

Gráficos Básicos en R

Módulo 1 - Exploración de datos mediante gráficas

Paula Reyes
AGROSAVIA

¿Qué hacen ustedes cuando tienen varias muestras de una misma variable y quieren ver la distribución?

1. Excel
2. SAS (licencia)

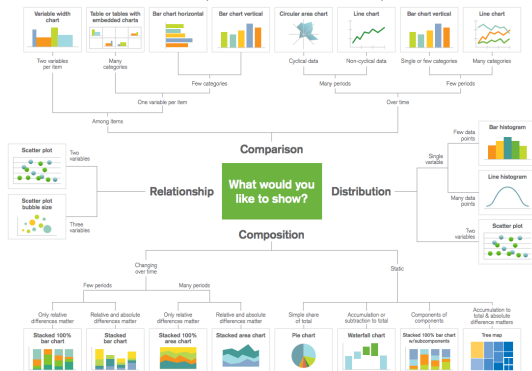
Usaremos un conjunto de datos de R *PlantGrowth* resultados de un experimento de pesos de plantas en tres condiciones. Inicialmente queremos ver habilidades iniciales de los participantes del curso.

Conjunto de datos [PlantGrowth](#)

De este conjunto de datos obtener gráficas de distribución de datos. Enviar figuras al correo academia@agrosavia.co. Figura primer día - Nombre.

Tipos de diagramas - ¿Cómo seleccionar?

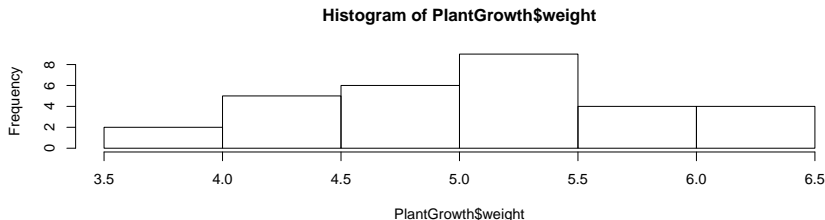
Cómo escogemos cada uno el tipo de gráfica a usar? Para poder convertir datos en conocimiento es importante poder presentar de la manera correcta (clara y concisa)



BigData is the New Black

En un histograma los datos están agrupados en rangos (por ejemplo 0-0.5, 0.5-1). Se grafican los rangos como barras continuas, cada barra representa un rango.

- El ancho de la barra es proporcional al ancho del rango.
- La altura de la barra es proporcional a la cantidad de datos en ese rango (frecuencia)



Es una representación gráfica de la distribución de los datos. Permite observar tendencias (eg. donde están agrupados la mayor parte de los datos).

Usaremos un conjunto de datos *PlantGrowth* resultado de un experimento de pesos de plantas en tres condiciones (control - ctrl, tratamiento1 -trt1, tratamiento2 -trt2)

```
str(PlantGrowth)
```

```
## 'data.frame':    30 obs. of  2 variables:  
## $ weight: num  4.17 5.58 5.18 6.11 4.5 4.61 5.17 4.53 5.33 5.14 ...  
## $ group : Factor w/ 3 levels "ctrl","trt1",...: 1 1 1 1 1 1 1 1 1 1 ...
```

Usaremos un conjunto de datos *PlantGrowth* resultados de un experimento de pesos de plantas en tres condiciones (control - ctrl, tratamiento1 -trt1, tratamiento2 -trt2)

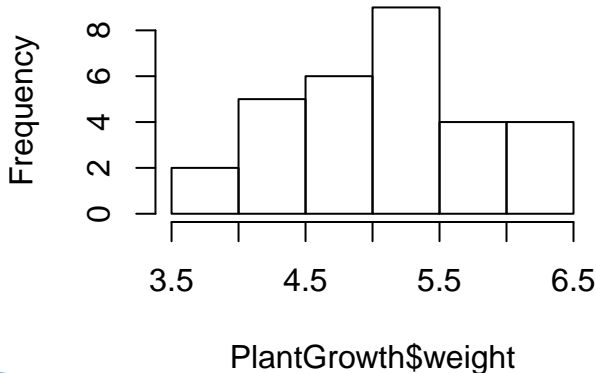
```
#PlantGrowth  
head(PlantGrowth)
```

```
##   weight group  
## 1    4.17  ctrl  
## 2    5.58  ctrl  
## 3    5.18  ctrl  
## 4    6.11  ctrl  
## 5    4.50  ctrl  
## 6    4.61  ctrl
```



```
hist(PlantGrowth$weight)
```

