

Article

Evolución de la actividad, ocupación para grupos de edad mayores de 16 años

Jeremy Joel Aguilar Marin, Jose Aguilar Milla, Javier Herrero Pérez

* Correspondence:

Simple Summary: A Simple summary goes here.**Keywords:** keyword 1; keyword 2; keyword 3 (list three to ten pertinent keywords specific to the article, yet reasonably common within the subject discipline.).

1. Introducción

Este artículo presenta un enfoque en R para el Análisis Exploratorio de Datos de la Encuesta de Población Activa (EPA), centrándose en la evolución histórica de la brecha de género en el mercado laboral español. Utilizando el ecosistema tidyverse, el estudio comienza con una limpieza y transformación de los datos empleando dplyr y el cálculo de Tasa de paro, actividad y ocupación, claves para el análisis. Seguido de la visualización con ggplot2 para ilustrar la evolución temporal. Por último, se validan los resultados con pruebas de *t Student* para confirmar la significancia estadística de los resultados.

1.1. Carga de Librerías

```
##  
## Adjuntando el paquete: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##     filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##     intersect, setdiff, setequal, union
```

Citation: . Evolución de la actividad, ocupación para grupos de edad mayores de 16 años. *Journal Not Specified* 2024, 1, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

Copyright: © 2025 by the authors. Submitted to *Journal Not Specified* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

2. Exploración inicial

2.1. Importación y verificación inicial

- **Codificación e importación:** Se verifica la codificación y se importa el archivo CSV delimitado por tabulaciones, se indica que el separador decimal es la coma y que los valores NA están codificados como "...".

```
## # A tibble: 3 x 2  
##   encoding  confidence  
##   <chr>        <dbl>  
## 1 UTF-8          1  
## 2 ISO-8859-1    0.98  
## 3 ISO-8859-2    0.28  
  
## Rows: 4446 Columns: 5  
## -- Column specification -----  
## Delimiter: "\t"
```

```

## chr (3): Sexo, Edad, Relación con la actividad económica      35
## dbl (1): Periodo                                         36
## num (1): Total                                         37
##
## i Use 'spec()' to retrieve the full column specification for this data. 39
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message 39

## # A tibble: 6 x 5                                         41
##   Sexo          Edad  `Relación con la actividad económica` Periodo Total
##   <chr>        <chr> <chr>                               <dbl> <dbl>
## 1 Ambos sexos Total Total                           2024 41566.
## 2 Ambos sexos Total Total                           2023 40983.
## 3 Ambos sexos Total Total                           2022 40367.
## 4 Ambos sexos Total Total                           2021 39926.
## 5 Ambos sexos Total Total                           2020 39579.
## 6 Ambos sexos Total Total                           2019 39269. 49

```

Se comprueba la estructura inicial de los datos

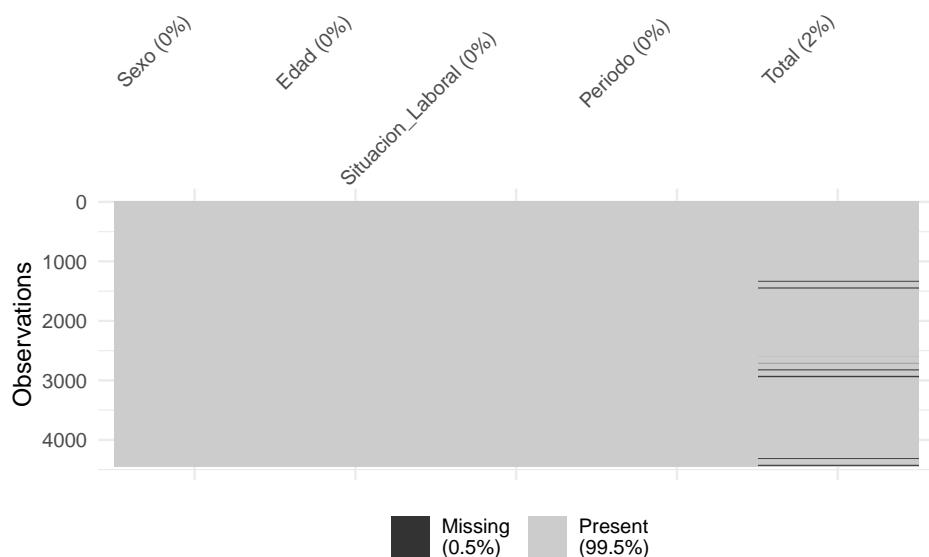
```

## spc_tbl_ [4,446 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)      51
## $ Sexo                      : chr [1:4446] "Ambos sexos" "Ambos sexos"
## $ Edad                      : chr [1:4446] "Total" "Total" "Total" ...
## $ Relación con la actividad económica: chr [1:4446] "Total" "Total" "Total" ...
## $ Periodo                    : num [1:4446] 2024 2023 2022 2021 2020 ...
## $ Total                      : num [1:4446] 41566 40983 40367 39926 39
## - attr(*, "spec")=           57
##   .. cols(                   58
##     .. Sexo = col_character(), 59
##     .. Edad = col_character(), 60
##     .. `Relación con la actividad económica` = col_character(), 61
##     .. Periodo = col_double(), 62
##     .. Total = col_number()    63
##   .. )
## - attr(*, "problems")=<externalptr> 65

```

2.2. Valores faltantes

Con el objetivo de ver la cantidad de los datos faltantes la variable que tiene estos datos se utiliza la siguiente función:

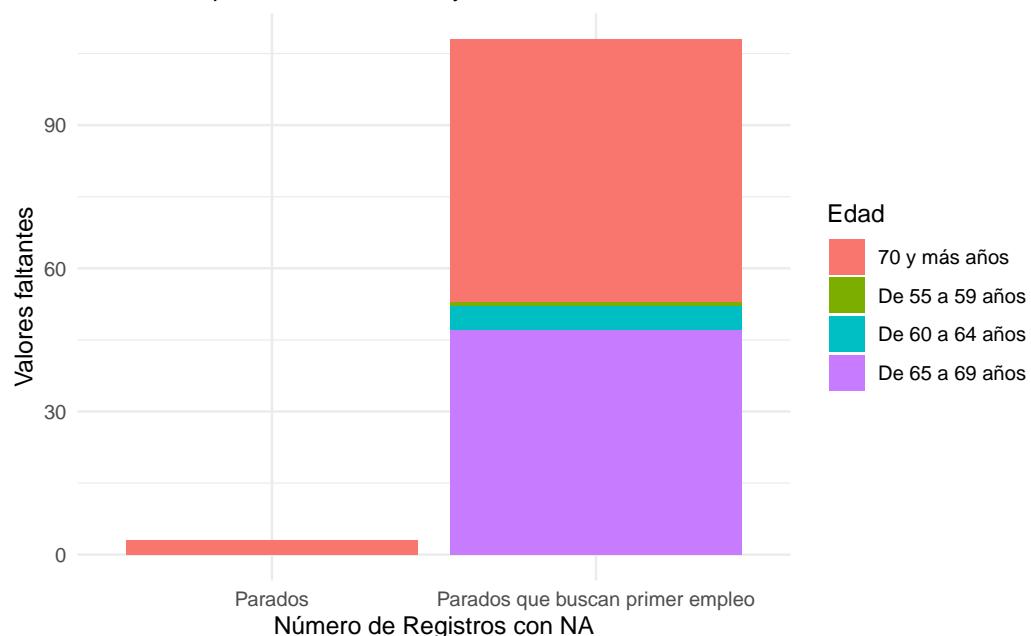


##	Sexo	Edad
##	0	0
## Relación con la actividad económica		70
##	0	71
##	Total	72
##	111	0
		73
		74
		75

Todos los valores ausentes se encuentran en Total

Frecuencia de los registros faltantes NA

Distribución por Situación Laboral y la edad



Se observa que los valores faltantes vienen asociados a grupos de edad altos y cuya relación con la actividad económica es "Parados que buscan primer empleo" o "Parados".

3. Definición de variables

El valor de la columna **Total** viene definido con la siguiente expresión:

Total = Activos + Inactivos

Y los **Activos** se dividen en Ocupados y Parados:

Activos = Ocupados + Parados

Fórmulas para el análisis (Tasas)

Las tasas permiten normalizar los datos y son la base para el análisis de brechas y series temporales:

La **tasa de paro** mide la proporción de la población activa que está desempleada:

$$\text{Tasa de paro} = \frac{\text{Parados}}{\text{Activos}} \cdot 100$$

La **tasa de actividad** mide la participación de una población en el mercado laboral:

$$\text{Tasa de actividad} = \frac{\text{Activos}}{\text{Total}} \cdot 100$$

La tasa de ocupación muestra qué parte de la población total de referencia tiene un empleo:

$$\text{Tasa de ocupación} = \frac{\text{Ocupados}}{\text{Total}} \cdot 100$$

4. Limpieza

- **Renombrar columnas:** Utilizando `rename()` para acortar "Relación con la actividad económica" por "Situacion_Laboral" para simplificar el manejo de columnas. 91
 - **Filtrado de doble conteo:** Se eliminan las instancias donde las variables representan la suma de otras instancias (Sexo="Ambos sexos" y Edad="Total") para evitar doble conteo y así verificar que cada registro sea una observación única. 92
 - **Se aplica `pivot_wider()`** utilizando los valores de la columna Situacion_Laboral para crear las nuevas columnas (`names_from`) y los valores numéricos de la columna Total. 93
 - **Renombrar variables:** Se renombra la columna "parados que buscan primer empleo" a un formato más corto Parados_sin_exp 94
 - **Imputación de NA a 0:** Los valores faltantes se concentran en categorías como Personas mayores de 65 en paro o que buscan su primer empleo. El INE suprime el dato por ser un valor insignificante, por lo que se opta por remplazar los valores nulos a 0 para no perder el resto de información válida de la instancia. 95
 - **Creación de tasas:** Se añaden las columnas Tasa_paro, Tasa_actividad y Tasa_ocupacion utilizando `mutate()` y las expresiones de la EPA. 96
 - **Ordenación en factores:** Se ordena la columna Edad como un factor ordinal y la columna Sexo como factor nominal. 97

Se comprueba que la limpieza esté bien hecha, primero viendo si el formato de los datos es el adecuado y asegurando que no quedan valores faltantes.

```
## # tibble [456 x 12] (S3: tbl_df/tbl/data.frame) 110
## # $ Sexo          : Factor w/ 2 levels "Hombres","Mujeres": 1 1 1 1 1 1 1 1 1111 1
## # $ Edad          : Ord.factor w/ 12 levels "16-19"  
"20-24"  
...: 1 1 1 1 1 1 1 1 1212 1
## # $ Periodo       : num [1:456] 2024 2023 2022 2021 2020 ... 113
## # $ Total         : num [1:456] 1095 1046 1019 984 992 ... 114
```

```

## $ Activos      : num [1:456] 189 167 161 144 142 ...
## $ Ocupados     : num [1:456] 117.2 94.7 88.1 72.4 71 ...
## $ Parados      : num [1:456] 71.6 72.4 73.1 71.4 71.3 67.5 75.2 82.5 81 ...
## $ Parados_sin_exp: num [1:456] 42.7 43.8 44.3 46.6 39 42.2 49 54.2 52.8 69 ...
## $ Inactivos    : num [1:456] 906 879 857 840 850 ...
## $ Tasa_paro    : num [1:456] 37.9 43.3 45.3 49.7 50.1 ...
## $ Tasa_actividad: num [1:456] 17.2 16 15.8 14.6 14.4 ...
## $ Tasa_ocupacion: num [1:456] 10.7 9.05 8.65 7.36 7.16 ...
##          Sexo        Edad       Periodo      Total      Activos
##          0           0           0           0           0           0
##          Ocupados    Parados   Parados_sin_exp  Inactivos  Tasa_paro
##          0           0           0           0           0           0
##          Tasa_actividad  Tasa_ocupacion
##          0           0

```

5. Evolución temporal de las tasas con la brecha

##Evolucion de la tasa de paro por género y evolución de la brecha de la tasa de paro por género

En este bloque de código se construyen varias representaciones gráficas con el fin de analizar la evolución de la tasa de paro y la brecha de género a lo largo del tiempo. Para ello, se parte de una tabla en la que se calcula la variable *brecha_paro*, definida como la diferencia entre la tasa de paro de mujeres y la de hombres. Posteriormente, los datos se agrupan por periodo para obtener las tasas medias anuales de paro por sexo y la brecha media correspondiente. IMPORTAMOS EL ARCHIVO CSV

```

library(readr)
df_clean <- read_csv("data/df_clean.csv")

## # Rows: 456 Columns: 12
## -- Column specification -----
## Delimiter: ","
## chr (2): Sexo, Edad
## dbl (10): Periodo, Total, Activos, Ocupados, Parados, Parados_sin_exp, Inact...
## 
## 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

```

Aggrupación de datos por periodo y por sexo

```

# Calculamos la media de la tasa de actividad para cada grupo
# Usamos na.rm = TRUE para ignorar valores faltantes
library(tidyverse)

```

```

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0
## v lubridate 1.9.4      v tibble    3.3.0
## v purrr     1.1.0
## -- 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

```

```

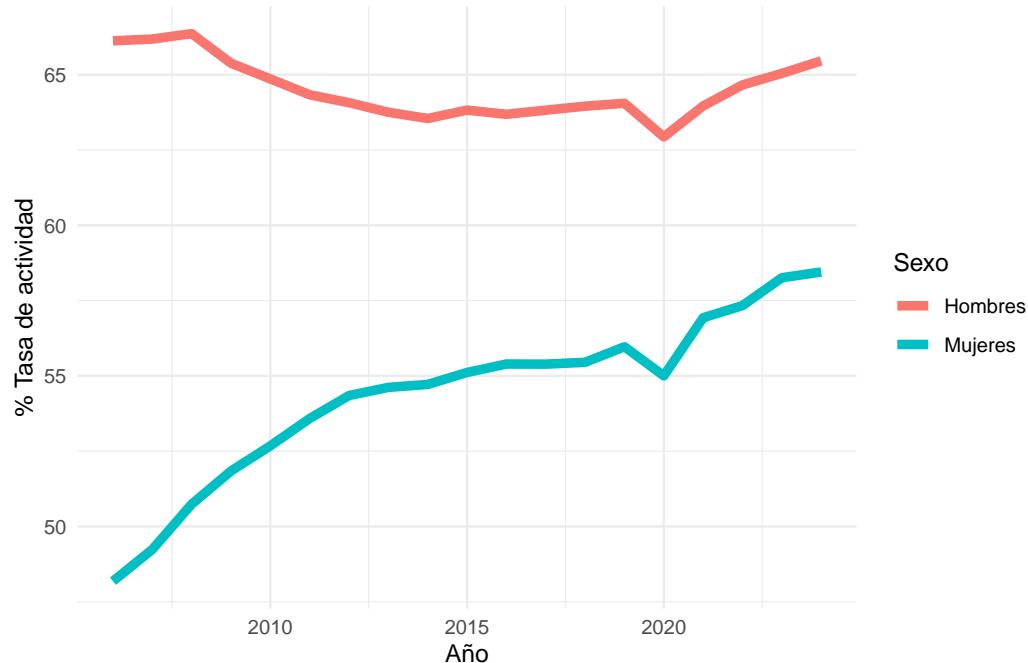
df_p_sexo <- df_clean %>% group_by(Periodo, Sexo) %>% summarise(tasa_actividad = mean(tasa_actividad),
, .groups = "drop")

```

#Que pretendemos? Calcular la media de la tasa de actividad de todos los grupos de

```
# Gráfica
library(ggplot2)
ggplot(df_p_sexo, aes(x = Periodo, y = tasa_actividad, color = Sexo)) + geom_line()
```

Tasa de actividad según sexo (2006–2024)



154

```
# Eje X: Periodo - Eje Y: tasa de actividad promedio
```

```
cat("La gráfica enseña la evolución de la tasa de actividad en España entre los años 2006 y 2024. Pese a esto, la actividad masculina en 2008 se desacelera economicamente, esto por el otro lado, la tasa de actividad femenina se muestra, en el periodo, ascen...")
```

```
## La gráfica enseña la evolución de la tasa de actividad en España entre los años 2006 y 2024. Pese a esto, la actividad masculina en 2008 se desacelera economicamente, esto por el otro lado, la tasa de actividad femenina se muestra, en el periodo, ascen...)
```

