

CURSO ANALISIS ESTADISTICO CON R

Caso estudio

Cargar la base de datos del curso “datos.curso1.RData”

```
load("datos/datos.curso1.RData")
```

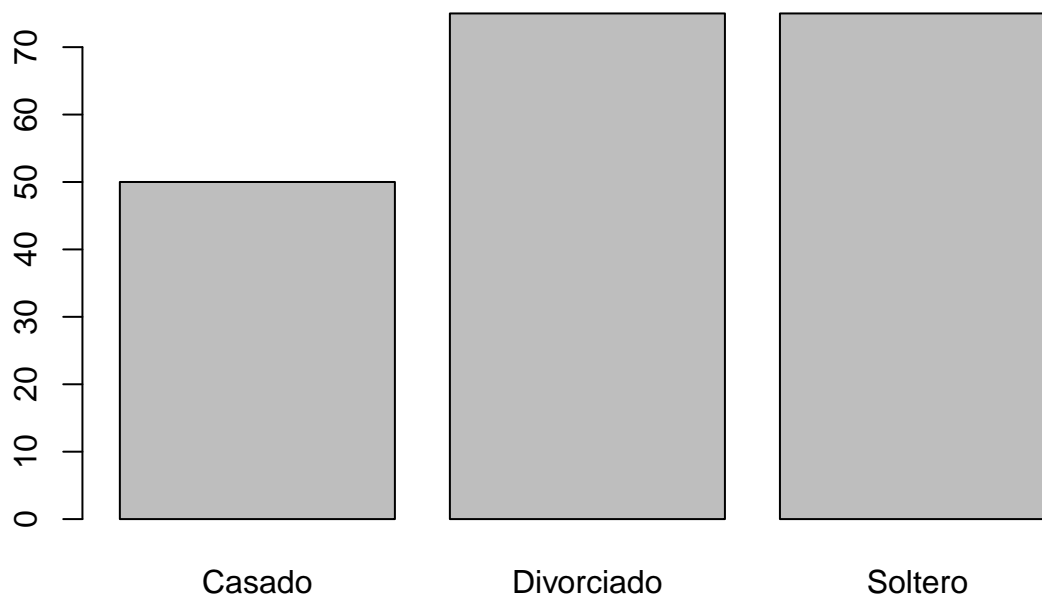
1. Análisis descriptivo de las variables categóricas: Estado civil y Diabetes

```
#####  
#### Estado civil ####  
#####  
  
class(datos$estado.civil)  
## [1] "character"  
  
table(datos$estado.civil,exclude=NULL)  
##  
##      Casado Divorciado      Soltero  
##       50         75         75  
  
prop.table(table(datos$estado.civil))  
##  
##      Casado Divorciado      Soltero  
##    0.250    0.375    0.375  
  
class(datos$sexo)  
## [1] "character"  
  
table(datos$sexo)  
##  
## Hombre  Mujer  
##    100    100  
  
table(datos$estado.civil,datos$sexo)  
##  
##              Hombre  Mujer  
## Casado           30     20  
## Divorciado       39     36  
## Soltero         31     44  
  
prop.table(table(datos$estado.civil,datos$sexo),1)  
##  
##              Hombre      Mujer  
## Casado      0.6000000 0.4000000  
## Divorciado  0.5200000 0.4800000
```

```
## Soltero 0.4133333 0.5866667

prop.table(table(datos$estado.civil,datos$sexo),2)
##
##           Hombre Mujer
## Casado      0.30  0.20
## Divorciado  0.39  0.36
## Soltero     0.31  0.44

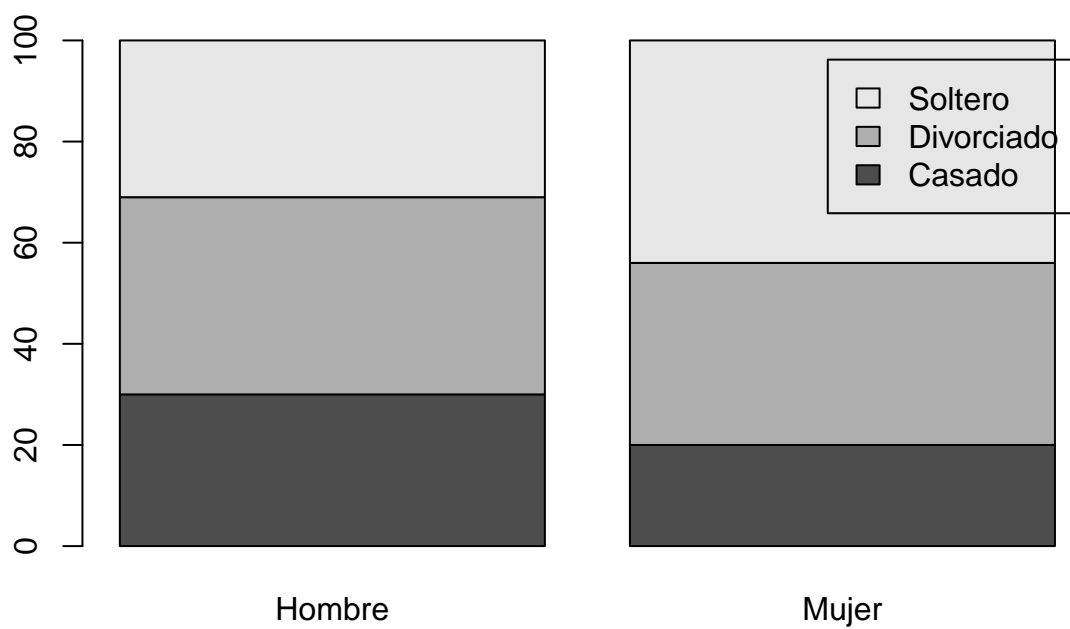
barplot(table(datos$estado.civil))
```



