

Visualización con ggplot I

Contenidos

- 01.** La importancia de la visualización
- 02.** Generando gráficos en R
- 03.** La sintaxis de ggplot
- 04.** Añadiendo elementos de visualización



01.

La importancia de la visualización



La visualización juega un rol importante en las etapas del análisis de datos:

- Exploración
- Modelamiento
- Comunicación

"Un simple gráfico ha traído más información a la mente de los analistas que cualquier otro dispositivo" (John Tukey)





Aspectos a considerar

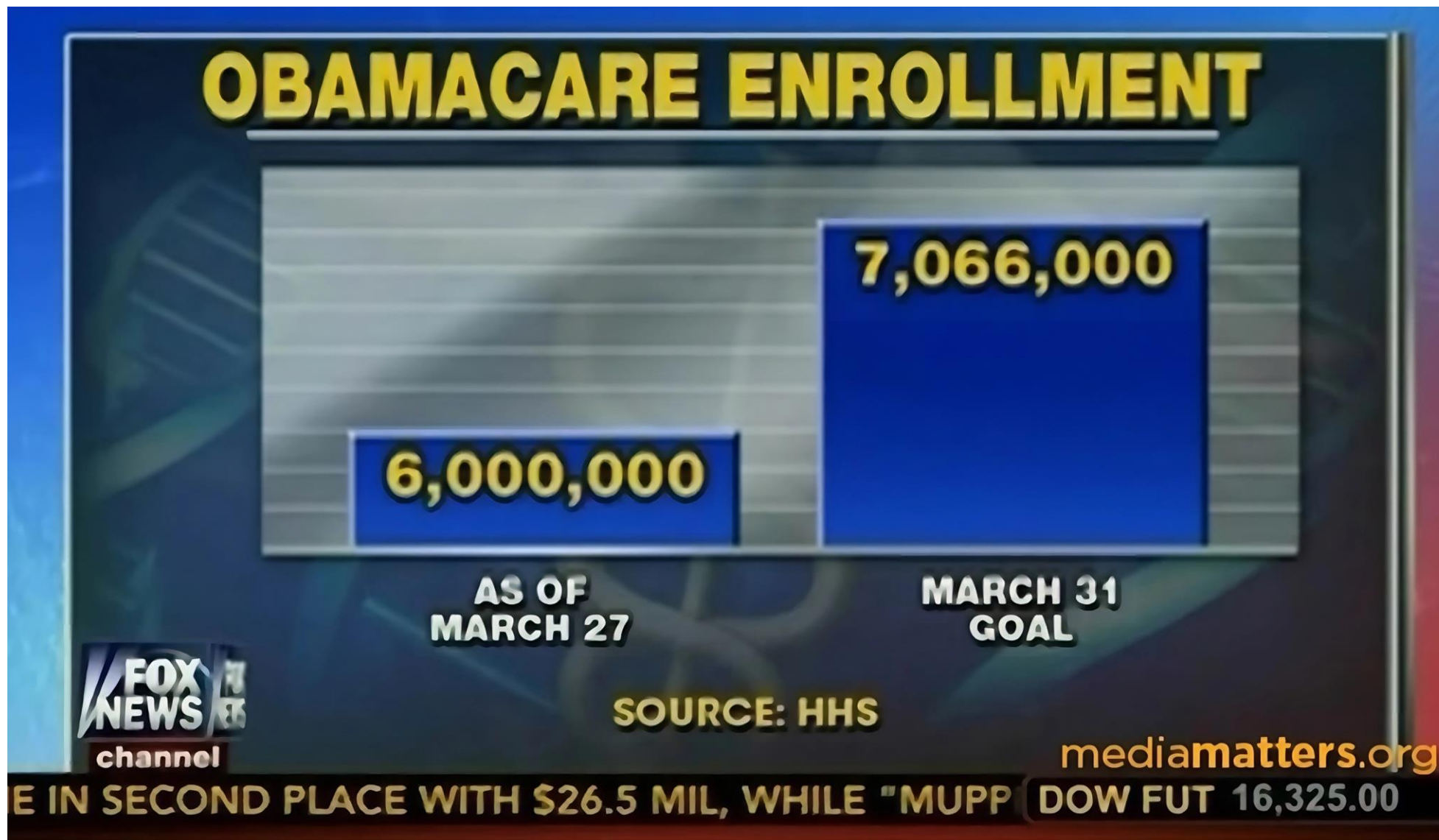
- Los gráficos nos permiten **comunicar y atraer la atención de una audiencia**: John Tukey
- Más allá de esto, debemos considerar que es necesaria la correcta utilización de las herramientas que disponemos, con la finalidad de que **nuestros gráficos reflejen, de la manera más fidedigna posible, la realidad que estamos intentando medir.**



