

ANÁLISIS ESTADÍSTICO Y VISUAL DE DELITOS EN COSTA RICA CON PYTHON

AUTORES

Andrey Calderon Vega

Melanie Gómez Rueda

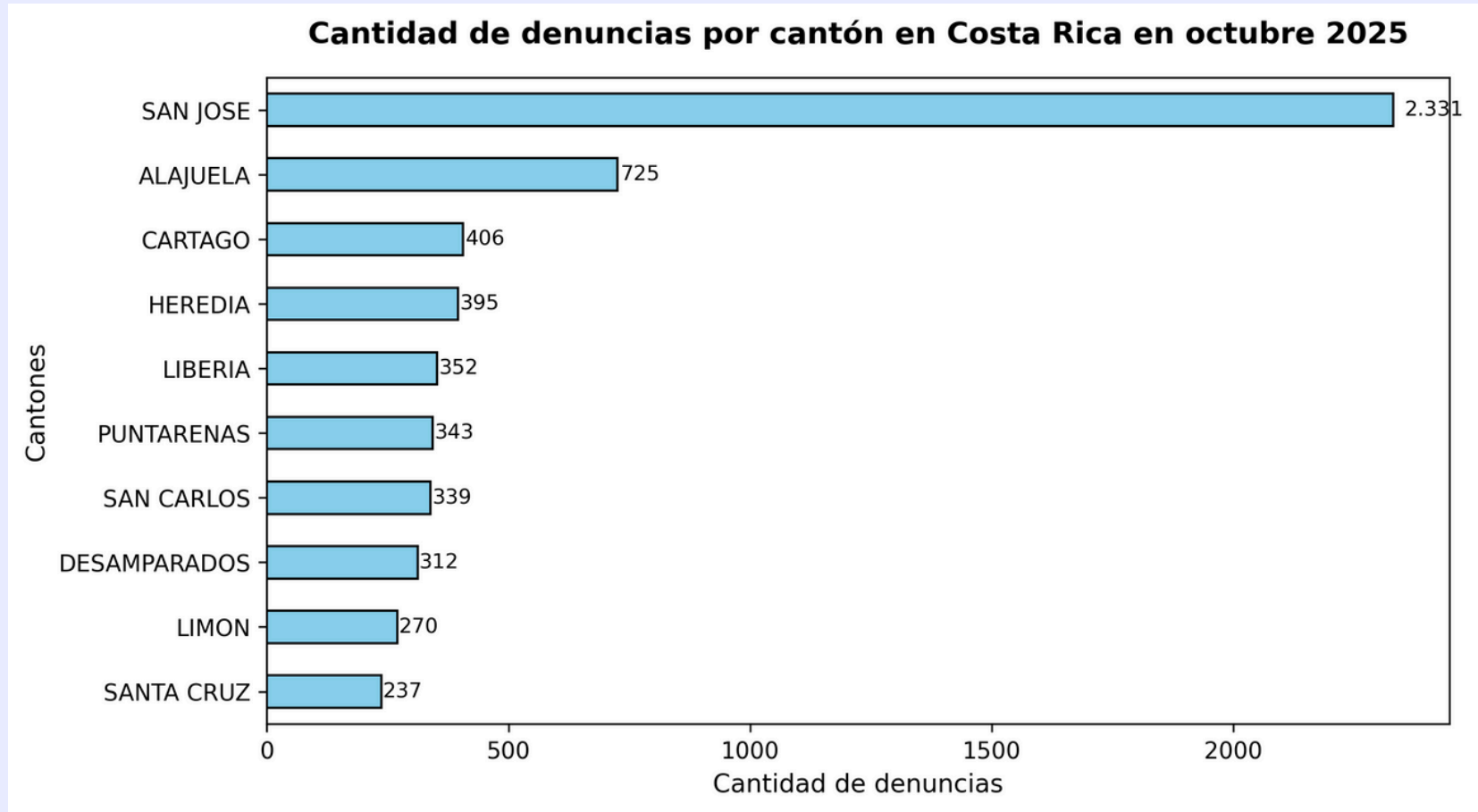
1) GRAFICO DE BARRAS HORIZONTALES

Definición:

El gráfico de barras horizontales es una representación visual que muestra y compara cantidades entre diferentes categorías mediante rectángulos dispuestos de manera horizontal.

Muestra:

Se utilizaron los registros de denuncias agrupados por cantón, considerando únicamente los 10 cantones con mayor cantidad de casos del conjunto de datos.



Uso:

Se utiliza cuando se desea comparar varias categorías o regiones y destacar cuáles tienen mayor o menor frecuencia. El formato horizontal es ideal cuando los nombres de las categorías son largos o cuando se necesita mejorar la legibilidad en comparación con las barras verticales.

Codigo:

```
if serie.empty():
    print("[AVISO] No hay datos para graficar por cantón. Revisa el archivo CSV.")
else:
    # Crear gráfico de barras horizontales
    ax = serie.sort_values().plot(kind="barh", figsize=(9, 5), color="skyblue",
    edgecolor="black")
    ax.set_title("Denuncias por Cantón (Top 10)", fontsize=13, weight="bold")
    ax.set_xlabel("Cantidad de Denuncias", fontsize=11)
    ax.set_ylabel("Cantones", fontsize=11)
```

