**1.) 1) Extract tweets for any user (try choosing a user who has more tweets)**

**2) Perform sentimental analysis on the tweets extracted from the above.**

Solution:

**1.) Extract tweets from tweeter**

Extracting tweets of Akshay Kumar

**Tweets\_AkshayKumar <- userTimeline('akshaykumar', n = 1000,includeRts = T)**

1000 tweets have been extracted

Tweets extracted is in list format

So need to convert into dataframe for further analysis

**Tweets\_Akshay\_KumarDF <- twListToDF(Tweets)**

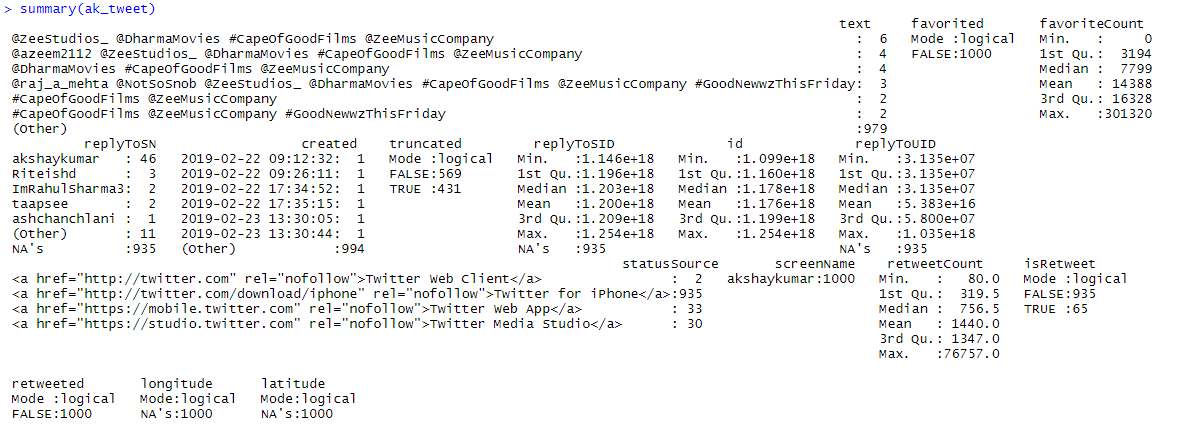
**It is saved in csv format (Tweets\_Akshay\_KumarDF.csv)**

**2.) Sentiment Analysis**

**Import Datasets**

**Tweets\_Akshay\_KumarDF.csv**

**Summary**



Only ‘text’ variable is required for analysis

Data pre processing

1. Converting to Corpus

ak\_tweet\_text <- Corpus(VectorSource(ak\_tweet\_text))



The above summarizes the corpus with 1000 documents.

2. Converting to Lower case

ak\_tweet\_text<- tm\_map(ak\_tweet\_text, tolower)

3. Removing Punctuation

ak\_tweet\_text <- tm\_map(ak\_tweet\_text, removePunctuation)

4. Removing Numbers

ak\_tweet\_text <- tm\_map(ak\_tweet\_text, removeNumbers)

5. Removing Stopwords

ak\_tweet\_text <- tm\_map(ak\_tweet\_text, removeWords, stopwords('english'))

6. Removing URL

removeURL <- function(z) gsub('http[[:alnum:]]\*', '', z)

ak\_tweet\_text <- tm\_map(ak\_tweet\_text, content\_transformer(removeURL))

7. Removing whitespace

ak\_tweet\_text <- tm\_map(ak\_tweet\_text, stripWhitespace)

8. Converting to Term Document Matrix

tdm <- TermDocumentMatrix(ak\_tweet\_text)

9. Converting to Document Term matrix

dtm <- t(tdm)

10. Converting document term matrix in matrix format

tdm <- as.matrix(tdm)

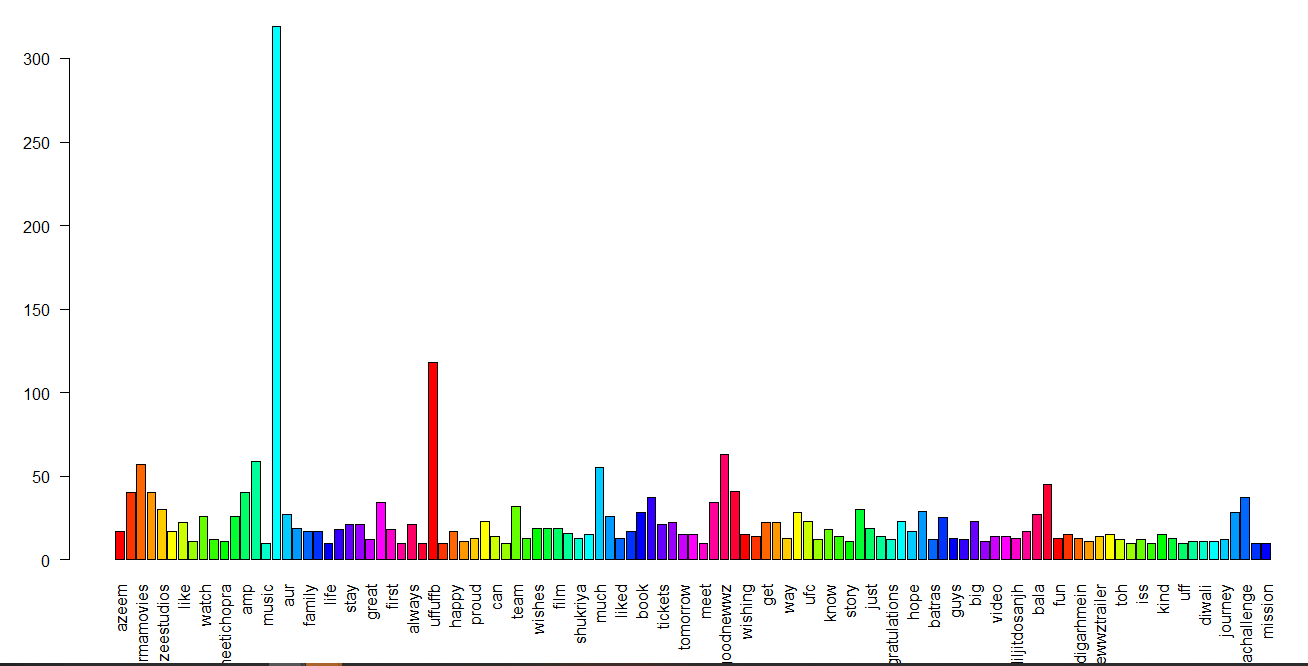
**Visualiztaion**

Lets first get the count of each entity

w <- rowSums(tdm)

Only for the words having frequency more then 10

Barplot



Few of the words can be removed as this does not contain any meaning.

# Removing menaing less word

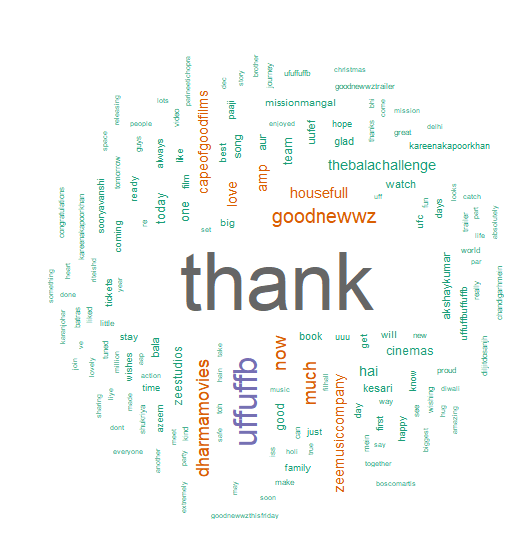
list\_word <- c("uffuffbuffuffb","diljitdosanjh\u0085", '\u0085',"ufuffuffb", "\u0092re","uuu", "riteishd")

ak\_tweet\_text <- tm\_map(ak\_tweet\_text, removeWords, list\_word)

ak\_tweet\_text <- tm\_map(ak\_tweet\_text, stripWhitespace)

**Creating the word clouds**

**wordcloud(words = names(y),freq = y, max.words = 150, random.order = F, min.freq = 5, colors=brewer.pal(8,"Dark2"), scale=c(5,0.3),rot.per = 0.7)**