Histogram of PlantGrowth\$weight

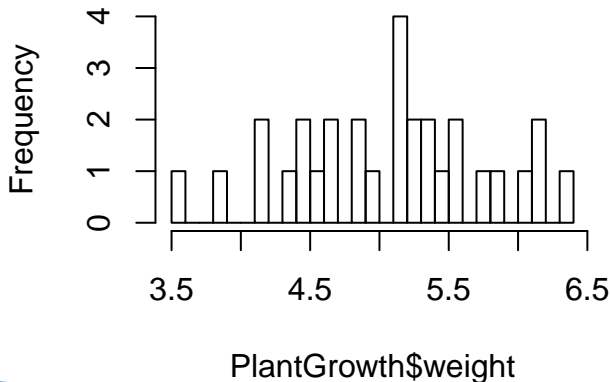


```
str(hist(PlantGrowth$weight,plot=FALSE))
```

```
## List of 6
## $ breaks : num [1:7] 3.5 4 4.5 5 5.5 6 6.5
## $ counts : int [1:6] 2 5 6 9 4 4
## $ density : num [1:6] 0.133 0.333 0.4 0.6 0.267 ...
## $ mids : num [1:6] 3.75 4.25 4.75 5.25 5.75 6.25
## $ xname : chr "PlantGrowth$weight"
## $ equidist: logi TRUE
## - attr(*, "class")= chr "histogram"
```

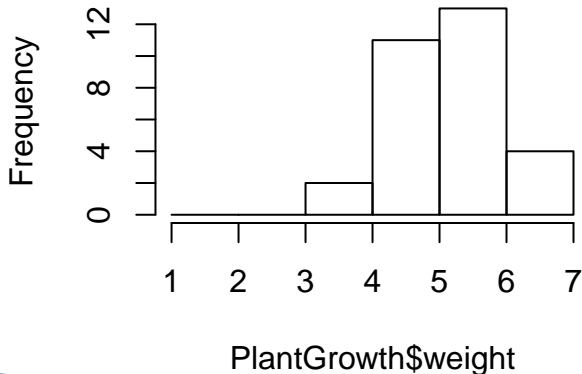
```
hist(PlantGrowth$weight,breaks=20)
```

Histogram of PlantGrowth\$weight



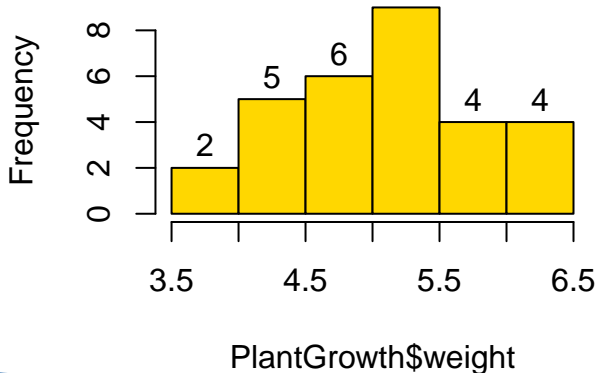
```
hist(PlantGrowth$weight,breaks=c(1:7))
```

Histogram of PlantGrowth\$weight



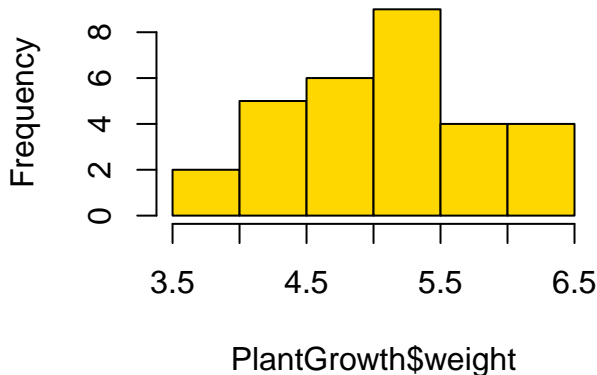
```
hist(PlantGrowth$weight,col=c("gold"),labels=TRUE)
```

Histogram of PlantGrowth\$weight



```
hist(PlantGrowth$weight,col=c("gold"))
```

Histogram of PlantGrowth\$weight



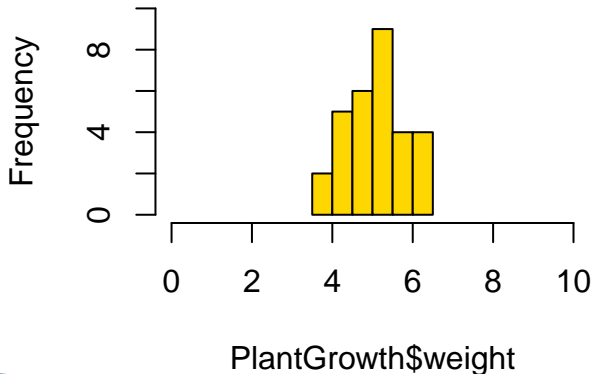
```
hist(PlantGrowth$weight,col=c("gold"),border="orange")
```

