

EDA1

David Criado Ramón

7/11/2019

1. Exploratory Data Analysis.

a. Ejemplo 1, hip dataset.

Descárgate el dataset hip con el siguiente comando:

```
hip <- read.table("http://astrostatistics.psu.edu/datasets/HIP_star.dat",
                  header=T, fill=T)
```

Una vez descargado, comprueba la dimensión y los nombres de las columnas del dataset. ¿Qué dimensión tiene? ¿qué tipos de datos alberga?

```
dim(hip)
```

```
## [1] 2719     9
sapply(hip, class)
```

```
##      HIP      Vmag       RA       DE      Plx      pmRA      pmDE
## "integer" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##      e_Plx      B.V
## "numeric" "numeric"
```

El dataset contiene 2719 muestras y tiene 9 atributos. Todos los datos son de tipo numérico, con 8 de los atributos números reales (variables continuas) y uno de ellos, HIP, son números enteros (variable discreta).

```
head(hip)
```

```
##   HIP  Vmag       RA       DE    Plx      pmRA      pmDE e_Plx  B.V
## 1   2  9.27 0.003797 -19.498837 21.90  181.21   -0.93  3.10  0.999
## 2  38  8.65 0.111047 -79.061831 23.84  162.30   -62.40  0.78  0.778
## 3  47 10.78 0.135192 -56.835248 24.45  -44.21 -145.90  1.97  1.150
## 4  54 10.57 0.151656  17.968956 20.97  367.14  -19.49  1.71  1.030
## 5  74  9.93 0.221873  35.752722 24.22  157.73  -40.31  1.36  1.068
## 6  81  8.57 0.243864 -4.932115 23.43 -184.70 -172.67  1.28  0.642
```

Muestra por pantalla la columna de la variable RA.

Sólo voy a mostrar los 100 primeros elementos de la columna, para que no ocupe demasiado espacio en el PDF.

```
head(hip$RA, n=100)
```

```
## [1]  0.003797 0.111047 0.135192 0.151656 0.221873 0.243864 0.348708
## [8]  0.426746 0.455182 0.478685 0.612287 0.696411 0.972063 1.099309
## [15] 1.102623 1.244275 1.281668 1.369764 1.423333 1.468617 1.843365
## [22] 1.966150 2.261459 2.315143 2.352249 2.431558 2.768701 2.878592
## [29] 2.898287 2.906145 3.125492 3.136756 3.287636 3.470117 3.499989
## [36] 3.517613 3.542473 3.561294 3.584257 3.588865 3.768990 4.047722
## [43] 4.101227 4.179702 4.342874 4.375348 4.376308 4.386210 4.387165
```

```

## [50] 4.401851 4.457699 4.582106 4.608357 4.901928 5.280108 5.398303
## [57] 5.604551 5.644953 5.735995 5.750338 5.770342 5.907064 5.913140
## [64] 5.947756 6.036948 6.091006 6.256391 6.266819 6.285408 6.667964
## [71] 6.930394 6.957396 7.292119 7.368555 7.427643 7.465869 7.575263
## [78] 7.608228 7.689831 7.801700 7.885670 8.033768 8.182476 8.613389
## [85] 8.622360 8.712882 8.739422 8.807178 8.908298 8.967503 8.970857
## [92] 9.007174 9.016931 9.028761 9.156804 9.342176 9.487066 9.561303
## [99] 9.638220 9.803365

```

Calcula las tendencias centrales de todos los datos del dataset (mean, mediana) utilizando la function apply.

```

print("Medias")

## [1] "Medias"

sapply(hip, mean, na.rm=T)

##          HIP        Vmag         RA          DE          Plx
## 56549.4828981    8.2593858  173.4529975 -0.1397663  22.1980213
##          pmRA       pmDE        e_Plx        B.V
## 5.3761346   -63.9419934   1.6267929   0.7615299

print("Medianas")

## [1] "Medianas"

sapply(hip, median, na.rm=T)

##          HIP        Vmag         RA          DE          Plx
## 56413.000000    8.2800000  173.369788  3.254234  22.100000
##          pmRA       pmDE        e_Plx        B.V
## 10.550000   -49.480000   1.140000   0.710500

```

Haz lo mismo para las medidas de dispersión mínimo y máximo. ¿Sería posible hacerlo con un único comando? ¿Qué hace la función range()?

```

print("Mínimos")

## [1] "Mínimos"

sapply(hip, min, na.rm=T)

##          HIP        Vmag         RA          DE          Plx
## 2.000000    0.450000  0.003797 -87.202730 20.000000
##          pmRA       pmDE        e_Plx        B.V
## -868.010000 -1392.300000  0.450000 -0.158000

print("Máximos")

## [1] "Máximos"

sapply(hip, max, na.rm=T)

##          HIP        Vmag         RA          DE          Plx
## 120003.00000 12.74000  359.95468  88.30268 25.000000
##          pmRA       pmDE        e_Plx        B.V
## 781.34000   481.19000  46.91000   2.80000

```

Sí, mediante la función **range**, que nos devuelve un vector conteniendo los mínimos (primer elemento de cada vector) y los máximos (segundo elemento de cada vector).

```
print("Mínimos y máximos")

## [1] "Mínimos y máximos"
sapply(hip, range, na.rm=T)

##          HIP   Vmag         RA        DE    Plx     pmRA     pmDE e_Plx      B.V
## [1,] 2 0.45 0.003797 -87.20273 20 -868.01 -1392.30 0.45 -0.158
## [2,] 120003 12.74 359.954685 88.30268 25 781.34 481.19 46.91 2.800
```

Sin embargo, las medidas más populares de dispersión son la varianza (**var()**), su desviación estándar (**sd()**) y la desviación absoluta de la mediana o MAD. Calcula estas medidas para los valores de RA.

```
print("Varianza, Desviación est\'andar, Desviaci\'on absoluta de la mediana")

## [1] "Varianza, Desviaci\'on est\'andar, Desviaci\'on absoluta de la mediana"
c(var(hip$RA, na.rm=T), sd(hip$RA, na.rm=T), mad(hip$RA, na.rm=T))

## [1] 11566.3248 107.5468 146.9334
```

Imagina que quieres calcular dos valores de una sola vez. ¿Te serviría este código?

```
f = function(x) c(median(x), mad(x))
f(hip[,3])
```

```
## [1] 173.3698 146.9334
```

Sí que sirve. El código encapsula la llamada a las funciones y nos permite aplicar ambas a un vector o una columna de un data frame. Podría ser problemática si tenemos “missing values” pues no se ha indicado que se ignoren dichos valores.

¿Cuál sería el resultado de aplicar **apply(hip,2,f)**?

```
apply(hip, 2, f)

##          HIP   Vmag         RA        DE    Plx     pmRA     pmDE
## [1,] 56413.00 8.280000 173.3698 3.254234 22.100000 10.5500 -49.48000
## [2,] 49090.37 1.882902 146.9334 43.984032 1.764294 141.6476 99.49729
##          e_Plx      B.V
## [1,] 1.140000    NA
## [2,] 0.489258    NA
```

Sería el resultado de aplicar la mediana y la desviación absoluta de la mediana a cada columna. Nótese que puesto que había “missing values” en el atributo B.V aparece NA como resultado de estas operaciones.

Vamos a medir la dispersión de la muestra utilizando el concepto de cuartiles. El percentil 90 es aquel dato que excede en un 10% a todos los demás datos. El cuartil (quantile) es el mismo concepto, solo que habla de proporciones en vez de porcentajes. De forma que el percentil 90 es lo mismo que el cuartil 0.90. La mediana “median” de un dataset es el valor más central, en otras palabras exactamente la mitad del dataset excede la media. Calcula el cuartil .10 y .50 para la columna RA del dataset hip. Sugerencia: quantile()

```
quantile(hip$RA, c(0.1, 0.5))
```

```
##      10%      50%
## 28.92324 173.36979
```

Los cuantiles 0.25 y 0.75 se conocen como el first quartile y el third quartile, respectivamente. Calcula los cuatro cuartiles para RA con un único comando.

```
quantile(hip$RA)
```

```
##      0%      25%      50%      75%      100%
## 0.003797 70.141368 173.369788 266.923319 359.954685
```

Otra medida de dispersion es la diferencia entre el primer y el tercer cuartil conocida como rango intercuartil (IQR) Inter Quantile Range. ¿Obtienes ese valor con la función summary()?

```
summary(hip)
```

```
##      HIP          Vmag          RA          DE
## Min.   : 2   Min.   :0.450   Min.   : 0.0038   Min.   :-87.2027
## 1st Qu.:21770 1st Qu.: 7.050   1st Qu.: 70.1414   1st Qu.:-31.3635
## Median :56413  Median : 8.280   Median :173.3698   Median : 3.2542
## Mean   :56549  Mean   : 8.259   Mean   :173.4530   Mean   : -0.1398
## 3rd Qu.:87096 3rd Qu.: 9.610   3rd Qu.:266.9233   3rd Qu.: 28.0705
## Max.   :120003 Max.   :12.740   Max.   :359.9547   Max.   : 88.3027
##
##      Plx          pmRA          pmDE          e_Plx
## Min.   :20.00  Min.   :-868.010  Min.   :-1392.30  Min.   : 0.450
## 1st Qu.:20.98  1st Qu.:-91.980  1st Qu.:-130.79  1st Qu.: 0.870
## Median :22.10  Median : 10.550  Median : -49.48  Median : 1.140
## Mean   :22.20  Mean   :  5.376  Mean   : -63.94  Mean   : 1.627
## 3rd Qu.:23.36  3rd Qu.: 103.870 3rd Qu.:    8.57  3rd Qu.: 1.680
## Max.   :25.00  Max.   : 781.340  Max.   : 481.19  Max.   :46.910
##
##      B.V
## Min.   :-0.1580
## 1st Qu.: 0.5600
## Median : 0.7105
## Mean   : 0.7615
## 3rd Qu.: 0.9530
## Max.   : 2.8000
## NA's   :41
```

No obtenemos esa información directamente pero si que tenemos el valor del primer y tercer cuartil, por lo que podemos restar el tercer con el primer cuartil para obtener el IQR.

Hasta ahora has ignorado la presencia de valores perdidos NA. La función any() devuelve TRUE si se encuentra al menos un TRUE en el vector que damos como argumento. Su combinación con is.na es muy útil. ¿qué obtienes cuando ejecutas el siguiente comando? ¿Cómo lo interpretas?

```
hasNA = function(x) any(is.na(x))
apply(hip, 2, hasNA)

##   HIP Vmag   RA   DE   Plx pmRA pmDE e_Plx   B.V
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
```

Al aplicar el comando la función hasNA es aplicada a cada columna, por lo que la salida nos indicará los nombre de las columnas y pondrá TRUE en aquellas que contengan “missing values” y false en el resto.

Prueba a ejecutar el siguiente comando.

```
min(hip$B.V)
```

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

Nos devuelve NA, porque existen “missing values” en la columna, hemos de indicar na.rm=T para que los ignore.

hip1 = na.omit(hip) Como has observado nos devuelve NA para toda la columna, normalmente queríamos poder usar la función sobre el resto de datos que no son NA: Para ello podemos utilizar la función na.omit. ¿Que ocurre cuando lo hacemos?. Usando apply calcula la media para hip. Intenta calcular la media de forma que solo cambie la de B.V cuando ignores los valores NA

```
apply(hip, 2, mean)

##          HIP        Vmag         RA         DE         Plx
## 56549.4828981  8.2593858 173.4529975 -0.1397663 22.1980213
##          pmRA       pmDE       e_Plx       B.V
## 5.3761346   -63.9419934    1.6267929        NA

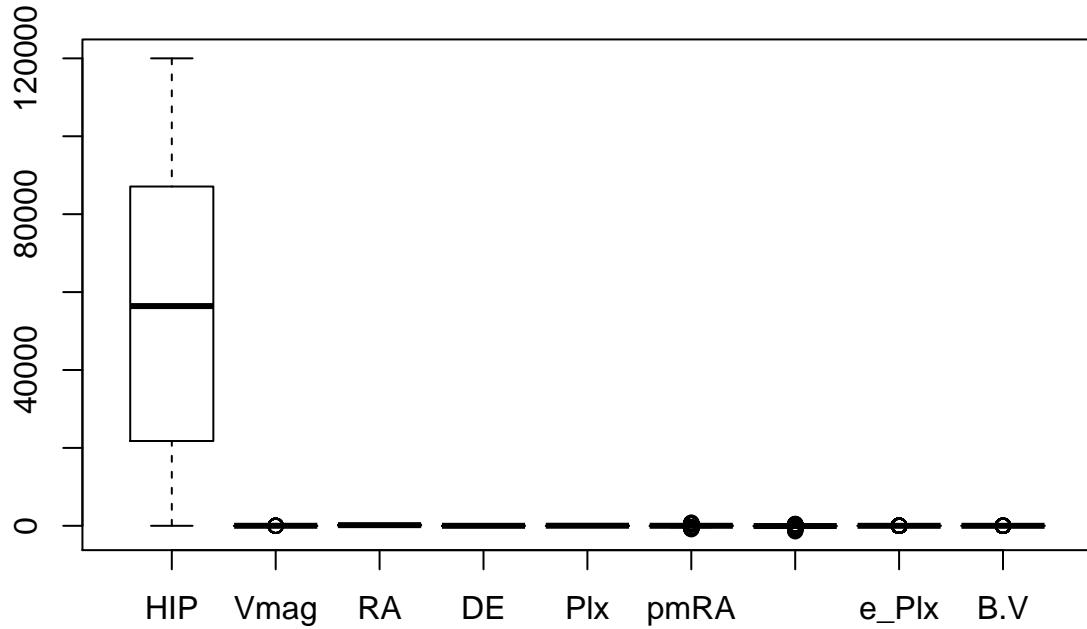
hip1 = na.omit(hip)
apply(hip1, 2, mean)

##          HIP        Vmag         RA         DE         Plx
## 56575.8050784  8.2147797 173.5284087 -0.2743560 22.1954033
##          pmRA       pmDE       e_Plx       B.V
## 5.5370575   -63.5345892    1.5449552    0.7615299
```

Al haber aplicado na.omit se eliminan las muestras que contienen na, por lo que el valor de la media de B.V ahora va a ser numérico en vez de NA.

Obtén una idea aproximada de tus datos mediante la creación de un boxplot del hip dataset.

```
boxplot(hip)
```

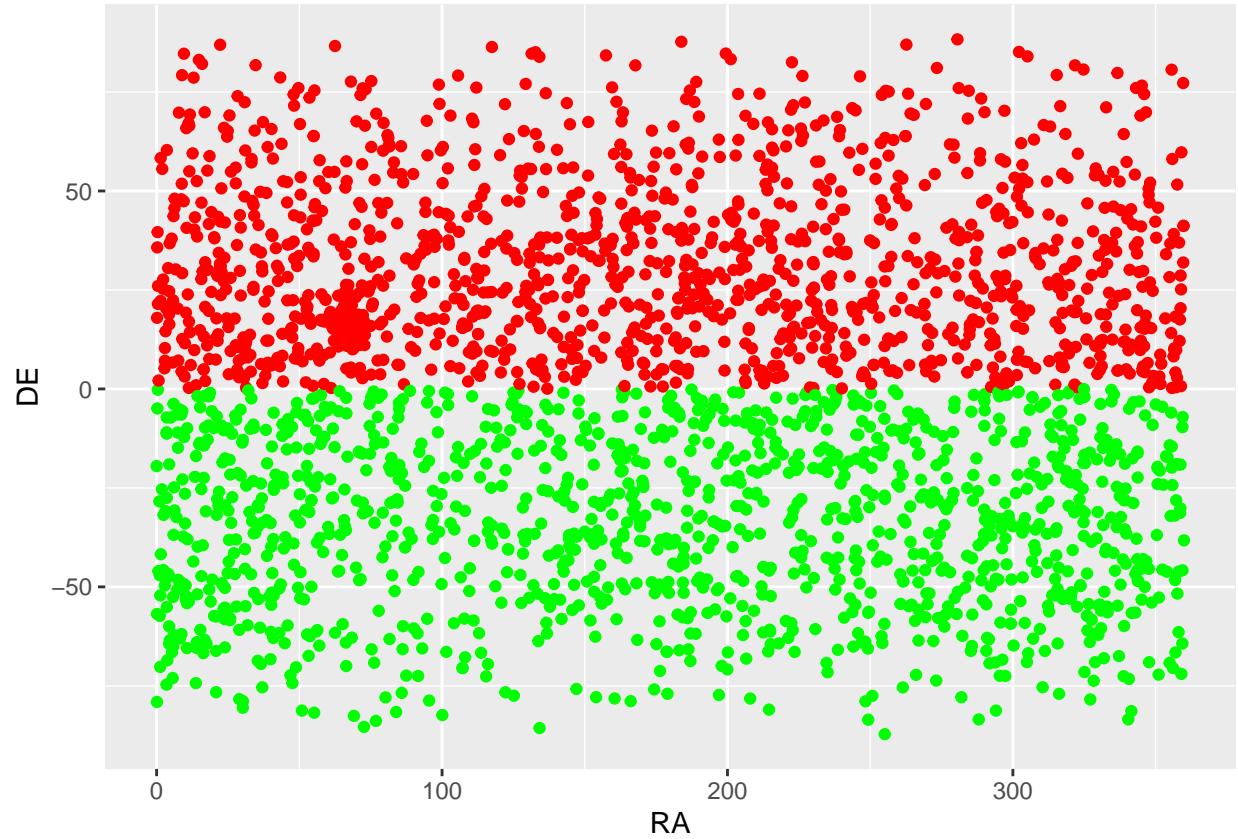


Podemos observar que HIP tiene una escala muy distinta del resto de variables y pmRA y pmDE tienen una cantidad abundante de “outliers”.

Crea un scatterplot que te compare los valores de RA y DE. Representa los puntos con el símbolo ‘?’ Y que estos puntos sean de color rojo si DE excede de 0. Sugerencia: ifelse()

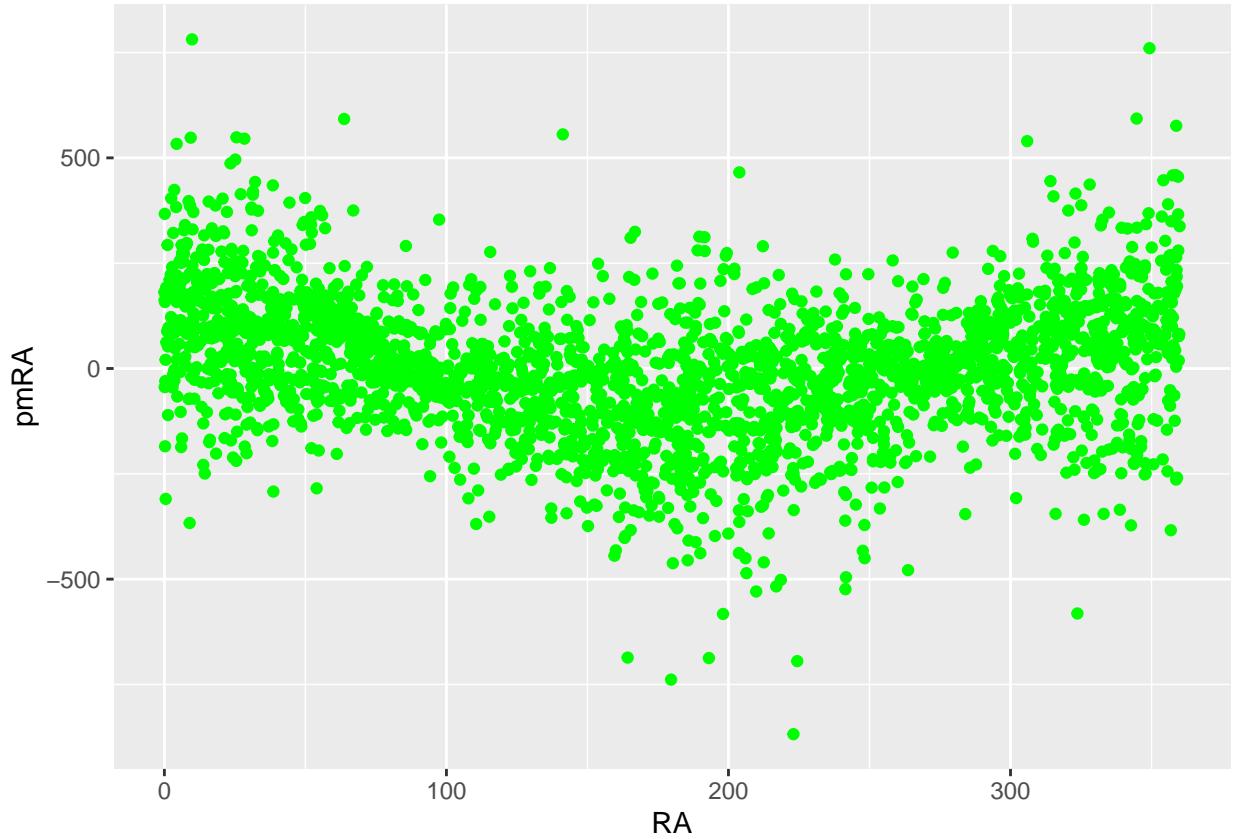
Vamos a representar de verde los puntos que tengan $DE \leq 0$ y de rojo los puntos que tenga $DE > 0$.

```
ggplot(hip, aes(x=RA, y=DE)) +
  geom_point(col=ifelse(hip$DE > 0, "red", "green"))
```



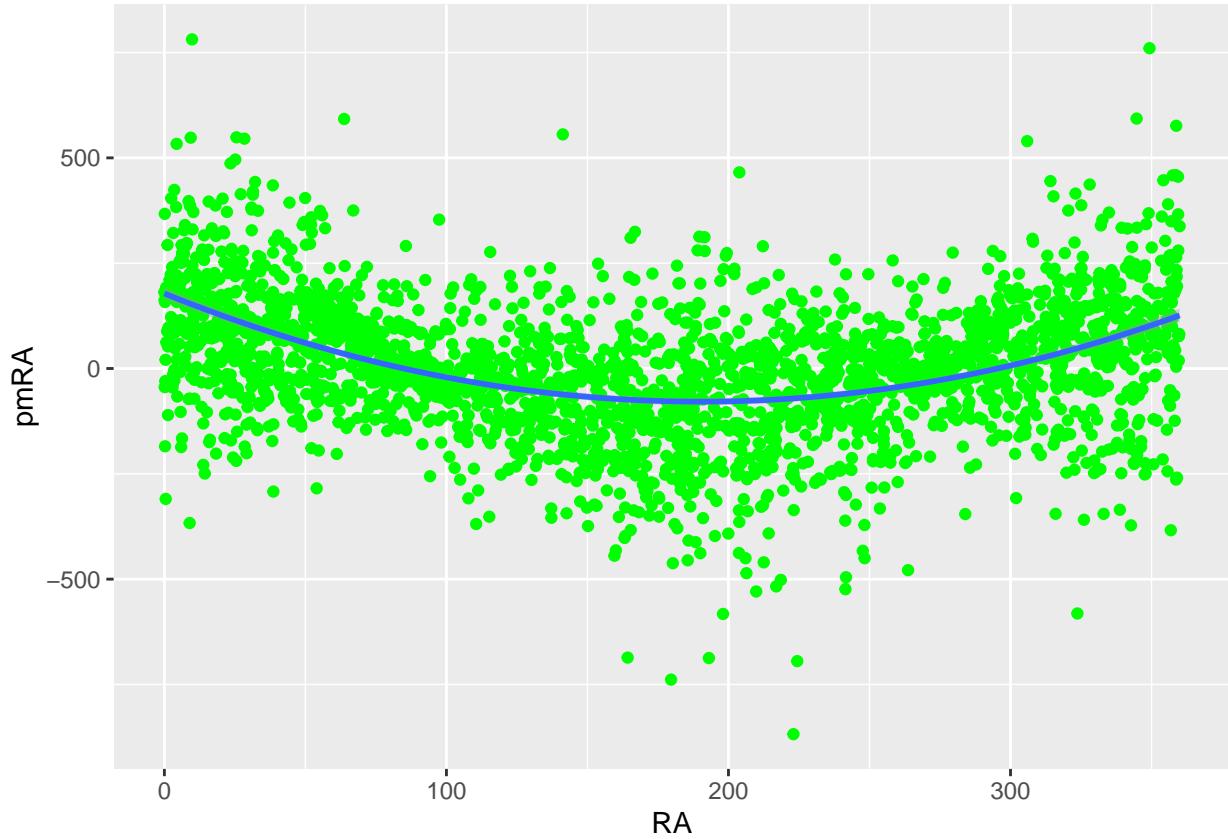
Haz un scatterplot de RA y pmRA. ¿Ves algún patrón?

```
ggplot(hip, aes(x=RA, y=pmRA)) + geom_point(col="green")
```



Parece que los puntos siguen una parábola. Observémoslo trazando una con geom_smooth.

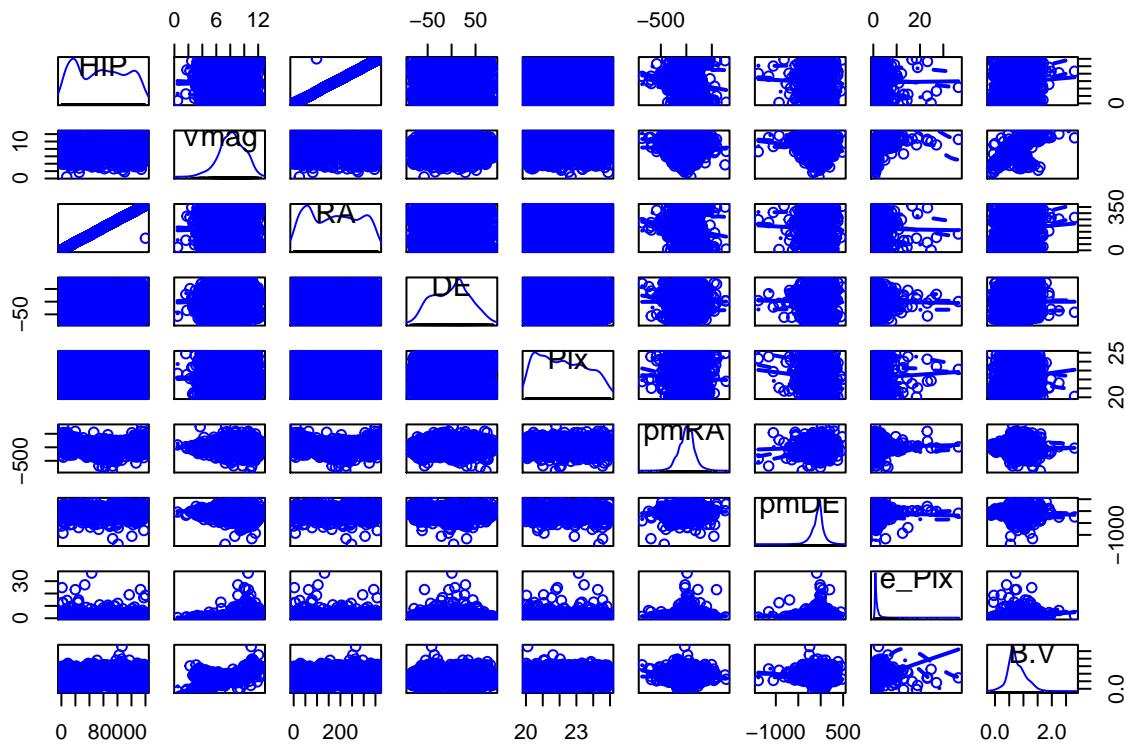
```
ggplot(hip, aes(x=RA, y=pmRA)) + geom_point(col="green") +  
  geom_smooth(method=lm, formula = y ~ x + I(x^2))
```



Parece que la distribución de los puntos sigue la trayectoria de la parábola trazada, pero hay presentes un número considerable de “outliers”.