Una mala construcción de gráficos puede generar visualizaciones no efectivas



Visualización no efectiva



- Cómo vimos en el gráfico anterior, **la manipulación de los datos afecta la interpretación de la información que podemos extraer de este.**
- Así, en base a la falta de precisión de la variable correspondiente al eje Y, **se obtiene una captura de la realidad distorsionada**, que no refleja el estado real del fenómeno que buscamos medir: la afiliación al programa de salud estadounidense *Obamacare*.
- Tomando en cuenta lo anterior, a continuación, presentaremos la librería ggplot, en base a la cual es posible crear visualizaciones de datos en el software R.





02.

Generando gráficos en R





Descripción del dataset

- Trabajaremos con un *dataset* de nombres inscritos en el registro civil
- Los datos se encuentran en el paquete `guaguas` de [Riva Quiroga](#).
- A continuación, escribimos nuestro script en R:

```
#install.packages("guaguas")  
library(guaguas)  
head(guaguas, 2)  
  
## # A tibble: 2 × 5  
##   año nombre sexo      n proporción  
##   <dbl> <chr> <chr> <dbl>     <dbl>  
## 1  1920 María  F     2130     0.104  
## 2  1920 José   M      984     0.0483
```



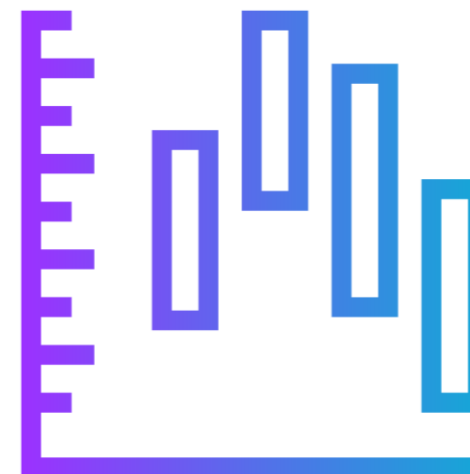


Generemos un gráfico de barras con la cantidad de mujeres y hombres

Para comenzar, llamemos a la librería *tidyverse* y creamos una tabla de datos, que será nuestra base para la visualización gráfica:

```
library(tidyverse)
tabla <- guaguas %>%
  group_by(sexo) %>%
  summarise(frecuencia = sum(n))
Tabla
```

```
## # A tibble: 3 × 2
##   sexo frecuencia
##   <chr>      <dbl>
## 1 F          10931372
## 2 I             344
## 3 M          11197384
```



Vemos que la variable sexo presenta 3 categorías:

- **Femenino (F)**
- **Indeterminado (I)**
- **Masculino (M)**



03.

La sintaxis de ggplot



ggplot funciona con un sistema de capas. Las revisamos a continuación:

