



Gráficos en R
Instituto Nacional de Estadística (INE)

David Pérez Ros

24/10/24

Indice

1	Introducción	5
1.1	Paquetes Principales	5
1.2	Datos	6
2	Truco para infografías	7
3	lattice	11
3.1	Características Principales	11
3.2	Datos de ejemplo	12
3.3	Realización manual	13
3.4	Realización automática (lattice)	14
3.5	Más información	16
4	plotly	17
4.1	Características Principales	17
4.2	Tipos de gráficos	17
4.3	Ejemplo de Uso	18
4.4	Más información	39
5	SF	41
6	jk	43
7	Leaflet	45
7.1	General	45
7.2	Mapas	46
	Anexos	55
A	Ineapir	55
A.1	Introducción	55

Capítulo 1

Introducción

La visualización de datos es una herramienta fundamental en el análisis estadístico y la ciencia de datos, ya que permite interpretar y comunicar información compleja de manera efectiva. Con el creciente volumen de datos disponibles en diversas disciplinas, la capacidad de visualizar patrones y tendencias se ha vuelto esencial para la toma de decisiones informadas. En este contexto, [R](#), un lenguaje de programación y entorno de software para análisis estadístico, ha desarrollado un ecosistema robusto y diverso de paquetes para la creación de gráficos.

Las funciones gráficas base de [R](#) ofrecían una solución sencilla y eficiente para la creación de gráficos básicos. Con el tiempo, la necesidad de visualizaciones más sofisticadas y personalizables llevó al desarrollo de paquetes adicionales que ampliaron significativamente las capacidades gráficas de [R](#).

Hoy en día, la comunidad de [R](#) dispone de una variedad de paquetes especializados que permiten desde la creación de gráficos simples hasta la construcción de visualizaciones interactivas y altamente personalizadas. Esta interactividad mejora significativamente la capacidad de los usuarios para explorar datos de manera dinámica, ofreciendo nuevas perspectivas y facilitando la comprensión de información compleja.

En este estado del arte, se revisan los paquetes más relevantes para la visualización de datos en [R](#), destacando sus características, ventajas y aplicaciones. Su versatilidad y su reciente incorporación de formatos como [Bookdown](#), [R Markdown](#), y [Quarto](#) lo hacen idóneo para realizar análisis de datos centrándose únicamente en dicho análisis y no perder tiempo en diseñar el entorno en el que se presentarán.

Este proyecto se ha realizado en [Bookdown](#)¹ con la idea de que sea fácilmente exportable a PDF o leído desde la web, además de poder evolucionar en el tiempo, incluyendo correcciones tipográficas o nuevos apartados. Se expondrán los diferentes tipos de paquetes para graficar en [R](#), clasificándolos según su propósito: visualización exploratoria, visualización de resultados estadísticos, gráficos interactivos, gráficos para informes y publicaciones, entre otros. Esto permitirá una comprensión integral de las herramientas disponibles y su adecuada aplicación en diversos contextos de análisis de datos.

1.1 Paquetes Principales

- **ggplot2**: permite construir visualizaciones complejas mediante una sintaxis declarativa y altamente personalizable². Su flexibilidad y capacidad para manejar grandes conjuntos de datos lo convierten en una herramienta esencial para analistas y científicos de datos.
- **plotly**: paquete popular que extiende la funcionalidad de [ggplot2](#) al ofrecer gráficos interactivos. Desarrollado inicialmente como una biblioteca para Python, su integración con [R](#) ha permitido a los usuarios crear visualizaciones dinámicas y envolventes que pueden ser fácilmente compartidas en plataformas web³.

¹Ver (Xie, 2017) para más información sobre [Bookdown](#).

²Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

³Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. CRC Press.

- **lattice**: paquete diseñado para gráficos multivariantes, basado en el concepto de gráficos en trellis. Proporciona un enfoque sistemático para la visualización de datos multivariados mediante el uso de paneles condicionados⁴. Es particularmente útil en el análisis exploratorio de datos complejos.
- **leaflet**, **sf** y **tmap**: para la visualización de datos espaciales, los paquetes **sf** y **tmap** proporcionan herramientas especializadas. **sf** facilita la manipulación de datos geoespaciales, mientras que **tmap** permite la creación de mapas temáticos interactivos y estáticos⁵. Estos paquetes son esenciales para el análisis geoespacial y la presentación de resultados en disciplinas como la geografía y la ecología.
- **Tendencias Actuales**: Las tendencias actuales en la visualización de datos con R incluyen un aumento en la demanda de gráficos interactivos y dashboards, la integración con herramientas de presentación web, y el desarrollo de paquetes que hacen más accesible la creación de gráficos para usuarios menos técnicos⁶.

1.2 Datos

En este proyecto la mayor parte de datos usados se han extraído directamente del servicio API del INE, a través del paquete **ineapir**. Para más información, (véase (Crespo, 2024)). Esto contribuye a la independencia del proyecto, sin necesidad de descargar datos de fuentes externas de manera “manual” para poder replicar los ejemplos. Del mismo modo, promociona y expone casos de uso para dicha herramienta.

Nota: Este espacio de trabajo se encuentra todavía en proceso de elaboración y por ello seguirá incluyendo información nueva de manera periódica.

⁴Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer.

⁵Lovelace, R., Nowosad, J., & Muenchow, J. (2019). *Geocomputation with R*. CRC Press.

⁶Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2021). *shiny: Web Application Framework for R*. R package version 1.7.1.

Capítulo 2

Truco para infografías

En este apartado se expondrá como exportar por capas un gráfico de R a Power Point/ Photoshop. Este método puede resultar muy útil para realizar ciertas infografías, ya que en la herramienta de destino se puede modificar el formato del gráfico fácilmente.

Tomemos los datos de la población de España desde 2010 hasta 2022, para ello usaremos la API desarrollada para la obtención de datos. Para más información sobre esta, véase (Crespo, 2024).

1. **Buscar identificador de la serie.** Abriendo los datos en la web del INE, vemos que el código identificador de la serie es ECP320 (véase cheatsheet de (Crespo, 2024)).
2. **Cargar los datos** usando función `ineapir::get_data_series()`.

```
# Verificar si el paquete está instalado
if (!require("ineapir")) remotes::install_github("es-ine/ineapir")

# Cargar paquete
library(ineapir)

# Cargar Serie Población ECP320
a <- get_data_series("ECP320", dateStart = "2002/01/01", unnest = TRUE, tip = "AM")
datos <- data.frame(
  fecha = as.Date(a$Fecha),
  pob = a$Valor
)

# Formateamos año y mes
datos$year <- format(datos$fecha, "%Y")
datos$month <- format(datos$fecha, "%m")

# Filtramos para tener sólo población a 1 de Enero
datos <- datos[datos$month == "01", ]
```

3. Crear gráfico que represente la población a lo largo de los años.

```
# Verificar si está instalado
if (!require("ggplot2")) install.packages("ggplot2")
library(ggplot2)
```

```
# Function to set numbers with marks and without scientific notation
marks_no_sci <- function(x) format(x, big.mark = ".", decimal.mark = ",", scientific = FALSE)

# Crear el gráfico de puntos
grafico <- ggplot(datos, aes(x = year, y = pob)) +
  geom_point(stat = "identity", color = "#457e76") +
  labs(title = "Población en España", x = "Año", y = "Población") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  scale_y_continuous(labels = marks_no_sci, limits = c(40000000, 50000000), breaks = seq(40000000, 50000000, 1000000))

grafico
```

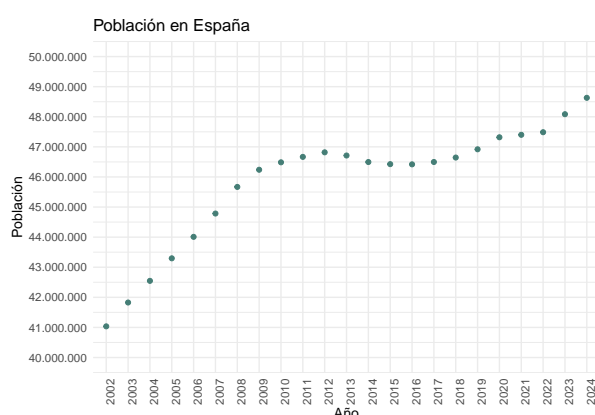


Figure 2.1: Evolución de la población a lo largo de los años.

Ahora imaginar que un equipo quiere realizar una infografía con este gráfico. Se puede guardar como .svg de tal manera que sea un gráfico vectorial y se pueda abrir en aplicaciones como Photoshop o Power Point.

Veamos:

```
# Verificar si el paquete está instalado
if (!require("svglite")) install.packages("svglite")
library(svglite)
ggsave(file = "/Users/davpero/Desktop/grarfico.svg", plot = grafico)
```

Por ejemplo: Pegamos el gráfico en Power Point y clickando con el botón derecho le damos a Grupo-> Desagrupar. Lo que conseguimos es dejar la imagen como elementos vectoriales independientes, tal y como se puede ver en la Figura 2.2

```
knitr::include_graphics(c("fig/capturas/plot1.png", "fig/capturas/plot2.png"))
```

Posteriormente, en Power Point, el equipo de infografías puede darle el formato deseado sin necesidad de tener que modificar el gráfico desde su código fuente (ggplot2). No obstante, estos cambios posibles se reducen a aspectos de formato tal como cambiar el tamaño de letra, colores, o eliminar elementos.

Veamos pues como eliminamos las rayas de fondo y cambiamos el tamaño y color de los textos:

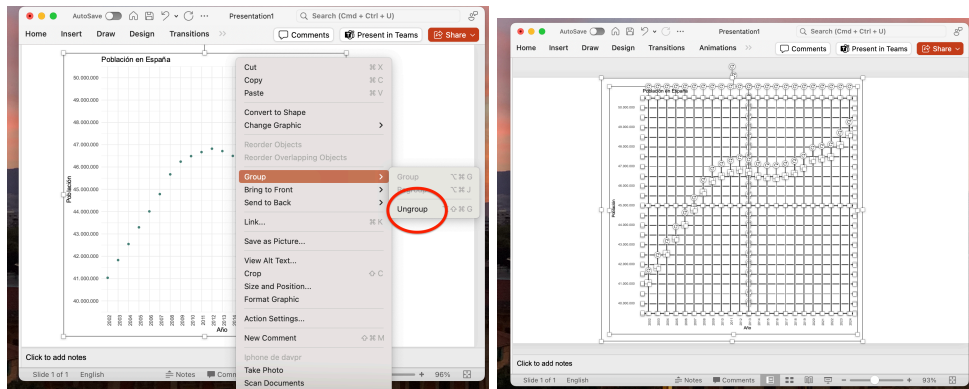


Figure 2.2: Pegado del gráfico como imagen vectorial

```
knitr::include_graphics("fig/capturas/plot3.png")
```

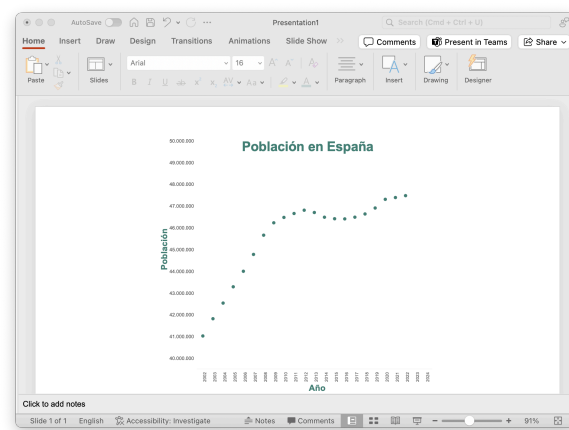


Figure 2.3: Modificación del gráfico en Power Point

Capítulo 3

lattice

El paquete **lattice** es una herramienta poderosa para la visualización de datos multivariantes en R. Desarrollado por Deepayan Sarkar, lattice está basado en el concepto de gráficos en trellis ¹, que facilita la visualización de relaciones complejas entre múltiples variables mediante la creación de gráficos condicionados. Esto es especialmente útil en análisis exploratorios de datos, donde es crucial comprender las interacciones y patrones en conjuntos de datos grandes y complejos.

