

GraficoTema3

Tratamiento de Datos, Grado en Ciencia de Datos - UV

Iker Cuadros Ruiz, Guillem García Garro, Carles Pascual i Sivera

2024-02-19

Contents

0. Comprobación e instalación de paquetes necesarios.	1
1. Importación de datos.	1
2. Filtrado de Datos	2
3. Creación del Gráfico	2

0. Comprobación e instalación de paquetes necesarios.

Preparación del entorno de trabajo asegurando que todas las librerías necesarias están instaladas y cargadas para el análisis y la visualización de los próximos datos.

1. Importación de datos.

Lectura y preparación de los datos sobre COVID-19 para su posterior análisis y visualización.

Se puede observar que se especifica el formato de la variable de fecha para que sea reconocido correctamente como fecha con un formato específico.

Posteriormente, convierte la variable de ubicación en una variable categórica, lo que facilitará ciertas operaciones y visualizaciones con los datos.

```
## Esta sentencia recoge los datos directamente desde la dirección web
# df.covid <- read_csv("https://covid.ourworldindata.org/data/owid-covid-data.csv",
#   col_types = cols(date = col_date(format = "%Y-%m-%d")),
#   na = "NA")

df.covid <- read_csv("data/owid-covid-data.csv",
  col_types = cols(date = col_date(format = "%Y-%m-%d")),
  na = "NA")
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
df.covid$location <- as.factor(df.covid$location)
```

2. Filtrado de Datos

2.1. Definición de Países y Colores

Preparación de los datos para una representación visual donde se comparan diferentes países utilizando una paleta de colores distintivos.

El código recopila datos sobre varios países y los organiza alfabéticamente. Luego, asigna colores aleatorios a cada país en la lista, excepto a España, que se define específicamente como “rojo”.

Por último, ajusta la lista de colores para asegurarse de que España tenga el color “rojo” en función de su posición en la lista ordenada de países.

```
# Se cogen los datos de los países a analizar
countries <- c("Spain", "United States", "United Kingdom", "Japan", "France", "Italy", "Germany", "Canada")
countries <- sort(countries, decreasing = FALSE)

# Definición de la lista de colores en función de la longitud de la lista
colores <- c()
for ( i in 1:(length(countries)-1) ) {
  colores <- c(colores, sample(colors(1), 1))
}

# Buscar la posición en la que se encuentra España, para asignarle
# el color rojo en la lista de valores
posEsp <- match("Spain", countries)

colores <- c(colores[1:posEsp-1], "red", colores[posEsp:length(colores)])
```

2.2. Filtrado de datos

Se realiza una selección de la información específica sobre COVID-19 del conjunto de datos importado.

En primer lugar, se eligen los datos de los países que están en la lista creada anteriormente (*countries*), donde aparecen los países cuyos datos se quieren mostrar. Adicionalmente, los datos tomados serán aquellos tomados a partir del 13 de febrero de 2020.

Luego, se extraen los datos solo para España desde el conjunto de datos filtrado anteriormente para poder ajustarle el color en el siguiente apartado.

```
df.covid.filtered <- df.covid %>%
  subset(location %in% countries
         & new_deaths > 0
         & date >= as.Date("2020-02-13"))

df.covid.filt.spain <- df.covid.filtered %>%
  subset(location == "Spain")
```

3. Creación del Gráfico

Creación de un gráfico haciendo uso de los datos filtrados de COVID-19, donde muestra la evolución de nuevas muertes suavizadas por millón de personas a lo largo del tiempo.

Además, se produce un resaltado sobre la línea de datos de España en rojo y personaliza las etiquetas del gráfico.

