

**UNIVERSIDAD DE EL SALVADOR
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE
DEPARTAMENTO DE MATEMÁTICA**



Licenciatura en Estadística

Control Estadístico del Paquete R

”UNIDAD TRES”

Práctica 09-Análisis de una variable bidimensional categórica.

**Alumna:
Martha Yoana Medina Sánchez**

**Fecha de elaboración
Santa Ana - 27 de noviembre de 2015**

Práctica 09-Análisis de unavariable bidimensional categ³rica.

REALICE UN ANÁLISIS ESTADÍSTICO DE LOS DATOS.

1. Activa tu directorio de trabajo.

```
getwd()

## [1] "C:/Users/User/Documents/PRACTICAS_YOANA_MEDINA/Yoana/PRACTICAS DE R"

setwd("C:/Users/User/Documents/PRACTICAS_YOANA_MEDINA/Yoana/PRACTICAS DE R")
```

2. Limpia de objetos el área de trabajo (Workspace).

```
ls()

## character(0)

rm(list=ls(all=TRUE))
ls()

## character(0)
```

3. Crea un nuevo Script y llámale "Script09-DatosBivariados1".
4. Crea en Excel una hoja de datos con dos columnas o variables

```
# Recuerda que al guardar la hoja, el tipo de archivo es de extension .csv
# (delimitado por comas).
# Llamale al archivo: HojaCat
```

```
# Otra forma de crear la hoja de datos es la siguiente (Vea la
# Practica 04): Primero crear las dos variables categoricas en un
# editor de texto como NotePad o WordPad, colocando nombre a cada
# columna, y llamandole "HojaCat.txt".
```

```
# Luego puede leer o recuperar este archivo con la función read.table()
HojaCat <- read.table("HojaCat.txt", header=TRUE)
HojaCat

## [1] VAR    X1      VAR.1 X2
## <0 rows> (or 0-length row.names)
```

5. Recupera desde el entorno de R la hoja de datos de Excel.

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

##      Estado.Civil  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
```

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

```
attach(HojaCat, pos=2)

# pos especifica la posicion donde buscar la conexion

search()

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

7. Crea una tabla de contingencia o de doble entrada

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

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

```
length(HojaCat)
```

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

```
# Note que esta instruccion no devuelve el numero de elementos, sino mas  
# bien el numero de variables o columnas consideradas en el conjunto de  
# datos.
```

```
# Encuentra la suma de cada fila de la tabla de contingencia
```

```
# Distribucion 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 filade la tabla de contingencia
```

```
# distribución marginal de
```

```
Y="Ocupaci\ 'on"
```

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

```
## desocupado    estudia    trabaja
```

```
##           6           6           6
```

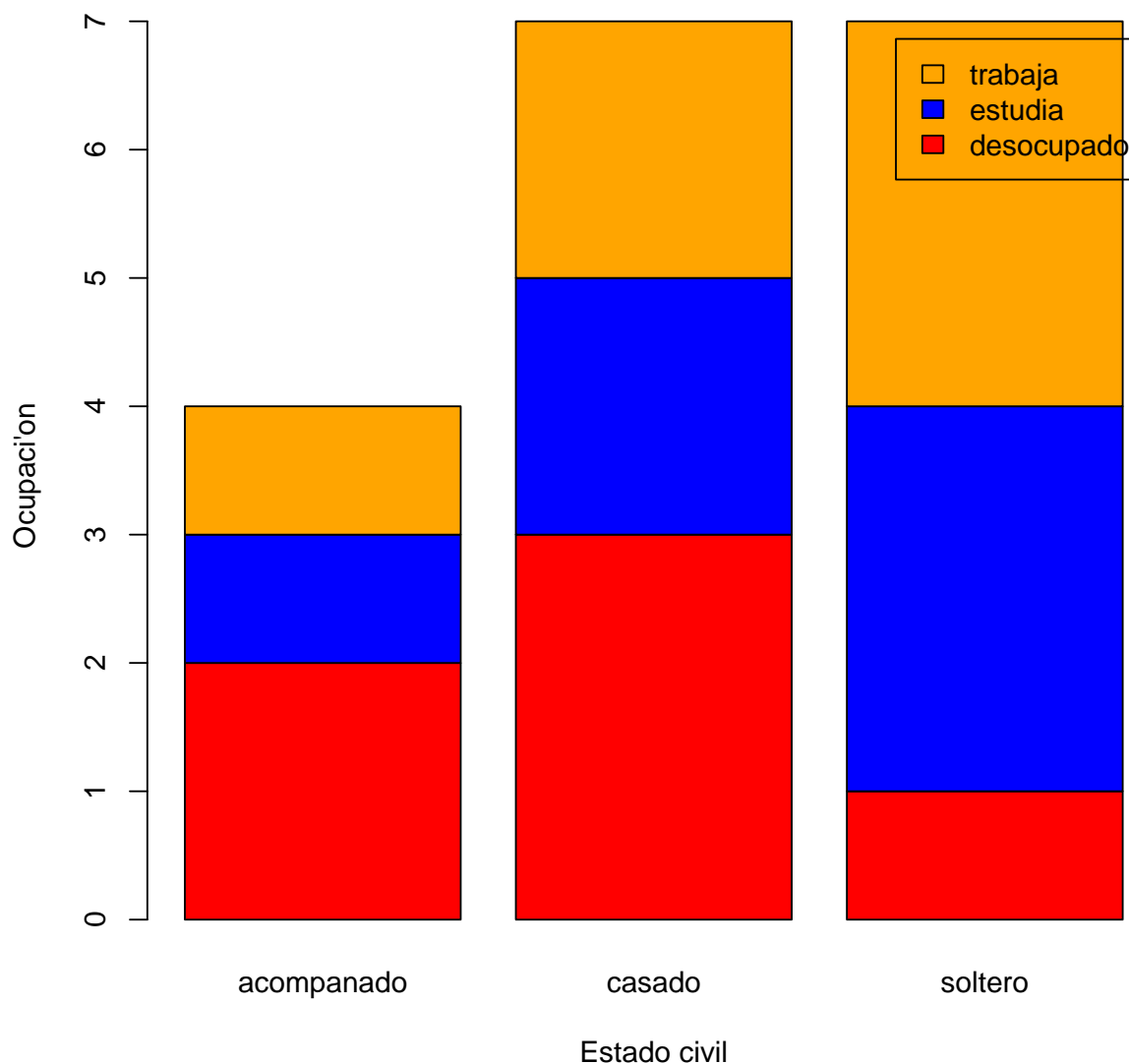
```
# 2 indica que son totales por columna
```

```
# Graficos de barras para tabla de contingencia.
```

```
# Barras apiladas
```

```
barplot(t(tablaCont), main="Gráfico de barras (Estado Civil,  
Ocupación)", xlab="Estado civil", ylab="Ocupación",  
col=c("red", "blue", "orange"), legend.text=TRUE)
```

