

# Taller R

## Sesion 7

Septiembre 2022

# Índice

- Pivot longer y pivot wider
- variables factor
- Capaz de gráficos
- Más gráficos

# Material de trabajo

Carpeta de drive:

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

Base a descargar ile\_completa

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

ID de reunión: 838 4169 2580

# Formato base de datos

Los datos pueden venir presentados en dos formatos: **largo** o **ancho**.

Los datasets con formato largo tienen pocas columnas y muchas filas, mientras los de formato ancho poseen muchas columnas y pocas filas.

Sin embargo, en ambas representaciones los datos son exactamente los mismos.

Representación de datos en formato largo y ancho.

# pivot\_wider y pivot\_longer

Tidyverse ofrece dos métodos que sirven para este propósito: `pivot_wider()` y `pivot_longer()` (estas funciones reemplazaron a `gather()` y a `spread()`, respectivamente).

- **`pivot_wider()`**: toma un conjunto de variables (vectores/columnas) y las colapsa en una sola columna con valores que resumen los datos de ese conjunto de variables. Hace que el data frame sea más largo.
- **`pivot_longer()`**: toma dos variables y las descompone entre múltiples variables (hace que el data frame sea más ancho).

Como pueden ver, no existe una medida absoluta de largo o ancho. Simplemente un dataset puede tener una representación más ancha o más larga.

Entonces... ¿las bases del SESNSP son formato largo o ancho?, ¿cómo conviene procesarla?

# En resumen

## wider a long

pais	anio	casos
Afganistán	1999	745
Afganistán	2000	2666
Brasil	1999	37737
Brasil	2000	80488
China	1999	212258
China	2000	213766

pais	1999	2000
Afganistán	745	2666
Brasil	37737	80488
China	212258	213766

Tabla 4

# Long a wider

**names\_from**      **values\_from**

plot_id	genus	mean_weight
1	Baiomys	7.00
2	Baiomys	6.00
3	Baiomys	8.61
1	Chaetodipus	22.20
2	Chaetodipus	25.11
3	Chaetodipus	24.64
1	Dipodomys	60.23
2	Dipodomys	55.68
3	Dipodomys	52.05

plot_id	Baiomys	Chaetodipus	Dipodomys
1	7.00	22.20	60.23
2	6.00	25.11	55.68
3	8.61	24.64	52.05

`data.frame`

column with new  
variable names

column of values  
for new  
variables

```
surveys_gw %>% pivot_wider(names_from = genus, values_from = mean_weight)
```

# Antes de comenzar recuerda:

## Llamar librerías:

```
#Cargar librerías
library(dplyr) # Manipulación de datos o funciones anidadas
library(tidyverse) # Paqueterías de tidy
library(ggplot2) # Generación de gráficos
library(esquisse) # Interfaz para generar gráfico de manera muy sencilla
library(readxl)
```

## Importar o cargar archivos:

```
base_2019 <- read_excel("G:/Mi unidad/Almacenamiento/Rstudio/Material/Sesion 7/201")
base_2020 <- read_excel("G:/Mi unidad/Almacenamiento/Rstudio/Material/Sesion 7/202")
base_2021 <- read_excel("G:/Mi unidad/Almacenamiento/Rstudio/Material/Sesion 7/202")
base_2022 <- read_excel("G:/Mi unidad/Almacenamiento/Rstudio/Material/Sesion 7/202")
```



```
#Dimensión de filas y columnas  
dim(base_2019)
```

```
## [1] 242158      21
```

```
#Nombre de columnas - 21 variables  
colnames(base_2019)
```

