

Visualización de datos

Importación de datos:

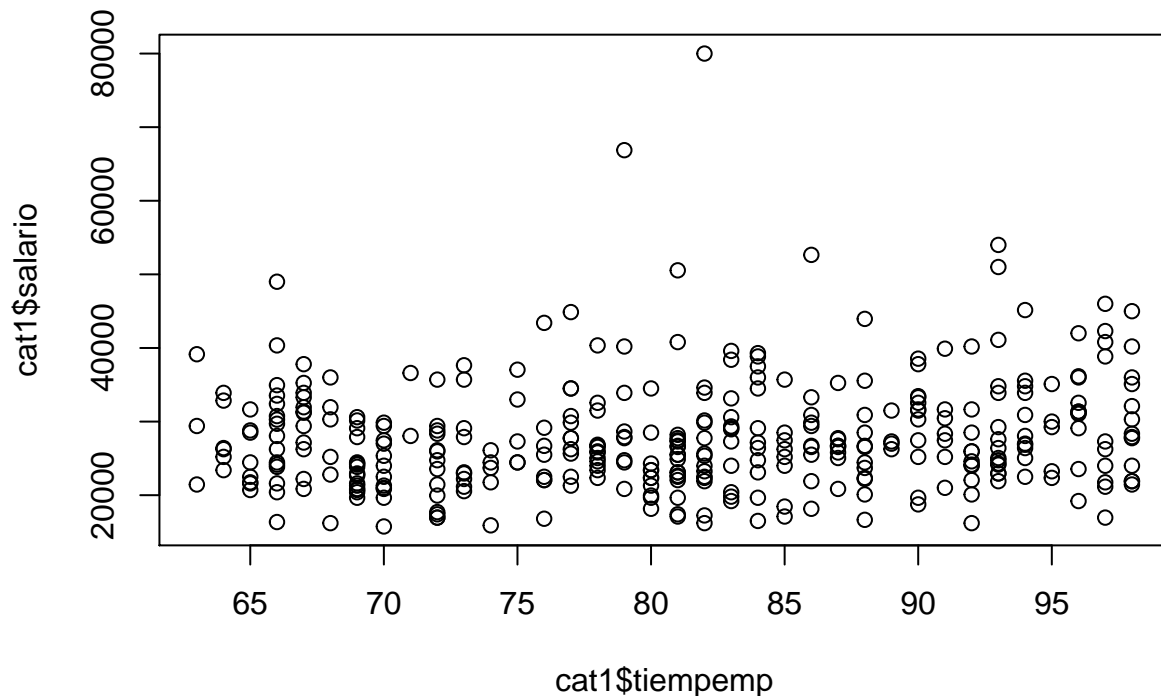
```
setwd("~/R/introstatsconr")
library(haven)
empleados <- read_sav("instrostat-v2/data/EMPLEADOS.sav")

head(empleados)

## # A tibble: 6 x 17
##   id sexo  fechnac      educ  catlab salario salini tiempemp expprev
##   <dbl> <chr+1> <date>      <dbl+1> <dbl+1> <dbl+1> <dbl+> <dbl+1b> <dbl+1>
## 1     1 h [Hom~ 1952-02-03 15 [15] 3 [Dir~   57000  27000     98    144
## 2     2 h [Hom~ 1958-05-23 16 [16] 1 [Adm~   40200  18750     98     36
## 3     3 h [Hom~ 1955-02-09 15 [15] 1 [Adm~   45000  21000     98    138
## 4     4 h [Hom~ 1958-08-22 15 [15] 1 [Adm~   32100  13500     98     67
## 5     5 h [Hom~ 1956-04-26 15 [15] 1 [Adm~   36000  18750     98    114
## 6     6 15 h [Hom~ 1962-08-29 12 [12] 1 [Adm~   27300  13500     97     66
## # ... with 8 more variables: minoría <dbl+1b1>, salinico <dbl>,
## #   sexo_rec <dbl+1b1>, PRE_1 <dbl>, RES_1 <dbl>, PRE_2 <dbl>,
## #   RES_2 <dbl>, `filter_$` <dbl+1b1>
```

