Table of Contents

Twitter :

Extracting Twitter Data:

For this Assignment, I prefer to Extract the tweets from NarendraModi, as he is one of the frequent twitter users in India.

DATA EXTRACTION PROCEDURE:

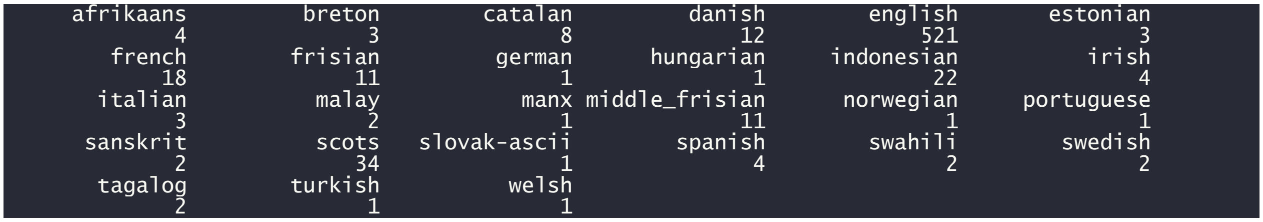
- After getting my appname, consumer_key, consumer_secret, access_token, access_secret from my developer account. I connect my twitter account using function setup_twitter_oauth from library **twitteR**.
- Using userTimeline function I extracted the top 1000 tweets from narendramodi excluding the retweets.
- Using twListToDF function I convert the extracted list of data to a proper data frame format.
- Then I saved it in my System for analysis purpose.

REMOVING HYPERLINKS AND TIME TAGS:

- In our Tweets we can see many video links are given, as well as location and time tag is present in our extracted data. So, we need to remove those words.
- To perform the desired task I used the gsub function as well as grepl .

LANGUAGE SELECTION:

- We all know that Narendramodi use Nepali, Hindi, and some time other unrecognizable languages in his tweets. I prefer to analyze the tweets, which are in pure English.
- To perform my desire result I used the function textcat from the library **textcat**.



```
      afrikaans         breton        catalan         danish        english       estonian
              4              3              8             12            521              3
         french        frisian         german      hungarian     indonesian          irish
             18             11              1              1             22              4
        italian          malay           manx middle_frisian      norwegian     portuguese
              3              2              1             11              1              1
       sanskrit          scots   slovak-ascii        spanish        swahili        swedish
              2             34              1              4              2              2
        tagalog        turkish          welsh
              2              1              1
```

- This is the classification of text cat. Maybe we can say there must be some miss classification. Among all I considered Scots and English classified languages.

LATENT DIRICHLET ALLOCATION

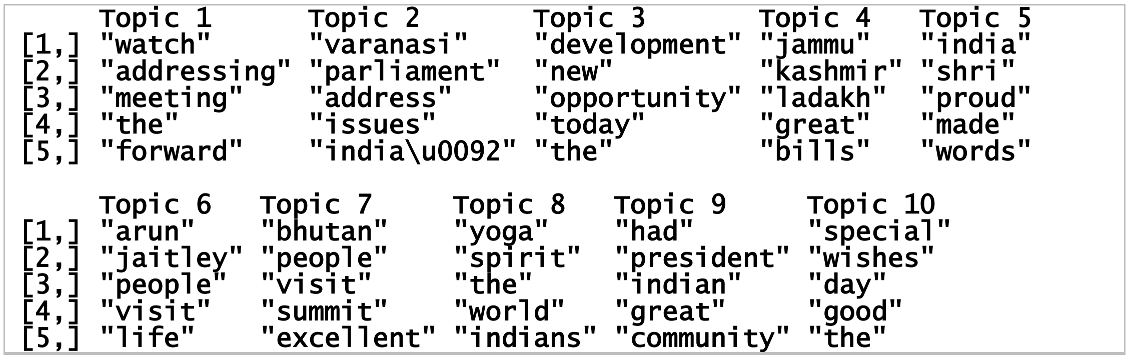
Here I make my extracted text pass through a long data cleaning procedure after converting it to a corpus.

Like removal of stop words, punctuation, numbers and whitespace using the function tm_map from the library tm.

Then I converted the cleaned Corpus to TermDocumentMetrix.

Then we considered only the words which appear more than 3 times in our TDM.

In this DocumentTermedMetrix we performed LDA and consider 5 most used terms in 10 topics.

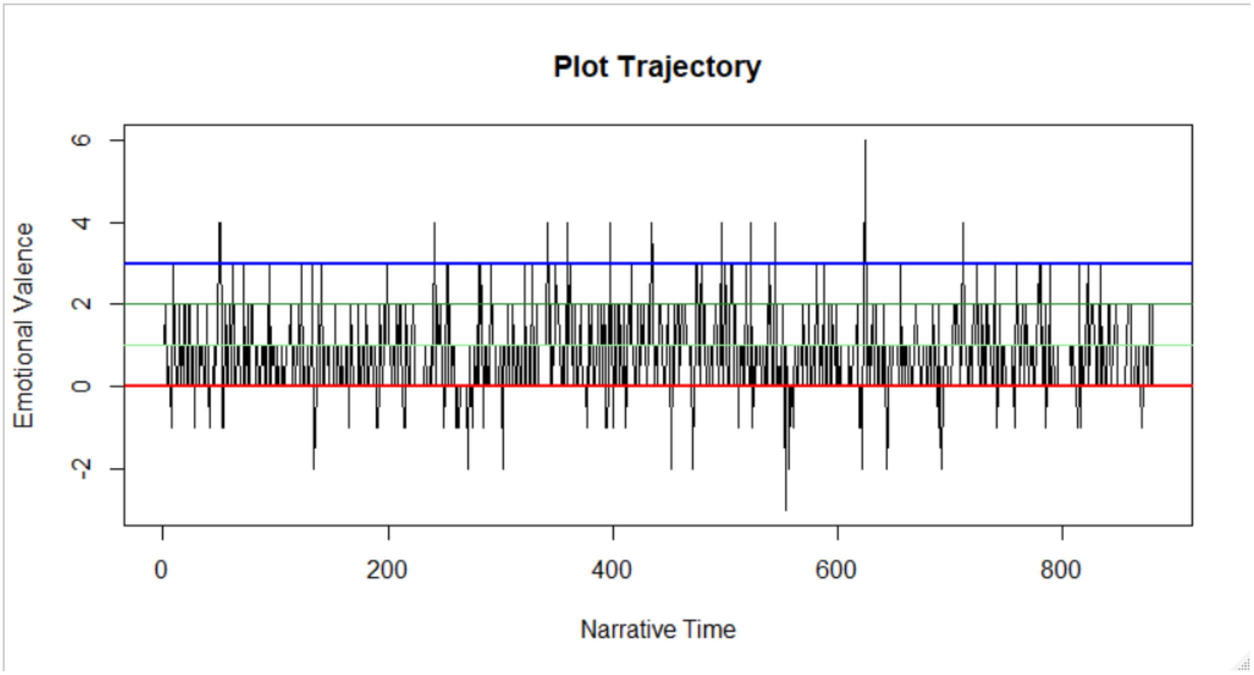


```
     Topic 1      Topic 2       Topic 3       Topic 4   Topic 5
[1,] "watch"      "varanasi"    "development" "jammu"   "india"
[2,] "addressing" "parliament"  "new"         "kashmir" "shri"
[3,] "meeting"    "address"     "opportunity" "ladakh"  "proud"
[4,] "the"        "issues"      "today"       "great"   "made"
[5,] "forward"    "india\u0092" "the"         "bills"   "words"

     Topic 6   Topic 7     Topic 8   Topic 9     Topic 10
[1,] "arun"    "bhutan"    "yoga"    "had"       "special"
[2,] "jaitley" "people"    "spirit"  "president" "wishes"
[3,] "people"  "visit"     "the"     "indian"    "day"
[4,] "visit"   "summit"    "world"   "great"     "good"
[5,] "life"    "excellent" "indians" "community" "the"
```

SENTAMENTAL ANALYSIS

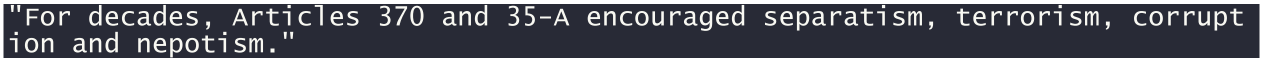
There are 6 methods for sentimental analysis. Those are "syuzhet", "afinn", "bing", "nrc","stanford", "custom".

Analyzing using the method “nrc”:



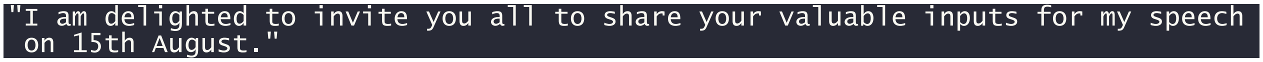
Here from the plot we can see that in his tweets majority of the sentences, contains positive vibes.