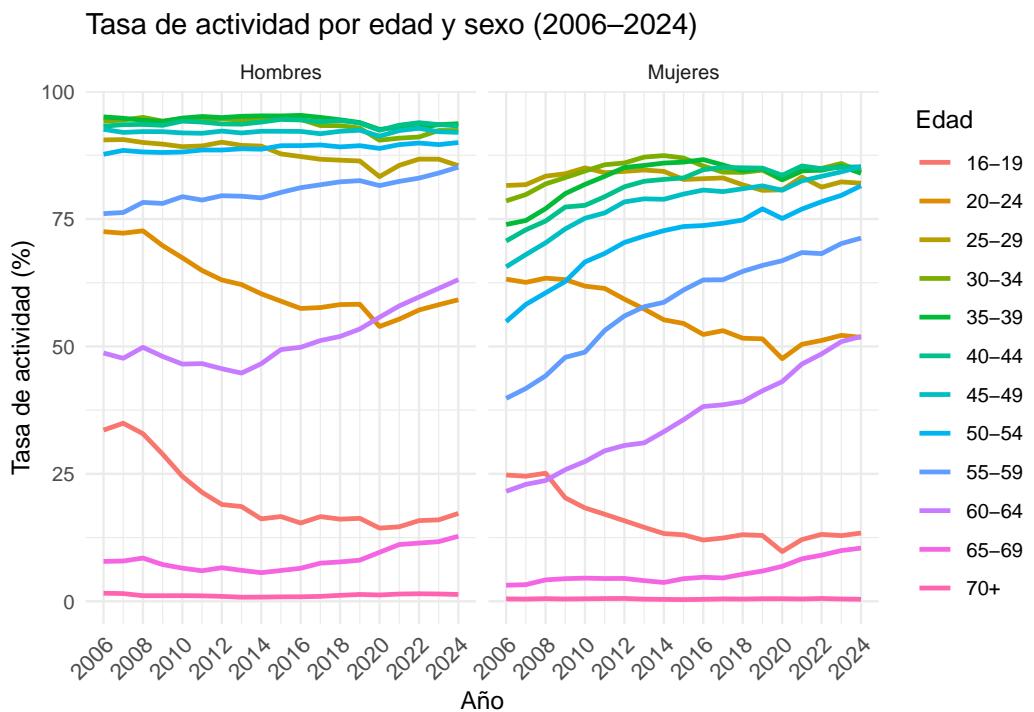
Tasa de actividad por edad y sexo

158

```
# Agrupamos por Periodo, sexo y Edad
```

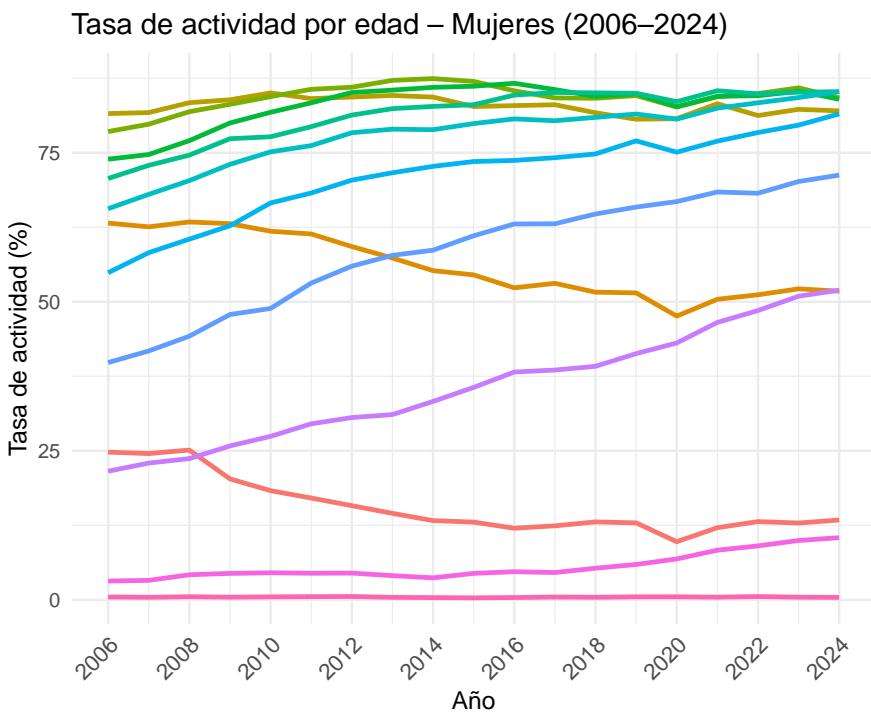
```
## Que pretendemos? Calcular la media de la tasa de actividad para cada combinación de sexo y edad
df_sexo_edad <- df_clean %>% group_by(Periodo, Sexo, Edad) %>% summarise(tasa_actividad = mean(tasa_actividad))
```

```
ggplot(df_sexo_edad, aes(x = Periodo, y = tasa_actividad, color = Edad)) + geom_line()
```



```
df_mujeres <- df_sexo_edad %>% filter(Sexo == "Mujeres")

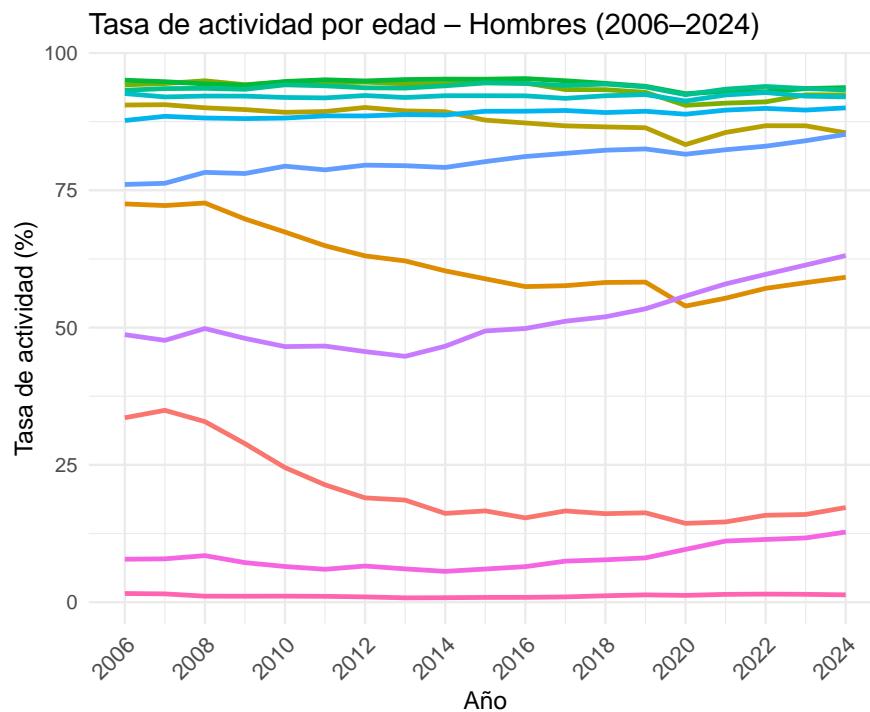
ggplot(df_mujeres, aes(x = Periodo, y = tasa_actividad, color = Edad)) +
  geom_line(linewidth = 1) +
  scale_x_continuous(breaks = seq(2006, 2024, by = 2)) +
  scale_y_continuous(labels = scales::comma) +
  labs(
    title = "Tasa de actividad por edad - Mujeres (2006-2024)",
    x = "Año",
    y = "Tasa de actividad (%)",
    color = "Edad"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 10))
```



160

```
df_hombres <- df_sexo_edad %>% filter(Sexo == "Hombres")

ggplot(df_hombres, aes(x = Periodo, y = tasa_actividad, color = Edad)) +
  geom_line(linewidth = 1) +
  scale_x_continuous(breaks = seq(2006, 2024, by = 2)) +
  scale_y_continuous(labels = scales::comma) +
  labs(
    title = "Tasa de actividad por edad - Hombres (2006-2024)",
    x = "Año",
    y = "Tasa de actividad (%)",
    color = "Edad"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 10))
```



TASA DE PARO POR SEXO

```

df_paro_sexo <- df_clean %>% group_by(Periodo, Sexo) %>% summarise(tasa_paro = mean)

ggplot(df_paro_sexo, aes(x = Periodo, y = tasa_paro, color = Sexo)) + geom_line(sil

```

161

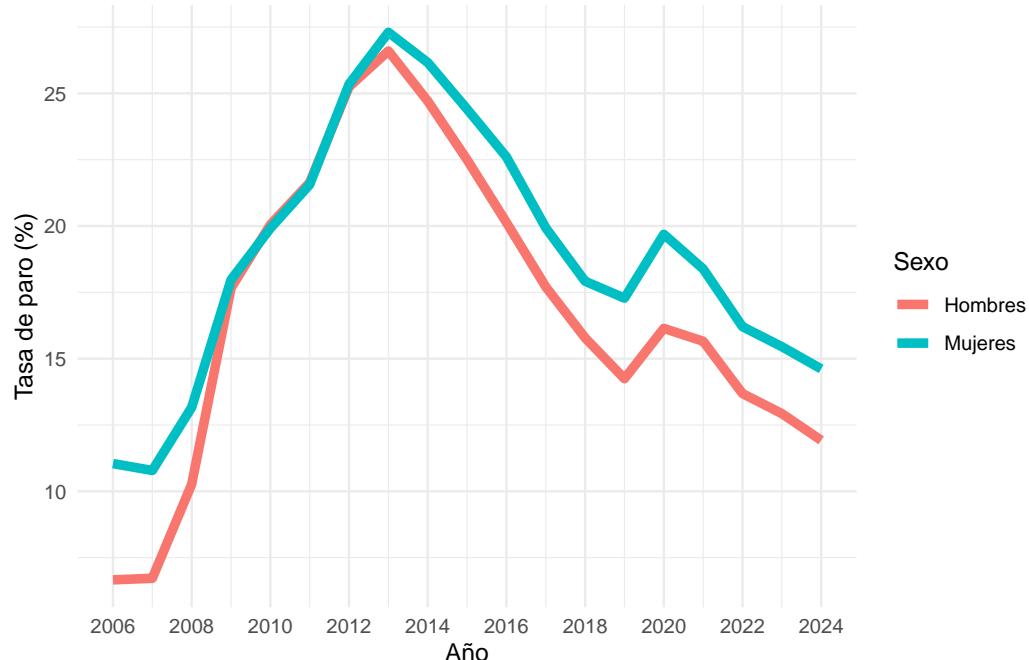
162

```

## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0. 163
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was 164
## generated. 165
166
167

```

Tasa de paro por sexo (2006–2024)



```

library(dplyr)
library(tidyr)
library(ggplot2)

# 1º crear una tabla con la variable brecha_paro
brecha <- df_clean %>%
  select(Periodo, Edad, Sexo, Tasa_paro) %>%
  pivot_wider(names_from = Sexo, values_from = Tasa_paro) %>%
  mutate(brecha_paro=Mujeres-Hombres)
brecha

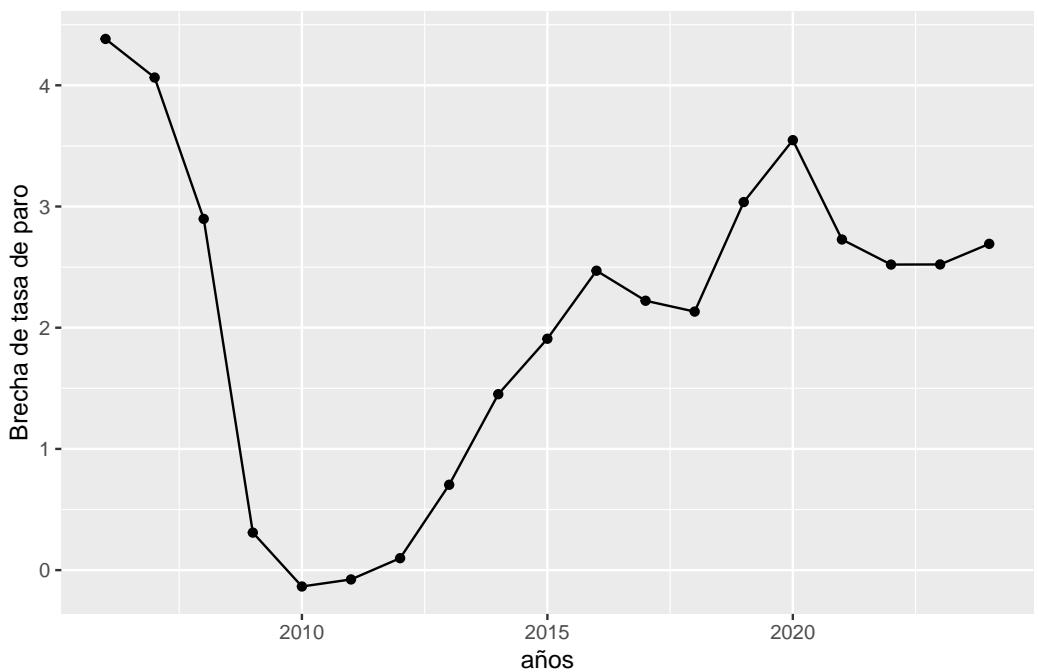
## # A tibble: 228 x 5
##   Periodo Edad   Hombres Mujeres brecha_paro
##   <dbl>   <chr>    <dbl>    <dbl>      <dbl>
## 1 2024 16-19    37.9    44.4     6.43
## 2 2023 16-19    43.3    42.8    -0.493
## 3 2022 16-19    45.3    44.4    -0.912
## 4 2021 16-19    49.7    52.3     2.69
## 5 2020 16-19    50.1    61.0     10.9
## 6 2019 16-19    42.8    48.4     5.59
## 7 2018 16-19    49.7    49.7     0.0375
## 8 2017 16-19    54.2    55.1     0.919
## 9 2016 16-19    58.9    61.6     2.74
## 10 2015 16-19   64.7    70.7     6.04
## # i 218 more rows
## # A tibble: 19 x 4
##   Periodo tasa_hombres tasa_mujeres brecha_paro_total
##   <dbl>        <dbl>        <dbl>          <dbl>
## 1 2006         6.66       11.0          4.38
## 2 2007         6.72       10.8          4.06
## 3 2008        10.3        13.2          2.90
## 4 2009        17.7        18.0          0.310
## 5 2010        20.1        19.9         -0.135
## 6 2011        21.6        21.6         -0.0770
## 7 2012        25.2        25.3          0.0984
## 8 2013        26.6        27.3          0.704
## 9 2014        24.7        26.2          1.45
## 10 2015       22.5        24.4          1.91
## 11 2016       20.1        22.6          2.47
## 12 2017       17.7        19.9          2.22
## 13 2018       15.8        17.9          2.13
## 14 2019       14.2        17.3          3.04
## 15 2020       16.1        19.7          3.55