En vez de crear los plots por separado para cada par de columnas, hazlos con un solo comando con el scatterplot matrix.

```
scatterplotMatrix(hip)
```



No se ve mucho en el PDF de este scatterplex hay demasiadas variables, si desea analizarlo es mejor visualizarlo en R y hacer zoom, probablemente para mostrarlo en un informe es necesario reducir el número de datos a comparar de forma simultánea.

Para poder acceder a las variables por su nombre usa `attach(hip)`. Vamos a seleccionar las estrellas Hyades del dataset aplicando los siguientes filtros:

- RA in the range (50,100)
- DE in the range (0,25)
- pmRA in the range (90,130)
- pmDE in the range (-60,-10)
- $e_{Plx} < 5$
- $Vmag > 4$ OR $B.V < 0.2$ (this eliminates 4 red giants)

Voy a hacerlo en el siguiente ejercicio utilizando `dplyr` y sin usar `attach`

Crea un nuevo dataset con la aplicación de estos filtro. El Nuevo dataset se llama hyades. ¿Que dimensiones tiene? Grafica un scatterplot de Vmag vs B.V

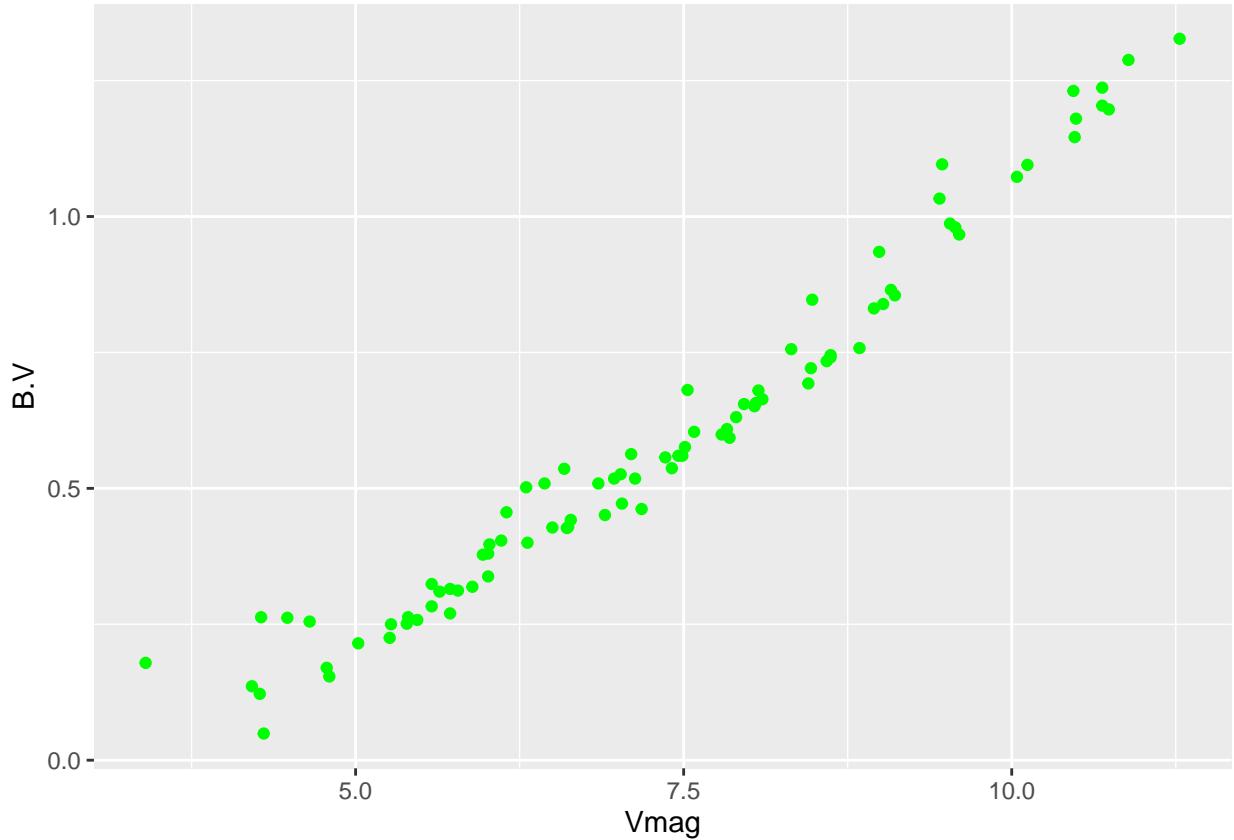
```
hyades <- hip %>% filter(between(RA, 50, 100) &
                           between(DE, 0, 25) &
                           between(pmRA, 90, 130) &
                           between(pmDE, -60, -10) &
                           e_Plx < 5 &
                           (Vmag > 4 | B.V < 0.2))

dim(hyades)
```

```
## [1] 88  9
```

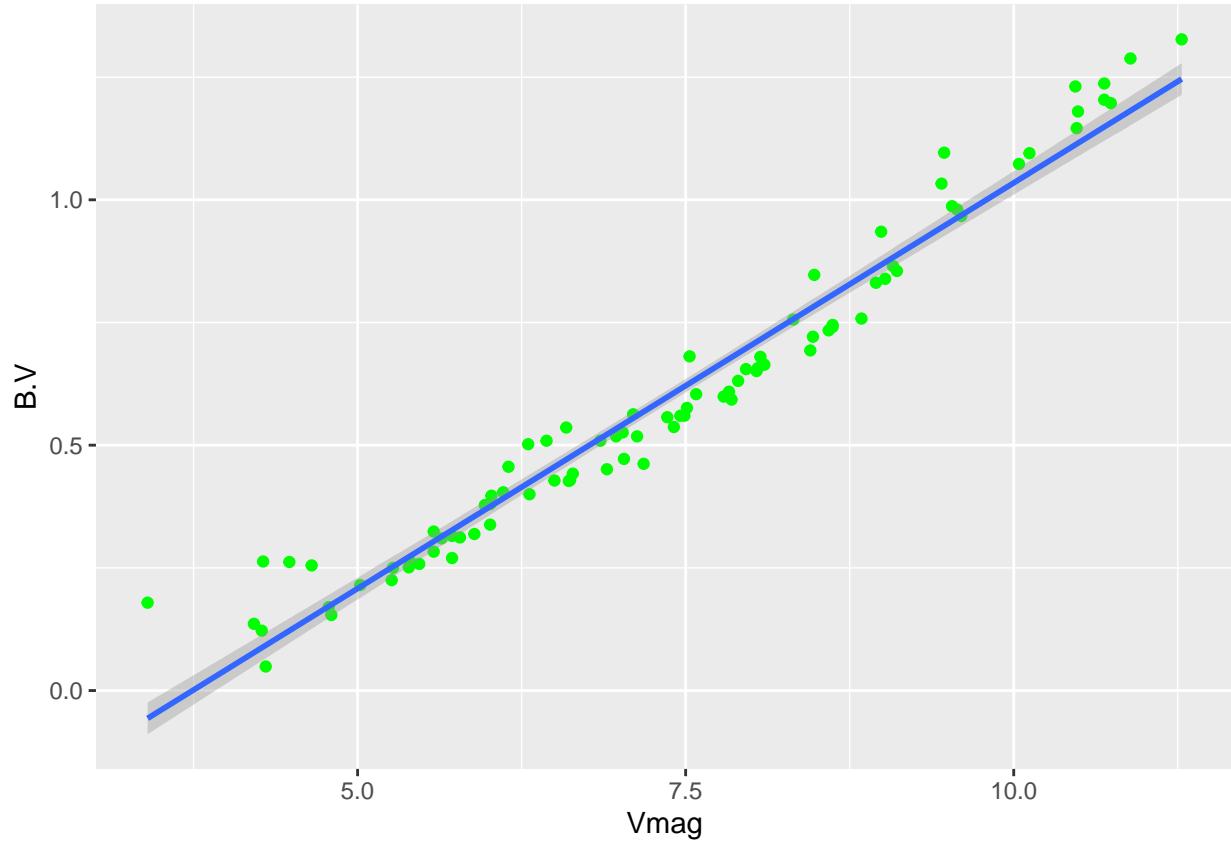
El nuevo dataset tiene 88 muestras y 9 atributos para cada muestra.

```
ggplot(hyades, aes(x=Vmag, y=B.V)) + geom_point(col="green")
```



En este caso parece que ambas variables siguen una relación lineal, utilicemos geom_smooth para observarlo con la recta.

```
ggplot(hyades, aes(x=Vmag, y=B.V)) + geom_point(col="green") +  
  geom_smooth(method="lm")
```



b. Ejemplo 2, iris dataset.

Vamos a utilizar el ejemplo del dataset iris que está incluido en la distribución de R. Este dataset fue creado por Douglas Fisher. Consta de tres clases y tipos de 3 clases de tipos de flores: *setosa*, *virginica* y *versicolor*. Cada una de ellas con cuatro atributos: *sepal width*, *sepal length*, *petal width*, *petal length*.

Inspecciona las primeras filas del dataset y calcula el summary() del mismo con cada atributo del dataset.

```
head(iris)

##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1       3.5        1.4       0.2  setosa
## 2          4.9       3.0        1.4       0.2  setosa
## 3          4.7       3.2        1.3       0.2  setosa
## 4          4.6       3.1        1.5       0.2  setosa
## 5          5.0       3.6        1.4       0.2  setosa
## 6          5.4       3.9        1.7       0.4  setosa
```

```
summary(iris)

##   Sepal.Length     Sepal.Width      Petal.Length      Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100  1st Qu.:2.800  1st Qu.:1.600  1st Qu.:0.300
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
```

```

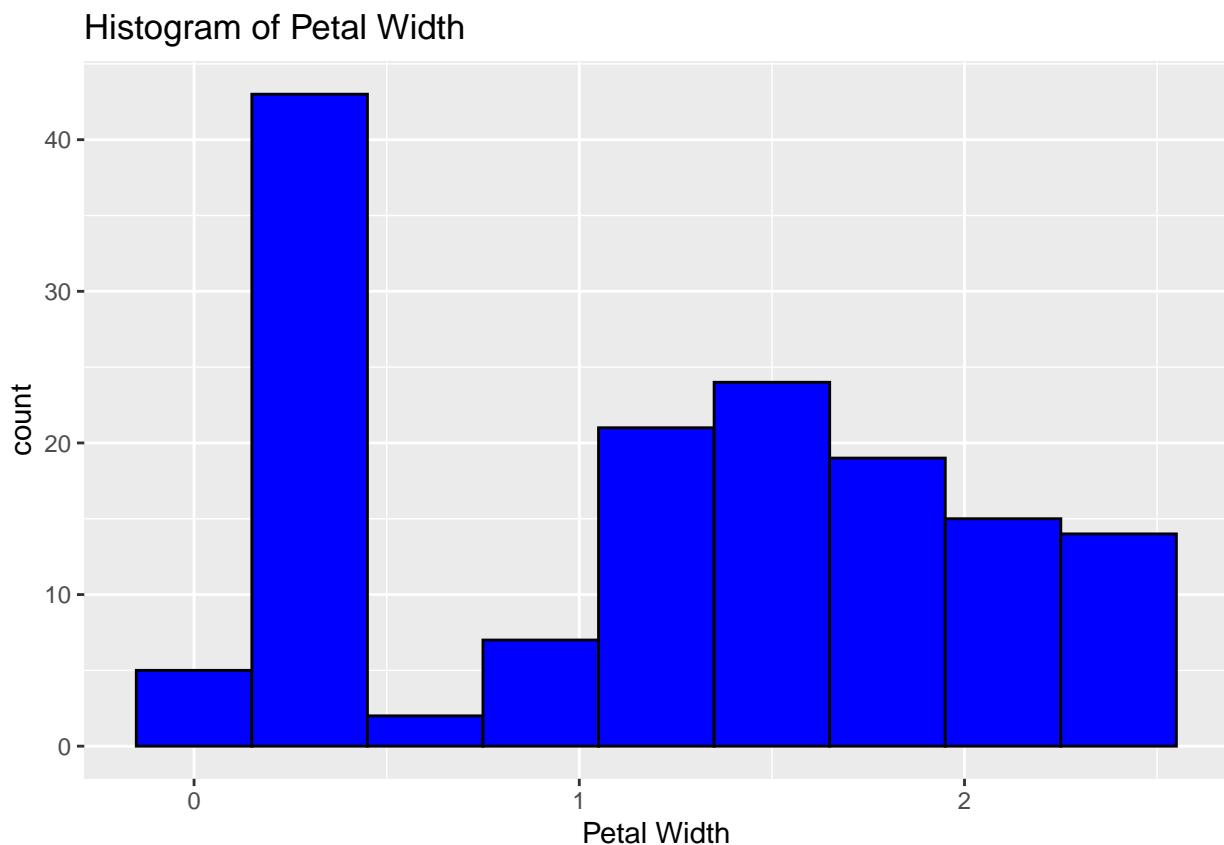
##   Max.    :7.900   Max.    :4.400   Max.    :6.900   Max.    :2.500
##   Species
##   setosa     :50
##   versicolor:50
##   virginica :50
##
##
```

Crea un histograma de *petal.width*, teniendo en cuenta que el número de *bins* es variable fija éste a 9. Añádele color y nombres al eje x “*Petal Width*” y al gráfico dale el nombre “*Histogram of Petal Width*”. Crea un histograma para cada variable.

- Petal Width

Nota personal: Por alguna razón, poner *xlab* dentro de *labs* no me cambia el título del eje X, así que lo pongo fuera.

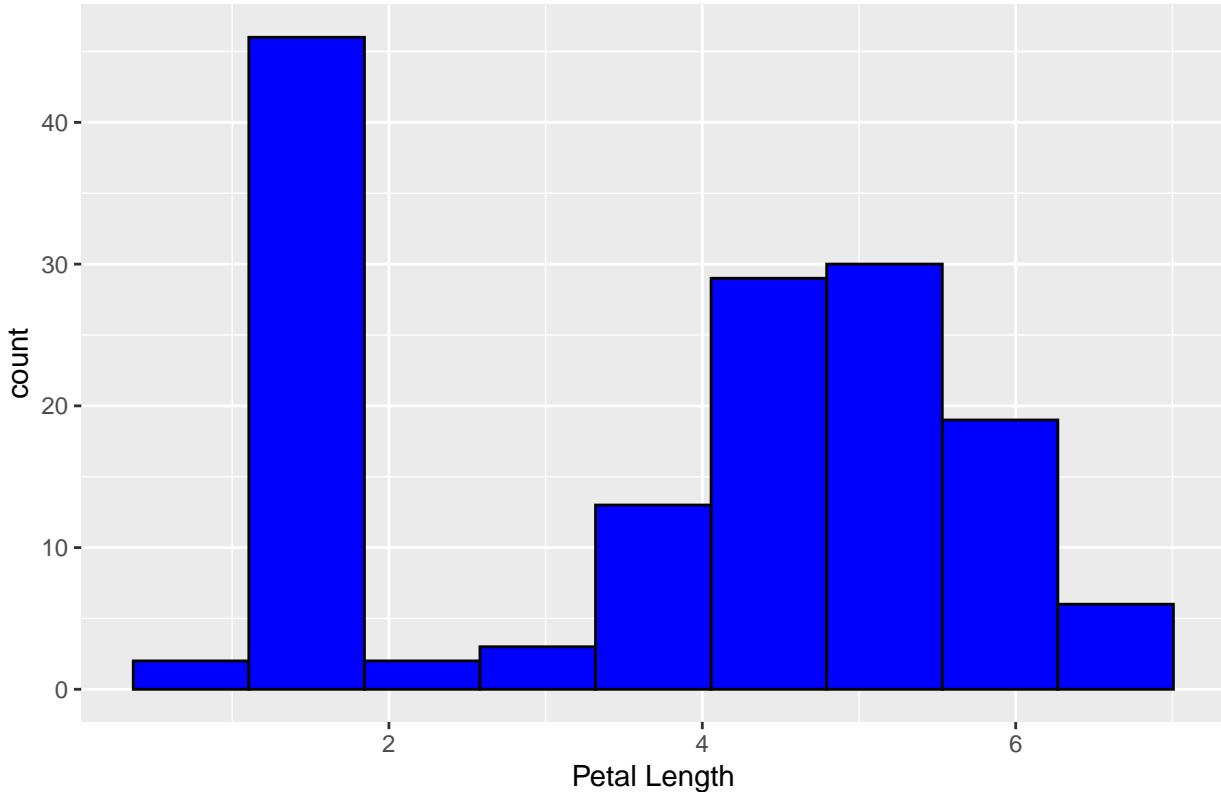
```
ggplot(iris, aes(x=Petal.Width)) +
  geom_histogram(bins = 9, fill="blue", col="black") +
  labs(title="Histogram of Petal Width") + xlab("Petal Width")
```



- Petal Length

```
ggplot(iris, aes(x=Petal.Length)) +
  geom_histogram(bins = 9, fill="blue", col="black") +
  labs(title="Histogram of Petal Length") + xlab("Petal Length")
```

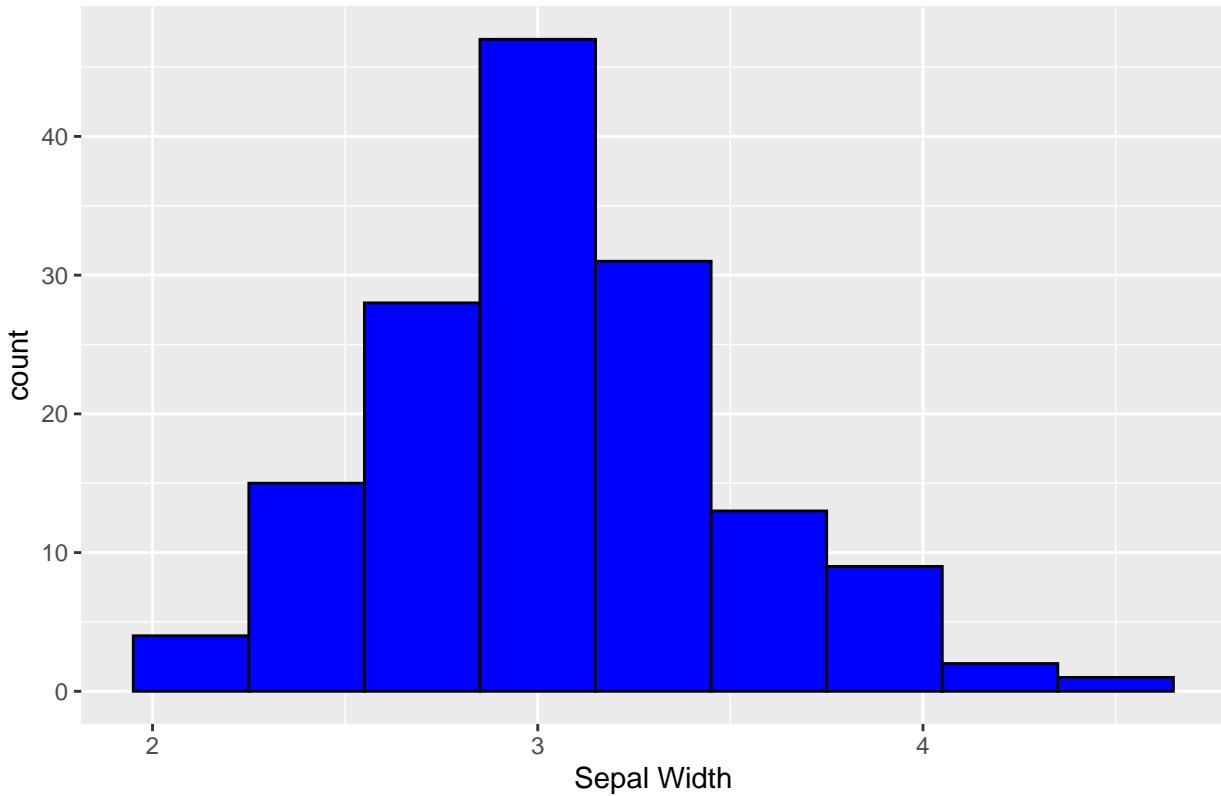
Histogram of Petal Length



- Sepal Width

```
ggplot(iris, aes(x=Sepal.Width)) +  
  geom_histogram(bins = 9, fill="blue", col="black") +  
  labs(title="Histogram of Sepal Width") + xlab("Sepal Width")
```

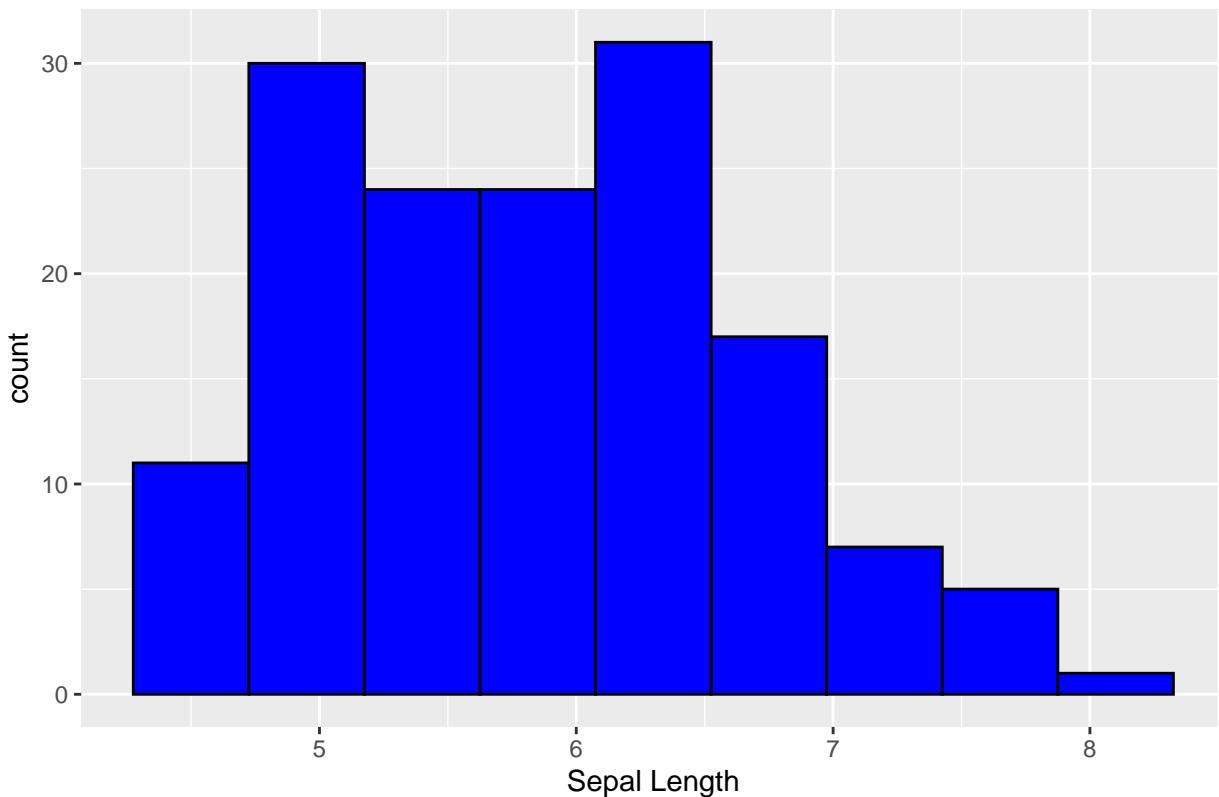
Histogram of Sepal Width



- Sepal Length

```
ggplot(iris, aes(x=Sepal.Length)) +  
  geom_histogram(bins = 9, fill="blue", col="black") +  
  labs(title="Histogram of Sepal Length") + xlab("Sepal Length")
```

