

# Covid 19 en México

Murpholinux Peligro

12 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
```

```
## --2020-05-12 20:04:52-- 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]
## --2020-05-12 20:04:53-- 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: 950607 (928K) [application/octet-stream]
## Saving to: 'full.csv'
##
##      OK ..... 5% 76.8K 11s
##     50K ..... 10% 118K 9s
##    100K ..... 16% 245K 7s
##    150K ..... 21% 216K 5s
##    200K ..... 26% 119K 5s
##    250K ..... 32% 33.3K 7s
##    300K ..... 37% 26.6K 9s
##    350K ..... 43% 24.1K 10s
##    400K ..... 48% 47.0K 9s
##    450K ..... 53% 26.8K 9s
##    500K ..... 59% 111K 7s
##    550K ..... 64% 39.4K 7s
```

```
##      600K ..... 70% 85.3K 5s
##      650K ..... 75% 27.2K 5s
##      700K ..... 80% 58.5K 4s
##      750K ..... 86% 58.1K 3s
##      800K ..... 91% 47.6K 2s
##      850K ..... 96% 186K 1s
##      900K ..... 100% 34.4K=18s
```

```
##
```

```
## 2020-05-12 20:05:12 (50.7 KB/s) - 'full.csv' saved [950607/950607]
```

```
# 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 9.922456    2.826998   3.510 0.000947 ***
```

```
## b 0.057858    0.006232   9.284 1.52e-12 ***
```

```
## ---
```

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

```
##
## Residual standard error: 33.19 on 51 degrees of freedom
##
## Number of iterations to convergence: 10
## Achieved convergence tolerance: 2.08e-06

a2 <- coef(expm)[1]
b2 <- coef(expm)[2]
# Obtiene el número de decesos acumulados
cum <- sum(y)

# 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 = 150,
    label = TeX(sprintf("$y = %.2f e^{%.2fx}$", a2, b2))) +
  # una etiqueta para el número de decesos acumulados,
  annotate("label", x = 10, y = 200,
    label = TeX(sprintf("Decesos acumulados, %d", cum))) +
  # los títulos necesarios,
  ylab("Decesos") + xlab("Día") +
  ggtitle("Decesos por Covid-19 en México, a partir del 20 de marzo.")
p2
```

Decesos por Covid-19 en México, a partir del 20 de marzo.

