

# Código Depuración

Rosalia

2025-07-30

## Contents

<b>1 Paquetes</b>	<b>1</b>
<b>2 Carga de datos</b>	<b>2</b>
<b>3 Combinación de variables</b>	<b>2</b>
<b>4 Modificando errores en los países</b>	<b>3</b>
<b>5 Seleccionando los países y creando la base de datos completa</b>	<b>4</b>
<b>6 Inicio del Análisis para el reto 1</b>	<b>5</b>
<b>7 Confirmando los campos y planteando preguntas.</b>	<b>5</b>
<b>8 Realizando gráficos generales</b>	<b>6</b>

## 1 Paquetes

```
library(tidyverse)
library(readxl)
library(dplyr)
library(ggplot2)
library(ggrepel)
library(countrycode)
library(scales)
```

## 2 Carga de datos

```
esperanzadevida <- read_excel("Datos/Base_Datos_Original/esperanza_de_vida_mujer.xlsx")
ProductoGDP <- read_excel("Datos/Base_Datos_Original/gdp_pcap.xlsx")
nacimientos <- read_excel("Datos/Base_Datos_Original/nacimientos_por_mujer.xlsx")
```

## 3 Combinación de variables

```
vida_long <- pivot_longer(esperanzadevida, cols= -country, names_to = "year", values_to = "esp")
producto_long <- pivot_longer(ProductoGDP, cols= -country, names_to = "year", values_to = "prod")
nacimientos_long <- pivot_longer(nacimientos, cols= -country, names_to = "year", values_to = "nac")

producto_long %>%
  count(country, year) %>%
  filter(n > 1)

## # A tibble: 0 x 3
## # i 3 variables: country <chr>, year <chr>, n <int>

n_distinct(vida_long$year)

## [1] 151

n_distinct(producto_long$year)

## [1] 301

n_distinct(nacimientos_long$year)

## [1] 301

setdiff(producto_long$country, vida_long$country)

##  [1] "Congo, Dem. Rep."           "Congo, Rep."
##  [3] "Micronesia, Fed. Sts."      "Uk"
##  [5] "Hong kong, China"          "kazakhstan"
##  [7] "kenya"                      "kyrgyz Republic"
##  [9] "kiribati"                   "St. kitts and Nevis"
## [11] "South korea"                "kuwait"
## [13] "St. Lucia"                  "North korea"
## [15] "SlovakRepublic"             "St. Vincent and the Grenadines"
```

## 4 Modificando errores en los países

```
producto_long <- producto_long %>%
  mutate(country = recode(country,
    "Uk" = "United Kingdom",
    "South korea" = "Korea, Rep.",
    "North korea" = "Korea, Dem. Rep.",
    "kazakhstan" = "Kazakhstan",
    "kenya" = "Kenya",
    "kuwait" = "Kuwait",
    "kiribati" = "Kiribati",
    "St. Lucia" = "Saint Lucia",
    "St. kitts and Nevis" = "Saint Kitts and Nevis",
    "St. Vincent and the Grenadines" = "Saint Vincent and the Grenadines",
    "Hong kong, China" = "Hong Kong SAR, China",
    "SlovakRepublic" = "Slovak Republic",
    "kyrgyz Republic" = "Kyrgyz Republic",
    "Micronesia, Fed. Sts." = "Federated States of Micronesia",
    "Micronesia" = "Federated States of Micronesia",
    "Congo, Rep." = "Congo, Rep.",
    "Congo, Dem. Rep." = "Congo, Dem. Rep.",
    "Virgin Islands (U,S,)" = "Virgin Islands (U.S.)"
  ))
```

```
nacimientos_long <- nacimientos_long %>%
  mutate(country = recode(country,
    "Uk" = "United Kingdom",
    "South korea" = "Korea, Rep.",
    "North korea" = "Korea, Dem. Rep.",
    "kazakhstan" = "Kazakhstan",
    "kenya" = "Kenya",
    "kuwait" = "Kuwait",
    "kiribati" = "Kiribati",
    "St. Lucia" = "Saint Lucia",
    "St. kitts and Nevis" = "Saint Kitts and Nevis",
    "St. Vincent and the Grenadines" = "Saint Vincent and the Grenadines",
    "Hong kong, China" = "Hong Kong SAR, China",
    "SlovakRepublic" = "Slovak Republic",
    "kyrgyz Republic" = "Kyrgyz Republic",
    "Micronesia, Fed. Sts." = "Federated States of Micronesia",
    "Micronesia" = "Federated States of Micronesia",
    "Congo, Rep." = "Congo, Rep.",
    "Congo, Dem. Rep." = "Congo, Dem. Rep.",
    "Virgin Islands (U,S,)" = "Virgin Islands (U.S.)"
  ))
```

```

vida_long <- vida_long %>%
  mutate(country = recode(country,
    "Uk" = "United Kingdom",
    "South korea" = "Korea, Rep.",
    "North korea" = "Korea, Dem. Rep.",
    "kazakhstan" = "Kazakhstan",
    "kenya" = "Kenya",
    "kuwait" = "Kuwait",
    "kiribati" = "Kiribati",
    "St. Lucia" = "Saint Lucia",
    "St. kitts and Nevis" = "Saint Kitts and Nevis",
    "St. Vincent and the Grenadines" = "Saint Vincent and the Grenadines",
    "Hong kong, China" = "Hong Kong SAR, China",
    "SlovakRepublic" = "Slovak Republic",
    "kyrgyz Republic" = "Kyrgyz Republic",
    "Micronesia, Fed. Sts." = "Federated States of Micronesia",
    "Micronesia" = "Federated States of Micronesia",
    "Congo, Rep." = "Congo, Rep.",
    "Congo, Dem. Rep." = "Congo, Dem. Rep.",
    "Virgin Islands (U.S.)" = "Virgin Islands (U.S.)"
  ))

```

## 5 Seleccionando los países y creando la base de datos completa

```

setdiff(nacimientos_long$country, vida_long$country)

## character(0)

paises_elegidos <- unique(vida_long$country)

producto_long <- producto_long %>%
  filter(country %in% paises_elegidos) %>%
  distinct(country, year, .keep_all = TRUE)

nacimientos_long <- nacimientos_long %>%
  filter(country %in% paises_elegidos) %>%
  distinct(country, year, .keep_all = TRUE)

base_completa <- vida_long %>%
  left_join(producto_long, by = c("country", "year")) %>%
  left_join(nacimientos_long, by = c("country", "year"))

```

## 6 Inicio del Análisis para el reto 1

```
#Resumen estadístico general de cada variable  
summary(base_completa$esperanza_de_vida)
```

```
##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.    NA's  
##    11.30    68.10   77.00    74.43   83.90   97.40       5
```

```
summary(base_completa$producto_por_capita)
```

```
##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.    NA's  
##    354     4260   13400   24637   38400  244000    7550
```

```
summary(base_completa$nacimientos_por_mujer)
```

```
##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.    NA's  
##    0.710    1.640   2.140    3.008   3.890   8.860   5889
```

## 7 Confirmando los campos y planteando preguntas.

```
# Asegurar que el campo "year" sea numérico  
base_completa$year <- as.numeric(base_completa$year)

# 1. País que más veces ha tenido la mayor esperanza de vida
lideres_por_ano <- base_completa %>%
  group_by(year) %>%
  filter(!is.na(esperanza_de_vida)) %>%
  slice_max(order_by = esperanza_de_vida, n = 1) %>%
  ungroup()

pais_mas_lider <- lideres_por_ano %>%
  count(country, sort = TRUE)

# Mostrar el país con más "primeros lugares" en esperanza de vida
head(pais_mas_lider, 1)
```

```
## # A tibble: 1 x 2
##   country     n
##   <chr>    <int>
## 1 Monaco     65
```

```

# Ver los principales 5 líderes históricos
head(pais_mas_lider, 5)

## # A tibble: 5 x 2
##   country      n
##   <chr>     <int>
## 1 Monaco      65
## 2 Japan       50
## 3 Andorra     27
## 4 Guernsey     9
## 5 San Marino    4

#Tabla resumen por año
resumen_por_anو <- base_completa %>%
  group_by(year) %>%
  summarise(
    promedio = mean(esperanza_de_vida, na.rm = TRUE),
    maximo = max(esperanza_de_vida, na.rm = TRUE),
    minimo = min(esperanza_de_vida, na.rm = TRUE),
    pais_max = country[which.max(esperanza_de_vida)],
    pais_min = country[which.min(esperanza_de_vida)]
  )

# Ver las primeras 10 filas
head(resumen_por_anو, 10)

## # A tibble: 10 x 6
##   year  promedio  maximo  minimo pais_max pais_min
##   <dbl>    <dbl>    <dbl>    <dbl> <chr>    <chr>
## 1 1950     51.9    74.9    21.1 San Marino North Korea
## 2 1951     52.2    75.1    25.3 Guernsey   North Korea
## 3 1952     53.0    75.2    29.7 San Marino Afghanistan
## 4 1953     53.6    76      30.2 Guernsey   Afghanistan
## 5 1954     54.3    75.9    30.4 Guernsey   Afghanistan
## 6 1955     54.8    75.9    30.8 Iceland    Mali
## 7 1956     55.3    79      31     Guernsey   Mali
## 8 1957     55.8    76.8    31.2 Guernsey   Mali
## 9 1958     56.4    76.4    30     Guernsey   South Sudan
## 10 1959    56.9    78      30.2 Guernsey   South Sudan

```

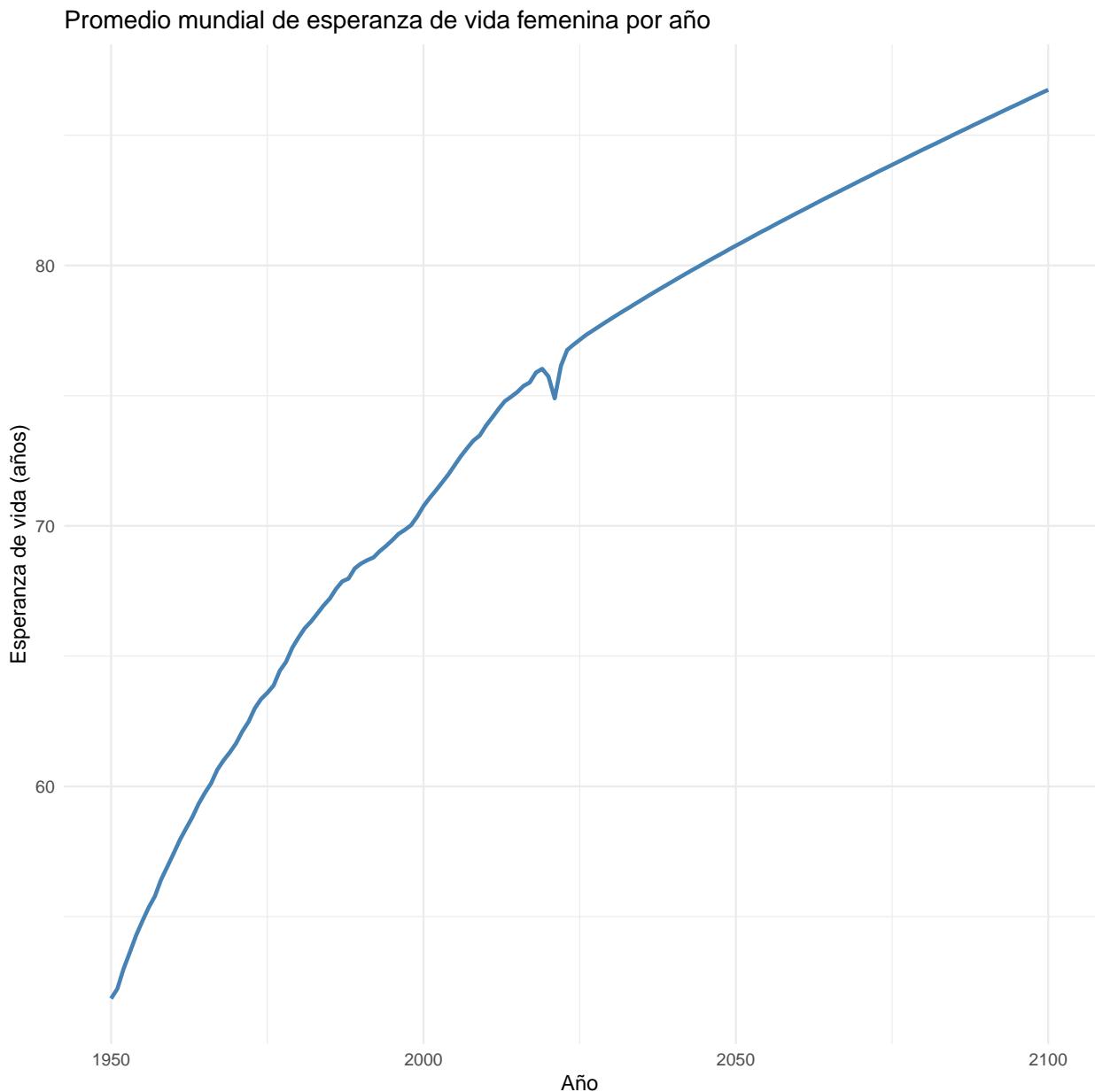
## 8 Realizando gráficos generales

```

ggplot(resumen_por_anو, aes(x = year, y = promedio)) +
  geom_line(color = "steelblue", size = 1)

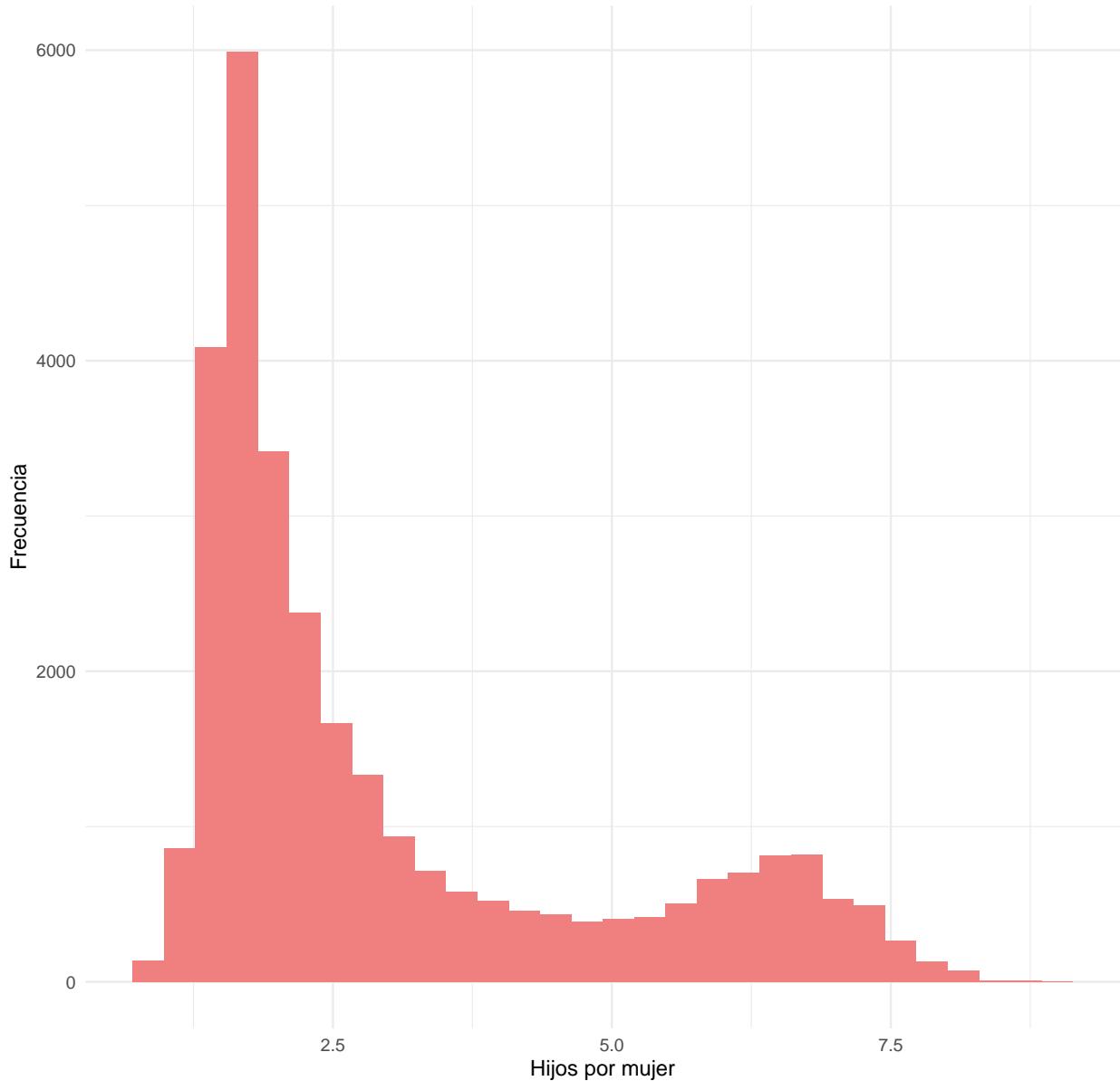
```

```
labs(title = "Promedio mundial de esperanza de vida femenina por año",
     x = "Año",
     y = "Esperanza de vida (años)") +
theme_minimal()
```



```
ggplot(base_completa, aes(x = nacimientos_por_mujer)) +
  geom_histogram(bins = 30, fill = "lightcoral") +
  labs(title = "Distribución de nacimientos por mujer", x = "Hijos por mujer", y = "Frecuencia")
  theme_minimal()
```

