

Informe del Proyecto

Rosalia

2025-08-08

```
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.4.3
## Warning: package 'ggplot2' was built under R version 4.4.3
## Warning: package 'tibble' was built under R version 4.4.3
## Warning: package 'tidyr' was built under R version 4.4.3
## Warning: package 'readr' was built under R version 4.4.3
## Warning: package 'purrr' was built under R version 4.4.3
## Warning: package 'dplyr' was built under R version 4.4.3
## Warning: package 'stringr' was built under R version 4.4.3
## Warning: package 'forcats' was built under R version 4.4.3
## Warning: package 'lubridate' was built under R version 4.4.3

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr    1.3.1
## v purrr    1.0.4
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readxl)
library(readr)
library(dplyr)
library(ggplot2)
library(ggrepel)

## Warning: package 'ggrepel' was built under R version 4.4.3
```

```

library(countrycode)

## Warning: package 'countrycode' was built under R version 4.4.3

library(scales)

## Warning: package 'scales' was built under R version 4.4.3

##
## Adjuntando el paquete: 'scales'
##
## The following object is masked from 'package:purrr':
##       discard
##
## The following object is masked from 'package:readr':
##       col_factor

library(maps)

## Warning: package 'maps' was built under R version 4.4.3

##
## Adjuntando el paquete: 'maps'
##
## The following object is masked from 'package:purrr':
##       map

library(DT)

## Warning: package 'DT' was built under R version 4.4.3

library(lorem)

## Warning: package 'lorem' was built under R version 4.4.3

library(png)
library(grid)
library(tinytex)

## Warning: package 'tinytex' was built under R version 4.4.3

```

Introducción Introducción

Esta propuesta de diseñar una investigación en ciencias de datos me brinda la libertad de explorar con mayor profundidad temas sociales que me atraviesan personalmente. Al revisar las bases de datos disponibles en Gapminder, encontré una gran variedad de información relacionada con las mujeres, lo que despertó en mí

una inquietud sobre la posible relación entre la longevidad femenina y la maternidad. En Costa Rica, existe una expresión popular según la cual “tener hijos te riega las bilis”, lo que en el imaginario colectivo podría asociarse con la idea de que la maternidad acorta la vida. Esta creencia me llevó a preguntarme si realmente existe una relación significativa entre la esperanza de vida de las mujeres y la cantidad promedio de hijos o hijas que tienen. En lo personal, mi familia ilustra esta interrogante de forma particular. Mi madre tuvo una sola hermana, quien no tuvo hijos y falleció a una edad relativamente baja para los estándares actuales. Por otro lado, espero que mi madre —quien sí tuvo hijos— viva muchos más años que el promedio nacional. Aunque anecdótica, esta observación familiar me motivó a profundizar en la investigación científica sobre el tema. En el proceso, encontré tres estudios que llamaron mi atención. Vaupel (2001) plantea que podría existir una relación positiva entre el número de hijos y la longevidad en las mujeres, especialmente cuando el último parto ocurre a una edad avanzada. Por su parte, Zhang et al. (2023) afirman que las mujeres que tienen entre tres y cinco hijos tienden a vivir más tiempo, mientras que aquellas con uno o ningún hijo presentan una esperanza de vida más corta. Me resulta interesante cómo estas investigaciones se reflejan, de alguna forma, en la experiencia de mi propia familia. A mis 35 años, nunca había considerado que la decisión de tener hijos pudiera estar vinculada con la longevidad. Aunque esta no ha sido una motivación personal, estos hallazgos abren la posibilidad de considerar la maternidad desde una perspectiva distinta. Por otro lado, Hampton (2021) analiza la relación entre desarrollo y fecundidad, señalando que, aunque existe la idea de que a mayor desarrollo económico se podría esperar un aumento de la fecundidad, en la práctica no se observa una asociación directa. Su estudio concluye que, al menos en la última década, los niveles elevados de desarrollo humano ya no están vinculados con repuntes en la tasa de natalidad.

Objetivos

Ante estas inquietudes el Objetivo General es:

Explorar la posible relación entre la esperanza de vida femenina, la fecundidad y el nivel de desarrollo económico en países del mundo, a partir de datos históricos entre 1950 y 2021 disponibles en Gapminder.

En cuanto a los objetivos específicos son:

1. Analizar patrones y tendencias en la relación entre esperanza de vida femenina y número promedio de hijos por mujer, identificando posibles asociaciones según región geográfica y a lo largo del tiempo.
2. Explorar la relación entre el producto interno bruto per cápita (ajustado por poder adquisitivo) y los niveles de fecundidad y longevidad femenina, para determinar si existen diferencias significativas entre países con distinto nivel de desarrollo económico.

Como se menciona en los objetivos, he decidido trabajar con tres conjuntos de datos disponibles en Gapminder para desarrollar un análisis exploratorio de datos:

- Life_expectancy_female: esperanza de vida al nacer de mujeres (en años).
- Children_per_women_total_fertility: número promedio de hijos por mujer a lo largo de su vida.
- Gross domestic product per person: producto interno bruto per cápita ajustado por paridad de poder adquisitivo al año 2021.

Estos conjuntos de datos cubren períodos desde 1800 hasta 210 vbg 0, por lo que el análisis se centrará en los datos comunes disponibles a partir de 1950.

Datos Utilizados

```
Datos_FEGDP <- readr::read_csv("../Datos/Base_Datos_depurada/base_completa.csv", locale = locale(encoding = "UTF-8"))
```

```
## Rows: 35636 Columns: 7
## -- Column specification -----
## Delimiter: ","
## chr (3): country, continente, gdp_grupo
## dbl (4): year, esperanza_de_vida, producto_por_capita, nacimientos_por_mujer
##
## 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.
```

```
Vars <- paste0(c("esperanza_de_vida", "producto_por_capita", "nacimientos_por_mujer", "continente", "gdp"))
```

Construcción del Dashboard

La creación de un dashboard permite visualizar información que, por otros medios, no sería posible explorar con la misma claridad. El dashboard titulado “Relación entre fecundidad, esperanza de vida de las mujeres y el Producto Interno Bruto (PIB)” fue desarrollado con el propósito de cumplir los objetivos planteados previamente. Para su diseño, se tomó como referencia el boceto propuesto en el Reto 1.

