

# Practica\_9

Cristian Zaldaña

2022-08-17

## Índice

<b>Práctica 09-Análisis de una variable bidimensional categórica.</b>	<b>2</b>
REALICE UN ANÁLISIS ESTADÍSTICO DE LOS DATOS. . . . .	2

## Práctica 09-Análisis de una variable bidimensional categórica.

### REALICE UN ANÁLISIS ESTADÍSTICO DE LOS DATOS.

- Crea en Excel una hoja de datos con dos columnas o variables Recuerda que al guardar la hoja, el tipo de archivo es de extensión .csv(delimitado por comas).\ Llámale al archivo: HojaCat
- Recupera desde el entorno de R la hoja de datos de Excel.

```
HojaCat <- read.csv("HojaCat.csv", strip.white=TRUE, sep = ";")  
HojaCat
```

```
##      Estado  Ocupacion  
## 1      casado desocupado  
## 2      soltero  estudia  
## 3      soltero  trabaja  
## 4      casado  estudia  
## 5 acompañado trabaja  
## 6      soltero desocupado  
## 7      casado trabaja  
## 8      casado  estudia  
## 9 acompañado desocupado  
## 10 acompañado estudia  
## 11      casado trabaja  
## 12      soltero  estudia  
## 13 acompañado desocupado  
## 14      casado desocupado  
## 15      soltero  estudia  
## 16      soltero trabaja  
## 17      casado desocupado  
## 18      soltero trabaja
```

- Conecta la hoja de datos a la segunda ruta o lista de búsqueda.

```
attach(HojaCat, pos=2) # pos especifica la posición donde buscar la conexión  
search()
```

```
## [1] ".GlobalEnv"      "HojaCat"           "package:stats"  
## [4] "package:graphics" "package:grDevices" "package:utils"  
## [7] "package:datasets" "package:methods"   "Autoloads"  
## [10] "package:base"
```

- Crea una tabla de contingencia o de doble entrada

```
tablaCont <- table(HojaCat)  
tablaCont
```

```
##      Ocupacion  
## Estado  desocupado estudia trabaja  
## acompañado      2      1      1  
## casado          3      2      2  
## soltero         1      3      3
```

```
length(HojaCat)
```

```
## [1] 2
```

Note que esta instrucción no devuelve el número de elementos, sino más bien el número de variables o columnas consideradas en el conjunto de datos.\

Encuentra la suma de cada fila de la tabla de contingencia\ Distribución marginal de X=Estado civil

```
suma.filas <- apply(tablaCont, 1, sum)
suma.filas
```

```
## acompañado    casado    soltero
##           4         7         7
```

El 1 indica que son totales por fila\

Encuentra la suma de cada fila de la tabla de contingencia\ distribución marginal de Y=Ocupación

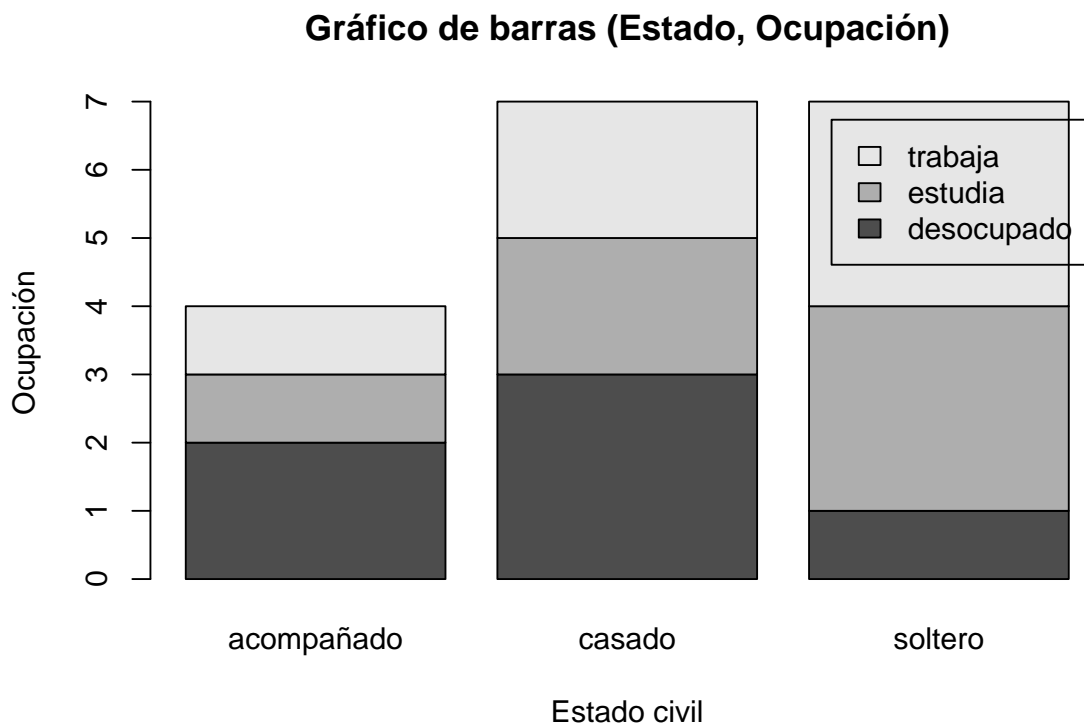
```
suma.columnas <- apply(tablaCont, 2, sum)
suma.columnas
```

```
## desocupado    estudia    trabaja
##           6         6         6
```