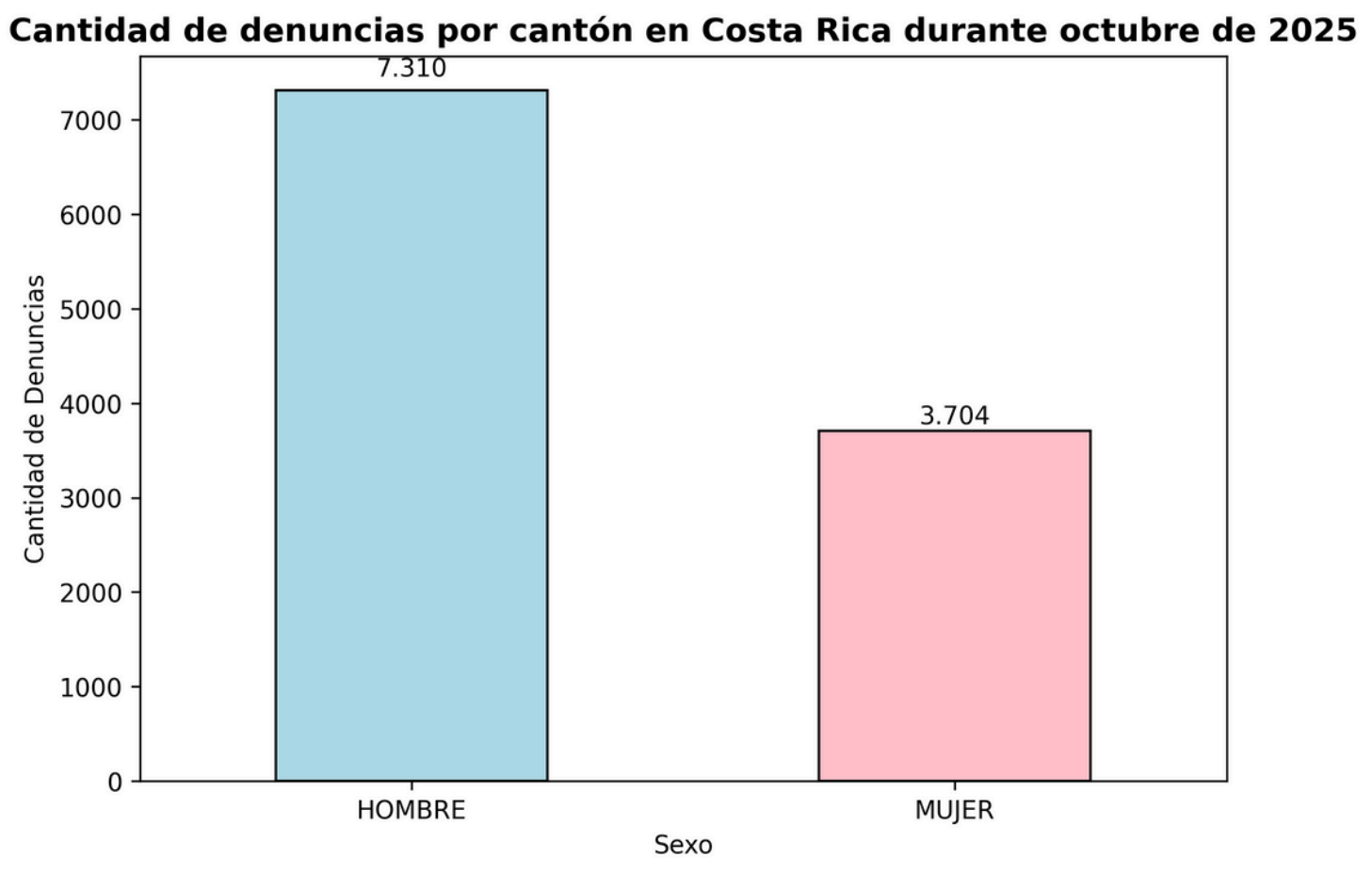
2) GRAFICO DE BARRAS VERTICALES

Definición:

El gráfico de barras horizontales es una herramienta visual que permite comparar valores entre distintas categorías mediante rectángulos dispuestos de forma horizontal.

Muestra:

Se tomaron las denuncias clasificadas según el sexo de la persona denunciante, agrupando los datos en las categorías masculino y femenino.



Uso:

se utiliza cuando se desea comparar categorías nominales o no numéricas, especialmente cuando los nombres de las categorías son largos o numerosos

Codigo:

```
if serie.empty():
    print("[AVISO] No hay datos para graficar por sexo. Revisa el archivo CSV.")
else:
    # Crear gráfico de barras verticales
    colores = ["lightblue" if s == "HOMBRE" else "pink" for s in serie.index]

    ax = serie.plot(kind="bar", figsize=(7, 5), color=colores, edgecolor="black")
    ax.set_title("Cantidad de denuncias por cantón en Costa Rica durante octubre de 2025", fontsize=13, weight="bold")
    ax.set_xlabel("Sexo")
    ax.set_ylabel("Cantidad de Denuncias")
```

INTRODUCCIÓN

Este trabajo analiza los delitos registrados en Costa Rica durante el año 2025. Se utiliza Python y R para limpiar, procesar y representar los datos de forma visual mediante gráficos estadísticos.

