```
Warning message in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
“collapsing to unique 'x' values”
Warning message in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
“collapsing to unique 'x' values”
Warning message in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
“collapsing to unique 'x' values”
Warning message in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
“collapsing to unique 'x' values”
Warning message in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
“collapsing to unique 'x' values”
Warning message in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
“collapsing to unique 'x' values”
```

Box-Percentile Plot

Práctica 3 - Actividad de aplicación

Ejemplo 3: El archivo DirectMarketing.csv contiene datos de un vendedor de marketing directo el cual vende sus productos solo a través de correos electrónicos directos. El vendedor envía catálogos a los clientes con las caractearísticas de los productos y estos ordenan directamente de los catálogos. El responsable de marketing ha desarrollado registros de clientes para aprender qué hace que algunos clientes gasten más que otros. El conjunto de datos incluye n=1000 clientes y las siguientes variables: Age (Edad del cliente; Adulta/Media/Joven); Gender (Género del cliente; masculino/femenino); OwnHome (Own/Rent); Married (soltero/casado); Location (lejos/cerca, en términos de distancia al local de una tienda que vende productos similares); Salary (sueldo anual de los clientes, en dólares); Children (número de hijos 0-3); History (del volumen de compra anterior, baja/media/alta/NA, NA significa que este cliente aún no ha adquirido ningún producto); Catalogs (número de catálogos enviados); y AmountSpent (gasto en dólares). El objetivo es explicar la variable AmountSpent en términos de las caractarísticas de los clientes.

```
#Abrimos el archivo de datos indicando que las observaciones están separados por ","
#Aplicamos stringAsFactor
dDirM<-read.csv("https://raw.githubusercontent.com/Emax1/R/master/DirectMarketing.csv",sep
dDirM
#Se tienen 10 variables con 1000 observaciones por cada una de ellas
```

A data.frame: 1000 × 10

Age	Gender	OwnHome	Married	Location	Salary	Children	History	Catalogs	Amou
<fct>	<fct>	<fct>	<fct>	<fct>	<int>	<int>	<fct>	<int>	
Old	Female	Own	Single	Far	47500	0	High	6	
Middle	Male	Rent	Single	Close	63600	0	High	6	
Young	Female	Rent	Single	Close	13500	0	Low	18	
Middle	Male	Own	Married	Close	85600	1	High	18	
Middle	Female	Own	Single	Close	68400	0	High	12	
Young	Male	Own	Married	Close	30400	0	Low	6	
Middle	Female	Rent	Single	Close	48100	0	Medium	12	
Middle	Male	Own	Single	Close	68400	0	High	18	
Middle	Female	Own	Married	Close	51900	3	Low	6	
Old	Male	Own	Married	Far	80700	0	NA	18	
Young	Male	Rent	Married	Close	43700	1	NA	12	
Middle	Male	Own	Married	Far	111800	3	High	18	
Middle	Female	Own	Married	Close	44100	1	Medium	24	
Middle	Male	Own	Married	Close	111400	0	High	12	
Old	Female	Own	Married	Far	110000	0	High	24	
Middle	Female	Own	Married	Far	83100	1	NA	12	
Old	Female	Own	Married	Close	82800	0	High	24	
Middle	Male	Own	Married	Close	90100	0	High	18	
Middle	Male	Rent	Single	Close	38600	1	Low	18	
Old	Female	Rent	Single	Far	14000	0	Low	12	
Young	Female	Rent	Single	Far	14300	2	Low	6	
Young	Male	Rent	Married	Close	49800	0	Medium	12	
Old	Female	Rent	Married	Far	60200	0	High	18	
Young	Male	Rent	Married	Far	43500	1	Medium	12	
Young	Female	Rent	Married	Close	41700	1	NA	6	
Middle	Female	Own	Married	Close	92300	3	Low	18	
Middle	Male	Own	Single	Close	62800	3	NA	6	
Middle	Female	Rent	Married	Far	72900	0	High	18	
Middle	Female	Rent	Married	Close	66900	2	Medium	12	
Young	Male	Rent	Single	Close	21700	1	NA	18	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Young	Female	Rent	Single	Close	25300	3	NA	6
Old	Female	Own	Married	Close	43600	0	Medium	12
Middle	Male	Own	Married	Close	83600	2	Medium	6
Middle	Female	Own	Single	Far	56200	0	High	18
Young	Male	Rent	Married	Close	45500	0	Medium	6
Middle	Male	Own	Married	Close	99300	1	High	12
Middle	Male	Own	Married	Far	108000	3	High	24
Middle	Female	Own	Married	Far	84900	1	High	24
Middle	Male	Own	Married	Far	93000	3	High	6
Middle	Female	Own	Married	Close	82000	0	High	18
Old	Female	Own	Single	Close	33000	3	NA	6
Middle	Female	Own	Single	Close	36900	2	Low	24
Middle	Female	Rent	Single	Close	33200	1	NA	6
Young	Male	Rent	Married	Close	49300	0	NA	24
Old	Male	Rent	Married	Far	66200	0	High	24
Middle	Male	Own	Married	Far	96800	2	High	18
Old	Male	Own	Single	Close	63200	0	High	24
Old	Male	Own	Married	Far	112900	0	High	24
Middle	Female	Rent	Single	Close	32100	2	Low	24
Middle	Male	Own	Married	Far	102700	1	High	18

```
#Las variables observadas son:
names(dDirM)
```

