Sentiment on Economy

Abhijith

6/28/2021

```
library(rtweet)
library(ggplot2)
library(dplyr)
library(tidytext)
library(tm)
library(textdata)
library(wordcloud)
library(syuzhet)
library(RColorBrewer)
library(SnowballC)
library(RCurl)
library(xml2)
library(wordcloud2)
library(qdapRegex)
library(stringr)
```

connecting to twitter API

```
twitter_token<-create_token(app = app_name,consumer_key=key,consumer_secret = secret,access_toke
n = access,access_secret = accesssecr)
twitter_token
```

```
## <Token>
## <oauth_endpoint>
##  request:   https://api.twitter.com/oauth/request_token
##  authorize: https://api.twitter.com/oauth/authenticate
##  access:    https://api.twitter.com/oauth/access_token
## <oauth_app> abhijith_senti
##   key:    tqJ3SCw4tEr2LlsXqQ47JlisV
##   secret: <hidden>
## <credentials> oauth_token, oauth_token_secret
## ---
```

To get all the tweets with “economy” keyword in it and I decided not to include retweets in this

```
tweet_econ<-search_tweets(q="economy",n=2000,include_rts = FALSE)
```

only taking text column and seperating it as individual vector

```
tweet_text<-str_c(tweet_econ$text,collapse="")
```

cleaning

```
tweet_text<-tweet_text %>% str_remove("\\n") %>% rm_twitter_url() %>% rm_url() %>%  str_remove_a
ll("@\\S+") %>% str_remove_all("#\\S+") %>% removeWords(stopwords("english")) %>% removeNumbers
() %>% stripWhitespace() %>% removeWords(c("amp","t.co")) %>% str_remove_all(".,") %>%removePunc
tuation() %>% wordStem(language = "english")
```

converting the cleaned document into a form of Matrix

```
text_summ <- Corpus(VectorSource(tweet_text)) %>% TermDocumentMatrix() %>% as.matrix()
text_summ<-sort(rowSums(text_summ),decreasing=TRUE)
```

Getting individual data frame of words with the count for wordcloud

```
text_summ<- data.frame(word=names(text_summ),freq=text_summ,row.names=NULL)
text_summ
```

Using wordcloud2 package, to create a wordcloud

```
word_cloud<-wordcloud2(data=text_summ,minRotation = 0,maxRotation = 0,ellipticity = 0.6)

word_cloud
```

Now since we got the wordcloud. We will get back to sentiment analysis Cleaning the document in total again, especially text column

```
econ_tweet<-tweet_econ%>% select(screen_name,text)
econ_tweet$s_text<-gsub("https","",econ_tweet$text)
econ_tweet$s_text<-gsub("t.co","",econ_tweet$s_text)
econ_tweet$s_text<-gsub("\\|","",econ_tweet$s_text)
econ_tweet$s_text<-gsub("@","",econ_tweet$s_text)
econ_tweet$s_text<-gsub("/","",econ_tweet$s_text)
```

Stemming,removing numbers and punctations using tm package,textdata package and snowballC

```
econ_tweet$s_text<-econ_tweet$s_text %>% wordStem(language="english") %>% removeNumbers() %>% re
movePunctuation() %>% removeWords(stopwords("english"))
```

Removing common english stopwords for the convenience of analysis

```
econ_twitter<-econ_tweet %>% select(s_text) %>% unnest_tokens(word,s_text)
economy_twitter<-econ_twitter %>% anti_join(stop_words)
```

```
## Joining, by = "word"
```

