

Taller de R

Sesión 5

agostos 2022

Material de trabajo

Carpeta de drive:

<https://drive.google.com/drive/u/0/folders/14QDyRnqnLUjosCBIAzBqXL-qgIOMjNJL>

Base a trabajar: víctimas del secretariado:[clic aquí](#)

Liga de zoom <https://us02web.zoom.us/j/83841692580>

ID de reunión: 838 4169 2580

Hoy

- Datos abierto: SESNSP
- `group_by()`, `filter()` y `summarise()`
- Paquetería `ggplot2`:
- estructura de graficos.
- visualizaciones
- Esquisse

Datos de víctimas del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (SESNSP)

Son datos abiertos que registran la incidencia en el país, lo integra en 3 grandes dataset o conjunto de datos:

- Cifras de Incidencia Delictiva Estatal, 2015 - julio 2022 (carpetas)
- Cifras de Víctimas del Fuero Común, 2015 - julio 2022 (víctimas)
- Cifras de Incidencia Delictiva Federal, 2012 - julio 2022 (carpetas)

Cada Fiscalía reporta al SESNSP los registros mensuales de más de 13 delitos, tiene un histórico de enero 2015 a julio 2022.

Aquí el sitio web: <https://www.gob.mx/sesnsp/acciones-y-programas/datos-abiertos-de-incidencia-delictiva>

Base de datos

Leer o cargar base de datos

```
victimas<-read.csv("https://docs.google.com/spreadsheets/d/e/2PACX-1vTrYBvTBhjBzuo
```

Nombres de columna de la base de víctimas

```
colnames(victimas)
```

```
## [1] "Año" "Clave_Ent" "Entidad"
## [4] "Bien.jurídico.afectado" "Tipo.de.delito" "Subtipo.de.delito"
## [7] "Modalidad" "Sexo" "Rango.de.edad"
## [10] "Enero" "Febrero" "Marzo"
## [13] "Abril" "Mayo" "Junio"
## [16] "Julio" "Agosto" "Septiembre"
## [19] "Octubre" "Noviembre" "Diciembre"
```

Dimensión de la base

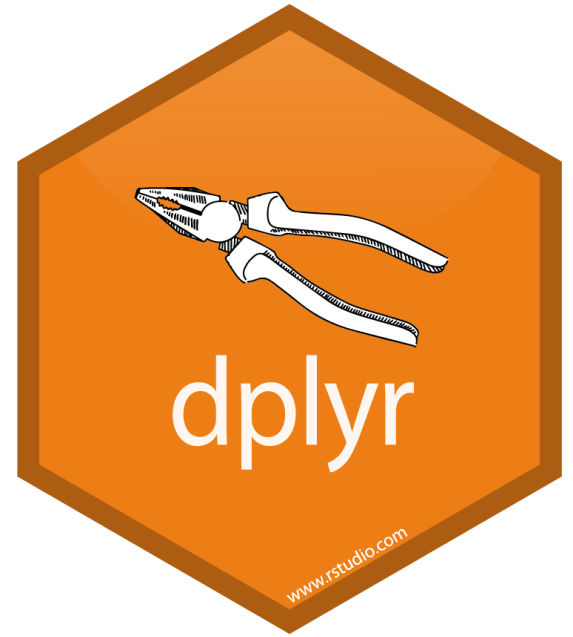
```
dim(victimas)
```

```
## [1] 58880 21
```

Más sobre la base

```
unique(victimas$Tipo.de.delito)
```

```
## [1] "Homicidio"  
## [2] "Lesiones"  
## [3] "Feminicidio"  
## [4] "Otros delitos que atentan contra la vida y la integridad corporal"  
## [5] "Secuestro"  
## [6] "Tráfico de menores"  
## [7] "Rapto"  
## [8] "Otros delitos que atentan contra la libertad personal"  
## [9] "Extorsión"  
## [10] "Corrupción de menores"  
## [11] "Trata de personas"  
## [12] "Otros delitos contra la sociedad"  
## [13] "Aborto"
```



dplyr: más funciones

group_by()