Gráfico de barras (Estado Civil, Ocupación)

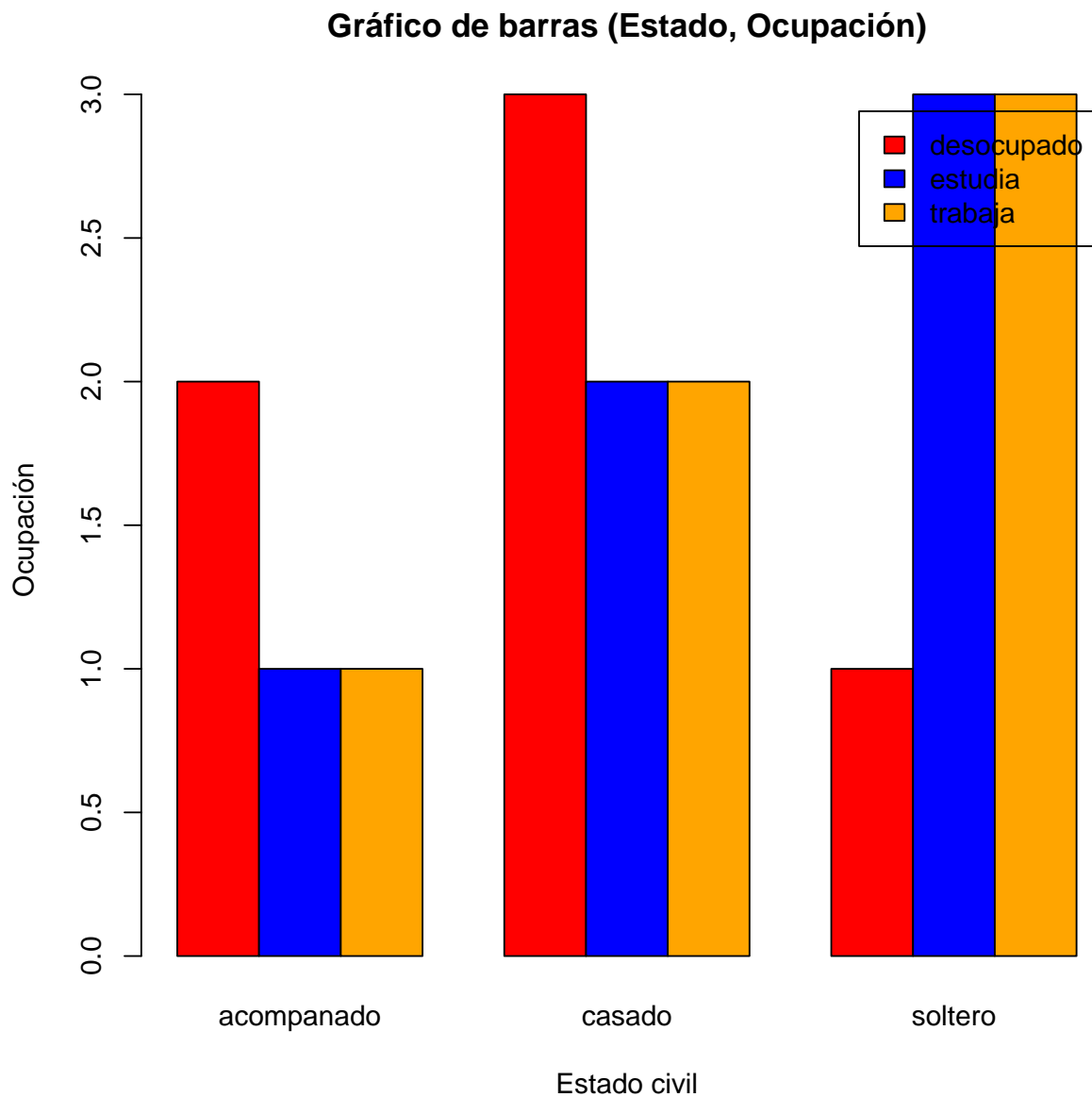


Note que `t(tablaCont)` indica que las barras representan el Estado civil de los encuestados y que estas se subdividen en cada una de las diferentes ocupaciones consideradas.

