M8L2\_Assignment

Xueyi Fan

July 10, 2016

# Assignment:

Create regular expressions for the patterns below:

* Match any of the following punctuation characters in the ASCII table: !"#$%&'()+,
* Create one regular expression to match all common misspellings of calendar (see <https://en.wikipedia.org/wiki/Wikipedia:Lists_of_common_misspellings/C>)
* Create one regular expression to match any character except line breaks.
* You need to validate a ZIP code (U.S. postal code), allowing both the five-digit and nine-digit (called ZIP+4) formats. The regex should match 02115 and 02115-5515, but not 2115, 2115-5515, 21155515,021155515, etc..
* You need to validate a legit any password for your website. Passwords have the following complexity requirements: Length between 8 and 32 characters, ASCII visible and space characters only, One or more uppercase letters, One or more lowercase letters, One or more special characters (ASCII punctuation)
* Load the file M08\_tweets.csv (it is online at '<http://nikbearbrown.com/YouTube/MachineLearning/M08/M08_tweets.csv>')
* Do the following:
  + Extract a list of the top 9 users (e.g. @NikBearBrown)
  + Extract a list of the top 9 hashtags (e.g. #Bear)
  + Find the top 5 most positve tweets
  + Find the top 5 most negative tweets
  + Create a world cloud of 100 related tweets
  + Which tweets could be classified as game development?

# Answer:

Create regular expressions for the patterns below:

## Match any of the following punctuation characters in the ASCII table: !"#$%&'()+,

regular expression:  
  
{[\!\"\#\$\%\&\'\(\)\+]}g

## Create one regular expression to match all common misspellings of calendar (see <https://en.wikipedia.org/wiki/Wikipedia:Lists_of_common_misspellings/C>)

There are four misspellings of calendar. They are "calandar", "calander", "calender", "colander".  
  
regular expression:  
  
{c[oa]l[ae]nd[ae]r}g

## Create one regular expression to match any character except line breaks.

regular expression:  
  
{.}g

## You need to validate a ZIP code (U.S. postal code), allowing both the five-digit and nine-digit (called ZIP+4) formats. The regex should match 02115 and 02115-5515, but not 2115, 2115-5515, 21155515,021155515, etc..

regular expression:  
  
{\b[0-9]{5}(?:-[0-9]{4})?\b}g

## You need to validate a legit any password for your website. Passwords have the following complexity requirements: Length between 8 and 32 characters, ASCII visible and space characters only, One or more uppercase letters, One or more lowercase letters, One or more special characters (ASCII punctuation)

Regular expression:  
  
#length between 8 and 32  
{^.{8,32}$}  
  
#ASCII visible and space characters only  
{^[\x20-\x7E]+$}  
  
#One or more uppercase letters  
{[A-Z]+}  
  
#one or more lowercase letters  
{[a-z]+}  
  
#one or more special characters  
{[\!\"\#\$\%\&\'\(\)\\*\+\,\\\-\.\/\:\;\<\=\>\?\@\[\\\]\^\\_\`\{\|\}\~]+}

## \* Load the file M08\_tweets.csv (it is online at '<http://nikbearbrown.com/YouTube/MachineLearning/M08/M08_tweets.csv>')

library("tm")

## Loading required package: NLP

library("wordcloud")

## Loading required package: RColorBrewer

library("RTextTools")

## Loading required package: SparseM

##   
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':  
##   
## backsolve

tweets\_url <- "http://nikbearbrown.com/YouTube/MachineLearning/M08/M08\_tweets.csv"  
  
tweets <- read.csv(url(tweets\_url), header = F)  
str(tweets)

## 'data.frame': 85250 obs. of 1 variable:  
## $ V1: Factor w/ 48476 levels " Busquets could miss the game on Wednesday http://t.co/1rwZgYtf9d #MCFC #reddit",..: 1349 41731 47643 18112 39202 37239 47793 42399 41348 46879 ...