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

```
# Valores único de delitos que reportan por carpetas
unique(base_2019$`Tipo de delito`)
```

```
## [1] "Homicidio"
## [2] "Lesiones"
## [3] "Feminicidio"
## [4] "Aborto"
## [5] "Otros delitos que atentan contra la vida y la integridad corporal"
## [6] "Secuestro"
## [7] "Tráfico de menores"
## [8] "Rapto"
## [9] "Otros delitos que atentan contra la libertad personal"
## [10] "Abuso sexual"
## [11] "Acoso sexual"
## [12] "Hostigamiento sexual"
## [13] "Violación simple"
## [14] "Violación equiparada"
## [15] "Incesto"
## [16] "Otros delitos que atentan contra la libertad y la seguridad sexual"
## [17] "Robo"
## [18] "Fraude"
## [19] "Abuso de confianza"
## [20] "Extorsión"
## [21] "Daño a la propiedad"
## [22] "Despojo"
## [23] "Otros delitos contra el patrimonio"
## [24] "Violencia familiar"
## [25] "Violencia de género en todas sus modalidades distinta a la violencia familiar"
## [26] "Incumplimiento de obligaciones de asistencia familiar"
## [27] "Otros delitos contra la familia"
## [28] "Corrupción de menores"
## [29] "Trata de personas"
## [30] "Otros delitos contra la sociedad"
## [31] "Narcomenudeo"
```

# Renombrar delito

```
#Año 2019
```

```
base_2019$`Tipo de delito`[base_2020$`Tipo de delito`=="Violación equiparada"] <-  
base_2019$`Tipo de delito`[base_2020$`Tipo de delito`=="Violación simple"] <- "Vio
```

```
#Año 2020
```

```
base_2020$`Tipo de delito`[base_2020$`Tipo de delito`=="Violación equiparada"] <-  
base_2020$`Tipo de delito`[base_2020$`Tipo de delito`=="Violación simple"] <- "Vio
```

```
#Año 2021
```

```
base_2021$`Tipo de delito`[base_2021$`Tipo de delito`=="Violación equiparada"] <-  
base_2021$`Tipo de delito`[base_2021$`Tipo de delito`=="Violación simple"] <- "Vio
```

```
#Año 2022
```

```
base_2022$`Tipo de delito`[base_2022$`Tipo de delito`=="Violación equiparada"] <-  
base_2022$`Tipo de delito`[base_2022$`Tipo de delito`=="Violación simple"] <- "Vio
```

```
#Vector con los delitos de interes
delitos<-c("Abuso sexual", "Feminicidio", "Violación", "Violencia familiar")
```

## Año 2019

```
#Filtar para 2019 los delitos de interés en Jalisco

base_2019 %>%
  filter(Entidad=="Jalisco",
         `Tipo de delito` %in% delitos) %>%
  group_by(Año, `Tipo de delito`, Municipio) %>%
  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))->Municipal_jal_2019
```

## Año 2020

```
#Filtrar para 2020 los delitos de interés en Jalisco

base_2020 %>%
  filter(Entidad=="Jalisco",
         `Tipo de delito` %in% delitos) %>%
  group_by(Año, `Tipo de delito`, Municipio) %>%
  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))->Municipal_jal_2020
```

## Año 2021

```
#Filtrar para 2021 los delitos de interés en Jalisco

base_2021 %>%
  filter(Entidad=="Jalisco",
         `Tipo de delito` %in% delitos) %>%
  group_by(Año, `Tipo de delito`, Municipio) %>%
  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))->Municipal_jal_2021
```

## Año 2022

```
#Filtrar para 2022 los delitos de interés en Jalisco

base_2022 %>%
  filter(Entidad=="Jalisco",
         `Tipo de delito` %in% delitos) %>%
  group_by(Año, `Tipo de delito`, Municipio) %>%
  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)) ->Municipal_jal_2022
```

# rbind

La función `rbind` en R, abreviatura de row-bind , se puede usar para combinar vectores, matrices y marcos de datos por filas.