En caso de usar únicamente `tablaCont`; las barras representarían las diferentes ocupaciones y estas estarían subdivididas en cada uno de los estados civiles.

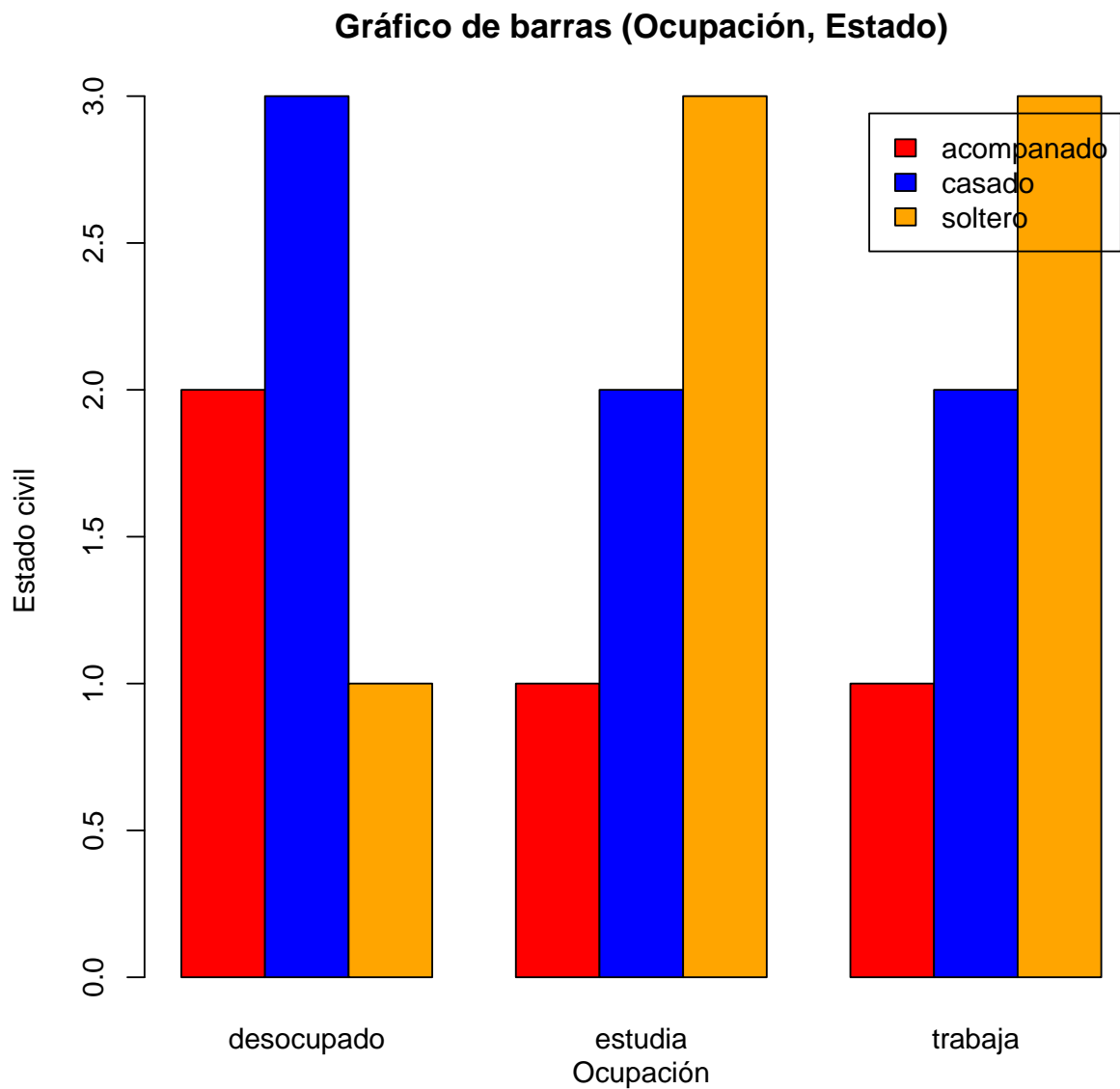
Barras agrupadas

```
barplot(t(tablaCont), main="Gráfico de barras (Estado, Ocupación)",
        xlab="Estado civil", ylab="Ocupación", beside=TRUE,
        col=c("red", "blue", "orange"), legend.text=TRUE)
```



*# Note que la instruccion beside =TRUE, indica que por cada una de las
diferentes ocupaciones se creara una barra para cada estado civil. Note
que al usar beside =FALSE se obtiene el mismo grafico de la instruccion
anterior.*

```
barplot(tablaCont, main="Gráfico de barras (Ocupación, Estado)",
        xlab="Ocupación", ylab="Estado civil", beside=TRUE,
        col=c("red", "blue", "orange"), legend.text=TRUE)
```



