

EDA del volumen de alimentos, clasificados en categorías generales, consumidos por los hogares españoles

David Moreno

2023-09-14

1. Objetivos

El presente análisis exploratorio de datos (EDA) se enmarca en un proyecto de mayor alcance destinado a investigar detalladamente el patrón de consumo de alimentos en los hogares españoles.

Los datos considerados en este estudio fueron recopilados a partir de la fuente oficial del Ministerio de Agricultura, Pesca y Alimentación. Esta información se encontraba distribuida en 22 archivos de Excel, cada uno de los cuales contenía hasta 14 hojas de datos. En particular, me centro en la hoja denominada “VOLUMEN”, que almacena información sobre la cantidad de alimentos consumidos por los hogares en diversas regiones de alimentos.

Por lo tanto, el propósito principal de este análisis es adentrarnos en el patrón de consumo de alimentos por parte de los hogares españoles, centrándonos en la cantidad consumida de las principales categorías de alimentos. Esto nos permitirá dar respuesta a las siguientes cuestiones fundamentales:

- ¿Cómo ha evolucionado el volumen de consumo de alimentos en España a lo largo de los años?
- ¿En qué medida varía el consumo de alimentos entre las distintas regiones de España?
- ¿Cuáles son las categorías de alimentos más populares y de qué manera han experimentado cambios a lo largo de las décadas estudiadas?

Adicionalmente, la información obtenida en este análisis nos proporcionará una base sólida para plantear nuevas preguntas que enriquecerán futuros análisis y contribuirán a una comprensión más completa de los hábitos alimentarios en los hogares españoles.

2. Preparación del entorno de trabajo.

Configuro el repositorio de CRAN de manera no interactiva

```
options(repos = c(CRAN = "https://cran.r-project.org"))
```

Instalo las librerías que voy a utilizar

```
install.packages("readxl")
```

```
## Installing package into 'C:/Users/gilga/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'readxl' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'readxl'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problema al copiar
## C:\Users\gilga\AppData\Local\R\win-library\4.3\00LOCK\readxl\libs\x64\readxl.dll
## a C:\Users\gilga\AppData\Local\R\win-library\4.3\readxl\libs\x64\readxl.dll:
## Permission denied
```

```

## Warning: restored 'readxl'

##
## The downloaded binary packages are in
## C:\Users\gilga\AppData\Local\Temp\RtmpY9qQNj\downloaded_packages
install.packages("httr")

## Installing package into 'C:/Users/gilga/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
## package 'httr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\gilga\AppData\Local\Temp\RtmpY9qQNj\downloaded_packages
install.packages("tidyverse")

## Installing package into 'C:/Users/gilga/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\gilga\AppData\Local\Temp\RtmpY9qQNj\downloaded_packages
install.packages("rmarkdown")

## Installing package into 'C:/Users/gilga/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
## package 'rmarkdown' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\gilga\AppData\Local\Temp\RtmpY9qQNj\downloaded_packages
install.packages("dplyr")

## Installing package into 'C:/Users/gilga/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
## package 'dplyr' successfully unpacked and MD5 sums checked
## Warning: cannot remove prior installation of package 'dplyr'
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problema al copiar
## C:\Users\gilga\AppData\Local\R\win-library\4.3\00LOCK\dplyr\libs\x64\dplyr.dll
## a C:\Users\gilga\AppData\Local\R\win-library\4.3\dplyr\libs\x64\dplyr.dll:
## Permission denied
## Warning: restored 'dplyr'

##
## The downloaded binary packages are in
## C:\Users\gilga\AppData\Local\Temp\RtmpY9qQNj\downloaded_packages
install.packages("ggplot2")

## Installing package into 'C:/Users/gilga/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
## package 'ggplot2' successfully unpacked and MD5 sums checked
##

```

```
## The downloaded binary packages are in
## C:\Users\gilga\AppData\Local\Temp\RtmpY9qQNj\downloaded_packages

install.packages("corrplot")

## Installing package into 'C:/Users/gilga/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'corrplot' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\gilga\AppData\Local\Temp\RtmpY9qQNj\downloaded_packages
```

Y, por último, se cargan las librerías.

```
library(readxl)
library(httr)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.3      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2    3.4.3      v tibble     3.2.1
## v lubridate  1.9.2      v tidyr      1.3.0
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(rmarkdown)
library(dplyr)
library(ggplot2)
library(corrplot)
```

```
## corrplot 0.92 loaded
```

3. Obtención de los datos

Hago una copia de df_volumen llamada data, con la que trabajaré a continuación

```
# Datos de la ruta
ruta <- "https://github.com/DMorgon/portafolios/raw/main"
rama <- "alimentacion/datos_procesados"
nombre_archivo <- "volumen_categorias.csv"

# Creo la ruta desde donde se descargaran los archivos
archivo_url <- paste0(ruta, "/", rama, "/", nombre_archivo)

# Creo el nombre del archivo local
archivo_local <- "volumen_categoria.csv"

# Descargar el archivo desde la URL
response <- GET(archivo_url, write_disk(archivo_local, overwrite = TRUE))

# Creo el df con los datos del archivo local
volumen_categorias <- read.csv(archivo_local)
```

```
# Borro el archivo local
file.remove(archivo_local)
```

```
## [1] TRUE
```

```
# Elimino las variables que ya no se utilizará.
```

```
rm(ruta, rama, nombre_archivo, archivo_url, archivo_local, response)
```

Hago una copia de df_volumen llamada data, con la que trabajaré a continuación

```
data <- volumen_categorias
```

4. Analisis descriptivo del conjunto de los datos

Obtengo una vista preliminar del marco de datos

```
View(data)
```

Examino la estructura interna del marco de datos:

```
str(data)
```

```
## 'data.frame': 14059 obs. of 4 variables:
## $ AÑO : chr "2000-01-01" "2000-01-01" "2000-01-01" "2000-01-01" ...
## $ CATEGORIAS: chr "Huevos" "Miel" "Carne" "Pesca" ...
## $ REGIONES : chr "Cataluña" "Cataluña" "Cataluña" "Cataluña" ...
## $ VOLUMEN : num 62709 2840 327730 153650 514570 ...
```

Creo un resumen estadístico de las variables del conjunto de datos

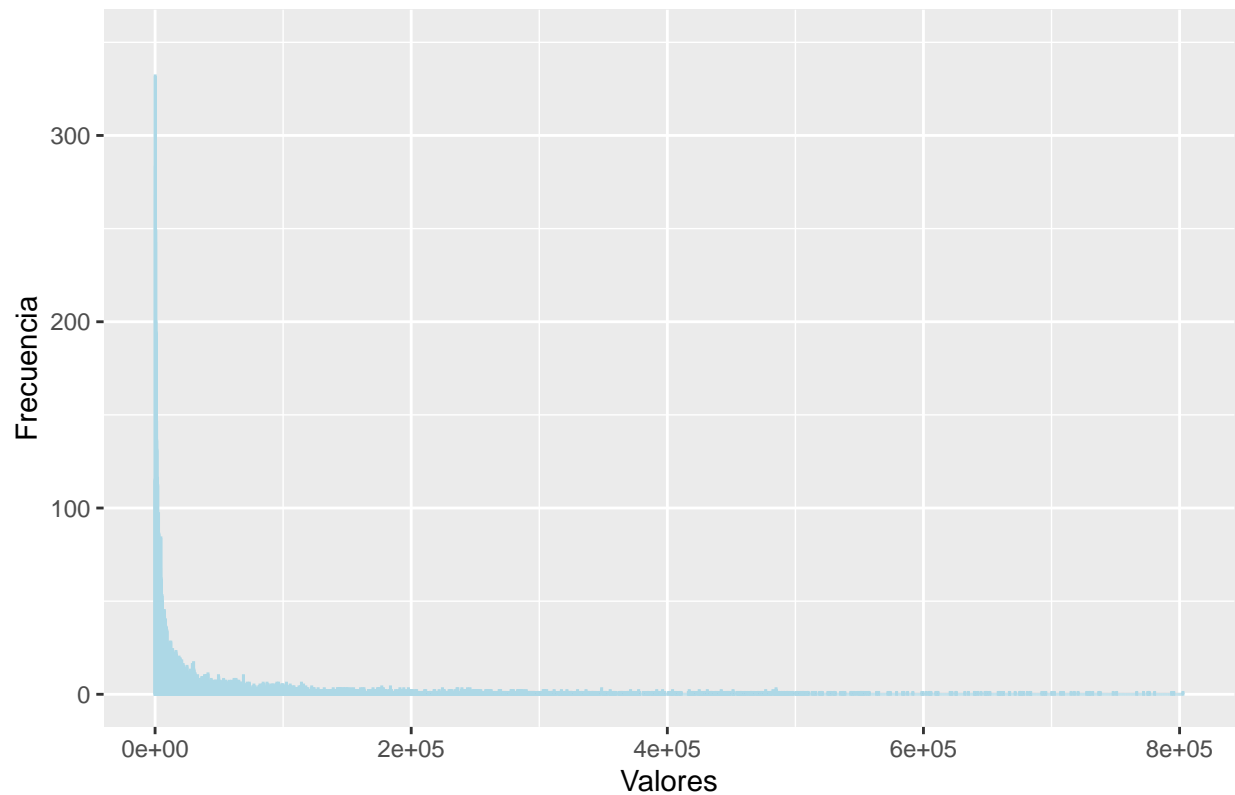
```
summary(data)
```

```
##      AÑO      CATEGORIAS      REGIONES      VOLUMEN
## Length:14059 Length:14059 Length:14059 Min. : 0
## Class :character Class :character Class :character 1st Qu.: 1950
## Mode :character Mode :character Mode :character Median : 8157
##                                     Mean : 44046
##                                     3rd Qu.: 39298
##                                     Max. :802585
```

Creo un histograma para examinar la distribución de los datos de la variable VOLUMEN

```
# Creo el histograma de VOLUMEN
ggplot(
  data = data,
  aes(
    x = VOLUMEN)) +
  geom_histogram(
    binwidth = 100,
    fill = "blue",
    color = "lightblue") +
  labs(x = "Valores",
       y = "Frecuencia",
       title = "Histograma de VOLUMEN") +
  ylim(0, 350)
```

Histograma de VOLUMEN

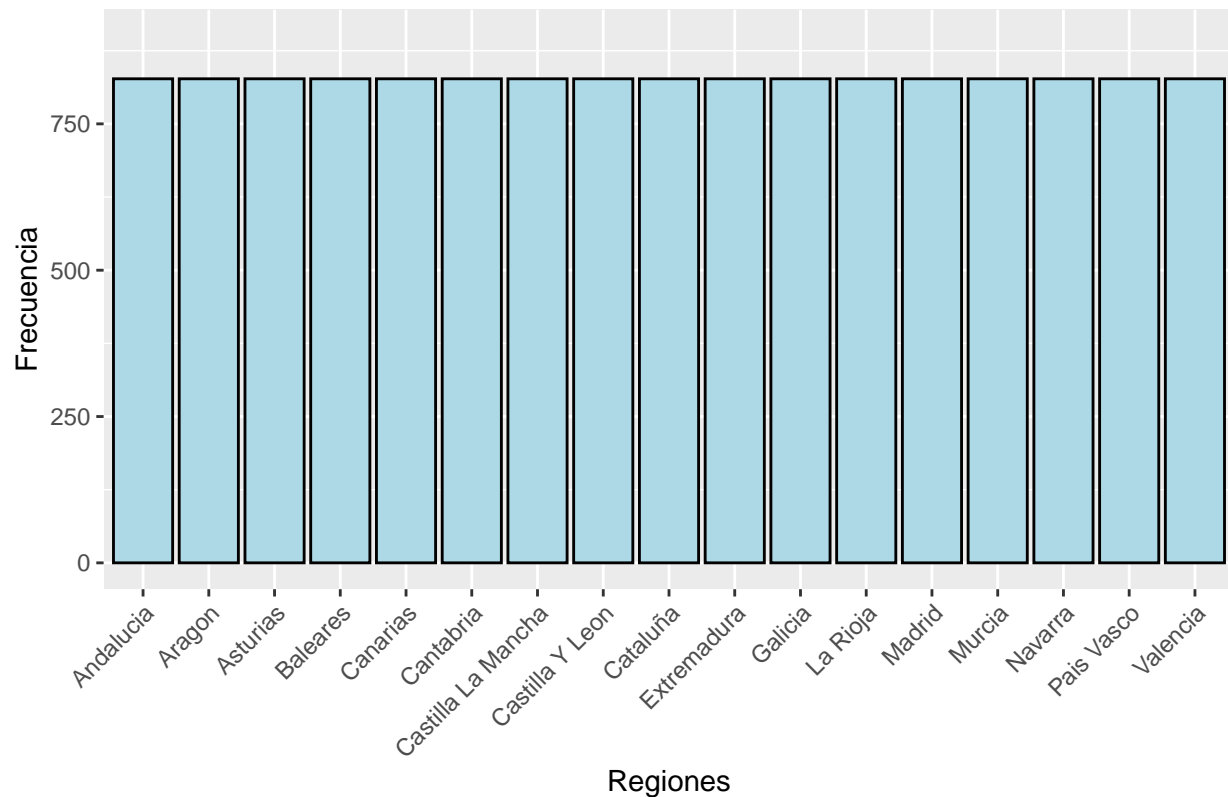


Creo un histograma para examinar la distribución de los datos de la variable REGIONES

Crear un gráfico de barras de la frecuencia de REGIONES

```
ggplot(
  data = data,
  aes(
    x = REGIONES)) +
  geom_bar(
    fill = "lightblue",
    color = "black",
    width = 0.9) +
  labs(
    x = "Regiones",
    y = "Frecuencia",
    title = "Gráfico de Barras de la frecuencia de cada región") +
  coord_cartesian(
    ylim = c(0, 900)) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1))
```

Gráfico de Barras de la frecuencia de cada región

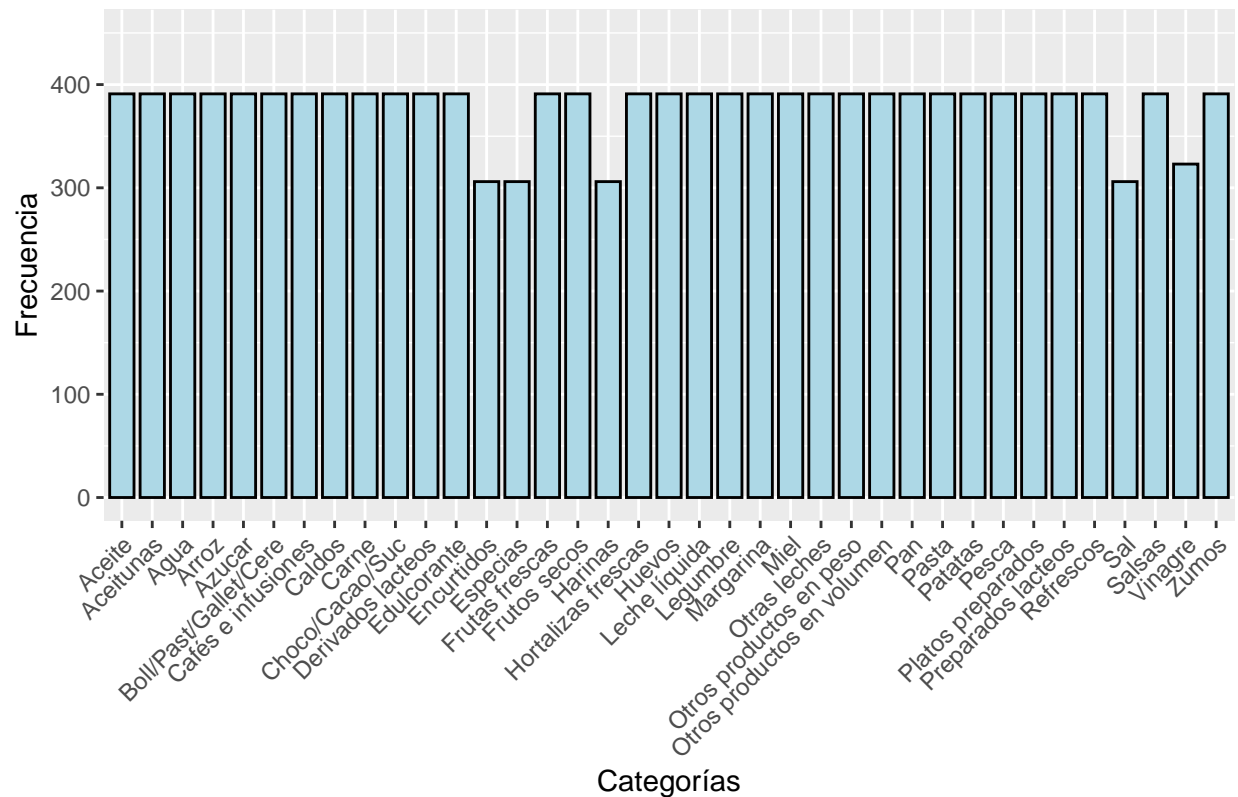


Creo un histograma para examinar la distribución de los datos de la variable CATEGORÍAS

Crear un gráfico de barras de la frecuencia de CATEGORIAS

```
ggplot(
  data = data,
  aes(
    x = CATEGORIAS)) +
  geom_bar(
    fill = "lightblue",
    color = "black",
    width = 0.8) +
  labs(
    x = "Categorías",
    y = "Frecuencia",
    title = "Gráfico de Barras de la frecuencia de cada categoria de alimentos") +
  coord_cartesian(
    ylim = c(0, 450)) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1))
```

Gráfico de Barras de la frecuencia de cada categoria de alimentos

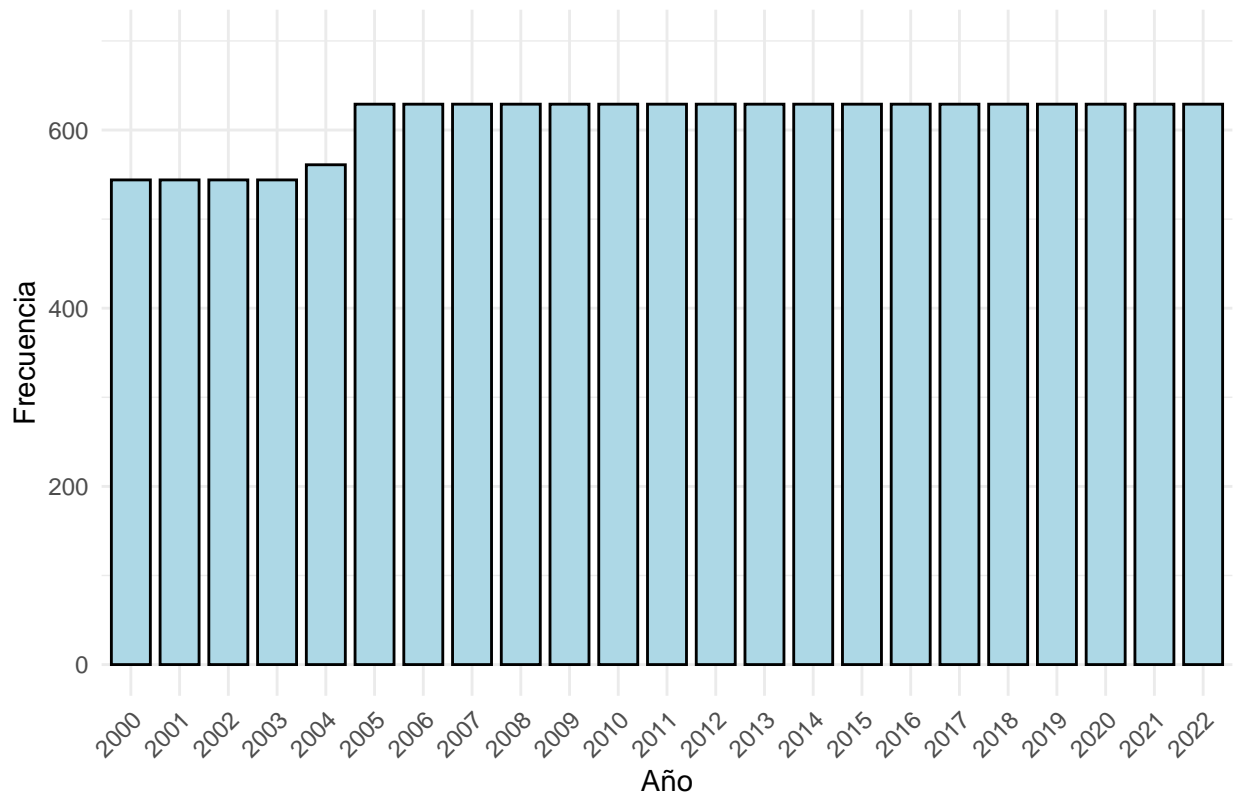


Creo un histograma para examinar la distribución de los datos de la variable AÑO

Crear un gráfico de barras de la frecuencia de cada año

```
ggplot(
  data = data,
  aes(
    x = factor(substr(AÑO, 1, 4))) +
  geom_bar(
    fill = "lightblue",
    color = "black",
    width = 0.8) +
  labs(
    x = "Año",
    y = "Frecuencia",
    title = "Gráfico de Barras de la frecuencia de cada año") +
  coord_cartesian(ylim = c(0, 700)) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Gráfico de Barras de la frecuencia de cada año



Tras examinar el marco de datos he llegado a la siguiente conclusión:

- El número total de observaciones del conjunto de datos es de 14,059 observaciones y 4 variables, que son:
 - **AÑO**: representa fechas en formato de texto (chr). La serie va desde el año 2000 hasta el 2022. Los años 2000, 2001, 2002, 2003 y 2004 tienen una frecuencia menor que el resto de año. Puede deberse a que se produjo un cambio de metodología y a partir del 2005 se introdujo nuevas categorías, no contabilizadas anteriormente.
 - **CATEGORÍAS**: representar las categorías a las que pertenecen los datos. Está en formato de texto (chr). El número de valores únicos son 37. Destacar que las categorías encurtidos, especias, harinas, sal y vinagre tienen menos frecuencia absolutas que el resto de las categorías. Esto, unido a que, gualmente, existen años con menos frecuencia, es una muestra de que estas categorías se introdujeron con posterioridad. Así, el vinagre pudo introducirse en el año 2004 y el resto en a partir del 2005.
 - **REGIONES**: representa las regiones a las que pertenecen los datos. También está en formato de texto (chr) y existen 17 regiones diferentes. Todas las regiones tienen la misma frecuencia, es decir, se repiten el mismo número de veces.
 - **VOLUMEN**: es una variable numérica (num) y representa una cantidad de los alimentos consumidos por los hogares españoles. Esta variable tiene un amplio rango de valores, que va desde el 0 hasta 802585. Además, un sesgo hacia la izquierda en su distribución, es decir, la mayoría de los valores se sitúan a la izquierda.

5. Ajuste de variables

Transformo los datos de la variable VOLUMEN a tipo numérico.


```
data$VOLUMEN <- as.numeric(data$VOLUMEN)
```

Y Ahora cambio AÑO a tipo de dato de fecha

```
# Suponiendo que "data" es tu marco de datos  
data$AÑO <- as.Date(data$AÑO, format = "%Y-%m-%d")
```

A continuación, tanto las variables REGIONES como CATEGORIAS las transformo en factores

```
data$REGIONES <- factor(data$REGIONES)  
data$CATEGORIAS <- factor(data$CATEGORIAS)
```

6. Detección y tratamiento de valores ausentes

Por lo que parece a continuación, no existen valores ausentes.

```
any(is.na(data))
```

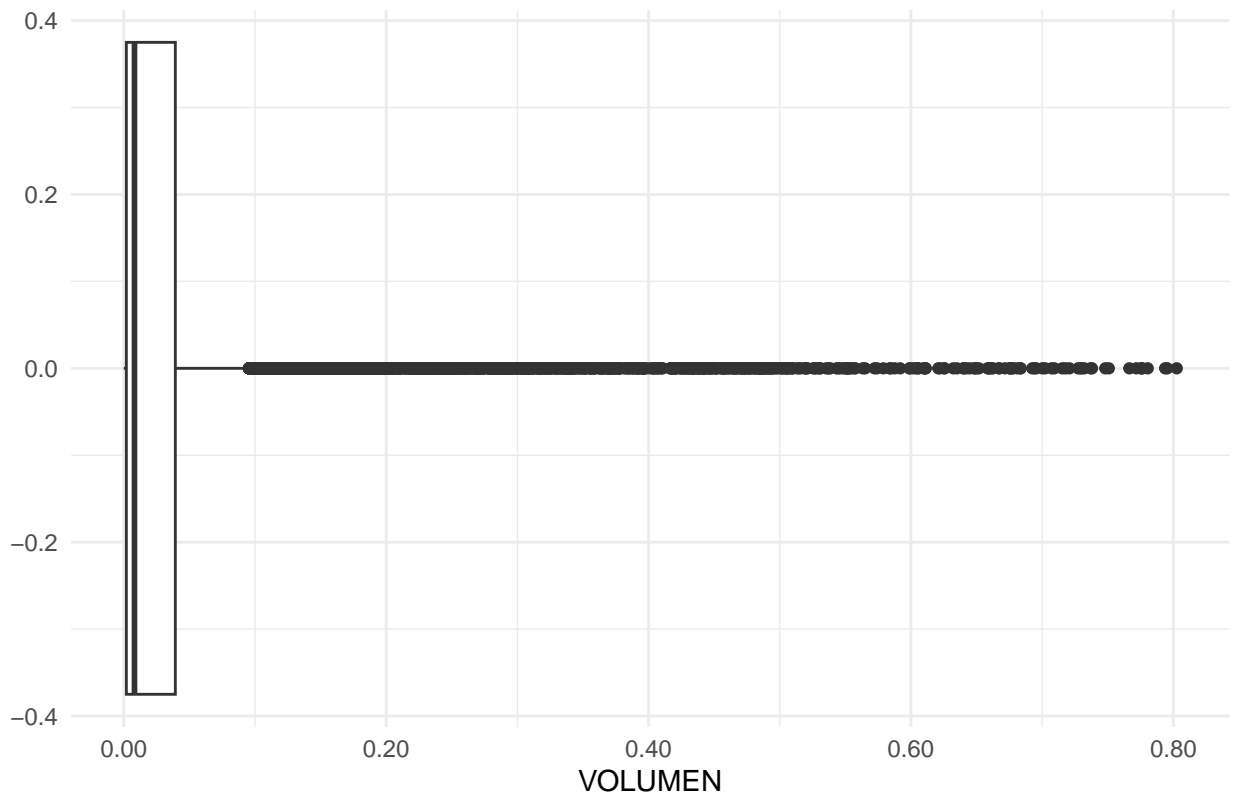
```
## [1] FALSE
```

7. Identificación de valores atípicos

A continuación realizo un gráfico de caja y bigotes para visualizar si existen datos atípicos.

```
# Calculo las estadísticas necesarias para reproducir el gráfico  
estadisticas <- boxplot.stats(data$VOLUMEN)  
  
# Construyo el gráfico de cajas y bigotes  
ggplot(  
  data = data,  
  aes(  
    x = VOLUMEN)) +  
  geom_boxplot() +  
  
  scale_x_continuous(  
    labels = scales::number_format(scale = 1e-6, accuracy = 0.01)) +  
  labs(  
    x = "VOLUMEN",  
    title = "Gráfico de caja y bigotes de VOLUMEN") +  
  theme_minimal()
```

Gráfico de caja y bigotes de VOLUMEN



```
rm(estadisticas)
```

Como se ve en gráfico, el rango de valores de la variable VOLUMEN es muy amplio, ya que va desde el 0 hasta más de los 800000. Sin embargo, como se puede apreciar, la mayoría de los valores están situados en la izquierda, es decir, su distribución presenta un sesgo hacia la izquierda.

Como era de esperar, la variable también presenta valores atípicos. Sin embargo, es importante destacar que estos valores atípicos no necesariamente indican errores en la recopilación de datos, sino que pueden reflejar características genuinas de los datos.

A continuación, se exponen algunas razones por las cuales se pueden encontrar valores atípicos en esta variable:

- **Diferencias en las Categorías de Alimentos:** Es completamente normal que las distintas categorías de alimentos se consuman en cantidades variadas. Esto puede dar lugar a diferencias significativas en la distribución de “VOLUMEN” entre las categorías. Al realizar análisis o visualizaciones específicas para cada categoría, es posible identificar valores atípicos en aquellas categorías que se consumen en cantidades mucho mayores o menores en comparación con otras.
- **Diferencias en las Regiones:** Las variaciones en el tamaño de la población y las características demográficas de las diferentes regiones pueden influir en los patrones de consumo de alimentos. Regiones con poblaciones más grandes pueden tener un consumo total de alimentos más elevado, lo que puede afectar la distribución de “VOLUMEN” para esas regiones. Al analizar los valores atípicos en función de las regiones, es posible comprender mejor las diferencias regionales.
- **Cambios en la Metodología y la Disponibilidad de Datos:** Cambios en la metodología de recopilación de datos o en la disponibilidad de datos a lo largo del tiempo pueden desempeñar un papel importante en la presencia de valores atípicos. Por ejemplo, si ciertas categorías de alimentos comenzaron a registrarse a partir del año 2005, es natural que existan menos datos disponibles para los

años anteriores. Esto puede resultar en valores atípicos en esos años anteriores, ya que los datos pueden ser menos representativos o menos precisos debido a la falta de información histórica.

En resumen, los valores atípicos en la variable “VOLUMEN” son el resultado de las diferencias en las categorías de alimentos, las variaciones regionales y los posibles cambios en la metodología y la disponibilidad de datos a lo largo del tiempo. Estos valores no necesariamente indican errores, sino que proporcionan información valiosa sobre la realidad subyacente de los datos.

8. Análisis de relaciones entre variables.

De las cuatro variables que tiene el conjunto de datos, selecciona la variable VOLUMEN como la variable respuesta y a continuación las comparo con el resto de variables

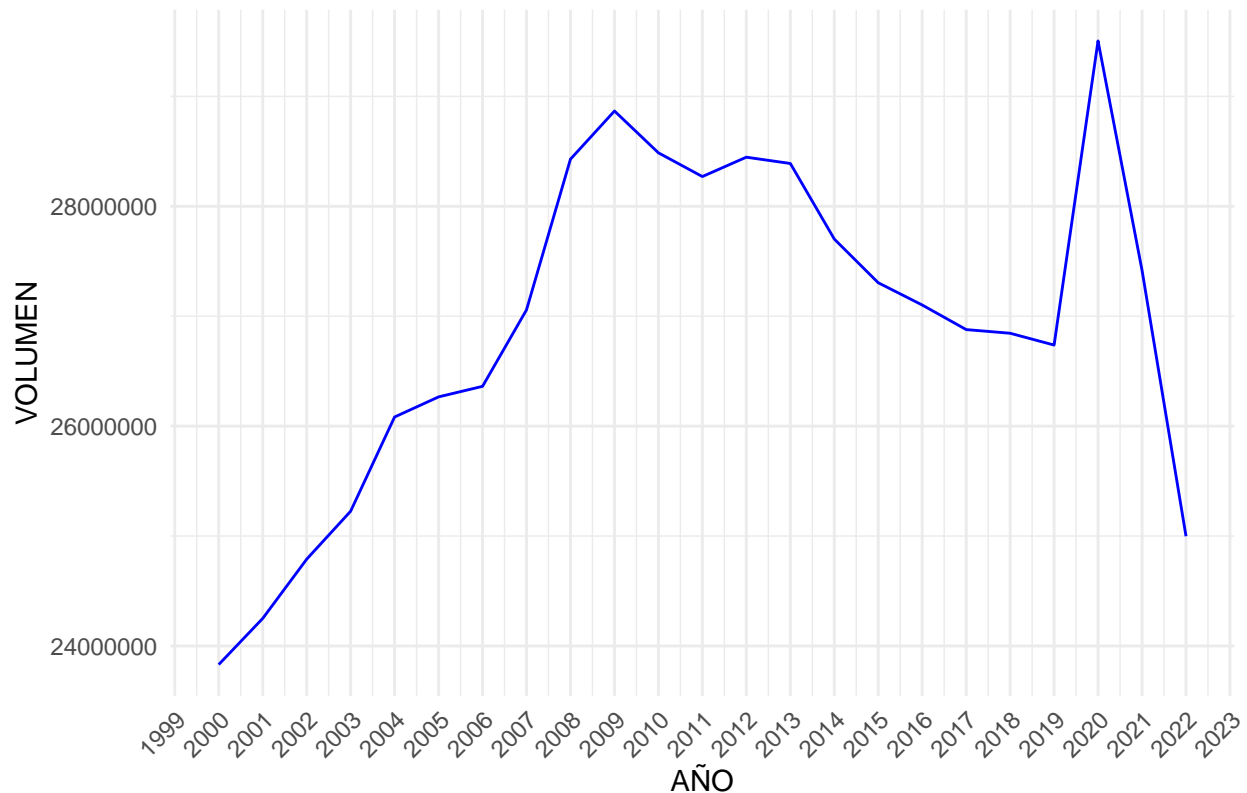
7.1. VOLUMEN VS AÑO.