3.1 Características Principales

1. **Gráficos Condicionados:** lattice permite crear gráficos que muestran relaciones entre variables condicionadas a los valores de otras variables. Esto facilita la visualización de patrones en subgrupos de datos.
2. **Paneles Múltiples:** Los gráficos pueden ser divididos en paneles múltiples, cada uno mostrando una porción diferente del conjunto de datos, lo que permite una comparación visual directa entre diferentes subgrupos.
3. **Fórmulas:** Utiliza una fórmula para especificar las relaciones entre las variables que se van a graficar, proporcionando una sintaxis clara y concisa.
4. **Temas Personalizables:** lattice permite la personalización de temas gráficos, incluyendo colores, tamaños y tipos de letra, lo que facilita la creación de visualizaciones estéticamente agradables.
5. **Integración con el Modelo de Trellis:** La integración con el modelo de Trellis permite la creación de gráficos consistentes y bien organizados.

Lo primero escribimos unas líneas de código que verifiquen si el paquete está instalado, y en caso negativo lo instalen. Posteriormente lo cargamos:

```
# Verificar si el paquete está instalado
if (!require("lattice")) install.packages("lattice")

# Cargar paquete
library(lattice)
```

Como se comentaba, la principal funcionalidad de este paquete es que permite diferenciar cualquier tipo de gráfico (Diagrama de dispersión, Histograma,...) a partir de una variable categórica mostrando diferentes gráficos o superpuestos en uno mismo. Para ilustrar lo ventajoso de este gráfico se expondrá un ejemplo:

¹Un **gráfico en trellis** es una visualización que presenta múltiples gráficos dispuestos en un conjunto de paneles, permitiendo comparar diferentes subconjuntos de datos de manera eficiente. Cada panel muestra el mismo tipo de gráfico, pero para diferentes segmentos de datos, facilitando la detección de patrones y tendencias en grupos distintos. Véase https://es.wikipedia.org/wiki/Gr%C3%A1fico_de_celos%C3%ADa

3.2 Datos de ejemplo

Vamos a usar el conjunto de datos `mtcars` que pertenece al paquete `datasets` el cual fue elaborado para la revista *Motor Trend US* en 1974 y que contiene el consumo de combustible y 10 aspectos relacionados con el diseño y rendimiento para 32 coches distintos. Es un dataset muy usado en el aprendizaje de R y que permite ejemplificar diversas técnicas estadísticas.

Variables del dataset `mtcars`:

1. **mpg**: Medida autonomía del coche. Millas recorridas por galón de combustible (miles per gallon)
2. **cyl**: Número de cilindros
3. **disp**: Desplazamiento (pulgadas cúbicas)
4. **hp**: Potencia (caballos de fuerza)
5. **drat**: Relación del eje trasero
6. **wt**: Peso del auto (miles de libras)
7. **qsec**: Tiempo de 1/4 de milla (en segundos)
8. **vs**: Tipo de motor (0 = V-shaped, 1 = Straight)
9. **am**: Tipo de transmisión (0 = Automática, 1 = Manual)
10. **gear**: Número de marchas
11. **carb**: Número de carburadores

```
data <- mtcars
data$gear <- factor(data$gear, levels = c(3, 4, 5))
data$cyl <- factor(data$cyl, levels = c(4, 6, 8))
head(data)
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant        18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Suponer que se desea mostrar la relación entre `mpg` y `hp` para ver si hay relación entre la velocidad y la potencia del coche. Por ello, parece razonable realizar un gráfico tipo:

```
plot(hp ~ mpg, data = data, main = " Autonomía (mpg) vs. Potencia (hp) ")
```

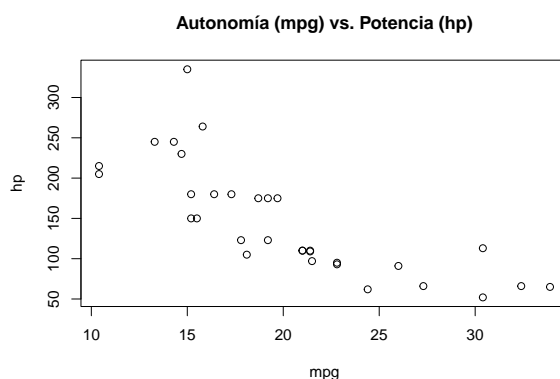


Figure 3.1: Gráfico de dispersión de potencia y autonomía de los automóviles.

Se observa que a medida que aumenta la potencia del motor (hp), disminuye el consumo de combustible medido en millas por galón (mpg), lo cual tiene sentido ya que a mayor potencia de un coche, más combustible se espera que gaste, y por tanto menor autonomía tendrá.

Ahora suponer que queremos hacer distinciones en función del número de cilindros del coche, es decir, el mismo **gráfico de dispersión condicionado por el número de cilindros**:

```
plot(hp ~ mpg, col = factor(cyl), data = data, main = " Autonomía (mpg) vs. Potencia (hp) ")
# Legend
legend("topright",
      title = "Cilindros",
      legend = levels(factor(data$cyl)),
      pch = 19,
      col = factor(levels(factor(data$cyl)))
)
```

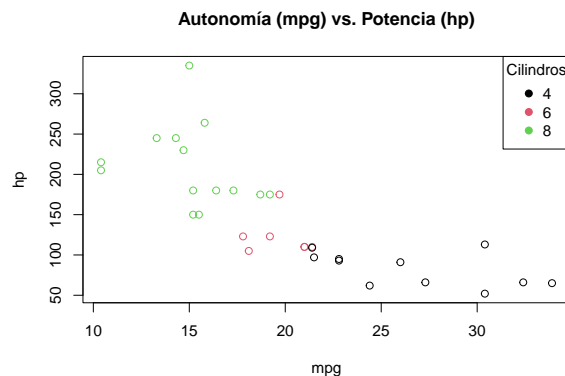


Figure 3.2: Gráfico de dispersión condicionado por el número de cilindros.

Se sigue observando una relación inversa entre potencia y autonomía de los automóviles, teniendo en cuenta que los de mayor cilindrada son los que mayor potencia y menor autonomía tienen. A medida que aumente el número de observaciones se tenderá a ver muy lleno el gráfico por lo que convendrá separar en varios gráficos dependiendo del número de cilindros.

3.3 Realización manual

Para separar estos tres gráficos, de manera “manual”, se procedería:

```
# Disposición de los gráficos
par(mfrow = c(1, 3))
# 4 cilindros
plot(hp ~ mpg, data = data[data$cyl == 4, ], main = "4 cilindros", col = "black")

# 6 cilindros
plot(hp ~ mpg, data = data[data$cyl == 6, ], main = "6 cilindros", col = "red")

# 8 cilindros
plot(hp ~ mpg, data = data[data$cyl == 8, ], main = "8 cilindros", col = "green")
```

Conforme aumente el número de categorías, realizar estos gráficos será tedioso y ahí es donde entra en juego el paquete `lattice`

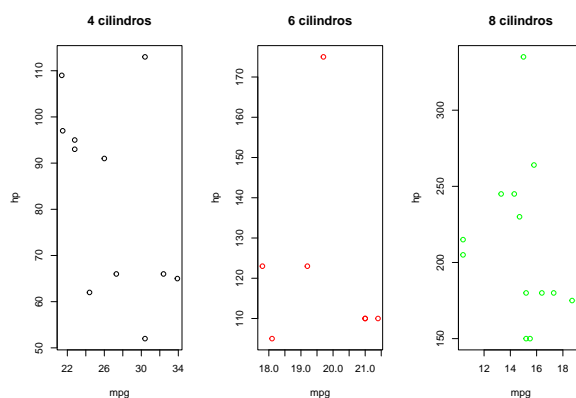


Figure 3.3: Gráfico de dispersión condicionado por el número de cilindros (separados).

3.4 Realización automática (lattice)

Véase que el siguiente gráfico realiza la misma tarea con mucho menos código:

```
# Gráfico por número de cilindros
xyplot(hp ~ mpg | cyl, group = cyl, data = data, scales = "free", aspect = "fill")
```

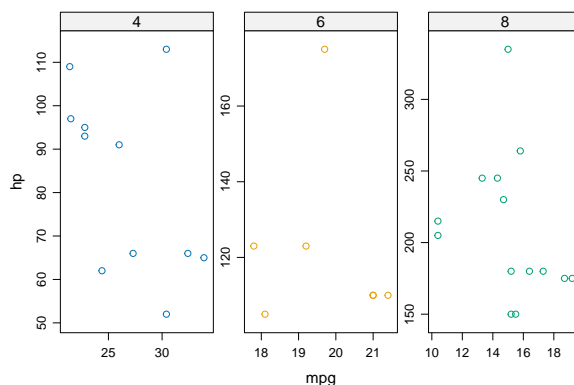


Figure 3.4: Gráfico de dispersión con lattice condicionado por el número de cilindros.

Además, se pueden combinar en un miso gráfico al igual que antes (quitando `| cyl`):

```
# Gráfico
xyplot(hp ~ mpg, group = cyl, data = data, scales = "free", aspect = "fill")
```

Es decir, veamos la forma general de esta función [Custom block ver info](#)

```
plot_function(y ~ x | g, group = g, data)
```

donde:

- **plot_function**: es cualquier función de graficar de `lattice`, por ejemplo
 - **xyplot**: Diagrama de dispersión
 - **density.plot**: Línea de densidad

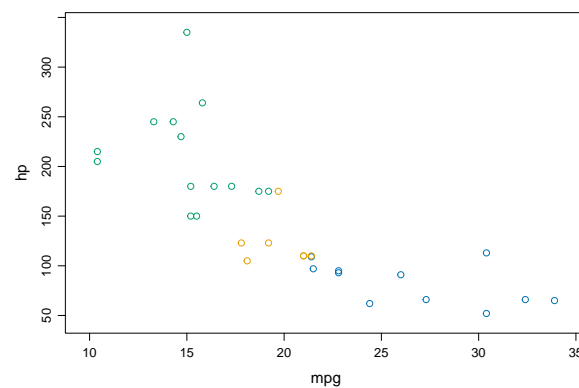


Figure 3.5: Gráfico de dispersión con lattice condicionado por el número de cilindros.

- **histogram**: Histograma
- **bwplot**: Diagramas de caja
- **| g** (OPCIONAL): Indica que el gráfico se va a dividir en tantos gráficos como valores tome la variable `data$g`. Es decir, nos mostrará el gráfico de $y \sim x$ para cada grupo de la variable `g`
- **group=g** (OPCIONAL): indica que pinte los elementos dibujados agrupando por la variable `g`.
- **data**: nombre del conjunto de datos que contiene las variables `x, y, g`.

En el **siguiente ejemplo**, queremos representar la densidad de la autonomía de los vehículos (mpg) distinguiendo el número de cilindros que tienen, por ello:

- Añadir `densityplot(~mpg, data = data)` para el gráfico de densidad.
- Añadir argumento `| cyl` para realizar un gráfico por cada número de cilindros distinto.
- Añadir argumento `group=cyl` para colorear dependiendo del número de cilindros del vehículo.

