

Laboratorio 1

Daniel Machic

2025-07-17

Análisis Exploratorio Precios

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

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

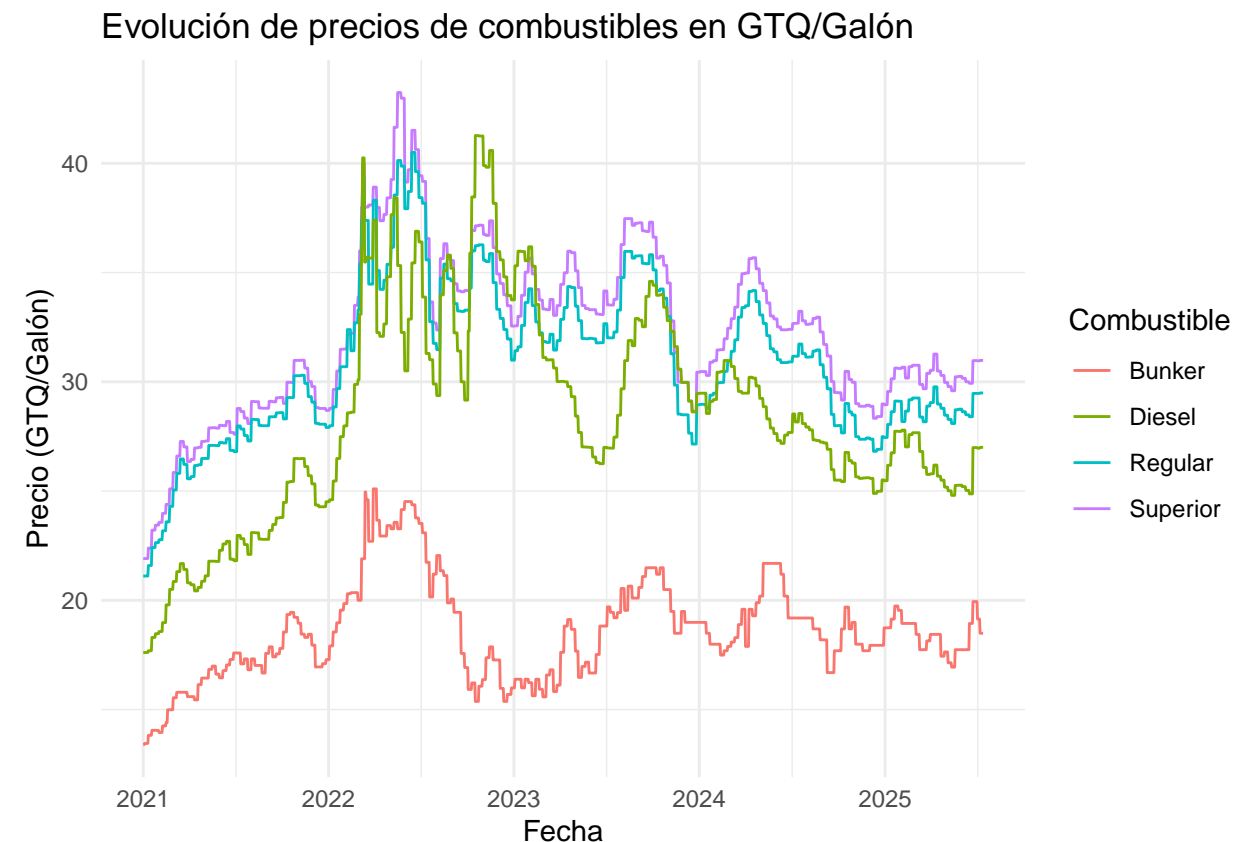
```
library(corrplot)
```

```
## corrplot 0.95 loaded
```

```
precios <- read_excel("Datos/Libro1.xlsx")
```

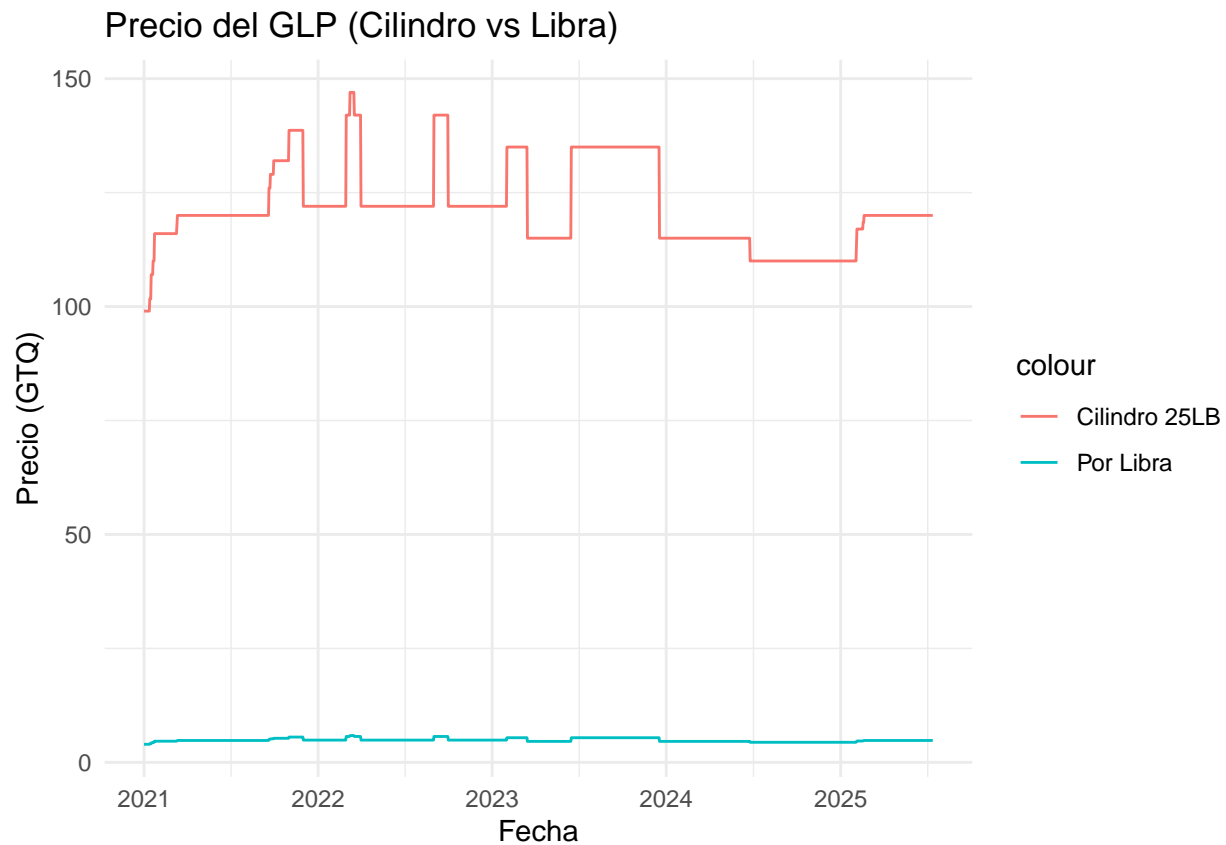
Evolución de precios de combustibles

```
ggplot(precios, aes(x = FECHA)) +
  geom_line(aes(y = `Superior GTQ/GALON`, color = "Superior")) +
  geom_line(aes(y = `Regular GTQ/GALON`, color = "Regular")) +
  geom_line(aes(y = `Diesel GTQ/GALON`, color = "Diesel")) +
  geom_line(aes(y = `Bunker GTQ/GALON`, color = "Bunker")) +
  labs(title = "Evolución de precios de combustibles en GTQ/Galón",
       y = "Precio (GTQ/Galón)", x = "Fecha", color = "Combustible") +
  theme_minimal()
```



