Demonitisation | UP Election | Win of BJP

Subham Agrawal

23 March 2017

Loading required packages

```
library(twitteR) # fetching the tweets
library(ROAuth) # R authentication
library(plyr) # breaking the data into manageable pieces
```

```
##
## Attaching package: 'plyr'
```

```
## The following object is masked from 'package:twitteR':
##
##     id
```

```
library(stringr) # string processing
library(ggplot2) # plotting results
library(RColorBrewer) # dependency
library(wordcloud) # for wordcloud
library(NLP) # NAtural Language Processing
```

```
##
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```
library(tm) # text mining
library(Rstem) # word stemming
library(SnowballC) # for stemming
```

```
##
## Attaching package: 'SnowballC'
```

```
## The following objects are masked from 'package:Rstem':
##
##     getStemLanguages, wordStem
```

```
library(sentiment) # sentiment analysis
```

Twitter API Credentials

```
reqURL <- "https://api.twitter.com/oauth/request_token"
accessURL <- "https://api.twitter.com/oauth/access_token"
authURL <- "https://api.twitter.com/oauth/authorize"
api_key <- "e2hAYurbg7h8oCCaK572HGzpb"
api_secret <- "em65eNpyeeBfRhZyNJ2pBW2wJvxxsl1tFyKXZlsdefrOonjElM"
access_token <- "836133291136692224-PXZFKrLmtzffzqEqDFVCLGaXUt35Rod"
access_token_secret <- "HzjHWBnParuUYkxxxcomhPhGajMRTiTbNb1p9AhEiS37R"
```

Authorization

```
setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)
```

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

```
authenticate <- OAuthFactory$new(consumerKey=api_key, consumerSecret=api_secret, requ
estURL=reqURL, accessURL=accessURL, authURL=authURL)
```

1. download.file(url="http://curl.haxx.se/ca/cacert.pem (http://curl.haxx.se/ca/cacert.pem)",destfile="cacert.pem")
2. authenticate$handshake(cainfo="cacert.pem") It navigates to the specified link to authorize app and click "Authorize App". In RStudio, type in the pin number. Save the object "cred" on your local machine as "twitter authentication.Rdata."
3. save(authenticate, file="twitter authentication.Rdata")

ANALYSIS FOR DEMONETISATION

Getting tweets and creating data frame with additional information on tweets

Here, number of tweets is 2000 on demonetisation

```
tweets=searchTwitter("demonetisation", n=1500, lang="en", since='2017-02-01',
until='2017-03-22')
```

Convert the tweets to a text format

```
tweets_txt=sapply(tweets, function(t) t$getText())
```

getting dates for each tweet

```
tweets_date=lapply(tweets, function(x) x$getCreated())
tweets_date=sapply(tweets_date,function(x) strftime(x, format="%Y-%m-%d %H:%M:%S",tz
= "UTC"))
```

getting retweet count

```
isretweet=sapply(tweets, function(x) x$getIsRetweet())
retweetcount=sapply(tweets, function(x) x$getRetweetCount())
```

getting favorite count

```
favoritecount=sapply(tweets, function(x) x$getFavoriteCount())
```

Creating dataframe with all information

```
data=as.data.frame(cbind(tweet=tweets_txt, date=tweets_date, isretweet=isretweet, ret
weetcount=retweetcount, favoritecount=favoritecount))
```

Importing text files for positive and negative words

```
hu.liu.pos=scan(file="positive-words.txt", what='character', comment.char=';')
hu.liu.neg=scan(file="negative-words.txt", what='character', comment.char=';')
```

Preparing text for analytics i.e. cleaning

define error handling function when trying tolower

```
tryTolower = function(x)
{
  # create missing value
  y = NA
  # tryCatch error
  try_error = tryCatch(tolower(x), error=function(e) e)
  # if not an error
  if (!inherits(try_error, "error"))
    y = tolower(x)
  # result
  return(y)
}
```

clean tweet function

```
cleantweets = function(tweets_txt)
{
  # remove retweet entities
  tweets_txt = gsub("(RT|via)((?:\\b\\W*@\\w+)+)", "", tweets_txt)
  # remove @people
  tweets_txt = gsub("@\\w+", "", tweets_txt)
  # remove hashtags with space so that words have space in between them
  tweets_txt = gsub("#"," ",tweets_txt)
  # remove /n
  tweets_txt = gsub("\n","",tweets_txt)
  # remove punctuation
  tweets_txt = gsub("[[:punct:]]", "", tweets_txt)
  # remove numbers
  tweets_txt = gsub("[[:digit:]]", "", tweets_txt)
  # remove html links
  tweets_txt = gsub("http\\w+", "", tweets_txt)
  # removes all besides the alphabets and numbers
  # it removes all the hindi in text as well
  tweets_txt = gsub("[^A-Za-z0-9]", " ", tweets_txt)
  # remove unnecessary spaces
  tweets_txt = gsub("[ \t]{2,}", " ", tweets_txt)
  tweets_txt = gsub("^\\s+|\\s+$", "", tweets_txt)
  # use tryTolower with sapply
  tweets_txt = tryTolower(tweets_txt)
  # result
  return(tweets_txt)
}
```

Now we check if our text is cleaned as required by looking at head of tweets before and after applying clean function.

```
head(tweets_txt)
```

```
## [1] "Law prevails over PM Modi’s word: A-G tells SC on note ban deadline https://
t.co/tlloPYVshc https://t.co/cjjsZeCgdV"
## [2] "Note ban most disruptive policy innovation since..#DeMonetisation, #NoteBan,#
Modi,#BJP,#RBI,#DeMonetisationDisaster. https://t.co/seX2FvEjwM"
## [3] "RT @Vidyut: What Bill Gates or other foreign influencers have to do with #DeM
onetisation - https://t.co/v6RJf4ot9I https://t.co/MmGieWeFxF"
## [4] "After Demonetisation, Digital Deluge, Banks still not optimised.#DeMonetisati
on,#Digital,#Digitisation,#NoteBan,https://t.co/LkC9HnsoLv"
## [5] "RT @3NovicesHyd: #3Novices : Law prevails over PM Modi’s word: A-G tells SC o
n note ban deadline March 22, 2017 at 04:59AM https://t.co/Uhp…"
## [6] "#3Novices : Law prevails over PM Modi’s word: A-G tells SC on note ban deadli
ne March 22, 2017 at 04:59AM https://t.co/UhpRomtEZj #News #H…"
```

```
tweets_txt = cleantweets(tweets_txt)
head(tweets_txt)
```

```
## [1] "law prevails over pm modis word ag tells sc on note ban deadline"

## [2] "note ban most disruptive policy innovation since demonetisation noteban modi
 bjp rbi demonetisationdisaster"
## [3] "what bill gates or other foreign influencers have to do with demonetisation"

## [4] "after demonetisation digital deluge banks still not optimised demonetisation
 digital digitisation noteban"
## [5] "novices law prevails over pm modis word ag tells sc on note ban deadline marc
h at am"
## [6] "novices law prevails over pm modis word ag tells sc on note ban deadline marc
h at am news h"
```

We can see that every word is separated with space in between and unwanted text is removed.

Scoring function using above cleantweets function

```
score.sentiment = function(sentences, pos.words, neg.words, .progress='none')
{
  # Parameters
  # sentences: vector of text to score
  # pos.words: vector of words of positive sentiment
  # neg.words: vector of words of negative sentiment
  # .progress: passed to laply() to control of progress bar
  # create a simple array of scores with laply
  scores = laply(sentences,
                 function(sentence, pos.words, neg.words)
                 {
                   # preparing text for analysis
                   sentence = cleantweets(sentence)

                   # split sentence into words with str_split (stringr package)
                   word.list = str_split(sentence, "\\s+")
                   words = unlist(word.list)

                   # compare words to the dictionaries of positive & negative terms
                   pos.matches = match(words, pos.words)
                   neg.matches = match(words, neg.words)

                   # get the position of the matched term or NA
                   # we just want a TRUE/FALSE
                   pos.matches = !is.na(pos.matches)
                   neg.matches = !is.na(neg.matches)

                   # final score
                   score = sum(pos.matches) - sum(neg.matches)
                   return(score)
                 }, pos.words, neg.words, .progress=.progress )

  # data frame with scores for each sentence
  scores.df = data.frame(text=sentences, score=scores)
  return(scores.df)
}
```

