Covid Casos

Ing. José Miguel Pereira

23/1/2022

# Casos Covid

En el presente proyecto se realiza la visualización de los casos de COVID-19 y sus variante por paises.

La data contiene la siguiente información:

* Localización (location): País.
* Fecha (date): Fecha a la entrada de la información.
* Variante (variant) : Variante correspondiente.
* Numero de secuencia (num\_sequences): número de secuencia correspondiente (por país, variante y fecha).
* Porcentaje de secuencia (perc\_sequences): Porcentaje de secuencias del número total de secuencias (para el país, la variante y fecha).
* Número de secuencias totales (numsequencestotal): Número total de secuencias (para el país, la variante y la fecha).

Cita:

Yam Peleg, “Omicron daily cases by country (COVID-19 variant).” Kaggle, 2022, doi: 10.34740/KAGGLE/DSV/3085368.

# Tabla de contenido

1. Librerías
2. Data
3. Manipulación de datos y resumen
4. Casos de variantes por país
5. Casos por fecha
6. Percepción de las variables de COVID-19 en los paises.
7. Variable Omicron

# 1. Librerías

Se procede a importar las librerías a emplear para el procesamiento y visualización de los datos.

# Librerias   
library(dplyr)  
library(lubridate)  
library(reshape2)  
library(openxlsx)  
# Gráficos  
library(ggplot2)  
library(gganimate)  
library(plotly)  
# Gráficos de correlación  
library(corrplot)  
# Resumen personalizado  
library(gtsummary)  
library(DataExplorer)  
# cluster result visualization with network diagram  
library(igraph)  
library(vioplot)  
library(treemapify)

# 2. Data

Con las librerias cargadas, se procede a cargar la base de datos.

location date variant num\_sequences perc\_sequences  
1 Angola 2020-07-06 Alpha 0 0  
2 Angola 2020-07-06 B.1.1.277 0 0  
3 Angola 2020-07-06 B.1.1.302 0 0  
4 Angola 2020-07-06 B.1.1.519 0 0  
5 Angola 2020-07-06 B.1.160 0 0  
6 Angola 2020-07-06 B.1.177 0 0  
 num\_sequences\_total  
1 3  
2 3  
3 3  
4 3  
5 3  
6 3

# 3. Manipulación de datos y resumen

Vamos a observar si existen datos nulos

location date variant num\_sequences   
 0 0 0 0   
 perc\_sequences num\_sequences\_total   
 0 0

[1] 0

Se observa que no existen datos faltantes.

'data.frame': 100416 obs. of 6 variables:  
 $ location : chr "Angola" "Angola" "Angola" "Angola" ...  
 $ date : chr "2020-07-06" "2020-07-06" "2020-07-06" "2020-07-06" ...  
 $ variant : chr "Alpha" "B.1.1.277" "B.1.1.302" "B.1.1.519" ...  
 $ num\_sequences : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ perc\_sequences : num 0 0 0 0 0 0 0 0 0 0 ...  
 $ num\_sequences\_total: int 3 3 3 3 3 3 3 3 3 3 ...

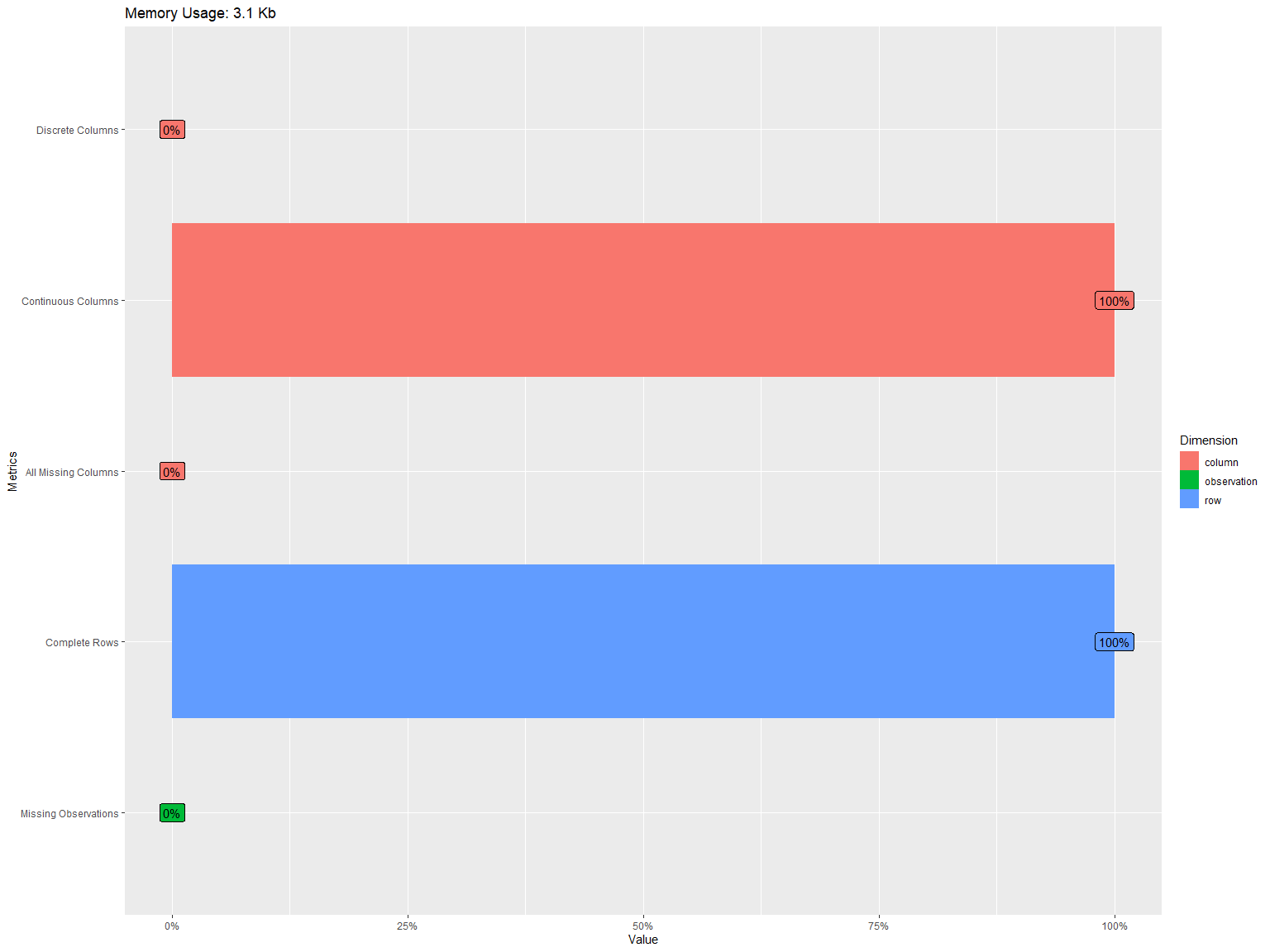
Se observa que la fecha no se encuentra en formato. Toca cambiarla.

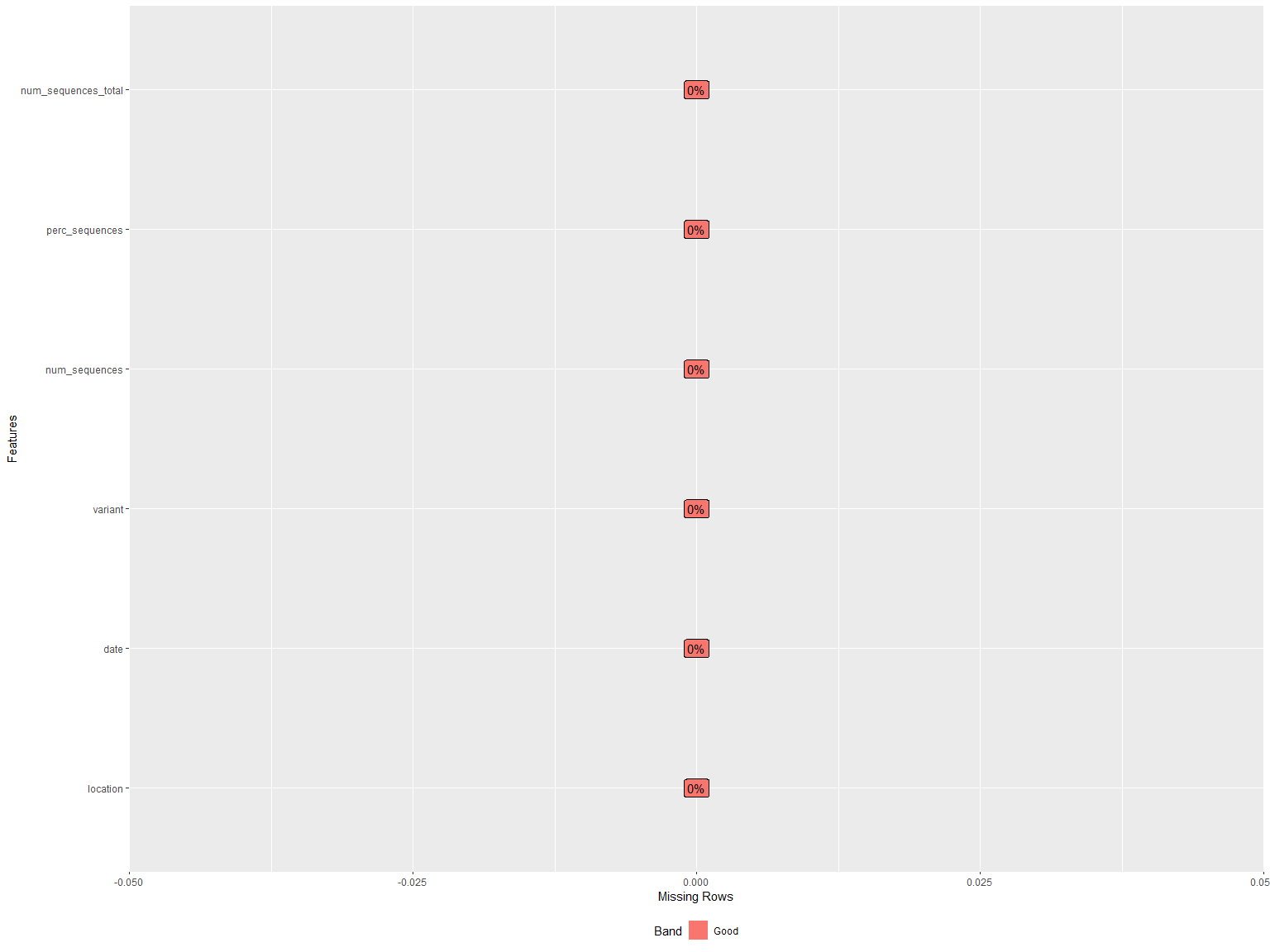
'data.frame': 100416 obs. of 6 variables:  
 $ location : chr "Angola" "Angola" "Angola" "Angola" ...  
 $ date : Date, format: "2020-07-06" "2020-07-06" ...  
 $ variant : chr "Alpha" "B.1.1.277" "B.1.1.302" "B.1.1.519" ...  
 $ num\_sequences : int 0 0 0 0 0 0 0 0 0 0 ...  
 $ perc\_sequences : num 0 0 0 0 0 0 0 0 0 0 ...  
 $ num\_sequences\_total: int 3 3 3 3 3 3 3 3 3 3 ...

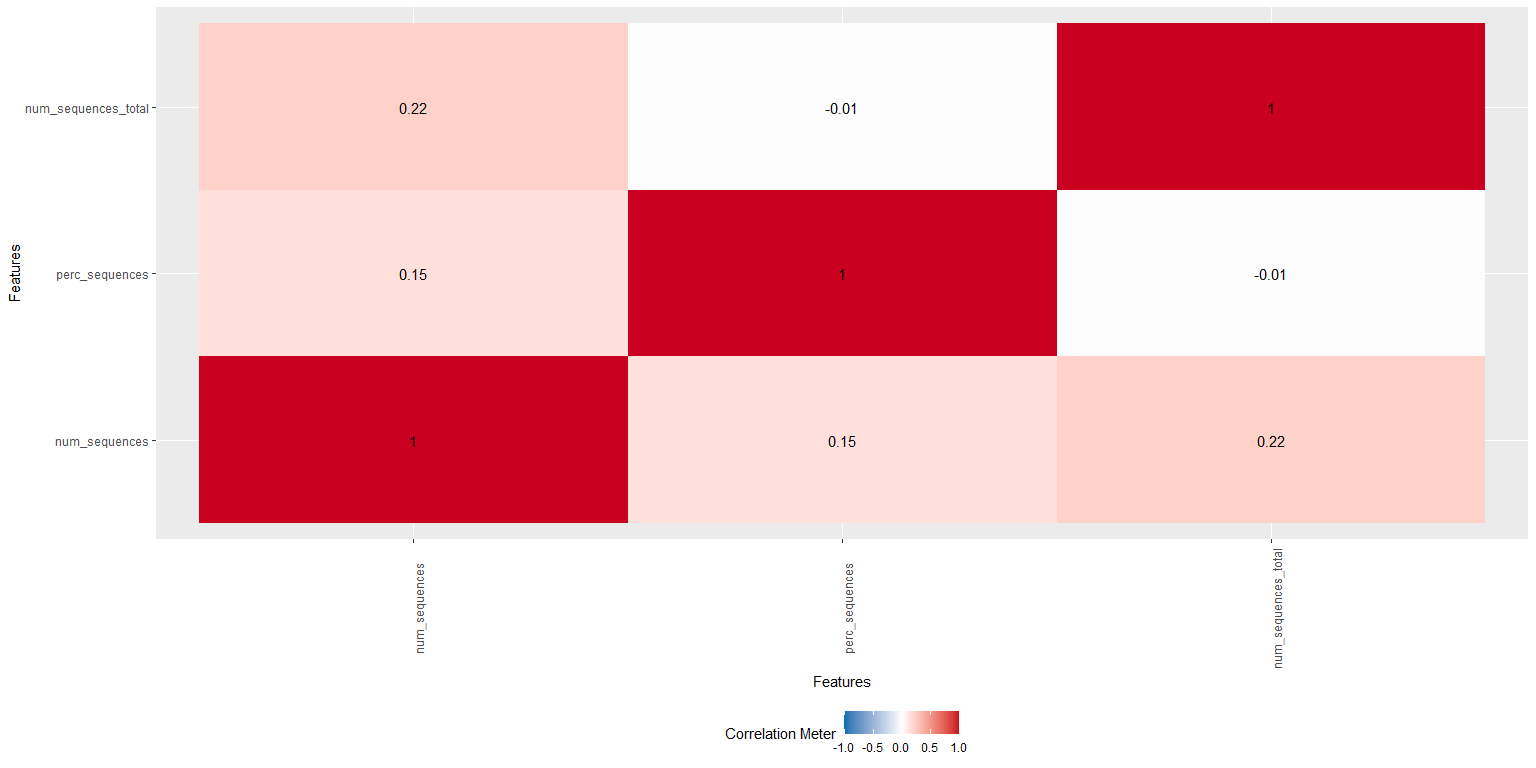
Los datos se encuentran correctos.

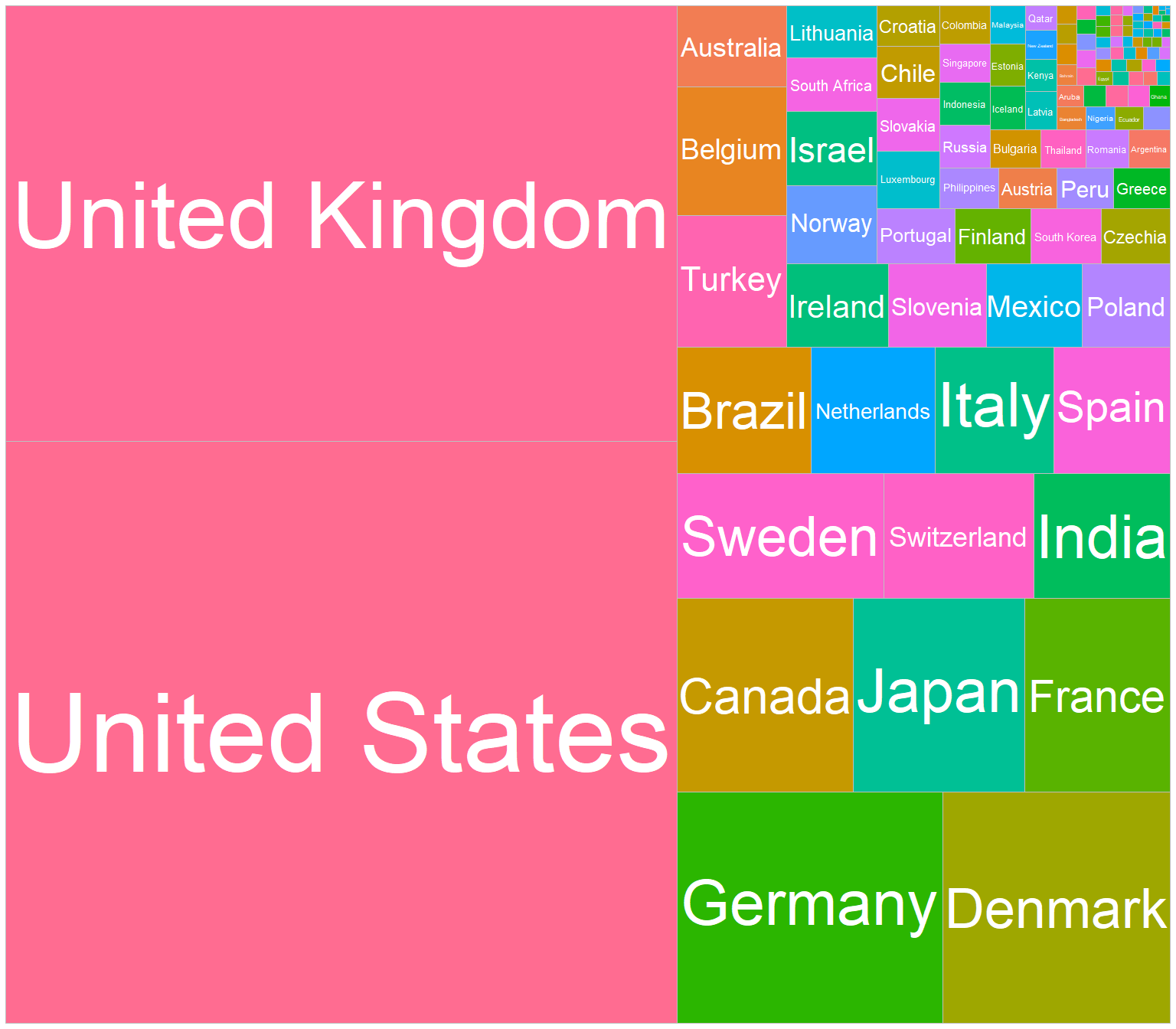
Vamos a observar cuantas observaciones de variantes se poseen.

variant n  
1 Alpha 4184  
2 B.1.1.277 4184  
3 B.1.1.302 4184  
4 B.1.1.519 4184  
5 B.1.160 4184  
6 B.1.177 4184  
7 B.1.221 4184  
8 B.1.258 4184  
9 B.1.367 4184  
10 B.1.620 4184  
11 Beta 4184  
12 Delta 4184  
13 Epsilon 4184  
14 Eta 4184  
15 Gamma 4184  
16 Iota 4184  
17 Kappa 4184  
18 Lambda 4184  
19 Mu 4184  
20 non\_who 4184  
21 Omicron 4184  
22 others 4184  
23 S:677H.Robin1 4184  
24 S:677P.Pelican 4184