La función **group_by()** agrupa un conjunto de filas seleccionado en un conjunto de filas de resumen de acuerdo con los valores de una o más columnas o expresiones.

city	particle size	amount ($\mu\text{g}/\text{m}^3$)
New York	large	23
New York	small	14



city	mean	sum	n
New York	18.5	37	2

London	large	22
London	small	16



London	19.0	38	2
--------	------	----	---

Beijing	large	121
Beijing	small	56



Beijing	88.5	177	2
---------	------	-----	---

summarise()

La función **summarise()** funciona de forma análoga a la función **mutate()**, excepto que en lugar de añadir nuevas columnas crea un nuevo data frame.

Así por ejemplo, ara calcular la mediana y la varianza de la variable **amount** en el conjunto de datos **pollution**:

summarise()

city	particle size	amount ($\mu\text{g}/\text{m}^3$)
New York	large	23
New York	small	14
London	large	22
London	small	16
Beijing	large	121
Beijing	small	56



median	variance
22.5	1731.6

summarise te puede devolver: suma, máximo, mínimos, varianza, mediana y desv standar.

na.rm=T

Excluye los valores faltantes al realizar estos cálculos, simplemente podemos incluir el argumento `na.rm = TRUE`

Por ejemplo:

Al momento de querer sumar el las 12 columnas de las variables para todos los años, los últimos meses de 2022 no estaran porque aún no hay registro y para que R entienda que es estos valores faltantes no afecta se utiliza el argumento `na.rm=T`

A transformar datos

```
#Sólo trabajaremos con Feminicidios y Homicidios dolosos de mujeres -filtramos del  
delitos<- c("Feminicidio", "Homicidio doloso")
```

```
#Hacemos sumatoria de los meses por los delitos que corresponden a mujeres en MX  
victimas%>% filter(Sexo=="Mujer",  
                  Subtipo.de.delito%in%delitos) %>%  
group_by(Año, Subtipo.de.delito) %>%  
summarise(ene=sum(Enero, na.rm = T),  
          feb=sum(Febrero, na.rm = T),  
          mar=sum(Marzo, na.rm = T),  
          abr=sum(Abril, na.rm = T),  
          may=sum(Mayo, na.rm = T),  
          jun=sum(Junio, na.rm = T),  
          jul=sum(Julio, na.rm = T),  
          ago=sum(Agosto, na.rm = T),  
          sep=sum(Septiembre, na.rm = T),  
          oct=sum(Octubre, na.rm = T),  
          nov=sum(Noviembre, na.rm = T),  
          dic=sum(Diciembre, na.rm = T),  
          Total=sum(ene+feb+mar+abr+  
                   may+jun+jul+ago+  
                   sep+oct+nov+dic)) %>%  
select(Año, Subtipo.de.delito, Total)->victimas_anual
```

Objeto creado llamado victiminas anual

```
victiminas_anual
```

```
## # A tibble: 16 × 3
## # Groups:   Año [8]
##   Año Subtipo.de.delito Total
##   <int> <chr>           <int>
## 1  2015 Feminicidio         427
## 2  2015 Homicidio doloso    1734
## 3  2016 Feminicidio         647
## 4  2016 Homicidio doloso    2188
## 5  2017 Feminicidio         766
## 6  2017 Homicidio doloso    2535
## 7  2018 Feminicidio         919
## 8  2018 Homicidio doloso    2759
## 9  2019 Feminicidio         972
## 10 2019 Homicidio doloso    2867
## 11 2020 Feminicidio         978
## 12 2020 Homicidio doloso    2798
## 13 2021 Feminicidio        1016
## 14 2021 Homicidio doloso    2745
## 15 2022 Feminicidio         545
## 16 2022 Homicidio doloso    1639
```

Usando summarise calculamos una métrica

```
victimas_anual %>% group_by(Año) %>%  
  summarise(muertes_dia=sum(Total/365))
```

```
## # A tibble: 8 × 2  
##   Año muertes_dia  
##   <int>      <dbl>  
## 1  2015         5.92  
## 2  2016         7.77  
## 3  2017         9.04  
## 4  2018        10.1  
## 5  2019        10.5  
## 6  2020        10.3  
## 7  2021        10.3  
## 8  2022         5.98
```

Muertes de mujeres en México al día **10 mujeres al día son asesinadas en México**. Se observa un incremento a 10 mujeres que se a mantenido constante.

