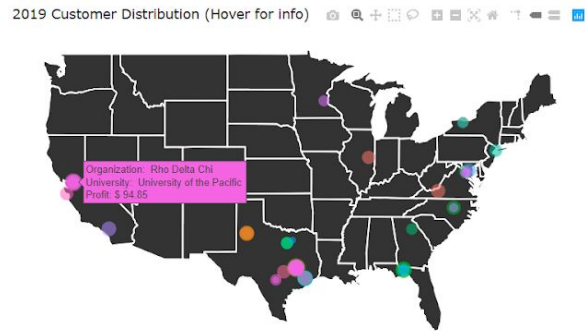


Figure 7.

Figure 8.

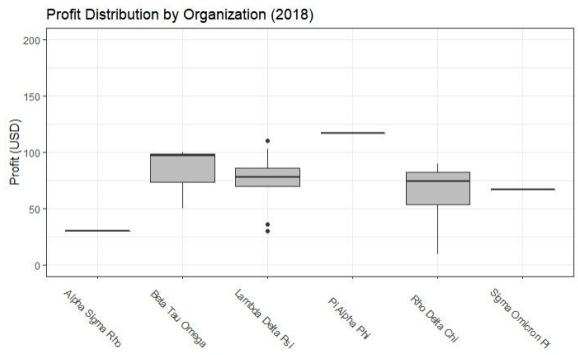


Figure 9.

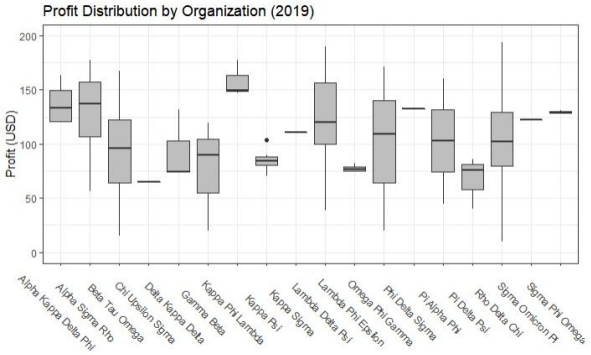


Figure 10.

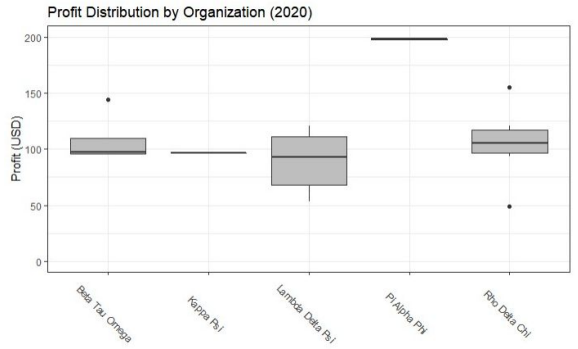


Figure 11.

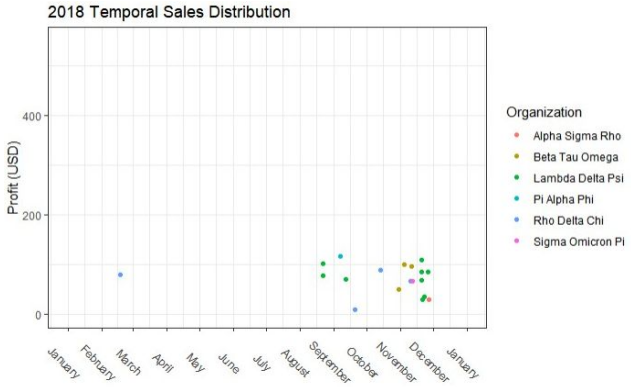


Figure 12.

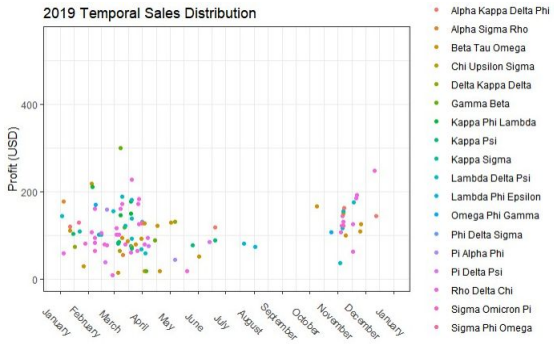


Figure 13.

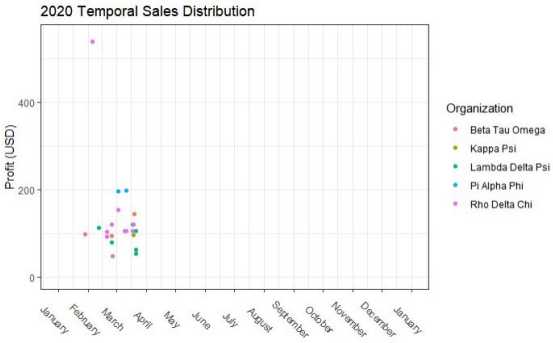


Figure 14.