8. Calcula tablas de proporciones o de probabilidades.

```
# Guardar las todas las opciones iniciales y modificar numero
# de decimales

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

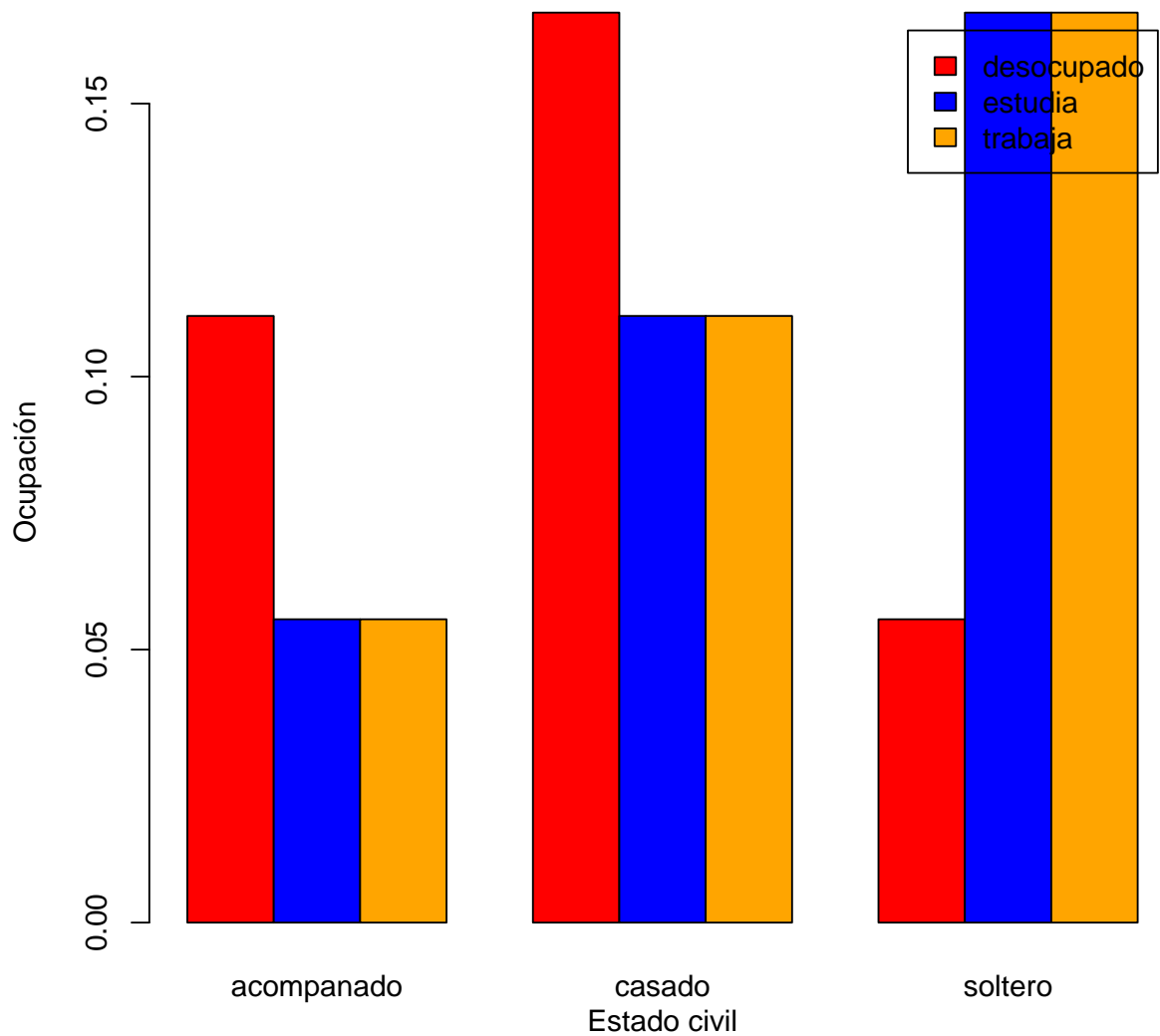
## $digits
## [1] 3
```

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

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

```
##              Ocupacion  
## Estado.Civil 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="Gráfico de barras (Estado, Ocupación)",  
        xlab="Estado civil\n", ylab="Ocupación", beside=TRUE,  
        col=c("red", "blue", "orange"), legend.text=TRUE)
```


Gráfico de barras (Estado, Ocupación)

