

Análisis de las publicaciones en Twitter de legisladores de Uruguay

Pablo Alvarez

30/9/2021

Contents

Comparando Tweets	1
Construyendo los datos.	1
Análisis del texto	6
Palabras más frecuentes	6
Asociación de palabras.	9
Análisis de sentimientos.	12
Final	14
Comentarios	14

Comparando Tweets

Este trabajo corresponde a la entrega final del Curso **R aplicado al análisis cualitativo/ FCS-UdelaR** del año 2021, dictado por la profesora Elina Gómez.

Utilizando herramientas presentadas en el curso se realiza un análisis de las publicaciones en la red social **“Twitter”** por parte de legisladores del Parlamento uruguayo.

Para ello se consideran dos grupos. Por un lado las publicaciones de las cuentas correspondientes a los legisladores de Cabildo Abierto, y por otro lado las publicaciones de legisladores del Frente Amplio en representación del Partido Comunista y el Partido Socialista.

La selección de los casos (legisladores) corresponde a trabajar con la totalidad de los legisladores de un sector y a la vez equilibrar el número de legisladores en cada grupo. Las cuentas de Twitter fueron obtenidas de la información de cada legislador en la web del Parlamento. No todos los legisladores tienen cuentas de Twitter.

Con la información obtenida se busca identificar, a través de las publicaciones en la red social Twitter de los legisladores de Cabildo Abierto, aspectos significativos en relación con las categorías en las que se suele ubicar en la literatura a las “nuevas derechas”: conservadurismo social, tradición/nación en contraposición a lo que considerar “foráneo”.

Además para enriquecer el análisis y poder contrastar la información se analiza también a legisladores del partido Frente Amplio, partido que se ubica en el otro extremo de la escala ideológica dentro del sistema de partidos de Uruguay. En este sentido será de interés las diferencias/semajanzas que puedan surgir en los aspectos mencionados anteriormente, así como las que puedan surgir sobre asuntos de coyuntura relacionados al gobierno.

La información surgida de los tweets se evaluará a través de distintas herramientas: **análisis de frecuencia, análisis de asociación de palabras y análisis de emociones y sentimientos.**

El sentido general del documento se orienta a visualizar el proceso de análisis a través de los paquetes y funciones presentados en el curso, más que en el análisis cualitativo propiamente dicho.

Construyendo los datos.

Scrapeo de Twitter.

Luego de obtener las cuentas de los legisladores se construyen dos objetos de R (vectores) con dichas cuentas según partido: “Cuentas_Tw_Ca” para las cuentas en Twitter de los legisladores de Cabildo Abierto, y “Cuentas_Ps_Pcu” para los legisladores del Partido Socialista y Partido Comunista.

Las cuentas relacionadas **Cabildo Abierto** son las siguientes:

```
## [1] "RodrigoAlberna3" "SebastianCal1" "BulminiNazmi" "Ecapillera"
## [5] "LustEduardo" "perezbonavita" "APerrone510" "SodanoDiputado"
## [9] "GuidoManiniRios" "DomenechMG" "raul_lozanob"
```

Las cuentas relacionadas con el **Partido Comunista** y el **Partido Socialista** son las siguiente:

```
## [1] "AitaUbaldo" "DanielGerhard" "Veronica_Mato" "gera_nunez1001"
## [5] "Ana01001" "Oandradelallana" "SilviaNaneFA" "GabyBarreiro_uy"
## [9] "gonzacivila" "EnzoMalanCastro" "DanielOlesker"
```

Se cargan los paquetes con los que se trabajará durante el análisis, estos son:

```
library(rtweet)
library(quanteda)
library(quanteda.textplots)
library(quanteda.textstats)
library(stringr)
library(RColorBrewer)
library(colorspace)
library(tidyverse)
library(syuzhet)
library(kableExtra)
library(knitr)
```

Para realizar el scraping utilizo la función *get_timelines()* del paquete *rtweet::*, ya que el interés está en lo que cada cuenta publica. Asigno a dos objetos *Tw_Cabildo* y *Tw_Ps_Pcu* los datos que arroja la función aplicada para las cuentas de cada partido.