Histogram of PlantGrowth\$weight



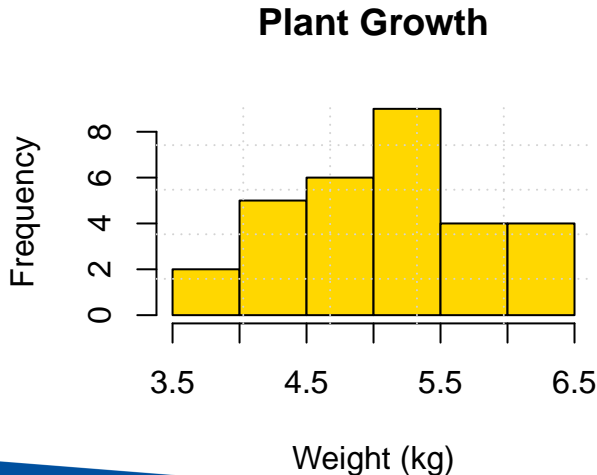
```
hist(PlantGrowth$weight,col=c("gold"),xlim=c(0,10),ylim=c(0,10))
```

Histogram of PlantGrowth\$weight



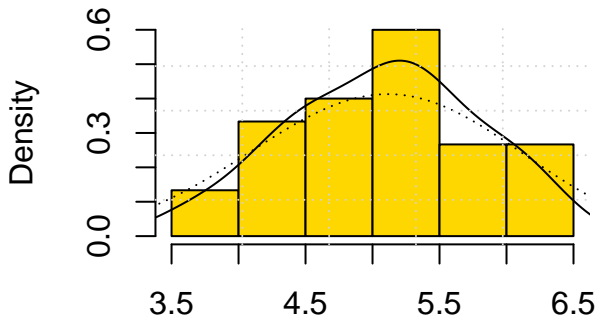
El histograma de frecuencias - título (gráfico y eje x)

```
hist(PlantGrowth$weight,col=c("gold"),main="Plant Growth",xlab="Weight (kg)")  
grid(5)
```



```
hist(PlantGrowth$weight,col=c("gold"),prob=TRUE)  
lines(density(PlantGrowth$weight))  
lines(density(PlantGrowth$weight,adjust = 2),lty="dotted")  
grid(5)
```

Histogram of PlantGrowth\$weight



PlantGrowth\$weight

```
levels(PlantGrowth$group)
```

```
## [1] "ctrl" "trt1" "trt2"
```

```
levels(PlantGrowth$group)[1]
```

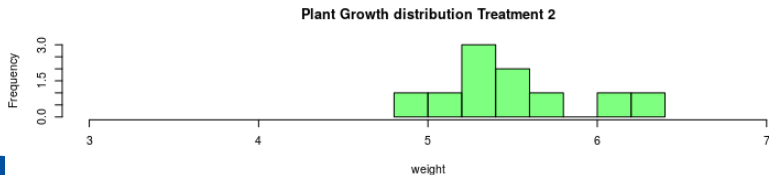
```
## [1] "ctrl"
```

```
PlantGrowthCtrl<-PlantGrowth[PlantGrowth$group==levels(PlantGrowth$group)[1],]  
PlantGrowthTrt1<-PlantGrowth[PlantGrowth$group==levels(PlantGrowth$group)[2],]  
PlantGrowthTrt2<-PlantGrowth[PlantGrowth$group==levels(PlantGrowth$group)[3],]
```

Histogramas por tratamiento

```
par(c(3,1))  
hist(PlantGrowthCtrl$weight, xlim=c(3,7), col=rgb(1,0,0,0.4),  
     xlab="weight", main="Plant Growth distribution Control")  
hist(PlantGrowthTrt1$weight, xlim=c(3,7), col=rgb(0,0,1,0.4),  
     xlab="weight", main="Plant Growth distribution Treatment 1")  
hist(PlantGrowthTrt2$weight, xlim=c(3,7), col=rgb(0,1,0,0.4),  
     xlab="weight", main="Plant Growth distribution Treatment 2")
```

