

Proyecto visualización: Imagen Covid

Grupo K

2024-03-05

En primer lugar, vamos a cargar todas las librerías que necesitamos. Y además, vamos a ‘quitar’ los warnings que nos puedan aparecer.

Seguidamente, importaremos la información con la que vamos a trabajar y la guardaremos en una variable (el cual, más adelante, será el data frame con el que vamos a trabajar) y este se llamará ‘datos’.

```
rm(list=ls())

#Cargamos las librerías

#Para que no salgan los warnings:
suppressWarnings({
  library(readr)
  library(ggplot2)
  library(plotly)
})

#Importamos los datos
datos <- read.csv("owid-covid-data.csv")

#Visualizamos los datos, está en comentario para que a la hora de compilar el
#documento no aparezcan reflejados dichos datos.

#head(datos)
```

A continuación, seleccionamos y cargamos los países que tenemos que mostrar en nuestro gráfico según el enunciado:

“United States”, “Italy”, “France”, “Germany”, “Canada”, “Japan”, “United Kingdom” y “Spain”.

```
#Cargamos los países que hemos de mostrar

us <- datos[datos$location == "United States",]
it <- datos[datos$location == "Italy",]
fr <- datos[datos$location == "France",]
ge <- datos[datos$location == "Germany",]
ca <- datos[datos$location == "Canada",]
ja <- datos[datos$location == "Japan",]
uk <- datos[datos$location == "United Kingdom",]
sp <- datos[datos$location == "Spain",]
```

Una vez ya hemos seleccionado los países que vamos a representar, creamos un nuevo data frame únicamente con la información de estos. Este nuevo data frame se va llamar ‘df’:

```
#Creamos un nuevo data frame con los datos que nos interesan
#(España lo añadimos más tarde, puesto que tiene un grosor distinto):
```

```
df <- rbind(us, it, fr, ge, ca, ja, uk)
```

```
#Lo mostramos, mismo motivo de comentario que en el caso anterior
#head(df)
```

Ahora, convertimos la columna 'date' de España a formato fecha.

```
#Convertimos la columna 'date' a formato de fecha
sp$date <- as.Date(sp$date, format = "%Y-%m-%d")
```

```
#Nos aseguramos de que la columna 'date' en el conjunto de datos df ya esté en
#formato de fecha.
df$date <- as.Date(df$date, format = "%Y-%m-%d")
```

El siguiente paso ya sería crear el gráfico. En el eje 'x' ponemos las fechas para que se vea en el paso del tiempo, y en el eje 'y' las muertes por millón. Para esta última variable cogemos smoothed (suavizadas), para que salga el gráfico más adecuado.

```
#Creamos el gráfico con la información que tenemos
gr <- ggplot(df, mapping = aes(x = date, y = new_deaths_smoothed_per_million,
                               group = location))
```

```
#Definimos los colores del enunciado
colores <- c("United States" = "mediumpurple3", "Italy" = "aquamarine3",
             "France" = "orange3", "Germany" = "green4", "Canada" = "tomato4",
             "Japan" = "steelblue4", "United Kingdom" = "violetred1")
```

```
#Añadimos línea con color a los países, color según país (más tarde
#añadiremos España)
```

```
gr_col <- gr + geom_line(aes(color = location)) +
```

```
  #Colores como en el ejemplo
  scale_color_manual(values = colores)
```

```
#Definimos las fechas que queremos mostrar en el eje x
fechas_mostrar <- as.Date(c("2020-05-13", "2021-04-08", "2022-03-24",
                           "2023-06-02", "2024-02-29"))
```

```
#Convertimos las fechas a formato que ggplot puede entender
fechas_labels <- format(fechas_mostrar, "%b %d, %Y")
```

```
# Ponemos como se va a mostrar la fecha y las fechas específicas que queremos
#mostrar
```

```
gr_col_fecha <- gr_col + scale_x_date(date_labels = "%b %d, %Y",
                                       breaks = fechas_mostrar, labels =
                                       fechas_labels) +
```

```
  #Para que vaya de 2 en 2 y el expand para que el 0 coincida con el eje x
  scale_y_continuous(breaks=seq(0, 18, by=2), expand = c(0, 0)) +
  geom_hline(yintercept = c(2, 4, 6, 8, 10, 12, 14, 16, 18), color = "gray80",
            linetype = "dashed", size = 0.2)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
#Añadimos el título y subtítulo. Tenemos en cuenta que, en este caso, no sería
#necesario poner nombres a los ejes pues el título del gráfico ya da la
#suficiente información.
```

```
#Además, establecemos 'caption' donde quedará reflejada la fuente de los datos:
```

```
gr_final <- gr_col_fecha + labs(title = "Daily new confirmed COVID-19 deaths per million people - Grupo
  subtitle = "7-day rolling average. For some countries the number of confirmed deaths is much lower than
  This is because of limited testing and challenges in the attribution of
```

```
  x = "", y = "",
  caption = "Source: Johns Hopkins University CSSE COVID-19 Data") +
```

```
#Añadimos España más grueso y de color rojo
```

```
geom_line(data = sp, aes(x = date, y = new_deaths_smoothed_per_million),
```

```
#Con la función 'theme_classic' ponemos el fondo en blanco
  color = "red", linewidth = 1.1) + theme_classic()
```

```
#Hacemos el subtítulo más pequeño para que se visualice
```

```
#de manera correcta:
```

```
gr_final <- gr_final + theme(plot.subtitle = element_text(size = 8))
```

```
#Visualizamos gráfico sin ser interactivo primero:
```