```
# Establecer el formato de etiqueta del eje x a decimal
options(scipen = 999)

data_año <- data %>%
  group_by(AÑO) %>%
  summarise(
    Volumen = sum(VOLUMEN))

# Crear un gráfico de líneas con todos los años en el eje de las X
ggplot(
  data_año,
  aes(
    x = AÑO,
    y = Volumen)) +
  geom_line(color = "blue") +
  labs(
    x = "AÑO",
    y = "VOLUMEN",
    title = "Gráfico de Líneas VOLUMEN vs. AÑO") +
  theme_minimal() +
  scale_x_date(
    date_labels = "%Y",
    date_breaks = "1 year") +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1))
```

Gráfico de Líneas VOLUMEN vs. AÑO



```
rm(data_año)
```

En este gráfico, se observa la evolución de la cantidad de alimentos consumidos en hogares españoles desde el año 2000 hasta el 2022.

- En el año 2000, el consumo comenzó en niveles mínimos, situándose por debajo de los 24 mil millones de kilos.
- Hasta 2009, la cantidad de alimentos consumidos experimentó un aumento constante, acercándose a los 29 mil millones de kilos.
- A partir del año 2010, se inició un declive que continuó hasta 2019, momento en el cual se situó por debajo de los 27 mil millones de kilos.
- En el año 2020, se produjo un aumento significativo, superando los 29 mil millones de kilos.
- Sin embargo, en los años 2021 y 2022, el consumo disminuyó, retornando a niveles aproximados a los del año 2003.

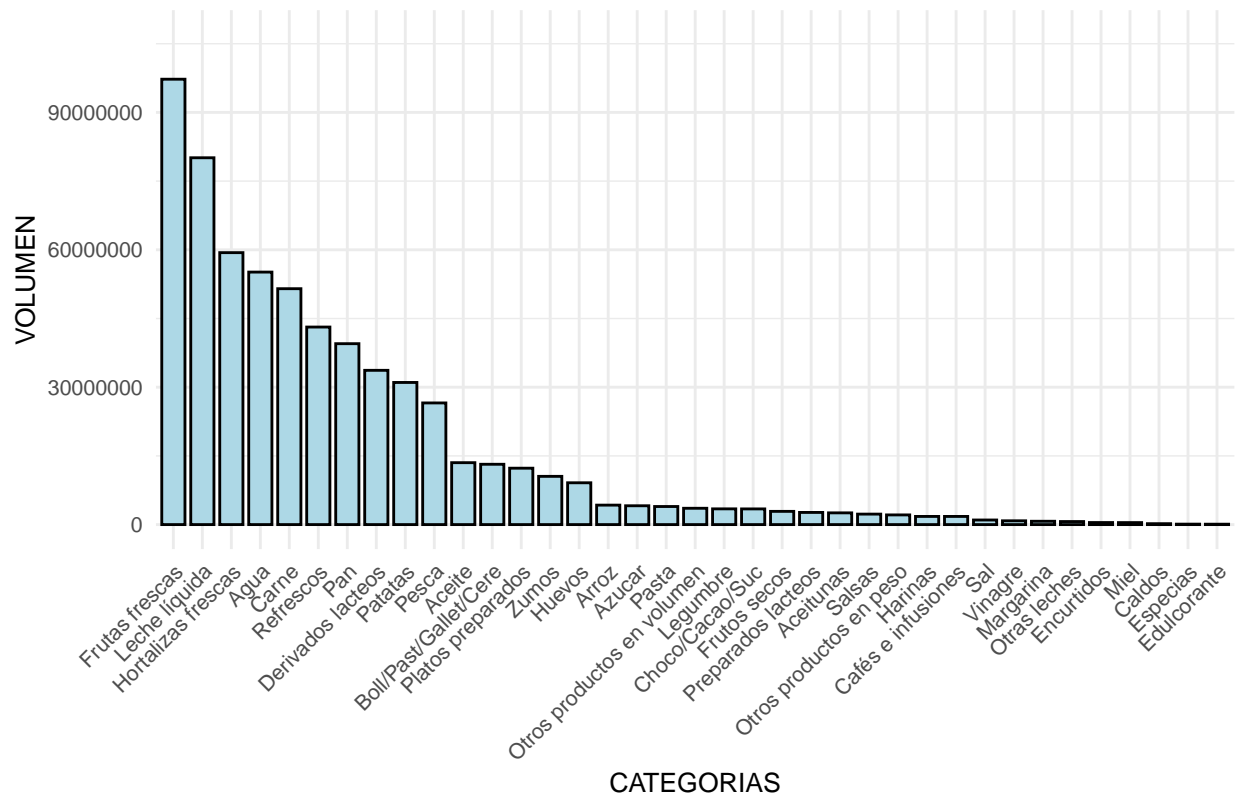
En resumen, la evolución de la cantidad de alimentos consumidos muestra cambios significativos a lo largo de estos años, con un periodo de descenso a partir de 2021. Esta información es crucial para comprender las tendencias en los hábitos de consumo alimentario en España.

7.3. VOLUMEN VS CATEGORIAS

```
data_categorias <- data %>%  
  group_by(CATEGORIAS) %>%  
  summarise(Volumen = sum(VOLUMEN)) %>%  
  arrange(desc(Volumen))
```

```
# Crea un gráfico de líneas
ggplot(
  data = data_categorias,
  aes(
    x = reorder(CATEGORIAS, -Volumen),
    y = Volumen)) +
  geom_bar(
    stat = "identity",
    position = "dodge",
    fill = "lightblue",
    color = "black",
    width = 0.8) +
  labs(
    x = "CATEGORIAS",
    y = "VOLUMEN",
    title = "Gráfico de Barras VOLUMEN vs. CATEGORIAS") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.text.y = element_text(size = 8),
        axis.title = element_text(size = 10),
        legend.position = "none") +
  ylim(0, max(data_categorias$Volumen * 1.1))
```

Gráfico de Barras VOLUMEN vs. CATEGORIAS



```
rm(data_categorias)
```

Tu descripción es clara y proporciona una visión general de la distribución del consumo de diferentes categorías

de alimentos según el gráfico de barras. Sin embargo, puedes mejorarla haciendo que la información sea más cohesiva y organizada. Aquí tienes una versión revisada:

En este gráfico, se destaca la variabilidad en el consumo de diversas categorías de alimentos:

- En el grupo de las categorías más consumidas se encuentran la fruta fresca y la leche líquida, destacando por su alta demanda.
- Le siguen de cerca las hortalizas frescas, el agua mineral embotellada y la carne, que también tienen un consumo significativo.
- En una categoría intermedia se encuentran los refrescos y el pan, con niveles de consumo notables.
- Posteriormente, se ubican las categorías de derivados lácteos, patatas, pescado y marisco, que se consumen con frecuencia.
- En un quinto grupo se incluyen el aceite, productos de bollería, galletas y cereales, platos preparados, zumos y huevos, con un nivel de consumo menor en comparación con las categorías mencionadas anteriormente.
- Finalmente, las demás categorías muestran niveles de consumo más bajos en general.

Esta representación visual ofrece una perspectiva clara de las preferencias de consumo de alimentos en cada categoría y es valiosa para comprender las tendencias alimentarias en la población.

7.4. VOLUMEN VS AÑO VS CATEGORIAS

```
# Agrupa los datos por categorías y años, calculando la suma del VOLUMEN para cada combinación
data_categorias <- data %>%
  group_by(CATEGORIAS, AÑO) %>%
  summarise(
    Volumen = sum(VOLUMEN))
```

```
## `summarise()` has grouped output by 'CATEGORIAS'. You can override using the
## `.groups` argument.
```

```
# Obtiene las categorías únicas presentes en los datos
categorias_unicas <- unique(data_categorias$CATEGORIAS)

# Crea un gráfico de líneas para cada categoría de alimentos
for (i in categorias_unicas) {

  # Filtra los datos para la categoría actual
  datos <- data_categorias %>%
    filter(CATEGORIAS == i)

  # Crea un gráfico de líneas para la categoría actual
  grafico <- ggplot(
    data = datos,
    aes(
      x = AÑO,
      y = Volumen)) +
    geom_line(
      color = "blue") +
    labs(
      x = "AÑO",
      y = "VOLUMEN",
      title = paste0("Gráfico de Líneas con el consumo de ", i)) +
    theme_minimal() +
```

```

scale_x_date(
  date_labels = "%Y",
  date_breaks = "1 year" ) +
theme(
  axis.text.x = element_text(angle = 45, hjust = 1))

