



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



EDUCACION PERMANENTE
UNIVERSIDAD DE LA REPÚBLICA

Trabajo final
Curso “R aplicado al análisis cualitativo”

2021

Docente: Elina Gómez

Muriel Taks
4.932.795-4

Índice

1. Introducción.....	p. 3
2. Twitter y política.....	p. 4
3. Etapa exploratoria - Tweets que mencionan a Joe Biden.....	p. 5
4. Etapa exploratoria - Tweets “likeados” por Joe Biden.....	p. 7
5. Comparación entre cuentas de Joe Biden y Donald Trump - análisis de sentimientos.....	p. 10
6. Identificación de temáticas en los tweets realizados por Joe Biden.....	p. 12
7. Reflexiones finales.....	p. 15
8. Bibliografía.....	p. 16
9. Anexo.....	p. 17

Introducción

En el marco del trabajo final del curso “R aplicado al análisis cualitativo”, se trabajó con información extraída de la red social Twitter, a través de herramientas de análisis y procesamiento de la información.

Aunque de manera exploratoria, se indagó en los tweets realizados por los últimos presidentes de Estados Unidos, poniendo el foco de atención en la cuenta del actual presidente, Joe Biden. Se optó por analizar información del contexto socioeconómico y político norteamericano con el objetivo de salirse de la coyuntura uruguaya actual y explorar formas diversas de hacer política. El caso de la política estadounidense se configura como uno de los más emblemáticos a nivel mundial, tanto por el alcance de sus repercusiones como el difundido conocimiento de las figuras públicas que la integran.

Se consideró pertinente utilizar información extraída de Twitter para llevar adelante este incipiente análisis teniendo en cuenta que se trata de una de las mayores redes sociales a nivel mundial. La misma fue creada en 2006 y desde entonces se ha encargado de servir como plataforma para la publicación de miles de millones de textos cortos en los que se incluye opiniones, enlaces, contenido multimedia, comunicados públicos, entre otros. Por lo tanto, Twitter se configura como terreno fértil para el análisis de big data, brindando a los científicos sociales acceso inmediato a enormes cantidades de información, ricas en contenido y en forma.

Se aclara que el acceso a la misma fue posible a través del permiso brindado por Twitter para “accounts developers” en el que se otorgó un API para extraer los datos necesarios, con fines exclusivamente académicos y de no pública difusión.

Twitter y política

Como fue mencionado en el apartado anterior, Twitter ha servido desde su creación como una plataforma para la manifestación de políticos/as alrededor del mundo. La difusión de los programas de gobierno así como el debate con oponentes políticos han sido algunos de los fenómenos que se han logrado evidenciar a través de esta red social, en relación al mundo de la política. El fácil acceso a la información, el carácter breve de los statements publicados y la rapidez de la difusión de los mismos han permitido que Twitter se convierta en un espacio ideal para que los políticos y políticas expresen sus ideas. De alguna manera ha democratizado información respecto a gobiernos y su forma de gobernar que en otras épocas era reservada únicamente a ciertos grupos socioeconómicos.

En el caso estadounidense puntualmente, Twitter se ha convertido en un arma poderosa para la competencia electoral y la puja por el poder político. A principios de 2021, ocurrió uno de los eventos más polémicos en la historia actual de la red social, a partir de la censura de la cuenta de Donald Trump y su posterior cierre como medida punitiva. La propagación de mensajes que incitaban a la violencia fue el principal motivo que sustentó dicha decisión. Estos mensajes habían sido difundidos por el ex mandatario en un clima caldeado a nivel político, social y económico, en el que Estados Unidos atravesaba un cambio de mandato enmarcado en la pandemia global del Covid-19.

El caso de censura hacia la cuenta de Trump se configuró como un punto de inflexión en cuanto a las discusiones acerca de la libertad de expresión y manifestación de los individuos. Remarca, por si hasta ese punto quedaba alguna duda, el enorme poder que las redes sociales poseen sobre todo tipo de individuos hoy en día, incluso sobre el propio presidente de Estados Unidos.