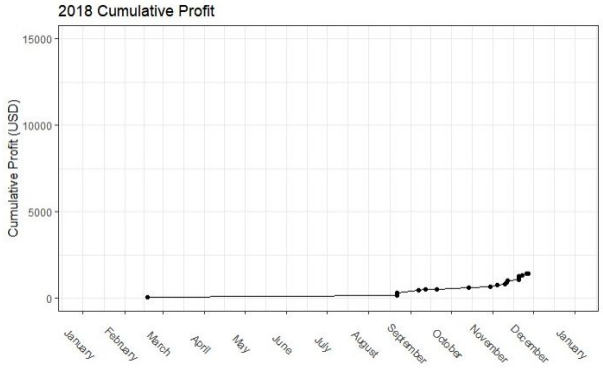


Figure 15.

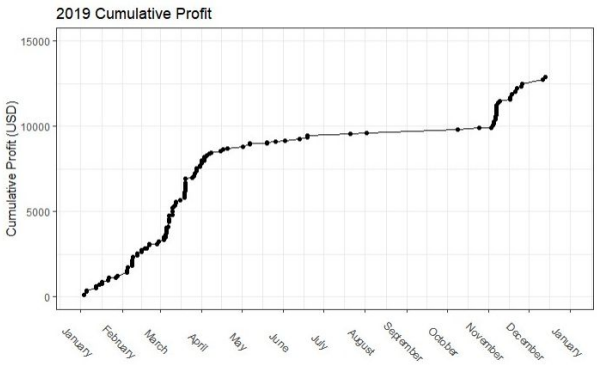


Figure 16.

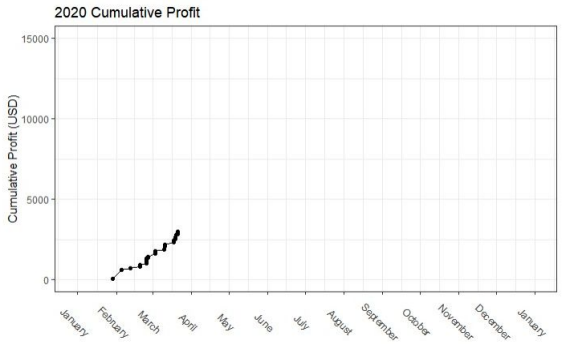


Figure 17.

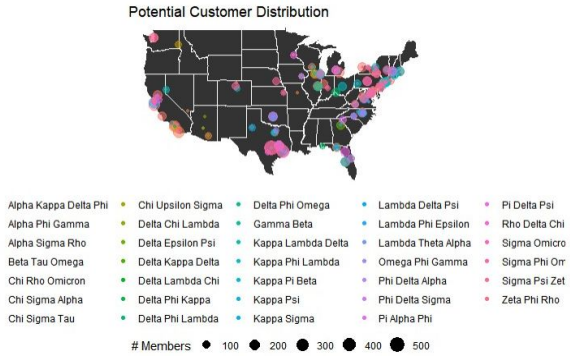


Figure 18.

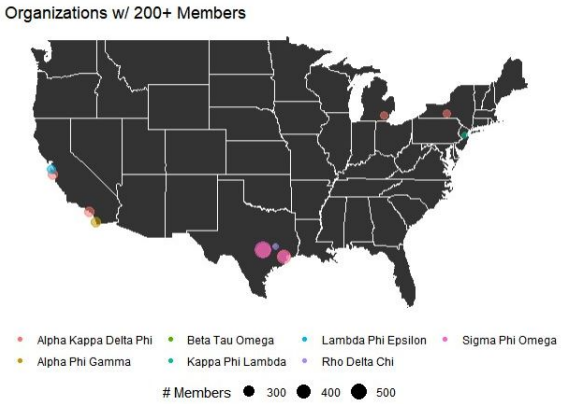


Figure 19.

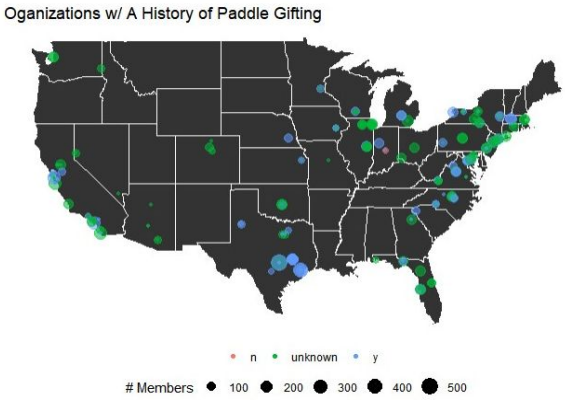


Figure 20.

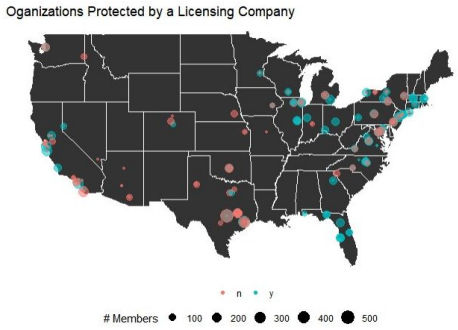


Figure 21.