#Realizo el scraping para cada vector de cuentas con la función get_timelines() del paquete rtweet.

```
Tw_Cabildo <- get_timelines(Cuentas_Tw_Ca, n=3200,include_rts = FALSE, retryonratelimit = TRUE)
```

```
Tw_Ps_Pcu <- get_timelines(Cuentas_Tw_Ps_Pcu, n=3200,include_rts = FALSE, retryonratelimit = TRUE)
```

A los efectos de no tener que correr la función cada vez, guardamos en un archivo la información del *scraping* realizado el 17 de setiembre de 2021.

```
saveRDS(Tw_Cabildo, file = "~/Curso R Cualitativo/Trabajo Final/Twitter de CABILDO.rds")  
saveRDS(Tw_Ps_Pcu, file = "~/Curso R Cualitativo/Trabajo Final/Twitter de PS_PCU.rds")
```

Por lo tanto recuperamos esa informacion a partir de cargar los archivos.

```
Tw_Cabildo <- readRDS("~/Curso R Cualitativo/Trabajo Final/Twitter de CABILDO.rds")  
Tw_Ps_Pcu <- readRDS("~/Curso R Cualitativo/Trabajo Final/Twitter de PS_PCU.rds")
```

Table 1: N° de Tweets de cada legislador

screen_name	numero_tweets
AitaUbaldo	332
AnaO1001	368
APerrone510	711
BulminiNazmi	59
DanielOlesker	2501
DomenechMG	3
ECapillera	2239
EnzoMalanCastro	1382
GabyBarreiro_uy	108
gera_nunez1001	1730
gonzacivila	1310
GuidoManiniRios	792
LustEduardo	433
Oandradelallana	2110
perezbonavita	1077
raul_lozanob	797
RodrigoAlberna3	229
SebastianCal1	148
SilviaNaneFA	2149
SodanoDiputado	144
Veronica_Mato	621

Cantidad de Tweets según cada usuario.

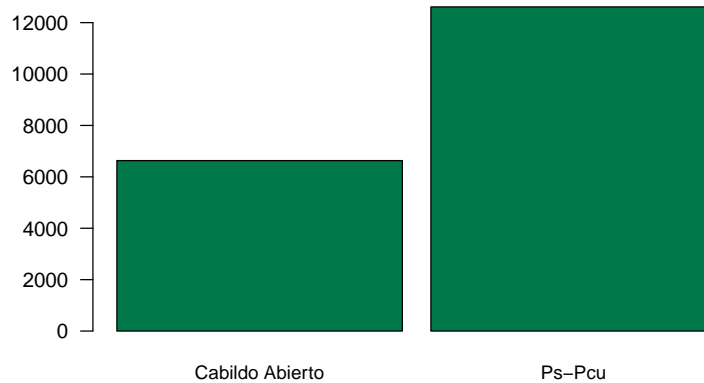
La Tabla 1 presenta la cantidad de tweets correspondiente a cada legislador, ordenados alfabéticamente por el nombre de usuario de la red:

La cantidad de tweets de cada grupo analizado se muestra en la Tabla 2. El Gráfico 1 permite ver fácilmente que las cuentas de los legisladores del FA duplican al numero de tweets de Cabildo Abierto. Esto está relacionado con la mayor actividad y también con la antigüedad de las cuentas. Por ejemplo la cuenta correspondiente al Senador Domenech es de reciente creación, y eso explica la cantidad de publicaciones de dicha cuenta.

Table 2: Cantidad de tweets por partidos

Cabildo Abierto	PS-PCU
6632	12611

Gráfico 1: Cantidad de Tweets por partido



Creación de *CORPUS*

Para comenzar a realizar el análisis del contenido de los tweets creo dos *corpus* con la información obtenida a través del scraping correspondiente al texto de los tweets. Para ello utilizo la función *corpus()* del paquete *quantda*:

Para la posterior limpieza del texto cargamos *stopes*, un listado de palabras que carecen de valor para un análisis de texto en español.

```
#Corpus----
Corpus_Tw_Ca <- corpus(Tw_Cabildo, text_field = "text")
Corpus_Tw_Ps_Pcu <- corpus(Tw_Ps_Pcu, text_field = "text")

stopes <- c("~/Curso R Cualitativo/stopes.csv")
```