En este análisis no se buscará establecer un desarrollo valorativo acerca de si este acto debe ser etiquetado como una acción responsable o como una violación a los derechos humanos. Por el contrario, el objetivo de este trabajo es indagar en la enorme potencialidad que Twitter ofrece como red social en términos de análisis de información. Se intentará evidenciar, de forma exploratoria, cómo fenómenos como

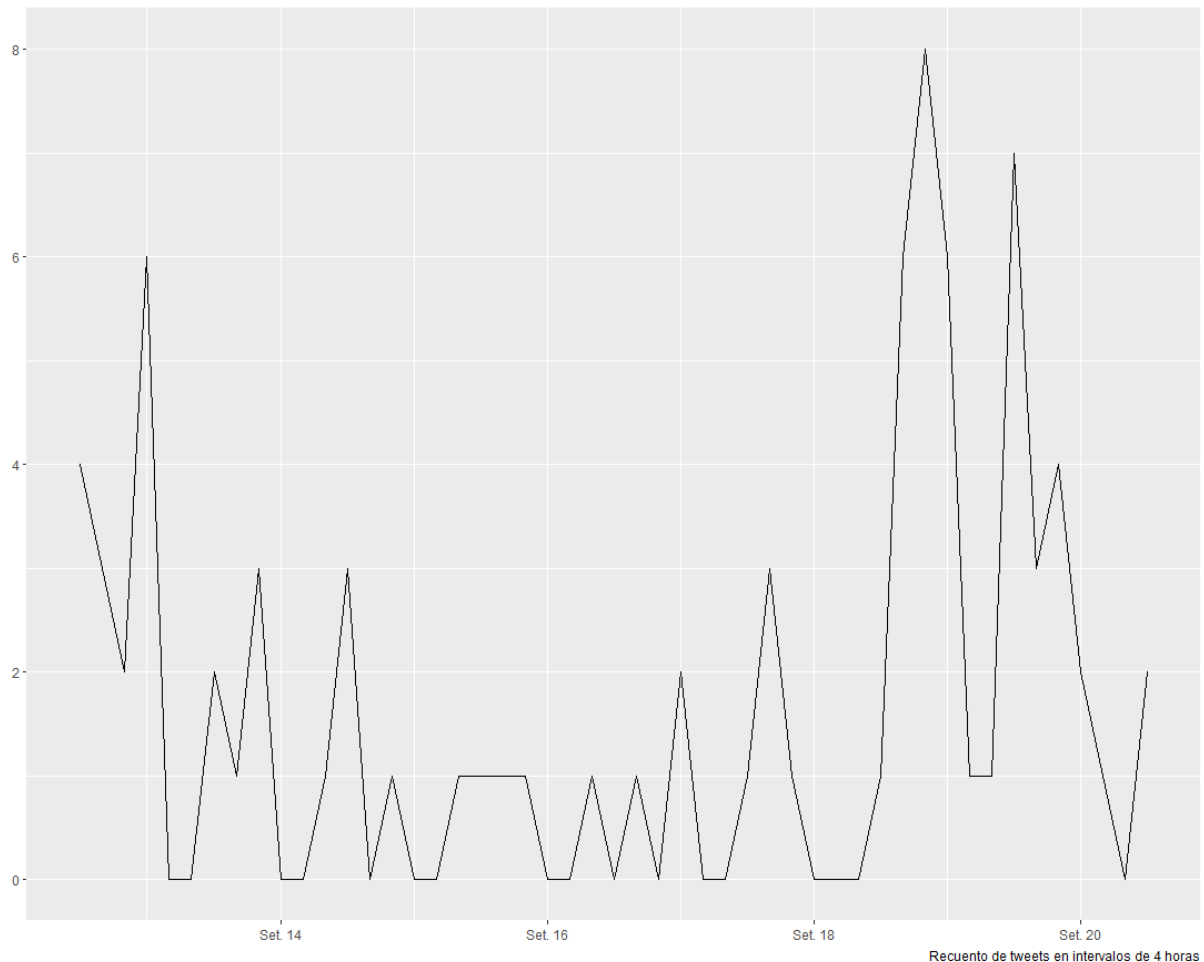
la censura de Trump, el ascenso al poder de Joe Biden o su relación con la primera dama pueden ser indagados desde la lupa de las Ciencias Sociales a través de entornos y lenguajes de programación como lo es R.

Etapla exploratoria - Tweets que mencionan a Joe Biden

A través del paquete “rtweet” que la plataforma de R ofrece, se buscaron alrededor de 80 tweets en los que se haya mencionado a Joe Biden en los últimos 8 días. Este límite es establecido por el propio programa y su respectivo paquete, por lo que no pudo ser realizada una búsqueda más amplia en términos temporales. Sin embargo, se consideró como una muestra “representativa” acerca de los principales tópicos y cuentas que mencionan a Biden en la red social.

Para retratar la cantidad absoluta de los tweets encontrados, y así poder evidenciar con qué frecuencia Biden es mencionado entre los usuarios de Twitter, se creó un gráfico a través del comando “ggplot2”. El siguiente gráfico muestra el recuento de tweets en intervalos de 4 horas, desde aproximadamente el 13 hasta el 20 de setiembre. La decisión de hacer el corte de 4 en 4 horas fue llevada a cabo por motivos prácticos, entendiendo que permite evidenciar una tendencia aproximada de los tweets realizados, sin que la información contenida en el gráfico resulte abrumadora. Se optó por pedir la extracción de los tweets de tipo “popular”, a partir de la cual se devuelven solamente los tweets más populares en la sección. Este último aspecto resultó pertinente con el fin de observar aquellos comentarios que cobraron relevancia en la red social, y no cualquier tipo de frase o párrafo que haya sido publicado por cualquier usuario en Twitter. Aunque se entiende que esto reduce la representatividad de los resultados, se asume este riesgo con el fin de analizar información que resulte relevante para los fines del trabajo.