The word thank is highest frequency word repeated

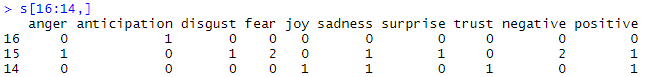
**Sentiment Analysis**

Obtaining Sentiment Scores

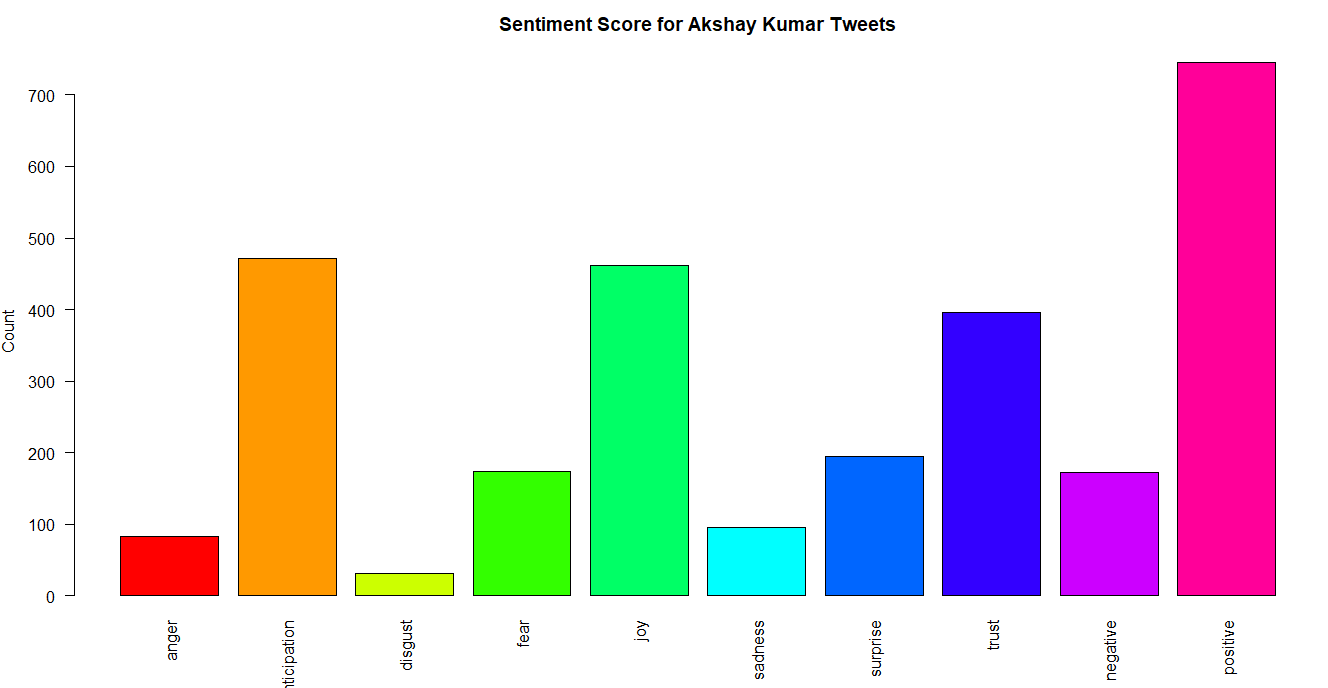
s <- get\_nrc\_sentiment(ak\_tweet\_text)

get\_nrc\_sentiment---- It is a dictionary used to calculate the presence of eight different emotions and their corresponding valence in a text file

The following is a glance of emotions and their corresponding valence from row no. 16, 15 ,14



Barplot



Form the above score it can be said that, the tweets that Akshay Kumar posted is mainly positive.

He expressed the feeling of anticipation and joy through his tweet. There is also a sense of trust.

In terms of anger and fear the his sentiments are very low.

Most Positive tweets

[1] "A very happy birthday to my friendly neighbor. Wishing you another

blockbuster year ahead @iHrithik.

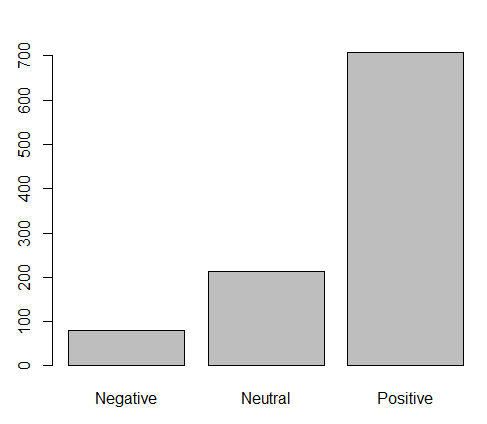
[2] "Trust you to do something so spontaneous my friend, took me

completely by surprise and a very pleasant one.

Most Negative tweets

[1] "Absolutely heartbreaking to know about the devastation by floods in Assam.All affected, humans or animals,deserve”

**Segregating positive and negative tweets**





There are 707 positive tweets, 79 negative tweets and 214 Neutral tweets.

**2.) 1) Extract reviews of any product from ecommerce website like snapdeal and amazon**

**2) Perform sentimental analysis**

**1.)** Extract reviews of any product from ecommerce website like snapdeal and amazon

# From snapdeal

**1.) Extract reviews from Snapdeal**

Product Name = Karbonn Aura (Mobile Phone)

aurl1 <- "https://www.snapdeal.com/product/karbonn-black-aura-1-worlds/663390882770/reviews?page"

aurl2 <- "&sortBy=HELPFUL#defRevPDP"

snap\_deal\_reviews <- NULL

for (i in 1:20){

print(i)

a\_link <- as.character(paste(aurl1,i,sep="="))

a\_link <- paste(a\_link,aurl2,sep="")

print(a\_link)

murl <- read\_html(a\_link)

rev <- murl %>%

html\_nodes(".head") %>%

html\_text(".head")

snap\_deal\_reviews <- c(snap\_deal\_reviews,rev)

Reviews extracted is in csv format

It is saved in csv format (karbonn\_aura.csv)

**2.) Sentiment Analysis**

**Import Datasets**

**Karbonn\_auro.csv**

Data pre processing

1. Converting to Corpus

karbonn\_aura\_text <- Corpus(VectorSource(krabonn\_aura$x))



The above summarizes the corpus with 269 documents.

2. Converting to Lower case

karbonn\_aura\_text<- tm\_map(karbonn\_aura\_text, tolower)

3. Removing Punctuation

karbonn\_aura\_text <- tm\_map(karbonn\_aura\_text, removePunctuation)

4. Removing Numbers

karbonn\_aura\_text <- tm\_map(karbonn\_aura\_text, removeNumbers)

5. Removing Stopwords

karbonn\_aura\_text <- tm\_map(karbonn\_aura\_text, removeWords, stopwords('english'))

6. Removing URL

removeURL <- function(z) gsub('http[[:alnum:]]\*', '', z)

karbonn\_aura\_text <- tm\_map(karbonn\_aura\_text, content\_transformer(removeURL))

7. Removing whitespace

karbonn\_aura\_text <- tm\_map(karbonn\_aura\_text, stripWhitespace)

8. Converting to Term Document Matrix

tdm <- TermDocumentMatrix(karbonn\_aura\_text)

9. Converting to Document Term matrix

dtm <- t(tdm)

10. Converting document term matrix in matrix format

tdm <- as.matrix(tdm)

