

# Twitter Project on Trader Joe's

*Yitong Zhang*

*December 16, 2017*

```
#Set up
api_key<- "DCnaooJkXNICrHNagogBv0kqr"
api_secret<- "lgUAQkxj6N7KgJ7e0GnZpOAJvNOMDBODiJp3uvAndter9jq9By"
access_token<- "927641256196526081-q3uJd67zY1Jc5YBIGRrIyroLl4kWQer"
access_token_secret<- "pPkNuBVInDZqBgLF00YK9gflL20T2Shv76qqrlnSgc0HK5"

setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)

## [1] "Using direct authentication"
```

## Twitter Locations in United States

From the US map we can see that most of the twitters are from California and east part of US.

```
yo <- searchTwitter("trader joe's", n = 5000,lang="en", geocode = '37.09,-95.71,1000mi')
yodf <- twListToDF(yo)
loc <- -1*is.na(yodf$longitude) + 1
sum(loc)
```

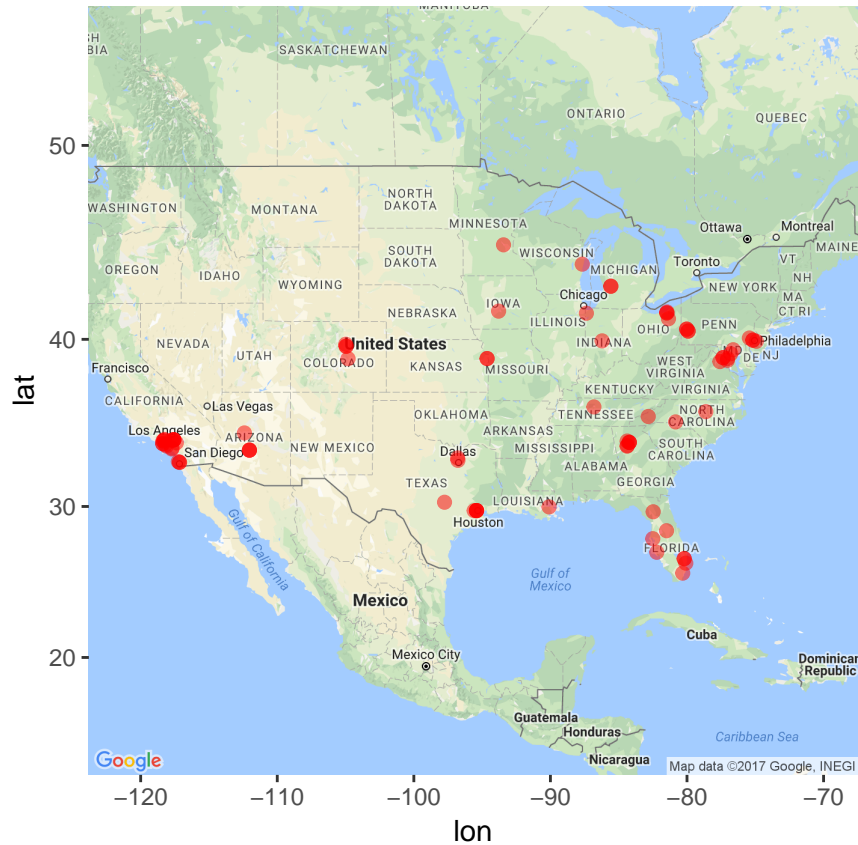
```
## [1] 82
```

```
loc1 <- which(loc==1)
locations <- data.frame(yodf$latitude[loc1], yodf$longitude[loc1])
locations$yodf.latitude.loc1<-as.numeric(as.character( locations$yodf.latitude.loc1))
locations$yodf.longitude.loc1<-as.numeric(as.character(locations$yodf.longitude.loc1))
names(locations)<-c("lat","lon")
write.csv(locations,"twitter_us.csv")
```

```
map <- ggmap(get_map(location = "United States", zoom = 4,
source = "google", maptype = "roadmap")) +
  geom_point(aes(x=lon,
y = lat), data = locations, alpha = 0.5, size = 2,
color = "red")
```

```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=United+States&zoom=4&size=640x640
```

```
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=United%20States&sens
map
```