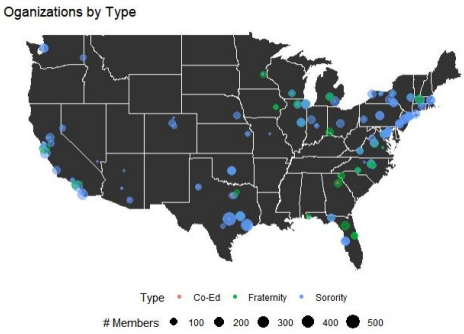


Figure 22.

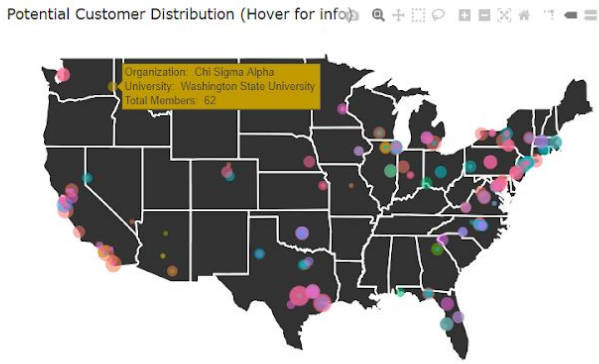


Figure 23.

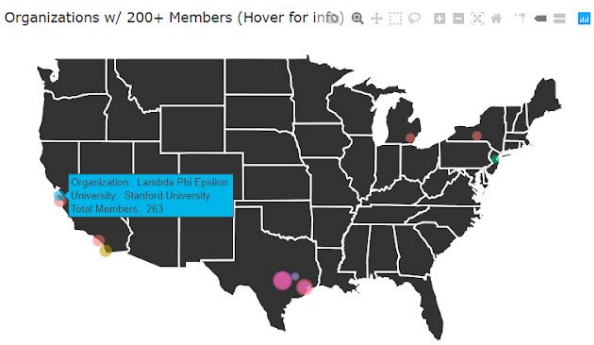


Figure 24.

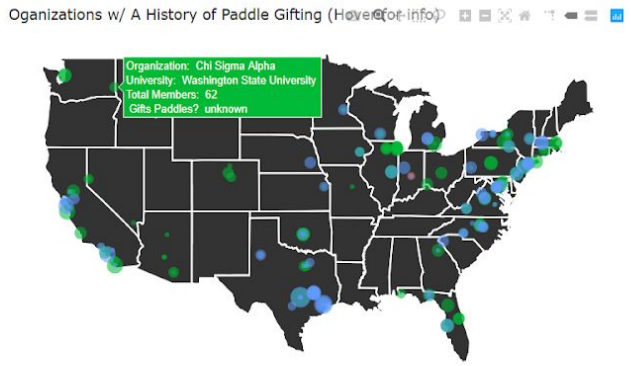


Figure 25.

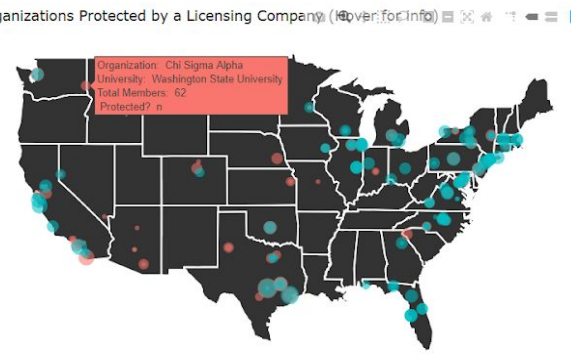


Figure 26.

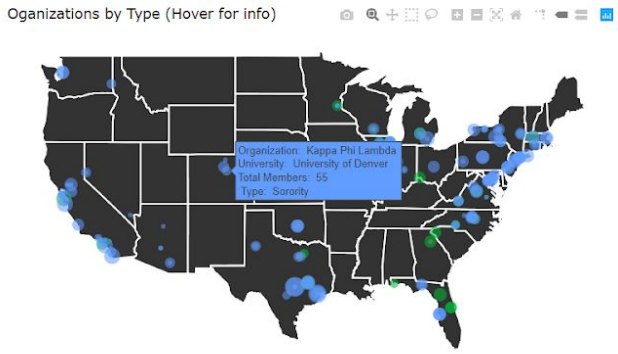


Figure 27.

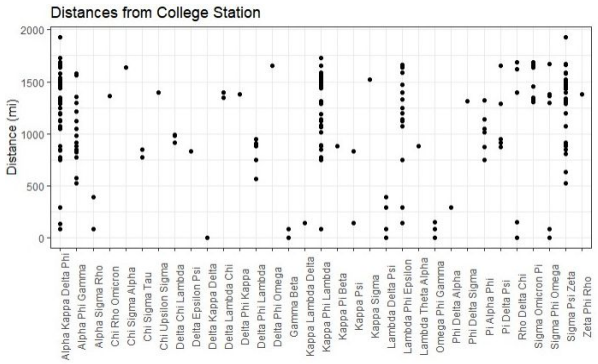


Figure 28.