'Age' · 'Gender' · 'OwnHome' · 'Married' · 'Location' · 'Salary' · 'Children' · 'History' · 'Catalogs' · 'AmountSpent'

```
#Estructura de la matriz
str(dDirM)
#Según la estructura de la matríz se tienen variables categóricas de 3 y 2 niveles además
```

```
'data.frame':   1000 obs. of  10 variables:
 $ Age        : Factor w/ 3 levels "Middle","Old",..: 2 1 3 1 1 3 1 1 1 2 ...
 $ Gender     : Factor w/ 2 levels "Female","Male": 1 2 1 2 1 2 1 2 1 2 ...
 $ OwnHome    : Factor w/ 2 levels "Own","Rent": 1 2 2 1 1 1 2 1 1 1 ...
 $ Married    : Factor w/ 2 levels "Married","Single": 2 2 2 1 2 1 2 2 1 1 ...
 $ Location   : Factor w/ 2 levels "Close","Far": 2 1 1 1 1 1 1 1 1 2 ...
 $ Salary     : int  47500 63600 13500 85600 68400 30400 48100 68400 51900 80700 ...
 $ Children   : int  0 0 0 1 0 0 0 0 3 0 ...
 $ History    : Factor w/ 3 levels "High","Low","Medium": 1 1 2 1 1 2 3 1 2 NA ...
 $ Catalogs   : int  6 6 18 18 12 6 12 18 6 18 ...
 $ AmountSpent: int  755 1318 296 2436 1304 495 782 1155 158 3034 ...
```

```
#Hacemos un resumen de los datos observados
```

```
summary(dDirM)
#Encontramos 5 variables categóricas de las cuales se muestra la media por catetoría
#De las demás se observan los estadísticos descriptivos de variables continuas
```

```
      Age         Gender    OwnHome       Married     Location
 Middle:508   Female:506   Own :516   Married:502   Close:710
 Old   :205   Male  :494   Rent:484   Single :498   Far  :290
 Young :287
```

```
     Salary          Children       History       Catalogs      AmountSpent
 Min.   : 10100   Min.   :0.000   High  :255   Min.   : 6.00   Min.   :  38.0
 1st Qu.: 29975   1st Qu.:0.000   Low   :230   1st Qu.: 6.00   1st Qu.: 488.2
 Median : 53700   Median :1.000   Medium:212   Median :12.00   Median : 962.0
 Mean   : 56104   Mean   :0.934   NA's  :303   Mean   :14.68   Mean   :1216.8
 3rd Qu.: 77025   3rd Qu.:2.000                3rd Qu.:18.00   3rd Qu.:1688.5
 Max.   :168800   Max.   :3.000                Max.   :24.00   Max.   :6217.0
```

```
#Se tienen 3 categorías de edad "Age" y se espera aplicar la media (mean) a cada uno de el
#Analizamos la categoría Age
#Tabla de dDirM$Age
table(dDirM$Age)
table(dDirM$Gender)
table(dDirM$OwnHome)
table(dDirM$Married)
table(dDirM$Location)
table(dDirM$History)
table(dDirM$Children)
#Respecto a la edad se puede observar que la mitad de los clientes corresponden a la categ
#Se observa además que gran parte de los clientes son de localidades cercanas;
#Prácticamente la mitad son clientes sin hijos y a medida que aumentan el número de hijos
#Respecto a las demás categorías encontramos prácticamente proporciones similares.
```

```
Middle    Old  Young
   508    205    287

Female   Male
   506    494

 Own Rent
 516  484

Married  Single
    502     498

Close   Far
  710   290

  High    Low Medium
   255    230    212