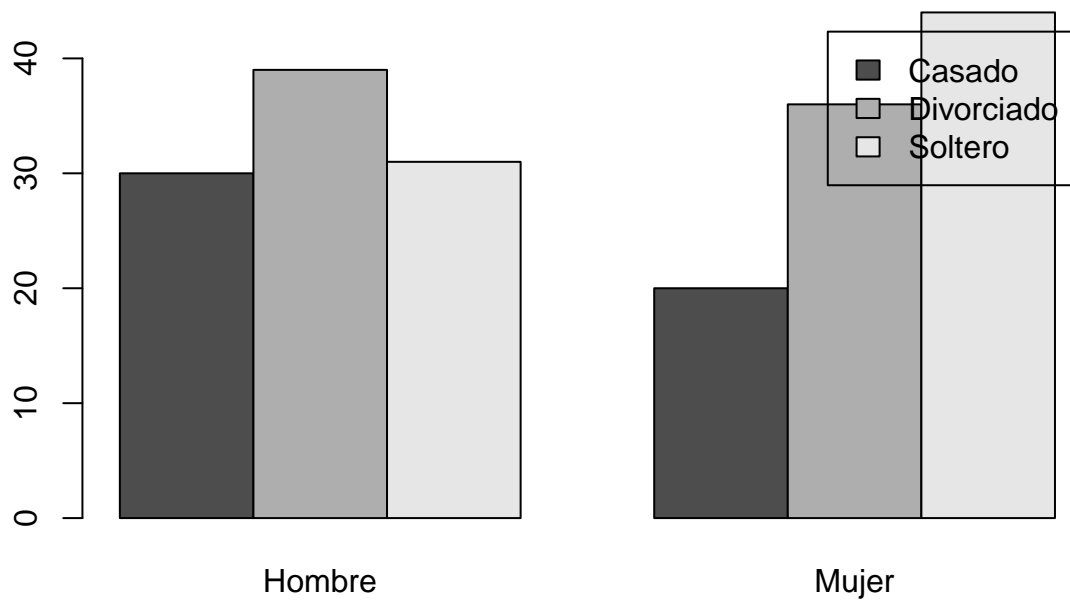
```
tabla_estado_sexo<-table(datos$estado.civil,datos$sexo)

tabla_estado_sexo
##
##           Hombre Mujer
## Casado      30   20
## Divorciado  39   36
## Soltero     31   44

barplot(tabla_estado_sexo,legend.text=rownames(tabla_estado_sexo))
```



```
barplot(tabla_estado_sexo, legend.text=rownames(tabla_estado_sexo), beside=T)
```



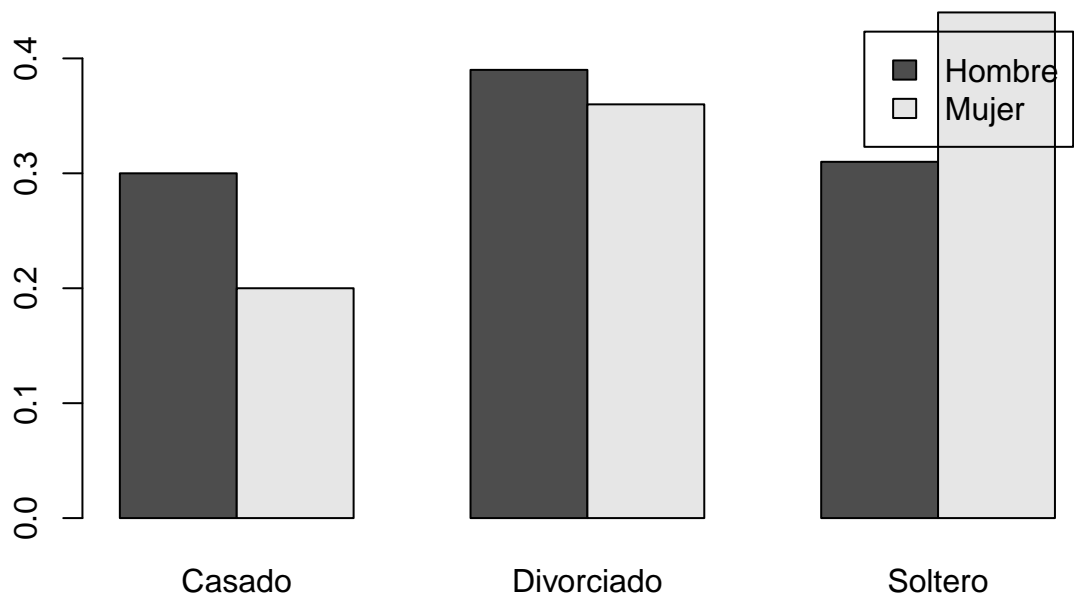
```

tabla_estado_sexo_prop<-prop.table(table(datos$estado.civil,datos$sexo),2)

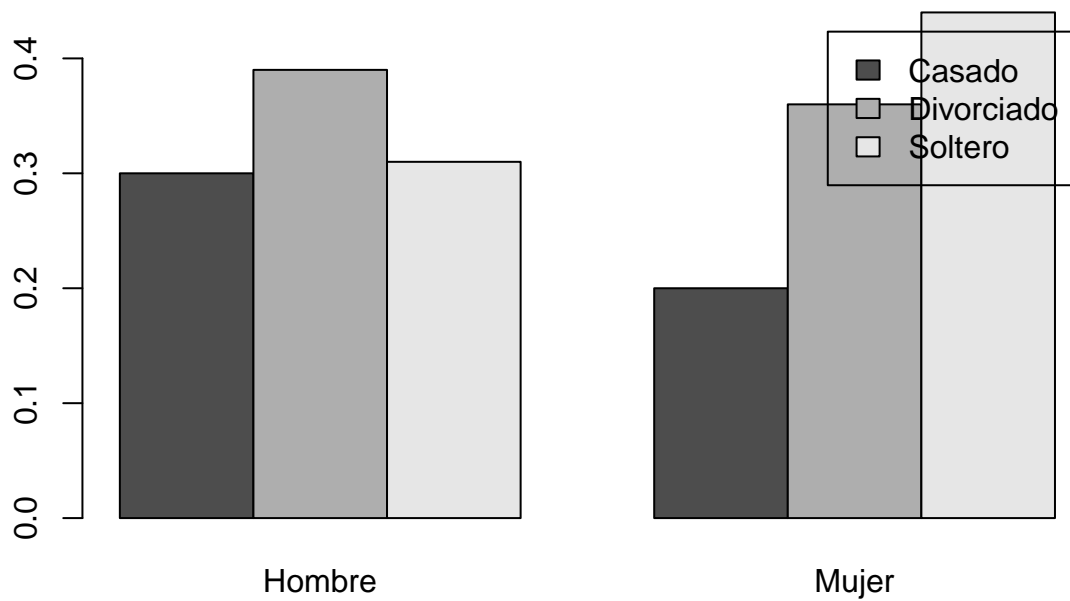
tabla_estado_sexo_prop
##
##      Hombre Mujer
## Casado    0.30  0.20
## Divorciado 0.39  0.36
## Soltero   0.31  0.44

barplot(t(tabla_estado_sexo_prop),legend.text=colnames(tabla_estado_sexo_prop),beside=T)