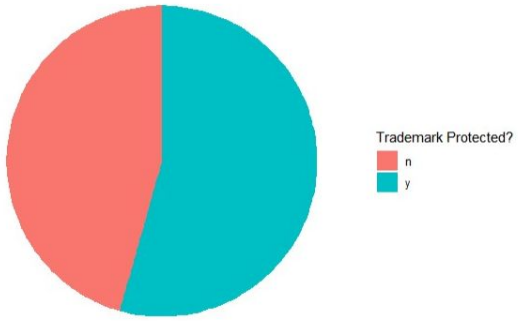


Figure 29.

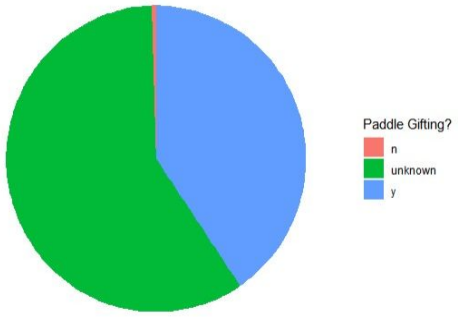
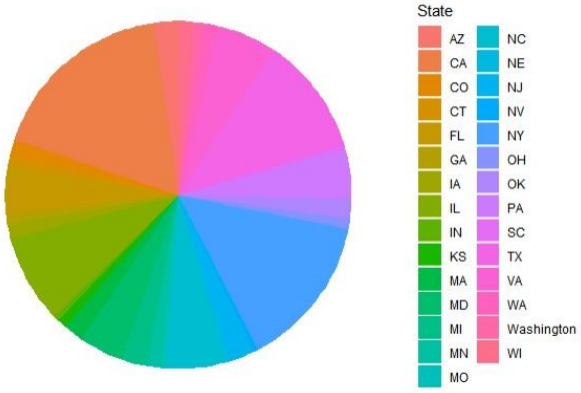
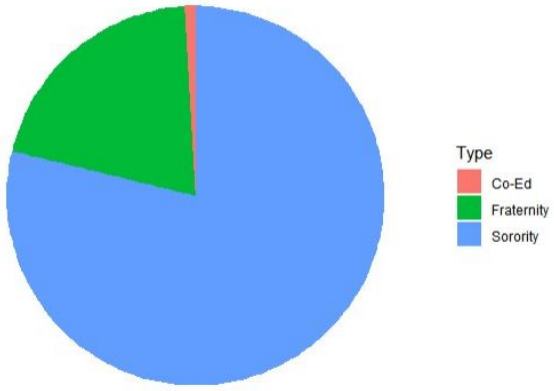


Figure 30.





Discussion

Before collecting and analyzing the data, we hypothesized that at least 70% of Asian-American Greek organizations in the contiguous United States are clients of the trademark-licensing company, "Affinity Licensing". Having found that this majority is true, with 71.6% of organizations being licensed through Affinity, we fail to reject the hypothesis. With this result, we can confirm that it would be in the benefit of TK Paddles to become licensed with Affinity, because it would expand the potential customer base enormously.

By mapping the locations of TK Paddles customers throughout previous years, we found that most customers have come from Texas. While this can likely be attributed partially to the fact that the company is based out of Texas, the underlying factor driving this clustering was made clear by mapping the locations of Asian-American Greek organizations that are protected by a licensing company. While most organizations in Texas are not protected, many organizations in the Great Lakes region, along the East Coast, and in California are. Acquiring this license would unlock customers literally all across the nation.

The primary goal of this study was to collect, analyze, and visualize data that would assist in creating a better marketing strategy for TK Paddles. The conclusion that obtaining an Affinity license would be profitable is important, but the data revealed other patterns that could help increase profit. Graphing the distribution of sales over time and the cumulative profit over time revealed a pattern of seasonality; March and December are by far and away the most profitable months of the year. With this information, we can suggest that it would be most effective to focus resources, such as paid advertising, during the months preceding and including March and

December. It may also be suggested to do the opposite over the summer, when sales are relatively flat.

During our study, the only limitation we came across was the lack of previous relevant research. This is likely due to the fact that Greek paddle-gifting is a niche market. This experiment seems to be the first of its topic, so there are improvements that could be made to future studies. One improvement would be collecting data on the percent of Affinity-licensed organizations that take part in paddle-gifting. This would give a better idea of if acquiring the license would be beneficial. Another improvement would be to analyze other licensing groups as well. A more thorough investigation of multiple licensing firms would show if it is worth getting licensed with a different one.

Through data collection and analysis, we have found that there are patterns and processes that TK Paddles can utilize to maximize their customer base and profit. Affinity Licensing is the most prominent licensing firm for Asian-American Greek organizations, with over 70% of the organizations licensed through them. For this reason, it would be a profitable decision for TK Paddles to obtain this license.

References

Badruddin, Bilal. "About." *National APIDA Panhellenic Association*, 2019, www.napahq.org/about/.

"Category:Asian-American Fraternities and Sororities." *Wikipedia*, Wikimedia Foundation, 11 Feb. 2019, en.wikipedia.org/wiki/Category:Asian-American_fraternities_and_sororities.