Se crear objeto que incluya los feminicidios en MX por año

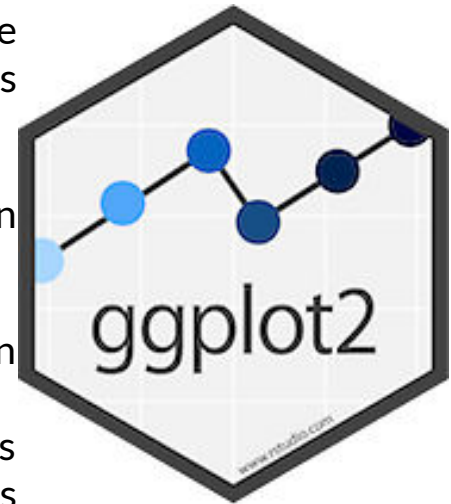
```
fem_anual<-victimas%>%  
  filter(Sexo=="Mujer",  
         Subtipo.de.delito=="Feminicidio") %>%  
  group_by(Año) %>%  
  summarise(ene=sum(Enero, na.rm = T),  
            feb=sum(Febrero, na.rm = T),  
            mar=sum(Marzo, na.rm = T),  
            abr=sum(Abril, na.rm = T),  
            may=sum(Mayo, na.rm = T),  
            jun=sum(Junio, na.rm = T),  
            jul=sum(Julio, na.rm = T),  
            ago=sum(Agosto, na.rm = T),  
            sep=sum(Septiembre, na.rm = T),  
            oct=sum(Octubre, na.rm = T),  
            nov=sum(Noviembre, na.rm = T),  
            dic=sum(Diciembre, na.rm = T),  
            Total=sum(ene+feb+mar+abr+  
                      may+jun+jul+ago+  
                      sep+oct+nov+dic))
```

