

Tema2_Ejercicio_02

Fran Camacho

2024-12-16

Tema2 - Ejercicio 2

Cargar datos del paquete ISLR

(Y el paquete gmodels, para las tablas cruzadas. <- no usadas al final).

```
install.packages("ISLR")
```

```
## The following package(s) will be installed:
## - ISLR [1.4]
## These packages will be installed into "~/git/masterML/renv/library/linux-debian-bookworm/R-4.4/x86_64
##
## # Installing packages -----
## - Installing ISLR ... OK [linked from cache]
## Successfully installed 1 package in 9.4 milliseconds.
```

```
library("ISLR")
#install.packages("gmodels")
#library("gmodels")
```

Examinar su contenido (estructura y algunos registros) del dataset Wage:

```
#Structure
str(Wage)
```

```
## 'data.frame': 3000 obs. of 11 variables:
## $ year : int 2006 2004 2003 2003 2005 2008 2009 2008 2006 2004 ...
## $ age : int 18 24 45 43 50 54 44 30 41 52 ...
## $ maritl : Factor w/ 5 levels "1. Never Married",...: 1 1 2 2 4 2 2 1 1 2 ...
## $ race : Factor w/ 4 levels "1. White", "2. Black",...: 1 1 1 3 1 1 4 3 2 1 ...
## $ education : Factor w/ 5 levels "1. < HS Grad",...: 1 4 3 4 2 4 3 3 3 2 ...
## $ region : Factor w/ 9 levels "1. New England",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ jobclass : Factor w/ 2 levels "1. Industrial",...: 1 2 1 2 2 2 1 2 2 2 ...
## $ health : Factor w/ 2 levels "1. <=Good", "2. >=Very Good": 1 2 1 2 1 2 2 1 2 2 ...
## $ health_ins: Factor w/ 2 levels "1. Yes", "2. No": 2 2 1 1 1 1 1 1 1 1 ...
## $ logwage : num 4.32 4.26 4.88 5.04 4.32 ...
## $ wage : num 75 70.5 131 154.7 75 ...
```