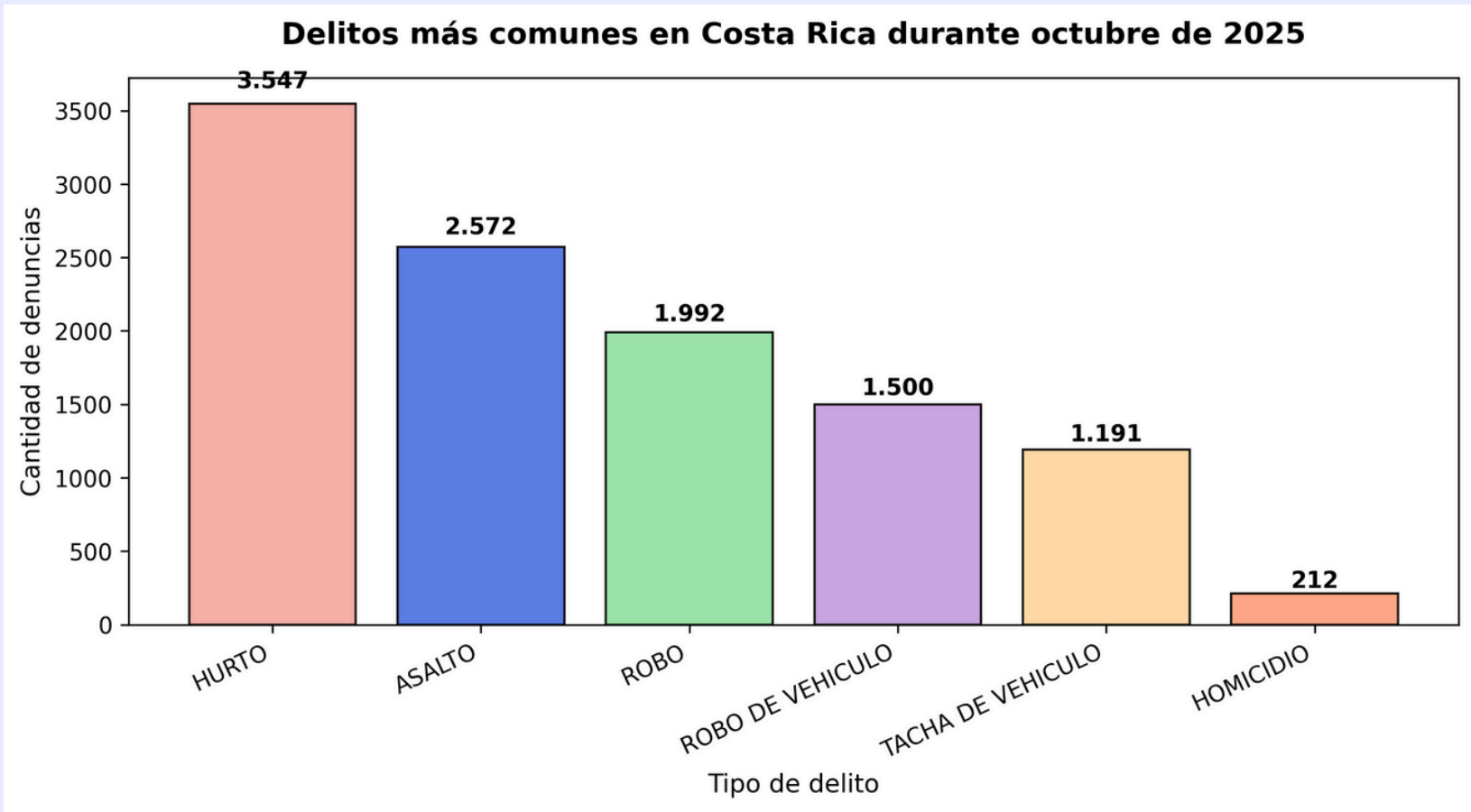
3) GRAFICO DE BARRAS SIMPLES

Definición:

Representa una sola variable categórica y permite comparar la frecuencia o cantidad de cada categoría de forma individual, cada barra muestra un único valor, facilitando la comparación directa entre diferentes categorías.

Muestra:

Se emplearon las frecuencias de denuncias agrupadas por tipo de delito, sin incluir subdivisiones por sexo u otras variables.



Uso:

Se usa cuando se desea comparar distintas categorías de una misma variable (por ejemplo, tipos de delitos) sin subdividirlas en grupos o combinarlas con otras dimensiones.

Codigo:

```
if serie.empty():
    print("[AVISO] No hay datos para graficar tipos de delito. Revisa el archivo CSV.")
else:
    # Colores personalizados (evita colores parecidos)
    colores = [
        "#F4A7A0", # HURTO
        "#4A6FE3", # ASALTO (negro)
        "#91E3A1", # ROBO (verde)
        "#E2B7EB", # ROBO DE VEHÍCULO (lila, más claro y distinto)
        "#FFD59E", # TACHA DE VEHÍCULO
        "#FF66A1" # HOMICIDIO
    ]

    plt.figure(figsize=(8, 6))
    plt.pie(
        serie.values,
        labels=serie.index,
        autopct=lambda p: f"{p:.1f}%\n({int(p * sum(serie.values) / 100):,})",
        startangle=90,
        colors=colores[:len(serie)],
        textprops={'fontsize': 10}
    )
```