```



```
barplot(tabla_estado_sexo_prop, legend.text=rownames(tabla_estado_sexo_prop), beside=T)
```



```
#####
#### Diabetes ####
#####

class(datos$diabetes)
## [1] "character"

table(datos$diabetes,exclude=NULL)
##
## No Si
## 110 90

prop.table(table(datos$diabetes))
##
## No Si
## 0.55 0.45

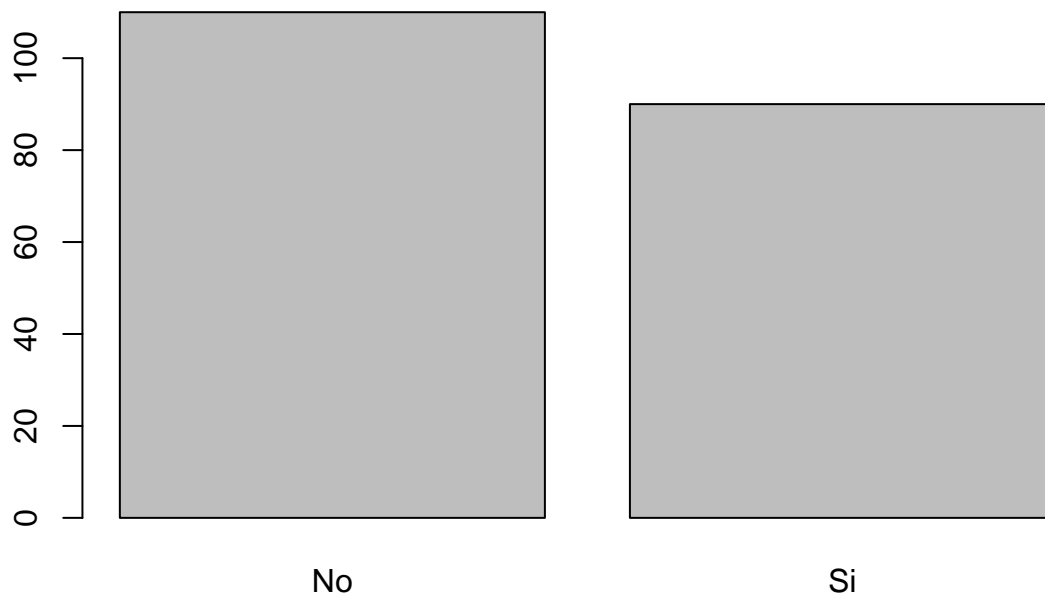
table(datos$diabetes,datos$sexo)
##
## Hombre Mujer
## No 55 55
## Si 45 45

prop.table(table(datos$diabetes,datos$sexo),1)
```

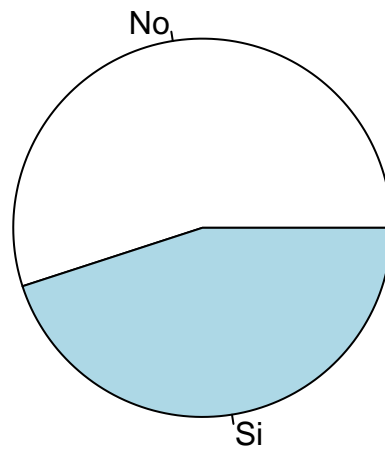
```
##
##      Hombre Mujer
## No    0.5    0.5
## Si    0.5    0.5

prop.table(table(datos$diabetes,datos$sexo),2)
##
##      Hombre Mujer
## No    0.55    0.55
## Si    0.45    0.45

barplot(table(datos$diabetes))
```



