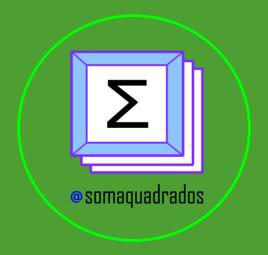
# Introducción al análisis de datos biológicos con R

Clase 4



Eliana F. Burgos

# Clase 4 CONTENIDO



- 1. Estadística descriptiva: frecuencias
- 2. Gráficos

# Frecuencias y distribución de frecuencias



- es la base de la inferencia estadística
- representa la cantidad de veces que un determinado valor se repite en nuestra muestra
- organiza la información disponible para describir los datos

lista de valores + frecuencias => distribución de frecuencias



# RETOMAMOS EL EJERCICIO 3 DE LA CLASE Σ ANTERIOR © somaquadrados

En el marco de la evaluación del estado de conservaciónn de mamíferos de la provincia, nos designaron la especie *Chironectes minimus* mejor conocida como **cuica de agua**, propia de la Selva Paranaense y que, hasta el momento, en nuestro país solo se encuentra en Misiones. Sabemos que en otras regiones su peso es menor en áreas degradadas respecto de las áreas conservadas. Realizamos la captura de esta especie (con ayuda de un veterinario, las medidas de bioseguridad adecuadas y los permisos correspondientes), donde registramos el peso y el sexo de los individuos en dos sitios completamente distintos: un área natural protegida y un cultivo de yerba con arroyos y tajamares.



## FRECUENCIAS ABSOLUTAS Y RELATIVAS



Frecuencia absoluta -> es el número de veces que se repite un valor

**Frecuencia relativa** -> la proporción que representa esa frecuencia absoluta en relación con el total.

**Frecuencias acumuladas** -> suma sucesiva de las frecuencias de los datos ordenados de menor a mayor,

#### **Ejemplo**

## 0.4871795 0.5128205

```
f.abs <- table(datos$uso_suelo)
f.abs

##
## cultivo reserva
## 19 20

prop.table(f.abs)

##
## cultivo reserva</pre>
```

### si quiero expresarla en porcentaje uso prop.table()\*100



```
##
## cultivo reserva
## 0.4871795 0.5128205

prop.table(f.abs)*100

##
## cultivo reserva
## 48.71795 51.28205
```



### **Ejemplo**

```
##
## 2014 2015
    30 9
##
##
##
   2014 2015
## 0.7692308 0.2307692
##
## h m
## 18 21
##
##
  h
                 m
## 0.4615385 0.5384615
```









• es la especificaciónn de las frecuencias correspondientes a cada uno de sus valores

Variable	Frec. absoluta	Frec. relativa
USO DEL SUELO		
reserva	20	0.5128205
cultivo	19	0.4871795
AÑO		
2014	30	0.7692308
2015	9	0.2307692
SEXO		
m	21	0.5384615
h	18	0.4615385



#### **Variables cuantitativas**



podemos calcular las frecuencias para cada valor

```
##
## 467 497 525 545 556 573 586 598 603 617 625 634 642 653 661 681 693 714
## 1 1 3 3 3 1 6 1 1 2 2 2 1 3
##
##
         467
                   497
                              525
                                        545
                                                   556
                                                             573
                                                                       586
## 0.02564103 0.02564103 0.07692308 0.07692308 0.07692308 0.02564103 0.15384615
##
         598
                   603
                              617
                                        625
                                                  634
                                                             642
                                                                       653
## 0.02564103 0.02564103 0.05128205 0.05128205 0.05128205 0.02564103 0.07692308
                   681
##
         661
                              693
                                        714
## 0.02564103 0.07692308 0.07692308 0.05128205
```



#### • Tabla de distribución de frecuencias

```
Somaquadrados
```

```
library(fdth)
tabla <- fdt(datos$peso, start=460, end=715, h=15)
tabla</pre>
```

