R – Data Structures (Part 1)

Abhishek Kumar
ItsAbhishekKumar.com
@MeAbhishekKumar

Outline

Atomic vector	Data frame	List
Matrix	Factor	Array

Data Structures

- **Collection of data elements**

- **Grouped under one name**

- **Container**
  - What kind of items to put in ?
  - How to arrange items ?

Data Structures

What kind of items to put in ?

Similar type of items	**Dissimilar** type of items
Homogeneous data structure	**Heterogeneous** data structure
Atomic vector, Matrix, Array	**List, Data frame**

How to arrange items ?

| **Atomic vector** 1 D | **Matrix** 2 D | **Array** n D | **List** 1 D | **Data frame** 2 D |

Basic Classes of Objects
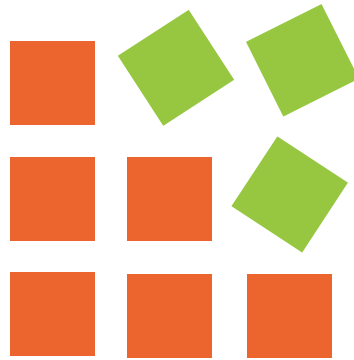
Atomic classes

Character	**Numeric**	**Integer**	**Logical**	**Complex**
"A", "c"	4.36, 7.42	3,5	True, False	1 + 7i , 8 - 2i

Atomic Vector

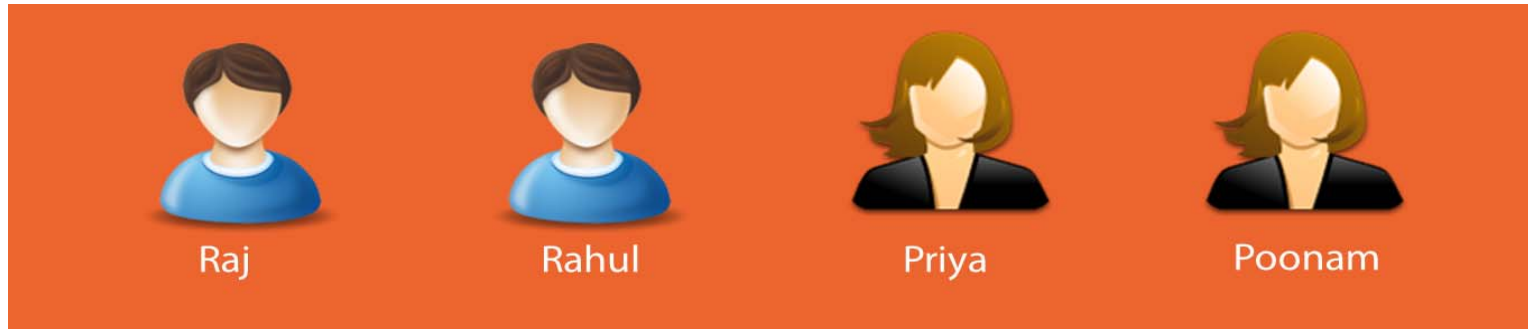- **Commonly knows as vectors**

- **Homogeneous** data structure

Character vector	Numeric vector	Integer vector	Logical vector	Complex vector
{"A","c"}	{ 4.36, 7.42 }	{ 3,5 }	{ True, False }	{ 1 + 7i , 8 - 2i }