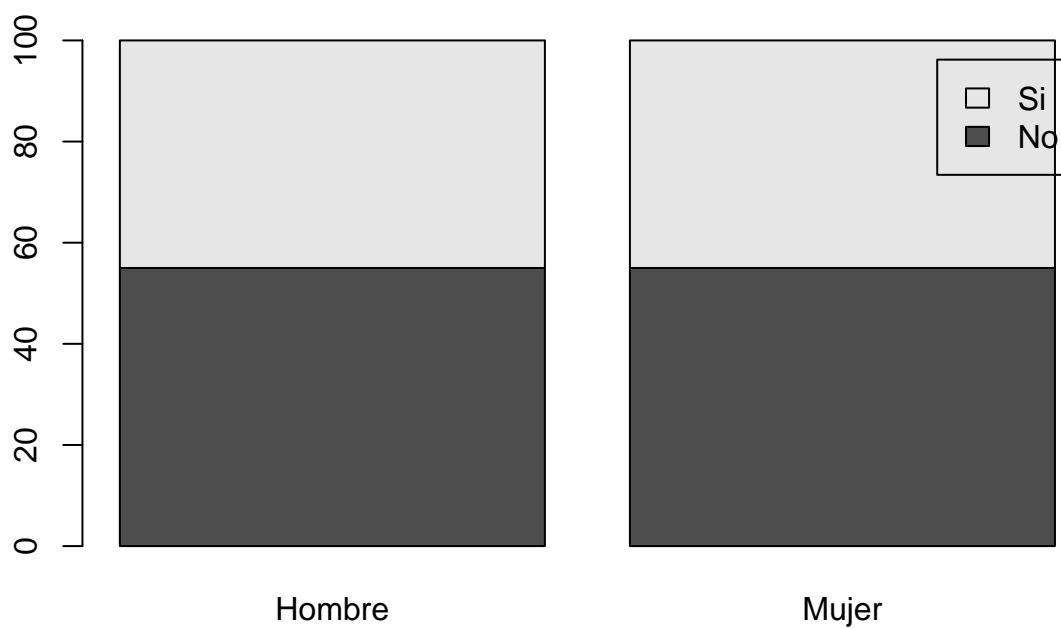
```
pie(table(datos$diabetes))
```



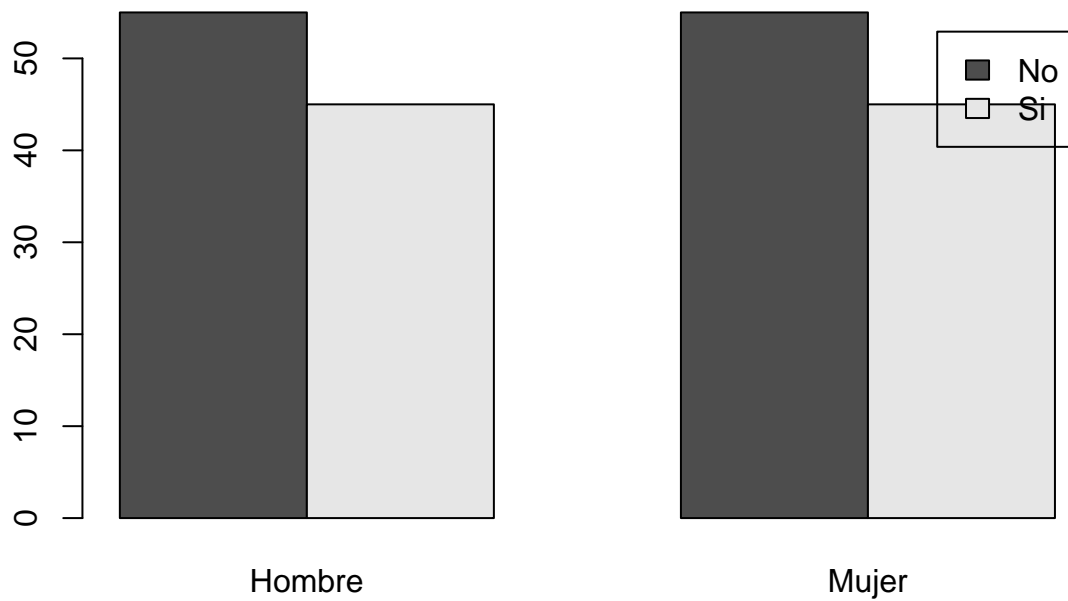
```
tabla_diabetes_sexo<-table(datos$diabetes,datos$sexo)

tabla_diabetes_sexo
##
##      Hombre Mujer
##  No      55    55
##  Si      45    45

barplot(tabla_diabetes_sexo,legend.text=rownames(tabla_diabetes_sexo))
```

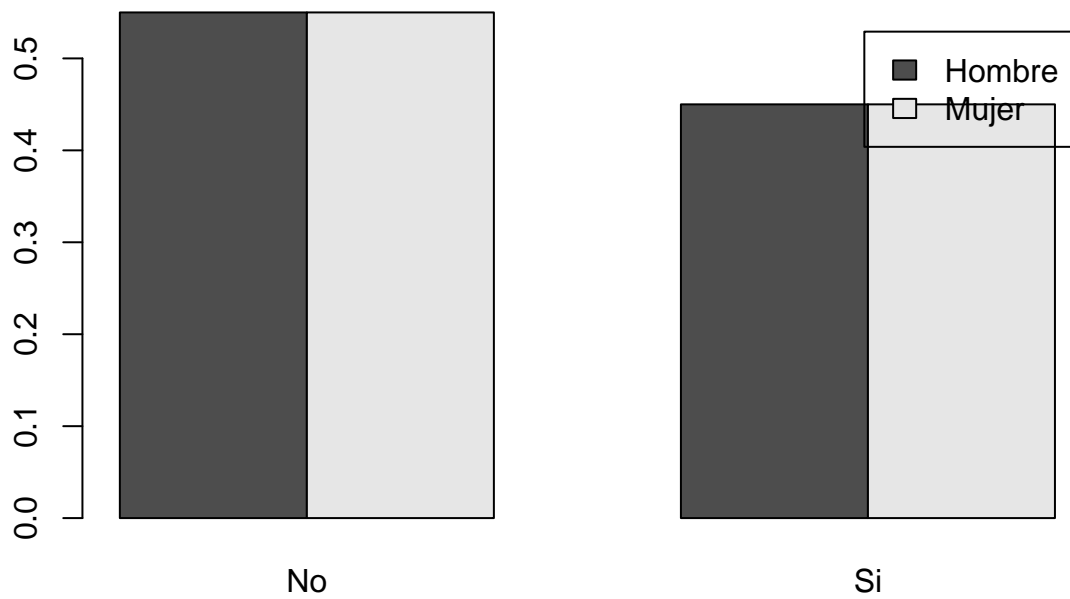
```
barplot(tabla_diabetes_sexo, legend.text=rownames(tabla_diabetes_sexo), beside=T)
```