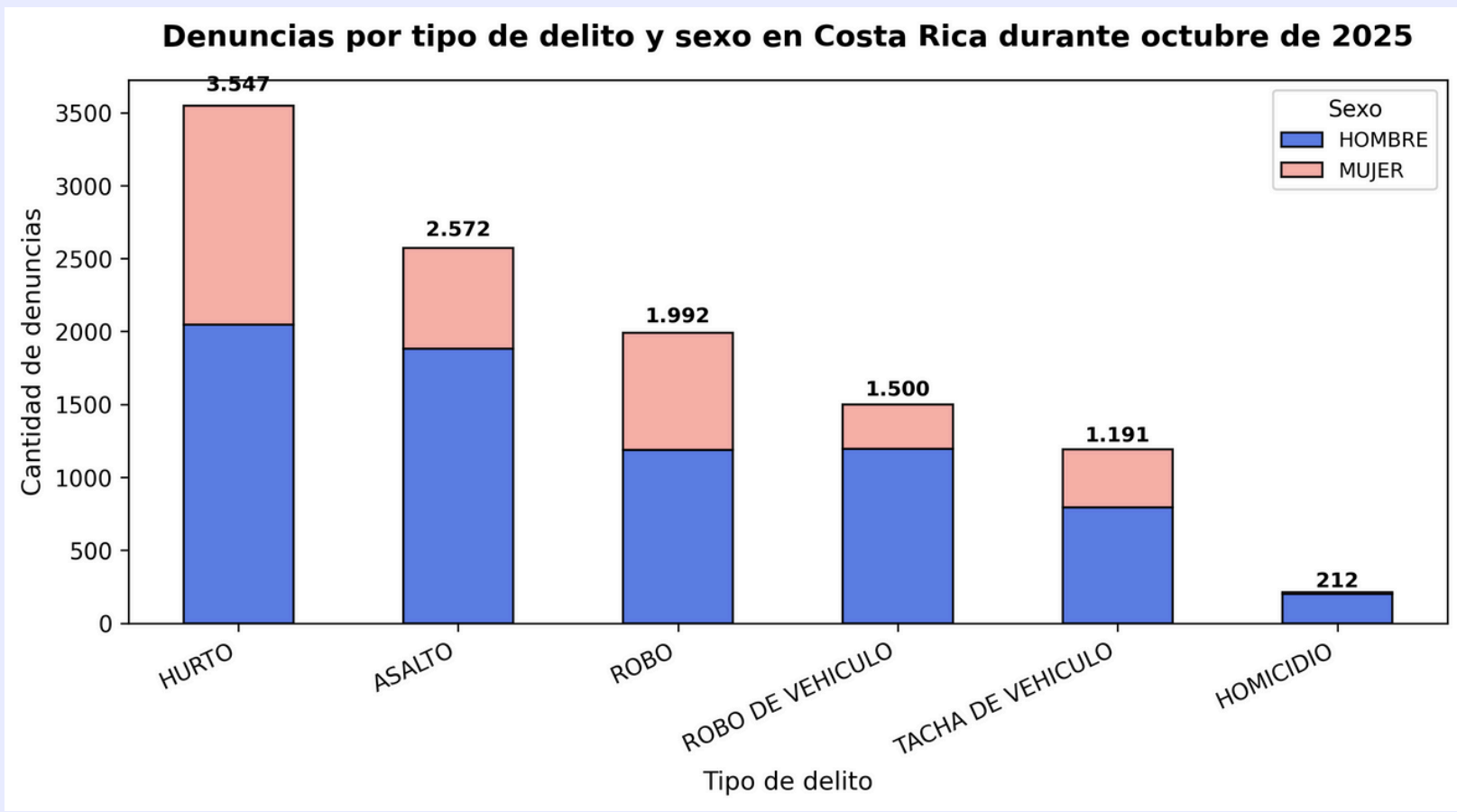
4) GRAFICO DE BARRAS COMPUESTAS

Definición:

El gráfico de barras compuestas (apiladas) es un tipo de gráfico que muestra varios conjuntos de datos dentro de la misma barra, donde cada segmento representa una parte del total.

Muestra:

Se usaron los datos de denuncias por tipo de delito, divididos según el sexo de las personas involucradas, para mostrar la proporción interna dentro de cada categoría.



Uso:

Se utiliza cuando se desea comparar totales y al mismo tiempo mostrar la distribución interna de cada categoría y para identificar qué grupos aportan más al total y analizar diferencias dentro de las categorías.

Codigo:

```
# Crear figura
fig, ax = plt.subplots(figsize=(9, 5))
# Graficar barras apiladas
tabla.plot(
    kind="bar",
    stacked=True,
    color=[colores.get(col, "gray") for col in tabla.columns],
    edgecolor="black",
    alpha=0.9,
    ax=ax
)
```