Do the following:

### Extract a list of the top 9 users (e.g. @NikBearBrown)

users\_re <- function(line){  
 m <- gregexpr("@([\_a-zA-Z1-9]+)", line)  
 list <- as.list(regmatches(line,m))  
 list  
}  
  
result <- apply(tweets, 1, users\_re)  
length(result)

## [1] 85250

user\_list <- as.list(NULL)  
  
for (i in (1:85250)){  
 if (!is.na(result[[i]]$V1[1])){  
 user\_list <- c(user\_list,result[[i]]$V1)   
 }  
}   
  
sort(table((unlist(user\_list))), decreasing = T)[1:9]

##   
## @YouTube @InfoDuniaBahasa @JuegoDeTronosTM @HoodieAllen   
## 1445 1392 583 493   
## @FreakingTrue @SimiMention\_ @NickiDaily @DarkestDungeon   
## 438 341 312 274   
## @NickiWorldNews   
## 254

The top nine of users are @Youtube, @InfoDuniaBahasa, @JuegoDeTronosTM, @HoodieAllen, @FreakingTrue, @SimiMention, @NickiDaily, @DarkestDungeon, @NickiWorldNews

### Extract a list of the top 9 hashtags (e.g. #Bear)

hashtags\_re <- function(line){  
 m <- gregexpr("#([\_a-zA-Z1-9]+)", line)  
 list <- as.list(regmatches(line,m))  
 list  
}  
  
result\_hashtag <- apply(tweets, 1, hashtags\_re)  
hashtag\_list <- as.list(NULL)  
  
for (i in (1:85250)){  
 if (!is.na(result\_hashtag[[i]]$V1[1])){  
 hashtag\_list <- c(hashtag\_list,result\_hashtag[[i]]$V1)   
 }  
}   
  
sort(table((unlist(hashtag\_list))), decreasing = T)[1:9]

##   
## #gamedev #indiedev #gameinsight #ipadgames #KCA   
## 9190 4465 1596 1029 693   
## #BigData #bigdata #GameDev #indiegame   
## 675 673 669 583

The top nine hashtags are #gamedev, #indiedev, #gameinsight, #ipadgames, #KCA, #BigData, #bigdata, #GameDev, #indiegame

However, among these 9 hashtags, like #gamedev & #GameDev, #BigData & #bigdata, there are actually same content expect the letter case. So I will change all the hashtag to lower case.

hashtag\_list <- tolower(hashtag\_list)  
sort(table((unlist(hashtag\_list))), decreasing = T)[1:9]

##   
## #gamedev #indiedev #gameinsight #bigdata #ipad   
## 10067 4681 1763 1398 1086   
## #ipadgames #kca #game #android   
## 1047 733 728 689

Finally, I got top nine hashtags: #gamedev, #indiedev, #gameinsight, #bigdata, #ipad, #ipadgames, #kca, #game, #android

# Sentiment Analysis and get the polarity score of each document

library("qdap")

## Warning: package 'qdap' was built under R version 3.2.5

## Loading required package: qdapDictionaries

## Loading required package: qdapRegex

## Loading required package: qdapTools

##   
## Attaching package: 'qdap'

## The following objects are masked from 'package:tm':  
##   
## as.DocumentTermMatrix, as.TermDocumentMatrix

## The following object is masked from 'package:NLP':  
##   
## ngrams

## The following object is masked from 'package:base':  
##   
## Filter

library("e1071")  
  
#pre-poccessing the data  
clean\_sub <- function(x){  
 #remove all punctuations  
 gsub("[[:punct:]]", "",  
 #remove all control characters  
 gsub("[[:cntrl:]]", "",  
 #remove all digit  
 gsub("\\d+","",  
 #remove the hashtag  
 gsub("#[\_a-zA-Z1-9]+", "",   
 #remove the users  
 gsub("@([\_a-zA-Z1-9]+)","",  
 #remove the url  
 gsub("(https?://[^ ]+)","",x))))))  
   
}  
  
tweets\_clean <- apply(tweets, 1, clean\_sub)  
tweets\_clean <-as.character(tweets\_clean)  
tweets\_clean <- stripWhitespace(tweets\_clean)  
tweets\_clean <- removeWords(tweets\_clean, stopwords())  
head(tweets\_clean)

## [1] " Im game either"   
## [2] "shoe game Santa Monica kingrickey "   
## [3] "Witness devastation great fast "   
## [4] "Game Started MNT AcW "   
## [5] "RT Kyle Walkers best highlight todays game "  
## [6] "RT Game day Lets go Spurs "

ps <- polarity(tweets\_clean)  
dim(ps$all)

## [1] 85250 6

ps$all[1:10,]

