text\_mining\_viz.R

Cameron Willliams

2020-02-15

library(tm)

## Loading required package: NLP

library(wordcloud)

## Loading required package: RColorBrewer

samplecsv <- read.csv("sample.csv", stringsAsFactors = FALSE)  
head(samplecsv)

## tweet\_id airline\_sentiment airline\_sentiment\_confidence negativereason  
## 1 5.7e+17 neutral 1.0000   
## 2 5.7e+17 positive 0.3486   
## 3 5.7e+17 neutral 0.6837   
## 4 5.7e+17 negative 1.0000 Bad Flight  
## 5 5.7e+17 negative 1.0000 Can't Tell  
## 6 5.7e+17 negative 1.0000 Can't Tell  
## negativereason\_confidence airline airline\_sentiment\_gold name  
## 1 NA Virgin America NA cairdin  
## 2 0.0000 Virgin America NA jnardino  
## 3 NA Virgin America NA yvonnalynn  
## 4 0.7033 Virgin America NA jnardino  
## 5 1.0000 Virgin America NA jnardino  
## 6 0.6842 Virgin America NA jnardino  
## negativereason\_gold retweet\_count  
## 1 NA 0  
## 2 NA 0  
## 3 NA 0  
## 4 NA 0  
## 5 NA 0  
## 6 NA 0  
## text  
## 1 @VirginAmerica What @dhepburn said.  
## 2 @VirginAmerica plus you've added commercials to the experience... tacky.  
## 3 @VirginAmerica I didn't today... Must mean I need to take another trip!  
## 4 @VirginAmerica it's really aggressive to blast obnoxious "entertainment" in your guests' faces &amp; they have little recourse  
## 5 @VirginAmerica and it's a really big bad thing about it  
## 6 @VirginAmerica seriously would pay $30 a flight for seats that didn't have this playing.\nit's really the only bad thing about flying VA  
## tweet\_coord tweet\_created tweet\_location user\_timezone  
## 1 2/24/2015 11:35 Eastern Time (US & Canada)  
## 2 2/24/2015 11:15 Pacific Time (US & Canada)  
## 3 2/24/2015 11:15 Lets Play Central Time (US & Canada)  
## 4 2/24/2015 11:15 Pacific Time (US & Canada)  
## 5 2/24/2015 11:14 Pacific Time (US & Canada)  
## 6 2/24/2015 11:14 Pacific Time (US & Canada)

samplecsv <- samplecsv$text  
head(samplecsv)

## [1] "@VirginAmerica What @dhepburn said."   
## [2] "@VirginAmerica plus you've added commercials to the experience... tacky."   
## [3] "@VirginAmerica I didn't today... Must mean I need to take another trip!"   
## [4] "@VirginAmerica it's really aggressive to blast obnoxious \"entertainment\" in your guests' faces &amp; they have little recourse"   
## [5] "@VirginAmerica and it's a really big bad thing about it"   
## [6] "@VirginAmerica seriously would pay $30 a flight for seats that didn't have this playing.\nit's really the only bad thing about flying VA"

words.vec <- VectorSource(samplecsv)  
words.corpus <- Corpus(words.vec)  
  
words.corpus

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 366

words.corpus <- tm\_map(words.corpus,  
 content\_transformer(tolower))

## Warning in tm\_map.SimpleCorpus(words.corpus, content\_transformer(tolower)):  
## transformation drops documents

words.corpus <- tm\_map(words.corpus,  
 removePunctuation)

## Warning in tm\_map.SimpleCorpus(words.corpus, removePunctuation): transformation  
## drops documents

words.corpus <- tm\_map(words.corpus,removeNumbers)

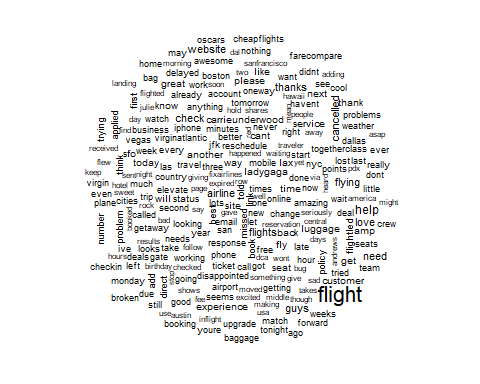
## Warning in tm\_map.SimpleCorpus(words.corpus, removeNumbers): transformation  
## drops documents

words.corpus <- tm\_map(words.corpus, removeWords,c(stopwords("english"),"can", "just"))

## Warning in tm\_map.SimpleCorpus(words.corpus, removeWords,  
## c(stopwords("english"), : transformation drops documents

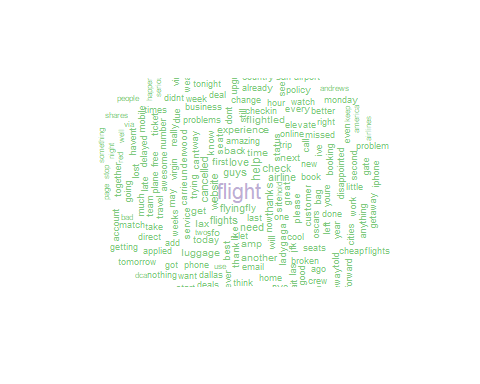
tdm <- TermDocumentMatrix(words.corpus)   
  
  
m <- as.matrix(tdm)  
  
  
wordcounts <- rowSums(m)  
wordcounts <- sort(wordcounts, decreasing = TRUE)  
  
  
cloudFrame <- data.frame( word = names(wordcounts),freq = wordcounts)  
  
wordcloud(cloudFrame$word, cloudFrame$freq)

## Warning in wordcloud(cloudFrame$word, cloudFrame$freq): virginamerica could not  
## be fit on page. It will not be plotted.



wordcloud(names(wordcounts), wordcounts, min.freq = 1,  
 max.words = 200, random.order = FALSE, rot.per = 0.45, colors = brewer.pal(5,"Accent"))

## Warning in wordcloud(names(wordcounts), wordcounts, min.freq = 1, max.words =  
## 200, : virginamerica could not be fit on page. It will not be plotted.



jpeg('p1.jepg')  
wordcloud(cloudFrame$word, cloudFrame$freq)  
dev.off()

## png   
## 2

A close up of a piece of paper

Description automatically generated