

Lab 3 - .qmd, gráficas, leer

Métodos Cuantitativos

Carlos Eduardo Molina Berumen

El Colegio de México

17 de octubre de 2025



Objetivos

- Usar nuestros propios datos
- Aprender a usar gráficas
- Entender .qmd

Hemos visto cómo se usan los datos ya precargados en R, pero ¿Cómo uso mis propios datos?

Debemos enseñarle a R a leer archivos con datos.

Descarguen en su computadora el archivo “Estados.csv”, guárdenlo en la carpeta donde está su proyecto.

¿Qué es un archivo .csv?

```
##install.packages(readr)  
  
library(readr)  
  
data <- read_csv("Estados.csv")
```

- ¿Qué datos contiene esta tabla?
- ¿Cuántas variables tiene?
- ¿Cuántas observaciones?
- ¿Qué valores puede tomar la variable “Entidad federativa”

```
ncol(data)
```

```
[1] 38
```

```
nrow(data)
```

```
[1] 192
```

```
unique(data$`Entidad federativa`)
```

```
[1] "01 Aguascalientes" "02 Baja California"
[3] "03 Baja California Sur" "04 Campeche"
[5] "05 Coahuila de Zaragoza" "06 Colima"
[7] "07 Chiapas" "08 Chihuahua"
[9] "09 Ciudad de México" "10 Durango"
[11] "11 Guanajuato" "12 Guerrero"
[13] "13 Hidalgo" "14 Jalisco"
[15] "15 México" "16 Michoacán de Oca
[17] "17 Morelos" "18 Nayarit"
```

Intentemos sacar el promedio de años de escolaridad de la población mexicana incluyendo todos los años.

```
prom_escolaridad <- mean(data$Años_escolaridad_15)
```

¿Qué sucedió?

Ahora intenten este código

```
prom_escolaridad <- mean(data$Años_escolaridad_15, na.rm= TRUE)
```

Interpreten el resultado.

Como ya vimos, data contiene datos de muchos años. Creemos una nueva base con solo dtaos de 2020.

```
library(dplyr)

data2020 <- data %>%
  filter(Año == "2020")
```

¿Cuál es el estado con la escolaridad máxima? ¿Cuántos años son lo máximo?

```
max(data2020$Años_escolaridad_15)
```

```
[1] 11.48204
```

```
data2020 %>%  
  filter(Años_escolaridad_15 == max(Años_escolaridad_15))
```

```
# A tibble: 1 x 38
```

```
  `Entidad federativa`    Año Porcentaje_alfabetizacion Años_es  
  <chr>                  <dbl>                <dbl>  
1 09 Ciudad de México    2020                98.2  
# i 34 more variables: Pob_hombres <dbl>, Pob_mujeres <dbl>,  
#   Edad_mediana <dbl>, `Pob_0-4_años` <dbl>, `Pob_10-14_años`  
#   `Pob_+100_años` <dbl>, `Pob_15-19_años` <dbl>, `Pob_20-24`  
#   `Pob_25-29_años` <dbl>, `Pob_30-34_años` <dbl>, `Pob_35-39`  
#   `Pob_40-44_años` <dbl>, `Pob_45-49_años` <dbl>, `Pob_05-9`
```

Hágamos una serie de gráficas sobre nuestros datos.

```
as.numeric(data2020$Edad_mediana)
```

```
[1] 27 30 29 29 29 30 24 29 35 27 28 27 30 29 30 28 30 29 30  
[26] 30 29 30 28 31 30 28
```

```
hist(data2020$Edad_mediana,  
      xlab = "Edad",  
      ylab = "FRecuencia")
```

Histogram of data2020\$Edad_mediana