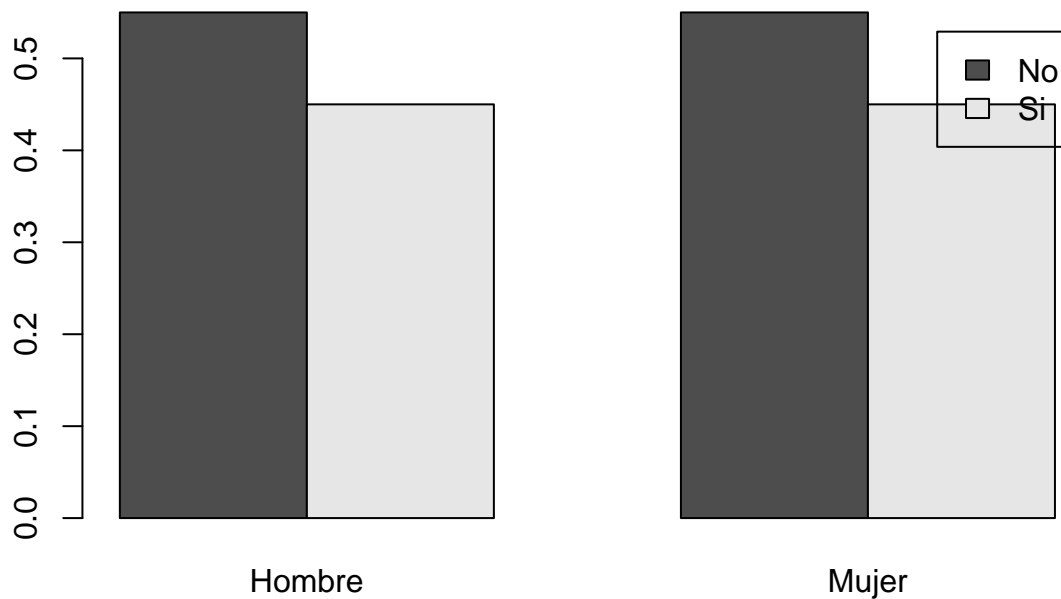
```
tabla_diabetes_sexo_prop<-prop.table(table(datos$diabetes,datos$sexo),2)

tabla_diabetes_sexo_prop
##
##      Hombre Mujer
##  No    0.55  0.55
##  Si    0.45  0.45

barplot(t(tabla_diabetes_sexo_prop),legend.text=colnames(tabla_diabetes_sexo_prop),beside=T)
```



```
barplot(tabla_diabetes_sexo_prop, legend.text=rownames(tabla_diabetes_sexo_prop), beside=T)
```



2. Análisis de asociación entre cáncer de mama en mujeres y las variables: Edad, Altura, Peso, bmi, Diabetes, Fumador, Estado civil, Nivel de estudios, Fumador teniendo en cuenta el bmi

```
# Outcome

class(datos$cancer.mama)
## [1] "character"

table(datos$cancer.mama,exclude=NULL)
##
##  No  Si <NA>
##  27  74  99

table(datos$cancer.mama,datos$sexo,exclude=NULL)
##
##      Hombre Mujer
##  No         0   27
##  Si         1   73
##  <NA>      99    0

datos.mujeres<-datos[datos$sexo%in%"Mujer",]

table(datos.mujeres$cancer.mama,exclude=NULL)
##
```

```
## No Si
## 27 73

datos.mujeres$outcome<-as.factor(datos.mujeres$cancer.mama)

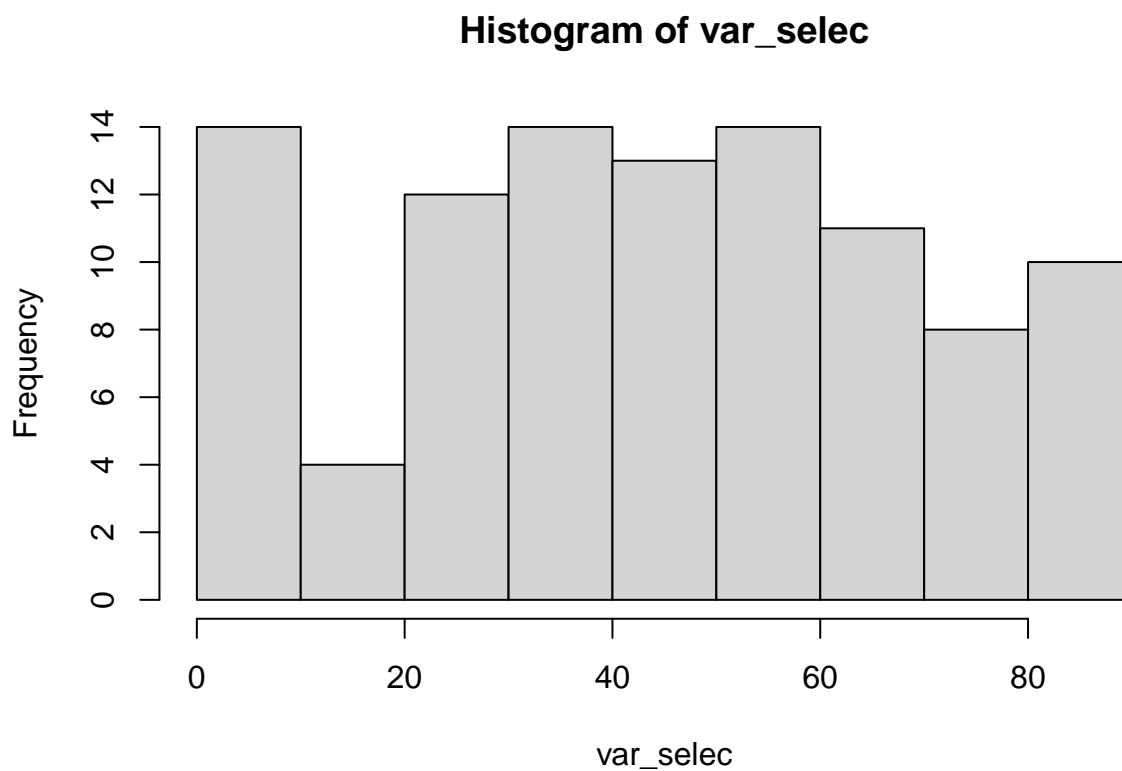
table(as.numeric(datos.mujeres$outcome),datos.mujeres$cancer.mama,exclude=NULL)
##
##      No Si
##    1 27  0
##    2  0 73

# Variable edad

var_selec<-datos.mujeres$edad

class(var_selec)
## [1] "numeric"

hist(var_selec)
```



```
range(var_selec)
## [1] 3 85
mean(var_selec)
## [1] 44.7
```

```

quantile(var_selec)
##      0%    25%    50%    75%   100%
##  3.00 27.75 43.50 65.25 85.00

modelo<-glm(outcome ~ var_selec, data=datos.mujeres,family=binomial)
exp(cbind(OR=coef(modelo),confint(modelo)))
## Waiting for profiling to be done...
##              OR      2.5 %    97.5 %
## (Intercept) 2.8903198 1.1728225 7.744539
## var_selec   0.9985146 0.9801947 1.017001

# Variable altura

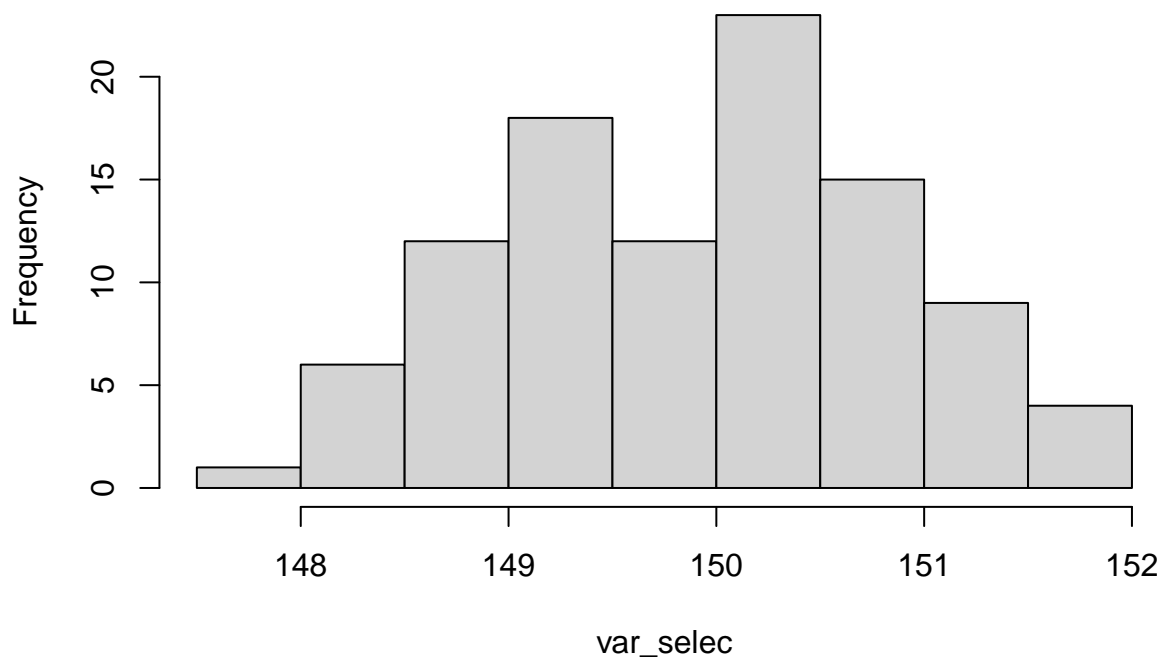
var_selec<-datos.mujeres$altura

class(var_selec)
## [1] "numeric"

hist(var_selec)

