

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



Licenciatura en Estadística

Control Estadístico del Paquete R

”UNIDAD TRES”

**Alumna:
Erika Beatríz Guillén Pineda**

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

1. ANÁLISIS DE UNA VARIABLE BIDIMENSIONAL CATEGÓRICA

Ejemplo 1: Se selecciona aleatoriamente una muestra de 18 personas adultas, para estudiar si existe relación entre su estado civil y su ocupación.

1.1. REALICE UN ANÁLISIS ESTADÍSTICO DE LOS DATOS.

1) Activa tu directorio de trabajo.

```
getwd()

## [1] "C:/Users/User/Documents/TODAS_PRACTICAS"

setwd("C:/Users/User/Documents/TODAS_PRACTICAS")
```

@ 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 llámalo "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 extensión .csv (delimitado por coma)
# Llámale al archivo: HojaCat
# Otra forma de crear la hoja de datos es la siguiente (Vea la Práctica 04):
# Primero crear las dos variables categóricas en un editor de texto como NotePad o Word
# Luego puede leer o recuperar este archivo con la función read.table()
HojaCat <- read.table("HojaCat.txt", header=TRUE)
HojaCat
```

	ESTADO	OCUPACIÓN
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
```

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

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

```
## ESTADO OCUPACIÓN
## 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 posición donde buscar la conexión
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

##                OCUPACIÃ\\N
## ESTADO        Desocupado Estudia Trabaja
## Acompañado      2          1          1
## Casado           3          2          2
## Soltero         1          3          3

length(HojaCat)

