

Día 1) Elementos básicos de R:

Datos y estructuras de datos

Los principales tipos de datos en R son numeric (que está reservado para los números naturales) y integer (reservado para números enteros), character (para las palabras o “strings”), logical (TRUE o FALSE). Algunos tipos de datos pueden ser convertidos en otros, así por ejemplo una variable puede tener como valor el número 1, o el string “1”. Sin embargo esto no ocurre en todos los casos, ¿se podría convertir el string “Hola” en un objeto de la clase numeric?

Los 5 tipos de estructuras de datos más importantes en R son: vector, matrix, array, list y data.frame. Éstos pueden ser datos de 1 dimensión (como vector y list), de 2 dimensiones (matrix y data.frame) o de más dimensiones (array). Por otro lado, las estructuras de datos también se diferencian de acuerdo a si los elementos que contienen son todos del mismo tipo o no. Mientras que en vector, matrix y array todos los datos tienen que ser del mismo tipo (por ej. numeric), en list y dataframe esto no es así (por ej. podemos tener un dataframe que en la primera columna tenga datos tipo character y en la segunda datos tipo numeric). Además de las estructuras mencionadas también tenemos los factores. Factor es un tipo especial de vector de caracteres que se reserva para asignar variables cualitativas, como por ejemplo los distintos tratamientos dentro de un experimento)

Ejercicios:

- 1) Asigne a la variable x el valor 2 y a la variable y el valor 10. Ahora asigne a la variable w el valor x+y y a la variable z el valor x == y (observá que tiene dos signos = y no sólo uno, ¿Por qué?). ¿De qué clase son estas dos últimas variables?. **Ayuda:** la asignación de una variable se realiza con el operador “<-”, así por ejemplo para asignar el valor “Hola Mundo” a la variable Saludo debo tipear Saludo<-“Hola Mundo”. Además, tipeá en la consola ?class
- 2) Creá una variable que se llame “GFP” y asigne el valor 509. ¿Qué clase de dato es? Ahora convierte la variable “GFP” en “character”. **Ayuda:** tipeá en la consola ?as.character
- 3) Asigne a la variable GFP el valor “verde” e intenta convertirla en la clase numeric ¿Qué ocurre?
- 4) Crea un vector que se llame “r1” que contenga 10 números aleatorios elegidos por vos entre el 0 y el 100 **Ayuda:** ?c.
- 5) Con la misma operación que en el ejercicio anterior crea un nuevo vector r2, ahora realiza las siguientes operaciones: r1+r2, r1*r2, r1/r2. ¿Cómo realiza R estas operaciones?
- 6) Intenten ordenar de mayor a menor y de menor a mayor el vector r1. ¿De qué largo es el vector r1? ¿Cómo podría agregarle un elemento al vector r1? **Ayuda:** ?sort, ?length
- 7) Creá dos vectores. Uno llamado fp que contenga los siguientes datos: "Sirius", "CFP", "GFP", "Citrine", y otro llamado nm que contenga los siguientes valores: 424, 476, 509, 528. A partir de estos dos vectores crea una matriz m_fp y un dataframe que se llame df_fp. Comprueba de qué clase es cada objeto. ¿Qué tienen de diferente? **Ayuda:** ?cbind y ?data.frame
- 8) Construí un vector con todos los números del 1 al 9. **Ayuda:** comienzo:fin
- 9) Construí un vector con todos los números impares del 1 al 9. **Ayuda:** ?seq