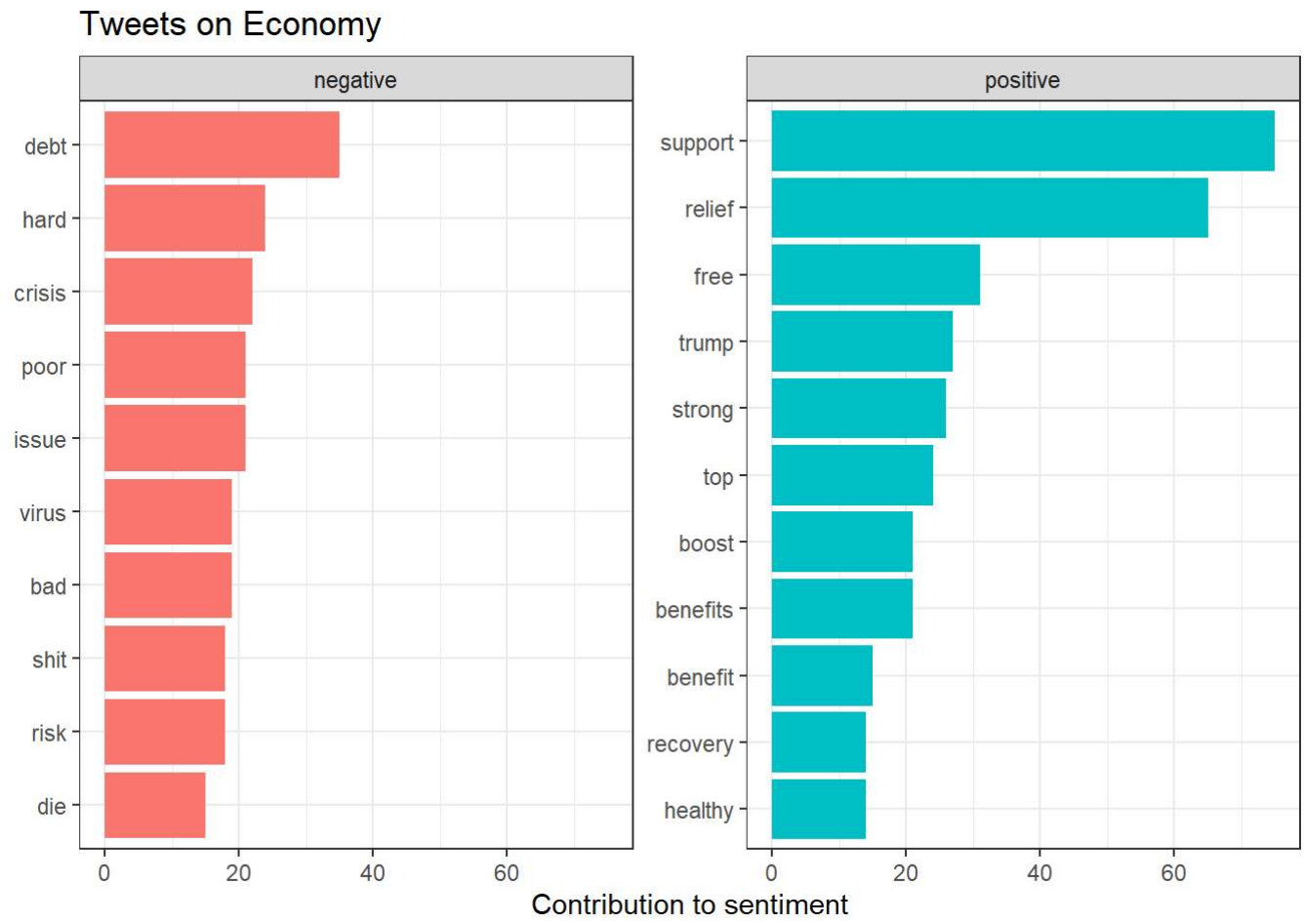
Using tidytext and dplyr by using bing,afinn lexicons we can run our sentiment analysis using bing lexicon and ggplot we can see what are most negative tweets and most positive tweets

```
bing_econ<-economy_twitter %>% inner_join(get_sentiments("bing")) %>% count(word,sentiment,sort=
TRUE) %>% ungroup()
```

```
## Joining, by = "word"
```

```
bing_econ%>% group_by(sentiment) %>% top_n(10) %>% ungroup() %>% mutate(word=reorder(word,n)) %
>% ggplot(aes(word,n,fill=sentiment))+geom_col(show.legend=FALSE)+facet_wrap(~sentiment,scales=
"free_y")+labs(title="Tweets on Economy",y="Contribution to sentiment",x=NULL)+coord_flip()+them
e_bw()
```

```
## Selecting by n
```




```
ggsave("Contribution to sentiment economy.png")
```

