

Final Project

Jiawei Li

2017/12/18

Text analysis

Topic #Rihanna

Topic #Taylor Swift

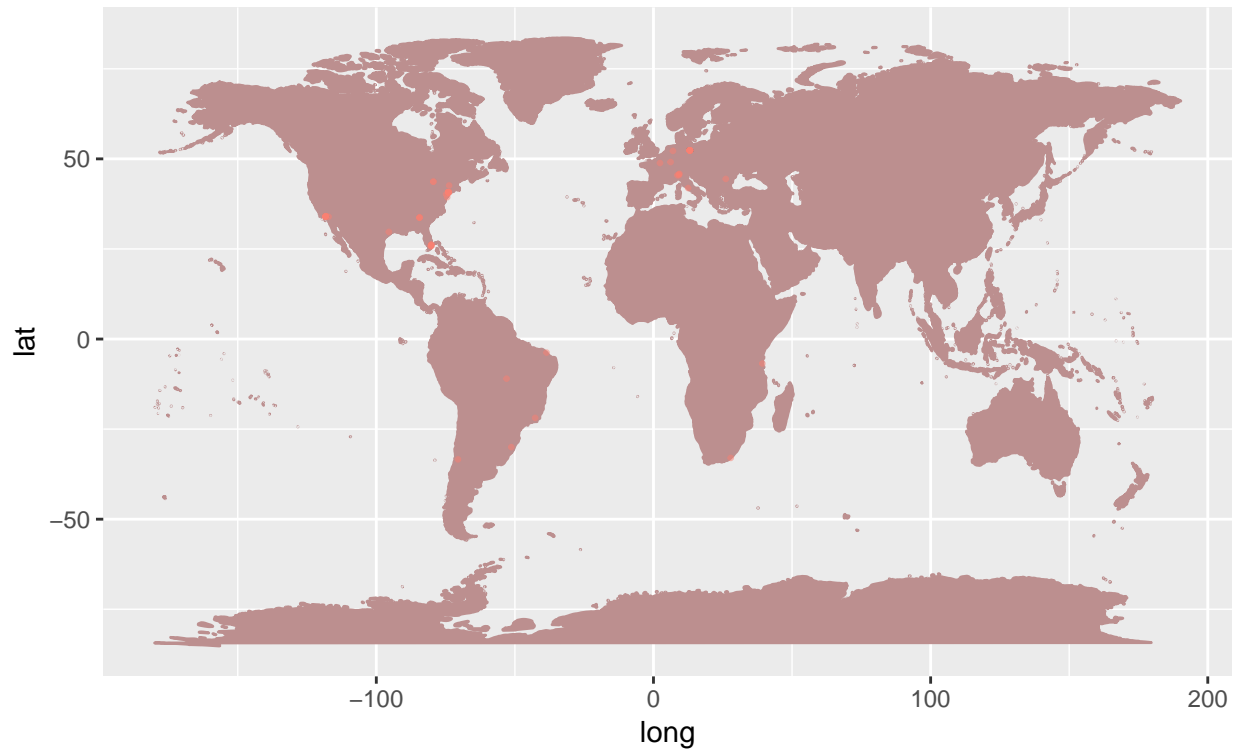
Search Topic

First search 5000 results of the topic #Rihanna from Twitter.

It looks like most of the data points I collected are in the area of Europe and United States.

```
## [1] 85
```