The most negative vibe from his tweet is:



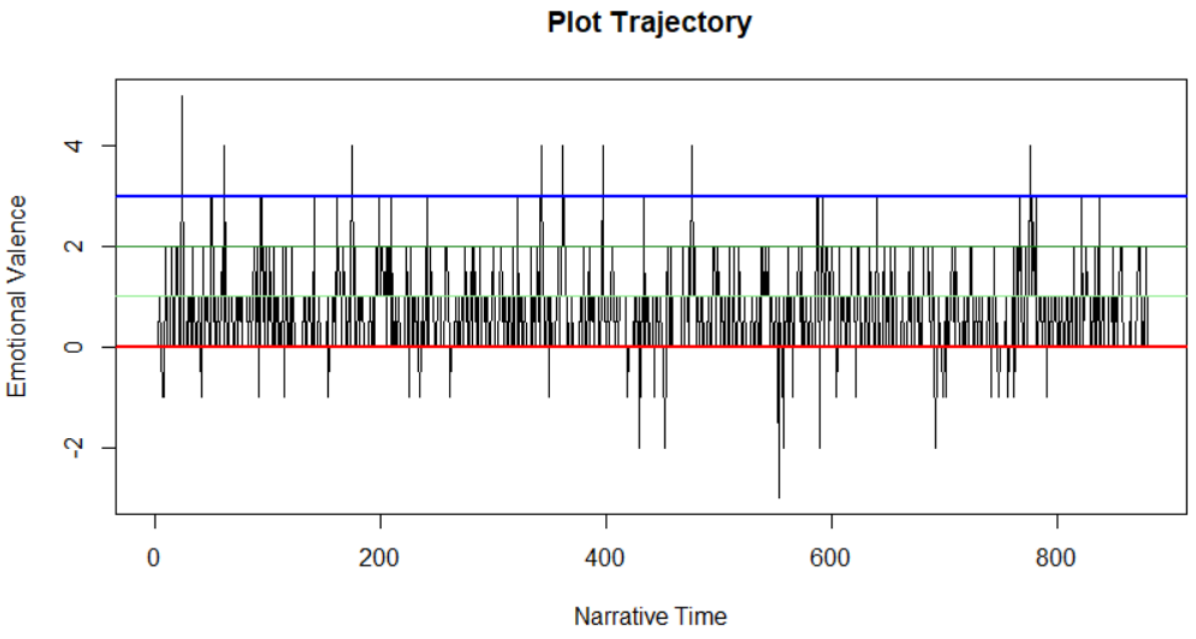
```
"For decades, Articles 370 and 35-A encouraged separatism, terrorism, corrupt
ion and nepotism."
```

The most positive vibe from his tweets is:

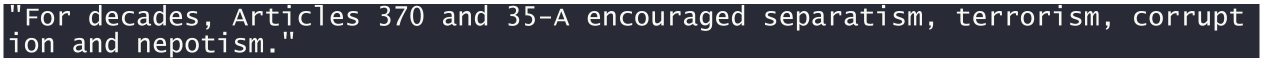


```
"I am delighted to invite you all to share your valuable inputs for my speech
 on 15th August."
```

Analyzing using the method “bing ”:

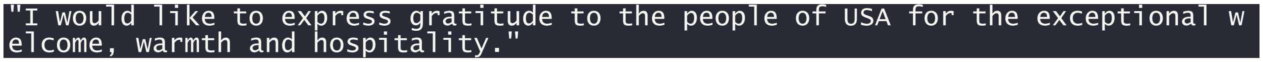


The most negative vibe from his tweets.

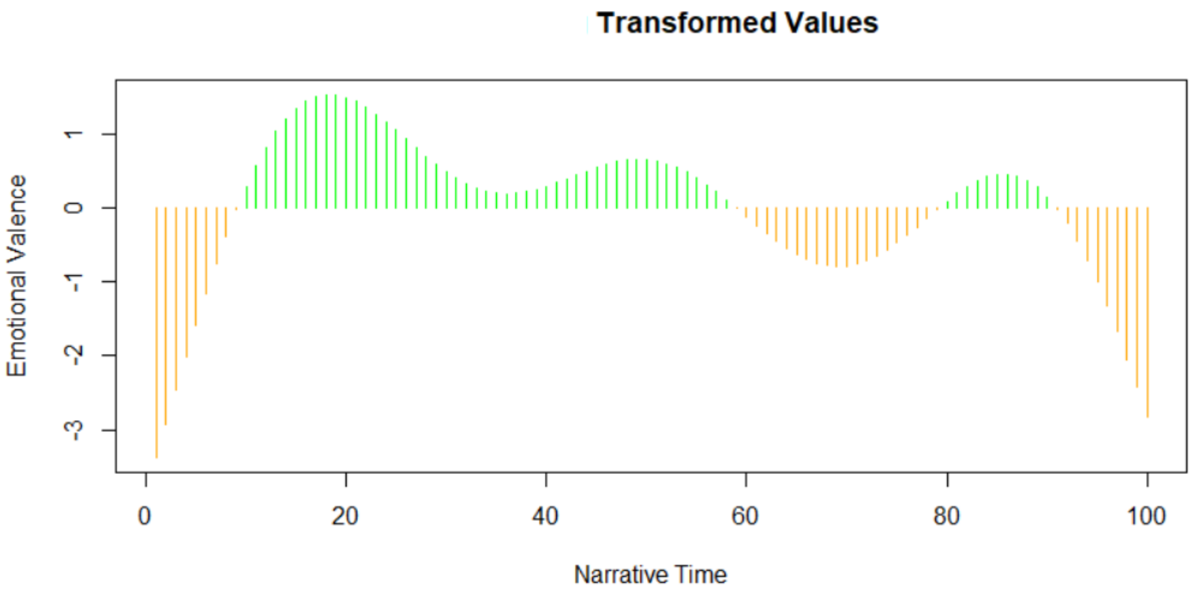


```
"For decades, Articles 370 and 35-A encouraged separatism, terrorism, corrupt
ion and nepotism."
```

The most positive vibe from his tweets.

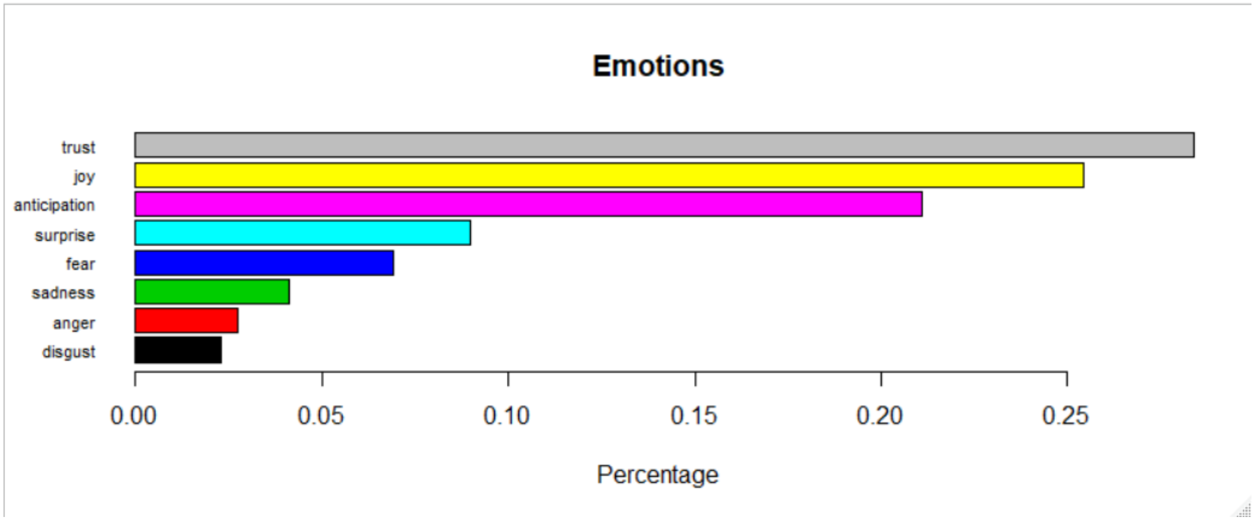


```
"I would like to express gratitude to the people of USA for the exceptional w
elcome, warmth and hospitality."
```



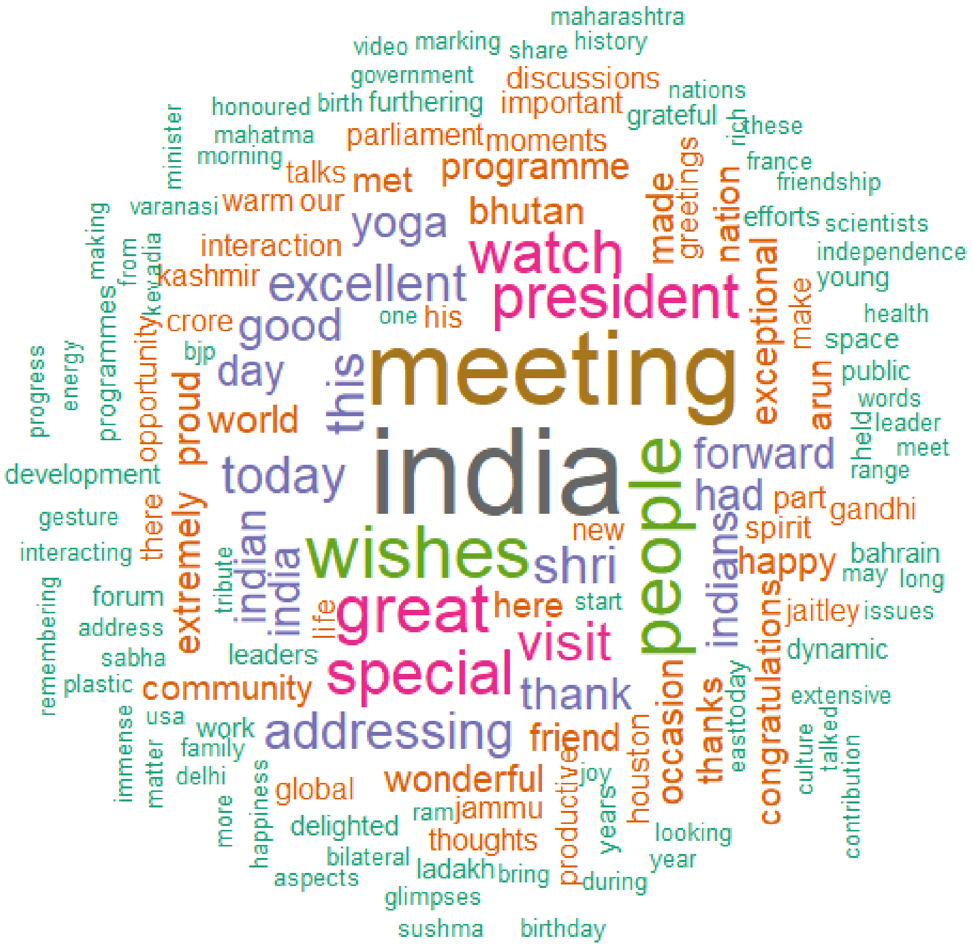
Emotion Analysis:

Now we will look for the Emotions in his tweets. There are 8 emotions recognized by our function ger_nrc_sentiments.