# Imprime el gráfico actual
print(grafico)
}

```

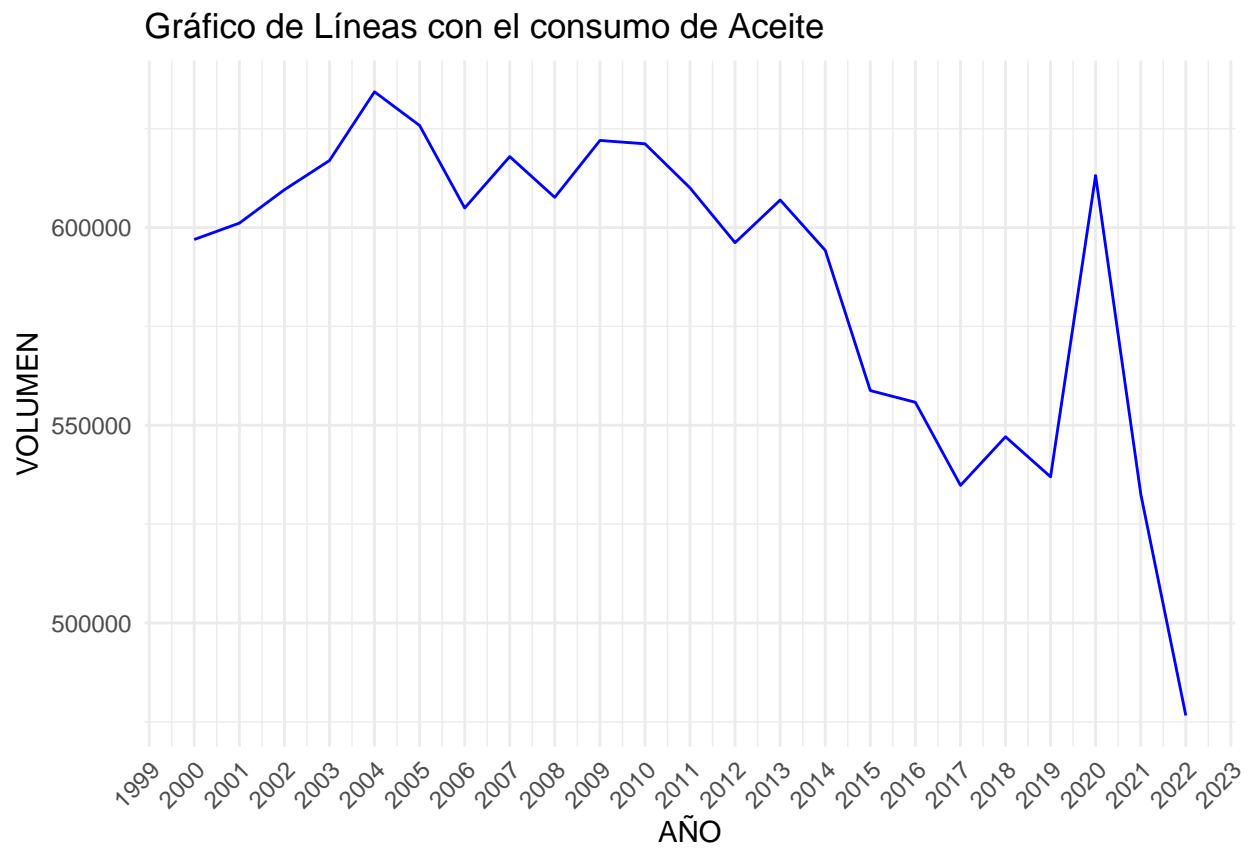


Gráfico de Líneas con el consumo de Aceitunas

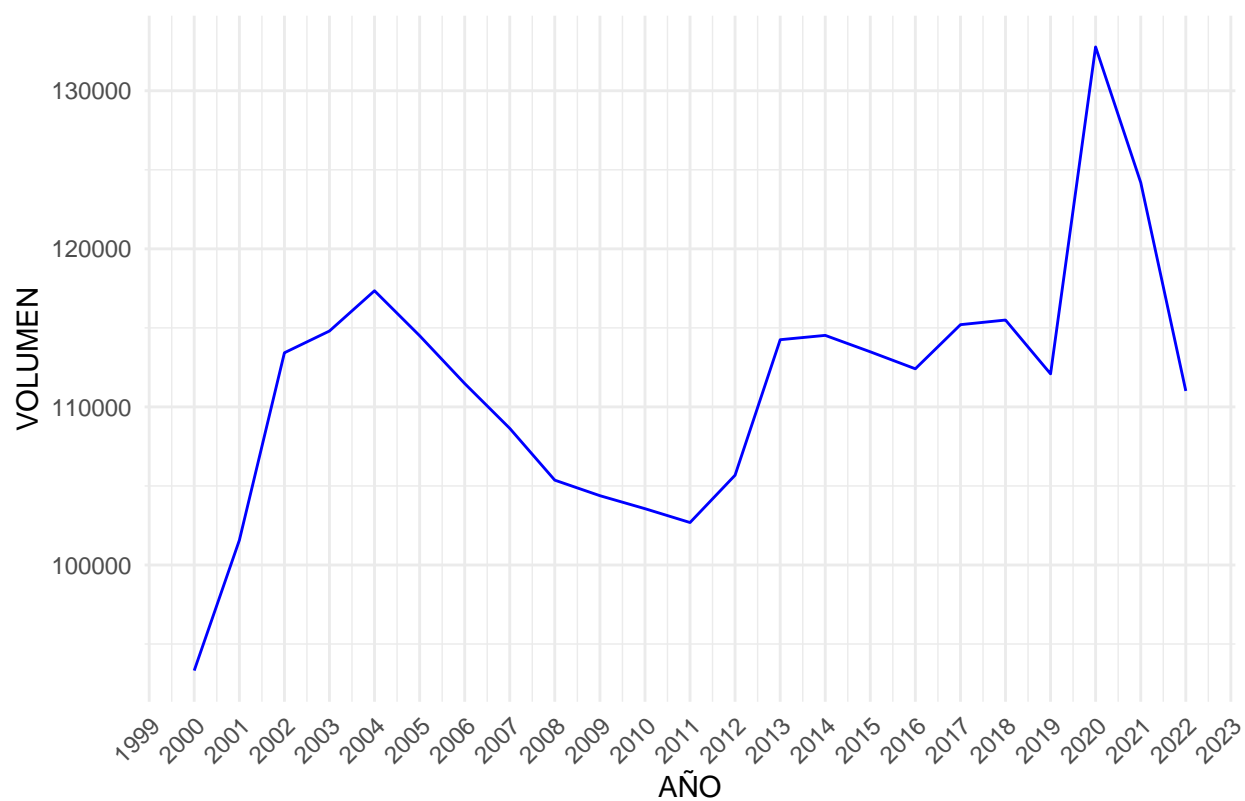
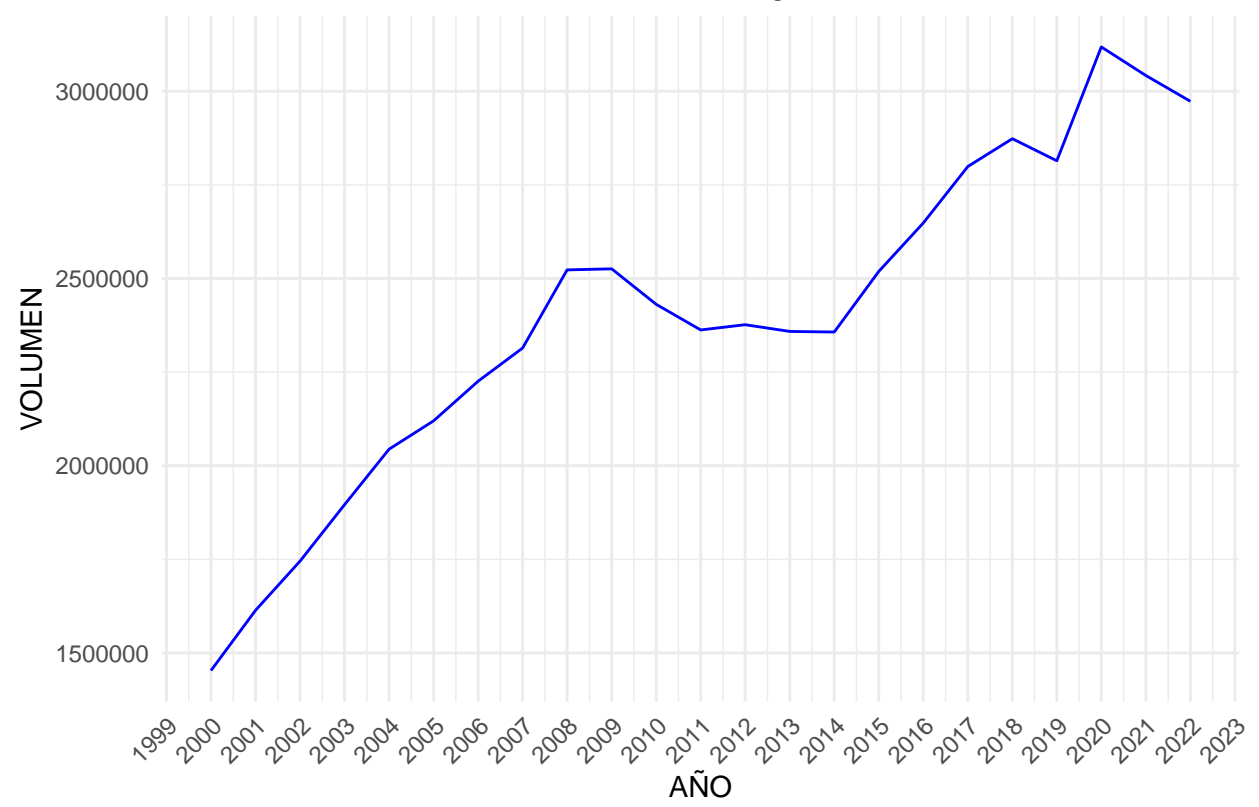
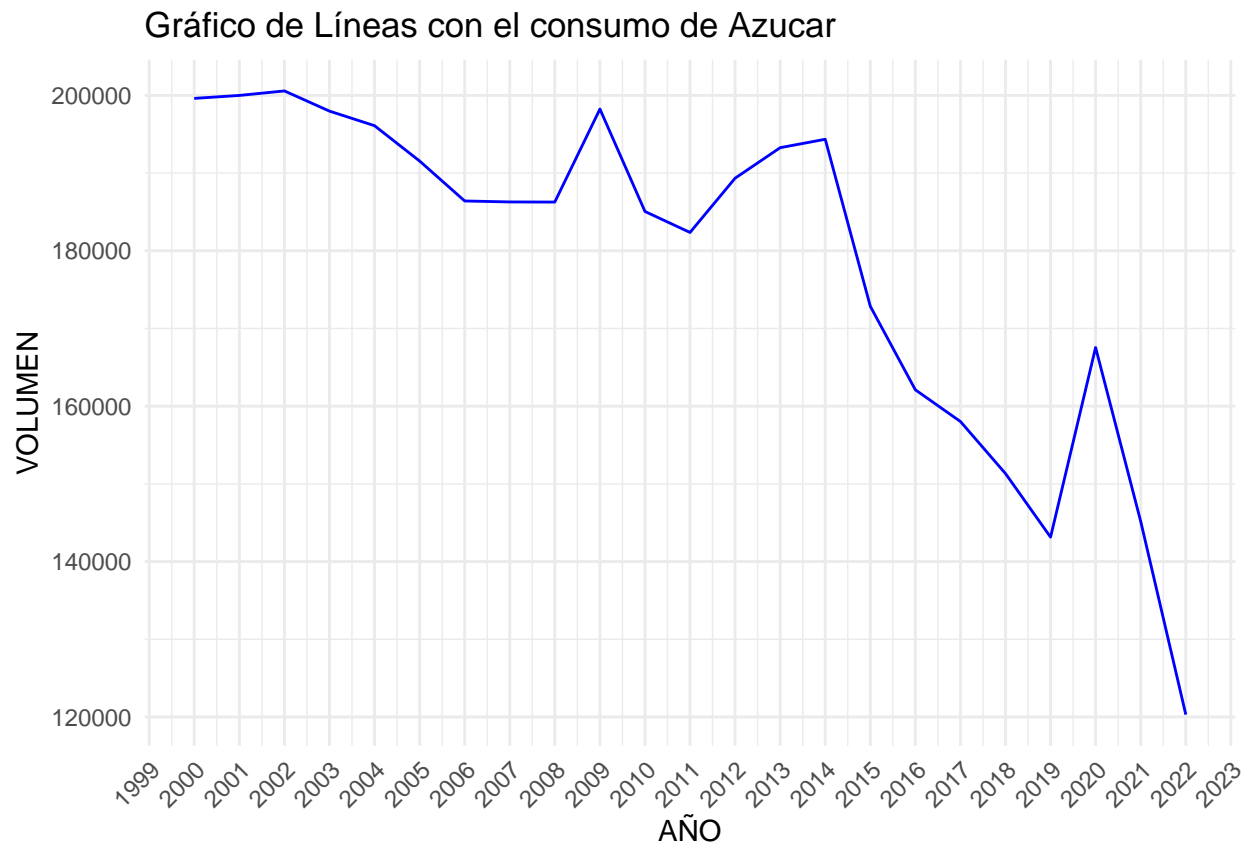


Gráfico de Líneas con el consumo de Agua







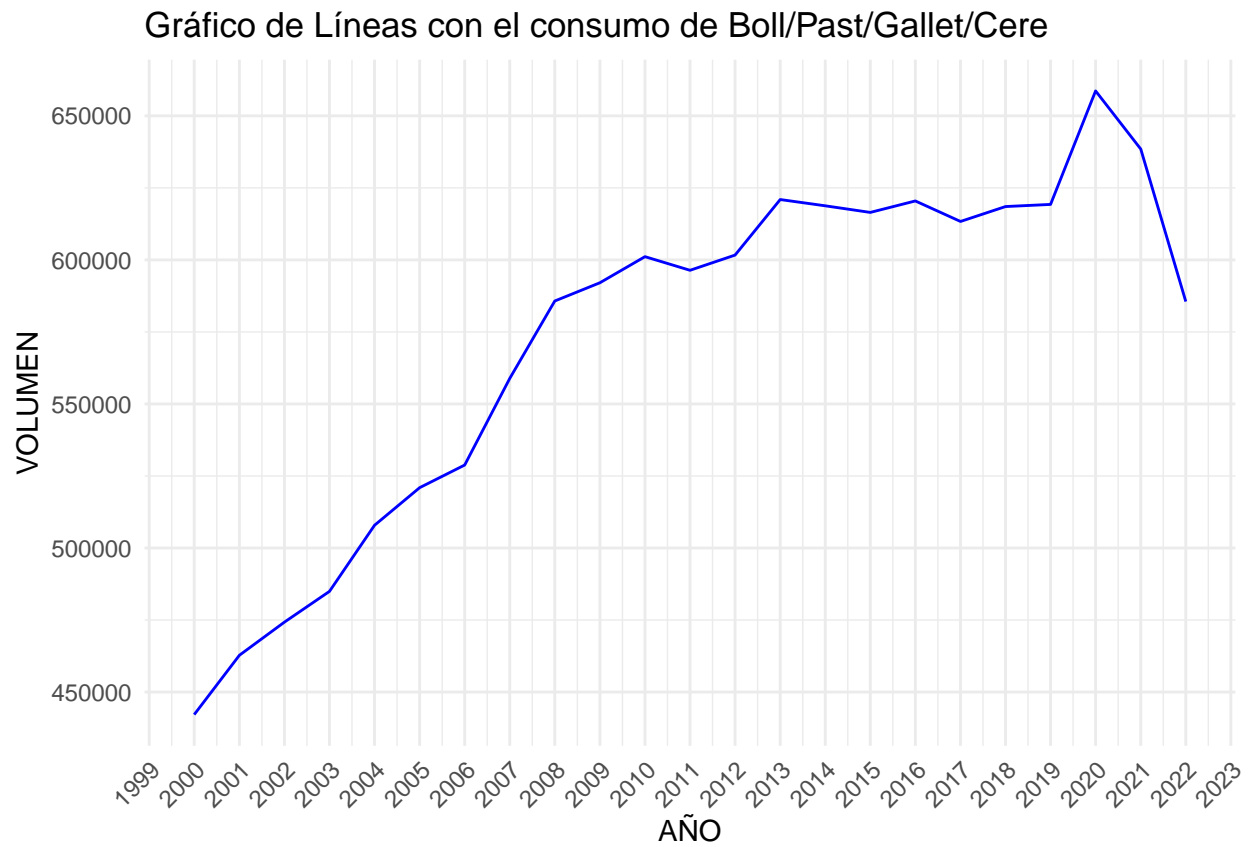


Gráfico de Líneas con el consumo de Cafés e infusiones

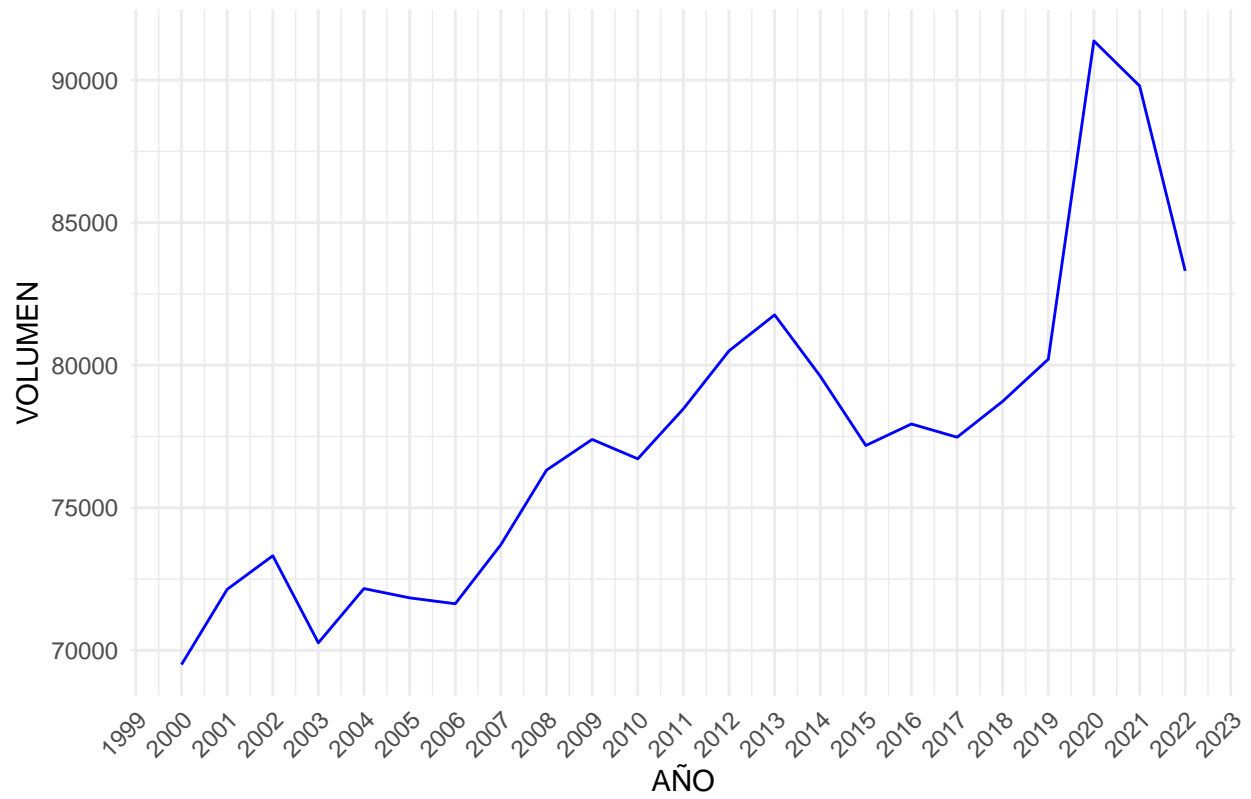


Gráfico de Líneas con el consumo de Caldos

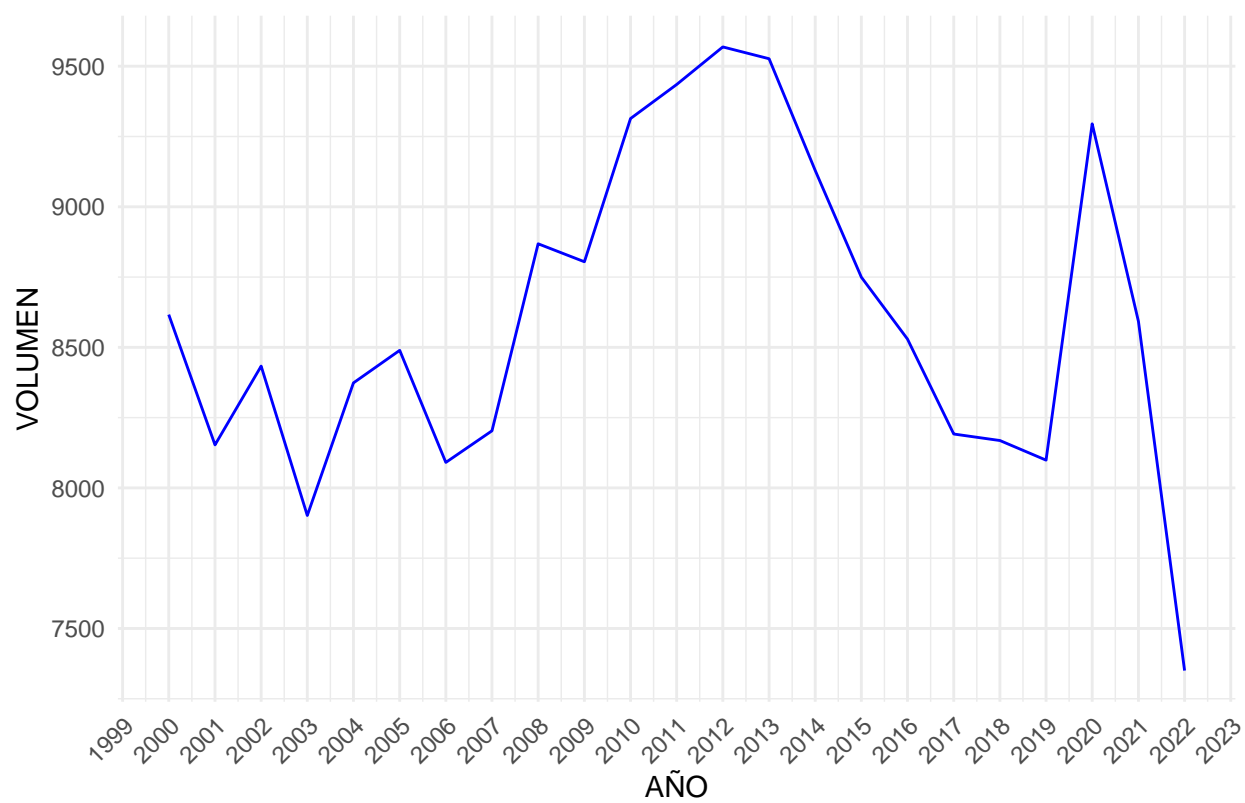


Gráfico de Líneas con el consumo de Carne

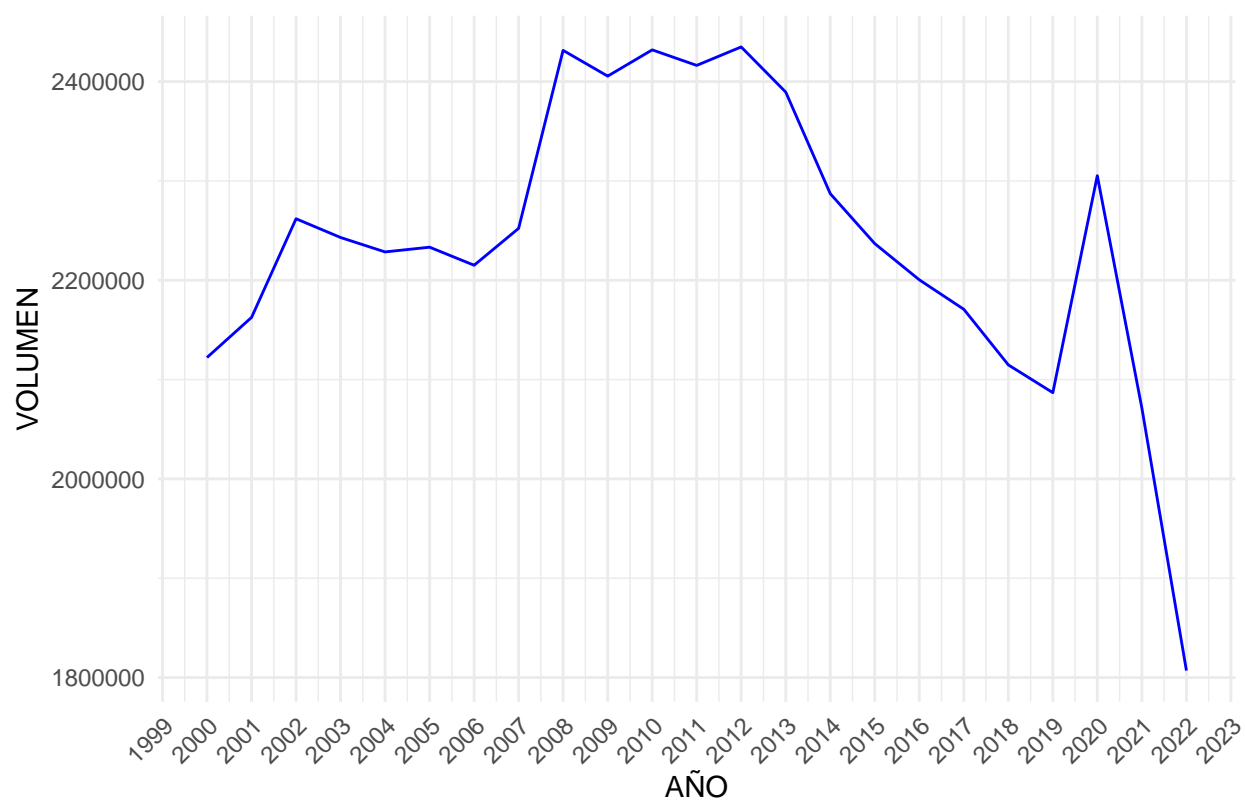


Gráfico de Líneas con el consumo de Choco/Cacao/Suc

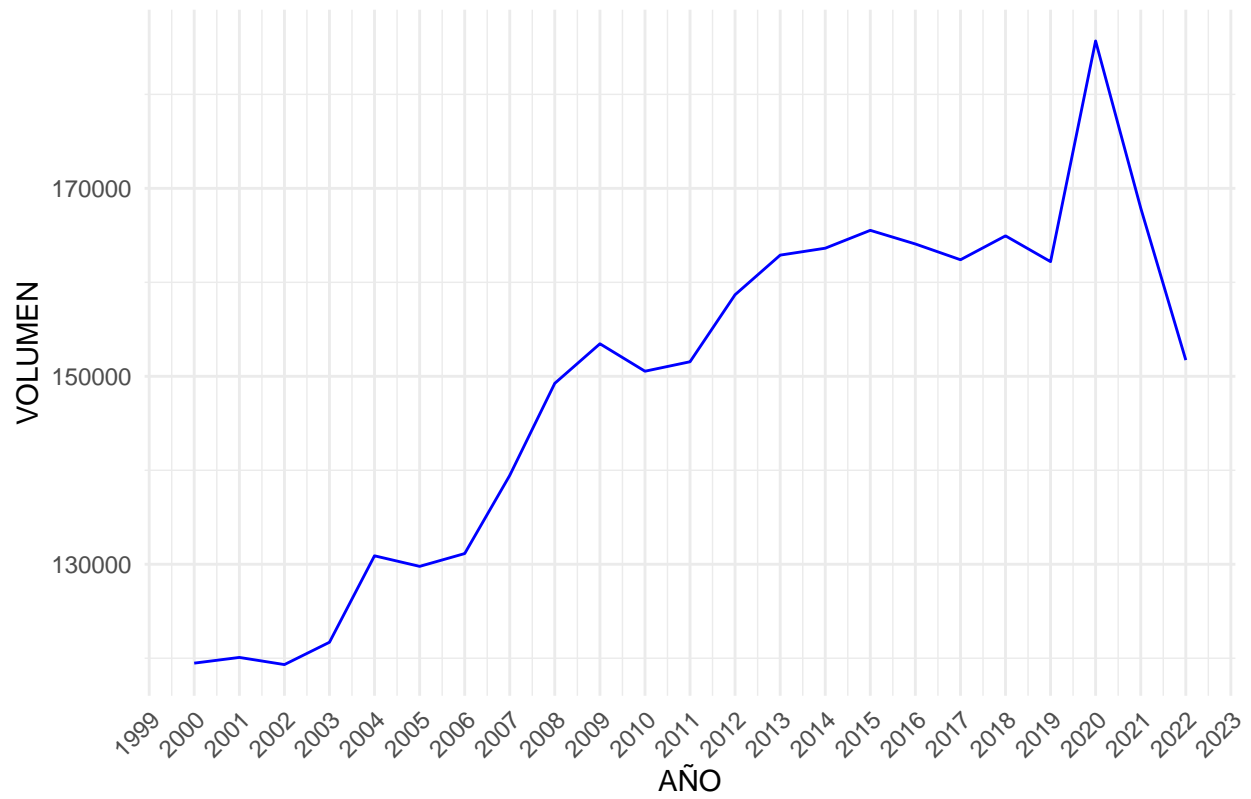


Gráfico de Líneas con el consumo de Derivados lacteos

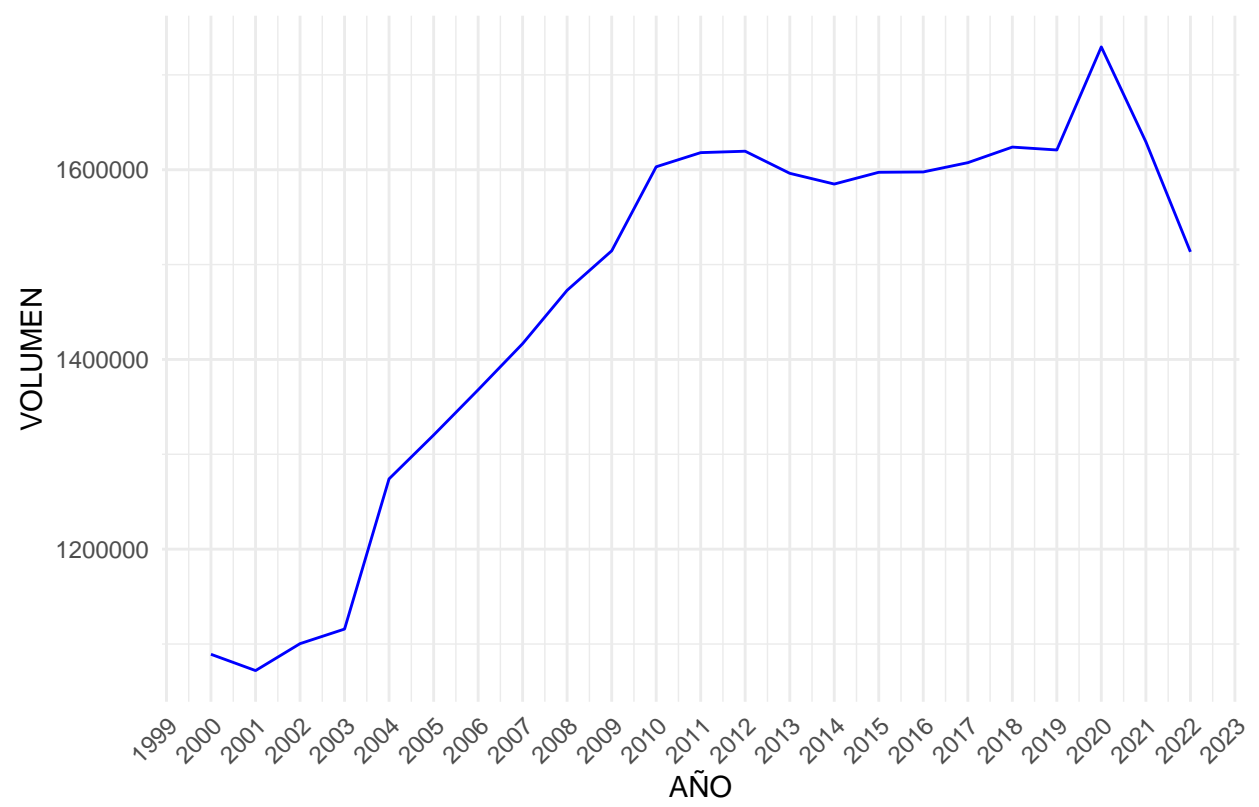


Gráfico de Líneas con el consumo de Edulcorante

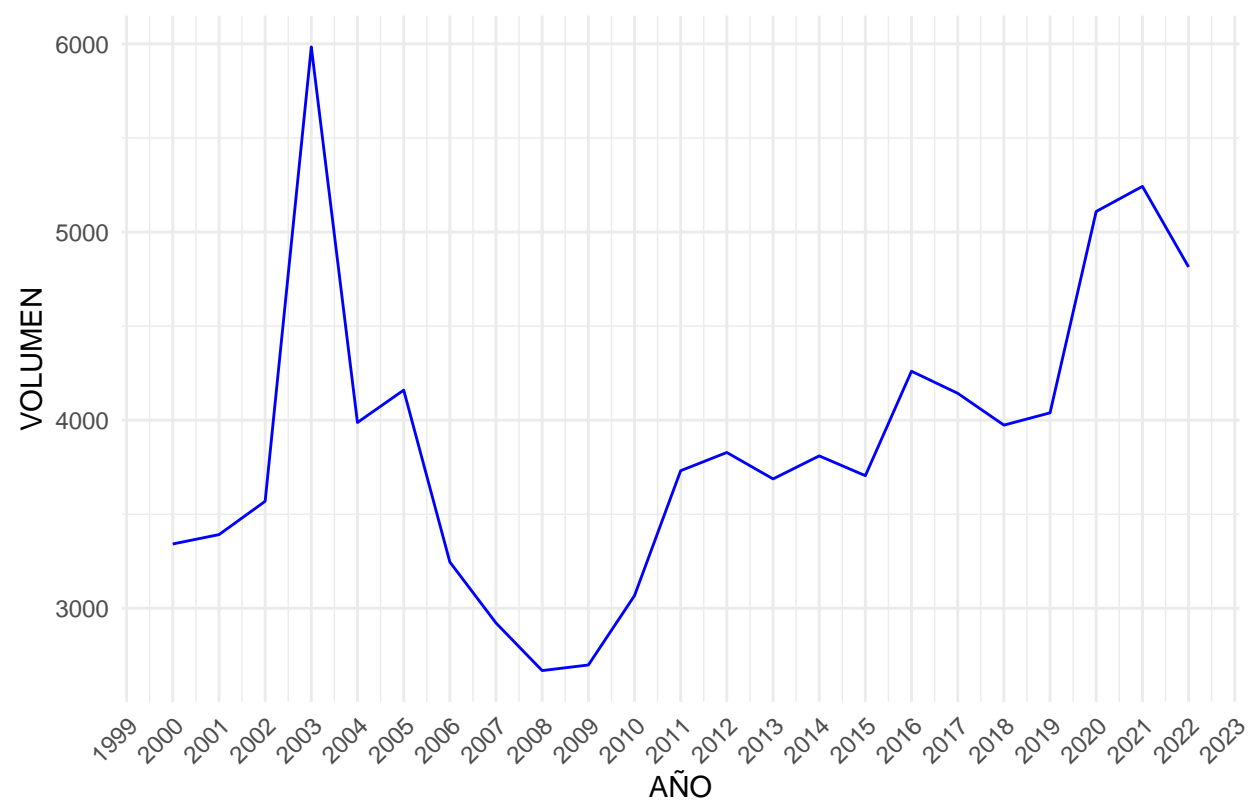


Gráfico de Líneas con el consumo de Encurtidos

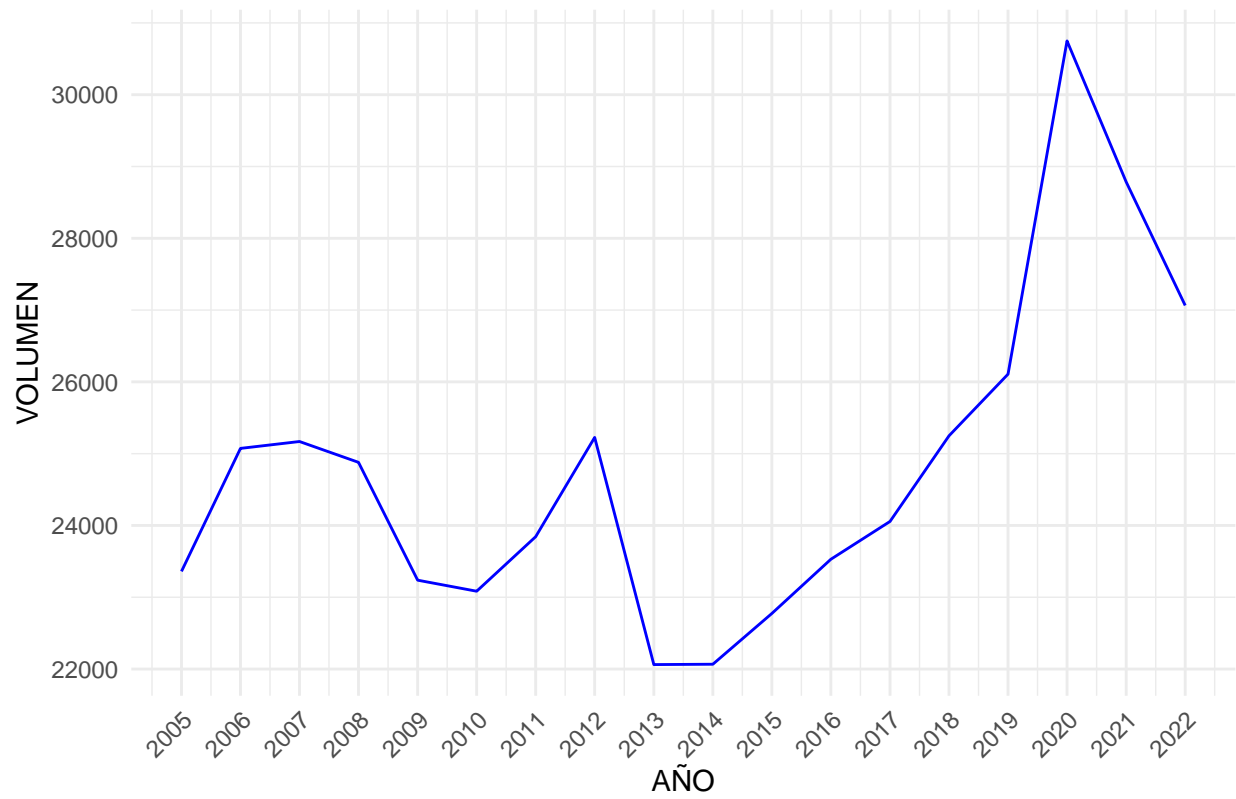


Gráfico de Líneas con el consumo de Especies

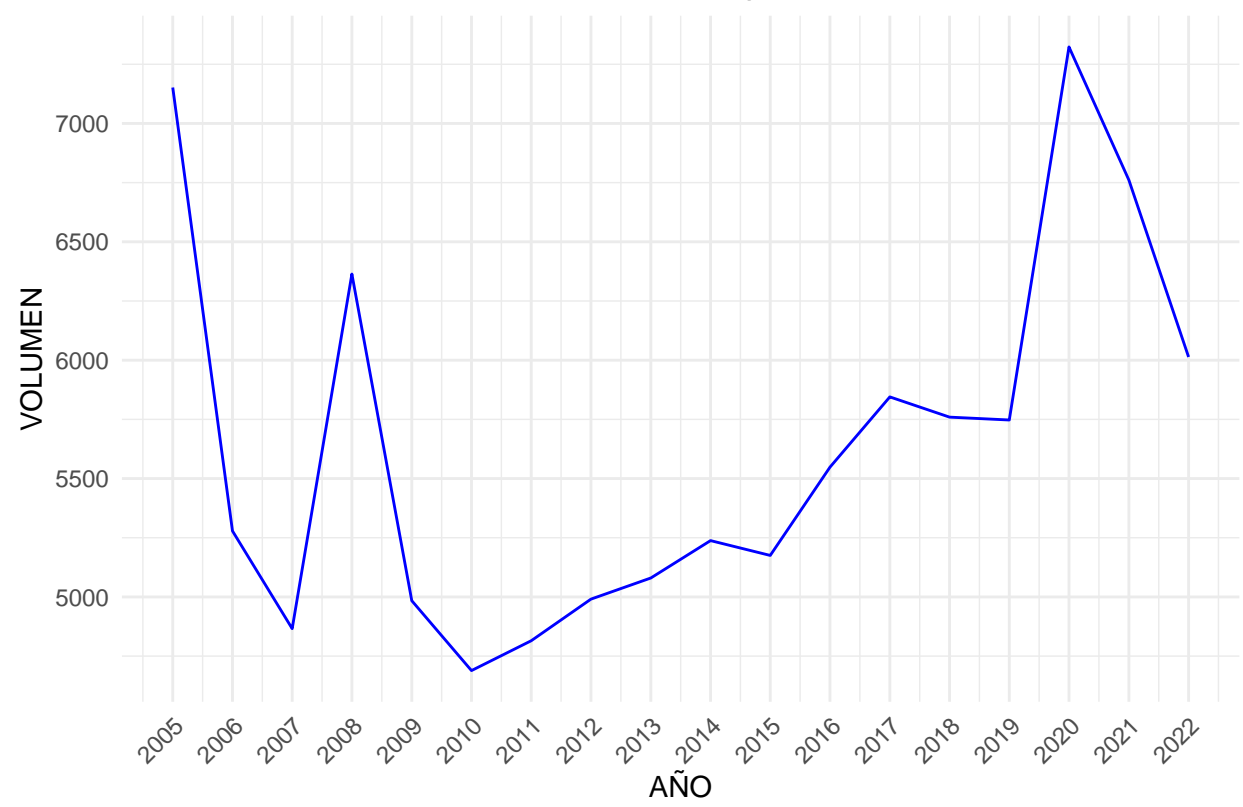
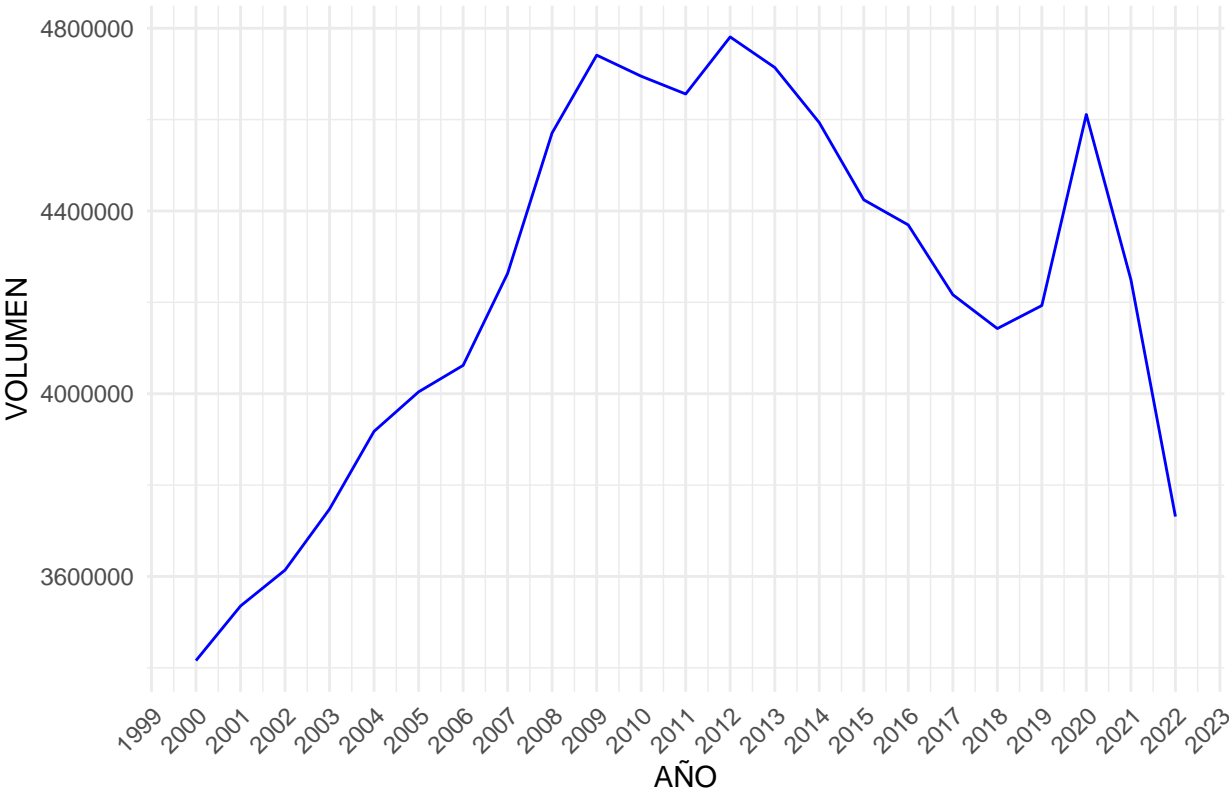
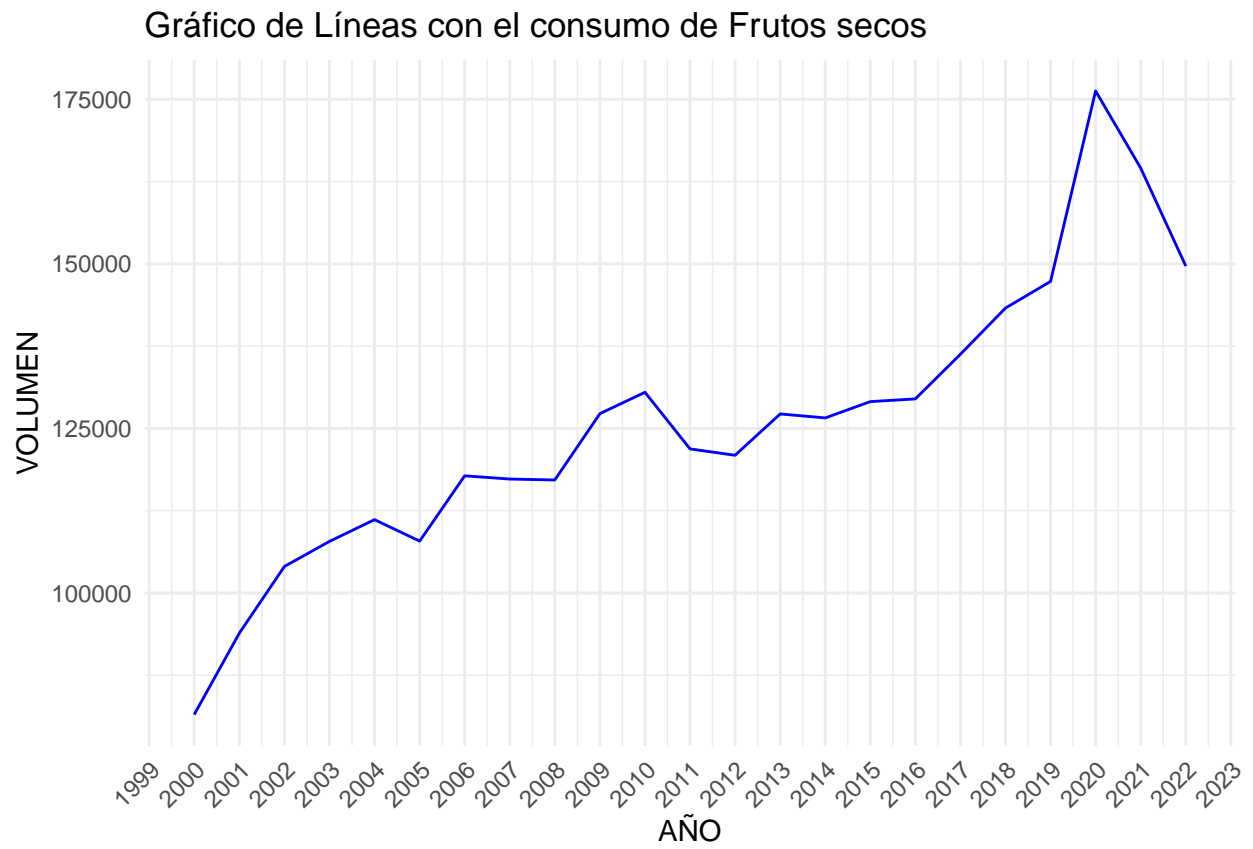


Gráfico de Líneas con el consumo de Frutas frescas





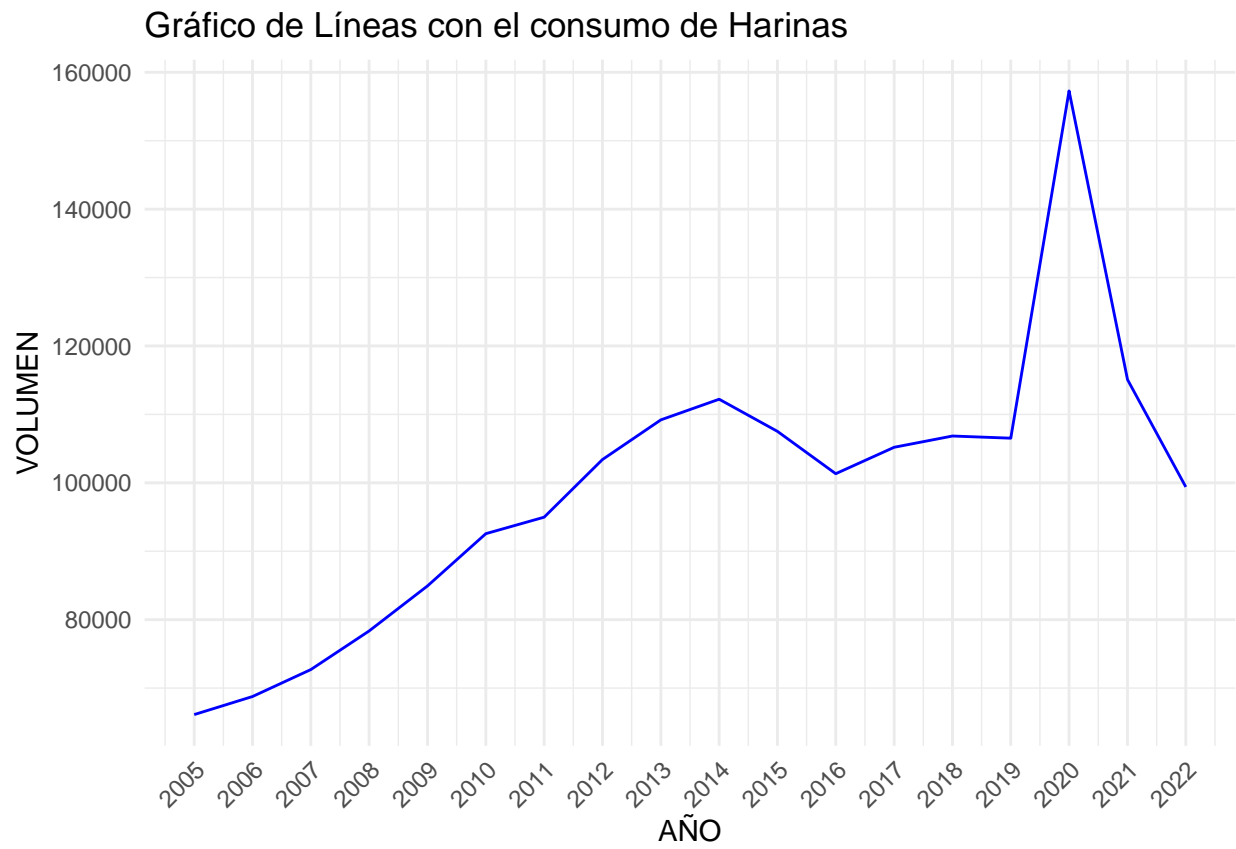


Gráfico de Líneas con el consumo de Hortalizas frescas

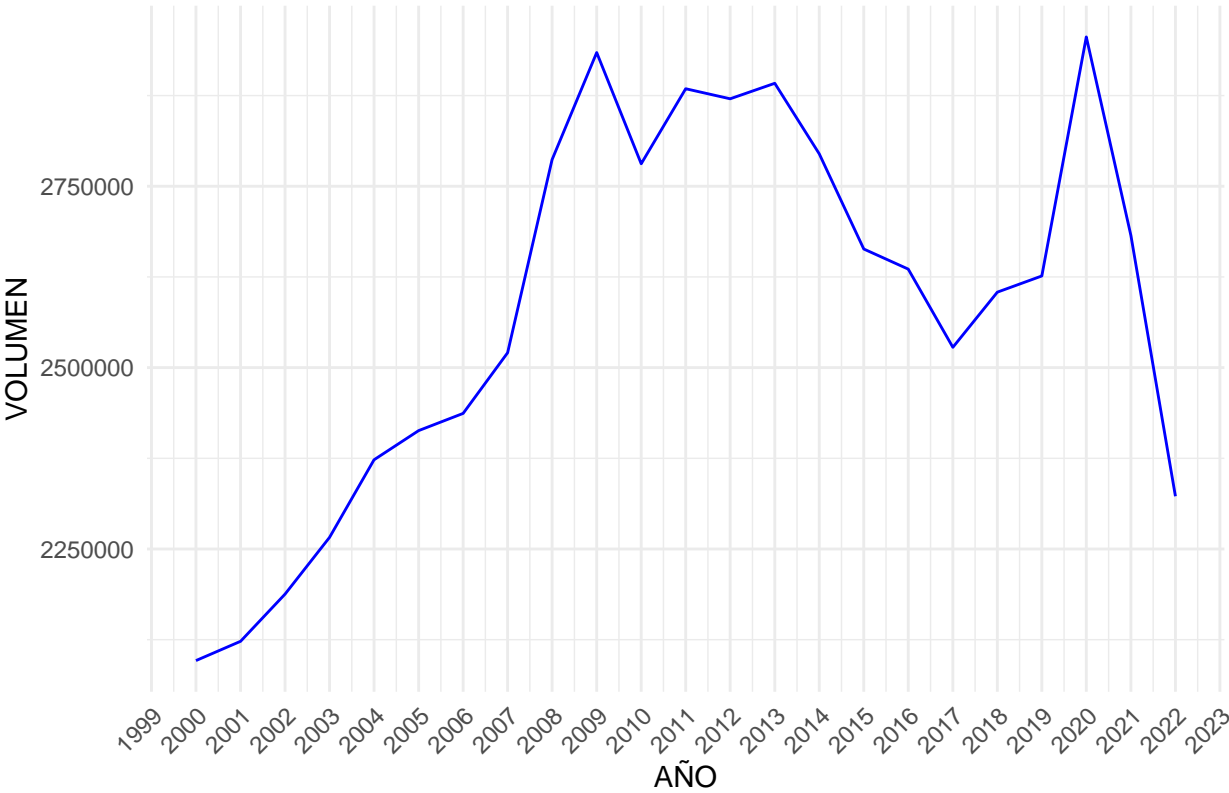
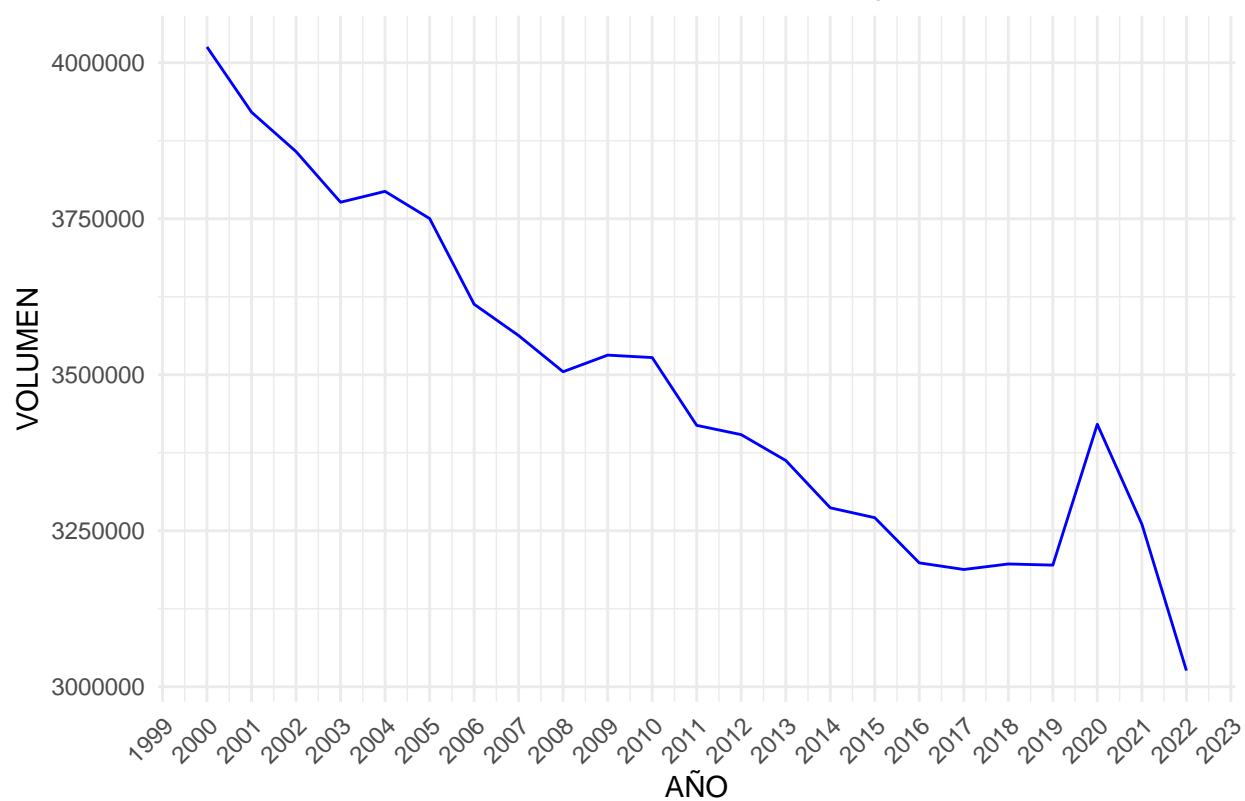




Gráfico de Líneas con el consumo de Leche líquida



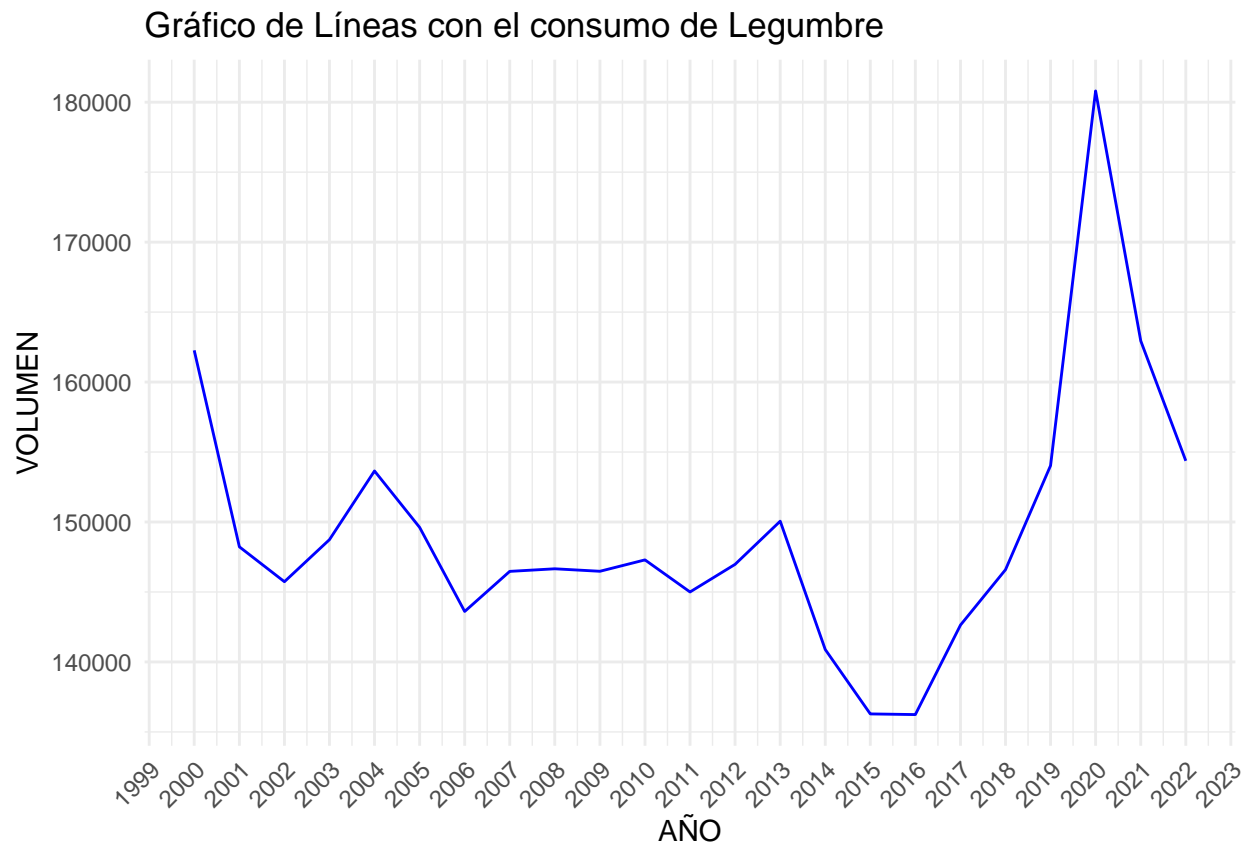


Gráfico de Líneas con el consumo de Margarina

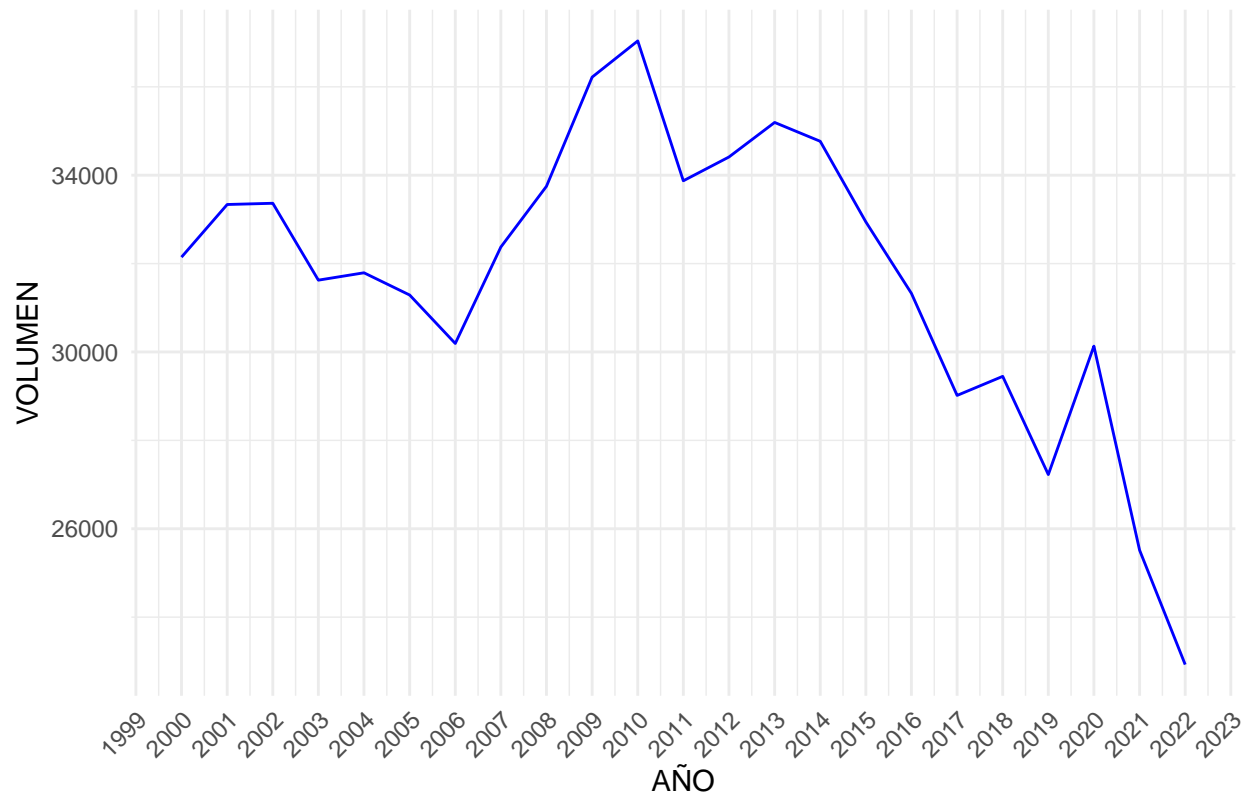


Gráfico de Líneas con el consumo de Miel

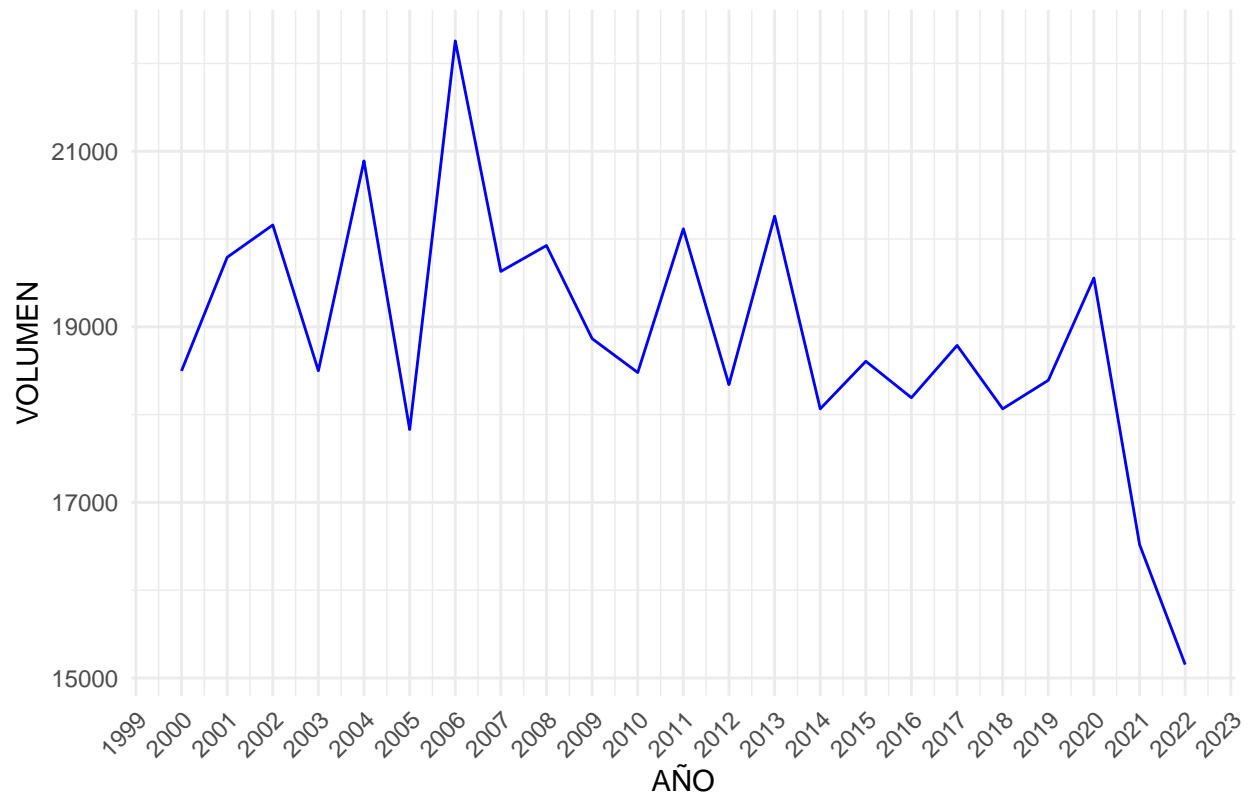
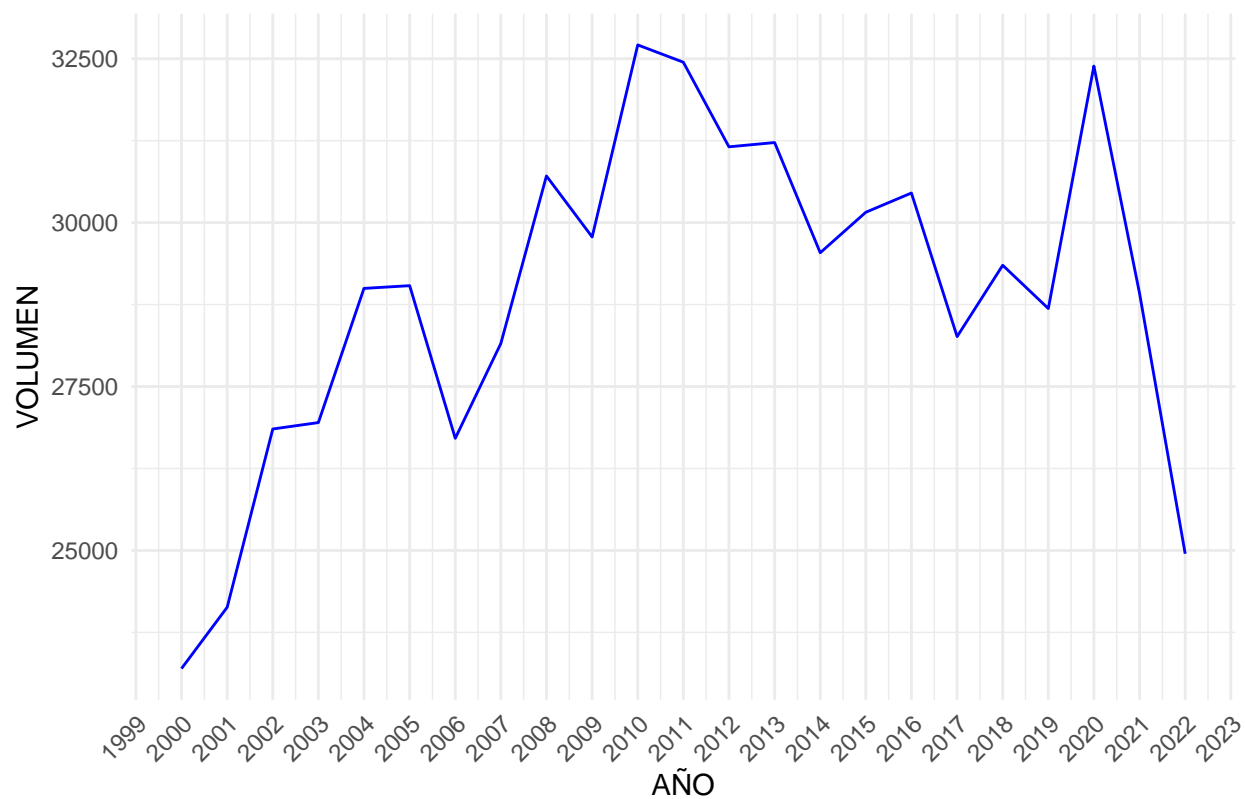
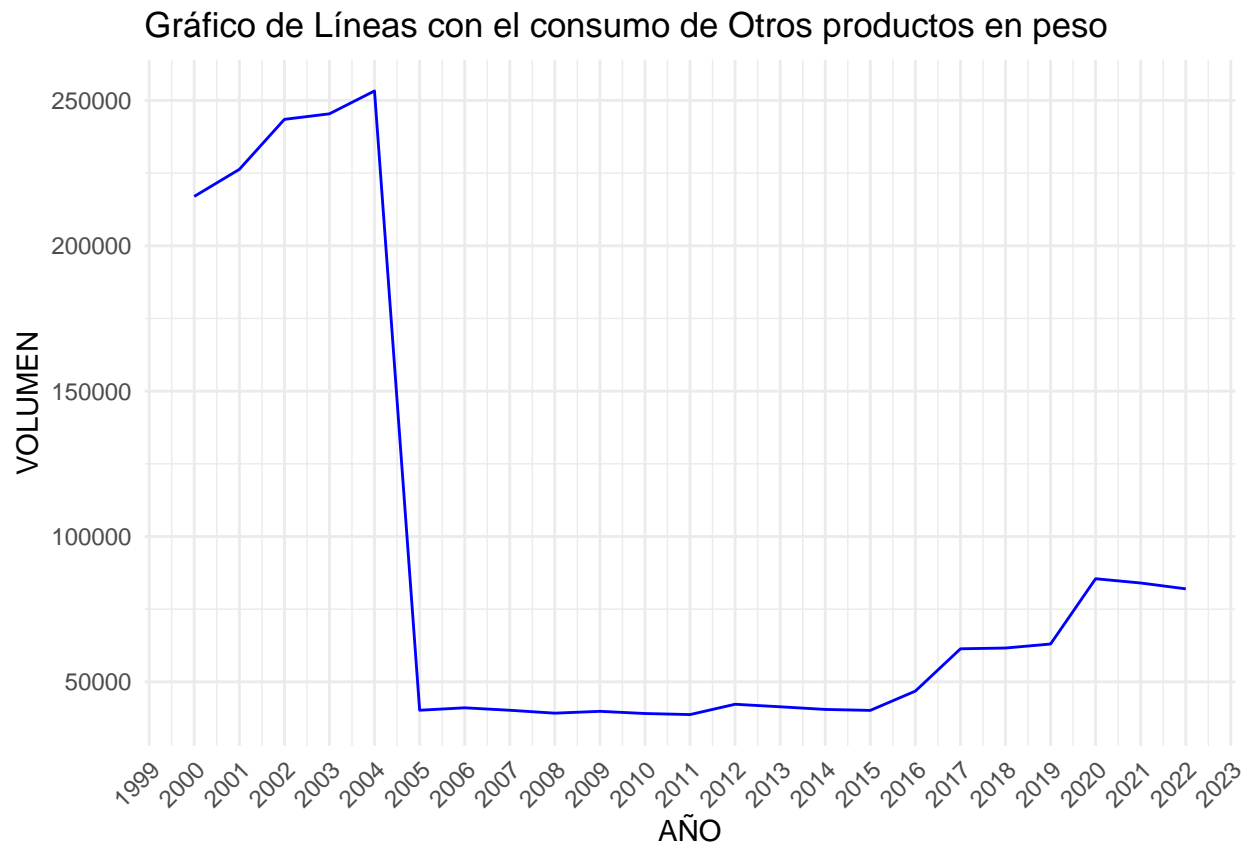


Gráfico de Líneas con el consumo de Otras leches





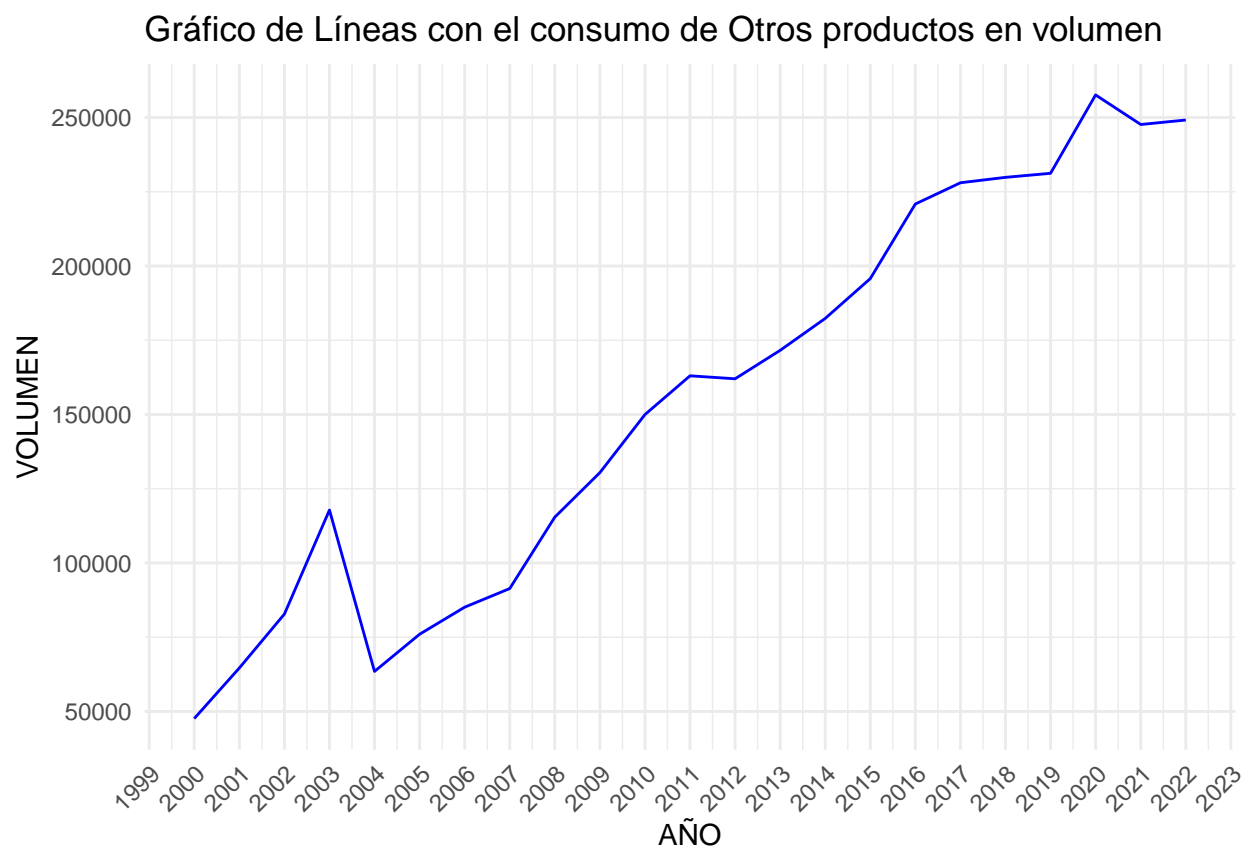


Gráfico de Líneas con el consumo de Pan

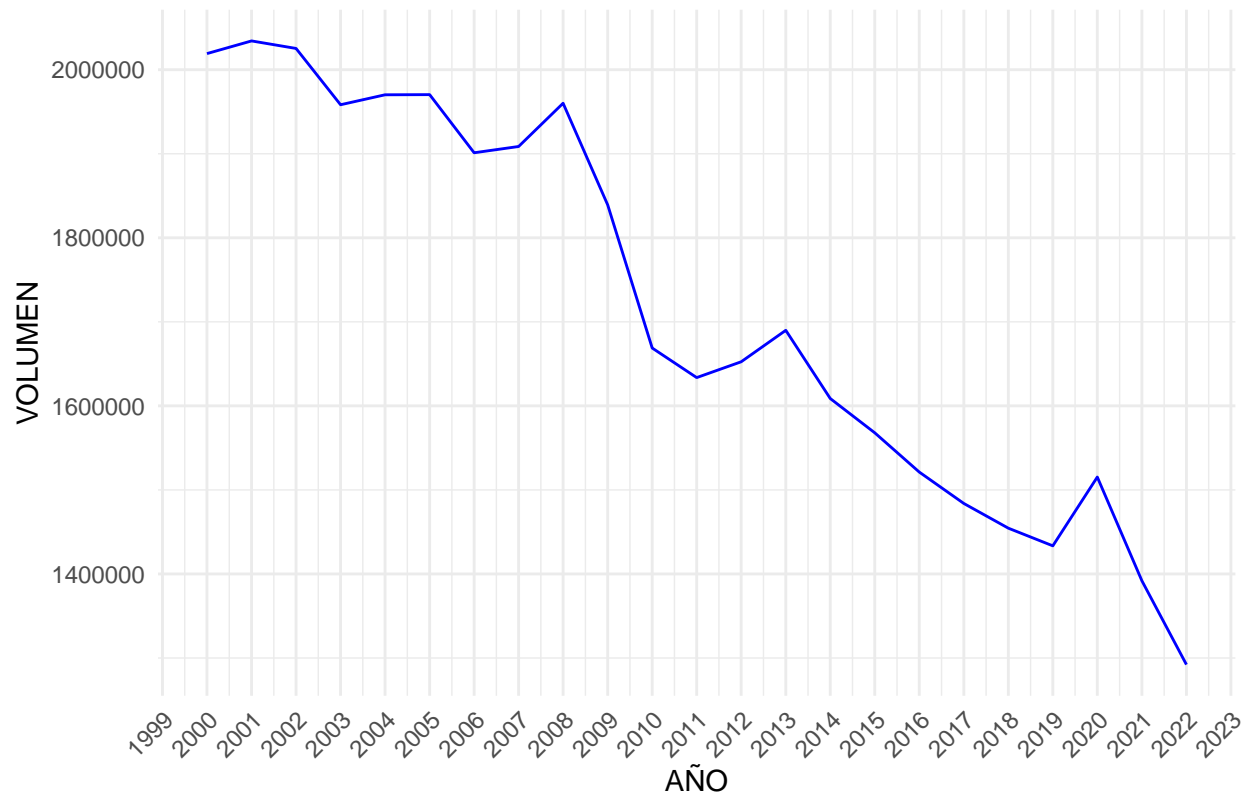


Gráfico de Líneas con el consumo de Pasta

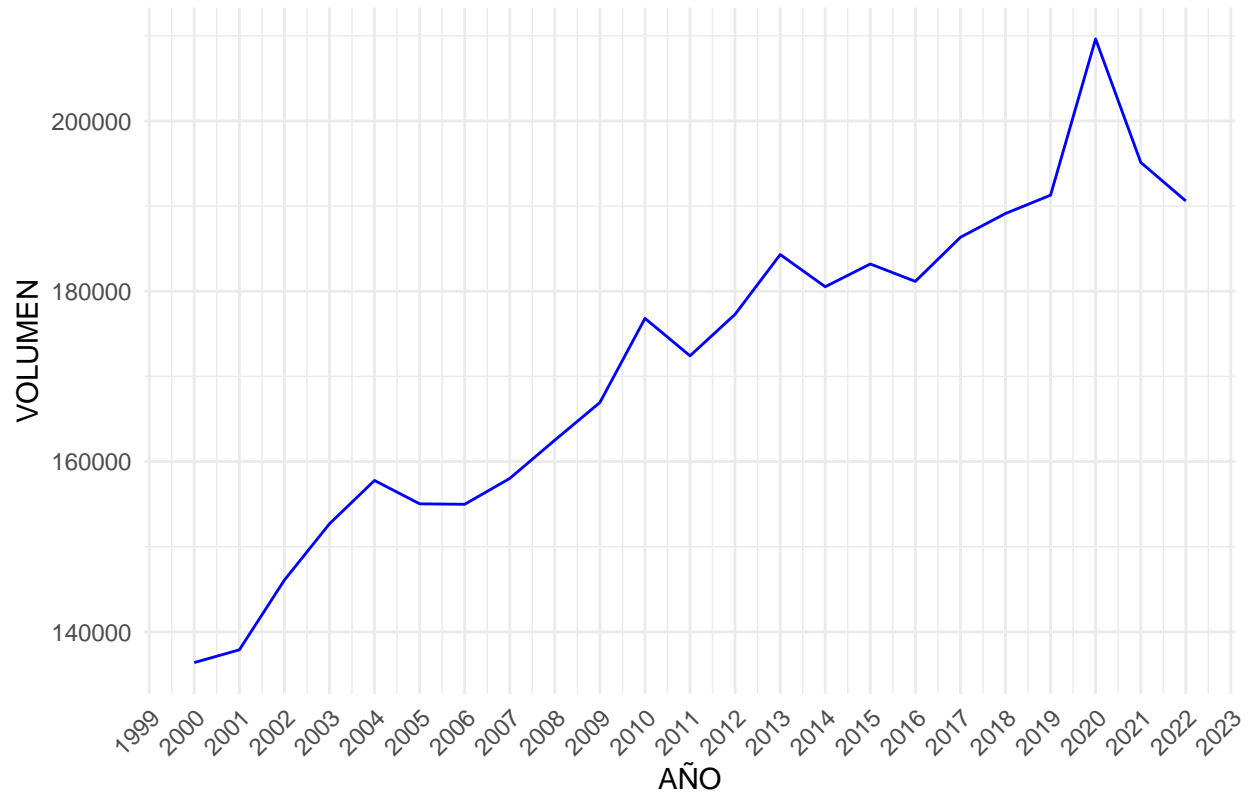
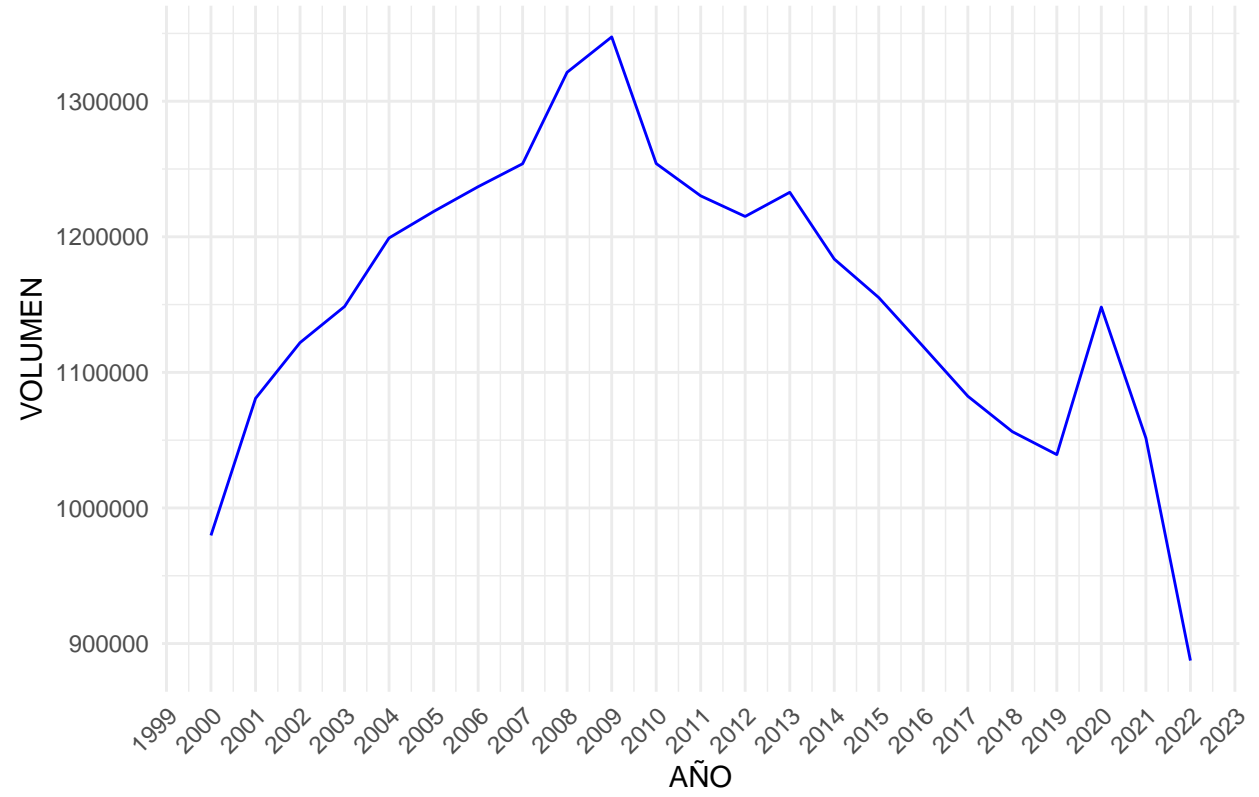




Gráfico de Líneas con el consumo de Pesca



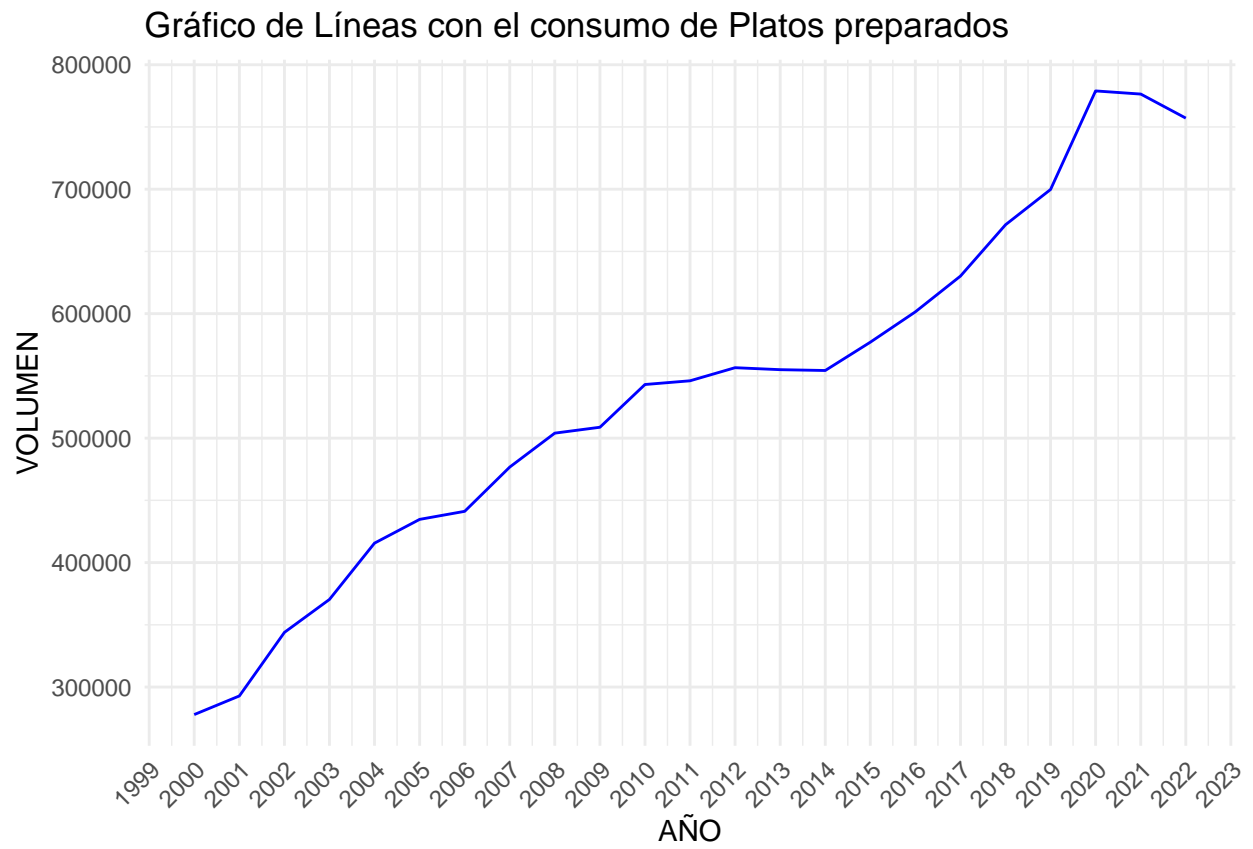


Gráfico de Líneas con el consumo de Preparados lacteos

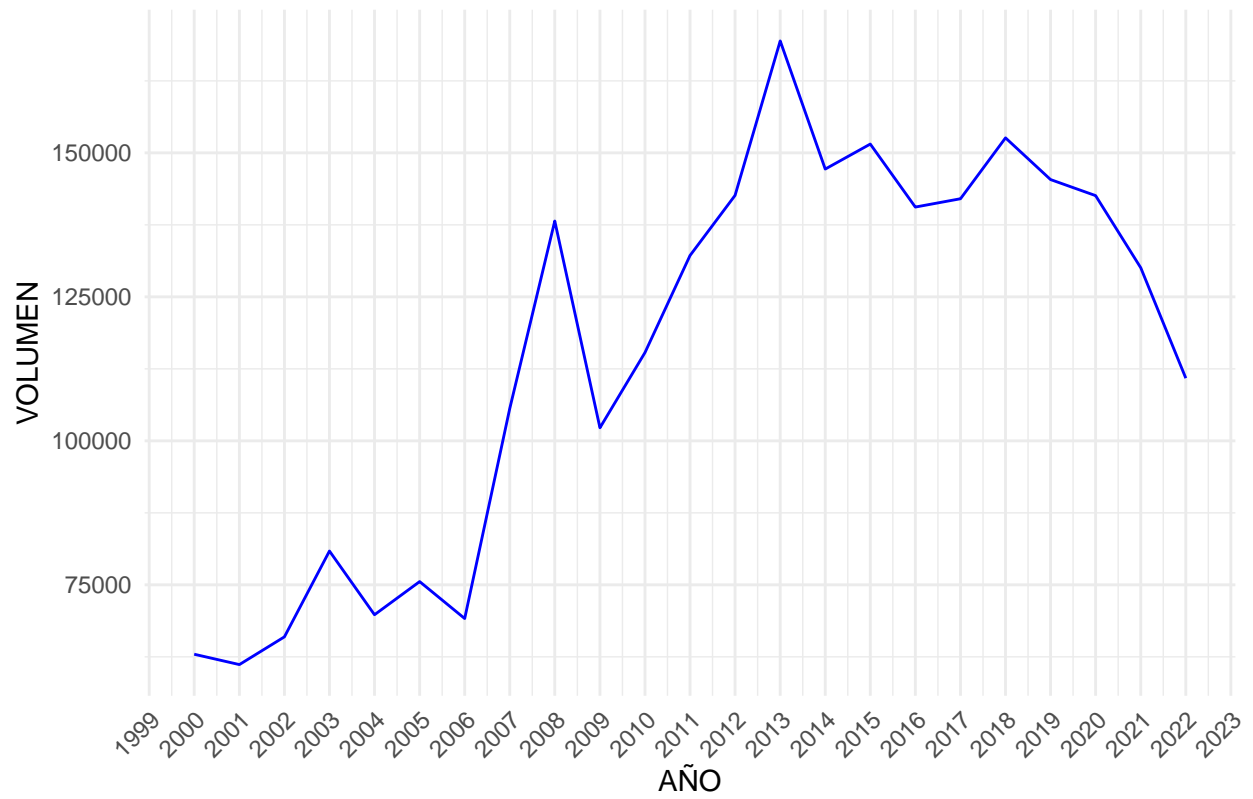


Gráfico de Líneas con el consumo de Refrescos

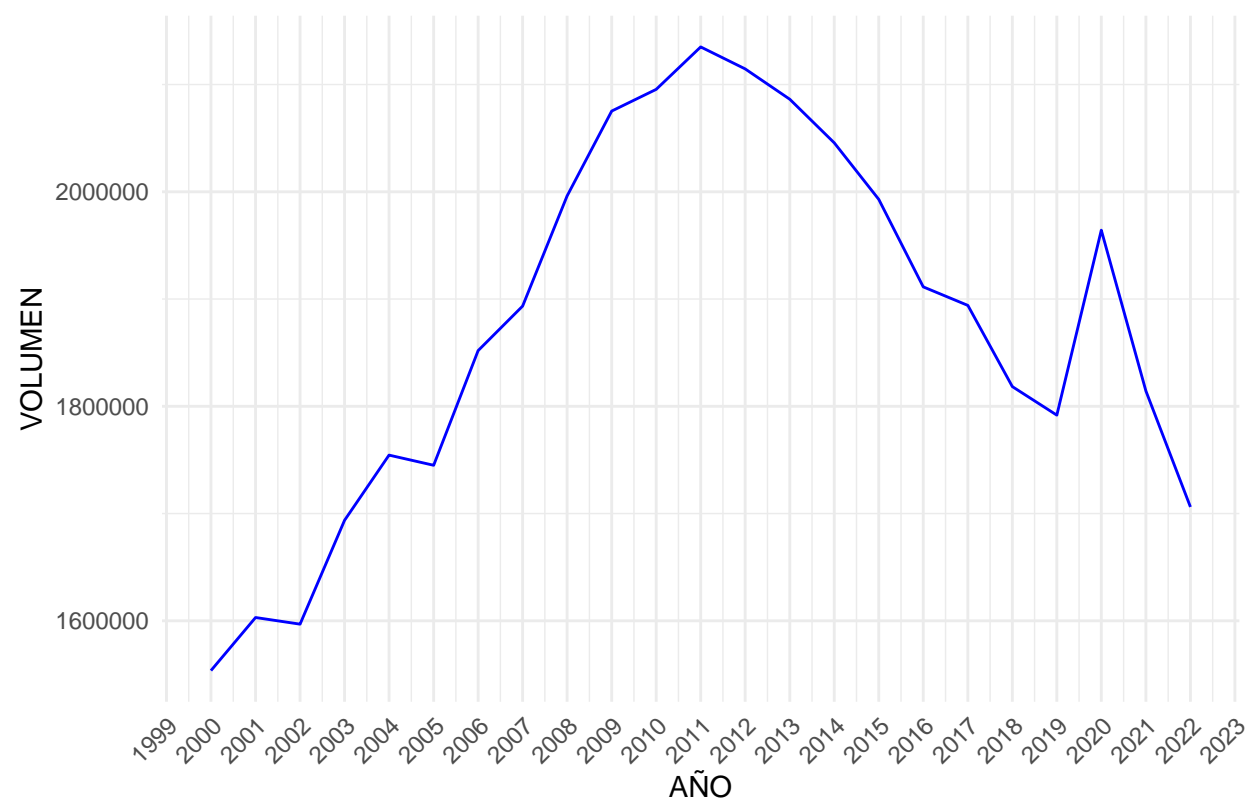
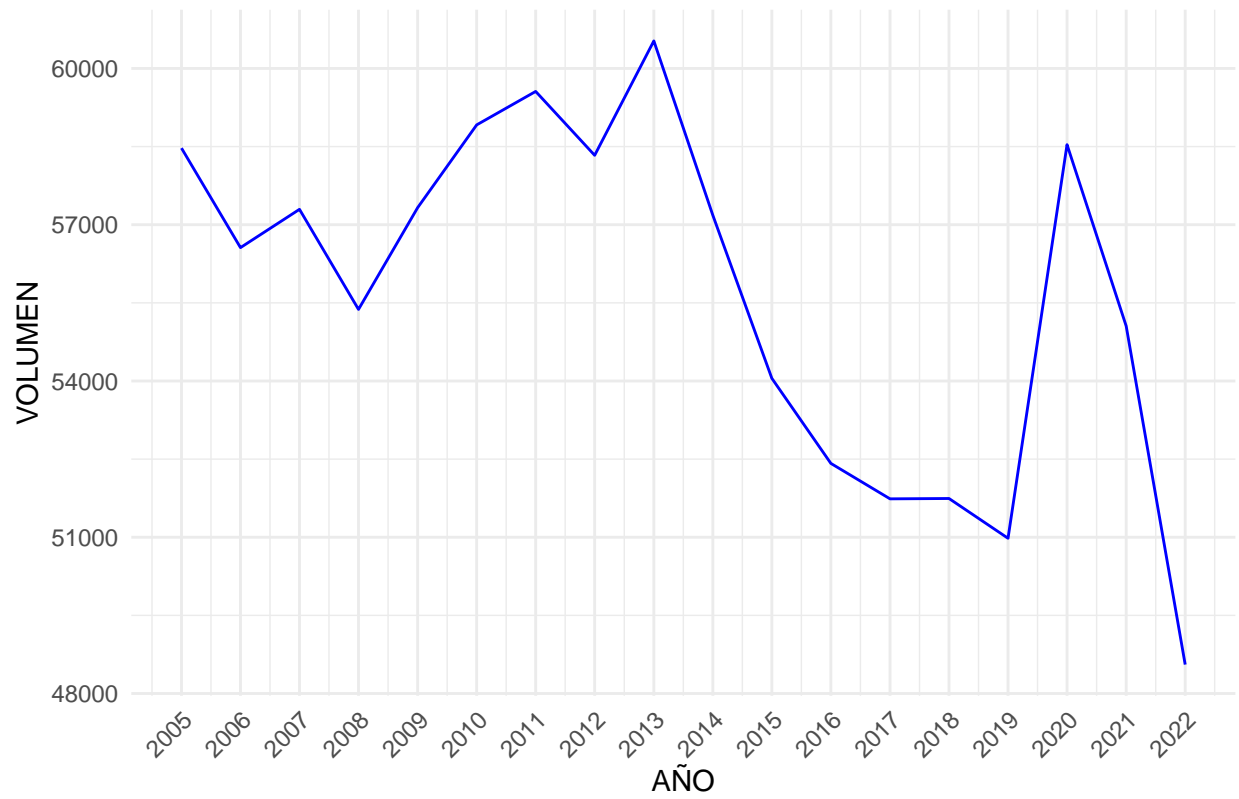


Gráfico de Líneas con el consumo de Sal



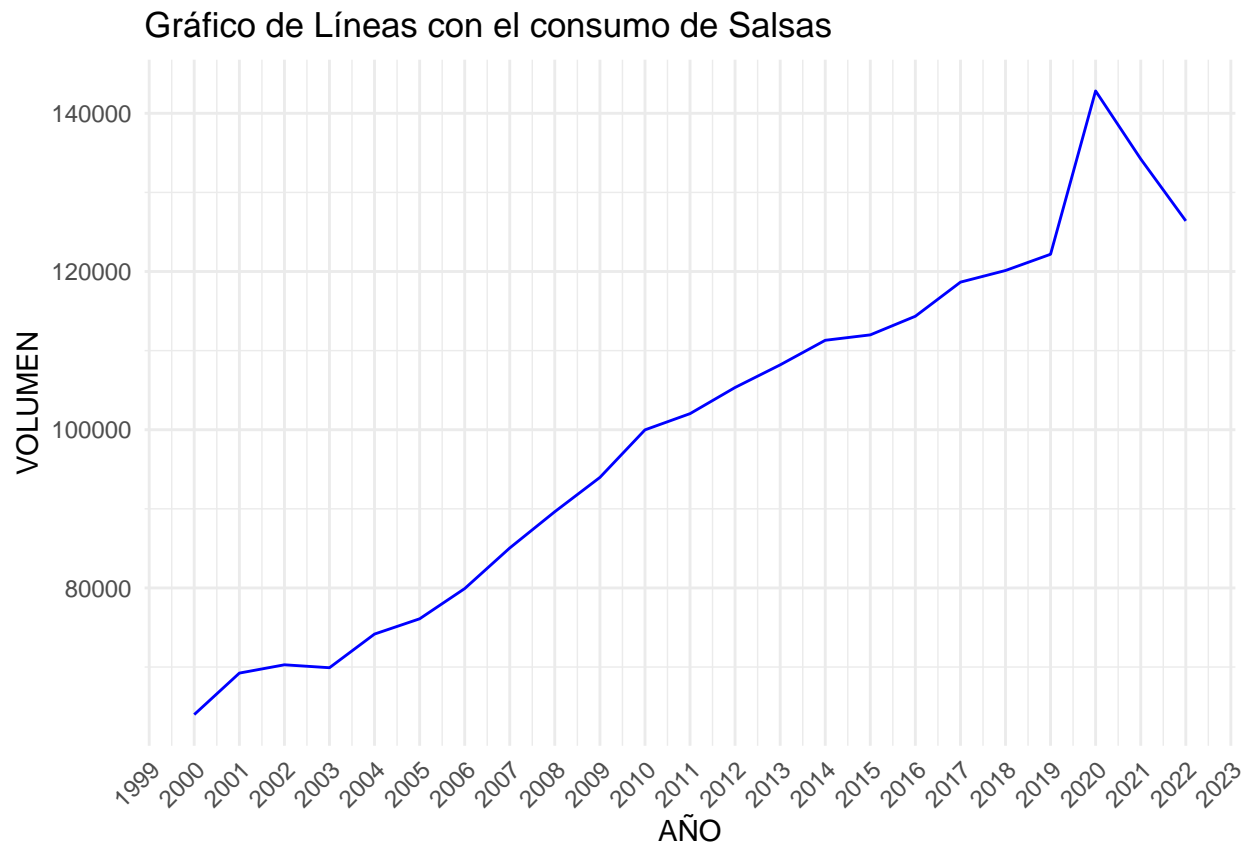


Gráfico de Líneas con el consumo de Vinagre

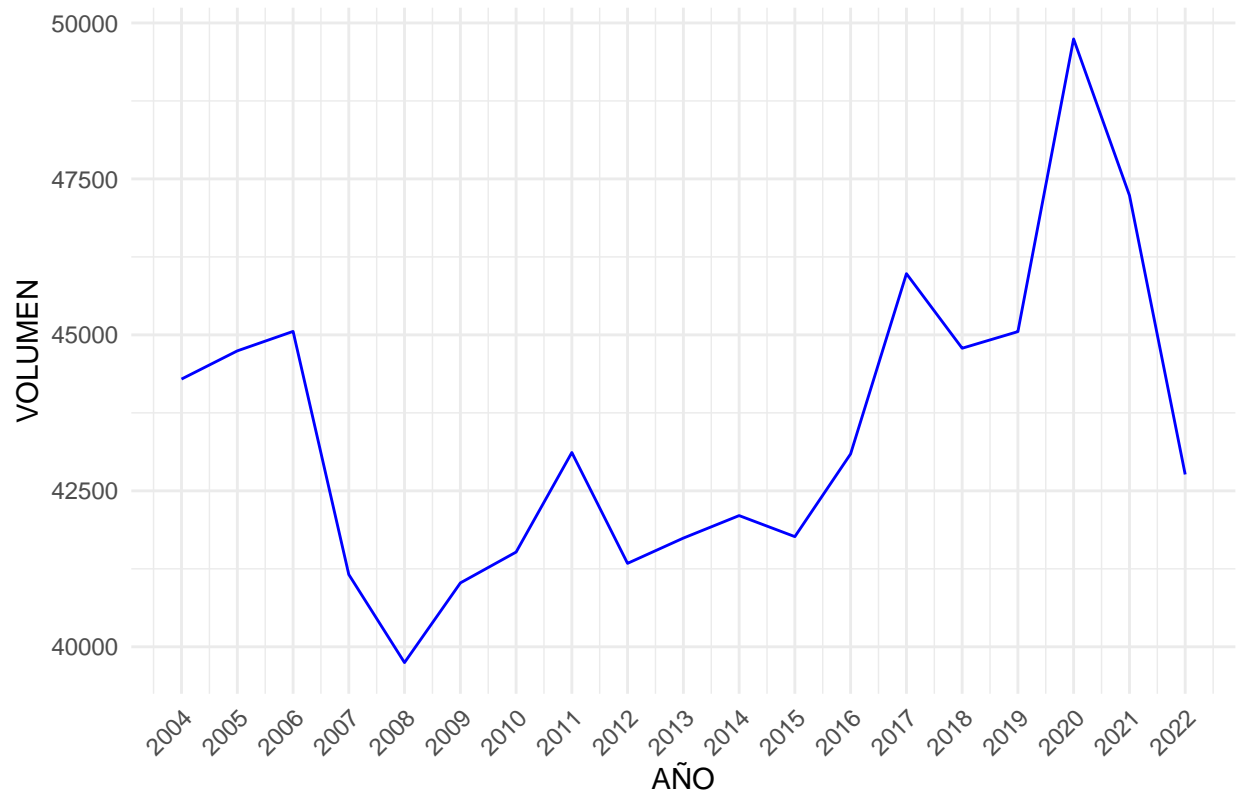
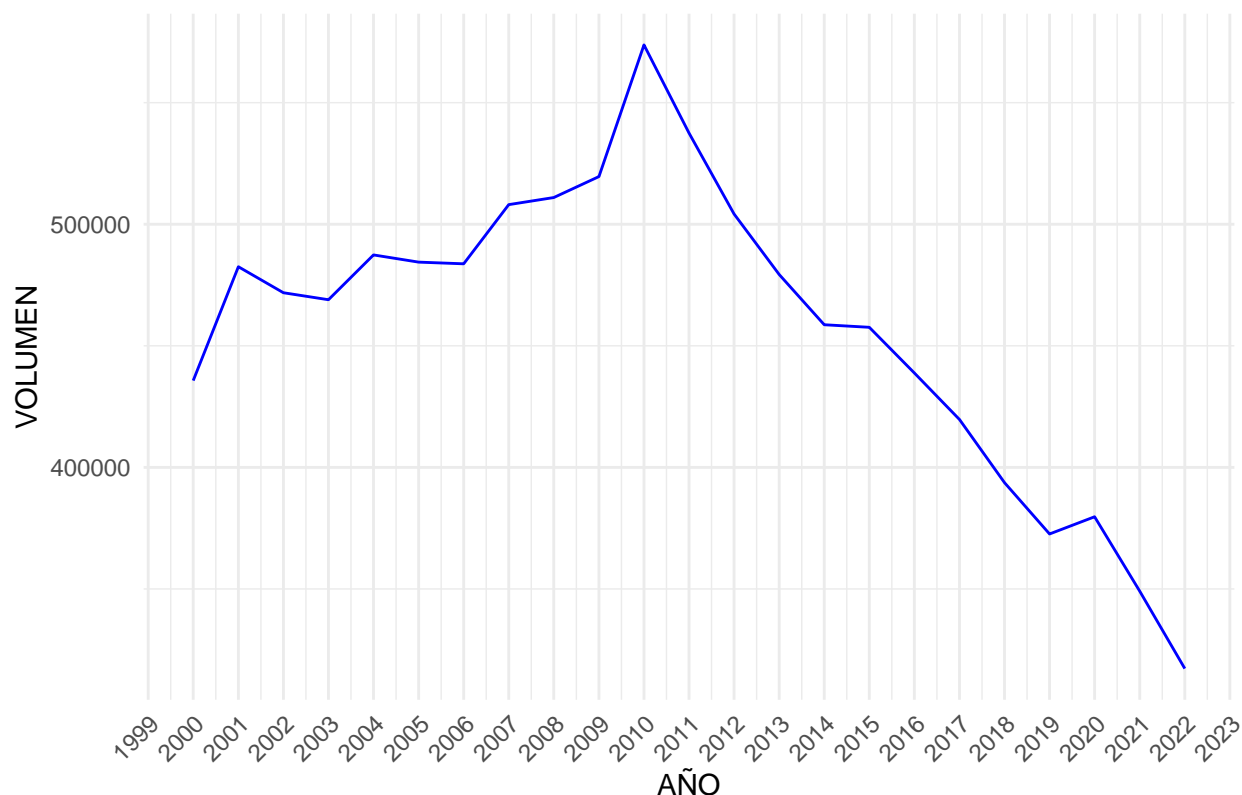


Gráfico de Líneas con el consumo de Zumos