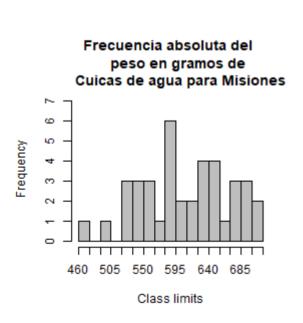
```
## Class limits f rf rf(%) cf cf(%)
       [460,475) 1 0.03 2.56 1 2.56
##
       [475,490) 0 0.00 0.00 1 2.56
##
##
      [490,505) 1 0.03 2.56 2 5.13
      [505,520) 0 0.00 0.00 2 5.13
##
##
      [520,535) 3 0.08 7.69 5 12.82
       [535,550) 3 0.08 7.69 8
##
                                20.51
##
       [550,565) 3 0.08 7.69 11 28.21
       [565,580) 1 0.03 2.56 12
##
                                30.77
##
       [580,595) 6 0.15 15.38 18 46.15
       [595,610) 2 0.05 5.13 20
##
                                51.28
##
       [610,625) 2 0.05 5.13 22
                                56.41
       [625,640) 4 0.10 10.26 26
                                66.67
##
##
       [640,655) 4 0.10 10.26 30 76.92
       [655,670) 1 0.03 2.56 31 79.49
##
##
       [670,685) 3 0.08 7.69 34 87.18
##
      [685,700) 3 0.08 7.69 37 94.87
##
       [700,715) 2 0.05 5.13 39 100.00
```

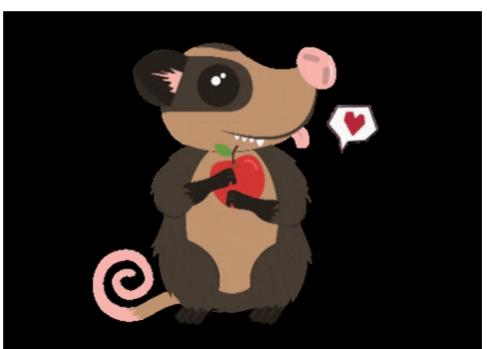
## Frecuencias relativas

Somaquadrados

• **Graficar**: Histograma

```
plot(tabla, type="fh",
    main="Frecuencia absoluta del
    peso en gramos de
    Cuicas de agua para Misiones")
```



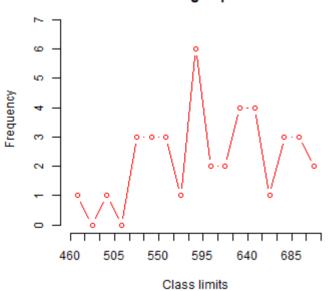


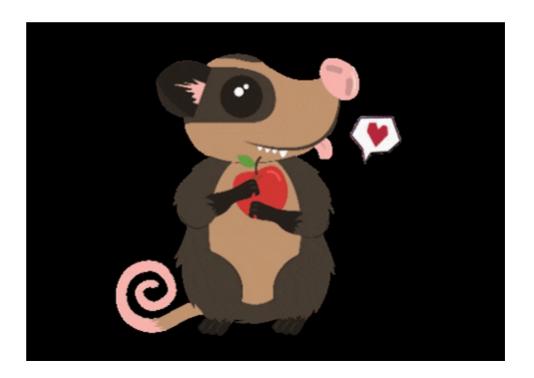
• **Graficar**: Polígono de frecuencias



```
plot(tabla, type="fp", col="red",
    main="Frecuencia absoluta del
    peso en gramos
    de Cuicas de agua para Misiones" )
```

#### Frecuencia absoluta del peso en gramos de Cuicas de agua para Misiones





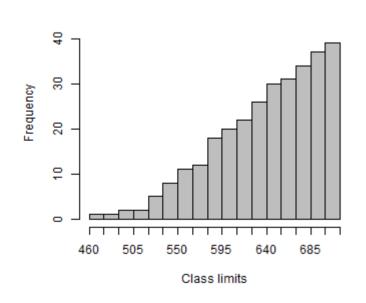
## Frecuencias acumuladas

# © somaquadrados

#### **Graficar**: Histograma

```
plot(tabla, type="cfh",
    main="Frecuencia acumulada del
    peso en gramos de
    Cuicas de agua para Misiones")
```

#### Frecuencia acumulada del peso en gramos de Cuicas de agua para Misiones

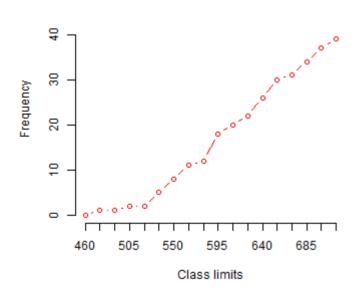




#### **Graficar**: Polígono de frecuencias



#### Frecuencia acumulada del peso en gramos de Cuicas de agua para Misiones







# **GRÁFICOS**



## **GRAFICAR EN R**

# © somaquadrados

### Paquete **stats**

- histogramas
- gráficos de densidad
- gráficos de línea
- gráficos de cajas
- gráficos de barra
- gráficos de torta