  0   1   2   3
462 267 146 125
```

```
#La media general de la variable de interés o variable respuesta AmountSpent
m<-mean(dDirM$AmountSpent)
```

```
m
```

```
1216.77
```

```
#Obtendremos la media y desviación estándar de la variable AmountSpent por categoría
#X - es el vector al cual se le buscan los estadísticos
#INDEX - la variable por la que se agrupan estos
#FUN - la  función que se utiliza para aplicar
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Age,FUN=mean)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Age,FUN=sd)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Gender,FUN=mean)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Gender,FUN=sd)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$OwnHome,FUN=mean)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$OwnHome,FUN=sd)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Married,FUN=mean)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Married,FUN=sd)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Location,FUN=mean)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Location,FUN=sd)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Children,FUN=mean)
tapply(X=dDirM$AmountSpent,INDEX=dDirM$Children,FUN=sd)
#Según estos datos las medias más altas de consumo corresponden a:
#Categoría de edades intermedias"Middle" seguidos muy de cerca por los jóvenes; Género más
#Al aumentnar el número de hijos se reduce la media de consumo
```

```
Middle:  1501.69094488189 Old:  1432.12682926829 Young:  558.623693379791
Middle:  959.273124800847 Old:  1045.93002686145 Young:  450.145948977348
Female:  1025.33992094862 Male:  1412.85020242915
Female:  910.529013389988 Male:  972.76995076405
Own:  1543.13565891473 Rent:  868.826446280992
Own:  1030.25047619346 Rent:  737.270349437999
Married:  1672.06972111554 Single:  757.813253012048
Married:  1037.68085748728 Single:  592.200791904786
Close:  1061.68591549296 Far:  1596.45862068966
Close:  808.456432896427 Far:  1177.12023828293
0:  1406.64069264069 1:  1219.5393258427 2:  940.746575342466 3:  831.488
0:  986.843367294233 1:  971.294048724052 2:  822.307064147458 3:
```

```
#Llamamos la librería (Hmisc) contiene funciones para análisis descriptivo
#library(Hmisc)
#Buscamos en ella la función "describe" para el análisis descriptivo de la variable Amount
describe(dDirM$AmountSpent)
```

```
dDirM$AmountSpent
       n  missing distinct     Info     Mean      Gmd      .05      .10
    1000        0      852        1     1217     1014    187.9    242.9
     .25      .50      .75      .90      .95
   488.2    962.0   1688.5   2551.4   3052.5

lowest :   38   43   47   62   65, highest: 5503 5564 5830 5878 6217
```

```
#Usaremos el comando "summary" para visualizar la información básica de algunas variables
dDirM[,c("AmountSpent","Age","Location","Children")]
```

