

Big Data



About..

컴퓨터소프트웨어공학과
김 원 일



R matrix – 1

- matrix 정의

- 1차원은 vector 자료 구조를 이용하여 저장, 2차원 데이터 전용 자료 처리를 위한 자료 구조
- 대부분의 데이터는 2차원 형태로 구성되므로 많이 사용
- 배열과 동일하게 동일한 다수의 자료형을 선언하고 사용하는 형태
- 선언 “변수 <- matrix(data, nrow, ncol, byrow, dimnames))”

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

그리고, R 또는 R 패키지들을 출판물에 인용하는 방법에 $

'demo()'를 입력하신다면 몇가지 데모를 보실 수 있으며, $
또한, 'help.start()'의 입력을 통하여 HTML 브라우저에 $
R의 종료를 원하시면 'q()'를 입력해주세요.

[이전에 저장한 작업공간을 복구하였습니다]

> mtx <- matrix( 1:20, nrow = 4, ncol = 5 )
> mtx
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    5    9   13   17
[2,]    2    6   10   14   18
[3,]    3    7   11   15   19
[4,]    4    8   12   16   20
```



• 데이터 저장 형식

- 컬럼(column) 우선으로 데이터가 저장되는 것이 기본 설정
- 열(row) 우선으로 저장하려면 byrow = T 로 설정해야 함
- 일반적인 2차원 배열이라고 볼 수 있음
- vector 처럼 시작 인덱스는 1부터 시작

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> mtx <- matrix( 1:20, nrow = 4, ncol = 5 )
> mtx
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    5    9   13   17
[2,]    2    6   10   14   18
[3,]    3    7   11   15   19
[4,]    4    8   12   16   20
> mtx2 <- matrix( 1:20, nrow = 4, ncol = 5, byrow = T )
> mtx2
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    2    3    4    5
[2,]    6    7    8    9   10
[3,]   11   12   13   14   15
[4,]   16   17   18   19   20
> |
```



R matrix - 3

- 원소 접근 방법 - 1

- 일반적인 배열 원소 접근과 다른 형식으로 접근해야 함
- 배열 원소 형식으로 접근하는 정의할 수 없는 값으로 출력

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말
[Icons: File Explorer, Run, Save, Print, etc.]

> ?matrix
starting httpd help server ... done
> mtx[ 1 ] [ 1 ]
[1] 1
> mtx[ 1 ] [ 2 ]
[1] NA
> mtx[ 1 ] [ 3 ]
[1] NA
> mtx[ 2 ] [ 3 ]
[1] NA
> mtx[ 2 ] [ 1 ]
[1] 2
> mtx[ 2 ] [ 2 ]
[1] NA
> |
```



• 원소 접근 방법 – 2

- “대표matrix명[row, column]” 형식으로 접근해야 정확한 접근 가능
- 접근을 통한 다양한 연산은 동일하게 사용 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> mtx
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    5    9   13   17
[2,]    2    6   10   14   18
[3,]    3    7   11   15   19
[4,]    4    8   12   16   20
> mtx[ 1, 2 ]
[1] 5
> mtx[ 2, 4 ]
[1] 14
> mtx[ 1, 3 ] * mtx[ 4, 1 ]
[1] 36
> |
```



