

TRABAJO PA PRESENTACION - ANÁLISIS DESCRIPTIVO Y VISUALIZACIONES DEL CASO DENGUE EN EL PERÚ, PERIODO 2000 - 2023

2025-01-18

INTEGRANTES:

- ALBINO HUERTAS EDER ALBERTO / 73036543@continental.edu.pe
(mailto:73036543@continental.edu.pe)
- HUILLCA ALARCON EDSON JOEL / 46406505@continental.edu.pe
(mailto:46406505@continental.edu.pe)
- NIÑO ALCANTARA ULISES EMILIO / 16737529@continental.edu.pe
(mailto:16737529@continental.edu.pe)
- TOLEDO CHÁVEZ SALLY DAMARIS / 74444594@continental.edu.pe
(mailto:74444594@continental.edu.pe)

Presentación

Este dataset contiene información sobre la notificación de casos de dengue al sistema de vigilancia en salud pública del Perú, gestionado y administrado por el Centro Nacional de Epidemiología, Prevención y Control de Enfermedades (CDC PERU). Los dengue_data provienen de la Red Nacional de Epidemiología (RENACE), que está conformada por 10 232 establecimientos de salud del Ministerio de Salud, EsSalud y otros del sector en los diferentes niveles de las Direcciones Regionales de Salud que tiene el Perú.

Descripción de los campos

CAMPO	TIPO	DENOMINACIÓN	VALORES
departamento	Caracter	Región geográfica	
provincia	Caracter	Provincia	
distrito	Caracter	Lugar probable de infección	
enfermedad	Caracter	Diagnóstico vigilado	
ano	Integer	Año	
semana	Integer	Semana de inicio de síntomas	
diagnostic	Caracter	CIE 10	
tipo_dx	Caracter	Tipo de diagnóstico	C= Confirmado, P = Probable, S = Sospechoso
diresa	Caracter	Dirección de salud que notifica	

CAMPO	TIPO	DENOMINACIÓN	VALORES
ubigeo	Caracter	Código del lugar probable de infección	
edad	Caracter	Edad del paciente	
tipo_edad	Caracter	Tipo de edad del paciente	A = Año, M = Mes, D = Días
sexo	Caracter	Sexo	M = Masculino, F = Femenino

```
dengue_data <- read_csv(paste0(getwd(), "/lib/datos_abiertos_vigilancia_dengue_2000_2023.csv"))
```

```
## Rows: 757890 Columns: 13
## — Column specification —————
## Delimiter: ","
## chr (10): departamento, provincia, distrito, enfermedad, diagnostic, tipo_dx...
## dbl (3): ano, semana, edad
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Resumen general de los datos

```
head(dengue_data)
```

```
## # A tibble: 6 × 13
##   departamento provincia    distrito enfermedad   ano semana diagnostic tipo_dx
##   <chr>         <chr>      <chr>      <chr>      <dbl> <dbl> <chr>      <chr>
## 1 HUANUCO      LEONCIO PRADO LUYANDO  DENGUE SI... 2000    47 A97.0      P
## 2 HUANUCO      LEONCIO PRADO LUYANDO  DENGUE SI... 2000    40 A97.0      P
## 3 HUANUCO      LEONCIO PRADO JOSE CR... DENGUE SI... 2000    48 A97.0      C
## 4 HUANUCO      LEONCIO PRADO JOSE CR... DENGUE SI... 2000    37 A97.0      P
## 5 HUANUCO      LEONCIO PRADO MARIANO... DENGUE SI... 2000    42 A97.0      C
## 6 HUANUCO      LEONCIO PRADO LUYANDO  DENGUE SI... 2000    40 A97.0      P
## # i 5 more variables: diresa <chr>, ubigeo <chr>, edad <dbl>, tipo_edad <chr>,
## #   sexo <chr>
```

```
summary(dengue_data)
```

```
## departamento      provincia      distrito      enfermedad
## Length:757890      Length:757890      Length:757890      Length:757890
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##
##
##
##      ano      semana      diagnostic      tipo_dx
## Min.   :2000   Min.    : 1.00   Length:757890      Length:757890
## 1st Qu.:2014   1st Qu.:13.00   Class :character    Class :character
## Median :2020   Median :20.00   Mode :character     Mode :character
## Mean   :2018   Mean    :22.47
## 3rd Qu.:2023   3rd Qu.:29.00
## Max.   :2023   Max.    :53.00
##      diresa      ubigeo      edad      tipo_edad
## Length:757890      Length:757890      Min.    : 0      Length:757890
## Class :character    Class :character    1st Qu.: 15      Class :character
## Mode :character     Mode :character    Median : 27      Mode :character
##                      Mean    : 189
##                      3rd Qu.: 42
##                      Max.    :71963641
##      sexo
## Length:757890
## Class :character
## Mode :character
##
##
##
```

ANÁLISIS DEL DATASET

```
# 2. Limpiar y preparar los datos
# Convertir campos relevantes a tipo factor
colnames(dengue_data)
```

```
## [1] "departamento" "provincia"      "distrito"      "enfermedad"    "ano"
## [6] "semana"        "diagnostic"     "tipo_dx"       "diresa"        "ubigeo"
## [11] "edad"          "tipo_edad"     "sexo"
```

```
cols_to_factor <- c("departamento", "provincia", "distrito", "enfermedad",
                    "diagnostic", "tipo_dx", "diresa", "ubigeo", "tipo_edad", "sexo")
dengue_data[cols_to_factor] <- lapply(dengue_data[cols_to_factor], as.factor)
View(dengue_data)

# Convertir campos de Año y Semana a fechas
# Supongamos que la semana comienza el Lunes
dengue_data <- dengue_data %>%
  mutate(
    Fecha_inicio = suppressWarnings(as.Date(paste(ano, semana, 1, sep = "-"), "%Y-%U-%u"))
  )

