

## Ejercicio 1. Lab 4

Para realizar esta tarea tendremos que instalar las librerías animation y pyramid. El primer paso es la lectura correcta de los datos, que pasa por un previo tratamiento de los mismos. Cuando los datos se descargan del Istac es recomendable eliminar el separador de los miles y agregarle , en el separador decimal. Esto es fácil de realizar usando ms excel.

Una vez terminado esto tendremos que poner las opciones de la animación y usaremos la función saveGIF para crear nuestra imagen.

Dentro de esta función recorreremos del 19 al 1 debido a que los años están al revés en los datos descargados e iremos confeccionando nuestra pirámide.

```
setwd("C:/Users/jrpen/OneDrive/Escritorio")
poblac <- read.csv2("istac.csv", header=F, col.names=c("Edad", rep(c("Hombre", "Mujer"), 20)), skip=10,

library(animation)
library(pyramid)

ani.options(interval=.5)

saveGIF({
  for (i in 19:1){
    piramide <- as.data.frame((poblac[,c((2*i), (2*i)+1, 1)]))
    names(piramide)<- c("Hombres", "Mujeres", "Edad")
    pyramid(piramide, Llab = "Hombres", Rlab = "Mujeres", Clab = "Edad",
            main=paste("Poblacion Canarias", 2019-i),
            Laxis=seq(0,100000, len=5), AxisFM = "d", Csize=0.8)
  }
}, movie.name="prueba.gif")

## Output at: prueba.gif
## [1] TRUE
```

## Ejercicio 3. Lab 4

setup)

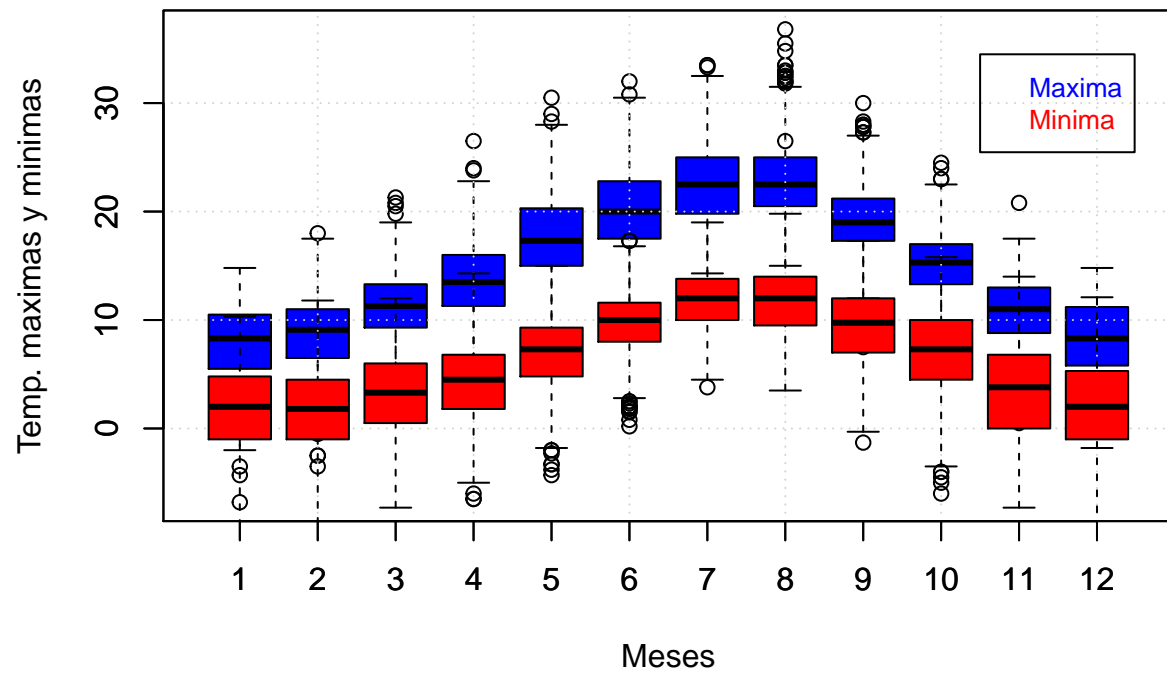
```
setwd("~/Escritorio/R/Datasets")
library(knitr)
library(ggplot2)
datos_c <- read.table("Tiempo_Clima.txt", sep="\t", dec = ".", header=T)
attach(datos_c)
kable(datos_c[c(1:3, 150:152, 1300:1304), ])
```

	T_Maxima	T_Minima	Precipitacion	Mes	Year
1	10.8	6.5	12.2	1	1987
2	10.5	4.5	1.3	1	1987
3	7.5	-1.0	0.1	1	1987
150	18.3	12.5	0.0	5	1987
151	19.3	10.8	11.7	5	1987
152	19.8	10.3	0.0	6	1987
1300	23.8	12.5	0.0	7	1990
1301	25.0	7.3	0.0	7	1990
1302	25.0	6.8	0.0	7	1990
1303	26.3	8.4	0.0	7	1990
1304	24.0	13.3	0.0	7	1990

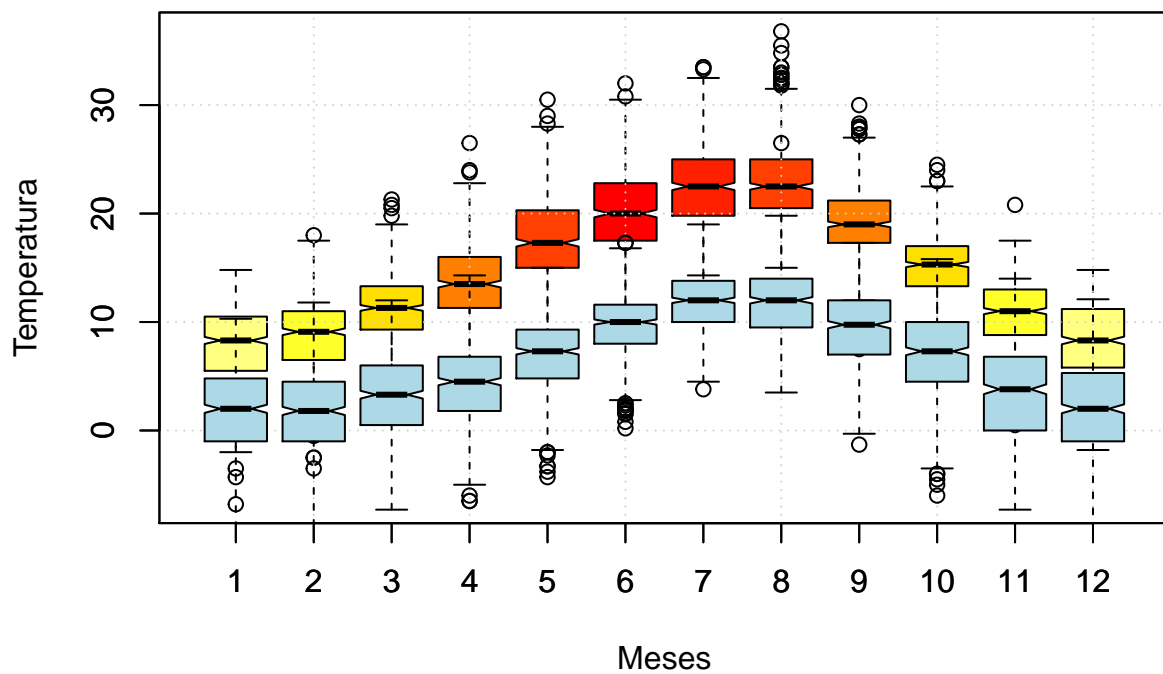
**Visualizar con R básico y con ggplot2 los cambios de temperatura máxima y mínima en función del mes de año.**

Para ello realizaremos una primera visión conjunta usando el paquete de R básico, realizaremos dos plot de las temperaturas máximas y las mínimas con una leyenda explicativa (se pueden apreciar dos, una mas personalizada y otra menos). Y a continuación usaremos ggplot2 para reproducir la misma situación.