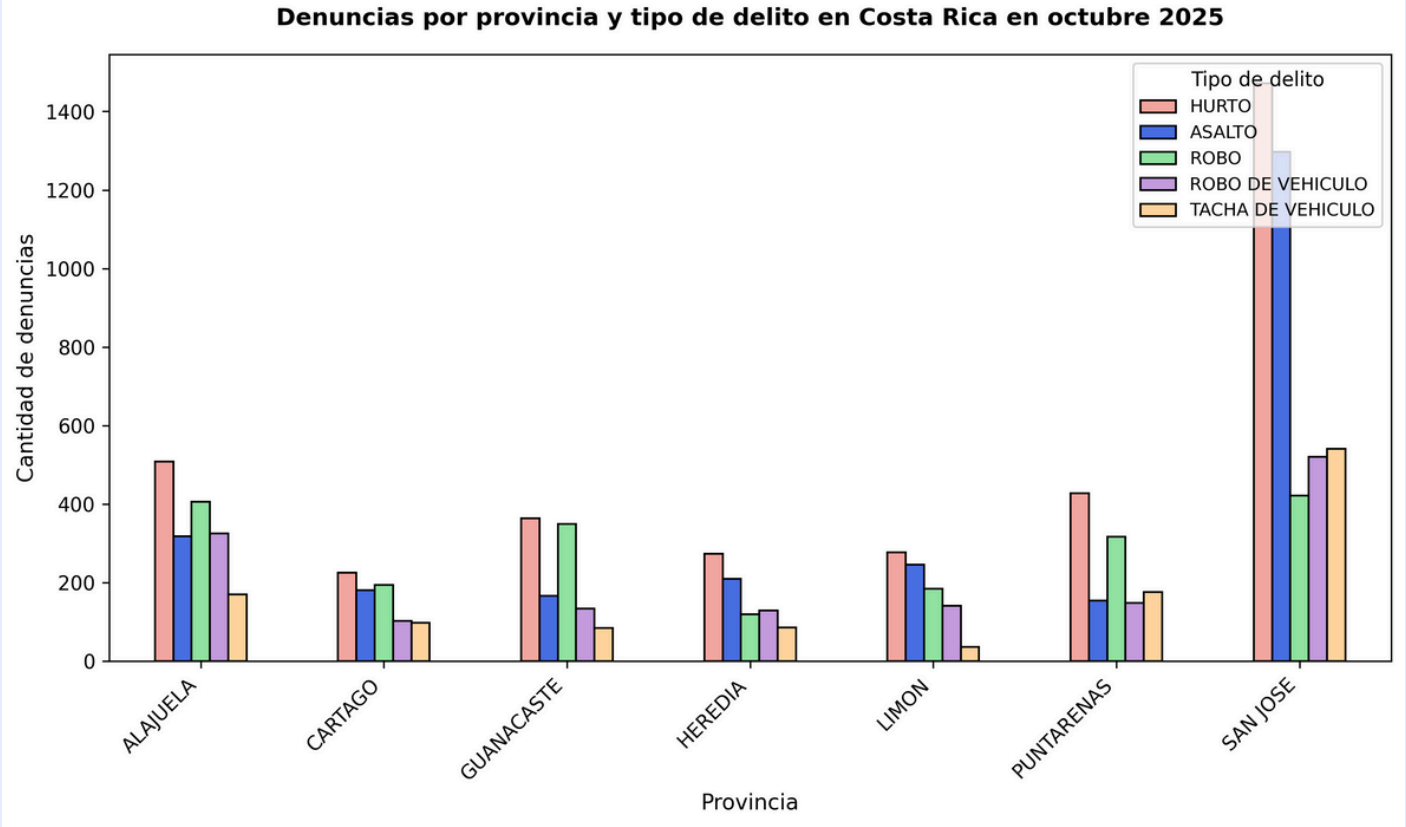

5) GRAFICO DE BARRAS COMPARATIVAS

Definición:

Representación visual que muestra varias variables dentro de una misma categoría, dividiendo cada barra en segmentos que representan diferentes componentes del total.

Muestra:

Se utilizaron los registros de denuncias clasificados por provincia y tipo de delito, agrupando los datos de manera que cada provincia muestra varias barras comparativas correspondientes a los diferentes tipos de delitos registrados durante octubre de 2025



Uso:

se utiliza cuando se desea comparar la distribución de distintas variables dentro de una misma categoría y Es útil para analizar cómo se componen los totales y detectar patrones o diferencias entre grupos.

Codigo:

```
if tabla.empty():
    print("[AVISO] No hay datos para graficar denuncias por provincia y delito. Revisa el CSV.")
else:
    # Crear gráfico de barras agrupadas
    ax = tabla.plot(
        kind="bar",
        figsize=(10, 6),
        color=colores[len(top_delitos)],
        edgecolor="black")
```