Histogramas por tratamiento

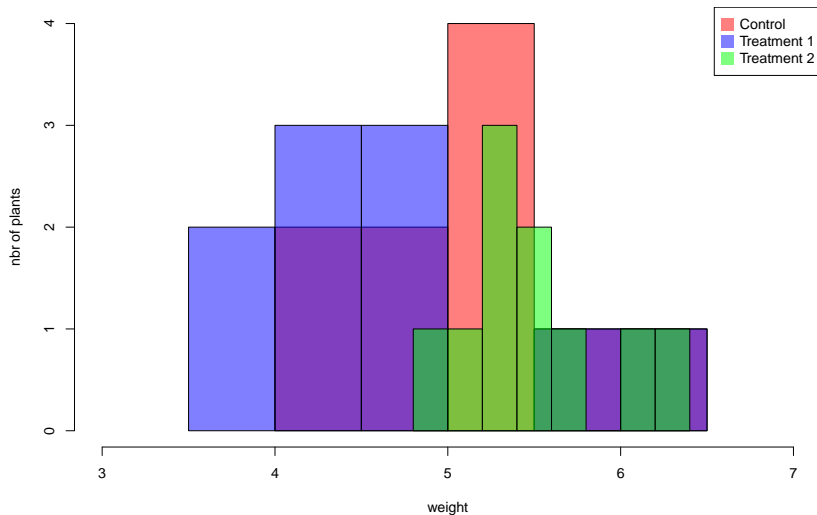


Histogramas por tratamiento

```
hist(PlantGrowthCtrl$weight, xlim=c(3,7), col=rgb(1,0,0,0.5),  
     xlab="weight",ylab="nbr of plants", main="Plant Growth distribution" )  
hist(PlantGrowthTrt1$weight, xlim=c(3,7), col=rgb(0,0,1,0.5), add=T)  
hist(PlantGrowthTrt2$weight, xlim=c(3,7), col=rgb(0,1,0,0.5), add=T)  
legend("topright", legend=c("Control","Treatment 1","Treatment 2"),  
       col=c(rgb(1,0,0,0.5),rgb(0,0,1,0.5),rgb(0,1,0,0.5)), pt.cex=2, pch=15 )
```

Histogramas por tratamiento

Plant Growth distribution

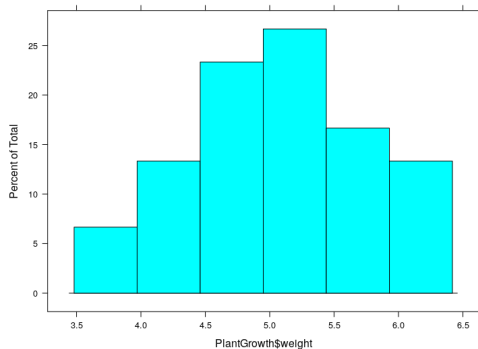


A continuación hay ejemplos de otras librerías. Sin embargo, es necesario instalar y cargar los paquetes.

- lattice
- ggplot2
- plotly

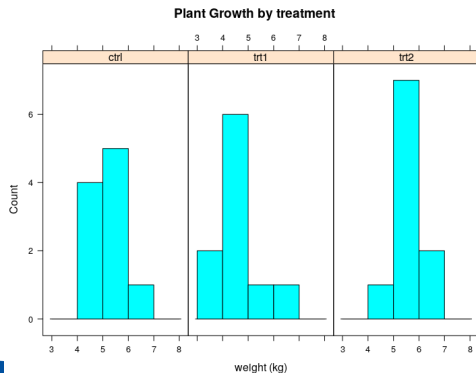
Histograma - libreria lattice

```
# instala libreria  
install.packages("lattice")  
# carga libreria  
library(lattice)  
histogram(~weight, data=PlantGrowth)
```



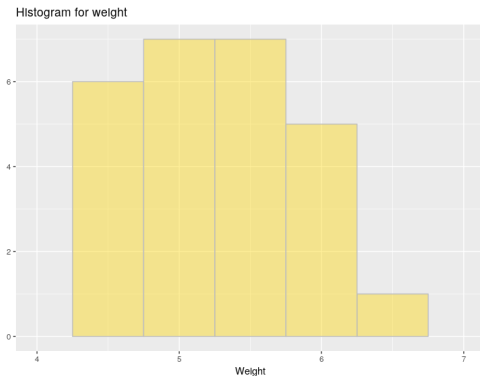
Histograma - libreria lattice divido por grupo

```
# instala libreria
install.packages("lattice")
# carga libreria
library(lattice)
histogram(~weight|group,data=PlantGrowth,type="count",xlab="weight (kg)",
          main="Plant Growth by treatment",breaks = 3:8,layout=c(3,1))
```