```
rm(data_categorias, categorias_unicas, i, grafico, datos)
```

En general, durante los años 2021 y 2022, se observa una tendencia a la disminución en el consumo en todas las categorías de alimentos.

Sin embargo, al analizar la serie temporal en su totalidad, se destacan varios patrones interesantes:

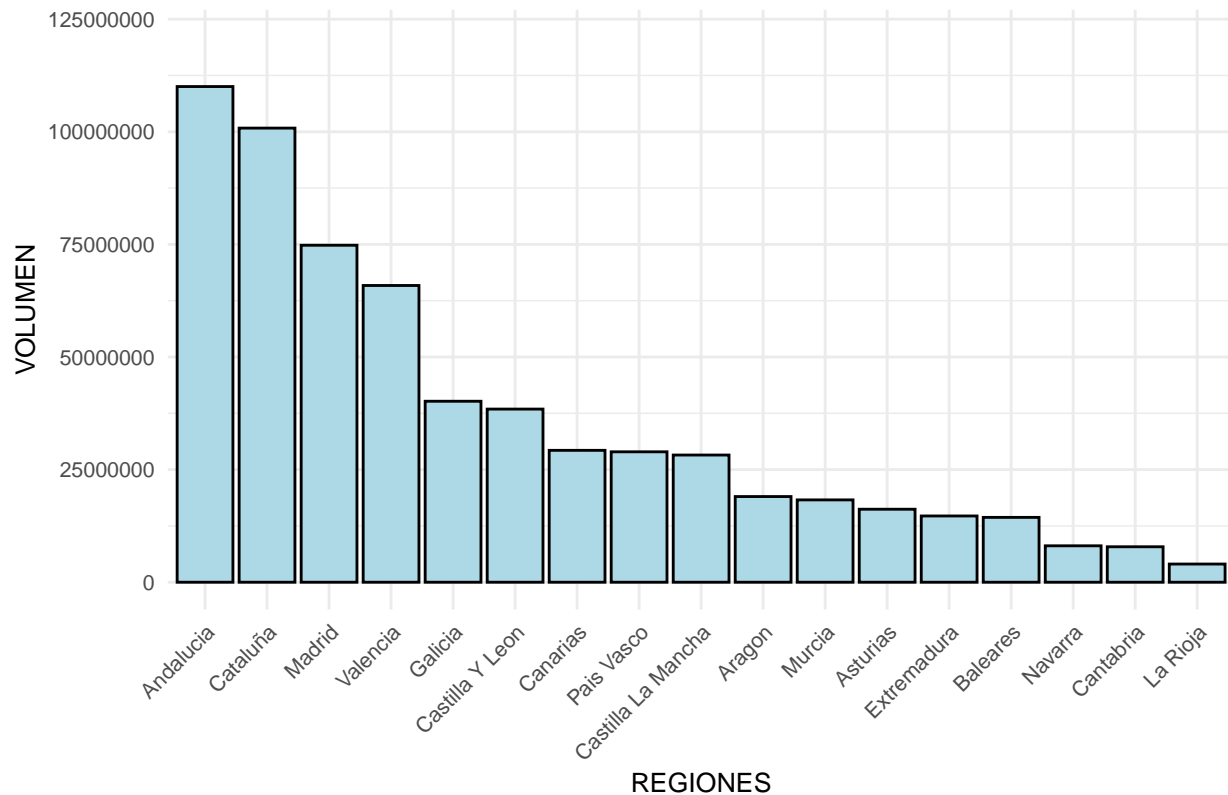
- Algunas categorías, como el agua, bollería, pastelería, galletas y cereales, cafés e infusiones, chocolates, cacao y sucedáneos, derivados lácteos, frutos secos, otros productos en volumen, pasta, platos preparados y salsas, comienzan con un consumo relativamente bajo en el año 2000 y experimentan un aumento constante a lo largo de los años.
- Otras categorías, como el aceite, el azúcar, la leche líquida y el pan, comienzan la serie con un consumo alto y experimentan una disminución gradual a lo largo del tiempo.
- Algunas categorías, como la fruta fresca, las hortalizas frescas, el pescado y los refrescos, comienzan con un consumo bajo y alcanzan un pico en los años 2009 y 2010, pero luego experimentan una disminución hasta 2019. En 2020, se observa un repunte temporal, seguido de una disminución pronunciada en 2021 y 2022.
- Otras categorías, como la sal, la margarina, la carne y los caldos, comienzan con un consumo moderado o alto, mantienen cierta estabilidad durante algunos años y luego experimentan un aumento hasta alcanzar un pico en 2009. Sin embargo, a partir de 2013, su consumo disminuye significativamente hasta 2019. En 2020, se observa nuevamente un repunte, pero luego su consumo cae en 2021 y 2022.
- Por último, el arroz y otros productos en peso comienzan con un consumo alto, pero experimentan una fuerte caída hacia 2003 y 2006. A partir de 2008, su consumo comienza a aumentar nuevamente.

7.4. VOLUMEN VS REGIONES