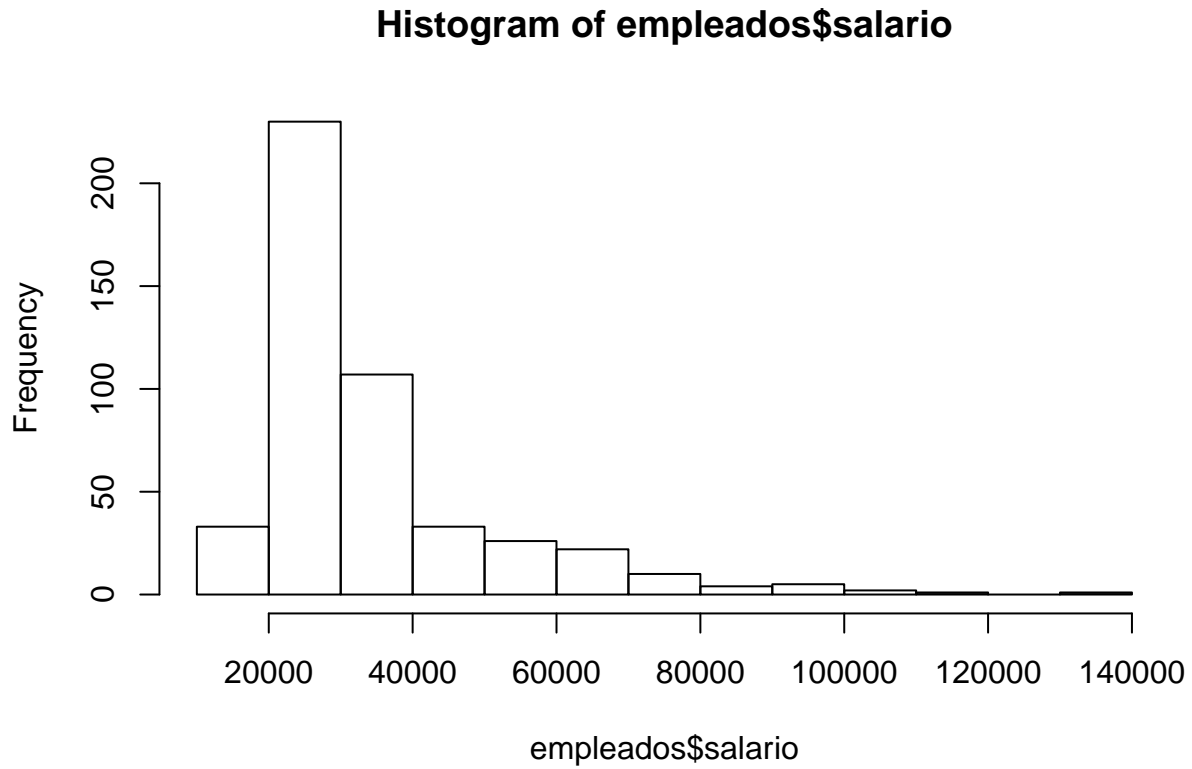
Ahora vamos a hacer un gráfico de dispersión para analizar la relación entre la experiencia previa y el salario inicial para trabajadores de categoría 1.

```
cat1 <- subset(empleados, catlab==1)
plot(cat1$tiempemp, cat1$salario)
```



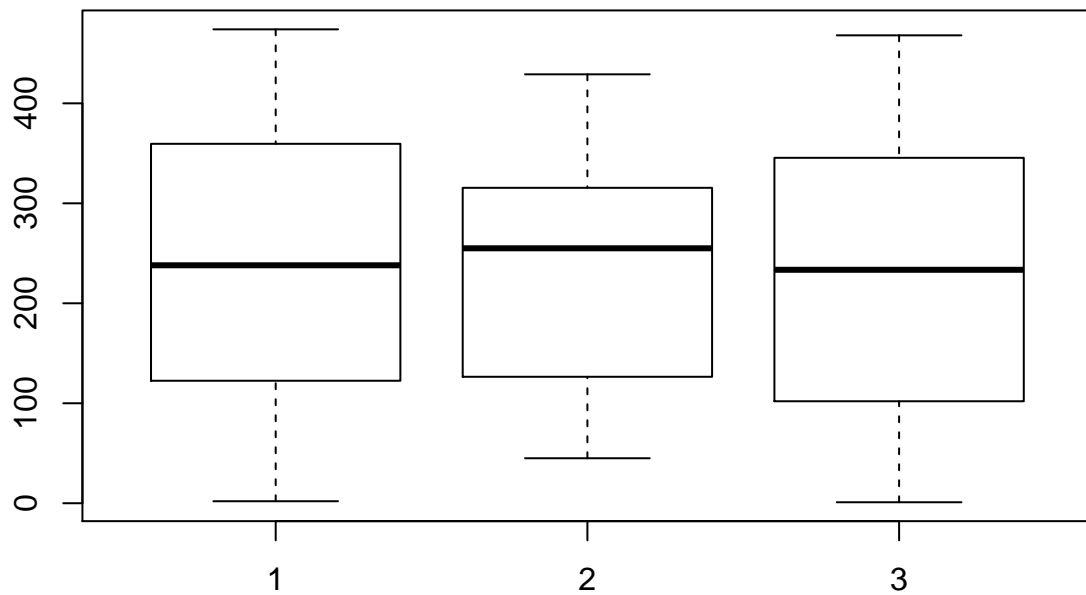
Los histogramas son representaciones gráficas muy útiles para analizar distribuciones. En este caso veremos cómo se distribuye el salario actual según el número de empleados:

```
hist(empleados$salario)
```