Aunque no se logró replicar visualmente el diseño del boceto original debido a las limitaciones propias en la herramienta en R, sí fue posible implementar una funcionalidad que permite consultar la información de forma efectiva y dinámica.

En la primera sección del dashboard, el objetivo era ofrecer una visión general de las tres variables principales. En lugar de utilizar gráficos, se optó por presentar los datos mediante tablas dinámicas que permiten al usuario consultar la información según el año seleccionado. Esta decisión respondió al interés de mantener la lógica de visualización del boceto original, que incluye botones para retroceder en intervalos de 10 años. Sin embargo, al implementar un input en R para la selección del año, se consideró que no era necesario replicar dichos botones, ya que el usuario puede seleccionar libremente el año deseado desde un menú desplegable.

Asimismo, para mejorar la visualización, se incorporó el nombre del país, el año seleccionado, la variable correspondiente y la bandera del país, aportando un componente visual atractivo y fácil de interpretar.

La segunda sección, dedicada a los gráficos comparativos de variables, presenta tres gráficos de dispersión con distintos colores, lo que facilita su diferenciación. Además, se añadió la opción de seleccionar un país o continente, lo que permite analizar cómo varía la correlación entre las tres variables seleccionadas según el contexto geográfico.

En la tercera sección, correspondiente a los gráficos de interacción avanzada, se representa la relación entre las tres variables mediante un gráfico de dispersión: el eje x representa el PIB (GDP), el eje y la tasa de fecundidad (hijos por mujer), mientras que el tamaño de los puntos indica la esperanza de vida. Esta visualización permite captar de forma integrada la interacción entre las variables clave del estudio.

Al lado derecho de la tercera sección, por curiosidad y siguiendo el diseño del boceto original, se incluyeron dos gráficos de dispersión adicionales que permiten explorar la relación inversa entre las variables. El primero permite al usuario seleccionar una cantidad de hijos deseada y visualizar la esperanza de vida asociada a ese nivel de fecundidad. El segundo gráfico permite seleccionar una edad específica (esperanza de vida) para observar cuántos hijos, en promedio, tienen las mujeres que alcanzan esa edad. Estas visualizaciones brindan un enfoque interactivo y didáctico para comprender los patrones entre fecundidad y longevidad femenina.

Para finalizar, se incluye una tabla dinámica que permite a las personas usuarias explorar libremente los datos, realizar diferentes combinaciones y responder a nuevas inquietudes o curiosidades que puedan surgir durante la interacción.

Resultados Generales

Ante las investigaciones presentadas en la introducción, surge la duda de si tener menos hijos realmente alarga la esperanza de vida femenina, o si, por el contrario, tener entre 2 y 5 hijos genera un efecto similar. Esta pregunta guía el análisis exploratorio presentado en este dashboard.

En la sección de gráficos de combinación de variables, el gráfico que relaciona fecundidad y esperanza de vida muestra una tendencia general decreciente: a mayor número de hijos por mujer, menor esperanza de vida.

```
datos_filtrados <- Datos_FEGDP

ggplot(datos_filtrados, aes(x = nacimientos_por_mujer, y = esperanza_de_vida)) +
  geom_point(alpha = 0.4, color = "darkred") +
  geom_smooth(method = "lm", se = FALSE, color = "black") +
```

```

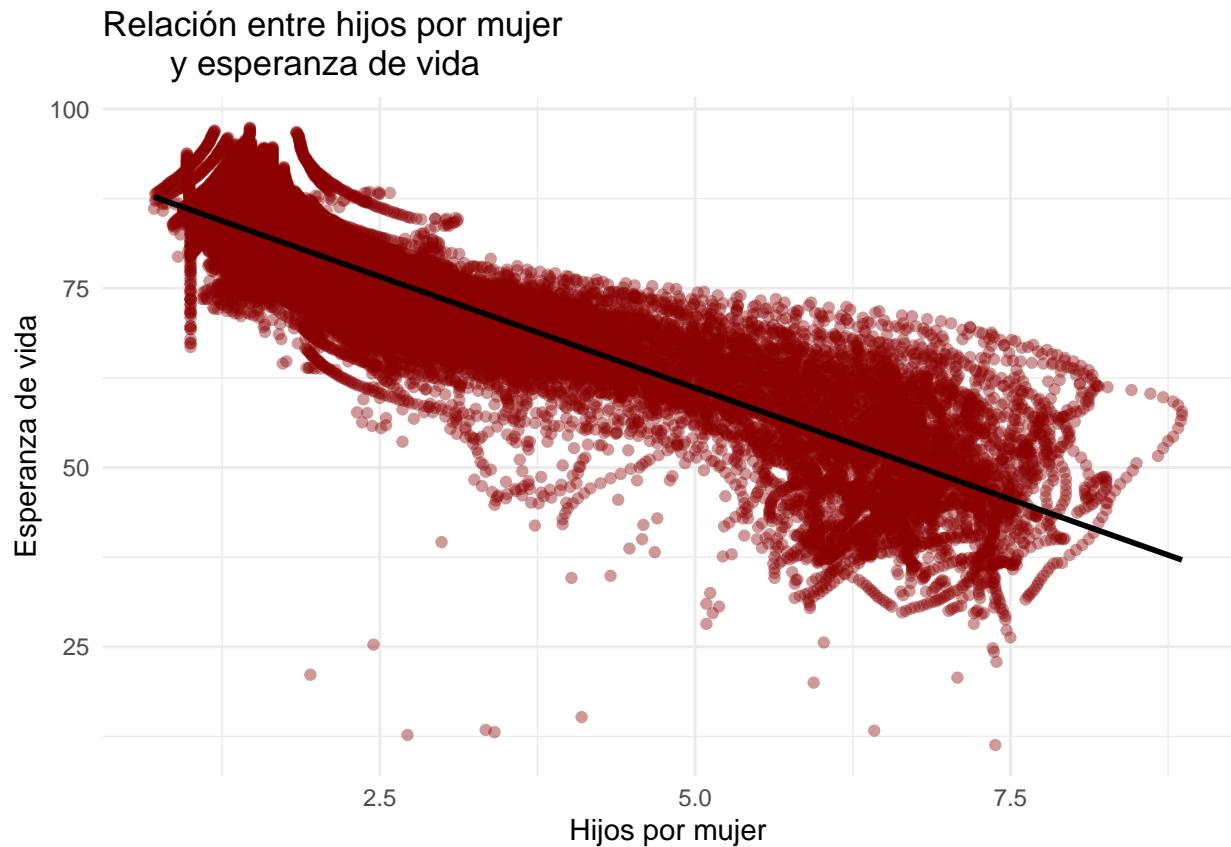
  labs(title = "Relación entre hijos por mujer
    y esperanza de vida", x = "Hijos por mujer", y = "Esperanza de vida") +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 5893 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 5893 rows containing missing values or values outside the scale range
## ('geom_point()').