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

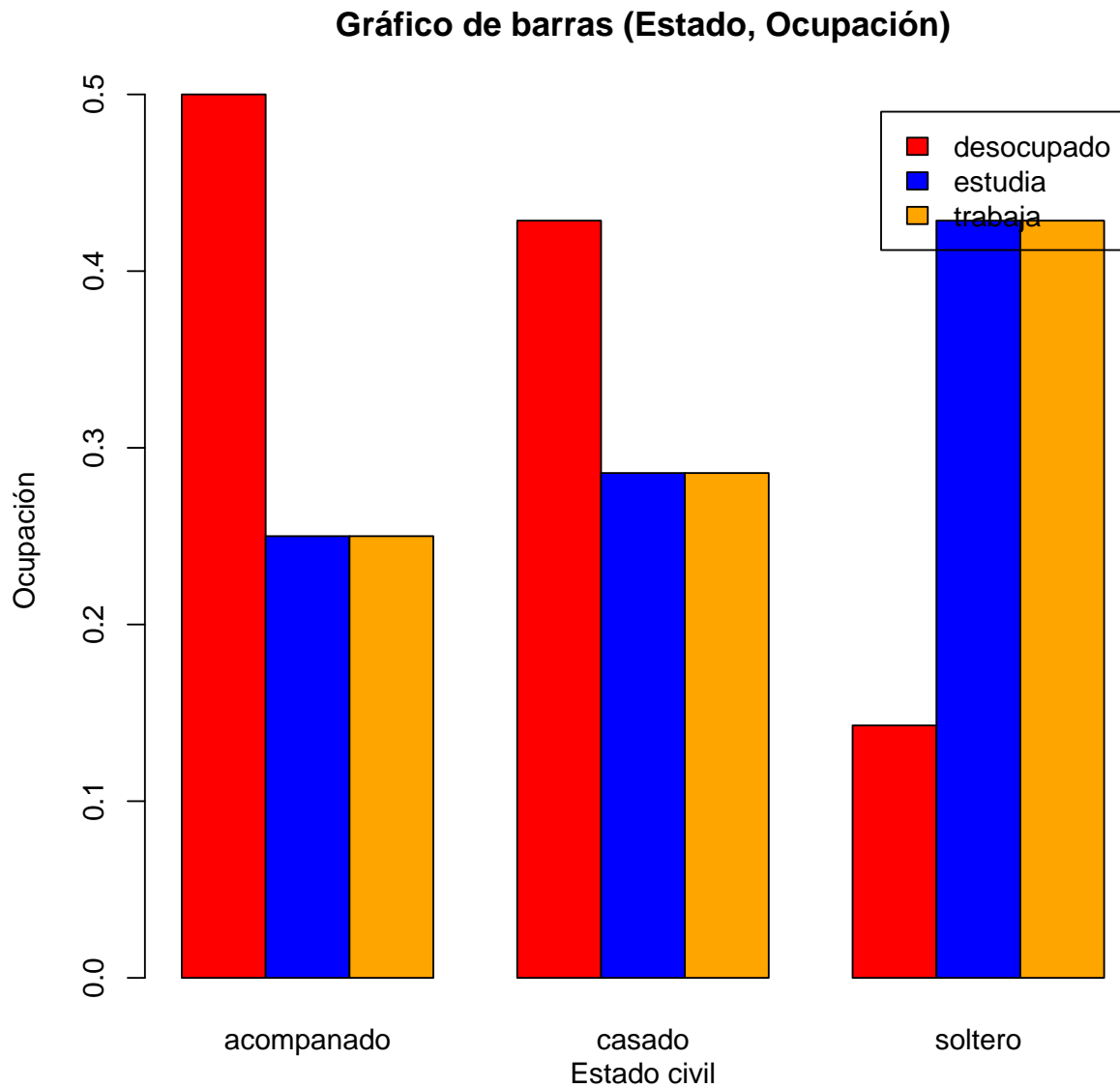
```
propFila <- prop.table(tablaCont, 1); propFila
```

```
##           Ocupacion
## Estado.Civil 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="Gráfico de barras (Estado, Ocupación)",
```

```
xlab="Estado civil\n", ylab="Ocupación", beside=TRUE,
col=c("red", "blue", "orange"), legend.text=TRUE)
```