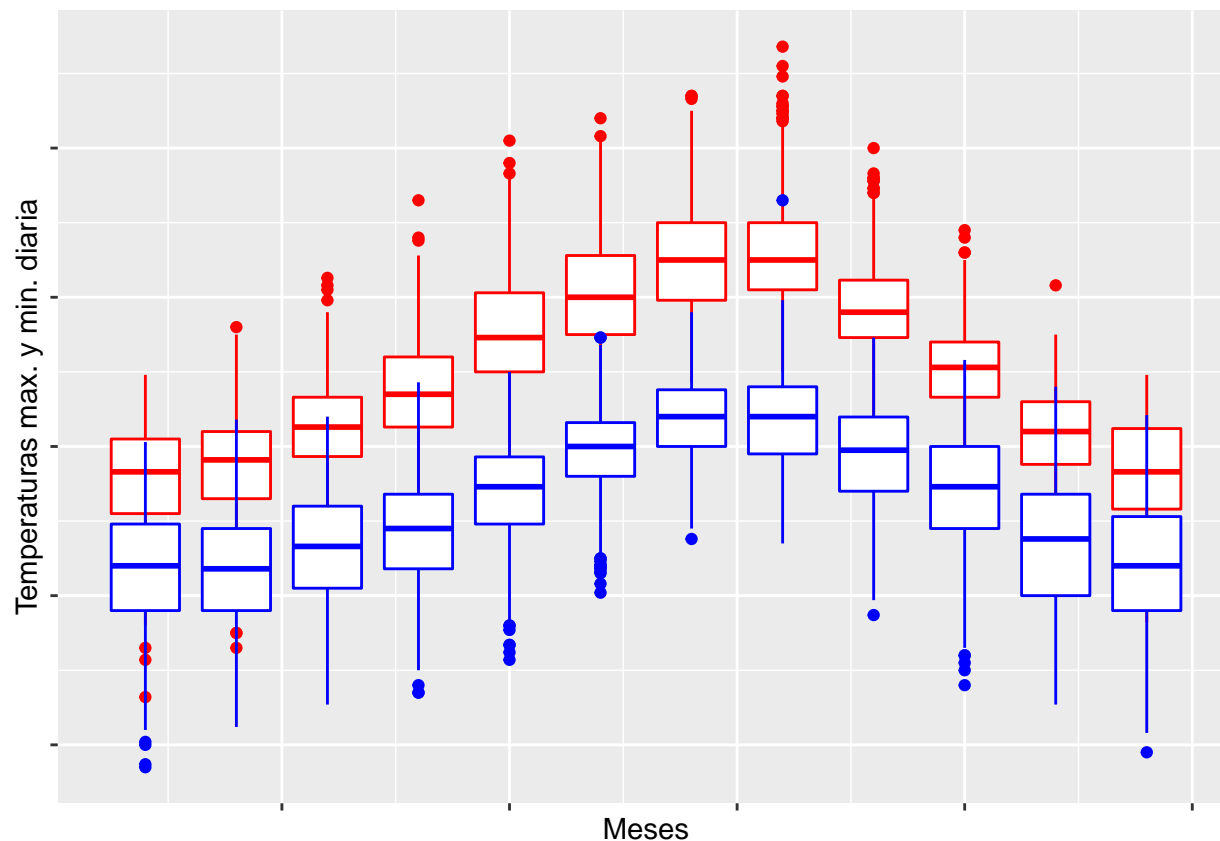
```
Mes<-factor(Mes)
plot(Mes, T_Maxima, col="blue", xlab="Meses", ylab="Temp. maximas y minimas");grid()
plot(Mes, T_Minima, col="red", add=T)
legend(10.5, 34.5, legend=c("Maxima", "Minima"), cex=0.85, text.col=c("blue", "red"))
```



```
estacion <- heat.colors(12)
paleta_sel <- c(11,10,8,5,3,1,2,3,5,8,10,11)
plot(Mes, T_Maxima, col=estacion[paleta_sel], xlab="Meses", ylab="Temperatura", notch=T); grid()
plot(Mes, T_Minima, col="lightblue", notch=T, add=T)
```



```
g<- ggplot(datos_c, aes(x=Mes, y=T_Maxima))
g+geom_boxplot(aes(group=Mes), col="red")+
  ylab("Temperaturas max. y min. diaria")+xlab("Meses")+
  theme(axis.text=element_blank())+
  geom_boxplot(aes(y=T_Minima, group=Mes), col="blue")
```



## Idem a) para el nivel de precipitaciones a lo largo de los meses de cada año.

Para ello lo primero que realizaremos es tratado de datos, usando la función `par`, la función `aggregate`... Y a continuación mostraremos usando R básico las precipitaciones a lo largo de los meses.

```
par(mfrow=c(1,1))
max_p <- aggregate(Precipitacion~Year, datos_c, max);max_p
```

```
##      Year Precipitacion
## 1  1987          54.3
## 2  1988          23.7
## 3  1989          29.0
## 4  1990          15.0
## 5  1991          30.3
## 6  1992          44.5
## 7  1993          59.5
## 8  1994          38.8
## 9  1995          32.5
## 10 1996          22.0
## 11 1997          32.0
## 12 1998          30.9
## 13 1999          42.5
## 14 2000          41.4
## 15 2001          26.3
## 16 2002          43.0
## 17 2003          29.5
## 18 2004          24.5
## 19 2005          22.0
```

```
plot(max_p$Year, max_p$Precipitacion, ylab="Precipitaciones maximas anuales", xlab="Años", col="red", t
```



## c) Crear un data frame con los temperaturas máximas y mínimas por año.

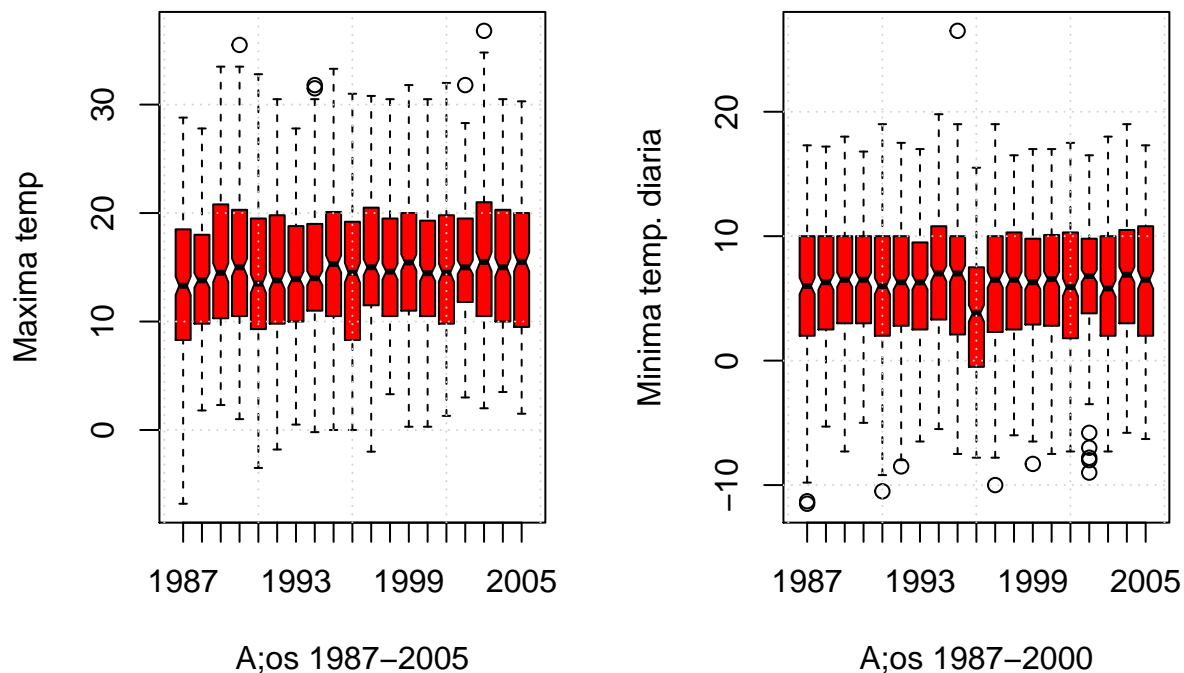
Lo primero que realizaremos es el data frame usando el comando aggregate y creando un nuevo data frame con data.frame. Por último lo mostraremos usando R básico.

```
max_temp <- aggregate(T_Maxima~Year, datos_c, max)
min_temp <- aggregate(T_Minima~Year, datos_c, min)
temp <- data.frame(Year=unique(Year), T_Maxima=max_temp[,2], T_Minima=min_temp[,2])
kable(temp)
```

Year	T_Maxima	T_Minima
1987	28.8	-11.5
1988	27.8	-5.3
1989	33.5	-7.3
1990	35.5	-5.0
1991	32.8	-10.5
1992	30.5	-8.5
1993	27.8	-6.5
1994	31.8	-5.5
1995	33.3	-7.5
1996	31.0	-7.8
1997	30.8	-10.0
1998	30.5	-6.0
1999	31.8	-8.3
2000	30.5	-7.5

Year	T_Maxima	T_Minima
2001	32.0	-7.3
2002	31.8	-9.0
2003	36.8	-7.3
2004	30.5	-5.8
2005	30.3	-6.3

```
Year<- factor(Year)
par(mfrow=c(1,2))
plot(Year, T_Maxima, ylab="Maxima temp", xlab="Años 1987-2005", col="red", notch=TRUE);grid()
plot(Year, T_Minima, ylab="Minima temp. diaria", xlab="Años 1987-2000", col="red", notch=TRUE); grid()
```



d) Utilizar la función `lm()` para ver la tendencia e influencia del cambio de temperaturas a lo largo de los años. Analizar si existe una relación con el fenómeno del calentamiento global. Razonar y justificar las respuestas.

