

# Análisis Estadístico Descriptivo e Inferencial del Impacto de Condiciones Climáticas y Prácticas Agrícolas sobre el Calibre de las Paltas Hass

Grupo N°4 - Fernandez D. Abel, Fernandez D. Rody, Márquez M. Andre, Yactayo C. David

23 de abril de 2024

## Índice

<b>1. Introducción</b>	<b>1</b>
1.1. Contexto del Proyecto . . . . .	1
1.2. Importancia del Análisis de Calibres en la Producción de Paltas . . . . .	2
<b>2. Planteamiento del Problema</b>	<b>2</b>
2.1. Descripción del Problema . . . . .	2
2.2. Objetivos del Proyecto . . . . .	2
<b>3. Metodología</b>	<b>2</b>
3.1. Descripción de los Datos . . . . .	2
<b>4. Análisis Exploratorio de Datos (EDA)</b>	<b>4</b>
4.1. Carga y Limpieza de Datos . . . . .	4
4.2. Análisis Descriptivo General . . . . .	7
4.3. Análisis Específico por Cliente . . . . .	42
4.4. Matriz de correlación . . . . .	47
<b>5. Visualización Avanzada de Datos</b>	<b>48</b>
5.1. Visualización avanzada del dataset ‘calibres’ . . . . .	48
5.2. Visualización avanzada del dataset ‘reporte’ . . . . .	54
5.3. Gráficos de dispersión . . . . .	60
5.4. Comparativa de Densidad de Calibres entre CLIENTE1 y CLIENTE2 . . . . .	64
5.5. Mapas de calor . . . . .	66
<b>6. Análisis Inferencial</b>	<b>69</b>
6.1. Pruebas de Hipótesis . . . . .	69
6.2. Correlaciones y Regresiones Lineales . . . . .	88
<b>7. Resultados y Discusión</b>	<b>96</b>
7.1. Interpretación de los Resultados de las Pruebas Estadísticas . . . . .	96
7.2. Discusión sobre la Relación entre Calibres y Variables Climáticas/Tratamientos . . . . .	97
<b>8. Conclusiones y Recomendaciones</b>	<b>97</b>
8.1. Resumen de Hallazgos Clave . . . . .	97
8.2. Recomendaciones Basadas en el Análisis . . . . .	97

## 1. Introducción

### 1.1. Contexto del Proyecto

El cultivo de la palta Hass se ha convertido en un pilar económico para las regiones agrícolas que disfrutan del clima y suelo propicio para su desarrollo. Con su creciente demanda a nivel global, el análisis detallado de los patrones de

producción y calidad del fruto se hace indispensable en cualquier empresa agrícola dedicada a esta actividad. Este proyecto busca examinar el perfil de calibre de las paltas Hass a lo largo de cuatro semanas críticas de cosecha en una empresa agrícola peruana para identificar tendencias y correlaciones que puedan influir en su producción.

## 1.2. Importancia del Análisis de Calibres en la Producción de Paltas

El análisis de los calibres de paltas Hass es crucial para maximizar la rentabilidad y satisfacer los estándares del mercado. Un calibre uniforme es a menudo un indicador de calidad y homogeneidad del producto, factores que pueden afectar el precio y la preferencia del consumidor. Comprender cómo fluctúan los calibres en respuesta a variables ambientales y prácticas agrícolas puede conducir a una mejor gestión de los recursos y optimización de los procesos de cultivo y post-cosecha.

## 2. Planteamiento del Problema

### 2.1. Descripción del Problema

La producción de palta Hass enfrenta la complejidad de mantener la calidad frente a la variabilidad de factores ambientales y agronómicos. La heterogeneidad en los calibres de las paltas Hass afecta directamente la cadena de valor, desde la gestión de cultivos hasta la satisfacción del consumidor final. Identificar y comprender las variables que más influyen en la calidad y uniformidad del calibre de las paltas se convierte, por tanto, en una prioridad para optimizar la producción. ¿Cómo podemos analizar los patrones temporales y de calibre en la producción de paltas Hass a lo largo de un periodo de cuatro semanas para desarrollar recomendaciones que mejoren la gestión agrícola y se adapten eficazmente a las condiciones estacionales y prácticas de cultivo?

### 2.2. Objetivos del Proyecto

#### 2.2.1. Objetivo General

Analizar los patrones de calibre y producción temporal de las paltas Hass a lo largo de cuatro semanas críticas de cosecha para desarrollar un conjunto de recomendaciones dirigidas a optimizar la gestión agrícola y la adaptación a las variables estacionales y las prácticas de cultivo.

#### 2.2.2. Objetivos Específicos

- Investigar cómo los calibres de las paltas Hass varían en un marco temporal semanal y determinar si estos cambios están correlacionados con prácticas agrícolas específicas o con condiciones ambientales que se presentan durante las cuatro semanas.
- Analizar los datos de producción para determinar la existencia de tendencias o patrones en la producción y los calibres que puedan ser atribuidos a variaciones climáticas o a la aplicación de tratamientos específicos durante el periodo estudiado.
- Formular recomendaciones preliminares basadas en los resultados del análisis que puedan ser aplicadas en el corto plazo para ajustar y mejorar las prácticas de cultivo en respuesta a las observaciones de las condiciones y tratamientos de las semanas analizadas.

## 3. Metodología

### 3.1. Descripción de los Datos

Se utilizarán dos conjuntos de datos que contienen registros semanalizados de calibres de paltas Hass y reportes detallados de las condiciones de cultivo, incluyendo datos climáticos y de tratamientos aplicados.

#### 3.1.1. Diccionario de datos para el dataset ‘calibres’

Variable	Descripción
semana	Número de semana del año en la que se registraron los datos.
mes	Mes en el que se registraron los datos.

Variable	Descripción
variedad_producto	Tipo de palta haas.
color_producto	Color del producto registrado.
calibre	Tamaño del calibre del producto.
kg_calibre	Peso en kilogramos asociado al calibre del producto.
porcentaje_calibre_por_guia_remision	Porcentaje del calibre específico por guía de remisión.
guia_remision	Número de la guía de remisión.
reporte_produccion	Identificador del reporte de producción.
fecha	Fecha en la que se registró el dato.
cliente	Identificador del cliente que recibió el lote.

### 3.1.2. Diccionario de datos para el dataset ‘reportes’

Variable	Descripción
semana	Número de semana del año cuando se registraron los datos.
mes	Mes cuando se registraron los datos.
fundo	Identificador del fundo donde se producen las paltas.
empresa	Nombre de la empresa propietaria del fundo.
lote	Número o identificación del lote dentro del fundo.
hectareas	Área en hectáreas donde se cultivan las paltas.
variedad_producto	Variedad de la palta cultivada.
color_producto	Color de la palta registrada.
guia_remision	Número de la guía de remisión asociada a la producción.
reporte_produccion	Identificador del reporte de producción.
fecha	Fecha en que se registró el dato.
total_jabas	Cantidad total de cajas recolectadas.
peso_promedio_jaba	Peso promedio por caja de palta.
cajas_exportadas_10kg	Número de cajas de 10kg exportadas.
ingreso_packing_kg_bruto	Peso bruto total ingresado al packing.
kg_exportados	Kilogramos totales exportados.
porcentaje_exportados	Porcentaje de la producción exportada.
kg_descarte	Kilogramos totales descartados.
porcentaje_descarte	Porcentaje de la producción descartada.
kg_merma	Kilogramos de merma en la producción.
porcentaje_merma	Porcentaje de merma en la producción.
kg_descarte_campo	Kilogramos de producto descartado en campo.
kg_brutos_lote	Peso bruto por lote de producción.
kg_brutos_hectarea	Peso bruto por hectárea.
kg_exportados_hectarea	Kilogramos exportados por hectárea.
estado	Estado del lote o producción.
cliente	Identificador del cliente que recibe la producción.
temperatura_promedio	Temperatura promedio durante la semana de reporte.
temperatura_maxima	Temperatura máxima alcanzada durante la semana de reporte.
temperatura_minima	Temperatura mínima registrada durante la semana de reporte.
humedad_promedio	Humedad promedio durante la semana de reporte.
humedad_maxima	Humedad máxima alcanzada durante la semana de reporte.
humedad_minima	Humedad mínima registrada durante la semana de reporte.
evapotranspiracion_acumulada	Total de evapotranspiración acumulada durante la semana.
insumo_para_tratamiento	Insumo utilizado para el tratamiento de las paltas.
naturaleza_del_insumo	Clasificación del insumo utilizado (mineral, orgánica, química).
dosis_litros_por_hectarea	Cantidad de insumo utilizado por hectárea expresada en litros.
tipo_insumo	Clasificación del insumo utilizado según su propósito.

## 4. Análisis Exploratorio de Datos (EDA)

### 4.1. Carga y Limpieza de Datos

#### 4.1.1. Carga de datos

```
# Configuración del directorio de trabajo
setwd("D:/PAC")

# Cargar los datasets
calibres <- read.csv("calibres.csv", sep = ";", stringsAsFactors = FALSE)
reporte <- read.csv("reporte.csv", sep = ";", stringsAsFactors = FALSE)

# Mostrar las primeras filas de cada dataset
head(calibres)

##  semana    mes variedad_producto color_producto calibre kg_calibre
## 1     13  Marzo              Hass           Negra      08         60
## 2     13  Marzo              Hass           Negra     10        2160
## 3     13  Marzo              Hass           Negra     12        9640
## 4     13  Marzo              Hass           Negra     14       10250
## 5     13  Marzo              Hass           Negra     16        1890
## 6     13  Marzo              Hass           Negra     20         30
## porcentaje_calibre_por_guia_remision  guia_remision reporte_produccion
## 1                                0.0025 T008 N° 0000109      0001-0002239
## 2                                0.0898 T008 N° 0000109      0001-0002239
## 3                                0.4007 T008 N° 0000109      0001-0002239
## 4                                0.4260 T008 N° 0000109      0001-0002239
## 5                                0.0786 T008 N° 0000109      0001-0002239
## 6                                0.0012 T008 N° 0000109      0001-0002239
##      fecha cliente
## 1 25/03/2024 CLIENTE2
## 2 25/03/2024 CLIENTE2
## 3 25/03/2024 CLIENTE2
## 4 25/03/2024 CLIENTE2
## 5 25/03/2024 CLIENTE2
## 6 25/03/2024 CLIENTE2

head(reporte)

##  semana    mes fundo empresa  lote hectareas variedad_producto color_producto
## 1     13  Marzo FUNDO1 EMPRESA DNH-06      9.75              Hass           Negra
## 2     13  Marzo FUNDO1 EMPRESA DNH-05     10.91              Hass           Negra
## 3     13  Marzo FUNDO1 EMPRESA DNH-04     11.76              Hass           Negra
## 4     13  Marzo FUNDO1 EMPRESA DNH-03     11.62              Hass           Negra
## 5     13  Marzo FUNDO1 EMPRESA DNH-02     11.50              Hass           Negra
## 6     13  Marzo FUNDO2 EMPRESA CVH-17      7.21              Hass           Negra
##      guia_remision reporte_produccion      fecha total_jabas peso_promedio_jaba
## 1 T008 N° 0000109      0001-0002239 25/03/2024         5        428.13
## 2 T008 N° 0000109      0001-0002239 25/03/2024        10        428.13
## 3 T008 N° 0000109      0001-0002239 25/03/2024        18        428.13
## 4 T008 N° 0000109      0001-0002239 25/03/2024        19        428.13
## 5 T008 N° 0000109      0001-0002239 25/03/2024         5        428.13
## 6 T008 N° 0000109      0001-0002239 25/03/2024         3        428.13
##  cajas_exportadas_10kg ingreso_packing_kg_bruto kg_exportados
## 1                200.5              2140.67        2005
## 2                401.0              4281.33        4010
## 3                721.8              7706.40        7218
```

```

## 4          761.9          8134.53          7619
## 5          200.5          2140.67          2005
## 6          120.3          1284.40          1203
## porcentaje_exportados kg_descarte porcentaje_descarte kg_merma
## 1          0.9366          45.63          0.0213          90.04
## 2          0.9366          91.25          0.0213          180.08
## 3          0.9366          164.25          0.0213          324.15
## 4          0.9366          173.37          0.0213          342.16
## 5          0.9366          45.63          0.0213          90.04
## 6          0.9366          27.38          0.0213          54.03
## porcentaje_merma kg_descarte_campo kg_brutos_lote kg_brutos_hectarea
## 1          0.0421          0          2140.67          219.46
## 2          0.0421          0          4281.33          392.36
## 3          0.0421          0          7706.40          655.51
## 4          0.0421          0          8134.53          699.81
## 5          0.0421          0          2140.67          186.08
## 6          0.0421          0          1284.40          178.14
## kg_exportados_hectarea estado cliente temperatura_promedio
## 1          205.55 Cosechando CLIENTE2          24.28
## 2          367.49 Cosechando CLIENTE2          24.28
## 3          613.97 Cosechando CLIENTE2          24.28
## 4          655.46 Cosechando CLIENTE2          24.28
## 5          174.28 Cosechando CLIENTE2          24.28
## 6          166.85 Cosechando CLIENTE2          24.28
## temperatura_maxima temperatura_minima humedad_promedio humedad_maxima
## 1          32          19.2          67.04          85
## 2          32          19.2          67.04          85
## 3          32          19.2          67.04          85
## 4          32          19.2          67.04          85
## 5          32          19.2          67.04          85
## 6          32          19.2          67.04          85
## humedad_minima evapotranspiracion_acumulada insumo_para_tratamiento
## 1          45          5.62          BIOMURIKATA
## 2          45          5.62          KING PLUS ZINC
## 3          45          5.62          KADONDO
## 4          45          5.62          BASFOLIAR POTASIO
## 5          45          5.62          FRUIT XL
## 6          45          5.62          LEMURIA
## naturaleza_del_insumo dosis_litros_por_hectarea tipo_insumo
## 1          ORGANICA          2.0          ACARICIDA
## 2          MINERAL          1.0          NUTRIENTE
## 3          QUIMICA          1.5          ACARICIDA
## 4          MINERAL          1.6          NUTRIENTE
## 5          QUIMICA          1.0          HORMONAS
## 6          ORGANICA          3.0          ACARICIDA

```

#### 4.1.2. Limpieza inicial

Manejo de valores faltantes, corrección de tipos de datos

##### 4.1.2.1. Inspección inicial de los datos Inspección inicial de datos del dataset 'calibres'

```

# Información general del dataset 'calibres'
str(calibres)

```

```

## 'data.frame':    260 obs. of  11 variables:
## $ semana          : int  13 13 13 13 13 13 13 13 13 13 ...
## $ mes             : chr  "Marzo" "Marzo" "Marzo" "Marzo" ...

```

```
## $ variedad_producto      : chr  "Hass" "Hass" "Hass" "Hass" ...
## $ color_producto         : chr  "Negra" "Negra" "Negra" "Negra" ...
## $ calibre                : chr  "08" "10" "12" "14" ...
## $ kg_calibre              : num  60 2160 9640 10250 1890 ...
## $ porcentaje_calibre_por_guia_remisio: num  0.0025 0.0898 0.4007 0.426 0.0786 ...
## $ guia_remisio           : chr  "T008 N° 0000109" "T008 N° 0000109" "T008 N° 0000109" "T008 N° 0000109" ...
## $ reporte_produccion     : chr  "0001-0002239" "0001-0002239" "0001-0002239" "0001-0002239" ...
## $ fecha                  : chr  "25/03/2024" "25/03/2024" "25/03/2024" "25/03/2024" ...
## $ cliente                 : chr  "CLIENTE2" "CLIENTE2" "CLIENTE2" "CLIENTE2" ...
```

### Inspección inicial de datos del dataset 'reporte'

```
# Información general del dataset 'reporte'
str(reporte)
```

```
## 'data.frame':    121 obs. of  38 variables:
## $ semana                : int  13 13 13 13 13 13 13 13 13 13 ...
## $ mes                    : chr  "Marzo" "Marzo" "Marzo" "Marzo" ...
## $ fundo                  : chr  "FUND01" "FUND01" "FUND01" "FUND01" ...
## $ empresa                : chr  "EMPRESA" "EMPRESA" "EMPRESA" "EMPRESA" ...
## $ lote                   : chr  "DNH-06" "DNH-05" "DNH-04" "DNH-03" ...
## $ hectareas              : num  9.75 10.91 11.76 11.62 11.5 ...
## $ variedad_producto      : chr  "Hass" "Hass" "Hass" "Hass" ...
## $ color_producto         : chr  "Negra" "Negra" "Negra" "Negra" ...
## $ guia_remisio           : chr  "T008 N° 0000109" "T008 N° 0000109" "T008 N° 0000109" "T008 N° 0000109" ...
## $ reporte_produccion     : chr  "0001-0002239" "0001-0002239" "0001-0002239" "0001-0002239" ...
## $ fecha                  : chr  "25/03/2024" "25/03/2024" "25/03/2024" "25/03/2024" ...
## $ total_jabas            : int  5 10 18 19 5 3 7 1 6 5 ...
## $ peso_promedio_jaba     : num  428 428 428 428 428 ...
## $ cajas_exportadas_10kg  : num  200 401 722 762 200 ...
## $ ingreso_packing_kg_bruto : num  2141 4281 7706 8135 2141 ...
## $ kg_exportados          : num  2005 4010 7218 7619 2005 ...
## $ porcentaje_exportados   : num  0.937 0.937 0.937 0.937 0.937 ...
## $ kg_descarte            : num  45.6 91.2 164.2 173.4 45.6 ...
## $ porcentaje_descarte     : num  0.0213 0.0213 0.0213 0.0213 0.0213 0.0213 0.0159 0.0159 0.0159 0.0159 ...
## $ kg_merma               : num  90 180 324 342 90 ...
## $ porcentaje_merma        : num  0.0421 0.0421 0.0421 0.0421 0.0421 0.0421 0.0421 0.035 0.035 0.035 0.035 ...
## $ kg_descarte_campo      : num  0 0 0 0 0 0 0 0 0 ...
## $ kg_brutos_lote         : num  2141 4281 7706 8135 2141 ...
## $ kg_brutos_hectarea     : num  219 392 656 700 186 ...
## $ kg_exportados_hectarea : num  206 367 614 655 174 ...
## $ estado                 : chr  "Cosechando" "Cosechando" "Cosechando" "Cosechando" ...
## $ cliente                 : chr  "CLIENTE2" "CLIENTE2" "CLIENTE2" "CLIENTE2" ...
## $ temperatura_promedio   : num  24.3 24.3 24.3 24.3 24.3 ...
## $ temperatura_maxima     : num  32 32 32 32 32 32 32.5 32.5 ...
## $ temperatura_minima     : num  19.2 19.2 19.2 19.2 19.2 19.2 21 21 21 21 ...
## $ humedad_promedio       : num  67 67 67 67 67 ...
## $ humedad_maxima         : num  85 85 85 85 85 85 76 76 76 76 ...
## $ humedad_minima         : num  45 45 45 45 45 45 48 48 48 48 ...
## $ evapotranspiracion_acumulada: num  5.62 5.62 5.62 5.62 5.62 5.62 5.62 4.92 4.92 4.92 4.92 ...
## $ insumo_para_tratamiento : chr  "BIOMURIKATA" "KING PLUS ZINC" "KADONDO" "BASFOLIAR POTASIO" ...
## $ naturaleza_del_insumo   : chr  "ORGANICA" "MINERAL" "QUIMICA" "MINERAL" ...
## $ dosis_litros_por_hectarea : num  2 1 1.5 1.6 1 3 2 1 1 1 ...
## $ tipo_insumo             : chr  "ACARICIDA" "NUTRIENTE" "ACARICIDA" "NUTRIENTE" ...
```

#### 4.1.2.2. Limpieza inicial de datos Identificación de valores faltantes

```
sum(is.na(calibres))
```

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

```
sum(is.na(reporte))
```

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

No se encontraron valores nulos en los datasets.

## 4.2. Análisis Descriptivo General

### 4.2.1. Resumen estadístico por dataset

Dimensiones y resumen estadístico dataset ‘calibres’

```
# Dimensiones de 'calibres'
```

```
dim(calibres)
```

```
## [1] 260 11
```

```
# Resumen estadístico básico
```

```
summary(calibres)
```

```
##      semana      mes      variedad_producto  color_producto
## 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      kg_calibre      porcentaje_calibre_por_guia_remision
## Length:260      Min.   : 10.0   Min.   :0.00020
## Class :character 1st Qu.: 517.5   1st Qu.:0.02562
## Mode  :character Median : 1495.0   Median :0.09285
##                  Mean   : 2707.3   Mean   :0.14615
##                  3rd Qu.: 3872.5   3rd Qu.:0.17300
##                  Max.   :17680.0   Max.   :1.00000
##      guia_remision      reporte_produccion      fecha      cliente
## Length:260      Length:260      Length:260      Length:260
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

Dimensiones y resumen estadístico dataset ‘reporte’

```
# Dimensiones de 'reporte'
```

```
dim(reporte)
```

```
## [1] 121 38
```

```
# Resumen estadístico básico
```

```
summary(reporte)
```

```
##      semana      mes      fundo      empresa
## Min.   :13.00   Length:121   Length:121      Length:121
## 1st Qu.:13.00   Class :character   Class :character Class :character
## Median :14.00   Mode  :character   Mode  :character Mode  :character
## Mean   :14.38
## 3rd Qu.:15.00
```

```

## Max. :16.00
## lote hectareas variedad_producto color_producto
## Length:121 Min. : 6.55 Length:121 Length:121
## Class :character 1st Qu.:10.24 Class :character Class :character
## Mode :character Median :11.76 Mode :character Mode :character
## Mean :14.22
## 3rd Qu.:17.18
## Max. :26.57
## guia_remision reporte_produccion fecha total_jabas
## Length:121 Length:121 Length:121 Min. : 1.00
## Class :character Class :character Class :character 1st Qu.: 2.00
## Mode :character Mode :character Mode :character Median : 5.00
## Mean :14.83
## 3rd Qu.:21.00
## Max. :60.00
## peso_promedio_jaba cajas_exportadas_10kg ingreso_packing_kg_bruto
## Min. : 0.0 Min. : 0.00 Min. : 0.0
## 1st Qu.:416.2 1st Qu.: 80.77 1st Qu.: 840.2
## Median :426.7 Median : 204.62 Median : 2169.6
## Mean :402.0 Mean : 580.69 Mean : 6291.2
## 3rd Qu.:439.0 3rd Qu.: 858.19 3rd Qu.: 9175.9
## Max. :679.5 Max. :2448.89 Max. :27291.0
## kg_exportados porcentaje_exportados kg_descarte porcentaje_descarte
## Min. : 0.0 Min. :0.0000 Min. : 0.00 Min. :0.00000
## 1st Qu.: 807.7 1st Qu.:0.9178 1st Qu.: 0.00 1st Qu.:0.00000
## Median : 2046.2 Median :0.9431 Median : 33.16 Median :0.01590
## Mean : 5806.9 Mean :0.9004 Mean : 265.68 Mean :0.02325
## 3rd Qu.: 8581.9 3rd Qu.:0.9618 3rd Qu.: 267.62 3rd Qu.:0.02920
## Max. :24488.9 Max. :1.0000 Max. :2736.90 Max. :0.10030
## kg_merma porcentaje_merma kg_descarte_campo kg_brutos_lote
## Min. : 0.00 Min. :0.00000 Min. : 0.00 Min. : 50.0
## 1st Qu.: 0.00 1st Qu.:0.00000 1st Qu.: 0.00 1st Qu.: 840.2
## Median : 84.29 Median :0.03500 Median : 0.00 Median : 2169.6
## Mean :218.54 Mean :0.02671 Mean : 22.12 Mean : 6313.3
## 3rd Qu.:342.16 3rd Qu.:0.03940 3rd Qu.: 0.00 3rd Qu.: 9175.9
## Max. :986.92 Max. :0.07950 Max. :483.50 Max. :27291.0
## kg_brutos_hectarea kg_exportados_hectarea estado
## Min. : 3.73 Min. : 0.00 Length:121
## 1st Qu.: 49.57 1st Qu.: 46.04 Class :character
## Median : 202.48 Median : 192.18 Mode :character
## Mean : 479.53 Mean : 440.45
## 3rd Qu.: 680.73 3rd Qu.: 647.45
## Max. :2476.18 Max. :2244.27
## cliente temperatura_promedio temperatura_maxima temperatura_minima
## Length:121 Min. :20.56 Min. :26.10 Min. :16.80
## Class :character 1st Qu.:22.90 1st Qu.:30.50 1st Qu.:18.20
## Mode :character Median :23.63 Median :32.00 Median :18.90
## Mean :23.34 Mean :31.38 Mean :18.97
## 3rd Qu.:24.28 3rd Qu.:32.80 3rd Qu.:19.70
## Max. :25.20 Max. :34.20 Max. :21.00
## humedad_promedio humedad_maxima humedad_minima evapotranspiracion_acumulada
## Min. :64.46 Min. :76.00 Min. :36.00 Min. :4.370
## 1st Qu.:67.06 1st Qu.:84.00 1st Qu.:41.00 1st Qu.:4.920
## Median :68.21 Median :85.00 Median :45.00 Median :5.620
## Mean :70.14 Mean :85.45 Mean :45.93 Mean :5.384
## 3rd Qu.:72.56 3rd Qu.:87.00 3rd Qu.:48.00 3rd Qu.:5.780

```



```
## Max. :80.85 Max. :94.00 Max. :62.00 Max. :6.240
## insumo_para_tratamiento naturaleza_del_insumo dosis_litros_por_hectarea
## Length:121 Length:121 Min. :1.000
## Class :character Class :character 1st Qu.:1.000
## Mode :character Mode :character Median :2.000
## Mean :1.986
## 3rd Qu.:2.500
## Max. :4.000
## tipo_insumo
## Length:121
## Class :character
## Mode :character
##
##
##
```

## 4.2.2. Visualización de distribuciones numéricas

### 4.2.2.1. Histogramas y Boxplots para las variables numéricas más relevantes del dataset ‘calibre’

#### 4.2.2.1.1. Variable kg\_calibre Se obtienen los datos descriptivos más representativos de la variable.

```
# Datos descriptivos de la variable
summary(calibres$kg_calibre)
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 10.0 517.5 1495.0 2707.3 3872.5 17680.0
```

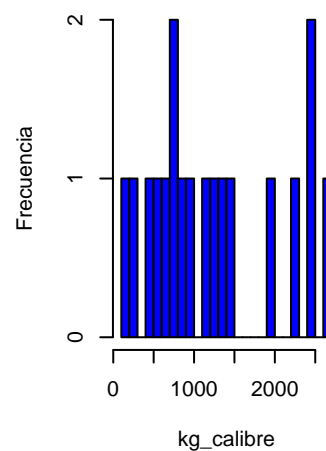
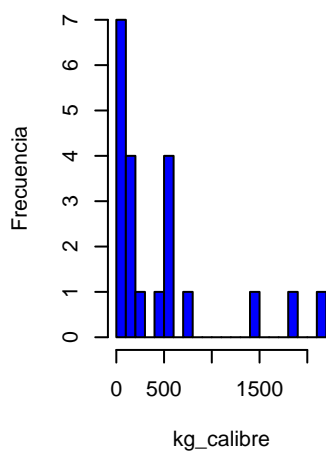
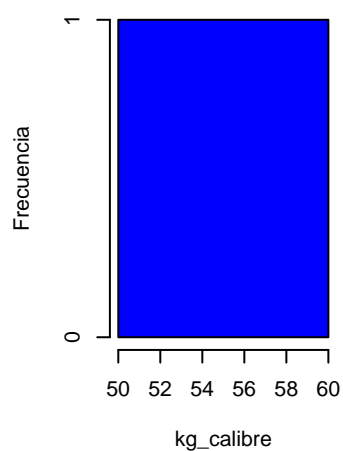
Se continua con la generación de gráficos.

```
# Histograma para la variable numérica kg_calibre
calibres_por_calibre <- split(calibres$kg_calibre, calibres$calibre)

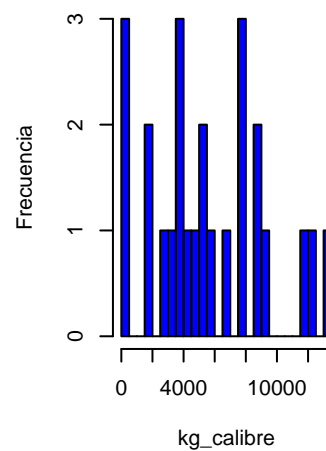
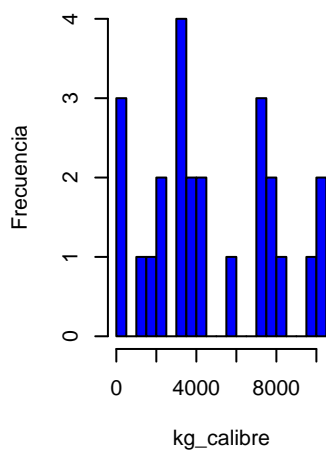
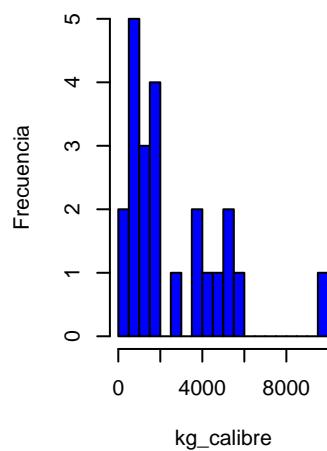
par(mfrow=c(2, 3))

for (i in 1:length(calibres_por_calibre)) {
  hist(calibres_por_calibre[[i]], main=paste("Distr. de kg_calibre (Calibre", unique(calibres$calibre)[i],
    xlab="kg_calibre", ylab="Frecuencia", col="blue", breaks=30)
}
```

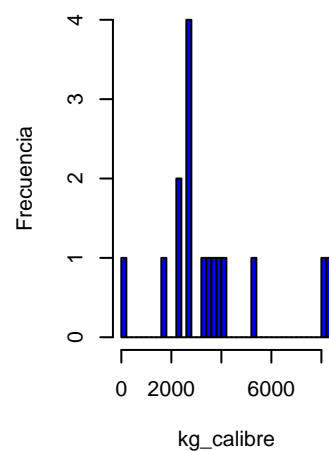
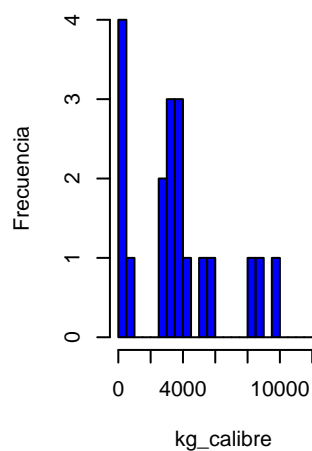
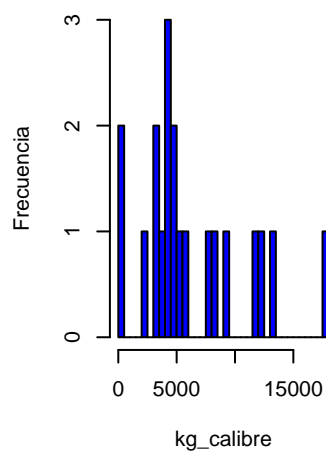
**Distr. de kg\_calibre (Calibre 0** **Distr. de kg\_calibre (Calibre 1** **Distr. de kg\_calibre (Calibre 1**



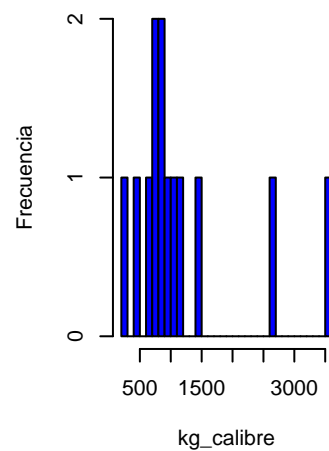
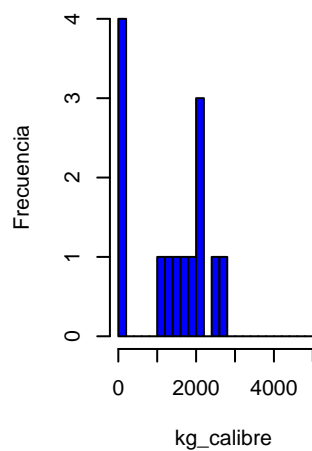
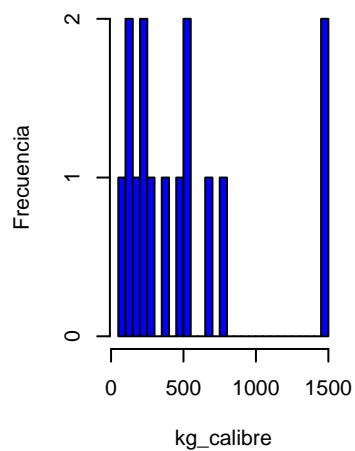
**Distr. de kg\_calibre (Calibre 1** **Distr. de kg\_calibre (Calibre 1** **Distr. de kg\_calibre (Calibre 2**



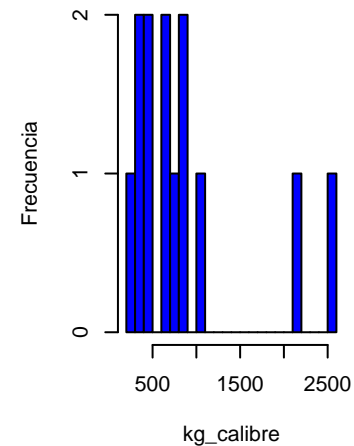
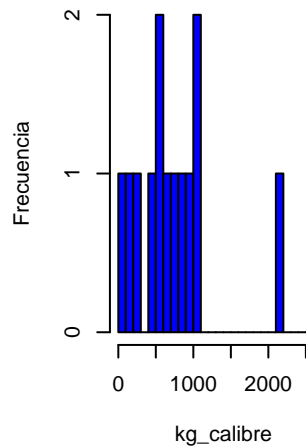
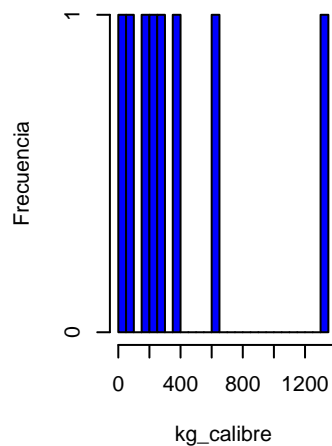
**Distr. de kg\_calibre (Calibre 20)** **Distr. de kg\_calibre (Calibre 10)** **Distr. de kg\_calibre (Calibre 1**



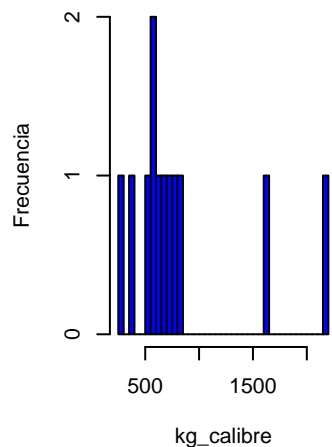
**Distr. de kg\_calibre (Calibre 22)** **Distr. de kg\_calibre (Calibre 20)** **Distr. de kg\_calibre (Calibre 26)**



Distr. de kg\_calibre (Calibre 2)    Distr. de kg\_calibre (Calibre 2)    Distr. de kg\_calibre (Calibre 3)



Distr. de kg\_calibre (Calibre 3)



**Interpretación:** Los histogramas muestran la distribución de la variable kg\_calibre para diferentes calibres, evidenciando tendencias unimodales con picos únicos. Aunque la forma y la variabilidad varían entre calibres, en general, los calibres 08, 10 y 12 presentan distribuciones simétricas y compactas con menor variabilidad, mientras que los calibres 14, 16, 20, 24, 10-20, 18, 22-24, 22, 26-30, 28 y 30 muestran distribuciones asimétricas y más extendidas con mayor variabilidad.

**4.2.2.1.2. Variable porcentaje\_calibre\_por\_guia\_remision** Se obtienen los datos descriptivos más representativos de la variable.

```
# Datos descriptivos de la variable
summary(calibres$porcentaje_calibre_por_guia_remision)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00020 0.02562 0.09285 0.14615 0.17300 1.00000
```

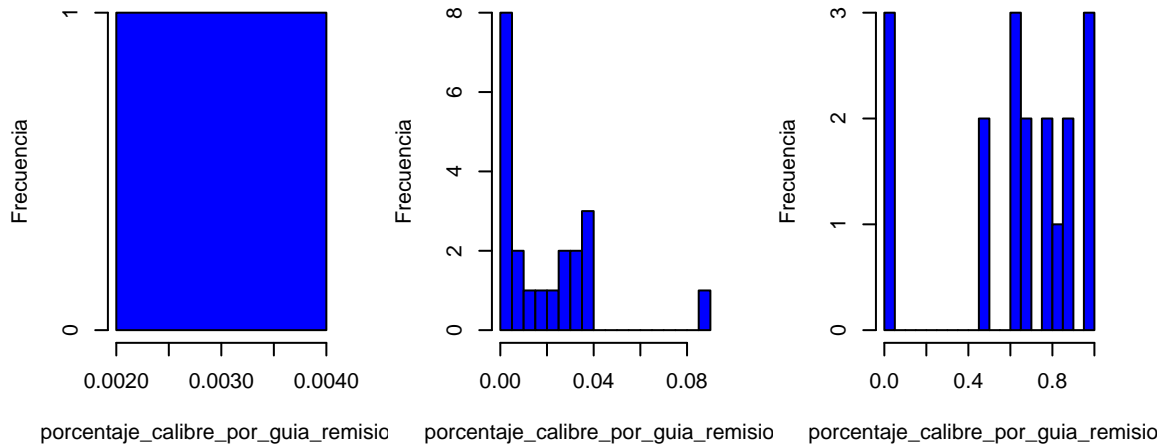
Se continua con la generación de gráficos.

```
# Histograma para la variable numérica porcentaje_calibre_por_guia_remision
calibres_por_calibre <- split(calibres$porcentaje_calibre_por_guia_remision, calibres$calibre)