### Distribución de nacimientos por mujer



```
#Esto me dice que la mayoría de las mujeres durante los años tienen aproximadamente 1
```

```
names(base_completa)
```

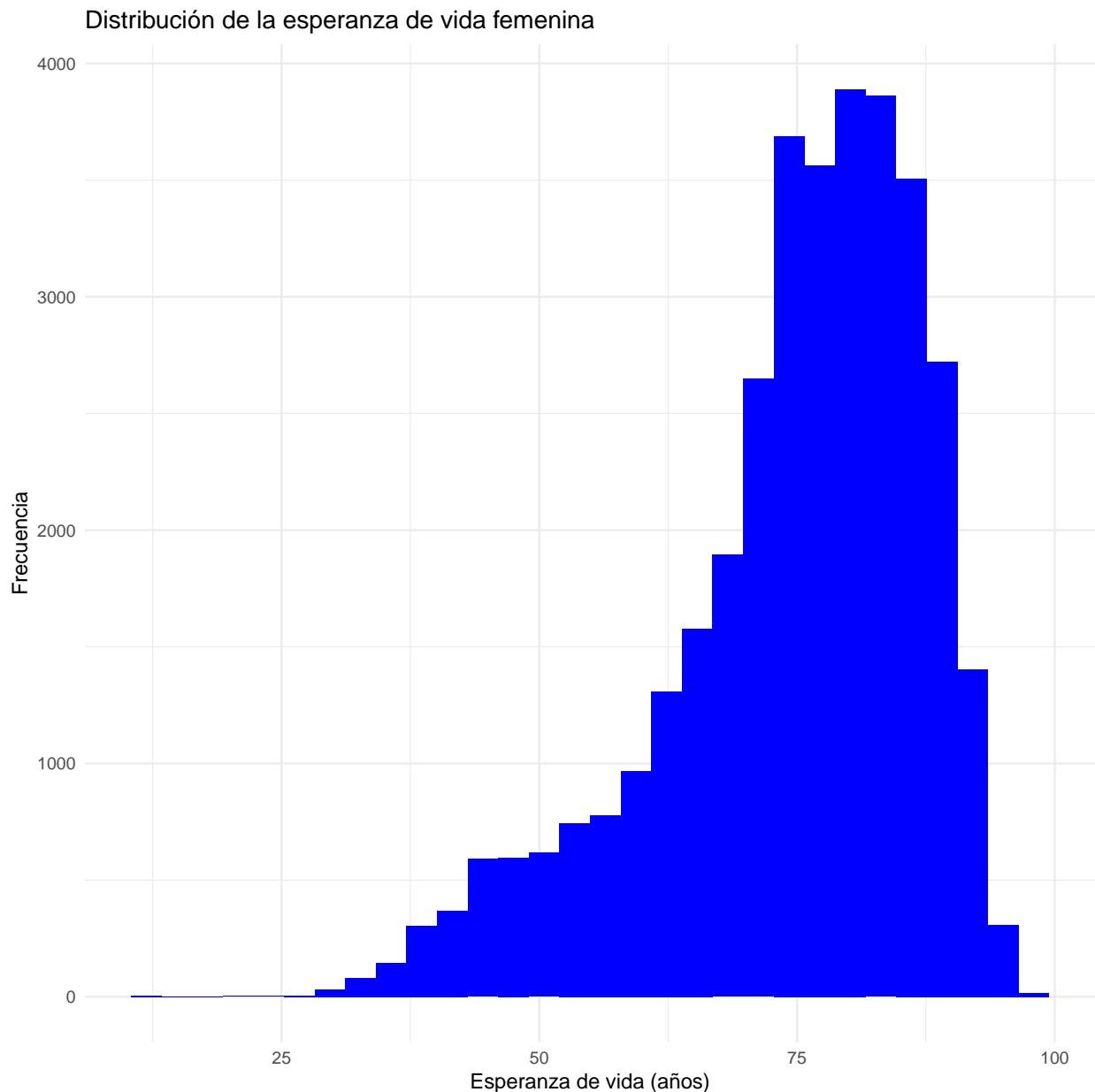
```
## [1] "country"                 "year"                   "esperanza_de_vida"  
## [4] "producto_por_capita"    "nacimientos_por_mujer"
```

```
# Histograma esperanza de vida  
ggplot(base_completa, aes(x = esperanza_de_vida)) +  
  geom_histogram(bins = 30, fill = "blue") +  
  labs(title = "Distribución de la esperanza de vida femenina",
```

```

x= "Esperanza de vida (años)",
y="Frecuencia") +
theme_minimal()

```



#Reto 1 solicitaba Saber si tenía Outliers

```

# Función para detectar outliers
detect_outliers <- function(x) {
  stats <- boxplot.stats(x)$out
  return(stats)
}

```

```

# Aplicar
out_vida <- detect_outliers(base_completa$esperanza_de_vida)
out_nacimientos <- detect_outliers(base_completa$nacimientos_por_muher)

length(out_vida)      # Cantidad de outliers en esperanza de vida

## [1] 1175

length(out_nacimientos) # Cantidad de outliers en fecundidad

## [1] 765

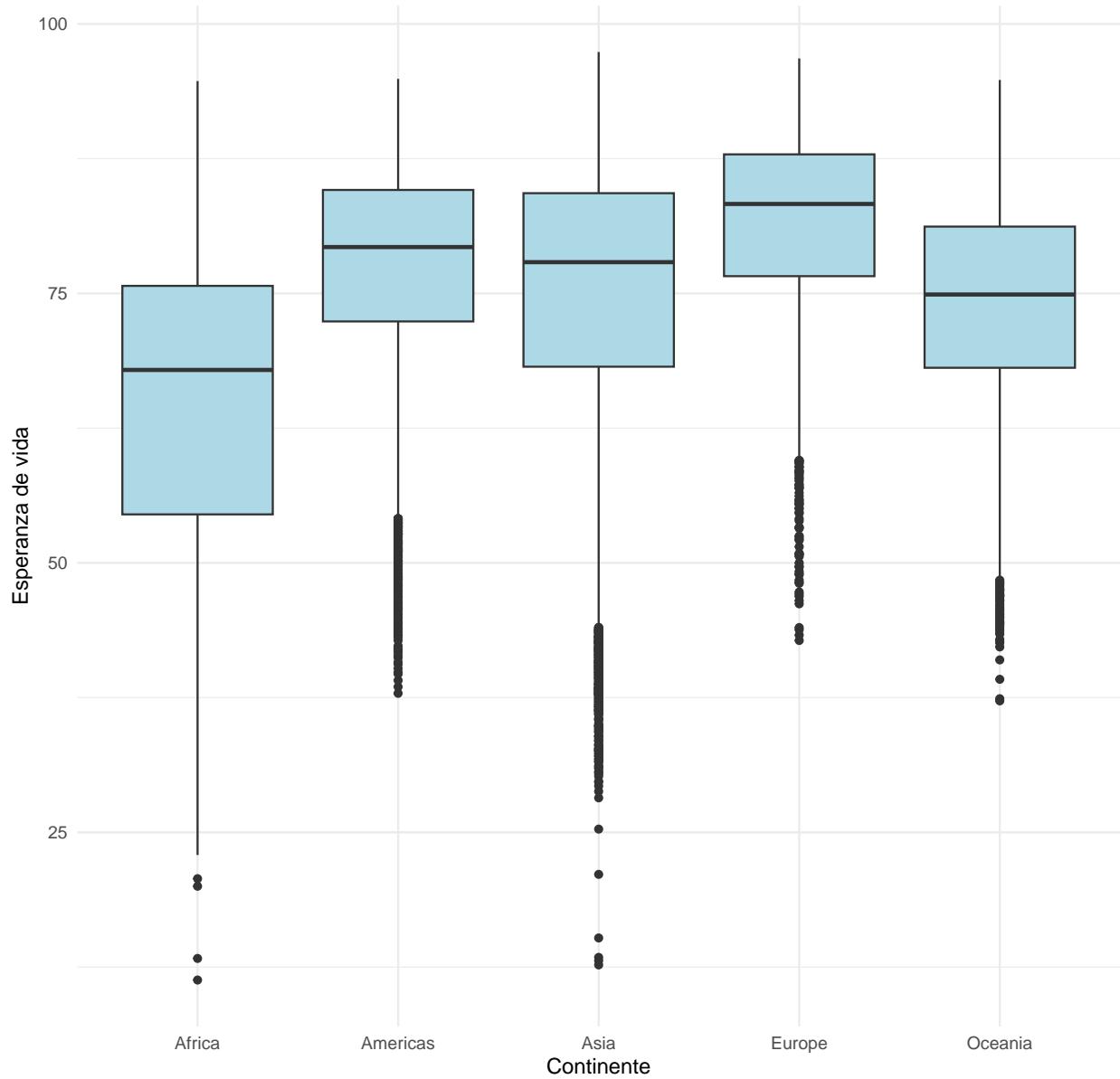
#Quiero dividirlos por continente #Agregué manualmente Kosovo porque no lo reconoce la libreria
countrycoud

base_completa <- base_completa %>%
  mutate(
    continente = countrycode(country, origin = "country.name", destination = "continent"),
    continente = ifelse(is.na(continente) & country == "Kosovo", "Europe", continente)
  )

ggplot(base_completa, aes(x = continente, y = esperanza_de_vida)) +
  geom_boxplot(fill = "lightblue") +
  labs(title = "Esperanza de vida por continente", y = "Esperanza de vida", x = "Continente") +
  theme_minimal()

