

A Survey on Data Manipulations in R

Dr. N. R. Ananthanarayanan

Associate Professor

Department of Computer Engineering

SCSVMV, Kanchipuram, India

Abstract— In engineering, reference is made to physical quantities, such as force, speed and time. For instance speed of a car, speed of a motorcycle. It is useful to separate these physical quantities into two types. Quantities which are referred to as first type are known as scalars. These can be fully described by a single number known as the magnitude. Quantities of the second type are those which require the specification of a direction, in addition to a magnitude, before they are completely described. These are known as vectors. Special methods have been developed for handling vectors in calculations, giving rise to subjects such as vector algebra, vector geometry and vector calculus. In this paper highlights on manipulating data in R programming.

Keywords: Data Manipulations in R, Numeric Vectors, Integer Vectors

I. INTRODUCTION

A vector is the simplest type of data structure in R. Simply put, a vector is a sequence of data elements of the same basic type. Members of a vector are called Components. There are several types of vectors, such as

- 1) Numeric vectors, containing all kind of numbers.
- 2) Integer vectors, containing integer values. (An integer vector is a special kind of numeric vector.)
- 3) Logical vectors, containing logical values (TRUE and/or FALSE)
- 4) Character vectors, containing text
- 5) Datetime vectors, containing dates and times in different formats
- 6) Factors, a special type of vector to work with categories.[1]

II. CREATION OF VECTOR IN R

Vectors are generally created using the c() function. Since, a vector must have elements of the same type, this function will try and coerce elements to the same type, if they are different.

```
x <- c(1, 5, 4, 9, 0)
```

typeof(x)

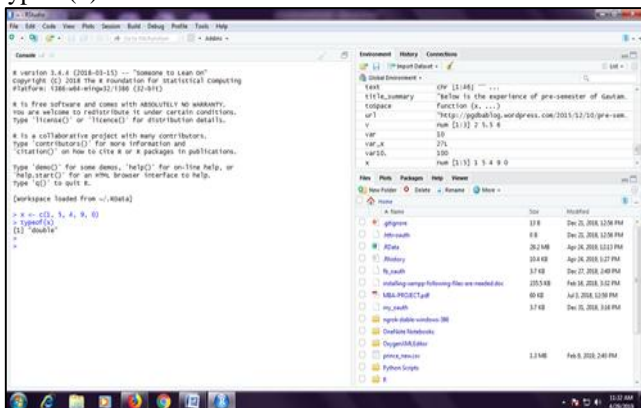


Fig. 1: Creating a Vector in R

III. CREATING A VECTOR USING: OPERATOR

To create a vector with consecutive numbers the : operator is used.

Program Code:

```
X<-1:7;x
```

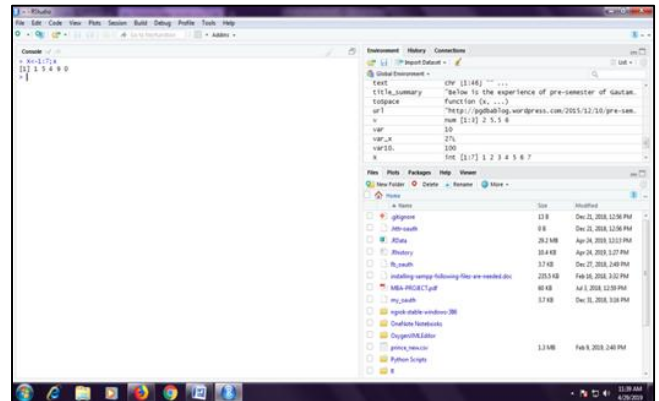


Fig. 2: Colon Operator for Sequencing Numbers

IV. SEQ() FUNCTION

More complex sequences can be created using the seq() function, like defining number of points in an interval, or the step size.

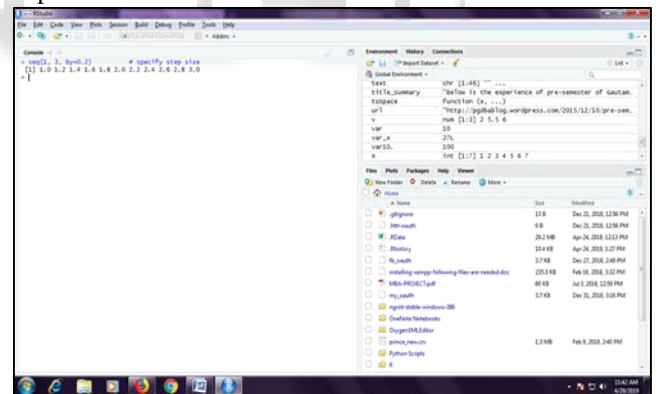


Fig. 3: Complex Sequence using Seq() Function

V. VECTOR INDEX

Vector index means selecting a subset of the elements in order to use them in further analysis or possibly change them. There are three types of vector indexing.