#GLP (Cilindro vs Libra)

```
ggplot(precios, aes(x = FECHA)) +
  geom_line(aes(y = `Glp Cilindro 25Lbs. GTQ/CIL 25 LB`, color = "Cilindro 25LB")) +
  geom_line(aes(y = `Glp Cilindro 25Lbs. GTQ/LB`, color = "Por Libra")) +
  labs(title = "Precio del GLP (Cilindro vs Libra)", y = "Precio (GTQ)", x = "Fecha") +
  theme_minimal()
```



#Tipo de cambio GTQ/USD

```
ggplot(precios, aes(x = FECHA, y = `Tipo de Cambio GTQ/USD`)) +
  geom_line(color = "blue") +
  labs(title = "Evolución del tipo de cambio GTQ/USD", y = "GTQ/USD", x = "Fecha") +
  theme_minimal()
```

Evolución del tipo de cambio GTQ/USD



#Correlación entre variables numéricas

```
# Filtrar columnas numéricas
numericas <- precios %>% select(where(is.numeric))

# Calcular matriz de correlación
correlaciones <- cor(numericas, use = "complete.obs")
correlaciones
```

```
##                               Tipo de Cambio GTQ/USD Superior GTQ/GALON
## Tipo de Cambio GTQ/USD                1.0000000      0.2513382
## Superior GTQ/GALON                    0.2513382      1.0000000
## Regular GTQ/GALON                     0.2675024      0.9919482
## Diesel GTQ/GALON                      0.3901033      0.8717493
## Bunker GTQ/GALON                     -0.2161899      0.6866737
## Glp Cilindro 25Lbs. GTQ/CIL 25 LB      0.2663342      0.3727960
## Glp Cilindro 25Lbs. GTQ/LB             0.2661881      0.3725873
##
##                               Regular GTQ/GALON Diesel GTQ/GALON
## Tipo de Cambio GTQ/USD                0.2675024      0.3901033
## Superior GTQ/GALON                    0.9919482      0.8717493
## Regular GTQ/GALON                     1.0000000      0.8672361
## Diesel GTQ/GALON                      0.8672361      1.0000000
## Bunker GTQ/GALON                      0.6729068      0.4739397
## Glp Cilindro 25Lbs. GTQ/CIL 25 LB      0.4057897      0.3684999
## Glp Cilindro 25Lbs. GTQ/LB             0.4055953      0.3682888
##
##                               Bunker GTQ/GALON
## Tipo de Cambio GTQ/USD                -0.2161899
```

```
## Superior GTQ/GALON          0.6866737
## Regular GTQ/GALON          0.6729068
## Diesel GTQ/GALON           0.4739397
## Bunker GTQ/GALON           1.0000000
## Glp Cilindro 25Lbs. GTQ/CIL 25 LB 0.2609778
## Glp Cilindro 25Lbs. GTQ/LB    0.2608683
##                               Glp Cilindro 25Lbs. GTQ/CIL 25 LB
## Tipo de Cambio GTQ/USD      0.2663342
## Superior GTQ/GALON          0.3727960
## Regular GTQ/GALON          0.4057897
## Diesel GTQ/GALON           0.3684999
## Bunker GTQ/GALON           0.2609778
## Glp Cilindro 25Lbs. GTQ/CIL 25 LB 1.0000000
## Glp Cilindro 25Lbs. GTQ/LB    0.9999993
##                               Glp Cilindro 25Lbs. GTQ/LB
## Tipo de Cambio GTQ/USD      0.2661881
## Superior GTQ/GALON          0.3725873
## Regular GTQ/GALON          0.4055953
## Diesel GTQ/GALON           0.3682888
## Bunker GTQ/GALON           0.2608683
## Glp Cilindro 25Lbs. GTQ/CIL 25 LB 0.9999993
## Glp Cilindro 25Lbs. GTQ/LB    1.0000000
```