2 indica que son totales por columna\

Gráficos de barras para tabla de contingencia.\ Barras apiladas

```
barplot(t(tablaCont), main="Gráfico de barras (Estado, Ocupación)", xlab="Estado civil",
ylab="Ocupación", legend.text=TRUE)
```

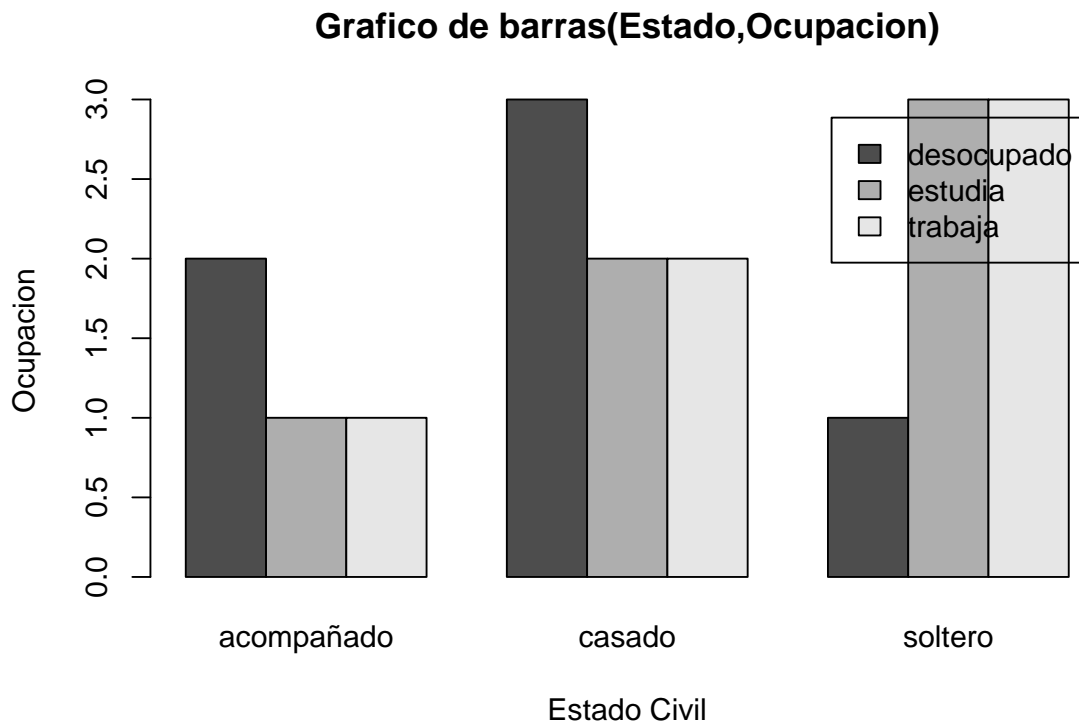


Note que `t(tablaCont)` indica que las barras representan el Estado civil de los encuestados y que éstas se subdividen en cada una de las diferentes ocupaciones consideradas. \ En caso de usar únicamente

tablaCont; las barras representarán las diferentes ocupaciones y éstas estarán subdividas en cada uno de los estados civiles.