Frecuencia de Joe Biden de los últimos 9 días



Se evidencia que el pico máximo de menciones hacia Biden se dio en las últimas horas del día 18 de setiembre. Al intentar buscar respuestas respecto al motivo de este pico, se encontró que ese mismo día Biden había anunciado la deportación de refugiados inmigrantes que se encontraban en la frontera entre Texas y México. Se calcularon alrededor de 15.000 inmigrantes que se encontraban en esta situación, por lo que es entendible que dicho anuncio haya causado revuelo en la red social.

Aunque Biden se caracteriza como un presidente con una postura más abierta hacia la inmigración que el ex-mandatario Donald Trump, varias de sus políticas apuntan a establecer “orden” con respecto a la situación de las migraciones. Parte de la sociedad estadounidense reclama acción hacia este tipo de problemáticas y Biden, como recién llegado a la Casa Blanca, comenzó a dar respuestas.

Etapla exploratoria - Tweets “likeados” por Joe Biden

Para continuar con un análisis exploratorio de la información, y utilizando las herramientas que R dispone para la extracción de información de Twitter, se indagó en los estados likeados por Biden. Se trata de aquellos tweets para los que la cuenta oficial de Joe Biden clickeó “me gusta”, afirmando cierta afinidad o sentimiento positivo con lo establecido en el tweet. Utilizando la función “get_favorites” se accedió a 20 estados likeados por Biden, los cuales se muestran a continuación:

created_at	screen_name	text	source
2020-08-25 18:44:44	cindymccain	@JoeBiden Thank you for your kind words.	Twitter for iPhone
2020-08-11 20:31:39	DrBiden	Hey @DouglasEmhoff. Are you ready?	Twitter Web App
2020-02-15 00:34:35	DrBiden	xo Jill https://t.co/EON3Ap3BHm https://t.co/jEngPM...	Twitter Web App
2019-08-21 13:50:32	DrBiden	First day of school! It's been over thirty years and I ...	Twitter for iPhone
2019-07-20 19:02:22	DrBiden	Every knock, every phone call, and every conversati...	Twitter for iPhone
2019-05-29 16:12:23	DrBiden	Our best policies don't come from politics, they com...	Twitter Web Client
2019-05-23 18:53:49	DrBiden	Again and again people have said to me, "Joe is the ...	Twitter Media Studio
2019-05-10 16:00:16	DrBiden	The women and men who are the backbone of milit...	Twitter for iPhone
2019-05-07 14:55:17	DrBiden	Teaching is my internal compass. From "Miss Jacobs...	Twitter for iPhone
2019-04-25 11:42:14	DrBiden	#NewProfilePic https://t.co/GXrG0pqB6T	Twitter for iPhone
2019-04-25 10:30:04	DrBiden	It's official! Excited and proud that @JoeBiden is run...	Twitter for iPhone
2019-03-13 19:03:58	DrBiden	Just found out @GirlsWhoCode is bringing their incr...	Twitter for iPhone
2019-03-08 18:30:01	DrBiden	Inspired from Guatemala to Vietnam, by these spirit...	Twitter for iPhone
2019-02-15 17:18:52	DrBiden	On #NationalCaregiversDay, and every day, we are ...	Twitter for iPhone
2019-02-08 16:41:31	DrBiden	Happy National School Counseling Week! Thanks @...	Twitter for iPhone
2019-02-08 01:32:41	DrBiden	Last night, I was proud to join @dcnewsbash in sup...	Twitter for iPhone
2019-02-07 16:30:35	DrBiden	Joe and I met Kev this summer along with the @sup...	Twitter for iPhone
2019-01-21 17:03:33	DrBiden	Uncommon acts of kindness have a transformative ...	TweetDeck
2019-01-17 22:37:10	DrBiden	Of course @MichelleObama shares a birthday with ...	TweetDeck
2019-01-17 15:02:41	DrBiden	Congratulations, Brian! Proud to work with you at t...	TweetDeck

La reducida cantidad de tweets buscados se debe a la practicidad y rapidez de la búsqueda; en el caso de haber aumentado el “n” requerido, la extracción se hacía más lenta y engorrosa. Por lo tanto, como se ha aclarado, se utiliza esta información entendiendo que no puede ser para nada generalizable, sino que únicamente explorada incipientemente.