A. Positional Indexing

Each element from the vector by placing its position in brackets:[2]

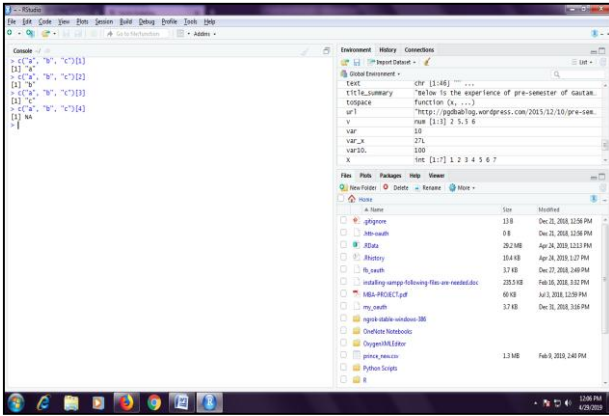


Fig. 4: Positional Indexing

B. Named Indexing

A second approach is vector elements are accessed by assigned names, such that each element has a value but also a name attached to it:[2]

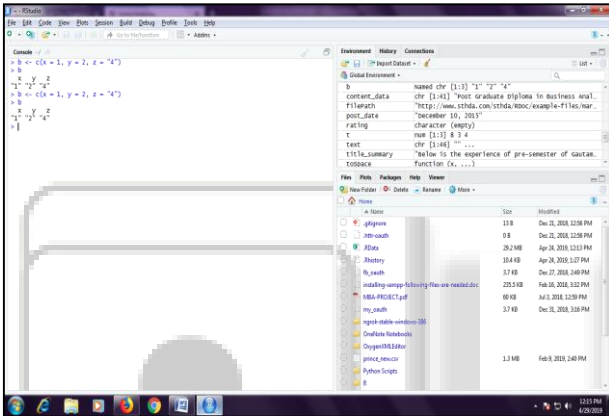


Fig. 5: Named Indexing

C. Logical Indexing

Logical indexing allows us to extract elements that meet specified criteria, as specified by an R logical expression. Thus, with a given vector, we could, for example, extract elements that are equal to a particular value:

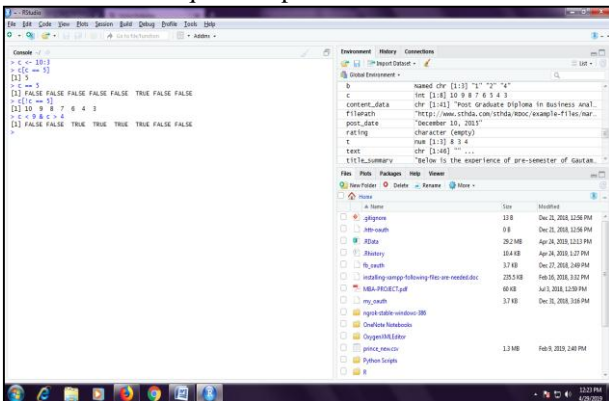


Fig. 6: Logical Indexing

1) Single Element Vector

Inserting one value in R becomes vector length 1[3]

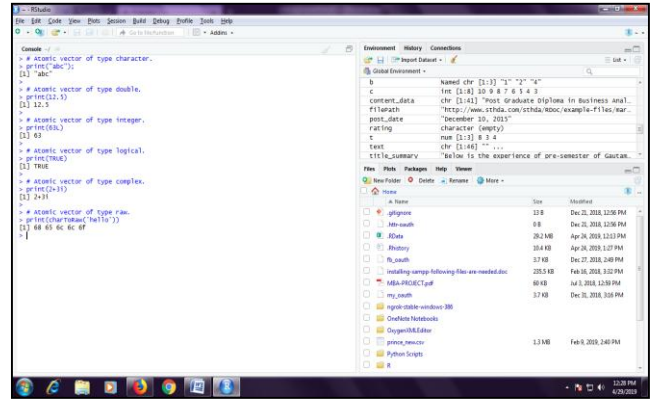


Fig. 7: Single Element Vector

2) Multiple Element Vector

Using colon operator with numeric data[3]

