manipalglobal
Academy of Data Science

Introduction to Data Types and Data Structures in R

PRESENTED BY

Vinod Raju
Data Science Practitioner

TOC

Data Types
Data Structure
Summary

Topic Number	Topic Name
1	Data Types
2	Data Structure
3	Summary

TOC

Data Types
Data Structure
Summary

Learning Objectives

By the end of this unit, you will be able to:

- Perform basic programming operations
- Explain various data types supported in R
- Identify various data structures

TOC

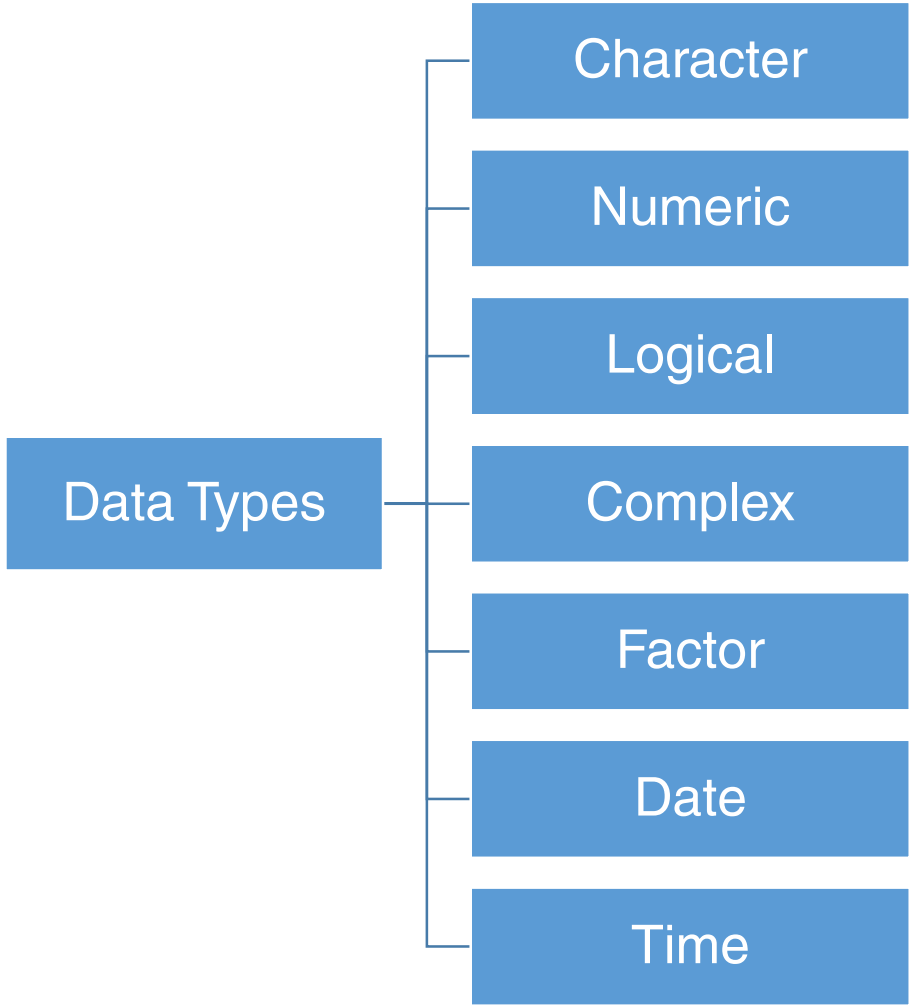
Data Types
Data Structure
Summary

Data Types

TOC

Data Types
