

*Kaggle tools*

---

# R: Subsetting & Vocabulary

---

Part 1 - 3rd

# Subsetting - Vector

## ❖ Vector

- 같은 형의 원소로 구성된 object

```
> v <- 1:10
> v
[1] 1 2 3 4 5 6 7 8 9 10
> v <- c(1, 3, 5, 7, 9)
> v
[1] 1 3 5 7 9
> v <- c("Fred", "Mary", "David")
> v
[1] "Fred" "Mary" "David"
>
```

- [ ]를 통해 indexing

```
> v <- c(1, 3, 5, 7, 9)
> v
[1] 1 3 5 7 9
> v[5]
[1] 9
> v[c(2,3)]
[1] 3 5
> |
```

\* c(): combine

# Subsetting - List

## ❖ List

- 이질적인 원소들로 구성되는 object

```
> lst <- list(3.14, "Moe", c(1, 1, 2, 3))
> lst
[[1]]
[1] 3.14

[[2]]
[1] "Moe"

[[3]]
[1] 1 1 2 3
```

- indexing

```
> lst <- list(name = "Fred", wife = "Mary", c(1, 3, 5, 10))
> lst
$name
[1] "Fred"

$wife
[1] "Mary"

[[3]]
[1] 1 3 5 10

> lst[[1]]
[1] "Fred"
> lst$name
[1] "Fred"
> lst[[3]]
[1] 1 3 5 10
> lst[1]
$name
[1] "Fred"
```



# Subsetting - Data Frame

- ❖ Data frame 내 indexing
  - []를 사용하여, 행번호 또는 열번호 지정 가능  
(순서는 [행(row), 열(column)])

```
> mtcars[2,5]
[1] 3.9
> mtcars[c(1:5), c(2,5)]
      cyl drat
Mazda RX4      6 3.90
Mazda RX4 Wag  6 3.90
Datsun 710     4 3.85
Hornet 4 Drive  6 3.08
Hornet Sportabout 8 3.15
```

# Subsetting - Data Frame

## ❖ Data frame 내 indexing

- 데이터프레임명\$리스트 key 형태로 얻어낼 수 있다

```
> mtcars$cyl
[1] 6 6 4 6 8 6 8 4 4 6 6 8 8 8 8 8 8 4 4 4 4 8 8 8 8 4 4 4 8 6 8 4
> head(mtcars[,mtcars$cyl])
```

	wt	wt.1	hp	wt.2	vs	wt.3	vs.1	hp.1	hp.2	wt.4	wt.5	vs.2	vs.3	vs.4	vs.5	vs.6	vs.7	hp.3
Mazda RX4	2.620	2.620	110	2.620	0	2.620	0	110	110	2.620	2.620	0	0	0	0	0	0	110
Mazda RX4 Wag	2.875	2.875	110	2.875	0	2.875	0	110	110	2.875	2.875	0	0	0	0	0	0	110
Datsun 710	2.320	2.320	93	2.320	1	2.320	1	93	93	2.320	2.320	1	1	1	1	1	1	93
Hornet 4 Drive	3.215	3.215	110	3.215	1	3.215	1	110	110	3.215	3.215	1	1	1	1	1	1	110
Hornet Sportabout	3.440	3.440	175	3.440	0	3.440	0	175	175	3.440	3.440	0	0	0	0	0	0	175
Valiant	3.460	3.460	105	3.460	1	3.460	1	105	105	3.460	3.460	1	1	1	1	1	1	105

  

	hp.4	hp.5	hp.6	vs.8	vs.9	vs.10	vs.11	hp.7	hp.8	hp.9	vs.12	wt.6	vs.13	hp.10
Mazda RX4	110	110	110	0	0	0	0	110	110	110	0	2.620	0	110
Mazda RX4 Wag	110	110	110	0	0	0	0	110	110	110	0	2.875	0	110
Datsun 710	93	93	93	1	1	1	1	93	93	93	1	2.320	1	93
Hornet 4 Drive	110	110	110	1	1	1	1	110	110	110	1	3.215	1	110
Hornet Sportabout	175	175	175	0	0	0	0	175	175	175	0	3.440	0	175
Valiant	105	105	105	1	1	1	1	105	105	105	1	3.460	1	105

# Basic Operator

연산자	뜻
+, -, *, /	더하기, 빼기, 곱하기, 나누기
log(), log10()	로그
exp()	exponential
sqrt()	제곱근( $\sqrt{\quad}$ )
==, !=	같다, 같지 않다
>, <, >=, <=	크다, 작다, 크거나 같다, 작거나 같다
round()	반올림
ceiling()	올림
floor()	버림
% / %	몫
%%	나머지
:	수열
&	and
	or



# Basic Functions

명령어	기능	예제
ls()	생성된 object 리스트를 보여줌	
rm()	object 삭제	rm(x), rm(list = ls())
gc()	garbage collection 호출하여 메모리 효율성을 올림	
c( , )	2개 이상의 원소 결합	c(x, y), c(1:3, 5), c("Dad", "Mom", "Dog")
rep( , )	원소를 반복	rep(5, 5), rep(0, 10), rep(1:3, 5)
seq( , , )	규칙이 있는 수열 생성	seq(1, 10, 2), seq(-10, 10, 3)
sort()	정렬	sort(x), sort(x, decreasing = TRUE)
order()	첨자정렬	order(x)
sample( , )	데이터 섞기, 랜덤추출	sample(1:10, 5), sample(1:10, 5, replace = TRUE)
paste( , , sep = )	문자열 연결	paste("Hi", "everyone", sep = " "), paste("A", 1:10, sep="")

# Data Mode

Mode	형태	변경 함수
numeric	3.14, 9, 5, -2, 0	as.numeric()
integer	0, 1, 2, 3, 4, 5	as.integer()
character	"Mary", "Dog", "I"	as.character()
factor	F, M / 1, 2, 3	as.factor()
logical	TRUE, FALSE	as.logical()
date	2015-07-07	as.Date(), as.Date(,format = "%m/%d/%Y")



---

# Attention!

---

- ❖ 백문이 불여일행 - 직접해보는 게 최고
- ❖ Something wrong?
  - 스펠링 또는 구두점
  - 대소문자 구분
  - 에러 메시지를 잘 읽어보자
  - help 기능 활용
  - Googling
- ❖ 안 배운 것도 조합해보자. 에러를 두려워말라.