

# **Análisis de datos de Twitter en el Día Mundial de la Salud**

---

**Xavier Vivancos García**

**Consultora: Dra. Laia Subirats Maté**

**Máster en Inteligencia de Negocio y Big Data**

---

---

# Índice

## 1. Objetivos

## 2. Estado del arte

## 3. Captura de datos

## 4. Análisis de datos

## 5. Conclusiones

---

## Objetivos

- Capturar *tweets* durante el Día Mundial de la Salud (7 de abril)
- Analizar el contenido de los *tweets*
- Determinar la actitud o sentimientos de los autores de los *tweets*
- Estudiar diferencias entre *tweets* escritos en inglés, castellano y catalán
- Dibujar y describir el grafo formado entre *tweets* y *retweets* (usuarios y *tweets* más relevantes)
- Determinar en qué países la concienciación en Twitter es mayor/menor

---

# Índice

1. Objetivos

2. Estado del arte

3. Captura de datos

4. Análisis de datos

5. Conclusiones

---

# Estado del arte

## 1. Análisis de sentimientos

Librerías de R tidyverse y tidytext

Lexicones AFINN, Bing y NRC

## 2. Palabras más frecuentes y elaboración de los *wordclouds*

Librerías de R tm y wordcloud2

## 3. Geolocalización de los *tweets*

Librerías de R ggmap y leaflet

## 4. Creación y visualización del grafo

Librería de R igraph

Software Gephi

---

# Índice

1. Objetivos

2. Estado del arte

**3. Captura de datos**

4. Análisis de datos

5. Conclusiones

---

# Captura de datos

- **Herramienta**  
RStudio
- **Librerías**  
ROAuth, streamR y twitterR
- **Hashtags o palabras clave**  
#WorldHealthDay  
#DiaMundialDeLaSalud  
#DiaMundialDeLaSalut  
#SalutPerATothom  
#HealthForAll  
#WHD2018
- **Duración**  
3 horas. Se inició el día 7 de mayo a las 15:17 y finalizó a las 18:17
- **Resultado**  
42.838 *tweets*

