

# Title of the project: **Twitter Sentiment Analysis - TEXT MINING**

Team Members: Aswin G E0119062 , Lohit Arun E0119042 , Pragadeeshwar N E0119003

Problem Statement:

***To analyse the mindset of the users over a period of time about the US Political scenario.***

Data Description :

Tweets of the Users about *DonaldTrump* and *JoeBiden*

Software Requirements and Platforms used: RStudio , Google Colab

***Installing the Packages :***

```
install.packages("ROAuth")
```

```
Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)
```

```
install.packages("plyr")
```

```
Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)
```

```
install.packages("stringr")
```

```
Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)
```

```
install.packages("RColorBrewer")
```

```
Installing package into '/usr/local/lib/R/site-library'
```

```
(as 'lib' is unspecified)
```

```
install.packages("ggplot2")
```

```
Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)
```

```
install.packages("RCurl")
```

```
Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)
```

```
install.packages("syuzhet")
```

```
Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)
```

```
install.packages("tm")
```

```
Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)
```

```
install.packages("wordcloud")
```

```
Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)
```

```
install.packages("twitter")
```

```
Installing package into '/usr/local/lib/R/site-library'  
(as 'lib' is unspecified)
```

### ***Importing the Libraries***

```
library(twitterR)  
library(ROAuth)  
library(plyr)  
library(stringr)  
library(ggplot2)  
library(RCurl)  
library(RColorBrewer)  
require('ROAuth')  
require('RCurl')  
library(syuzhet)
```

```
library(tm)
library(wordcloud)
```

Attaching package: 'plyr'

The following object is masked from 'package:twitter':

id

Loading required package: NLP

Attaching package: 'NLP'

The following object is masked from 'package:ggplot2':

annotate

## Linking with the Twitter Developer Account:

```
requestURL <- "https://api.twitter.com/oauth/request_token"
accessURL = "https://api.twitter.com/oauth/access_token"
authURL = "https://api.twitter.com/oauth/authorize"
```

```
consumerKey = "JQECH17A2ds6yRPXB3eG2VqBX"
consumerSecret = "9laFAxNgMjzJTNaWi7nvTGueUy9kKhjkWKXqje0InIgTyMbA7"
accessToken<- "1437806033355755528-x9Z8bIy8BTpIstiNW6Gn6vGxOqMRsn"
accessTokenSecret <- "320Sp9Yy0IjwQjGoWvv9hUAdh3AbymdTXdkVveVgMHOyG"
```

```
setup_twitter_oauth(consumerKey ,consumerSecret, accessToken, accessTokenSecret )
```

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

```
cred <- OAuthFactory$new(consumerKey= consumerKey,
                        consumerSecret=consumerSecret,
                        requestURL=requestURL,
                        accessURL=accessURL,
                        authURL=authURL)
```

For ***DonaldTrump*** :

## ▼ C01 : Know your Data

```
tweet_trump <- searchTwitter('#DonaldTrump',n=500,lang = "en",since = "2020-09-14",until =  
tweet_c = twListToDF(tweet_trump) #Creating the DataFrame  
write.csv(tweet_c,file = "/content/sample_data/trump.csv",row.names = F) #Writing it into  
  
tweet_text <- tweet_c$text #Selecting the specific "Text" column from the dataset  
View(tweet_text)
```

```
'Random Trump https://t.co/KVKhrDt06i - #randomtrump #trump #donaldtrump #giphy #gif #gifs
https://t.co/VuZqiiOokh' ·
'RT @Ghostly_Host: This Is My President! https://t.co/9jBbD7Bpfv #Trump #NRA #America #MAGA
#KAG #DonaldTrump #Trump2024 #PleaseRT' ·
'#DonaldTrump #donaldjtrump Little Man Big Loser https://t.co/6U0Uaypouv https://t.co/tL79H151Tb' ·
'RT @Red_N_Right: Trump warns Taliban-seized U.S. weapons will be given to China, Russia | Lee
Brown https://t.co/2UMNwLluWF ~ #DonaldTrump...' ·
'Random Trump https://t.co/FMNmGaRLDP - #randomtrump #trump #donaldtrump #giphy #gif #gifs
https://t.co/6U0Uaypouv https://t.co/tL79H151Tb' ·
```