```
# Agrupa los datos por regiones calculando la suma del VOLUMEN
data_regiones <- data %>%
  group_by(REGIONES) %>%
  summarise(Volumen = sum(VOLUMEN)) %>%
  arrange(desc(Volumen))

# Crea un gráfico de líneas
ggplot(
  data = data_regiones,
  aes(
    x = reorder(REGIONES, -Volumen),
    y = Volumen)) +
  geom_bar(
    stat = "identity",
    position = "dodge",
    fill = "lightblue",
    color = "black",
    width = 0.9) +
  labs(
    x = "REGIONES",
    y = "VOLUMEN",
    title = "Gráfico de Barras VOLUMEN vs. REGIONES") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
        axis.text.y = element_text(size = 8),
        axis.title = element_text(size = 10),
        legend.position = "none") +
  ylim(0, max(data_regiones$Volumen * 1.1))
```

Gráfico de Barras VOLUMEN vs. REGIONES



```
rm(data_regiones)
```

Como se puede apreciar en el gráfico, las regiones muestran un patrón de consumo de alimentos bastante dispar:

Las regiones de Andalucía y Cataluña lideran en consumo de alimentos, destacándose por su alta demanda de productos.

Les siguen de cerca en términos de consumo las comunidades de Madrid y Valencia, que también presentan cifras considerables.

Galicia y Castilla y León ocupan el tercer lugar en cuanto al volumen de consumo de alimentos.

En cuarto lugar se encuentran las comunidades autónomas de Canarias, País Vasco y Castilla-La Mancha, que tienen un nivel de consumo similar.

Aragón y Murcia se ubican en el siguiente grupo, mostrando cifras moderadas de consumo.

Asturias, Extremadura y Baleares presentan un nivel de consumo relativamente menor en comparación con las regiones mencionadas anteriormente.

Por último, las comunidades de Navarra, Cantabria y, en menor medida, La Rioja muestran los niveles más bajos de consumo de alimentos en el conjunto de datos.

7.4. VOLUMEN VS AÑO VS REGIONES

```
# Agrupa los datos por categorías y años, calculando la suma del VOLUMEN para cada combinación
data_categorias <- data %>%
  group_by(REGIONES, AÑO) %>%
```

```

summarise(
  Volumen = sum(VOLUMEN))

## `summarise()` has grouped output by 'REGIONES'. You can override using the
## `.groups` argument.

# Obtiene las categorías únicas presentes en los datos
regiones_unicas <- unique(data_categorias$REGIONES)

# Crea un gráfico de líneas para cada categoría de alimentos
for (i in regiones_unicas) {

  # Filtra los datos para la categoría actual
  datos <- data_categorias %>%
    filter(REGIONES == i)

  # Crea un gráfico de líneas para la categoría actual
  grafico <- ggplot(
    data = datos,
    aes(
      x = AÑO,
      y = Volumen)) +
    geom_line(
      color = "lightblue") +
    labs(
      x = "AÑO",
      y = "VOLUMEN",
      title = paste0("Gráfico de Líneas con el consumo de ", i)) +
    theme_minimal() +
    scale_x_date(
      date_labels = "%Y",
      date_breaks = "1 year") +
    theme(
      axis.text.x = element_text(angle = 45, hjust = 1))

