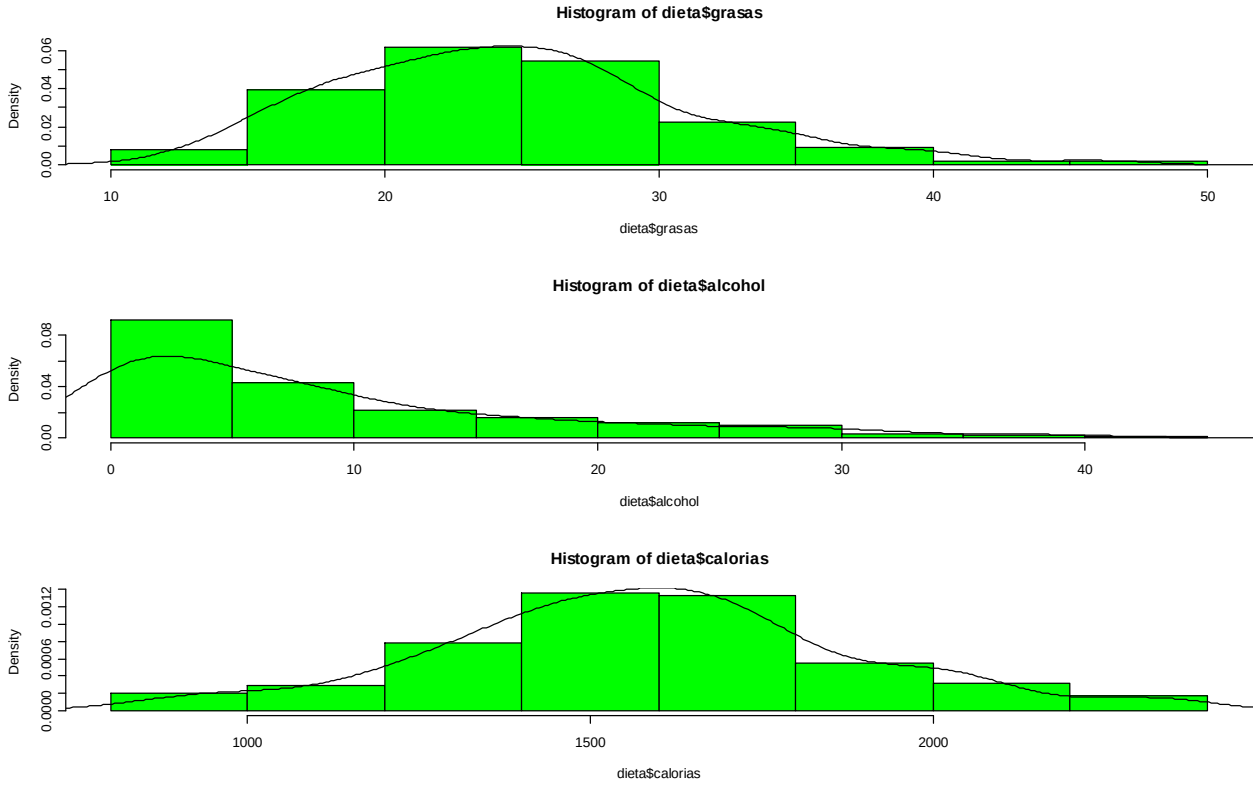
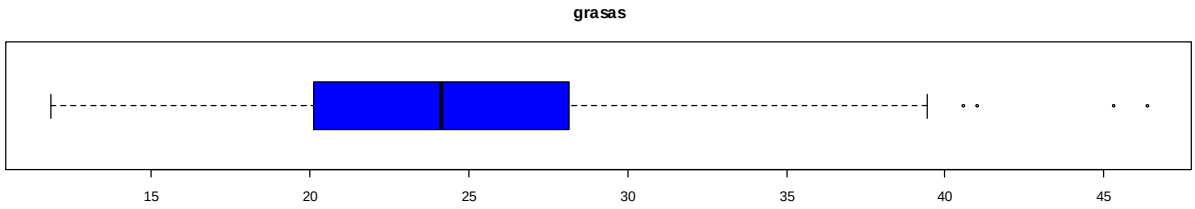
```
# Asignamos un nombre al archivo

dieta<-read.csv2(file.choose(),header=T)
class(dieta)
[1] "data.frame"

# Reemplazamos los valores faltantes (999.9) por NA

dieta$grasas[172]<-NA
dieta$alcohol[22]<-NA
dieta$alcohol[25]<-NA
dieta$alcohol[163]<-NA

# Calculamos valores resumen

# la media
apply(dieta[,1:3],2,mean,na.rm=T)
   grasas   alcohol  calorias
  24.7743    8.8320 1584.5549

# la mediana
apply(dieta[,1:3],2,median,na.rm=T)
   grasas  alcohol calorias
  24.130    5.905 1585.000

# los cuartiles
apply(dieta[,1:3],2,quantile,na.rm=T)
      grasas alcohol calorias
0%   11.8200   0.000      800
25%  20.1675   1.780     1400
50%  24.1300   5.905     1585
75%  28.0950  12.965     1761
100% 46.3600  40.110     2376

# la distancia intercuartil
apply(dieta[,1:3],2,quantile,na.rm=T)[4,]-
apply(dieta[,1:3],2,quantile,na.rm=T)[2,]
  grasas  alcohol calorias
  7.9275  11.1850 361.0000

# la varianza
apply(dieta[,1:3],2,var,na.rm=T)
     grasas     alcohol    calorias
   42.69246    84.52749 93074.41121

# la mad
apply(dieta[,1:3],2,mad,na.rm=T)
    grasas    alcohol   calorias
  5.922987   7.346283 271.315800

# las medidas resumen
apply(dieta[,1:3],2,summary,na.rm=T)
$grasas
