**Muriel Taks - 4.932.795-4**

**Script de R:**

library(RGtk2)

library(RQDA)

library(dplyr)

library(ggplot2)

library(rtweet)

library(quanteda)

library(stringr)

library(rtweet)

library(dplyr)

token = create\_token(app = "RcodingMu",

consumer\_key = "rMRA2MVKp3NZRgbMeJsvmFQS7",

consumer\_secret = "6yhAgFkHfWfPW2gRvPbOSAutsYItc20RCnBx7ndiUOlKXF6feC",

access\_token = "1359268974228369416-9rd2T6M8sCPPmu54jNfRNsG4mGLY8X",

access\_secret = "Qk1HBrLtKun4OR3jlN3KtGeW59Gj0VhInEwjV3OR0ebI1",

set\_renv = TRUE)

JoeBiden <- search\_tweets(

'"Joe Biden"', n = 1000, include\_rts = FALSE,retryonratelimit=FALSE,type="popular")

#subir un poco el n y probar de forma distintas las comillas

JoeBiden %>%

ts\_plot("4 hours") +

ggplot2::theme(plot.title = ggplot2::element\_text(face = "bold")) +

ggplot2::labs(x = NULL, y = NULL,

title = "Frecuencia de Joe Biden de los últimos 9 días",

caption = "Recuento de tweets en intervalos de 4 horas")

#Agarro los estados más likeados del usuario. Son estados que el ha dado likes

dm\_fav = get\_favorites("JoeBiden", n = 100)

serie <- get\_timelines(c("JoeBiden", "DrBiden"), n = 3200, retryonratelimit=FALSE)

serie %>%

dplyr::filter(created\_at > "2021-01-01") %>%

dplyr::group\_by(screen\_name) %>%

ts\_plot("weeks") +

ggplot2::geom\_point() +

ggplot2::theme\_minimal() +

ggplot2::theme(

legend.title = ggplot2::element\_blank(),

legend.position = "bottom",

plot.title = ggplot2::element\_text(face = "bold")) +

ggplot2::labs(

x = NULL, y = NULL,

title = "Frecuencia de los estados de Twitter publicados",

subtitle = "Recuento de estados de Twitter agregados por semana - Enero/Setiembre 2021"

)

serie2 <- get\_timelines(c("FLOTUS", "POTUS"), n = 3200, retryonratelimit=FALSE)

serie2 %>%

dplyr::filter(created\_at > "2021-01-01") %>%

dplyr::group\_by(screen\_name) %>%

ts\_plot("weeks") +

ggplot2::geom\_point() +

ggplot2::theme\_minimal() +

ggplot2::theme(

legend.title = ggplot2::element\_blank(),

legend.position = "bottom",

plot.title = ggplot2::element\_text(face = "bold")) +

ggplot2::labs(

x = NULL, y = NULL,

title = "Frecuencia de los estados de Twitter publicados",

subtitle = "Recuento de estados de Twitter agregados por semana - Enero/Setiembre 2021"

)

DonaldTrump <- search\_tweets(

'"Donald Trump"', n = 1000, include\_rts = FALSE,retryonratelimit=FALSE,type="popular")

DonaldTrump$Presidente="Donald Trump"

JoeBiden$Presidente="Joe Biden"

#Analisis de sentimientos

tweets\_df = rbind(JoeBiden,DonaldTrump)

install.packages("syuzhet")

library(syuzhet)

Sentiment <- get\_nrc\_sentiment(tweets\_df$text, language = "english")

tweets\_df\_senti <- cbind(tweets\_df, Sentiment)

tweets\_df\_senti$puntaje<-tweets\_df\_senti$positive-tweets\_df\_senti$negative

tweets\_df\_senti$sentimiento=ifelse(tweets\_df\_senti$puntaje<0,"Negativo","Positivo")

tweets\_df\_senti$sentimiento=ifelse(tweets\_df\_senti$puntaje==0,"Neutral",tweets\_df\_senti$sentimiento)

tweets\_sent <- tweets\_df\_senti %>% group\_by(Presidente,sentimiento) %>% summarise(count=n()) %>%

mutate(per = round(prop.table(count)\*100,1))

ggplot(tweets\_sent, aes(x=Presidente, y=per, fill=sentimiento))+

geom\_bar(position="dodge", stat="identity")+

scale\_fill\_manual(values = c("#EB594D", "#FFFAA4","#98E898"))+

geom\_text(data = tweets\_sent,

aes(x = Presidente, y = per, label = per),position=position\_dodge(width=0.9), vjust=-0.25)

