

# Análisis de Estadísticas Delictivas (Delitos Sexuales) en Colombia (2018-2024)

Centro Nacional Colombo Aleman SENA Sede TIC- Grupo TAD-7

2024-09-18

## Introducción

En este análisis estadístico, se examinan los 18 tipos de delitos reportados en Colombia desde el año 2018 hasta la fecha actual de Agosto de 2024, utilizando las bases de datos oficiales proporcionadas por la Policía Nacional. El estudio se enfoca en comparar las tasas de criminalidad a lo largo de diferentes periodos para evaluar las variaciones en la incidencia y distribución de cada delito bajo distintas administraciones.

Mediante un análisis descriptivo, se identifican tendencias y patrones criminales tanto a nivel nacional como específicamente en el municipio de Barranquilla y en todo el departamento del Atlántico. Los resultados obtenidos están destinados a informar a ciudadanos, empresas, entidades y servidores públicos, brindándoles una visión integral del comportamiento delictivo y facilitando la implementación de estrategias de prevención más efectivas. Así, el estudio contribuye a un entendimiento más profundo de los desafíos que enfrenta Colombia en términos de seguridad y justicia.

## Contexto del Estudio

En Colombia, las lesiones personales son un problema grave y persistente que afecta a diversas poblaciones. La información obtenida del sistema SIEDCO ha sido analizada para proporcionar una visión más clara sobre cómo varían los casos reportados en términos de género, edad y ubicación geográfica. Este análisis es crucial para identificar áreas con mayores tasas de incidencia y para evaluar cómo las intervenciones y políticas actuales están influyendo en la reducción de estos delitos.

## Metodología

Los datos utilizados en este estudio se han extraído de la página oficial de la Policía Nacional de Colombia y abarcan el período de 2018 a agosto de 2024. Se han analizado varias dimensiones de los datos:

1. **Distribución por Género:** Se ha calculado la cantidad total de víctimas por género y se ha visualizado mediante un gráfico de pastel para mostrar la proporción de víctimas masculinas y femeninas.
2. **Distribución por Edad:** Se ha determinado la cantidad total de víctimas en diferentes rangos de edad, con gráficos de pastel que reflejan el porcentaje de víctimas en cada grupo etario.
3. **Tendencias Anuales:** Se ha analizado la cantidad de abusos reportados cada año para identificar tendencias a lo largo del tiempo, utilizando gráficos de líneas para representar las fluctuaciones anuales.
4. **Distribución Geográfica:** Se ha examinado la cantidad de reportes por departamento y municipio, con un enfoque especial en el Atlántico y Barranquilla. Gráficos de barras y pasteles ilustran la concentración de casos en las principales áreas.

# Resultados

## Análisis

```
library(readxl)
library(ggplot2)
library(dplyr)
library(tidyr)

# Leer el archivo Excel
file_path <- "C:/Users/ASOU/Desktop/SENA/R/LESION_TRANSITO/BD_LESION_TRANSITO.xlsx"
BD_LESION_TRANSITO <- read_excel(file_path, sheet = "BD_LESION_TRANSITO")
```

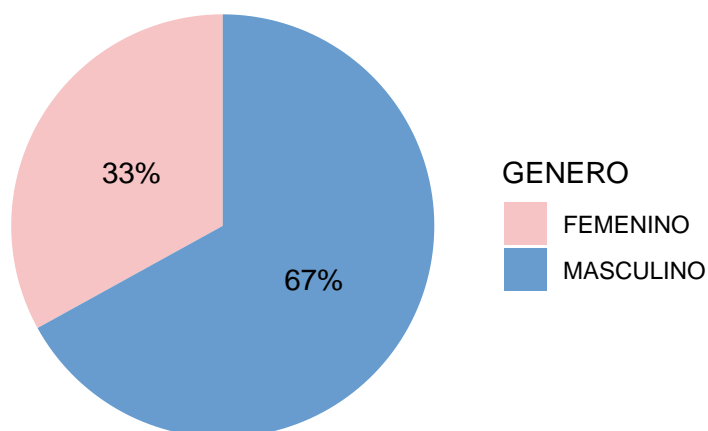
## Distribución por Género

En esta sección, el análisis estadístico de lesiones personales examina la proporción de víctimas desglosada por género. Esta visualización tiene como objetivo proporcionar una comprensión clara de cómo se distribuyen las víctimas de lesiones personales según su género.

```
# Filtrar los datos para eliminar "N/D"
datos_filtrados <- BD_LESION_TRANSITO %>%
  filter(GENERO != "N/D")
# Calcular la cantidad de lesiones por género
datos_genero_lesion_transito <- datos_filtrados %>%
  group_by(GENERO) %>%
  summarise(CANTIDAD_TOTAL = sum(CANTIDAD))

datos_genero_lesion_transito <- datos_genero_lesion_transito %>%
  mutate(PORCENTAJE = CANTIDAD_TOTAL / sum(CANTIDAD_TOTAL) * 100)
```

## Porcentaje de lesiones por género



La gráfica de pastel que presentamos ilustra la distribución de las lesiones personales en Colombia entre 2018 y 2024, segmentada por género. De acuerdo con los datos, se observa que el 33% de las lesiones personales corresponde a personas de género femenino, mientras que el 67% corresponde a personas de género masculino.