## all wc polarity pos.words neg.words  
## 1 all 3 0.0000000 - -  
## 2 all 5 0.0000000 - -  
## 3 all 4 0.5000000 great, fast devastation  
## 4 all 4 0.0000000 - -  
## 5 all 7 0.3779645 best -  
## 6 all 6 0.0000000 - -  
## 7 all 12 0.2886751 wow -  
## 8 all 12 0.0000000 - -  
## 9 all 14 0.0000000 - -  
## 10 all 10 -0.6324555 - hell, foul  
## text.var  
## 1 Im game either  
## 2 shoe game Santa Monica kingrickey   
## 3 Witness devastation great fast   
## 4 Game Started MNT AcW   
## 5 RT Kyle Walkers best highlight todays game   
## 6 RT Game day Lets go Spurs   
## 7 Wow runs scored scored one inningif game continued Oregon demolished Oregon state  
## 8 Sony PlayStation GB Video Game System Jet Black CUHA Full read eBay   
## 9 Saudades meu amor vir joga um game aki em casa ainda em domingo desse  
## 10 What hell UCONN For second time game foul shot clock

### Find the top 5 most positve tweets

pos\_5 <- order(ps$all[,3], decreasing=T) [1:5]  
pos\_5

## [1] 69189 70876 58289 56576 75032

tweets[pos\_5,]

## [1] @Vauxhall\_Wales @FAWales After that great game against Ireland this would be a great prize to cherish that win   
## [2] Congratulations @chorleyfc. Great game and great result. #Chorley #COYM   
## [3] Great game @NHLJets! Nice work @OndrejPavelec31!   
## [4] @UserDpj32 Great win on Sat. Great game I like it when you play with that motor turn up and relax enough to finish   
## [5] Valiant Hearts - The Great Wars is an exceptionally good game! Great story telling with history #psplus #ValiantHearts  
## 48476 Levels: Busquets could miss the game on Wednesday http://t.co/1rwZgYtf9d #MCFC #reddit ...

According to the polarity scores, the top five most positive tweets are the tweets with highest scores. Their position are 69189, 70876, 58289, 56576, 75032. The tweets are as below:

[1] @Vauxhall\_Wales @FAWales After that great game against Ireland this would be a great prize to cherish that win  
[2] Congratulations @chorleyfc. Great game and great result. #Chorley #COYM  
[3] Great game @NHLJets! Nice work @OndrejPavelec31!  
[4] @UserDpj32 Great win on Sat. Great game I like it when you play with that motor turn up and relax enough to finish

[5] Valiant Hearts - The Great Wars is an exceptionally good game! Great story telling with history #psplus #ValiantHearts

### Find the top 5 most negative tweets

neg\_5 <- order(ps$all[,3], decreasing=F) [1:5]  
neg\_5

## [1] 63370 22294 45076 74408 57731

tweets[neg\_5,]

## [1] RT @Vauncey: Massive PR fail from @atari as they threaten to sue @llamasoft\_ox over TxK http://t.co/M9vHEDah21 #gamedev   
## [2] Pathetic shitty ass fucking joke of a fucking team get bent fuck this game is off u guys r a fucking joke   
## [3] Pathetic shitty ass fucking joke of a fucking team get bent fuck this game is off u guys r a fucking joke   
## [4] @EASPORTSFIFA you guys are fucking bitches and retarded fuck faces. We pay to fucking buy your fucking game and you fucking shit on our face  
## [5] @B\_Rich145 @Finnesotan91 just fucking with ya but sure as hell makes up for that game he lost us   
## 48476 Levels: Busquets could miss the game on Wednesday http://t.co/1rwZgYtf9d #MCFC #reddit ...

According to the polarity scores, the top five most positive tweets are the tweets with highest scores. Their position are 63370 22294 45076 74408 57731. The tweets are as below:

[1] RT @Vauncey: Massive PR fail from @atari as they threaten to sue @llamasoft\_ox over TxK <http://t.co/M9vHEDah21> #gamedev

[2] Pathetic shitty ass fucking joke of a fucking team get bent fuck this game is off u guys r a fucking joke

[3] Pathetic shitty ass fucking joke of a fucking team get bent fuck this game is off u guys r a fucking joke

[4] @EASPORTSFIFA you guys are fucking bitches and retarded fuck faces. We pay to fucking buy your fucking game and you fucking shit on our face

[5] @B\_Rich145 @Finnesotan91 just fucking with ya but sure as hell makes up for that game he lost us

### Create a world cloud of 100 related tweets

Here, I want to use the top 100 positie tweets with the highest polarity score. According to the previous results, I have calculated all the score of tweets. So I choose the top 100 tweets.