Se puede observar gracias a las aproximaciones realizadas y los datos mostrados que la temperatura a lo largo de los años esta aumentando, parece que es un cambio poco drástico y aun así no tenemos suficientes datos para dar una opinión rigurosa y certera. Sin embargo un cambio en la temperatura, por muy pequeño que sea, influye de forma bestial en la biosfera terrestre. También cabe mencionar que a lo largo de los años el aumento de la población humana, el uso de recursos fósiles que contienen dióxido de carbono, etc... puede haber influido significativamente.

```
par(mfrow=c(1,1))
plot(max_temp$Year, max_temp$T_Maxima, ylab="Temp. maximas", xlab="Años", col="red", type="b");grid()
```

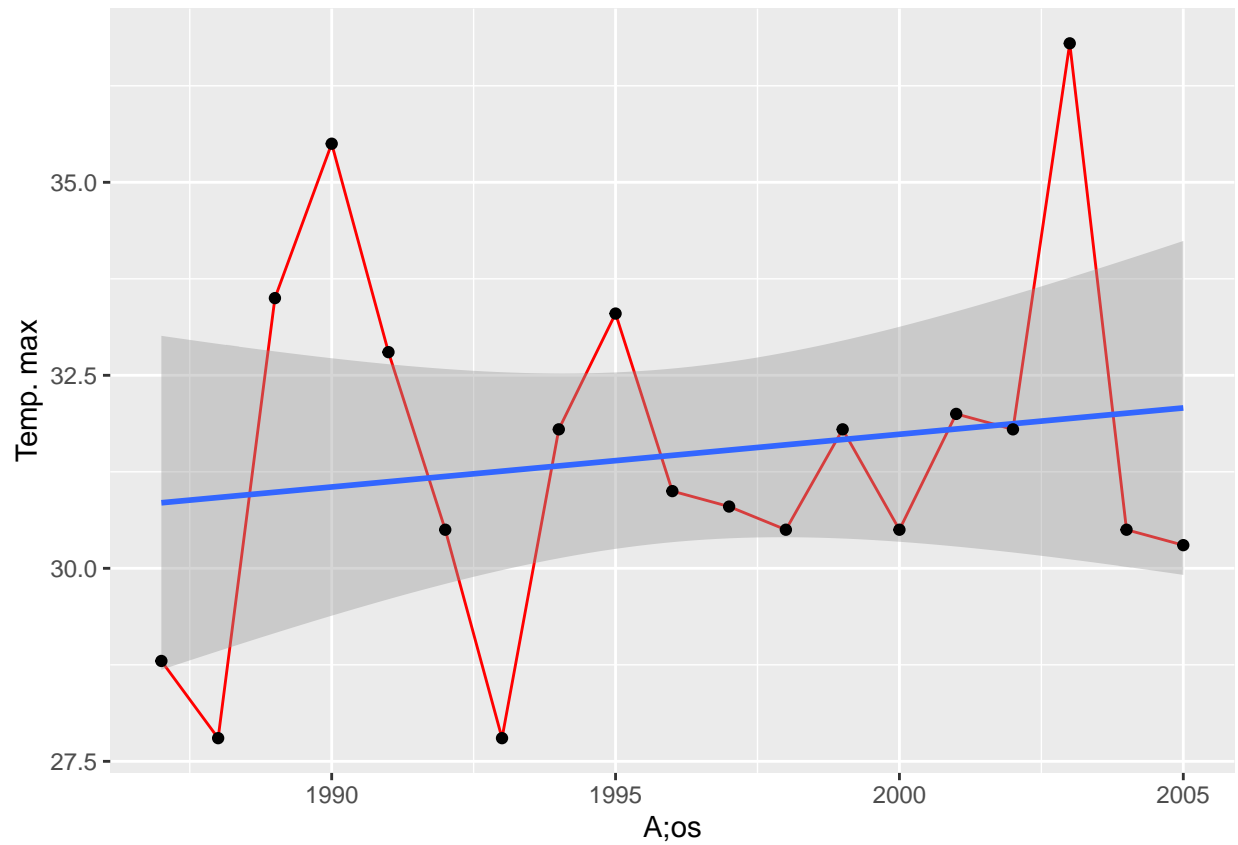
```
mod<- lm(max_temp$T_Maxima~max_temp$Year)
abline(mod, col="green", lwd=2)
```



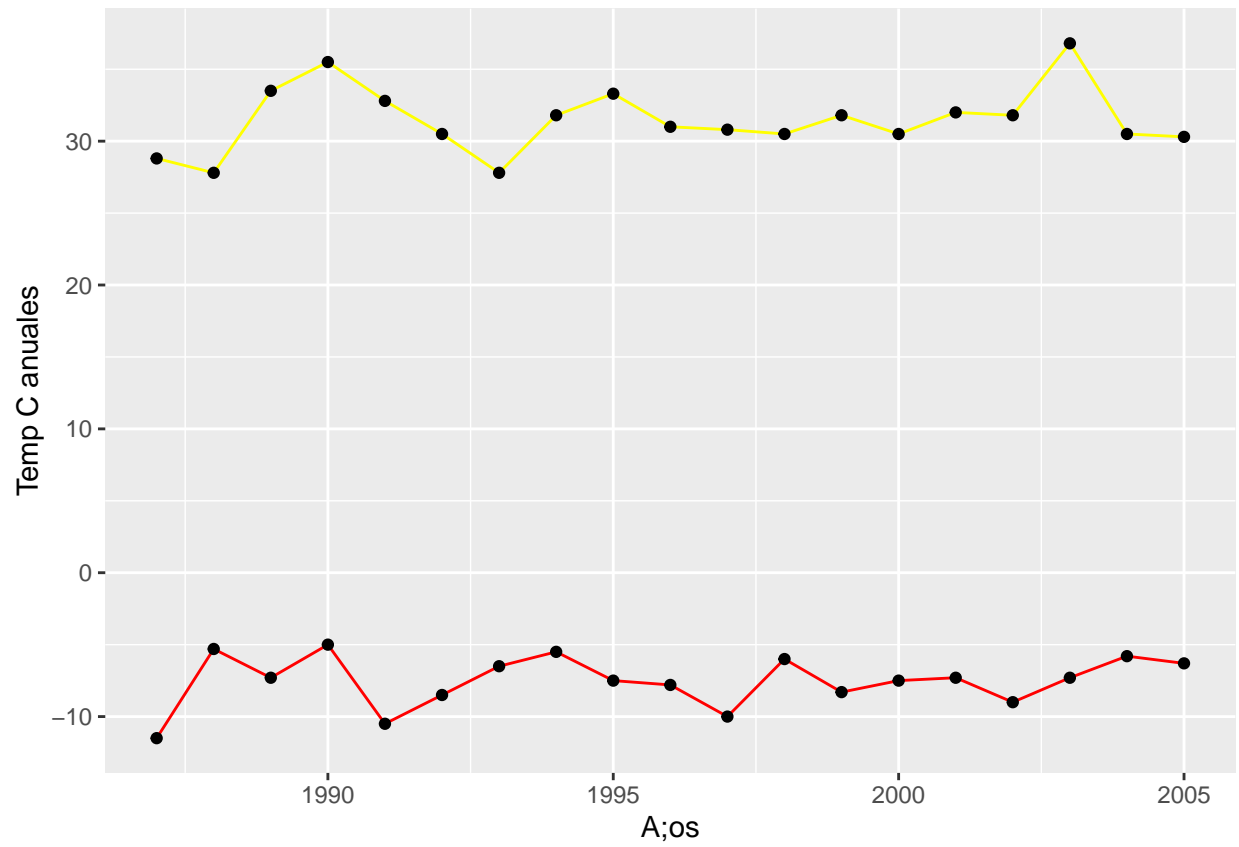
```
g2<-ggplot(max_temp, aes(x=Year, y=T_Maxima))+ylab("Temp. max")+xlab("A;os")
g2+geom_line(col="red")+geom_smooth(method="lm")+geom_point()
```