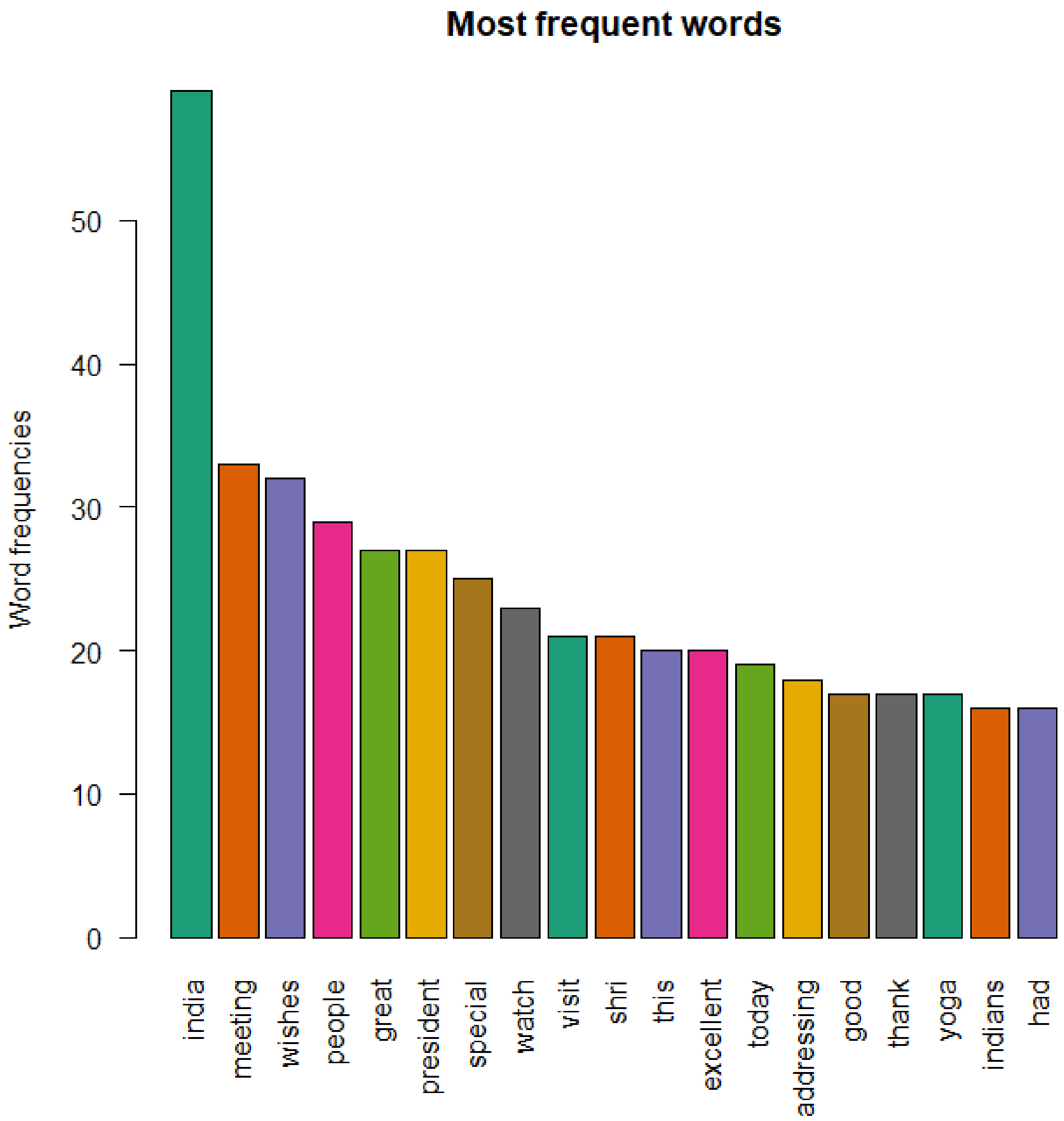


From this Bar plot we can see that, In his tweets they highly focus on public trusts.

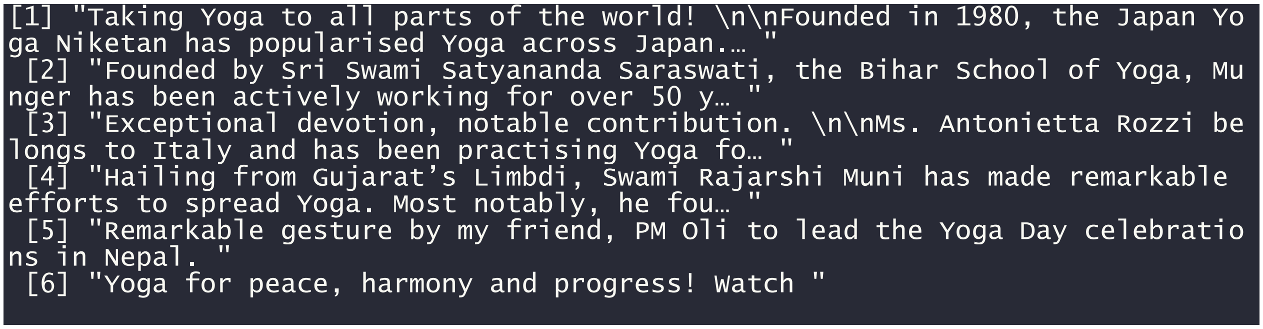
WORDCLOUD:



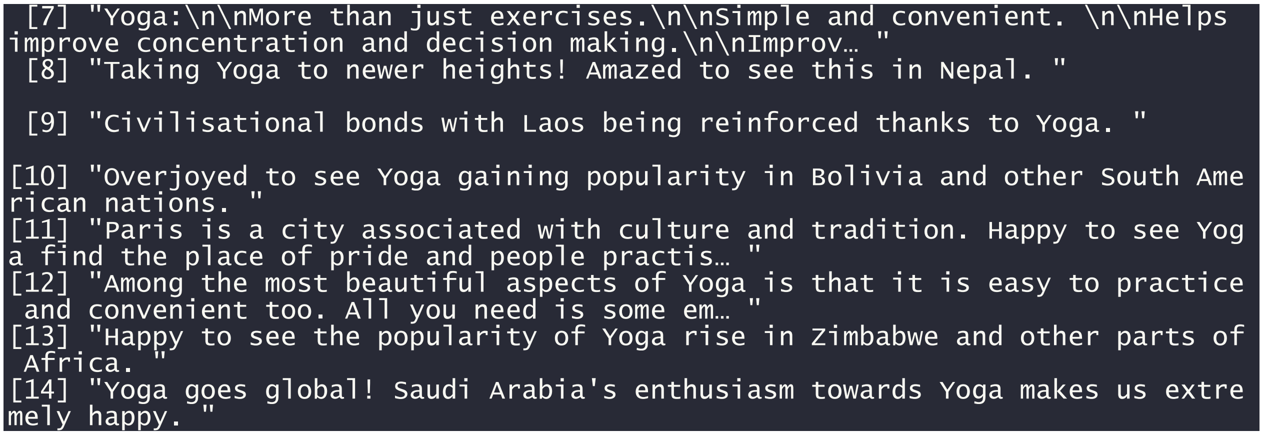
Based on this Bar plot we can get more Information regarding his frequent words in his twitter tweets.



From this I was curious why he used the word Yoga, so to I find out the tweets



```
[1] "Taking Yoga to all parts of the world! \n\nFounded in 1980, the Japan Yo
ga Niketan has popularised Yoga across Japan.… "
 [2] "Founded by Sri Swami Satyananda Saraswati, the Bihar School of Yoga, Mu
nger has been actively working for over 50 y… "
 [3] "Exceptional devotion, notable contribution. \n\nMs. Antonietta Rozzi be
longs to Italy and has been practising Yoga fo… "
 [4] "Hailing from Gujarat’s Limbdi, Swami Rajarshi Muni has made remarkable
efforts to spread Yoga. Most notably, he fou… "
 [5] "Remarkable gesture by my friend, PM Oli to lead the Yoga Day celebratio
ns in Nepal. "
 [6] "Yoga for peace, harmony and progress! Watch "
```



```
 [7] "Yoga:\n\nMore than just exercises.\n\nSimple and convenient. \n\nHelps
improve concentration and decision making.\n\nImprov… "
 [8] "Taking Yoga to newer heights! Amazed to see this in Nepal. "

 [9] "Civilisational bonds with Laos being reinforced thanks to Yoga. "

[10] "Overjoyed to see Yoga gaining popularity in Bolivia and other South Ame
rican nations. "
[11] "Paris is a city associated with culture and tradition. Happy to see Yog
a find the place of pride and people practis… "
[12] "Among the most beautiful aspects of Yoga is that it is easy to practice
 and convenient too. All you need is some em… "
[13] "Happy to see the popularity of Yoga rise in Zimbabwe and other parts of
 Africa. "
[14] "Yoga goes global! Saudi Arabia's enthusiasm towards Yoga makes us extre
mely happy. "
```

Using the function grepl.