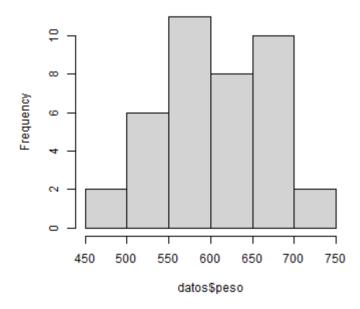


### Histogramas básico para una variable continua



hist(datos\$peso, main = "Frecuencia del peso en gramos de Cuicas de agua para Misiones")

## Frecuencia del peso en gramos de Cuicas de agua para Misiones

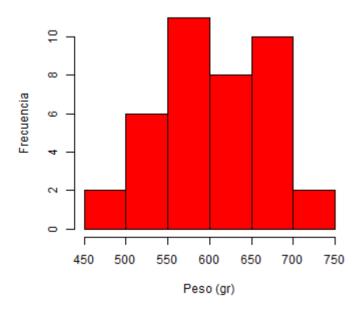


### Histogramas básico para una variable continua

```
Z
```

```
hist(datos$peso, main = "Frecuencia del peso en gramos
    de Cuicas de agua para Misiones",
    xlab = "Peso (gr)",
    ylab = "Frecuencia",
    col = "red",
    border = "black",
)
```

## Frecuencia del peso en gramos de Cuicas de agua para Misiones

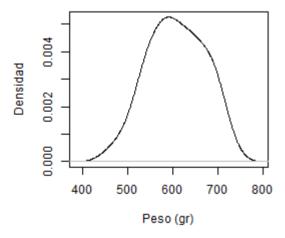


#### Gráficos de densidad



```
densidad <- density(datos$peso)
plot(densidad,
    main = "Función de densidad del peso en gramos
    de Cuicas de agua para Misiones",
    xlab = "Peso (gr)",
    ylab = "Densidad")</pre>
```

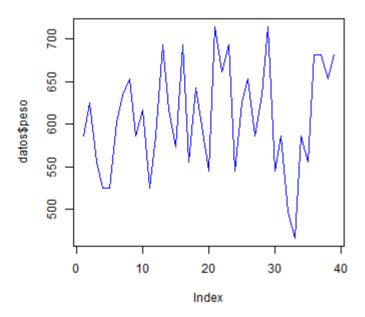
## Función de densidad del peso en gramo de Cuicas de agua para Misiones



#### **Gráficos de línea**



plot(datos\$peso, type="l", col="blue")

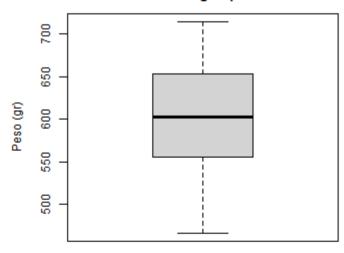


#### **Gráficos de caja**



```
boxplot(datos$peso, main = "Boxplot del peso en gramos
    de Cuicas de agua para Misiones", ylab="Peso (gr)",
        outline = TRUE)
```

## Boxplot del peso en gramos de Cuicas de agua para Misiones

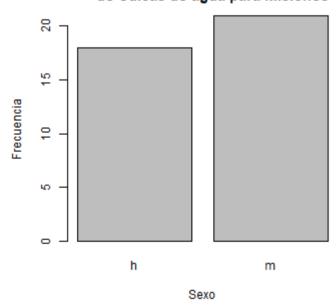


#### **Gráficos de barra**



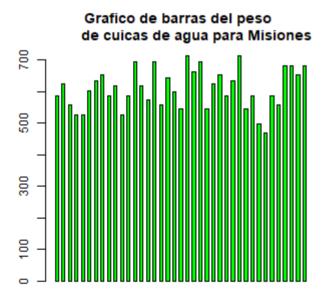
```
plot(datos$sexo, main = "Grafico de barras del sexo
    de Cuicas de agua para Misiones",
    xlab = "Sexo", ylab = "Frecuencia")
```

#### Grafico de barras del sexo de Cuicas de agua para Misiones



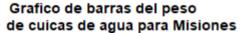
#### **Gráficos de barra**

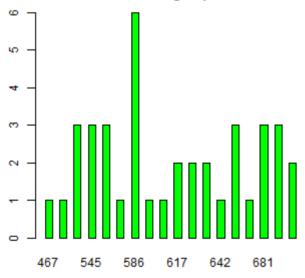




#### **Gráficos de barra**







#### **Gráficos de torta**



```
porcentajes <- as.numeric(round(((prop.table(table(datos$sexo)))*100),2))
porcentajes

## [1] 46.15 53.85

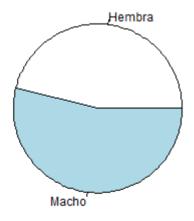
etiquetas <- c("Hembra", "Macho")
etiquetas

## [1] "Hembra" "Macho"</pre>
```



```
pie(porcentajes, etiquetas,
    main = "Proporción de sexos de las
    Cuicas de agua en Misiones",
    sub = "")
```