The Map of the Data Points–Rihanna

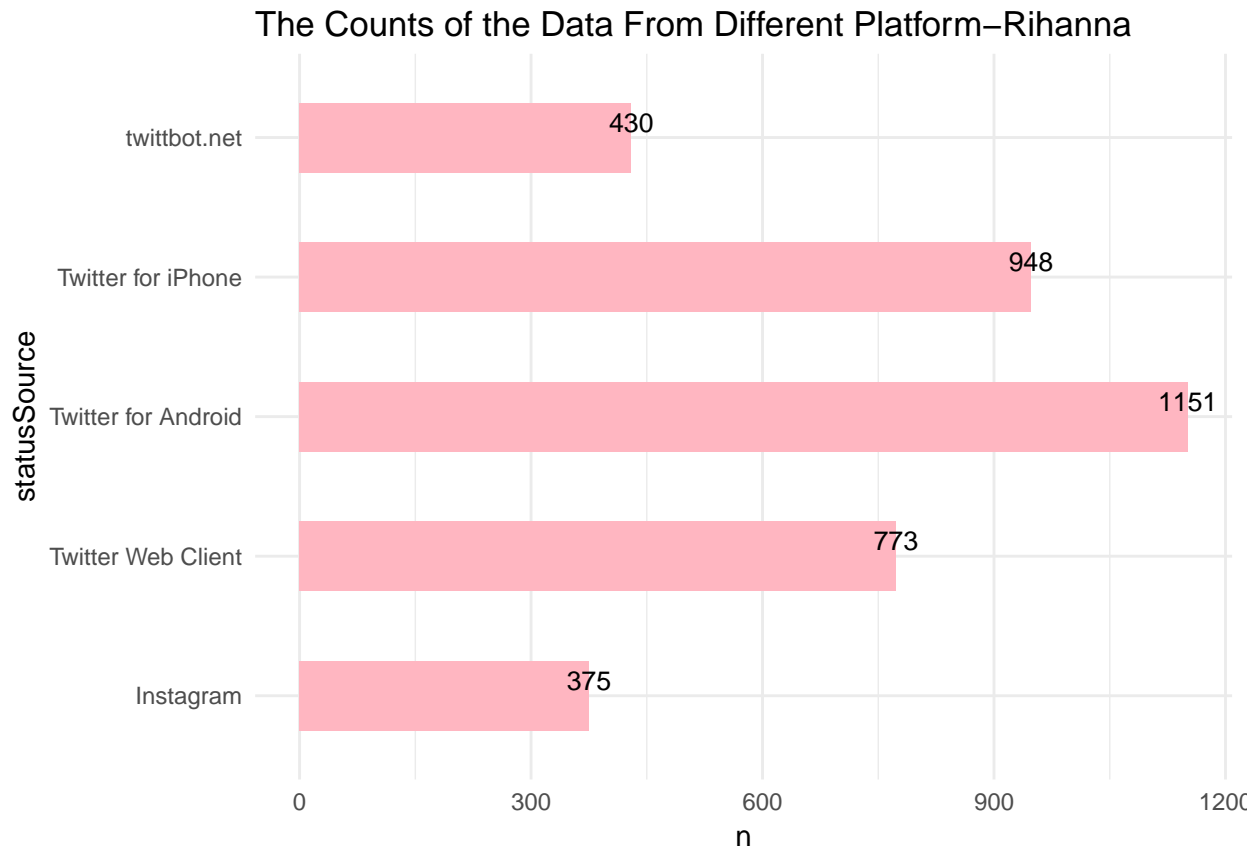


Calculate the count of tweets by platform.

Here is the barplot of the count of the source. Most of the tweets are from Android, and then is Iphone.

