Reporte

Grupo No4

2024-04-20

Contents

Set working directory

```
# Set working directory
setwd("D:/1.Maestria Ciencia Datos/03. INTRODUCCIÓN A LOS MODELOS ESTADÍSTICOS-23MCDAP002-PSMA-00609-191
# Importing the dataset
data <- read_excel("FINAL TOTAL.xlsx", sheet = "Reporte")
# Drop the last row
data <- data[-nrow(data), ]

#Head
head(data)
```

```
## # A tibble: 6 x 38
##   Semana Mes   Fundo    Empresa Lote     Ha Variedad Color `N° Guia de Remisión`
##    <dbl> <chr> <chr>    <chr>   <chr> <dbl> <chr>    <chr> <chr>
## 1     13 Marzo Don Nico Agrico~ DNH-~  9.75 Hass     Negra T008 N° 0000109
## 2     13 Marzo Don Nico Agrico~ DNH-~ 10.9  Hass     Negra T008 N° 0000109
## 3     13 Marzo Don Nico Agrico~ DNH-~ 11.8  Hass     Negra T008 N° 0000109
## 4     13 Marzo Don Nico Agrico~ DNH-~ 11.6  Hass     Negra T008 N° 0000109
## 5     13 Marzo Don Nico Agrico~ DNH-~ 11.5  Hass     Negra T008 N° 0000109
## 6     13 Marzo Cuatro ~ Agrico~ CVH-~  7.21 Hass     Negra T008 N° 0000109
## # i 29 more variables: `N° Reporte de Producción` <chr>, Fecha <dttm>,
## #   `Total Jabas` <dbl>, `Peso Promedio Jaba (Kg)` <dbl>,
## #   `Cajas  Exportadas (10 Kg)` <dbl>, `Ingreso Packing (Kg Bruto)` <dbl>,
## #   `Kg Exportados` <dbl>, `% Exportado` <dbl>, `Kg Descarte` <dbl>,
## #   `% Descarte` <dbl>, `Kg Merma` <dbl>, `% Merma` <dbl>,
## #   `Kg Descarte de Campo` <dbl>, `Kg Brutos Lote` <dbl>, `Kg Brutos Ha` <dbl>,
## #   `Kg Exportado Ha` <dbl>, Status <chr>, CLIENTE <chr>, ...
```

```
names(data)
```

```
##  [1] "Semana"                    "Mes"
##  [3] "Fundo"                     "Empresa"
##  [5] "Lote"                      "Ha"
##  [7] "Variedad"                  "Color"
##  [9] "N° Guia de Remisión"       "N° Reporte de Producción"
## [11] "Fecha"                     "Total Jabas"
```

```
## [13] "Peso Promedio Jaba (Kg)"    "Cajas  Exportadas (10 Kg)"
## [15] "Ingreso Packing (Kg Bruto)" "Kg Exportados"
## [17] "% Exportado"                "Kg Descarte"
## [19] "% Descarte"                 "Kg Merma"
## [21] "% Merma"                    "Kg Descarte de Campo"
## [23] "Kg Brutos Lote"             "Kg Brutos Ha"
## [25] "Kg Exportado Ha"            "Status"
## [27] "CLIENTE"                    "TEMP PROM"
## [29] "TEMP. MAX"                  "TEMP. MIN"
## [31] "HUM. PROM"                  "HUM. MAX"
## [33] "HUM. MIN"                   "ET ACUMULADA"
## [35] "PRODUCTO"                   "NATURALEZA"
## [37] "DOSIS - L/HA"               "TIPO"
```

```r
# Rename specific columns using dplyr's rename() function
data <- data %>%
  rename(
    NoGR = `N° Guia de Remisión`,
    NoRepProd = `N° Reporte de Producción`,
    TotJabas = `Total Jabas`,
    PesoPromJkg = `Peso Promedio Jaba (Kg)`,
    CajasExp10kg = `Cajas  Exportadas (10 Kg)`,
    IngPakBrkg = `Ingreso Packing (Kg Bruto)`,
    KgExp = `Kg Exportados`,
    Perc_Exp = `% Exportado`,
    KgDesc = `Kg Descarte`,
    Perc_Desc = `% Descarte`,
    KgMerm = `Kg Merma`,
    Perc_Merm = `% Merma`,
    KgDescCamp = `Kg Descarte de Campo`,
    KgBruLt = `Kg Brutos Lote`,
    KgBruHa = `Kg Brutos Ha`,
    KgExpHa = `Kg Exportado Ha`,
    Cliente = `CLIENTE`,
    TempProm = `TEMP PROM`,
    TempMax = `TEMP. MAX`,
    TempMin = `TEMP. MIN`,
    HumProm = `HUM. PROM`,
    HumMax = `HUM. MAX`,
    HumMin = `HUM. MIN`,
    ETAcum = `ET ACUMULADA`,
    Producto = `PRODUCTO`,
    Naturaleza = `NATURALEZA`,
    DosisLtxHa = `DOSIS - L/HA`,
    Tipo = `TIPO`,
    # Continue renaming as needed
  )
#Names
colnames(data)
```

```
##  [1] "Semana"       "Mes"          "Fundo"        "Empresa"      "Lote"
##  [6] "Ha"           "Variedad"     "Color"        "NoGR"         "NoRepProd"
## [11] "Fecha"        "TotJabas"     "PesoPromJkg"  "CajasExp10kg" "IngPakBrkg"
## [16] "KgExp"        "Perc_Exp"     "KgDesc"       "Perc_Desc"    "KgMerm"
## [21] "Perc_Merm"    "KgDescCamp"   "KgBruLt"      "KgBruHa"      "KgExpHa"
```

```
## [26] "Status"       "Cliente"      "TempProm"     "TempMax"      "TempMin"
## [31] "HumProm"      "HumMax"       "HumMin"       "ETAcum"       "Producto"
## [36] "Naturaleza"   "DosisLtxHa"   "Tipo"
```

```
#STR
str(data)
```

```
## tibble [121 x 38] (S3: tbl_df/tbl/data.frame)
##  $ Semana      : num [1:121] 13 13 13 13 13 13 13 13 13 13 ...
##  $ Mes         : chr [1:121] "Marzo" "Marzo" "Marzo" "Marzo" ...
##  $ Fundo       : chr [1:121] "Don Nico" "Don Nico" "Don Nico" "Don Nico" ...
##  $ Empresa     : chr [1:121] "Agricola Guili S.A.C" "Agricola Guili S.A.C" "Agricola Guili S.A.C" "Ag
##  $ Lote        : chr [1:121] "DNH-06" "DNH-05" "DNH-04" "DNH-03" ...
##  $ Ha          : num [1:121] 9.75 10.91 11.76 11.62 11.5 ...
##  $ Variedad    : chr [1:121] "Hass" "Hass" "Hass" "Hass" ...
##  $ Color       : chr [1:121] "Negra" "Negra" "Negra" "Negra" ...
##  $ NoGR        : chr [1:121] "T008 N° 0000109" "T008 N° 0000109" "T008 N° 0000109" "T008 N° 0000109"
##  $ NoRepProd   : chr [1:121] "0001-0002239" "0001-0002239" "0001-0002239" "0001-0002239" ...
##  $ Fecha       : POSIXct[1:121], format: "2024-03-25" "2024-03-25" ...
##  $ TotJabas    : num [1:121] 5 10 18 19 5 3 7 1 6 5 ...
##  $ PesoPromJkg : num [1:121] 428 428 428 428 428 ...
##  $ CajasExp10kg: num [1:121] 201 401 722 762 201 ...
##  $ IngPakBrkg  : num [1:121] 2141 4281 7706 8135 2141 ...
##  $ KgExp       : num [1:121] 2005 4010 7218 7619 2005 ...
##  $ Perc_Exp    : num [1:121] 0.937 0.937 0.937 0.937 0.937 ...
##  $ KgDesc      : num [1:121] 45.6 91.2 164.2 173.4 45.6 ...
##  $ Perc_Desc   : num [1:121] 0.0213 0.0213 0.0213 0.0213 0.0213 ...
##  $ KgMerm      : num [1:121] 90 180 324 342 90 ...
##  $ Perc_Merm   : num [1:121] 0.0421 0.0421 0.0421 0.0421 0.0421 ...
##  $ KgDescCamp  : num [1:121] 0 0 0 0 0 0 0 0 0 0 ...
##  $ KgBruLt     : num [1:121] 2141 4281 7706 8135 2141 ...
##  $ KgBruHa     : num [1:121] 219 392 656 700 186 ...
##  $ KgExpHa     : num [1:121] 206 367 614 655 174 ...
##  $ Status      : chr [1:121] "Cosechando" "Cosechando" "Cosechando" "Cosechando" ...
##  $ Cliente     : chr [1:121] "BAIKA" "BAIKA" "BAIKA" "BAIKA" ...
##  $ TempProm    : num [1:121] 24.3 24.3 24.3 24.3 24.3 ...
##  $ TempMax     : num [1:121] 32 32 32 32 32 32 32.5 32.5 32.5 32.5 ...
##  $ TempMin     : num [1:121] 19.2 19.2 19.2 19.2 19.2 19.2 21 21 21 21 ...
##  $ HumProm     : num [1:121] 67 67 67 67 67 ...
##  $ HumMax      : num [1:121] 85 85 85 85 85 85 76 76 76 76 ...
##  $ HumMin      : num [1:121] 45 45 45 45 45 45 48 48 48 48 ...
##  $ ETAcum      : num [1:121] 5.62 5.62 5.62 5.62 5.62 5.62 4.92 4.92 4.92 4.92 ...
##  $ Producto    : chr [1:121] "BIOMURIKATA" "KING PLUS ZINC" "KADONDO" "BASFOLIAR POTASIO" ...
##  $ Naturaleza  : chr [1:121] "ORGANICA" "MINERAL" "QUIMICA" "MINERAL" ...
##  $ DosisLtxHa  : num [1:121] 2 1 1.5 1.6 1 3 2 1 1 1 ...
##  $ Tipo        : chr [1:121] "ACARICIDA" "NUTRIENTE" "ACARICIDA" "NUTRIENTE" ...
```

```
# Check for missing data
sum(is.na(data))
```

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

```
colnames(data)
```

```
##  [1] "Semana"       "Mes"          "Fundo"        "Empresa"      "Lote"
##  [6] "Ha"           "Variedad"     "Color"        "NoGR"         "NoRepProd"
## [11] "Fecha"        "TotJabas"     "PesoPromJkg"  "CajasExp10kg" "IngPakBrkg"
```

```
## [16] "KgExp"         "Perc_Exp"     "KgDesc"       "Perc_Desc"    "KgMerm"
## [21] "Perc_Merm"     "KgDescCamp"   "KgBruLt"      "KgBruHa"      "KgExpHa"
## [26] "Status"        "Cliente"      "TempProm"     "TempMax"      "TempMin"
## [31] "HumProm"       "HumMax"       "HumMin"       "ETAcum"       "Producto"
## [36] "Naturaleza"    "DosisLtxHa"   "Tipo"
```

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

```
##      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                Ha           Variedad            Color
##  Length:121         Min.   : 6.553   Length:121         Length:121
##  Class :character   1st Qu.:10.244   Class :character   Class :character
##  Mode  :character   Median :11.756   Mode  :character   Mode  :character
##                     Mean   :14.222
##                     3rd Qu.:17.183
##                     Max.   :26.570
##      NoGR            NoRepProd             Fecha
##  Length:121         Length:121         Min.   :2024-03-25 00:00:00.00
##  Class :character   Class :character   1st Qu.:2024-03-27 00:00:00.00
##  Mode  :character   Mode  :character   Median :2024-04-06 00:00:00.00
##                                        Mean   :2024-04-05 06:56:31.74
##                                        3rd Qu.:2024-04-12 00:00:00.00
##                                        Max.   :2024-04-16 00:00:00.00
##     TotJabas      PesoPromJkg     CajasExp10kg       IngPakBrkg
##  Min.   : 1.00   Min.   :  0.0   Min.   :   0.00   Min.   :    0.0
##  1st Qu.: 2.00   1st Qu.:416.2   1st Qu.:  80.77   1st Qu.:  840.2
##  Median : 5.00   Median :426.7   Median : 204.62   Median : 2169.6
##  Mean   :14.83   Mean   :402.0   Mean   : 580.69   Mean   : 6291.2
##  3rd Qu.:21.00   3rd Qu.:439.0   3rd Qu.: 858.19   3rd Qu.: 9175.9
##  Max.   :60.00   Max.   :679.5   Max.   :2448.89   Max.   :27291.0
##      KgExp            Perc_Exp          KgDesc          Perc_Desc
##  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   :0.02326
##  3rd Qu.: 8581.9   3rd Qu.:0.9618   3rd Qu.: 267.62   3rd Qu.:0.02917
##  Max.   :24488.9   Max.   :1.0000   Max.   :2736.90   Max.   :0.10029
##      KgMerm          Perc_Merm         KgDescCamp        KgBruLt
##  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 :0.03495   Median :  0.00   Median : 2169.6
##  Mean   :218.54   Mean   :0.02671   Mean   : 22.12   Mean   : 6313.3
##  3rd Qu.:342.16   3rd Qu.:0.03941   3rd Qu.:  0.00   3rd Qu.: 9175.9
##  Max.   :986.92   Max.   :0.07947   Max.   :483.50   Max.   :27291.0
##     KgBruHa            KgExpHa           Status            Cliente
##  Min.   :   3.729   Min.   :   0.00   Length:121         Length:121
##  1st Qu.:  49.568   1st Qu.:  46.04   Class :character   Class :character
##  Median : 202.479   Median : 192.18   Mode  :character   Mode  :character
```

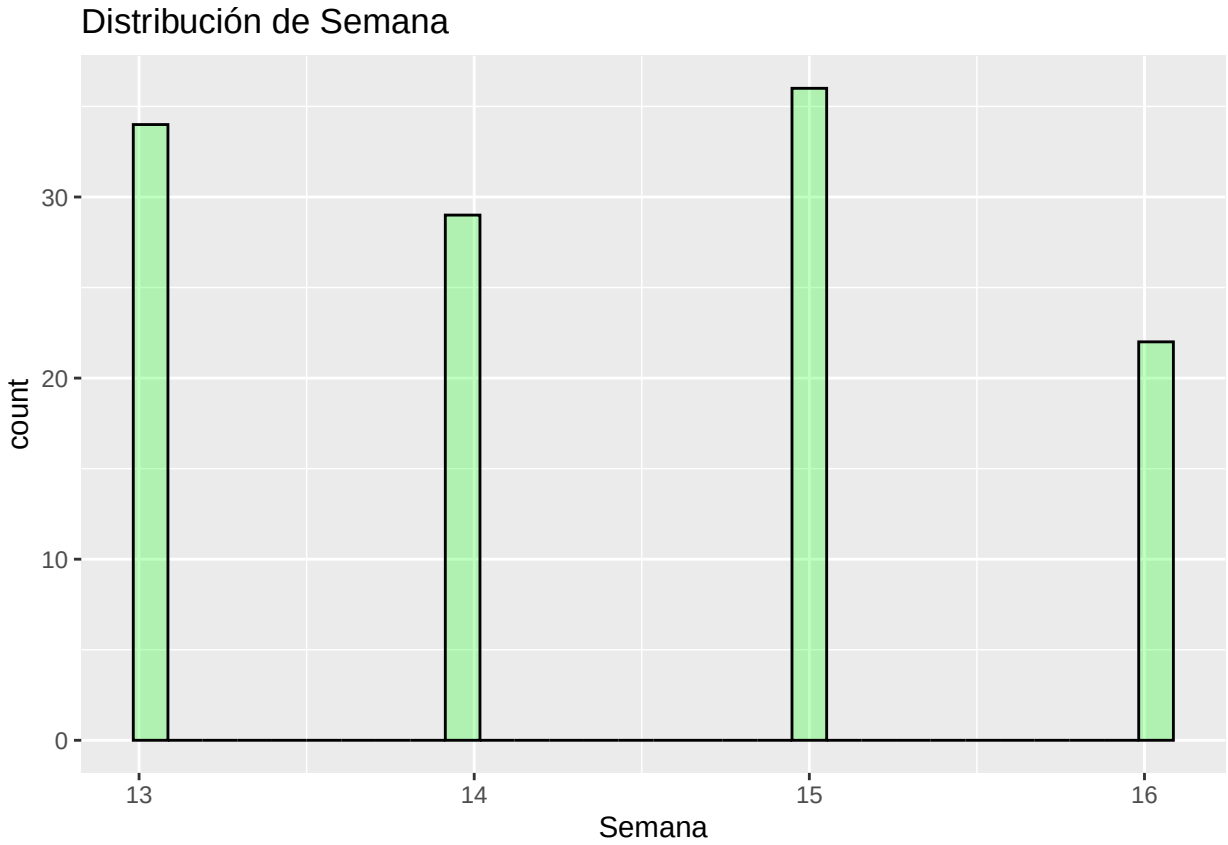
```
##  Mean   : 479.526   Mean   : 440.45
##  3rd Qu.: 680.728   3rd Qu.: 647.45
##  Max.   :2476.181   Max.   :2244.27
##     TempProm        TempMax         TempMin         HumProm
##  Min.   :20.56   Min.   :26.10   Min.   :16.80   Min.   :64.46
##  1st Qu.:22.90   1st Qu.:30.50   1st Qu.:18.20   1st Qu.:67.06
##  Median :23.63   Median :32.00   Median :18.90   Median :68.21
##  Mean   :23.34   Mean   :31.38   Mean   :18.97   Mean   :70.14
##  3rd Qu.:24.28   3rd Qu.:32.80   3rd Qu.:19.70   3rd Qu.:72.56
##  Max.   :25.20   Max.   :34.20   Max.   :21.00   Max.   :80.85
##      HumMax          HumMin          ETAcum        Producto
##  Min.   :76.00   Min.   :36.00   Min.   :4.370   Length:121
##  1st Qu.:84.00   1st Qu.:41.00   1st Qu.:4.920   Class :character
##  Median :85.00   Median :45.00   Median :5.620   Mode  :character
##  Mean   :85.45   Mean   :45.93   Mean   :5.384
##  3rd Qu.:87.00   3rd Qu.:48.00   3rd Qu.:5.780
##  Max.   :94.00   Max.   :62.00   Max.   :6.240
##   Naturaleza          DosisLtxHa        Tipo
##  Length:121         Min.   :1.000   Length:121
##  Class :character   1st Qu.:1.000   Class :character
##  Mode  :character   Median :2.000   Mode  :character
##                     Mean   :1.986
##                     3rd Qu.:2.500
##                     Max.   :4.000
```