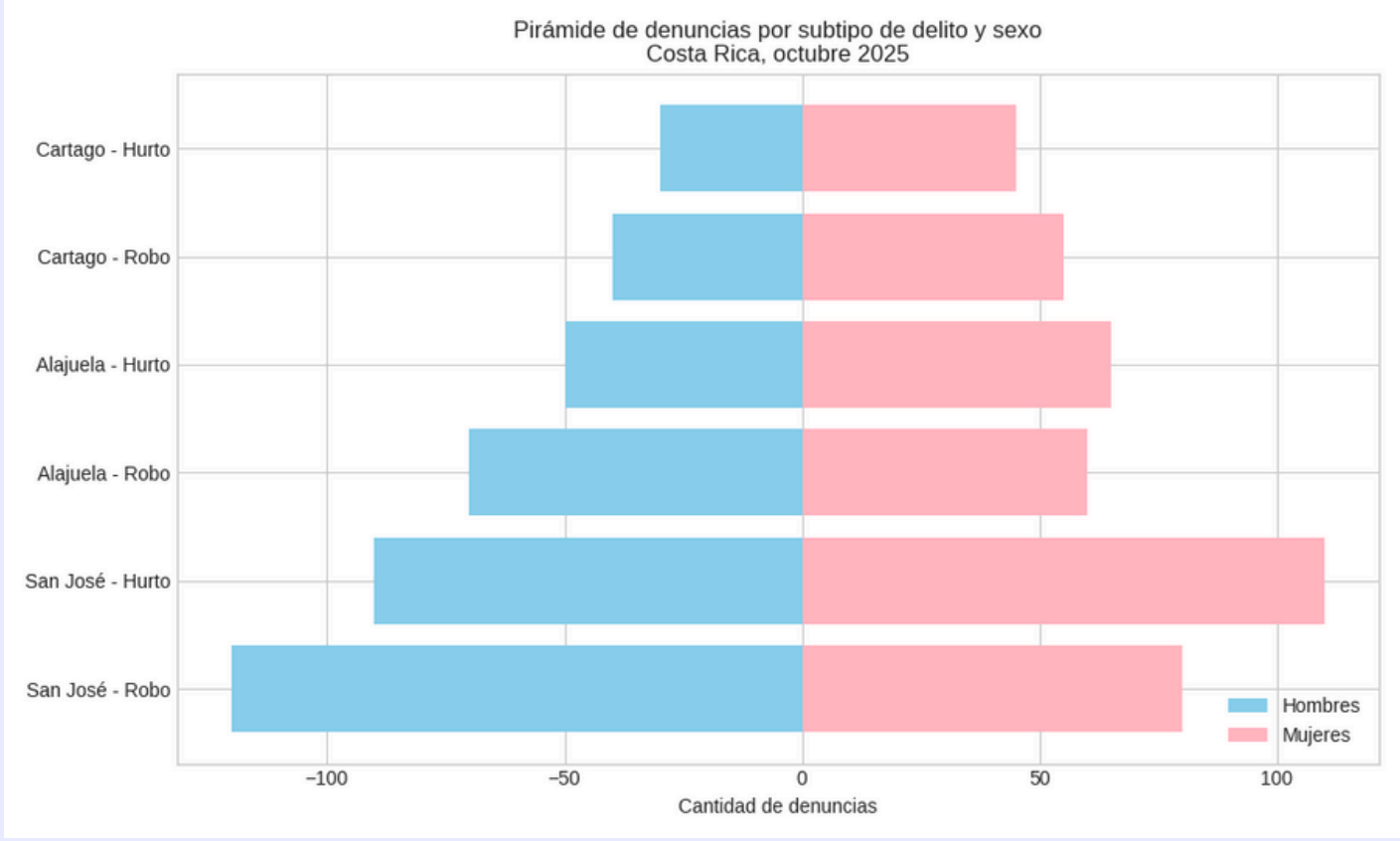
6) GRAFICO DE PIRAMIDE

Definición:

Un gráfico de pirámide es una visualización triangular dividida en secciones horizontales que se utiliza para mostrar una estructura jerárquica o una relación basada en fundamentos

Muestra:

Compuesta por un total de 1,090 denuncias, distribuidas entre hombres y mujeres en tres provincias (San José, Alajuela y Cartago) y dos subtipos de delito (Robo y Hurto).



Uso:

Los gráficos de pirámide se utilizan para mostrar estructuras jerárquicas, distribuciones de datos y comparaciones.

```
Codigo: # Datos de ejemplo
datos <- data.frame(
  Grupo = c("San José - Robo", "San José - Hurto", "Alajuela - Robo",
    "Alajuela - Hurto", "Cartago - Robo", "Cartago - Hurto"),
  Hombre = c(-120, -90, -70, -50, -40, -30),
  Mujer = c(80, 110, 60, 65, 55, 45)
)# Convertir a formato largo
datos_largo <- reshape2::melt(datos, id.vars = "Grupo", variable.name = "Sexo", value.name = "Cantidad")
# Gráfico de pirámide
ggplot(datos_largo, aes(x = Grupo, y = Cantidad, fill = Sexo)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  scale_fill_manual(values = c("Hombre" = "skyblue", "Mujer" = "pink")) +
  labs(title = "Frecuencia de subtipos de delito por provincia en Costa Rica en octubre de 2025",
    x = "Provincia - Subtipo de delito", y = "Cantidad de denuncias") +
  theme_minimal()
```

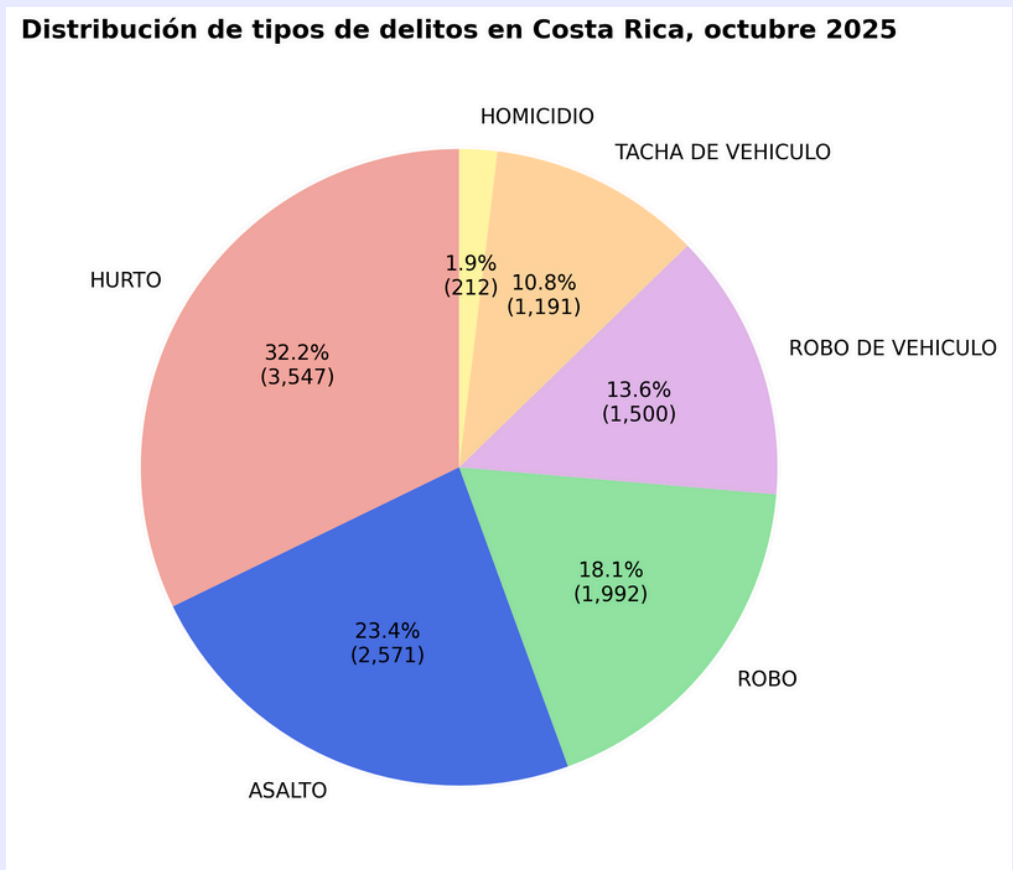
7) GRAFICO DE PASTEL

Definición:

representación circular que muestra cómo se divide un conjunto de datos en proporciones. Cada sector del círculo representa una categoría, y su tamaño es proporcional a la cantidad o porcentaje que ocupa dentro del total.

