

Manipulación de datos

Juan Felipe Acevedo Pérez

Monitor Junior

Correo: uniic_bog@unal.edu.co

Tel: 3165000 **Ext:** 12301

Contenido



**Tipos de
secuencia**



**Tipos de
objetos en
Python**



**Tipos de
objetos en R**

1

Tipos de secuencia / Estructuras de datos

Correo: uniic_bog@unal.edu.co
Teléfono: 3165000 ext 12301

UIFCE
UNIDAD DE INFORMÁTICA

Tipos de secuencia Python

- **Cadenas de caracteres:** Con ' o ""; es inmutable (no puede ser cambiada).
- **Listas** Paréntesis cuadrados []. Son mutables.
- **Data Frames** tiene la siguiente estructura:

```
import pandas as pd  
df = pd.DataFrame()
```
- **Tuplas:** Por los paréntesis redondos (). Son inmutables.
- **Diccionarios:** Entre { }. La llave es inmutable pero el valor sí es mutable.

Estructuras de datos R

- **Cadena de caracteres:** Con ' o "".
- **Listas:** Se reconoce por el comando **"list()"**.
- **Data Frames:** Se reconoce porque tiene el comando **"data.frame()"**.
- **Vectores:** se reconocen por el comando **c()**.
- **Matrices y Arrays:** Se reconocen por tener el comando **"matrix()"**.

2

Tipos de objetos en Python


Correo: uniic_bog@unal.edu.co
Teléfono: 3165000 ext 12301

UIFCE
UNIDAD DE INFORMÁTICA

Cadenas de caracteres

```
cadena = "Hola, mundo. Tengo un poco de texto sin sentido que servirá"+  
"para explicar cómo buscar cadenas dentro de cadenas"
```