Plot of tweet score distribution

```
tweet_score=score.sentiment(tweets_txt, hu.liu.pos, hu.liu.neg, .progress='text')
```

```
##
  |
  |                                                                      |   0%
  |
  |                                                                      |   1%
  |
  |=                                                                     |   1%
  |
  |=                                                                     |   2%
  |
  |==                                                                    |   2%
  |
  |==                                                                    |   3%
  |
  |==                                                                    |   4%
  |
  |===                                                                   |   4%
  |
  |===                                                                   |   5%
  |
  |====                                                                  |   5%
  |
  |====                                                                  |   6%
  |
  |====                                                                  |   7%
  |
  |=====                                                                 |   7%
  |
  |=====                                                                 |   8%
  |
  |======                                                                |   8%
  |
  |======                                                                |   9%
  |
  |======                                                                |  10%
  |
  |=======                                                               |  10%
  |
  |=======                                                               |  11%
  |
  |=======                                                               |  12%
  |
  |========                                                              |  12%
  |
  |========                                                              |  13%
  |
  |=========                                                             |  13%
  |
  |=========                                                             |  14%
  |
  |=========                                                             |  15%
  |
  |==========                                                            |  15%
  |
  |==========                                                            |  16%
  |
  |===========                                                           |  16%
```

```
|
|===========                                                    |  17%
|
|===========                                                    |  18%
|
|============                                                   |  18%
|
|============                                                   |  19%
|
|=============                                                  |  19%
|
|=============                                                  |  20%
|
|=============                                                  |  21%
|
|==============                                                 |  21%
|
|==============                                                 |  22%
|
|===============                                                |  22%
|
|===============                                                |  23%
|
|===============                                                |  24%
|
|================                                               |  24%
|
|================                                               |  25%
|
|=================                                              |  25%
|
|=================                                              |  26%
|
|=================                                              |  27%
|
|==================                                             |  27%
|
|==================                                             |  28%
|
|===================                                            |  28%
|
|===================                                            |  29%
|
|===================                                            |  30%
|
|====================                                           |  30%
|
|====================                                           |  31%
|
|====================                                           |  32%
|
|=====================                                          |  32%
|
|=====================                                          |  33%
|
|======================                                         |  33%
|
|======================                                         |  34%
```

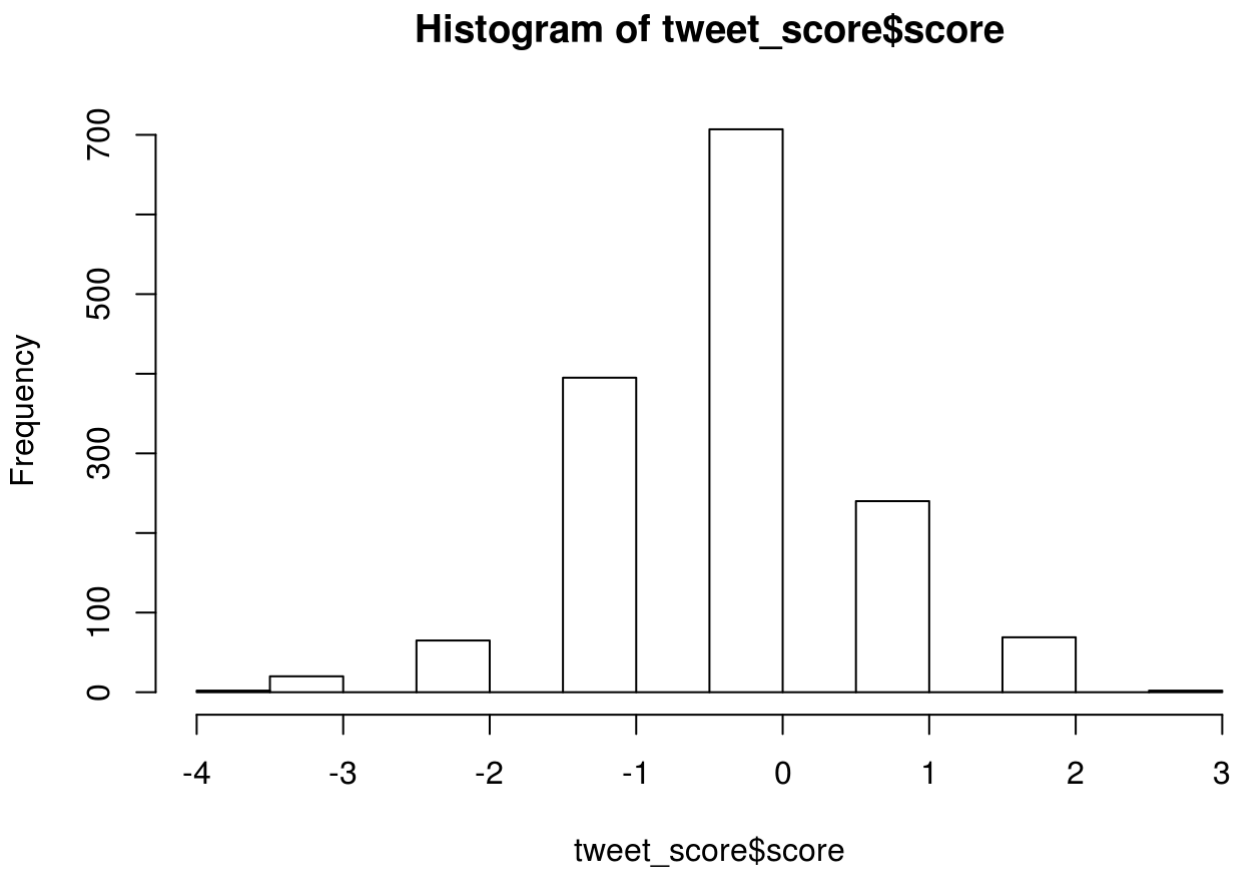
```
|
|======================                                          |  35%
|
|======================                                          |  35%
|
|======================                                          |  36%
|
|======================                                          |  36%
|
|======================                                          |  37%
|
|======================                                          |  38%
|
|======================                                          |  38%
|
|======================                                          |  39%
|
|======================                                          |  39%
|
|======================                                          |  40%
|
|======================                                          |  41%
|
|======================                                          |  41%
|
|======================                                          |  42%
|
|======================                                          |  42%
|
|======================                                          |  43%
|
|======================                                          |  44%
|
|======================                                          |  44%
|
|======================                                          |  45%
|
|======================                                          |  45%
|
|======================                                          |  46%
|
|======================                                          |  47%
|
|======================                                          |  47%
|
|======================                                          |  48%
|
|======================                                          |  48%
|
|======================                                          |  49%
|
|======================                                          |  50%
|
|======================                                          |  50%
|
|======================                                          |  51%
|
|======================                                          |  52%
```

```
|
|==================================                                    |  52%
|
|==================================                                    |  53%
|
|===================================                                   |  53%
|
|===================================                                   |  54%
|
|===================================                                   |  55%
|
|====================================                                  |  55%
|
|====================================                                  |  56%
|
|====================================                                  |  56%
|
|=====================================                                 |  57%
|
|=====================================                                 |  58%
|
|=====================================                                 |  58%
|
|======================================                                |  59%
|
|======================================                                |  59%
|
|======================================                                |  60%
|
|=======================================                               |  61%
|
|=======================================                               |  61%
|
|=======================================                               |  62%
|
|========================================                              |  62%
|
|========================================                              |  63%
|
|========================================                              |  64%
|
|=========================================                             |  64%
|
|=========================================                             |  65%
|
|=========================================                             |  65%
|
|==========================================                            |  66%
|
|==========================================                            |  67%
|
|==========================================                            |  67%
|
|===========================================                           |  68%
|
|============================================                          |  68%
|
|============================================                          |  69%
```

```
|
|=============================================                    |  70%
|
|=============================================                    |  70%
|
|=============================================                    |  71%
|
|=============================================                    |  72%
|
|=============================================                    |  72%
|
|=============================================                    |  73%
|
|=============================================                    |  73%
|
|=============================================                    |  74%
|
|=============================================                    |  75%
|
|=============================================                    |  75%
|
|=============================================                    |  76%
|
|=============================================                    |  76%
|
|=============================================                    |  77%
|
|=============================================                    |  78%
|
|=============================================                    |  78%
|
|=============================================                    |  79%
|
|=============================================                    |  79%
|
|=============================================                    |  80%
|
|=============================================                    |  81%
|
|=============================================                    |  81%
|
|=============================================                    |  82%
|
|=============================================                    |  82%
|
|=============================================                    |  83%
|
|=============================================                    |  84%
|
|=============================================                    |  84%
|
|=============================================                    |  85%
|
|=============================================                    |  85%
|
|=============================================                    |  86%
|
|=============================================                    |  87%
```

```
  |
  |=============================================================         |  87%
  |
  |=============================================================         |  88%
  |
  |==============================================================        |  88%
  |
  |==============================================================        |  89%
  |
  |==============================================================        |  90%
  |
  |===============================================================       |  90%
  |
  |===============================================================       |  91%
  |
  |===============================================================       |  92%
  |
  |================================================================      |  92%
  |
  |================================================================      |  93%
  |
  |================================================================      |  93%
  |
  |================================================================      |  94%
  |
  |================================================================      |  95%
  |
  |=================================================================     |  95%
  |
  |=================================================================     |  96%
  |
  |=================================================================     |  96%
  |
  |=================================================================     |  97%
  |
  |=================================================================     |  98%
  |
  |=================================================================     |  98%
  |
  |=================================================================     |  99%
  |
  |=================================================================     |  99%
  |
  |=================================================================     | 100%
```