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

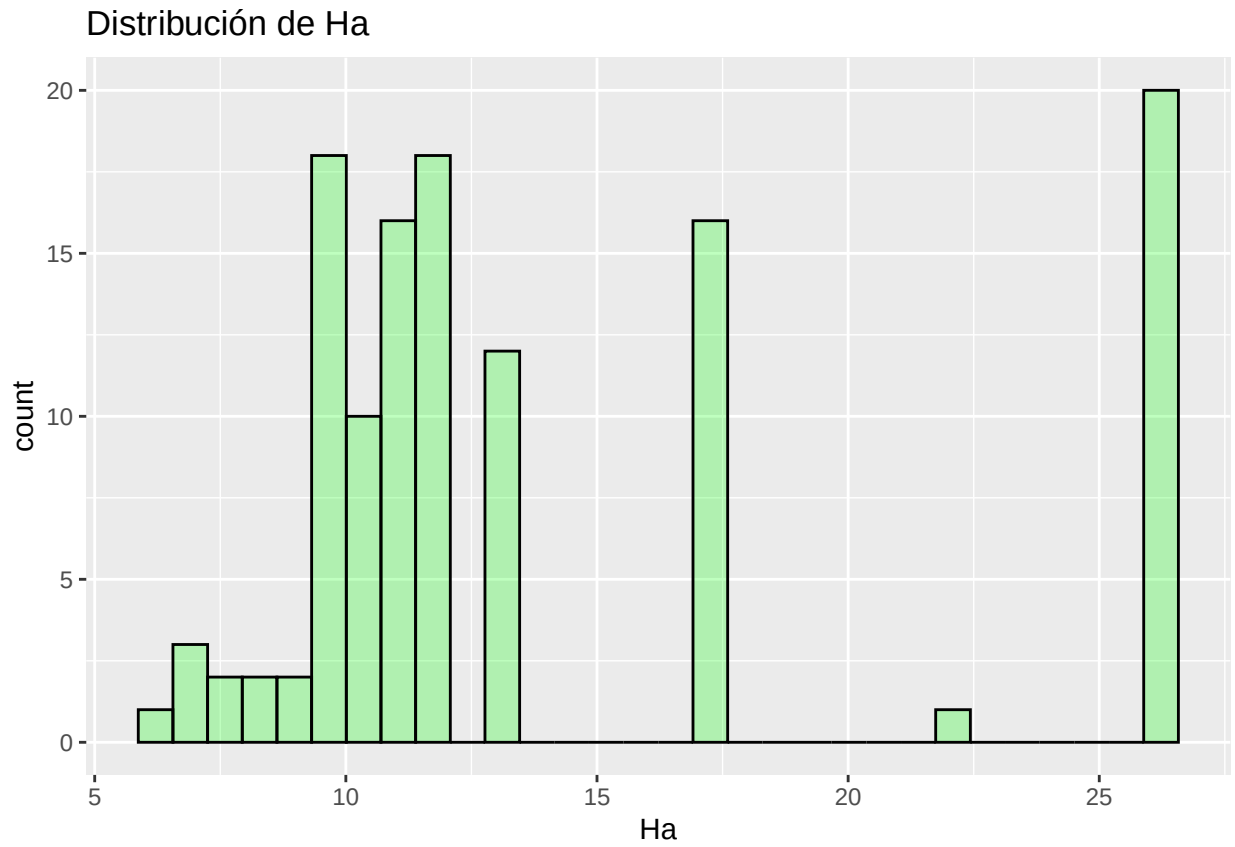
```
##       Semana          Mes        Fundo      Empresa         Lote           Ha
##         TRUE        FALSE        FALSE        FALSE        FALSE         TRUE
##     Variedad        Color         NoGR    NoRepProd        Fecha     TotJabas
##        FALSE        FALSE        FALSE        FALSE        FALSE         TRUE
##  PesoPromJkg CajasExp10kg   IngPakBrkg        KgExp     Perc_Exp       KgDesc
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
##    Perc_Desc       KgMerm    Perc_Merm   KgDescCamp      KgBruLt      KgBruHa
##         TRUE         TRUE         TRUE         TRUE         TRUE         TRUE
##      KgExpHa       Status      Cliente     TempProm      TempMax      TempMin
##         TRUE        FALSE        FALSE         TRUE         TRUE         TRUE
##      HumProm       HumMax       HumMin       ETAcum     Producto   Naturaleza
##         TRUE         TRUE         TRUE         TRUE        FALSE        FALSE
##   DosisLtxHa         Tipo
##         TRUE        FALSE
```

```r
library(rlang)  # for the sym() function
# Generating histograms using aes() with tidy evaluation
lapply(names(data_n), function(x) {
  ggplot(data, aes(x = !!sym(x))) +  # Use tidy evaluation to interpret x
    geom_histogram(alpha = 0.25, bins = 30, fill = "green", color = "black") +
    labs(title = paste("Distribución de", x))
})
```

