PAC - Mejora de la eficiencia operativa y maximización de la rentabilidad de la producción de Palta Haas en la empresa Agricola Guili S.A.C.

Grupo Nº4 - Fernandez D. Abel, Fernandez D. Rody, Márquez M. Andre, Yactayo C. David

20 de abril de 2024

Contents

I. Problemática

Frente al contexto competitivo y dinámico del mercado actual, la empresa Agricola Guili S.A.C. se plantea una interrogante esencial, ¿Cómo puede la empresa mejorar la eficiencia operativa a través del análisis estadístico de la producción y distribución de calibres para optimizar la asignación de recursos y maximizar la rentabilidad? Esta pregunta se convierte en el eje de una investigación exhaustiva, buscando desentrañar las complejidades del proceso productivo y logístico mediante el uso riguroso de herramientas analíticas y estadísticas.

Con el fin de abordar este desafío, la empresa se ha propuesto un objetivo claro y definido, Examinar las características y tendencias en la producción y distribución de calibres de palta para identificar áreas clave que permitan optimizar la asignación de recursos, con el propósito final de potenciar la eficiencia operativa y la productividad de la empresa. La consecución de este objetivo será fundamental para informar decisiones estratégicas que impulsen un desarrollo sostenible y mejoren la posición competitiva de la empresa en la industria.

Este estudio se centra en el análisis de un conjunto de datos exhaustivo que recopila información sobre la producción y comercialización de paltas por parte de diversos clientes a lo largo de un período determinado. El objetivo principal es comprender las características y patrones de la producción de calibres de paltas, con el fin de identificar posibles áreas de mejora y optimización. El conjunto de datos incluye variables como la semana, el mes, el tipo de palta (variedad), el color, el calibre específico y la cantidad en kilogramos.

II. EDA: Análisis Exploratorio de Datos

2.1. Descripción de la data

Se ha utilizado una data que esta conformado por dos datasets. Cada dataset provee datos específicos para evaluar distintas fases y aspectos de la producción agrícola, desde la clasificación y calibre de los frutos hasta la eficiencia y resultados de las operaciones de cosecha y proyecciones de producción.

2.1.1. Dataset “CALIBRE”

Registros: 260

Variables:

- Semana: Número de la semana del año.
- Mes: Mes de registro.
- Variedad: Tipo de palta, en este caso, Hass.
- Color: Color del fruto.
- Calibre: Tamaño del fruto.
- Kg Calibre: Peso en kilogramos correspondiente al calibre.
- % Calibre: Porcentaje que representa el calibre dentro de la producción total.
- N° Guia de Remisión: Número identificador de la guía de remisión.
- N° Reporte de Producción: Número identificador del reporte de producción.
- Fecha: Fecha de registro.
- CLIENTE: Nombre del cliente

2.1.2. Dataset “REPORTE”

Registros: 121

Variables:

- Semana, Mes, Fundo, Empresa, Lote, Ha: Información de tiempo y lugar de producción.
- Variedad, Color: Características del fruto cosechado.
- N° Guia de Remisión, N° Reporte de Producción, Fecha: Identificadores y fecha de reporte.

- Total Jabas, Peso Promedio Jaba (Kg), Cajas Exportadas (10 Kg): Detalles de empaque y exportación.
- Ingreso Packing (Kg Bruto), Kg Exportados, % Exportado, Kg Descarte, % Descarte, Kg Merma, % Merma, Kg Descarte de Campo: Pesos y porcentajes de clasificación post-cosecha.
- Kg Brutos Lote, Kg Brutos Ha, Kg Exportado Ha: Producción total y por hectárea.
- Status, CLIENTE: Estado actual del lote y cliente asociado.
- Variables climáticas y de tratamiento: TEMP PROM, HUM. PROM, ET ACUMULADA, PRODUCTO, NATURALEZA, DOSIS - L/HA, TIPO.

2.2. Configuración del Entorno

Como paso previo al análisis, se establece el directorio de trabajo adecuado. A continuación, se importa el conjunto de datos denominado "FINAL TOTAL.xlsx" desde la hoja "Calibres". Para garantizar la integridad de la información, se procede a eliminar la última fila del conjunto de datos.

```
# Configuración del directorio de trabajo
setwd("D:/1.Maestria Ciencia Datos/03. INTRODUCCIÓN A LOS MODELOS ESTADÍSTICOS-23MCDAP002-PSMA-00609-19

# Importando el conjunto de datos
data <- read_excel("FINAL TOTAL.xlsx", sheet = "Calibres")

# Eliminando la última fila
data <- data[-nrow(data), ]
```

2.3. Exploración Inicial del Conjunto de Datos

Con el fin de familiarizarse con la estructura y el contenido del conjunto de datos, se realiza una exploración inicial. En primer lugar, se observan las primeras filas del conjunto de datos para obtener una idea general de las variables y sus valores. Posteriormente, se imprimen los nombres de las columnas para comprender mejor la organización de la información.

```
# Explorando el conjunto de datos

# Muestra las primeras filas del conjunto de datos para obtener una visión general de su contenido
head(data)
```

```
## # A tibble: 6 x 11
##   Semana Mes   Variedad Color Calibre `Kg Calibre` `% Calibre`
##    <dbl> <chr> <chr>    <chr> <chr>          <dbl>       <dbl>
## 1     13 Marzo Hass     Negra 08              60.0     0.00249
## 2     13 Marzo Hass     Negra 10            2160       0.0898
## 3     13 Marzo Hass     Negra 12            9640       0.401
## 4     13 Marzo Hass     Negra 14           10250.      0.426
## 5     13 Marzo Hass     Negra 16            1890.      0.0786
## 6     13 Marzo Hass     Negra 20              30.0     0.00125
## # i 4 more variables: `N° Guia de Remisión` <chr>,
## #   `N° Reporte de Producción` <chr>, Fecha <dttm>, CLIENTE <chr>
```

```
# Imprime los nombres de las columnas del conjunto de datos
names(data)
```

```
##  [1] "Semana"                    "Mes"
##  [3] "Variedad"                  "Color"
##  [5] "Calibre"                   "Kg Calibre"
##  [7] "% Calibre"                 "N° Guia de Remisión"
##  [9] "N° Reporte de Producción" "Fecha"
## [11] "CLIENTE"
```

2.4. Renombrado de Columnas

Para mejorar la claridad y facilitar el manejo de las variables, se renombran algunas columnas del conjunto de datos. Se utiliza el paquete dplyr para realizar esta tarea de manera eficiente. Los nuevos nombres de las columnas se muestran en el código.

```
# Renombrando columnas (usando dplyr)
data <- data %>%
  rename(
    KgCal = `Kg Calibre`,
    Perc_Cal = `% Calibre`,
    NoGR = `N° Guia de Remisión`,
    NoRepProd = `N° Reporte de Producción`,
    Cliente = `CLIENTE`,
    # Se puede continuar renombrando según sea necesario
  )

# Verificando los nuevos nombres de columna
colnames(data)
```

```
##  [1] "Semana"    "Mes"       "Variedad"  "Color"     "Calibre"   "KgCal"
##  [7] "Perc_Cal"  "NoGR"      "NoRepProd" "Fecha"     "Cliente"
```

2.5. Exploración Avanzada de la Estructura

Para profundizar en la comprensión de la estructura del conjunto de datos, se utiliza la función str(). Esta función proporciona información detallada sobre el tipo de datos de cada columna y los primeros valores de cada una. Además, se verifica la presencia de datos faltantes utilizando la función sum(is.na(data)).

```
# Explorando la estructura del conjunto de datos
str(data)
```

```
## tibble [260 x 11] (S3: tbl_df/tbl/data.frame)
##  $ Semana   : num [1:260] 13 13 13 13 13 13 13 13 13 13 ...
##  $ Mes      : chr [1:260] "Marzo" "Marzo" "Marzo" "Marzo" ...
##  $ Variedad : chr [1:260] "Hass" "Hass" "Hass" "Hass" ...
##  $ Color    : chr [1:260] "Negra" "Negra" "Negra" "Negra" ...
##  $ Calibre  : chr [1:260] "08" "10" "12" "14" ...
##  $ KgCal    : num [1:260] 60 2160 9640 10250 1890 ...
##  $ Perc_Cal : num [1:260] 0.00249 0.08978 0.40067 0.42602 0.07855 ...
##  $ NoGR     : chr [1:260] "T008 N° 0000109" "T008 N° 0000109" "T008 N° 0000109" "T008 N° 0000109" ..
##  $ NoRepProd: chr [1:260] "0001-0002239" "0001-0002239" "0001-0002239" "0001-0002239" ...
##  $ Fecha    : POSIXct[1:260], format: "2024-03-25" "2024-03-25" ...
##  $ Cliente  : chr [1:260] "BAIKA" "BAIKA" "BAIKA" "BAIKA" ...
```

```
# Buscando datos faltantes
sum(is.na(data))
```

```
## [1] 0
```

2.6. Verificación Final de Nombres de Columna

Finalmente, se realiza una última verificación de los nombres de las columnas

```
# Verificando los nombres de las columnas después de todas las manipulaciones
colnames(data)
```

```
##  [1] "Semana"    "Mes"       "Variedad"  "Color"     "Calibre"   "KgCal"
```

```
##  [7] "Perc_Cal"  "NoGR"      "NoRepProd" "Fecha"     "Cliente"
```

III. Análisis Descriptivo Inicial

3.1. Resumen estadístico

Se llevó a cabo un resumen estadístico inicial con el objetivo de obtener una comprensión básica de las distribuciones de las variables numéricas, como "KgCal" (Kilogramos por Calibre) y "Perc_Cal" (Porcentaje por Calibre).

```
# Basic summary statistics
summary(data)
```

```
##      Semana          Mes              Variedad           Color
##  Min.   :13.00   Length:260         Length:260         Length:260
##  1st Qu.:14.00   Class :character   Class :character   Class :character
##  Median :15.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   :14.76
##  3rd Qu.:15.00
##  Max.   :16.00
##    Calibre             KgCal            Perc_Cal             NoGR
##  Length:260         Min.   :   10.0   Min.   :0.0002174   Length:260
##  Class :character   1st Qu.:  517.5   1st Qu.:0.0256626   Class :character
##  Mode  :character   Median : 1495.0   Median :0.0928920   Mode  :character
##                     Mean   : 2707.3   Mean   :0.1461538
##                     3rd Qu.: 3872.5   3rd Qu.:0.1730048
##                     Max.   :17680.0   Max.   :1.0000000
##   NoRepProd             Fecha                          Cliente
##  Length:260         Min.   :2024-03-25 00:00:00.00   Length:260
##  Class :character   1st Qu.:2024-04-05 00:00:00.00   Class :character
##  Mode  :character   Median :2024-04-10 00:00:00.00   Mode  :character
##                     Mean   :2024-04-08 12:49:50.76
##                     3rd Qu.:2024-04-13 00:00:00.00
##                     Max.   :2024-04-16 00:00:00.00
```

```
numeric_columns <- sapply(data, is.numeric)
data_n <- data[, numeric_columns]

print(numeric_columns)
```

```
##    Semana       Mes  Variedad     Color   Calibre     KgCal  Perc_Cal      NoGR
##      TRUE     FALSE     FALSE     FALSE     FALSE      TRUE      TRUE     FALSE
## NoRepProd     Fecha   Cliente
##     FALSE     FALSE     FALSE
```