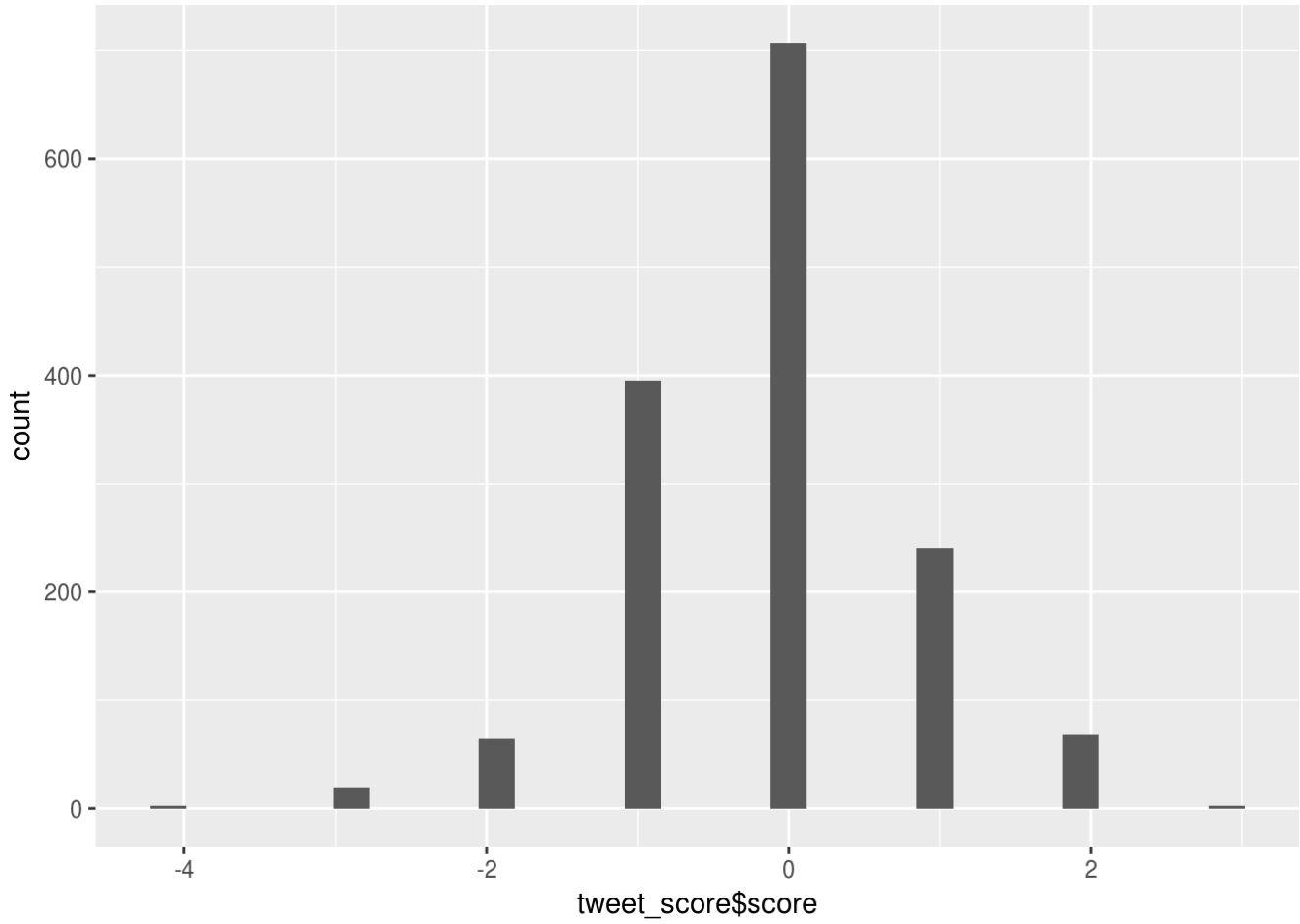
```
hist(tweet_score$score)
```



Plot using ggplot2 for better graphics

```
qplot(tweet_score$score)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



From the above plot, we can see that people have neutral opinion about demonetiation. As the score of -1 is higher than +1, we can say people are pessimistic and spectacular about demonetisation. But they don't have very negative neither very positive opinions on it. We will find out more from further analysis.

Ignore middle and calculate score

```
tweet_score$very.pos=as.numeric(tweet_score$score>=2)
tweet_score$very.neg=as.numeric(tweet_score$score<=-2)
pos.count=sum(tweet_score$very.pos)
neg.count=sum(tweet_score$very.neg)
all.count=pos.count+neg.count
score=round(100*pos.count/all.count)
score
```

```
## [1] 45
```

This score further justifies that people are slightly negative towards demonetisation. Let's further analyze the sentiments of people using sentiment package and compare results.

Sentiment analysis using sentiment package in R

classify_emotion funtion classifies the emotions of people in joy, anger, disgust, surprise, fear and sadness.

```
class_emo = classify_emotion(tweets_txt, algorithm="bayes", prior=1.0)
```

get emotion best fit

```
emotion = class_emo[,7]
```

We'll put unknown instead of NAs for tweets we couldn't claasify the emotion.

```
# substitute NA's by "unknown"
emotion[is.na(emotion)] = "unknown"
```

classify polarity

```
class_pol = classify_polarity(tweets_txt, algorithm="bayes")
```

get polarity best fit

```
polarity = class_pol[,4]
```

data frame with results

```
sentiment_df = data.frame(text=tweets_txt, emotion=emotion, polarity=polarity, string
sAsFactors=FALSE)
```

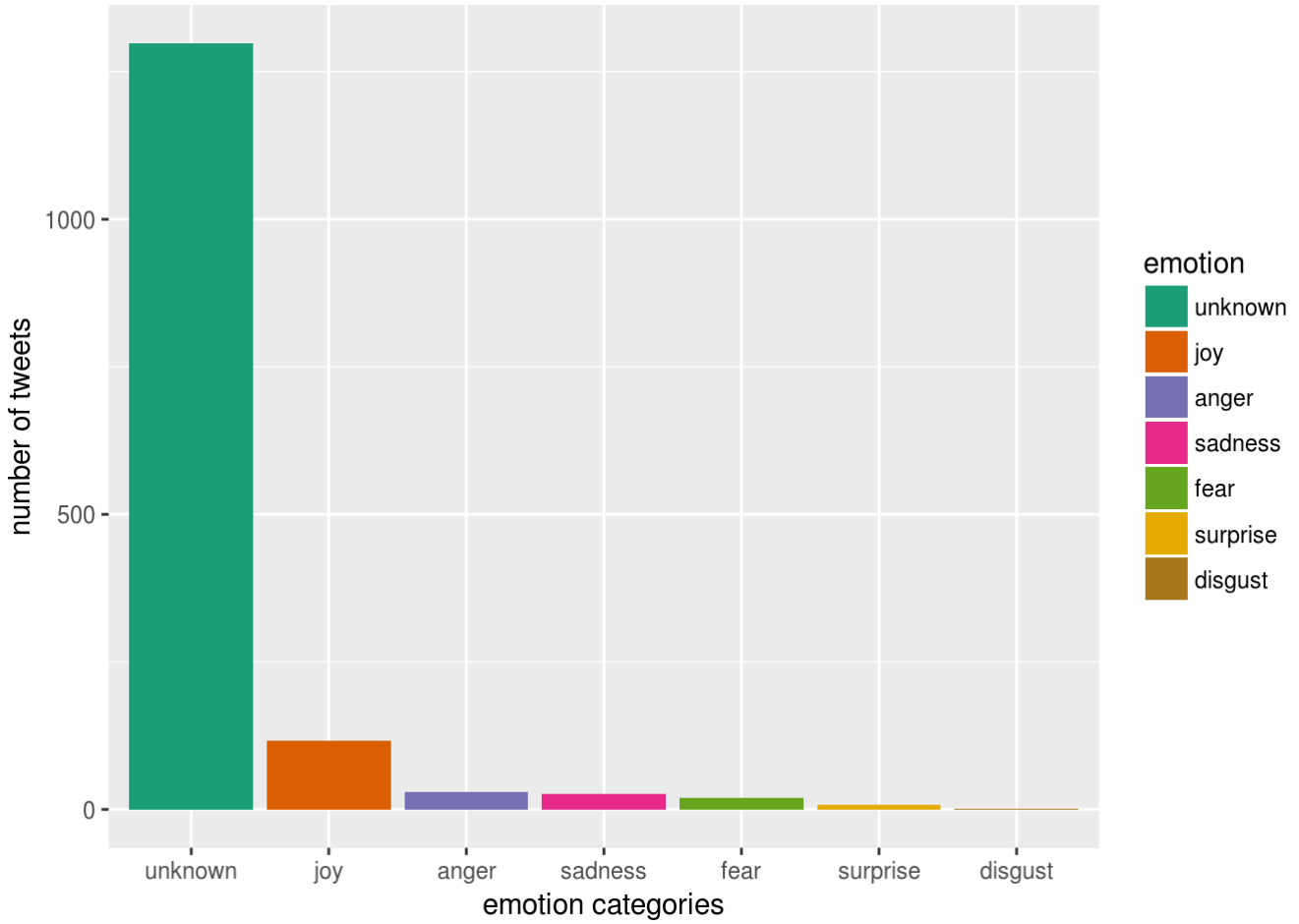
sort data frame according to emotions for a better plot

