

## 5. Bases de Datos

### 5.1. Estructuras

R maneja diferentes estructuras de datos que se identifican con base en su tamaño y contenido. Las tres más utilizadas para el análisis estadístico son: vectores, matrices y tablas (Data Frames).

Un **vector** es un objeto que puede almacenar un solo dato o un conjunto de ellos, siempre y cuando sean del mismo tipo, lo que significa que no se pueden combinar valores numéricos con texto. Las **matrices** son un conjunto de vectores y para definir una matriz es necesario especificar el número de columnas y filas. Los datos (atributos) que se almacenan en estas estructuras se acomodan de manera ordenada y forman un rectángulo. Al igual que sucede en matemáticas, se pueden efectuar operaciones como la suma o multiplicación y modificar los valores de sus celdas.

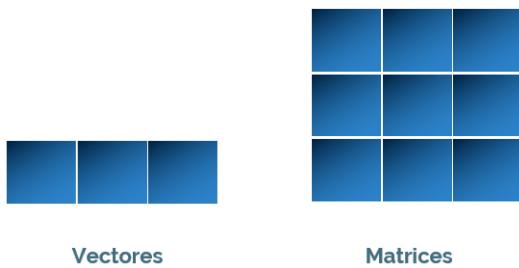


Figura 9: Estructuras de datos

Los **dataframes** son muy similares a las matrices y tienen las siguientes características: los renglones admiten distintos tipos de datos, pero sus columnas conservan la restricción de contener sólo un tipo de datos. Por ejemplo, en una columna se puede guardar el nombre de las personas y en otra su edad como sigue:


**Tabla (Data Frame)**

Figura 10: DataFrame

## 5.2. Subconjuntos

Para ubicar segmentos específicos de las bases de datos (M) se toma como referencia los números de filas y columnas **M[COLUMNAS , FILAS]**, similar a lo que sucede con un eje coordenado (x, y).

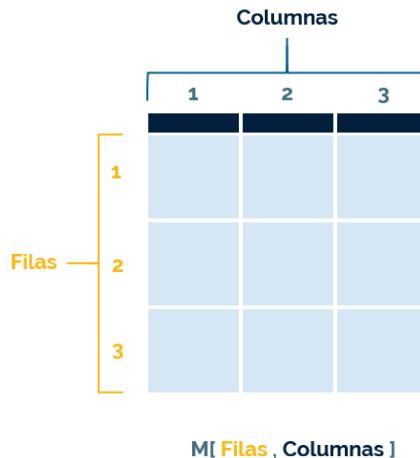


Figura 11: Organización de los DataFrame

De esta forma, es posible ubicar a columnas o filas de manera independiente, así como segmentos específicos de la base de datos.

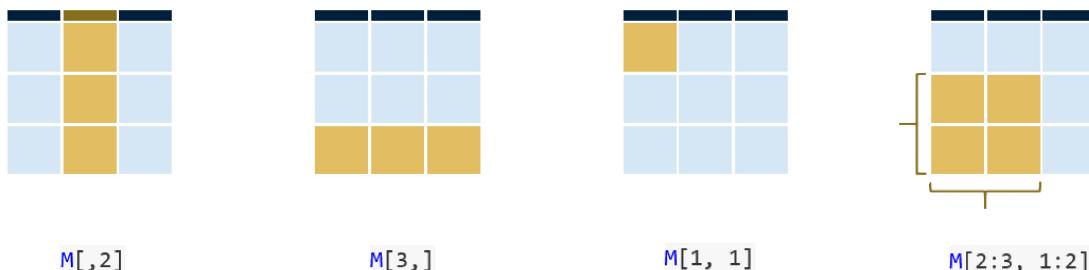


Figura 12: Subconjuntos

Otras formas de extraer una variable (columna) en particular son:

- Escribir el símbolo `$` y después el nombre del objeto
- Hacer referencia a la posición del objeto usando índices y corchetes dobles.

nombre	edad	sexo

**M\$edad**

**M[[2]]**

Figura 13: Selección de una columna

### 5.3. Operadores

Los operadores son símbolos que se utilizan para manipular los datos y desarrollar tareas específicas. Esto depende del tipo de dato que se esté manejando, por ejemplo, la suma solo es válida si se aplica a números, pero es irrelevante si se consideran texto. En la tabla siguiente se muestran los operadores más comunes que se utilizan en el análisis cuantitativo.

Clasificación	Símbolo	Tipo de operación	Ejemplo	Resultado
Aritméticos	+	Suma	$5 + 3$	8
	-	Resta	$9 - 8$	1
	*	Multiplicación	$5 * 3$	15
	/	División	$10 / 2$	5
	^	Potencia	$2 ^ 2$	4
Relacionales	<	Menor que	$4 < 5$	TRUE
	$\leq$	Menor igual que	$8 \leq 2$	FALSE
	=	Exactamente igual a	$3 == 3$	TRUE

### 5.4. Importar y exportar datos

Una de las características que ha hecho popular a R es la facilidad para establecer un vínculo con otros paquetes estadísticos. Esto lo hace mediante la importación de datos que se encuentran codificados en diferentes formatos, de esta manera puede trabajar con archivos de Stata (.dta), bases de datos (.dbf) o con archivos de uso más común como hojas de Excel.

Al romper esta barrera que imponen los formatos, R abre la posibilidad de analizar cualquier base de datos sin prestar atención a su formato. De manera análoga, R puede exportar los datos a importar a un sinfín de formatos, aunque los archivos nativos tienen extensión RDS.