Interpretación de las variables numéricas

- **Semana**: La semana mínima es 13 y la máxima es 16. Esto indica que el conjunto de datos abarca un período de 4 semanas. La media de 14.76 sugiere que la mayoría de las observaciones se encuentran en la segunda mitad del período.
- **KgCal**: La cantidad mínima de kilogramos por calibre es 10.0 y la máxima es 17680.0. Esto indica una amplia variabilidad en la producción de calibres de paltas. La media de 2707.3 kg sugiere que la producción promedio se encuentra en la parte superior del rango.
- **Perc_Cal**: El porcentaje mínimo por calibre es 0.0002174 y el máximo es 1.0000000. Esto indica que la proporción de cada calibre en la producción total varía considerablemente. La media de 0.1461538

sugiere que, en promedio, los calibres representan el 14.6% de la producción total.

IV. Visualización de Datos

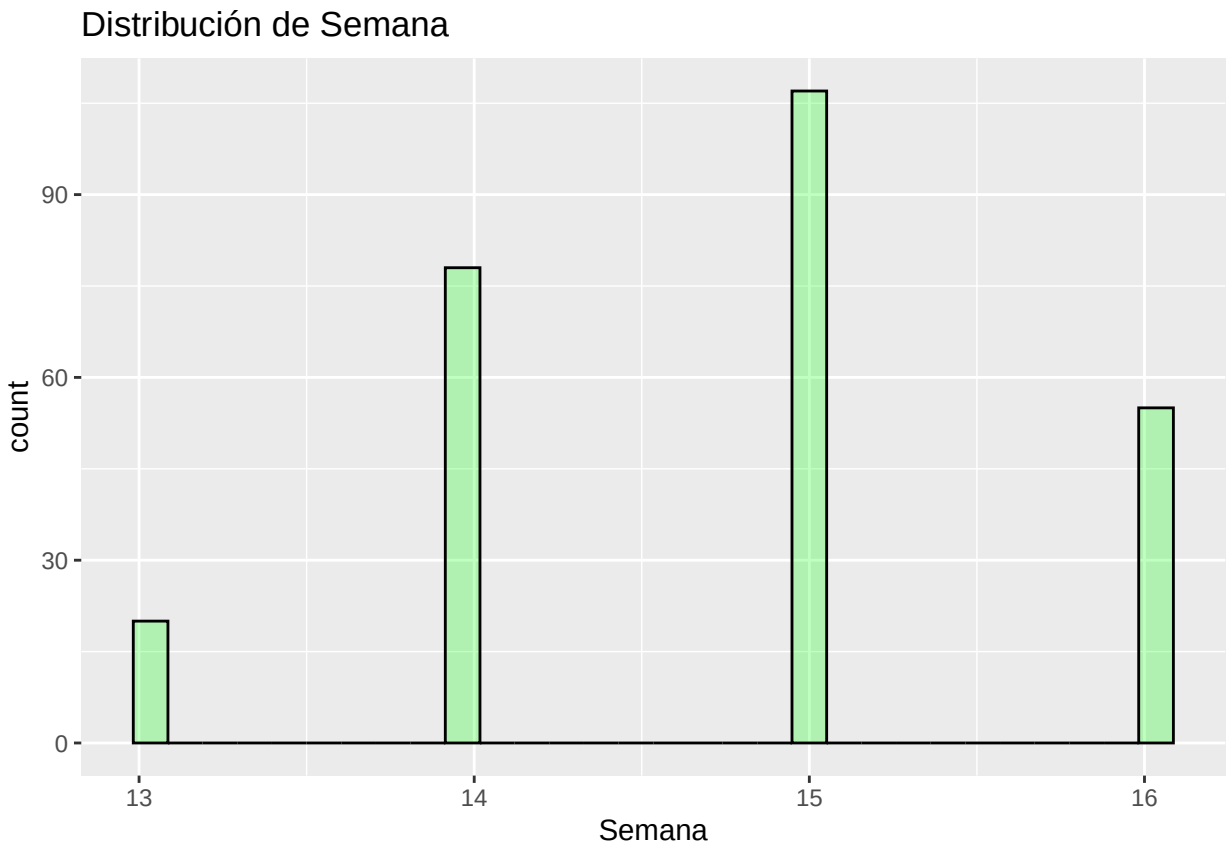
4.1. Visualización de Variables Numéricas

Las visualizaciones juegan un papel crucial en la interpretación de los datos. Se generaron histogramas para observar la distribución de las semanas, los kilogramos por calibre, y el porcentaje por calibre. Estos gráficos revelan cómo se comportan los datos a lo largo de las distintas semanas del periodo estudiado y proporcionan una visión directa de la variabilidad en la producción.