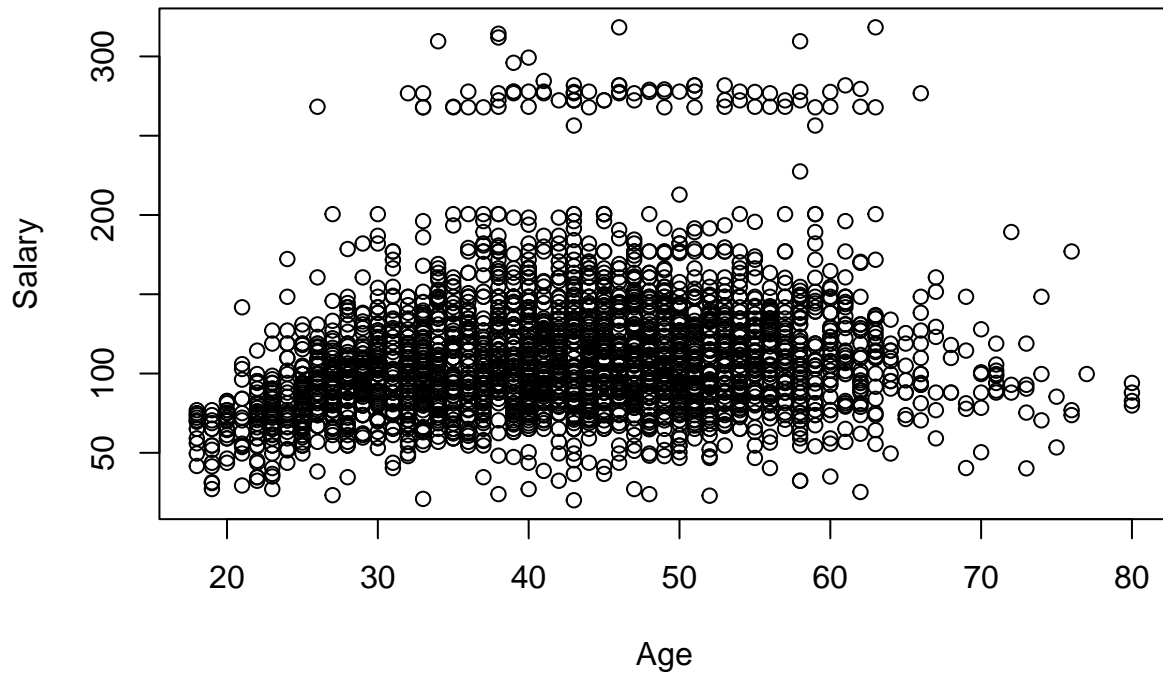
```
#Have a look at some data
head(Wage,10)
```

```
##      year age      maritl      race      education      region
## 231655 2006  18 1. Never Married 1. White      1. < HS Grad 2. Middle Atlantic
## 86582  2004  24 1. Never Married 1. White 4. College Grad 2. Middle Atlantic
## 161300 2003  45      2. Married 1. White 3. Some College 2. Middle Atlantic
## 155159 2003  43      2. Married 3. Asian 4. College Grad 2. Middle Atlantic
## 11443  2005  50      4. Divorced 1. White      2. HS Grad 2. Middle Atlantic
## 376662 2008  54      2. Married 1. White 4. College Grad 2. Middle Atlantic
## 450601 2009  44      2. Married 4. Other 3. Some College 2. Middle Atlantic
## 377954 2008  30 1. Never Married 3. Asian 3. Some College 2. Middle Atlantic
## 228963 2006  41 1. Never Married 2. Black 3. Some College 2. Middle Atlantic
## 81404  2004  52      2. Married 1. White      2. HS Grad 2. Middle Atlantic
##      jobclass      health health_ins logwage      wage
## 231655 1. Industrial      1. <=Good      2. No 4.318063 75.04315
## 86582  2. Information 2. >=Very Good      2. No 4.255273 70.47602
## 161300 1. Industrial      1. <=Good      1. Yes 4.875061 130.98218
## 155159 2. Information 2. >=Very Good      1. Yes 5.041393 154.68529
## 11443  2. Information      1. <=Good      1. Yes 4.318063 75.04315
## 376662 2. Information 2. >=Very Good      1. Yes 4.845098 127.11574
## 450601 1. Industrial 2. >=Very Good      1. Yes 5.133021 169.52854
## 377954 2. Information      1. <=Good      1. Yes 4.716003 111.72085
## 228963 2. Information 2. >=Very Good      1. Yes 4.778151 118.88436
## 81404  2. Information 2. >=Very Good      1. Yes 4.857332 128.68049
```

Relación entre edad y salario:

```
plot(x = Wage$age, y = Wage$wage, main = "Scatterplot of Age vs. Salary",
     xlab = "Age",
     ylab = "Salary")
```