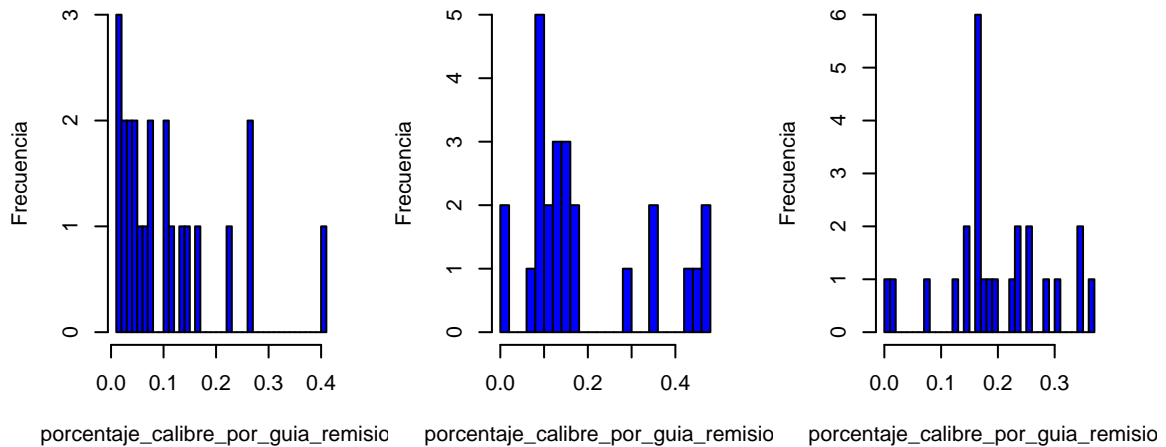
par(mfrow=c(2, 3))
```

```
for (i in 1:length(calibres_por_calibre)) {
  hist(calibres_por_calibre[[i]], main=paste("Distribución de porcentaje_calibre_por_guia_remision (Calibre", i, ")"),
       xlab="porcentaje_calibre_por_guia_remision", ylab="Frecuencia", col="blue", breaks=30)
}
```

porcentaje\_calibre\_por\_guia\_remision porcentaje\_calibre\_por\_guia\_remision porcentaje\_calibre\_por\_guia\_remision

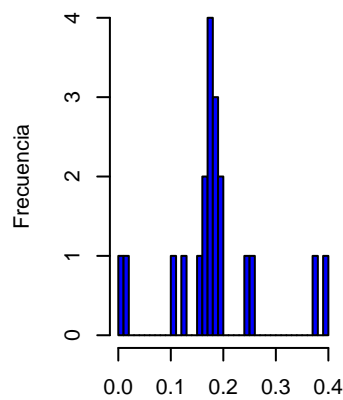


porcentaje\_calibre\_por\_guia\_remision porcentaje\_calibre\_por\_guia\_remision porcentaje\_calibre\_por\_guia\_remision

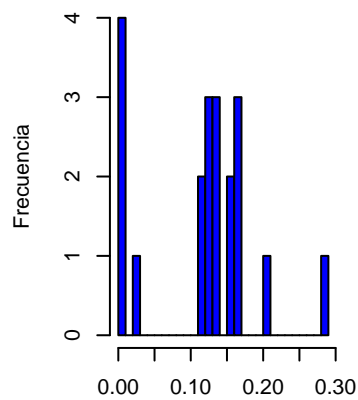


porcentaje\_calibre\_por\_guia\_remision porcentaje\_calibre\_por\_guia\_remision porcentaje\_calibre\_por\_guia\_remision

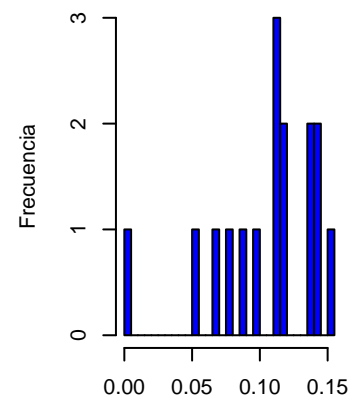
porcentaje\_calibre\_por\_guia\_r porcentaje\_calibre\_por\_guia\_re porcentaje\_calibre\_por\_guia\_r



porcentaje\_calibre\_por\_guia\_remisio

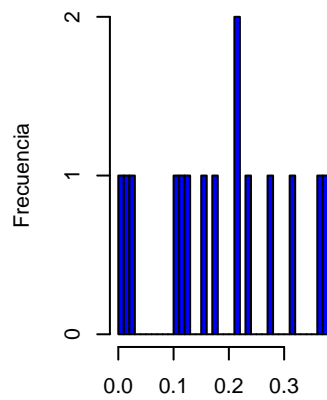


porcentaje\_calibre\_por\_guia\_remisio

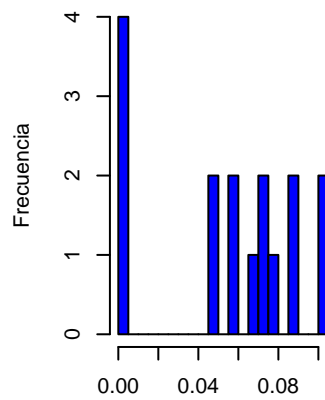


porcentaje\_calibre\_por\_guia\_remisio

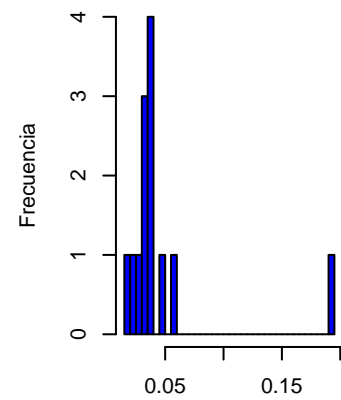
porcentaje\_calibre\_por\_guia\_remisio porcentaje\_calibre\_por\_guia\_remisio porcentaje\_calibre\_por\_guia\_remisio



porcentaje\_calibre\_por\_guia\_remisio

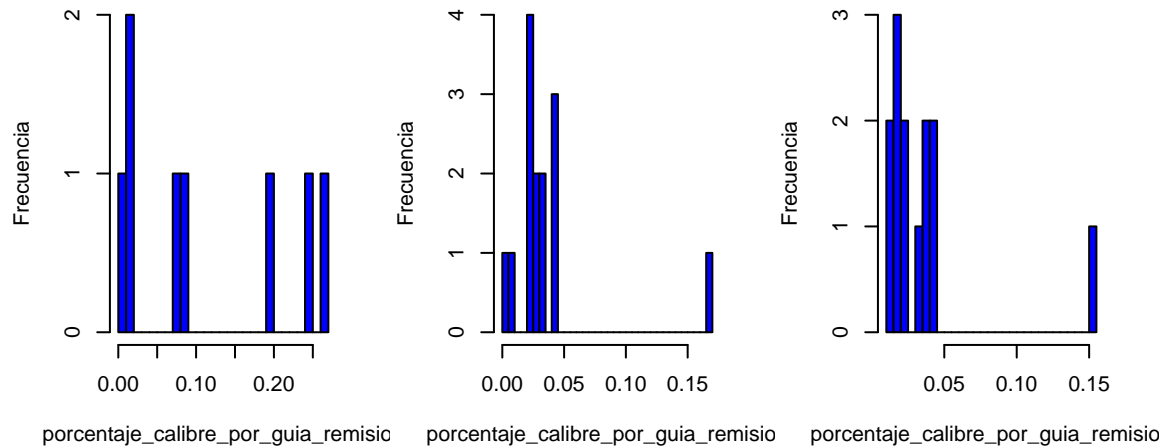


porcentaje\_calibre\_por\_guia\_remisio

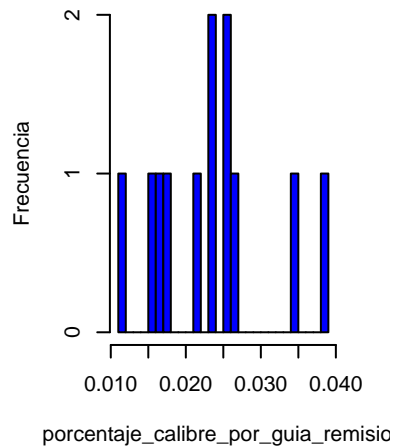


porcentaje\_calibre\_por\_guia\_remisio

porcentaje\_calibre\_por\_guia\_roorcentaje\_calibre\_por\_guia\_roorcentaje\_calibre\_por\_guia\_r



porcentaje\_calibre\_por\_guia\_r



**Interpretación:** Los histogramas muestran la distribución de la variable porcentaje\_calibre\_por\_guia\_remision para seis calibres diferentes (08, 10, 12, 14, 16 y 20), revelando distribuciones mayormente unimodales con algunos sesgos hacia la derecha. Calibres como el 08, 10 y 12 exhiben distribuciones simétricas y compactas, mientras que los calibres 14, 16 y 20 muestran distribuciones más asimétricas y extendidas. La variabilidad también varía entre calibres, siendo menor para los calibres más bajos y mayor para los más altos.

#### 4.2.2.2. Histogramas y Boxplots para las variables numéricas más relevantes del dataset Reporte

**4.2.2.2.1. Variable hectareas** Se obtienen los datos descriptivos más representativos de la variable.

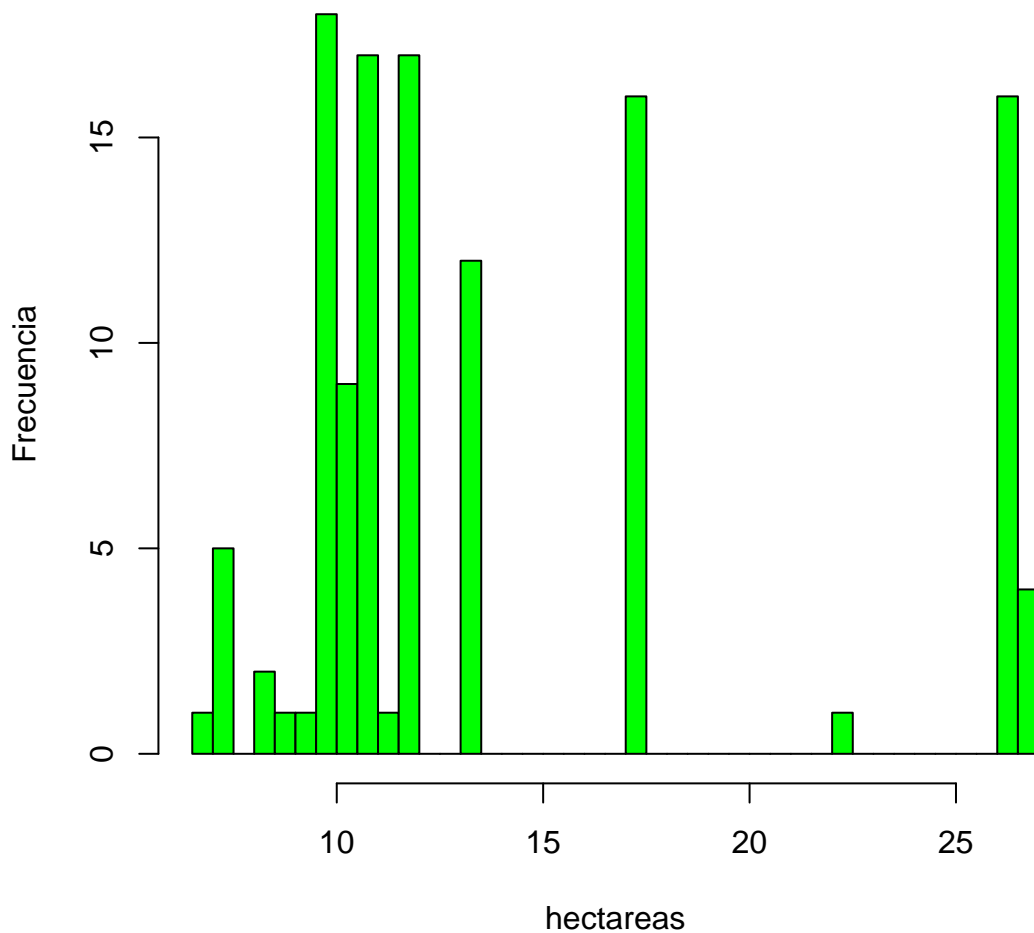
```
# Datos descriptivos de la variable
summary(reporte$hectareas)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      6.55  10.24   11.76   14.22  17.18   26.57
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica hectareas
hist(reporte$hectareas, main="Distribución de Variable hectareas", xlab="hectareas", ylab="Frecuencia", col="blue")
```

## Distribución de Variable hectareas

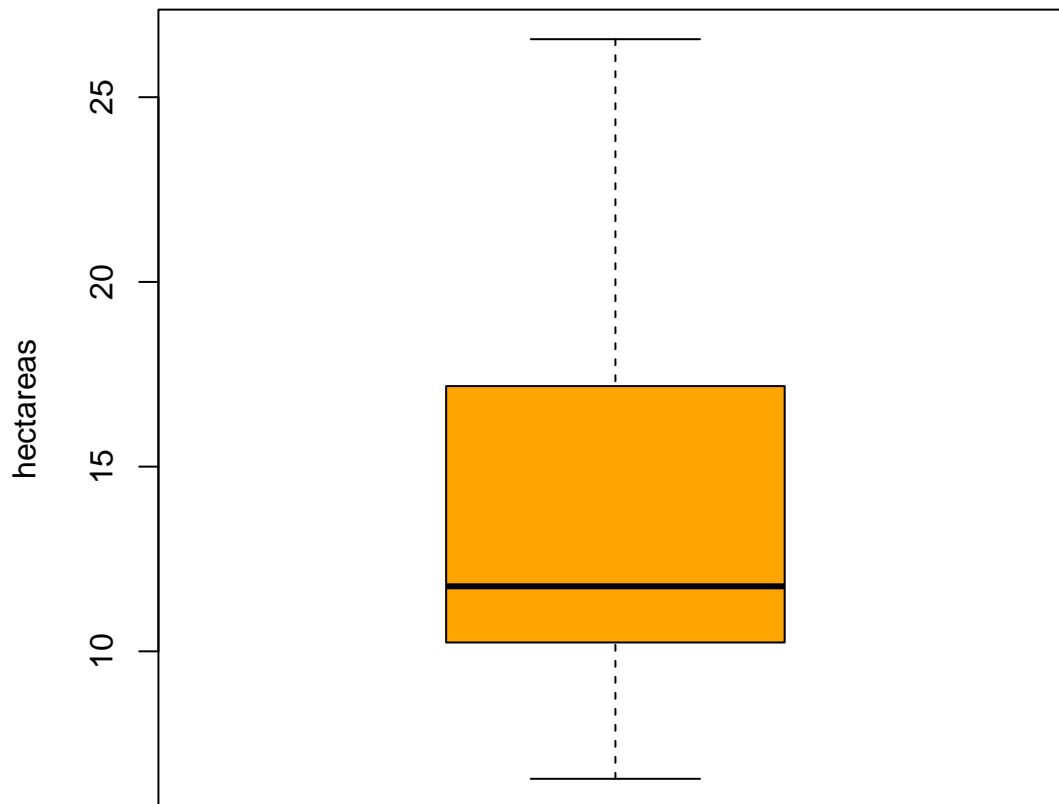


**Interpretación:** El histograma muestra una distribución de las hectáreas cultivadas de paltas Hass. La mayoría de las observaciones se encuentran en el rango de 10 a 15 hectáreas, con un pico prominente alrededor de 12 hectáreas. Esto indica que la mayoría de las fincas en el dataset tienden a tener un tamaño relativamente pequeño. Hay también algunos valores en el extremo superior de la distribución, cerca de las 25 hectáreas, que podrían ser fincas más grandes o posibles outliers. La distribución no es simétrica y parece tener una ligera cola hacia la derecha, indicando que hay menos fincas con áreas más grandes.

```
# Boxplot para la variable numérica hectareas  
boxplot(reporte$hectareas, main="Boxplot de Variable hectareas", ylab="hectareas", col="orange")
```



## Boxplot de Variable hectareas



**Interpretación:** El boxplot proporciona una visualización de la distribución de hectáreas desde otra perspectiva. La mediana de hectáreas cultivadas, indicada por la línea dentro de la caja, está cerca de 12 hectáreas, lo que coincide con el pico observado en el histograma. La caja en sí, que representa el rango intercuartílico (IQR), muestra que la mitad central de las fincas tiene entre aproximadamente 10 y 17 hectáreas. No hay puntos fuera de los 'bigotes', lo que sugiere que no hay valores extremos o atípicos significativos en esta variable. La caja es más ancha en la parte superior, lo que indica una mayor variabilidad en las fincas más grandes en comparación con las más pequeñas.

**4.2.2.2.2. Variable kg\_brutos\_lote** Se obtienen los datos descriptivos más representativos de la variable.

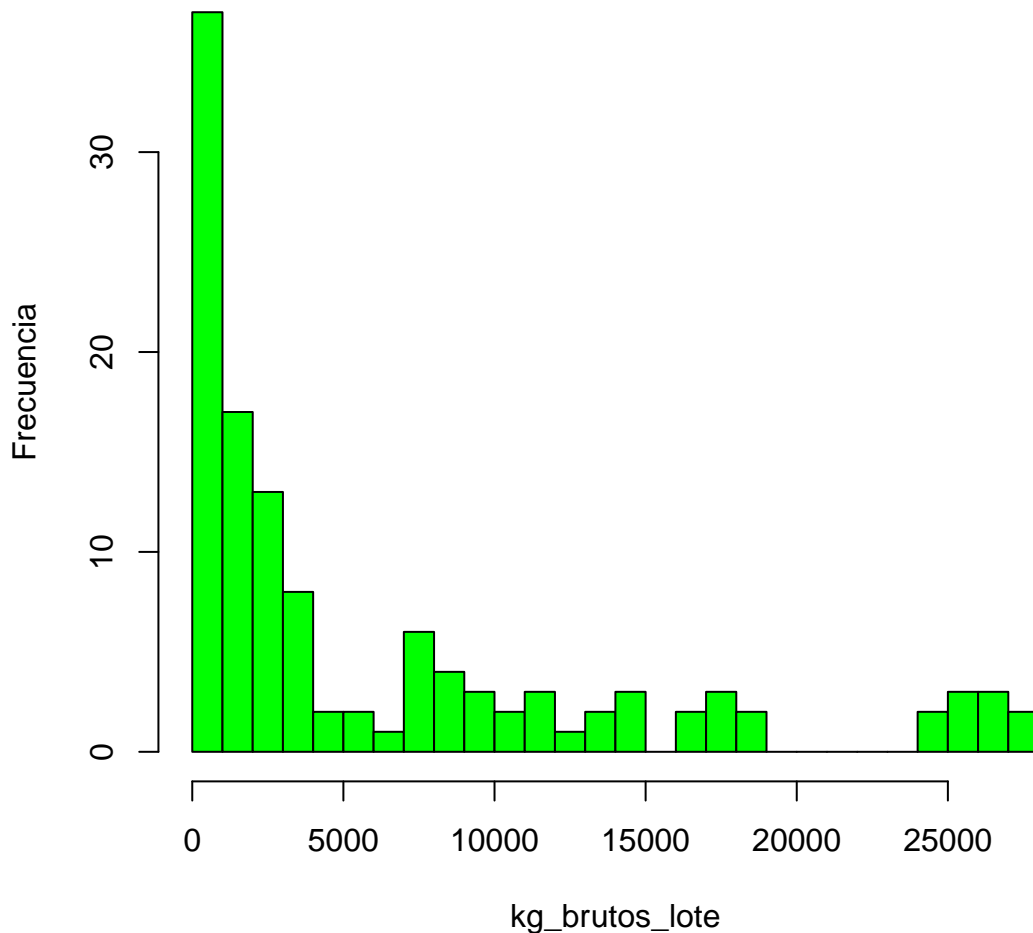
```
# Datos descriptivos de la variable
summary(reporte$kg_brutos_lote)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      50.0   840.2  2169.6  6313.3  9175.9 27291.0
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica kg_brutos_lote
hist(reporte$kg_brutos_lote, main="Distribución de Variable kg_brutos_lote", xlab="kg_brutos_lote", ylab="")
```

## Distribución de Variable kg\_brutos\_lote

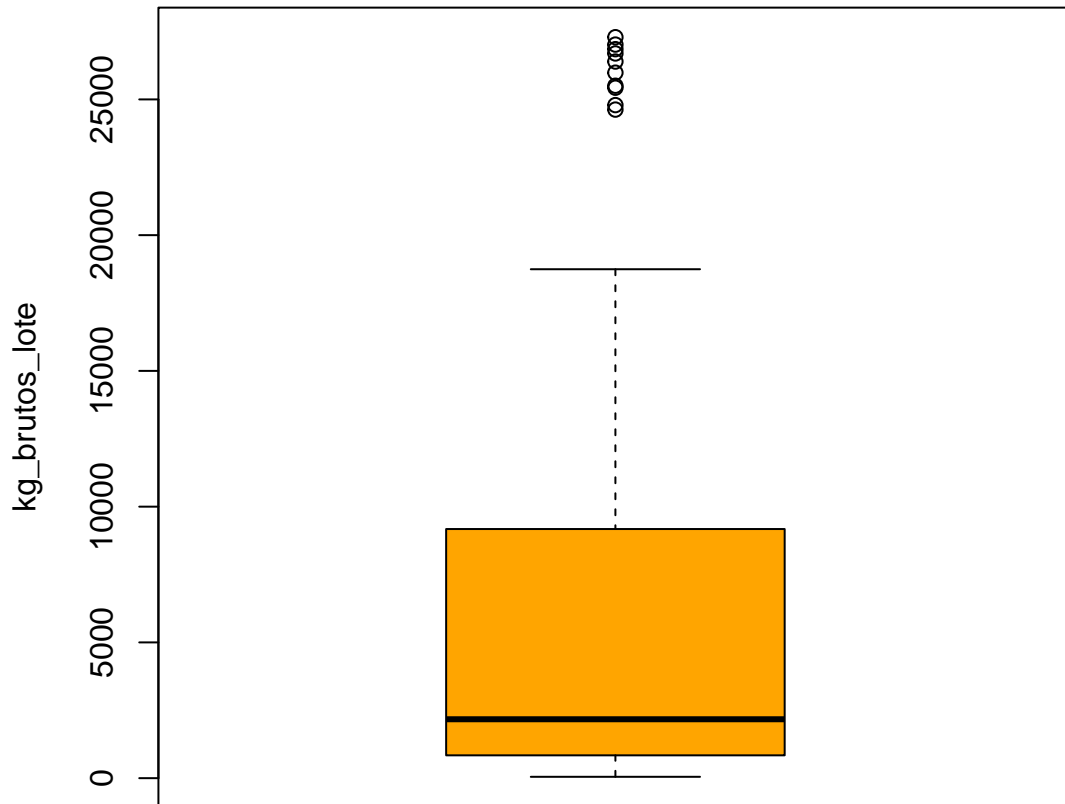


**Interpretación:** El histograma muestra la distribución de los kilogramos brutos por lote de paltas Hass. Observamos que la mayoría de los lotes tienen un peso bruto bajo, con un pico significativo en el rango más bajo de la escala (50 kg aproximadamente). La frecuencia disminuye rápidamente a medida que aumenta la cantidad de kg brutos por lote, indicando que hay menos lotes con pesos mayores. La distribución es asimétrica a la derecha, con algunos lotes que llegan hasta aproximadamente 27,291 kg, lo cual es coherente con los valores extremos observados. La forma de la distribución sugiere que la producción de paltas por lote es predominantemente de baja a media, con unos pocos lotes de producción alta.

```
# Boxplot para la variable numérica kg_brutos_lote
```

```
boxplot(reporte$kg_brutos_lote, main="Boxplot de Variable kg_brutos_lote", ylab="kg_brutos_lote", col="orange")
```

### Boxplot de Variable kg\_brutos\_lote



**Interpretación:** El boxplot ofrece una representación visual de la distribución de los kg brutos por lote. El rango intercuartílico, representado por la caja, muestra que la mitad central de los datos varía de aproximadamente 840 a 9176 kg por lote. La mediana está en 2169.6 kg, lo que indica que la mitad de los lotes tiene un peso bruto menor a este valor y la otra mitad, mayor. La línea que se extiende desde la caja, conocida como bigote, va hasta cerca de 27291 kg, que es el valor máximo no atípico. Sin embargo, el gráfico muestra también puntos fuera del bigote superior, lo que indica la presencia de valores atípicos que podrían ser lotes con pesos excepcionalmente altos.

**4.2.2.2.3. Variable kg\_exportados** Se obtienen los datos descriptivos más representativos de la variable.

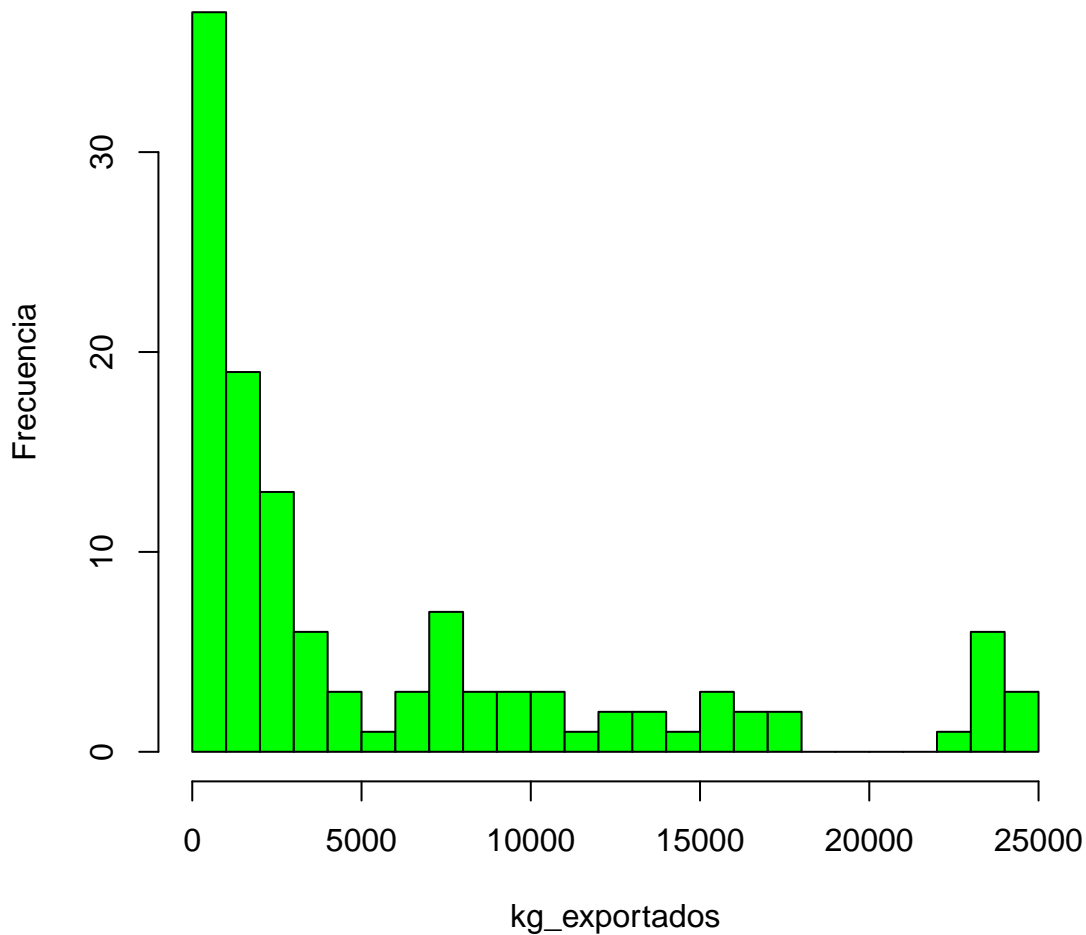
```
# Datos descriptivos de la variable
summary(reporte$kg_exportados)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   807.7   2046.2   5806.9   8581.9  24488.9
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica kg_exportados
hist(reporte$kg_exportados, main="Distribución de Variable kg_exportados", xlab="kg_exportados", ylab="Fre
```

## Distribución de Variable kg\_exportados

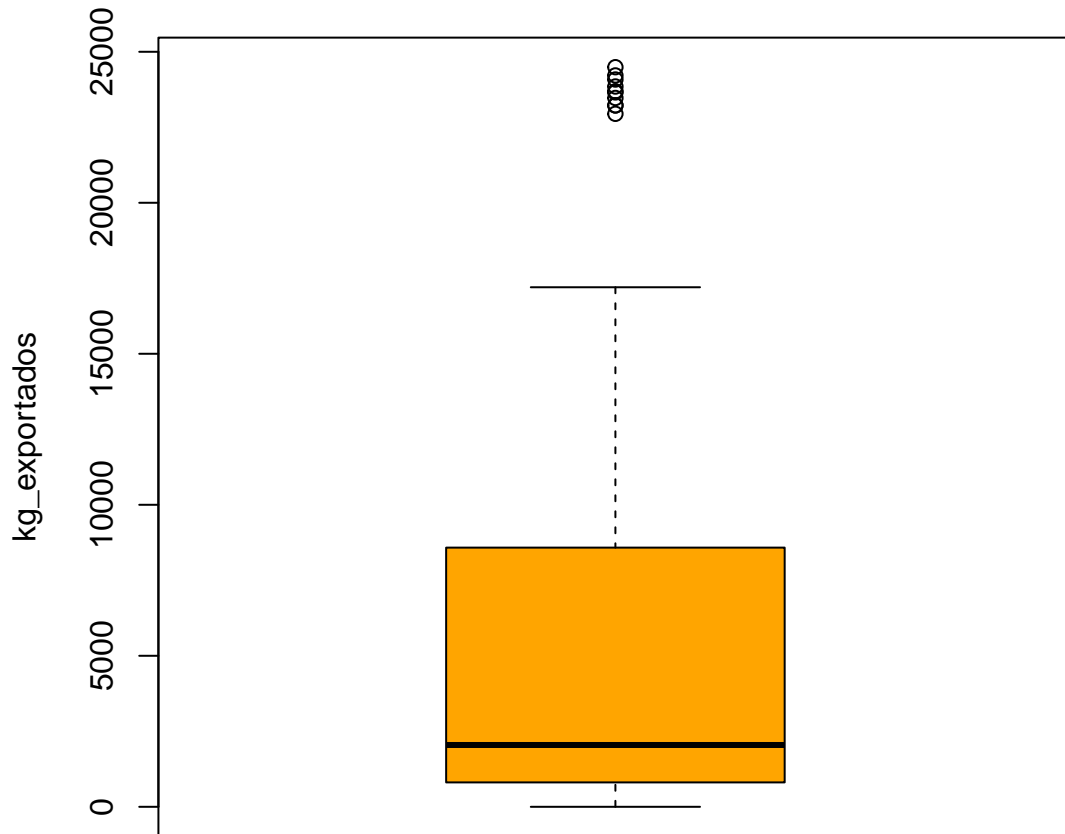


**Interpretación:** Observamos que la mayoría de los datos se concentran en el extremo inferior de la escala, con un número significativo de lotes que tienen pocos kilogramos exportados. Esto se refleja en la alta barra cerca del cero y sugiere que hay una gran cantidad de lotes con volúmenes de exportación bajos. La distribución se extiende hacia la derecha con una larga cola, indicando que hay menos lotes con volúmenes de exportación altos. Esta distribución asimétrica a la derecha sugiere que, aunque la mayoría de los lotes exporta una cantidad menor, hay algunos lotes que exportan cantidades mucho mayores, lo cual puede indicar la presencia de operaciones de exportación de mayor escala o eventos puntuales de exportación masiva.

```
# Boxplot para la variable numérica kg_exportados
```

```
boxplot(reporte$kg_exportados, main="Boxplot de Variable kg_exportados", ylab="kg_exportados", col="orange")
```

### Boxplot de Variable kg\_exportados



**Interpretación:** El boxplot proporciona un resumen visual de los kilogramos de paltas exportados. La línea dentro de la caja, que representa la mediana, está en 2046.2 kg, lo que indica que la mitad de los lotes exporta menos que esta cantidad y la otra mitad, más. El rango intercuartílico se extiende desde aproximadamente 808 hasta 8582 kg, mostrando la variabilidad de los datos. Notamos que hay varios puntos fuera del bigote superior, lo que indica la presencia de valores atípicos. Esto puede significar que algunos lotes tienen volúmenes de exportación significativamente más altos que la mayoría de los otros lotes. La presencia de estos valores atípicos puede tener un impacto en la media, que es de 5806.9 kg, haciéndola más alta que la mediana.

**4.2.2.2.4. Variable kg\_exportados\_hectarea** Se obtienen los datos descriptivos más representativos de la variable.

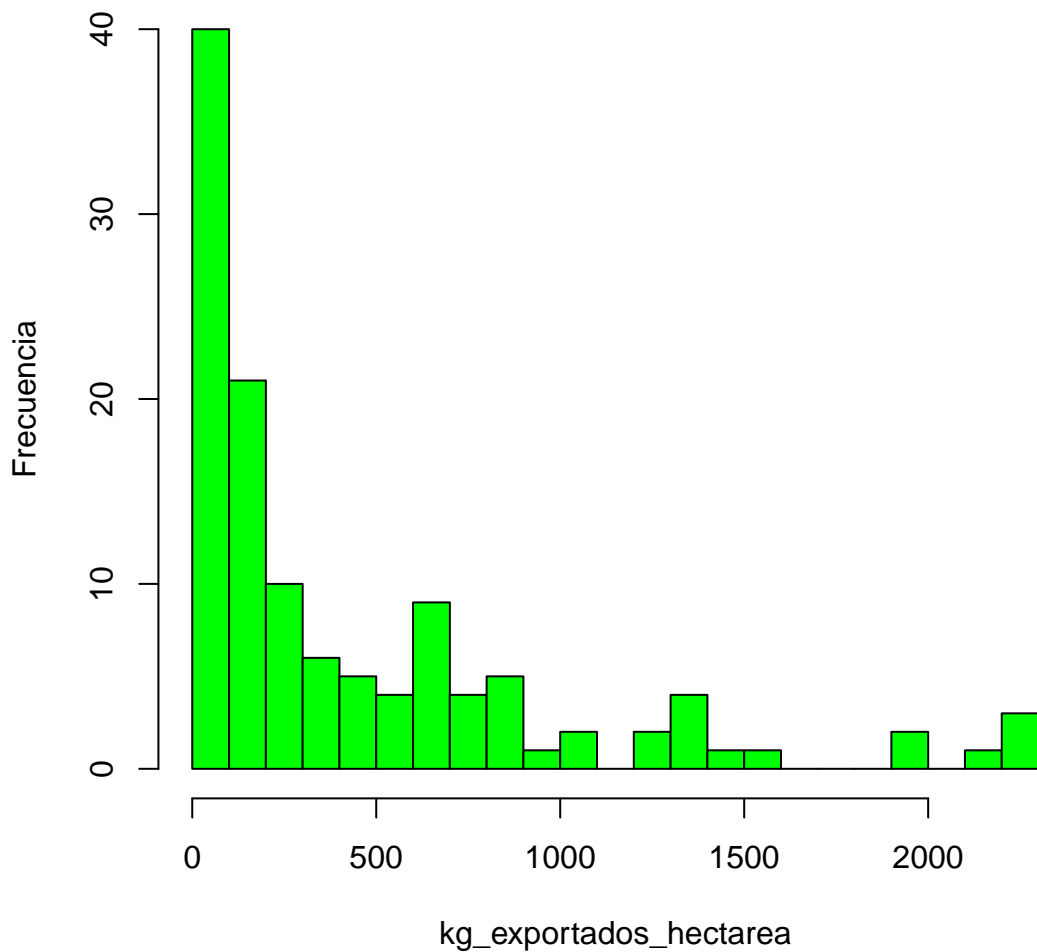
```
# Datos descriptivos de la variable
summary(reporte$kg_exportados_hectarea)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   46.04   192.18   440.45   647.45  2244.27
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica kg_exportados_hectarea
hist(reporte$kg_exportados_hectarea, main="Distribución de Variable kg_exportados_hectarea", xlab="kg_exp")
```

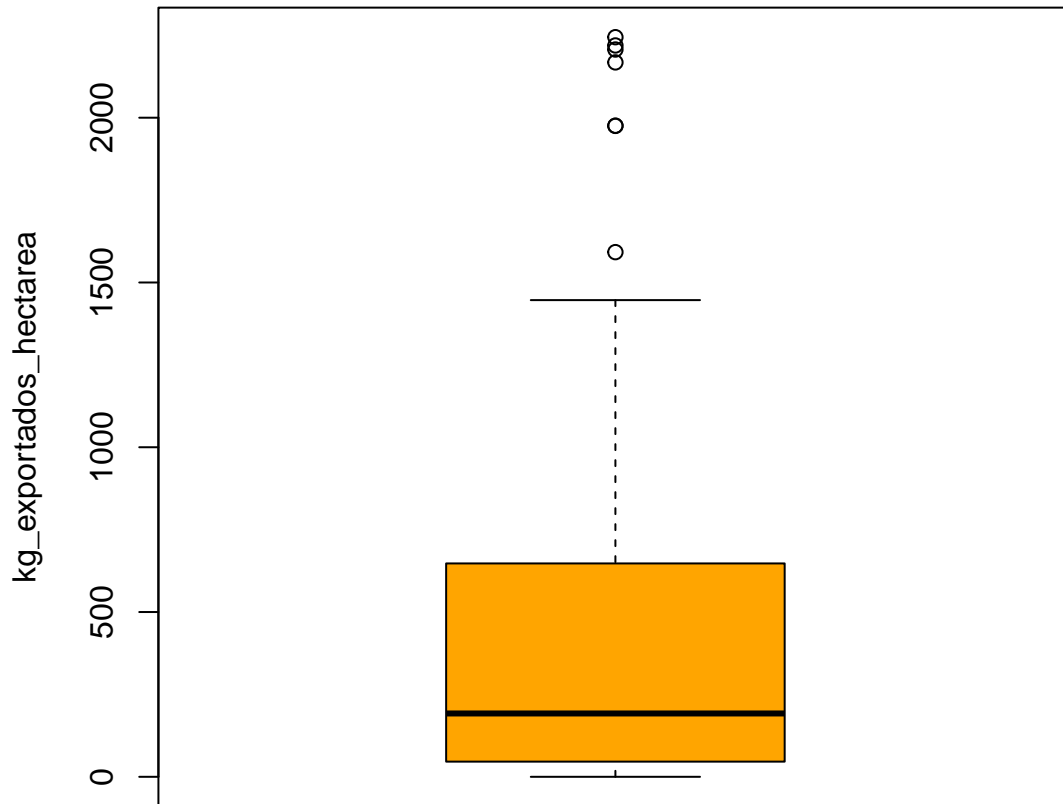
## Distribución de Variable kg\_exportados\_hectarea



**Interpretación:** El histograma de los kilogramos exportados por hectárea muestra una clara concentración de valores en el extremo inferior del eje x. Con la mayoría de los datos agrupados cerca del origen, hay una fuerte asimetría hacia la derecha, indicando que hay menos parcelas con valores altos de exportación por hectárea. La barra más alta está cerca del cero, lo que podría sugerir que hay una cantidad considerable de hectáreas con muy bajos niveles de exportación. Esto puede reflejar limitaciones en la productividad o problemas en ciertas áreas que están afectando la capacidad de exportación.

```
# Boxplot para la variable numérica kg_exportados_hectarea
boxplot(reporte$kg_exportados_hectarea, main="Boxplot de Variable kg_exportados_hectarea", ylab="kg_exportados_hectarea")
```

### Boxplot de Variable kg\_exportados\_hectarea



**Interpretación:** El boxplot para los kilogramos exportados por hectárea muestra una mediana de 192.18 kg por hectárea, lo que indica que la mitad de las hectáreas tienen un rendimiento de exportación por debajo de esta cifra y la otra mitad por encima. El rango intercuartílico se extiende desde aproximadamente 46 a 647 kg por hectárea, lo que sugiere que hay una variación moderada en los rendimientos de exportación entre la mayoría de las hectáreas. La presencia de varios valores atípicos indica que hay algunas hectáreas con rendimientos excepcionalmente altos, mucho más allá del promedio. La media es de 440.45 kg por hectárea, esto podría estar influenciado por estos valores atípicos, y es más alta que la mediana, lo que sugiere una distribución sesgada.

**4.2.2.2.5. Variable temperatura\_promedio** Se obtienen los datos descriptivos más representativos de la variable.

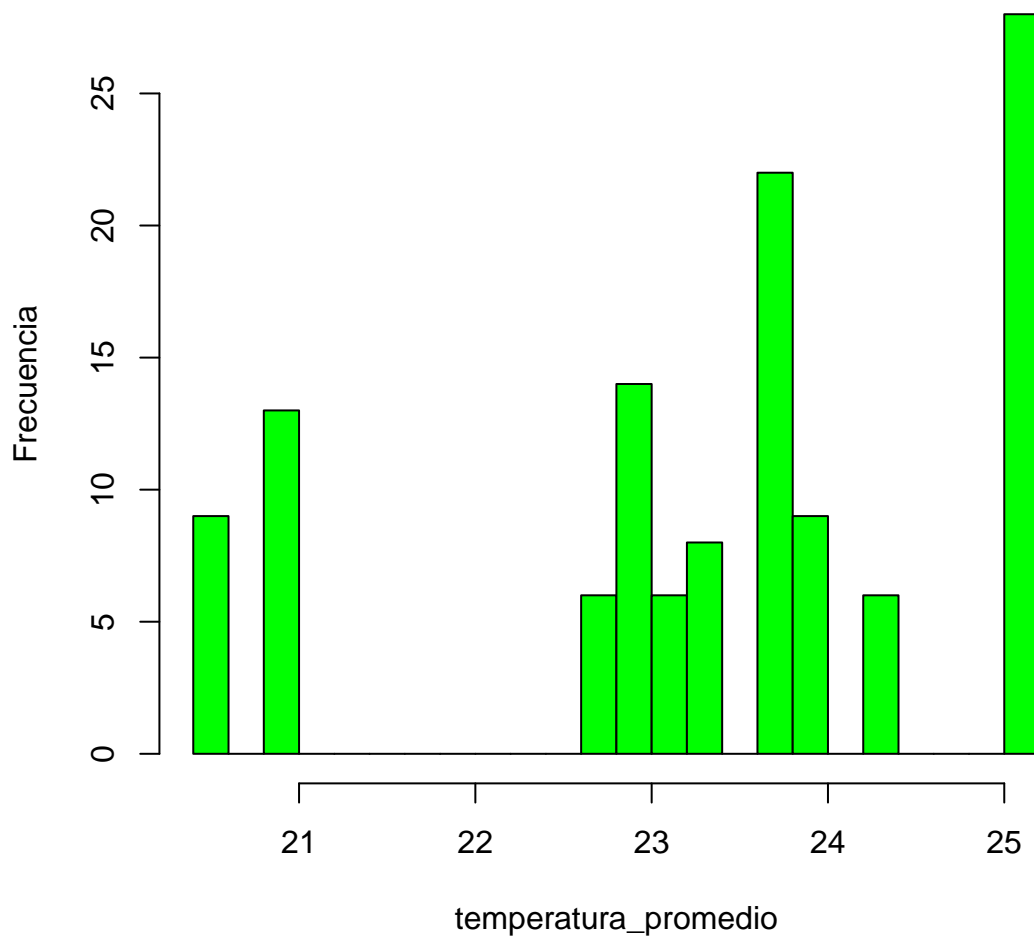
```
# Datos descriptivos de la variable
summary(reporte$temperatura_promedio)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  20.56   22.90   23.63   23.34   24.28   25.20
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica temperatura_promedio
hist(reporte$temperatura_promedio, main="Distribución de Variable temperatura_promedio", xlab="temperatura_promedio")
```

## Distribución de Variable temperatura\_promedio



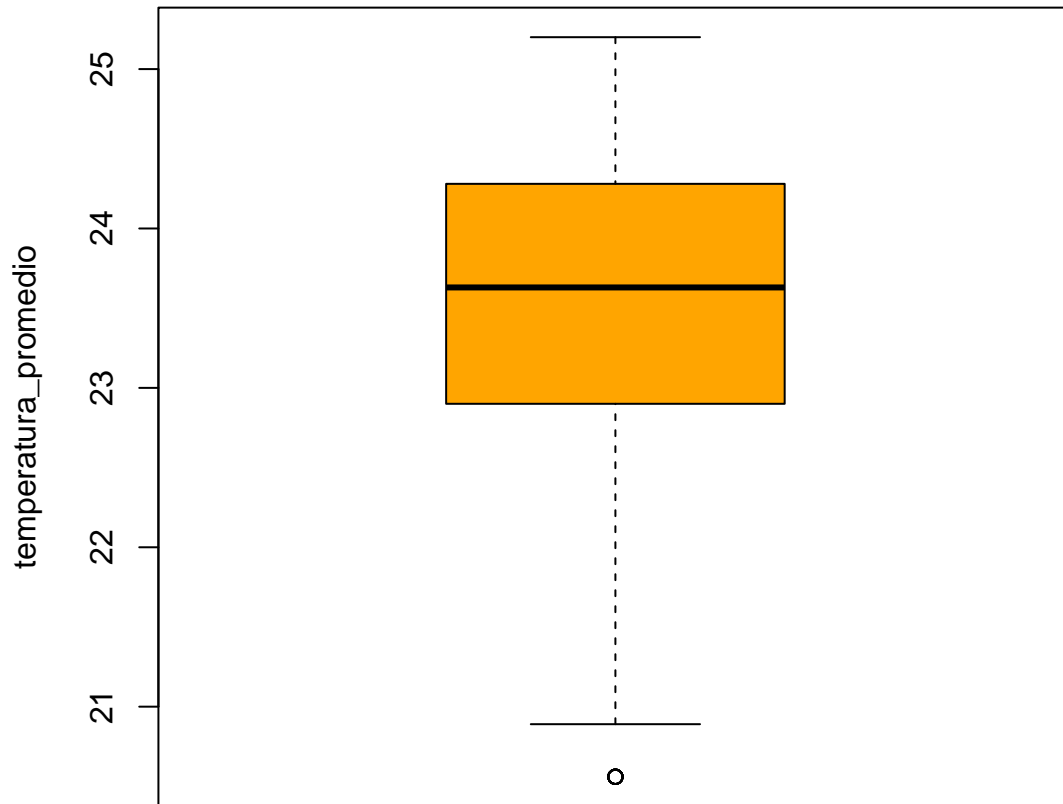
**Interpretación:** Los datos muestran una distribución que se aproxima a una forma normal con la mayoría de las temperaturas concentradas alrededor del rango medio, cerca de la mediana y la media, que están muy cercanas entre sí. Esto indica una distribución consistente y un promedio de temperatura que se mantiene estable.

```
# Boxplot para la variable numérica temperatura_promedio
```

```
boxplot(reporte$temperatura_promedio, main="Boxplot de Variable temperatura_promedio", ylab="temperatura_p
```



### Boxplot de Variable temperatura\_promedio



**Interpretación:** Se observa que la caja, que representa desde el primer cuartil hasta el tercer cuartil, abarca temperaturas aproximadamente entre 23 y 24 grados, con una línea en el medio que indica la mediana alrededor de 23.5 grados. No hay signos evidentes de valores atípicos y la distribución de los datos parece ser simétrica en torno a la mediana.

**4.2.2.2.6. Variable temperatura\_maxima** Se obtienen los datos descriptivos más representativos de la variable.

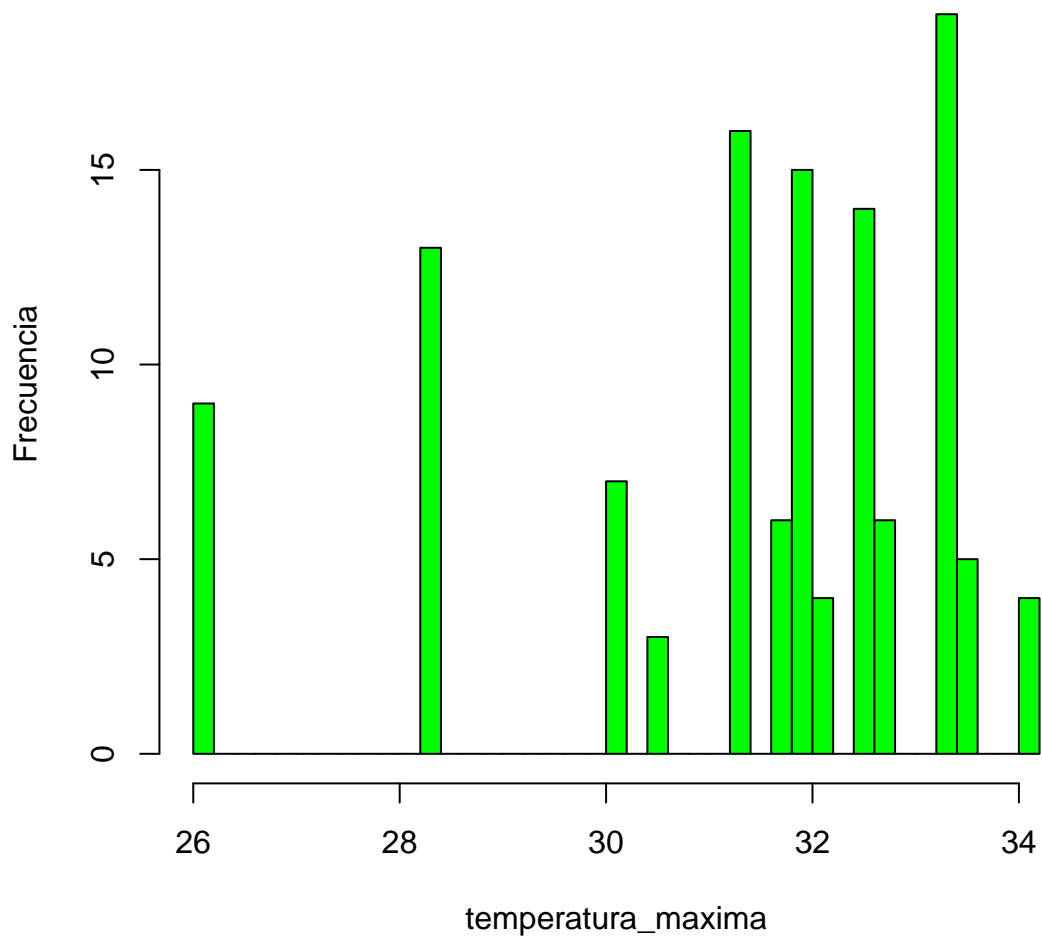
```
# Datos descriptivos de la variable
summary(reporte$temperatura_maxima)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  26.10   30.50   32.00   31.38   32.80   34.20
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica temperatura_maxima
hist(reporte$temperatura_maxima, main="Distribución de Variable temperatura_maxima", xlab="temperatura_maxima")
```

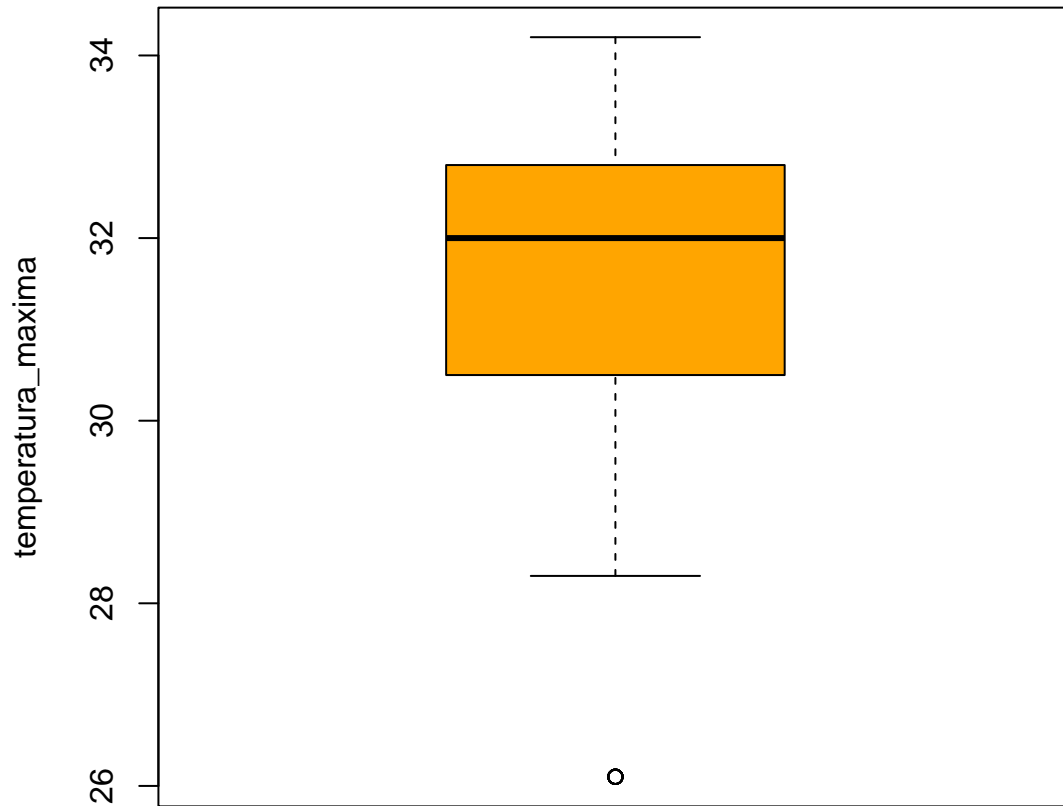
## Distribución de Variable temperatura\_maxima



**Interpretación:** El histograma de la variable muestra una distribución de frecuencias con múltiples modas, reflejando posiblemente diferentes grupos o condiciones que afectan las temperaturas máximas. Las temperaturas parecen agruparse alrededor de varios rangos, con picos notables alrededor de los valores de 28°C, 32°C y 34°C, lo que podría indicar variaciones estacionales o microclimas dentro de las áreas de cultivo.