## [1] 2

# sino m\ 'as bien el n\ 'umero de variables o columnas consideradas
# en el conjunto de datos.

# Encuentra la suma de cada fila de la tabla de contingencia
# Distribuci\ 'on 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
# distribuci\ 'on marginal de Y=Ocupaci\ 'on

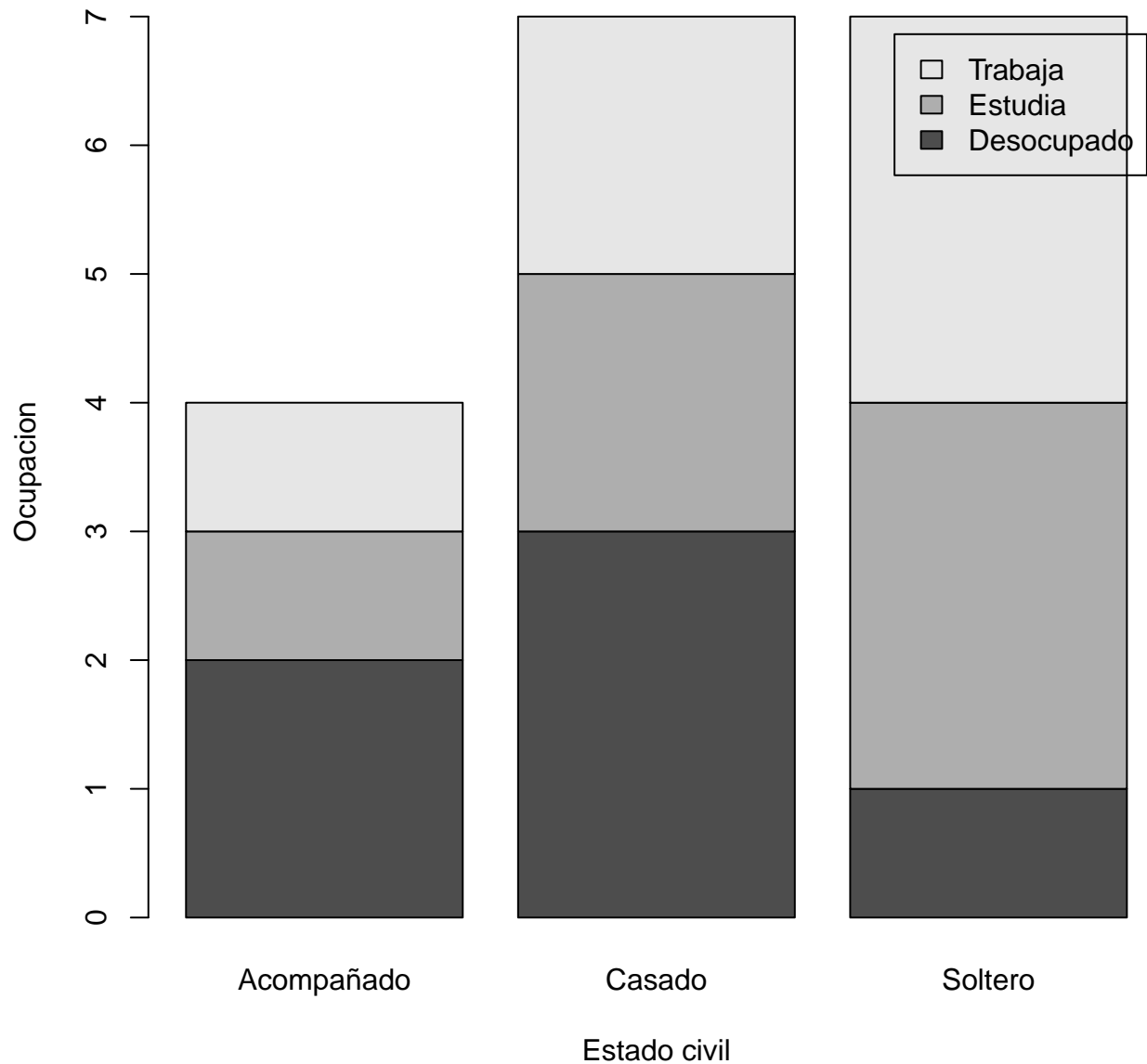
suma.columnas <- apply(tablaCont, 2, sum);
suma.columnas

## Desocupado      Estudia      Trabaja
##          6          6          6

# 2 indica que son totales por columna
# Gr\ 'aficos de barras para tabla de contingencia.
# Barras apiladas

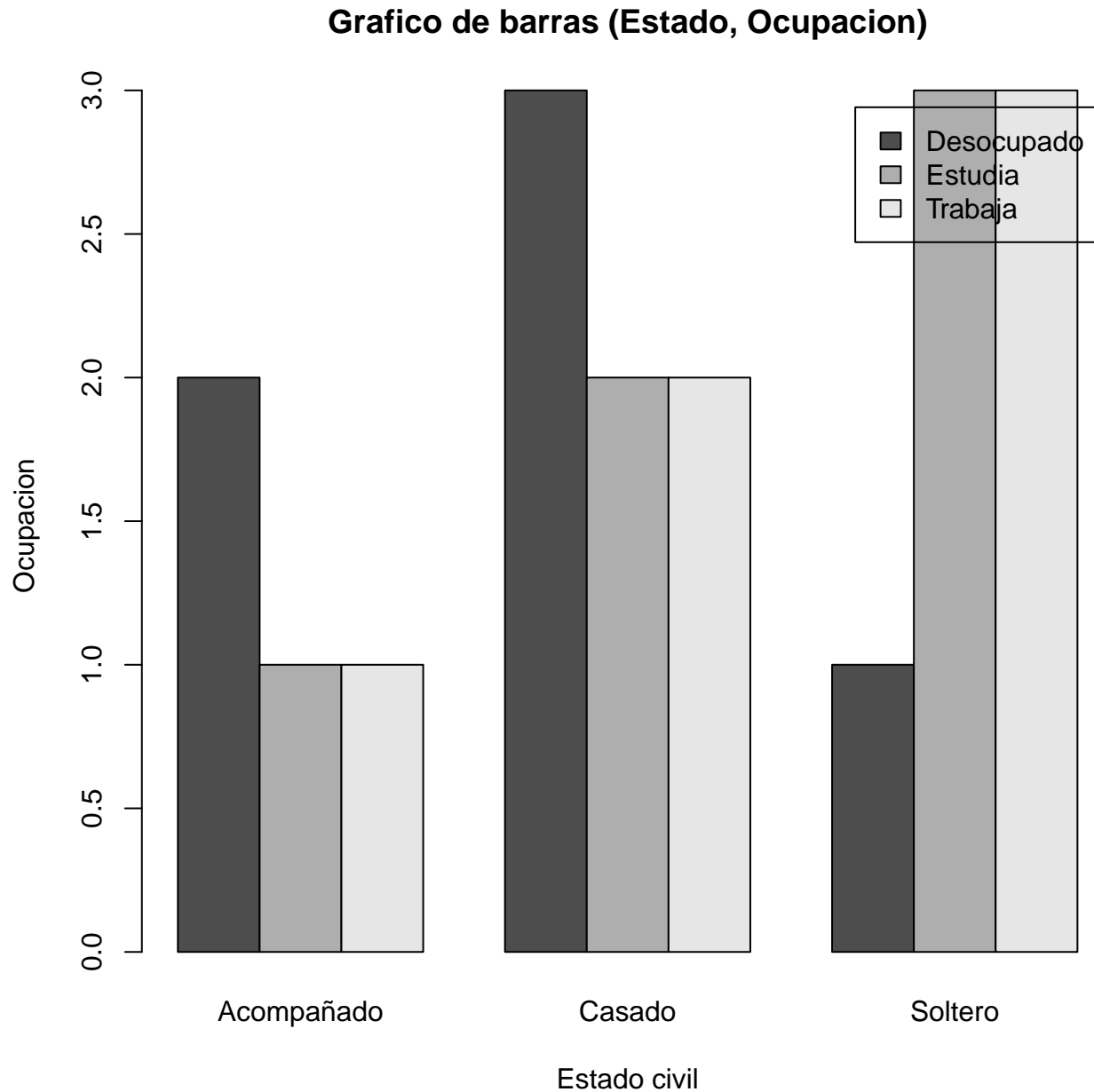
barplot(t(tablaCont), main="Grafico de barras (Estado, Ocupacion)",
        xlab="Estado civil",
        ylab="Ocupacion", legend.text=TRUE)