"Cultural Interest Fraternities and Sororities." *Wikipedia*, Wikimedia Foundation, 16 Apr. 2020, en.wikipedia.org/wiki/Cultural_interest_fraternities_and_sororities.

Dooley, Carlo. *TK Demographic Data - 2019*, 2019, docs.google.com/spreadsheets/d/1iXL2YS7bh_0q5Jtia57-AG1ApFJdXwm5AES2ZjBnx_c/edit.

Dooley, Carlo. *TK Demographic Data - Potential*, 2020, docs.google.com/spreadsheets/d/1JBPqzHCxWqZUFS31unugrVHMD54ogRoElMPbQu13_Zs/edit#gid=0.

Dooley, Carlo. *TK Demographic Data 2020*, 2020, docs.google.com/spreadsheets/d/1K_PAj75N6fET-opqMP839yJnLqmUjg195jZHvZj1m-w/edit#gid=0.

Dooley, Carlo. *TK Paddles Demographic Data - 2018*, 2018, docs.google.com/spreadsheets/d/1GdnWUMDaBQ8OOb0c7Bg95nf-rfnGDhKfgdKetliIMbE/edit

Dosono, B, et al. "History of Asian American Fraternal Organizations." 2019.

Dosono, Bryan, et al., editors. "NAPA Resource Guide." 2019.

Hayward, N. “Asian American Greek Survey.” 2008.

R-Code

````
---
title: "TK Paddles LLC Customer Data Analysis"
author: "Sapir Dooley, Emily Gill, Austin Pigg, Raiden Utt"
date: "4/17/2020"
output:
  html_document:
    df_print: paged
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:


```{r import packages}
# install.packages("ggplot2")
library("ggplot2")
# install.packages("dplyr")
library("dplyr")
# install.packages("maps")
library("maps")
# install.packages("viridis")
library("viridis")
#install.packages("hrbrthemes")
library("hrbrthemes")
# install.packages("plotly")
library("plotly")
#install.packages("tidyr")
library("tidyr")
#install.packages("lubridate")
library("lubridate")
#install.packages("zoo")
library("zoo")
#install.packages("scales")
library(scales)

```


```{r load data}
# Create paths
path = "/MyFiles/University/SP_2020/GEOG_312/Assignments/finalProject"
eighteenPath <- file.path(path, "demographics2018.csv")
nineteenPath <- file.path(path, "demographics2019.csv")
twentyPath <- file.path(path, "demographics2020.csv")
futurePath <- file.path(path, "futureDemographics.csv")
````
````

```
#read in data
eighteen <- read.csv(eighteenPath, header = T)
nineteen <- read.csv(nineteenPath, header = T)
twenty <- read.csv(twentyPath, header = T)
future <- read.csv(futurePath, header = T)

#convert profit to factor
eighteen$profit <- as.numeric(eighteen$profit)
nineteen$profit <- as.numeric(nineteen$profit)
twenty$profit <- as.numeric(twenty$profit)

#convert dates to dates
eighteen$date <- as.Date(eighteen$date)
eighteen$month_day <- as.Date(eighteen$month_day)
nineteen$date <- as.Date(nineteen$date)
nineteen$month_day <- as.Date(nineteen$month_day)
twenty$month_day <- as.Date(twenty$month_day)

#add cumulative sum and mean profit
eighteen <- eighteen %>%
  mutate(cumsum = cumsum(profit)) %>%
  drop_na(cumsum)
  eighteen$meanProfit <- mean(eighteen$profit)
nineteen <- nineteen %>%
  mutate(cumsum = cumsum(profit)) %>%
  drop_na(cumsum)
  nineteen$meanProfit <- mean(nineteen$profit)
twenty <- twenty %>%
  mutate(cumsum = cumsum(profit)) %>%
  drop_na(cumsum)
  twenty$meanProfit <- mean(twenty$profit)

# add data IDs
future$ID <- seq.int(nrow(data))

# calculate distances from college station
future$ZZZ=1
temp <- future %>% full_join(future,c("ZZZ"="ZZZ")) %>%
        filter(ID.x != ID.y) %>%
                              mutate(dist=(sqrt(((longitude.x-longitude.y)^2    +
(latitude.x-latitude.y)^2))*100)*0.621371)
distances  <-  subset(new,organization.x  ==  "Beta  Tau  Omega",  select  =
c("organization.x", "city.x","organization.y", "city.y", "dist"))


head(eighteen)
head(nineteen)
head(twenty)
head(future)
#head(future1)
head(distances)
```
````{r variables}
usa <- map_data("usa")
states <- map_data("state")
```

```
future1 <- subset(future, totalMembers > 200, select = c("latitude","longitude","totalMembers", "organization", "university"))
```

```{r static maps}
sm1 <- ggplot() +
  geom_polygon(data=states, aes(long, lat, group = group), color = "white") + coord_fixed(1.3) + guides(fill=FALSE) +
   geom_point(data=eighteen, aes(longitude, latitude, size = profit, color = organization, alpha = 1/20)) +
  ggtitle("2018 Customer Distribution") +
  theme_void() +
  theme(panel.border = element_blank(), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), axis.line = element_blank(),
    legend.position = "bottom", legend.box = "vertical") +
  guides(alpha = "none") + labs(size = "Profit (USD)", color = "Organization")
sm1
sm2 <- ggplot() +
  geom_polygon(data=states, aes(long, lat, group = group), color = "white") + coord_fixed(1.3) + guides(fill=FALSE) +
   geom_point(data=nineteen, aes(longitude, latitude, size = profit, color = organization, alpha = 1/20)) +
  ggtitle("2019 Customer Distribution") +
    theme_void() + theme(panel.border = element_blank(), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), axis.line = element_blank(),
    legend.position = "bottom", legend.box = "vertical") +
  guides(alpha = "none", color = guide_legend(nrow = 5)) + labs(size = "Profit (USD)", color = "Organization")
sm2

sm3 <- ggplot() +
  geom_polygon(data=states, aes(long, lat, group = group), color = "white") + coord_fixed(1.3) + guides(fill=FALSE) +
   geom_point(data=twenty, aes(longitude, latitude, size = profit, color = organization, alpha = 1/20)) +
  ggtitle("2020 Customer Distribution") +
    theme_void() + theme(panel.border = element_blank(), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), axis.line = element_blank(),
    legend.position = "bottom", legend.box = "vertical") +
  guides(alpha = "none") + labs(size = "Profit (USD)", color = "Organization")
sm3