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  11.82   20.17   24.13   24.77   28.10   46.36    1.00

$alcohol
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  0.000   1.780   5.905   8.832  12.960  40.110   3.000
```

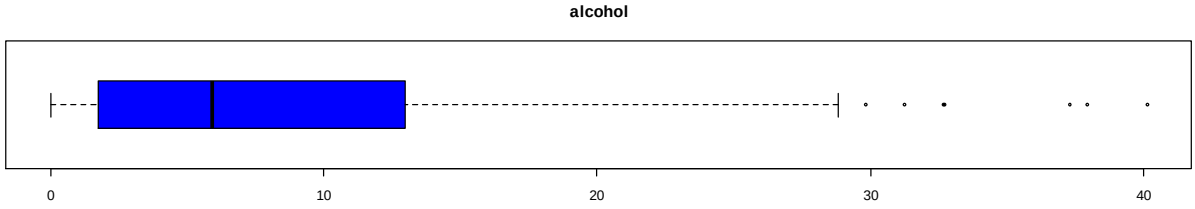
```
$calorias
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    800    1400    1585    1585    1761    2376
```

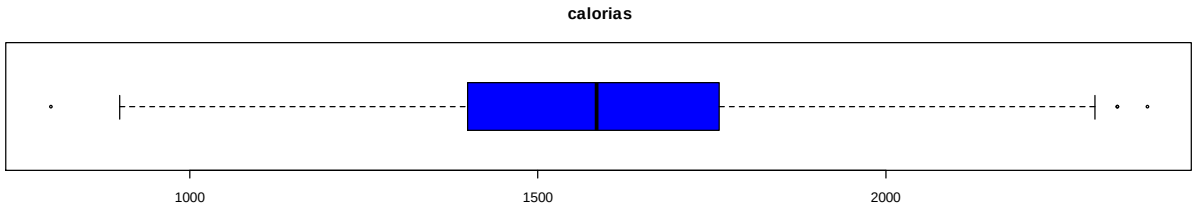
```
# graficamos los 3 histogramas junto con la estimación de la densidad
par(mfrow=c(3,1))
hist(dieta$grasas,probability=T,col="green")
lines(density(dieta$grasas,na.rm=T))
hist(dieta$alcohol,probability=T,col="green")
lines(density(dieta$alcohol,na.rm=T))
hist(dieta$calorias,probability=T,col="green")
lines(density(dieta$calorias,na.rm=T))
```




```
# graficamos los 3 boxplots
boxplot(dieta$grasas,col="blue",horizontal=T)
title("grasas")
boxplot(dieta$alcohol,col="blue",horizontal=T)
title("alcohol")
boxplot(dieta$calorias,col="blue",horizontal=T)
title("calorias")
```



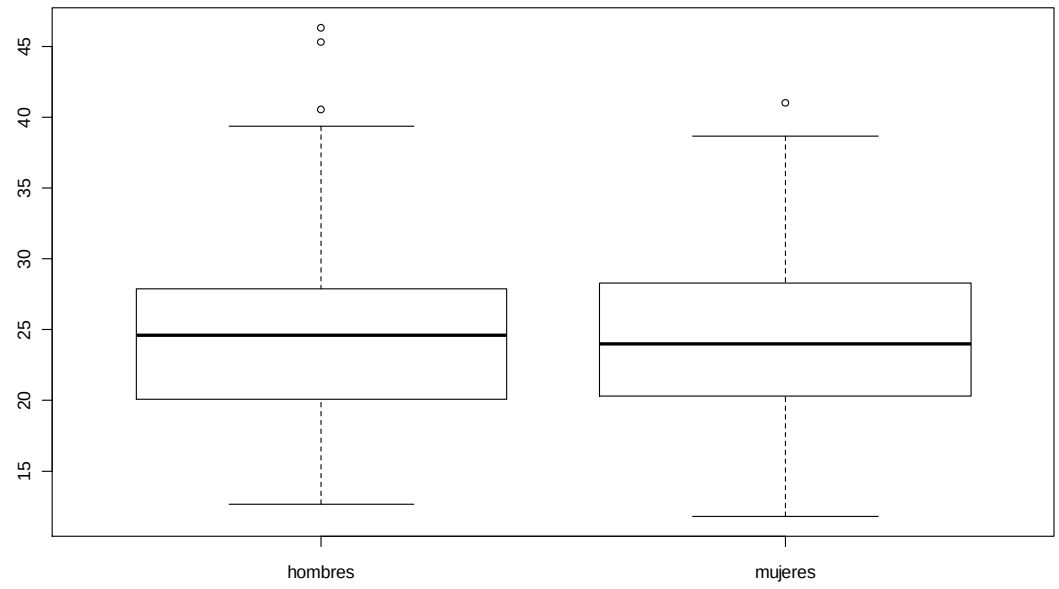




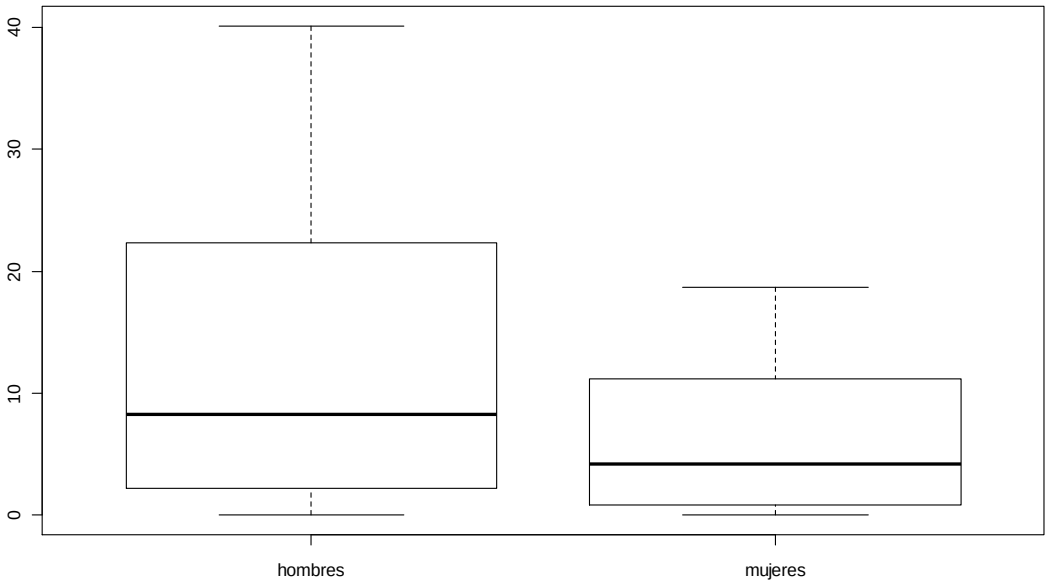

```
# los diagramas tallo hoja
stem(dieta$grasas)
  The decimal point is at the |

  10 | 8
  12 | 7
  14 | 044672349
  16 | 000677903567779
  18 | 013666688890788