## ▼ C02: Data Wrangling

```
#KAG #DonaldTrump #trump2024 #PleaseRT' ·
```

```
tweet_text <- tolower(tweet_text)
tweet_text <- gsub("rt","",tweet_text)
tweet_text <- gsub("@\\w+","",tweet_text)
tweet_text <- gsub("[[:punct:]]","",tweet_text)
tweet_text <- gsub("http\\w+","",tweet_text)
tweet_text <- gsub("[\\t]{2,}","",tweet_text)
```

```
tweet_text <- gsub("^","",tweet_text)
tweet_text <- gsub("$","",tweet_text)
```

```
'RT @Ghostly_Host: This Is My President! https://t.co/9jBbD7Bpfv #Trump #NRA #America #MAGA
```

```
View(tweet_text)
```

```
'random trump randomtrump trump donaldrump giphy gif gifs ' ·
' this is my president trump nra america maga kag donaldrump trump2024 please' ·
'donaldrump donaldjtrump little man big loser ' ·
' trump warns talibanseized us weapons will be given to china russia lee brown donaldrump' ·
'random trump randomtrump trump donaldrump giphy gif gifs ' ·
' can somebody please tell me how the senate didnt do today interview in person ' ·
' this is my president trump nra america maga kag donaldrump trump2024 please' ·
'donaldrump is eager to lose again theloser ' ·
'israeli antivaxx leader dies of covid19 \n\n \n\nleftwing world trump politics news maga ' ·
' this is my president trump nra america maga kag donaldrump trump2024 please' ·
'why would he bother with prince andrew when he was in such close company with manhattan pussy
grabber misogynist ' ·
'its not right trudeau vows to criminalize blocking access to hospitals amid protests \n\n ' ·
'random trump randomtrump trump donaldrump giphy gif gifs ' ·
' this is my president trump nra america maga kag donaldrump trump2024 please' ·
'larryelder refuses to commit to accepting californiarecallection results in interview hes following in ' ·
'random trump randomtrump trump donaldrump giphy gif gifs ' ·
'amy coney barrett is the product of a corrupt and politicized supreme cou nomination process ' ·
' this is my president trump nra america maga kag donaldrump trump2024 please' ·
'if donaldrump should be triedimprisoned for not committing a crimewhat does that say about hillar ' ·
' theres hope folks \nmaga\ndonaldrump \nimpeachbidennow \n ' ·
'random trump randomtrump trump donaldrump giphy gif gifs ' ·
'trump wouldnt dare let another riot at the capitol happen sator would hecan you trust him lets go wit ' ·
' this is my president trump nra america maga kag donaldrump trump2024 please' ·
'random trump randomtrump trump donaldrump giphy gif gifs ' ·
' this is my president trump nra america maga kag donaldrump trump2024 please'
```

Corpus creation:

```
' this is my president trump nra america maga kag donaldrump trump2024 please' ·
```

```
corp1 <- Corpus(VectorSource(tweet_text)) #Creating the corpus
tweet_c.text.corpus <- tm_map(corp1,function(x)removeWords(x,stopwords())) #Removing the
```

```
Warning message in tm_map.SimpleCorpus(corp1, function(x) removeWords(x, stopwords())
"transformation drops documents"
```

```
'random trump randomtrump trump donaldrump giphy gif gifs ' ·
```

## ▼ CO3: Data Modeling and Analysis

```
' in the old days we called people like donaldrump amp larryelderorcagov sore losers ' ·
```