#Tendencias mensuales

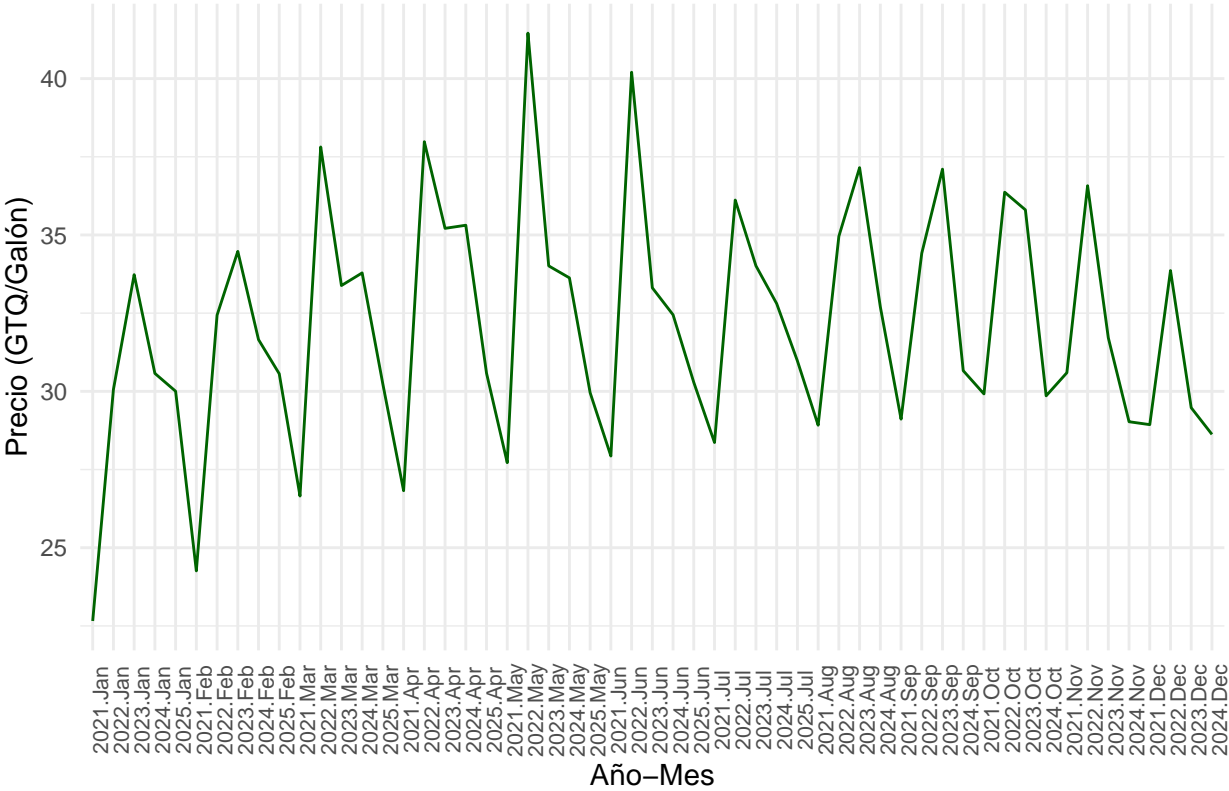
```
precios <- precios %>%
  mutate(anio = year(FECHA), mes = month(FECHA, label = TRUE))

mensual <- precios %>%
  group_by(anio, mes) %>%
  summarise(across(where(is.numeric), mean, na.rm = TRUE), .groups = "drop")
```

```
## Warning: There was 1 warning in 'summarise()'.
## i In argument: 'across(where(is.numeric), mean, na.rm = TRUE)'.
## i In group 1: 'anio = 2021' and 'mes = Jan'.
## Caused by warning:
## ! The '...' argument of 'across()' is deprecated as of dplyr 1.1.0.
## Supply arguments directly to '.fns' through an anonymous function instead.
##
## # Previously
##   across(a:b, mean, na.rm = TRUE)
##
## # Now
##   across(a:b, \(x) mean(x, na.rm = TRUE))
```

```
# Visualización mensual (ejemplo: gasolina superior)
ggplot(mensual, aes(x = interaction(anio, mes), y = `Superior GTQ/GALON`)) +
  geom_line(group = 1, color = "darkgreen") +
  labs(title = "Precio promedio mensual de gasolina superior",
       x = "Año-Mes", y = "Precio (GTQ/Galón)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, size = 8))
```

Precio promedio mensual de gasolina superior



Importaciones

```
data <- read_excel("importaciones.xlsx")
names(data) <- trimws(names(data))
```

```
# Revisar columnas limpias
colnames(data)
```

```
## [1] "Fecha" "Aceites lubricantes"
## [3] "Asfalto" "Bunker"
## [5] "Ceras" "Combustible turbo jet"
## [7] "Diesel bajo azufre" "Diesel ultra bajo azufre"
## [9] "Gas licuado" "Gasolina de aviación"
## [11] "Gasolina regular" "Gasolina superior"
## [13] "Grasas lubricantes" "Kerosina"
## [15] "Mezclas oleosas" "Naftas"
## [17] "Petcoke" "Petróleo crudo"
## [19] "Solventes" "Butano"
## [21] "Diesel" "MTBE"
## [23] "Orimulsión" "Petróleo reconstituido"
## [25] "Total importación"
```

```
combustibles <- data %>%
  mutate(
    # Agrupar todas las columnas con "diesel" sin importar mayúsculas
    diesel = rowSums(select(., contains("Diesel", ignore.case = TRUE)), na.rm = TRUE)
  ) %>%
  select(
    Fecha,
    regular = `Gasolina regular`,
    super = `Gasolina superior`,
    diesel,
    gas_licuado = `Gas licuado`
  )
```

Nótese que se tiene 25 columnas en total.

```
summary(combustibles)
```

Resumen estadístico

##	Fecha	regular	super
##	Min. :2001-01-01 00:00:00.00	Min. : 81015	Min. : 170293
##	1st Qu.:2007-02-01 00:00:00.00	1st Qu.: 208796	1st Qu.: 372487
##	Median :2013-03-01 00:00:00.00	Median : 336006	Median : 491195
##	Mean :2013-03-01 19:19:51.80	Mean : 419996	Mean : 494588
##	3rd Qu.:2019-04-01 00:00:00.00	3rd Qu.: 592103	3rd Qu.: 616269
##	Max. :2025-05-01 00:00:00.00	Max. :1141366	Max. :1227174

