

3. More Data structures in R

Principles of Data Science with R

Dr. Uma Ravat

PSTAT 10

Announcement

1. **HW1(last problem)** was updated on Saturday morning. Check that you have the latest version with corrected names.
2. **Extended Quiz 1** will be available after lecture till this Friday 8am for replacing Quiz 1 score.
 - **Academic Integrity:** It's not okay to see anybody's complete solution even you mention the collaboration. It's best to talk verbally and not show any solutions.
 - **Email Policy:** : Check page titled **Course Staff Office hours** which has detailed email policy.
3. Worksheet 2 - most were able to complete.

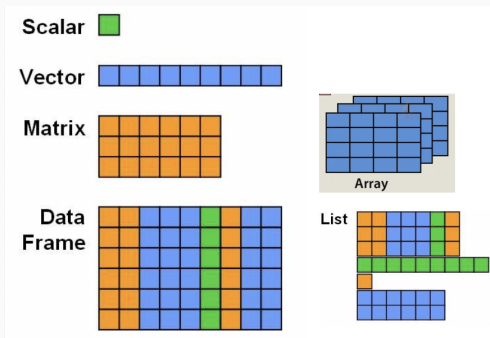
4. Please use **ULA/TA office hours**.
5. HW 1 is open and due this Wednesday
 - HW clinic will be on Wednesday 6pm - 8pm.
6. Quiz 2 will be this Friday.

Next we will see. . .

- More data structures
 - vectors
 - matrix
 - array
 - factor
 - logical operators

Matrices, Arrays, and Lists

- Data with dimensionality
- Often times, a vector (1-D) is not enough.



Matrix : two dimensional vector

Matrix : two dimensional vector

- All elements must be of the same data type
- ordered as [row, column]
- ?matrix: syntax and arguments
- **Syntax :** matrix(data, nrow, ncol, byrow, dimnames)

Description

matrix creates a matrix from the given set of values.

as.matrix attempts to turn its argument into a matrix.

is.matrix tests if its argument is a (strict) matrix.

Usage

```
matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE,  
        dimnames = NULL)
```

```
as.matrix(x, ...)  
## S3 method for class 'data.frame'  
as.matrix(x, rownames.force = NA, ...)
```

```
is.matrix(x)
```

Arguments

data	an optional data vector (including a list or expression vector). Non-atomic classed R objects are coerced by as.vector and all attributes discarded.
nrow	the desired number of rows.
ncol	the desired number of columns.
byrow	logical. If FALSE (the default) the matrix is filled by columns, otherwise the matrix is filled by rows.
dimnames	A dimnames attribute for the matrix: NULL or a list of length 2 giving the row and column names respectively. An empty list is treated as NULL, and a list of length one as row names. The list can be named, and the list names will be used as names for the dimensions.
x	an R object.
...	additional arguments to be passed to or from methods.

What we did

1. Create matrix
 - 1.2 Get matrix dimensions
 - 1.3 Assigning row and column names
 - 1.4 Get row and column names
 - 1.5 Creating a matrix by binding to rows or columns
 2. Subsetting
 - 2.1 Using [] : provide an index to each dimension.
 - order in a matrix is given as [row , col]
 - **Omitting** an index returns all elements in that dimension
 - 2.2 Subsetting by using row/column names
- Add anything else that we did*


```
(x <- matrix(1:9, nrow=3, ncol=3))
```

```
##      [,1] [,2] [,3]
## [1,]    1    4    7
## [2,]    2    5    8
## [3,]    3    6    9
```

Always print your object in exercises you submit. If you create an object but don't print it, the grader can't verify you successfully did the required task.

What is the result of `x[1:2, -3]`?

Create a 4×5 matrix `mx` of integers 1 through 17 (inclusive).
Print the matrix.