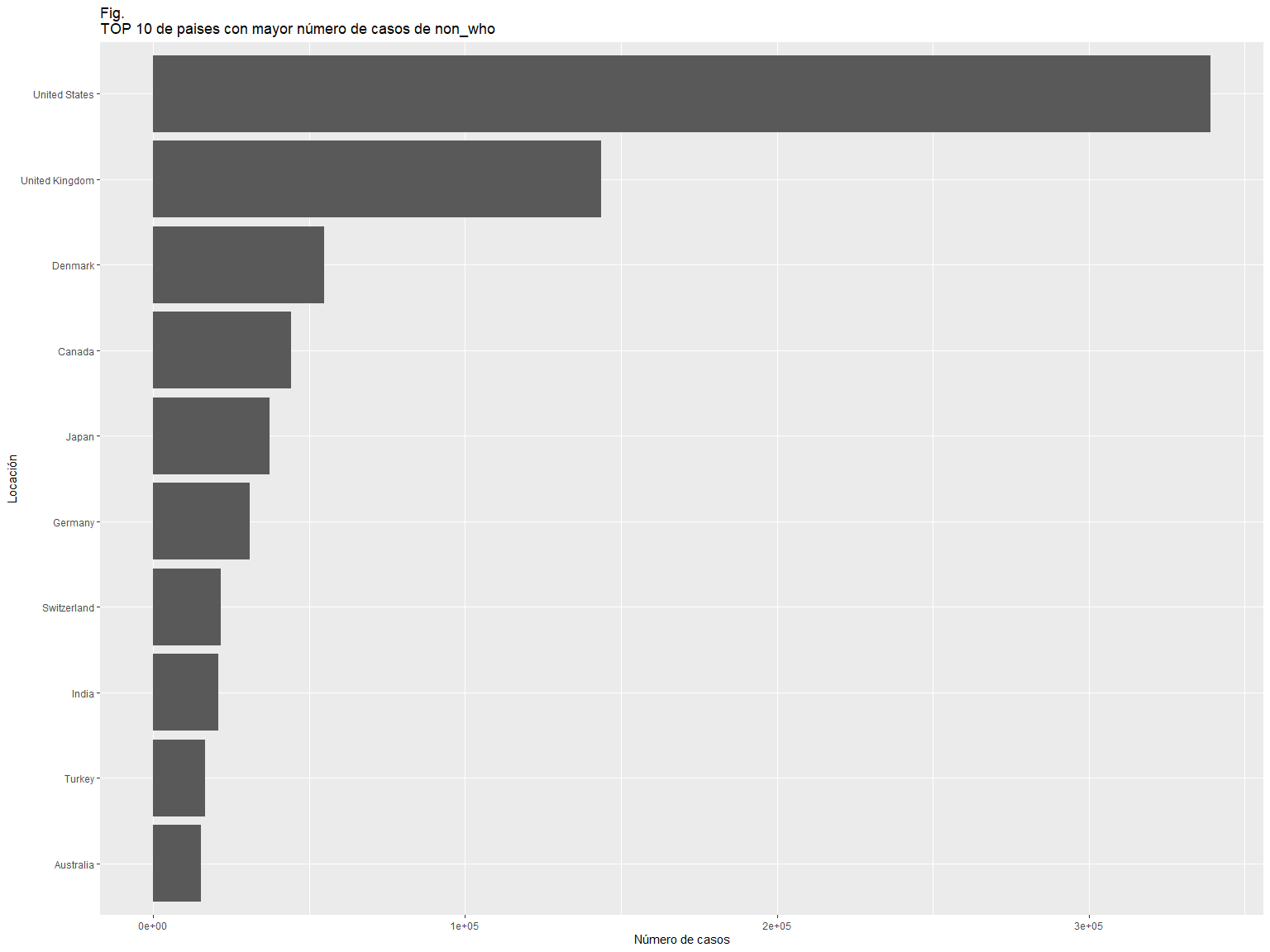
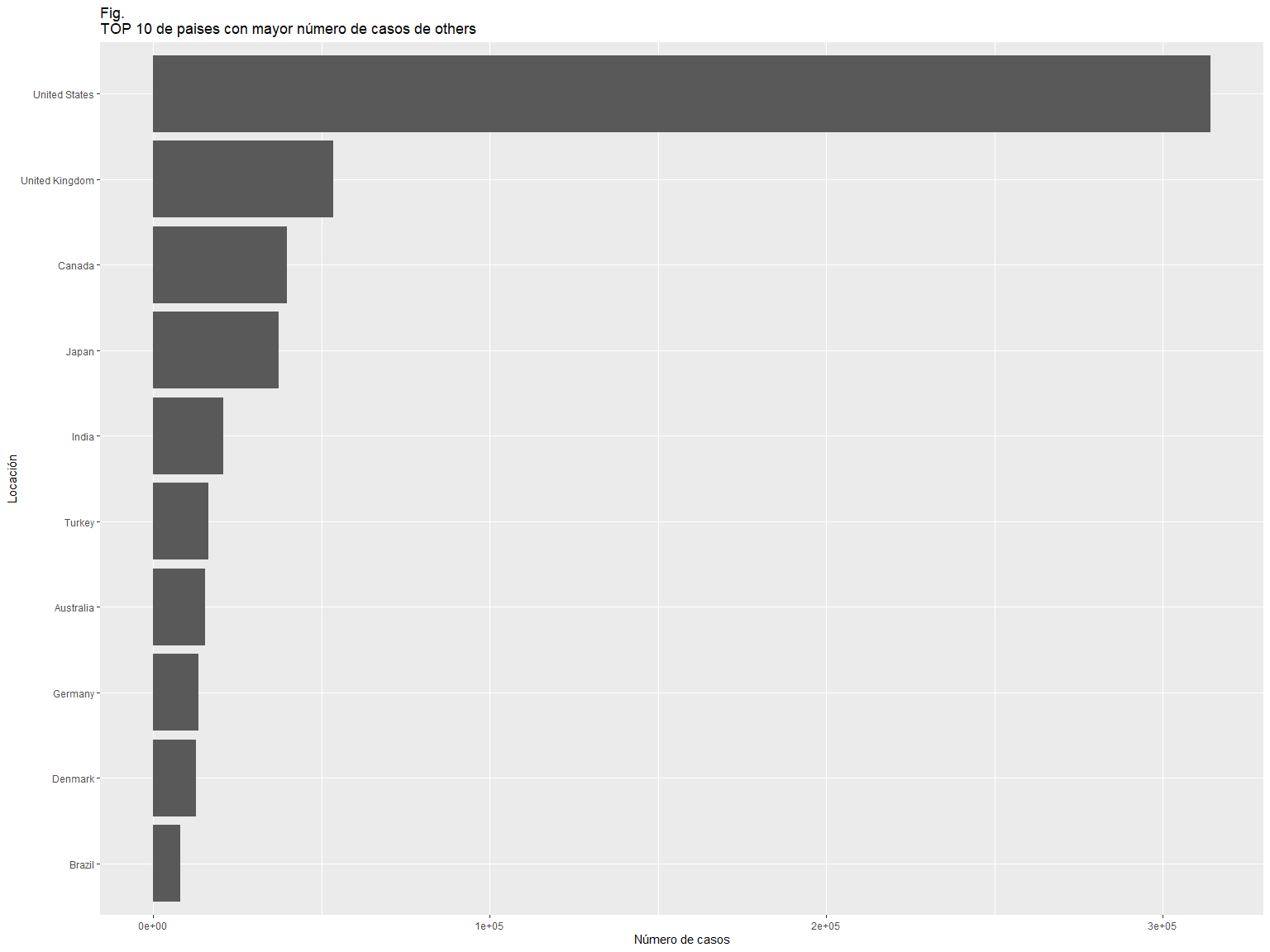
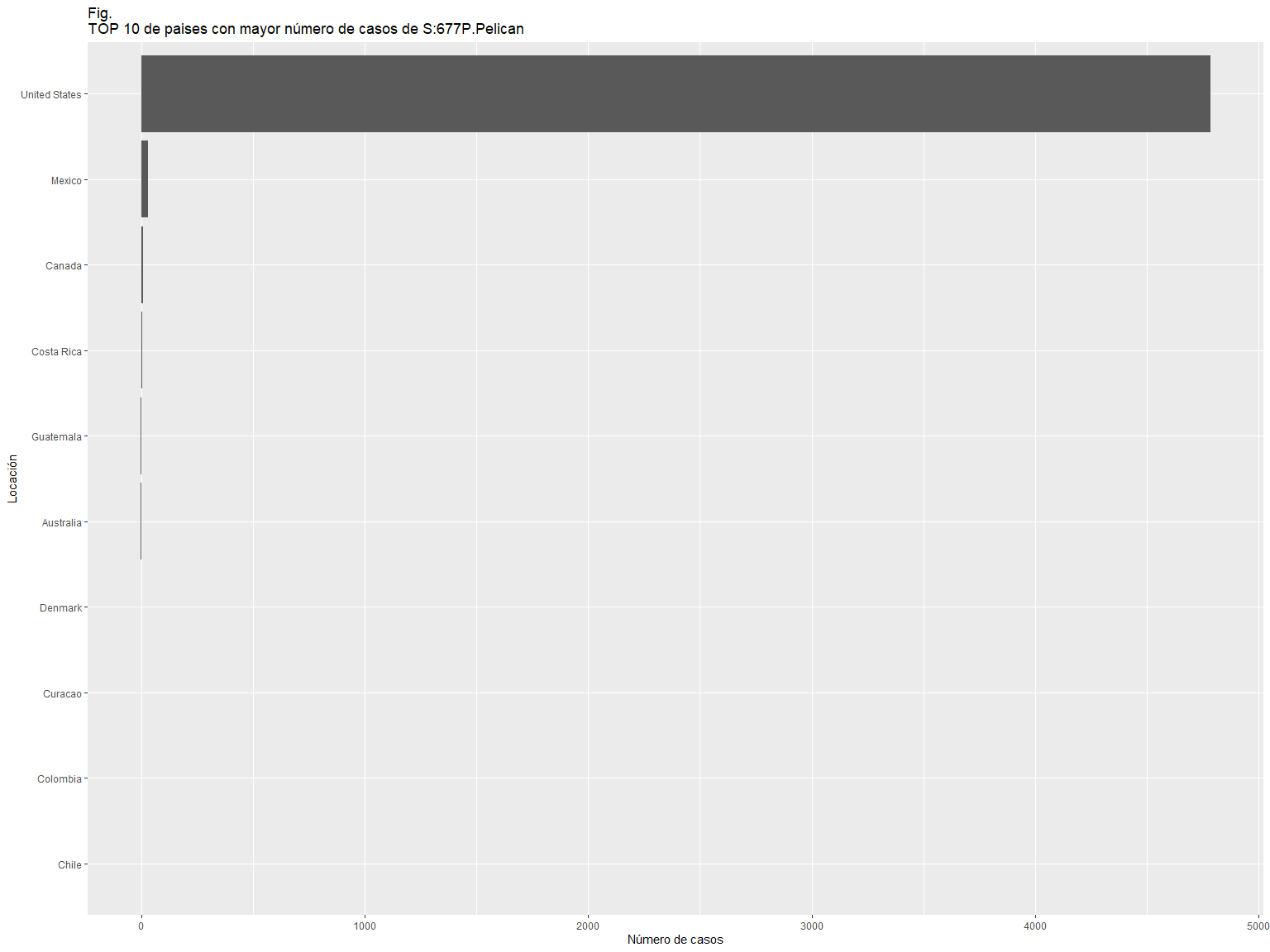
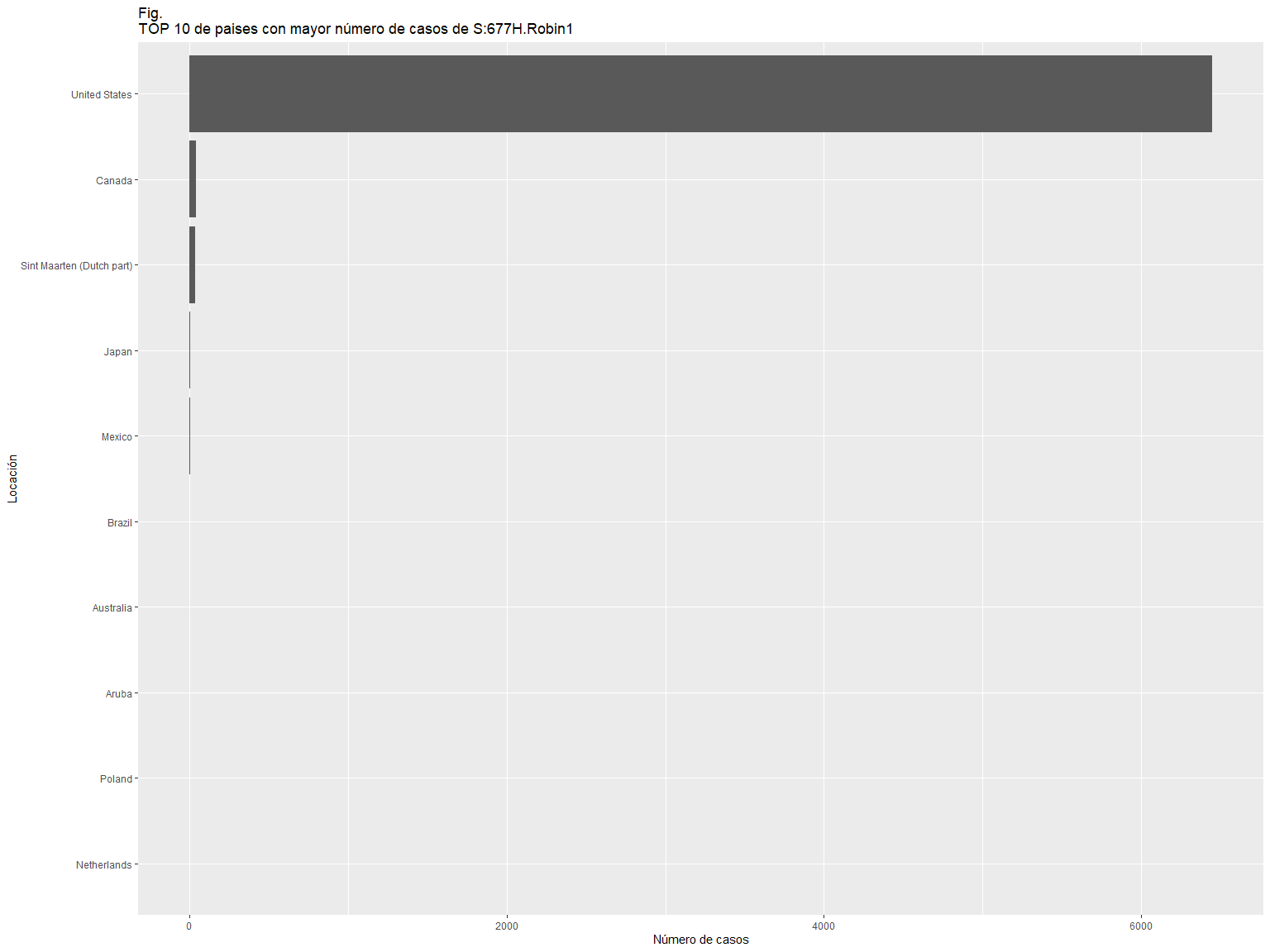
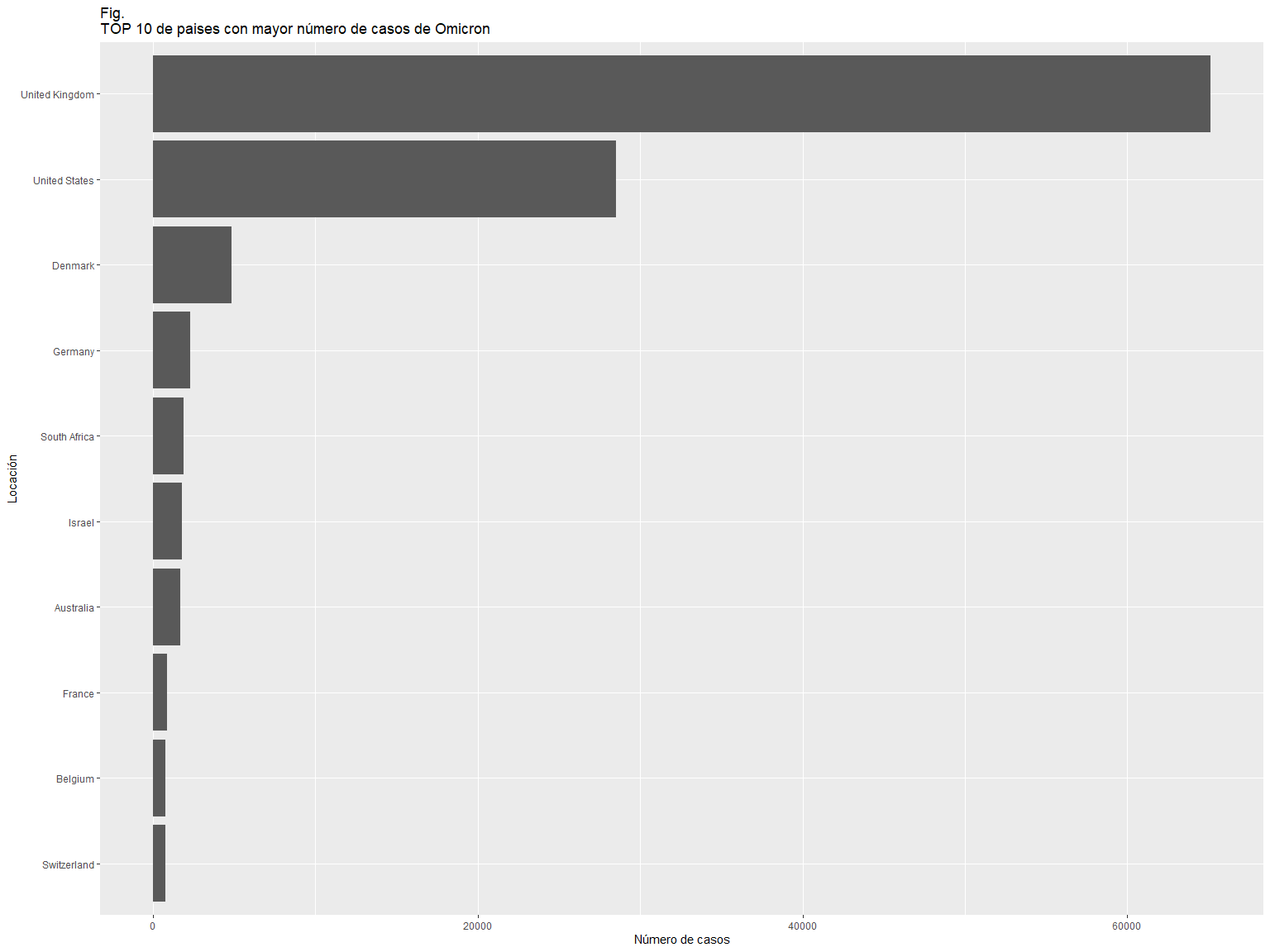
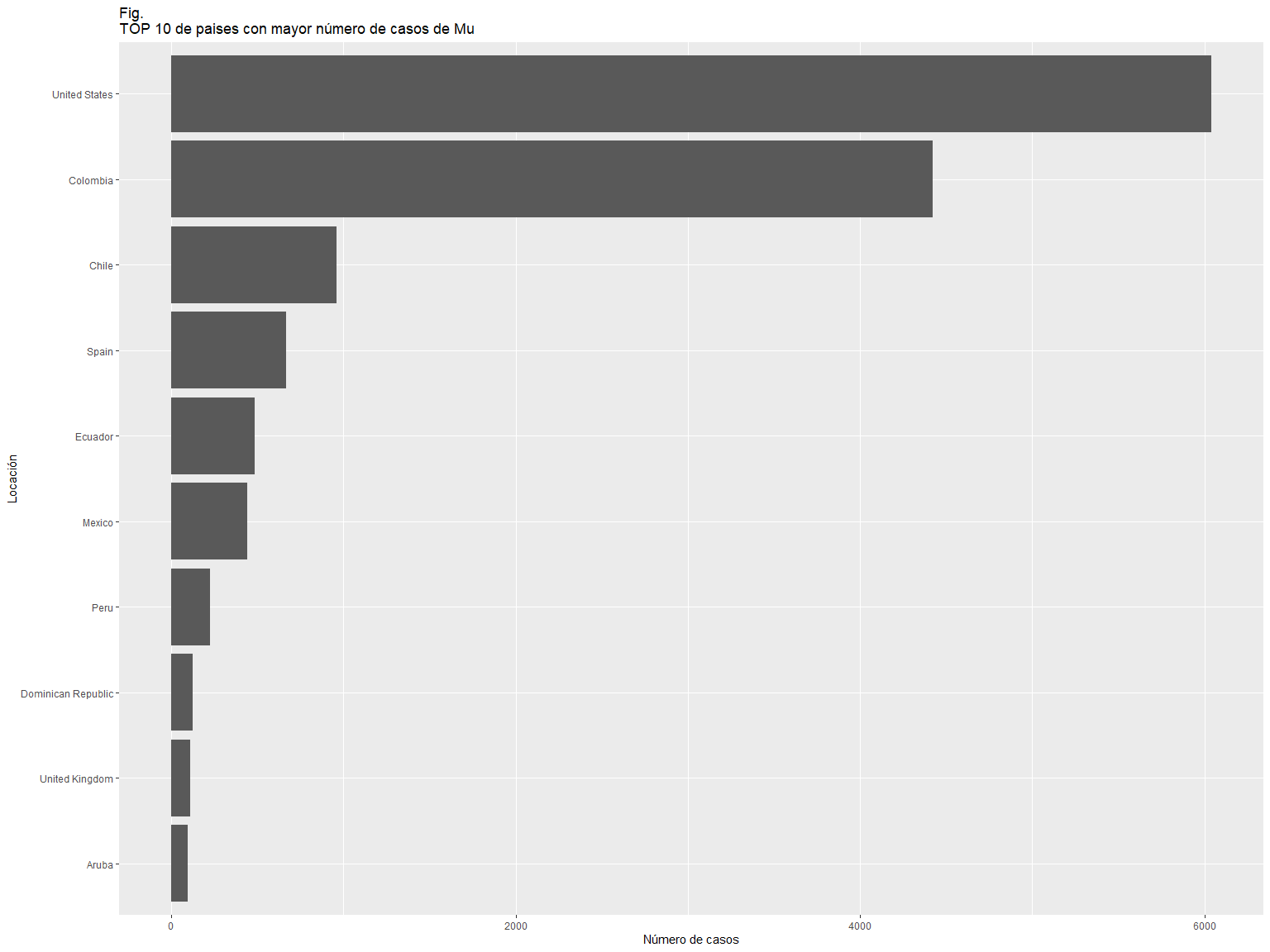
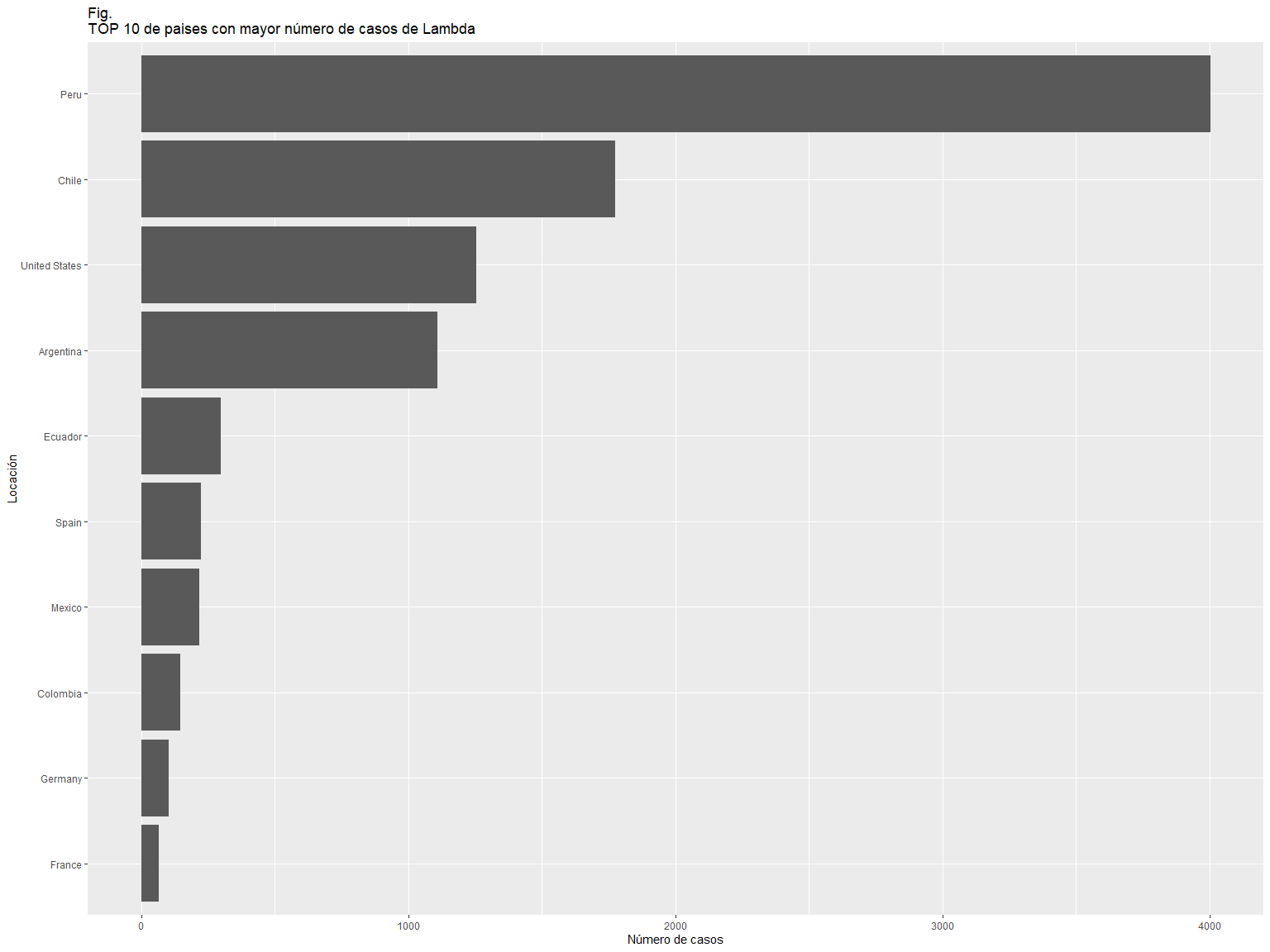
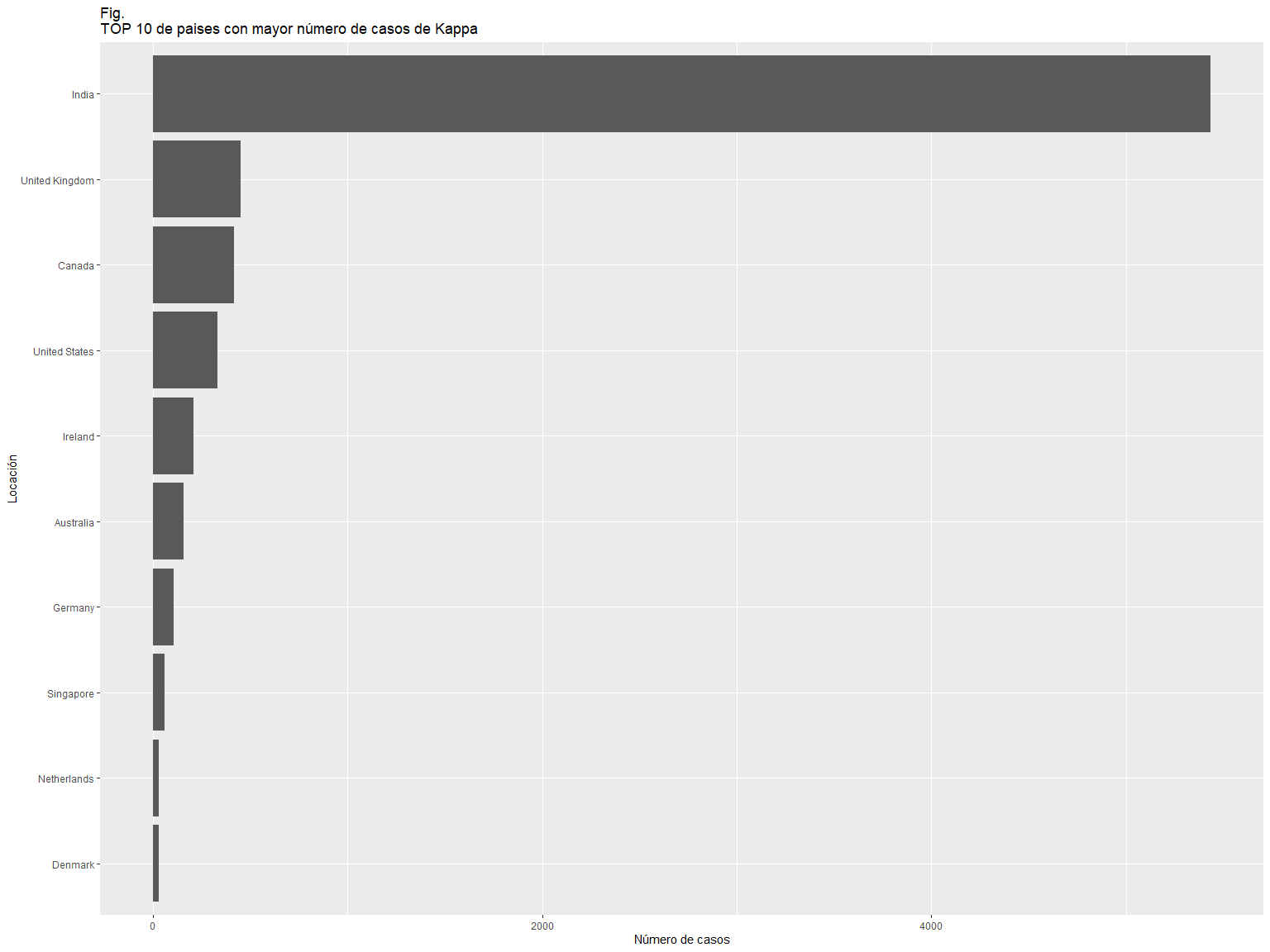
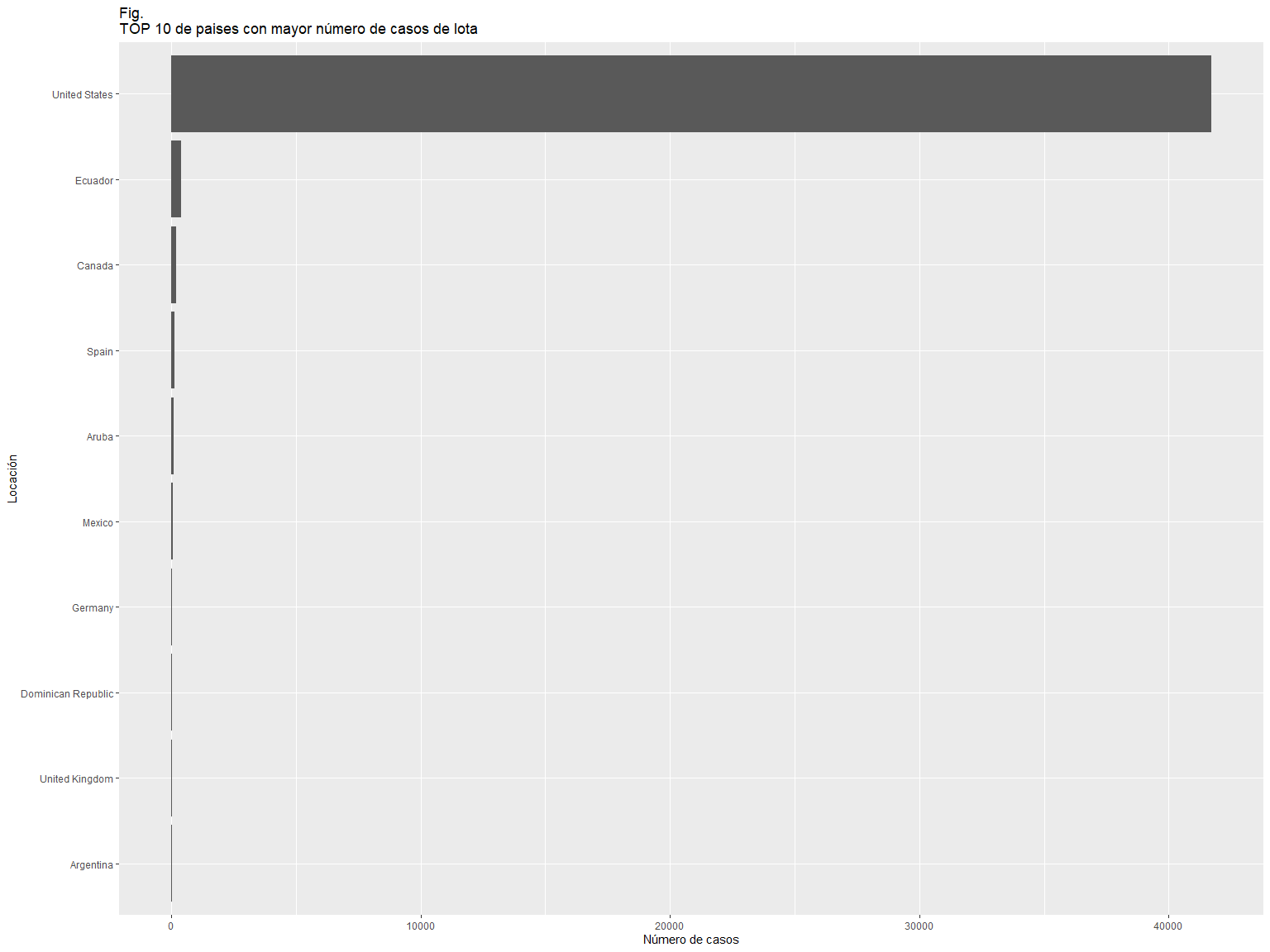
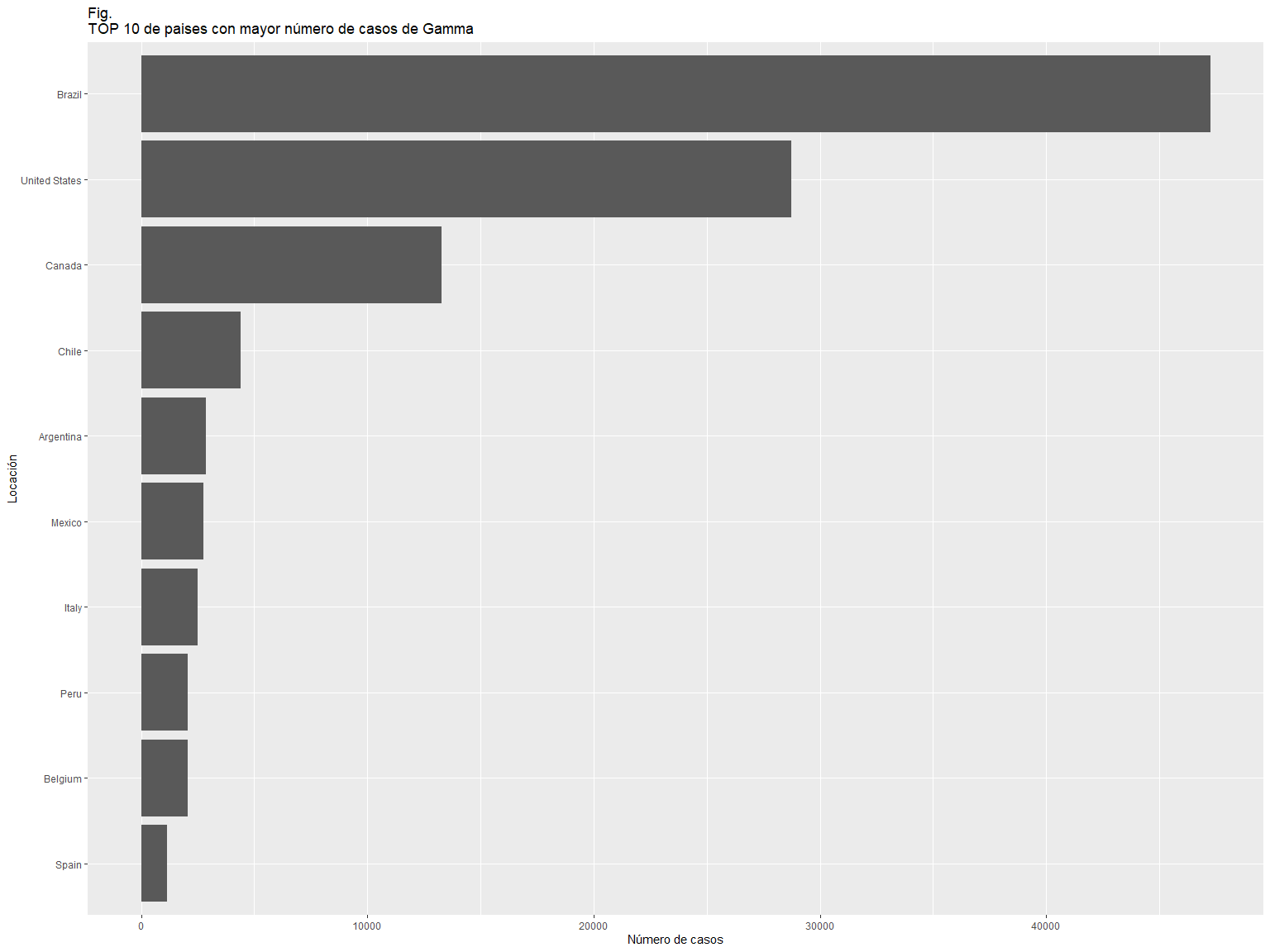
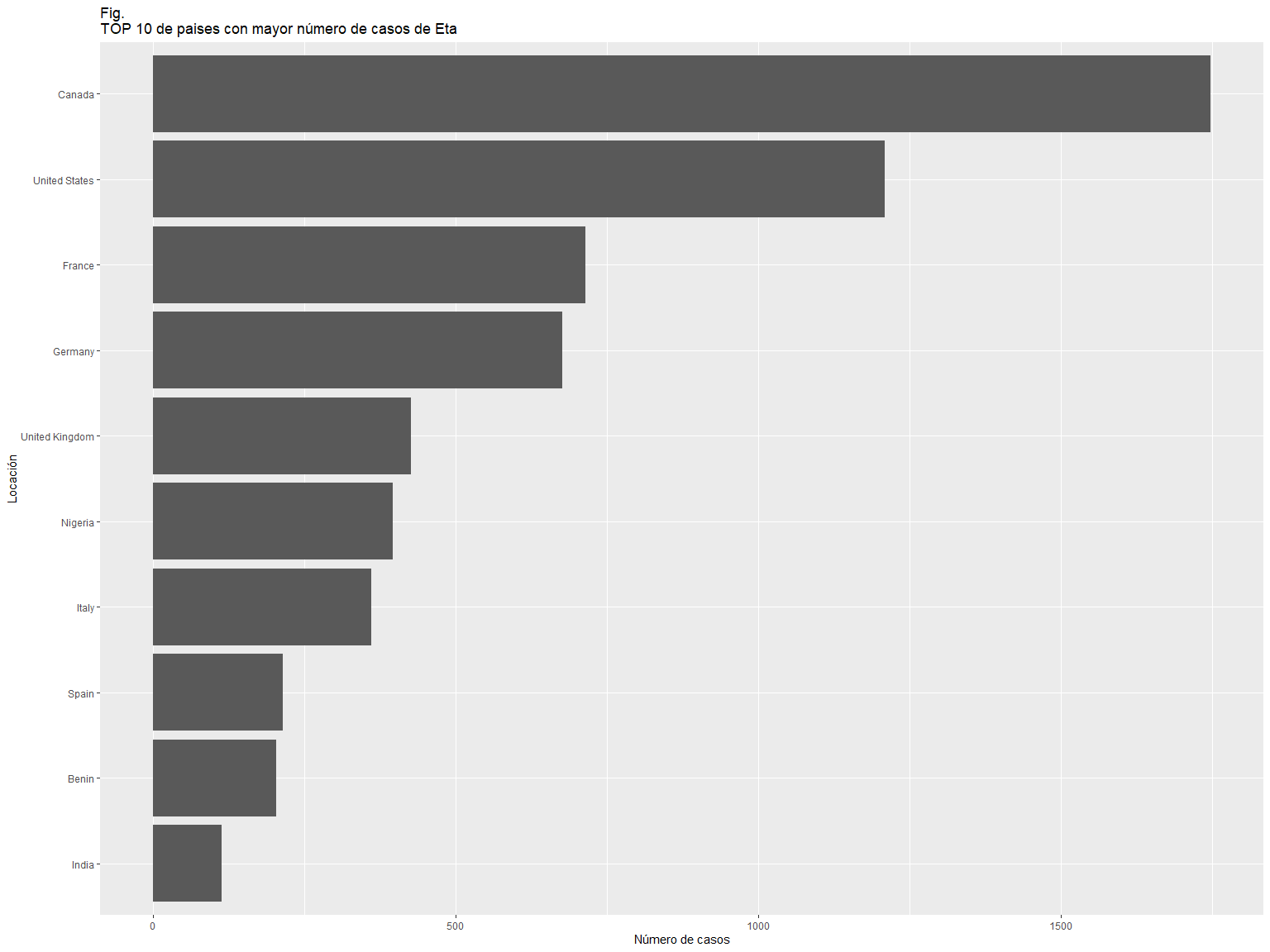
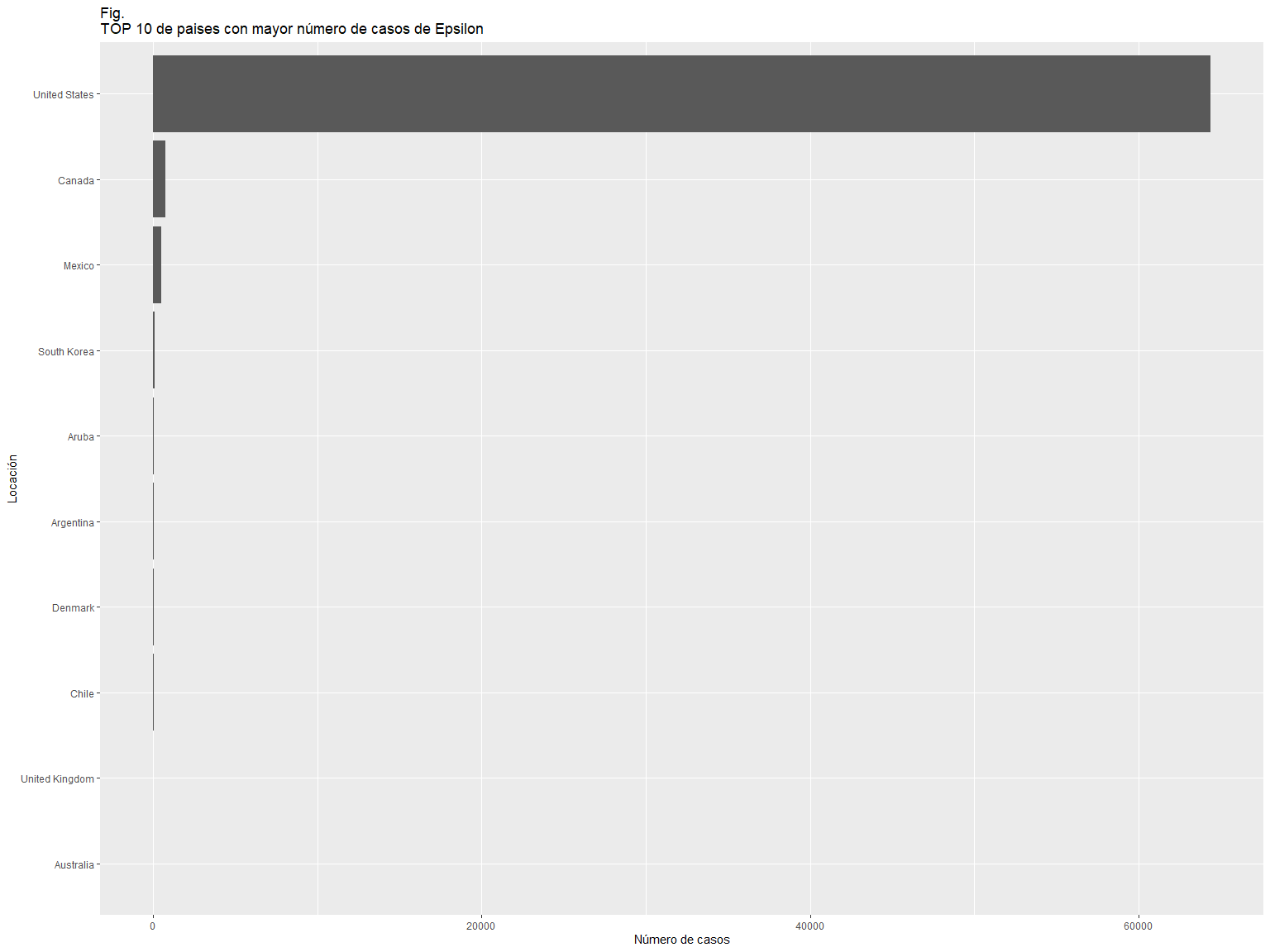
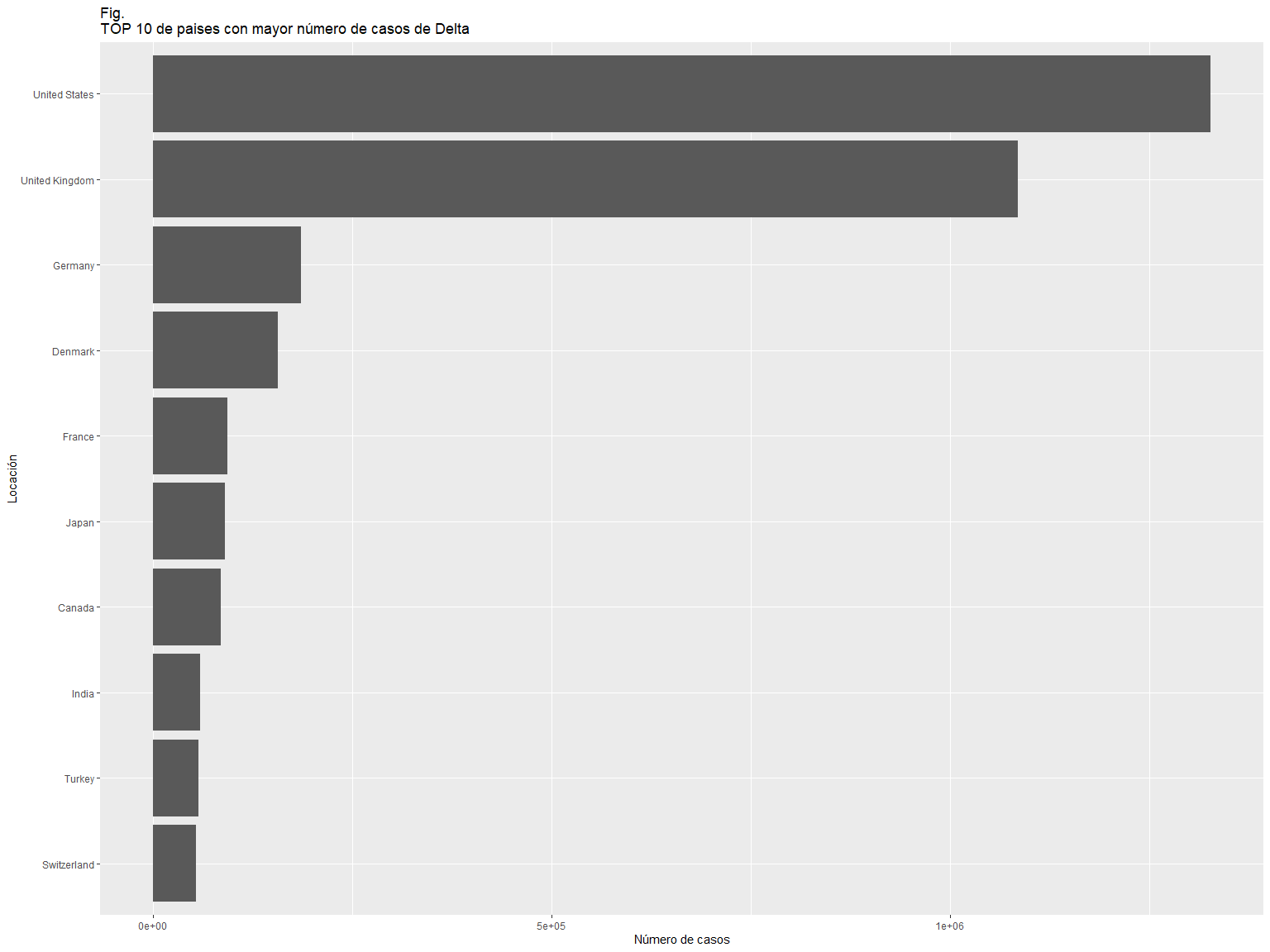