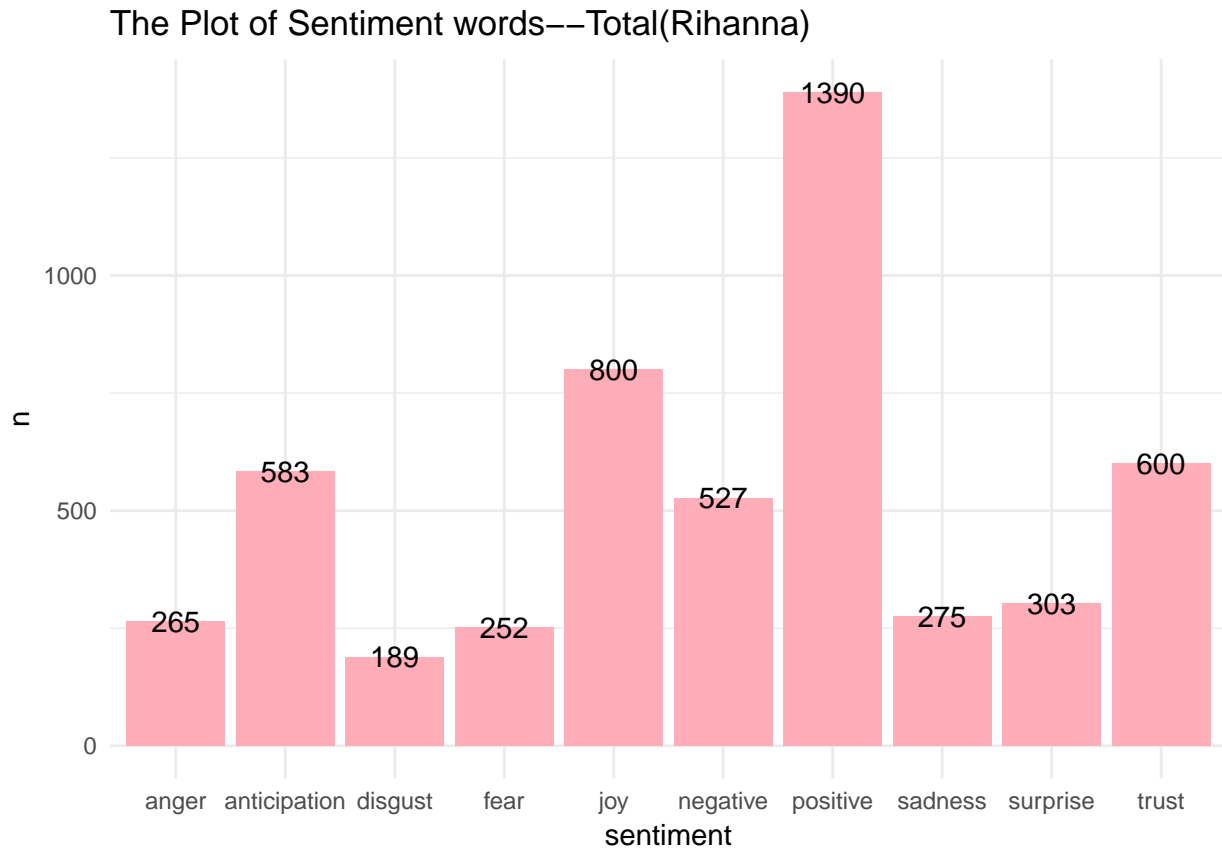
statusSource	n	percent_of_tweets
Twitter for Android	1151	0.230

statusSource	n	percent_of_tweets
Twitter for iPhone	948	0.190
Twitter Web Client	773	0.155
twittbot.net	430	0.086
Instagram	375	0.075