Unir los 3 objetos para tener todo el histórico desde 2019 hasta julio 2022

Deben tener el mismo nombre y número de columnas

```
rbind(Municipal_jal_2019, Municipal_jal_2020, Municipal_jal_2021, Municipal_jal_20
```



# Exportar datos en csv

Para exportar una base de datos que tenemos en el environment podemos hacerlo con la función **write.csv**, el comando es el siguiente, primero va el nombre de la base en el environment, luego el nombre del archivo como quieres que se guarde en tu laptop, y por último el encoding.

Mi archivo se llama **base\_jalisco\_delitos.csv**

```
write.csv(base_jalisco, "base_jalisco_delitos.csv",  
          fileEncoding="UTF8")
```

# Pivot

Formato PIVOT de base para hacer bases anchas a largas. Se reduce variables

```
base_jalisco %>%  
  pivot_longer("ene":"dic", names_to = "Mes")->base_jalisco_largo
```

# Transformación

Agrupamos y contamos la violencia familiar por año en Jalisco

```
#Gráfico por facet de año violencia familiar
```

```
#Agrupar y sumar violencia familiar
```

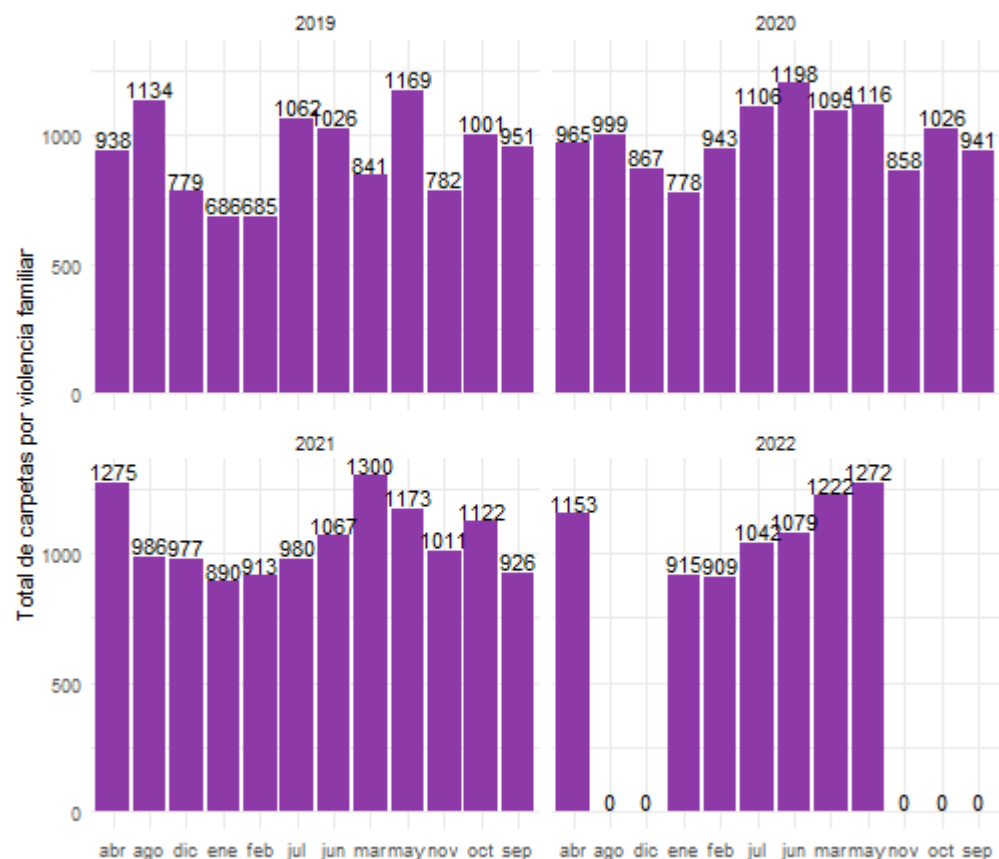
```
base_jalisco_largo %>%  
  filter(`Tipo de delito`=="Violencia familiar") %>%  
  group_by(Mes, Año) %>%  
  summarise(total=sum(value))->base_vf
```

```
#Gráfico de barra
```

```
ggplot(base_vf) +  
  aes(x = Mes, y = total) +  
  geom_col(fill = "#8e39a8") +  
  theme_minimal() +  
  geom_text(aes(label = total, vjust=-0.2))+  
  labs(title = "Registros de violencia familiar en la entidad de Jalisco",  
        subtitle = "Periodo 2019 - julio 2022",  
        caption = "Elaborado con los datos del Secretariado Ejecutivo del Sistema N  
ylab("Total de carpetas por violencia familiar") +  
  xlab("")+  
  facet_wrap(vars(Año))->graf_vf
```

