

El objetivo de esta guía es repasar algunos conceptos relativos a la sintaxis y manejo de datos usando R, siguiendo la propuesta presentada en *Foundations of Statistics with R*, de Darrin Speegle y Bryan Clair. Haciendo clicke aquí podrá acceder al libro. Buena suerte!

1. a) Sea

```
x <- c(1,2,3)
y <- c(6,5,4).
```

Predecir el resultado de correr cada una de las siguientes líneas y después chequear las respuestas.

```
x * 2
x * y
x[1] * y[2]
1/x
(1:10) * x[2]
rep(c(1,1,2), times = 2)
seq(from = 0, to = 10, length.out = 5) + 1:10
```

- b) ¿Qué ocurre si se multiplican 2 vectores de distinta longitud?

2. a) Crear un vector llamado tratamiento con coordenadas  $A$ ,  $B$  y  $C$  de manera que  $A$  aparezca 20 veces,  $B$  18 y  $C$  22.

- b) Definir un vector  $J$  que sea una secuencia de 1 a 30 con un incremento de 2 y luego sumar la primera y la octava coordenadas.

3. Calcular la suma de los números naturales del 1 al 100 usando R.

4. Calcular la suma de los cuadrados de los números naturales del 1 al 100 usando R.

5. Considerar el conjunto de datos `airquality`, incluido en la librería `datasets`. Pida ayuda con el comando `?airquality` para obtener información.

- a) ¿Cuántas observaciones tiene el conjunto de datos? ¿Cuántas variables?

- b) ¿Cuáles son los nombres de las variables?



- c) ¿Qué variables tienen datos faltantes?

- d) ¿Cuántas observaciones corresponden al mes de mayo?

6. Considerar el conjunto de datos `mtcars` en R. (Ver `help(mtcars)`)



- a) ¿Qué autos tienen 4 velocidades?

- b) ¿Qué subconjunto de `mtcars` es `mtcars[mtcars$disp > 150 & mtcars$mpg > 20,]`?



- c) ¿Qué autos tienen 4 velocidades y transmisión manual?

- d) Hallar la cantidad media de millas por galón de los autos con 2 carburadores.

7. El conjunto de datos `arbolado-en-espacios-verdes.csv` contiene datos de 2011 de todos los árboles de los espacios verdes de la ciudad de Buenos Aires. Los datos actualizados pueden encontrarse aquí, junto con una descripción de los mismos .

- a) ¿Cuántas observaciones tiene el conjunto de datos? ¿Cuántas variables?
- b) ¿Cuáles son los nombres de las variables?
- c) Calcular la altura promedio de los árboles de la ciudad.
- d) ¿Cuántos árboles había en 2011 en la plaza Arenales?
- e) Construya un `data.frame` llamado `arboles_cercanos` que contenga sólo las filas correspondientes a su espacio verde favorito. Puede utilizar el comando `sort` para ordenar alfabéticamente un vector de caracteres.
- f) Utilizando el comando `unique` averigüe los nombres de los árboles presentes en su espacio verde favorito en 2011.

8. Los datos TITANIC3 del paquete PASWR2 contienen información sobre los pasajeros del Titanic, incluyendo clase, sexo y si sobrevivieron o no, entre otras características.



- a) Determine la proporción de sobrevivientes por clase.
- b) Calcule la proporción de sobrevivientes por clase y sexo. ¿Quién tuvo una tasa más alta de supervivencia: los varones de 1ra clase o las mujeres de 3ra?
- c) ¿Cuál era la edad de la mujer más grande que sobrevivió?

9. El conjunto de datos CARS2004 del paquete PASWR2 contiene datos de automóviles en Europa del año 2004. ( ver `help(CARS2004)`)



- a) Calcular cantidad total de autos en cada país.
- b) Calcular la tasa de mortalidad de automovilistas para cada país como el número total de muertes de automovilistas dividido el número total de autos.



- c) Hacer un gráfico de barras que indique la tasa de mortalidad en accidentes de tránsito de cada país de la Unión Europea. Ordene las barras de forma creciente.
- d) ¿Qué país tiene la menor tasa de mortalidad de automovilistas y qué país la más baja?
- e) Haga un gráfico de cantidad total de autos vs población. ¿Cómo describiría la relación?



- f) Haga un gráfico de cantidad total de autos vs tasa de mortalidad de automovilistas. ¿Cómo describiría la relación?