Starbucks' Growth and Spread

Dylan Vener

Introduction

Starbucks, which opened it's first store in 1971 in Seattle, has, in the last 20 years, become a global phenomenom, and a cultural touchstone. After an initial slow growth in the 1980s, Starbucks started spreading across the nation, and now has over 23 thousand locations worldwide.

Winter, a blogger, has made a personal mission of visiting every company owned Starbucks, and while he is nowhere near finished, he has gathered data from the approximately 3000 stores hes has visited, including their date of opening.

Using this data, we can visualize just how frantic their growth has been.

Methodology

To start my analysis, I first scraped Winter's website to gather the data.

```
#Scrape the website and put it into a dataframe
url <- 'http://www.starbuckseverywhere.net/StoreOpeningDates.htm'
temp.pg <- read_html(url)
star.df <- as_tibble((temp.pg %>% html_nodes("table") %>% html_table(header=TRUE))[[1]])
```

Next, I had to do some minor cleaning to more easily integrate it.

```
#Grab the first section of the market to get the State
star.df$Market <- gsub("\\\\.*","",as.matrix(star.df$Market))

#Convert the state name into the state abbreviation
star.df$state <- state.abb[match(star.df$Market,state.name)]

#Parse the dates into a workable format
star.df$Opened <- parse_date(star.df$Opened,'%m/%d/%Y')
```

Next, I made use of the geocode function, which maps a city and state to a longitude and latitude.

```
#Using the geocode function, find a longitude and latitude for each starbucks, to the city
star.latlong.df <- as_tibble(mapply(geocode,star.df$City,star.df$state))
#Join this data to the dataframe
star.df$lat <- as.numeric(unname(sapply(t(star.latlong.df)[,1],'[',1)))
star.df$long <-  as.numeric(unname(sapply(t(star.latlong.df)[,2],'[',1)))
#Keep only cases where we have all the required information
star.df <- star.df[complete.cases(star.df),]
```

Once we have this information, we're ready to begin creating the map.

```
#Grabbing a map of the Continental United States
US.map <- get_map("United States",zoom = 4)

#Function to draw the map of the US, with the location of each starbucks up to and including that date
#Stores opened up to 7 days before the date are twice as large, and non-transparent
draw <- function(date){
```

```
  g <- ggmap(US.map) +
    geom_point(data=star.df[star.df$Opened<date-7,],aes(x=lat,y=long),alpha=0.5,size=1) +
    geom_point(data=star.df[((star.df$Opened>=date-7) & (star.df$Opened<=date)),],aes(x=lat,y=long),siz
    labs(x="Latitude",y="Longitude",title=format(date,format='%B %d, %Y'))
  print(g)
}
```

Then, to create the animation, we just call that function for a range of dates, for clarity, I started from 1991, the second old date in our dataset, and went up till the newest, taking a snapshot at every week.

```
#Draw the above function for every week from the second oldest date, 1991, until the most recent, 2016
draw.animate <- function(){
  lapply(seq(head(tail(star.df,2),1)$Opened,max(star.df$Opened),by='week'),draw)
}
```
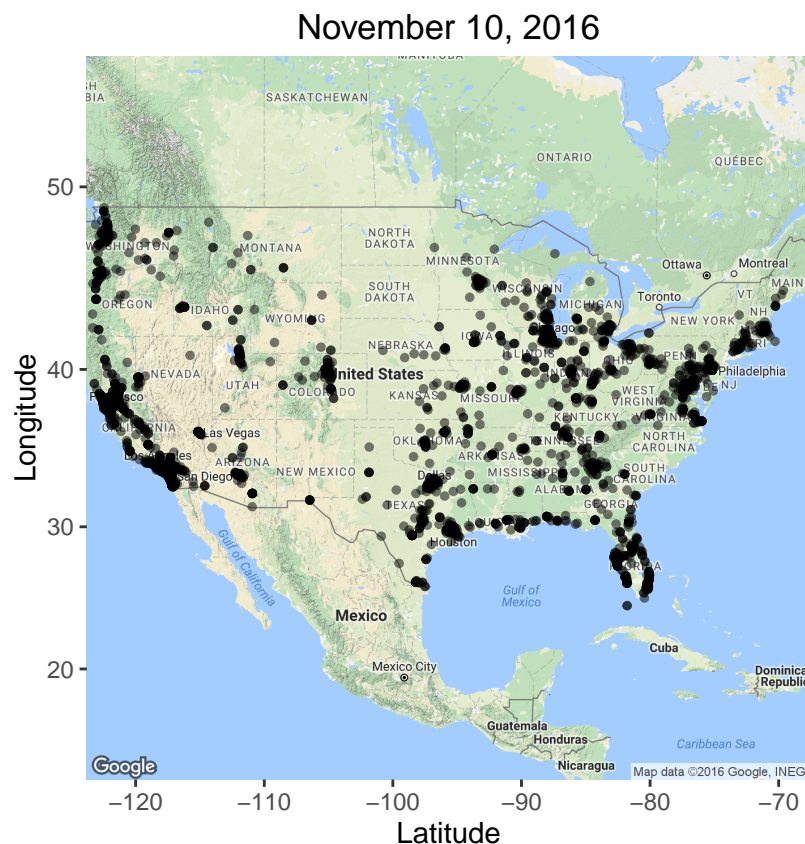
To finish the process, we save the realization of this function into a gif.

```
#Save the above into a GIF
saveGIF(draw.animate(), interval = .05, movie.name = 'Starbuck_big.gif')
```