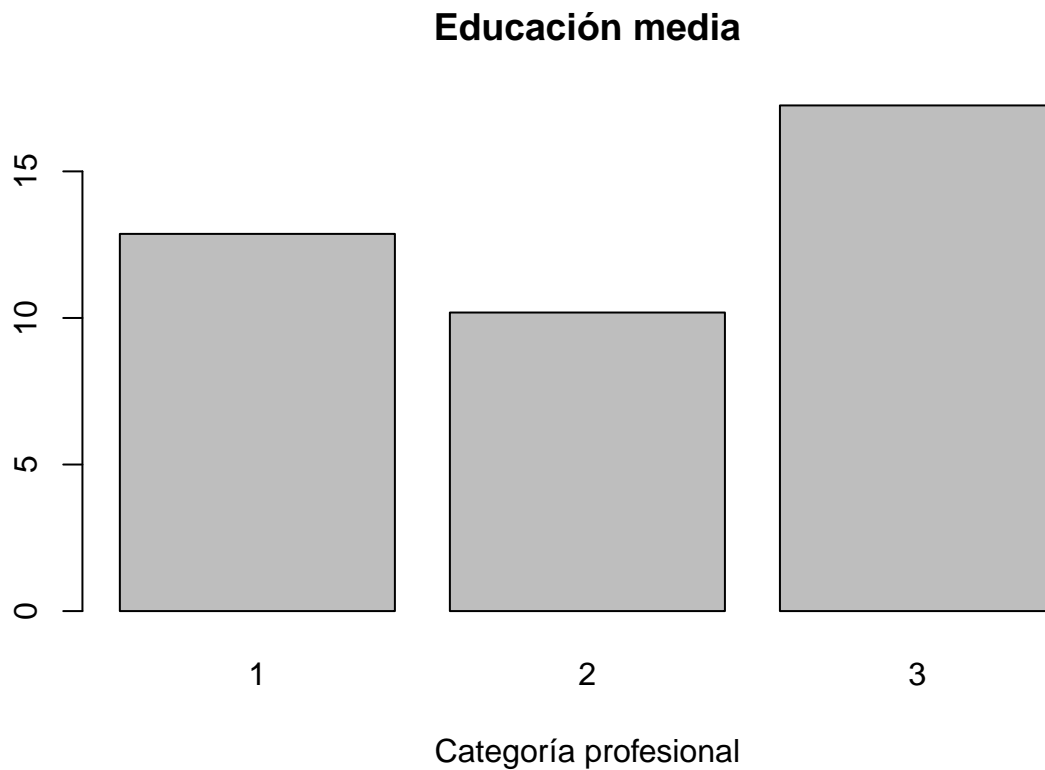
Otra forma de representar distribuciones es empleando gráficos de caja y bigotes.

```
boxplot(id~catlab, data = empleados)
```



Finalmente, vamos a analizar el nivel educativo por categoría del empleado utilizando un gráfico de barras

```
mean.educ <- aggregate(educ~catlab, data = empleados, mean)
mean.educ <- mean.educ[,2]
names(mean.educ) <- c("1", "2", "3")
barplot(mean.educ, main="Educación media",
        xlab = "Categoría profesional")
```



ggplot2

Lo primero que hay que hacer es instalar el paquete `ggplot2`.

