

Data Frame

Data frames:

- **Data frames:** Tabla de doble entrada, formadas por variables en columnas y observaciones en filas.
- **data():** para abrir una ventana con la lista de los objetos de datos a los que tenemos acceso en la sesión actual de R.
- Si entramos a `data(package=.packages(all.available = TRUE))` obtendremos la lista de todos los objetos de datos a los que tenemos acceso.

Data Frame de Iris

```
df = iris
head(df, 5)
```

```
##   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)
```

```
##   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
```

Obteniendo información de un dataframe

- **head(df,n)** para mostrar las primeras n filas. Por defecto muestra 6
- **tail(df,n)** para mostrar las últimos n filas. Por defecto muestra 6
- **str(df)** muestra estructura global de un dataframe
- **names(df)** para producir un vector con los nombres de las columnas.

```
str(df)
```

```
## '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 ...
```

```
names(df)
```

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

Dataset Orange

```
str(Orange)
```

```
## Classes 'nfnGroupedData', 'nfGroupedData', 'groupedData' and 'data.frame': 35 obs. of 3 variables
## $ Tree : Ord.factor w/ 5 levels "3"<"1"<"5"<"2"<...: 2 2 2 2 2 2 2 4 4 4 ...
## $ age : num 118 484 664 1004 1231 ...
## $ circumference: num 30 58 87 115 120 142 145 33 69 111 ...
## - attr(*, "formula")=Class 'formula' language circumference ~ age | Tree
## .. ..- attr(*, ".Environment")=<environment: R_EmptyEnv>
## - attr(*, "labels")=List of 2
## ..$ x: chr "Time since December 31, 1968"
## ..$ y: chr "Trunk circumference"
## - attr(*, "units")=List of 2
## ..$ x: chr "(days)"
## ..$ y: chr "(mm)"
```

```
head(Orange, 5)
```

```
## Tree age circumference
## 1 1 118 30
## 2 1 484 58
## 3 1 664 87
## 4 1 1004 115
## 5 1 1231 120
```

```
tail(Orange, 5)
```

```
## Tree age circumference
## 31 5 664 81
## 32 5 1004 125
## 33 5 1231 142
## 34 5 1372 174
## 35 5 1582 177
```

Obteniendo información de dataframe

- **rownames(df)**: produce un vector con identificadores de las filas.
- **colnames(df)**: produce un vector con identificadores de las columnas
- **dimnames(df)**: produce una list formada por dos vectores (el de los identificadores de las filas y de los nombres de las columnas)
- **nrow(df)**: número de filas de un data frame
- **ncol(df)**: para consultar el número de columnas de un data frame
- **dim(df)**: para producir un vector con el número de filas y columnas.
- **df\$nombre_variable**: para obtener una columna concreta de un dataframe.
 - El resultado será un vector o factor, según como esté definida la columna dentro del data frame.
 - Las variables de un dataframe son internas, no están definidas en el entorno global de trabajo de R.

```
rownames(df)
```

```
## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11"
## [12] "12" "13" "14" "15" "16" "17" "18" "19" "20" "21" "22"
## [23] "23" "24" "25" "26" "27" "28" "29" "30" "31" "32" "33"
## [34] "34" "35" "36" "37" "38" "39" "40" "41" "42" "43" "44"
## [45] "45" "46" "47" "48" "49" "50" "51" "52" "53" "54" "55"
## [56] "56" "57" "58" "59" "60" "61" "62" "63" "64" "65" "66"
## [67] "67" "68" "69" "70" "71" "72" "73" "74" "75" "76" "77"
## [78] "78" "79" "80" "81" "82" "83" "84" "85" "86" "87" "88"
## [89] "89" "90" "91" "92" "93" "94" "95" "96" "97" "98" "99"
## [100] "100" "101" "102" "103" "104" "105" "106" "107" "108" "109" "110"
```

```
## [111] "111" "112" "113" "114" "115" "116" "117" "118" "119" "120" "121"
## [122] "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143"
## [144] "144" "145" "146" "147" "148" "149" "150"
```

```
colnames(df) # equivalente a names
```

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

```
dimnames(df)
```

```
## [[1]]
## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11"
## [12] "12" "13" "14" "15" "16" "17" "18" "19" "20" "21" "22"
## [23] "23" "24" "25" "26" "27" "28" "29" "30" "31" "32" "33"
## [34] "34" "35" "36" "37" "38" "39" "40" "41" "42" "43" "44"
## [45] "45" "46" "47" "48" "49" "50" "51" "52" "53" "54" "55"
## [56] "56" "57" "58" "59" "60" "61" "62" "63" "64" "65" "66"
## [67] "67" "68" "69" "70" "71" "72" "73" "74" "75" "76" "77"
## [78] "78" "79" "80" "81" "82" "83" "84" "85" "86" "87" "88"
## [89] "89" "90" "91" "92" "93" "94" "95" "96" "97" "98" "99"
## [100] "100" "101" "102" "103" "104" "105" "106" "107" "108" "109" "110"
## [111] "111" "112" "113" "114" "115" "116" "117" "118" "119" "120" "121"
## [122] "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143"
## [144] "144" "145" "146" "147" "148" "149" "150"
##
## [[2]]
## [1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width"
## [5] "Species"
```

```
nrow(df)
```

```
## [1] 150
```

```
ncol(df)
```

```
## [1] 5
```

```
dim(df)
```

```
## [1] 150 5
```

```
df$Species[1:10]
```

```
## [1] setosa setosa setosa setosa setosa setosa setosa setosa setosa setosa
## Levels: setosa versicolor virginica
```

```
df$Petal.Length[1:10] # Es un factor
```

```
## [1] 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5
```

Sub Data Frame

df[n,m]: Extraer trozos de dataframe por filas y columnas, donde n y m pueden definirse como: + intervalos + condiciones + números naturales + no poner nada + si solo queremos definir la subtabla quedandonos con algunas variables basta aplicar el nombre del data frame al vector de variables. + Estas construcciones se pueden usar tambien para reordenar las filas o columnas.

```
df[1:10,]
```

```
##      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
## 6          5.4         3.9         1.7         0.4   setosa
## 7          4.6         3.4         1.4         0.3   setosa
## 8          5.0         3.4         1.5         0.2   setosa
## 9          4.4         2.9         1.4         0.2   setosa
## 10         4.9         3.1         1.5         0.1   setosa
```

```
df[1:10, 2:4]
```

```
##      Sepal.Width Petal.Length Petal.Width
## 1          3.5         1.4         0.2
## 2          3.0         1.4         0.2
## 3          3.2         1.3         0.2
## 4          3.1         1.5         0.2
## 5          3.6         1.4         0.2
## 6          3.9         1.7         0.4
## 7          3.4         1.4         0.3
## 8          3.4         1.5         0.2
## 9          2.9         1.4         0.2
## 10         3.1         1.5         0.1
```

```
df[df$Species == "setosa" & df$Sepal.Width >4, ]
```

```
##      Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 16          5.7         4.4         1.5         0.4   setosa
## 33          5.2         4.1         1.5         0.1   setosa
## 34          5.5         4.2         1.4         0.2   setosa
```

```
df[df$Species == "setosa" & df$Sepal.Width >4, ][c(1,3), c(2,5)]
```

```
##      Sepal.Width Species
## 16          4.4   setosa
## 34          4.2   setosa
```