1. Subset `mx` to create the following matrix:

```
result
```

```
##           [,1] [,2] [,3]
## [1,]         1     5    17
## [2,]         4     8     3
```

2. Set all entries of `mx` greater than 10 to zero.

**Arrays : Higher (>2) dimensional
vectors**

Arrays : Higher (>2) dimensional vectors

- **Arrays** are matrices in layers.
 - ordered as [row, column, level]

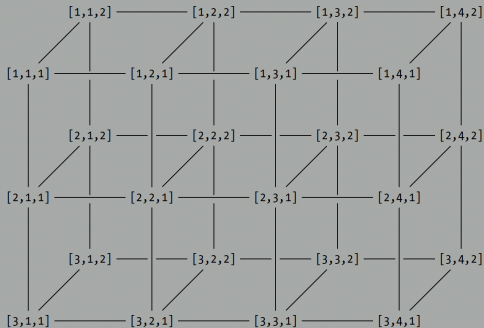


Figure 3-3: A conceptual diagram of a $3 \times 4 \times 2$ array. The index of each element is given at the corresponding position. These indexes are provided in the strict order of [row, column, layer].

Array creation Syntax

```
array(data, dim, dimnames)
```

```
A = array(data = 1:18, dim = c(3, 3, 2))
```

What we did

1. Create array
 - 1.2 Get array dimensions
 - 1.3 Assigning row, column, layer names
 - 1.4 Get row, column, layer names
 - 1.5 Creating a matrix by binding to rows or columns 4 Subsetting
 - 1.6 Using []
 - 1.7 Subsetting by using dimnames names
2. apply function

Add anything else that we did

Data structures in R

Homogeneity ↑	Dimensions →		
	1 D	2 D	Multi-D
	Homogeneous	Matrix	Array
Heterogeneous	List	Dataframe	

Factor data structure

Factor data structure

- when data belong to one of a finite number of distinct categories

A factor consists of

1. a set of **values**
2. a set of **valid levels** (the different categories)
 - the levels can be **ordered** (ordinal) or **unordered** (nominal)

Example:

- Ordinal
 - Grades: A, B, C
 - Month of the year : Jan, Feb, Mar, ... Dec
- Nominal
 - Sex: Male, Female
 - Color of Hair : Brown, Black, Blonde, Red, Other

Ordered factor : Two options in R

1. `ordered()` function for ordinal data.
2. `factor()` function with the argument `ordered = TRUE`

Logical Operators

Logical Operators

Single comparison operator	Interpretation
==	Equal to
!=	Not equal to
>	Greater than
<	Less than
>=	Greater than or equal to
<=	Less than or equal to

```
5+6 == 4
```

```
## [1] FALSE
```

Comparing Two Logical Values

Operator	Interpretation	Results
&	AND (element-wise)	TRUE & TRUE is TRUE
		TRUE & FALSE is FALSE
		FALSE & TRUE is FALSE
		FALSE & FALSE is FALSE
&&	AND (single comparison)	Same as & above
	OR (element-wise)	TRUE TRUE is TRUE
		TRUE FALSE is TRUE
		FALSE TRUE is TRUE
		FALSE FALSE is FALSE
	OR (single comparison)	Same as above
!	NOT	!TRUE is FALSE
		!FALSE is TRUE

`which()`

`which()` function returns index positions that satisfy a logical condition

For example, looking at the `state.name` dataset, we see that "Alaska" is the second element.

```
which(state.name == "Alaska")
```

```
## [1] 2
```

Example: Using the state data set

How many US states have an area less than 10,000 **OR** greater than 100,000 square miles?

How many US states have an area less than 10,000 **AND** greater than 100,000 square miles?

questions you should be able to answer

- “What are the different data types in R?”
- “What are the different data structures in R?”
- “How do I access data within the various data structures?”

Summary:

More data structures

- matrices and arrays. (Textbook Ch3)
- Factors (Textbook Chapter 4)
- Logical values (Textbook Chapter 4)

Maintain a glossary of functions used.

Next we will see. . .

- More data structures
 - Lists
 - Data frames

Learning Programming is HARD!



E. Kale Edmiston PhD

@EKaleEdmiston

Follow



A friend/colleague who is an excellent programmer offhandedly told me the other day that coding is 90% googling error messages & 10% writing code. Until this point, I thought that all the time I spent googling error messages meant I was bad at coding. What a perspective change!

8:12 AM - 4 Jan 2019

151 Retweets 1,069 Likes



27



151



1.1K