Se observó a partir de la búsqueda realizada que la mayoría de los estados likeados por Biden son tweets publicados por su esposa, Jill Biden (cuyo usuario de Twitter es @DrBiden). Entra en el panorama del análisis una nueva actriz, tratándose de la primera dama de Estados Unidos.

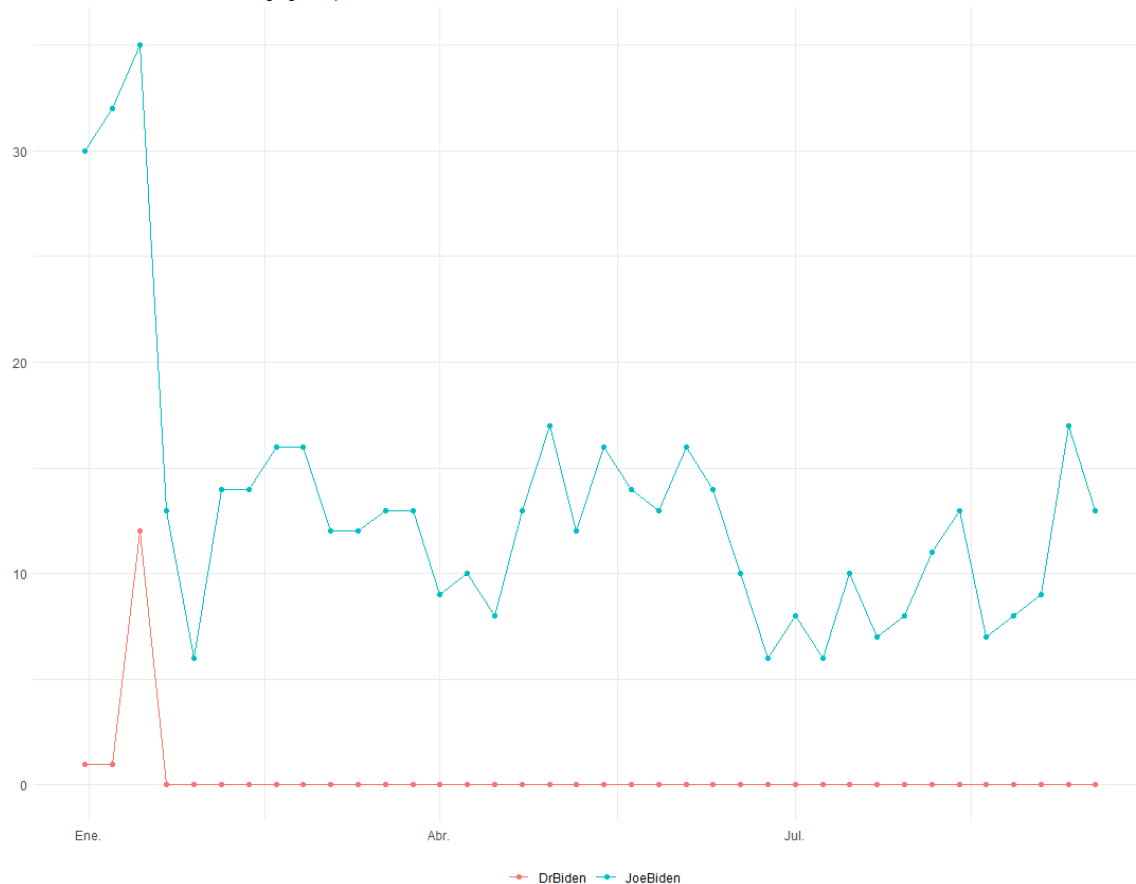
Dicho país no ha tenido presidentas mujeres, ni tampoco presidentes varones cuya pareja oficial haya sido otro hombre, por lo que las primeras damas son figuras que han acompañado la trayectoria política de Estados Unidos hace cientos de años. En muchos casos, las esposas de los mandatarios representan en sí mismas un rol político con gran relevancia. Uno de los casos más actuales al respecto es el de Michelle Obama, esposa de Barack Obama, quien fue siempre identificada con un rol activo a nivel social y cultural.

En el caso de Jill Biden, comenzó a ser mundialmente reconocida a partir de la campaña política de su esposo, para posteriormente “ascender” al cargo de primera dama del país el 20 de enero de 2021. Luego de observar que la gran mayoría de los tweets likeados del actual presidente eran, justamente, publicados por su esposa, emergió la curiosidad por indagar en la actividad “twittera” de dicha mujer.