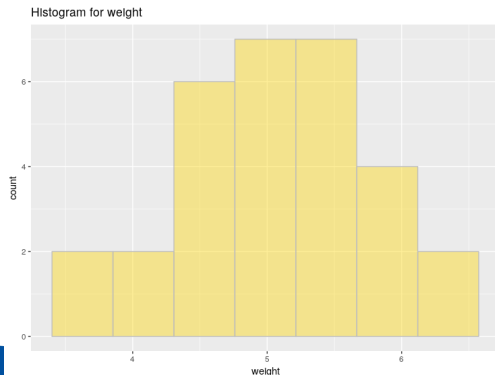
Histograma - libreria ggplot2 - qplot()

```
# instala libreria
install.packages("ggplot2")
# carga libreria
library(ggplot2)
qplot(PlantGrowth$weight, geom="histogram", bins=7, main="Histogram for weight",
      xlab="Weight", fill=I("gold"), col=I("gray"), alpha=I(.4), xlim=c(4,7))
```



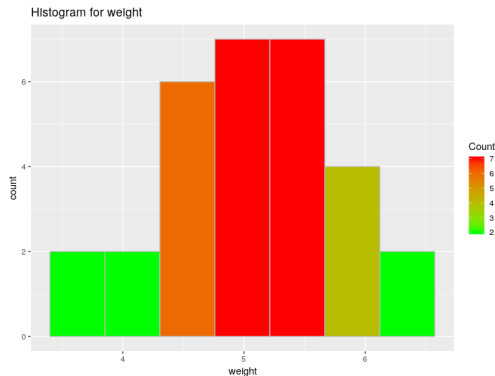
Histograma - libreria ggplot2 - ggplot()

```
# instala libreria
install.packages("ggplot2")
# carga libreria
library(ggplot2)
ggplot(data=PlantGrowth,aes(PlantGrowth$weight))+
  geom_histogram(bins = 7,col="gray",fill="gold",alpha=0.4)+
  labs(title = "Histogram for weight",x="weight",y="count",xlim(c(4,7)))
```



Histograma - libreria ggplot2 - forma 2 - rango color

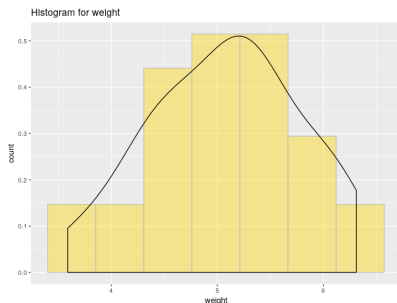
```
ggplot(data=PlantGrowth,aes(PlantGrowth$weight))+  
geom_histogram(bins = 7,col="gray",aes(fill=..count..))+  
labs(title = "Histogram for weight",x="weight",y="count",xlim(c(4,7)))+  
scale_fill_gradient("Count",low="green",high = "red")
```



Histograma - libreria ggplot2 qplot() vs ggplot()

qplot() tiene una sintaxis más sencilla con respecto a ggplot(). qplot viene de “quick plot”. Sin embargo, ggplot() es más flexible y permite configurar otros parámetros. Generalmente se usa ggplot()

```
ggplot(data=PlantGrowth,aes(PlantGrowth$weight))+  
geom_histogram(bins = 7,col="gray",fill="gold",alpha=0.4,aes(y=..density..))+  
labs(title = "Histogram for weight",x="weight",y="count",xlim(c(4,7)))+  
geom_density(col=1)
```

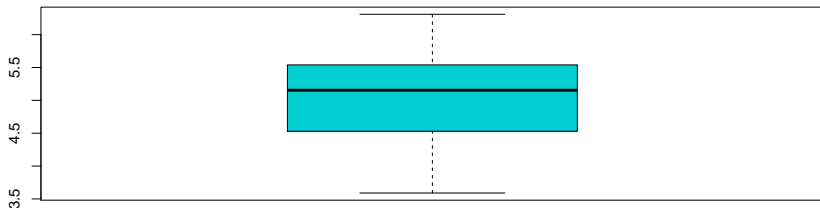


```
install.packages("plotly")  
library(plotly)  
#configurar usuario y key  
#Sys.setenv("plotly_username"=usuario)  
#Sys.setenv("plotly_api_key"=keygenerada)  
p<-plot_ly(x=PlantGrowth$weight,type="histogram")  
char_link =api_create(p,filename = "histogram_basic")  
char_link
```