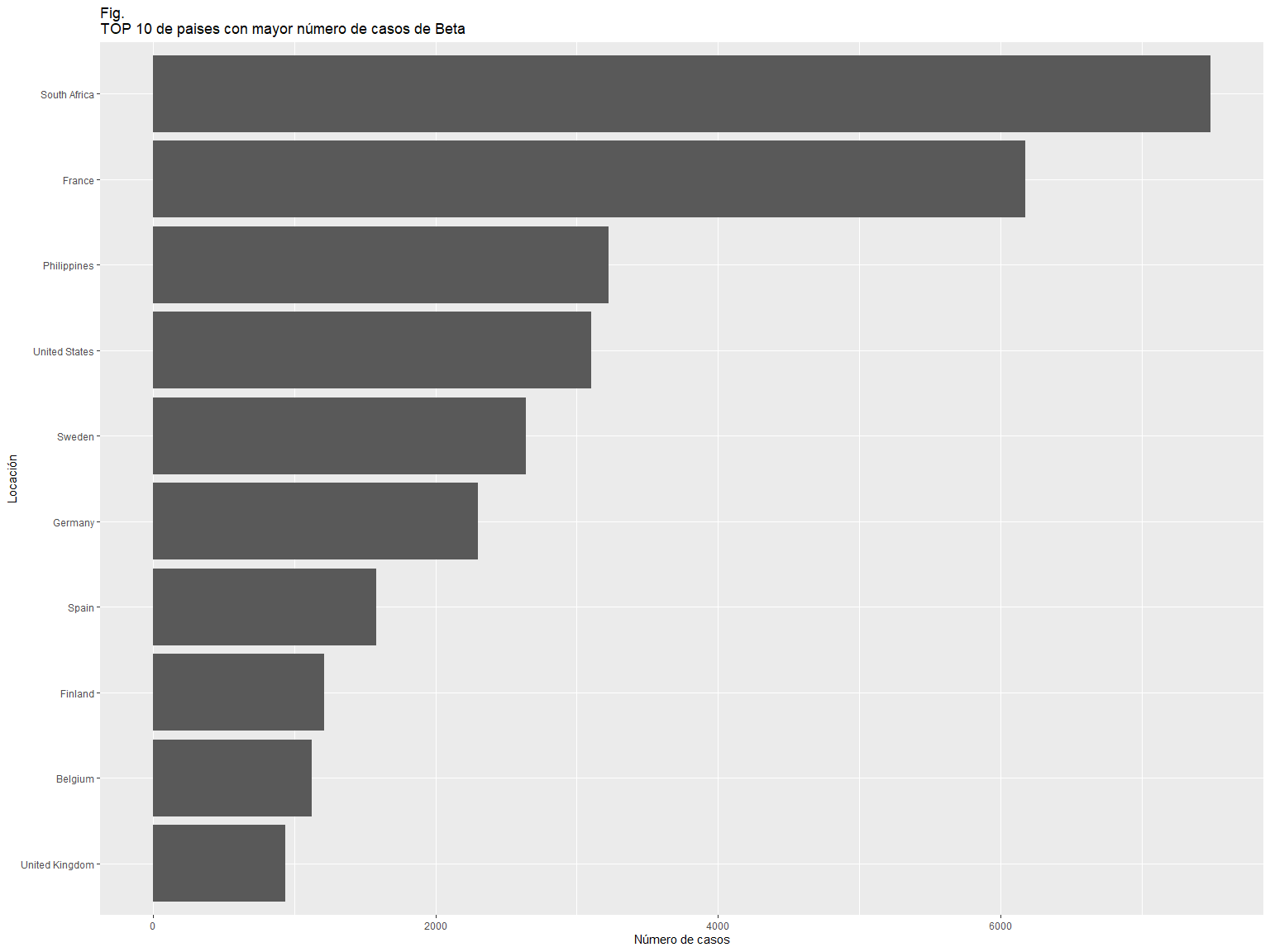
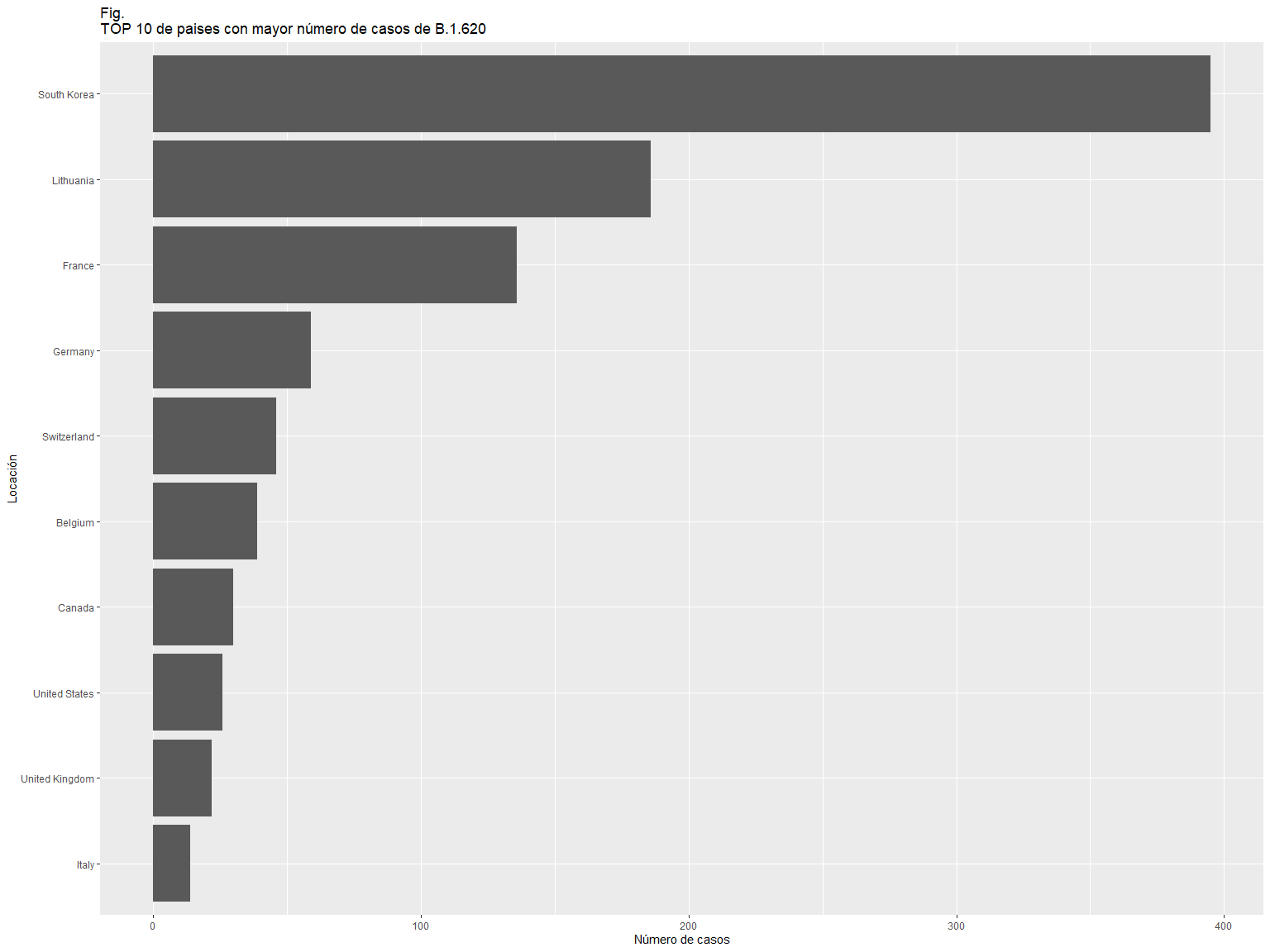
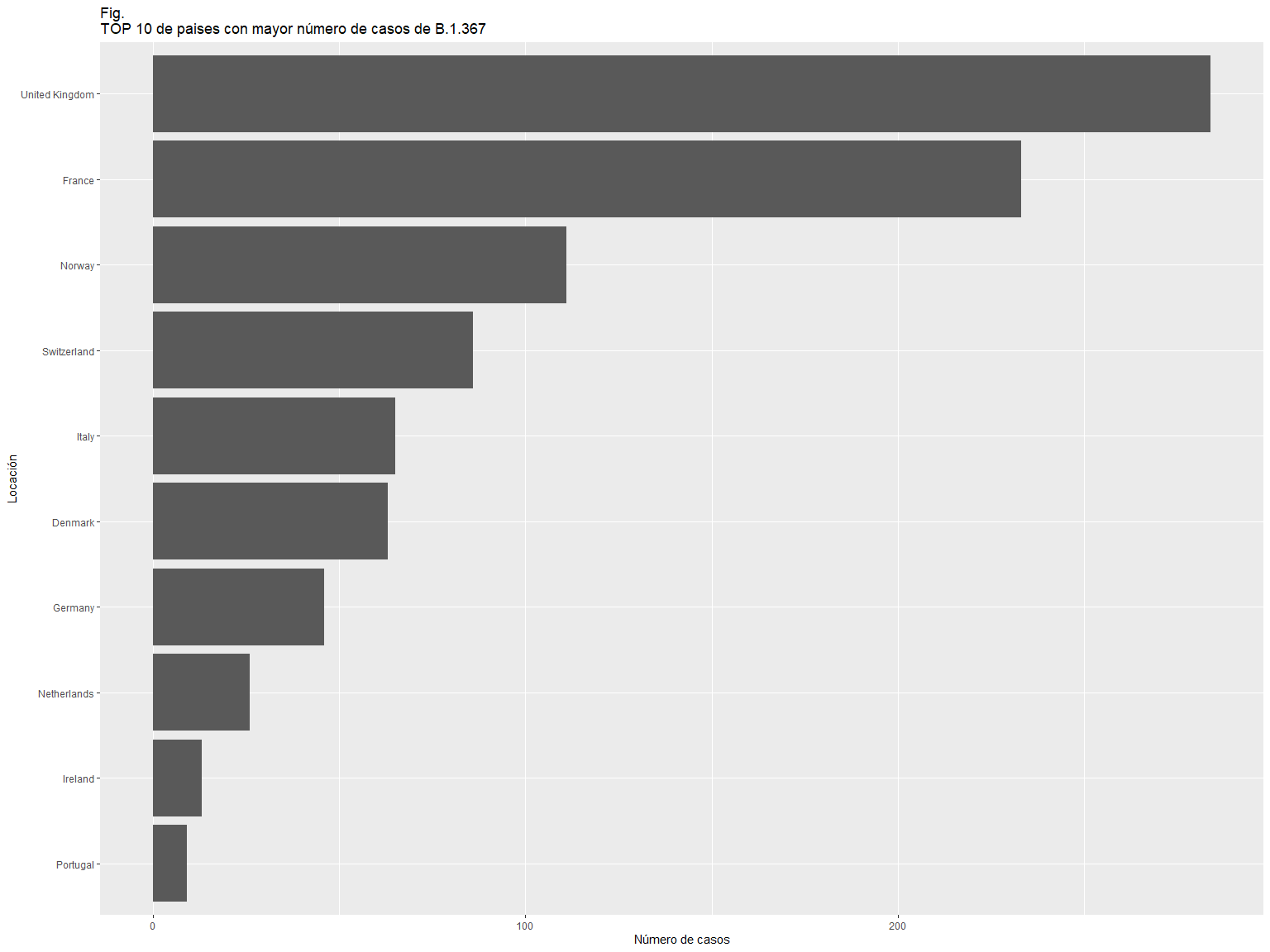
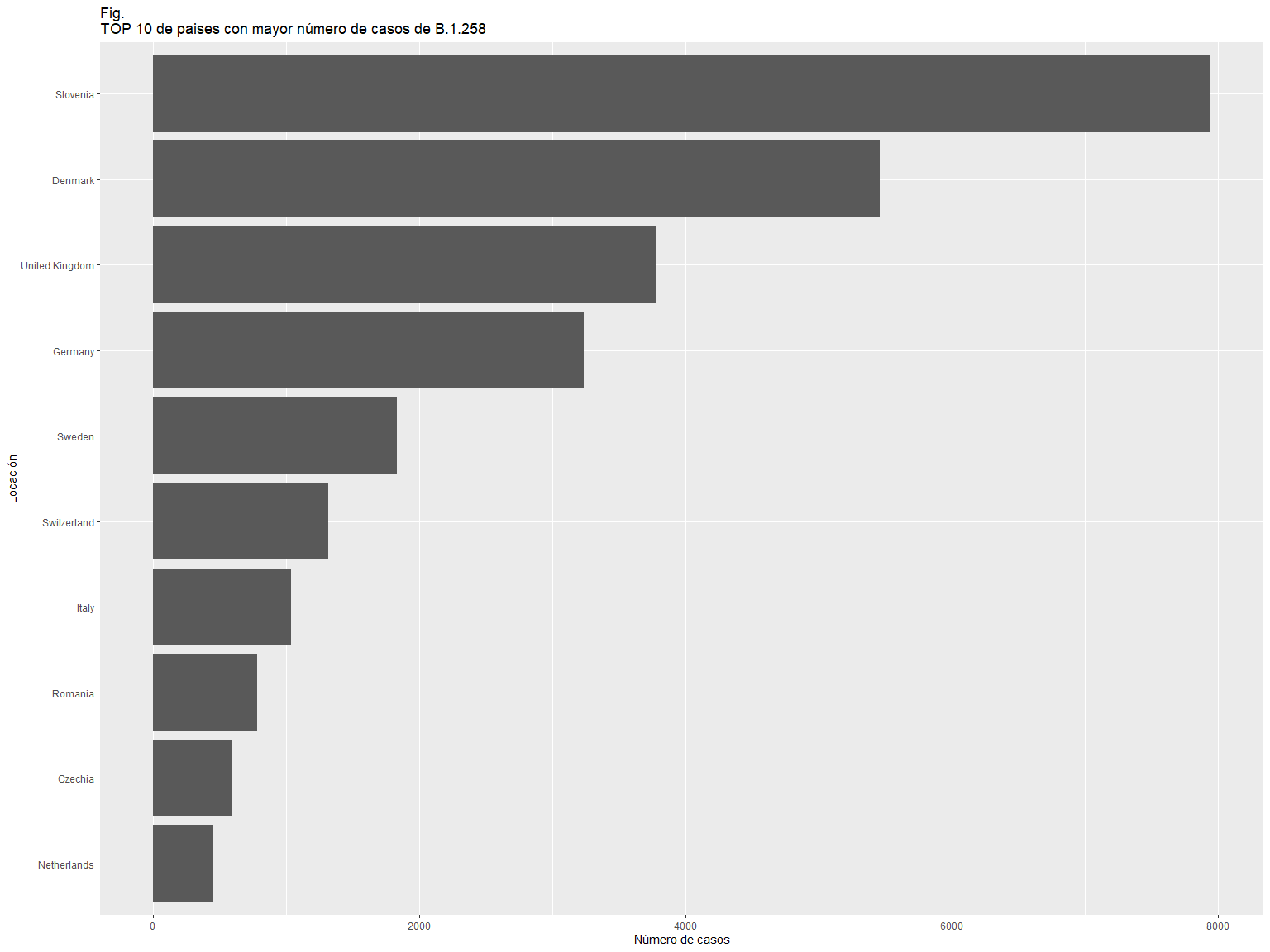
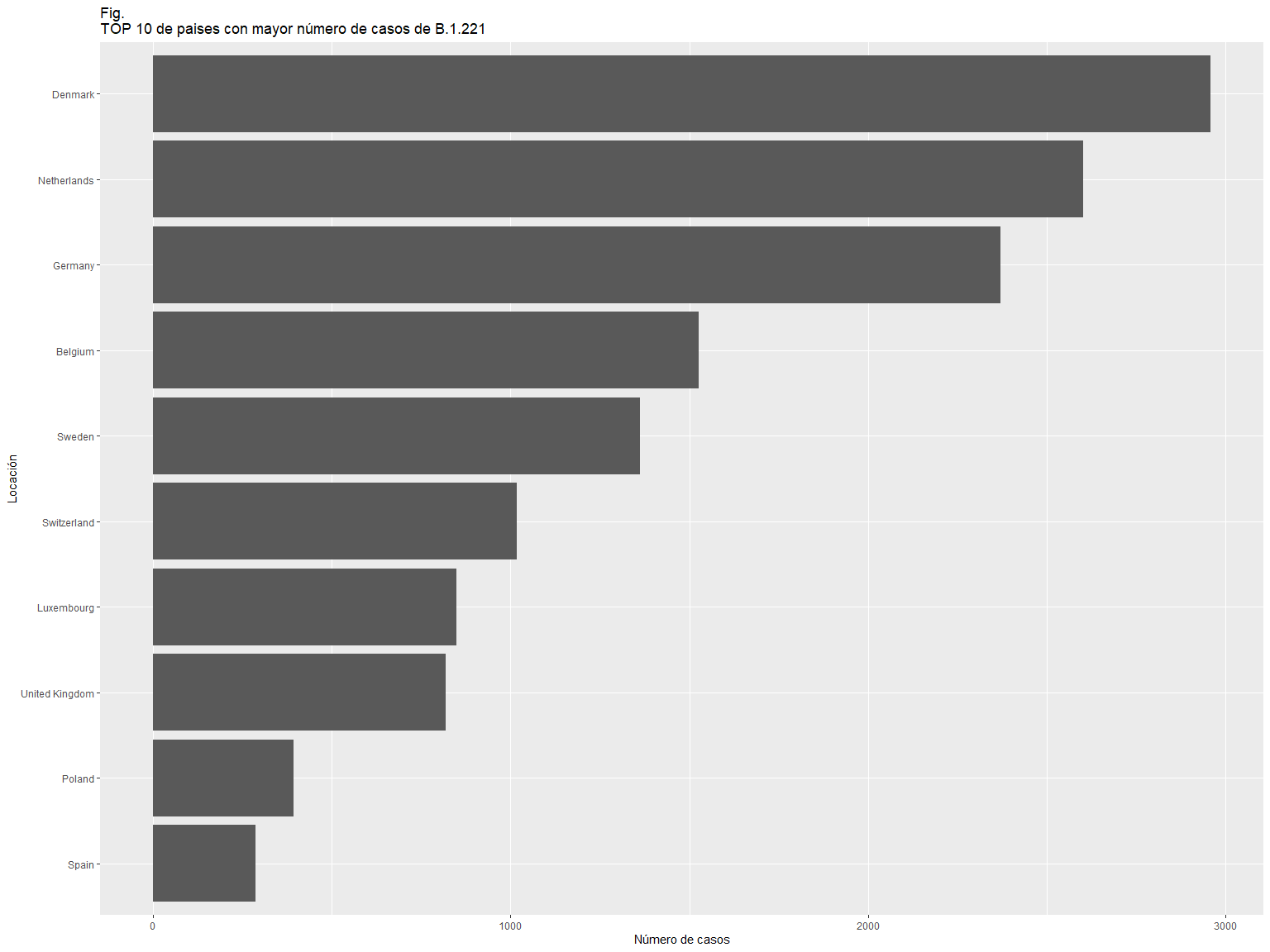
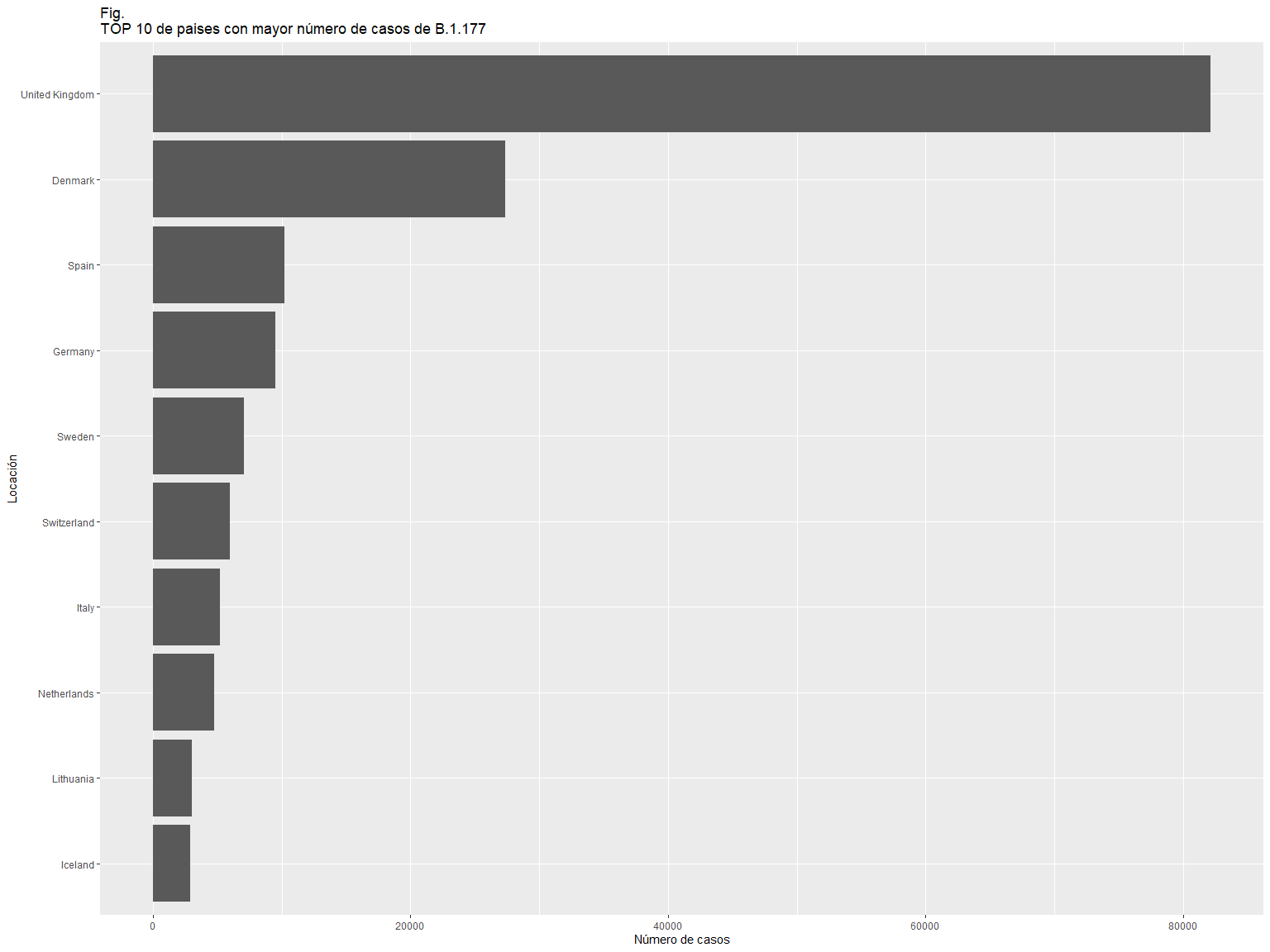
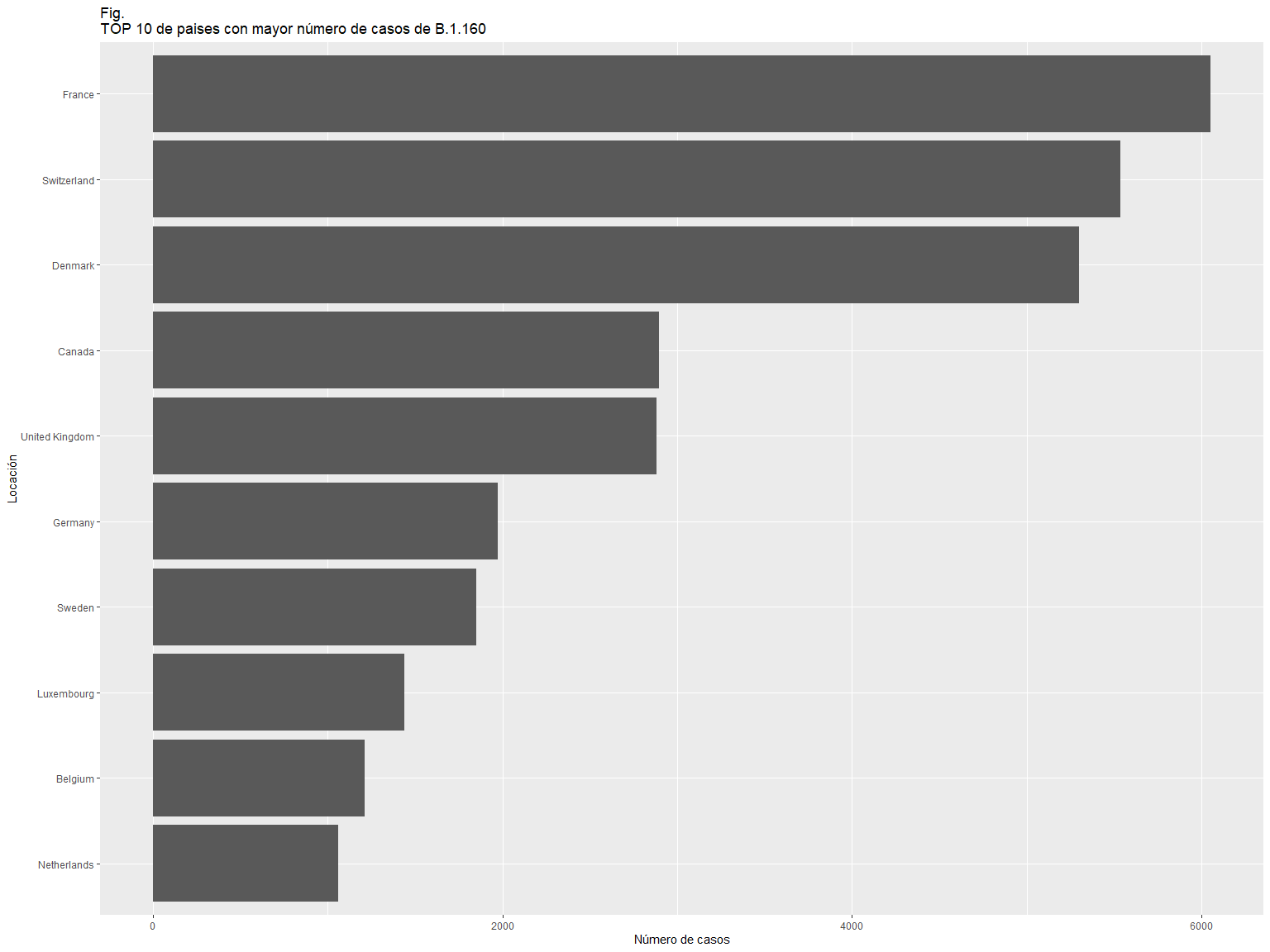
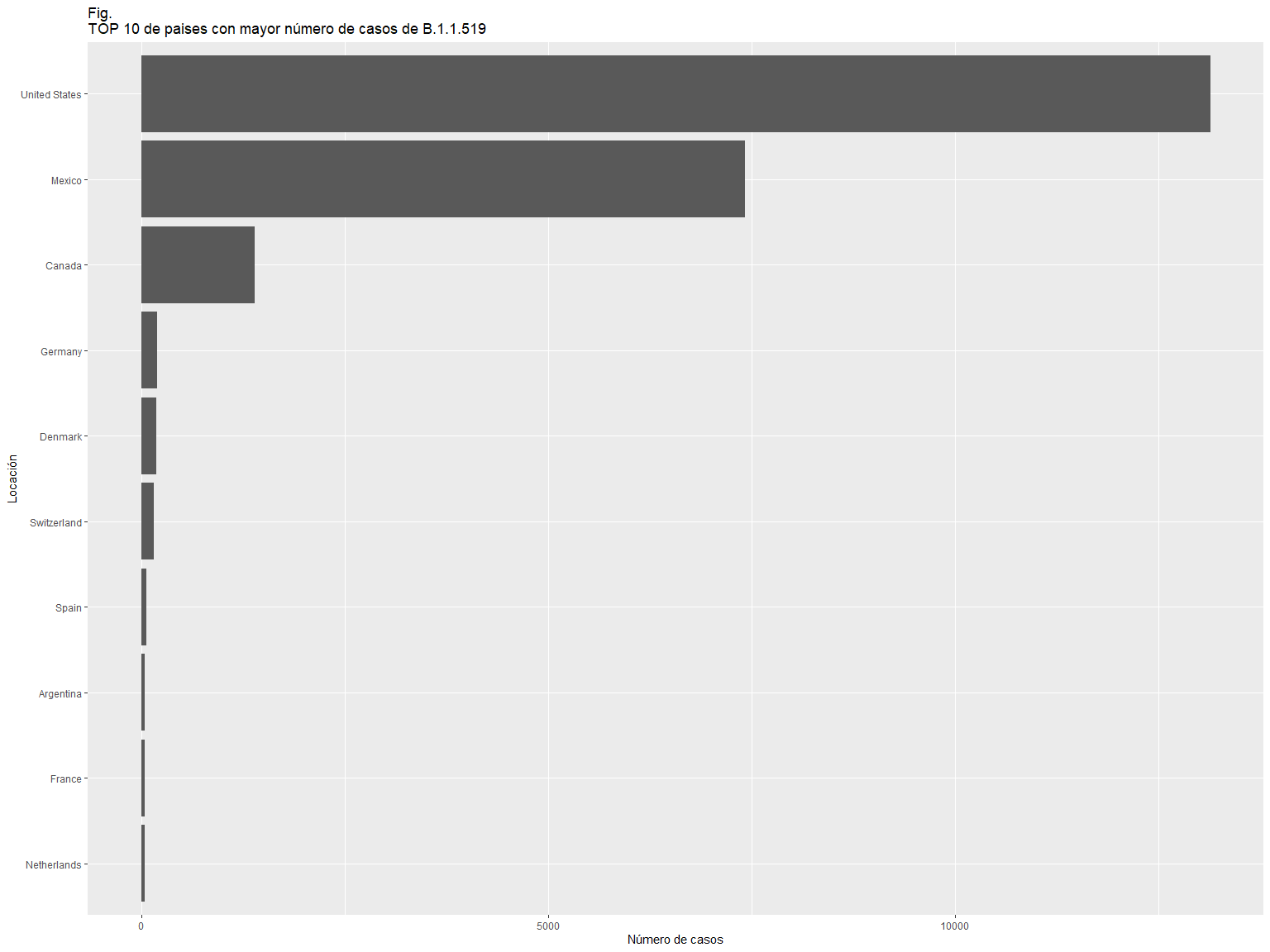
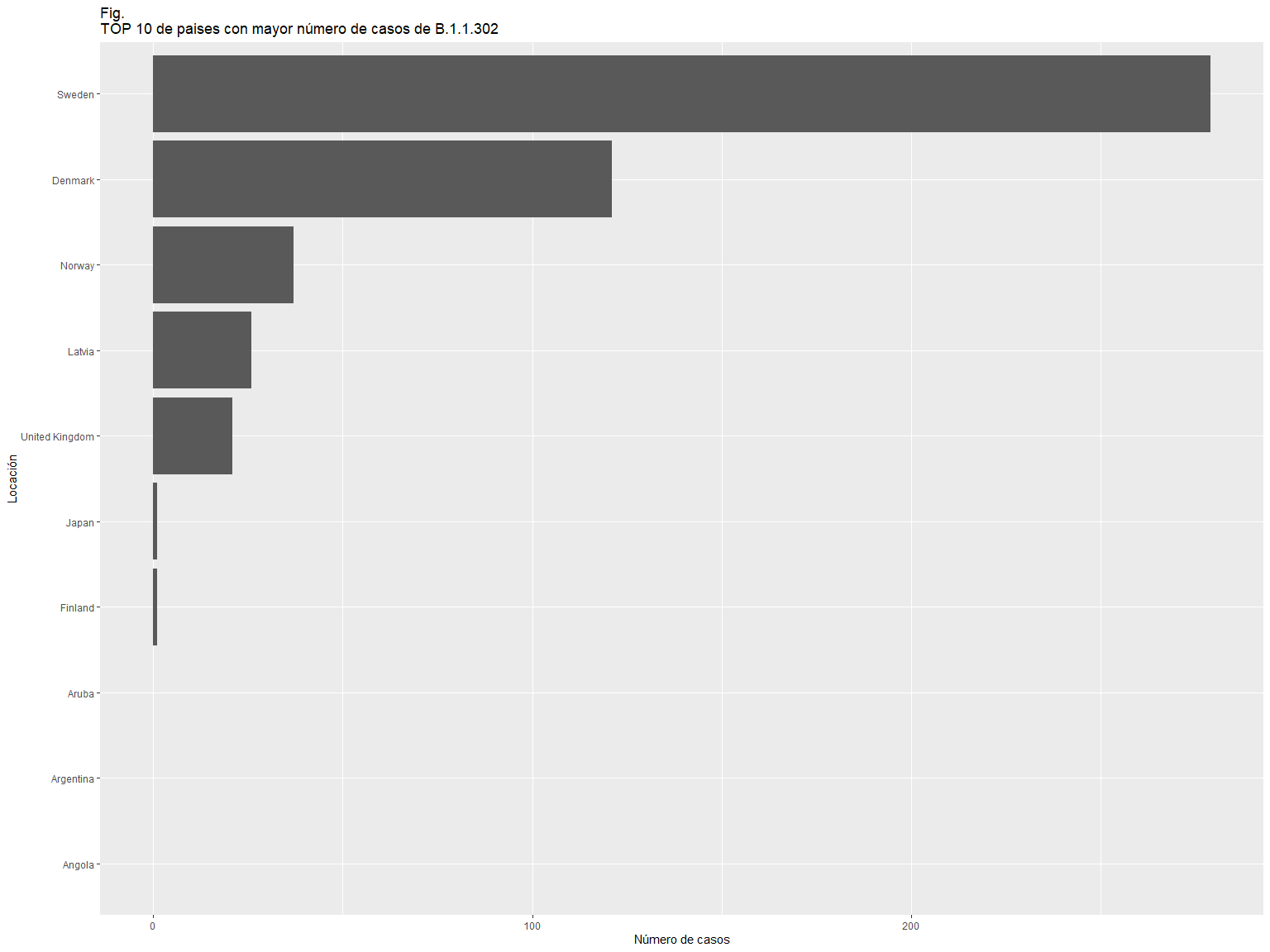
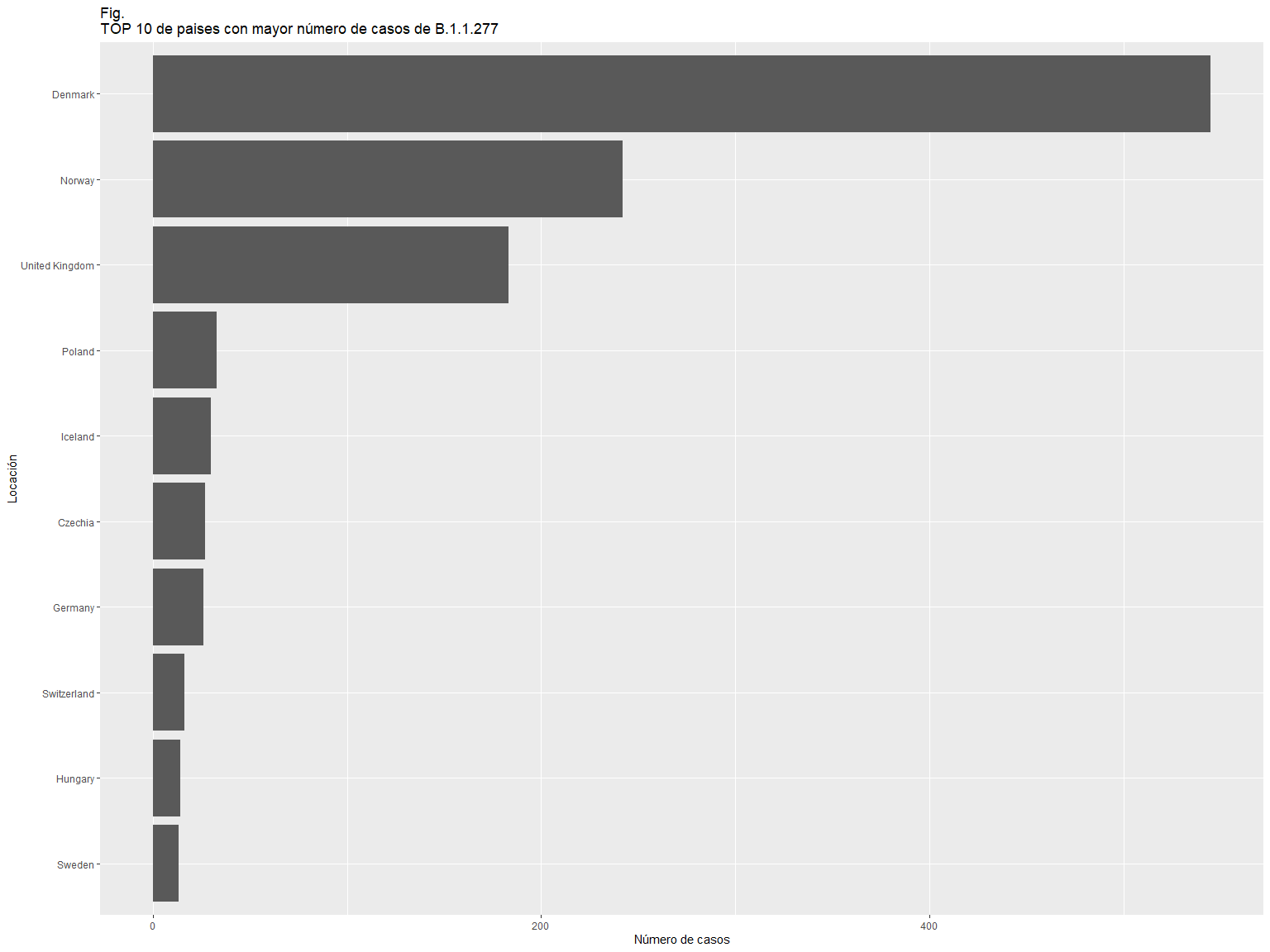
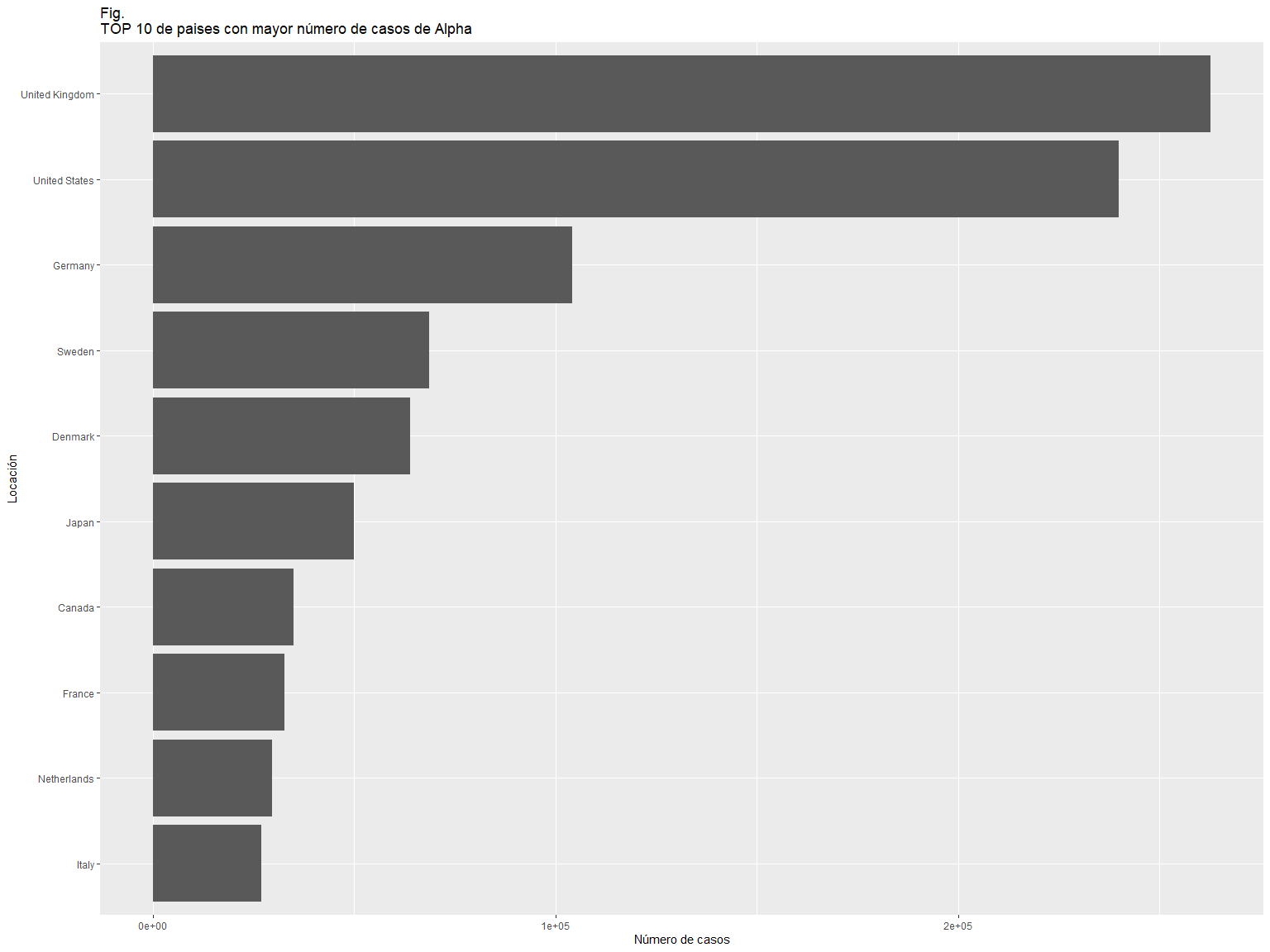
Se logra apreciar en la ilustración anterior como se encuentran los casos según los paises.

