amazon

dz1397

12/2/2020

```
library(quanteda)
## Package version: 2.1.2
## Parallel computing: 2 of 12 threads used.
## See https://quanteda.io for tutorials and examples.
## Attaching package: 'quanteda'
## The following object is masked from 'package:utils':
##
##
       View
library(tidyr)
library(tidytext)
library(wordcloud)
## Loading required package: RColorBrewer
library(tm)
## Loading required package: NLP
##
## Attaching package: 'NLP'
## The following objects are masked from 'package:quanteda':
##
##
       meta, meta<-
## Attaching package: 'tm'
  The following objects are masked from 'package:quanteda':
##
##
       as.DocumentTermMatrix, stopwords
```

PART 1: Exploratory Data Analysis & Text Analytics

```
data1 <- read.csv("USvideos.csv")</pre>
```

1. How many complaints have been generated?

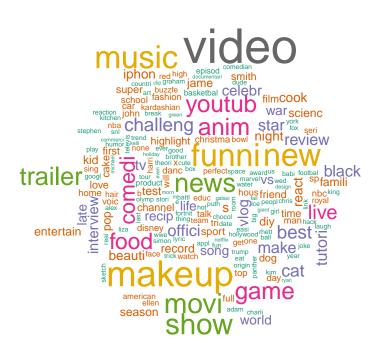
```
dim(data1)
```

```
## [1] 40949
                16
colnames (data1)
## [1] "video_id"
                                 "trending_date"
                                                          "title"
## [4] "channel_title"
                                 "category_id"
                                                          "publish_time"
## [7] "tags"
                                 "views"
                                                          "likes"
## [10] "dislikes"
                                 "comment_count"
                                                          "thumbnail_link"
## [13] "comments_disabled"
                                 "ratings_disabled"
                                                          "video_error_or_removed"
## [16] "description"
process_list<- list("description","tags","title")</pre>
precorpus <- data1
dim(precorpus) # dim of file
## [1] 40949
for (i in process list){
  precorpus[[i]] <- gsub("'", "", precorpus[[i]]) # remove apostrophes</pre>
precorpus[[i]] <- gsub("[[:cntrl:]]", " ", precorpus[[i]]) # replace control characters with space</pre>
precorpus[[i]] <- gsub("^[[:space:]]+", "", precorpus[[i]]) # remove whitespace at beginning of documen</pre>
precorpus[[i]] <- gsub("[[:space:]]+$", "", precorpus[[i]]) # remove whitespace at end of documents
precorpus[[i]] <- gsub("[^a-zA-Z -]", " ", precorpus[[i]]) # allows only letters</pre>
precorpus[[i]] <- tolower(precorpus[[i]])</pre>
newscorpus_des<- corpus(precorpus$description,</pre>
                    docnames=precorpus$X,
                    docvar=data.frame(pos=precorpus$video_id,
                                      date=precorpus$trending_date,
                                      loc = precorpus$views))
newscorpus_tag<- corpus(precorpus$tags,</pre>
                    docnames=precorpus$X,
                    docvar=data.frame(pos=precorpus$video_id,
                                      date=precorpus$trending_date,
                                      loc = precorpus$views))
newscorpus_title<- corpus(precorpus$title,</pre>
                    docnames=precorpus$X,
                    docvar=data.frame(pos=precorpus$video id,
                                      date=precorpus$trending_date,
                                      loc = precorpus$views))
newscorpus<-newscorpus_des
#explore the corpus
#summary(newscorpus) #summary of corpus
# create a custom dictonary
swlist = c("www", "http", "https", "n", "com")
dfm.stem<- dfm(newscorpus,</pre>
               remove = c(swlist,stopwords("english")),
               verbose=F,
               stem=TRUE)
```

```
topfeatures(dfm.stem, n=50)
##
           ly
                     bit
                              youtub
                                        twitter instagram
                                                                 video
                                                                         facebook
##
                               48900
                                                      38833
        54462
                    51884
                                          41557
                                                                 37651
                                                                             35484
##
                                          watch
                                                    nfollow
       nhttps
                       s
                               nhttp
                                                                    us
                                                                               goo
                               24997
                                          20410
##
        28966
                    27429
                                                      17901
                                                                 17773
                                                                             16810
##
           gl
                        С
                                   v
                                          music
                                                    channel
                                                                   new
                                                                             youtu
##
                                          16017
                                                      15983
        16785
                    16483
                               16266
                                                                 14927
                                                                             13432
##
                                       smarturl
                       t
                            subscrib
                                                       show
                                                                  amzn
                                                                               use
          get
##
        13375
                    13074
                               13065
                                          12533
                                                      12017
                                                                 11152
                                                                             10967
##
     ntwitter ninstagram
                                make
                                           like
                                                       list nsubscrib nfacebook
##
        10923
                    10705
                               10661
                                          10439
                                                      10380
                                                                 10143
                                                                              9782
##
                 product
                                news
                                                                  time
                                                                              live
          can
                                            one
                                                        now
                                                                              8355
##
         9426
                    9392
                                9247
                                           8660
                                                       8611
                                                                  8507
##
         love
                               world
                                         tumblr
                                                                nwatch
                                                                              link
                      go
                                                          m
##
         7959
                                           7613
                                                                  7082
                                                                              7035
                    7802
                                7777
                                                       7237
##
         nthe
##
         6950
dfm.simple<- dfm(newscorpus,</pre>
               remove = c(swlist,stopwords("english")),
               verbose=F,
               stem=F)
set.seed(142)
                #keeps cloud' shape fixed
freq<-topfeatures(dfm.simple, n=500)</pre>
wordcloud(names(freq),
          freq, max.words=200,
          scale=c(2, .7),
          colors=brewer.pal(8, "Dark2"))
```

```
products world playknow facebook products world playknow facebook products world playknow facebook products world playknow full repinterest nfollow channel full repinterest nfollow channel producer news itunes shop watching online with production social nsnapchatuse amazon voice home family need every thanks series nmy never late episodes content can netflixnfor pgx much it next both on e film can netflixnfor pgx much it next both on news watching online episode on neverlate episodes content can netflixnfor pgx much it today ever make today ever make watch sinc wantone depisode both on nike income and it is now the play it is now the play it is now to camerawfi think ox song to come the play it is now to camerawfi think album best playing it is now to camerawfi think album best playing it is now to check back im nabout jimmy makeup eliqid w see big m both or check back im nabout jimmy movie via also life fox available please free way live game link not inght go apple things star time youtunninstagram
```

```
newscorpus<-newscorpus_tag
swlist = c("www", "http", "https", "n", "com")
dfm.stem<- dfm(newscorpus,
                remove = c(swlist,stopwords("english")),
                verbose=F,
                stem=TRUE)
topfeatures(dfm.stem, n=50)
                                                                 show
##
       video
                  makeup
                              funni
                                                                            movi
                                         music
                                                      new
                                                                                       news
##
       10938
                    7958
                               7323
                                          7307
                                                      7108