Data Structure
Summary

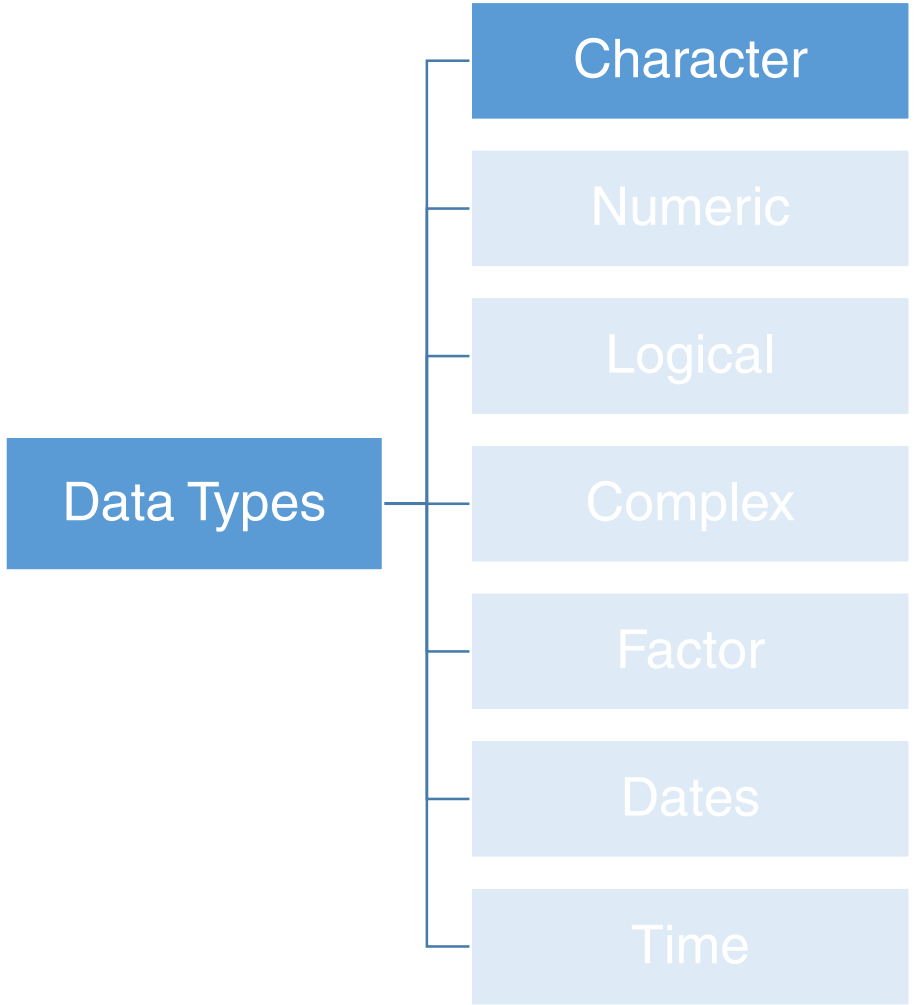
Data Types



TOC

Data Types
Data Structure
Summary

Data Types



Letters, words, group of words or any value enclosed in " " double quotes or single quotes, are treated as a character literals in R.

Example:

```
s <- "string"
class(s)
```

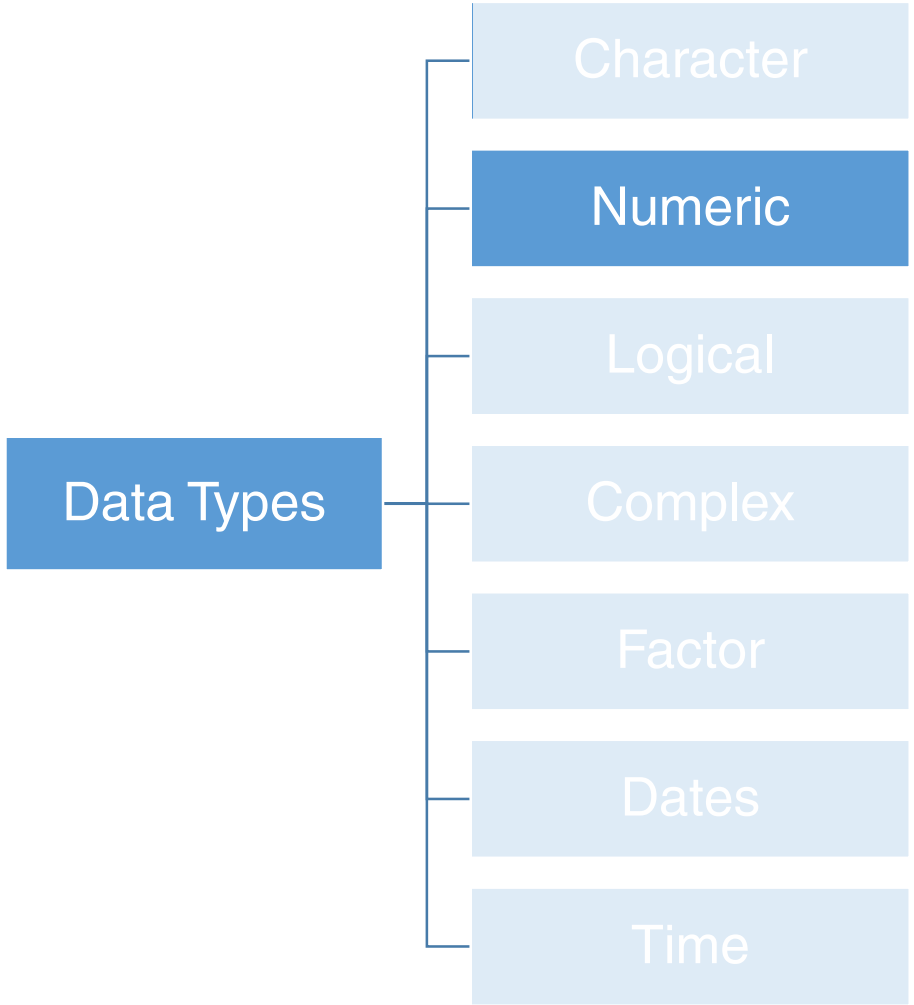
Output would be …

```
[1] "character"
```

TOC

Data Types
Data Structure
Summary

Data Types



Numeric type is the base class that includes integers and decimal numbers under its family. This data type is used to represent numbers.