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





```
g2<-ggplot(max_temp, aes(x=Year, y=T_Maxima))+
  ylab("Temp C anuales")+xlab("Años")
g3<-g2+geom_line(col="yellow")+geom_point()
g4<-g3+geom_line(aes(x=min_temp$Year, y=min_temp$T_Minima), col="red")+
  geom_point(aes(x=min_temp$Year, y=min_temp$T_Minima));g4
```



```
detach(datos_c)
```

## Ejercicio 4. Lab 4

```
setwd("~/Escritorio/")
poblac <- read.csv("ejercicio4.csv", header=F, col.names=c("Edad", "Hombre", "Mujer"))

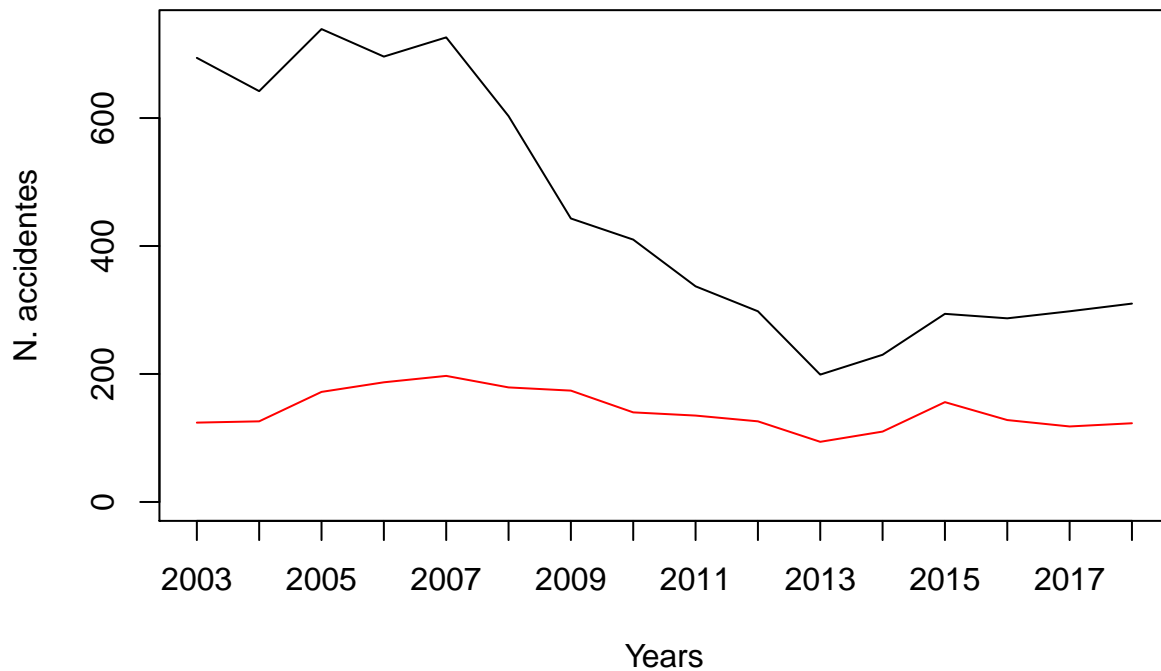
attach(poblac)
```

a)

Para realizar este trabajo tendremos que tener en cuenta 2 cosas primero que los datos de los años no se han descargado ordenados (tendremos que tenerlo en consideración a la hora de realizar los gráficos) y la comparativa a realizar entre hombres y mujeres.

A su vez hemos visualizado una tabla con los años máximos y mínimos segun mujer y hombre.

```
plot(Edad[16:1], Hombre[16:1], xaxt="n", type="l",
     ylab="N. accidentes", xlab="Years", ylim=c(0, max(Hombre)))
years <- as.character(seq(2003, 2018, 1))
axis(side = 1, at=Edad[16:1], labels=years)
points(Edad[16:1], Mujer[16:1], col="red", type="l")
```



```

H_max <- Edad[which.max(Hombre)]; H_max

## [1] 2005
H_min <- Edad[which.min(Hombre)]; H_min

## [1] 2013
M_max <- Edad[which.max(Mujer)]; M_max

## [1] 2007
M_min <- Edad[which.min(Mujer)]; M_min

## [1] 2013
v_sing<- data.frame(Sexo=c("hombres", "mujeres"
), max_year=c(H_max, M_max), min_year=c(H_min, M_min));v_sing

##      Sexo max_year min_year
## 1 hombres    2005    2013
## 2 mujeres    2007    2013

```

b)

Hemos realizado el mismo proceso anteriormente descrito pero con los datos totales proporcionados por la ISTAC.

```

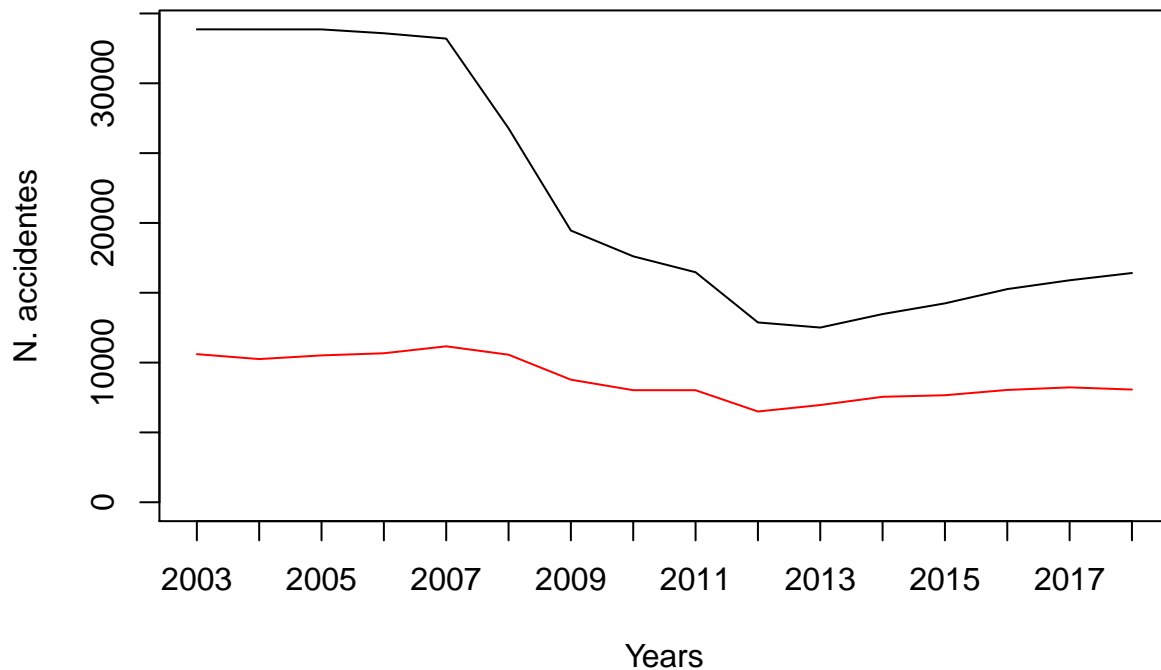
poblac <- read.csv("Ejercicio41.csv", header=F, col.names=c("Edad", "Hombre", "Mujer"))

attach(poblac)

## The following objects are masked from poblac (pos = 3):
##
##      Edad, Hombre, Mujer

plot(Edad[16:1], Hombre[16:1], xaxt="n", type="l",
      ylab="N. accidentes", xlab="Years", ylim=c(0, max(Hombre)))
years <- as.character(seq(2003, 2018, 1))
axis(side = 1, at=Edad[16:1], labels=years)
points(Edad[16:1], Mujer[16:1], col="red", type="l")

```



```
H_max <- Edad[which.max(Hombre)]; H_max
```

```
## [1] 2003
```

```
H_min <- Edad[which.min(Hombre)]; H_min
```

```
## [1] 2013
```

```
M_max <- Edad[which.max(Mujer)]; M_max
```

```
## [1] 2007
```

```
M_min <- Edad[which.min(Mujer)]; M_min
```

```
## [1] 2012
```

```
v_sing<- data.frame(Sexo=c("hombres", "mujeres"),
  max_year=c(H_max, M_max), min_year=c(H_min, M_min));v_sing
```

```
##      Sexo max_year min_year
## 1 hombres    2003    2013
## 2 mujeres    2007    2012
```

c)

Podemos observar que definitivamente se han producido bastantes mejoras en los accidentes sobretodo en el entorno masculino puesto que la diferencia entre los datos observadoa antes de 2007 y los de 2017 es de casi 10.000 accidentes. Esto es un claro signo de que se esta abordando el problema con soluciones eficaces.

d)

Hemos usado la librería mgcv para predecir cuales son los sucesos que pasarán en 2019 y 2020 como podemos observar hemos realizado dos modelos uno lineal y otro cuadrático.