```
# Gráficos independientes
densityplot(~ mpg | cyl, group = cyl, data = data, plot.points = TRUE)
```

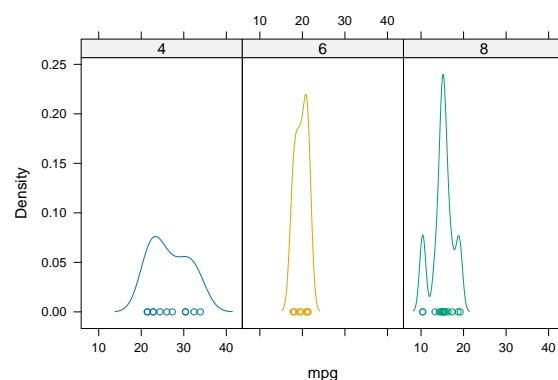


Figure 3.6: Densidad de la autonomía condicionada al número de cilindros (en gráficos distintos).

Ahora vamos a mostrarlos todos superpuestos en un mismo gráfico, eliminando el argumento `| cyl`:

```
# Superpuestos
densityplot(~mpg, group = cyl, data = data, plot.points = FALSE)
```

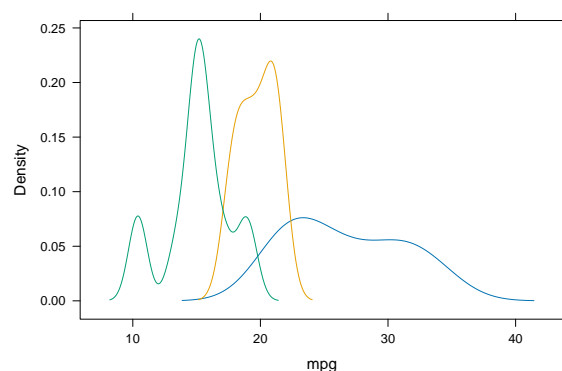


Figure 3.7: Densidad de la autonomía condicionada al número de cilindros.

3.5 Más información

Para más información acerca del paquete `lattice` y casos de uso, consultar:

- [Página del paquete en CRAN:](#)
 - Sarkar, Deepayan. *Lattice: Multivariate Data Visualization with R*. Springer, 2008. Este libro proporciona una cobertura completa de las capacidades de `lattice` para la visualización de datos multivariantes en R.
 - [R Documentation - lattice](#). Página de documentación que proporciona ejemplos de uso, detalles de funciones y comentarios de la comunidad.
 - [Quick-R: Lattice Graphs](#) Un tutorial práctico que cubre los conceptos básicos de los gráficos `lattice` y proporciona ejemplos de código.
-

Capítulo 4

plotly

El paquete [plotly](#) es una poderosa herramienta para la creación de gráficos interactivos en R. Desarrollado inicialmente como una biblioteca para Python, plotly ha sido adaptado para R, permitiendo a los usuarios beneficiarse de sus capacidades de visualización dinámica y envolvente. La interactividad de los gráficos generados con plotly facilita que se puedan integrar en sitios web, y en combinación con otros paquetes como [ineapir](#) (véase (Crespo, 2024)) se puede proporcionar un análisis directo e interactivo de datos extraídos en bruto del INE.

4.1 Características Principales

1. **Interactividad:** plotly destaca por sus capacidades interactivas, permitiendo a los usuarios hacer zoom, desplazar, y obtener información detallada mediante “hover” sobre los elementos del gráfico.
2. **Compatibilidad con ggplot2:** plotly puede convertir gráficos estáticos de ggplot2 en gráficos interactivos sin necesidad de reescribir el código original.
3. **Soporte para gráficos en 3D:** plotly facilita la creación de gráficos tridimensionales, como superficies y dispersión en 3D, que son difíciles de lograr con otras herramientas.
4. **Integración con R Markdown y Shiny:** los gráficos de plotly pueden integrarse fácilmente en documentos R Markdown y aplicaciones Shiny, mejorando la presentación de informes y el desarrollo de aplicaciones interactivas.
5. **Variedad de tipos de gráficos:** desde gráficos de líneas y barras hasta mapas y gráficos de superficie, plotly soporta una amplia gama de visualizaciones.

4.2 Tipos de gráficos

La base de *plotly* es la interactividad de los gráficos que permite exploraciones más profundas y detalladas. En general, se pueden realizar:

- **Mapas interactivos:** Permiten explorar datos geoespaciales por diferentes áreas.
- **Gráficos de barras:** Facilitan la comparación de categorías y pueden permitir la selección dinámica de subcategorías.
- **Series temporales:** Posibilitan filtrar por años y observar tendencias a lo largo del tiempo.
- **Histogramas interactivos:** Permiten ajustar el número de bins y explorar la distribución de los datos.
- **Box plots:** Ofrecen interactividad para seleccionar y resaltar datos atípicos o rangos específicos.
- **Gráficos de dispersión:** Permiten hacer zoom y seleccionar áreas específicas para análisis detallado.
- **Gráficos de calor:** Visualizan matrices de datos con interactividad para resaltar valores específicos.

- **Gráficos de burbujas:** Facilitan la visualización de relaciones entre múltiples variables con capacidad de ajuste de tamaño y color de las burbujas.

4.2.1 Nota

Debido a que los gráficos generados con `plotly` son interactivos, se genera un conflicto a la hora de guardar estos como pdf ya que se debe seleccionar una de las posibles vistas estáticas de dichos gráficos. La siguiente línea de código incluye el paquete `webshot` que se encarga de guardar capturas (imágenes estáticas) de dichos gráficos interactivos a la hora de imprimir la página como `.pdf`.

4.3 Ejemplo de Uso

4.3.1 Gráfico interactivo

A continuación se muestra un ejemplo básico de cómo crear un gráfico interactivo con `plotly` a partir de un conjunto de datos:

```
# Instalar y cargar el paquete plotly
# install.packages("plotly")

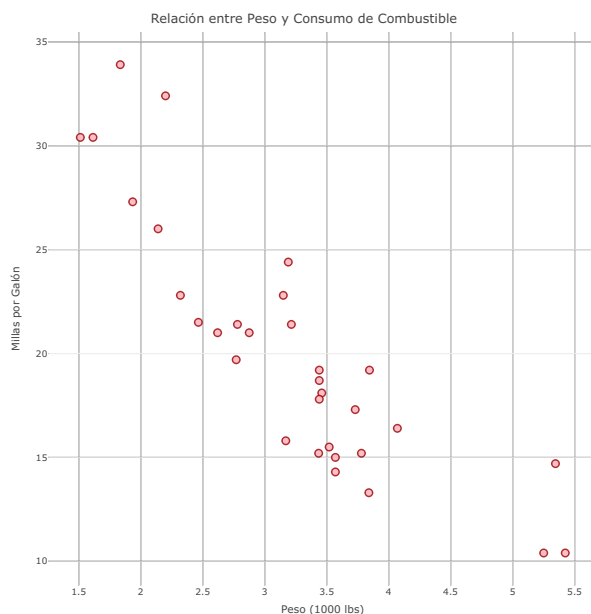
# Verificar si está instalado
if (!require("plotly")) install.packages("plotly")
if (!require("dplyr")) install.packages("dplyr")
if (!require("webshot")) install.packages("webshot")
if (!require("ineapir")) install.packages("ineapir")
```

```
# Cargar paquete
library(dplyr)
library(plotly)
library(webshot)
library(ineapir)

# Datos de ejemplo
data <- mtcars
```

```
# Crear un gráfico interactivo de dispersión
p <- plot_ly(data,
  x = ~wt, y = ~mpg, width = 800, height = 800, type = "scatter", mode = "markers",
  marker = list(size = 10, color = "rgba(255, 182, 193, .9)", line = list(color = "rgba(152, 0, 0, .9)"))
) %>%
  layout(
    title = "Relación entre Peso y Consumo de Combustible",
    xaxis = list(title = "Peso (1000 lbs)"),
    yaxis = list(title = "Millas por Galón")
  )

# Mostrar el gráfico
p
```



En este ejemplo, se utiliza el conjunto de datos `mtcars` para crear un gráfico de dispersión interactivo que muestra la relación entre el peso del vehículo y el consumo de combustible. El argumento `marker` permite personalizar la apariencia de los puntos en el gráfico, mientras que `layout` se utiliza para añadir títulos y etiquetas a los ejes.

4.3.2 Tree Maps

En este apartado se va a exponer como crear los [Tree Maps](#) un tipo de gráfico muy útil para ver como se distribuye una cierta característica entre diferentes categorías. Por ejemplo, en este caso, como se distribuye el gasto medio por persona en España dentro de los grupos ECOICOP¹.

La ventaja de usar `plotly` es que nos permite construir TreeMaps interactivos, es decir, con distintos niveles de profundidad. En este caso en particular si apretamos en como se distribuye el gasto para el **ECOICOP 01: Agricultura, ganadería, caza y servicios relacionados con las mismas**, nos permite inspeccionar como se distribuye el gasto medio dentro de ese subnivel.

1. En la web del ine [buscamos la tabla que contiene los datos deseados](#), relativos a *Gasto total, gastos medios y distribución del gasto de los hogares*, que tiene por id 25143. (Véase <https://www.ine.es/jaxiT3/Tabla.htm?t=25143&L=0>)
2. Usando `ineapir` extraemos los datos usando el servicio API del INE. Ejecutando `get_metadata_table_varval(25143)` vemos por que variables debemos filtrar para obtener los datos deseados.

```
get_metadata_table_varval(25143)
```

```
##      Id Fk_Variable
## 1  16473         349
## 2   8997          70
## 3   8998          70
## 4   8999          70
## 5   9000          70
## 6   9001          70
```

¹La [ECOICOP](#) es una clasificación europea de consumo, que facilita la consulta conjunta con otras estadísticas como el IPC, y a su vez permite la comparabilidad de datos con otros países. Clasifica los bienes y servicios en 12 grandes grupos de gasto (del 01 al 12).

## 7	9002	70
## 8	9003	70
## 9	9004	70
## 10	9005	70
## 11	9006	70
## 12	9007	70
## 13	9008	70
## 14	9009	70
## 15	9010	70
## 16	9011	70
## 17	9012	70
## 18	9013	70
## 19	9015	70
## 20	8995	70
## 21	304092	762
## 22	304093	762
## 23	304094	762
## 24	304095	762
## 25	304096	762
## 26	304097	762
## 27	304098	762
## 28	304099	762
## 29	304100	762
## 30	304101	762
## 31	304102	762
## 32	304103	762
## 33	304104	762
## 34	72	3
## 35	8641	3
## 36	8642	3
## 37	8856	57
## 38	8857	57
## 39	8858	57
## 40	8859	57
## 41	8860	57

##	Nombre
## 1	Total Nacional
## 2	Andalucía
## 3	Aragón
## 4	Asturias, Principado de
## 5	Balears, Illes
## 6	Canarias
## 7	Cantabria
## 8	Castilla y León
## 9	Castilla - La Mancha
## 10	Cataluña
## 11	Comunitat Valenciana
## 12	Extremadura
## 13	Galicia
## 14	Madrid, Comunidad de
## 15	Murcia, Región de
## 16	Navarra, Comunidad Foral de
## 17	País Vasco
## 18	Rioja, La
## 19	Ceuta
## 20	Melilla
## 21	Índice general

## 22	Alimentos y bebidas no alcohólicas
## 23	Bebidas alcohólicas y tabaco
## 24	Vestido y calzado
## 25	Vivienda, agua, electricidad, gas y otros combustibles
## 26	Muebles, artículos del hogar y artículos para el mantenimiento corriente del hogar
## 27	Sanidad
## 28	Transporte
## 29	Comunicaciones
## 30	Ocio y cultura
## 31	Enseñanza
## 32	Restaurantes y hoteles
## 33	Otros bienes y servicios
## 34	Dato base
## 35	Variación respecto al año anterior
## 36	Variación respecto al año base
## 37	Gasto total
## 38	Distribución porcentual
## 39	Gasto medio por hogar
## 40	Gasto medio por persona
## 41	Gasto medio por unidad de consumo
##	Codigo
## 1	00
## 2	01
## 3	02
## 4	03
## 5	04
## 6	05
## 7	06
## 8	07
## 9	08
## 10	09
## 11	10
## 12	11
## 13	12
## 14	13
## 15	14
## 16	15
## 17	16
## 18	17
## 19	18
## 20	19
## 21	00
## 22	01
## 23	02
## 24	03
## 25	04
## 26	05
## 27	06
## 28	07
## 29	08
## 30	09
## 31	10
## 32	11
## 33	12
## 34	
## 35	
## 36	

```
## 37
## 38
## 39
## 40
## 41
```

Vemos que necesitamos filtrar por:

- **Geografía:** Total Nacional.
- **ECOICOP** (2 dígitos): Todos.
- **Tipo de dato:** Dato base.
- **Gasto:** Gasto medio por persona.

Es decir, lo ponemos en formato de lista para poder pasarlo por la función `get_data_table()` como filtro.