  20 | 1124457779034446668899
  22 | 00234562344455666678
  24 | 01124456688112335679
  26 | 02334445556880114556789
  28 | 000333667124666
  30 | 0902444
  32 | 029233
  34 | 0123364
  36 | 234
  38 | 7734
  40 | 60
  42 |
  44 | 3
  46 | 4
```

```
stem(dieta$alcohol)

  The decimal point is at the |

   0 | 000000000000000011444456666888888899013446888899
   2 | 022223366789001445788
   4 | 224566678890222448
   6 | 0067890012338888
   8 | 11222335784579
  10 | 01349122
  12 | 025901689
  14 | 214677
  16 | 562389
  18 | 74
  20 | 1578
  22 | 34794
  24 | 07
  26 | 369
  28 | 3388
  30 | 3
  32 | 67
  34 |
  36 | 39
  38 |
  40 | 1

stem(dieta$calorias)

  The decimal point is 2 digit(s) to the right of the |

   8 | 0
   9 | 01246
  10 | 033
  11 | 012568
  12 | 00223356899
  13 | 00122333466777899
  14 | 000233344555556677788999
  15 | 000122334566666678899999
  16 | 11111222334445567788888999
  17 | 000001223334666777889
  18 | 11222378899
  19 | 44577899
  20 | 113455599
  21 | 04
  22 | 69
  23 | 0338

# separamos la base de datos en dos grupos según el sexo y asignamos un nombre a
cada uno
split(dieta,dieta$sexo)
hombres<- split(dieta,dieta$sexo)$M
mujeres<- split(dieta,dieta$sexo)$F

#calculamos valores resumen por separado
apply(hombres[,1:3],2,summary,na.rm=T)