Vemos una tendencia ascendente en las predicciones salvo en la cuadrática masculina, que primero realiza un descenso y después vuelve a pronunciarse.

```
library(mgcv)

## Loading required package: nlme
## This is mgcv 1.8-33. For overview type 'help("mgcv-package")'.

modelo1<-gam(Hombre~s(Edad))
xv<-c(2019, 2020)
yv<-predict(modelo1, list(Edad=xv))
plot(Edad[16:1], Hombre[16:1], xaxt="n", type="l",
      ylab="N. accidentes", xlab="Years", ylim=c(0, max(Hombre)), xlim=c(2003, 2020))
years <- as.character(seq(2003, 2018, 1))
axis(side = 1, at=Edad[16:1], labels=years)
points(Edad[16:1], Mujer[16:1], col="red", type="l")
points(xv, yv, col="blue", lwd=2)

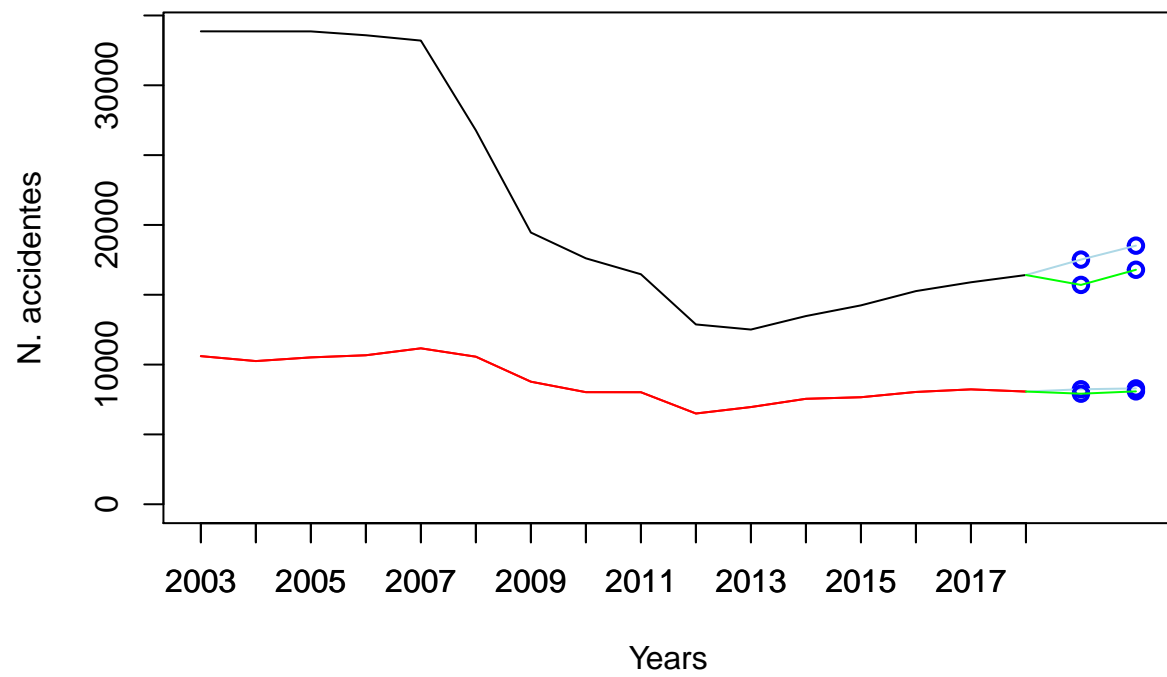
lines(c(2018, xv), c(Hombre[Edad==2018], yv), col='lightblue')

modelo2<-gam(Mujer~s(Edad))
xv<-c(2019, 2020)
yv<-predict(modelo2, list(Edad=xv))
points(xv, yv, col="blue", lwd=2)
lines(c(2018, xv), c(Mujer[Edad==2018], yv), col='lightblue')

modelo1<-lm(Hombre~Edad+I(Edad^2))
xv<-c(2019, 2020)
yv<-predict(modelo1, list(Edad=xv))
years <- as.character(seq(2003, 2018, 1))
axis(side = 1, at=Edad[16:1], labels=years)
points(Edad[16:1], Mujer[16:1], col="red", type="l")
points(xv, yv, col="blue", lwd=2)

lines(c(2018, xv), c(Hombre[Edad==2018], yv), col='green')

modelo2<-lm(Mujer~Edad+I(Edad^2))
xv<-c(2019, 2020)
yv<-predict(modelo2, list(Edad=xv))
points(xv, yv, col="blue", lwd=2)
lines(c(2018, xv), c(Mujer[Edad==2018], yv), col='green')
```



## Lab 5

### Ejercicio 1:

Para la realización de esta práctica usaremos diferentes funciones del paquete MASS. Una de ellas la `fitdist` que nos permitira ajustar las distribuciones.

#### a y b)

Lo primero que haremos sera cargar el archivo correspondiente y crear las diferentes distribuciones tanto gamma como normal

Para ello usaremos la función `fitdist` como habíamos comentado anteriormente. De esta forma facilmente podremos calcular las distribuciones pedidas usando como parametro su nombre.

```
setwd("~/Escritorio/R/Datasets")
dist <- read.csv2("Datos_Distribucion_Lab5_1.txt", sep = "\t", header = TRUE)

library(MASS)
ajuste.gamma <- fitdistr(as.numeric(dist$Litros), "gamma")
ajuste.gamma
```

```
##      shape      rate
## 4.4898333 0.8673706
## (0.6128239) (0.1252656)
```

```
ajuste.normal <- fitdistr(as.numeric(dist$Litros), "normal")
ajuste.normal
```

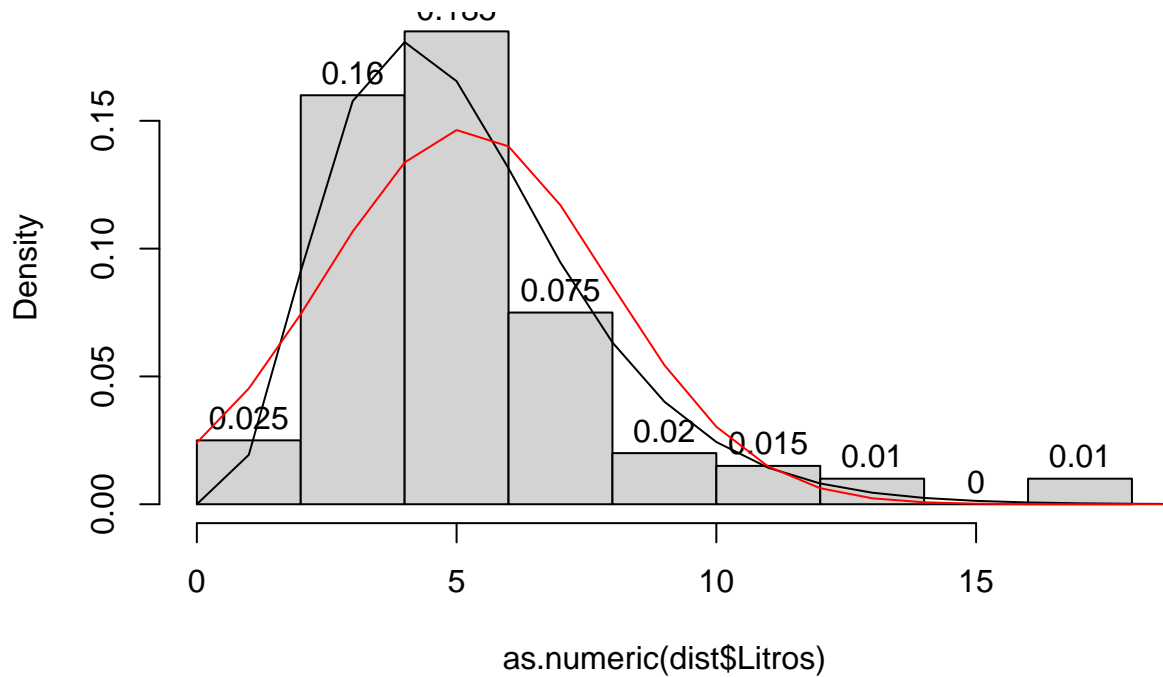
```
##      mean      sd
## 5.1763600 2.7271362
## (0.2727136) (0.1928376)
```