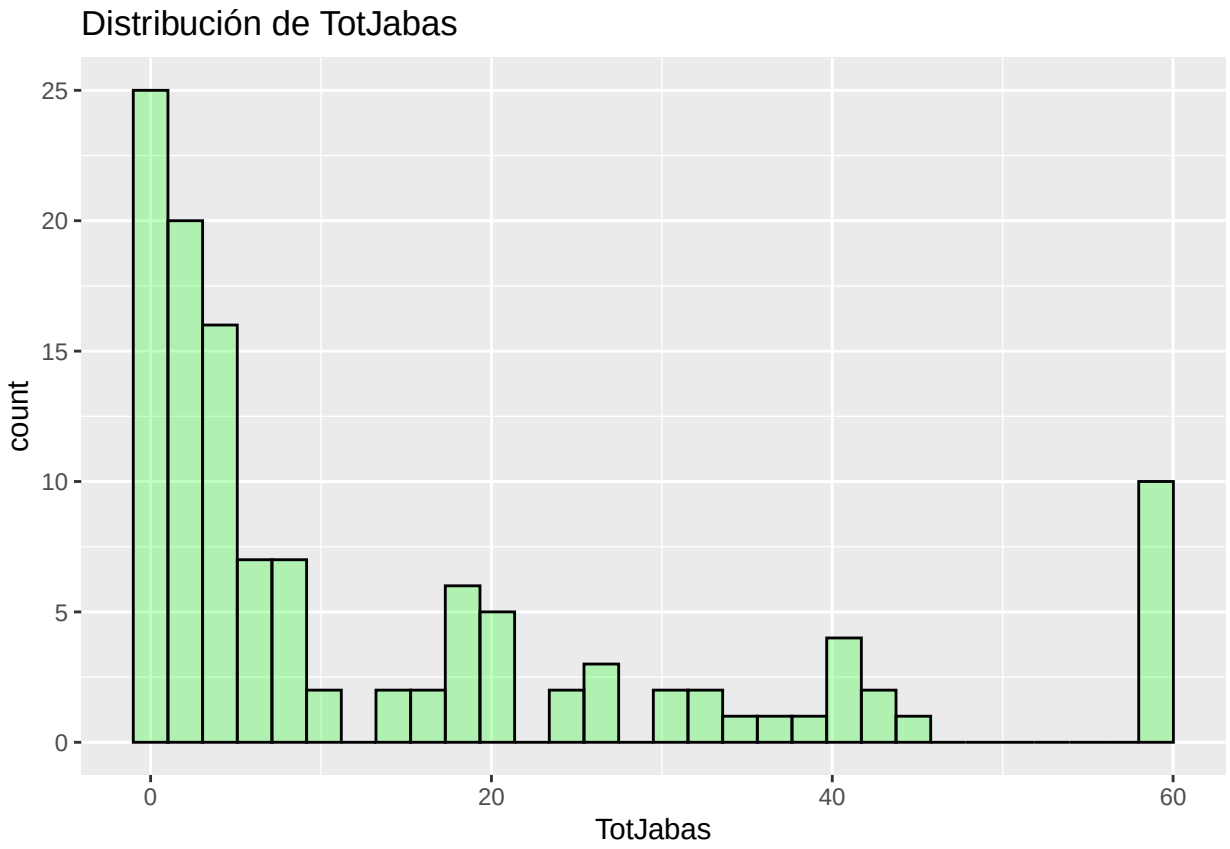
```
## [[1]]
```



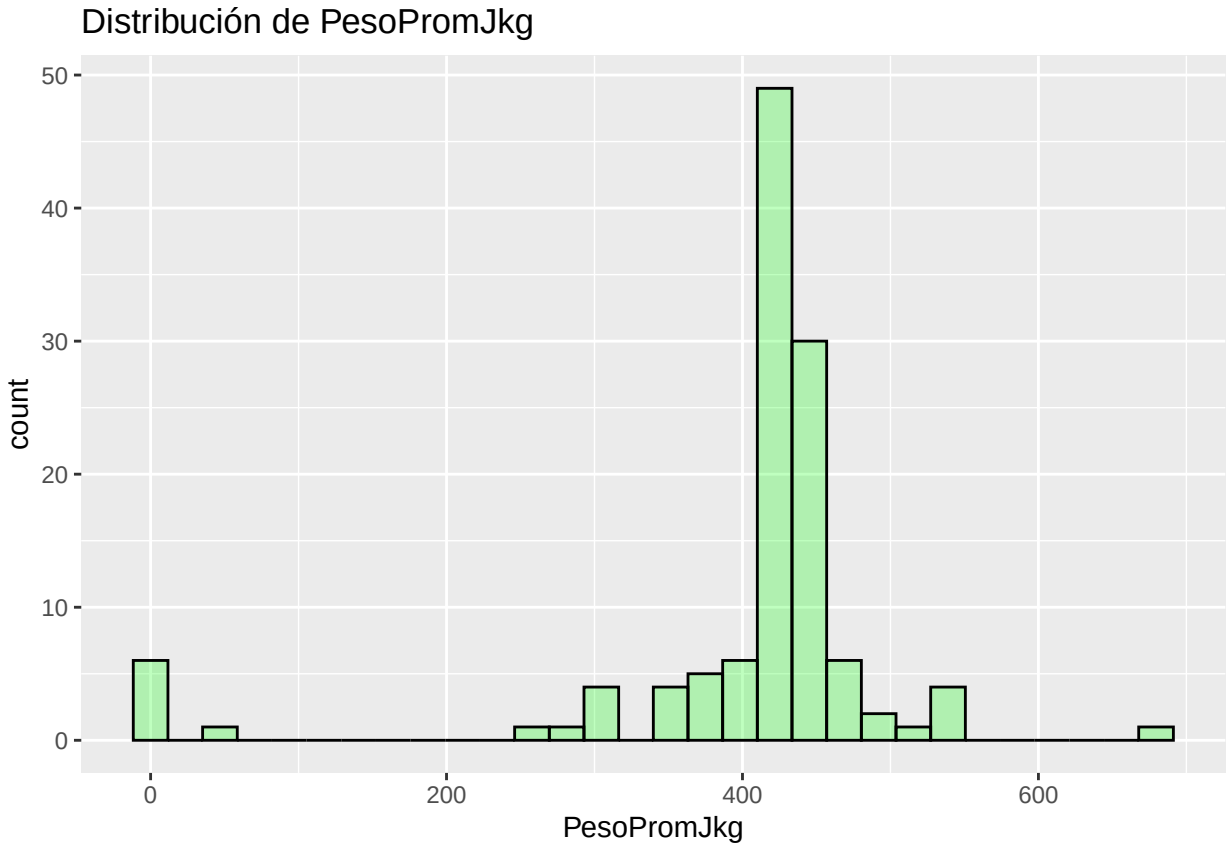

```
##
## [[2]]
```



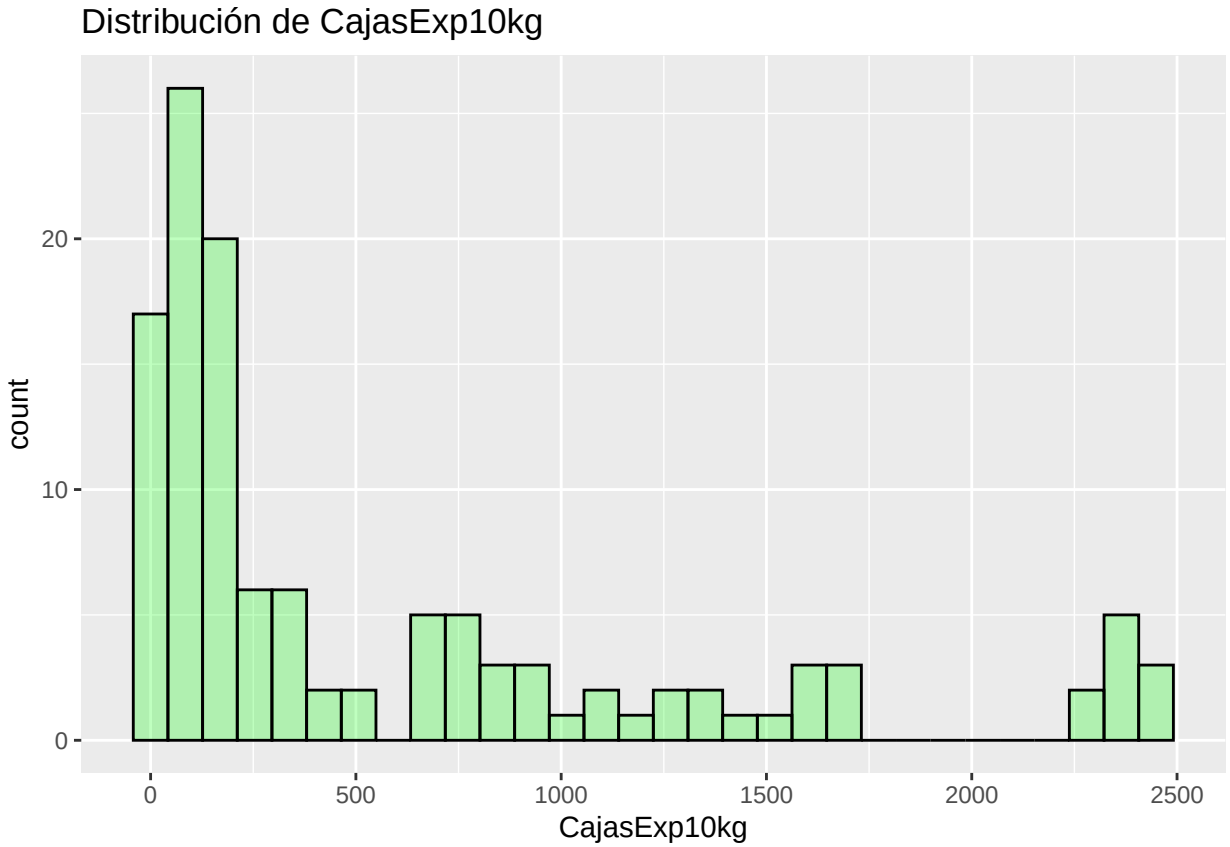

```
##
## [[3]]
```


Distribución de TotJabas
count
TotJabas
0
5
10
15
20
25
0
20
40
60

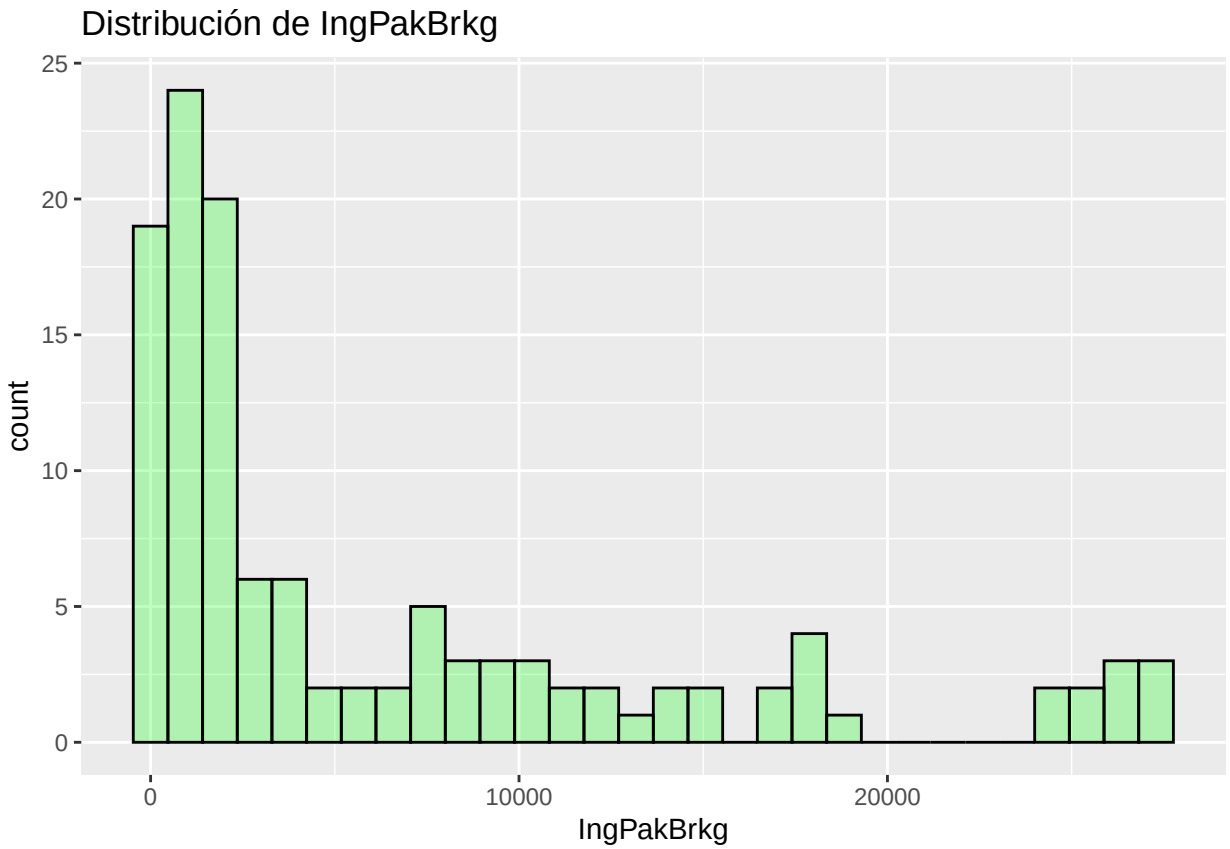
```
##
## [[4]]
```


Distribución de PesoPromJkg
count
PesoPromJkg
0
10
20
30
40
50
0
200
400
600

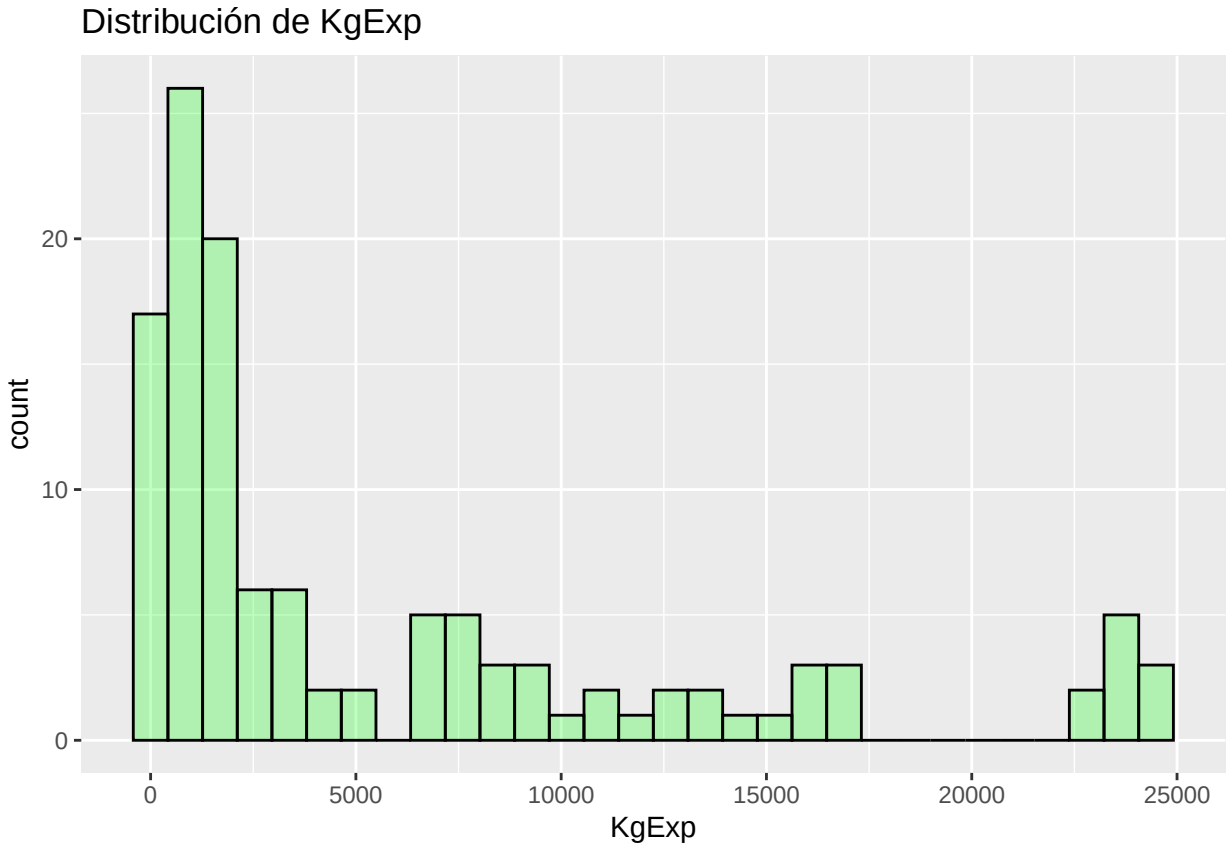
```
##
## [[5]]
```


Distribución de CajasExp10kg
count
20
10
0
0
500
1000
1500
2000
2500
CajasExp10kg

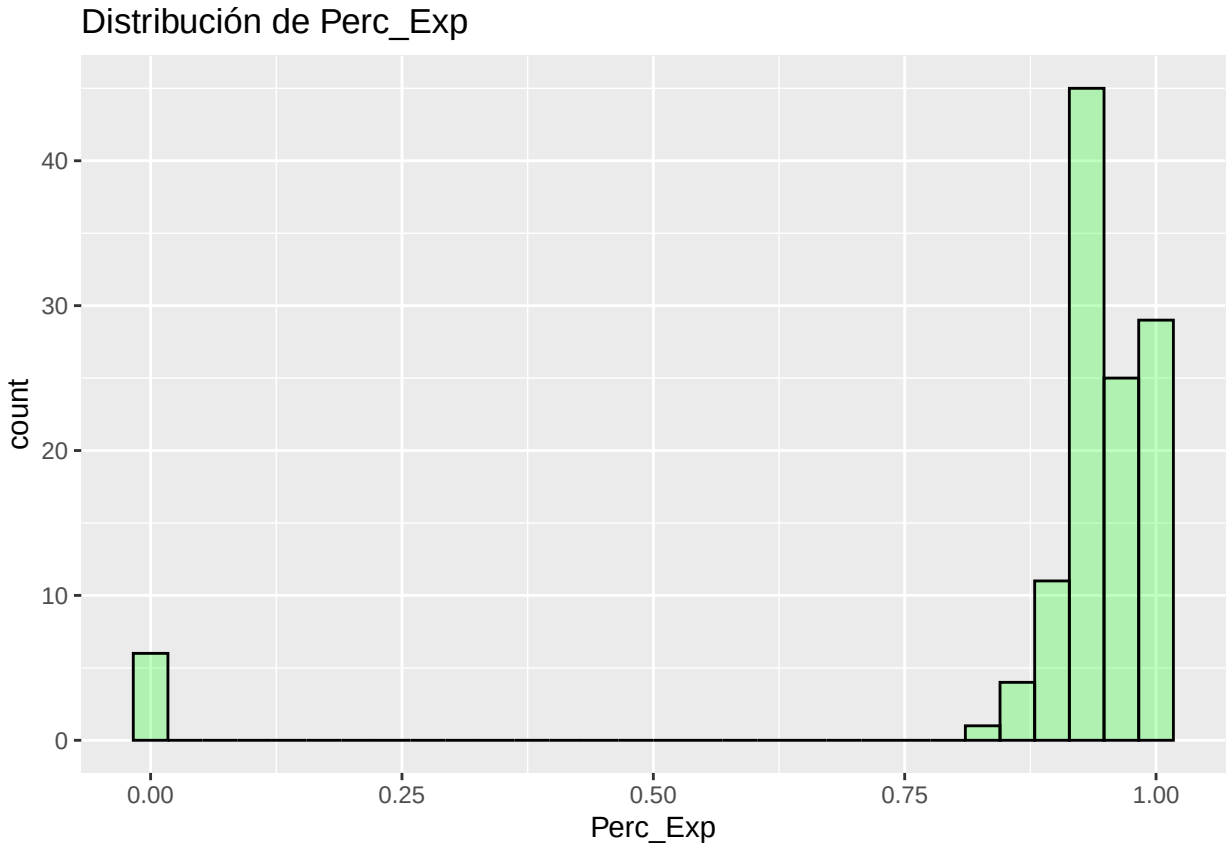
```
##
## [[6]]
```


Distribución de IngPakBrkg
25
20
15
10
5
0
count
0
10000
20000
IngPakBrkg

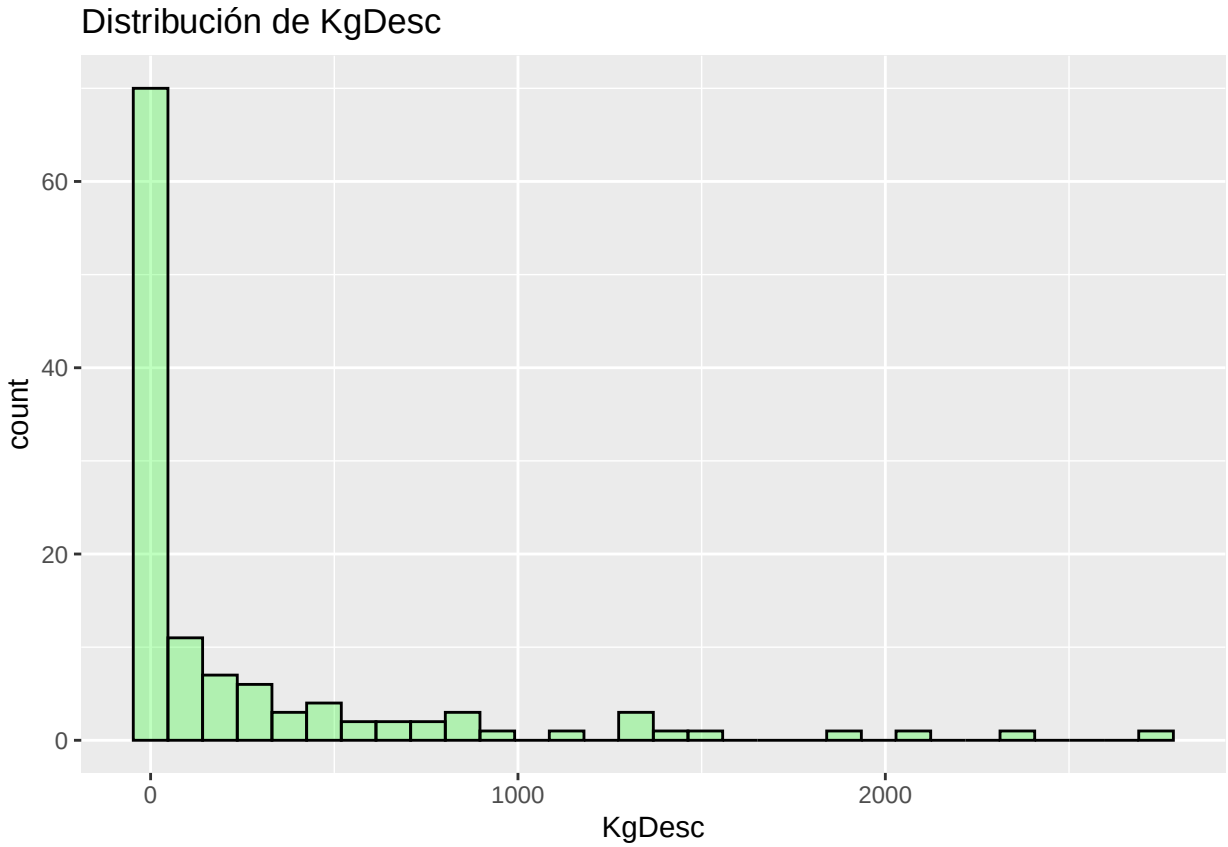
```
##
## [[7]]
```


Distribución de KgExp
count
20
10
0
0
5000
10000
15000
20000
25000
KgExp

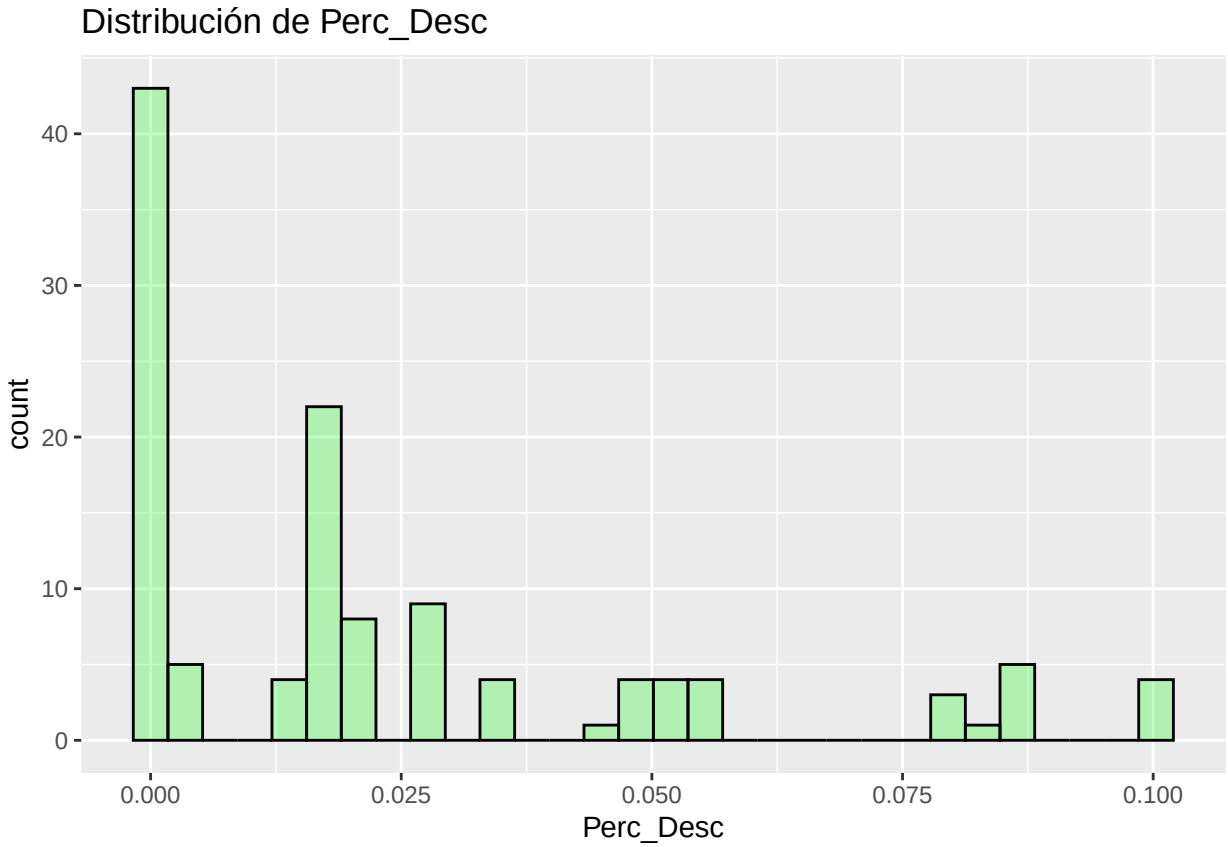
```
##
## [[8]]
```


Distribución de Perc_Exp
count
40
30
20
10
0
0.00
0.25
0.50
0.75
1.00
Perc_Exp

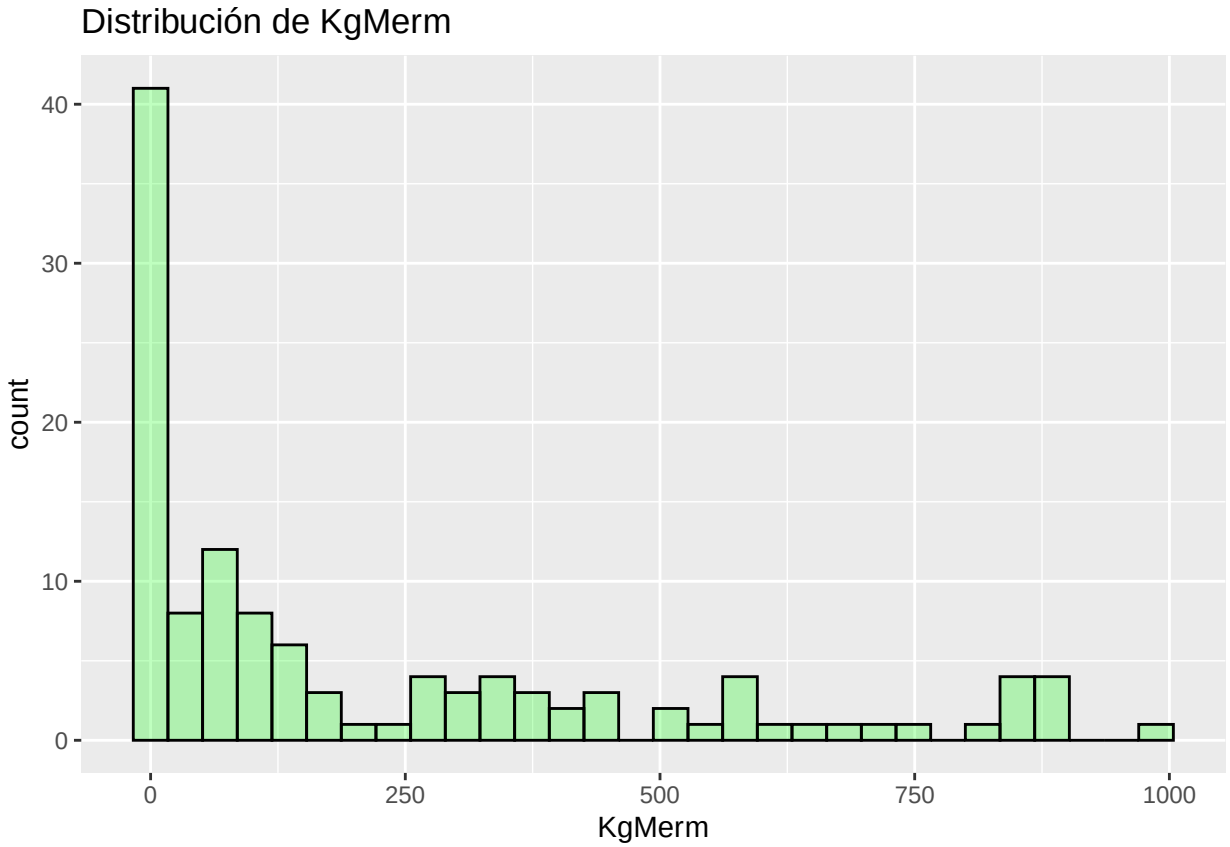
```
##
## [[9]]
```



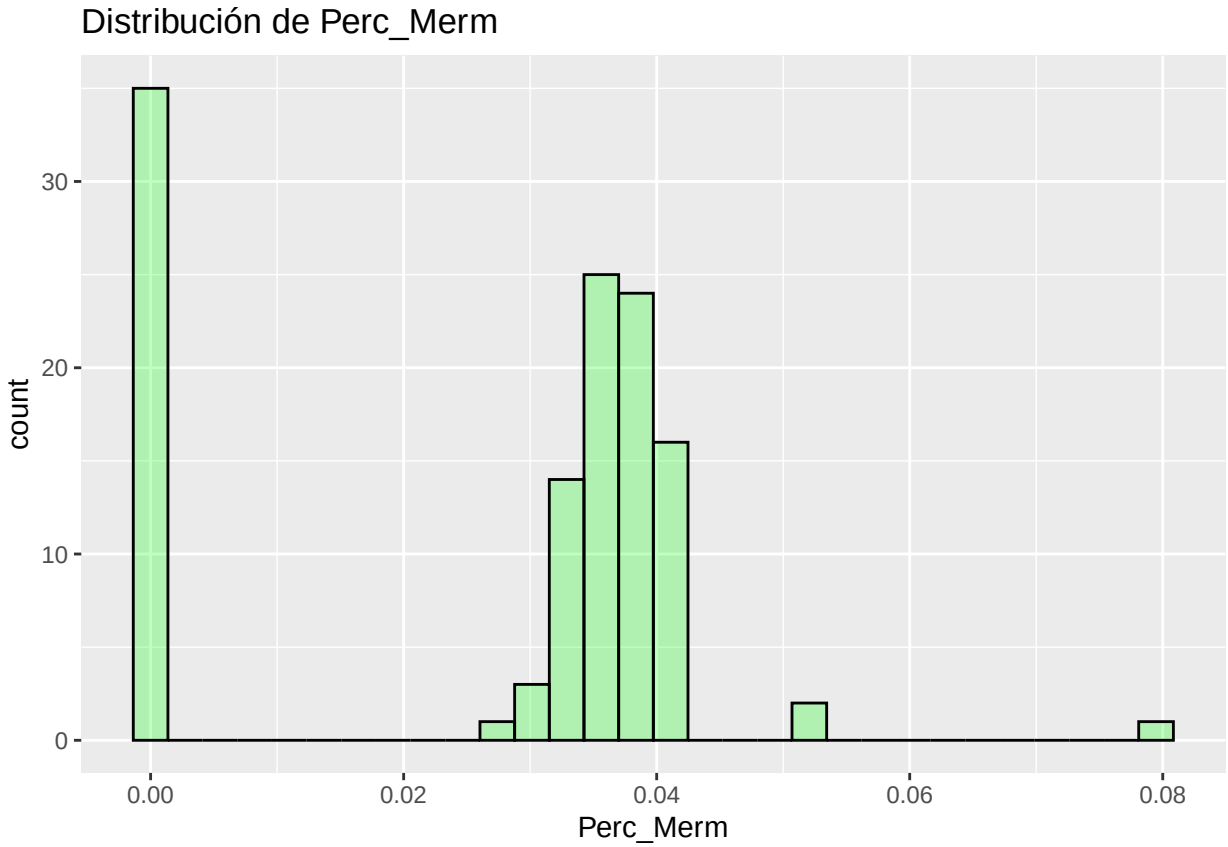

```
##
## [[10]]
```


Distribución de Perc_Desc
count
40
30
20
10
0
0.000
0.025
0.050
0.075
0.100
Perc_Desc