Histogram of Sepal Length



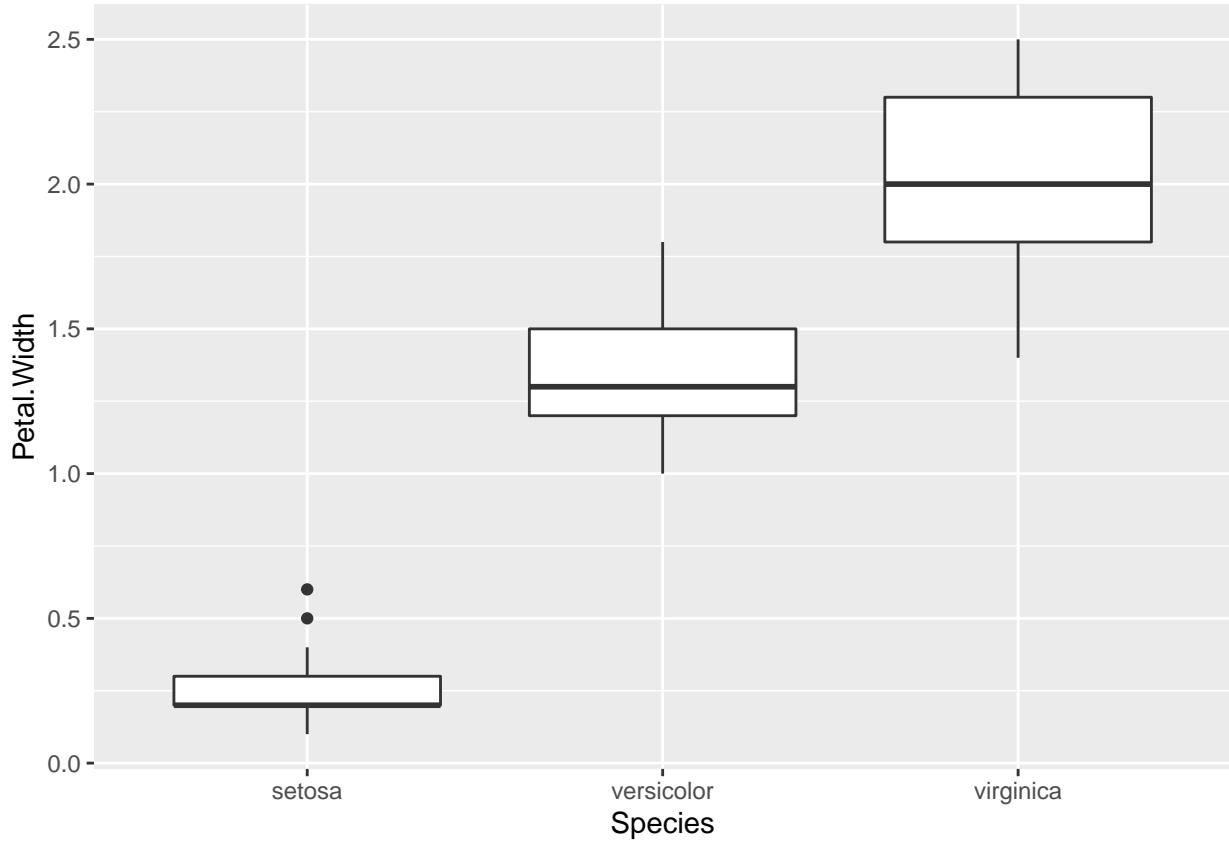
Crea los cuartiles del dataset.

```
# No tomamos la especie para los cuartiles.  
apply(iris[,1:4], 2, quantile)
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width  
## 0%          4.3        2.0       1.00      0.1  
## 25%         5.1        2.8       1.60      0.3  
## 50%         5.8        3.0       4.35      1.3  
## 75%         6.4        3.3       5.10      1.8  
## 100%        7.9        4.4       6.90      2.5
```

Representa en un boxplot la variable ancho de hoja dependiendo del tipo de hoja que tengan.

```
ggplot(iris, aes(x=Species, y=Petal.Width)) + geom_boxplot()
```



Crea los cuartiles para cada tipo de iris y represéntalos en un plot como líneas cada una de un color

```
df_setosa <- subset(iris[-5], iris$Species=="setosa")
df_versicolor <- subset(iris[-5], iris$Species=="versicolor")
df_virginica <- subset(iris[-5], iris$Species=="virginica")
q_setosa <- apply(df_setosa, 2, quantile)
q_versicolor <- apply(df_versicolor, 2, quantile)
q_virginica <- apply(df_virginica, 2, quantile)
colnames(q_setosa) <- paste0(colnames(q_setosa), ".Setosa")
colnames(q_versicolor) <- paste0(colnames(q_versicolor), ".Versicolor")
colnames(q_virginica) <- paste0(colnames(q_virginica), ".Virginica")
df_quantiles_iris <- as.data.frame(cbind(q_setosa, q_versicolor, q_virginica))
df_quantiles_iris$Percentil <- seq(0,100,25)
df_quantiles_iris

##      Sepal.Length.Setosa Sepal.Width.Setosa Petal.Length.Setosa
## 0%                  4.3          2.300        1.000
## 25%                 4.8          3.200        1.400
## 50%                 5.0          3.400        1.500
## 75%                 5.2          3.675        1.575
## 100%                5.8          4.400        1.900
##      Petal.Width.Setosa Sepal.Length.Versicolor Sepal.Width.Versicolor
## 0%                  0.1              4.9          2.000
## 25%                 0.2              5.6          2.525
## 50%                 0.2              5.9          2.800
```

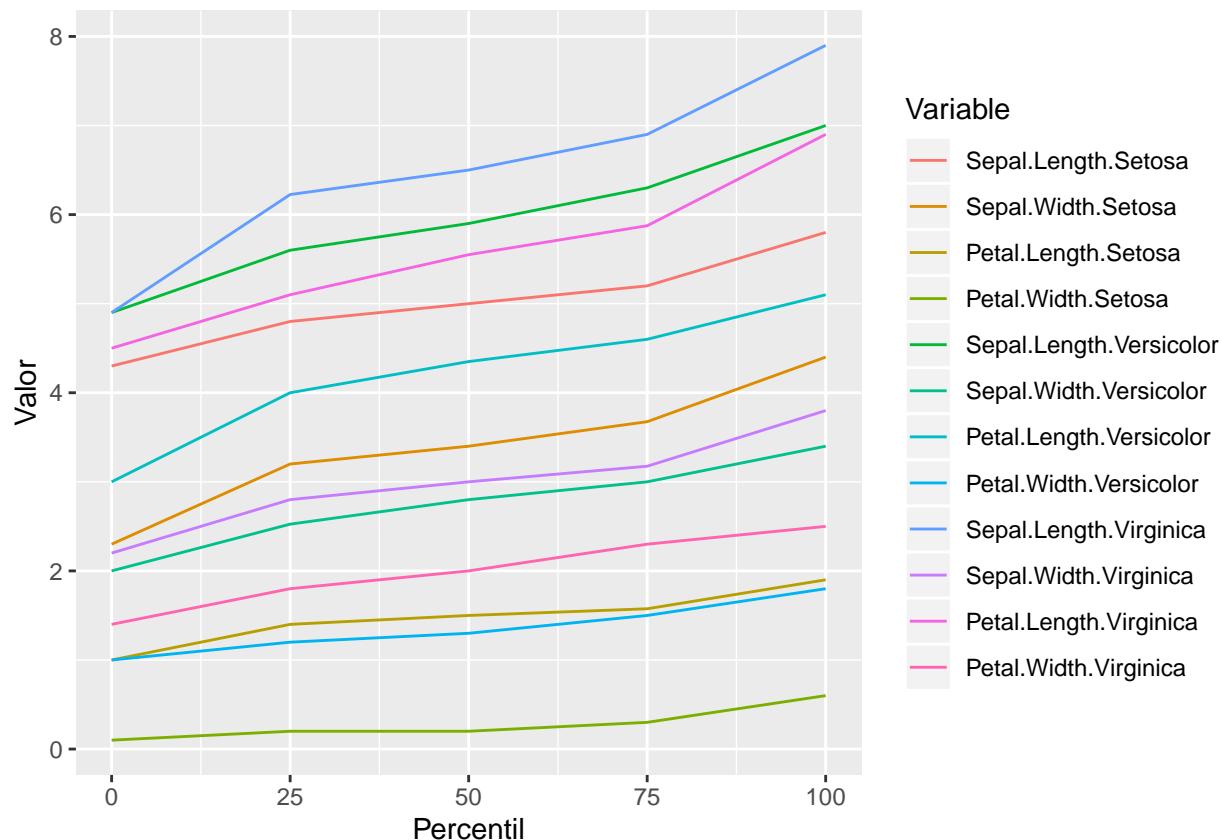
```

## 75%          0.3          6.3          3.000
## 100%         0.6          7.0          3.400
## Petal.Length.Versicolor Petal.Width.Versicolor Sepal.Length.Virginica
## 0%            3.00          1.0          4.900
## 25%           4.00          1.2          6.225
## 50%           4.35          1.3          6.500
## 75%           4.60          1.5          6.900
## 100%          5.10          1.8          7.900
## Sepal.Width.Virginica Petal.Length.Virginica Petal.Width.Virginica
## 0%            2.200         4.500         1.4
## 25%           2.800         5.100         1.8
## 50%           3.000         5.550         2.0
## 75%           3.175         5.875         2.3
## 100%          3.800         6.900         2.5
## Percentil
## 0%            0
## 25%           25
## 50%           50
## 75%           75
## 100%          100

library(reshape2)
# Utilizo melt de reshape2 para facilitar pintar muchos gráficos de líneas.
df_quantiles_iris2 <- melt(df_quantiles_iris, id.var="Percentil")
names(df_quantiles_iris2) <-c("Percentil", "Variable", "Valor")

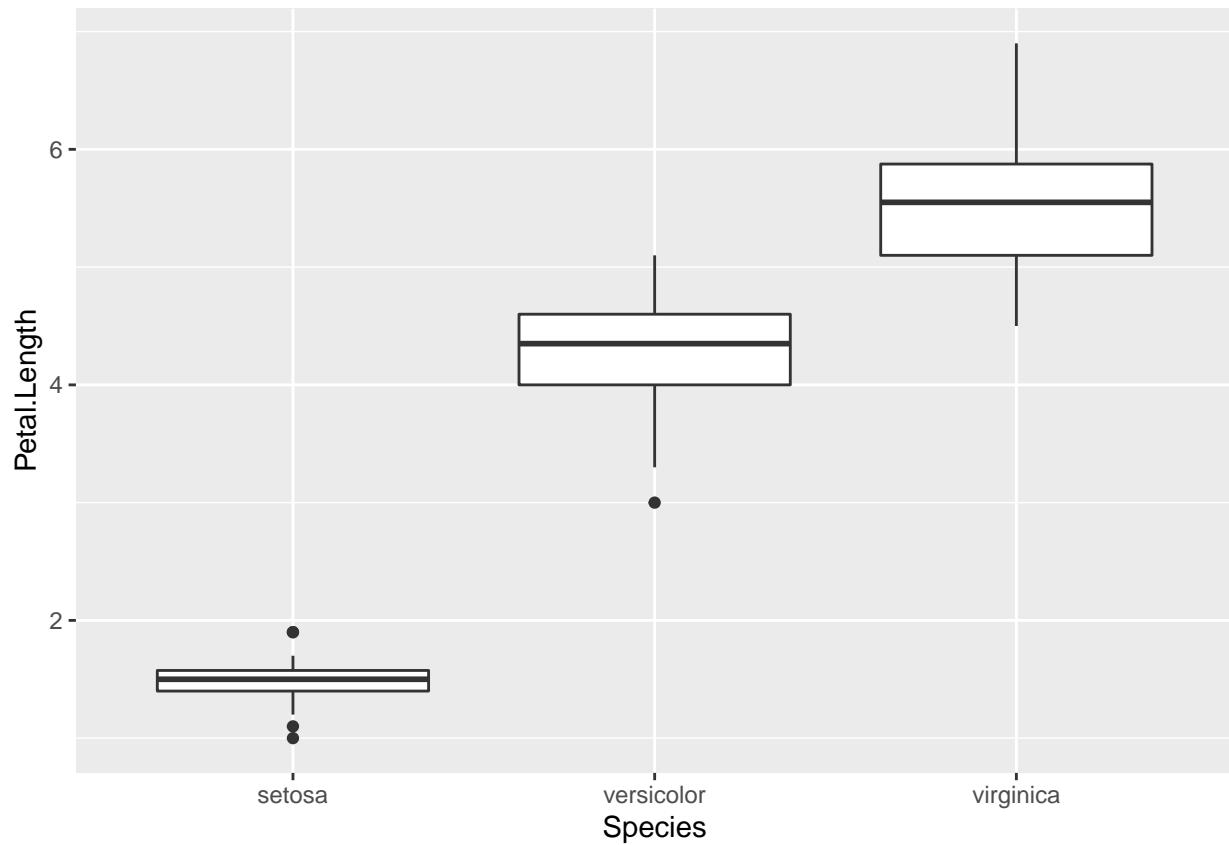
ggplot(df_quantiles_iris2, aes(x=Percentil, y=Valor, colour=Variable)) +
  geom_line()

```



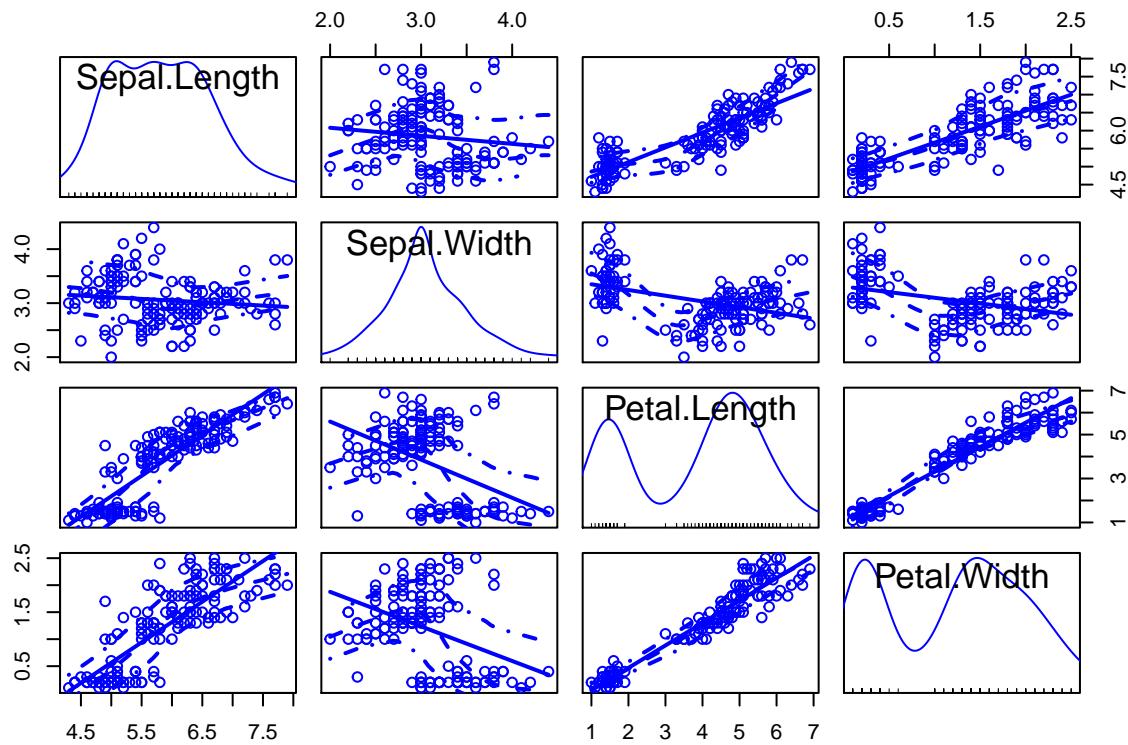
Crea los boxplot de la longitud del pétalo en función de la especie de Iris.

```
ggplot(iris, aes(x=Species, y=Petal.Length)) + geom_boxplot()
```



Compara con scatter plots las variables entre sí.

```
scatterplotMatrix(iris[-5])
```



Crea una nueva columna llamada proporción que es el ratio entre Sepal.Length y Sepal.Width. Podeis hacerlo en R base o usando el paquete dplyr. Los que no lo conozcan tienen informacion al final del documento.

```
iris %>% mutate(proporcion=Sepal.Length/Sepal.Width)
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 1	5.1	3.5	1.4	0.2	setosa
## 2	4.9	3.0	1.4	0.2	setosa
## 3	4.7	3.2	1.3	0.2	setosa
## 4	4.6	3.1	1.5	0.2	setosa
## 5	5.0	3.6	1.4	0.2	setosa
## 6	5.4	3.9	1.7	0.4	setosa
## 7	4.6	3.4	1.4	0.3	setosa
## 8	5.0	3.4	1.5	0.2	setosa
## 9	4.4	2.9	1.4	0.2	setosa
## 10	4.9	3.1	1.5	0.1	setosa
## 11	5.4	3.7	1.5	0.2	setosa
## 12	4.8	3.4	1.6	0.2	setosa
## 13	4.8	3.0	1.4	0.1	setosa
## 14	4.3	3.0	1.1	0.1	setosa
## 15	5.8	4.0	1.2	0.2	setosa
## 16	5.7	4.4	1.5	0.4	setosa
## 17	5.4	3.9	1.3	0.4	setosa
## 18	5.1	3.5	1.4	0.3	setosa
## 19	5.7	3.8	1.7	0.3	setosa

## 20	5.1	3.8	1.5	0.3	setosa
## 21	5.4	3.4	1.7	0.2	setosa
## 22	5.1	3.7	1.5	0.4	setosa
## 23	4.6	3.6	1.0	0.2	setosa
## 24	5.1	3.3	1.7	0.5	setosa
## 25	4.8	3.4	1.9	0.2	setosa
## 26	5.0	3.0	1.6	0.2	setosa
## 27	5.0	3.4	1.6	0.4	setosa
## 28	5.2	3.5	1.5	0.2	setosa
## 29	5.2	3.4	1.4	0.2	setosa
## 30	4.7	3.2	1.6	0.2	setosa
## 31	4.8	3.1	1.6	0.2	setosa
## 32	5.4	3.4	1.5	0.4	setosa
## 33	5.2	4.1	1.5	0.1	setosa
## 34	5.5	4.2	1.4	0.2	setosa
## 35	4.9	3.1	1.5	0.2	setosa
## 36	5.0	3.2	1.2	0.2	setosa
## 37	5.5	3.5	1.3	0.2	setosa
## 38	4.9	3.6	1.4	0.1	setosa
## 39	4.4	3.0	1.3	0.2	setosa
## 40	5.1	3.4	1.5	0.2	setosa
## 41	5.0	3.5	1.3	0.3	setosa
## 42	4.5	2.3	1.3	0.3	setosa
## 43	4.4	3.2	1.3	0.2	setosa
## 44	5.0	3.5	1.6	0.6	setosa
## 45	5.1	3.8	1.9	0.4	setosa
## 46	4.8	3.0	1.4	0.3	setosa
## 47	5.1	3.8	1.6	0.2	setosa
## 48	4.6	3.2	1.4	0.2	setosa
## 49	5.3	3.7	1.5	0.2	setosa
## 50	5.0	3.3	1.4	0.2	setosa
## 51	7.0	3.2	4.7	1.4	versicolor
## 52	6.4	3.2	4.5	1.5	versicolor
## 53	6.9	3.1	4.9	1.5	versicolor
## 54	5.5	2.3	4.0	1.3	versicolor
## 55	6.5	2.8	4.6	1.5	versicolor
## 56	5.7	2.8	4.5	1.3	versicolor
## 57	6.3	3.3	4.7	1.6	versicolor
## 58	4.9	2.4	3.3	1.0	versicolor
## 59	6.6	2.9	4.6	1.3	versicolor
## 60	5.2	2.7	3.9	1.4	versicolor
## 61	5.0	2.0	3.5	1.0	versicolor
## 62	5.9	3.0	4.2	1.5	versicolor
## 63	6.0	2.2	4.0	1.0	versicolor
## 64	6.1	2.9	4.7	1.4	versicolor
## 65	5.6	2.9	3.6	1.3	versicolor
## 66	6.7	3.1	4.4	1.4	versicolor
## 67	5.6	3.0	4.5	1.5	versicolor
## 68	5.8	2.7	4.1	1.0	versicolor
## 69	6.2	2.2	4.5	1.5	versicolor
## 70	5.6	2.5	3.9	1.1	versicolor
## 71	5.9	3.2	4.8	1.8	versicolor
## 72	6.1	2.8	4.0	1.3	versicolor
## 73	6.3	2.5	4.9	1.5	versicolor

## 74	6.1	2.8	4.7	1.2 versicolor
## 75	6.4	2.9	4.3	1.3 versicolor
## 76	6.6	3.0	4.4	1.4 versicolor
## 77	6.8	2.8	4.8	1.4 versicolor
## 78	6.7	3.0	5.0	1.7 versicolor
## 79	6.0	2.9	4.5	1.5 versicolor
## 80	5.7	2.6	3.5	1.0 versicolor
## 81	5.5	2.4	3.8	1.1 versicolor
## 82	5.5	2.4	3.7	1.0 versicolor
## 83	5.8	2.7	3.9	1.2 versicolor
## 84	6.0	2.7	5.1	1.6 versicolor
## 85	5.4	3.0	4.5	1.5 versicolor
## 86	6.0	3.4	4.5	1.6 versicolor
## 87	6.7	3.1	4.7	1.5 versicolor
## 88	6.3	2.3	4.4	1.3 versicolor
## 89	5.6	3.0	4.1	1.3 versicolor
## 90	5.5	2.5	4.0	1.3 versicolor
## 91	5.5	2.6	4.4	1.2 versicolor
## 92	6.1	3.0	4.6	1.4 versicolor
## 93	5.8	2.6	4.0	1.2 versicolor
## 94	5.0	2.3	3.3	1.0 versicolor
## 95	5.6	2.7	4.2	1.3 versicolor
## 96	5.7	3.0	4.2	1.2 versicolor
## 97	5.7	2.9	4.2	1.3 versicolor
## 98	6.2	2.9	4.3	1.3 versicolor
## 99	5.1	2.5	3.0	1.1 versicolor
## 100	5.7	2.8	4.1	1.3 versicolor
## 101	6.3	3.3	6.0	2.5 virginica
## 102	5.8	2.7	5.1	1.9 virginica
## 103	7.1	3.0	5.9	2.1 virginica
## 104	6.3	2.9	5.6	1.8 virginica
## 105	6.5	3.0	5.8	2.2 virginica
## 106	7.6	3.0	6.6	2.1 virginica
## 107	4.9	2.5	4.5	1.7 virginica
## 108	7.3	2.9	6.3	1.8 virginica
## 109	6.7	2.5	5.8	1.8 virginica
## 110	7.2	3.6	6.1	2.5 virginica
## 111	6.5	3.2	5.1	2.0 virginica
## 112	6.4	2.7	5.3	1.9 virginica
## 113	6.8	3.0	5.5	2.1 virginica
## 114	5.7	2.5	5.0	2.0 virginica
## 115	5.8	2.8	5.1	2.4 virginica
## 116	6.4	3.2	5.3	2.3 virginica
## 117	6.5	3.0	5.5	1.8 virginica
## 118	7.7	3.8	6.7	2.2 virginica
## 119	7.7	2.6	6.9	2.3 virginica
## 120	6.0	2.2	5.0	1.5 virginica
## 121	6.9	3.2	5.7	2.3 virginica
## 122	5.6	2.8	4.9	2.0 virginica
## 123	7.7	2.8	6.7	2.0 virginica
## 124	6.3	2.7	4.9	1.8 virginica
## 125	6.7	3.3	5.7	2.1 virginica
## 126	7.2	3.2	6.0	1.8 virginica
## 127	6.2	2.8	4.8	1.8 virginica

```

## 128      6.1      3.0      4.9      1.8  virginica
## 129      6.4      2.8      5.6      2.1  virginica
## 130      7.2      3.0      5.8      1.6  virginica
## 131      7.4      2.8      6.1      1.9  virginica
## 132      7.9      3.8      6.4      2.0  virginica
## 133      6.4      2.8      5.6      2.2  virginica
## 134      6.3      2.8      5.1      1.5  virginica
## 135      6.1      2.6      5.6      1.4  virginica
## 136      7.7      3.0      6.1      2.3  virginica
## 137      6.3      3.4      5.6      2.4  virginica
## 138      6.4      3.1      5.5      1.8  virginica
## 139      6.0      3.0      4.8      1.8  virginica
## 140      6.9      3.1      5.4      2.1  virginica
## 141      6.7      3.1      5.6      2.4  virginica
## 142      6.9      3.1      5.1      2.3  virginica
## 143      5.8      2.7      5.1      1.9  virginica
## 144      6.8      3.2      5.9      2.3  virginica
## 145      6.7      3.3      5.7      2.5  virginica
## 146      6.7      3.0      5.2      2.3  virginica
## 147      6.3      2.5      5.0      1.9  virginica
## 148      6.5      3.0      5.2      2.0  virginica
## 149      6.2      3.4      5.4      2.3  virginica
## 150      5.9      3.0      5.1      1.8  virginica
##     proporcion
## 1      1.457143
## 2      1.633333
## 3      1.468750
## 4      1.483871
## 5      1.388889
## 6      1.384615
## 7      1.352941
## 8      1.470588
## 9      1.517241
## 10     1.580645
## 11     1.459459
## 12     1.411765
## 13     1.600000
## 14     1.433333
## 15     1.450000
## 16     1.295455
## 17     1.384615
## 18     1.457143
## 19     1.500000
## 20     1.342105
## 21     1.588235
## 22     1.378378
## 23     1.277778
## 24     1.545455
## 25     1.411765
## 26     1.666667
## 27     1.470588
## 28     1.485714
## 29     1.529412
## 30     1.468750

```

```
## 31      1.548387
## 32      1.588235
## 33      1.268293
## 34      1.309524
## 35      1.580645
## 36      1.562500
## 37      1.571429
## 38      1.361111
## 39      1.466667
## 40      1.500000
## 41      1.428571
## 42      1.956522
## 43      1.375000
## 44      1.428571
## 45      1.342105
## 46      1.600000
## 47      1.342105
## 48      1.437500
## 49      1.432432
## 50      1.515152
## 51      2.187500
## 52      2.000000
## 53      2.225806
## 54      2.391304
## 55      2.321429
## 56      2.035714
## 57      1.909091
## 58      2.041667
## 59      2.275862
## 60      1.925926
## 61      2.500000
## 62      1.966667
## 63      2.727273
## 64      2.103448
## 65      1.931034
## 66      2.161290
## 67      1.866667
## 68      2.148148
## 69      2.818182
## 70      2.240000
## 71      1.843750
## 72      2.178571
## 73      2.520000
## 74      2.178571
## 75      2.206897
## 76      2.200000
## 77      2.428571
## 78      2.233333
## 79      2.068966
## 80      2.192308
## 81      2.291667
## 82      2.291667
## 83      2.148148
## 84      2.222222
```

```
## 85      1.800000
## 86      1.764706
## 87      2.161290
## 88      2.739130
## 89      1.866667
## 90      2.200000
## 91      2.115385
## 92      2.033333
## 93      2.230769
## 94      2.173913
## 95      2.074074
## 96      1.900000
## 97      1.965517
## 98      2.137931
## 99      2.040000
## 100     2.035714
## 101     1.909091
## 102     2.148148
## 103     2.366667
## 104     2.172414
## 105     2.166667
## 106     2.533333
## 107     1.960000
## 108     2.517241
## 109     2.680000
## 110     2.000000
## 111     2.031250
## 112     2.370370
## 113     2.266667
## 114     2.280000
## 115     2.071429
## 116     2.000000
## 117     2.166667
## 118     2.026316
## 119     2.961538
## 120     2.727273
## 121     2.156250
## 122     2.000000
## 123     2.750000
## 124     2.333333
## 125     2.030303
## 126     2.250000
## 127     2.214286
## 128     2.033333
## 129     2.285714
## 130     2.400000
## 131     2.642857
## 132     2.078947
## 133     2.285714
## 134     2.250000
## 135     2.346154
## 136     2.566667
## 137     1.852941
## 138     2.064516
```

```

## 139 2.000000
## 140 2.225806
## 141 2.161290
## 142 2.225806
## 143 2.148148
## 144 2.125000
## 145 2.030303
## 146 2.233333
## 147 2.520000
## 148 2.166667
## 149 1.823529
## 150 1.966667

```

El conjunto de datos “swiss” contiene una medida estandarizada de fecundidad y varios indicadores socioeconómicos para cada una de las 47 provincias francófonas de Suiza.