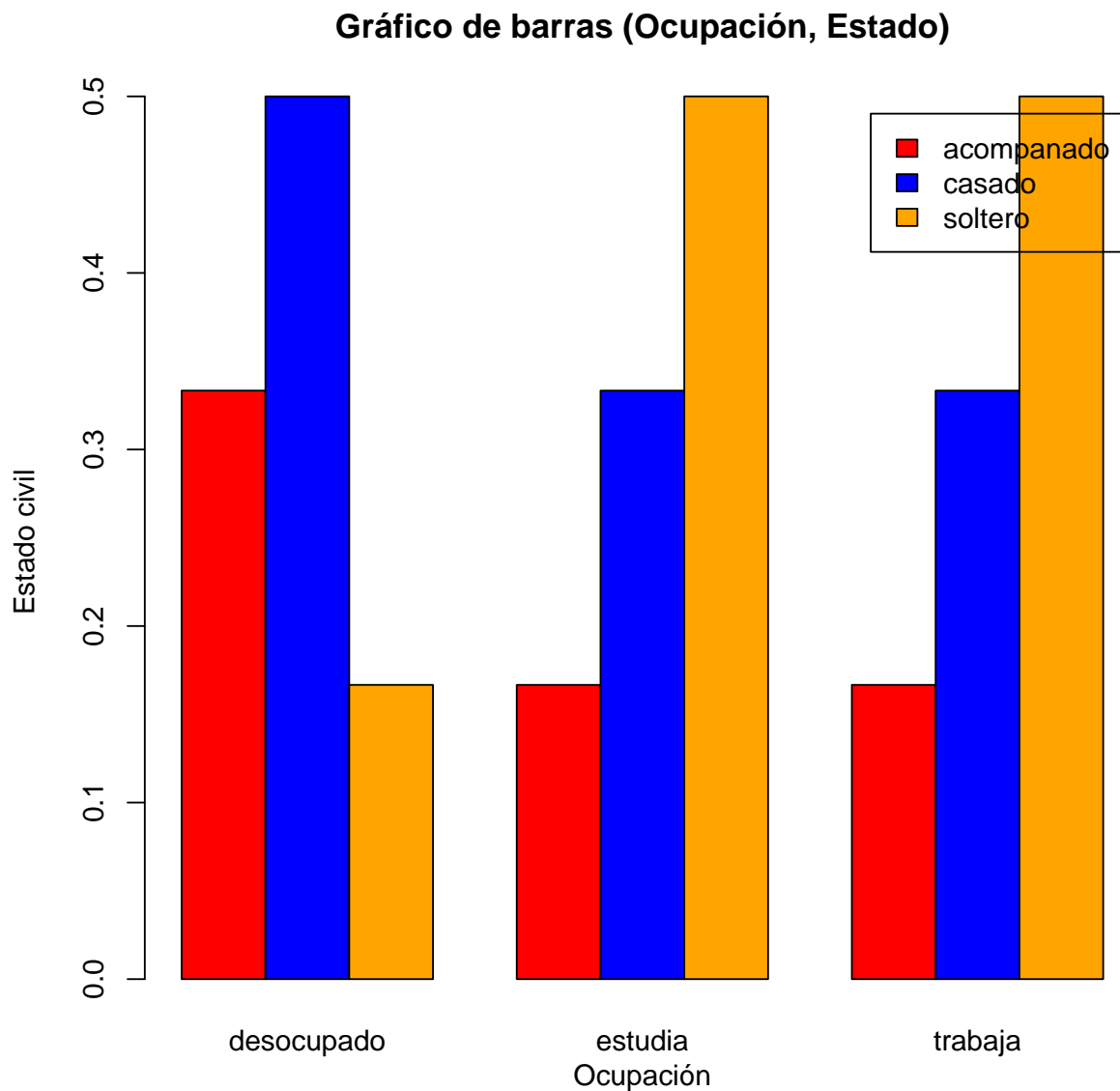
Proporciones basadas en el total por columna, cada columna suma 1.

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

```
##          Ocupacion
## Estado.Civil 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="Gráfico de barras (Ocupación, Estado)",
        xlab="Ocupación", ylab="Estado civil", beside=TRUE,
        col=c("red", "blue", "orange"), legend.text=TRUE)
```

