

---

Chapter 2


# 변수



# 변수

## 변수(Variable)

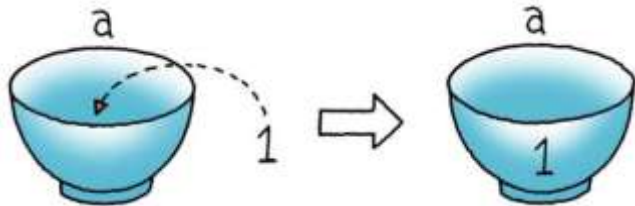
- 다양한 값을 지니고 있는 하나의 속성
- 변수는 데이터 분석의 대상



소득	성별	학점	국적
1,000만 원	남자	3.8	대한민국
2,000만 원	남자	4.2	대한민국
3,000만 원	여자	2.6	대한민국
4,000만 원	여자	4.5	대한민국

# 변수

## 변수 생성하기



```
a <- 1  
a  
## [1] 1  
  
b <- 2  
b  
## [1] 2  
  
c <- 3  
c  
## [1] 3  
  
d <- 3.5  
d  
## [1] 3.5
```

# 연산자

## 연산자

산술 연산자

+, -, \*, /

논리 연산자

>, <, <=, >=, ==, !=

```
a+b
```

```
## [1] 3
```

```
a+b+c
```

```
## [1] 6
```

```
4/b
```

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

```
5*b
```

```
## [1] 10
```

# 연산자

## 연산자

산술 연산자

+, -, \*, /

논리 연산자

>, <, <=, >=, ==, !=

```
a+b
```

```
## [1] 3
```

```
a+b+c
```

```
## [1] 6
```

```
4/b
```

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

```
5*b
```

```
## [1] 10
```

# 변수

## 연속 변수 만들기

`c()`

```
var1 <- c(1, 2, 5, 7, 8)
```

```
var1
```

```
## [1] 1 2 5 7 8
```

```
var2 <- c(1:5)
```

```
var2
```

```
## [1] 1 2 3 4 5
```

## 연속 문자 변수 만들기

```
str4 <- c("a", "b", "c")
```

```
str4
```

```
## [1] "a" "b" "c"
```

```
str5 <- c("Hello!", "World", "is", "good!")
```

```
str5
```

```
## [1] "Hello!" "World"  "is"     "good!"
```



# 변수

## 연속 변수 만들기

seq()

```
var3 <- seq(1, 5)
```

```
var3
```

```
## [1] 1 2 3 4 5
```

```
var4 <- seq(1, 10, by = 2)
```

```
var4
```

```
## [1] 1 3 5 7 9
```

```
var5 <- seq(1, 10, by = 3)
```

```
var5
```

```
## [1] 1 4 7 10
```

## 변수삭제

```
rm(변수명)
```

---

Chapter 2

# 데이터 타입





# 데이터 타입

## 주요 데이터 유형

숫자(numeric) – 숫자로 이루어져 있으며 정수(integer) 실수(double)을 모두 포함

문자(character)- 문자로 이루어져 있으며 길이에 관계없이 형식이 같음

논리형(logical): - TRUE, FALSE. 즉 참 거짓

범주형(factor): - 명목/순서형 변수

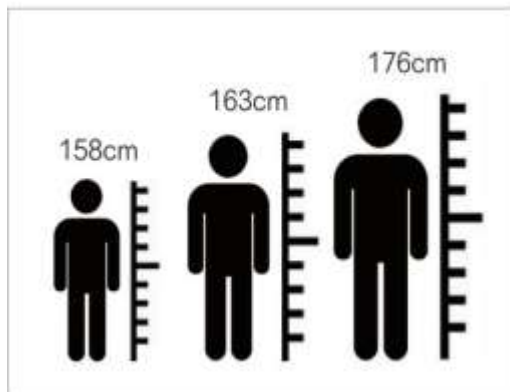
## 특수 데이터 유형

NA -결측값(값이 존재하지 않음)

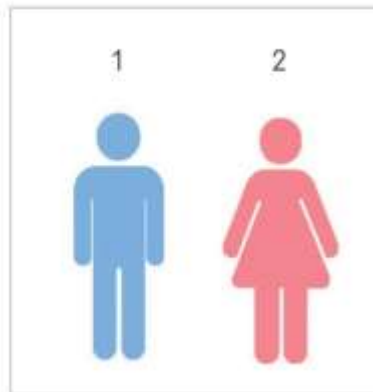
NaN- 수학적으로 정의가 불가능한 수

Inf- 무한값

연속 변수



범주 변수



함수

기능

class()

타입확인

str()

구조확인

levels()

범주형 확인



# 데이터 타입

## 숫자로 구성된 factor 변수

```
var1 <- c(1,2,3,1,2)           # numeric 변수 생성
var2 <- factor(c(1,2,3,1,2))    # factor 변수 생성

var1      # numeric 변수 출력
## [1] 1 2 3 1 2

var2      # factor 변수 출력
## [1] 1 2 3 1 2
## Levels: 1 2 3
```

## 문자로 구성된 factor 변수

```
var3 <- c("a", "b", "b", "c")
var4 <- factor(c("a", "b", "b", "c"))

var3
## [1] "a" "b" "b" "c"

var4
Levels: a b b c
```