```



Sin embargo, existen excepciones. Algunos países con 1 o 2 hijos mantienen niveles altos de esperanza de vida, lo que indica que no se puede afirmar de forma categórica que tener menos hijos siempre incrementa la longevidad. Los datos sugieren que a partir de 2.5 hijos por mujer, la esperanza de vida comienza a descender de manera más marcada.

```

hijos <- 1.5

# Filtrar datos según 2 hijos
datos_filtrados <- Datos_FEGDP %>%
  filter(!is.na(nacimientos_por_mujer),
         !is.na(esperanza_de_vida)
  ) %>%

```

```

filter(abs(nacimientos_por_mujer - hijos) < 0.2)

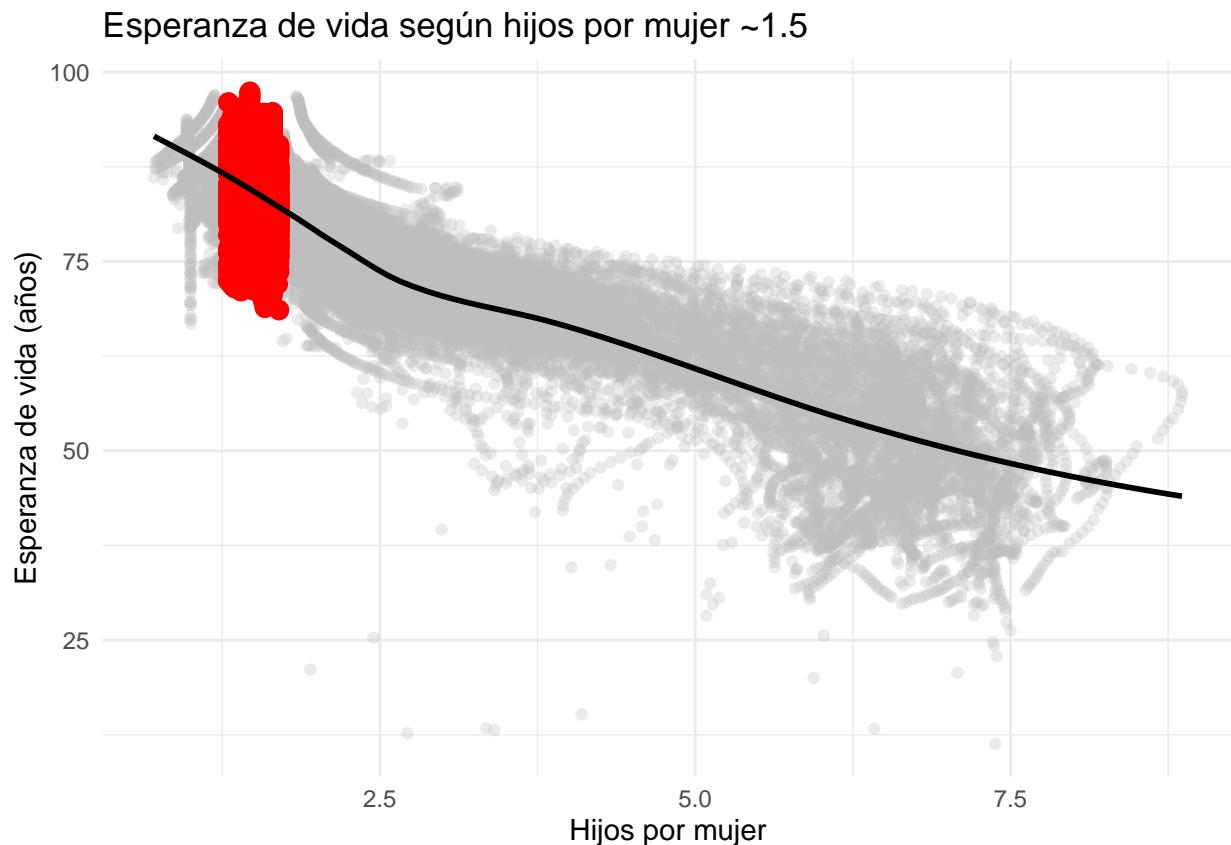
#Gráfico
ggplot(Datos_FEGDP, aes(x = nacimientos_por_mujer, y = esperanza_de_vida)) +
  geom_point(alpha = 0.3, color = "gray") +
  geom_point(data = datos_filtrados, aes(x = nacimientos_por_mujer, y = esperanza_de_vida),
             color = "red", size = 3) +
  geom_smooth(method = "loess", se = FALSE, color = "black") +
  labs(
    title = paste("Esperanza de vida según hijos por mujer",
                  if (!is.na(hijos)) paste0("~", hijos) else ""),
    x = "Hijos por mujer",
    y = "Esperanza de vida (años)"
  ) +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 5893 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 5893 rows containing missing values or values outside the scale range
## (`geom_point()`).

