

EJERCICIOS DE R: DATA FRAME

Una de las aplicaciones más útiles del lenguaje de programación R sería el almacenamiento de datos, para lo que va a utilizar la instrucción `data.frame`.

En este documento se van a proponer una serie de ejercicios que tendrán como finalidad practicar y mejorar el uso de dicha instrucción. Además, se incluirán las soluciones de cada ejercicio, indicando posibles errores que se pueden cometer.

- 1) En un grupo de amigos, Adriana tiene 18 años y mide 1,55m, Ismael tiene 17 años y mide 1,83m y Pedro, que mide 1,88m, tiene 19 años. Represente esa información en una tabla en la que las columnas sean “amigo”, “edad” y “estatura”.

RESOLUCIÓN:

```
> rm(list=ls())
> amigo=c("Adriana", "Ismael","Pedro")
> edad=c(18,17,19)
> estatura=c(155,183,188)
> AA=data.frame(amigo, edad, estatura)
> AA
```

	amigo	edad	estatura
1	Adriana	18	155
2	Ismael	17	183
3	Pedro	19	188

* En caso de que tengamos valores almacenados previamente en la memoria y deseemos borrarlos antes de empezar a hacer el ejercicio, usaremos la instrucción `rm(list=ls())` antes de definir los diferentes objetos

- Un error común en este tipos de ejercicios sería no llamar de la misma manera a los elementos que se han definido inicialmente y a los elementos que se han introducido en la instrucción `data.frame`, lo que hará que de error el programa:

```
> rm(list=ls())
> amigo=c("Adriana", "Ismael","Pedro")
> edad=c(18,17,19)
> estatura=c(155,183,188)
> AA=data.frame(Amigo, edad, estatura)
Error: object 'Amigo' not found
> AA
Error: object 'AA' not found
```

- 2) Realice una tabla que contenga las columnas “ciudad”, “pais”, “continente” e “idioma” a partir de la siguiente información:
- Londres se encuentra en Inglaterra, es del continente europeo y su idioma oficial es el inglés.
 - Copenhague se encuentra en Dinamarca, en Europa, y su idioma oficial es el danés.
 - Brasilia está en Brasil, en América, y su idioma oficial es el portugués.
 - Sidney es una ciudad australiana, perteneciente al continente de Oceanía, donde se habla inglés.
 - Tokio es la capital de Japón, un país de Asia en el que se habla japonés.

RESOLUCIÓN

```
> rm(list=ls())
> ciudad=c("Londres", "Copenhage", "Brasilia", "Sidney", "Tokio")
> pais=c("Inglaterra", "Dinamarca", "Brasil", "Australia", "Japon")
> continente=c("Europa", "Europa", "America", "Oceania", "Asia")
> idioma=c("ingles", "danes", "portugues", "ingles", "japones")
> AA=data.frame(ciudad,pais,continente,idioma)
> AA
```

	ciudad	pais	continente	idioma
1	Londres	Inglaterra	Europa	ingles
2	Copenhage	Dinamarca	Europa	danes
3	Brasilia	Brasil	America	portugues
4	Sidney	Australia	Oceania	ingles
5	Tokio	Japon	Asia	japones

A la hora de introducir los objetos, es importante recordar que el lenguaje de programación R no reconoce las tildes.

- 3) En una clase hay 8 chicos y 6 chicas que juegan al fútbol, 5 chicos y 9 chicas que hacen baloncesto, 3 chicos y 4 chicas que hacen boxeo y 4 chicos y 5 chicas que juegan al tenis. Representa la información dada en una tabla en la que aparezcan los deportes y el número de jugadores y jugadoras.

RESOLUCIÓN:

```
> rm(list=ls())
> deporte=c("futbol", "baloncesto", "boxeo", "tenis")
> jugadores=c(8,5,3,4)
> jugadoras=c(6,9,4,5)
> BB=data.frame(deporte, jugadores, jugadoras)
> BB
```

	deporte	jugadores	jugadoras
1	futbol	8	6
2	baloncesto	5	9
3	boxeo	3	4
4	tenis	4	5

- 4) Cuatro alumnos han recibido sus calificaciones de varios exámenes, junto con un comentario de su tutor:
- Alicia ha sacado un 6,9 en lengua, un 5.7 en matemáticas, un 9.8 en inglés, un 8 en biología y como comentario le han puesto “bastante bien”.
 - María ha sacado un 7,6 en lengua, un 9 en matemáticas, un 7,2 en inglés, un 7 en biología y como comentario le han puesto “bastante bien”.