```

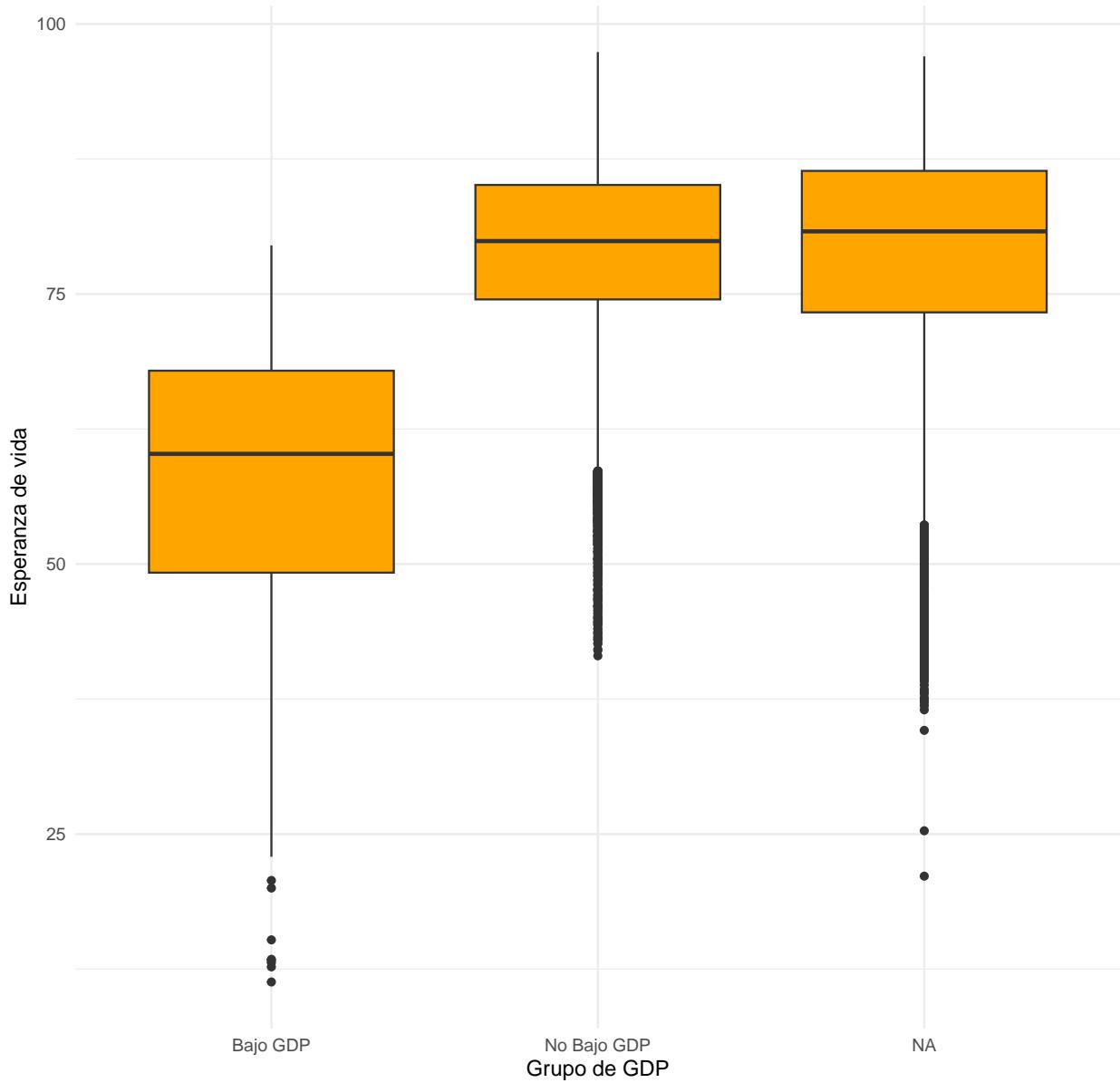
Esperanza de vida por continente



```
base_completa <- base_completa %>%
  mutate(gdp_bajo = ifelse(producto_por_capita < 5000, "Bajo GDP", "No Bajo GDP"))

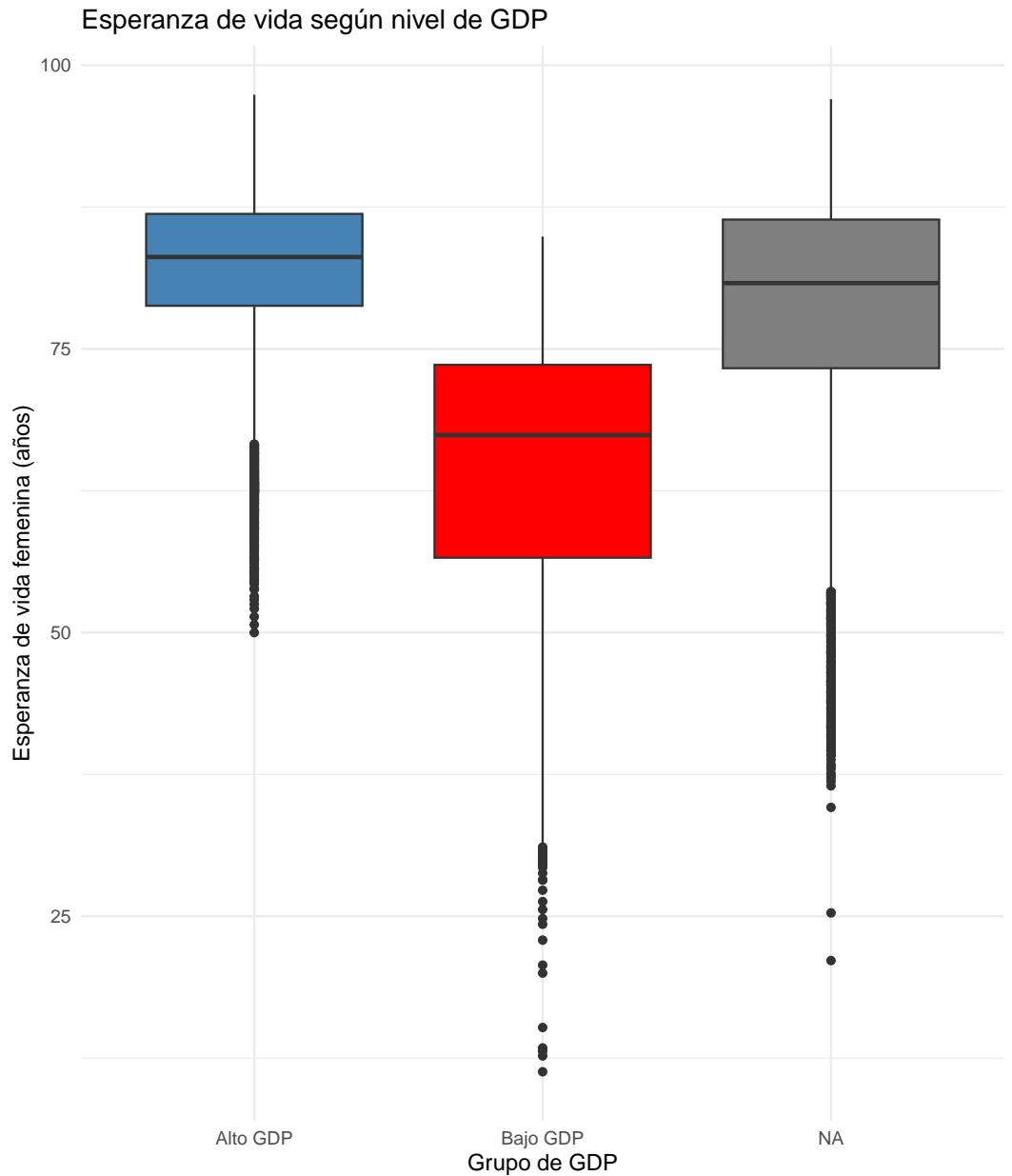
ggplot(base_completa, aes(x = gdp_bajo, y = esperanza_de_vida)) +
  geom_boxplot(fill = "orange") +
  labs(title = "Esperanza de vida según nivel de GDP", y = "Esperanza de vida", x = "Grupo de GDP")
  theme_minimal()
```

### Esperanza de vida según nivel de GDP

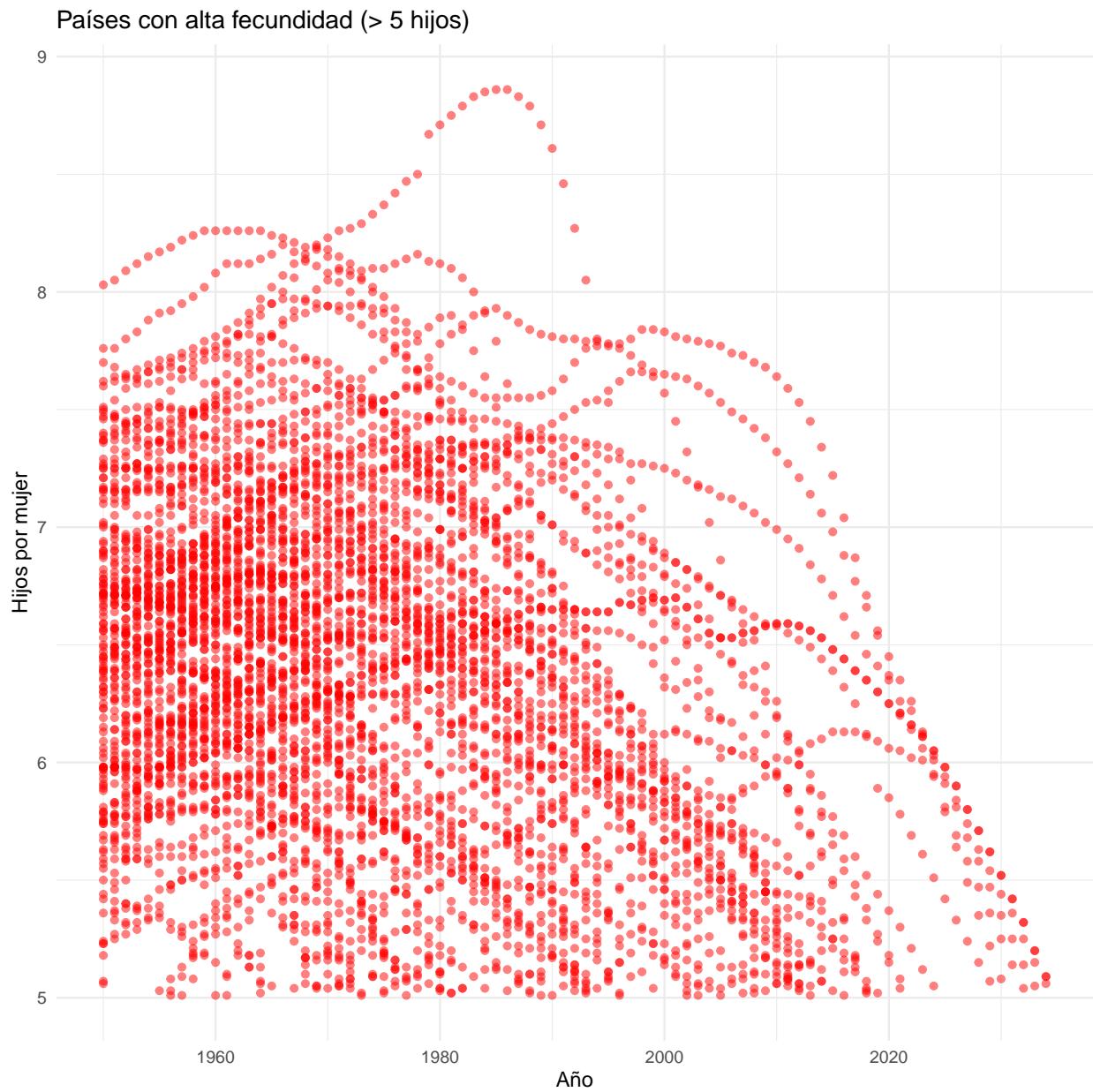


```
base_completa <- base_completa %>%
  mutate(gdp_grupo = ifelse(producto_por_capita > 13845, "Alto GDP", "Bajo GDP"))

ggplot(base_completa, aes(x = gdp_grupo, y = esperanza_de_vida, fill = gdp_grupo)) +
  geom_boxplot() +
  labs(title = "Esperanza de vida según nivel de GDP",
       x = "Grupo de GDP",
       y = "Esperanza de vida femenina (años)") +
  scale_fill_manual(values = c("Bajo GDP" = "red", "Alto GDP" = "steelblue")) +
  theme_minimal()
```

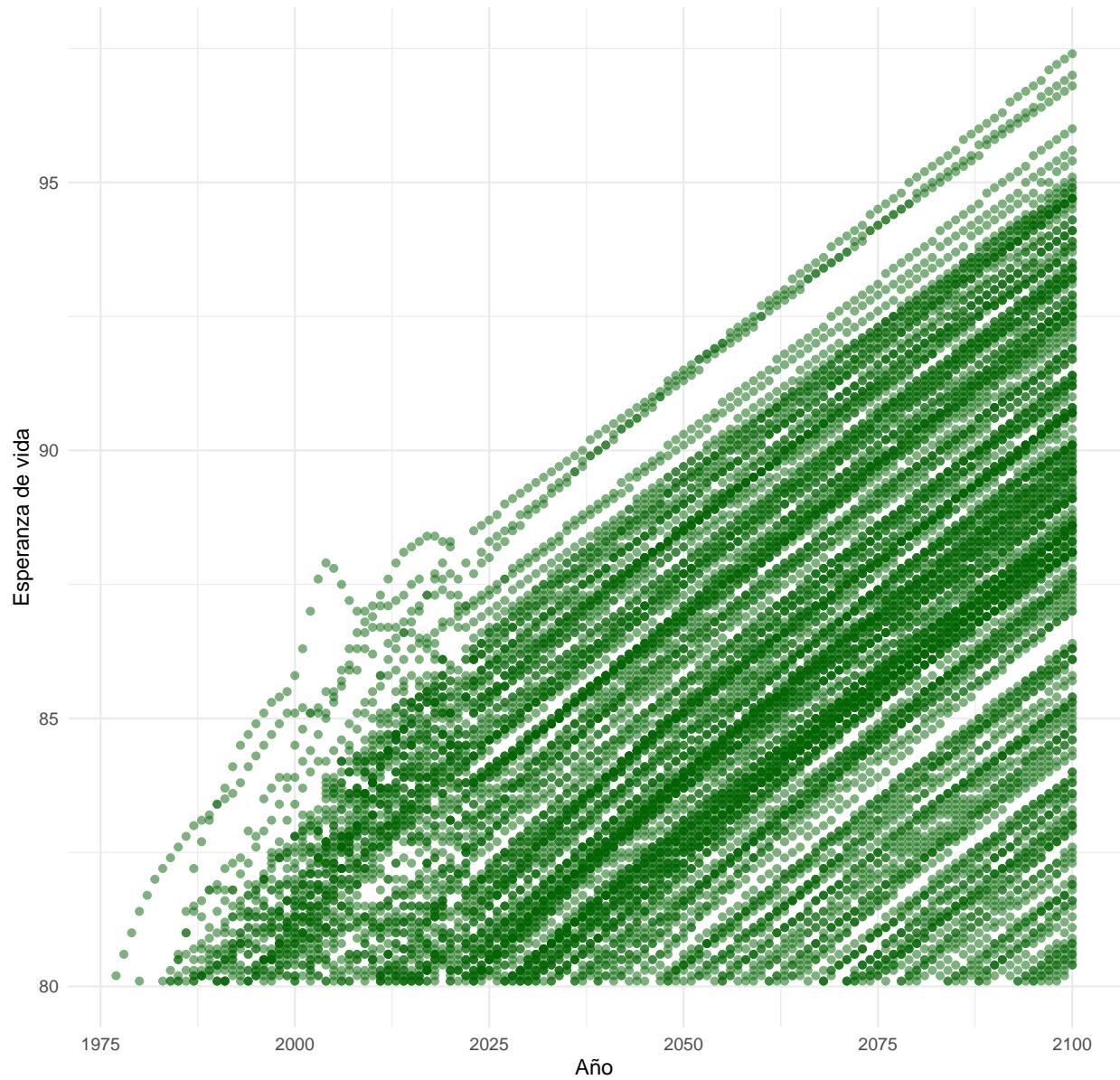


```
base_completa %>%
  filter(nacimientos_por_mujer > 5) %>%
  ggplot(aes(x = as.numeric(year), y = nacimientos_por_mujer)) +
  geom_point(alpha = 0.5, color = "red") +
  labs(title = "Países con alta fecundidad (> 5 hijos)", x = "Año", y = "Hijos por mujer") +
  theme_minimal()
```