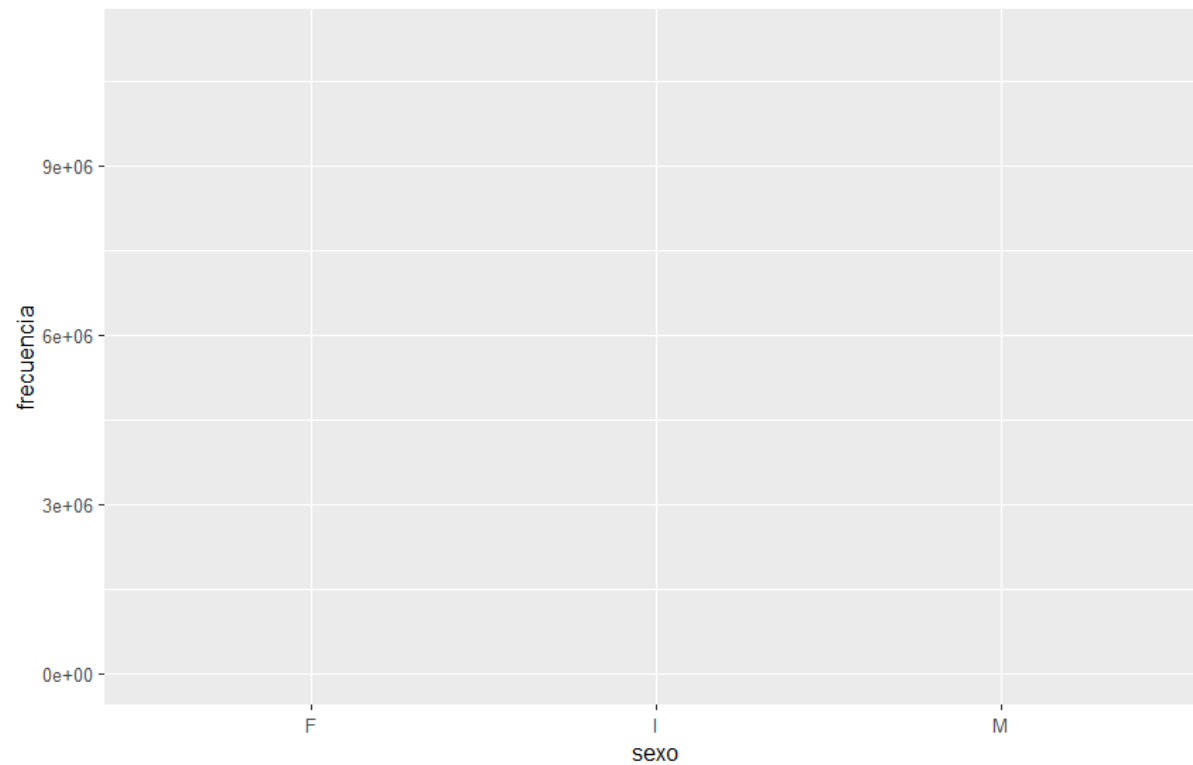
- **data**: tabla para la cual queremos generar el gráfico.
- **aes**: viene de aesthetics. Contiene los elementos estéticos de nuestro gráfico
 - **x**: variable que será mapeada al eje x
 - **y**: variable que será mapeada al eje y



Llamamos librería y ejecutamos código

```
library(ggplot2)
```

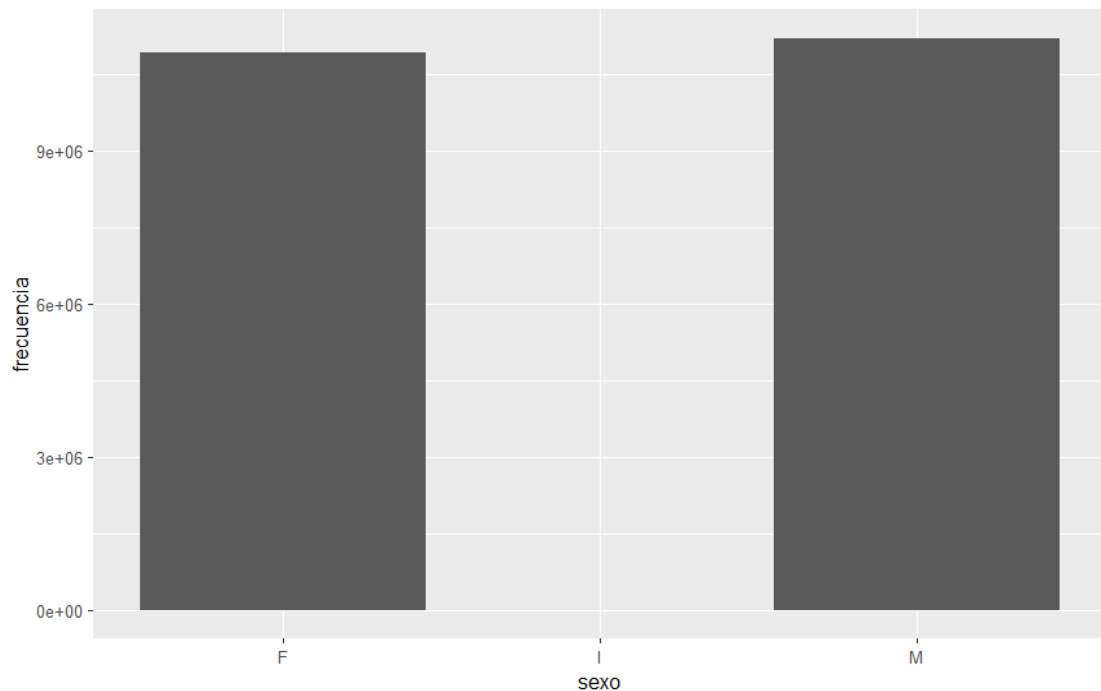
```
ggplot(data = tabla, aes(x = sexo, y = frecuencia))
```