#Get the 100 tweets  
tweets\_100\_index <- order(ps$all[,3], decreasing=T) [1:100]  
tweets\_100\_index

## [1] 69189 70876 58289 56576 75032 50851 63332 66389 50717 55193 4389  
## [12] 4669 27171 27451 63906 58088 1719 24501 49680 56480 71769 51392  
## [23] 63342 15247 15349 16239 18121 22289 38029 38131 39021 40903 45071  
## [34] 68362 75847 77890 53171 51562 73026 73245 74110 11640 34422 65187  
## [45] 65189 65192 65209 76346 75432 488 19194 23270 41976 57218 57670  
## [56] 80706 80792 82046 8707 31489 15795 38577 20873 43655 2496 25278  
## [67] 3616 26398 77054 21436 44218 72565 1417 17907 24199 40689 55027  
## [78] 77572 6466 29248 66547 78511 50602 70166 17381 40163 76632 55168  
## [89] 3300 26082 47790 58020 64665 69787 71036 9110 31892 77060 15527  
## [100] 38309

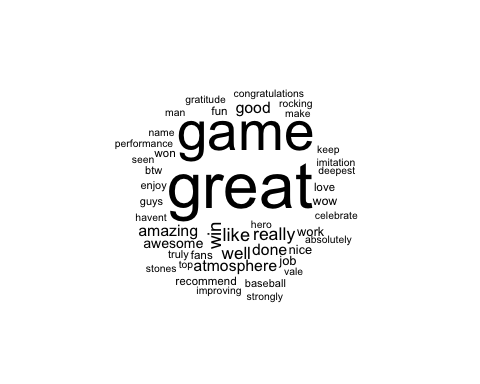
tweets\_100 <- tweets\_clean[tweets\_100\_index]  
str(tweets\_100)

## chr [1:100] " After great game Ireland great prize cherish win" ...

#create corpus  
corpus <-Corpus(VectorSource(tweets\_100))  
  
#Convert to lower-case  
corpus <- tm\_map(corpus, tolower)  
  
#remove stopwords  
corpus <- tm\_map(corpus, function(x)removeWords(x,stopwords()))  
  
#convert corpus to a plain text document  
corpus <- tm\_map(corpus, PlainTextDocument)  
  
inspect(corpus)

## <<VCorpus>>  
## Metadata: corpus specific: 0, document level (indexed): 0  
## Content: documents: 100  
##   
## [[1]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 52  
##   
## [[2]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 41  
##   
## [[3]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 21  
##   
## [[4]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 76  
##   
## [[5]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 82  
##   
## [[6]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 63  
##   
## [[7]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 61  
##   
## [[8]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 95  
##   
## [[9]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 46  
##   
## [[10]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 42  
##   
## [[11]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 50  
##   
## [[12]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 50  
##   
## [[13]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 50  
##   
## [[14]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 50  
##   
## [[15]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 61  
##   
## [[16]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 57  
##   
## [[17]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 29  
##   
## [[18]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 29  
##   
## [[19]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 30  
##   
## [[20]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 69  
##   
## [[21]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 53  
##   
## [[22]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 65  
##   
## [[23]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 76  
##   
## [[24]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[25]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[26]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[27]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[28]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[29]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[30]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[31]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[32]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[33]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[34]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[35]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[36]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 50  
##   
## [[37]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 41  
##   
## [[38]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 45  
##   
## [[39]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 85  
##   
## [[40]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 56  
##   
## [[41]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 56  
##   
## [[42]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 33  
##   
## [[43]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 33  
##   
## [[44]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 37  
##   
## [[45]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 37  
##   
## [[46]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 37  
##   
## [[47]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 37  
##   
## [[48]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 24  
##   
## [[49]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 90  
##   
## [[50]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 43  
##   
## [[51]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 49  
##   
## [[52]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 43  
##   
## [[53]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 49  
##   
## [[54]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 57  
##   
## [[55]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 44  
##   
## [[56]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 41  
##   
## [[57]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 41  
##   
## [[58]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 41  
##   
## [[59]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 88  
##   
## [[60]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 88  
##   
## [[61]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 89  
##   
## [[62]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 89  
##   
## [[63]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 58  
##   
## [[64]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 58  
##   
## [[65]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 92  
##   
## [[66]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 92  
##   
## [[67]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 37  
##   
## [[68]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 37  
##   
## [[69]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 33  
##   
## [[70]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 61  
##   
## [[71]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 61  
##   
## [[72]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 86  
##   
## [[73]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 61  
##   
## [[74]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 40  
##   
## [[75]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 61  
##   
## [[76]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 40  
##   
## [[77]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 40  
##   
## [[78]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 49  
##   
## [[79]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 27  
##   
## [[80]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 27  
##   
## [[81]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 94  
##   
## [[82]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 85  
##   
## [[83]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 26  
##   
## [[84]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 29  
##   
## [[85]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 95  
##   
## [[86]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 95  
##   
## [[87]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 86  
##   
## [[88]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 88  
##   
## [[89]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 46  
##   
## [[90]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 46  
##   
## [[91]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 70  
##   
## [[92]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 51  
##   
## [[93]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 52  
##   
## [[94]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 50  
##   
## [[95]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 43  
##   
## [[96]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 44  
##   
## [[97]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 44  
##   
## [[98]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 35  
##   
## [[99]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 26  
##   
## [[100]]  
## <<PlainTextDocument>>  
## Metadata: 7  
## Content: chars: 26