  # Imprime el gráfico actual
  print(grafico)
}

```

Gráfico de Líneas con el consumo de Andalucía

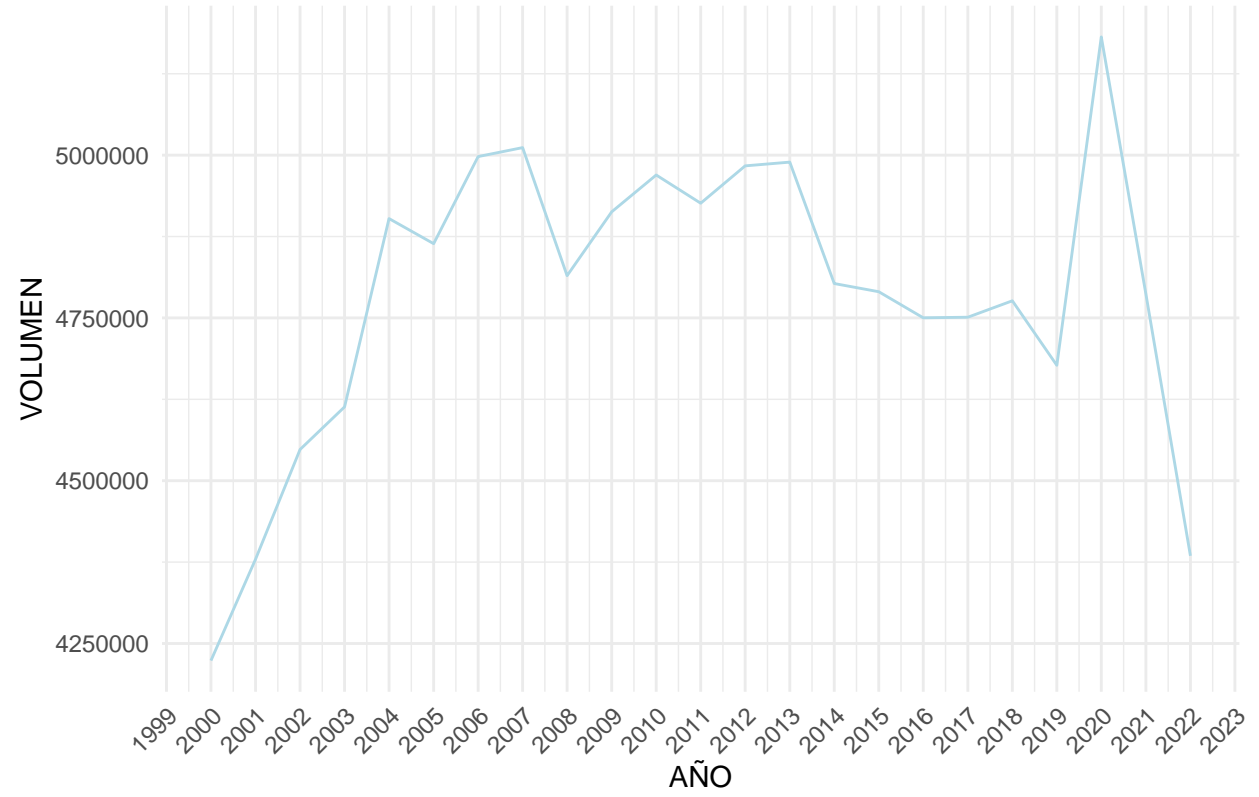


Gráfico de Líneas con el consumo de Aragon

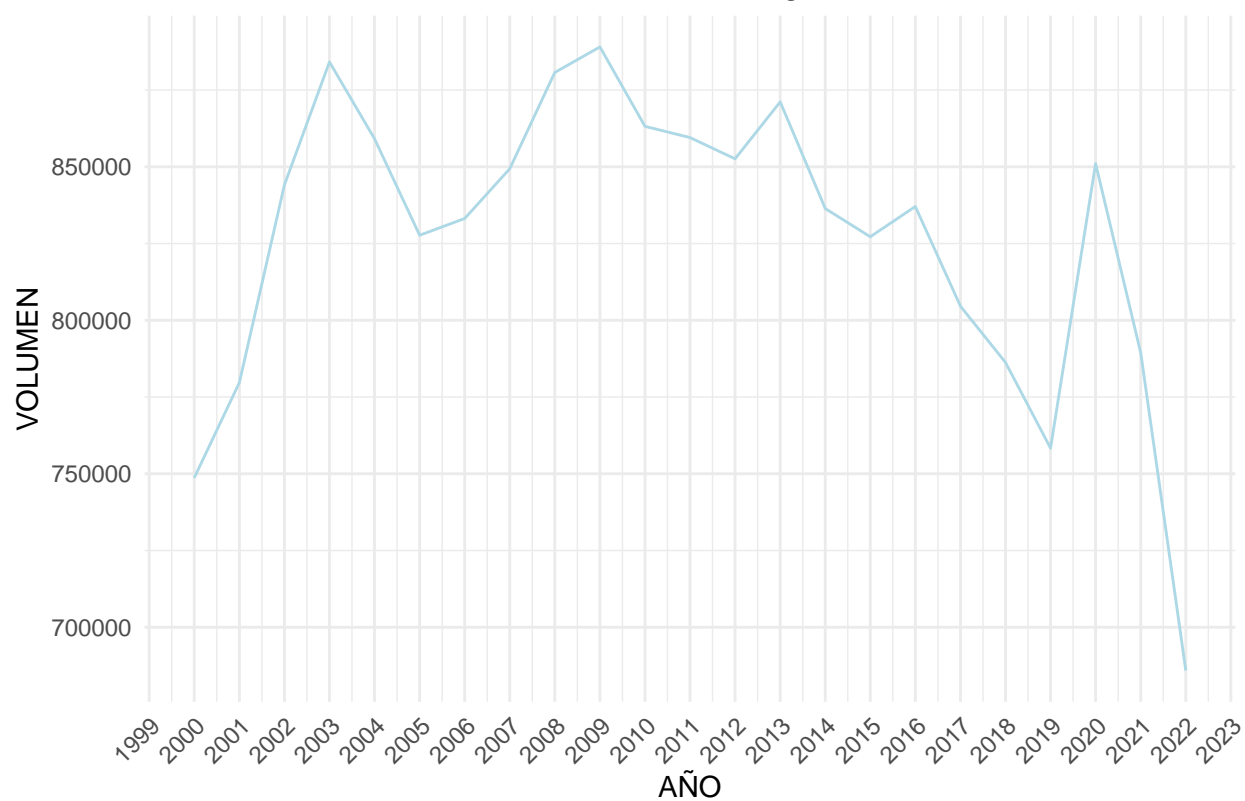


Gráfico de Líneas con el consumo de Asturias



Gráfico de Líneas con el consumo de Baleares

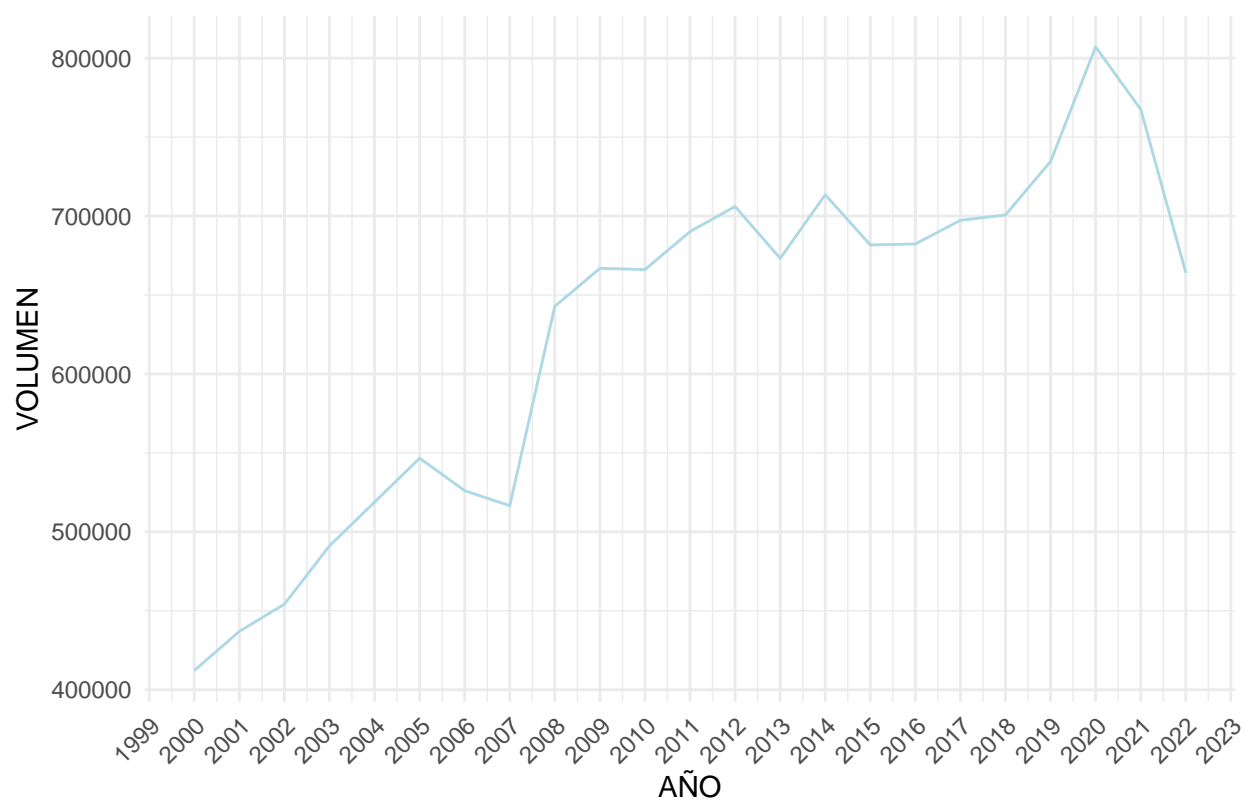


Gráfico de Líneas con el consumo de Canarias

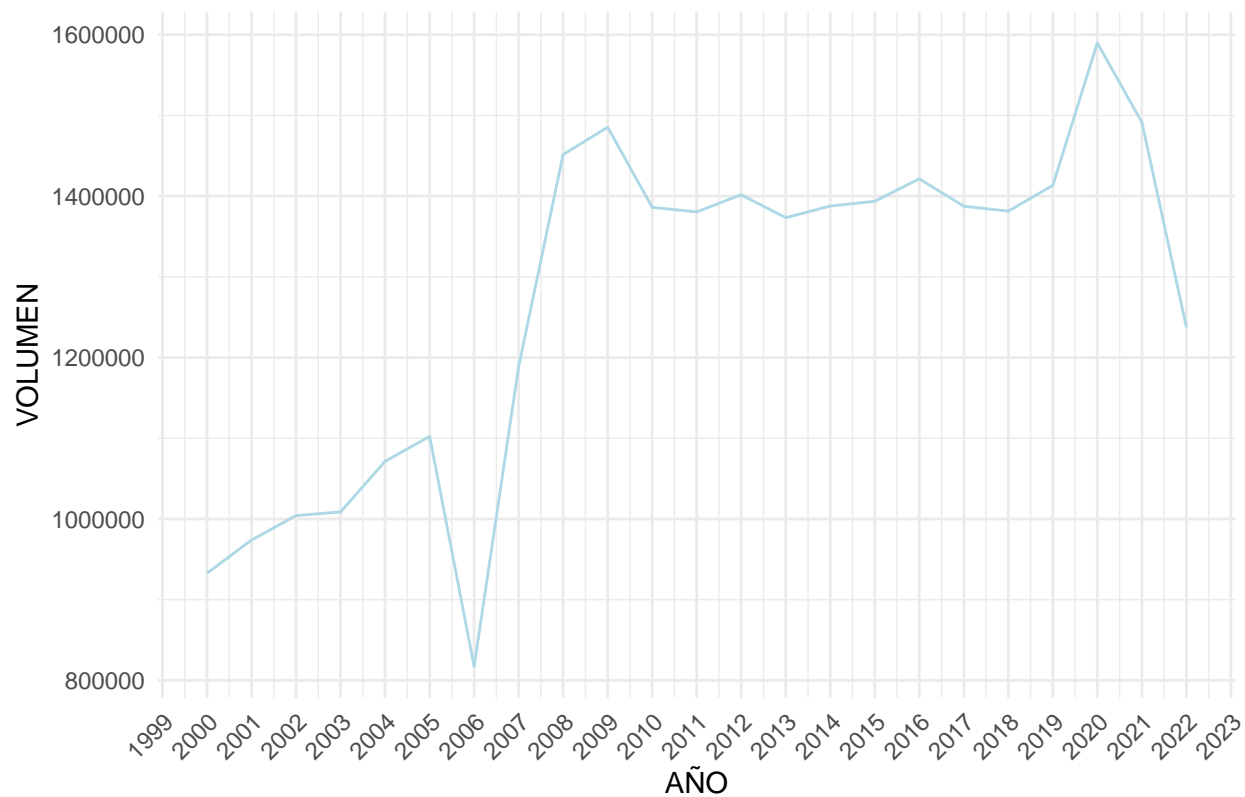


Gráfico de Líneas con el consumo de Cantabria

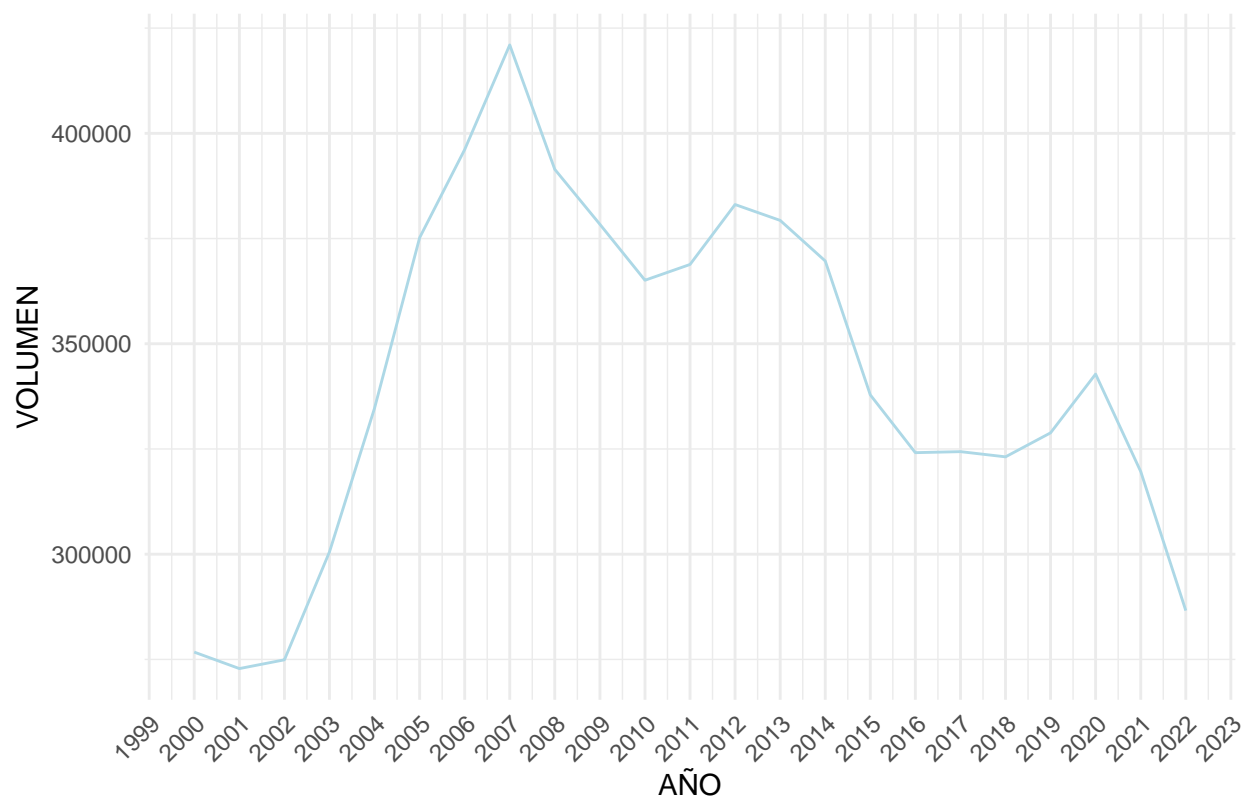


Gráfico de Líneas con el consumo de Castilla La Mancha

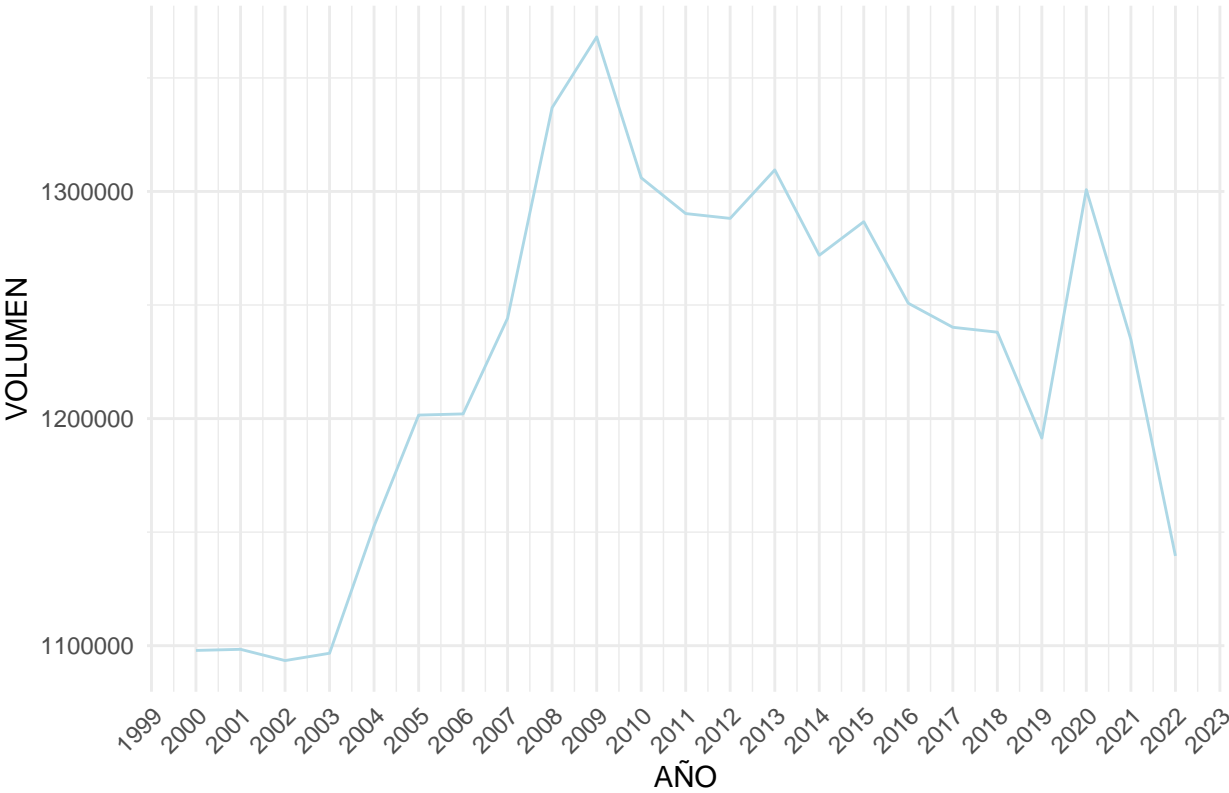
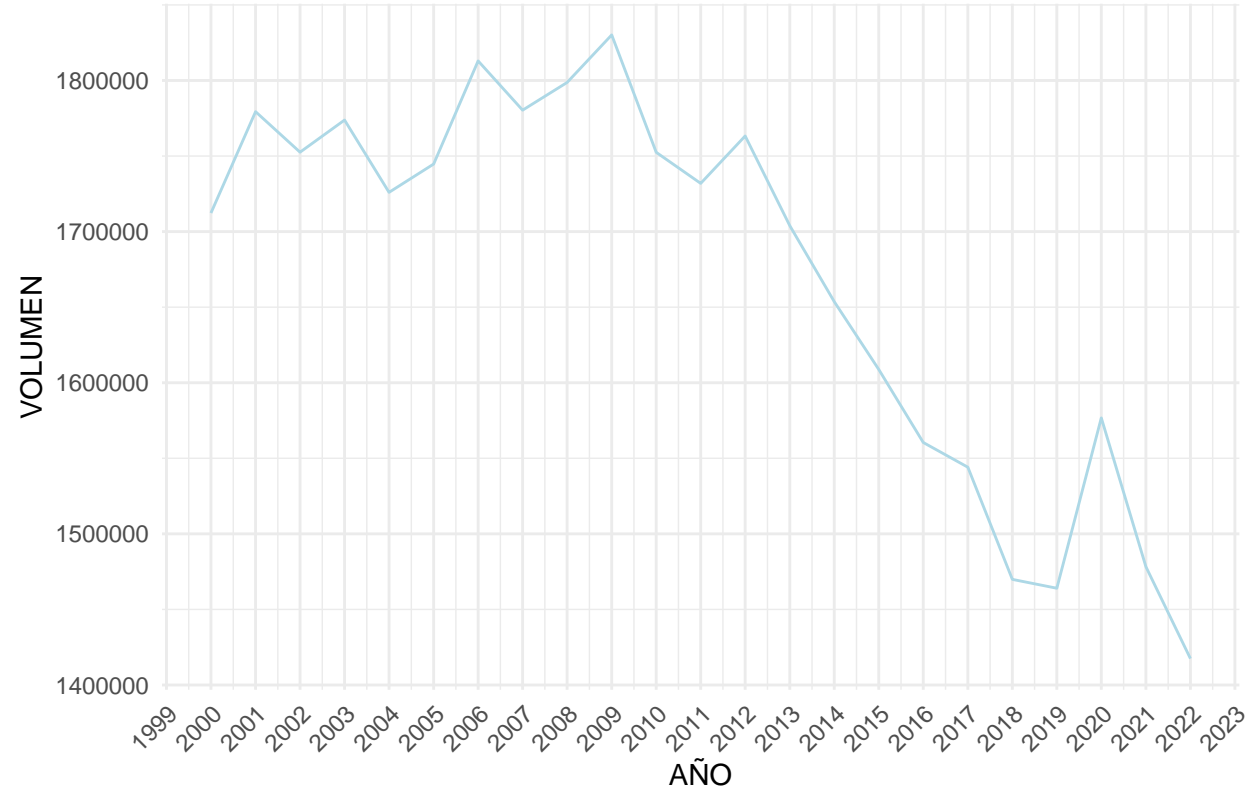


Gráfico de Líneas con el consumo de Castilla Y Leon



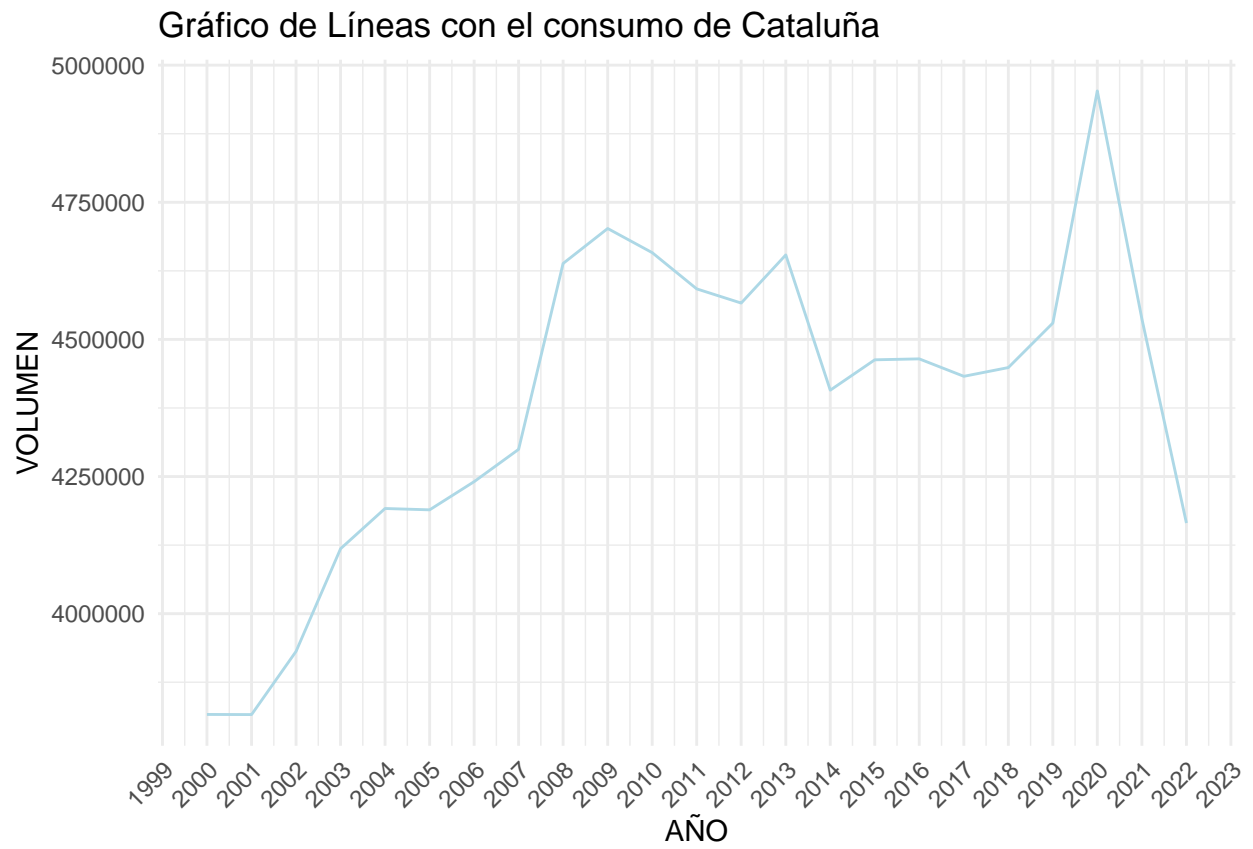


Gráfico de Líneas con el consumo de Extremadura

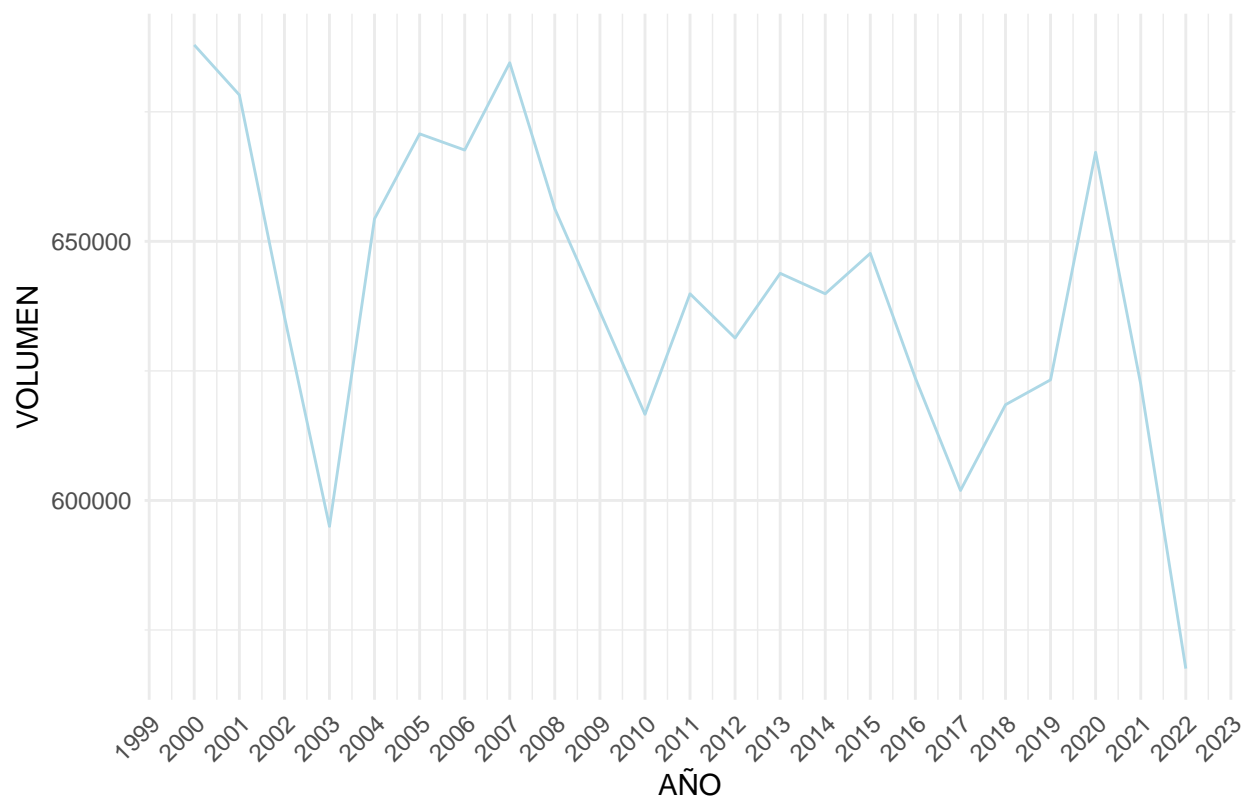


Gráfico de Líneas con el consumo de Galicia

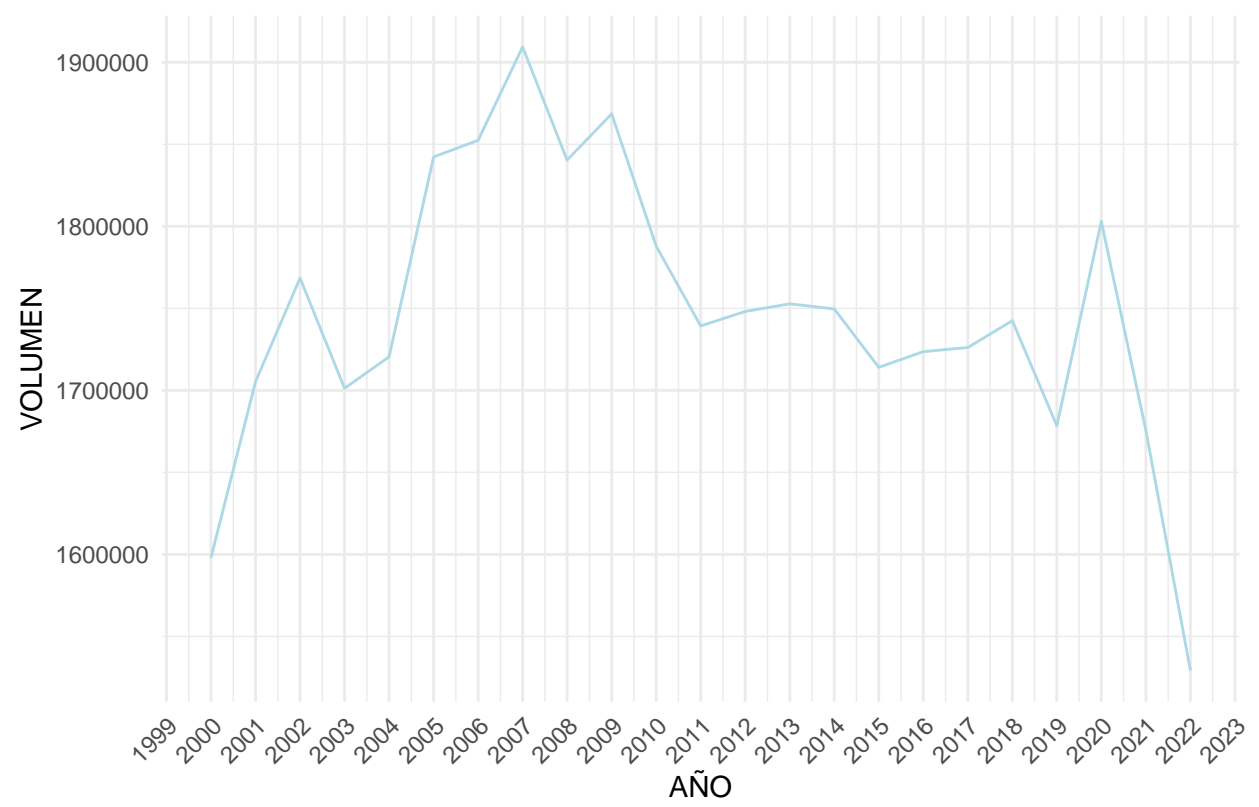


Gráfico de Líneas con el consumo de La Rioja

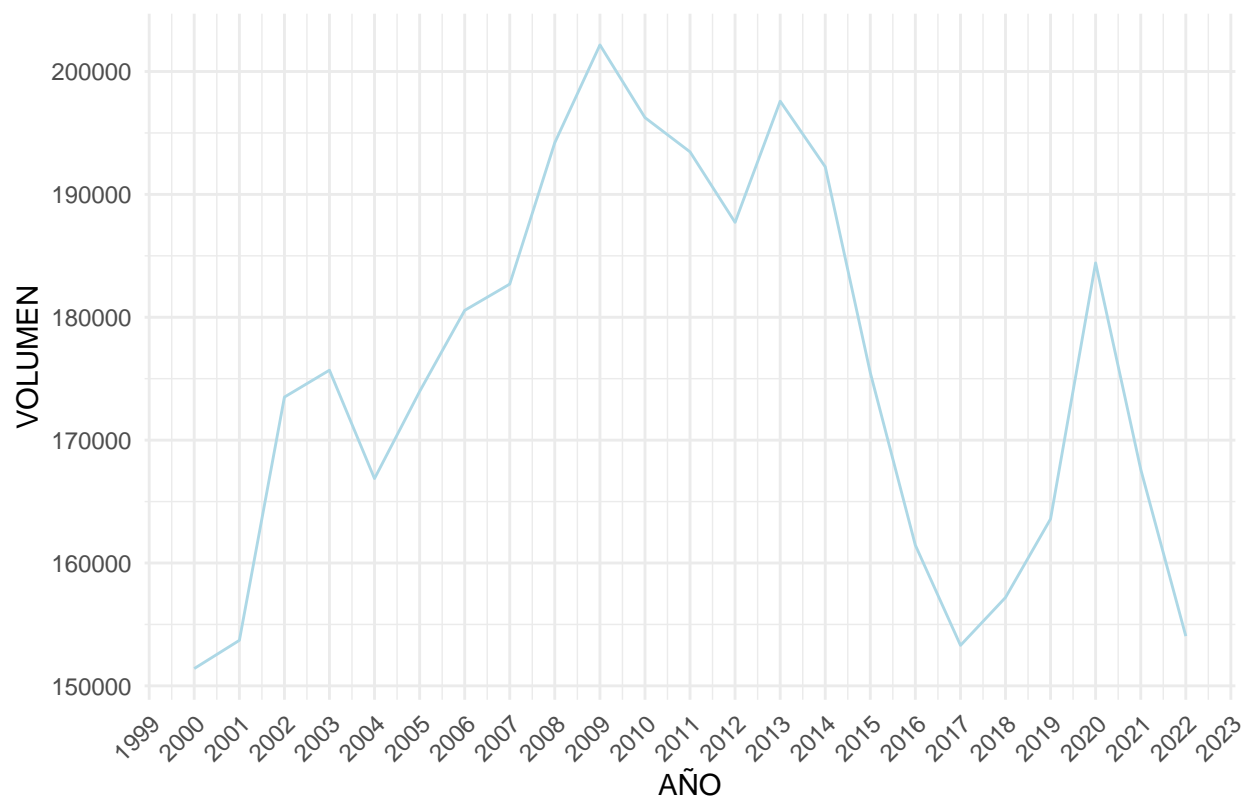


Gráfico de Líneas con el consumo de Madrid

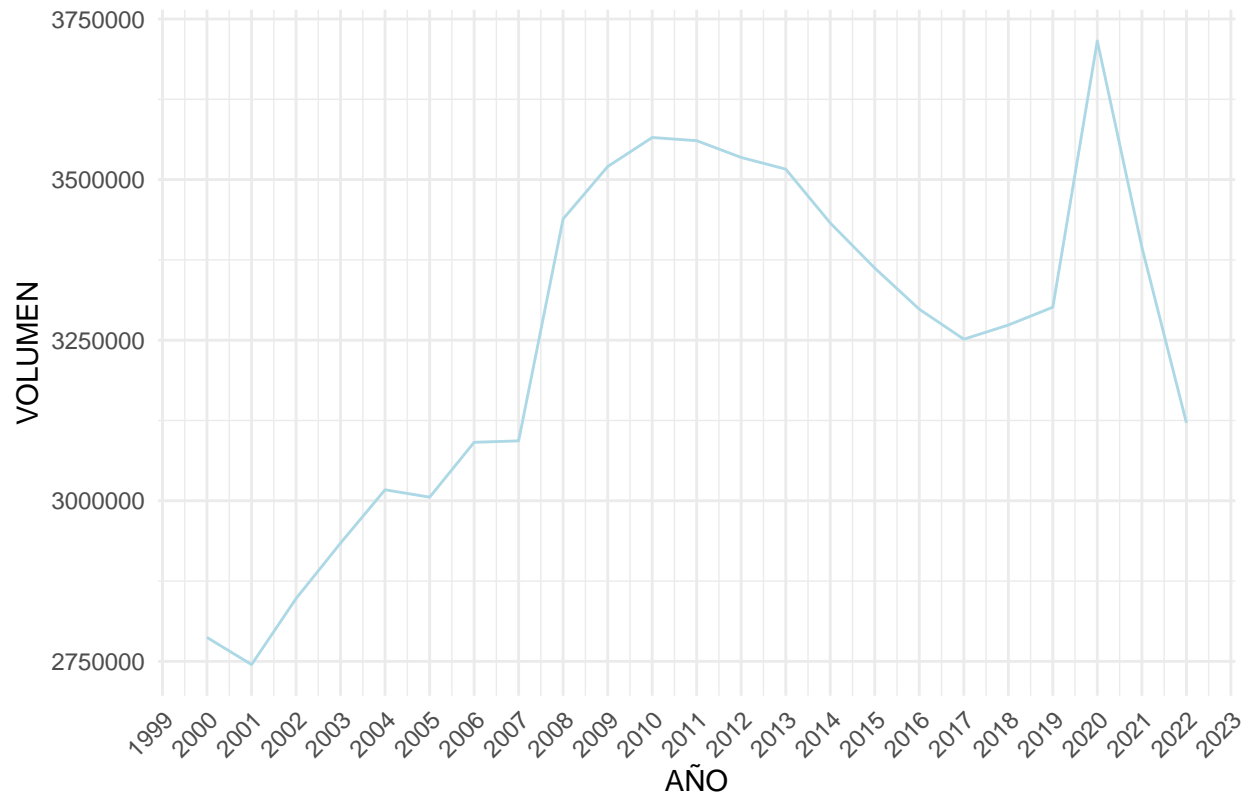
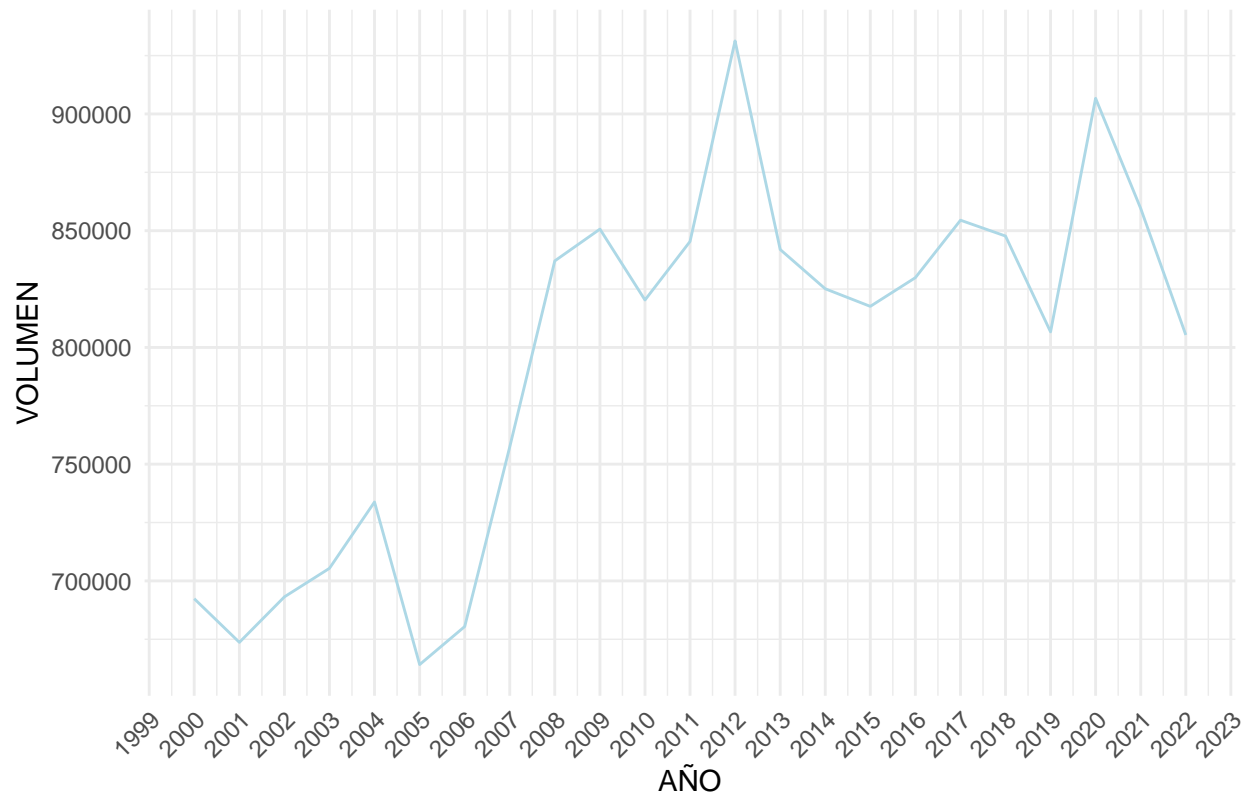


Gráfico de Líneas con el consumo de Murcia



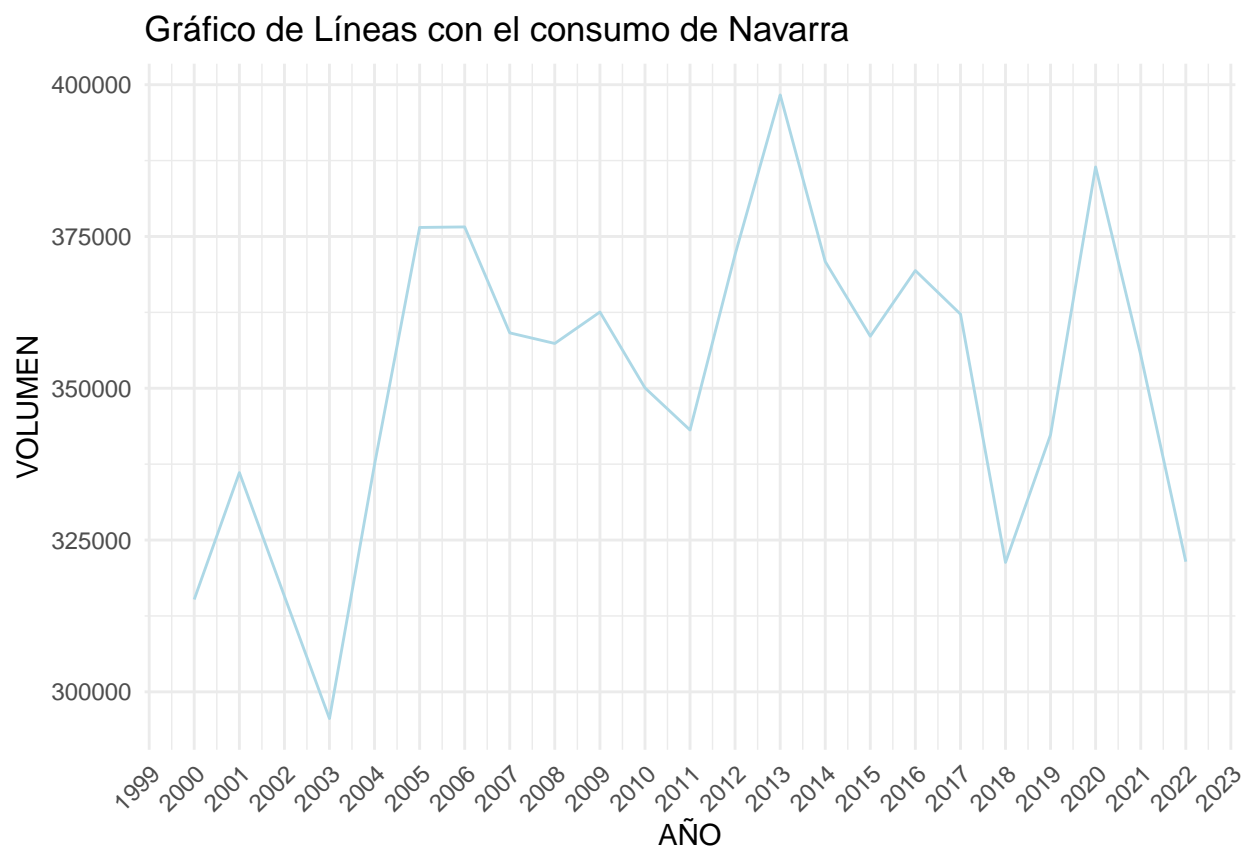


Gráfico de Líneas con el consumo de Pais Vasco

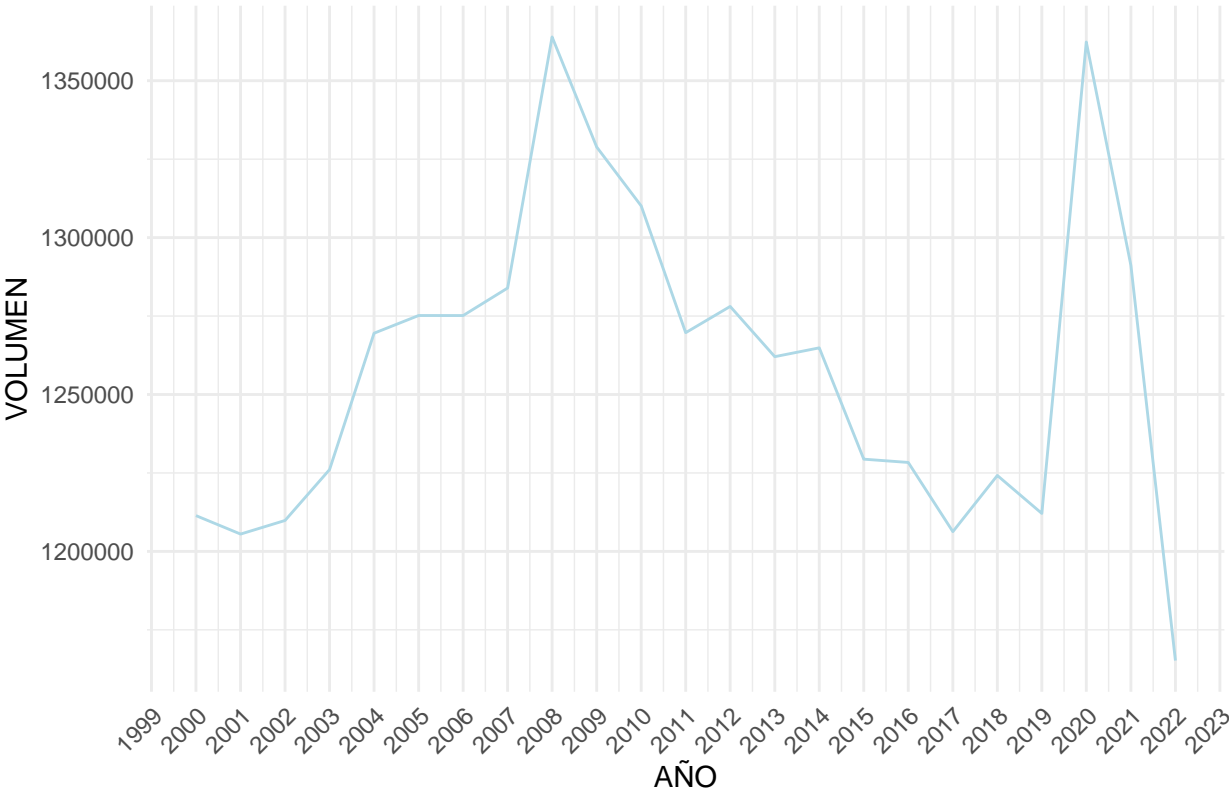
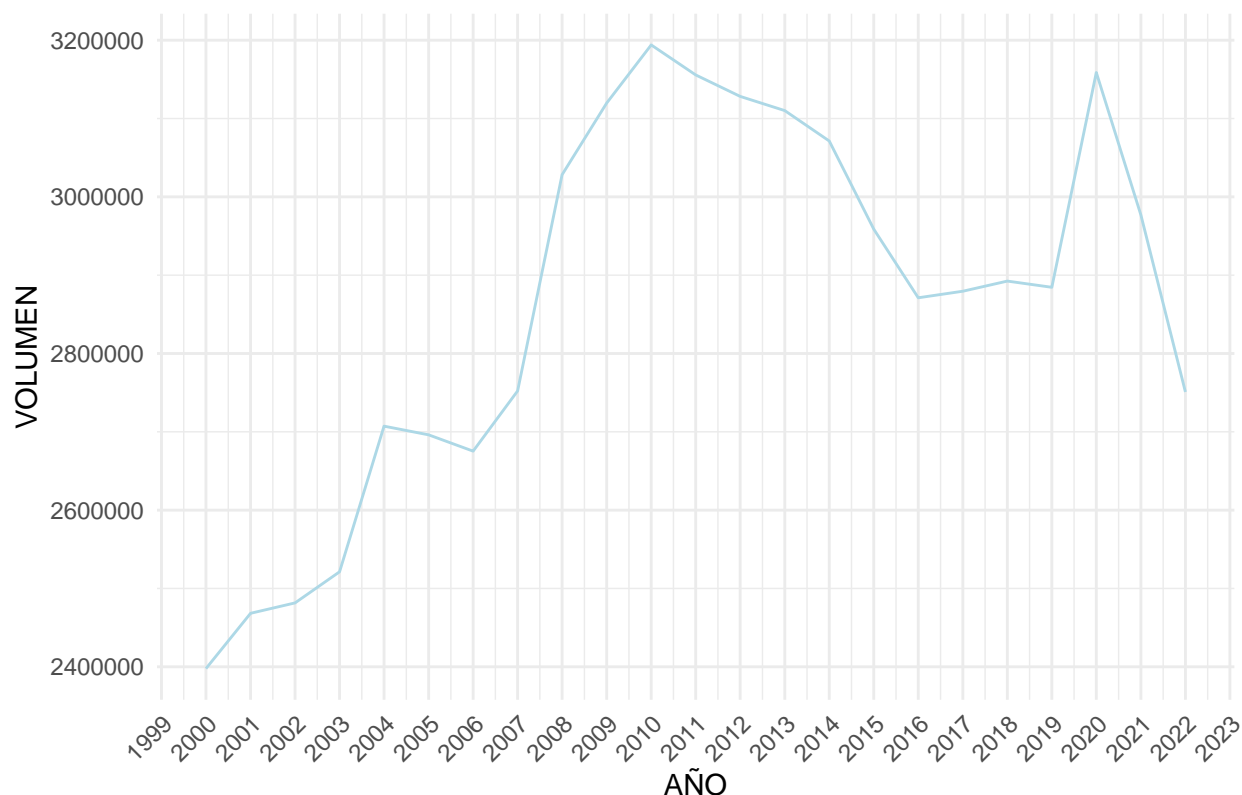


Gráfico de Líneas con el consumo de Valencia



```
rm(data_categorias, regiones_unicas, datos, i, grafico)
```

En términos generales, se observa que en el año 2020 todas las regiones experimentaron un repunte en el consumo de alimentos por parte de los hogares españoles, seguido de una disminución pronunciada en los años 2021 y 2022.

En particular, se pueden destacar las tendencias de consumo de las siguientes regiones:

- Asturias y Castilla y Leon: Inició la serie con un consumo de alimentos elevado y experimentó una disminución casi lineal a lo largo del tiempo.
- Baleares: Comenzó la serie con un consumo bajo y experimentó un crecimiento constante a lo largo de los años.
- Luego, existen un conjunto de regiones que como Cantabria, Castillas La Macha, Valencia, País Vasco, Madrid, La Rioja o Galicia, que muestran un comportamiento similar al general. Empiezan la serie con un consumo bajo o medio. Luego experimentan una subida en el consumo hasta el 2009 para luego caer paulatinamente hasta el 2019. En el 2020 se produce un repunte, y luego, en el 2021 y 2022 cae.

7.4. VOLUMEN VS REGIONES VS CATEGORÍAS

Crea gráficos de barras que muestran el volumen de consumo de diferentes categorías de alimentos por region.

