StarbucksHistoryInvestigation

Ruinan(Victor) Zhang

11/9/2016

# Background

This task is to investigate growth of chain companies such as Starbucks. The approach is to construct a dynamic graphic to illustrate Starbuck stores' expension. The date is provided by Winter who setted his mission drink coffeefrom every company-owned (including partner companies) Starbucks in the world. The openning dates of the starbucks are provided on [his website](http://www.starbuckseverywhere.net/StoreOpeningDates.htm):

# Approach

To download the data from Winter's website, I used the web scrapping tools from rvest package to download and clean the informations about every starbucks stores he visited by 11/03/2016.

# read hdml  
url = "http://www.starbuckseverywhere.net/StoreOpeningDates.htm"  
html <- html(url)  
  
# retrieve data from Winter's website and save it  
df <- html %>% html\_nodes("table") %>% html\_table()  
save(df,file = "starbucks\_data.Rda")

Based on the address of each store, I used geocode() function to request the longtitude and latitude of every store.

geo\_locations <- data.frame() # df for storing all the geo\_locations of starbucks   
for(i in 1:nrow(df)){  
 row <- df[i,]  
 location\_str <- paste(row$Market,row$City,row$Name) # location str  
 geo<-geocode(location\_str,source = "dsk") # call geo\_code() to get locations  
 geo\_locations<- rbind(geo\_locations,geo)  
}  
geo\_locations

After achieving the locations in longitude and latitude, I sorted all the stores by there openning date and construct a dynamic graphic using the "animation" package and stored the output as a gif file.

# sort df by year, month, and data  
date\_info <- as.data.frame(matrix(date\_info,ncol=3,byrow = T))  
names(date\_info) <- c("month",'day','year')  
df <- cbind(df,date\_info)  
df <- (df[order(df$year,df$month,df$day),])  
  
# drop cases with NA for locations  
geo\_locations = geo\_locations[complete.cases(geo\_locations), ]  
  
# get the base satellite map of US  
mapgilbert <- get\_map(location = c(lon = -98.5795, lat = 39.8282), zoom = 4,  
 maptype = "satellite", scale = 2)  
g<-ggmap(mapgilbert)+ggtitle("Starbucks in US")+guides(fill=FALSE)+xlab("longitude")+ylab("latitude")  
g  
  
# plot function for the animation  
plotfoo<- function(){  
 i=1  
 while(i<nrow(geo\_locations)){  
 # the increment is dynamic  
 increment = ceiling(i/2)  
 if(increment>50){  
 increment=50  
 }  
 bound = i+increment  
 if(bound>nrow(geo\_locations)){  
 bound = nrow(geo\_locations)  
 }  
 each\_location <- geo\_locations[i:bound,] # for each iteraton, plot i to i+increment rows  
 print(i)  
 i=i+increment  
 g <- g+geom\_point(data=each\_location,aes(x = lon, y = lat, fill = "red"), size = 2, shape = 21,alpha=0.8)  
 g <- g+ggtitle(paste("Starbucks in US by",geo\_locations[i,]$year))  
 print(g)  
 }  
}  
  
saveGIF(plotfoo(),interval = 0.1,movie.name = "starbucks.gif")

# Result

Based on the animation, it is clear that majority of the starbucks are opened in large cities. Starbucks is an organization initiated from Seattle in Washington state and quickly expend to both east and west coasts. The growth of starbucks really bursted from 2006 to 2008. However, for state like Wisconsin and Kansas, it seems like Starbucks have very fiew stores in those areas. There are definitely many stores in most area of east cost and midwest and major cities on west coast.

The dynamic graph is attached as a seperate gif file.

# Full Code

library(ggmap)  
library(ggplot2)  
library(rvest)  
library(dplyr)  
library(maps)  
library(animation)  
library(stringr)  
  
url = "http://www.starbuckseverywhere.net/StoreOpeningDates.htm"  
  
html <- html(url)  
  
# retrieve data from Winter's website  
df <- html %>% html\_nodes("table") %>% html\_table()  
save(df,file = "starbucks\_data.Rda")  
 load("starbucks\_data.Rda")  
  
 df <- df[[1]]  
 colnames(df) <- df[1,]  
 df <- df[2:nrow(df),]  
 # clean backslahes in teh Market column  
 df$Market <-str\_replace(df$Market,"\\\\"," ")  
  
  
row = df[1,]  
# clean out the date infos  
date\_info<- str\_split(df$Opened,"/")  
date\_info <- unlist(date\_info)  
date\_info <- as.numeric(date\_info)  
  
# sort df by year, month, and data  
date\_info <- as.data.frame(matrix(date\_info,ncol=3,byrow = T))  
names(date\_info) <- c("month",'day','year')  
df <- cbind(df,date\_info)  
df <- (df[order(df$year,df$month,df$day),])  
  
  
geo\_locations <- data.frame() # df for storing all the geo\_locations of starbucks   
for(i in 1:nrow(df)){  
 row <- df[i,]  
 location\_str <- paste(row$Market,row$City,row$Name)  
 geo<-geocode(location\_str,source = "dsk")  
 geo\_locations<- rbind(geo\_locations,geo)  
}  
geo\_locations <- cbind(df,geo\_locations)  
  
# save the geo\_locations  
save(geo\_locations,file = "geo\_locations.Rda")  
load("geo\_locations.Rda")  
  
# drop cases with NA for locations  
geo\_locations = geo\_locations[complete.cases(geo\_locations), ]  
  
# get the base satellite map of US  
mapgilbert <- get\_map(location = c(lon = -98.5795, lat = 39.8282), zoom = 4,  
 maptype = "satellite", scale = 2)  
g<-ggmap(mapgilbert)+ggtitle("Starbucks in US")+guides(fill=FALSE)+xlab("longitude")+ylab("latitude")  
g  
  
# plot function for the animation  
plotfoo<- function(){  
 i=1  
 while(i<nrow(geo\_locations)){  
 # the increment is dynamic  
 increment = ceiling(i/2)  
 if(increment>50){  
 increment=50  
 }  
 bound = i+increment  
 if(bound>nrow(geo\_locations)){  
 bound = nrow(geo\_locations)  
 }  
 each\_location <- geo\_locations[i:bound,] # for each iteraton, plot i to i+increment rows  
 print(i)  
 i=i+increment  
 g <- g+geom\_point(data=each\_location,aes(x = lon, y = lat, fill = "red"), size = 2, shape = 21,alpha=0.8)  
 g <- g+ggtitle(paste("Starbucks in US by",geo\_locations[i,]$year))  
 print(g)  
 }  
}  
  
saveGIF(plotfoo(),interval = 0.1,movie.name = "starbucks.gif")