Creamos los *dfm* de *quantda* para limpiar los datos y poder trabajar sobre ellos.

```
#dfm----

dfm_Tw_Ca <- dfm(Corpus_Tw_Ca,
  stem=FALSE,
  tolower=TRUE,
  remove=c("@*", "#*", stopwords(language="spanish"), stopes),
  remove_url=TRUE,
  remove_punct=TRUE,
  remove_numbers=TRUE,
  verbose=FALSE) %>%
  dfm_remove(min_nchar=3)

dfm_Tw_Ps_Pcu <- dfm(Corpus_Tw_Ps_Pcu,
  stem=FALSE,
  tolower=TRUE,
  remove=c("@*", "#*", stopwords(language="spanish"), stopes),
  remove_url=TRUE,
  remove_punct=TRUE,
```

```
remove_numbers=TRUE,  
verbose=FALSE) %>%  
dfm_remove(min_nchar=3)
```

Table 3: Cabildo Abierto

	Frecuencia
hoy	601
gracia	575
día	425
año	370
cabildo	325
junto	281
gent	277
much	275
paí	254
felicitation	249

Análisis del texto

Con la información obtenida durante el proceso previo procedemos a realizar los tres análisis indicados.

En primer lugar veremos cuáles son las palabras más frecuentes dentro de cada grupo político y lo visualizaremos en una nube de palabras.

En segundo término observaremos el grado de asociación entre palabras. En este caso nos interesan un grupo de palabras/conceptos que a priori se considera que pueden evidenciar relaciones representativas con categorías asociadas a las “nuevas derechas”. Estas palabras son: **ideología, género, familia, patria, gobierno y luc**

Por último se realizará un análisis de sentimiento utilizando el diccionario **nrc** del paquete **syuzhet::** se observará de forma comparada como cada grupo responde a dicho análisis.

Palabras más frecuentes

Tablas de Frecuencia

Las Tablas 2 y 3 muestran la lista de las 10 palabras más frecuentes para cada grupo analizado. Lo primero que llama la atención es que en ambos casos aparece la palabra **hoy** como la de mayor frecuencia. Esto podría explicar el tono inmediato de los tweets, donde la referencia está sobre lo que acontece, el presente, el día mismo en que se expresa. La inmediatez de la comunicación en las redes sociales podría ser una forma de interpretar esta situación.

También entre las primeras 5 palabras más frecuentes ambos grupos comparten otras dos palabras: **gracias** y **año**.

Sin embargo alguna diferencia surge entre ambos grupos. Considerando las 5 palabras más frecuentes **gobierno** y **política** aparecen para los legisladores del Frente Amplio, en tanto que para Cabildo Abierto tres se corresponden a una mención temporal además de “año” y “hoy” se suma **día**, y la otra palabra hace mención a su organización, **cabildo**.

Table 4: Partido Comunista y Partido Socialista

	Frecuencia
hoy	916
gobierno	712
año	648
gracia	527
política	523
ahora	492
social	479
trabajador	420
siempr	419
ley	411

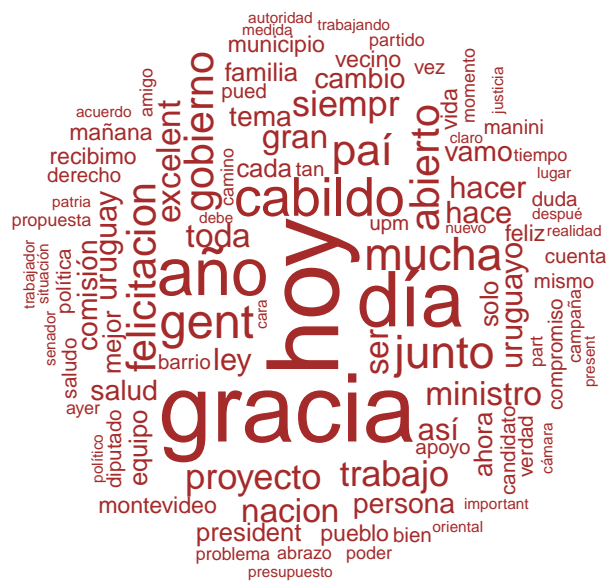
Nubes de Palabras

La misma información se presenta como **nube de palabras**. En este caso se visualizan las 100 palabras cuya frecuencia sea mayor que 2. De esta forma la imagen permite establecer más fácilmente la relación entre la frecuencia de las palabras.