# 4. Casos de variantes por pais.

Se procede a mostrar la incidencia de las variables de COVID-19 en los paises.

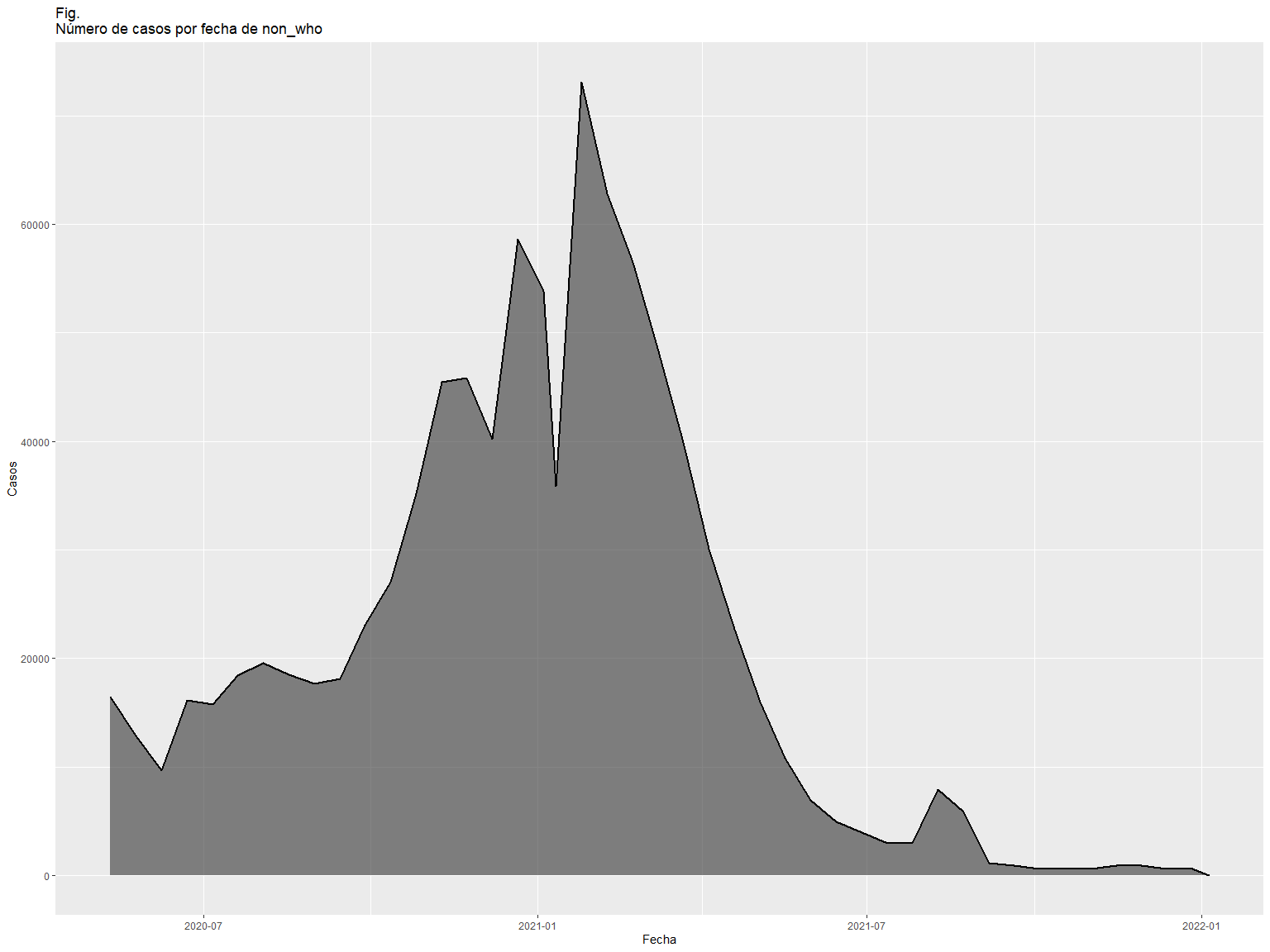
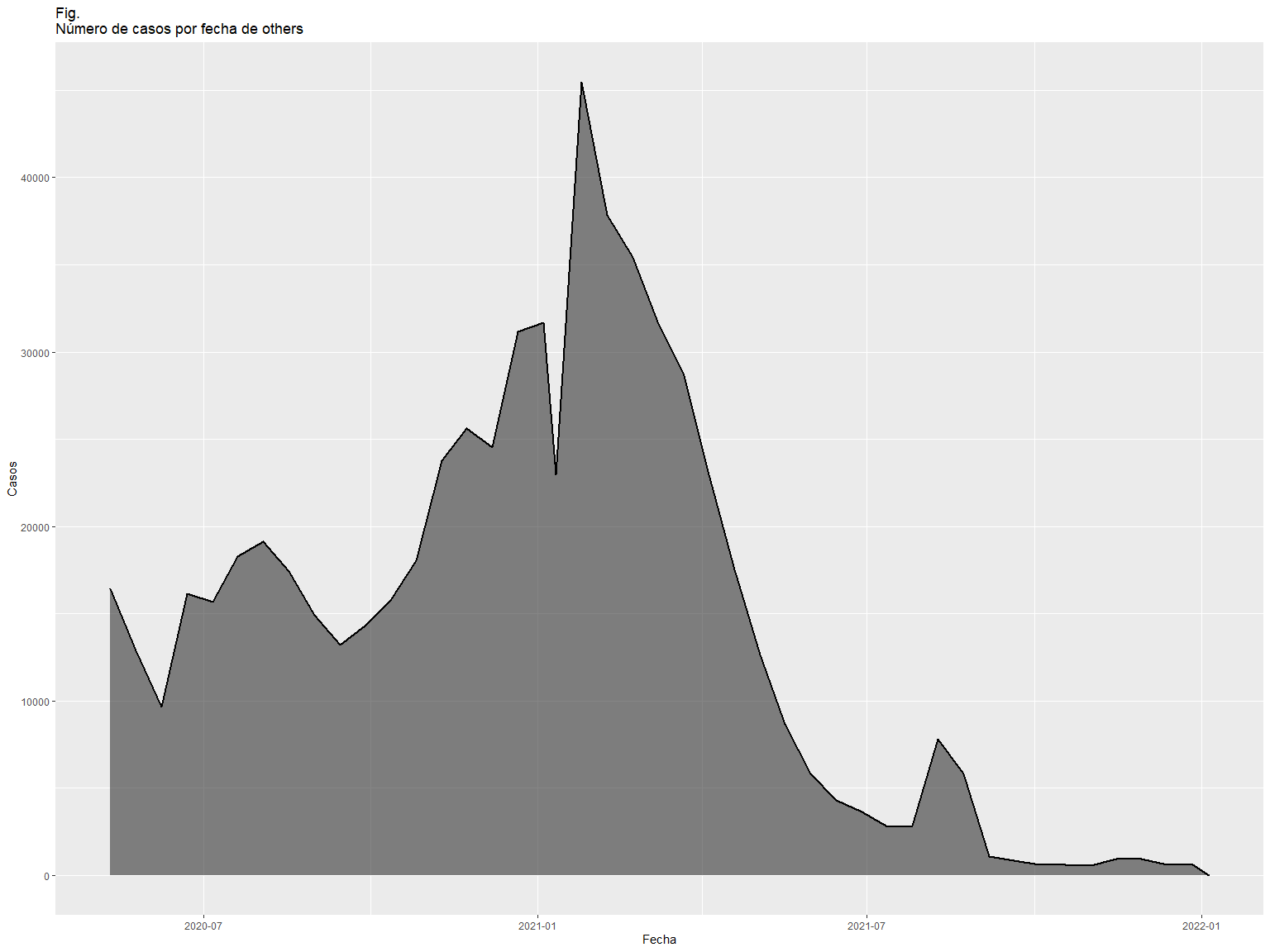
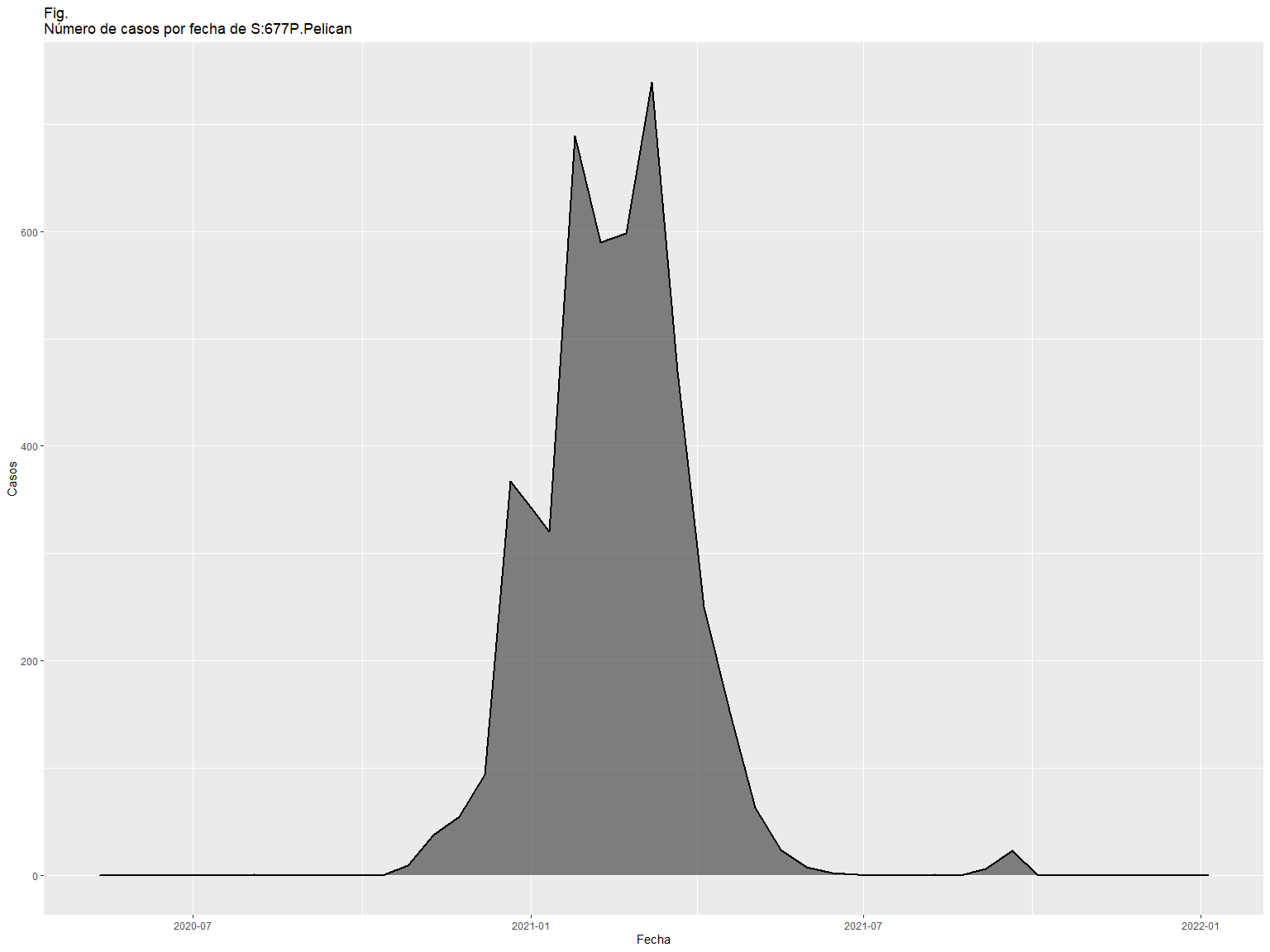
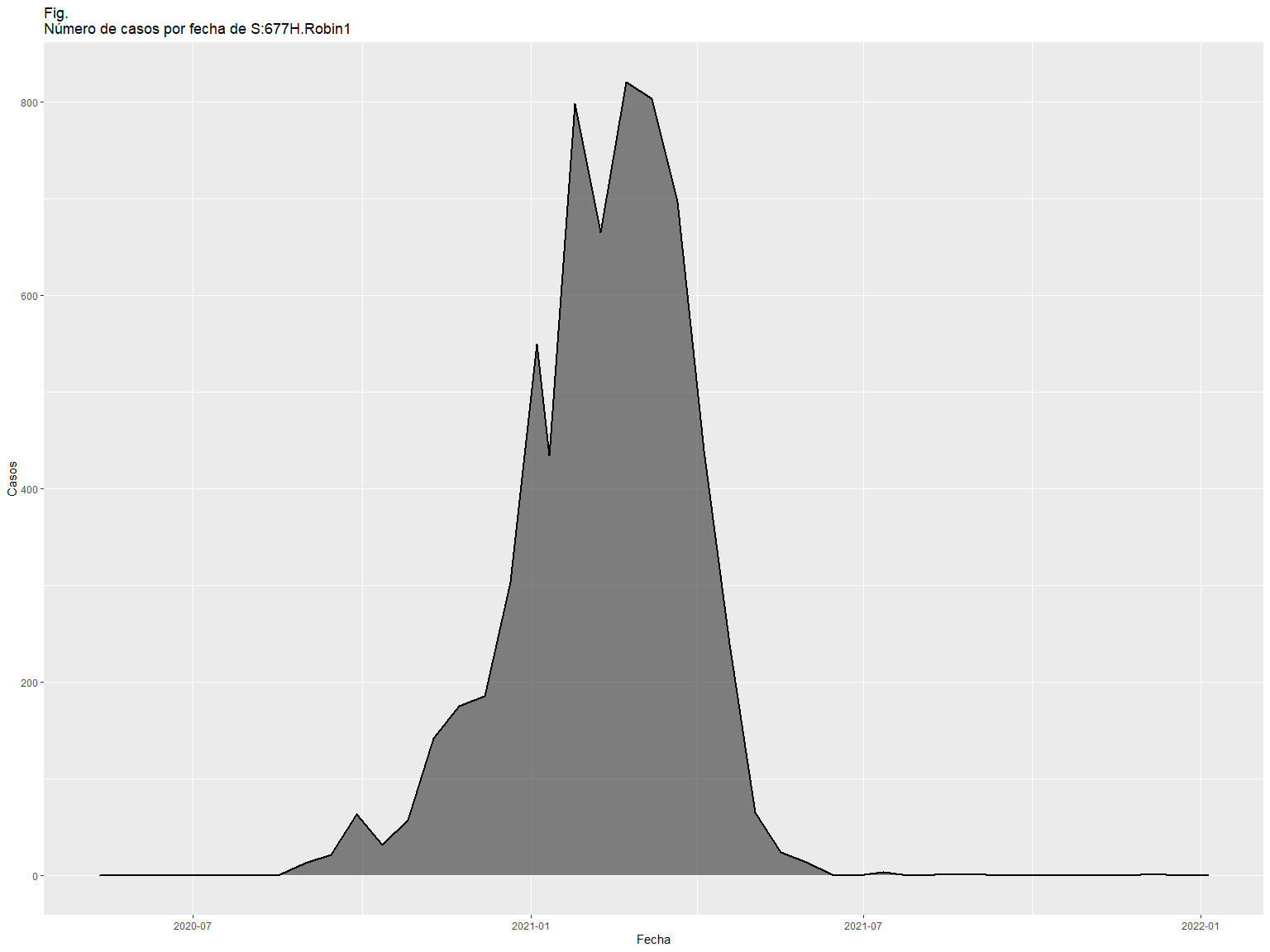
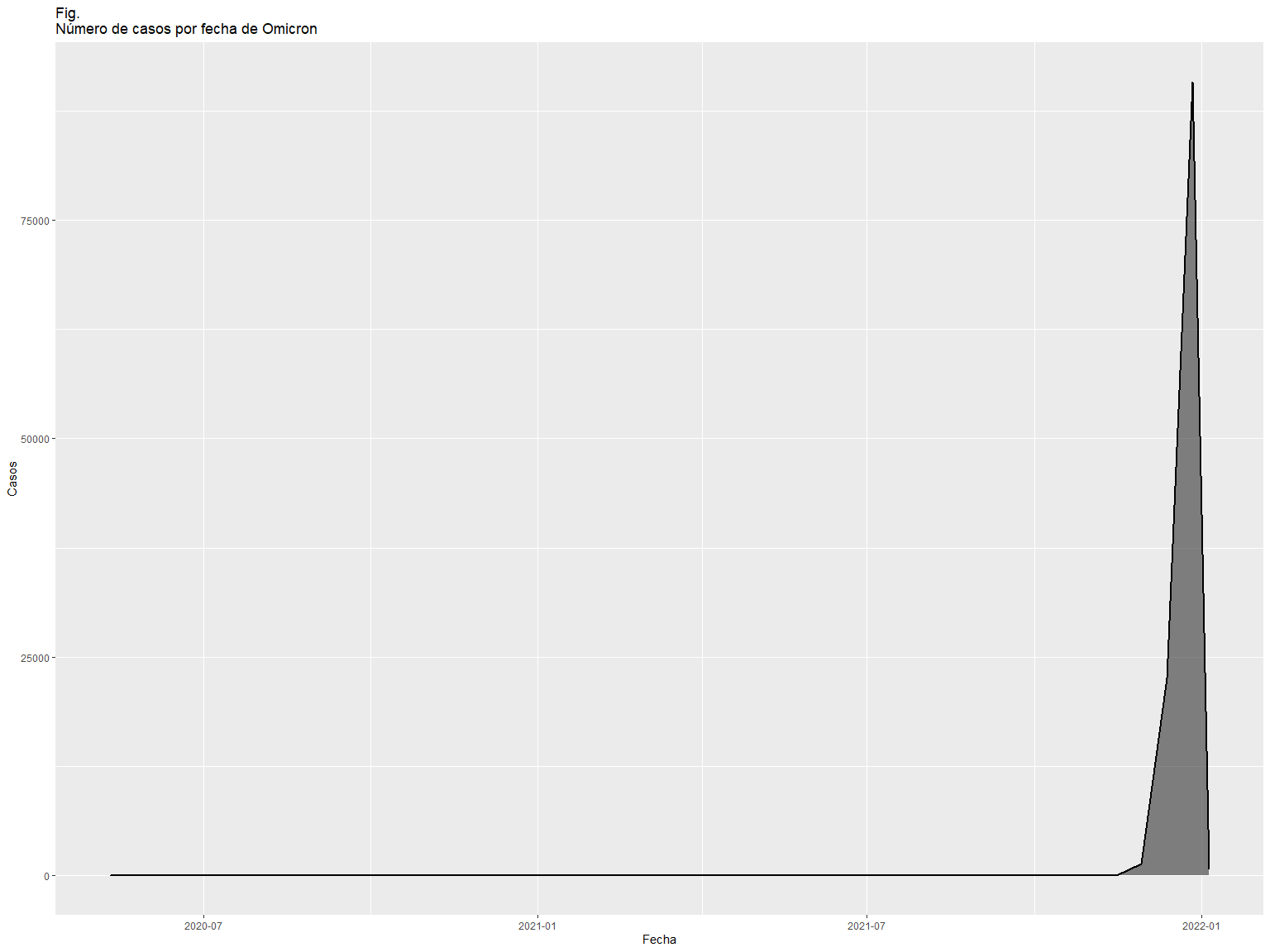
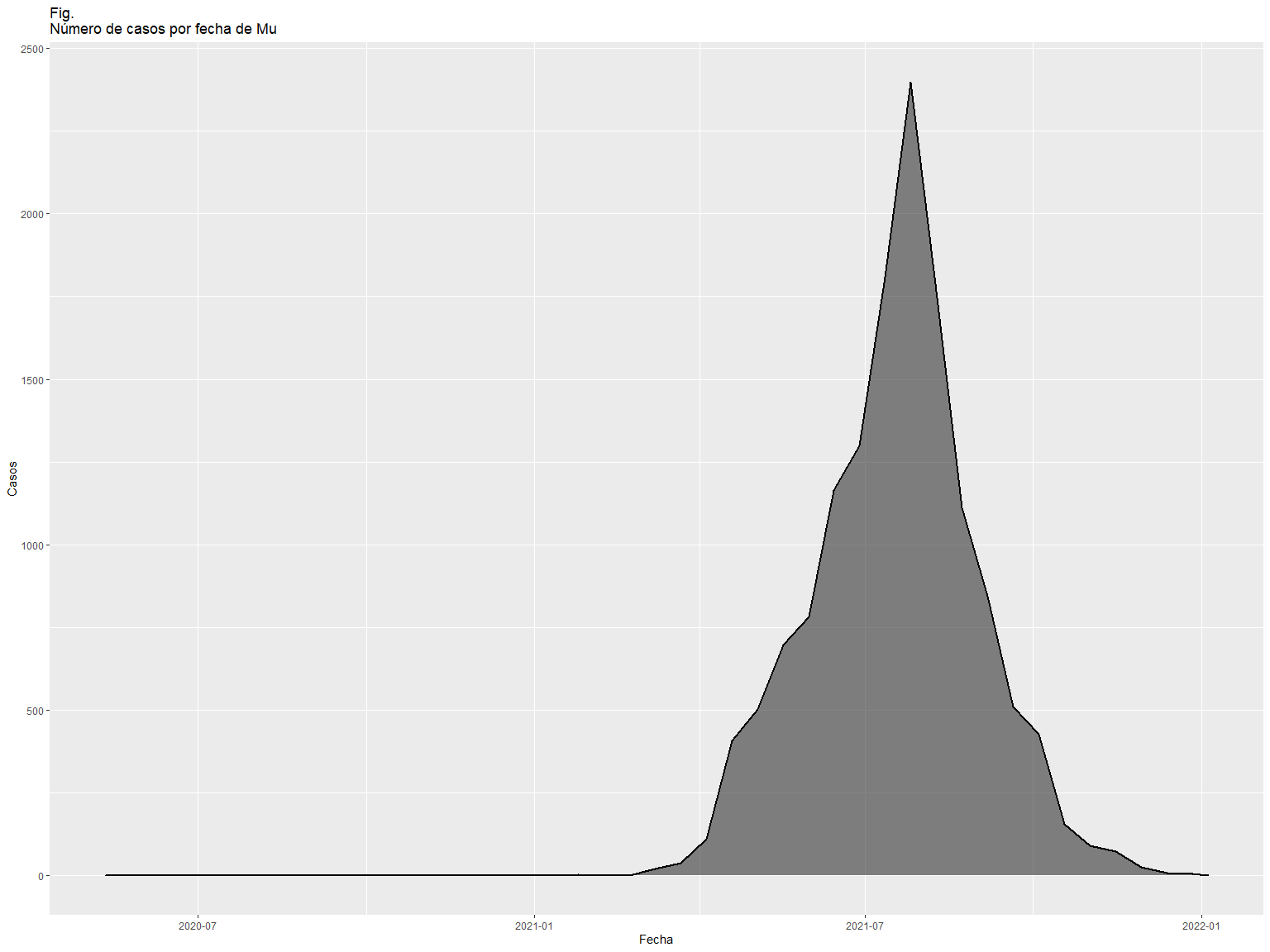
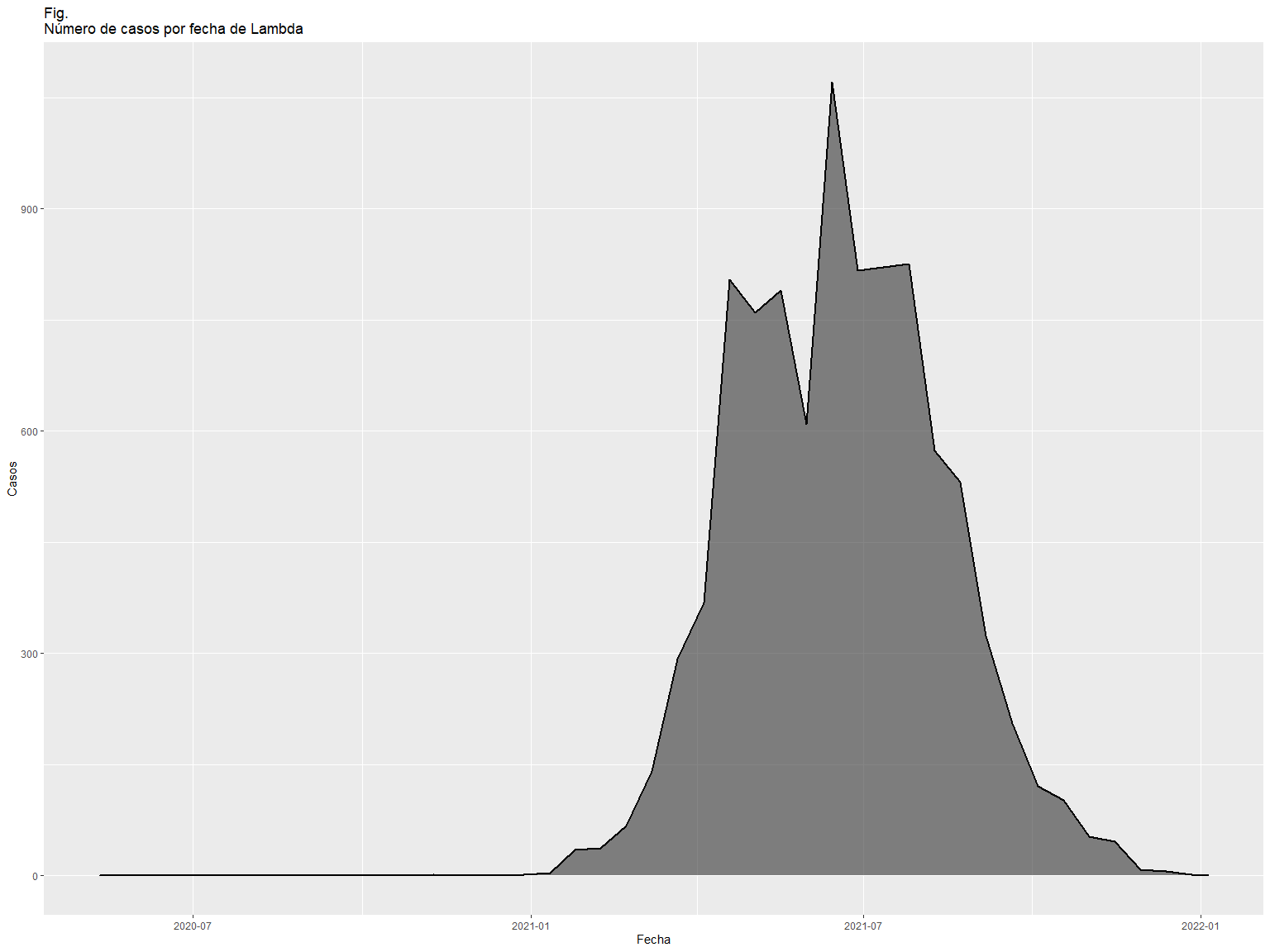
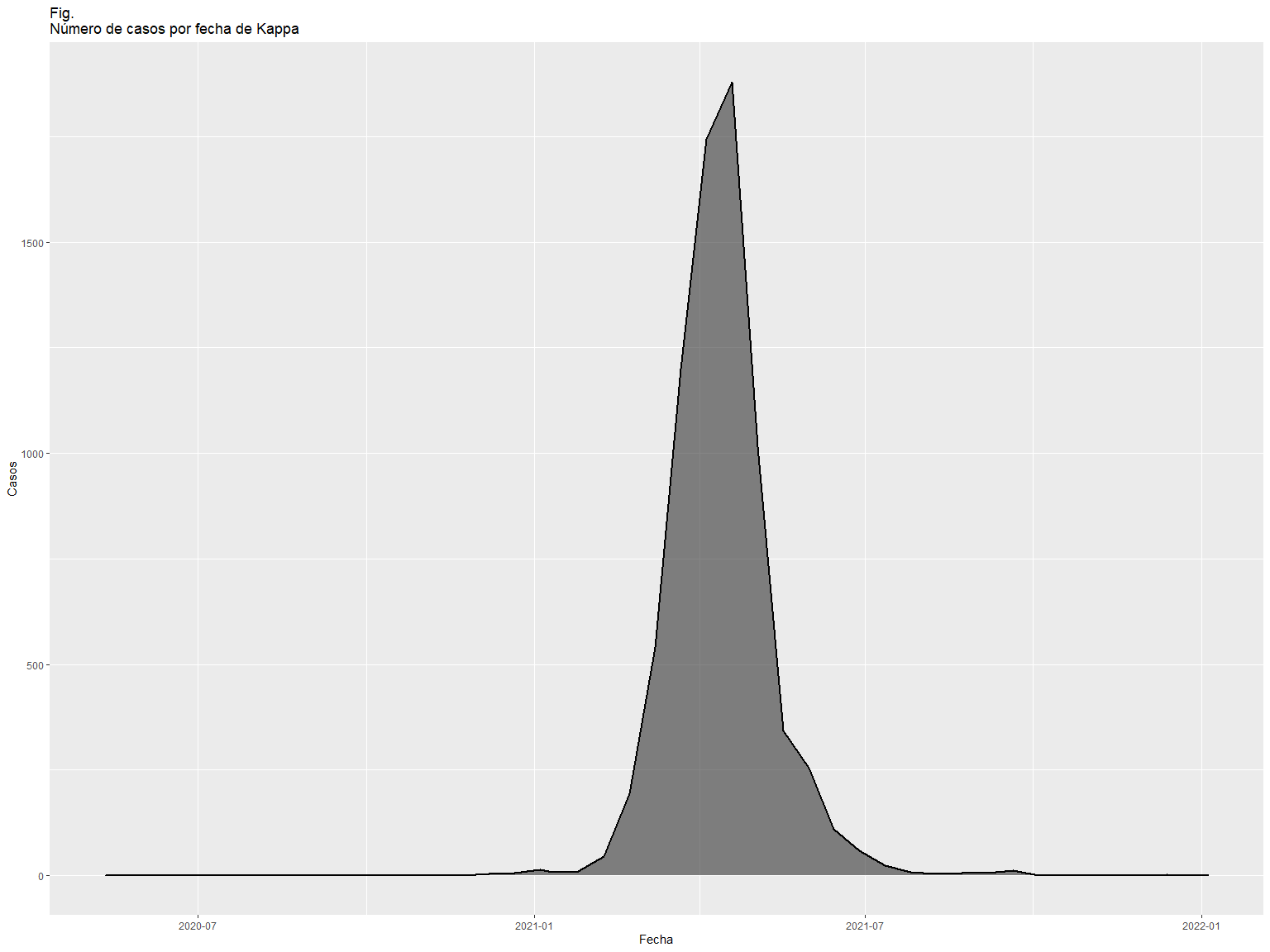
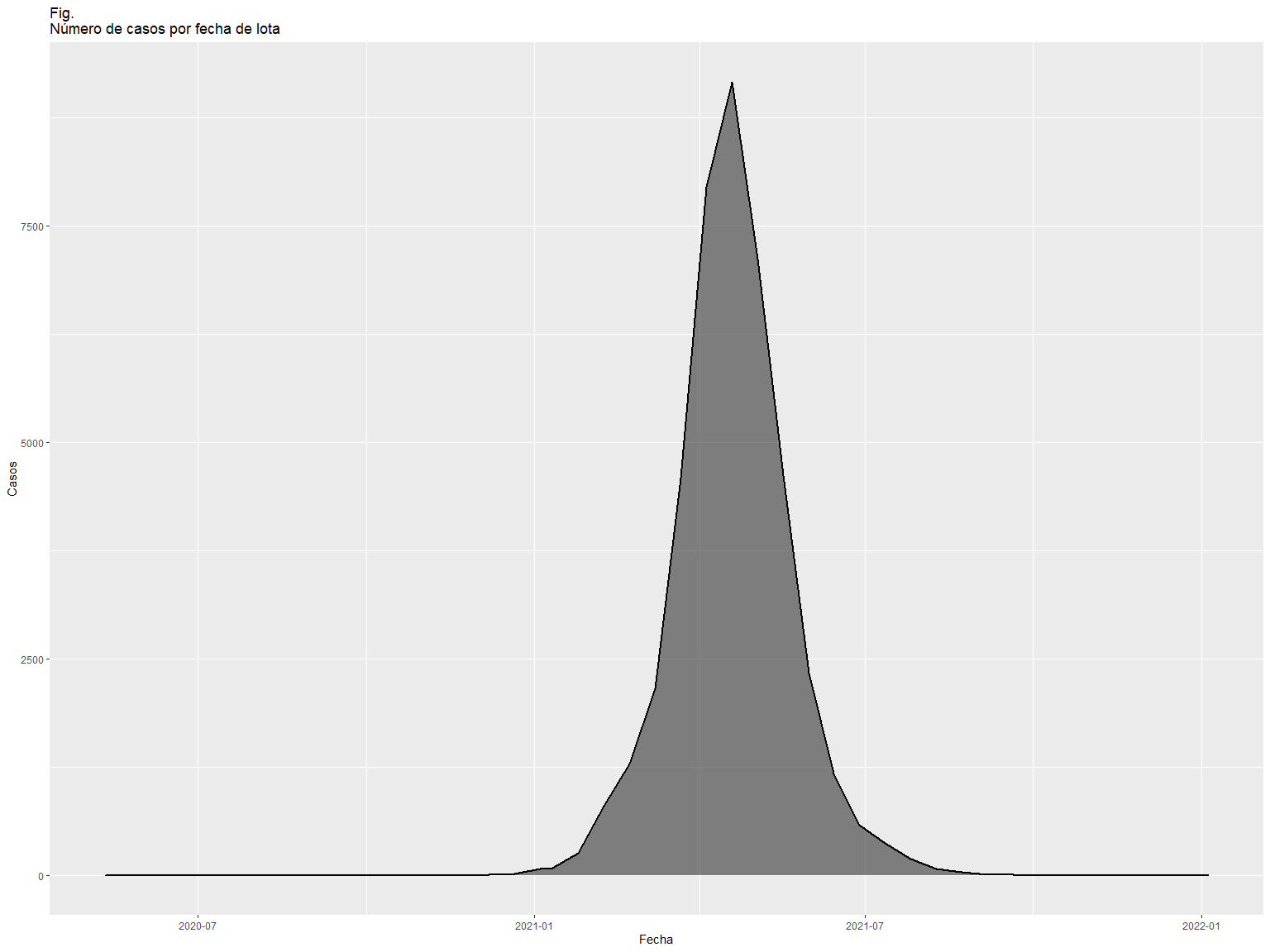
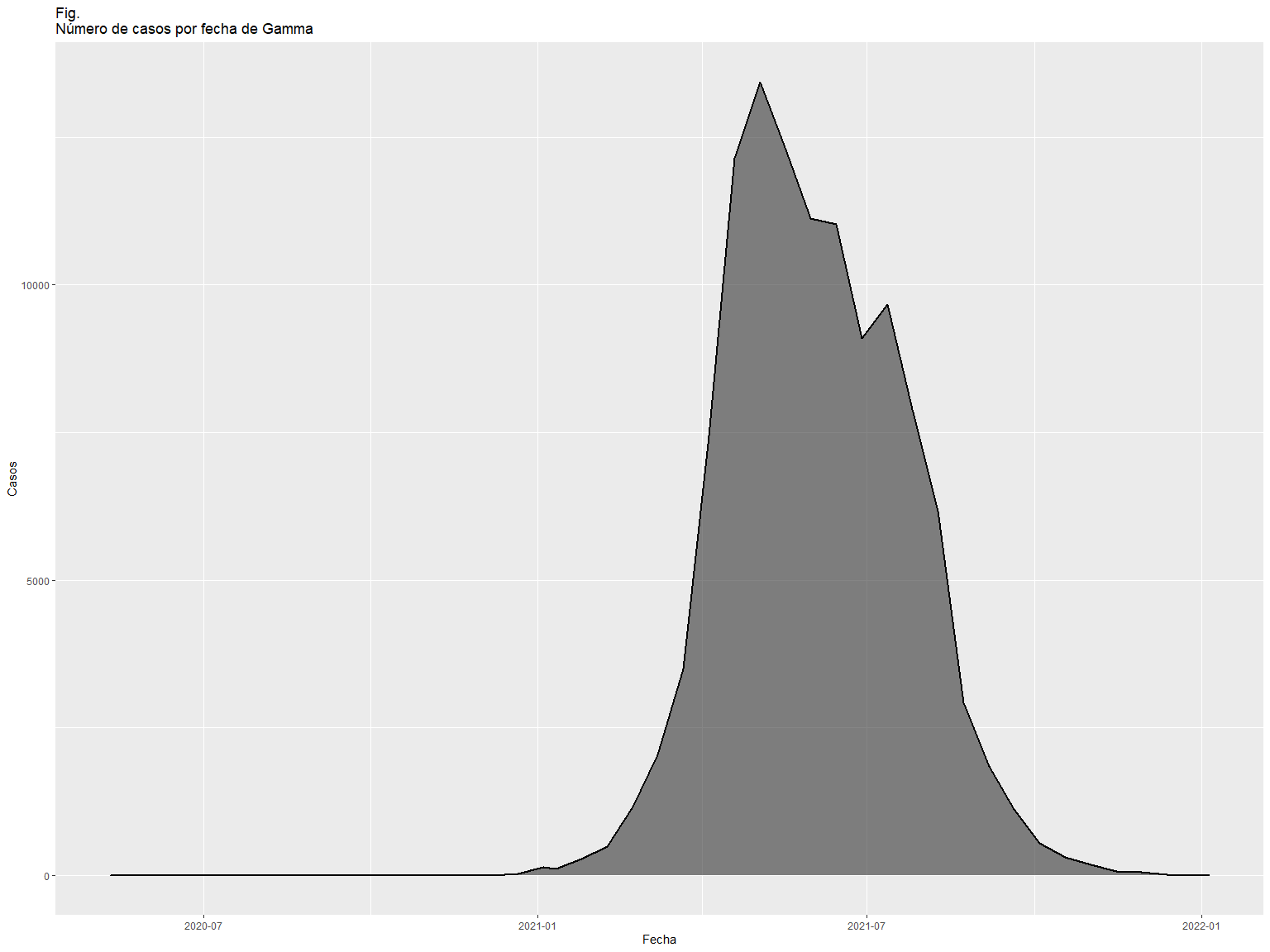
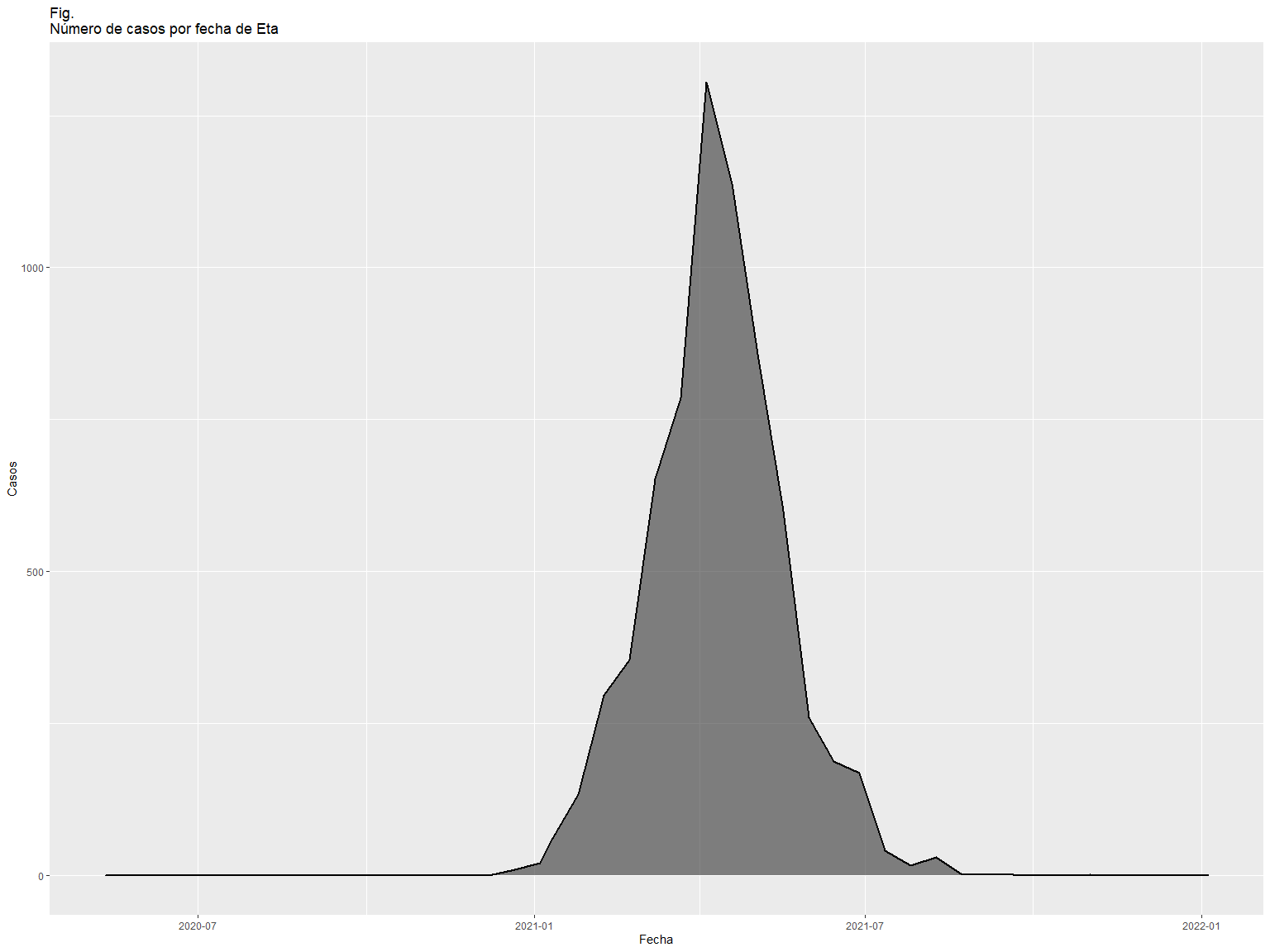
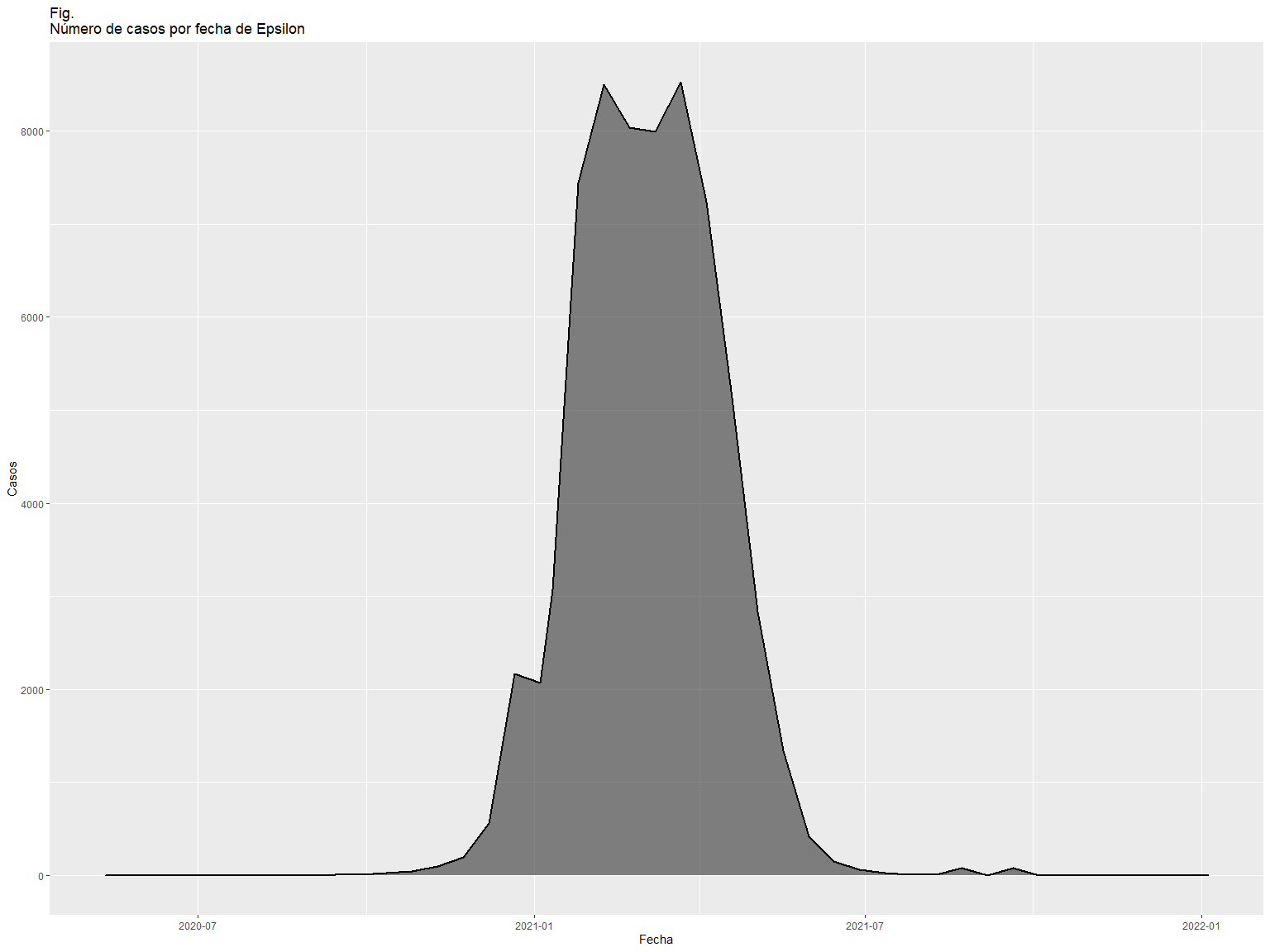