Añadiendo capas...

```
ggplot(data = tabla, aes(x = sexo, y = frecuencia)) +  
  geom_bar(stat = "identity")
```



- **geom_bar**: capa de figura geométrica
- **"identity"** le dice a ggplot que no queremos hacer nada con los datos, es decir, graficar la tabla sin modificaciones

¿Qué es exactamente identity?

- **geom_bar** por defecto llama a la función **stat_count**
- **stat_count** espera como input una sola variable
- Escribimos y ejecutamos código:

```
ggplot(data = tabla, aes(x = sexo, y = frecuencia)) +  
  geom_bar()
```

```
## Error in `geom_bar()`:  
## ! Problem while computing stat.  
## i Error occurred in the 1st layer.  
## Caused by error in `setup_params()`:  
## ! `stat_count()` must only have an x or y aesthetic.
```

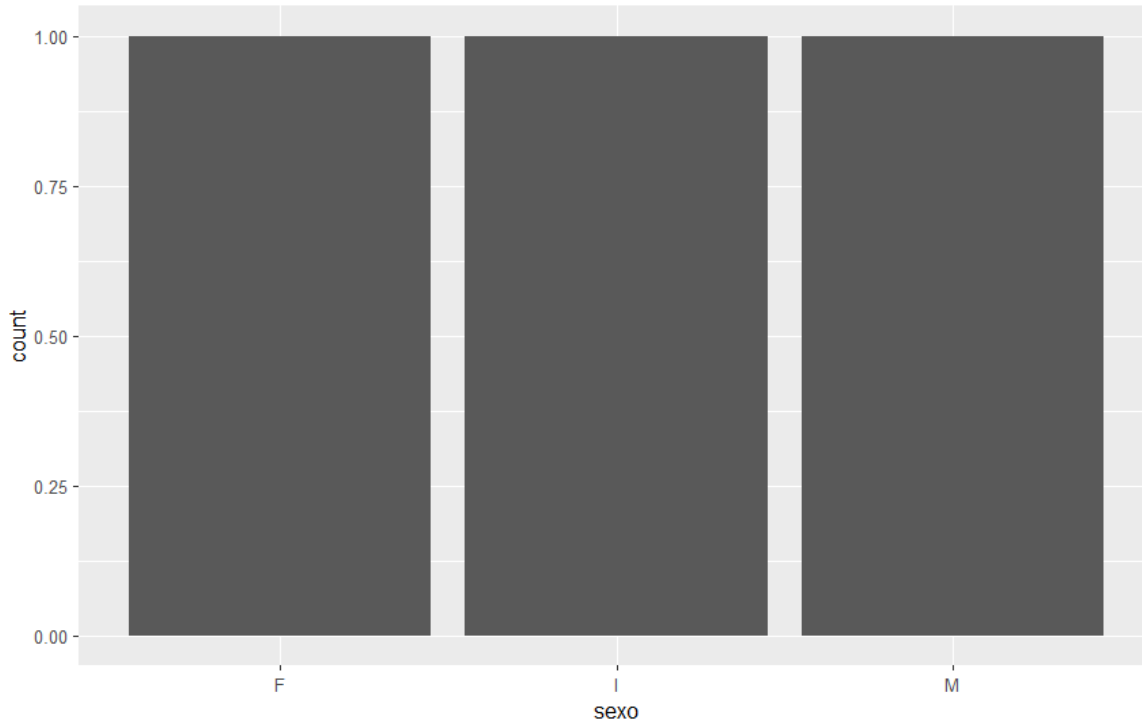




Sigamos la instrucción de ggplot

Escribimos y ejecutamos código:

```
ggplot(data = tabla, aes(x = sexo)) +  
  geom_bar()
```