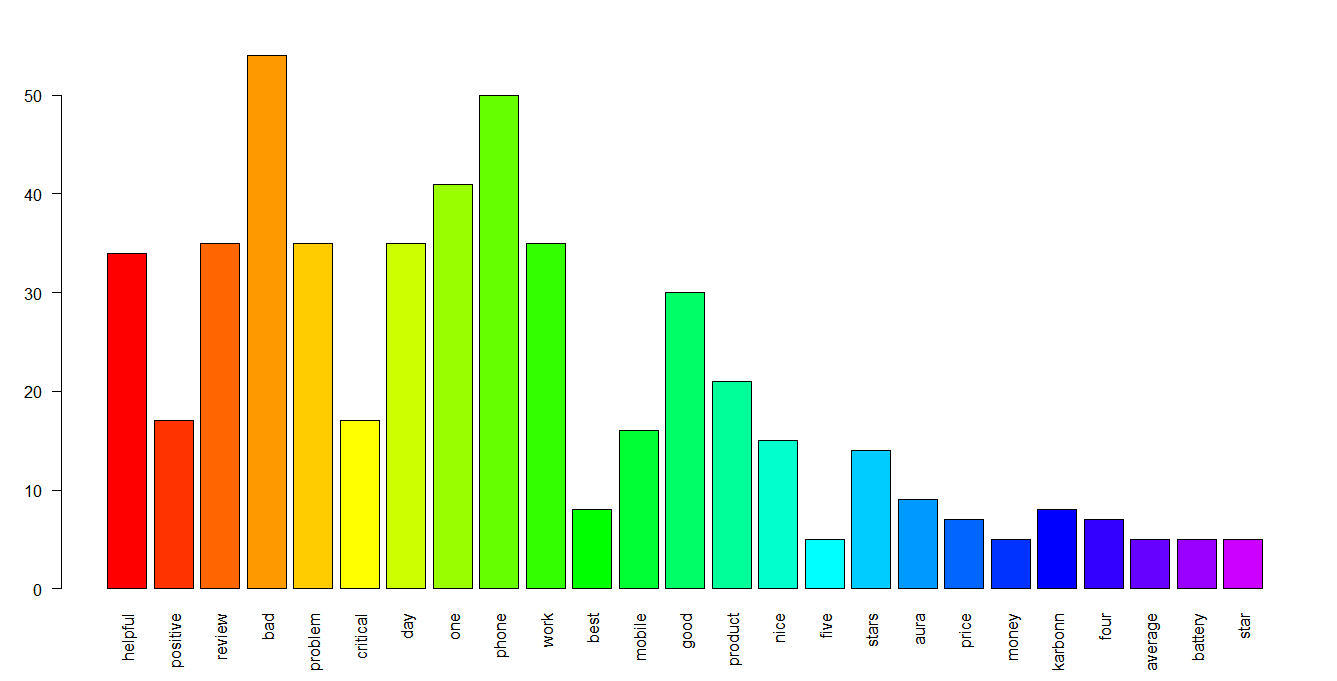
**Visualiztaion**

Lets first get the count of each entity

w <- rowSums(tdm)

Only for the words having frequency more then 5

Barplot



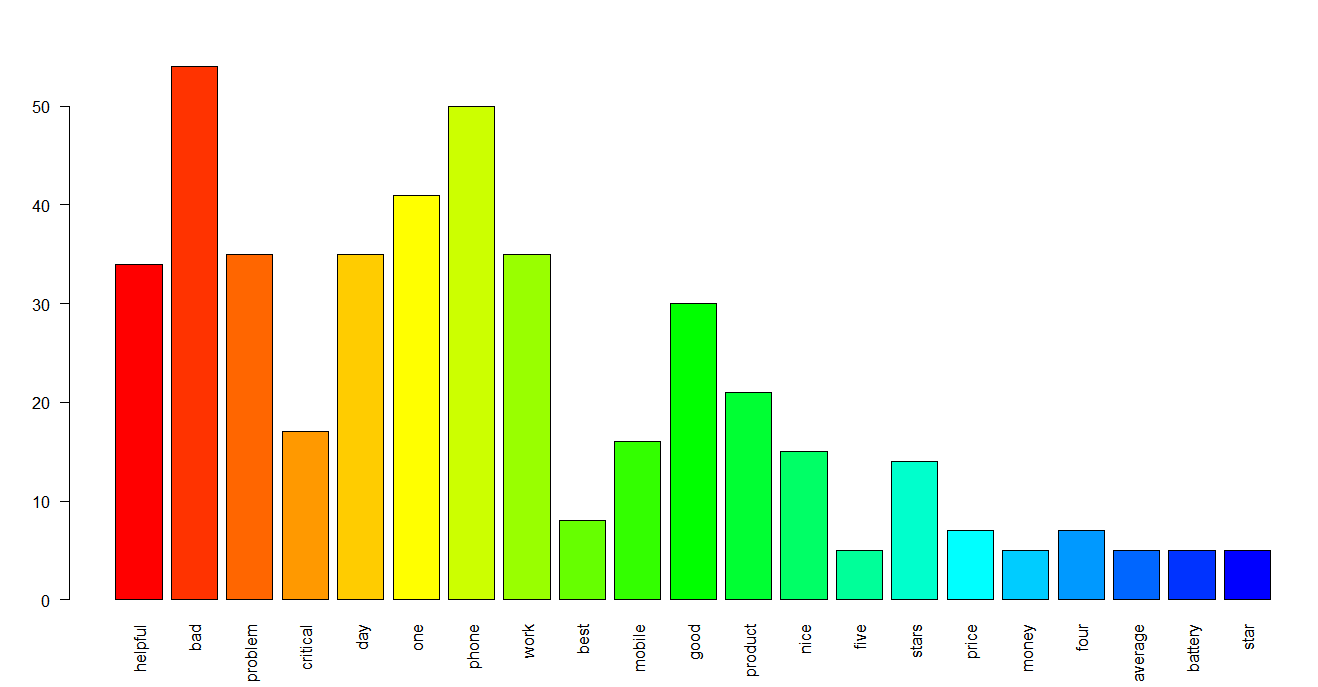
Few of the words can be removed as this does not contain any meaning.

# Removing some common word

list\_word <-c(stopwords('english'), "review","positive","karbonn","aura")

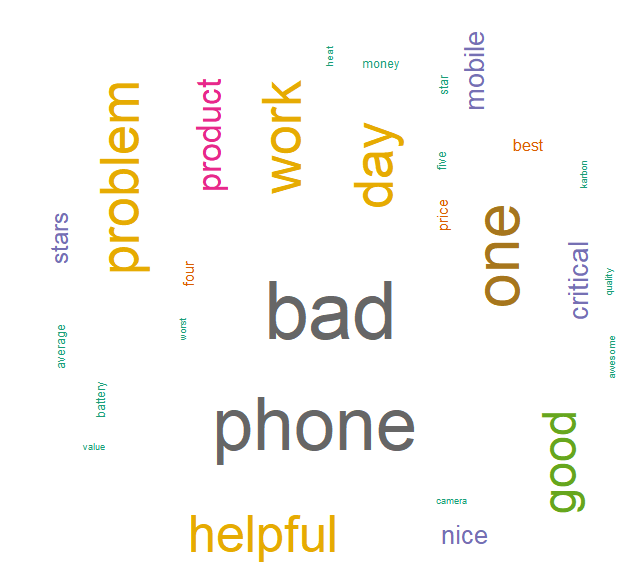
karbonn\_aura\_text <- tm\_map(karbonn\_aura\_text, stopwords(list\_word))

karbonn\_aura\_text <- tm\_map(karbonn\_aura\_text, stripWhitespace)



**Creating the word clouds**

**wordcloud(words = names(y),freq = y, max.words = 150, random.order = F, min.freq = 3, colors=brewer.pal(8,"Dark2"), scale=c(5,0.3),rot.per = 0.7)**