Para esto, se realizó un gráfico comparando los tweets realizados tanto por Joe como por Jill Biden en el período de enero-setiembre de 2021, a partir de la función “get_timelines” y “ggplot2”. De esta forma, se pensó que se podría indagar en cuán activos o inactivos resulta ser cada una de estas figuras políticas en la red social. Como se evidencia en el gráfico, el resultado sorprendió por la ausencia de tweets realizados por Jill desde enero de 2021. Más precisamente, desde el 20 de enero de dicho mes. Se concluyó que no podía tratarse de una mera casualidad, teniendo en cuenta que ese fue el día en el que Joe asumió como 46° presidente de Estados Unidos.

Frecuencia de los estados de Twitter publicados

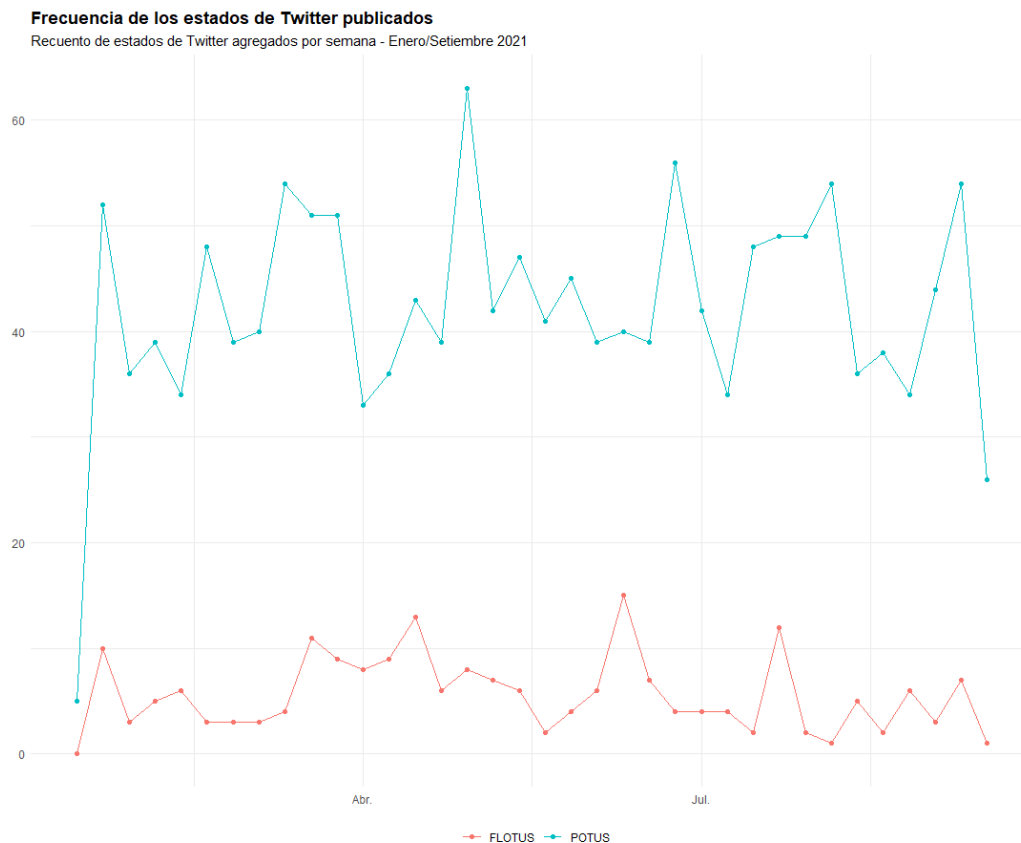
Recuento de estados de Twitter agregados por semana - Enero/Setiembre 2021



Al ahondar en las razones a las cuales se debía esta ausencia, se percató de la existencia de dos cuentas de Twitter oficiales: POTUS (President of the United States) y FLOTUS (First Lady of the United States). Al asumir el cargo, tanto el presidente como su esposa, acceden a dichas cuentas para realizar tweets relacionados a su mandato y de pública difusión. Esto no significa que deban abandonar totalmente sus cuentas anteriores, pero en el caso de Jill, parece haberse tomado esa decisión. Como se explica en el artículo periodístico publicado por El País “Twitter entrega la cuenta @POTUS a Biden como presidente de Estados Unidos” (Godoy, 2021) dicha cuenta es reseteada digitalmente, para que el nuevo mandatario “comience de cero” en cuanto a la publicación de tweets.

Donald Trump, caracterizado por ser una figura polémica, se negó a adoptar la cuenta POTUS para publicar sus statements políticos, y continuó usando su cuenta personal, en la cual aumentó millones de seguidores de forma continua durante todo su mandato (Godoy, 2021).

