

1. { 자료 구조 }

(R Study group 자료)

데이터 종류 (data)

정형 데이터 : 특정한 구조에 담겨어 있는 데이터, 가장 흔한 구조
ex) 엑셀, CSV, Data Base (DB)

비정형 데이터 : 특정 형태 없이 저장된 데이터
ex) 페이스북 텍스트 데이터, 인스타그램 사진 데이터

반정형 데이터 : 특정 형태는 존재, 연산 불가능
ex) web log-data, HTML, XML

(R Study group 자료)

데이터 구조 in ' R '

1) 벡터 (vector)

a <- 5 # 단일 벡터(= 스칼라)

a_text <- "a"

b <- c(1, 3, 5) # 벡터

b_text <- c("a", "b", "c")

기본구조 이름 = **c(숫자, 숫자2, ...)** # " 문자"

데이터 구조 in ' R '

1) 벡터 (vector)

```
b <- c(1, 3, 5) # 벡터
```

```
b_text <- c("a", "b", "c")
```

b[1]	b[2]	b_text[3]
> 1	> 3	> "c"

기본구조 이름[순번] # 순번에 해당하는 값 출력

데이터 구조 in ' R '

2) 팩터(factor) : 범주형 데이터 (= 명목척도)
ex) 학년, 성별, 지역 etc.

```
region <- c("창원시", "마산시", "진해시")  
region <- as.factor(region)
```

기본구조 이름 = as.factor(이름) # 팩터화

(R Study group 자료)

데이터 구조 in ' R '

2) 팩터(factor) : 범주형 데이터 (= 명목척도)
ex) 학년, 성별, 지역 etc.

```
gender <- c(1, 2) # 여성 : 1, 남성 : 2  
gender <- as.factor(gender)
```

기본구조 이름 = as.factor(이름) # 팩터화

(R Study group 자료)

데이터 구조 in ' R '

3) 데이터프레임(data frame) :

둘 이상의 벡터 모음 **BUT ! 모든 벡터의 길이는 같음**

id	gender	region
1	남성	창원
2	여성	마산
3	남성	진해

(R Study group 자료)

데이터 구조 in ' R '

3) 데이터프레임(data frame) :

```
id <- c(1, 2, 3)
```

```
gender <- c("남성", "여성", "남성")
```

```
region <- c("창원", "마산", "진해")
```

```
df <- data.frame(id, gender, region)
```

기본구조 이름 = data.frame(벡터, 벡터, ...)

데이터 구조 in ' R '

3) 데이터프레임(data frame) :

```
df <- data.frame(id = c(1, 2, 3),  
                 gender = c("남성", "여성", "남성"),  
                 region = c("창원", "마산", "진해"))
```

기본구조 이름 = data.frame(벡터 = 값,
 벡터2 = 값, ...)

데이터 구조 in ' R '

3) 데이터프레임(data frame) 의 값을 보는 방법 ex) region 변수(벡터)를 보고싶음

df[,3]

df[, "region"]

df\$region

기본구조 이름[, 열] # 보려는 열 위치 입력
이름[, "변수 이름"]
이름\$변수이름 # " " (x)

(R Study group 자료)

데이터 구조 in ' R '

4) 리스트(list) : 주머니, 혹은 가방의 개념 (?)
즉 정리해서 넣기 나름이다.

```
a_list <- list(1, "a", 3)
```

```
a_list2 <- list(c(1, "a", 3), 1, 2, "b")
```

```
a_list3 <- list(c(1, "a", 3), list(1, 2, "b"), 2, 3)
```

기본구조 이름 <- list("문자", 숫자, list(), ...)

문자, 숫자 통일성 없이 다 넣을 수 있음

데이터 구조 in ' R '

4) 리스트(list)

```
a_list <- list(c(1, 2, "b"), "a", 3)
```

```
a_list[1] > [[1]] 1, 2, "b"
```

```
a_list[[1]][2] > 2
```

기본구조 이름 <- list("문자", 숫자, list(), ...)

문자, 숫자 통일성 없이 다 넣을 수 있음

데이터 구조 in ' R '

5) 매트릭스(matrix) : 행렬구조

```
a <- c(1, 2, 3, 4)  
a_matrix <- matrix(a, 2, 2)
```

!! matrix 데이터는 모두 같은 속성을 가져야한다.

기본구조 이름 <- matrix(벡터, n행, n열, ...)

데이터 구조 in ' R '

5) 매트릭스(matrix) : 행렬구조

```
a <- c(1, "a", 3, "b")  
a_matrix <- matrix(a, 2, 2)
```

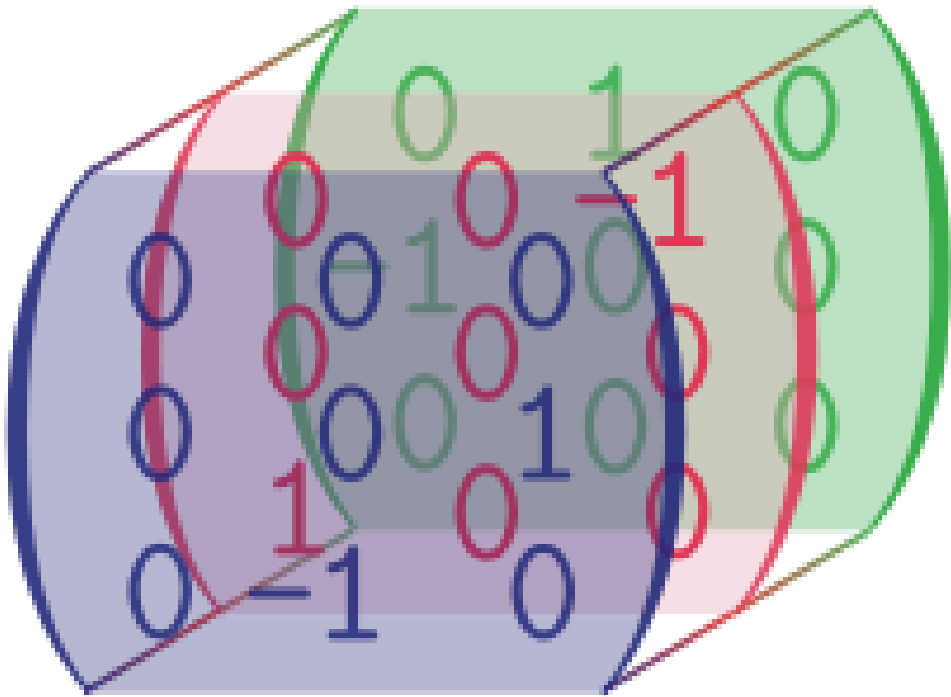
› 모두 문자로 변환됨

기본구조 이름 <- matrix(벡터, n행, n열, ...)

데이터 구조 in ' R '

6) 어레이(array) : 차원이 있는 행렬구조

ex) 음성 데이터, 그림 데이터, 영상 데이터 etc.



(주로 딥러닝 (Deep learning) 에서 많이 사용되는 구조)

(R Study group 자료)