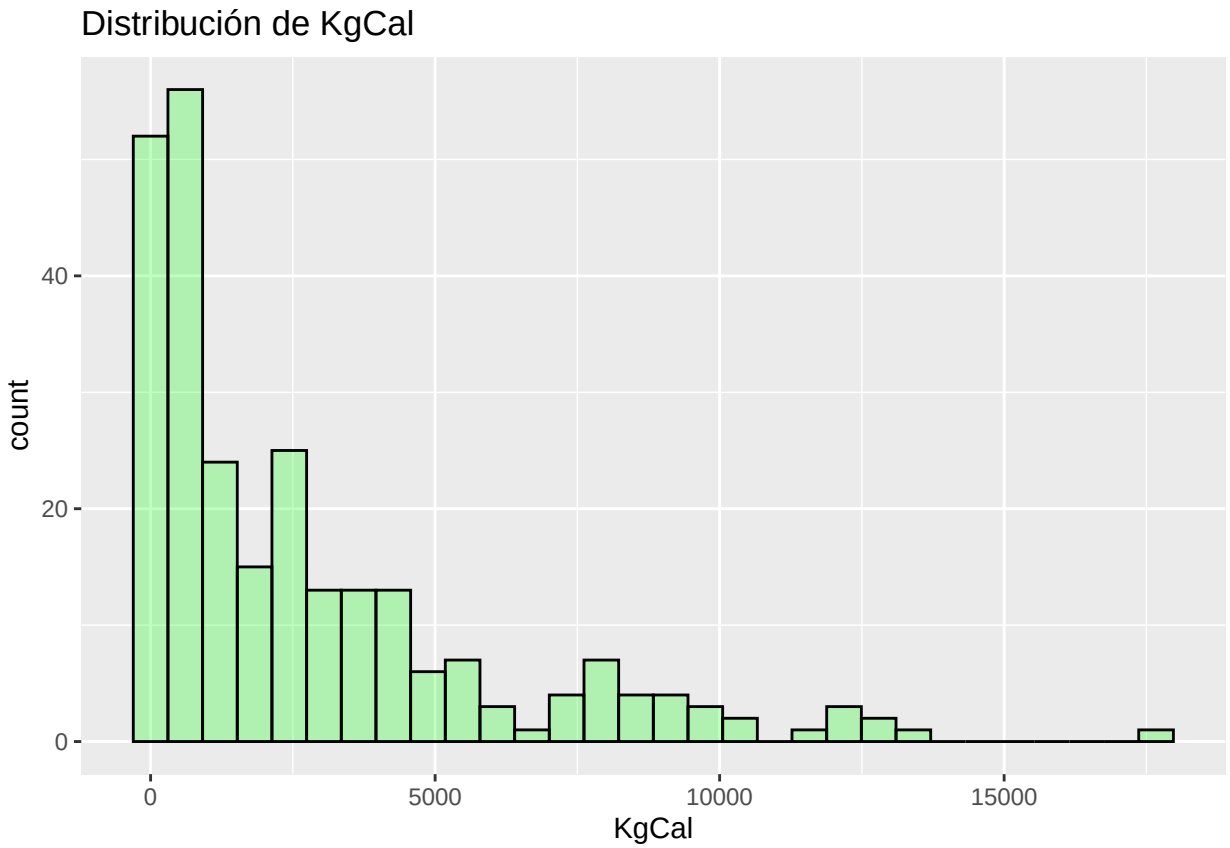
```
# Generando histogramas
lapply(names(data_n), function(x) {
  ggplot(data, aes_string(x = x)) +
    geom_histogram(alpha=0.25, bins = 30, fill = "green", color = "black") +
    labs(title = paste("Distribución de", x))
})
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## i Please use tidy evaluation idioms with `aes()`.
## i See also `vignette("ggplot2-in-packages")` for more information.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

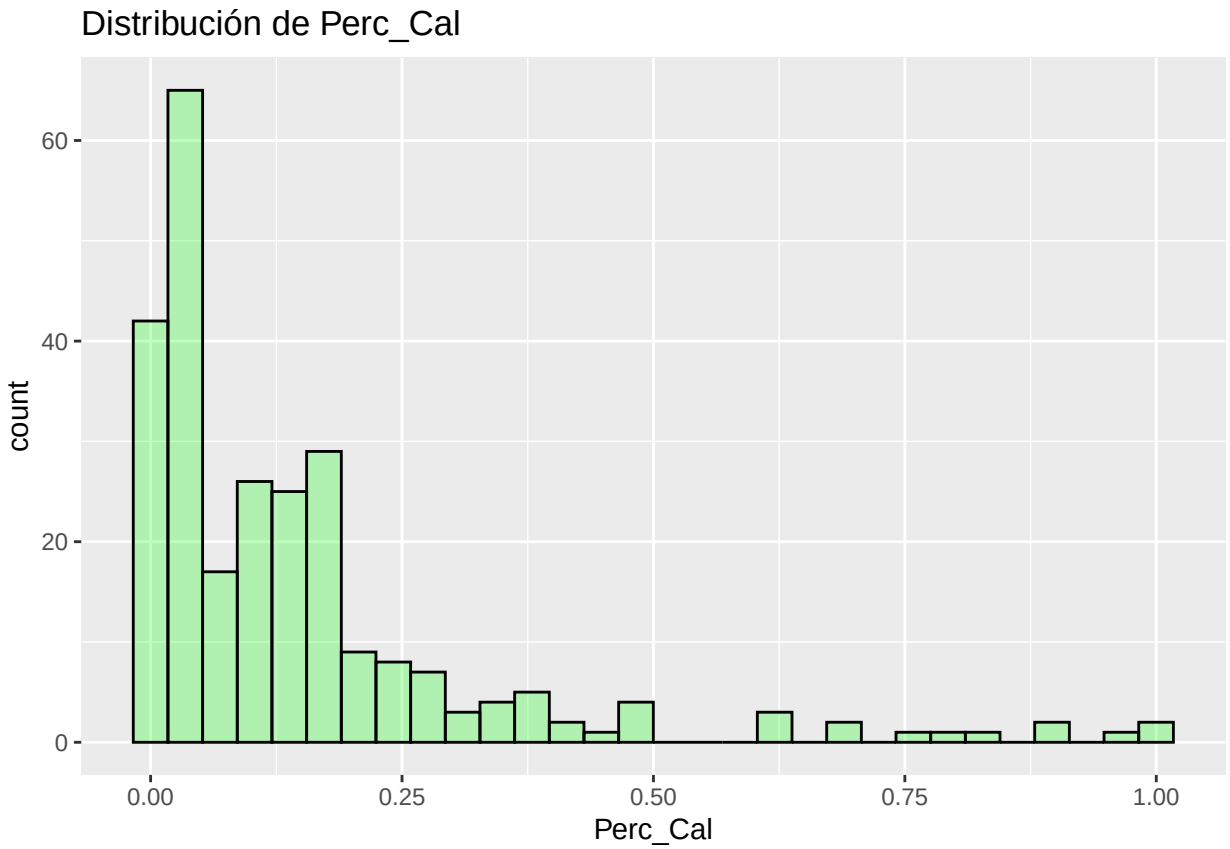
```
## [[1]]
```



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




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



4.1.1. Distribución de la Semana

- **Forma de la distribución**: La distribución de la semana presenta una tendencia hacia la derecha, con una mayor concentración de observaciones en las semanas 14 y 15. Esto indica que la mayor parte de la producción de paltas se registró en estas dos semanas.
- **Media y mediana**: La media de la distribución se encuentra alrededor de la semana 14.76, mientras que la mediana se ubica en la semana 15. Esta diferencia entre la media y la mediana sugiere que la distribución está ligeramente sesgada hacia la derecha.ç
- **Variabilidad**: La distribución muestra una variabilidad moderada en la semana de producción. Las semanas van desde la 13 hasta la 16, con una dispersión moderada entre las observaciones.
- **Valores atípicos**: Es poco probable que existan valores atípicos en la distribución, ya que las semanas están representadas por números enteros.

Interpretación

- La concentración de la producción en las semanas 14 y 15 podría estar relacionada con factores climáticos o de maduración de las paltas.
- Es importante analizar la distribución de la semana en conjunto con otras variables, como KgCal o Perc_Cal, para identificar posibles relaciones.

4.1.2. Distribución de KgCal

- **Forma de la distribución**: La distribución de KgCal presenta una asimetría positiva, con una cola más larga hacia los valores altos. Esto indica que una proporción significativa de las observaciones tiene una producción de KgCal por encima de la media.
- **Media y mediana**: La media de la distribución se encuentra aproximadamente en 2700 KgCal, mientras que la mediana se ubica alrededor de 1500 KgCal. Esta diferencia entre la media y la mediana

sugiere que la distribución está ligeramente sesgada hacia la derecha.
- **Variabilidad**: La distribución muestra una considerable variabilidad en la producción de KgCal. Los valores de KgCal van desde 10.0 hasta 17680.0, con una gran dispersión entre las observaciones.
- **Valores atípicos**: Es posible que existan valores atípicos en la distribución, es decir, observaciones que se alejan significativamente del resto de los datos. Estos valores podrían afectar la interpretación del análisis.

Interpretación

- La variabilidad en la producción de KgCal podría estar relacionada con la variedad de calibres, las condiciones climáticas, las prácticas agrícolas o la madurez de la fruta.
- Es importante analizar la distribución de KgCal en conjunto con otras variables, como la semana o el calibre, para identificar posibles relaciones.
- Los valores atípicos podrían indicar errores en la medición o casos excepcionales que requieren un análisis más profundo.

4.1.3. Distribución de Perc_Cal

- **Forma de la distribución**: La distribución de Perc_Cal presenta una asimetría positiva, con una cola más larga hacia los valores altos. Esto indica que una proporción considerable de las observaciones tiene un porcentaje de calibre por encima de la media.
- **Media y mediana**: La media de la distribución se encuentra aproximadamente en 0.15, mientras que la mediana se ubica alrededor de 0.10. Esta diferencia entre la media y la mediana sugiere que la distribución está ligeramente sesgada hacia la derecha.
- **Variabilidad**: La distribución muestra una considerable variabilidad en el porcentaje por calibre. Los valores de Perc_Cal van desde 0.0002 hasta 1.0000, con una gran dispersión entre las observaciones.
- **Valores atípicos**: Es posible que existan valores atípicos en la distribución, es decir, observaciones que se alejan significativamente del resto de los datos. Estos valores podrían afectar la interpretación del análisis.

Interpretación

- La variabilidad en el porcentaje por calibre podría estar relacionada con la variedad de calibres, las condiciones climáticas, las prácticas agrícolas o la madurez de la fruta.
- Es importante analizar la distribución de Perc_Cal en conjunto con otras variables, como la semana o el calibre, para identificar posibles relaciones.
- Los valores atípicos podrían indicar errores en la medición o casos excepcionales que requieren un análisis más profundo.

```
# Conjunto de datos
names(data)
```

```
##  [1] "Semana"    "Mes"       "Variedad"  "Color"     "Calibre"   "KgCal"
##  [7] "Perc_Cal"  "NoGR"      "NoRepProd" "Fecha"     "Cliente"
```

```
# Subconjunto de datos
names(data_n)
```

```
## [1] "Semana"   "KgCal"    "Perc_Cal"
```

V. Visualización Avanzada con Boxplots

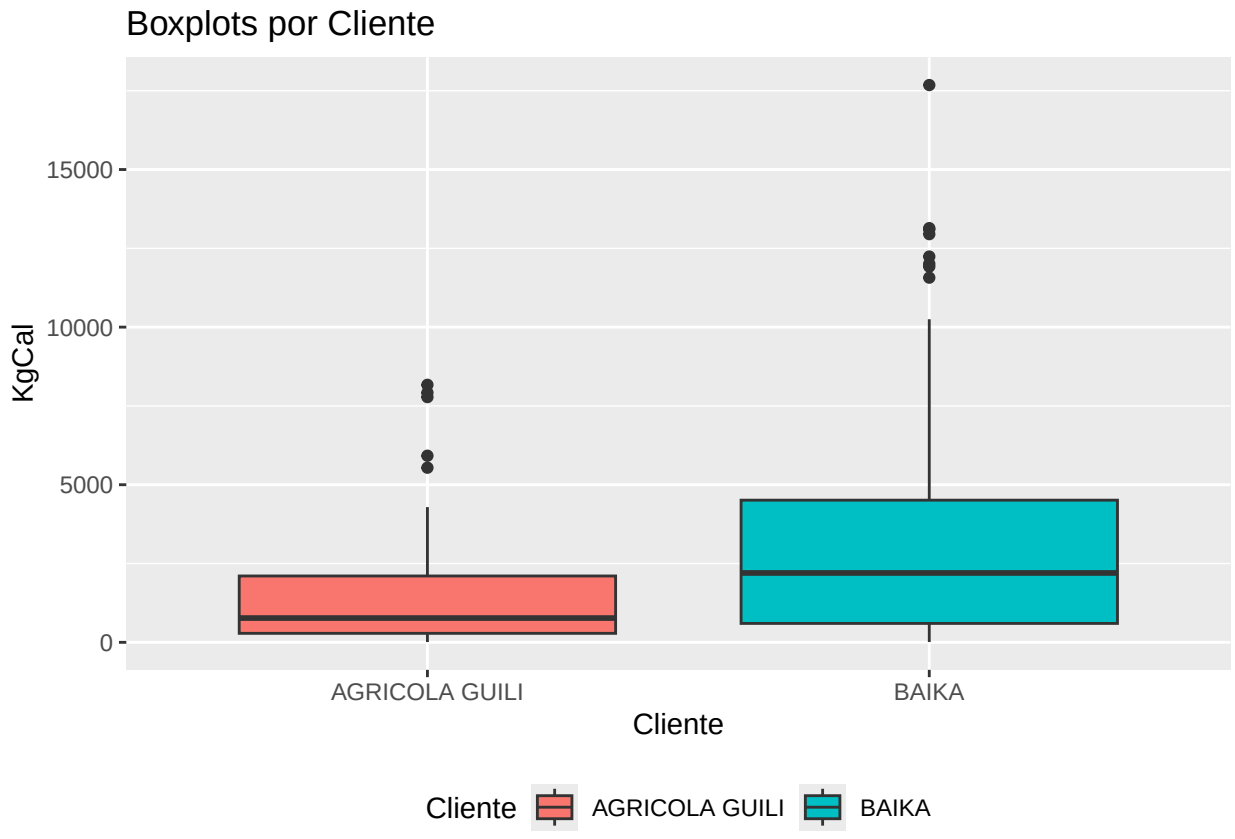
Para comparar la variabilidad de los kilogramos por calibre entre diferentes clientes y calibres, se elaboraron boxplots. Estos gráficos destacan las medianas, los rangos intercuartílicos y los valores atípicos, facilitando la comparación entre grupos. Por ejemplo, los boxplots por cliente y por calibre muestran cómo difieren

las distribuciones, lo cual es esencial para identificar clientes o calibres con comportamientos de producción distintos.

5.1. Analizando la Producción de KgCal por Cliente con Boxplots

Este código crea un boxplot para visualizar la distribución de kilogramos por calibre (KgCal) en función del cliente (Cliente) en el conjunto de datos. El boxplot permite comparar la producción de KgCal entre diferentes clientes de manera gráfica y eficiente.