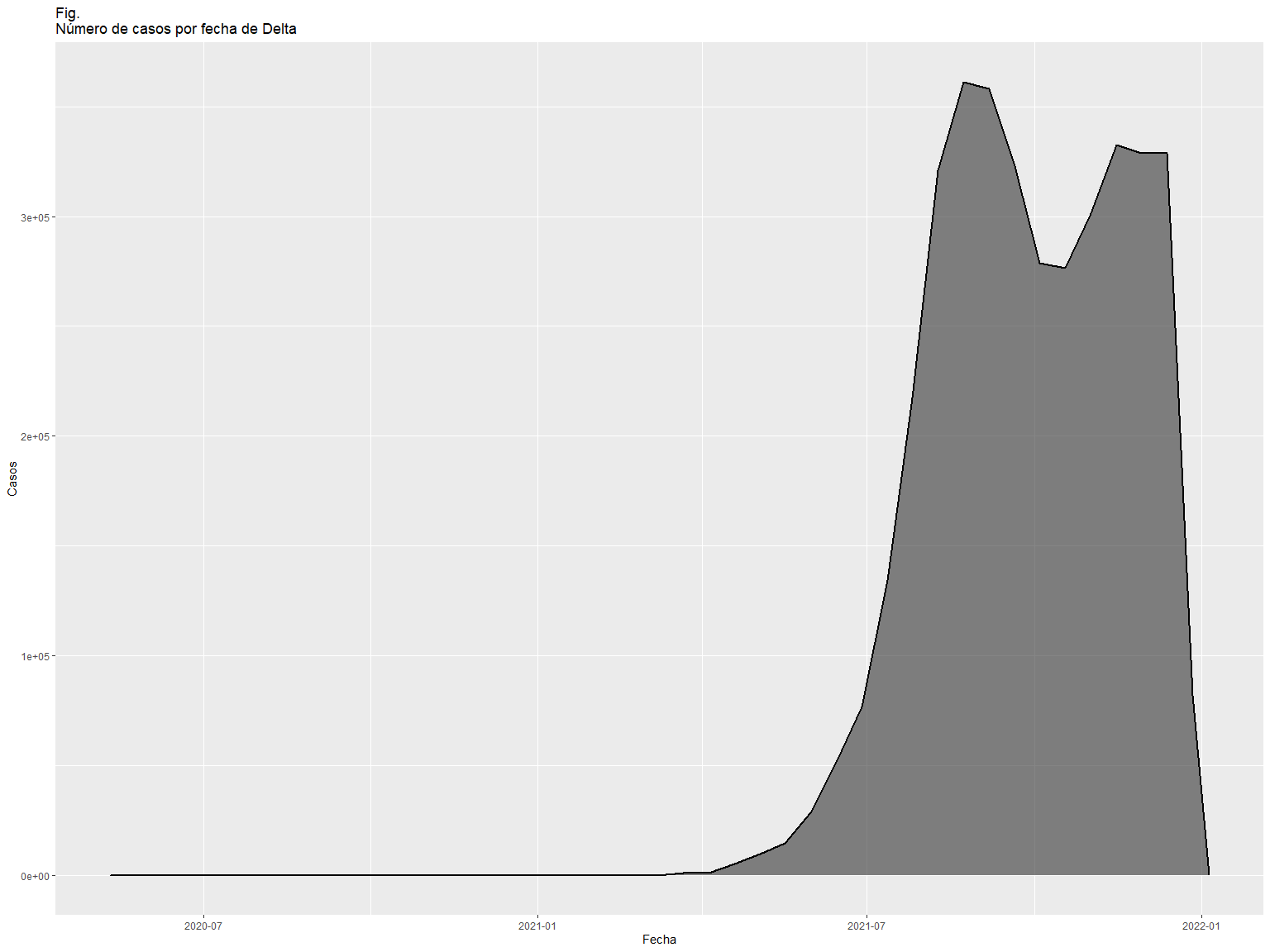
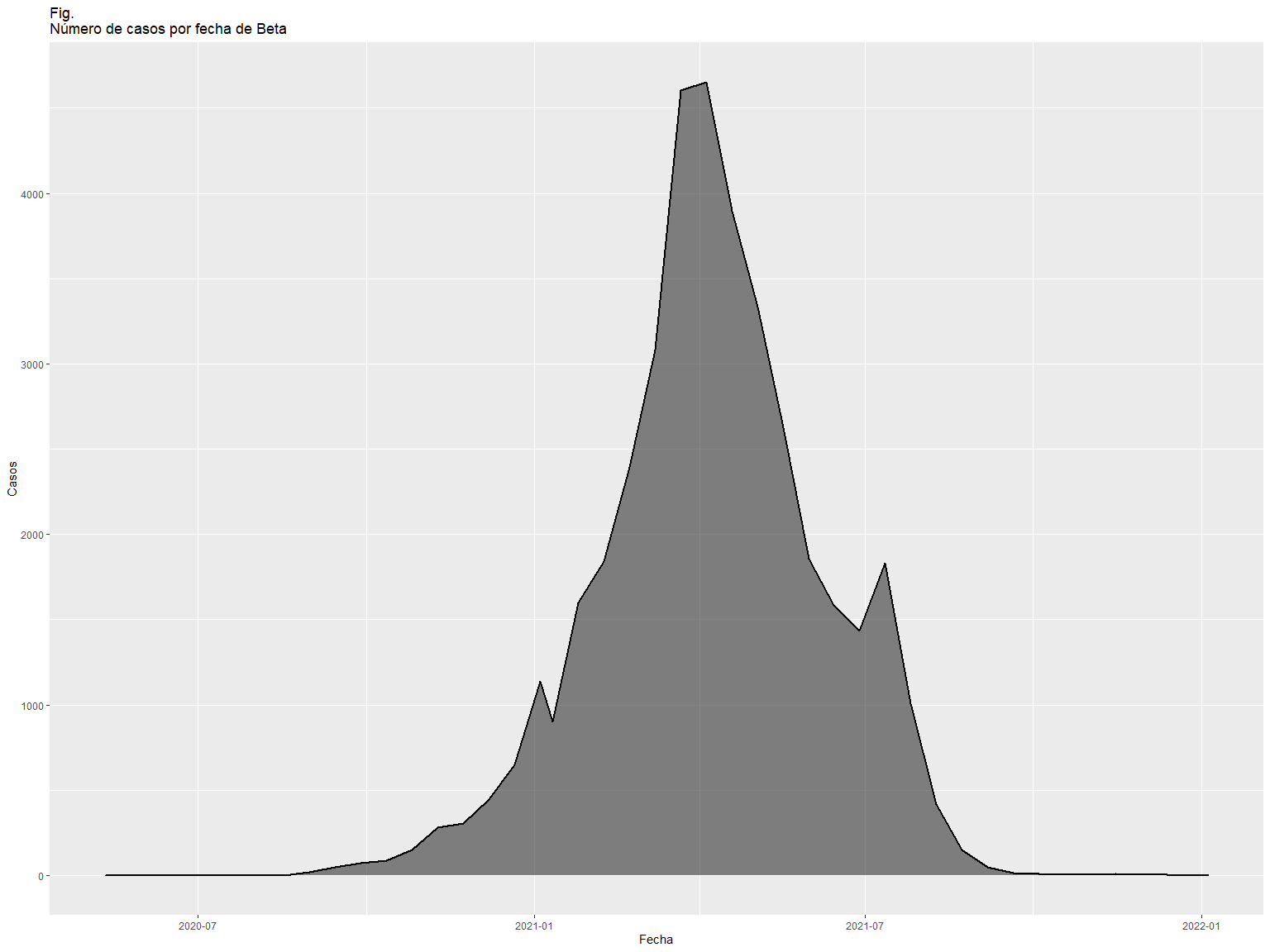
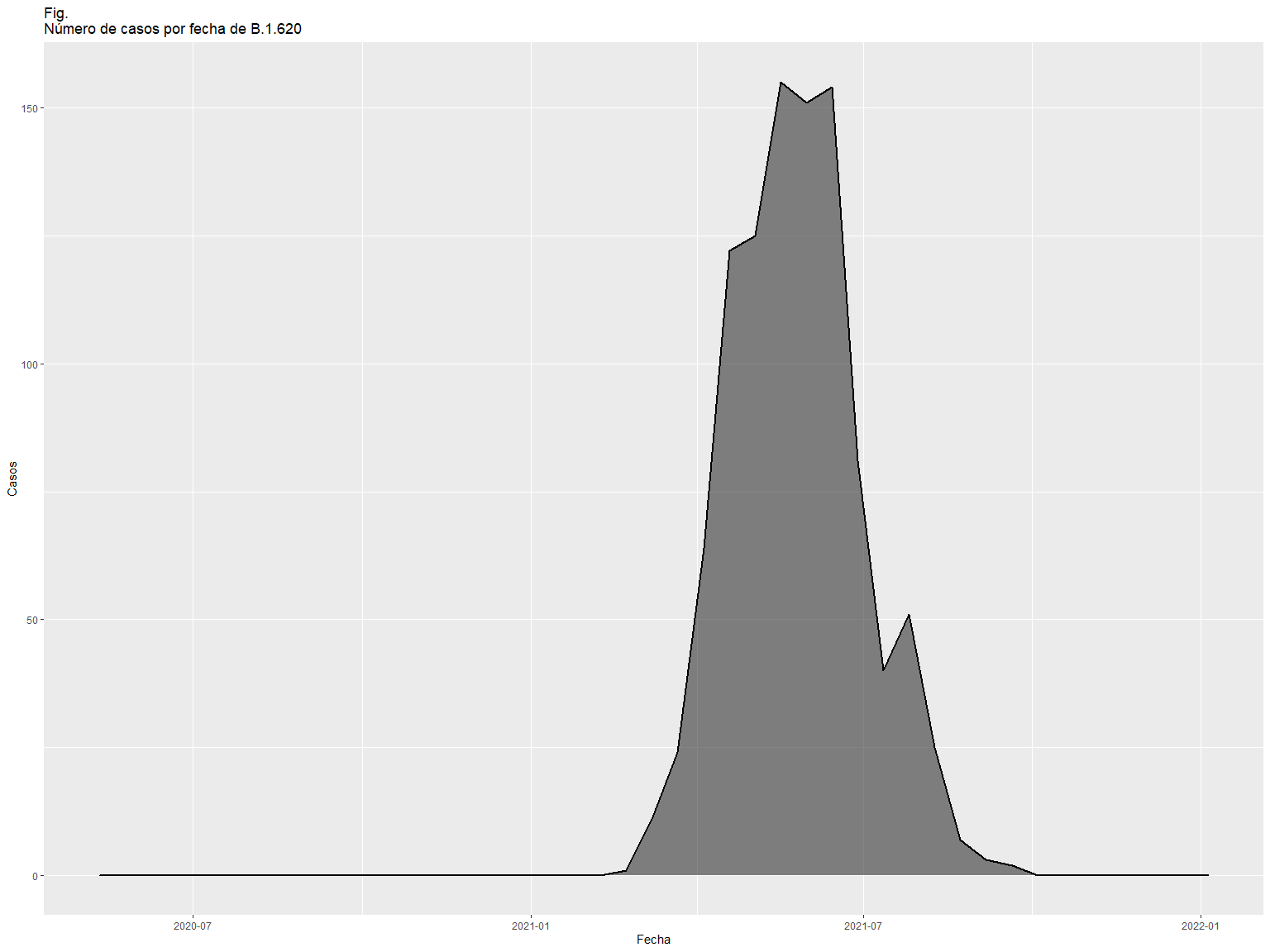
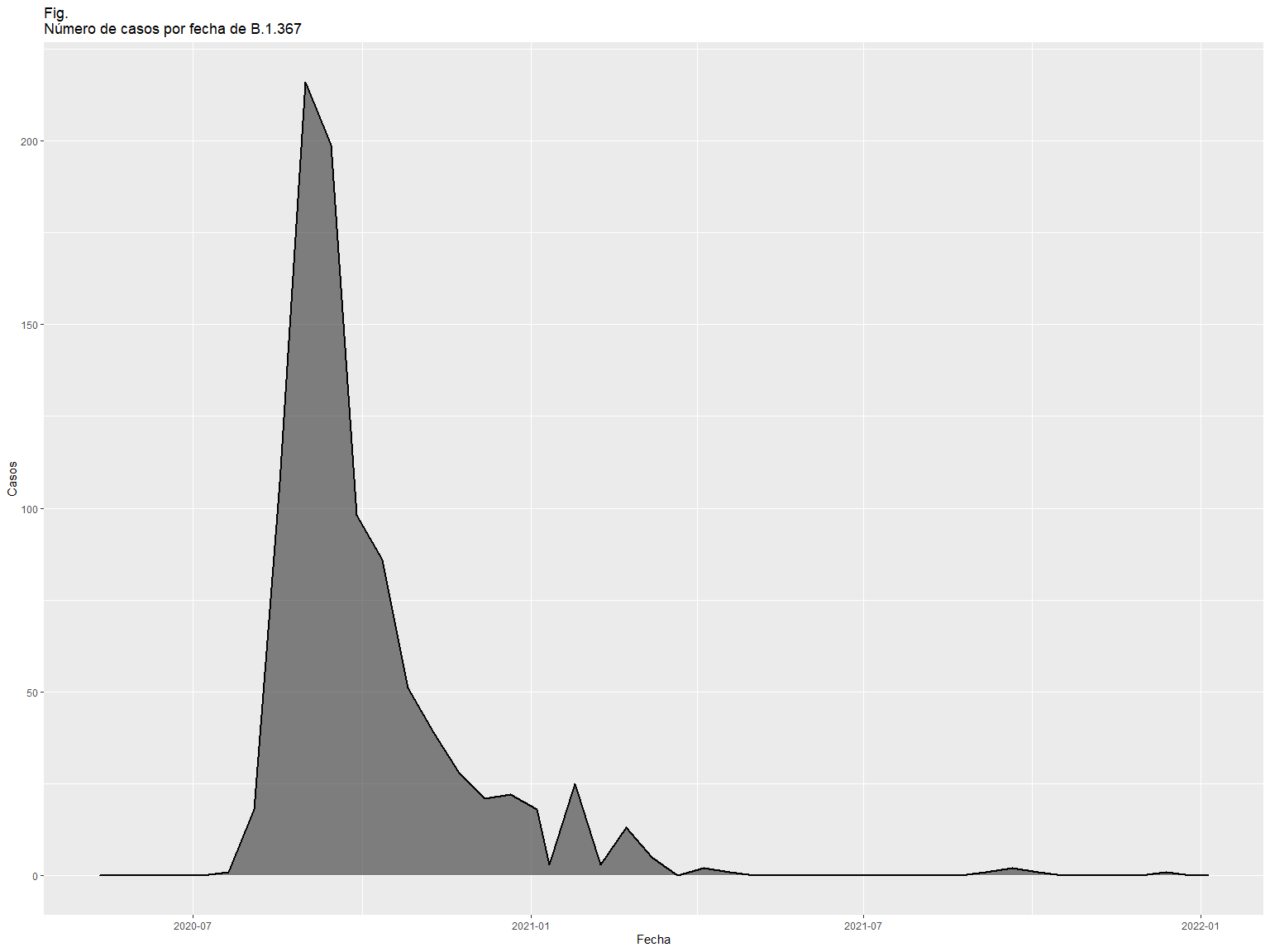
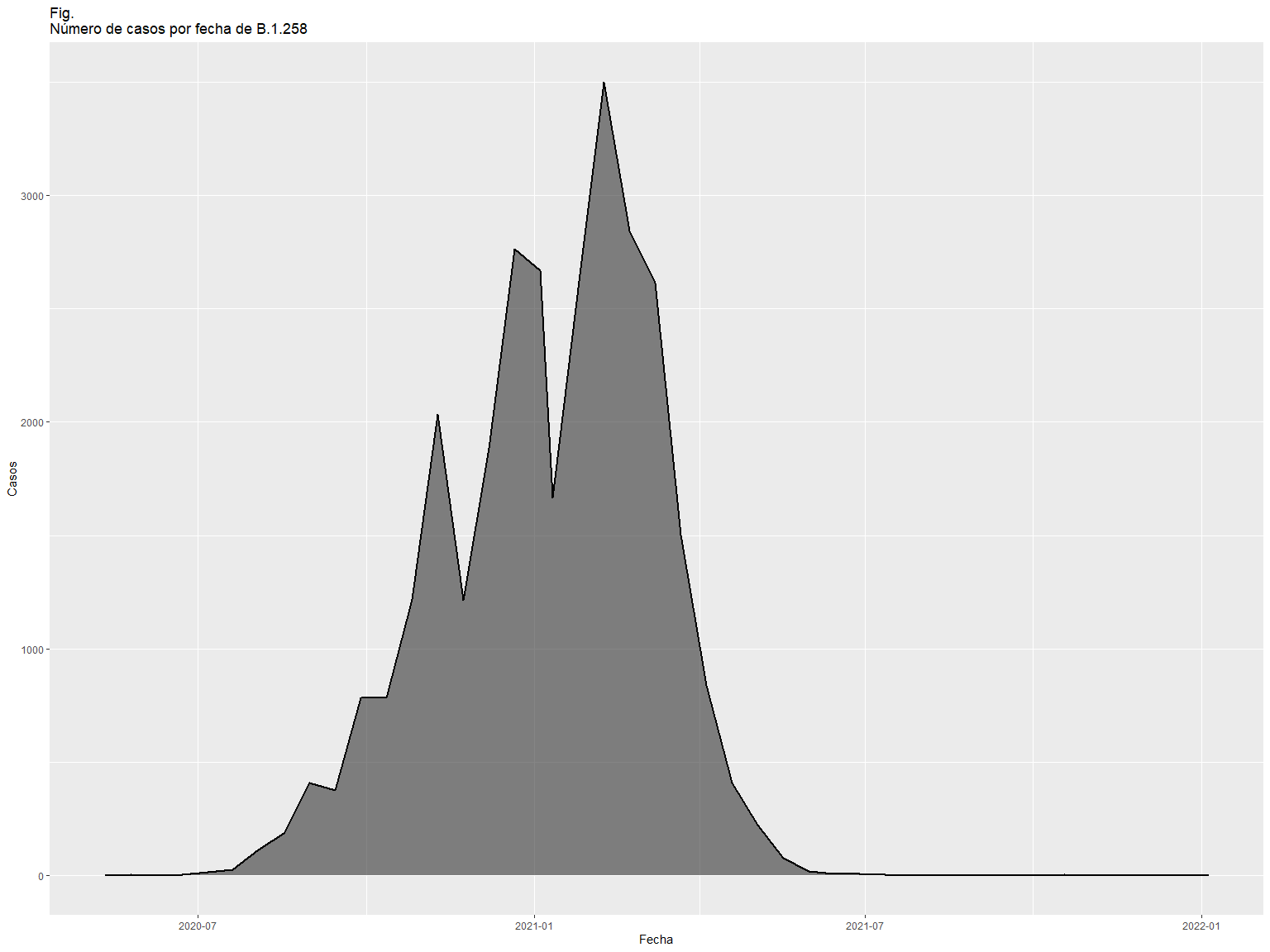
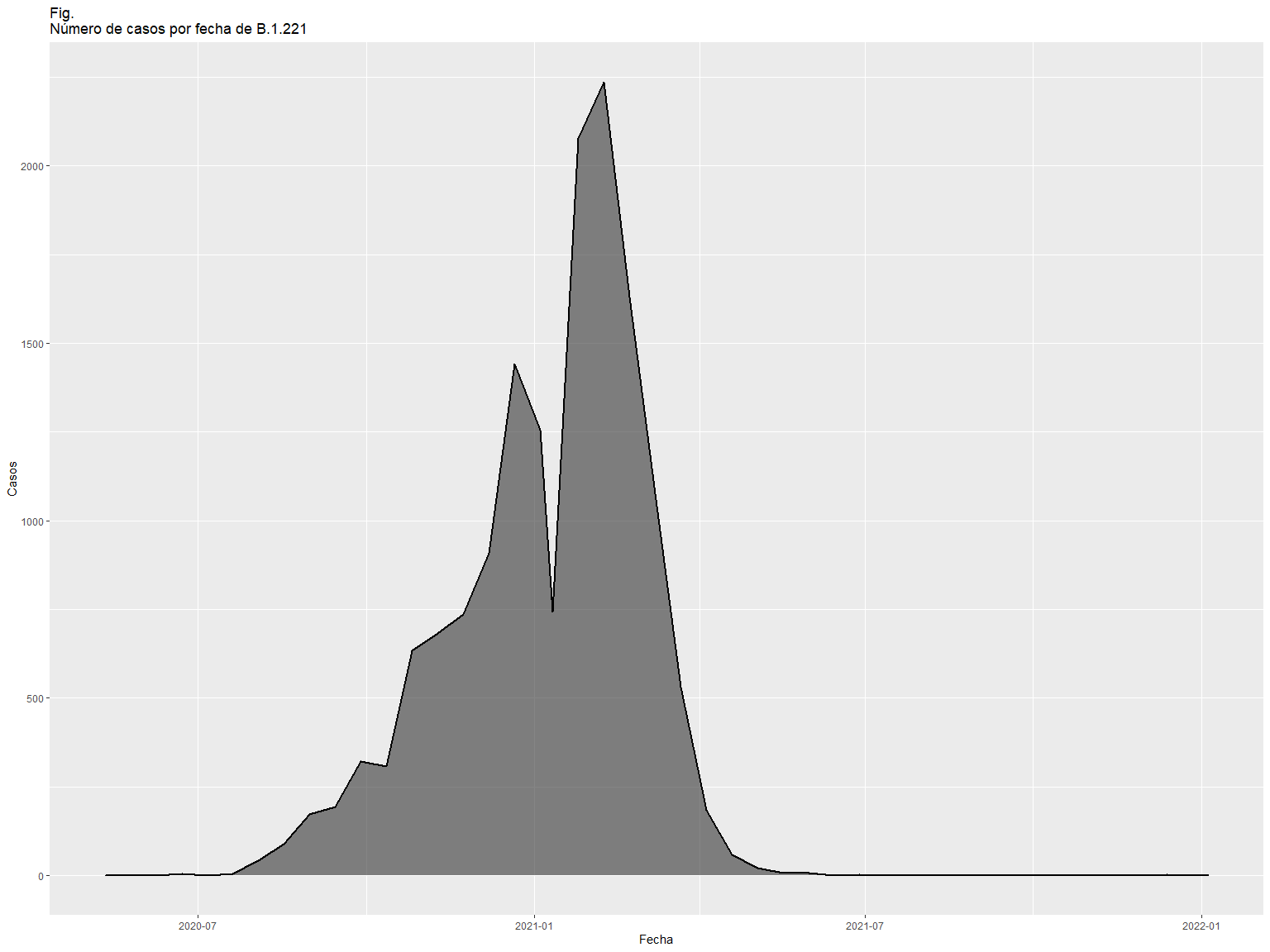
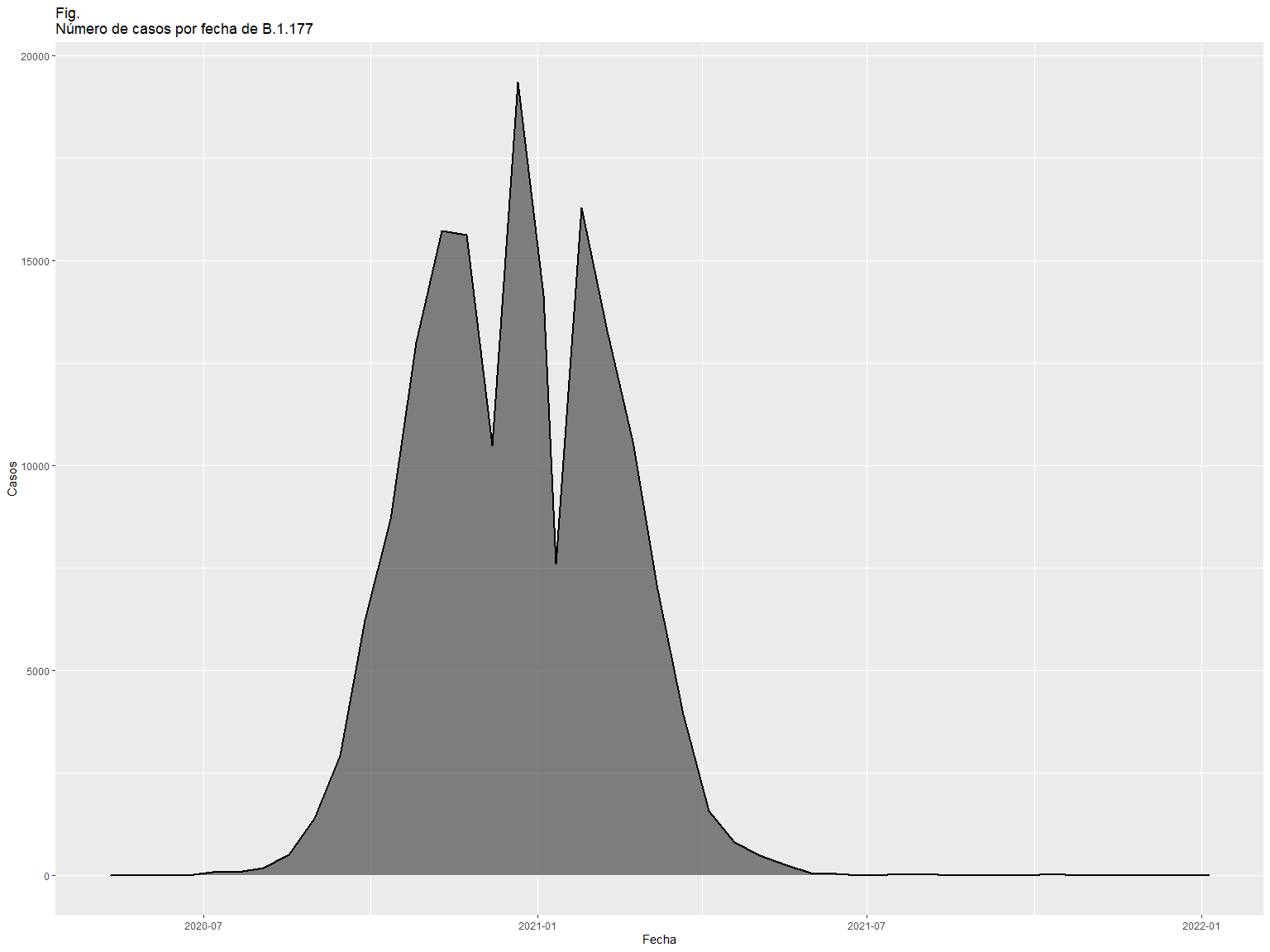
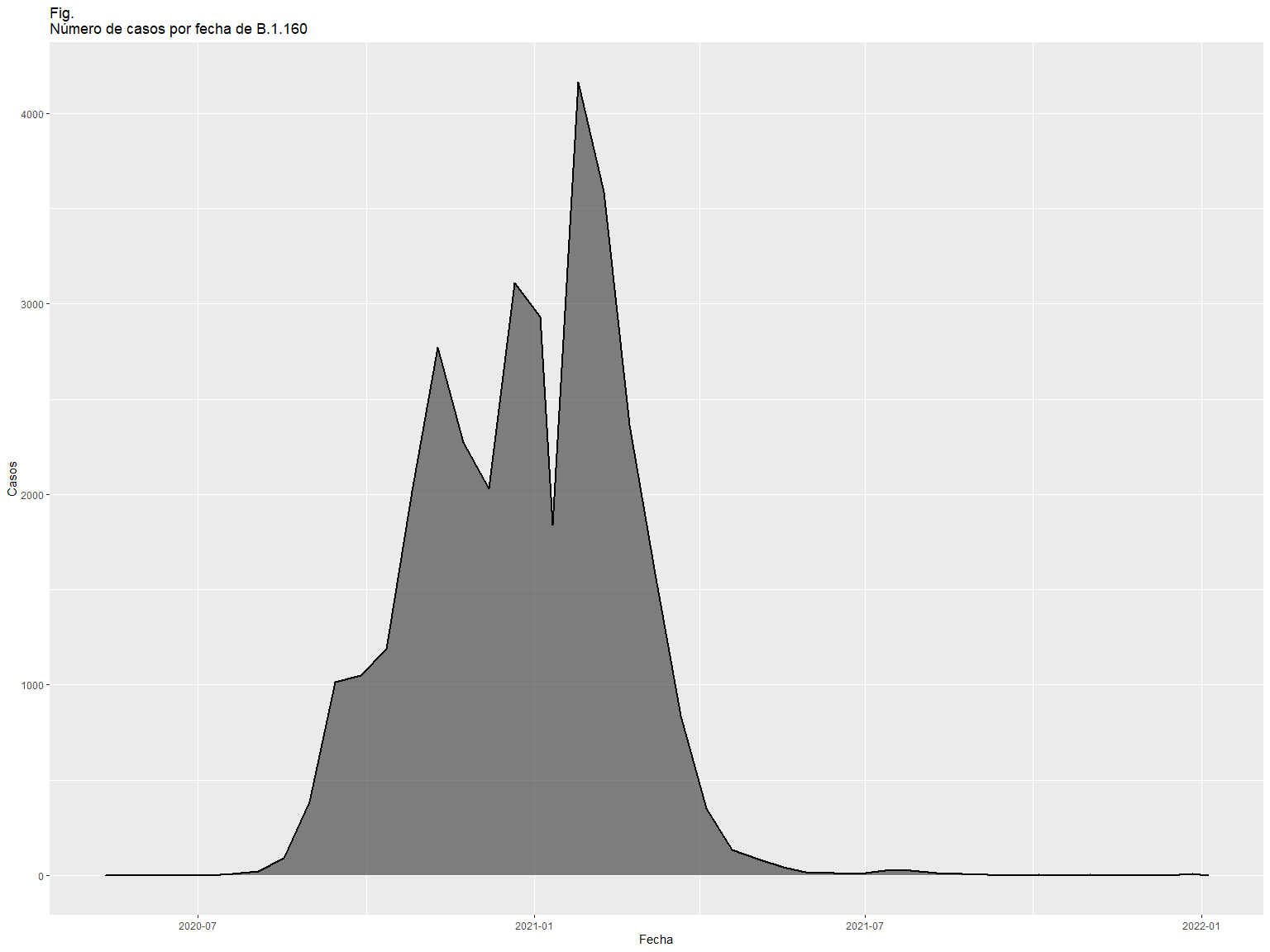
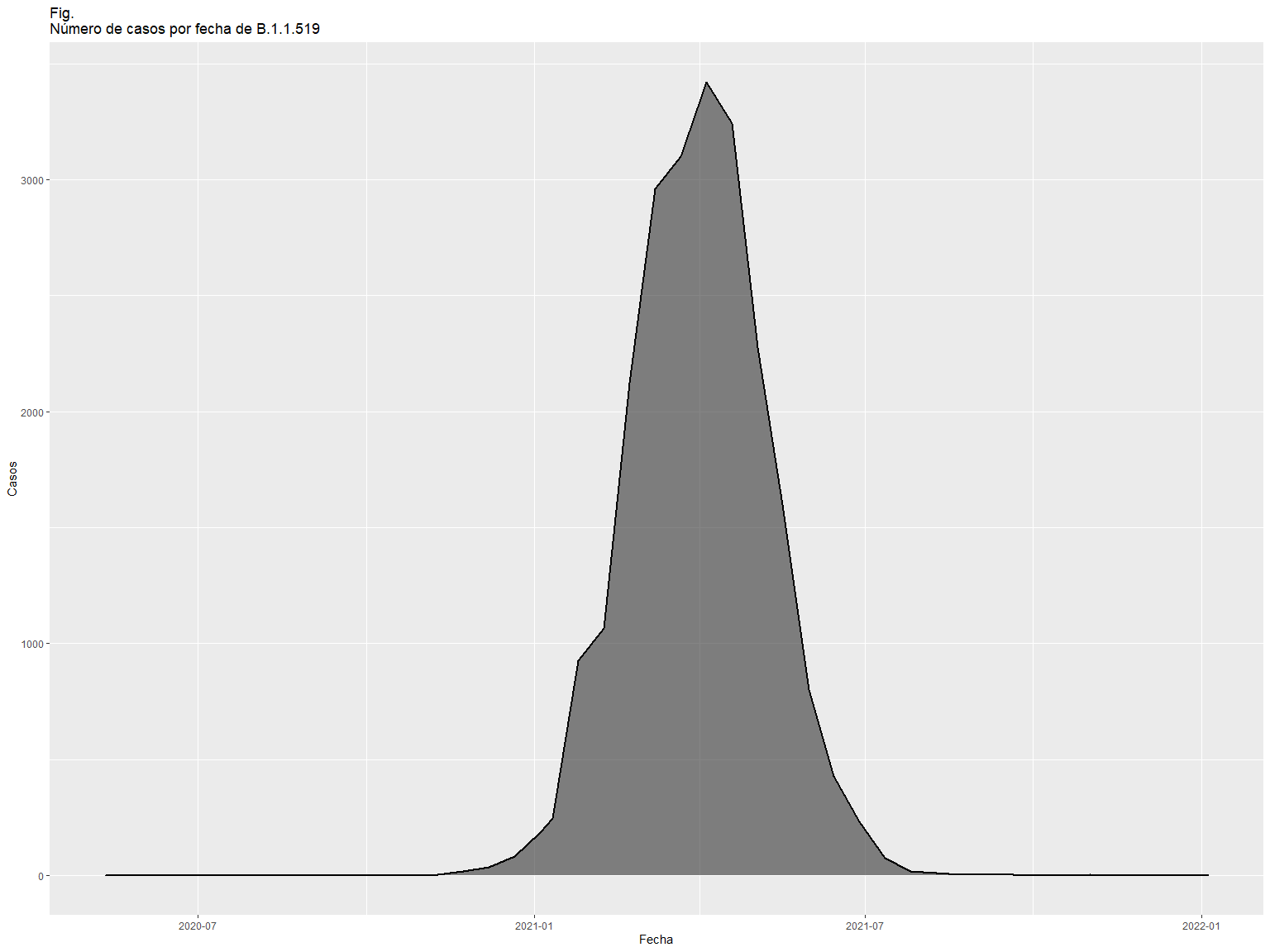
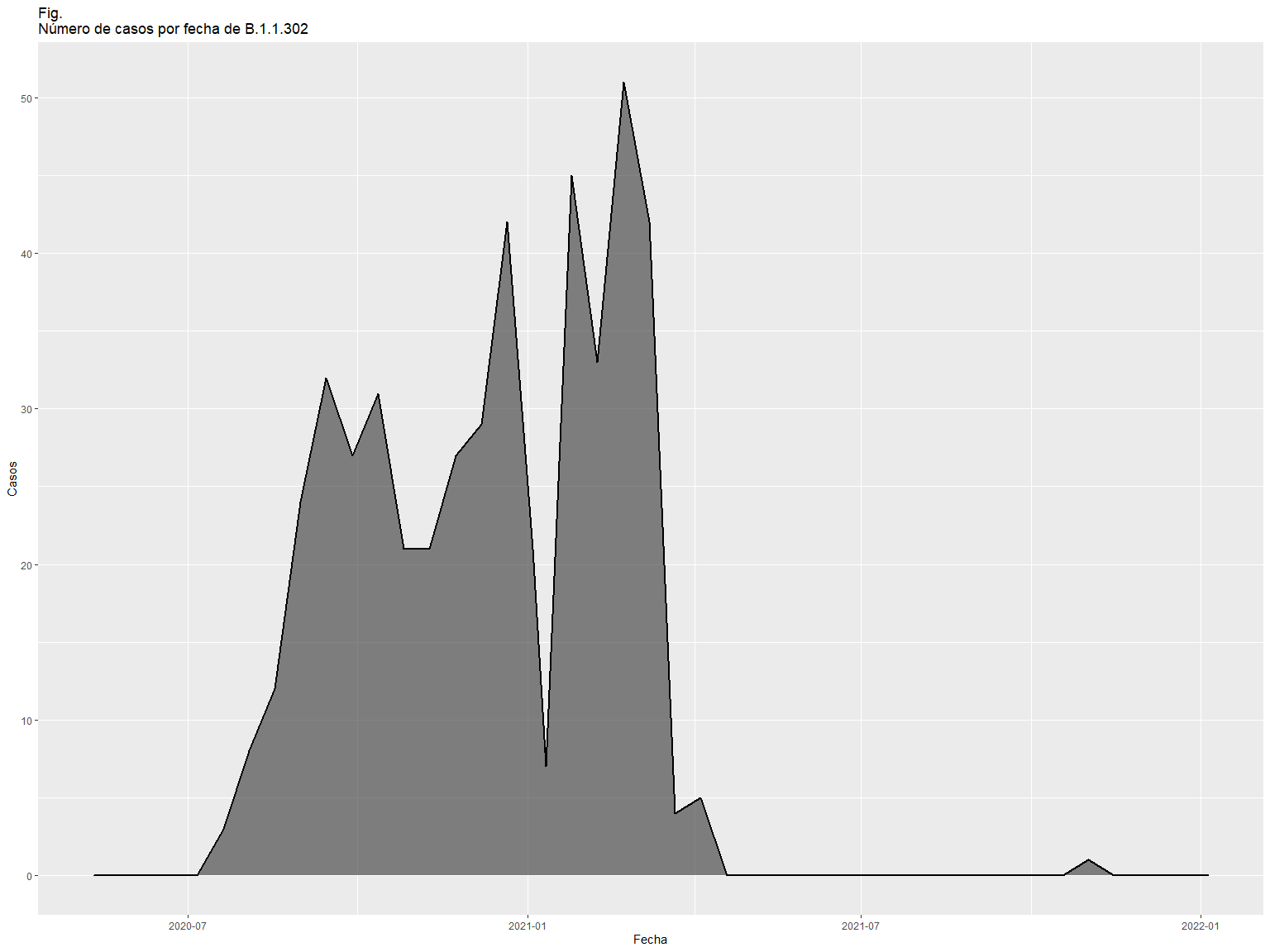
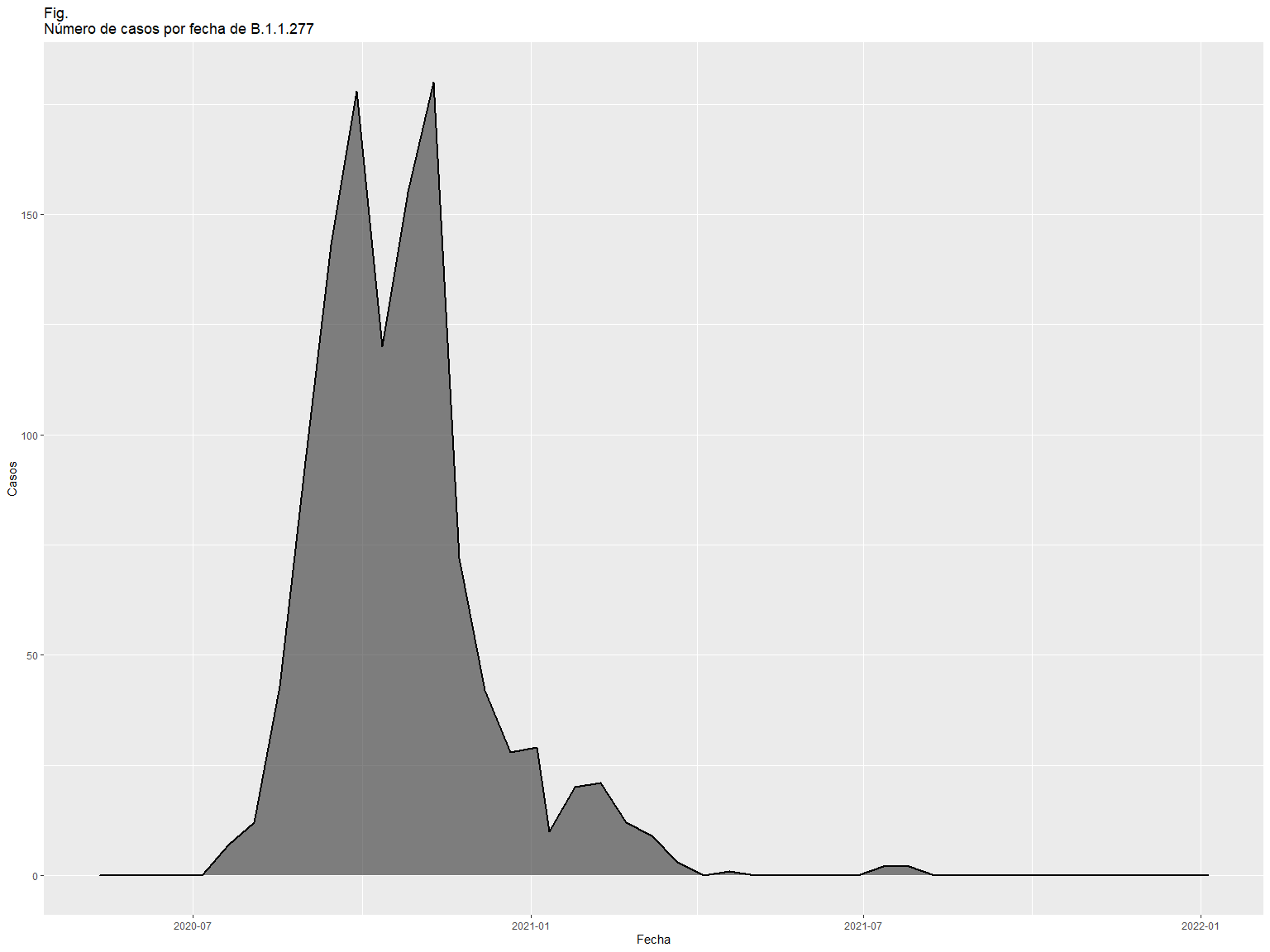
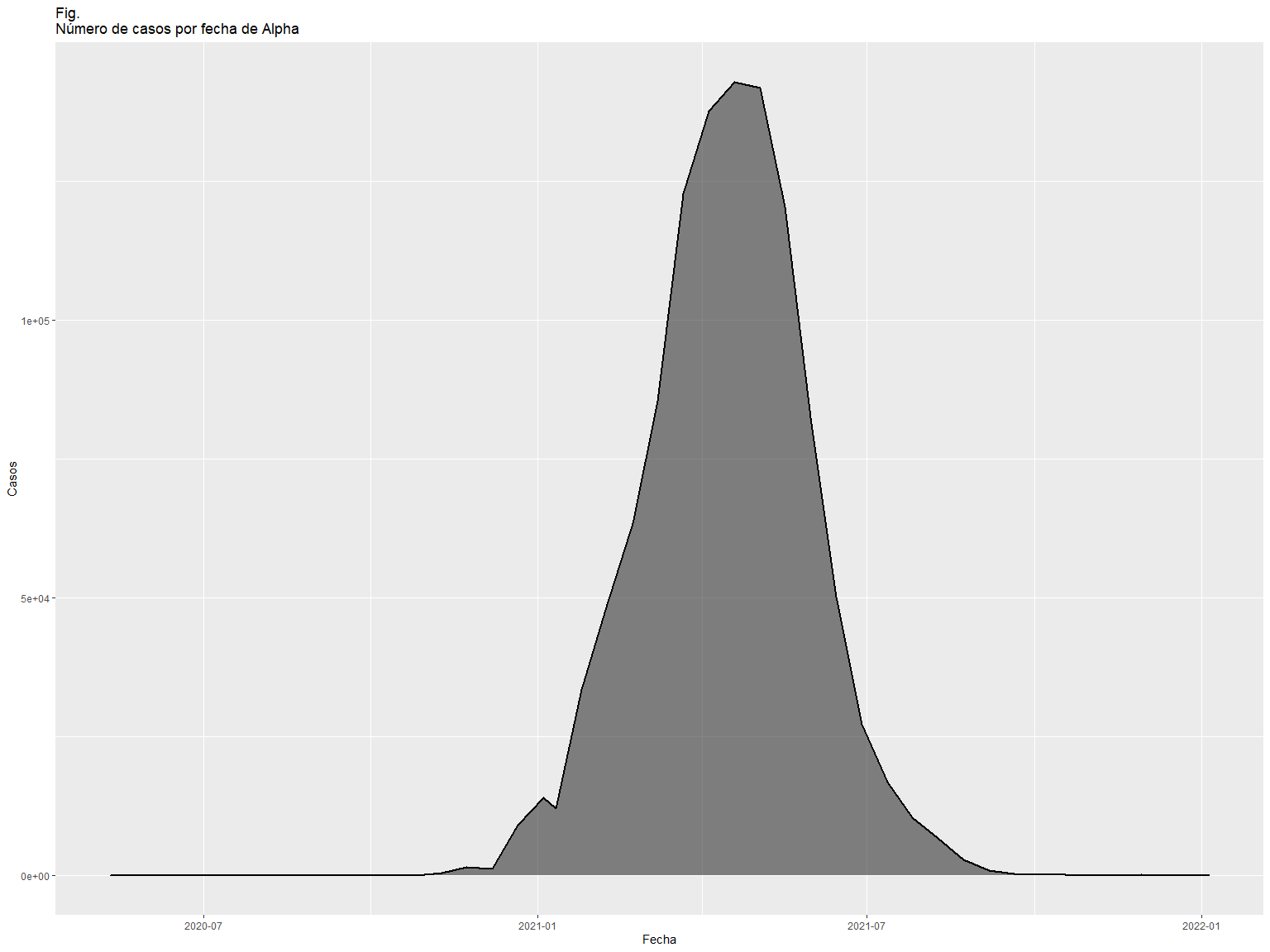
Para lo cual, se realizará una gráfica de los 10 paises con mayor número de casos por cada una de estas cepas.