```
sentiment_df = within(sentiment_df, emotion <- factor(emotion, levels=names(sort(tabl
e(emotion), decreasing=TRUE))))
```

Plots of the obtained result

plot distribution of emotions

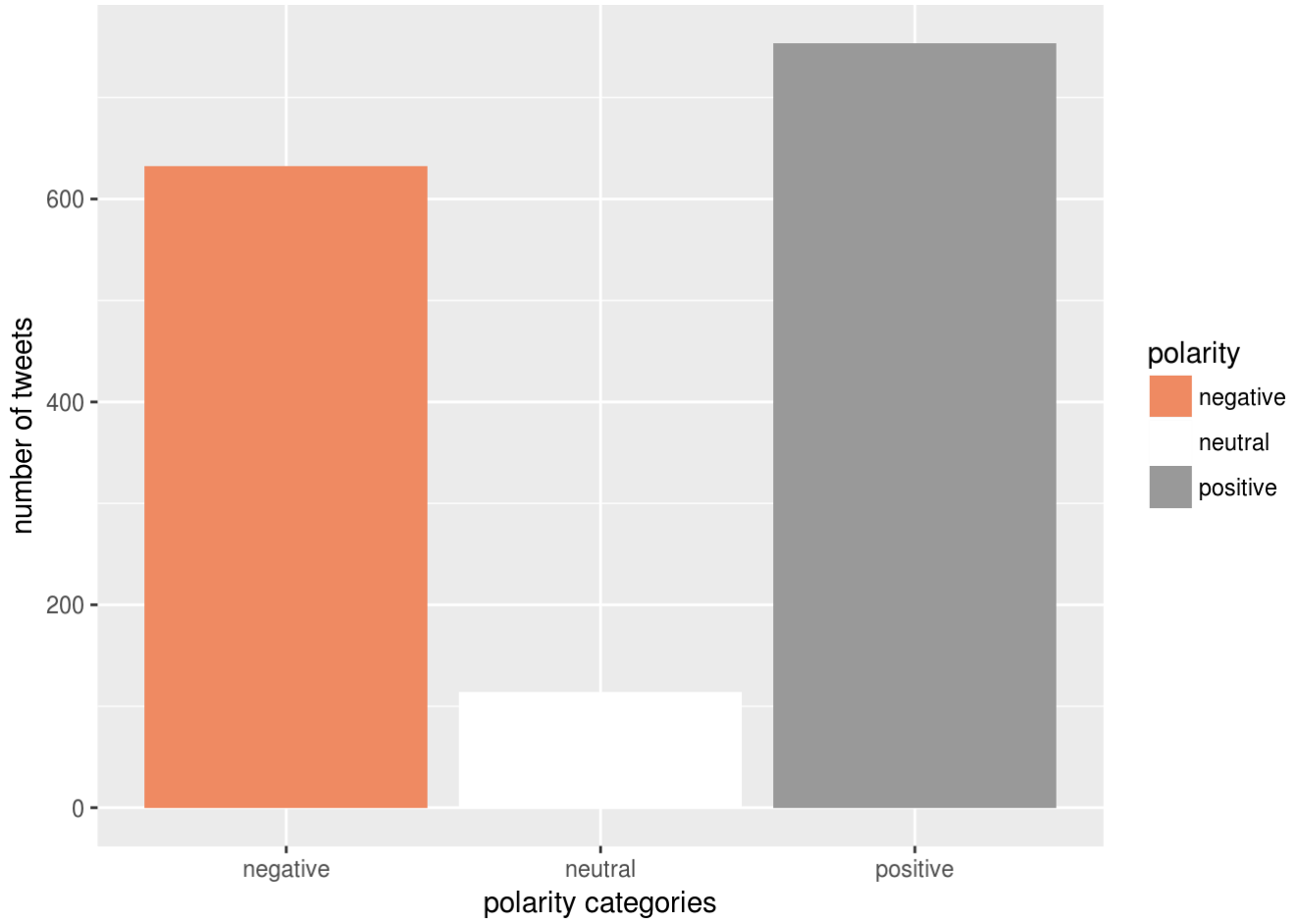
```
ggplot(sentiment_df, aes(x=emotion)) +
  geom_bar(aes(y=..count.., fill=emotion)) +
  scale_fill_brewer(palette="Dark2") +
  labs(x="emotion categories", y="number of tweets")
```



We can observe that people have all kinds of emotions i.e. joy,anger, sadness, fear, surprise and disgust. We couldn't classify emotions for more than half of tweets in data. But we can still see that joy is prominent. It tells that Indian citizens are happy for corrupt free India even after the problems they faced due to note ban. There is one more significant reason i.e. twitter is mostly used by middle class(lower and upper) and youths of country. This is the group which is least affected of demonetisation. Group of poor people who are adversely affected by it are not active in twitter.

plot distribution of polarity

```
ggplot(sentiment_df, aes(x=polarity)) +
  geom_bar(aes(y=..count.., fill=polarity)) +
  scale_fill_brewer(palette="RdGy") +
  labs(x="polarity categories", y="number of tweets")
```



From above plot, we see that poeple are more positive about demonitisation. Positive bar is higher than the negative bar. There are very people who are neutral.

This result contradicts to what we get from plotting score. But we can conclude that people don't have very strong opinions on demonetisation. Lets plot the word cloud to get the reasons behind each emotion and make a better understanding of all.

Separate the words by emotion and visualize the words with a comparison cloud

separating text by emotion

```
emos = levels(factor(sentiment_df$emotion))
nemo = length(emos)
emo.docs = rep("", nemo)
for (i in 1:nemo)
{
  tmp = tweets_txt[emotion == emos[i]]
  emo.docs[i] = paste(tmp, collapse=" ")
}
```

create corpus

```
corpus = Corpus(VectorSource(emo.docs))
```

remove stopwords and word 'demonetisation'

```
corpus = tm_map(corpus, removeWords, c("demonetisation",stopwords("english")))
```

We can stem the document using package SnowballC (so variations of same words can be removed)

```
#corpus = tm_map(corpus, stemDocument)
```

Form Document Term Matrix

```
tdm = TermDocumentMatrix(corpus)
tdm = as.matrix(tdm)
colnames(tdm) = emos
```

comparison word cloud