```

Grafico de barras (Estado, Ocupacion)

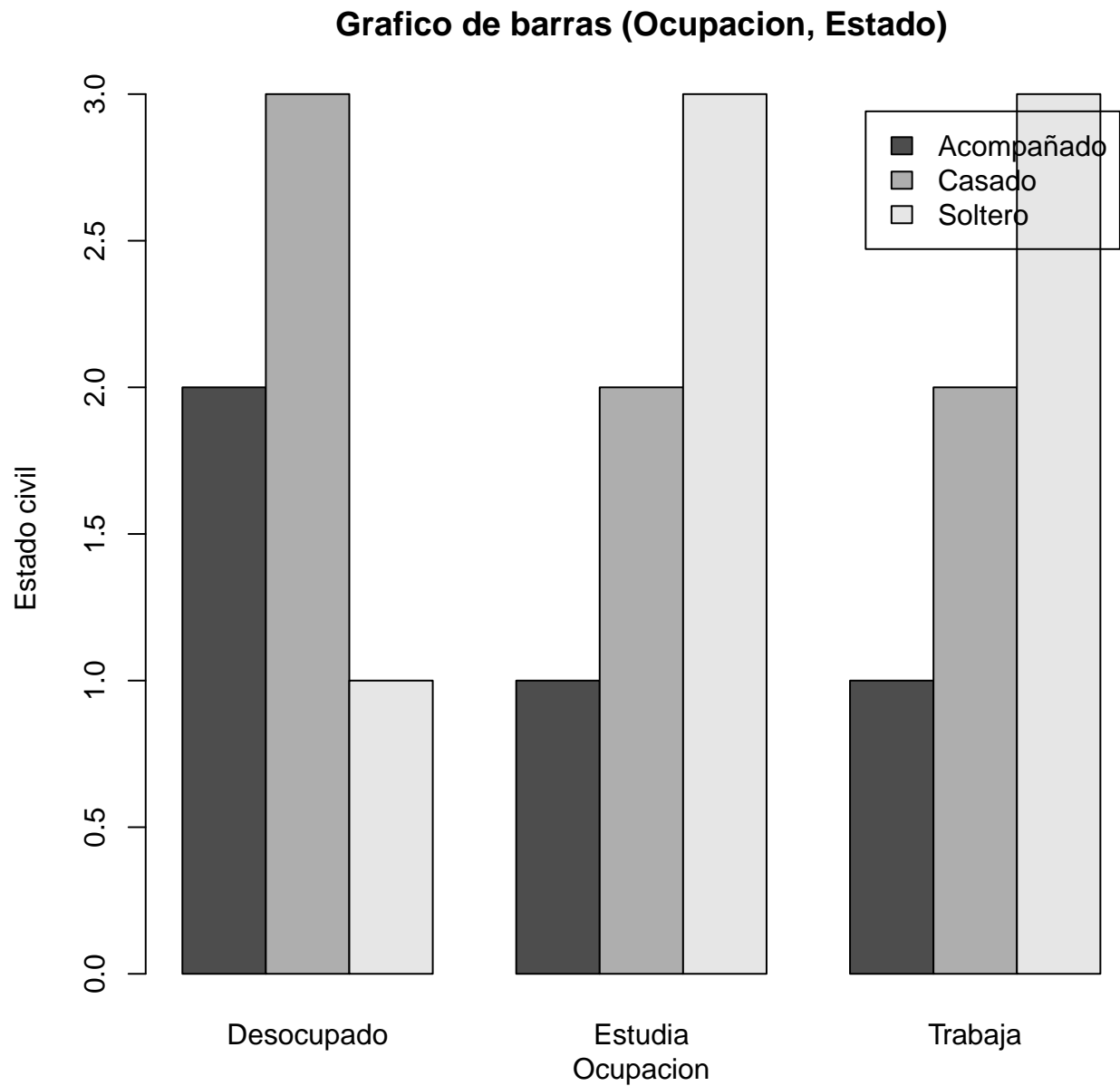
```
# Note que t(tablaCont) indica que las barras representan el
# Estado civil de los encuestados y que \'estas se subdividen en
# cada una de las diferentes ocupaciones consideradas.
# En caso de usar \'unicamente tablaCont; las barras representar\'on
# las diferentes ocupaciones y ?stas estar\'an 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 instrucci\ 'on beside =TRUE, indica que por cada una de
# las diferentes ocupaciones se crear? una barra para cada estado civil.
# Note que al usar beside =FALSE se obtiene el mismo gr\ 'afico de la
# instrucci\ 'on anterior.
```

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



8) Calcula tablas de proporciones o de probabilidades.