Example:

```
s <- 23.5
class(s)
```

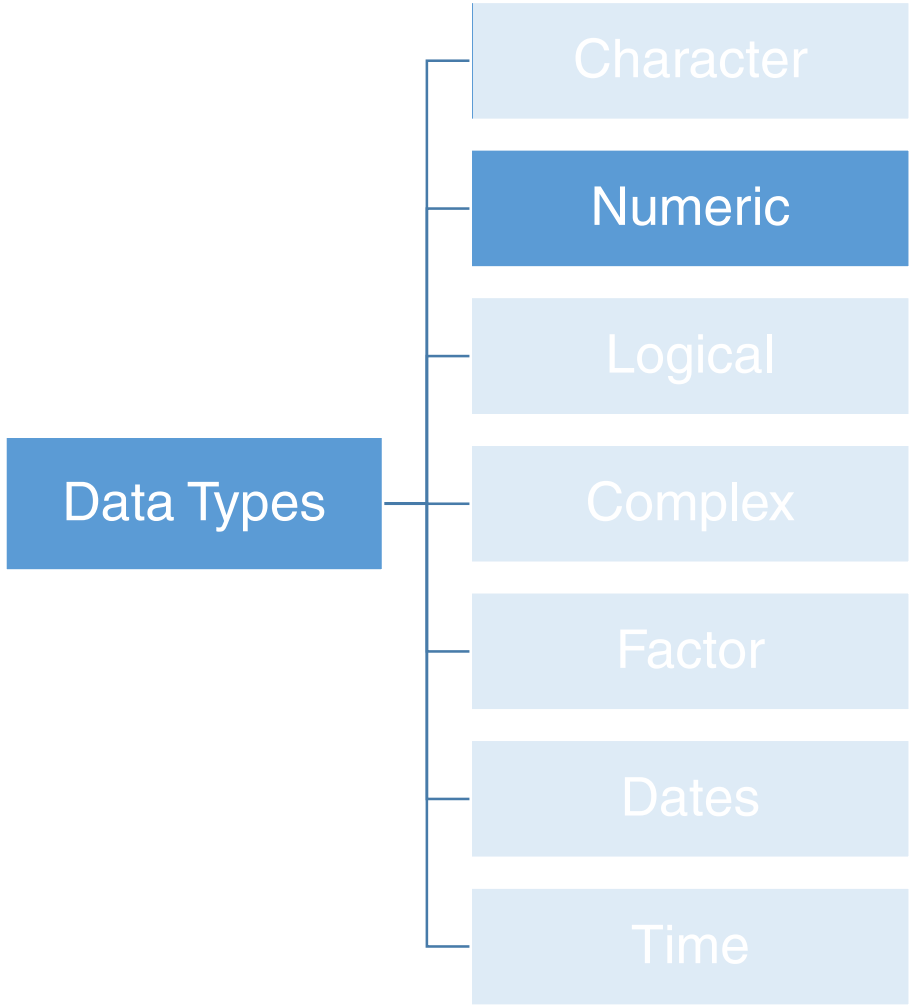
Output would be …

[1] “numeric”

TOC

Data Types
Data Structure
Summary

Data Types



Integers are also treated as Numeric data type.

Explanation:

- as.integer – To explicitly typecast it to integer
- is.integer – To check if a variable is integer or not

