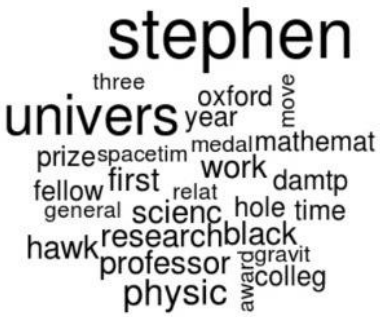
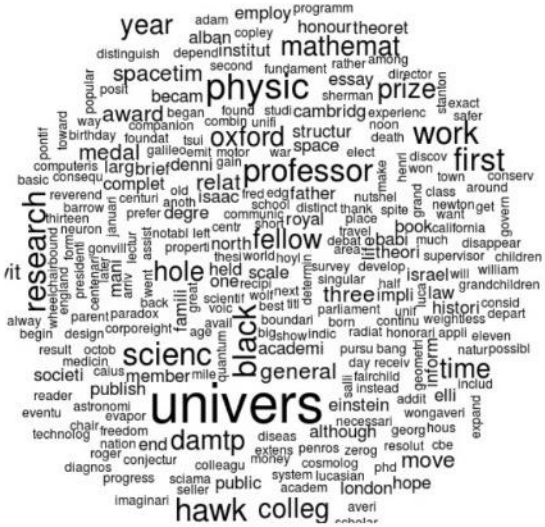
```
# Name : Manvi Pandya

#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)
```

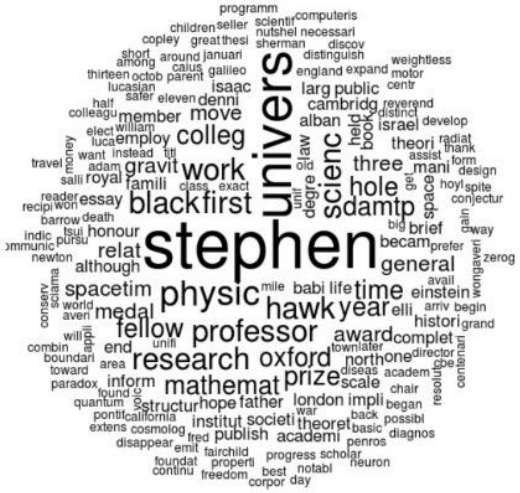



```
wordcloud(words=d$word,freq = d$freq,min.freq = 1)
```



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"))


stephen
univers
physic
research
black
hole
scienc
first
hawk
professor
work
oxford
fellow
mathemat
general
relat
medal
degre
prize
colleg
damtp
year
time
spacetim
award
london
isaac
histori
although
essay
israel
member
move
scale
three
structur
book
honour
societi
theoret
cambridg
public
academi
gravit
north
alban
theori
mani
denni
father
inform
law
publish
royal
space
complet
institut
einstein
employ
impli
becam
cosmolog
paradox
supervisor
california
zerog
presidenti
sherman
conserv
elect
hous
recipi
imaginari
singular
receiv
extens
freedom
short
place
back
nutshel
arriv
debat
diagnosi
galileo
reverend
design
children
gain
among
conjectur
notabl
thesi
toward
sali
england
indic
centr
geometri
radiat
wheelchairbound
weightless
began
fairchild
way
disappear
centenari
companion
system
make
unifi
thirteen
distinguish
caius
parliament
stanton
pursu
spite
penros
eleven
phd
distinct
natur
bang
neuron
grandchildren
eventu
sciama
luca
roger
develop
gramm
erienc
popular
travel
nation
great
discov
depend
progress
fundament
found
posit
war
studi
octob
gonvill
wongaveri
rather
school
basic
william
properti
director
necessari
honorari
includ
corpor
day
possibl
astronomi
lectur
second
continu
addit
survey
quantum
combin
big
prefer
tsui
larg
diseas
end
old
age
famili
life
elli
held
henri
world
show
appli
get
left
one
exact
went
hope
depart
brief
much
will
area
babi
next
avail
anoth
instead