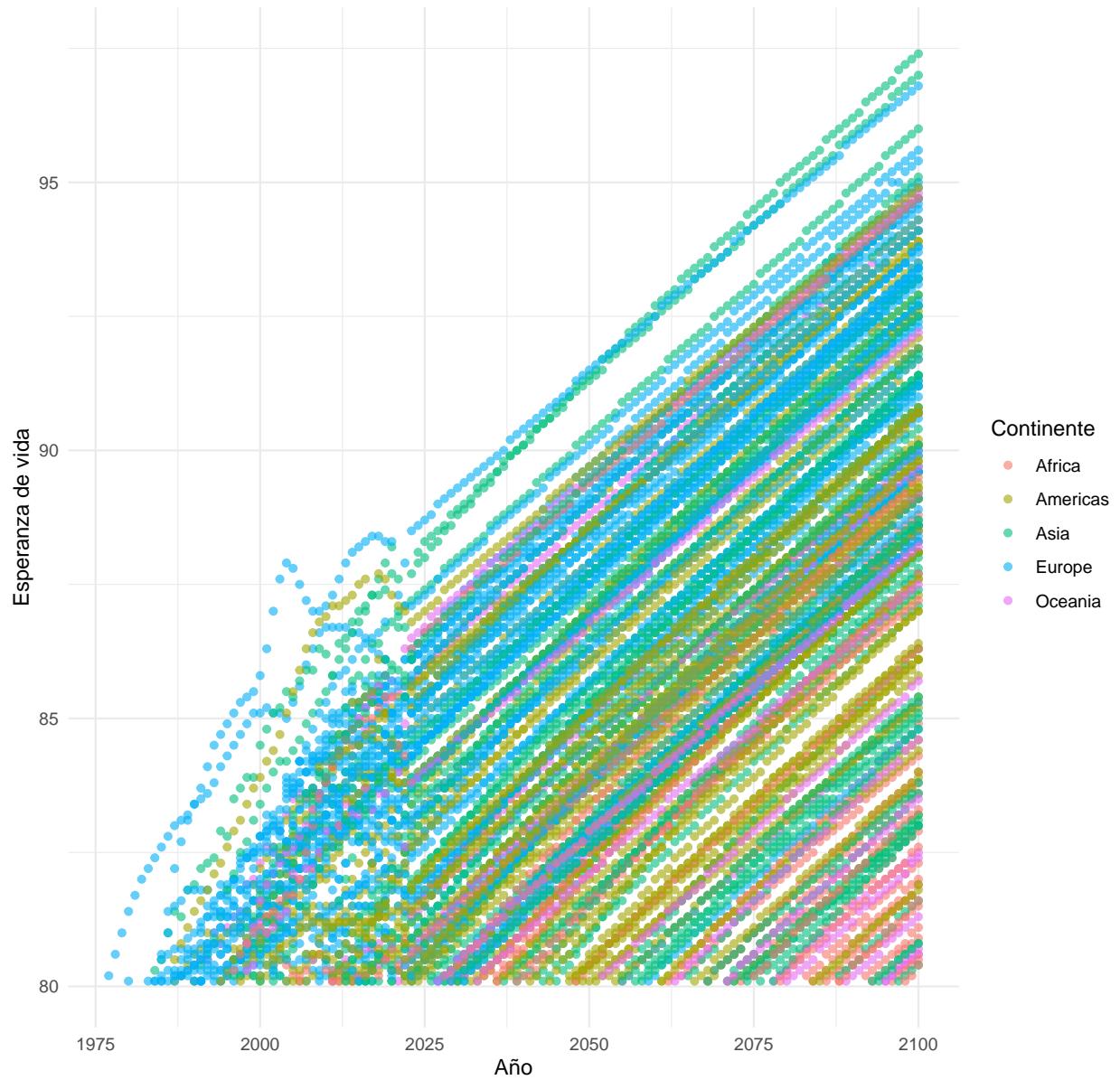
```
base_completa %>%
  filter(esperanza_de_vida > 80) %>%
  ggplot(aes(x = as.numeric(year), y = esperanza_de_vida)) +
  geom_point(alpha = 0.5, color = "darkgreen") +
  labs(title = "Países con alta esperanza de vida (> 80 años)", x = "Año", y = "Esperanza de vida")
  theme_minimal()
```

### Países con alta esperanza de vida (> 80 años)



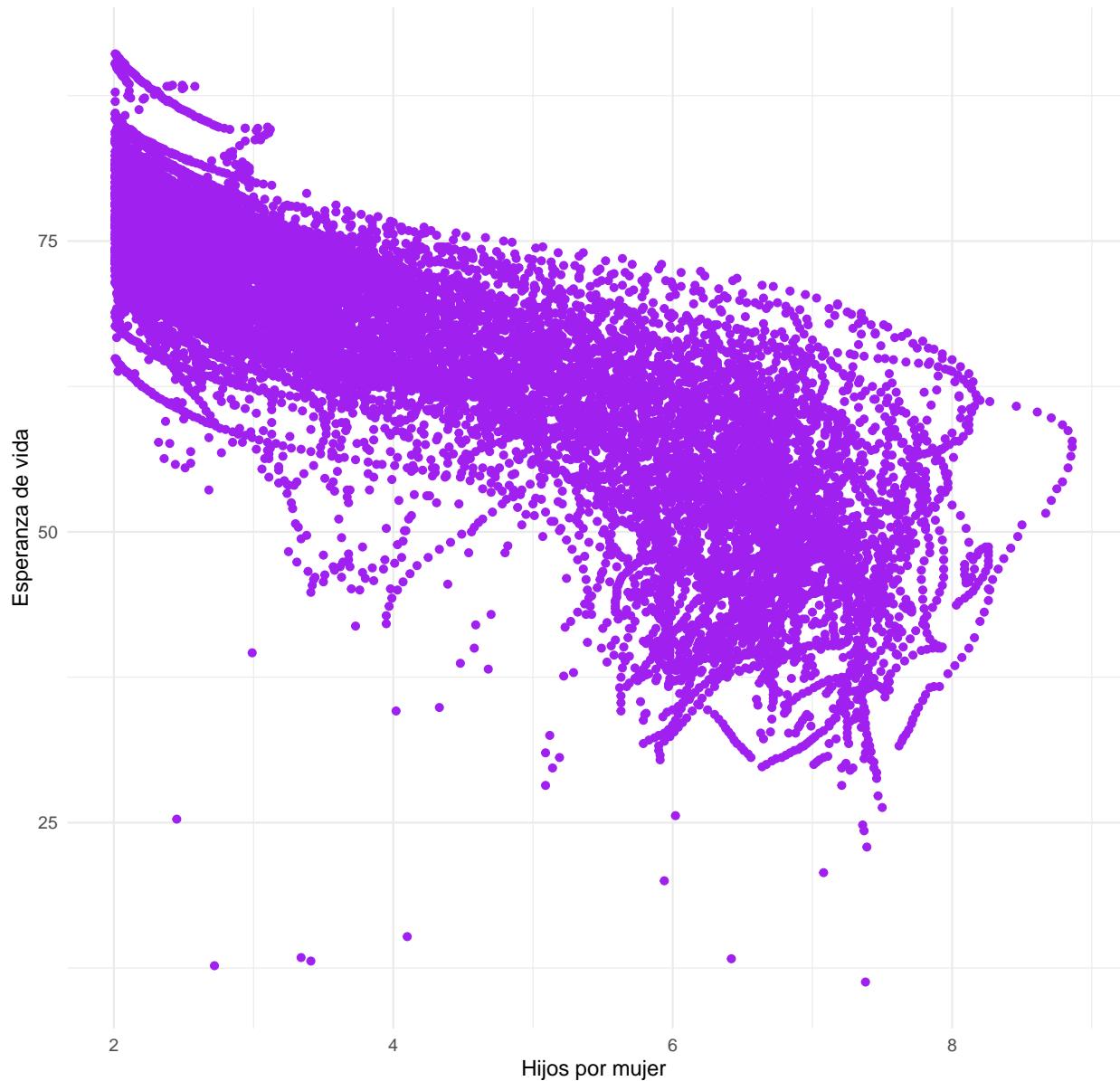
```
base_completa %>%
  filter(esperanza_de_vida > 80) %>%
  ggplot(aes(x = as.numeric(year), y = esperanza_de_vida, color = continente)) +
  geom_point(alpha = 0.6) +
  labs(title = "Países con alta esperanza de vida(>80) por continente",
       x = "Año",
       y = "Esperanza de vida",
       color = "Continente") +
  theme_minimal()
```

### Países con alta esperanza de vida(>80) por continente



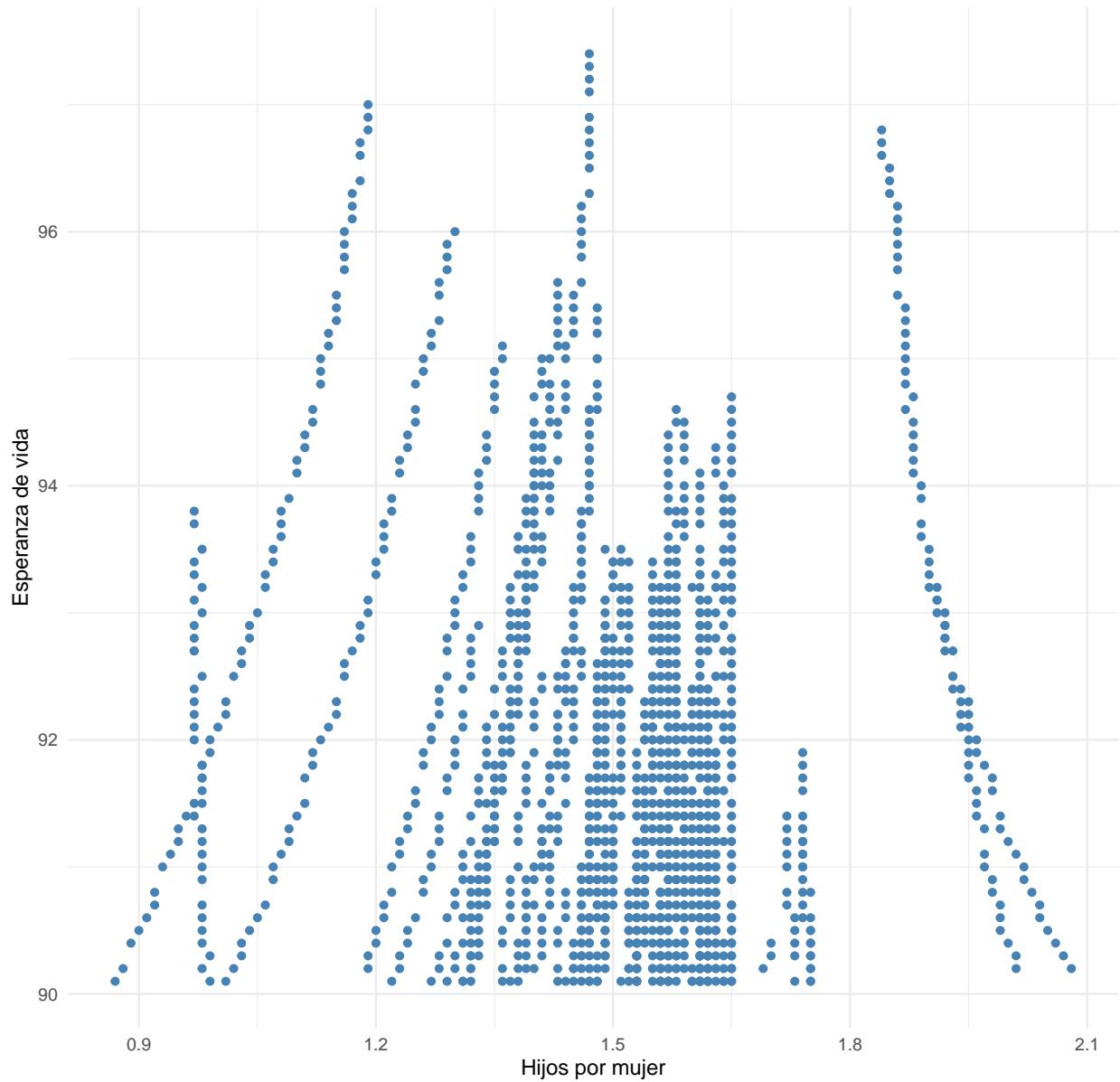
```
base_completa %>%
  filter(nacimientos_por_mujer > 2, esperanza_de_vida < 100) %>%
  ggplot(aes(x = nacimientos_por_mujer, y = esperanza_de_vida)) +
  geom_point(color = "purple") +
  labs(title = "Alta Fecundidad y baja esperanza de vida", x = "Hijos por mujer", y = "Esperanza de vida")
  theme_minimal()
```

### Alta Fecundidad y baja esperanza de vida



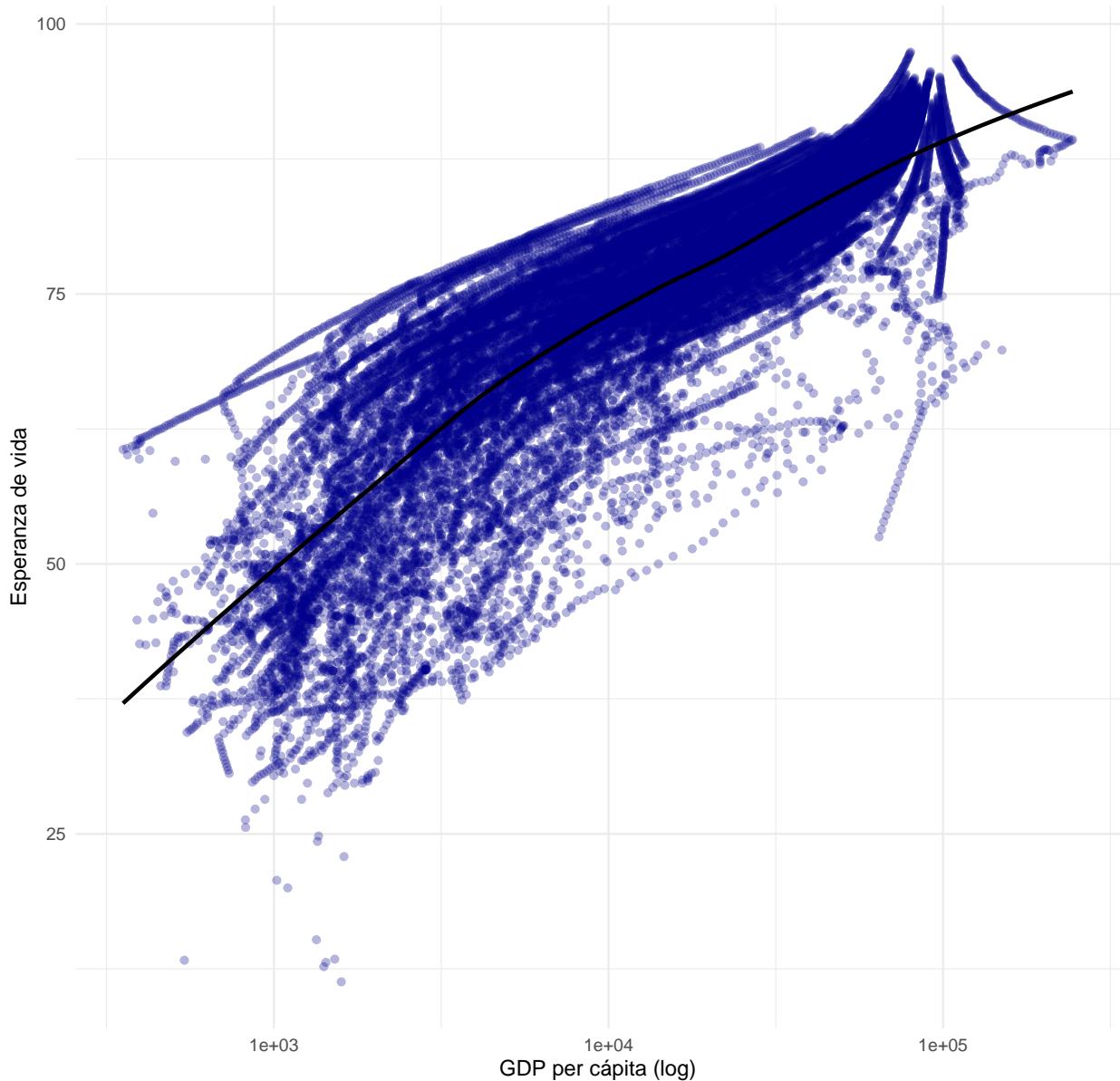
```
base_completa %>%
  filter(nacimientos_por_mujer < 3, esperanza_de_vida > 90) %>%
  ggplot(aes(x = nacimientos_por_mujer, y = esperanza_de_vida)) +
  geom_point(color = "steelblue") +
  labs(title = "Baja fecundidad y alta esperanza de vida", x = "Hijos por mujer", y = "Esperanza de vida") +
  theme_minimal()
```