Este análisis sugiere que los hombres son más propensos a sufrir lesiones personales en comparación con las mujeres durante el período mencionado. Varios factores podrían contribuir a esta tendencia, como patrones de movilidad, comportamientos en situaciones de riesgo y la exposición a circunstancias que aumentan la probabilidad de lesiones. Es posible que los hombres, en general, estén más expuestos a situaciones que resultan en lesiones, lo que podría llevar a un mayor número de casos reportados.

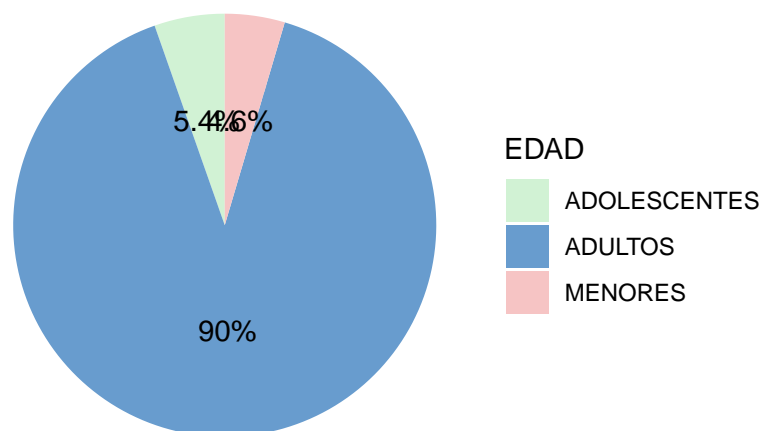
## Distribución por Edad

Se analiza la prevalencia de lesiones personales entre diferentes rangos etarios, utilizando un gráfico de pastel para representar visualmente los datos. Esta gráfica muestra cómo se distribuyen las víctimas de lesiones personales según sus grupos de edad: - Menores: entre 0años a 12 años - Adolescentes: entre 13 años y 17 años - Adultos: 18 años en adelante Agrupación referente a la clasificación del código de infancia y adolescencia ley 1098 de del 8 de noviembre de 2006, en su artículo 3.

```
# Filtrar datos para excluir las filas donde EDAD es "N/D"
datos_filtrados_edad <- BD_LESION_TRANSITO %>%
  filter(EDAD != "N/D")
# Calcular la cantidad de lesiones por edad
datos_edad_lesiones_transito <- datos_filtrados_edad %>%
  group_by(EDAD) %>%
  summarise(CANTIDAD_TOTAL = sum(CANTIDAD), .groups = 'drop')

datos_genero_lesion_transito <- datos_filtrados %>%
  group_by(GENERO) %>%
  summarise(CANTIDAD_TOTAL = sum(CANTIDAD))
# Calcular el porcentaje
datos_edad_lesiones_transito <- datos_edad_lesiones_transito %>%
  mutate(PORCENTAJE = CANTIDAD_TOTAL / sum(CANTIDAD_TOTAL) * 100)
# Crear el gráfico de pastel con porcentaje
ggplot(datos_edad_lesiones_transito, aes(x = "", y = CANTIDAD_TOTAL, fill = EDAD)) +
  geom_bar(stat = "identity", width = 1) + # Crear barras de tamaño 1 para gráfico de pastel
  coord_polar("y", start = 0) + # Convertir el gráfico de barras en un gráfico de pastel
  labs(title = "Porcentaje de lesiones por rango de edad") + # Título del gráfico
  theme_void() + # Eliminar fondo y ejes
  scale_fill_manual(values = c("#D1F2D4", "#689cce", "#F6C4C4")) + # Especificar colores, ajusta seg
  geom_text(aes(label = paste0(round(PORCENTAJE, 1), "%")), # Etiquetas con porcentajes
    position = position_stack(vjust = 0.5)) +
  theme(plot.title = element_text(hjust = 0.5)) # Centrar el título
```

## Porcentaje de lesiones por rango de edad



La gráfica de pastel que se presenta muestra la distribución de lesiones personales en Colombia entre 2018 y 2024, segmentada por rango de edad. Los datos indican que el 5.4% de las lesiones corresponden a

adolescentes, el 4.6% a menores, y un notable 90% a adultos.