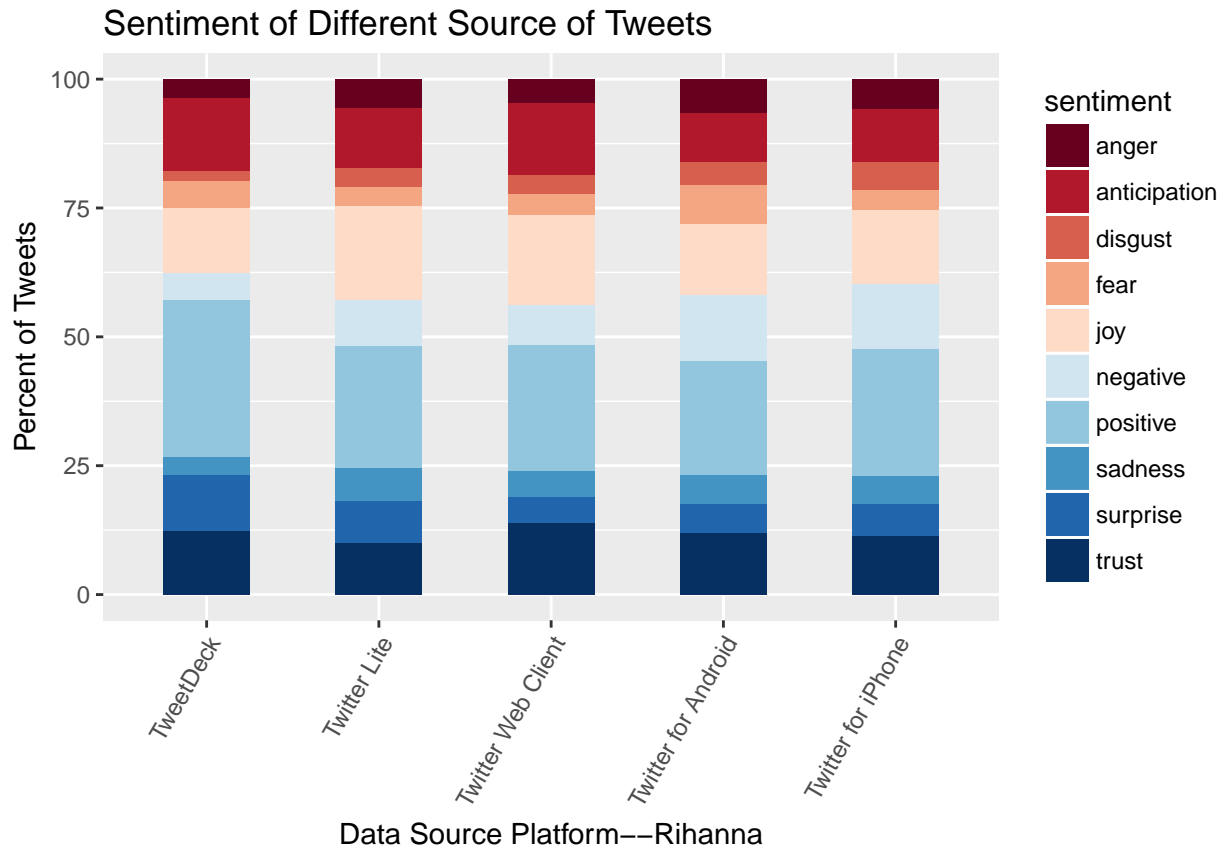
Sentiment of total data

After clean the data, plot the Sentiment of Topic #Rihanna in total. We can see that most of the words are positive in the topic of #Rihanna. The least is disgust.



Sentiment of top 5 data source

Calculate and plot the Sentiment of Topic #Rihanna from different data source platform. By comparing the sentiment from different data platforms, we can see that positive and joy words have biggest part in the bar.



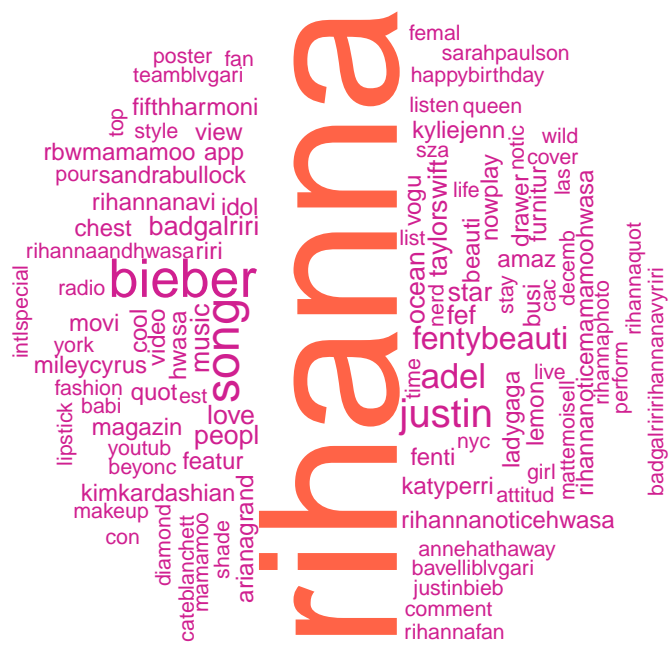
Wordclouds Analysis--Top 10 words in the Topic(#Rihanna)