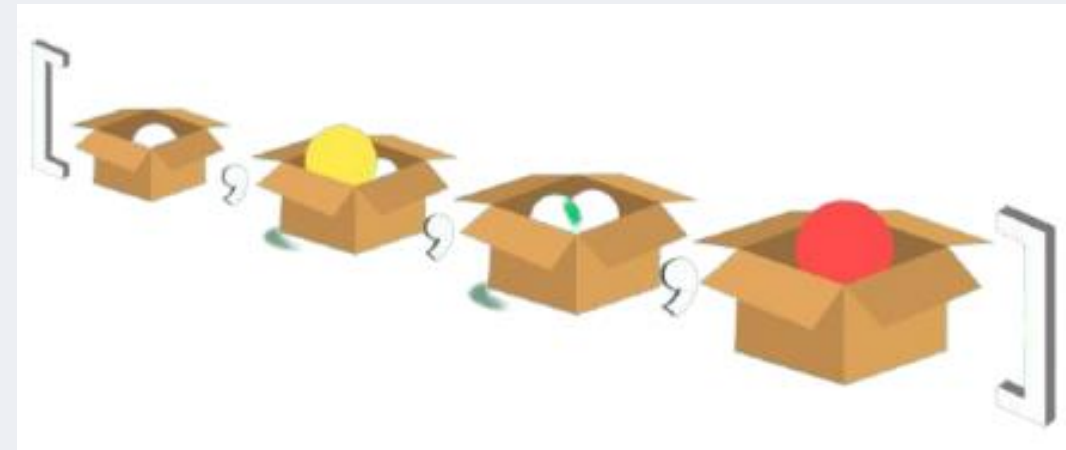
H o l a a t o d o s

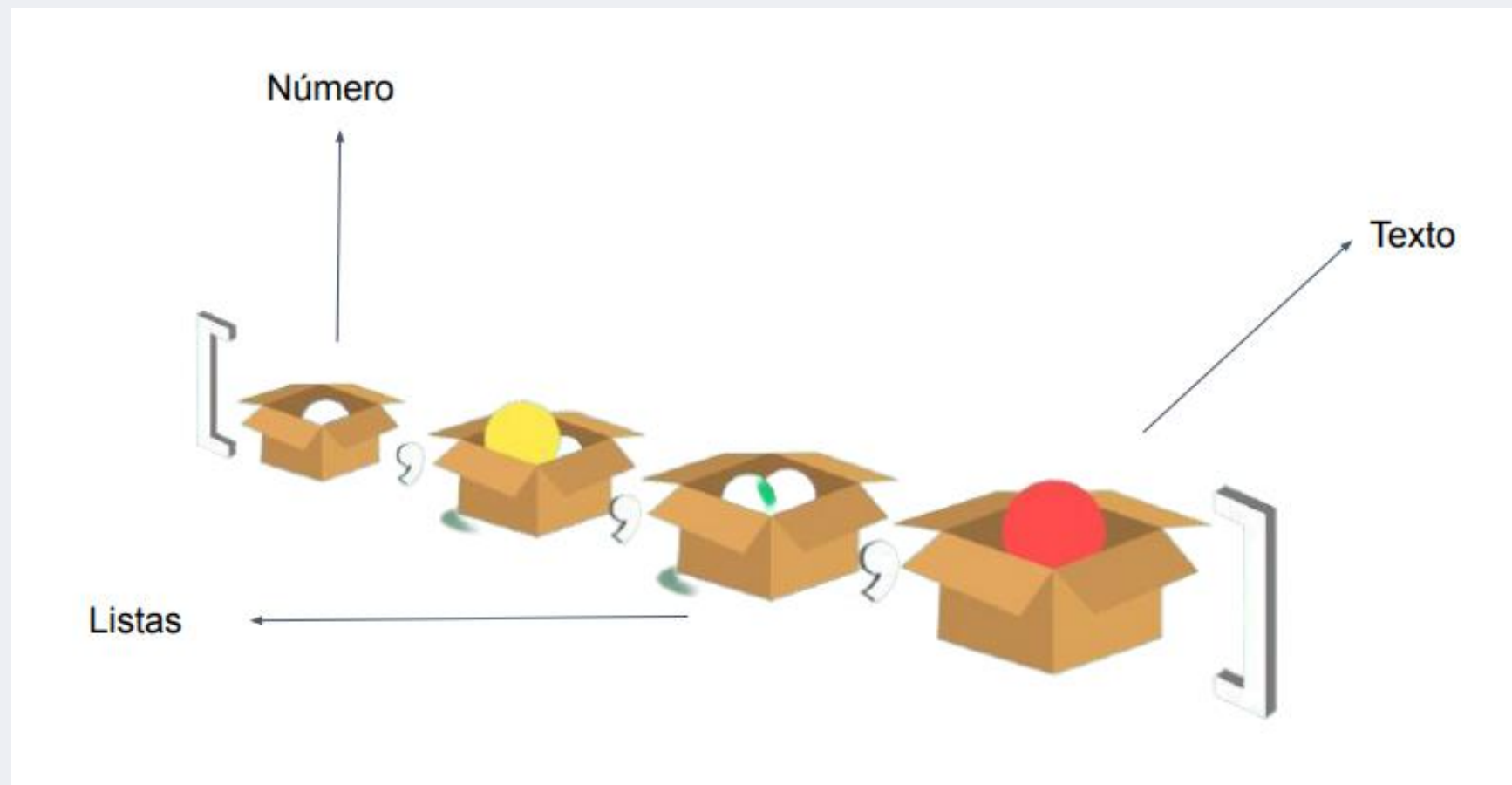


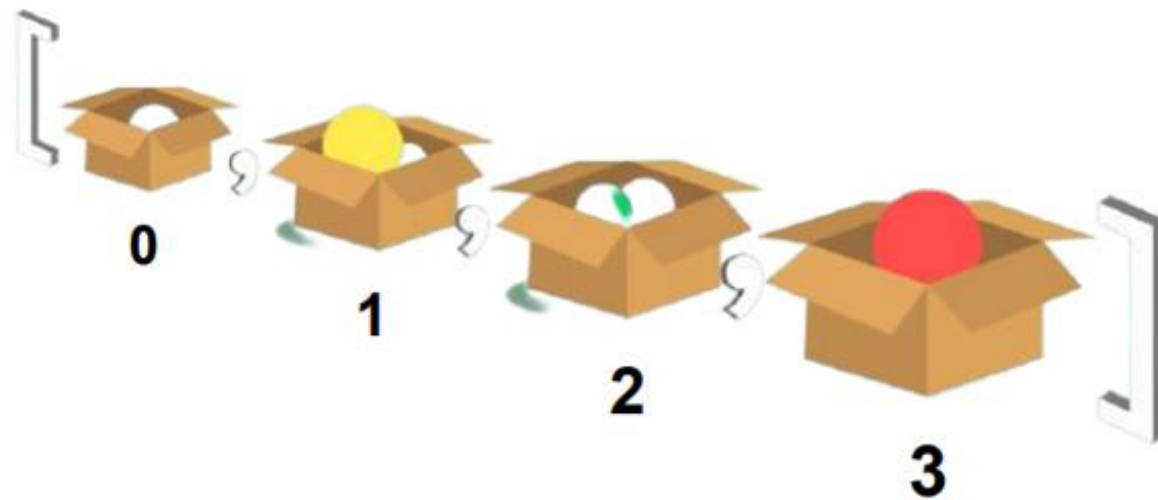
0 1 2 3 4 5 6 7 8 9 10 11

Listas Python

- Las listas son conjuntos ordenados de elementos (números, cadenas, listas, etc).
- Las listas se delimitan por corchetes ([]) y los elementos se separan por comas.



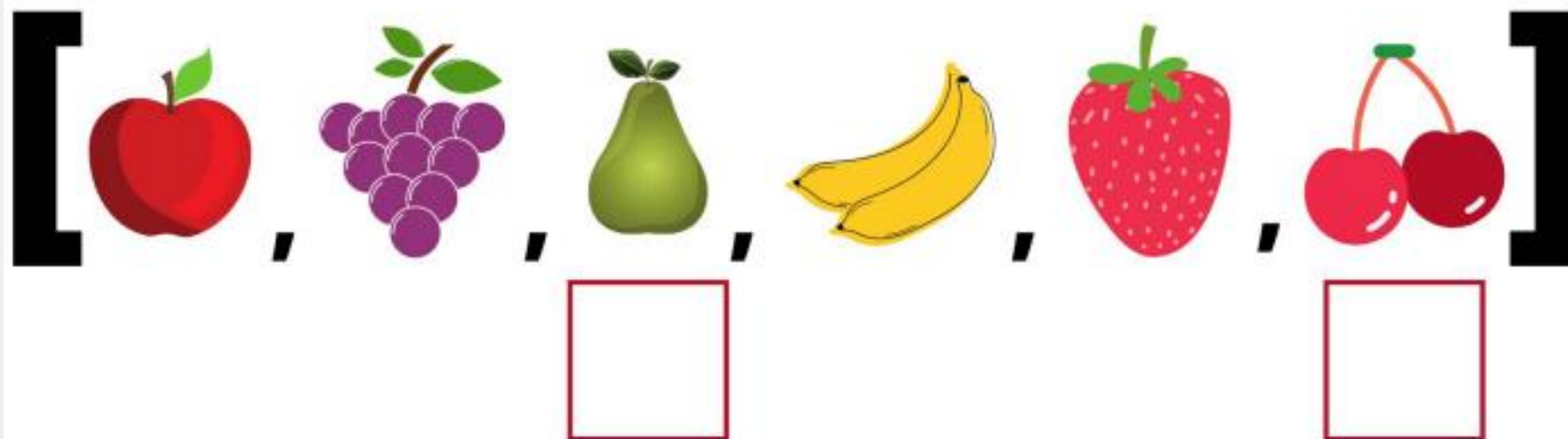




Correo: uniic_bog@unal.edu.co

Teléfono: 3165000 ext 12301

**Completa el cuadro con la posición
indicada**



Tuplas

Las tuplas son como las listas, excepto que son inmutables (sus elementos no pueden ser modificados). Se identifican fácilmente porque en vez de usar corchetes, se definen entre paréntesis. En lo demás, funcionan igual a las listas y al resto de las secuencias.

```
In [15]: f = (1, 2.3, "Python", True)
         type(f)
```

```
Out[15]: tuple
```

```
In [16]: print(f)
         (1, 2.3, 'Python', True)
```

Diccionarios

Los diccionarios son objetos de tipo mapping (un objeto contenedor que permite recuperado de claves arbitrarias). Los diccionarios se pueden crear colocando una lista separada por comas de pares key: value entre llaves, por ejemplo:

`{'jack': 1996, 'juan': 2000}` o
`{1996: 'jack', 2000: 'juan'}`, o
por el constructor `dict`.