Barras agrupadas

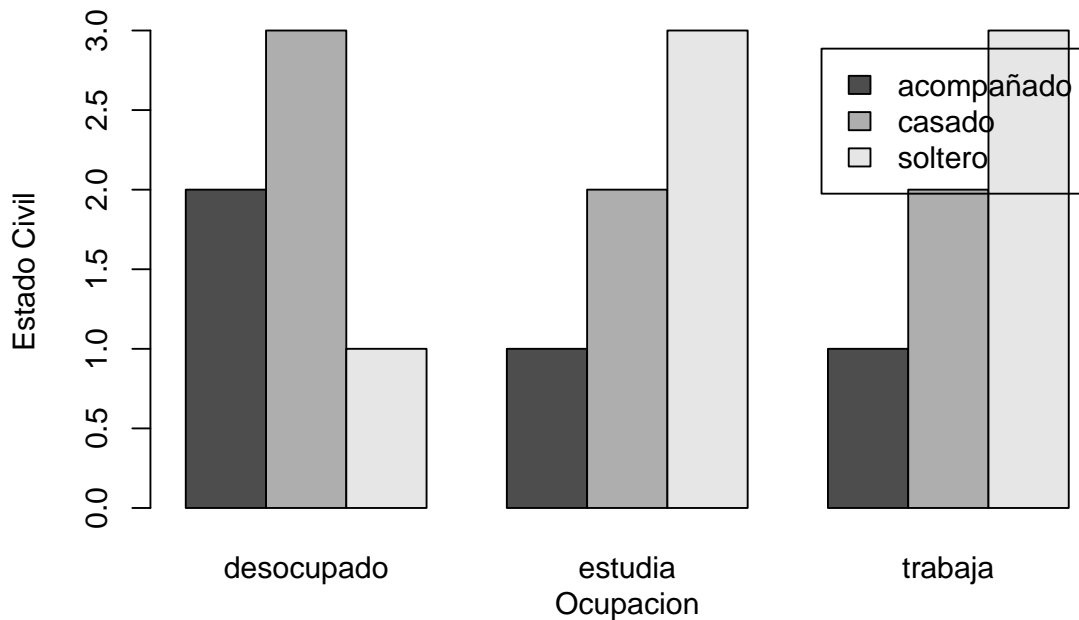
```
barplot(t(tablaCont), main = "Grafico de barras(Estado,Ocupacion)", xlab="Estado Civil",
ylab="Ocupacion", beside=TRUE, legend.text=TRUE)
```



Note que la instrucción `beside = TRUE`, indica que por cada una de las diferentes ocupaciones se creará una barra para cada estado civil. Note que al usar `beside = FALSE` se obtiene el mismo gráfico de la instrucción anterior.

```
barplot(tablaCont, main = "Grafico de barras(Ocupacion, Estado)", xlab="Ocupacion\n",
ylab="Estado Civil", beside=TRUE, legend.text=TRUE)
```

**Grafico de barras(Ocupacion, Estado)**



- Calcula tablas de proporciones o de probabilidades.\ Guardar las todas las opciones iniciales y modificar número de decimales\

```
op <- options()
options(digits=3) # sólo imprime 3 lugares decimales
options('digits')
```

```
## $digits
## [1] 3
```