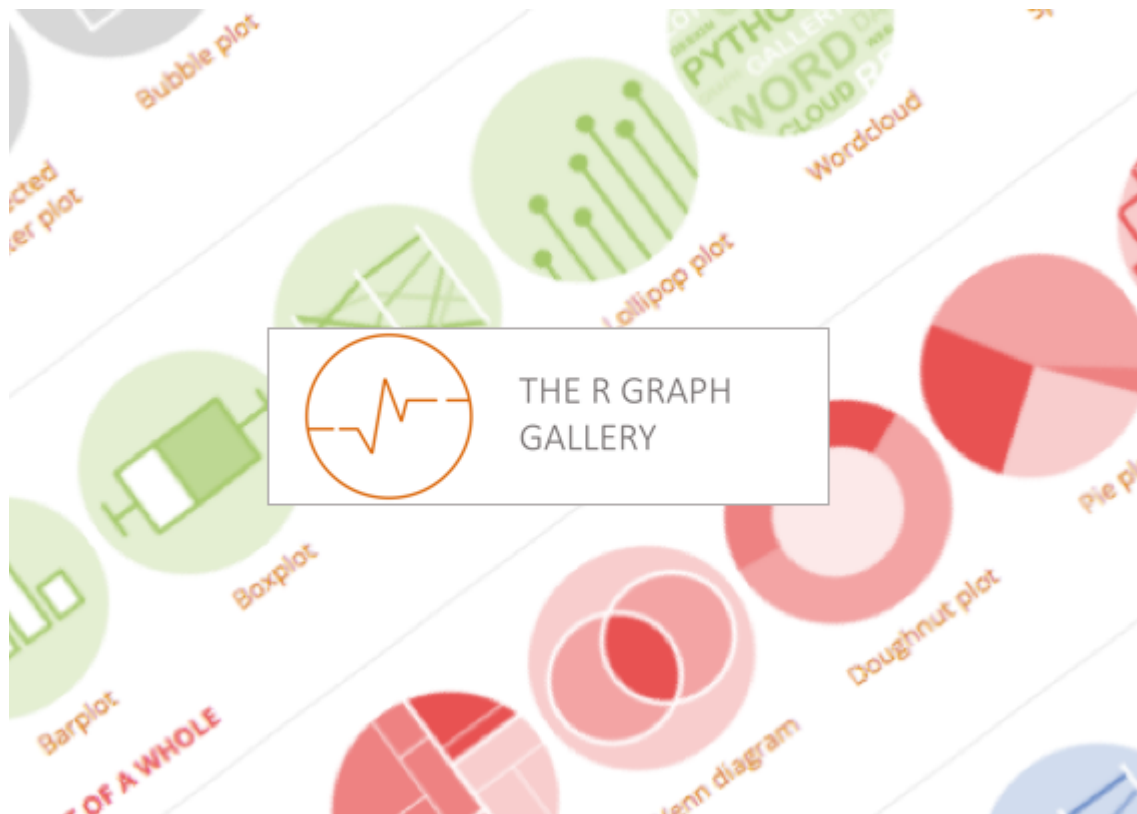
Paquetería ggplot2

La librería ggplot2 de R es un sistema organizado de visualización de datos. Forma parte del conjunto de librerías llamado **tidyverse**.

Los elementos necesarios para representar un gráfico con ggplot2 son los siguientes:

- Un **data** frame que contiene los datos que se quieren visualizar.
- Los **aesthetics**, es decir, una lista de relaciones entre las variables del fichero de datos y determinados aspectos del gráfico (como por ejemplo coordenadas, formas o colores).
- Los **geoms**, que especifican los elementos geométricos (puntos, líneas, círculos, etc) que se van a representar.
- Esta basado en **The grammar of graphics** de Leland Wilkinson.





Este enlace te llevará conocer más gráficos en R.

Grammar of Graphics by L. Wilkinson



Describes non-data ink: Design elements!