```

```{r interactive maps}
im1 <- ggplot() +
  geom_polygon(data=states,aes(long, lat, group = group), color = "white") + coord_fixed(1.3) + guides(fill=FALSE) +
   geom_point(data=eighteen, aes(longitude, latitude, size = profit, color = organization, alpha = 1/20, group = 1,
               text = paste("Organization: ", organization, "<br>University: ", university, "<br>Profit: $", profit))) +
  ggtitle("2018 Customer Distribution (Hover for info)") +
```

```
    theme_void() + theme(panel.border = element_blank(), panel.grid.major =
element_blank(),
         panel.grid.minor = element_blank(), axis.line = element_blank(),
legend.position = "none")
im1 <- ggplotly(im1, tooltip = "text")
im1 <- im1 %>% layout(legend = list(orientation = 'h', y = -0.05))
im1


im2 <- ggplot() +
   geom_polygon(data=states,aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
    geom_point(data=nineteen, aes(longitude, latitude, size = profit, color =
organization, alpha = 1/20, group = 1,
               text = paste("Organization: ", organization, "<br>University: ",
university, "<br>Profit: $", profit))) +
  ggtitle("2019 Customer Distribution (Hover for info)") +
    theme_void() + theme(panel.border = element_blank(), panel.grid.major =
element_blank(),
         panel.grid.minor = element_blank(), axis.line = element_blank(),
legend.position = "none")
im2 <- ggplotly(im2, tooltip = "text")
im2 <- im2 %>% layout(legend = list(orientation = 'h', y = -0.05))
im2

im3 <- ggplot() +
   geom_polygon(data=states,aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
    geom_point(data=twenty, aes(longitude, latitude, size = profit, color =
organization, alpha = 1/20, group = 1,
               text = paste("Organization: ", organization, "<br>University: ",
university, "<br>Profit: $", profit))) +
  ggtitle("2019 Customer Distribution (Hover for info)") +
    theme_void() + theme(panel.border = element_blank(), panel.grid.major =
element_blank(),
         panel.grid.minor = element_blank(), axis.line = element_blank(),
legend.position = "none")
im3 <- ggplotly(im3, tooltip = "text")
im3 <- im3 %>% layout(legend = list(orientation = 'h', y = -0.05))
im3



```


```{r boxplots}
b1 <- ggplot(eighteen, aes(x=organization, y=profit)) +
  geom_boxplot(fill="gray") +
   labs(title="Profit Distribution by Organization (2018)",x=element_blank(), y
= "Profit (USD)") +
  ylim(0,200) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45))
b1

b2 <- ggplot(nineteen, aes(x=organization, y=profit)) +
```

```
  geom_boxplot(fill="gray") +
   labs(title="Profit Distribution by Organization (2019)",x=element_blank(), y
= "Profit (USD)") +
  ylim(0,200) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45))
b2

b3 <- ggplot(twenty, aes(x=organization, y=profit)) +
  geom_boxplot(fill="gray") +
   labs(title="Profit Distribution by Organization (2020)",x=element_blank(), y
= "Profit (USD)") +
  ylim(0,200) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45))
b3

```

```{r line plots}
lp1 <- ggplot(eighteen, aes(x=date, y=profit)) +
  geom_point(aes(color = organization)) +
       scale_x_date(date_breaks    =    "1    month",    date_labels    =    "%B",
limit=c(as.Date("2018-01-01"),as.Date("2018-12-31"))) +
  ylim(0,550) +
  labs(x=element_blank(), y="Profit (USD)", color="Organization") +
  ggtitle("2018 Temporal Sales Distribution") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45))
lp1

lp2 <- ggplot(nineteen, aes(x=date, y=profit)) +
  geom_point(aes(color = organization)) +
       scale_x_date(date_breaks    =    "1    month",    date_labels    =    "%B",
limit=c(as.Date("2019-01-01"),as.Date("2019-12-31"))) +
  ylim(0,550) +
  labs(x=element_blank(), y="Profit (USD)", color="Organization") +
  ggtitle("2019 Temporal Sales Distribution") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45))
lp2

lp3 <- ggplot(twenty, aes(x=month_day, y=profit)) +
  geom_point(aes(color = organization)) +
       scale_x_date(date_breaks    =    "1    month",    date_labels    =    "%B",
limit=c(as.Date("2020-01-01"),as.Date("2020-12-31"))) +
  ylim(0,550) +
  labs(x=element_blank(), y="Profit (USD)", color="Organization") +
  ggtitle("2020 Temporal Sales Distribution") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45))
lp3