tweets.tdm <- TermDocumentMatrix(corpus,control=list(removePunctuation=TRUE, tolower=T, minWordLength=1, stopwords(kind="en")))  
m <- as.matrix(tweets.tdm)  
v<- sort(rowSums(m), decreasing=T)  
myNames <- names(v)  
d <- data.frame(word=myNames,freq=v)  
wordcloud(d$word, d$freq, min.freq=4, random.color=T,random.order = F)



### Which tweets could be classified as game development?

I will use key words: #gamedev, game development to find the tweets

#use regular expression to find tweets containing "game develpment"  
#remove all url  
#pre-poccessing the data  
clean\_sub <- function(x){  
 #remove the url  
 gsub("(https?://[^ ]+)","",x)  
}  
tweets\_no\_url <- apply(tweets, 1, clean\_sub)  
#to alter all tweets as lower case   
tweets\_no\_url <- as.data.frame(tolower(tweets\_no\_url))  
  
head(tweets\_no\_url)

## tolower(tweets\_no\_url)  
## 1 @big9incherforu @hardnfast12345 @gemsdoubledees i'm game for either!  
## 2 shoe game s/o @sirjonesmag santa monica kingrickey20 n8o100   
## 3 witness the devastation of the #bombs, great for #exploring fast! #terraria #gamedev   
## 4 game started | mnt at acw   
## 5 rt @simply\_spurs: kyle walker's best highlight from today's game   
## 6 rt @marcobelinelli: game day!! lets go spurs!!!! #gospursgo

find\_gamedev <- function(x){  
 grepl("game[ ]?dev", x)  
}  
  
has\_gamedev <- apply(tweets\_no\_url, 1,find\_gamedev)  
tweets\_gamedev <- tweets[which(has\_gamedev),]  
  
#among tweets\_gamedev, remove all tweets with "game developer"  
find\_developer <- function(x){  
 grepl("game developer", x)  
}  
tweets\_gamedev <- as.data.frame(tweets\_gamedev)  
has\_developer <- apply(tweets\_gamedev, 1,find\_developer)  
game\_dev <- tweets\_gamedev[which(!has\_developer),]  
  
length(game\_dev)

## [1] 10144

head(game\_dev,100)