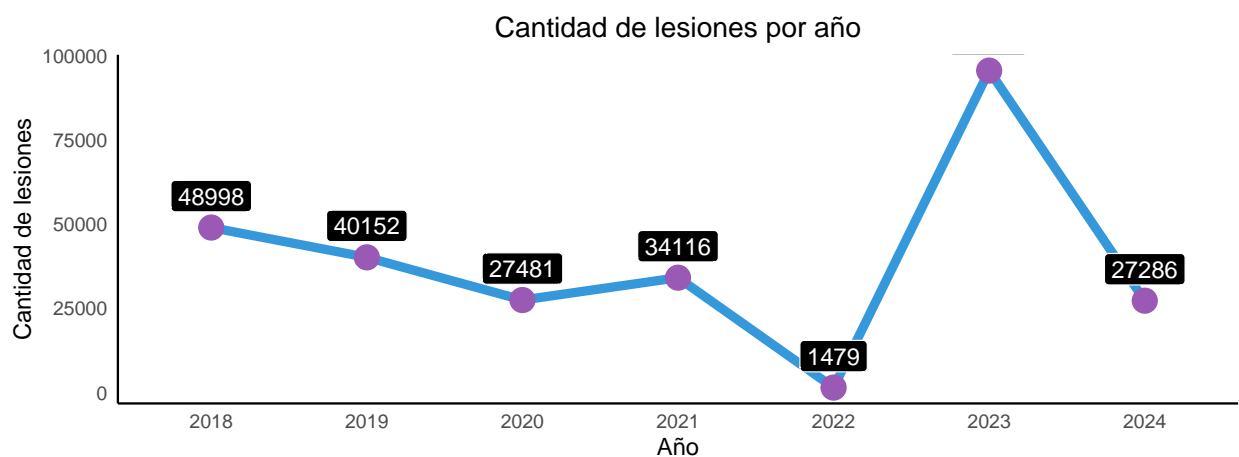
Esta gráfica revela que una abrumadora mayoría de las lesiones personales afecta a la población adulta. Este hallazgo sugiere que los adultos son los más afectados por las lesiones personales, lo que podría estar relacionado con varios factores, como patrones de actividad, mayor exposición a situaciones de riesgo, y diferencias en el comportamiento en comparación con los menores y adolescentes

## Tendencias Anuales

```
BD_LESION_TRANSITO <- BD_LESION_TRANSITO %>%
  mutate(AÑO = as.factor(AÑO))

# Agrupar los datos por año y sumar la cantidad de abusos
datos_año_lesiones_transito <- BD_LESION_TRANSITO %>%
  group_by(AÑO) %>%
  summarise(CANTIDAD_TOTAL = sum(CANTIDAD, na.rm = TRUE))

# Crear el gráfico de línea
ggplot(datos_año_lesiones_transito, aes(x = AÑO, y = CANTIDAD_TOTAL)) +
  geom_line(aes(group = 1), color = "#3498db", size = 2) + # Aumentar el grosor de la línea a 2
  geom_point(color = "#9b59b6", size = 5) + # Ajustar el tamaño de los puntos
  geom_label(aes(label = CANTIDAD_TOTAL), # Agregar etiquetas con las cantidades
    vjust = -0.5, # Ajustar verticalmente para evitar superposición con los puntos
    hjust = 0.5, # Ajustar horizontalmente
    fill = "black", # Fondo blanco para las etiquetas
    color = "white") + # Color del texto
  labs(title = "Cantidad de lesiones por año",
    x = "Año",
    y = "Cantidad de lesiones") +
  scale_x_discrete(breaks = unique(datos_año_lesiones_transito$AÑO), # Asegurarse de que todos los años estén representados
    labels = unique(datos_año_lesiones_transito$AÑO)) + # Usar los mismos valores para las etiquetas
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5), # Centrar el título
    panel.grid.major = element_blank(), # Eliminar las líneas de cuadrícula mayores
    panel.grid.minor = element_blank(), # Eliminar las líneas de cuadrícula menores
    panel.border = element_blank(), # Eliminar el borde del panel
    axis.line = element_line(color = "black") # Opcional: Añadir línea del eje si es necesario
  )
)
```



La gráfica muestra la evolución del número de lesiones personales en Colombia entre 2018 y 2024, destacando las siguientes tendencias:

Tendencia Descendente Inicial (2018-2020): Durante los años 2018 a 2020, se observa una disminución gradual en el número de lesiones, con una reducción notable del 2018 al 2020. Este descenso podría haber sido influenciado por diversos factores, incluyendo políticas de prevención o cambios en la conducta social.

Aumento en 2021 y 2023: A partir de 2021, se registra un aumento en el número de lesiones, posiblemente relacionado con la recuperación de la movilidad post-pandemia y la reactivación de actividades sociales y económicas. Este aumento puede reflejar un retorno a niveles más altos de actividad y exposición a riesgos.