---

# Índice

1. Objetivos

2. Estado del arte

3. Captura de datos

4. Análisis de datos

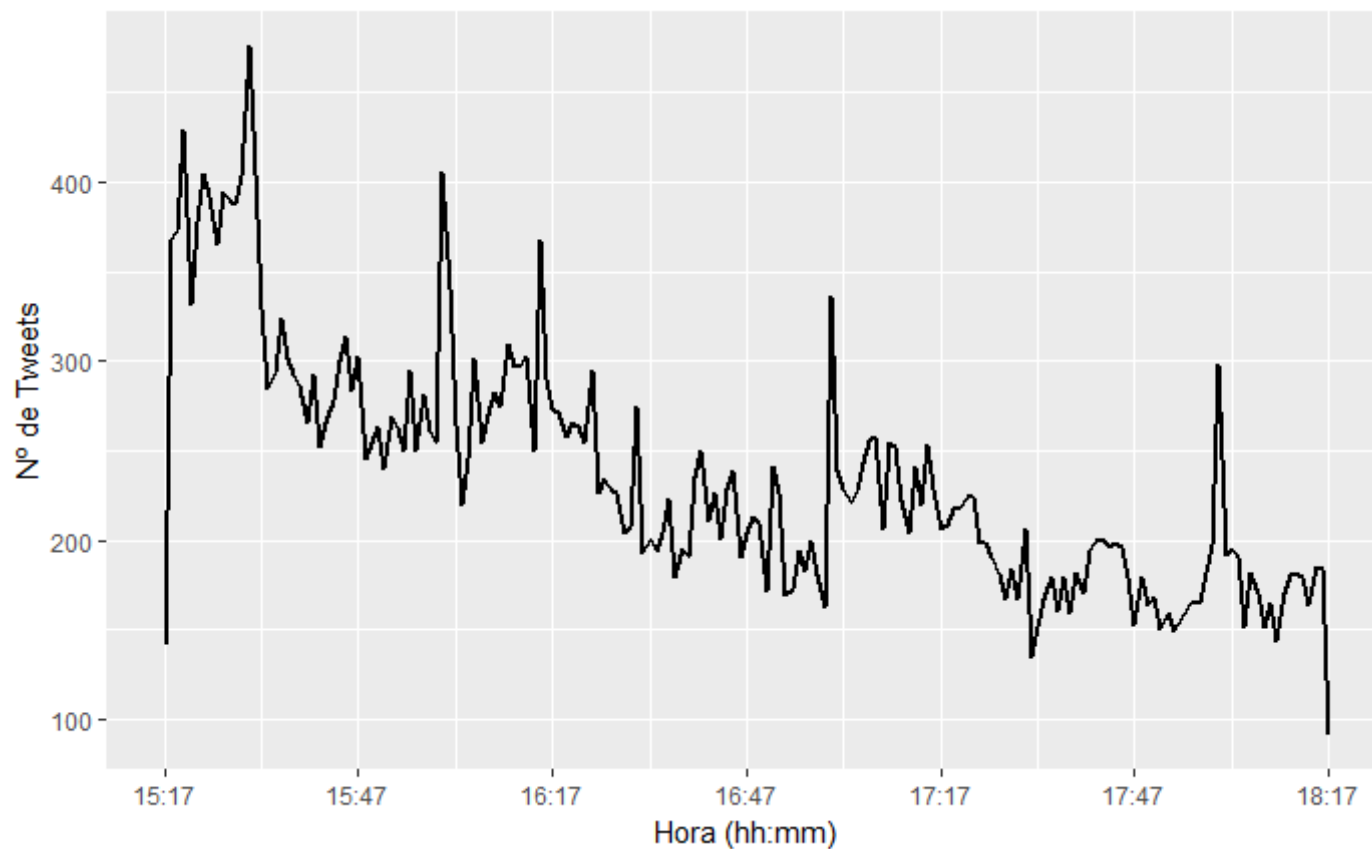
5. Conclusiones

---



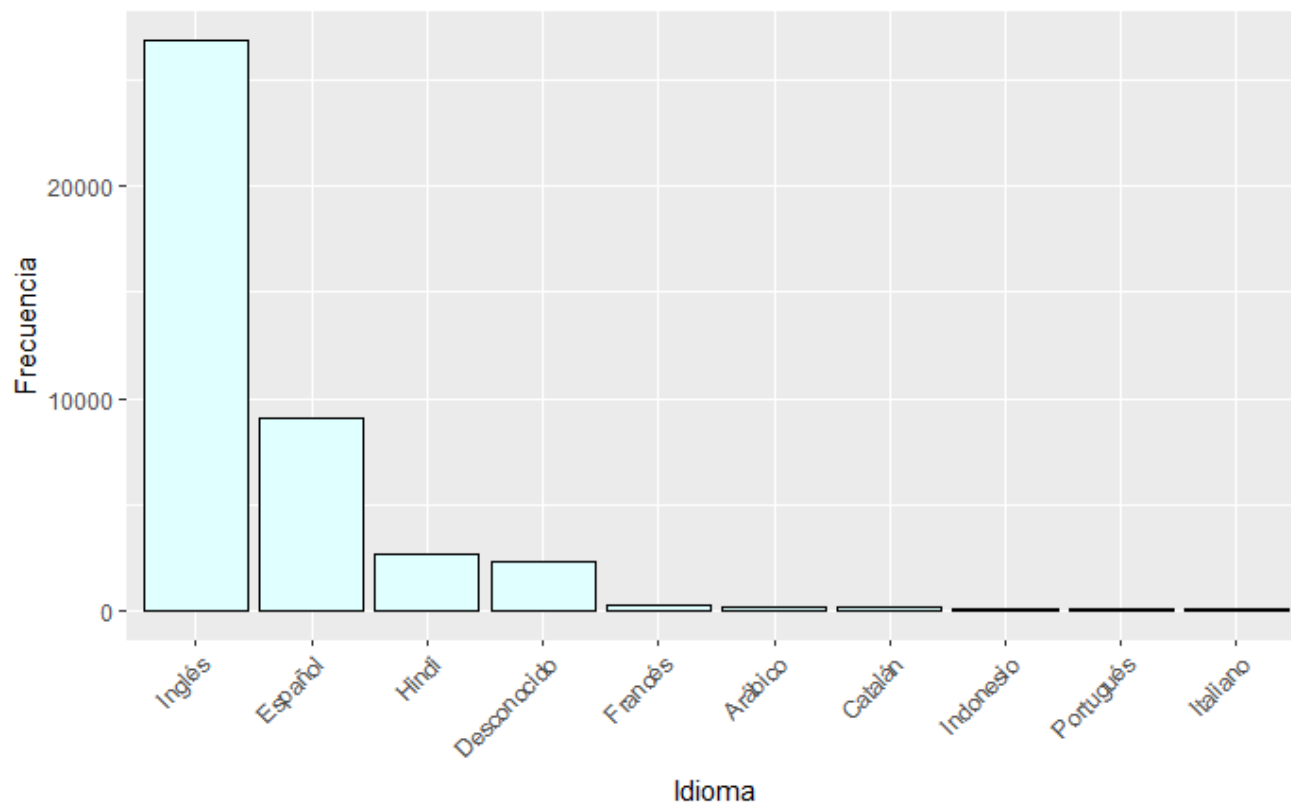
## Análisis de datos

Número de *tweets* durante la captura



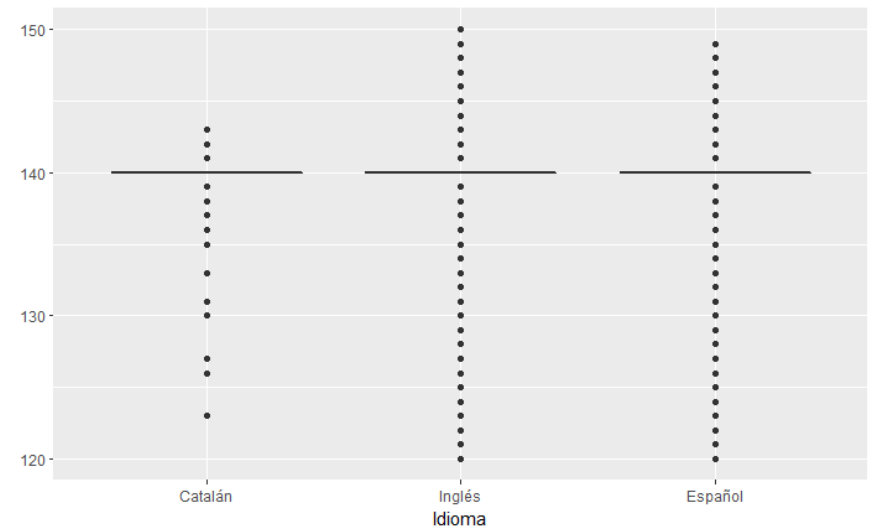
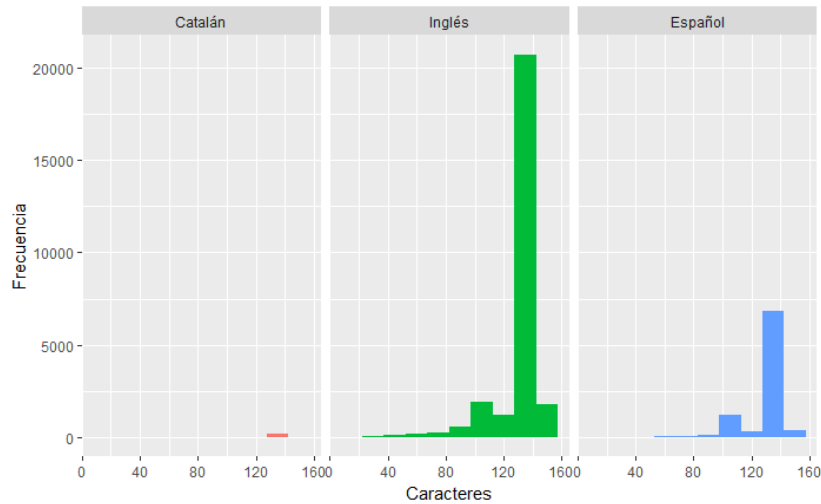
# Análisis de datos