#### Proporción de sexos de las Cuicas de agua en Misiones





```
etiquetas <- paste(etiquetas, porcentajes)
etiquetas <- paste(etiquetas, "%", sep = "")
pie(porcentajes, etiquetas,
    main = "Proporción de sexos de las
    Cuicas de agua en Misiones",
    sub = "")</pre>
```





# Paquete "tidyverse"



Es un conjunto de paquetes que comparten algunas estructuras y funciones para importar, manipular, modelar y graficar datos



# ggplot2



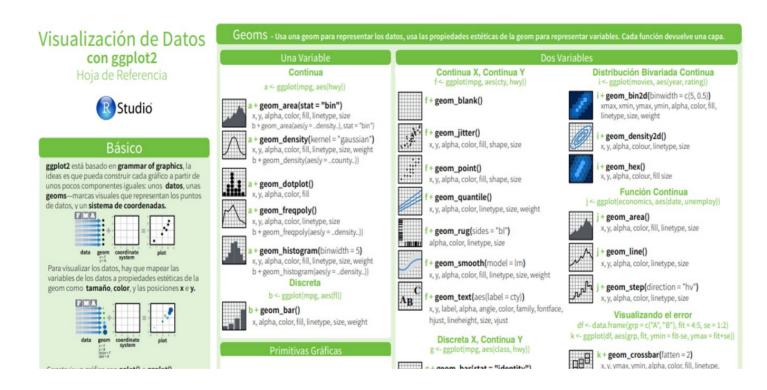
library(ggplot2)

#### **Estructura**

- datos -> base de datos, tabla
- aes -> asocia los elementos a graficar
- geometrías -> capas, define el tipo de gráfico (histograma, barras, etc)
- facetas -> permite dividir un gráico en dos o mas partes siguiendo las mismas escalas

# ggplot2





https://www.rstudio.com/wp-content/uploads/2015/04/ggplot2-spanish.pdf

# ggplot2



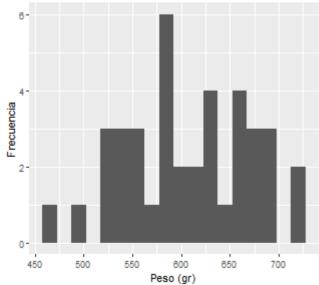
- histogramas
- gráficos de densidad
- gráficos de puntos
- gráficos de línea
- gráficos de cajas
- gráficos de barra
- gráficos de torta



# Histogramas



#### Histograma del peso en gramos de Cuicas de agua en Misiones

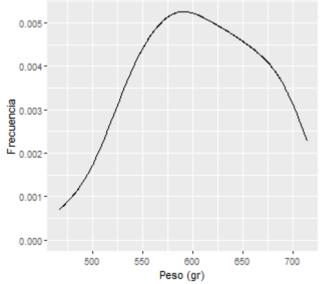




## Gráficos de densidad



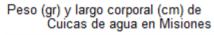
#### Peso en gramos de Cuicas de agua en Misiones