La primer nube corresponde a las publicaciones de los legisladores de Cabildo Abierto.

```
#Nube de Palabras CA---

Nube_CA <- textplot_wordcloud(dfm_Tw_Ca,
                              freq = frecuentes_Ca,
                              min.freq = 2,
                              max.words = 100,
                              colors = c("Brown"),
                              random.order = F,
                              random.color = F,
                              scale = c(5, 0.1),
                              rot.per = 0.3)
```

La siguiente nube de palabras corresponde a legisladores del Frente Amplio.



Asociación de palabras.

A través de la función `textstat_simil()` del paquete `quanteda::` se realiza un análisis de correlación entre las palabras seleccionadas y el resto de palabras utilizadas en las publicaciones. Se realiza el análisis para las 20 palabras con mayor valor de correlación, pero se muestran las primeras 10.

Las palabras a evaluar son: **ideología**, **género**, **familia**, **patria**, **gobierno** y **luc**.

Estas palabras se seleccionaron para observar tres grandes áreas en el marco de un estudio sobre posicionamiento de Cabildo Abierto, como partido de derecha. Las palabras **ideología**, **género** y **familia** se eligen para acercarse al posicionamiento en un campo “conservador” en términos sociales. La selección de palabra **patria** está orientado a buscar referencias en el sentido del eje “tradición” y “nacionalismo”. Por último las palabras **gobierno** y **luc** responden a elementos de coyuntura, pero podrían dar cuenta de su posición en la coalición gobernante.

En este trabajo se muestran los resultados para las primeras tres palabras mencionadas. Sin embargo se mantienen en el archivo **Rmd** los análisis estadísticos para las otras palabras.

Ideología

```
ideologia_CA <- dfm_Tw_Ca %>%
  textstat_simil(selection = c("ideología"),
    margin = "feature", method = "correlation") %>%
  as.data.frame() %>%
  arrange(-correlation) %>%
  top_n(20)

kbl(head(ideologia_CA,10), caption = "Asociación con 'ideología':Cabildo Abierto") %>%
  kable_styling(latex_options = c("striped", "hold_position")) %>%
  row_spec(0,color="white", background = "#007849")
```

Table 5: Asociación con 'ideología':Cabildo Abierto

feature1	feature2	correlation
perciball	ideología	0.3647902
género	ideología	0.2726824
arraigo	ideología	0.2579262
peligrosísimo	ideología	0.2579262
arranquen	ideología	0.2579262
inculqu	ideología	0.2579262
convirtiendo	ideología	0.2579262
deriva	ideología	0.2579262
encomi	ideología	0.2579262
miramo	ideología	0.2579262

Table 6: Asociación con 'ideología':PS y PCU

feature1	feature2	correlation
ocultarlo	ideología	0.3535185
vistió	ideología	0.3535185
progr	ideología	0.3535185
influencia	ideología	0.3535185
pura	ideología	0.2641763
valoramo	ideología	0.2497074
orientan	ideología	0.2497074
cere	ideología	0.2036665
sesgado	ideología	0.2036665
vivirán	ideología	0.1765628

En asociación con la palabra **ideología**, destaca (de forma esperada) la mayor correlación con la palabra **género** en los legisladores y legisladoras de Cabildo Abierto, junto con la mención **peligrosísimo**. Esto puede entenderse como una evidencia del uso del concepto “ideología de género” y su valoración negativa/peligrosa para los legisladores de Cabildo Abierto, en tanto es el partido político que más se ha mostrado conteste a esta conceptualización y ha sido firme en la defensa de arreglos heteronormativos a los que considera amenazados por la “ideología de género”.

