

Estadística para la Ciencia de Datos

Exámen parcial: Análisis estadístico exploratorio

Manuel Alejandro Serrano Macias

Fecha de modificación: 2025-10-30 23:51:44.541514

Contents

Reporte de Análisis Estadístico Exploratorio de Dengue	1
Prerequisitos	1
1. Resumen edad promedio, desviación estandar por sexo	2
2. Gráfico comparativo de edad por estado y sexo	3
3. Gráfico de densidad	4
4. Gráfico de densidad de los municipios con mayor cantidad de casos	5
5. Gráficos de barras de frecuencia mensual (Enero a Septiembre)	8

Reporte de Análisis Estadístico Exploratorio de Dengue

A continuación se presenta un análisis estadístico exploratorio de los casos de dengue en los estados de Michoacán, Colima, Nayarit, Jalisco y Guerrero.

Prerequisitos

1. Se instalan y preparan las librerías a utilizar durante el análisis.

```
# install.packages("readxl")  
# install.packages("lubridate")
```

```
library(readxl)  
library(lubridate)
```

2. Captura y limpieza de datos para el análisis. Se obtienen las entidades federativas que se incluirán en el estudio eligiendo los estados de Michoacán, Colima, Nayarit, Jalisco y Guerrero mediante las siguientes instrucciones:

```
dengue_datos_base <- read.csv("dengue_abierto.csv")  
  
catalogo_sexo <- read_excel("Catálogos_Dengue.xlsx", sheet = "CATÁLOGO SEXO")  
catalogo_entidad <- read_excel("Catálogos_Dengue.xlsx", sheet = "CATÁLOGO ENTIDAD")  
catalogo_municipio <- read_excel("Catálogos_Dengue.xlsx", sheet = "CATÁLOGO MUNICIPIO")  
  
names(catalogo_sexo) <- c("CLAVE_SEXO", "DESCRIPCIÓN_CATALOGO_SEXO")  
catalogo_entidad$CLAVE_ENTIDAD <- as.numeric(catalogo_entidad$CLAVE_ENTIDAD)  
estados_estudio <- c(16, 6, 18, 14, 12)  
  
dengue_datos_estudio <- dengue_datos_base[dengue_datos_base$ENTIDAD_RES %in% estados_estudio, ]
```

```
dengue_datos_estudio <- merge(dengue_datos_estudio, catalogo_sexo,
                              by.x = "SEXO", by.y = "CLAVE_SEXO")

dengue_datos_estudio <- merge(dengue_datos_estudio, catalogo_entidad,
                              by.x = "ENTIDAD_RES", by.y = "CLAVE_ENTIDAD")
```

1. Resumen edad promedio, desviación estandar por sexo

Se calculan la edad promedio y desviación estandar, por sexo, de las entidades federativas incluidas en el análisis.

```
edad_promedio <- aggregate(EDAD_ANOS ~ ENTIDAD_FEDERATIVA + DESCRIPCIÓN_CATALOGO_SEXO,
                           data = dengue_datos_estudio,
                           FUN = mean,
                           na.rm = TRUE)

desviacion_estandar <- aggregate(EDAD_ANOS ~ ENTIDAD_FEDERATIVA + DESCRIPCIÓN_CATALOGO_SEXO,
                                 data = dengue_datos_estudio,
                                 FUN = sd,
                                 na.rm = TRUE)

names(edad_promedio)[names(edad_promedio) == "EDAD_ANOS"] <- "EDAD_PROMEDIO"
names(desviacion_estandar)[names(desviacion_estandar) == "EDAD_ANOS"] <- "DESVIACION_ESTANDAR"

tabla_resumen_edad <- merge(edad_promedio, desviacion_estandar,
                             by = c("ENTIDAD_FEDERATIVA", "DESCRIPCIÓN_CATALOGO_SEXO"))
tabla_resumen_edad
```