```
# Configuraciçon del tamaño del gráfico
options(repr.plot.width=18, repr.plot.height=6)

# Creación del boxplot
boxplot <- ggplot(data, aes(x = Cliente, y = KgCal, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Cliente")

# Estableciendo la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")

# Mostrando el boxplot
print(boxplot)
```

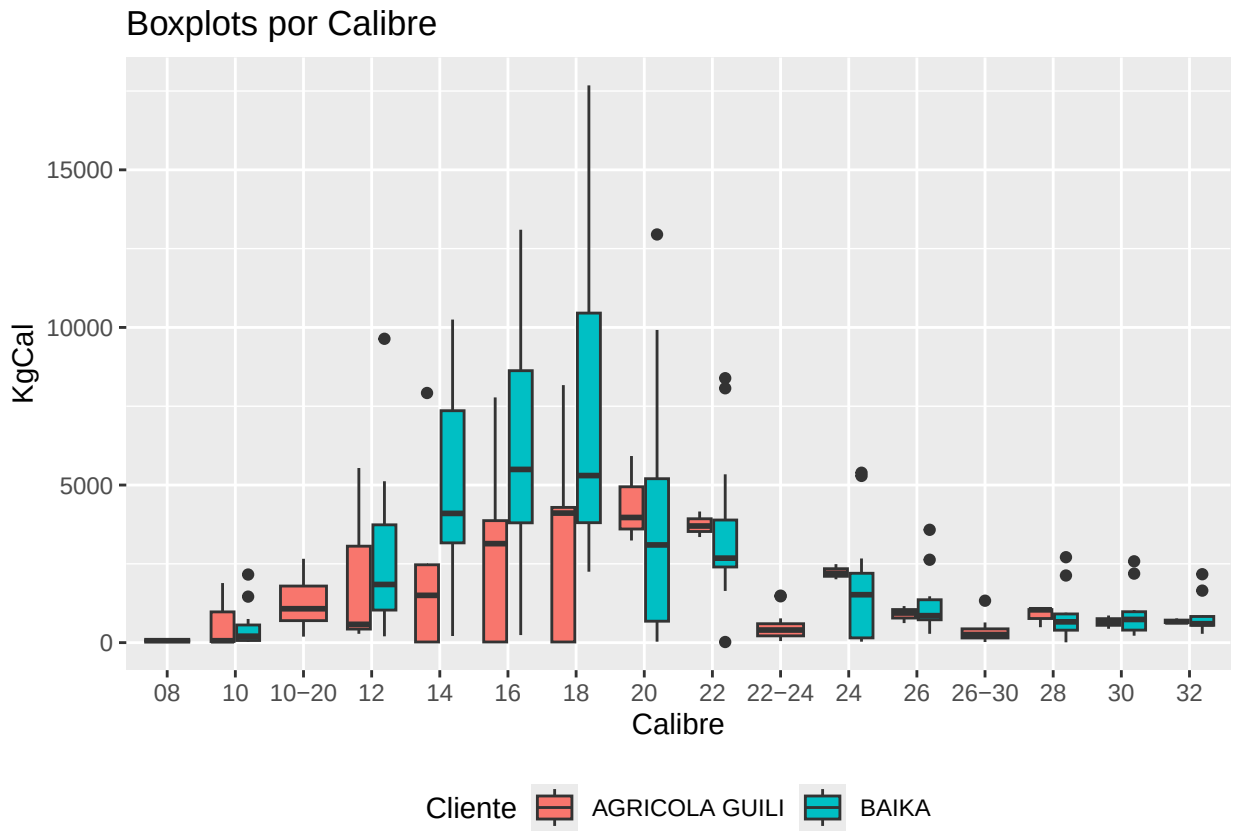


5.2. Analizando la Distribución de KgCal por Calibre con Boxplots

Este código crea un boxplot para visualizar la distribución de kilogramos por calibre (KgCal) en función del calibre (Calibre) en el conjunto de datos. El boxplot permite comparar la producción de KgCal entre

diferentes calibres de paltas de manera gráfica y eficiente.

```
# Creando el boxplot
boxplot <- ggplot(data, aes(x = Calibre, y = KgCal, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Calibre")

# Estableciendo la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")

# Mostrando el boxplot
print(boxplot)
```

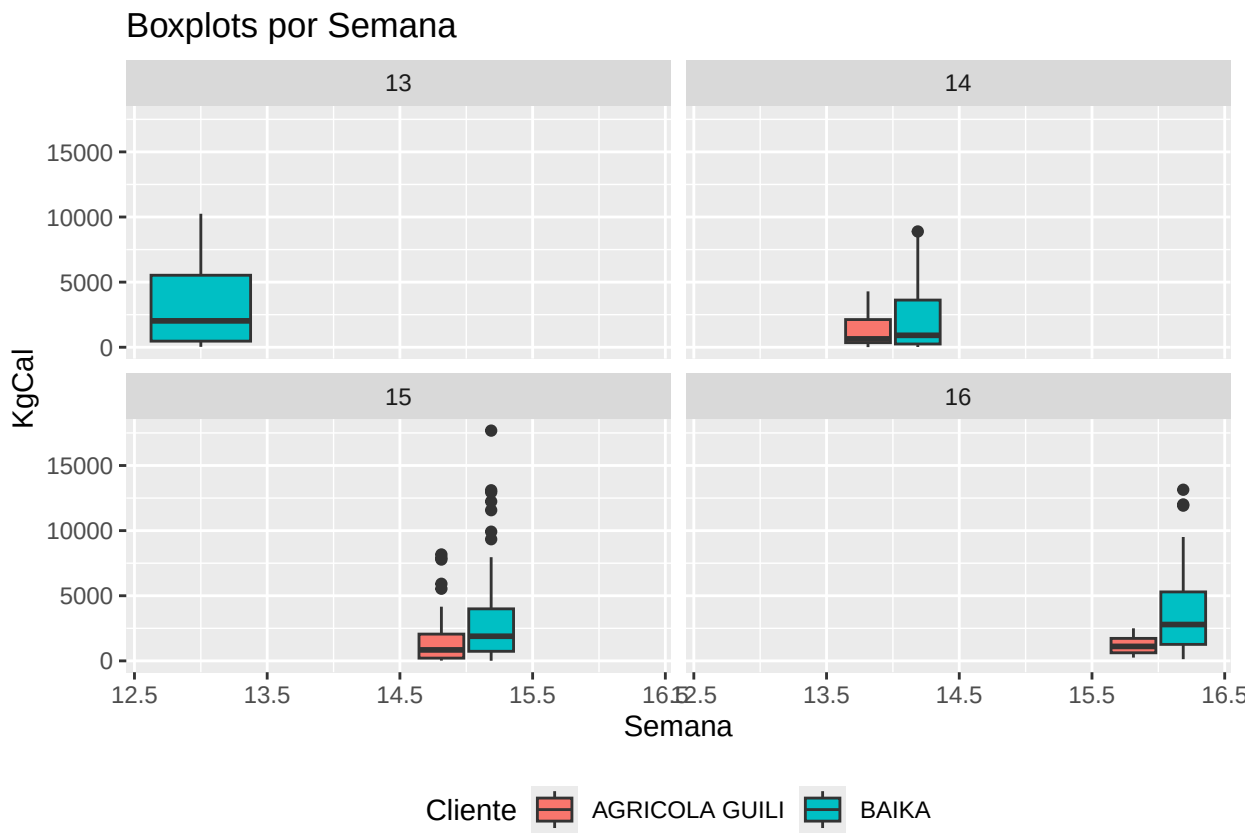


5.3. Analizando la Distribución de KgCal por Semana con Boxplots Faceteados

Este código crea un conjunto de boxplots para visualizar la distribución de kilogramos por calibre (KgCal) a lo largo de diferentes semanas (Semana) y clientes (Cliente) en el conjunto de datos. Los boxplots están faceteados por Semana, lo que permite una comparación más detallada de la distribución de KgCal dentro de cada semana y entre distintos clientes.