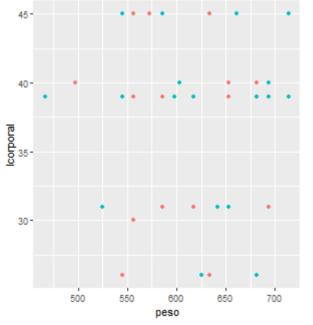




# Gráficos de puntos





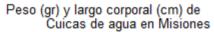


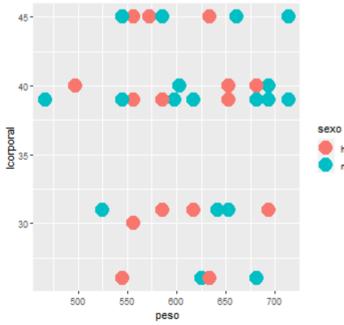


sexo

# Gráficos de puntos



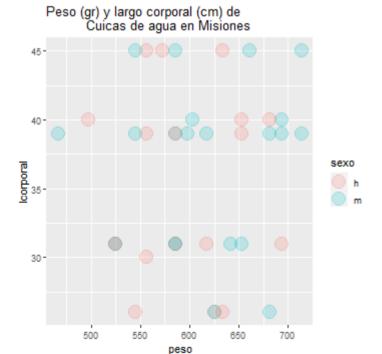






# Gráficos de puntos

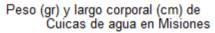


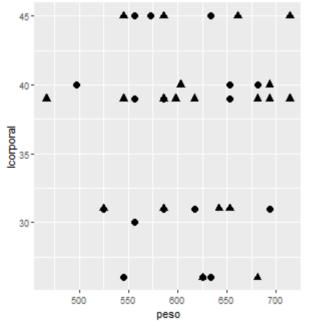




# Gráficos de puntos





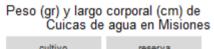


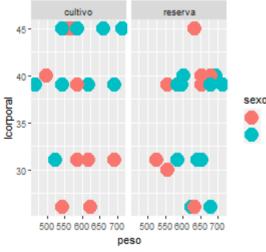


sexo

# Gráficos de puntos





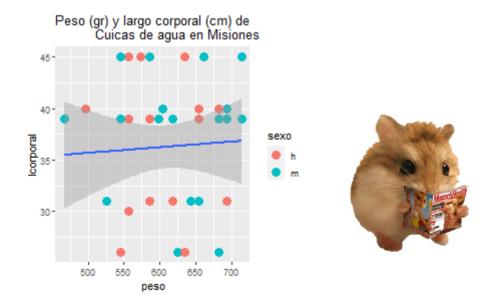




### Transformación estadística



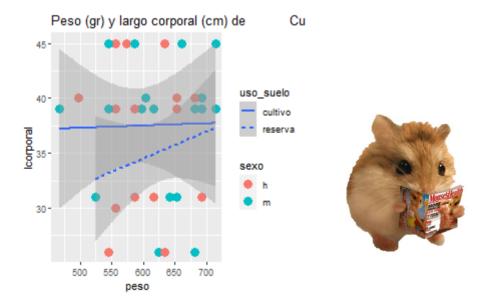
## `geom\_smooth()` using formula 'y ~ x'



## Transformación estadística



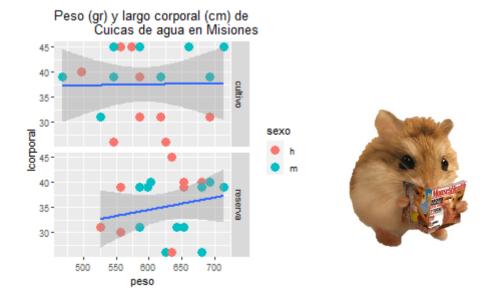
## `geom\_smooth()` using formula 'y ~ x'



## Transformación estadística



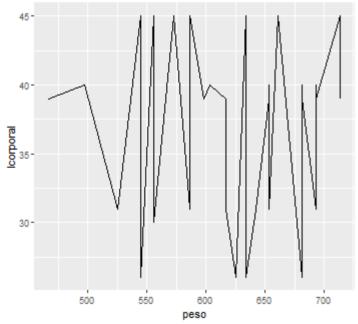
## `geom\_smooth()` using formula 'y ~ x'



# Gráficos de línea



#### Peso (gr) y largo corporal (cm) de Cuicas de agua en Misiones





# Gráficos de línea



```
ggplot(datos) +
    geom_line(aes(x=peso, y=lcorporal, colour=uso_suelo), linetype=1, size=1) + ggt
    Cuicas de agua en Misiones")
```

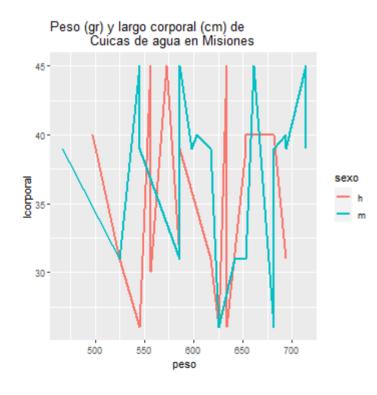




## Gráficos de línea

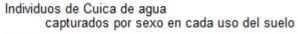


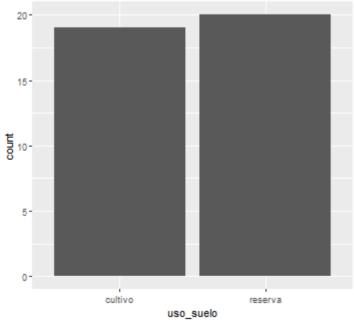
```
ggplot(datos) +
    geom_line(aes(x=peso, y=lcorporal, colour=sexo), linetype=1, size=1) + ggtitle
    Cuicas de agua en Misiones")
```