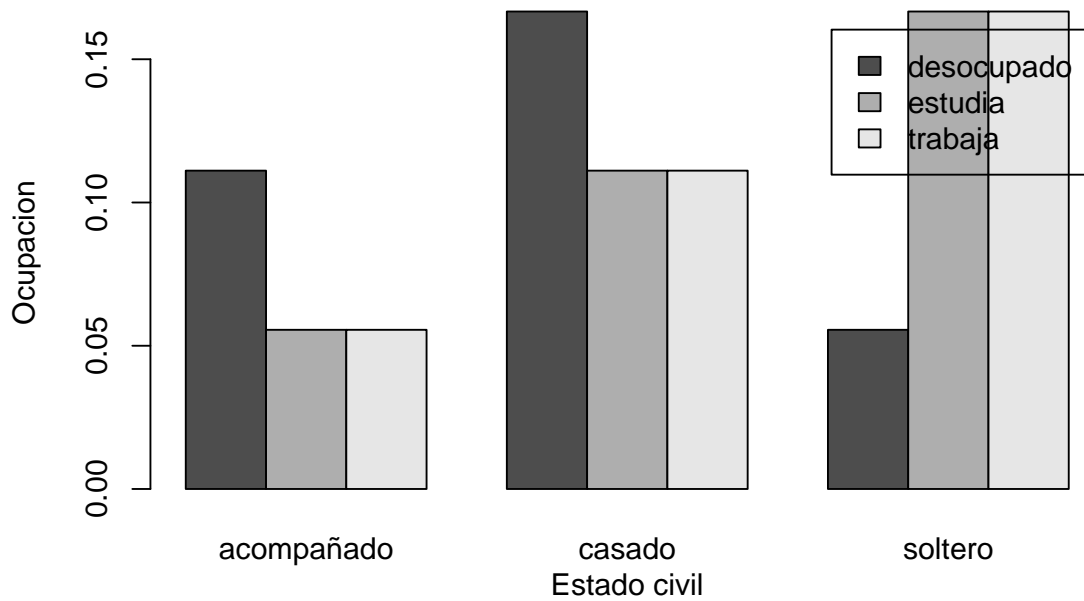
Proporciones basadas en el total de la muestra, la suma de filas y columnas suman 1.

```
propTotal <- prop.table(tablaCont)
propTotal
```

```
##          Ocupacion
## Estado   desocupado estudia trabaja
## acompañado  0.1111  0.0556  0.0556
## casado      0.1667  0.1111  0.1111
## soltero     0.0556  0.1667  0.1667
```

```
barplot(t(propTotal), main="Grafico de barras (Estado, Ocupacion)", xlab="Estado civil\n",
ylab="Ocupacion", beside=TRUE, legend.text=TRUE)
```

**Grafico de barras (Estado, Ocupacion)**



Proporciones basadas en el total por fila, cada fila suma 1.

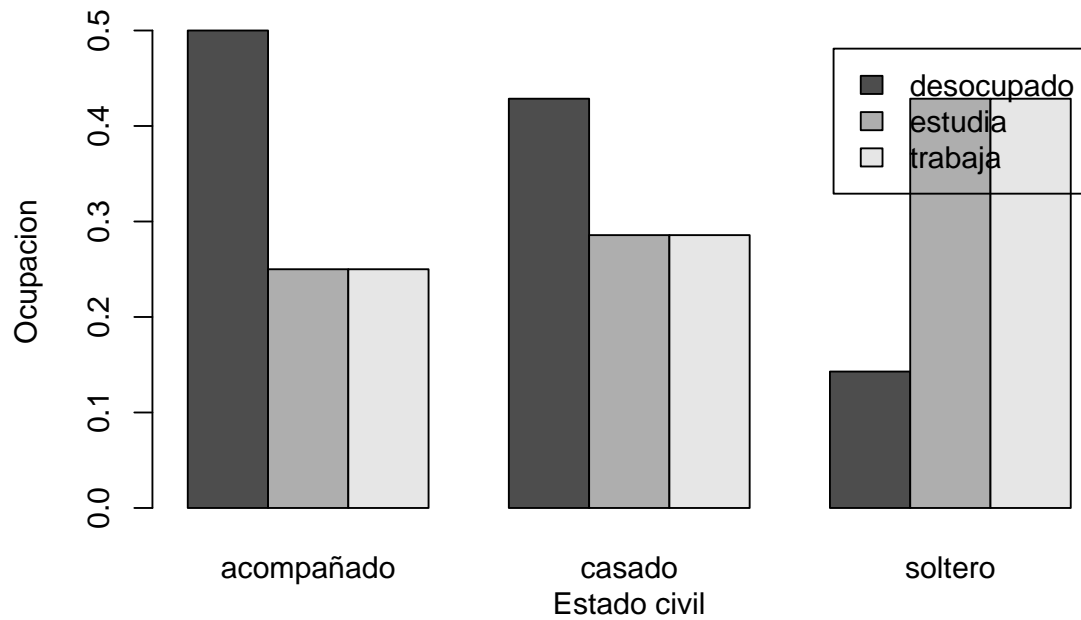
```
propFila <- prop.table(tablaCont, 1)
propFila
```

```
##          Ocupacion
## Estado  desocupado estudia trabaja
## acompañado    0.500   0.250   0.250
## casado         0.429   0.286   0.286
## soltero        0.143   0.429   0.429
```