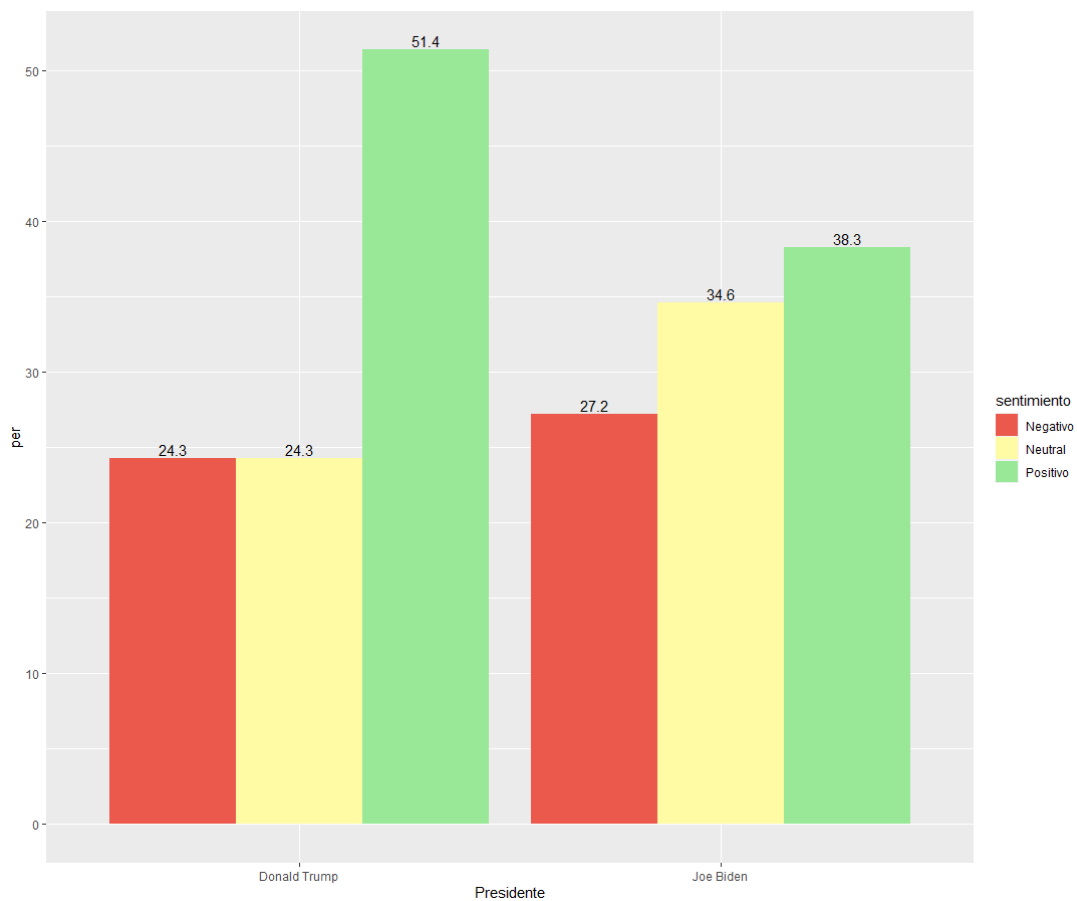
Se realizó por lo tanto una segunda comparación, esta vez entre las cuentas de POTUS y FLOTUS, para observar comparativamente la actividad de cada una de estas figuras en los últimos meses. Se evidenció una disminución de la brecha que distanciaba enormemente la actividad de Joe en contraste con la de Jill, aunque el nivel de “twittero” del presidente continúa excediendo por lejos el de la primera dama.



Comparación entre cuentas de Joe Biden y Donald Trump - análisis de sentimientos

Entendiendo que Donald Trump resultó ser un presidente polémico y caracterizado por sus extremas posturas referentes a temáticas socioeconómicas, se consideró interesante realizar una comparación entre los tweets que lo mencionan y los que mencionan al actual presidente, Biden. No se procuró únicamente hacer una búsqueda de la cantidad de tweets en los que se menciona a uno u otro, sino que un análisis de sentimientos de dichas publicaciones. Para llevar

adelante este procedimiento, luego de haber extraído 80 tweets del tipo “popular” que mencionan a Trump, se utilizó el paquete “syuzhet”. Se utilizó el diccionario brindado por dicho paquete en el idioma inglés, a partir del cual se identifican ciertos términos con un sentimiento concreto. Se aclara que durante todos estos procedimientos se realizó una “limpieza” de los tweets, eliminando enlaces, hashtags, retweets y otro tipo de elementos que pueden llegar a estorbar el manejo de la información. A partir de las herramientas mencionadas, es posible evaluar el tono reflejado en cada uno de los tweets. Los resultados se presentan a continuación, a través del siguiente gráfico y tabla:



Presidente	sentimiento	count	per
Donald Trump	Negativo	17	24.3
Donald Trump	Neutral	17	24.3
Donald Trump	Positivo	36	51.4
Joe Biden	Negativo	22	27.2
Joe Biden	Neutral	28	34.6
Joe Biden	Positivo	31	38.3

Se evidencia, tal vez de forma contraintuitiva, una mayor presencia de tweets positivos relacionados a Donald Trump (51.4% contra 38.3% de Joe Biden). A su vez, análogamente, existe una diferencia de tres puntos porcentuales en los que Trump supera a Biden en tweets de sentimiento positivo. Por su parte, el tono “neutral” adquiere una mayor presencia en los tweets que mencionan a Biden.