```



En términos regionales, Europa presenta un comportamiento particular. Se observa una caída en la cantidad de hijos por mujeres y con esto también la esperanza de vida en países con 2 hijos por mujer, aunque no se trata de una reducción sistemática. De hecho, varios países con solo 1 hijo muestran menor esperanza de vida que aquellos con 2, lo cual podría deberse a factores contextuales como el envejecimiento poblacional o las políticas de salud.

```
datos_filtrados <- Datos_FEGDP %>%
  filter(continente == "Europe")

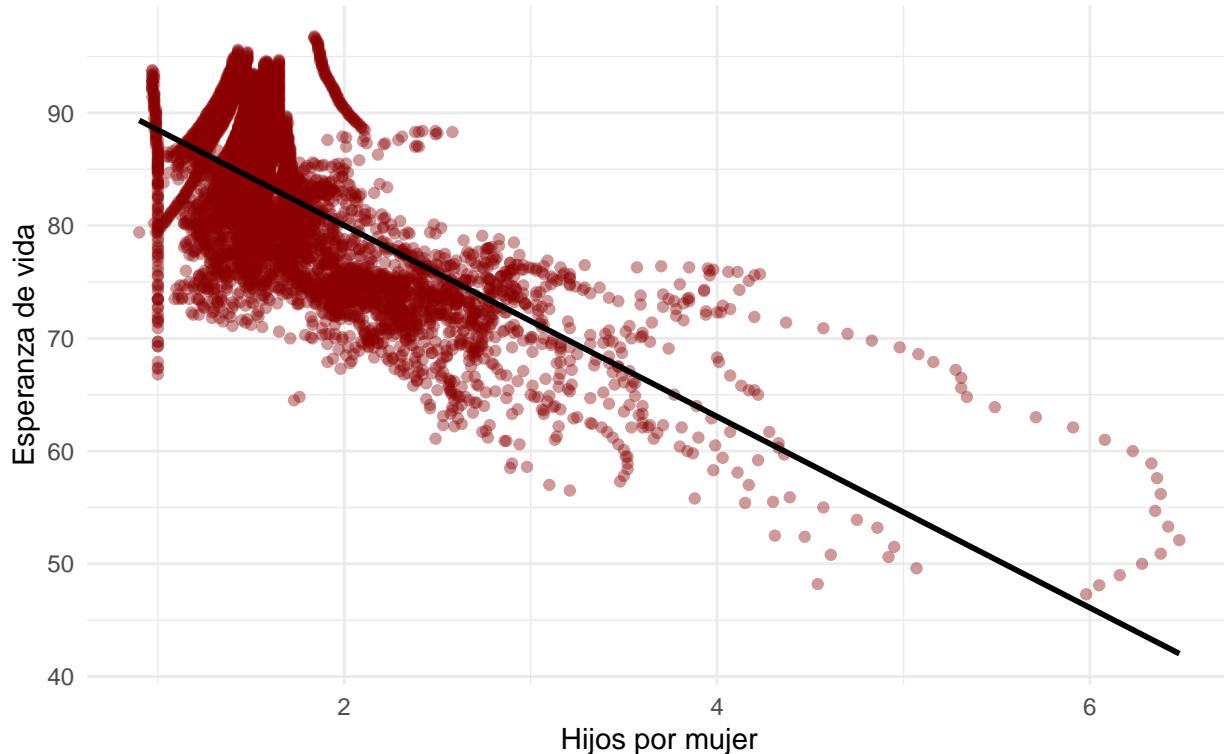
ggplot(datos_filtrados, aes(x = nacimientos_por_mujer, y = esperanza_de_vida)) +
  geom_point(alpha = 0.4, color = "darkred") +
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  labs(title = "Relación entre hijos por mujer  
y esperanza de vida de Europa", x = "Hijos por mujer", y = "Esperanza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

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 906 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 906 rows containing missing values or values outside the scale range
## ('geom_point()').
```