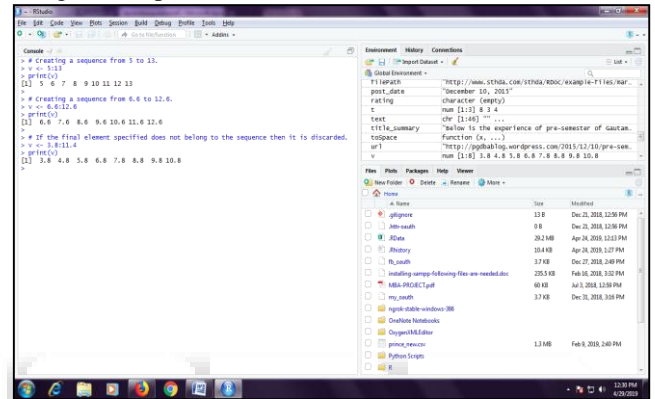


Fig. 8: Multiple Element Vector

3) Matrix in R

A matrix is a two-dimensional rectangular data set and thus it can be created using vector input to the matrix function. In addition, a matrix is a collection of numbers arranged into a fixed number of rows and columns. Usually, the numbers are the real numbers. Hence, the data elements must be of the same basic type[4].

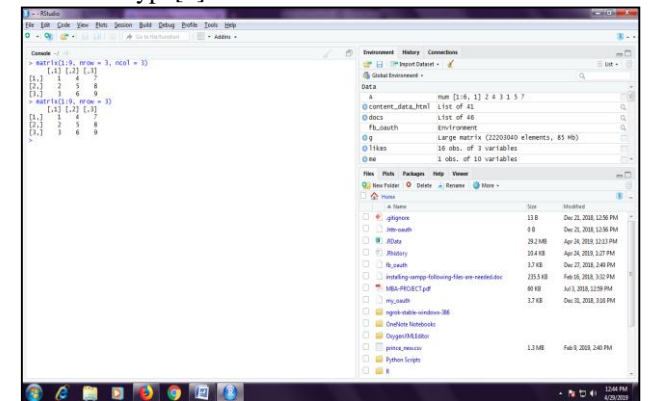


Fig. 9: Creating values using Matrix() in R

One can see that the matrix is filled column-wise. This can be reversed to row-wise filling by passing TRUE to the argument byrow.[5]

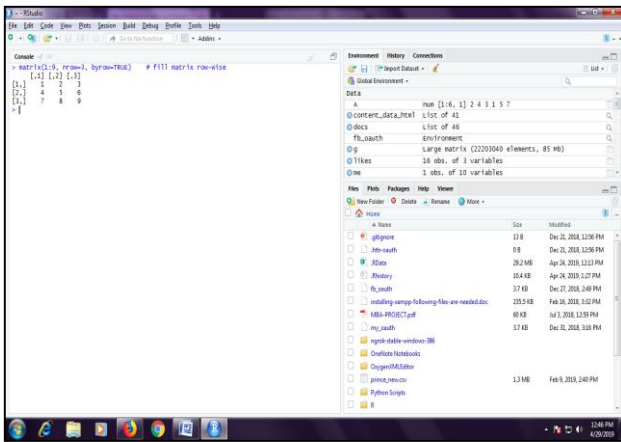


Fig. 10: Matrix Determinant

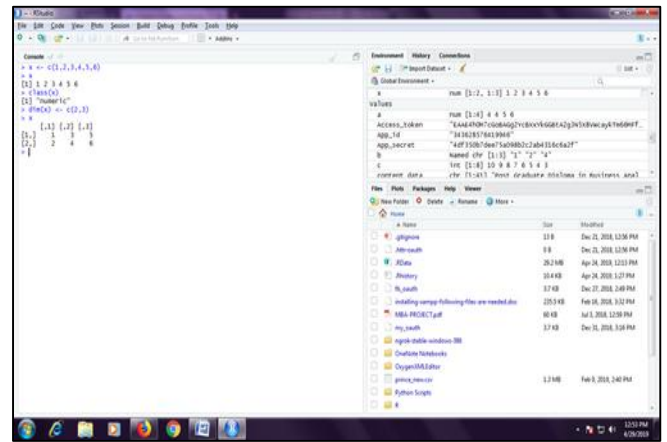


Fig. 12: Matrix creation using dim statement

4) Naming Rows and Columns in a matrix

It is possible to name the rows and columns of matrix during creation by passing a 2 element list to the argument `dimnames[5]`