```
swiss
```

	Fertility	Agriculture	Examination	Education	Catholic
## Courteley	80.2	17.0	15	12	9.96
## Delemont	83.1	45.1	6	9	84.84
## Franches-Mnt	92.5	39.7	5	5	93.40
## Moutier	85.8	36.5	12	7	33.77
## Neuveville	76.9	43.5	17	15	5.16
## Porrentruy	76.1	35.3	9	7	90.57
## Broye	83.8	70.2	16	7	92.85
## Glane	92.4	67.8	14	8	97.16
## Gruyere	82.4	53.3	12	7	97.67
## Sarine	82.9	45.2	16	13	91.38
## Veveyse	87.1	64.5	14	6	98.61
## Aigle	64.1	62.0	21	12	8.52
## Aubonne	66.9	67.5	14	7	2.27
## Avenches	68.9	60.7	19	12	4.43
## Cossonay	61.7	69.3	22	5	2.82
## Echallens	68.3	72.6	18	2	24.20
## Grandson	71.7	34.0	17	8	3.30
## Lausanne	55.7	19.4	26	28	12.11
## La Vallee	54.3	15.2	31	20	2.15
## Lavaux	65.1	73.0	19	9	2.84
## Morges	65.5	59.8	22	10	5.23
## Moudon	65.0	55.1	14	3	4.52
## Nyone	56.6	50.9	22	12	15.14
## Orbe	57.4	54.1	20	6	4.20
## Oron	72.5	71.2	12	1	2.40
## Payerne	74.2	58.1	14	8	5.23
## Paysd'enhaut	72.0	63.5	6	3	2.56
## Rolle	60.5	60.8	16	10	7.72
## Vevey	58.3	26.8	25	19	18.46
## Yverdon	65.4	49.5	15	8	6.10
## Conthey	75.5	85.9	3	2	99.71
## Entremont	69.3	84.9	7	6	99.68
## Herens	77.3	89.7	5	2	100.00
## Martigny	70.5	78.2	12	6	98.96

## MontHEY	79.4	64.9	7	3	98.22
## St Maurice	65.0	75.9	9	9	99.06
## Sierre	92.2	84.6	3	3	99.46
## Sion	79.3	63.1	13	13	96.83
## Boudry	70.4	38.4	26	12	5.62
## La Chauxdfnd	65.7	7.7	29	11	13.79
## Le Locle	72.7	16.7	22	13	11.22
## Neuchatel	64.4	17.6	35	32	16.92
## Val de Ruz	77.6	37.6	15	7	4.97
## ValdeTravers	67.6	18.7	25	7	8.65
## V. De Geneve	35.0	1.2	37	53	42.34
## Rive Droite	44.7	46.6	16	29	50.43
## Rive Gauche	42.8	27.7	22	29	58.33
## Infant.Mortality					
## Courtelary		22.2			
## Delemont		22.2			
## Franches-Mnt		20.2			
## Moutier		20.3			
## Neuveville		20.6			
## Porrentruy		26.6			
## Broye		23.6			
## Glane		24.9			
## Gruyere		21.0			
## Sarine		24.4			
## Veveyse		24.5			
## Aigle		16.5			
## Aubonne		19.1			
## Avenches		22.7			
## Cossonay		18.7			
## Echallens		21.2			
## Grandson		20.0			
## Lausanne		20.2			
## La Vallee		10.8			
## Lavaux		20.0			
## Morges		18.0			
## Moudon		22.4			
## Nyone		16.7			
## Orbe		15.3			
## Oron		21.0			
## Payerne		23.8			
## Paysd'enhaut		18.0			
## Rolle		16.3			
## Vevey		20.9			
## Yverdon		22.5			
## Conthey		15.1			
## Entremont		19.8			
## Herens		18.3			
## Martigwy		19.4			
## MontHEY		20.2			
## St Maurice		17.8			
## Sierre		16.3			
## Sion		18.1			
## Boudry		20.3			
## La Chauxdfnd		20.5			

```

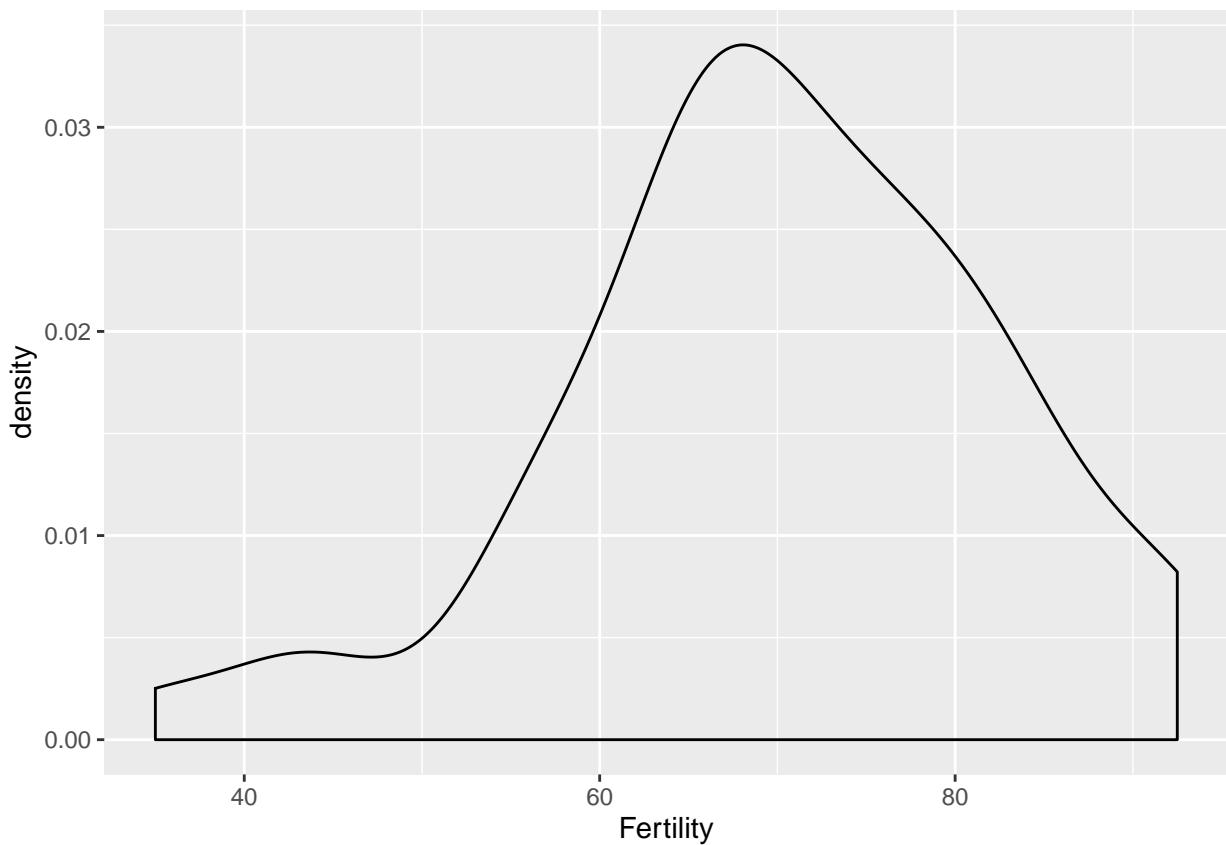
## Le Locle           18.9
## Neuchatel        23.0
## Val de Ruz       20.0
## ValdeTravers     19.5
## V. De Geneve     18.0
## Rive Droite      18.2
## Rive Gauche      19.3

```

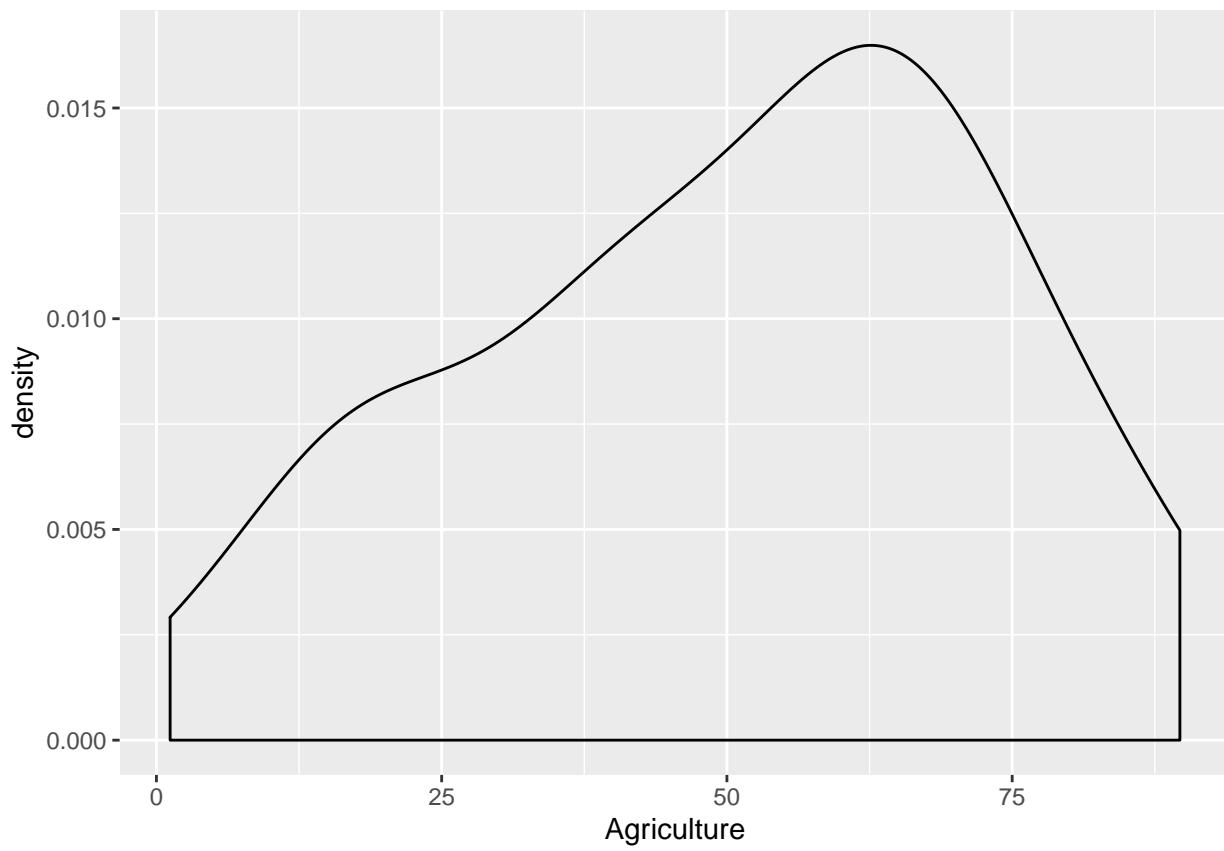
¿Qué diagrama dibujaría para mostrar la distribución de todos los valores? ¿Qué conclusiones sacarías?

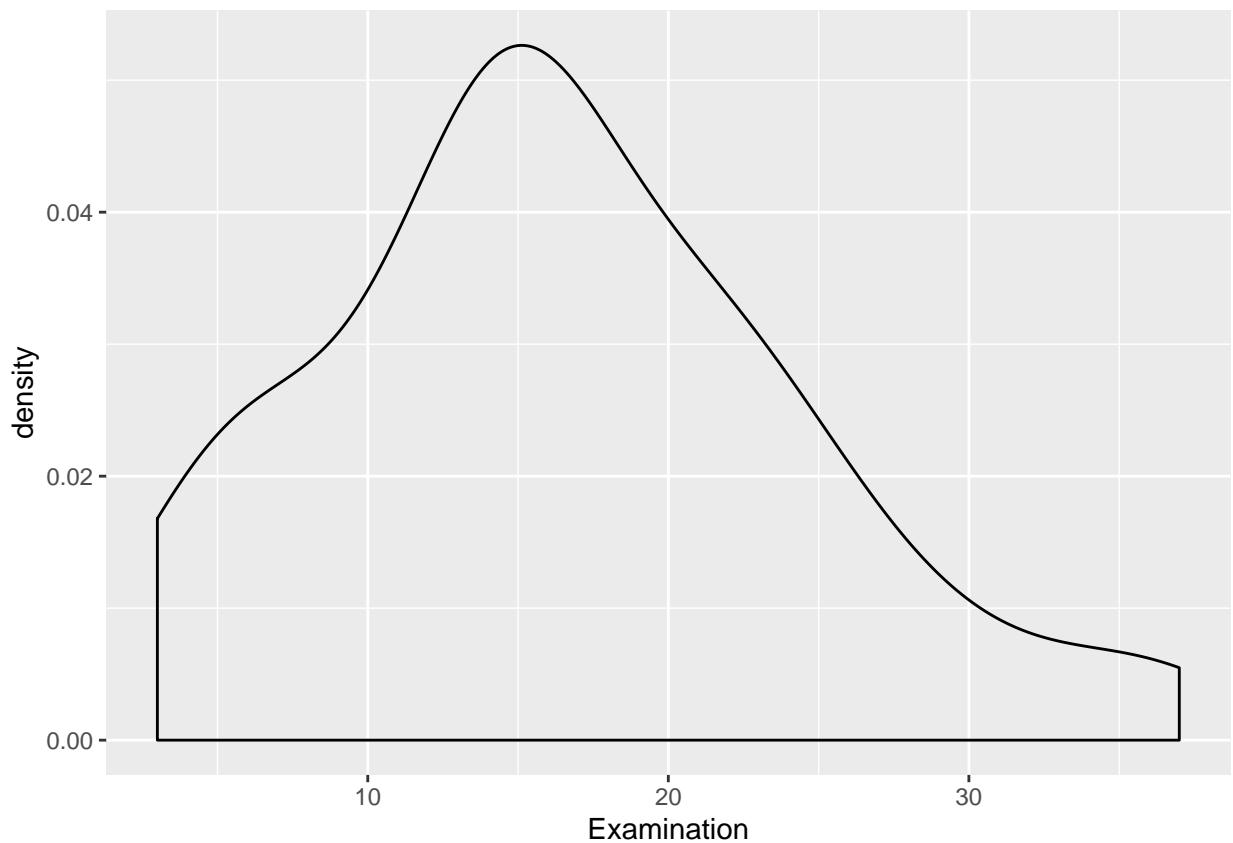
Podríamos utilizar un histograma, pero puesto que el número de muestras es relativamente bajo (47 muestras), vamos a optar por pintar una curva de densidad de distribución.

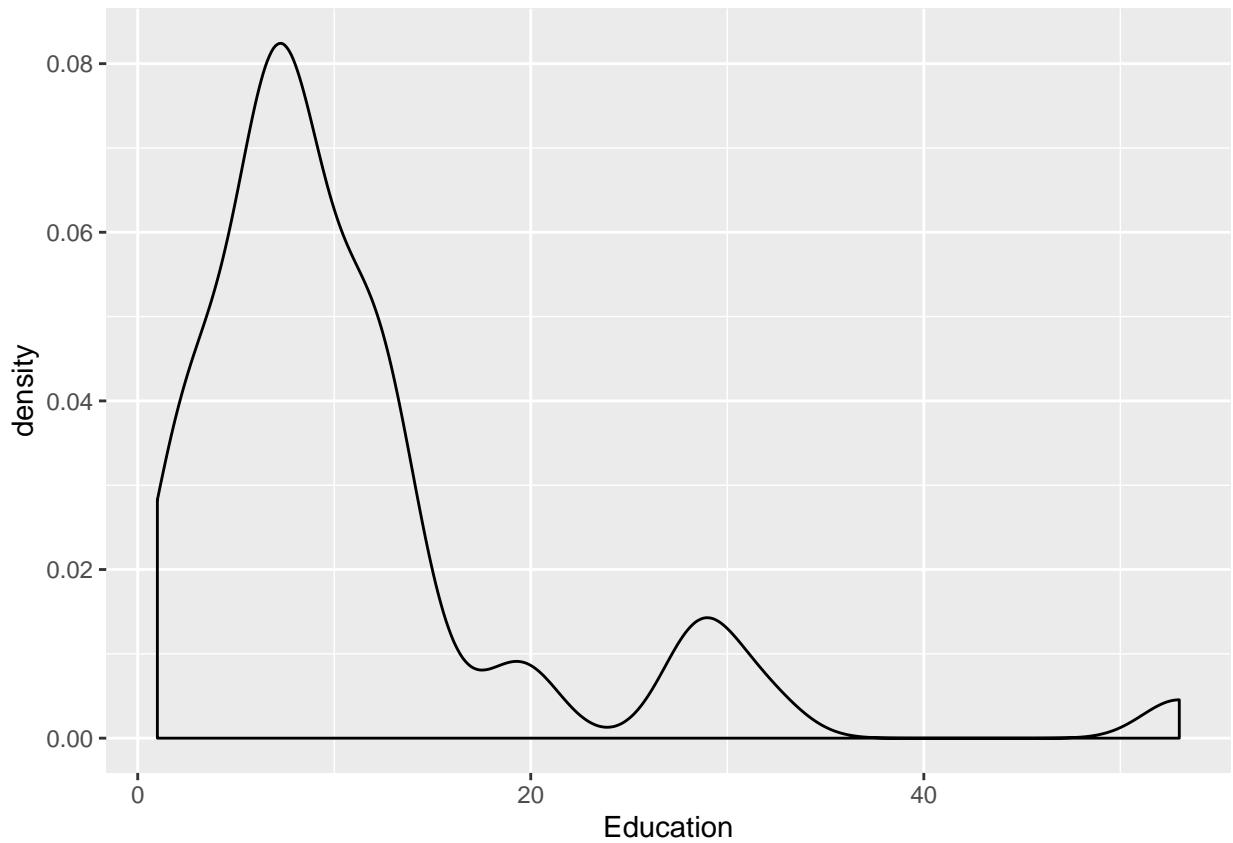
```
ggplot(swiss, aes(x=Fertility)) + geom_density()
```



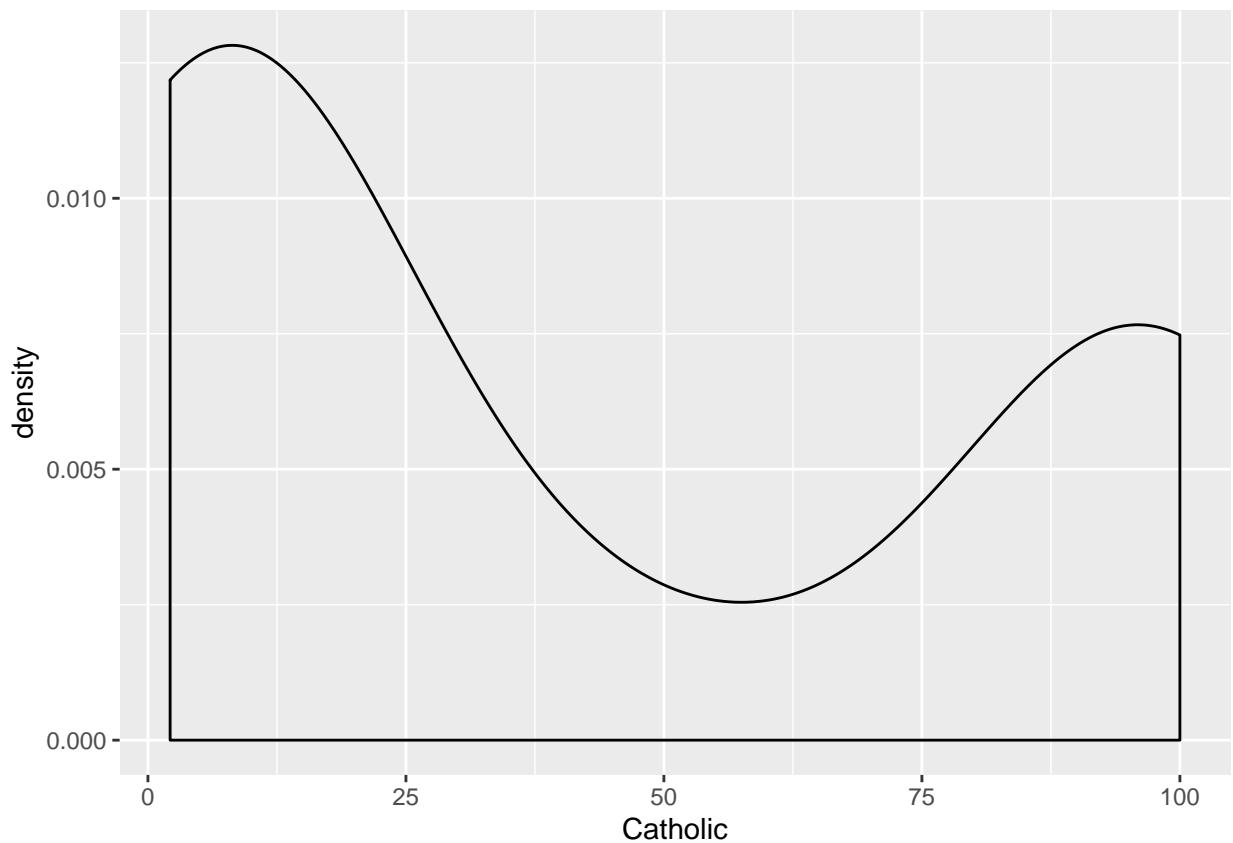
```
ggplot(swiss, aes(x=Agriculture)) + geom_density()
```



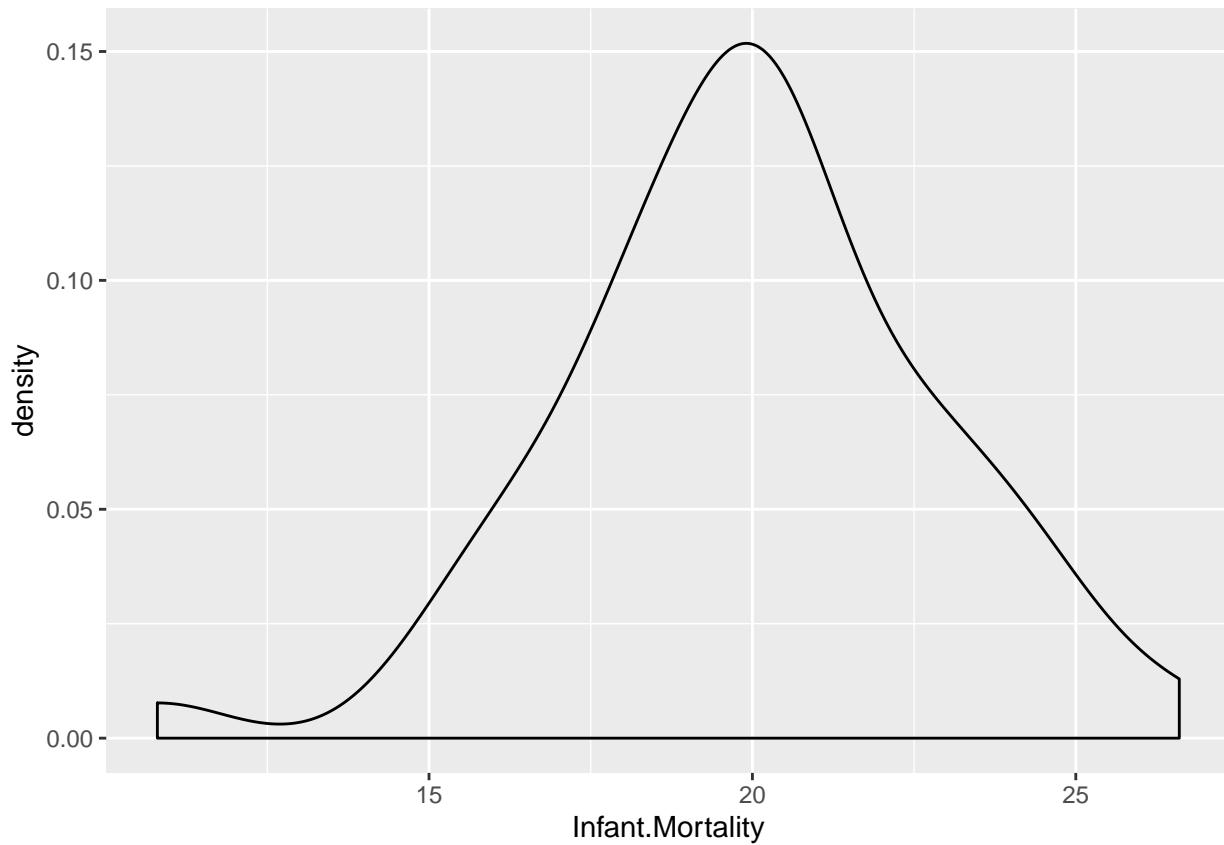