## Idiomas más frecuentes



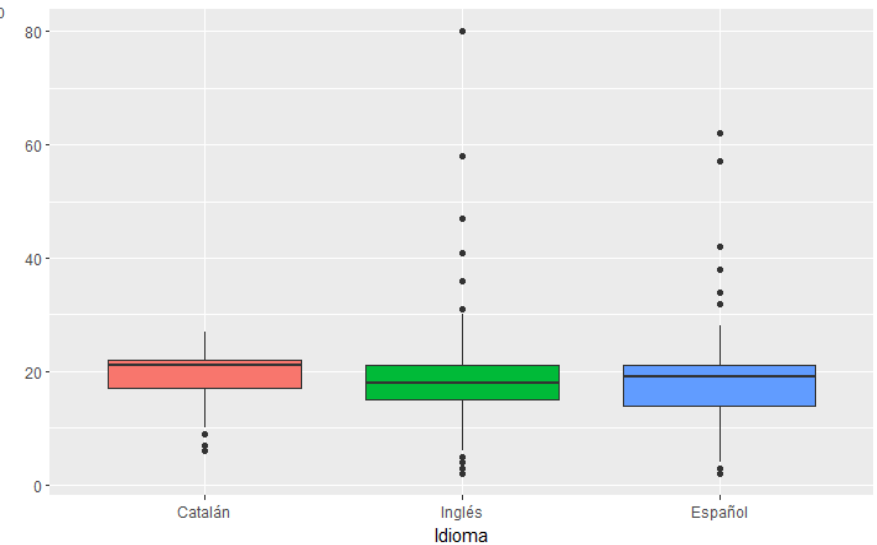
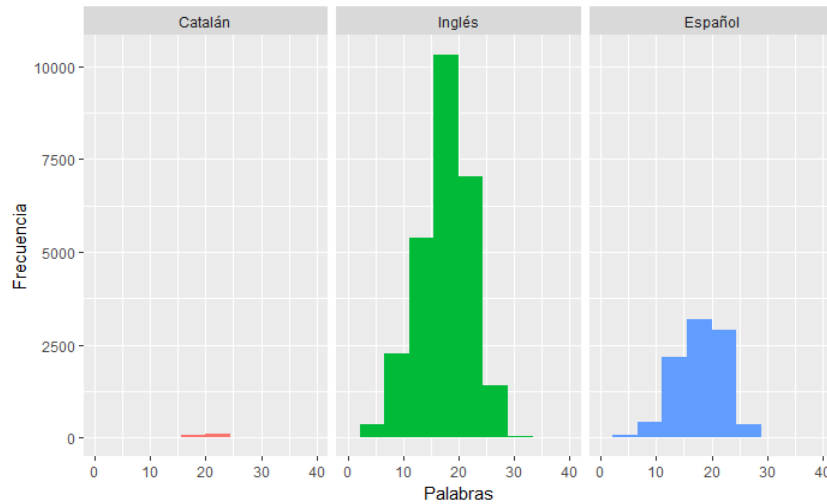
# Análisis de datos

Histogramas y *boxplots* del número de caracteres de los *tweets*



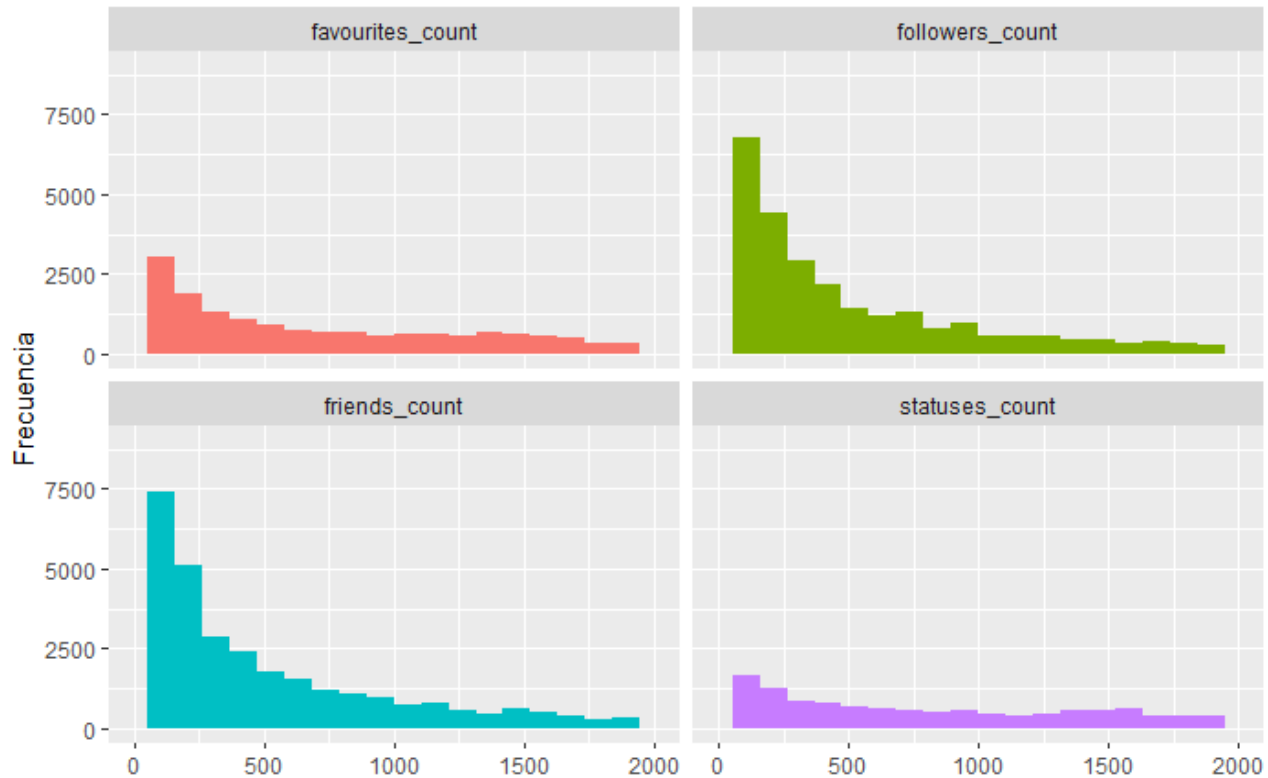
# Análisis de datos

Histogramas y *boxplots* del número de palabras de los *tweets*



# Análisis de datos

Histogramas de algunos atributos de los usuarios



## Análisis de datos

*Wordcloud* con las palabras más frecuentes en la descripción del perfil de los usuarios