```

## 16	2021	15.7	18.4	2.73	201
## 17	2022	13.7	16.2	2.52	202
## 18	2023	12.9	15.5	2.52	203
## 19	2024	11.9	14.6	2.69	204

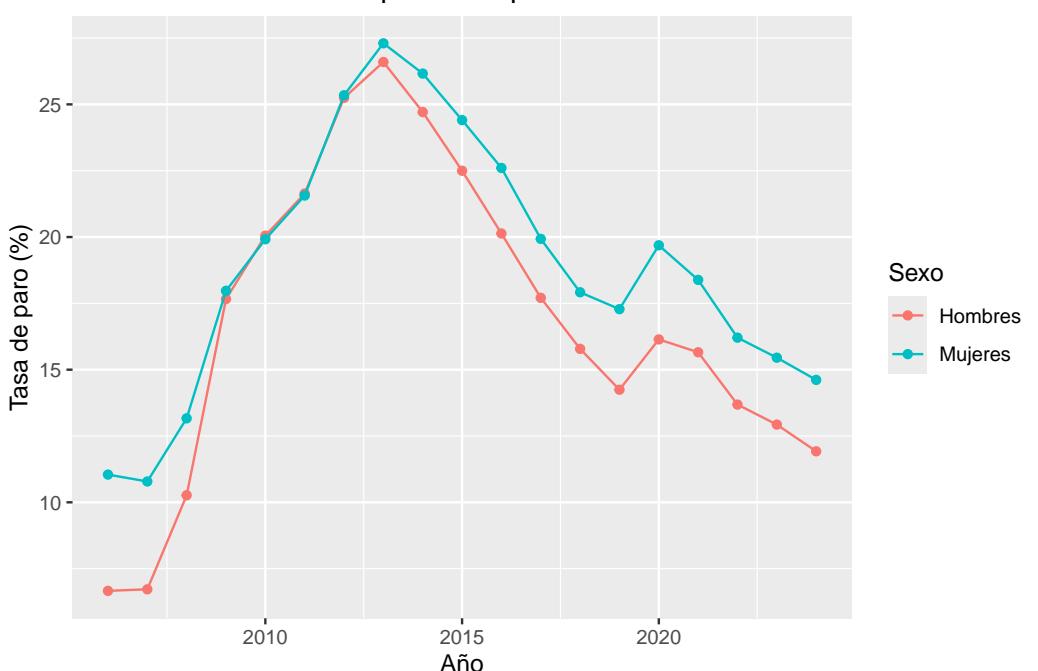
```
evo_brecha_total <- ggplot(brechas_total,aes(x=Periodo,y=brecha_paro_total)) + geom_line(aes(y = brecha_paro_total))
```

Evolución de la brecha de paro total



```
evo_tasa_total <- ggplot(brechas_total, aes(x = Periodo)) + geom_line(aes(y = tasa_paro_total))
```

Evolución de la tasa de paro total por sexo

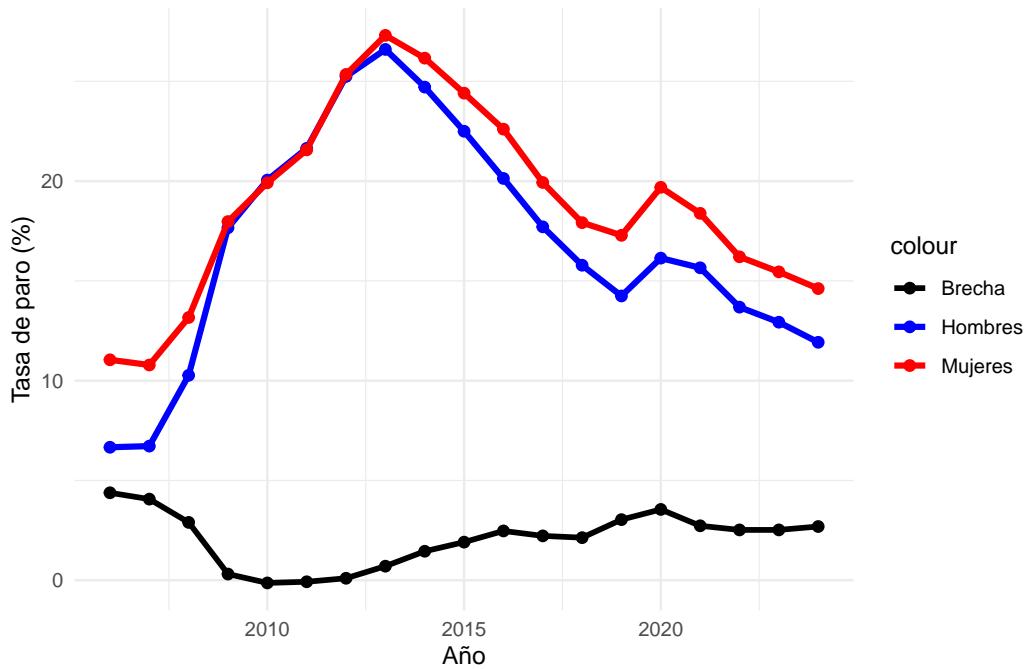


Sexo

- Hombres
- Mujeres

```
evo_tasa_y_brecha<-ggplot(brechas_total, aes(x = Periodo)) +geom_line(aes(y = tasa))
evo_tasa_y_brecha
```

Evolución de la tasa de paro total por sexo y brecha



El primer gráfico (evo_brecha_total) representa la evolución de la brecha de paro total tanto con líneas como con puntos para mostrar la tendencia y los valores anuales. En él se observa cómo la diferencia entre hombres y mujeres en la tasa de paro varía a lo largo del tiempo, con un aumento notable en los años posteriores a la crisis de 2008 y una ligera reducción en la última etapa analizada. Por su parte, el segundo gráfico (evo_tasa_total) muestra la evolución de la tasa de paro de hombres y mujeres de forma independiente, utilizando colores diferenciados para cada grupo. Esta visualización permite comparar directamente la evolución de ambos sexos, destacando la persistencia de una mayor tasa de paro entre las mujeres en prácticamente todos los años del periodo estudiado.

Por último, el tercer gráfico (evo_tasa_y_brecha) combina ambas perspectivas en una única figura. Se incluyen las líneas de hombres, mujeres y brecha, diferenciadas mediante colores específicos definidos con `scale_color_manual()`. Este enfoque conjunto facilita observar la relación entre las dos tasas y su diferencia a lo largo del tiempo, ofreciendo una visión global de la evolución de las desigualdades de género en el desempleo. En conjunto, el código y las representaciones resultantes permiten analizar de forma visual y comparativa la magnitud y la persistencia de la brecha de género en el mercado laboral español.

5.1. Evolucion de la tasa de paro y su brecha por grupo de edad

En este bloque de código se elaboran una serie de gráficos exploratorios con el objetivo de analizar la evolución de la brecha de paro entre mujeres y hombres desagregada por grupos de edad. A partir del conjunto de datos principal, se construye una tabla denominada `brecha_edades` que contiene la tasa de paro por sexo y edad, y se calcula la variable `brecha_paro` como la diferencia entre las tasas de mujeres y hombres. Posteriormente, los grupos de edad se clasifican en intervalos de diez años, y los datos se dividen en distintos subconjuntos (`df_1`, `df_2`, `df_3`) para facilitar la representación y mejorar la legibilidad de los gráficos.

```

brecha_edades <- df_clean %>%
  select(Sexo, Edad, Periodo, Tasa_paro) %>%
  pivot_wider(names_from = Sexo, values_from = Tasa_paro) %>%
  mutate(brecha_paro=Mujeres-Hombres)
brecha_edades

## # A tibble: 228 x 5
##   Edad Periodo Hombres Mujeres brecha_paro
##   <chr>  <dbl>    <dbl>    <dbl>      <dbl>
## 1 16-19  2024     37.9    44.4      6.43
## 2 16-19  2023     43.3    42.8     -0.493
## 3 16-19  2022     45.3    44.4     -0.912
## 4 16-19  2021     49.7    52.3      2.69
## 5 16-19  2020     50.1    61.0      10.9
## 6 16-19  2019     42.8    48.4      5.59
## 7 16-19  2018     49.7    49.7      0.0375
## 8 16-19  2017     54.2    55.1      0.919
## 9 16-19  2016     58.9    61.6      2.74
## 10 16-19 2015     64.7    70.7      6.04
## # i 218 more rows
## # i 218 more rows

unique(brecha_edades$Edad)

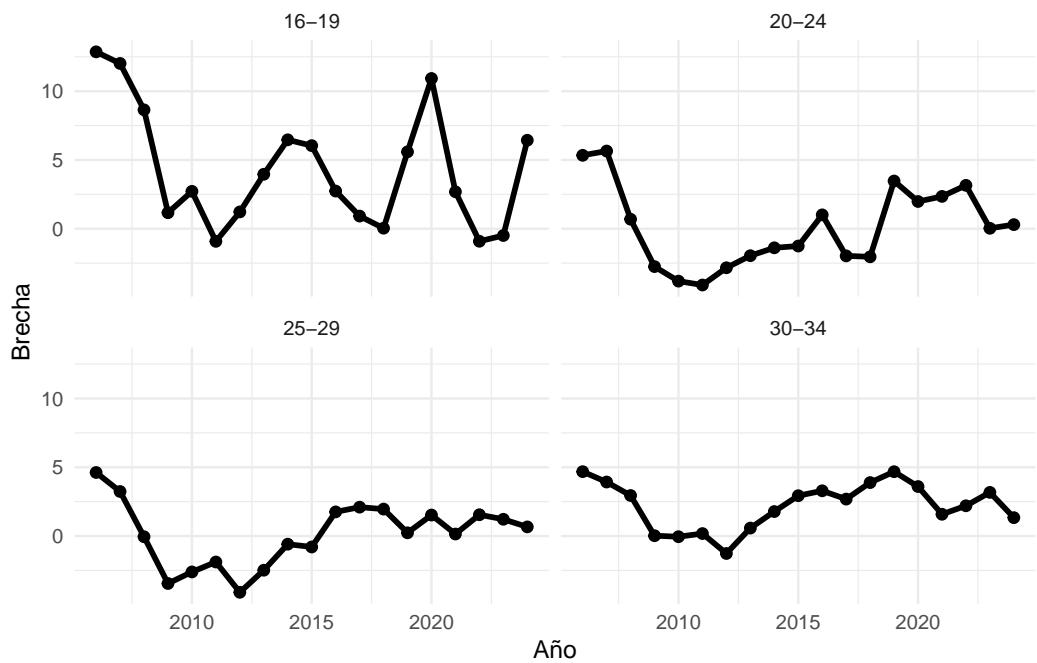
## [1] "16-19" "20-24" "25-29" "30-34" "35-39" "40-44" "45-49" "50-54" "55-59"
## [10] "60-64" "65-69" "70+"
library(ggplot2)
#Para hacerlo con un unico dataframe, es muy complejo de graficar, asi que es mejor
edades_1 <- c("16-19", "20-24", "25-29", "30-34")
edades_2 <- c("35-39", "40-44", "45-49", "50-54")
edades_3 <- c("55-59", "60-64", "65-69", "70+")

df_1 <- brecha_edades %>% filter(Edad %in% edades_1)
df_2 <- brecha_edades %>% filter(Edad %in% edades_2)
df_3 <- brecha_edades %>% filter(Edad %in% edades_3)

ggplot(df_1, aes(x = Periodo, y = brecha_paro)) +
  geom_line( size = 1.2) +
  geom_point(size = 2) +
  facet_wrap(~ Edad, ncol = 2) + # ← cuadrícula 2x2
  labs(title = "Evolución de la brecha de paro (Mujeres - Hombres)", x = "Año", y =
  theme_minimal())

```

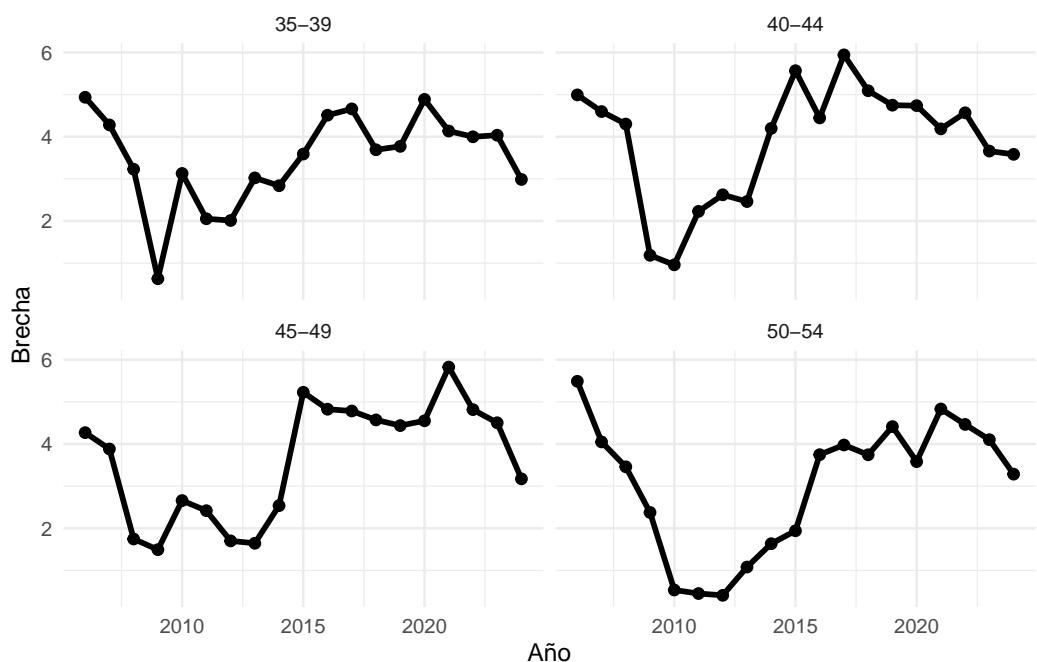
Evolución de la brecha de paro (Mujeres – Hombres)



250

```
ggplot(df_2, aes(x = Periodo, y = brecha_paro)) +
  geom_line( size = 1.2) +
  geom_point(size = 2) +
  facet_wrap(~ Edad, ncol = 2) +
  labs(title = "Evolución de la brecha de paro (Mujeres – Hombres)", x = "Año", y = theme_minimal())
```

Evolución de la brecha de paro (Mujeres – Hombres)

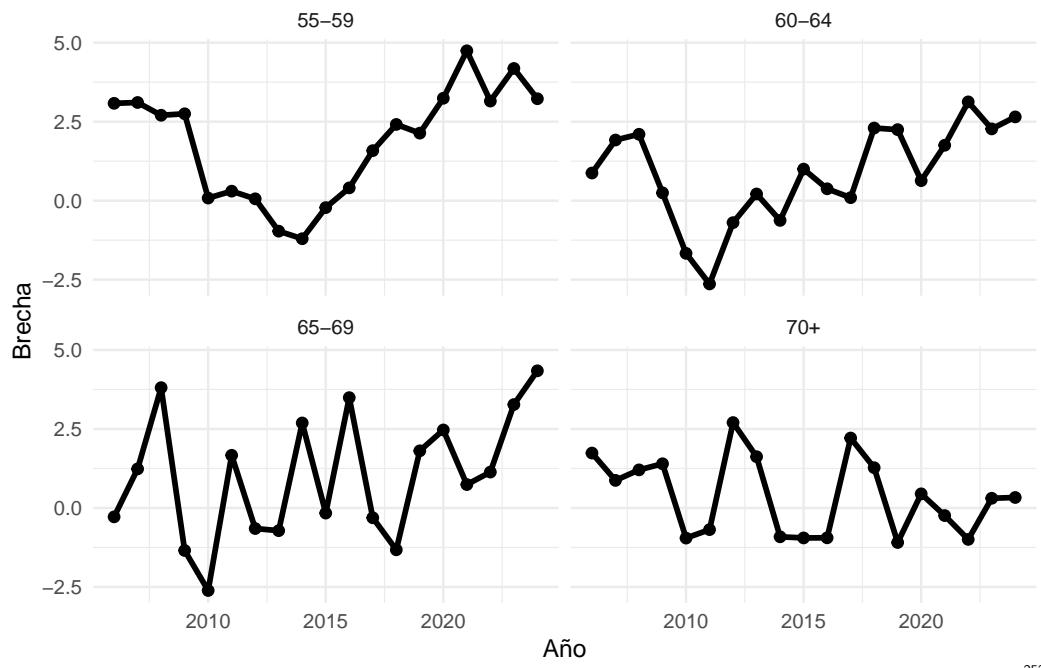


251

```
ggplot(df_3, aes(x = Periodo, y = brecha_paro)) +
  geom_line( size = 1.2) +
```