```
#Filtro
filter_ecoicop_1 <- list(
  "349" = "16473", # Total Nacional
  "762" = "", # Todos GRUPOS de 2 dígitos del ECOICOP
  "3" = "72", # Dato Base
  "57" = "8859" # Gasto medio por persona
) # Dato base

# Obtenemos datos
# nlast=1 para devolver sólo para el último periodo
data_ecoicop_1 <- get_data_table(
  filter = filter_ecoicop_1, idTable = 25143, nlast = 1, tip = "AM", unnest = TRUE,
  metacodes = TRUE, validate = FALSE, metanames = TRUE
)
```

Como hemos dicho que queremos que en el gráfico se pueda mostrar cómo se distribuye el gasto también dentro de cada grupo de ECOICOP, necesitamos un dígito más de desagregación en la clasificación ECOICOP. Buscamos de nuevo en la web del INE en que tabla se encuentran estos datos y filtramos de nuevo por lo pertinente.

```
# Vemos por que variables se puede filtrar
get_metadata_table_varval(25144)
```

```
##      Id Fk_Variable
## 1  16473      349
## 2   8997       70
## 3   8998       70
## 4   8999       70
## 5   9000       70
## 6   9001       70
## 7   9002       70
## 8   9003       70
## 9   9004       70
## 10  9005       70
## 11  9006       70
## 12  9007       70
## 13  9008       70
## 14  9009       70
## 15  9010       70
```

## 16	9011	70
## 17	9012	70
## 18	9013	70
## 19	9015	70
## 20	8995	70
## 21	311041	796
## 22	310446	796
## 23	310447	796
## 24	310448	796
## 25	310449	796
## 26	310450	796
## 27	310451	796
## 28	310452	796
## 29	310453	796
## 30	310454	796
## 31	310455	796
## 32	310456	796
## 33	310457	796
## 34	310458	796
## 35	310459	796
## 36	310460	796
## 37	310461	796
## 38	310462	796
## 39	310463	796
## 40	310464	796
## 41	310465	796
## 42	310466	796
## 43	310467	796
## 44	310468	796
## 45	310469	796
## 46	310470	796
## 47	310471	796
## 48	310472	796
## 49	310473	796
## 50	310474	796
## 51	310475	796
## 52	310476	796
## 53	310477	796
## 54	310478	796
## 55	310479	796
## 56	310480	796
## 57	310481	796
## 58	310482	796
## 59	310483	796
## 60	310484	796
## 61	310485	796
## 62	72	3
## 63	8641	3
## 64	8642	3
## 65	8856	57
## 66	8857	57
## 67	8858	57
## 68	8859	57
## 69	8860	57
##		
## 1		
## 2		

## 3	
## 4	Asturi
## 5	
## 6	
## 7	
## 8	
## 9	Cas
## 10	
## 11	Com
## 12	
## 13	
## 14	Mad
## 15	
## 16	Navarra, C
## 17	
## 18	
## 19	
## 20	
## 21	
## 22	
## 23	Bebid
## 24	Be
## 25	
## 26	
## 27	
## 28	Alquileres r
## 29	Alquileres impu
## 30	Conservación y reparaci
## 31	Suministro de agua y otros servicios relacionado
## 32	Electricidad, gas y o
## 33	Muebles y accesorios, alfombras y otros revesti
## 34	Artículos texti
## 35	Ap
## 36	Menaje y herra
## 37	Bienes y servicios para el mantenimiento co
## 38	Productos, aparatos
## 39	Servicios médicos, ambulatorio
## 40	Adquisi
## 41	Utilización de veh
## 42	Servic
## 43	Equipos d
## 44	Servicios postales, d
## 45	Equipos audiovisuales, fotográficos y de procesamien
## 46	Otros bienes para el ocio, el deporte y la cultura (juegos, instrumentos musicales, jardinería
## 47	Servicios recreativos, deport
## 48	Prensa, libros y artíc
## 49	Pa
## 50	Enseñanza In
## 51	Enseñanza Secundaria y Postsecun
## 52	E
## 53	Enseñanzas no d
## 54	Restaur
## 55	Servici
## 56	
## 57	Efectos pe
## 58	
## 59	

## 60		Servicios fin
## 61		Otros s
## 62		
## 63		Variación respect
## 64		Variación res
## 65		
## 66		Distri
## 67		Gast
## 68		Gasto :
## 69		Gasto medio por
##	Codigo	
## 1	00	
## 2	01	
## 3	02	
## 4	03	
## 5	04	
## 6	05	
## 7	06	
## 8	07	
## 9	08	
## 10	09	
## 11	10	
## 12	11	
## 13	12	
## 14	13	
## 15	14	
## 16	15	
## 17	16	
## 18	17	
## 19	18	
## 20	19	
## 21	00	
## 22	01.1 T	
## 23	01.2 T	
## 24	02.1 T	
## 25	02.2 T	
## 26	03.1 T	
## 27	03.2 T	
## 28	04.1 T	
## 29	04.2 T	
## 30	04.3 T	
## 31	04.4 T	
## 32	04.5 T	
## 33	05.1 T	
## 34	05.2 T	
## 35	05.3 T	
## 36	05.4 T	
## 37	05.5 T	
## 38	06.1 T	
## 39	06.2 T	
## 40	07.1 T	
## 41	07.2 T	
## 42	07.3 T	
## 43	08.1 T	
## 44	08.2 T	
## 45	09.1 T	
## 46	09.2 T	

```
## 47 09.3 T
## 48 09.4 T
## 49 09.5 T
## 50 10.1 T
## 51 10.2 T
## 52 10.3 T
## 53 10.4 T
## 54 11.1 T
## 55 11.2 T
## 56 12.1 T
## 57 12.2 T
## 58 12.3 T
## 59 12.4 T
## 60 12.5 T
## 61 12.6 T
## 62
## 63
## 64
## 65
## 66
## 67
## 68
## 69
```

```
filter_ecoicop_2 <- list(
  "349" = "16473", # Total Nacional
  "796" = "", # Todos GRUPOS de 2 dígitos del ECOICOP
  "3" = "72", # Dato base
  "57" = "8859" # Gasto medio por persona
)

# Obtenemos datos
# nlast=1 para devolver sólo para el último periodo
data_ecoicop_2 <- get_data_table(filter = filter_ecoicop_2, idTable = 25144, nlast = 1, tip = "AM",
```

Una vez obtenidos todos los datos, debemos juntar los datos en un data frame que contenga la siguiente estructura:

- **id**: único para cada fila
- **label**: etiqueta con cada grupo de ECOICOP
- **parent**: indica quien es el superior jerárquico de dicha fila. Por ejemplo de ECOICOP 01.1 su parent es 01. Es necesario para poder mapear bien el gráfico.
- **values**: indica el valor que toma dicha entrada.

```
# Datos para ECOICOP con 3 dígitos
data_ecoicop_2 <- data_ecoicop_2[, c("Subgrupos.de.gasto.(3.dígitos)", "Subgrupos.de.gasto.(3.dígitos)

# quitamos el total
data_ecoicop_2 <- data_ecoicop_2[data_ecoicop_2[, "Subgrupos.de.gasto.(3.dígitos).Codigo"] != "00", ]

# Función que a cada elemento de ECOICOP 3 dígitos, asocia su grupo de agregación de 2 dígitos
get_label <- function(inputs, label) {
```

```

# Extrae los primeros dos caracteres de cada elemento del vector de entrada
prefixes <- substr(inputs, 1, 2) #tomamos sólo 2 primeros eltos Ej: 01.1 tomamos 01
# Usa sapply para aplicar la función a cada prefijo
sapply(prefixes, function(prefix) {
  if (prefix == "01") {
    return(label[2])
  } else if (prefix == "02") {
    return(label[3])
  } else if (prefix == "03") {
    return(label[4])
  } else if (prefix == "04") {
    return(label[5])
  } else if (prefix == "05") {
    return(label[6])
  } else if (prefix == "06") {
    return(label[7])
  } else if (prefix == "07") {
    return(label[8])
  } else if (prefix == "08") {
    return(label[9])
  } else if (prefix == "09") {
    return(label[10])
  } else if (prefix == "10") {
    return(label[11])
  } else if (prefix == "11") {
    return(label[12])
  } else if (prefix == "12") {
    return(label[13])
  } else {
    return(NA) # Devuelve NA si el prefijo no coincide con ninguno
  }
})
}

```

```

# Aplicamos funcion a ECOICOP con 3 dígitos y con label las de ECOICOP 2 dígitos
data_ecoicop_2$parent <- get_label(data_ecoicop_2$`Subgrupos.de.gasto.(3.dígitos).Codigo`, label = d

```

```

# Etiquetas
labels <- c(data_ecoicop_1$`Grupos.de.gasto.(2.dígitos)`, data_ecoicop_2$`Subgrupos.de.gasto.(3.dígitos)`)

# parents, el del primer elemento es , es decir, el total. Los del primer nivel
# el parent es Índice Total=labels[1]
# En el segundo nivel, ya lo que corresponda
aux <- data_ecoicop_1$`Grupos.de.gasto.(2.dígitos)`
parents<-c("", rep(aux[1], length(aux) - 1), data_ecoicop_2$parent)

```

```

fig <- plot_ly(
  type = "treemap",
  labels = labels,
  parents = parents,
  textinfo = "label+value+percent entry+percent root",
  hoverinfo = "label+value+percent entry+percent root",
  values = as.integer(c(data_ecoicop_1$Valor, data_ecoicop_2$Valor),0),
  maxdepth = 2,
  branchvalues='total',
  tiling = list(squarifyratio = 2))

```

```
# Importante ajustar este numero para que quede bontito
#Quitarlo para ver que pasa sin el
```

```
fig
```

Índice general

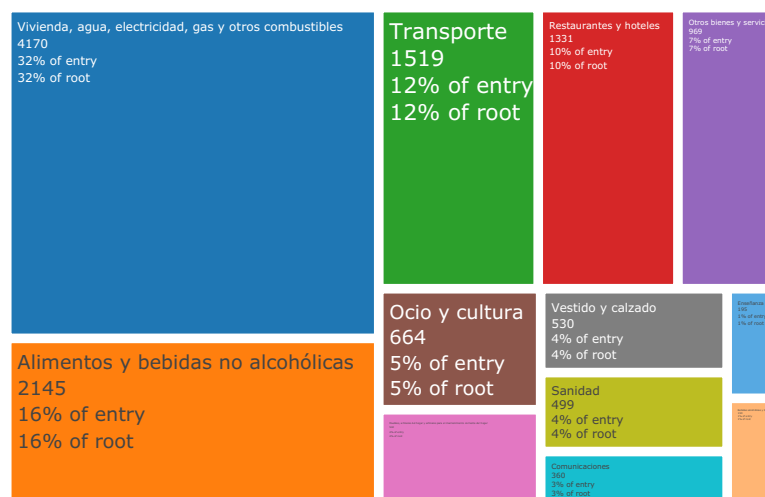


Figure 4.1: sgf

4.3.3 Series Temporales

Plotly ofrece potentes herramientas para la visualización de series temporales, permitiendo a los usuarios interactuar con los datos de manera dinámica y obtener información valiosa de forma rápida y efectiva. Su capacidad para crear gráficos atractivos y funcionales lo convierte en una opción excelente para el análisis de datos temporales.

Características Clave de Plotly para Series Temporales

- **Interactividad:**
 - **Zoom y Pan:** Los usuarios pueden acercar y alejar secciones del gráfico.
 - **Hover:** Información detallada aparece al pasar el cursor sobre los puntos de datos.
 - **Selección de Rangos:** Rangos deslizantes y botones de selector para ajustar la ventana temporal visualizada.
- **Elementos Visuales**
 - **Líneas y Marcadores:** Visualización clara de tendencias y puntos específicos.
 - **Colores y Estilos Personalizables:** Facilitan la distinción entre diferentes series de datos.
- **Funciones Avanzadas:**
 - **Rangoselector:** Botones para seleccionar periodos específicos (e.g., últimos 3 meses, último año).

- **Rangeslider**: Barra deslizante para ajustar el rango de fechas mostrado.
- **Anotaciones y Formateo**: Posibilidad de agregar notas y ajustar el formato de fechas.

Los datos los obtendremos usando `ineapir`, para más información (véase (Crespo, 2024)).

1. En la web del ine **buscamos la serie que contiene los datos deseados**, población de españa, que tiene por id ECP319(hombres) y ECP318(mujeres).
2. Usando `ineapir` extraemos los datos usando el servicio API del INE.