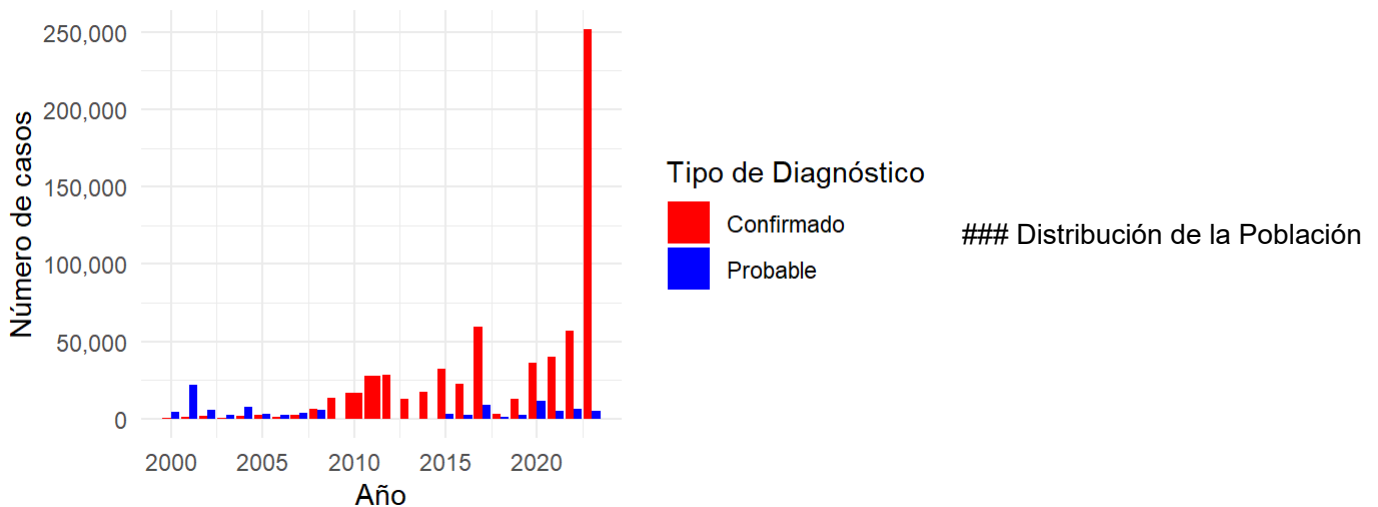
# Convertir Edad a numérica para análisis
# (Manejo de posibles casos como NA en datos no válidos)
dengue_data$edad <- as.numeric(dengue_data$edad)
```

Análisis descriptivo y visualizaciones

```
# Total de casos por año y tipo de diagnóstico
total_por_año <- dengue_data %>%
  group_by(ano, tipo_dx) %>%
  summarise(Casos = n(), .groups = 'drop')

ggplot(total_por_año, aes(x = ano, y = Casos, fill = tipo_dx)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "Total de casos por año y tipo de diagnóstico",
       x = "Año", y = "Número de casos", fill = "Tipo de Diagnóstico") +
  scale_fill_manual(
    values = c("C" = "red", "P" = "blue"),
    labels = c("C" = "Confirmado", "P" = "Probable")
  ) +
  scale_y_continuous(labels = comma)
```

Total de casos por año y tipo de diagnóstico

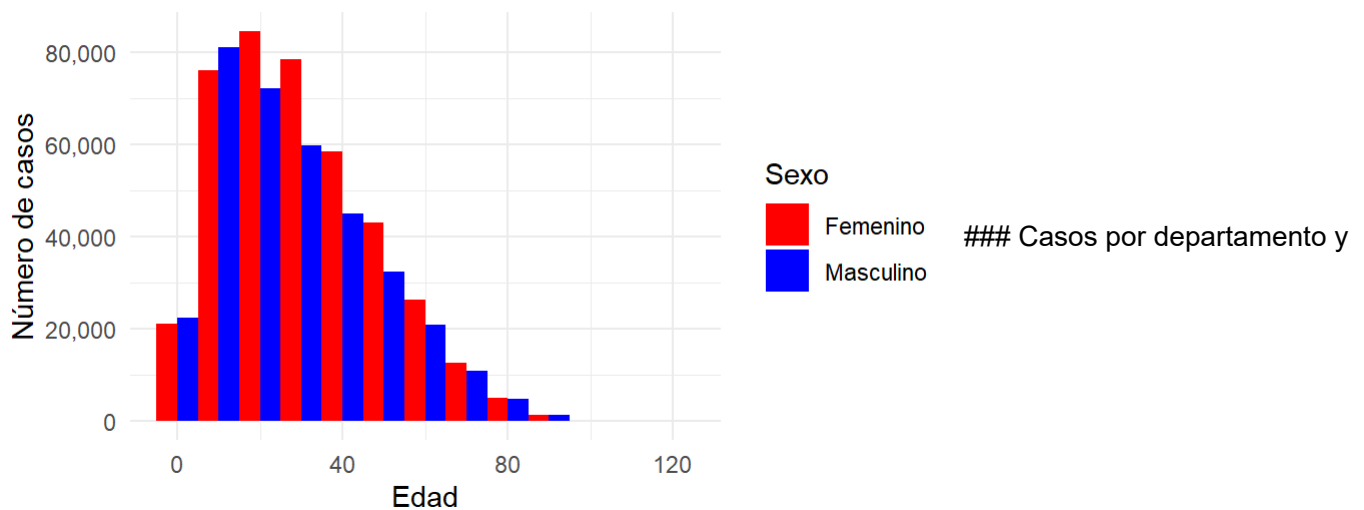


por edad y sexo