The word bad is highest frequency word

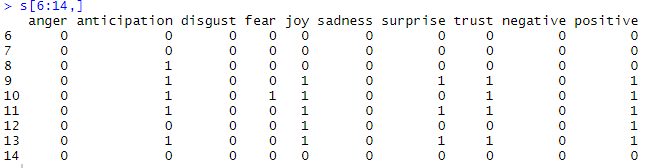
**Sentiment Analysis**

Obtaining Sentiment Scores

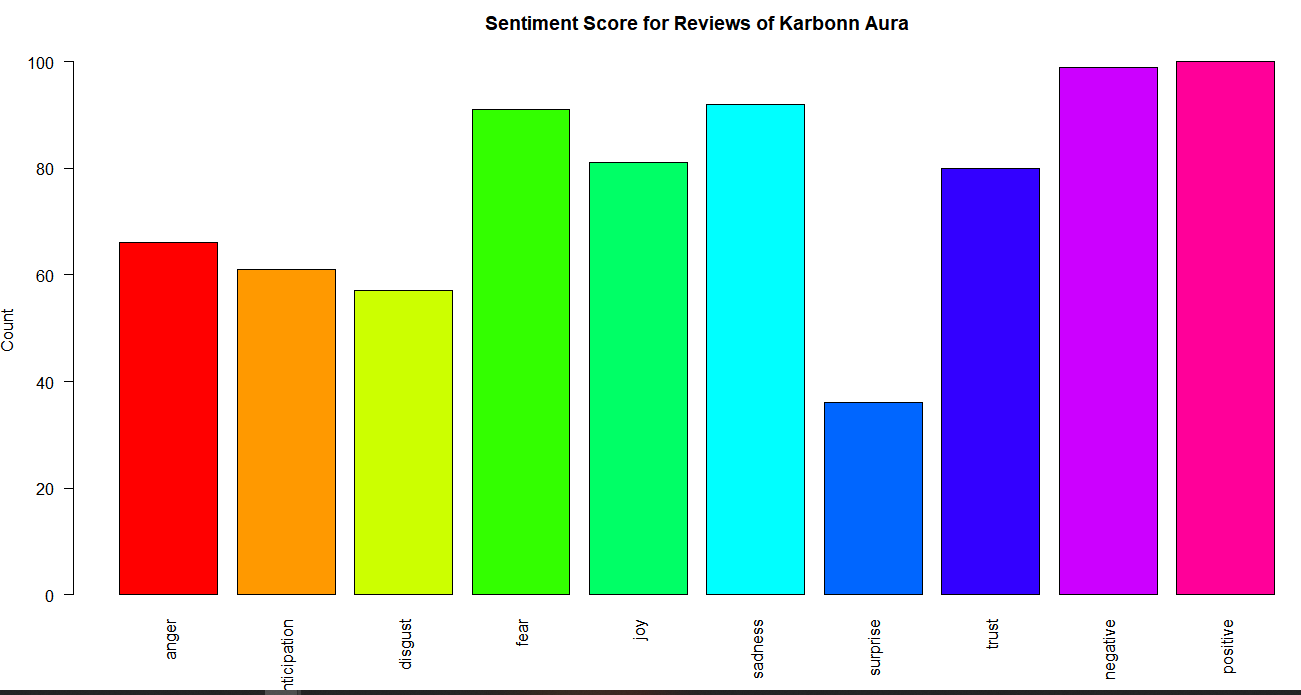
s <- get\_nrc\_sentiment(karbonn\_aura\_review)

get\_nrc\_sentiment---- It is a dictionary used to calculate the presence of eight different emotions and their corresponding valence in a text file

The following is a glance of emotions and their corresponding valence from row no. 6 to 14



Barplot



Form the above score it can be said that, that there is a combination of both positive and negative reviews from customers in almost equal proportion.

They expressed the feeling of sadness and fear through their reviews. There is also a sense of joy.

They also express anger about this phone. So there is mixed reviews about the phone.

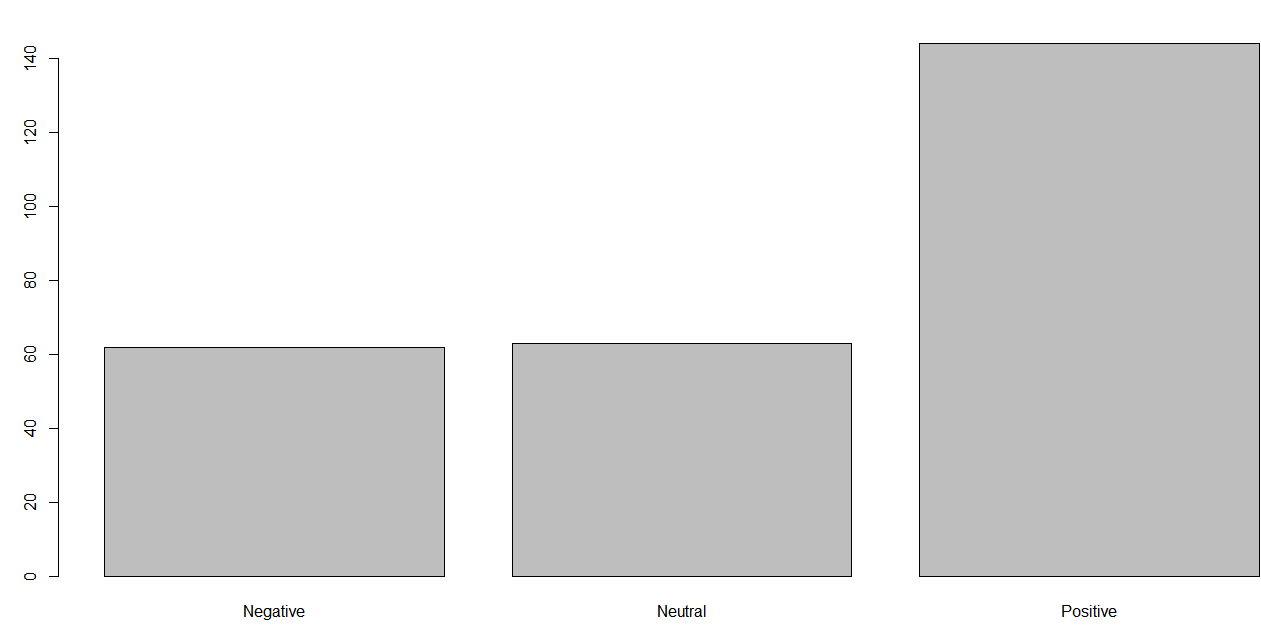
Most Positive reviews

"Received before time. And satisfied so far. Worth buying"

Most Negative reviews

"Bad problem”

**Segregating positive and negative reviews**





There are 144 positive reviews, 63 neutral reviews and 62 negative reviews.

# From Amazon

Extracting rewiews for the product Canon EOS 1500D (Camera)

**aurl <- "https://www.amazon.in/Canon-1500D-Digital-Camera-S18-55/product-reviews/B07BS4TJ43/ref=cm\_cr\_getr\_d\_paging\_btm\_prev\_1?pageNumber"**

**amazon\_reviews <- NULL**

**for (i in 1:20){**

**a\_link <- as.character(paste(aurl,i,sep="="))**

**murl <- read\_html(a\_link)**

**rev <- murl %>%**

**html\_nodes(".review-text") %>%**

**html\_text()**

**amazon\_reviews <- c(amazon\_reviews,rev)**

**}**

**write\_csv <- write.csv(amazon\_reviews,"canon\_camera.csv")**

Reviews extracted is in csv format

It is saved in csv format (canon\_camera.csv)

**2.) Sentiment Analysis**

**Import Datasets**

**Canon\_camera.csv**

Data pre processing

1. Converting to Corpus

cannon\_camera\_review <- Corpus(VectorSource(cannon\_camera$x))



The above summarizes the corpus with 200 documents.

2. Converting to Lower case

cannon\_camera\_review<- tm\_map(cannon\_camera\_review, tolower)

3. Removing Punctuation

cannon\_camera\_review<- tm\_map(cannon\_camera\_review, removePunctuation)

4. Removing Numbers

cannon\_camera\_review<- tm\_map(cannon\_camera\_review, removeNumbers)

5. Removing Stopwords

cannon\_camera\_review<- tm\_map(cannon\_camera\_review, removeWords, stopwords('english'))

6. Removing URL

removeURL <- function(z) gsub('http[[:alnum:]]\*', '', z)

cannon\_camera\_review<- tm\_map(cannon\_camera\_review, content\_transformer(removeURL))

7. Removing whitespace

cannon\_camera\_review<- tm\_map(cannon\_camera\_review, stripWhitespace)

8. Converting to Term Document Matrix

tdm <- TermDocumentMatrix(cannon\_camera\_review)

9. Converting to Document Term matrix

dtm <- t(tdm)

10. Converting document term matrix in matrix format

tdm <- as.matrix(tdm)