A data.frame: 1000 × 4

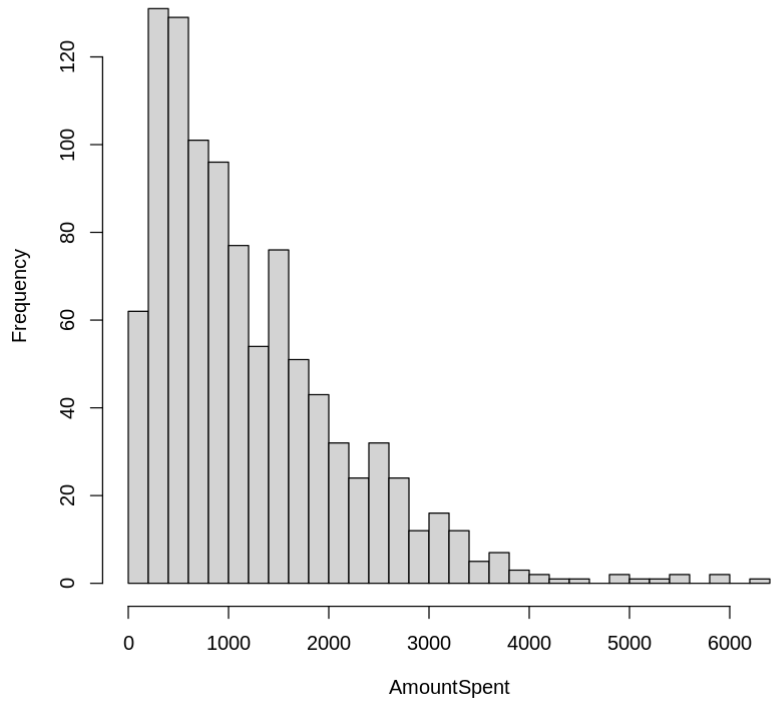
AmountSpent	Age	Location	Children
<int>	<fct>	<fct>	<int>
755	Old	Far	0
1318	Middle	Close	0
296	Young	Close	0
2436	Middle	Close	1
1304	Middle	Close	0
495	Young	Close	0
782	Middle	Close	0
1155	Middle	Close	0
158	Middle	Close	3
3034	Old	Far	0
927	Young	Close	1
2065	Middle	Far	3
704	Middle	Close	1
2136	Middle	Close	0
5564	Old	Far	0
2766	Middle	Far	1
3010	Old	Close	0
1956	Middle	Close	0
542	Middle	Close	1
410	Old	Far	0
194	Young	Far	2
827	Young	Close	0
2328	Old	Far	0
808	Young	Far	1
521	Young	Close	1
655	Middle	Close	3
707	Middle	Close	3
2075	Middle	Far	0
769	Middle	Close	2
774	Young	Close	1
⋮	⋮	⋮	⋮

510	Young	Close	3
1047	Old	Close	0
538	Middle	Close	2
1546	Middle	Far	0
754	Young	Close	0
1504	Middle	Close	1
1849	Middle	Far	3
3072	Middle	Far	1
1046	Middle	Far	3
1800	Middle	Close	0
700	Old	Close	3
514	Middle	Close	2
320	Middle	Close	1
1692	Young	Close	0
2057	Old	Far	0
2299	Middle	Far	2
1507	Old	Close	0