```
dengue_data_edad_sexo <- dengue_data %>% filter(edad >= 0 & edad <= 120)

ggplot(dengue_data_edad_sexo, aes(x = edad, fill = sexo)) +
  geom_histogram(binwidth = 10, position = "dodge") +
  theme_minimal() +
  labs(title = "Distribución de Casos por Edad y Sexo",
       x = "Edad", y = "Número de casos", fill = "Sexo")+
  scale_fill_manual(
    values = c("M" = "blue", "F" = "red"),
    labels = c("M" = "Masculino", "F" = "Femenino")
  ) +
  scale_y_continuous(labels = comma)
```

Distribución de Casos por Edad y Sexo

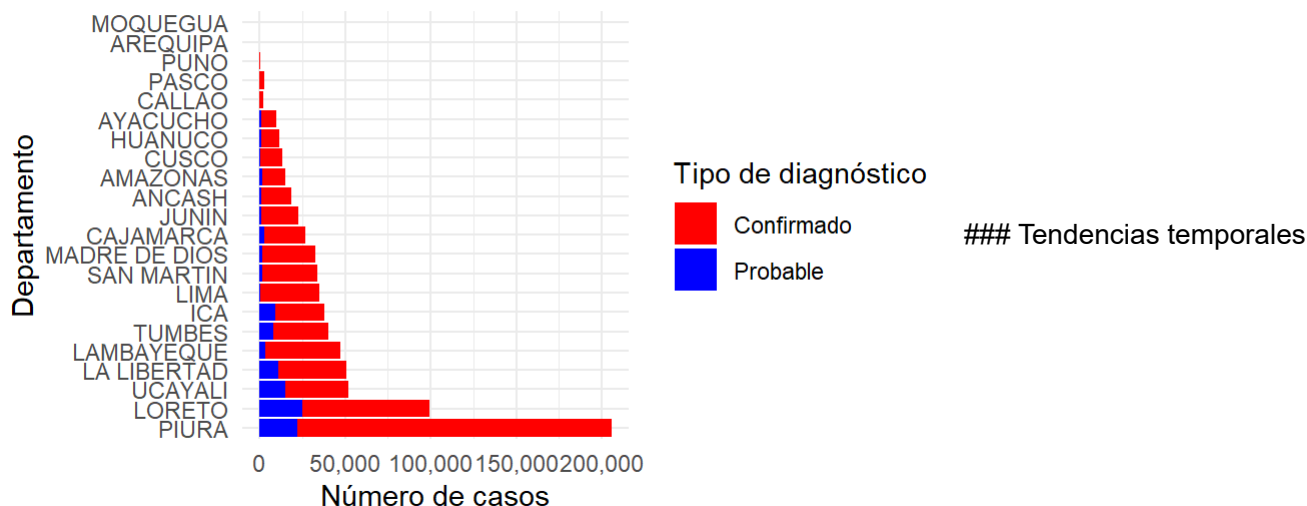


tipo de diagnóstico

```
casos_departamento <- dengue_data %>%
  group_by(departamento, tipo_dx) %>%
  summarise(Casos = n(), .groups = 'drop')

ggplot(casos_departamento, aes(x = reorder(departamento, -Casos), y = Casos, fill = tipo_dx)) +
  geom_bar(stat = "identity", position = "stack") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Casos por Departamento y Tipo de Diagnóstico",
       x = "Departamento", y = "Número de casos", fill = "Tipo de diagnóstico") +
  scale_fill_manual(
    values = c("C" = "red", "P" = "blue"),
    labels = c("C" = "Confirmado", "P" = "Probable")
  ) +
  scale_y_continuous(labels = comma)
```

Casos por Departamento y Tipo de Diagnóstico



Casos por semana

tendencia_semanal <- dengue_data %>%

group_by(Fecha_inicio) %>%

summarise(Casos = n(), .groups = 'drop')

ggplot(tendencia_semanal, aes(x = Fecha_inicio, y = Casos)) +

geom_line(color = "blue") +

theme_minimal() +

labs(title = "Tendencia semanal de casos",

x = "Fecha de Inicio de Síntomas", y = "Número de Casos") +

scale_x_date(date_labels = "%b %Y", date_breaks = "2 year") +

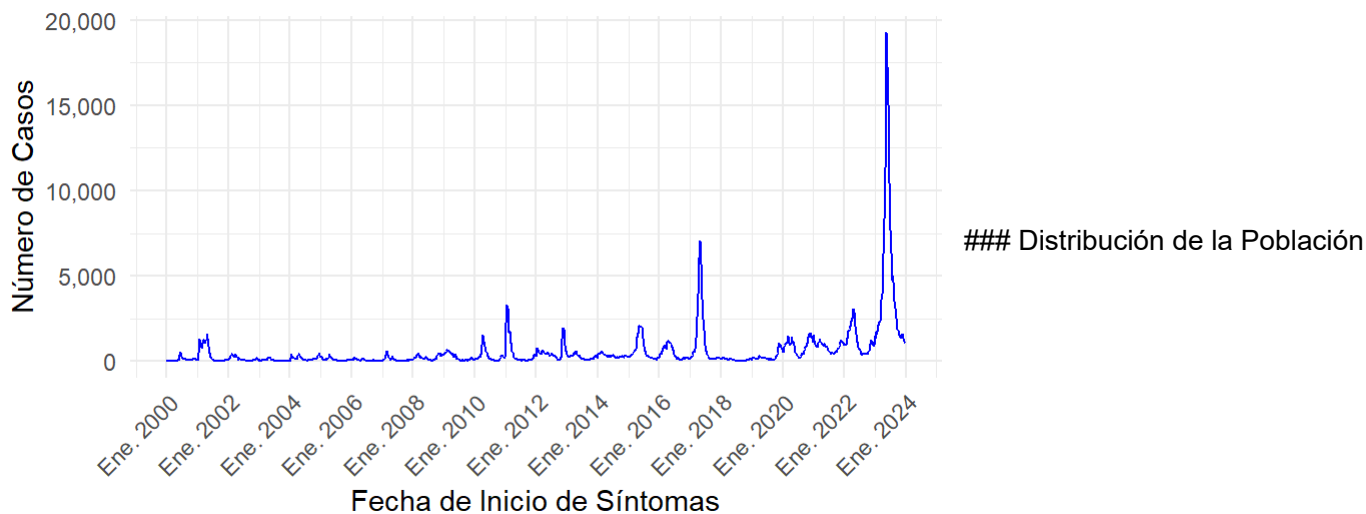
scale_y_continuous(labels = comma)+

theme(axis.text.x = element_text(angle = 45, hjust = 1))

Warning: Removed 1 row containing missing values or values outside the scale range

(`geom_line()`).

Tendencia semanal de casos



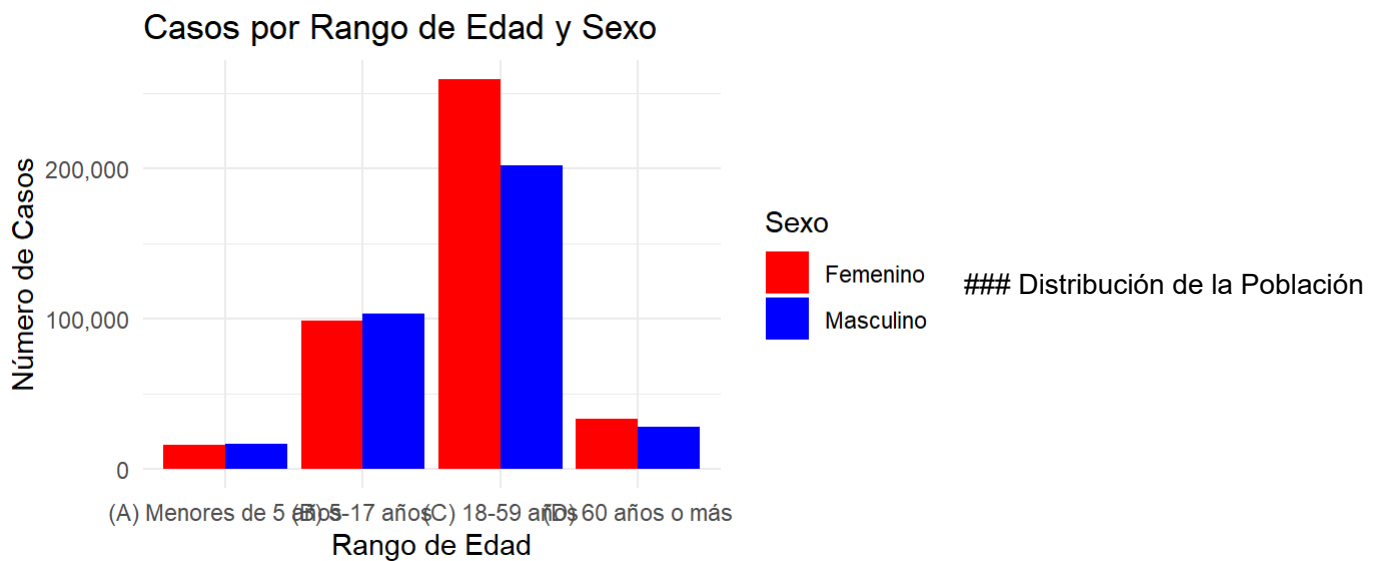
Distribución de la Población

por sexo y rango de edad