AMAZON

Data Extraction from Amazon.in

For my analysis, I choose the Samsumg M30 smart phone.

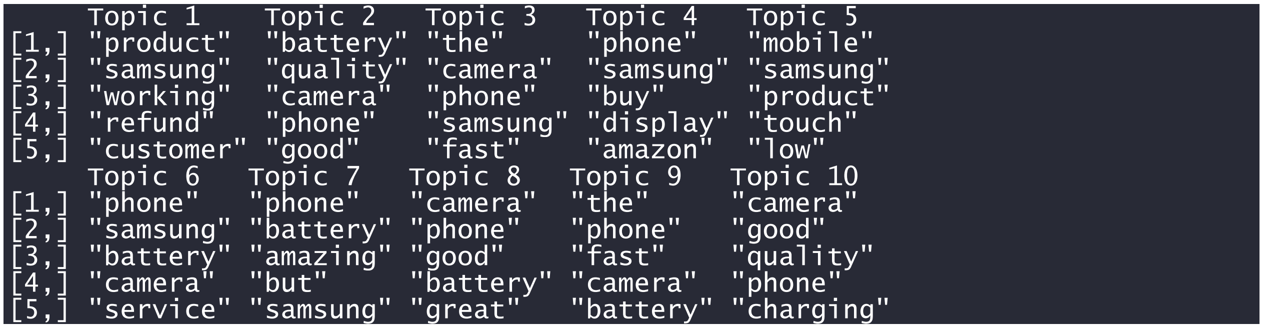
Extraction Procedure

- In Amazon.in I search for the product and see some of his review.
- After that I search for particular html_node for the reviews, here I used the extension of opera which is nothing but an Add on called **SelectorGadget**. My node was .review-text .
- Using the library **rvest** I extracted the review for top 10 pages and save it as text format for future references.

Data Cleaning Process:

- The Whole Text data processed through same steps as done in the narendramodi tweets.
- Removing time tag, Location tag, extra hyperlinks, and other unnecessary characters.
- Considering only English textual reviews, using **textcat package and function.**
- Removing stopwords, punctuation, numbers and stripping white space using the package tm_map, using the library tm.

LATENT DIRICHLET ALLOCATION



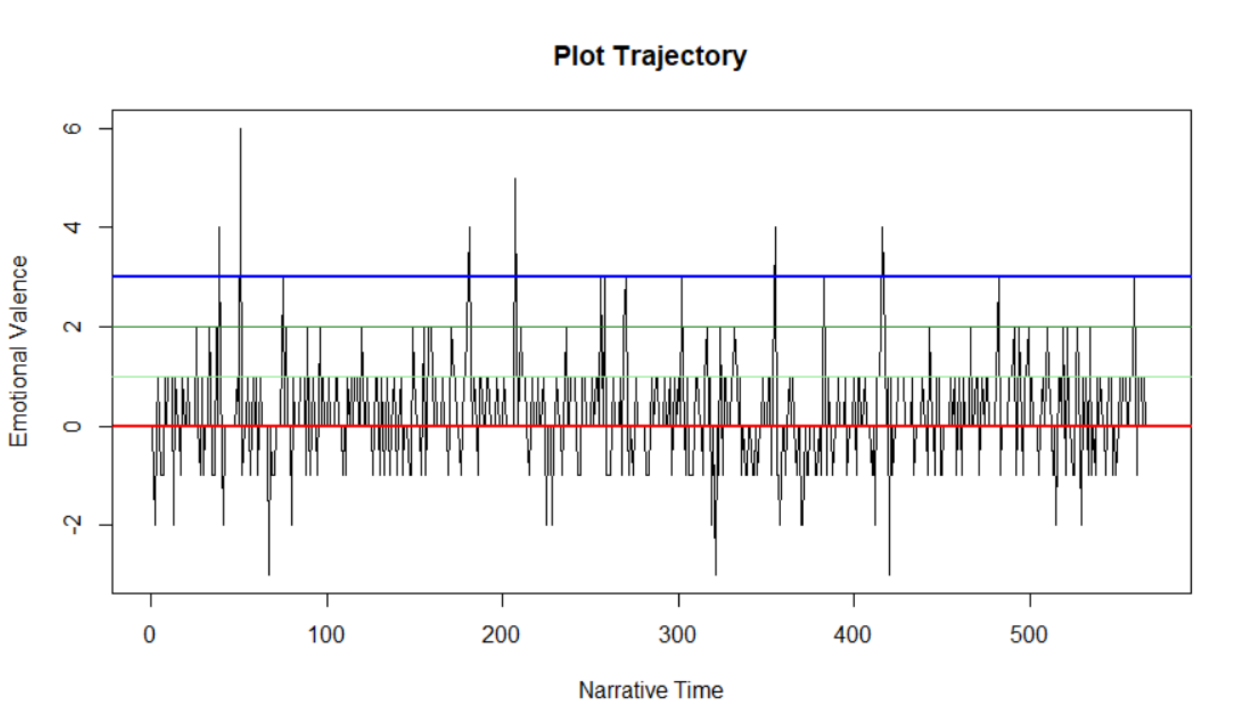
```
     Topic 1    Topic 2   Topic 3   Topic 4   Topic 5  
[1,] "product"  "battery" "the"     "phone"   "mobile" 
[2,] "samsung"  "quality" "camera"  "samsung" "samsung"
[3,] "working"  "camera"  "phone"   "buy"     "product"
[4,] "refund"   "phone"   "samsung" "display" "touch"  
[5,] "customer" "good"    "fast"    "amazon"  "low"    
     Topic 6   Topic 7   Topic 8   Topic 9   Topic 10  
[1,] "phone"   "phone"   "camera"  "the"     "camera"  
[2,] "samsung" "battery" "phone"   "phone"   "good"    
[3,] "battery" "amazing" "good"    "fast"    "quality" 
[4,] "camera"  "but"     "battery" "camera"  "phone"   
[5,] "service" "samsung" "great"   "battery" "charging"
```

Here we have considered 10 topics and 5 terms for each topics.

SENTAMENTAL ANALYSIS

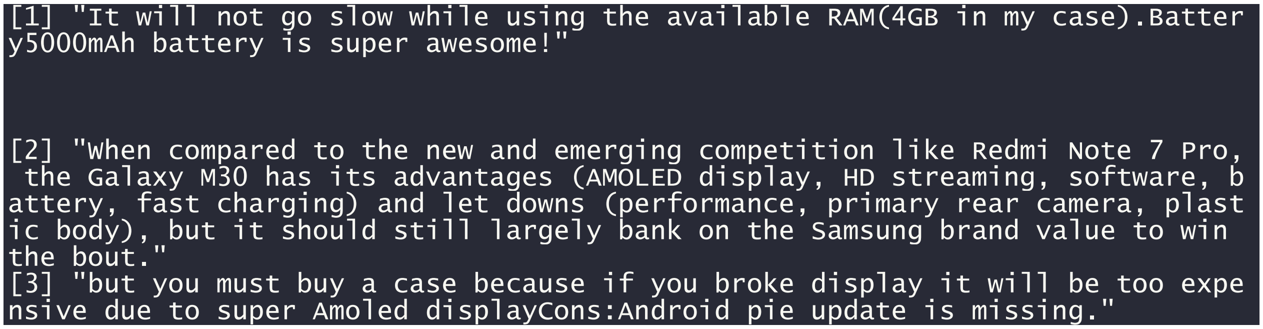
There are 6 methods for sentimental analysis. Those are "syuzhet", "afinn", "bing", "nrc","stanford", "custom".

Analyzing using the method “nrc”:



From this plot, we can clearly see that, getting a -ve or positive review is equally likely.

Lets see the most negative review according to our NLP technique.

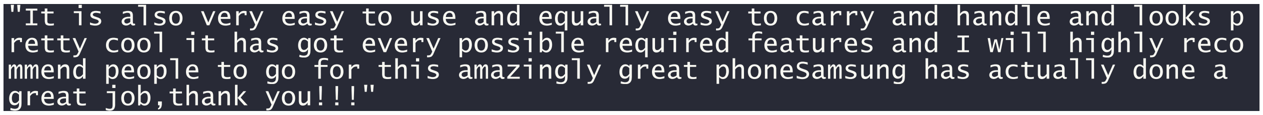


```
[1] "It will not go slow while using the available RAM(4GB in my case).Batter
y5000mAh battery is super awesome!"



[2] "When compared to the new and emerging competition like Redmi Note 7 Pro,
 the Galaxy M30 has its advantages (AMOLED display, HD streaming, software, b
attery, fast charging) and let downs (performance, primary rear camera, plast
ic body), but it should still largely bank on the Samsung brand value to win
the bout."
[3] "but you must buy a case because if you broke display it will be too expe
nsive due to super Amoled displayCons:Android pie update is missing."
```

But it seems like the review contains negative sentence like :