R matrix – 5

• 문자열 matrix – 1

- 문자열로 matrix를 초기화할 때 오류가 발생
- 입력되는 데이터가 함수 인수로 인식되어 발생하는 오류

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

matrix("str1", "str2", "str3", "444", nrow = 3, ncol = 3)에서 다음과 $
  사용되지 않은 인자 ("444")
> stx <- matrix( 'str1', "str2", 'str3', "444", nrow = 2, ncol = 2 )
matrix("str1", "str2", "str3", "444", nrow = 2, ncol = 2)에서 다음과 $
  사용되지 않은 인자 ("444")
> stx <- matrix( 'str1', "str2", 'str3', "444", nrow = 2, ncol = 3 )
matrix("str1", "str2", "str3", "444", nrow = 2, ncol = 3)에서 다음과 $
  사용되지 않은 인자 ("444")
> stx <- matrix( 'str1', "str2", 'str3', "444", nrow = 3, ncol = 2 )
matrix("str1", "str2", "str3", "444", nrow = 3, ncol = 2)에서 다음과 $
  사용되지 않은 인자 ("444")
> stx <- matrix( 'str1', "str2", 'str3', "444", nrow = 2, ncol = 2 )
```



R matrix – 6

• 문자열 matrix – 2

- vector 또는 별도의 자료형을 이용하여 하나의 데이터 셋으로 입력해야 함
- vector 생성을 위한 `c()` 함수를 이용하여 문자열 직접 입력 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> stx <- matrix( c( 'str1', "str2", str3, 444 ), nrow = 2, ncol = 2 )
matrix(c("str1", "str2", str3, 444), nrow = 2, ncol = 2)에서 다음과 같은 $
객체 'str3'를 찾을 수 없습니다
> stx <- matrix( c( 'str1', "str2", 'str3', "444" ), nrow = 2, ncol = 2 )
> stx
      [,1] [,2]
[1,] "str1" "str3"
[2,] "str2" "444"
> |
```



R matrix – 7

- 문자열 matrix – 3

- 이미 생성되어 있는 문자열 vector를 이용하여 초기화
- vector의 개수에 맞도록 설정하면 그대로 입력 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> str <- c( "str1", "str2", 'str3', 'str4', 234, 555 )
> str
[1] "str1" "str2" "str3" "str4" "234"  "555"
> stx <- matrix( str, nrow = 2, ncol = 3 )
> stx
      [,1] [,2] [,3]
[1,] "str1" "str3" "234"
[2,] "str2" "str4" "555"
> |
```




R matrix – 8

- matrix가 vector 보다 큰 범위를 갖는 경우
 - 비어 있는 공간에 vector가 다시 순서대로 재 입력

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

[ ,1] [ ,2] [ ,3]
[1,] "str1" "str3" "234"
[2,] "str2" "str4" "555"
> stx1 <- matrix ( str, nrow = 3, ncol = 3 )
> stx1
[ ,1] [ ,2] [ ,3]
[1,] "str1" "str4" "str1"
[2,] "str2" "234" "str2"
[3,] "str3" "555" "str3"
> stx2 <- matrix ( str, nrow = 3, ncol = 4 )
> stx2
[ ,1] [ ,2] [ ,3] [ ,4]
[1,] "str1" "str4" "str1" "str4"
[2,] "str2" "234" "str2" "234"
[3,] "str3" "555" "str3" "555"
> |
```



R matrix – 9

- matrix가 vector 보다 작은 범위를 갖는 경우
 - vector의 개수가 row와 column 개수의 배수가 되면 입력 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> stx3 <- matrix ( str, nrow = 3, ncol = 1 )
> stx3
      [,1]
[1,] "str1"
[2,] "str2"
[3,] "str3"
> stx3 <- matrix ( str, nrow = 2, ncol = 2 )
> stx3
      [,1] [,2]
[1,] "str1" "str3"
[2,] "str2" "str4"
> stx4 <- matrix( str, nrow = 4, ncol = 1 )
경고메시지 (들) :
matrix(str, nrow = 4, ncol = 1)에서:
  데이터의 길이[6]가 행의 개수[4]의 배수가 되지 않습니다
> |
```



R matrix – 10

- matrix에 vector 결합 – 1
 - cbind(), rbind() 함수를 이용하여 결합 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

> cd <- 1:4
> rd <- 5:8
> cd
[1] 1 2 3 4
> rd
[1] 5 6 7 8
> m1 <- cbind( cd, rd )
> m1
      cd rd
[1,]  1  5
[2,]  2  6
[3,]  3  7
[4,]  4  8
> m2 <- rbind( cd, rd )
> m2
      [,1] [,2] [,3] [,4]
cd       1    2    3    4
rd       5    6    7    8
> |
```

Windows 정품 인증
[설정]으로 이동하여 Windows를 정품 인증합니다.