virus <- unique(DB$variant)  
for (i\_virus in virus){  
 data\_nueva<-group\_by(DB[which(DB$variant == i\_virus),],   
 location) %>%  
 summarise(num\_sequences = sum(num\_sequences))   
   
 data\_nueva <- arrange(data\_nueva, desc(data\_nueva$num\_sequences))  
 data\_nueva <- data.frame(head(data\_nueva,n=10))  
   
 # Grafico  
 print(  
 ggplot(data=data\_nueva,   
 aes(x = num\_sequences,   
 y = reorder(location, num\_sequences))) +  
 geom\_bar(stat="identity")+   
 ggtitle(paste("Fig. \nTOP 10 de paises con mayor número de casos de",i\_virus))+  
 xlab("Número de casos")+  
 ylab("Locación"))  
}

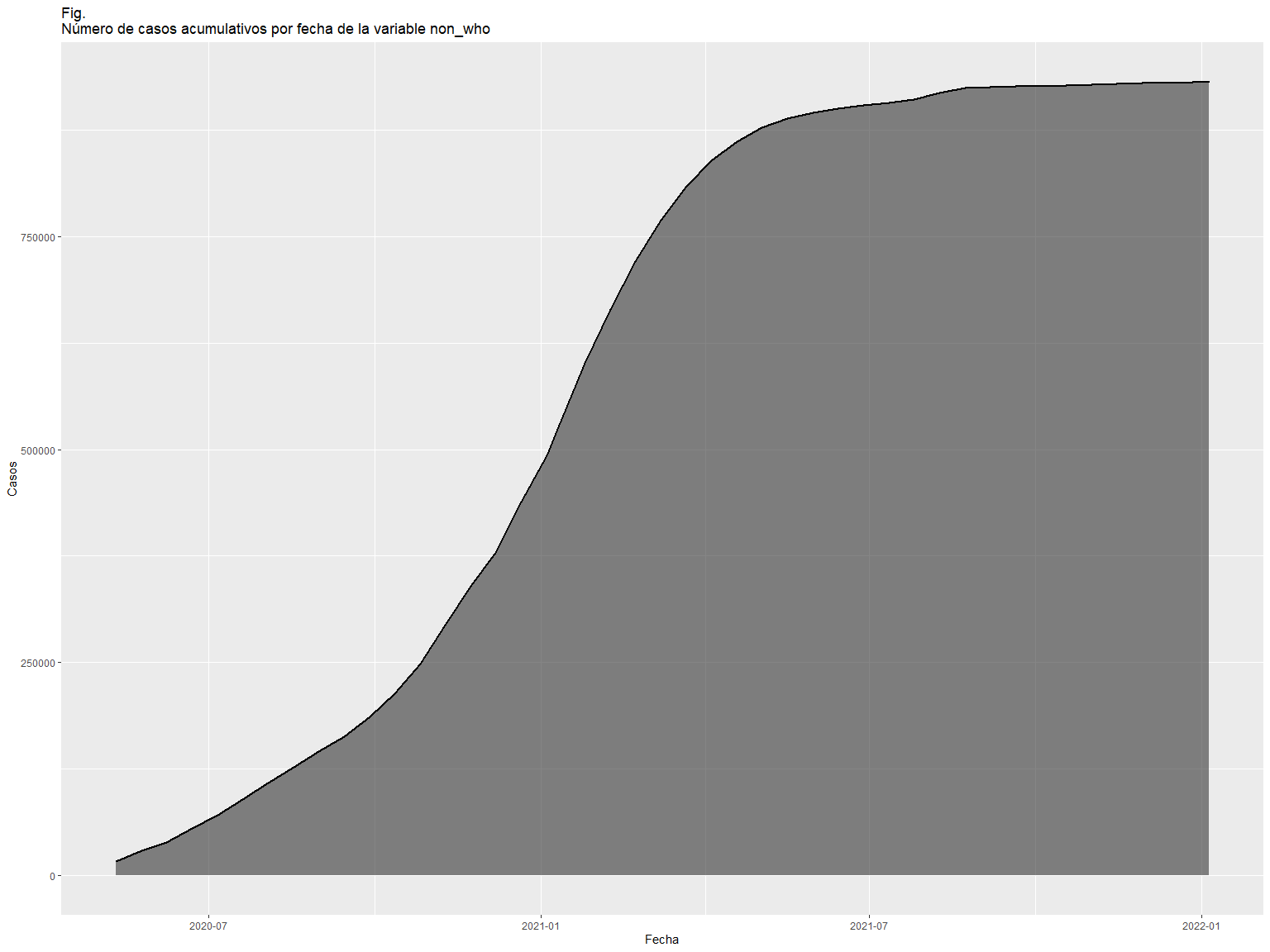
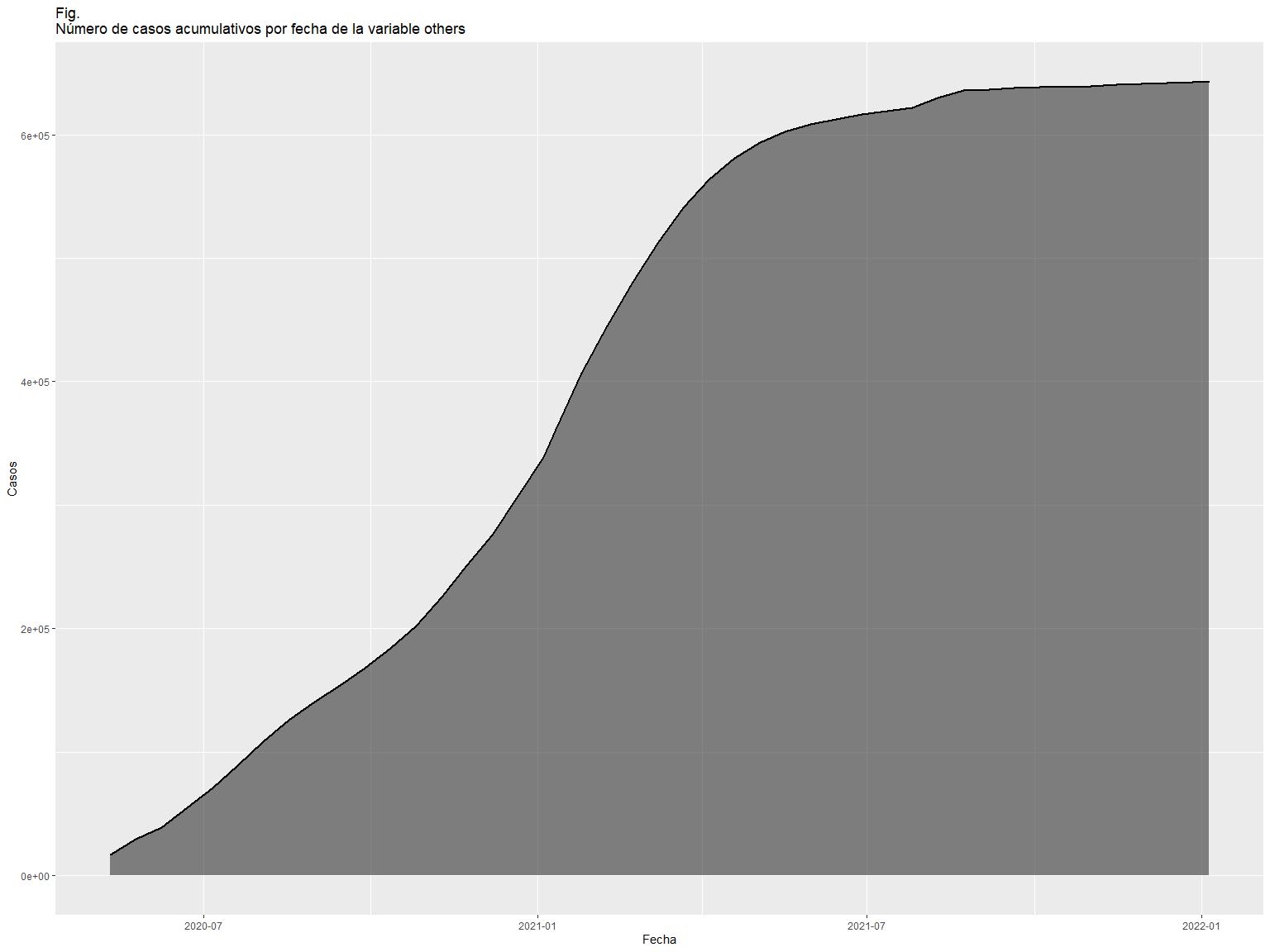
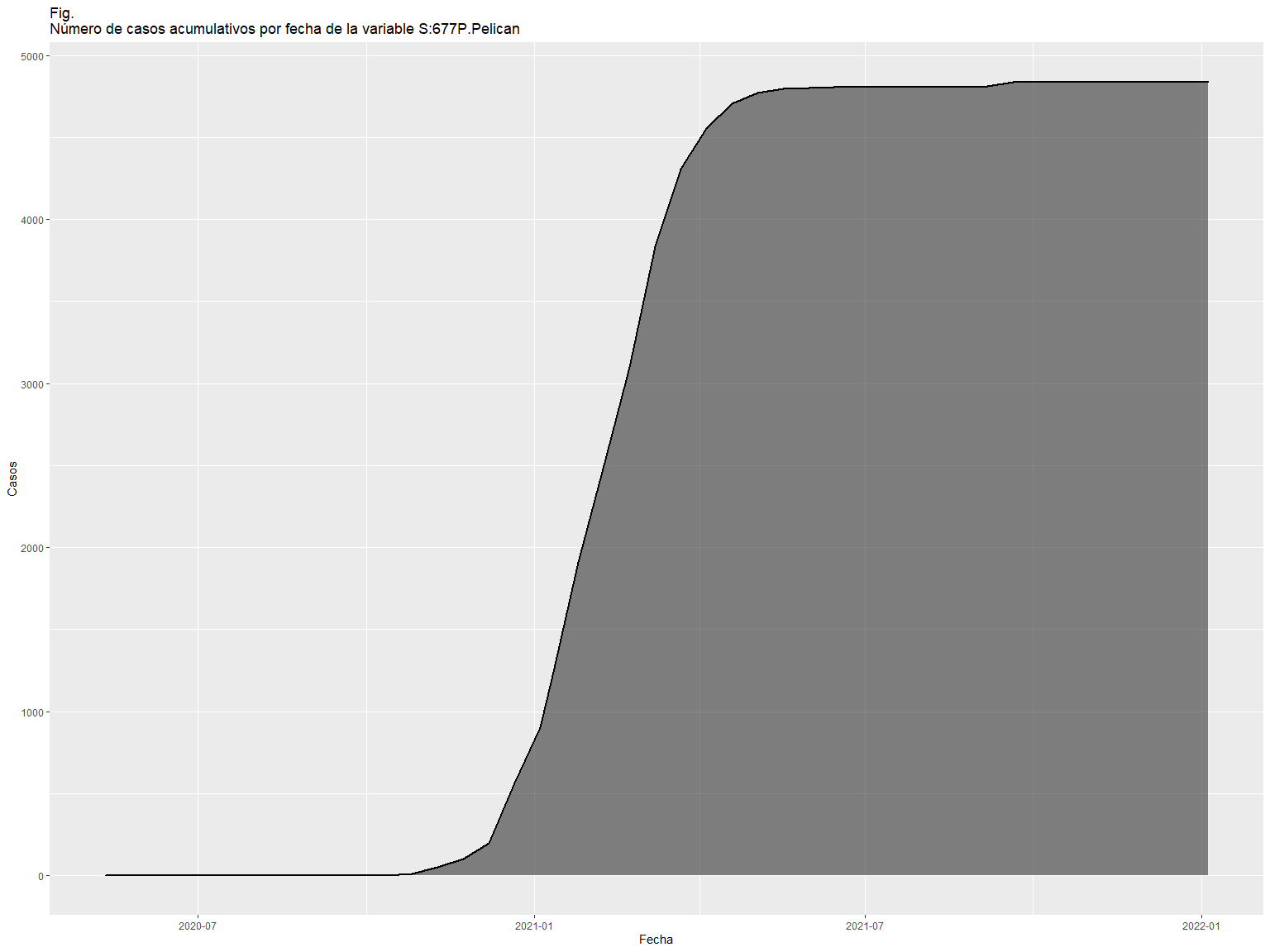
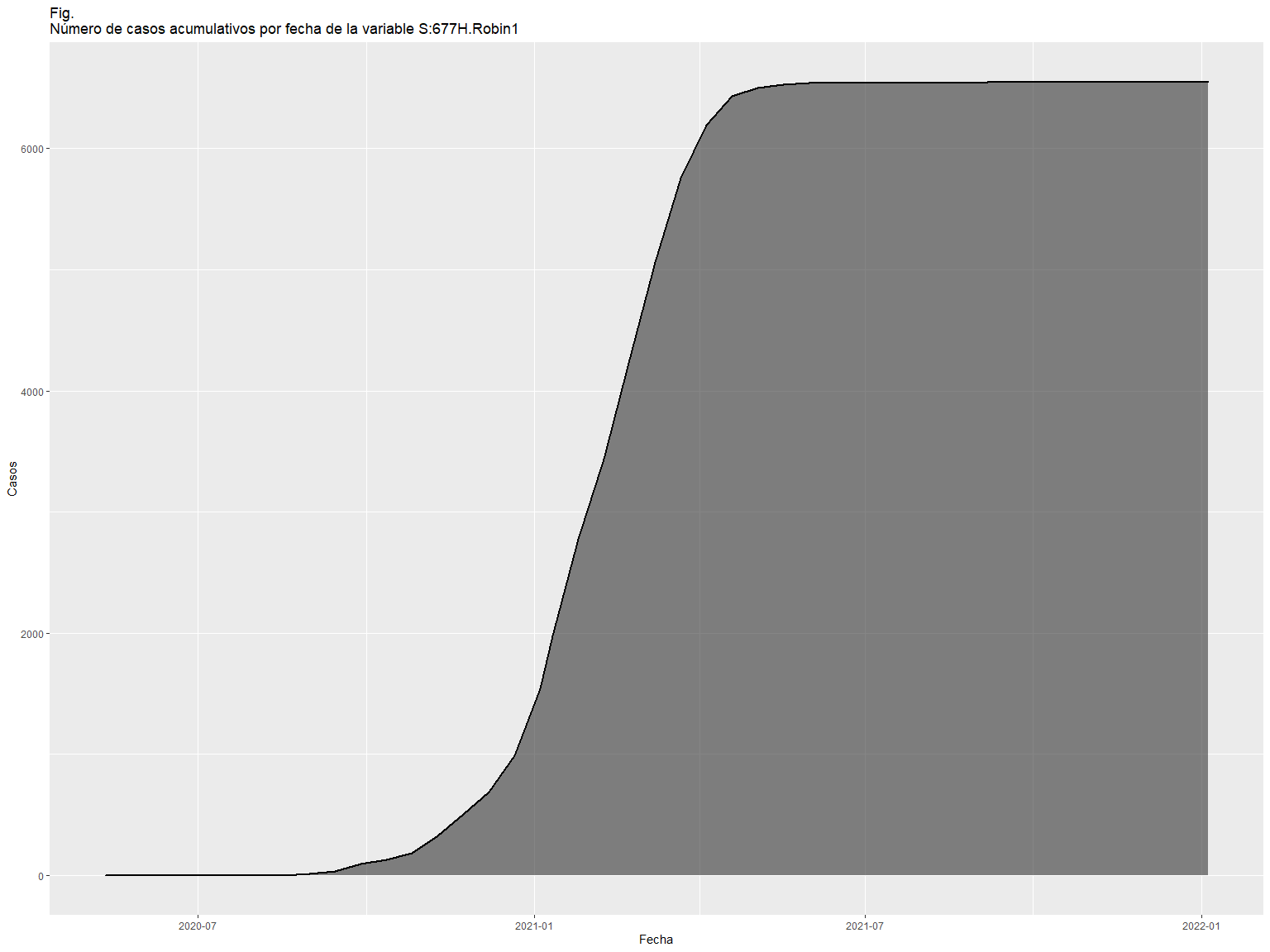
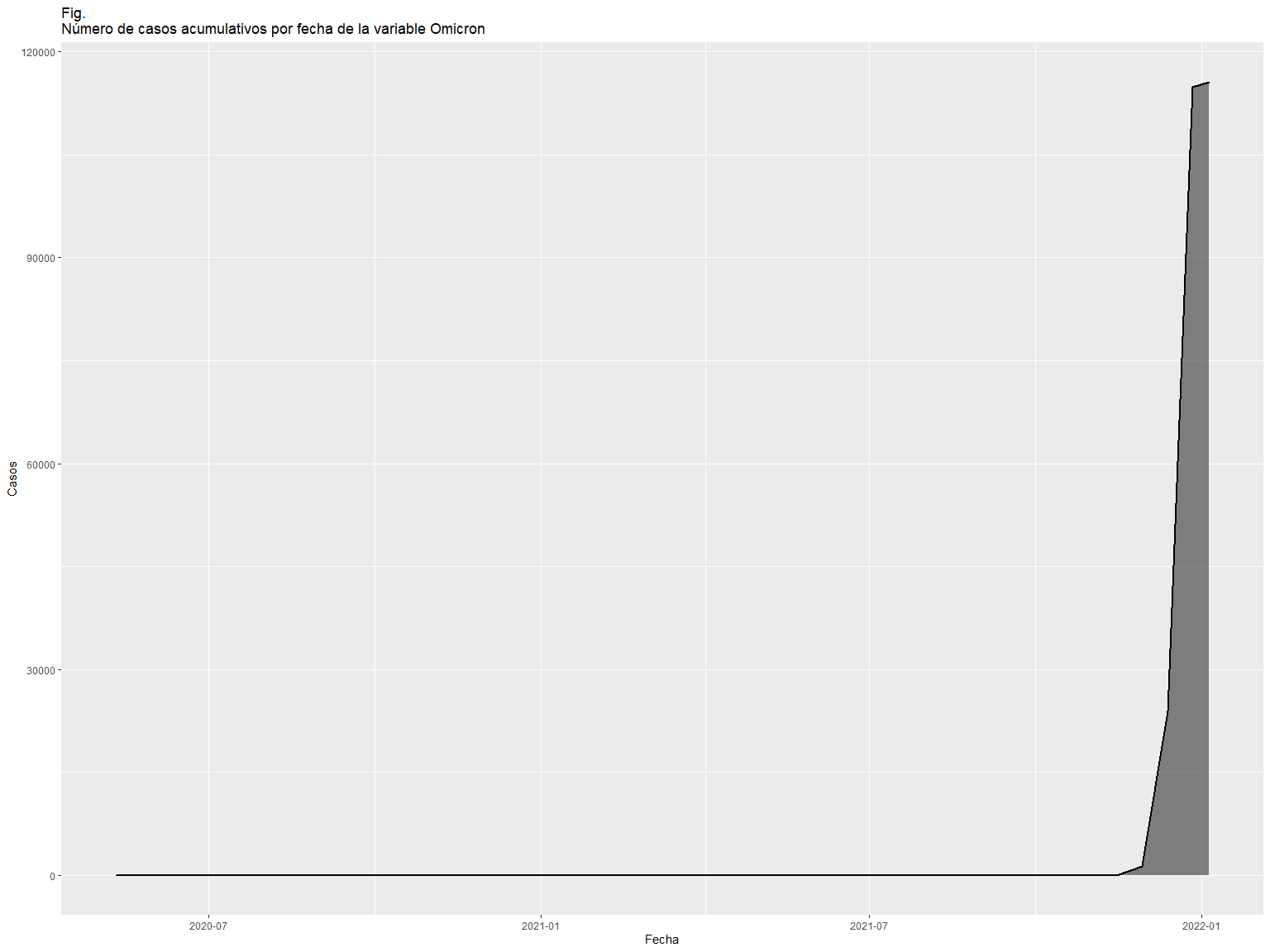