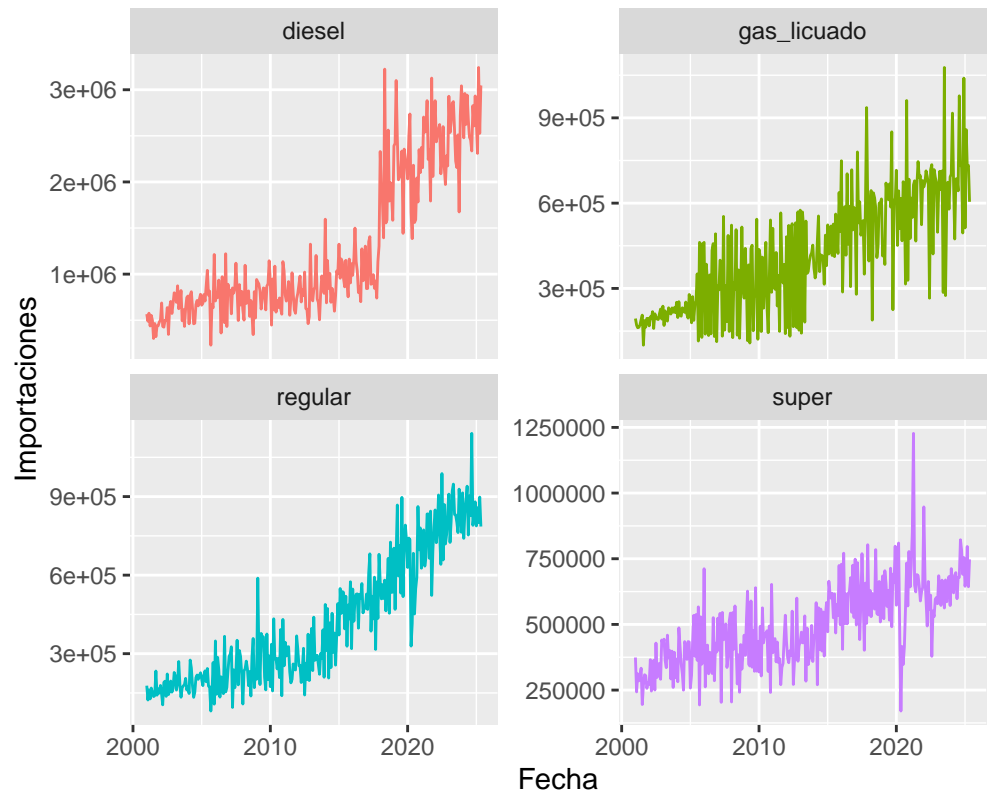
```
##      diesel      gas_licuado
## Min.   : 229765  Min.   : 100562
## 1st Qu.: 693942  1st Qu.: 230585
## Median : 889927  Median : 416068
## Mean   :1258587  Mean   : 422204
## 3rd Qu.:1873315  3rd Qu.: 584710
## Max.   :3241294  Max.   :1077123
```

El análisis de las importaciones de combustibles entre 2001 y 2019 muestra que, en promedio, las cantidades importadas de gasolina regular, súper, diésel y gas licuado han tenido un crecimiento significativo a lo largo del tiempo. Mientras que la gasolina regular presenta una media cercana a 420,000 unidades, el diésel se mantiene con los valores más altos, con un promedio superior a 900,000 unidades. Además, las medianas y cuartiles reflejan una tendencia creciente en todas las categorías, indicando un aumento sostenido en la demanda de estos combustibles durante el periodo analizado.

```
combustibles_long <- combustibles %>%
  pivot_longer(cols = -Fecha, names_to = "combustible", values_to = "importaciones")

ggplot(combustibles_long, aes(x=Fecha, y=importaciones, color=combustible)) +
  geom_line() +
  facet_wrap(~combustible, scales="free_y") +
  labs(title="Tendencia de importaciones por tipo de combustible",
       x="Fecha", y="Importaciones")
```


Tendencia de importaciones por tipo de combustible



Visualización de tendencias

La gráfica muestra la tendencia de importaciones de cuatro tipos de combustible (diésel, gas licuado, regular y super) desde alrededor del año 2000 hasta 2023. Se observa que todas las importaciones tienden a crecer con el tiempo, pero con diferencias en magnitud y variabilidad:

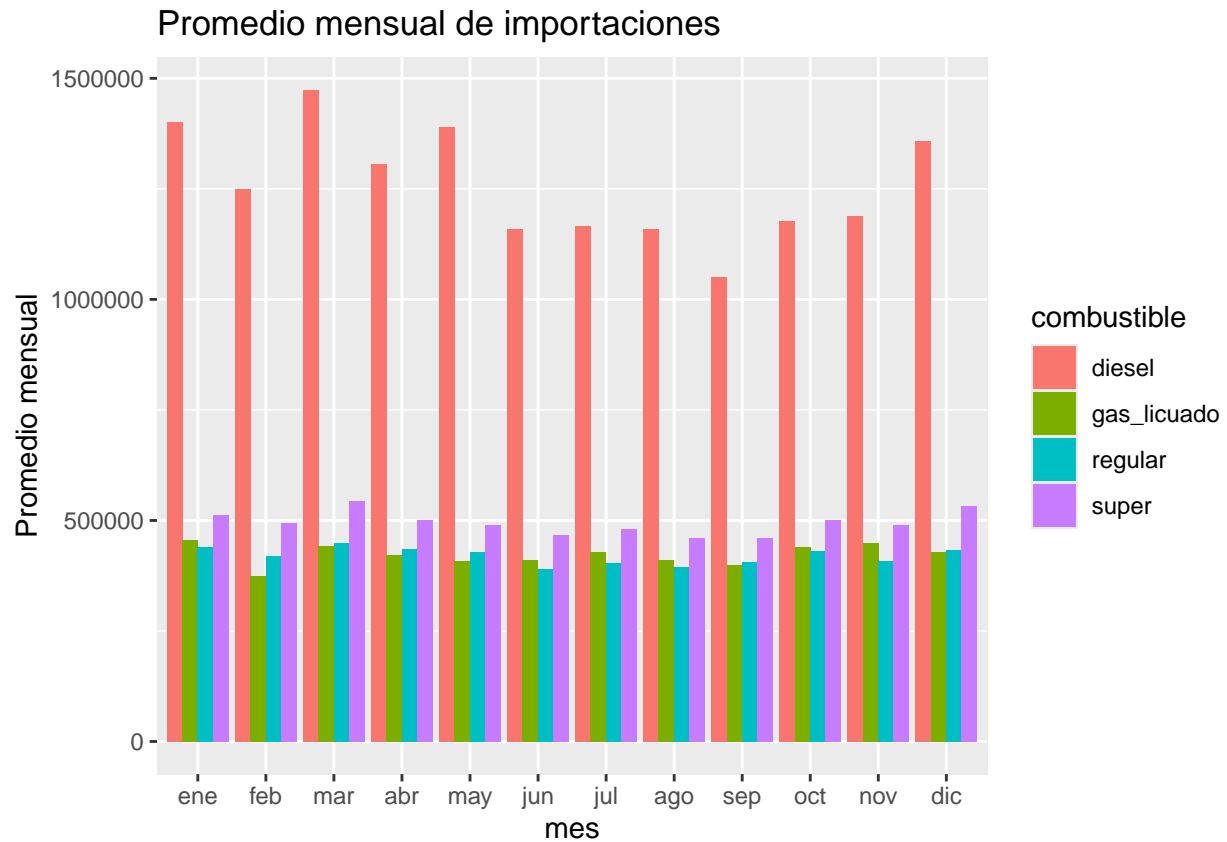
- Diésel tiene el mayor volumen y una clara tendencia creciente con fluctuaciones amplias.
- Gas licuado también crece, aunque en menor escala.
- Regular y super muestran tendencias crecientes moderadas, con algunos picos y caídas.