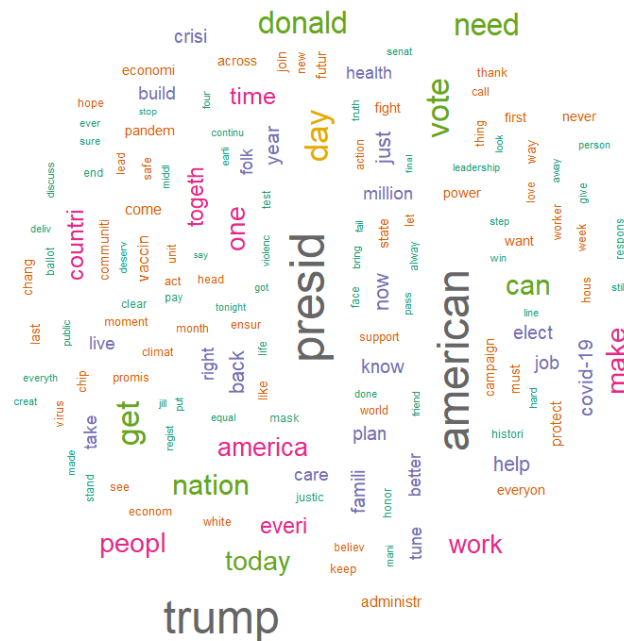
Se considera pertinente incluir estos hallazgos en el análisis entendiendo que este tipo de herramientas colaboran a develar la información que se encuentra subyacente al sentido común o conocimiento cotidiano. Aunque se podría pensar que la figura de Trump es usualmente asociada a una connotación negativa, aún más luego de la forma en que terminó su mandato, se observa un fuerte apoyo entre los usuarios de Twitter que lo mencionan. Esto puede deberse a una enorme cantidad de factores acerca de los cuales no se indagará en este trabajo, pero resulta necesario tomar conciencia del apoyo mantenido que este líder tuvo y aún tiene en la sociedad estadounidense.

Identificación de temáticas en los tweets realizados por Joe Biden

Dejando de lado la sección de carácter comparativo, se ahondará en esta etapa en los tweets que fueron realizados desde la cuenta oficial de Joe Biden, la cual no fue abandonada a pesar de adoptar también la cuenta de POTUS. Específicamente, interesa identificar cuáles fueron las principales temáticas abordadas en dichos tweets. Como fue mencionado en la parte introductoria de este trabajo, Twitter ofrece una plataforma de enorme potencial político para la manifestación de ideas y creencias, casi siempre vinculadas a una ideología partidaria. Joe Biden no queda al margen de dicha dinámica, por lo que, como tantos otros presidentes a lo largo de la historia actual, utiliza la red social para comunicar su posición con respecto a temáticas sociales, económicas y culturales.

Antes de pasar a la identificación de temáticas, se realizó un análisis primario de la información, para observar cuáles eran los términos más repetidos entre los tweets realizados por Biden. Se creó previamente una lista de stop words en inglés para incluir únicamente términos relevantes en dicha búsqueda. Los resultados se

presentan visualmente en la siguiente nube de palabras, con el objetivo de hacer la información más accesible:



Salta a la vista la palabra “Trump” como una de las más mencionadas entre los tweets, así como también “Donald” aunque en menor medida. Por otro lado, “American” es otro de los términos con mayor presencia, y “vote” aunque en menor medida. Para ahondar en este punto, se optó por identificar las temáticas en la cuales todas estas palabras se encuentran englobadas.

R ofrece la posibilidad de indagar en este aspecto a través de la utilización de diccionarios. Anteriormente fue empleado un diccionario de sentimientos, mientras que en esta ocasión se utilizará un diccionario personalizado. El mismo fue creado de manera emergente, a medida que se leyó rápidamente los tweets realizados por Biden. Identificando ciertas palabras que se repetían y otras que fueron incluidas más bien por sentido común, se crearon tres categorías temáticas: “Social”, “Economía” y “Pandemia”. La categoría “Social” fue construida a partir de términos como "black", "social", "school", "politics", "society", "family", "culture", "education", entre otras. La categoría “Economía” engloba términos como "economics",

"economy", "jobs", "unemployment", "employment", "shares", "dolar*", entre otras. Por último, la categoría "Pandemia" refiere a términos como "pandemic", "covid", "vaccine", "pfizer", "death", "hospitals", "virus", etc.