```
#1 Crear los boxplots
boxplot <- ggplot(data, aes(x = Semana, y = KgCal, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Semana") +
  facet_wrap(~Semana)
# Establecer la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")
```

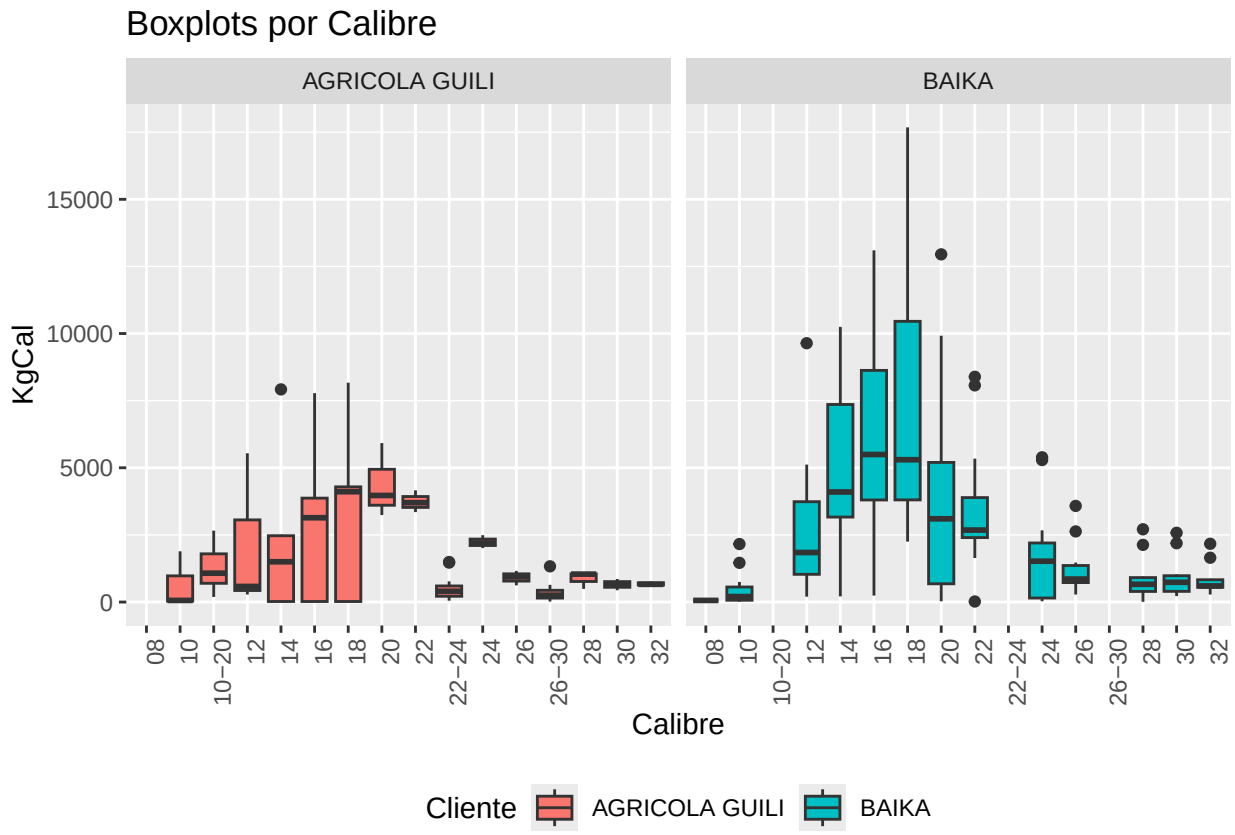
```r
# Mostrar los boxplots
print(boxplot)
```



5.4. Analizando la Distribución de KgCal por Calibre con Boxplots Faceteados y Etiquetas Giradas

Este código crea un conjunto de boxplots faceteados para visualizar la distribución de kilogramos por calibre (KgCal) a lo largo de diferentes calibres (Calibre) y clientes (Cliente) en el conjunto de datos. Los boxplots están faceteados por Cliente, lo que permite una comparación más detallada de la distribución de KgCal dentro de cada cliente y entre diferentes calibres. Adicionalmente, las etiquetas del eje X están rotadas 90 grados para una mejor legibilidad.

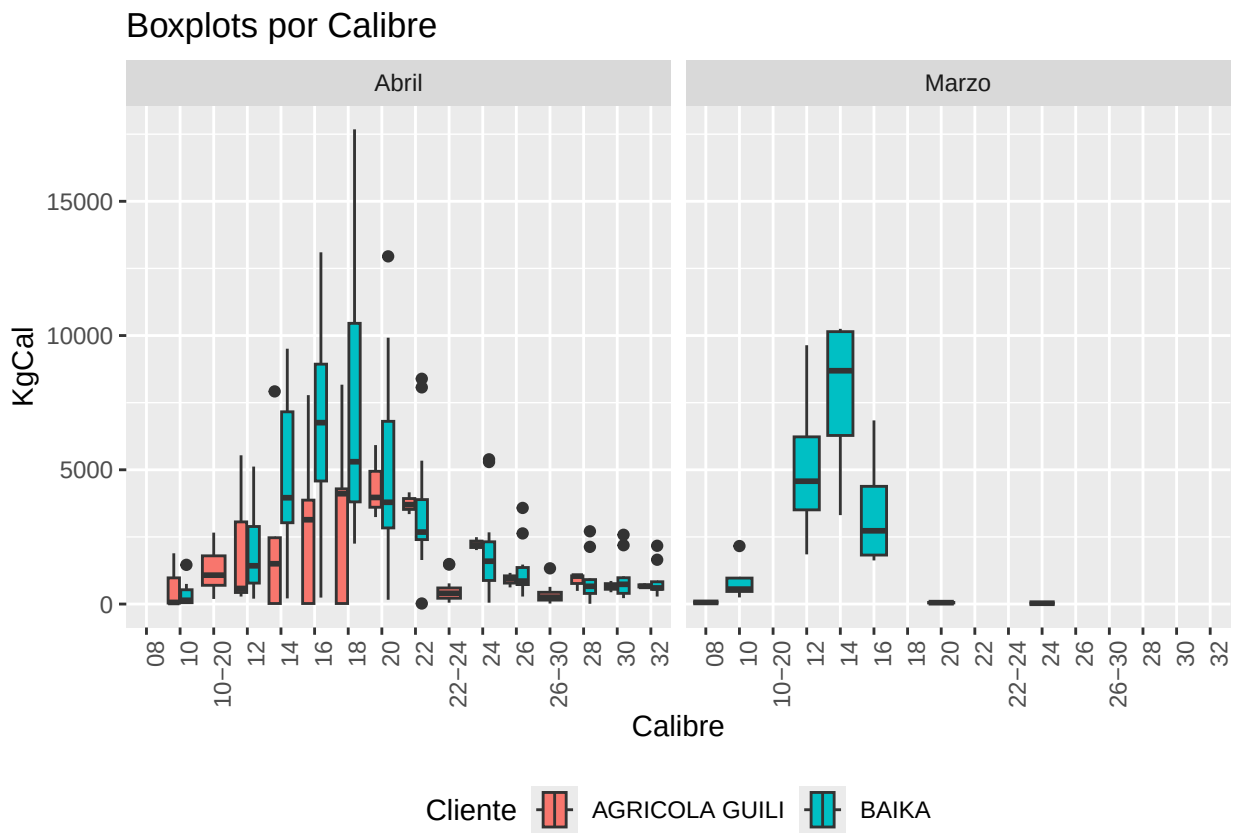
```r
#2 Crear los boxplots
options(repr.plot.width=15, repr.plot.height=5) # Tamaño original
boxplot <- ggplot(data, aes(x = Calibre, y = KgCal, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Calibre") +
  facet_wrap(~Cliente)+# Cambia "Otra_Variable" por el nombre de la variable que deseas usar para divid
  theme(axis.text.x = element_text(angle = 90, hjust = 1))  # Girar los labels del eje x 90 grados
# Establecer la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")
# Mostrar los boxplots
print(boxplot)
```



5.5. Explorando la Distribución de Calibre por KgCal con Boxplots Faceteados y Gráfico Rotado

Este código genera un conjunto de boxplots faceteados para visualizar la distribución de calibres (Calibre) a lo largo de diferentes kilogramos por calibre (KgCal) y meses (Mes) en el conjunto de datos. Los boxplots están faceteados por Mes, permitiendo una comparación más detallada de la distribución del Calibre dentro de cada mes y entre diferentes KgCal.

```
#3 Crear los boxplots
options(repr.plot.width=15, repr.plot.height=5) # Tamaño original
boxplot <- ggplot(data, aes(y = Calibre, x = KgCal, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Calibre") +
  facet_wrap(~Mes) + # Cambia "Otra_Variable" por el nombre de la variable que deseas usar para dividir
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +  # Girar los labels del eje x 90 grados
  coord_flip() # Esto gira el gráfico para que los boxplots sean horizontales
# Establecer la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")
# Mostrar los boxplots
print(boxplot)
```



VI. Matriz de Correlación

Se calculó una matriz de correlación para examinar las relaciones lineales entre las variables numéricas.

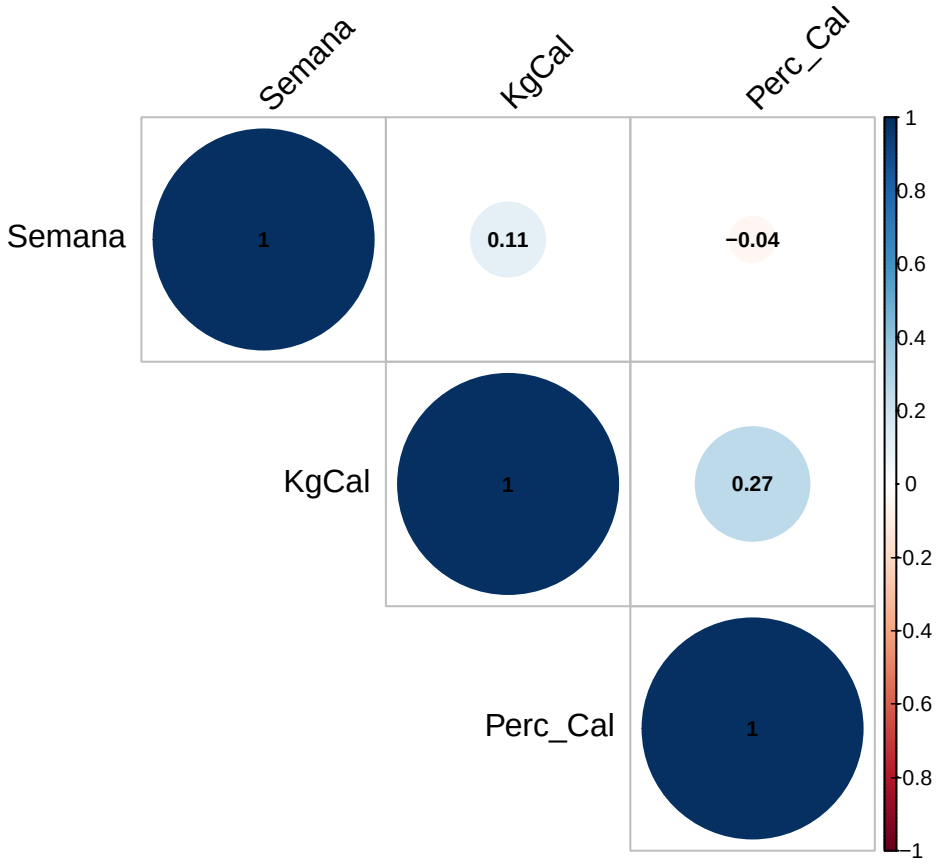
Este código calcula y visualiza una matriz de correlación para las variables numéricas en el conjunto de datos data_n. La matriz de correlación permite identificar relaciones lineales entre pares de variables.