ENTIDAD_FEDERATIVA	DESCRIPCIÓN_CATALOGO_SEXO	EDAD_PROMEDIO	DESVIACION_ESTANDAR
COLIMA	HOMBRE	23.91371	17.26406
COLIMA	MUJER	25.85981	17.36246
GUERRERO	HOMBRE	22.40599	18.06166
GUERRERO	MUJER	24.80821	18.17591
JALISCO	HOMBRE	30.91933	18.22385
JALISCO	MUJER	34.13147	18.10152
MICHOACÁN DE OCAMPO	HOMBRE	27.28045	18.57407
MICHOACÁN DE OCAMPO	MUJER	31.27589	18.59600
NAYARIT	HOMBRE	29.89813	18.94738
NAYARIT	MUJER	31.71562	18.26106

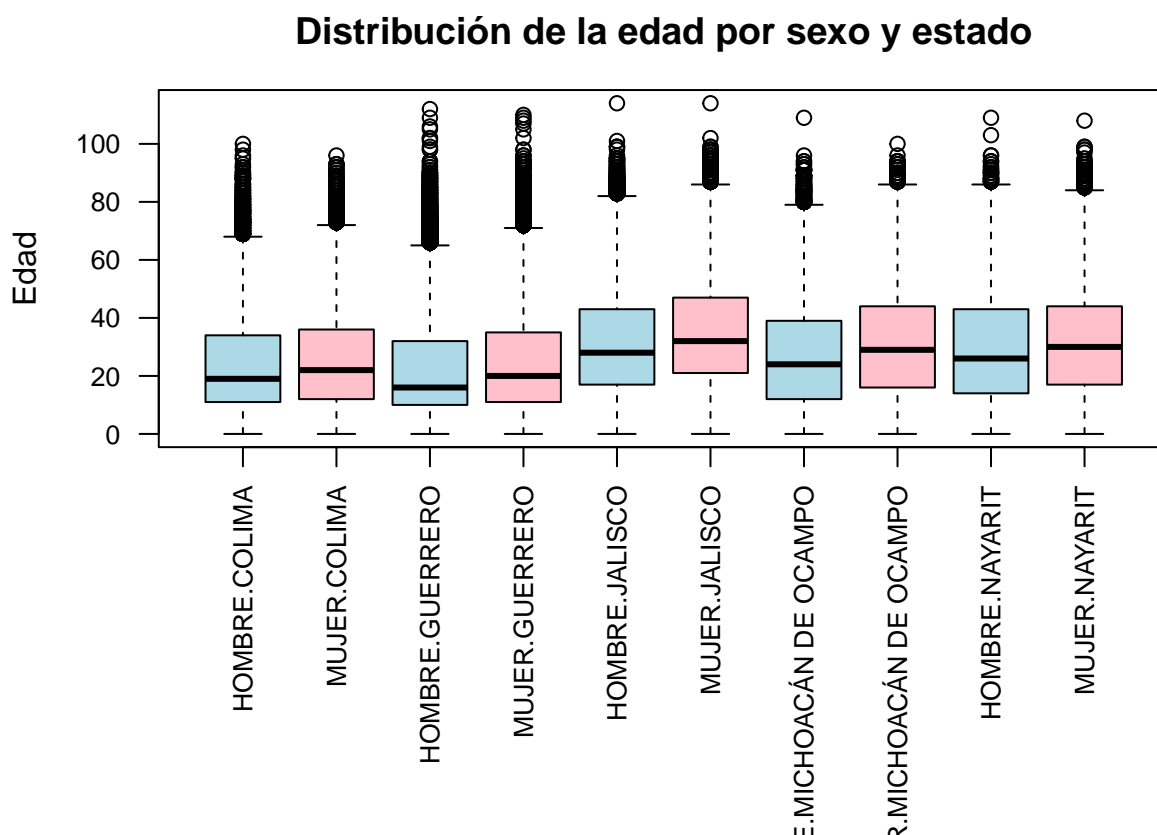
Observaciones

- En todos los estados, la edad promedio de mujeres es ligeramente superior a la de los hombres.
- La desviación estándar es similar entre ambos sexos en todos los estados, indicando una variación de edad comparable en la población de casos de dengue.
- Los casos de Michoacán y Nayarit tienden a tener un promedio de edad ligeramente menor que el resto de los estados.

