

```
# Trabajo final - 23 a 26 Septiembre 2023
```

```
install.packages("readxl")  
library(readxl)
```

```
# =====  
# Los pasos a continuación hacen referencia a la preparacion de la data con el  
# archivo .xlsx que se encontraba en plataforma. Esto con el fin de evidenciar  
# las dificultades al decodificar al español
```

```
# Importar los datos  
datos <- read_xlsx("violencia.xlsx", sheet = 1, col_names = FALSE)  
head(datos)
```

```
# Obtener nombre columnnas  
columns_name <- strsplit(datos[[1]][1], ',')  
columns_name[[1]][2:length(columns_name[[1]])]  
columns_name <- columns_name[[1]][2: length(columns_name[[1]])]  
columns_name
```

```
# Dividir los datos en columnas  
datos <- as.character(datos)  
columnas <- strsplit(datos[[1]], ",")  
columnas_clean_na <- na.omit(columnas)  
head(columnas_clean_na)  
# columnas_clean_na[[4]][9]  
length(columnas_clean_na)  
# Formatear cada columna  
for (i in 1:length(columnas_clean_na)) {  
  columnas_clean_na[[i]][9] <- as.numeric(columnas_clean_na[[i]][9])  
  columnas[[i]] <- as.factor(columnas_clean_na[[i]])  
  columnas_clean_na[[i]][6] <- as.Date(columnas_clean_na[[i]][6])  
  columnas_clean_na[[i]][2] <- enc2utf8(as.character(columnas_clean_na[[i]][2]))  
  columnas_clean_na[[i]][3] <- enc2utf8(as.character(columnas_clean_na[[i]][3]))  
}
```

```
head(columnas_clean_na)  
# Combinar las filas formateadas en un único vector  
datos <- do.call(rbind, columnas_clean_na)  
# crear data frame  
datos <- data.frame(datos)
```

```
# Visualizar los datos  
head(datos)
```

```
# remover columna 1 generada automaticamente con NA o error  
datos <- datos[, -c(1)]  
head(datos)
```

```
# remover datos con NA  
datos <- na.omit(datos)  
head(datos)  
summary(datos)
```

```
# resetear enumeacion de fila  
rownames(datos) <- 1:nrow(datos)
```

```
# renombrar columnas  
colnames(datos) <- columns_name
```

```

head(datos)
tail(datos)
summary(datos)
datos

# apply encoding utf-8 to "departamento" & "municipio" columns
colnames(datos)
# [1] "departamento" "municipio" "codigo_dane" "armas_medios" "fecha_hecho"
# [6] "genero" "grupo_etario" "cantidad"
length(datos[1,1])
datos[1,"departamento"]
datos[1,"municipio"]
for (i in 1:length(datos[,1])) {
  print(i)
  datos[i,"departamento"] <- enc2utf8(as.character(datos[i,"departamento"]))
  datos[i,"municipio"] <- enc2utf8(as.character(datos[i,"municipio"]))
}

head(datos)

# =====

#####
##### RESOLUCION TRABAJO FINAL #####
#####

##### PREPARACION DE DATA #####

# Importacion datos
data_main <- read.csv("violencia.csv", encoding="UTF-8", fileEncoding="UTF-8")
head(data_main)

# remover primera columna con enumeracion desde 0
data_main <- data_main[, -c(1)]
head(data_main)
tail(data_main)

# Eliminar filas con NA's
data_main <- na.omit(data_main)

# se corrige errores de formato o espaciado
data_main$departamento <- gsub(" ", " ", data_main$departamento)
data_main$municipio <- gsub(" ", " ", data_main$municipio)
data_main$armas_medios <- gsub(" ", " ", data_main$armas_medios)
head(data_main)

# Conversion de tipos de datos
data_main$departamento <- as.factor(data_main$departamento)
data_main$municipio <- as.factor(data_main$municipio)
data_main$codigo_dane <- as.factor(data_main$codigo_dane)
data_main$armas_medios <- as.factor(data_main$armas_medios)
data_main$fecha_hecho <- as.Date(
  data_main$fecha_hecho, format = "%d/%m/%Y"
)
data_main$genero <- as.factor(data_main$genero)
data_main$grupo_etario <- as.factor(data_main$grupo_etario)
data_main$cantidad <- as.numeric(data_main$cantidad)
# verificar data
head(data_main)