## [1] Witness the devastation of the #bombs, great for #exploring fast! #terraria #gamedev http://t.co/p2DVQq9323   
## [2] RT @AV\_metaldemon: Here's a trailer video thingy for my #7DRL game: https://t.co/Axz7G63Eqz #gamedev cc @Enichan   
## [3] #Programming #GameDev #UE4 @UnrealEngine   
## [4] RT @JoeCreates: Going to try to animate that smoke. http://t.co/rKrM1TNNSc #pixelart #gamedev http://t.co/lw3TG84XsL   
## [5] #Programming #GameDev #UE4 @UnrealEngine   
## [6] #TRPSGAME #gamedev #indiedev #b3d #alldev http://t.co/uOAXRPswlR   
## [7] @akanimh We'd love for you to play the demo too! Let us know what you think: http://t.co/b6cGuG7QYO #indiedev #gamedev   
## [8] RT @LTP\_ATS: More gross shaders on low poly critters #gamedev http://t.co/pyhw8oForu   
## [9] RT @IraGame: @akanimh We'd love for you to play the demo too! Let us know what you think: http://t.co/b6cGuG7QYO #indiedev #gamedev   
## [10] New online game launch set: MAATA Games, a Netherland-based game development company, toget... http://t.co/N82Tl5iiBB #outsourcing, #bpo   
## [11] Sneak up on your prey with Invisibility in "That Rock Paper Scissor Game!" #TRPSGAME #gamedev #alldev http://t.co/USfSuKb9Gl   
## [12] RT @LTP\_ATS: More gross shaders on low poly critters #gamedev http://t.co/pyhw8oForu   
## [13] RT @Ogardonix: Do you prefer dwarfs or goblins? Enter the world of @Ralingame http://t.co/QjIKJfdKZm #indiedev #gamedev   
## [14] @AuerbachKeller "We also libel millions of gamers for clicks and threaten to smear game devs who try to stand up for them"-- @voxdotcom   
## [15] #TRPSGAME #gamedev #indiedev #b3d #alldev http://t.co/uOAXRPswlR   
## [16] RT @JoeCreates: Going to try to animate that smoke. http://t.co/rKrM1TNNSc #pixelart #gamedev http://t.co/lw3TG84XsL   
## [17] I forgot to ask so if any #gamedev's out there have cool Android games they're having issues getting coverage for: mc -at- gamewalkers . com   
## [18] Concept I am working on... started it a while ago but been busy... #indiedev #gamedev http://t.co/9GnPF5Ipax   
## [19] #gamedev is hard, help out a #indiedev. Share and up vote our @reddit post. http://t.co/KZk1x3k4NF   
## [20] RT @AJohanson89: Concept I am working on... started it a while ago but been busy... #indiedev #gamedev http://t.co/9GnPF5Ipax   
## [21] #TRPSGAME #gamedev #indiedev #b3d #alldev http://t.co/uOAXRPswlR   
## [22] RT @armedgelatinous: #gamedev is hard, help out a #indiedev. Share and up vote our @reddit post. http://t.co/KZk1x3k4NF   
## [23] #gamedev #gamedesign #indie #indiedev #games #design   
## [24] RT @armedgelatinous: #gamedev is hard, help out a #indiedev. Share and up vote our @reddit post. http://t.co/KZk1x3k4NF   
## [25] @starrmazer: Booting up Ronny Raygon for #sxsw! @robotloveskitty #indiedev #gamedev http://t.co/izZUy0yfQS   
## [26] The Achievement list is completed for my IOS Game, all that's left are the levels, and testing. #indiedev #gamedev #UnrealEngine #iOS82   
## [27] RT @armedgelatinous: #gamedev is hard, help out a #indiedev. Share and up vote our @reddit post. http://t.co/KZk1x3k4NF   
## [28] RT @AJohanson89: Concept I am working on... started it a while ago but been busy... #indiedev #gamedev http://t.co/9GnPF5Ipax   
## [29] RT @RobotLovesKitty: "@starrmazer: Booting up Ronny Raygon for #sxsw! @robotloveskitty #indiedev #gamedev http://t.co/izZUy0yfQS"   
## [30] This 90sec trailer takes the classic Rock Paper Scissor to the next level! #gamedev #TRPSGAME #alldev #indiedev http://t.co/M7nVid8Xx4   
## [31] RT @AJohanson89: Concept I am working on... started it a while ago but been busy... #indiedev #gamedev http://t.co/9GnPF5Ipax   
## [32] Agu The Game LightSaber Attack #gamedev #indiedev #screenshotsaturday #graphicdesign #photoshop #illustrator #ux #UI http://t.co/cYHOsQwkCq   
## [33] RT @MediaL0C: For those who missed it on Saturday - How British video games became a billion pound industry http://t.co/wZCsqi6pQ1 #gamedev   
## [34] #Programming #GameDev #UE4 @UnrealEngine   
## [35] @tree\_fortress As a game dev studio, thought you'd like "#GameDev w/ @ScruffyFurn" on IoT: http://t.co/4niVqQVDEE   