### Baja fecundidad y alta esperanza de vida



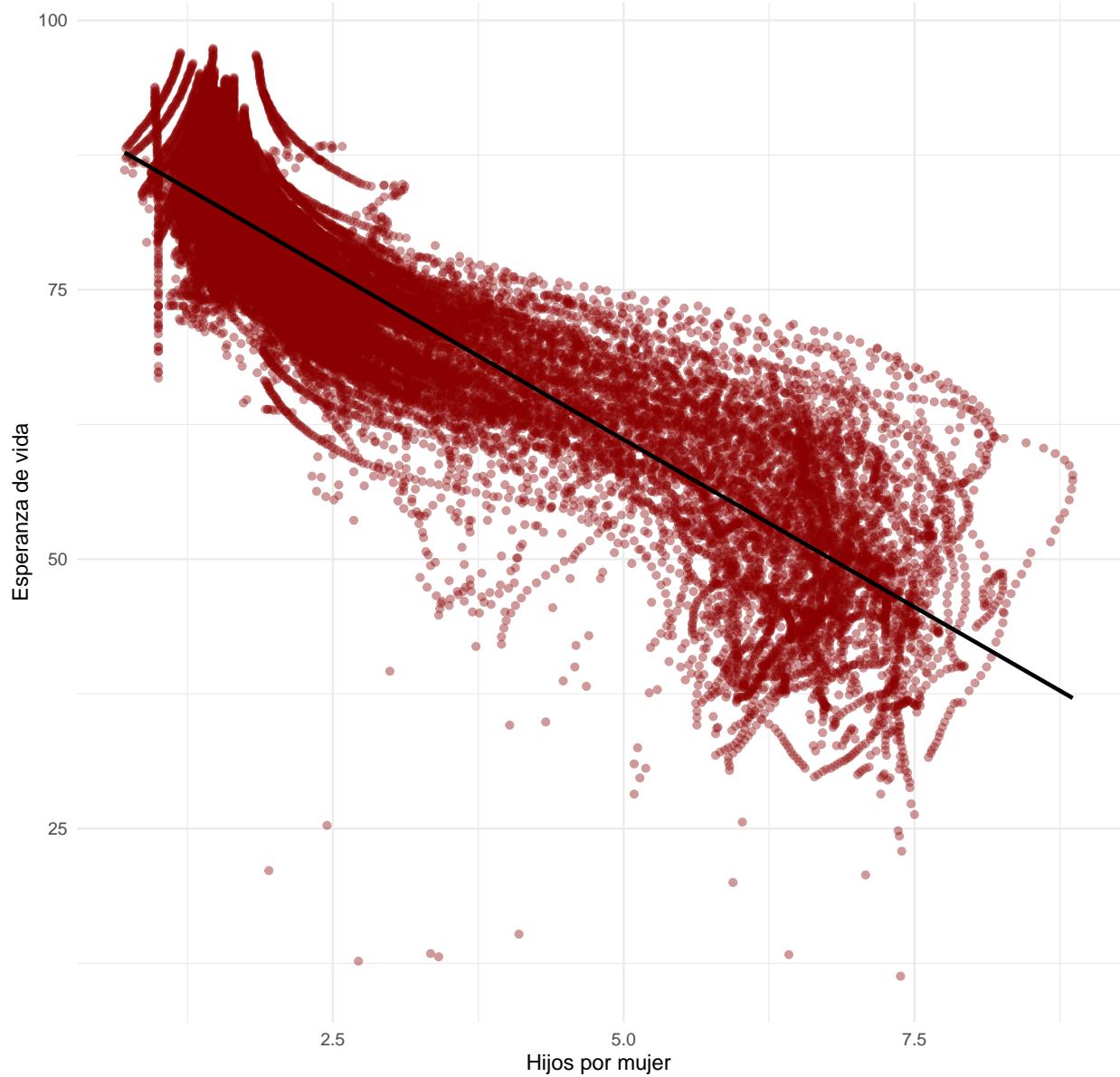
```
ggplot(base_completa, aes(x = producto_por_capita, y = esperanza_de_vida)) +  
  geom_point(alpha = 0.3, color = "darkblue") +  
  scale_x_log10() +  
  geom_smooth(method = "loess", se = FALSE, color = "black") +  
  labs(title = "Relación entre GDP per cápita y esperanza de vida", x = "GDP per cápita (log)",  
    theme_minimal())
```

### Relación entre GDP per cápita y esperanza de vida



```
ggplot(base_completa, aes(x = nacimientos_por_mujer, y = esperanza_de_vida)) +  
  geom_point(alpha = 0.4, color = "darkred") +  
  geom_smooth(method = "lm", se = FALSE, color = "black") +  
  labs(title = "Relación entre hijos por mujer y esperanza de vida", x = "Hijos por mujer", y = "Esperanza de vida")
```

### Relación entre hijos por mujer y esperanza de vida



```
ggplot(base_completa, aes(x = producto_por_capita, y = nacimientos_por_mujer)) +  
  geom_point(alpha = 0.4, color = "darkgreen") +  
  scale_x_log10() +  
  geom_smooth(method = "loess", se = FALSE, color = "black") +  
  labs(title = "Relación entre GDP y fecundidad", x = "GDP per cápita (log)", y = "Hijos por mu  
theme_minimal()
```

## Relación entre GDP y fecundidad



```
base_completa %>%
  filter(nacimientos_por_mujer > 5) %>%
  distinct(country) # te da solo los países únicos
```

```
## # A tibble: 138 x 1
##   country
##   <chr>
## 1 Afghanistan
## 2 Angola
## 3 Albania
## 4 UAE
## 5 Azerbaijan
```

```

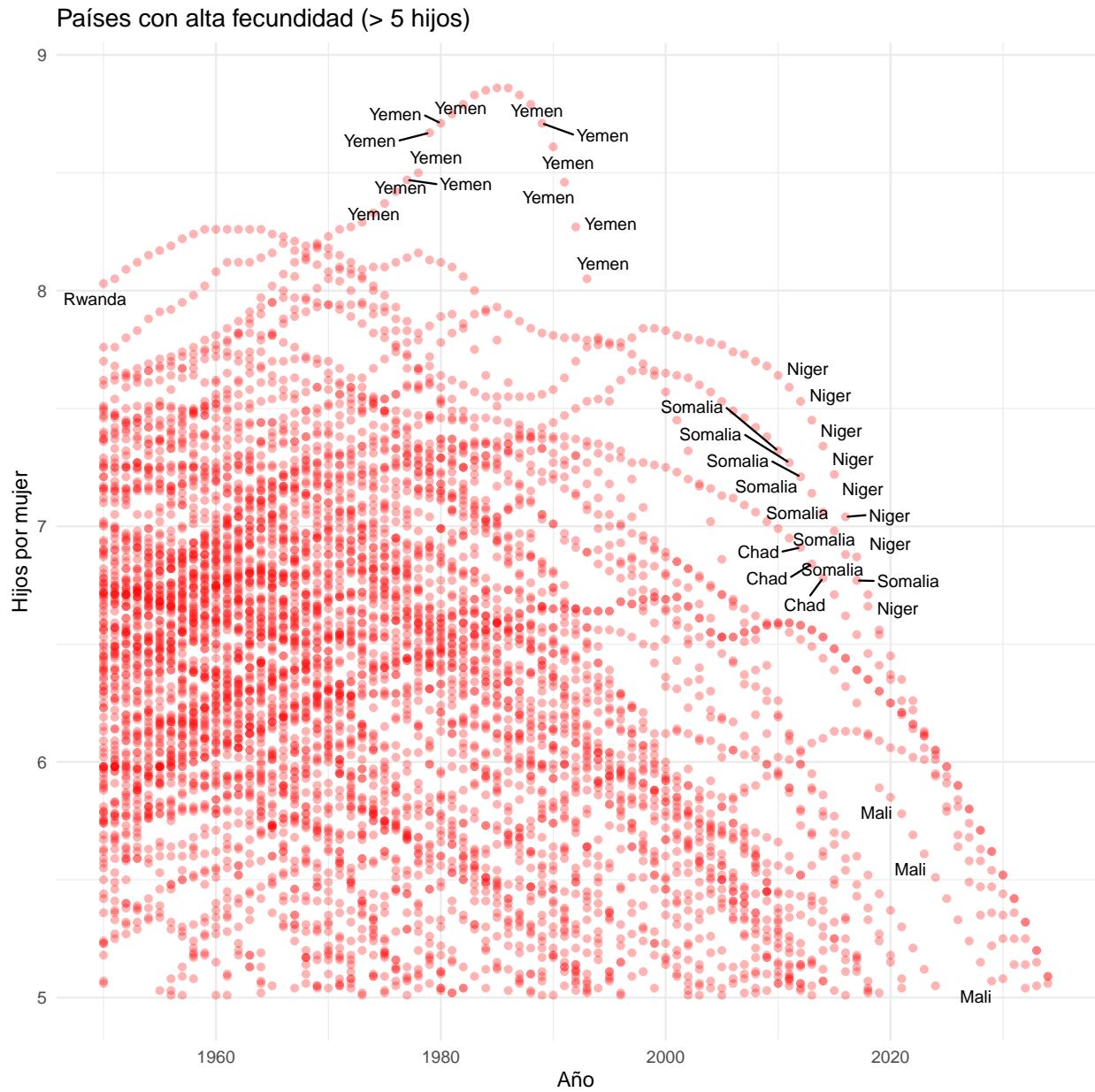
## 6 Burundi
## 7 Benin
## 8 Burkina Faso
## 9 Bangladesh
## 10 Bahrain
## # i 128 more rows

base_completa %>%
  filter(nacimientos_por_mujer > 5) %>%
  select(country, year, nacimientos_por_mujer) %>%
  arrange(desc(nacimientos_por_mujer)) # ordenado de mayor a menor

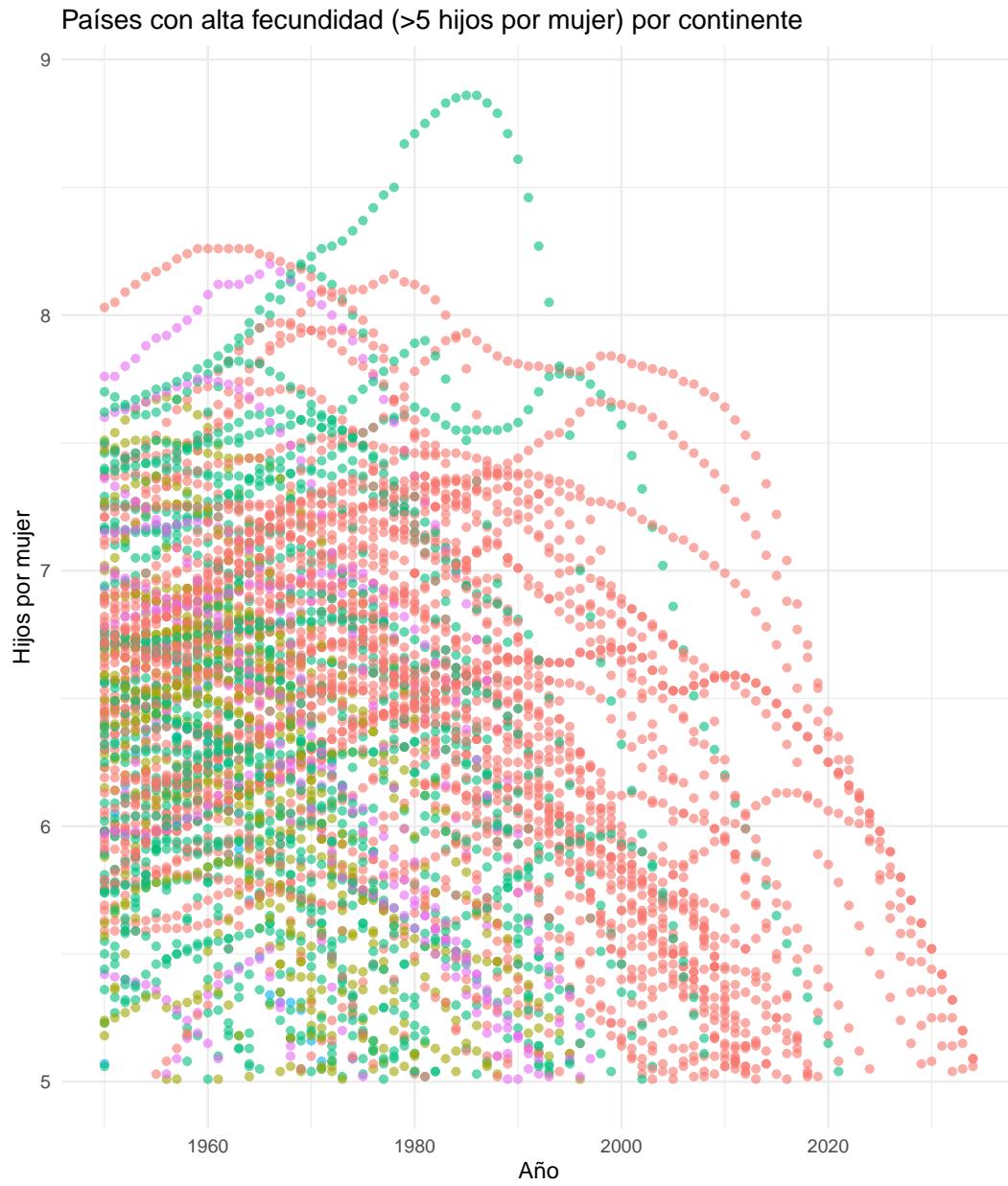
## # A tibble: 5,717 x 3
##   country   year nacimientos_por_mujer
##   <chr>     <dbl>                <dbl>
## 1 Yemen      1985                 8.86
## 2 Yemen      1986                 8.86
## 3 Yemen      1984                 8.85
## 4 Yemen      1983                 8.83
## 5 Yemen      1987                 8.83
## 6 Yemen      1982                 8.79
## 7 Yemen      1988                 8.79
## 8 Yemen      1981                 8.75
## 9 Yemen      1980                 8.71
## 10 Yemen     1989                 8.71
## # i 5,707 more rows

base_completa %>%
  filter(nacimientos_por_mujer > 5) %>%
  ggplot(aes(x = as.numeric(year), y = nacimientos_por_mujer, label = country)) +
  geom_point(alpha = 0.3, color = "red") +
  geom_text_repel(size = 3, max.overlaps = 10) +
  labs(title = "Países con alta fecundidad (> 5 hijos)", x = "Año", y = "Hijos por mujer") +
  theme_minimal()