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

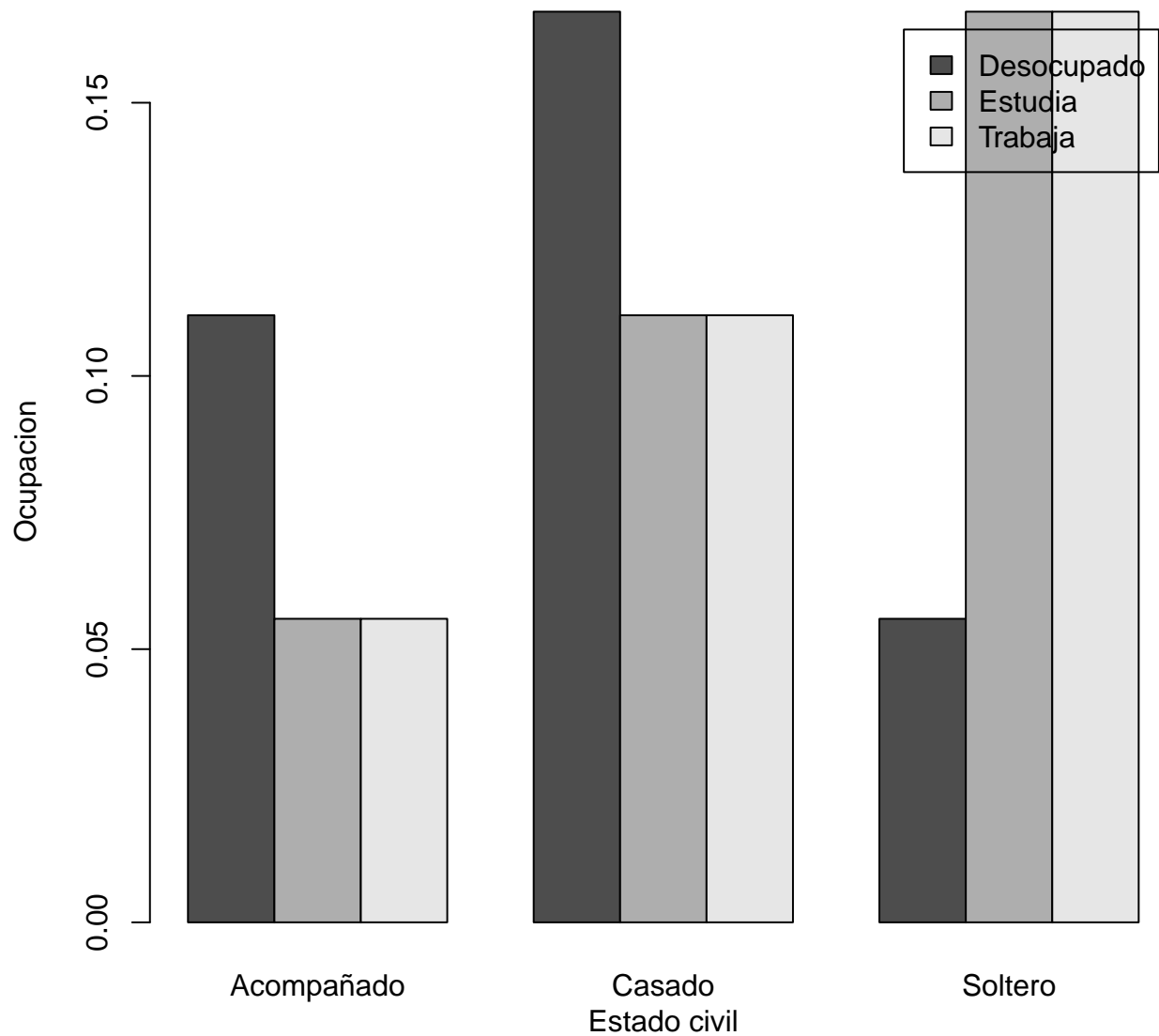
## $digits
## [1] 3

# Proporciones basadas en el total de la muestra, la suma de filas
# y columnas suman 1.
propTotal <- prop.table(tablaCont); propTotal
```

