

# Lectura de Datos desde un fichero

Welton Vieira dos Santos

28/2/2020

## Leyendo tablas de datos

- **read.table()**: para definir un data frame a partir de una tabla de datos contenida en un fichero
  - Este fichero puede estar guardado en nuestro ordenador o bien podemos conocer su URL. Sea cual sea el caso, se aplica la función al nombre del fichero o la dirección entre comillas

## Carga de ficheros local

Como ejemplo, vamos a cargar un fichero de datos llamado *bulls.dat* que se encuentra en la carpeta *datos* de ese ambiente de trabajo.

```
#Cargar la tabla del fichero bulls.dat a la variable datos
datos <- read.table("../data/bulls.dat")
```

```
#Para visualizar el contenido de la variable datos
head(datos, 10)#Las 10 primeras filas
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
1	1	2200	51.0	1128	70.9	7	0.25	54.8	1720
2	1	2250	51.9	1108	72.1	7	0.25	55.3	1575
3	1	1625	49.9	1011	71.6	6	0.15	53.1	1410
4	1	4600	53.1	993	68.9	8	0.35	56.4	1595
5	1	2150	51.2	996	68.6	7	0.25	55.0	1488
6	1	1225	49.2	985	71.4	6	0.15	51.4	1500
7	1	2250	51.0	959	72.1	7	0.20	54.0	1522
8	1	4000	51.5	1060	69.3	7	0.30	55.6	1765
9	1	1600	50.1	979	71.2	6	0.25	51.5	1365
10	1	1525	49.6	1083	75.8	6	0.30	54.6	1640

```
tail(datos, 10)#Las 10 últimas filas
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
67	8	1550	51.2	1057	74.8	7	0.10	55.5	1520
68	8	1250	50.8	1040	74.5	6	0.10	55.8	1516
69	8	1350	52.7	1079	75.5	7	0.15	56.1	1595
70	8	1725	51.4	1034	71.2	7	0.10	56.0	1655
71	8	1750	50.7	1012	71.6	6	0.10	54.3	1480
72	8	1450	51.4	997	73.4	7	0.10	55.2	1454
73	8	1200	49.8	991	70.8	6	0.15	54.6	1475
74	8	1425	50.0	928	70.8	6	0.10	53.9	1375
75	8	1250	50.1	990	71.0	6	0.10	54.9	1564
76	8	1500	51.7	992	70.6	7	0.15	55.1	1458

## Parámetros de la función `read.table()`

- **header = TRUE**: para indicar si la tabla que importamos tiene una primera fila con los nombres de las columnas. El valor por defecto es **FALSE**
- **col.names = c(...)**: para especificar el nombre de las columnas. No olvidar que cada nombre debe ir entre comillas.
- **sep**: para especificar las separaciones entre columnas en el fichero (si no es un espacio en blanco). Si es así, hay que introducir el parámetro pertinente entre comillas. Para verificar los tipos de separación que proporciona R utilice la función `?read.tables`
- **dec**: para especificar el signo que separa la parte entera de la decimal (si no es un punto. Si es así, hay que introducir el parámetro pertinente entre comillas)

```
df <- read.table("../data/bulls.dat",
                 header = FALSE,
                 col.names = c("breed", "sale_price", "shoulder",
                              "fat_free", "percent_ff", "frame_scale",
                              "back_fat", "sale_height", "sale_weight"))
names(df)

[1] "breed"      "sale_price" "shoulder"   "fat_free"   "percent_ff"
[6] "frame_scale" "back_fat"    "sale_height" "sale_weight"
```

Para verificar que los datos están bien cargados, utilizar la función para ver la estructura del data frame

```
str(df)

'data.frame': 76 obs. of 9 variables:
 $ breed      : int  1 1 1 1 1 1 1 1 1 1 ...
 $ sale_price : int  2200 2250 1625 4600 2150 1225 2250 4000 1600 1525 ...
 $ shoulder   : num  51 51.9 49.9 53.1 51.2 49.2 51 51.5 50.1 49.6 ...
 $ fat_free   : int  1128 1108 1011 993 996 985 959 1060 979 1083 ...
 $ percent_ff : num  70.9 72.1 71.6 68.9 68.6 71.4 72.1 69.3 71.2 75.8 ...
 $ frame_scale: int  7 7 6 8 7 6 7 7 6 6 ...
 $ back_fat   : num  0.25 0.25 0.15 0.35 0.25 0.15 0.2 0.3 0.25 0.3 ...
 $ sale_height: num  54.8 55.3 53.1 56.4 55 51.4 54 55.6 51.5 54.6 ...
 $ sale_weight: int  1720 1575 1410 1595 1488 1500 1522 1765 1365 1640 ...
```