

Data Frames

Code ▾

Vamos a descargar de la pagina de UCI Machine Learning repository la base de datos de Adults.

Lo primero que hay que verificar es nuestro directorio actual de trabajo y modificarlo si necesario.

Hide

```
getwd() # directorio actual
```

```
[1] "C:/Users/mbtec/Documents/GitHub/PropedeuticoDataScience2017/CuadernosR/2_realdata"
```

Con el comando `setwd` podemos cambiar el directorio de trabajo. Por ejemplo

Hide

```
setwd("C:/Users/mbtec/Documents/GitHub/PropedeuticoDataScience2017/CuadernosR/Cuaderno2")
```

```
Error in setwd("C:/Users/mbtec/Documents/GitHub/PropedeuticoDataScience2017/CuadernosR/Cuaderno2") :  
  cannot change working directory
```

En RStudio la maneta facil tambien es dar click en los 3 puntos ... que aparecen en la ventana de Files, navegar al destino, y luego dar click en `More` y seleccionar la opcion

`Set As Working Directory`. Alternativamente (mas facil) en el menu principal en `Session`. Otra opcion es trabajar siempre con proyectos que “fijan” el working directory. El working directory lo pueden cambiar cuantas veces quieran.

Bajar los datos

Hide

```
download.file("https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data", "adultsdata.csv")
```

```
trying URL 'https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data'  
Content type 'text/plain; charset=UTF-8' length 3974305 bytes (3.8 MB)  
downloaded 3.8 MB
```

Alternativamente se puede descargar manualmente. Ahora hay que leerlo en R.

Hide

```
adults <- read.csv("adultsdata.csv", header = FALSE)
head(adults) # imprime las primeras líneas
```

Podemos acceder a las variables por número de columna o por nombre. En este caso como usamos `header = FALSE` automáticamente R eligió los nombres V1, V2, etc.

Manualmente podemos elegir nombres

Hide

```
names(adults) <- c("age", "workclass", "fnlwgt", "education",
"education_num", "marital_status", "occupation", "relationship",
"race", "sex", "capital_gain", "capital_loss", "hours_per_week", "native_country", "unemployed")
head(adults)
```