```
ggplot(swiss, aes(x=Catholic)) + geom_density()
```



```
ggplot(swiss, aes(x=Infant.Mortality)) + geom_density()
```

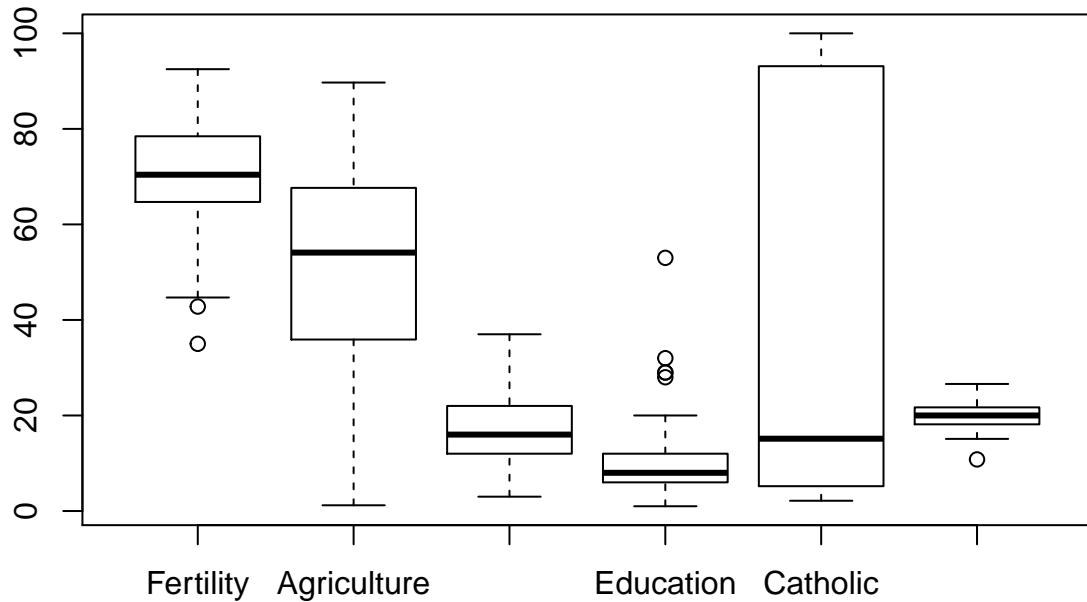


Las variables Fertility, Agriculture e Infant Mortality son “negative skewed”. Las distribuciones de Fertility e Infant Mortality se asemejan bastante. Examination es “positive skewed”. Education y Catholic son multimodales, lo que nos podría indicar la presencia de varias clases. En concreto, Catholic tiene 3 puntos extremos bastante definidos, mientras que en Education no queda tan claro.

Dibuje gráficos para cada variable. ¿Qué puede concluir de las distribuciones con respecto a su forma y posibles valores atípicos?

Hemos hablado brevemente de la forma en el apartado anterior, para ver la escala y los posibles valores atípicos vamos a usar un boxplot.

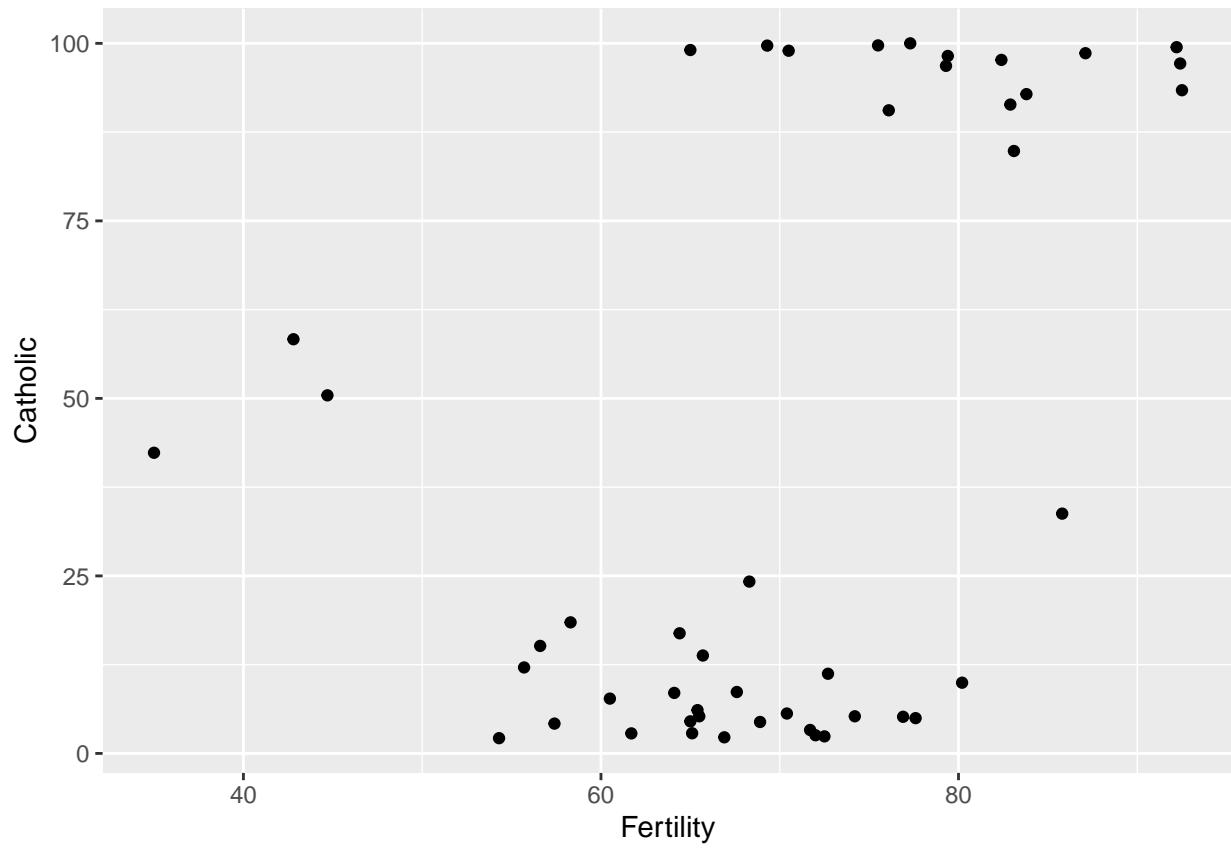
```
boxplot(swiss)
```



Podemos ver que todas las variables se encuentran dentro del rango [0, 100], presentan diferentes escalas pero son comparables en el boxplot. La variable con más variabilidad en los datos es educación donde la mayoría de los datos se encuentran alrededor del 10 pero encontramos posibles “outliers” en “Education”. Por otro lado, sólo vemos dos “outliers” en Fertility y uno en “Infant.Mortality”.

Dibuja un diagrama de dispersión de Fertilidad frente a % Catholic. ¿Qué tipo de áreas tienen las tasas de fertilidad más bajas?

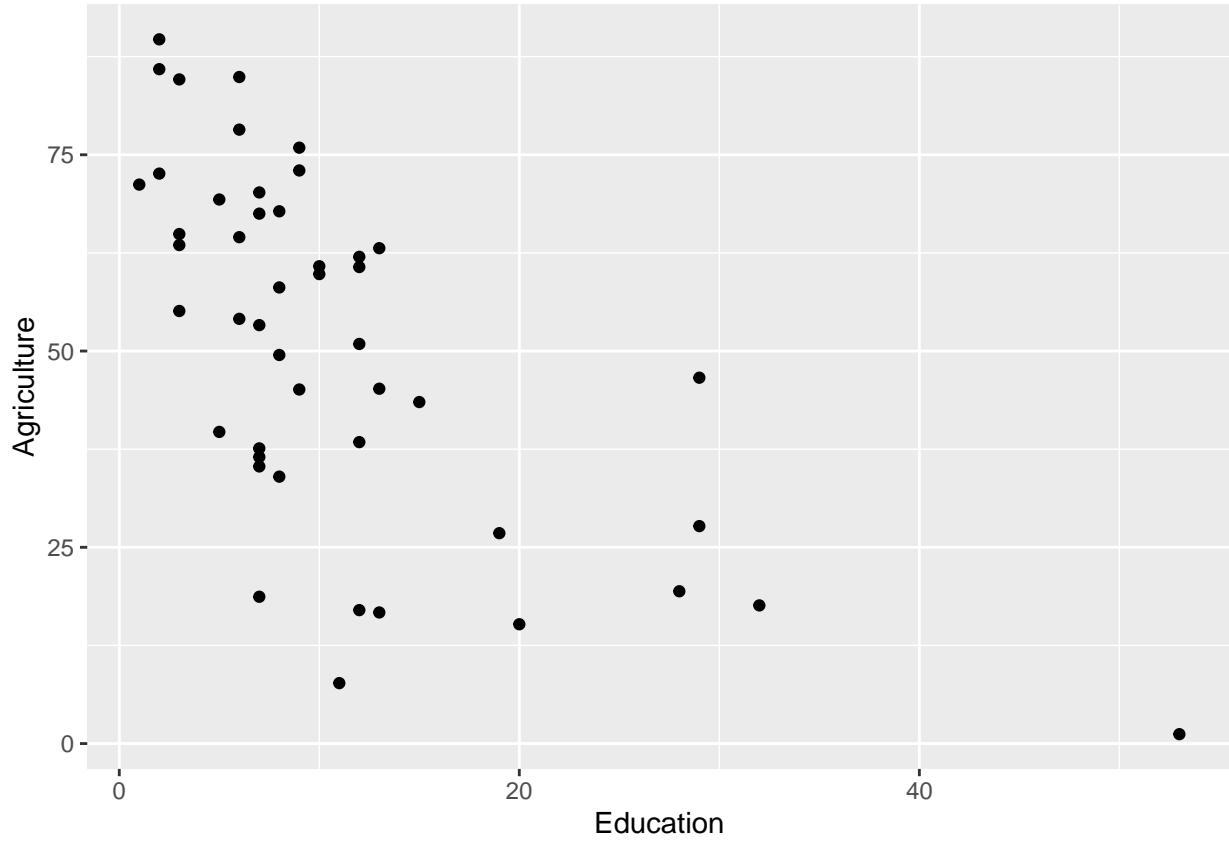
```
ggplot(swiss, aes(x=Fertility, y=Catholic)) + geom_point()
```



Las zonas menos fértiles se corresponden a aquellas donde alrededor la mitad de la población es católica, aunque sólo tenemos 3 muestras. Estas van seguidas de las zonas donde la población católica es inferior al 25 %.

¿Qué tipo de relación existe entre las variables Educación y Agricultura?

```
ggplot(swiss, aes(x=Education, y=Agriculture)) + geom_point()
```



La tendencia es que a menor nivel de educación, más porcentaje de hombres trabajando en la agricultura.

El conjunto de datos de aceites de oliva es bien conocido y se puede encontrar en varios paquetes, por ejemplo, como aceitunas en extracat.. La fuente original de los datos es el artículo [Forina et al., 1983].

El paquete extracat ha sido eliminado de CRAN así que voy a usar otro paquete que contiene el dataset