**Relación entre hijos por mujer
y esperanza de vida de Europa**



En el caso de África, se registra la esperanza de vida más baja en comparación con otros continentes, manteniéndose la tendencia decreciente a pesar de una alta fecundidad.

```

datos_filtrados <- Datos_FEGDP %>%
  filter(continente == "Africa")

ggplot(datos_filtrados, 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 de África", x = "Hijos por mujer", y = "Esperanza de vida") +
  theme_minimal()

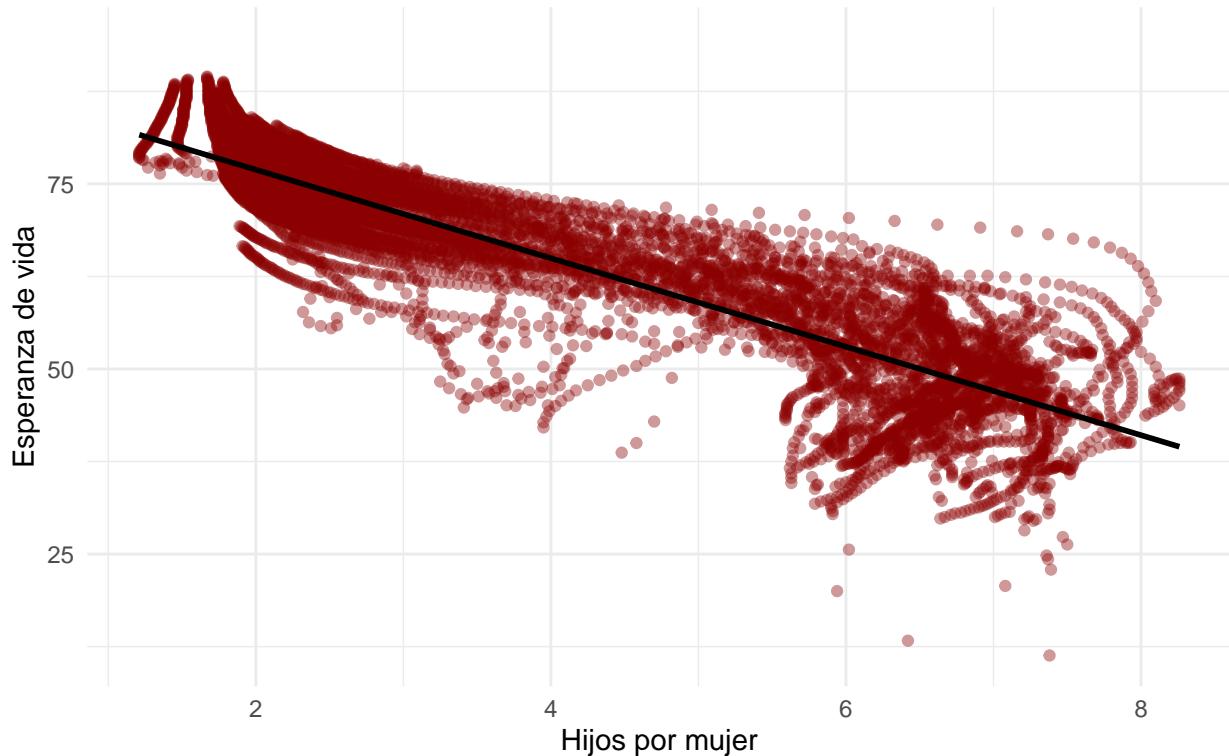
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 606 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 606 rows containing missing values or values outside the scale range
## ('geom_point()').

```

Relación entre hijos por mujer y esperanza de vida de África



En esta región, el bajo nivel del PIB per cápita también resalta como un elemento que podría influir en los resultados, reflejando bajos niveles de desarrollo económico.

```

datos_filtrados <- Datos_FEGDP %>%
  filter(continente == "Africa")

ggplot(datos_filtrados, 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 de África", x = "GDP per cápita (log)", y = "Esperanza de vida") +
theme_minimal()+
theme(
  plot.title = element_text(size = 20, face = "bold"),
  axis.title = element_text(size = 15)
)

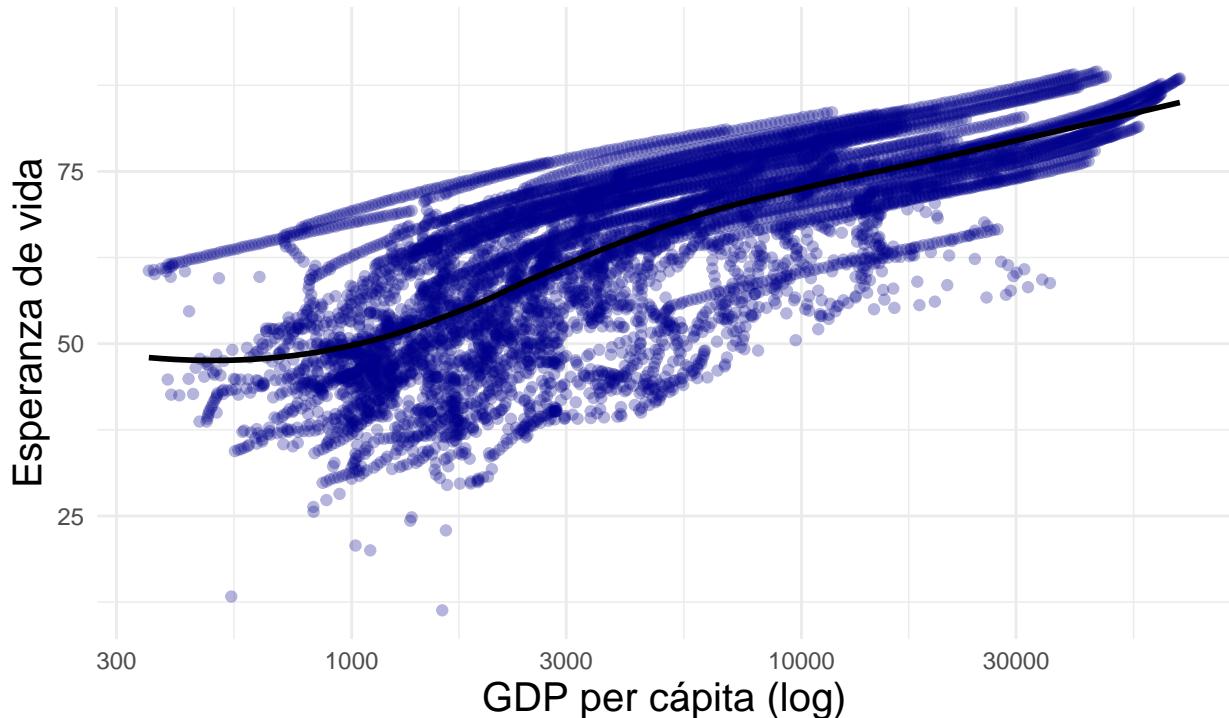
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 906 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 906 rows containing missing values or values outside the scale range
## (`geom_point()`).