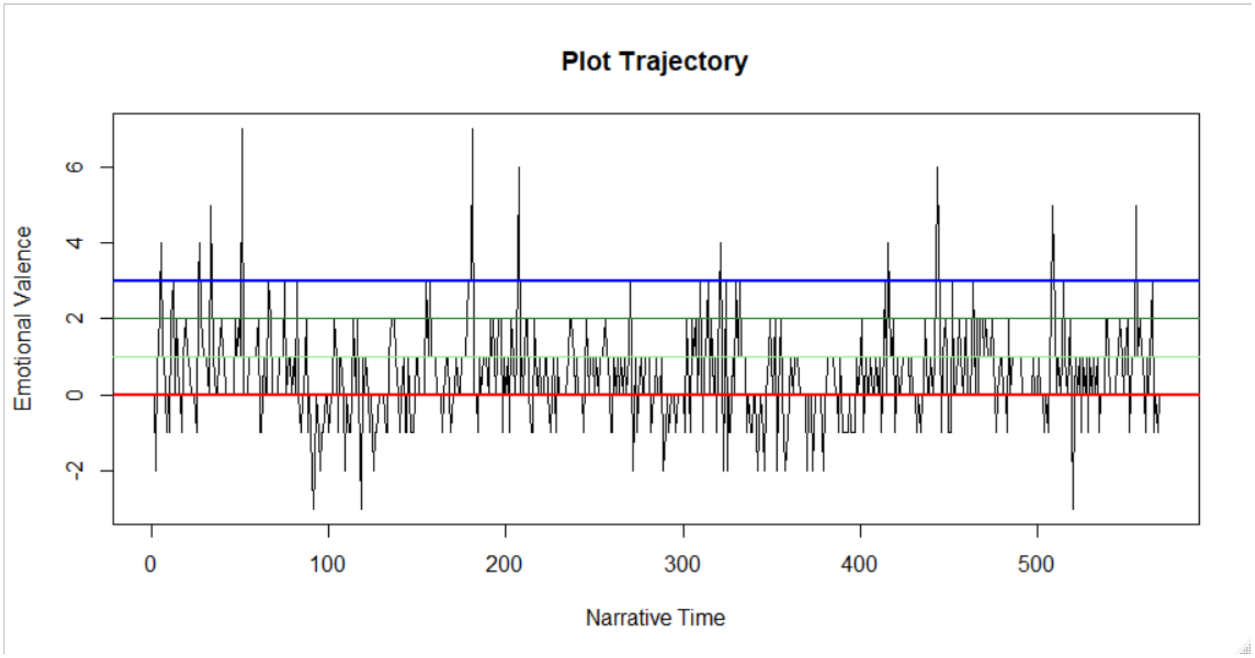
Pie update is missing, if you broke display, too expensive. Compare to . so it is choosen as negative response using our nrc method.

And positive review was



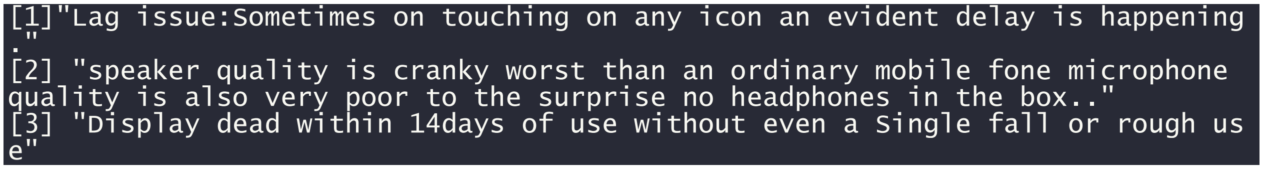
```
"It is also very easy to use and equally easy to carry and handle and looks p
retty cool it has got every possible required features and I will highly reco
mmend people to go for this amazingly great phoneSamsung has actually done a
great job,thank you!!!"
```

Analyzing using the method “bing”:



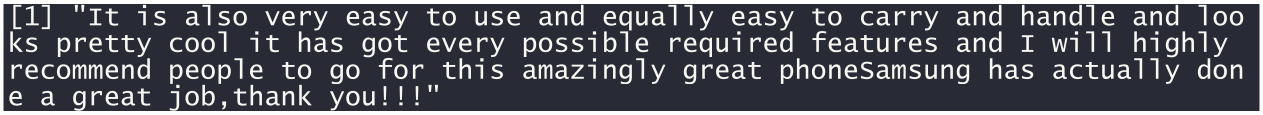
From the above plot we can say extreme positive review are more in counts as compare to extreme negative reviews.

Here are some extreme negative reviews:



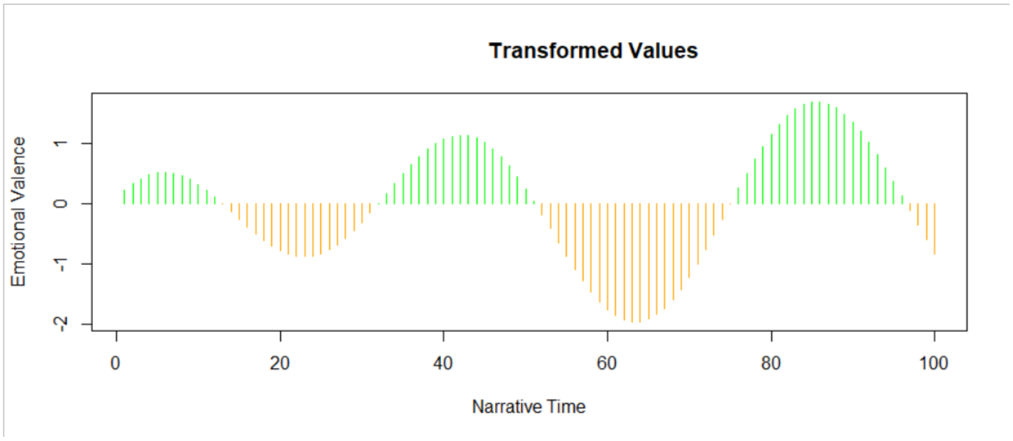
```
[1]"Lag issue:Sometimes on touching on any icon an evident delay is happening
."
[2] "speaker quality is cranky worst than an ordinary mobile fone microphone
quality is also very poor to the surprise no headphones in the box.."
[3] "Display dead within 14days of use without even a Single fall or rough us
e"
```

Here are some Extreme positive reviews:



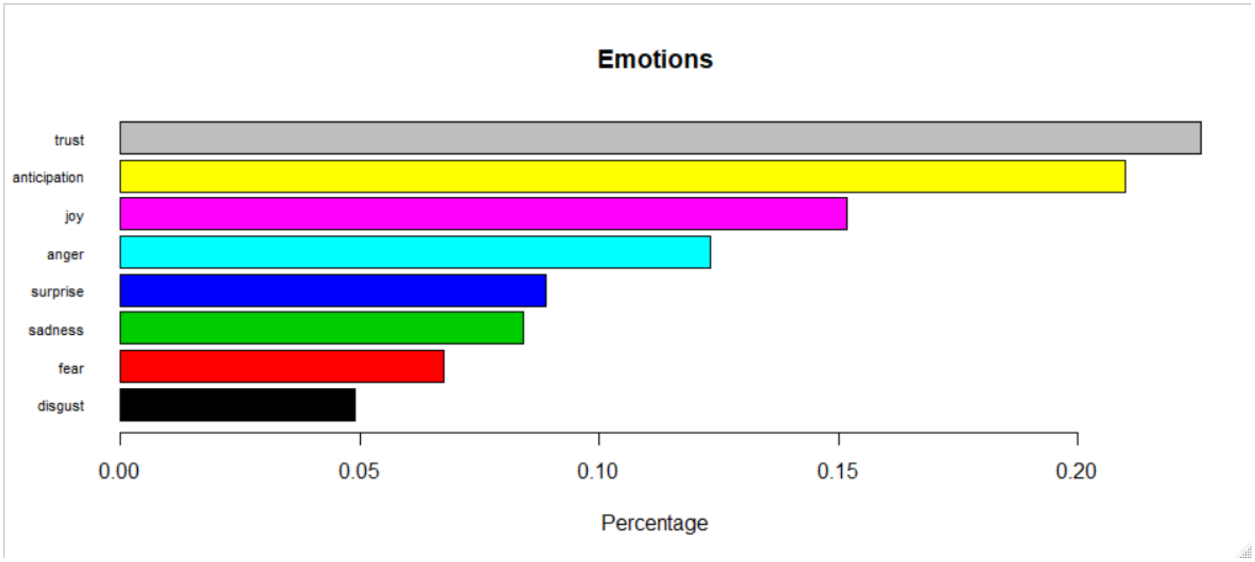
```
[1] "It is also very easy to use and equally easy to carry and handle and loo
ks pretty cool it has got every possible required features and I will highly
recommend people to go for this amazingly great phoneSamsung has actually don
e a great job,thank you!!!"
```

In our nrc as well as bing method, this particular sentence is represented as the most positive review among all.

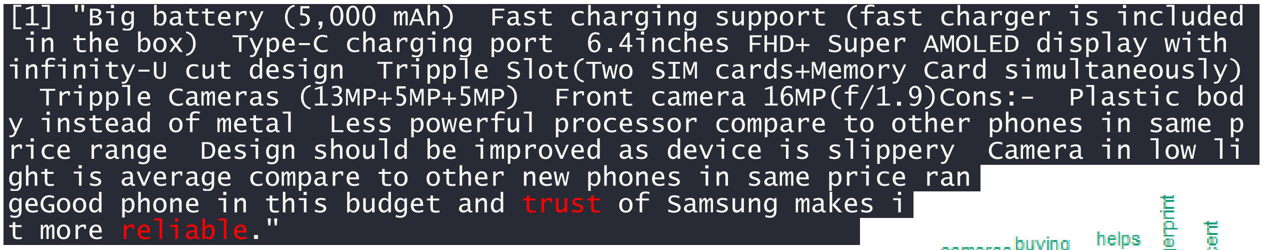


Emotion Analysis:

Now we will look for the Emotions in his tweets. There are 8 emotions recognized by our function ger_nrc_sentiments.