6217	Old	Far	0
405	Middle	Close	2
3785	Middle	Far	1
3537	Middle	Close	2
540	Old	Far	0
941	Middle	Close	1
5503	Middle	Far	0
273	Young	Close	0
384	Young	Close	1
1073	Middle	Far	1

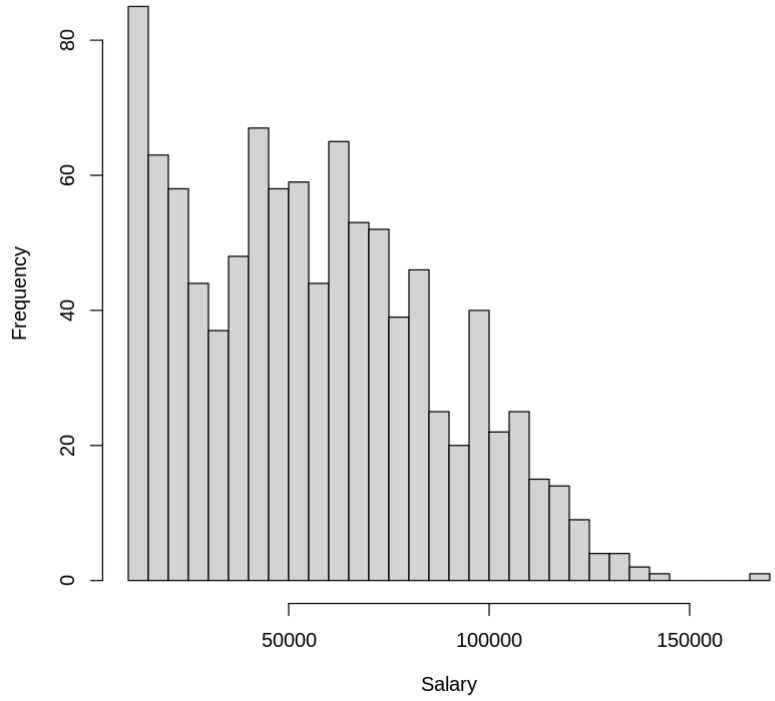
```
#Vemos un resumen de la base de datos obtenida
summary(dDirM[,c("Age","Location","Children","Salary","AmountSpent")])
```

```
      Age       Location      Children         Salary        AmountSpent   
 Middle:508   Close:710   Min.   :0.000   Min.   : 10100   Min.   :  38.0  
 Old   :205   Far  :290   1st Qu.:0.000   1st Qu.: 29975   1st Qu.: 488.2  
 Young :287               Median :1.000   Median : 53700   Median : 962.0  
                          Mean   :0.934   Mean   : 56104   Mean   :1216.8  
                          3rd Qu.:2.000   3rd Qu.: 77025   3rd Qu.:1688.5  
                          Max.   :3.000   Max.   :168800   Max.   :6217.0  
```

```
#Examinamos la variable "AmountSpent"
hist(dDirM["AmountSpent"]); hist(dDirM["Salary"])
```