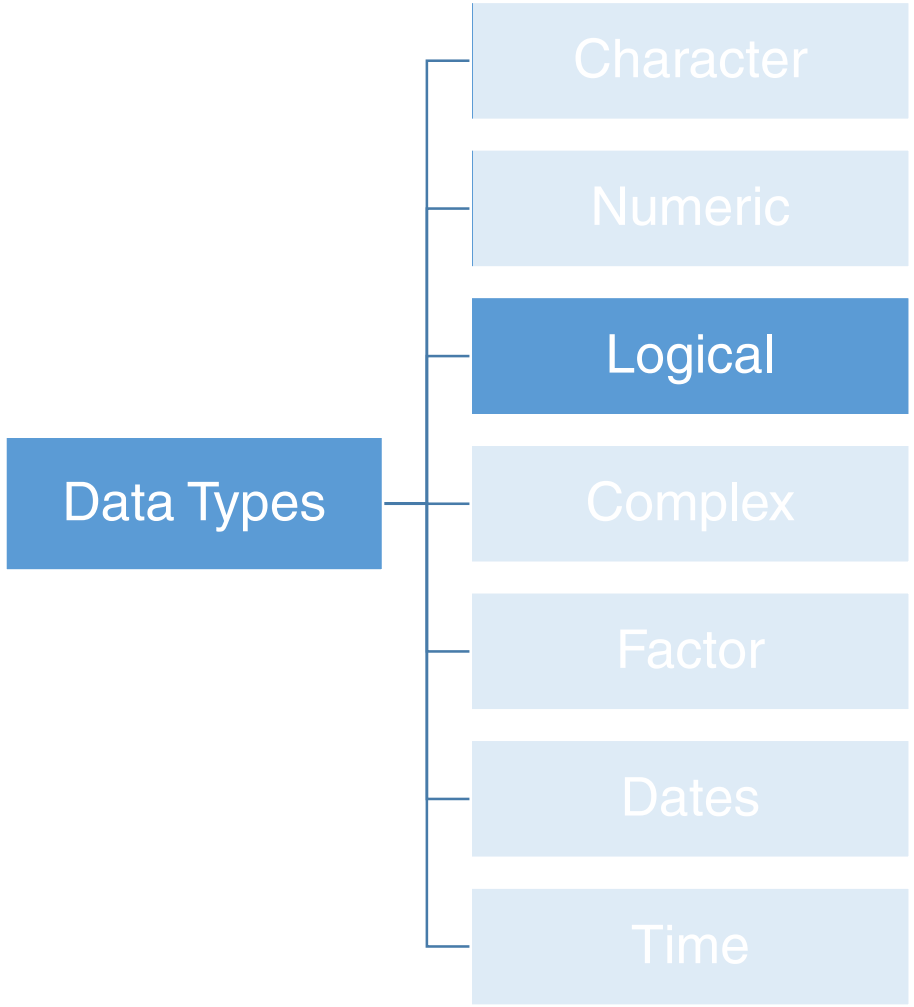
TOC

Data Types

Data Structure

Summary

Data Types



Logical data type takes only two values, either TRUE or FALSE, which has to be in upper case only.

Example:

```
s <-TRUE
class(s)
```

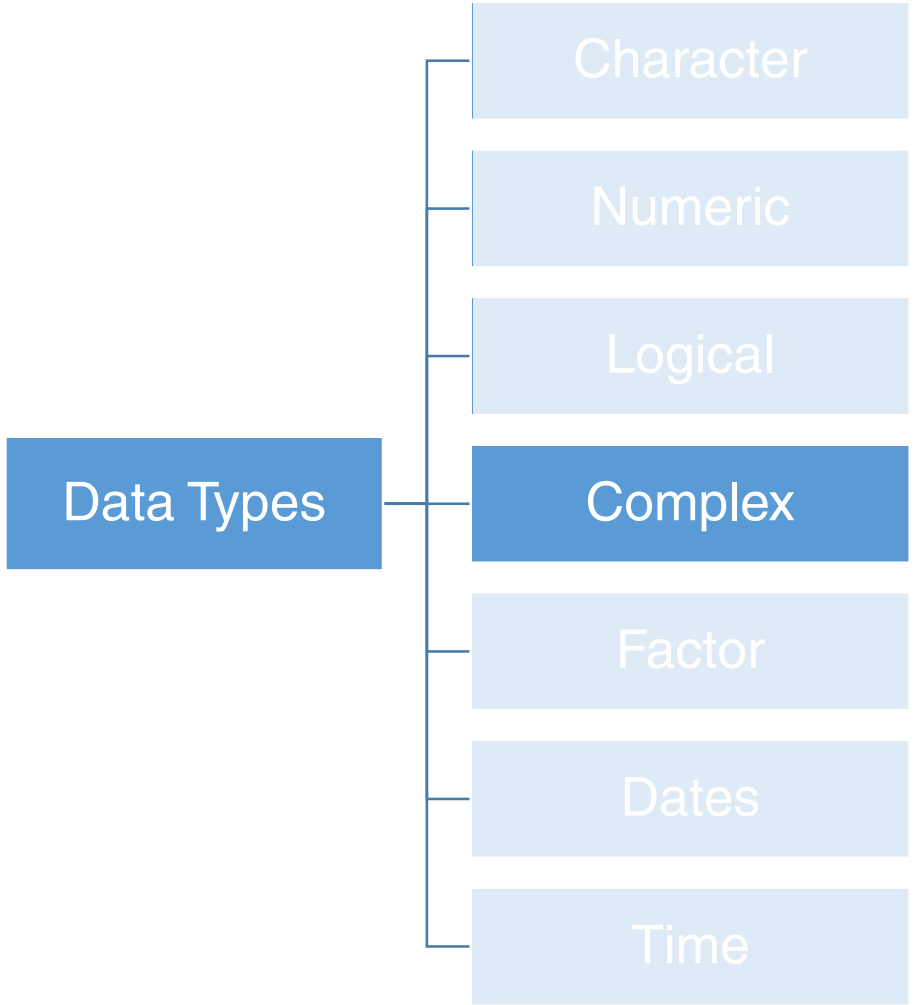
Output would be …

```
TRUE
```

TOC

Data Types
Data Structure
Summary

Data Types



Complex data types are of the type 3 + 2i, that is, they are used to represent complex numbers in mathematics.