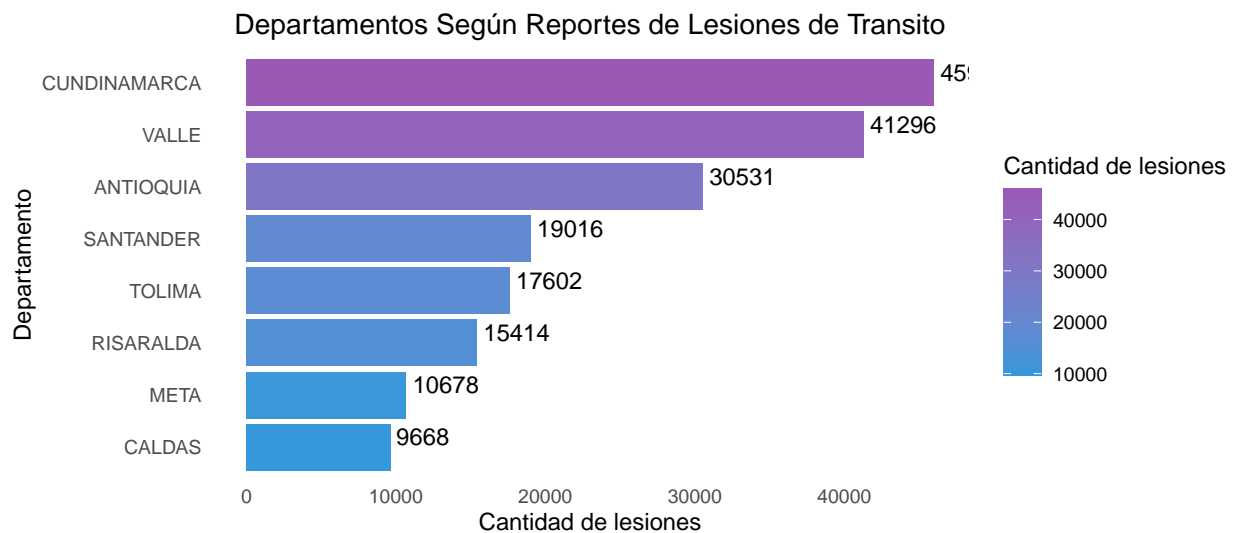
Disminución Drástica en 2022 y Recuperación Incompleta en 2024: La disminución drástica en 2022, con solo 1,479 lesiones, es notable y podría deberse a un año atípico, cambios en la metodología de reporte, o una disminución real en los incidentes. La recuperación incompleta en 2024 sugiere que el número de lesiones no ha vuelto a los niveles previos al 2022, indicando la necesidad de investigar más a fondo las causas subyacentes.



## Distribución Geográfica para Atlántico

```
# Agrupar por DEPARTAMENTO y sumar CANTIDAD
# Seleccionar los 10 departamentos con mayor cantidad de lesiones
top_departamentos <- BD_LESION_TRANSITO %>%
  group_by(DEPARTAMENTO) %>%
  summarise(CANTIDAD_TOTAL = sum(CANTIDAD, na.rm = TRUE)) %>%
  arrange(desc(CANTIDAD_TOTAL)) %>%
  slice_max(order_by = CANTIDAD_TOTAL, n = 8)

# Crear gráfico de barras
ggplot(top_departamentos, aes(x = reorder(DEPARTAMENTO, CANTIDAD_TOTAL), y = CANTIDAD_TOTAL, fill = CANTIDAD_TOTAL)) +
  geom_bar(stat = "identity") + # Crear barras
  geom_text(aes(label = CANTIDAD_TOTAL), # Etiquetas con las cantidades
            vjust = 0.5, # Ajustar verticalmente para centrar en la barra
            hjust = -0.1, # Ajustar horizontalmente para que el texto quede al lado de la barra
            color = "black", # Color del texto
            nudge_x = 0.2) + # Desplazar el texto un poco hacia la derecha
  scale_fill_gradient(low = "#3498db", high = "#9b59b6") + # Degradado de azul a rosa
  coord_flip() + # Girar el gráfico
  labs(title = "Departamentos Según Reportes de Lesiones de Transito",
       x = "Departamento",
       y = "Cantidad de lesiones",
       fill = "Cantidad de lesiones") + # Etiqueta de la leyenda
  theme_minimal() + # Usar un tema limpio para una mejor presentación
  theme(
    plot.title = element_text(hjust = 0.5), # Centrar el título
    panel.grid.major = element_blank(), # Eliminar las líneas de cuadrícula mayores
    panel.grid.minor = element_blank(), # Eliminar las líneas de cuadrícula menores
    panel.border = element_blank() # Eliminar el borde del panel
  )
```



El mayor numero de casos son de Cundinamarca: Con 45,984 lesiones, Cundinamarca es el departamento con el mayor número de reportes de lesiones en accidentes de tránsito. Esto indica que Cundinamarca, al ser uno de los departamentos más poblados y con una gran actividad económica, enfrenta una alta incidencia de lesiones en accidentes de tránsito.