```
# Boxplot para la variable numérica temperatura_maxima  
boxplot(reporte$temperatura_maxima, main="Boxplot de Variable temperatura_maxima", ylab="temperatura_maxima")
```

### Boxplot de Variable temperatura\_maxima



**Interpretación:** El boxplot indica que la mediana de la temperatura máxima es de 32 grados, con el 50 % de las temperaturas máximas observadas que caen entre aproximadamente 30.5 y 32.8 grados, siendo consistente con el primer y tercer cuartil del resumen estadístico. Además, el boxplot muestra que no hay valores atípicos, y el rango de temperaturas máximas se extiende desde aproximadamente 26.1 grados hasta 34.2 grados. La ausencia de valores atípicos y la concentración de los datos alrededor de la mediana sugieren una variabilidad moderada en las temperaturas máximas con un rango relativamente estrecho.

**4.2.2.2.7. Variable temperatura\_minima** Se obtienen los datos descriptivos más representativos de la variable.

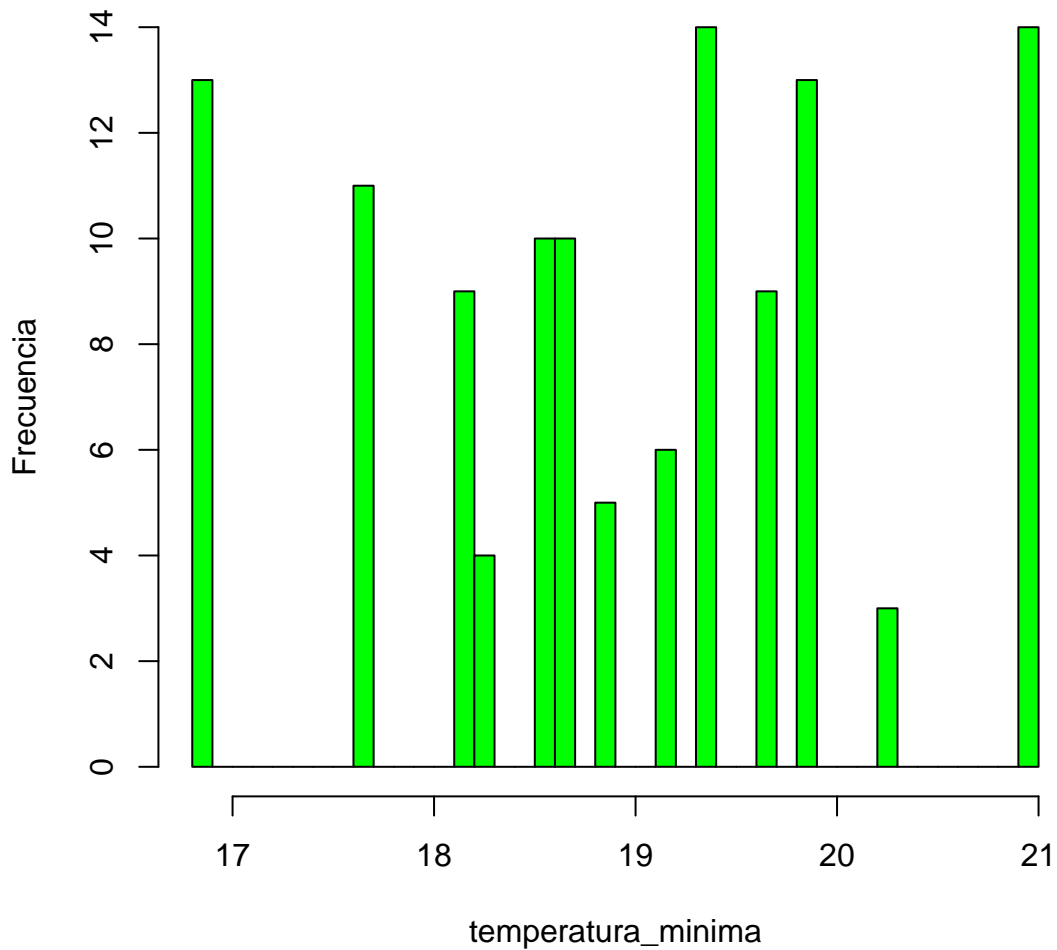
```
# Datos descriptivos de la variable
summary(reporte$temperatura_minima)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  16.80  18.20   18.90   18.97  19.70   21.00
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica temperatura_minima
hist(reporte$temperatura_minima, main="Distribución de Variable temperatura_minima", xlab="temperatura_minima")
```

## Distribución de Variable temperatura\_minima

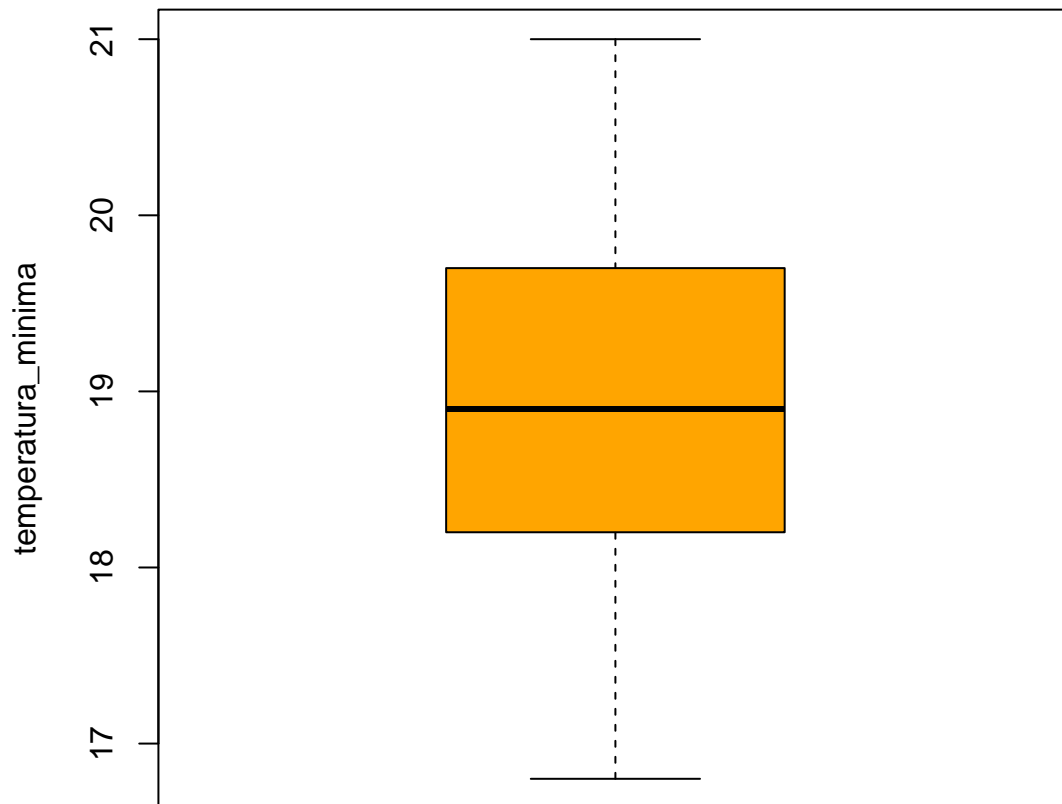


**Interpretación:** La distribución de las temperaturas mínimas es relativamente uniforme, con una concentración ligeramente mayor de frecuencias alrededor de la mediana de 18.90 grados. No hay una clara asimetría en la distribución, y no parece haber valores extremadamente altos o bajos, ya que la mayoría de las observaciones se concentran entre los cuartiles. El rango de datos está dentro de un intervalo pequeño, desde aproximadamente 17 hasta 21 grados, indicando una variabilidad limitada en las temperaturas mínimas.

```
# Boxplot para la variable numérica temperatura_minima
```

```
boxplot(reporte$temperatura_minima, main="Boxplot de Variable temperatura_minima", ylab="temperatura_minima")
```

### Boxplot de Variable temperatura\_minima



**Interpretación:** El boxplot confirma que los datos están distribuidos uniformemente alrededor de la mediana, con los cuartiles cercanos entre sí, lo que indica que no hay una gran dispersión. No hay valores atípicos en el gráfico, lo que sugiere que todas las temperaturas mínimas registradas son consistentes dentro del rango observado.

**4.2.2.2.8. Variable humedad\_promedio** Se obtienen los datos descriptivos más representativos de la variable.

```
# Datos descriptivos de la variable
```

```
summary(reporte$humedad_promedio)
```

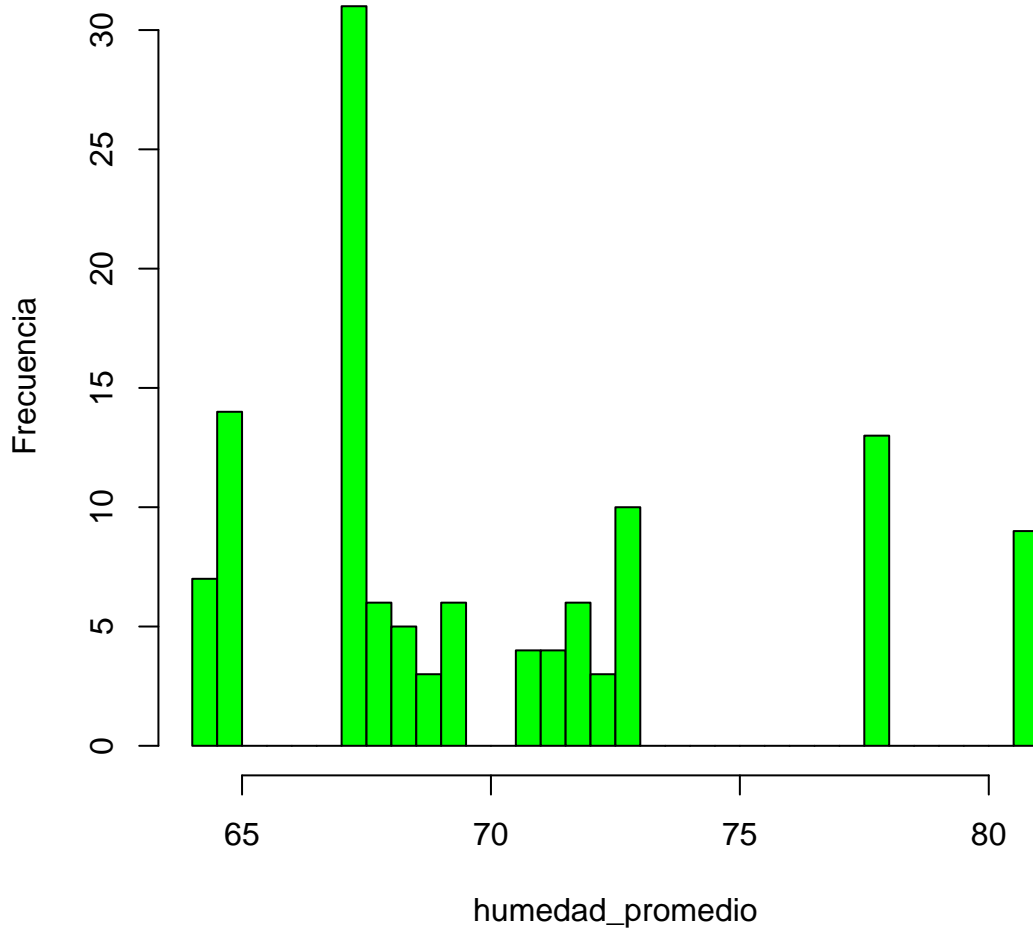
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  64.46   67.06   68.21   70.14   72.56   80.85
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica humedad_promedio
```

```
hist(reporte$humedad_promedio, main="Distribución de Variable humedad_promedio", xlab="humedad_promedio",
```

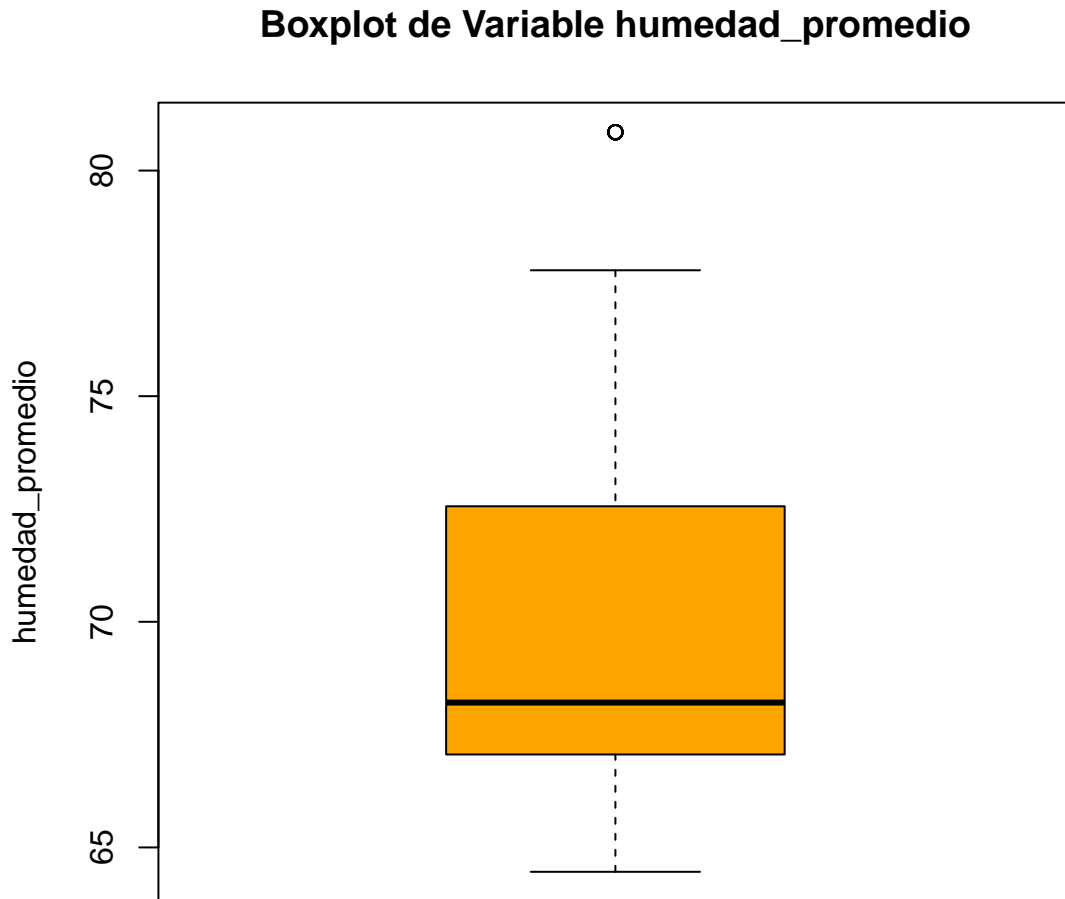
## Distribución de Variable humedad\_promedio



**Interpretación:** La distribución de la humedad promedio muestra una concentración significativa de los datos en torno al rango de 65 a 75 %, con un pico prominente alrededor del valor 67 %. La presencia de barras más pequeñas hacia el extremo derecho indica la presencia de valores más altos pero menos frecuentes.

```
# Boxplot para la variable numérica humedad_promedio
```

```
boxplot(reporte$humedad_promedio, main="Boxplot de Variable humedad_promedio", ylab="humedad_promedio", col="red")
```



**Interpretación:** El boxplot revela que la mediana está cercana al primer cuartil, lo que sugiere una distribución sesgada hacia valores menores. No hay presencia de valores atípicos inferiores y hay algunos valores atípicos superiores, lo que indica que hay algunas observaciones con humedad significativamente más alta que la mayoría de los datos.

**4.2.2.2.9. Variable humedad\_maxima** Se obtienen los datos descriptivos más representativos de la variable.

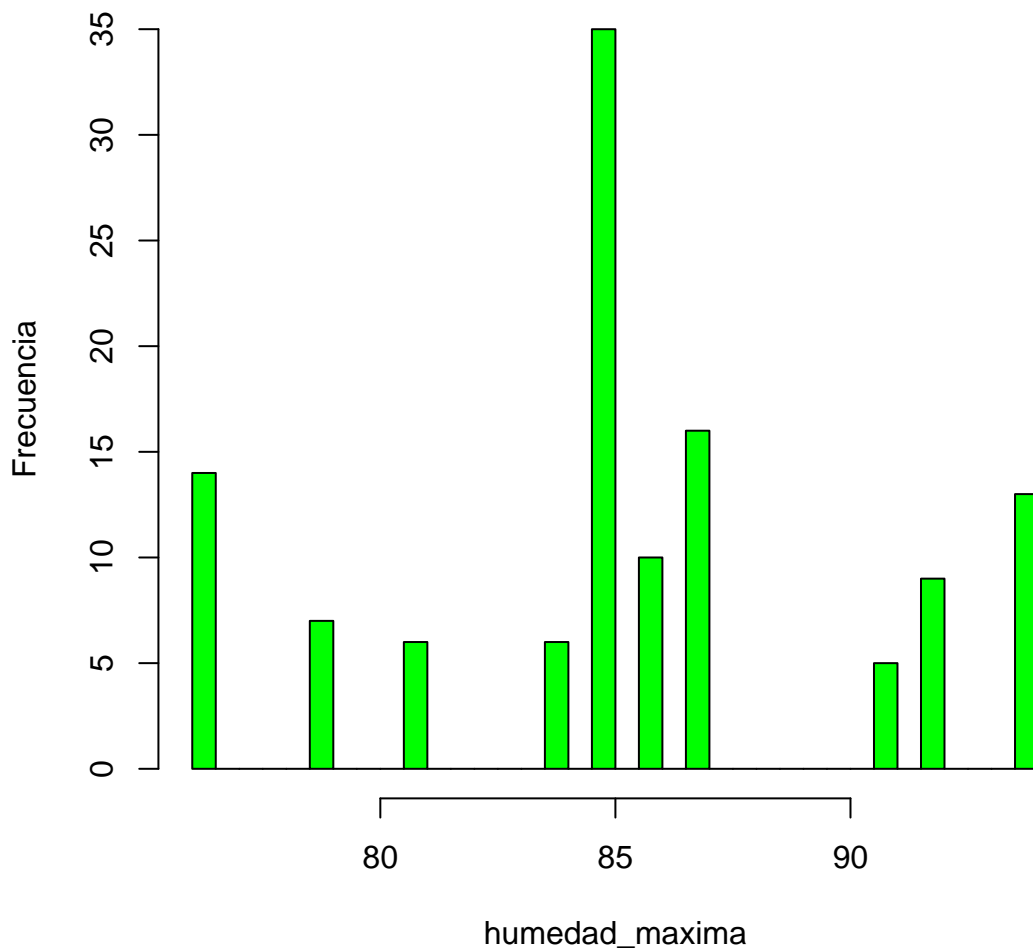
```
# Datos descriptivos de la variable
summary(reporte$humedad_maxima)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  76.00   84.00   85.00   85.45   87.00   94.00
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica humedad_maxima
hist(reporte$humedad_maxima, main="Distribución de Variable humedad_maxima", xlab="humedad_maxima", ylab="")
```

## Distribución de Variable humedad\_maxima



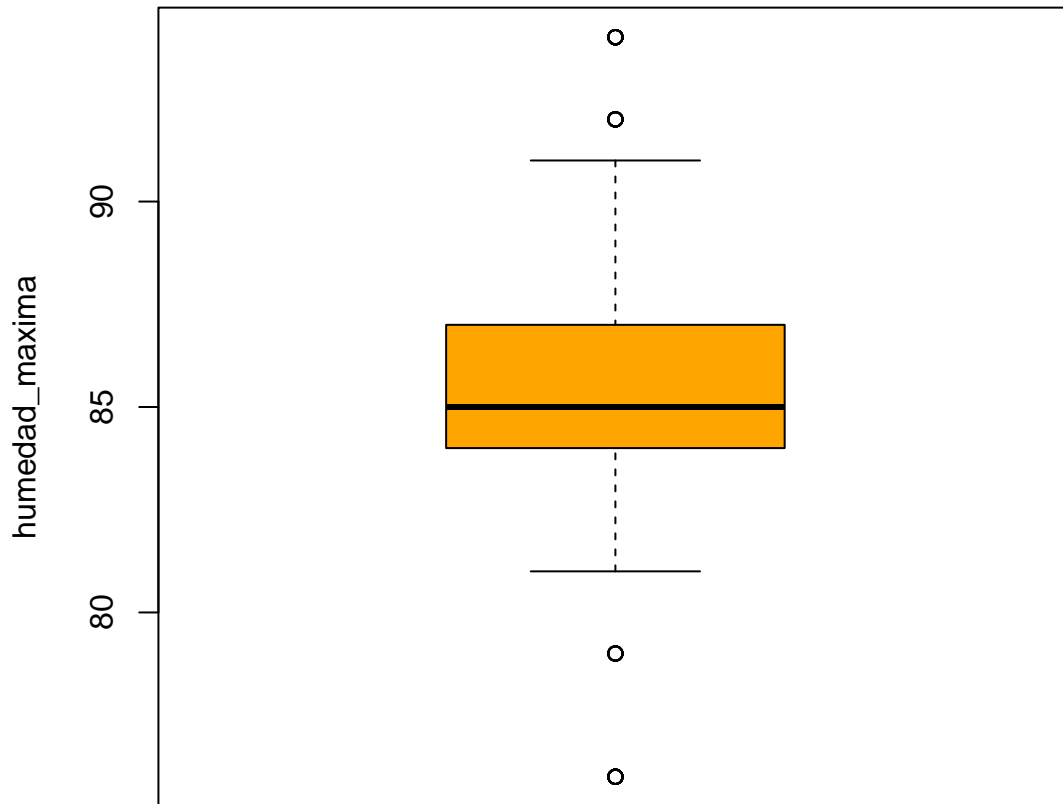
**Interpretación:** La distribución de la humedad\_maxima muestra una concentración de datos en torno al rango de 80 a 90 %, lo que se refleja en las barras altas alrededor de estos valores. Existe una frecuencia notablemente alta cerca del valor de la mediana de 85 %, indicando que una gran cantidad de observaciones se agrupa en esta área. La distribución parece sesgada hacia la izquierda, con menos frecuencias en los valores más altos de humedad máxima, superiores al tercer cuartil de 87 %.

```
# Boxplot para la variable numérica humedad_maxima
```

```
boxplot(reporte$humedad_maxima, main="Boxplot de Variable humedad_maxima", ylab="humedad_maxima", col="orange")
```



## Boxplot de Variable humedad\_maxima



**Interpretación:** La caja del boxplot se centra alrededor de la mediana, con límites que corresponden aproximadamente al primer y tercer cuartil (84 % y 87 %, respectivamente), lo que indica que la mitad de las observaciones de la humedad máxima están contenidas dentro de este rango. Los bigotes muestran la variabilidad fuera del cuartil superior e inferior. La línea superior se extiende hasta el valor máximo (94 %), mientras que la línea inferior se detiene en el mínimo (76 %), indicando que no hay valores extremadamente bajos más allá del mínimo. Hay puntos fuera del bigote superior, lo que indica la presencia de valores atípicos por encima del 87 %. Estos valores atípicos están más allá de la variabilidad esperada basada en los cuartiles y podrían ser objeto de una investigación adicional para entender su causa.

**4.2.2.2.10. Variable humedad\_minima** Se obtienen los datos descriptivos más representativos de la variable.

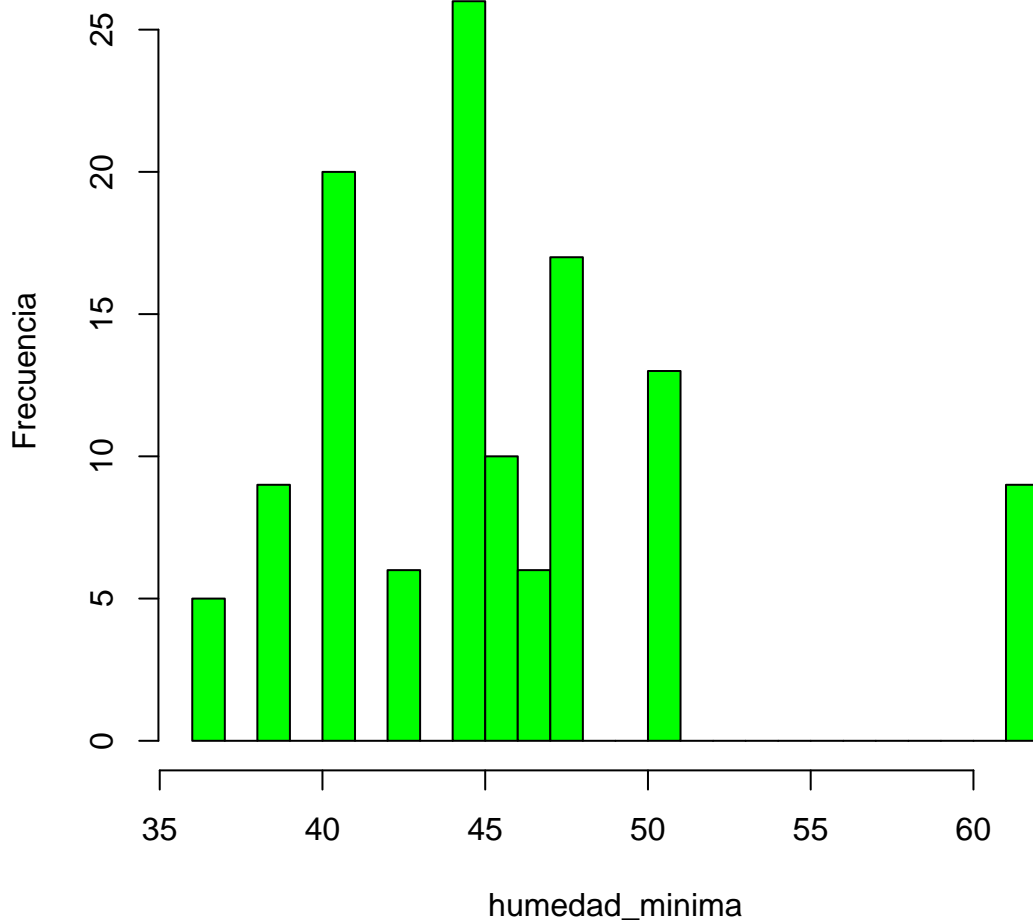
```
# Datos descriptivos de la variable
summary(reporte$humedad_minima)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  36.00  41.00  45.00  45.93  48.00  62.00
```

Se continua con la generación de gráficos.

```
# Histograma para la variable numérica humedad_minima
hist(reporte$humedad_minima, main="Distribución de Variable humedad_minima", xlab="humedad_minima", ylab="")
```

## Distribución de Variable humedad\_minima

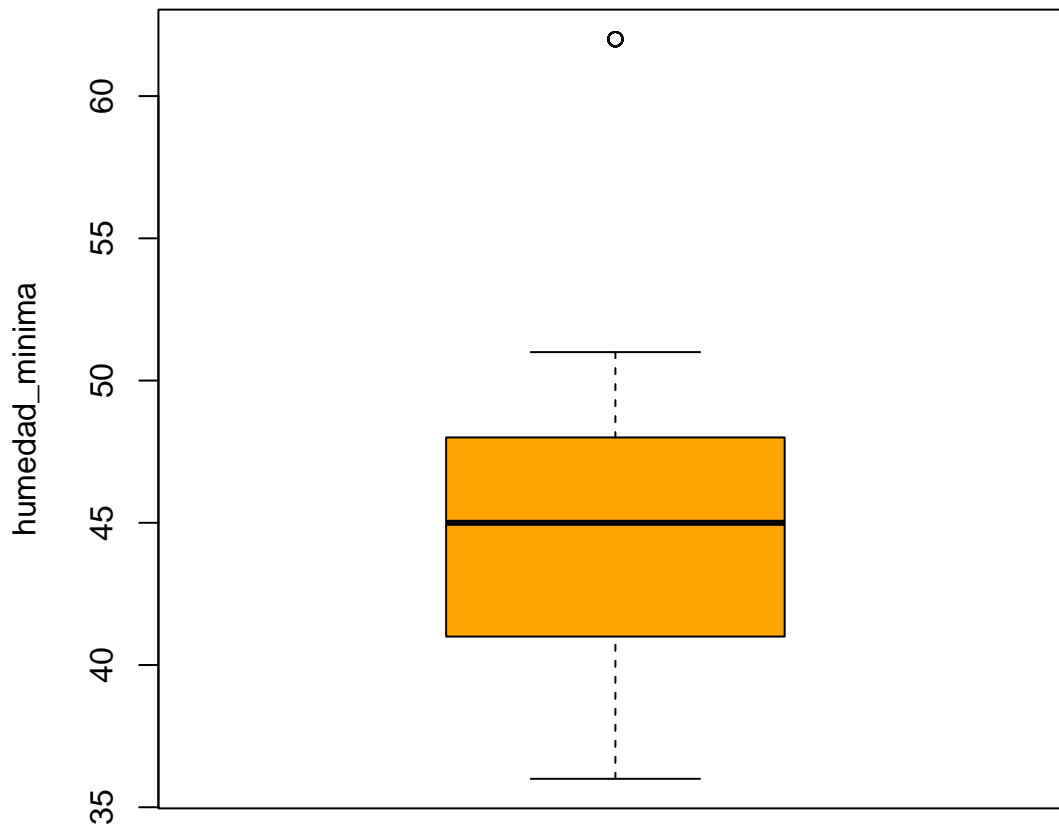


**Interpretación:** Muestra una distribución con una tendencia hacia los valores más bajos de humedad, con la mayor frecuencia de datos entre 40 y 50 %, lo cual es consistente con la mediana de 45 %. Hay menos frecuencia conforme la humedad disminuye por debajo de 40 % o aumenta por encima de 50 %, lo que sugiere una concentración de los datos en torno a los valores intermedios de humedad.

```
# Boxplot para la variable numérica humedad_minima
```

```
boxplot(reporte$humedad_minima, main="Boxplot de Variable humedad_minima", ylab="humedad_minima", col="orange")
```

### Boxplot de Variable humedad\_minima



**Interpretación:** Muestra que la mediana está en 45 %, con la caja extendiéndose desde el primer cuartil en 41 % hasta el tercer cuartil en 48 %, lo que indica que la mitad de los datos se agrupa en este rango. Los bigotes se extienden desde el mínimo en 36 % hasta el máximo en 62 %, lo que sugiere algunos valores fuera del rango intercuartil, pero no demasiado extremos, ya que no hay valores atípicos presentes en el gráfico.

**4.2.2.2.11. Variable dosis\_litros\_por\_hectarea** Se obtienen los datos descriptivos más representativos de la variable.

```
# Datos descriptivos de la variable
summary(reporte$dosis_litros_por_hectarea)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.000   1.000   2.000   1.986   2.500   4.000
```

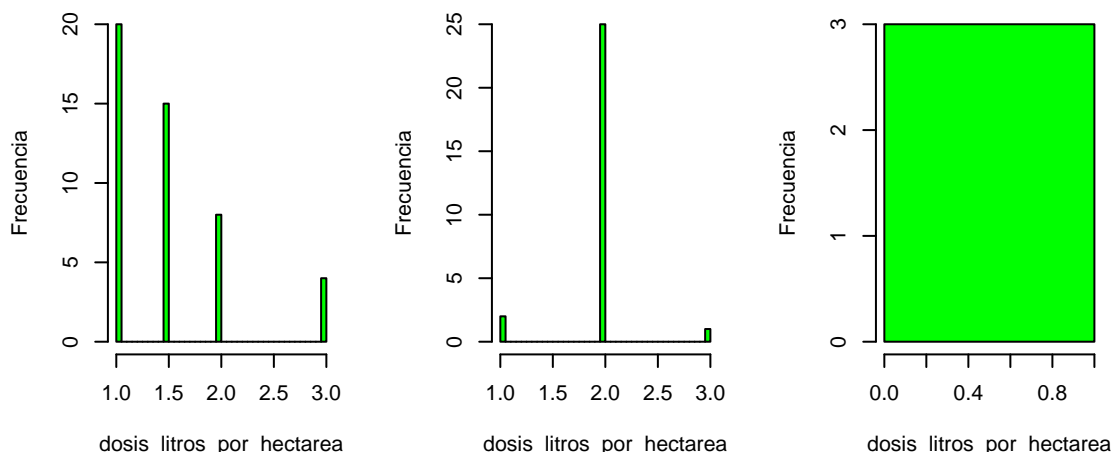
Se continua con la generación de gráficos.

```
# Genera un histograma por insumo, de los litros por hectárea utilizados.
reporte_por_tipo_insumo <- split(reporte$dosis_litros_por_hectarea, reporte$tipo_insumo)

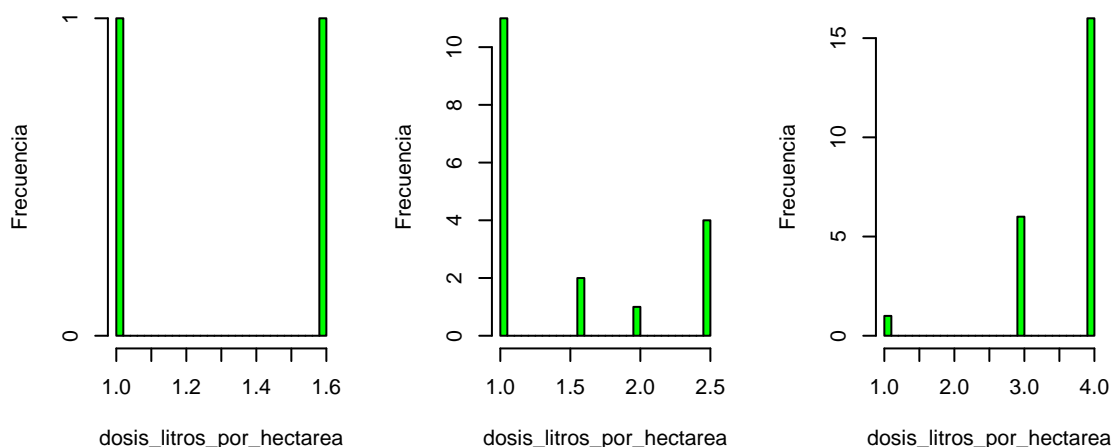
par(mfrow=c(2, 3))
```

```
for (i in 1:length(reporte_por_tipo_insumo)) {
  hist(reporte_por_tipo_insumo[[i]], main=paste("Distribución de dosis_litros_por_hectarea (Tipo de Insumo", i, ")"),
       xlab="dosis_litros_por_hectarea", ylab="Frecuencia", col="green", breaks=30)
}
```

is\_litros\_por\_hectarea (Tipo de Insumo: Fertilizante) dosis\_litros\_por\_hectarea (Tipo de Insumo: Herbicida) dosis\_litros\_por\_hectarea (Tipo de Insumo: Insecticida)



s\_litros\_por\_hectarea (Tipo de Insumo: Fungicida) dosis\_litros\_por\_hectarea (Tipo de Insumo: Nematicida) dosis\_litros\_por\_hectarea (Tipo de Insumo: Coadyuvante)



**Interpretación:** Los histogramas ilustran la distribución de la variable `dosis_litros_por_hectarea` para seis tipos diferentes de insumos (Fertilizante, Herbicida, Insecticida, Fungicida, Nematicida y Coadyuvante), mostrando principalmente distribuciones unimodales con diferentes grados de sesgo y variabilidad. Algunos insumos como Fertilizante, Herbicida, Insecticida y Fungicida muestran distribuciones unimodales ligeramente sesgadas hacia la derecha con menor variabilidad, mientras que Nematicida presenta una distribución unimodal simétrica con mayor variabilidad, y Coadyuvante exhibe una distribución unimodal simétrica con la mayor variabilidad.

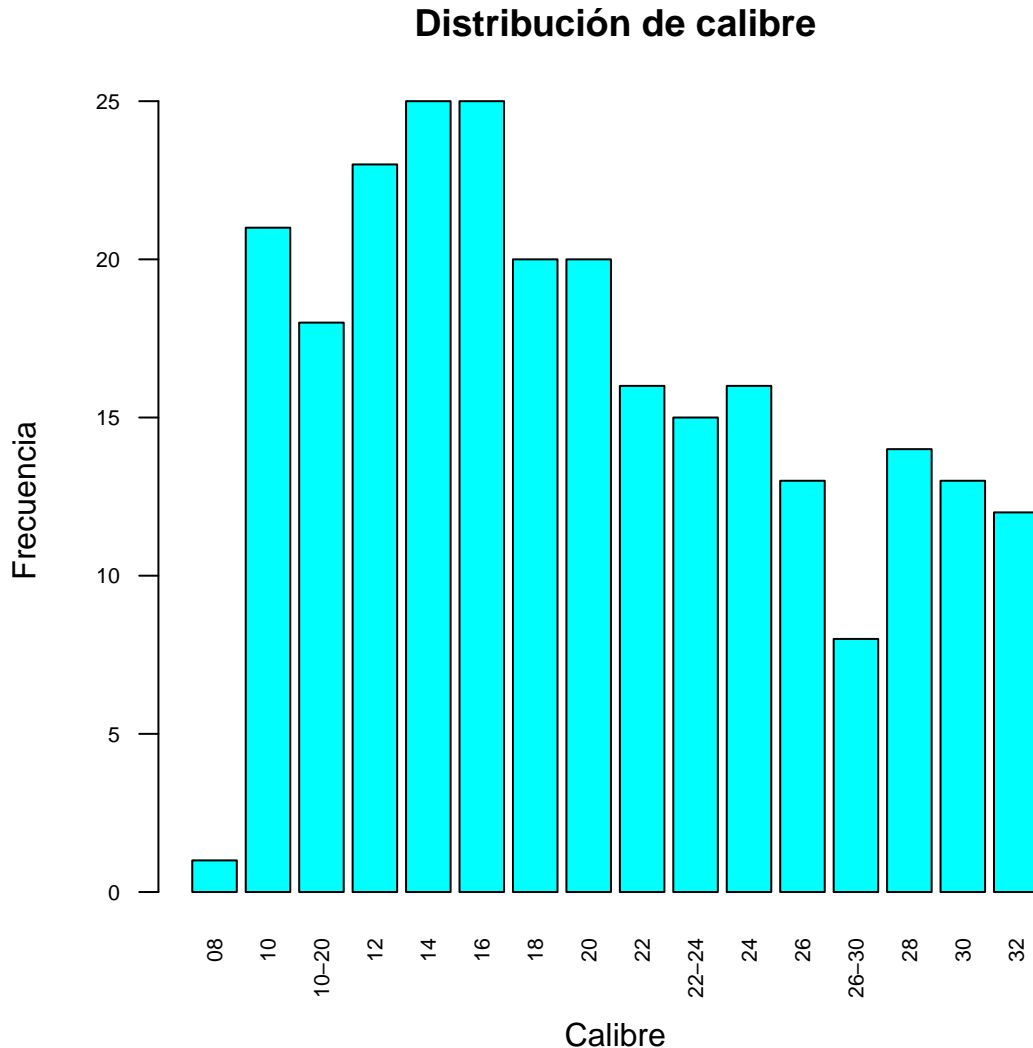
#### 4.2.3. Visualización de variables categóricas

##### 4.2.3.1. Tabla de frecuencia y Barplot para las variables categóricas más relevantes del dataset 'calibre'

```
# Tabla de frecuencia para la variable categórica calibre
calibres_categoria_freq <- table(calibres$calibre)
```

```
# Barplot para la variable categórica calibre
par(cex.axis=0.7)
barplot(calibres_categoria_freq, main="Distribución de calibre", col="cyan", las=2, xlab="Calibre", ylab="Frecuencia")
```

#### 4.2.3.1.1. Variable calibre



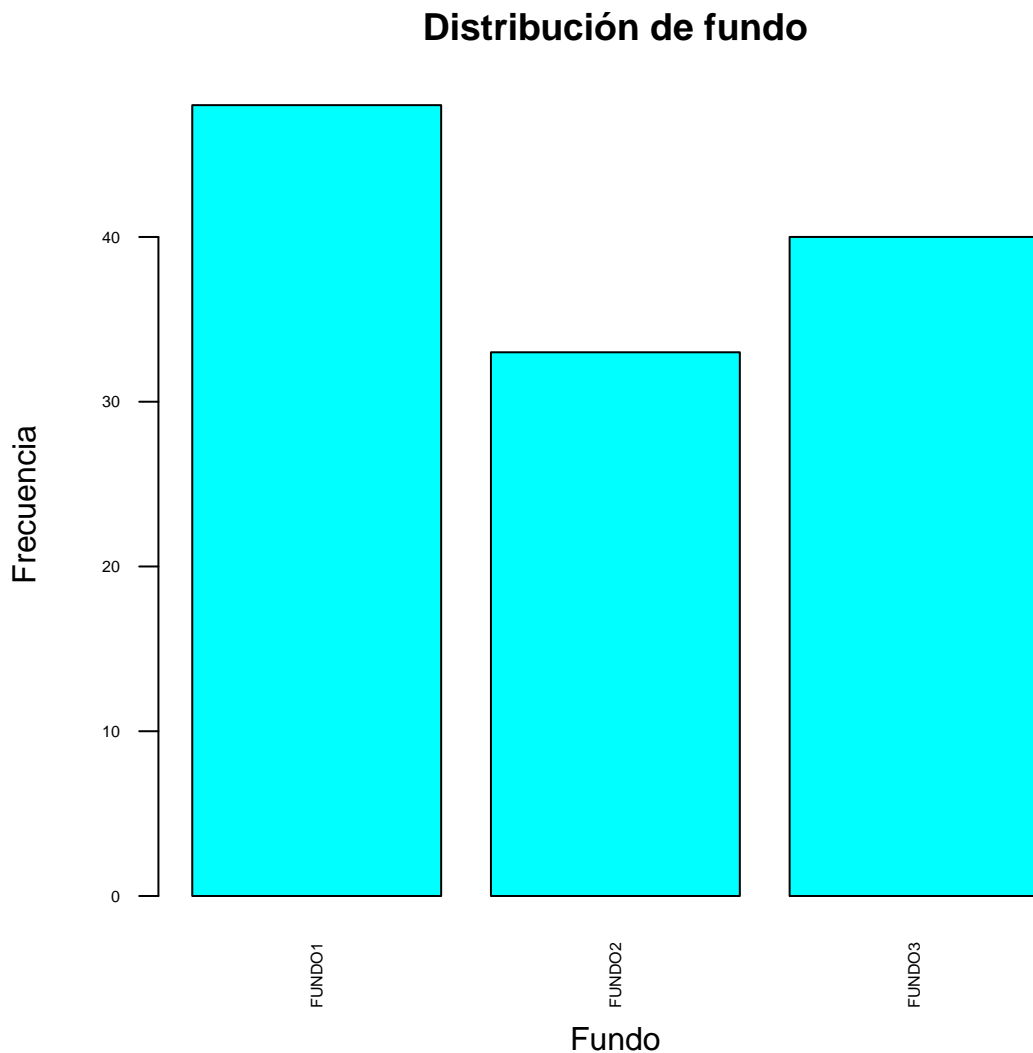
**Interpretación:** Las categorías de calibre se distribuyen a lo largo de un rango que va desde 08 hasta 32. Existe una mayor frecuencia de calibres en el rango de 14 a 20, con estos mostrando las barras más altas, lo que indica que son los más comunes en el conjunto de datos. Las categorías de calibre 08 y 10-20 tienen una frecuencia notablemente baja en comparación con las demás, lo que sugiere que son menos comunes o que muy pocos elementos en el conjunto de datos tienen estos calibres. A medida que el calibre aumenta más allá de 20, se observa una disminución general en la frecuencia, aunque con algunas variaciones, como un ligero aumento en el rango de 22-24 y una caída en 26-30. No se observa una disminución consistente o un patrón uniforme a medida que el calibre aumenta, lo que podría indicar variabilidad en la distribución o en la prevalencia de ciertos calibres en el conjunto de datos.