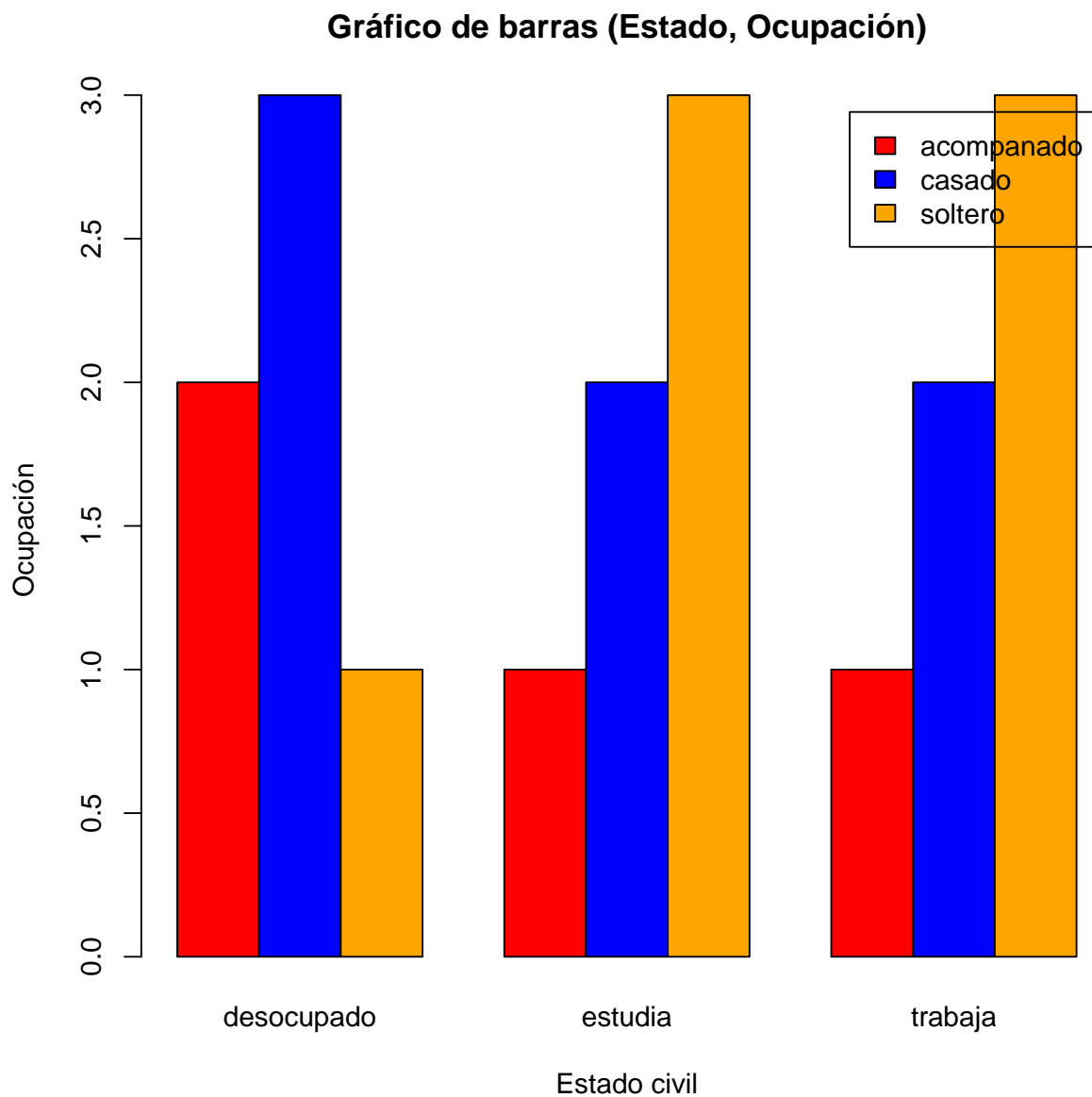


9. Otra forma de elaborar los gráficos de barras para el vector bidimensional categórico.

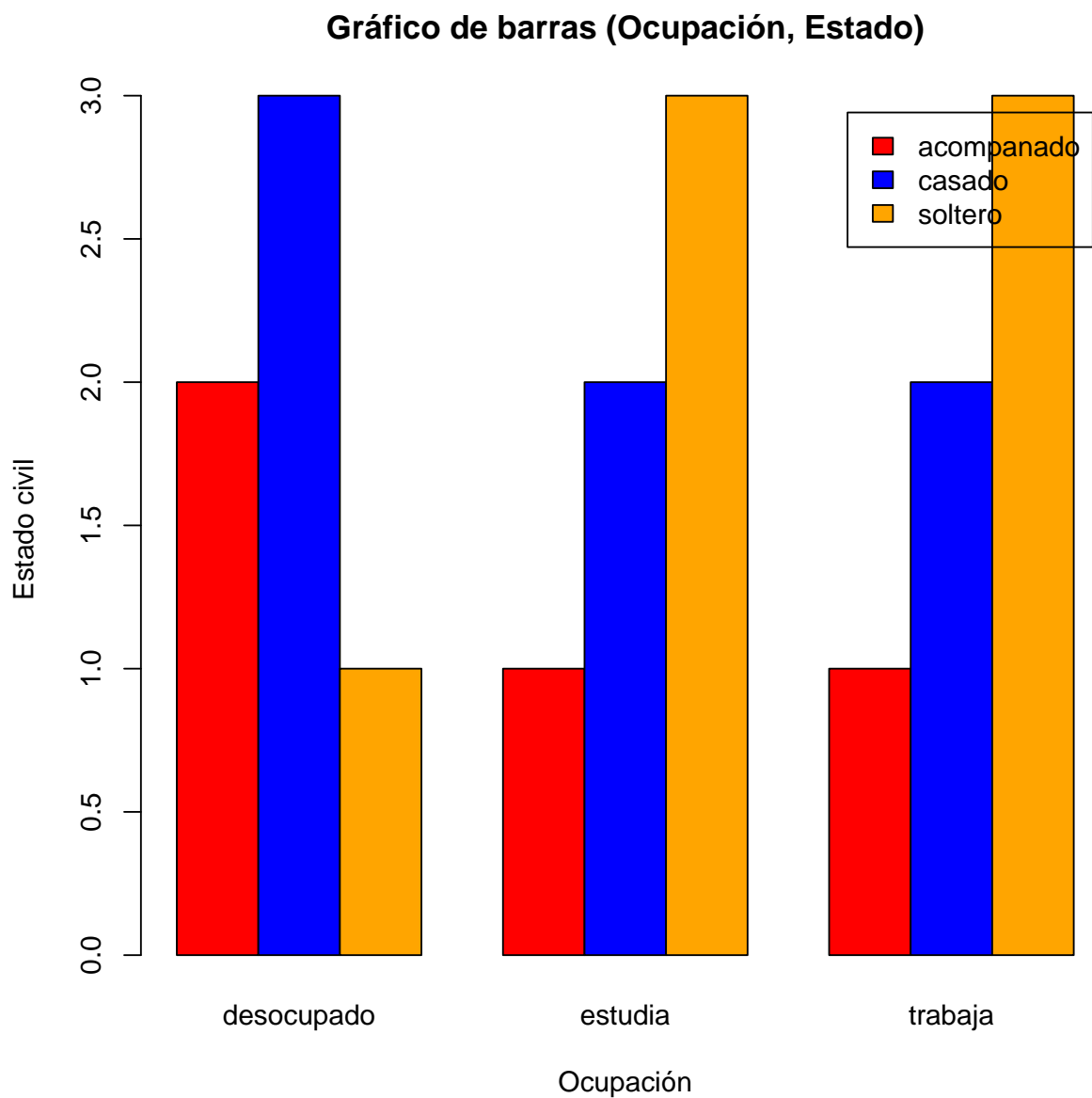
```
# Gráfico de barras no apiladas y colocación de leyenda
```

```
barplot(table(HojaCat), main="Gráfico de barras (Estado, Ocupación)",
        xlab = "Estado civil", ylab="Ocupación", beside=TRUE,
```

