


# **TWITTER: CÓMO OBTENER SU DATA**

**(¡MUCHA DATA!)**





A decorative wavy line in yellow and white on the left side of the image.

# **ACCEDIENDO AL API DE TWITTER**

# ¿QUÉ ES UN API?

- Los API (*application programming interface*) son repositorios digitales que sirven para, entre otras cosas, facilitar el intercambio de información entre el usuario y los “dueños” de la información.
- En el sentido más básico, el API te permite acceder a la página web del repositorio y acceder a su data (no estructurada).
- Hay veces que el acceso es directo y abierto. Hay otras, como con Twitter, que el acceso es directo, pero requiere un registro previo (si todavía no lo haces, puedes encontrar las instrucciones para hacerlo [aquí](#)).

# **LAS DOS FUNCIONES DEL API DE TWITTER**

- A través del API de Twitter puedes acceder a la data de dos maneras: Twitter Rest y Twitter Stream.
- Veamos para qué sirve cada una...

# ACCEDIENDO A LA DATA DE TWITTER I

- Para acceder al API de Twitter, necesitamos primero cargar el paquete *rtweet*. Si aún no instalas el paquete, debes correr el siguiente código:

```
7 install.packages("rtweet") # solamente una vez
```

- Una vez instalado, cargamos los paquete necesarios:

```
8  
9 library(rtweet) # rtweet nos permite acceder a la data de TW  
10 library(tidyverse) # tidyverse nos ayuda con la minupulación de datos  
11
```

# ACCEDIENDO A LA DATA DE TWITTER II

- Como aprendimos hace poco, algunos APIs necesitan de credenciales. Para obtener las credenciales de API de Twitter, debes:
  - Tener una cuenta de Twitter
  - Aplicar a una cuenta de desarrollador
- Si todavía no lo has hecho, el paquete *rtweet* tiene un buen tutorial (además del que está en nuestro GitHub):

```
17 vignette("auth", package = "rtweet") # Te lleva al tutorial
```

# ACCEDIENDO A LA DATA DE TWITTER III

- Una vez que tienes las credenciales, se las das a R y ¡listo! Ya tienes acceso a la data de Twitter:

```
21 create_token( # ¡Recuerda no compartir con nadie!  
22   app = "tw_svallejo", # El nombre que le diste al app  
23   consumer_key = [REDACTED] ¡SECRETO! , # Copia y pega de tu página de developer de TW  
24   consumer_secret = [REDACTED] ¡SECRETO! ,  
25   access_token = [REDACTED] ¡SECRETO! ,  
26   access_secret = [REDACTED] ¡SECRETO!  
27 )
```

- Esto queda grabado en el *environment* de R, así que solo lo tienes que hacer una vez.



- Una vez que corras el código, deberías ver algo así en tu consola:

```
<Token>
<oauth_endpoint>
  request:  https://api.twitter.com/oauth/request_token
  authorize: https://api.twitter.com/oauth/authenticate
  access:   https://api.twitter.com/oauth/access_token
<oauth_app> tw_svallejo
  key:      cnTng4aQMlVTyfGtV04CKT5Sf
  secret:   <hidden>
<credentials> oauth_token, oauth_token_secret
---
```

# TWITTER REST: LOS TUIITS QUE YA FUERON

- Con el Rest API puede acceder a:
  - Tuits de los últimos 6 - 9 días
  - 18,000 tuits cada 15 minutos
- Lo único que debes hacer es dar un o varios términos de búsqueda y comienzas a bajarlos a tu computadora.

```
33 duque_tw <- search_tweets(q = 'duque', # El término que quiere buscar
34                             n = 100, # Cuántos tuits quiere que me devuelve el API
35                             include_rts = T, # Si quiere que incluya RTs (siempre sí... es mejor tener data de más que de menos)
36                             retryonratelimit = TRUE) # Si quiero que automáticamente vuelva a intentar si llego al límite de tuits por 15 min.
37
```

```
> duque_tw %>% as_tibble() # Me permite visualizar mejor la base de datos
```

```
# A tibble: 17,997 x 90
```

	user_id	status_id	created_at	screen_name	text	source
	<chr>	<chr>	<dtm>	<chr>	<chr>	<chr>
1	77912932...	143681501...	2021-09-11 22:12:48	GodosMRobi...	"Resultab...	Twitte...
2	12482831...	143681501...	2021-09-11 22:12:48	pisaslow	"De @Bajo...	Twitte...
3	12482831...	143677988...	2021-09-11 19:53:13	pisaslow	"Daniel M...	Twitte...
4	12482831...	143678094...	2021-09-11 19:57:26	pisaslow	"El proye...	Twitte...
5	12482831...	143669379...	2021-09-11 14:11:07	pisaslow	"¿Dónde e...	Twitte...
6	12482831...	143669690...	2021-09-11 14:23:29	pisaslow	"El rostr...	Twitte...
7	46286192...	143681500...	2021-09-11 22:12:46	fabian0612...	"Venezuel...	Twitte...
8	292673016	143681432...	2021-09-11 22:10:03	GomezSandr...	"Duque ti...	Twitte...
9	292673016	143681500...	2021-09-11 22:12:46	GomezSandr...	"El gobie...	Twitte...
10	84927989...	143678030...	2021-09-11 19:54:52	Fernando27...	"Iván Duq...	Twitte...

```
# ... with 17,987 more rows, and 84 more variables:
```

```
> duque_tw$text[1] # El texto de un tuit
```


```
[1] "Resultaba tan predecible la próxima jugada de Duque y su banda de narcotraficantes instalados en el palacio de Nariño, que el escenario del falso positivo no podía faltar luego de su atormentado intento de torpedear la mesa de Diálogo en México. https://t.co/udYYN1UMUo"
```