```
comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"),
                 scale = c(3,.4), random.order = FALSE, title.size = 1.5)
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : udaipurpeople could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : tremendous could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"),
## scale = c(3, : demonetisationall could not be fit on page. It will not be
## plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : especially could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : farming could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : govts could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : housewife could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : insurance could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : insurers could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : knows could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : leaving could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : letter could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : loans could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : ndtv could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : overim could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : partyshame could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : planned could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : politically could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : poorly could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : products could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : prulife could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : pulled could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : raghuram could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : rajasthan could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : redviolets could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : reform could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : regarding could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : resident could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : roses could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : sandeep could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : send could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : served could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : shame could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : shoddily could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : signature could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : sleep could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : sleeps could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : spirit could not be fit on page. It will not be plotted.
```

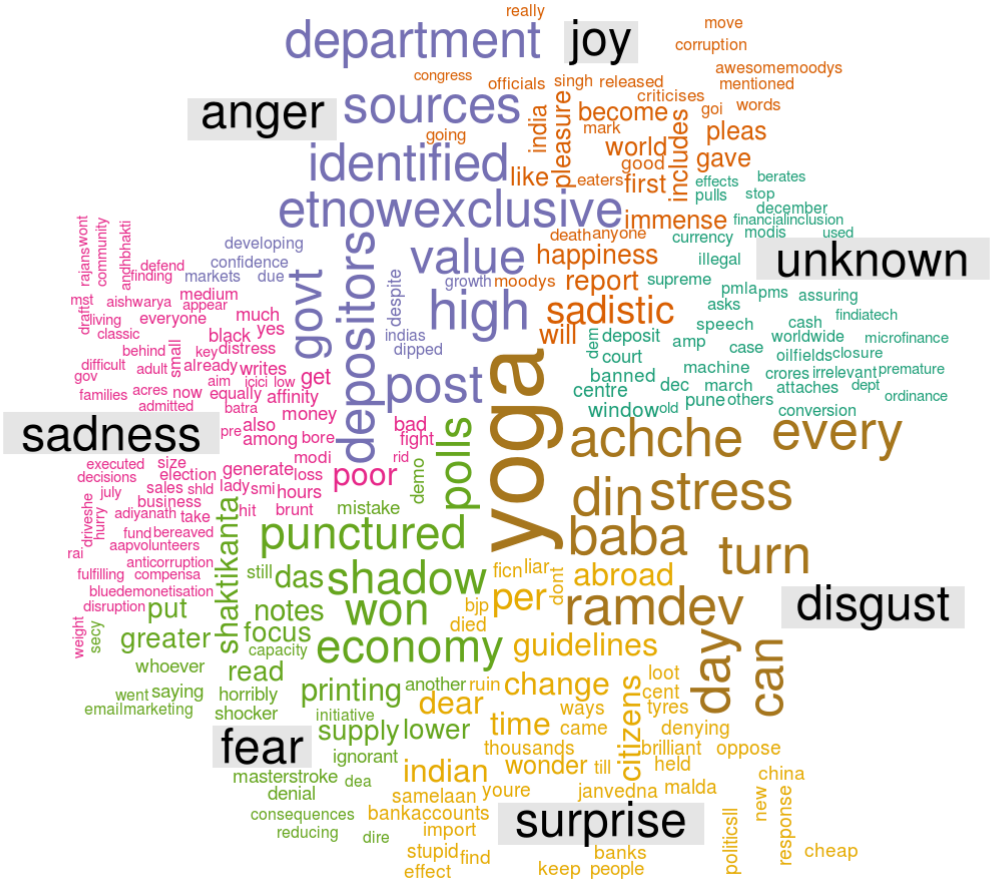
```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : standing could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : transferred could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : views could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : vigilantism could not be fit on page. It will not be plotted.
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(3, : worldpoetryday could not be fit on page. It will not be plotted.
```



Lets look at words plotted for each emotion oone by one. 1. Surprise: We have words such as citizens, response, oppose, wonder, time, politics. Its obvious that people were surprised by the decision of Modi government on note ban. Some people opposed it whereas others were spectacular of the change. 2. Fear: words are economy, printing, supply, notes, consequences, ignorant, etc. These tells us that some poeple fear that we've ignorant government or there might be an effect on economy of India or time it will take to reprint & supply notes. All these were reasons of fear for citizens. 3. Sadness: words are poor, business, election, farming, govt disruption. It shows the sadness people express for farmers and poors. Also how business and election is disrupted because of demonetisation. 4. Anger: markets, developing, growth, depositors, govt, etc. These words shows the anger of people due to poor market and limiting of growth because of demonetisation for government. 5. Joy: good, moodys, awesomemoodys, etc. It shows the optimism of people for Modis's government and people are happy leading towards corrupt free India.

ANALYSIS FOR EFFECT OF DEMONETISATION IN UP ELECTION

Using above code for demonetisation and UP as search word

```
tweets=searchTwitter("demonetisation+UP", n=500, lang="en", since='2017-03-01',
until='2017-03-22')
tweets_txt=sapply(tweets, function(t) t$getText())
tweet_score=score.sentiment(tweets_txt, hu.liu.pos, hu.liu.neg, .progress='text')
```

```
##
  |
  |                                                                      |   0%
  |
  |                                                                      |   1%
  |
  |=                                                                     |   1%
  |
  |=                                                                     |   2%
  |
  |==                                                                    |   2%
  |
  |==                                                                    |   3%
  |
  |==                                                                    |   4%
  |
  |===                                                                   |   4%
  |
  |===                                                                   |   5%
  |
  |====                                                                  |   5%
  |
  |====                                                                  |   6%
  |
  |====                                                                  |   7%
  |
  |=====                                                                 |   7%
  |
  |=====                                                                 |   8%
  |
  |======                                                                |   9%
  |
  |======                                                                |  10%
  |
  |=======                                                               |  10%
  |
  |=======                                                               |  11%
  |
  |========                                                              |  12%
  |
  |========                                                              |  13%
  |
  |=========                                                             |  13%
  |
  |=========                                                             |  14%
  |
  |=========                                                             |  15%
  |
  |==========                                                            |  15%
  |
  |==========                                                            |  16%
  |
  |===========                                                           |  16%
  |
  |===========                                                           |  17%
  |
  |===========                                                           |  18%
```

```
|
|============                                                     |  18%
|
|============                                                     |  19%
|
|=============                                                    |  19%
|
|=============                                                    |  20%
|
|=============                                                    |  21%
|
|==============                                                   |  21%
|
|==============                                                   |  22%
|
|===============                                                  |  22%
|
|===============                                                  |  23%
|
|===============                                                  |  24%
|
|================                                                 |  24%
|
|================                                                 |  25%
|
|=================                                                |  25%
|
|=================                                                |  26%
|
|=================                                                |  27%
|
|==================                                               |  27%
|
|==================                                               |  28%
|
|===================                                              |  29%
|
|===================                                              |  30%
|
|====================                                             |  30%
|
|====================                                             |  31%
|
|=====================                                            |  32%
|
|=====================                                            |  33%
|
|======================                                           |  33%
|
|======================                                           |  34%
|
|======================                                           |  35%
|
|=======================                                          |  35%
|
|=======================                                          |  36%
|
|=======================                                          |  36%
```

```
|
|========================                                 |  37%
|
|========================                                 |  38%
|
|=========================                                |  38%
|
|=========================                                |  39%
|
|=========================                                |  39%
|
|==========================                               |  40%
|
|==========================                               |  41%
|
|==========================                               |  41%
|
|===========================                              |  42%
|
|===========================                              |  42%
|
|============================                             |  43%
|
|============================                             |  44%
|
|============================                             |  44%
|
|=============================                            |  45%
|
|=============================                            |  45%
|
|=============================                            |  46%
|
|==============================                           |  47%
|
|==============================                           |  47%
|
|==============================                           |  48%
|
|===============================                          |  49%
|
|===============================                          |  50%
|
|===============================                          |  50%
|
|================================                         |  51%
|
|================================                         |  52%
|
|=================================                        |  53%
|
|=================================                        |  53%
|
|=================================                        |  54%
|
|==================================                       |  55%
|
|==================================                       |  55%
```

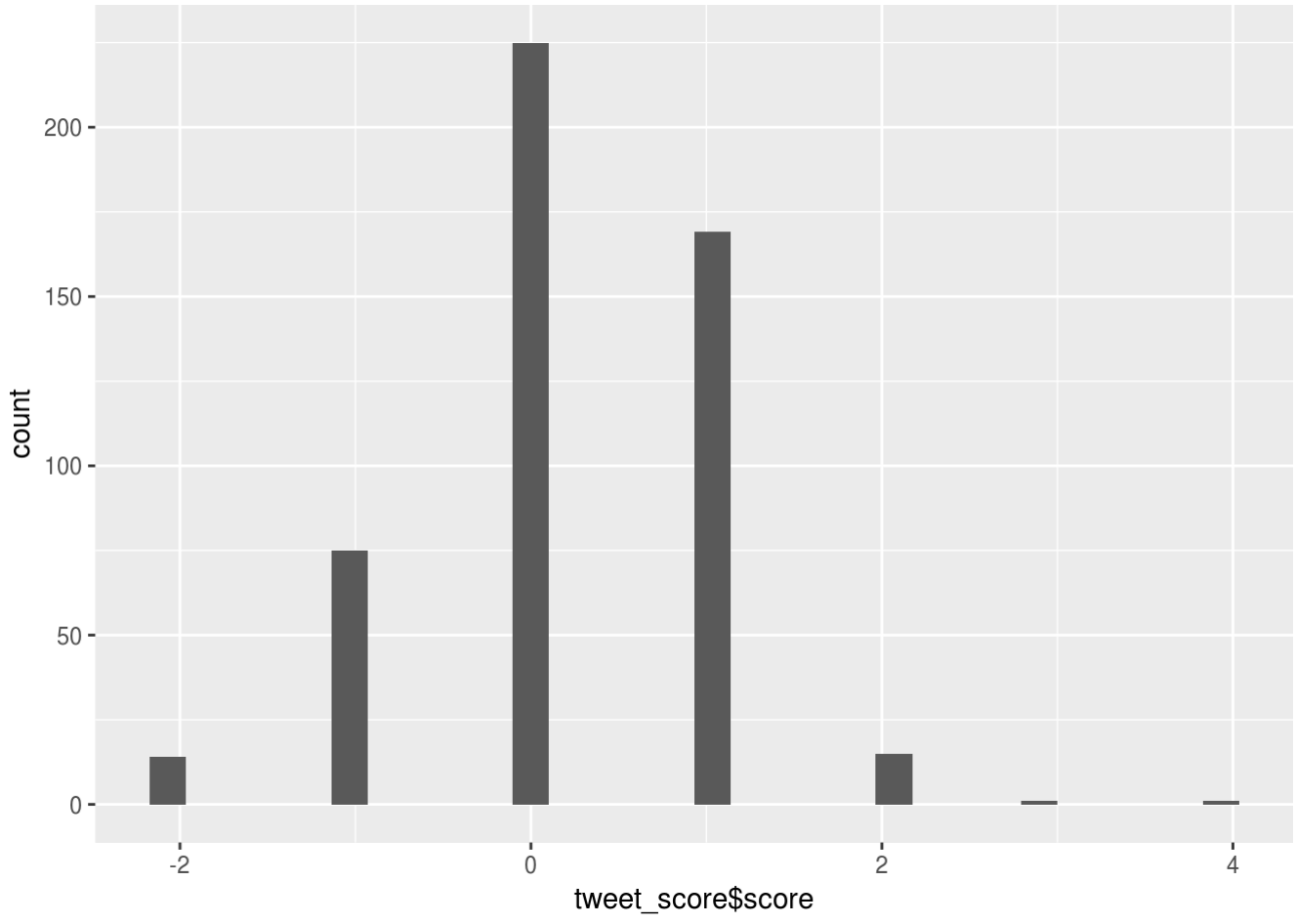
```
|
|==================================                                    |  56%
|
|==================================                                    |  56%
|
|===================================                                   |  57%
|
|===================================                                   |  58%
|
|====================================                                  |  58%
|
|====================================                                  |  59%
|
|=====================================                                 |  59%
|
|=====================================                                 |  60%
|
|======================================                                |  61%
|
|======================================                                |  61%
|
|=======================================                               |  62%
|
|=======================================                               |  62%
|
|========================================                              |  63%
|
|========================================                              |  64%
|
|=========================================                             |  64%
|
|=========================================                             |  65%
|
|==========================================                            |  65%
|
|==========================================                            |  66%
|
|==========================================                            |  67%
|
|===========================================                           |  67%
|
|===========================================                           |  68%
|
|============================================                          |  69%
|
|============================================                          |  70%
|
|============================================                          |  70%
|
|=============================================                         |  71%
|
|=============================================                         |  72%
|
|==============================================                        |  73%
|
|==============================================                        |  73%
|
|==============================================                        |  74%
```