## [36] Pls vote with 1 click for Rize @pitchatpalace - http://t.co/DjeFSwR3Eq every click counts! #gamedev #health #MondayMotivation #entrepreneur   
## [37] RT @cdndevs: @tree\_fortress As a game dev studio, thought you'd like "#GameDev w/ @ScruffyFurn" on IoT: http://t.co/4niVqQVDEE   
## [38] RT @xDavidLeon: Canvaleon, the stealthy platformer from @OXiABgs is coming soon to Wii U! Yay :D http://t.co/OsoRTWLWIC #gamedev   
## [39] RT @RobotLovesKitty: "@starrmazer: Booting up Ronny Raygon for #sxsw! @robotloveskitty #indiedev #gamedev http://t.co/izZUy0yfQS"   
## [40] RT @cdndevs: @tree\_fortress As a game dev studio, thought you'd like "#GameDev w/ @ScruffyFurn" on IoT: http://t.co/4niVqQVDEE   
## [41] Drawing the interior of a lighthouse... I'm assuming it looks like the interior of a nuclear submarine. #gamedev #indiedev #lighthouse   
## [42] RT @gamasutra: Timeless: 20 open-world games that every #gamedev should play/study [2007] - http://t.co/gqjAyjeS1q http://t.co/7mwi93oKuc   
## [43] Ep.11 of #Divercast Multiplayer Judgement http://t.co/J7MG7zLRnU #GameDev #Indiedev And be sure to send us your questions at #AskDDD   
## [44] A pic of the art style for our game. We are going for Castle Crashers mixed with Mario. Enjoy! #GameDev #IndieGame http://t.co/iGVrrIp1vc   
## [45] RT @tigrisgames: New game screenshots on http://t.co/5QmQl5otD8 (Tigris Games on #Facebook) #gamedev #indiedev   
## [46] #indiedev #gamedev #unity3d #unity   
## [47] RT @shipsandmonster: Help us decide The Huntress' New Haircolor! http://t.co/BGtuclGWL9 #mobilegame #gamedev #haircolor #shipsandmonsters   
## [48] This skeleton raid ended in flames. #gamedev #gaming #skeletons #games http://t.co/cPQjlaG2yA   
## [49] #indiedev #gamedev #unity3d #unity   
## [50] RT @SimpleFunGames: This skeleton raid ended in flames. #gamedev #gaming #skeletons #games http://t.co/cPQjlaG2yA   
## [51] RT @SimpleFunGames: This skeleton raid ended in flames. #gamedev #gaming #skeletons #games http://t.co/cPQjlaG2yA   
## [52] RT @TallGamer: Kiddos watching the video I am working on for my channel - http://t.co/zLcheMP5Xe #gamedev #indiedev http://t.co/trhTuAtWwL   
## [53] HyperGalaxyTeam: We're adding robots to The Specimen. Wanna see? #indiedev #gamedev #gamers #gameart http://t.co/ppeEiHROpg   
## [54] MichalIvanceky: looking through some old game concepts we've had #gamedev #indiedev http://t.co/ZVjKM4sgwX   
## [55] #indiedev #gamedev #unity3d #unity   
## [56] tigrisgames: New game screenshots on http://t.co/k6aPiDeGw8; (Tigris Games on #Facebook) #gamedev #indiedev   
## [57] RT @HyperGalaxyTeam: We're adding robots to The Specimen. Wanna see? #indiedev #gamedev #gamers #gameart http://t.co/Asxh5QK2cl   
## [58] SimpleFunGames: This skeleton raid ended in flames. #gamedev #gaming #skeletons #games http://t.co/QVnB7qyjrv   
## [59] TallGamer: Kiddos watching the video I am working on for my channel - http://t.co/TsMksqp7ap; #gamedev #indiedev   
## [60] RT @Allegorithmic: Substance Painter 1.3 GDC demo: Last week's stream, now in HD! #gamedev #gameart https://t.co/HMTZOk9H5g   
## [61] RT @TopcatsLair: #GamerGate No respect for game devs. Kotaku making clickbait stories out of literally nothing. http://t.co/SB9OWi5ECs   
## [62] #indiedev #gamedev #unity3d #unity   
## [63] RT @Parodossy: MichalIvanceky: looking through some old game concepts we've had #gamedev #indiedev http://t.co/ZVjKM4sgwX   
## [64] RT @Parodossy: HyperGalaxyTeam: We're adding robots to The Specimen. Wanna see? #indiedev #gamedev #gamers #gameart http://t.co/ppeEiHROpg   
## [65] RT @SimpleFunGames: This skeleton raid ended in flames. #gamedev #gaming #skeletons #games http://t.co/cPQjlaG2yA   
## [66] Enjoy retro indie games? Like our Facebook page for a chance to win! http://t.co/vQGuXfNE4c #indiedev #gamedev http://t.co/cUK1olQ8EO   
## [67] RT @vladanese: It would be nice someday finish this idea. A parody. Im just a beginner for now. XD #pixelart #gamedev http://t.co/PpEkQhfAs1   
## [68] RT @Parodossy: tigrisgames: New game screenshots on http://t.co/k6aPiDeGw8; (Tigris Games on #Facebook) #gamedev #indiedev   
## [69] @KillLaGoon @CHOBITCOIN @legobutts man I hate feeling unethical whenever I play grab-a-dick with five guys to further my work as a game dev.   