# OBTENER LOS TIMELINES

```
> timelines_pols <- get_timelines(c("IvanDuque", "petrogustavo", "ClaudiaLopez"), n=100)
```

```
> timelines_pols %>% as_tibble()
```

```
# A tibble: 300 x 90
```

	user_id	status_id	created_at	screen_name	text	source
	<chr>	<chr>	<dtm>	<chr>	<chr>	<chr>
1	77653794	143681108...	2021-09-11 21:57:12	IvanDuque	#AEstaHora...	Twitte...
2	77653794	143680892...	2021-09-11 21:48:36	IvanDuque	Este golpe...	Twitte...
3	77653794	143680886...	2021-09-11 21:48:22	IvanDuque	En operaci...	Twitte...
4	77653794	143680408...	2021-09-11 21:29:22	IvanDuque	 #EnVivo ...	Twitte...
5	77653794	143667823...	2021-09-11 13:09:18	IvanDuque	He ordenad...	Twitte...
6	77653794	143667812...	2021-09-11 13:08:51	IvanDuque	Lamentamos...	Twitte...
7	77653794	143667629...	2021-09-11 13:01:36	IvanDuque	El Plan Na...	Twitte...
8	77653794	143665068...	2021-09-11 11:19:48	IvanDuque	Hoy expres...	Twitte...
9	77653794	143665068...	2021-09-11 11:19:48	IvanDuque	Tolerarlo,...	Twitte...
10	77653794	143665067...	2021-09-11 11:19:48	IvanDuque	Hace 20 añ...	Twitte...

```
# ... with 290 more rows, and 84 more variables: display_text_width <dbl>,
```

```
> timelines_pols$text[1] # tuit de Duque
```

```
[1] "#AEstaHora Participamos en la entrega del Laboratorio de Diagnóstico Fitosanitario en la Zona Bananera, Magdalena, que permitirá un mayor acompañamiento a los productores, por parte d el @ICACOLOMBIA para la producción de frutos bajo estrictos protocolos de seguridad alimentari a. https://t.co/0Zm35hf0EV"
```

```
> timelines_pols$text[101] # tuit de Petro
```

```
[1] "Con la P de Pacto, de pueblo, de progreso y de Paz.\n\nBarranquilla comienza la construcc ión del gran cambio histórico de Colombia. https://t.co/ugwTJn04tK"
```

```
> timelines_pols$text[201] # tuit de López
```

```
[1] "Gracias, mi querido @luchogarzon! https://t.co/03MjuMIvt9"
```



# INFORMACIÓN SOBRE LOS USUARIOS

```
> usuarios_col <- lookup_users(c("IvanDuque", "petrogustavo", "ClaudiaLopez"))
```

```
> usuarios_col$description
```

```
[1] "Presidente de la República de Colombia \U0001f1e8\U0001f1f4 2018-2022. Legalidad, Emprendimiento y Equidad. #ElFuturoEsDeTodos"  
[2] "Perfil Oficial del dirigente político progresista colombiano Gustavo Petro. Por una Colombia Humana con justicia social y en Paz"  
[3] "Primera Alcaldesa de Bogotá. Orgullosa bogotana y ciudadana. ¡Inspiramos, unimos y ganamos! #ClaudiaAlcaldesa \U0001f49a\U0001f49c #JuntosCuidamosBogotá"
```

# LIKES RECIENTES


```
> like_col <- get_favorites(c("IvanDuque", "petrogustavo", "ClaudiaLopez"))
```

```
> table(like_col$screen_name[like_col$favorited_by=="IvanDuque"]) # Estas son las cuentas a las que Duque le ha dado like
```

AlokSharma_RDG	COMANDANTE_EJC	DanielPalam	DiegoMesaP
1	1	2	7
Fruizgomez	infopresidencia	invimacolombia	IvanDuque
1	375	1	2
karenabudi	linaarbelaes	LucenErnesto	MinAgricultura
2	1	1	1
MincomercioCo	MintrabajoCol	MyriamCMartinez	PosconflictoCO
1	1	1	1
VictoriaBxE			
1			

# SEGUIDORES DE UN USUARIO

```
> follow_duque <- get_followers("IvanDuque") # solo los primeros 5000
> follow_duque %>% as_tibble() # Solo te da el ID.
# A tibble: 5,000 x 1
  user_id
  <chr>
1 1425205601592827906
2 1436798036504719367
3 1436811673944342528
4 1436711912356687943
5 1436809295232999424
6 1405536238287745042
7 1436795148755685384
8 1388135429174611974
9 715639755535421441
10 1433572679148310536
# ... with 4,990 more rows
```



```
> follow_duque_usuario <- lookup_users("1388135429174611974")
> follow_duque_usuario$description
[1] "Humano, Terrícola y venezolano (en ese orden)"
```

# API STREAM: EL FLUIR DE LOS TUIITS

- Si queremos saber en tiempo real los tuits que están circulando, podemos utilizar la función *stream\_tweets*:

```
> colombia_stream <- stream_tweets("colombia", # El término que quiero buscar. Si utilizo "" m
e devuelve un sample de TODA la data de TW
+                               timeout = 100) # Por cuánto tiempo quiero que corra.
Streaming tweets for 100 seconds...
Finished streaming tweets!
```



- ¡Listo! Con estas herramientas tienes todo lo necesario para bajar la data que vas a necesitar en tu camino hacia el análisis de las redes sociales.
- (Pregunten sobre el Academic API de Twitter)