- **geom_bar** cuenta el número de filas en cada categorías
- En ciertas ocasiones será útil

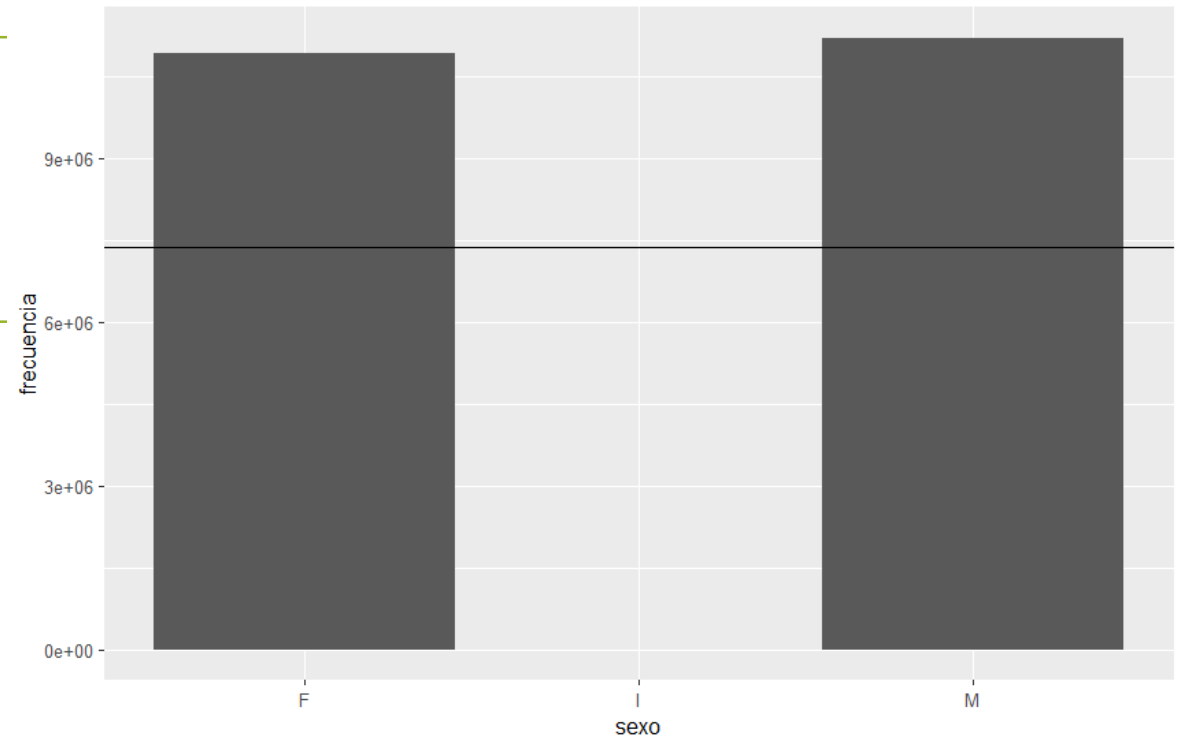


Otras capas a gráficos

Escribimos y ejecutamos código:

```
ggplot(data = tabla, aes(x = sexo, y = frecuencia)) +  
  geom_bar(stat = "identity") +  
  geom_hline(aes(yintercept = mean(frecuencia)) )
```

