

Lecture 10 - R Software

Ashutosh Rajput
Department of Mathematics

Rajdhani College
University of Delhi

<https://rfidcard.cards>

The Structure of Data Items

The data can exist in a variety of forms (or structures). R recognizes several forms of data and these forms each have their own particular uses. Data can exist as numerical or character data, which can also be constructed and put together in a variety of ways. This section looks at these different structures: **vector**, **matrix**, **data frame**, and **list**.

- **Vector Items:** A vector can be thought of as a one-dimensional object.

Here are two simple vectors:

```
> data8 = scan(file.choose(), what = 'char', sep = ',')
[1] "Jan" "Feb" "Mar" "Apr" "May" "Jun" "Jul" "Aug" "Sep" "Oct"
     "Nov" "Dec"
```

```
> data7 = scan(file.choose())
Read 15 items
> data7
[1] 23.0 17.0 12.5 11.0 17.0 12.0 14.5 9.0 11.0 9.0 12.5 14.5 17.0 8.0 21.0
```

We can also have a vector of factors, we use **cut**(categorical variables)

```
> cut
[1] mow mow mow mow mow unmow unmow unmow unmow
Levels: mow unmow
```

Factors: Factors are variables in **R** which take on a limited number of different values; such variables are often referred to as categorical variables.

Datasets

In Datasets, some important terms are:

- Observations
- Variables

Example: People

- Each person = observation
- `properties(name, age,...)` = variables

Name	Age	Married
Anne	28	FALSE
Pete	30	TRUE
Frank	21	TRUE
Julia	39	FALSE
Cath	35	TRUE

Data Frames

A data frame is a two-dimensional object, that is, it has rows and columns. **R** treats the columns as separate samples or variables, and the rows represent the replicates or observations.

Some key points are:

- Specially for Datasets
- Rows = Observations(Eg: Persons)
- Columns = Variables(Eg: name, age,...)
- Contain Elements of different types(Character and Numeric)

Create Data frame

- Import from data source(CSV,SQL,Excel,...)
- We create this by using `data.frame()`

```
> name = c("Anne", "Pete", "Frank", "Julia", "Cath")
```

```
> age = c(28, 30, 21, 39, 35)
```

```
> married = c( FALSE, TRUE, TRUE, FALSE, TRUE)
```

```
> df = data.frame( name, age , married)
```

```
> df
  name age married
1 Anne  28  FALSE
2 Pete  30   TRUE
3 Frank 21   TRUE
4 Julia 39  FALSE
5 Cath  35   TRUE
```