## Registros de violencia familiar en la entidad de Jalisco

Periodo 2019 - julio 2022



Elaborado con los datos del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (SESNSP)

# Variable mes a factor

Con factor se busca que los meses ya tengan el orden habitual que conocemos y no por orden alfabetico como lo hace por default R.

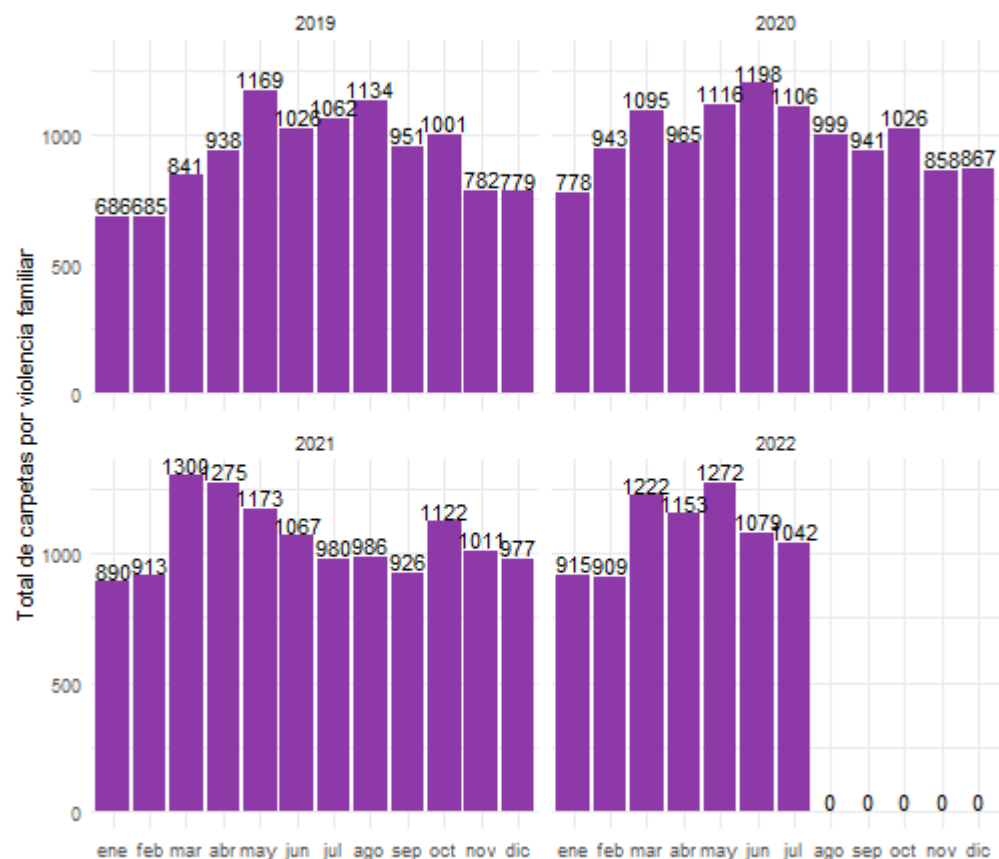
```
#Tratar meses a factor
base_vf <- base_vf %>%
  mutate(Mes=factor(Mes,
                     levels=c("ene", "feb", "mar", "abr", "may", "jun", "jul",
                              "ago", "sep", "oct", "nov", "dic")))
```