## Análisis de datos

Wordclouds con las palabras más frecuentes de los tweets



Inglés



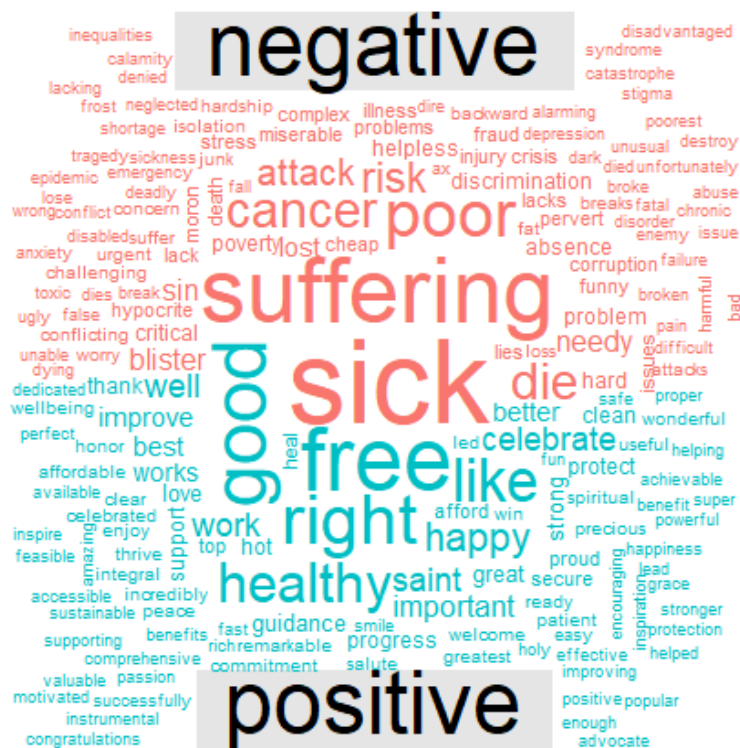
## Castellano



## Catalán

# Análisis de datos

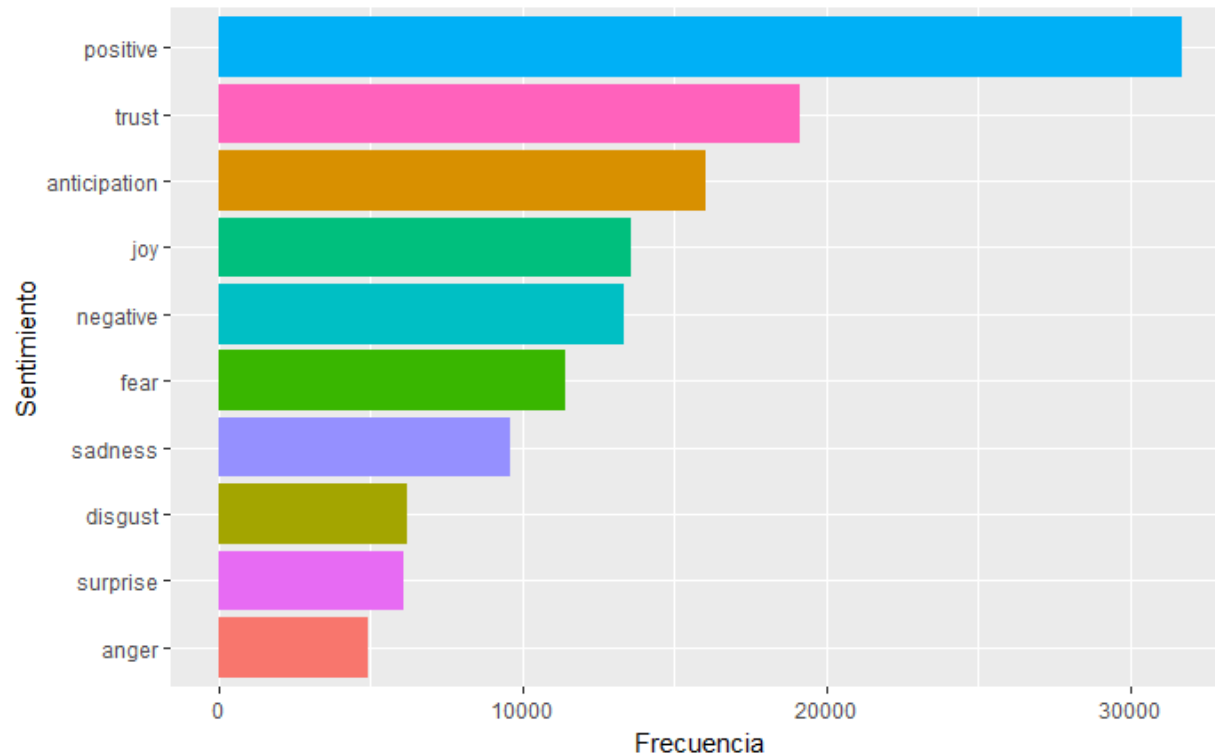
Palabras positivas y negativas (lexicón Bing)