#### c)

Para mostrar las gráficas correspondientes sera sencillo usando un `hist` y `points`. Observamos como ambas funciones se adaptan de buena forma a nuestro histograma.

```
hist(as.numeric(dist$Litros), freq=FALSE, main="", labels=TRUE)
points(0:99, dgamma(0:99, 4.489, 0.867), type="l")
points(0:99, dnorm(0:99, 5.17, 2.72), type="l", col="red")
```





d y e)

A continuación calcularemos las medias para poder comprobar cual de las distribuciones es mas válida.

```
mu_norm <- mean(dnorm(0:15, 5.17, 2.72)); mu_norm
```

```
## [1] 0.0613693
```

```
sd_norm <- sd(dnorm(0:15, 5.17, 2.72)); sd_norm
```

```
## [1] 0.05372984
```

```
mu_gamma<- mean(dgamma(0:15, 4.489, 0.867));mu_gamma
```

```
## [1] 0.0624284
```

```
sd_gamma<- sd(dgamma(0:15, 4.489, 0.867));sd_gamma
```

```
## [1] 0.06531116
```

```
mu_base <- mean(c(0.025, 0.16, 0.185, 0.075, 0.02, 0.015, 0.01, 0,0.01));mu_base
```

```
## [1] 0.05555556
```

```
sd_base <- sd(c(0.025, 0.16, 0.185, 0.075, 0.02, 0.015, 0.01, 0,0.01));sd_base
```

```
## [1] 0.06993052
```

```
mu_norm - mu_base
```

```
## [1] 0.005813742
```

```
mu_gamma - mu_base
```

```
## [1] 0.00687284
```

```
sd_norm - sd_base
```

```
## [1] -0.01620068
```

```
sd_gamma - sd_base
```

```
## [1] -0.004619357
```

Como podemos observar la distribución que menor error ostenta es la normal, por lo tanto la que mejor se ajusta al resultado.

## Ejercicio 2:

Comenzaremos leyendo el fichero de datos correspondientes. Como indica el enunciado debemos de calcular la nota media poblacional con un intervalo de confianza del 90%.

```
setwd("~/Escritorio/R/Datasets")
notas <- read.csv("Notas_2019_20.csv", sep = ",")
notas$Notas <- as.numeric(notas$Notas)
```

```
notas <- as.numeric(notas$Notas)
```

```
ci <- mean(as.numeric(notas)) - qt(0.90, length(notas)) * sd(as.numeric(notas))/sqrt(length(notas)); ci
```

```
## [1] 4.664232
```

```
cs <- mean(as.numeric(notas)) + qt(0.90, length(notas)) * sd(as.numeric(notas))/sqrt(length(notas)); cs
```

```
## [1] 5.271566
```

A continuación realizaremos un intervalo de confianza para la desviación estandar y varianza al igual que arriba del 90%.

```
ci <- length(notas)* var(as.numeric(notas))/qchisq(0.90, length(notas));ci
```

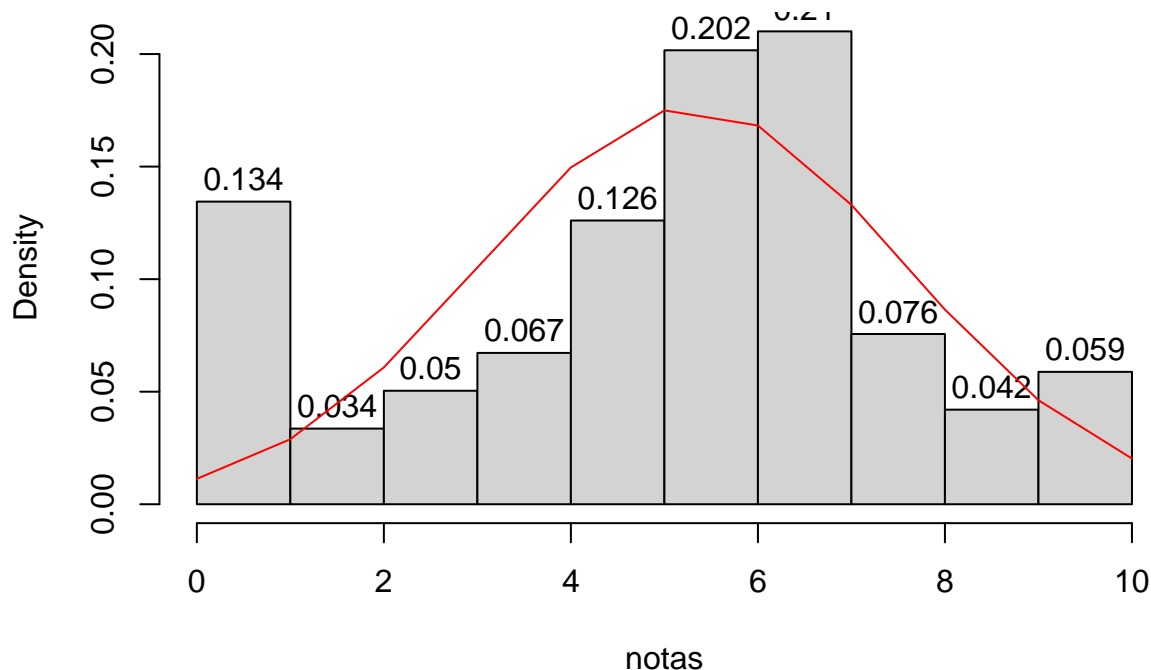
```
## [1] 5.650671
```

```
cs <- length(notas)* var(as.numeric(notas))/qchisq(1-0.90, length(notas));cs
```

```
## [1] 7.886005
```

Y por último mostraremos los datos con un histograma y los compararemos con la distribución dada

```
hist(notas, freq=FALSE, main="", labels=TRUE)
points(0:10, dnorm(0:10, 5.3, 2.26), type="l", col="red")
```



Como conclusiones podemos asumir que en promedio la mayoría de alumnos o bien suspenden o aprueban con un 5 a penas. A su vez si observamos la d. estandar vemos como es amplia por lo tanto existiran alumnos con muy buenas notas y alumnos con muy malas notas.