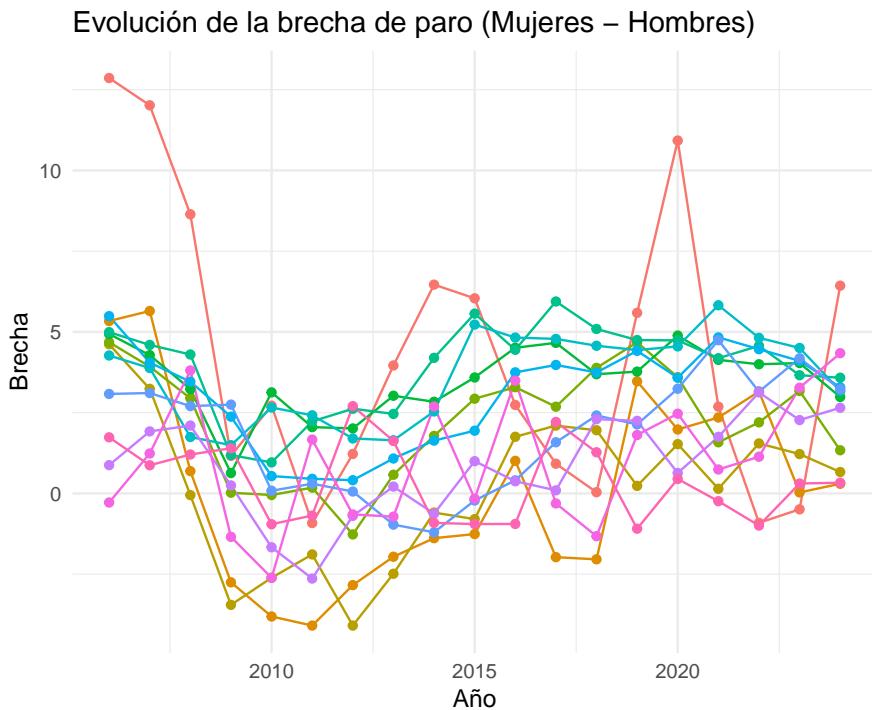
```
geom_point(size = 2) +
facet_wrap(~ Edad, ncol = 2) +
labs(title = "Evolución de la brecha de paro (Mujeres - Hombres)", x = "Año", y =
theme_minimal()
```

Evolución de la brecha de paro (Mujeres – Hombres)



252

```
colores_vivos <- c(
  "#FF4B00", "#00FFFF", "#FFD300", "#FF1493",
  "#00FF00", "#1E90FF", "#FF7F50", "#ADFF2F",
  "#DA70D6", "#FF4500", "#00CED1", "#FFFF66"
)
ggplot(brecha_edades, aes(x = Periodo, y = brecha_paro, colour = Edad)) +
  geom_line( size = 0.5) +
  geom_point() +
  labs(title = "Evolución de la brecha de paro (Mujeres - Hombres)", x = "Año", y =
  theme_minimal()
```



Los primeros gráficos utilizan `facet_wrap()` para mostrar la evolución temporal de la brecha de paro dentro de cada grupo de edad. Cada panel representa un intervalo etario y permite observar cómo la diferencia entre hombres y mujeres ha variado a lo largo del tiempo. Estas visualizaciones, evidencian que la brecha no se comporta de manera homogénea entre generaciones: mientras que en los grupos más jóvenes las diferencias tienden a fluctuar con mayor intensidad, en los grupos intermedios y de mayor edad la brecha presenta una evolución más estable, aunque persiste una desigualdad estructural en la mayoría de los casos.

Por último, se incluye una representación global en la que se superponen todas las series de edad en un mismo gráfico, utilizando una paleta de colores vivos definida manualmente. Este gráfico tiene un carácter eminentemente exploratorio y permite comparar visualmente la magnitud y la variabilidad de la brecha de paro entre cohortes. En conjunto, este bloque de código contribuye a profundizar en el análisis intergeneracional de la desigualdad laboral, revelando que las diferencias por género en el desempleo no afectan por igual a todos los grupos de edad y que ciertos tramos, especialmente los más jóvenes, muestran una mayor volatilidad a lo largo del tiempo.

5.2. Gráfica de Barras comparación grupos de edades con mayor brecha en 2006 y en 2024

En este bloque de código se elaboran una serie de gráficos exploratorios de barras destinados a analizar la evolución de la brecha de paro por grupos de edad a lo largo del tiempo. Para ello, a partir del conjunto `brecha_edades`, que contiene la diferencia entre las tasas de paro femeninas y masculinas, se genera un conjunto de gráficos individuales por año mediante la función `lapply()`. En cada uno de ellos, las barras representan la magnitud de la brecha en los distintos grupos de edad, mientras que la línea discontinua negra indica la media general de la brecha para ese año. De esta forma, el código permite observar en paralelo la estructura intergeneracional de las desigualdades laborales y cómo varía la media de brecha entre períodos.

```
a1<- brecha_edades %>% filter(Periodo %in% 2024)
```

```
library(ggplot2)
library(dplyr)
```

```
rango_y <- range(brecha_edades$brecha_paro)
años<- unique(df_clean$Periodo)

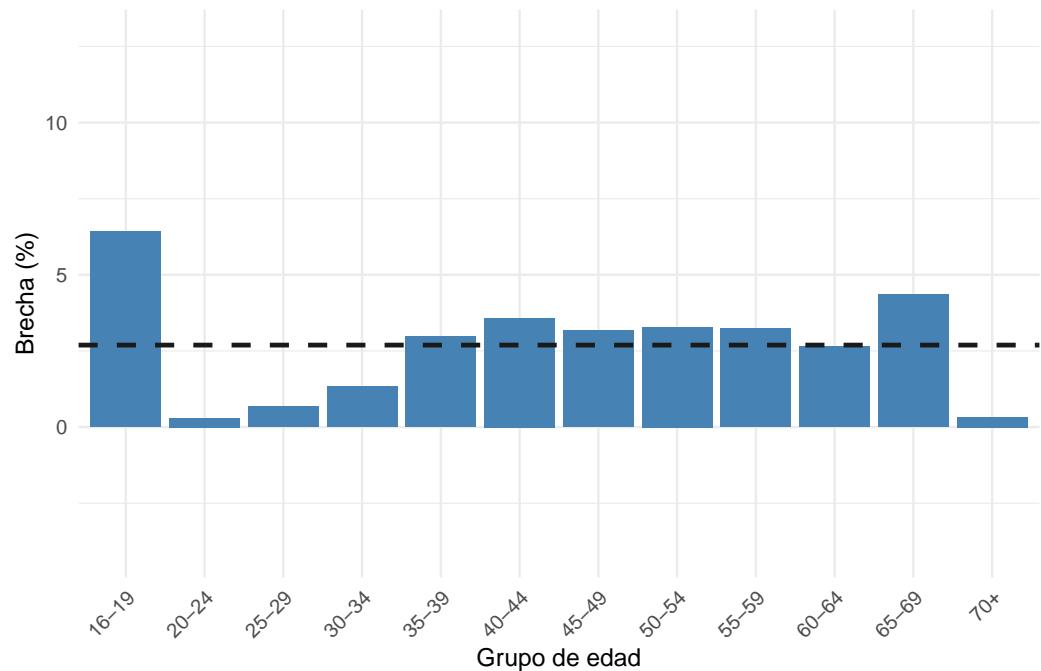
# Definir los años que quieres graficar
años <- c(2024, 2020, 2010, 2007, 2006)

# Calcular rango del eje Y (opcional, si ya lo tienes definido)
rango_y <- range(brecha_edades$brecha_paro, na.rm = TRUE)

# Generar gráficos solo para esos años
graficos_brecha_paro <- lapply(años, function(año) {
  brecha_edades %>%
    filter(Periodo == año) %>%
    ggplot(aes(x = factor(Edad, levels = unique(Edad)), y = brecha_paro)) +
    geom_col(fill = "steelblue") +
    ylim(rango_y) +
    labs(
      title = paste("Brecha de paro por grupo de edad - ", año),
      x = "Grupo de edad",
      y = "Brecha (%)"
    ) +
    geom_hline(
      yintercept = mean(brecha_edades$brecha_paro[brecha_edades$Periodo == año]),
      color = "grey10",
      linetype = "dashed",
      linewidth = 1
    ) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
})

graficos_brecha_paro

## [[1]]
```

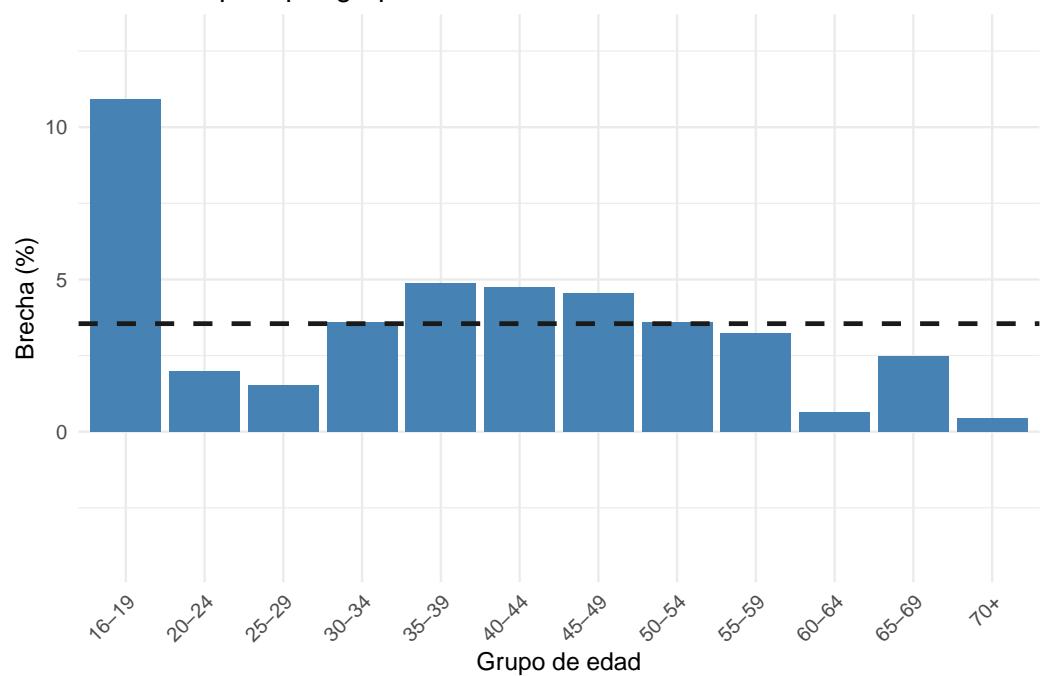
Brecha de paro por grupo de edad – 2024

281

```
##  
## [[2]]
```

282

283

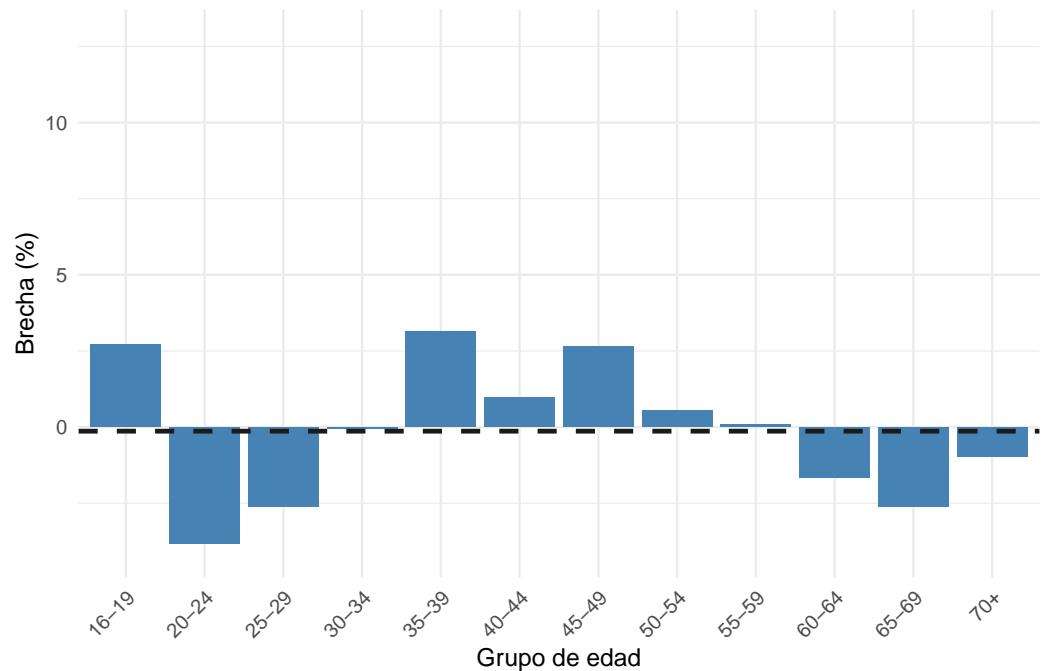
Brecha de paro por grupo de edad – 2020

284

```
##  
## [[3]]
```

285

286

Brecha de paro por grupo de edad – 2010

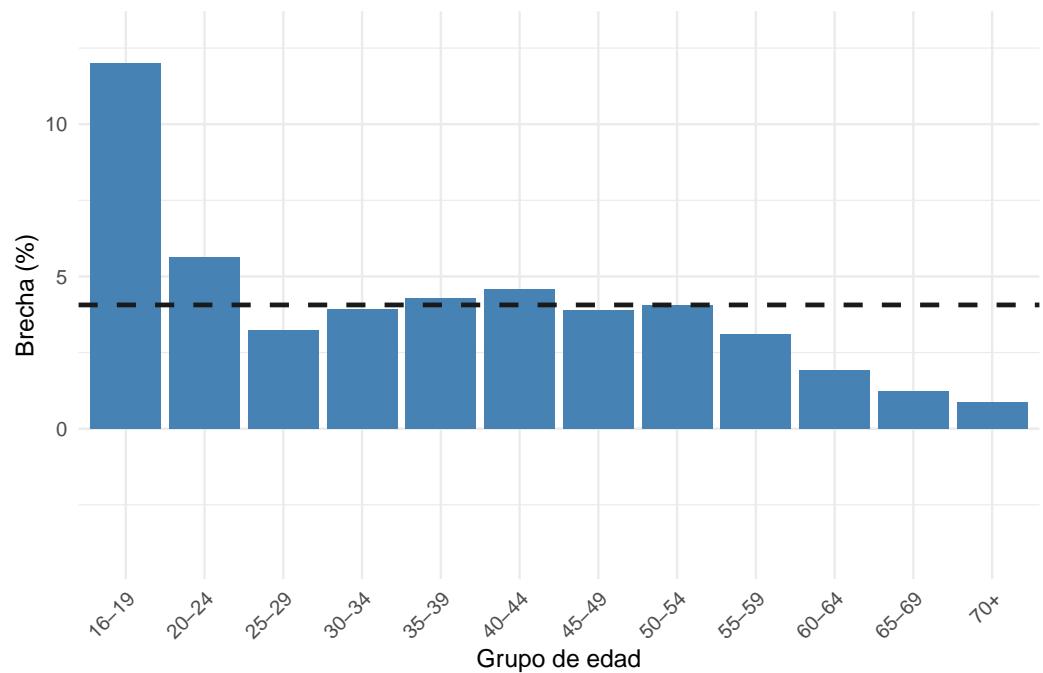
```
##
```

```
## [[4]]
```

287

288

289

Brecha de paro por grupo de edad – 2007

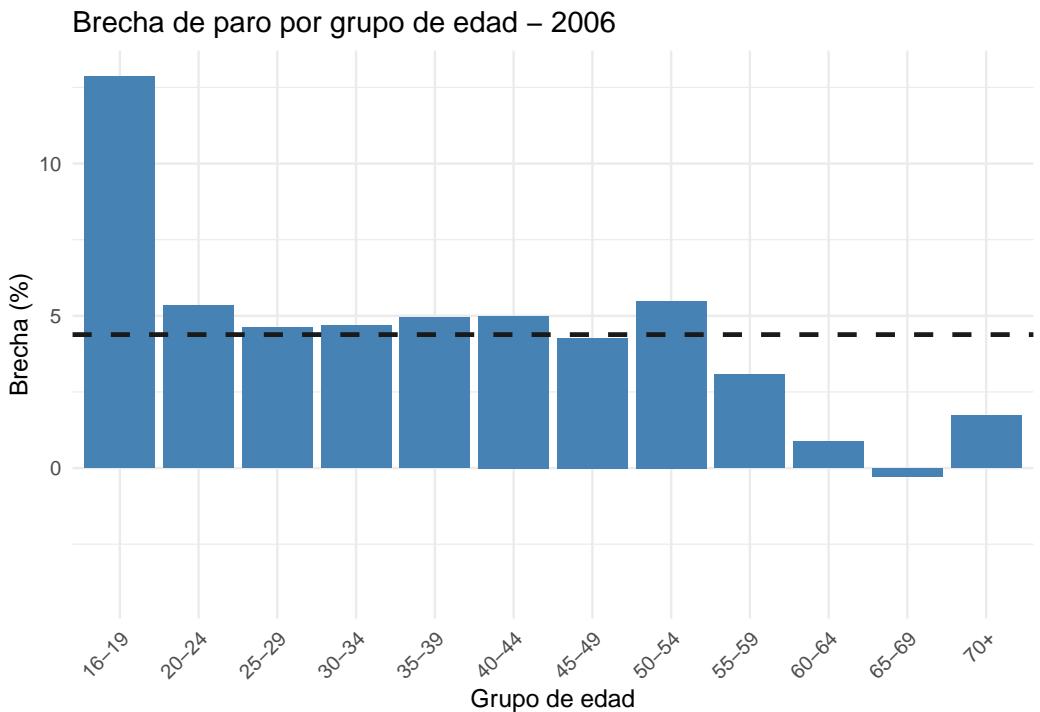
```
##
```

```
## [[5]]
```

290

291

292



Los gráficos resultantes muestran, de forma clara y comparativa, cómo la intensidad de la brecha de género en el desempleo varía según la edad y el momento temporal. En los primeros años analizados, la brecha es especialmente elevada entre los jóvenes, lo que evidencia una mayor vulnerabilidad del empleo femenino en los primeros tramos de edad. En cambio, en los años más recientes para esos tramos de edad, las diferencias tienden a moderarse y presentan un patrón más equilibrado entre grupos, aunque persisten desigualdades estructurales en los segmentos intermedios y de mayor edad.