```

```{r profit over time line plots}
lp4 <- ggplot(eighteen, aes(x=date, y=cumsum)) +
```

```
  geom_line() +
  geom_point() +
       scale_x_date(date_breaks    =    "1    month",    date_labels    =    "%B",
limit=c(as.Date("2018-01-01"),as.Date("2018-12-31"))) +
  ylim(0,15000) +
  labs(x=element_blank(), y="Cumulative Profit (USD)") +
  ggtitle("2018 Cumulative Profit") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45))
lp4

lp5 <- ggplot(nineteen, aes(x=date, y=cumsum)) +
  geom_line() +
  geom_point() +
       scale_x_date(date_breaks    =    "1    month",    date_labels    =    "%B",
limit=c(as.Date("2019-01-01"),as.Date("2019-12-31"))) +
  ylim(0,15000) +
  labs(x=element_blank(), y="Cumulative Profit (USD)") +
  ggtitle("2019 Cumulative Profit") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45))
lp5

lp6 <- ggplot(twenty, aes(x=month_day, y=cumsum)) +
  geom_line() +
  geom_point() +
       scale_x_date(date_breaks    =    "1    month",    date_labels    =    "%B",
limit=c(as.Date("2020-01-01"),as.Date("2020-12-31"))) +
  ylim(0,15000) +
  labs(x=element_blank(), y="Cumulative Profit (USD)") +
  ggtitle("2020 Cumulative Profit") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = -45))
lp6
```
```{r static future bubble maps}
sm4 <- ggplot() +
   geom_polygon(data=states, aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
  geom_point(data=future, aes(longitude, latitude, size = totalMembers, color =
organization, alpha = 1/20)) +
  ggtitle("Potential Customer Distribution") +
  theme_void() +
  theme(panel.border = element_blank(), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), axis.line = element_blank(),
    legend.position = "bottom", legend.box = "vertical") +
  guides(alpha = "none") + labs(size = "# Members", color = element_blank())
sm4

sm5 <- ggplot() +
   geom_polygon(data=states, aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
   geom_point(data=future1, aes(longitude, latitude, size = totalMembers, color
= organization, alpha = 1/20)) +
  ggtitle("Organizations w/ 200+ Members") +
  theme_void() +
  theme(panel.border = element_blank(), panel.grid.major = element_blank(),
```

```
    panel.grid.minor = element_blank(), axis.line = element_blank(),
    legend.position = "bottom", legend.box = "vertical") +
  guides(alpha = "none") + labs(size = "# Members", color = element_blank())
sm5

sm6 <- ggplot() +
   geom_polygon(data=states, aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
  geom_point(data=future, aes(longitude, latitude, size = totalMembers, color =
paddles.y.n., alpha = 1/20)) +
  ggtitle("Oganizations w/ A History of Paddle Gifting") +
  theme_void() +
  theme(panel.border = element_blank(), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), axis.line = element_blank(),
    legend.position = "bottom", legend.box = "vertical") +
  guides(alpha = "none") + labs(size = "# Members", color = element_blank())
sm6

sm7 <- ggplot() +
   geom_polygon(data=states, aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
  geom_point(data=future, aes(longitude, latitude, size = totalMembers, color =
affinityClient, alpha = 1/20)) +
  ggtitle("Oganizations Protected by a Licensing Company") +
  theme_void() +
  theme(panel.border = element_blank(), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), axis.line = element_blank(),
    legend.position = "bottom", legend.box = "vertical") +
  guides(alpha = "none") + labs(size = "# Members", color = element_blank())
sm7

sm8 <- ggplot() +
   geom_polygon(data=states, aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
  geom_point(data=future, aes(longitude, latitude, size = totalMembers, color =
type, alpha = 1/20)) +
  ggtitle("Oganizations by Type") +
  theme_void() +
  theme(panel.border = element_blank(), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), axis.line = element_blank(),
    legend.position = "bottom", legend.box = "vertical") +
  guides(alpha = "none") + labs(size = "# Members", color = "Type")
sm8