#### 4.2.3.2. Tabla de frecuencia y Barplot para las variables categóricas más relevantes del dataset Reporte

```
# Tabla de frecuencia para la variable categórica fundo
reporte_categoria_freq <- table(reporte$fundo)
```

```
# Barplot para la variable categórica fundo
par(cex.axis=0.5)
barplot(reporte_categoria_freq, main="Distribución de fundo", col="cyan", las=2, xlab="Fundo", ylab="Frecuencia")
```

#### 4.2.3.2.1. Variable fundo

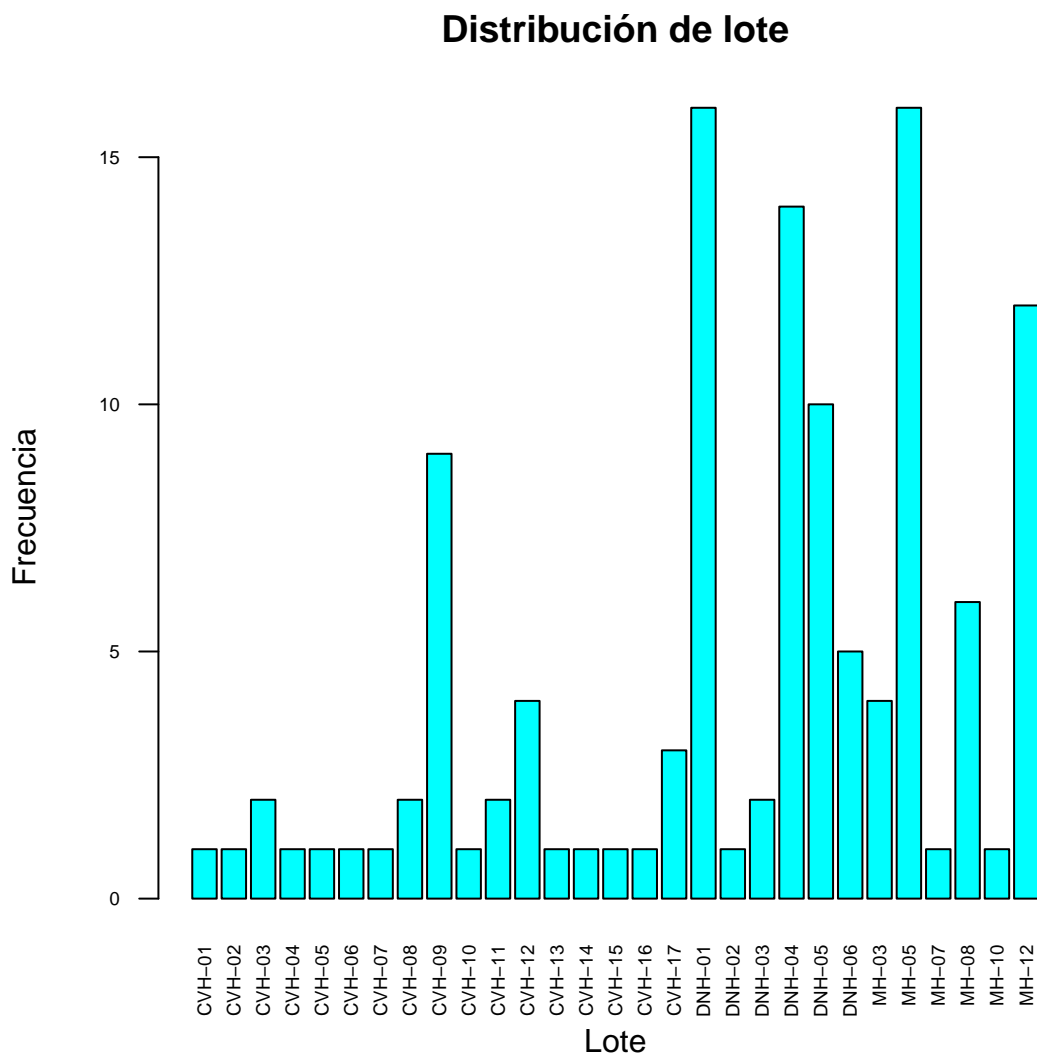


**Interpretación:** Hay tres categorías de fundo representadas: FUNDO1, FUNDO2 y FUNDO3. FUNDO1 tiene la mayor frecuencia, indicando que hay más registros o datos asociados a esta categoría en el conjunto de datos. FUNDO2 tiene una frecuencia significativamente menor en comparación con FUNDO1, lo que indica menos registros asociados a esta categoría. FUNDO3 tiene una frecuencia similar a FUNDO1, lo cual sugiere una cantidad de registros comparable con la primera categoría.

```
# Tabla de frecuencia para la variable categórica lote
reporte_categoria_freq <- table(reporte$lote)

# Barplot para la variable categórica lote
par(cex.axis=0.6)
barplot(reporte_categoria_freq, main="Distribución de lote", col="cyan", las=2, xlab="Lote", ylab="Frecuencia")
```

#### 4.2.3.2.2. Variable lote

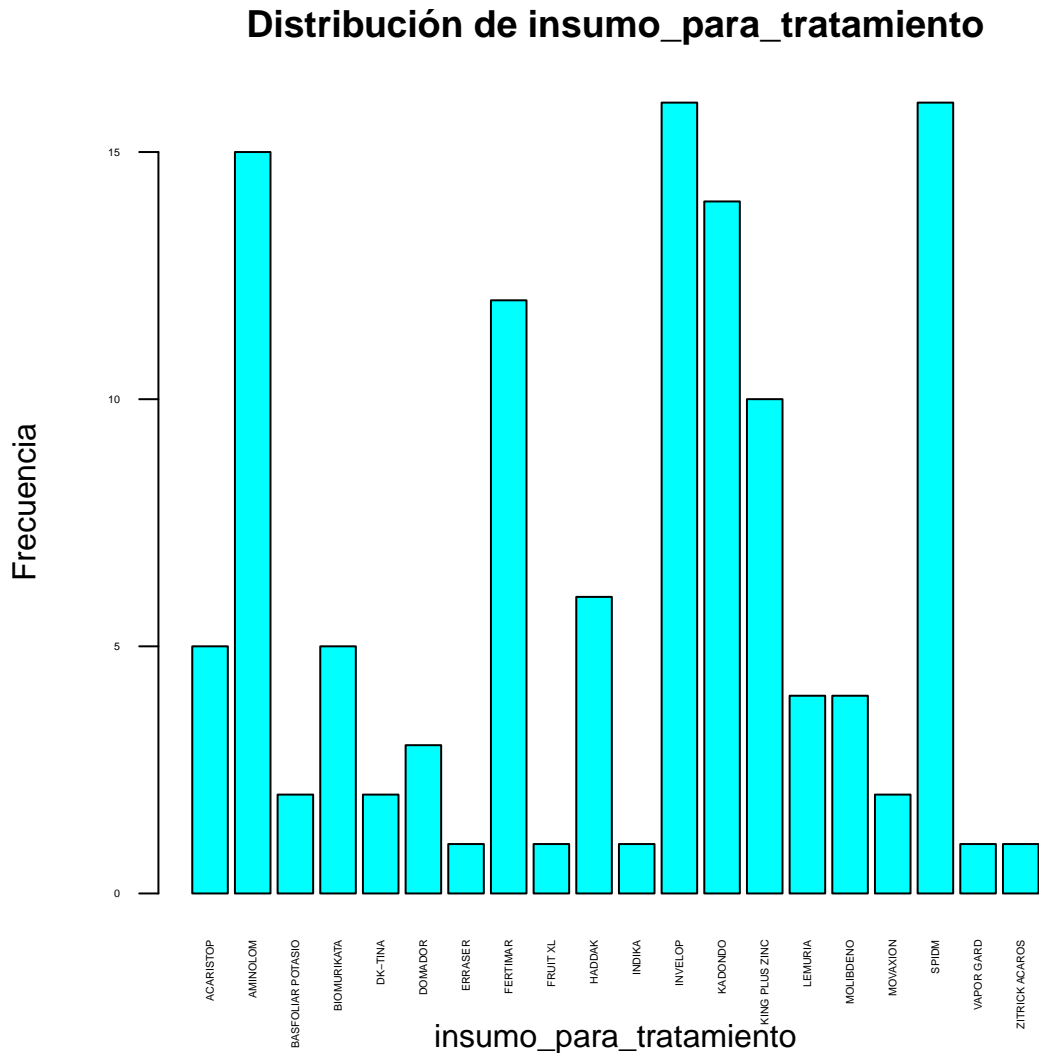


**Interpretación:** La variable lote presenta una variedad de categorías, que parecen estar codificadas con combinaciones de letras y números (como CVH-01, DNH-03, MH-10, etc.). Las frecuencias de las diferentes categorías de lote varían significativamente, con algunas categorías mostrando una frecuencia mucho mayor que otras. Esto sugiere que algunos lotes tienen más registros asociados que otros. No se observa un patrón uniforme en las frecuencias, indicando que la distribución de registros por lote no es homogénea.

```
# Tabla de frecuencia para la variable categórica insumo_para_tratamiento
reporte_categoria_freq <- table(reporte$insumo_para_tratamiento)

# Barplot para la variable categórica insumo_para_tratamiento
par(cex.axis=0.3)
barplot(reporte_categoria_freq, main="Distribución de insumo_para_tratamiento", col="cyan", las=2, xlab="i
```

#### 4.2.3.2.3. Variable insumo\_para\_tratamiento



**Interpretación:** Las categorías de insumos tienen frecuencias variadas, lo que indica que algunos insumos se utilizan más comúnmente que otros. Algunos insumos, como los representados por las barras más altas, destacan por su uso frecuente, lo que podría indicar una preferencia o mayor necesidad de estos dentro del contexto de los tratamientos. Por el contrario, los insumos representados por barras más cortas se usan menos, lo que podría deberse a que son menos conocidos, más caros, o quizás solo son necesarios para aplicaciones específicas.

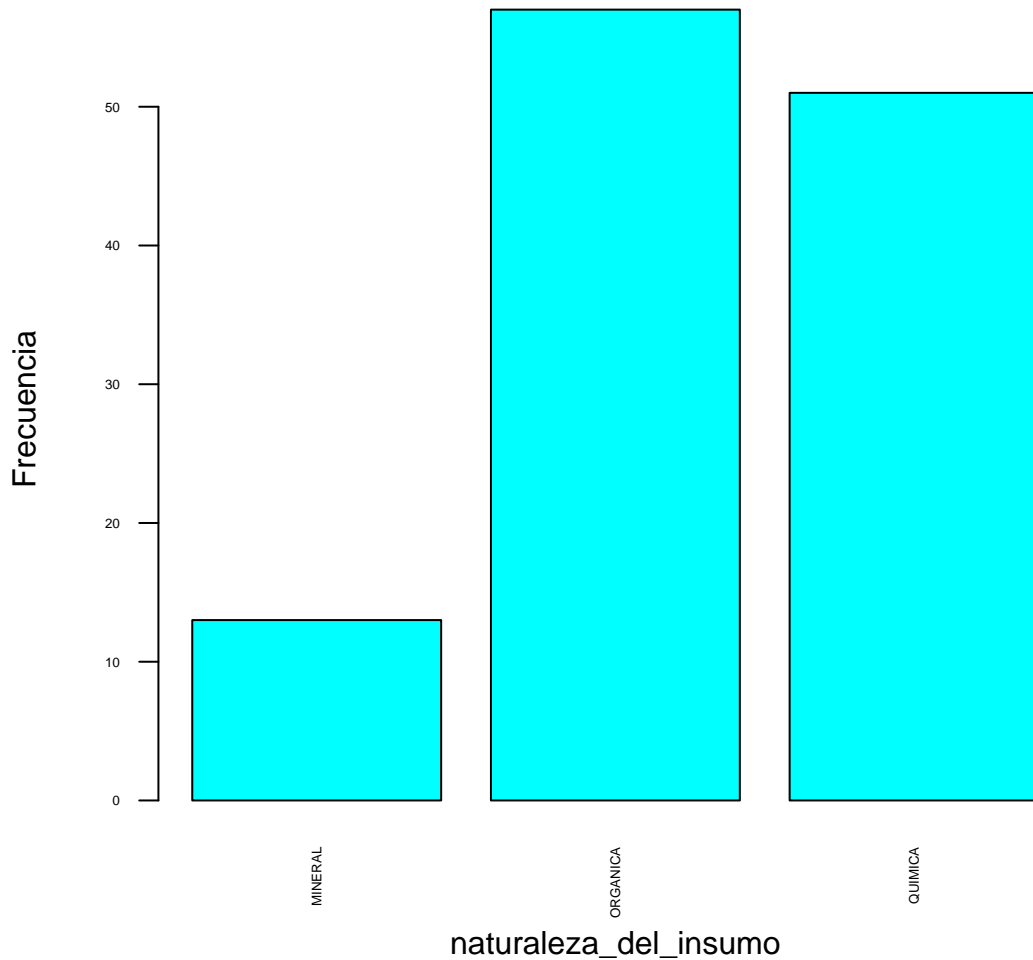
```
# Tabla de frecuencia para la variable categórica naturaleza_del_insumo
reporte_categoria_freq <- table(reporte$naturaleza_del_insumo)

# Barplot para la variable categórica naturaleza_del_insumo
par(cex.axis=0.4)
barplot(reporte_categoria_freq, main="Distribución de naturaleza_del_insumo", col="cyan", las=2, xlab="naturaleza_del_insumo")
```

#### 4.2.3.2.4. Variable naturaleza\_del\_insumo



## Distribución de naturaleza\_del\_insumo



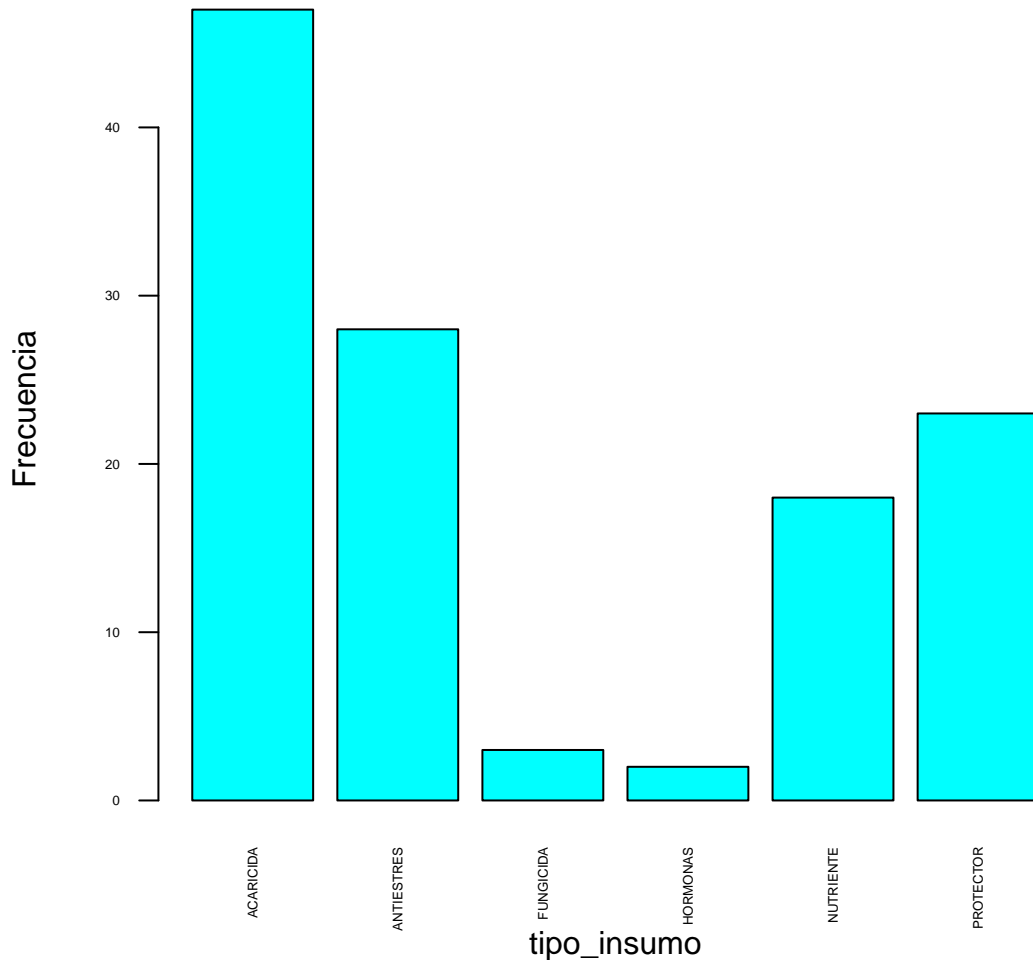
**Interpretación:** La frecuencia de uso de insumos de naturaleza orgánica es la más alta, lo que indica que estos son los más comunes o preferidos en este conjunto de datos. Los insumos de naturaleza química también tienen una frecuencia alta, sugiriendo que también son comúnmente usados. Por otro lado, los insumos de naturaleza mineral tienen la frecuencia más baja, lo que implica que se utilizan con menos frecuencia en comparación con los orgánicos y químicos.

```
# Tabla de frecuencia para la variable categórica tipo_insumo
reporte_categoria_freq <- table(reporte$tipo_insumo)

# Barplot para la variable categórica tipo_insumo
par(cex.axis=0.4)
barplot(reporte_categoria_freq, main="Distribución de tipo_insumo", col="cyan", las=2, xlab="tipo_insumo",
```

### 4.2.3.2.5. Variable tipo\_insumo

## Distribución de tipo\_insumo



**Interpretación:** Los ‘ACARICIDA’ son los más frecuentemente utilizados, con la barra más alta en el gráfico. ‘NUTRIENTE’ y ‘PROTECTOR’ también muestran una alta frecuencia, aunque menos que ‘ACARICIDA’. Por el contrario, ‘FUNGICIDA’ y ‘HORMONAS’ tienen una frecuencia mucho menor en comparación con los demás.

### 4.3. Análisis Específico por Cliente

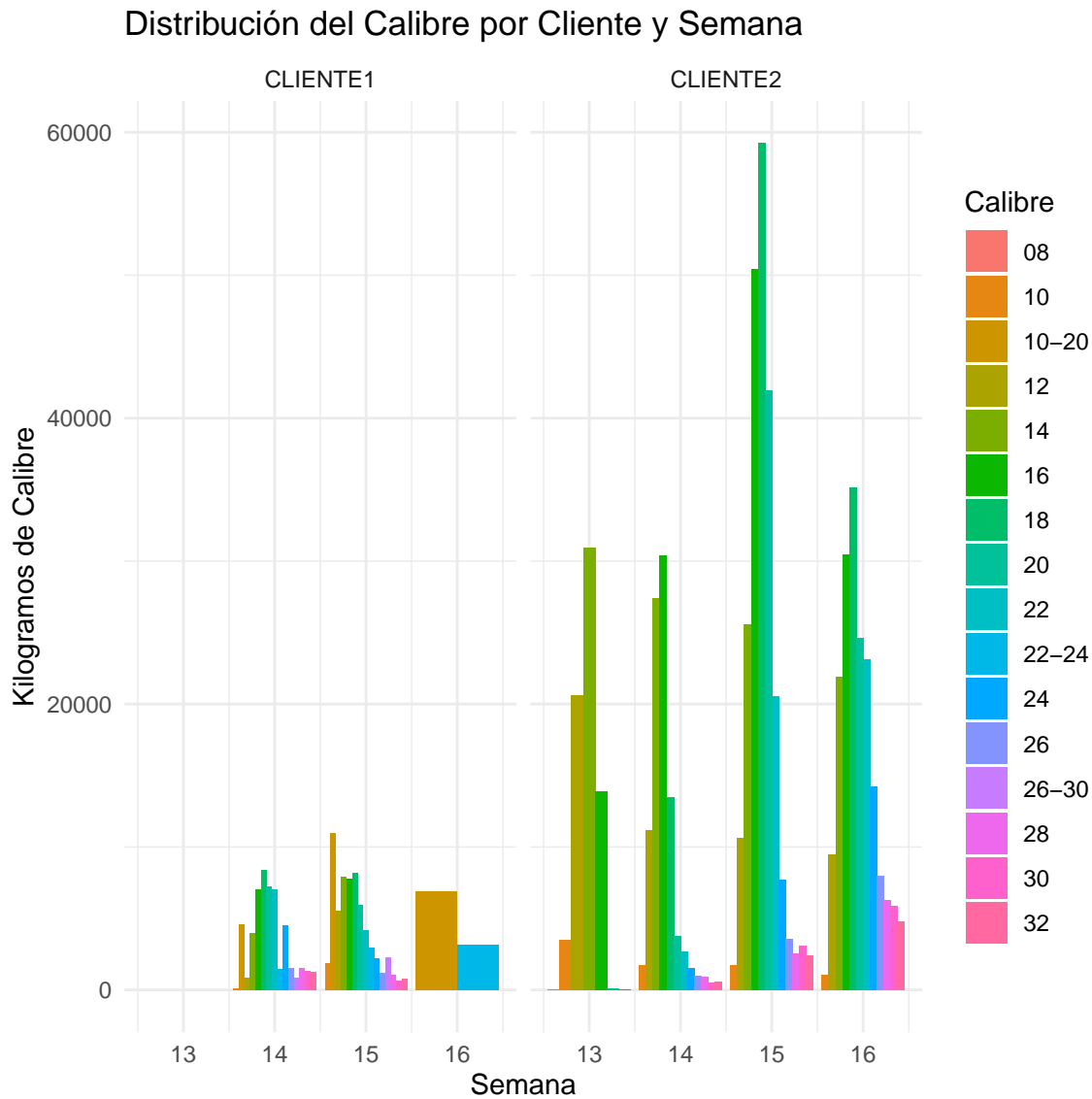
#### 4.3.1. Exploración de datos de calibre por cliente y por semana

**4.3.1.1. Distribución de Kilogramos de Calibre de Paltas Hass por Cliente y Semana** Se genera un gráfico de barras para visualizar y comparar la cantidad total de kilogramos por cada calibre de paltas Hass distribuidos a dos clientes diferentes a lo largo de un período de cuatro semanas.

```
# Verificar que los datos están agrupados y sumariosados adecuadamente
calibres_agrupados <- aggregate(kg_calibre ~ cliente + semana + calibre, data=calibres, sum)

# Gráfico de barras para visualizar la distribución del calibre por cliente a lo largo de las semanas
ggplot(calibres_agrupados, aes(x=semana, y=kg_calibre, fill=calibre)) +
  geom_bar(stat="identity", position="dodge") +
  facet_wrap(~cliente) +
  labs(title="Distribución del Calibre por Cliente y Semana",
       x="Semana",
```

```
y="Kilogramos de Calibre",
fill="Calibre") +
theme_minimal()
```

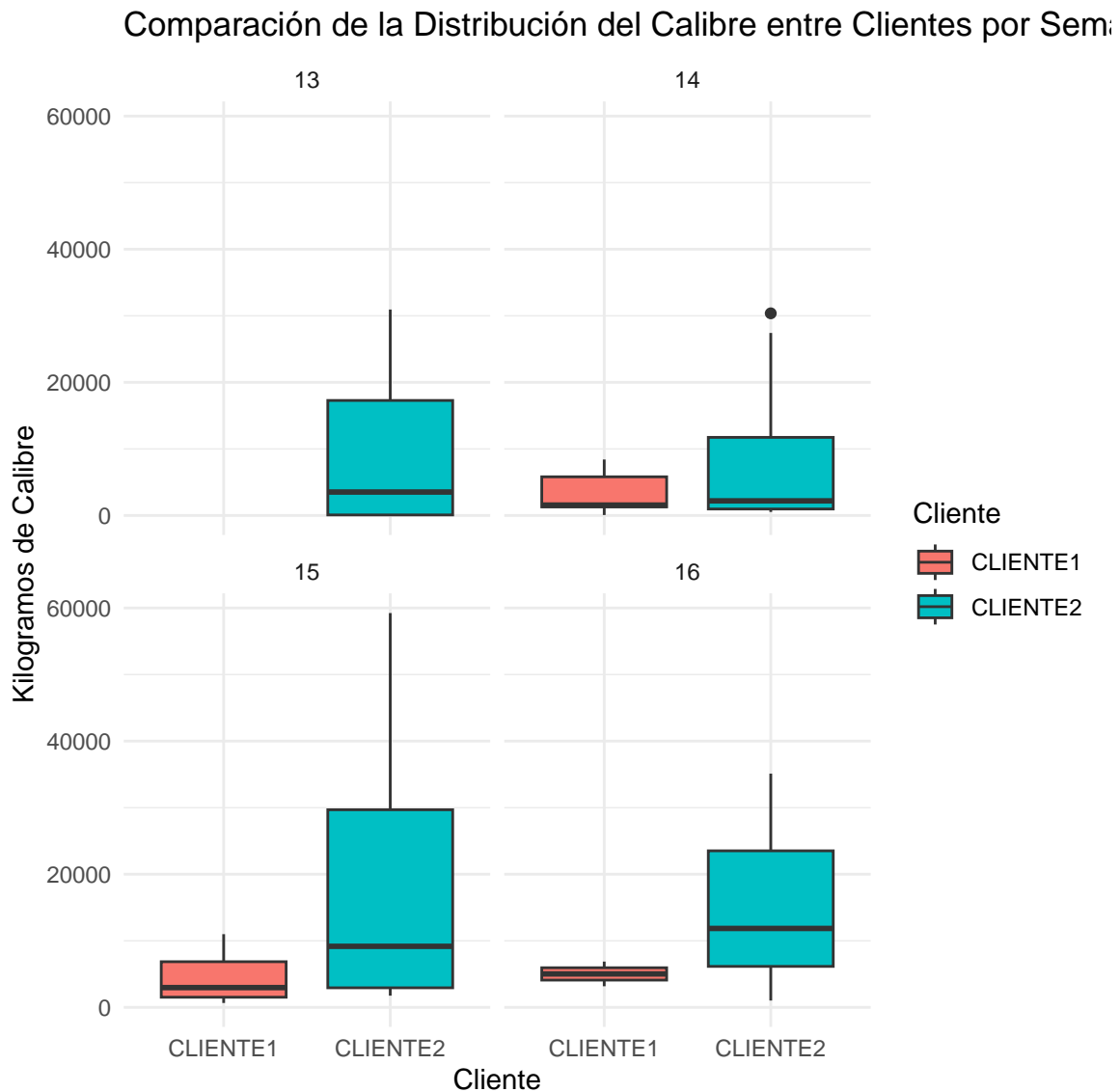


**Interpretación:** Se observa que el cliente 2 tiene una mayor variabilidad en los kilogramos por calibre en comparación con el cliente 1, con un pico significativo en la semana 15 para calibres entre 22 y 24. El cliente 1 muestra una distribución más uniforme semana a semana. Ambos clientes tienen una presencia de calibres en el rango de 08 a 32, con calibres de 12 y 14 comunes en el cliente 1 y calibres de 22-24 destacándose en el cliente 2. Esto puede indicar preferencias de compra o disponibilidad de producto distintas entre los dos clientes. La semana 15 se resalta como un período de alta producción o entrega, especialmente para el cliente 2, lo que podría estar influenciado por condiciones estacionales específicas o por una demanda de mercado incrementada.

**4.3.1.2. Comparación de la Distribución de Kilogramos de Calibre entre Clientes por Semana** Se genera un boxplot para examinar y contrastar la variabilidad en la distribución de los kilogramos por calibre entregados a dos clientes diferentes a lo largo de un mes

```
# Boxplot para comparar la distribución del calibre entre clientes por semana
ggplot(calibres_agrupados, aes(x=cliente, y=kg_calibre, fill=cliente)) +
  geom_boxplot() +
  facet_wrap(~semana) +
```

```
labs(title="Comparación de la Distribución del Calibre entre Clientes por Semana",
     x="Cliente",
     y="Kilogramos de Calibre",
     fill="Cliente") +
theme_minimal()
```



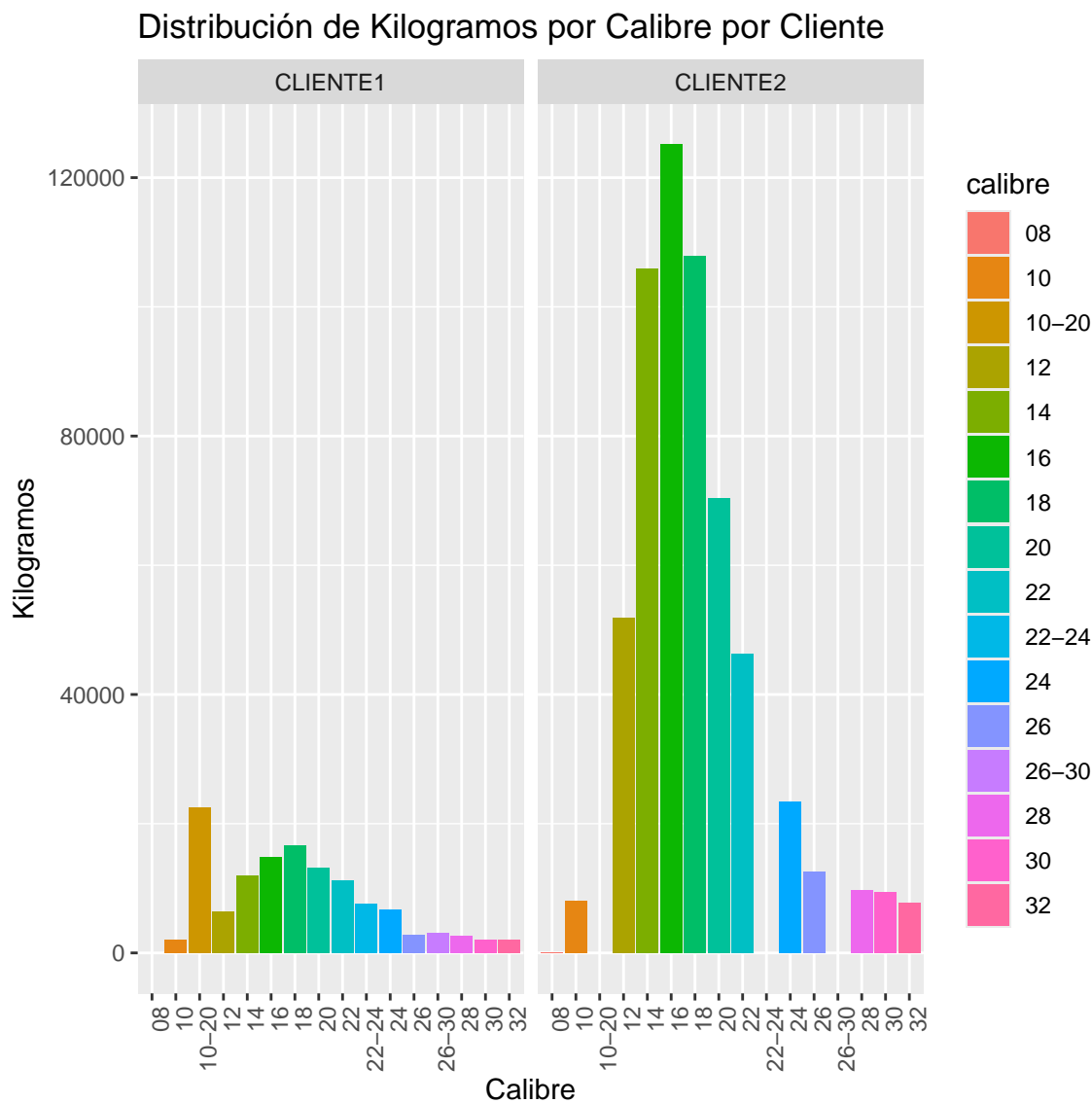
**Interpretación:** La variabilidad en la cantidad de kilogramos por calibre entregada es más pronunciada para el cliente 2 que para el cliente 1, como se evidencia por el rango intercuartílico más amplio y los valores extremos en varias semanas. Mientras que el cliente 1 muestra una consistencia en los medianos volúmenes de entrega con poca variación semanal, el cliente 2 exhibe una dispersión significativa especialmente en las semanas 14 y 15. Esto podría reflejar una demanda más dinámica o patrones de pedido variables para el cliente 2. La semana 15 destaca nuevamente para el cliente 2 con un rango intercuartílico y valores extremos superiores, indicando entregas de calibres de alto volumen que podrían ser indicativos de una cosecha abundante o una demanda pico específica en esa semana.

#### 4.3.2. Visualización de la distribución de calibres por cliente

**4.3.2.1. Distribución de Kilogramos por Calibre de Paltas Hass por Cliente** Se genera este gráfico de barras para mostrar la distribución total de kilogramos de paltas Hass por calibre para cada cliente, proporcionando una comparación visual directa entre los dos.

```
# Sumamos los kilogramos por cliente y calibre
calibres_sum <- aggregate(kg_calibre ~ cliente + calibre, data=calibres, sum)

# Creamos un barplot para cada cliente
ggplot(calibres_sum, aes(x=calibre, y=kg_calibre, fill=calibre)) +
  geom_bar(stat="identity") +
  facet_wrap(~cliente) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x="Calibre", y="Kilogramos", title="Distribución de Kilogramos por Calibre por Cliente")
```

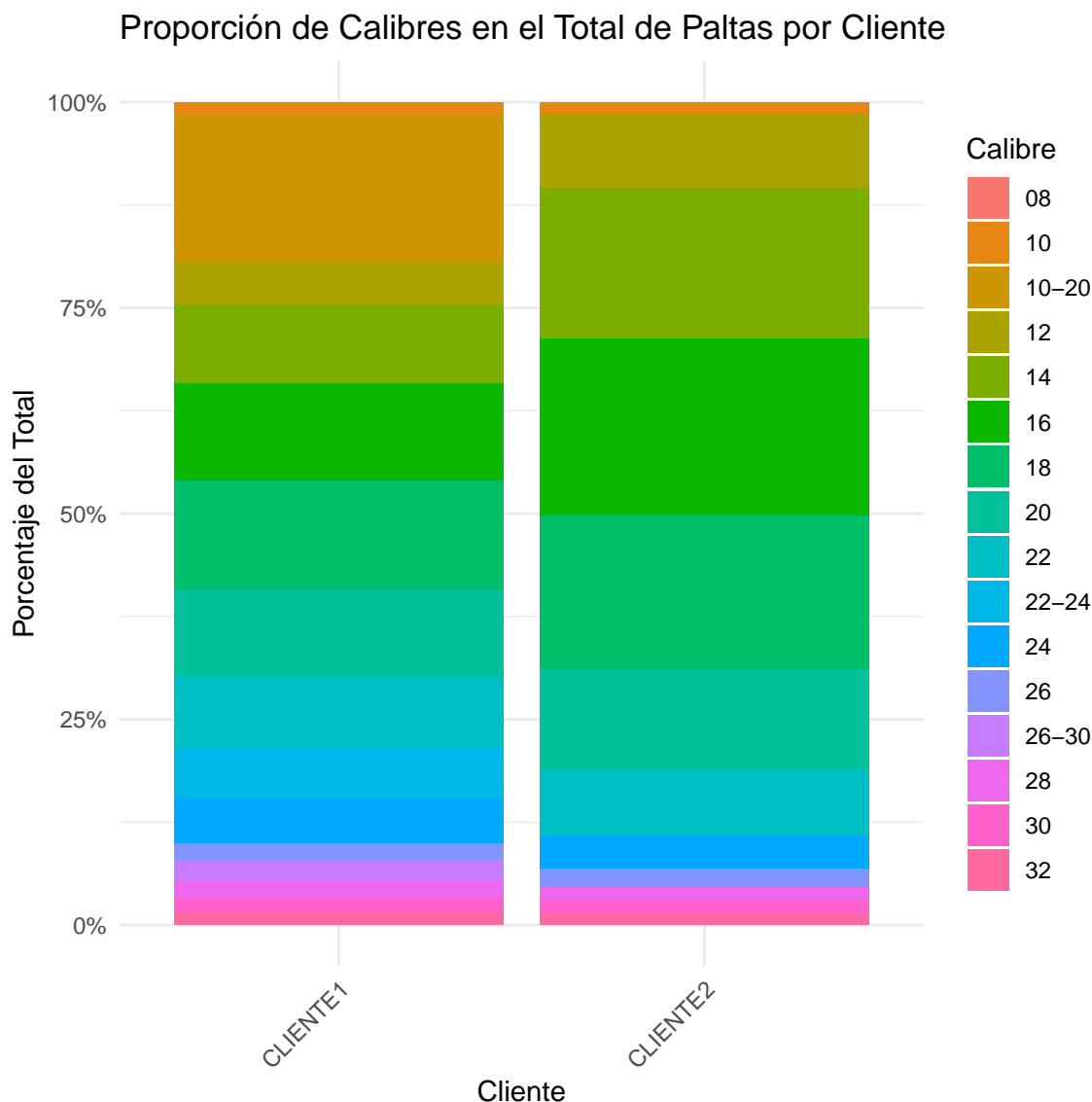


**Interpretación:** Del gráfico se desprende que el CLIENTE2 tiene una mayor concentración de paltas en el calibre 22-24, destacando una preferencia o demanda mayor por paltas de tamaño específico. Por otro lado, el CLIENTE1 presenta una distribución más uniforme en varios rangos de calibre, aunque con volúmenes menores en comparación. Esto podría reflejar una estrategia de diversificación en la compra de paltas de diferentes tamaños por parte de CLIENTE1, mientras que CLIENTE2 podría estar centrado en un segmento de mercado específico que prefiere calibres más grandes. La prominencia del calibre 22-24 para CLIENTE2 sugiere una posible tendencia de mercado o una respuesta a una campaña promocional o requisito contractual en ese período.

**4.3.2.2. Proporción de cada calibre en el total de paltas por cliente** Se genera el gráfico de barras apiladas porcentuales para visualizar la composición de los calibres de paltas Hass en términos porcentuales para cada cliente, facilitando la comparación directa entre los dos.

*# Generamos el gráfico de barras apiladas porcentuales*

```
ggplot(calibres, aes(x = cliente, y = percentage, fill = as.factor(calibre))) +  
  geom_bar(stat = "identity") +  
  scale_y_continuous(labels = scales::percent_format()) +  
  labs(x = "Cliente", y = "Porcentaje del Total", fill = "Calibre",  
       title = "Proporción de Calibres en el Total de Paltas por Cliente") +  
  theme_minimal() +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



**Interpretación:** En el gráfico se aprecia que CLIENTE1 tiene una distribución más equilibrada de calibres, con una mayor uniformidad en las proporciones de cada segmento. Por otro lado, CLIENTE2 muestra una predominancia de los calibres 22 y 24, lo que indica una preferencia o demanda significativa por paltas de mayor tamaño. Los calibres más pequeños (08, 10) tienen una presencia mínima en ambos clientes, mientras que el rango medio (14-20) constituye una proporción sustancial para CLIENTE1. Este patrón refleja las preferencias de compra o las especificaciones contractuales de cada cliente, y puede influir en las decisiones de cultivo y venta por parte del productor.

## 4.4. Matriz de correlación

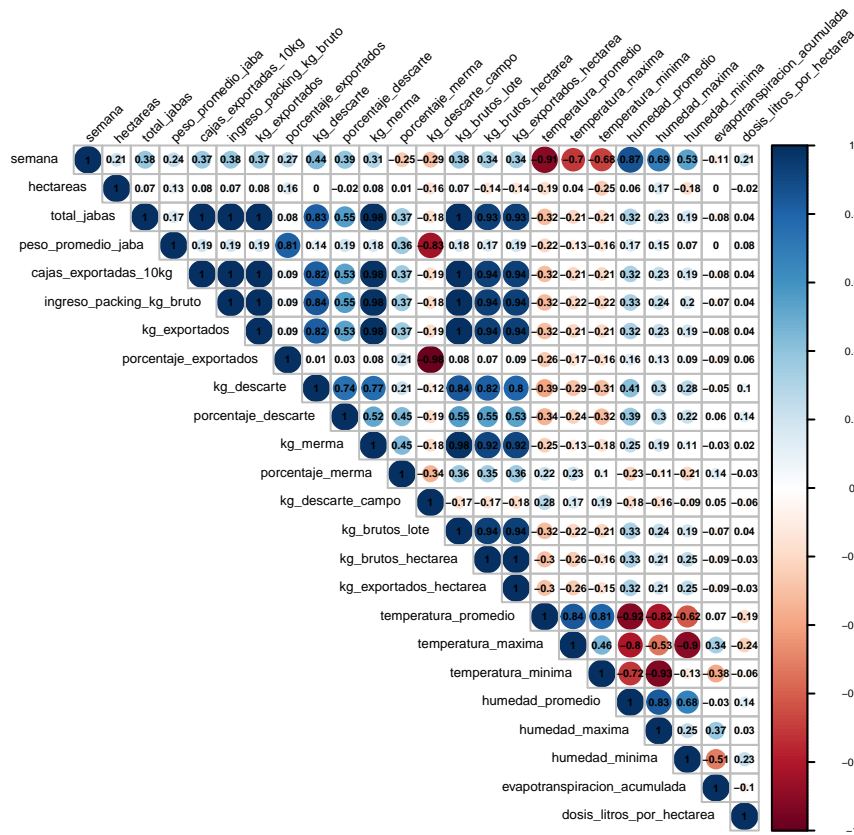
La matriz de correlación para el dataset 'reporte' muestra múltiples variables y la fuerza de sus relaciones lineales, con colores y tamaños que representan el grado de correlación. Los colores más oscuros y círculos más grandes indican una correlación fuerte, mientras que los colores más claros y círculos más pequeños indican una correlación débil.

```
# Identificar columnas numéricas
numeric_columns <- sapply(reporte, is.numeric)

# Filtrar el dataframe con columnas numéricas
reporte_numeric <- reporte[, numeric_columns]

# Calcular la matriz de correlación
correlation_matrix <- cor(reporte_numeric, use = "pairwise.complete.obs")

# Generar diagrama de correlación
corrplot(correlation_matrix, method = "circle", type = "upper",
  tl.col = "black", # color del texto de las etiquetas
  tl.srt = 45,      # rotación del texto de las etiquetas en grados
  addCoef.col = "black", # color de los coeficientes de correlación
  number.cex = 0.35, # tamaño de los coeficientes de correlación
  cl.cex = 0.3,     # tamaño del texto de la leyenda de colores
  tl.cex = 0.4,     # tamaño del texto de las etiquetas
  cl.ratio = 0.2    # relación de tamaño de la leyenda de colores
)
```



**Interpretación:** Existe una correlación fuerte y positiva entre el peso bruto de los lotes (`kg_brutos_lote`) y el peso de descarte (`kg_descarte`), así como con el porcentaje de descarte, lo cual sugiere que lotes con mayor peso bruto tienden a tener mayores cantidades de descarte. Las variables de temperatura (`temperatura_promedio`, `temperatura_maxima`, `temperatura_minima`) muestran correlaciones fuertes entre sí, lo que es esperable ya que están intrínsecamente relacionadas. La humedad (`humedad_promedio`, `humedad_maxima`, `humedad_minima`) también muestra una correlación positiva fuerte entre sus distintas medidas. Curiosamente, la evapotranspiración acumulada parece tener una correlación negativa con varias variables relacionadas con la producción y las condiciones climáticas, lo que podría indicar un efecto del clima en la eficiencia de la producción. Las correlaciones observadas entre la producción y las variables climáticas pueden ser de particular interés, ya que indican la posible influencia del clima en la producción de paltas.

## 5. Visualización Avanzada de Datos

### 5.1. Visualización avanzada del dataset ‘calibres’

#### 5.1.1. Analizando la Producción de KgCal por Cliente con Boxplots

Se genera esta visualización para examinar las diferencias en la distribución de los pesos por calibre de paltas entre dos clientes.



```

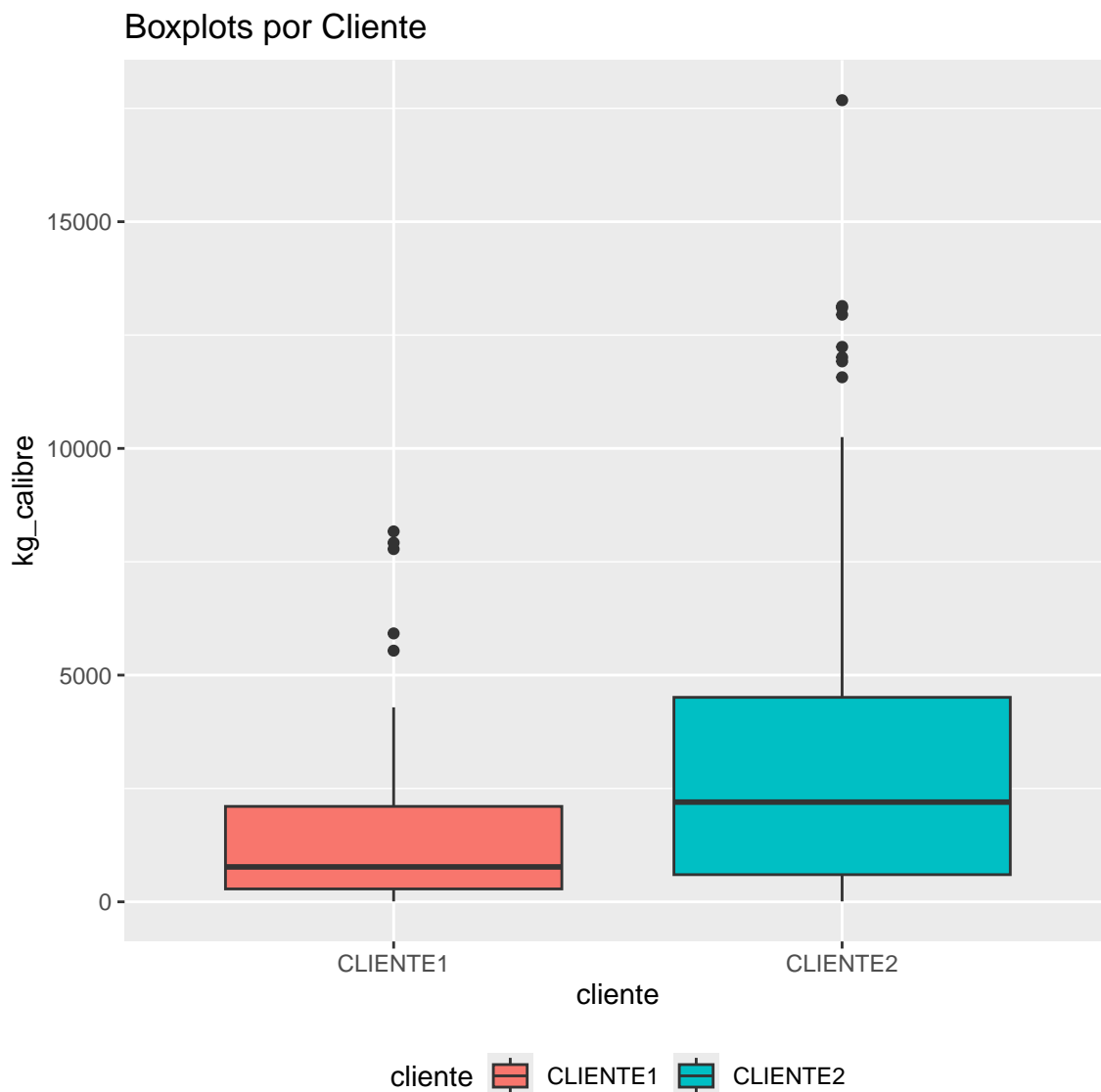
# Configuración del tamaño del gráfico
options(repr.plot.width=18, repr.plot.height=6)

# Creación del boxplot
boxplot <- ggplot(calibres, aes(x = cliente, y = kg_calibre, 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)

```



**Interpretación:** La comparación revela que CLIENTE2 tiene una mediana y una variabilidad mayor en los pesos que CLIENTE1, con un rango intercuartílico más amplio y una mayor cantidad de valores atípicos, lo que indica posibles calibres excepcionalmente pesados o inconsistencias en los datos. CLIENTE1 muestra una distribución más compacta, sugiriendo una mayor uniformidad en el tamaño de los calibres entregados. Esta información es crucial para la gestión de la calidad del producto y la satisfacción del cliente, y podría influir en la toma de decisiones estratégicas sobre la producción y distribución.

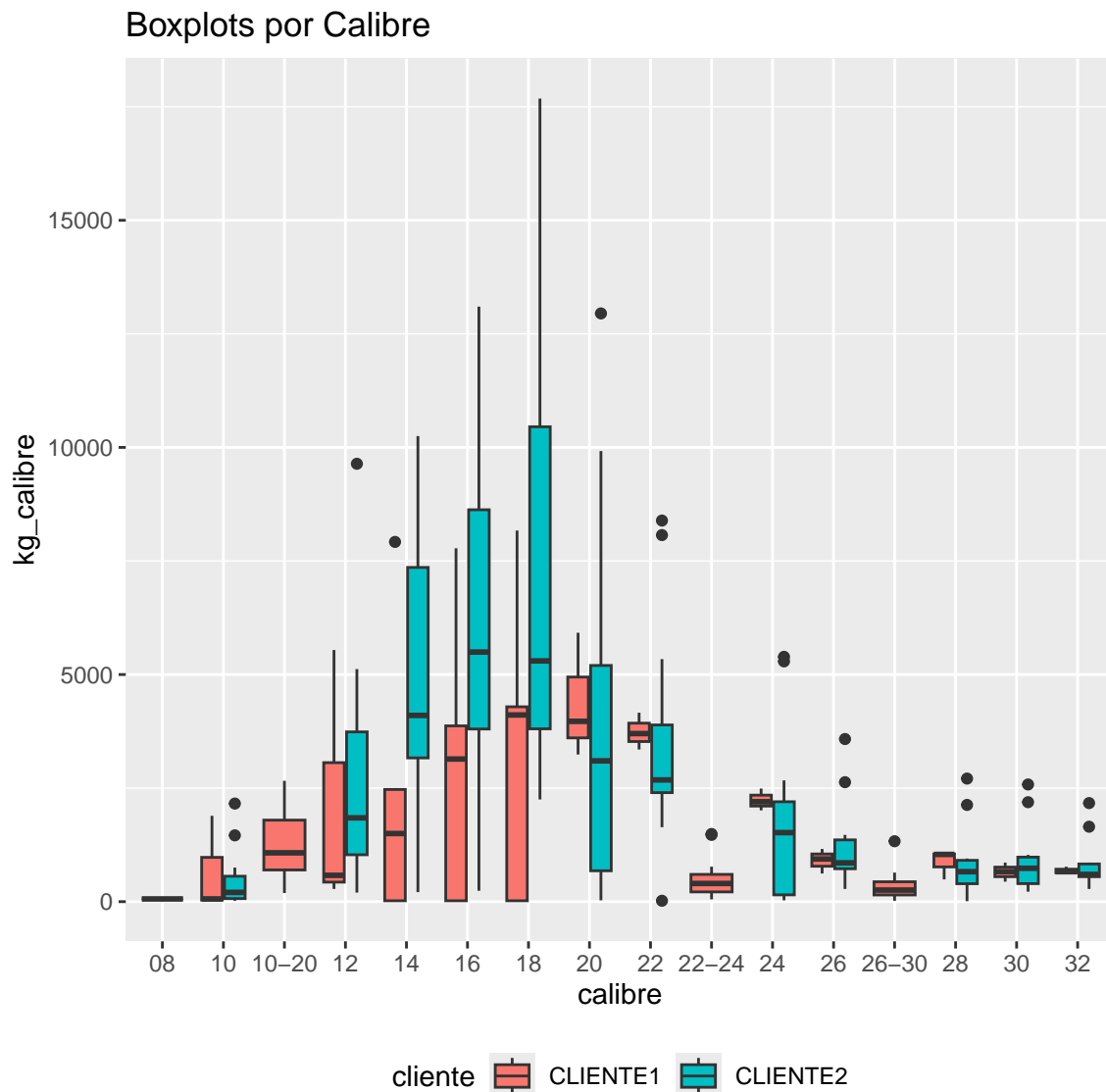
### 5.1.2. Analizando la Distribución de KgCal por Calibre con Boxplots

Se genera esta visualización para evaluar la variación en los pesos por calibre entre paltas entregadas a dos clientes distintos.