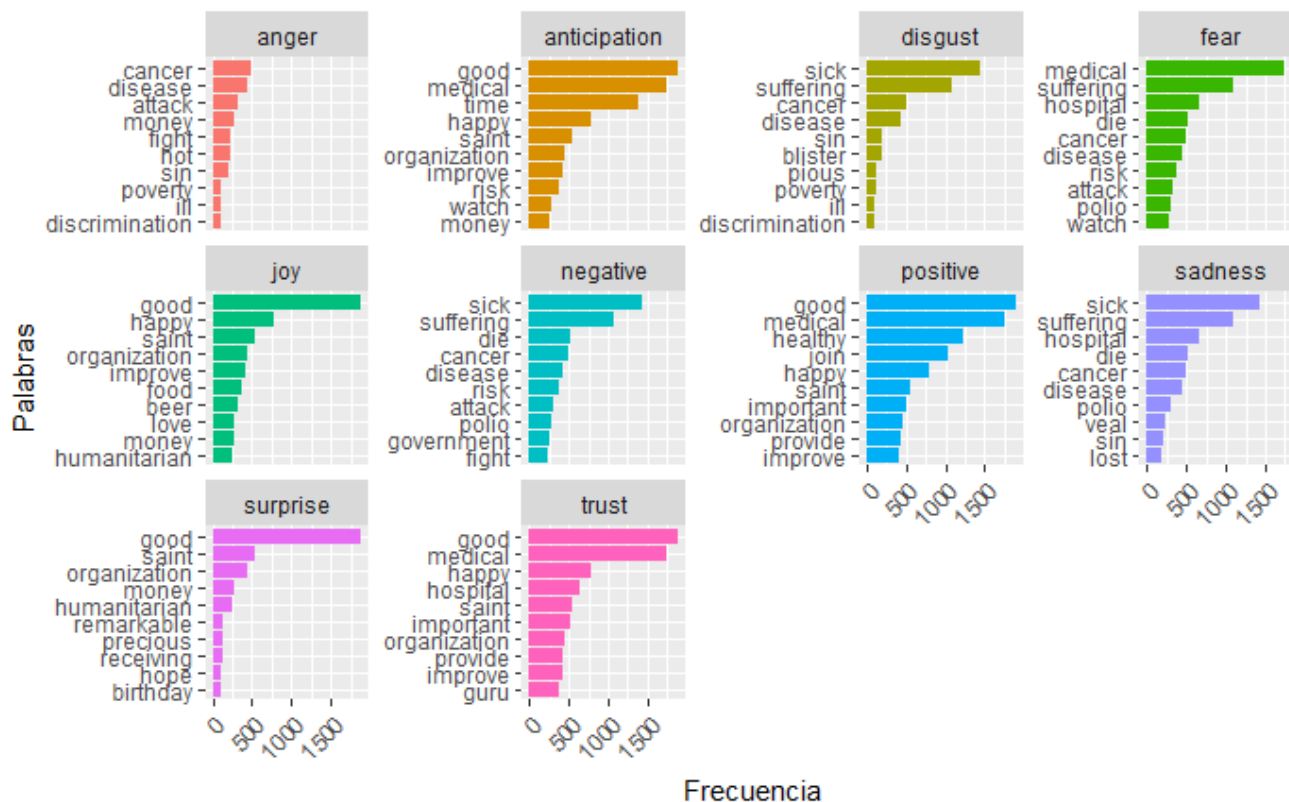
# Análisis de datos

Frecuencia de cada sentimiento en los *tweets* (lexicón NRC)

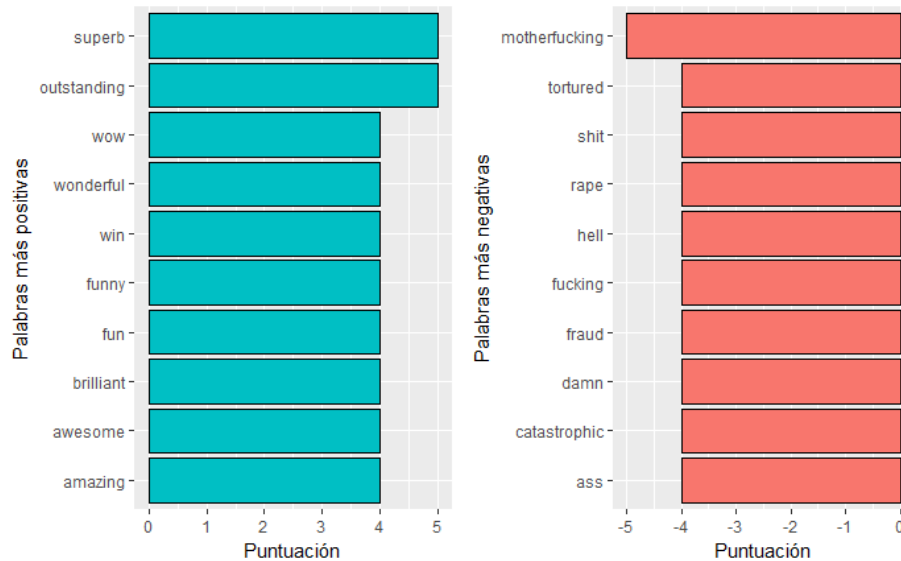


# Análisis de datos

Palabras más frecuentes para cada sentimiento (lexicón NRC)