2. Gráfico comparativo de edad por estado y sexo

Se genera el gráfico comparativo mediante la función `boxplot` con la fórmula para comparar edad por género y por estado.

```
par(mar = c(10, 4, 3, 2) + 0.1)
boxplot(EDAD_ANOS ~ DESCRIPCIÓN_CATALOGO_SEXO * ENTIDAD_FEDERATIVA,
  data = dengue_datos_estudio,
  main = "Distribución de la edad por sexo y estado",
  xlab = "",
  ylab = "Edad",
  col = rep(c("lightblue", "pink"), 5),
  las = 2,
  cex.axis = 0.8)
```



```
par(mar = c(5, 4, 4, 2) + 0.1)
```

Observaciones

- La mediana de la edad en todos los grupos se sitúa alrededor de los 25 años, lo que indica que la mitad de los casos de dengue se concentran en la población joven adulta.
- La forma de las cajas muestra una distribución de edad con una ligera asimetría hacia edades mayores. El rango intercuartílico abarca edades de aproximadamente 18 a 40 años en la mayoría de los grupos, confirmando alta incidencia en este rango.
- Valores Atípicos. Se observa una gran cantidad de puntos que representan valores atípicos, especialmente en edades avanzadas (más de 60 años).
- Comparación por sexo y estado: Las cajas de mujeres están colocadas un poco más arriba que las de hombres en cada estado, confirmando que la edad promedio y la mediana de las pacientes es ligeramente mayor.

mayor, tal como se observó en la tabla resumen del punto anterior.

3. Gráfico de densidad

Se genera el gráfico de densidad para visualizar la distribución de la edad para cada estado, permitiendo la comparación de las formas de las distribuciones. Se separan los datos de edades por estado, posteriormente se calcula la densidad y finalmente se genera el gráfico de densidades agregando cada estado.

```
colima_edad <-