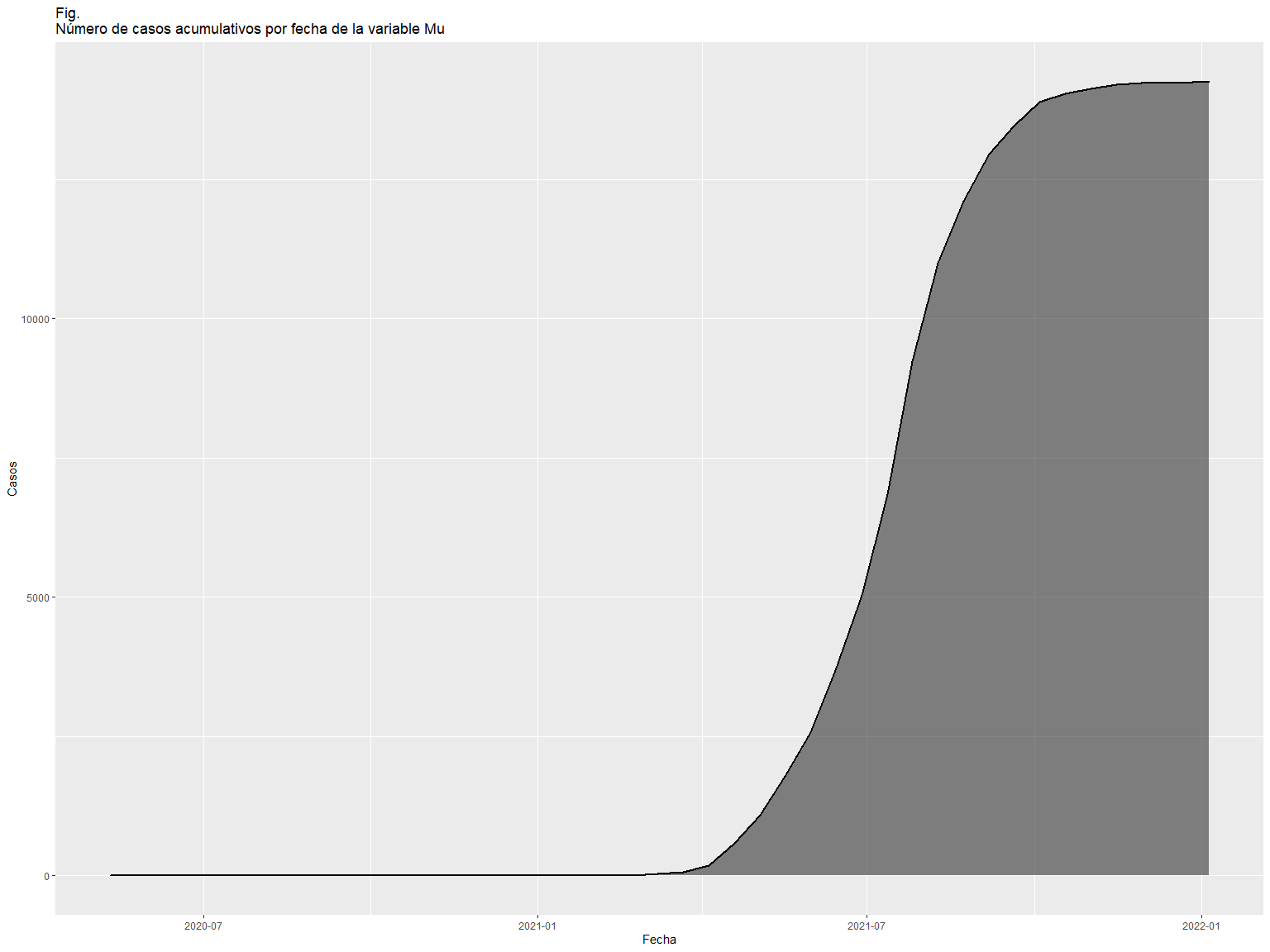
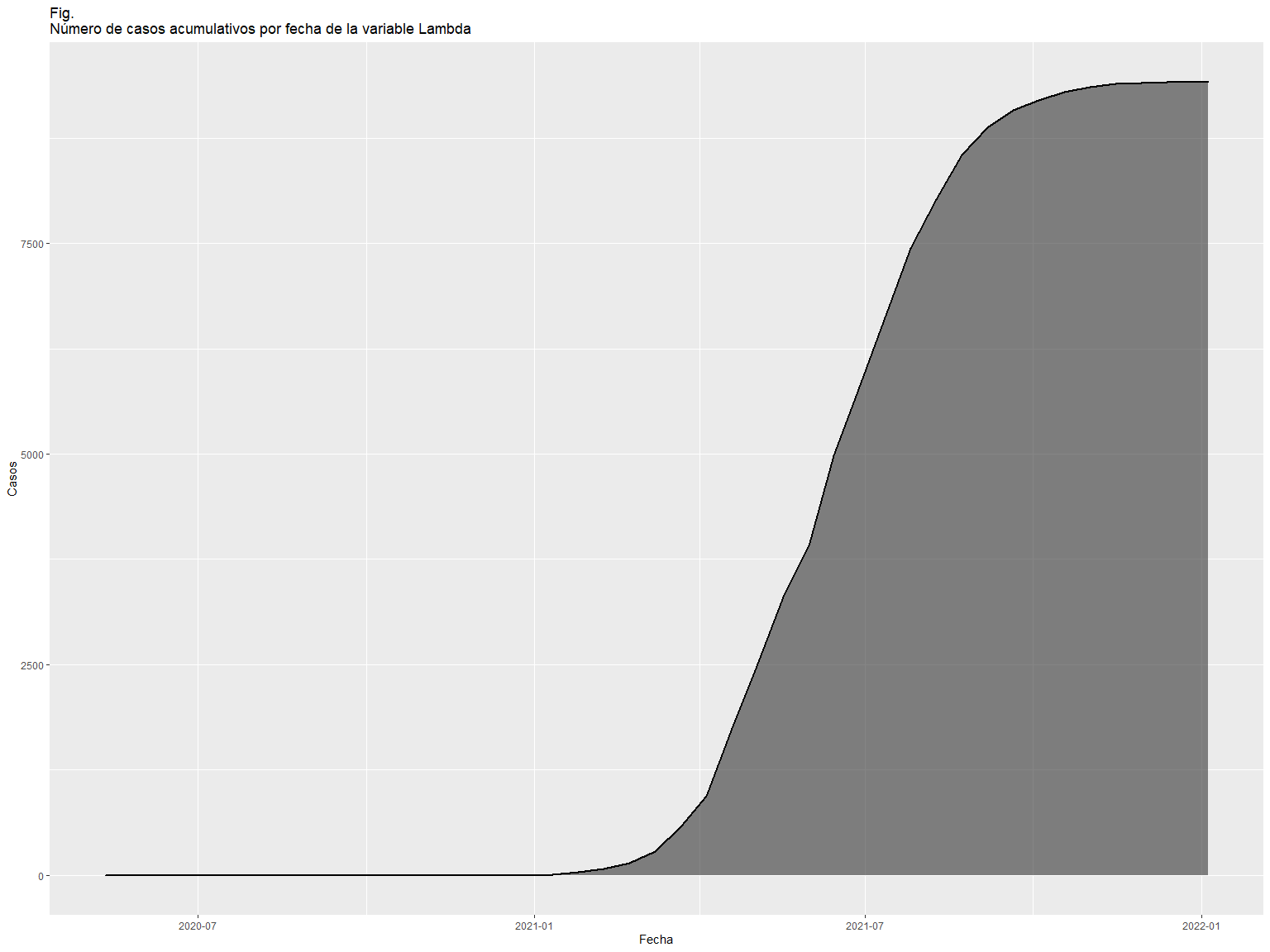
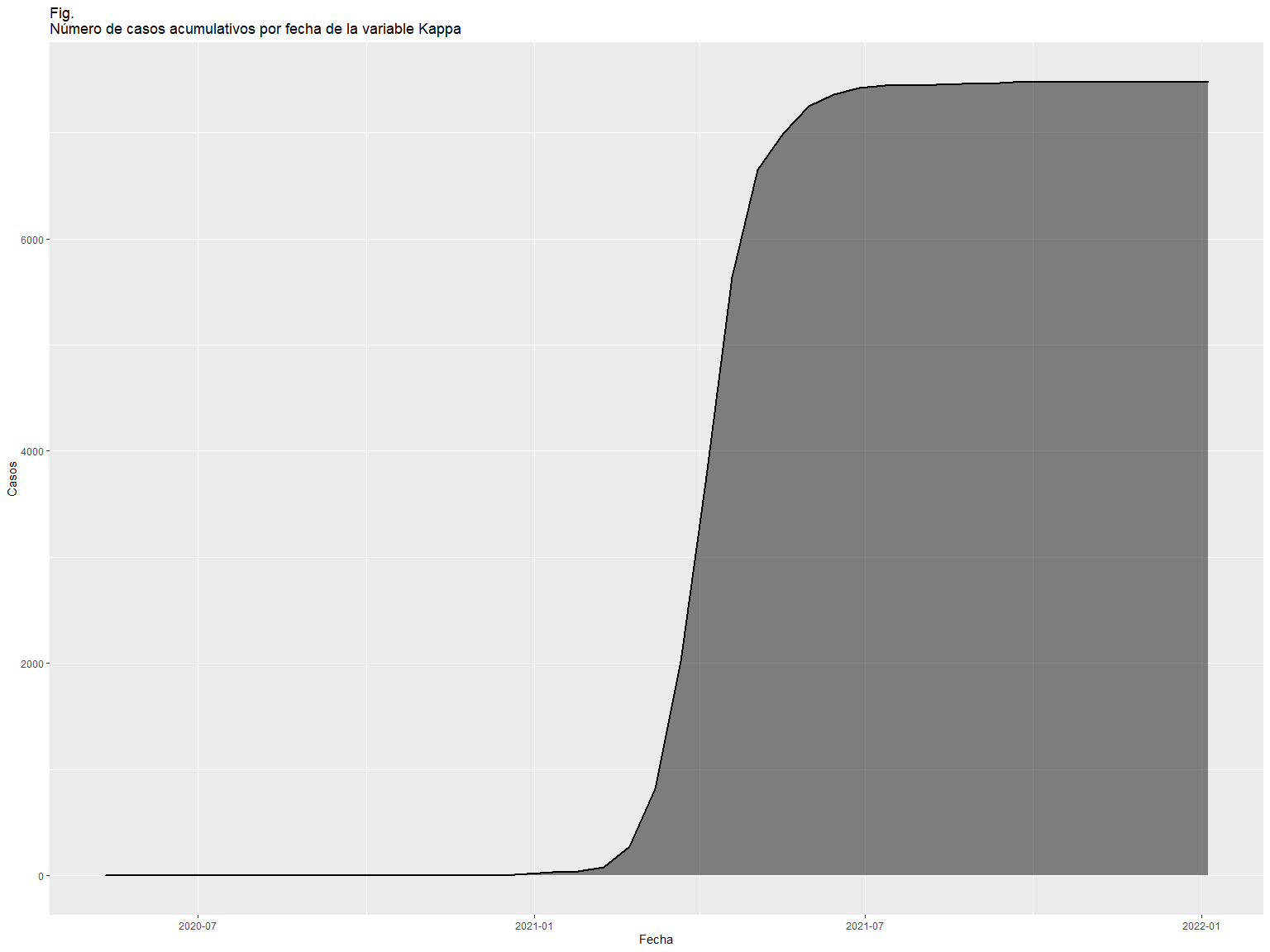
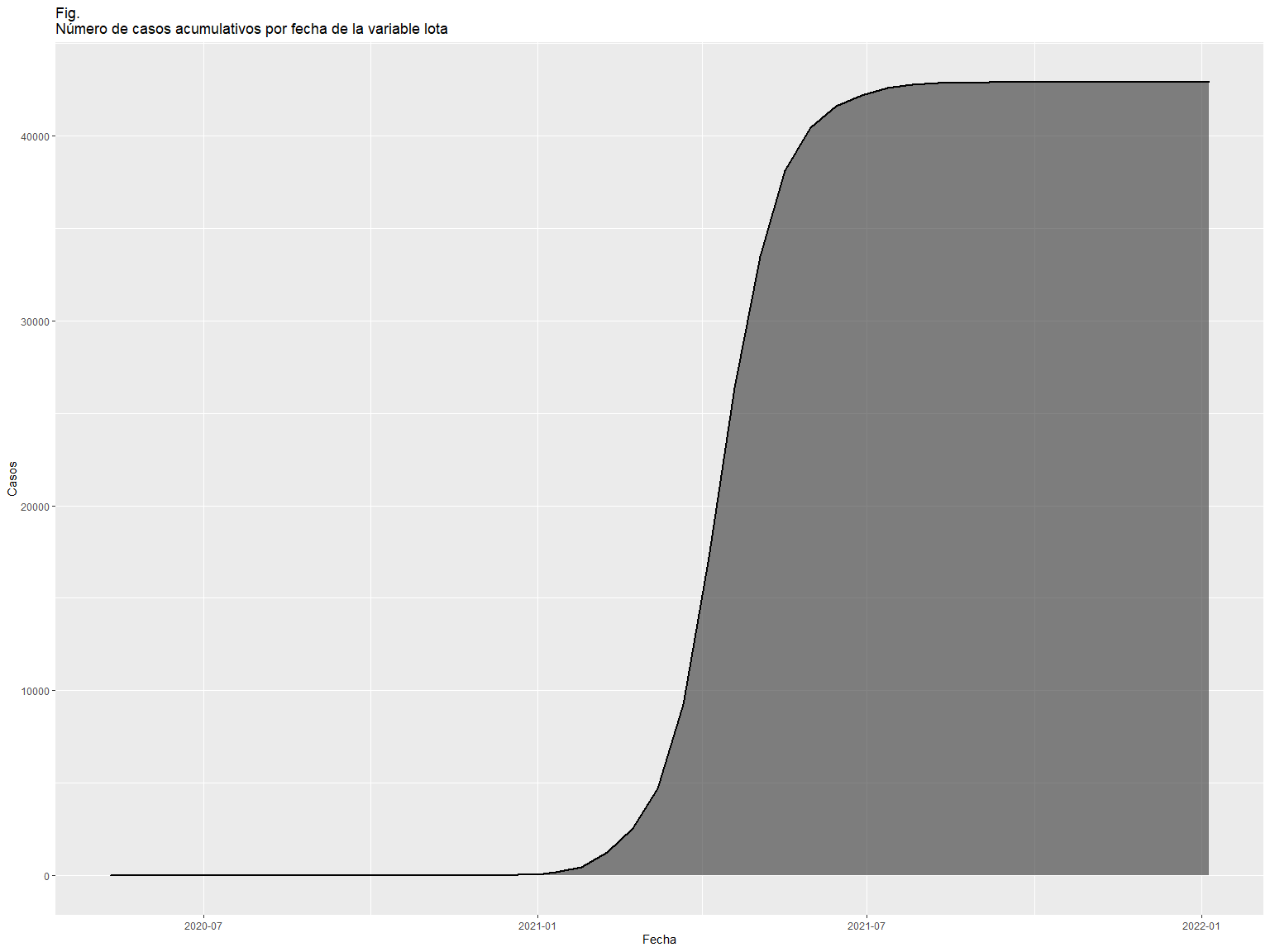
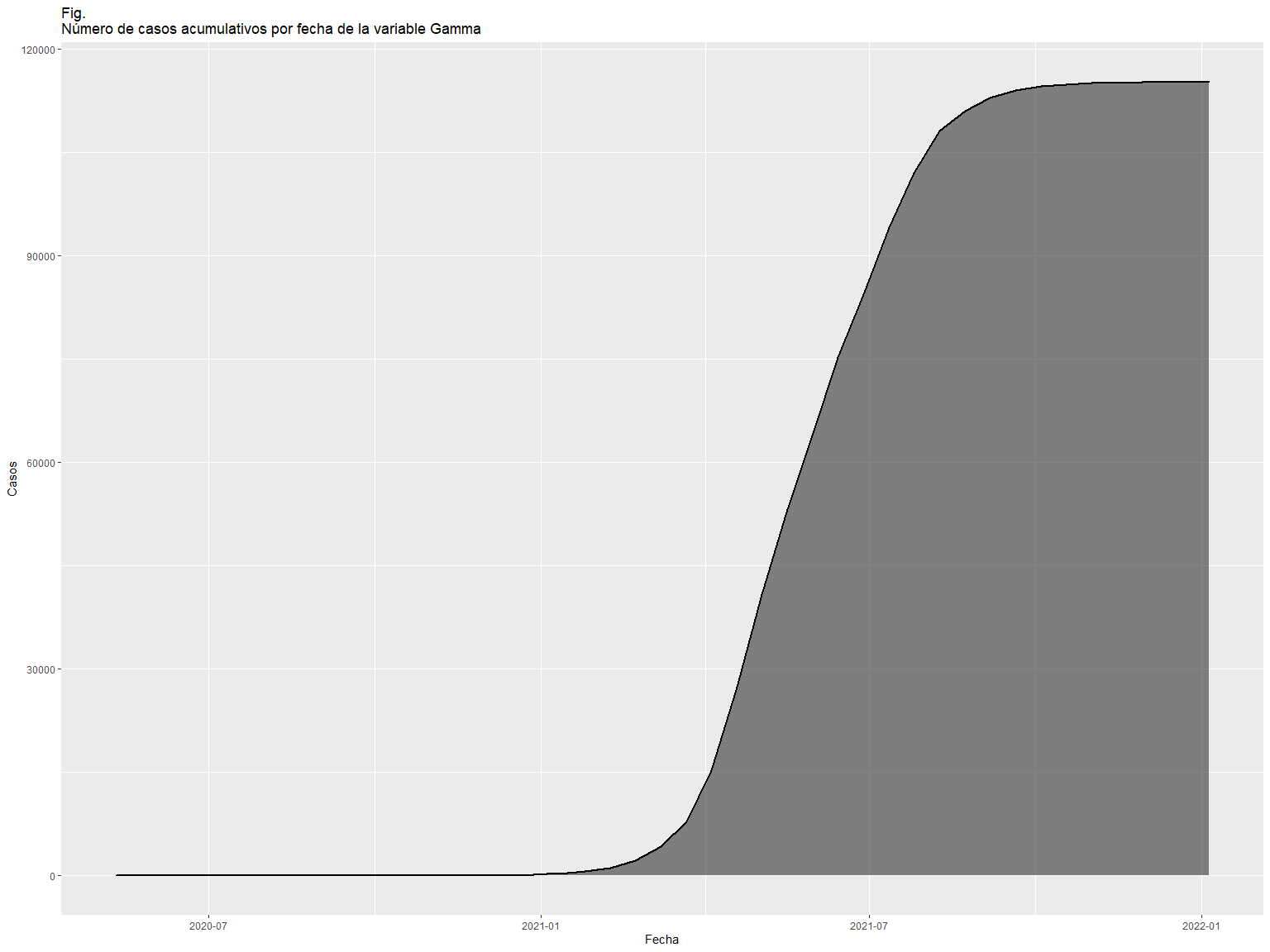
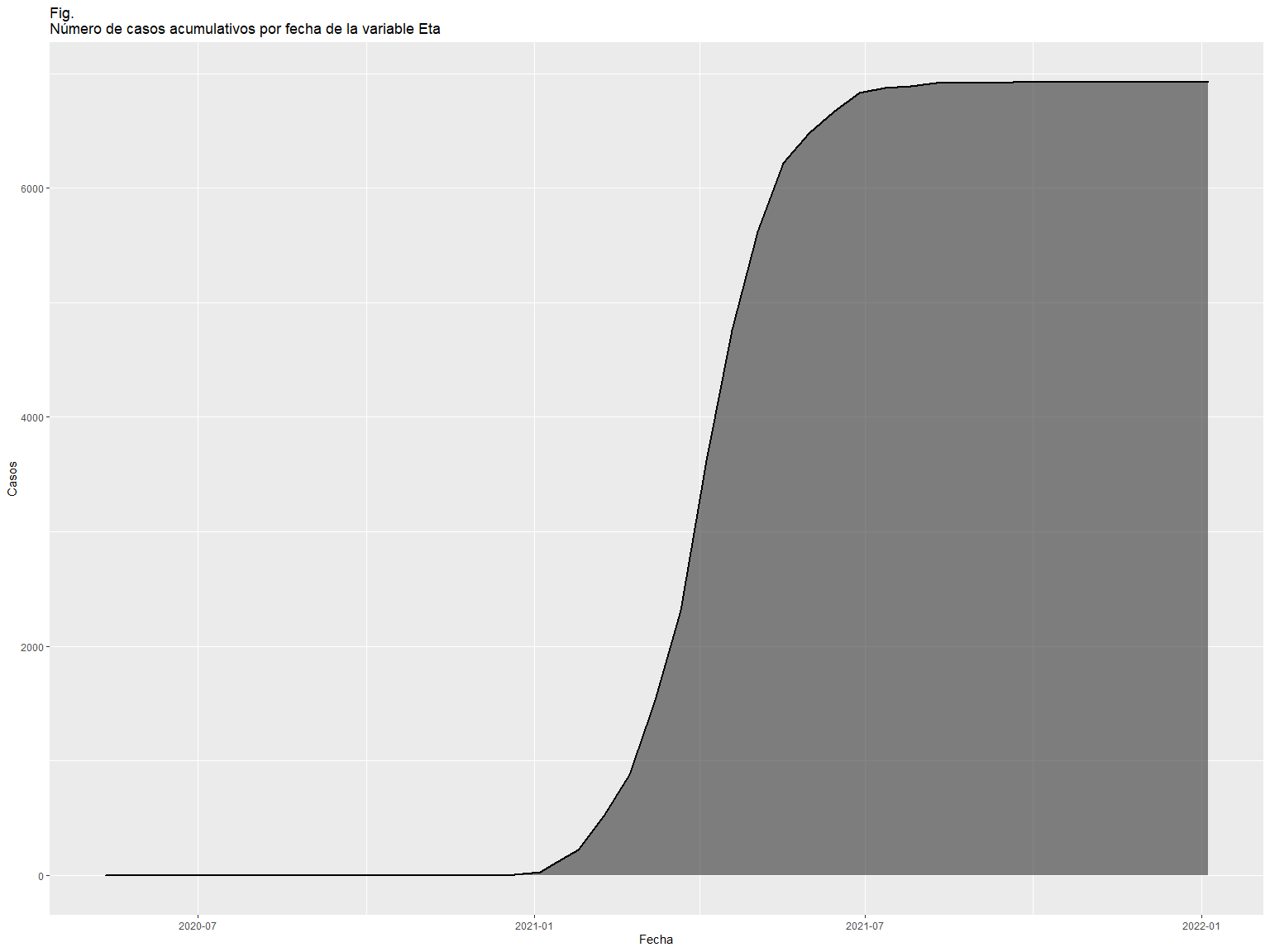
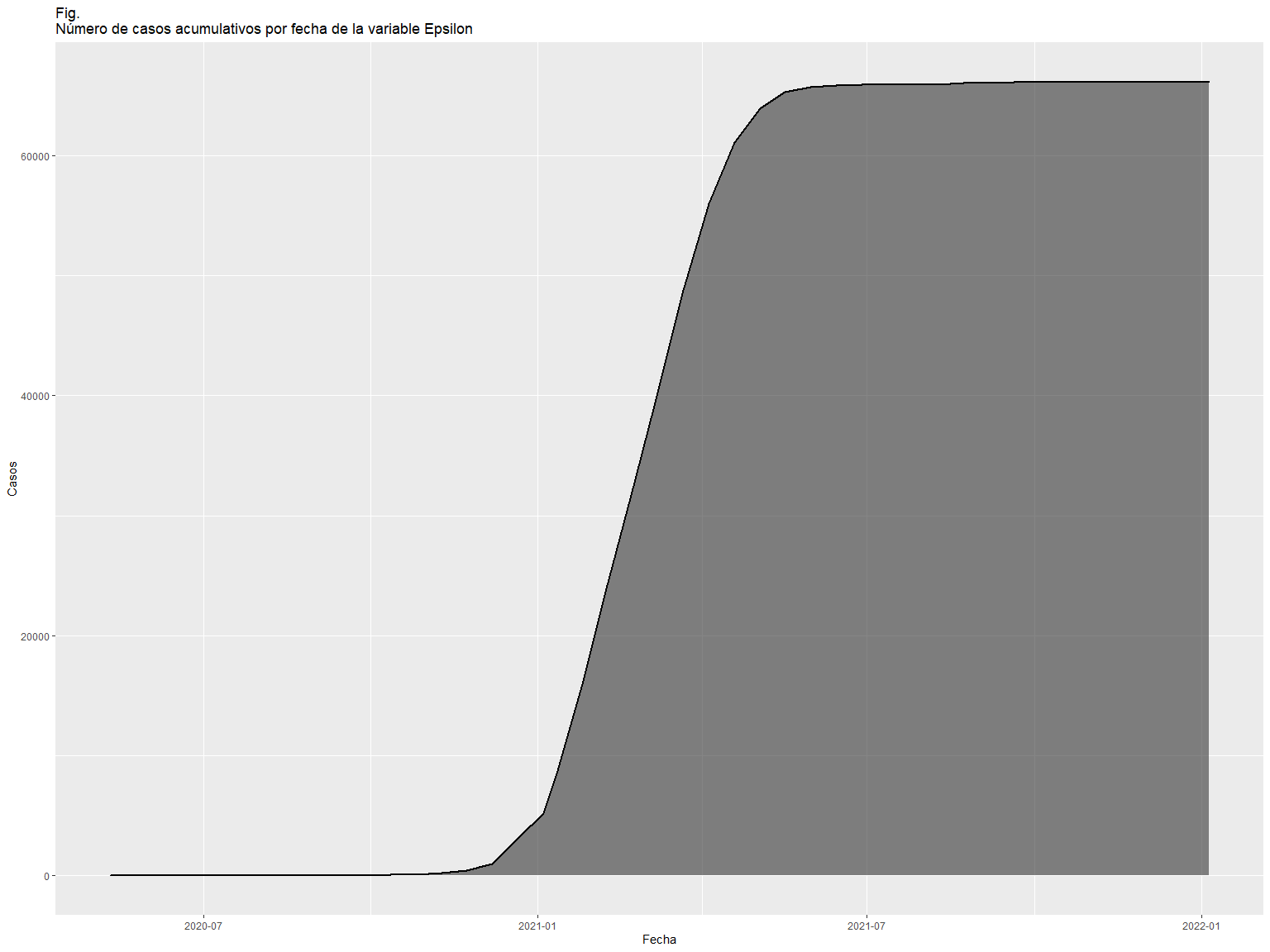
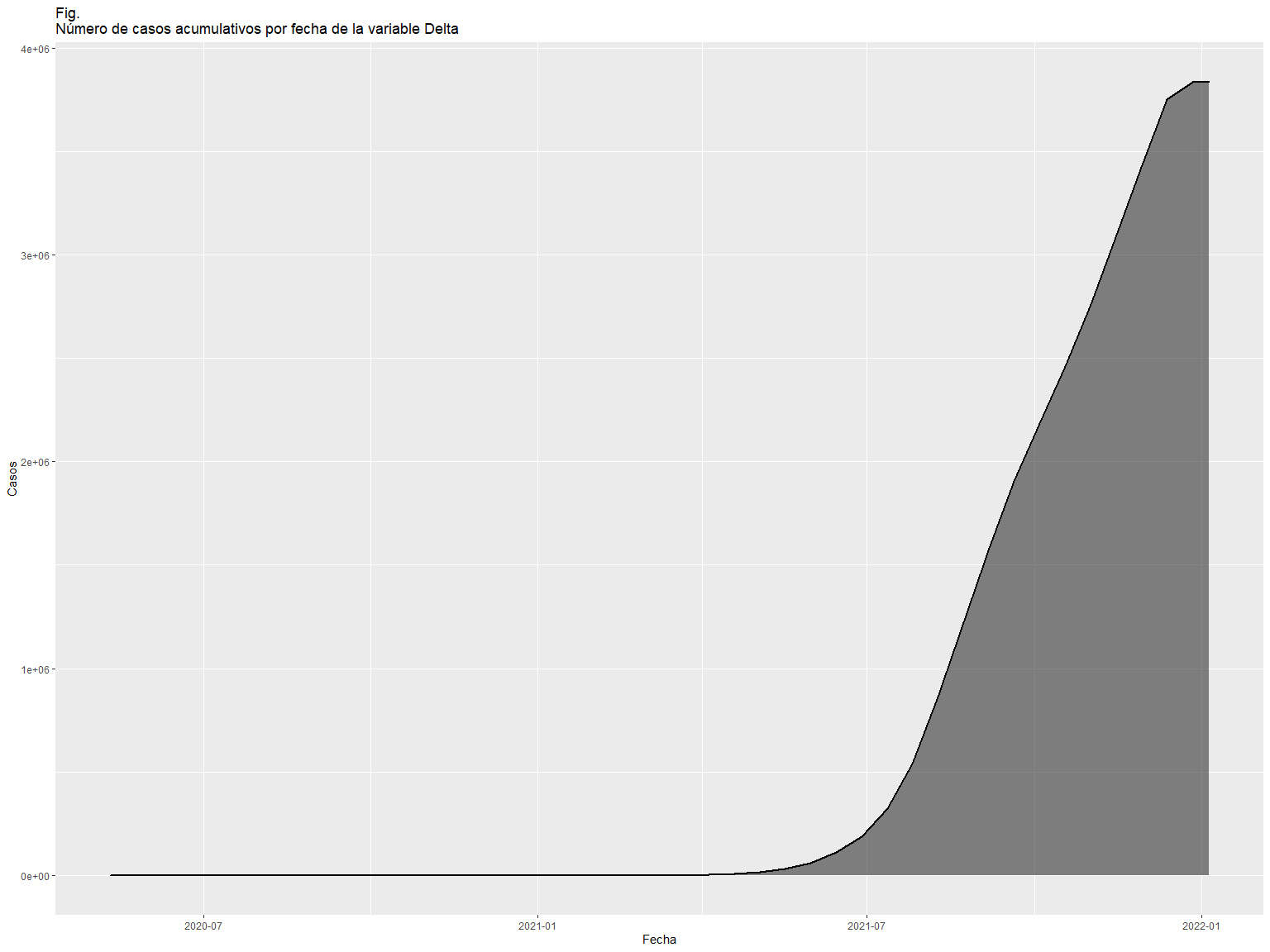
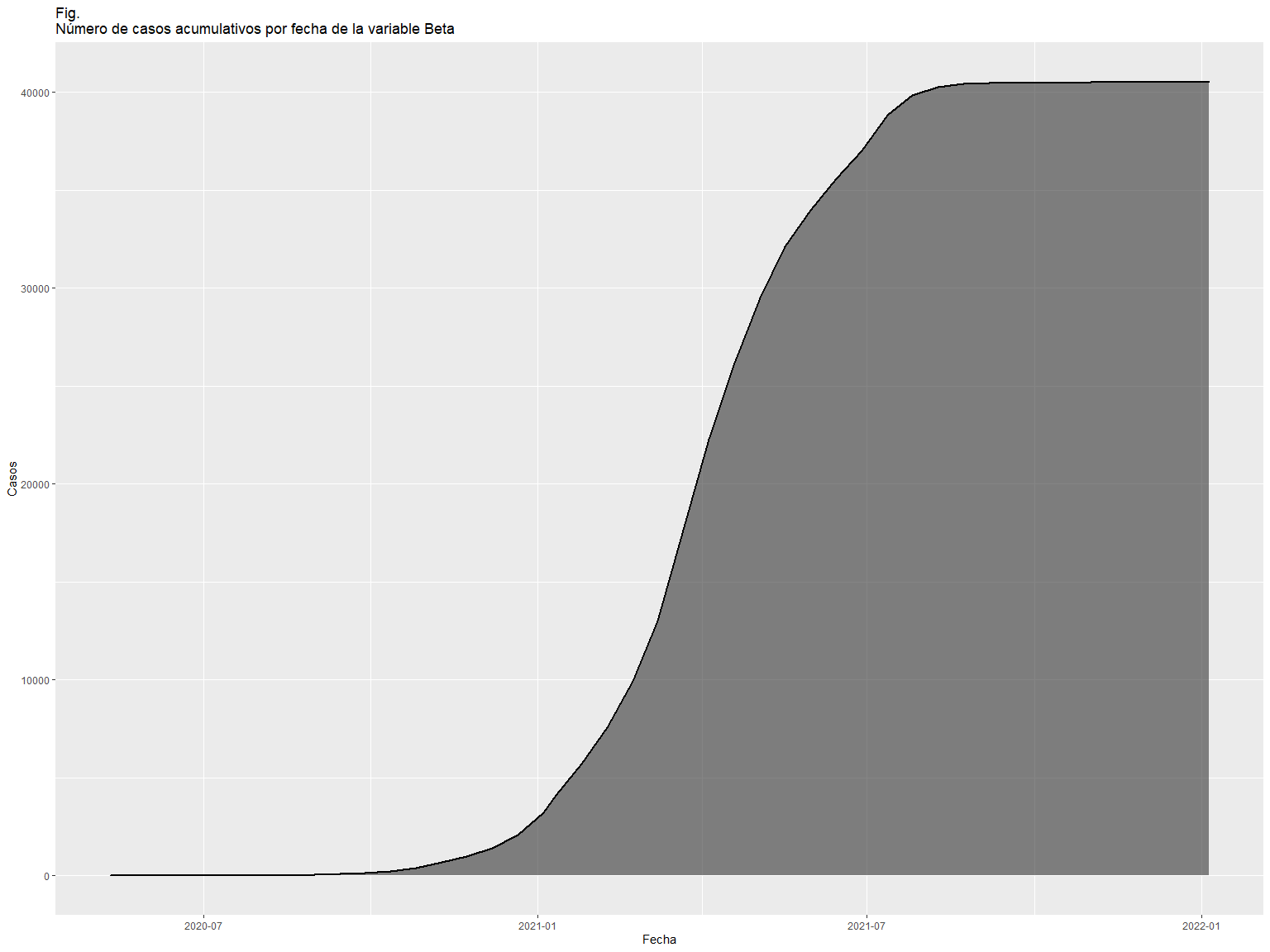
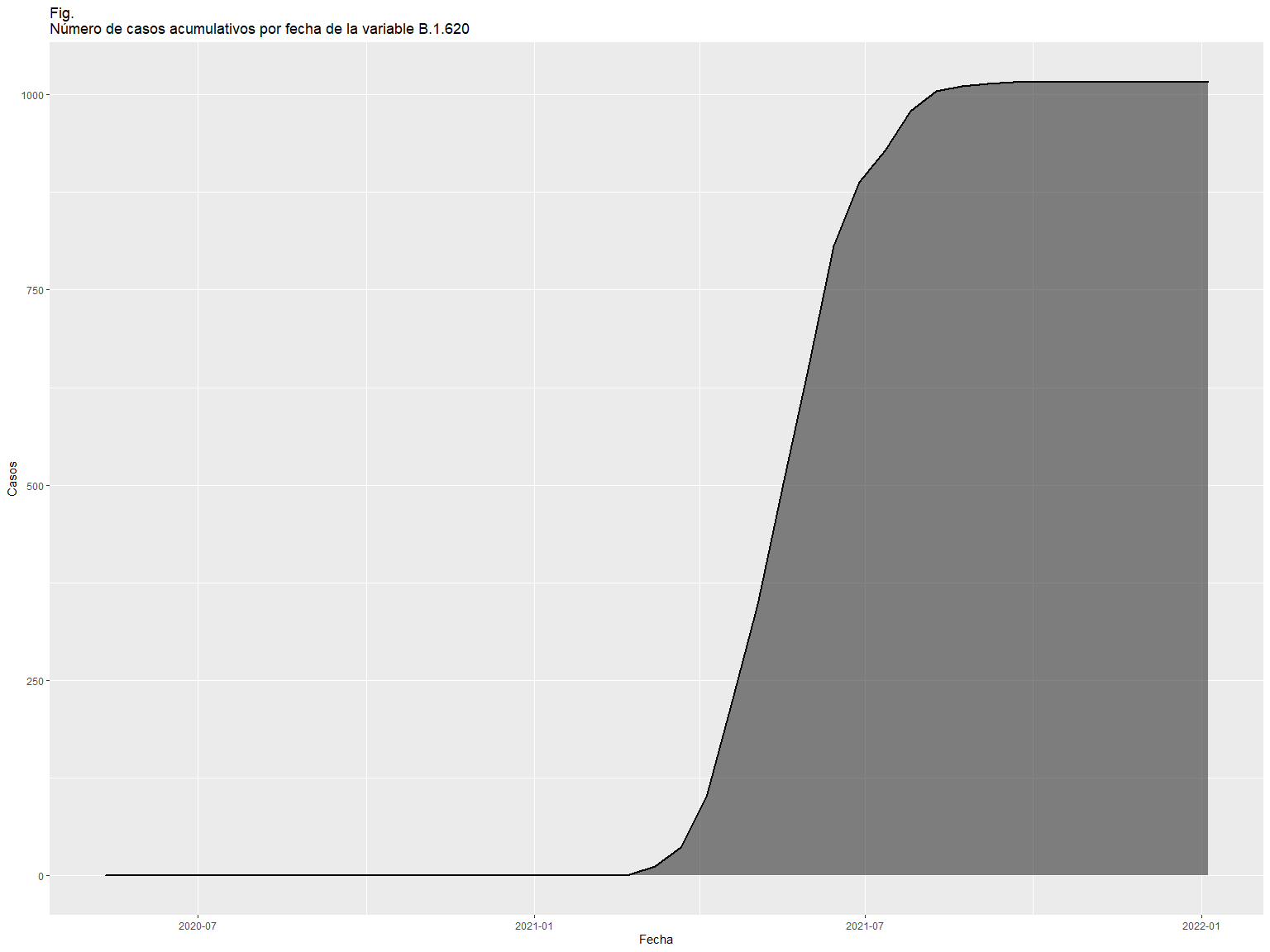
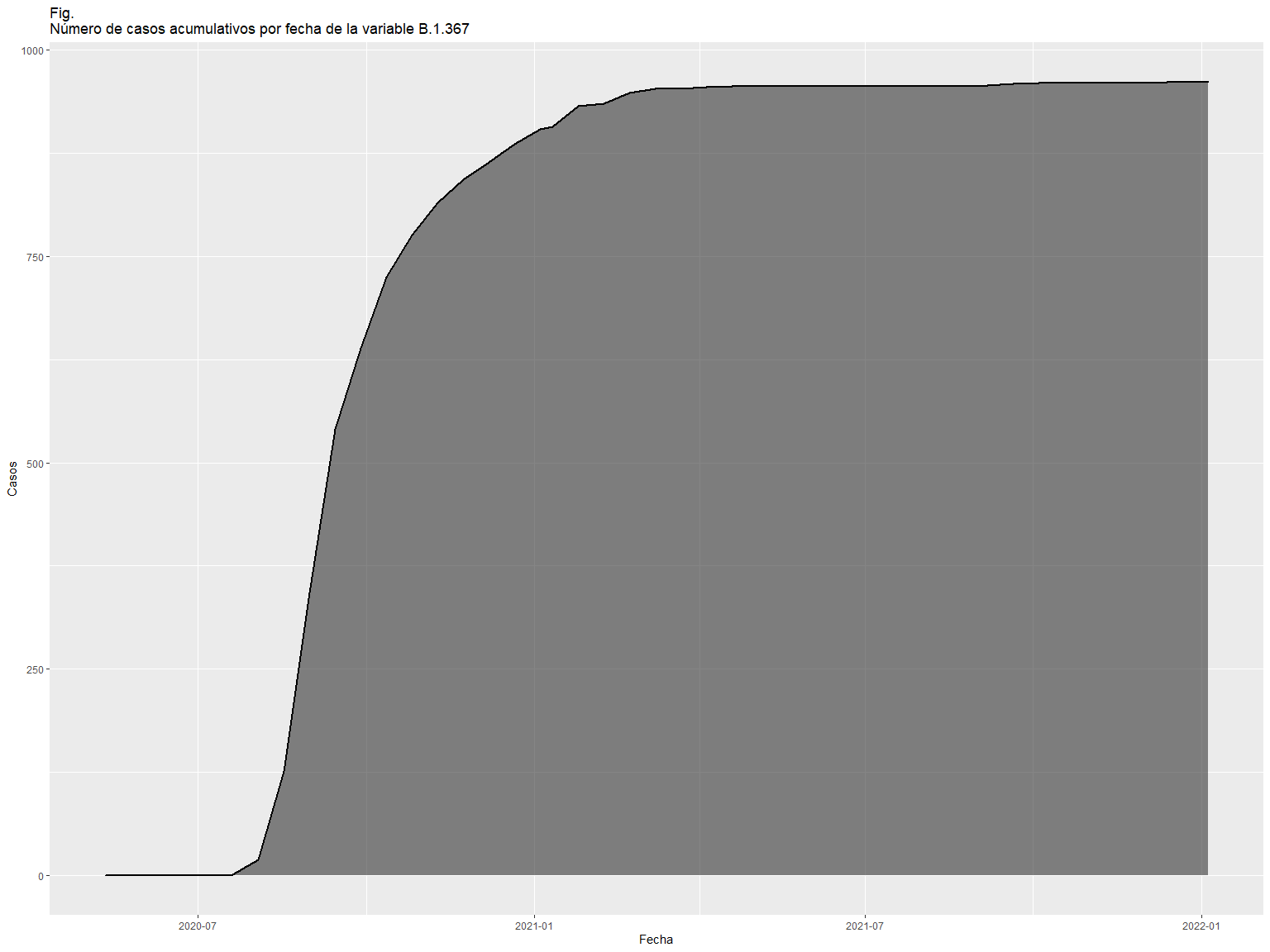
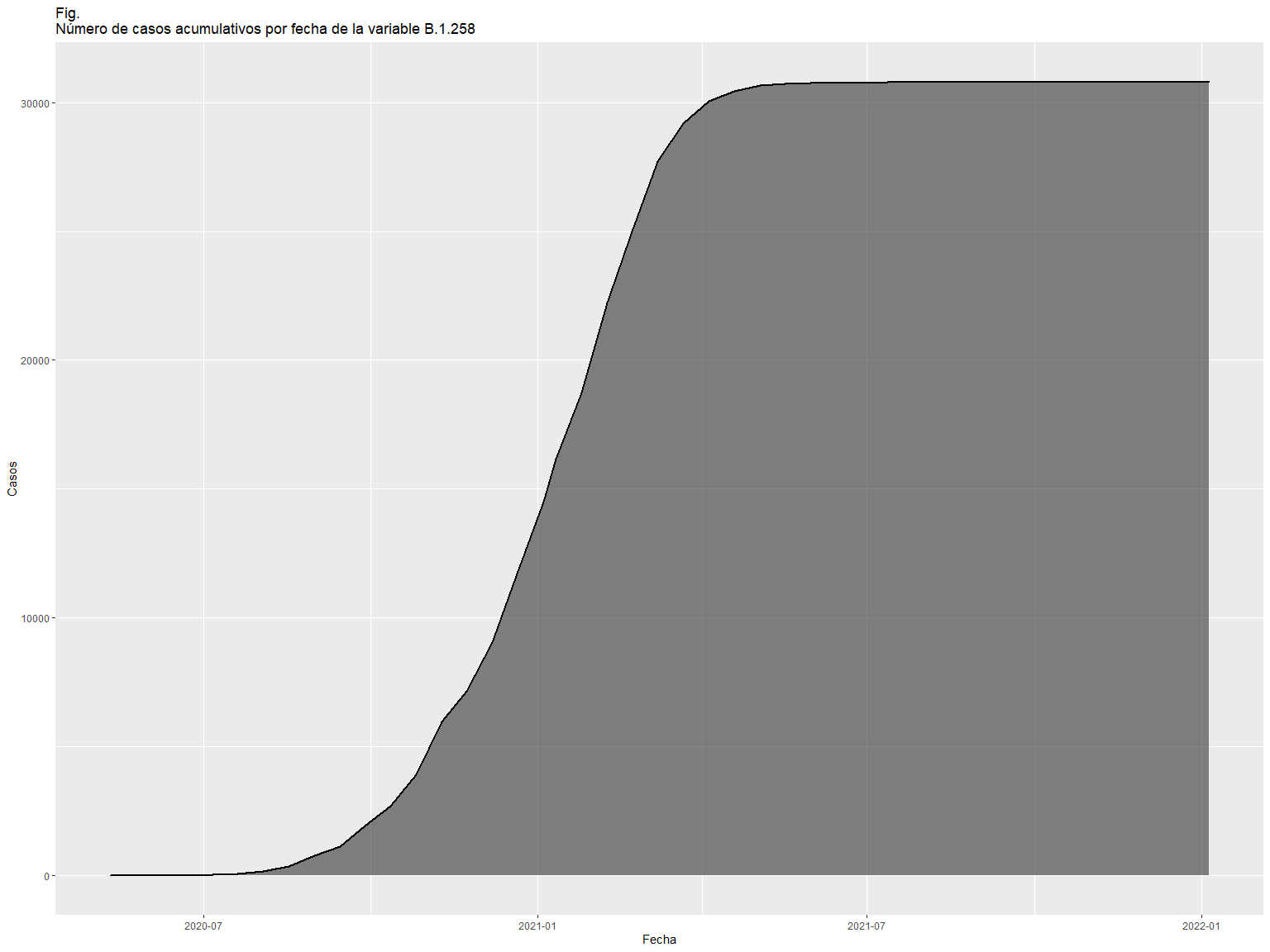
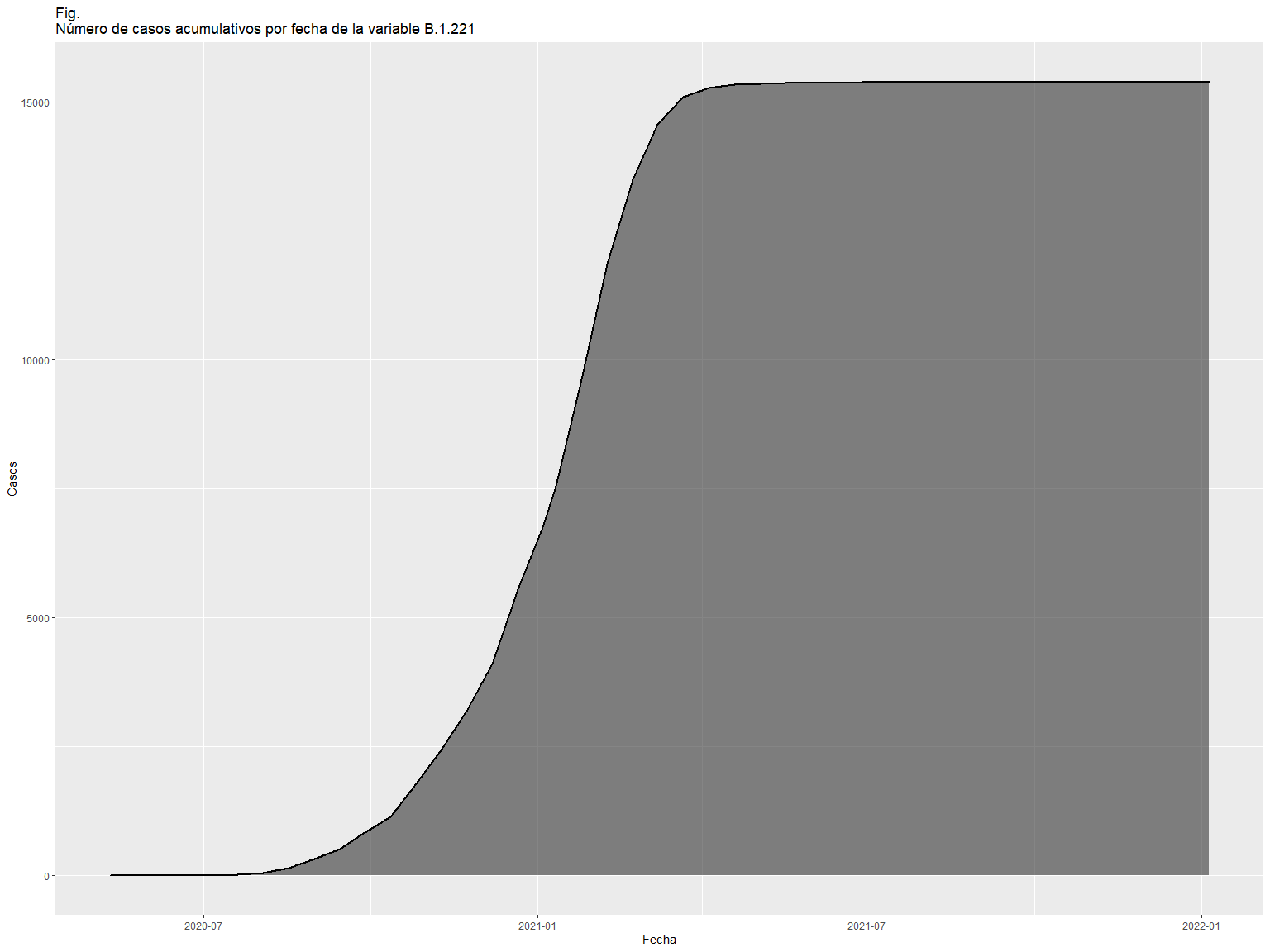
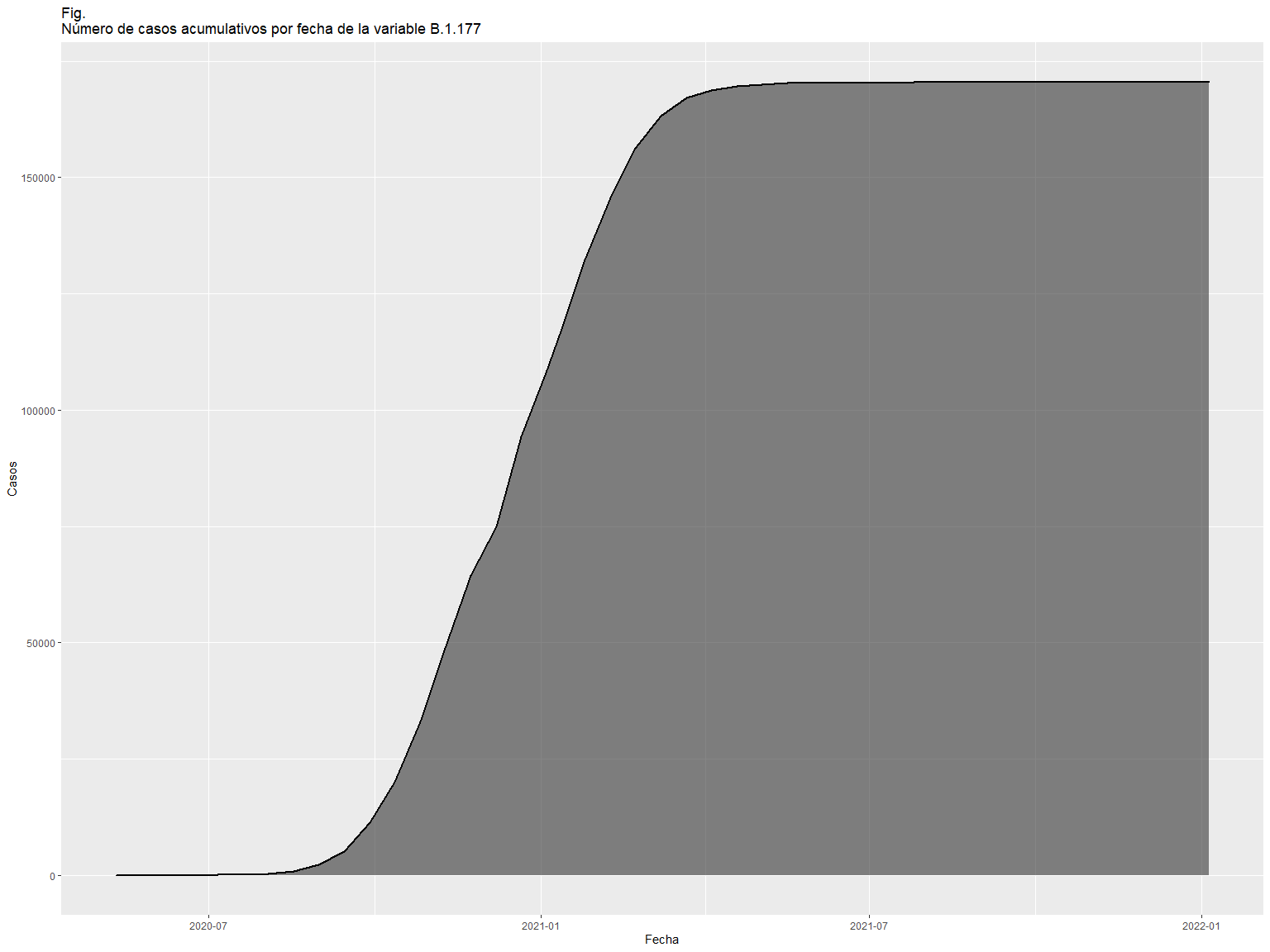
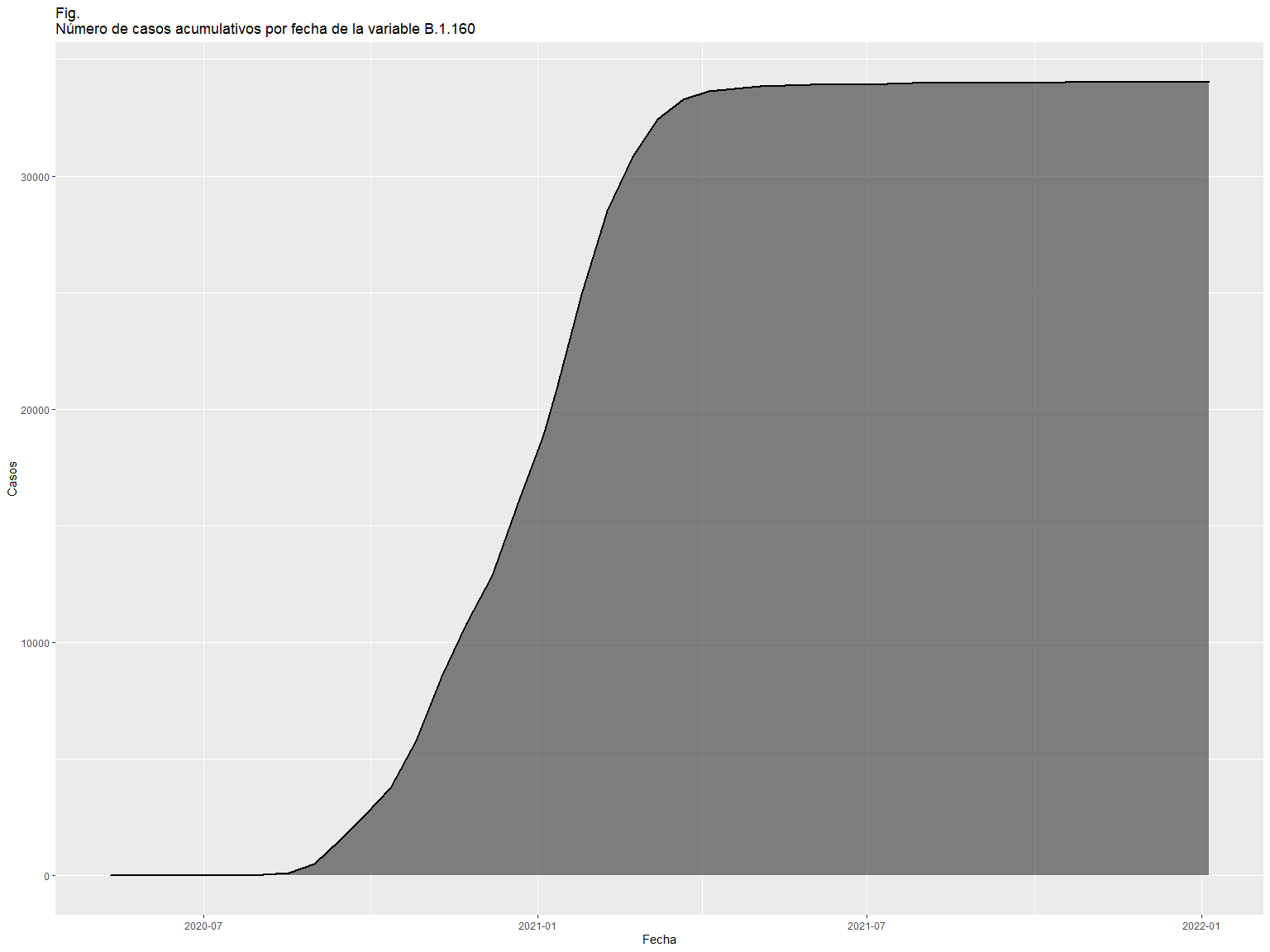
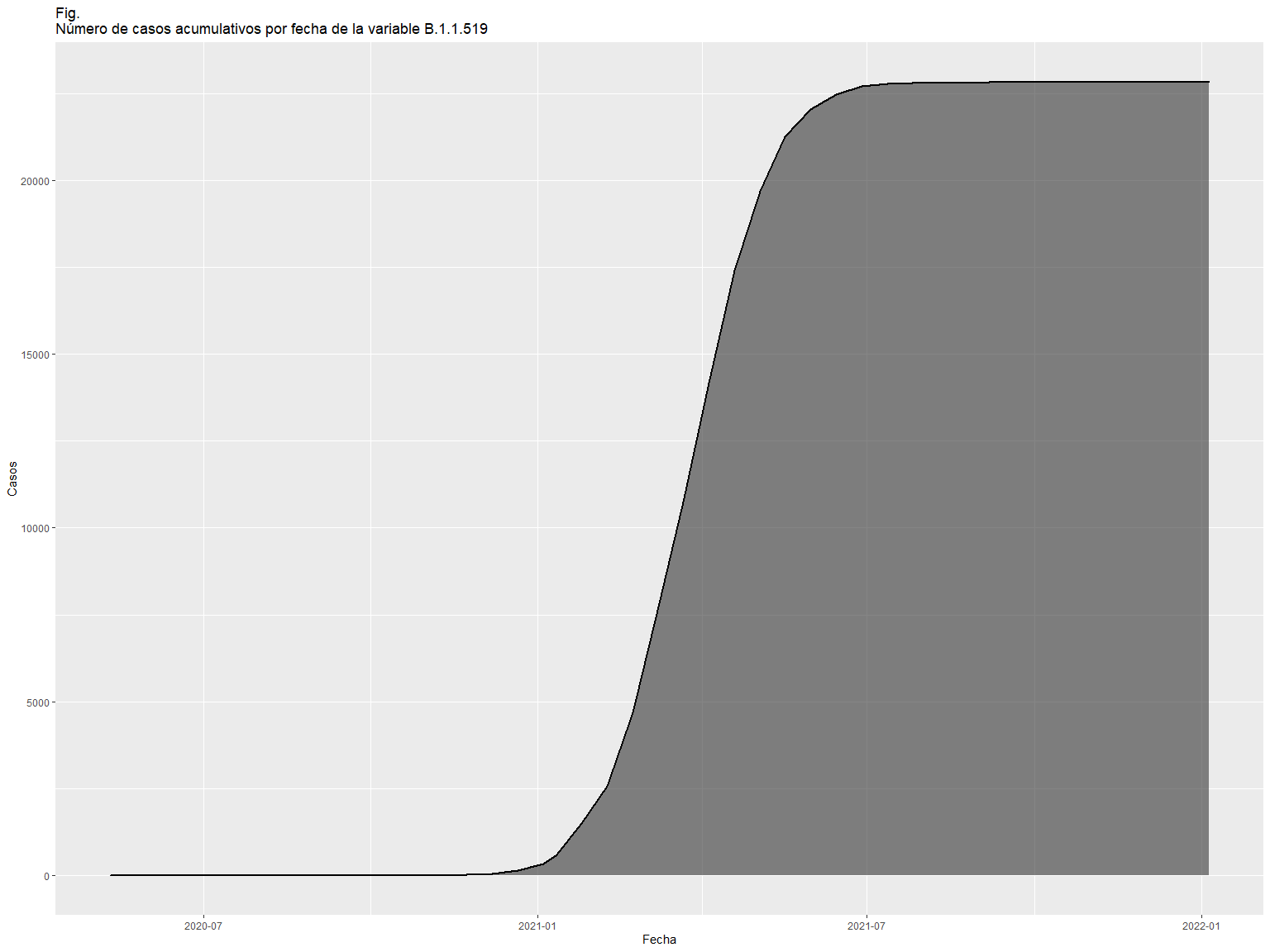
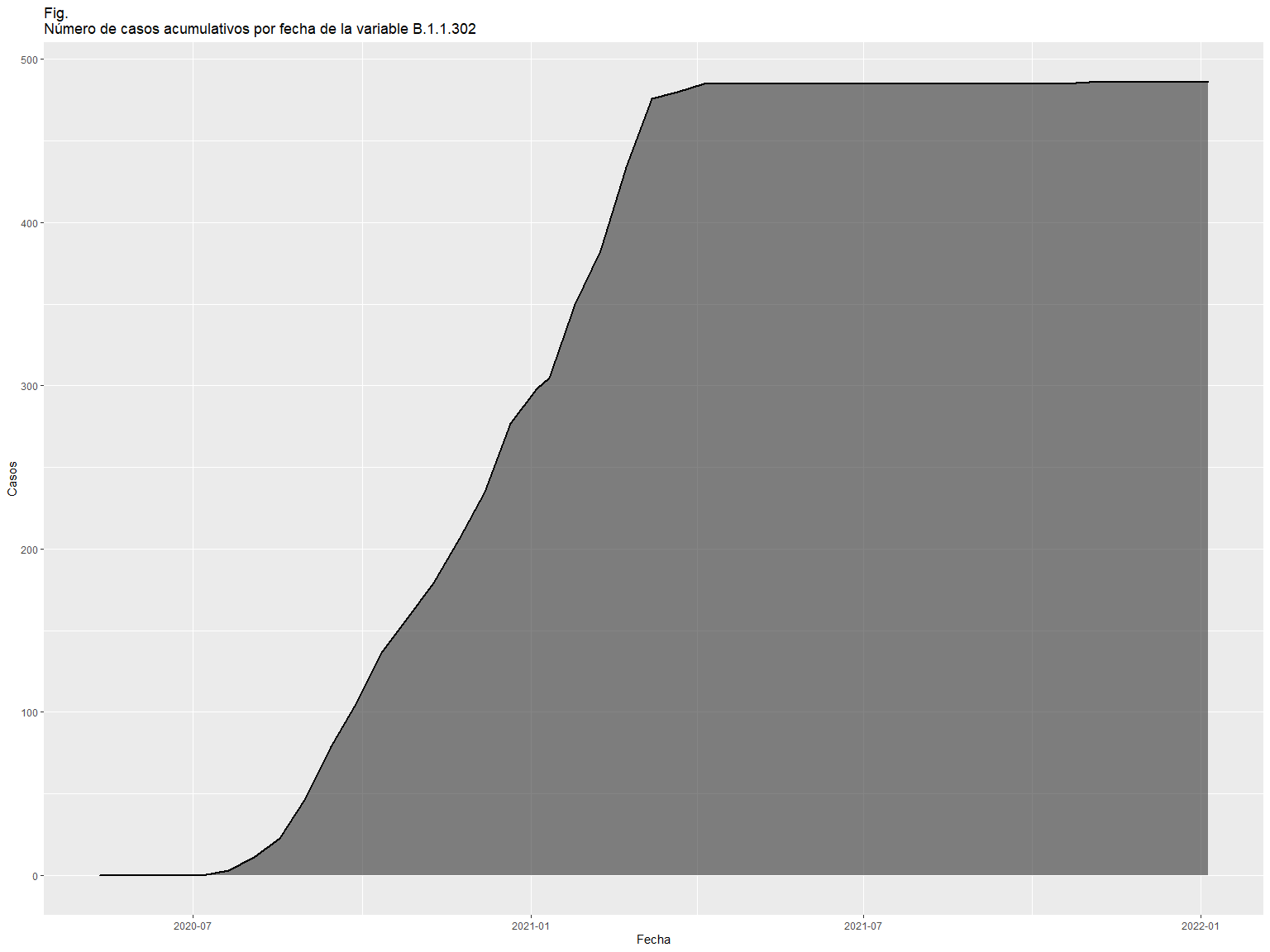
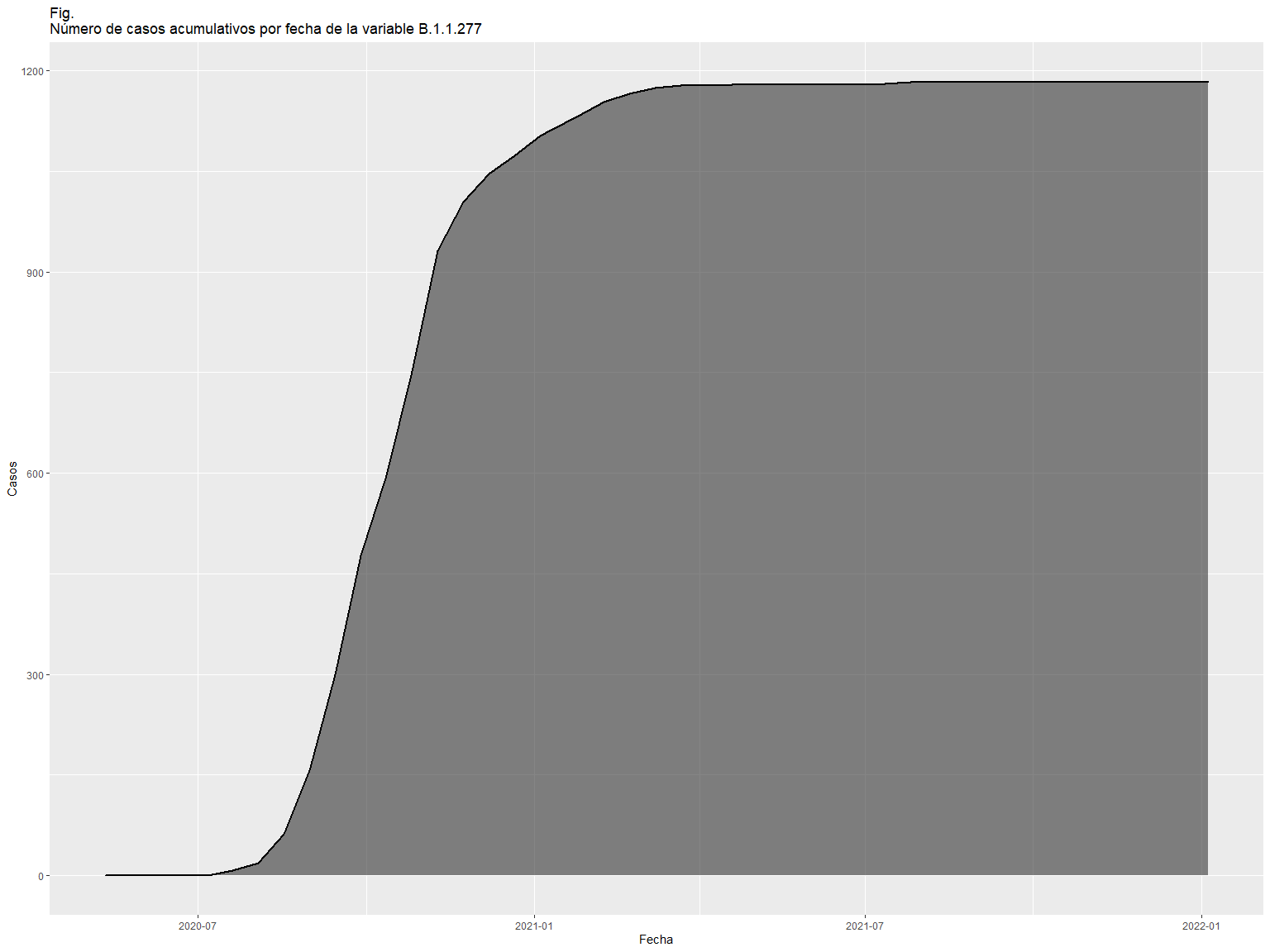
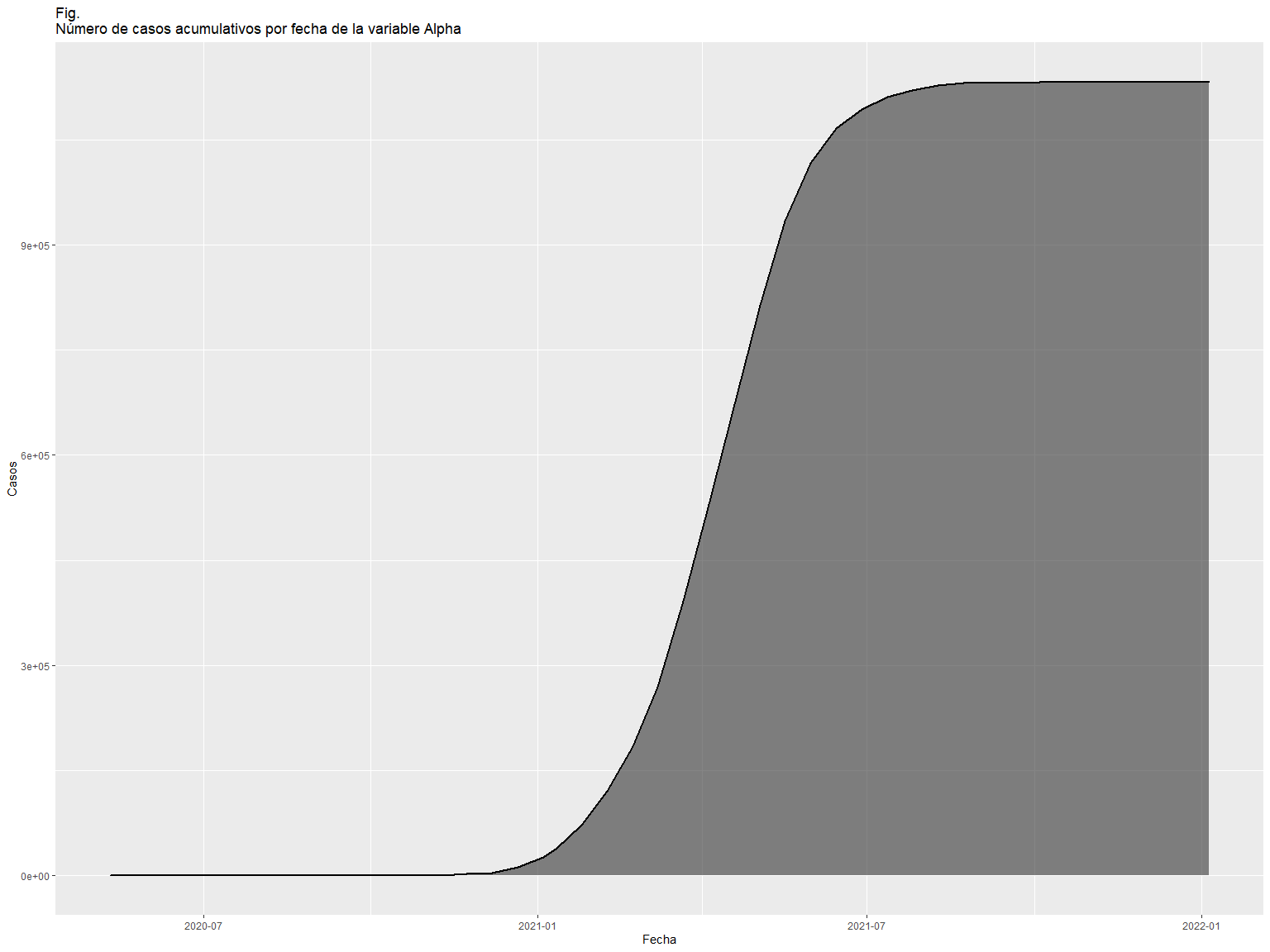