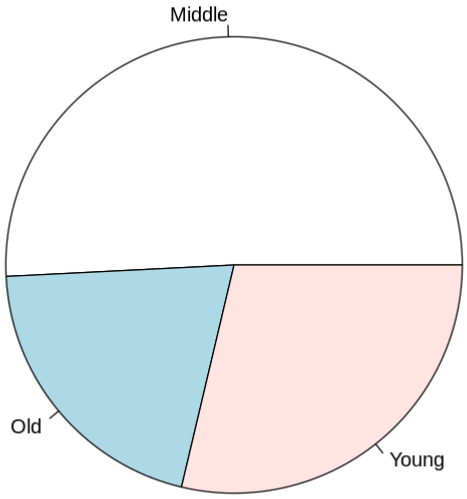


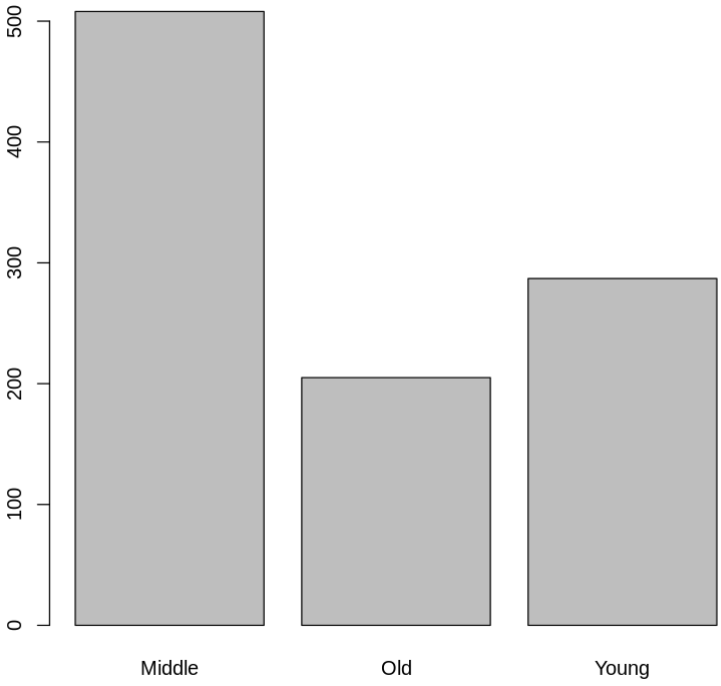
n:1000 m:0

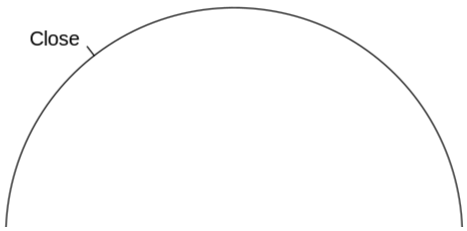


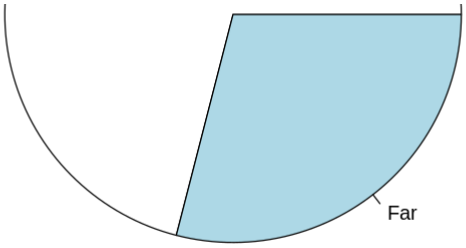
n:1000 m:0

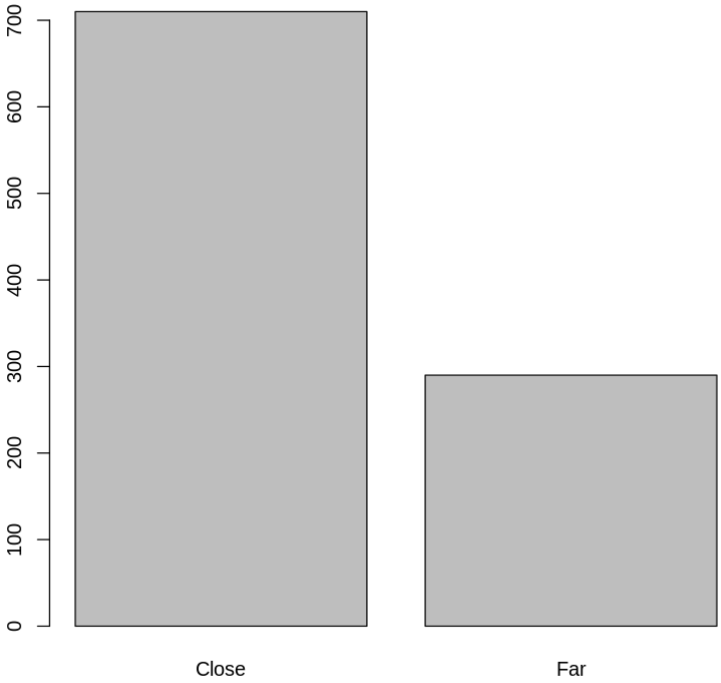
```
#Gráfica de las tablas de datos de las variables; Gráfico de pastel y de barras
pie(table(dDirM["Age"])); barplot(table(dDirM["Age"]))
pie(table(dDirM["Location"])); barplot(table(dDirM["Location"]))
pie(table(dDirM["Children"])); barplot(table(dDirM["Children"]))
```