En conjunto, este bloque busca ofrecer una visión desagregada y evolutiva de la brecha de paro por edad. Gracias a la representación mediante diagramas de barras, se facilita la comparación entre grupos generacionales y se identifican con claridad los grupos que más contribuyen a mantener las disparidades laborales entre hombres y mujeres a lo largo del tiempo.

Estadísticas

6. desv est y varianza

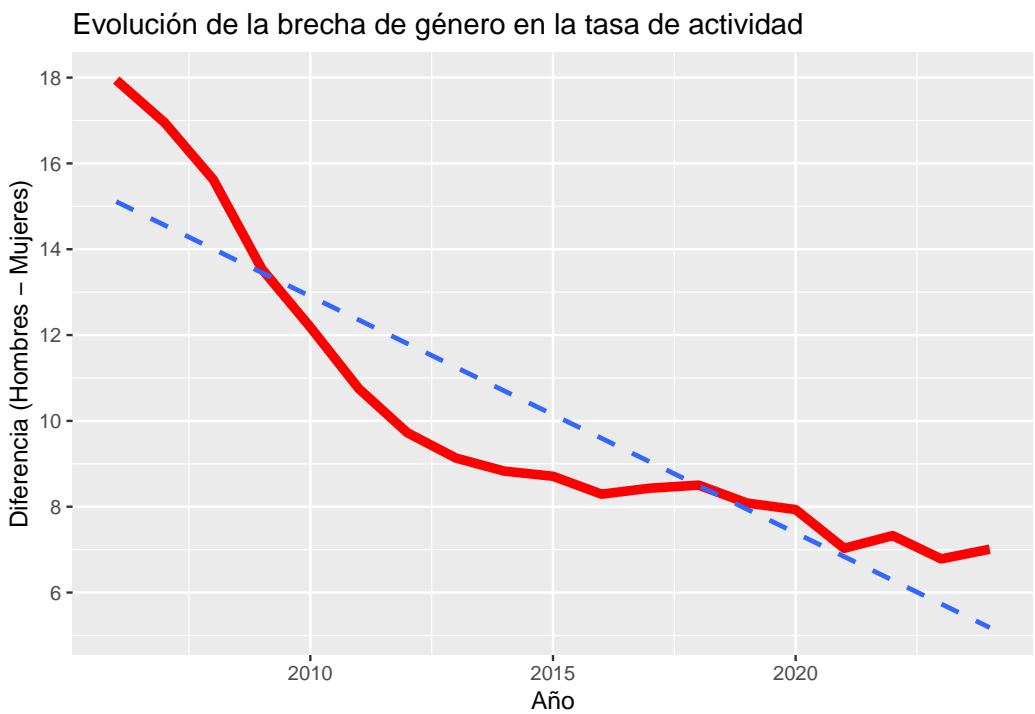
```
df_clean %>% group_by(Sexo) %>% summarise(sd_ocupados = sd(Ocupados, na.rm = TRUE)

## # A tibble: 2 x 3
##   Sexo     sd_ocupados var_ocupados
##   <chr>        <dbl>       <dbl>
## 1 Hombres      591.     349261.
## 2 Mujeres      490.     240316.

df_clean %>%
  group_by(Sexo) %>%
  summarise(media_ocupados = mean(Ocupados, na.rm = TRUE),
            media_inactivos = mean(Inactivos, na.rm = TRUE)) %>%
  summarise(brecha_ocup = diff(media_ocupados),
            brecha_inact = diff(media_inactivos))

## # A tibble: 1 x 2
##   brecha_ocup brecha_inact
```

```
##           <dbl>           <dbl>
## 1       -166.          247.          315
## 
## df_brecha_genero <- df_clean %>% group_by(Periodo, Sexo) %>% summarise(tasa = mean
## df_brecha_genero
## 
## # # A tibble: 19 x 4
##   Periodo Hombres Mujeres Brecha
##   <dbl>    <dbl>    <dbl>    <dbl> 317
## 1 2006     66.1    48.2    17.9 318
## 2 2007     66.2    49.2    16.9 319
## 3 2008     66.4    50.7    15.6 320
## 4 2009     65.4    51.8    13.5 321
## 5 2010     64.9    52.7    12.2 322
## 6 2011     64.3    53.6    10.7 323
## 7 2012     64.1    54.3    9.72 324
## 8 2013     63.8    54.6    9.13 325
## 9 2014     63.5    54.7    8.83 326
## 10 2015    63.8    55.1    8.71 327
## 11 2016    63.7    55.4    8.29 328
## 12 2017    63.8    55.4    8.43 329
## 13 2018    64.0    55.5    8.50 330
## 14 2019    64.1    56.0    8.09 331
## 15 2020    62.9    55.0    7.93 332
## 16 2021    64.0    56.9    7.03 333
## 17 2022    64.7    57.3    7.33 334
## 18 2023    65.0    58.3    6.78 335
## 19 2024    65.5    58.5    7.01 336
## 
## ggplot(df_brecha_genero, aes(x = Periodo, y = Brecha)) +
##   geom_line(color = "red", linewidth = 2) +
##   geom_smooth(method = "lm", se = FALSE, linetype = "dashed") + scale_y_continuous
##   labs(title = "Evolución de la brecha de género en la tasa de actividad",
##        x = "Año", y = "Diferencia (Hombres - Mujeres)") 337
## 
## ## `geom_smooth()` using formula = 'y ~ x' 338
```



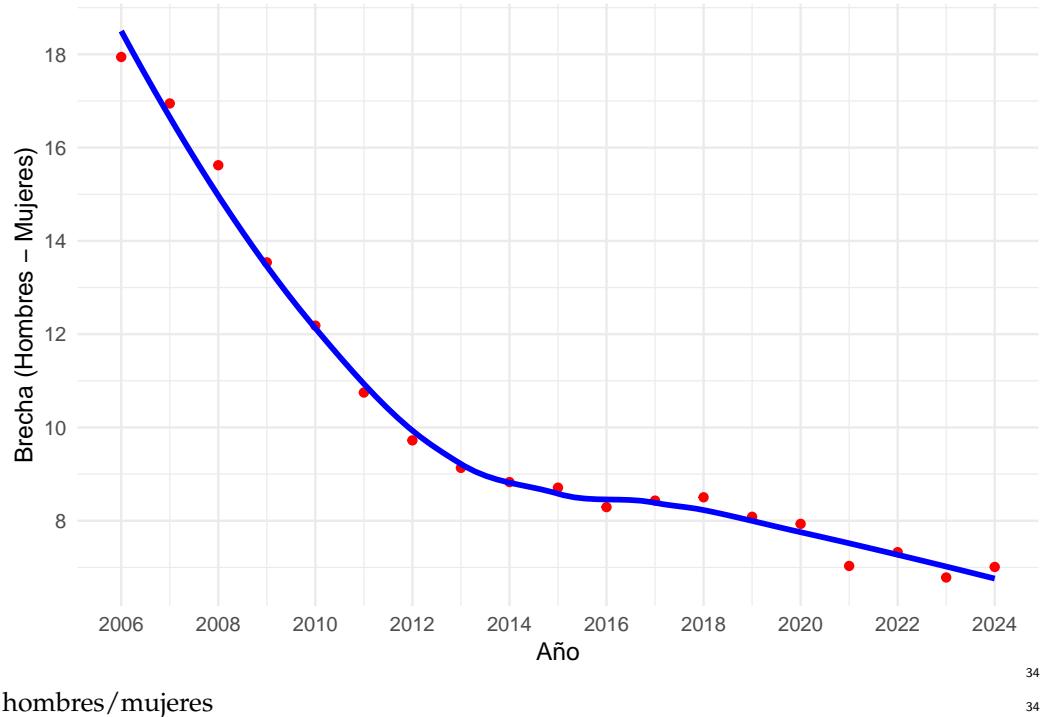
340

```
ggplot(df_brecha_genero, aes(x = Periodo, y = Brecha)) +
  geom_point(color = "red") +
  geom_smooth(method = "loess", se = FALSE, color = "blue", linewidth = 1.2) +
  scale_y_continuous(limits = c(6, 18), labels =
    c("6", "8", "10", "12", "14", "16", "18"))
  labs(
    title = "Interpolación de la brecha de género NO LINEAL",
    x = "Año",
    y = "Brecha (Hombres - Mujeres)"
  ) +
  theme_minimal()
```

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

341

Interpolación de la brecha de género NO LINEAL



```
brecha_relativa <- df_clean %>%
  group_by(Periodo, Sexo) %>%
  summarise(Tasa = mean(Tasa_actividad, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = Sexo, values_from = Tasa) %>%
  mutate(Razon = Hombres / Mujeres)
```

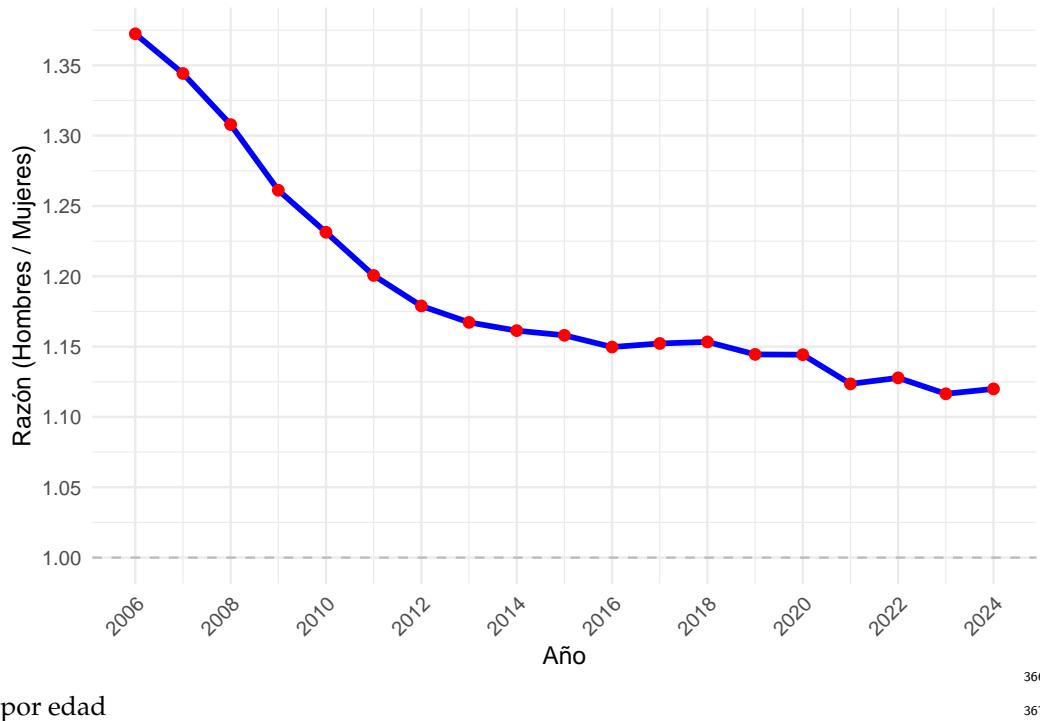
```
brecha_relativa
```

```
## # A tibble: 19 x 4
##   Periodo Hombres Mujeres Razon
##       <dbl>    <dbl>    <dbl>    <dbl>
## 1     2006    66.1    48.2    1.37
## 2     2007    66.2    49.2    1.34
## 3     2008    66.4    50.7    1.31
## 4     2009    65.4    51.8    1.26
## 5     2010    64.9    52.7    1.23
## 6     2011    64.3    53.6    1.20
## 7     2012    64.1    54.3    1.18
## 8     2013    63.8    54.6    1.17
## 9     2014    63.5    54.7    1.16
## 10    2015    63.8    55.1    1.16
## 11    2016    63.7    55.4    1.15
## 12    2017    63.8    55.4    1.15
## 13    2018    64.0    55.5    1.15
## 14    2019    64.1    56.0    1.14
## 15    2020    62.9    55.0    1.14
## 16    2021    64.0    56.9    1.12
## 17    2022    64.7    57.3    1.13
## 18    2023    65.0    58.3    1.12
## 19    2024    65.5    58.5    1.12
```

```
#j
#l

ggplot(brecha_relativa, aes(x = Periodo, y = Razon)) +
  geom_line(color = "blue", linewidth = 1.2) +
  geom_point(color = "red", size = 2) +
  geom_hline(yintercept = 1, linetype = "dashed", color = "gray") + # línea de igualdad
  scale_y_continuous(breaks = seq(1, 1.5, by = 0.05)) + # marcas cada 0.05
  scale_x_continuous(breaks = seq(2006, 2024, by = 2)) + # años cada 2
  labs(
    title = "Evolución de la razón Hombres/Mujeres",
    x = "Año",
    y = "Razón (Hombres / Mujeres)"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(face = "bold")
  )
)
```

Evolución de la razón Hombres/Mujeres



7. Distribución de la población en el mercado laboral

7.1. Composición del mercado laboral

En este bloque de código se genera un gráfico exploratorio con el propósito de comprobar la visualización de la composición del mercado laboral antes de automatizar el proceso para todos los años. Para ello, se filtran los datos del año 2024 y se seleccionan las variables relevantes relacionadas con el estado laboral (*Activos*, *Ocupados*, *Parados e Inactivos*), junto con la edad y el sexo. Posteriormente, mediante la función `pivot_longer()`, se reorganiza la tabla a formato largo, lo que permite representar las distintas categorías laborales como valores dentro de una misma variable (*Estado*), facilitando así la creación de un gráfico apilado en `ggplot2`.