  dengue_datos_estudio[dengue_datos_estudio$ENTIDAD_FEDERATIVA == "COLIMA", ]$EDAD_ANOS

guerrero_edad <-
  dengue_datos_estudio[dengue_datos_estudio$ENTIDAD_FEDERATIVA == "GUERRERO", ]$EDAD_ANOS

jalisco_edad <-
  dengue_datos_estudio[dengue_datos_estudio$ENTIDAD_FEDERATIVA == "JALISCO", ]$EDAD_ANOS

michoacan_edad <-
  dengue_datos_estudio[dengue_datos_estudio$ENTIDAD_FEDERATIVA == "MICHOACÁN DE OCAMPO", ]$EDAD_ANOS

nayarit_edad <-
  dengue_datos_estudio[dengue_datos_estudio$ENTIDAD_FEDERATIVA == "NAYARIT", ]$EDAD_ANOS

densidad_colima <- density(colima_edad, na.rm = TRUE)
densidad_guerrero <- density(guerrero_edad, na.rm = TRUE)
densidad_jalisco <- density(jalisco_edad, na.rm = TRUE)
densidad_michoacan <- density(michoacan_edad, na.rm = TRUE)
densidad_nayarit <- density(nayarit_edad, na.rm = TRUE)

rango_x <- range(c(densidad_colima$x,
  densidad_guerrero$x,
  densidad_jalisco$x,
  densidad_michoacan$x,
  densidad_nayarit$x))

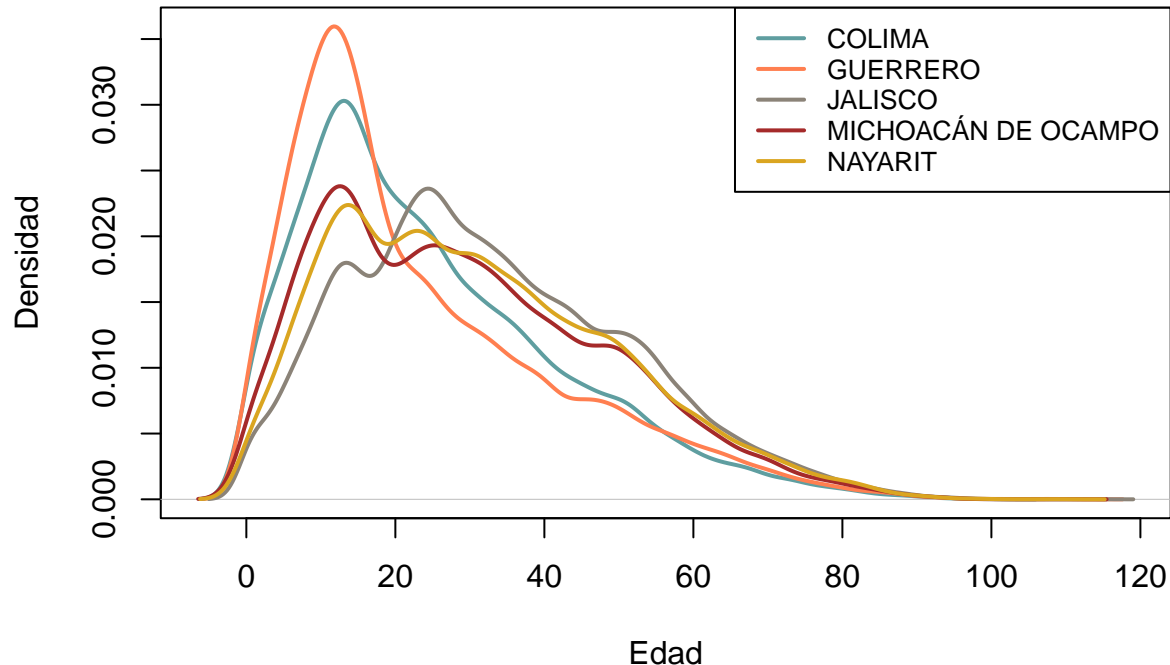
rango_y <- range(c(densidad_colima$y,
  densidad_guerrero$y,
  densidad_jalisco$y,
  densidad_michoacan$y,
  densidad_nayarit$y))

plot(densidad_colima, main = "Gráfico de densidad de edad de casos por estado",
  xlab = "Edad",
  ylab = "Densidad",
  xlim = rango_x,
  ylim = rango_y,
  col = "cadetblue",
  lwd = 2)
lines(densidad_guerrero, col = "coral", lwd = 2)
lines(densidad_jalisco, col = "antiquewhite4", lwd = 2)
lines(densidad_michoacan, col = "brown", lwd = 2)
lines(densidad_nayarit, col = "goldenrod", lwd = 2)

legend("topright",
  legend = c("COLIMA", "GUERRERO", "JALISCO", "MICHOACÁN DE OCAMPO", "NAYARIT"),
```

```
col = c("cadetblue", "coral", "antiquewhite4", "brown", "goldenrod"),
lty = 1, lwd = 2, cex = 0.8)
```

