

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

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
```

```
## # A tibble: 6 x 5
##   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.
```

2.2. Valores faltantes

Todos los valores ausentes se encuentran en Total

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

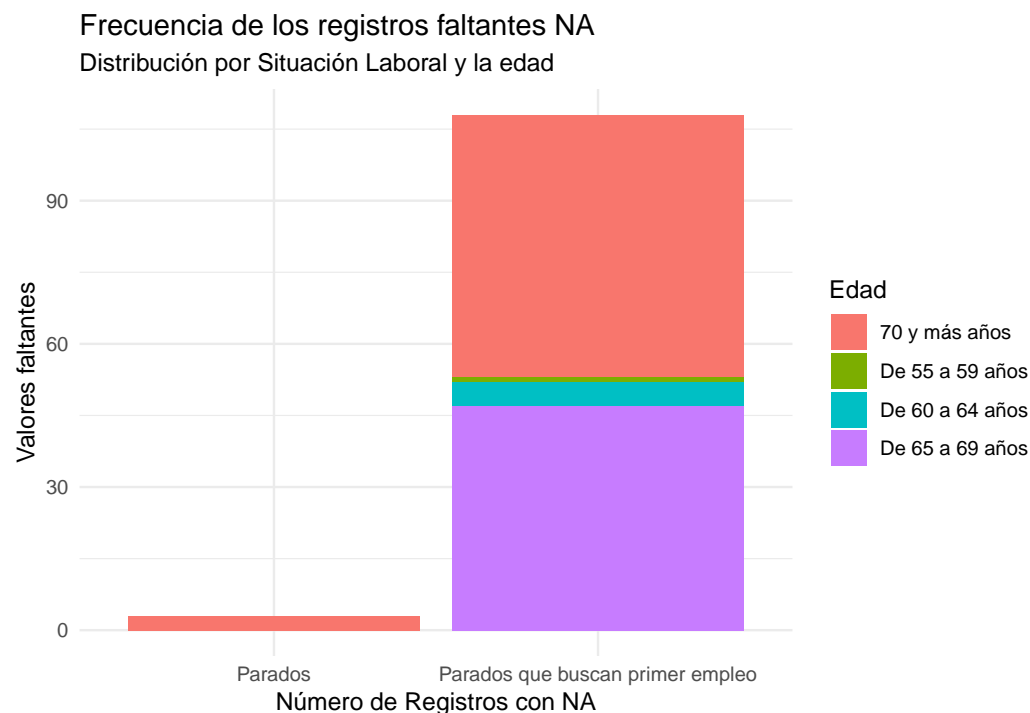
Received:

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/>).



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:

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

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

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:

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

4. Limpieza

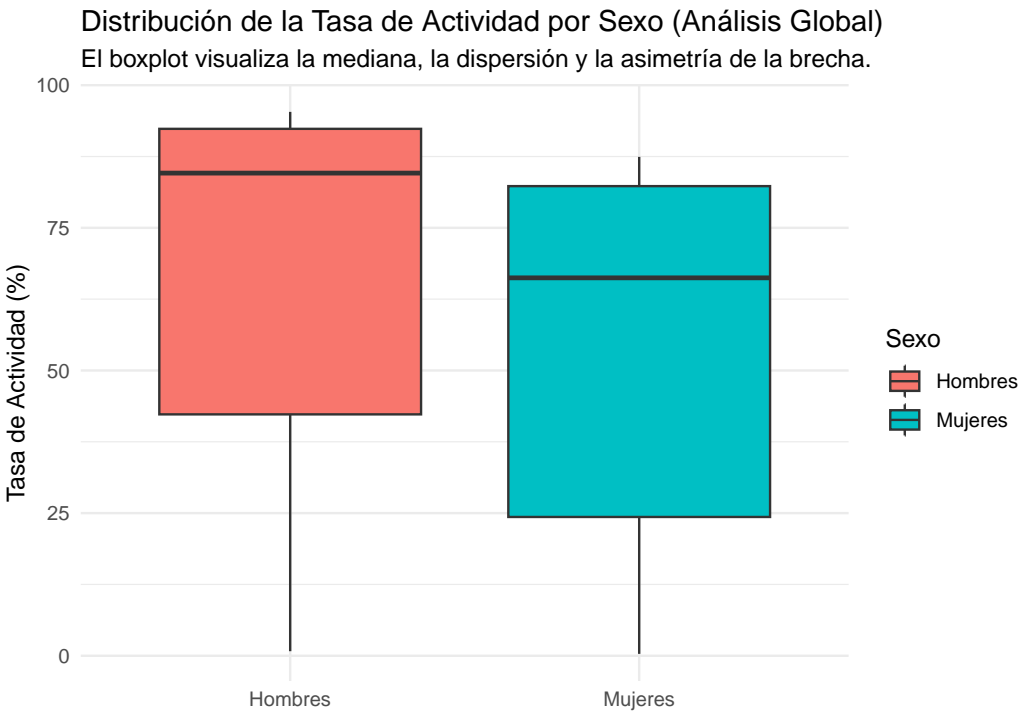