Total por fila se indica en 1

```
barplot(t(propFila), main="Grafico de barras (Estado, Ocupacion)", xlab="Estado civil\n",
ylab="Ocupacion", beside=TRUE, legend.text=TRUE)
```

**Grafico de barras (Estado, Ocupacion)**



Proporciones basadas en el total por columna, cada columna suma 1.

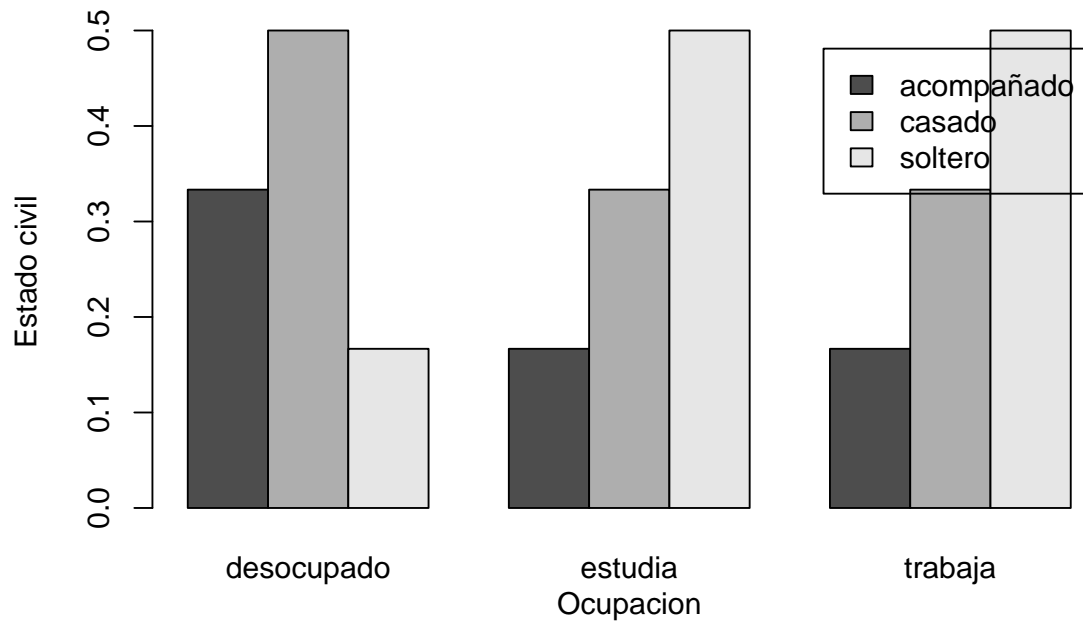
```
propColum <- prop.table(tablaCont, 2)
propColum
```

```
##          Ocupacion
## Estado   desocupado estudia trabaja
## acompañado    0.333   0.167   0.167
## casado         0.500   0.333   0.333
## soltero        0.167   0.500   0.500
```

Total por columna se indica en 2

```
barplot(propColum, main="Grafico de barras (Ocupacion, Estado)", xlab="Ocupacion\n",
ylab="Estado civil", beside=TRUE, legend.text=TRUE)
```