Gráfico de densidad de edad de casos por estado



```
par(mar = c(5, 4, 4, 2) + 0.1)
```

Observaciones

- Las curvas de densidad son similares en forma y ubicación para los cinco estados. Esto sugiere que el patrón de la edad de los casos de dengue es consistente en esos estados del país.
- Todas las curvas tienen un pico de densidad entre los 15 y 25 años. Esto confirma que los adolescentes y jóvenes son el grupo de edad con mayor registro de casos.
- Las colas de las distribuciones se extienden más hacia la derecha (edades mayores), indicando que, aunque la mayoría de los casos son jóvenes, existe una dispersión significativa de casos en adultos y adultos mayores.

4. Gráfico de densidad de los municipios con mayor cantidad de casos

Se genera el gráfico de densidad para visualizar la distribución de los municipios con mayor cantidad de caso por estado, permitiendo la comparación de las formas de las distribuciones. Primero se identifican los municipios con mayor cantidad de casos. Posteriormente, se separan los datos por los municipios identificados y se calcula la densidad. Finalmente se genera el gráfico de densidades agregando cada municipio identificado.

```
top_municipio <- function(datos, codigo_estado) {
  datos_estado <- datos[datos$ENTIDAD_RES == codigo_estado, ]
  conteo_municipios <- table(datos_estado$MUNICIPIO_RES)
  top_mun_code <- names(conteo_municipios)[which.max(conteo_municipios)]
  return(top_mun_code)
}

colima_top_municipio <- top_municipio(dengue_datos_estudio, 6)
```

```

guerrero_top_municipio <- top_municipio(dengue_datos_estudio, 12)
jalisco_top_municipio <- top_municipio(dengue_datos_estudio, 14)
michoacan_top_municipio <- top_municipio(dengue_datos_estudio, 16)
nayarit_top_municipio <- top_municipio(dengue_datos_estudio, 18)

datos_top_municipios <- data.frame(
  ENTIDAD_RES = c(6, 12, 14, 16, 18),
  MUNICIPIO_RES = c(colima_top_municipio,
                    guerrero_top_municipio,
                    jalisco_top_municipio,
                    michoacan_top_municipio,
                    nayarit_top_municipio)
)

datos_top_municipios$MUNICIPIO_RES <- as.numeric(datos_top_municipios$MUNICIPIO_RES)

catalogo_municipio$CLAVE_MUNICIPIO <- as.numeric(catalogo_municipio$CLAVE_MUNICIPIO)
catalogo_municipio$CLAVE_ENTIDAD <- as.numeric(catalogo_municipio$CLAVE_ENTIDAD)

## Warning: NAs introduced by coercion
top_municipios_nombres <- merge(datos_top_municipios, catalogo_municipio,
                                by.x = c("ENTIDAD_RES", "MUNICIPIO_RES"),
                                by.y = c("CLAVE_ENTIDAD", "CLAVE_MUNICIPIO"),
                                all.x = TRUE)

top_municipios_nombres <-
  merge(top_municipios_nombres, catalogo_entidad, by.x = "ENTIDAD_RES", by.y = "CLAVE_ENTIDAD")
top_municipios_nombres$MUNICIPIO_ESTADO <-
  paste(top_municipios_nombres$MUNICIPIO, " (", top_municipios_nombres$ENTIDAD_FEDERATIVA, ")", sep="")

mun1 <- dengue_datos_estudio$ENTIDAD_RES == 6 & dengue_datos_estudio$MUNICIPIO_RES == colima_top_municipio
mun2 <- dengue_datos_estudio$ENTIDAD_RES == 12 & dengue_datos_estudio$MUNICIPIO_RES == guerrero_top_municipio
mun3 <- dengue_datos_estudio$ENTIDAD_RES == 14 & dengue_datos_estudio$MUNICIPIO_RES == jalisco_top_municipio
mun4 <- dengue_datos_estudio$ENTIDAD_RES == 16 & dengue_datos_estudio$MUNICIPIO_RES == michoacan_top_municipio
mun5 <- dengue_datos_estudio$ENTIDAD_RES == 18 & dengue_datos_estudio$MUNICIPIO_RES == nayarit_top_municipio

datos_top_mun <- dengue_datos_estudio[mun1 | mun2 | mun3 | mun4 | mun5, ]

datos_top_mun$ID_COMPUESTO <- paste(datos_top_mun$ENTIDAD_RES, datos_top_mun$MUNICIPIO_RES, sep = "_")
top_mun_vista <- top_municipios_nombres[, c("ENTIDAD_RES", "MUNICIPIO_RES", "MUNICIPIO_ESTADO")]

top_mun_vista$ID_COMPUESTO <- paste(top_mun_vista$ENTIDAD_RES, top_mun_vista$MUNICIPIO_RES, sep = "_")

datos_top_mun_final <- merge(datos_top_mun, top_mun_vista[, c("ID_COMPUESTO", "MUNICIPIO_ESTADO")],
                             by = "ID_COMPUESTO", all.x = TRUE)

muni_list <- unique(datos_top_mun_final$MUNICIPIO_ESTADO)
densidades_muni <- list()
colores <- c("cadetblue", "coral", "antiquewhite4", "brown", "goldenrod")
max_y <- 0

colima_mun_edad <- datos_top_mun_final[datos_top_mun_final$ENTIDAD_FEDERATIVA == "COLIMA", ]$EDAD_ANOS
densidad_colima_mun <- density(colima_mun_edad, na.rm = TRUE)