```
# Matriz de correlación (si aplica)
# Instalar el paquete 'corrplot' si aún no está instalado
if (!require("corrplot")) install.packages("corrplot")
```

```
## Loading required package: corrplot
```

```
## corrplot 0.92 loaded
```

```
library(corrplot)
correlation_matrix <- cor(data_n, use = "pairwise.complete.obs")
corrplot(correlation_matrix, method = "circle", type = "upper", #order = "hclust",
         tl.col = "black",  # color del texto de la etiqueta
         tl.srt = 45,       # rotación de la etiqueta del texto en grados
         addCoef.col = "black", # color de los coeficientes de correlación
         number.cex = 0.7,  # tamaño de los coeficientes de correlación
         cl.cex = 0.7,      # tamaño del texto de la leyenda de colores
         cl.ratio = 0.1     # proporción del tamaño de la leyenda de colores
)
```




```r
colnames(data)
```

```
##  [1] "Semana"    "Mes"       "Variedad"  "Color"     "Calibre"   "KgCal"
##  [7] "Perc_Cal"  "NoGR"      "NoRepProd" "Fecha"     "Cliente"
```

VII. Análisis Visual Avanzado de Calibre vs KgCal con Facetas y Líneas

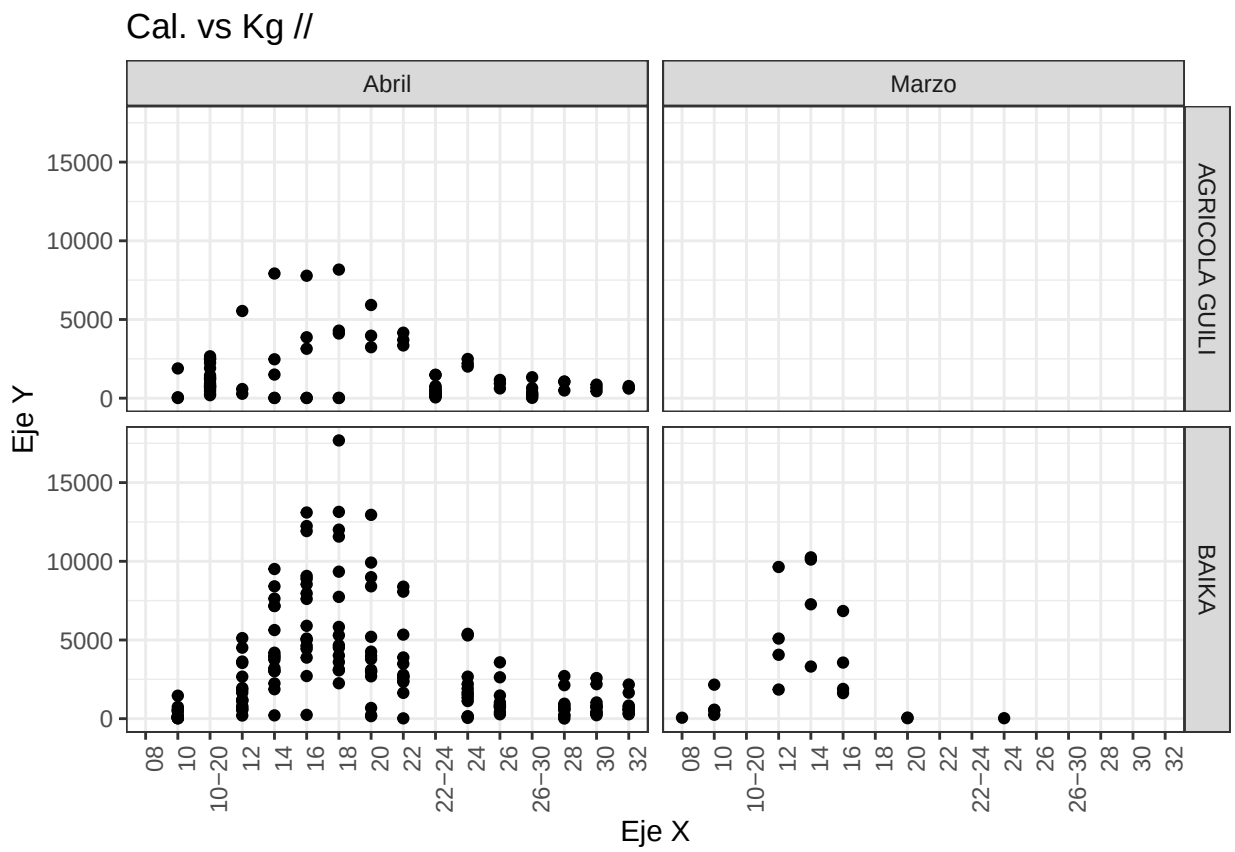
Esta sección complementa el análisis de la distribución de Calibre y KgCal con dos nuevas visualizaciones: un gráfico de dispersión faceteado y un gráfico de líneas. Estas visualizaciones adicionales permiten explorar las relaciones entre Calibre y KgCal con mayor detalle, considerando factores como el Cliente y el Mes.

7.1. Gráfico de Dispersión Faceteado

Este gráfico utiliza la función ggplot() para crear un diagrama de dispersión donde los puntos de datos se agrupan por Cliente y Mes. Esto permite observar cómo la distribución de Calibre y KgCal varía entre diferentes clientes y meses.

```r
### FACET #
ggplot(data, aes(x = Calibre, y = KgCal)) +
  geom_point() +
  facet_grid(Cliente ~ Mes) +
  labs(title = "Cal. vs Kg // ",
       x = "Eje X",
       y = "Eje Y") +
  theme_bw() +
```

```
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))  # Girar los labels del eje x 90 grados
```




```
# Ahora exportar después de cambiar el directorio
write.csv(data, "my_datac.csv", row.names = FALSE)

# Lineas
data$Cliente <- as.factor(data$Cliente)
data$Calibre <- as.numeric(data$Calibre)
```

```
## Warning: NAs introducidos por coerción
```

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

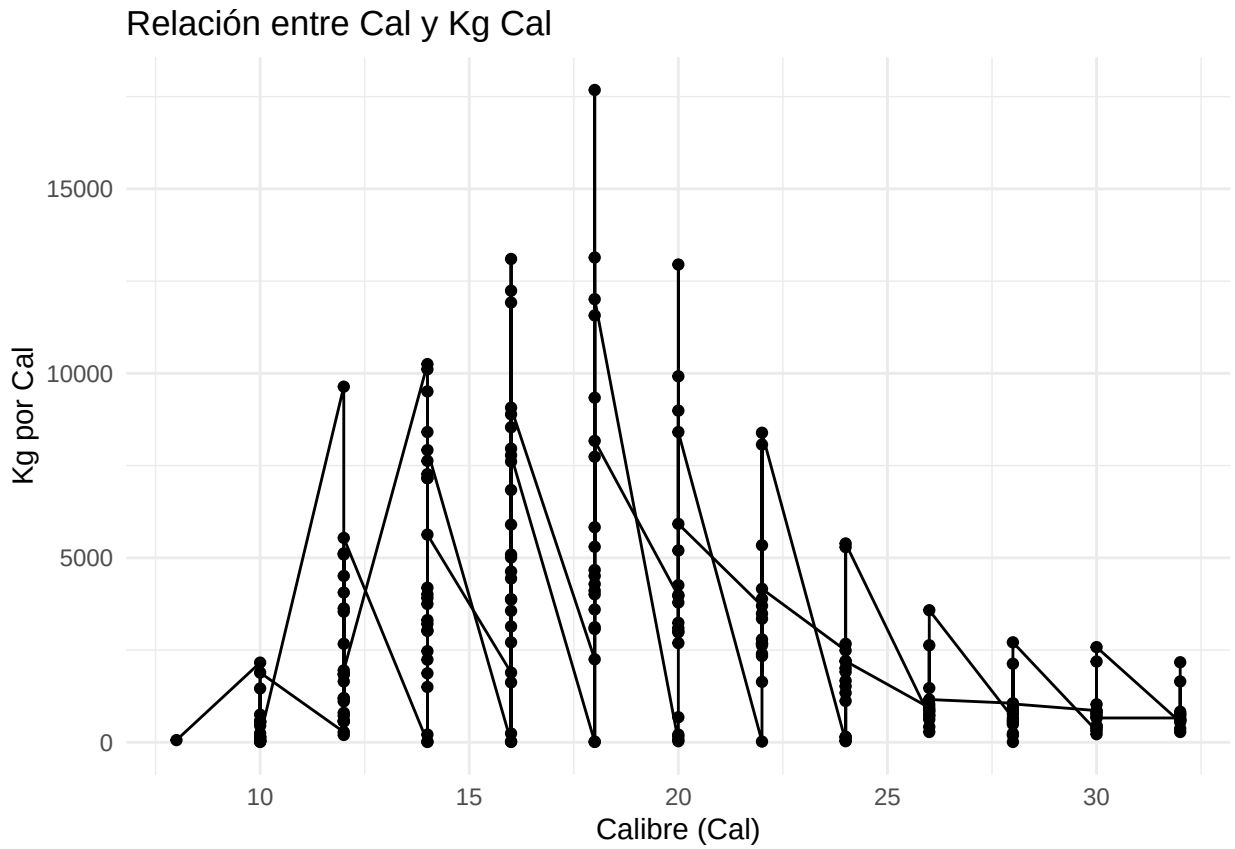
7.2. Gráfico de Líneas:

Este gráfico utiliza la función ggplot() para crear un gráfico de líneas que muestra la relación entre Calibre y KgCal, diferenciando por Cliente. Los puntos se agregan para mejorar la visualización de los datos.

```
# Crear el gráfico de líneas
ggplot(data, aes(x=Calibre, y=KgCal, colr=Cliente, group=Cliente)) +
  geom_line() +
  geom_point() +  # Agregar puntos para mejor visualización de los datos
  labs(title="Relación entre Cal y Kg Cal",
       x="Calibre (Cal)", y="Kg por Cal") +
  scale_color_viridis_d(option = "inferno") +  # Usar una paleta de colores para diferenciar los Fundos
  theme_minimal()  # Tema minimalista
```

```
## Warning: Removed 41 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

```
## Warning: Removed 41 rows containing missing values or values outside the scale range
## (`geom_point()`).
```




```
#

names(data)
```

```
##  [1] "Semana"    "Mes"       "Variedad"  "Color"     "Calibre"   "KgCal"
##  [7] "Perc_Cal"  "NoGR"      "NoRepProd" "Fecha"     "Cliente"
```

```
names(data_n)
```

```
## [1] "Semana"   "KgCal"    "Perc_Cal"
```