Muestra:

Se tomaron los valores totales de denuncias agrupados por tipo de delito, representando el porcentaje que aporta cada delito al total general durante octubre de 2025.



Uso:

se utiliza cuando se desea representar cómo se distribuye un total entre diferentes categorías. Es ideal para mostrar porcentajes o proporciones,

```
Codigo: if serie.empty():
    print("[AVISO] No hay datos para graficar tipos de delito. Revisa el archivo CSV.")
else:
    # Colores personalizados (evita colores parecidos)
    colores = {
        "#F4A7A0", # HURTO
        "#4A6FE3", # ASALTO (negro)
        "#91E3A1", # ROBO (verde)
        "#E2B7EB", # ROBO DE VEHÍCULO (lila, más claro y distinto)
        "#FFD59E", # TACHA DE VEHÍCULO
        "#FFF6A1" # HOMICIDIO
    }

plt.figure(figsize=(8, 6))
plt.pie(
    serie.values,
    labels=serie.index,
    autopct=lambda p: f'{p:.1f}%\n(int(p *
sum(serie.values) / 100):,))",
    startangle=90,
    colors=colores[len(serie)],
    textprops={'fontsize': 10}
)
```

CONCLUSION

El análisis estadístico permitió visualizar de forma clara las denuncias registradas en Costa Rica durante octubre de 2025.

Mediante diferentes tipos de gráficos, se identificaron patrones, comparaciones y distribuciones que facilitan la comprensión de los datos y apoyan la toma de decisiones basadas en evidencia.

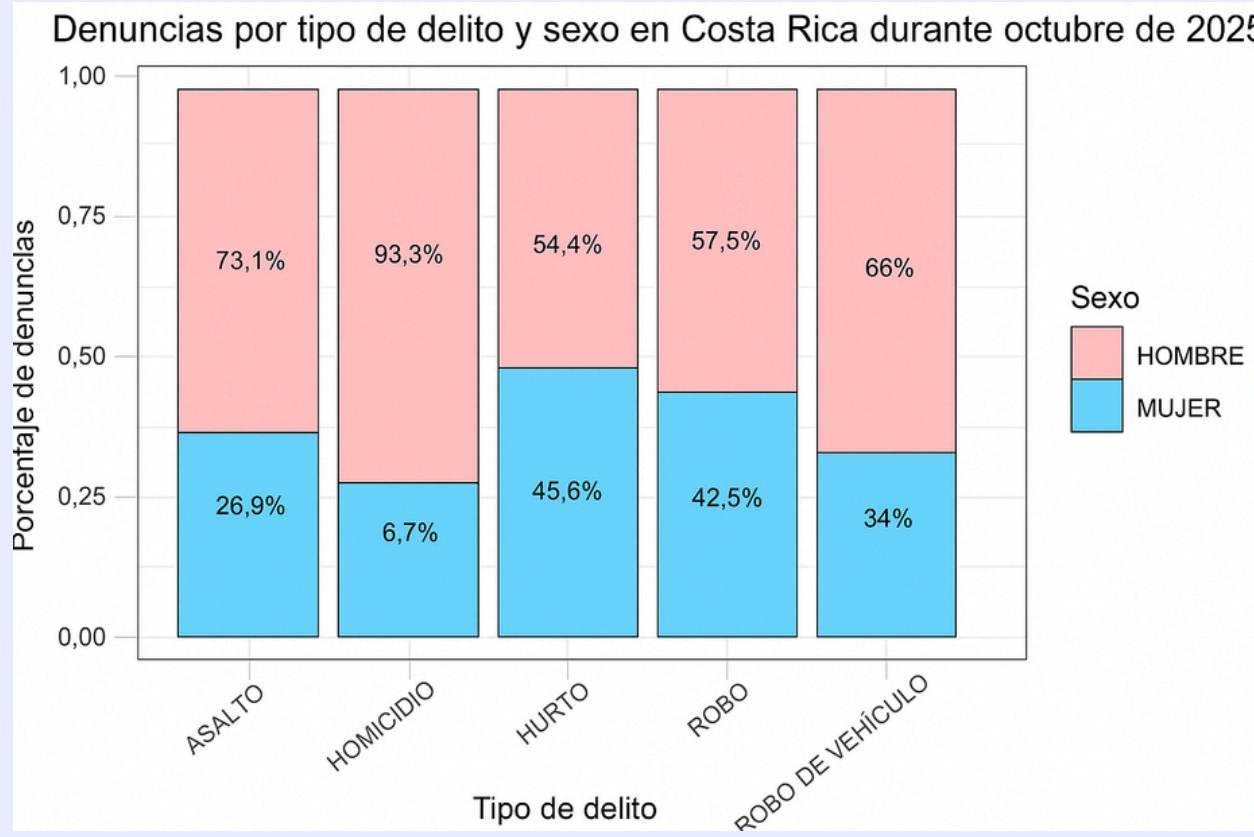
8) GRAFICO DE BARRAS 100%

Definición:

Un gráfico de barras 100% muestra la proporción relativa de categorías dentro de un total, normalizando cada barra al 100% para facilitar la comparación porcentual entre grupos.

Muestra:

Conjunto total de denuncias registradas en Costa Rica durante octubre de 2025, desglosadas por tipo de delito y sexo.