**Euclidean Distance:**

```
'random trump randomtrump trump donaldrump giphy gif gifs ' ·
```

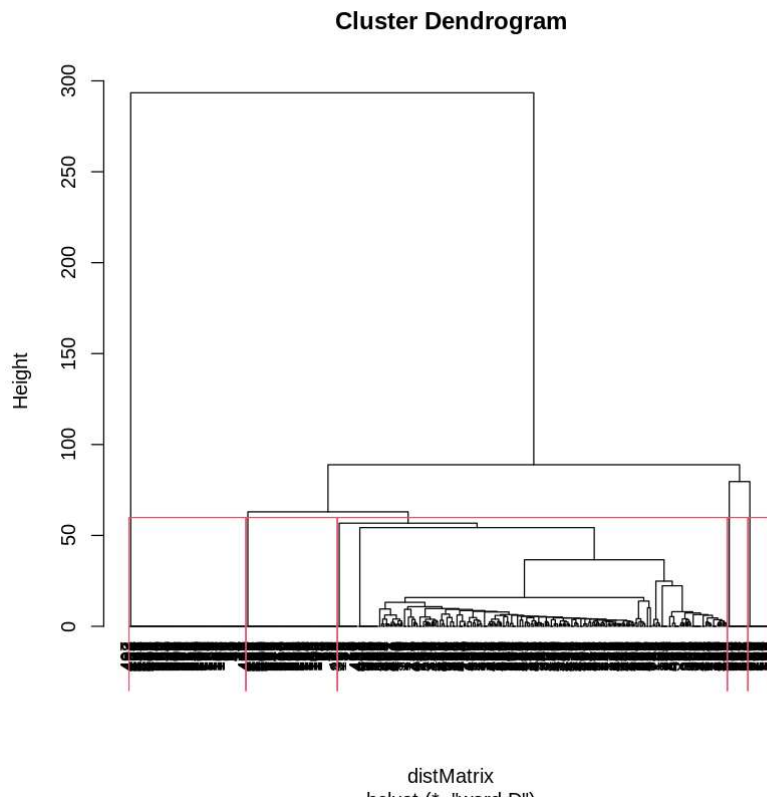
```
m <- as.matrix(dtm)
distMatrix <- dist(m, method="euclidean")
```

```
'random trump giphy gif gifs ' ·
```

**Heirarchial Clustering - Dendogram :**

```
' 2016 trump lost popular vote by 3 million\n2020 trump lost popu ' ·
```

```
groups <- hclust(distMatrix,method="ward.D")
plot(groups, cex=0.9, hang=-1)
rect.hclust(groups, k=5)
```



## ▼ C04: Mining Techniques in R

### TF-IDF

```
dtm.matrix = as.matrix(dtm)
wordcloud(colnames(dtm.matrix), dtm.matrix[, ], max.words = 50)
```

**Document Term matrix (DTM):**

```

ndocs <- length(corp1) #Checking the length of the corpus
ndocs

500

minTermFreq <- ndocs * 0.01
maxTermFreq <- ndocs * .5

dtm = DocumentTermMatrix(corp1,
                          control = list(
                            stopwords = TRUE,
                            wordLengths=c(4, 15),
                            removePunctuation = T,
                            removeNumbers = T,
                            #stemming = T,
                            bounds = list(global = c(minTermFreq, maxTermFreq))
                          ))
dtm

<<DocumentTermMatrix (documents: 500, terms: 115)>>
Non-/sparse entries: 1736/55764
Sparsity           : 97%
Maximal term length: 13
Weighting          : term frequency (tf)

```

## CO5: Implementing case study in R : Describe the scenario taken for analysis

**WORDCLOUD CREATION**

```

library(wordcloud) #Importing the WordCloud Library
wordcloud(tweet_c.text.corpus,min.freq = 10,colors=brewer.pal(8,"Dark2"),random.color=TRUE

```





*So, from this WordCloud we can see that the word "donaldtrump" has the most number of frequency along with the words trump, randomtrump.*

### Sentiment Analysis:

```
mysentiment_trump <- get_nrc_sentiment((tweet_text))

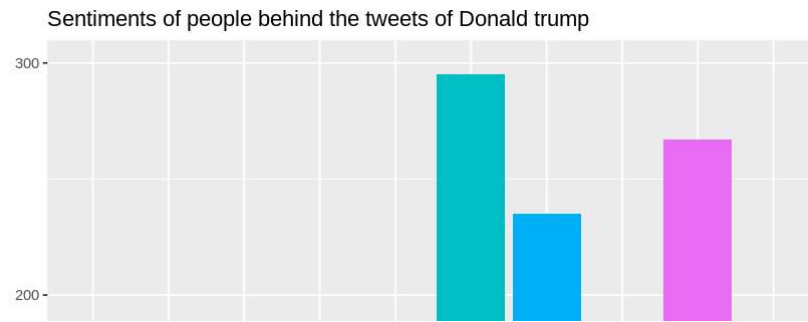
sentimentscores_trump <- data.frame(colSums(mysentiment_trump[,]))

names(sentimentscores_trump) <- "Score"

sentimentscores_trump <- cbind("sentiment"= rownames(sentimentscores_trump),sentimentscore

rownames(sentimentscores_trump) <- NULL

ggplot(data = sentimentscores_trump,aes(x=sentiment,y=Score)) +geom_bar(aes(fill=sentiment
  theme(legend.position = "none")+xlab("Sentiments")+ylab("scores")+ggtitle("Sentiments of
```



For **JoeBiden** :