- **geom_hline**: línea horizontal





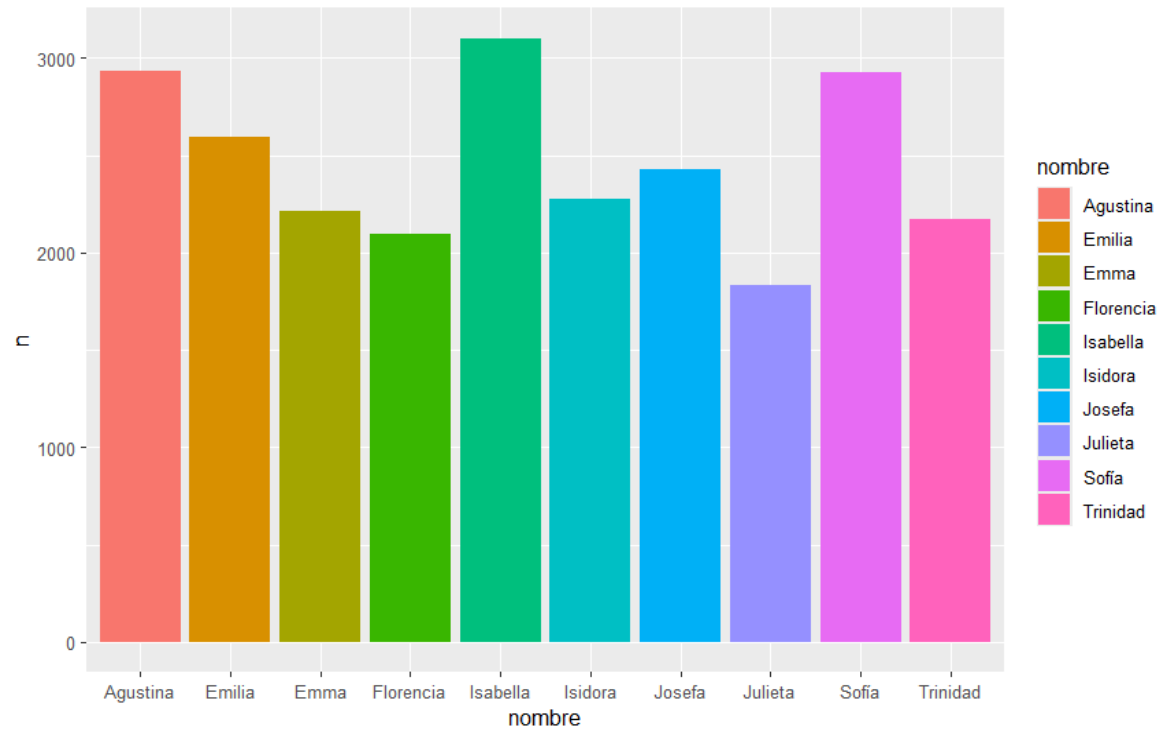
04.

**Agregando más
atributos**





Nombres de mujer más populares de 2019



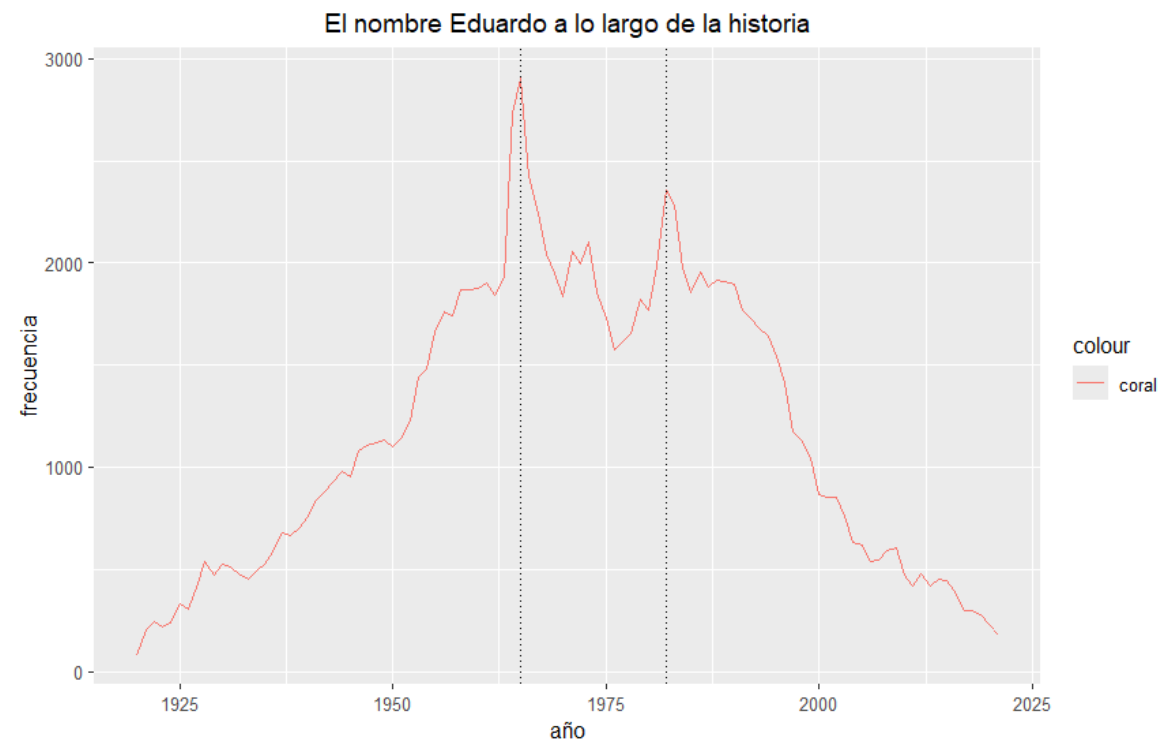
Escribimos y ejecutamos código:

```
top10 <- guaguas %>%  
  filter(anio == 2019 & sexo == "F") %>%  
  arrange(desc(n)) %>%  
  slice(1:10)  
ggplot(top10, aes(x = nombre, y = n, fill = nombre)) +  
  geom_bar(stat = "identity")
```