```
library("classifly")
olives
```

##	Region	Area	palmitic	palmitoleic	stearic	oleic	linoleic
## 1	1	North-Apulia	1075	75	226	7823	672
## 2	1	North-Apulia	1088	73	224	7709	781
## 3	1	North-Apulia	911	54	246	8113	549
## 4	1	North-Apulia	966	57	240	7952	619
## 5	1	North-Apulia	1051	67	259	7771	672
## 6	1	North-Apulia	911	49	268	7924	678
## 7	1	North-Apulia	922	66	264	7990	618
## 8	1	North-Apulia	1100	61	235	7728	734
## 9	1	North-Apulia	1082	60	239	7745	709
## 10	1	North-Apulia	1037	55	213	7944	633
## 11	1	North-Apulia	1051	35	219	7978	605
## 12	1	North-Apulia	1036	59	235	7868	661
## 13	1	North-Apulia	1074	70	214	7728	747
## 14	1	North-Apulia	875	52	243	8018	655
## 15	1	North-Apulia	952	49	254	7795	780

## 16	1	North-Apulia	1155	98	201	7606	816
## 17	1	North-Apulia	943	94	183	7840	788
## 18	1	North-Apulia	1278	69	205	7344	957
## 19	1	North-Apulia	961	70	195	7958	742
## 20	1	North-Apulia	952	77	258	7820	736
## 21	1	North-Apulia	1074	67	236	7692	716
## 22	1	North-Apulia	995	46	288	7806	679
## 23	1	North-Apulia	1056	53	247	7703	700
## 24	1	North-Apulia	1065	39	234	7876	703
## 25	1	North-Apulia	1065	45	245	7779	696
## 26	1	Calabria	1315	139	230	7299	832
## 27	1	Calabria	1321	136	217	7174	950
## 28	1	Calabria	1359	115	246	7234	874
## 29	1	Calabria	1378	111	272	7127	940
## 30	1	Calabria	1295	109	245	7253	903
## 31	1	Calabria	1275	121	215	7285	892
## 32	1	Calabria	1336	120	318	7083	915
## 33	1	Calabria	1309	122	241	7257	870
## 34	1	Calabria	1340	114	189	7337	820
## 35	1	Calabria	1299	116	253	7309	823
## 36	1	Calabria	1221	107	221	7441	798
## 37	1	Calabria	1245	72	283	7395	829
## 38	1	Calabria	1285	129	244	7323	819
## 39	1	Calabria	1248	107	313	7299	840
## 40	1	Calabria	1356	106	236	7209	866
## 41	1	Calabria	1260	102	228	7354	870
## 42	1	Calabria	1261	121	312	7238	877
## 43	1	Calabria	1304	124	279	7160	928
## 44	1	Calabria	1344	117	287	7129	897
## 45	1	Calabria	1323	96	300	7351	757
## 46	1	Calabria	1292	117	215	7351	839
## 47	1	Calabria	1254	118	244	7394	786
## 48	1	Calabria	1312	131	259	7167	939
## 49	1	Calabria	1213	109	301	7261	925
## 50	1	Calabria	1359	98	351	7262	780
## 51	1	Calabria	1266	97	263	7435	743
## 52	1	Calabria	1298	99	311	7311	787
## 53	1	Calabria	1272	116	279	7258	872
## 54	1	Calabria	1278	87	332	7379	771
## 55	1	Calabria	1184	112	311	7391	819
## 56	1	Calabria	1382	110	268	7241	828
## 57	1	Calabria	1183	146	292	7580	618
## 58	1	Calabria	1261	153	219	7355	818
## 59	1	Calabria	1198	136	239	7639	633
## 60	1	Calabria	1225	134	232	7658	616
## 61	1	Calabria	1339	166	208	7190	923
## 62	1	Calabria	1132	157	240	7641	638
## 63	1	Calabria	1381	183	245	7385	609
## 64	1	Calabria	1409	128	257	7257	759
## 65	1	Calabria	1306	127	250	7254	869
## 66	1	Calabria	1372	120	250	7355	702
## 67	1	Calabria	1336	113	242	7293	855
## 68	1	Calabria	1401	151	238	7164	857
## 69	1	Calabria	1390	119	234	7236	823

## 70	1	Calabria	1432	152	281	7029	949
## 71	1	Calabria	1412	124	298	7182	790
## 72	1	Calabria	1366	147	291	7197	783
## 73	1	Calabria	1383	118	273	7282	738
## 74	1	Calabria	1283	102	263	7400	763
## 75	1	Calabria	1296	136	260	7380	780
## 76	1	Calabria	1287	108	287	7343	826
## 77	1	Calabria	1351	159	296	7229	810
## 78	1	Calabria	1241	97	268	7499	709
## 79	1	Calabria	1267	101	300	7230	898
## 80	1	Calabria	1235	138	252	7322	861
## 81	1	Calabria	1255	103	223	7395	848
## 82	1	South-Apulia	1454	183	196	7057	1014
## 83	1	South-Apulia	1347	194	197	7277	895
## 84	1	South-Apulia	1364	204	225	6929	1084
## 85	1	South-Apulia	1410	199	216	7130	955
## 86	1	South-Apulia	1384	178	208	7105	999
## 87	1	South-Apulia	1412	185	217	6842	1203
## 88	1	South-Apulia	1410	232	280	6715	1233
## 89	1	South-Apulia	1509	209	257	6647	1240
## 90	1	South-Apulia	1317	197	256	7036	1067
## 91	1	South-Apulia	1286	192	203	7132	1053
## 92	1	South-Apulia	1273	191	202	6862	1303
## 93	1	South-Apulia	1463	183	183	6747	1307
## 94	1	South-Apulia	1399	187	191	6861	1233
## 95	1	South-Apulia	1413	193	208	6875	1202
## 96	1	South-Apulia	1369	206	203	6953	1168
## 97	1	South-Apulia	1488	172	170	6920	1144
## 98	1	South-Apulia	1323	160	205	6911	1298
## 99	1	South-Apulia	1311	166	170	6902	1312
## 100	1	South-Apulia	1286	163	183	7040	1230
## 101	1	South-Apulia	1380	173	188	7038	1139
## 102	1	South-Apulia	1394	164	223	7086	1042
## 103	1	South-Apulia	1324	174	198	6863	1289
## 104	1	South-Apulia	1290	157	192	7000	1263
## 105	1	South-Apulia	1361	163	196	6888	1273
## 106	1	South-Apulia	1387	182	242	6913	1101
## 107	1	South-Apulia	1369	180	181	7000	1130
## 108	1	South-Apulia	1303	165	175	7025	1243
## 109	1	South-Apulia	1346	160	169	7072	1151
## 110	1	South-Apulia	1369	171	184	6937	1246
## 111	1	South-Apulia	1305	172	169	7004	1260
## 112	1	South-Apulia	1351	179	186	6935	1243
## 113	1	South-Apulia	1283	151	182	7000	1271
## 114	1	South-Apulia	1449	175	198	6883	1162
## 115	1	South-Apulia	1310	180	183	7054	1202
## 116	1	South-Apulia	1360	163	176	6901	1280
## 117	1	South-Apulia	1300	187	196	6920	1253
## 118	1	South-Apulia	1368	171	218	7010	1057
## 119	1	South-Apulia	1207	151	156	7159	1234
## 120	1	South-Apulia	1348	154	183	6917	1277
## 121	1	South-Apulia	1334	186	229	7261	827
## 122	1	South-Apulia	1301	156	207	7003	1229
## 123	1	South-Apulia	1226	181	213	6961	1230

## 124	1	South-Apulia	1201	168	190	7100	1216
## 125	1	South-Apulia	1297	153	177	7004	1260
## 126	1	South-Apulia	1248	163	158	7103	1222
## 127	1	South-Apulia	1335	159	197	6974	1220
## 128	1	South-Apulia	1219	167	171	7087	1254
## 129	1	South-Apulia	1318	179	177	7030	1194
## 130	1	South-Apulia	1264	167	166	7130	1187
## 131	1	South-Apulia	1201	175	201	7129	1193
## 132	1	South-Apulia	1252	180	181	7055	1214
## 133	1	South-Apulia	1273	182	209	6965	1191
## 134	1	South-Apulia	1351	179	170	7034	1154
## 135	1	South-Apulia	1336	155	212	7103	1086
## 136	1	South-Apulia	1499	201	182	6803	1204
## 137	1	South-Apulia	1425	198	193	7032	1041
## 138	1	South-Apulia	1358	204	227	6962	1109
## 139	1	South-Apulia	1346	181	257	7147	933
## 140	1	South-Apulia	1392	186	256	6732	1278
## 141	1	South-Apulia	1311	166	222	7006	1147
## 142	1	South-Apulia	1314	171	229	6923	1198
## 143	1	South-Apulia	1409	200	207	6842	1224
## 144	1	South-Apulia	1342	174	221	6993	1147
## 145	1	South-Apulia	1387	182	206	7100	1020
## 146	1	South-Apulia	1413	202	205	6920	1165
## 147	1	South-Apulia	1430	209	225	6800	1200
## 148	1	South-Apulia	1336	185	223	6956	1155
## 149	1	South-Apulia	1372	200	200	6916	1189
## 150	1	South-Apulia	1330	157	228	7055	1108
## 151	1	South-Apulia	1412	207	208	6822	1239
## 152	1	South-Apulia	1321	209	217	6948	1178
## 153	1	South-Apulia	1401	200	217	6980	1073
## 154	1	South-Apulia	1401	214	217	6734	1293
## 155	1	South-Apulia	1457	168	242	6724	1266
## 156	1	South-Apulia	1451	199	221	6835	1177
## 157	1	South-Apulia	1438	206	248	6806	1183
## 158	1	South-Apulia	1462	204	237	6644	1309
## 159	1	South-Apulia	1529	215	203	6602	1310
## 160	1	South-Apulia	1510	189	245	6752	1188
## 161	1	South-Apulia	1437	222	184	6803	1240
## 162	1	South-Apulia	1327	129	247	7024	1157
## 163	1	South-Apulia	1438	172	252	6630	1380
## 164	1	South-Apulia	1447	176	189	6849	1180
## 165	1	South-Apulia	1355	144	214	6972	1198
## 166	1	South-Apulia	1369	156	241	6890	1209
## 167	1	South-Apulia	1471	188	276	6697	1269
## 168	1	South-Apulia	1456	179	240	6738	1267
## 169	1	South-Apulia	1314	140	207	7020	1220
## 170	1	South-Apulia	1408	176	192	6909	1195
## 171	1	South-Apulia	1397	172	191	7107	1018
## 172	1	South-Apulia	1413	191	186	6937	1180
## 173	1	South-Apulia	1539	194	213	6764	1178
## 174	1	South-Apulia	1304	159	234	7019	1174
## 175	1	South-Apulia	1341	160	231	7033	1069
## 176	1	South-Apulia	1508	208	249	6641	1311
## 177	1	South-Apulia	1515	226	257	6595	1287

## 178	1	South-Apulia	1262	165	235	7120	1113
## 179	1	South-Apulia	1307	197	238	7003	1144
## 180	1	South-Apulia	1294	159	253	7009	1190
## 181	1	South-Apulia	1460	187	215	6843	1172
## 182	1	South-Apulia	1476	187	203	6837	1197
## 183	1	South-Apulia	1482	178	197	6814	1201
## 184	1	South-Apulia	1388	176	185	7008	1111
## 185	1	South-Apulia	1367	172	235	7066	1054
## 186	1	South-Apulia	1272	207	205	7152	1098
## 187	1	South-Apulia	1323	157	234	7132	1022
## 188	1	South-Apulia	1206	218	242	7193	1002
## 189	1	South-Apulia	1383	157	217	7018	1090
## 190	1	South-Apulia	1521	190	238	6956	986
## 191	1	South-Apulia	1350	168	227	6986	1165
## 192	1	South-Apulia	1422	181	218	6813	1230
## 193	1	South-Apulia	1298	166	224	6986	1162
## 194	1	South-Apulia	1447	236	245	6607	1336
## 195	1	South-Apulia	1347	197	211	6795	1300
## 196	1	South-Apulia	1339	170	253	6989	1110
## 197	1	South-Apulia	1388	183	216	6867	1208
## 198	1	South-Apulia	1527	260	232	6488	1370
## 199	1	South-Apulia	1495	237	236	6571	1318
## 200	1	South-Apulia	1487	246	251	6504	1390
## 201	1	South-Apulia	1399	180	232	6855	1190
## 202	1	South-Apulia	1489	215	242	6777	1145
## 203	1	South-Apulia	1339	166	226	6928	1198
## 204	1	South-Apulia	1482	246	238	6444	1462
## 205	1	South-Apulia	1434	172	255	6646	1354
## 206	1	South-Apulia	1347	156	214	6850	1313
## 207	1	South-Apulia	1340	158	233	6848	1272
## 208	1	South-Apulia	1453	180	244	6752	1238
## 209	1	South-Apulia	1306	149	226	7082	1097
## 210	1	South-Apulia	1349	161	217	6997	1138
## 211	1	South-Apulia	1254	151	205	7319	947
## 212	1	South-Apulia	1168	144	220	7230	1109
## 213	1	South-Apulia	1346	167	224	6959	1111
## 214	1	South-Apulia	1390	184	212	6898	1189
## 215	1	South-Apulia	1283	149	224	7077	1104
## 216	1	South-Apulia	1214	137	232	7269	1005
## 217	1	South-Apulia	1491	227	205	6941	988
## 218	1	South-Apulia	1479	218	207	7039	887
## 219	1	South-Apulia	1445	174	228	6875	1123
## 220	1	South-Apulia	1439	183	218	6775	1226
## 221	1	South-Apulia	1387	154	204	6991	1090
## 222	1	South-Apulia	1426	169	192	7025	1043
## 223	1	South-Apulia	1451	200	208	6980	1006
## 224	1	South-Apulia	1493	204	188	6913	1044
## 225	1	South-Apulia	1419	192	207	6996	1014
## 226	1	South-Apulia	1342	177	199	7172	952
## 227	1	South-Apulia	1349	152	236	7145	949
## 228	1	South-Apulia	1440	196	208	6938	1070
## 229	1	South-Apulia	1460	215	197	6918	1081
## 230	1	South-Apulia	1249	133	205	7417	827
## 231	1	South-Apulia	1348	159	238	7017	1081

## 232	1	South-Apulia	1341	155	244	6958	1144
## 233	1	South-Apulia	1398	149	204	7182	907
## 234	1	South-Apulia	1454	200	199	6910	1090
## 235	1	South-Apulia	1334	153	219	6928	1214
## 236	1	South-Apulia	1438	204	189	7107	910
## 237	1	South-Apulia	1303	138	212	7170	1016
## 238	1	South-Apulia	1323	147	210	7108	1070
## 239	1	South-Apulia	1417	169	207	6875	1184
## 240	1	South-Apulia	1360	167	225	6883	1220
## 241	1	South-Apulia	1420	179	214	6923	1121
## 242	1	South-Apulia	1472	218	214	6724	1238
## 243	1	South-Apulia	1368	174	205	7042	1066
## 244	1	South-Apulia	1367	173	228	6948	1141
## 245	1	South-Apulia	1403	173	209	6843	1210
## 246	1	South-Apulia	1413	197	206	6737	1387
## 247	1	South-Apulia	1201	138	207	7011	1269
## 248	1	South-Apulia	1359	180	207	6895	1203
## 249	1	South-Apulia	1518	198	225	6681	1243
## 250	1	South-Apulia	1434	185	189	6771	1269
## 251	1	South-Apulia	1367	162	179	6772	1368
## 252	1	South-Apulia	1461	181	197	6783	1246
## 253	1	South-Apulia	1368	161	198	7030	1095
## 254	1	South-Apulia	1419	159	215	6862	1193
## 255	1	South-Apulia	1514	162	298	6725	1119
## 256	1	South-Apulia	1328	171	253	6987	1030
## 257	1	South-Apulia	1469	160	337	6675	1127
## 258	1	Sicily	1222	133	227	7425	824
## 259	1	Sicily	1639	172	331	6510	1124
## 260	1	Sicily	1345	133	272	6801	1194
## 261	1	Sicily	1339	170	275	6838	1060
## 262	1	Sicily	1194	135	263	7277	889
## 263	1	Sicily	1112	68	375	7770	448
## 264	1	Sicily	1222	70	329	7605	566
## 265	1	Sicily	1136	72	341	7616	661
## 266	1	Sicily	926	41	277	7815	784
## 267	1	Sicily	1105	69	373	7714	532
## 268	1	Sicily	1109	79	305	7576	763
## 269	1	Sicily	1284	93	265	7235	893
## 270	1	Sicily	1120	69	277	7416	946
## 271	1	Sicily	916	52	281	7870	694
## 272	1	Sicily	905	49	288	7747	812
## 273	1	Sicily	1206	55	287	7329	935
## 274	1	Sicily	1457	182	267	7020	863
## 275	1	Sicily	1327	140	193	7328	823
## 276	1	Sicily	1303	100	251	7045	1049
## 277	1	Sicily	1444	175	259	6876	1027
## 278	1	Sicily	1505	243	226	6962	858
## 279	1	Sicily	1429	162	223	6917	1041
## 280	1	Sicily	1491	162	211	6994	928
## 281	1	Sicily	1393	128	211	7189	870
## 282	1	Sicily	1404	134	210	7110	923
## 283	1	Sicily	1222	130	214	7374	856
## 284	1	Sicily	1153	74	316	7593	705
## 285	1	Sicily	1169	76	307	7553	728

## 286	1	Sicily	1369	104	237	7375	775
## 287	1	Sicily	993	58	267	7743	773
## 288	1	Sicily	980	53	254	7719	815
## 289	1	Sicily	967	55	273	7692	833
## 290	1	Sicily	1128	73	354	7527	728
## 291	1	Sicily	1188	85	273	7445	814
## 292	1	Sicily	1257	95	247	7405	812
## 293	1	Sicily	1262	88	301	7471	704
## 294	1	South-Apulia	1283	153	196	7107	1115
## 295	1	South-Apulia	1263	155	199	7140	1148
## 296	1	South-Apulia	1369	158	215	7160	958
## 297	1	South-Apulia	1353	172	175	6965	1212
## 298	1	South-Apulia	1187	139	185	7427	952
## 299	1	South-Apulia	1732	231	156	6437	1313
## 300	1	South-Apulia	1620	255	166	6628	1212
## 301	1	South-Apulia	1543	172	193	6740	1157
## 302	1	South-Apulia	1498	170	195	6804	1206
## 303	1	South-Apulia	1399	169	171	7011	1100
## 304	1	South-Apulia	1293	156	191	7101	1111
## 305	1	South-Apulia	1420	175	152	7004	1149
## 306	1	South-Apulia	1721	238	255	6300	1350
## 307	1	South-Apulia	1742	221	156	6415	1315
## 308	1	South-Apulia	1391	187	189	6975	1062
## 309	1	South-Apulia	1517	206	249	6680	1205
## 310	1	South-Apulia	1269	157	193	7140	1148
## 311	1	South-Apulia	1577	204	208	6732	1183
## 312	1	South-Apulia	1590	241	195	6705	1149
## 313	1	South-Apulia	1621	280	197	6608	1179
## 314	1	South-Apulia	1753	275	236	6367	1214
## 315	1	South-Apulia	1679	260	177	6568	1191
## 316	1	South-Apulia	1419	203	176	6973	1083
## 317	1	South-Apulia	1693	236	174	6499	1204
## 318	1	South-Apulia	1692	270	234	6499	1196
## 319	1	South-Apulia	1638	252	215	6570	1199
## 320	1	South-Apulia	1497	247	219	6621	1270
## 321	1	South-Apulia	1442	222	194	6677	1314
## 322	1	South-Apulia	1680	270	170	6440	1310
## 323	1	South-Apulia	1463	164	185	6909	1154
## 324	2	Inland-Sardinia	1129	120	222	7272	1112
## 325	2	Inland-Sardinia	1042	135	210	7376	1116
## 326	2	Inland-Sardinia	1103	96	210	7380	1085
## 327	2	Inland-Sardinia	1118	97	221	7279	1154
## 328	2	Inland-Sardinia	1052	95	215	7388	1126
## 329	2	Inland-Sardinia	1116	102	231	7290	1168
## 330	2	Inland-Sardinia	1108	132	231	7319	1101
## 331	2	Inland-Sardinia	1129	108	212	7386	1074
## 332	2	Inland-Sardinia	1085	91	223	7384	1126
## 333	2	Inland-Sardinia	1104	103	233	7322	1147
## 334	2	Inland-Sardinia	1098	88	212	7338	1140
## 335	2	Coast-Sardinia	1135	98	251	7120	1314
## 336	2	Coast-Sardinia	1158	108	245	7065	1326
## 337	2	Coast-Sardinia	1133	110	241	7080	1342
## 338	2	Coast-Sardinia	1095	125	250	7120	1305
## 339	2	Coast-Sardinia	1201	87	238	6990	1383

## 340	2	Coast-Sardinia	1213	112	245	7007	1335
## 341	2	Inland-Sardinia	1108	92	231	7367	1110
## 342	2	Inland-Sardinia	1075	103	207	7413	1096
## 343	2	Inland-Sardinia	1059	96	228	7386	1128
## 344	2	Inland-Sardinia	1176	92	207	7347	1057
## 345	2	Inland-Sardinia	1159	98	213	7320	1108
## 346	2	Inland-Sardinia	1132	80	201	7398	1095
## 347	2	Inland-Sardinia	1107	75	220	7399	1096
## 348	2	Inland-Sardinia	1092	104	234	7355	1126
## 349	2	Inland-Sardinia	1119	81	219	7409	1057
## 350	2	Inland-Sardinia	1106	93	212	7381	1104
## 351	2	Inland-Sardinia	1047	101	238	7385	1120
## 352	2	Inland-Sardinia	1165	99	214	7331	1101
## 353	2	Inland-Sardinia	1158	84	201	7327	1123
## 354	2	Inland-Sardinia	1095	88	203	7415	1093
## 355	2	Inland-Sardinia	1176	75	205	7396	1107
## 356	2	Inland-Sardinia	1103	109	220	7335	1140
## 357	2	Inland-Sardinia	1112	92	209	7356	1125
## 358	2	Inland-Sardinia	1091	93	222	7377	1113
## 359	2	Inland-Sardinia	1080	98	219	7371	1125
## 360	2	Inland-Sardinia	1051	108	227	7403	1114
## 361	2	Inland-Sardinia	1096	84	211	7415	1091
## 362	2	Inland-Sardinia	1142	97	225	7341	1101
## 363	2	Inland-Sardinia	1047	96	236	7399	1107
## 364	2	Inland-Sardinia	1114	86	210	7359	1116
## 365	2	Inland-Sardinia	1140	93	241	7324	1098
## 366	2	Inland-Sardinia	1075	91	200	7410	1107
## 367	2	Inland-Sardinia	1092	106	219	7427	1125
## 368	2	Inland-Sardinia	1076	95	204	7408	1130
## 369	2	Inland-Sardinia	1178	89	201	7381	1099
## 370	2	Inland-Sardinia	1095	104	223	7367	1111
## 371	2	Coast-Sardinia	1166	97	272	6971	1390
## 372	2	Coast-Sardinia	1154	119	257	7130	1253
## 373	2	Coast-Sardinia	1177	111	241	6882	1470
## 374	2	Coast-Sardinia	1160	96	240	7043	1357
## 375	2	Coast-Sardinia	1122	104	241	7145	1313
## 376	2	Coast-Sardinia	1132	99	257	7065	1362
## 377	2	Coast-Sardinia	1096	100	260	7162	1282
## 378	2	Coast-Sardinia	1131	87	233	7144	1307
## 379	2	Coast-Sardinia	1184	105	258	7020	1340
## 380	2	Coast-Sardinia	1135	94	235	7123	1320
## 381	2	Coast-Sardinia	1084	96	240	7164	1330
## 382	2	Coast-Sardinia	1086	127	252	7159	1285
## 383	2	Coast-Sardinia	1140	95	258	7085	1347
## 384	2	Coast-Sardinia	1138	101	254	7103	1310
## 385	2	Coast-Sardinia	1159	110	261	7068	1297
## 386	2	Inland-Sardinia	1051	78	211	7421	1146
## 387	2	Inland-Sardinia	1048	79	213	7439	1130
## 388	2	Inland-Sardinia	1061	86	220	7421	1102
## 389	2	Inland-Sardinia	1105	88	210	7353	1142
## 390	2	Inland-Sardinia	1145	35	237	7208	1118
## 391	2	Inland-Sardinia	1049	96	219	7303	1168
## 392	2	Inland-Sardinia	1105	120	218	7302	1158
## 393	2	Inland-Sardinia	1030	84	214	7403	1177

## 394	2	Inland-Sardinia	1070	98	215	7280	1240
## 395	2	Inland-Sardinia	1103	81	208	7310	1177
## 396	2	Inland-Sardinia	1040	101	205	7368	1176
## 397	2	Inland-Sardinia	1100	95	210	7320	1113
## 398	2	Inland-Sardinia	1118	85	199	7415	1060
## 399	2	Inland-Sardinia	1065	98	230	7345	1163
## 400	2	Inland-Sardinia	1131	78	221	7358	1120
## 401	2	Inland-Sardinia	1080	120	218	7296	1145
## 402	2	Inland-Sardinia	1075	86	231	7403	1109
## 403	2	Inland-Sardinia	1040	103	228	7364	1173
## 404	2	Inland-Sardinia	1128	82	203	7320	1148
## 405	2	Inland-Sardinia	1060	111	231	7363	1149
## 406	2	Inland-Sardinia	1103	78	220	7365	1149
## 407	2	Inland-Sardinia	1110	91	201	7318	1185
## 408	2	Inland-Sardinia	1091	108	218	7383	1183
## 409	2	Inland-Sardinia	1094	96	220	7341	1127
## 410	2	Coast-Sardinia	1131	87	208	7170	1308
## 411	2	Coast-Sardinia	1175	108	214	7076	1307
## 412	2	Coast-Sardinia	1076	77	202	7243	1305
## 413	2	Coast-Sardinia	1120	90	240	7068	1383
## 414	2	Coast-Sardinia	1152	111	238	7080	1372
## 415	2	Coast-Sardinia	1141	95	250	7035	1388
## 416	2	Coast-Sardinia	1098	103	267	7135	1301
## 417	2	Coast-Sardinia	1126	100	236	7062	1380
## 418	2	Coast-Sardinia	1087	89	243	7200	1302
## 419	2	Coast-Sardinia	1115	96	236	7085	1372
## 420	2	Coast-Sardinia	1178	92	241	7006	1376
## 421	2	Coast-Sardinia	1162	106	242	7025	1368
## 422	3	Umbria	1085	70	180	7955	605
## 423	3	Umbria	1085	70	185	7955	600
## 424	3	Umbria	1090	60	190	7950	600
## 425	3	Umbria	1080	65	189	7960	602
## 426	3	Umbria	1090	60	195	7955	600
## 427	3	Umbria	1105	55	200	7900	600
## 428	3	Umbria	1060	75	175	7975	610
## 429	3	Umbria	1050	70	170	7977	605
## 430	3	Umbria	1100	55	198	7905	600
## 431	3	Umbria	1065	65	178	7965	605
## 432	3	Umbria	1085	60	188	7955	602
## 433	3	Umbria	1080	65	180	7960	605
## 434	3	Umbria	1085	60	190	7955	602
## 435	3	Umbria	1075	68	195	7960	602
## 436	3	Umbria	1090	58	192	7950	600
## 437	3	Umbria	1095	60	198	7945	600
## 438	3	Umbria	1090	58	195	7950	600
## 439	3	Umbria	1095	58	198	7950	602
## 440	3	Umbria	1090	58	195	7940	600
## 441	3	Umbria	1095	58	198	7945	600
## 442	3	Umbria	1095	55	200	7940	600
## 443	3	Umbria	1080	70	188	7965	608
## 444	3	Umbria	1090	60	195	7950	600
## 445	3	Umbria	1105	55	200	7900	595
## 446	3	Umbria	1110	50	205	7900	595
## 447	3	Umbria	1075	70	198	7978	608

## 448	3	Umbria	1075	65	185	7980	608
## 449	3	Umbria	1065	75	180	7975	610
## 450	3	Umbria	1070	75	188	7980	602
## 451	3	Umbria	1070	75	188	7980	602
## 452	3	Umbria	1100	70	200	7910	610
## 453	3	Umbria	1075	70	185	7960	610
## 454	3	Umbria	1050	78	175	7990	610
## 455	3	Umbria	1090	60	198	7945	600
## 456	3	Umbria	1050	78	188	7990	608
## 457	3	Umbria	1075	70	190	7975	605
## 458	3	Umbria	1098	54	202	7945	595
## 459	3	Umbria	1105	15	198	8005	575
## 460	3	Umbria	1110	75	220	7915	510
## 461	3	Umbria	1058	50	178	7988	626
## 462	3	Umbria	1115	30	225	7955	600
## 463	3	Umbria	1105	30	198	7995	570
## 464	3	Umbria	1072	49	178	7980	615
## 465	3	Umbria	1110	15	210	7990	570
## 466	3	Umbria	1110	80	215	7910	525
## 467	3	Umbria	1055	60	175	7985	620
## 468	3	Umbria	1100	80	215	7930	535
## 469	3	Umbria	1105	55	205	7965	600
## 470	3	Umbria	1095	50	210	7948	600
## 471	3	Umbria	1110	50	220	7950	600
## 472	3	Umbria	1092	37	210	7955	600
## 473	3	East-Liguria	1290	60	260	7550	670
## 474	3	East-Liguria	1170	80	230	7690	720
## 475	3	East-Liguria	1100	90	250	7680	760
## 476	3	East-Liguria	1120	70	240	7720	730
## 477	3	East-Liguria	1160	70	250	7650	750
## 478	3	East-Liguria	1200	50	210	7770	690
## 479	3	East-Liguria	1140	50	200	7990	580
## 480	3	East-Liguria	1220	80	240	7610	760
## 481	3	East-Liguria	1180	90	250	7520	800
## 482	3	East-Liguria	1210	70	250	7560	780
## 483	3	East-Liguria	1220	80	220	7540	770
## 484	3	East-Liguria	1180	100	190	7520	820
## 485	3	East-Liguria	1160	90	220	7580	790
## 486	3	East-Liguria	1130	100	240	7620	780
## 487	3	East-Liguria	1080	100	260	7710	750
## 488	3	East-Liguria	1090	90	280	7730	720
## 489	3	East-Liguria	1020	100	270	7770	710
## 490	3	East-Liguria	1090	90	250	7680	760
## 491	3	East-Liguria	1120	100	260	7720	680
## 492	3	East-Liguria	1080	80	240	7830	670
## 493	3	East-Liguria	1160	70	230	7860	640
## 494	3	East-Liguria	1100	80	240	7820	670
## 495	3	East-Liguria	1050	100	250	7930	630
## 496	3	East-Liguria	1090	90	270	7780	690
## 497	3	East-Liguria	1120	80	260	7750	680
## 498	3	East-Liguria	1120	100	250	7680	730
## 499	3	East-Liguria	1190	90	230	7670	710
## 500	3	East-Liguria	1170	110	250	7620	740
## 501	3	East-Liguria	1120	100	230	7720	730

## 502	3	East-Liguria	1190	80	270	7690	720
## 503	3	East-Liguria	1400	90	270	7420	800
## 504	3	East-Liguria	1350	80	250	7520	760
## 505	3	East-Liguria	1090	60	220	7890	670
## 506	3	East-Liguria	1150	90	230	7790	650
## 507	3	East-Liguria	1240	90	220	7820	590
## 508	3	East-Liguria	1220	100	240	7890	530
## 509	3	East-Liguria	1180	80	250	7870	580
## 510	3	East-Liguria	1170	110	240	7730	630
## 511	3	East-Liguria	1170	100	280	7710	640
## 512	3	East-Liguria	1180	80	220	7790	680
## 513	3	East-Liguria	1200	90	240	7820	590
## 514	3	East-Liguria	1140	90	240	7880	570
## 515	3	East-Liguria	1160	70	210	7870	580
## 516	3	East-Liguria	1130	80	250	7780	650
## 517	3	East-Liguria	1150	80	240	7800	630
## 518	3	East-Liguria	1110	70	240	7820	670
## 519	3	East-Liguria	1150	70	220	7850	620
## 520	3	East-Liguria	1180	80	240	7760	670
## 521	3	East-Liguria	1020	80	250	7920	680
## 522	3	East-Liguria	610	80	230	8410	650
## 523	3	West-Liguria	1190	150	290	7340	1020
## 524	3	West-Liguria	1110	130	210	7550	1000
## 525	3	West-Liguria	1020	100	220	7530	1030
## 526	3	West-Liguria	1070	120	210	7600	990
## 527	3	West-Liguria	1010	90	350	7480	1050
## 528	3	West-Liguria	1060	140	240	7680	830
## 529	3	West-Liguria	1060	140	270	7620	880
## 530	3	West-Liguria	1030	100	230	7740	900
## 531	3	West-Liguria	1120	130	250	7530	970
## 532	3	West-Liguria	1030	110	220	7760	980
## 533	3	West-Liguria	1070	100	230	7600	990
## 534	3	West-Liguria	1140	180	220	7610	850
## 535	3	West-Liguria	1090	180	230	7590	860
## 536	3	West-Liguria	980	110	300	7720	910
## 537	3	West-Liguria	980	90	330	7540	1040
## 538	3	West-Liguria	960	90	200	7810	940
## 539	3	West-Liguria	990	90	210	7780	930
## 540	3	West-Liguria	1060	120	210	7600	1010
## 541	3	West-Liguria	1240	150	250	7610	730
## 542	3	West-Liguria	1060	90	310	7850	690
## 543	3	West-Liguria	1020	100	290	7620	960
## 544	3	West-Liguria	970	90	220	7700	1020
## 545	3	West-Liguria	1180	130	220	7450	1010
## 546	3	West-Liguria	1060	140	240	7690	850
## 547	3	West-Liguria	990	100	250	7630	1030
## 548	3	West-Liguria	1010	90	350	7630	940
## 549	3	West-Liguria	1040	90	250	7780	820
## 550	3	West-Liguria	1040	90	250	7810	810
## 551	3	West-Liguria	1020	90	350	7620	920
## 552	3	West-Liguria	1020	90	260	7620	1010
## 553	3	West-Liguria	1010	90	350	7610	930
## 554	3	West-Liguria	920	110	340	7720	910
## 555	3	West-Liguria	1030	100	250	7710	900

## 556	3	West-Liguria	960	90	300	7820	830
## 557	3	West-Liguria	1030	110	210	7810	840
## 558	3	West-Liguria	1010	100	240	7710	910
## 559	3	West-Liguria	1020	90	240	7800	850
## 560	3	West-Liguria	1120	90	300	7650	830
## 561	3	West-Liguria	1090	90	290	7710	800
## 562	3	West-Liguria	1100	120	280	7630	770
## 563	3	West-Liguria	1090	80	240	7820	760
## 564	3	West-Liguria	1150	90	250	7720	810
## 565	3	West-Liguria	1110	90	230	7810	750
## 566	3	West-Liguria	1010	110	210	7720	950
## 567	3	West-Liguria	1070	100	220	7730	870
## 568	3	West-Liguria	1280	110	290	7490	790
## 569	3	West-Liguria	1060	100	270	7740	810
## 570	3	West-Liguria	1010	90	210	7720	970
## 571	3	West-Liguria	990	120	250	7750	870
## 572	3	West-Liguria	960	80	240	7950	740
##		linolenic	arachidic	eicosenoic			
## 1		36	60	29			
## 2		31	61	29			
## 3		31	63	29			
## 4		50	78	35			
## 5		50	80	46			
## 6		51	70	44			
## 7		49	56	29			
## 8		39	64	35			
## 9		46	83	33			
## 10		26	52	30			
## 11		21	65	24			
## 12		30	62	44			
## 13		50	79	33			
## 14		41	79	32			
## 15		50	75	41			
## 16		32	60	29			
## 17		42	75	31			
## 18		45	70	28			
## 19		46	75	30			
## 20		43	78	33			
## 21		56	83	45			
## 22		56	86	40			
## 23		54	89	51			
## 24		42	74	26			
## 25		47	82	38			
## 26		42	60	32			
## 27		43	63	30			
## 28		45	63	18			
## 29		46	64	23			
## 30		43	62	38			
## 31		40	68	41			
## 32		50	70	38			
## 33		46	72	35			
## 34		48	72	21			
## 35		40	69	27			
## 36		54	70	28			

## 37	44	67	28
## 38	57	65	36
## 39	46	66	33
## 40	48	75	36
## 41	49	64	28
## 42	47	65	25
## 43	48	61	37
## 44	51	65	41
## 45	47	54	26
## 46	48	61	32
## 47	46	71	24
## 48	41	69	20
## 49	47	65	31
## 50	41	56	16
## 51	45	69	29
## 52	45	67	23
## 53	43	72	27
## 54	44	53	24
## 55	48	57	28
## 56	39	60	30
## 57	38	51	23
## 58	52	70	26
## 59	27	55	19
## 60	36	49	26
## 61	40	69	25
## 62	45	60	31
## 63	47	70	25
## 64	43	57	16
## 65	47	68	24
## 66	44	68	28
## 67	38	60	18
## 68	45	72	36
## 69	40	62	41
## 70	39	55	25
## 71	45	68	28
## 72	51	70	34
## 73	45	68	29
## 74	54	65	28
## 75	48	51	18
## 76	44	44	23
## 77	36	60	22
## 78	52	69	36
## 79	74	65	34
## 80	54	66	36
## 81	47	56	30
## 82	27	46	19
## 83	25	46	15
## 84	21	50	14
## 85	21	48	19
## 86	29	67	26
## 87	34	72	32
## 88	32	60	24
## 89	42	62	30
## 90	40	60	22

## 91	38	65	28
## 92	43	70	28
## 93	36	60	24
## 94	38	60	17
## 95	30	60	18
## 96	35	50	16
## 97	37	54	14
## 98	24	50	17
## 99	41	69	28
## 100	29	57	12
## 101	31	44	14
## 102	24	43	23
## 103	36	70	21
## 104	26	51	19
## 105	37	58	24
## 106	44	68	30
## 107	39	45	24
## 108	31	41	16
## 109	39	48	15
## 110	30	48	15
## 111	28	50	11
## 112	36	50	19
## 113	40	52	21
## 114	40	70	22
## 115	26	32	12
## 116	28	65	27
## 117	41	76	25
## 118	41	54	26
## 119	27	51	14
## 120	48	56	16
## 121	34	56	20
## 122	41	48	14
## 123	47	74	26
## 124	43	64	16
## 125	35	60	16
## 126	31	60	14
## 127	36	60	17
## 128	35	50	16
## 129	35	42	25
## 130	22	52	12
## 131	36	49	15
## 132	31	59	38
## 133	43	74	23
## 134	35	66	10
## 135	33	55	20
## 136	30	56	24
## 137	31	52	17
## 138	41	65	34
## 139	40	60	36
## 140	53	64	29
## 141	41	80	27
## 142	47	76	42
## 143	31	60	27
## 144	36	64	23

## 145	34	54	17
## 146	36	46	13
## 147	32	59	27
## 148	56	73	16
## 149	33	50	22
## 150	42	55	25
## 151	36	51	28
## 152	42	62	23
## 153	40	68	21
## 154	44	69	27
## 155	54	59	30
## 156	37	51	29
## 157	34	57	28
## 158	42	54	28
## 159	45	69	27
## 160	36	52	28
## 161	43	56	16
## 162	38	56	22
## 163	40	64	24
## 164	42	64	26
## 165	33	60	24
## 166	42	63	30
## 167	34	51	16
## 168	41	65	14
## 169	28	59	12
## 170	45	50	25
## 171	36	50	29
## 172	31	46	13
## 173	38	58	16
## 174	38	53	19
## 175	40	67	33
## 176	25	43	20
## 177	41	63	16
## 178	32	51	21
## 179	37	50	24
## 180	30	52	13
## 181	35	56	32
## 182	36	48	22
## 183	40	64	24
## 184	48	53	31
## 185	35	45	26
## 186	37	52	22
## 187	38	58	31
## 188	37	54	25
## 189	40	60	37
## 190	36	50	23
## 191	29	58	17
## 192	30	59	21
## 193	34	65	31
## 194	33	51	21
## 195	32	59	34
## 196	29	63	23
## 197	28	61	21
## 198	31	45	20

## 199	32	58	26
## 200	29	53	19
## 201	32	66	22
## 202	30	60	22
## 203	30	60	23
## 204	27	50	20
## 205	27	59	25
## 206	25	48	19
## 207	32	63	25
## 208	34	54	23
## 209	33	61	24
## 210	31	62	23
## 211	28	54	23
## 212	31	52	28
## 213	30	49	23
## 214	29	44	19
## 215	30	57	32
## 216	32	55	23
## 217	33	68	34
## 218	36	65	36
## 219	29	69	31
## 220	32	66	29
## 221	34	74	32
## 222	31	64	27
## 223	30	62	31
## 224	32	61	35
## 225	36	70	36
## 226	34	65	33
## 227	35	75	29
## 228	32	61	26
## 229	28	55	23
## 230	33	72	33
## 231	31	67	25
## 232	32	68	26
## 233	29	76	30
## 234	30	62	25
## 235	33	66	24
## 236	33	63	27
## 237	34	69	25
## 238	33	61	20
## 239	34	57	27
## 240	31	55	27
## 241	33	56	27
## 242	29	53	23
## 243	31	57	26
## 244	32	53	24
## 245	33	63	33
## 246	34	60	31
## 247	37	64	35
## 248	33	61	30
## 249	29	57	24
## 250	30	62	25
## 251	33	64	27
## 252	26	57	23

## 253	33	59	31
## 254	35	60	31
## 255	45	93	30
## 256	38	83	39
## 257	44	94	36
## 258	36	69	35
## 259	46	91	32
## 260	48	83	37
## 261	46	88	43
## 262	44	95	41
## 263	52	69	45
## 264	48	67	43
## 265	49	65	32
## 266	45	65	25
## 267	51	68	37
## 268	45	64	36
## 269	43	77	46
## 270	42	59	36
## 271	42	64	58
## 272	49	71	56
## 273	44	74	42
## 274	41	84	37
## 275	36	87	35
## 276	40	86	40
## 277	34	78	32
## 278	30	72	27
## 279	37	77	40
## 280	37	97	38
## 281	38	93	40
## 282	40	101	43
## 283	38	89	45
## 284	42	64	32
## 285	43	69	32
## 286	39	70	15
## 287	41	62	44
## 288	44	69	47
## 289	45	63	47
## 290	44	76	38
## 291	44	73	42
## 292	43	70	35
## 293	43	71	31
## 294	37	60	28
## 295	31	42	18
## 296	38	69	32
## 297	28	75	19
## 298	29	56	22
## 299	45	62	23
## 300	29	62	27
## 301	52	87	34
## 302	35	66	23
## 303	36	72	16
## 304	32	60	31
## 305	27	50	20
## 306	35	70	28

## 307	43	82	23
## 308	52	70	45
## 309	33	80	27
## 310	31	40	18
## 311	20	52	20
## 312	27	68	21
## 313	28	58	27
## 314	23	61	27
## 315	30	59	33
## 316	38	78	27
## 317	51	102	37
## 318	31	59	15
## 319	39	53	29
## 320	36	73	32
## 321	36	72	38
## 322	31	62	28
## 323	49	58	17
## 324	43	98	2
## 325	35	90	3
## 326	32	94	3
## 327	35	94	2
## 328	31	92	1
## 329	26	66	1
## 330	20	66	2
## 331	28	62	3
## 332	28	62	3
## 333	27	61	2
## 334	28	67	1
## 335	20	61	2
## 336	22	75	1
## 337	21	68	3
## 338	21	83	1
## 339	25	75	3
## 340	22	65	3
## 341	29	62	3
## 342	32	68	2
## 343	25	72	2
## 344	35	82	1
## 345	38	64	1
## 346	27	67	2
## 347	29	90	1
## 348	28	58	2
## 349	33	81	2
## 350	35	68	1
## 351	28	89	1
## 352	22	67	3
## 353	29	77	2
## 354	37	78	1
## 355	33	74	2
## 356	28	59	2
## 357	32	73	2
## 358	20	53	2
## 359	33	78	1
## 360	30	66	3

## 361	30	71	2
## 362	28	65	1
## 363	32	80	3
## 364	31	83	2
## 365	23	74	1
## 366	36	80	1
## 367	33	77	1
## 368	27	79	2
## 369	34	87	2
## 370	43	56	2
## 371	20	83	3
## 372	22	61	1
## 373	22	95	2
## 374	24	79	2
## 375	15	58	1
## 376	24	90	3
## 377	25	74	2
## 378	25	72	3
## 379	26	66	2
## 380	24	67	2
## 381	28	57	1
## 382	28	62	2
## 383	23	71	3
## 384	25	68	1
## 385	27	77	2
## 386	30	82	2
## 387	28	61	2
## 388	29	79	3
## 389	28	72	1
## 390	20	46	2
## 391	22	47	2
## 392	23	45	3
## 393	21	70	1
## 394	28	68	3
## 395	30	90	3
## 396	25	85	3
## 397	22	72	3
## 398	36	86	3
## 399	24	74	1
## 400	22	69	2
## 401	35	105	2
## 402	22	73	3
## 403	25	66	2
## 404	30	88	1
## 405	20	65	1
## 406	20	65	2
## 407	24	74	2
## 408	28	88	3
## 409	26	96	2
## 410	28	57	2
## 411	33	85	2
## 412	29	67	1
## 413	23	75	1
## 414	25	81	2

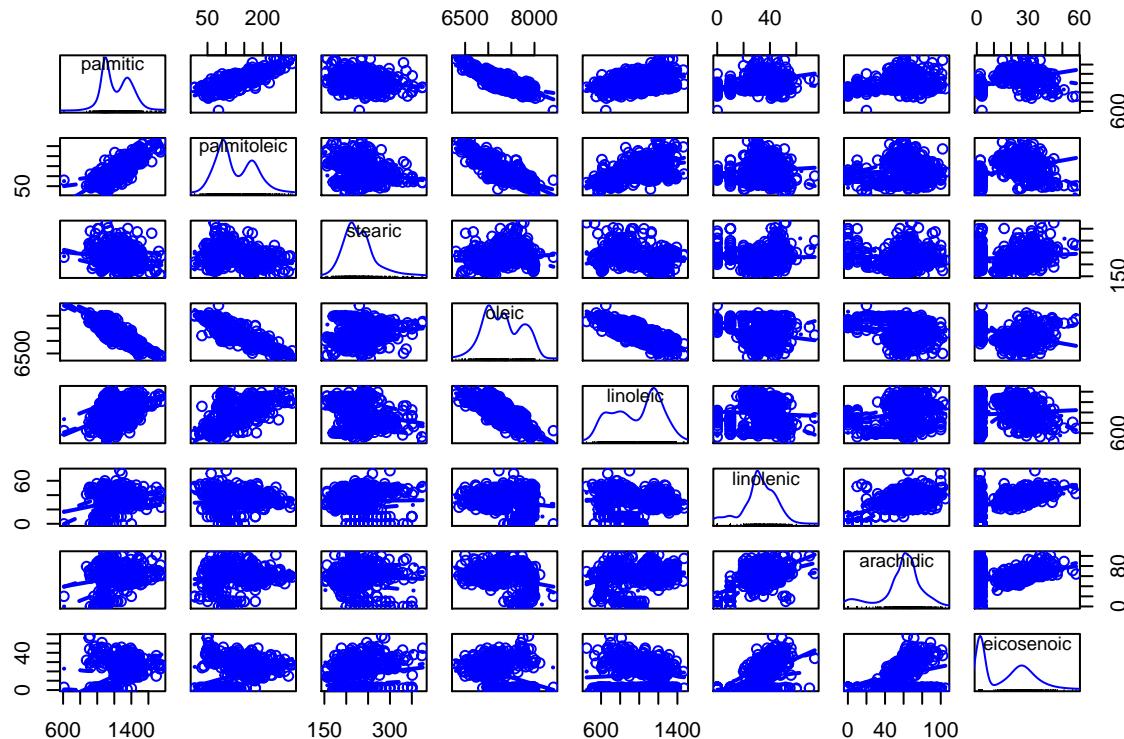
## 415	22	68	2
## 416	24	76	2
## 417	26	69	1
## 418	18	60	1
## 419	20	75	2
## 420	22	84	1
## 421	25	71	2
## 422	20	50	1
## 423	25	55	1
## 424	28	47	2
## 425	35	20	1
## 426	28	42	2
## 427	37	55	2
## 428	20	55	2
## 429	28	65	1
## 430	35	50	3
## 431	22	65	2
## 432	30	50	2
## 433	25	55	1
## 434	30	53	1
## 435	20	40	3
## 436	35	40	3
## 437	38	34	2
## 438	30	42	2
## 439	35	32	1
## 440	35	42	2
## 441	35	34	1
## 442	35	45	3
## 443	28	36	3
## 444	32	38	2
## 445	39	56	1
## 446	40	52	1
## 447	28	33	2
## 448	35	42	3
## 449	25	50	3
## 450	22	45	2
## 451	22	45	1
## 452	39	44	1
## 453	22	58	2
## 454	18	59	3
## 455	32	35	2
## 456	28	23	3
## 457	28	27	3
## 458	42	32	2
## 459	52	20	2
## 460	55	65	2
## 461	40	55	3
## 462	55	15	2
## 463	52	20	3
## 464	48	48	2
## 465	50	20	2
## 466	50	60	1
## 467	45	50	1
## 468	45	60	2

## 469	25	20	2
## 470	25	35	2
## 471	52	10	2
## 472	40	40	3
## 473	70	100	2
## 474	40	70	1
## 475	30	80	2
## 476	40	80	2
## 477	30	90	1
## 478	20	50	3
## 479	10	20	1
## 480	30	60	2
## 481	50	100	2
## 482	40	90	2
## 483	60	100	2
## 484	50	100	1
## 485	40	90	1
## 486	30	90	1
## 487	20	70	2
## 488	50	100	1
## 489	40	90	1
## 490	60	80	1
## 491	30	80	2
## 492	30	70	2
## 493	10	20	1
## 494	20	70	2
## 495	10	30	3
## 496	30	50	3
## 497	30	80	3
## 498	40	70	2
## 499	30	80	2
## 500	20	90	1
## 501	20	70	1
## 502	10	40	2
## 503	0	20	2
## 504	10	30	1
## 505	10	60	2
## 506	30	60	1
## 507	10	30	1
## 508	0	10	2
## 509	10	30	2
## 510	30	90	1
## 511	20	70	3
## 512	10	40	1
## 513	10	50	2
## 514	20	60	3
## 515	30	80	3
## 516	40	60	3
## 517	30	70	2
## 518	20	70	3
## 519	20	40	2
## 520	20	50	2
## 521	10	30	3
## 522	0	20	3

## 523	0	10	2
## 524	0	0	1
## 525	0	0	3
## 526	0	10	3
## 527	10	10	1
## 528	10	40	2
## 529	10	20	1
## 530	0	0	2
## 531	0	0	3
## 532	0	0	2
## 533	10	0	1
## 534	10	10	2
## 535	10	40	2
## 536	10	0	3
## 537	0	0	2
## 538	0	0	2
## 539	0	0	2
## 540	0	0	1
## 541	10	10	1
## 542	0	0	2
## 543	0	10	2
## 544	0	0	3
## 545	0	10	2
## 546	10	10	1
## 547	0	0	3
## 548	10	0	3
## 549	10	10	1
## 550	10	10	2
## 551	10	0	3
## 552	0	0	3
## 553	10	0	3
## 554	0	0	3
## 555	0	10	2
## 556	0	0	3
## 557	0	0	1
## 558	10	20	2
## 559	0	0	2
## 560	0	10	1
## 561	10	0	2
## 562	10	10	2
## 563	10	0	2
## 564	0	10	3
## 565	0	10	2
## 566	0	0	1
## 567	10	10	2
## 568	10	10	2
## 569	10	10	3
## 570	0	0	2
## 571	10	10	2
## 572	10	20	2

Dibuje un scatterplot de las ocho variables continuas. ¿Cuáles de los ácidos grasos están fuertemente asociados positivamente y cuáles fuertemente asociados negativamente?

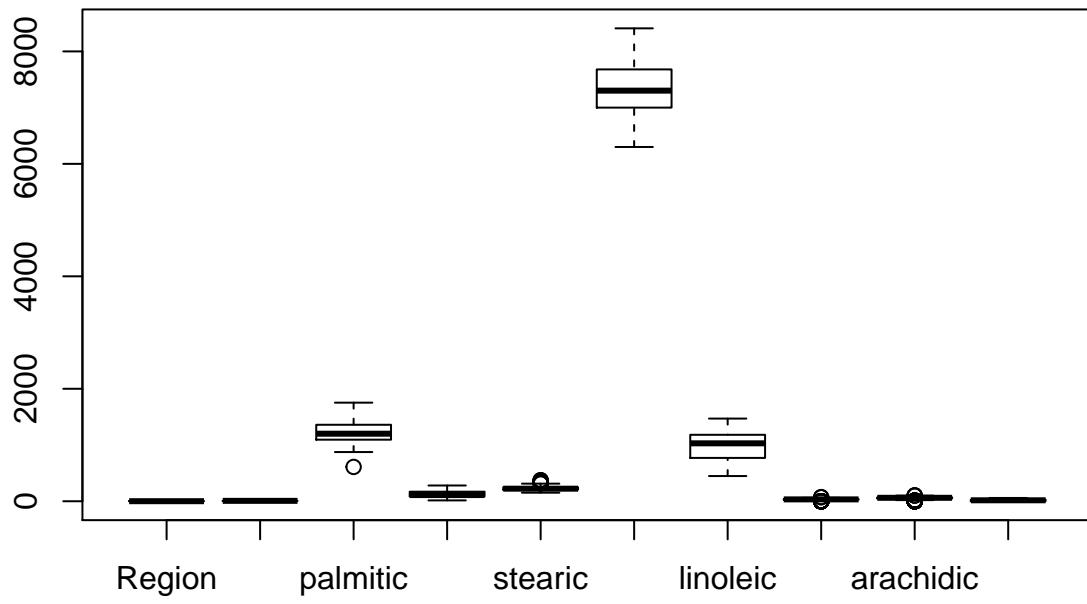
```
scatterplotMatrix(olives[,3:length(olives)])
```



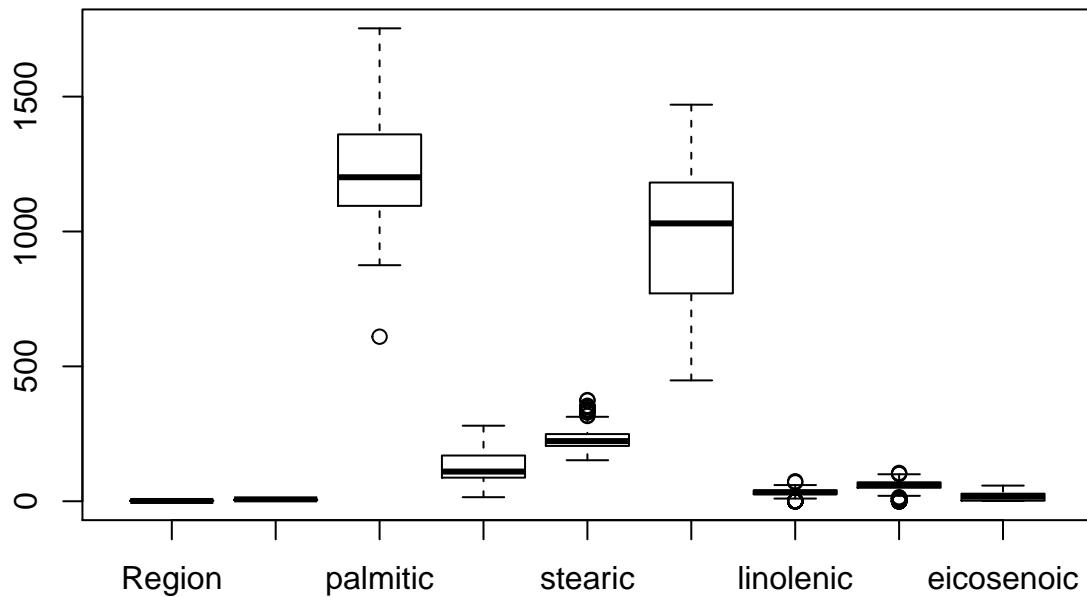
- Tienen una relación lineal fuerte positiva palmitic con palmitoleic, tienen una relación lineal fuerte negativa palmitic con oleic y oleic con linoleic.

¿Hay valores atípicos u otras características que valga la pena mencionar?