```
# Población Hombres
# Mirando en la web del INE cual es el código de la operación
# ECP319

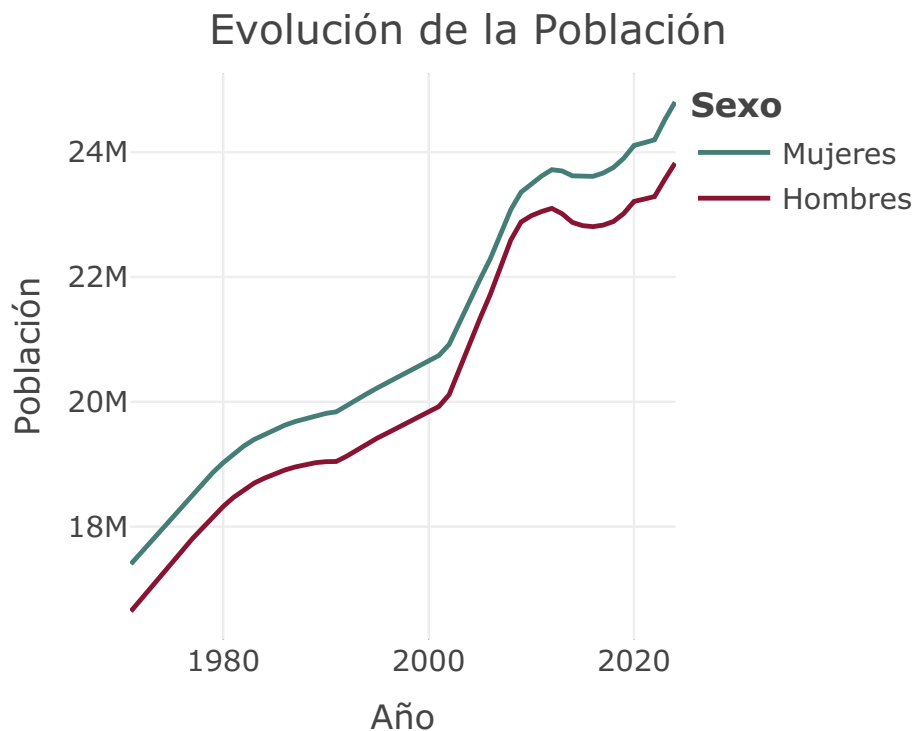
# Hombres
a <- get_data_series(codSeries = "ECP319", tip = "AM", dateStart = "1971/01/01", unnest = TRUE)
# Mujeres
b <- get_data_series(codSeries = "ECP318", tip = "AM", dateStart = "1971/01/01", unnest = TRUE)

data <- data.frame(fecha = as.Date(a$Fecha), Year = a$Anyo, Hombres = a$Valor, Mujeres = b$Valor)

# Filtramos
data <- data %>%
  filter(format(as.Date(fecha), "%m") == "01")
```

```
# Crear el gráfico interactivo
fig <- plot_ly(data, x = ~Year) %>%
  add_lines(y = ~Mujeres, name = "Mujeres", line = list(color = "#457e76")) %>%
  add_lines(y = ~Hombres, name = "Hombres", line = list(color = "#881333")) %>%
  layout(
    title = "Evolución de la Población",
    xaxis = list(
      title = "Año"
    ),
    yaxis = list(title = "Población"),
    legend = list(title = list(text = "<b>Sexo</b>")),
    hovermode = "closest"
  ) %>%
  config(locale = 'es')

# Mostrar el gráfico
fig
```



4.3.3.1 Evolución IPC

Ahora vamos a mostrar la evolución del IPC y el IPC Subyacente a lo largo de los años. Para ello, lo primero buscamos en la web del INE en que tabla se encuentran dichos datos, siendo esta la “50902”. A continuación, con `get_metadata_table_varval(50902)` vemos que variables y valores hay que tomar para filtrar IPC por Índice General y por Variación Anual.

1. En la web del ine **buscamos la tabla que contiene los datos deseados**, que tiene por id 50902. (Véase <https://www.ine.es/jaxiT3/Tabla.htm?t=50902&L=0>)
2. Usando `ineapir` extraemos los datos usando el servicio API del INE. Ejecutando `get_metadata_table_varval(50902)` vemos que debemos filtrar por Índice general, y variación anual.

```
get_metadata_table_varval(50902)
```

```
##          Id Fk_Variable
## 1  304092          762
## 2  304093          762
## 3  304094          762
## 4  304095          762
## 5  304096          762
## 6  304097          762
## 7  304098          762
## 8  304099          762
## 9  304100          762
## 10 304101          762
## 11 304102          762
## 12 304103          762
## 13 304104          762
## 14   83             3
## 15   84             3
## 16   74             3
```

```
## 17      87      3
##
##                                     Nombre
## 1                                     Índice general
## 2                                     Alimentos y bebidas no alcohólicas
## 3                                     Bebidas alcohólicas y tabaco
## 4                                     Vestido y calzado
## 5                                     Vivienda, agua, electricidad, gas y otros combustibles
## 6 Muebles, artículos del hogar y artículos para el mantenimiento corriente del hogar
## 7                                     Sanidad
## 8                                     Transporte
## 9                                     Comunicaciones
## 10                                    Ocio y cultura
## 11                                    Enseñanza
## 12                                    Restaurantes y hoteles
## 13                                    Otros bienes y servicios
## 14                                    Índice
## 15                                    Variación mensual
## 16                                    Variación anual
## 17                                    Variación en lo que va de año
##      Codigo
## 1      00
## 2      01
## 3      02
## 4      03
## 5      04
## 6      05
## 7      06
## 8      07
## 9      08
## 10     09
## 11     10
## 12     11
## 13     12
## 14      0
## 15      1
## 16      2
## 17      5
```

```
# get_metadata_variables()
# Filter cpi overall index
# Todos los approaches de filtrado son similares
# 762(grupo)= 304092(Indie general).      3(tipodato)=74(Variacion anual)
filter <- list(values = c("variación anual", "índice general"))
filter <- list(grupo = "304092", tipodato = "74")
filter <- list("762" = "304092", "3" = "74")

# IPC GENERAL
# Table url: https://www.ine.es/jaxiT3/Tabla.htm?t=50902&L=0
general <- get_data_table(
  idTable = 50902, filter = filter, unnest = TRUE,
  tip = "A", validate = FALSE
)

# IPC SUBYACENTE
get_metadata_table_varval(50907)
```

##	Id	Fk_Variable
## 1	304092	762
## 2	12859	269
## 3	12833	269
## 4	12834	269
## 5	12835	269
## 6	12836	269
## 7	12860	269
## 8	12837	269
## 9	12848	269
## 10	12838	269
## 11	12839	269
## 12	12849	269
## 13	12850	269
## 14	12840	269
## 15	12851	269
## 16	12841	269
## 17	12852	269
## 18	12842	269
## 19	12853	269
## 20	12843	269
## 21	12854	269
## 22	12844	269
## 23	12845	269
## 24	12855	269
## 25	12856	269
## 26	12846	269
## 27	12857	269
## 28	12847	269
## 29	12858	269
## 30	303820	269
## 31	83	3
## 32	84	3
## 33	74	3
## 34	87	3

##	Nombre
## 1	Índice general
## 2	Alimentos con elaboración, bebidas y tabaco
## 3	Alimentos sin elaboración
## 4	Bienes industriales
## 5	Servicios sin alquiler de vivienda
## 6	Carburantes y combustibles
## 7	Bienes industriales duraderos
## 8	Productos energéticos
## 9	General sin alimentos, bebidas y tabaco
## 10	General sin alquiler de vivienda
## 11	General sin carburantes ni combustibles
## 12	General sin productos energéticos
## 13	General sin alimentos no elaborados ni productos energéticos
## 14	General sin servicios (incluye alquiler de vivienda)
## 15	General sin servicios ni alquiler de vivienda
## 16	General sin alimentos ni bebidas ni tabaco ni alquiler de vivienda
## 17	General sin alimentos ni bebidas ni tabaco ni carburantes ni combustibles
## 18	General sin alimentos ni bebidas ni tabaco ni productos energéticos
## 19	General sin alimentos ni bebidas ni tabaco ni servicios (incluido alquiler de vivienda)
## 20	General sin alimentos ni bebidas ni tabaco ni servicios ni alquiler de vivienda
## 21	Bienes industriales no duraderos


```

## 22                                     General sin tabaco
## 23                                     Alimentos y bebidas
## 24                                     Alimentos elaborados
## 25                                     Bienes industriales sin energía
## 26                                     Servicios
## 27                                     Alimentos, bebidas y tabaco
## 28                                     Alimentos no elaborados y productos energéticos
## 29                                     Bienes industriales sin productos energéticos
## 30                                     General sin carburantes ni combustibles líquidos
## 31                                     Índice
## 32                                     Variación mensual
## 33                                     Variación anual
## 34                                     Variación en lo que va de año
##      Codigo
## 1      00
## 2      01
## 3      02
## 4      03
## 5      04
## 6      05
## 7      06
## 8      07
## 9      08
## 10     09
## 11     10
## 12     11
## 13     12
## 14     13
## 15     14
## 16     15
## 17     16
## 18     17
## 19     18
## 20     19
## 21     20
## 22     21
## 23     22
## 24     23
## 25     24
## 26     25
## 27     26
## 28     27
## 29     28
## 30     29
## 31      0
## 32      1
## 33      2
## 34      5

```

```

filter <- list(
  "3" = "74", # Fk_variable=Id
  "269" = "12850" # Fk_variable=Id
)
# Filter core cpi
filter <- list(values = c(
  "variación anual",
  "General sin alimentos no elaborados ni productos energéticos"

```

```

))

# Request data of core cpi
# Table url: https://www.ine.es/jaxiT3/Tabla.htm?t=50907&L=0
subyacente <- get_data_table(
  idTable = 50907, filter = filter, unnest = TRUE,
  tip = "A", validate = FALSE
)