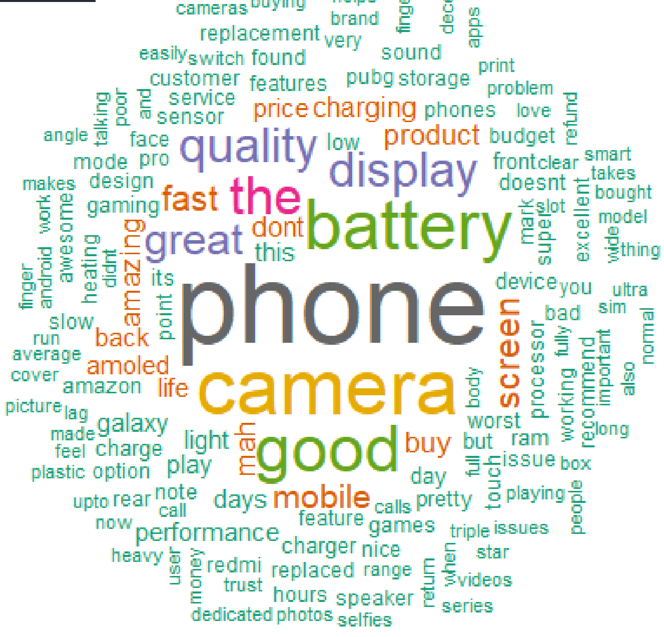
Lets have a look on the review contains the most trustworthy words.

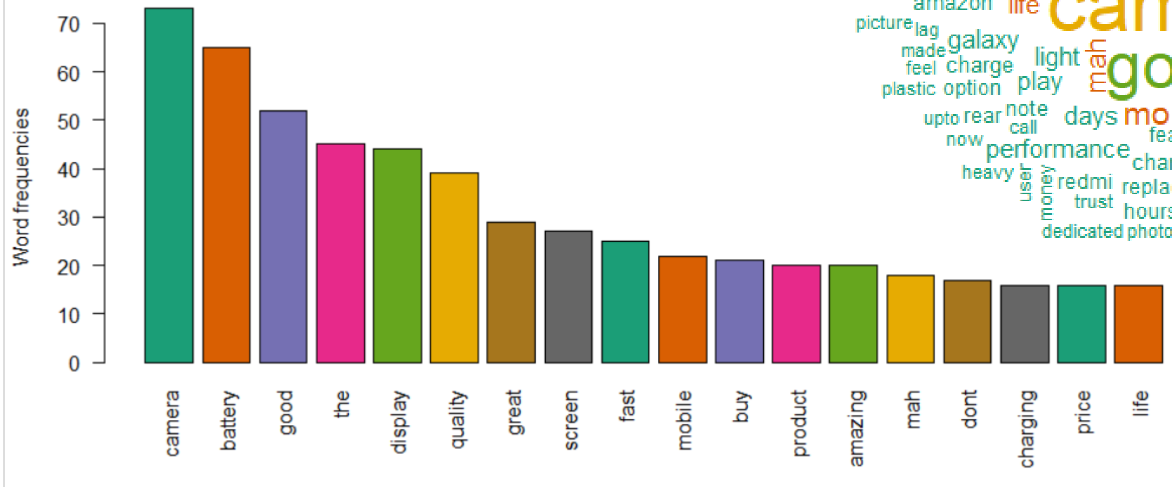


```
[1] "Big battery (5,000 mAh)  Fast charging support (fast charger is included
 in the box)  Type-C charging port  6.4inches FHD+ Super AMOLED display with
infinity-U cut design  Tripple Slot(Two SIM cards+Memory Card simultaneously)
  Tripple Cameras (13MP+5MP+5MP)  Front camera 16MP(f/1.9)Cons:-  Plastic bod
y instead of metal  Less powerful processor compare to other phones in same p
rice range  Design should be improved as device is slippery  Camera in low li
ght is average compare to other new phones in same price ran
geGood phone in this budget and trust of Samsung makes i
t more reliable."
```

WORDCLOUD:

Here From the wordcloud we can see that apart from the word phone and Camera we can see another word battery, so I am curious about that why every one are talking about battery, and came to know that, this smartphone contains 5000mah of battery life.





IMDB:

Extracting IMDB Movie Reviews, based on rating pages:

This is a 7.3 rated IMDb film, It's a American romantic dark fantasy film directed by Guillermo del Toro and written by del Toro and Vanessa Taylor. As many of my friends say that Oscar winner movies are too much boring, so I thought to take this Oscar winner movie ***The Shape Of Water*** to perform my Sentimental Analysis.

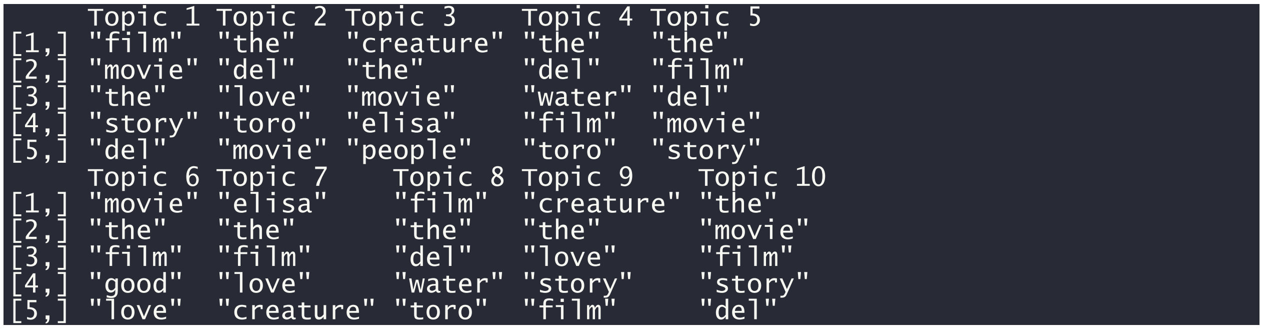
Extraction Procedure

- In IMDb I searched for The Shape Of Water to see some of his review.
- After that I search for particular html_node for the reviews, here I used the extension of opera which is nothing but an Add on called **SelectorGadget**. My node was .show-more__control.
- Using the library **rvest** I extracted the top worthful review with spoilers for all 1-10 ratings and save it as text format for future references.

Data Cleaning Process:

- The Whole Text data processed through same steps as done in the previous extracted reviews from Samsung M30.
- Removing time tag, Location tag, extra hyperlinks, and other unnecessary characters.
- Considering only English textual reviews, using **textcat package and function.**
- Removing stopwords, punctuation, numbers and stripping white space using the package tm_map, using the library tm.

LATENT DIRICHLET ALLOCATION

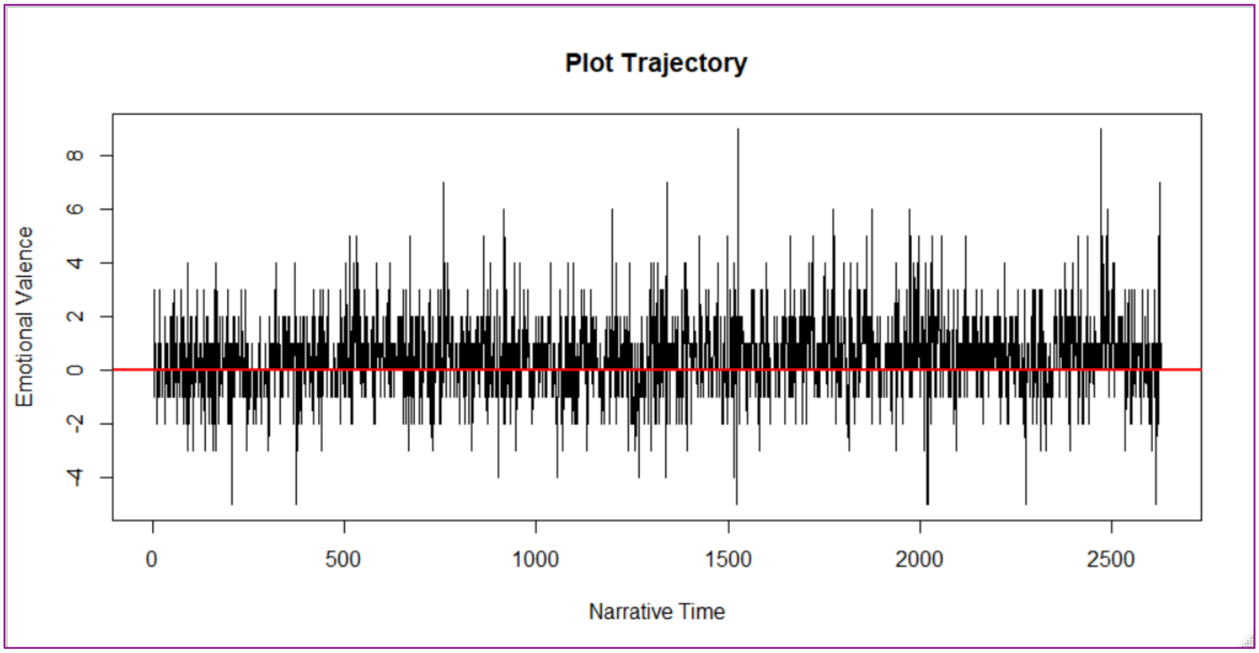


```
     Topic 1 Topic 2 Topic 3    Topic 4 Topic 5
[1,] "film"  "the"   "creature" "the"   "the"
[2,] "movie" "del"   "the"      "del"   "film"
[3,] "the"   "love"  "movie"    "water" "del"
[4,] "story" "toro"  "elisa"    "film"  "movie"
[5,] "del"   "movie" "people"   "toro"  "story"
     Topic 6 Topic 7    Topic 8 Topic 9    Topic 10
[1,] "movie" "elisa"    "film"  "creature" "the"
[2,] "the"   "the"      "the"   "the"      "movie"
[3,] "film"  "film"     "del"   "love"     "film"
[4,] "good"  "love"     "water" "story"    "story"
[5,] "love"  "creature" "toro"  "film"     "del"
```

SENTAMENTAL ANALYSIS

There are 6 methods for sentimental analysis. Those are "syuzhet", "afinn", "bing", "nrc","stanford", "custom".