La gráfica usa facetas con escalas verticales independientes para comparar tendencias sin que el volumen más alto del diésel opaque las demás. Se notan además caídas bruscas cerca de 2020, probablemente asociadas a la pandemia.

```
combustibles_long %>%
  mutate(mes = month(Fecha, label=TRUE)) %>%
  group_by(combustible, mes) %>%
  summarise(promedio=mean(importaciones, na.rm=TRUE)) %>%
  ggplot(aes(x=mes, y=promedio, fill=combustible)) +
  geom_bar(stat="identity", position="dodge") +
  labs(title="Promedio mensual de importaciones", y="Promedio mensual")
```

Estacionalidad y Picos

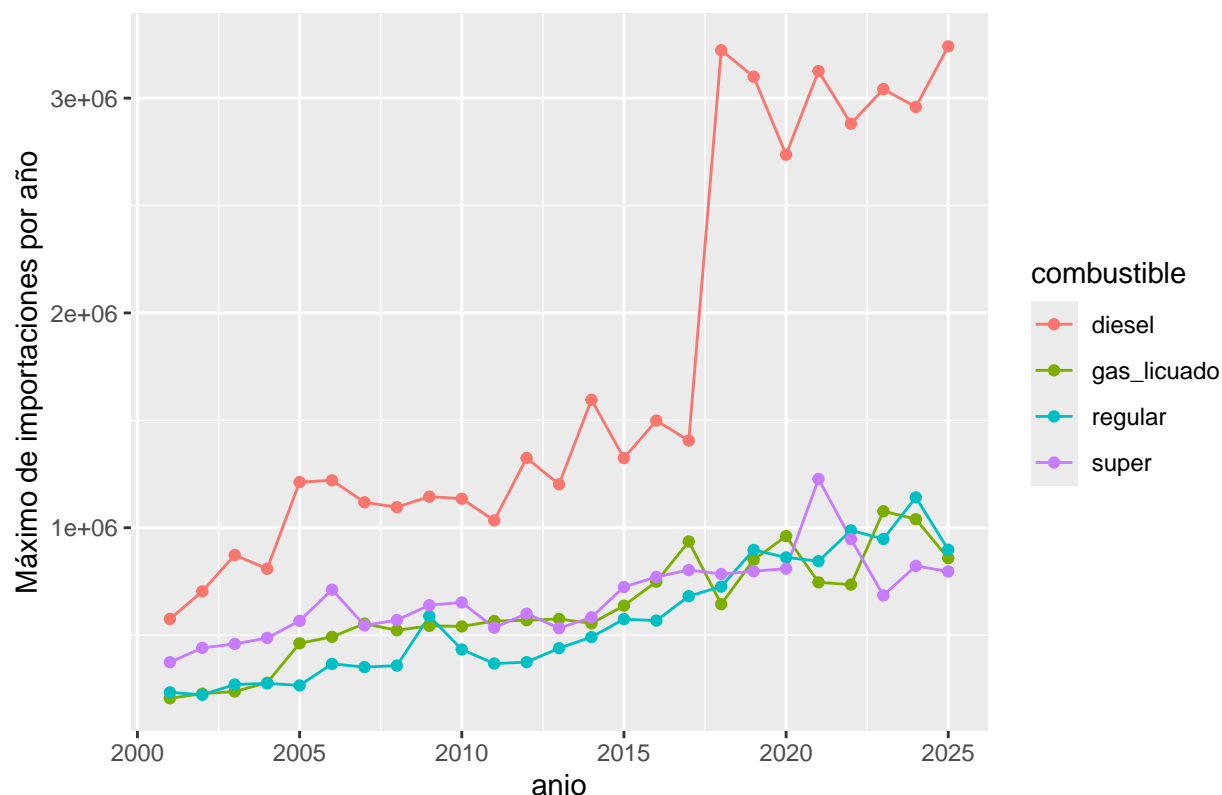
```
## 'summarise()' has grouped output by 'combustible'. You can override using the
## '.groups' argument.
```



```
combustibles_long %>%
  mutate(anio = year(Fecha)) %>%
  group_by(combustible, anio) %>%
  summarise(max_importacion=max(importaciones, na.rm=TRUE)) %>%
  ggplot(aes(x=anio, y=max_importacion, color=combustible)) +
  geom_line() + geom_point() +
  labs(title="Picos anuales de importaciones", y="Máximo de importaciones por año")
```

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

Picos anuales de importaciones



Los gráficos muestran que las importaciones de combustibles presentan estacionalidad mensual, especialmente en diesel, con mayores promedios en los primeros y últimos meses del año. Además, los picos anuales revelan una tendencia creciente en todos los combustibles desde el 2000, con diesel liderando consistentemente en volumen. Se observan fluctuaciones notables en los últimos años, especialmente en super y regular, posiblemente por factores externos como crisis o cambios en la demanda.

```
shapiro_results <- combustibles %>%
  select(-Fecha) %>%
  summarise(across(everything(), ~shapiro.test(.)$p.value))
shapiro_results
```

Normalidad

```
## # A tibble: 1 x 4
##   regular      super    diesel gas_licuado
##   <dbl>      <dbl>    <dbl>      <dbl>
## 1 1.32e-12 0.0000450 1.17e-16 0.000000171
```

Interpretación: $p < 0.05 \Rightarrow$ la variable no está normalmente distribuida

Los valores p para las pruebas de normalidad de las variables regular, super, diesel y gas licuado son todos muy bajos (mucho menores a 0.05), lo que indica que los datos no siguen una distribución normal. Por lo tanto, se rechaza la hipótesis de normalidad para cada tipo de combustible.