- 10) Crea una matriz `m` de 3x2 (3 filas y 2 columnas) que contenga en todas las posiciones el valor 0. **Ayuda:** `?matrix`
- 11) Crea una matriz `m` de 3x3 que contenga en la posición [1, 1] el número 1, en la [1, 2] el número 2 y así hasta el número 9 en la posición [3, 3]. Los elementos de una matriz son [fila, columna]. **Ayuda:** `?matrix` y ejercicio 8).
- 12) A la matriz del punto anterior cambiale el valor de la primera fila y columna por “Esto es un string” con el siguiente código: `m[1, 1] <- “Esto es un string”` (¿Podrías explicar cómo funciona esto?). Imprimir el resultado en pantalla, ¿Qué ocurrió con el resto de los valores de la matriz? ¿Por qué?
- 13) Solo si llegamos a ver factores en clase: a partir de un vector “Co2” con los siguientes elementos: “Bajo”, “Medio”, “Medio”, “Bajo”, “Bajo” y “Alto”; crea un objeto de tipo factor. ¿Qué niveles (o levels) tiene? Ahora cambia el primer valor del objeto por “Medio” de la siguiente forma `Co2[1] <- “Medio”` ¿Cambiaron los niveles del objeto Co2? Ahora si quiero cambiar el primer valor por “Muy bajo”, ¿Qué ocurre? ¿Por qué? **Ayuda:** `?factor`, `?levels`

Subsetting

- 14) Generen dos listas a partir de las siguientes sintaxis: `nombres1 <- list(c("Juana", "Pedro", "Camila"))` y `nombres2 <- list("Juana", "Pedro", "Camila")`. ¿Cuántos elementos tienen cada una de ellas? Para cada caso intenten imprimir en pantalla solamente el nombre Camila

A partir del paquete “datasets” generen un `data.frame` “air” que contiene datos climáticos de la ciudad de Nueva York, usando el siguiente comando: `air <- datasets::airquality`. (Hay algunos paquetes precargados en R, como por ejemplo datasets, a los que se accede poniendo nombre del paquete:).

- 15) Imprime en pantalla solamente las primeras filas de `air` y luego las últimas. **Ayuda:** `?head`, `?tail`
- 16) Imprime en pantalla todos los valores de Temperaturas registradas y luego solamente el valor de Temperatura que se registró el tercer día.
- 17) Seleccioná todas las filas de `air` del mes de mayo.
- 18) ¿En qué día hubo menor radiación solar? ¿De cuánto fue esa radiación? **Ayuda:** `?which.min` y `?min`. Al correr el comando `min`, observamos que el resultado devuelto es NA. ¿Qué significa? ¿Por qué es el mínimo?. Probar nuevamente con `min` pero pasando el parámetro `na.rm = TRUE`, es decir, `min(vector, na.rm=TRUE)`. ¿Qué obtenemos ahora?
- 19) ¿Cuál fue la temperatura el 27 de agosto? **Ayuda:** `?which`, `&` (este símbolo se llama ampersand y es equivalente a “y”. Permite concatenar dos condiciones, por ejemplo, quiero que el día sea el 27 y el mes sea agosto. Solo aquellos registros que cumplan ambas condiciones van a ser incluidos.)
- 20) ¿Cuál fue la temperatura los meses de agosto y septiembre? **Ayuda:** `?which`, `|` (este símbolo se llama pipe o tubería y es equivalente a “o”. Permite concatenar dos condiciones, por

ejemplo, quiero que el mes sea agosto o el mes sea septiembre. Solo aquellos registros que cumplan alguna de las dos condiciones van a ser incluidos.)

21) Seleccioná todas las filas de air del mes de mayo cuya radiación solar sea mayor a 150.

22) La temperatura en el data.frame está expresada en grados Fahrenheit, generá una nueva columna que se llame temp_c con las temperaturas en grados Celsius. **Ayuda:** $^{\circ}\text{C} = (^{\circ}\text{F} - 32)/1.8$

23) Genera un nuevo data.frame llamado “calor” que contenga la información de los días calurosos, por ejemplo, en los que hizo más de 30°C. Luego, utilizando la función table contá cuantos días calurosos hubo en cada mes. ¿Qué mes es el que tuvo mayor cantidad de días calurosos?. **Ayuda:** La función table permite contar casos. Por ejemplo, aplicando table al vector meses <- c(“julio”, “julio”, “agosto”, “marzo”, “agosto”, “julio”), obtenemos la siguiente tabla:

agosto	julio	marzo
2	3	1

Es decir, table nos dice que en el vector meses, agosto aparece 2 veces, julio 3 y marzo 1.