```
# Calculo el resumen de datos por regiones y categorías
datos_regiones_categoria <- data %>%
  group_by(REGIONES, CATEGORIAS) %>%
  summarise(Volumen = sum(VOLUMEN)) %>%
  arrange(desc(Volumen))
```

`summarise()` has grouped output by 'REGIONES'. You can override using the

```
## `.groups` argument.
# Obtengo el nombre de las regiones únicas
regiones_unicas <- unique(datos_regiones_categoria$REGIONES)

# Itero a través de cada región para crear gráficos de barras por región.
for (i in regiones_unicas){

  # Filtro los datos para la región actual
  datos <- datos_regiones_categoria %>%
    filter(REGIONES == i)

  # Creo un gráfico de barras para la región actual
  grafico <- ggplot(
    data = datos,
    aes(
      x = reorder(CATEGORIAS, -Volumen),
      y = Volumen)) +
    geom_bar(
      stat = "identity",
      position = "dodge",
      fill = "lightblue",
      color = "black",
      width = 0.8) +
    labs(
      x = "CATEGORÍAS",
      y = "VOLUMEN",
      title = paste0("Gráfico de Barras VOLUMEN vs CATEGORÍAS en ", i)) +
    theme_minimal() +
    theme(
      axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
      axis.text.y = element_text(size = 8),
      axis.title = element_text(size = 10),
      legend.position = "none") +
    ylim(0, max(datos$Volumen) * 1.1)

  print(grafico)
}
```


Gráfico de Barras VOLUMEN vs CATEGORÍAS en Cataluña

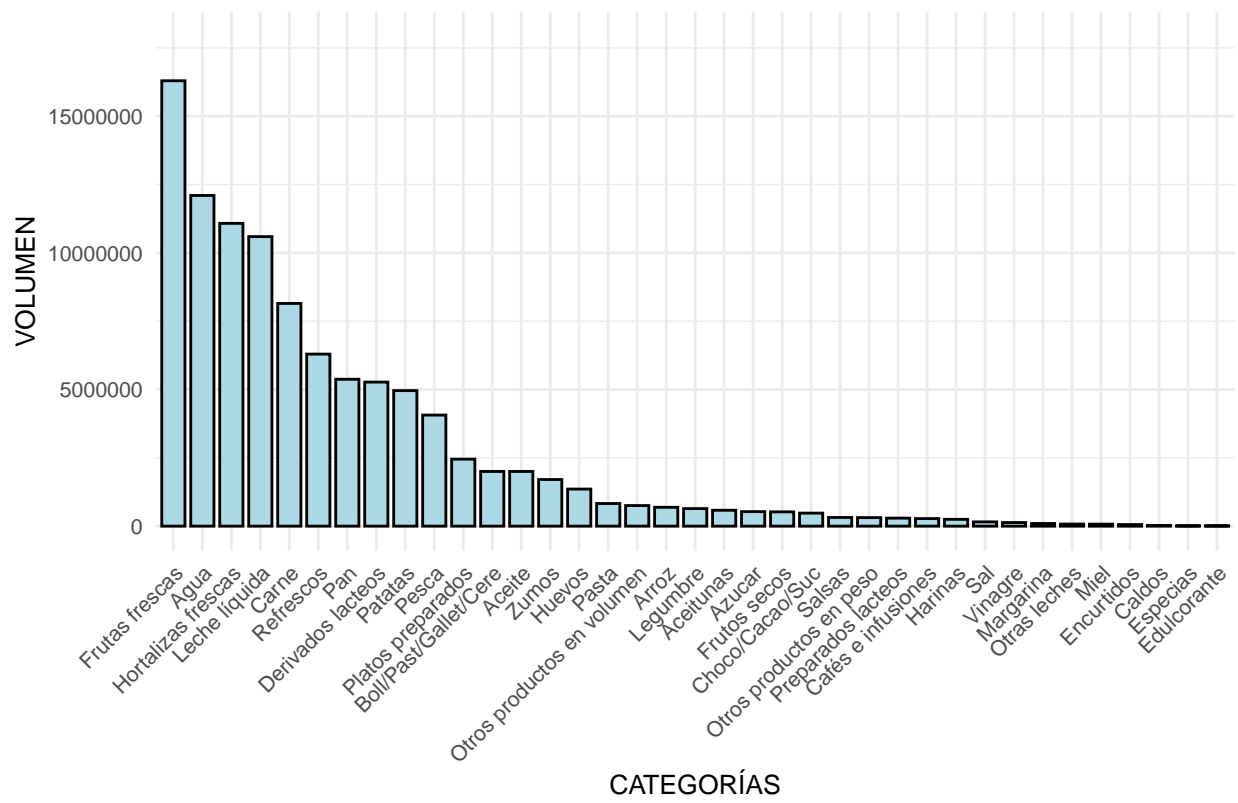


Gráfico de Barras VOLUMEN vs CATEGORÍAS en Andalucía

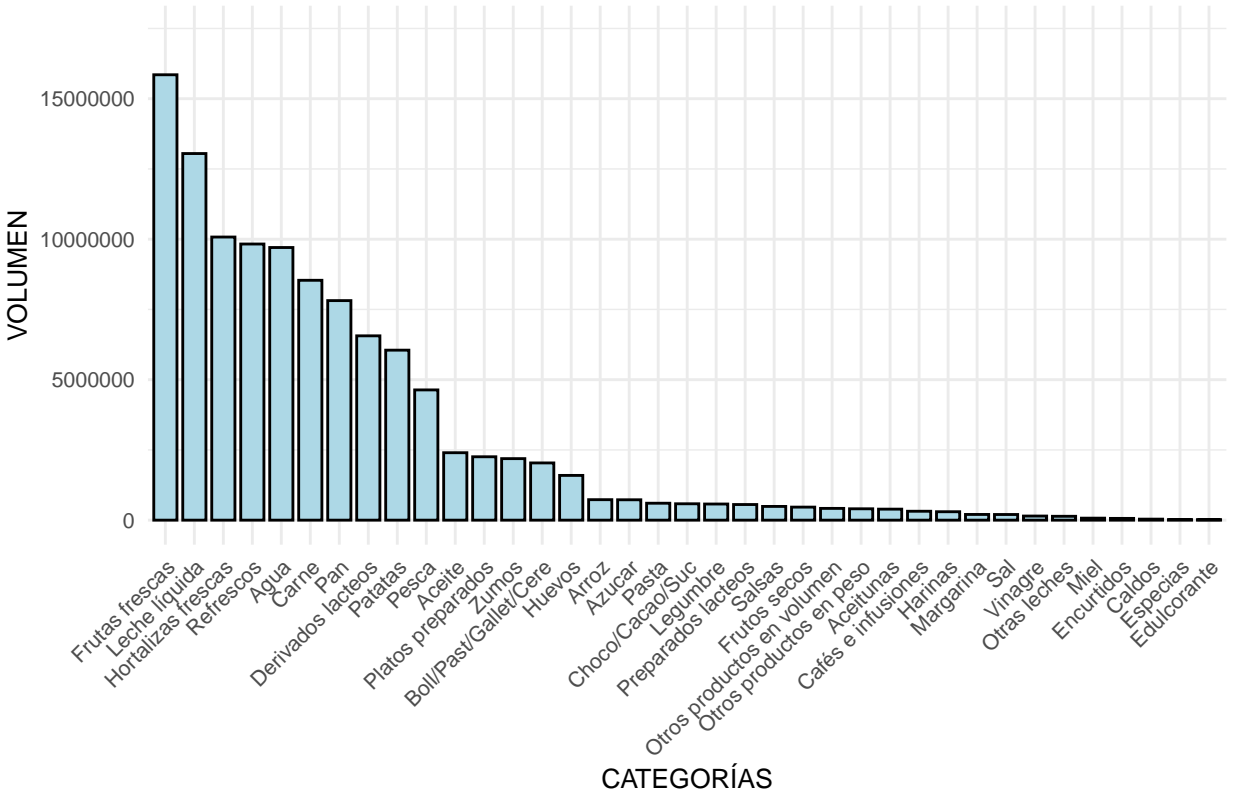


Gráfico de Barras VOLUMEN vs CATEGORÍAS en Madrid

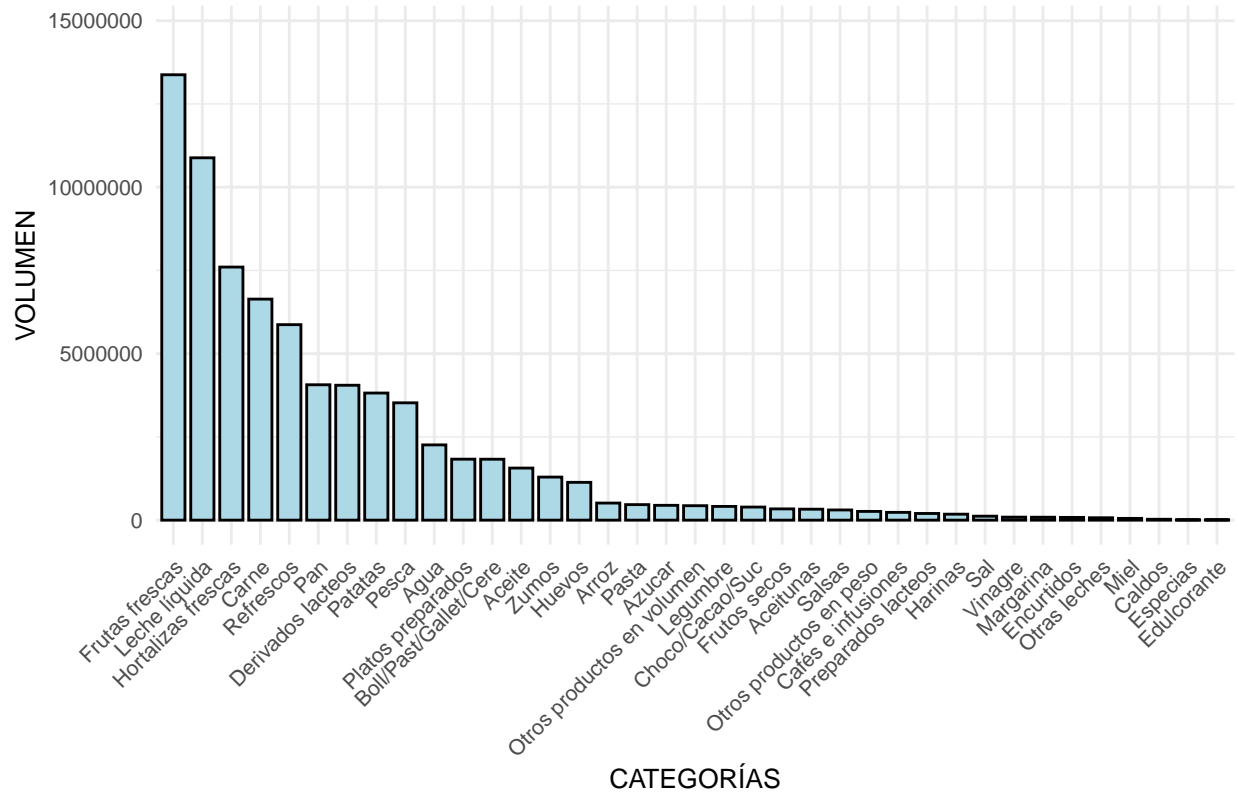


Gráfico de Barras VOLUMEN vs CATEGORÍAS en Valencia

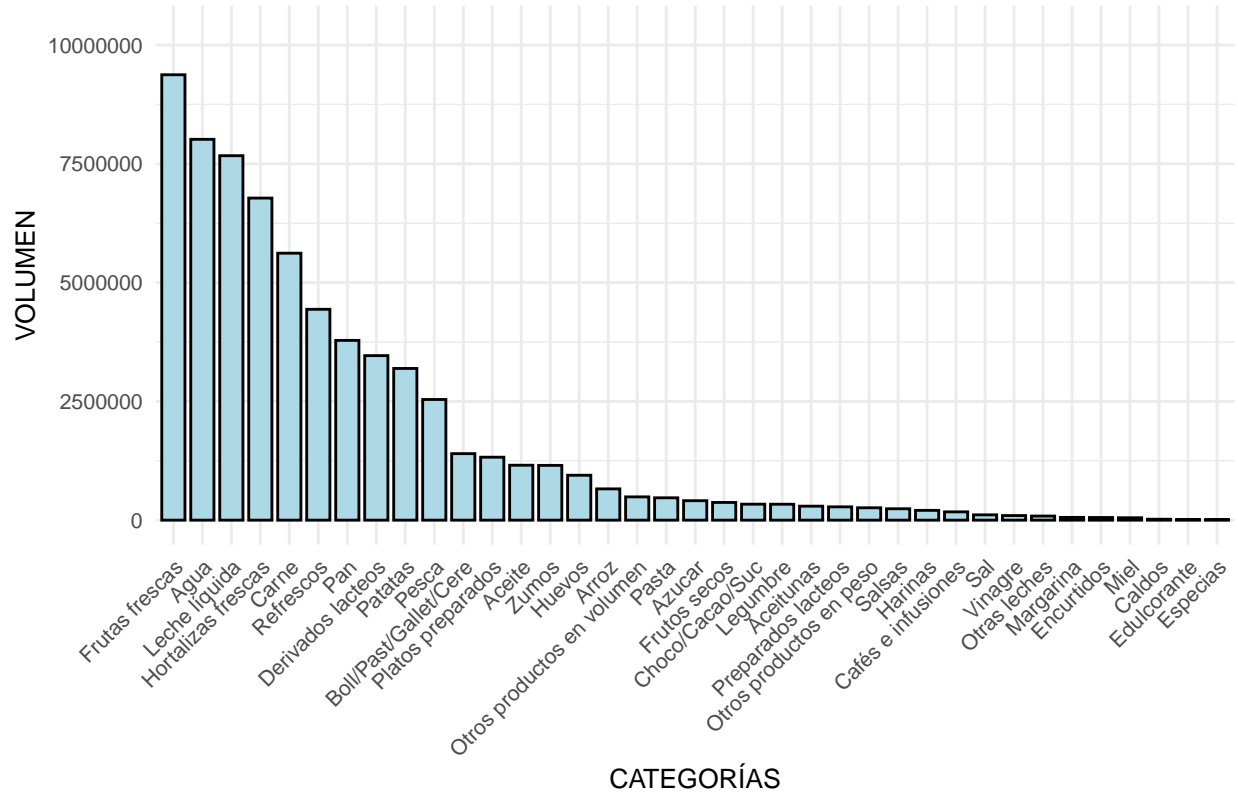


Gráfico de Barras VOLUMEN vs CATEGORÍAS en Castilla Y Leon

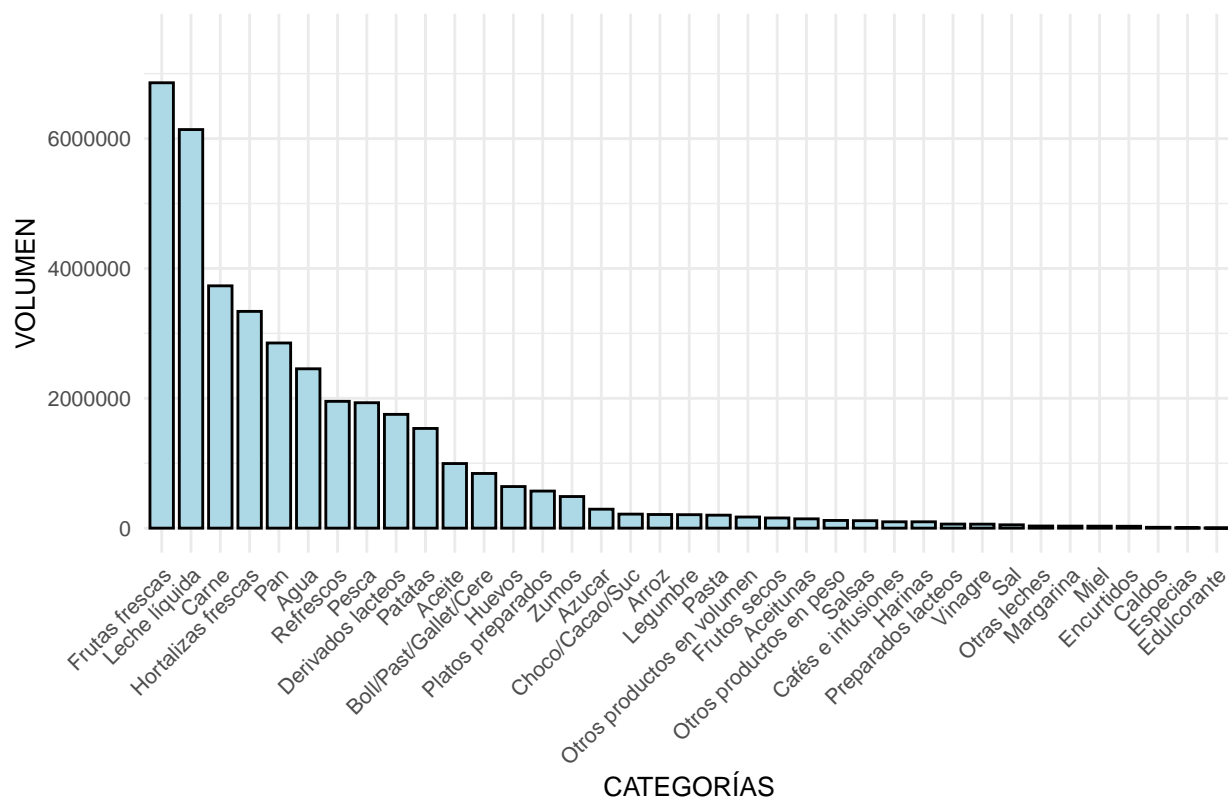
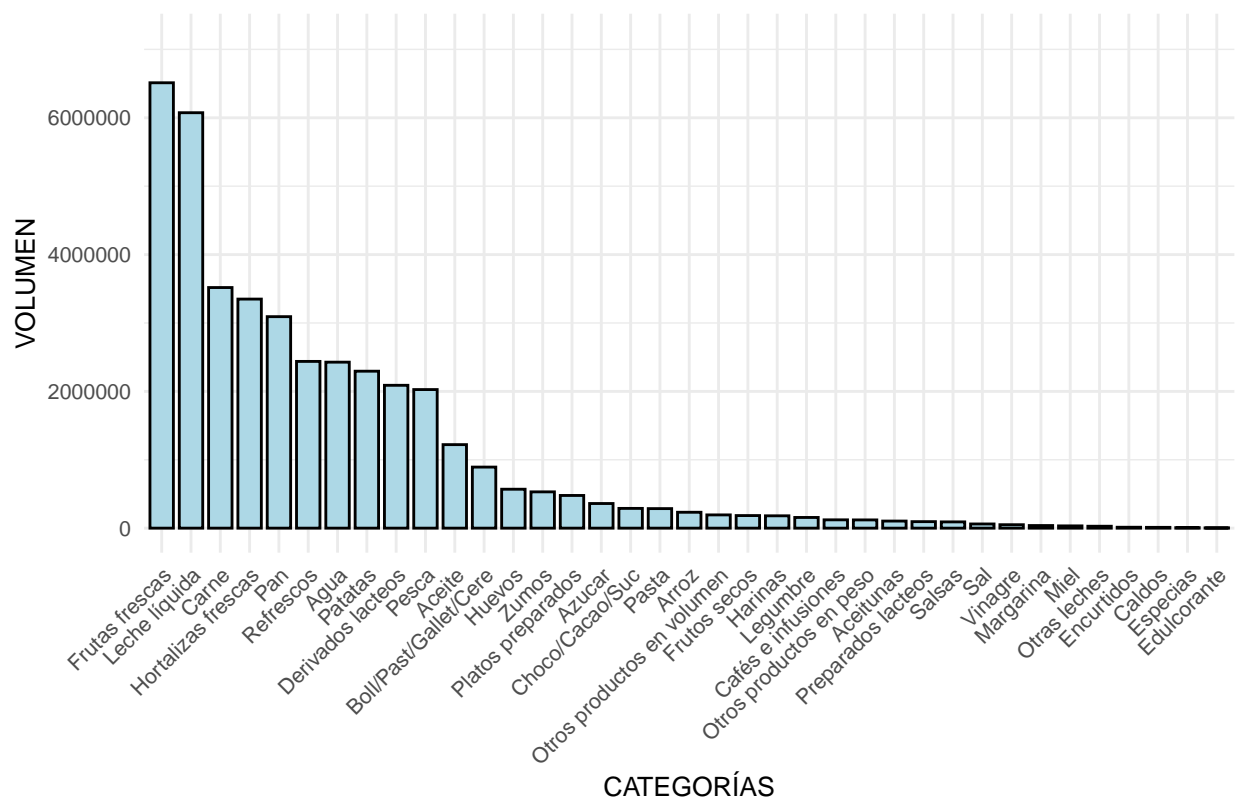
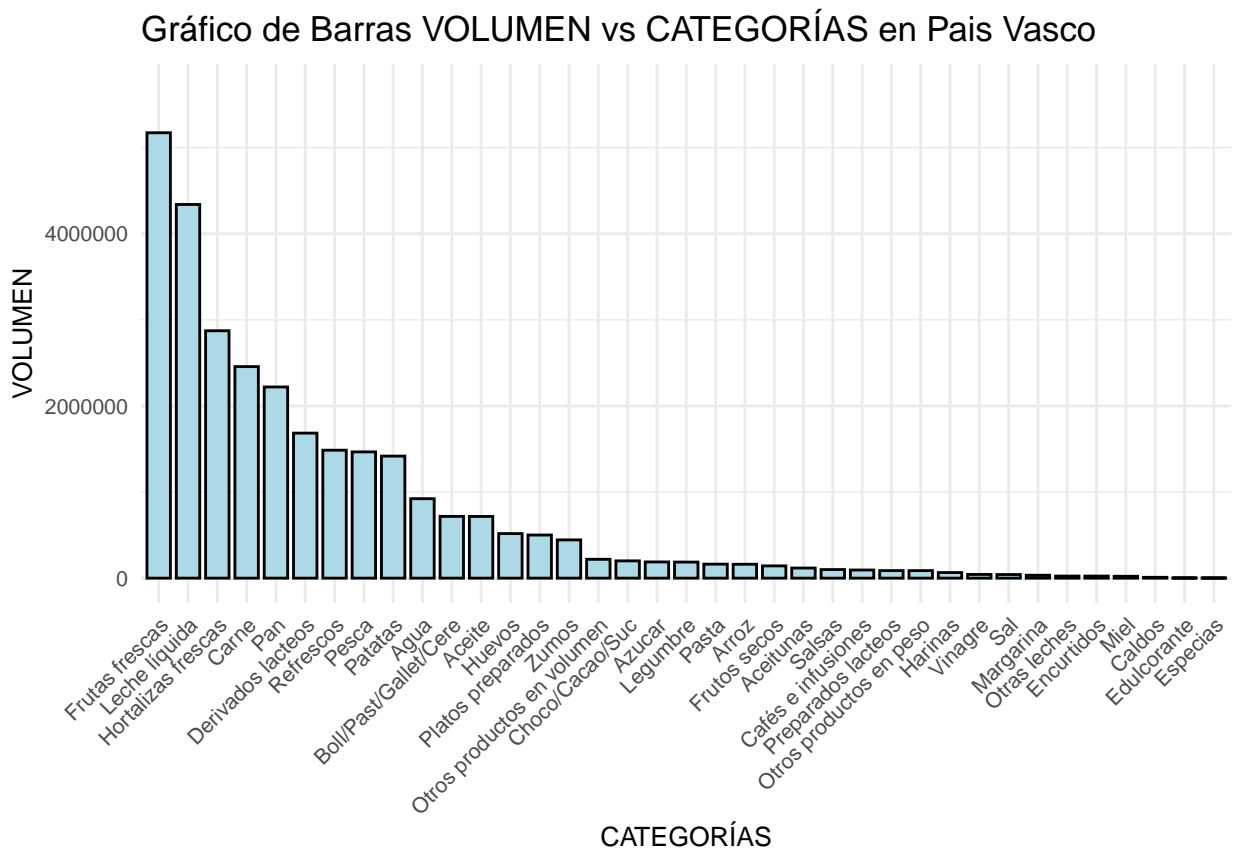
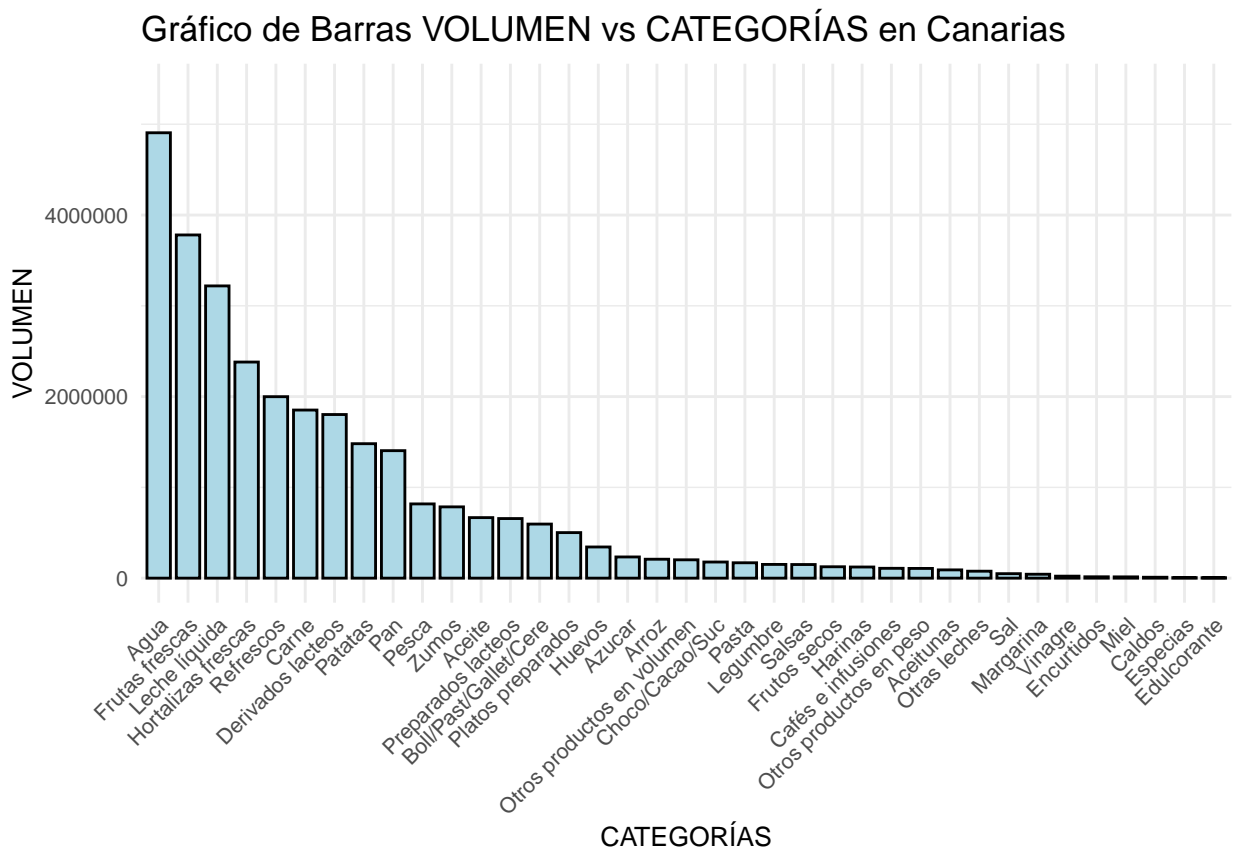
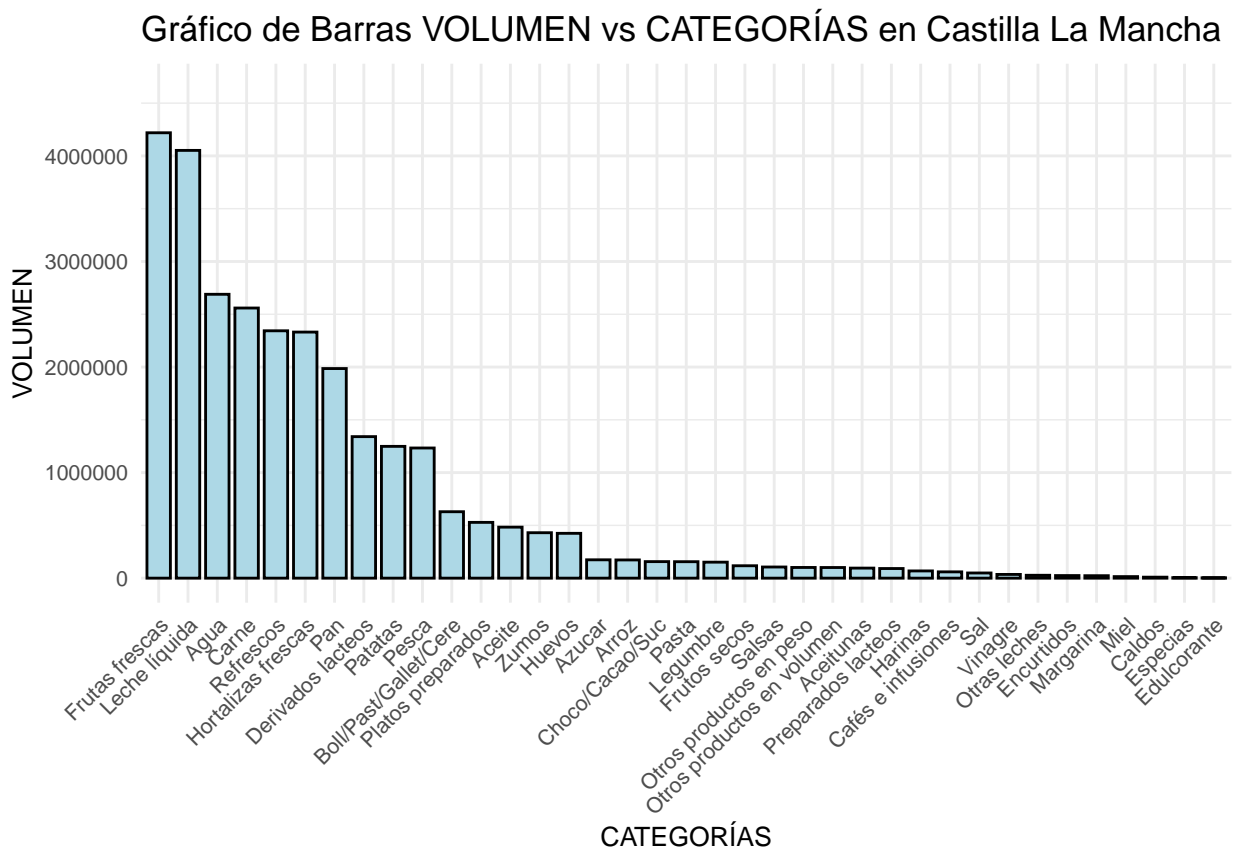


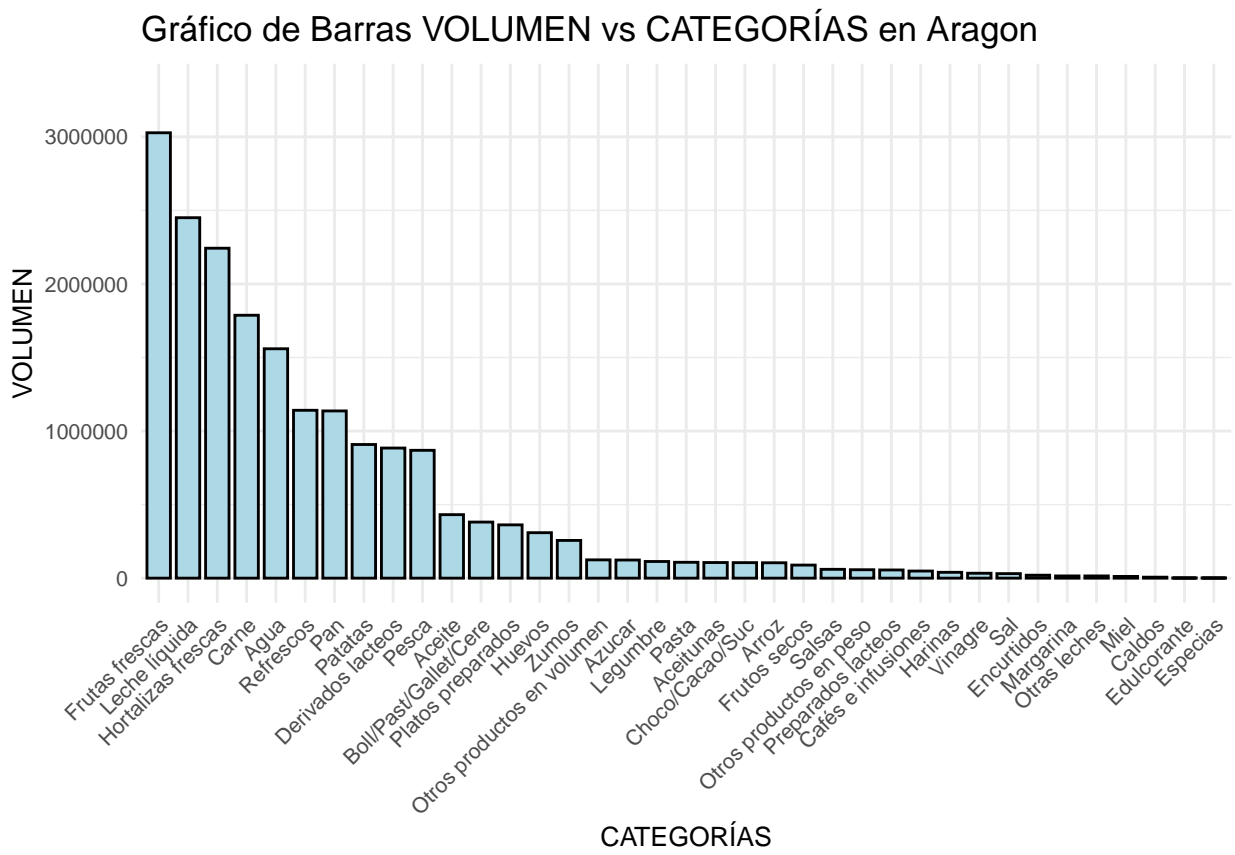
Gráfico de Barras VOLUMEN vs CATEGORÍAS en Galicia











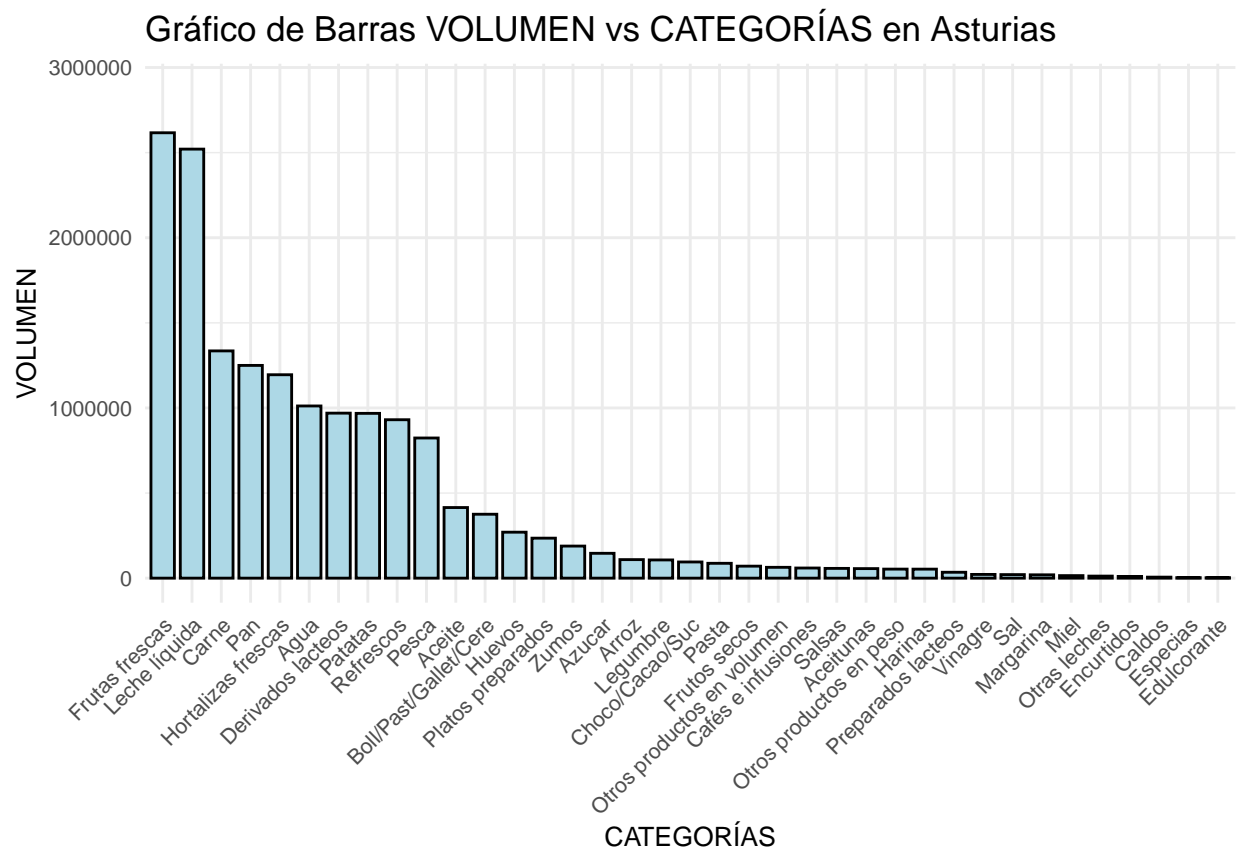


Gráfico de Barras VOLUMEN vs CATEGORÍAS en Murcia

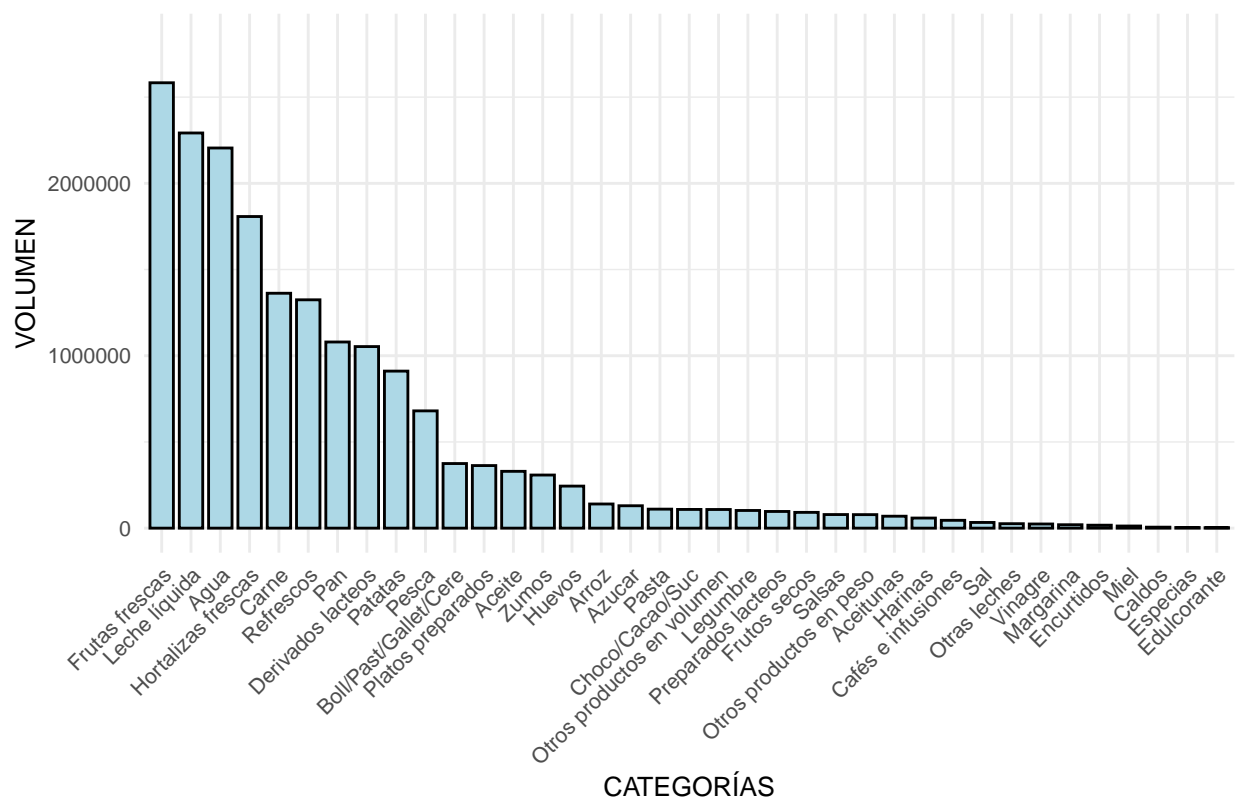
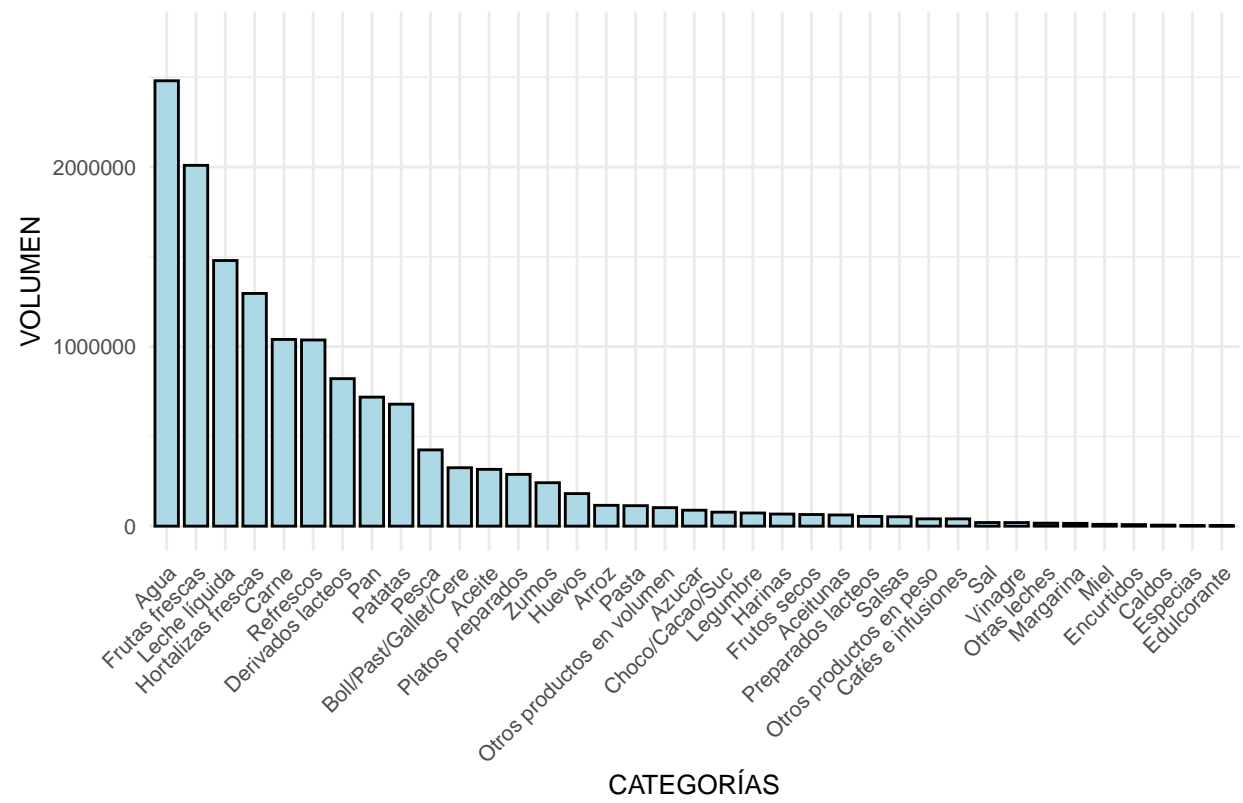
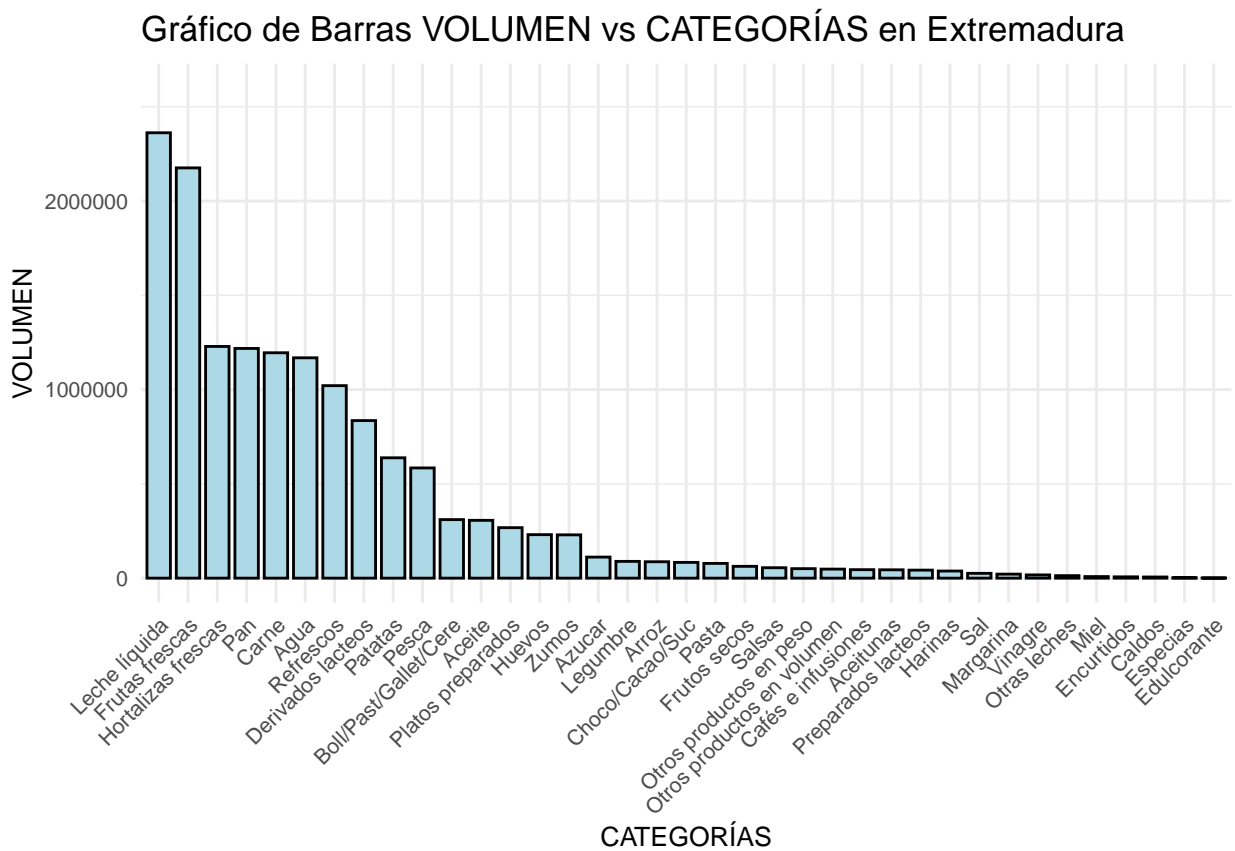


Gráfico de Barras VOLUMEN vs CATEGORÍAS en Baleares





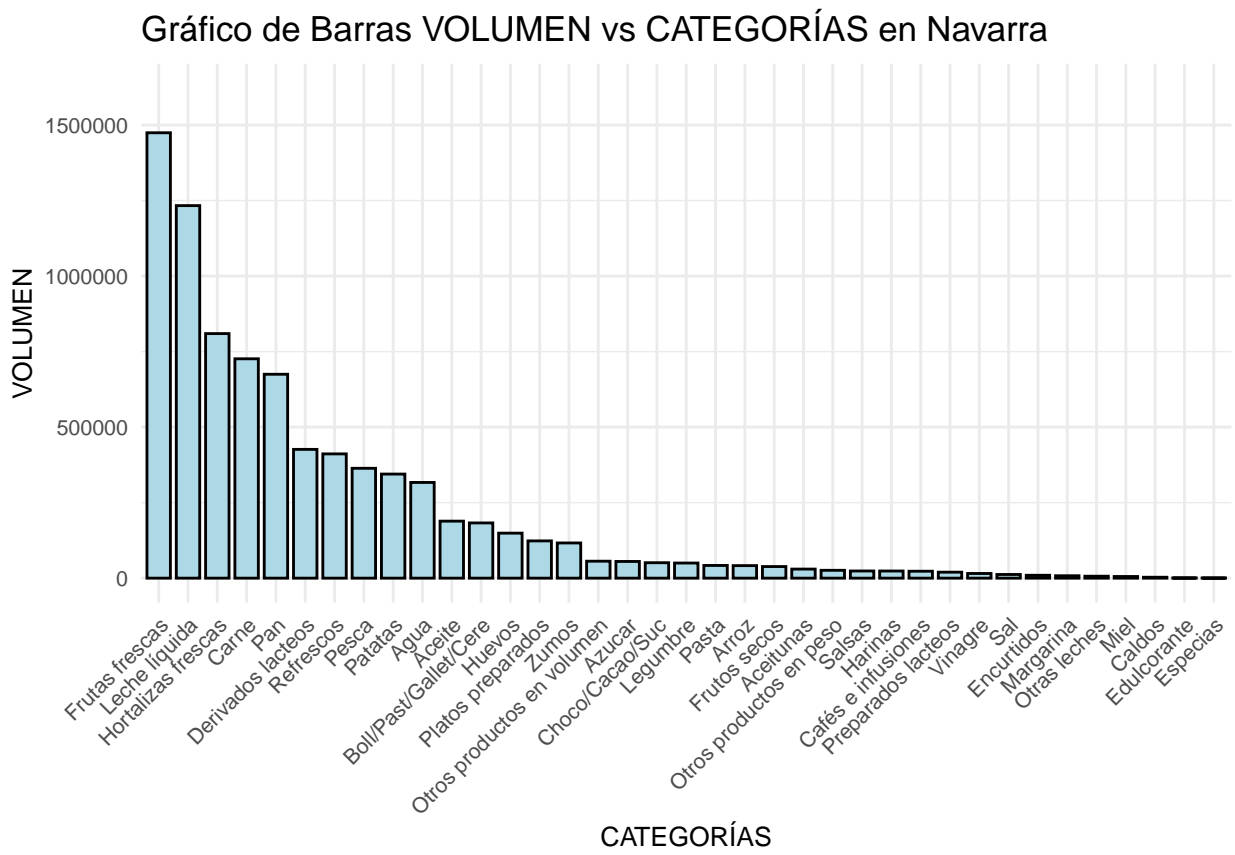


Gráfico de Barras VOLUMEN vs CATEGORÍAS en Cantabria

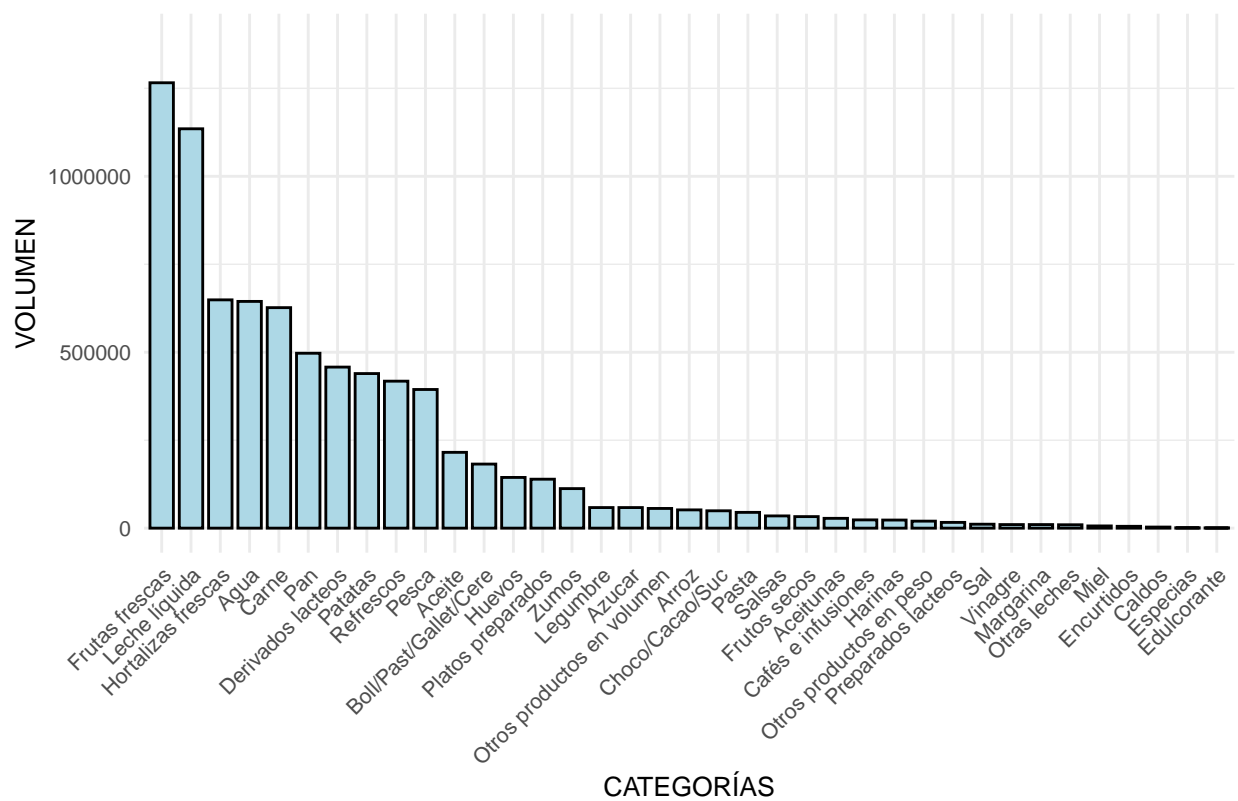
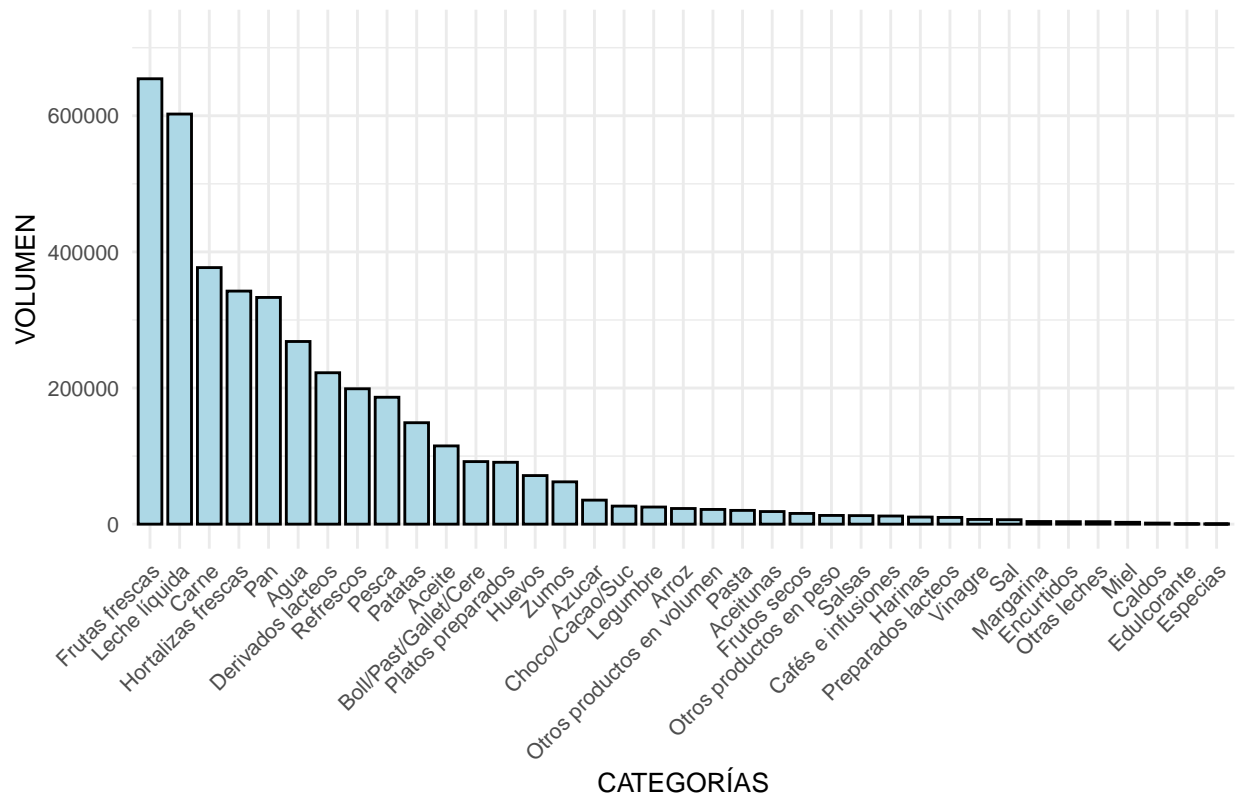


Gráfico de Barras VOLUMEN vs CATEGORÍAS en La Rioja