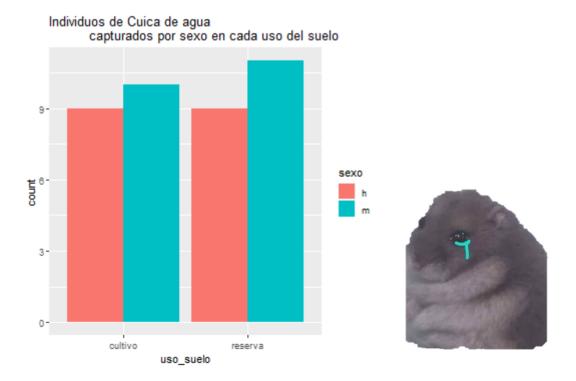






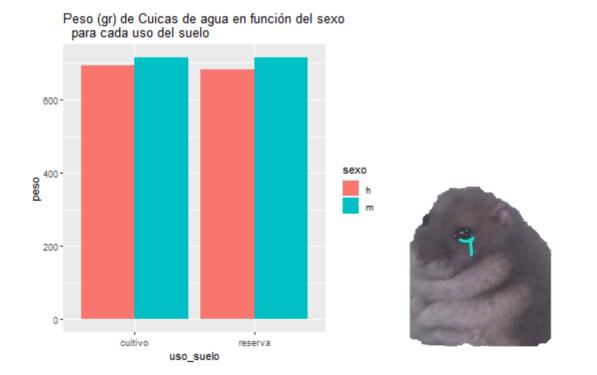








```
ggplot(datos, aes(x=uso_suelo, y= peso, fill=sexo)) +
  geom_bar(position="dodge", stat= "identity")+
  ggtitle("Peso (gr) de Cuicas de agua en función del sexo
  para cada uso del suelo")
```





**geom\_bar()** y **geom\_col()** grafican conteos, proporciones o frecuencias de los datos

si queremos graficar los valores crudos de una variable cuantitativa, es mejor usar un gráfico de lineas o de puntos

para estas variables también podemos usar boxplot que ademas nos da una idea de las medidas de posicion y dispersion de nuestra variable



# Gráficos de caja

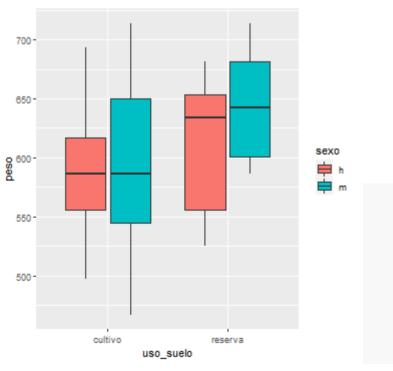


```
ggplot(datos) +
  geom_boxplot(aes(x = uso_suelo, y = peso))+
    geom_jitter(aes(x = uso_suelo, y = peso),
        size = 2,
        alpha = 0.5,
        width = 0.1)
```

# Gráficos de caja



```
ggplot(datos) +
  geom_boxplot(aes(x = uso_suelo, y = peso, fill=sexo))
```

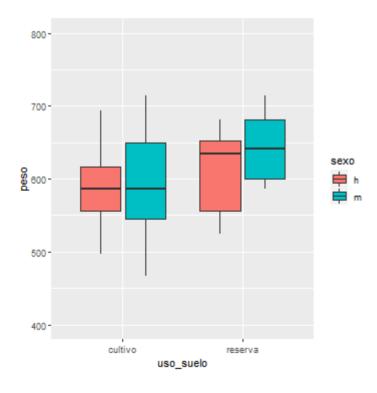




# Gráficos de caja



```
ggplot(datos) +
  geom_boxplot(aes(x = uso_suelo, y = peso, fill=sexo))+
  scale_y_continuous(limits=c(400,800))
```