```
# Crear rangos de edad
dengue_data <- dengue_data %>%
  mutate(Rango_edad = case_when(
    edad < 5 ~ "(A) Menores de 5 años",
    edad >= 5 & edad < 18 ~ "(B) 5-17 años",
    edad >= 18 & edad < 60 ~ "(C) 18-59 años",
    edad >= 60 ~ "(D) 60 años o más",
    TRUE ~ "Desconocido"
  ))

# Casos por rango de edad y sexo
casos_por_rango <- dengue_data %>%
  group_by(Rango_edad, sexo) %>%
  summarise(Casos = n(), .groups = 'drop')

ggplot(casos_por_rango, aes(x = Rango_edad, y = Casos, fill = sexo)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "Casos por Rango de Edad y Sexo",
       x = "Rango de Edad", y = "Número de Casos", fill = "Sexo")+
  scale_fill_manual(
    values = c("M" = "blue", "F" = "red"),
    labels = c("M" = "Masculino", "F" = "Femenino")
  ) +
  scale_y_continuous(labels = comma)
```



por departamento y sexo:

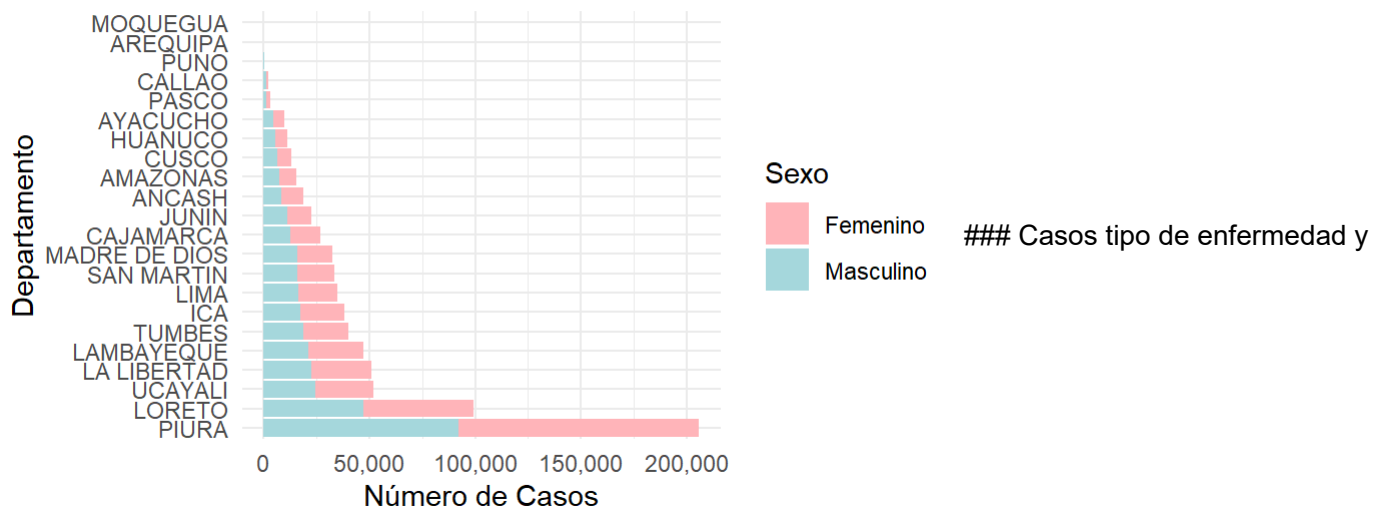
```