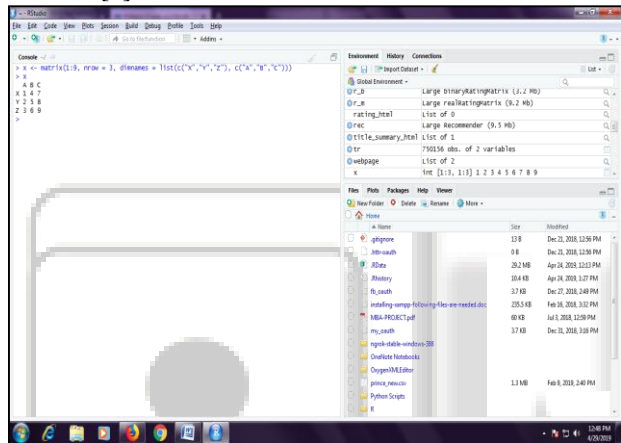


Fig. 11: Rows and column names using dimnames

5) Creating a matrix using Cbind() and Rbind() Function[5]

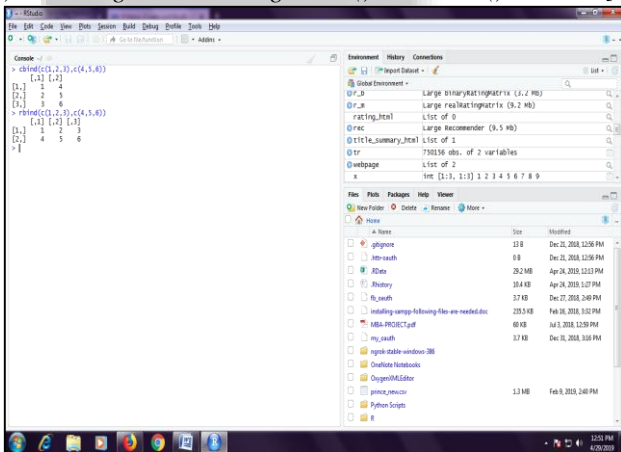


Fig. 11: Cbind() and Rbind() Function

Matrix of vectors can be created by setting dimension using `dim()` statement[5]

6) Accessing Vector Elements in Matrix using

One can access elements of a matrix using the square bracket [indexing method]. Elements can be accessed as `var[row, column]`. Here rows and columns are vectors. [5]

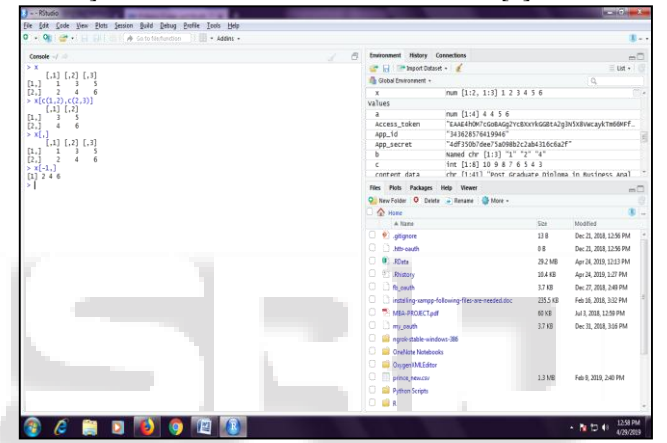


Fig. 13: Indexing Method in Matrix

if the matrix returned after indexing is a row matrix or column matrix, the result is given as a vector[5].

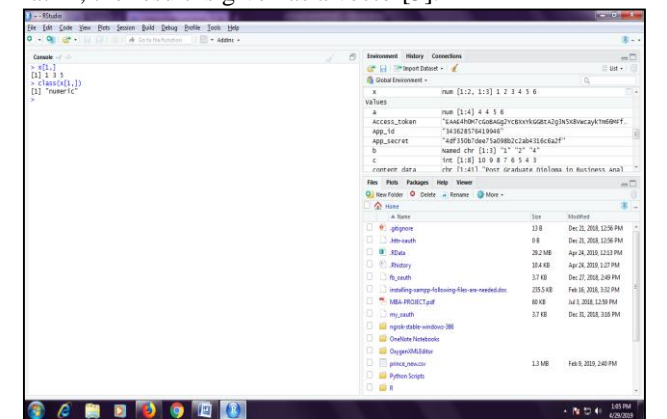


Fig. 14: Row matrix

7) Matrix Computations

Various mathematical operations are performed on the matrices using the R operators. The result of the operation is also a matrix. The dimensions (number of rows and columns) should be same for the matrices involved in the operation[6].

