DATA STRUCTURES IN R CHEAT SHEET

Data Structures In R

Data Structure

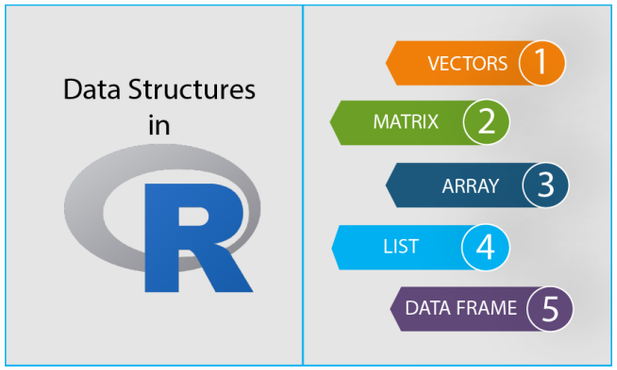
It is a way of organizing data that contains the items stored and their relationship to each other

R Programming

It is a **programming language** which is mainly used by Data Scientists, it is a preferred by the people who are good at Statistics and mathematics. In this language functions and codes are stored in a package inside the library

Data Tables

It extends and enhances the functionality of Data Frames



Syntax for the use of R Data Structures

Vector:

- To create vector:
  `v1 < - c (1,2,3)`
- Get length
  `length(v1)`
- Check if all or any is true
  `all(v1); any(v1)`
- Integer indexing
  `v1[1:3]; v1[c (1<6)]`
- Boolean indexing
  `v1[is.na(v1)] < - 0`
- Naming:
  `c(first = 'a', ..) or names(v1) < -c('first', ..)`

Lists:

- To create the list:
  `list1 < - list (first = 'a', … )`
- Create empty list
  `vector (mode = 'list' , Length = 3)`
- Get element:
  `list1[[1]] or list1[['First']]`
- Append using numeric index
  `list1 [[6]] < - 2`

Matrix:

- To create Matrix:
  `matrix1 < - matrix(1:10, nrow = 5 )`
  # fills rows 1 to 5, column 1 with 1:5, and column 2 with 6:10
- Matrix multiplication:
  `matrix1 %*% t (matrix2)`
  # where t() is transpose

Data Frames:

- To create data frame:
  `df1 < - data.frame (col1=v1, col2=v2, v3)`
- Dimension:
  `nrow(df1); ncol(df1); dim(df1)`
- Get/set column names
  `rownames(df1)`
  `rownames(df1) < - c(… )`
- Preview
  `head(df1, n=10) ; tail(… )`
- Get data types:
  `class(df1) # is data.frame`
- Index by columns
  df1['col1'] or df1[1]
  `df1[ c('col1', 'col3')] or df1[ c(1,3)]`
- Index by rows or columns
  `df1[ c(1,3), 2:3]`
  # returns data from rows 1,3 and columns 2,3
- To create data table from data.frame:
  `data.table(df1)`
- Index by columns:
  `dt1[, 'col1' , with= FALSE] or dt1[, list (col1)]`
- Show info for each data.table in memory:
  `tables()`
- Show keys in data.table:
  `key(dt1)`
- Create index for col1 and reorder data according to col1:
  `setkey(dt1,col1)`
- Use key to select data:
  `dt1[c('col1value1', 'col1value2',]`
- Multiple key select:
  `dt1[J('1', c('2', '3')), ]`
- Aggregation:
  `dt1[, list(col1=mean(col1)), by = col2 ]`
  `dt1[, list(col1=mean(col1), col2sum= sum(col2)), by = list(col3, col4) ]`

Types of R objects

- **Vector**: The basic data structure in R is Vector, it comes in two parts
  - Atomic vector and
  - List
  - Basic way of using vectors is by c () functions E.g.: C (1,2,3)
- **Matrix**: A matrix is a collection of numbers arranged into an affixed number of rows and columns. By using a matrix function we can reproduce a memory representation of the matrix in R
- **Array**: In R it is called multi-dimensional data structure. Here, the data is stored in the form of matrices. Array in R is the data object which can store data in more than two dimensions
- **List**: These are the objects which contain elements of different types like string, numbers, vectors and another list inside it. It can be created using list () function
- **Data Frames**: It refers to the tabular form of data, representing the cases (rows), each of which consists of number of observations or measurements (columns). It is used for storing data tables, it is a list of vectors of equal length

rajeshkch4u@gmail.com