casos_por_departamento_sexo <- dengue_data %>%
  group_by(departamento, sexo) %>%
  summarise(Casos = n(), .groups = 'drop')

ggplot(casos_por_departamento_sexo, aes(x = reorder(departamento, -Casos), y = Casos, fill =
sexo)) +
  geom_bar(stat = "identity", position = "stack") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Casos de Dengue por Departamento y Sexo",
       x = "Departamento", y = "Número de Casos", fill = "Sexo") +
  scale_fill_manual(
    values = c("M" = "#A8DADC", "F" = "#FFB6B9"),
    labels = c("M" = "Masculino", "F" = "Femenino")
  ) +
  scale_y_continuous(labels = comma)

```

Casos de Dengue por Departamento y Sexo



edad

```

dengue_data <- dengue_data %>%
  mutate(Rango_edad = case_when(
    edad <= 2 ~ "(A) Lactantes (0-2 años)",
    edad > 2 & edad <= 12 ~ "(B) Infante (3-12 años)",
    edad > 12 & edad <= 20 ~ "(C) Adolescente (13-20 años)",
    edad > 20 & edad <= 26 ~ "(D) Juventud (21-26 años)",
    edad > 26 & edad <= 59 ~ "(E) Adulto (27-59 años)",
    edad >= 60 ~ "(F) Anciano (60+ años)",
    TRUE ~ "Desconocido"
  ))

casos_por_enfermedad_y_edad <- dengue_data %>%
  group_by(Rango_edad, enfermedad) %>%
  summarise(Casos = n(), .groups = 'drop')

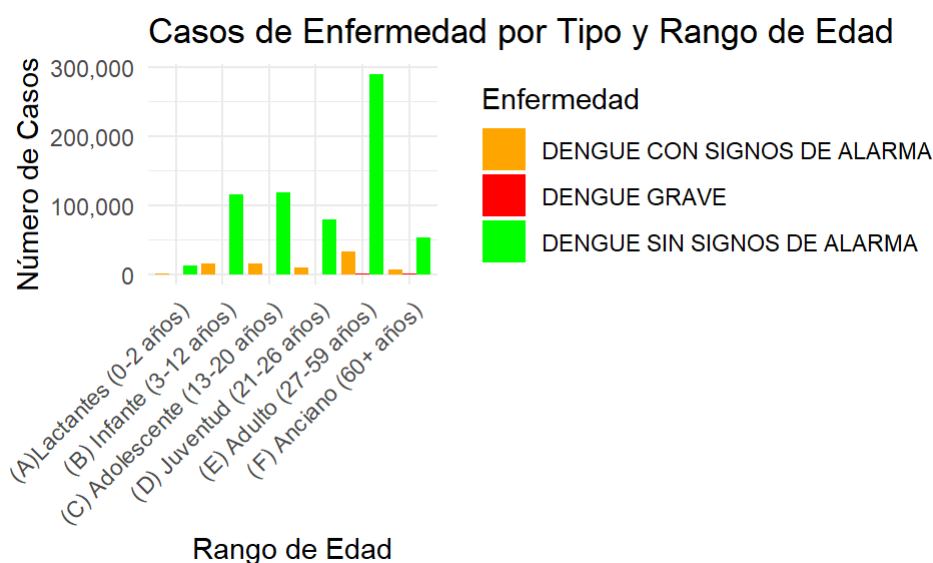
head(casos_por_enfermedad_y_edad)

```



```
## # A tibble: 6 × 3
##   Rango_edad      enfermedad      Casos
##   <chr>          <fct>          <int>
## 1 (A) Lactantes (0-2 años) DENGUE CON SIGNOS DE ALARMA 1565
## 2 (A) Lactantes (0-2 años) DENGUE GRAVE 51
## 3 (A) Lactantes (0-2 años) DENGUE SIN SIGNOS DE ALARMA 13288
## 4 (B) Infante (3-12 años) DENGUE CON SIGNOS DE ALARMA 16437
## 5 (B) Infante (3-12 años) DENGUE GRAVE 531
## 6 (B) Infante (3-12 años) DENGUE SIN SIGNOS DE ALARMA 115263
```

```
ggplot(casos_por_enfermedad_y_edad, aes(x = Rango_edad, y = Casos, fill = enfermedad)) +
  geom_bar(stat = "identity", position = "dodge") +
  theme_minimal() +
  labs(title = "Casos de Enfermedad por Tipo y Rango de Edad",
       x = "Rango de Edad", y = "Número de Casos", fill = "Enfermedad") +
  scale_fill_manual(values = c(
    "DENGUE CON SIGNOS DE ALARMA" = "orange",
    "DENGUE GRAVE" = "red",
    "DENGUE SIN SIGNOS DE ALARMA" = "green"
  )) +
  scale_y_continuous(labels = scales::comma) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



ANÁLISIS DE CASOS

CONFIRMADOS:

```
# Datos confirmados
dengue_data_confirmado <- dengue_data[dengue_data$tipo_dx == 'C', ]
```