Plot the word clouds of #Rihanna. Top 1 word is rihanna, which is obviously. But it is suprise that the secon word is bieber. We can also see that fentybeauty—the cosmetics brand establised by Rihanna herself is also in top 10.

```
## chr [1:39621] "videos" "100m" "views" "#rihanna" "#taylorswift" ...
```

Word	Frequency (approx.)	factor(word)
rihanna	3200	rihanna
bieber	700	bieber
song	600	song
justin	450	justin
adel	350	adel
fentybeauti	250	fentybeauti
ocean	200	ocean
fef	150	fef
star	150	star
love	150	love

Here is the word clouds of the topic #Rihanna.

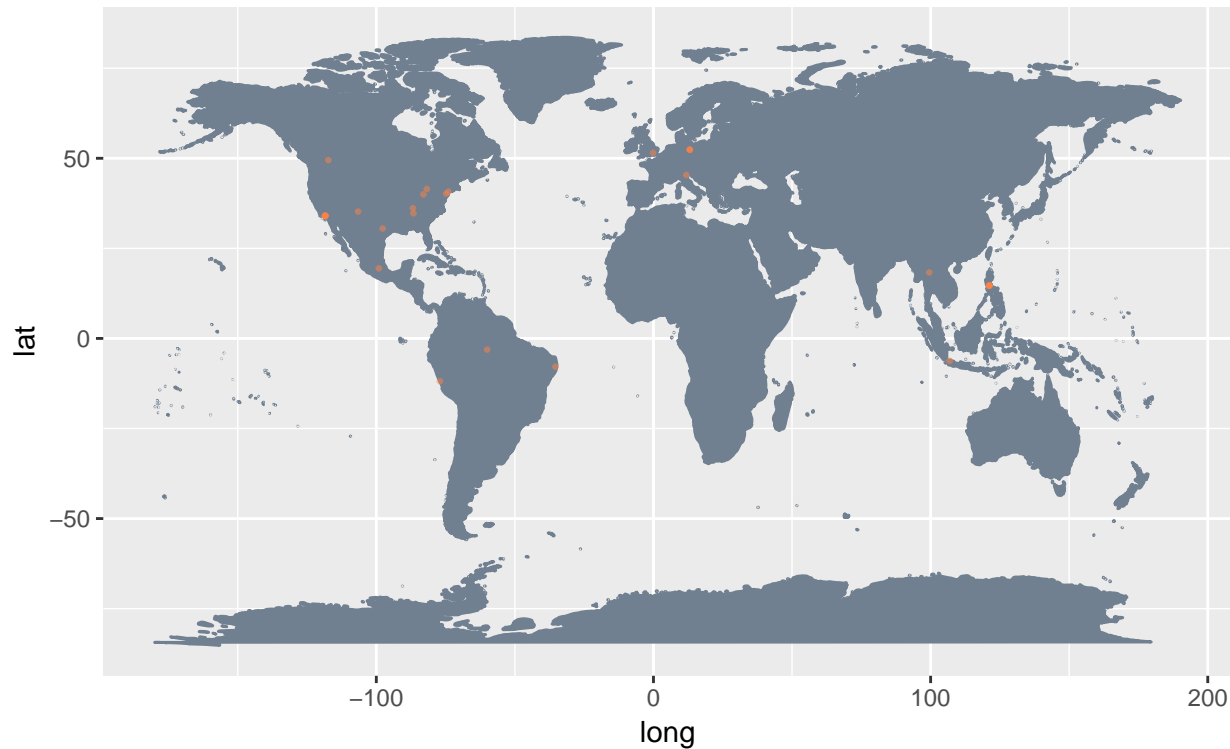


Analysis of Taylor Swift

Search Topic

First search 5000 results of the topic #Taylor Swift from Twitter. It looks like most of the majority part of points I collected are in the United States. There is no points in the Africa.

```
## [1] 32
```



Calculate the count of tweets by platform.