<https://plot.ly/~phreyes/1#plot>

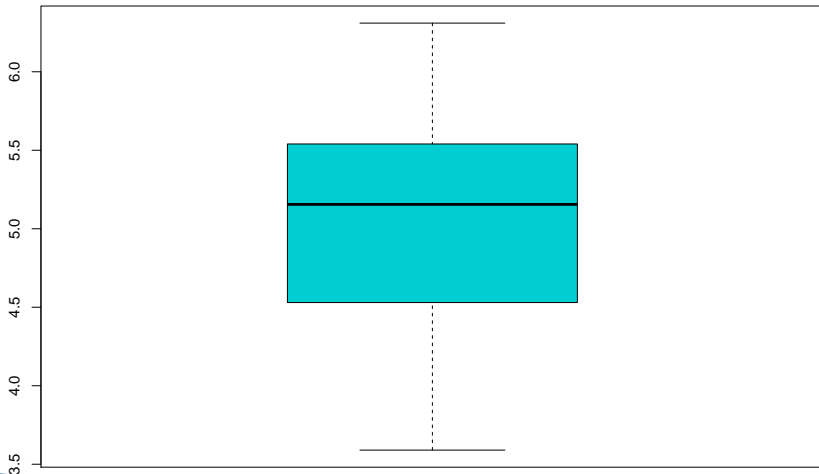
Es uno de los gráficos más comunes. Da un resumen bueno de una o varias variables numéricas.

- La línea que divide la caja en la mitad representa la mediana.
- Los finales de la caja representan el primer Q_1 y el tercer cuartil Q_3 .
- Las líneas externas representan el valor más alto y más bajo (excluyendo valores atípicos)



Diagramas de caja - Boxplot

```
boxplot(PlantGrowth$weight,col = "darkturquoise")
```



```
str(boxplot(PlantGrowth$weight,plot = FALSE))
```

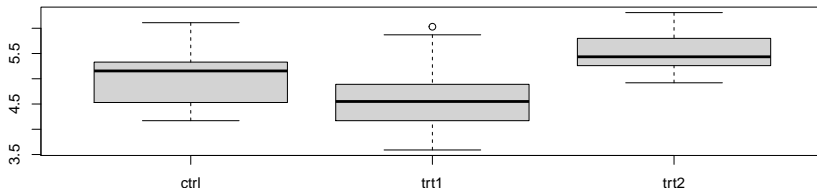
```
## List of 6
## $ stats: num [1:5, 1] 3.59 4.53 5.15 5.54 6.31
## $ n : num 30
## $ conf : num [1:2, 1] 4.86 5.45
## $ out : num(0)
## $ group: num(0)
## $ names: chr "1"
```

```
quantile(PlantGrowth$weight,c(0,0.25,0.5,0.75,1))
```

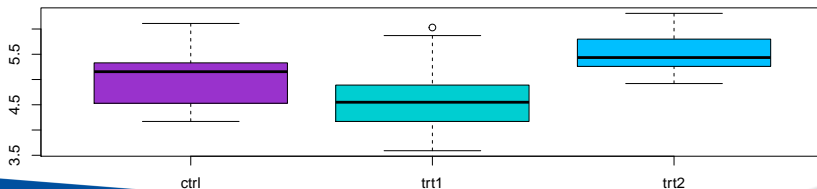
```
##      0%    25%    50%    75%   100%
## 3.590 4.550 5.155 5.530 6.310
```

Diagramas de caja - Boxplot dividiendo por grupo

```
boxplot(weight ~ group, data = PlantGrowth, col = "lightgray")
```



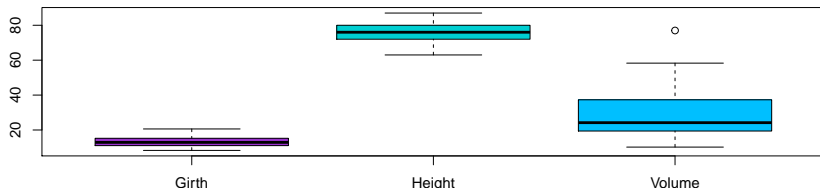
```
boxplot(weight ~ group, data = PlantGrowth, col=c("darkorchid", "darkturquoise",  
"deepskyblue"))
```



Diagramas de caja - Boxplot para un dataframe

Otro conjunto de datos *trees*. Medidas del diámetro, altura y volumen de 31 arboles de mora.

