

Big Data



About..

컴퓨터소프트웨어공학과
김 원 일



- matrix 데이터 접근 – 1

- vector와 다르게 첨자 범위를 벗어나면 에러 발생

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

> 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
> |
```

Big Data



About..

컴퓨터소프트웨어공학과
김 원 일



Data Frame – 1

• 또 다른 2차원 배열을 다루는 자료형

- matrix와 달리 다양한 자료형을 포함할 수 있는 matrix
- matrix는 기본적으로 모든 자료형이 동일한 경우에만 사용
- 일반적인 구조체나 클래스의 자료형 선언과 동일
- "변수 <- data.frame(data1, data2, data3 ...) [ENTER]" 형식으로 생성

```
> num <- c( 1, 3, 5, 7 )
> log <- c( T, T, F, T )
> char <- c( "a", "b", "c", "d" )
> df <- data.frame( num, log, char )
> df
```

	num	log	char
1	1	TRUE	a
2	3	TRUE	b
3	5	FALSE	c
4	7	TRUE	d

```
> |
```



Data Frame – 2

- 새로운 데이터 추가

- “변수\$컬럼명 <- c(...) [ENTER]” 형태로 컬럼 추가 가능

```
RGui (64-bit) - [R Console]
파일 편집 보기 기타 패키지 윈도우 도움말
[Icons]

> df$name <- c( "kim", "park", "lee", "choi" )
> df
  num  log char name
1   1 TRUE   a  kim
2   3 TRUE   b park
3   5 FALSE  c  lee
4   7 TRUE   d choi
> str( df )
'data.frame':   4 obs. of  4 variables:
 $ num : num  1 3 5 7
 $ log : logi TRUE TRUE FALSE TRUE
 $ char: chr  "a" "b" "c" "d"
 $ name: chr  "kim" "park" "lee" "choi"
> |
```



Data Frame – 3

- data frame 간 결합

- rbind(), cbind()를 이용하여 matrix와 같이 결합 가능

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

> df1 <- data.frame( num, log )
> num1 <- c( 9, 11, 13, 15 )
> log1 <- c( T, F, T, T )
> df2 <- data.frame( num1, log1 )
> df1
```

```
  num  log
1    1 TRUE
2    3 TRUE
3    5 FALSE
4    7 TRUE
```

```
> df2
  num1 log1
1     9  TRUE
2    11 FALSE
3    13  TRUE
4    15  TRUE
> |
```