# Volver a ejecutar el gráfico

```
ggplot(base_vf) +  
  aes(x = Mes, y = total) +  
  geom_col(fill = "#8e39a8") +  
  theme_minimal() +  
  geom_text(aes(label = total, vjust=-0.2))+  
  labs(title = "Registros de violencia familiar en la entidad de Jalisco",  
        subtitle = "Periodo 2019 - julio 2022",  
        caption = "Elaborado con los datos del Secretariado Ejecutivo del Sistema N.  
ylab("Total de carpetas por violencia familiar") +  
  xlab("")+  
  facet_wrap(vars(Año))->graf_vf
```

## Registros de violencia familiar en la entidad de Jalisco

Periodo 2019 - julio 2022



Elaborado con los datos del Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública (SESNSP)

# Abuso sexual

```
base_jalisco_largo %>%  
  filter(`Tipo de delito`=="Abuso sexual") %>%  
  group_by(Mes, Año) %>%  
  summarise(total=sum(value))->asi
```

## `summarise()` has grouped output by 'Mes'. You can override using the `.groups`  
## argument.

## Vemos el resultado

```
asi
```

```
## # A tibble: 48 × 3  
## # Groups:   Mes [12]  
##   Mes      Año total  
##   <chr> <dbl> <dbl>  
## 1 abr    2019   229  
## 2 abr    2020   176  
## 3 abr    2021   224  
## 4 abr    2022   255  
## 5 ago    2019   226  
## 6 ago    2020   179  
## 7 ago    2021   202  
## 8 ago    2022     0  
## 9 dic    2019   170  
## 10 dic   2020   145
```



# Gráfico abuso sexual en barra y facet

```
ggplot(asi) +  
  aes(x = Mes, y = total) +  
  geom_col(fill = "#a61f70") +  
  theme_minimal() +  
  geom_text(aes(label = total, vjust=-0.2))+  
  labs(title = "Registros de abuso sexual en la entidad de Jalisco",  
        subtitle = "Periodo 2019 - julio 2022",  
        caption = "Elaborado con los datos del Secretariado Ejecutivo del Sistema N.  
ylab("Total de carpetas por abuso sexual") +  
  xlab("")+  
  facet_wrap(vars(Año))->graf_asi
```

# Gráfico asi

graf\_asi

# Factor mes

```
#Tratar meses a factor
asi <- asi %>%
  mutate(Mes=factor(Mes,
                     levels=c("ene", "feb", "mar", "abr", "may", "jun", "jul",
                              "ago", "sep", "oct", "nov", "dic")))
```

```
#Gráfico de barra asi
ggplot(asi) +
  aes(x = Mes, y = total) +
  geom_col(fill = "#a61f70") +
  theme_minimal() +
  geom_text(aes(label = total, vjust=-0.2))+
  labs(title = "Registros de abuso sexual en la entidad de Jalisco",
        subtitle = "Periodo 2019 - julio 2022",
        caption = "Elaborado con los datos del Secretariado Ejecutivo del Sistema N",
        ylab("Total de carpetas por abuso sexual") +
        xlab(""))+
  facet_wrap(vars(Año))->graf_asi
```