VIII. Análisis de varianza (ANOVA)

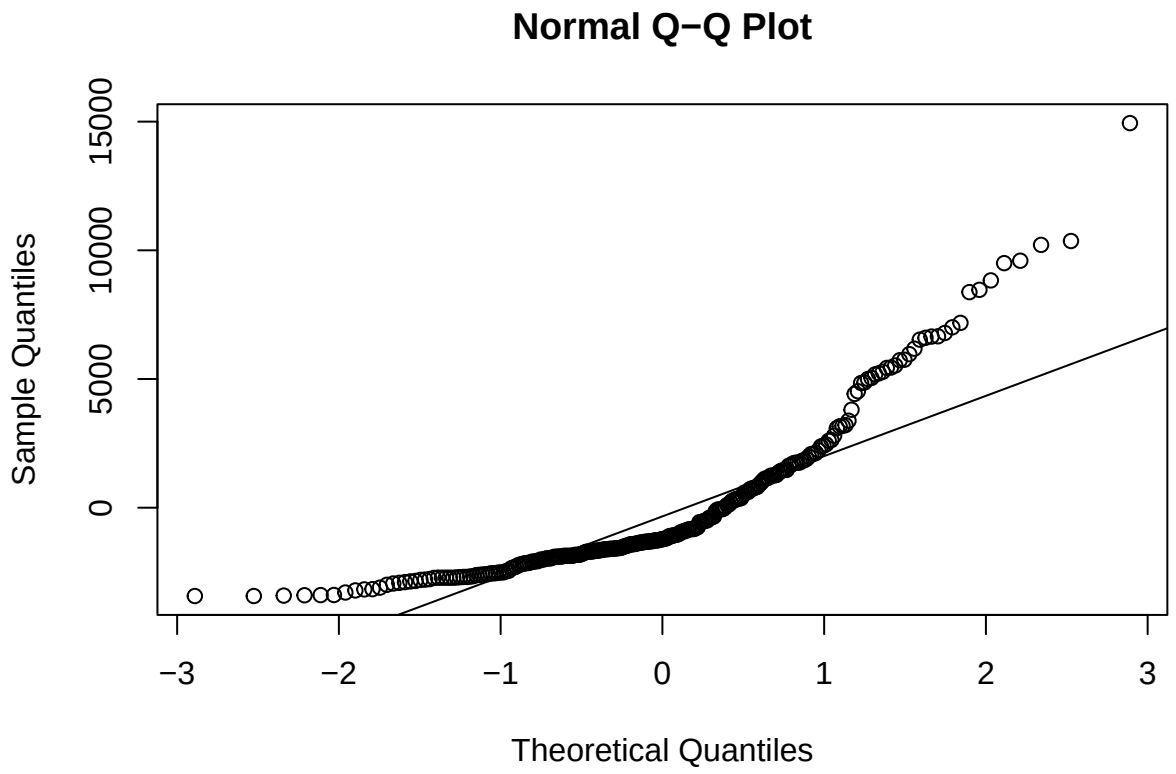
Se realizó un análisis de varianza (ANOVA) para determinar si existen diferencias estadísticamente significativas en los kilogramos por calibre entre las diferentes semanas. Los resultados del ANOVA indicaron que hay variaciones significativas, lo cual sugiere que la semana influye en la cantidad de producción.

```
### ANOVA ###
# Ensure that 'Fundo' is a factor and 'KgExpxHa' is numeric
data$Semana <- as.factor(data$Semana)
data$KgCal <- as.numeric(data$KgCal)
# ANOVA to compare 'KgExpxHa' across different 'Fundo'
```

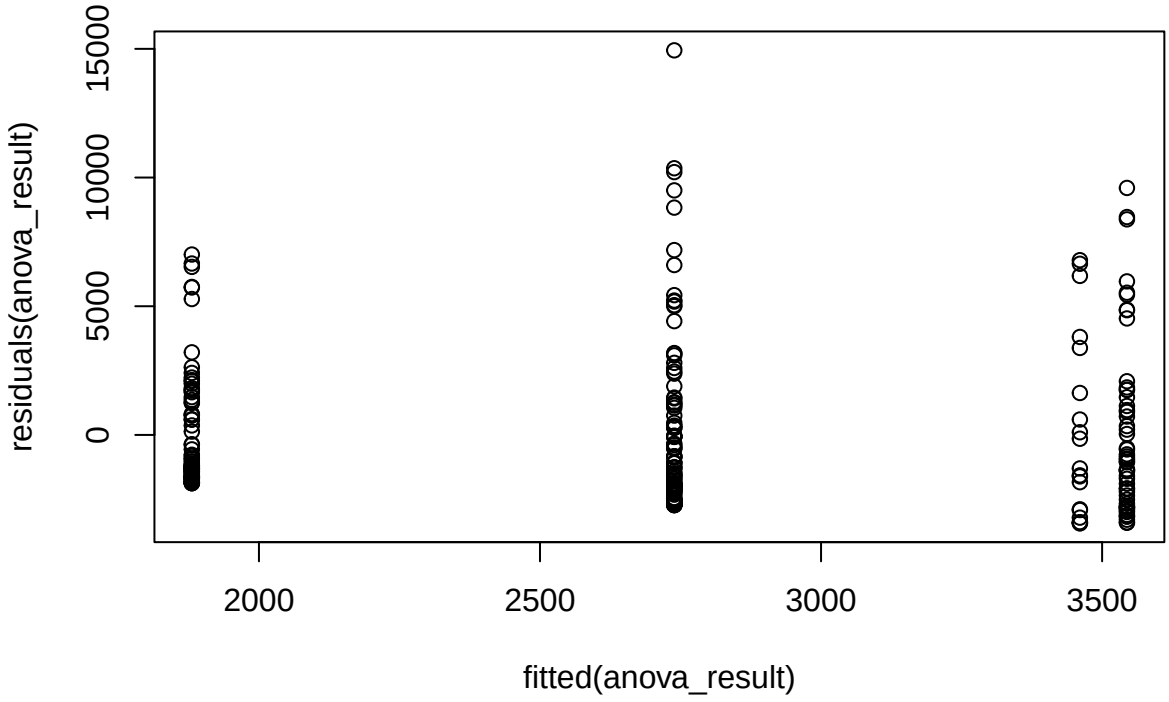
```r
anova_result <- aov(KgCal ~ Semana, data = data)
# Check the summary of the ANOVA
summary(anova_result)
```

```
##              Df    Sum Sq  Mean Sq F value Pr(>F)
## Semana        3 1.033e+08 34425265   3.615 0.0138 *
## Residuals   256 2.438e+09  9523384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Check for assumptions: Normality
qqnorm(residuals(anova_result))
qqline(residuals(anova_result))
```



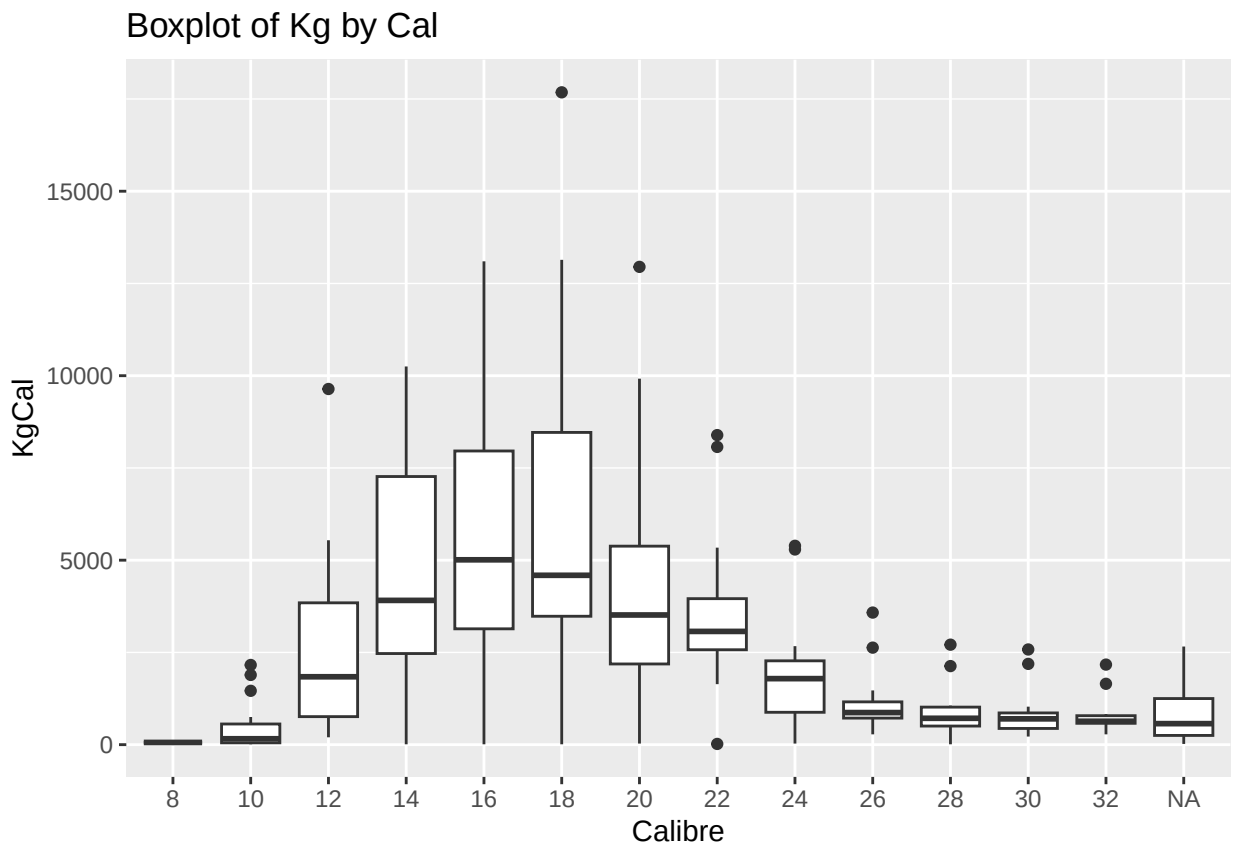

```r
# Homogeneity of variances
plot(residuals(anova_result) ~ fitted(anova_result))
```




```
# If ANOVA is significant, conduct post-hoc tests
TukeyHSD(anova_result)
```

