

## Ejercicio con data frames

En un fichero llamado *muestra.RData* se tienen los datos de una muestra de personas referentes a las siguientes variables:

- **Sexo**, codificada como 1 Hombre y 6 Mujer
- **Edad**
- **CPMN**, municipio de nacimiento
- **Altura** en metros
- **Peso** en Kg
- **Situacion**. codificada como 1 Ocupada, 2 Parada y 3 Inactiva

1. Cargar el fichero en Rstudio.

```
load ("muestra.RData")
```

2. Crear una nueva variable denominada *fsexo* que sea un factor de la variable Sexo con la codificación adecuada. (1 Hombre y 6 Mujer)

```
factor (muestra$Sexo) -> fsexo  
levels(fsexo) <- c('Hombre', 'Mujer')  
cbind(muestra, fsexo) -> muestra
```

3. ¿Cuántas personas hay en la muestra? ¿Cuántos hombres y mujeres hay?

```
length(muestra$Sexo)  
[1] 797  
  
length(muestra$fsexo [muestra$fsexo == 'Hombre'])  
[1] 368  
  
length(muestra$fsexo [muestra$fsexo == 'Mujer'])  
[1] 429
```

4. Convertir la variable Altura en centímetros (en la misma variable).

```
muestra$Altura <- muestra$Altura*100
```

5. Calcular el peso y la altura media de las personas por sexo (usando el factor creado) en una misma orden.

```
aggregate(cbind(Altura, Peso)~fsexo, FUN=mean, data=muestra)
```

6. Crear un subconjunto con las personas paradas que pesen más de 75 Kg.

```
subset(muestra, Situacion==2&Peso>75, 1:7)
```

7. Crear una nueva variable denominada *fsituacion* que sea un factor de la variable Trabaja con la codificación adecuada. (1 Ocupada, 2 Parada y 3 Inactiva).

```
factor (muestra$Situacion) -> fsituacion  
levels (fsituacion) <- c('Ocupada', 'Parada', 'Inactiva')  
cbind(muestra, fsituacion) -> muestra
```

8. Hacer una tabla de frecuencias conjunta de las variables fsexo y ftrabaja.

```
table(muestra$fsexo,muestra$fsituacion)
```

9. ¿Cuál es el municipio de nacimiento más frecuente entre las personas de la muestra?

```
tail(names(sort(table(muestra$CPMN))), 1)
```