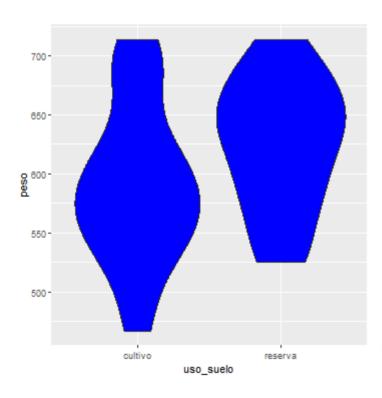




# Gráficos de violin



```
ggplot(datos, aes(x=uso_suelo, y=peso)) +
  geom_violin(fill="blue")
```





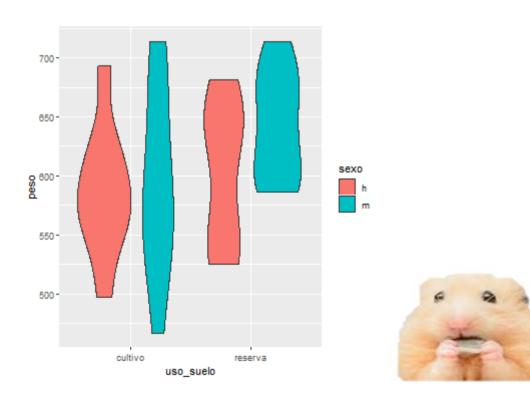
## Gráficos de violin



# Gráficos de violin



```
ggplot(datos, aes(x=uso_suelo, y=peso, fill=sexo)) +
  geom_violin()
```



## **Gráficos**



Para ver mas estilos de gráficos y cómo programarlos

-> https://www.r-graph-gallery.com/index.html



# Gráficos temporales

## [11] "2015-10-01" "2015-02-01"



```
library(tidyverse)
library(lubridate)

set.seed(1984)
    f <- sample(seq(as.Date('2015-01-01'), as.Date('2021-01-01'), by = "month"), 12)

# Dato de tipo 'date' (tiempo)
fecha <- c(f) %>%
    as.Date()
fecha

## [1] "2018-04-01" "2016-08-01" "2019-10-01" "2017-08-01" "2016-05-01"
## [6] "2020-08-01" "2017-12-01" "2018-02-01" "2018-09-01" "2020-05-01"
```



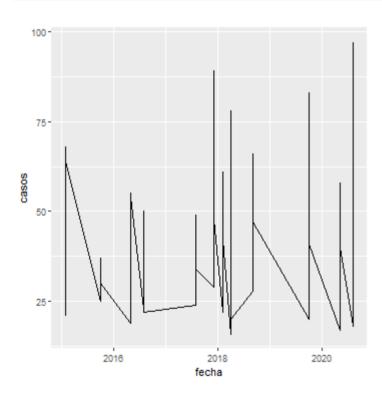
```
rural <- sample(15:30, 12)
urbano <- sample(30:100, 12)
periurbano <- sample(15:70, 12)</pre>
data <- data.frame(rural, urbano, periurbano, fecha) %>%
  pivot_longer(-fecha,
               names_to = "uso_tierra",
               values_to = "casos")
data
## # A tibble: 36 x 3
## fecha uso_tierra casos
## <date> <chr>
                           <int>
## 1 2018-04-01 rural
                              16
## 2 2018-04-01 urbano
                              78
## 3 2018-04-01 periurbano
                              20
## 4 2016-08-01 rural
                              23
## 5 2016-08-01 urbano
                              50
## 6 2016-08-01 periurbano
                              22
## 7 2019-10-01 rural
                              20
## 8 2019-10-01 urbano
                              83
## 9 2019-10-01 periurbano
                              41
## 10 2017-08-01 rural
                               24
```

## # ... with 26 more rows

### **Gráfico de línea**



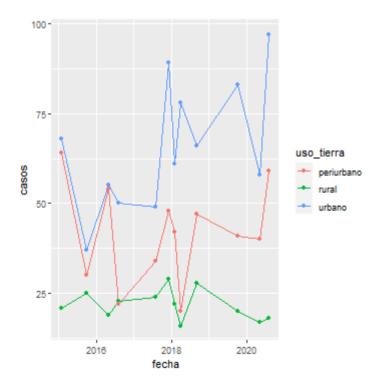
```
ggplot(data, aes(x=fecha, y=casos)) +
  geom_line()
```



### **Gráfico de línea**

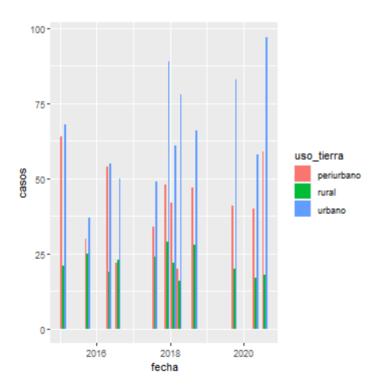


```
data %>%
  ggplot(aes(x = fecha, y = casos, color = uso_tierra)) +
  geom_line() +
  geom_point()
```