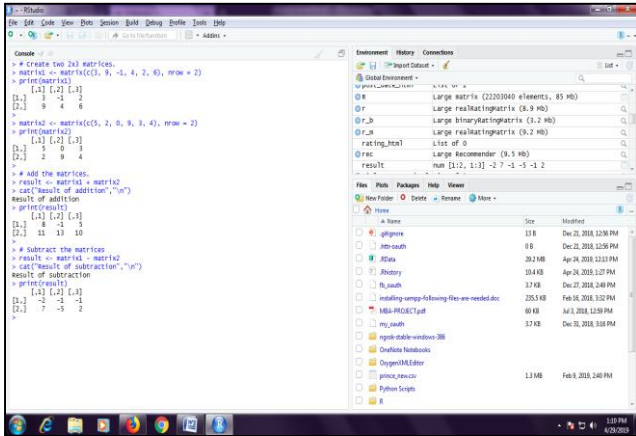


Fig. 15: Matrix addition and subtraction

8) Factors in R

Factors are the data objects which are used to categorize the data and store it as levels. They can store both strings and integers. They are useful in the columns which have a limited number of unique values. Like "Male", "Female" and True, False etc. They are useful in data analysis for statistical modeling. Factors are created using the factor () function by taking a vector as input.[7]

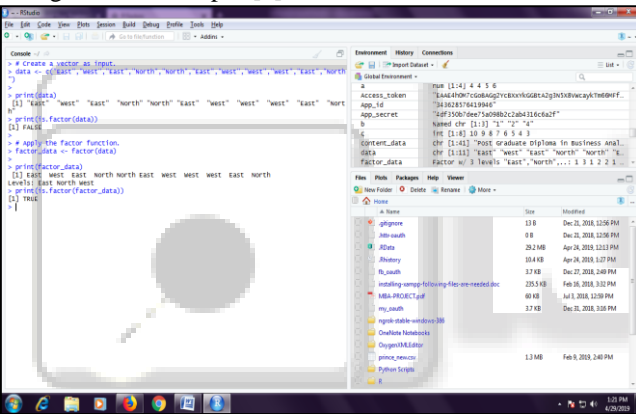


Fig. 16: Using Factor Method

9) Factors in Data Frame

On creating any data frame with a column of text data, R treats the text column as categorical data and creates factors on it [7].

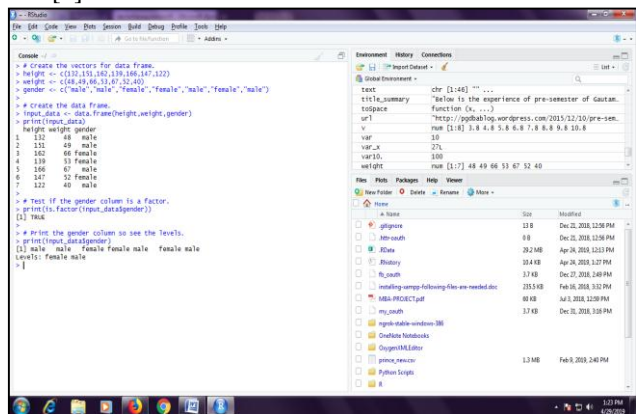


Fig. 16: Using Factors in Data Frame

10) Changing order of levels using Factors[7]

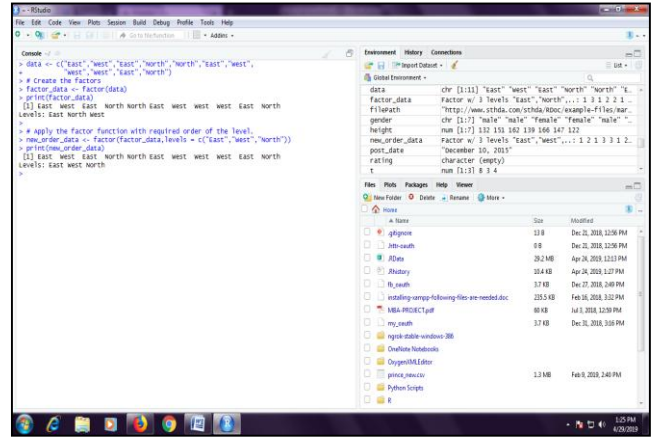


Fig. 17: Factor Levels in R

REFERENCES

- [1] <https://datascienceplus.com/vectors-and-functions-in-r/>, Vectors and Functions in R Published on March 28,2018 at 8.00 a.m.
- [2] <https://thomasleeper.com/Rcourse/Tutorials/vectorindexing.html>
- [3] https://www.tutorialspoint.com/r/r_vectors.htm
- [4] <https://data-flair.training/blogs/r-matrix-operations-applications/>
- [5] <https://www.datamentor.io/r-programming/matrix/>
- [6] https://www.tutorialspoint.com/r/r_matrices.htm
- [7] https://www.tutorialspoint.com/r/r_factors.htm