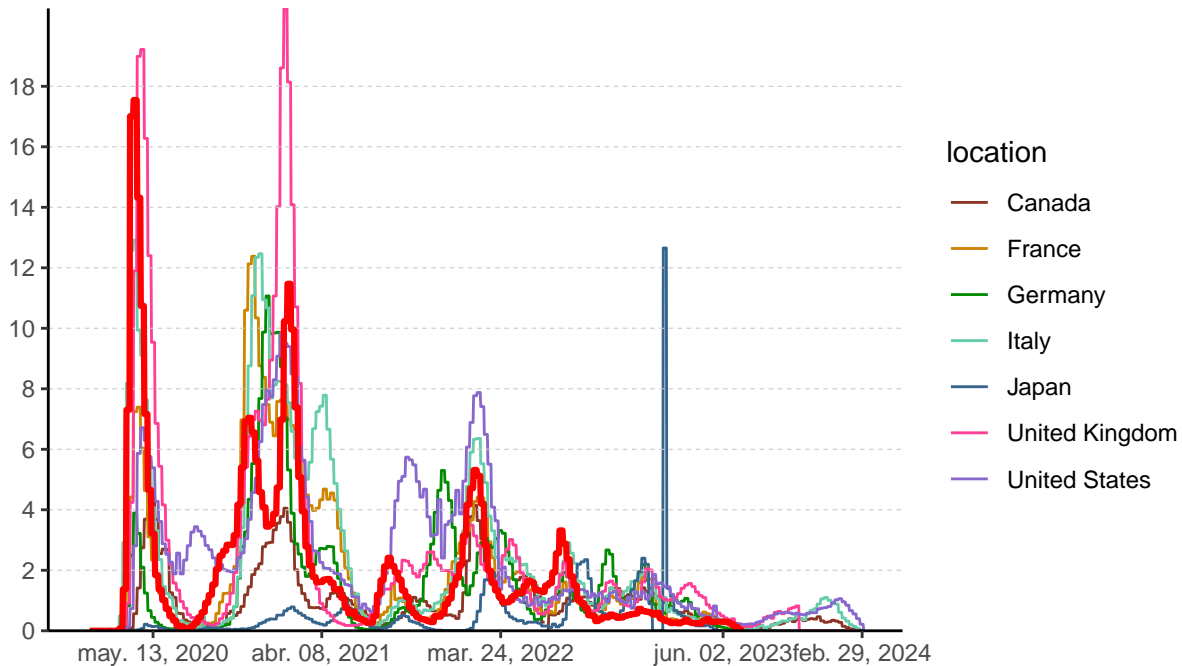
```
gr_final
```

```
## Warning: Removed 519 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

```
## Warning: Removed 243 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

Daily new confirmed COVID-19 deaths per million people – Grupo K

7-day rolling average. For some countries the number of confirmed deaths is much lower than the true number of deaths. This is because of limited testing and challenges in the attribution of the cause of death



Source: Johns Hopkins University CSSE COVID-19 Data

A continuación, una vez hemos terminado nuestro gráfico, vamos a hacerlo interactivo.

Esto lo conseguimos con la función ‘ggplotly()’ pasándole como argumento el nombre de nuestro gráfico, en este caso: “gr_final”. De esta forma, podemos observar de manera más detallada la información que este contiene, como por ejemplo, las muertes exactas en un determinado país en un día en concreto. Esto nos permite extraer conclusiones mucho más específicas.

#Convertimos a gráfico interactivo con plotly.

Añadimos subtítulo y leyenda al gráfico interactivo, pues a este no se añaden automáticamente habiéndolas definido en el normal.

```
final <- ggplotly(gr_final) %>%
```

```
  layout( #Bajamos un poco a leyenda para que no choque con el subtítulo
```

```
    legend = list(title = list(text = "Country"),
```

```
                  x = 1.05, y = 0.5, #posición de la leyenda
```

```
                  xanchor = "left", yanchor = "middle"), #anclaje de la leyenda
```

```
    title = list(text = "Daily new confirmed COVID-19 deaths per million people - Grupo K"),
```

```
    annotations = list(
```

```
      list(
```

```
        text = "7-day rolling average. For some countries the number of confirmed deaths is much lower than the true number of deaths. This is because of limited testing and challenges in the attribution of the cause of death",
```

```
        x = 0, #ubicación del subtítulo
```

```
        y = 1.05,
```

```
        xref = "paper",
```

```
        yref = "paper",
```

```
        showarrow = FALSE,
```

```

    font = list(size = 10) #tamaño del subtítulo
),

#Además, añadimos el caption por el mismo motivo
list(
  text = "Source: Johns Hopkins University CSSE COVID-19 Data",
  x = 1.35, #ubicación del caption
  y = -0.1,
  xref = "paper",
  yref = "paper",
  showarrow = FALSE,
  font = list(size = 10) #tamaño del caption
)
)
)

#Mostramos el gráfico interactivo, para compilarlo en pdf lo comentamos, pues
#este tipo de archivo no permite los gráficos interactivos. Se muestra como
#tal en el html.
#final

```

A través de nuestro gráfico, podemos extraer una serie de conclusiones:

En el 2020, España y Reino Unido tuvieron algunos de sus picos más altos, es decir, un mayor número de muertes. En el caso de Reino Unido, esto se repitió en 2021.

Por otra parte la tendencia de muerte de los demás países es muy similar entre todos estos, el 2021 es donde la mayoría se encuentran más afectados. En concreto, Japón despunta en el 2023 con otro pico considerablemente alto de muertes, mientras que los otros países siguen una tendencia de descenso, en cuanto a muertes por COVID se refiere. La cual, se mantiene en la mayoría de los casos en este 2024.