```
|
|================================================                  |  75%
|
|================================================                  |  75%
|
|================================================                  |  76%
|
|================================================                  |  76%
|
|================================================                  |  77%
|
|================================================                  |  78%
|
|================================================                  |  78%
|
|================================================                  |  79%
|
|================================================                  |  79%
|
|================================================                  |  80%
|
|================================================                  |  81%
|
|================================================                  |  81%
|
|================================================                  |  82%
|
|================================================                  |  82%
|
|================================================                  |  83%
|
|================================================                  |  84%
|
|================================================                  |  84%
|
|================================================                  |  85%
|
|================================================                  |  85%
|
|================================================                  |  86%
|
|================================================                  |  87%
|
|================================================                  |  87%
|
|================================================                  |  88%
|
|================================================                  |  89%
|
|================================================                  |  90%
|
|================================================                  |  90%
|
|================================================                  |  91%
|
|================================================                  |  92%
|
|================================================                  |  93%
```

```
  |
  |=============================================================     |  93%
  |
  |==============================================================    |  94%
  |
  |===============================================================   |  95%
  |
  |================================================================  |  95%
  |
  |================================================================  |  96%
  |
  |================================================================= |  96%
  |
  |================================================================= |  97%
  |
  |================================================================= |  98%
  |
  |================================================================== |  98%
  |
  |================================================================== |  99%
  |
  |==================================================================|  99%
  |
  |==================================================================| 100%
```

```
qplot(tweet_score$score)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



We can see that voters have positive opinions about demonetisation for election.

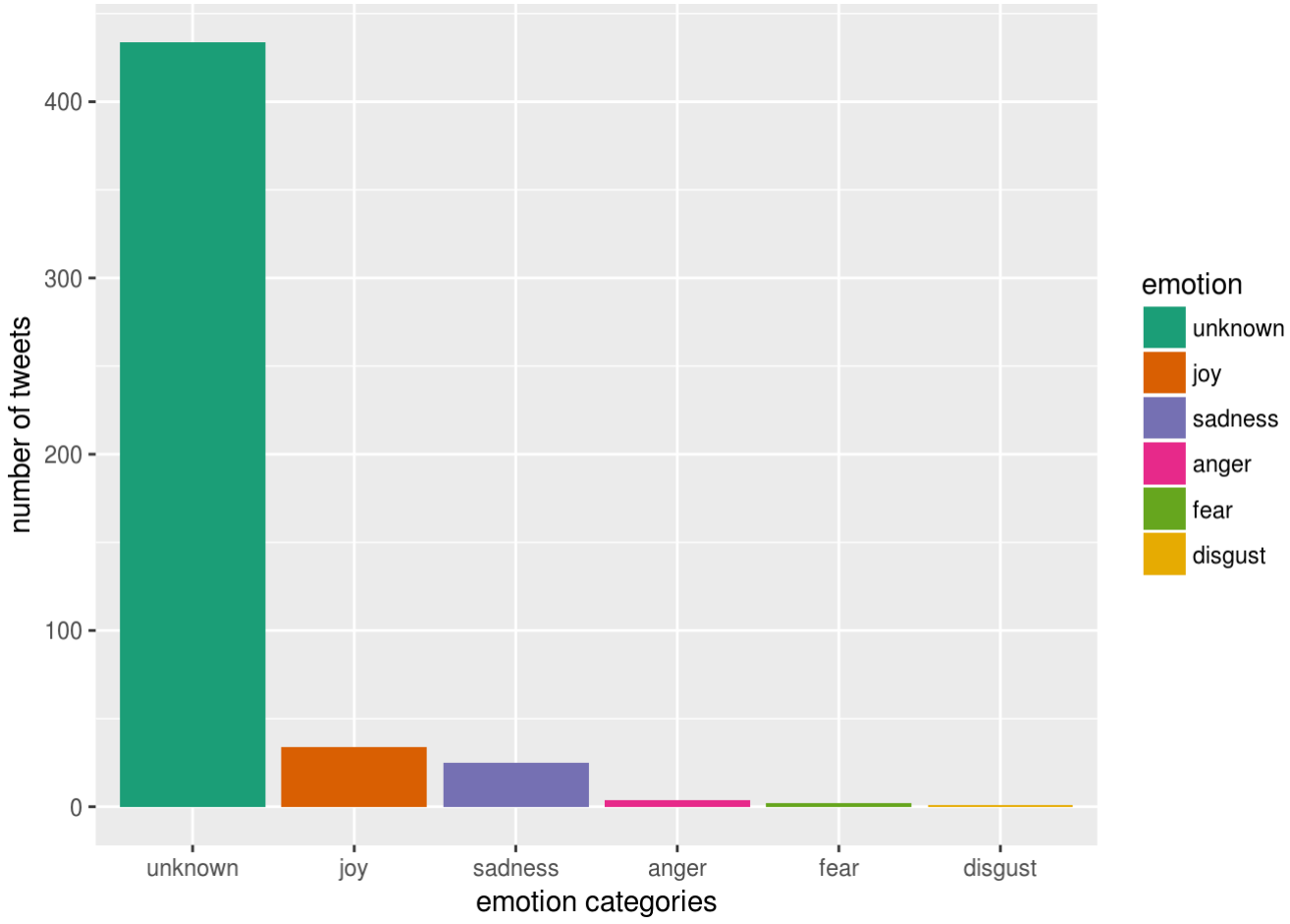
```
tweets_txt = cleantweets(tweets_txt)
class_emo = classify_emotion(tweets_txt, algorithm="bayes", prior=1.0)
emotion = class_emo[,7]
emotion[is.na(emotion)] = "unknown"

class_pol = classify_polarity(tweets_txt, algorithm="bayes")
polarity = class_pol[,4]

sentiment_df = data.frame(text=tweets_txt, emotion=emotion, polarity=polarity, string
sAsFactors=FALSE)
sentiment_df = within(sentiment_df, emotion <- factor(emotion, levels=names(sort(tabl
e(emotion), decreasing=TRUE))))
```