## Ejercicio 4

##a)

```
library(ggplot2)
library(knitr)
setwd("~/Escritorio/R/Datasets")
nivel_ozono <- read.csv("Nivel_Ozono.csv", header = T, sep = ",", dec='.')

attach(nivel_ozono)

m_oeste <- mean(Ozono[Jardin.localizacion=="Oeste"]);m_oeste

## [1] 61.26

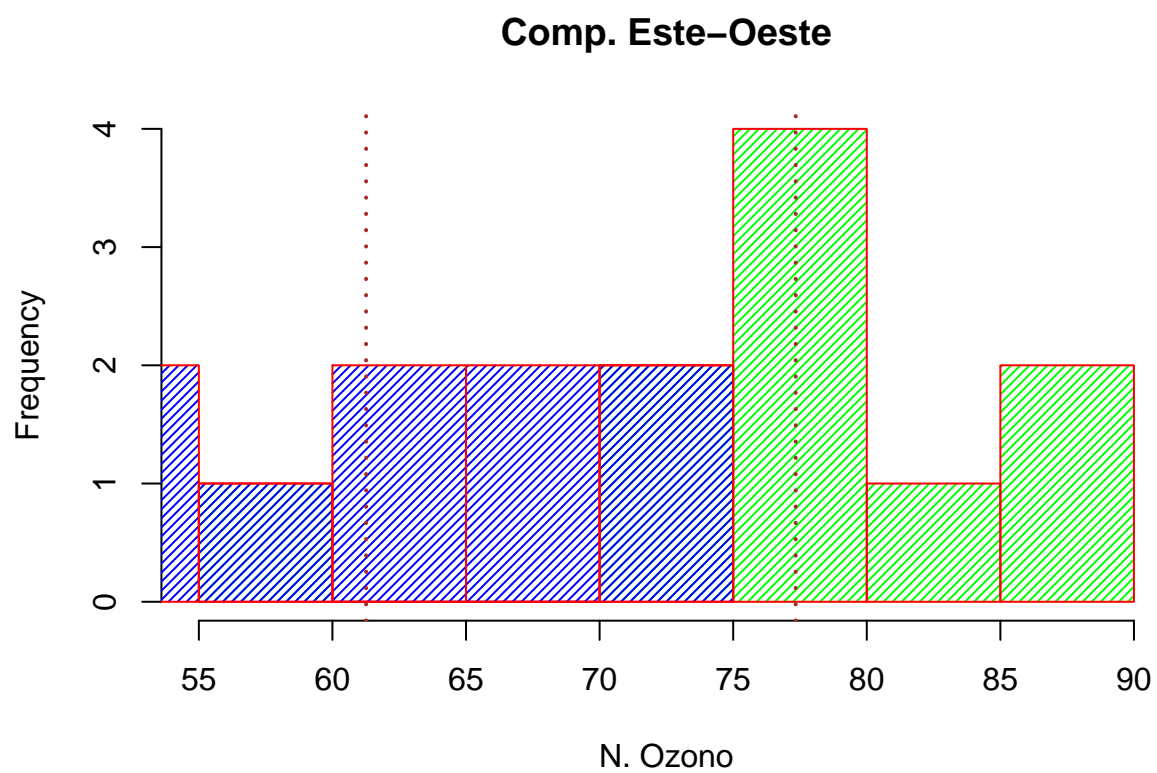
m_este <- mean(Ozono[Jardin.localizacion=="Este"]);m_este

## [1] 77.34

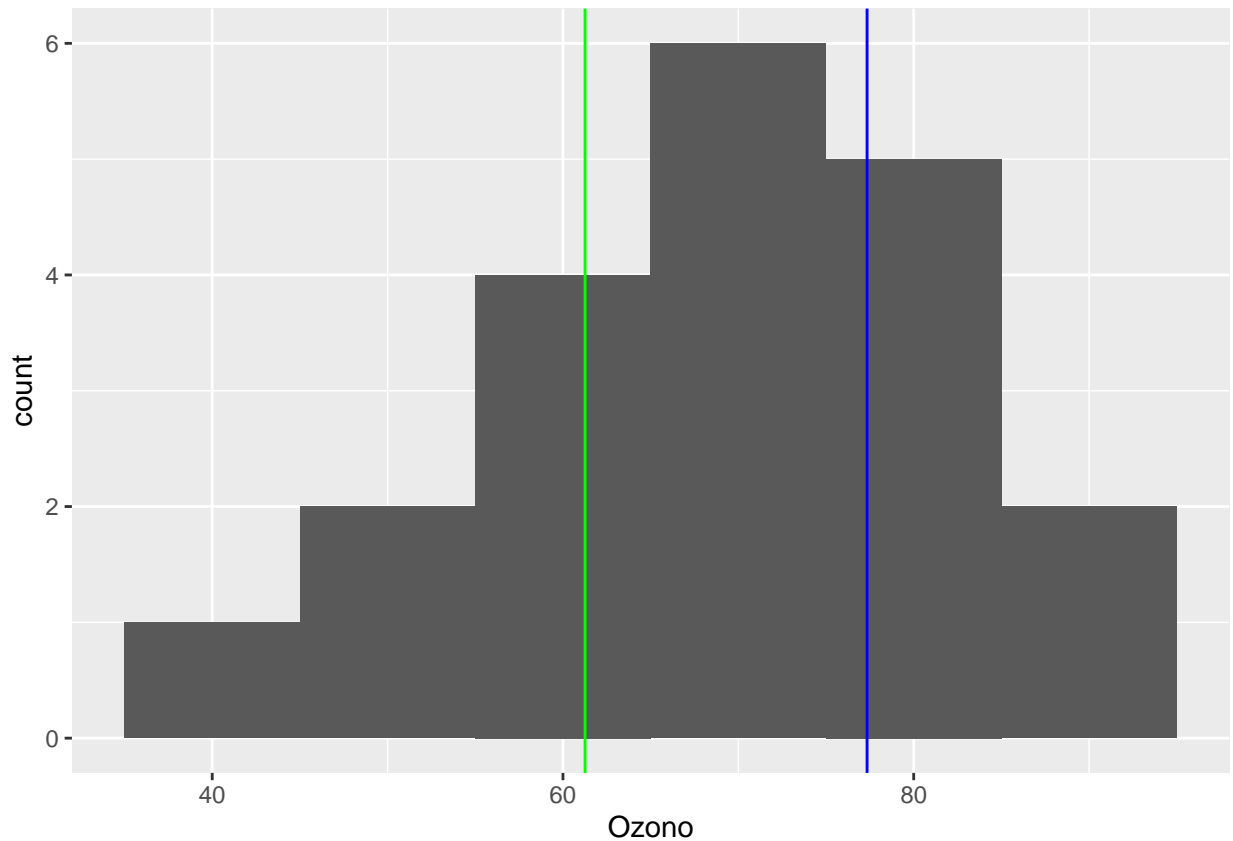
hist(Ozono[Jardin.localizacion == "Este", col="green", density=30, border="red", xlab="N. Ozono",
      main = "Comp. Este-Oeste")
hist(Ozono[Jardin.localizacion=="Oeste", add=T, col="blue", density=30, border="red")

abline(v=m_oeste, col="brown", lty=3, lwd=2)
```

```
abline(v=m_este, col="brown", lty=3, lwd=2)
```



```
g<-ggplot(nivel_ozono, aes(x=Ozono))
g+geom_histogram(binwidth = 10)+ geom_vline(xintercept=c(m_oeste, m_este), colour=c("green", "blue"))
```



##b y c) Para ello usaremos la función `t.test` y `var.test` que nos brinda el paquete de R desde el cual podremos facilmente analizar el ozono por la localización del jardin de forma muy sencilla como vemos abajo.

Podemos observar que existen claras diferencias en ambas medias y varianzas.

```
t.test(Ozono~Jardin.localizacion, data=nivel_ozono)
```

```
##
##  Welch Two Sample t-test
##
## data:  Ozono by Jardin.localizacion
## t = 4.2363, df = 17.656, p-value = 0.0005159
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   8.094171 24.065829
## sample estimates:
## mean in group Este mean in group Oeste
##           77.34           61.26
```

```
var.test(Ozono~Jardin.localizacion, data=nivel_ozono)
```

```
##
##  F test to compare two variances
##
## data:  Ozono by Jardin.localizacion
## F = 0.75503, num df = 9, denom df = 9, p-value = 0.6823
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
```

```
## 0.1875386 3.0397437
## sample estimates:
## ratio of variances
## 0.7550293
```

## Ejercicio 5:

a)

```
setwd("~/Escritorio/R/Datasets")
vacas <- read.csv("Vacas.csv", sep = ",", dec=".")
```