Valle del Cauca: Valle del Cauca sigue a Cundinamarca con una diferencia relativamente estrecha en el número de reportes. A partir de este punto, se observa una disminución significativa en las cifras de lesiones, reduciéndose en aproximadamente 10,000 reportes y cayendo a menos de 20,000 en los departamentos siguientes. Este patrón sugiere que, aunque Valle del Cauca también enfrenta un número elevado de lesiones, la incidencia en otros departamentos es considerablemente menor.

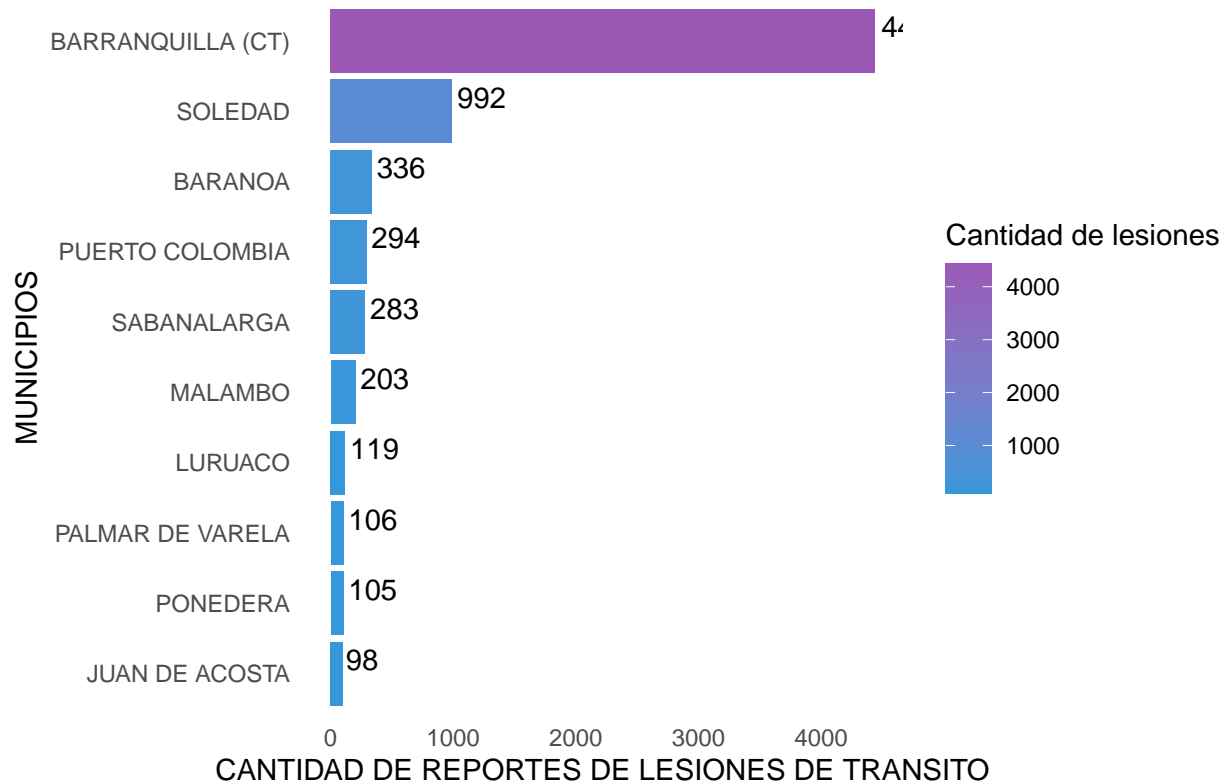
## Atlántico

```
#TOP 10 DE MUNICIPIOS DEL ATLANTICO CON MAYOR REPORTES
#GRAFICO EN BARRA

# Filtrar datos para el Departamento del Atlántico
df_lesiones_transito_Atlantico <- BD_LESION_TRANSITO %>%
  filter(DEPARTAMENTO == "ATLÁNTICO") %>% # Filtrar por el departamento del Atlántico
  group_by(MUNICIPIO) %>% # Agrupar por municipio
  summarise(CANTIDAD_TOTAL = sum(CANTIDAD, na.rm = TRUE)) %>% # Sumar la cantidad total de abusos por municipio
  arrange(desc(CANTIDAD_TOTAL)) %>% # Ordenar en orden descendente
  slice_max(order_by = CANTIDAD_TOTAL, n = 10) # Seleccionar los 10 municipios con mayor cantidad

# Crear gráfico de barras
ggplot(df_lesiones_transito_Atlantico, aes(x = reorder(MUNICIPIO, CANTIDAD_TOTAL), y = CANTIDAD_TOTAL, fill = fill)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.7)) + # Crear barras
  geom_text(aes(label = CANTIDAD_TOTAL), # Etiquetas con las cantidades
            vjust = 0.5, # Ajustar verticalmente para centrar en la barra
            hjust = -0.1, # Ajustar horizontalmente para que el texto quede al lado de la barra
            color = "black", # Color del texto
            nudge_x = 0.2) + # Desplazar el texto un poco hacia la derecha
  scale_fill_gradient(low = "#3498db", high = "#9b59b6") + # Escala de colores
  coord_flip() + # Girar el gráfico para una mejor legibilidad
  labs(title = "TOP 10 MUNICIPIOS CON MAYOR REPORTES DE LESIONES DE TRANSITO EN EL ATLÁNTICO",
       x = "MUNICIPIOS",
       y = "CANTIDAD DE REPORTES DE LESIONES DE TRANSITO",
       fill = "Cantidad de lesiones") + # Etiqueta de la leyenda
  theme_minimal() + # Usar un tema limpio para una mejor presentación
  theme(
    plot.title = element_text(hjust = 0.5), # Centrar el título
    panel.grid.major = element_blank(), # Eliminar las líneas de cuadrícula mayores
    panel.grid.minor = element_blank(), # Eliminar las líneas de cuadrícula menores
    panel.border = element_blank(), # Eliminar el borde del panel
  )
```