The plotting space you are using

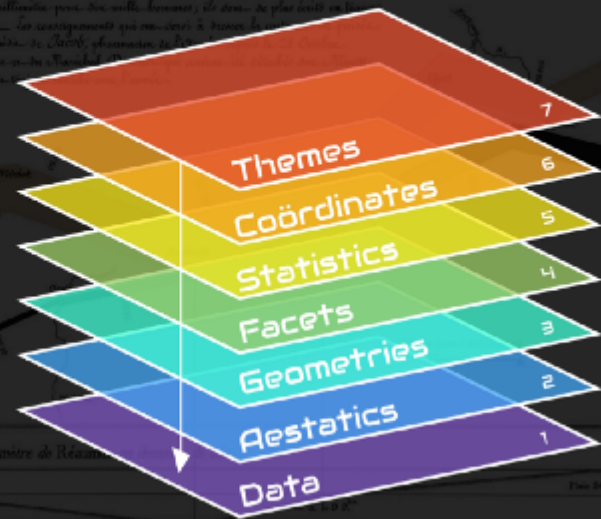
Statistical models & summaries

Rows and columns of sub-plots

Shapes used to represent your data

The scales on which the data is mapped

The actual variables to be plotted



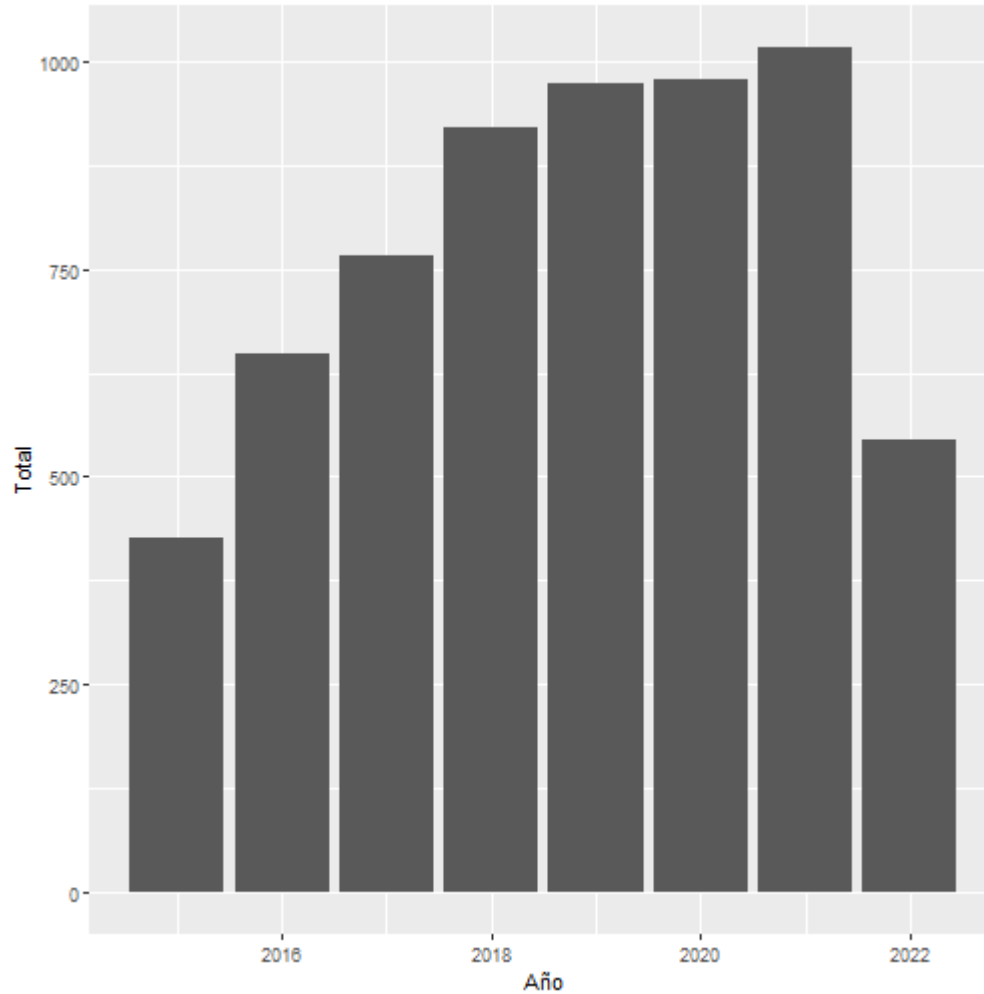
Gramática de gráficos

7 capas

1. **Datos:** Todo comienza aquí, con las variables trazadas que desea visualizar.
2. **Estética:** una capa de escalas (por ejemplo, ejes xey) en la que se asignan los datos.
3. **Geometría:** formas utilizadas para representar sus datos, por ejemplo, barras, líneas, puntos, etc.
4. **Facetas:** cuando hay demasiados puntos de datos en un gráfico, puede dividir las variables para obtener una vista más clara.
5. **Estadísticas:** datos agregados (resúmenes) o tendencias utilizando modelos estadísticos.
6. **Coordenadas:** describe el espacio de trazado que está utilizando.
7. **Temas:** describe la tinta sin datos. La capa de tema le permite diseñar con una identidad visual particular utilizando fuentes, colores y otros elementos de diseño.

La idea central de esta es que se puedan generar gráficos a partir de una sintaxis determinada.