## Twitter Locations in California

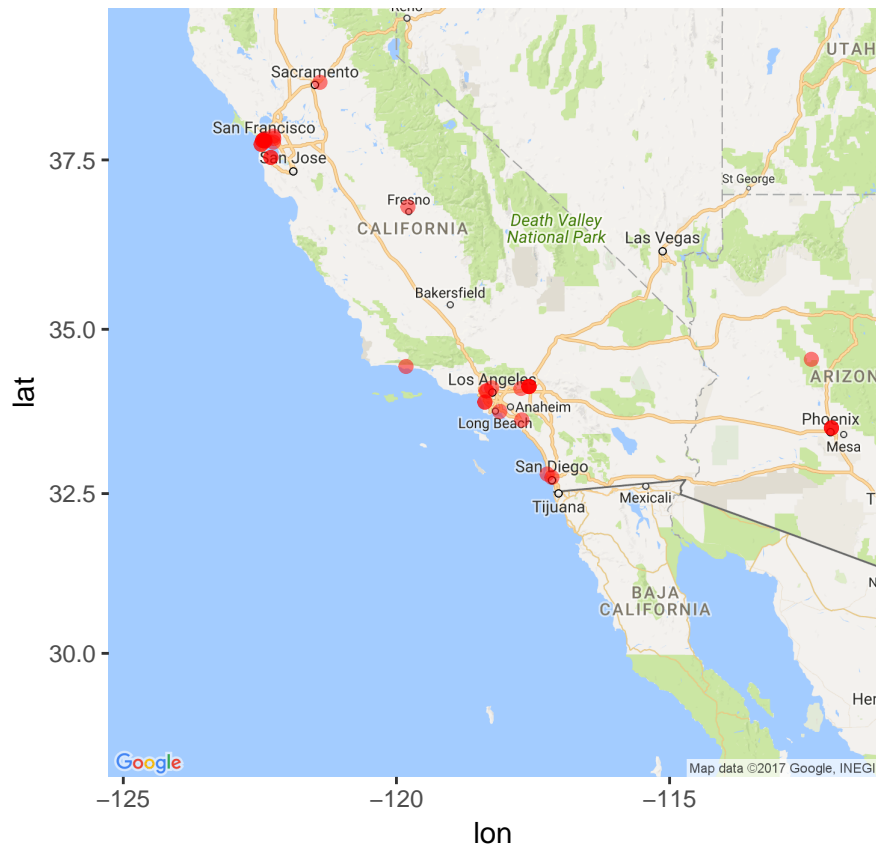
From the map of California we can see that the tweets are concentrated in Los Angeles and San Francisco.

```
## [1] "Rate limited .... blocking for a minute and retrying up to 119 times ..."
```

```
## [1] 95
```

```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=Los+Angelas&zoom=6&size=640x640&
```

```
## Information from URL : http://maps.googleapis.com/maps/api/geocode/json?address=Los%20Angelas&sensor=
```