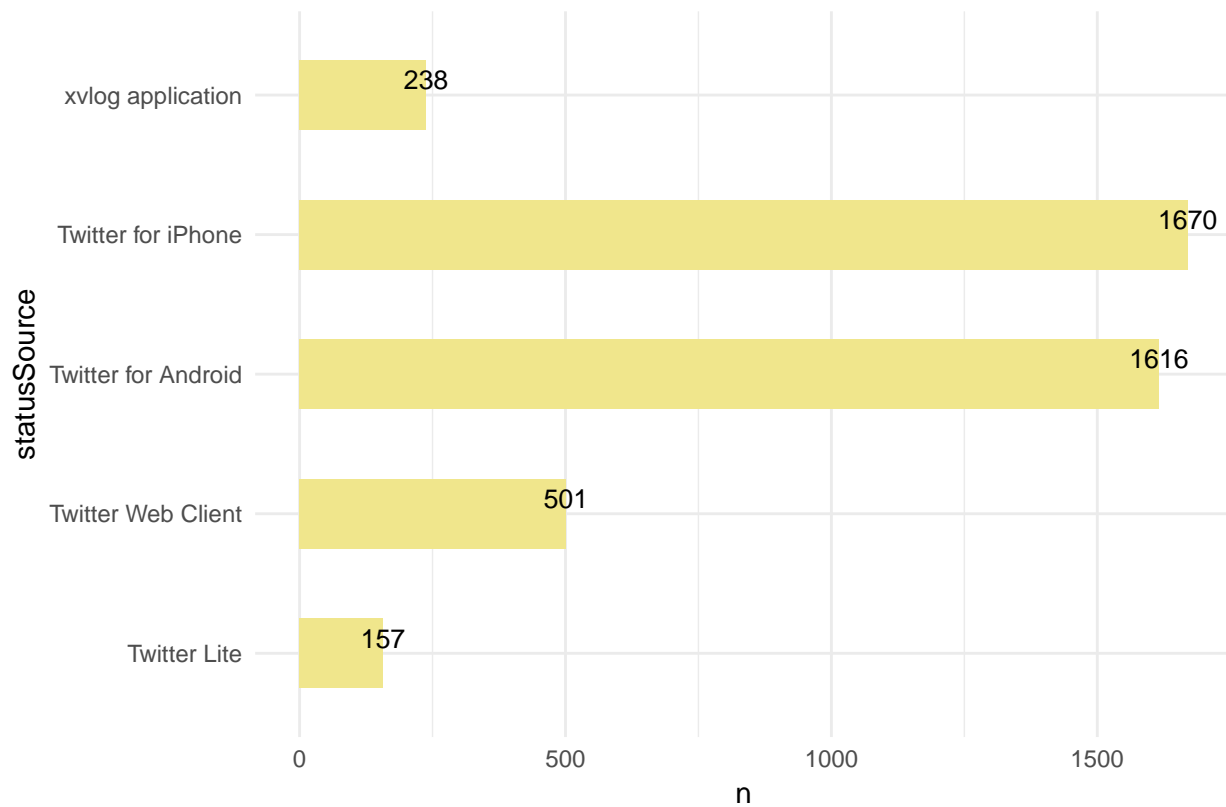
Here is the table of the count of the source. Different from #Rihanna, most of the tweets contain topic #Taylor Swift are sent from iPhone.

