

Data Frame

Data Frames

Un **Data Frame** es una tabla de doble entrada, formada por variables en las columnas y observaciones de esas variables en las filas, de manera que cada fila contiene los valores de las variables para un mismo caso o un mismo individuo.

- **Data()**: para abrir una ventana con la lista de los objetos de datos a los que tenemos acceso en la sesión actual de R (los que lleva la instalación básica de R y los que aportan los paquetes que tengamos cargados).
 - Si entramos **data(package = .packages(all.available = TRUE))** obtendremos la lista de todos los objetos de datos a los que tenemos acceso, incluyendo los de los paquetes que tengamos instalados, pero que no estén cargados en la sesión actual.

Obteniendo información del data frame

- **head(DataFrame, n)**: para mostrar las n primeras filas del data frame. Por defecto se muestran las 6 primeras filas.
- **tail(DataFrame, n)**: para mostrar las n últimas filas del data frame. Por defecto se muestran las 6 últimas filas.
- **str(DataFrame)**: para conocer la estructura global de un data frame.
- **names(DataFrame)**: para producir un vector con los nombres de las columnas.
- **rownames(DataFrame)**: para producir un vector con los identificadores de las filas.
 - Recuerda siempre que estos identificadores son palabras, aunque sean números, de ahí que los imprima entre comillas.
- **colnames(DataFrame)**: para producir un vector con los identificadores de las columnas.
- **dimnames(DataFrame)**: para producir una lista formada por dos vectores (el de los identificadores de las filas y el de los nombres de las columnas).
- **nrow(DataFrame)**: para consultar el número de filas de una data frame.
- **ncol(DataFrame)**: para consultar el número de columnas de una data frame.
- **dim(DataFrame)**: para producir un vector con el número de filas y el de columnas de una data frame.
- **DataFrame.\$nombre_variable**: para obtener una columna concreta de un data frame
 - El resultado será un vector o un factor, según cómo esté definida la columna dentro del data frame.
 - Las variables de un data frame son internas, no están definidas en el entorno global de trabajo de R.
- **DataFrame[n, m]**: para extraer “trozos” del data frame por filas y columnas (funciona exactamente igual que en matrices) donde n y m pueden definirse como:
 - intervalos
 - condiciones

- números naturales
- no poner nada
- Si sólo queremos definir la subtabla quedándonos con algunas variables, basta aplicar el nombre del data frame al vector de variables.
- Estas construcciones se pueden usar también para reordenar las filas o columnas.

Data Frame de Iris

Añadir una base de datos de iris

```
df <- iris #Cargando los datos de Iris a la variable df
head(df, 5) #Leer df y sólo muestra los 5 primeros datos
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa

```
tail(df, 5) #Mostrar los 5 últimos datos de df
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
146	6.7	3.0	5.2	2.3	virginica
147	6.3	2.5	5.0	1.9	virginica
148	6.5	3.0	5.2	2.0	virginica
149	6.2	3.4	5.4	2.3	virginica
150	5.9	3.0	5.1	1.8	virginica

Para consultar la información básica del Data Frame **df**, es decir, sólo muestra el nombre de las columnas de la tabla.

```
names(df) #Para ver los nombres de las variables o columnas del Data Frame
```

```
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
```

Para ver su estructura:

```
str(df) #Para ver la estructura del Data Frame
```

```
'data.frame':  150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

Como se puede observar, la estructura tiene 3 vectores numéricos y un factor de 3 niveles.