# Format Fecha column as date
general$Fecha <- as.Date(general$Fecha)
subyacente$Fecha <- as.Date(subyacente$Fecha)

```

Una vez hemos preparado los datos, los dibujamos en un gráfico interactivo:

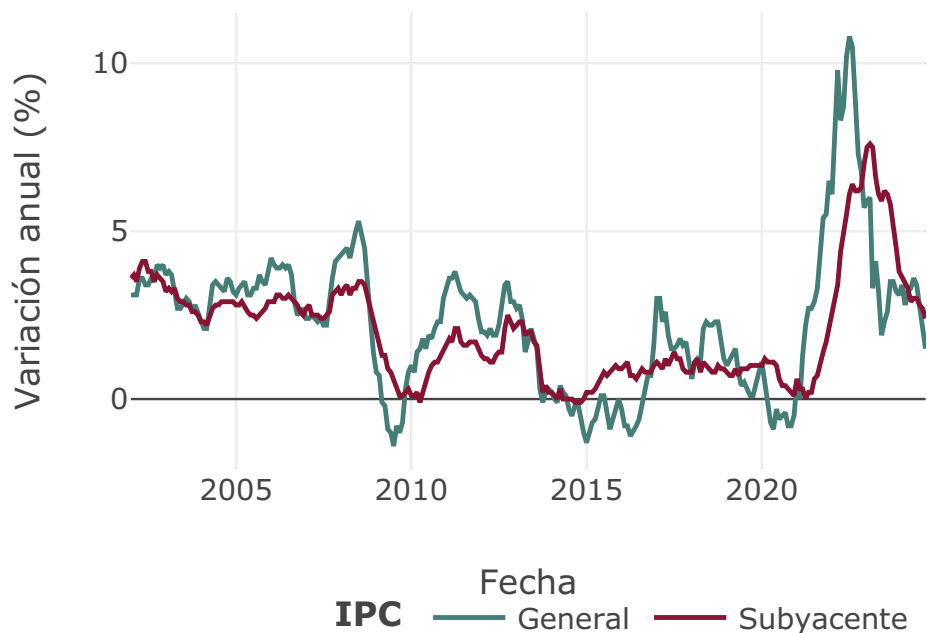
```

# Plot cpi overall index
fig <- plot_ly(general,
  x = ~Fecha, y = ~Valor, name = "General",
  type = "scatter", mode = "lines", line = list(color = "#457e76", dash = "dashed")
)

## Plot core cpi
fig <- fig %>%
  add_trace(
    y = ~ subyacente$Valor, name = "Subyacente",
    mode = "lines", line = list(color = "#881333", dash = "dashed")
  ) %>%
  layout(
    yaxis = list(title = "Variación anual (%)" ),
    legend = list(
      title = list(text = "<b> IPC </b>"),
      x = 0.25,
      y = -0.25,
      orientation = "h"
    ),
    hovermode = "x"
  ) %>%
  config(displayModeBar = FALSE) %>%
  config(locale = 'es')

fig

```



4.3.4 Grafico de barras

Los gráficos de barras en Plotly permiten comparar fácilmente valores entre diferentes categorías. Pueden ser verticales u horizontales, y se pueden personalizar con colores específicos para cada barra. También soportan la agrupación de barras (bar charts) y la apilación (stacked bar charts). La orientación, orden y formato de los ejes pueden ser ajustados para mejorar la legibilidad, y se pueden añadir anotaciones y etiquetas detalladas para proporcionar contexto adicional a los datos visualizados.

1. En la web del ine **buscamos la tabla que contiene los datos deseados**, que tiene por id 2915. (Véase <https://www.ine.es/jaxiT3/Tabla.htm?t=2915&L=0>)
2. Usando `ineapir` extraemos los datos usando el servicio API del INE. Ejecutando `get_metadata_table_varval(2915)` vemos que debemos filtrar por todos municipios

```
get_metadata_table_varval(2915)
```

##	Id	Fk_Variable	Nombre	Codigo
## 1	16473	70	Total	00
## 2	8997	70	Andalucía	01
## 3	8998	70	Aragón	02
## 4	8999	70	Asturias, Principado de	03
## 5	9000	70	Balears, Illes	04
## 6	9001	70	Canarias	05
## 7	9002	70	Cantabria	06
## 8	9003	70	Castilla y León	07
## 9	9004	70	Castilla - La Mancha	08
## 10	9005	70	Cataluña	09
## 11	9006	70	Comunitat Valenciana	10
## 12	9007	70	Extremadura	11
## 13	9008	70	Galicia	12
## 14	9009	70	Madrid, Comunidad de	13
## 15	9010	70	Murcia, Región de	14

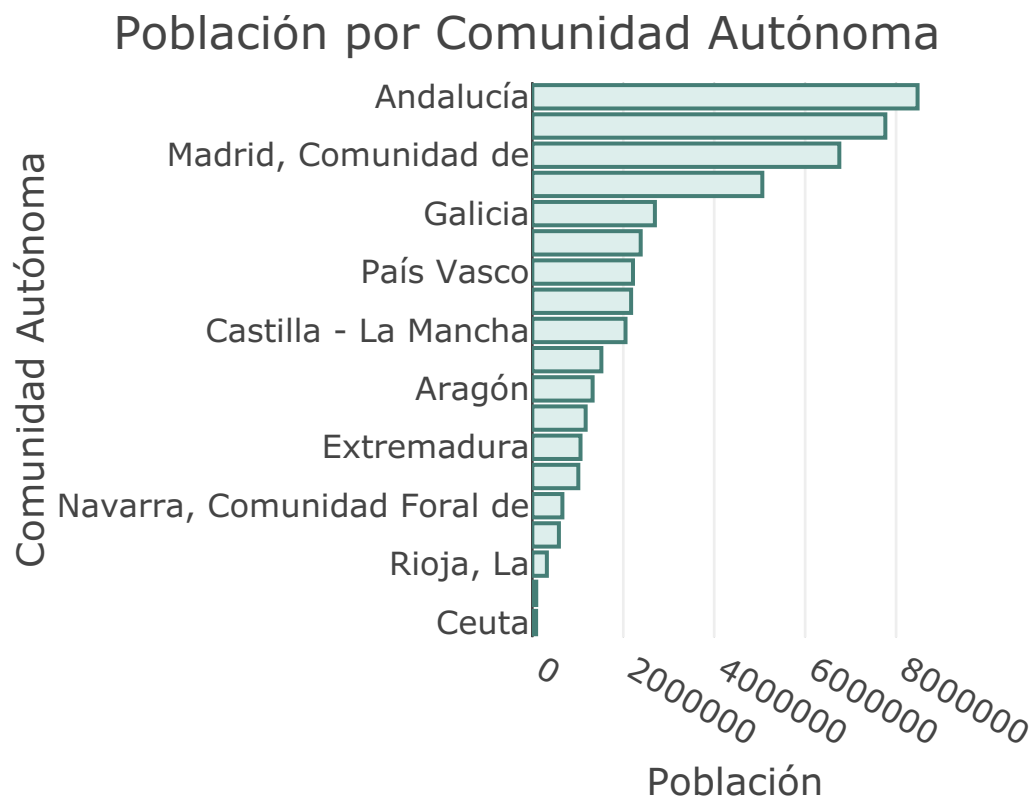
## 16	9011	70	Navarra, Comunidad Foral de	15
## 17	9012	70	País Vasco	16
## 18	9013	70	Rioja, La	17
## 19	9015	70	Ceuta	18
## 20	8995	70	Melilla	19
## 21	20275	34	Total	A
## 22	20267	34	Menos de 101	B
## 23	20264	34	De 101 a 500	C
## 24	20270	34	De 501 a 1.000	D
## 25	20260	34	De 1.001 a 2.000	E
## 26	20262	34	De 2.001 a 5.000	F
## 27	20274	34	De 5.001 a 10.000	G
## 28	20273	34	De 10.001 a 20.000	H
## 29	20263	34	De 20.001 a 50.000	I
## 30	20265	34	De 50.001 a 100.000	J
## 31	20266	34	De 100.001 a 500.000	K
## 32	20268	34	Más de 500.000	L

```
filter <- list("34" = "20275") # 34(Tamaño municipios)=20275(Total)
data2 <- get_data_table(2915, filter = filter, unnest = TRUE, tip = "AM", nlast = 1, metanames = TRUE)
data2 <- data2[-1, ] # Quitamos primera observación. Total España
```

```
# Ordenar el dataframe por la columna 'poblacion'
data2 <- data.frame(poblacion = data2$Valor, ccaa = data2$Comunidades.y.Ciudades.Autónomas)
data2 <- data2[order(data2$poblacion), ]

# Crear el gráfico de barras horizontales
fig <- plot_ly(data2,
  x = ~poblacion, y = ~ factor(ccaa, levels = data2$ccaa), type = "bar", orientation = "h",
  marker = list(color = "#ddeeec", line = list(color = "#457e76", width = 1.5))
) %>%
  layout(
    title = "Población por Comunidad Autónoma",
    xaxis = list(
      title = "Población",
      tickformat = ".,", # Formatear los números para que se muestren con punto como separador de miles
      tickprefix = " ", # Espacio para mejorar la legibilidad
      separatethousands = TRUE # Asegurar que los miles se separen correctamente
    ),
    yaxis = list(title = "Comunidad Autónoma"),
    margin = list(l = 150) # Ajustar el margen izquierdo para evitar que las etiquetas se corten
  ) %>%
  config(locale = 'es')
```

```
# Mostrar el gráfico
fig
```



4.3.5 Gráfico Sankey

Este tipo de gráfico es poco conocido pero muy útil para visualizar ciertos fenómenos sociales/demográficos. Por ejemplo: mostrar la evolución de la intención de voto, representando como se han desplazado las intenciones entre diferentes periodos; mostrar la evolución de la situación laboral de la comunidad, mostrando los flujos entre distintos estados laborales en distintos periodos.

Veamos pues este último gráfico. El objetivo es intentar reproducir la aplicación del INE [Flujos de personas en el mercado laboral](#). En este caso, dentro de bookdown, la aplicación no se puede hacer tan dinámica como la presentada en la web del INE, ya que el formato no es el adecuado para ello ².

1. En la web del ine **buscamos la tabla que contiene los datos deseados**, que tiene por id 13930. (Véase <https://www.ine.es/jaxiT3/Tabla.htm?t=66257&L=0>)
2. Usando `ineapir` extraemos los datos usando el servicio API del INE. Ejecutando `get_metadata_table_varval(66257)` vemos que debemos filtrar por ambos sexos

```
# Tabla 66257. Información del mercado laboral
# Vemos sus metadatos para ver por que filtrar
get_metadata_table_varval(66257)
```

##	Id	Fk_Variable	Nombre	Codigo
## 1	454	18	Ambos sexos	
## 2	452	18	Hombres	1
## 3	453	18	Mujeres	6
## 4	283949	386	Total (trimestre actual)	
## 5	21332	386	Ocupados (trimestre actual)	
## 6	21335	386	Parados (trimestre actual)	

²Para ver una aplicación dinámica como la presente en la web del INE, ver apartado de Shiny Apps o FlexDashboards.

```
## 7    20279          386    Inactivos (trimestre actual)
## 8    283949         539      Total (trimestre anterior)
## 9     21332         539    Ocupados (trimestre anterior)
## 10   21335         539    Parados (trimestre anterior)
## 11   20279         539 Inactivos (trimestre anterior)
## 12  283997         539 No consta (trimestre anterior)
```

```
filter <- list("18" = "454") # "Sexo"="Ambos sexos"

datos <- get_data_table(66257, filter = filter, metanames = TRUE, tip = "AM", metacodes = TRUE, unne

# Quitamos los datos totales
datos <- datos %>%
  filter(Relación.con.la.actividad.en.el.trimestre.actual.Id != "283949") %>%
  filter(Relación.con.la.actividad.en.trimestre.anterior.Id != "283949")
```

Para el sankey plot hay que crear un “diccionario” de etiquetas para saber como pintarlas. Véase documentación para más información.