366

367

368

369

370

371

372

373

374

375

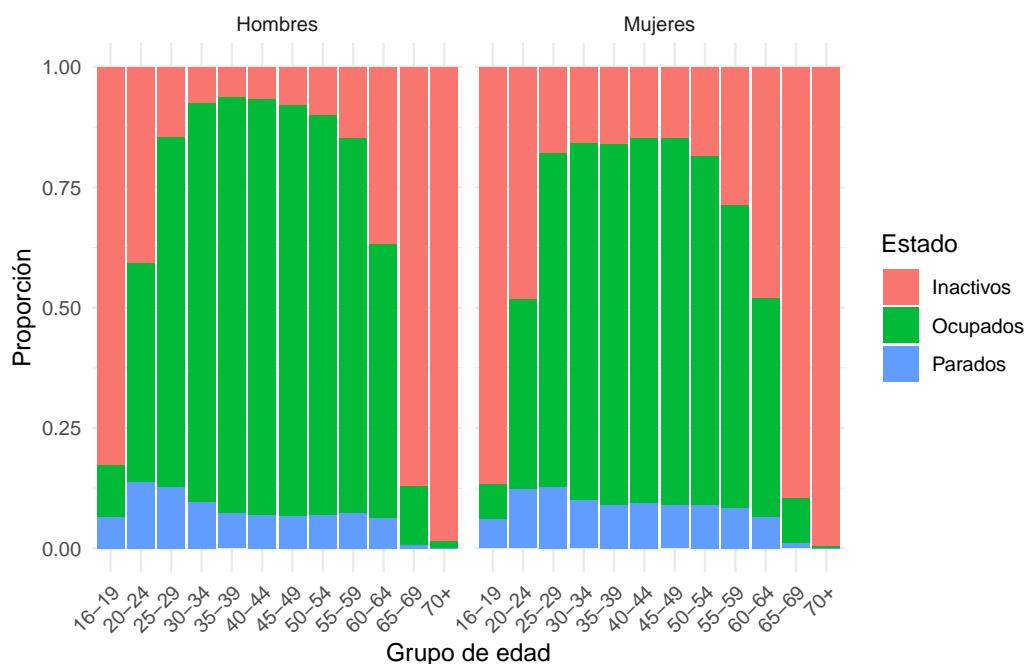
376

377

El gráfico resultante muestra la distribución proporcional de la población según su situación laboral y grupo de edad, diferenciando entre hombres y mujeres mediante `facet_wrap(~Sexo)`. Se emplea `geom_col(position = "fill")` para que las barras representen proporciones y no valores absolutos, y se aplican ajustes estéticos para mejorar la legibilidad, como la rotación de etiquetas del eje X y el uso de `theme_minimal()`. Este gráfico sirvió como prueba inicial para evaluar la claridad y estructura de la visualización antes de desarrollar la función que posteriormente generó automáticamente el mismo tipo de gráfico para todos los años del periodo analizado.

```
378 datos_composicion_2024 <- df_clean %>%
379   filter(Periodo==2024) %>%
380   select(Periodo, Edad, Sexo, Activos, Ocupados, Parados, Inactivos) %>%
381   pivot_longer(cols = c(Inactivos, Ocupados, Parados), names_to = "Estado", values_to =
382
383 datos_composicion_2024 %>%
384   ggplot(aes(x = Edad, y = N_Personas, fill = Estado)) +
385     geom_col(position = "fill") +
386     labs(
387       title = "Composición del mercado laboral por grupo de edad en 2024",
388       x = "Grupo de edad",
389       y = "Proporción",
390       fill = "Estado"
391     ) +
392     theme_minimal() +
393     theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
394     facet_wrap(~Sexo)
```

Composición del mercado laboral por grupo de edad en 2024



Tras la comprobación inicial con los datos de 2024, se elaboró este bloque de código para automatizar la generación del mismo tipo de gráfico para todos los años del conjunto de datos. Para ello, se utiliza la función `lapply()`, que aplica el mismo procedimiento de visualización a cada año contenido en la variable `años`. Dentro de la función, los datos se filtran por periodo y se representa la composición del mercado laboral mediante

geom_col(position = "fill"), mostrando la proporción de cada estado laboral (ocupados, parados e inactivos) dentro de cada grupo de edad.

El código utiliza facet_wrap(~Sexo) para mostrar en paneles separados la información correspondiente a hombres y mujeres. El título de cada gráfico se personaliza dinámicamente con el año correspondiente mediante paste(). Este procedimiento permitió obtener una serie de gráficos anuales, facilitando así el análisis comparativo de la evolución de la estructura del mercado laboral a lo largo del tiempo y entre sexos.

```

datos_composicion<- df_clean %>%
  select(Periodo,Edad,Sexo,Activos,Ocupados,Parados,Inactivos) %>%
  pivot_longer(cols = c(Inactivos,Ocupados,Parados),names_to = "Estado",values_to =
  años <- c(2024, 2020, 2010, 2007, 2006)

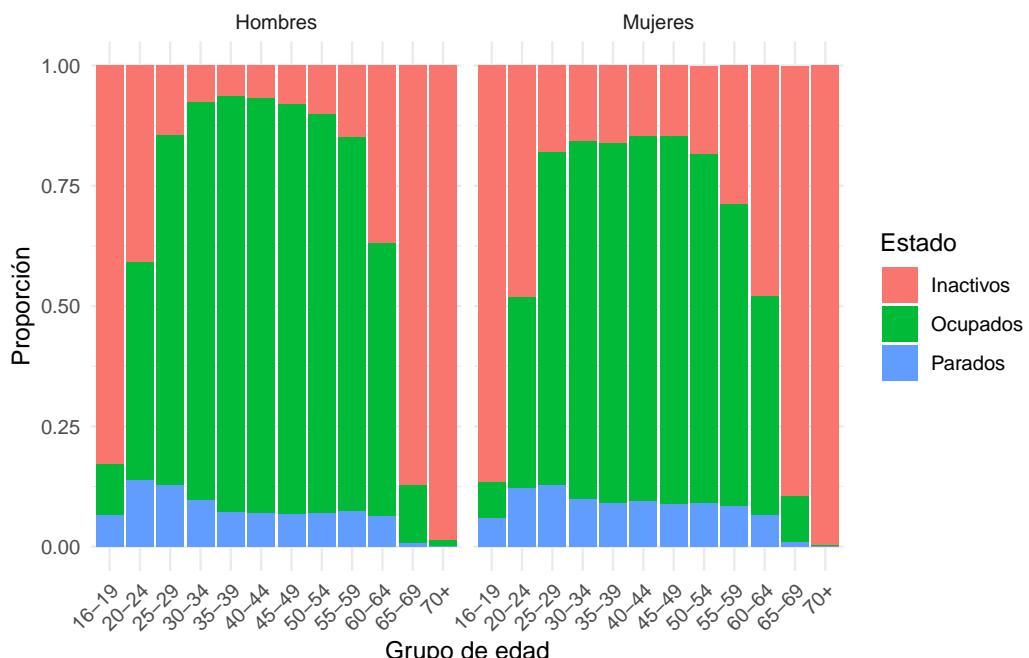
graficos_composicion_por_año <- lapply(años, function(año) {
  datos_composicion %>%
    filter(Periodo == año) %>%
    ggplot(aes(x = factor(Edad, levels = unique(Edad)), y = N_Personas, fill = Estado)) +
    geom_col(position = "fill") +
    labs(
      title = paste("Composición del mercado laboral por grupo de edad - ", año),
      x = "Grupo de edad",
      y = "Proporción",
      fill = "Estado"
    ) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    facet_wrap(~Sexo)#Añadir o quitar según queramos comparar hombres y mujeres o
  })

```

graficos_composicion_por_año

[[1]]

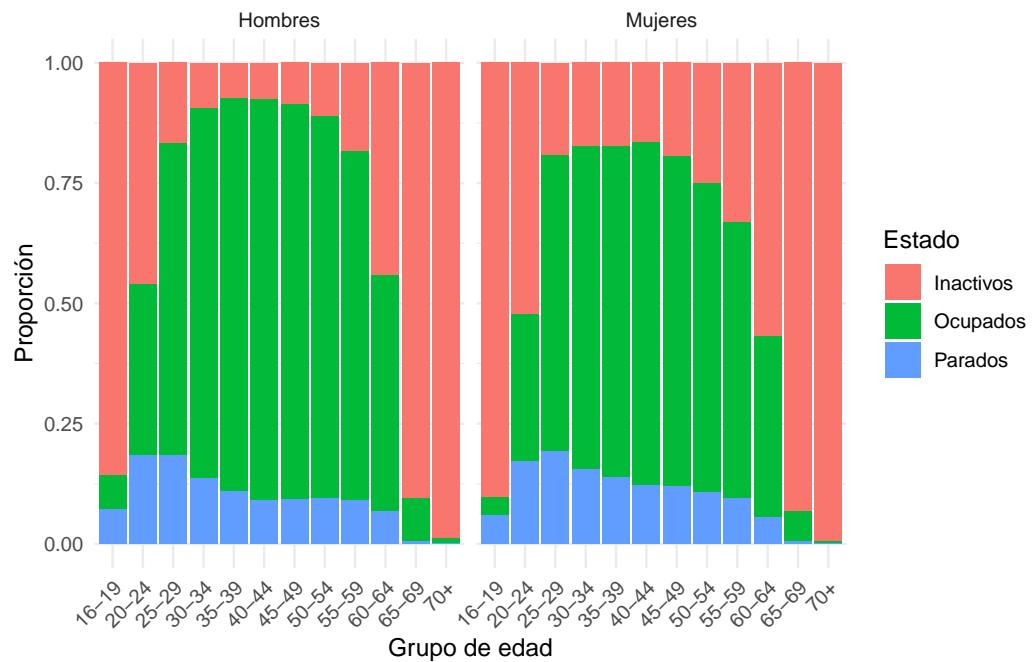
Composición del mercado laboral por grupo de edad – 2024



```
##  
## [[2]]
```

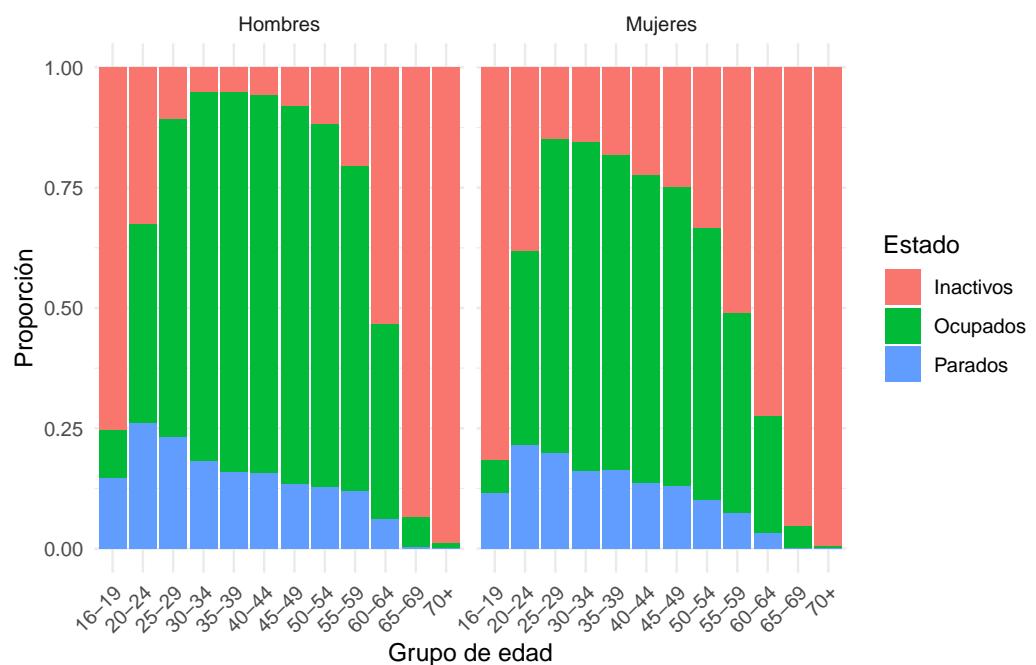
401
402

Composición del mercado laboral por grupo de edad – 2020

403
404
405

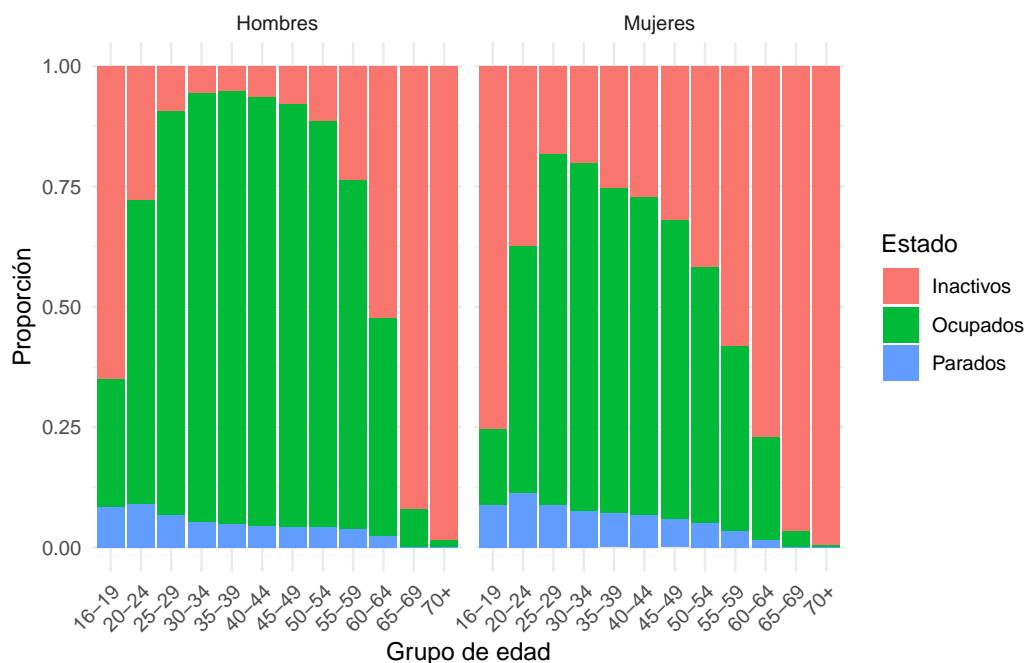
```
##  
## [[3]]
```

Composición del mercado laboral por grupo de edad – 2010

406
407
408

```
##  
## [[4]]
```

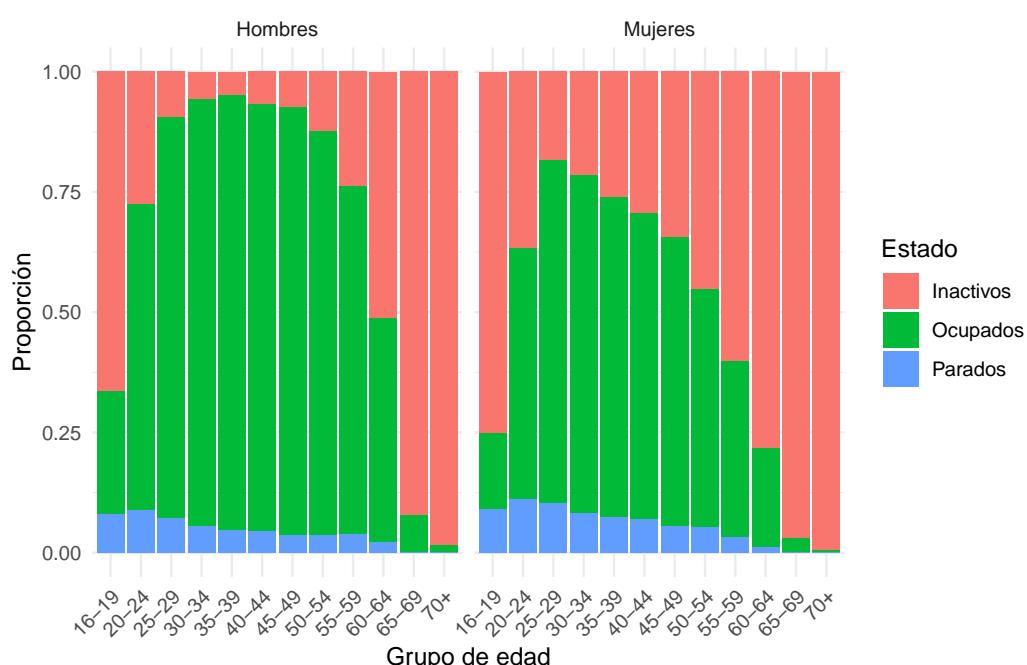
Composición del mercado laboral por grupo de edad – 2007



##

[[5]]

Composición del mercado laboral por grupo de edad – 2006



Del análisis de los gráficos se observa una estructura bastante estable del mercado laboral por edad y sexo a lo largo del tiempo, con ligeras variaciones asociadas a los ciclos económicos. En ambos sexos, el grupo central de edad (30-54 años) concentra la mayor proporción de personas ocupadas, mientras que los extremos —jóvenes y mayores— presentan una mayor presencia de inactivos. Las mujeres muestran sistemáticamente una menor proporción de ocupación y una mayor de inactividad en comparación con los hombres, aunque esta diferencia tiende a reducirse en los años más recientes. Asimismo, la proporción de personas paradas es más significativa en los grupos jóvenes, reflejando una

413

414

415

416

417

418

419

420

mayor vulnerabilidad del empleo juvenil. En conjunto, los gráficos ponen de manifiesto que, pese a los avances en participación femenina y la estabilidad general del empleo en edades medias, persisten desigualdades de género e importantes diferencias intergeneracionales dentro del mercado laboral.