$grasas
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  12.71   20.06   24.61   25.01   27.92   46.36    1.00
```

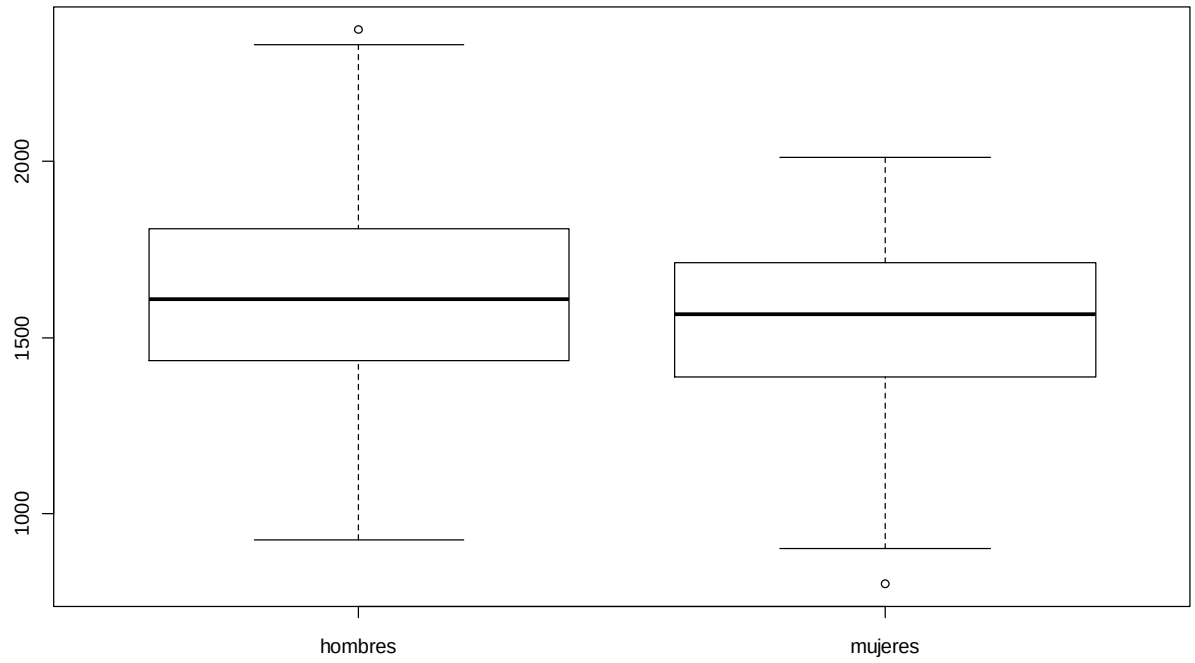
```
$alcohol
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00    2.20    8.26   12.44   22.31   40.11    1.00

$calorias
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    923    1435    1610    1642    1804    2376

apply(mujeres[,1:3],2,summary,na.rm=T)
$grasas
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  11.82   20.30   23.98   24.59   28.30   41.01

$alcohol
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  0.000   0.800   4.210   5.844  11.150  18.690   2.000

$calorias
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    800    1388    1568    1537    1714    2013

#calculamos varianzas por separado
> apply(hombres[,1:3],2,var,na.rm=T)
      grasas      alcohol     calorias
    48.20431    123.75365 113291.16217
> apply(mujeres[,1:3],2,var,na.rm=T)
     grasas     alcohol    calorias
   38.61012    33.11029 72486.33438
#realizamos boxplots para cada grupo
boxplot(hombres$grasas,mujeres$grasas,names=c("hombres","mujeres"))
```




```
boxplot(hombres$alcohol,mujeres$alcohol,names=c("hombres","mujeres"))
```




```