```

Histogram of var_selec



```

range(var_selec)
## [1] 147.8264 151.8925
mean(var_selec)
## [1] 149.9015

```

```

quantile(var_selec)
##           0%           25%           50%           75%           100%
## 147.8264 149.1347 150.0546 150.5714 151.8925

modelo<-glm(outcome ~ var_selec, data=datos.mujeres,family=binomial)
exp(cbind(OR=coef(modelo),confint(modelo)))
## Waiting for profiling to be done...
##              OR              2.5 %              97.5 %
## (Intercept) 5.917569e+14 2.160737e-16 2.026074e+46
## var_selec   8.023520e-01 4.946029e-01 1.280615e+00

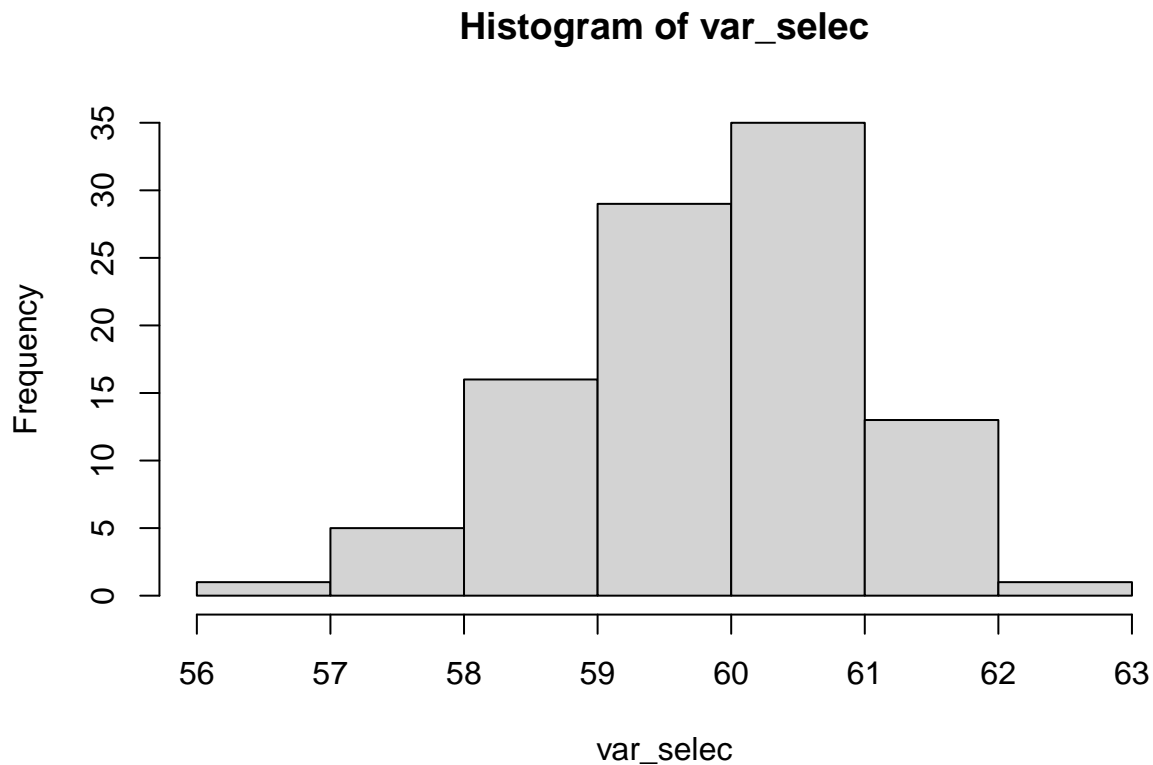
# Variable peso

var_selec<-datos.mujeres$peso

class(var_selec)
## [1] "numeric"

hist(var_selec)