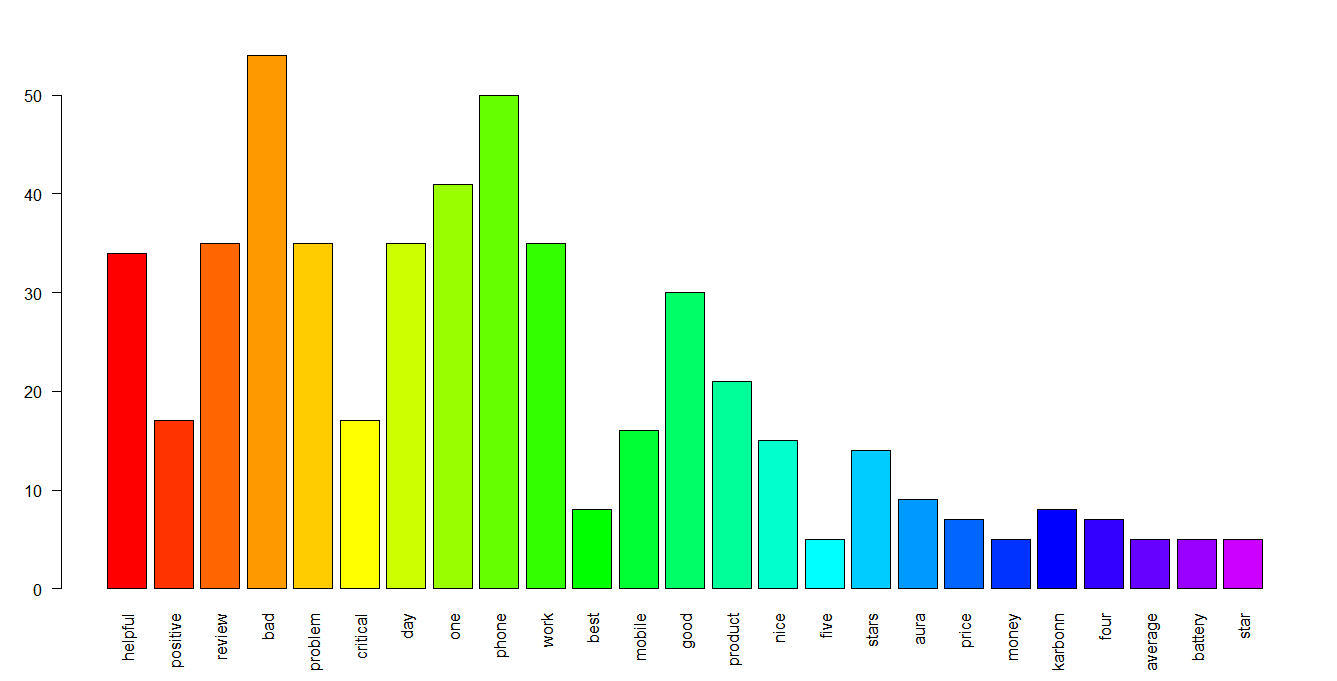
**Visualiztaion**

Lets first get the count of each entity

w <- rowSums(tdm)

Only for the words having frequency more then 10

Barplot



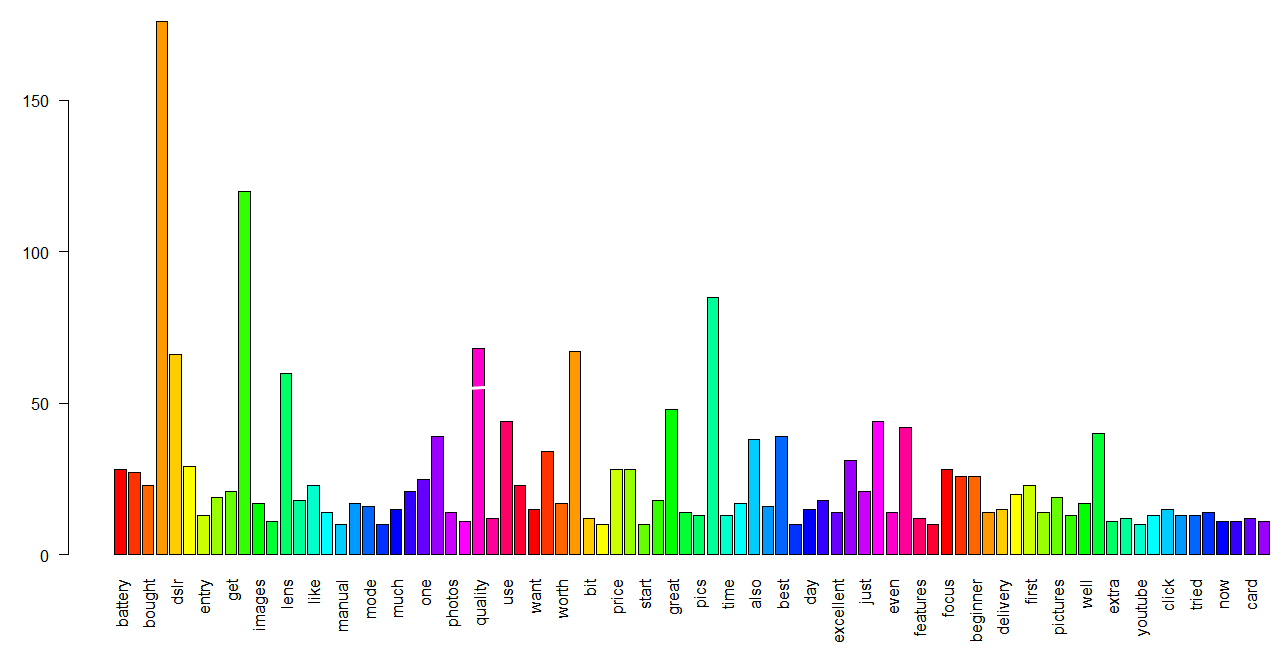
Few of the words can be removed as this does not contain any meaning.

# Removing some common word

list\_word <-c(stopwords('english'), "canon",”amazon”)

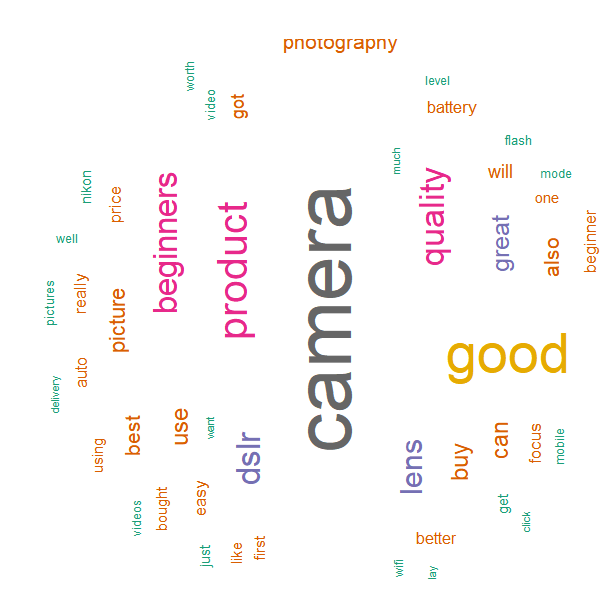
cannon\_camera\_review<- tm\_map(cannon\_camera\_review, stopwords(list\_word))

cannon\_camera\_review<- tm\_mapcannon\_camera\_review, stripWhitespace)



**Creating the word clouds**

**wordcloud(words = names(y),freq = y, max.words = 150, random.order = F, min.freq = 3, colors=brewer.pal(8,"Dark2"), scale=c(5,0.3),rot.per = 0.7)**