statusSource	n	percent_of_tweets
Twitter for iPhone	1670	0.334
Twitter for Android	1616	0.323
Twitter Web Client	501	0.100
xvlog application	238	0.048
Twitter Lite	157	0.031

Barplot of the count of the source

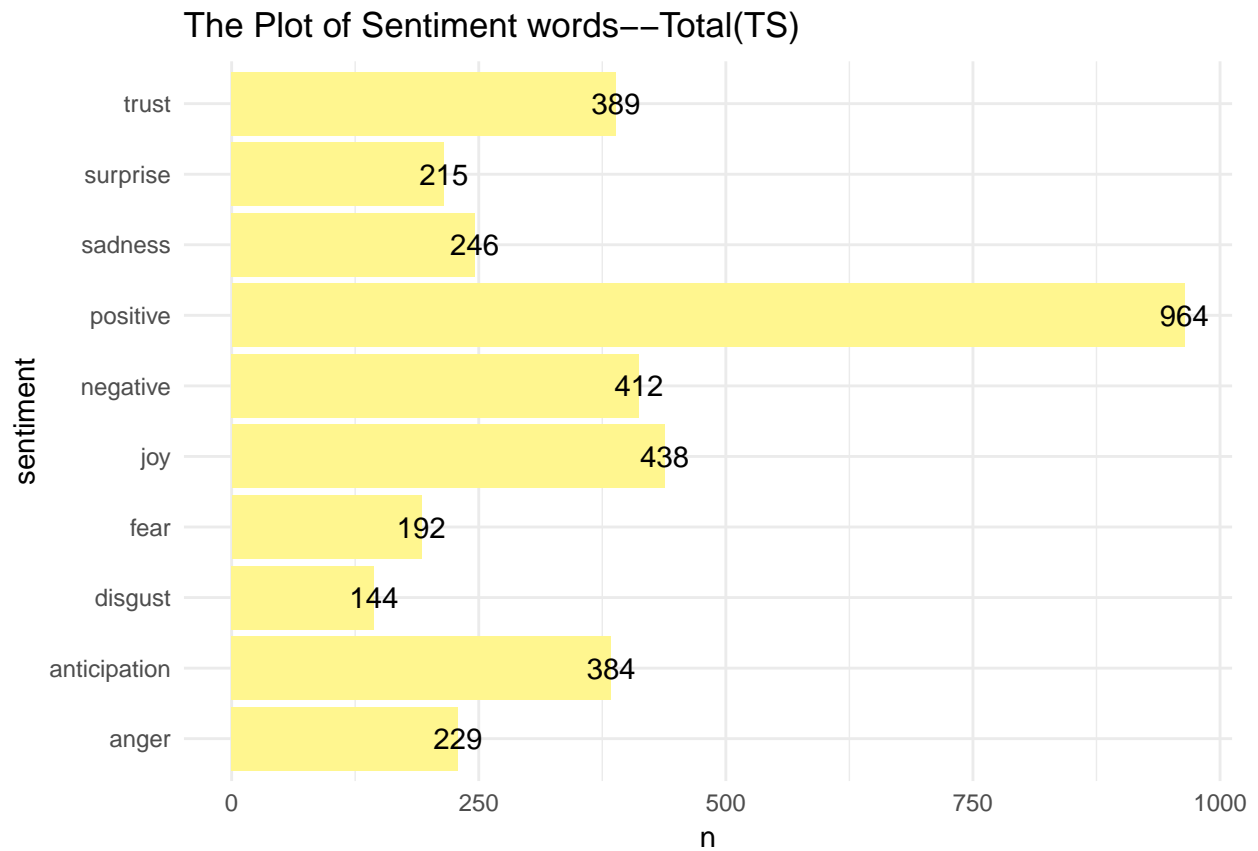
Here is the barplot of the count of the source.