Scatterplot of Age vs. Salary



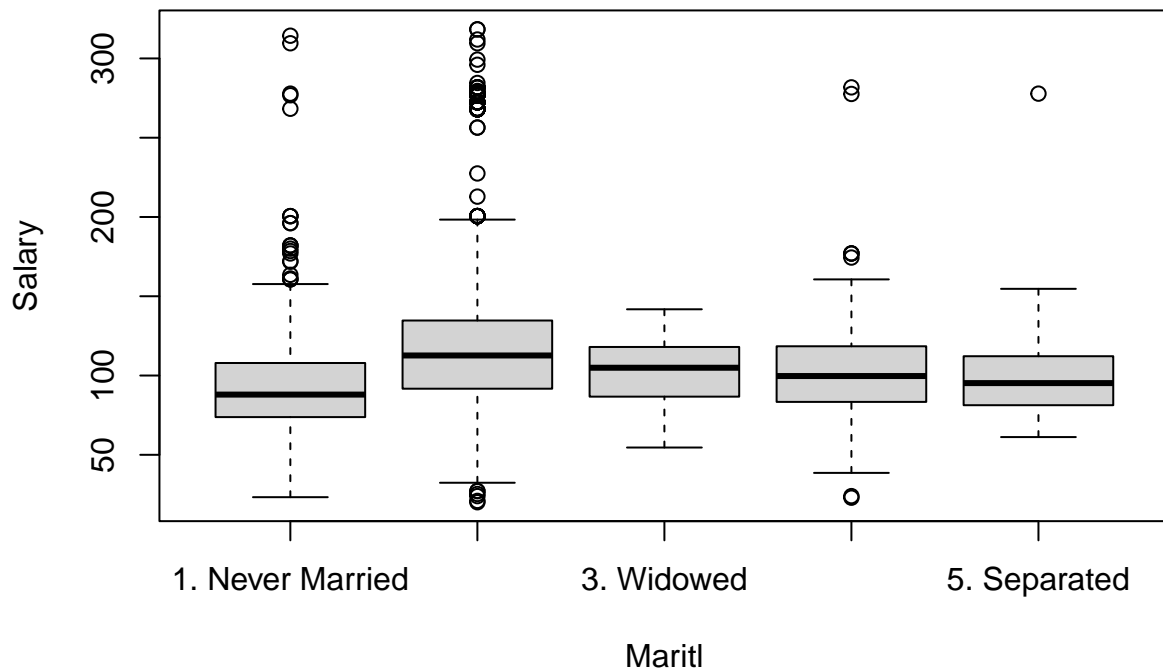
No se aprecia una influencia importante de la edad en el salario (se ve que los trabajadores más jóvenes tienen salarios más bajos).

El resto de variables que pueden influir en el salario son categóricas. Si utilizamos el mismo diagrama para intentar relacionar el estado civil con el salario, se obtiene lo siguiente:

Relación entre estado civil y salario

```
plot(x = Wage$maritl, y = Wage$wage, main = "Scatterplot of Maritl vs. Salary",  
xlab = "Maritl",  
ylab = "Salary")
```

Scatterplot of Maritl vs. Salary



Se puede ver que los casados tienen mejores sueldos, y los nunca casados los peores. Esto es coherente con la relación entre edad y salario: hay en general menos casados entre los hombres más jóvenes.

Relación entre estado civil y salario usando tablas cruzadas

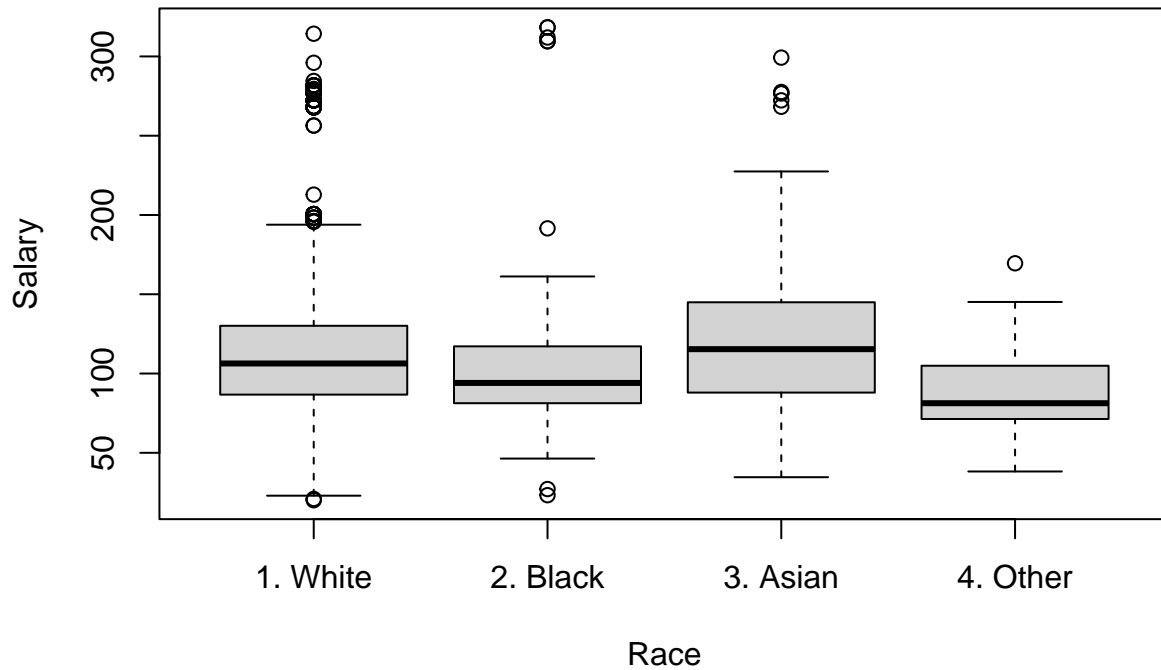
No apropiada para variables numéricas, las dos variables tienen que ser categóricas para utilizar tablas cruzadas.

```
#CrossTable(x = Wage$maritl, y = Wage$wage)
```