```

Relación entre GDP per cápita y esperanza de vida de África



En las Américas, se conserva la tendencia general: los países con mayor fecundidad tienden a registrar niveles más bajos de esperanza de vida. Este patrón refuerza la hipótesis de una correlación negativa entre estas dos variables, aunque no necesariamente causal.

```

datos_filtrados <- Datos_FEGDP %>%
  filter(continente == "Americas")

ggplot(datos_filtrados, 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 de America", x = "Hijos por mujer", y = "Esperanza de vida") +
  theme_minimal()

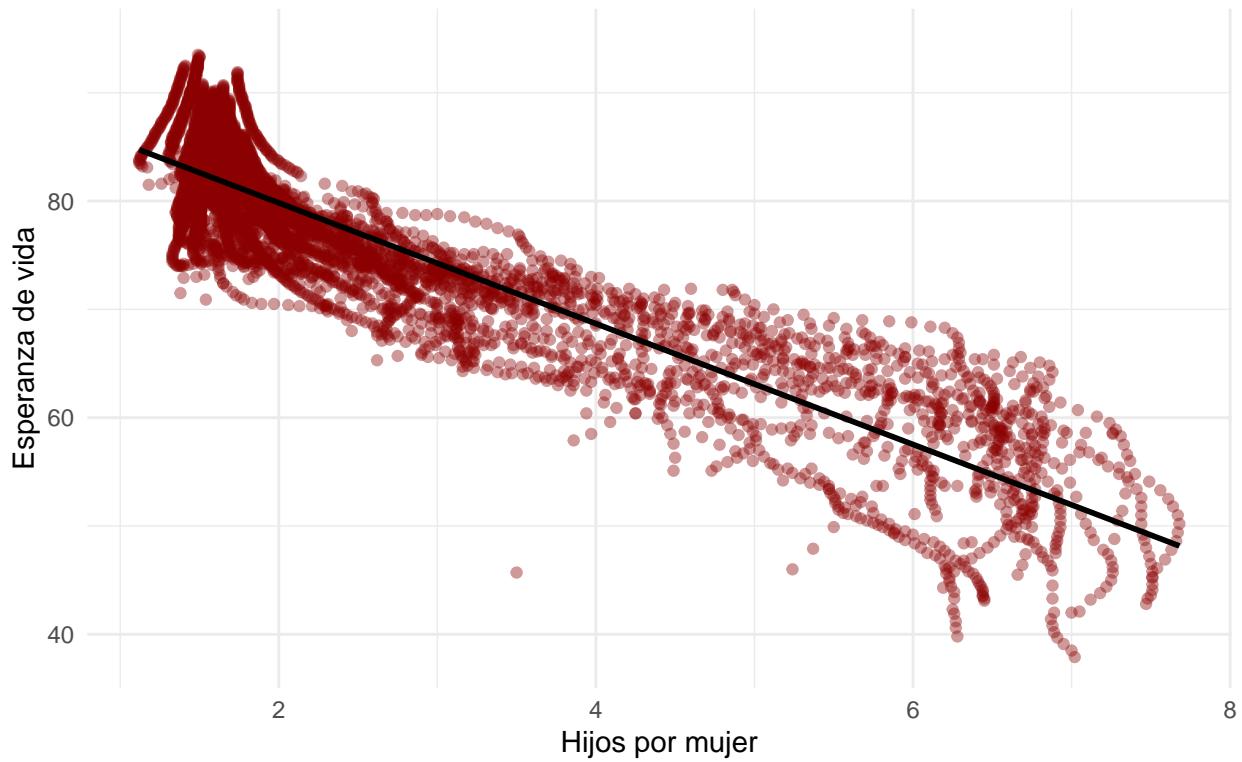
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 2869 rows containing non-finite outside the scale range
## ('stat_smooth()').

## Warning: Removed 2869 rows containing missing values or values outside the scale range
## ('geom_point()').

```

**Relación entre hijos por mujer
y esperanza de vida de America**



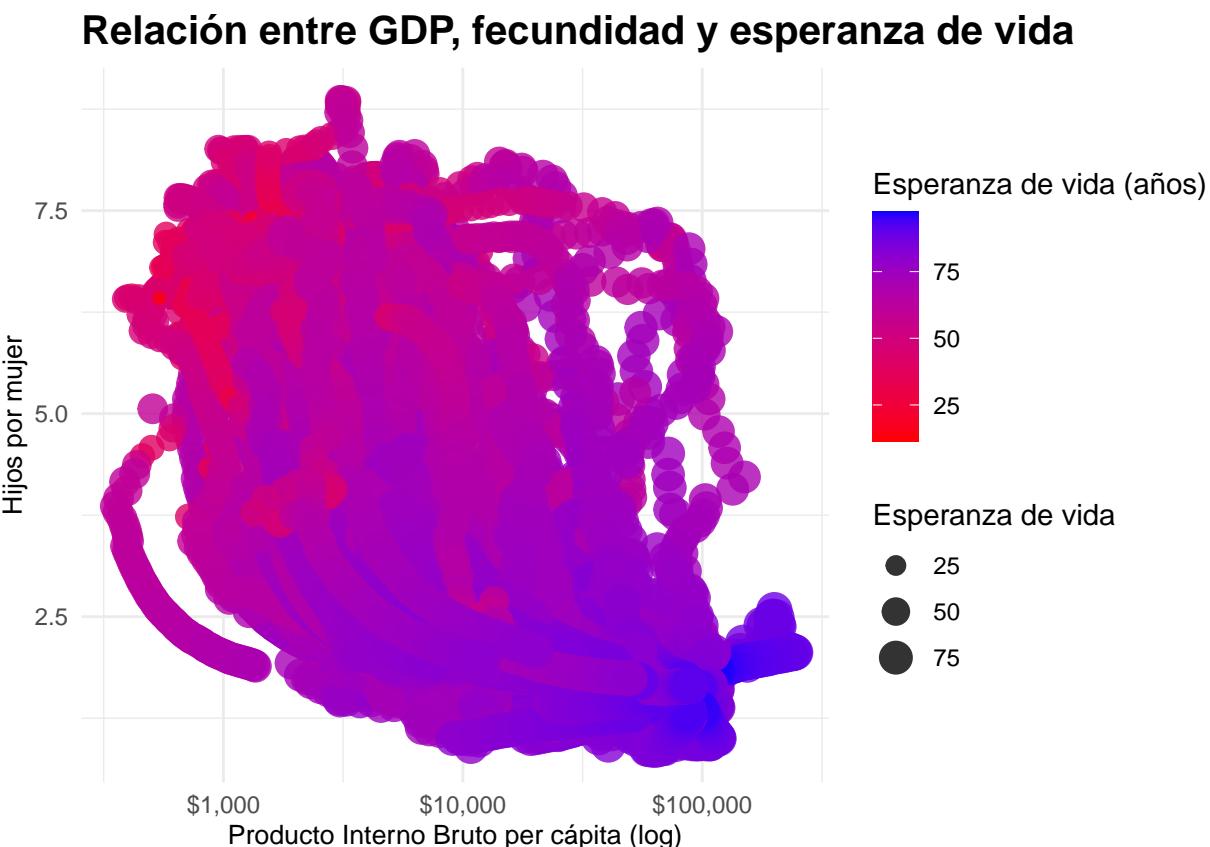
En la sección de gráficos de interacción avanzada, se identifican valores altos de esperanza de vida asociados a tasas de fecundidad cercanas a 0.7 hijos por mujer, con una esperanza de vida promedio entre 86 y 89 años. Por otro lado, los países con una fecundidad de 1 hijo muestran una esperanza de vida aún mayor, entre 87 y 96 años. Esto sugiere que los países con baja fecundidad tienden a presentar mayores niveles de longevidad. Sin embargo, no se observa un patrón lineal estricto: no puede afirmarse que a menor número de hijos por mujer siempre corresponda una mayor esperanza de vida.

```

datos_filtrados <- Datos_FEGDP %>%
  filter(
    !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",
       x = "Producto Interno Bruto per cápita (log)",
       y = "Hijos por mujer",
       color = "Esperanza de vida (años)",
       size = "Esperanza de vida") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 15, face = "bold"),
    axis.title = element_text(size = 10)
  )