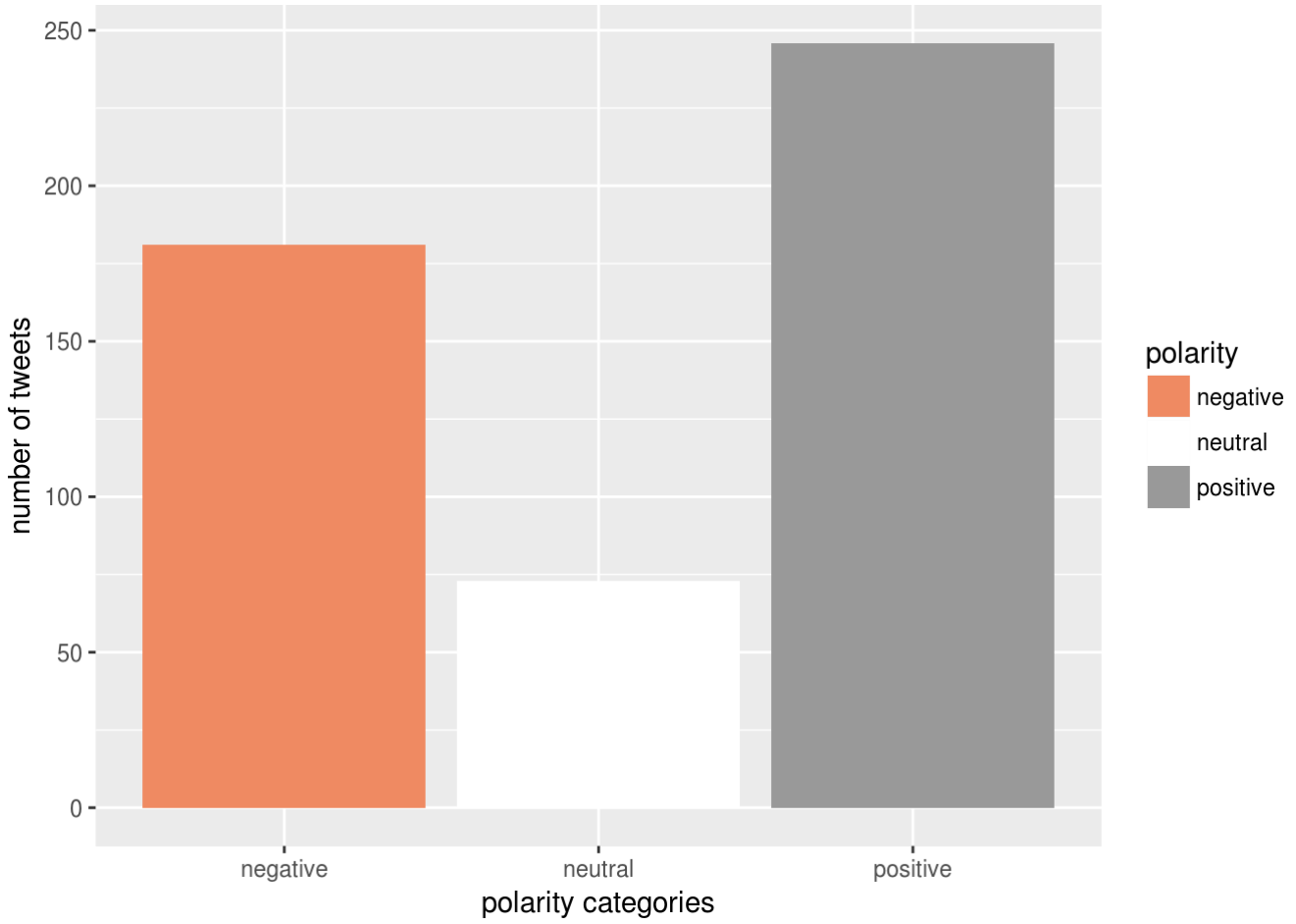
Plots

```
ggplot(sentiment_df, aes(x=emotion)) +
  geom_bar(aes(y=..count.., fill=emotion)) +
  scale_fill_brewer(palette="Dark2") +
  labs(x="emotion categories", y="number of tweets")
```



It shows that voters in UP have both emotions, joy as well as sadness because of demonetisation.

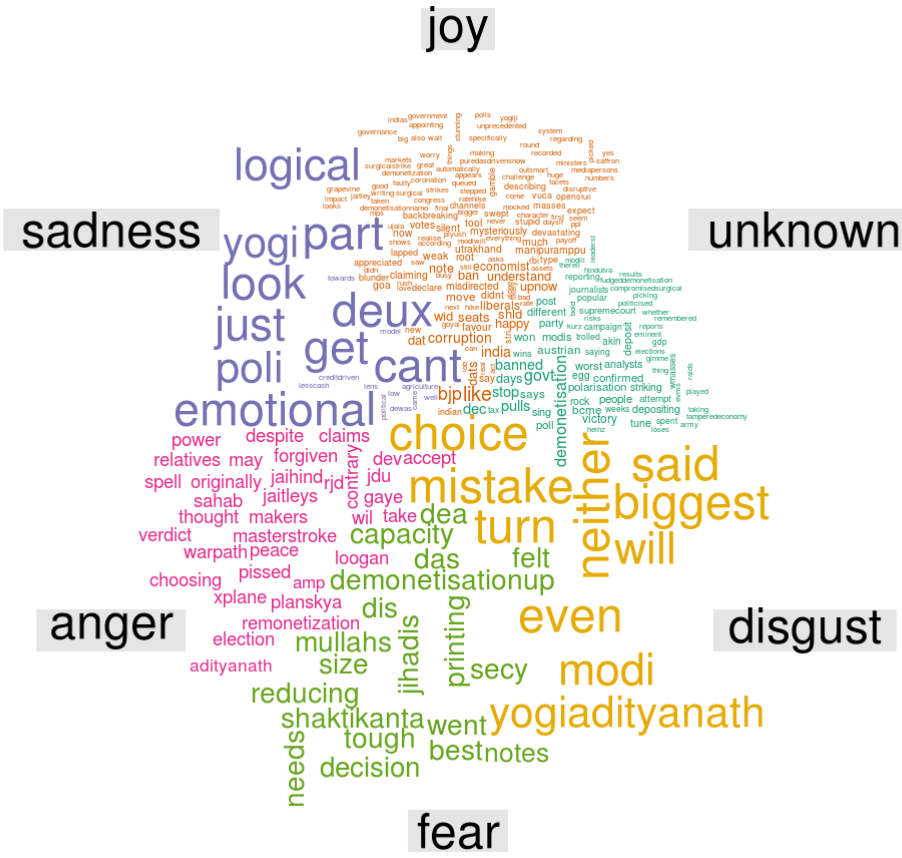
```
ggplot(sentiment_df, aes(x=polarity)) +
  geom_bar(aes(y=..count.., fill=polarity)) +
  scale_fill_brewer(palette="RdGy") +
  labs(x="polarity categories", y="number of tweets")
```



It further shows the positive opinions of people on demonetisation. This time we get same result from both analysis. As people are positive about demonetisation, we will see if BJP gains from it in our further analysis.

Word cloud

```
emos = levels(factor(sentiment_df$emotion))
nemo = length(emos)
emo.docs = rep("", nemo)
for (i in 1:nemo)
{
  tmp = tweets_txt[emotion == emos[i]]
  emo.docs[i] = paste(tmp, collapse=" ")
}
corpus = Corpus(VectorSource(emo.docs))
corpus = tm_map(corpus, removeWords, stopwords("english"))
#corpus = tm_map(corpus, stemDocument)
tdm = TermDocumentMatrix(corpus)
tdm = as.matrix(tdm)
colnames(tdm) = emos
comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"),
                 scale = c(1.5,.2), random.order = FALSE, title.size = 1.5)
```



The words such as mullah, jihadis in fear shows polarisation(i.e. hindu muslim) involved in elections. It also shows disgust of people for both the parties (modi and yogiadityanath in disgust). People are not able to make a choice because of lack of good candidates.

ANALYSIS OD ROLE OF DEMONETISATION IN WIN OF BJP

Using above code for demonetisation + UP + BJP as search word

```
tweets=searchTwitter("demonetisation+UP", n=400, lang="en", since='2017-03-01',
until='2017-03-22')
tweets_txt=sapply(tweets, function(t) t$getText())
```

