

Session 4: Data Structures in R

Jamuna S Murthy
Assistant Professor
Department of CSE

Academic Year - 2025-26

1 Introduction

R offers various data structures to store and manipulate data. Understanding these is fundamental to effective programming in R.

2 Vector

A **vector** is a one-dimensional array that holds data of the same type.

```
1 # Numeric vector
2 v <- c(1, 2, 3, 4)
3
4 # Character vector
5 names <- c("Alice", "Bob", "Charlie")
6
7 # Logical vector
8 flags <- c(TRUE, FALSE, TRUE)
```

3 Matrix

A **matrix** is a two-dimensional array with elements of the same type.

```
1 m <- matrix(1:9, nrow=3, ncol=3)
2 print(m)
```

4 Matrix

A **matrix** is a two-dimensional array with elements of the same type.

```
1 m <- matrix(1:9, nrow=3, ncol=3)
2 print(m)
```

A matrix is created using the `matrix()` function, where:

-
- `1:9` generates a sequence of numbers from 1 to 9.
 - `nrow=3` sets the number of rows to 3.
 - `ncol=3` sets the number of columns to 3.

R fills the matrix column-wise by default.

5 Array

An **array** can have more than two dimensions, with all elements of the same type.

```
1 a <- array(1:12, dim = c(2, 3, 2))
2 print(a)
```

The `array()` function takes a vector of values and arranges them into multidimensional structures.

- `1:12` is the data.
- `dim = c(2, 3, 2)` creates a 3D array with 2 rows, 3 columns, and 2 layers.

Arrays are useful when working with higher-dimensional structured data.

6 List

A **list** can hold elements of different types (like a mixed collection).

```
1 my_list <- list(name = "Alice", age = 25, scores = c(90, 85, 88))
2 print(my_list)
```

A list can contain elements such as numbers, strings, vectors, or even other lists. Useful when dealing with grouped but heterogeneous data.

7 Data Frame

A **data frame** is a two-dimensional table where each column can have a different data type.

```
1 df <- data.frame(Name = c("Alice", "Bob"),
2 Age = c(25, 30),
3 Passed = c(TRUE, FALSE))
4 print(df)
```

Each column can be a different type: numeric, character, logical, etc. It's one of the most used structures for real-world tabular datasets.

8 Factor

A **factor** is used to represent categorical data.

```
1 gender <- factor(c("Male", "Female", "Female", "Male"))
2 levels(gender)
```

Factors are used to store categorical data efficiently. Internally, they are stored as integers with labels. Useful for grouping, statistical modeling, and plotting.

9 Summary Table

Structure	Dimension	Homogeneous?	Use Case
Vector	1D	Yes	Simple sequences
Matrix	2D	Yes	Mathematical computations
Array	nD	Yes	Multidimensional data
List	1D	No	Mixed-type storage
Data Frame	2D	No (by column)	Tabular data
Factor	1D	Categorical	Representing categories

Practice Questions

1. Create a numeric, character, and logical vector in R.
2. Create a 3x3 matrix with values 1 to 9.
3. Construct a list with your name, age, and a vector of scores.
4. Create a data frame with student names, ages, and pass status. Print it.
5. Convert a character vector of gender values into a factor and display its levels.