

Covid 19 en México

Murpholinox Peligro

19 de abril del 2020

```
# Carga los paquetes necesarios.
library(ggplot2)
library(ggdark)
library(latex2exp)
library(lubridate)
library(dplyr)
# Configura el directorio de trabajo.
setwd("/home/murphy/Repos/plotcovid19mx")
```

```
# Descarga datos del European CDC.
wget -O full.csv https://opendata.ecdc.europa.eu/covid19/casedistribution/csv
# Obtiene las líneas correspondientes a México.
grep Mex full.csv > mex.csv
# Solo ocupamos la columna 1 y 6 (fecha y decesos por día).
awk -F "," '{print $1"," $6}' mex.csv > clean.csv
# Añade nombre a las columnas.
echo "fecha,decesos" >> clean.csv
# Revierte los datos en la lista.
tac clean.csv > clean_r.csv
```

```
## --2020-04-23 00:21:39-- https://opendata.ecdc.europa.eu/covid19/casedistribution/csv
## Resolving opendata.ecdc.europa.eu (opendata.ecdc.europa.eu)... 212.181.0.63
## Connecting to opendata.ecdc.europa.eu (opendata.ecdc.europa.eu)|212.181.0.63|:443... connected.
## HTTP request sent, awaiting response... 301 Moved Permanently
## Location: https://opendata.ecdc.europa.eu/covid19/casedistribution/csv/ [following]
## --2020-04-23 00:21:39-- https://opendata.ecdc.europa.eu/covid19/casedistribution/csv/
## Reusing existing connection to opendata.ecdc.europa.eu:443.
## HTTP request sent, awaiting response... 200 OK
## Length: 720081 (703K) [application/octet-stream]
## Saving to: 'full.csv'
##
##      OK ..... 7% 226K 3s
##    50K ..... 14% 235K 3s
##   100K ..... 21% 242K 2s
##   150K ..... 28% 241K 2s
##   200K ..... 35% 247K 2s
##   250K ..... 42% 2.41M 1s
##   300K ..... 49% 246K 1s
##   350K ..... 56% 221K 1s
##   400K ..... 63% 85.4K 1s
##   450K ..... 71% 68.8M 1s
##   500K ..... 78% 216K 1s
##   550K ..... 85% 250K 0s
```

```
##      600K ..... 92% 239K 0s
##      650K ..... 99% 251K 0s
##      700K ... 100% 24.1M=2.9s
##
## 2020-04-23 00:21:43 (240 KB/s) - 'full.csv' saved [720081/720081]
```

```
# Carga los datos limpios a R.
datos <- read.csv("~/Repos/plotcovid19mx/clean_r.csv")
# Cambia el formato de la fecha de d/m/y a y-m-d.
datos$newdate <- lubridate::dmy(datos$fecha)

# Crea una nueva variable con un nuevo formato para la fecha (a números enteros).
# Esto se hace porque x con formato de fecha impide que el algoritmo para
# obtener el modelo exponencial llegue a una solución.
xmax <- max(length(datos$fecha))
datos$number <- seq(0,xmax-1)
# Necesitamos los días del brote en México (después del 20-marzo).
smalldf<-datos %>%
  filter(number >= 72)
# Ordena los datos a usar en un tibble.
x<-smalldf$number
x<-x-71
y<-smalldf$decesos
y2<-y+1
nice<-tibble(x,y)
dummy<-tibble(x,y2)
# Guarda datos finales en formato CSV.
write.csv(nice, file=~ /Repos/plotcovid19mx/nice.csv")
# Crea una gráfica base.
p <- ggplot(data = nice, aes(x=x, y=y)) + geom_point()

# Crea modelo lineal.
linm <- lm(log(y2)~x, data = dummy)
# Obtiene los parámetros del modelo exponencial.
a <- exp(coef(linm)[1])
b <- coef(linm)[2]
# Aplica los parámetros del modelo exponencial en la creación del mismo.
expm <- nls(y ~ a * exp(b * x), start = list(a=a, b=b), data = nice)
summary(expm)
```

```
##
## Formula: y ~ a * exp(b * x)
##
## Parameters:
##      Estimate Std. Error t value Pr(>|t|)
## a   2.27566    1.35475    1.68   0.103
## b   0.10902    0.01968    5.54 4.12e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.54 on 32 degrees of freedom
##
## Number of iterations to convergence: 5
## Achieved convergence tolerance: 4.603e-06
```

```
# Agrega el ajuste exponencial a la gráfica base,
p2 <- p +
  stat_smooth(method = 'nls', formula = y ~ a * exp(b * x), se=FALSE,
              method.args = list(start = list(a = a, b = b))) +
  # la ecuación de la exponencial,
  # annotate("label", x = 5, y = 115, label = TeX("$y = a e ^{bx}$")) +
  # los títulos necesarios,
  ylab("Decesos") + xlab("Día") + ggtitle("Decesos por Covid-19 (20-03/21-04)")
p2
```