```



Es importante aclarar que los datos analizados provienen de una investigación exploratoria con información agregada a nivel país, por lo que no permiten establecer relaciones causales individuales. Lo que se observa son asociaciones a nivel macro, que pueden orientar futuras investigaciones.

Aunque no se puede concluir que tener hijos necesariamente alarga o acorta la vida, los gráficos permiten identificar tendencias consistentes, especialmente al analizar los datos por regiones o niveles de desarrollo económico.

En cuanto al PIB per cápita, el gráfico de interacción avanzada muestra cómo esta variable se asocia tanto con fecundidad como con esperanza de vida. En el gráfico “Relación entre GDP, fecundidad y esperanza de vida”, se observa que los países con PIB superiores a 100,000 dólares se ubican principalmente en la franja de 2 a 2.5 hijos por mujer, y presentan una esperanza de vida alta (representada en color azul). Esto evidencia que, si bien la baja fecundidad puede asociarse con mayor longevidad, el nivel de desarrollo económico es un factor modulador clave.

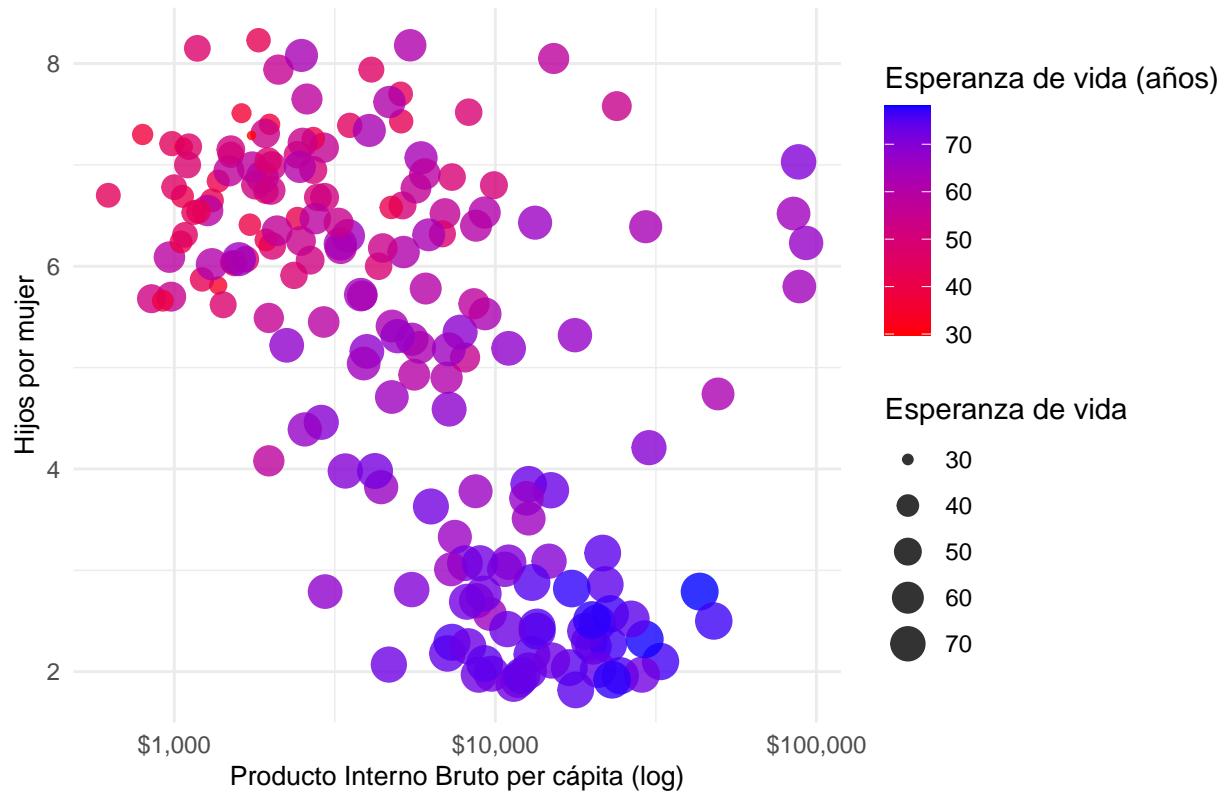
Al comparar países según sus niveles de ingreso, se observa que en aquellos con mayor PIB per cápita, la relación entre baja fecundidad y alta esperanza de vida es más clara y estable. En contraste, los países con ingresos bajos presentan relaciones más dispersas, posiblemente influenciadas por factores externos como conflictos, pobreza estructural o debilidad institucional.