Género

Table 7: Asociación con 'género':Cabildo Abierto

feature1	feature2	correlation
arraigo	género	0.3533667
peligrosísimo	género	0.3533667
arranquen	género	0.3533667
inculqu	género	0.3533667
implicar	género	0.3533667
interperson	género	0.3533667
erradicar	género	0.3533667
trauma	género	0.3533667
relacionamiento	género	0.3533667
presupuestal	género	0.3533667

Table 8: Asociación con 'género':PS y PCU

feature1	feature2	correlation
violencia	género	0.4254129
basada	género	0.3130506
juzgado	género	0.2159275
multimateria	género	0.1884000
raza	género	0.1777427
equidad	género	0.1503013
mujer	género	0.1263609
obligart	género	0.1256781
ejercerlo	género	0.1256781
paleolítico	género	0.1256781

En el análisis de los legisladores de Cabildo Abierto, la palabra **género** aparece correlacionada con igual

valor para todas las palabras. Sin embargo se puede percibir que las palabras correlacionadas se identifican de forma negativa con la palabra **género**. Entre ellas aparecen: *peligrosísimo*, *trauma*.

En un sentido diferente aparece correlacionada la palabra *género* en los tweets de los legisladores del Frente Amplio. En este caso las principales menciones se asocian con conceptos que se relacionan a una perspectiva de derechos. En este sentido la mayor correlación está con la palabra *violencia* y la segunda es la palabra *basada*. Esto permite suponer que la referencia hace mención a la idea de “violencia basada en género”. La palabra **multimateria** seguramente se relaciona con la discusión presupuestal en torno a los juzgados multimateria que está en relación también a los temas de violencia basada en género. Aparece también la mención a la palabra *equidad* que se relaciona positivamente en una perspectiva de derechos. Por tanto, según los tweets analizados, para los legisladores del FA la palabra **género** fue en mayor medida utilizada en referencia a la violencia basada en género. Para Cabildo Abierto la idea género aparece como “amenaza” en tanto que para el Frente Amplio aparece relacionada a los “derechos” (equidad, contra la violencia basada en género, etc.)

Familia

Para esta palabra, las mayores correlaciones en los tweets de Cabildo Abierto están asociadas a menciones relacionadas (aparentemente) con fallecimientos. También para los legisladores del FA esto aparece. Sin embargo en el análisis de Cabildo Abierto hay dos palabras que tienen relación a nuestro interés y en particular la palabra *base*. Es esperable que esto remita a la noción de la “familia” como “base” de la sociedad, un concepto fuertemente arraigado en las perspectivas conservadoras. La otra palabra es *policia*, seguramente en conexión con la idea de “familia policial”, aspectos también relevantes en el discurso conservador.

Table 9: Asociación con ‘familia’:Cabildo Abierto

feature1	feature2	correlation
condolencia	familia	0.2821273
pésame	familia	0.2664245
fallecimiento	familia	0.2565163
emprendió	familia	0.2319659
estara	familia	0.2319659
amigo	familia	0.2202960
contención	familia	0.2179187
base	familia	0.2157006
polici	familia	0.1947559
dolor	familia	0.1874028

Table 10: Asociación con ‘familia’:PS y PCU

feature1	feature2	correlation
pésame	familia	0.2246845
amigo	familia	0.1840825
condolencia	familia	0.1823286
fallecimiento	familia	0.1721243
apretado	familia	0.1544761
adoptarlo	familia	0.1486226
carné	familia	0.1486226
heterosexu	familia	0.1486226
respetemo	familia	0.1486226
desbordaba	familia	0.1486226

Análisis de sentimientos.

Aplicando la función

Utilizando la función `get_nrc_sentiment()` del paquete *syuzhet* se aplica el diccionario de sentimientos a los **corpus** creados.

Igual que fuera realizado para el scraping de las cuentas de Twitter, guardamos el resultado de aplicar esta función en archivos para poder cargarlos directamente.

Para realizar el análisis se usa el siguiente código:

```
sentimiento_Ca <- get_nrc_sentiment(Corpus_Tw_Ca, language = "spanish")  
  
sentimiento_Ps_Pcu <- get_nrc_sentiment(Corpus_Tw_Ps_Pcu, language = "spanish")
```

A partir de la función utilizada obtenemos las medias de cada una de las ocho “emociones” y los dos “sentimientos” para ambos grupos y colocamos los resultados en una misma tabla. La información así dispuesta permite un primer acercamiento para el análisis.

Table 11: Media de sentimientos nrc

	ira	anticipación	disgusto	miedo	alegría	tristeza	sorpres	confianza	negativo	positivo
CA	0.247	0.404	0.161	0.403	0.458	0.411	0.212	0.883	0.704	1.347
PS-PCU	0.378	0.408	0.205	0.514	0.429	0.496	0.221	0.895	0.989	1.368

A través de la Tabla 10 se puede observar que en general los valores son similares en ambos grupos, pero también hay algunas diferencias.