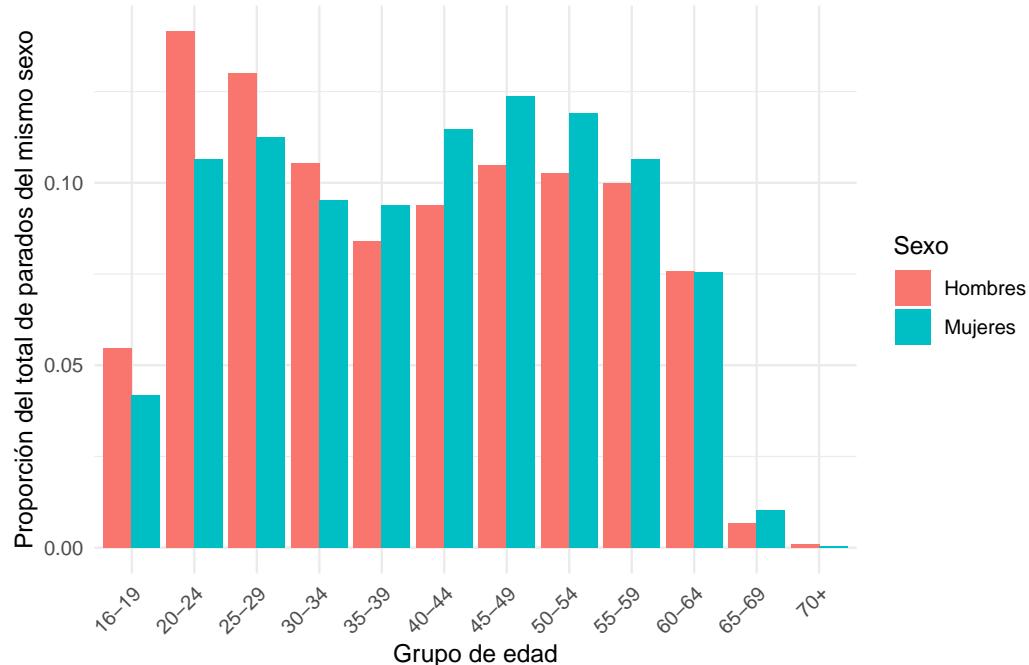
##Proporción de parados En este bloque de código se genera una serie de gráficos que representan la distribución del paro dentro de cada sexo por grupo de edad y año. Para ello, se utiliza una estructura, que permite crear automáticamente un gráfico por cada año disponible en el conjunto de datos. Primero, los datos se filtran por periodo y se agrupan por sexo y año, tras lo cual se calcula la proporción que representa cada grupo de edad dentro del total de personas paradas del mismo sexo (`Parados / sum(Parados)`).

A continuación, se construyen los gráficos con `ggplot2`, empleando `geom_col(position = "dodge")` para comparar visualmente la distribución entre hombres y mujeres dentro de cada grupo de edad. El objetivo de este análisis es observar cómo se reparte el desempleo dentro de cada sexo a lo largo del tiempo, identificando posibles diferencias estructurales entre hombres y mujeres en función de la edad. De esta manera, se obtiene una visión más detallada del peso relativo de cada grupo de edad en el total de personas paradas de cada sexo y su evolución temporal.

```
graficos_parados_proporcion_por_año<- lapply(años, function(año){
  df_clean %>%
    filter(Periodo==año) %>%
    group_by(Sexo, Periodo) %>%
    mutate(Proporcion_parados_en_sexo = Parados / sum(Parados)) %>%
    ungroup() %>%
    ggplot(aes(x = factor(Edad, levels = unique(Edad)), y = Proporcion_parados_en_sexo,
               geom_col(position = "dodge") +
               labs(
                 title = paste("Distribución del paro dentro de cada sexo - ",año),
                 x = "Grupo de edad",
                 y = "Proporción del total de parados del mismo sexo",
                 fill = "Sexo"
               ) +
               theme_minimal() +
               theme(axis.text.x = element_text(angle = 45, hjust = 1))
  })
  graficos_parados_proporcion_por_año

## [[1]]
```

Distribución del paro dentro de cada sexo – 2024



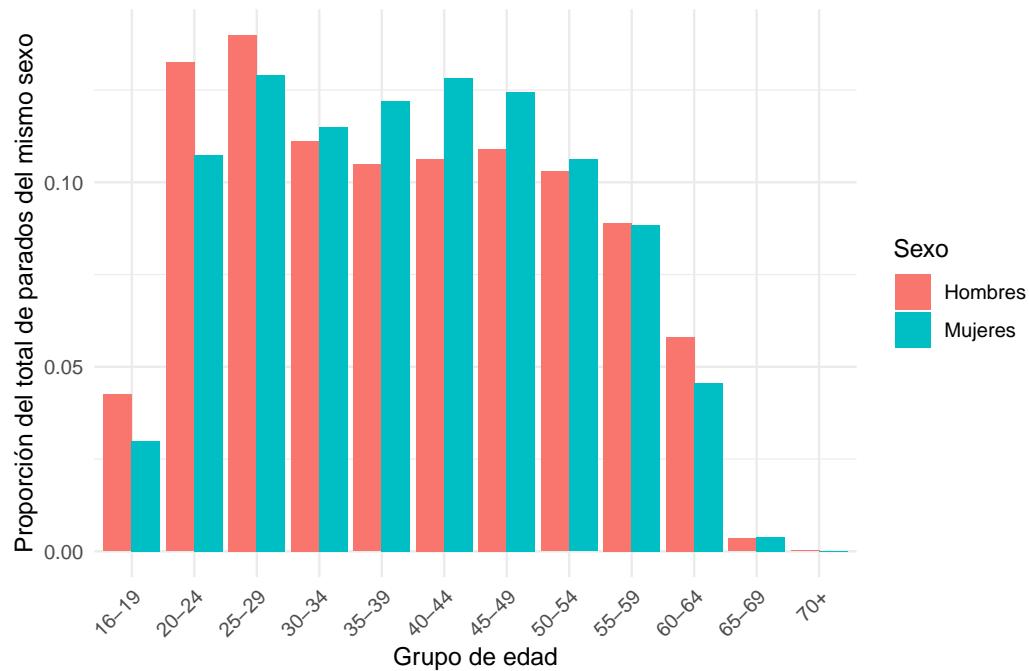
439

```
##  
## [[2]]
```

440

441

Distribución del paro dentro de cada sexo – 2020



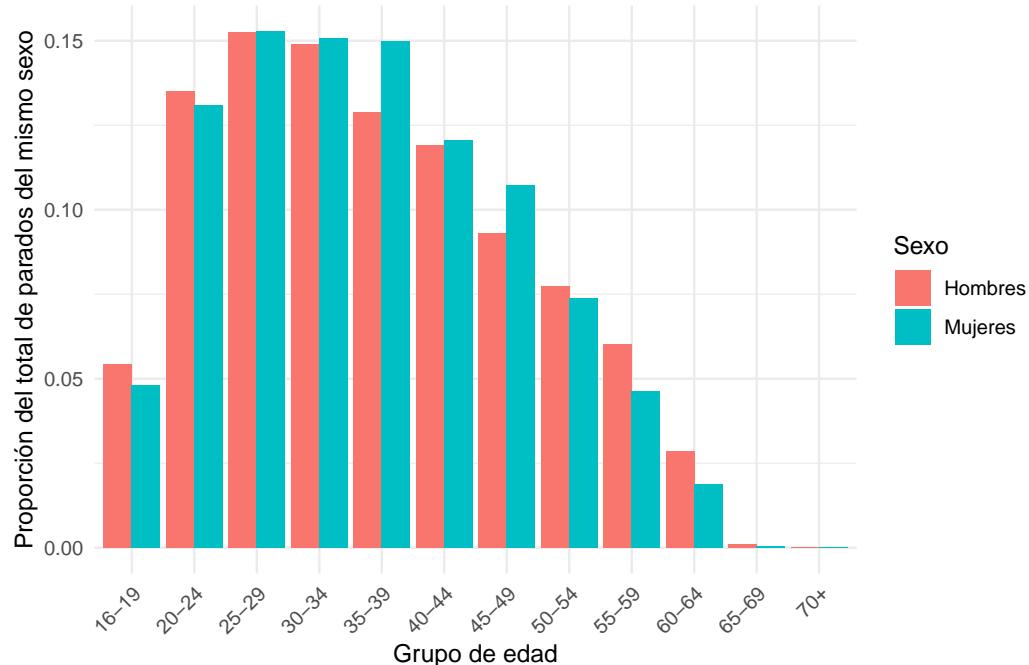
442

```
##  
## [[3]]
```

443

444

Distribución del paro dentro de cada sexo – 2010



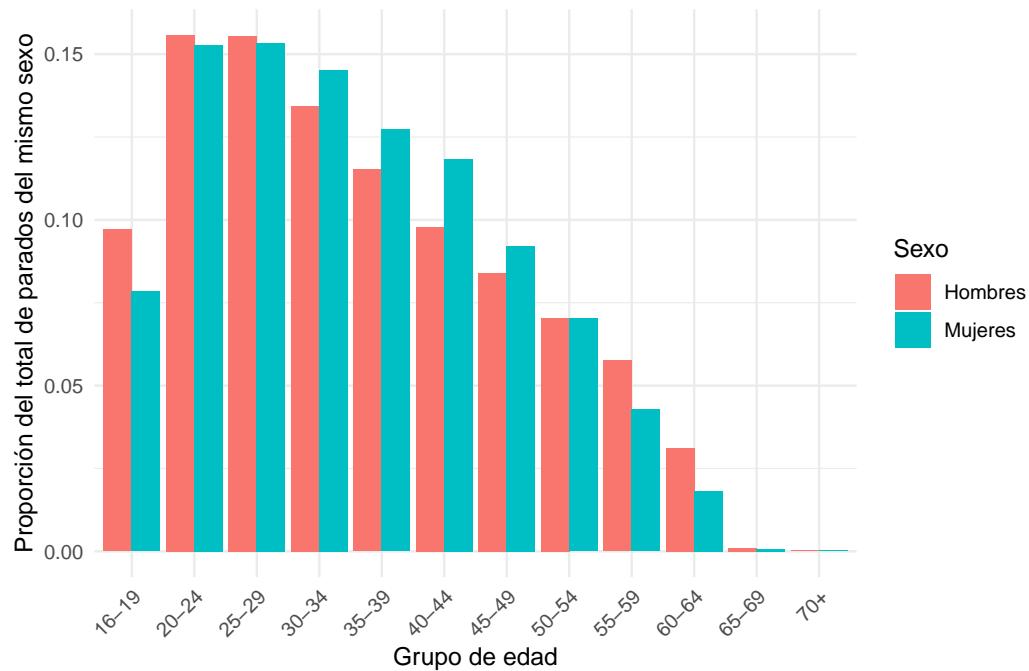
445

```
##  
## [[4]]
```

446

447

Distribución del paro dentro de cada sexo – 2007



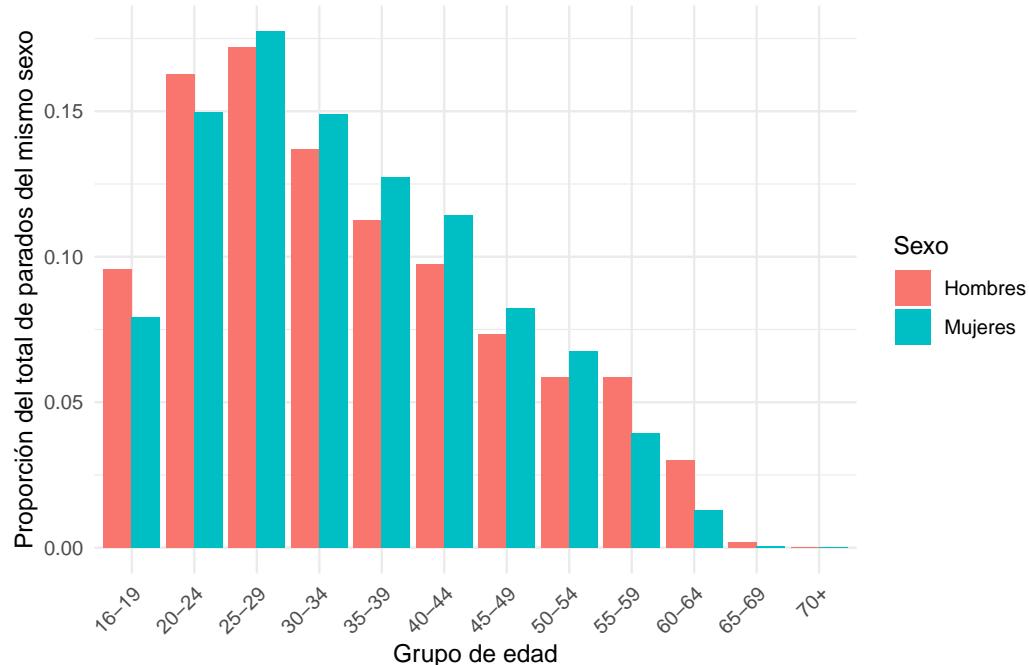
448

```
##  
## [[5]]
```

449

450

Distribución del paro dentro de cada sexo – 2006



En estas gráficas se aprecia cómo la distribución del paro dentro de cada sexo ha cambiado en los últimos años, influida por transformaciones estructurales del mercado laboral y demográficas. La elevación progresiva de la edad de jubilación y el retraso en la salida definitiva del mercado laboral han provocado un aumento en la proporción de personas activas en edades más avanzadas, lo que se traduce también en una mayor representación del paro entre los grupos de 55 años o más. Esto indica que el desempleo ya no es un fenómeno concentrado exclusivamente en edades intermedias, sino que se extiende también hacia los trabajadores de mayor edad, quienes permanecen más tiempo en el mercado laboral. Por el contrario, los grupos más jóvenes (de 16 a 24 años) han reducido su presencia relativa en el total de personas desempleadas, debido principalmente a una incorporación más tardía al mercado laboral como resultado del alargamiento de la etapa educativa y del aumento de la formación superior.

En cuanto al análisis por género, se observa que la distribución del paro dentro de cada sexo mantiene ciertas similitudes, pero también muestra matices importantes. En los grupos de edad intermedia (25-54 años), las mujeres tienden a concentrar una mayor proporción del desempleo dentro de su propio sexo, lo que refleja una persistente desigualdad estructural en el acceso y la estabilidad del empleo femenino, especialmente en etapas donde la conciliación familiar y laboral juega un papel determinante. Sin embargo, entre los hombres, el paro se distribuye de forma algo más homogénea entre los distintos grupos de edad, con un incremento más visible en las edades avanzadas, vinculado al impacto de los sectores más cíclicos (como la construcción o la industria) y a la dificultad de reinserción laboral tras la pérdida del empleo.