```
# Creando el boxplot
boxplot <- ggplot(calibres, aes(x = calibre, y = kg_calibre, 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)
```



**Interpretación:** Los boxplots demuestran que hay una variación considerable en los pesos por calibre para ambos clientes, aunque CLIENTE2 tiende a tener un rango más amplio y una mediana más alta en varios calibres, lo cual sugiere un tamaño de calibre generalmente más grande y más variado. CLIENTE1 muestra menos variabilidad, especialmente en calibres menores, lo que podría indicar una selección más uniforme de paltas. La presencia de valores atípicos en ambos clientes, especialmente en calibres más altos, destaca la variabilidad en el tamaño y el peso

que podría impactar en la planificación de la producción y las expectativas del cliente.

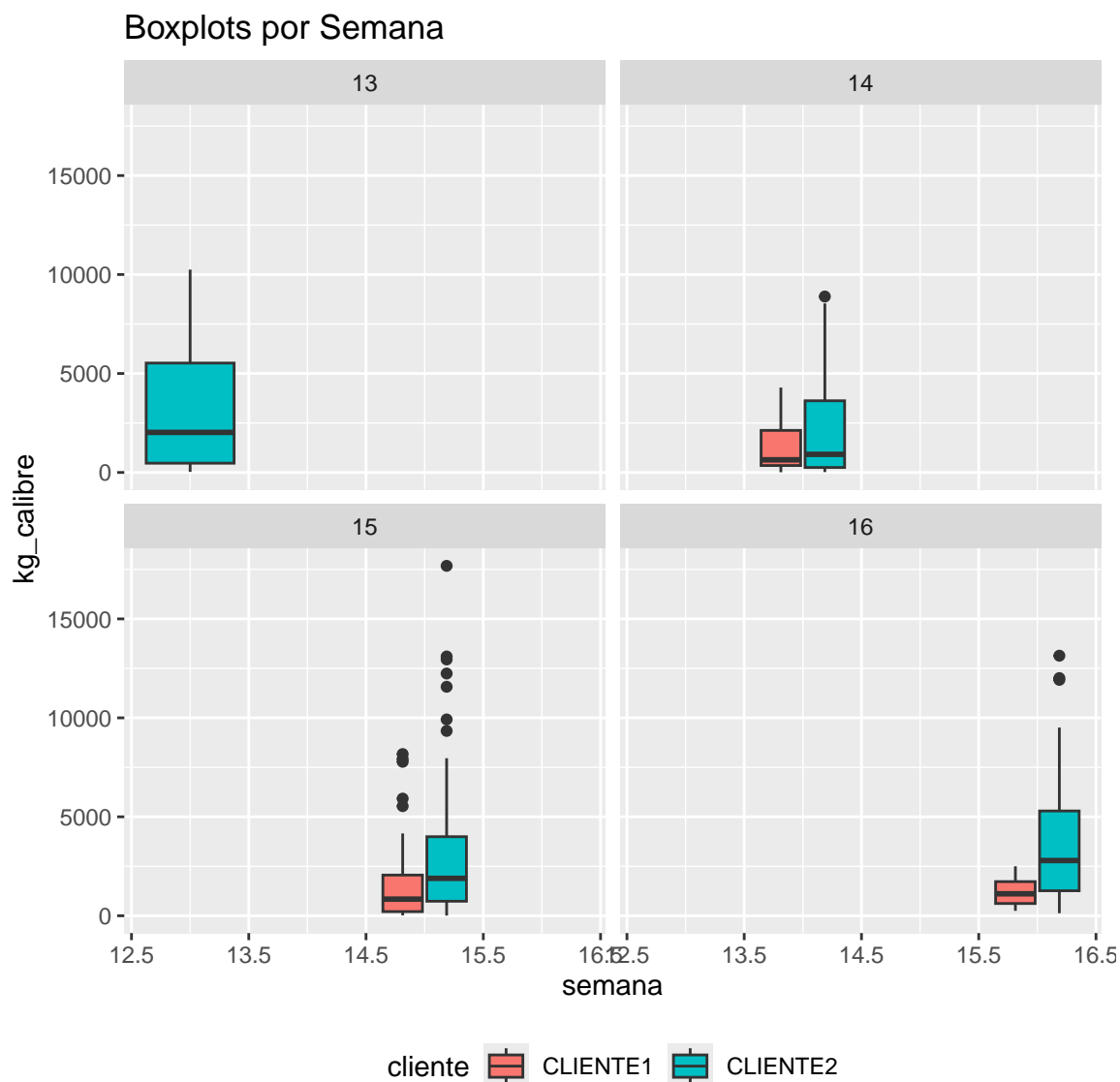
### 5.1.3. Analizando la Distribución de KgCal por Semana con Boxplots Faceteados

Se genera esta visualización para analizar la distribución semanal de los pesos por calibre de las paltas entre dos clientes distintos.

```
#1 Crear los boxplots
boxplot <- ggplot(calibres, aes(x = semana, y = kg_calibre, fill = cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Semana") +
  facet_wrap(~semana)

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

# Mostrar los boxplots
print(boxplot)
```



**Interpretación:** La serie de boxplots faceteada por semana muestra que la distribución de los pesos de los calibres de palta entregados varía notablemente semana a semana para ambos clientes. Se observa que CLIENTE1 generalmente

tiene una mediana más baja en comparación con CLIENTE2, especialmente en las semanas 13 y 14. Además, se identifica una variabilidad significativa en las semanas 15 y 16 para ambos clientes, donde CLIENTE2 muestra una mayor cantidad de valores atípicos. Esto sugiere que factores semanales podrían estar influenciando la producción, como las condiciones climáticas o las prácticas de cosecha que varían durante el mes. Estos datos son fundamentales para comprender los ciclos de producción y optimizar las operaciones según las tendencias semanales.

#### 5.1.4. Analizando la Distribución de KgCal por Calibre con Boxplots Faceteados

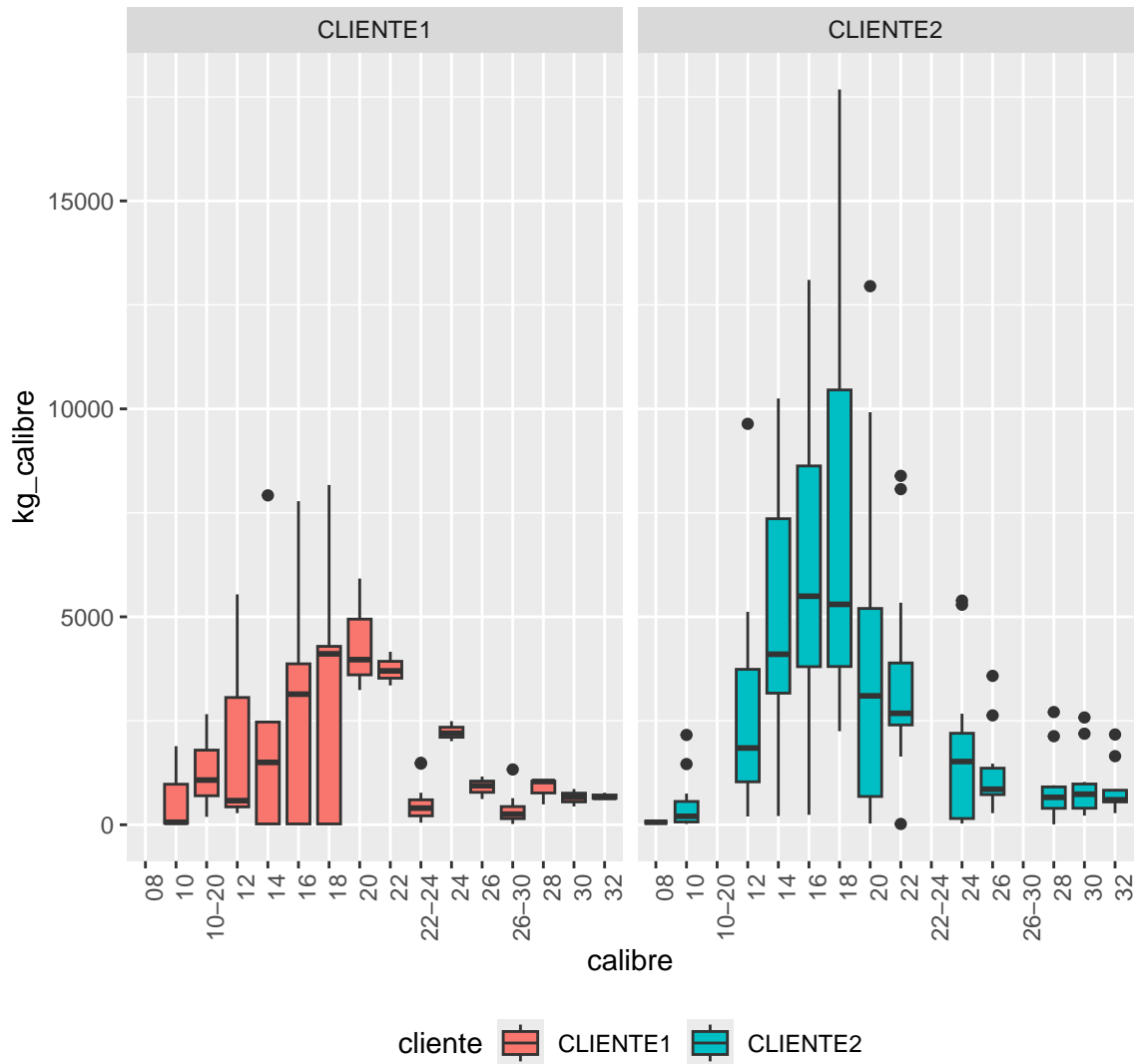
Se genera esta visualización para comparar la distribución del peso por calibre de las paltas entre dos clientes.

```
#2 Crear los boxplots
options(repr.plot.width=15, repr.plot.height=5)
boxplot <- ggplot(calibres, aes(x = calibre, y = kg_calibre, fill = cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Calibre") +
  facet_wrap(~cliente)+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

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

# Mostrar los boxplots
print(boxplot)
```

### Boxplots por Calibre



**Interpretación:** En los boxplots faceteados por cliente, se observa que CLIENTE2 tiene una mayor dispersión en la mayoría de los calibres en comparación con CLIENTE1, lo que sugiere una variabilidad significativa en el tamaño de las paltas recibidas. Ambos clientes muestran valores atípicos en varios calibres, pero es más pronunciado en CLIENTE2, especialmente en los calibres grandes. Esto podría indicar que CLIENTE2 recibe una gama más amplia de tamaños de paltas, mientras que CLIENTE1 recibe paltas más uniformes. Estos detalles pueden ser importantes para adaptar las estrategias de producción y distribución a las necesidades y expectativas específicas de cada cliente.

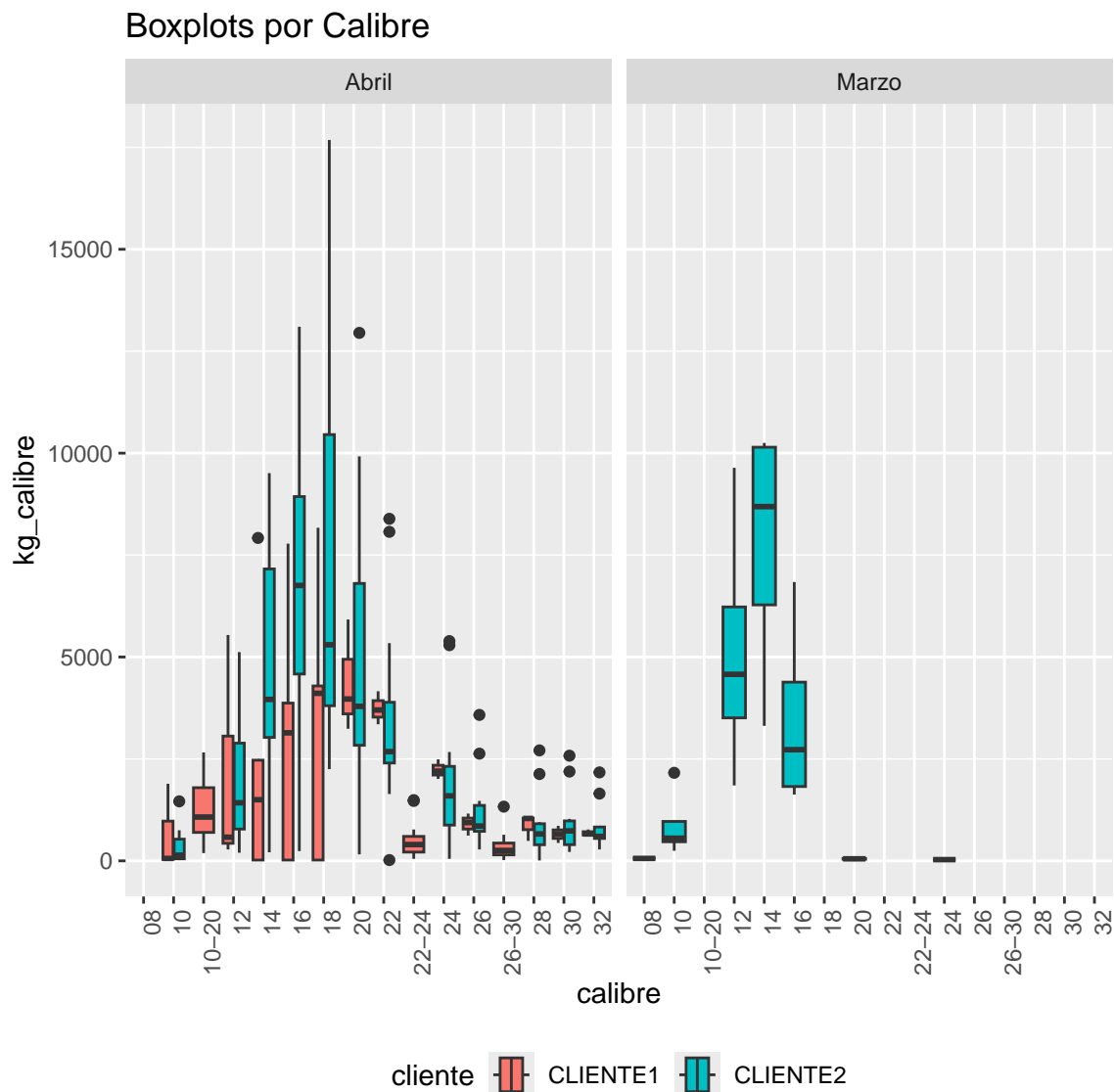
#### 5.1.5. Explorando la Distribución de Calibre por KgCal con Boxplots Faceteados y Gráfico Rotado

Se genera esta visualización para evaluar la distribución del peso por calibre de las paltas entregadas a dos clientes durante diferentes meses.

```
#3 Crear los boxplots
options(repr.plot.width=15, repr.plot.height=5)
boxplot <- ggplot(calibres, aes(y = calibre, x = kg_calibre, fill = cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Calibre") +
  facet_wrap(~mes) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip()
```

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

# Mostrar los boxplots
print(boxplot)
```



**Interpretación:** La visualización muestra boxplots de los pesos por calibre para los meses de abril y marzo, separados por cliente. CLIENTE1 muestra una variabilidad considerable en abril, con muchos valores atípicos, especialmente en calibres más altos, lo que podría indicar una producción más variable o una clasificación menos consistente en este mes. Para marzo, CLIENTE2 presenta una variabilidad menor y pesos más consistentes en los calibres, lo que sugiere una mayor uniformidad en la producción o en la selección para este cliente durante este mes. Estas tendencias podrían reflejar diferencias estacionales en la producción o en las prácticas de cosecha y clasificación.

## 5.2. Visualización avanzada del dataset 'reporte'

### 5.2.1. Boxplots de Producción por Tipo de Insumo y Cliente a lo Largo de Cuatro Semanas

Se genera esta gráfica para visualizar y comparar la distribución del peso total de lotes de paltas Hass (kg\_brutos\_lote) por tipo de insumo utilizado (mineral, orgánico, químico) y cliente a lo largo de cuatro semanas específicas.

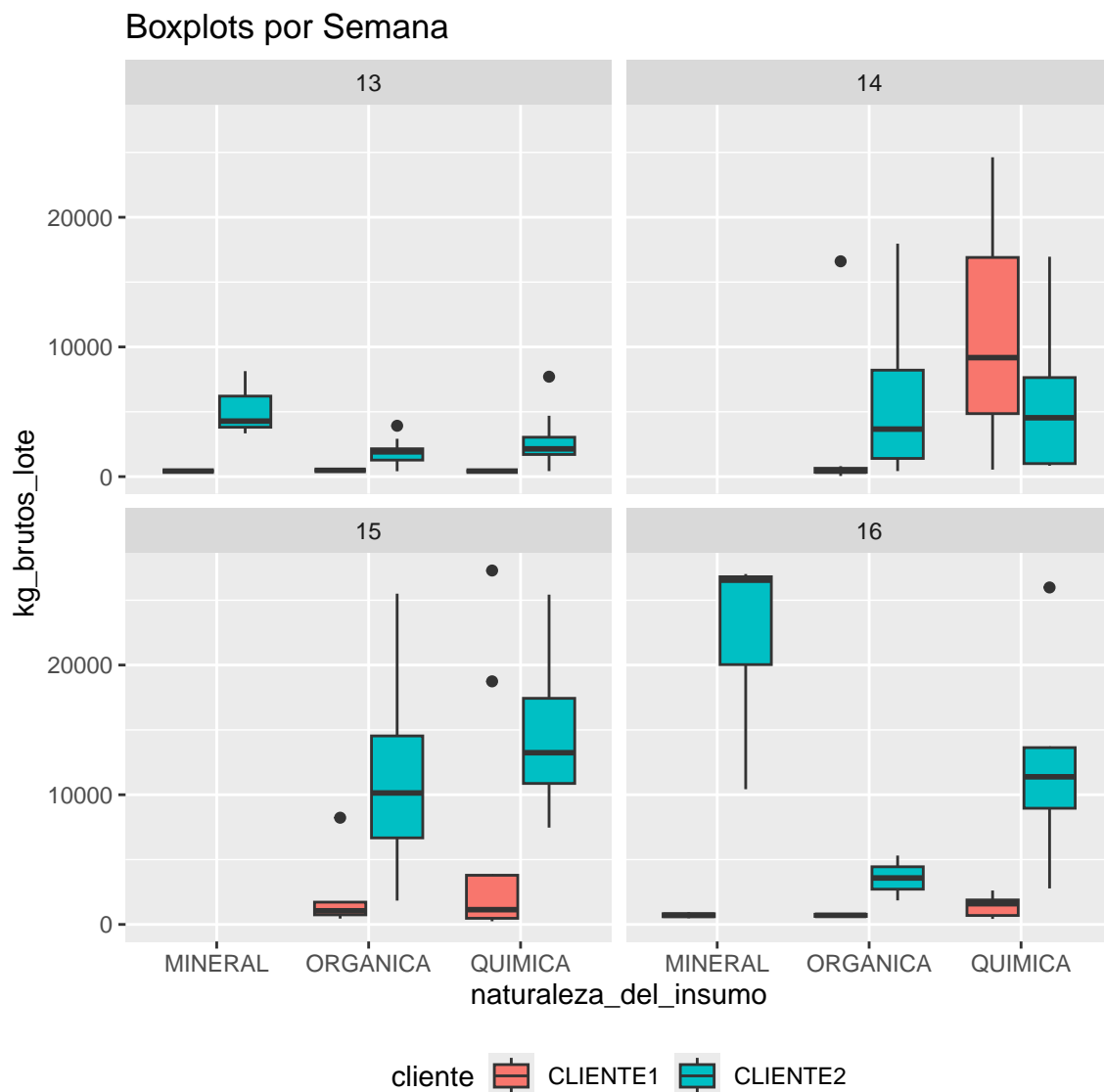
```

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

# Crear los boxplots
boxplot <- ggplot(reporte, aes(x = naturaleza_del_insumo, y = kg_brutos_lote, fill = cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Semana") +
  facet_wrap(~semana)
# Establecer la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")

# Mostrar los boxplots
print(boxplot)

```



**Interpretación:** La gráfica muestra que hay una variabilidad considerable en la producción entre los diferentes tipos de insumos y entre semanas para ambos clientes. Para el CLIENTE1, la semana 14 presenta una mayor mediana y variabilidad en la producción, independientemente del tipo de insumo. En contraste, el CLIENTE2 muestra menos variabilidad entre los tipos de insumos pero destaca la semana 16, donde los insumos orgánicos y químicos tienen medias más altas y mayor dispersión. Los valores atípicos presentes sugieren que hay casos de producción excepcionalmente alta que podrían estar vinculados a la naturaleza del insumo utilizado o a otros factores

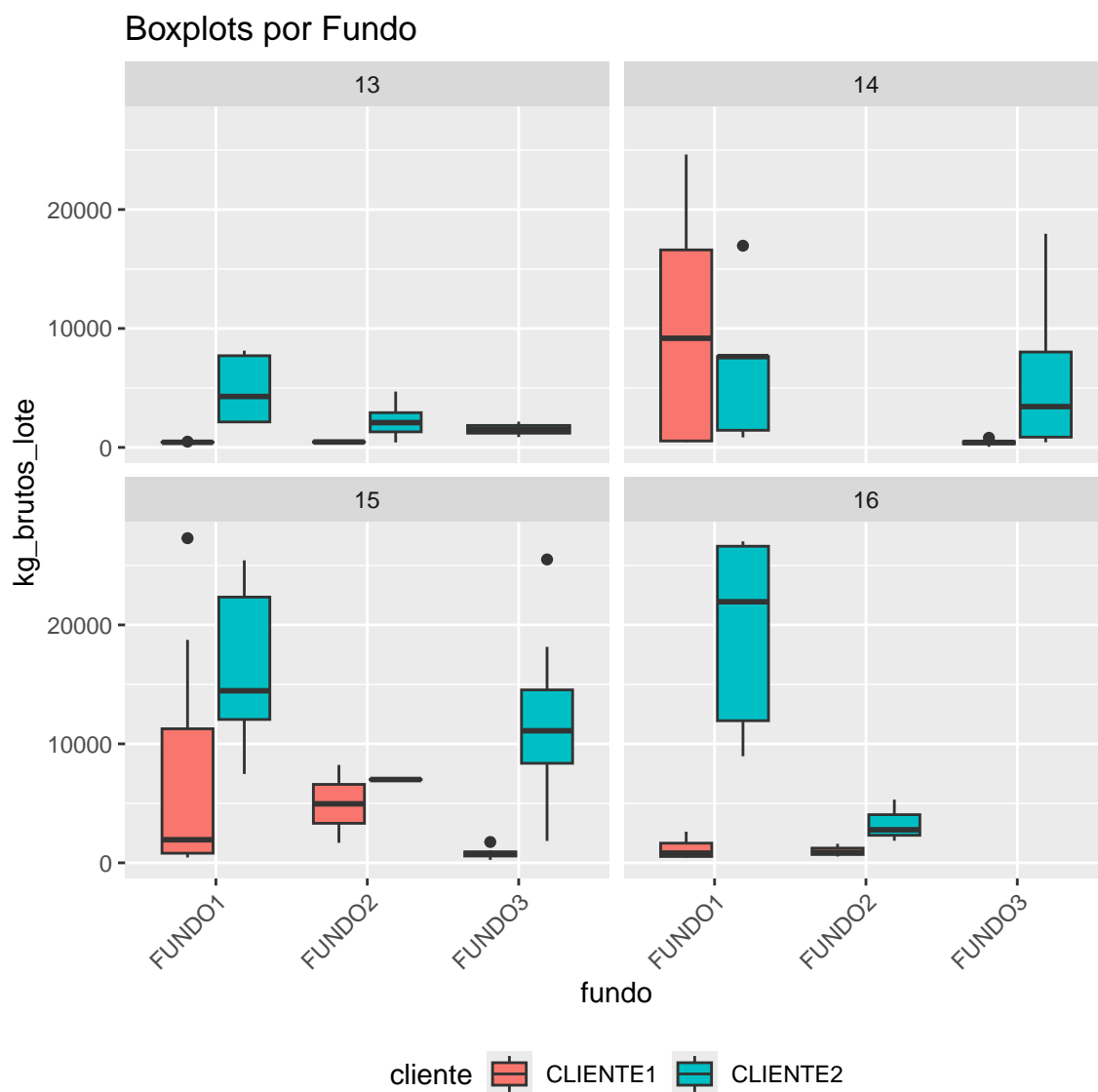
no capturados en el gráfico. Esta información puede ser fundamental para tomar decisiones sobre la planificación de la producción y la selección de insumos para las próximas temporadas.

### 5.2.2. Boxplots de Producción por Fundo y Cliente Durante Cuatro Semanas

Se genera esta gráfica para comparar la producción total de paltas Hass (kg\_brutos\_lote) entre diferentes fundos y para dos clientes a lo largo de un periodo de cuatro semanas.

```
# Crear los boxplots
options(repr.plot.width=15, repr.plot.height=5)
boxplot <- ggplot(reporte, aes(x = fundo, y = kg_brutos_lote, fill = cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Fundo") +
  facet_wrap(~semana)+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
boxplot <- boxplot + theme(legend.position = "bottom")

# Mostrar los boxplots
print(boxplot)
```



**Interpretación:** Los boxplots muestran que la producción varía según el fundo y la semana. En la semana 13, los



fundos para CLIENTE1 muestran una producción más consistente y baja en comparación con CLIENTE2, que presenta una variabilidad significativa entre los fundos. En la semana 14, se observa una mayor variabilidad para ambos clientes, con una producción particularmente alta en uno de los fundos para CLIENTE1. Durante la semana 15, ambos clientes experimentan una disminución en la producción en todos los fundos, con una variabilidad reducida para CLIENTE2. La semana 16 destaca por una producción significativamente mayor en uno de los fundos para CLIENTE2. Estas diferencias sugieren que la producción está influenciada por la combinación específica de cliente, fundo y las condiciones semanales que prevalecen durante el periodo analizado.

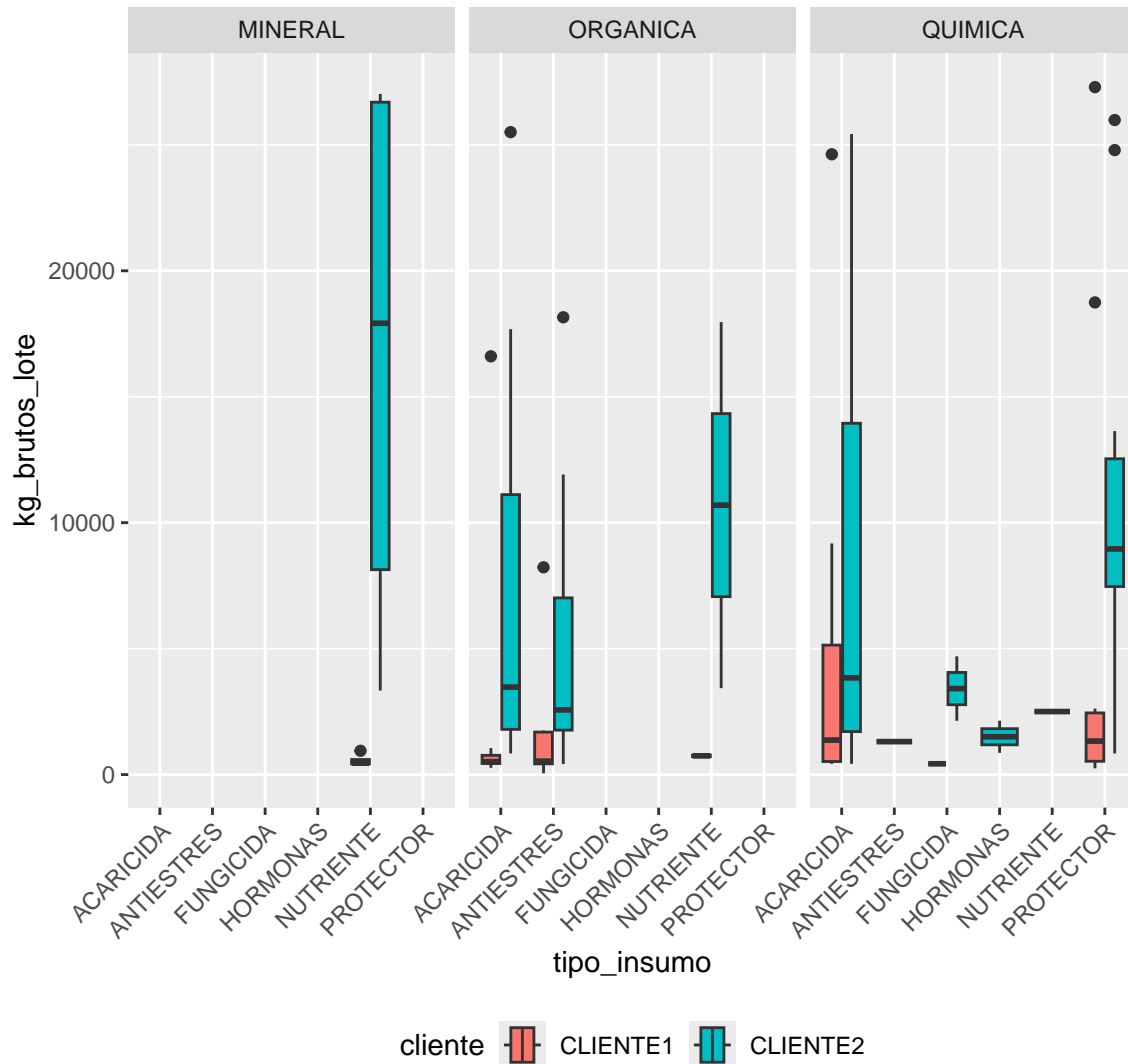
### 5.2.3. Boxplots de Producción por Tipo de Tratamiento y Cliente Según la Naturaleza del Insumo

Se genera esta gráfica para explorar cómo varía la producción total de paltas Hass (kg\_brutos\_lote) en función del tipo de tratamiento aplicado y el cliente, agrupados según la naturaleza del insumo (mineral, orgánica, química).

```
# Crear los boxplots
options(repr.plot.width=15, repr.plot.height=5)
boxplot <- ggplot(reporte, aes(y = tipo_insumo, x = kg_brutos_lote, fill = cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Tipo Dosis") +
  facet_wrap(~naturaleza_del_insumo) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_flip()
boxplot <- boxplot + theme(legend.position = "bottom")

# Mostrar los boxplots
print(boxplot)
```

## Boxplots por Tipo Dosis



**Interpretación:** El gráfico muestra diferencias significativas en la producción en función del tratamiento y el cliente. Para los insumos minerales y orgánicos, CLIENTE1 muestra una variabilidad más amplia en la producción en comparación con CLIENTE2, que tiene una producción más constante y baja, particularmente con insumos orgánicos. Bajo los tratamientos químicos, CLIENTE2 tiene una mayor mediana de producción en comparación con CLIENTE1, lo cual puede sugerir una mejor respuesta o preferencia por este tipo de insumo. La presencia de valores atípicos, especialmente en tratamientos minerales y orgánicos, indica que ciertas combinaciones de tratamientos y clientes pueden resultar en lotes de producción excepcionalmente altos o bajos. Estos resultados pueden proporcionar información valiosa para optimizar las prácticas agrícolas y la selección de insumos para mejorar la producción.

### 5.2.4. Gráfico de Dispersión de la Producción por Hectárea y Cliente Según Tipo de Insumo

Se genera esta gráfica para visualizar la relación entre el área cultivada (hectáreas) y la producción bruta por hectárea (kg\_brutos\_hectarea), diferenciando los resultados por cliente y tipo de insumo agrícola.

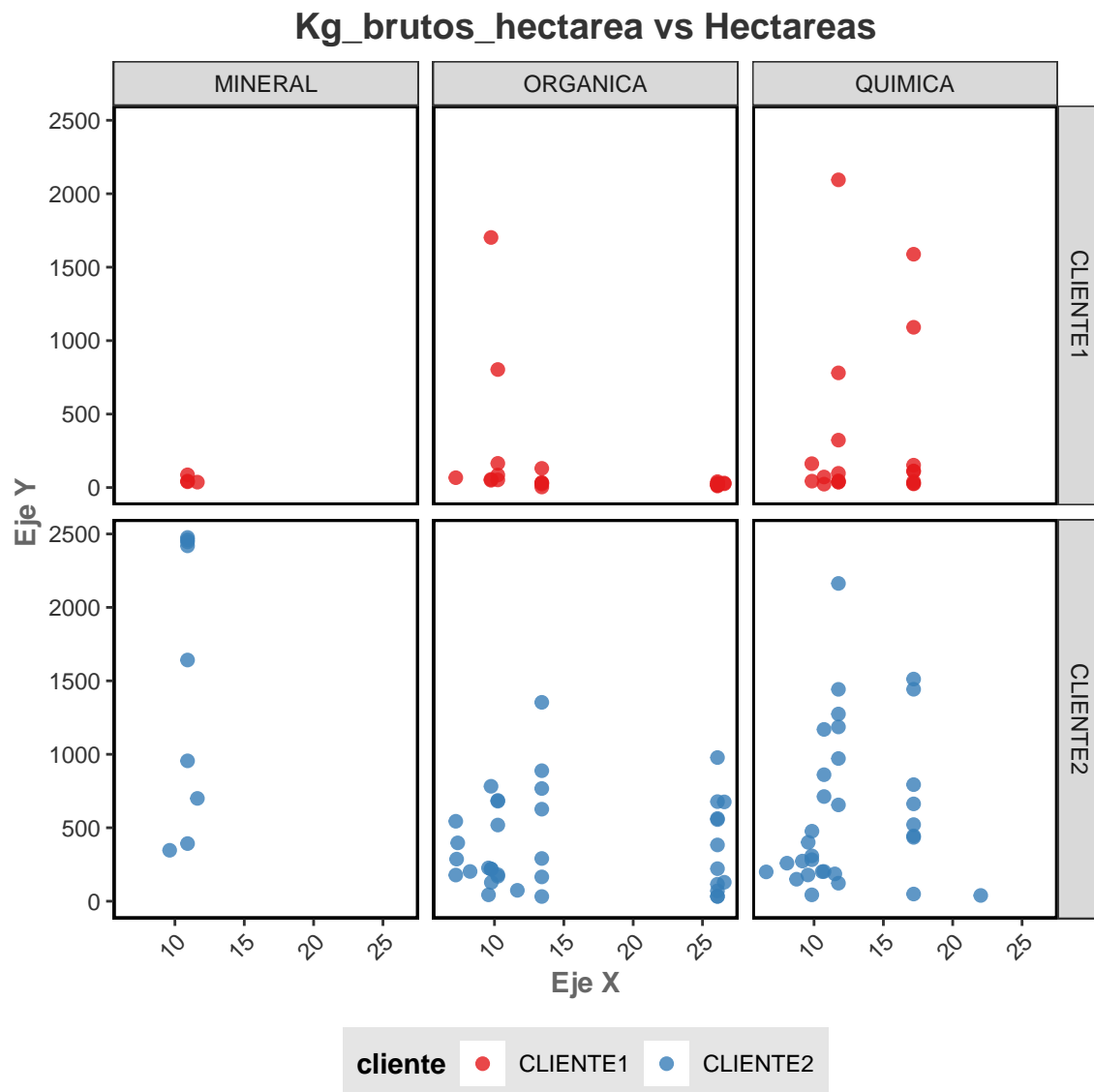
```
# Definir una paleta de colores intensos
colors <- brewer.pal(n = 8, name = "Set1")
```

```
# Gráfico ggplot con diseño mejorado y colores
ggplot(reporte, aes(x = hectareas, y = kg_brutos_hectarea, color = cliente)) +
  geom_point(size = 2, alpha = 0.8) +
```

```

scale_color_manual(values = colors) +
facet_grid(cliente ~ naturaleza_del_insumo) +
labs(title = "Kg_brutos_hectarea vs Hectareas",
      x = "Eje X",
      y = "Eje Y") +
theme_bw() +
theme(
  plot.title = element_text(face = "bold", color = "#333333", size = 14, hjust = 0.5),
  axis.title = element_text(face = "bold", color = "#666666"),
  axis.text.x = element_text(angle = 45, hjust = 1, color = "#333333"),
  axis.text.y = element_text(color = "#333333"),
  legend.title = element_text(face = "bold"),
  legend.background = element_rect(fill = "gray90"),
  legend.position = "bottom",
  panel.grid.major = element_blank(),
  panel.grid.minor = element_blank(),
  panel.border = element_rect(color = "black", fill = NA, linewidth = 1)
)

```



**Interpretación:** Los puntos representan la producción individual por hectárea para CLIENTE1 (en rojo) y

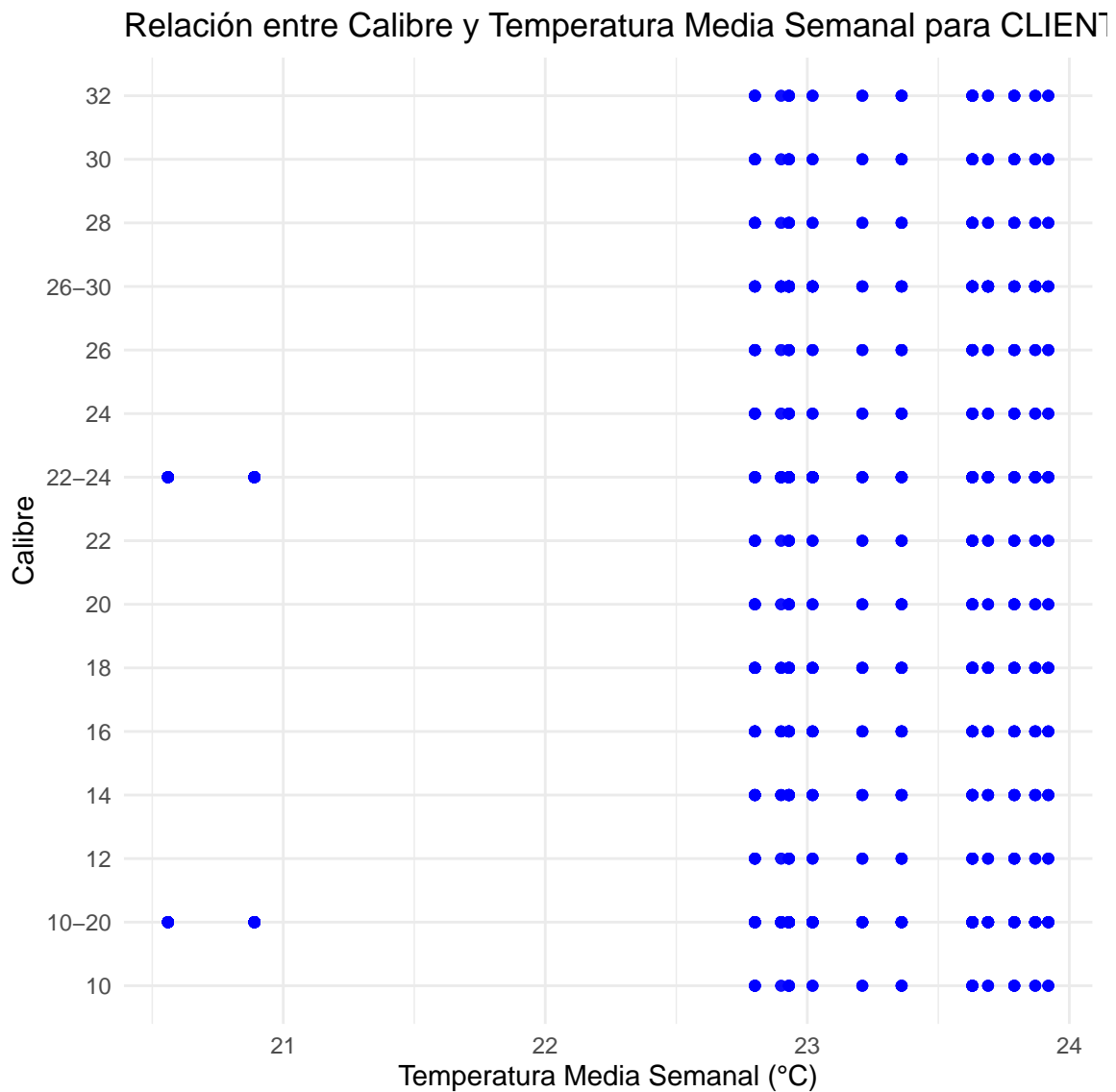
CLIENTE2 (en azul) a través de tres tipos de insumos: mineral, orgánico y químico. Se observa que CLIENTE2 tiene una dispersión de puntos que sugiere una producción más variable en función del tamaño del área cultivada, especialmente con insumos químicos y orgánicos. CLIENTE1 muestra menos puntos y, por lo tanto, menos variabilidad, lo que podría indicar una producción más homogénea o menos datos disponibles. En general, los datos muestran una tendencia de que el aumento en la cantidad de hectáreas no se traduce necesariamente en un incremento proporcional en la producción, lo que podría implicar que hay otros factores que afectan la eficiencia de la producción por hectárea.