La media de sentimientos negativos es mayor para el caso de los legisladores del Partido Socialista y el Partido Comunista que para los de Cabildo Abierto. También se da la misma situación para las emociones *miedo* e *ira*.

En menor medida, también es mayor la presencia de sentimiento *positivo* en los tweets de los legisladores del Frente Amplio. En cuanto a *confianza* y *tristeza* también la media del FA es superior. Sólo el valor correspondiente a la emoción *alegría* es mayor en los legisladores de Cabildo Abierto en relación con el FA.

Considerando la relación entre sentimiento *positivo* y *negativo* ambos grupos presentan mayor sentimiento *positivo*.

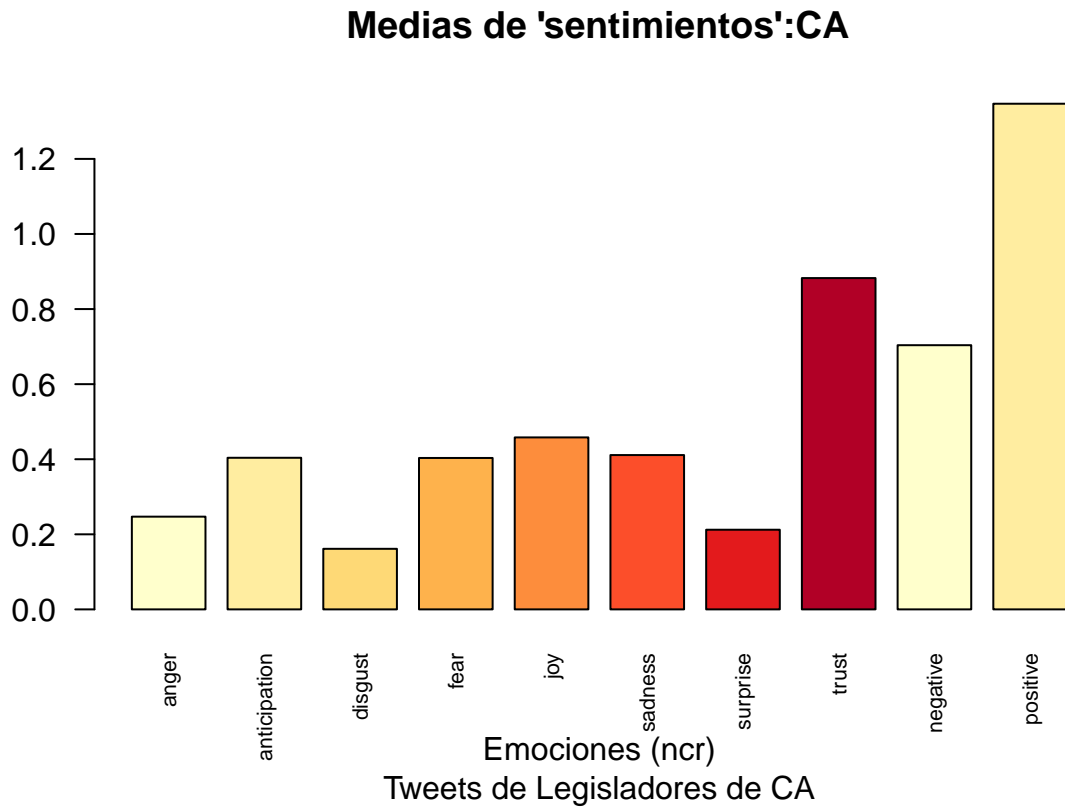
Analizando dentro de cada grupo, se percibe que para los legisladores de ambos grupos la emoción más presente en relación al resto es la *confianza*. En tanto la emoción menos presente también es la misma para los legisladores de Cabildo Abierto y del Frente Amplio, pero en este caso es la de *disgusto*.

Graficando los sentimientos.