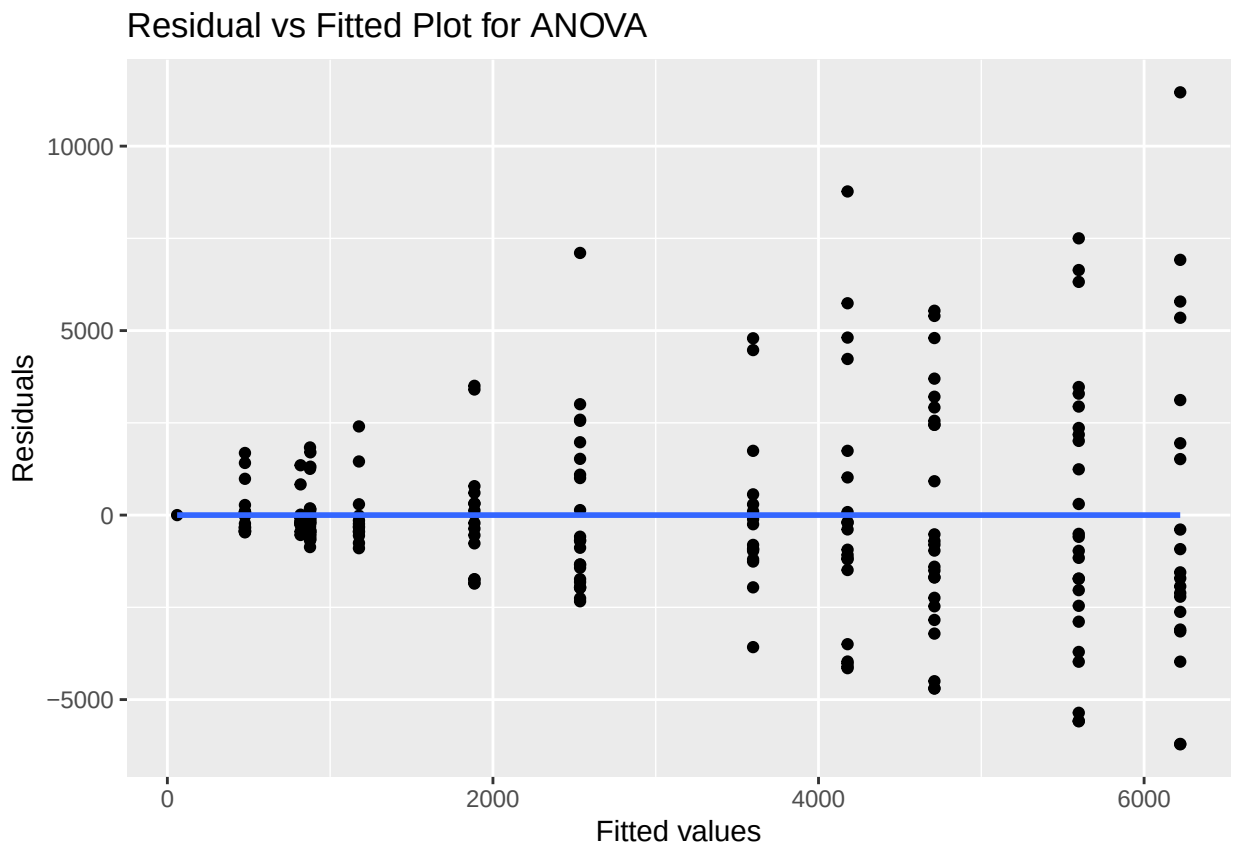
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = KgCal ~ Semana, data = data)
##
## $Semana
##              diff        lwr       upr     p adj
## 14-13 -1579.85897 -3580.0541  420.3362 0.1752378
## 15-13  -721.62150 -2665.7126 1222.4696 0.7723158
## 16-13    83.68182 -2000.1177 2167.4813 0.9995982
## 15-14   858.23748  -329.9044 2046.3794 0.2445009
## 16-14  1663.54079   258.4033 3068.6782 0.0129406
## 16-15   805.30331  -518.7513 2129.3580 0.3959995
```

```
## ANOVA #2
# Convert 'Fundo' to a factor if necessary
data$Calibre <- as.factor(data$Calibre)

# Boxplot for 'KgExpxHa' across different 'Fundo'
ggplot(data, aes(x=Calibre, y=KgCal)) +
  geom_boxplot() +
  labs(title="Boxplot of Kg by Cal", x="Calibre", y="KgCal")
```

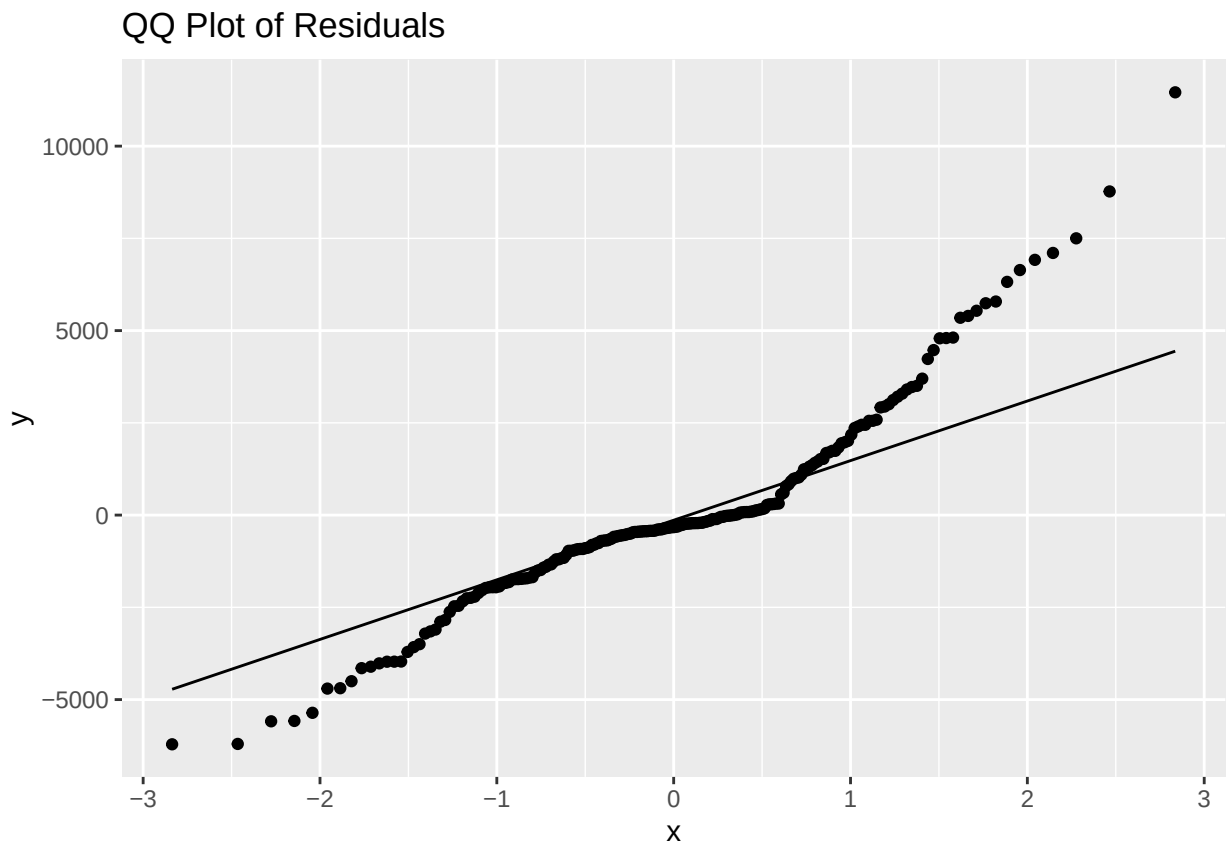



```
# Performing the ANOVA
anova_result <- aov(KgCal ~ Calibre, data = data)

# Residual plot
res_data <- data.frame(residuals= residuals(anova_result), fitted=fitted(anova_result))
ggplot(res_data, aes(x=fitted, y=residuals)) +
  geom_point() +
  geom_smooth(method="lm", se=FALSE) +
  labs(title="Residual vs Fitted Plot for ANOVA", x="Fitted values", y="Residuals")
```

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




```
# QQ plot of residuals
ggplot(res_data, aes(sample=residuals)) +
  geom_qq() +
  geom_qq_line() +
  labs(title="QQ Plot of Residuals")
```




```
###################################
```

IX. Conclusiones

Este estudio exhaustivo analiza la producción y comercialización de paltas por parte de diversos clientes durante un período específico. El objetivo principal es comprender las características y patrones de la producción de calibres de paltas para identificar áreas de mejora y optimización. El análisis se basa en un conjunto de datos completo que incluye información sobre la semana, el mes, la variedad de palta, el color, el calibre específico y la cantidad en kilogramos.

Análisis Exploratorio de Datos

- Se exploró la estructura y el contenido del conjunto de datos, identificando variables numéricas y categóricas.
- Se renombraron algunas columnas para mejorar la claridad y facilitar el manejo de las variables.
- Se verificó la presencia de datos faltantes, encontrando una mínima cantidad que no afecta el análisis.
- Se exploraron las distribuciones de las variables numéricas mediante histogramas y resúmenes estadísticos.

Visualización de Datos

- Se generaron histogramas para observar la distribución de las semanas, los kilogramos por calibre y el porcentaje por calibre.
- Se identificaron patrones como la concentración de la producción en las semanas 14 y 15, la asimetría positiva en las distribuciones de KgCal y Perc_Cal, y la variabilidad en la producción según el calibre.
- Se crearon boxplots para comparar la producción de KgCal por cliente, por calibre, por semana y por cliente-semana.
- Los boxplots revelaron diferencias significativas en la distribución de KgCal entre clientes, calibres y semanas.

- Se observó una interacción entre el cliente y la semana, con patrones de producción distintos para cada combinación.
- Se utilizaron boxplots faceteados para analizar la distribución de KgCal por calibre y cliente, identificando variaciones en la producción dentro de cada cliente y entre calibres.
- Se emplearon gráficos de dispersión faceteados para visualizar la relación entre Calibre y KgCal, considerando el cliente y el mes.
- Se observaron patrones de dispersión y agrupamiento específicos por cliente y mes, revelando información sobre la relación entre Calibre y KgCal en diferentes contextos.
- Se crearon gráficos de líneas para mostrar la tendencia de KgCal a medida que aumenta el Calibre, diferenciando por cliente.
- Las pendientes y formas de las líneas proporcionaron información sobre las diferencias en la relación entre Calibre y KgCal entre clientes.

Análisis de Varianza (ANOVA)

- Se realizó un análisis ANOVA para determinar si existen diferencias estadísticamente significativas en los kilogramos por calibre entre las diferentes semanas. Los resultados del ANOVA indicaron que hay variaciones significativas, lo cual sugiere que la semana influye en la cantidad de producción.

Insights

- La producción de calibres de paltas presenta una variabilidad considerable en términos de kilogramos por calibre, porcentaje por calibre y distribución a lo largo de las semanas.
- Se identificaron patrones de producción distintos para diferentes clientes, calibres, semanas y combinaciones - de cliente-semana.
- La semana influye significativamente en la cantidad de producción de calibres de paltas.
- La relación entre Calibre y KgCal varía entre clientes, con patrones de dispersión y agrupamiento específicos.
- Se observan diferencias en la pendiente y forma de las líneas de tendencia entre clientes, lo que indica - variaciones en la relación entre Calibre y KgCal.

X. Recomendaciones

- Considerar la variabilidad de la producción al realizar pronósticos y planificar la comercialización.
- Analizar las diferencias en la producción por cliente, calibre, semana y combinaciones de cliente-semana para identificar oportunidades de mejora y optimizar estrategias.
- Profundizar en el análisis de la relación entre Calibre y KgCal para cada cliente, considerando factores adicionales como la variedad de palta o las condiciones climáticas.
- Implementar sistemas de monitoreo y control de la producción para optimizar la calidad y cantidad de calibres de paltas.

XI. Repositorio

Para acceder al repositorio en GitHub ingresar por el siguiente enlace: https://github.com/Abel-Fernandez-D/PAC_metodos_estadisticos