## MUNICIPIOS CON MAYOR REPORTES DE LESIONES DE TRANSITO EN EL ATL



**Dominio de Barranquilla:** La gráfica revela una amplia diferencia en el número de lesiones reportadas entre Barranquilla y otros municipios. Barranquilla, al ser una ciudad significativamente más grande y con una mayor densidad poblacional, presenta cifras de lesiones que superan ampliamente a las de otros municipios.

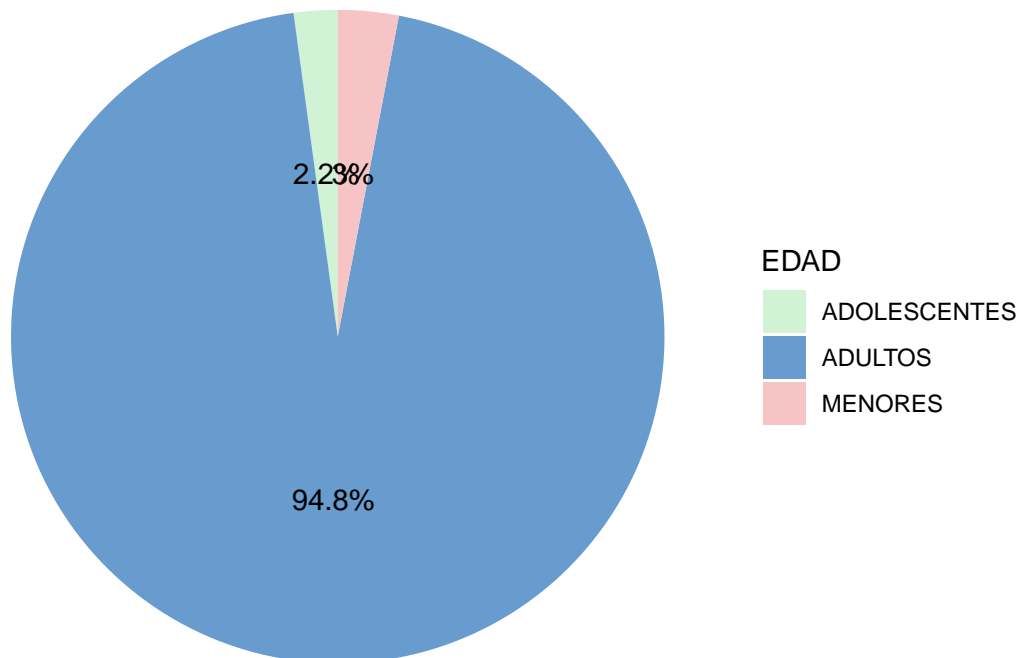
**Comparación con Otros Municipios:** La gran disparidad en los datos sugiere que Barranquilla, debido a su tamaño y población, experimenta un mayor número de lesiones en accidentes de tránsito en comparación con municipios de menor tamaño. Esta diferencia destaca la correlación entre la densidad de población y la incidencia de lesiones en accidentes.