Uso:

El gráfico de barras 100% se usa para comparar la contribución porcentual relativa de cada serie de datos dentro de un grupo tota

```
Codigo: ggplot(datos, aes(x = Delito, y = Porcentaje, fill = Sexo)) +
  geom_bar(stat = "identity", position = "fill", color = "black") +
  geom_text(aes(label = paste0(round(Porcentaje, 1), "%")),
    position = position_fill(vjust = 0.5), size = 3, color = "black") +
  scale_y_continuous(labels = percent_format(scale = 1)) +
  labs(
    title = "Denuncias por tipo de delito y sexo en Costa Rica durante octubre de 2025",
    x = "Tipo de delito", y = "Porcentaje de denuncias", fill = "Sexo"
  ) +
  theme_minimal(base_size = 12) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(hjust = 0.5, face = "bold"))
```

9) GRAFICO LINEAL

Definición:

Un gráfico lineal es una representación gráfica de datos que conecta puntos de datos individuales con segmentos de línea recta. Se utiliza principalmente para mostrar la tendencia de una variable a lo largo del tiempo o la relación entre dos variables.

Muestra:

La muestra de denuncias por tipo de víctima en Costa Rica durante octubre de 2025 corresponde al total de denuncias registradas en ese mes, agrupadas según el tipo de víctima.



Uso:

Se usa el gráfico lineal para mostrar tendencias y cambios continuos en los datos a lo largo del tiempo o de un rango continuo, como el tiempo, la distancia, el peso o la temperatura.

```
Codigo: # Leer y filtrar datos
datos <- read.csv("Estadisticas_limpio.csv", sep = ";")
datos_filtrados <- subset(datos, Anio == 2025 & Mes == 10)
# Agrupar por tipo de víctima
datos_victima <- aggregate(Cantidad ~ TipoVictima, data = datos_filtrados, FUN = sum)
# Crear gráfico lineal
ggplot(datos_victima, aes(x = TipoVictima, y = Cantidad, group = 1)) +
  geom_line(color = "steelblue", size = 1.2) +
  geom_point(color = "skyblue", size = 3) +
  labs(title = "Denuncias por tipo de víctima en Costa Rica durante octubre de 2025",
    x = "Tipo de víctima",
    y = "Cantidad de denuncias") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

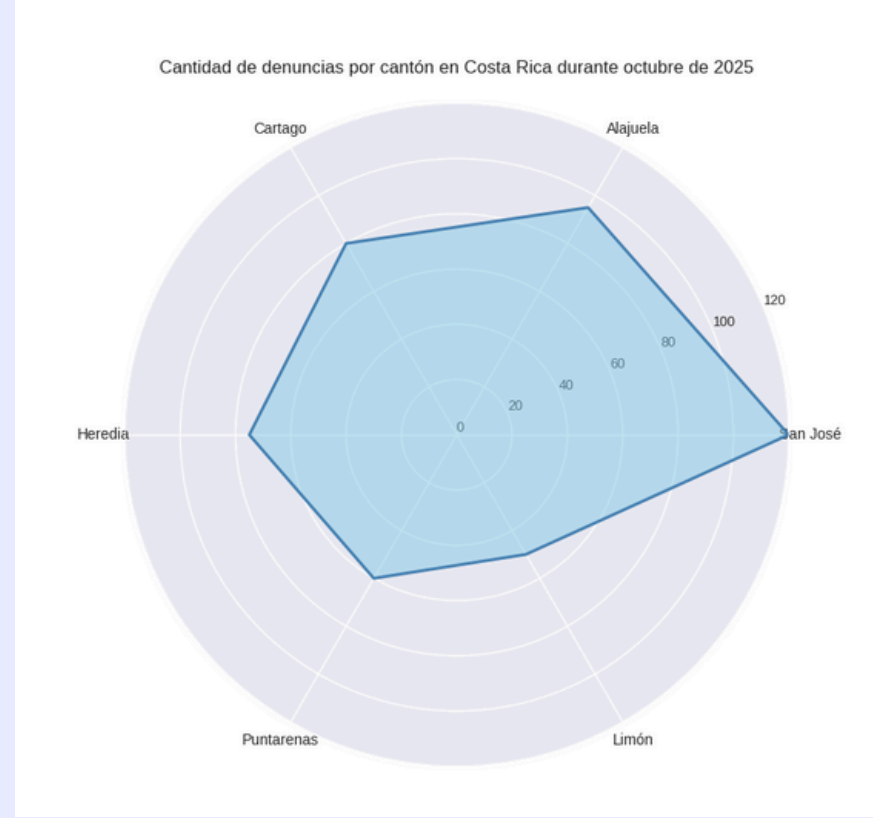
10) GRAFICO DE ARAÑA / RADAR

Definición:

Un gráfico de radar es una forma de visualizar datos multivariados en un gráfico bidimensional, comparando múltiples variables cuantitativas a través de ejes radiales que parten de un punto central.

Muestra:

corresponde al total de denuncias registradas en los seis cantones con mayor cantidad de casos en Costa Rica durante octubre de 2025.



Uso:

Son ideales para comparar rendimientos (deportistas, empleados), características de productos,, y para analizar la competencia en un mercado.

```
Codigo: # Datos de ejemplo
cantones <- c("San José", "Alajuela", "Cartago", "Heredia", "Puntarenas", "Limón")
denuncias <- c(120, 95, 80, 75, 60, 50)
# Preparar datos para radar
max_val <- max(denuncias)
datos_radar <- as.data.frame(rbind(rep(max_val, 6), rep(0, 6), denuncias))
colnames(datos_radar) <- cantones
# Graficar
radarchart(datos_radar,
  axistype = 1,
  pcol = "steelblue",
  pfcol = alpha("skyblue", 0.5),
  plwd = 2,
  cglcol = "gray",
  caxislabs = seq(0, max_val, 20),
  vlce = 0.9,
  title = "Denuncias por cantón en octubre 2025")
```

GIT HUB PAGE

<https://github.com/andreyac1397>