The Counts of the Data From Different Platform–TS



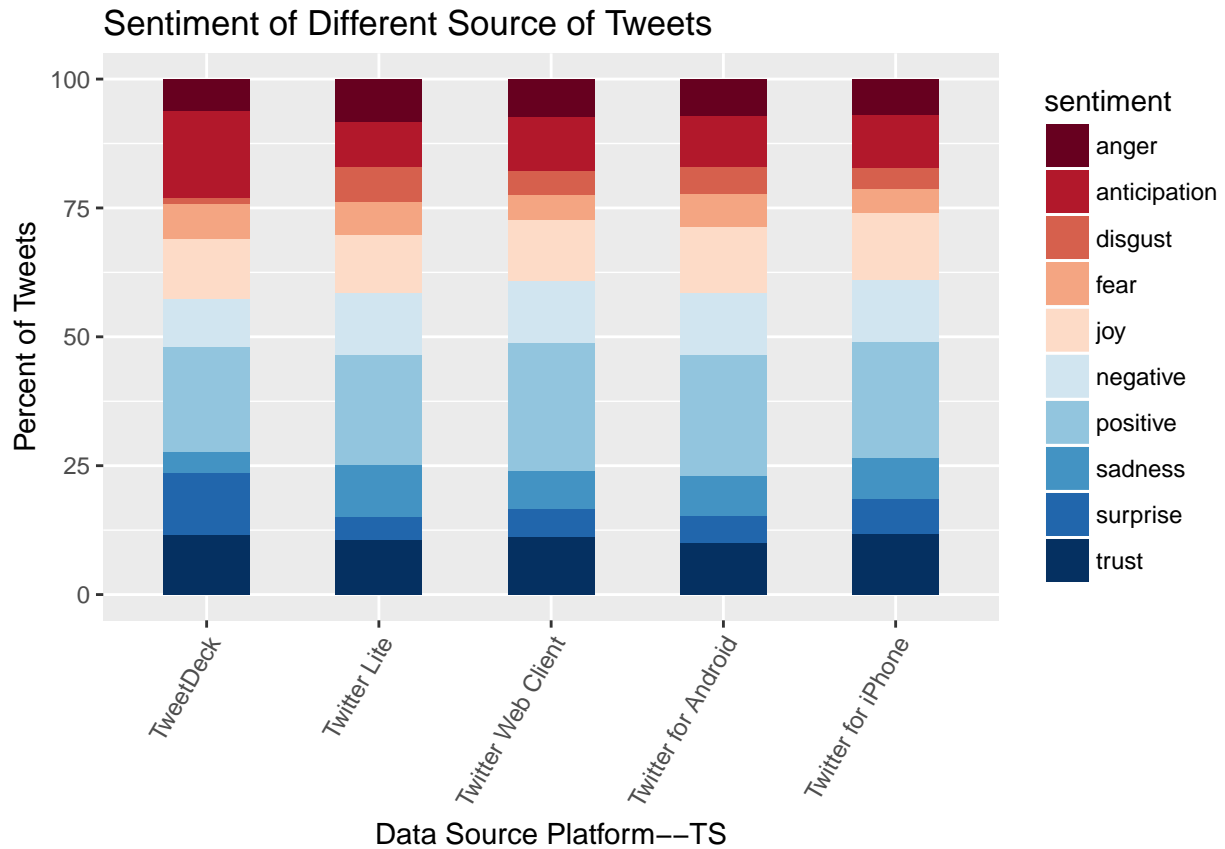
Sentiment of total

After clean the data, plot the Sentiment of Topic #Taylor Swift in total. The result below shows that positive words appear frequently in the topic of Taylor Swift.



Sentiment of top 5 data source

Calculate and plot the Sentiment of Topic #Taylor Swift from different data source platform. The results is similar to the topic of #Rihanna.



Wordclouds Analysis--Top 10 words in the Topic(#Taylor Swift)

The character of "Taylor Swift" appears a lot in the top 10 words.

```
## chr [1:24497] "#demilovato" "#taylorswift" "#beyonc" "00e9" ...
```

