install.packages("readr") # solo la primera vez

library(readr)

datos <- read\_csv("C:/Users/MI PC/OneDrive/Desktop/COVID19.csv")

**Manipulación de datos**

library(dplyr)

library(tidyr)

library(stringr)

library(readr)

head(datos)

colnames(datos)

# Ver estructura general del dataframe

glimpse(datos)

**Filtrado de valores perdidos y países**

datos\_esp <- datos %>%

filter(country\_code == "ES" & !is.na(country\_code))

library(stringr)

Nos quedamos con una base de 24000 tweets

**Instalación de paquetes**

install.packages("tidytext")

install.packages("stopwords")

library(tidytext)

library(stopwords)

library(dplyr)

**Buscar palabras clave relacionadas**

#Palabras asociadas a la ansiedad

palabras\_ansiedad <- c("ansiedad", "ansiosa", "ansioso", "nervios", "nerviosa", "nervioso", "ataque", "pánico")

# Palabras asociadas a depresión

palabras\_depresion <- c("depresión", "triste", "tristeza", "vacío", "llorar", "llanto", "desesperanza", "infeliz", "abatido", "deprimido", "deprimida")

# Unimos ambas listas

palabras\_clave <- c(palabras\_ansiedad, palabras\_depresion)

# Filtramos los tweets que contienen al menos una palabra clave

library(ggplot2)

tweets\_relacionados <- datos\_esp %>%

filter(str\_detect(str\_to\_lower(text), str\_c(palabras\_clave, collapse = "|")))

colnames(tweets\_relacionados)

tweets\_relacionados <- tweets\_relacionados %>%

mutate(fecha = as.Date(created\_at))

frecuencia\_diaria <- tweets\_relacionados %>%

count(fecha)

#Gráfico de línea

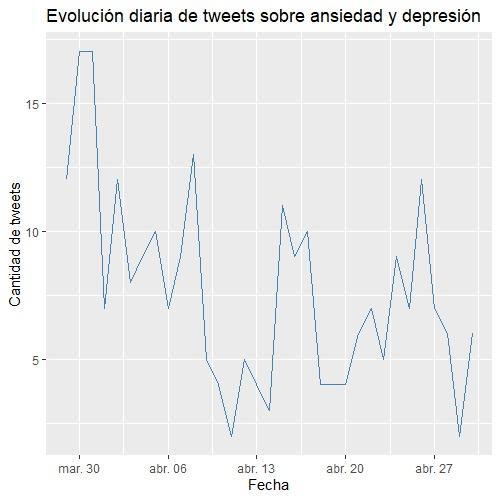
ggplot(frecuencia\_diaria, aes(x = fecha, y = n)) +

geom\_line(color = "steelblue") +

labs(title = "Evolución diaria de tweets sobre ansiedad y depresión",

x = "Fecha",

y = "Cantidad de tweets")



**Nube de palabras**

install.packages("wordcloud")

install.packages("RColorBrewer")

install.packages("stringr")

library(wordcloud)

library(RColorBrewer)

library(dplyr)

library(tidytext)

library(stopwords)

library(stringr)

tokens\_emocionales <- tweets\_relacionados %>%

select(text) %>%

unnest\_tokens(word, text) %>%

filter(!word %in% stopwords("es")) %>%

filter(str\_detect(word, "[a-z]"))

# Conteo de frecuencia

frecuencias <- tokens\_emocionales %>%

count(word, sort = TRUE)

library(wordcloud)

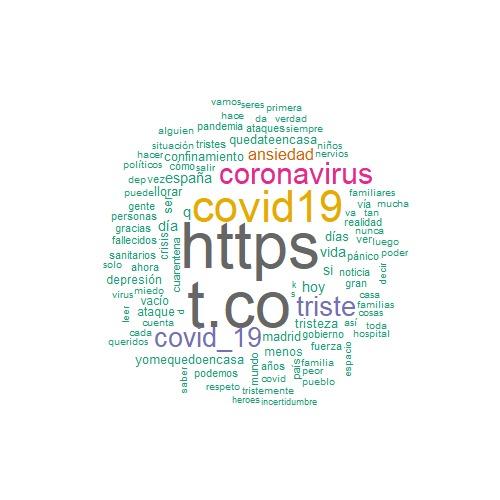
wordcloud(words = frecuencias$word,

freq = frecuencias$n,

max.words = 100,

random.order = FALSE,

colors = brewer.pal(8, "Dark2"))



**Análisis de emociones a través de un diccionario**

nrc <- read.csv("C:/Users/marie/Documents/Data Sciencie/Trabajo-Final-/nrc\_espanol.csv", encoding = "UTF-8")

nrc <- nrc %>%

filter(valor == 1) %>%

select(palabra, sentimiento)

# Cruzamos los tokens con las emociones del diccionario

sentimientos <- tokens\_emocionales %>%

inner\_join(nrc, by = c("word" = "palabra"))

#Contamos las emociones encontradas

conteo\_emociones <- sentimientos %>%

count(sentimiento, sort = TRUE)

print(conteo\_emociones)

#Visualizador en gráfico

library(ggplot2)

ggplot(conteo\_emociones, aes(x = reorder(sentimiento, n), y = n)) +

geom\_col(fill = "steelblue") +

coord\_flip() +

labs(title = "Emociones detectadas en los tweets relacionados con salud mental",

x = "Emoción",

y = "Frecuencia")