```
##          OCUPACIÓN\N
## 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)
```

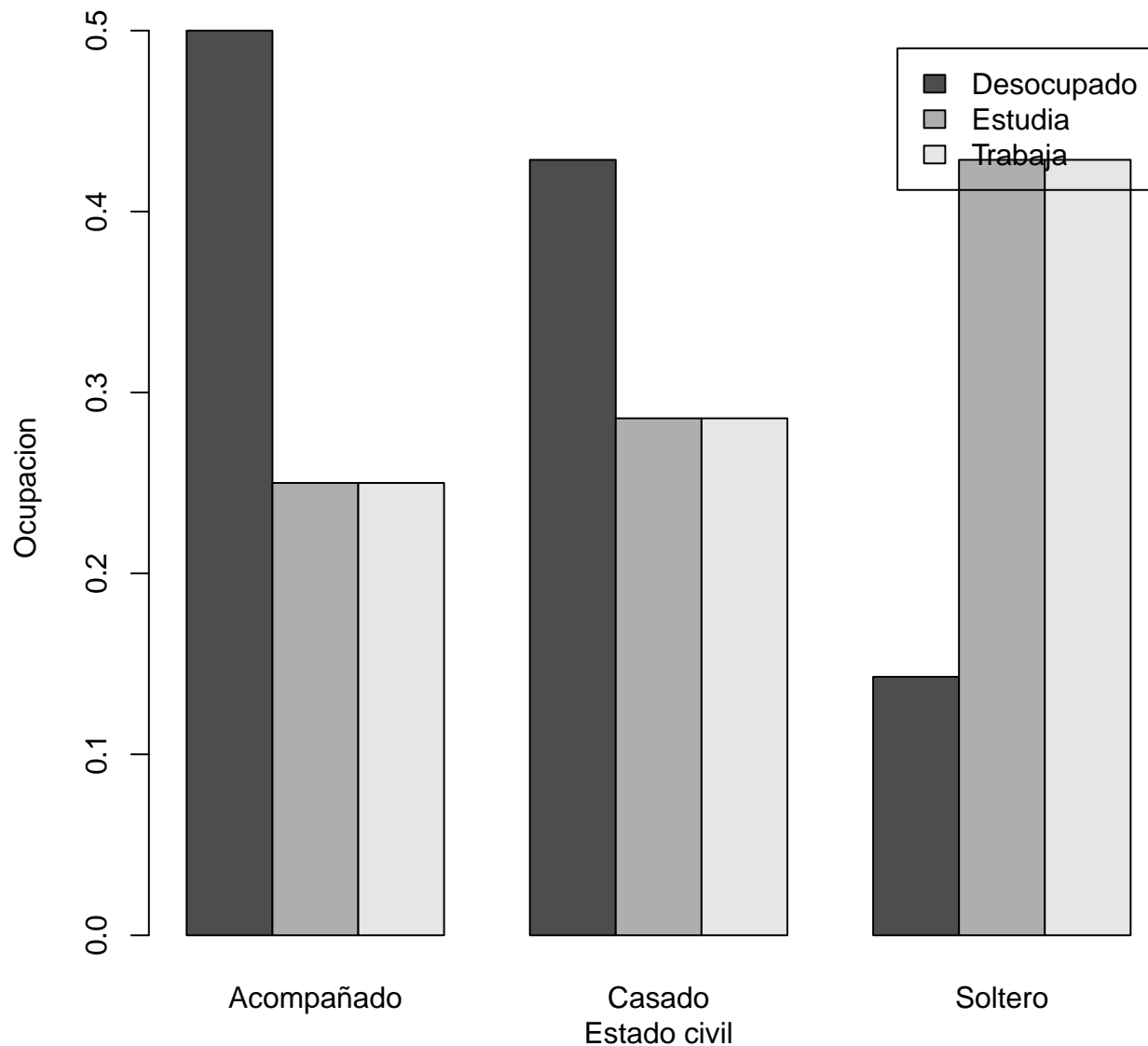
Grafico de barras (Estado, Ocupacion)




```
# Proporciones basadas en el total por fila, cada fila suma 1.
propFila <- prop.table(tablaCont, 1);
propFila