```

```

max_y <- max(max_y, max(densidad_colima_mun$y))

guerrero_mun_edad <- datos_top_mun_final[datos_top_mun_final$ENTIDAD_FEDERATIVA == "GUERRERO", ]$EDAD_AÑOS
densidad_guerrero_mun <- density(guerrero_mun_edad, na.rm = TRUE)
max_y <- max(max_y, max(densidad_guerrero_mun$y))

jalisco_mun_edad <- datos_top_mun_final[datos_top_mun_final$ENTIDAD_FEDERATIVA == "JALISCO", ]$EDAD_AÑOS
densidad_jalisco_mun <- density(jalisco_mun_edad, na.rm = TRUE)
max_y <- max(max_y, max(densidad_jalisco_mun$y))

michoacan_mun_edad <- datos_top_mun_final[datos_top_mun_final$ENTIDAD_FEDERATIVA == "MICHOACÁN DE OCAMPO", ]$EDAD_AÑOS
densidad_michoacan_mun <- density(michoacan_mun_edad, na.rm = TRUE)
max_y <- max(max_y, max(densidad_michoacan_mun$y))

nayarit_mun_edad <- datos_top_mun_final[datos_top_mun_final$ENTIDAD_FEDERATIVA == "NAYARIT", ]$EDAD_AÑOS
densidad_nayarit_mun <- density(nayarit_mun_edad, na.rm = TRUE)
max_y <- max(max_y, max(densidad_nayarit_mun$y))

top_mun_vista

```

ENTIDAD_RES	MUNICIPIO_RES	MUNICIPIO_ESTADO	ID_COMPUESTO
6	7	MANZANILLO (COLIMA)	6_7
12	1	ACAPULCO DE JUÁREZ (GUERRERO)	12_1
14	39	GUADALAJARA (JALISCO)	14_39
16	53	MORELIA (MICHOACÁN DE OCAMPO)	16_53
18	17	TEPIC (NAYARIT)	18_17

```

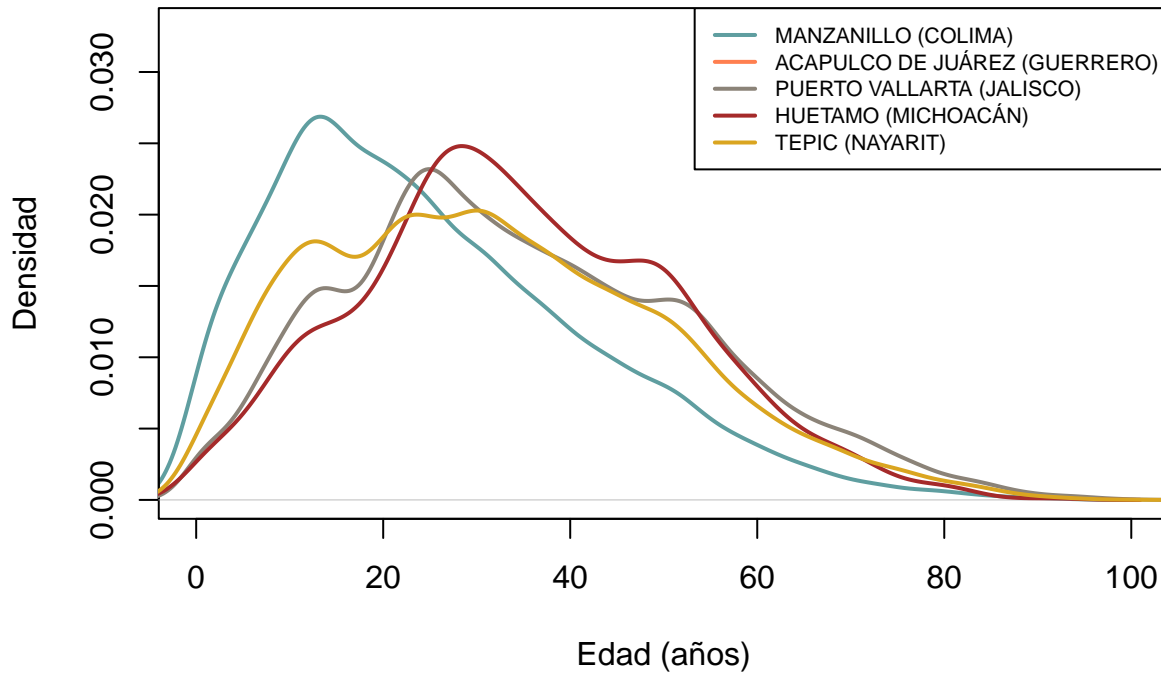
plot(densidad_colima_mun,
     main = "Densidad de edad en los municipios con más casos",
     xlab = "Edad (años)",
     ylab = "Densidad",
     xlim = c(0, 100),
     ylim = c(0, max_y * 1.05),
     col = "cadetblue",
     lwd = 2)

# lines(guerrero_mun_edad, col = "coral", lwd = 2) # Revisar antes de entregar
lines(densidad_jalisco_mun, col = "antiquewhite4", lwd = 2)
lines(densidad_michoacan_mun, col = "brown", lwd = 2)
lines(densidad_nayarit_mun, col = "goldenrod", lwd = 2)

legend("topright",
     legend = c("MANZANILLO (COLIMA)", "ACAPULCO DE JUÁREZ (GUERRERO)", "PUERTO VALLARTA (JALISCO)", "MORÉLIA (MICHOACÁN DE OCAMPO)", "TEPIC (NAYARIT)"),
     col = colores,
     lty = 1, lwd = 2, cex = 0.7)