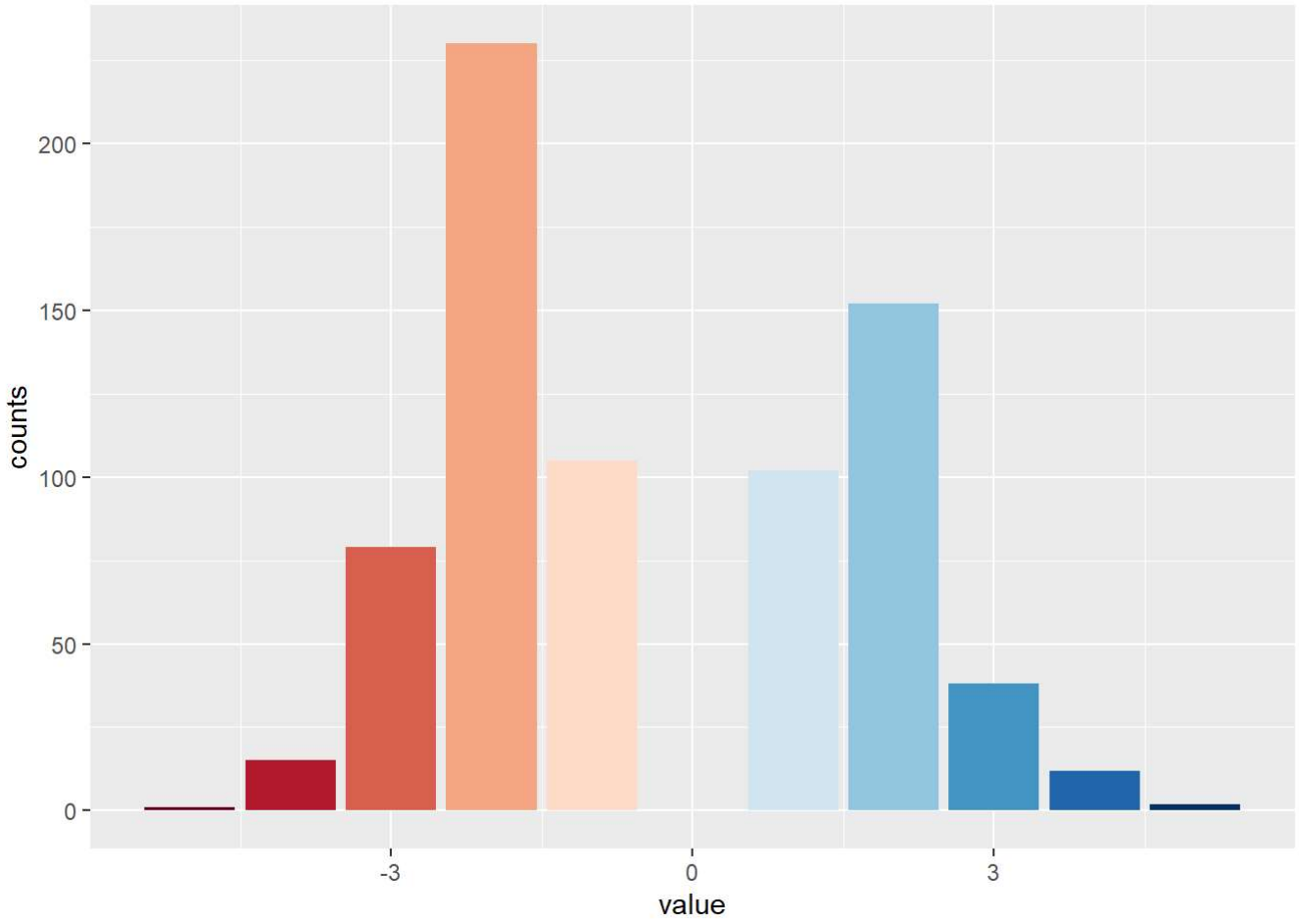
```
## Saving 7 x 5 in image
```

Using affin lexicon and ggplot we can create barplot where we can observe whether what type of tweets are more in number, 0 to -4 are negative tweets and 0 to +4 are positive tweets

```
affin_econ<-economy_twitter %>% inner_join(get_sentiments("afinn")) %>% count(word,value,sort=TR
UE) %>% ungroup()
```

```
## Joining, by = "word"
```

```
affin_econ<-affin_econ %>% group_by(value) %>% summarise(counts=n())
affin_econ %>% ggplot(aes(x=value,y=counts,fill=as.factor(value)))+geom_bar(stat="identity",posi
tion=position_stack(),show.legend = FALSE)+scale_fill_brewer(type="Diverging",palette="RdBu",dir
ection=1,aesthetics="fill")
```


200
150
100
50
0
counts
-3
0
3
value