Analyzing using the method “nrc”:



Here we can see extreme +ve review cross the limit of 8, where as extreme -ve review is up to -6.

From this plot we can see extreme darker on the +ve side of y axis i.e. large numbers of +ve review on the movie as compare to -ve reviews.

Lets look on some of the extreme -ve reviews.

[1] "The story is absurd, with moments of evil from the cartoonish antagonist that are simultaneously laughable and disgusting."

[2] "Once the river is at full tide, the plan is to remove the creature from the apartment and bring it to safety, in the watery refuge.Meanwhile, the creature's dark side becomes slightly manifest when it attempts to take a bite out of a house cat (animal lovers do not despair: the creature is unsuccessful in significantly injuring the little kitty!)."

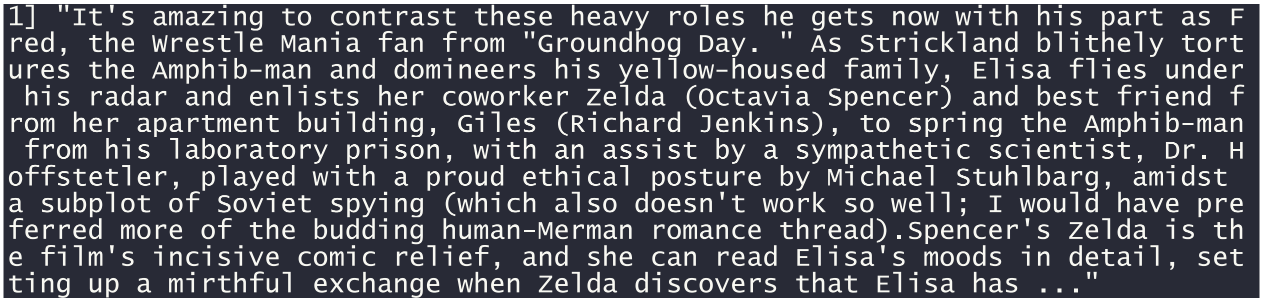
[3] "That thread is mostly window dressing, it doesn't work so well as did the fight between partisans and fascists in the Spanish Civil War context that was so deftly applied to "Pan's Labyrinth," but it serves to lay out a twisted homicidal antagonist, a true believer, Col. Strickland, played by Michael Shannon with sadistic perversion."

[4] "A mute custodian with a case of curiosity killed the cat syndrome (Sally Hawkins) becomes entangled in a tug-of-war between American and Soviet government powers after she discovers a creature being held hostage by the facility she works at, and her endlessly kind heart won't allow herself to let them use it as their pawn."

[5] "Torture, intimidation, emotional manipulation, racism, verbal and physical abuse and sexual harrassment are all employed by him, even a lack of basic hygiene is displayed as he fails to wash his hands after urinating in the first scene the audience is properly introduced to him."

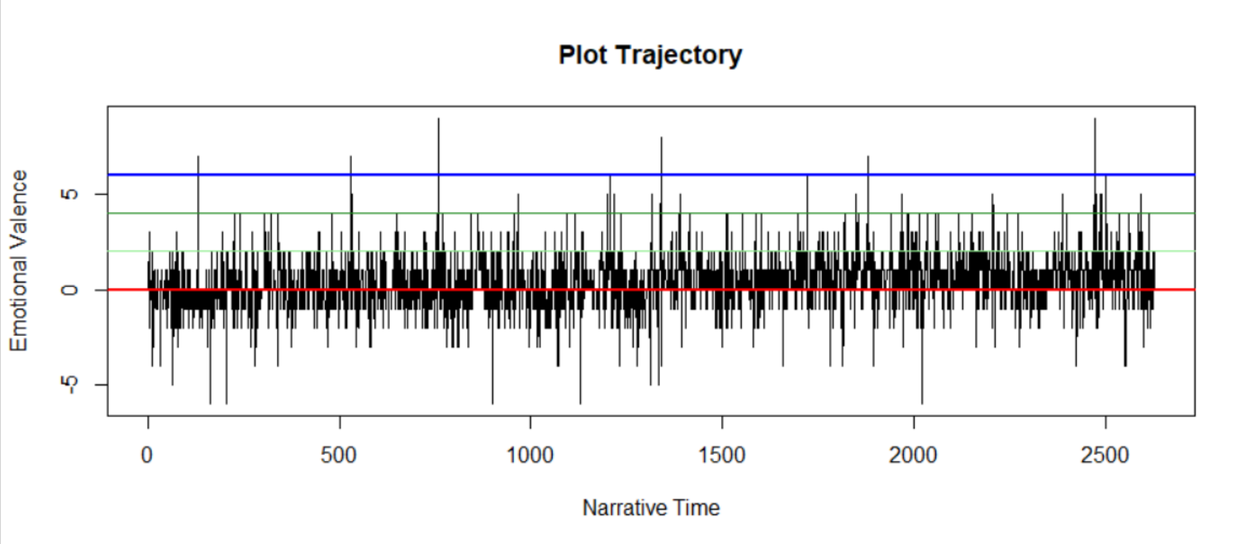
Looking at these reviews we can see that these are more like spoilers not -ve reviews. As the inside review those negative words are present.

Now Lets look at the extreme +ve reviews.



```
1] "It's amazing to contrast these heavy roles he gets now with his part as F
red, the Wrestle Mania fan from "Groundhog Day. " As Strickland blithely tort
ures the Amphib-man and domineers his yellow-housed family, Elisa flies under
 his radar and enlists her coworker Zelda (Octavia Spencer) and best friend f
rom her apartment building, Giles (Richard Jenkins), to spring the Amphib-man
 from his laboratory prison, with an assist by a sympathetic scientist, Dr. H
offstetler, played with a proud ethical posture by Michael Stuhlbarg, amidst
a subplot of Soviet spying (which also doesn't work so well; I would have pre
ferred more of the budding human-Merman romance thread).Spencer's Zelda is th
e film's incisive comic relief, and she can read Elisa's moods in detail, set
ting up a mirthful exchange when Zelda discovers that Elisa has ..."
```

Analyzing using the method "bing":



Here in this method we can see equally likely classified positive as well as -ve reviews.

Lets Have a look on the reviews with extreme negative vibes.

[1] "There are a couple of questionable violent scenes (torturing a dying man by dragging him around via a bullet wound to the cheek had a touch of the old GDT that we know and love) but the plot literally has no surprises whatsoever."

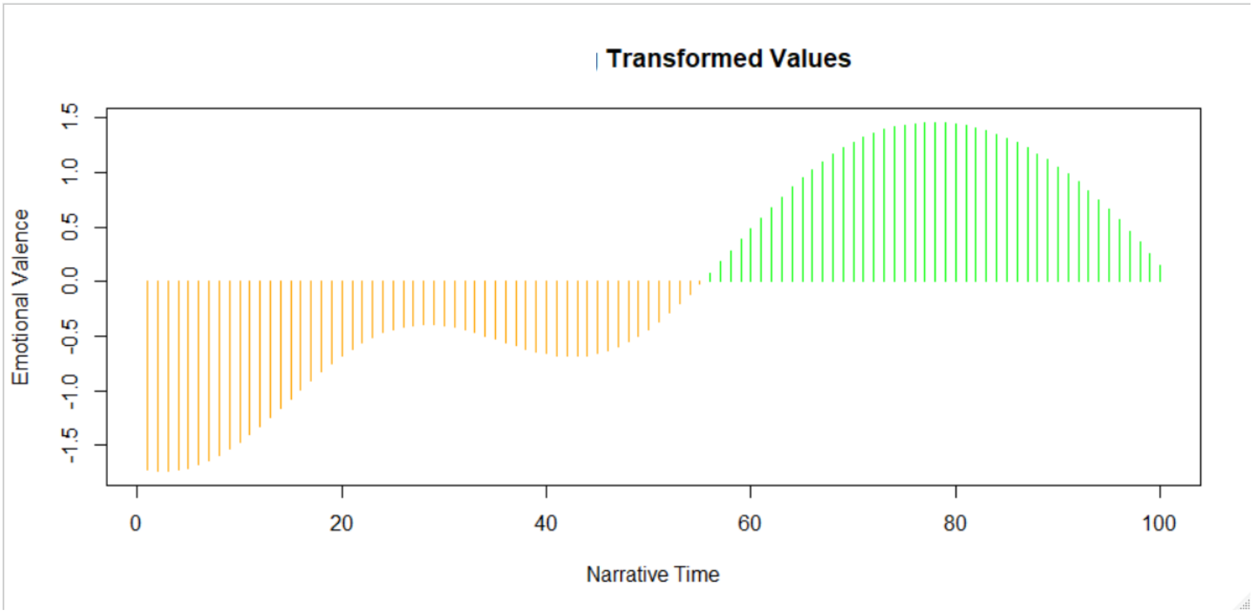
[2] "The story is absurd, with moments of evil from the cartoonish antagonist that are simultaneously laughable and disgusting."