Example:

```
s = 1 + 8i
class(s)  # Checking the class of “s”
```

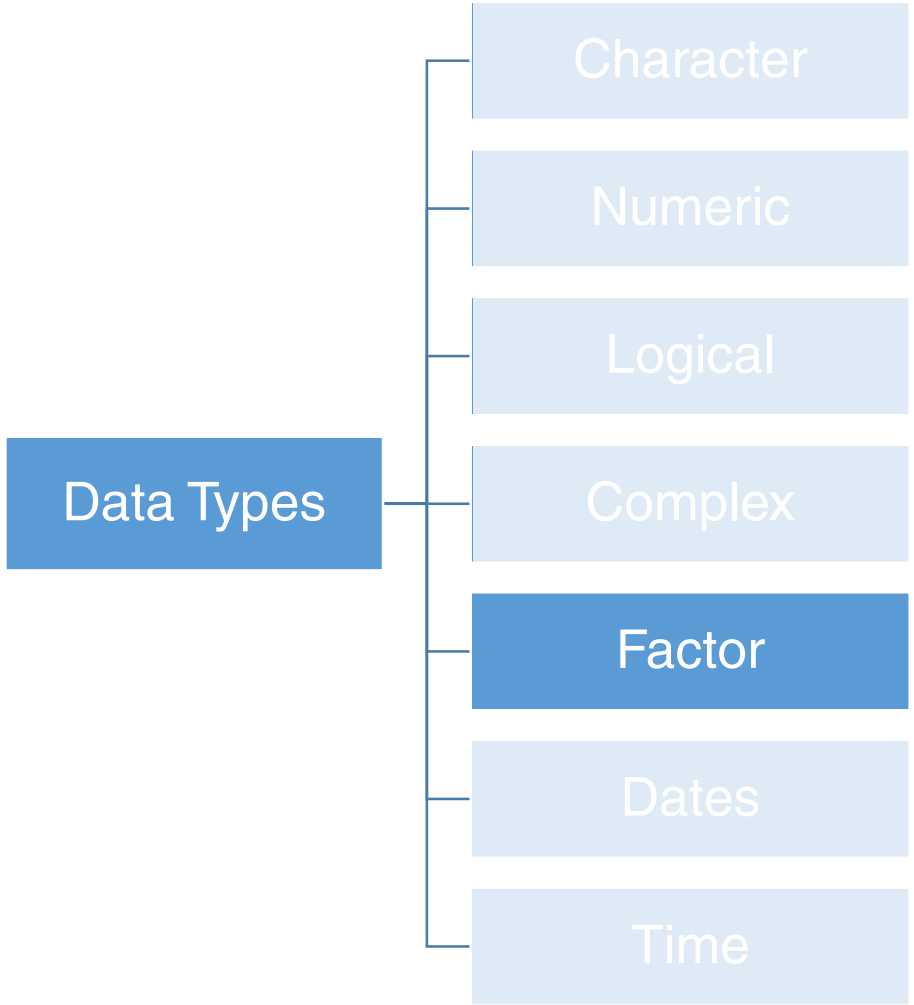
Output would be …

```
[1] "complex"
```

TOC

Data Types
Data Structure
Summary

Data Types



Factors are the data objects, which are used to categorise the data and store it as levels. They can store both strings and integers. Factors are created using the factor() function by taking a vector as input.