En conjunto, estas gráficas ponen de manifiesto un proceso de envejecimiento del desempleo y una convergencia progresiva entre hombres y mujeres en la estructura interna del paro, aunque todavía persisten diferencias derivadas de los roles laborales y sociales tradicionales. Mientras los hombres mayores ganan peso dentro del paro total por el retraso en la jubilación, las mujeres siguen mostrando una mayor vulnerabilidad en las edades centrales, reflejando las desigualdades de género que aún caracterizan el mercado laboral español.

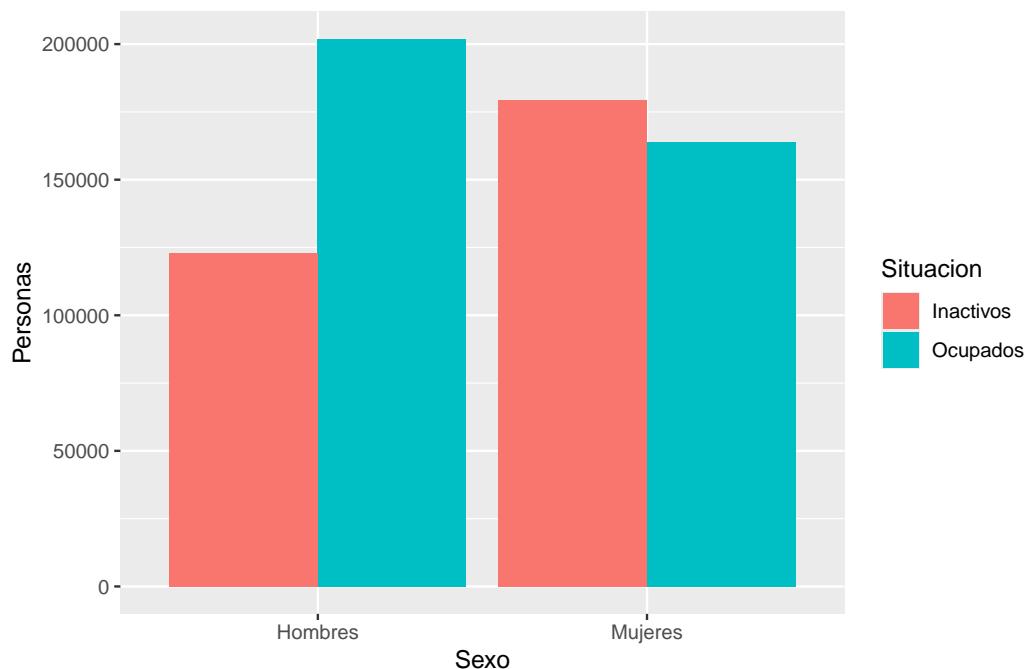
#Distribución de ocupados e inactivos por sexo

```

df_ocup_inact <- df_clean %>%
  group_by(Sexo) %>%
  summarise(
    Ocupados = sum(Ocupados, na.rm = TRUE),
    Inactivos = sum(Inactivos, na.rm = TRUE)
  ) %>%
  pivot_longer(cols = c(Ocupados, Inactivos), names_to = "Situacion", values_to =
  Personas)

ggplot(df_ocup_inact, aes(x = Sexo, y = Personas, fill = Situacion)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Distribución de ocupados e inactivos por sexo", x = "Sexo", y = "Personas")
  
```

Distribución de ocupados e inactivos por sexo



#act inact

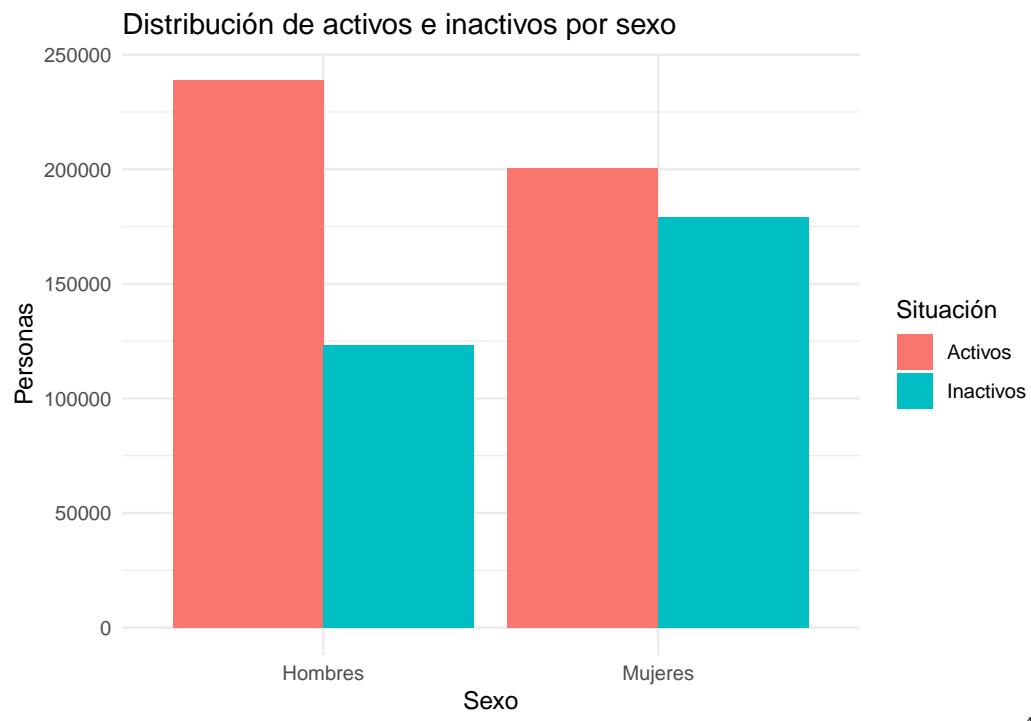
```

df_act_inact <- df_clean %>%
  group_by(Sexo) %>%
  summarise(
    Activos = sum(Activos, na.rm = TRUE),
    Inactivos = sum(Inactivos, na.rm = TRUE)
  ) %>%
  pivot_longer(cols = c(Activos, Inactivos), names_to = "Situacion", values_to =
  Personas)

ggplot(df_act_inact, aes(x = Sexo, y = Personas, fill = Situacion)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Distribución de activos e inactivos por sexo",
    x = "Sexo",
    y = "Personas",
    fill = "Situación"
  ) +
  theme_minimal()
  
```

482

483



#grafico violin

```

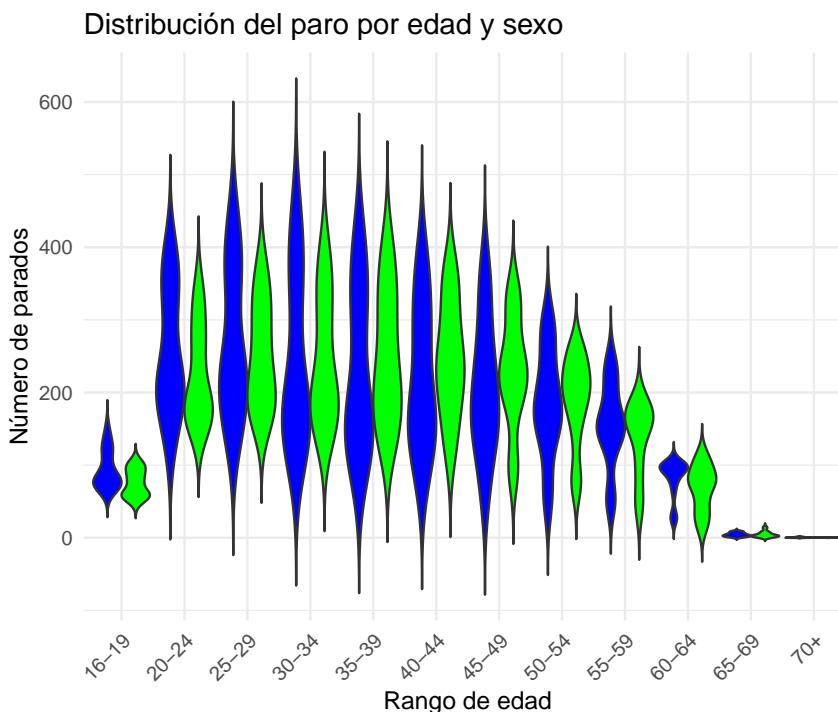
library(dplyr)
library(ggplot2)

# Preparar los datos: convertir Edad a factor con orden lógico
df_clean <- df_clean %>%
  mutate(Edad = factor(Edad, levels = unique(Edad)))

# Crear gráfico de violín
ggplot(df_clean, aes(x = Edad, y = Parados, fill = Sexo)) +
  geom_violin(trim = FALSE, scale = "width") +
  labs(
    title = "Distribución del paro por edad y sexo",
    x = "Rango de edad",
    y = "Número de parados",
    fill = "Sexo"
  ) +
  theme_minimal() +
  scale_fill_manual(values = c("Hombres" = "blue", "Mujeres" = "green")) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
  
```

484

485



#feminizacion

```
df_feminizacion <- df_clean %>%
  group_by(Periodo, Edad, Sexo) %>%
  summarise(Ocupados = sum(Ocupados), .groups = "drop") %>%
  pivot_wider(names_from = Sexo, values_from = Ocupados) %>%
  mutate(Indice_feminizacion = Mujeres / (Hombres + Mujeres) * 100)
```

df_feminizacion

```
## # A tibble: 228 x 5
##   Periodo Edad   Hombres   Mujeres Indice_feminizacion
##   <dbl> <fct>     <dbl>     <dbl>             <dbl>
## 1 2006 16-19     244.     142.            36.8
## 2 2006 20-24     928.     727.            43.9
## 3 2006 25-29    1594.    1285.            44.6
## 4 2006 30-34    1790.    1327.            42.6
## 5 2006 35-39    1709.    1198.            41.2
## 6 2006 40-44    1565.    1101.            41.3
## 7 2006 45-49    1399.    947.             40.4
## 8 2006 50-54    1116.    668.             37.5
## 9 2006 55-59     871.    459.             34.5
## 10 2006 60-64    491.    230.             31.9
## # i 218 more rows
```

```
library(dplyr)
library(tidyr)
library(ggplot2)
```

```
ggplot(df_feminizacion, aes(x = Periodo, y = Indice_feminizacion, color = Edad)) +
  geom_line(size = 1.2) +
  labs(
    title = "Índice de feminización en ocupación por edad (2006-2024)",
```

486

487

488

489

490

491

492

493

494

495

496

497

498

499

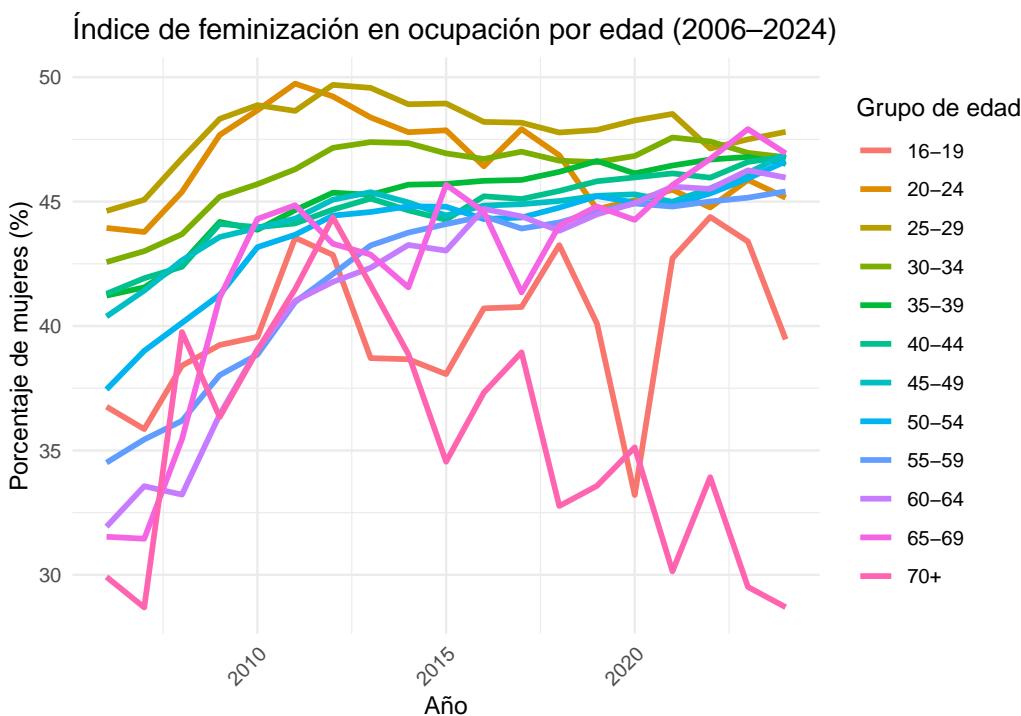
500

501

```

x = "Año",
y = "Porcentaje de mujeres (%)",
color = "Grupo de edad"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))

```



las mujeres tienen mayor presencia en la ocupación en los grupos de edad entre 30 y 54 años, especialmente desde 2010. los jóvenes (16–24 años) y mayores (65+) presentan menor participación femenina y más fluctuaciones.

8. Resultados

9. Conclusión

10. _____ JOE

10.1. GRAFICAS JOE

11. _____

12. _____ JOSE _____

#_____

502

503

504

505

506

507

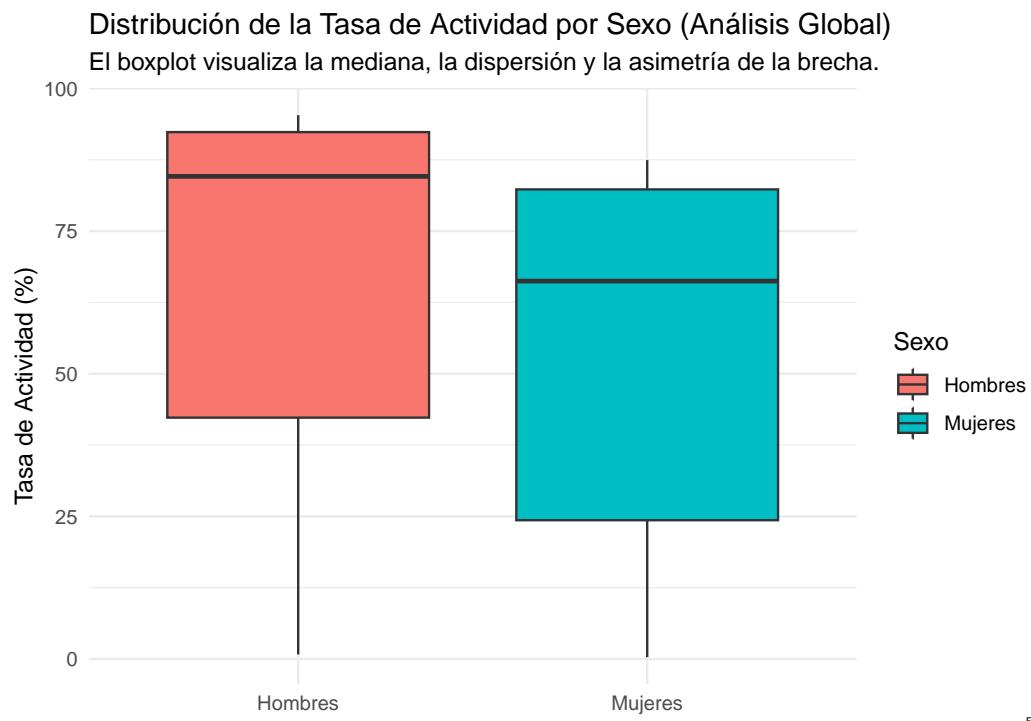
508

509

510

511

512



```
## 514
## Welch Two Sample t-test 515
## 516
## data: Tasa_ocupacion by Sexo 517
## t = 3.767, df = 7.4, p-value = 0.00633 518
## alternative hypothesis: true difference in means between group Hombres and group 519
## 95 percent confidence interval: 520
## 3.368217 14.401655 521
## sample estimates: 522
## mean in group Hombres mean in group Mujeres 523
## 82.77577 73.89084 524
```

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

525
526
527