`vacas`

```
##      suplemento  dieta ganancia
## 1      sup1    trigo 17.37125
## 2      sup1    trigo 16.81489
## 3      sup1    trigo 18.08184
## 4      sup1    trigo 15.78175
## 5     control    trigo 17.70656
## 6     control    trigo 18.22717
## 7     control    trigo 16.08650
## 8     control    trigo 17.60184
## 9      sup2    trigo 20.78861
## 10     sup2    trigo 20.00858
## 11     sup2    trigo 19.30287
## 12     sup2    trigo 18.57330
## 13     sup3    trigo 19.54688
## 14     sup3    trigo 17.96440
## 15     sup3    trigo 21.43738
## 16     sup3    trigo 19.60763
## 17     sup1   avena 19.36092
## 18     sup1   avena 20.64938
## 19     sup1   avena 19.60615
## 20     sup1   avena 19.03557
## 21     control   avena 21.56473
## 22     control   avena 20.25976
## 23     control   avena 19.20455
## 24     control   avena 20.94562
## 25     sup2   avena 21.14743
## 26     sup2   avena 23.00609
## 27     sup2   avena 21.90279
## 28     sup2   avena 21.38462
## 29     sup3   avena 23.58744
## 30     sup3   avena 22.31517
## 31     sup3   avena 22.37653
## 32     sup3   avena 24.91439
## 33     sup1  cebada 23.82942
## 34     sup1  cebada 23.75688
## 35     sup1  cebada 21.34873
## 36     sup1  cebada 20.92947
## 37     control  cebada 24.69354
## 38     control  cebada 24.32013
## 39     control  cebada 22.02811
```

```
## 40 control cebada 22.14482
## 41 sup2 cebada 24.62988
## 42 sup2 cebada 23.08774
## 43 sup2 cebada 27.79490
## 44 sup2 cebada 26.78869
## 45 sup3 cebada 26.04248
## 46 sup3 cebada 25.28337
## 47 sup3 cebada 25.03889
## 48 sup3 cebada 29.02916
```

```
sup1 <- vacas$ganancia[vacas$suplemento=="sup1"]
sup2 <- vacas$ganancia[vacas$suplemento=="sup2"]
sup3 <- vacas$ganancia[vacas$suplemento=="sup3"]
sup4 <- vacas$ganancia[vacas$suplemento=="control"]
```

```
Datos<-data.frame(Variable=c(sup1, sup2,sup3, sup4), Suplemento=factor(c(rep(1,length(sup1)),rep(2,length(sup2)),rep(3,length(sup3)),rep(4,length(sup4)))))
```

```
ANOVA<-aov(Variable~Suplemento, data=Datos)
summary(ANOVA)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Suplemento   3   91.9   30.627    3.823 0.0161 *
## Residuals    44  352.5    8.011
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
qf(0.90, 4-1, 44-4)
```

```
## [1] 2.226092
```

Al ser qf menor que F value podemos afirmar que se rechaza la hipotesis nula. Por lo tanto no afecta el tipo de suplemento que se use de manera significativa.

```
sup1 <- vacas$ganancia[vacas$dieta=="trigo"]
sup2 <- vacas$ganancia[vacas$dieta=="avena"]
sup3 <- vacas$ganancia[vacas$dieta=="cebada"]
```

```
Datos<-data.frame(Variable=c(sup1, sup2,sup3), Suplemento=factor(c(rep(1,length(sup1)),rep(2,length(sup2)),rep(3,length(sup3)))))
```

```
ANOVA<-aov(Variable~Suplemento, data=Datos)
summary(ANOVA)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Suplemento   2  287.2   143.59   41.11 7e-11 ***
## Residuals    45  157.2    3.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
qf(0.90, 3-1, 44-4)
```

```
## [1] 2.440369
```

al ser qf bastante inferior que F value afirmamos la hipotesis nula por lo tanto la dieta influye a la hora de las ganancias de peso

## Lab 6

### Ejercicio1

a)

No se ha encontrado el dataset `estudiantes_ausentes` dentro del zip de los diferentes datasets.

```
setwd("~/Escritorio/R/Datasets")
```

b)

### Ejercicio2

Para realizar este ejercicio primero crearemos una tabla con los totales por fila y por columna (d. marginales en x e y) y realizaremos un contraste  $\chi^2$  de independencia.

```
library(knitr)
xy <- matrix(c(510, 116, 37, 578, 78, 48, 10, 141, 175, 62, 18, 314), ncol=3, nrow=4, byrow=F);
colnames(xy)<- c("Europa", "America", "Africa")
rownames(xy)<- c("A", "B", "AB", "0")

xy <- as.table(xy)
kable(xy)
```

	Europa	America	Africa
A	510	78	175
B	116	48	62
AB	37	10	18
0	578	141	314

```
xy_a <- addmargins(xy)
kable(xy_a)
```

	Europa	America	Africa	Sum
A	510	78	175	763
B	116	48	62	226
AB	37	10	18	65
0	578	141	314	1033
Sum	1241	277	569	2087

```
ni <-xy_a[5,];ni
```

```
##  Europa America  Africa    Sum
##   1241     277    569   2087
```



```

nj<- xy_a[,4];nj

##      A      B      AB      0      Sum
## 763  226    65 1033 2087
N<-as.numeric(xy_a[5,4]);N

## [1] 2087
pxy<- xy^2;pxy

##      Europa America Africa
## A  260100    6084 30625
## B   13456    2304  3844
## AB   1369     100   324
## 0  334084   19881 98596
suma<-0

for(i in 1:3){
  for(j in 1:4){
    suma<-suma+as.numeric(pxy[j,i]/(ni[i]*nj[j]))
  }
}

CHI2<-N*(suma-1)
CHI2

## [1] 37.16161
gl<-(nrow(xy)-1)*(ncol(xy)-1); gl

## [1] 6
qchisq(0.95, gl)

## [1] 12.59159

```

Como observamos el valor 37.16 es mayor que el límite 12.59 el estadístico esta dentro de la RC y se rechaza la hipótesis de independencia

## Ejercicio 3:

Para ello usaremos la función `chisqtest` y así calcular a su vez la corrección de Yates.

```

library(knitr)
xy <- matrix(c(65, 43, 16, 12), ncol=2, nrow=2, byrow=F);
colnames(xy)<- c("MedicacionA", "MedicacionB")
rownames(xy)<- c("ConRecuperacion", "SinRecuperacion")

test<-chisq.test(xy, correct=TRUE);test

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  xy
## X-squared = 0.0058148, df = 1, p-value = 0.9392

```

```
qchisq(0.95, 1)
```

```
## [1] 3.841459
```

```
test$expected
```

```
##           MedicacionA MedicacionB
## ConRecuperacion    64.32353    16.67647
## SinRecuperacion    43.67647    11.32353
```

```
round(test$residuals^2, 2)
```

```
##           MedicacionA MedicacionB
## ConRecuperacion      0.01      0.03
## SinRecuperacion      0.01      0.04
```

El valor de  $\chi^2$  esta por debajo del límite calculado usando la función `qchisq` con los grados de libertad dados en el test. Por lo tanto aceptamos la hipótesis de independencia.

## Ejercicio 4:

```
library(knitr)
xy <- matrix(c(125, 16, 22, 31), ncol=2, nrow=2, byrow=F);
colnames(xy)<- c("aprobado", "suspendido")
rownames(xy)<- c("aprobado", "suspendido")
xy
```

```
##           aprobado suspendido
## aprobado      125          22
## suspendido     16          31
```

```
test<- mcnemar.test(xy, y=NULL, correct=TRUE);test
```

```
##
## McNemar's Chi-squared test with continuity correction
##
## data: xy
## McNemar's chi-squared = 0.65789, df = 1, p-value = 0.4173
test$residuals
```

```
## NULL
```

A la luz de los resultados teniendo un nivel de significancia del 5% podemos afirmar que rechazamos la hipótesis nula por lo tanto la evaluación de los profesores es diferente.

## Ejercicio 5

A)

```
d1 <- c(9, 9, 9, 8, 7, 10, 9, 9, NA)
d2 <- c(8, 9, 8, 6, 7, 8, 9, NA, NA)
d3 <- c(10, 12, 10, 8, 9, 11, 10, 9, 8)
```

```
data<-data.frame(valores = c(d1,d2,d3), condicion=c("d1","d1","d1","d1","d1","d1","d1","d1","d1","d1","d2","d2","d2","d2","d2","d2","d2","d2","d2","d2","d3","d3","d3","d3","d3","d3","d3","d3","d3","d3"))
```

```
data$condicion

## [1] "d1" "d1" "d1" "d1" "d1" "d1" "d1" "d1" "d1" "d2" "d2" "d2" "d2" "d2" "d2"
## [16] "d2" "d2" "d2" "d3" "d3" "d3" "d3" "d3" "d3" "d3" "d3" "d3" "d3"

kruskal.test(data$valores~data$condicion, data=data)

##
## Kruskal-Wallis rank sum test
##
## data: data$valores by data$condicion
## Kruskal-Wallis chi-squared = 7.3821, df = 2, p-value = 0.02495
```

Como podemos observar el test encuentra significancia en la diferencia de almenos dos grupos.

## B)

```
pairwise.wilcox.test(data$valores, data$condicion, p.adjust.method = "holm")

## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot compute
## exact p-value with ties

## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot compute
## exact p-value with ties

## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot compute
## exact p-value with ties

##
## Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data: data$valores and data$condicion
##
##      d1      d2
## d2 0.192 -
## d3 0.192 0.051
##
## P value adjustment method: holm
```

Esos grupos que hablabamos antes se pueden observar usando comparaciones por pares en las cuales se reflejan diferencias significativas entre d2 y d3.