```
#Necesario para el sankey plot
# Crear etiquetas únicas
labels <- unique(c(
  datos$Relación.con.la.actividad.en.trimestre.anterior,
  datos$Relación.con.la.actividad.en.el.trimestre.actual
))
labels <- c(labels[3], labels[1], labels[2], labels[4], labels[7], labels[5], labels[6])
# Crear un dataframe para mapear las etiquetas a índices
label_map <- data.frame(label = labels, id = seq_along(labels) - 1)
```

```
# Hacemos join con las etiquetas
filtered_data <- datos %>%
  left_join(label_map, by = c("Relación.con.la.actividad.en.trimestre.anterior" = "label")) %>%
  rename(source = id) %>%
  left_join(label_map, by = c("Relación.con.la.actividad.en.el.trimestre.actual" = "label")) %>%
  rename(target = id)
```

```
# Asignar colores a los enlaces (opcional)
link_colors <- ifelse(filtered_data$source == 0, "#FFD700", #Inactivos
  ifelse(filtered_data$source == 1, "#CD853F", #Ocupados
    ifelse(filtered_data$source == 2, "#FF4500", #Parados
      "#9e3a26" #No consta
    )))

fig <- plot_ly(
  type = "sankey",
  orientation = "h",
  node = list(
    label = labels,
    color = c("#FFD700", "#CD853F", "#FF4500", "#9e3a26", "#FFD700", "#CD853F", "#FF4500", "#9e3a26"),
    pad = 15,
    thickness = 20,
    line = list(color = "black", width = 0.5)
  ),
  link = list(
    source = filtered_data$source,
```

```

    target = filtered_data$target,
    value = filtered_data$Valor,
    color = link_colors
  )
)

fig <- fig %>%
  layout(
    title = "Transición de Actividades Económicas entre Trimestres",
    font = list(size = 10)
  )%>%
  config(locale = 'es')
fig

```

Transición de Actividades Económicas entre Trimestres

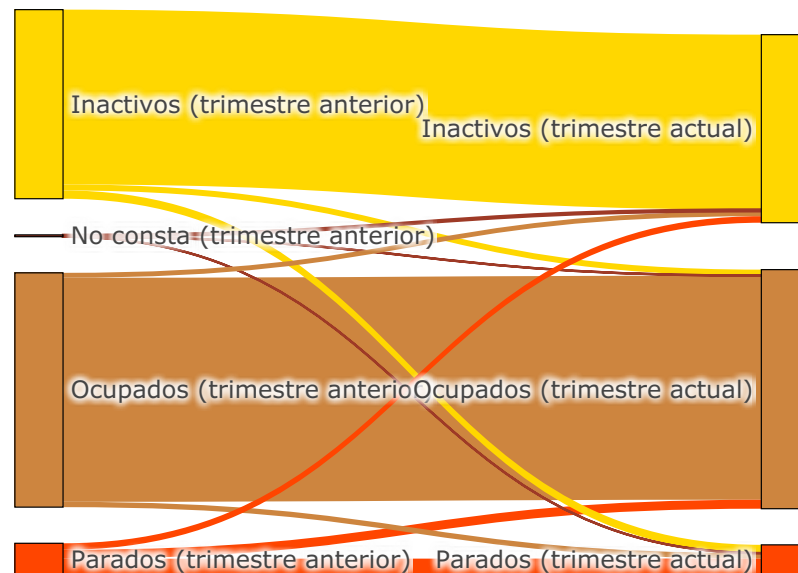


Figure 4.2: Flujos de personas en el mercado laboral en España 2024 T2

4.4 Más información

Para más información sobre el paquete plotly y sus posibles funciones, buscar en la siguiente web:

<https://plotly.com/r/getting-started/>

Capítulo 5

SF

fsda

Capítulo 6

jk

fdas

Capítulo 7

Leaflet

[Leaflet](#) es un paquete que permite integrar mapas interactivos en aplicaciones web y en documentos R. Utiliza la biblioteca JavaScript Leaflet, lo que proporciona una gran cantidad de funcionalidades para visualizar datos geoespaciales de forma intuitiva. Con Leaflet, puedes agregar marcadores, polígonos, capas de datos y muchos otros elementos a tus mapas. Su simplicidad y flexibilidad lo hacen ideal para visualizaciones geográficas en proyectos de análisis de datos.

7.1 General

7.1.1 Características Principales de Leaflet

1. **Interactividad:** Leaflet permite crear mapas interactivos donde los usuarios pueden acercar, alejar y hacer clic en elementos para obtener información adicional a través de ventanas emergentes (popups).
2. **Capas Múltiples:** Puedes agregar diferentes tipos de capas al mapa, como capas de imagen, capas de mosaico, y datos vectoriales, lo que permite superponer información.
3. **Marcadores Personalizables:** Leaflet ofrece la posibilidad de personalizar los marcadores (íconos, colores, tamaños) y añadir popups informativos para cada uno.
4. **Soporte para GeoJSON:** Permite la importación y visualización de datos geoespaciales en formato GeoJSON, facilitando la representación de geometrías complejas.
5. **Simplicidad de Uso:** La sintaxis es intuitiva y sencilla, lo que permite a los usuarios crear mapas rápidamente con un código mínimo.
6. **Compatibilidad con Dispositivos Móviles:** Los mapas son responsivos y funcionan bien en dispositivos móviles, garantizando una buena experiencia de usuario.
7. **Extensibilidad:** Leaflet puede ser ampliado mediante complementos (plugins) que permiten agregar nuevas funcionalidades, como gráficos de tiempo, clusters de puntos, y más.
8. **Temas Personalizables:** Se pueden aplicar diferentes estilos y temas a los mapas para adaptarlos a la visualización deseada.

7.1.2 Tipos de Gráficos en Leaflet

1. **Mapas Base:**
 - Mapas de mosaico (Tiles): Proporcionan una capa base de fondo (OpenStreetMap, Google Maps, etc.).

- Mapas de satélite y otros estilos personalizados.

2. Marcadores:

- **Marcadores Simples:** Puntos que indican ubicaciones específicas en el mapa.
- **Marcadores Agrupados:** Agrupan varios marcadores en una sola vista para evitar el desorden visual (clustering).

3. Polígonos y Líneas:

- **Polígonos:** Representan áreas geográficas específicas (ejemplo: límites de comunidades autónomas).
- **Líneas:** Representan rutas o trayectorias, como carreteras o caminos.

4. Capas de Calor (Heatmaps):

- Visualizan la densidad de puntos en un área específica, utilizando colores para representar la intensidad.

5. Gráficos de Tiempos:

- Permiten visualizar datos a lo largo del tiempo, añadiendo un componente temporal a los mapas.

6. Visualización de Datos:

- **Data Binding:** Vincula datos a los marcadores o polígonos para mostrar información adicional (por ejemplo, población, estadísticas).

7. Mapas de Coropletas:

- Representan datos cuantitativos a través de variaciones en el color de las áreas geográficas, ideal para visualizar indicadores como la población o la densidad de un fenómeno.

7.2 Mapas

7.2.1 Ejemplo esperanza de vida

Vamos a representar la esperanza de vida por comunidades autónomas en un mapa de España. Para ello necesitamos **extraer primero los datos**, que como ya se ha comentado previamente se obtienen a partir del paquete `ineapir`.

1. En la web del ine **buscamos la tabla que contiene los datos deseados**, que tiene por id 27154. (Véase <https://www.ine.es/jaxiT3/Tabla.htm?t=13930&L=0>)
2. Usando `ineapir` extraemos los datos usando el servicio API del INE. Aplicando `get_metadata_table_varval(27154)` vemos que necesitamos filtrar por “sexo”=“total”, “0 años” y “esperanza de vida”, todo ello para todas comunidades autónomas.

```
# Filtrar por esperanza de vida
# "" para indicar que tome todos los valores
get_metadata_table_varval(27154)
```

##	Id	Fk_Variable	Nombre	Codigo
## 1	8997	70	Andalucía	01
## 2	8998	70	Aragón	02
## 3	8999	70	Asturias, Principado de	03
## 4	9000	70	Balears, Illes	04
## 5	9001	70	Canarias	05
## 6	9002	70	Cantabria	06
## 7	9003	70	Castilla y León	07
## 8	9004	70	Castilla - La Mancha	08
## 9	9005	70	Cataluña	09
## 10	9006	70	Comunitat Valenciana	10
## 11	9007	70	Extremadura	11
## 12	9008	70	Galicia	12
## 13	9009	70	Madrid, Comunidad de	13
## 14	9010	70	Murcia, Región de	14
## 15	9011	70	Navarra, Comunidad Foral de	15
## 16	9012	70	País Vasco	16
## 17	9013	70	Rioja, La	17
## 18	9015	70	Ceuta	18
## 19	8995	70	Melilla	19
## 20	451	18	Total	
## 21	452	18	Hombres	1
## 22	453	18	Mujeres	6
## 23	15319	355	0 años	Y0
## 24	15659	360	De 1 a 4 años	Y1T4
## 25	15664	360	De 5 a 9 años	Y5T9
## 26	15660	360	De 10 a 14 años	Y10T14
## 27	15665	360	De 15 a 19 años	Y15T19
## 28	15661	360	De 20 a 24 años	Y20T24
## 29	15666	360	De 25 a 29 años	Y25T29
## 30	15662	360	De 30 a 34 años	Y30T34
## 31	15658	360	De 35 a 39 años	Y35T39
## 32	15663	360	De 40 a 44 años	Y40T44
## 33	15667	360	De 45 a 49 años	Y45T49
## 34	15648	360	De 50 a 54 años	Y50T54
## 35	15649	360	De 55 a 59 años	Y55T59
## 36	15650	360	De 60 a 64 años	Y60T64
## 37	15651	360	De 65 a 69 años	Y65T69
## 38	15652	360	De 70 a 74 años	Y70T74
## 39	15653	360	De 75 a 79 años	Y75T79
## 40	15654	360	De 80 a 84 años	Y80T84
## 41	15655	360	De 85 a 89 años	Y85T89
## 42	284344	357	90 y más años	Y-GE90
## 43	15656	360	De 90 a 94 años	Y90T94
## 44	15657	357	95 y más años	Y-GE95
## 45	22189	260	Tasa de mortalidad	
## 46	311510	260	Promedio de años vividos el último año de vida	
## 47	311511	260	Riesgo de muerte	
## 48	311512	260	Supervivientes	
## 49	311513	260	Defunciones teóricas	
## 50	311514	260	Población estacionaria	
## 51	311515	260	Tiempo por vivir	
## 52	284290	260	Esperanza de vida	

```
filter = list(ccaa = "", sexo = "total", values = c("0 años", "esperanza de vida"))
```

Tabl de mortalidad por año, ccaa, ciudadaes, sexo, edad y funciones.