Relación entre raza y salario

```
plot(x = Wage$race, y = Wage$wage, main = "Scatterplot of Race vs. Salary",  
     xlab = "Race",  
     ylab = "Salary")
```

Scatterplot of Race vs. Salary

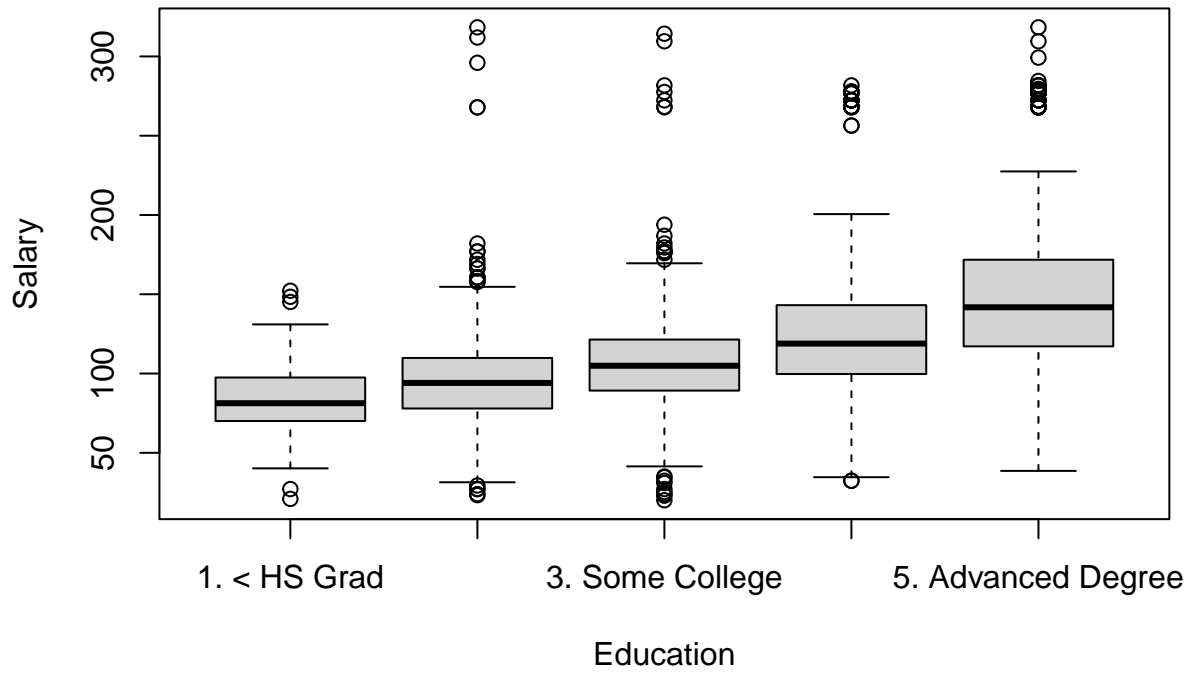


Se puede apreciar como en este dataset los trabajadores de origen asiático tienen en general mejores sueldos, pero como en cambio entre las personas blancas se dan más casos de salarios atípicamente altos.

Relación entre educación y salario

```
plot(x = Wage$education, y = Wage$wage, main = "Scatterplot of Education vs. Salary",  
xlab = "Education",  
ylab = "Salary")
```

Scatterplot of Education vs. Salary



Aquí se puede ver claramente una correlación positiva entre educación y salario: a mejor educación, mejor salario.