```
# Elimino las variables que no utilizaré
rm(datos_regiones_categoria, regiones_unicas, datos, i, grafico)
```

A continuación realizo etiqueta a las categorías basado en cuartiles para diferentes regiones.

```
# Calculo el resumen de datos por regiones y categorías
datos_regiones_categoria <- data %>%
  group_by(REGIONES, CATEGORIAS) %>%
  summarise(Volumen = sum(VOLUMEN)) %>%
  arrange(desc(Volumen))
```

```
## `summarise()` has grouped output by 'REGIONES'. You can override using the
## `.groups` argument.
```

```
# Obtengo el nombre de las regiones
regiones_unicas <- unique(datos_regiones_categoria$REGIONES)

# Itero a través de cada región para realizar análisis por región.
for(i in regiones_unicas){

  # Filtrar los datos para la región actual
  datos_region <- datos_regiones_categoria %>%
    filter(REGIONES == i)

  # Calculo las estadísticas necesarias para reproducir el gráfico
  estadisticas <- boxplot.stats(datos_region$Volumen)

  # Etiqueto las categorías en función de los cuartiles
```

```

datos_region <- datos_region %>%
  mutate(
    Estadisticas = case_when(
      Volumen <= estadisticas$stats[2] ~ "Primer cuartil",
      Volumen > estadisticas$stats[2] & Volumen <= estadisticas$stats[3] ~ "Segundo cuartil",
      Volumen > estadisticas$stats[3] & Volumen <= estadisticas$stats[4] ~ "Tercer cuartil",
      Volumen > estadisticas$stats[4] & Volumen < min(estadisticas$out) ~ "Cuarto cuartil",
      Volumen %in% estadisticas$out ~ "Valores atípicos",
      TRUE ~ "Otros"))

# Imprimir la tabla con etiquetas
print(datos_region)
}

```

```

## # A tibble: 37 x 4
## # Groups:   REGIONES [1]
##   REGIONES CATEGORIAS      Volumen Estadisticas
##   <fct>    <fct>          <dbl> <chr>
## 1 Cataluña Frutas frescas 16296364. Valores atípicos
## 2 Cataluña Agua          12100158. Valores atípicos
## 3 Cataluña Hortalizas frescas 11078650. Valores atípicos
## 4 Cataluña Leche líquida 10593599. Valores atípicos
## 5 Cataluña Carne          8149313. Cuarto cuartil
## 6 Cataluña Refrescos      6295062. Cuarto cuartil
## 7 Cataluña Pan            5374720. Cuarto cuartil
## 8 Cataluña Derivados lacteos 5271553. Cuarto cuartil
## 9 Cataluña Patatas        4959041. Cuarto cuartil
## 10 Cataluña Pesca         4063405. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups:   REGIONES [1]
##   REGIONES CATEGORIAS      Volumen Estadisticas
##   <fct>    <fct>          <dbl> <chr>
## 1 Andalucía Frutas frescas 15852624. Valores atípicos
## 2 Andalucía Leche líquida 13049052. Valores atípicos
## 3 Andalucía Hortalizas frescas 10077219. Cuarto cuartil
## 4 Andalucía Refrescos      9828534. Cuarto cuartil
## 5 Andalucía Agua          9704390. Cuarto cuartil
## 6 Andalucía Carne          8537014. Cuarto cuartil
## 7 Andalucía Pan            7814616. Cuarto cuartil
## 8 Andalucía Derivados lacteos 6560188. Cuarto cuartil
## 9 Andalucía Patatas        6050805. Cuarto cuartil
## 10 Andalucía Pesca         4635702. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups:   REGIONES [1]
##   REGIONES CATEGORIAS      Volumen Estadisticas
##   <fct>    <fct>          <dbl> <chr>
## 1 Madrid   Frutas frescas 13375484. Valores atípicos
## 2 Madrid   Leche líquida 10881693. Valores atípicos
## 3 Madrid   Hortalizas frescas 7600312. Valores atípicos
## 4 Madrid   Carne          6637682. Valores atípicos
## 5 Madrid   Refrescos      5871086. Valores atípicos
## 6 Madrid   Pan            4065865. Cuarto cuartil

```

```

## 7 Madrid Derivados lacteos 4051328. Cuarto cuartil
## 8 Madrid Patatas 3817464. Cuarto cuartil
## 9 Madrid Pesca 3523501. Cuarto cuartil
## 10 Madrid Agua 2260287. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 Valencia Frutas frescas 9373928. Valores atípicos
## 2 Valencia Agua 8014900. Valores atípicos
## 3 Valencia Leche líquida 7670503. Valores atípicos
## 4 Valencia Hortalizas frescas 6778725. Valores atípicos
## 5 Valencia Carne 5618348. Cuarto cuartil
## 6 Valencia Refrescos 4437629. Cuarto cuartil
## 7 Valencia Pan 3783103. Cuarto cuartil
## 8 Valencia Derivados lacteos 3462723. Cuarto cuartil
## 9 Valencia Patatas 3192742. Cuarto cuartil
## 10 Valencia Pesca 2539459. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 Castilla Y Leon Frutas frescas 6859577. Valores atípicos
## 2 Castilla Y Leon Leche líquida 6138611. Valores atípicos
## 3 Castilla Y Leon Carne 3731919. Cuarto cuartil
## 4 Castilla Y Leon Hortalizas frescas 3338767. Cuarto cuartil
## 5 Castilla Y Leon Pan 2851812. Cuarto cuartil
## 6 Castilla Y Leon Agua 2454357. Cuarto cuartil
## 7 Castilla Y Leon Refrescos 1953203. Cuarto cuartil
## 8 Castilla Y Leon Pesca 1932130. Cuarto cuartil
## 9 Castilla Y Leon Derivados lacteos 1752364. Cuarto cuartil
## 10 Castilla Y Leon Patatas 1535778. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 Galicia Frutas frescas 6511886. Valores atípicos
## 2 Galicia Leche líquida 6073929. Valores atípicos
## 3 Galicia Carne 3517936. Cuarto cuartil
## 4 Galicia Hortalizas frescas 3349037. Cuarto cuartil
## 5 Galicia Pan 3092213. Cuarto cuartil
## 6 Galicia Refrescos 2438064. Cuarto cuartil
## 7 Galicia Agua 2427168. Cuarto cuartil
## 8 Galicia Patatas 2295026. Cuarto cuartil
## 9 Galicia Derivados lacteos 2088776. Cuarto cuartil
## 10 Galicia Pesca 2026331. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>

```

```
## 1 Pais Vasco Frutas frescas      5172164. Valores atípicos
## 2 Pais Vasco Leche líquida      4339410. Valores atípicos
## 3 Pais Vasco Hortalizas frescas 2873147. Valores atípicos
## 4 Pais Vasco Carne              2456518. Valores atípicos
## 5 Pais Vasco Pan                2220036. Valores atípicos
## 6 Pais Vasco Derivados lacteos 1684419. Cuarto cuartil
## 7 Pais Vasco Refrescos          1485894. Cuarto cuartil
## 8 Pais Vasco Pesca              1466148. Cuarto cuartil
## 9 Pais Vasco Patatas            1417223. Cuarto cuartil
## 10 Pais Vasco Agua              922918. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups:   REGIONES [1]
##   REGIONES CATEGORIAS      Volumen Estadísticas
##   <fct>    <fct>          <dbl> <chr>
## 1 Canarias Agua          4906002. Valores atípicos
## 2 Canarias Frutas frescas 3780397. Valores atípicos
## 3 Canarias Leche líquida 3219052. Valores atípicos
## 4 Canarias Hortalizas frescas 2380436. Valores atípicos
## 5 Canarias Refrescos     1998955. Valores atípicos
## 6 Canarias Carne         1851764. Cuarto cuartil
## 7 Canarias Derivados lacteos 1802339. Cuarto cuartil
## 8 Canarias Patatas       1480996. Cuarto cuartil
## 9 Canarias Pan           1404403. Cuarto cuartil
## 10 Canarias Pesca        817727. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups:   REGIONES [1]
##   REGIONES CATEGORIAS      Volumen Estadísticas
##   <fct>    <fct>          <dbl> <chr>
## 1 Castilla La Mancha Frutas frescas 4219229. Valores atípicos
## 2 Castilla La Mancha Leche líquida 4052274. Valores atípicos
## 3 Castilla La Mancha Agua          2689064. Cuarto cuartil
## 4 Castilla La Mancha Carne         2558898. Cuarto cuartil
## 5 Castilla La Mancha Refrescos     2343332. Cuarto cuartil
## 6 Castilla La Mancha Hortalizas frescas 2331253. Cuarto cuartil
## 7 Castilla La Mancha Pan           1985646. Cuarto cuartil
## 8 Castilla La Mancha Derivados lacteos 1340598. Cuarto cuartil
## 9 Castilla La Mancha Patatas       1248514. Cuarto cuartil
## 10 Castilla La Mancha Pesca        1232938. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups:   REGIONES [1]
##   REGIONES CATEGORIAS      Volumen Estadísticas
##   <fct>    <fct>          <dbl> <chr>
## 1 Aragon Frutas frescas      3027479. Valores atípicos
## 2 Aragon Leche líquida      2450122. Valores atípicos
## 3 Aragon Hortalizas frescas 2242785. Valores atípicos
## 4 Aragon Carne              1787007. Cuarto cuartil
## 5 Aragon Agua              1559142. Cuarto cuartil
## 6 Aragon Refrescos         1140988. Cuarto cuartil
## 7 Aragon Pan               1136671. Cuarto cuartil
## 8 Aragon Patatas           908410. Cuarto cuartil
## 9 Aragon Derivados lacteos 884453. Cuarto cuartil
```

```

## 10 Aragon Pesca 869058. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 Asturias Frutas frescas 2616598. Valores atípicos
## 2 Asturias Leche líquida 2520468. Valores atípicos
## 3 Asturias Carne 1335011. Cuarto cuartil
## 4 Asturias Pan 1249902. Cuarto cuartil
## 5 Asturias Hortalizas frescas 1195055. Cuarto cuartil
## 6 Asturias Agua 1011641. Cuarto cuartil
## 7 Asturias Derivados lacteos 969431. Cuarto cuartil
## 8 Asturias Patatas 968197. Cuarto cuartil
## 9 Asturias Refrescos 930740. Cuarto cuartil
## 10 Asturias Pesca 823472. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 Murcia Frutas frescas 2583134. Valores atípicos
## 2 Murcia Leche líquida 2291927. Valores atípicos
## 3 Murcia Agua 2204789. Valores atípicos
## 4 Murcia Hortalizas frescas 1807329. Valores atípicos
## 5 Murcia Carne 1362384. Cuarto cuartil
## 6 Murcia Refrescos 1324335. Cuarto cuartil
## 7 Murcia Pan 1079651. Cuarto cuartil
## 8 Murcia Derivados lacteos 1053053. Cuarto cuartil
## 9 Murcia Patatas 910754. Cuarto cuartil
## 10 Murcia Pesca 680262. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 Baleares Agua 2480369. Valores atípicos
## 2 Baleares Frutas frescas 2009332. Valores atípicos
## 3 Baleares Leche líquida 1479246. Valores atípicos
## 4 Baleares Hortalizas frescas 1296025. Valores atípicos
## 5 Baleares Carne 1039782. Valores atípicos
## 6 Baleares Refrescos 1036872. Valores atípicos
## 7 Baleares Derivados lacteos 821378. Cuarto cuartil
## 8 Baleares Pan 718570. Cuarto cuartil
## 9 Baleares Patatas 679220. Cuarto cuartil
## 10 Baleares Pesca 424689. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 Extremadura Leche líquida 2361883. Valores atípicos
## 2 Extremadura Frutas frescas 2175678. Valores atípicos
## 3 Extremadura Hortalizas frescas 1228685. Cuarto cuartil

```

```
## 4 Extremadura Pan 1218083. Cuarto cuartil
## 5 Extremadura Carne 1195223. Cuarto cuartil
## 6 Extremadura Agua 1168505. Cuarto cuartil
## 7 Extremadura Refrescos 1020838. Cuarto cuartil
## 8 Extremadura Derivados lacteos 835658. Cuarto cuartil
## 9 Extremadura Patatas 638234. Cuarto cuartil
## 10 Extremadura Pesca 584524. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 Navarra Frutas frescas 1474425. Valores atípicos
## 2 Navarra Leche líquida 1233200. Valores atípicos
## 3 Navarra Hortalizas frescas 809598. Valores atípicos
## 4 Navarra Carne 726151. Cuarto cuartil
## 5 Navarra Pan 674949. Cuarto cuartil
## 6 Navarra Derivados lacteos 426299. Cuarto cuartil
## 7 Navarra Refrescos 411391. Cuarto cuartil
## 8 Navarra Pesca 363698. Cuarto cuartil
## 9 Navarra Patatas 344521. Cuarto cuartil
## 10 Navarra Agua 316852. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 Cantabria Frutas frescas 1265818. Valores atípicos
## 2 Cantabria Leche líquida 1134970. Valores atípicos
## 3 Cantabria Hortalizas frescas 648784. Cuarto cuartil
## 4 Cantabria Agua 644432. Cuarto cuartil
## 5 Cantabria Carne 626673. Cuarto cuartil
## 6 Cantabria Pan 497095. Cuarto cuartil
## 7 Cantabria Derivados lacteos 457755. Cuarto cuartil
## 8 Cantabria Patatas 439306. Cuarto cuartil
## 9 Cantabria Refrescos 417747. Cuarto cuartil
## 10 Cantabria Pesca 394075. Tercer cuartil
## # i 27 more rows
## # A tibble: 37 x 4
## # Groups: REGIONES [1]
## REGIONES CATEGORIAS Volumen Estadisticas
## <fct> <fct> <dbl> <chr>
## 1 La Rioja Frutas frescas 654337. Valores atípicos
## 2 La Rioja Leche líquida 602490. Valores atípicos
## 3 La Rioja Carne 376920. Valores atípicos
## 4 La Rioja Hortalizas frescas 342420. Cuarto cuartil
## 5 La Rioja Pan 333074. Cuarto cuartil
## 6 La Rioja Agua 268369. Cuarto cuartil
## 7 La Rioja Derivados lacteos 222447. Cuarto cuartil
## 8 La Rioja Refrescos 198896. Cuarto cuartil
## 9 La Rioja Pesca 186441. Cuarto cuartil
## 10 La Rioja Patatas 149017. Tercer cuartil
## # i 27 more rows
```

```
rm(datos_regiones_categoria, regiones_unicas, estadisticas, datos_region, i)
```

Por regla general, las categorías de alimentos más consumidas en todas las regiones son la fruta fresca y la leche líquida. Le sigue las hortalizas frescas, la carne el pan y los refrescos. Otras categorías, pero en menor medida que las anteriores, pueden ser los derivados lácteos, las patatas y la pesca.

Las categorías que menos se consumen son los edulcorantes, las especias, los caldos, la miel, otras leches y los encurtidos.

El agua envasada dependiendo de la región tiene una prioridad u otras. Así, por ejemplo, en Canarias y Baleares es la categoría de alimentos que más se consume. Le sigue Cataluña y Valencia, que es el segundo, y después Castilla y León y Murcia, que es el tercero. Donde menos prioridad le dan es en Madrid, País Vasco y Navarra, donde ocupa la décima posición.

A priori, consideraba que la pesca, es decir, el pescado y el marisco, tenía mayor prioridad en zonas de costa, como Galicia o Cantabria, y en las islas Baleares y Canarias. Sin embargo, no se ha evidenciado esta hipótesis. Así, la pesca, por ejemplo ocupa la decima posición tanto en Extremadura, Baleares como en Cataluña.

9. Conclusiones

Como he indicado al comienzo, el presente análisis exploratorio tenía como finalidad comprender los patrones de consumo de alimentos por parte de los hogares españoles.

Con el fin de ayudarme en el desarrollo del análisis, formule una serie de preguntas guías que me permitirían cumplir con el objetivo marcado. Estas preguntas son:

- ¿Cómo ha evolucionado el volumen de consumo de alimentos en España a lo largo de los años?
- ¿En qué medida varía el consumo de alimentos entre las distintas regiones de España?
- ¿Cuáles son las categorías de alimentos más populares y de qué manera han experimentado cambios a lo largo de las décadas estudiadas?

Para hayar la respuesta a estas preguntas decidí utilizar un conjunto de datos, denominado “volumen_categorías”. Este conjunto de datos, constaba de 14.059 observaciones y 4 variables que son: AÑO, CATEGORIAS, REGIONES y VOLUMEN.

Las conclusiones a las que he podido llegar son las siguientes:

1. Evolución del Volumen de Consumo de Alimentos:

- Observamos que desde el año 2000 hasta 2022, en términos generales, la cantidad de alimentos consumidos por los hogares españoles ha experimentado un crecimiento, pasando de casi 24 millones a más de 25 millones de kilos. Sin embargo, este crecimiento no ha sido lineal y ha estado marcado por fluctuaciones significativas.
- El año 2023 comenzo en una etapa de caída del consumo de alimentos.

2. Variabilidad Regional en el Consumo de Alimentos:

- En general, durante los años 2021 y 2022, se observa una tendencia a la disminución en el consumo en todas regiones.
- He identificado diferencias notables en el consumo de alimentos entre las distintas regiones de España. Andalucía y Cataluña lideran en consumo, seguidas de cerca por Madrid y Valencia. En contraste, La Rioja, Cantabria y Navarra presentan los niveles de consumo más bajos.

3. Categorías de Alimentos Populares y Cambios en el Tiempo:

- En general, durante los años 2021 y 2022, se observa una tendencia a la disminución en el consumo en todas las categorías de alimentos.

- Al analizar las categorías de alimentos, Observé que la fruta fresca y la leche líquida son las más consumidas, tanto a nivel general como regional. Otras categorías populares incluyen las hortalizas frescas, la carne, el pan y los refrescos.
- Además, indentifiqué categorías menos consumidas, como edulcorantes, especias, caldos y encurtidos.
- El consumo de agua envasada varía según la región. Por ejemplo, en Canarias y Baleares, esta categoría ocupa el primer lugar, seguida de Cataluña y Valencia en segundo lugar. En contraste, Madrid, País Vasco y Navarra la ubican en la décima posición en sus preferencias.
- Contrariamente a la hipótesis inicial que sugería un mayor consumo de pescado y marisco en regiones costeras como Galicia y Cantabria, así como en las islas Baleares y Canarias, los datos no respaldan esta suposición. Estas categorías ocupan la décima posición en regiones como Extremadura, Baleares y Cataluña.