## 5.5. Estructuras de control

Las estructuras de control son condicionales que evalúan un argumento y, dependiendo de la respuesta, continúan con un determinado flujo. Existen varias estructuras en R, sin embargo, las dos más utilizadas son las denominadas como IF-ELSE y FOR. La primera de ellas (IF-ELSE) es útil cuando deseamos que una acción se ejecute únicamente cuando se satisface cierta condición.

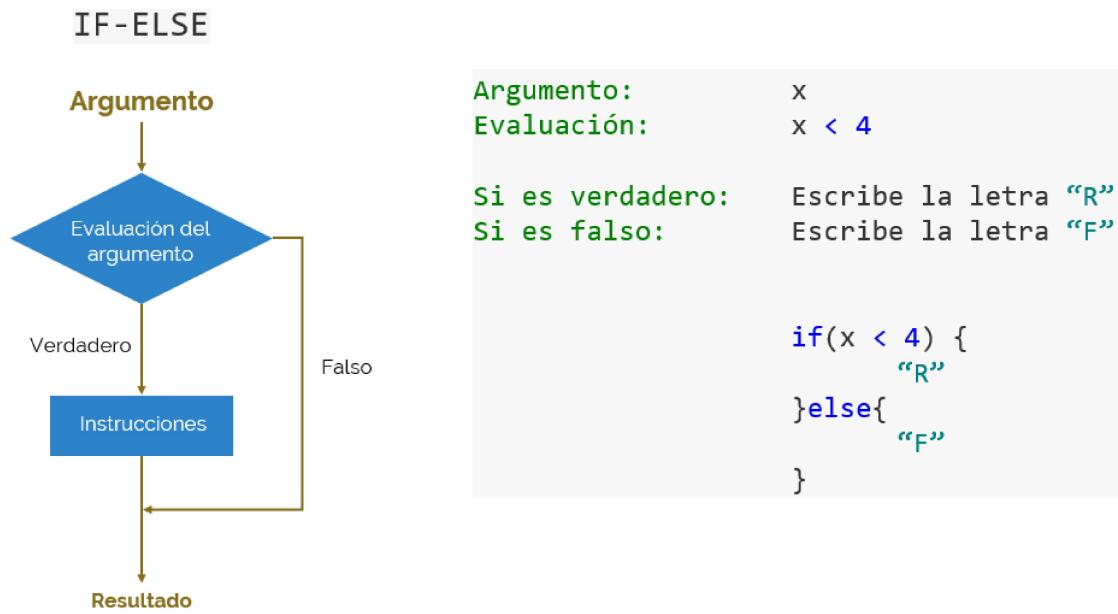


Figura 14: Ciclo IF

La segunda es la secuencia FOR, la cual nos permite hacer un bucle (loop) y ejecutar el código que está dentro de él, tantas veces como se le indique.

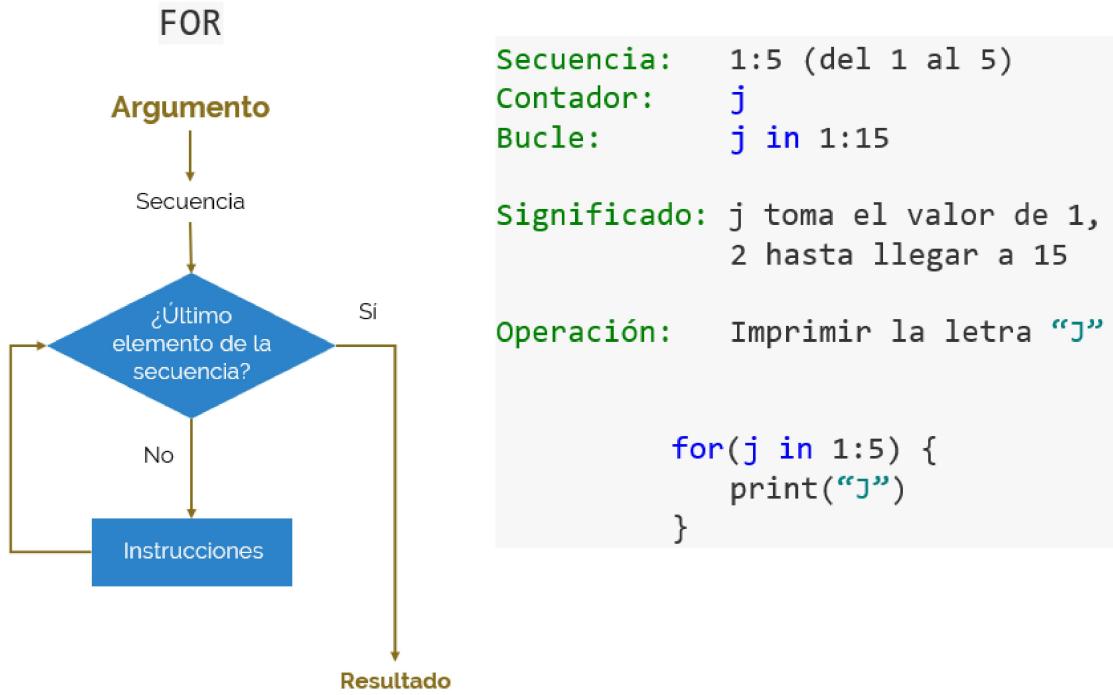


Figura 15: Ciclo FOR