Again, here also we are getting same reviews as we get in our previous method.

Extreme Positive review.

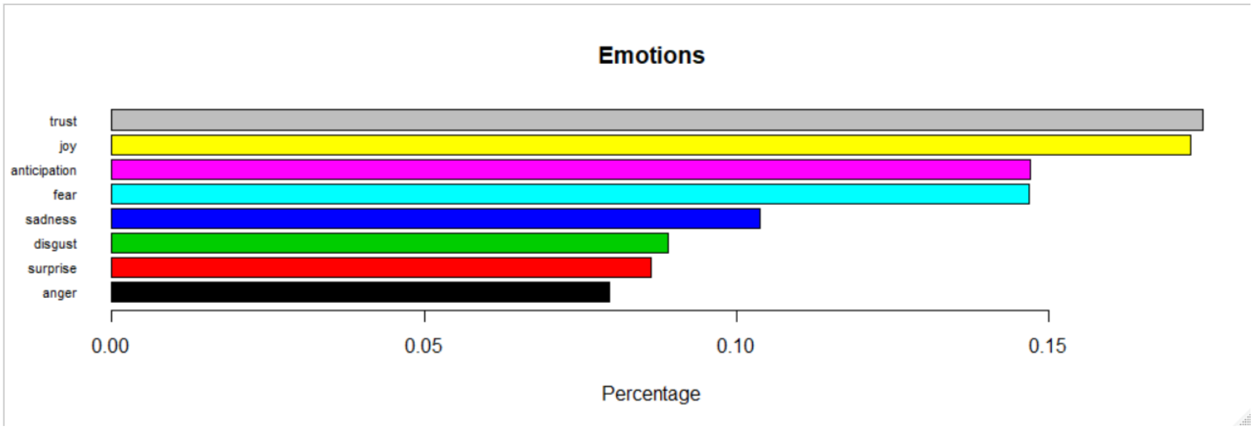
[1] "A wonderful amalgamation of sumptuous production design, lush camerawork, composed direction & excellent lead performance, The Shape of Water continues Guillermo del Toro's fascination with monsters and is an unconventional story of love that's pure, perceptive & poetic."

[2] Without false decorations, there is no doubt that "The Shape of Water" is an eye-catching drop-dead gorgeous masterpiece, Del Toro's last Magnum opus is a eerie and charming homage to ordinary-people-do-extraordinary-things stories and, without hyperbolizing my extreme love for this filmmaker- although that's just the truth -his film is one of the most clever, stirring and visually handsome experiences, a deeply moving and impressive modern-cinema classic, an artistically perfect, beautifully acted, directed and scripted impossible-love movie that I have seen in my whole life."



Emotion Analysis:

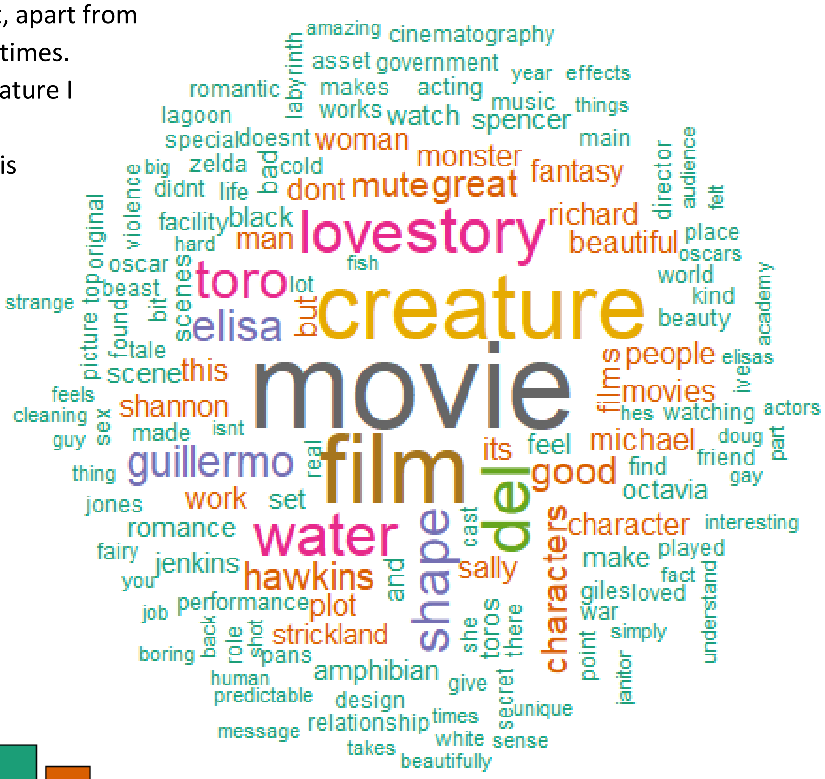
Now we will look for the Emotions in his tweets. There are 8 emotions recognized by our function ger_nrc_sentiments.

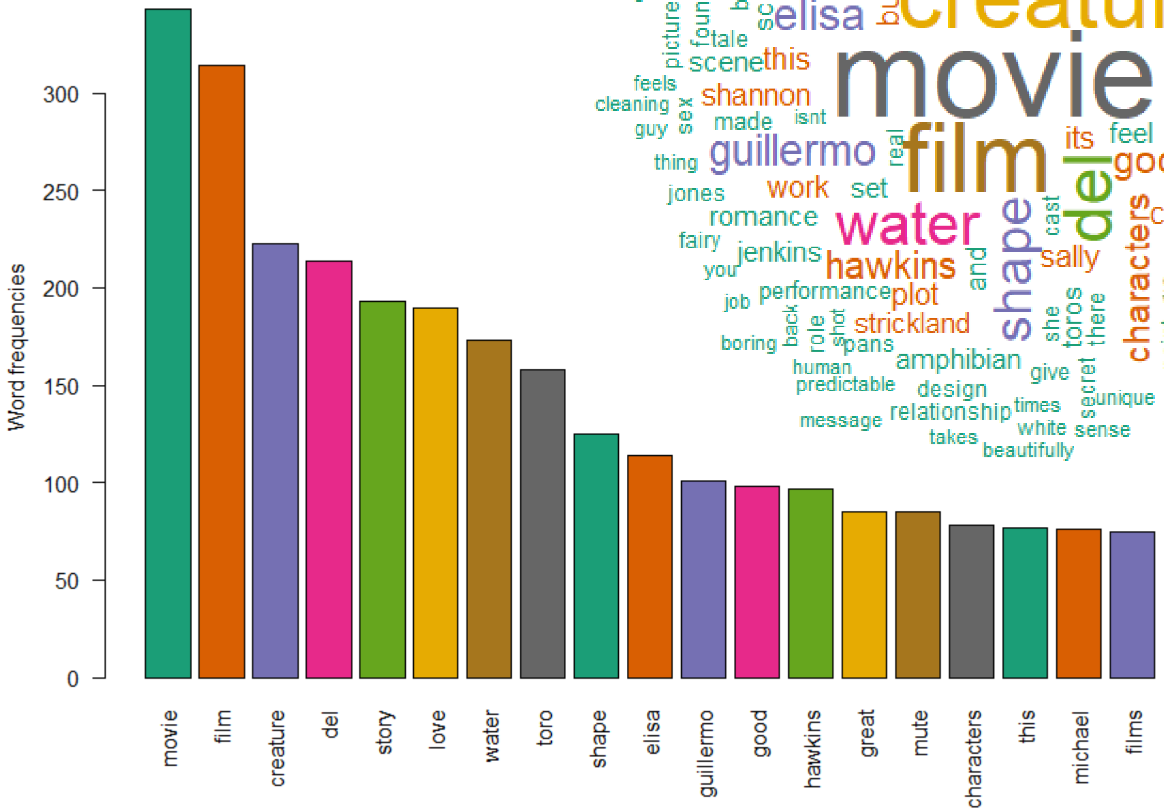


Here we can see more joy and trust category of emotions in the reviews.

WORDCLOUD:

From this Wordcloud and bar plot we can see that, apart from movie, the word creature is repeated most of the times. With some searches for the review containing Creature I came to know that, the movie is based on the Creature who lives in water. As well as this movie is based on lovestory between the Elisa and the creature so In most reply these words come up, resulting frequent used words in comment.


amazing
cinematography
labyrinth
asset
government
romantic
makes
acting
year
effects
lagoon
works
watch
music
spencer
things
special
doesnt
woman
main
big
zelda
bad
cold
monster
fantasy
director
audience
violence
didnt
life
dont
mute
great
felt
original
facility
black
richard
hard
man
lovestory
beautiful
place
oscars
oscar
fish
world
top
beast
scenes
toro
lot
kind
academy
strange
bit
elisa
but
creature
beauty
picture
found
tale
people
elisas
scene
this
films
ive
feels
movie
movies
cleaning
sex
shannon
hes
watching
actors
guy
made
isnt
michael
doug
part
real
its
feel
friend
thing
guillermo
film
good
find
gay
jones
work
set
del
octavia
romance
cast
character
interesting
fairy
water
shape
characters
make
played
jenkins
hawkins
and
sally
fact
you
giles
loved
job
performance
plot
war
boring
back
role
shot
strickland
she
toros
there
point
simply
understand
pans
janitor
human
amphibian
give
predictable
secret
design
unique
message
relationship
times
white
sense
takes
beautifully


Word frequencies
movie
film
creature
del
story
love
water
toro
shape
elisa
guillermo
good
hawkins
great
mute
characters
this
michael
films