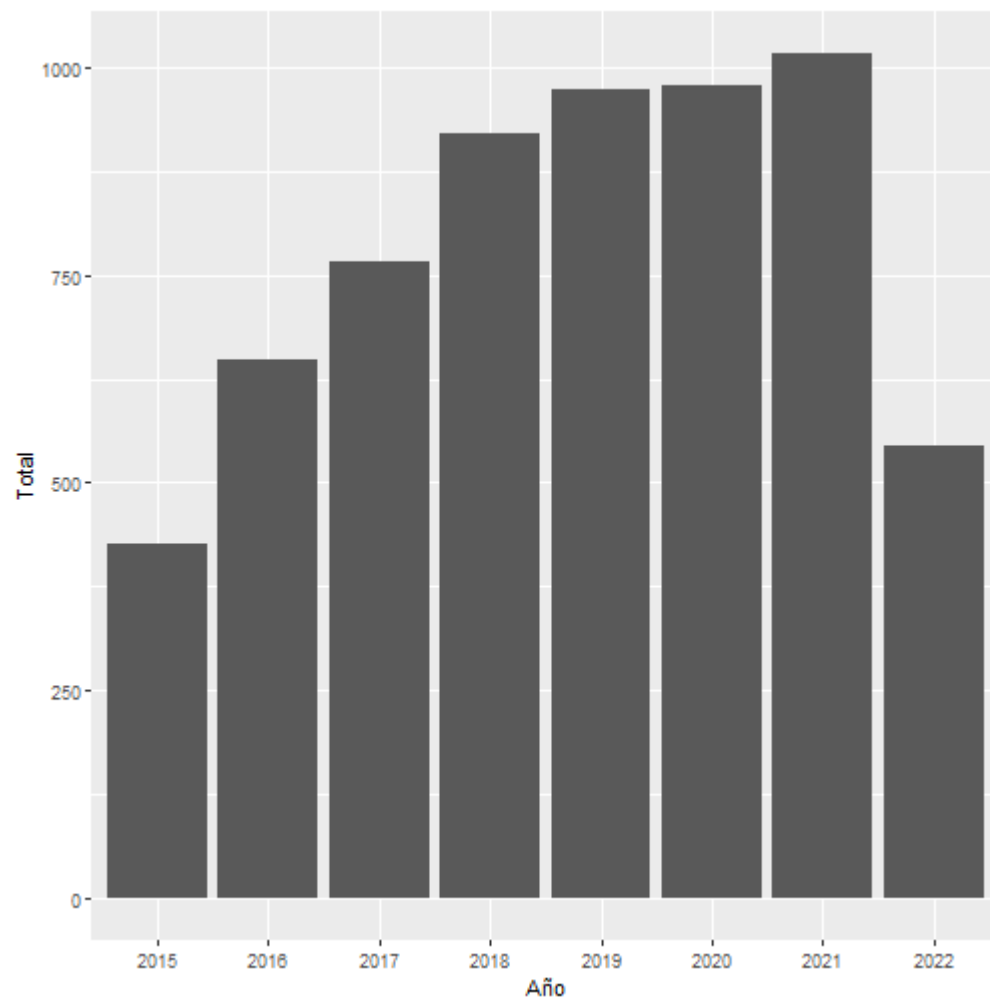
```
#Gráfico de columnas de feminicidios - Primeras 3 capas  
ggplot(fem_anual)+  
  aes(x=Año, y=Total)+  
  geom_col()
```



```
#Mostrar valor a cada columna de mi X transformo variable año a factor
fem_anual <- fem_anual %>%
  mutate(Año=factor(Año,
                     levels=c("2015", "2016", "2017", "2018", "2019", "2020", "2021", "2022")
  )
fem_anual
```

```
## # A tibble: 8 × 14
##   Año      ene    feb    mar    abr    may    jun    jul    ago    sep    oct    nov    dic
##   <fct> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>
## 1 2015      33     30     33     36     28     28     37     40     44     48     30     40
## 2 2016      41     54     70     63     52     60     43     46     57     48     63     50
## 3 2017      51     69     63     62     69     77     75     70     56     61     58     55
## 4 2018      71     69     72     82     64     79     84     65     79     85     68    101
## 5 2019      72     70     83     68     80     79     88     94     89     72     81     96
## 6 2020      75     92     77     73     75     97     77     75     82     82     89     84
## 7 2021      76     78    103     82    111     86     72    112     70     71     74     81
## 8 2022      79     81     75     83     85     88     54      0      0      0      0      0
## # ... with 1 more variable: Total <int>
```

```
#De nuevo el gráfico anterior  
ggplot(fem_anual)+  
  aes(x=Año, y=Total)+  
  geom_col()
```



```
#De nuevo el gráfico anterior
#Añadir relleno o color y etiquetas de valores a las columnas del gráfico
ggplot(fem_anual)+
  aes(x=Año, y=Total)+
  geom_col(fill = "#925cbf")+
  geom_text(aes(y = Total, label = Total, vjust = -0.5))
```

```
#De nuevo el gráfico anterior
#Añadir títulos y subtítulos al gráfico
ggplot(fem_anual)+
  aes(x=Año, y=Total)+
  geom_col(fill = "#925cbf")+
  geom_text(aes(y = Total, label = Total, vjust = -0.5))+
  labs(title = "Registros de feminicidios por año en México",
        subtitle = "Periodo 2015 - agosto 2022",
        caption = "Elaborado con los datos del Secretariado Ejecutivo del
        Sistema Nacional de Seguridad Pública (SESNSP))+
  ylab("Total de registro") +
  xlab("Año")
```