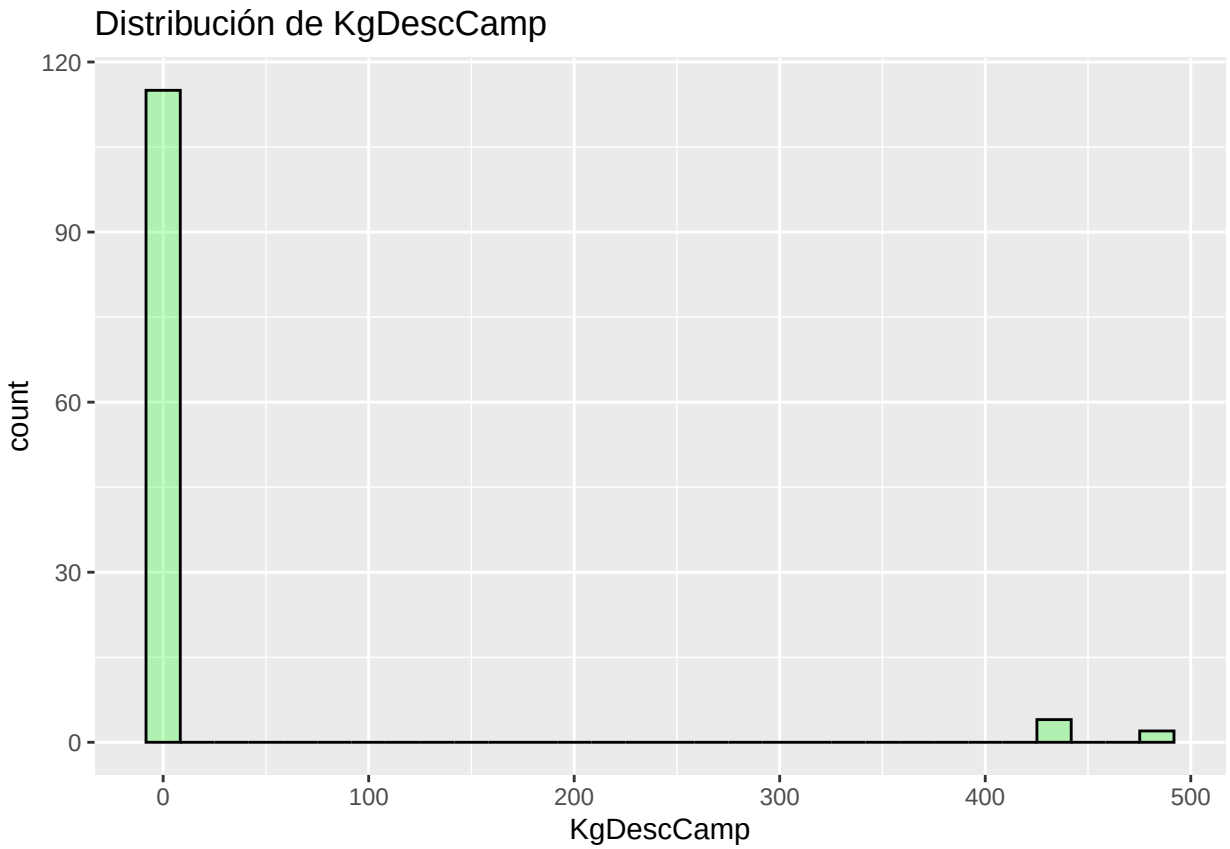
```
##
## [[11]]
```


Distribución de KgMerm
count
KgMerm
0
10
20
30
40
0
250
500
750
1000

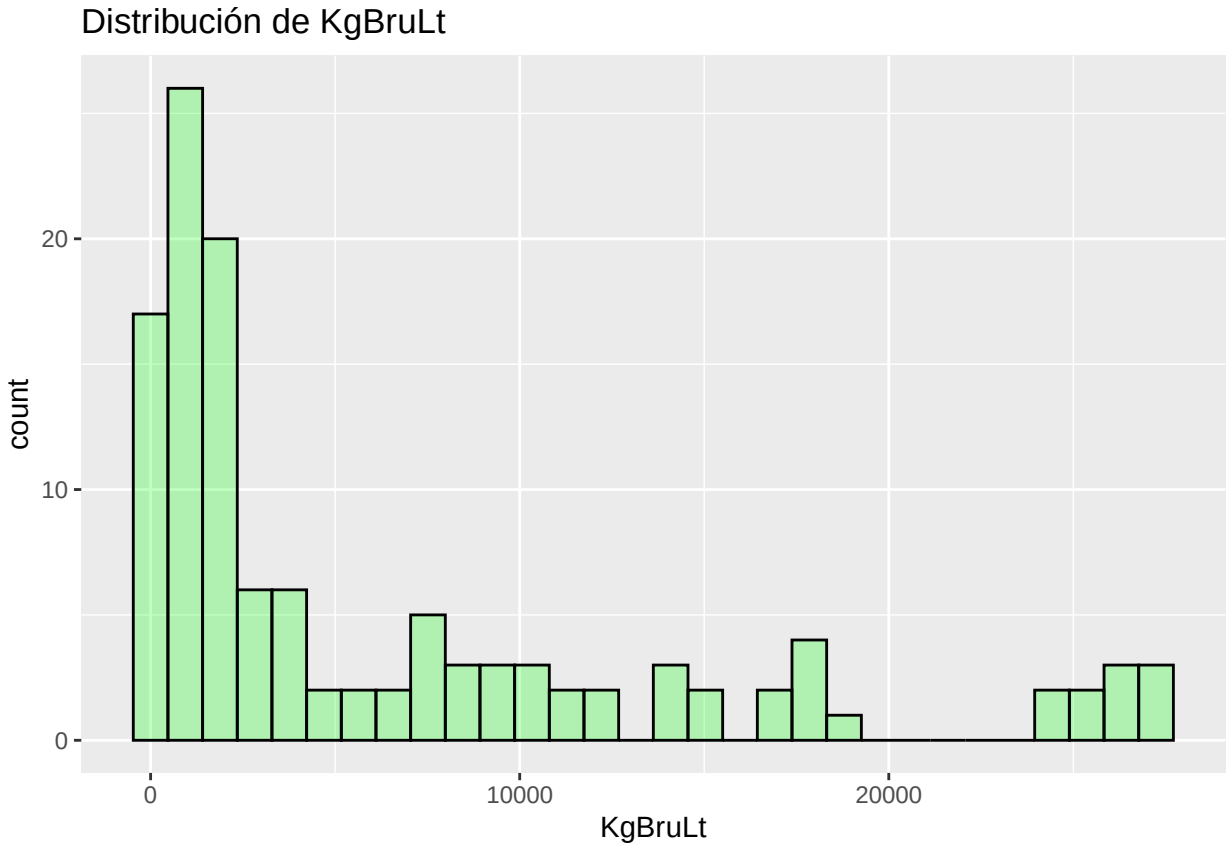
```
##
## [[12]]
```


Distribución de Perc_Merm
count
30
20
10
0
0.00
0.02
0.04
0.06
0.08
Perc_Merm

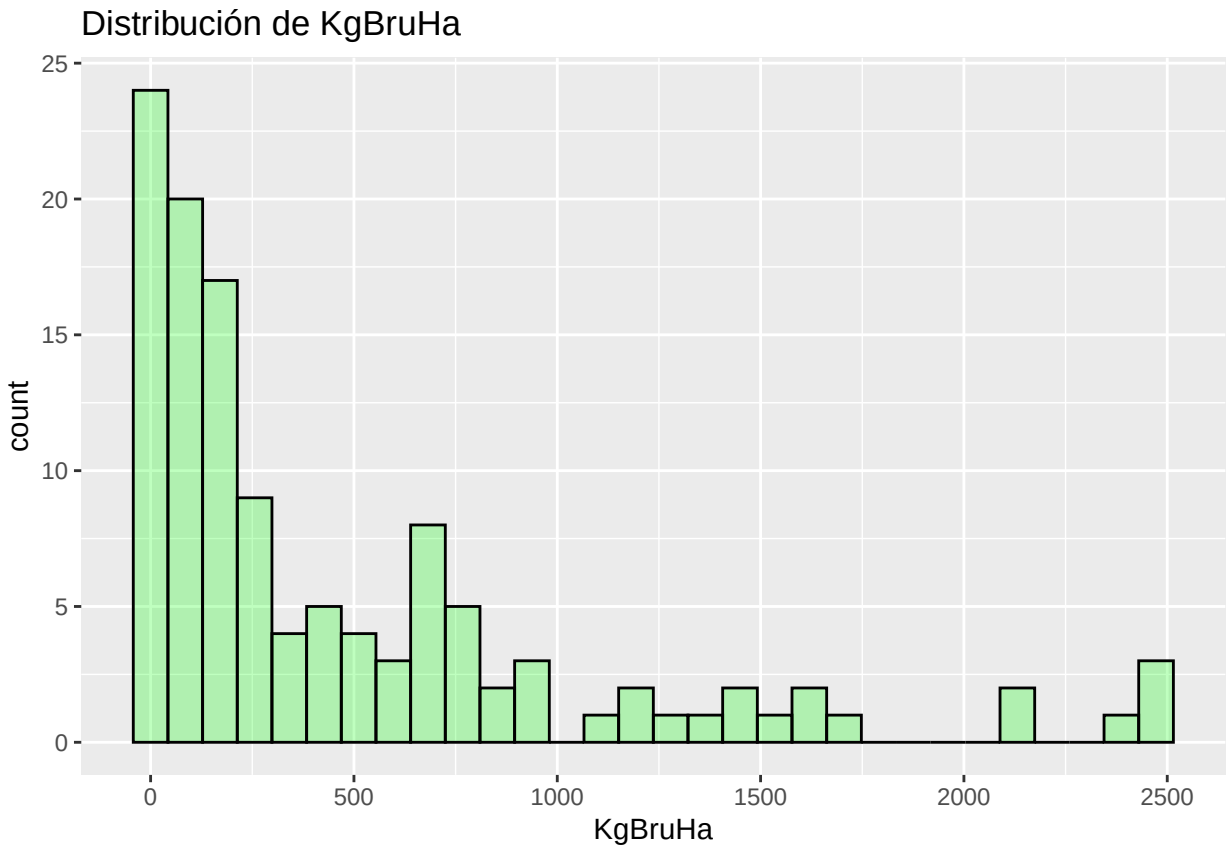
```
##
## [[13]]
```


Distribución de KgDescCamp
120
90
60
30
0
count
0
100
200
300
400
500
KgDescCamp

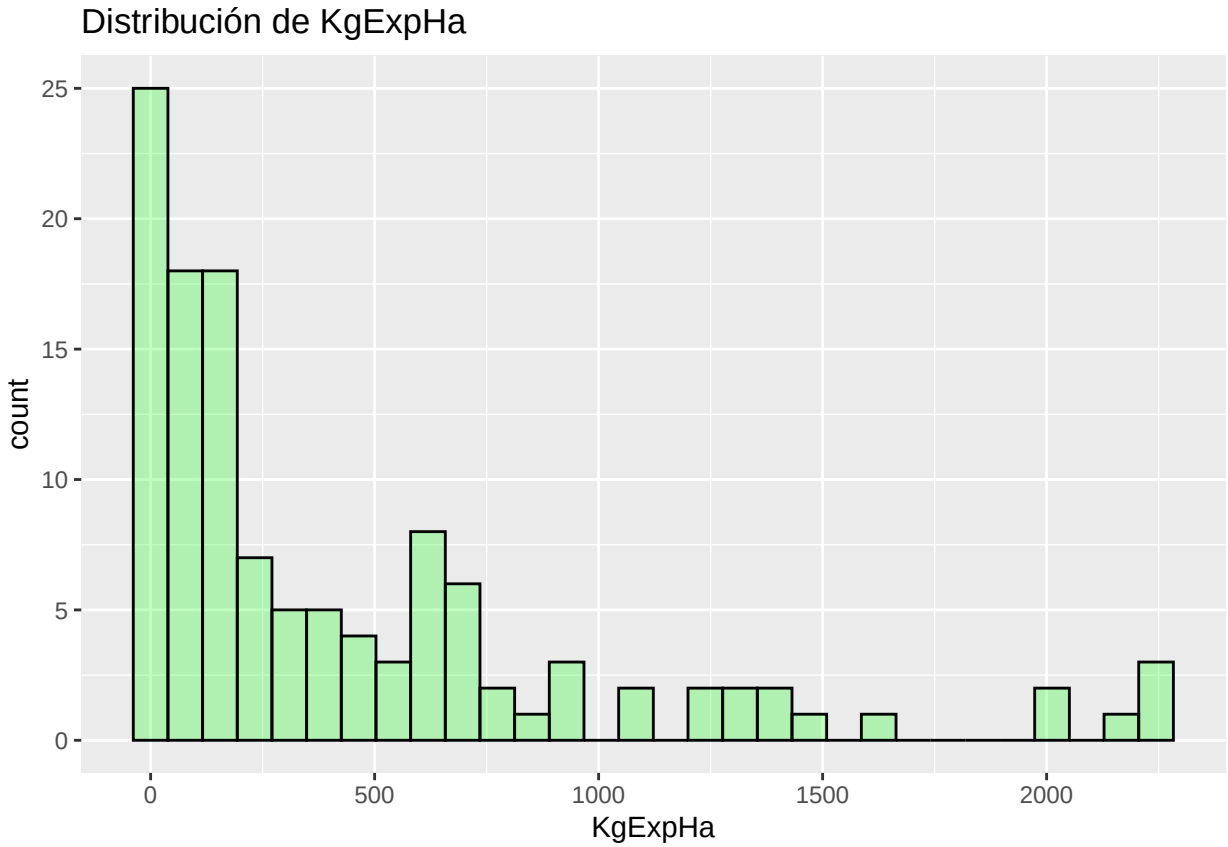
```
##
## [[14]]
```



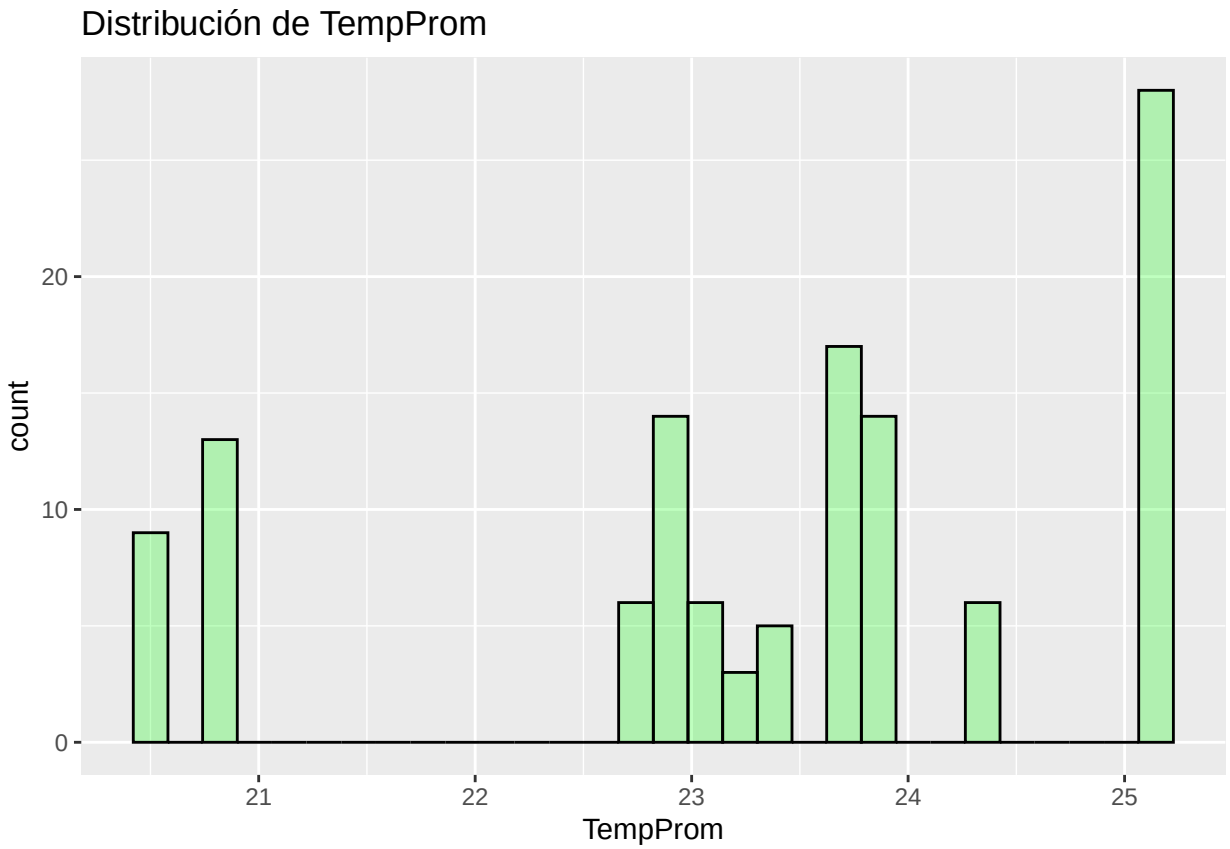

```
##
## [[15]]
```


Distribución de KgBruHa
25
20
15
10
5
0
count
0
500
1000
1500
2000
2500
KgBruHa

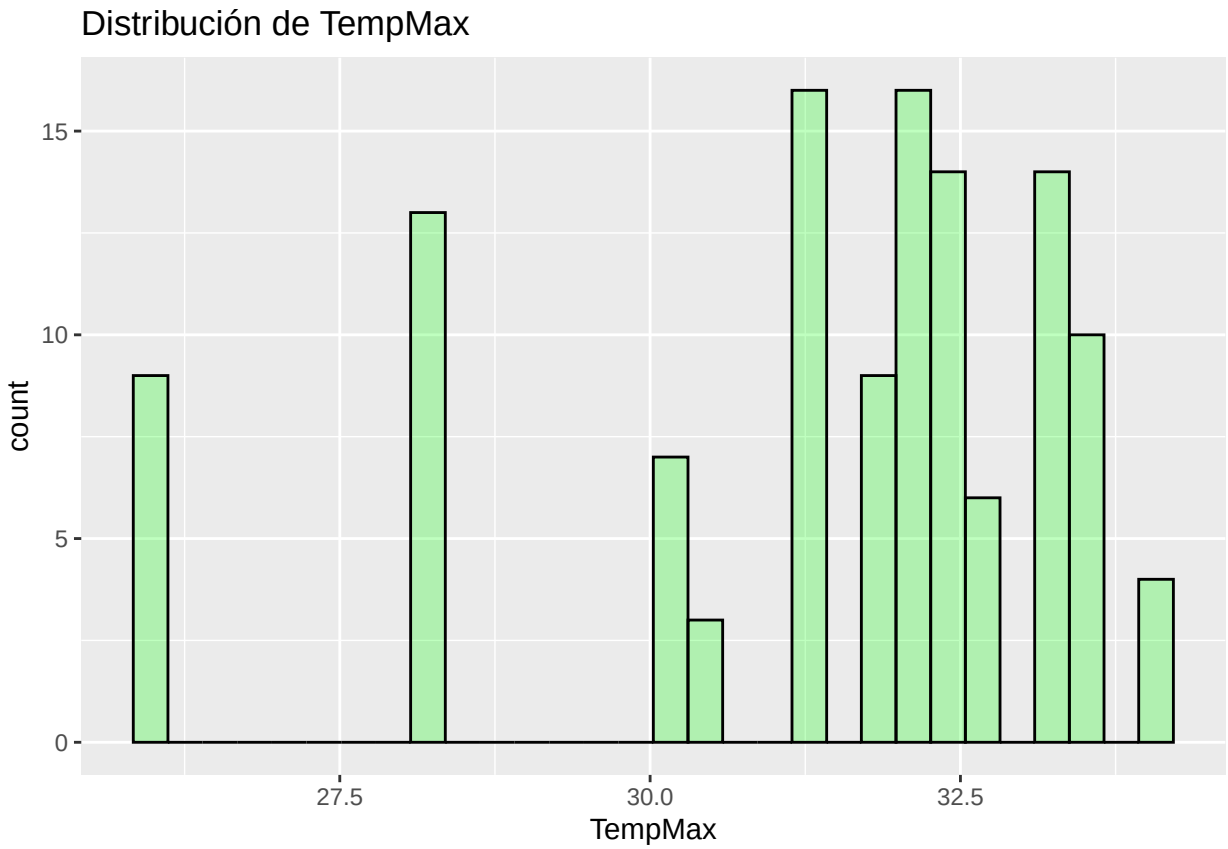
```
##
## [[16]]
```


Distribución de KgExpHa
25
20
15
10
5
0
count
0
500
1000
1500
2000
KgExpHa

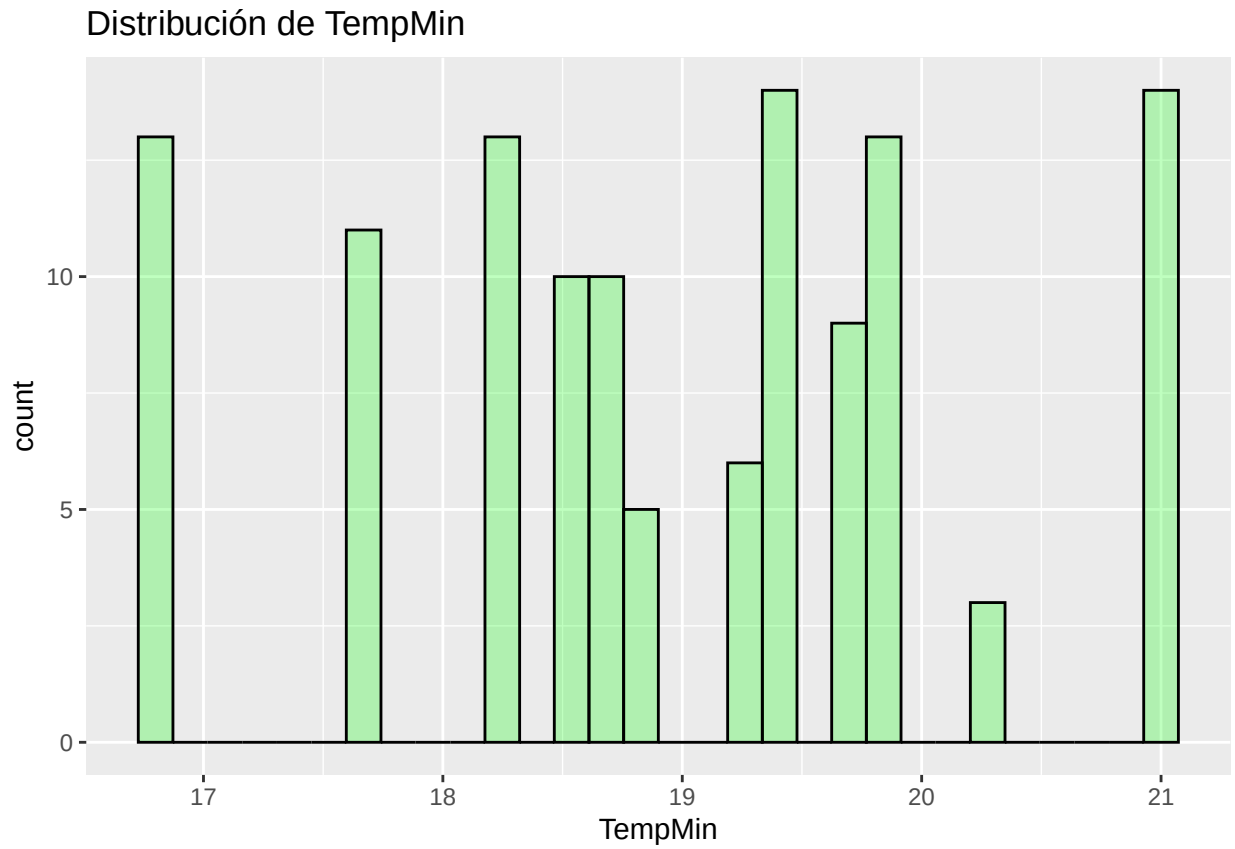
```
##
## [[17]]
```


Distribución de TempProm
count
20
10
0
21
22
23
24
25
TempProm

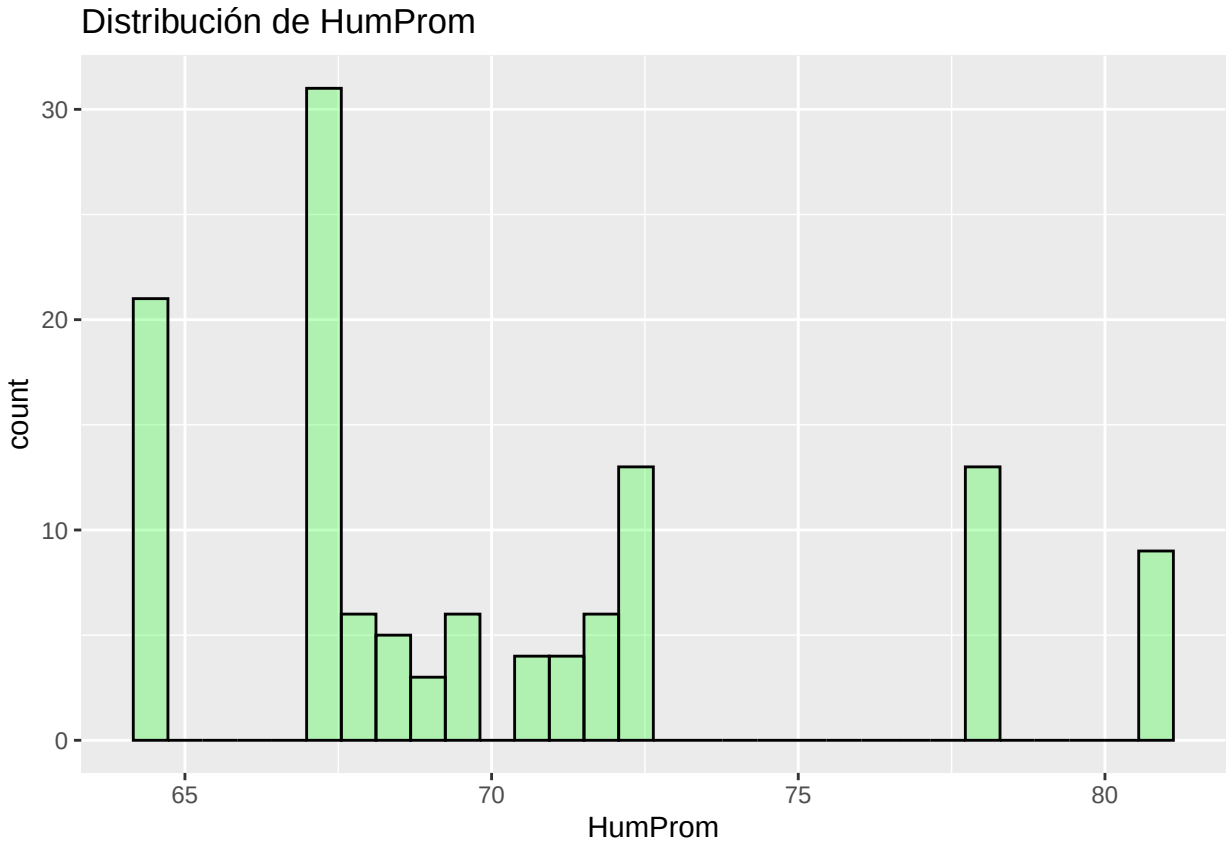
```
##
## [[18]]
```



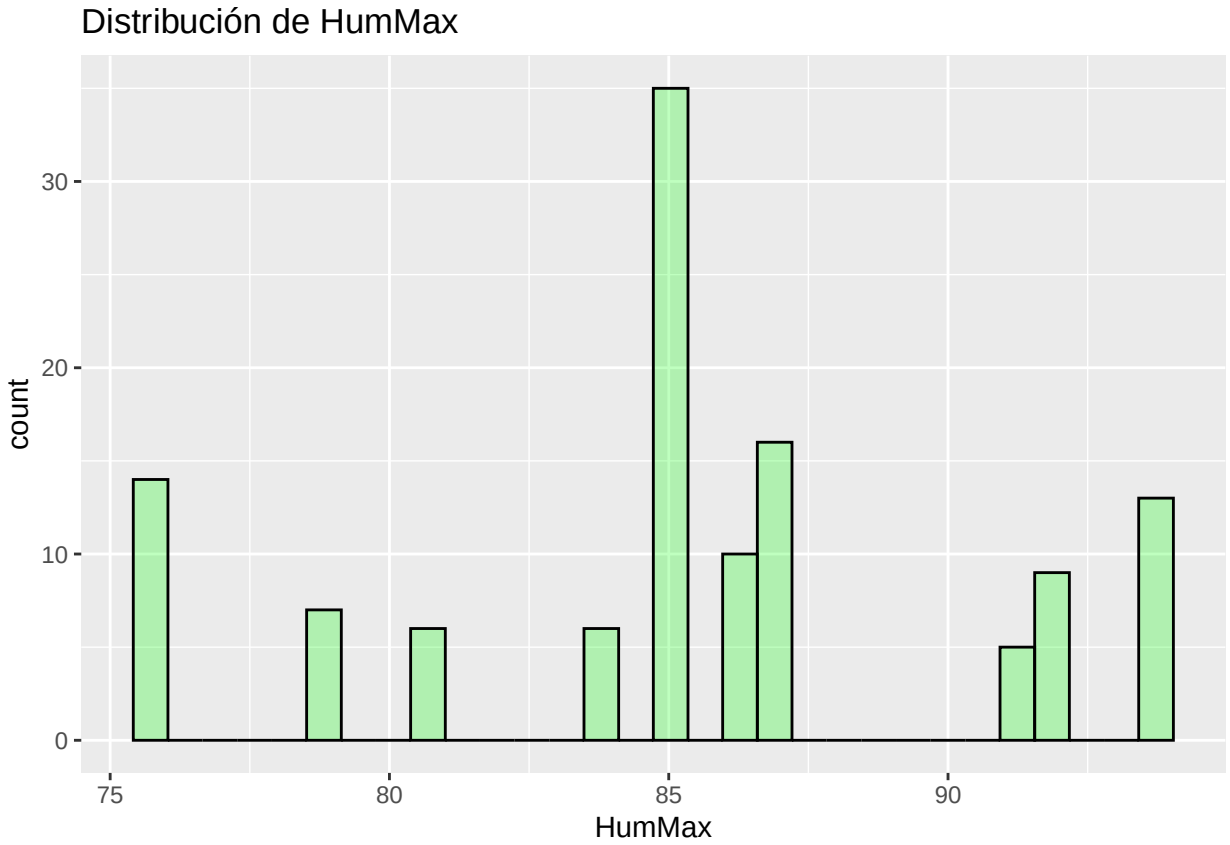

```
##
## [[19]]
```