```



```
base_completa %>%
  filter(nacimientos_por_mujer > 5) %>%
  ggplot(aes(x = as.numeric(year), y = nacimientos_por_mujer, color = continente)) +
  geom_point(alpha = 0.6) +
  labs(title = "Países con alta fecundidad (>5 hijos por mujer) por continente",
       x = "Año",
       y = "Hijos por mujer",
       color = "Continente") +
  theme_minimal()
```



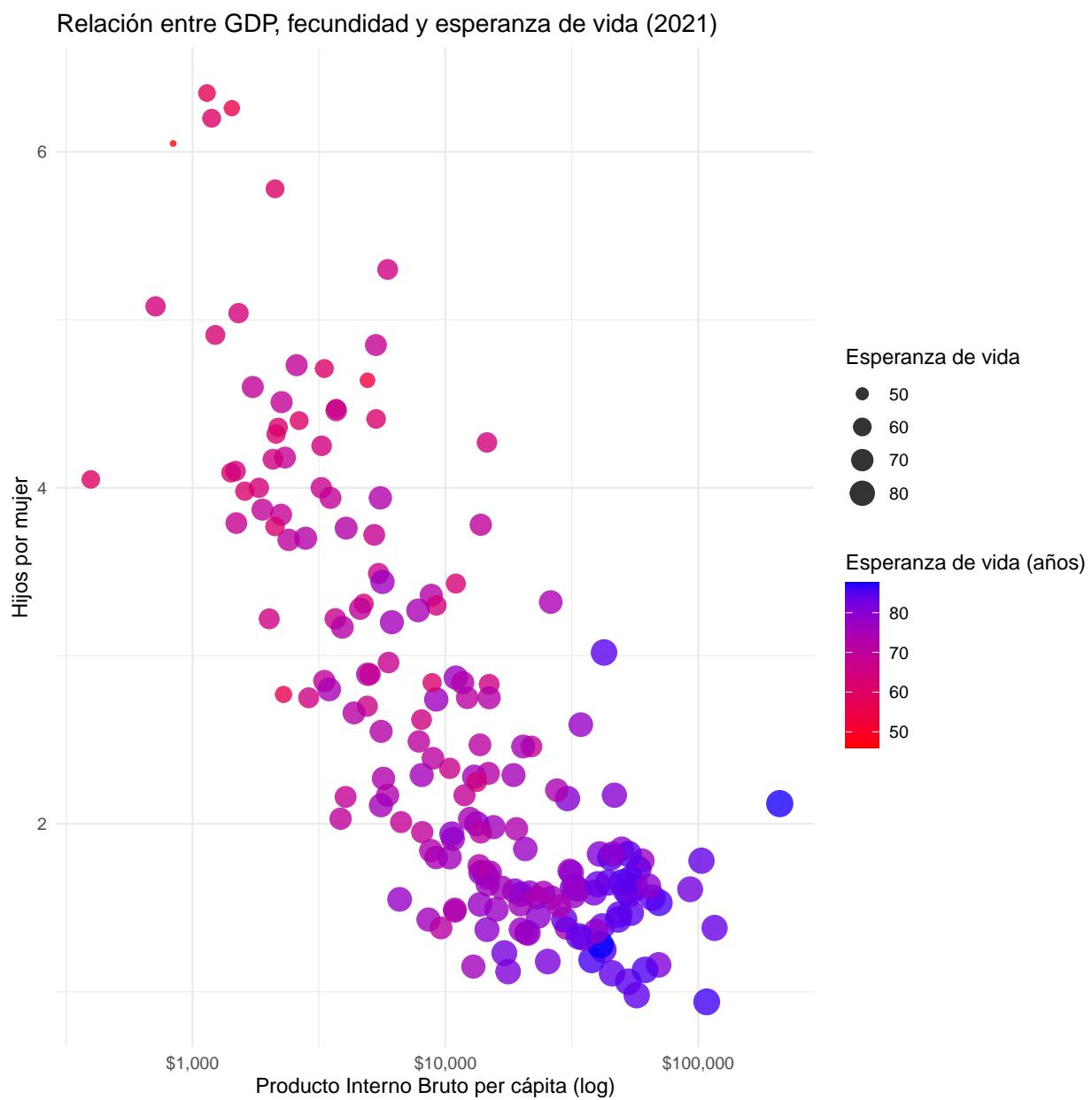
```
# Filtrar para un año específico (por ejemplo 2021)
datos_filtrados <- base_completa %>%
  filter(year == 2021,
        !is.na(producto_por_capita),
        !is.na(nacimientos_por_mujer),
        !is.na(esperanza_de_vida))

# Gráfico
ggplot(datos_filtrados, aes(x = producto_por_capita,
                               y = nacimientos_por_mujer,
                               size = esperanza_de_vida,
                               color = esperanza_de_vida)) +
```

```

geom_point(alpha = 0.8) +
scale_x_log10(labels = dollar_format(prefix = "$")) +
scale_color_gradient(low = "red", high = "blue") +
labs(title = "Relación entre GDP, fecundidad y esperanza de vida (2021)",
x = "Producto Interno Bruto per cápita (log)",
y = "Hijos por mujer",
color = "Esperanza de vida (años)",
size = "Esperanza de vida") +
theme_minimal()

```



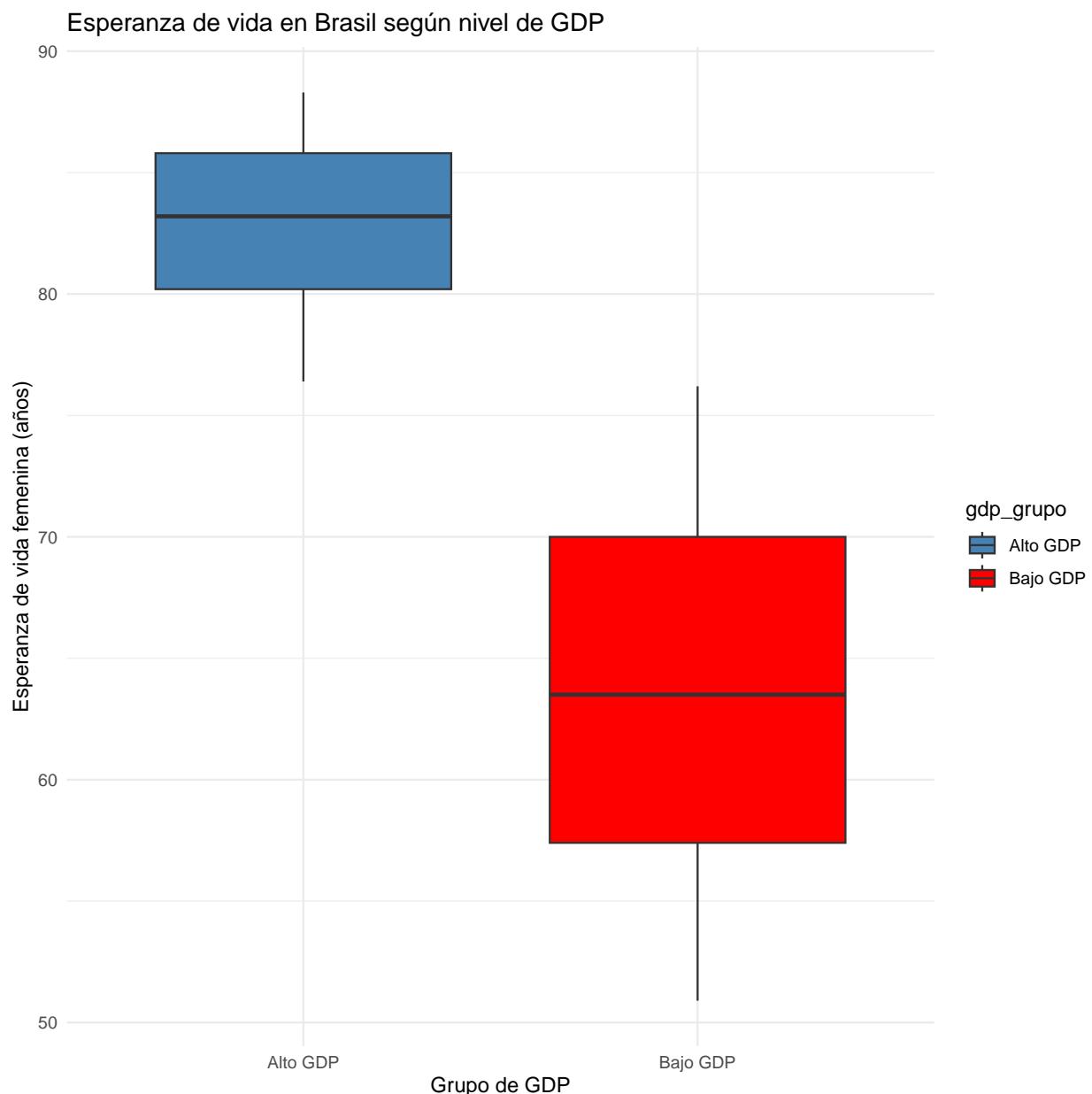
# Probando para visualización

```

brasil_data <- base_completa %>%
  filter(country == "Brazil") %>%
  mutate(gdp_grupo = ifelse(producto_por_capita > 13845, "Alto GDP", "Bajo GDP"))

# Crear el boxplot solo para Brasil
ggplot(brasil_data, aes(x = gdp_grupo, y = esperanza_de_vida, fill = gdp_grupo)) +
  geom_boxplot() +
  labs(title = "Esperanza de vida en Brasil según nivel de GDP",
       x = "Grupo de GDP",
       y = "Esperanza de vida femenina (años)") +
  scale_fill_manual(values = c("Bajo GDP" = "red", "Alto GDP" = "steelblue")) +
  theme_minimal()

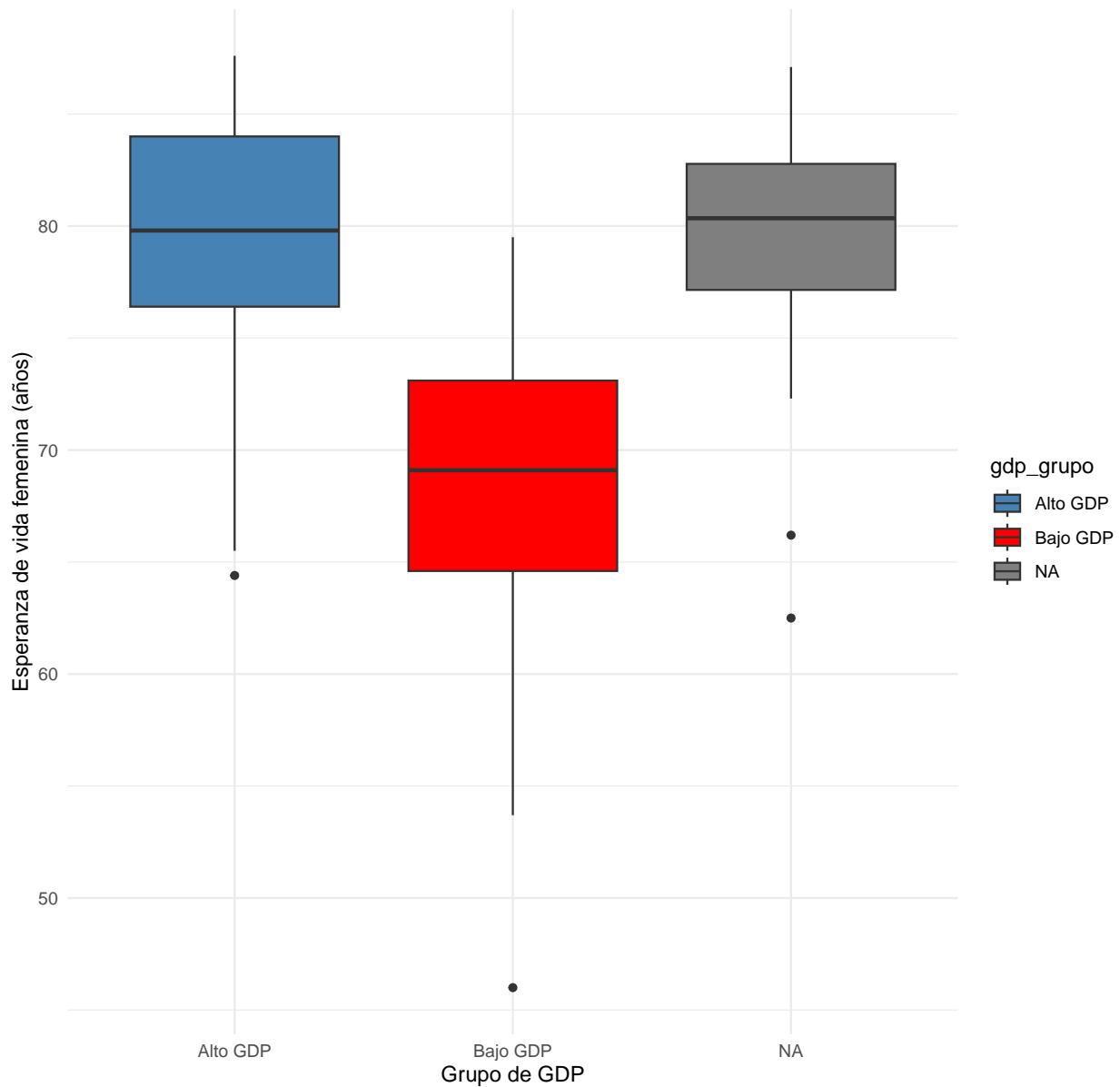
```



```
# Filtrar datos solo del año 2021
datos_2021 <- base_completa %>%
  filter(year == 2021) %>%
  mutate(gdp_grupo = ifelse(producto_por_capita > 13845, "Alto GDP", "Bajo GDP"))

# Crear el boxplot solo para 2021
ggplot(datos_2021, aes(x = gdp_grupo, y = esperanza_de_vida, fill = gdp_grupo)) +
  geom_boxplot() +
  labs(title = "Esperanza de vida femenina en 2021 según nivel de GDP",
       x = "Grupo de GDP",
       y = "Esperanza de vida femenina (años)") +
  scale_fill_manual(values = c("Bajo GDP" = "red", "Alto GDP" = "steelblue")) +
  theme_minimal()
```