```
boxplot(olives)
```



```
boxplot(olives[-6])
```



Vemos que hay bastantes outliers en stearic, linoleic y arachidic.

El conjunto de datos se llama Lanza del paquete HSAUR2

```
library("HSAUR2")

## Loading required package: tools
Lanza

##   study treatment classification
## 1 Misoprostol 1
## 2 Misoprostol 1
## 3 Misoprostol 1
## 4 Misoprostol 1
## 5 Misoprostol 1
## 6 Misoprostol 1
## 7 Misoprostol 1
## 8 Misoprostol 1
## 9 Misoprostol 1
## 10 Misoprostol 1
## 11 Misoprostol 1
## 12 Misoprostol 1
## 13 Misoprostol 1
## 14 Misoprostol 1
## 15 Misoprostol 1
## 16 Misoprostol 1
## 17 Misoprostol 1
```

## 18	I Misoprostol	1
## 19	I Misoprostol	1
## 20	I Misoprostol	1
## 21	I Misoprostol	1
## 22	I Misoprostol	2
## 23	I Misoprostol	2
## 24	I Misoprostol	3
## 25	I Misoprostol	3
## 26	I Misoprostol	3
## 27	I Misoprostol	3
## 28	I Misoprostol	4
## 29	I Misoprostol	4
## 30	I Placebo	1
## 31	I Placebo	1
## 32	I Placebo	2
## 33	I Placebo	2
## 34	I Placebo	3
## 35	I Placebo	3
## 36	I Placebo	3
## 37	I Placebo	3
## 38	I Placebo	4
## 39	I Placebo	4
## 40	I Placebo	4
## 41	I Placebo	4
## 42	I Placebo	4
## 43	I Placebo	4
## 44	I Placebo	4
## 45	I Placebo	4
## 46	I Placebo	4
## 47	I Placebo	5
## 48	I Placebo	5
## 49	I Placebo	5
## 50	I Placebo	5
## 51	I Placebo	5
## 52	I Placebo	5
## 53	I Placebo	5
## 54	I Placebo	5
## 55	I Placebo	5
## 56	I Placebo	5
## 57	I Placebo	5
## 58	I Placebo	5
## 59	I Placebo	5
## 60	II Misoprostol	1
## 61	II Misoprostol	1
## 62	II Misoprostol	1
## 63	II Misoprostol	1
## 64	II Misoprostol	1
## 65	II Misoprostol	1
## 66	II Misoprostol	1
## 67	II Misoprostol	1
## 68	II Misoprostol	1
## 69	II Misoprostol	1
## 70	II Misoprostol	1
## 71	II Misoprostol	1

## 72	II Misoprostol	1
## 73	II Misoprostol	1
## 74	II Misoprostol	1
## 75	II Misoprostol	1
## 76	II Misoprostol	1
## 77	II Misoprostol	1
## 78	II Misoprostol	1
## 79	II Misoprostol	1
## 80	II Misoprostol	2
## 81	II Misoprostol	2
## 82	II Misoprostol	2
## 83	II Misoprostol	2
## 84	II Misoprostol	3
## 85	II Misoprostol	3
## 86	II Misoprostol	3
## 87	II Misoprostol	3
## 88	II Misoprostol	3
## 89	II Misoprostol	3
## 90	II Placebo	1
## 91	II Placebo	1
## 92	II Placebo	1
## 93	II Placebo	1
## 94	II Placebo	1
## 95	II Placebo	1
## 96	II Placebo	1
## 97	II Placebo	1
## 98	II Placebo	2
## 99	II Placebo	2
## 100	II Placebo	2
## 101	II Placebo	2
## 102	II Placebo	3
## 103	II Placebo	3
## 104	II Placebo	3
## 105	II Placebo	3
## 106	II Placebo	3
## 107	II Placebo	3
## 108	II Placebo	3
## 109	II Placebo	3
## 110	II Placebo	3
## 111	II Placebo	4
## 112	II Placebo	4
## 113	II Placebo	4
## 114	II Placebo	4
## 115	II Placebo	5
## 116	II Placebo	5
## 117	II Placebo	5
## 118	II Placebo	5
## 119	II Placebo	5
## 120	III Misoprostol	1
## 121	III Misoprostol	1
## 122	III Misoprostol	1
## 123	III Misoprostol	1
## 124	III Misoprostol	1
## 125	III Misoprostol	1

## 126	III Misoprostol	1
## 127	III Misoprostol	1
## 128	III Misoprostol	1
## 129	III Misoprostol	1
## 130	III Misoprostol	1
## 131	III Misoprostol	1
## 132	III Misoprostol	1
## 133	III Misoprostol	1
## 134	III Misoprostol	1
## 135	III Misoprostol	1
## 136	III Misoprostol	1
## 137	III Misoprostol	1
## 138	III Misoprostol	1
## 139	III Misoprostol	1
## 140	III Misoprostol	2
## 141	III Misoprostol	2
## 142	III Misoprostol	2
## 143	III Misoprostol	2
## 144	III Misoprostol	3
## 145	III Misoprostol	3
## 146	III Misoprostol	3
## 147	III Misoprostol	4
## 148	III Misoprostol	5
## 149	III Misoprostol	5
## 150	III Placebo	2
## 151	III Placebo	2
## 152	III Placebo	3
## 153	III Placebo	3
## 154	III Placebo	3
## 155	III Placebo	3
## 156	III Placebo	3
## 157	III Placebo	4
## 158	III Placebo	4
## 159	III Placebo	4
## 160	III Placebo	4
## 161	III Placebo	4
## 162	III Placebo	5
## 163	III Placebo	5
## 164	III Placebo	5
## 165	III Placebo	5
## 166	III Placebo	5
## 167	III Placebo	5
## 168	III Placebo	5
## 169	III Placebo	5
## 170	III Placebo	5
## 171	III Placebo	5
## 172	III Placebo	5
## 173	III Placebo	5
## 174	III Placebo	5
## 175	III Placebo	5
## 176	III Placebo	5
## 177	III Placebo	5
## 178	III Placebo	5
## 179	IV Misoprostol	1

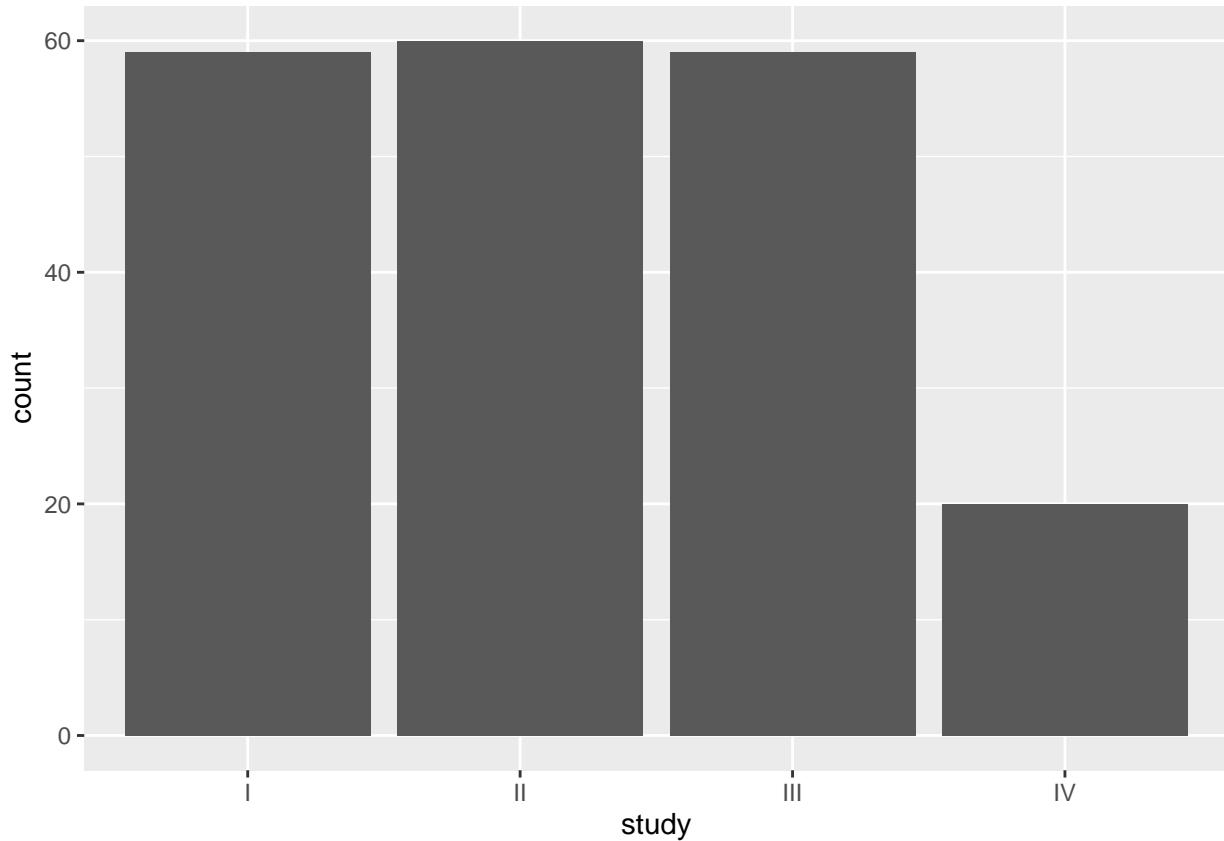
```

## 180    IV Misoprostol      2
## 181    IV Misoprostol      2
## 182    IV Misoprostol      2
## 183    IV Misoprostol      2
## 184    IV Misoprostol      3
## 185    IV Misoprostol      3
## 186    IV Misoprostol      3
## 187    IV Misoprostol      3
## 188    IV Misoprostol      3
## 189    IV Placebo          4
## 190    IV Placebo          4
## 191    IV Placebo          4
## 192    IV Placebo          4
## 193    IV Placebo          5
## 194    IV Placebo          5
## 195    IV Placebo          5
## 196    IV Placebo          5
## 197    IV Placebo          5
## 198    IV Placebo          5

```

Se informan los datos de cuatro estudios. Dibuje un diagrama para mostrar si los cuatro estudios son igualmente grandes.

```
ggplot(Lanza, aes(x=study)) + geom_bar()
```

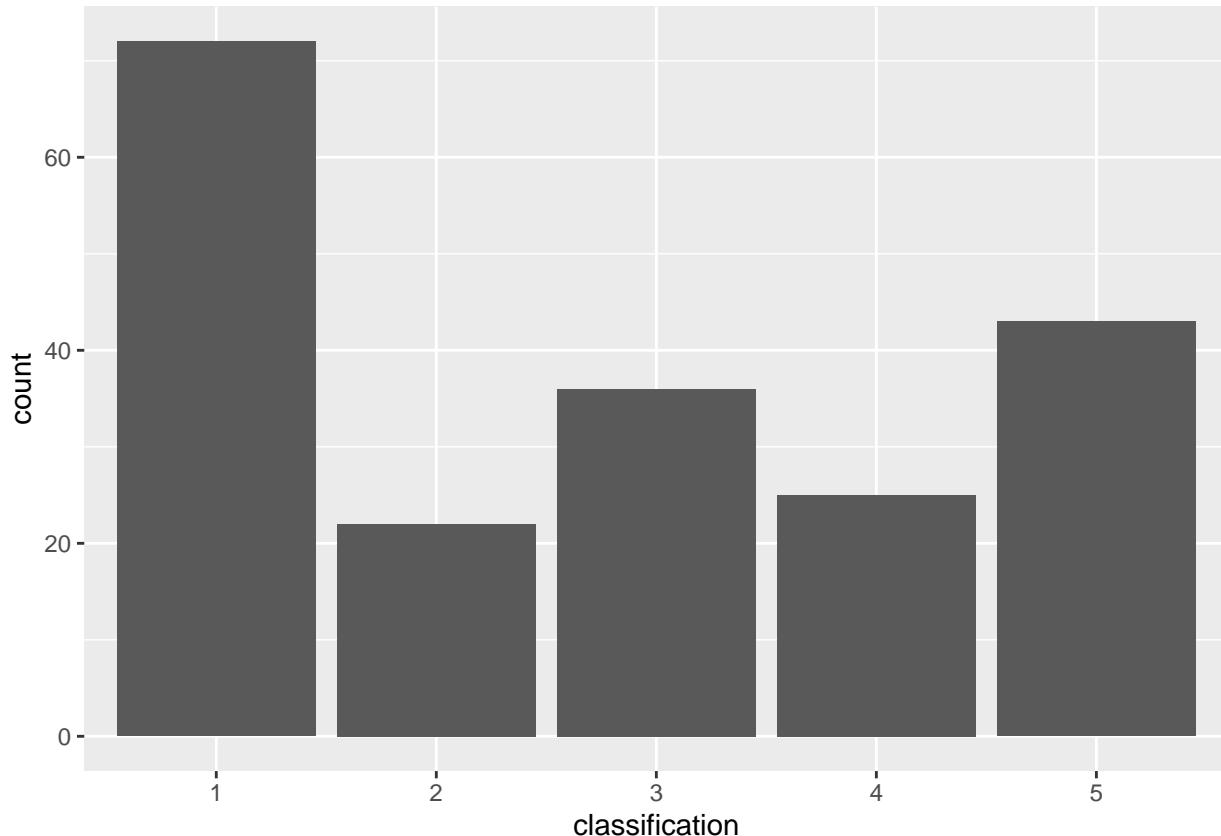


Tenemos un número bastante similar de muestras en los tres primeros estudios pero muchos menos en el

último.

El resultado se mide por la clasificación de la variable con puntuaciones de 1 (mejor) a 5 (peor). ¿Cómo describirías la distribución?