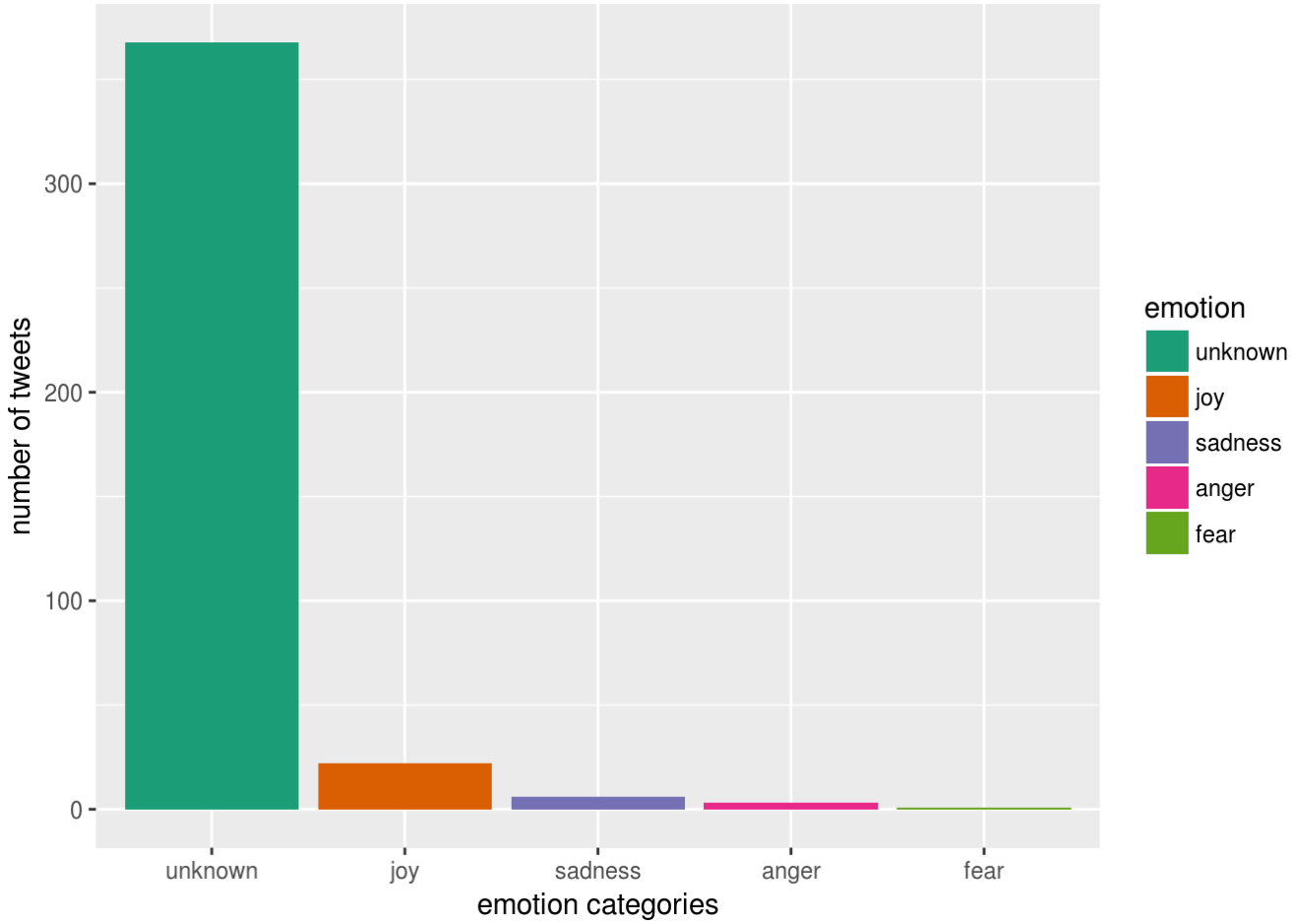
```
tweets_txt = cleantweets(tweets_txt)
class_emo = classify_emotion(tweets_txt, algorithm="bayes", prior=1.0)
emotion = class_emo[,7]
emotion[is.na(emotion)] = "unknown"

class_pol = classify_polarity(tweets_txt, algorithm="bayes")
polarity = class_pol[,4]

sentiment_df = data.frame(text=tweets_txt, emotion=emotion, polarity=polarity, string
sAsFactors=FALSE)
sentiment_df = within(sentiment_df, emotion <- factor(emotion, levels=names(sort(tabl
e(emotion), decreasing=TRUE))))
```

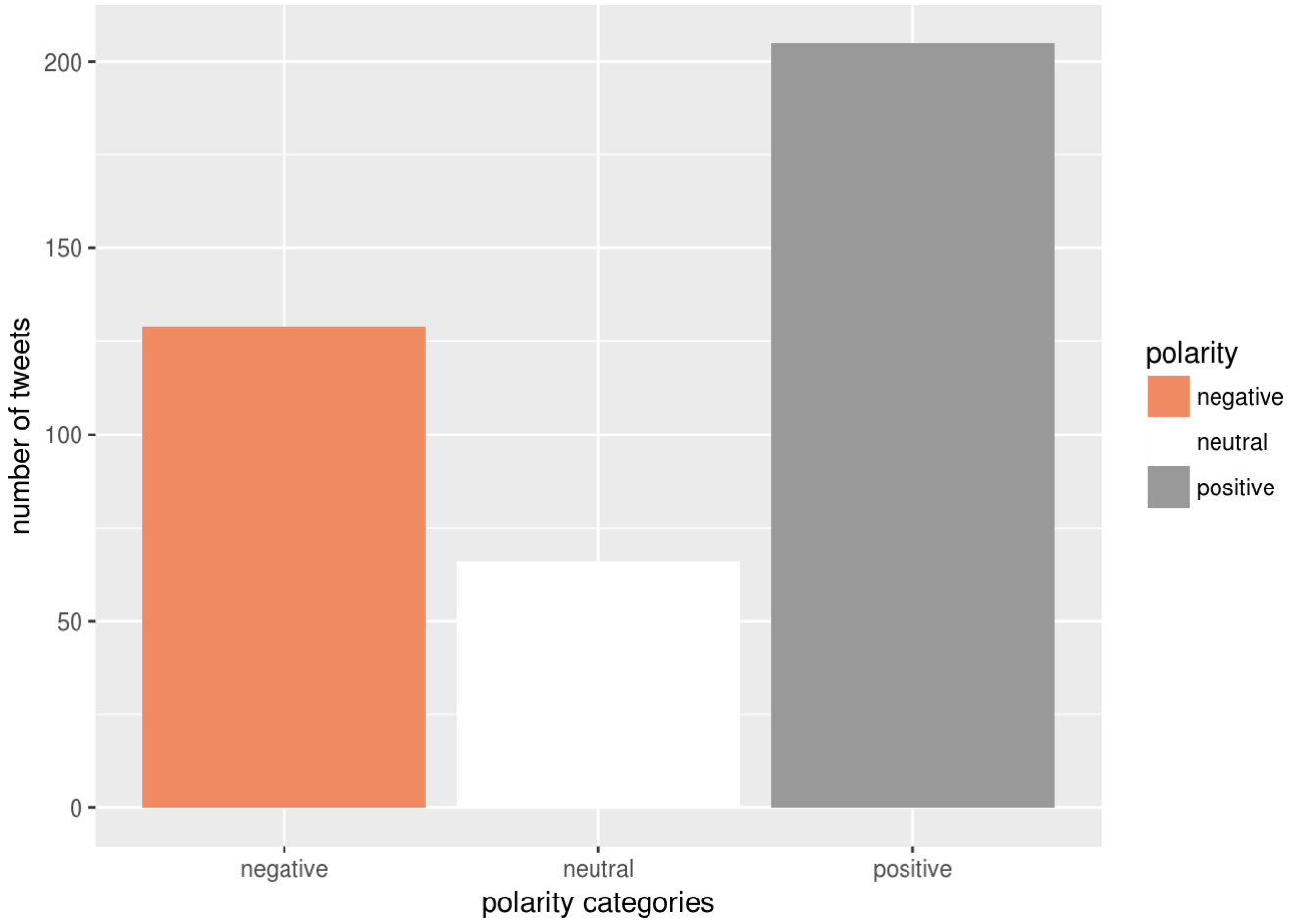
Plots

```
ggplot(sentiment_df, aes(x=emotion)) +
  geom_bar(aes(y=..count.., fill=emotion)) +
  scale_fill_brewer(palette="Dark2") +
  labs(x="emotion categories", y="number of tweets")
```



Most prominent emotion is joy. It shows that people are positive about demonetisation and BJP in UP election.

```
ggplot(sentiment_df, aes(x=polarity)) +
  geom_bar(aes(y=..count.., fill=polarity)) +
  scale_fill_brewer(palette="RdGy") +
  labs(x="polarity categories", y="number of tweets")
```



It can be observed that tweets are mostly positive. Its more positive than the analysis done on demonetisation. It shows how smartly BJP marketed demonetisation as positive to people in UP before elections. It justifies good impact of demonetisation in election for BJP.

Word cloud

```
emos = levels(factor(sentiment_df$emotion))
nemo = length(emos)
emo.docs = rep("", nemo)
for (i in 1:nemo)
{
  tmp = tweets_txt[emotion == emos[i]]
  emo.docs[i] = paste(tmp, collapse=" ")
}
corpus = Corpus(VectorSource(emo.docs))
corpus = tm_map(corpus, removeWords, stopwords("english"))
#corpus = tm_map(corpus, stemDocument)
tdm = TermDocumentMatrix(corpus)
tdm = as.matrix(tdm)
colnames(tdm) = emos
comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"),
                 scale = c(2,.35), random.order = FALSE, title.size = 1.5)
```

```
## Warning in comparison.cloud(tdm, colors = brewer.pal(nemo, "Dark2"), scale
## = c(2, : adityanath could not be fit on page. It will not be plotted.
```

Words such as remonetization, jaitley and masterstroke in anger section expresses anger of people towards finance minister Arun Jaitley due to delay in remonetization. In joy section, we can see words Uttarakhand and Goa. It shows joy among BJP supporters for win in other states as well. Thus, we can conclude that worcloud gives a good description of the situation.

NOTE: Inclusion of maps could have been done. But as most poeple in India don't have the GPS on in their devices, MAPS API doesn't give latitude and longitude of tweets.

Thus we conclude that sentiment analysis can be done to identify and extract subjective information from the source materials.