## Datos Barranquilla

```
# Filtrar los datos para la ciudad de Barranquilla
df_lesiones_transito_Barranquilla <- BD_LESION_TRANSITO %>%
  filter(MUNICIPIO == "BARRANQUILLA (CT)")
# Filtrar datos para excluir las filas donde EDAD es "N/D"
datos_filtrados_edad_barranquilla <- df_lesiones_transito_Barranquilla %>% filter(EDAD != "N/D")
# Calcular la cantidad de víctimas por edad
datos_edad_lesiones_transito_barranquilla <- datos_filtrados_edad_barranquilla %>%
  group_by(EDAD) %>%
  summarise(CANTIDAD_TOTAL = sum(CANTIDAD), .groups = 'drop')
# Calcular el porcentaje
datos_edad_lesiones_transito_barranquilla <- datos_edad_lesiones_transito_barranquilla %>%
  mutate(PORCENTAJE = CANTIDAD_TOTAL / sum(CANTIDAD_TOTAL) * 100)

# Crear el gráfico de pastel con porcentaje
ggplot(datos_edad_lesiones_transito_barranquilla, aes(x = "", y = CANTIDAD_TOTAL, fill = EDAD)) +
  geom_bar(stat = "identity", width = 1) + # Crear barras de tamaño 1 para gráfico de pastel
  coord_polar("y", start = 0) + # Convertir el gráfico de barras en un gráfico de pastel
  labs(title = "Porcentaje de lesiones por rango de edad en Barranquilla") + # Título del gráfico
  theme_void() + # Eliminar fondo y ejes
  scale_fill_manual(values = c("#D1F2D4", "#689cce", "#F6C4C4")) + # Especificar colores
  geom_text(aes(label = paste0(round(PORCENTAJE, 1), "%")), # Etiquetas con porcentajes
    position = position_stack(vjust = 0.5)) +
  theme(plot.title = element_text(hjust = 0.5)) # Centrar el título
```

## Porcentaje de lesiones por rango de edad en Barranquilla



Concentración en Adultos: Los datos presentados en la gráfica muestran un patrón consistente con el observado a nivel regional. Al igual que en el resto del Atlántico, los adultos son el grupo etario que concentra la mayor parte de las lesiones por accidentes de tránsito. Esto sugiere que, independientemente de las diferencias locales, la tendencia general en el Atlántico es que los adultos son los más afectados por estos incidentes.