**Grafico de barras (Ocupacion, Estado)**

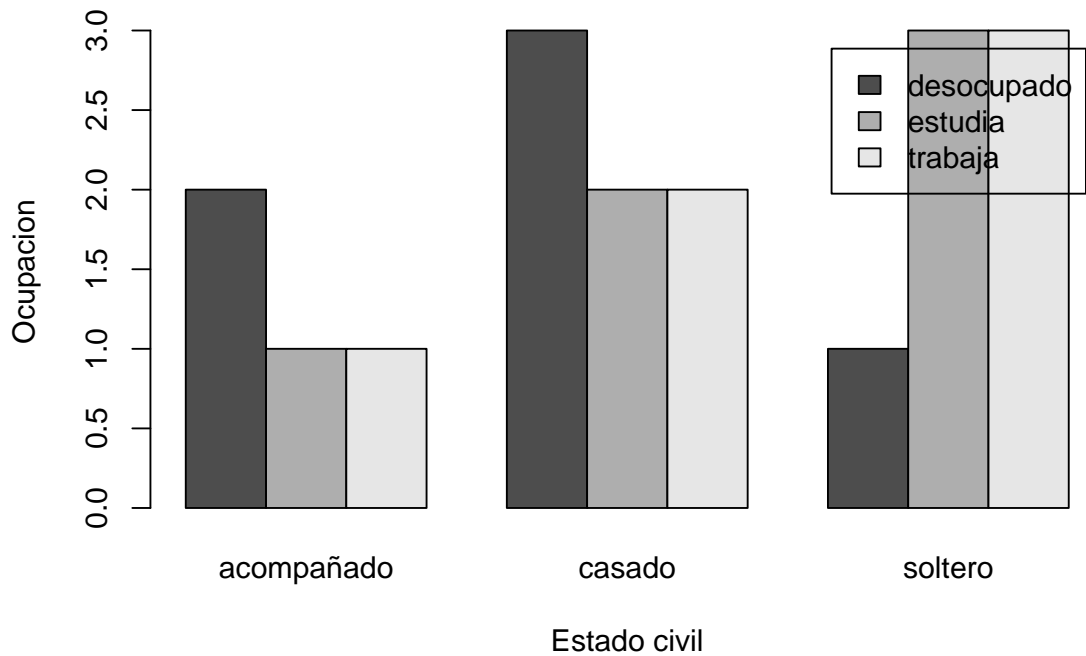


- Otra forma de elaborar los gráficos de barras para el vector bidimensional categórico. \ Gráfico de barras no apiladas y colocación de leyenda

```
barplot(table(Ocupacion, Estado), main="Grafico de barras (Estado, Ocupacion)", xlab =  
"Estado civil", ylab="Ocupacion", beside=TRUE, legend.text=T)
```