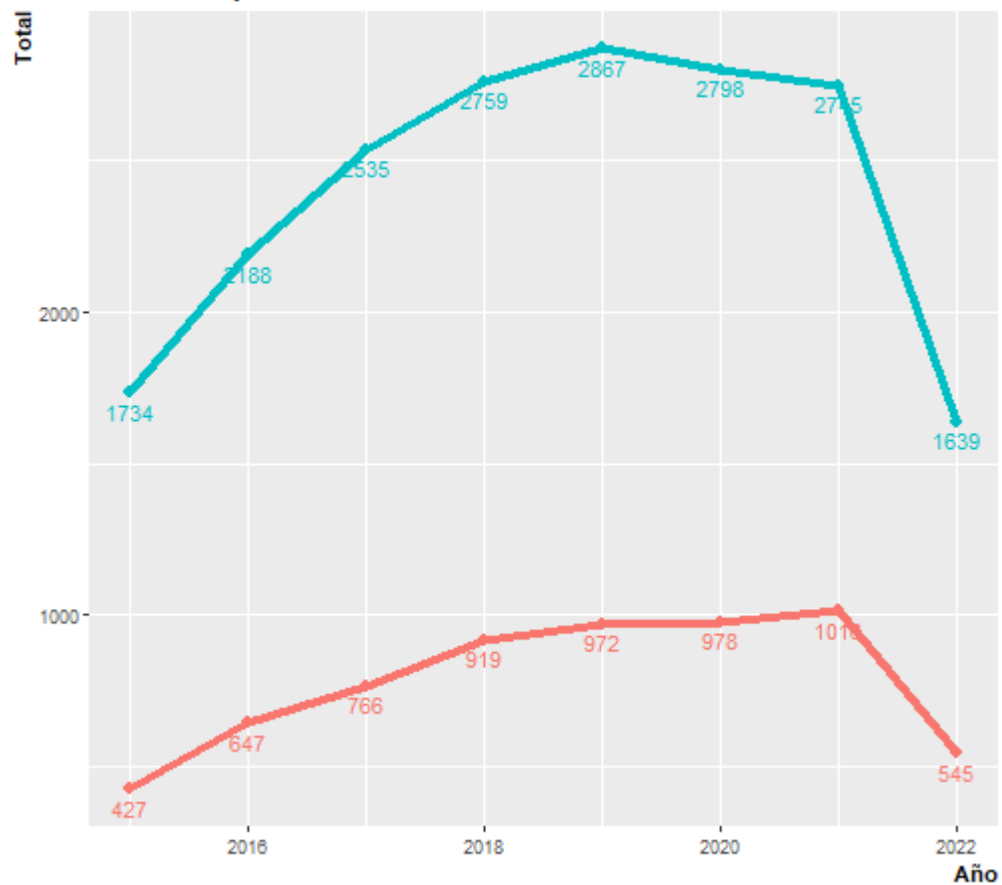
```
#Añadir theme
ggplot(fem_anual)+
  aes(x=Año, y=Total)+
  geom_col(fill = "#925cbf")+
  geom_text(aes(y = Total, label = Total, vjust = -0.5))+
  labs(title = "Registros de feminicidios por año en México",
        subtitle = "Periodo 2015 - agosto 2022",
        caption = "Elaborado con los datos del Secretariado Ejecutivo del
        Sistema Nacional de Seguridad Pública (SESNSP)")+
  ylab("Total de registro") +
  xlab("Año")+
  theme_minimal()
```


