

María Paula Llano

Departamento de Ciencias de la
Atmósfera y los Océanos

Laboratorio de Procesamiento de Información Meteorológica/Oceanográfica

data.frame

Estructura recursiva - **data.frame**

Muy frecuentemente, los datos se disponen en tablas: las hojas de cálculo, las bases de datos, etc. contienen, esencialmente, tablas. Muchos de los métodos estadísticos trabajan con la información organizada en tablas.

Las hojas de cálculo son herramientas que prácticamente todos hemos usado para manipular tablas con datos. Las hojas de cálculo más conocidas son las de Excel.

Un **data frame** es una lista, cuyos componentes pueden ser vectores, matrices o factores, con la única salvedad de que las longitudes, o número de “renglones”, en el caso de matrices, deben coincidir en todos los componentes.

La apariencia de un data frame es la de una tabla y una forma de crearlos es mediante la función data.frame()

Trabajar con data.frame implica muchas veces, trabajar con grandes volúmenes de información.

Entonces existen funciones que nos muestran esa información.

Volvemos a usar los datos “iris”.

iris #aparece TODA la información

print(iris) # pasa lo mismo

Otras funciones útiles para inspeccionar son:

plot(iris) # la representa gráficamente

summary(iris) # resumen estadístico de las columnas

str(iris) # "representación textual" del objeto

head(iris) # primeras seis filas

tail(iris) # últimas seis filas

Estructura recursiva - data.frame

La apariencia de un data frame es la de una tabla y una forma de crearlos es mediante la función **data.frame()**

Para leer una tabla (que es un data frame) lo hacemos mediante la función **read.table()**

Acceso a las variables con \$, [[]], []

midataframe\$var2

midataframe[[2]]

midataframe[2]

midataframe[3, 2]

Más sobre data.frame

Para seleccionar una columna puedo emplear `$` o `[[]]`.

NO puedo llamar por el nombre de una columna.

Para poder “llamar” a cada componente utilizo `attach()` y la operación inversa es `detach()`

`attach(midataframe)`

`detach(midataframe)`

También puedo seleccionar según ciertas características mediante `subset()`

Transformaciones

Podemos modificar o manipular cualquier objeto con operadores (aritméticos, lógicos y comparativos) y funciones.

Aclaración: Cualquier modificación que hagamos con las variables de un banco de datos adjuntado (attached) no afectarán al propio data.frame.

Si queremos modificar el contenido interno de los data.frames podemos utilizar la función **transform()**.

Practicar SOLOS

Usando los datos de “airquality” ESCRIBIR las sentencias para calcular:

#¿cuál es la temperatura media de todos los días?

#¿cuál es la temperatura media en el mes de mayo?

#¿cuál fue el día más ventoso?

escribir la salida que nos diga día y mes del viento máximo

ordenar la información según el criterio de velocidad del viento ascendente.

crear una tabla adicional seleccionando solo 2 de las variables para el mes de mayo.

graficar estas variables seleccionadas.