

Covid 19 en México

Murpholinox Peligro

04 05 2020

```
# Carga los paquetes necesarios.
library(ggplot2)
#library(ggdark) # No necesario si solo se crea el PDF.
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
```

```
## --2020-05-04 18:32:55-- 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]
## --2020-05-04 18:33:01-- 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: 865554 (845K) [application/octet-stream]
## Saving to: 'full.csv'
##
##      OK ..... 5% 123K 6s
##    50K ..... 11% 241K 5s
##   100K ..... 17% 247K 4s
##   150K ..... 23% 246K 3s
##   200K ..... 29% 252K 3s
##   250K ..... 35% 245K 3s
##   300K ..... 41% 947K 2s
##   350K ..... 47% 299K 2s
##   400K ..... 53% 278K 2s
##   450K ..... 59% 248K 1s
##   500K ..... 65% 248K 1s
##   550K ..... 70% 780K 1s
```

```
##      600K ..... 76% 249K 1s
##      650K ..... 82% 339K 1s
##      700K ..... 88% 240K 0s
##      750K ..... 94% 244K 0s
##      800K ..... 100% 2.98M=3.1s
```

```
##
```

```
## 2020-05-04 18:33:04 (276 KB/s) - 'full.csv' saved [865554/865554]
```

```
# 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)
```

```
# Se crean dos data frames, porque el modelo lineal con logaritmos se quiebra
```

```
# al tener ceros en y, por eso tenemos y2.
```

```
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 con el data frame dummy que contiene y2=y+1.
```

```
linm <- lm(log(y2)~x, data = dummy)
```

```
# Obtiene los parámetros del modelo lineal.
```

```
a1 <- exp(coef(linm)[1])
```

```
b1 <- coef(linm)[2]
```

```
# Aplica los parámetros del modelo lineal en la creación del modelo exponencial.
```

```
expm <- nls(y ~ a * exp(b * x), start = list(a=a1, b=b1), data = nice)
```

```
summary(expm)
```

```
##
```

```
## Formula: y ~ a * exp(b * x)
```

```
##
```

```
## Parameters:
```

```
##      Estimate Std. Error t value Pr(>|t|)
```

```
## a   9.04370      2.90233   3.116  0.00322 **
```

```
## b   0.05964      0.00821   7.265 4.69e-09 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 26.9 on 44 degrees of freedom
```

```
##
## Number of iterations to convergence: 11
## Achieved convergence tolerance: 3.773e-06

a2 <- coef(expm)[1]
b2 <- coef(expm)[2]

# 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 = a2, b = b2))) +
  # la ecuación de la exponencial,
  annotate("label", x = 10, y = 100,
    label = TeX(sprintf("$y = %.2f e^{%.2fx}$", a2, b2))) +
  # los títulos necesarios,
  ylab("Decesos") + xlab("Día") +
  ggtitle("Decesos por Covid-19 en México, a partir del 20 de marzo.")
p2
```