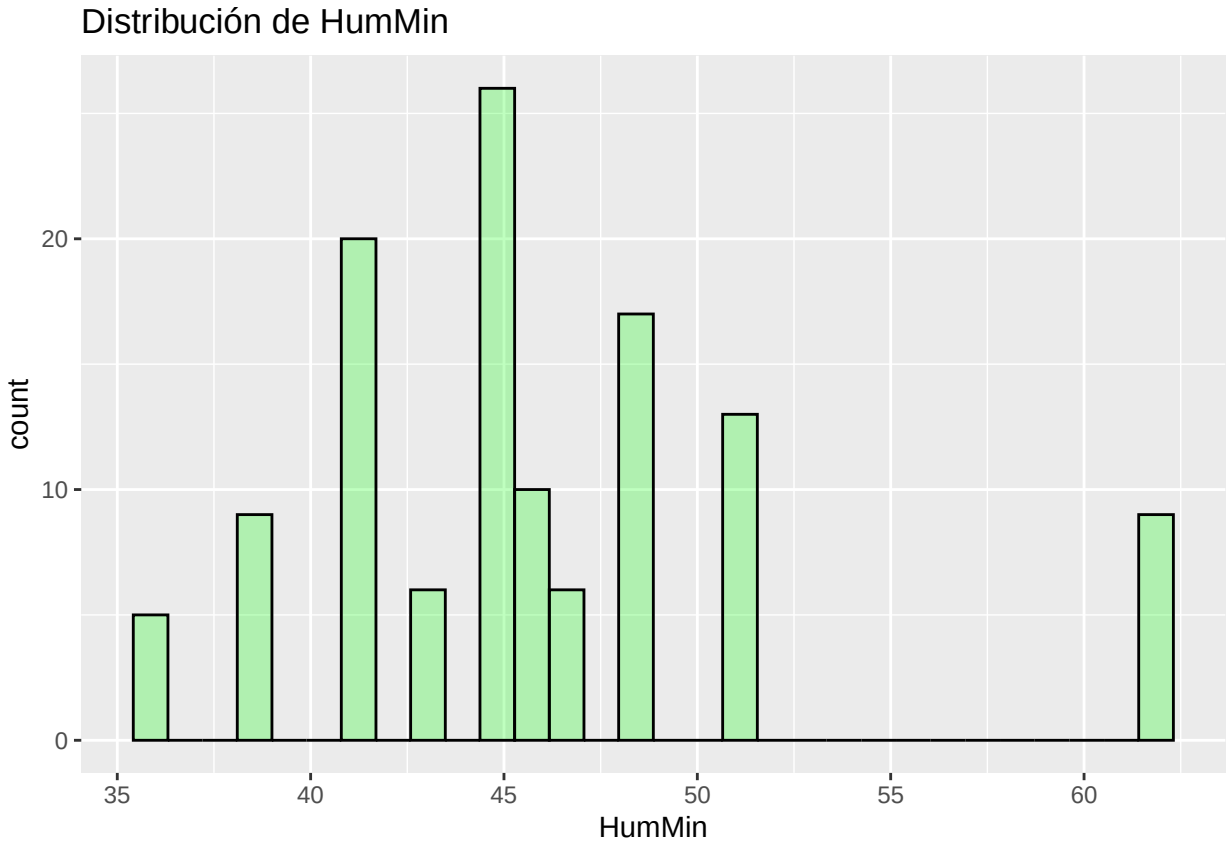

```
##
## [[20]]
```



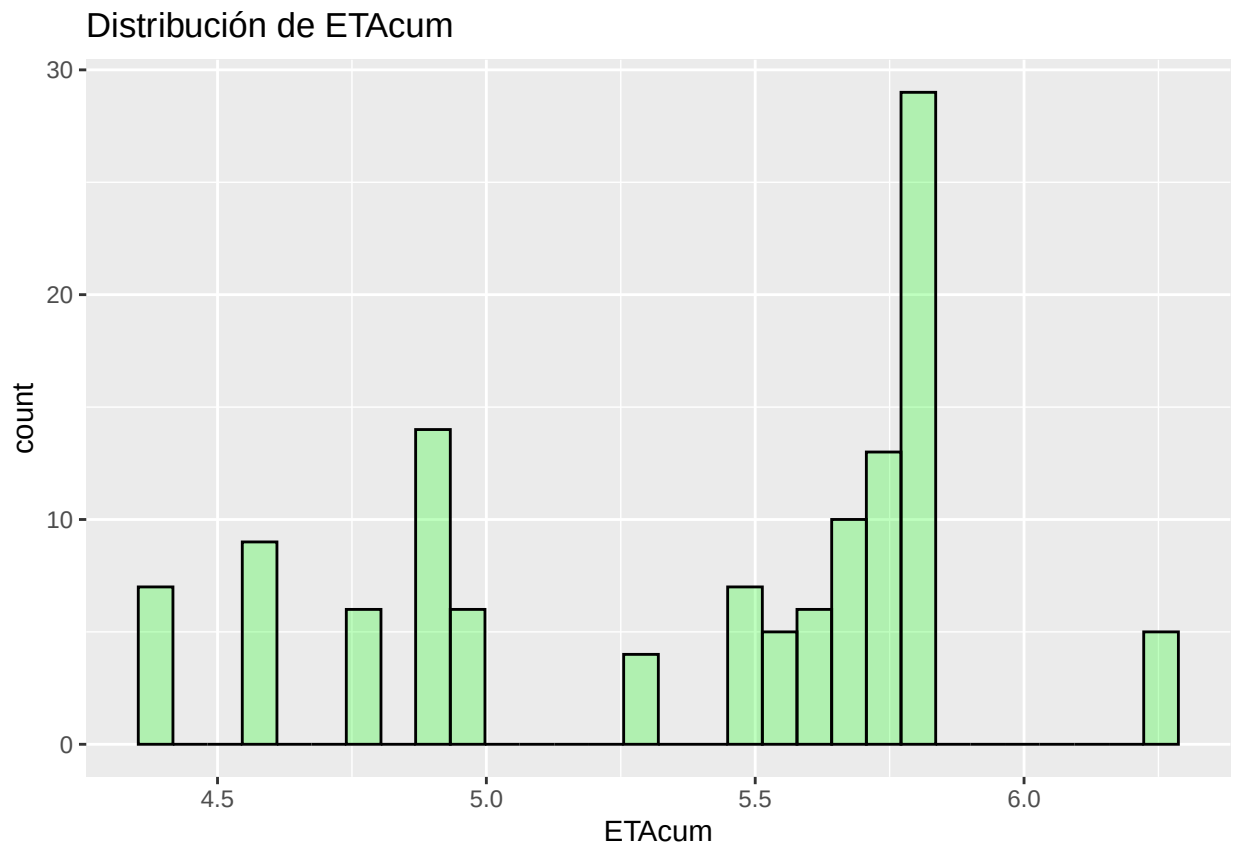

```
##
## [[21]]
```



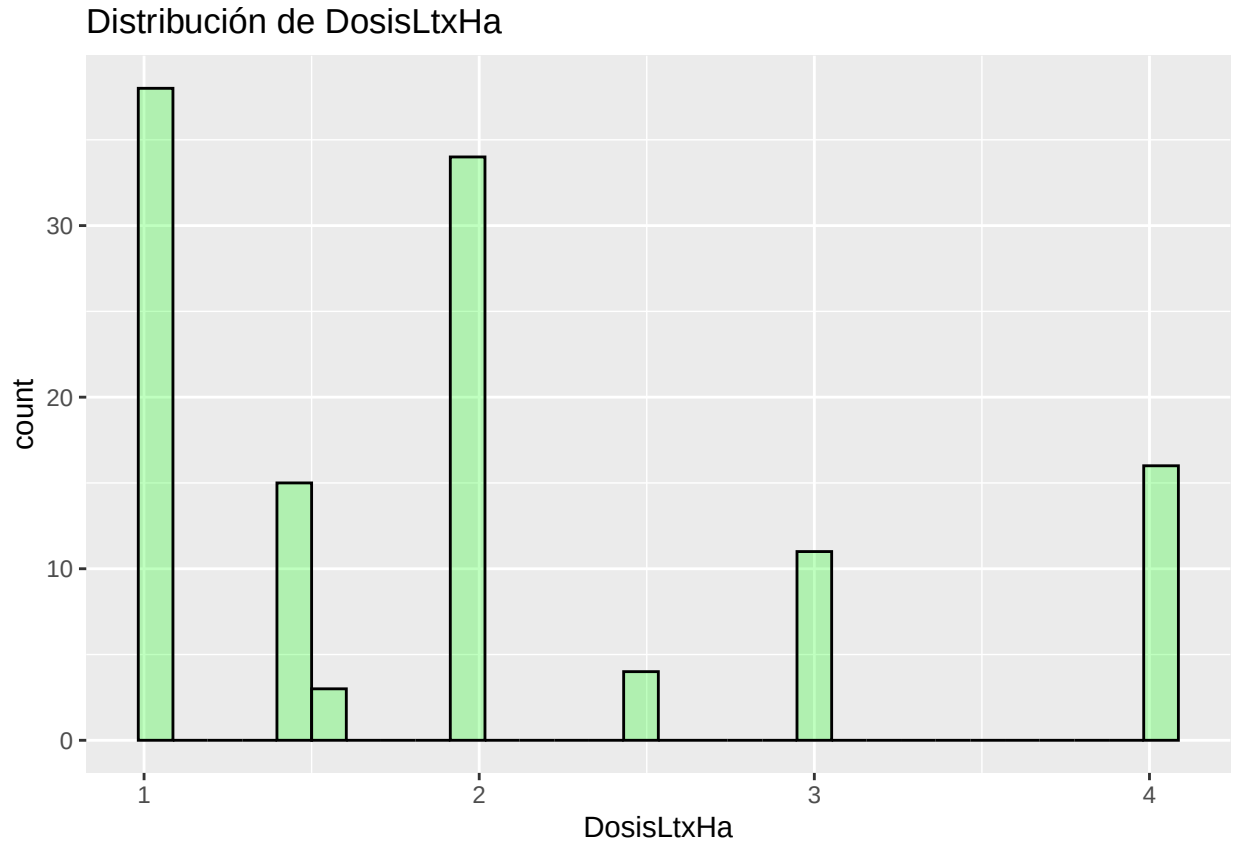

```
##
## [[22]]
```




```
##
## [[23]]
```




```
##
## [[24]]
```



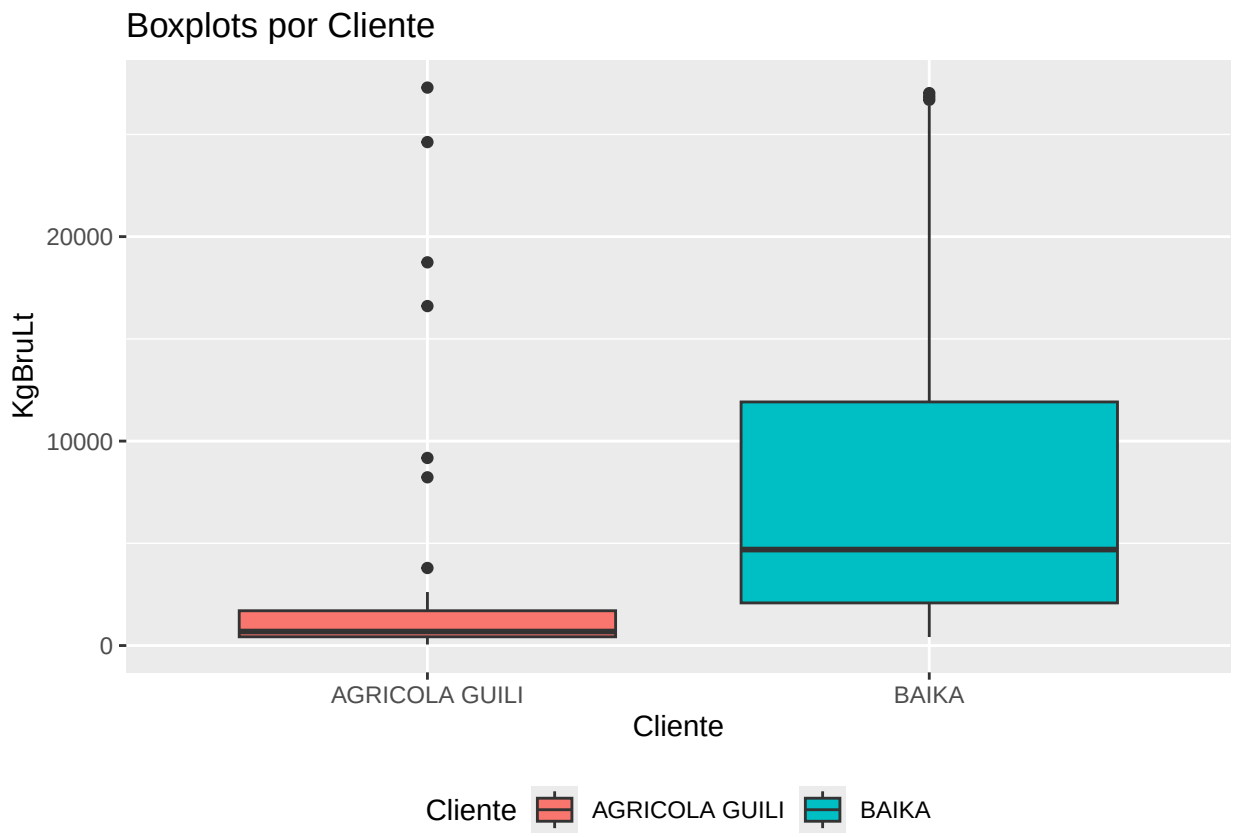

```
names(data)
```

```
##  [1] "Semana"       "Mes"          "Fundo"        "Empresa"      "Lote"
##  [6] "Ha"           "Variedad"     "Color"        "NoGR"         "NoRepProd"
## [11] "Fecha"        "TotJabas"     "PesoPromJkg"  "CajasExp10kg" "IngPakBrkg"
## [16] "KgExp"        "Perc_Exp"     "KgDesc"       "Perc_Desc"    "KgMerm"
## [21] "Perc_Merm"    "KgDescCamp"   "KgBruLt"      "KgBruHa"      "KgExpHa"
## [26] "Status"       "Cliente"      "TempProm"     "TempMax"      "TempMin"
## [31] "HumProm"      "HumMax"       "HumMin"       "ETAcum"       "Producto"
## [36] "Naturaleza"   "DosisLtxHa"   "Tipo"
```

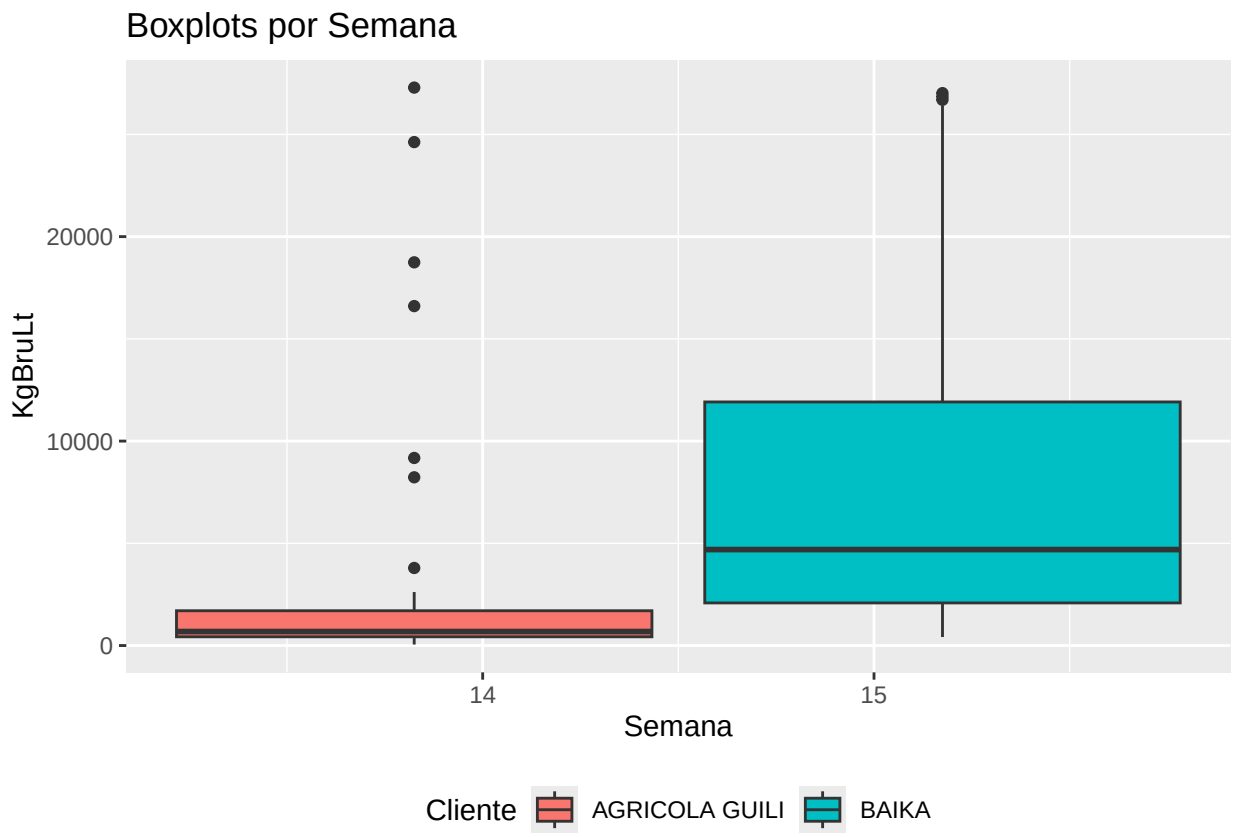
```
names(data_n)
```

```