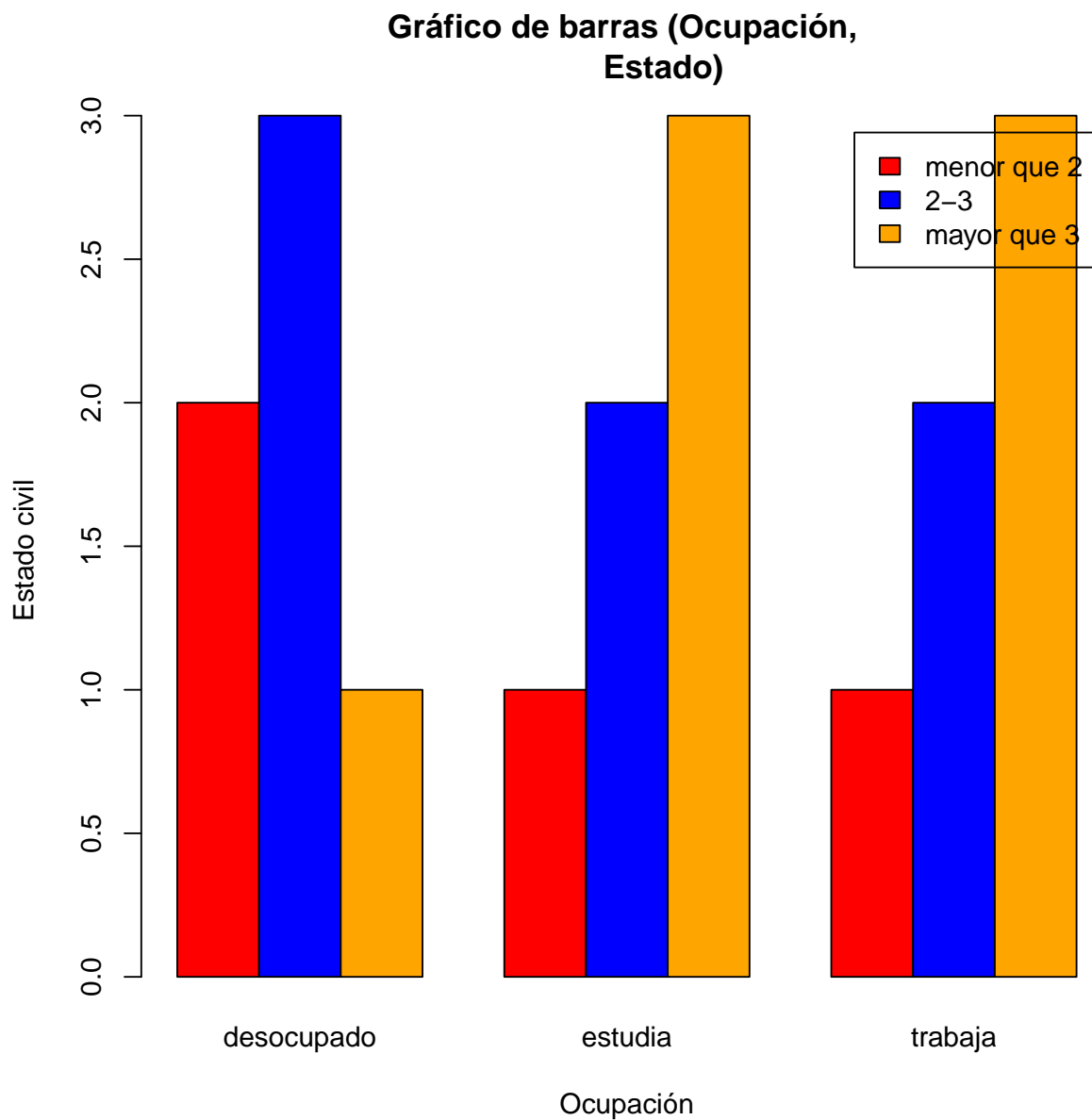
```
col=c("red", "blue", "orange"), legend.text=T)
```



```
barplot(table(HojaCat), main="Gráfico de barras (Ocupación, Estado)",
  xlab = "Ocupación", ylab="Estado civil", beside=TRUE,
  col=c("red", "blue", "orange"), legend.text=TRUE)
```



```
barplot(table(HojaCat), main="Gráfico de barras (Ocupación, Estado)", xlab="Ocupación", ylab="Estado civil",
  beside=TRUE, col=c("red", "blue", "orange"),
  legend.text=c("menor que 2", "2-3", "mayor que 3"))
```



*# Note que se puede definir a conveniencia la leyenda que se desea
incorporar en el grafico con la instruccion legend.text*

10. Realizar la prueba o contraste Chi-cuadrado de independencia

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

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

##
## 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 #  $f_{ij} = f_{i.}/\text{No. column}$ 

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