# 데이터 타입

## 변환 함수(Coercion Function)

함수	기능
as.numeric()	numeric으로 변환
as.factor()	factor로 변환
as.character()	character로 변환
as.data.frame()	Data Frame으로 변환
as.Date(표준서식)	날짜형 변환

# 객체형 데이터 타입

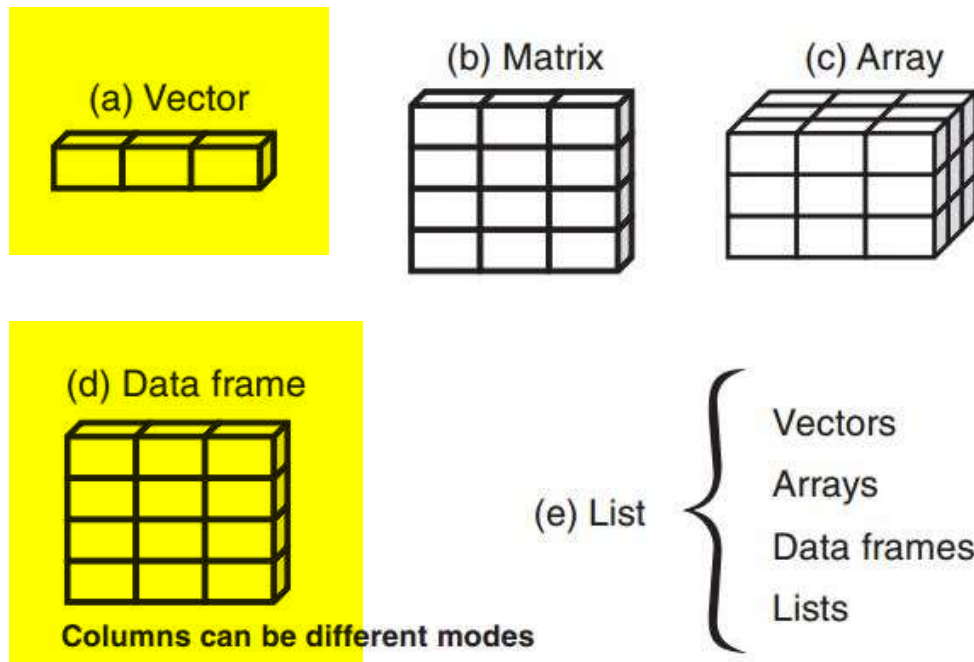


Figure 2.1 R data structures

- 데이터 프레임 외에도 다양한 데이터 구조가 있음
- 데이터 구조에 따라 활용 방법 다름

데이터 구조	차원	특징
벡터(Vector)	1차원	한 가지 변수 타입으로 구성
데이터 프레임(Data Frame)	2차원	다양한 변수 타입으로 구성
매트릭스(Matrix)	2차원	한 가지 변수 타입으로 구성
어레이(Array)	다차원	2차원 이상의 매트릭스
리스트(List)(완전혼합형)	다차원	서로 다른 데이터 구조 포함

# 객체형 데이터 타입

## 1. 벡터(Vectr)

- 하나 또는 여러 개의 값으로 구성된 데이터 구조
- 여러 타입을 섞을 수 없고, 한 가지 타입으로만 구성 가능

```
# 벡터 만들기
a <- 1
a

## [1] 1

b <- "hello"
b

## [1] "hello"

# 데이터 구조 확인
class(a)

## [1] "numeric"

class(b)

## [1] "character"
```

# 객체형 데이터 타입

## 2. 데이터 프레임(Data Frame)

- 행과 열로 구성된 2차원 데이터 구조
- 다양한 변수 타입으로 구성 가능

# 데이터 프레임 만들기

```
x1 <- data.frame(var1 = c(1,2,3),  
                 var2 = c("a","b","c"))
```

x1

```
##   var1 var2  
## 1    1   a  
## 2    2   b  
## 3    3   c
```

# 데이터 구조 확인

```
class(x1)
```

```
## [1] "data.frame"
```

# 객체형 데이터 타입

## 3. 매트릭스(Matrix)

- 행과 열로 구성된 2차원 데이터 구조
- 한 가지 타입으로만 구성 가능

# 매트릭스 만들기 - 1~12 로 2 열

```
x2 <- matrix(c(1:12), ncol = 2)
```

x2

```
##      [,1] [,2]  
## [1,]    1    7  
## [2,]    2    8  
## [3,]    3    9  
## [4,]    4   10  
## [5,]    5   11  
## [6,]    6   12
```

# 데이터 구조 확인

```
class(x2)
```

```
## [1] "matrix"
```

# 객체형 데이터 타입

## 5. 리스트(List)

- 모든 데이터 구조를 포함하는 데이터 구조
- 여러 데이터 구조를 합해 하나의 리스트로 구성 가능

# 리스트 생성 - 앞에서 생성한 데이터 구조 활용

```
x4 <- list(f1 = a,    # 벡터
          f2 = x1,    # 데이터 프레임
          f3 = x2,    # 매트릭스)
```

x4





## Chapter 2

# 수고하셨습니다