```

Densidad de edad en los municipios con más casos



Observaciones

- La distribución de la edad en los municipios con mayores casos sigue un patrón similar al estatal, aunque con un rango más amplio en la población joven (15-25 años) y adulta (25 - 35) años.
- La curva del municipio principal de Nayarit, Tepic, muestra una mayor densidad en el rango de 20-45 años. Esto podría indicar que la población afectada en Tepic es ligeramente mayor en promedio que en los demás municipios principales.
- La curva del municipio principal de Colima, Manzanillo, la más desplazada a la izquierda, lo que indica una mayor concentración de casos en la población más joven (cerca de 15 años).

5. Gráficos de barras de frecuencia mensual (Enero a Septiembre)

Se analiza la frecuencia de registro de nuevos casos, basándose en la fecha de inicio de síntomas. Se preparan los datos extrayendo la fecha y filtrando los meses de Enero a Septiembre que son los de interés para este ejercicio. Finalmente se generan los gráficos por cada estado en los meses indicados.

```
dengue_datos_estudio$FECHA <- dmy(dengue_datos_estudio$FECHA_SIGN_SINTOMAS)

dengue_datos_estudio$MES_NUMERO <- month(dengue_datos_estudio$FECHA)
dengue_datos_estudio$MES_NOMBRE <-
  month(dengue_datos_estudio$FECHA, label = TRUE, abbr = FALSE, locale = "es_ES.UTF-8")

datos_meses <-
  dengue_datos_estudio[dengue_datos_estudio$MES_NUMERO >= 1 & dengue_datos_estudio$MES_NUMERO <= 9, ]
frecuencia_mensual <- table(datos_meses$ENTIDAD_FEDERATIVA, datos_meses$MES_NOMBRE)

par(mfrow = c(2, 3), mar = c(4, 4, 3, 1))
colores_estado <- c("cadetblue", "coral", "antiquewhite4", "brown", "goldenrod")
orden_meses <- c("enero", "febrero", "marzo", "abril", "mayo", "junio", "julio", "agosto", "septiembre")
```



```