Operaciones comunes

Operación	Resultado
<code>x in s</code>	<code>True</code> si un elemento de <code>s</code> es igual a <code>x</code> , <code>False</code> en caso contrario
<code>x not in s</code>	<code>False</code> si un elemento de <code>s</code> es igual a <code>x</code> , <code>True</code> en caso contrario
<code>s + t</code>	la concatenación de <code>s</code> y <code>t</code>
<code>s * n</code> o <code>n * s</code>	equivale a concatenar <code>s</code> consigo mismo <code>n</code> veces
<code>s[i]</code>	El elemento <i>i</i> -ésimo de <code>s</code> , empezando a contar en 0
<code>s[i:j]</code>	la rebanada de <code>s</code> desde <code>i</code> hasta <code>j</code>
<code>s[i:j:k]</code>	la rebanada de <code>s</code> desde <code>i</code> hasta <code>j</code> , con paso <code>j</code>
<code>len(s)</code>	longitud de <code>s</code>
<code>min(s)</code>	el elemento más pequeño de <code>s</code>
<code>max(s)</code>	el elemento más grande de <code>s</code>
<code>s.index(x[, i[, j]])</code>	índice de la primera ocurrencia de <code>x</code> en <code>s</code> (en la posición <code>i</code> o superior, y antes de <code>j</code>)
<code>s.count(x)</code>	número total de ocurrencias de <code>x</code> en <code>s</code>

3

Tipos de objetos en R

Correo: uniic_bog@unal.edu.co

Teléfono: 3165000 ext 12301

Tipos de objetos: **atómicos**

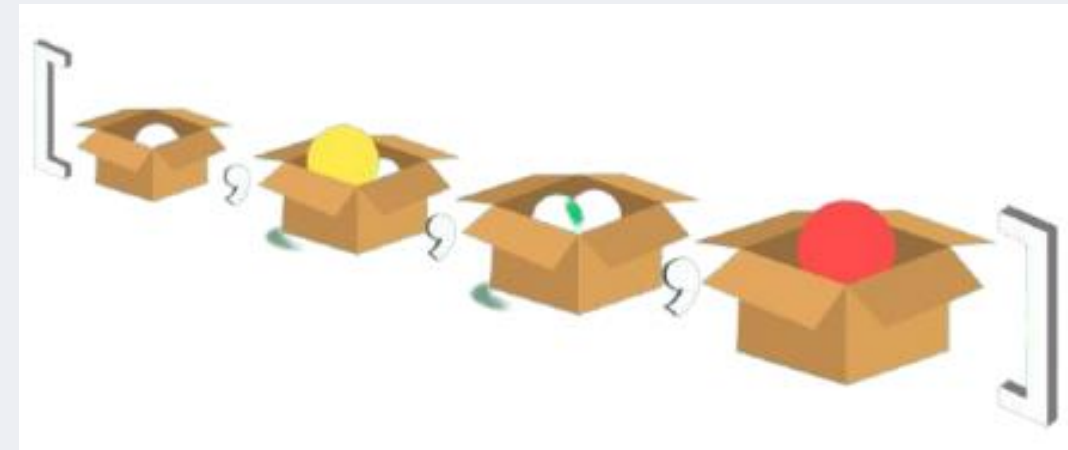
Objetos	Información
Vector	Cadenas unidimensionales (es decir una sola <i>columna</i> o <i>fila</i>) de un tipo único de valores (numéricos, caracteres, etc.) <code>c(1, 2, 3, 4, 5, 6)</code>
Matrices	Estructura bidimensional donde todos los datos deben ser del mismo tipo . <code>matrix(ejemplo, nrow=3, ncol=2)</code>
Factores	Este tipo de objeto es fundamental para el análisis estadístico pues es la forma como se tratan las variables categóricas. <code>> genero <- c("femenino", "masculino", "femenino", "femenino", "masculino", "femenino")</code> <code>> f <- factor(genero, levels=c("femenino", "masculino"))</code>