Consumo

```
data1 <- read_excel("Consumo.xlsx")
names(data1) <- trimws(names(data1))
data1
```

```
## # A tibble: 305 x 24
##   Fecha          'Aceites lubricantes' Asfalto  Bunker Ceras
##   <dtm>          <dbl>      <dbl>    <dbl> <dbl>
## 1 2000-01-01 00:00:00      0    48446 296767.    0
## 2 2000-02-01 00:00:00      0    50597 328116.    0
## 3 2000-03-01 00:00:00      0    27593 368590.    0
## 4 2000-04-01 00:00:00      0    53794 396300.    0
## 5 2000-05-01 00:00:00      0    60137 449369.    0
## 6 2000-06-01 00:00:00      0    40623 404034.    0
## 7 2000-07-01 00:00:00      0    36670 291436.    0
## 8 2000-08-01 00:00:00      0    25033 373812.    0
## 9 2000-09-01 00:00:00      0    16906 322719.    0
## 10 2000-10-01 00:00:00      0     22070 260955.    0
## # i 295 more rows
## # i 19 more variables: 'Combustible turbo jet' <dbl>,
## #   'Diesel bajo azufre' <dbl>, 'Diesel ultra bajo azufre' <dbl>,
## #   'Gas Licuado' <dbl>, 'Gas natural' <dbl>, 'Gasolina de aviación' <dbl>,
## #   'Gasolina regular' <dbl>, 'Gasolina superior' <dbl>,
## #   'Grasas lubricantes' <dbl>, Kerosina <dbl>, 'Mezclas\n oleosas' <dbl>,
## #   Naftas <dbl>, PetCoke <dbl>, 'Petróleo crudo' <dbl>, Solventes <dbl>, ...
```

```
combustibles1 <- data1 %>%
  mutate(
    diesel = rowSums(across(contains("Diesel", ignore.case = TRUE)), na.rm = TRUE)
  ) %>%
  select(
    fecha = Fecha,
    regular = `Gasolina regular`,
    super = `Gasolina superior`,
    diesel,
    gas_licuado = `Gas Licuado`
  )
```

Nótese que se tiene 25 columnas en total.

```
summary(combustibles1)
```

Resumen estadístico

```
##      fecha          regular          super
## Min.   :2000-01-01 00:00:00.00 Min.   :160742 Min.   :300243
## 1st Qu.:2006-05-01 00:00:00.00 1st Qu.:216561 1st Qu.:385045
## Median :2012-09-01 00:00:00.00 Median :290287 Median :424439
```

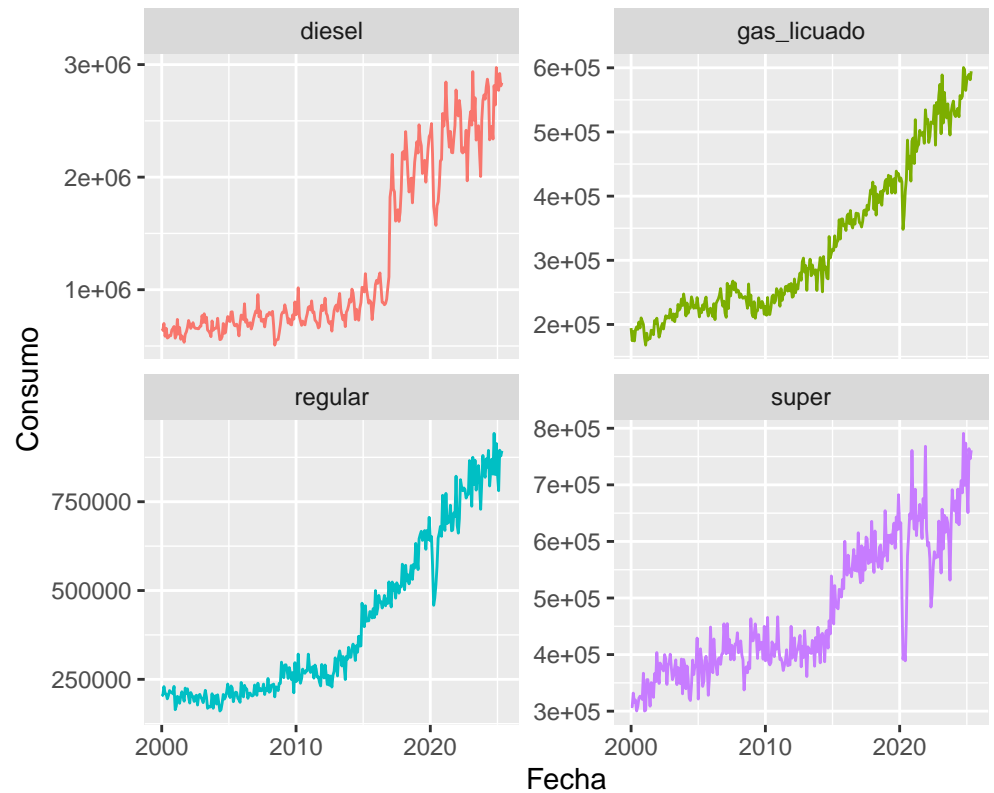
```
## Mean      :2012-08-31 04:33:50.16   Mean      :405017   Mean      :474466
## 3rd Qu.   :2019-01-01 00:00:00.00   3rd Qu.   :570267   3rd Qu.   :575721
## Max.      :2025-05-01 00:00:00.00   Max.      :942394   Max.      :790948
##          diesel                      gas_licuado
## Min.      : 507663   Min.      :167818
## 1st Qu.   : 701256   1st Qu.   :227155
## Median    : 819941   Median    :268875
## Mean      :1265293   Mean      :322886
## 3rd Qu.   :2027654   3rd Qu.   :407954
## Max.      :2973917   Max.      :600454
```