# Gráfico mensual de abuso sexual

graf\_asi

# Abuso sexual anual

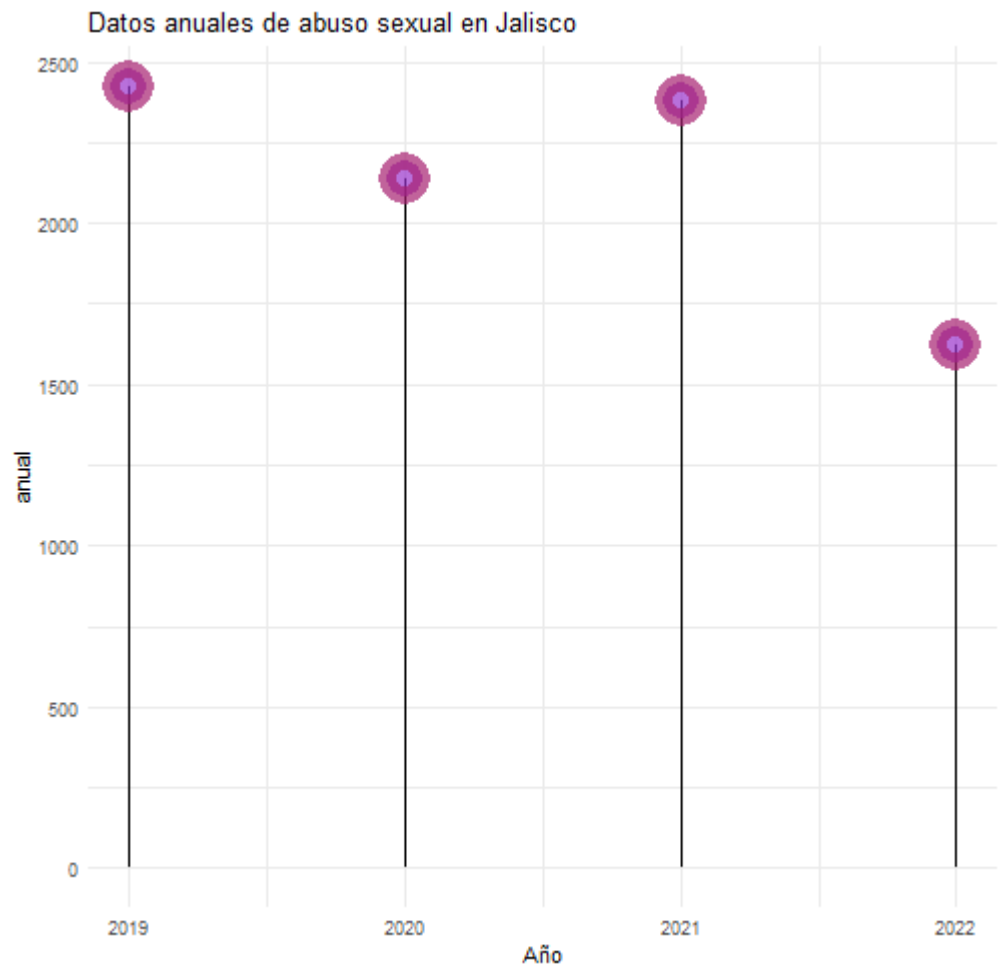
```
#ASI anual
asi %>%
  group_by(Año) %>%
  summarise(anual=sum(total))->asi_anual
```

```
asi_anual
```

```
## # A tibble: 4 × 2
##   Año anual
##   <dbl> <dbl>
## 1  2019  2423
## 2  2020  2137
## 3  2021  2378
## 4  2022  1625
```

# Lollipop

```
#Gráfico lollipop
ggplot(asi_anual)+
  aes(x=Año, y=anual)+
  geom_segment( aes(x=Año, xend=Año, y=0, yend=anual)) +
  geom_point( size=5, color="#a61f70", fill=alpha("darkorchid", 0.3), alpha=0.7, s
  labs(title="Datos anuales de abuso sexual en Jalisco",
        caption = "Elaboración propia, con datos del SESNSP")+
  theme_minimal()->asi_lollipop
```



# Gráfico punto y linea anual

```
### Gráfico linea y punto
asi_anual %>%
  ggplot( aes(x=Año, y=anual)) +
  geom_line( color="grey") +
  geom_point(shape=21, color="black", fill="#69b3a2", size=6)+
  theme_minimal()+
  labs(title="Abuso sexual en Jalisco",
        caption= "Con datos de SESNSP")+
  ylab("Registro de carpetas")+
  xlab("")->linea_asi
```



linea\_asi