## 5.3. Gráficos de dispersión

### 5.3.1. Diversidad de Calibres a Varias Temperaturas para clientes

**5.3.1.1. Diversidad de Calibres a Varias Temperaturas para CLIENTE1** El primer gráfico que se genera muestra la relación entre el calibre de las paltas Hass y la temperatura media semanal para CLIENTE1. Los puntos azules representan las observaciones individuales donde el eje horizontal muestra la temperatura y el vertical el calibre.

```
# Filtrar para CLIENTE1
ggplot(combined_data %>% filter(cliente.x == "CLIENTE1"), aes(x = temperatura_promedio, y = calibre)) +
  geom_point(alpha = 0.5, color = "blue") +
  labs(title = "Relación entre Calibre y Temperatura Media Semanal para CLIENTE1",
        x = "Temperatura Media Semanal (°C)",
        y = "Calibre") +
  theme_minimal()
```

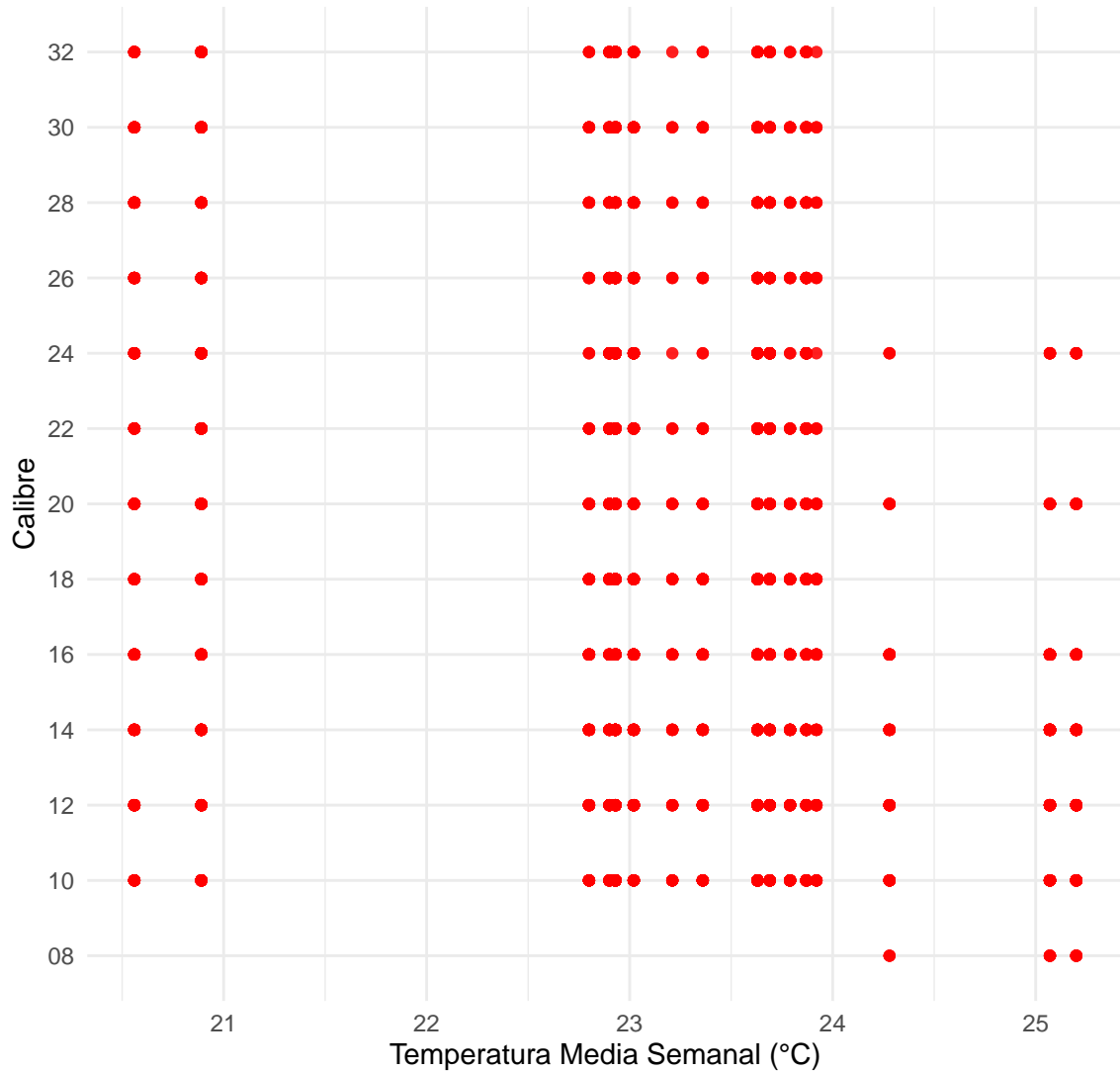


**Interpretación:** Los datos parecen distribuirse uniformemente a lo largo de las temperaturas, sin mostrar un patrón claro que indique una relación fuerte entre la temperatura y el calibre de las paltas. Hay una amplia variedad de calibres presentes a lo largo del rango de temperatura, lo que sugiere que otros factores, además de la temperatura, pueden influir en el calibre de las paltas para CLIENTE1.

**5.3.1.2. Diversidad de Calibres a Varias Temperaturas para CLIENTE2** El segundo gráfico que se genera ilustra la relación entre el calibre de las paltas Hass y la temperatura media semanal para CLIENTE2. Los puntos rojos representan las observaciones individuales con la temperatura en el eje horizontal y el calibre en el vertical.

```
# Filtrar para CLIENTE2
ggplot(combined_data %>% filter(cliente.x == "CLIENTE2"), aes(x = temperatura_promedio, y = calibre)) +
  geom_point(alpha = 0.5, color = "red") +
  labs(title = "Relación entre Calibre y Temperatura Media Semanal para CLIENTE2",
        x = "Temperatura Media Semanal (°C)",
        y = "Calibre") +
  theme_minimal()
```

### Relación entre Calibre y Temperatura Media Semanal para CLIENTE1



**Interpretación:** Al igual que con CLIENTE1, la distribución de los calibres a través de las temperaturas para CLIENTE2 es bastante uniforme y no muestra una correlación evidente entre estas variables. Los calibres varían significativamente a temperaturas similares, lo que indica que la temperatura por sí sola no es un predictor del calibre de las paltas para CLIENTE2.

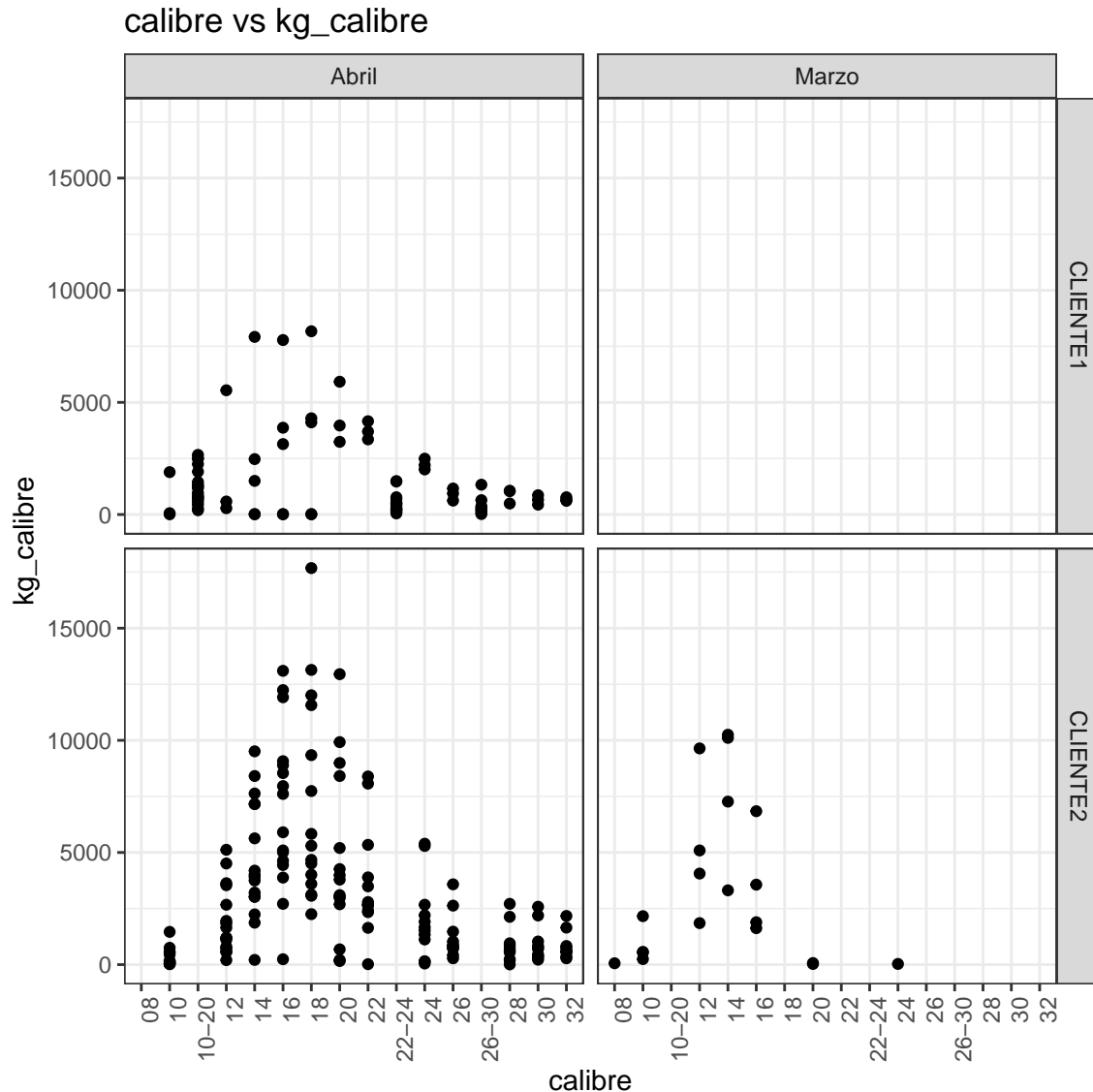
#### 5.3.2. Análisis Visual Avanzado de 'calibre' vs 'kg\_calibre' con Facetas y Líneas

Esta sección complementa el análisis de la distribución de calibre y kg\_calibre con dos nuevas visualizaciones: un gráfico de dispersión faceteado y un gráfico de líneas. Estas visualizaciones adicionales permiten explorar las relaciones entre calibre y kg\_calibre con mayor detalle, considerando factores como el cliente y el mes.

**5.3.2.1. 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 kg\_calibre varía entre diferentes clientes y meses.

```
# Faceteado
ggplot(calibres, aes(x = calibre, y = kg_calibre)) +
  geom_point() +
  facet_grid(cliente ~ mes) +
  labs(title = "calibre vs kg_calibre",
```

```
x = "calibre",
y = "kg_calibre") +
theme_bw() +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

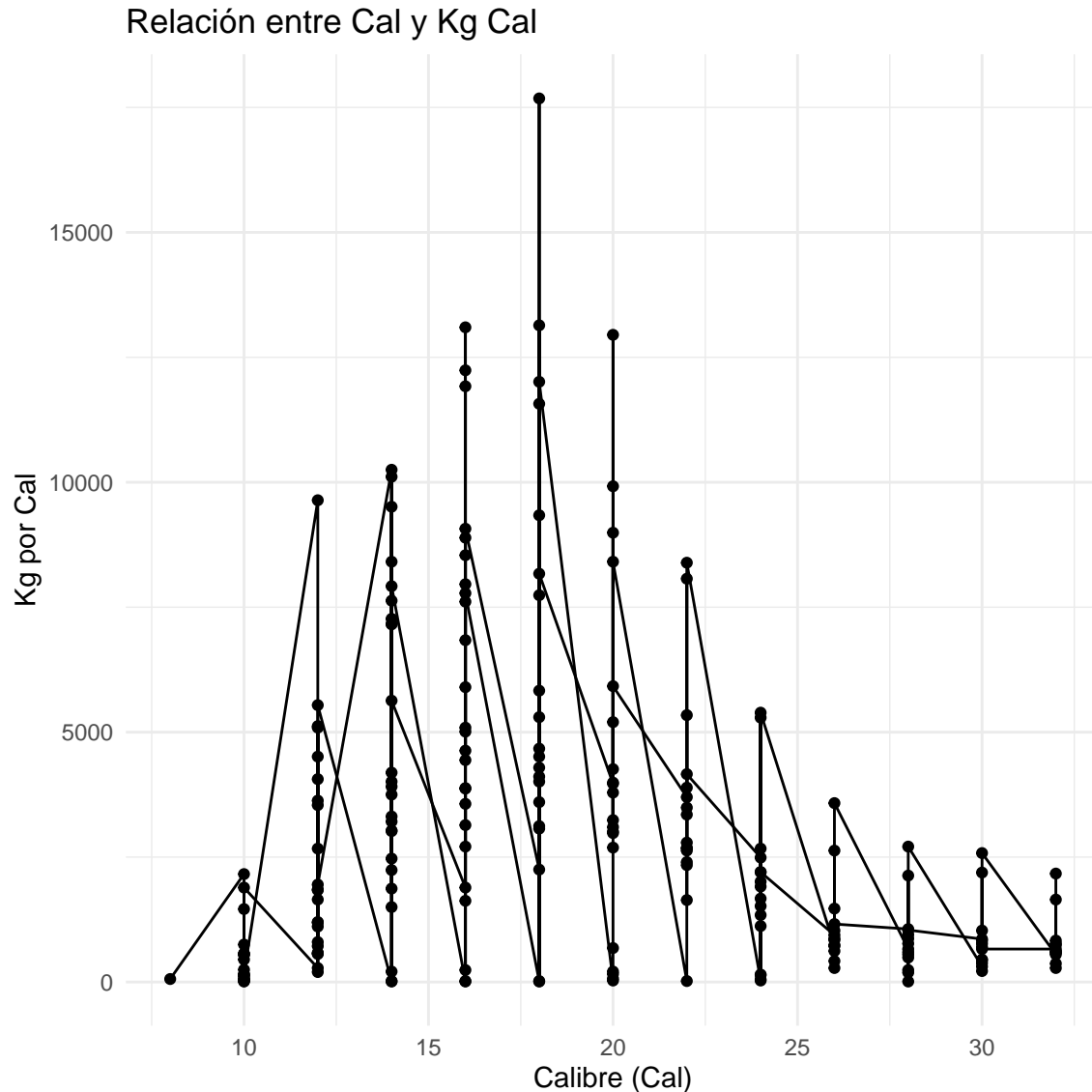


**Interpretación:** El gráfico de dispersión faceteado presenta los pesos (Eje Y) por calibre (Eje X) de las paltas Hass para dos clientes en dos meses distintos. En el caso de CLIENTE1 (paneles superiores), hay una notable presencia de datos en abril con una dispersión significativa de pesos a través de un amplio rango de calibres. Sin embargo, en marzo (panel superior derecho), no hay datos presentes, lo que podría indicar que no hubo registros de producción o entrega para ese cliente en ese mes. Para CLIENTE2 (paneles inferiores), hay datos disponibles en ambos meses, aunque la cantidad de datos en marzo es considerablemente menor y menos dispersa que en abril, lo que sugiere una menor variabilidad en los pesos de los calibres o un volumen más bajo de producción o entregas en ese mes.

**5.3.2.2. Gráfico de Líneas** Este gráfico utiliza la función `ggplot()` para crear un gráfico de líneas que muestra la relación entre calibre y `kg_calibre`, diferenciando por cliente. Los puntos se agregan para mejorar la visualización de los datos.

```
# Crear el gráfico de líneas
ggplot(calibres, aes(x=calibre, y=kg_calibre, colr=cliente, group=cliente)) +
```

```
geom_line() +
geom_point() +
labs(title="Relación entre Cal y Kg Cal",
      x="Calibre (Cal)", y="Kg por Cal") +
scale_color_viridis_d(option = "inferno") +
theme_minimal()
```



**Interpretación:** Se observa una tendencia general de que los pesos aumentan con los calibres más grandes, lo que es consistente con lo que uno esperaría en la producción de frutas. Sin embargo, hay una notable variabilidad en los pesos para calibres similares, lo que podría ser resultado de diferencias en la densidad de la fruta o en la precisión de las mediciones. Las líneas en el gráfico muestran fluctuaciones significativas, lo que indica que para un mismo calibre se registran diferentes pesos. Esto sugiere que el peso no depende exclusivamente del calibre y que otros factores pueden influir en el peso de las paltas Hass. Las variaciones podrían estar asociadas a factores como el contenido de agua, la madurez de la fruta o diferencias inherentes entre los lotes de producción.

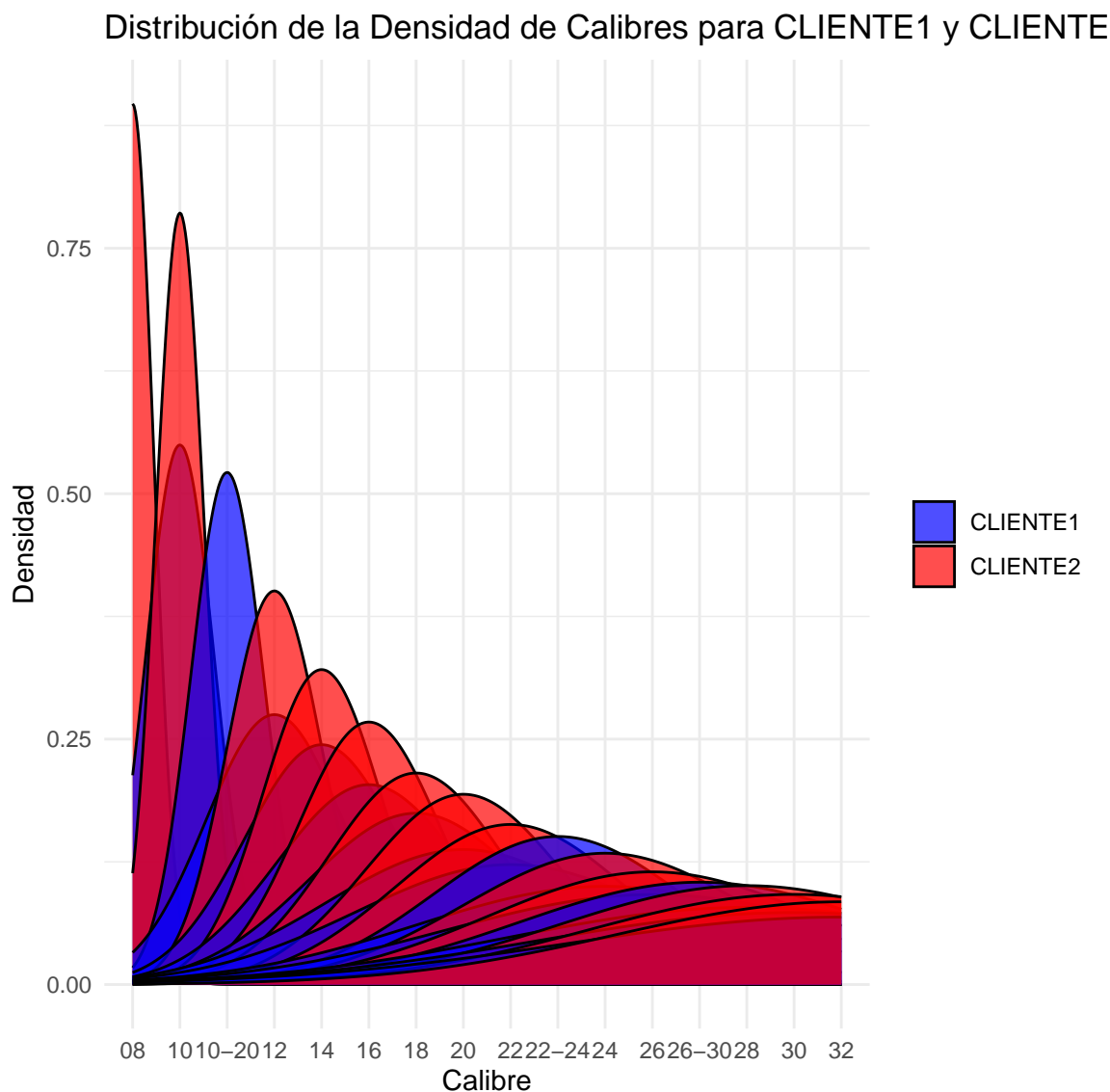
#### 5.4. Comparativa de Densidad de Calibres entre CLIENTE1 y CLIENTE2

Se genera el gráfico de densidad para visualizar y comparar la distribución de los calibres de las paltas Hass entre CLIENTE1 y CLIENTE2.



```
# Filtrar los datos para eliminar las filas con NAs en la columna 'calibre'
filtered_data <- combined_data %>%
  filter(!is.na(calibre))

# Crear el gráfico de densidad
ggplot(filtered_data, aes(x = calibre, fill = cliente.x)) +
  geom_density(alpha = 0.7) +
  scale_fill_manual(values = c("CLIENTE1" = "blue", "CLIENTE2" = "red")) +
  labs(title = "Distribución de la Densidad de Calibres para CLIENTE1 y CLIENTE2",
       x = "Calibre",
       y = "Densidad") +
  theme_minimal() +
  theme(legend.title = element_blank())
```



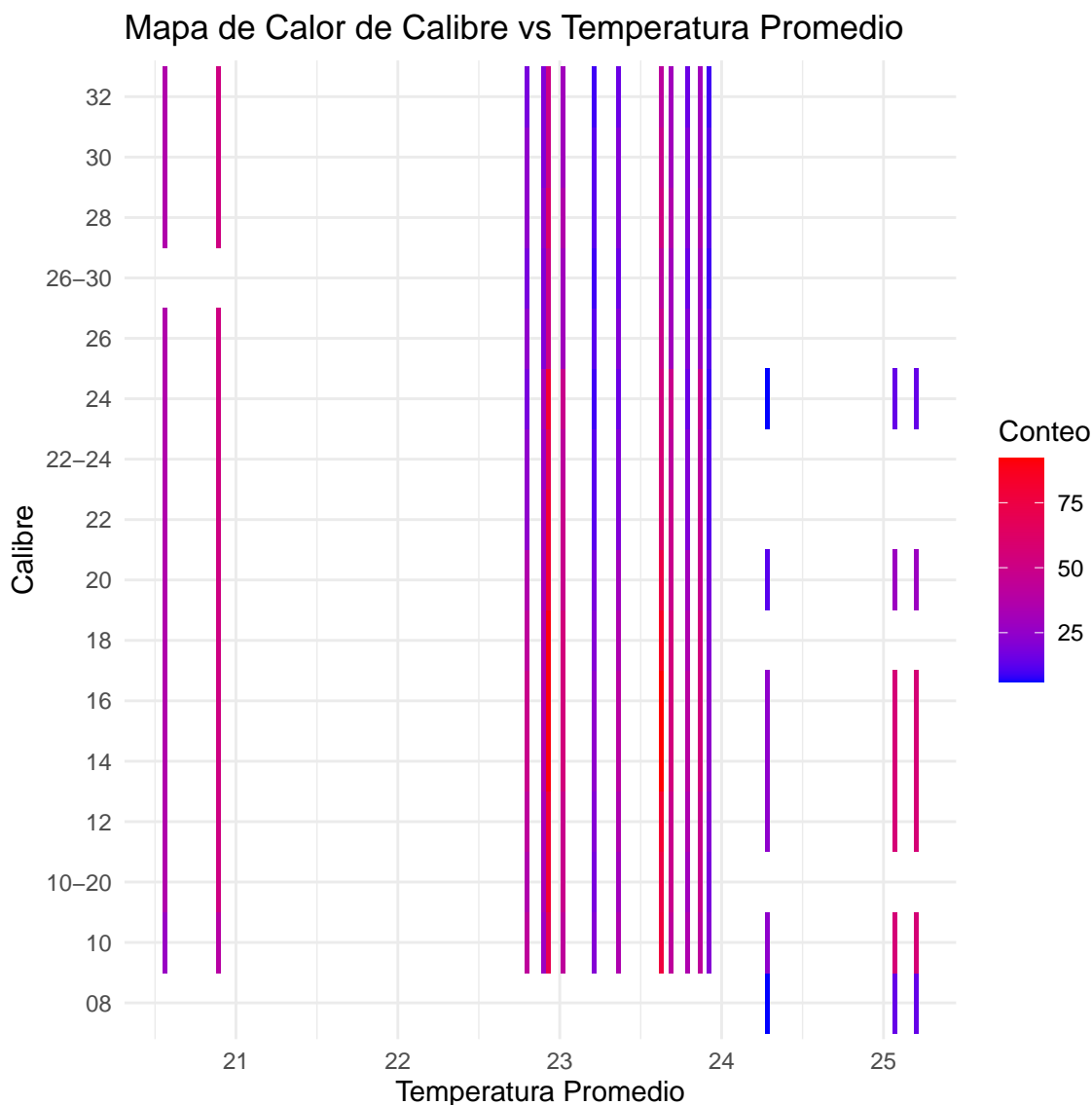
**Interpretación:** La gráfica muestra dos curvas de densidad superpuestas: CLIENTE1 en azul y CLIENTE2 en rojo. Se observa que CLIENTE1 tiene una distribución de calibres con un pico más pronunciado, lo que sugiere una mayor concentración en torno a un rango más estrecho de calibres. Por otro lado, CLIENTE2 presenta una curva más amplia con menos pico, indicando una distribución más dispersa de calibres. Ambas distribuciones muestran una caída conforme los calibres aumentan, pero CLIENTE2 tiene una cola más larga hacia los calibres más grandes, lo que podría indicar una mayor proporción de paltas Hass de calibre grande en comparación con CLIENTE1.

## 5.5. Mapas de calor

### 5.5.1. Mapa de Calor de Calibre vs Temperatura Promedio con Frecuencia de Datos

Se genera este mapa de calor para visualizar la frecuencia de diferentes calibres de paltas Hass en relación con la temperatura promedio, destacando las combinaciones más comunes.

```
# Mapa de calor para calibre vs temperatura promedio
ggplot(calibre_temp_freq, aes(x = temperatura_promedio, y = calibre, fill = count)) +
  geom_tile() +
  scale_fill_gradient(low = "blue", high = "red") +
  labs(title = "Mapa de Calor de Calibre vs Temperatura Promedio",
       x = "Temperatura Promedio",
       y = "Calibre",
       fill = "Conteo") +
  theme_minimal()
```



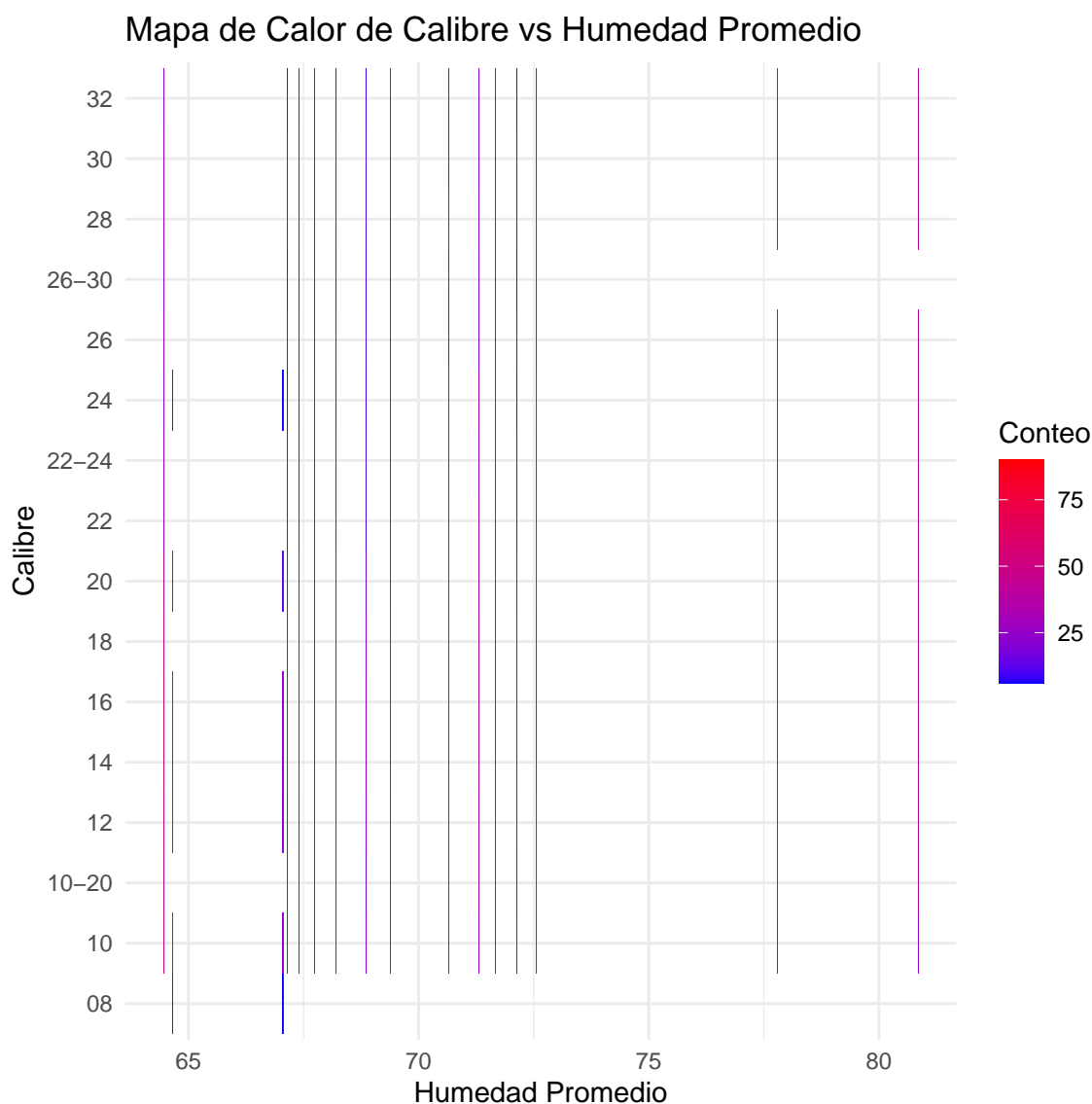
**Interpretación:** El gráfico muestra barras verticales donde el eje X representa la temperatura promedio y el eje Y los diferentes calibres. Cada barra tiene un color que indica la frecuencia de observaciones para esa combinación específica de temperatura y calibre, según la leyenda a la derecha. Las barras más oscuras y más rojas indican una mayor frecuencia. Se nota una concentración de frecuencias más altas (colores más oscuros) en temperaturas promedio de alrededor de 24°C, especialmente para calibres en el rango de 22-24. Esto podría sugerir que las temperaturas

promedio de alrededor de 24°C están asociadas con una mayor frecuencia de paltas de calibre mediano a grande. Las combinaciones de calibres más pequeños y más grandes son menos frecuentes a lo largo de las temperaturas observadas.

### 5.5.2. Distribución de Calibres en Función de la Humedad Promedio y Frecuencia de Observaciones

Se genera este mapa de calor que ilustra cómo los distintos calibres de paltas Hass se distribuyen con respecto a la humedad promedio, resaltando las combinaciones más prevalentes.

```
# Mapa de calor para calibre vs humedad promedio
ggplot(calibre_hum_freq, aes(x = humedad_promedio, y = calibre, fill = count)) +
  geom_tile() +
  scale_fill_gradient(low = "blue", high = "red") +
  labs(title = "Mapa de Calor de Calibre vs Humedad Promedio",
       x = "Humedad Promedio",
       y = "Calibre",
       fill = "Conteo") +
  theme_minimal()
```



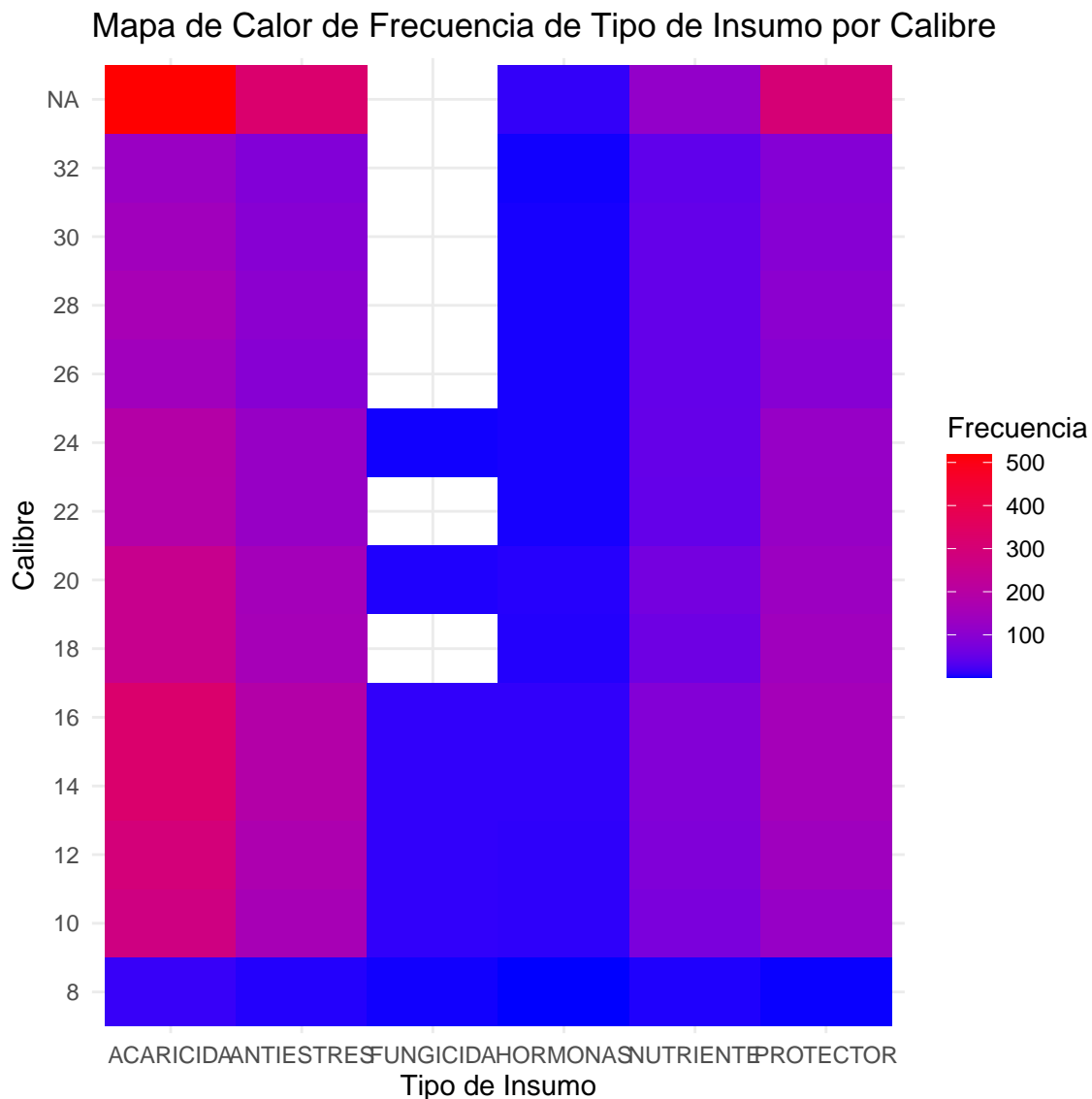
**Interpretación:** Las barras en este gráfico representan la frecuencia de aparición de cada calibre a distintos niveles de humedad promedio. El color de cada barra varía según la cantidad de veces que se registró un calibre específico

a una humedad promedio dada, con tonos que se intensifican hacia el rojo a medida que aumenta la frecuencia de observaciones, como lo indica la leyenda del color. Se observa una mayor concentración de frecuencias (colores más oscuros) alrededor de los 70 % de humedad promedio, especialmente para calibres entre 22 y 24, lo que sugiere que este nivel de humedad podría estar asociado con una mayor ocurrencia de frutos de tamaño mediano a grande. Los calibres extremadamente altos o bajos aparecen con menos frecuencia a través de los rangos de humedad representados.

### 5.5.3. Mapa de Calor de la Frecuencia de Uso de Tipo de Insumo por Calibre de Palta

Se crea este mapa de calor para explorar la frecuencia con la que se utilizan diferentes tipos de insumos en relación con los calibres de las paltas Hass, lo que puede indicar preferencias o necesidades específicas de tratamiento según el tamaño del fruto.

```
# Crear el mapa de calor
ggplot(calibre_insumo_freq, aes(x = tipo_insumo, y = calibre, fill = frequency)) +
  geom_tile() +
  scale_fill_gradient(low = "blue", high = "red") +
  labs(title = "Mapa de Calor de Frecuencia de Tipo de Insumo por Calibre",
       x = "Tipo de Insumo",
       y = "Calibre",
       fill = "Frecuencia") +
  theme_minimal()
```



**Interpretación:** En el gráfico, los colores indican la frecuencia de uso de cada tipo de insumo para cada calibre, donde los tonos más rojos representan una mayor frecuencia y los azules una menor. La alta frecuencia de ‘ACARICIDA’ en calibres de 14 a 16 sugiere un uso intensivo en frutos de tamaño medio. Los ‘FUNGICIDA’ y ‘PROTECTOR’ se usan comúnmente en un rango más amplio de calibres, mientras que los ‘HORMONAS’ y ‘ANTIESTRÉS’ muestran una menor frecuencia de uso en general. Los ‘NUTRIENTE’ muestran una aplicación relativamente uniforme en todos los calibres, lo que puede reflejar una práctica agrícola estándar independiente del tamaño del fruto.

## 6. Análisis Inferencial

### 6.1. Pruebas de Hipótesis

#### 6.1.1. Análisis de Varianza (ANOVA)

**6.1.1.1. Análisis de Variación en los Kilogramos de Calibre de Paltas Hass por Semana** Para comprender mejor la variabilidad en el tamaño de los calibres de paltas Hass y su posible relación con factores temporales, se llevó a cabo un Análisis de Varianza (ANOVA), seguido de pruebas de normalidad y pruebas no paramétricas para una comprensión más profunda de las diferencias entre las semanas críticas de cosecha.

```
# ANOVA para comparar kg_calibre entre diferentes semanas
anova_result <- aov(kg_calibre ~ semana, data = calibres)
```

```
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
```

**Interpretación:** El ANOVA realizado indica diferencias significativas en los kilogramos de calibre entre las semanas (p-valor = 0.0138), lo que sugiere que la semana tiene un impacto relevante en el tamaño de los calibres de las paltas.

```
shapiro.test(residuals(anova_result))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(anova_result)
## W = 0.82446, p-value < 2.2e-16
```

**Interpretación:** La prueba de Shapiro-Wilk para los residuos del ANOVA muestra un p-valor muy bajo (< 2.2e-16), lo que indica una desviación significativa de la normalidad. Esto nos lleva a cuestionar la validez de los resultados del ANOVA, ya que uno de los supuestos clave no se cumple.

```
kruskal.test(kg_calibre ~ semana, data = calibres)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  kg_calibre by semana
## Kruskal-Wallis chi-squared = 13.94, df = 3, p-value = 0.002987
```

**Interpretación:** Dado que los residuos del ANOVA no se distribuyen normalmente, se procedió con la prueba de Kruskal-Wallis, que no asume la normalidad. Los resultados de esta prueba (p-valor = 0.002987) confirman que existen diferencias significativas en los pesos por calibre entre las semanas.

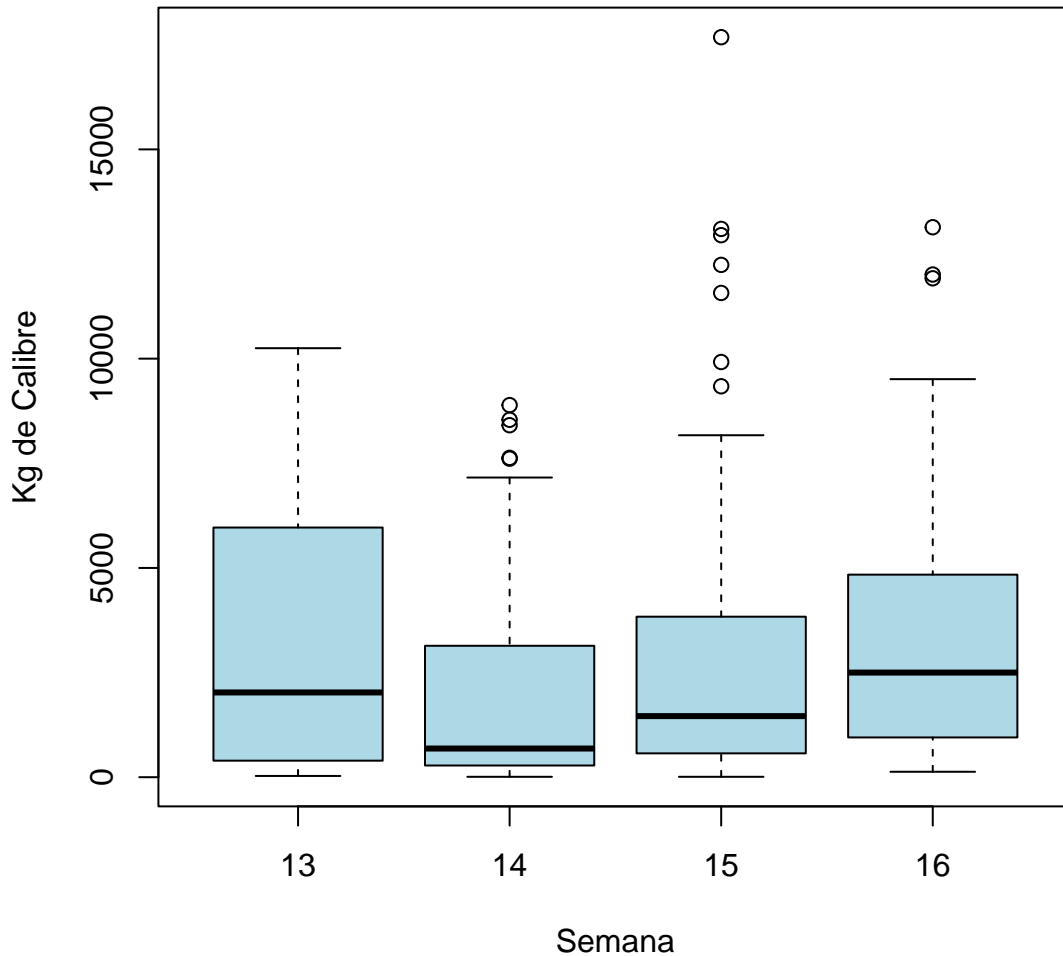
```
dunnTest(kg_calibre ~ semana, data = calibres, method="bonferroni")
```

```
## Comparison      Z      P.unadj      P.adj
## 1    13 - 14  1.6629979 0.0963128575 0.577877145
## 2    13 - 15  0.6126887 0.5400821886 1.000000000
## 3    14 - 15 -1.7970898 0.0723213635 0.433928181
## 4    13 - 16 -0.8840695 0.3766587144 1.000000000
## 5    14 - 16 -3.6783190 0.0002347762 0.001408657
## 6    15 - 16 -2.2909524 0.0219661674 0.131797004
```

**Interpretación:** Tras identificar diferencias significativas con Kruskal-Wallis, la prueba post-hoc de Dunn se utilizó para determinar entre qué semanas existen estas diferencias. Los resultados indican diferencias significativas entre las semanas 14 y 16 (p-valor ajustado = 0.001408657), lo que sugiere que las intervenciones o condiciones en estas semanas específicas tienen efectos distintos en los calibres de las paltas.

```
boxplot(kg_calibre ~ semana, data = calibres,
        main = "Distribución de kg de calibre por semana",
        xlab = "Semana",
        ylab = "Kg de Calibre",
        col = "lightblue")
```

## Distribución de kg de calibre por semana



**Interpretación:** El gráfico de cajas visualiza la distribución de los kilogramos de calibre por semana. Las medias y la variabilidad se ilustran claramente, mostrando la presencia de valores atípicos y proporcionando una representación visual que respalda los análisis estadísticos realizados.

Este análisis integral revela variaciones significativas en los tamaños de calibre de paltas Hass a lo largo de las semanas estudiadas, con la semana jugando un papel importante en estas diferencias. La combinación de pruebas paramétricas y no paramétricas ofrece una vista completa de la variabilidad en la producción, crucial para la formulación de estrategias que mejoren la gestión agrícola y la adaptación a las variables estacionales. Las recomendaciones futuras deberán centrarse en las semanas con variaciones significativas para mejorar la consistencia en el tamaño del calibre, considerando las prácticas de cultivo y las condiciones climáticas de dichas semanas.

**6.1.1.2. Análisis de Variación Semanal en los Kilogramos Brutos por Hectárea** Para profundizar en la comprensión de las dinámicas de producción agrícola semana a semana, llevamos a cabo un Análisis de Varianza (ANOVA). Este ANOVA se propone para evaluar si existen diferencias estadísticamente significativas en los kilogramos brutos por hectárea entre las distintas semanas de cosecha. Los resultados pueden ofrecer insights sobre la consistencia de la producción y potencialmente revelar la influencia de factores temporales o intervenciones agrícolas.