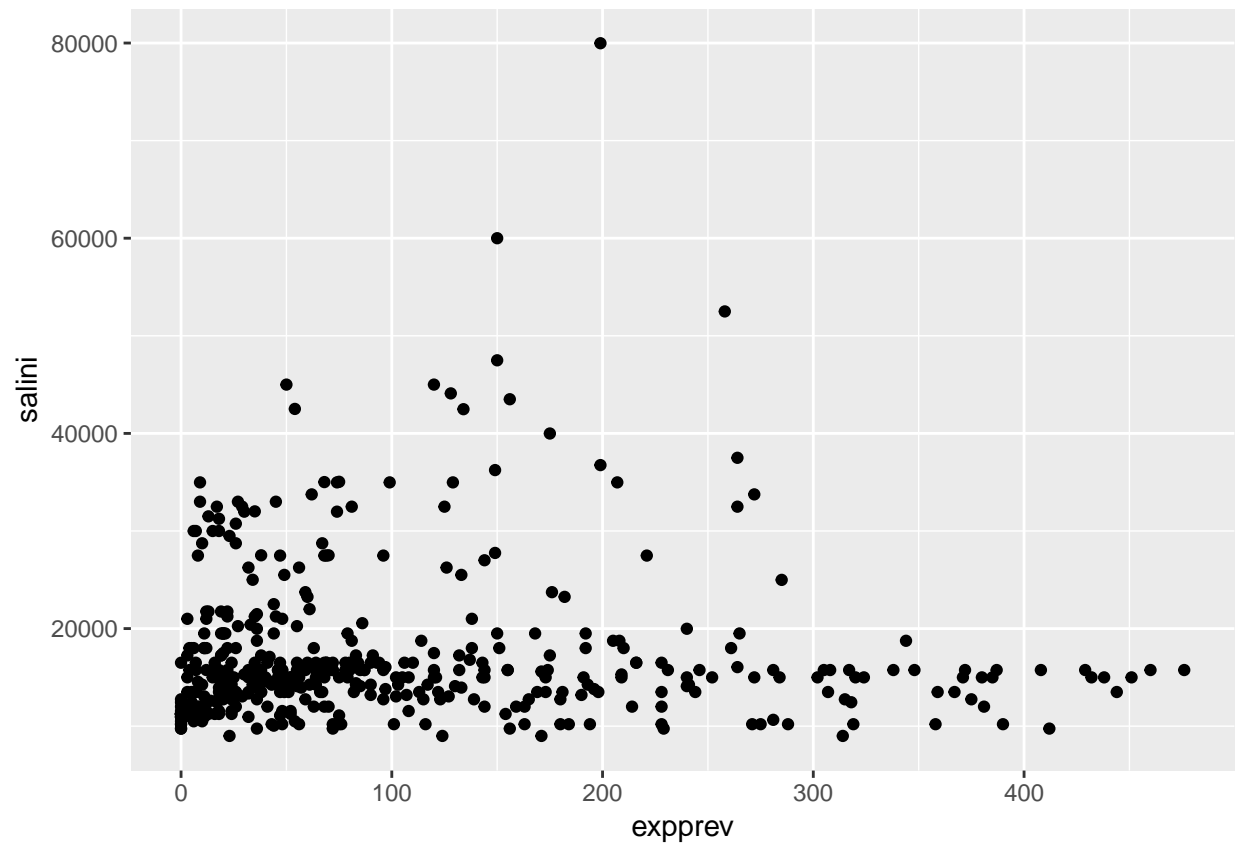
Ahora vamos a hacer nuestro primer gráfico de dispersión:

```
library(ggplot2)

# Error por el tipo de clase que es cada variable, hay que convertirlas
#ggplot(empleados) +
# geom_point(aes(x= expprev, y= salini))

empleados2 <- empleados
empleados2$expprev <- as.numeric(empleados2$expprev)
empleados2$salini <- as.numeric(empleados2$salini)

ggplot(empleados2) +
  geom_point(aes(x= expprev, y= salini))
```



Podemos seguir añadiendo capas y variables para mostrar más información en la gráfica

```
empleados2$catlab <- as.factor(empleados2$catlab)
empleados2$sexo <- as.factor(empleados2$sexo)

ggplot(empleados2) +
  geom_point(aes(x=expprev, y=salini, color=sexo, size=catlab))
```

```
## Warning: Using size for a discrete variable is not advised.
```