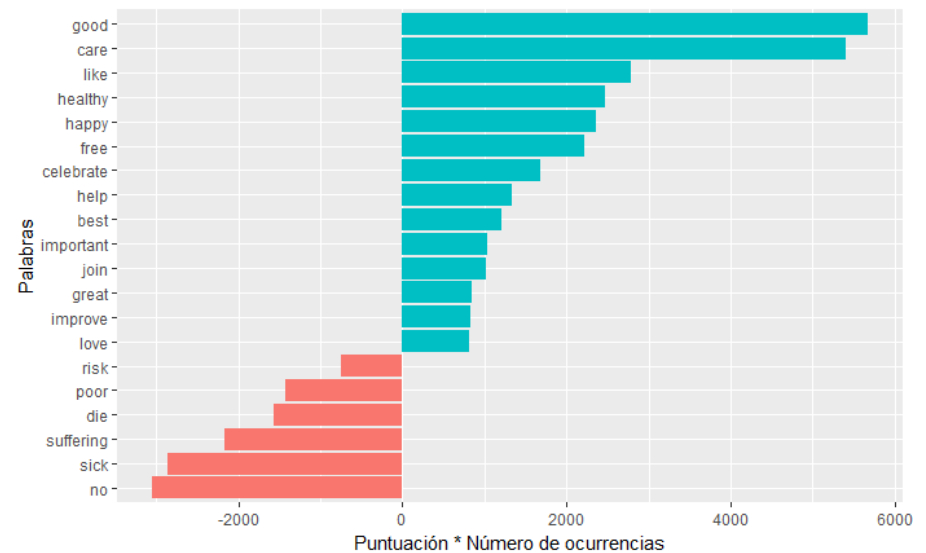


## Análisis de datos



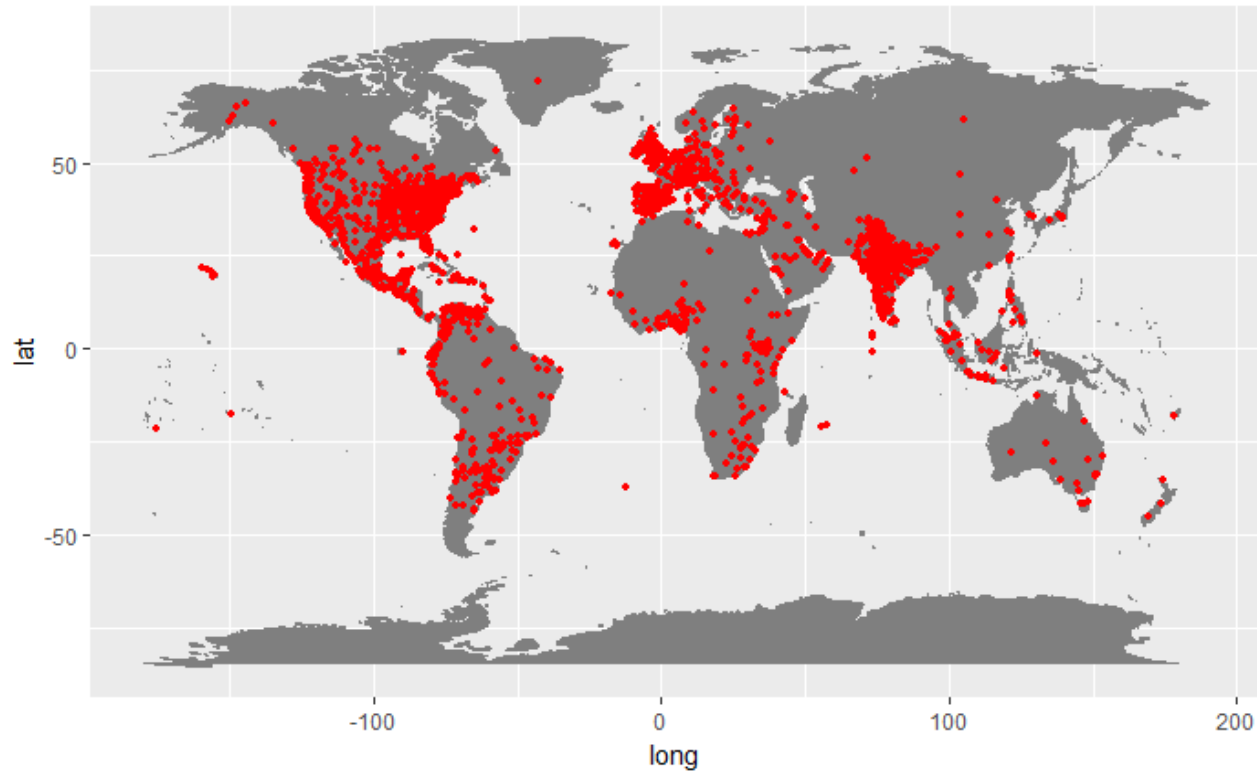
### Palabras más positivas y negativas (lexicón AFINN)

Palabras con más contribución  
(puntuación según el lexicón AFINN \* frecuencia)



## Análisis de datos

¿Desde dónde se han escrito más *tweets*?



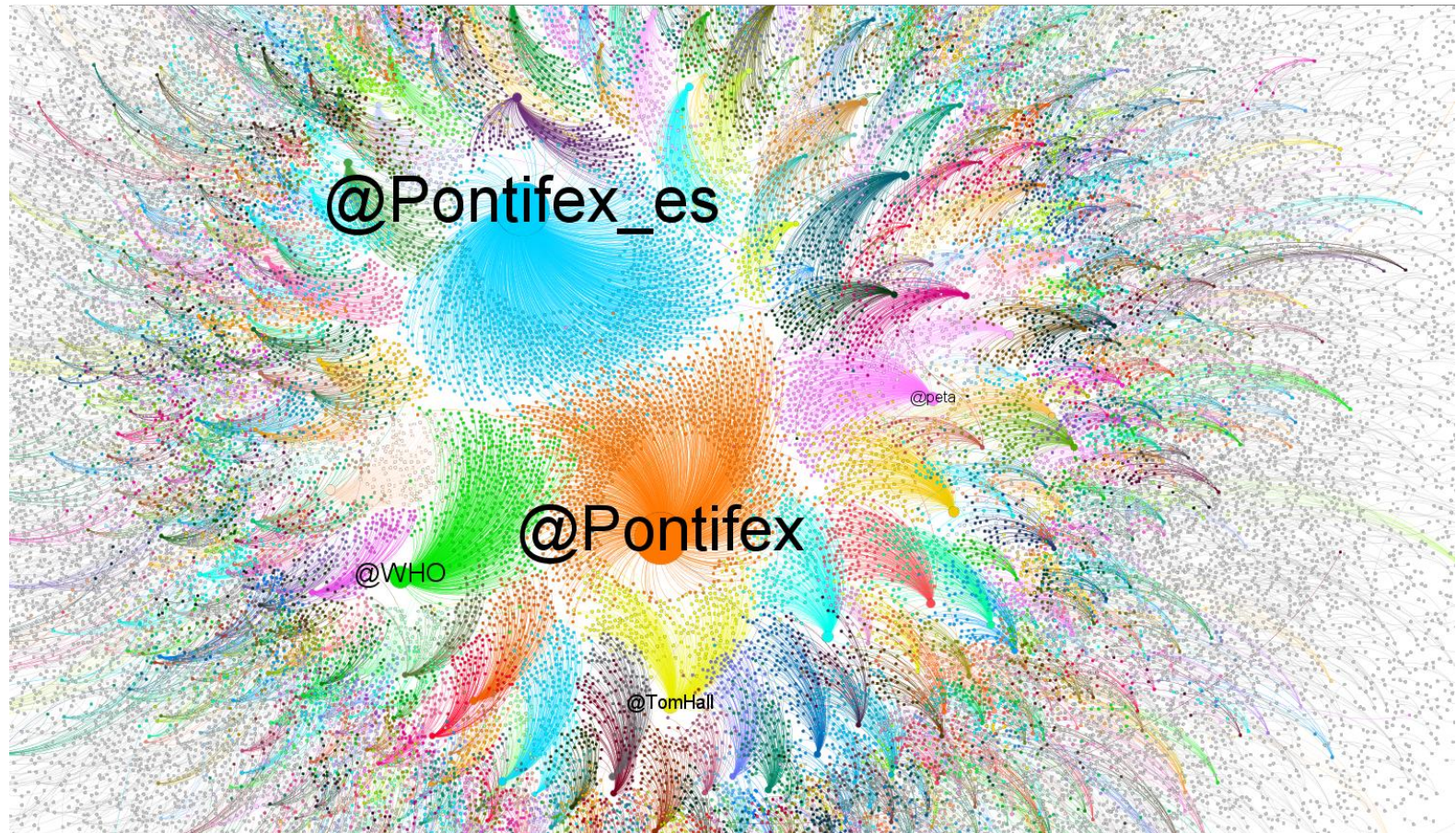
Mapa interactivo en <http://rpubs.com/xvivancos/tfm>