                                                                 6412
                                                                            6202
                                                                                       5889
##
     trailer
                  youtub
                                           anim
                                                      food
                                                               comedi
                                                                            live
                                                                                       best
                               game
##
                               5194
                                          5155
                                                                                       4073
         5590
                    5261
                                                      4977
                                                                 4768
                                                                            4317
##
                             offici
    challeng
                     cat
                                        tutori
                                                   review
                                                                black
                                                                            star
                                                                                       make
##
         3905
                    3786
                               3526
                                          3485
                                                      3413
                                                                 3325
                                                                            3314
                                                                                       3288
##
   interview
                    vlog
                             celebr
                                             tv
                                                    world
                                                                   vs
                                                                            life
                                                                                      recip
##
         3226
                    3162
                               3104
                                          2942
                                                      2911
                                                                 2906
                                                                            2893
                                                                                       2877
##
         late
                    song
                                war
                                         react
                                                   beauti
                                                                    s
                                                                            cook
                                                                                      night
##
         2819
                    2818
                               2741
                                          2727
                                                      2720
                                                                 2680
                                                                            2647
                                                                                       2643
##
       iphon
                                           kid
                    test
                                                   season
                                                                  div entertain
                                                                                       jame
                                pop
         2618
                    2595
                                          2585
                                                      2576
##
                               2591
                                                                 2557
                                                                            2512
                                                                                       2466
##
      record
                  famili
         2455
                    2351
##
set.seed(142)
                  #keeps cloud' shape fixed
freq<-topfeatures(dfm.stem, n=500)</pre>
```



```
newscorpus<-newscorpus_title
#explore the corpus
#summary(newscorpus) #summary of corpus
# create a custom dictonary
swlist = c("www", "http","https", "n", "com")
dfm.stem<- dfm(newscorpus,</pre>
                remove = c(swlist,stopwords("english")),
                verbose=F,
                stem=TRUE)
topfeatures (dfm.stem, n=50)
##
     offici
                video trailer
                                       ft
                                                 s
                                                        make
                                                                    VS
                                                                             new
##
       4039
                 2952
                           2258
                                     1315
                                               1192
                                                        1072
                                                                  1046
                                                                            1008
##
       live
                music
                        makeup
                                    audio
                                                          hd
                                                                 first
                                                                            star
                                               day
##
        983
                  941
                            876
                                      872
                                                         796
                                                                   762
                                                                             754
                                               817
##
                                                                            full
        get
                 game
                            tri challeng
                                               show
                                                        time
                                                                   war
##
        749
                  749
                            700
                                                                             621
                                      675
                                               666
                                                         665
                                                                   641
##
       love
                 movi
                          lyric
                                    black
                                             world
                                                        test
                                                                   one
                                                                          season
##
        615
                  609
                            607
                                      593
                                               580
                                                         547
                                                                   546
                                                                             520
```

```
##
               review
                           talk
                                               best
                                                       teaser
                                                                   like christma
       vear
                                    react
##
        515
                  505
                            482
                                      480
                                                468
                                                          462
                                                                    445
                                                                              443
##
       look
                 life
                           will
                                    super
                                                  х
                                                        watch
                                                                   jame
                                                                              cat
        430
                  427
                            423
                                      422
                                                          413
                                                                    405
                                                                              405
##
                                                416
##
       face
                 last
##
        405
                  395
set.seed(142)
                 #keeps cloud' shape fixed
freq<-topfeatures(dfm.stem, n=500)</pre>
wordcloud(names(freq),
           freq, max.words=200,
           scale=c(3, .7),
           colors=brewer.pal(8, "Dark2"))
```

```
jame tour everyth highlight school

ball wire perform Cat kardashian interview let know open game got music bad trump battl appl old like wrong super open game got music bad trump battl appl old like wrong super smith of cast dead one wrong super live diyjohn topfriend tutori danc girl to call life trailer ever short real adam to call life trailer ever short real adam one were come fan theori tv babi in one show graceblack beauti of giant teaser light break room war goe perfect infin season true meghan of the show part of the show part of the show part of the show part of the show week panther answer dont makeup of the show part of the show of th
```