

# Trabajo práctico 1: HelloR

## LABORATORIO DE DATOS

Facultad de Ciencias Exactas y Naturales - Universidad de Buenos Aires

Verano 2022

Este trabajo práctico debe entregarse en un *script* de **R**. Intercale texto comentado con el código para describir lo que usted realiza en su *script*, mencionar las magnitudes que calcula y las conclusiones que saca.

1. Tomando los datos de nombres del 2000 al 2004 (disponibles en [este link](#)) realice los siguientes ejercicios:

- a) Cargue los datos y visualice los primeros elementos *¿Qué información contiene el dataset?*. Convierta todos los caracteres de los nombres a minúscula.
- b) Cuente cuantos nombres distintos hay en total, y cuantos dentro de cada año. En base a esto, ¿cuánto se solapan los nombres de distintos años?
- c) El coeficiente de Jaccard entre dos conjuntos  $A$  y  $B$  se define como:

$$\frac{|A \cap B|}{|A \cup B|}$$

Construya una matriz  $J$  que tenga en  $J[i, j]$  el coeficiente de Jaccard entre el  $i$ -ésimo y el  $j$ -ésimo año. ¿Confirma esta matriz su suposición previa?

*Tip: recuerde aprovechar las funciones [unique](#), [intersect](#), [union](#) y [length](#).*

- d) Encuentre los 10 nombres más comunes en cada año. Compárelos. ¿Hay nombres repetidos? Encuentre los 10 nombres menos comunes en cada año. Compárelos. ¿Hay algo que los caracterice?
- e) Construya un nuevo dataset que abarque los 5 años, y que cuente la cantidad de veces que cada nombre fue utilizado en los mismos. *Tip: Si el tiempo de cálculo es muy grande, explore la función [aggregate](#)*

2. Una caminata al azar es una formalización matemática de la trayectoria que resulta de hacer sucesivos pasos aleatorios. Por ejemplo, la ruta trazada por una molécula mientras viaja por un líquido o un gas, el camino que sigue un animal en su búsqueda de comida, el precio de una acción fluctuante y la situación financiera de un jugador pueden tratarse, bajo ciertas hipótesis, como una caminata aleatoria.

En este punto construiremos un caminante al azar. Para esto:

- a) Usando la función [runif](#)( $n$ ,  $\min$ ,  $\max$ ) que genera  $n$  números al azar entre  $\min$  y  $\max$ , construya una función [da\\_el\\_paso](#)( $x_0$ ) que tome una posición inicial  $x_0$  y le sume un desplazamiento al azar entre -1 y 1 generado con [runif](#).
- b) Arme una función [esta\\_entre](#)( $x_0$ ,  $T_0$ ,  $T_1$ ) que verifique si  $x_0$  está entre  $T_0$  y  $T_1$ .
- c) Arme una función [camina](#)( $x_0$ ,  $T_0$ ,  $T_1$ ) que reciba  $x_0$ ,  $T_0$  y  $T_1$  y haga que el caminante se desplace desde  $x_0$  hasta tocar un borde  $T_0$  o  $T_1$ . Esta función debe retornar cuantos pasos dio el caminante y a qué posición llegó ( $T_0$  o  $T_1$ ).
- d) Genere 1000 caminatas con  $x_0=0$ ,  $T_0=-1$ ,  $T_1=10$ . ¿Cuántas caminatas llegaron a  $T_0$  y cuántas a  $T_1$ ? ¿Cuánto duran las caminatas que llegan a  $T_0$  y a  $T_1$ ?
- e) Arme una función [calcular\\_recorrido](#)( $x_0$ ,  $T_0$ ,  $T_1$ ) que retorne el recorrido de la caminata. Gráfiquelo haciendo [plot](#)([recorrido](#)). ¿Concuerda con lo obtenido en el ítem d)?