The highest frequency word is camera and 2nd highest is good

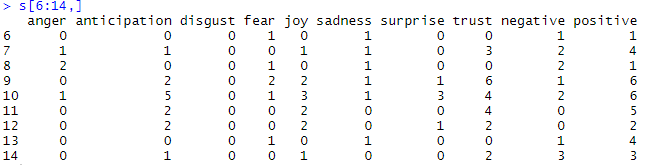
**Sentiment Analysis**

Obtaining Sentiment Scores

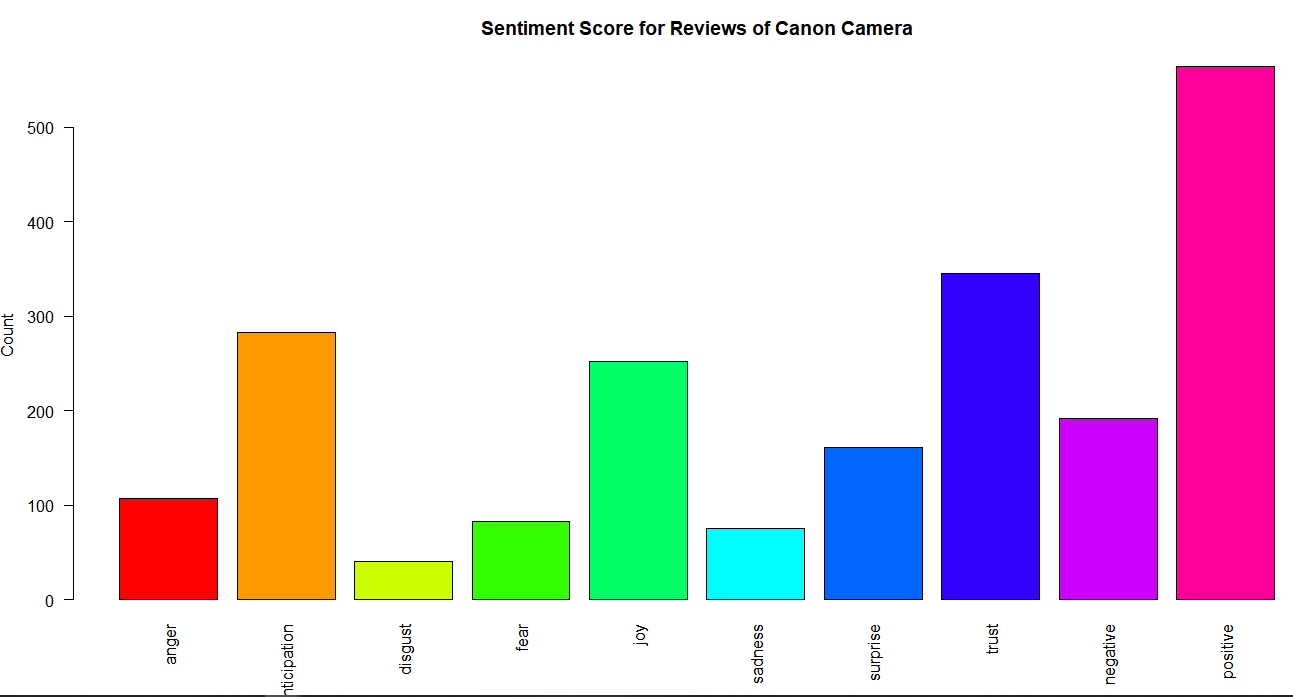
s <- get\_nrc\_sentiment(karbonn\_aura\_review)

get\_nrc\_sentiment---- It is a dictionary used to calculate the presence of eight different emotions and their corresponding valence in a text file

The following is a glance of emotions and their corresponding valence from row no. 6 to 14



Barplot



Form the above score it can be said that, there is more of positive reviews from customers then that of negative reviews.

There is a sense of trust and joy among the customer for the product.

Very few reviews are there which expresses anger and sadness

Most Positive reviews

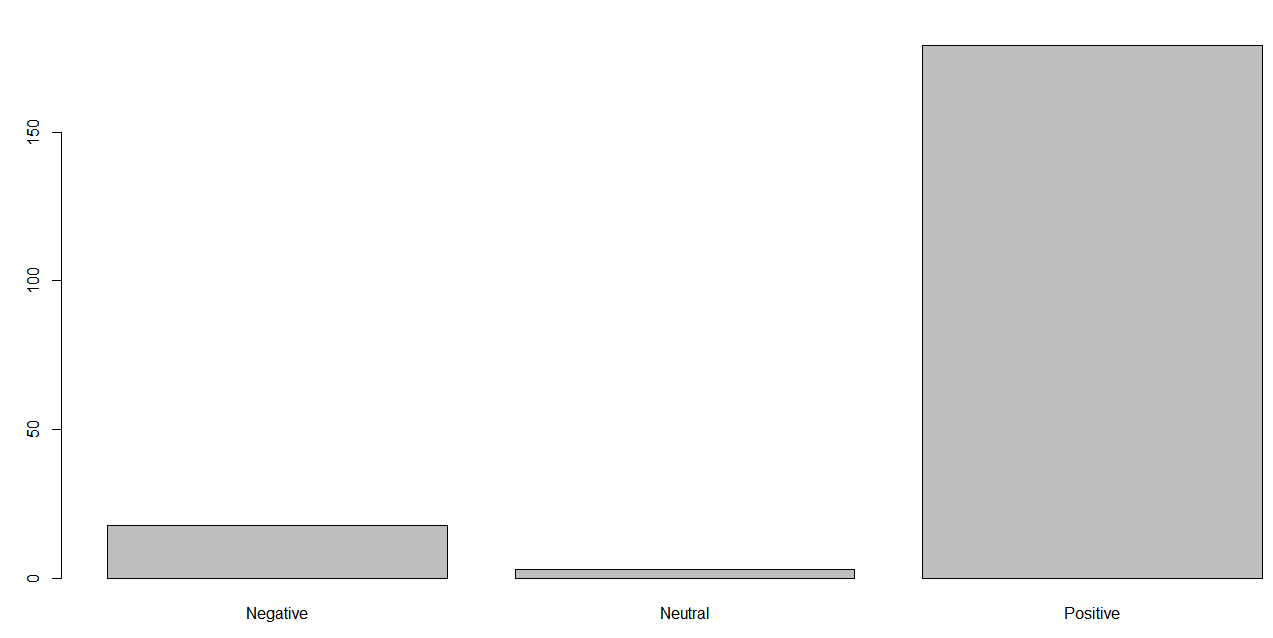
“Thanks to amazon and the courier partner for the quick delivery. Well it was an awesome deal from amazon purchased it for around 21.5k with a free gift Motorola wired headphone worth 650.Coming to the camera it’s a great learning camera especially for the beginners who love photography and also want themselves to enhance here. My personal experience after using it for a week for now with the combination of shutter speed, aperture and ISO with a desired white balance you can be click practically any of the desired picture that you want to click.The camera comes with a 18-55 lens that is selfacclaimed for beginners. Do master your parameters with this lens then you can obviously go for desired range of lens available in the market. The only shortcoming that I personally found was the focus which need to be adjusted manually/you need to have patience while clicking raw pictures. But this is again not a shortcoming for sure this is the area you need to improve yourself over the camera.”

Most Negative reviews

"Highly disappointed with the problems I'm facing with the camera, the

flash of the camera stops working and start showing error 50. Restarted many times but the problem remained same”

**Segregating positive and negative reviews**





There are 179 positive reviews, 3 neutral reviews and 18 negative reviews.

**3.) 1) Extract movie reviews for any movie from IMDB and perform sentimental analysis**

**2) Extract anything you choose from the internet and do some research on how we extract**

**using R. Programming and perform sentimental analysis.**

**1) Extract movie reviews for any movie from IMDB and perform sentimental analysis**

**Solution:-**

**Extracting reviews from IMDB**

Movie Name = “Frozen(2010)”

imdb\_reviews<-NULL

aurl<-"https://www.imdb.com/title/tt1323045/reviews?ref\_=tt\_ql\_3"

url<-read\_html(aurl)

ping<-url %>%

html\_nodes(".show-more\_\_control") %>%

html\_text()

imdb\_reviews<-c(imdb\_reviews,ping)

length(imdb\_reviews)

ping1<-url %>%

html\_nodes(".title") %>%

html\_text()

imdb\_reviews<-c(imdb\_reviews,ping1)

write.csv(imdb\_reviews,file="Frozen.csv")

Reviews extracted is in csv format

It is saved in csv format (Fraozen.csv)

**2.) Sentiment Analysis**

**Import Datasets**

**Frozen.csv**

Data pre processing

1. Converting to Corpus

frozen\_review<- Corpus(VectorSource(frozen $x))



The above summarizes the corpus with 76 documents.

