#########################################################################

# ECONOMIA CON MANZANITAS

#########################################################################

#::::::::::::::::::::GRAFICO DE NUBE DE PALABRAS::::::::::::::::

#quieres mas Videos de R!!!! Dale click al Sgt enlace:

https://www.youtube.com/watch?v=i0nkPfVUges

#representar graficamente un texto, para dar una idea de lo que contiene

#se dibuja las palabaras mas frecuentes, + palabras-> + grande el texto

#instalamos todos los paquetes que necesitamos

install.packages("tm")

install.packages("NLP")

install.packages("SnowballC")

install.packages("wordcloud")

install.packages("RColorBrewer")

#cargamos las librerias correspondientes

library(tm)

library(NLP)

library(SnowballC)

library(RColorBrewer)

library(wordcloud)

#DISCURSO DE NO VIOLENCIA DE GANDHI

#Que palabras hay con mayor frecuencia en el discurso de Gandhi?

#nuestro texto lo guardamos en bloc de notas en formato txt

texto <- readLines("D:/no\_violencia.txt")

texto = iconv(texto, to="ASCII//TRANSLIT")

texto = Corpus(VectorSource(texto))

########### LIMPIAMOS NUESTRO TEXTO CON EL COMANDO tm\_map

#ponemos todos los datos a minuscula (A!=a)

discurso=tm\_map(texto, tolower)

#quitamos los espacios en blanco

discurso =tm\_map(discurso, stripWhitespace)

#quitamos la puntuacion

discurso = tm\_map(discurso, removePunctuation)

#quitamos los numeros

discurso = tm\_map(discurso, removeNumbers)

#mostramos palabras vacias y genericas

stopwords("spanish")

#quitamos palabras genericas

discurso=tm\_map(discurso, removeWords,stopwords("spanish"))

#tambien podemos tener nuestra propia lista de palabras a quitar

############### DATA FRAME DE PALABRAS CON SU FRECUENCIA

#Creamos matriz de letras

letras= TermDocumentMatrix(discurso)

findFreqTerms(letras, lowfreq=5)

matrix=as.matrix(letras)

#lo ordenamos y sumamos las letras de nuestra matriz

vector <- sort(rowSums(matrix),decreasing=TRUE)

#creamos la data con las palabras y su frecuencia

dataletras <- data.frame(word= names(vector),frequencia=vector)

# lo nombra y le da formato de data.frame

################ GRAFICAMOS FRECUENCIA DE LAS PALABRAS

barplot(dataletras[1:10,]$freq, las = 2, names.arg = dataletras[1:10,]$word,

col ="blue", main ="PALABRAS MÁS FRECUENTES", ylab = "Frecuencia de palabras")

############ GRAFICAMOS LA NUBE DE PALABRAS

wordcloud(words = dataletras$word, freq = dataletras$freq, min.freq = 2,

max.words=70)

#en el centro la palabra mas importante,

wordcloud(words = dataletras$word, freq = dataletras$freq, min.freq = 2,

max.words=70, random.order=FALSE, rot.per=0.2,

colors=brewer.pal(8, "Dark2"))

########################################################################

# No te olvides del codigo mas importante de R: SUSCRIBIRTE Y DALE LIKE

#######################################################################