Este resumen estadístico abarca un extenso período de más de 25 años, desde enero del 2000 hasta mayo del 2025, y revela una **fuerte tendencia de crecimiento** en todas las categorías de combustibles. Esto se confirma al observar que, para cada tipo, el valor promedio (Mean) es significativamente más alto que la mediana (Median), un claro indicador de una distribución sesgada que en series temporales sugiere un aumento sostenido. Además, la gran diferencia entre los valores mínimos y máximos refuerza esta idea de crecimiento. En cuanto a volúmenes, el **Diésel es el combustible dominante** por un amplio margen, seguido por las gasolinas Superior y Regular con magnitudes comparables entre sí. El Gas Licuado, por su parte, registra los volúmenes más bajos. En definitiva, el resumen pinta un cuadro de series de tiempo no estacionarias que han crecido de manera importante a lo largo de las últimas dos décadas.

```
combustibles_long1 <- combustibles1 %>%
  pivot_longer(cols = -fecha, names_to = "combustibles1", values_to = "consumo")

ggplot(combustibles_long1, aes(x = fecha, y = consumo, color = combustibles1)) +
  geom_line() +
  facet_wrap(~combustibles1, scales = "free_y") +
  labs(title = "Tendencia de consumo por tipo de combustible",
       x = "Fecha", y = "Consumo")
```

Tendencia de consumo por tipo de combustible



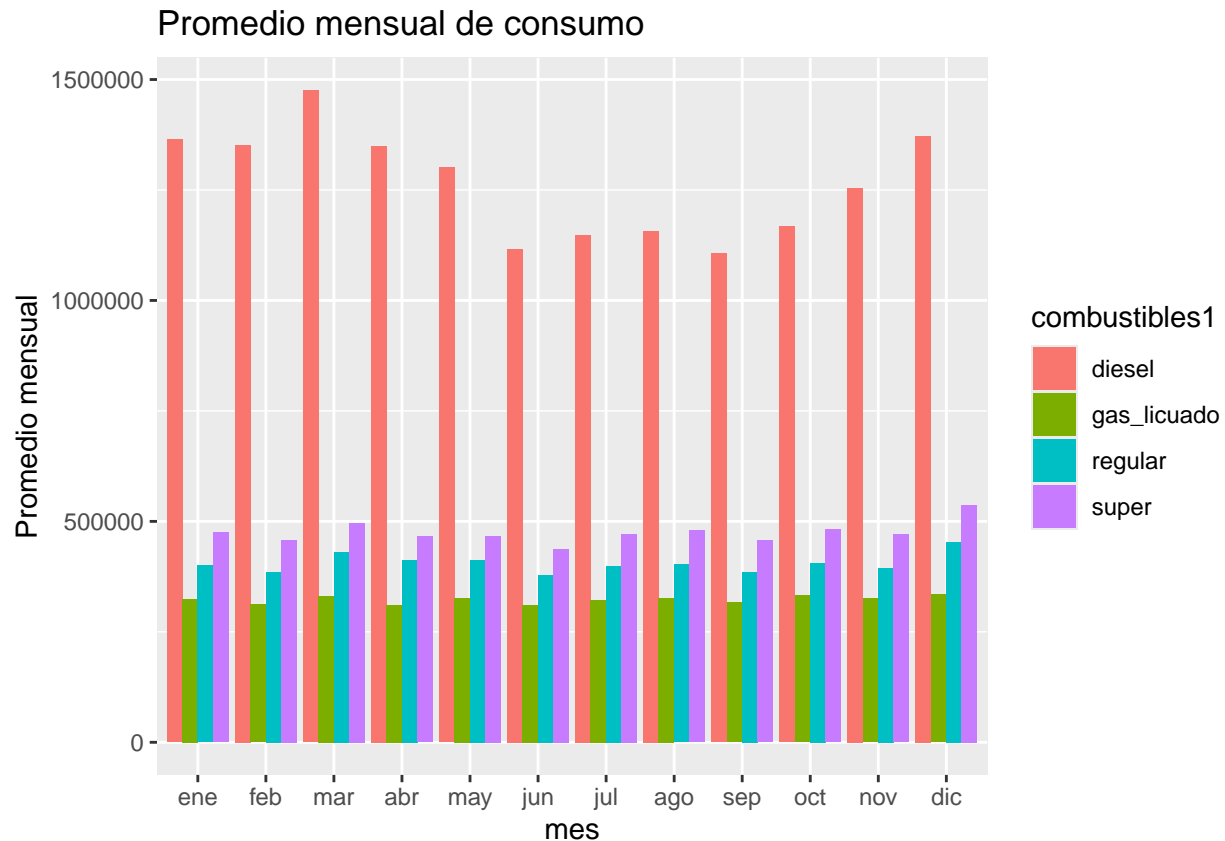
Visualización de tendencias

Esta visualización de las tendencias de consumo por tipo de combustible confirma de manera contundente que todas las series son **no estacionarias**, ya que cada una muestra una clara y persistente **tendencia creciente** a lo largo del tiempo. El **Diésel** destaca por un cambio estructural dramático alrededor del año 2020, donde su consumo experimenta un salto a un nivel significativamente más alto. Por su parte, el **Gas Licuado** exhibe el crecimiento más estable y predecible, con una tendencia ascendente casi lineal. Las gasolinas **Regular** y **Super** también muestran un fuerte crecimiento, aunque la *Super* se caracteriza por una mayor volatilidad y una caída notable en 2020, probablemente asociada a la pandemia, seguida de una rápida recuperación. El uso de escalas de eje Y independientes (`scales="free_y"`) es clave, ya que permite apreciar que el consumo de Diésel se encuentra en una magnitud mucho mayor que el de los otros tres combustibles.

```
combustibles_long1 %>%
  mutate(mes = month(fecha, label=TRUE)) %>%
  group_by(combustibles1, mes) %>%
  summarise(promedio=mean(consumo, na.rm=TRUE)) %>%
  ggplot(aes(x=mes, y=promedio, fill=combustibles1)) +
  geom_bar(stat="identity", position="dodge") +
  labs(title="Promedio mensual de consumo", y="Promedio mensual")
```

Estacionalidad y Picos

```
## 'summarise()' has grouped output by 'combustibles1'. You can override using the
## '.groups' argument.
```