Data Cleaning

The following are the cleaning method applied

1. Converting to Lower case

2. Removing Punctuation

3 . Removing Numbers

4. Removing Stopwords

5. Removing URL

6. Removing whitespace

7. Converting to Term Document Matrix

tdm <- TermDocumentMatrix(frozen\_review)

9. Converting to Document Term matrix

dtm <- t(tdm)

10. Converting document term matrix in matrix format

tdm <- as.matrix(tdm)

**Creating the word clouds**

**wordcloud(words = names(y),freq = y, max.words = 150, random.order = F, min.freq = 3, colors=brewer.pal(8,"Dark2"), scale=c(5,0.3),rot.per = 0.7)**



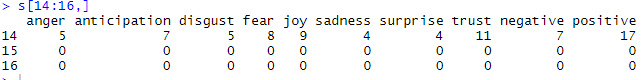
**Sentiment Analysis**

Obtaining Sentiment Scores

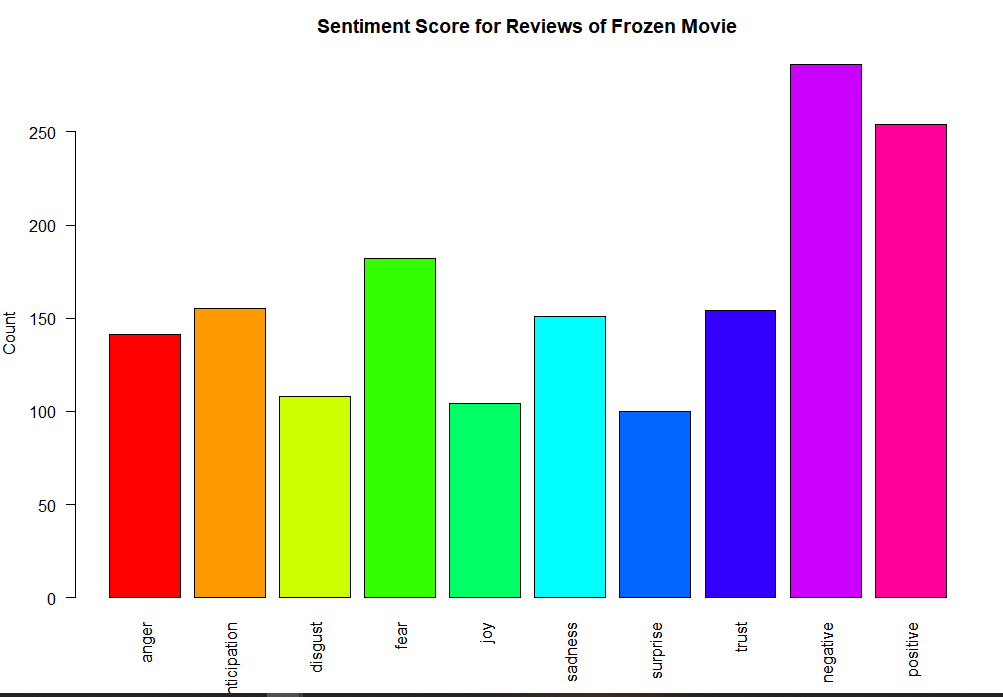
s <- get\_nrc\_sentiment(karbonn\_aura\_review)

get\_nrc\_sentiment---- It is a dictionary used to calculate the presence of eight different emotions and their corresponding valence in a text file

The following is a glance of emotions and their corresponding valence from row no. 14 to 16



Barplot



Form the above score it can be said that, that there is a combination of both positive and negative reviews from customers, but negative reviews are more to a little extent.

They expressed the feeling of anger, anticipation and fear through their reviews. There is also a sense of sadness and disgust to some extent.

There is also a sense of trust and joy in the reviews to medium extent

So there is mixed reviews from audience giving more weightage to negative sentiments.

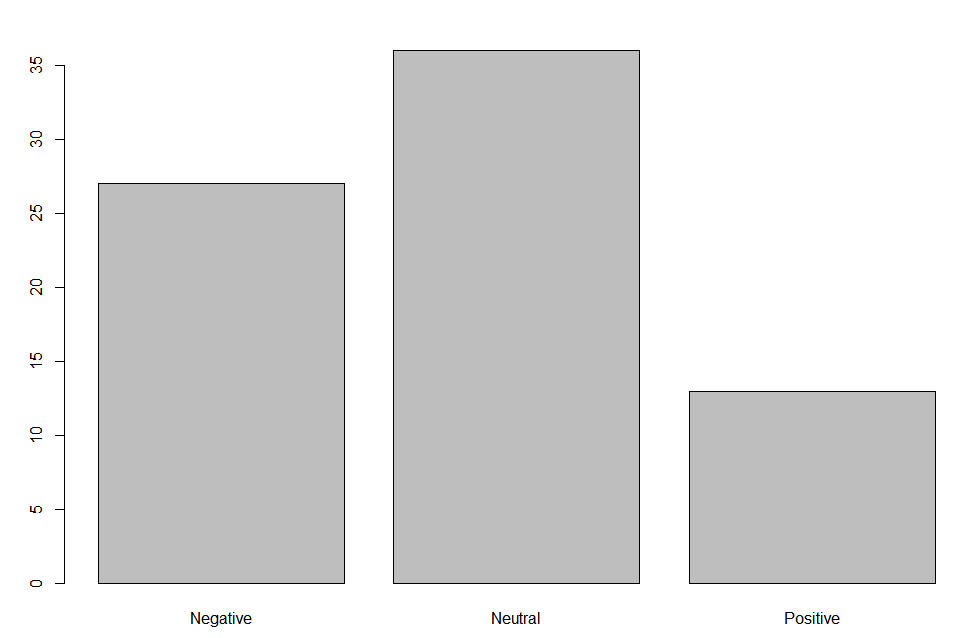
Most Positive reviews

"I'm not jumping up and down for this one but I thought it was a very well done survival/thriller that had some shocking, gruesome scenes as well as some thrills and definitely chills. The acting was pretty decent but the stand out here is newcomer Emma Bell who gave an heartbreaking performance as one of the stranded skiers on a ski lift, man I felt all of her pain. I thought the director did a great job using different techniques with the camera and used his one location effectively that makes you think and feel that you're right there with them on that ski lift, the thought just terrifies me but doesn't make me fear going skiing just makes me think twice about going on a ski lift. I think it's a tad overrated saying it does for skiing what Jaws did for swimming, not really and to be honest it's more Open Water than Jaws just with more action and a different situation, sorry it just didn't put the fear in me, it just had me feeling sorry for them and rooting for them than actually being scared. It will have you thinking throughout questioning what would you do in that situation, what lengths would you go in order to survive and some of it to me wasn't plausible while other scenes were very realistic and will have you on the edge of your seat. What makes this so special is the fact that they took a simple idea and went to great lengths to make it powerful and had three leads that could be anyone's friend, sister, brother, lover and it makes you feel for them even more and makes the whole situation more unsettling and shattering with some scenes that will make you a little teary eyed, good effort. This film actually has a heart and is refreshing from most horror offerings lately but it's more of a drama/thriller and I can't wait to see what director Adam Green has up his sleeve next! Overall it could of had a slightly better conclusion and I don't think it's going to be the best genre flick of the year because I just don't see that happening but if you want to see something different, new and bone chilling go support Adam Green's latest effort Frozen, though more of a renter on a snowy night to me. Recommended!"

Most Negative reviews

"When you lack imagination or vision, just make up a kitsch movie about the 'Big Bad Wolf'. Then throw in a brainless 'Little Red Riding Hood' bimbo and you can call yourself a Hollywood legend.Dan, the surviving college student finally tells Parker, the constantly crying & sobbing girl, that all you need to do is CLAP and SCREAM in a threatening manner because 'the wolves are just as afraid of you as you are of them'. So...why didn't he tell this to Lynch his buddy (the grandma symbol) who broke his leg jumping off of the ski lift? No, the filmmaker had to stage a wolf attack for false savagery, gratuitous violence and lack of any other creative story line. Who ever owns these PET WOLVES and rented them out should be more than ashamed of themselves for pandering and once again stereotyping these magnificent animals into target practice in the wild. Wolves are actually shy, elusive creatures that humans rarely see because wolves know how dangerous and unpredictable humans really are. Frozen's just another sad day for endangered wolves who always end up as the primal fear image in amateur showcases of this worn out, ancient children's myth."

**Segregating positive and negative reviews**





There are 13 positive reviews, 36 neutral reviews and 27 negative reviews.