```
ggplot(Lanza, aes(classification)) + geom_bar()
```



Vemos que tenemos muchas más muestras con classification 1 que del resto (más de 70 muestras), viene seguida por las de classification 5 (más de 40 muestras), a continuación de 3 (35-36 muestras) y para finalizar 2 y 4 (entre 20-25). Dado el número de muestra es claro que existen demasiadas muestras de la clase 1 con respecto al resto, 2 y 4 son muy parecidas e incluso podríamos considerar que 3 está dentro de un margen razonable. Estamos ante un caso de clases desbalanceadas.

El paquete vcdExtra incluye datos de un viejo estudio de cáncer de mama sobre la supervivencia o muerte de 474 pacientes.

```
library("vcdExtra")  
  
## Loading required package: vcd  
## Loading required package: grid  
## Loading required package: gnm  
##  
## Attaching package: 'gnm'  
##  
## The following object is masked from 'package:HSAUR2':  
##
```

```

##      voting
##
## Attaching package: 'vcdExtra'
##
## The following object is masked from 'package:carData':
## 
##      Burt
##
## The following object is masked from 'package:dplyr':
## 
##      summarise

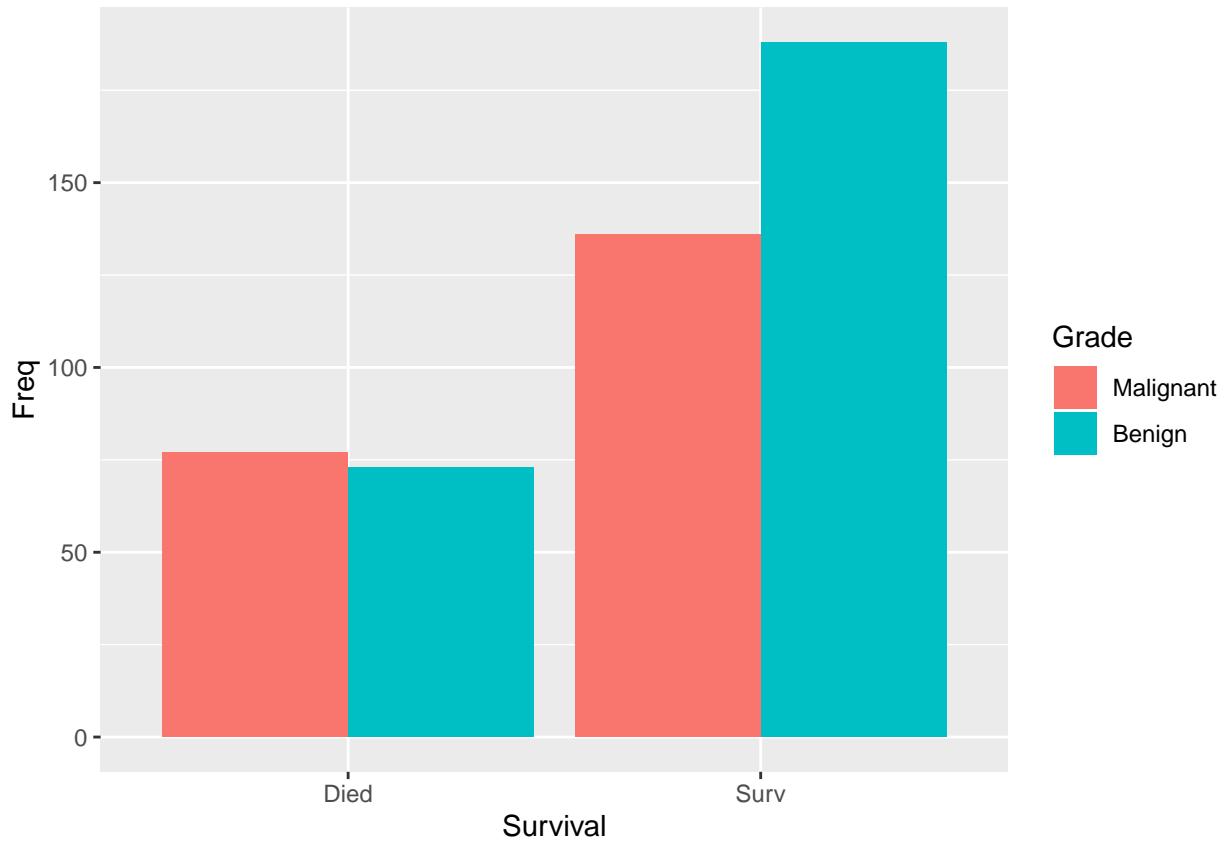
```

Convierta los datos en un data frame y dibuje gráficos para comparar las tasas de supervivencia, primero, por grado de malignidad y, en segundo lugar, por centro de diagnóstico.

```

df_cancer <- as.data.frame(Cancer)
df_cancer1 <- df_cancer %>% group_by(Survival, Grade) %>% summarize(Freq=sum(Freq))
df_cancer2 <- df_cancer %>% group_by(Survival, Center) %>% summarize(Freq=sum(Freq))
ggplot(df_cancer1, aes(x=Survival, y=Freq, group=Grade, fill=Grade)) +
  geom_bar(stat="identity", position="dodge")

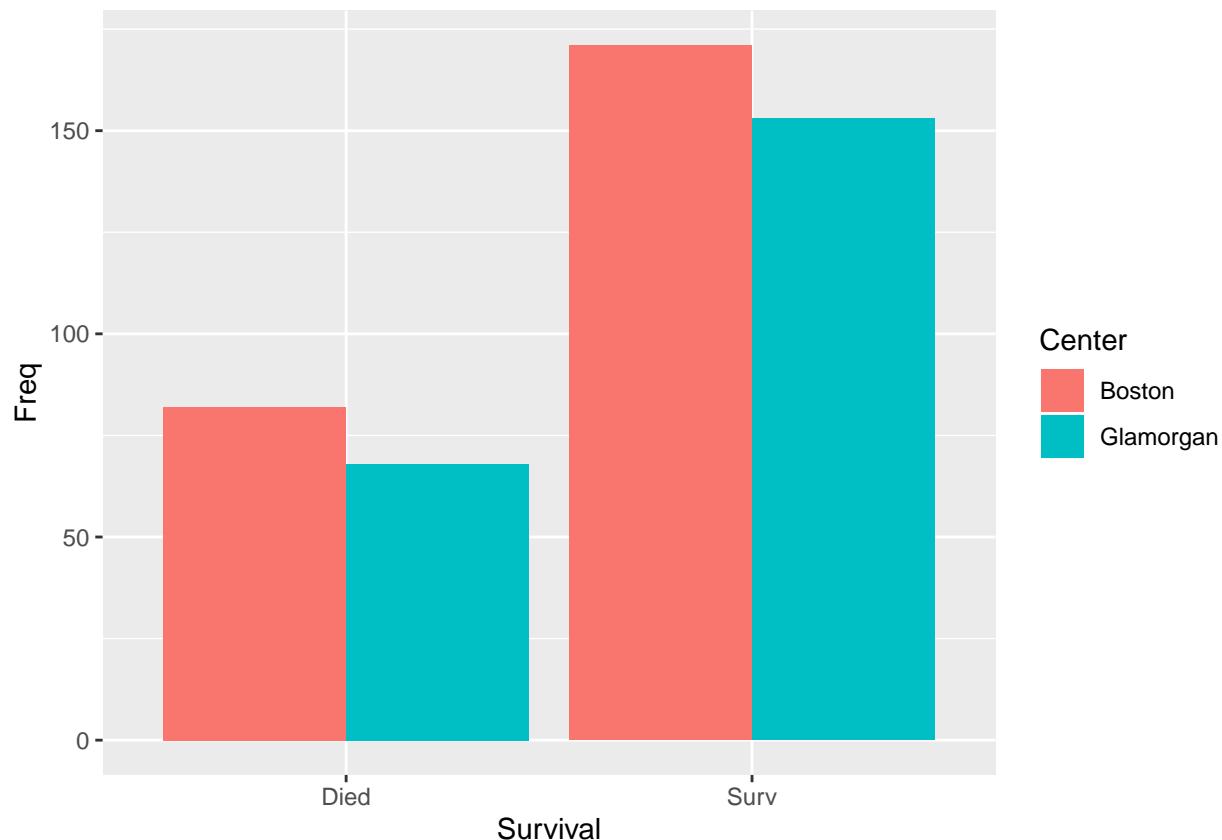
```



```

ggplot(df_cancer2, aes(x=Survival, y=Freq, group=Center, fill=Center)) +
  geom_bar(stat="identity", position="dodge")

```



¿Qué diagrama dibujaría para comparar las tasas de supervivencia tanto por grado de malignidad como por centro de diagnóstico? ¿Importa el orden de las variables explicativas?

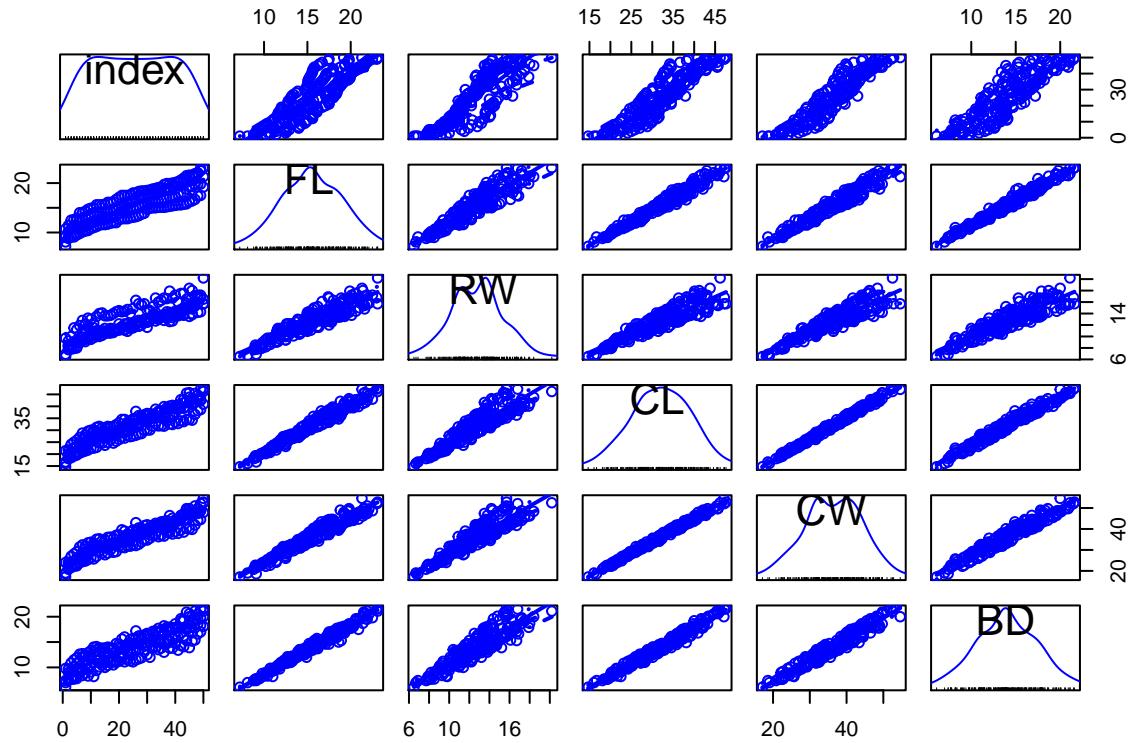
El diagrama presentado en el apartado anterior es suficientemente explicativo para comparar ambas variables. Lo importante del orden de esas variables es que no implique una mala legibilidad del gráfico.

Dataset crabs (del paquete MASS) [Venables y Ripley, 2002]. Los autores inicialmente se transforman a una escala logarítmica y luego escriben que:

“The data are very highly correlated and scatterplot matrices and brush plots [i.e. interactive graphics] are none too revealing.”.

```
library("MASS")
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##     select
```

```
scatterplotMatrix(crabs[,3:length(crabs)])
```



Utilizando gráficos generales, comente si la transformación logarítmica fue una buena idea y si está de acuerdo con su afirmación sobre las correlaciones.

La necesidad de una transformación logarítmica es un indicativo de que queremos reducir una “skewness” positiva, sin embargo, a vista de las curvas de densidad (presentes en la diagonal principal) no vemos ninguna que sea considerablemente asimétrica, dejando una cola más larga hacia la derecha (asimetría positiva).

Extra 2. Cómo crear subgrupos de datos en R

Busca información sobre la function `cut()`. Para ilustrar su uso vamos a utilizar el dataset `state.x77`. Si no lo tienes instalado instala el paquete `R-Datasets`. Usa la función `head()` para ver como son tus datos.

- Extrae la columna Frost y asigna el resultado a la variable `frost`

```
frost <- as.data.frame(state.x77)$Frost
```

- Tu Nuevo objeto es un vector numérico

```
class(frost)
```

```
## [1] "numeric"
```

- Ahora intenta agrupar los datos en `frost` en tres niveles. Para crear bins en tus datos puedes utilizar la función `cut()`.

```
cut(frost, 3)
```

```
## [1] [-0.188,62.7] (125,188]      (-0.188,62.7] (62.7,125]     (-0.188,62.7]
## [6] (125,188]      (125,188]      (62.7,125]     (-0.188,62.7] (-0.188,62.7]
## [11] (-0.188,62.7] (125,188]     (125,188]      (62.7,125]     (125,188]
## [16] (62.7,125]     (62.7,125]     (-0.188,62.7] (125,188]     (62.7,125]
## [21] (62.7,125]     (62.7,125]     (125,188]      (-0.188,62.7] (62.7,125]
## [26] (125,188]      (125,188]      (125,188]      (125,188]     (62.7,125]
## [31] (62.7,125]     (62.7,125]     (62.7,125]     (125,188]     (62.7,125]
## [36] (62.7,125]     (-0.188,62.7] (125,188]     (125,188]     (62.7,125]
## [41] (125,188]      (62.7,125]     (-0.188,62.7] (125,188]     (125,188]
## [46] (62.7,125]     (-0.188,62.7] (62.7,125]     (125,188]     (125,188]
## Levels: (-0.188,62.7] (62.7,125] (125,188]
```

- ¿Qué obtienes como nombres de los niveles? Obtengo como nombres el intervalo que representa el rango de valores para cada uno de los niveles.
- En la realidad no existen estados que tengan frost en días negativos. Esto es porque R añade un poco de padding. Prueba a solucionar el problema utilizando el parámetro include.lowest=TRUE en cut()

```
cut(frost, 3, include.lowest=TRUE)
```

```
## [1] [-0.188,62.7] (125,188]      [-0.188,62.7] (62.7,125]     [-0.188,62.7]
## [6] (125,188]      (125,188]      (62.7,125]     [-0.188,62.7] [-0.188,62.7]
## [11] [-0.188,62.7] (125,188]     (125,188]      (62.7,125]     (125,188]
## [16] (62.7,125]     (62.7,125]     [-0.188,62.7] (125,188]     (62.7,125]
## [21] (62.7,125]     (62.7,125]     (125,188]      [-0.188,62.7] (62.7,125]
## [26] (125,188]      (125,188]      (125,188]      (125,188]     (62.7,125]
## [31] (62.7,125]     (62.7,125]     (62.7,125]     (125,188]     (62.7,125]
## [36] (62.7,125]     [-0.188,62.7] (125,188]     (125,188]     (62.7,125]
## [41] (125,188]      (62.7,125]     [-0.188,62.7] (125,188]     (125,188]
## [46] (62.7,125]     [-0.188,62.7] (62.7,125]     (125,188]     (125,188]
## Levels: [-0.188,62.7] (62.7,125] (125,188]
```

No conseguimos arreglarlo con include.lowest. Sólo ha cambiado el intervalo de abierto a cerrado.

- Los nombres de los niveles no son demasiado informativos, especifica nuevos nombres para los niveles.

```
frost <- cut(frost, 3, include.lowest=TRUE, labels=c("Bajo", "Medio", "Alto"))
```

- Despues de este paso has creado un factor que clasifica los estados en bajo, medio y alto según el numero de heladas.

```
frost
```

```
## [1] Bajo Alto Bajo Medio Bajo Alto Alto Medio Bajo Bajo Bajo
## [12] Alto Alto Medio Alto Medio Medio Bajo Alto Medio Medio Medio
## [23] Alto Bajo Medio Alto Alto Alto Medio Medio Medio Medio Medio
## [34] Alto Medio Medio Bajo Alto Alto Medio Alto Medio Bajo Alto
## [45] Alto Medio Bajo Medio Alto Alto
## Levels: Bajo Medio Alto
```

- Ahora cuenta el número de estados que hay en cada uno de los niveles. PISTA: utiliza la función table()

```
table(frost)
```

```
## frost
## Bajo Medio Alto
##    11    19    20
```

Extra 3. Cómo ordenar tablas (dplyr)

```
library("dslabs")
library("tidyverse")

## -- Attaching packages ----- tidyverse 1.2.1 --
## v tibble  2.1.3     v purrr   0.3.3
## v tidyverse 1.0.0    v stringr  1.4.0
## v readr   1.3.1    v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter()      masks stats::filter()
## x dplyr::lag()         masks stats::lag()
## x car::recode()        masks dplyr::recode()
## x MASS::select()       masks dplyr::select()
## x purrr::some()        masks car::some()
## x vcdExtra::summarise() masks dplyr::summarise()
```

Utiizando el paquete dplyr realiza el siguiente análisis sobre el dataset murders. Para aquellos que no hicieron el curso de inicio a la programación para Ciencia de Datos he puesto en PRADO unas diapositivas explicando el funcionamiento.

```
murders
```

```
##           state abb      region population total
## 1          Alabama AL        South  4779736   135
## 2          Alaska AK        West   710231    19
## 3         Arizona AZ        West  6392017   232
## 4        Arkansas AR        South 2915918    93
## 5      California CA        West 37253956  1257
## 6        Colorado CO        West 5029196    65
## 7    Connecticut CT    Northeast 3574097   97
## 8        Delaware DE        South 897934    38
## 9 District of Columbia DC        South 601723    99
## 10       Florida FL        South 19687653  669
## 11       Georgia GA        South 9920000  376
## 12       Hawaii HI        West 1360301     7
## 13       Idaho ID        West 1567582   12
## 14      Illinois IL North Central 12830632  364
## 15       Indiana IN North Central 6483802  142
## 16       Iowa IA North Central 3046355   21
## 17       Kansas KS North Central 2853118   63
## 18      Kentucky KY        South 4339367  116
## 19      Louisiana LA        South 4533372  351
## 20       Maine ME    Northeast 1328361   11
## 21       Maryland MD        South 5773552  293
## 22 Massachusetts MA    Northeast 6547629  118
## 23       Michigan MI North Central 9883640  413
## 24      Minnesota MN North Central 5303925   53
## 25 Mississippi MS        South 2967297  120
## 26      Missouri MO North Central 5988927  321
## 27       Montana MT        West 989415    12
## 28      Nebraska NE North Central 1826341   32
## 29       Nevada NV        West 2700551   84
```

```

## 30      New Hampshire NH      Northeast    1316470    5
## 31      New Jersey   NJ      Northeast    8791894   246
## 32      New Mexico   NM          West    2059179    67
## 33      New York     NY      Northeast   19378102   517
## 34      North Carolina NC          South   9535483   286
## 35      North Dakota ND North Central  672591     4
## 36      Ohio          OH North Central 11536504   310
## 37      Oklahoma     OK          South   3751351   111
## 38      Oregon        OR          West    3831074    36
## 39      Pennsylvania PA      Northeast 12702379   457
## 40      Rhode Island RI      Northeast 1052567    16
## 41      South Carolina SC          South   4625364   207
## 42      South Dakota SD North Central 814180     8
## 43      Tennessee    TN          South   6346105   219
## 44      Texas         TX          South  25145561   805
## 45      Utah          UT          West    2763885    22
## 46      Vermont       VT      Northeast 625741     2
## 47      Virginia     VA          South   8001024   250
## 48      Washington   WA          West    6724540    93
## 49      West Virginia WV          South   1852994    27
## 50      Wisconsin    WI North Central 5686986   97
## 51      Wyoming      WY          West    563626     5

```

- Averigua cual es el estado con la mayor población

```
murders %>% filter(population==max(population))
```

```

##           state abb region population total
## 1 California CA    West   37253956 1257

```

- Averigua cual es el estado con el mayor ratio de asesinatos

```
murders %>% mutate(ratio=total/population) %>%
  filter(ratio==max(ratio))
```

```

##           state abb region population total      ratio
## 1 District of Columbia DC    South   601723    99 0.0001645275

```

- Averigua cual es el estado con el menor ratio de asesinatos

```
murders %>% mutate(ratio=total/population) %>%
  filter(ratio==min(ratio))
```

```

##           state abb region population total      ratio
## 1 Vermont   VT Northeast 625741     2 3.196211e-06

```

- Crea una tabla ordenada alfabeticamente por region y por el ratio de asesinatos

```
murders %>% mutate(ratio=total/population) %>%
  arrange(region, ratio)
```

```

##           state abb region population total      ratio
## 1 Vermont   VT Northeast 625741     2 3.196211e-06
## 2 New Hampshire NH Northeast 1316470    5 3.798036e-06
## 3 Maine     ME Northeast 1328361   11 8.280881e-06
## 4 Rhode Island RI Northeast 1052567   16 1.520093e-05
## 5 Massachusetts MA Northeast 6547629  118 1.802179e-05
## 6 New York   NY Northeast 19378102  517 2.667960e-05
## 7 Connecticut CT Northeast 3574097   97 2.713972e-05

```

```

## 8      New Jersey NJ      Northeast 8791894 246 2.798032e-05
## 9      Pennsylvania PA      Northeast 12702379 457 3.597751e-05
## 10     West Virginia WV      South 1852994 27 1.457101e-05
## 11     Kentucky KY      South 4339367 116 2.673201e-05
## 12     Alabama AL      South 4779736 135 2.824424e-05
## 13     Oklahoma OK      South 3751351 111 2.958934e-05
## 14     North Carolina NC      South 9535483 286 2.999324e-05
## 15     Virginia VA      South 8001024 250 3.124600e-05
## 16     Arkansas AR      South 2915918 93 3.189390e-05
## 17     Texas TX      South 25145561 805 3.201360e-05
## 18     Florida FL      South 19687653 669 3.398069e-05
## 19     Tennessee TN      South 6346105 219 3.450936e-05
## 20     Georgia GA      South 9920000 376 3.790323e-05
## 21     Mississippi MS      South 2967297 120 4.044085e-05
## 22     Delaware DE      South 897934 38 4.231937e-05
## 23     South Carolina SC      South 4625364 207 4.475323e-05
## 24     Maryland MD      South 5773552 293 5.074866e-05
## 25     Louisiana LA      South 4533372 351 7.742581e-05
## 26 District of Columbia DC      South 601723 99 1.645275e-04
## 27     North Dakota ND North Central 672591 4 5.947151e-06
## 28     Iowa IA North Central 3046355 21 6.893484e-06
## 29     South Dakota SD North Central 814180 8 9.825837e-06
## 30     Minnesota MN North Central 5303925 53 9.992600e-06
## 31     Wisconsin WI North Central 5686986 97 1.705649e-05
## 32     Nebraska NE North Central 1826341 32 1.752137e-05
## 33     Indiana IN North Central 6483802 142 2.190073e-05
## 34     Kansas KS North Central 2853118 63 2.208111e-05
## 35     Ohio OH North Central 11536504 310 2.687123e-05
## 36     Illinois IL North Central 12830632 364 2.836961e-05
## 37     Michigan MI North Central 9883640 413 4.178622e-05
## 38     Missouri MO North Central 5988927 321 5.359892e-05
## 39     Hawaii HI      West 1360301 7 5.145920e-06
## 40     Idaho ID      West 1567582 12 7.655102e-06
## 41     Utah UT      West 2763885 22 7.959810e-06
## 42     Wyoming WY      West 563626 5 8.871131e-06
## 43     Oregon OR      West 3831074 36 9.396843e-06
## 44     Montana MT      West 989415 12 1.212838e-05
## 45     Colorado CO      West 5029196 65 1.292453e-05
## 46     Washington WA      West 6724540 93 1.382994e-05
## 47     Alaska AK      West 710231 19 2.675186e-05
## 48     Nevada NV      West 2700551 84 3.110476e-05
## 49     New Mexico NM      West 2059179 67 3.253724e-05
## 50     California CA      West 37253956 1257 3.374138e-05
## 51     Arizona AZ      West 6392017 232 3.629527e-05

```

- Crea una tabla con la media del ratio de asesinatos por region

```

murders %>% mutate(ratio=total/population) %>%
  group_by(region) %>%
  summarize(mean(ratio))

```

```

## # A tibble: 4 x 2
##   region      `mean(ratio)`
##   <fct>          <dbl>
## 1 Northeast 0.0000185

```

```
## 2 South      0.0000442
## 3 North Central 0.0000218
## 4 West       0.0000183
```