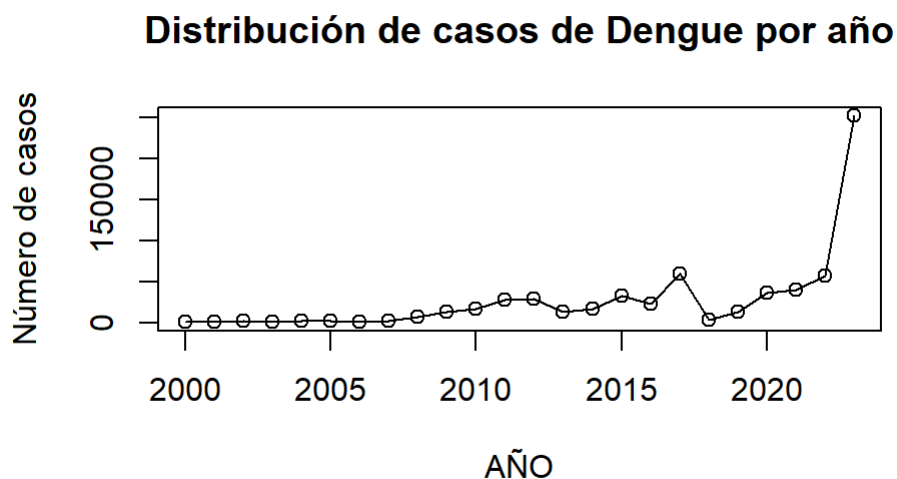
Distribución del dengue por año:

```
tabla_anio <- dengue_data_confirmado %>%
  count(ano)

print(tabla_anio)
```

```
## # A tibble: 24 × 2
##   ano      n
##   <dbl> <int>
## 1  2000    893
## 2  2001   1427
## 3  2002   2077
## 4  2003    508
## 5  2004   1650
## 6  2005   2448
## 7  2006   1219
## 8  2007   2767
## 9  2008   6761
## 10 2009  13395
## # i 14 more rows
```

```
plot(tabla_anio, type="o", main="Distribución de casos de Dengue por año", xlab = "AÑO", ylab = "Número de casos")
```



Observamos que la tendencia en casos de dengue confirmados ha ido incrementado a medida que pasa los años, sobre todo en el ultimo año **2023** que es el ultimo año registrado en el dataset.

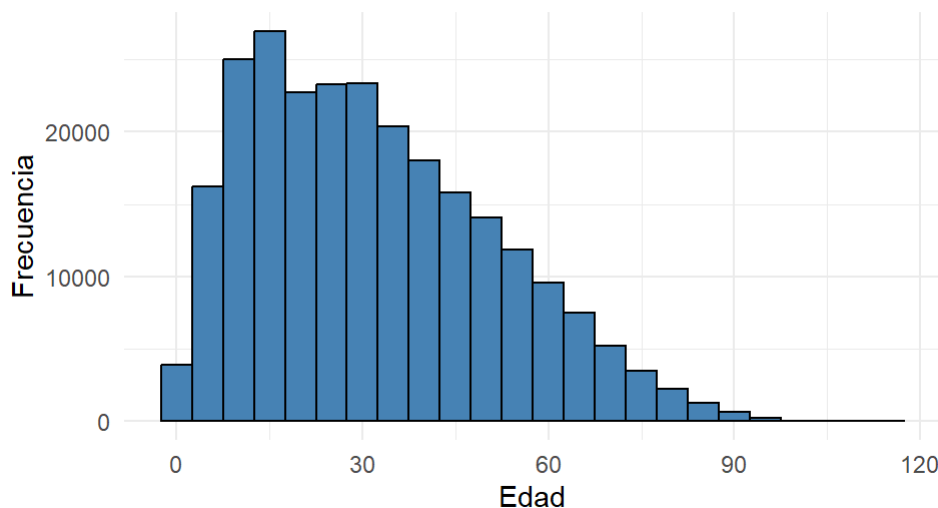
Distribución del dengue por edad año 2023: casos confirmados

```
dengue_data_edad_2023 <- dengue_data_confirmado[dengue_data_confirmado$edad >= 0 & dengue_data_confirmado$edad <= 120 & dengue_data_confirmado$ano=="2023",]
head(dengue_data_edad_2023)
```

```
## # A tibble: 6 × 15
##   departamento provincia    distrito enfermedad  ano semana diagnostic tipo_dx
##   <fct>          <fct>      <fct>    <fct>      <dbl>  <dbl> <fct>      <fct>
## 1 HUANUCO      LEONCIO PRADO RUPA-RU... DENGUE SI...  2023     1 A97.0      C
## 2 CAJAMARCA    JAEN          JAEN      DENGUE SI...  2023     1 A97.0      C
## 3 HUANUCO      LEONCIO PRADO LUYANDO  DENGUE CO...  2023     1 A97.1      C
## 4 JUNIN        CHANCHAMAYO  SAN RAM... DENGUE CO...  2023     1 A97.1      C
## 5 CAJAMARCA    SAN IGNACIO   SAN IGN... DENGUE SI...  2023     1 A97.0      C
## 6 JUNIN        CHANCHAMAYO  SAN RAM... DENGUE SI...  2023     1 A97.0      C
## # i 7 more variables: diresa <fct>, ubigeo <fct>, edad <dbl>, tipo_edad <fct>,
## #   sexo <fct>, Fecha_inicio <date>, Rango_edad <chr>
```

```
# Histograma de edades
ggplot(dengue_data_edad_2023, aes(x = edad)) +
  geom_histogram(binwidth = 5, fill = "steelblue", color = "black") +
  labs(title = "Distribución de la población por edades, año 2023", x = "Edad", y = "Frecuencia") +
  theme_minimal()
```

Distribución de la población por edades, año 2023



Durante el año 2023, la

población más afectada por el dengue está ubicada entre los 5 a 60 años.

