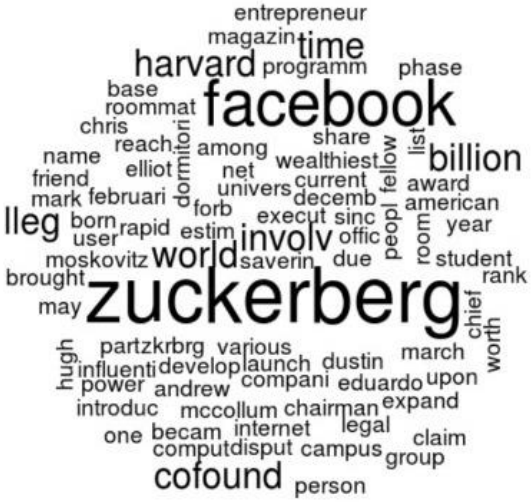
```
# Name : Manvi Pandya

#Batch : L-10

#Roll. No. : 33235


#install packages

install.packages("tm")

install.packages("wordcloud")

install.packages("RColorBrewer")

install.packages("corpus")


#Load packages

#text mining library used to highlight most frequently used words in a paragraph of text

library("tm")

library("wordcloud") #Analyse and visualize words as word cloud


text <- readLines(file.choose())

text


#Corpus is a format which is used by tm to clean text. It is a collection of docs with large structure of text

#VectorSource : creating a vector of character vectors

docs<- Corpus(VectorSource(text))


#Transform text by cleaning

docs <- tm_map(docs,tolower) #tm is case sensitive

docs <- tm_map(docs,removeNumbers) #remove numbers

docs <- tm_map(docs,removePunctuation) #remove Punctuations

docs <- tm_map(docs,removeWords,stopwords("english")) #Remove all stop words

docs <- tm_map(docs,stripWhitespace) #Remove whitespaces

docs <- tm_map(docs,stripWhitespace) #Remove whitespaces
```

```
toSpace <- content_transformer(function(x,pattern) gsub(pattern, " ",x))
docs <- tm_map(docs, toSpace, "\\W") #match any non word charcter

#Build document matrix
tdm <- TermDocumentMatrix(docs) #table with frequency of all words
m <- as.matrix(tdm) #convert to matrix

sorted_doc <- sort(rowSums(m),decreasing = TRUE)

#convert text data to dataframe
d <- data.frame(word=names(sorted_doc),freq=sorted_doc)
head(d,10)

wordcloud(words=d$word,freq = d$freq)
```


```
wordcloud(words=d$word,freq = d$freq,min.freq = 1)
```

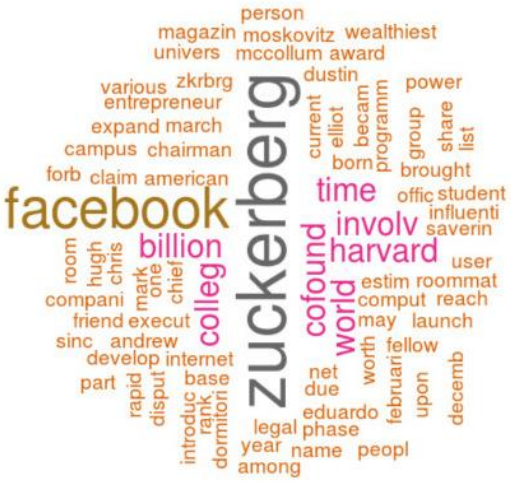
wordcloud(words=d$word,freq = d$freq,min.freq = 1,max.words = 200)



wordcloud(words=d$word,freq = d$freq,min.freq = 1,max.words = 200,random.order = FALSE)

wordcloud(words=d$word,freq = d$freq,min.freq = 1,max.words = 200,random.order = FALSE,rot.per = 0.35)

wordcloud(words=d$word, freq = d$freq , min.freq = 1, max.words = 200, random.order = FALSE,rot.per=0.35, colors=brewer.pal(8,"Dark2"))


person
magazin
moskovitz
wealthiest
univers
mccollum
award
dustin
various
zkrbrg
power
entrepreneur
current
elliot
becam
programm
group
share
list
expand
march
campus
chairman
born
brought
forb
claim
american
time
offic
student
facebook
zuckerberg
involv
influenti
saverin
cofound
harvard
billion
room
hugh
chris
mark
one
chief
user
estim
roommat
compani
comput
reach
colleg
world
friend
execut
may
launch
sinc
andrew
fellow
develop
internet
net
worth
part
rapid
disput
base
due
februari
upon
decemb
introduc
rank
dormitori
eduardo
legal
phase
year
name
peopl
among