## ▼ C01 : Know your Data



```
tweet_biden <- searchTwitter('#JoeBiden',n=500,lang = "en",since = "2020-09-14",until = "2020-10-14")
tweet_d = twListToDF(tweet_biden)
write.csv(tweet_d,file = "/content/sample_data/biden.csv",row.names = F)
```

```
tweet_text1 <- tweet_d$text
View(tweet_text1)
```

'RT @delictum: @JoeBiden #JoeBiden @potus #POTUS @DrBiden @FLOTUS #FLOTUS @WhiteHouse #WhiteHouse @KamalaHarris @vp @PressSec @jrpsaki see...' ·  
 '@bsarwary Good work #JoeBiden' ·  
 '@JoeBiden #JoeBiden @potus #POTUS @DrBiden @FLOTUS #FLOTUS @WhiteHouse #WhiteHouse @KamalaHarris @vp @PressSec... <https://t.co/DXsnINh0lj>' ·  
 'RT @LeeroyPress: #happeningnow on the Brooklyn Bridge in NYC , protesters against the mandatory shot chant " F@CK JOE BIDEN " \n#JoeBiden #V...' ·  
 'RT @LeeroyPress: #happeningnow on the Brooklyn Bridge in NYC , protesters against the mandatory shot chant " F@CK JOE BIDEN " \n#JoeBiden #V...' ·  
 'RT @ada\_cr\_: Joe Biden receives warm welcome in Boise #JoeBiden #idpol #idgov #Boise <https://t.co/3JEFRps0H5>' ·  
 '#JoeBiden & the #DemocraticParty are supporting the torture of #ZakariaZubaidi & the Palestinian political prisoner... <https://t.co/dA8h9s4gCg>' ·  
 'RT @LeeroyPress: #happeningnow on the Brooklyn Bridge in NYC , protesters against the mandatory shot chant " F@CK JOE BIDEN " \n#JoeBiden #V...' ·  
 '@AOC #JoeBiden is #theoldman. Feel better #AOC @AOC' ·  
 'RT @LeeroyPress: #happeningnow on the Brooklyn Bridge in NYC , protesters against the mandatory shot chant " F@CK JOE BIDEN " \n#JoeBiden #V...' ·  
 'Hey #DEMOCRATS your #INstalled #President #JackASSJoe #JoeBiden. You must be proud!!! <https://t.co/trOVW2GzON>' ·  
 'RT @DrGem2015: Big Tech is the engine of the US economy and saved the society in the pandemic. Biden's illiterate progressive mold wants to...' ·  
 '@mmpadellan No I hope not in fact all senators and representatives including #JoeBiden who is over 60 need to resig... <https://t.co/yeo5ybWBzk>' ·  
 'RT @LeeroyPress: #happeningnow on the Brooklyn Bridge in NYC , protesters against the mandatory shot chant " F@CK JOE BIDEN " \n#JoeBiden #V...' ·  
 'Immigrant children are still being abused while in custody. #Immigration #immigrant #daca #Dreamers #abolishice... <https://t.co/tNiriS666e>' ·  
 'RT @... #JoeBiden #V...' ·

## ▼ CO2: Data Wrangling

SHOT CHANT " F@CK JOE BIDEN " \n#JOEBIDEN #V...' ·

#Data Wrangling:

```
tweet_text1 <- tolower(tweet_text1)
tweet_text1 <- gsub("rt","",tweet_text1)
tweet_text1 <- gsub("@\\w+", "",tweet_text1)
tweet_text1 <- gsub("[[:punct:]]", "",tweet_text1)
tweet_text1 <- gsub("http\\w+", "",tweet_text1)
tweet_text1 <- gsub("[|\\t]{2,}", "",tweet_text1)
```

```
tweet_text1 <- gsub("^","",tweet_text1)
tweet_text1 <- gsub("$","",tweet_text1)
```