Esperanza de vida femenina en 2021 según nivel de GDP



```
# Filtrar solo datos de Brasil en el año 2021
brasil_2021 <- base_completa %>%
  filter(year == 2021, country == "Brazil") %>%
  mutate(gdp_grupo = ifelse(producto_por_capita > 13845, "Alto GDP", "Bajo GDP"))
```

```
# Ver tabla si solo quieres ver el dato
print(brasil_2021)
```

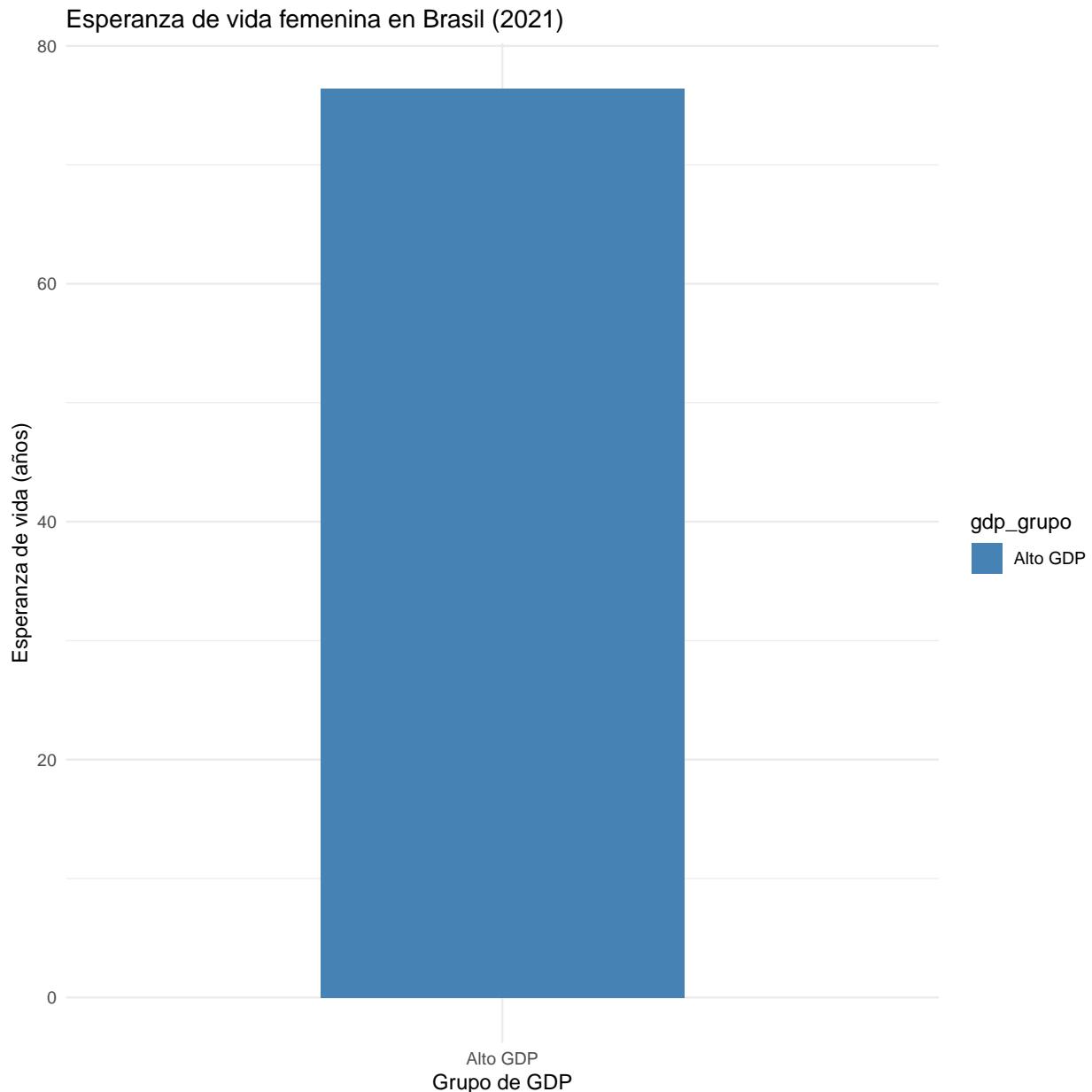
```
## # A tibble: 1 x 8
##   country year esperanza_de_vida producto_por_capita nacimientos_por_mu
##   <chr>    <dbl>            <dbl>                <dbl>
## 1 Brazil    2021             76.4                 14700
##   jor         <dbl>
```

```

## # i 3 more variables: continente <chr>, gdp_bajo <chr>, gdp_grupo <chr>

# D crear un gráfico de punto único
ggplot(brasil_2021, aes(x = gdp_grupo, y = esperanza_de_vida, fill = gdp_grupo)) +
  geom_col(width = 0.5) +
  labs(title = "Esperanza de vida femenina en Brasil (2021)",
       x = "Grupo de GDP",
       y = "Esperanza de vida (años)") +
  scale_fill_manual(values = c("Bajo GDP" = "red", "Alto GDP" = "steelblue")) +
  theme_minimal()

```



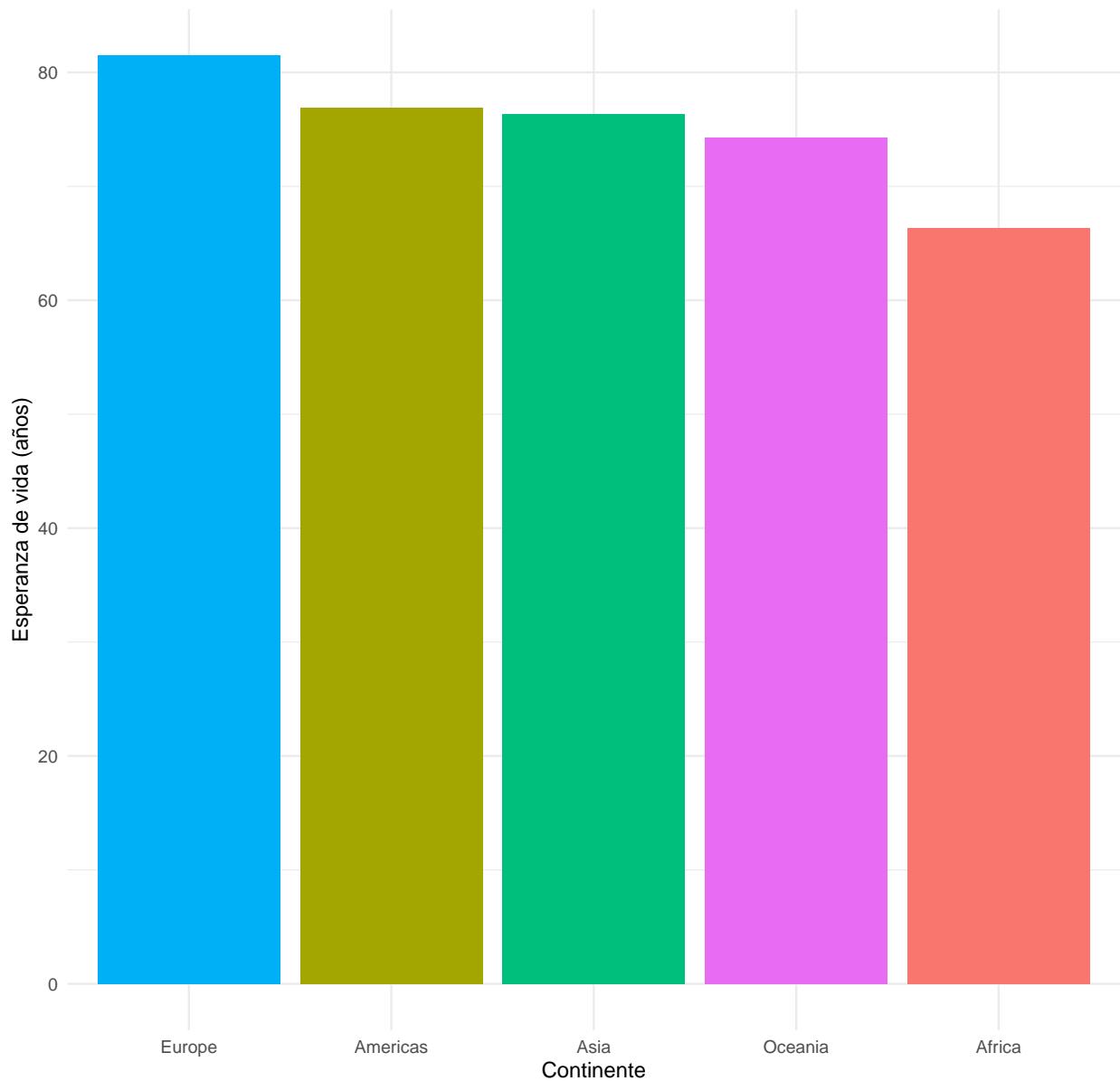
```

continentes_2021 <- base_completa %>%
  filter(year == 2021, !is.na(esperanza_de_vida), !is.na(continente)) %>%
  group_by(continente) %>%
  summarise(esperanza_media = mean(esperanza_de_vida, na.rm = TRUE)) %>%
  arrange(desc(esperanza_media))

# Gráfico
ggplot(continentes_2021, aes(x = reorder(continente, -esperanza_media), y = esperanza_media, f
  geom_col() +
  labs(title = "Esperanza de vida femenina promedio por continente (2021)",
       x = "Continente",
       y = "Esperanza de vida (años)") +
  theme_minimal() +
  theme(legend.position = "none")

```

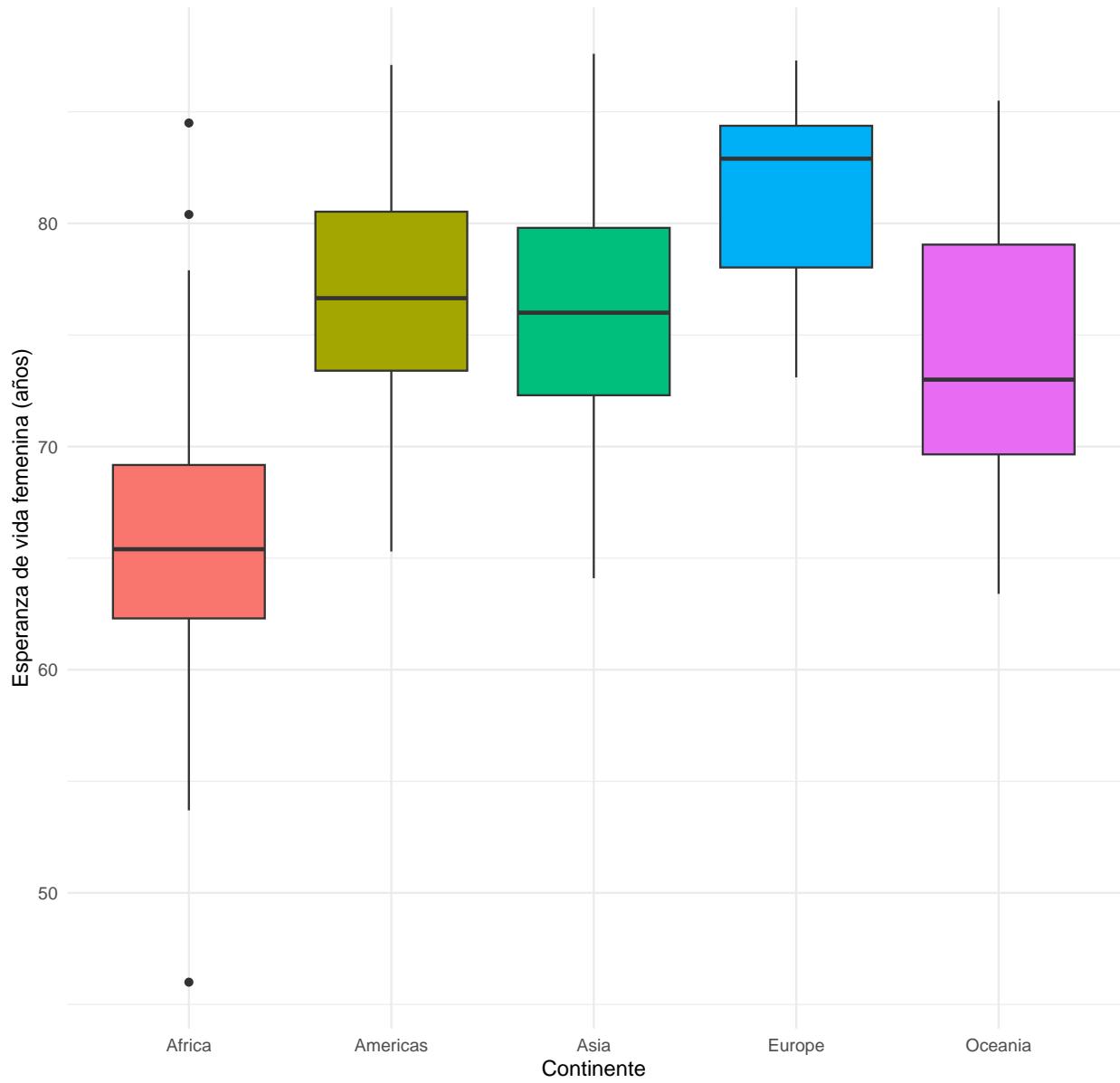
Esperanza de vida femenina promedio por continente (2021)



```
# Filtrar solo datos del año 2021 y no NA
base_2021 <- base_completa %>%
  filter(year == 2021, !is.na(esperanza_de_vida), !is.na(continente))

# Crear boxplot
ggplot(base_2021, aes(x = continente, y = esperanza_de_vida, fill = continente)) +
  geom_boxplot() +
  labs(title = "Distribución de la esperanza de vida femenina por continente (2021)",
       x = "Continente",
       y = "Esperanza de vida femenina (años)") +
  theme_minimal() +
  theme(legend.position = "none")
```