Example:

```
f = c("male","female","male","female")
f=factor(f)
print(f)
```

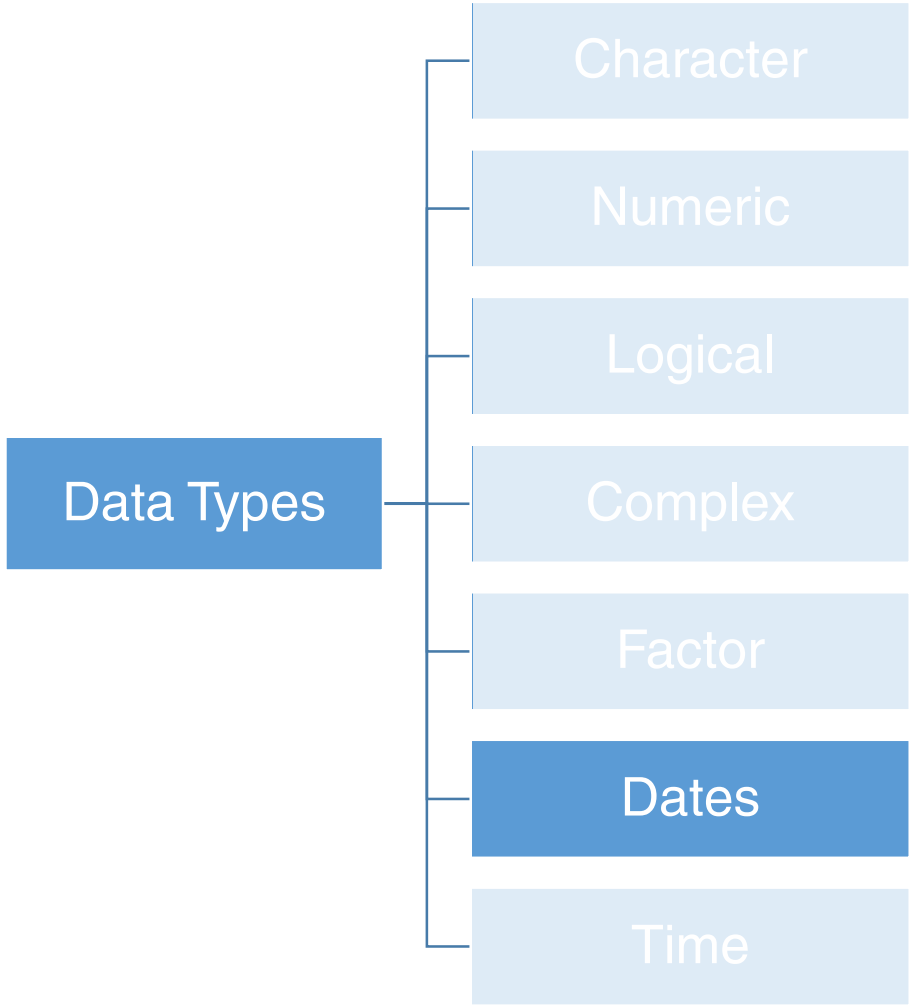
Output would be …

```
[1] male   female male   female
Levels: female male
```

TOC

Data Types
Data Structure
Summary

Data Types



R contains a set of object types for holding date and time information. The system time and date can also be requested.
Dates are represented as the number of days since 1970-01-01, with negative values for earlier dates.

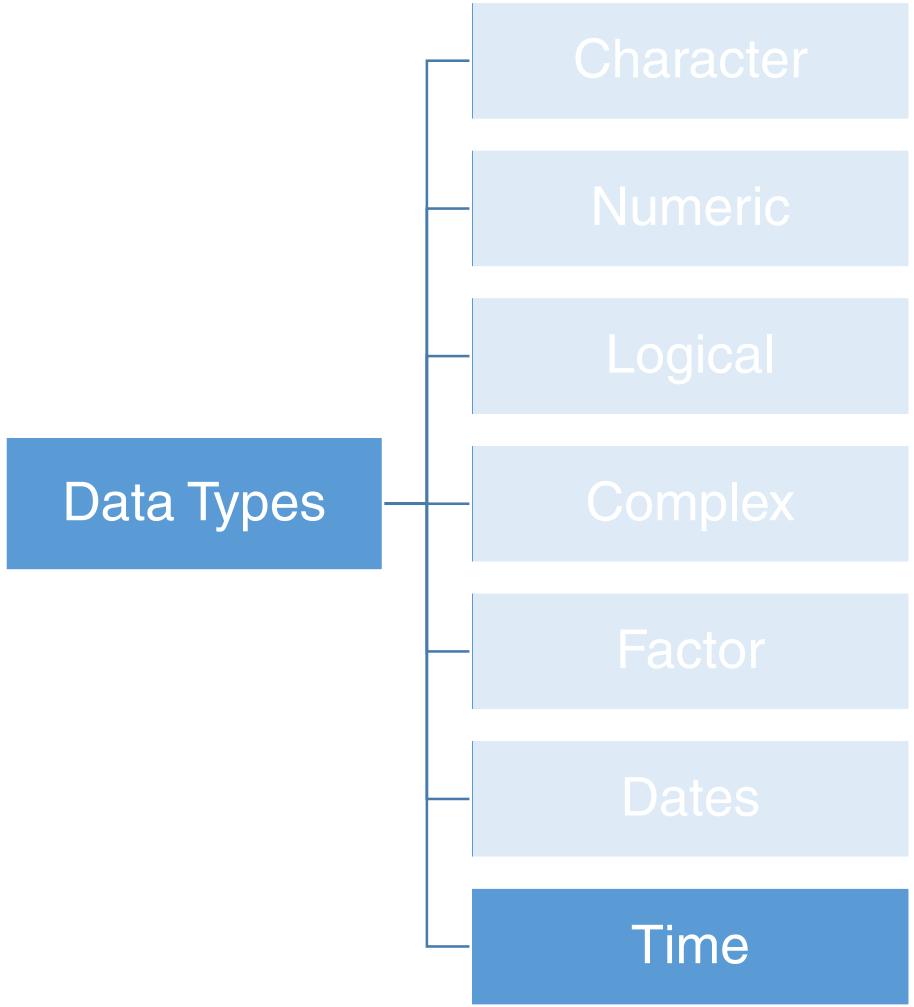
Example:
Sys.Date() #returns the current Date
Format.Date() #for conversion to and from character strings
Mydate=as.Date("31-08-2017")

Now check the class of "Mydate" object!