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

> df3 <- cbind( df1, df2 )
> df3
  num  log num1 log1
1    1 TRUE     9  TRUE
2    3 TRUE    11 FALSE
3    5 FALSE   13  TRUE
4    7 TRUE    15  TRUE
> |
```




Utilities - 1

- `ls()`
 - 작업 공간 내에 선언된 모든 변수 목록의 확인
- `str(변수 / 데이터)`
 - 변수 / 데이터의 종류와 값을 확인할 수 있음
- `length(변수 / 데이터)`
 - 데이터의 길이를 확인할 수 있음

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

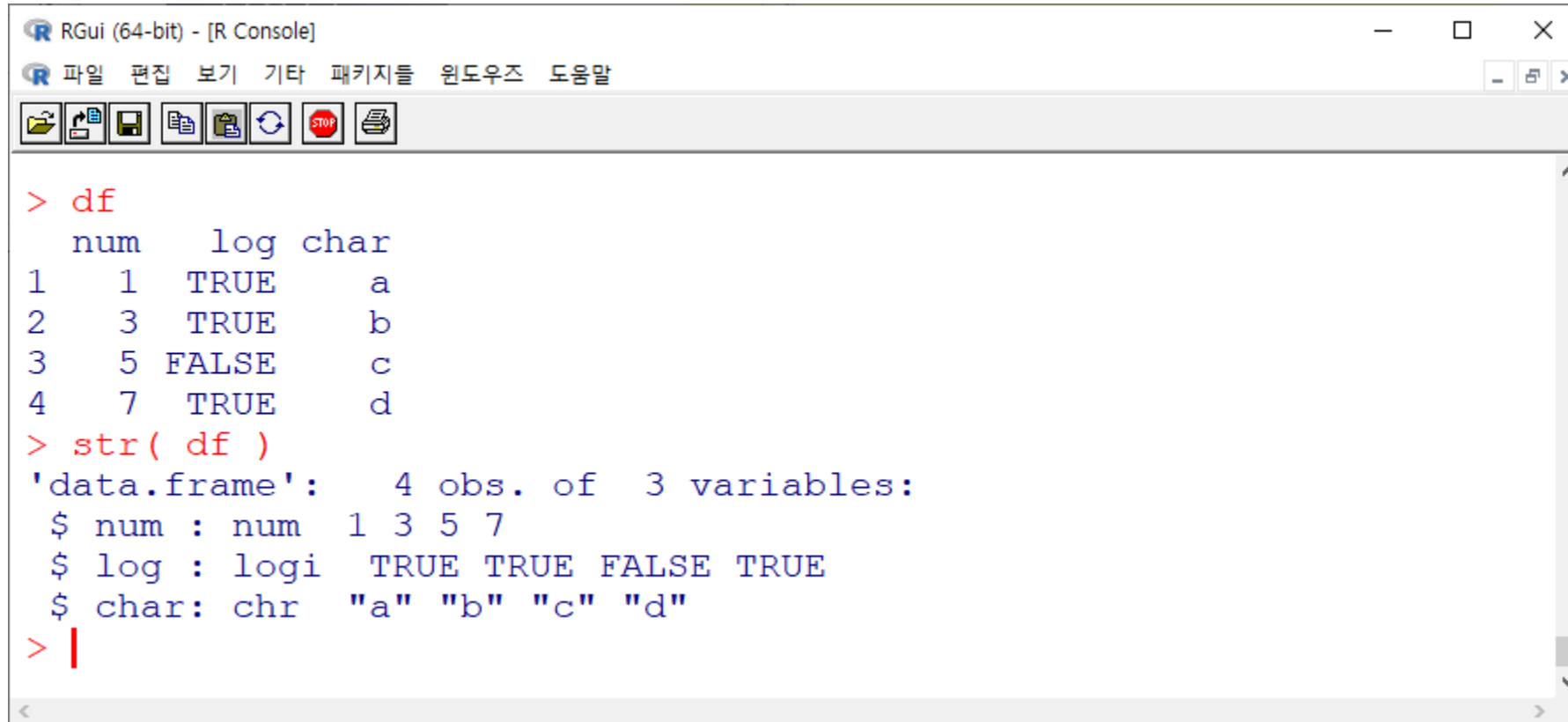
> ls( )
[1] "a"          "b"          "c"
[4] "d"          "mtx"        "userInfo.addr"
[7] "userInfo.name" "userInfo.phone" "vec"

> vec <- c( 1, 2, 3 : 7 )
> mtx <- matrix( vec )
> str( vec )
num [1:7] 1 2 3 4 5 6 7
> str( mtx )
num [1:7, 1] 1 2 3 4 5 6 7
> |
```



Utilities - 2

- **str() 함수로 내부 객체 정보 확인**
 - str(R-object) 형식으로 자료형과 내용을 확인할 수 있음



```
> df
  num  log char
1   1 TRUE   a
2   3 TRUE   b
3   5 FALSE  c
4   7 TRUE   d
> str(df)
'data.frame':   4 obs. of  3 variables:
 $ num : num  1 3 5 7
 $ log : logi TRUE TRUE FALSE TRUE
 $ char: chr  "a" "b" "c" "d"
> |
```



Utilities - 3

- `str()`로 각종 정보 확인
 - 모든 R 변수의 정보를 확인할 수 있음

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

> str( num )
num [1:4] 1 3 5 7
> str( char )
chr [1:4] "a" "b" "c" "d"
> str( log )
logi [1:4] TRUE TRUE FALSE TRUE
> str( a )
num 2.68e+08
> str( d )
num 69708
> str( userInfo.name )
chr "wikim"
> |
```



- length()로 데이터의 길이 확인
 - 모든 종류의 데이터 길이 확인이 가능