boxplot(hombres$calorias,mujeres$calorias,names=c("hombres","mujeres"))
```




```
# definimos una variable categórica según el nivel de calorías
cal.cat<-rep(0,173)
for(i in 1:173) if(dieta$calorias[i]<1100) cal.cat[i]<-1
for(i in 1:173) if(dieta$calorias[i]>=1100 & dieta$calorias[i]<1700)
cal.cat[i]<-2
for(i in 1:173) if(dieta$calorias[i]>=1700) cal.cat[i]<-3
```

```
cal.cat<-as.factor(cal.cat)

#armamos un nuevo data frame con esa columna y la variable alcohol
alcohol<-data.frame(dieta$alcohol,cal.cat)

#separamos en 3 grupos según el consumo de calorías y asignamos nombres
split(alcohol,cal.cat)
cal1<- split(alcohol,cal.cat)$`1`
cal2<- split(alcohol,cal.cat)$`2`
cal3<- split(alcohol,cal.cat)$`3`

# calculamos medidas resumen en cada categoria
summary(cal1$dieta.alcohol,na.rm=T)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
 0.0300  0.1175  0.7600  0.7288  1.3000  1.4300  2.0000
summary(cal2$dieta.alcohol,na.rm=T)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   0.925   3.700   4.133   7.055  10.860
summary(cal3$dieta.alcohol,na.rm=T)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   13.36   17.76   19.15   24.85   40.11    1.00

# graficamos los boxplots
boxplot(cal1$dieta.alcohol, cal2$dieta.alcohol,
cal3$dieta.alcohol,names=c("nivel 1", "nivel 2", "nivel 3"))
```