```

```

# Extraer año de columna fecha_hecho para analizar data por año
data_main["anio"] <- format(
  data_main$fecha_hecho, format = "%Y"
)
# verificar data
head(data_main)

# carga de librerias requeridas
library(dplyr)
library(ggplot2)
library(plotly)

# filtrado departamento
violencia_valle <- data_main %>% filter(
  departamento == "VALLE"
)
head(violencia_valle)

#####
##### CUESTIONARIO #####
#####

# ¿Cómo ha variado el número de casos reportados de violencia doméstica
# en los municipios del departamento asignado entre el primer y último año
# disponible en la base de datos?. Realice su visualización e Interprete
# los resultados

# Se calcula el número de casos reportados por año
violencia_valle_reportados <- violencia_valle %>%
  group_by(anio) %>%
  summarize(cantidad = sum(cantidad))

summary(violencia_valle_reportados)

# Gráfico
valle_reportados_grafico <- violencia_valle_reportados %>%
  ggplot(
    aes(x = anio, y = cantidad,
        fill = anio,
        text = sprintf("Cantidad: %s", cantidad)
    )
  ) +
  geom_bar(stat = "sum", position = "dodge") +
  labs(title = "Número de casos reportados de violencia doméstica en el departamento del
VALLE por año",
        x = "Año",
        y = "Número de casos",
        fill = "Año", size = NULL
  ) +
  theme_minimal() +
  theme(
    axis.text = element_text(size = 7, angle = 90),
    legend.position = "bottom"
  )

valle_reportados_grafico

ggplotly(valle_reportados_grafico, tooltip = "text")

```

```

# Compare la distribución del número de víctimas para el primer y último
# año de la base de datos. Responda esta pregunta con un gráfico interactivo,
# muestre en el hover la leyenda "Cantidad:" y el valor correspondiente.
# Interprete sus resultados.

# Se filtra los datos para el primer y último año

# Año 2010
violencia_valle_2010 <- violencia_valle[violenicia_valle$anio == 2010, ] %>%
  group_by(grupo_etario) %>%
  summarise(cantidad = sum(cantidad))
# head(violencia_valle_2010)
summary(violencia_valle_2010)

# Año 2023
violencia_valle_2023 <- violencia_valle[violenicia_valle$anio == 2023, ] %>%
  group_by(grupo_etario) %>%
  summarise(cantidad = sum(cantidad))
# head(violencia_valle_2023)
summary(violencia_valle_2023)

# Gráfico de distribucion

# Año 2010
violencia_valle_2010_grafico <- violencia_valle_2010 %>%
  ggplot(
    aes(
      x = grupo_etario, y = cantidad,
      color = grupo_etario,
      text = sprintf("Cantidad: %s", cantidad)
    )
  ) +
  geom_segment(
    aes(x = grupo_etario, xend = grupo_etario,
      y = 0, yend = cantidad
    )
  ) +
  geom_point(aes( size = 25)) +
  theme_minimal() +
  labs(title = "Distribución del número de víctimas en el Valle según grupo etario en
2010",
    x = "Grupo etario",
    y = "Número de víctimas",
    colour = "Grupo etario", size = NULL
  )
violencia_valle_2010_grafico
ggplotly(violencia_valle_2010_grafico, tooltip = "text")

# Año 2023
violencia_valle_2023_grafico <- violencia_valle_2023 %>%
  ggplot(
    aes(
      x = grupo_etario, y = cantidad,
      color = grupo_etario,
      text = sprintf("Cantidad: %s", cantidad)
    )
  ) +
  geom_segment(

```

```

    aes(x = grupo_etario, xend = grupo_etario,
        y = 0, yend = cantidad
    )
  ) +
  geom_point(aes( size = 25)) +
  theme_minimal() +
  labs(title = "Distribución del número de víctimas en el Valle según grupo etario en
2023",
        x = "Grupo etario",
        y = "Número de víctimas",
        colour = "Grupo etario", size = NULL
    )

```

```

violencia_valle_2023_grafico
ggplotly(violencia_valle_2023_grafico, tooltip = "text")

```

```

# ¿Qué municipios del departamento asignado reportaron más hechos violentos
# en el último año? Realice su visualización e interprete sus resultados.

```

```

# Se calcula el número de hechos violentos por municipio
violencia_valle_municipios_2023 <- violencia_valle_2023 %>%
  group_by(municipio) %>%
  summarise(cantidad = sum(cantidad))

```

```

# Se ordena los datos por número de casos
violencia_valle_municipios_2023 <- violencia_valle_municipios_2023[
  order(
    violencia_valle_municipios_2023$cantidad, decreasing = TRUE
  ),
]

```

```

summary(violencia_valle_municipios_2023)

```

```

# Se muestra el top 10 de municipips
top_valle_municipios_2023 <- violencia_valle_municipios_2023[1:10,]
top_valle_municipios_2023

```

```

# Gráfico
top_valle_municipios_2023_grafico <- top_valle_municipios_2023 %>%
  ggplot(
    aes(
      x = municipio, y = cantidad, color = municipio,
      text = sprintf("Cantidad: %s<br>Municipio: %s",
        cantidad, municipio)
    )
  ) +
  geom_segment(
    aes(x = municipio, xend = municipio,
        y = 0, yend = cantidad)
  ) +
  geom_point(aes( size = 25,)) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(size = 7, angle = 45),
    legend.position = "bottom"
  ) +
  labs(title = "Municipios con más hechos violentos del Valle en 2023",