- Renombrar columnas:** Utilizando `rename()` para acortar "Relación con la actividad económica" por "Situacion_Laboral" para simplificar el manejo de columnas.
- Filtrado de doble conteo:** Se eliminan las instancias donde las variables representan la suma de otras instancias (`Sexo="Ambos sexos"` y `Edad="Total"`) para evitar doble conteo y así verificar que cada registro sea una observación única.

- **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`.535455
- **Renombrar variables:** Se renombra la columna "parados que buscan primer empleo" a un formato más corto `Parados_sin_exp`5657
- **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.58596061
- **Creación de tasas:** Se añaden las columnas `Tasa_paro`, `Tasa_actividad` y `Tasa_ocupacion` utilizando `mutate()` y las expresiones de la EPA.6263
- **Ordenación en factores:** Se ordena la columna `Edad` como un factor ordinal y la columna `Sexo` como factor nominal.6465

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.6667

5. Análisis de las tasas68

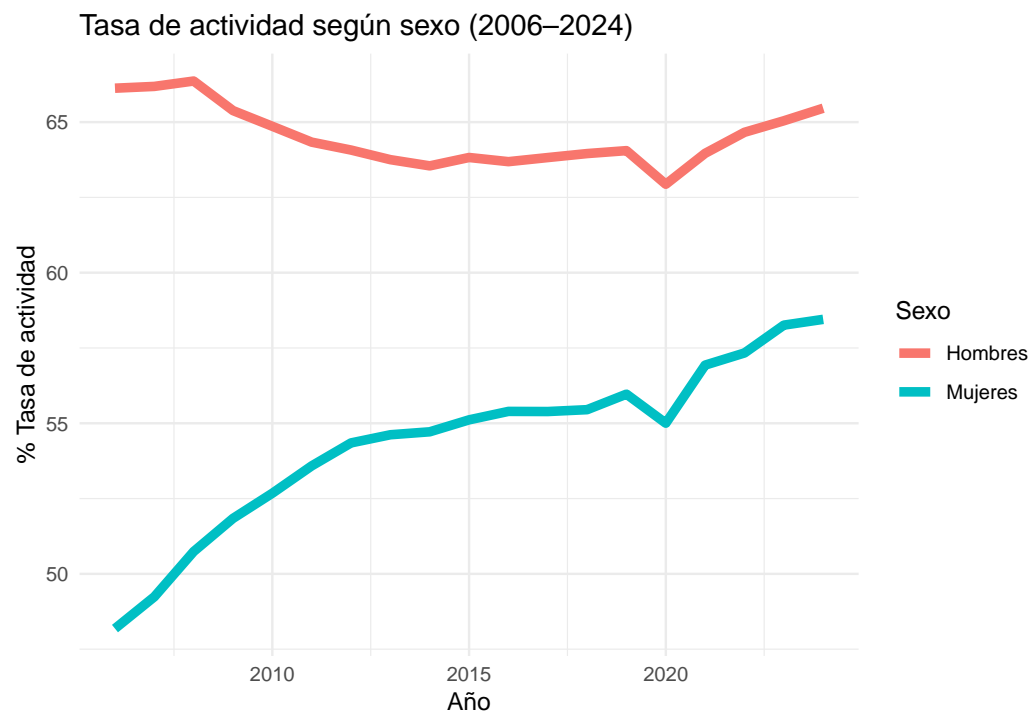
5.1. Distribución de la tasa de actividad por sexo69



5.2. Evolucion de la tasa de paro por género y evolución de la brecha de la tasa de paro por género71

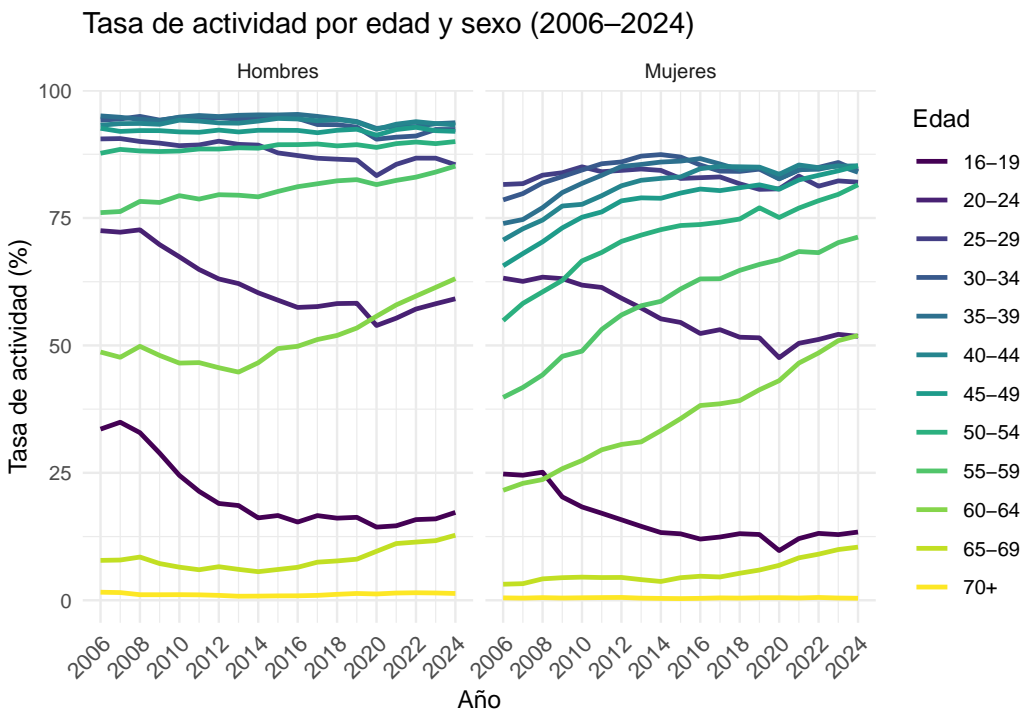
En este bloque de código se construyen varias representaciones gráficas con el fin de analizar la evolución de la tasa de paro y la brecha de género a lo largo del tiempo. Para ello, se parte de una tabla en la que se calcula la variable `brecha_paro`, definida como la diferencia entre la tasa de paro 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.727374757677

Agrupación de datos por periodo y por sexo78

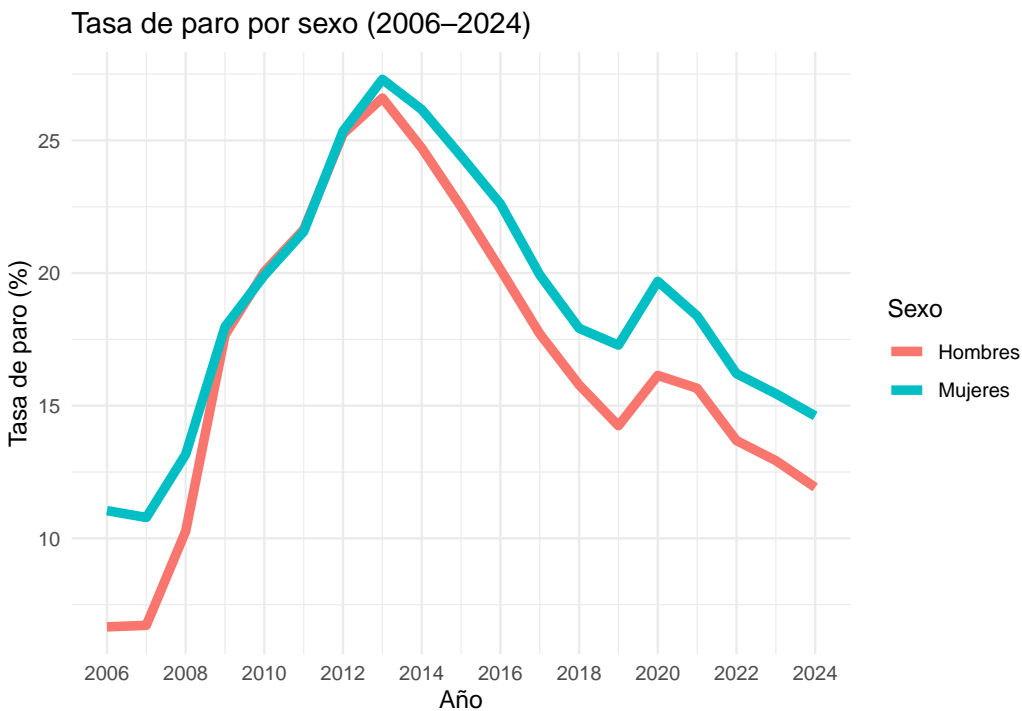


La gráfica enseña la evolución de la tasa de actividad en España entre los años 2006 y 2024, diferenciada por sexo. Por un lado, la proporción de hombres es mucho mayor a la de mujeres, cosa que va ligada a que los hombres siempre han tenido mayor participacion en muchos mas sectores, como lo puede ser la construcción, y a que las mujeres se ligan mas al cuidado familiar. Pese a esto, la actividad masculina en 2008 se desacelera economicamnete, esto provocado por la crisis financiera que se vivió en la epoca. Es en 2017/18 que comienza su