- c) Iker ha sacado un 8 en lengua, un 6,9 en matemáticas, un 8,8 en inglés, un 9 en biología y como comentario le han puesto “muy bien”.
- d) Carlota ha sacado un 9,3 en lengua, un 8,5 en matemáticas, un 10 en inglés, un 8,5 en biología y como comentario le han puesto “muy bien”.

Represente dicha información en una tabla que contenga las columnas “alumno”, “lengua”, “matemáticas”, “inglés”, “biología” y “comentario”

RESOLUCIÓN:

```
> rm(list=ls())
> alumnos=c("Alicia","Maria","Iker","Carlota")
> lengua=c(6.9,7.6,8,9.3)
> matematicas=c(5.7,9,6.9,8.5)
> ingles=c(9.8,7.2,8.8,10)
> biologia=c(8,7,9,8.5)
> comentarios=c("bastante bien","bastante bien","muy bien","muy bien")
> CC=data.frame(alumnos,lengua,matematicas,ingles,biologia,comentarios)
> CC
```

	alumnos	lengua	matematicas	ingles	biologia	comentarios
1	Alicia	6.9	5.7	9.8	8.0	bastante bien
2	Maria	7.6	9.0	7.2	7.0	bastante bien
3	Iker	8.0	6.9	8.8	9.0	muy bien
4	Carlota	9.3	8.5	10.0	8.5	muy bien

Es importante recordar que a la hora de introducir datos numéricos, el lenguaje de programación R va a reconocer la notación decimal con puntos, no con comas:

```
> rm(list=ls())
> alumnos=c("Alicia","Maria","Iker","Carlota")
> lengua=c(6,9,7,6,8,9,3)
> matematicas=c(5.7,9,6.9,8.5)
> ingles=c(9,8,7,2,8,8,10)
> biologia=c(8,7,9,8.5)
> comentarios=c("bastante bien","bastante bien","muy bien","muy bien")
> CC=data.frame(alumnos,lengua,matematicas,ingles,biologia,comentarios)
Error in data.frame(alumnos, lengua, matematicas, ingles, biologia, comentarios) :
  arguments imply differing number of rows: 4, 7
```

Al introducir números decimales con comas en vez de con números, el programa va a interpretar dicho número como dos distintos (en vez de 7.2, interpretará 7 y 2), por lo que no coincidirán las cantidades de valores totales que se han introducido en cada objeto y la tabla no se podrá realizar

5) Se ha conoce la siguiente información sobre tres de las redes sociales más populares actualmente:

- a) Instagram fue fundado por Kevin Systrom en 2010 y tiene 2.000.000.000 de usuarios
- b) Twitter fue fundado por Jack Dorsey en 2006 y cuenta con 200.000.000 usuarios activos

- c) Spotify fue fundado en 2006 por Daniel Ek y actualmente tiene 550.000.000 de usuarios.

Utilice esa información para representar una tabla en la que aparezcan cuatro columnas: “aplicación”, “creador”, “fundación” y “usuarios”

RESOLUCIÓN:

```
> rm(list=ls())
> aplicacion=c("Instagram","Twitter","Spotify")
> creador=c("Kevin Systrom", "Jack Dorsey", "Daniel Ek")
> fundacion=c(2010,2006,2006)
> usuarios=c(2e9,2e8,5.5e8)
> DD=data.frame(aplicacion,creador,fundacion,usuarios)
> DD
```

	aplicacion	creador	fundacion	usuarios
1	Instagram	Kevin Systrom	2010	2.0e+09
2	Twitter	Jack Dorsey	2006	2.0e+08
3	Spotify	Daniel Ek	2006	5.5e+08

Recordemos que a la hora de introducir números muy grandes, como 2.000.000.000, podemos expresarlo como 2×10^9 e introducirlo con la notación 2e9. A continuación comprobamos que obtenemos la misma tabla, independientemente de la manera en la que metamos esos valores:

```
> rm(list=ls())
> aplicacion=c("Instagram","Twitter","Spotify")
> creador=c("Kevin Systrom", "Jack Dorsey", "Daniel Ek")
> fundacion=c(2010,2006,2006)
> usuarios=c(2000000000,200000000,550000000)
> DD=data.frame(aplicacion,creador,fundacion,usuarios)
> DD
```

	aplicacion	creador	fundacion	usuarios
1	Instagram	Kevin Systrom	2010	2.0e+09
2	Twitter	Jack Dorsey	2006	2.0e+08
3	Spotify	Daniel Ek	2006	5.5e+08