frecuencia_mensual <- frecuencia_mensual[, orden_meses]

estados_ordenados <- rownames(frecuencia_mensual)

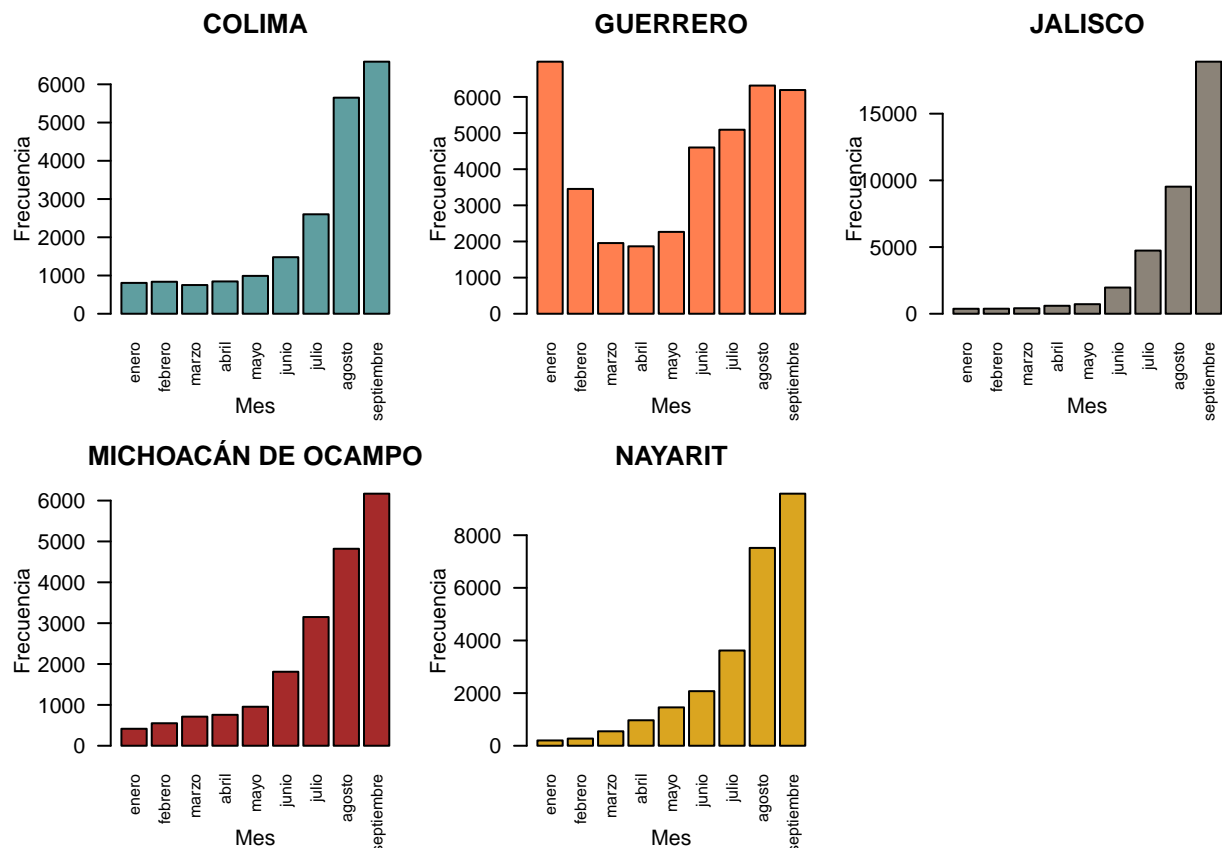
colima_freq <- frecuencia_mensual["COLIMA", ]
barplot(colima_freq, main = estados_ordenados[1], col = colores_estado[1],
        ylab = "Frecuencia", xlab = "Mes", las = 2, cex.names = 0.7)

guerrero_freq <- frecuencia_mensual["GUERRERO", ]
barplot(guerrero_freq, main = estados_ordenados[2], col = colores_estado[2],
        ylab = "Frecuencia", xlab = "Mes", las = 2, cex.names = 0.7)

jalisco_freq <- frecuencia_mensual["JALISCO", ]
barplot(jalisco_freq, main = estados_ordenados[3], col = colores_estado[3],
        ylab = "Frecuencia", xlab = "Mes", las = 2, cex.names = 0.7)

michoacan_freq <- frecuencia_mensual["MICHOACÁN DE OCAMPO", ]
barplot(michoacan_freq, main = estados_ordenados[4], col = colores_estado[4],
        ylab = "Frecuencia", xlab = "Mes", las = 2, cex.names = 0.7)

nayarit_freq <- frecuencia_mensual["NAYARIT", ]
barplot(nayarit_freq, main = estados_ordenados[5], col = colores_estado[5],
        ylab = "Frecuencia", xlab = "Mes", las = 2, cex.names = 0.7)
```



Observaciones

- En casi todos los estados se observa un patrón estacional con una baja incidencia o muy pocos casos

registrados en los meses de Enero a Mayo.

- Por otro lado, Guerrero muestra una alta incidencia en Enero y Febrero, posteriormente, en Marzo - Abril se reportan menos casos registrados.
- Un aumento sostenido y acelerado a partir de Junio y Julio en todos los estados.
- Los picos de registro de nuevos casos se alcanzan en los meses de Agosto y Septiembre, que coinciden con la etapa final de la temporada de lluvias y las altas temperaturas, condiciones favorables para la proliferación del mosquito del dengue.