recuperación, pero nunca a los niveles previos a la crisis. En 2020, debido al Covid-19, se observa otra caída en la tendencia (porque el Covid-19 afectó a muchos sectores, como la industria por ejemplo). Aunque se observa también su recuperació, tampoco volvió a los nieveles previos de la crisis de 2008 por el otro lado, la tasa de actividad femenina se muestra, en el periodo, ascendente ya que han ido incorporandose cada vez mas en el mercado laboral español

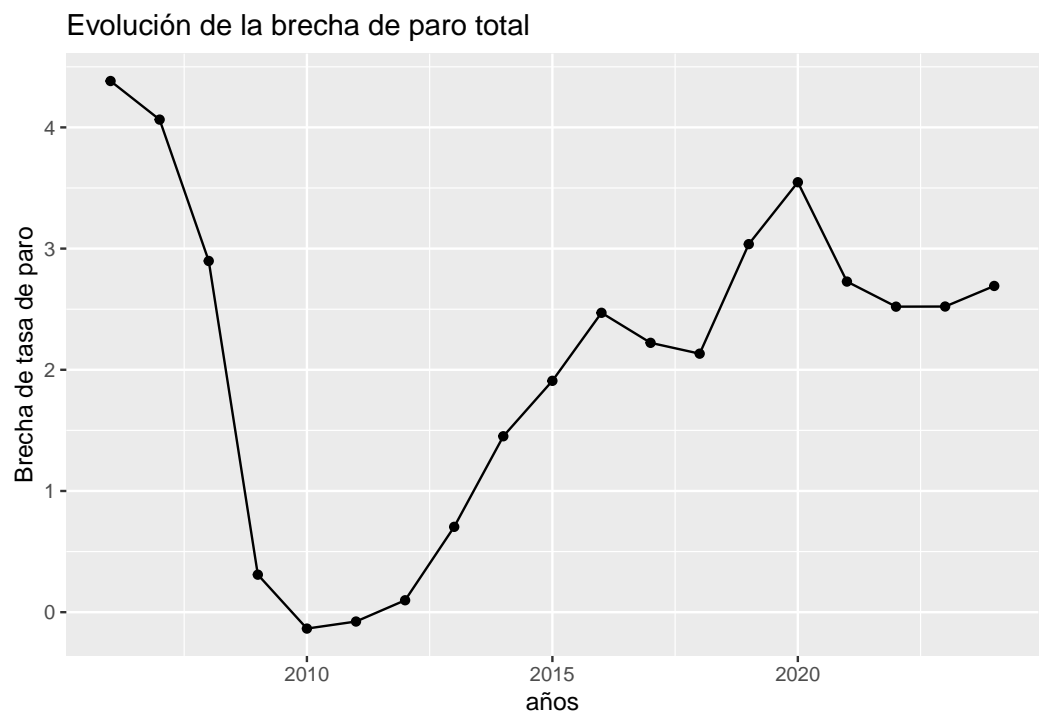
Tasa de actividad por edad y sexo



93



94



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.

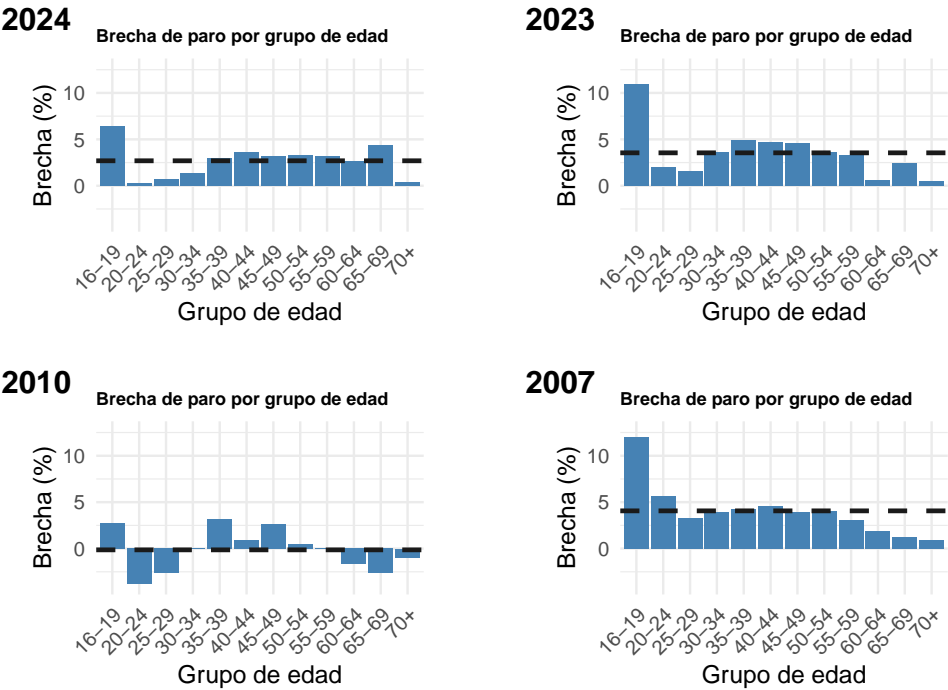
5.3. Evolucion de la tasa de paro y su brecha por grupo de edad

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.