# 5. Casos por fecha

virus <- unique(DB$variant)  
for (i\_virus in virus){  
 data\_nueva1 <-   
 group\_by(DB[which(DB$variant == i\_virus),],date) %>%  
 summarise(num\_sequences = sum(num\_sequences))  
   
 # Grafico  
 print(  
 ggplot(data=data\_nueva1,   
 aes(x = date,   
 y = num\_sequences)) +  
 geom\_area(alpha = 0.6,  
 size = 1,  
 color = "black")+   
 ggtitle(paste("Fig. \nNúmero de casos por fecha de",i\_virus))+  
 xlab("Fecha")+  
 ylab("Casos"))  
}



virus <- unique(DB$variant)  
for (i\_virus in virus){  
 data\_nueva1 <-   
 group\_by(DB[which(DB$variant == i\_virus),],date) %>%  
 summarise(num\_sequences = sum(num\_sequences)) %>%  
 mutate(num\_sequences = cumsum(num\_sequences))  
   
 # Grafico  
 print(  
 ggplot(data=data\_nueva1,   
 aes(x = date,  
 y = num\_sequences)) +  
 geom\_area(aes(y = num\_sequences),  
 alpha = 0.6,  
 size = 1,  
 color = "black")+   
 ggtitle(paste("Fig. \nNúmero de casos acumulativos por fecha de la variable",i\_virus))+  
 xlab("Fecha")+  
 ylab("Casos"))  
 }



Para poder visualizar los casos por las fechas, se realizó el gráfico interactivo.

# 6. Percepción de las variables de COVID-19 en los paises

data.nueva2 <- DB[c('location','variant','num\_sequences')]  
data.nueva2 <- group\_by(.data = data.nueva2,   
 location)  
  
temp.1 <- group\_by(data.nueva2, location) %>%   
 summarise( num\_sequences= sum(num\_sequences))  
temp.2 <- group\_by(data.nueva2,   
 location,  
 variant) %>%   
 summarise( num\_sequences= sum(num\_sequences))  
paises <- unique(DB$location)  
nueva.data <- temp.2[1,]  
nueva.data <- nueva.data[-1,]  
for (pais in paises) {  
 temp.3 <- group\_by(temp.2[which(temp.2$location == pais),],location)   
 val <- group\_by(temp.1[which(temp.1$location == pais),],location)  
 temp.3$num\_sequences <- temp.3$num\_sequences /val$num\_sequences  
   
 # Modificar el data frame  
 nueva.data <- rbind(nueva.data,temp.3)   
}

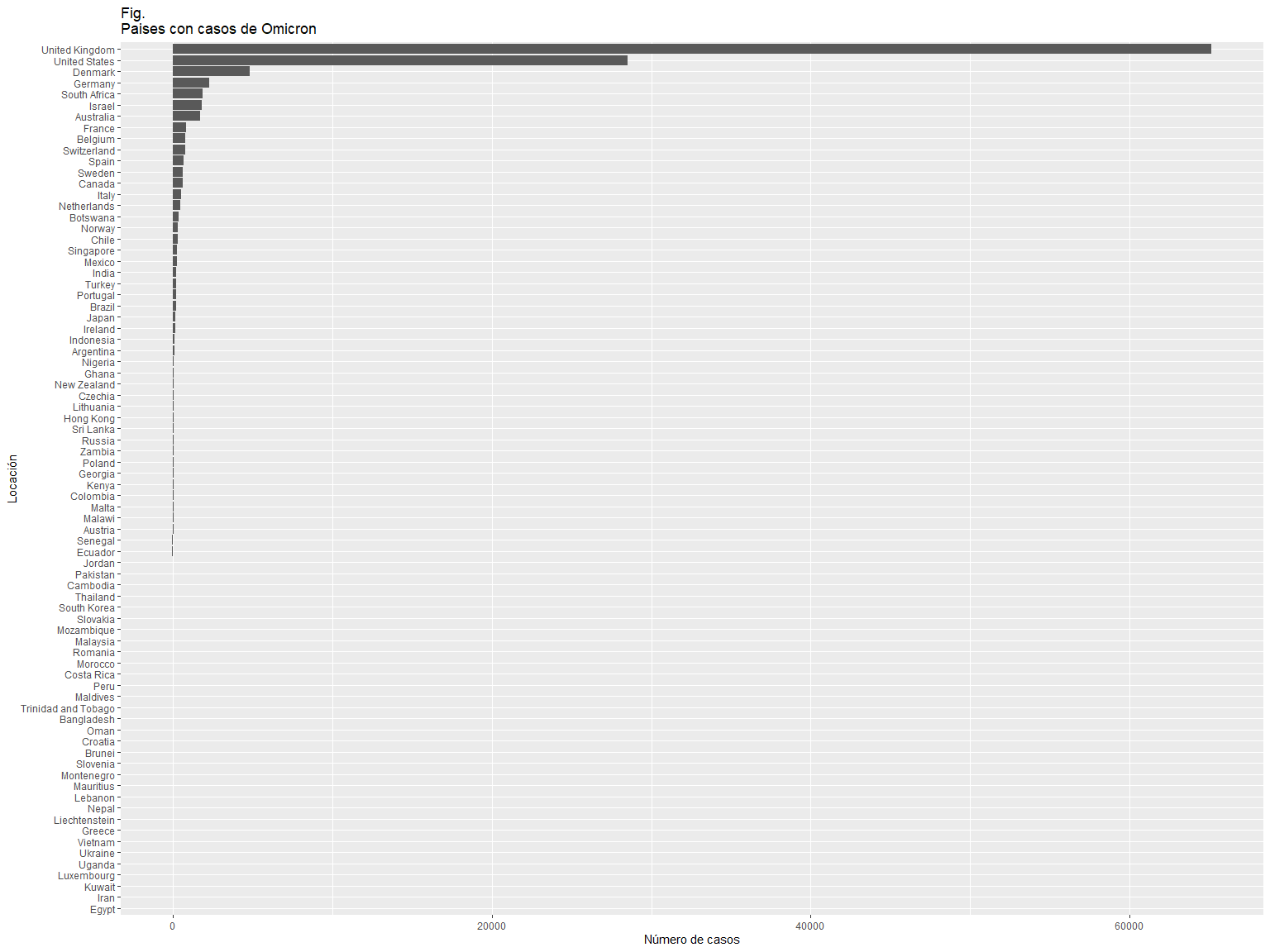
# Grafica de plot acumulada  
ggplot(data = nueva.data,   
 aes(x = location,   
 fill = variant,  
 y = num\_sequences)) +  
 geom\_bar(stat = "identity")+   
 labs(title = "Percepción de las cepas en los paises",  
 subtitle = "Relación entre 0 y 1",  
 x = "Paises",  
 y = "Porcentaje",  
 fill = 'Variantes')+  
 theme(plot.title=element\_text(family='Anton', size=10),  
 text = element\_text(size=10),  
 axis.text.x = element\_text(angle = 90,   
 vjust = 0.5,   
 hjust=1,  
 size = 7),  
 legend.title=element\_text(family='Anton', face='bold', size=14, color='black', hjust=0.5),  
 legend.position = 'bottom',  
 legend.key.size = unit(0.5, units = "cm"),  
 legend.box.background = element\_rect())



# 7. Variable Omicron

Omicron siendo la última variable de COVID-19 se analiza su incidencia en los paises.

data.nueva3 <- DB[c('location','variant','num\_sequences')]  
data.nueva3 <- filter(data.nueva3,   
 data.nueva3$variant == "Omicron")  
  
temp.1 <- group\_by(data.nueva3, location) %>%   
 summarise( num\_sequences= sum(num\_sequences))  
temp.1 <- filter(temp.1,  
 num\_sequences > 0)  
temp.1 <- arrange(temp.1, desc(num\_sequences))  
# Grafica de barra  
ggplot(data=temp.1,   
 aes(x = num\_sequences,   
 y = reorder(location, num\_sequences))) +  
 geom\_bar(stat="identity")+   
 ggtitle("Fig. \nPaises con casos de Omicron")+  
 xlab("Número de casos")+  
 ylab("Locación")



Se logra apreciar en la figura anterior, los paises que poseen al menos 1 caso de Omicron.

El país con mayor incidencia de la variante Omicron es United Kingdom, seguido de Estados Unidos, Dinamarca, Alemanía y Sudáfrica.