```

```

      x = "Municipio",
      y = "Número de hechos violentos",
      color = "Municipio", size = NULL
    )

top_valle_municipios_2023_grafico

ggplotly(top_valle_municipios_2023_grafico, tooltip = "text")

#### Adicional
# Listado top Sin municipio == Cali
top_valle_municipios_2023[top_valle_municipios_2023$municipio != "CALI (CT)", ]

# Gráfico
top_valle_municipios_2023_no_cali_grafico <- top_valle_municipios_2023[
  top_valle_municipios_2023$municipio != "CALI (CT)",
] %>%
  ggplot(
    aes(
      x = municipio, y = cantidad, color = municipio,
      text = sprintf("Cantidad: %s<br>Municipio: %s",
                      cantidad, municipio)
    )
  ) +
  geom_segment(
    aes(x = municipio, xend = municipio,
        y = 0, yend = cantidad)
  ) +
  geom_point(aes( size = 25,)) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(size = 7, angle = 45),
    legend.position = "bottom"
  ) +
  labs(title = "Municipios con más hechos violentos del Valle en 2023 - Sin Cali (CT)",
       x = "Municipio",
       y = "Número de hechos violentos",
       color = "Municipio", size = NULL
  )

top_valle_municipios_2023_no_cali_grafico

ggplotly(top_valle_municipios_2023_no_cali_grafico, tooltip = "text")


# Muestre la composición de hechos violentos para el último año en el
# departamento asignado. Interprete sus resultados. Realice en este caso
# un treemap.

library(treemapify)

# Se Obtiene armas/medios empleados en violencia domestica
medios_violencia_valle_2023 <- violencia_valle[violenca_valle$anio == 2023, ] %>%
  group_by(armas_medios, municipio) %>%
  summarise(cantidad = sum(cantidad))

summary(medios_violencia_valle_2023)

# Gráfico treemap

```

```
medios_violencia_valle_2023_grafico <- medios_violencia_valle_2023 %>%
  ggplot(
    aes(
      area = cantidad, fill = armas_medios
    )
  ) +
  geom_treemap() +
  geom_treemap_text(
    aes(
      label = paste0(
        municipio, "\n",
        "Cant.: ", cantidad, "\n",
        "(",
        round(cantidad / sum(medios_violencia_valle_2023$cantidad) * 100, 2),
        "%)"
      ),
      size = 8,
      color = "black",
      position = "identity",
    ) +
    labs(title = "Composición de hechos violentos en el departamento del Valle en 2023",
          x = "",
          y = "",
          fill = "Armas / medios"
    )
  )
```

```
medios_violencia_valle_2023_grafico
```

```
# Composicion hechos violentos en el departamento por municipio
violencia_valle[violenca_valle$anio == 2023, ] %>%
  group_by(armas_medios) %>%
  summarise(cantidad = sum(cantidad)) %>%
  ggplot(
    aes(
      area = cantidad, fill = armas_medios,
    )
  ) +
  geom_treemap() +
  geom_treemap_text(
    aes(
      label = paste0(
        #municipio, "\n",
        "Cant.: ", cantidad, "\n",
        "(",
        round(cantidad / sum(medios_violencia_valle_2023$cantidad) * 100, 2),
        "%)"
      ),
      size = 8,
      color = "black",
      position = "identity",
    ) +
    labs(title = "Composición de hechos violentos en el departamento del Valle en 2023",
          x = "",
          y = "",
          fill = "Armas / medios"
    )
  )
```

```
# Muestre la evolución de la cantidad de víctimas en la capital del departamento
# asignado a través de un gráfico de líneas interactivo
```

```
# Se filtra datos para la capital del departamento del valle
violencia_cali <- violencia_valle[violenicia_valle$municipio == "CALI (CT)", ] %>%
  group_by(anio, armas_medios) %>%
  summarize(cantidad = sum(cantidad))
#head(violencia_cali)
summary(violencia_cali)
```

```
# Gráfico de líneas
violencia_cali_grafico <- violencia_cali %>%
  ggplot(
    aes(
      x = anio, y = cantidad,
      text = sprintf(
        "Año: %s<br>Cantidad: %s<br>Arma/medio: %s",
        anio, cantidad, armas_medios
      )
    ) +
    geom_point(aes(x = anio, y= cantidad, color = armas_medios)) +
    geom_line(aes(group = armas_medios, color = armas_medios)) +
    labs(title = "Evolución de la cantidad de víctimas en Cali",
      x = "Año",
      y = "Cantidad de víctimas",
      color = "Armas/medios"
    ) +
    theme_light()

violencia_cali_grafico

ggplotly(violencia_cali_grafico, tooltip = "text")
```