##  [1] "Semana"       "Ha"           "TotJabas"     "PesoPromJkg"  "CajasExp10kg"
##  [6] "IngPakBrkg"   "KgExp"        "Perc_Exp"     "KgDesc"       "Perc_Desc"
## [11] "KgMerm"       "Perc_Merm"    "KgDescCamp"   "KgBruLt"      "KgBruHa"
## [16] "KgExpHa"      "TempProm"     "TempMax"      "TempMin"      "HumProm"
## [21] "HumMax"       "HumMin"       "ETAcum"       "DosisLtxHa"
```

```
########################################
# Configurar tamaño del gráfico
options(repr.plot.width=18, repr.plot.height=6)
# Crear el boxplot
boxplot <- ggplot(data, aes(x = Cliente, y = KgBruLt, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Cliente")
boxplot <- boxplot + theme(legend.position = "bottom")# Establecer la ubicación de la leyenda
print(boxplot) # Mostrar el boxplot
```



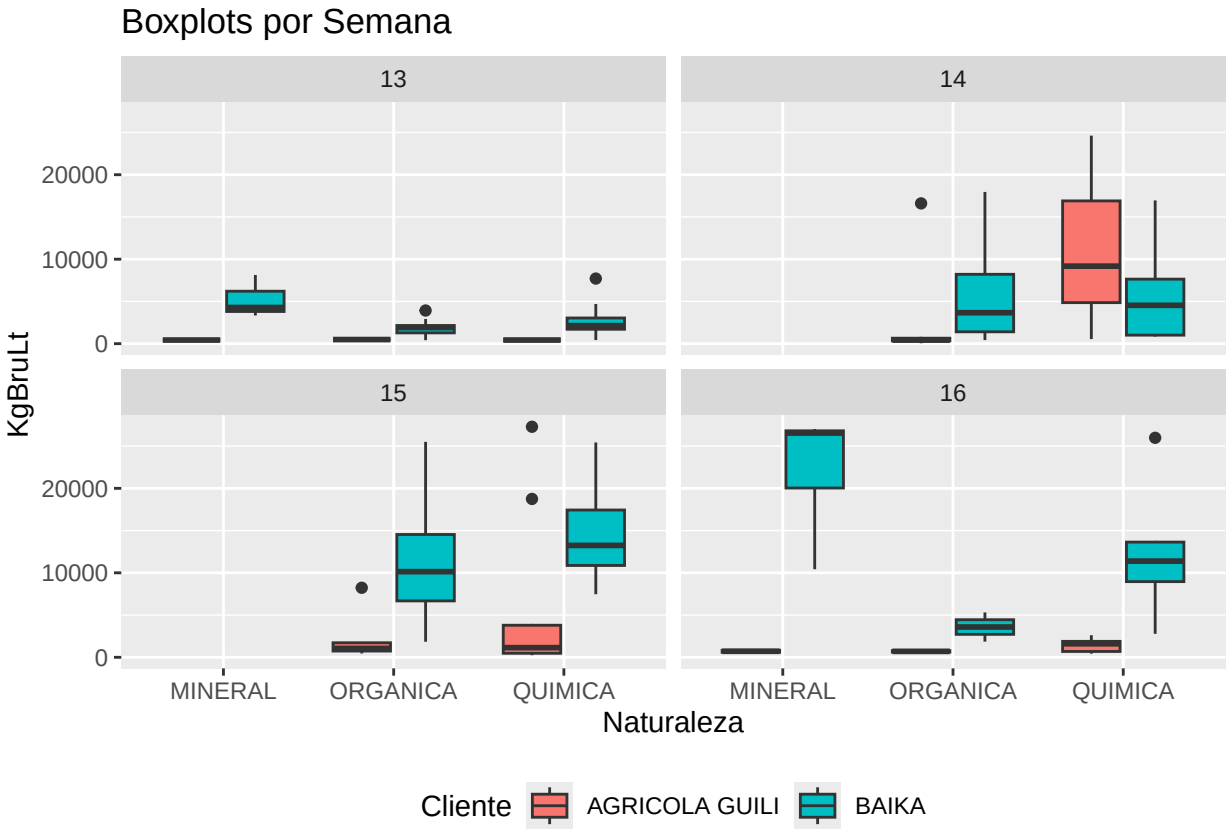

```
# Crear el boxplot
boxplot <- ggplot(data, aes(x = Semana, y = KgBruLt, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Semana")
# Establecer la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")
# Mostrar el boxplot
print(boxplot)
```