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

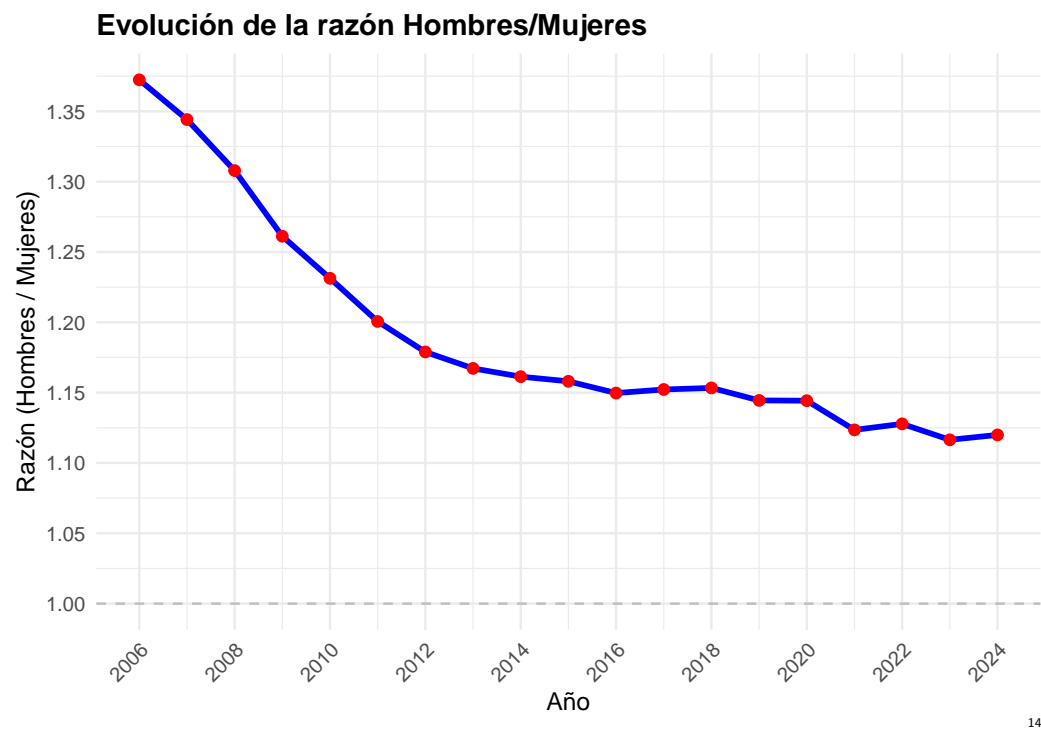
En este bloque de código se elaboran una serie de gráficos exploratorios de barras destinados a analizar la evolución de la brecha de paro por grupos de edad a lo largo del tiempo. Para ello, a partir del conjunto brecha_edades, que contiene la diferencia entre las

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 periodos.



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.