#Tweets HECHOS por Biden, no twitteados acerca de él

BIDEN <- get\_timeline(user="@JoeBiden",n = 5000)

##alguna limpieza previa de tweets:

#sacar espacios

BIDEN$text <- str\_replace\_all(BIDEN$text," "," ")

# sacar URLs

BIDEN$text <- str\_replace\_all(BIDEN$text, "http[[:alnum;]]\*","")

BIDEN$text <- str\_replace\_all(BIDEN$text, "s://t.co/[[:alnum;]]\*","")

# sacar toda referencia a RT

BIDEN$text <- str\_replace(BIDEN$text,"RT @[a-z,A-Z]\*: ","")

# sacar hashtags

BIDEN$text <- str\_replace\_all(BIDEN$text,"#[a-z,A-Z]\*","")

# sacar referencias a otros screen\_names

BIDEN$text <- str\_replace\_all(BIDEN$text,"@[a-z,A-Z]\*","")

##creo corpus y limpio: BIDEN

BIDEN\_corpus <- corpus(BIDEN,text\_field = "text")

#creo Dfm

mydfm\_BIDEN <- dfm(BIDEN\_corpus,

stem = FALSE,

tolower = TRUE,

remove\_punct = TRUE,

remove\_numbers = TRUE,

verbose = TRUE)%>%

dfm\_remove(min\_nchar=3)

### Armo un diccionario según mi interés

midic <- dictionary(list(social = c("president", "black", "social","school", "politics","society", "family","culture", "education", "health","democracy","democratic","climate change"),

economia = c("republicans", "economics","economy", "jobs", "unemployment", "employment","shares","dolar\*","debt\*","tax\*","wealthy"),

pandemia=c("pandemic","covid","vaccine","vaccinated","pfizer", "death", "hospitals", "virus")))

### Aplico el diccionario en mi dfm y saco el porcentaje

midic\_result\_BIDEN<-dfm\_lookup(mydfm\_BIDEN,dictionary=midic,nomatch="no\_aparece")

midic\_result\_BIDEN=convert(midic\_result\_BIDEN, to = "data.frame")

midic\_result\_BIDEN$Presidente="Joe Biden"

#hacemos tabla para comparar las menciones de cada categoría de palabras

tabla = midic\_result\_BIDEN %>%

summarise(social = sum(social), social\_prop = (sum(social)/sum(social,economia,pandemia,no\_aparece))\*100,

economia = sum(economia), economia\_prop = (sum(economia)/sum(social,economia,pandemia,no\_aparece))\*100,

pandemia = sum(pandemia), pandemia\_prop = (sum(pandemia)/sum(social,economia,pandemia,no\_aparece))\*100)

##abro un archivo con stopwords propias y modismos

stop = read.csv("C:\\Users\\Usuario\\Desktop\\R\\Curso R - Ed.Permanente\\Trabajo final\\stop\_words\_english.txt", sep = ";")

##me quedo con el vector

vector = as.character(stop$X0)

#aplico la funcion dfm con los argumentos para limpiar el texto

mydfm <- dfm(BIDEN\_corpus,

stem = FALSE,

tolower = TRUE,

remove = c(stopwords("english")),

remove\_punct = TRUE,

remove\_numbers = TRUE,

verbose = TRUE)%>%

dfm\_remove(min\_nchar=3)

#Nube de palabras

remotes::install\_github("quanteda/quanteda.textplots")

library(quanteda.textplots)

textplot\_wordcloud(mydfm, min.count = 4,max\_words = 150,random.order = TRUE,

rot.per = .50, colors = RColorBrewer::brewer.pal(8,"Dark2"))