## Wordcloud

```

tweets <- searchTwitter("trader joe's", n=5000, lang="en")
tweets.text <- sapply(tweets, function(x) x$getText())

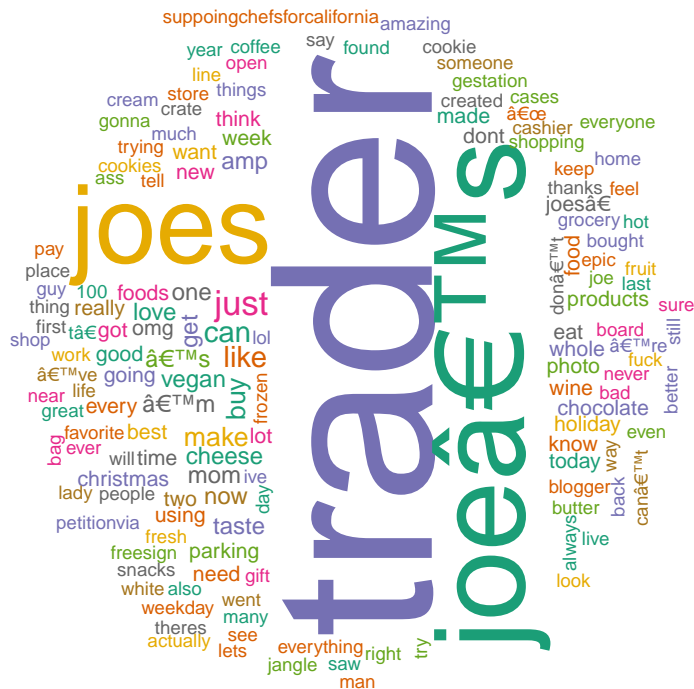
tweets.text<- str_replace_all(tweets.text,"[:graph:]", " ")

#tweets.text<-tm_map(tweets.test, function(x) iconv(enc2utf8(x), sub = "byte"))

tweets.text <- tolower(tweets.text)
tweets.text <- gsub("rt", "", tweets.text)
tweets.text <- gsub("@\\w+", "", tweets.text)
tweets.text <- gsub("[:punct:]", "", tweets.text)
tweets.text <- gsub("http\\w+", "", tweets.text)
tweets.text <- gsub("[ \\t]{2,}", "", tweets.text)
tweets.text <- gsub("^ ", "", tweets.text)
tweets.text <- gsub(" $", "", tweets.text)

tweets.text.corpus <- Corpus(VectorSource(tweets.text))
tweets.text.corpus <- tm_map(tweets.text.corpus, function(x)removeWords(x,stopwords()))
wordcloud(tweets.text.corpus,min.freq = 2, scale=c(7,0.5),colors=brewer.pal(8, "Dark2"),
          random.color= TRUE, random.order = FALSE, max.words = 150)

```



## Word Frequency Table and Plot

From the word frequency table and the plot we can see that the most frequent words are related to the name of Trader Joe's. Besides, the popular products in Trader Joe's includes wine, cheese and chocolate.

```
docs<- Corpus(VectorSource(tweets.text))
docs <- tm_map(docs, function(x)removeWords(x,stopwords()))

dtm <- TermDocumentMatrix(docs)

mat <- as.matrix(dtm)

v <- sort(rowSums(mat), decreasing = TRUE)

d<- data.frame(word= names(v), freq=v)
d50<-d[1:50,]
head(d,50)
```

```
##           word freq
## trader      trader 4575
## joeâ€œs      joeâ€œs 2383
## joes        joes 2030
## just        just 391
## can         can 272
## like        like 264
## buy         buy 257
```

## make	make	212	
## one	one	200	
## â<U+0080><U+0099>s			â<U+0080><U+0099>s 197
## get	get	182	
## vegan	vegan	173	
## cheese	cheese	172	
## â<U+0080><U+0099>m			â<U+0080><U+0099>m 169
## mom	mom	168	
## amp	amp	161	
## love	love	157	
## now	now	152	
## taste	taste	148	
## omg	omg	145	
## going	going	143	
## good	good	141	
## wine	wine	136	
## photo	photo	133	
## best	best	132	
## whole	whole	127	
## time	time	124	
## foods	foods	123	
## parking	parking	122	
## eat	eat	122	
## holiday	holiday	111	
## today	today	110	
## products	products	110	
## food	food	109	
## dont	dont	108	
## know	know	107	
## week	week	106	
## new	new	103	
## really	really	102	
## christmas	christmas	97	
## chocolate	chocolate	97	
## got	got	95	
## lot	lot	94	
## want	want	94	
## every	every	92	
## two	two	92	
## using	using	92	
## joesâ<U+0080>	joesâ<U+0080>	90	
## made	made	89	
## think	think	88	

```
plot<-ggplot(data=d50)+geom_col(aes(x=word,y=freq))+ggtitle("Word Frequency")
plot + theme(axis.text.x = element_text(angle = 60, hjust = 1))
```

Word Frequency