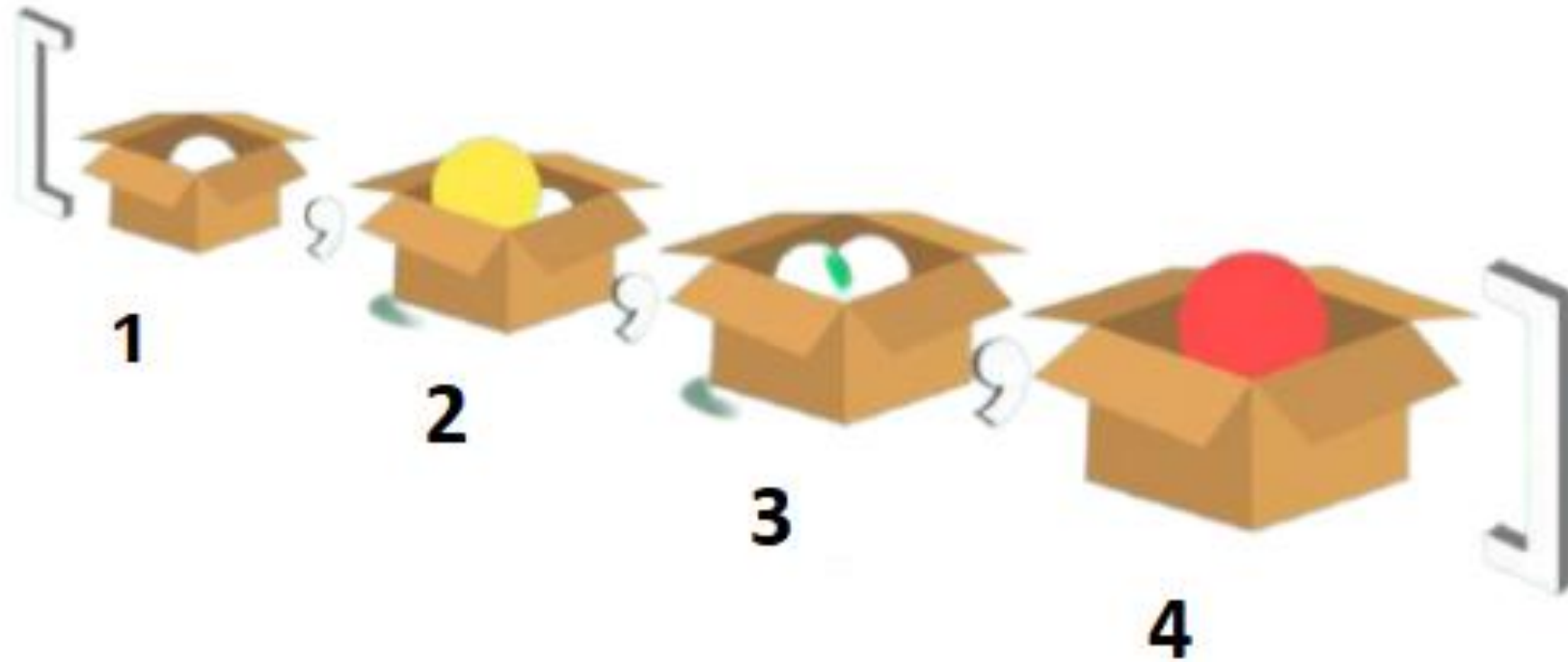
Tipos de objetos: **recursivos**

Objetos	Información
Listas	Las listas son objetos genéricos que corresponden a colecciones de distintos objetos que pueden ser de tipos distintos, a diferencia de los vectores y las matrices, donde los elementos deben ser del mismo tipo,
Data Frames	Como una matriz pero pudiendo contener distintos tipos de datos . Es lo que clásicamente se ha llamado <i>base de datos</i> o <i>hoja de datos</i> .