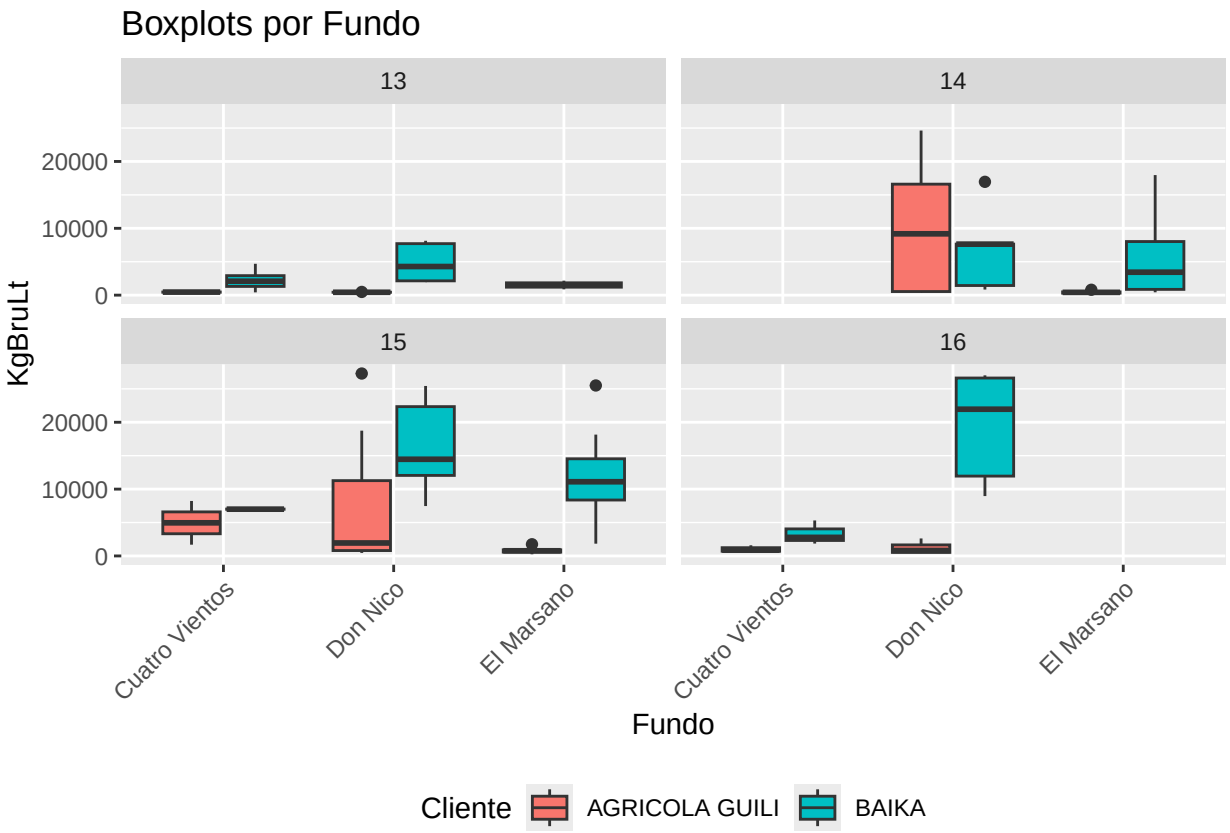

```
##########################################
names(data)
```

```
##  [1] "Semana"       "Mes"          "Fundo"        "Empresa"      "Lote"
##  [6] "Ha"           "Variedad"     "Color"        "NoGR"         "NoRepProd"
## [11] "Fecha"        "TotJabas"     "PesoPromJkg"  "CajasExp10kg" "IngPakBrkg"
## [16] "KgExp"        "Perc_Exp"     "KgDesc"       "Perc_Desc"    "KgMerm"
## [21] "Perc_Merm"    "KgDescCamp"   "KgBruLt"      "KgBruHa"      "KgExpHa"
## [26] "Status"       "Cliente"      "TempProm"     "TempMax"      "TempMin"
## [31] "HumProm"      "HumMax"       "HumMin"       "ETAcum"       "Producto"
## [36] "Naturaleza"   "DosisLtxHa"   "Tipo"
```

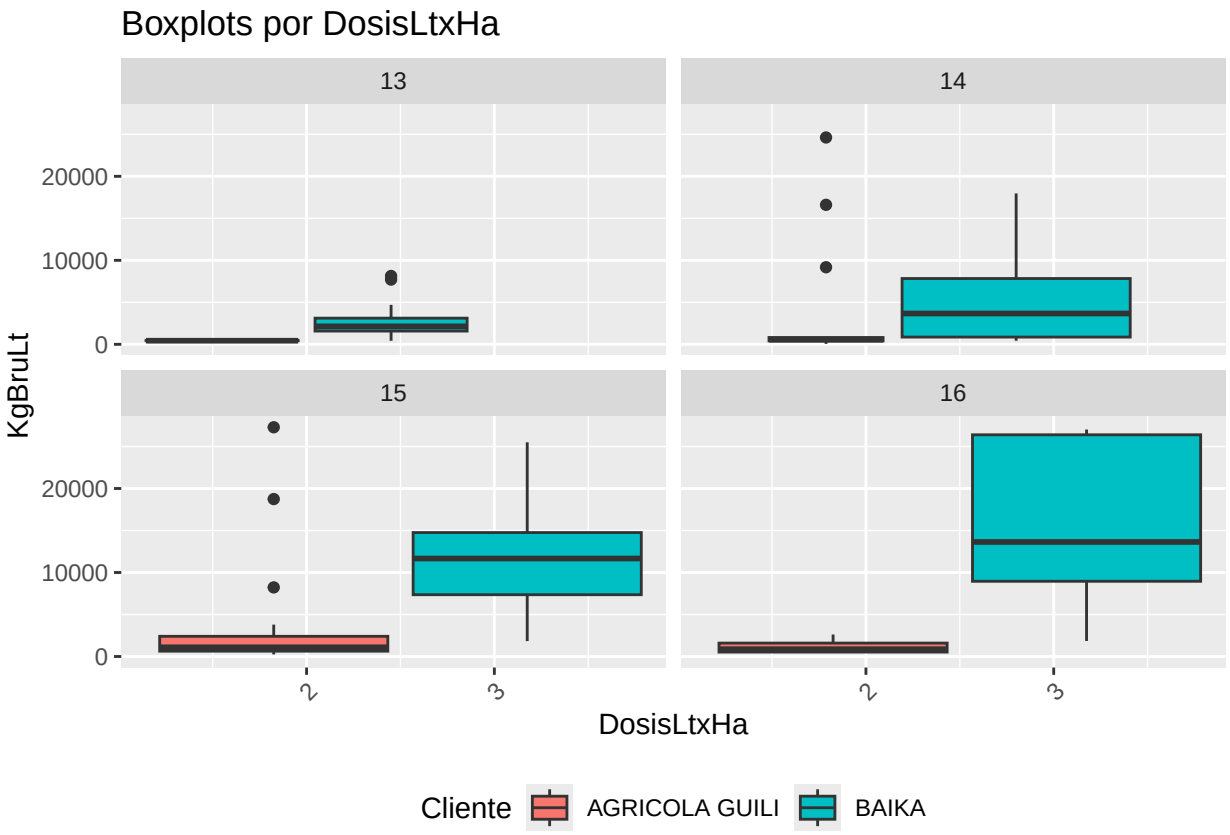
```
#1 Crear los boxplots
boxplot <- ggplot(data, aes(x = Naturaleza, y = KgBruLt, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Semana") +
  facet_wrap(~Semana)  # Cambia "Otra_Variable" por el nombre de la variable que deseas usar para divid
# Establecer la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")
# Mostrar los boxplots
print(boxplot)
```



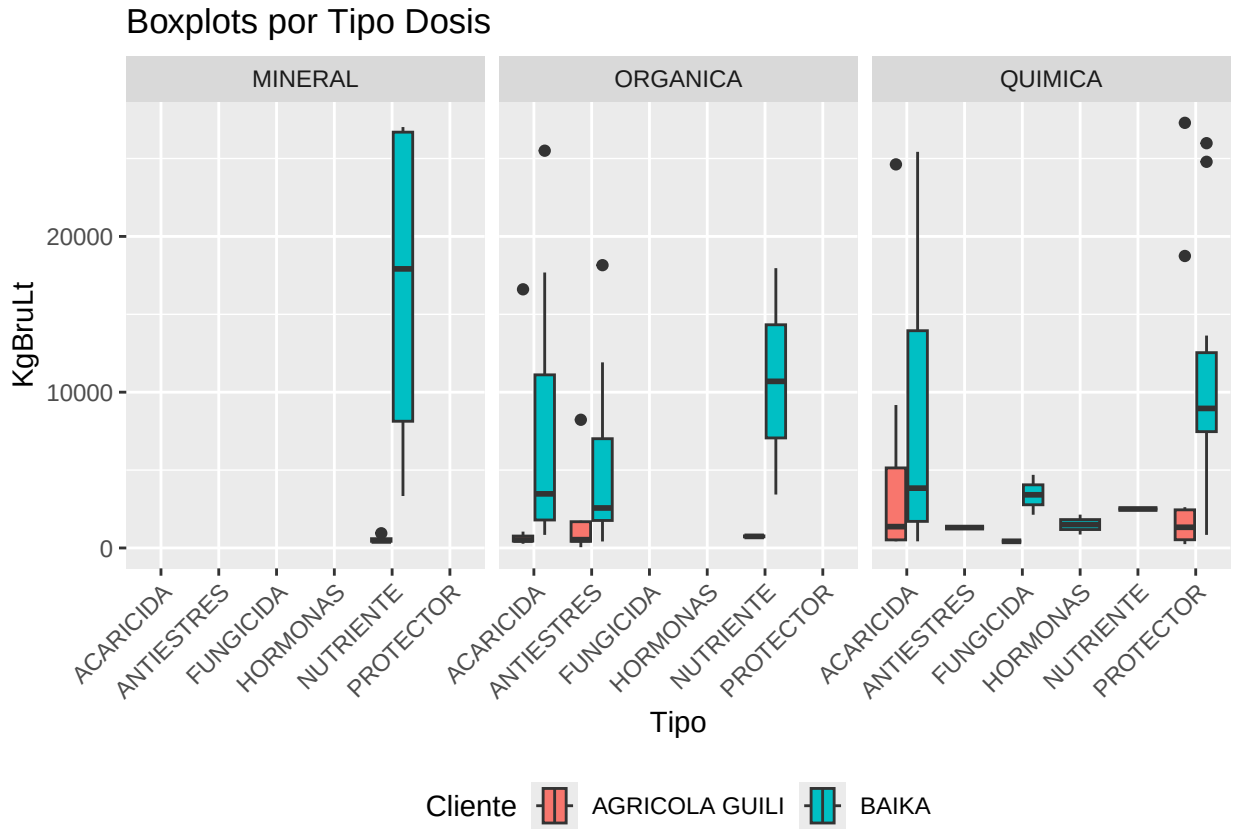

```
#2 Crear los boxplots
options(repr.plot.width=15, repr.plot.height=5) # Tamaño original
boxplot <- ggplot(data, aes(x = Fundo, y = KgBruLt, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Fundo") +
  facet_wrap(~Semana)+# Cambia "Otra_Variable" por el nombre de la variable que deseas usar para dividi
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # Girar los labels del eje x 90 grados
# Establecer la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")
# Mostrar los boxplots
print(boxplot)
```




```
#2.2 Crear los boxplots
options(repr.plot.width=15, repr.plot.height=5) # Tamaño original
boxplot <- ggplot(data, aes(x = DosisLtxHa, y = KgBruLt, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por DosisLtxHa") +
  facet_wrap(~Semana)+# Cambia "Otra_Variable" por el nombre de la variable que deseas usar para dividi
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # Girar los labels del eje x 90 grados
# Establecer la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")
# Mostrar los boxplots
print(boxplot)
```




```
#3 Crear los boxplots
options(repr.plot.width=15, repr.plot.height=5) # Tamaño original
boxplot <- ggplot(data, aes(y = Tipo, x = KgBruLt, fill = Cliente)) +
  geom_boxplot() +
  labs(title = "Boxplots por Tipo Dosis") +
  facet_wrap(~Naturaleza) + # Cambia "Otra_Variable" por el nombre de la variable que deseas usar para
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +  # Girar los labels del eje x 90 grados
  coord_flip() # Esto gira el gráfico para que los boxplots sean horizontales
# Establecer la ubicación de la leyenda
boxplot <- boxplot + theme(legend.position = "bottom")
# Mostrar los boxplots
print(boxplot)
```

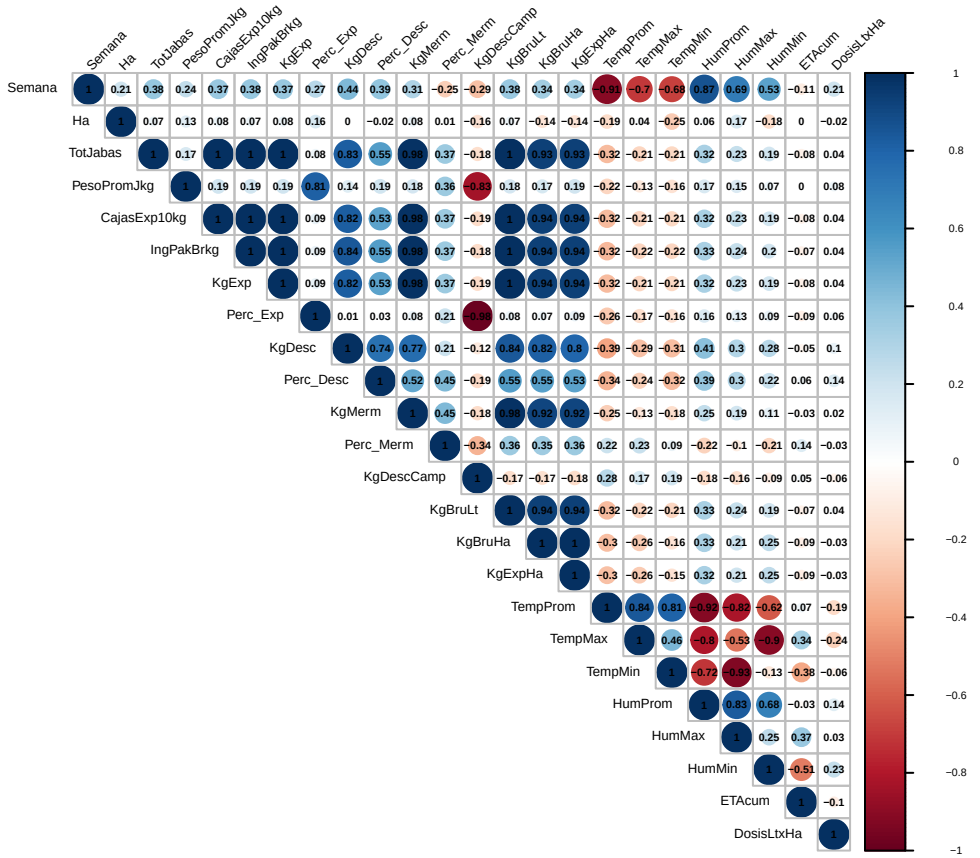



```
# Correlation matrix (if applicable)
# Install the 'corrplot' package if not already installed
if (!require("corrplot")) install.packages("corrplot")
```

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

```
## corrplot 0.92 loaded
```

```
library(corrplot)
correlation_matrix <- cor(data_n, use = "pairwise.complete.obs")
corrplot(correlation_matrix, method = "circle", type = "upper", #order = "hclust",
         tl.col = "black",  # text label color
         tl.srt = 45,       # text label rotation in degrees
         addCoef.col = "black", # color of the correlation coefficients
         number.cex = 0.35,  # size of the correlation coefficients
         cl.cex = 0.3,      # size of the color legend text
         tl.cex = 0.4,
         cl.ratio = 0.2     # ratio of the color legend size
)
```




```
colnames(data)
```