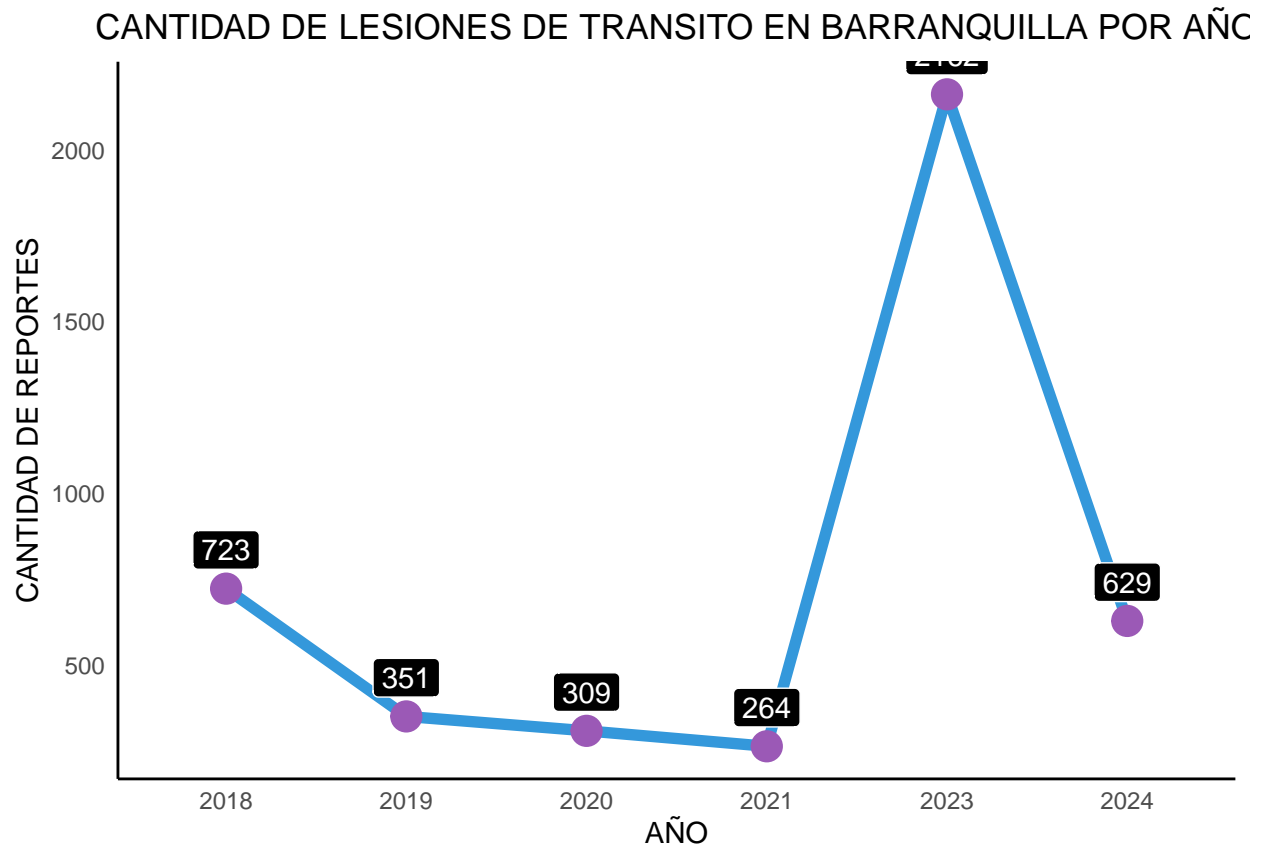
## Comportamiento a través del tiempo en Barranquilla

```
#GRAFICO LINEAL

# Agrupar los datos por año y sumar la cantidad de abusos
datos_año_lesiones_transito_Barranquilla <- df_lesiones_transito_Barranquilla %>%
  group_by(AÑO) %>%
  summarise(CANTIDAD_TOTAL = sum(CANTIDAD, na.rm = TRUE))

# Asegúrate de que AÑO sea un factor para la visualización
datos_año_lesiones_transito_Barranquilla<- datos_año_lesiones_transito_Barranquilla %>%
  mutate(AÑO = as.factor(AÑO))

# Crear el gráfico de línea
ggplot(datos_año_lesiones_transito_Barranquilla, aes(x = AÑO, y = CANTIDAD_TOTAL)) +
  geom_line(aes(group = 1), color = "#3498db", size = 2) + # Aumentar el grosor de la línea a 2
  geom_point(color = "#9b59b6", size = 5) + # Ajustar el tamaño de los puntos si es necesario
  geom_label(aes(label = CANTIDAD_TOTAL), # Agregar etiquetas con las cantidades
    vjust = -0.5, # Ajustar verticalmente para evitar superposición con los puntos
    hjust = 0.5, # Ajustar horizontalmente
    fill = "black", # Fondo blanco para las etiquetas
    color = "white") + # Color del texto
  labs(title = "CANTIDAD DE LESIONES DE TRANSITO EN BARRANQUILLA POR AÑO",
    x = "AÑO",
    y = "CANTIDAD DE REPORTE") +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5), # Centrar el título
    panel.grid.major = element_blank(), # Eliminar las líneas de cuadrícula mayores
    panel.grid.minor = element_blank(), # Eliminar las líneas de cuadrícula menores
    panel.border = element_blank(), # Eliminar el borde del panel
    axis.line = element_line(color = "black")# Opcional: Añadir línea del eje si es necesario
  )
```



La gráfica muestra una tendencia fluctuante en la cantidad de lesiones de tránsito reportadas en Barranquilla entre los años 2018 y 2024. Se observa un aumento significativo en el número de lesiones reportadas en el año 2023. Sin embargo, a pesar de este pico en 2023, la tendencia general desde 2018 hasta 2021 ha sido a la baja en el número de lesiones.



## Conclusión

El análisis de las estadísticas delictivas sobre lesiones en accidentes de tránsito en Colombia revela:

**Distribución por Género:** El 67% de las lesiones afectan a hombres y el 33% a mujeres. Esto sugiere que los hombres son más propensos a sufrir lesiones en accidentes de tránsito.

**Distribución por Edad:** El 90% de las lesiones afectan a adultos, mientras que solo el 5.4% y el 4.6% corresponden a adolescentes y menores, respectivamente.

**Tendencias Anuales:** Se observa una disminución de lesiones entre 2018 y 2020, un aumento en 2021 y 2023, y una disminución drástica en 2022, posiblemente por un cambio en la metodología de reporte.

**Distribución Geográfica:** Cundinamarca tiene el mayor número de reportes de lesiones, con Valle en segundo lugar. Los números disminuyen significativamente en otros departamentos.

**Barranquilla:** Muestra una tendencia fluctuante con un notable aumento en 2023, pero una tendencia general a la baja desde 2018 hasta 2021.

Estos hallazgos ayudan a identificar áreas críticas y a formular estrategias para mejorar la seguridad vial en Colombia. Estos resultados evidencian la necesidad de reforzar las políticas de protección, especialmente en las áreas con mayor incidencia, y sugieren que se deben implementar más campañas de sensibilización orientadas a reducir la violencia de género.

## Referencias

Policía Nacional de Colombia. Acceso a Datos Estadísticos.