Nombre Eduardo a lo largo de la historia

Escribimos y ejecutamos código:

```
guaguas %>%
  filter(nombre == "Eduardo" & sexo == "M") %>%
  ggplot(aes(x = anio, y = n, color = "coral"))
+
  geom_line() +
  geom_vline(xintercept = c(1965,
1982), linetype="dotted" ) +
  labs(title="El nombre Eduardo a lo largo de la
historia",
x ="año", y = "frecuencia") +
  theme(plot.title = element_text(hjust = 0.5))
```





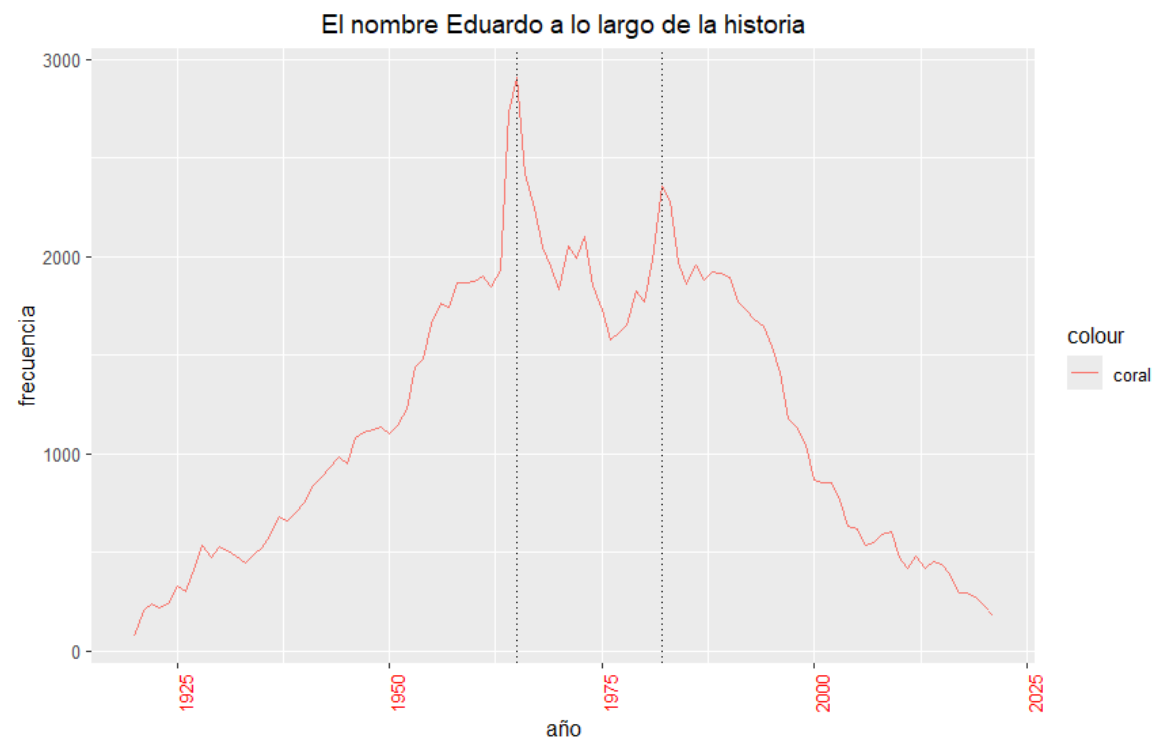
```

guaguas %>%
  filter(nombre == "Eduardo" & sexo == "M") %>%
  ggplot(aes(x = anio, y = n, color = "coral")) +
  geom_line() +
  geom_vline(xintercept = c(1965, 1982), linetype="dotted" ) +
  labs(title="El nombre Eduardo a lo largo de la historia",
        x = "año", y = "frecuencia") +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 90, color = "red" ))