6. Distribución de la población en el mercado laboral

6.1. Composición del mercado laboral

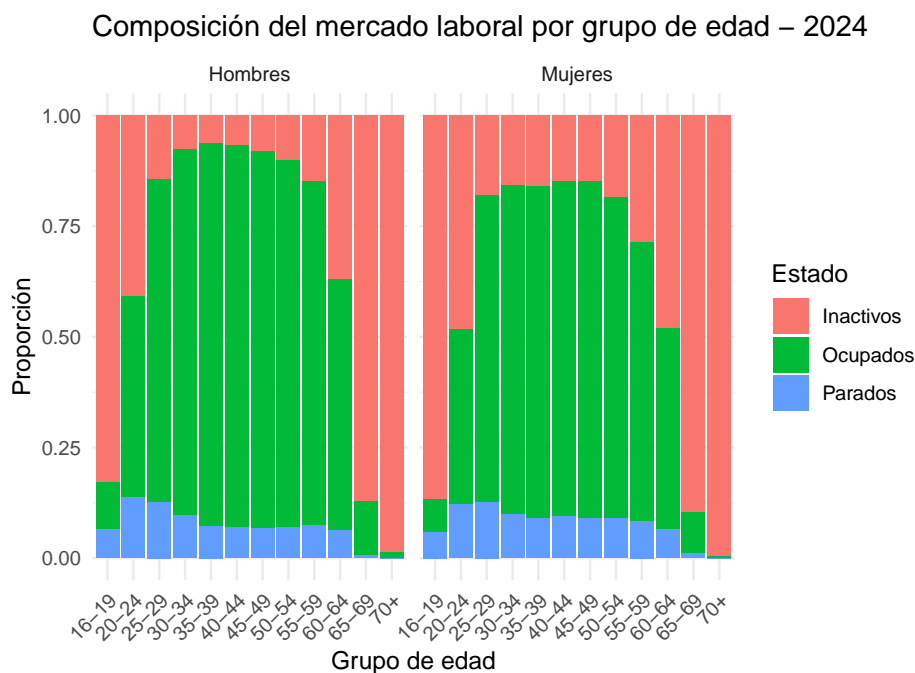
En este bloque de código se genera un gráfico exploratorio con el propósito de comprobar la visualización de la composición del mercado laboral antes de automatizar el proceso para todos los años. Para ello, se filtran los datos del año 2024 y se seleccionan las variables relevantes relacionadas con 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*.

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.

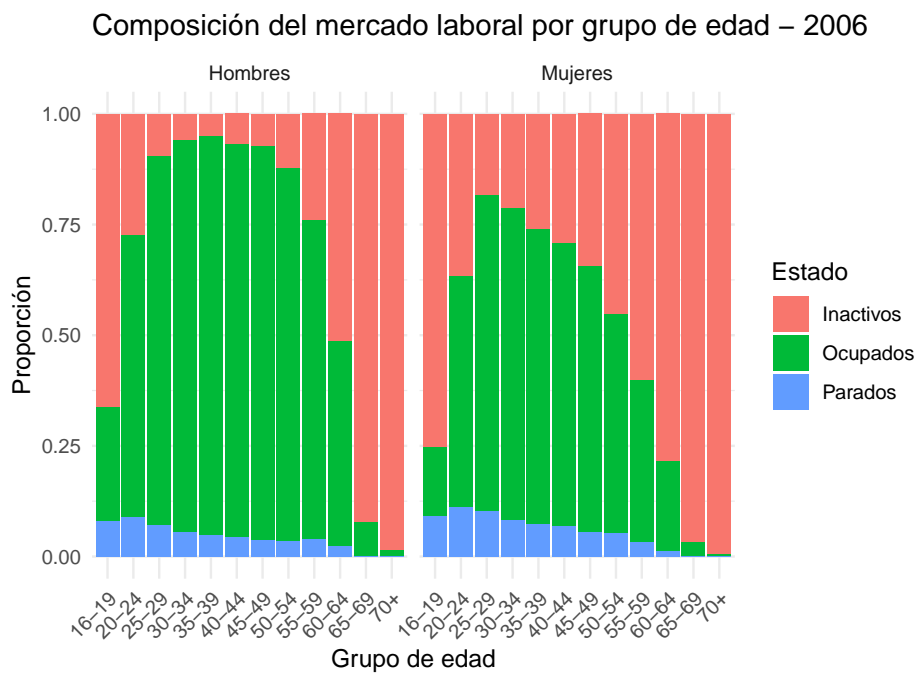
Tras la comprobación inicial con los datos de 2024, se elaboró este bloque de código para automatizar la generación del mismo tipo de gráfico para todos los años del conjunto de datos. Para ello, se utiliza la función `lapply()`, que aplica el mismo procedimiento de visualización a cada año contenido 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.