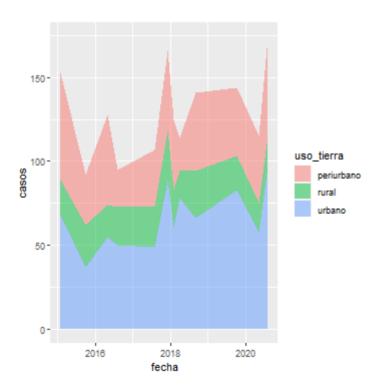
```
data %>%
  ggplot(aes(x = fecha, y = casos, fill = uso_tierra)) +
  geom_bar(position="dodge", stat="identity")
```



### Gráfico de área



```
data %>%
  ggplot(aes(x = fecha, y = casos, fill = uso_tierra)) +
  geom_area(alpha=0.5)
```



# Mapas interactivos



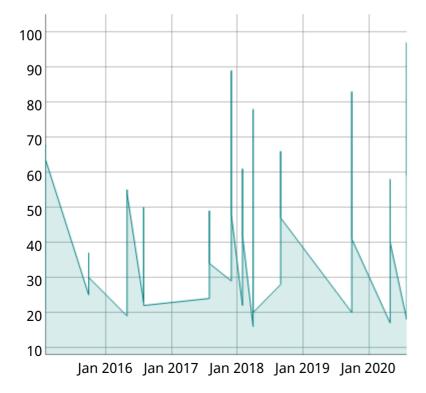
```
library(dygraphs)
library(xts)
```

```
data1 <- xts(x = data$casos, order.by = data$fecha)</pre>
```

# Mapas interactivos



```
p <- dygraph(data1) %>%
  dyOptions( fillGraph=TRUE )
p
```



### Gráfico de red



También llamados "Spider", gráficos de "Web", "Polar" o "Radarchart"

#### Base de datos

- -> fila -> una entidad
- -> columng -> variable cuantitativa

Las dos primeras filas proporcionan el mínimo y el máximo que se utilizará para cada variable

Sirve para graficar **RELACIONES** entre los datos

library(fmsb)

## Gráfico de red



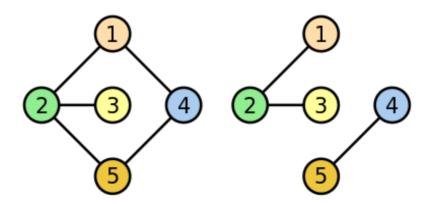


### **Teoria de grafos**

**Grafos** son un conjunto de **nodos** y **aristas** 

Permiten reprensentar relaciones binarias dentro de un conjunto

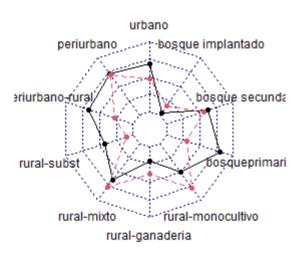
Relaciones dirigidas o no dirigidas



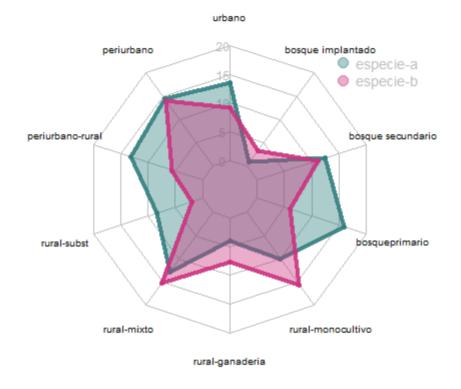
### **Ejemplo**



### radarchart(data)







### **Ejemplo**

```
Irados
```

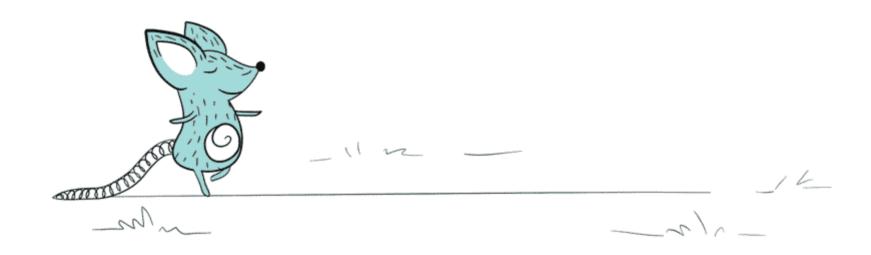
```
radarchart( casos, axistype=1 ,
   cglcol="grey", cglty=1, axislabcol="grey",
   caxislabels=seq(0,20,5), cglwd=0.8)
```



### **Dudas y/o consultas**



efburgos@conicet.gov.ar; mariliabioufpr@gmail.com



**Ejercicios ->** Ejercicios\_Clase4.pdf