TOC

Data Types
Data Structure
Summary

Data Types



Sys.time returns an absolute date-time value, which can be converted in various time zones and may return different days.

Example:
Sys.time() # would return the system time

Output would be …
[1] "2017-09-21 19:06:21 IST"

TOC

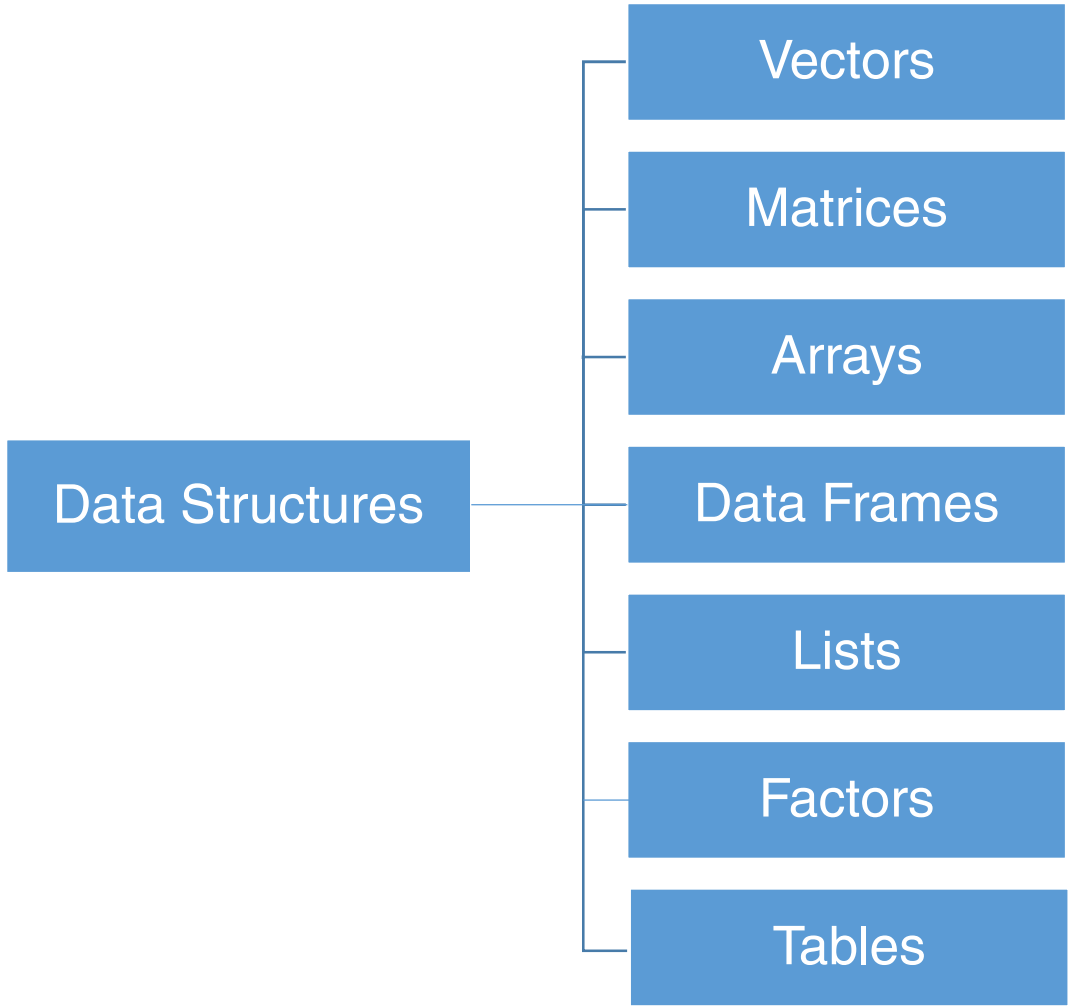
Data Types
Data Structure
Summary

Data Structures

TOC

Data Types
Data Structure
Summary

Data Structures



TOC

Data Types

Data Structure

Summary

Data Structures in R

Data structure in R includes:

Vector	A sequence of numbers or characters, or higher-dimensional arrays such as matrices
List	A collection of objects that may themselves be complicated
Factor	A sequence assigning a category to each index
data.frame	A table-like structure (experimental results often collected in this form)

TOC

Data Types

Data Structure

Summary

Vectors

Vectors are three types:

Numeric	*a <- c(2, 2.4, 3.0, 8.9,10.0)*
Character	*b <- c("Alpha", "Beta", "Gamma")*
Logical	*c <- c(TRUE, FALSE, TRUE, TRUE, TRUE, FALSE)*

Refer the element(s) of a Vector by Subscript(s):

a[1] → returns '2', the first element of Vector a

a[c(2,4)]→returns '2.4' and '8.9', the 2nd and 4th elements of vector 'a'

TOC

Data Types

Data Structure

Summary

Arrays

Arrays are similar to matrices, except that, they have more than two dimensions.