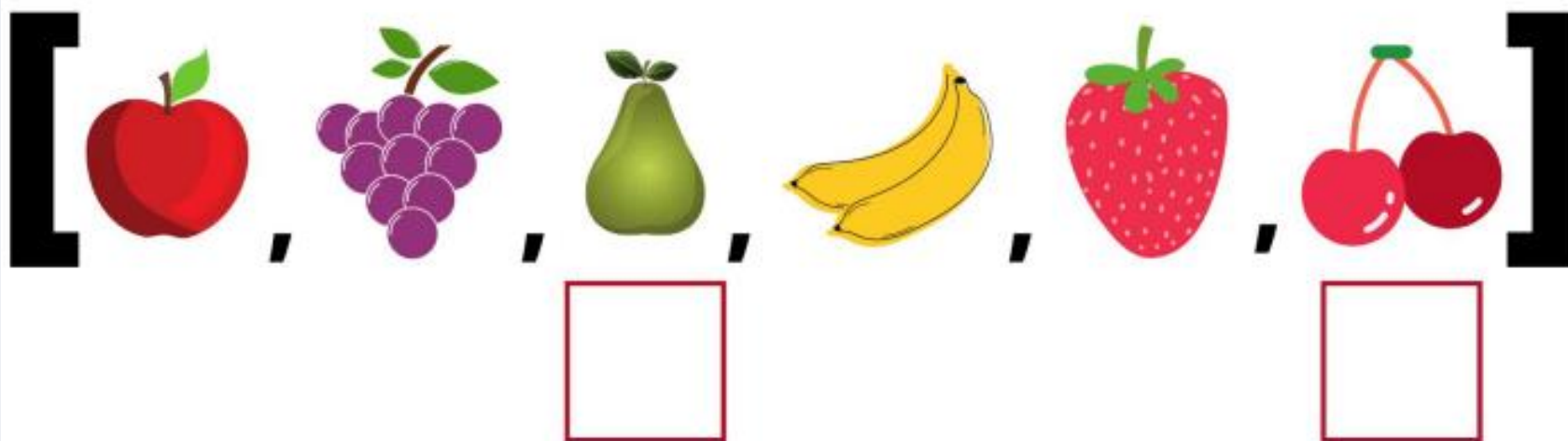
Listas R

- Podemos tener listas que contengan datos atómicos, vectores, matrices, arrays, data frames u otras listas.
- Para crear una lista usamos la función `list()`, que nos pedirá los elementos que deseamos incluir en nuestra lista. No importan las dimensiones.





**Completa el cuadro con la posición
indicada**



Data Frames

Los renglones en un data frame representan casos, individuos u observaciones, mientras que las columnas representan atributos, rasgos o variables.

##		Sepal.Length	Sepal.Width
##	1	5.1	3.5
##	2	4.9	3.0
##	3	4.7	3.2
##	4	4.6	3.1
##	5	5.0	3.6

Data Frames

Para crear un data frame
usamos la función
`data.frame()`. Esta función
nos pedirá un número de
vectores igual al número de
columnas que deseemos.
Está compuesto de vectores

##		Sepal.Length	Sepal.Width
##	1	5.1	3.5
##	2	4.9	3.0
##	3	4.7	3.2
##	4	4.6	3.1
##	5	5.0	3.6

Coerción

Al igual que con los datos, cuando intentamos hacer operaciones con una estructura de datos, R intenta coercirla al tipo apropiado para poder llevarlas a cabo con éxito.

Función	Coerciona a	Coerciona exitosamente a
<code>as.vector()</code>	Vector	Matrices
<code>as.matrix()</code>	Matrices	Vectores, Data frames
<code>as.data.frame()</code>	Data frame	Vectores, Matrices
<code>as.list()</code>	Lista	Vectores, Matrices, Data frames