```
# Table url: https://www.ine.es/jaxiT3/Tabla.htm?t=27154&L=0
esp <- get_data_table(idTable = 27154, filter = filter, nlast = 1, unnest = TRUE,
                      metacodes = TRUE, tip = "AM", validate = FALSE)

# Seleccionamos columnas de interés
esp <- subset(esp, select = c("Comunidades.y.Ciudades.Autónomas.Id", "T3_Periodo",
                              "Anyo", "Valor"))
```

7.2.1.1 Contornos

Una vez obtenidos los datos que se quieren incluir en el mapa, es necesario **disponer de los contornos geográficos** sobre los cuales se quieren representar dichos datos. En este caso, los contornos de las comunidades autónomas.

El ine dispone acceso de contornos para los límites geográficos en la siguiente web:



Web <https://www.ine.es/wstempus/geojs/ES/CONTORNOS/XXXX>
sustituyendo XXXX por el código del ámbito geográfico necesario, siendo estos:

- 349: total nacional
- 70: ccaa
- 115: provincias
- 19: municipios

Para colorear el mapa, se desea que se muestren colores para intervalos de Esperanza de Vida (tomando los cuartiles), para ello será necesario usar la función `colorBin()`. Por tanto, tenemos:

```
# Verificar si los paquetes están instalados
if (!require("leaflet")) install.packages("leaflet")
if (!require("sf")) install.packages("sf")
if (!require("htmltools")) install.packages("htmltools")

# Cargar paquete
library(leaflet)
library(sf)
library(htmltools)

# Contornos de las ccaa
ccaa <- read_sf("https://www.ine.es/wstempus/geojs/ES/CONTORNOS/70")

# join de los contornos y el dataset
ccaa <- merge(ccaa, esp, by.x = "id_region",
              by.y = "Comunidades.y.Ciudades.Autónomas.Id" )

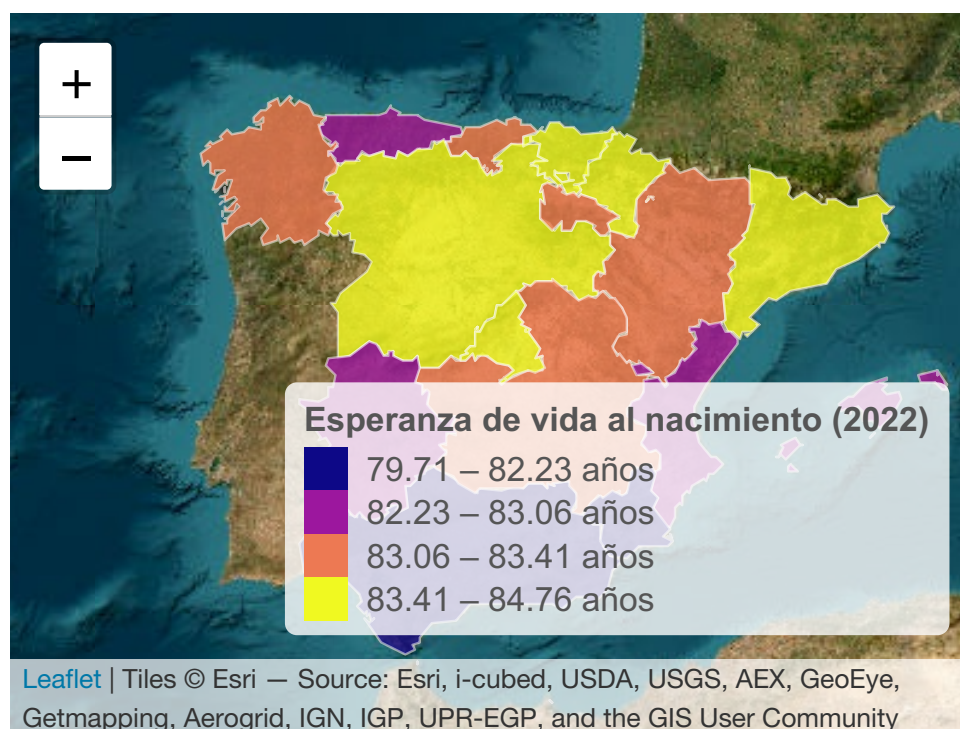
# Creamos colores para cada rango de esperanza de vida
pal <- colorBin("plasma", domain = NULL, bins = c(quantile(ccaa$Valor)))

# Labels of the map
labels <- sprintf(
  "<strong>%s</strong><br/> Esperanza de vida al nacimiento: %.2f años ",
  ccaa$nom_region, ccaa$Val
) %>% lapply(htmltools::HTML)
```



```
# Create the map
m <- leaflet(ccaa) %>%
  addProviderTiles(providers$Esri.WorldImagery) %>%
  setView(-4, 40, zoom = 5) %>%
  addPolygons(fillOpacity = 0.8,
              fillColor = ~pal(Valor),
              weight = 1,
              color = "white",
              label = labels,
              labelOptions = labelOptions(
                style = list("font-weight" = "normal", padding = "3px 8px"),
                textsize = "15px",
                direction = "auto"
              ),
              highlightOptions = highlightOptions(fillOpacity = 1, bringToFront = TRUE,
                                                    weight = 2, color = "white")
            ) %>%
  addLegend(pal = pal, values = ~Valor, opacity = 1.0, position = "bottomright",
            labFormat = labelFormat(suffix = " años", digits = 2),
            title = sprintf("Esperanza de vida al nacimiento (%s)",
                            esp$Anyo[1])) %>%
  config(locale = 'es')
```

m



7.2.2 Ejemplo salario medio

Ahora vamos a representar en un mapa el salario bruto mensual medio en cada comunidad autónoma para el año 2022. Para ello necesitamos:

1. En la web del ine **buscamos la tabla que contiene los datos deseados**, que tiene por id 13930. (Véase <https://www.ine.es/jaxiT3/Tabla.htm?t=13930&L=0>)

2. Usando `ineapir` extraemos los datos usando el servicio API del INE. Ejecutando `get_metadata_table_varval(13930)` vemos que debemos filtrar por jornada a tiempo completo, todas ccaa, y total decil.

```
# Filtrar por esperanza de vida
# "" para indicar que tome todos los valores
get_metadata_table_varval(13930)
```

##	Id	Fk_Variable	Nombre	Codigo
## 1	10757	120	Total	
## 2	10758	120	Jornada a tiempo completo	
## 3	10759	120	Jornada a tiempo parcial	
## 4	16473	349	Total Nacional	00
## 5	8997	70	Andalucía	01
## 6	8998	70	Aragón	02
## 7	8999	70	Asturias, Principado de	03
## 8	9000	70	Balears, Illes	04
## 9	9001	70	Canarias	05
## 10	9002	70	Cantabria	06
## 11	9003	70	Castilla y León	07
## 12	9004	70	Castilla - La Mancha	08
## 13	9005	70	Cataluña	09
## 14	9006	70	Comunitat Valenciana	10
## 15	9007	70	Extremadura	11
## 16	9008	70	Galicia	12
## 17	9009	70	Madrid, Comunidad de	13
## 18	9010	70	Murcia, Región de	14
## 19	9011	70	Navarra, Comunidad Foral de	15
## 20	9012	70	País Vasco	16
## 21	9013	70	Rioja, La	17
## 22	9015	70	Ceuta	18
## 23	8995	70	Melilla	19
## 24	298931	684	Total decil	
## 25	298932	684	1	
## 26	298933	684	2	
## 27	298934	684	3	
## 28	298935	684	4	
## 29	298936	684	5	
## 30	298937	684	6	
## 31	298938	684	7	
## 32	298939	684	8	
## 33	298940	684	9	
## 34	298941	684	10	

```
filter2 = list(
  "120" = "10758", # Jornada a tiempo completo
  "70" = "",      # Todas ccaa
  "684" = "298931" # Total decil
)

# Tabl de mortalidad por año, ccaa, ciudadaes, sexo, edad y funciones.
# Table url: https://www.ine.es/jaxiT3/Tabla.htm?t=27154&L=0
esp2 <- get_data_table(idTable = 13930, filter = filter2, nlast = 1, unnest = TRUE,
  metacodes = TRUE, tip = "AM", validate = FALSE)

# Seleccionamos columnas de interés
esp2 <- subset(esp2, select = c("Comunidades.y.Ciudades.Autonómas.Id", "Anyo", "Valor"))
```

7.2.2.1 Contornos

Una vez obtenidos los datos que se quieren incluir en el mapa, es necesario **disponer de los contornos geográficos** sobre los cuales se quieren representar dichos datos. En este caso, los contornos de las comunidades autónomas.

El ine dispone acceso de contornos para los límites geográficos en la siguiente web:



Web <https://www.ine.es/wstempus/geojs/ES/CONTORNOS/XXXX> sustituyendo XXXX por el código del ámbito geográfico necesario, siendo estos:

- 349: total nacional
- 70: ccaa
- 115: provincias
- 19: municipios

En nuestro caso necesitamos el de las comunidades autónomas.

```
# Contornos de las ccaa
ccaa2 <- read_sf("https://www.ine.es/wstempus/geojs/ES/CONTORNOS/70")
```

Para colorear el mapa, se desea que se muestren colores para intervalos de Esperanza de Vida (tomando los cuartiles), para ello será necesario usar la función `colorBin()`. Por tanto, tenemos:

```
# join de los contornos y el dataset
ccaa2 <- merge(ccaa2, esp2, by.x = "id_region",
               by.y = "Comunidades.y.Ciudades.Autonómas.Id" )

# Creamos colores para cada rango de esperanza de vida
pal2 <- colorNumeric("plasma", domain= NULL)

# Labels of the map
labels <- sprintf(
  "<strong>%s</strong><br/> Sueldo bruto mensual %s € ",
  ccaa2$nom_region, format(ccaa2$Val, big.mark = ".", decimal.mark = ",", nsmall = 2)
) %>% lapply(htmltools::HTML)

# Create the map
m2 <- leaflet(ccaa2) %>%
  addProviderTiles(providers$Esri.WorldImagery) %>%
  setView(-4, 40, zoom = 5) %>%
  addPolygons(fillOpacity = 0.8,
              fillColor = ~pal2(Valor),
              weight = 1,
              color = "white",
              label = labels,
              labelOptions = labelOptions(
                style = list("font-weight" = "normal", padding = "3px 8px"),
                textsize = "15px",
                direction = "auto"
              ),
              highlightOptions = highlightOptions(fillOpacity = 1, bringToFront = TRUE,
                                                    weight = 2, color = "white")) %>%
  addLegend(pal = pal2, values = ~Valor, opacity = 1.0, position = "bottomleft",
            labFormat = labelFormat(suffix = "€", digits = 2, big.mark = "."),
```

```
title = sprintf("Sueldo bruto mensual (%s)",
                esp$Anyo[1]))%>%
config(locale = 'es')
```

Anexos

Appendix A

Ineapir

A.1 Introducción

En el ámbito de la investigación y el análisis de datos socioeconómicos, el acceso a información precisa y actualizada es fundamental. **ineapir** es una herramienta diseñada para facilitar el acceso y la manipulación de datos proporcionados por el INE. Este paquete de R permite a los usuarios interactuar de manera eficiente con la API del INE, simplificando la tarea de obtener datos detallados sobre diversos aspectos de la economía, demografía y sociedad española.

En los siguientes apartados se abordará:

- **Introducción a la API del INE:** Una visión general de la API del INE, sus funcionalidades y cómo **ineapir** facilita su uso.
- **Instalación y Configuración de ineapir:** Pasos detallados para la instalación del paquete y la configuración inicial para comenzar a utilizarlo.
- **Obtención de Datos:** Instrucciones sobre cómo realizar consultas específicas a la API del INE y obtener datos en formatos utilizables.
- **Manipulación y Visualización de Datos:** Técnicas para limpiar, manipular y visualizar los datos obtenidos, incluyendo ejemplos prácticos.
- **Análisis Avanzados:** Métodos y herramientas avanzadas para el análisis de datos socioeconómicos utilizando **ineapir**.

A.1.1 Beneficios de Usar ineapir

ineapir no solo facilita el acceso a los datos del INE, sino que también mejora la eficiencia y precisión en el análisis de estos datos. Entre sus principales ventajas se encuentran:

- **Acceso Simplificado:** Permite realizar consultas complejas de manera sencilla y rápida.
- **Integración con R:** Aprovecha las capacidades de R para análisis estadísticos y visualización de datos.
- **Actualización Continua:** Facilita el acceso a los datos más recientes publicados por el INE.
- **Versatilidad:** Puede ser utilizado en diversos campos de estudio, desde la economía y la sociología hasta la demografía y las ciencias políticas.

A.1.2 Público Objetivo

Esta herramienta está dirigida a:

- Investigadores y académicos que trabajan con datos socioeconómicos.

- Profesionales del análisis de datos y la estadística.
- Estudiantes de ciencias sociales y económicas.
- Cualquier persona interesada en el análisis de datos proporcionados por el INE.

Bibliography

- Crespo, D. (2024). *Ineapir: Obtaining data published by the national statistics institute* [R package version 0.0.0.9000, <https://es-ine.github.io/ineapir/>]. <https://github.com/es-ine/ineapir>
- Xie, Y. (2017). *Authoring books and technical documents with R markdown* (1st ed.). Chapman & Hall. <https://bookdown.org/yihui/bookdown/>