##                OCUPACIÃ\\N
## 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
```

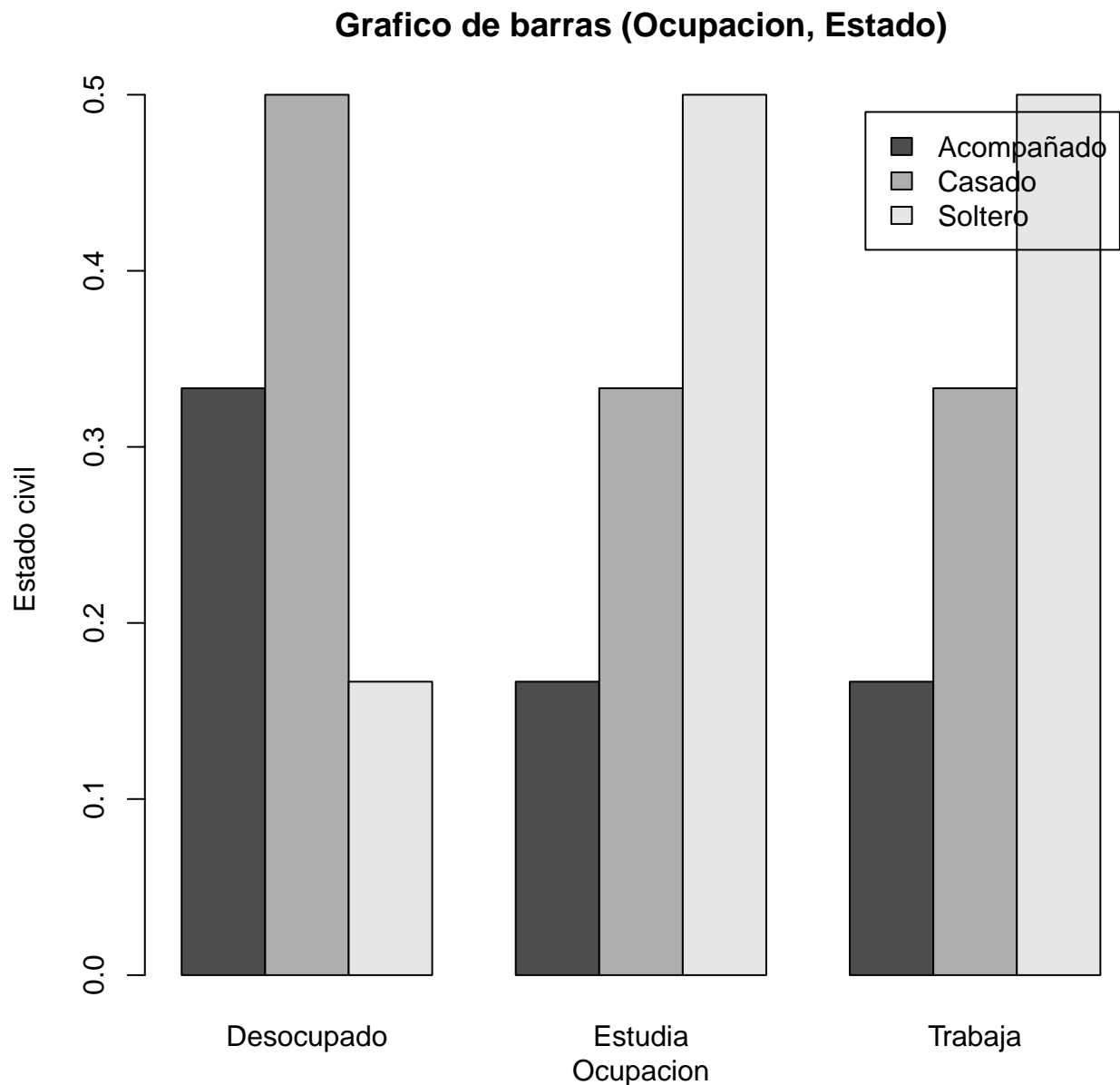
```
##          OCUPACIÓN\N
## ESTADO      Desocupado Estudia Trabaja
## Acompañado      0,333   0,167   0,167
## Casado           0.500   0.333   0.333
## Soltero          0.167   0.500   0.500
```

```
# 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)

```



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

## Error in table(Ocupacion, Estado): objeto 'Ocupacion' no encontrado

```

```

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

## Error in table(Estado, Ocupacion): objeto 'Estado' no encontrado

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"))

## Error in table(Estado, Ocupacion): objeto 'Estado' no encontrado

# Note que se puede definir a conveniencia la leyenda que se desea incorporar
# en el gráfico con la instrucción 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 #  $f_{ij} = f_{i.}/No. \text{ column}$ 

##
## OCUPACIÓN\N
## 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

```