

# Clase 1.1

# Estructuras de datos

Marcos Rosetti y Luis Pacheco-Cobos

Estadística y Manejo de Datos con R (EMDR) — Virtual

# Marcos de Datos

# Marcos de Datos

- Estructuras bidimensionales que combinan argumentos numéricos y cadenas de texto dentro de la misma entidad.

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))  
df
```

```
##      a b          c  
## 1    1 a  0.5403023  
## 2    2 b -0.4161468  
## 3    3 c -0.9899925  
## 4    4 d -0.6536436  
## 5    5 e  0.2836622  
## 6    6 f  0.9601703  
## 7    7 g  0.7539023  
## 8    8 h -0.1455000  
## 9    9 i -0.9111303  
## 10 10 j -0.8390715
```

# Marcos de Datos

- Podemos hacer referencia con los corchetes [ ]

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))
df[2]
```

```
##      b
## 1    a
## 2    b
## 3    c
## 4    d
## 5    e
## 6    f
## 7    g
## 8    h
## 9    i
## 10   j
```

# Marcos de Datos

- Podemos usar el número o el nombre (entre comillas).

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))  
df["c"][1:5, 1]
```

```
## [1] 0.5403023 -0.4161468 -0.9899925 -0.6536436 0.2836622
```

# Marcos de Datos

- Podemos usar el operador compacto `$` como referencia.

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))  
df$c
```

```
## [1] 0.5403023 -0.4161468 -0.9899925 -0.6536436 0.2836622 0.9601703  
## [7] 0.7539023 -0.1455000 -0.9111303 -0.8390715
```

```
df$c[1:5]
```

```
## [1] 0.5403023 -0.4161468 -0.9899925 -0.6536436 0.2836622
```

# Marcos de Datos

- ¿Qué diferencia puedes ver entre la referenciación usando de corchetes...

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))  
df["c"]
```

```
##           c  
## 1  0.5403023  
## 2 -0.4161468  
## 3 -0.9899925  
## 4 -0.6536436  
## 5  0.2836622  
## 6  0.9601703  
## 7  0.7539023  
## 8 -0.1455000  
## 9 -0.9111303  
## 10 -0.8390715
```

# Marcos de Datos

- ...y la referenciación usando el operador compacto?

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))  
df$c
```

```
## [1] 0.5403023 -0.4161468 -0.9899925 -0.6536436 0.2836622 0.9601703  
## [7] 0.7539023 -0.1455000 -0.9111303 -0.8390715
```



# Marcos de Datos

- Funciones que nos ayudan a conocer mejor nuestro marco de datos.

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))  
head(df)
```

```
##      a b          c  
## 1 1 a  0.5403023  
## 2 2 b -0.4161468  
## 3 3 c -0.9899925  
## 4 4 d -0.6536436  
## 5 5 e  0.2836622  
## 6 6 f  0.9601703
```

# Marcos de Datos

- Funciones que nos ayudan a conocer mejor nuestro marco de datos.

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))  
tail(df)
```

```
##      a b      c  
## 5    5 e  0.2836622  
## 6    6 f  0.9601703  
## 7    7 g  0.7539023  
## 8    8 h -0.1455000  
## 9    9 i -0.9111303  
## 10  10 j -0.8390715
```

# Marcos de Datos

- Funciones que nos ayudan a conocer mejor nuestro marco de datos.

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))
str(df)
```

```
## 'data.frame':  10 obs. of  3 variables:
## $ a: int  1 2 3 4 5 6 7 8 9 10
## $ b: chr  "a" "b" "c" "d" ...
## $ c: num  0.54 -0.416 -0.99 -0.654 0.284 ...
```

# Marcos de Datos

- Facilita pensar en el marco de datos como un equivalente a la hoja de cálculo.

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))
```

	A	B	C
1	a	b	c
2		1a	0.540302306
3		2b	-0.416146837
4		3c	-0.989992497
5			

# Marcos de Datos

- Los nombres de columna (a.k.a headers) serán los nombres de los componentes del marco de datos.

```
df <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))
colnames(df) <- c("Numeros", "Letras", "Coseno")
df
```

```
##      Numeros Letras      Coseno
## 1          1      a  0.5403023
## 2          2      b -0.4161468
## 3          3      c -0.9899925
## 4          4      d -0.6536436
## 5          5      e  0.2836622
## 6          6      f  0.9601703
## 7          7      g  0.7539023
## 8          8      h -0.1455000
## 9          9      i -0.9111303
## 10         10      j -0.8390715
```

# Marcos de Datos

- Se pueden agregar columnas o unir marcos de datos de manera horizontal con `cbind.data.frame()`

```
df1 <- data.frame(a = 1:10, b = letters[1:10], c = cos(1:10))
df2 <- data.frame(s = sin(1:10), t= tan(1:10))
cbind.data.frame(df1, df2)
```

```
##      a b      c      s      t
## 1  1 a  0.5403023  0.8414710  1.5574077
## 2  2 b -0.4161468  0.9092974 -2.1850399
## 3  3 c -0.9899925  0.1411200 -0.1425465
## 4  4 d -0.6536436 -0.7568025  1.1578213
## 5  5 e  0.2836622 -0.9589243 -3.3805150
## 6  6 f  0.9601703 -0.2794155 -0.2910062
## 7  7 g  0.7539023  0.6569866  0.8714480
## 8  8 h -0.1455000  0.9893582 -6.7997115
## 9  9 i -0.9111303  0.4121185 -0.4523157
## 10 10 j -0.8390715 -0.5440211  0.6483608
```

# Marcos de Datos

- Se pueden agregar filas, o unir marcos de datos de manera vertical con `rbind.data.frame()`

```
df1 <- data.frame(a = 1:5, b = letters[1:5], c = cos(1:5))
df2 <- data.frame(a = 11:16, b = letters[11:16], c = cos(11:16))
rbind.data.frame(df1, df2)
```

```
##      a b      c
## 1    1 a  0.540302306
## 2    2 b -0.416146837
## 3    3 c -0.989992497
## 4    4 d -0.653643621
## 5    5 e  0.283662185
## 6   11 k  0.004425698
## 7   12 l  0.843853959
## 8   13 m  0.907446781
## 9   14 n  0.136737218
## 10  15 o -0.759687913
## 11  16 p -0.957659480
```

# Licencia CC BY



Estadística y Manejo de Datos con R (EMDR) por Marcos F. Rosetti S. y Luis Pacheco-Cobos se distribuye bajo una [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).