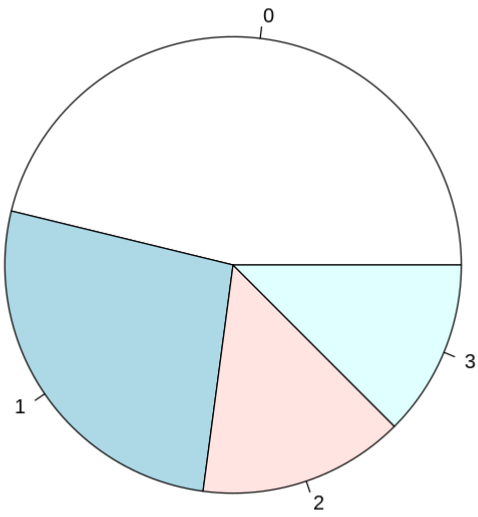

Middle
Old
Young

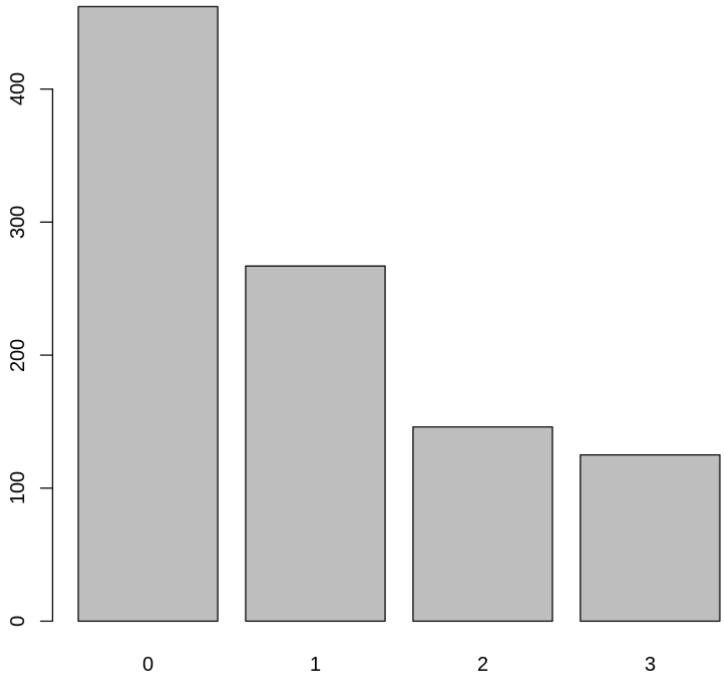

0
100
200
300
400
500
Middle
Old
Young


Close


Far


0
100
200
300
400
500
600
700
Close
Far


0
1
2
3



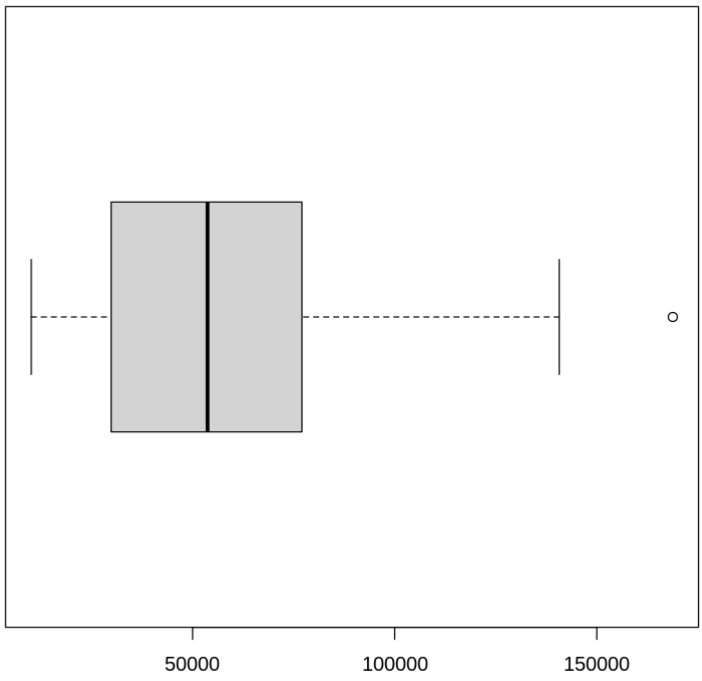

```
#Instalamos la librería "mice"
#install.packages("mice")
#Llamamos la librería (mice)
#library(mice)
#Aplicamos la función
md.pattern(dDirM[,1:10])
#La Fnción "md.pattern" nos permite visualizar los valores perdidos de una variable
```