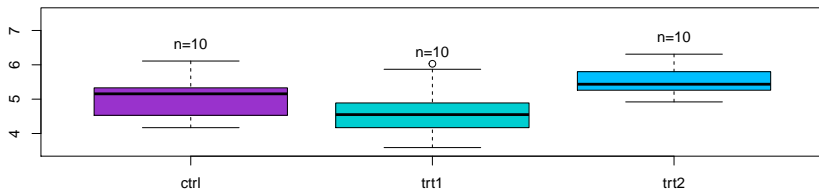
```
boxplot(trees,col=c("darkorchid","darkturquoise","deepskyblue"))
```



En caso de usar una matriz se realiza el boxplot por columnas.
El diámetro esta etiquetado incorrectamente como Girth

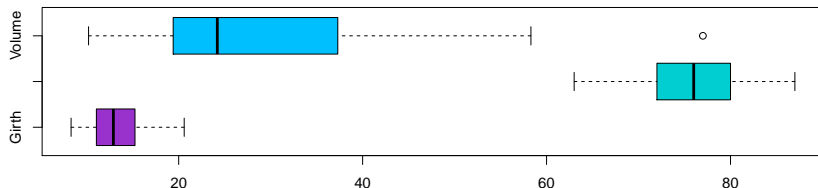
Diagramas de caja - Boxplot número de muestras

```
a=boxplot(PlantGrowth$weight~PlantGrowth$group,  
          col=c("darkorchid","darkturquoise","deepskyblue"),  
          ylim=c(3.5,7.5))  
text(c(1:nlevels(PlantGrowth$group)),a$stats[nrow(a$stats),]+0.5,  
     paste("n=",a$n[1:nlevels(PlantGrowth$group)],sep=""))
```



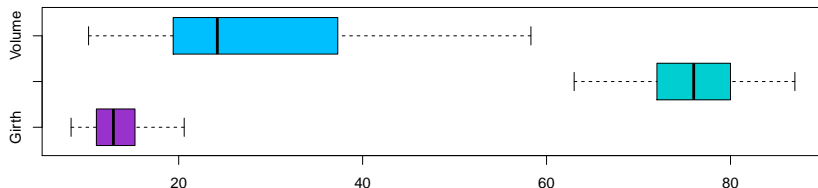
Diagramas de caja - Horizontal

```
boxplot(trees, horizontal=TRUE, col=c("darkorchid", "darkturquoise",  
                                     "deepskyblue"))
```



Diagramas de caja - Horizontal

```
boxplot(trees, horizontal=TRUE, col=c("darkorchid", "darkturquoise",  
                                     "deepskyblue"), outline = FALSE)
```



Diagramas de caja - 2 variables dividiendo por grupo

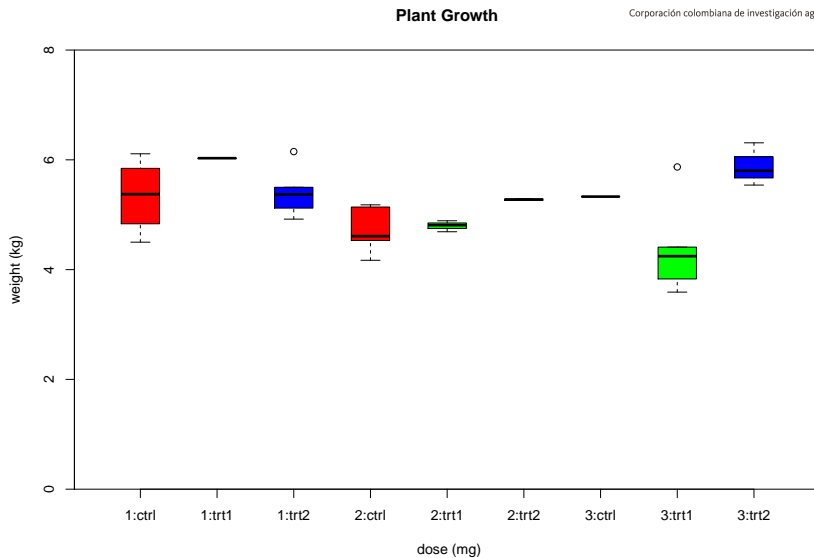
```
# Creemos columna dosis para ver otro ejemplo tratamientos y dosis  
# Números aleatorios para la dosis por esto la gráfica va a cambiar  
# para cada uno
```

```
PlantGrowth$dose=sample(1:3,30,replace=TRUE)
```