```

Un poco más sobre theme

theme nos permite modificar elementos estéticos:



```

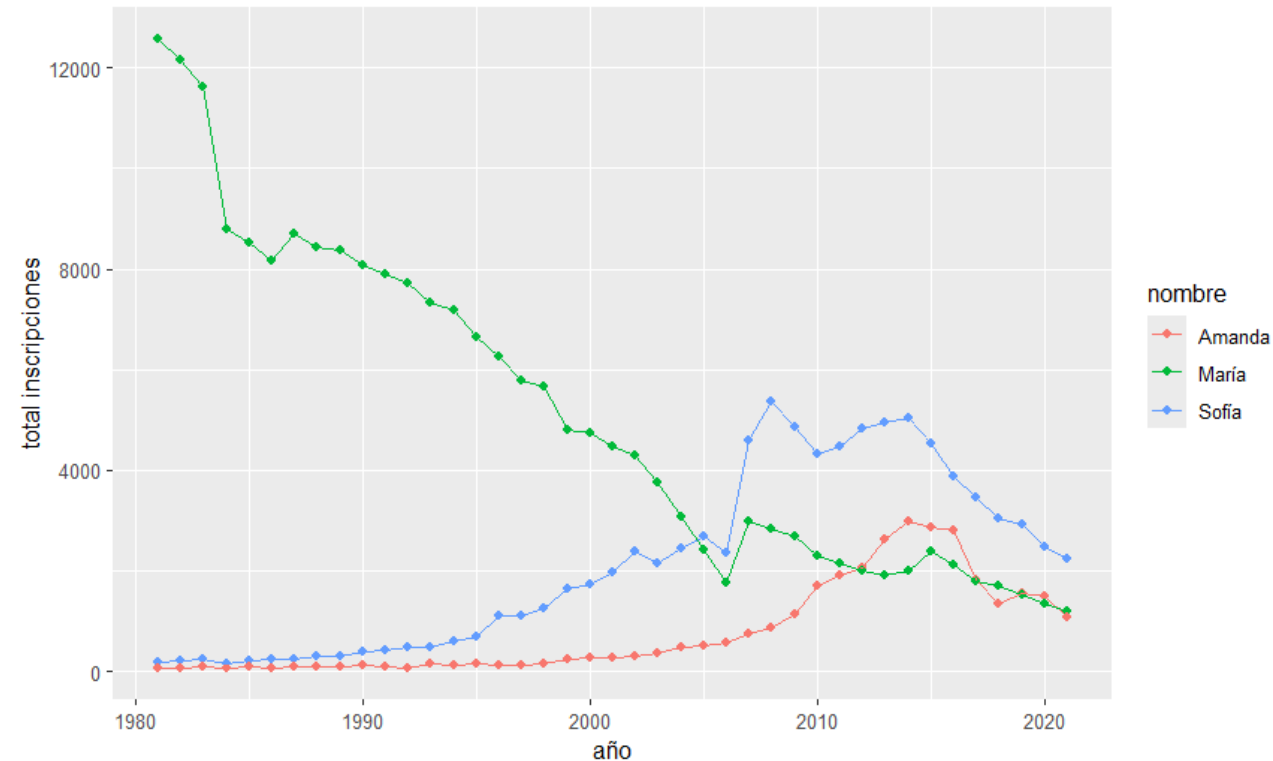
guaguas %>%
  filter(sexo == "F") %>%
  filter(nombre %in% c("Sofía", "María", "Amanda") & anio > 1980) %>%
  ggplot(aes(anio, n, color = nombre)) +
  geom_point() +
  geom_line() +
  labs(x = "año", y = "total inscripciones",
  title = "Algunos nombre de mujer a lo largo de la historia") +
  theme(plot.title = element_text(hjust = 0.5, size = 16, color = "blue"))

```

Escribimos y ejecutamos código:

Agreguemos más cosas al gráfico

Algunos nombre de mujer a lo largo de la historia



¡Muchas gracias!





www.ine.gob.cl