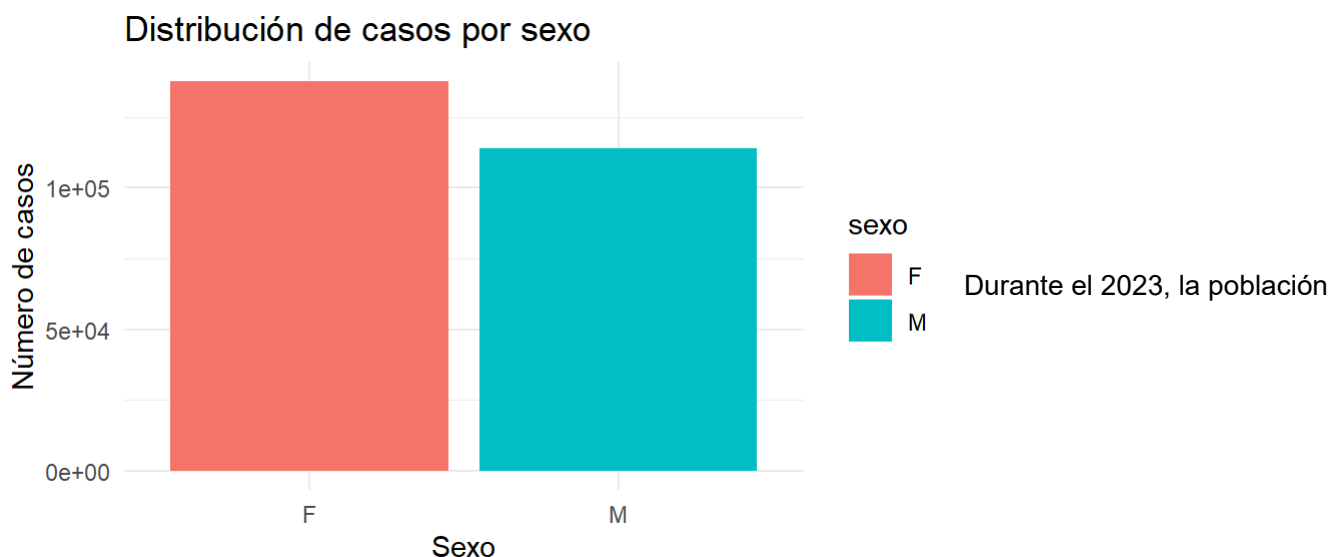
Distribución del dengue por sexo año 2023: casos confirmados

```
# Contar casos por sexo
tabla_sexo <- dengue_data_confirmado[dengue_data_confirmado$ano=="2023",] %>%
  count(sexo)

print(tabla_sexo)
```

```
## # A tibble: 2 × 2
##   sexo      n
##   <fct> <int>
## 1 F      137590
## 2 M      114015
```

```
# Gráfico de barras
ggplot(tabla_sexo, aes(x = sexo, y = n, fill = sexo)) +
  geom_bar(stat = "identity") +
  labs(title = "Distribución de casos por sexo", x = "Sexo", y = "Número de casos") +
  theme_minimal()
```



mas afectada por el dengue fueron mujeres.

Distribución del dengue por estado de la enfermedad año 2023: casos confirmados

```
conteo_enfermedad <- dengue_data_confirmado[dengue_data_confirmado$ano == "2023", ] %>%
  group_by(enfermedad) %>%
  summarise(conteo = n()) %>%
  arrange(conteo) # Ordenar por conteo

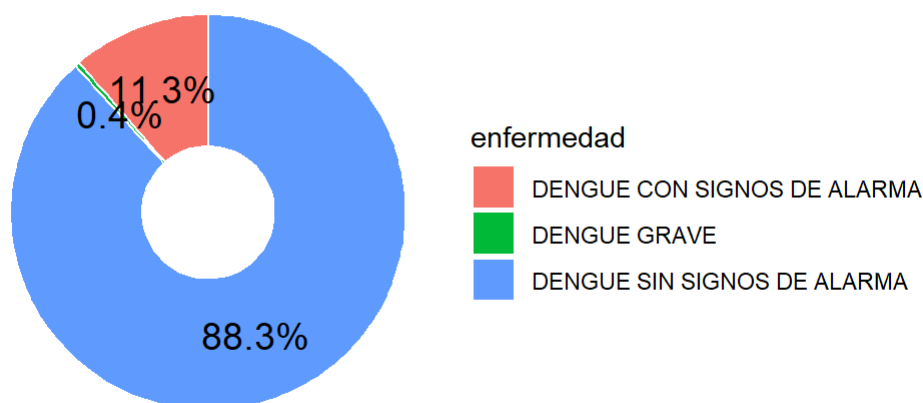
print(conteo_enfermedad)
```

```
## # A tibble: 3 × 2
##   enfermedad      conteo
##   <fct>          <int>
## 1 DENGUE GRAVE      999
## 2 DENGUE CON SIGNOS DE ALARMA 28538
## 3 DENGUE SIN SIGNOS DE ALARMA 222068
```

```
conteo_enfermedad$porcentaje <- round(conteo_enfermedad$conteo / sum(conteo_enfermedad$conteo) * 100, 1)
```

```
#Crear gráfico de anillo
ggplot(conteo_enfermedad, aes(x = "", y = conteo, fill = enfermedad)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) + # Coordinar en forma de pastel
  xlim(c(0, "")) + # Crear el hueco central (anillo)
  geom_text(aes(label = paste0(porcentaje, "%")),
            position = position_stack(vjust = 0.5), size = 5) + # Añadir etiquetas
  labs(title = "Gráfico de Anillo con Porcentajes") +
  theme_void() # Eliminar ejes y fondo
```