## Análisis de datos

Grafo de *tweets* y sus *retweets*

Nodos → *Tweets* implicados

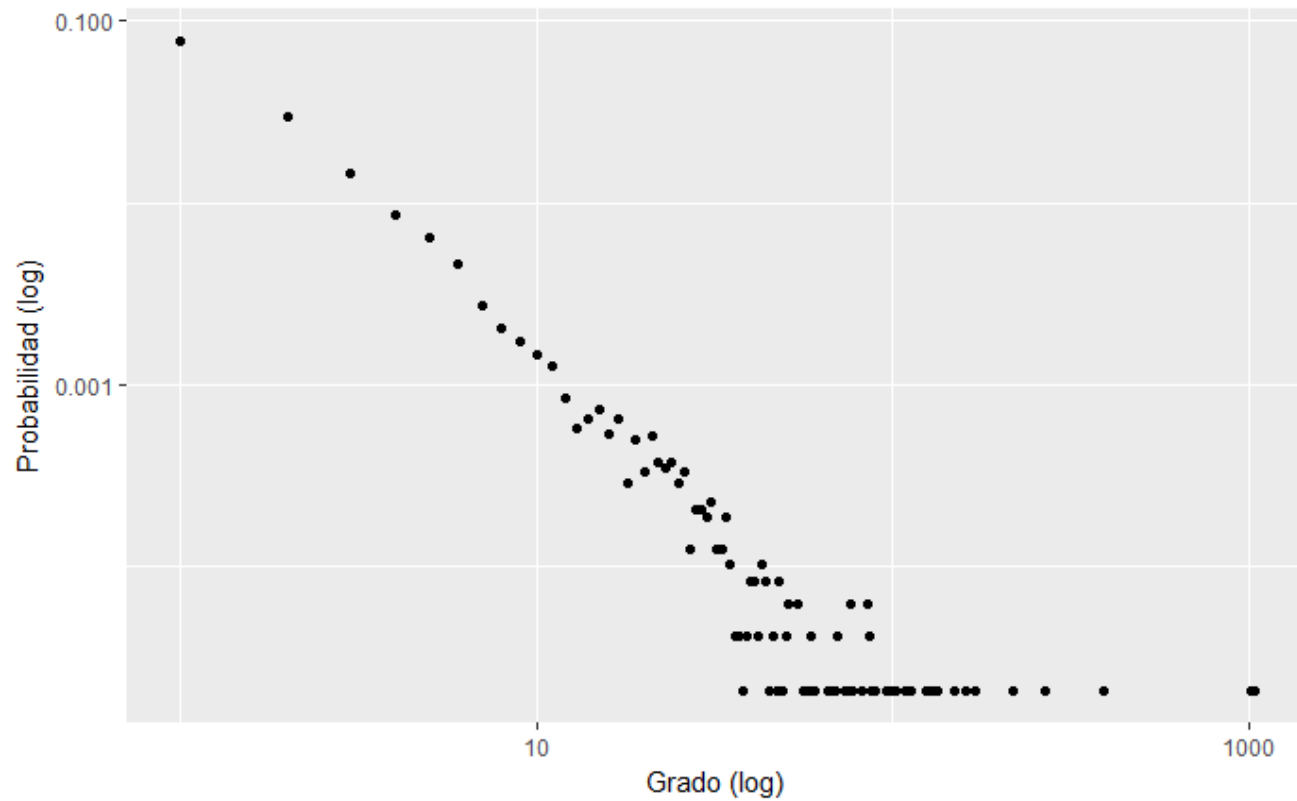
Arcos → Existencia de un *retweet* entre dos *tweets*





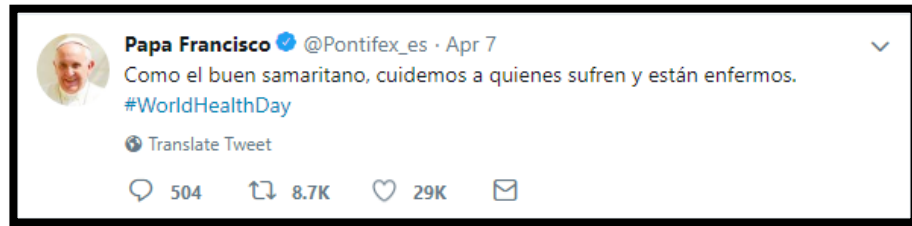
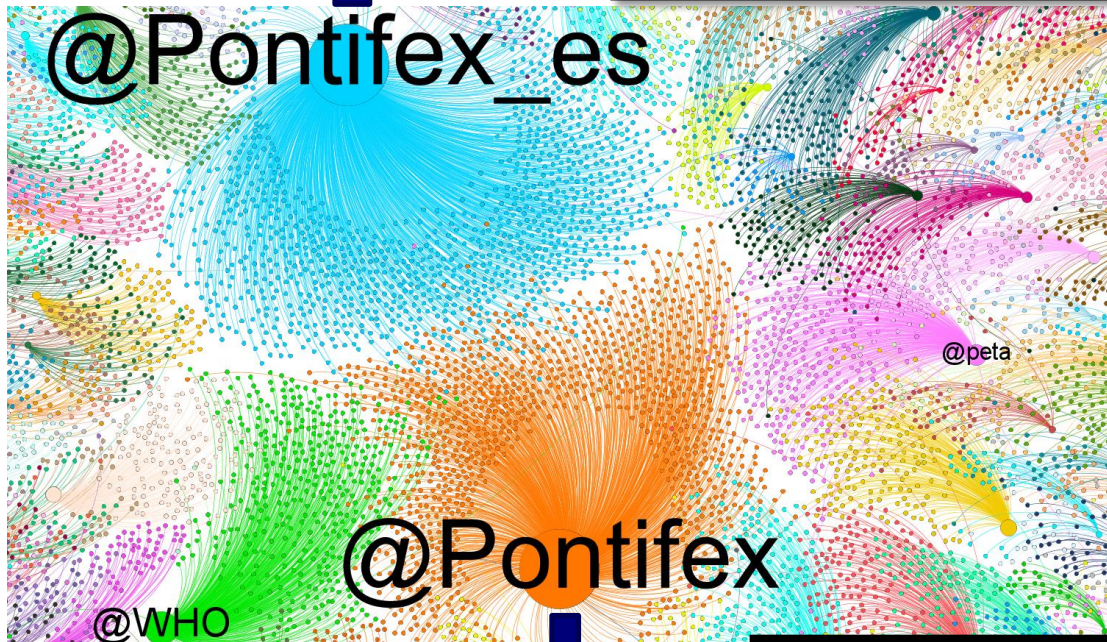
# Análisis de datos