Distribución de la esperanza de vida femenina por continente (2021)



```
# Filtrar solo datos del año 2021
base_2021 <- base_completa %>%
  filter(year == 2021) %>%
  drop_na(esperanza_de_vida, producto_por_capita, continente)

# Crear variable de grupo de GDP
base_2021 <- base_2021 %>%
  mutate(gdp_grupo = ifelse(producto_por_capita > 13845, "Alto GDP", "Bajo GDP"))

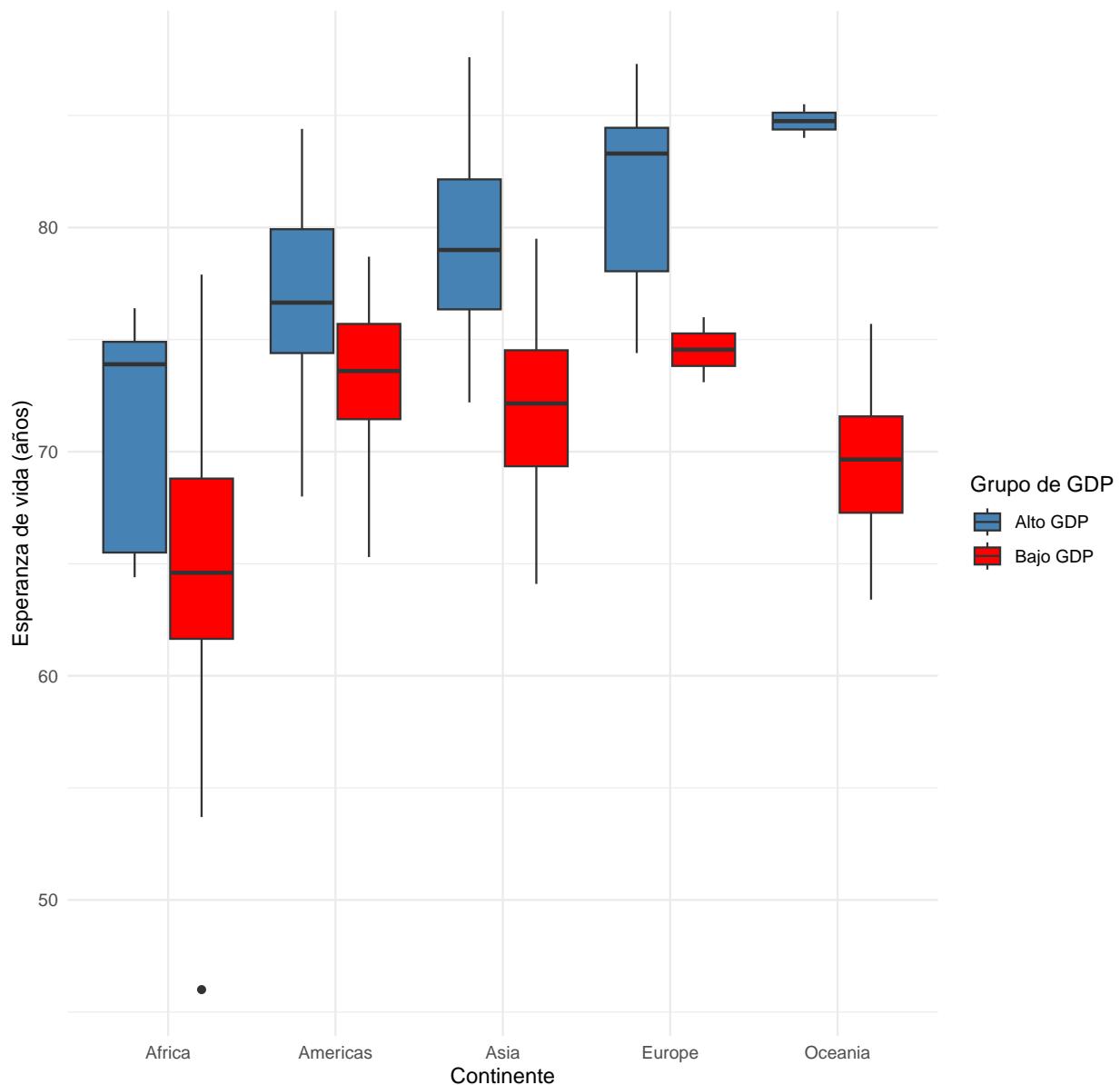
# Boxplot
ggplot(base_2021, aes(x = continente, y = esperanza_de_vida, fill = gdp_grupo)) +
  geom_boxplot(position = position_dodge(0.8)) +
```

```

  labs(
    title = "Esperanza de vida femenina por continente y nivel de GDP (2021)",
    x = "Continente",
    y = "Esperanza de vida (años)",
    fill = "Grupo de GDP"
  ) +
  scale_fill_manual(values = c("Bajo GDP" = "red", "Alto GDP" = "steelblue")) +
  theme_minimal()

```

Esperanza de vida femenina por continente y nivel de GDP (2021)



```

# Filtrar datos del año 2021 y eliminar valores faltantes
base_2021 <- base_completa %>%

```

```

filter(year == 2021) %>%
drop_na(esperanza_de_vida, nacimientos_por_mujer)

# Escalar ambas variables (para compararlas en el mismo gráfico)
base_barras <- base_2021 %>%
  select(country, esperanza_de_vida, nacimientos_por_mujer) %>%
  mutate(
    esperanza_escalada = scale(esperanza_de_vida)[,1],
    fecundidad_escalada = scale(nacimientos_por_mujer)[,1]
  ) %>%
  pivot_longer(cols = c(esperanza_escalada, fecundidad_escalada),
               names_to = "variable", values_to = "valor")

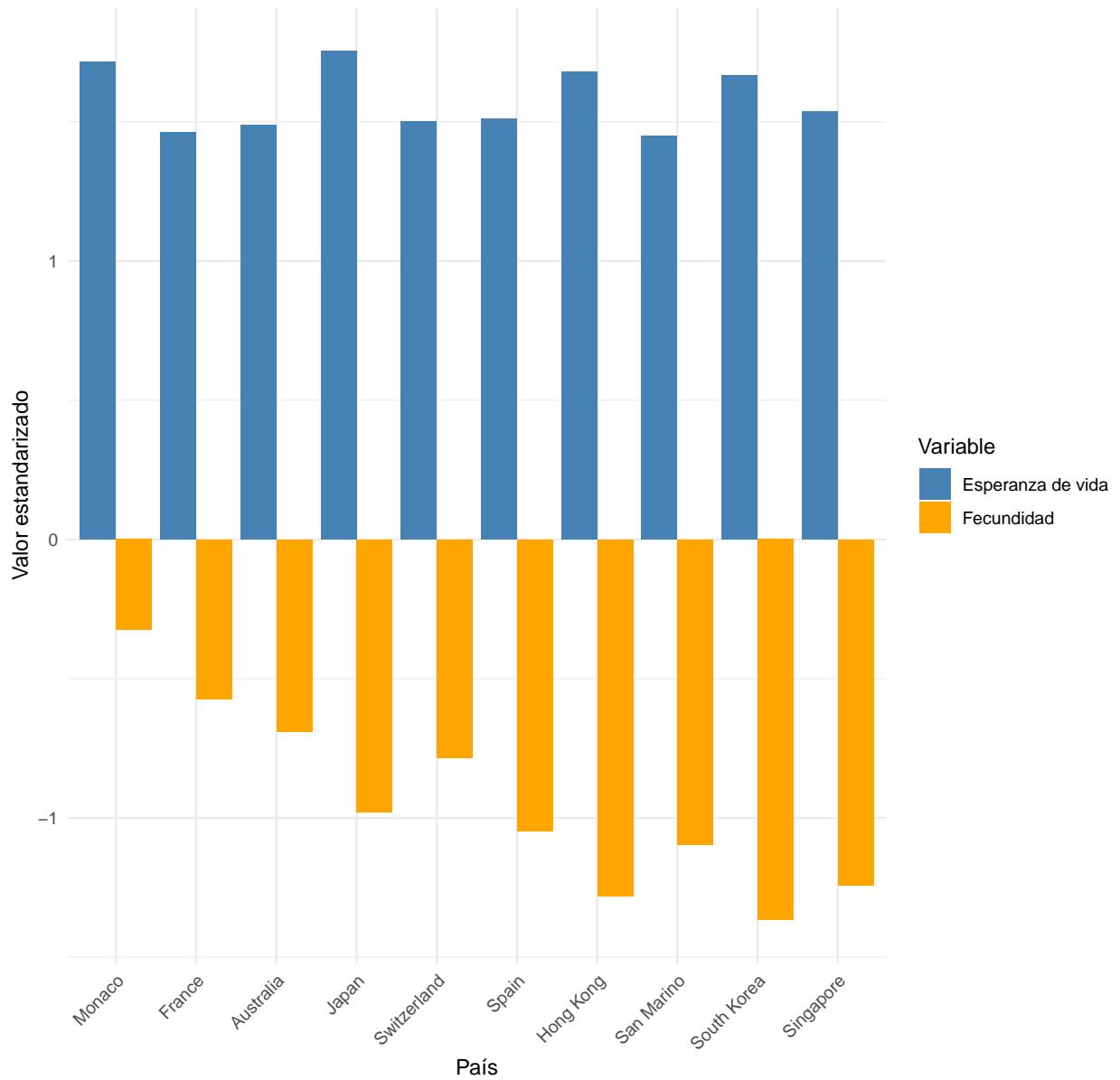
# Seleccionar los países con esperanza de vida más alta (o puedes cambiar a los más bajos)
top_paises <- base_2021 %>%
  arrange(desc(esperanza_de_vida)) %>%
  slice(1:10) %>% pull(country)

# Filtrar solo los países seleccionados
base_barras <- base_barras %>%
  filter(country %in% top_paises)

# Gráfico de barras
ggplot(base_barras, aes(x = reorder(country, -valor), y = valor, fill = variable)) +
  geom_col(position = "dodge") +
  labs(
    title = "Comparación estandarizada entre esperanza de vida y fecundidad (Top 10 países)",
    x = "País",
    y = "Valor estandarizado",
    fill = "Variable"
  ) +
  scale_fill_manual(
    values = c("esperanza_escalada" = "steelblue", "fecundidad_escalada" = "orange"),
    labels = c("Esperanza de vida", "Fecundidad")
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

Comparación estandarizada entre esperanza de vida y fecundidad (Top 10 países)



```
# Filtrar y limpiar datos del año 2021
base_2021 <- base_completa %>%
  filter(year == 2021) %>%
  drop_na(esperanza_de_vida, nacimientos_por_muher, continente)

# Agrupar por continente y calcular promedio
base_continente <- base_2021 %>%
  group_by(continente) %>%
  summarise(
    esperanza_media = mean(esperanza_de_vida),
    fecundidad_media = mean(nacimientos_por_muher)
  ) %>%
```

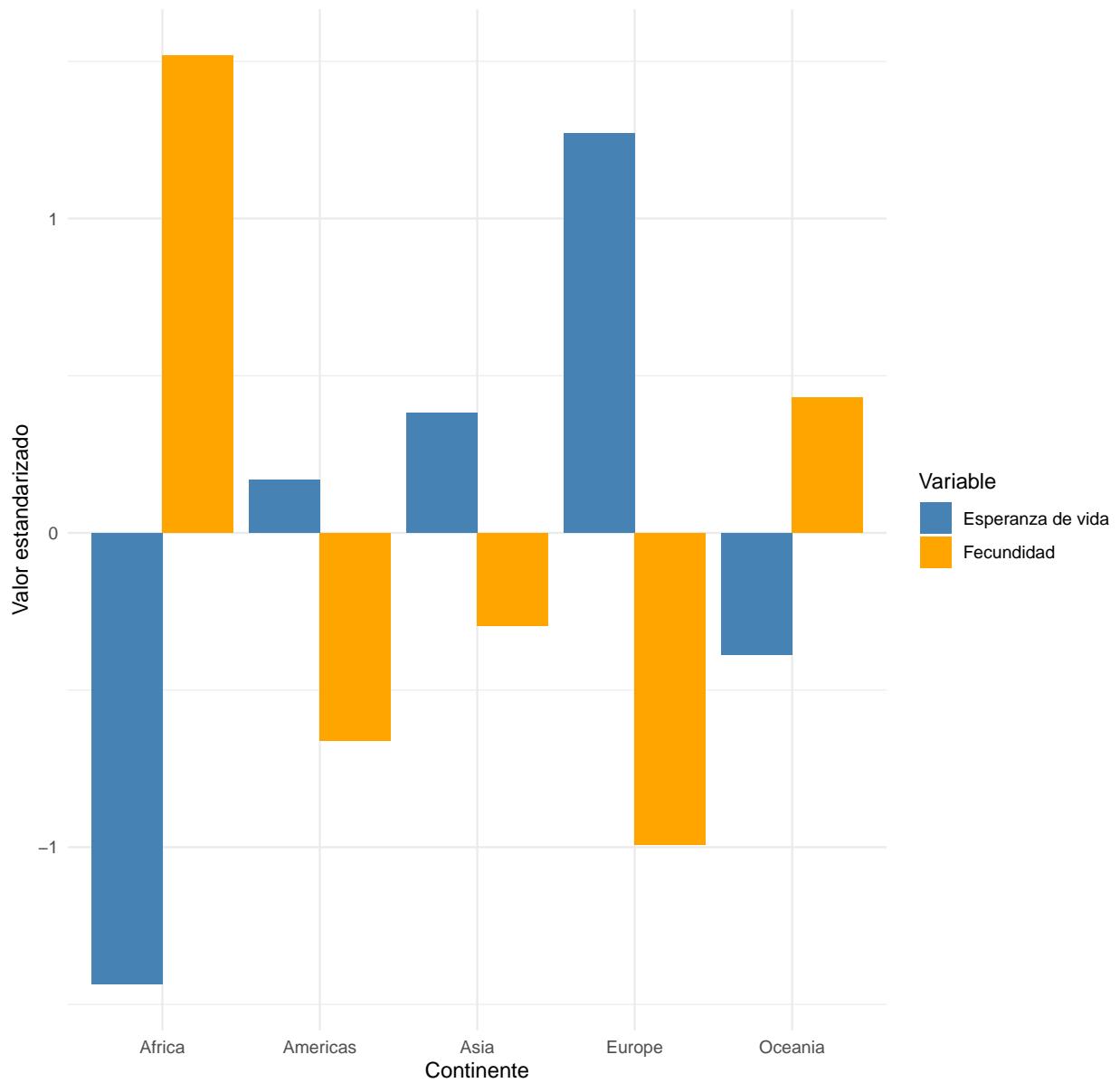
```

# Escalar ambas variables
mutate(
  esperanza_escalada = scale(esperanza_media)[,1],
  fecundidad_escalada = scale(fecundidad_media)[,1]
) %>%
pivot_longer(cols = c(esperanza_escalada, fecundidad_escalada),
              names_to = "variable", values_to = "valor")

# Gráfico de barras por continente
ggplot(base_continente, aes(x = continente, y = valor, fill = variable)) +
  geom_col(position = "dodge") +
  labs(
    title = "Comparación estandarizada de esperanza de vida y fecundidad por continente (2021)",
    x = "Continente",
    y = "Valor estandarizado",
    fill = "Variable"
  ) +
  scale_fill_manual(
    values = c("esperanza_escalada" = "steelblue", "fecundidad_escalada" = "orange"),
    labels = c("Esperanza de vida", "Fecundidad")
  ) +
  theme_minimal()

```

Comparación estandarizada de esperanza de vida y fecundidad por continente (2021)



```
write.csv(base_completa, "Datos/Base_Datos_depurada/base_completa.csv", row.names = FALSE)
```