charge. Feel like... <https://t.co/2DSEBd16K0>...

```
View(tweet_text1)
```

' joeiden potus flotus whitehouse see' · ' good work joeiden' ·  
 'joeiden potus flotus whitehouse ' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 ' joe Biden receives warm welcome in boise joeiden idpol idgov boise ' ·  
 'joeiden amp the democraticpay are suppoing the toure of zakariazubaidi amp the palestinian political  
 prisoner' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 'joeiden is theoldman feel better aoc ' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 'hey democrats your installed president jackassjoe joeiden you must be proud ' ·  
 ' big tech is the engine of the us economy and saved the society in the pandemic bidens illiterate  
 progressive mold wants to' ·  
 ' no i hope not in fact all senators and representatives including joeiden who is over 60 need to resig ' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 'immigrant children are still being abused while in custody immigration immigrant daca dreamers  
 abolishice ' ·  
 ' there comes the icecream truck Biden joeiden ' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 '125 لاالتطعيم الاجباري\n\nnocompulsoryvaccination\n\nunvaxxed\n\nvaxxed\n\nvaccinated\n\nunvaccinated\n\nnjoeiden\n\n  
 ' ·  
 ' it is 743pm est and joeiden and the Bidenregime are still holding office \n\nneach minute they remain ' ·  
 ' disgusting joe \n\n msnbc foxnews foxbusiness joeiden bloomberg cnbc cnn abc cbs whitehou's ' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 'hes fucking reading his fucking speech again joeiden senile california californiarecall fuckgavin' ·  
 'the white house abruptly cut the feed of president bidens briefing makes you wonder who is in charge  
 foolish p ' ·  
 'no republican not one would speak to the state of california and commitment to it as ' ·  
 'joeidenmustresign \nreligious extemptions for covid must be monitored what wheres religious  
 freedom joeiden' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·  
 ' why is joeiden campaigning in california when we are being invaded through our southern border and  
 there are six plan' ·  
 ' happeningnow on the brooklyn bridge in nyc protesters against the mandatory shot chant f joe Biden  
 \njoeiden v' ·

```
corp2 <- Corpus(VectorSource(tweet_text1))
```

```
tweet_d.text.corpus <- tm_map(corp2,function(x)removeWords(x,stopwords()))
```

Warning message in tm\_map.SimpleCorpus(corp2, function(x) removeWords(x, stopwords()  
 “transformation drops documents”

◀ [A horizontal scrollbar with a grey track and a white slider, positioned at the left end.] ▶

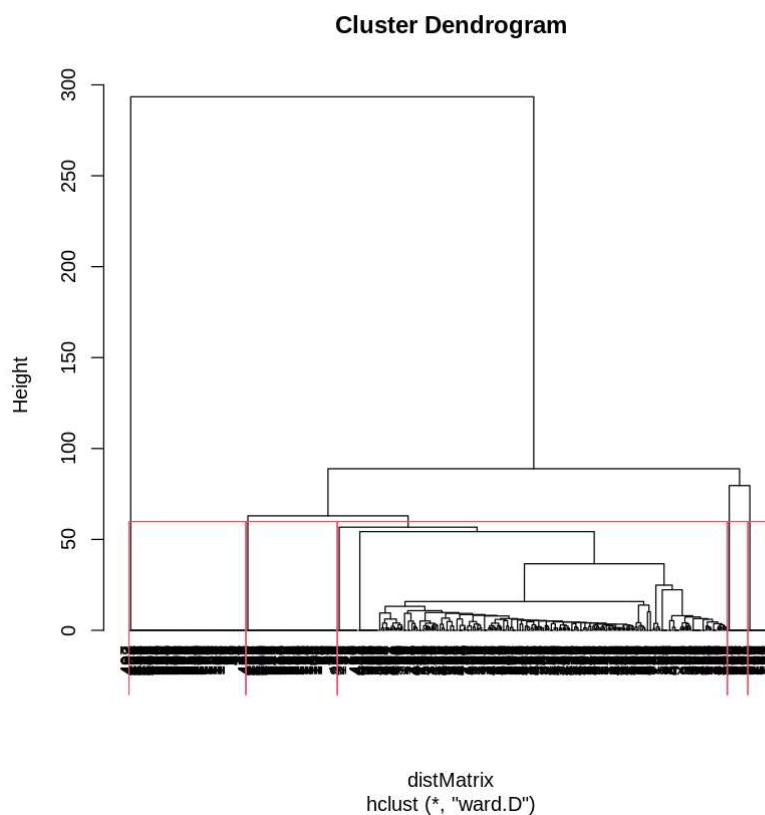
## ▼ C03: Data Modeling and Analysis

### Euclidean Distance:

```
m <- as.matrix(dtm)
distMatrix <- dist(m, method="euclidean")
```

### Hierarchical Clustering:

```
groups <- hclust(distMatrix,method="ward.D")
plot(groups, cex=0.9, hang=-1)
rect.hclust(groups, k=5)
```