Gráfico sumando dos geometrias línea y punto

```
#Gráfico de línea de tendencia
graf_muertes_vio<-ggplot(data=victimas_anual) + #leer datos
  aes(x = Año, y = Total, color = Subtipo.de.delito) + #xy del gráfico
  geom_line(size = 1.5) + #figura línea
  geom_point(size = 2.5)+ #figura puntos
  geom_text(label=victimas_anual$Total, vjust=1.5)+ #etiqueta de valores
  labs(title = "Muertes violentas de mujeres",
        subtitle = "Periodo 2015 - junio 2022",
        caption = "Elaboración propia con datos del SESNSP")+ #Títulos y subtítulos
  theme(legend.position='none', #Sin leyenda
        plot.title=element_text(family='Anton', face='bold',size=20), #formato tex
        plot.subtitle=element_text(family='Anton'), #formato texto del subtítulo
        axis.title.x = element_text(family='Anton', face='bold', hjust=1), #format
        axis.title.y = element_text(family='Anton', face='bold', hjust=1)) #format
```

Muertes violentas de mujeres

Periodo 2015 - junio 2022

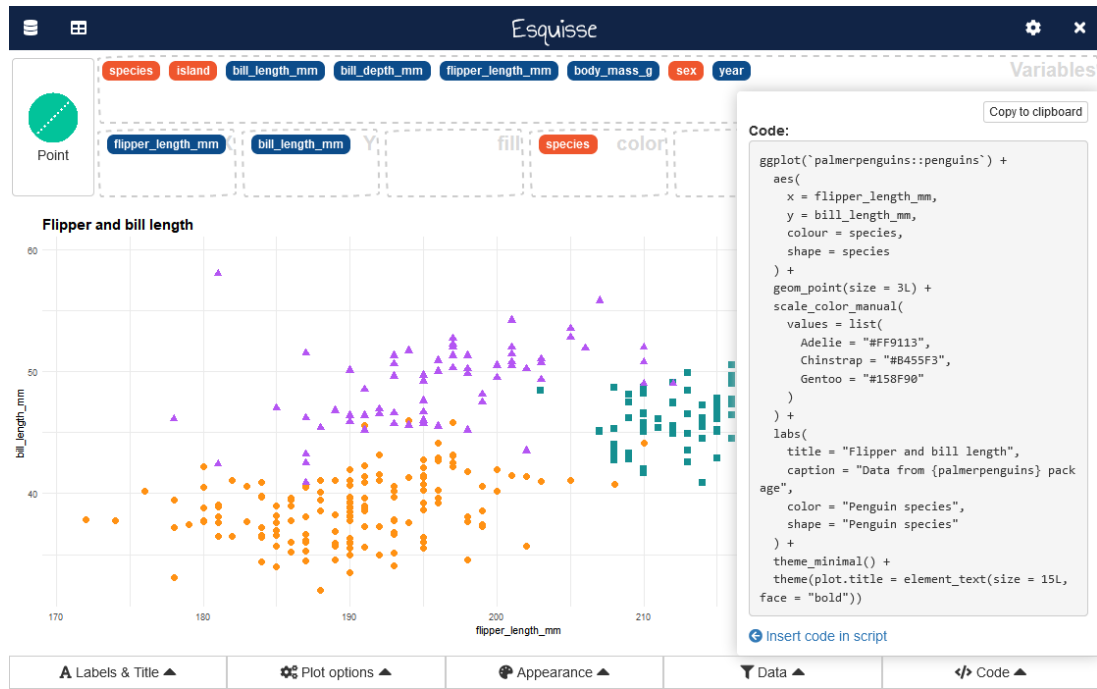


Elaboración propia con datos del SESNSP

Paquetería esquisse

Este complemento le permite explorar interactivamente sus datos visualizándolos con el paquete ggplot2 . Le permite dibujar diagramas de barras, curvas, diagramas de dispersión, histogramas, diagramas de caja y objetos sf , luego exportar el gráfico o recuperar el código para reproducir el gráfico.`{r} library(esquisse)`





Interfaz de esquisse