In arrays, the vector elements are arranged into rows and columns as specified by the user.

Example:

```
a=1:4
d=c(2,2)
myarr=array(a,d) #This would create a 2 by 2 array (2 rows and 2 columns)
```

Matrices are more popular in R than arrays.

TOC

Data Types

Data Structure

Summary

Matrices

The column elements of a matrix should have the same mode, such as numeric, character and the same length, that is, they must be homogenous.

Pseudocode:

```
matrixSample <- matrix(vector, nrow=r, ncol =c, byrow=FALSE,
    dimnames=list(char_vector_rownames, char_vector_colnames))
```

- byrow=TRUE indicates matrix should be filled by rows.
- byrow=FALSE indicates matrix should be filled by columns (default).
- dimnames specifies labels for rows and columns (optional).

TOC

Data Types

Data Structure

Summary

Data Frames

- Data frames are similar to a matrix, except that, different columns in a data frame can have different data types (numeric, character, factor, and so on).
- Data frames can be seen as a collection of vectors, arranged into columns placed sequentially, with the column names same as vector names.

Example:

```
a <- c(101,102,103,104,105)
b <- c("VICTOR", "JOHN", "MARY", "CHLARA", NA)
c <- c(TRUE, TRUE, TRUE, FALSE, FALSE)
sampData <- data.frame(a,b,c)
```

sampData$ColumnName – Refer to a particular column in the data frame

TOC

Data Types

Data Structure

Summary

Data Frames

Example:

```
a = 1:5 #The object a, holds a sequence of integers from 1 to 5
b = 5:10
c = 11:16
data.frame(a,b,c) -> df
```

Accessing the data from the data frame

The data frame which was created (df) is composed of 3 vectors (a,b,c) which are 3 columns of the data frame.

- *print(df$a)* prints the first column from the data frame
- *print(dfa,dfb)* prints the first 2 columns from the data frame

TOC

Data Types

Data Structure

Summary

Data Frames

Accessing elements of a data frame:

Commands

Exract the rows from a data frame	*row1 = df[1,]*
Extract the first 2 rows from the data frame	*row1to2 = df[1:2,]*
Print all the elements of row 1 and 5	*row1and5 = df[c(1,5),]*
Pick only the 3rd coloumn elements of row 1 and 5	*row15andcol3 = df[c(1,5),3]*

TOC

Data Types

Data Structure

Summary

Lists

- A list is an ordered collection of objects.
- List allows collection of different varieties of objects under one name.

Example:

```
v1 =c("true","false");
v2= 1:5
v3 = c("cat","dog")
my_list=list(v1,v2,v3)
```

Output:

```
[[1]]
[1] "true" "false"
[[2]]
[1] 1 2 3 4 5
[[3]]
[1] "cat" "dog"
```

TOC

Data Types

Data Structure

Summary

Factors

- A factor contains a set of numeric codes with character valued levels.
- A variable in R can be made ‘nominal’ by declaring it as a ‘factor’.
- The factor variable then stores these nominal values as a set of integers in the range [1,2,3..].

Example:

```
d1 <- c("M", "F", "M", "F", "F", "F")
d2 <- factor(d1)
d2
```

Output:

```
[1] M F M F F F
Levels: F M
```

TOC

Data Types

Data Structure

Summary

Tables

Table() function is used to create tabular results of categorical variables. It is essentially used to interactively analyse the data and gain insights from it.

Example:

The vector which is created below is given as an input to the *table()* command.

```
a=c(1,1,3,4,5,5,5,3,3,2,10)
table(a)
```

Output:

```
1  2  3  4  5 10
2  1  3  1  3  1
```

The Table() function gives the frequency count of how many times each number is repeated in the above vector “a”.

TOC

Data Types

Data Structure

Summary

Tables

Tables created using a matrix:

Example:

```
sample = matrix(c(10,22,1,2),ncol=2)
colnames(sample) <- c('fruits', 'vegetables')
rownames(sample) <- c('fresh', 'not_fresh')
samp.table <-as.table(sample)
samp.table
```

Output:

	fruits	*vegetables*
fresh	*10*	*1*
not_fresh	*22*	*2*

TOC

Data Types
Data Structure
Summary

Summary

TOC

Data Types

Data Structure

Summary

Summary

In this unit, you learnt:

- Programming involves data types and data structure.
- Data types are:
  - Character
  - Numeric
  - Logical
  - Complex
  - Factor
  - Dates and time
- Data structure in R include:
  - Vectors
  - List
  - Factor
  - Data frame
  - Tables

manipalglobal
Academy of Data Science