Atomic Vector

- **1-Dimensional** arrangement of items

Element 1	Element 2	Element 3

Atomic Vector


```
student.names <- c("Raj","Rahul","Priya","Poonam")# Character

student.weights <- c( 60.5, 72.5 , 45.2,  47.5)# Numeric

student.physics.marks <- c( 70L , 75L , 80L,  85L)# Integer

student.physics.interest <- c(FALSE, F, TRUE, T) # Logical
```
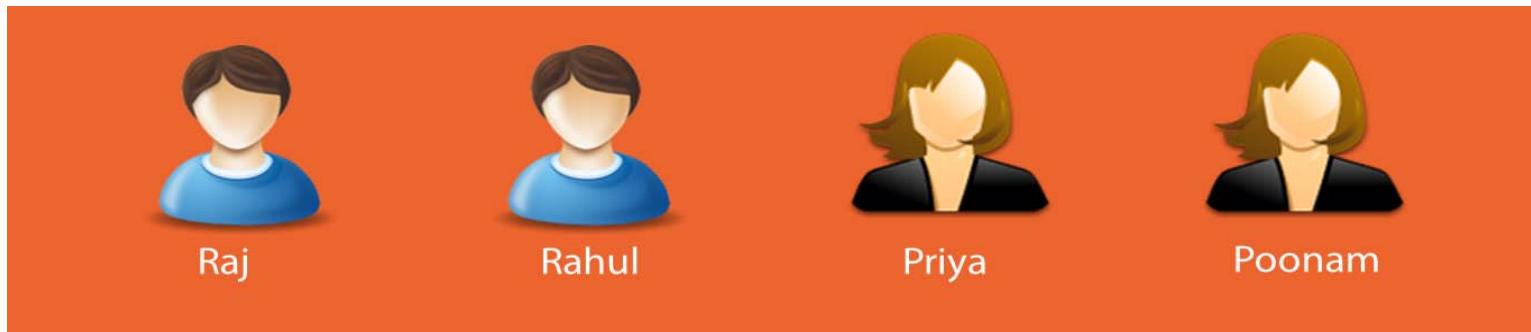
Common Operations on Atomic Vectors

Arithmetic & logical | Subsetting | Coercion

Factor

- Special case of vector used to store nominal values


```
# Using character vector
student.genders <- c("Male","Male","Female","Female")
# Using factor
student.genders <- factor(c("Male","Male","Female","Female"))
```

Factor

- More efficient than character vectors

- Self-describing than integer vectors

- Additional levels can be created

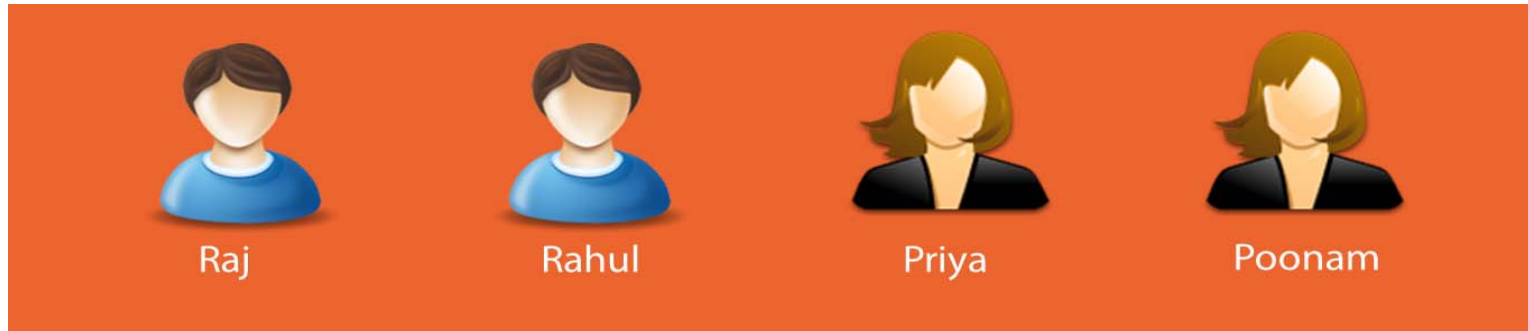- Some algorithms explicitly demand factors instead of character vectors

List

- **Heterogeneous** data structure

- **Can contain items of different classes**

- **1-Dimensional arrangement**

Element 1	Element 2	Element 3

List


```
student.names <- c("Raj","Rahul","Priya","Poonam")# Character

student.weights <- c( 60.5, 72.5 , 45.2,  47.5) # Numeric

student.genders <-
factor(c("Male","Male","Female","Female")) # Factor

student.physics.marks <- c( 70L , 75L , 80L,  85L)# Integer

student.chemistry.marks <- c(60L, 70L, 85L, 70L) # Integer
```

List


```
1  >  student.names[1]
[1] "Raj"
2  > student.weights[1]
[1] 60.5
3  > student.genders[1]
[1] Male
Levels : Female Male
4  > student.physics.marks[1]
[1] 70
5  > student.chemistry.marks[1]
[1] 60
```

Summary

Data structure

Homogeneous

Heterogeneous

Atomic vector
1 D

List
1 D

Factor
1 D