[[1]]



[[2]]



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

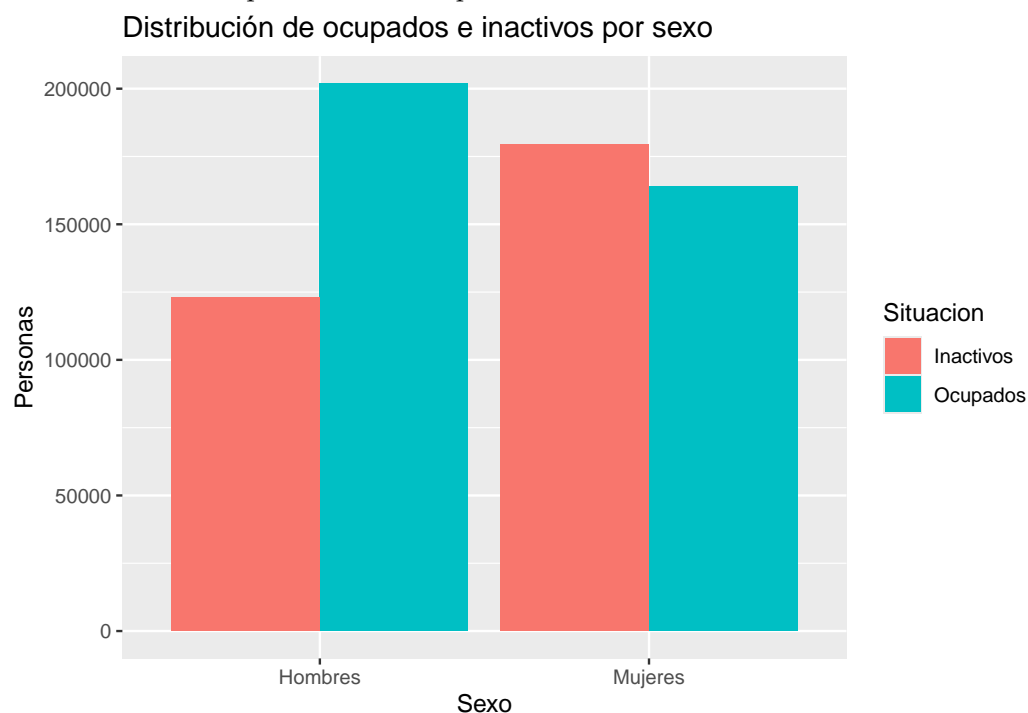
mayor vulnerabilidad del empleo juvenil. En conjunto, los gráficos ponen de manifiesto que, pese a los avances en participación femenina y la estabilidad general del empleo en edades medias, persisten desigualdades de género e importantes diferencias intergeneracionales dentro del mercado laboral.

##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 ($\text{Parados} / \text{sum}(\text{Parados})$).

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.

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



7. act inact

AQUI HAY QUE CALCULAR EL COCIENTE ENTRE ACTIVOS E INACTIVOS

8. Resultados

8.1. desv est y varianza

A tibble: 2 x 3

```

##      Sexo      sd_ocupados var_ocupados
##      <fct>          <dbl>          <dbl>
## 1 Hombres          591.          349261.
## 2 Mujeres          490.          240316.

## # A tibble: 1 x 2
##   brecha_ocup brecha_inact
##         <dbl>         <dbl>
## 1         -166.          247.

##
## Welch Two Sample t-test
##
## data:  Tasa_ocupacion by Sexo
## t = 3.767, df = 7.4, p-value = 0.00633
## alternative hypothesis: true difference in means between group Hombres and group
## 95 percent confidence interval:
##   3.368217 14.401655
## sample estimates:
## mean in group Hombres mean in group Mujeres
##           82.77577           73.89084

```

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.