Day 4 data manipulation

Paul

7/15/2020

library(tm) #text mining package

## Warning: package 'tm' was built under R version 4.0.2

## Loading required package: NLP

library(SnowballC)#text stemming package  
library(wordcloud)#word cloud generator

## Warning: package 'wordcloud' was built under R version 4.0.2

## Loading required package: RColorBrewer

library(RColorBrewer)#color palettes  
  
  
  
  
#step 3: text mining  
#load the text  
text<-readLines("CHAPTER ONE.txt")  
docs<-Corpus(VectorSource(text))  
inspect(docs)

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 9  
##   
## [1] CHAPTER ONE: INTRODUCTION   
## [2]   
## [3] 1.1 Background of the Problem   
## [4]   
## [5] The use of social media has advanced greatly over the years and become such an integral part of our everyday engagement in areas of communication, general social interaction and information sharing giving us platforms for political discourse and granting us access to social networking sites (SNSs) such as Facebook to get informed and engage ourselves in political issues. The focus of this research is investigating use of social media to influence government decisions in South Sudan. The study will be examining the significance of Facebook and Twitter to determine how each platform compares with each other in terms of government planning policies (Husain 2015).  
## [6] Given that Facebook and Twitter are really very popular with youths, it is believed that in a country like South Sudan barely out of civil war and pressed in by hard economic problems which are compounded further by the poor infrastructure among other issues, social media could provide the most inexpensive solution to addressing its communication needs with its citizenry particularly he youth. Facebook is enthusiastically embraced by political engagements as strategy for communication (Husain 2015).   
## [7] According to Streglitz and Xuan (2012), the interest in Facebook may be because positive political engagements or participation is possible with the ability to be influenced too easily to a directed group. But it is equally possible that the popular rise in the use of such platforms and services such as Facebook, WhatsApp, Twitter, Telegram, etc. may be due to their general characteristics perceived to offer ubiquity, immediacy and availability.   
## [8]   
## [9]

#text transformation - replace special characters  
library(magrittr)  
docs %>% tm\_map(removeNumbers) %>%   
 tm\_map(removePunctuation) %>%  
 tm\_map(stripWhitespace)

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

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

## Warning in tm\_map.SimpleCorpus(., stripWhitespace): transformation drops  
## documents

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 9

docs<-tm\_map(docs, content\_transformer(tolower))

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

docs<-tm\_map(docs, removeWords, stopwords("english"))

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

#Text stemming  
docs<-tm\_map(docs, stemDocument)

## Warning in tm\_map.SimpleCorpus(docs, stemDocument): transformation drops  
## documents

#Step 4: build term-document  
docs2<-TermDocumentMatrix(docs)  
docs3<-as.matrix(docs2)  
docs4<-sort(rowSums(docs3), decreasing = TRUE)  
docs5<-data.frame(word=names(docs4), feq=docs4)  
head(docs5, 10)

## word feq  
## facebook facebook 5  
## social social 5  
## engag engag 4  
## polit polit 4  
## media media 3  
## platform platform 3  
## use use 3  
## problem problem 2  
## (husain (husain 2  
## 2015). 2015). 2

#Step 5: Generate word cloud  
set.seed(234)  
dc<-wordcloud(words=docs5$word, freq=docs5$feq, min.freq   
 = 1, max.words = 20, random.order = TRUE, rot.per = 0.35,  
 colors=brewer.pal(8, "Dark2"))