```


```{r interactive future bubble maps}
im4 <- ggplot() +
   geom_polygon(data=states,aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
  geom_point(data=future, aes(longitude, latitude, size = totalMembers, color =
organization, alpha = 1/20, group = 1,
               text = paste("Organization: ", organization, "<br>University: ",
university, "<br>Total Members: ", totalMembers))) +
  ggtitle("Potential Customer Distribution (Hover for info)") +
```

```
    theme_void() + theme(panel.border = element_blank(), panel.grid.major =
element_blank(),
         panel.grid.minor = element_blank(), axis.line = element_blank(),
legend.position = "none")
im4 <- ggplotly(im4, tooltip = "text")
im4

im5 <- ggplot() +
   geom_polygon(data=states,aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
  geom_point(data=future1, aes(longitude, latitude, size = totalMembers, color
= organization, alpha = 1/20, group = 1,
               text = paste("Organization: ", organization, "<br>University: ",
university, "<br>Total Members: ", totalMembers))) +
  ggtitle("Organizations w/ 200+ Members (Hover for info)") +
    theme_void() + theme(panel.border = element_blank(), panel.grid.major =
element_blank(),
         panel.grid.minor = element_blank(), axis.line = element_blank(),
legend.position = "none")
im5 <- ggplotly(im5, tooltip = "text")
im5

im6 <- ggplot() +
   geom_polygon(data=states,aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
  geom_point(data=future, aes(longitude, latitude, size = totalMembers, color =
paddles.y.n., alpha = 1/20, group = 1,
               text = paste("Organization: ", organization, "<br>University: ",
university, "<br>Total Members: ", totalMembers, "<br> Gifts Paddles? ",
paddles.y.n.))) +
  ggtitle("Oganizations w/ A History of Paddle Gifting (Hover for info)") +
    theme_void() + theme(panel.border = element_blank(), panel.grid.major =
element_blank(),
         panel.grid.minor = element_blank(), axis.line = element_blank(),
legend.position = "none")
im6 <- ggplotly(im6, tooltip = "text")
im6

im7 <- ggplot() +
   geom_polygon(data=states,aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
  geom_point(data=future, aes(longitude, latitude, size = totalMembers, color =
affinityClient, alpha = 1/20, group = 1,
               text = paste("Organization: ", organization, "<br>University: ",
university, "<br>Total Members: ", totalMembers, "<br> Protected? ",
affinityClient))) +
  ggtitle("Oganizations Protected by a Licensing Company (Hover for info)") +
    theme_void() + theme(panel.border = element_blank(), panel.grid.major =
element_blank(),
         panel.grid.minor = element_blank(), axis.line = element_blank(),
legend.position = "none")
im7 <- ggplotly(im7, tooltip = "text")
im7

im8 <- ggplot() +
   geom_polygon(data=states,aes(long, lat, group = group), color = "white") +
coord_fixed(1.3) + guides(fill=FALSE) +
```

```
  geom_point(data=future, aes(longitude, latitude, size = totalMembers, color = type, alpha = 1/20, group = 1,
                text = paste("Organization: ", organization, "<br>University: ", university, "<br>Total Members: ", totalMembers, "<br> Type: ", type))) +
  ggtitle("Oganizations by Type (Hover for info)") +
    theme_void() + theme(panel.border = element_blank(), panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(), axis.line = element_blank(), legend.position = "none")
im8 <- ggplotly(im8, tooltip = "text")
im8

```
```{r proportion inferential analysis}

future2 <- subset(future, totalMembers > 100, select = c("latitude","longitude","totalMembers", "organization", "university", "affinityClient"))

levels(future2$affinityClient)

liscenced_yes_no <- future2$affinityClient == "y"

table(liscenced_yes_no)

p_hat <- mean(liscenced_yes_no)
p_hat

std_err <- sqrt(p_hat*(1-p_hat)/length(liscenced_yes_no))
z_star <- qnorm(.975)
z_star
lb <- p_hat - z_star*std_err
ub <- p_hat + z_star*std_err
c(lb,ub)
```

```{r}

```



```{r future analysis}
#Distances Dot Plot
sp1 <- ggplot(distances, aes(x=organization.y, y=dist)) +
  geom_point() +
  labs(x=element_blank(), y="Distance (mi)") +
  ggtitle("Distances from College Station") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90))
sp1

#Licensing Pie Chart
blank_theme <- theme_minimal()+
  theme(
  axis.title.x = element_blank(),
  axis.title.y = element_blank(),
```

```
  panel.border = element_blank(),
  panel.grid=element_blank(),
  axis.ticks = element_blank(),
  plot.title=element_text(size=14, face="bold")
  )

ggplot(future, aes(x=factor(1), fill=affinityClient)) +
  geom_bar(width = 1) + blank_theme +
  theme(axis.text.x=element_blank()) +
  coord_polar("y") +
  labs(fill = "Trademark Protected?")

# Paddle Gifting
ggplot(future, aes(x=factor(1), fill=paddles.y.n.)) +
  geom_bar(width = 1) + blank_theme +
  theme(axis.text.x=element_blank()) +
  coord_polar("y") +
  labs(fill = "Paddle Gifting?")

# State
ggplot(future, aes(x=factor(1), fill=state)) +
  geom_bar(width = 1) + blank_theme +
  theme(axis.text.x=element_blank()) +
  coord_polar("y") +
  labs(fill = "State")

# Type
ggplot(future, aes(x=factor(1), fill=type)) +
  geom_bar(width = 1) + blank_theme +
  theme(axis.text.x=element_blank()) +
  coord_polar("y") +
  labs(fill = "Type")

```