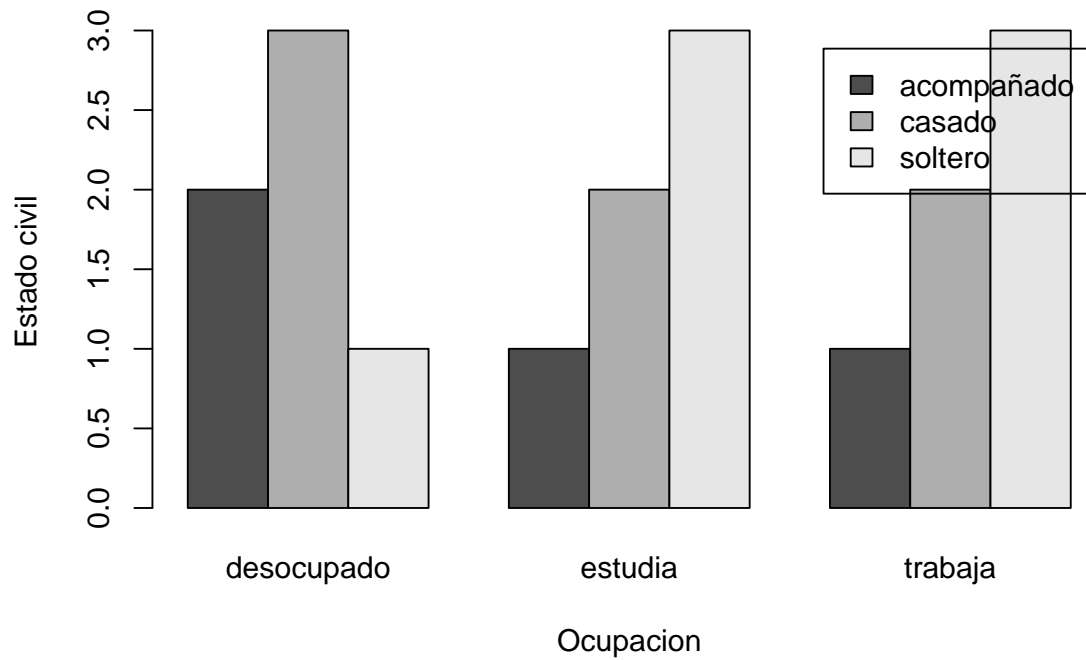


**Grafico de barras (Estado, Ocupacion)**



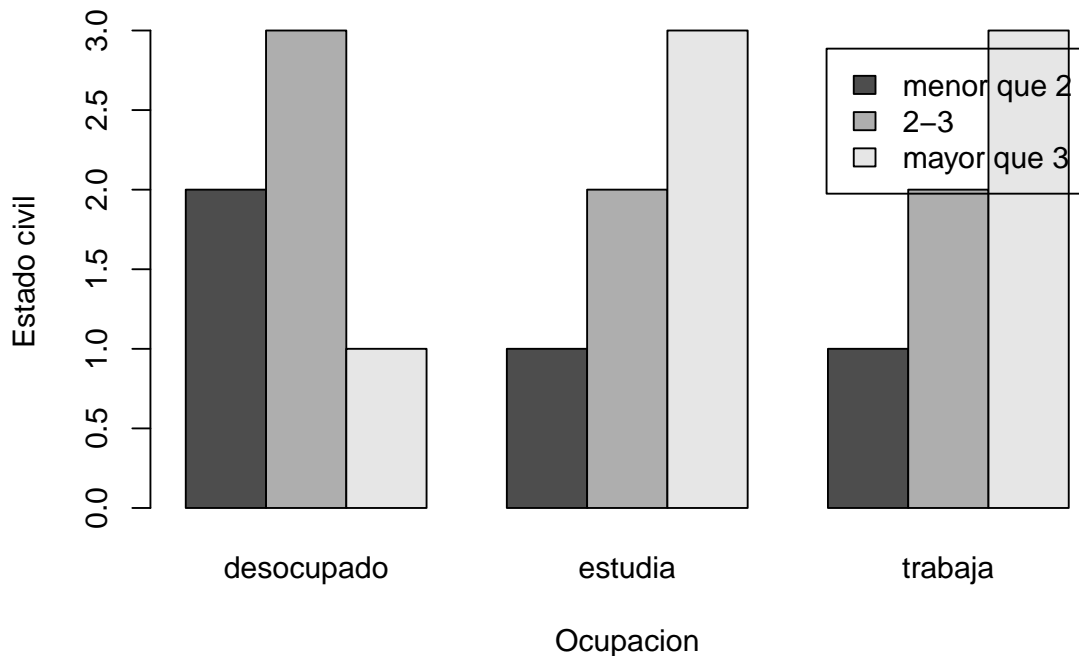
```
barplot(table(Estado, Ocupacion), main="Grafico de barras (Ocupacion, Estado)", xlab =  
"Ocupacion", ylab="Estado civil", beside=TRUE, legend.text=TRUE)
```

**Grafico de barras (Ocupacion, Estado)**



```
barplot(table(Estado, Ocupacion), main="Grafico de barras (Ocupacion, Estado)",  
xlab="Ocupacion", ylab="Estado civil", beside=TRUE, legend.text=c("menor que 2", "2-3",  
"mayor que 3"))
```

**Grafico de barras (Ocupacion, Estado)**



Note que se puede definir a conveniencia la leyenda que se desea incorporar en el gráfico con la instrucción `legend.text`

- Realizar la prueba o contraste Chi-cuadrado de independencia

```
prueba <- chisq.test(tablaCont)
```

```
## Warning in chisq.test(tablaCont): Chi-squared approximation may be incorrect
```

```
prueba
```

```
##
## Pearson's Chi-squared test
##
## data:  tablaCont
## X-squared = 2, df = 4, p-value = 0.7
```

Tenga en cuenta que las frecuencias esperadas deben ser todas mayores a 5\ Frecuencias absolutas esperadas para la prueba Chi-cuadrada\

```
prueba$expected # fij = fi./No. column
```

```
##          Ocupacion
## Estado  desocupado estudia trabaja
## acompañado    1.33    1.33    1.33
## casado         2.33    2.33    2.33
## soltero        2.33    2.33    2.33
```