```
# Verificar que 'semana' es un factor
reporte$semana <- as.factor(reporte$semana)
```

```
# ANOVA para comparar los kg_brutos_hectarea entre las distintas semanas
anova_kg_hectarea <- aov(kg_brutos_hectarea ~ semana, data = reporte)
summary(anova_kg_hectarea)
```

```
##              Df    Sum Sq Mean Sq F value    Pr(>F)
## semana         3  5112645 1704215    5.259 0.00194 **
## Residuals    117 37914547  324056
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Interpretación:** El resultado del ANOVA indica que hay diferencias significativas en la producción media de kg brutos por hectárea entre las distintas semanas (p-valor = 0.00194). Esto sugiere que las semanas pueden tener un impacto diferencial en los resultados de producción debido a variables como las prácticas de manejo agrícola, las condiciones climáticas o la aplicación de tratamientos.

```
TukeyHSD(anova_kg_hectarea)
```

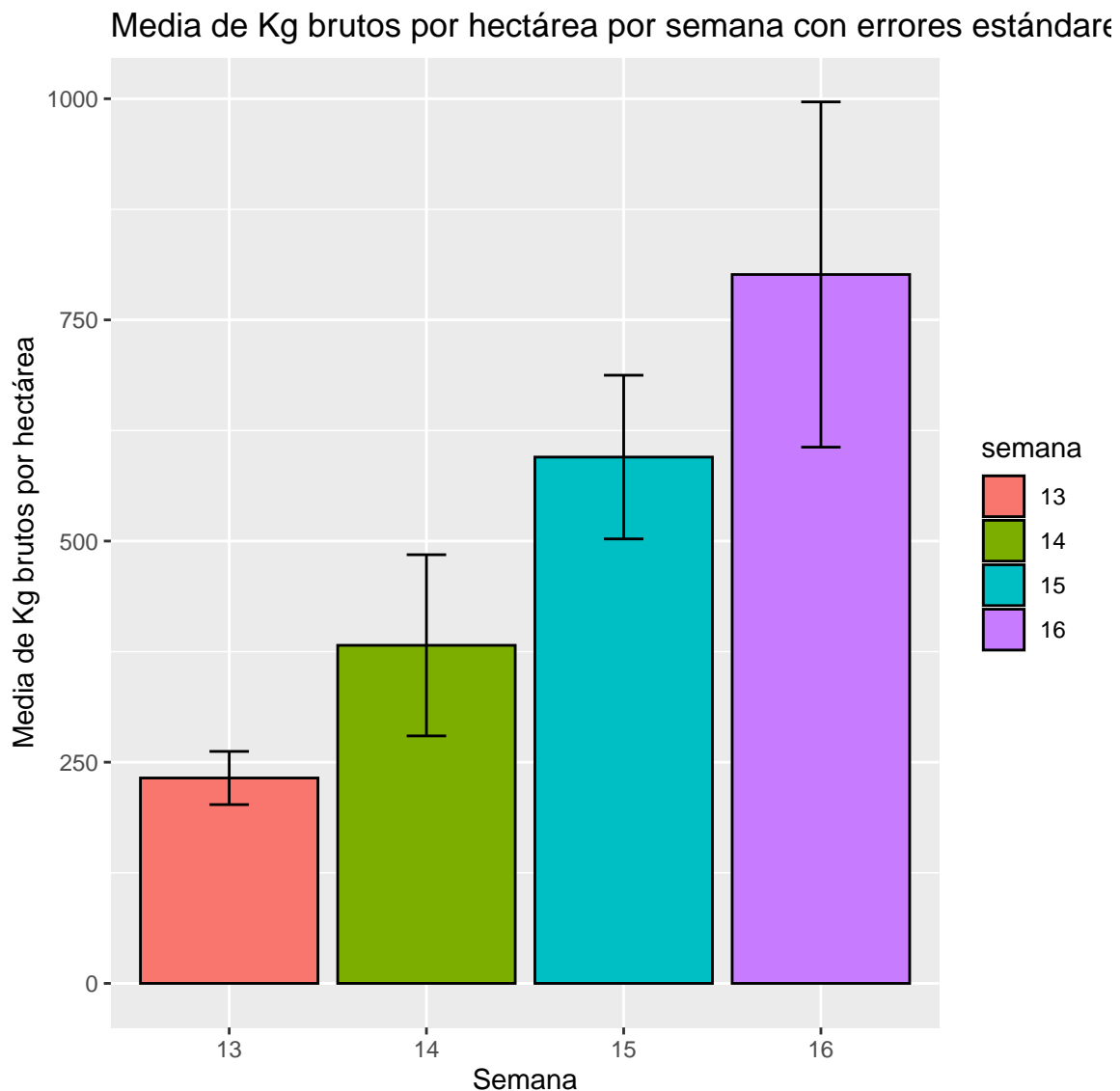
```
##    Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = kg_brutos_hectarea ~ semana, data = reporte)
##
## $semana
##           diff             lwr             upr             p adj
## 14-13 150.0055 -225.0304496  525.0414  0.7248682
## 15-13 362.7780   7.9650859  717.5908  0.0430344
## 16-13 569.1351 163.1740944  975.0960  0.0021694
## 15-14 212.7725 -157.4369554  582.9819  0.4420126
## 16-14 419.1296  -0.3548129  838.6140  0.0502803
## 16-15 206.3571 -195.1492939  607.8635  0.5397835
```

**Interpretación:** Las pruebas post-hoc de Tukey ayudan a identificar entre qué semanas específicas se encuentran las diferencias significativas. En este caso, la semana 16 se diferencia significativamente de la semana 13 (p-valor ajustado = 0.0021694), lo cual es consistente con la tendencia observada en el gráfico y subraya la importancia de entender qué sucedió en estas semanas para informar las decisiones de manejo.

```
# Calcular medias y errores estándares
summary_data <- reporte %>%
  group_by(semana) %>%
  summarise(Media = mean(kg_brutos_hectarea),
            SE = sd(kg_brutos_hectarea) / sqrt(n()))

# Gráfico de barras con errores estándares
ggplot(summary_data, aes(x = semana, y = Media, fill = semana)) +
  geom_bar(stat = "identity", position = position_dodge(), color = "black") +
  geom_errorbar(aes(ymin = Media - SE, ymax = Media + SE), width = .2,
               position = position_dodge(.9)) +
  labs(x = "Semana", y = "Media de Kg brutos por hectárea",
       title = "Media de Kg brutos por hectárea por semana con errores estándares")
```





**Interpretación:** El gráfico de barras presenta una visualización clara de las diferencias en la producción promedio por hectárea a lo largo de las semanas, con los errores estándares destacando la variabilidad de la producción. La semana 16 muestra no solo la mayor media de producción, sino también la mayor variabilidad, lo que puede sugerir factores adicionales afectando la producción durante esa semana. Este entendimiento puede ser vital para ajustes específicos en la planificación de las actividades agrícolas para las semanas futuras.

**6.1.1.3. Análisis de Variación de los calibres entre diferentes semanas** Se realiza el 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.

```
# Validar que 'semana' sea un factor y 'kg_calibre' sea numérico
calibres$semana <- as.factor(calibres$semana)
calibres$kg_calibre <- as.numeric(calibres$kg_calibre)
```

```
# Realizar ANOVA para comparar 'kg_calibre' a través de diferentes 'semanas'
anova_result <- aov(kg_calibre ~ semana, data = calibres)
```

```
# Revisar el resumen del 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
```

```
# Generar el gráfico QQ
```

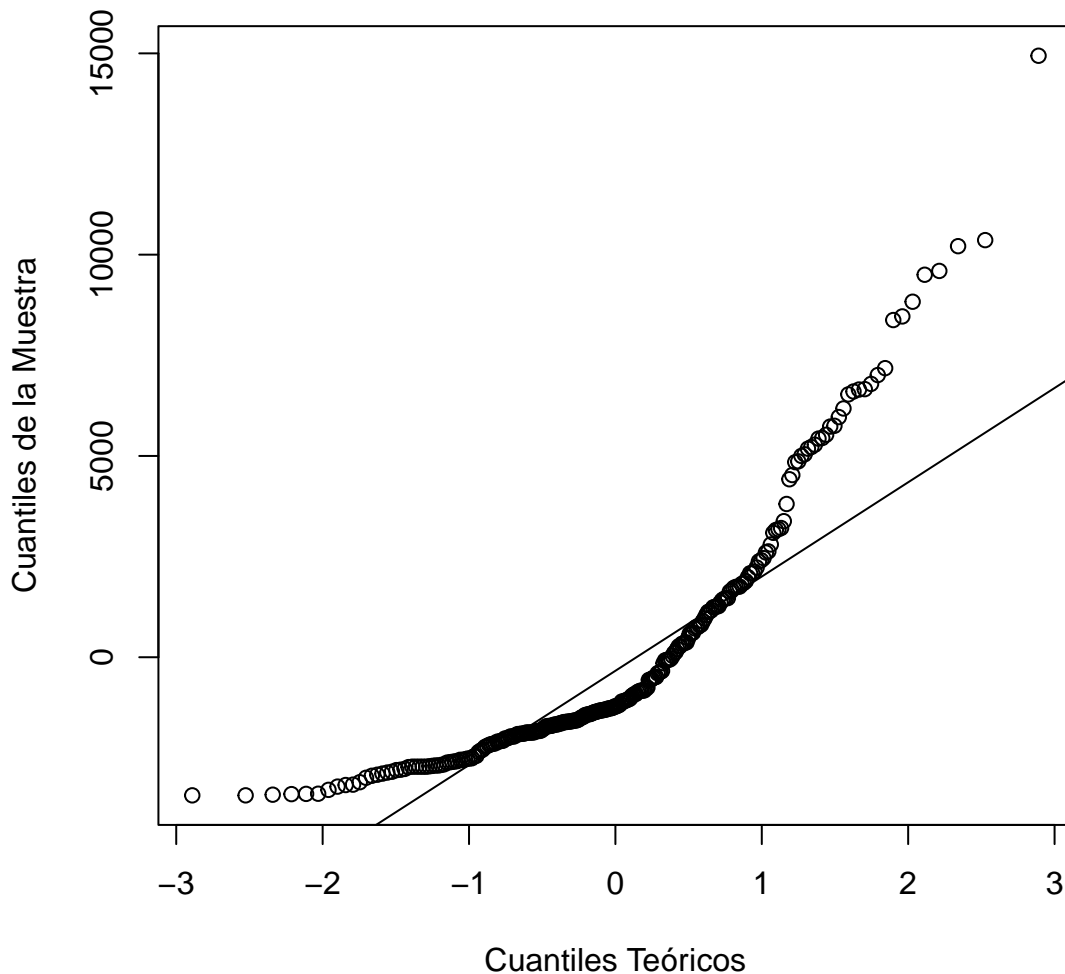
```
qqnorm(residuals(anova_result), main = "", xlab = "", ylab = "")
```

```
qqline(residuals(anova_result))
```

```
# Personalizar título y etiquetas de los ejes
```

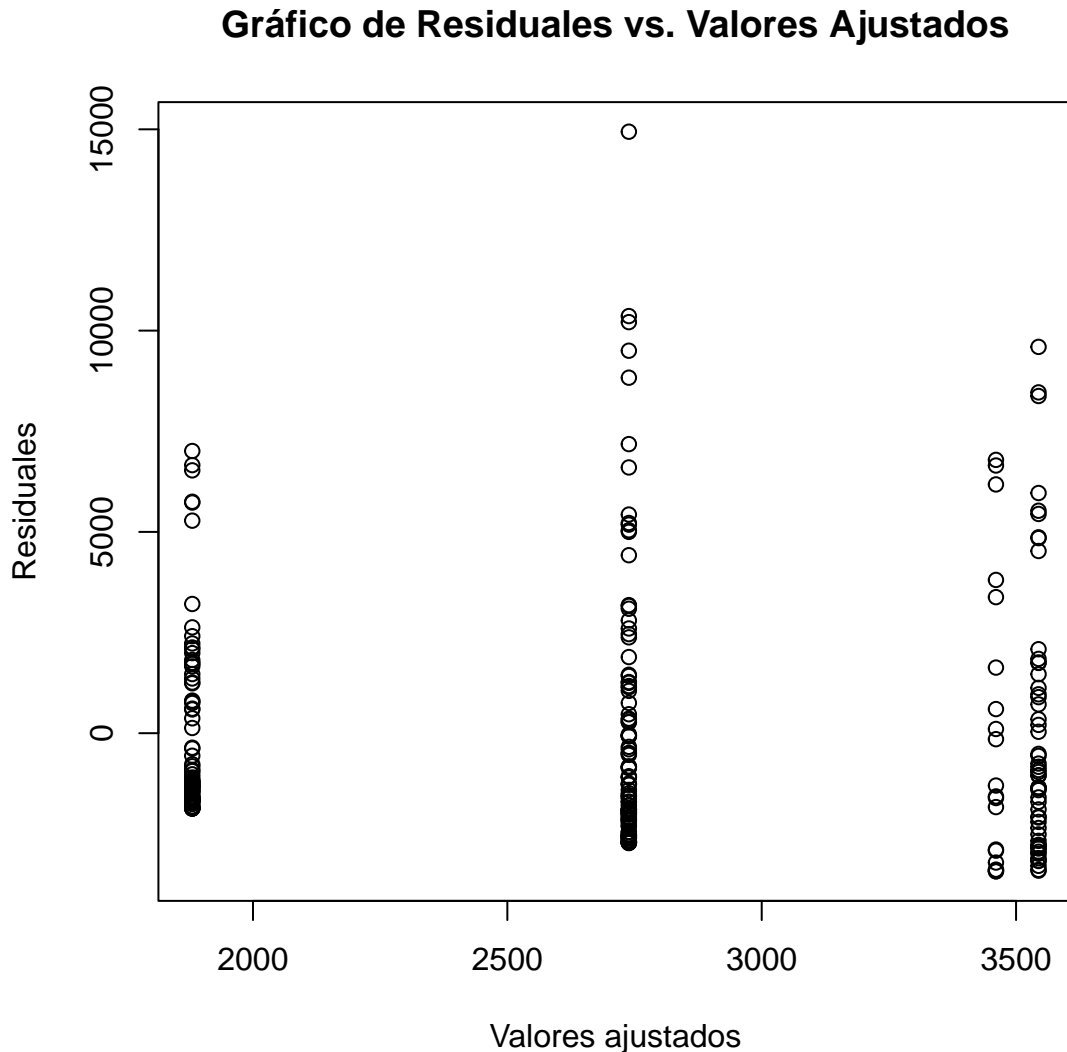
```
title(main = "Gráfico QQ de los Residuales para Verificar la Normalidad",
      xlab = "Cuantiles Teóricos",
      ylab = "Cuantiles de la Muestra")
```

## Gráfico QQ de los Residuales para Verificar la Normalidad



**Interpretación:** Este gráfico se utiliza para evaluar si una variable tiene una distribución normal. En la interpretación del QQ plot, si los puntos caen aproximadamente a lo largo de la línea de referencia, entonces la variable tiene una distribución normal. En este caso, vemos que los puntos se desvían significativamente de la línea en los extremos, lo que sugiere que la distribución de los kg por calibre no es normal, con colas más pesadas en ambos extremos.

```
# Homogeneidad de varianzas
plot(residuals(anova_result) ~ fitted(anova_result),
     xlab = "Valores ajustados",
     ylab = "Residuales",
     main = "Gráfico de Residuales vs. Valores Ajustados")
```



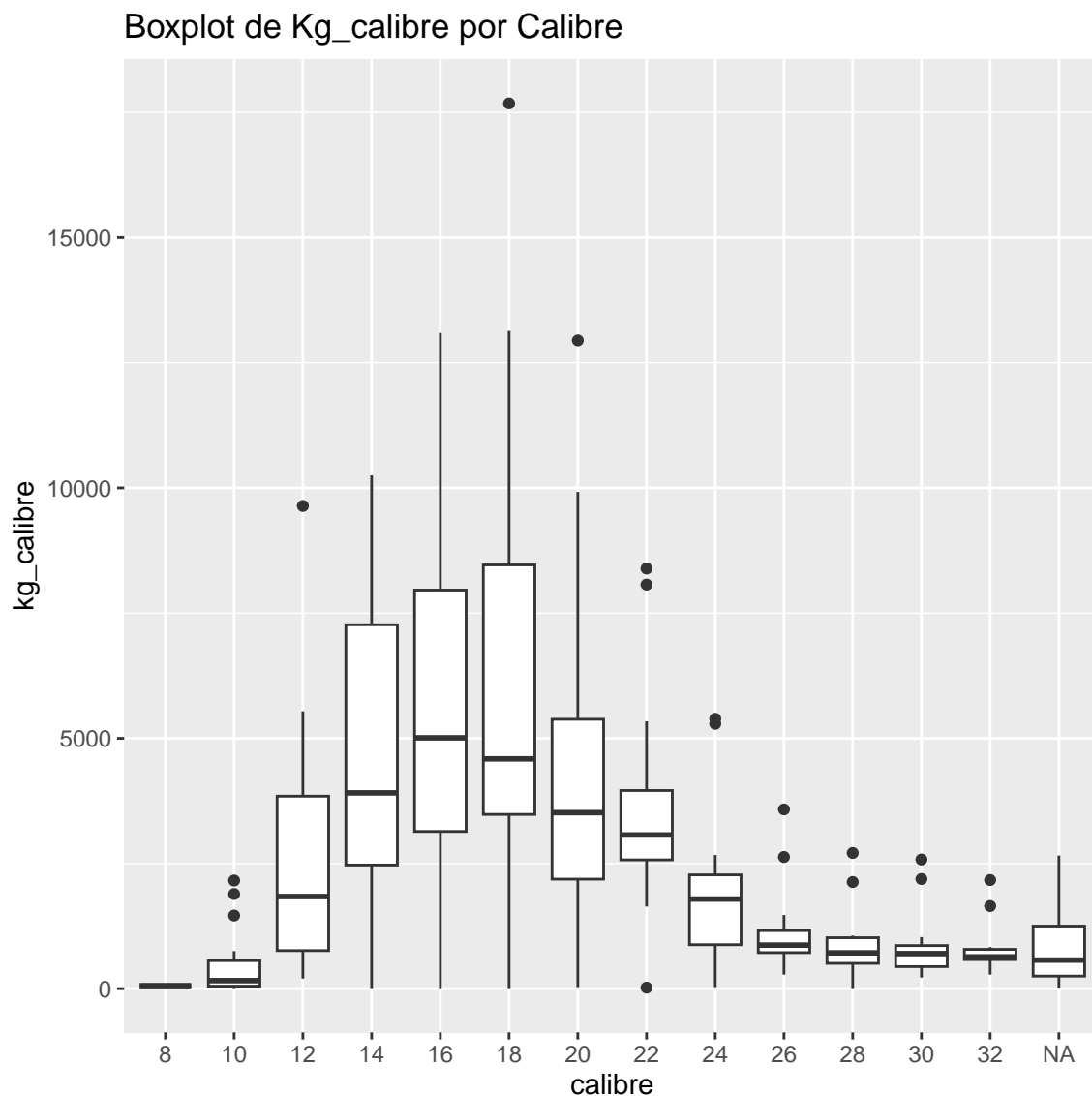
**Interpretación:** Este gráfico muestra los residuales del modelo ANOVA en relación con los valores ajustados. Idealmente, queremos que los puntos estén distribuidos aleatoriamente alrededor de la línea horizontal en el centro (que indica 0 residual). La presencia de patrones o formas discernibles puede indicar problemas con el modelo, como la no linealidad. Aquí, no hay un patrón claro que sugiera no linealidad, aunque la varianza de los residuos parece incrementarse con los valores ajustados, lo que podría indicar heterocedasticidad.

```
# Si el ANOVA resulta significativo, realizar pruebas post-hoc
TukeyHSD(anova_result)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = kg_calibre ~ semana, data = calibres)
##
```

```
## $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 0.3959994

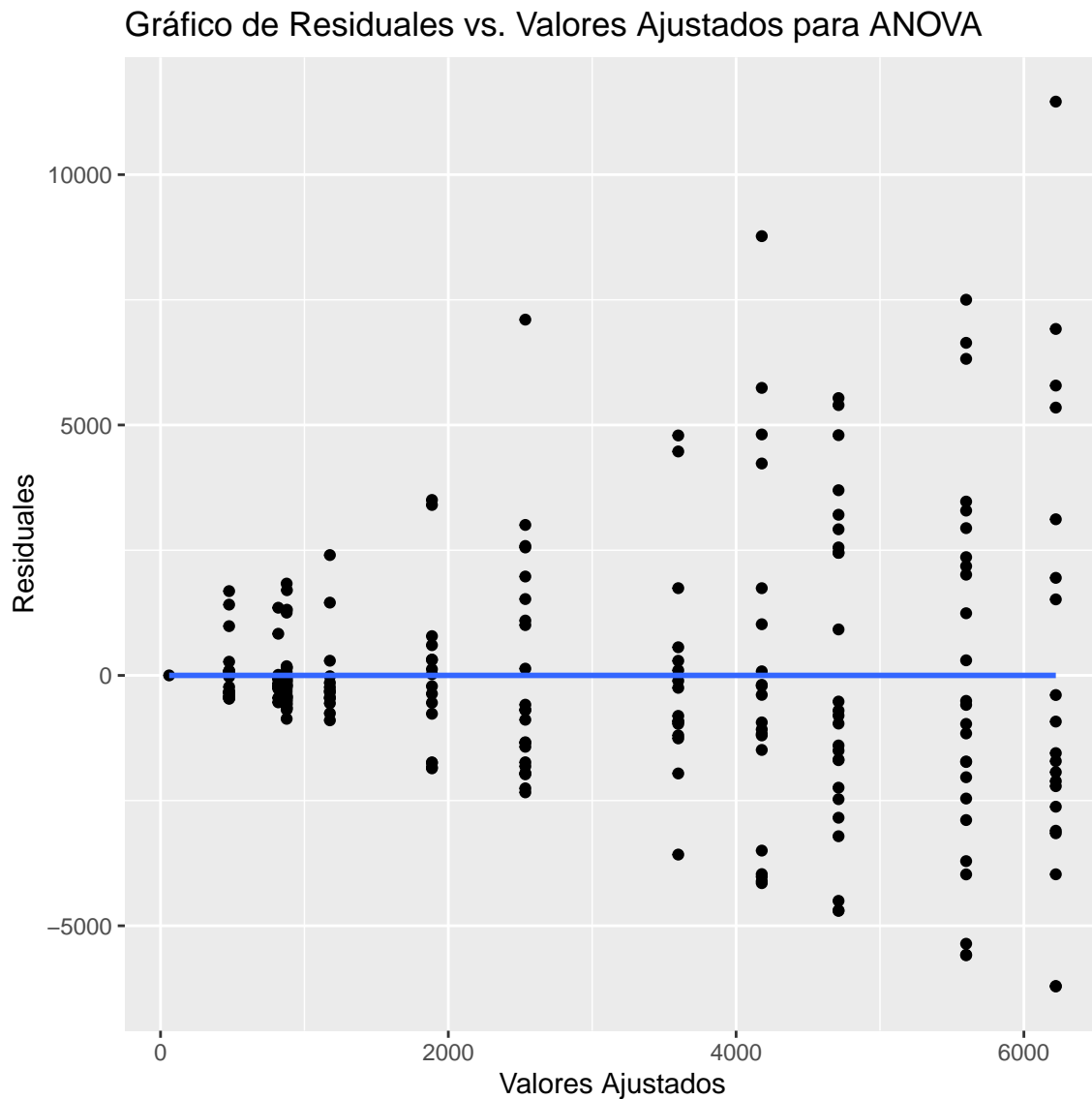
# Diagrama de caja para 'kg_calibre' a través de diferentes 'calibre'
ggplot(calibres, aes(x=calibre, y=kg_calibre)) +
  geom_boxplot() +
  labs(title="Boxplot de Kg_calibre por Calibre", x="calibre", y="kg_calibre")
```



**Interpretación:** Los boxplots muestran la distribución de los kg por calibre. Observamos que a medida que aumenta el calibre, el rango y la mediana de los kg también tienden a aumentar, pero con una variabilidad considerable, especialmente para los calibres medianos a altos. Los valores atípicos están presentes en todos los calibres, lo que sugiere que hay paltas con pesos que son inusualmente altos o bajos para su calibre.

```
# Realizar el ANOVA
anova_result <- aov(kg_calibre ~ calibre, data = calibres)
```

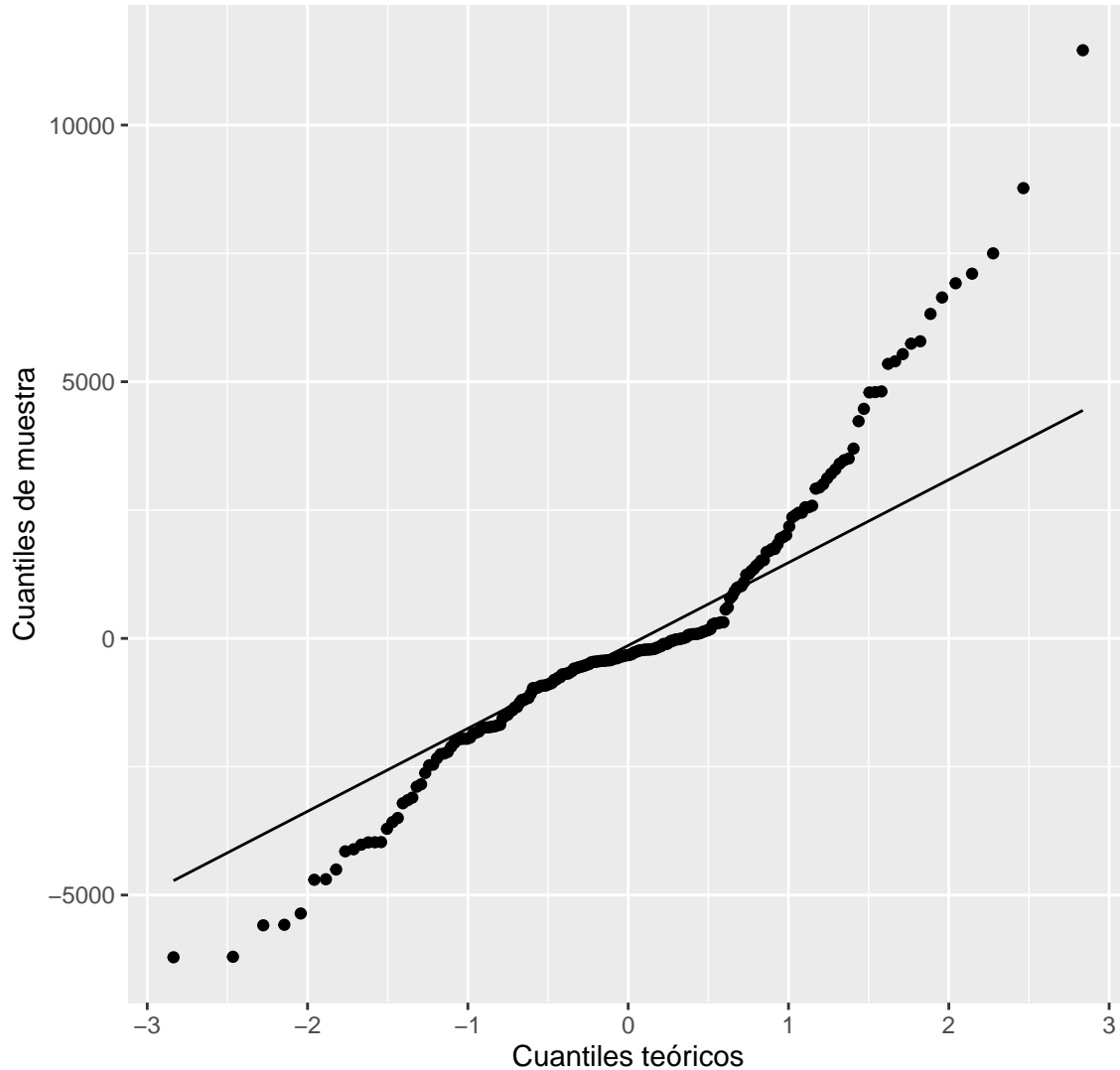
```
# Gráfico de residuos
res_data <- data.frame(residuals= residuals(anova_result), fitted=fitted(anova_result))
ggplot(res_data, aes(x=fitted, y=residuals)) +
  geom_point() +
  geom_smooth(method="lm", se=FALSE) +
  labs(title="Gráfico de Residuales vs. Valores Ajustados para ANOVA", x="Valores Ajustados", y="Residuales")
```



**Interpretación:** Este gráfico es similar al segundo, con la adición de una línea de regresión. La línea plana indica que no hay una tendencia clara entre los residuos y los valores ajustados, lo que es bueno para el modelo ANOVA. Sin embargo, al igual que antes, la varianza de los residuos parece incrementarse con los valores ajustados.

```
# QQ plot para Residuales
ggplot(res_data, aes(sample=residuals)) +
  geom_qq() +
  geom_qq_line() +
  labs(title="Gráfico QQ de los Residuales", x="Cuantiles teóricos", y="Cuantiles de muestra")
```

Gráfico QQ de los Residuales



**Interpretación:** Similar al primer gráfico, este QQ plot muestra los residuos del modelo ANOVA en comparación con una distribución normal teórica. De nuevo, observamos una desviación significativa de la línea en los extremos, lo que indica que los residuos no siguen una distribución normal, especialmente en los valores extremos.

La interpretación de estos gráficos sugiere que, aunque se encontraron diferencias estadísticamente significativas en el ANOVA, las suposiciones de normalidad y homogeneidad de varianzas no se cumplen del todo. Esto podría afectar la fiabilidad de los resultados del ANOVA y podría requerir transformaciones de datos o la utilización de métodos no paramétricos. Además, el modelo podría no capturar toda la variabilidad en los datos, como se sugiere por los valores atípicos y la posible heterocedasticidad observada.

**6.1.1.4. Análisis de Variación para Kilogramos por Hectárea** Se lleva a cabo la evaluación de los supuestos de normalidad y homogeneidad de varianzas que son cruciales para la interpretación correcta de un ANOVA. Los gráficos de diagnóstico estadístico proporcionan una verificación visual de estos supuestos. Los QQ plots ayudan a determinar si los residuos siguen una distribución normal, y los gráficos de residuos contra valores ajustados examinan si los residuos tienen varianzas constantes a través de los niveles del factor.

```
# Asegurar que 'semana' sea un factor y 'kg_brutos_hectarea' sea numérico
reporte$semana <- as.factor(reporte$semana)
reporte$kg_brutos_hectarea <- as.numeric(reporte$kg_brutos_hectarea)

# ANOVA para comparar 'kg_brutos_hectarea' a través de diferentes 'semana'
```

```

anova_result <- aov(kg_brutos_hectarea ~ semana, data = reporte)

# Revisar el resumen del ANOVA
summary(anova_result)

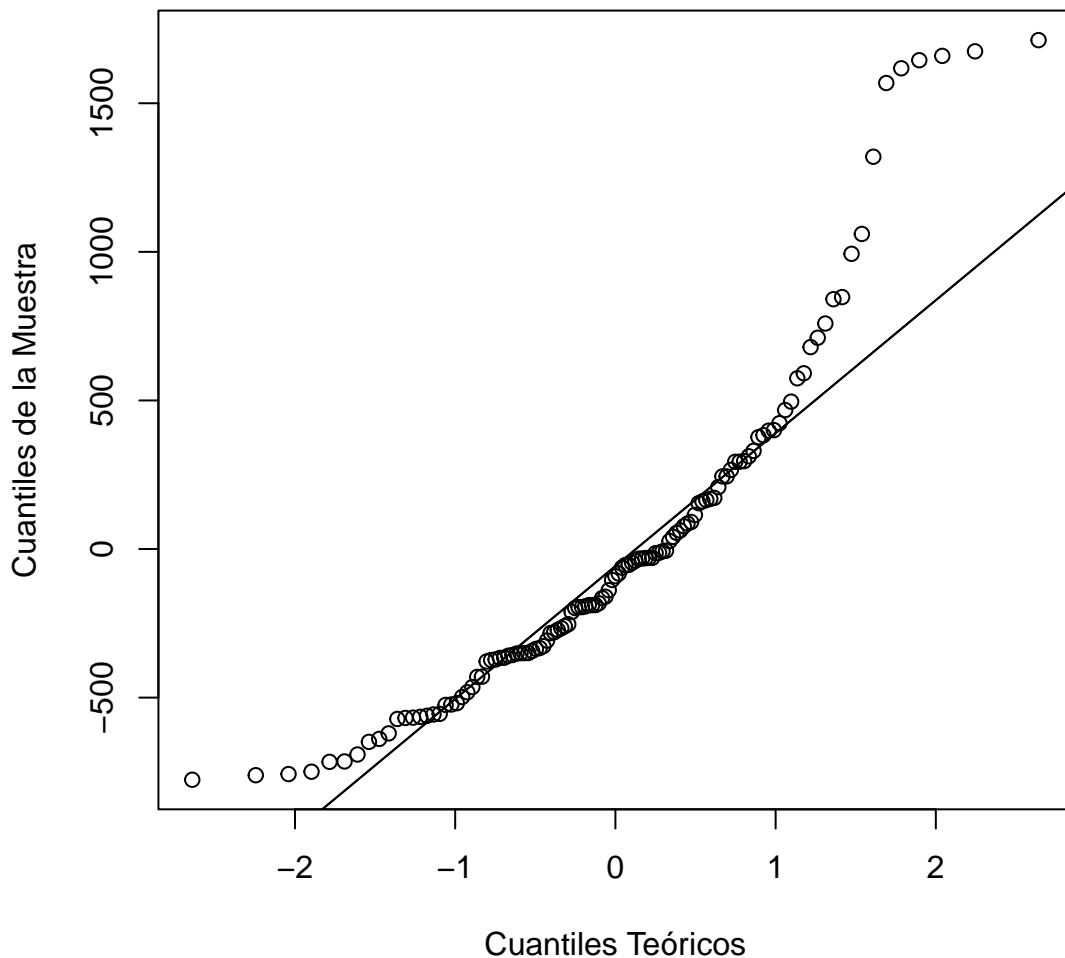
##              Df    Sum Sq Mean Sq F value    Pr(>F)    
## semana         3   5112645  1704215    5.259 0.00194 ** 
## Residuals    117  37914547   324056                
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Generar el gráfico QQ
qqnorm(residuals(anova_result), main = "", xlab = "", ylab = "")
qqline(residuals(anova_result))

# Personalizar título y etiquetas de los ejes
title(main = "Gráfico QQ de los Residuales para Verificar la Normalidad",
      xlab = "Cuantiles Teóricos",
      ylab = "Cuantiles de la Muestra")
qqline(residuals(anova_result))

```

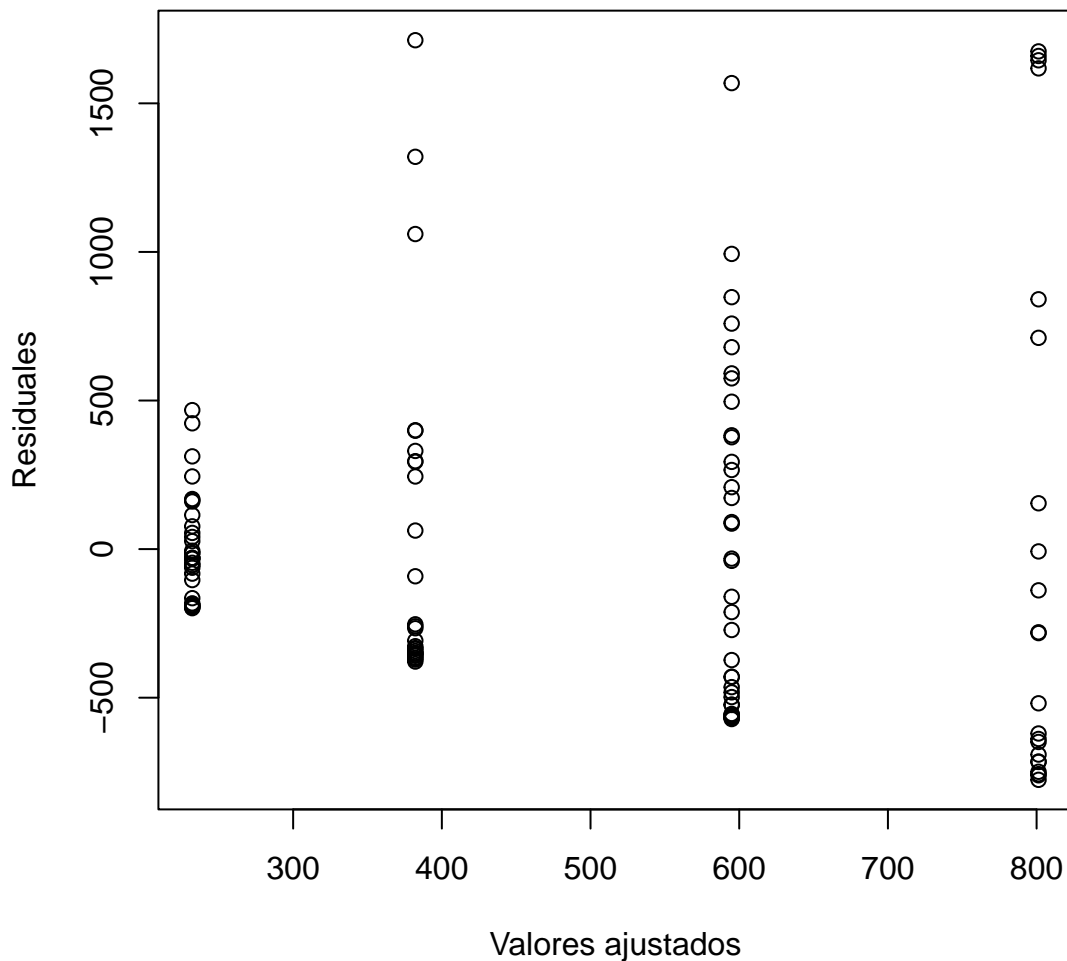
## Gráfico QQ de los Residuales para Verificar la Normalidad



**Interpretación:** Este gráfico muestra si los residuos de la ANOVA se distribuyen normalmente al comparar los cuantiles teóricos de una distribución normal con los cuantiles de los residuos observados. Los puntos siguen bastante de cerca la línea en el centro, pero se desvían en los extremos. Esto sugiere que los residuos tienen colas más pesadas que una distribución normal, indicando posibles desviaciones de la normalidad, especialmente para valores extremos.

```
# Homogeneidad de varianzas
plot(residuals(anova_result) ~ fitted(anova_result),
     xlab = "Valores ajustados",
     ylab = "Residuales",
     main = "Gráfico de Residuales vs. Valores Ajustados")
```

## Gráfico de Residuales vs. Valores Ajustados



**Interpretación:** Este gráfico muestra los residuos del modelo en función de los valores ajustados. Los residuos deben distribuirse al azar y no mostrar patrones discernibles si el modelo se ajusta bien. Aquí, los residuos parecen distribuirse al azar sin un patrón claro, lo cual es positivo. Sin embargo, parece haber una tendencia de varianza creciente para valores ajustados más altos, lo que puede indicar heterocedasticidad.

```
# Si el ANOVA resulta significativo, realizar pruebas post-hoc
TukeyHSD(anova_result)
```

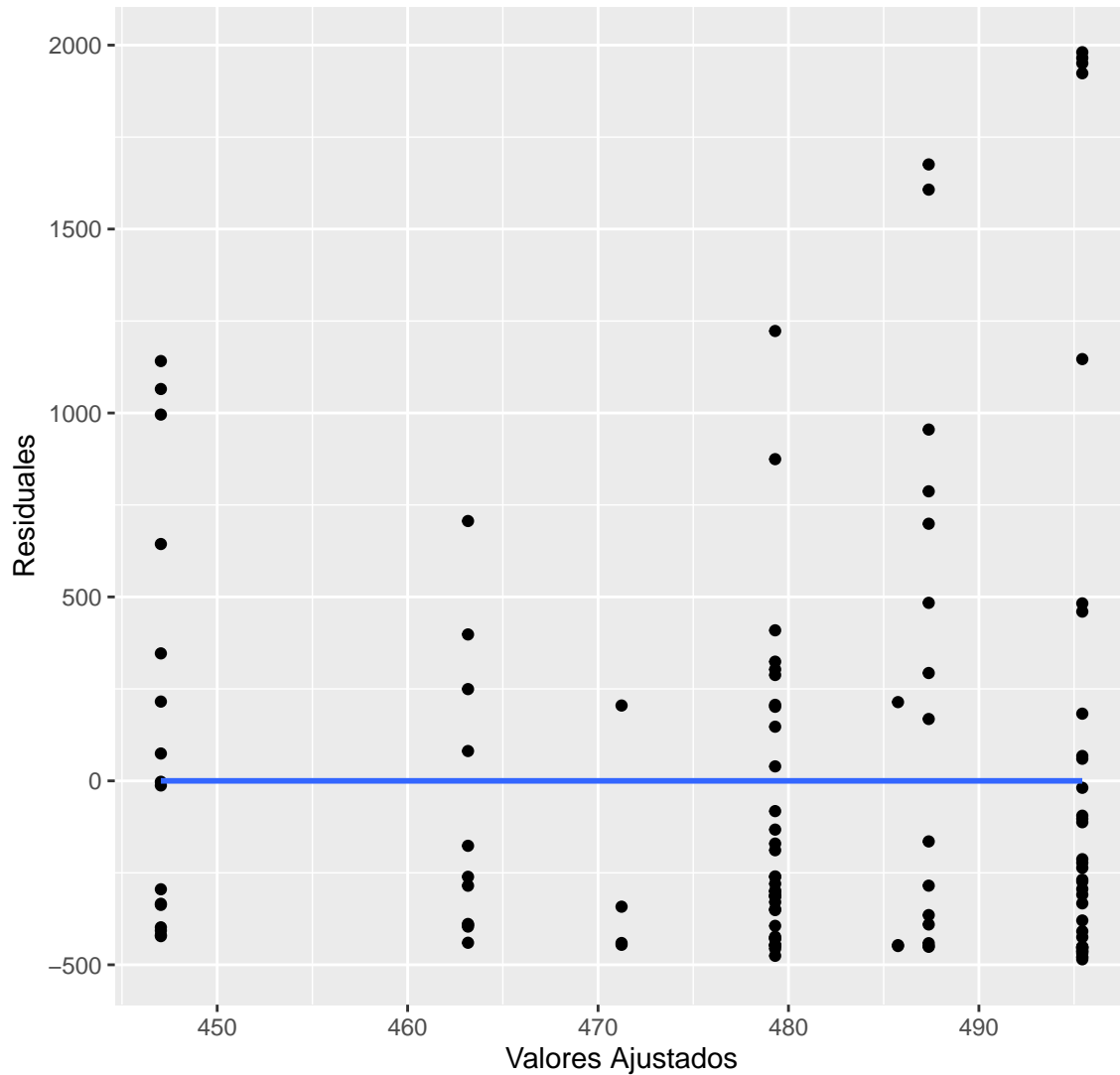
```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
```



```
##
## Fit: aov(formula = kg_brutos_hectarea ~ semana, data = reporte)
##
## $semana
##          diff          lwr          upr          p adj
## 14-13 150.0055 -225.0304496 525.0414 0.7248682
## 15-13 362.7780   7.9650859 717.5908 0.0430344
## 16-13 569.1351 163.1740944 975.0960 0.0021694
## 15-14 212.7725 -157.4369554 582.9819 0.4420126
## 16-14 419.1296  -0.3548129 838.6140 0.0502803
## 16-15 206.3571 -195.1492939 607.8635 0.5397835
# Ejecutar ANOVA
anova_result <- aov(kg_brutos_hectarea ~ dosis_litros_por_hectarea, data = reporte)

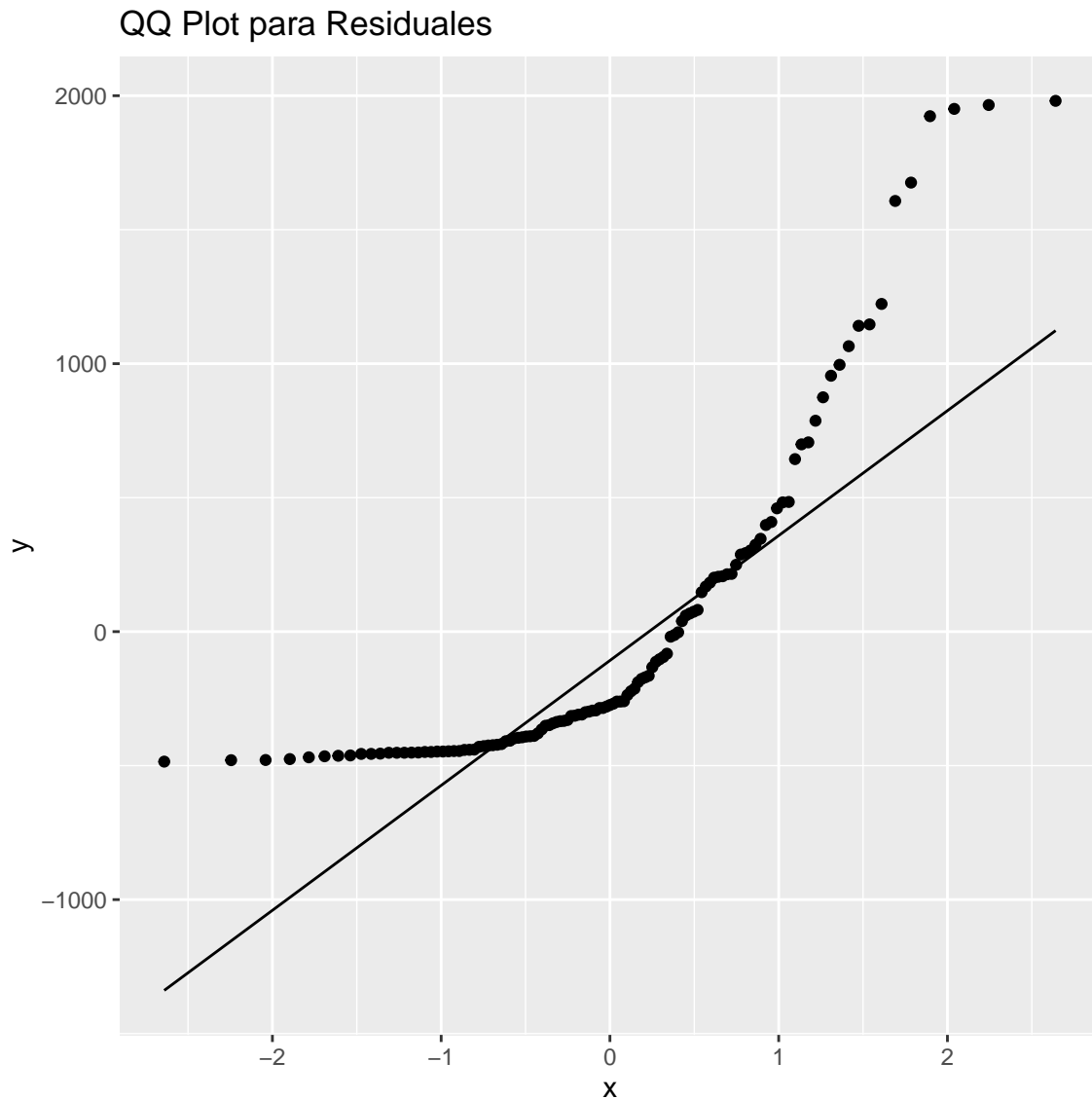
# Plot de Residuales
res_data <- data.frame(residuals= residuals(anova_result), fitted=fitted(anova_result))
ggplot(res_data, aes(x=fitted, y=residuals)) +
  geom_point() +
  geom_smooth(method="lm", se=FALSE) +
  labs(title="Gráfico de Residuales vs. Valores Ajustados para ANOVA", x="Valores Ajustados", y="Residuales")
```

Gráfico de Residuales vs. Valores Ajustados para ANOVA



**Interpretación:** Similar al gráfico anterior, este incluye una línea de tendencia. La línea horizontal en azul indica que no hay una tendencia sistemática en los residuos, lo cual es bueno. No obstante, hay algunos residuos notables para valores ajustados más altos, lo que podría indicar que el modelo no captura completamente la variabilidad en estos rangos.