Gráfico de Anillo con Porcentajes



Distribución del dengue por departamento año 2023: casos confirmados

```
conteo_departamento <- dengue_data_confirmado[dengue_data_confirmado$ano=="2023",] %>%
  group_by(departamento) %>%
  summarise(conteo = n())

conteo_departamento <- conteo_departamento %>%
  arrange(conteo)

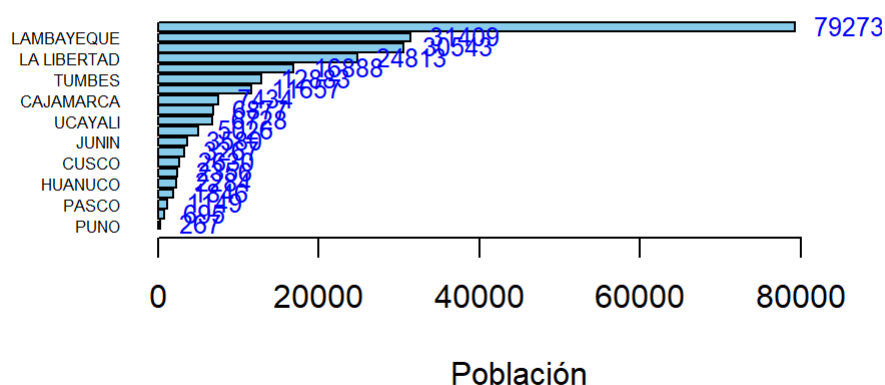
print(conteo_departamento)
```

```
## # A tibble: 20 × 2
##   departamento conteo
##   <fct>         <int>
## 1 PUNO           267
## 2 AYACUCHO       695
## 3 PASCO         1149
## 4 MADRE DE DIOS  1846
## 5 HUANUCO       2284
## 6 CALLAO        2356
## 7 CUSCO         2630
## 8 AMAZONAS      3267
## 9 JUNIN         3580
## 10 LORETO       5026
## 11 UCAYALI      6728
## 12 SAN MARTIN   6877
## 13 CAJAMARCA    7434
## 14 ANCASH      11657
## 15 TUMBES      12883
## 16 ICA         16888
## 17 LA LIBERTAD  24813
## 18 LIMA         30543
## 19 LAMBAYEQUE   31409
## 20 PIURA      79273
```

```
bd <- barplot(conteo_departamento$conteo,
  names.arg = conteo_departamento$departamento, # Etiquetas de las barras
  horiz = TRUE, # Hacerlo horizontal
  las = 1, # Texto de etiquetas en horizontal
  cex.names = 0.5, # Reducir tamaño del texto de las etiquetas
  col = "skyblue", # Color de las barras
  xlim = c(0, 90000), # Extender eje X
  main = "Distribución del dengue por departamento: 2023",
  xlab = "Población")

# Añadir los valores al final de las barras
text(conteo_departamento$conteo, bd, labels = conteo_departamento$conteo, pos = 4, cex = 0.8,
col = "blue")
```

Distribución del dengue por departamento: 2023



ANALISIS DEL

DEPARTAMENTO MAS AFECTADO POR EL DENGUE EN EL AÑO 2023: PIURA ##### Distribución del dengue por provincia: casos confirmados

```
conteo_provincia <- dengue_data_confirmado[dengue_data_confirmado$ano=="2023" & dengue_data_c
onfirmado$departamento=="PIURA",] %>%
  group_by(provincia) %>%
  summarise(conteo = n())

conteo_provincia <- conteo_provincia %>%
  arrange(conteo)

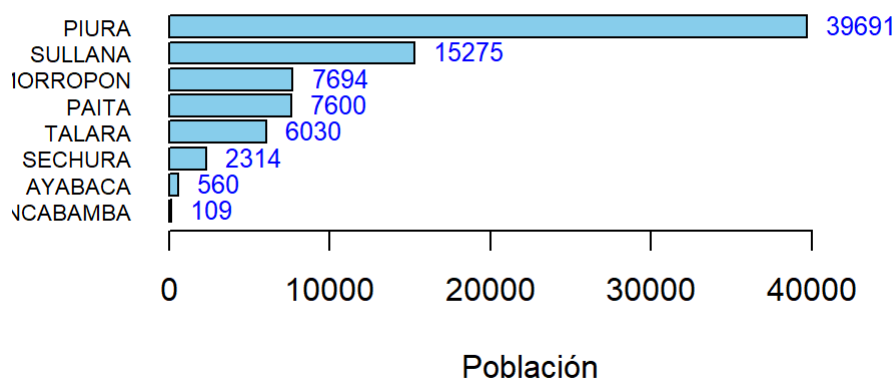
print(conteo_provincia)
```

```
## # A tibble: 8 × 2
##   provincia  conteo
##   <fct>      <int>
## 1 HUANCABAMBA    109
## 2 AYABACA        560
## 3 SECHURA      2314
## 4 TALARA        6030
## 5 PAITA         7600
## 6 MORROPON      7694
## 7 SULLANA      15275
## 8 PIURA       39691
```

```
bp <- barplot(conteo_provincia$conteo,
  names.arg = conteo_provincia$provincia, # Etiquetas de las barras
  horiz = TRUE,                          # Hacerlo horizontal
  las = 1,                               # Texto de etiquetas en horizontal
  cex.names = 0.7,                       # Reducir tamaño del texto de las etiquetas
  col = "skyblue",                       # Color de las barras
  xlim = c(0, 45000),                   # Extender eje X
  main = "Distribución del dengue por provincia: caso Piura",
  xlab = "Población")

# Añadir los valores al final de las barras
text(conteo_provincia$conteo, bp, labels = conteo_provincia$conteo, pos = 4, cex = 0.8, col =
"blue")
```

Distribución del dengue por provincia: caso Piura



Distribución del dengue

por distrito: casos confirmados

```
casos_por_distrito <- dengue_data %>%
  group_by(distrito) %>%
  summarise(Casos = n()), .groups = 'drop') %>%
  arrange(desc(Casos))

distrito_mas_afectado <- casos_por_distrito[1, ]

dengue_distrito_afectado <- dengue_data %>%
  filter(distrito == distrito_mas_afectado$distrito & tipo_dx=="C")

# Casos por distrito y sexo en gráfico circular con porcentajes y colores pastel
ggplot(dengue_distrito_afectado, aes(x = "", fill = sexo)) +
  geom_bar(stat = "count", width = 1) +
  coord_polar(theta = "y") +
  theme_minimal() +
  labs(title = paste("Casos de Dengue Sexo vs Distrito de", distrito_mas_afectado$distrito),
       fill = "Sexo") +
  scale_fill_manual(values = c("M" = "#A8DADC", "F" = "#FFB6B9")) +
  theme(
    axis.text = element_blank(),
    axis.title = element_blank(),
    plot.title = element_text(color = "#3E4A59", size = 18, face = "bold")
  ) +
  geom_text(aes(label = paste0(round(..count../sum(..count..)*100, 1), "%")),
            stat = "count",
            position = position_stack(vjust = 0.5),
            color = "#FFFFFF", size = 6)
```

```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(count)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Casos de Dengue Sexo vs Distrito