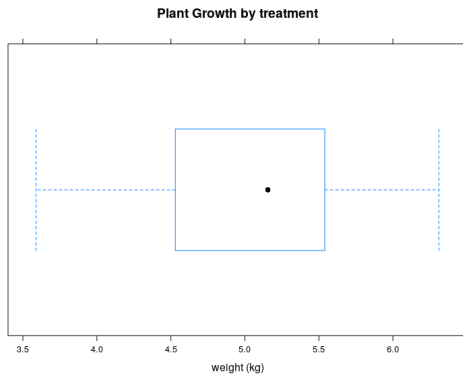
```
# Boxplot dividiendo por grupo
```

```
boxplot(weight ~ dose:group, data = PlantGrowth,  
boxwex = 0.5, col = c("red", "green","blue"),  
main = "Plant Growth",  
xlab = " dose (mg)", ylab = "weight (kg)",  
sep = ":", lex.order = TRUE, ylim = c(0, 8), yaxs = "i")
```

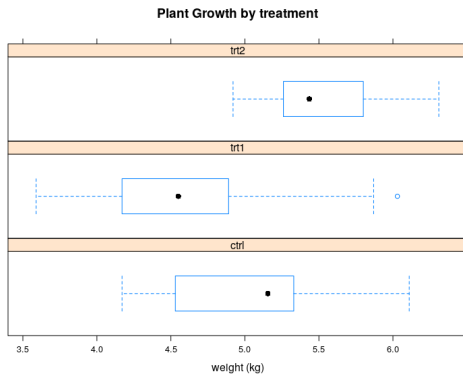

Diagramas de caja - agregar



```
bwplot(~weight, data=PlantGrowth,xlab="weight (kg)",  
main="Plant Growth by treatment")
```

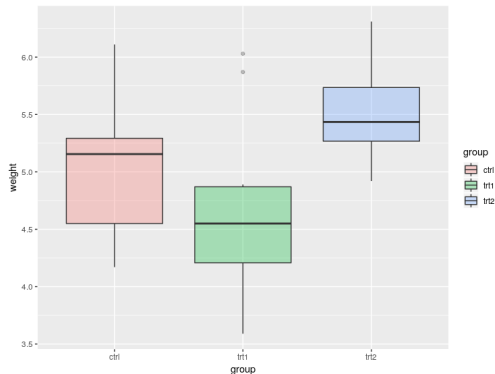


```
bwplot(~weight | group, data=PlantGrowth, xlab="weight (kg)",  
main="Plant Growth by treatment", layout=c(3,1))
```



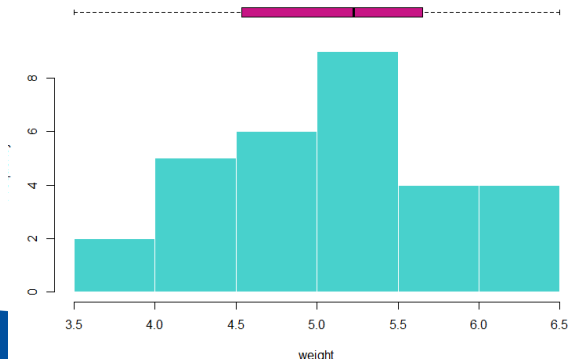
Diagramas de caja - libreria ggplot2

```
ggplot(PlantGrowth,aes(x=group,y=weight,fill=group))+geom_boxplot(alpha=0.3)
```



Histograma y Diagramas de caja - libreria ggplot2

```
#Dividir la pantalla
layout(mat=matrix(c(1,2),2,1,byrow=TRUE),height=c(1,8))
#Dibujar el boxplot y el histograma
par(mar=c(0,3.1,1.1,2.1))
boxplot(PlantGrowth$weight,horizontal = TRUE,xaxt="n",
        col="mediumvioletred",frame=F)
par(mar=c(4,3.1,1.1,2.1))
hist(PlantGrowth$weight,col="mediumturquoise",border=F,main="",xlab="weight")
```



```
p<-plot_ly(PlantGrowth,y=~weight,color=~group,type="box")  
chart_link = api_create(p, filename="box-multiple")  
chart_link
```

<https://plot.ly/~phreyes/3/>

- Inicialmente trabajaremos con un conjunto de datos predefinido en R.
- Si queda tiempo la idea es trabajar con un conjunto de datos de cada uno.

Conjunto de datos *ToothGrowth*. Este conjunto de datos consiste en la longitud de los dientes de cuyes. Se dividieron en grupos al los que se les dieron diferentes dosis de vitamina C (0.5, 1, y 2 mg) mediante dos métodos diferentes (jugo de naranja OJ o ácido ascórbico VC).

- Probar con este nuevo conjunto de datos a realizar las gráficas que se encuentran en las diapositivas. De histogramas y diagramas de caja.

- help r
- [The R Graph Gallery](#)