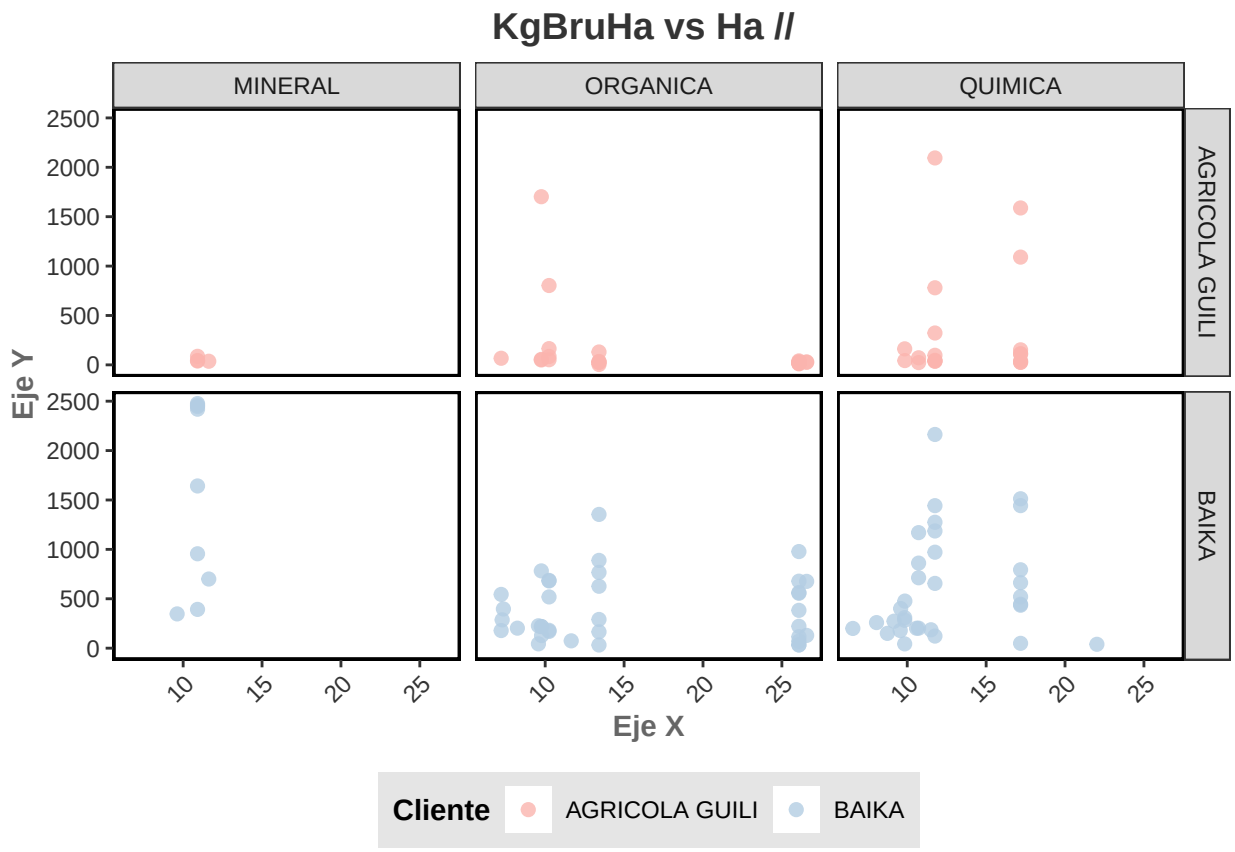
```
##  [1] "Semana"       "Mes"          "Fundo"        "Empresa"      "Lote"
##  [6] "Ha"           "Variedad"     "Color"        "NoGR"         "NoRepProd"
## [11] "Fecha"        "TotJabas"     "PesoPromJkg"  "CajasExp10kg" "IngPakBrkg"
## [16] "KgExp"        "Perc_Exp"     "KgDesc"       "Perc_Desc"    "KgMerm"
## [21] "Perc_Merm"    "KgDescCamp"   "KgBruLt"      "KgBruHa"      "KgExpHa"
## [26] "Status"       "Cliente"      "TempProm"     "TempMax"      "TempMin"
## [31] "HumProm"      "HumMax"       "HumMin"       "ETAcum"       "Producto"
## [36] "Naturaleza"   "DosisLtxHa"   "Tipo"
```

```
colnames(data_n)
```

```
##  [1] "Semana"       "Ha"           "TotJabas"     "PesoPromJkg"  "CajasExp10kg"
##  [6] "IngPakBrkg"   "KgExp"        "Perc_Exp"     "KgDesc"       "Perc_Desc"
## [11] "KgMerm"       "Perc_Merm"    "KgDescCamp"   "KgBruLt"      "KgBruHa"
## [16] "KgExpHa"      "TempProm"     "TempMax"      "TempMin"      "HumProm"
## [21] "HumMax"       "HumMin"       "ETAcum"       "DosisLtxHa"
```

```
##
library(ggplot2)
library(RColorBrewer)
# Definir una paleta de colores pastel
colors <- brewer.pal(n = 8, name = "Pastel1")
# Gráfico ggplot con diseño mejorado y colores pastel
ggplot(data, aes(x = Ha, y = KgBruHa, color = Cliente)) +
  geom_point(size = 2, alpha = 0.8) +  # Ajustar tamaño y transparencia de los puntos
  scale_color_manual(values = colors) +  # Usar colores pastel definidos
```

```
  facet_grid(Cliente ~ Naturaleza) +  # Rejilla de gráficos por Cliente y Naturaleza
  labs(title = "KgBruHa vs Ha // ",
       x = "Eje X",
       y = "Eje Y") +
  theme_bw() +  # Tema con fondo blanco
  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"),  # Girar y ajustar texto del
    axis.text.y = element_text(color = "#333333"),
    legend.title = element_text(face = "bold"),
    legend.background = element_rect(fill = "gray90"),
    legend.position = "bottom",  # Mover la leyenda al fondo
    panel.grid.major = element_blank(),  # Eliminar las líneas de la rejilla principales
    panel.grid.minor = element_blank(),  # Eliminar las líneas de la rejilla secundarias
    panel.border = element_rect(color = "black", fill = NA, linewidth = 1)  # Cambiar 'size' a 'linewid
  )
```




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

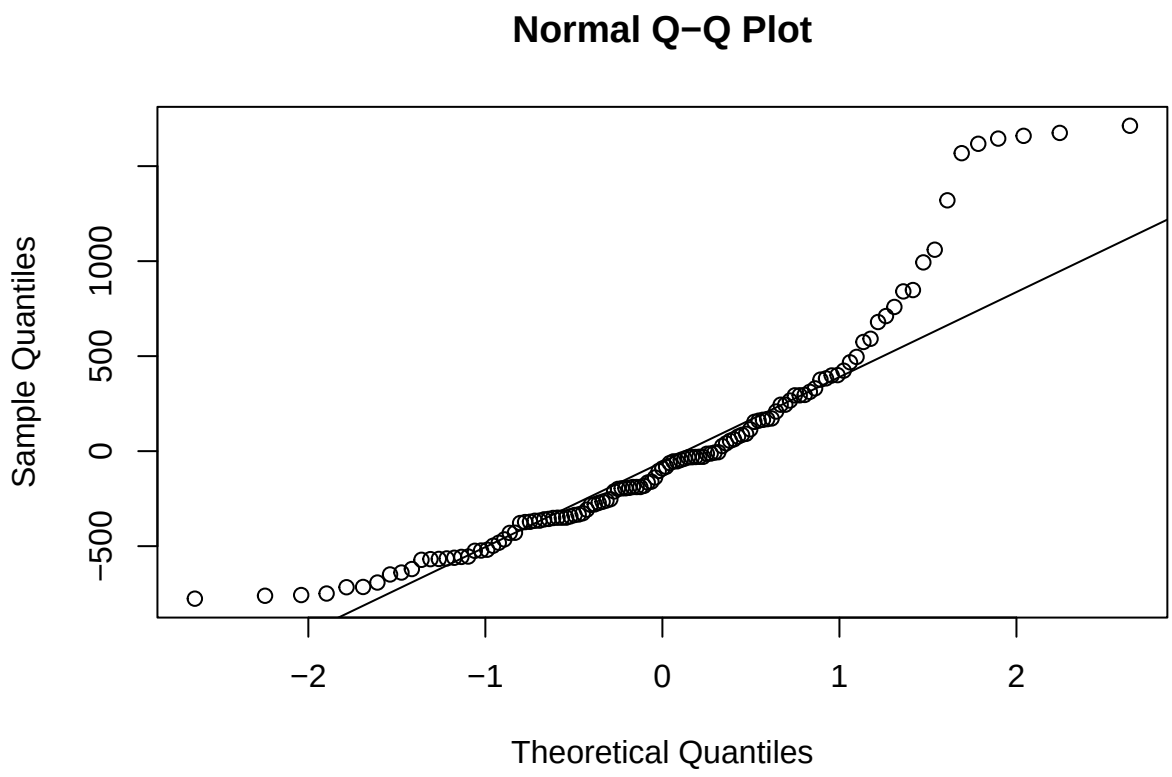
```
##  [1] "Semana"       "Ha"           "TotJabas"     "PesoPromJkg"  "CajasExp10kg"
##  [6] "IngPakBrkg"   "KgExp"        "Perc_Exp"     "KgDesc"       "Perc_Desc"
## [11] "KgMerm"       "Perc_Merm"    "KgDescCamp"   "KgBruLt"      "KgBruHa"
```

```
## [16] "KgExpHa"      "TempProm"     "TempMax"      "TempMin"      "HumProm"
## [21] "HumMax"       "HumMin"       "ETAcum"       "DosisLtxHa"
```

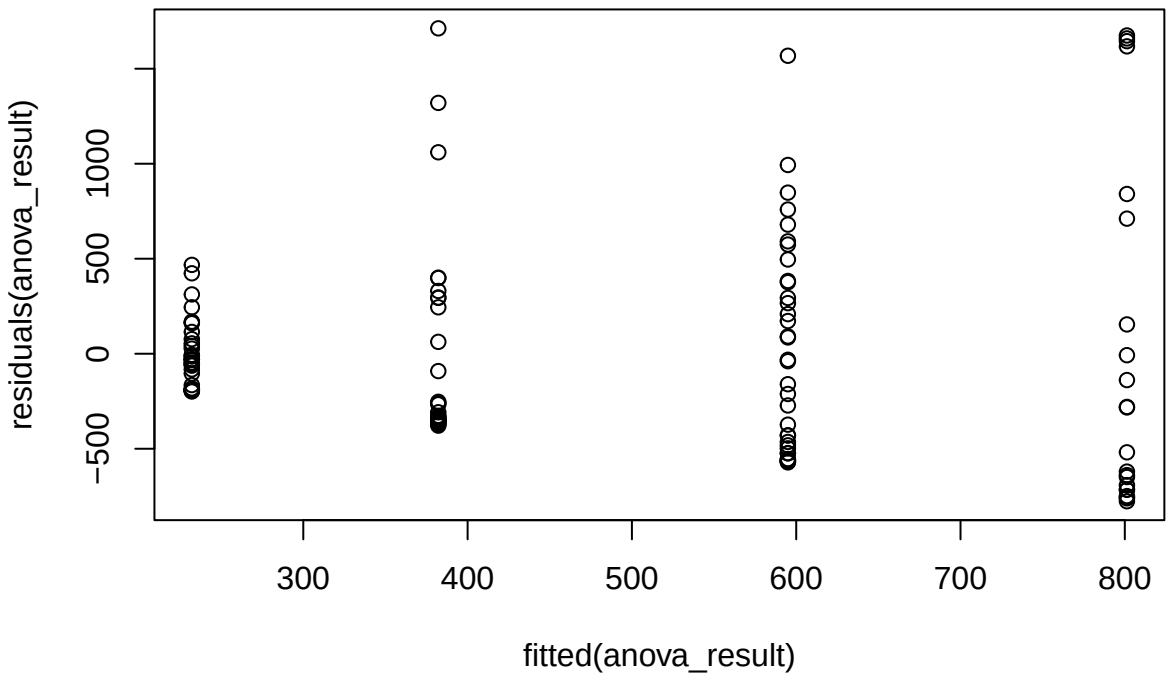
```r
### ANOVA ###
# Ensure that 'Fundo' is a factor and 'KgExpxHa' is numeric
data$Semana <- as.factor(data$Semana)
data$KgBruHa <- as.numeric(data$KgBruHa)
# ANOVA to compare 'KgExpxHa' across different 'Fundo'
anova_result <- aov(KgBruHa ~ Semana, data = data)
# Check the summary of the ANOVA
summary(anova_result)
```

```
##              Df   Sum Sq Mean Sq F value  Pr(>F)
## Semana        3  5112634 1704211   5.259 0.00194 **
## Residuals   117 37914535  324056
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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




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




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

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = KgBruHa ~ Semana, data = data)
##
## $Semana
##           diff          lwr      upr     p adj
## 14-13 150.0050 -225.0308789 525.0409 0.7248701
## 15-13 362.7772    7.9644076 717.5900 0.0430349
## 16-13 569.1345  163.1735800 975.0954 0.0021694
## 15-14 212.7722 -157.4371458 582.9816 0.4420135
## 16-14 419.1295   -0.3548395 838.6138 0.0502803
## 16-15 206.3573 -195.1490772 607.8636 0.5397828
```

```
## ANOVA #2
names(data)
```

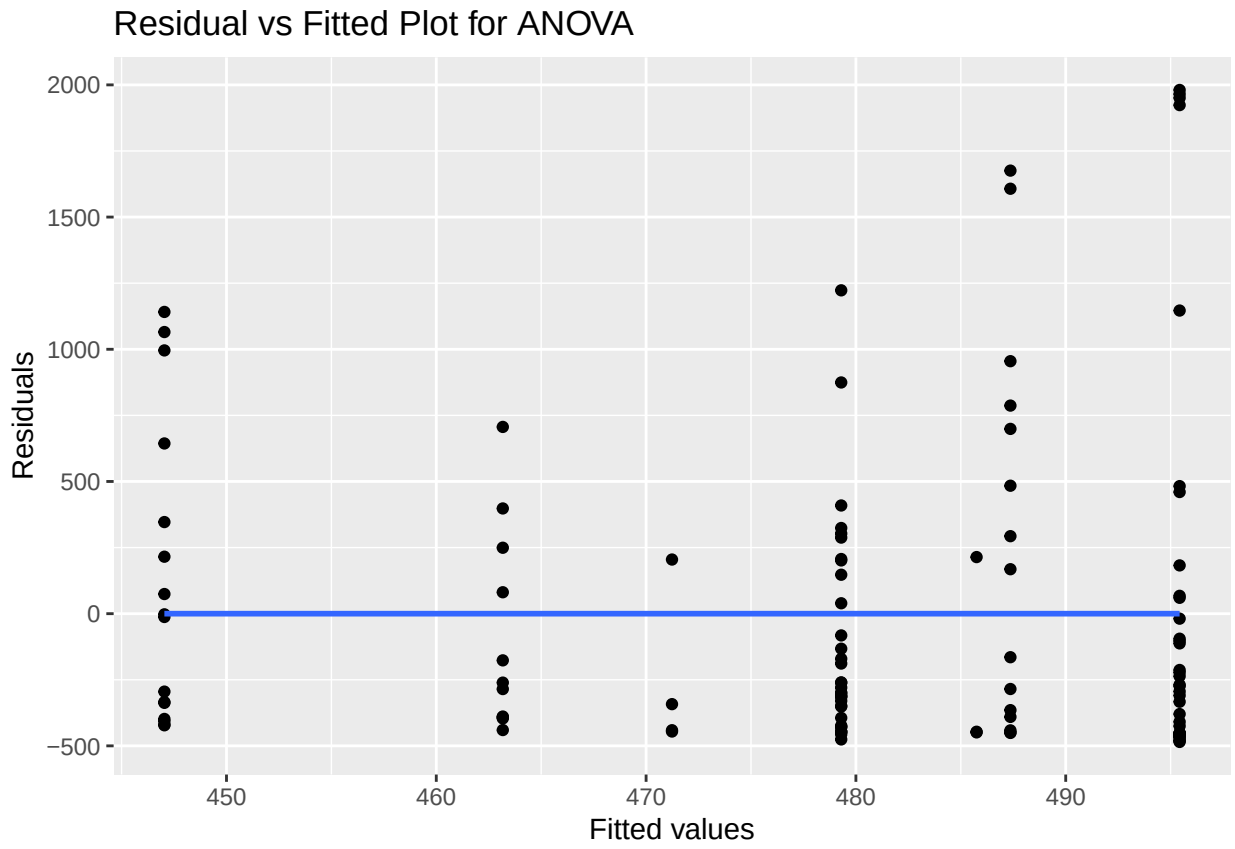
```
##  [1] "Semana"       "Mes"          "Fundo"        "Empresa"      "Lote"
##  [6] "Ha"           "Variedad"     "Color"        "NoGR"         "NoRepProd"
## [11] "Fecha"        "TotJabas"     "PesoPromJkg"  "CajasExp10kg" "IngPakBrkg"
## [16] "KgExp"        "Perc_Exp"     "KgDesc"       "Perc_Desc"    "KgMerm"
## [21] "Perc_Merm"    "KgDescCamp"   "KgBruLt"      "KgBruHa"      "KgExpHa"
## [26] "Status"       "Cliente"      "TempProm"     "TempMax"      "TempMin"
## [31] "HumProm"      "HumMax"       "HumMin"       "ETAcum"       "Producto"
```

```
## [36] "Naturaleza"   "DosisLtxHa"   "Tipo"
```

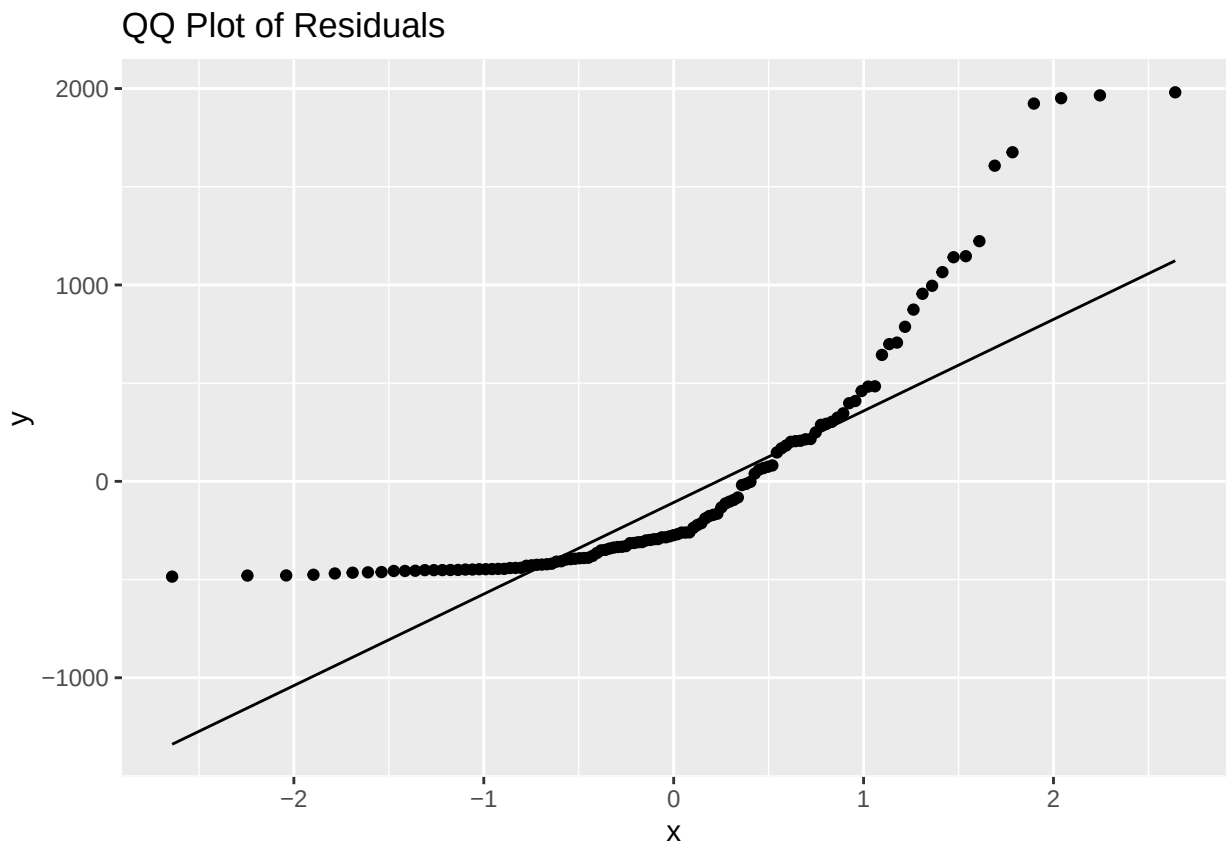
```r
# Performing the ANOVA
anova_result <- aov(KgBruHa ~ DosisLtxHa, data = data)

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

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




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



Notas:

1. **Packages**: Ensure you have the necessary R packages (`readxl`, `ggplot2`, `dplyr`) installed. If not, you can install them using `install.packages()`.

This markdown script provides a structure for conducting an EDA with visualizations, and you can further expand upon this as needed.

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.