R matrix – 11

- matrix에 vector 결합 – 2
 - 단일 vector를 이용하여 matrix 생성도 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

cd      1      2      3      4
rd      5      6      7      8
> m3 <- cbind( cd )
> m3
      cd
[1,]  1
[2,]  2
[3,]  3
[4,]  4
> m4 <- rbind( cd )
> m4
      [,1] [,2] [,3] [,4]
cd      1   2   3   4
> |
```



R matrix – 12

- matrix에 vector 결합 – 3
 - vector를 인수로 다중 데이터 입력 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

> m5 <- cbind( cd, cd, rd, cd )
> m5
      cd cd rd cd
[1,]  1  1  5  1
[2,]  2  2  6  2
[3,]  3  3  7  3
[4,]  4  4  8  4
> m6 <- rbind( rd, rd, cd, cd )
> m6
      [,1] [,2] [,3] [,4]
rd       5     6     7     8
rd       5     6     7     8
cd       1     2     3     4
cd       1     2     3     4
> |
```



- matrix에 matrix 결합 – 1
 - 이미 존재하는 matrix를 이용하여 matrix를 재 구성 가능
 - 마찬가지로 row와 column 개수의 배수가 되면 입력 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

> m7 <- rbind( m3, m4 )
rbind(m3, m4)에서 다음과 같은 에러가 발생했습니다:
  각각의 행렬이 가진 열의 개수는 반드시 서로 일치해야 함$
> m7 <- rbind( m3, m5 )
rbind(m3, m5)에서 다음과 같은 에러가 발생했습니다:
  각각의 행렬이 가진 열의 개수는 반드시 서로 일치해야 함$
> m7 <- rbind( m4, m5 )
> m7
      cd cd rd cd
cd   1  2  3  4
     1  1  5  1
     2  2  6  2
     3  3  7  3
     4  4  8  4
> |
```



R matrix – 14

- matrix에 matrix 결합 – 2
 - matrix 자신을 매개 변수로 row, column 추가도 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말
[Icons: File Explorer, Print, Save, Copy, Paste, Undo, Redo, Stop, Run]

> m6
      [,1] [,2] [,3] [,4]
rd      5      6      7      8
rd      5      6      7      8
cd      1      2      3      4
cd      1      2      3      4
> m6 <- rbind( m6, cd )
> m6
      [,1] [,2] [,3] [,4]
rd      5      6      7      8
rd      5      6      7      8
cd      1      2      3      4
cd      1      2      3      4
cd      1      2      3      4
> |
```



- matrix 데이터 접근 – 1

- vector와 다르게 첨자 범위를 벗어나면 에러 발생

```
RGU (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

> m1
      cd rd
[1,]  1  5
[2,]  2  6
[3,]  3  7
[4,]  4  8
> m1[ 1, 3 ]
m1 [1, 3]에서 다음과 같은 에러가 발생했습니다: 첨자의 허용$
> m1[ 5, 2 ]
m1 [5, 2]에서 다음과 같은 에러가 발생했습니다: 첨자의 허용$
> |
```




- matrix 데이터 접근 – 2

- row 또는 column을 명시하지 않으면 전체를 나타냄

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지들 윈도우즈 도움말

> m5
      cd cd rd cd
[1,]  1  1  5  1
[2,]  2  2  6  2
[3,]  3  3  7  3
[4,]  4  4  8  4
> m5[ 1, ]
cd cd rd cd
 1  1  5  1
> m5[ , 3 ]
[1] 5 6 7 8
> |
```



- matrix 값에 대한 접근

- row, column 표시 없이 접근해도 모든 연산 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

> m5
      cd cd rd cd
[1,]  1  1  5  1
[2,]  2  2  6  2
[3,]  3  3  7  3
[4,]  4  4  8  4
> m5[ 1, ]
cd cd rd cd
 1  1  5  1
> m5[ , 3 ]
[1] 5 6 7 8
> m5[ , 2 ] + m5[ , 3 ]
[1]  6  8 10 12
> |
```

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지를 윈도우즈 도움말

> m5
      cd cd rd cd
[1,]  1  1  5  1
[2,]  2  2  6  2
[3,]  3  3  7  3
[4,]  4  4  8  4
> m6
      [,1] [,2] [,3] [,4]
rd      5   6   7   8
rd      5   6   7   8
cd      1   2   3   4
cd      1   2   3   4
> m5[ , 2 ] + m6[ , 3 ]
rd rd cd cd
 8  9  6  7
> |
```