```



```

range(var_selec)
## [1] 56.56025 62.30988
mean(var_selec)

```

```
## [1] 59.90394
quantile(var_selec)
##      0%      25%      50%      75%     100%
## 56.56025 59.09829 59.97703 60.67012 62.30988

modelo<-glm(outcome ~ var_selec, data=datos.mujeres,family=binomial)
exp(cbind(OR=coef(modelo),confint(modelo)))
## Waiting for profiling to be done...
##      OR      2.5 %      97.5 %
## (Intercept) 0.02318102 3.752411e-13 1.907803e+09
## var_selec   1.08271511 7.120841e-01 1.640722e+00

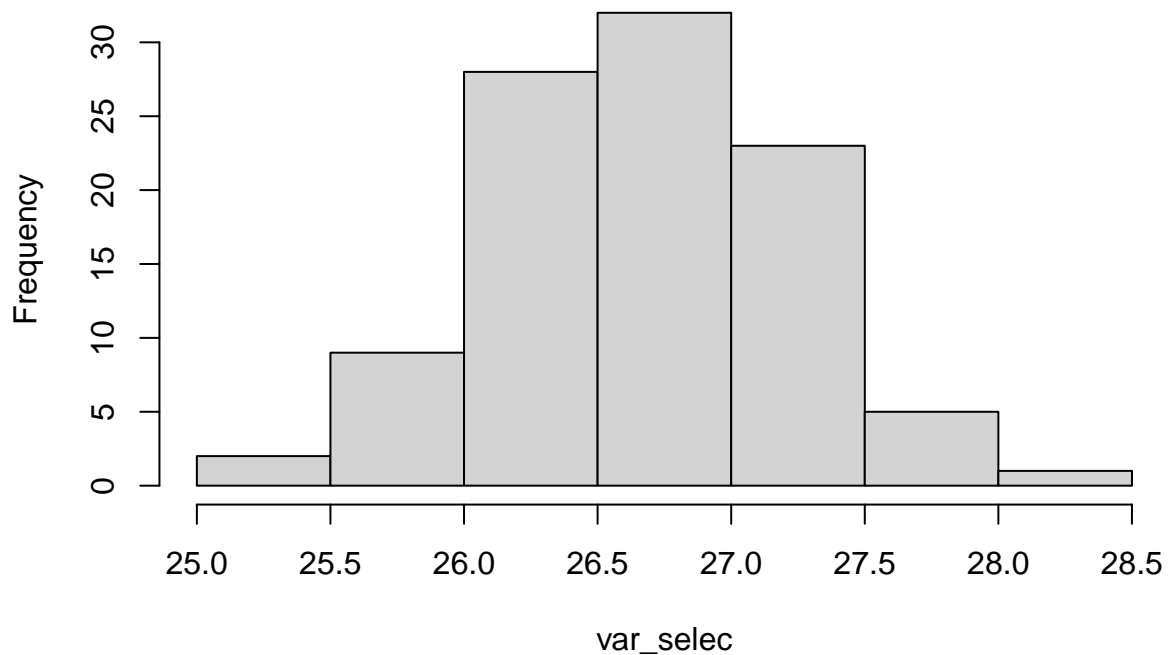
# Variable bmi

var_selec<-datos.mujeres$peso/c((datos.mujeres$altura/100)^2)

class(var_selec)
## [1] "numeric"

hist(var_selec)
```