En cuanto al eje temporal, en el gráfico “Relación entre GDP, fecundidad y esperanza de vida”, se aprecia que en 1970 la esperanza de vida promedio mundial rondaba los 80 años.

```
datos_filtrados <- Datos_FEGDP %>%
  filter( year == 1970,
         !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",
       x = "Producto Interno Bruto per cápita (log)",
       y = "Hijos por mujer",
       color = "Esperanza de vida (años)",
       size = "Esperanza de vida") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 15, face = "bold"),
    axis.title = element_text(size = 10)
  )
```

Relación entre GDP, fecundidad y esperanza de vida

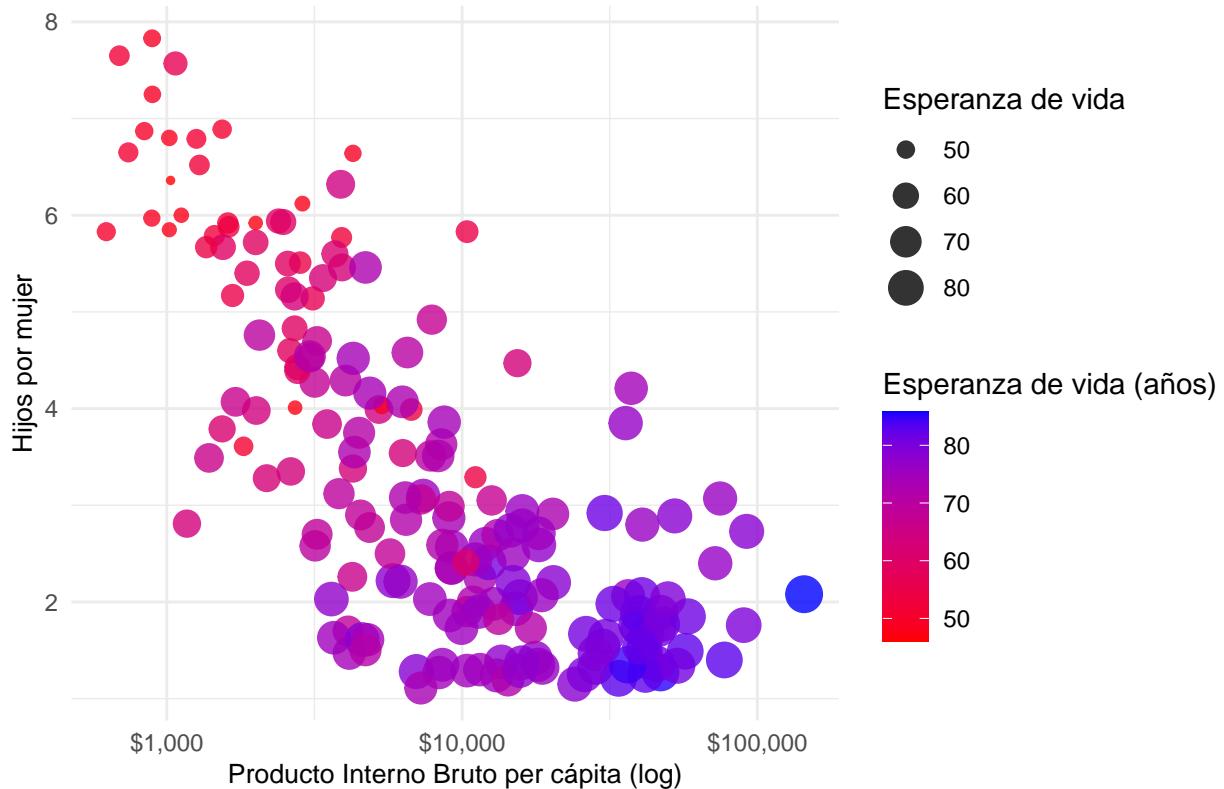


Para el año 2000, esta cifra aumentó casi en una década, acercándose a los 90 años. Este incremento evidencia mejoras globales en salud, condiciones de vida y acceso a servicios básicos.

```
datos_filtrados <- Datos_FEGDP %>%
  filter( year == 2000,
  !is.na(producto_por_capita),
  !is.na(nacimientos_por_muher),
  !is.na(esperanza_de_vida)
)

# Gráfico
ggplot(datos_filtrados, aes(x = producto_por_capita,
                             y = nacimientos_por_muher,
                             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",
       x = "Producto Interno Bruto per cápita (log)",
       y = "Hijos por mujer",
       color = "Esperanza de vida (años)",
       size = "Esperanza de vida") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 15, face = "bold"),
    axis.title = element_text(size = 10)
```

Relación entre GDP, fecundidad y esperanza de vida



Aunque la fecundidad no es un determinante único en la decisión de tener hijos ni en la esperanza de vida, constituye un elemento relevante en el análisis con perspectiva de género, ya que las decisiones reproductivas están influenciadas por factores sociales, económicos y culturales.

En conclusión, los datos permiten identificar patrones generales entre esperanza de vida, fecundidad y nivel de desarrollo económico. No obstante, esta investigación representa solo un punto de partida para análisis más precisos. En el futuro, podrían desarrollarse modelos predictivos que, tomando en cuenta el número de hijos y el nivel económico, estimen la esperanza de vida de las mujeres, sin considerar aún factores médicos o genéticos. Además, será necesario utilizar otros tipos de gráficos, técnicas estadísticas e indicadores adicionales para abordar nuevas preguntas y profundizar en las causas detrás de estas asociaciones.