En la siguiente tabla se resume la presencia de cada una de estas categorías en los tweets extraídos:

social	social_prop	economia	economia_prop	pandemia	pandemia_prop
304	0.3920254	56	0.0722152	166	0.2140665

Se observa que la categoría con mayor peso en términos proporcionales es "Social", seguida por "Pandemia" y en último lugar "Economía". El peso de cada una de estas categorías se encuentra determinado por las palabras que fueron asignadas a cada una de ellas a través de cierta decisión teórica. Se asume que de cambiarse el término "jobs" de "Economía" a "Social", por ejemplo, el resultado de la búsqueda y el peso relativo de cada una puede alterarse enormemente. Se remarca por lo tanto la necesidad de sustentar teóricamente cada una de estas decisiones metodológicas, dado que puede conducir a resultados sesgados en cierta manera.

En este caso, se tomaron dichas decisiones de manera más bien intuitiva, buscando indagar de forma superficial en el carácter de los tweets realizados por el actual presidente. El protagonismo de la categoría "Social" podría llegar a evidenciar la impronta del gobierno de Biden, el cual se diferencia radicalmente del de Trump en términos de propuestas políticas. Joe Biden, aunque mantiene ciertos puntos de contacto con su predecesor, se vuelca en mayor medida a la generación de políticas sociales que buscan atender problemáticas de larga data en Estados Unidos. No por esto debe concluirse que se trata de un gobierno de carácter progresista ni mucho menos, pero sí demuestra poner el foco de atención en la dimensión social del país en mayor medida que el gobierno anterior. Por otra parte, el gran peso de la categoría "Pandemia" en cierta forma refleja la postura política de Biden frente a la situación de emergencia sanitaria. La presencia de términos como "vaccines" y "pfizer" entre sus tweets evidencian una mayor responsabilidad adquirida frente al combatimiento del virus, en contraste con el anterior presidente.

Reflexiones finales

A lo largo del presente trabajo se buscó indagar en el comportamiento de Joe Biden en la red social Twitter, en comparación a otras figuras públicas como su esposa o el ex-presidente Donald Trump. Se realizó un procedimiento desde lo micro hasta lo macro, comenzando por algunas cuestiones más bien cuantitativas acerca de cantidades de tweets, hasta aspectos de carácter cualitativo como el análisis de sentimiento.

Aunque de ninguna manera los hallazgos presentados pueden considerarse representativos ni dignos de ser difundidos como evidencia empírica científica, a través de los mismos se evidencia un acercamiento al enorme potencial que la programación en R ofrece para el análisis de información.

Nunca ha sido tan fácil para el cientista social el acceso a tales cantidades de información, la cual refleja comportamientos, sentimientos, opiniones, intereses e ideologías de la sociedad. Por lo tanto, en este trabajo se buscó transmitir lo esencial de la búsqueda por nuevas herramientas que permitan hacerse provecho de dicha información. De esta forma, se considera que se podrá construir un tipo de conocimiento basado en la empiria “digital”, por llamarla de alguna forma, el cual dispondrá cada vez de menos limitaciones metodológicas y prácticas.

El acercamiento a Tweets y la comparación a través de gráficos que puede realizarse con la “simple” aplicación de una fórmula en R coding, demuestra que aún existe un enorme campo por descubrir en cuanto al análisis cualitativo de la información desde la programación.

Bibliografía

Godoy, J.D. (2021). *Twitter entrega la cuenta @POTUS a Biden como presidente de Estados Unidos*. Recuperado de: <https://elpais.com/tecnologia/2021-01-20/twitter-entrega-la-cuenta-potus-a-joe-biden-como-presidente-de-estados-unidos.html>

Anexo

Script de R:

```
library(RGtk2)
library(RQDA)
library(dplyr)
library(ggplot2)
library(rtweet)
library(quantda)
library(stringr)
library(rtweet)
library(dplyr)

token = create_token(app = "RcodingMu",
  consumer_key = "rMRA2MVKp3NZRgBMeJsvmFQS7",
  consumer_secret = "6yhAgFkHfWfPW2gRvPbOSAutsYItc20RCnBx7ndiUOIKXF6feC",
  access_token = "1359268974228369416-9rd2T6M8sCPPmu54jNfRNsG4mGLY8X",
  access_secret = "Qk1HBrLtKun4OR3jIN3KtGeW59Gj0VhInEwjV3OR0eb1",
  set_env = 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"

#Análisis 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"))

```