**2) Extract anything you choose from the internet and do some research on how we extract using R. Programming and perform sentimental analysis.**

Solution:-

**Extracting Comments from Youtube**

**Here for downloading youtube video comments, “”vosonSML” package is installed and invoked**

Video name = “Chicago Speech of Swami Vivekananda at the World Parliament of Religions for ¦ RSM ¦ SRM”

**youtube authentication will set the api key (API Ket should be generated from registering and creating projects in google developer console)**

youtubeAuth <- Authenticate("youtube", apiKey = "XXXXXXXX")

**To get youtube video id from url**

youtubeVideoIds <- GetYoutubeVideoIDs("https://www.youtube.com/watch?v=ii7obeNa20g")

**This will extratct the information about the videos**

youtubeData <- youtubeAuth %>%

Collect(videoIDs = youtubeVideoIds, maxComments = 100,

verbose = FALSE)

Need to analyse comments

swami\_vivekananda\_comment <- youtubeData$Comment

**2.) Sentiment Analysis**

**Import Datasets**

**Frozen.csv**

Data pre processing

1. Converting to Corpus

swami\_vivekananda\_comment <- Corpus(VectorSource(swami\_vivekananda\_comment))



The above summarizes the corpus with 124 documents.

**Data Cleaning**

The following are the cleaning method applied

1. Converting to Lower case

2. Removing Punctuation

3. Removing Numbers

4. Removing Stopwords

5. Removing URL

6. Removing whitespace

7. Converting to Term Document Matrix

tdm <- TermDocumentMatrix(swami\_vivekananda\_comment)

9. Converting to Document Term matrix

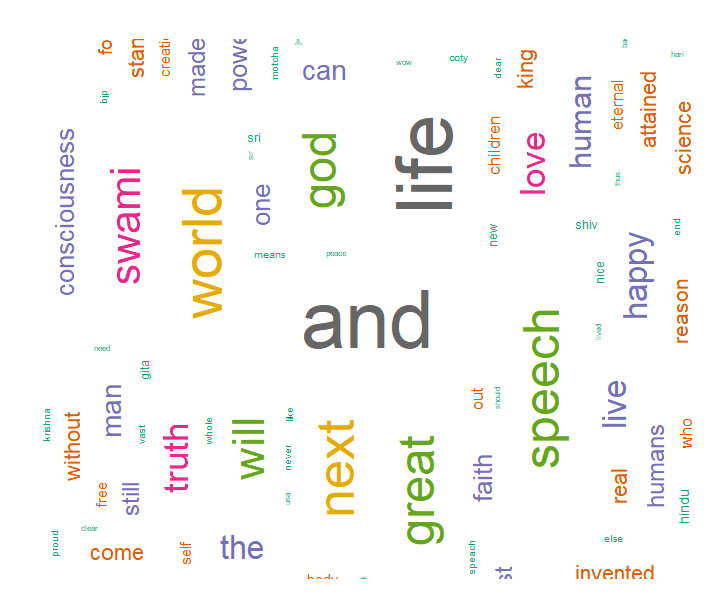
dtm <- t(tdm)

10. Converting document term matrix in matrix format

tdm <- as.matrix(tdm)

**Creating the word clouds**

**wordcloud(words = names(y),freq = y, max.words = 250, random.order = F, min.freq = 2, colors=brewer.pal(8,"Dark2"), scale=c(5,0.3),rot.per = 0.7)**



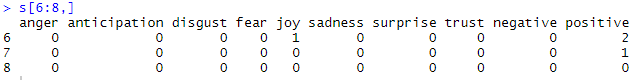
**Sentiment Analysis**

Obtaining Sentiment Scores

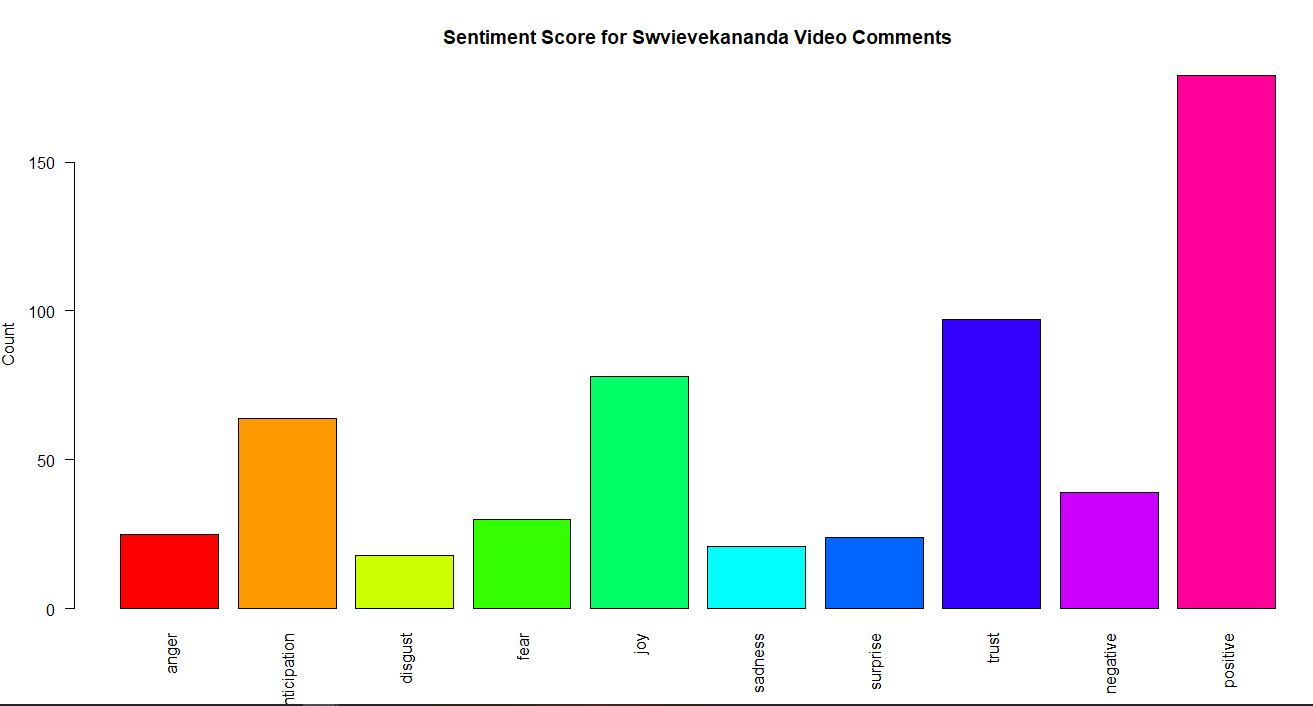
s <- get\_nrc\_sentiment(karbonn\_aura\_review)

get\_nrc\_sentiment---- It is a dictionary used to calculate the presence of eight different emotions and their corresponding valence in a text file

The following is a glance of emotions and their corresponding valence from row no. 6 to 8



Barplot



Form the above score it can be said that, viewers mostly expressed positive feelings about the video content. They expressed the feeling of trust and joy. There is a less feeling of sadness and disgust and anger .Overall the viewers expresses their positive sentiments over the video.

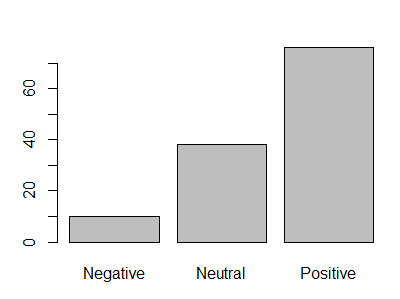
Most Positive Comments

"Wow Not just tolerance but acceptance of eachother. purity , style, quality ,eloquence, clarity , perception, wisdom , deep insights of Life, creator ,& creation is one & same , knowledge of cosmic energy. Self Realization is to evolve mankind to Truth . Idiots live miserable life , wakeup from delusions . Death is only to body. Spirit is Free & eternal , children of immortal Bliss. Absolute Oneness in universal consciousness. Love for love sake ...... Real monks power of vak shudhi , power of Brahmacharya in 1893 Amazing ,pranams swami”

Most Negative Comments

“Why would some one dislike this man”

**Segregating positive and negative comments**





There are 76 positive reviews, 38 neutral reviews and 10 negative comments.