## [70] RT @Parodossy: TallGamer: Kiddos watching the video I am working on for my channel - http://t.co/TsMksqp7ap; #gamedev #indiedev   
## [71] RT : MichalIvanceky: looking through some old game concepts we've had #gamedev #indiedev http://t.co/YFnErjBa7Q   
## [72] #indiedev #gamedev #unity3d #unity   
## [73] RT @ECbotOfficial: RT : MichalIvanceky: looking through some old game concepts we've had #gamedev #indiedev http://t.co/YFnErjBa7Q   
## [74] RT : SimpleFunGames: This skeleton raid ended in flames. #gamedev #gaming #skeletons #games http://t.co/MYZLFrnJaB   
## [75] #indiedev #gamedev #unity3d #unity   
## [76] RT : This skeleton raid ended in flames. #gamedev #gaming #skeletons #games http://t.co/MYZLFrnJaB   
## [77] Making a lot of buildings for the last "room" of the tutorial level. First one #CODeX #gamedev #indiedev http://t.co/Q7rzFeFSpd   
## [78] #WHITEPAPER: Increase the ROI with the Help of Mobile Test Automation http://t.co/AczEOzv0Ps #AppDev #GameDev #WebDev   
## [79] RT @FoxTwice: A-STAR Pathfinding AI for HTML5 Canvas Games http://t.co/rdLprDkXfA #gameDev   
## [80] RT @Parodossy: Lewpen: Alien Terrain .. #gamedev #webgl #screenshotsaturday http://t.co/wDyjQthUVX   
## [81] 3 NJIT young alum form a game development company and fund a successful Kickstarter! http://t.co/2OH39OlQOG #AlumMonday #NJIT #sneakyninja   
## [82] Some Two Brothers news coming this week. #gamedev   
## [83] RT @bitbar: #WHITEPAPER: Increase the ROI with the Help of Mobile Test Automation http://t.co/AczEOzv0Ps #AppDev #GameDev #WebDev   
## [84] RT @OfficialQSR: Shout out to @ShamsJorjani who treats his customers and Players the best! #gamedev #indiedev need to learn from him.   
## [85] RT @AckkStudios: Some Two Brothers news coming this week. #gamedev   
## [86] RT @bitbar: #WHITEPAPER: Increase the ROI with the Help of Mobile Test Automation http://t.co/AczEOzv0Ps #AppDev #GameDev #WebDev   
## [87] @runefondling are you trying to do gamedev for money?   
## [88] In the middle of prototyping new ideas, my favorite stage of game dev; really allows the creative juices to flow... Anyone have any ideas?   
## [89] @liamtwose @ProNoob13 You think the situation changed much? There were already free options &amp; #gamedev still requires incredible persistence  
## [90] RT @esperware: Official web site now open! http://t.co/wQfvzlRhR6 #indiegame #indiegamedev #gamedev #gaming @Clickteam @FusionScene   
## [91] #MMO #indiedev #gamedev #indiegame   
## [92] That awkward moment when someone asks to see your work, and everything you've created gives off a sense of utter insanity. #gamedev   
## [93] RT @Parodossy: MichalIvanceky: looking through some old game concepts we've had #gamedev #indiedev http://t.co/ZVjKM4sgwX   
## [94] #MMO #indiedev #gamedev #indiegame   
## [95] RT @bitbar: How to Instantly Performance Profile Any Android Game http://t.co/Nk60toUULm #GameDev #indiedev http://t.co/NZqf9GBsDv   
## [96] Sometimes skeletons end in flames. Other times they are killed by lizard people: http://t.co/OuRhyQT921 #gamedev #gaming #games   
## [97] RT @bitbar: #WHITEPAPER: Increase the ROI with the Help of Mobile Test Automation http://t.co/AczEOzv0Ps #AppDev #GameDev #WebDev   
## [98] RT @AckkStudios: Some Two Brothers news coming this week. #gamedev   
## [99] Game developer: &amp;quot;I decided to prove her wrong&amp;quot; #kidslit http://t.co/AwHq9jQq8v   
## [100] RT DarkSpaceGames: It's alive! Checkout http://t.co/t2sf2xCvoe and let me know what you guys think. #gamedev (via TheIndieSloth)   
## 48476 Levels: Busquets could miss the game on Wednesday http://t.co/1rwZgYtf9d #MCFC #reddit ...

I find 10144 tweets contains key word #gamedev, or game development, not game developer. For example:

[1] Witness the devastation of the #bombs, great for #exploring fast! #terraria #gamedev <http://t.co/p2DVQq9323>

[2] RT @AV\_metaldemon: Here's a trailer video thingy for my #7DRL game: <https://t.co/Axz7G63Eqz> #gamedev cc @Enichan