## ▼ C04: Mining Techniques in R

### TF-IDF:

```
dtm.matrix = as.matrix(dtm)
wordcloud(colnames(dtm.matrix), dtm.matrix[, ], max.words = 50)
```

# whitehouse

# potus

## Document Term Matrix

```
ndocs <- length(corp2)
ndocs

500

minTermFreq <- ndocs * 0.01
maxTermFreq <- ndocs * .5

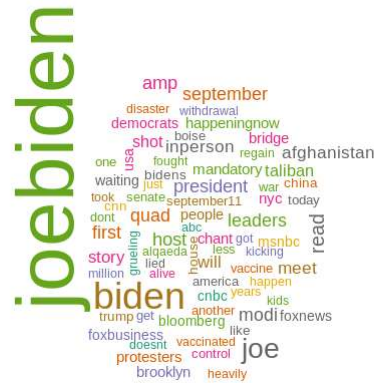
dtm = DocumentTermMatrix(corp2,
                          control = list(
                            stopwords = TRUE,
                            wordLengths=c(4, 15),
                            removePunctuation = T,
                            removeNumbers = T,
                            #stemming = T,
                            bounds = list(global = c(minTermFreq, maxTermFreq))
                          ))

dtm

<<DocumentTermMatrix (documents: 500, terms: 137)>>
Non-/sparse entries: 1725/66775
Sparsity           : 97%
Maximal term length: 15
Weighting          : term frequency (tf)
```

CO5: Implementing case study in R : Describe the scenario  
taken for analysis

```
wordcloud(tweet_d.text.corpus,min.freq = 10,colors=brewer.pal(8,"Dark2"),random.color=TRUE
```



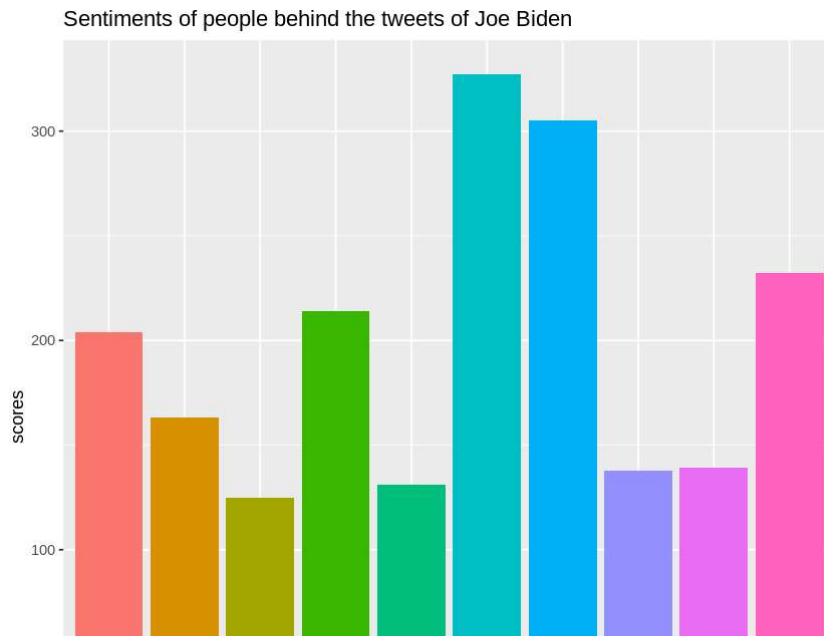
So, from this WordCloud we can see that the word "JoeBiden" has the most number of frequency along with the words *biden,democrats*.

### Sentiment Analysis:

```
mysentiment_biden <- get_nrc_sentiment((tweet_text1))
sentimentscores_biden <- data.frame(colSums(mysentiment_biden[,]))

names(sentimentscores_biden) <- "Score"
sentimentscores_biden <- cbind("sentiment"= rownames(sentimentscores_biden),sentimentscore
rownames(sentimentscores_biden) <- NULL

ggplot(data = sentimentscores_biden,aes(x=sentiment,y=Score)) +geom_bar(aes(fill=sentiment
  theme(legend.position = "none")+xlab("Sentiments")+ylab("scores")+ggtitle("Sentiments of
```



▼ **Conclusion:**



From the above two sentiment analysis, we can find the various difference of opinion of the users over the past year about Trump and Biden. Biden almost has similar negative and positive sentiments whereas Trump has more negative sentiment from the users. Likewise, Biden has created more trust amongst the users which is better than Trump