Histogram of var_selec



```
range(var_selec)
## [1] 25.21604 28.19508
mean(var_selec)
```



```
## [1] 26.66173
quantile(var_selec)
##          0%          25%          50%          75%          100%
## 25.21604 26.24606 26.61994 27.04814 28.19508

modelo<-glm(outcome ~ var_selec, data=datos.mujeres,family=binomial)
exp(cbind(OR=coef(modelo),confint(modelo)))
## Waiting for profiling to be done...
##          OR          2.5 %          97.5 %
## (Intercept) 0.0002712335 9.516078e-14 3.722368e+05
## var_selec   1.4129174635 6.420428e-01 3.206556e+00

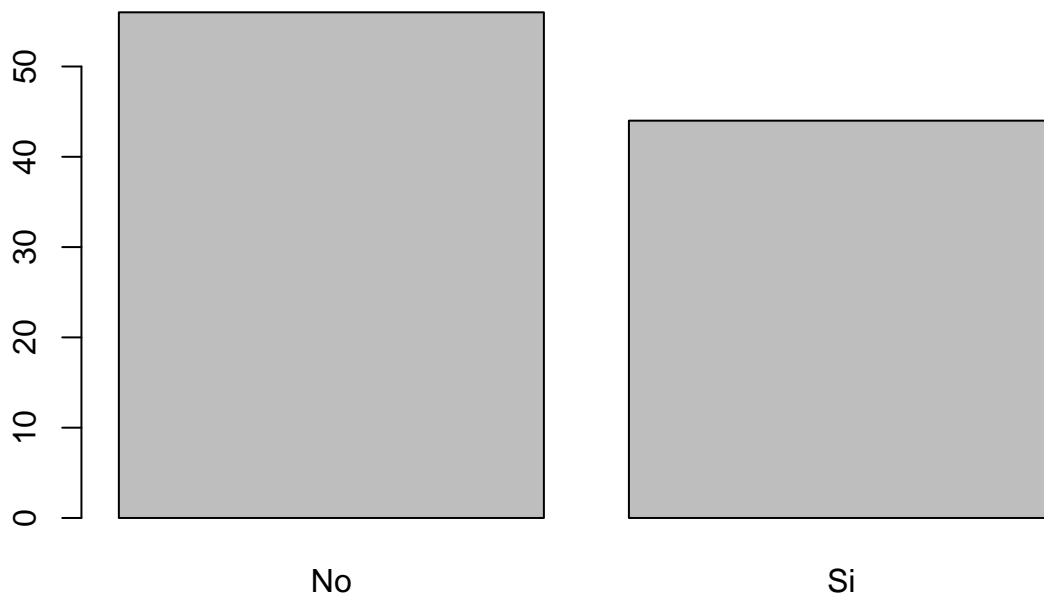
# Variable Fumador

var_selec<-datos.mujeres$fumador

class(var_selec)
## [1] "character"

var_selec<-as.factor(var_selec)

barplot(table(var_selec))
```



```
table(var_selec)
## var_selec
```

```
## No Si
## 56 44
prop.table(table(var_selec))
## var_selec
## No Si
## 0.56 0.44

modelo<-glm(outcome ~ var_selec, data=datos.mujeres,family=binomial)
exp(cbind(OR=coef(modelo),confint(modelo)))
## Waiting for profiling to be done...
##          OR      2.5 %   97.5 %
## (Intercept) 4.0909091 2.1973891 8.324706
## var_selecSi 0.4277778 0.1699309 1.043323

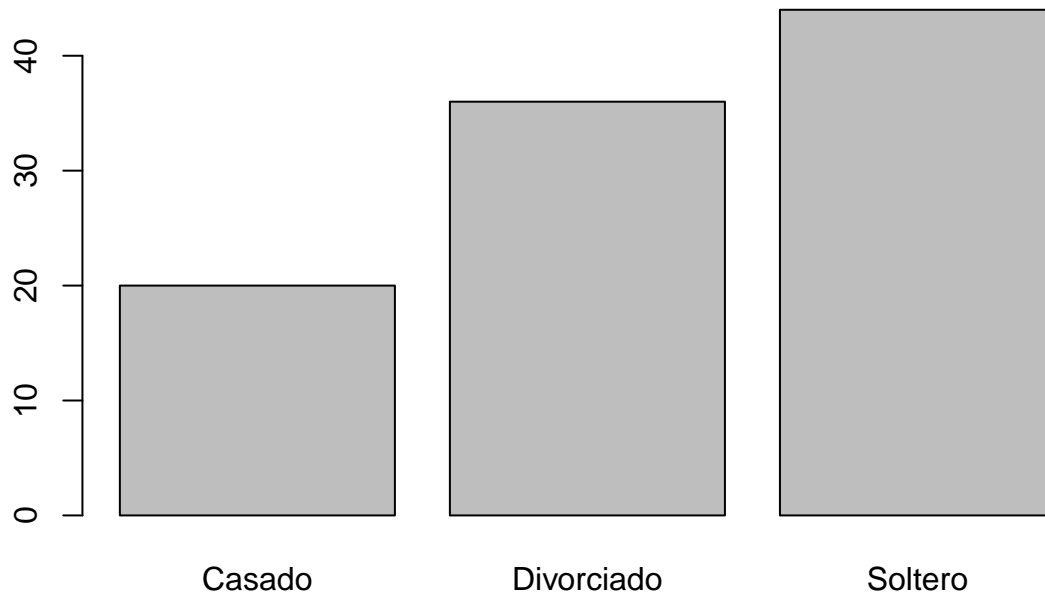
# Estado Civil

var_selec<-datos.mujeres$estado.civil

class(var_selec)
## [1] "character"

var_selec<-as.factor(var_selec)

barplot(table(var_selec))
```



```

table(var_selec)
## var_selec
##      Casado Divorciado      Soltero
##           20          36          44
prop.table(table(var_selec))
## var_selec
##      Casado Divorciado      Soltero
##           0.20          0.36          0.44

modelo<-glm(outcome ~ var_selec, data=datos.mujeres,family=binomial)
exp(cbind(OR=coef(modelo),confint(modelo)))
## Waiting for profiling to be done...
##              OR      2.5 %      97.5 %
## (Intercept)      8.9999998 2.59863458 56.6252111
## var_selecDivorciado 0.2888889 0.04114562 1.2641086
## var_selecSoltero   0.2148148 0.03147461 0.8825794

modelo<-glm(outcome ~ as.numeric(var_selec), data=datos.mujeres,family=binomial)
exp(cbind(OR=coef(modelo),confint(modelo)))
## Waiting for profiling to be done...
##              OR      2.5 %      97.5 %
## (Intercept)     11.3078452 2.5395244 64.8993487
## as.numeric(var_selec) 0.5401856 0.2743942 0.9968218

# Fumador teniendo en cuenta el bmi

datos.mujeres$bmi<-datos.mujeres$peso/c((datos.mujeres$altura/100)^2)

var_selec1<-datos.mujeres$fumador
class(var_selec1)
## [1] "character"
var_selec1<-as.factor(var_selec1)

var_selec2<-datos.mujeres$bmi

tapply(var_selec2,var_selec1,mean)
##      No      Si
## 26.62540 26.70798
t.test(var_selec2[var_selec1=="Si"],var_selec2[var_selec1=="No"])
##
## Welch Two Sample t-test
##
## data:  var_selec2[var_selec1 %in% "Si"] and var_selec2[var_selec1 %in% "No"]
## t = 0.74407, df = 97.993, p-value = 0.4586
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.1376671 0.3028312
## sample estimates:
## mean of x mean of y
## 26.70798 26.62540

```

```

modelo<-glm(outcome ~ var_selec1, data=datos.mujeres,family=binomial)
exp(cbind(OR=coef(modelo),confint(modelo)))
## Waiting for profiling to be done...
##              OR      2.5 %    97.5 %
## (Intercept)  4.0909091 2.1973891 8.324706
## var_selec1Si 0.4277778 0.1699309 1.043323

modelo<-glm(outcome ~ var_selec1 + var_selec2, data=datos.mujeres,family=binomial)
exp(cbind(OR=coef(modelo),confint(modelo)))
## Waiting for profiling to be done...
##              OR      2.5 %    97.5 %
## (Intercept)  3.432365e-05 3.828907e-15 1.228873e+05
## var_selec1Si 4.060845e-01 1.585066e-01 1.001510e+00
## var_selec2   1.552425e+00 6.796244e-01 3.691562e+00

```