Finalmente, guarda el gráfico como un archivo PDF y genera una versión interactiva utilizando plotly.

```
# Creación del grafico
grafico <- ggplot(df.covid.filtered, aes(x = date, y = new_deaths_smoothed_per_million)) +
  geom_line(aes(col= location),
            linewidth = ifelse(df.covid.filtered$location == "Spain",1,0.5),
            alpha = ifelse(df.covid.filtered$location == "Spain", 1, 0.6)) +
  geom_line(data = df.covid.filt.spain, col = "red") +
  scale_color_manual(values = colores) +
  scale_x_date(date_labels = ("%b %d,%Y"), breaks = "9 month") +
  scale_y_continuous(limits = c(0,21), breaks = seq(0, 18, by = 2))

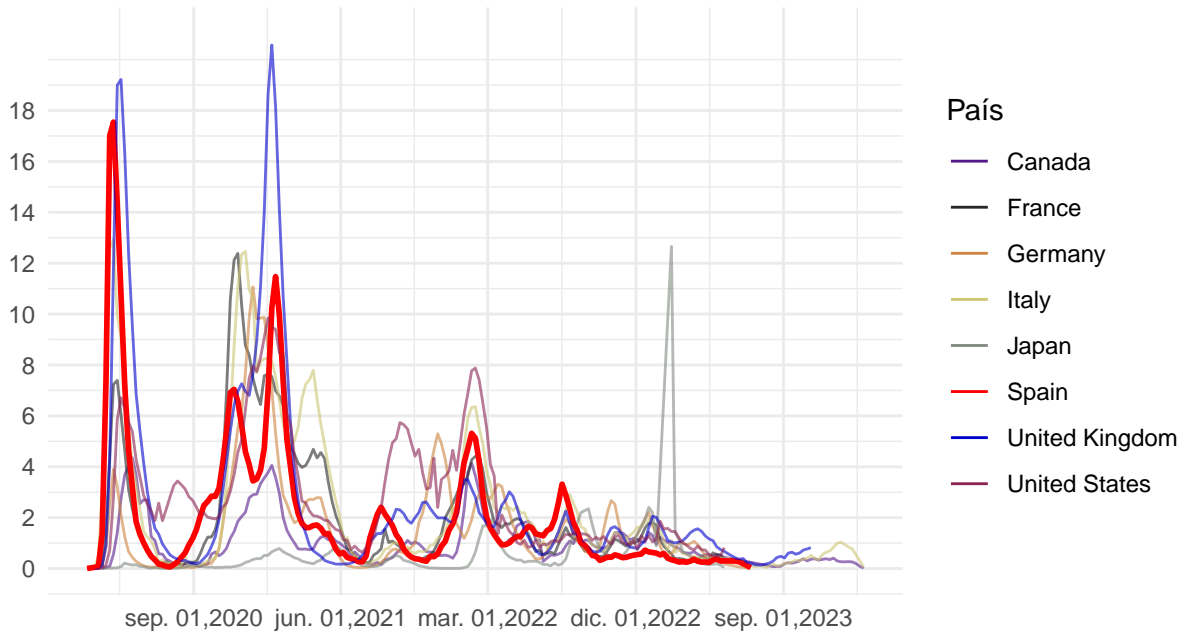
# Adición de las etiquetas necesatias
grafico <- grafico +
  labs(title = "Grupo N: Daily new confirmed COVID-19 deaths per million people",
        subtitle = "7-day rolling average.\nFor some countries the number of confirmed deaths is much low",
        x = "",
        y = "",
        col = "País",
        alt = "Deaths evolution graphic",
        caption = "Source: Johns Hopkins University CSSE COVID-19 Data") +
  theme_minimal()

# Cambio del tipo de letra de las etiquetas
grafico <- grafico +
  theme(text = element_text(family = "sans"),
        plot.title = element_text(size = 12, face = "bold"),
        plot.subtitle = element_text(size = 10),
        plot.caption = element_text(size = 8, face = "italic", hjust = 0))
grafico
```

Grupo N: Daily new confirmed COVID-19 deaths per million people

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

```
ggsave(filename = "./figures/prueba.pdf", plot = grafico, width = 8, height = 6)
```