## Distribución de los grados de entrada



## Análisis de datos

Tweets más relevantes



# Análisis de datos

## Top 5 tweets en castellano más relevantes

**Papa Francisco** @Pontifex\_es · Apr 7

Como el buen samaritano, cuidemos a quienes sufren y están enfermos.  
#WorldHealthDay

Translate Tweet

505 8.7K 29K

**Miguel Pizarro** @Miguel\_Pizarro · Apr 7

Venezuela en 1990: Un país pionero en salud y tratamientos de distintas enfermedades. Venezuela 2018: Un país con muertes diarias por la ausencia de medicinas, tratamientos, insumos, equipos y personal calificado. Hoy #DiaMundialDeLaSalud vale la pena recordar porque luchamos.

Translate Tweet

14 301 140

Show this thread

**Guardia Civil** @guardiacivil · Apr 7

#DiaMundialdeLaSalud

¿Dice tu colega que con pastis aguanta más tiempo estudiando? Ningún examen vale más que tu vida

Translate Tweet



13 162 275

**Fighter** @Orion\_fighter · Apr 7

La cacareada jornada de vacunacion (controlada por el carnet de la patria) es otra mentira del narcoregimen. Esta es la declaracion de un padre en #ElParaíso "NO VAN A COLOCAR VACUNAS, ES MENTIRA" #DiaMundialDeLaSalud en Venezuela #7Abr

Translate Tweet



"Hoy 6 de abril hasta el 13 de mayo arranca en Venezuela un proceso de vacunación general totalmente gratuito con más de once millones de dosis gratuitas. Nuestra meta es vacunar a nueve millones de personas (...) a nivel nacional", anunció Maduro en un acto de Gobierno transmitido por el canal estatal VTV.

**jufran** @juantatu · 1m  
Replying to @Orion\_fighter  
La jornada de vacunación que anuncio el gobierno es pura paja señores ni vayan estamos regresando de alla no hay ninguna vacuna de fiebre amalla. Otro engaña bobos de este goniermo tan util.

3 288

**Unidad Ciudadana** @UniCiudadanaAR · Apr 7

Desde que Vidal asumió no inauguró ningún centro de salud y mantiene cerrados seis hospitales de alta complejidad que construyó la gestión anterior.

Nota de @nkreplak  
#DiaMundialDeLaSalud #VidalCierraHospitales

Translate Tweet



**Vidal y sus elefantes blancos**  
unidadciudadana.org

41 566 398



# Análisis de datos

## Top 5 tweets en catalán más relevantes



**Socialistes (PSC)** @socialistes\_cat · Apr 7

Les famílies no haurien d'haver de triar entre gastar en salut o en altres necessitats, com aliments i educació.

#DiaMundialdeLaSalut #WorldHealthDay

Ana SB Retweeted



**Gemma Tarafa** @GemmaTarafa · Apr 7

En el #DiaMundialDeLaSalut ens sumem a la crida de la OMS per la cobertura universal #SalutPerATothom

Cal defensar més que mai l'accés universal a la Sanitat front l'ofensiva del Govern central i 155

Translate Tweet



**Antoni Trilla** @ToniTrilla · Apr 7

Avui es el Dia Mundial de la Salut.

La Salut es un Dret Huma.

Necessitem que hi hagi cobertura sanitaria universal. Cobertura per tothom i a tot arreu.

[bit.ly/WHDHealthForAll](https://bit.ly/WHDHealthForAll)

@WHO

#HealthForAll

#WorldHealthDay

@hospitalclinic

@UniBarcelona

@COMBarcelona

Translate Tweet



### Universal Health Care

The world has the right to Universal Health Coverage!

#HealthForAll

[apps.who.int](https://apps.who.int)

20



106



168



Sandra Retweeted

**Vall d'Hebron Campus** @vallhebron · Apr 7

Som un hospital públic de referència compromès amb una assistència sanitària universal de màxima qualitat. Posem tot el talent, la recerca, la docència i la innovació a l'abast de tothom. Avui i cada dia #DiaMundialdeLaSalut #SalutPerATothom #HealthForAll #WorldHealthDay

Translate Tweet



carol domenech Retweeted

**Bemba Saoco** @BembaSaoco · Apr 7

El #RumbaLlarena ha arribat a #Baqueira

@BembaSaoco is in da haus!

#Puigdemont #NoCallarem #DiaMundialDeLaSalut #LlibertatPresosPolitics

#llibertatdexpressió

Translate Tweet



2



61



99



---

# Índice

1. Objetivos
  2. Estado del arte
  3. Captura de datos
  4. Análisis de datos
  - 5. Conclusiones**
-

## Conclusiones

- El número de *tweets* por minuto descende a medida que pasan las horas
- El idioma más frecuente y con amplia diferencia es el inglés
- Las distribuciones del número de caracteres y palabras son muy parecidas para los tres idiomas analizados
- Las palabras más repetidas son similares para los tres idiomas, y se observan términos religiosos con bastante frecuencia (pontífice o samaritano, por ejemplo)
- La mayoría de palabras están asociadas con sentimientos positivos: confianza, esperanza y alegría
- Los sentimientos más tóxicos (disgusto, enfado y tristeza) tienen muy poca representatividad en los *tweets*
- En el mapa, encontramos mucha actividad principalmente desde el oeste de Europa, la India, Estados Unidos, México y Venezuela
- La religión tiene mucho peso durante la jornada, perteneciendo los dos *tweets* más relevantes del análisis al Papa Francisco
- Tres de los cinco *tweets* más relevantes en castellano critican la gestión del sistema sanitario en Venezuela y Argentina
- En los *tweets* más relevantes en catalán se acentúa la importancia de la cobertura sanitaria universal y se observan algunas críticas hacia el 155 y el encarcelamiento de los políticos catalanes

# ¡Muchas gracias!