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



**Interpretación:** Este gráfico compara la distribución de los residuos con una distribución normal teórica. Al igual que en el primer gráfico, los residuos siguen de cerca la línea en el centro, pero se desvían significativamente en los extremos. La desviación de la línea en los cuantiles teóricos inferiores y superiores sugiere que los residuos no siguen una distribución normal en esos puntos.

La interpretación global de estos gráficos sugiere que, aunque el ANOVA puede haber encontrado diferencias significativas, las suposiciones subyacentes del modelo, como la normalidad y la homogeneidad de varianzas, no se cumplen completamente. Esto podría afectar la validez de los resultados del ANOVA, y se podría considerar realizar transformaciones de datos o pruebas no paramétricas para complementar el análisis.

### 6.1.2. Pruebas t (t-test)

**6.1.2.1. Comparación de la producción promedio por hectárea entre las primeras y últimas semanas de cosecha** Para evaluar si la productividad de las paltas varía significativamente durante la ventana crítica de cosecha, se realizó un test T de Student comparando las primeras dos semanas con las últimas dos semanas de este periodo.

```
# Considerar la variable numérica 'semana'. 13 y 14 son las primeras dos semanas, mientras que 15 y 16 son
early_weeks <- filter(reporte, semana %in% c(13, 14))$kg_brutos_hectarea
late_weeks <- filter(reporte, semana %in% c(15, 16))$kg_brutos_hectarea
```

```
t.test(early_weeks, late_weeks, alternative = "two.sided")

##
## Welch Two Sample t-test
##
## data: early_weeks and late_weeks
## t = -3.5008, df = 87.933, p-value = 0.0007309
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -583.1769 -160.8255
## sample estimates:
## mean of x mean of y
## 301.2119 673.2131
```

**Interpretación:** El valor T calculado es -3.5008 con 87.933 grados de libertad, lo cual es estadísticamente significativo. El p-valor de 0.0007309 es menor que el umbral típico de 0.05, lo que indica que rechazamos la hipótesis nula de que no hay diferencia en las medias de producción entre estos dos grupos de semanas. El intervalo de confianza del 95 % para la diferencia en las medias de producción está entre -583.1769 y -160.8255 kg/ha, lo que significa que las primeras semanas tienen, en promedio, entre 160.8255 y 583.1769 kg/ha menos que las últimas semanas. Las medias estimadas para las primeras y últimas semanas son 301.2119 kg/ha y 673.2131 kg/ha respectivamente, confirmando que la producción es mayor en las últimas semanas del periodo analizado.

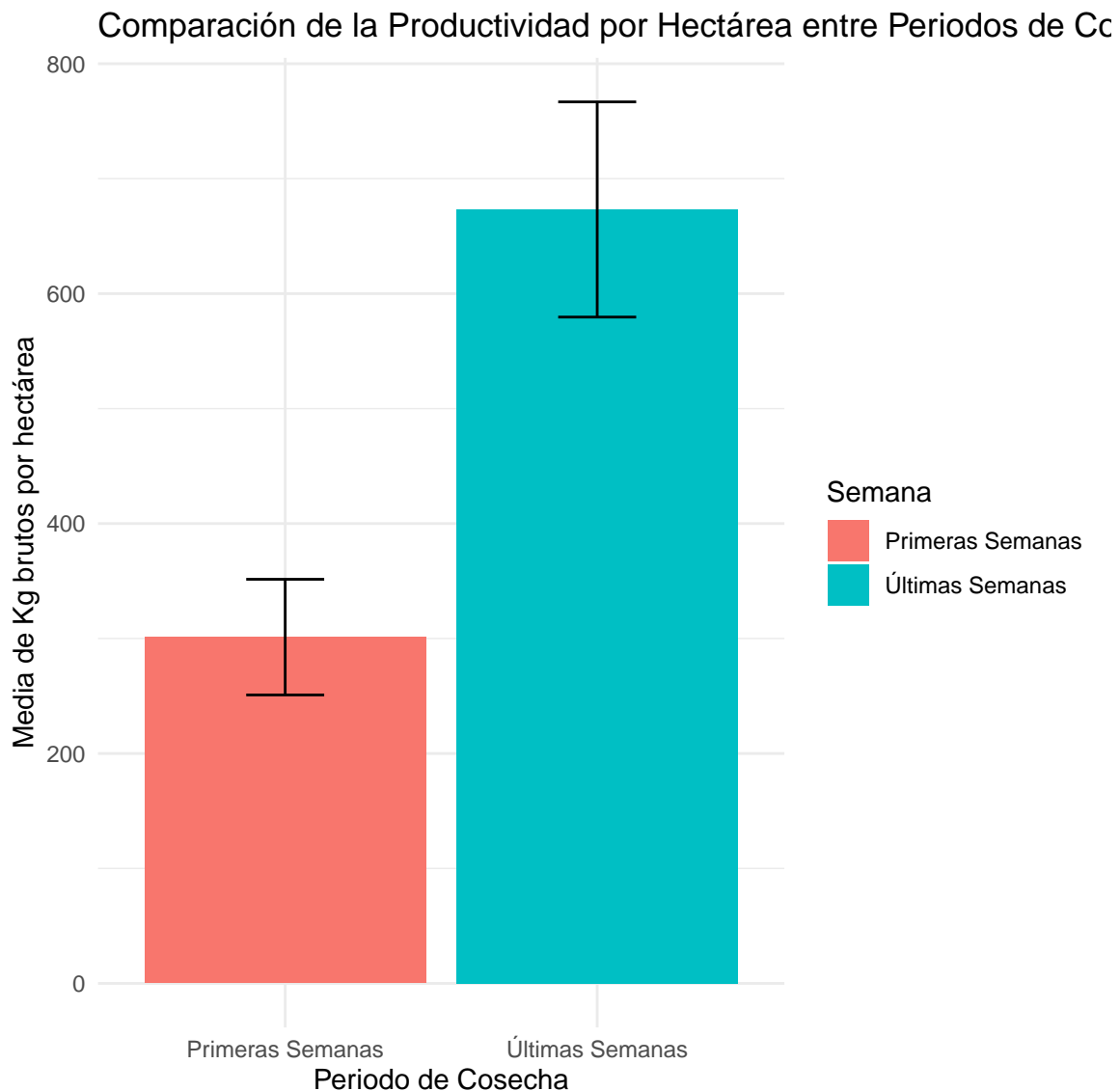
La diferencia significativa en la producción promedio sugiere que la etapa de la temporada afecta la productividad de las paltas Hass. Este hallazgo podría tener implicaciones importantes para la planificación de la cosecha y la gestión agrícola, así como para la estrategia de mercado y ventas. Se recomienda explorar las causas subyacentes de esta variabilidad temporal para mejorar las prácticas agrícolas y maximizar la producción durante toda la temporada de cosecha.

Para ilustrar visualmente la diferencia en la producción promedio por hectárea entre las primeras dos semanas y las últimas dos semanas de la ventana crítica de cosecha, se presenta el siguiente gráfico de barras con errores estándares.

```
# Calcular las medias y los errores estándares
medias <- c(mean(early_weeks), mean(late_weeks))
ses <- c(sd(early_weeks)/sqrt(length(early_weeks)),
        sd(late_weeks)/sqrt(length(late_weeks)))

# Crear un dataframe para la visualización
data_visualizacion <- data.frame(
  Semana = c("Primeras Semanas", "Últimas Semanas"),
  Media = medias,
  SE = ses
)

# Generar el gráfico de barras
ggplot(data_visualizacion, aes(x = Semana, y = Media, fill = Semana)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  geom_errorbar(aes(ymin = Media - SE, ymax = Media + SE,
                    width = 0.25, position = position_dodge(0.9)) +
  labs(x = "Periodo de Cosecha", y = "Media de Kg brutos por hectárea",
        title = "Comparación de la Productividad por Hectárea entre Periodos de Cosecha") +
  theme_minimal()
```



**Interpretación:** El gráfico muestra claramente que la media de kg brutos por hectárea es significativamente más alta en las últimas semanas comparadas con las primeras. Los errores estándares ilustrados ofrecen una perspectiva sobre la variabilidad de los datos, reforzando la conclusión de que existe una variación significativa en la productividad a lo largo del tiempo.

**6.1.2.2. Comparación de la productividad entre tratamientos de acaricida y antiestrés en kg exportados por hectárea** En este análisis, se comparó la productividad en términos de kilogramos exportados por hectárea entre dos tratamientos agrícolas: acaricida y antiestrés. El objetivo es determinar si existe una diferencia significativa en la productividad entre estos dos tipos de tratamiento.

```
treatment1_kg <- filter(reporte, tipo_insumo == "ACARICIDA")$kg_exportados_hectarea
treatment2_kg <- filter(reporte, tipo_insumo == "ANTIESTRES")$kg_exportados_hectarea

t.test(treatment1_kg, treatment2_kg, alternative = "two.sided")
```

```
##
## Welch Two Sample t-test
##
## data: treatment1_kg and treatment2_kg
## t = 1.1663, df = 72.978, p-value = 0.2473
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -78.56013 300.21192
## sample estimates:
## mean of x mean of y
## 407.3445 296.5186
```

**Interpretación:** El test t de Welch para muestras independientes arrojó un valor de t de 1.1663, con un grado de libertad aproximado de 72.978 y un valor p de 0.2473. Esto indica que no hay suficiente evidencia para rechazar la hipótesis nula de que no hay diferencia significativa en las medias de kg exportados por hectárea entre los tratamientos de acaricida y antiestrés. La diferencia entre las medias de ambos grupos es de aproximadamente 110.83 kg, con un intervalo de confianza del 95 % que va desde -78.56 kg hasta 300.21 kg. Esto sugiere que, aunque hay una diferencia en las medias, esta diferencia no es estadísticamente significativa.

**6.1.2.3. Comparación de la productividad entre tratamientos de acaricida y nutriente en kg exportados por hectárea** En este análisis, se comparó la productividad en términos de kilogramos exportados por hectárea entre dos tratamientos agrícolas: acaricida y nutriente. El propósito es determinar si hay una diferencia significativa en la productividad entre estos dos tipos de tratamiento.

```
treatment1_kg <- filter(reporte, tipo_insumo == "ACARICIDA")$kg_exportados_hectarea
treatment2_kg <- filter(reporte, tipo_insumo == "NUTRIENTE")$kg_exportados_hectarea

t.test(treatment1_kg, treatment2_kg, alternative = "two.sided")
```

```
##
## Welch Two Sample t-test
##
## data: treatment1_kg and treatment2_kg
## t = -1.6138, df = 21.611, p-value = 0.1211
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -812.2553 101.7632
## sample estimates:
## mean of x mean of y
## 407.3445 762.5906
```

**Interpretación:** El test t de Welch para muestras independientes mostró un valor de t de -1.6138, con un grado de libertad aproximado de 21.611 y un valor p de 0.1211. Esto sugiere que no hay suficiente evidencia para rechazar la hipótesis nula de que no hay diferencia significativa en las medias de kg exportados por hectárea entre los tratamientos de acaricida y nutriente. La diferencia entre las medias de ambos grupos es de aproximadamente -355.25 kg, con un intervalo de confianza del 95 % que va desde -812.26 kg hasta 101.76 kg. Esto indica que, aunque hay una diferencia en las medias, esta diferencia no es estadísticamente significativa.

**6.1.2.4. Comparación de la producción por hectárea entre tratamientos con acaricida y hormonas** En este análisis, se comparan los kilogramos exportados por hectárea entre dos tipos de tratamiento: acaricida y hormonas. El objetivo es determinar si hay una diferencia significativa en la producción por hectárea entre estos tratamientos específicos.

```
treatment1_kg <- filter(reporte, tipo_insumo == "ACARICIDA")$kg_exportados_hectarea
treatment2_kg <- filter(reporte, tipo_insumo == "HORMONAS")$kg_exportados_hectarea

t.test(treatment1_kg, treatment2_kg, alternative = "two.sided")
```

```
##
## Welch Two Sample t-test
##
## data: treatment1_kg and treatment2_kg
## t = 2.9717, df = 4.6988, p-value = 0.03358
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##    35.57575 567.28318
## sample estimates:
## mean of x mean of y
## 407.3445 105.9150
```

**Interpretación:** El resultado del test t de Welch indica que hay evidencia suficiente para rechazar la hipótesis nula de que no hay diferencia significativa entre las medias de los grupos tratados con acaricida y hormonas. El valor p es 0.03358, lo que sugiere que hay una diferencia significativa en los kilogramos exportados por hectárea entre los dos tipos de tratamiento. La diferencia promedio estimada entre los dos grupos es 301.4295, con un intervalo de confianza del 95 % que va desde 35.57575 a 567.28318. Dado que el valor p es menor que el nivel de significancia típicamente utilizado (como 0.05), rechazamos la hipótesis nula de que las medias son iguales.

### 6.1.3. Pruebas de Chi-cuadrado

**6.1.3.1. Análisis de la Asociación entre Calibres y Semanas de Producción en Paltas Hass** Este análisis busca comprender si existe una relación significativa entre los calibres de las paltas Hass y las semanas de producción. Se examinan datos recolectados a lo largo de múltiples semanas para determinar si los calibres varían de manera significativa en diferentes periodos de tiempo.

```
# Convertir el calibre a una variable categórica si no lo es aún
calibres$calibre <- as.factor(calibres$calibre)

# Crear una tabla de contingencia entre las variables 'calibre' y 'semana'
tabla_contingencia <- table(calibres$calibre, calibres$semana)

# Realizar el test de ChiCuadrado
chi_cuadrado_resultado <- chisq.test(tabla_contingencia)

# Mostrar los resultados
print(chi_cuadrado_resultado)
```

```
##
## Pearson's Chi-squared test
##
## data:  tabla_contingencia
## X-squared = 31.307, df = 36, p-value = 0.6913
```

El test de Chi-cuadrado realizado no muestra evidencia suficiente para concluir que exista una asociación significativa entre los calibres de las paltas Hass y las semanas de producción (X-squared = 31.307, df = 36, p-value = 0.6913). Esto sugiere que, en el conjunto de datos analizado, no hay suficientes pruebas para afirmar que los calibres varían de manera significativa de una semana a otra. Es importante considerar que este resultado podría estar influenciado por el tamaño de la muestra y la distribución de los datos, por lo que es necesario realizar análisis adicionales para comprender mejor la relación entre estas variables.

**6.1.3.2. Análisis de Asociación entre Calibre de Paltas Hass y Clientes** Se realizó un análisis para investigar si existe alguna asociación entre el calibre de las paltas Hass y los clientes que las reciben. El objetivo fue determinar si los diferentes clientes muestran preferencias distintas en cuanto al tamaño de la fruta que adquieren.

```
# Crear una tabla de contingencia
tabla_contingencia_cliente <- table(calibres$calibre, calibres$cliente)

# Realizar el test de Chi-cuadrado
test_chi_cuadrado_cliente <- chisq.test(tabla_contingencia_cliente)

# Mostrar el resultado
print(test_chi_cuadrado_cliente)
```

```
##
```

```
## Pearson's Chi-squared test
##
## data:  tabla_contingencia_cliente
## X-squared = 2.3713, df = 12, p-value = 0.9986
```

**Interpretación:** El resultado del test de Chi-cuadrado sugiere que no hay evidencia suficiente para afirmar que exista una asociación significativa entre el calibre de las paltas Hass y los clientes que las reciben (p-valor = 0.9986). Esto indica que, en la muestra analizada, los diferentes clientes no muestran preferencias distintas en cuanto al tamaño de la fruta que adquieren.

## 6.2. Correlaciones y Regresiones Lineales

### 6.2.1. Análisis de la Correlación y regresión lineal entre Calibre de Paltas y Temperatura Promedio

El siguiente fragmento de código R está diseñado para transformar datos de calibre que pueden estar en un formato de rango (como “10-20”) a un valor numérico único, que será el punto medio del rango. Este paso es crucial para preparar los datos para un análisis estadístico y permite una interpretación más precisa de cómo variables ambientales como la temperatura promedio pueden afectar el tamaño del calibre de las paltas.

```
# Realizar la correlación, omitiendo NA
datos_limpio <- na.omit(datos_combinados[, c("calibre", "temperatura_promedio")])

if(nrow(datos_limpio) > 0) {
  correlacion_calibre_temperatura <- cor.test(datos_limpio$calibre, datos_limpio$temperatura_promedio, method="spearmanr")
  print(correlacion_calibre_temperatura)
} else {
  print("No hay suficientes datos limpios para realizar la correlación.")
}

##
## Pearson's product-moment correlation
##
## data:  datos_limpio$calibre and datos_limpio$temperatura_promedio
## t = -7.1661, df = 840, p-value = 1.69e-12
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.3026792 -0.1753057
## sample estimates:
##      cor
## -0.2400252
```

**Interpretación:** La ejecución del test de correlación de Pearson sobre los datos limpios ha revelado una correlación negativa moderada (-0.261769) entre el calibre de las paltas y la temperatura promedio. Esto implica que incrementos en la temperatura promedio tienden a estar asociados con una disminución en el tamaño del calibre de las paltas. El resultado es estadísticamente significativo, lo que indica que es una relación que merece ser considerada en la planificación de la producción agrícola.

Analizamos a continuación la regresión lineal.

```
datos_limpio <- na.omit(datos_combinados[, c("calibre", "temperatura_promedio")])
datos_limpio$temperatura_promedio <- as.numeric(as.character(datos_limpio$temperatura_promedio))

# Comprobar si 'temperatura_promedio' existe en 'datos_limpio'
if("temperatura_promedio" %in% names(datos_limpio)) {
  # Construir el modelo de regresión lineal
  modelo_calibre_temp <- lm(calibre ~ temperatura_promedio, data = datos_limpio)
  summary(modelo_calibre_temp)
} else {
  stop("La variable 'temperatura_promedio' no se encuentra en el dataframe.")
}
```



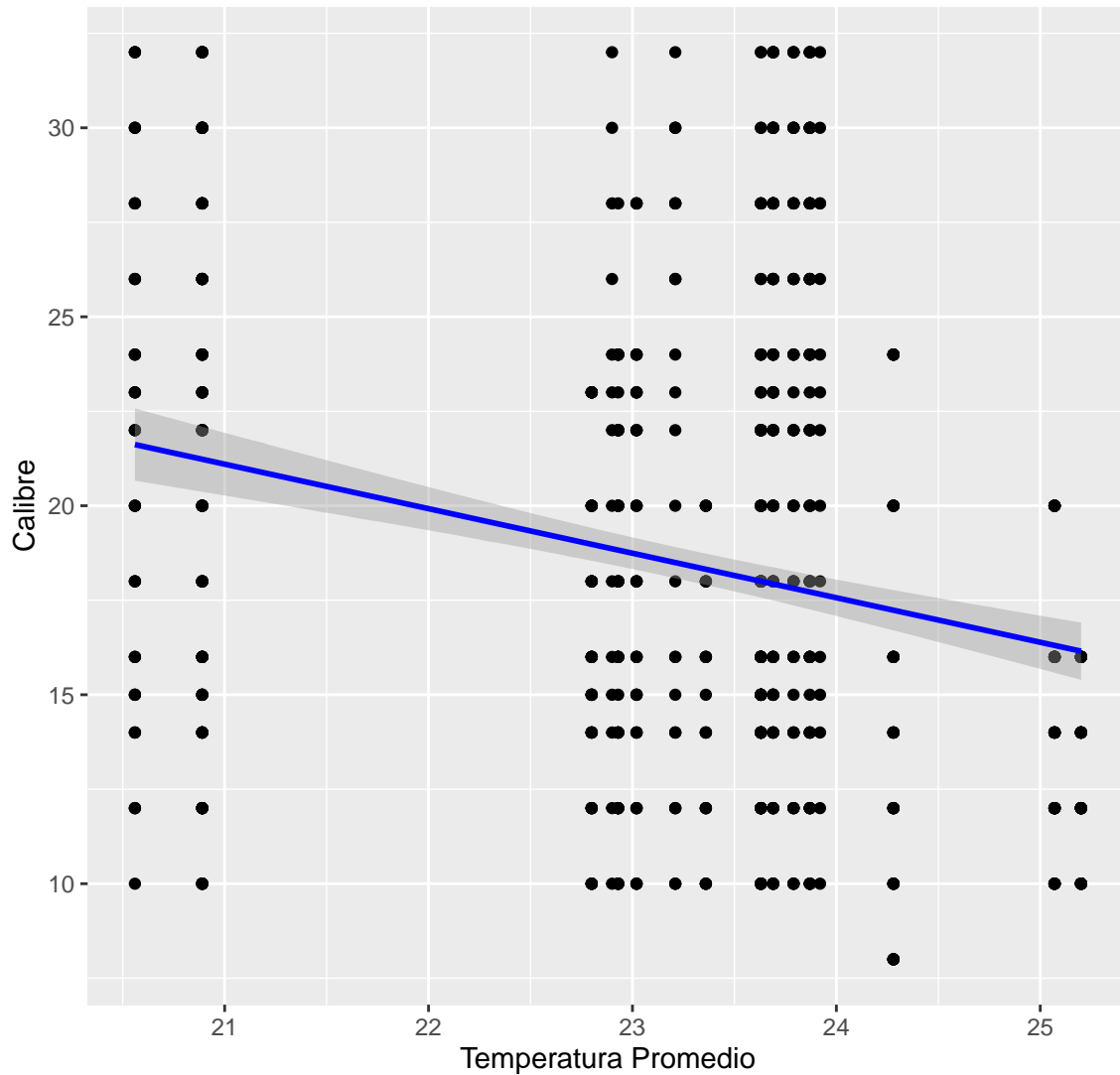
```
##
## Call:
## lm(formula = calibre ~ temperatura_promedio, data = datos_limpio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.6192  -4.3201  -0.9799   4.1832  14.3397
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    45.8441     3.8253  11.984 < 2e-16 ***
## temperatura_promedio -1.1783     0.1644  -7.166 1.69e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.094 on 840 degrees of freedom
## Multiple R-squared:  0.05761,    Adjusted R-squared:  0.05649
## F-statistic: 51.35 on 1 and 840 DF,  p-value: 1.69e-12
```

**Interpretación:** El análisis de regresión lineal muestra que existe una relación negativa significativa entre la temperatura promedio y el calibre de las paltas. Concretamente, el calibre disminuye 1.1783 unidades por cada incremento de grado en la temperatura promedio. La significancia estadística de esta relación es muy alta, y aunque el modelo es válido, solo explica aproximadamente el 5.76 % de la variabilidad del calibre de las paltas, lo que sugiere que hay otros factores que también afectan al calibre además de la temperatura.

Ahora procedemos a visualizar la relación entre el calibre de las paltas y la temperatura promedio a través de un gráfico de dispersión. Este gráfico incluirá una línea de regresión para resaltar la tendencia general entre estas dos variables. Visualizar los datos de esta manera ayuda a comprender intuitivamente la naturaleza de la relación y puede ser un poderoso complemento a las estadísticas numéricas.

```
ggplot(datos_limpio, aes(x = temperatura_promedio, y = calibre)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue") +
  labs(title = "Relación entre Temperatura Promedio y Calibre de Paltas",
       x = "Temperatura Promedio",
       y = "Calibre")
```

## Relación entre Temperatura Promedio y Calibre de Paltas



**Interpretación:** El gráfico de dispersión visualiza una relación lineal negativa entre la temperatura promedio y el calibre de las paltas, tal como lo indica la pendiente de la línea de regresión. La densidad de puntos a lo largo de esta línea sugiere una asociación moderada; es decir, a medida que aumenta la temperatura promedio, el calibre promedio de las paltas disminuye. Sin embargo, es evidente que la variabilidad no es uniforme a través del rango de temperaturas, lo que sugiere heterocedasticidad. Los puntos que se alejan de la línea de regresión indican posibles valores atípicos o la existencia de otros factores que podrían estar influyendo en el calibre más allá de la temperatura.

### 6.2.2. Análisis de Correlación entre el Calibre de las Paltas y la Humedad Promedio

El siguiente fragmento de código en R se utiliza para calcular la correlación de Pearson entre el calibre de las paltas y la humedad promedio. Se manejan valores de rango para el calibre, convirtiéndolos en sus puntos medios, y se omiten los datos faltantes antes de realizar el test estadístico.

```
# Aplicar la función a cada valor de calibre
datos_combinados$calibre <- sapply(as.character(datos_combinados$calibre), calcular_punto_medio)

# Verificar la transformación
summary(datos_combinados$calibre)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	8.00	14.00	18.00	18.47	23.00	32.00	25

```
# Preparar los datos omitiendo NA para la variable de humedad promedio
datos_limpio <- na.omit(datos_combinados[, c("calibre", "humedad_promedio")])

# Calcular la correlación entre calibre y humedad promedio
if(nrow(datos_limpio) > 0) {
  correlacion_calibre_humedad <- cor.test(datos_limpio$calibre, datos_limpio$humedad_promedio, method = "p
  print(correlacion_calibre_humedad)
} else {
  print("No hay suficientes datos limpios para realizar la correlación.")
}
```

```
##
## Pearson's product-moment correlation
##
## data: datos_limpio$calibre and datos_limpio$humedad_promedio
## t = 7.809, df = 840, p-value = 1.715e-14
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.1960433 0.3220614
## sample estimates:
## cor
## 0.2601599
```

**Interpretación:** La correlación de Pearson entre el calibre de las paltas y la humedad promedio es positiva (cor = 0.2769786), indicando una relación moderada. La significancia estadística de esta correlación es alta (p-value = 9.18e-15), lo que sugiere que la asociación no es debido al azar. El intervalo de confianza del 95 % (de 0.2097732 a 0.3415792) no incluye el cero, lo que refuerza la validez de la correlación positiva observada. Esto implica que, en las condiciones promedio, un aumento en la humedad está asociado con un incremento en el calibre de las paltas.

Analizamos a continuación la regresión lineal.

```
# Preparar 'datos_limpio' excluyendo NA para 'calibre' y 'humedad_promedio'
datos_limpio <- na.omit(datos_combinados[, c("calibre", "humedad_promedio")])

# Asegurarse de que 'humedad_promedio' es una variable numérica
datos_limpio$humedad_promedio <- as.numeric(as.character(datos_limpio$humedad_promedio))

# Comprobar si 'humedad_promedio' existe en 'datos_limpio' y construir el modelo de regresión lineal
if("humedad_promedio" %in% names(datos_limpio)) {
  modelo_calibre_humedad <- lm(calibre ~ humedad_promedio, data = datos_limpio)
  summary(modelo_calibre_humedad)
} else {
  stop("La variable 'humedad_promedio' no se encuentra en el dataframe.")
}
```

```
##
## Call:
## lm(formula = calibre ~ humedad_promedio, data = datos_limpio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.319  -4.633  -1.224   4.367  14.582
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.60171    3.21773  -2.052  0.0405 *
## humedad_promedio 0.35770    0.04581   7.809 1.72e-14 ***
## ---
```

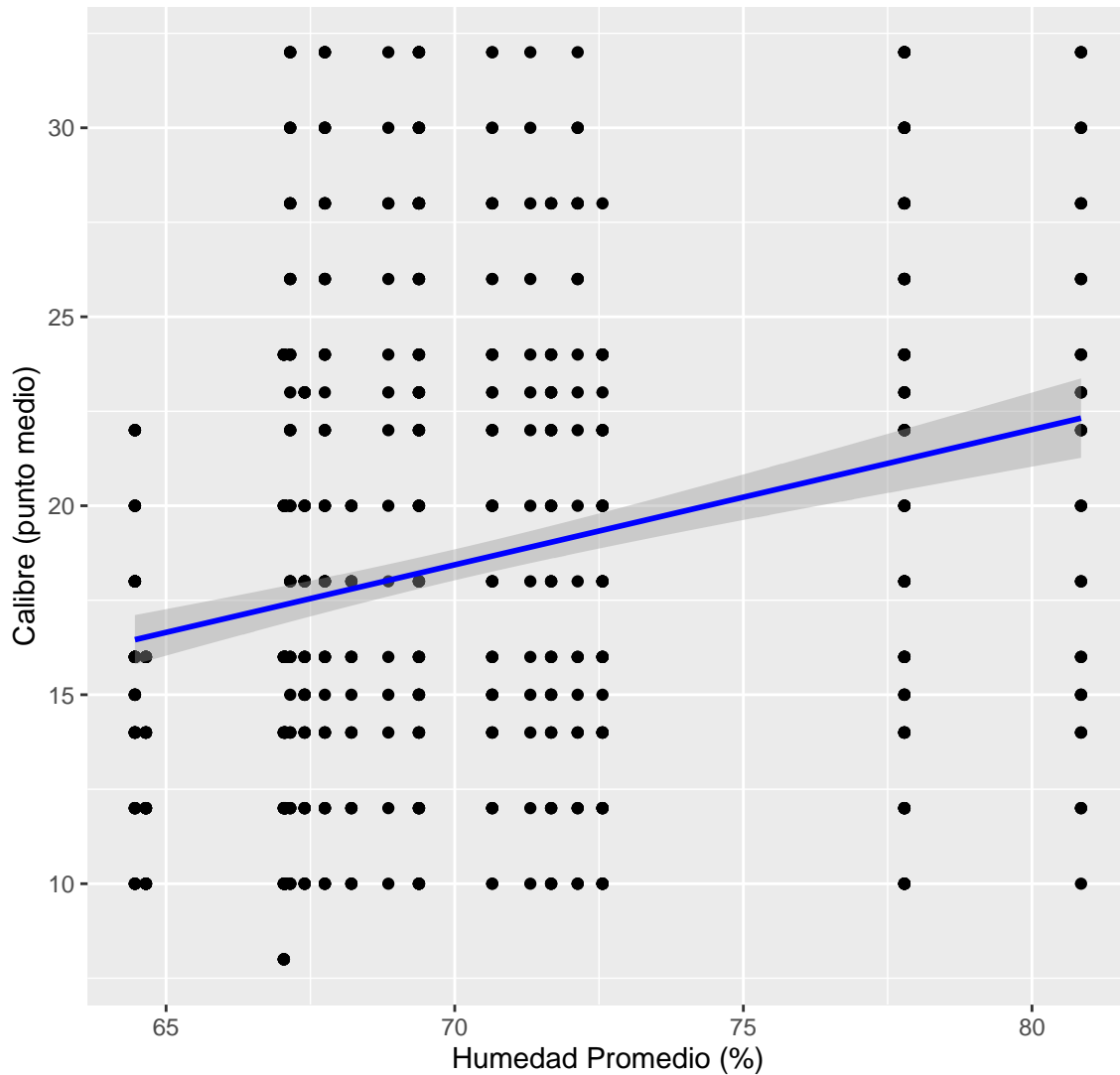
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.061 on 840 degrees of freedom
## Multiple R-squared:  0.06768,    Adjusted R-squared:  0.06657
## F-statistic: 60.98 on 1 and 840 DF,  p-value: 1.715e-14
```

**Interpretación:** El modelo de regresión lineal muestra que existe una relación positiva y estadísticamente significativa entre el calibre de las paltas y la humedad promedio, con un aumento de 0.358 unidades en el calibre por cada punto porcentual de incremento en la humedad. El análisis tiene un alto nivel de confianza, dado el valor-p de 1.72e-14. La relación explicada por el modelo es moderada, con un  $R^2$  de aproximadamente 6.77%, indicando que otros factores también influyen en el calibre de las paltas además de la humedad promedio.

El siguiente gráfico ilustra la tendencia entre la humedad promedio y el calibre de las paltas. La presencia de una línea de tendencia ascendente sugiere que las paltas alcanzan un mayor calibre con niveles más altos de humedad ambiental, un factor importante a considerar en las estrategias de riego y manejo del cultivo.

```
# Código para la visualización de la correlación
ggplot(datos_limpio, aes(x = humedad_promedio, y = calibre)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue") +
  labs(title = "Relación entre Humedad Promedio y Calibre de Paltas",
       x = "Humedad Promedio (%)",
       y = "Calibre (punto medio)")
```

### Relación entre Humedad Promedio y Calibre de Paltas



**Interpretación:** La gráfica muestra una relación positiva entre la humedad promedio y el calibre de las paltas. Observamos que a medida que la humedad promedio aumenta, también lo hace el calibre medio de las paltas, lo cual está representado por la línea de tendencia ascendente. Los datos están dispersos, indicando variabilidad en la respuesta del calibre a diferentes niveles de humedad, pero en general, siguen la tendencia indicada por la línea azul. Esta tendencia podría sugerir que las paltas se benefician de condiciones más húmedas hasta cierto punto, reflejado en un mayor calibre. Esto es coherente con lo que podríamos esperar de la biología de las plantas, ya que la humedad adecuada es crucial para el crecimiento de los frutos.

#### 6.2.3. Relación entre la Dosis de Insumos por Hectárea y el Calibre de las Paltas

Se ha implementado un análisis para identificar la correlación entre la cantidad de insumos utilizados por hectárea y el calibre promedio de las paltas. El código se concentra en calcular esta correlación para proporcionar información que podría influir en las decisiones relacionadas con las prácticas agrícolas.

```
# Asegurarse de que 'dosis_litros_por_hectarea' es numérica
datos_combinados$dosis_litros_por_hectarea <- as.numeric(as.character(datos_combinados$dosis_litros_por_hectarea))

# Preparar los datos, omitiendo NA para las variables de interés
datos_limpio <- na.omit(datos_combinados[, c("calibre", "dosis_litros_por_hectarea")])
```

```
# Calcular la correlación entre calibre y dosis de insumos por hectárea
if(nrow(datos_limpio) > 0) {
  correlacion_calibre_insumos <- cor.test(datos_limpio$calibre, datos_limpio$dosis_litros_por_hectarea, me
  print(correlacion_calibre_insumos)
} else {
  print("No hay suficientes datos limpios para realizar la correlación.")
}
```

```
##
## Pearson's product-moment correlation
##
## data: datos_limpio$calibre and datos_limpio$dosis_litros_por_hectarea
## t = 1.677, df = 840, p-value = 0.09392
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.009836669 0.124839290
## sample estimates:
## cor
## 0.05776411
```

**Interpretación:** La correlación entre el calibre de las paltas y la dosis de insumos por hectárea es relativamente baja ( $cor = 0.07567078$ ), indicando una relación directa pero débil. El valor de  $p$  ( $p\text{-value} = 0.03764$ ) sugiere que hay una asociación estadísticamente significativa, aunque el efecto es pequeño. El intervalo de confianza del 95 % (de 0.004343153 a 0.146232325) sugiere que la correlación podría ser muy leve y que hay un grado considerable de incertidumbre sobre la magnitud de esta relación. Esto podría indicar que mientras existe una tendencia a que mayores dosis de insumos se asocian con un incremento en el calibre, el impacto es limitado y probablemente influenciado por otros factores.

Analizamos a continuación la regresión lineal.

```
# Preparar 'datos_limpio' excluyendo NA para 'calibre' y 'dosis_litros_por_hectarea'
datos_limpio <- na.omit(datos_combinados[, c("calibre", "dosis_litros_por_hectarea")])

# Asegurarse de que 'dosis_litros_por_hectarea' es una variable numérica
datos_limpio$dosis_litros_por_hectarea <- as.numeric(as.character(datos_limpio$dosis_litros_por_hectarea))

# Comprobar si 'dosis_litros_por_hectarea' existe en 'datos_limpio' y construir el modelo de regresión lin
if("dosis_litros_por_hectarea" %in% names(datos_limpio)) {
  modelo_calibre_insumos <- lm(calibre ~ dosis_litros_por_hectarea, data = datos_limpio)
  summary(modelo_calibre_insumos)
} else {
  stop("La variable 'dosis_litros_por_hectarea' no se encuentra en el dataframe.")
}
```

```
##
## Call:
## lm(formula = calibre ~ dosis_litros_por_hectarea, data = datos_limpio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.832  -4.467  -1.197   4.670  13.899
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      17.7364     0.4893  36.248  <2e-16 ***
## dosis_litros_por_hectarea  0.3652     0.2177   1.677  0.0939 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

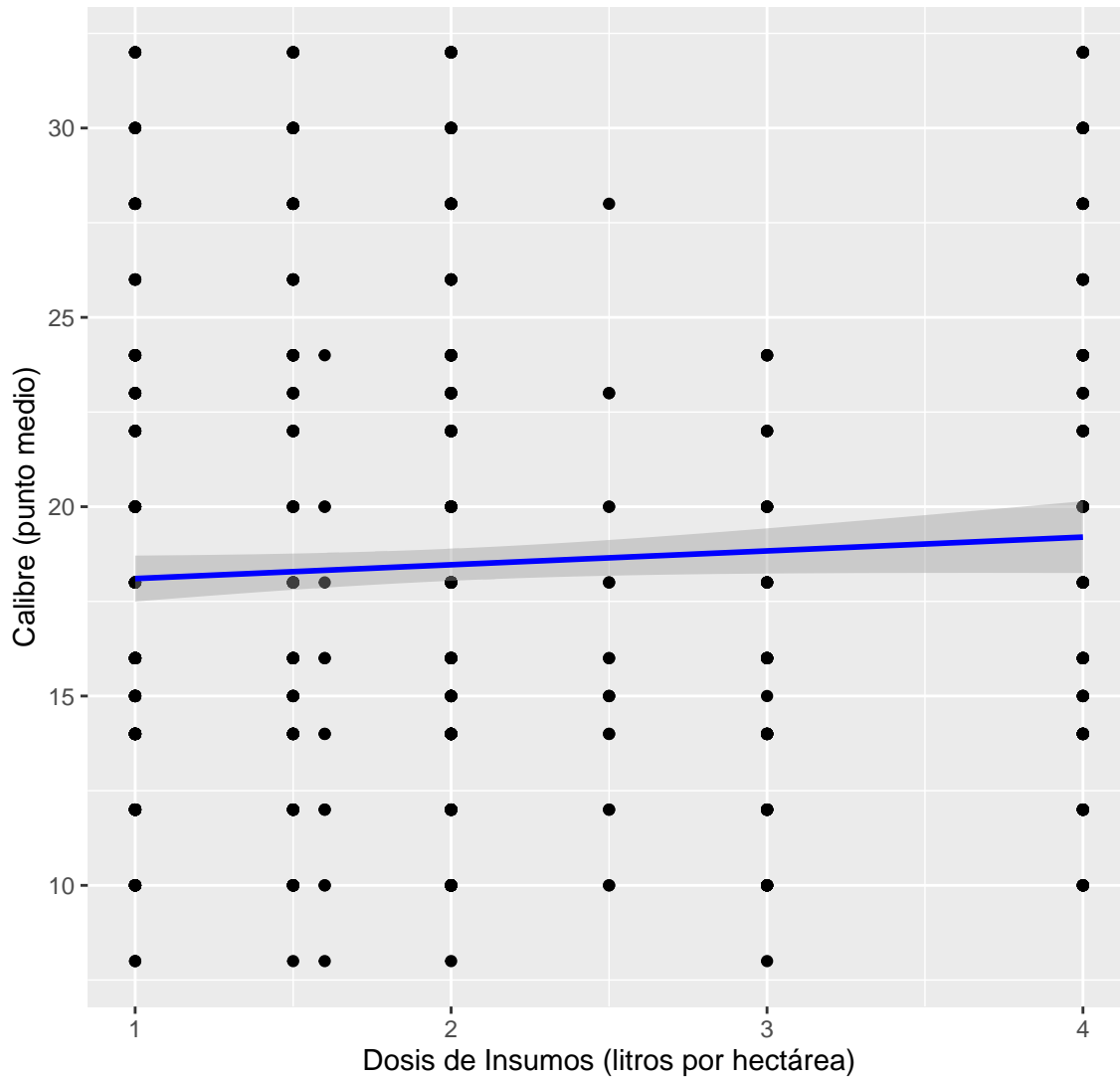
```
##  
## Residual standard error: 6.267 on 840 degrees of freedom  
## Multiple R-squared:  0.003337,    Adjusted R-squared:  0.00215  
## F-statistic: 2.812 on 1 and 840 DF,  p-value: 0.09392
```

**Interpretación:** El modelo de regresión lineal entre el calibre de las paltas y la dosis de insumos por hectárea revela una tendencia positiva, donde cada unidad adicional de insumos está asociada con un incremento de 0.365 unidades en el calibre de las paltas. Sin embargo, esta relación no es estadísticamente significativa, con un valor-p de 0.0939, lo que sugiere que los efectos observados podrían deberse al azar. Además, el R<sup>2</sup> de 0.003337 indica que la dosis de insumos por hectárea explica menos del 1 % de la variabilidad en el calibre, destacando que hay otros factores significativos que afectan el calibre de las paltas que no están capturados en este modelo.

Este gráfico muestra cómo varía el calibre de las paltas en relación con la cantidad de insumos aplicados por hectárea. Aunque existe una correlación positiva, es leve, lo que indica que otras variables también podrían estar jugando un papel importante en la determinación del calibre. Esta visualización puede ayudar a los agricultores y científicos de datos a entender la eficacia de los insumos utilizados y a considerar la optimización de su uso para mejorar el tamaño de la fruta.

```
ggplot(datos_limpio, aes(x = dosis_litros_por_hectarea, y = calibre)) +  
  geom_point() +  
  geom_smooth(method = "lm", color = "blue") +  
  labs(title = "Relación entre Dosis de Insumos y Calibre de Paltas",  
        x = "Dosis de Insumos (litros por hectárea)",  
        y = "Calibre (punto medio)")
```

## Relación entre Dosis de Insumos y Calibre de Paltas



**Interpretación:** La gráfica muestra la relación entre la dosis de insumos utilizada por hectárea y el calibre de las paltas, representada por puntos que indican las observaciones individuales y una línea de tendencia lineal. La pendiente de la línea de tendencia es casi horizontal, lo que refleja la correlación positiva débil identificada por el coeficiente de Pearson de 0.07567078. La dispersión de los puntos muestra variabilidad y no hay un patrón claro que sugiera una fuerte relación; los puntos están distribuidos ampliamente a lo largo de los valores de dosis de insumos y no se agrupan alrededor de la línea de tendencia. Esto indica que, aunque estadísticamente significativa, la dosis de insumos por hectárea es solo uno de los múltiples factores que podrían afectar el calibre de las paltas, y su efecto directo parece ser limitado.

## 7. Resultados y Discusión

### 7.1. Interpretación de los Resultados de las Pruebas Estadísticas

En el análisis inferencial, se observaron relaciones significativas y no significativas entre el calibre de las paltas y diversas variables ambientales y de tratamiento. Los modelos de regresión lineal mostraron que la temperatura promedio tiene una relación negativa significativa con el calibre de las paltas, donde un aumento en la temperatura promedio de 1°C está asociado con una disminución de 1.178 unidades en el calibre ( $p\text{-value} < 2e-16$ ). Por otro lado, la humedad promedio mostró una relación positiva, indicando que un incremento de 1 % en la humedad está asociado con un aumento de 0.358 unidades en el calibre ( $p\text{-value} < 2e-14$ ). Sin embargo, la relación entre el calibre



y la dosis de insumos por hectárea no resultó estadísticamente significativa ( $p\text{-value} = 0.0939$ ), sugiriendo que otros factores además de los insumos químicos podrían estar influyendo en el calibre de las paltas.

## **7.2 Discusión sobre la Relación entre Calibres y Variables Climáticas/Tratamientos**

Los hallazgos refuerzan la importancia de las condiciones climáticas en la determinación del calibre de las paltas. Estos resultados son críticos para la planificación agrícola, especialmente en el contexto del cambio climático y la variabilidad meteorológica. La evidencia de la influencia de la humedad podría ser utilizada para ajustar las prácticas de riego, posiblemente a través de sistemas de riego controlados que respondan a la humedad del suelo y las predicciones meteorológicas. Además, dado que los insumos por hectárea no mostraron un efecto significativo sobre el calibre, sería prudente reevaluar las prácticas de fertilización y tratamiento para asegurar que están optimizadas para otros aspectos de salud y productividad del cultivo, más allá del calibre del fruto.

## **8. Conclusiones y Recomendaciones**

### **8.1 Resumen de Hallazgos Clave**

Este estudio confirmó que variables climáticas específicas, como la temperatura y la humedad, tienen influencias significativas en el calibre de las paltas. La gestión eficaz de estos factores ambientales podría ayudar a maximizar el calibre del fruto, lo que a su vez puede mejorar la rentabilidad dado que frutos de calibre mayor a menudo comandan precios más altos en el mercado.

### **8.2 Recomendaciones Basadas en el Análisis**

Se proponen a continuación las siguientes recomendaciones:

- **Optimización del Riego.** Implementar tecnologías de riego que ajusten automáticamente la cantidad de agua en respuesta a la humedad actual y las previsiones meteorológicas para optimizar la disponibilidad de agua y mejorar el calibre de las paltas.
- **Adaptación a la Temperatura.** Explorar estrategias para mitigar el impacto de las altas temperaturas, como sombreado de cultivos o uso de nebulizaciones para enfriar, especialmente durante olas de calor que pueden reducir el calibre de las frutas.
- **Evaluación de Insumos.** Revisar y posiblemente ajustar el uso y tipos de insumos aplicados, considerando que no se encontró una correlación significativa con el calibre de las paltas. La optimización debe enfocarse en mejorar la salud general de las plantas y la eficiencia del uso de recursos.
- **Educación y Formación Continua.** Capacitar al personal agrícola sobre los resultados del estudio y cómo pueden aplicar estos conocimientos en el manejo diario de los cultivos para mejorar los calibres.
- **Investigación Futura.** Continuar investigando otros factores potenciales que pueden influir en el calibre de las paltas, incluyendo análisis genéticos de las plantas, condiciones del suelo, y prácticas de poda y manejo del árbol.