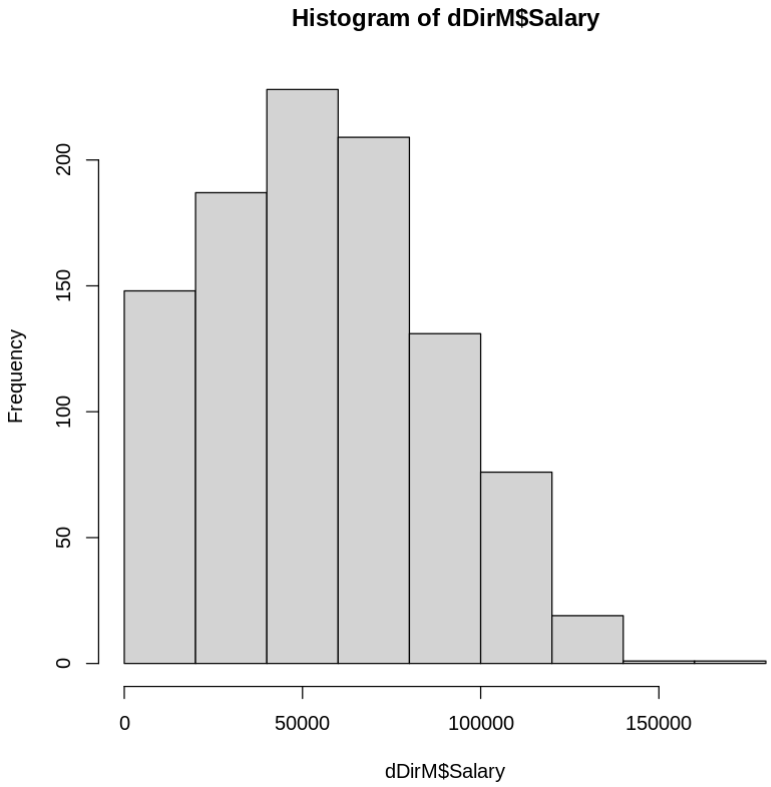
A matrix: 3 × 11 of type dbl

	Age	Gender	OwnHome	Married	Location	Salary	Children	Catalogs	AmountSpent
697	1	1	1	1	1	1	1	1	1
303	1	1	1	1	1	1	1	1	1
	0	0	0	0	0	0	0	0	0

```
#Hacemos un gráfico de cajas y bigotes
boxplot(dDirM$AmountSpent,horizontal=TRUE)
hist(dDirM$AmountSpent)
```

```
boxplot(dDirM$Salary,horizontal=TRUE)
hist(dDirM$Salary)
```





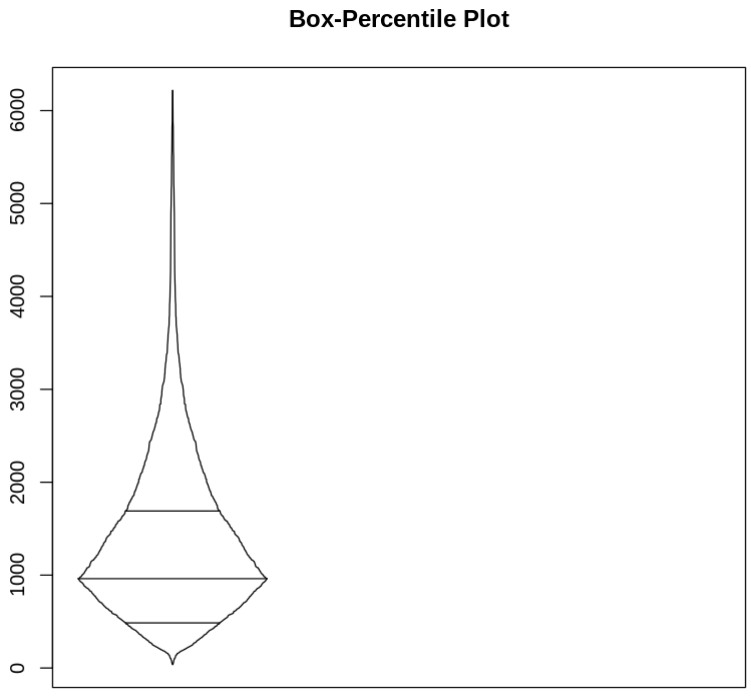

```
#Examinamos una nueva librería para gráficos
#library(Hmisc)
h3<-dDirM$AmountSpent
hn<-h3[dDirM$AmountSpent=="No"]
```

```
hy<-h3[dDirM$AmountSpent=="Yes"]
ds<-list(h3,hn,hy)
bpplot(ds)
```

```
Warning message in FUN(X[[i]], ...):
“no non-missing arguments to max; returning -Inf”
Warning message in FUN(X[[i]], ...):
“no non-missing arguments to max; returning -Inf”
Warning message in FUN(X[[i]], ...):
“no non-missing arguments to min; returning Inf”
Warning message in FUN(X[[i]], ...):
“no non-missing arguments to min; returning Inf”
Warning message in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
“collapsing to unique 'x' values”
Warning message in regularize.values(x, y, ties, missing(ties), na.rm = na.rm):
“collapsing to unique 'x' values”
Error in seq.default(delta, 1 - delta, delta): wrong sign in 'by' argument
Traceback:

1. bpplot(ds)
2. bpx(all.x[[i]], centers[i])
3. seq(delta, 1 - delta, delta)
4. seq.default(delta, 1 - delta, delta)
5. stop("wrong sign in 'by' argument")
```

SEARCH STACK OVERFLOW



Nueva sección

```
install.packages('arules')
```