Conclusion

The above map, which shows the status of the dataset up until November 10th, 2016, shows Starbucks' massive penetration and control of the US market.

The included gif shows just how incredibly swift this transformation was, opening several stores each week, for two decades has made Starbucks one of the most incredible success stories in corporate America, and in the world.

As the gif shows, Starbucks tends to grow target large cities, and then expand outward in that area. Importantly, they have no regard for density, placing many stores in a city, and Anecdotally, without regard for competition with themselves, as I personally have seen three in one intersection.

They're also verymuch not limited to only large cities, covering large swathes of several states. Their strategy appears to be quantity over any other metric, and from their current state, it appears to have succeeded.

The biggest shortcoming of this analysis is a limit on information, Starbucks does not publish any information about their store's opening dates, and so we must rely on Winter's dataset, which while well collected, is only about a 1/7th of the total number.

Code

```r
#To run on a different computer, the only thing required is the following packages, as well as ImageMag
library('readr')
library('rvest')
library('jsonlite')
library('tibble')
library('dplyr')
library('ggplot2')
library('tidyr')
library('lubridate')
library('htmltab')
library('XML')
library('ggmap')
library('localgeo')
library('animation')

#Scrape the website and put it into a dataframe
url <- 'http://www.starbuckseverywhere.net/StoreOpeningDates.htm'
temp.pg <- read_html(url)
star.df <- as_tibble((temp.pg %>% html_nodes("table") %>% html_table(header=TRUE))[[1]])

#Grab the first section of the market to get the State
star.df$Market <- gsub("\\\\.*","",as.matrix(star.df$Market))

#Convert the state name into the state abbreviation
star.df$state <- state.abb[match(star.df$Market,state.name)]

#Parse the dates into a workable format
star.df$Opened <- parse_date(star.df$Opened,'%m/%d/%Y')

#Using the geocode function, find a longitude and latitude for each starbucks, to the city
star.latlong.df <- as_tibble(mapply(geocode,star.df$City,star.df$state))
#Join this data to the dataframe
star.df$lat <- as.numeric(unname(sapply(t(star.latlong.df)[,1],'[',1)))
star.df$long <-  as.numeric(unname(sapply(t(star.latlong.df)[,2],'[',1)))

#Keep only cases where we have all the required information
star.df <- star.df[complete.cases(star.df),]

#Grab a map of the Continental United States
US.map <- get_map("United States",zoom = 4)
```

```r
#Function to draw the map of the US, with the location of each starbucks up to and including that date
#Stores opened up to 7 days before the date are twice as large, and non-transparent
draw <- function(date){
  g <- ggmap(US.map) +
    geom_point(data=star.df[star.df$Opened<date-7,],aes(x=lat,y=long),alpha=0.5,size=1) +
    geom_point(data=star.df[((star.df$Opened>=date-7) & (star.df$Opened<=date)),],aes(x=lat,y=long),size
    labs(x="Latitude",y="Longitude",title=format(date,format='%B %d, %Y'))
  print(g)
}

#Draw the above function for every week from the second oldest date, 1991, until the most recent, 2016
draw.animate <- function(){
  lapply(seq(head(tail(star.df,2),1)$Opened,max(star.df$Opened),by='week'),draw)
}

#Save the above into a GIF
saveGIF(draw.animate(), interval = .05, movie.name = 'Starbuck_big.gif')
```