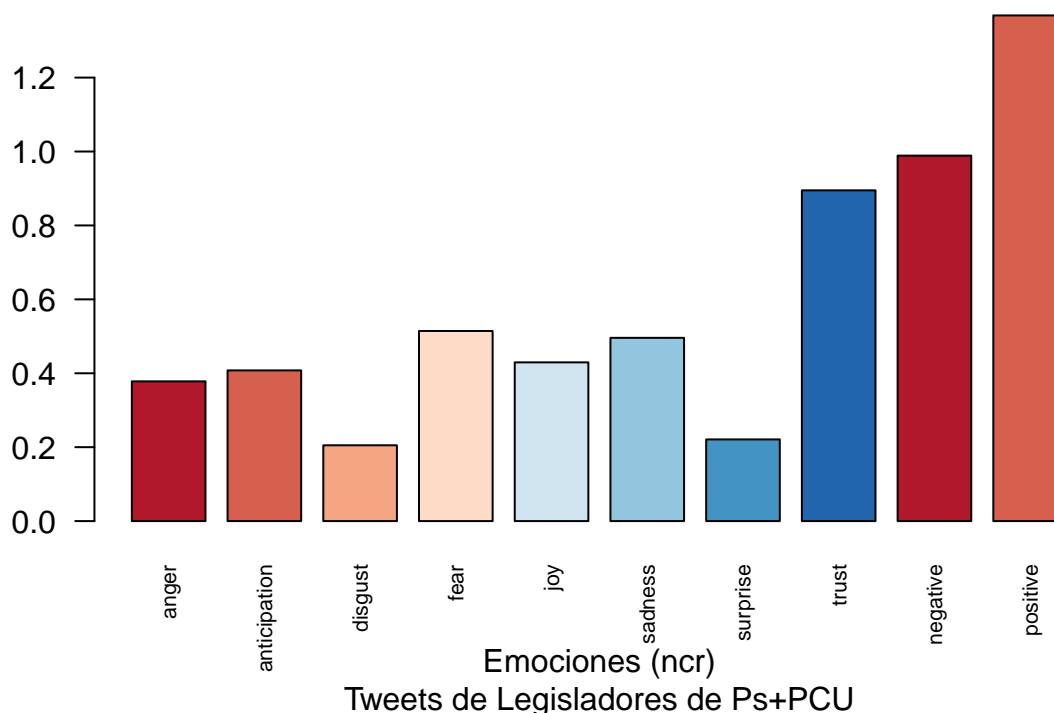
Otra forma de ver la información que hemos descrito es a través de gráficos de barra donde es más fácil percibir el peso de cada una de las emociones y sentimientos.

```
#Gráficos---  
  
graficoCA <- barplot(media_sent_Ca,  
                      space=0.3,  
                      horiz = FALSE,  
                      las = 2,  
                      cex.names = 0.7,
```

```
col= brewer.pal(n=8, name = "YlOrRd"),
main = "Medias de 'sentimientos':CA",
sub = "Tweets de Legisladores de CA",
xlab = "Emociones (ncr)",
ylab = NULL)
```



Medias de 'sentimientos: Ps+Pcu'



Final

Se ha presentado un breve análisis de los tweets de los legisladores de Cabildo Abierto y los Legisladores del Partido Socialista y el Partido Comunista. Se utilizaron varias de las herramientas del curso para construir una base de datos a partir del scraping de Tweet. Luego obtener los datos se realizó una limpieza de los mismos para eliminar información irrelevante y facilitar el trabajo.

Se realizaron análisis de **frecuencia**, **de correlación** y de **emociones y sentimientos**.

A lo largo del trabajo se presentó información a través de diferentes formas, en particular **tablas**, **nube de palabras** y **gráfico de barras**.

Si bien el centro del trabajo estuvo en mostrar el manejo de los paquetes y funciones de R, igual fue posible realizar un análisis cualitativo simple a partir de la información surgida de las distintas herramientas de análisis.

Comentarios

Entre las principales dificultades encontradas la que llevó mayor cantidad de tiempo fue la de diseño en Rmarkdown, fundamentalmente en cuanto a la ubicación de las tablas y gráficos en relación con el texto. *Correr* cada vez el **knit** para observar dónde estaban ubicadas las tablas insumió buena parte del trabajo, y sin duda es un debe en el trabajo. La posibilidad de aplicar *float* solicitaba cargar un paquete del LaTeX que no pude realizar, lo cual dificultó más el diseño del texto de salida.