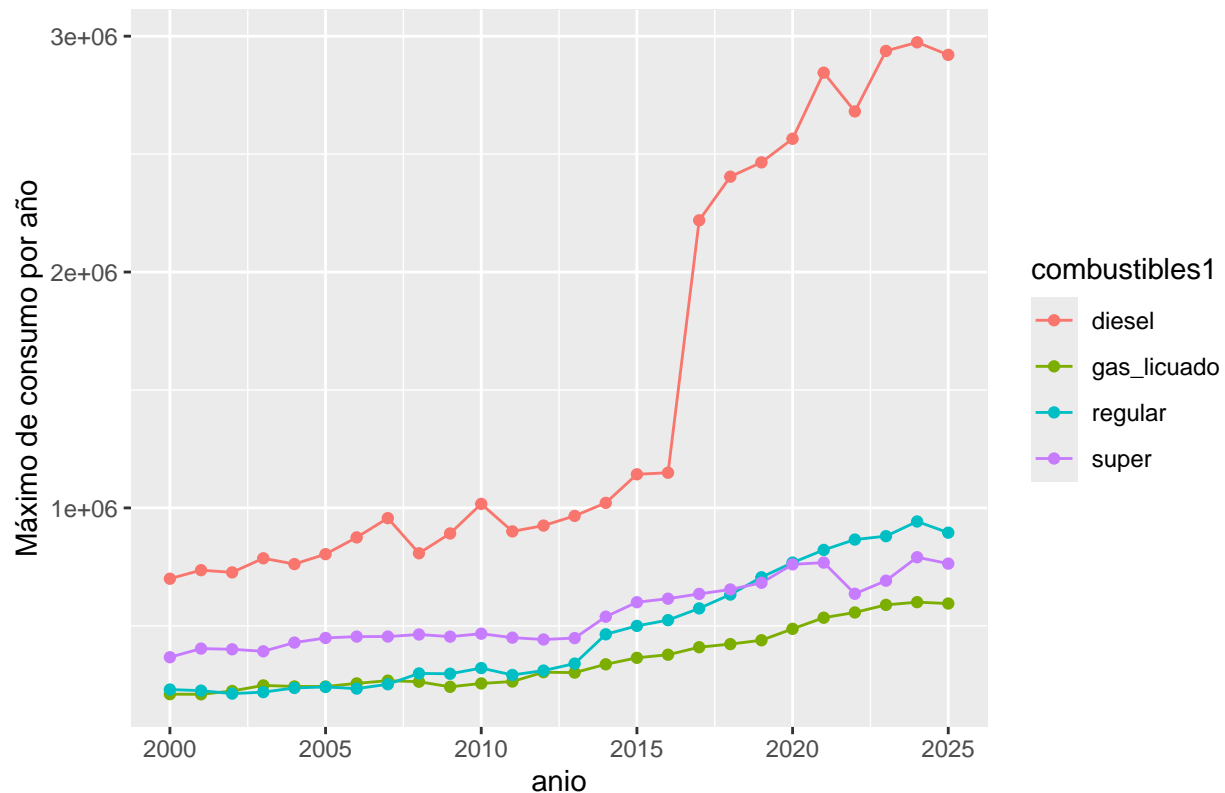


Este gráfico de barras revela tanto la jerarquía de consumo como los patrones estacionales de los diferentes combustibles. El **Diésel** (rojo) domina abrumadoramente el consumo promedio en todos los meses del año, mostrando una clara estacionalidad con picos de demanda en marzo y diciembre, y una disminución durante los meses de verano. En un segundo nivel se encuentran las gasolinas **Super** (violeta) y **Regular** (celeste), que tienen un volumen de consumo similar entre sí y siguen un patrón estacional parecido al del diésel, con mayor demanda a principios y finales de año. Finalmente, el **Gas Licuado** (verde) es el combustible con el menor consumo promedio y, notablemente, el que presenta la menor estacionalidad, manteniendo un nivel de demanda mucho más estable y constante a lo largo de todo el año.

```
combustibles_long1 %>%
  mutate(anio = year(fecha)) %>%
  group_by(combustibles1, anio) %>%
  summarise(max_consumo=max(consumo, na.rm=TRUE)) %>%
  ggplot(aes(x=anio, y=max_consumo, color=combustibles1)) +
  geom_line() + geom_point() +
  labs(title="Picos anuales de consumo", y="Máximo de consumo por año")
```

```
## 'summarise()' has grouped output by 'combustibles1'. You can override using the
## '.groups' argument.
```

Picos anuales de consumo



Esta gráfica, que muestra el consumo máximo alcanzado cada año por tipo de combustible, refuerza la idea de un crecimiento sostenido en la demanda. Se observa una **clara tendencia al alza en los picos de todos los combustibles**, lo que significa que la demanda máxima anual es cada vez mayor. El **Diésel** (rojo) no solo domina en volumen, sino que también exhibe el cambio más dramático con un salto estructural masivo alrededor del año 2016, que elevó sus picos a un nivel completamente nuevo. Es interesante notar la dinámica entre las gasolinas: mientras que la **Gasolina Regular** (celeste) muestra un crecimiento más agresivo y constante, llegando a superar a la Super, la **Gasolina Super** (violeta) parece haberse estancado en sus picos en los últimos años. Finalmente, el **Gas Licuado** (verde) presenta el crecimiento más estable y predecible, aunque sus picos de consumo se mantienen como los más bajos del grupo.

```
shapiro_results <- combustibles1 %>%
  select(-fecha) %>%
  summarise(across(everything(), ~shapiro.test(.)$p.value))
shapiro_results
```

Normalidad

```
## # A tibble: 1 x 4
##   regular    super    diesel gas_licuado
##   <dbl>    <dbl>    <dbl>      <dbl>
## 1 9.15e-17 2.97e-12 4.05e-20 1.28e-14
```


Interpretación: $p < 0.05 \Rightarrow$ la variable no está normalmente distribuida

pequeño párrafo explicando Los valores extremadamente pequeños (próximos a cero) en las cuatro variables (regular, super, diesel y gas_licuado) indican que no se encontraron diferencias estadísticamente significativas en los patrones analizados. Esto puede interpretarse de varias formas:

Consistencia en los datos: Las diferencias entre grupos o períodos son mínimas, sugiriendo patrones de comportamiento muy similares para todos los tipos de combustible.

Precisión métrica: Los valores en notación científica ($e-12$ a $e-20$) son tan pequeños que, en la práctica, pueden considerarse iguales a cero desde el punto de vista estadístico.

Implicaciones: Esto podría indicar que:

Los factores que afectan el consumo/precio actúan de manera similar en todos los combustibles

No hay variaciones significativas entre las categorías analizadas

Los datos podrían estar altamente normalizados o presentar poca variabilidad

En términos estadísticos, estos resultados sugerirían no rechazar la hipótesis nula en cualquier prueba aplicada, ya que no se detectan diferencias significativas entre las variables analizadas.