

데이터 병합

열방향으로 두 데이터프레임을 병합할 때 `merge()` 또는 `cbind()`를 사용할 수 있다.

병합 예제에서 사용할 세 개의 데이터프레임을 준비한다.

```
> stu1 <- data.frame(no = c(1,2,3), midterm = c(100,90,80))
> stu2 <- data.frame(no = c(1,2,3), finalterm = c(100,90,80))
> stu3 <- data.frame(no = c(1,4,5), quiz = c(99,88,77))
```

`merge()`는 병합 대상 데이터프레임들의 동일 컬럼명의 동일 데이터 행끼리 병합한다. 다음은 `stu1`과 `stu2`을 `merge()`로 병합한 예이다. `stu1`와 `stu2`의 동일 컬럼은 `no` 이므로, `no` 컬럼의 데이터가 같은 행끼리 병합된다.

```
> stu1
  no midterm
1  1     100
2  2      90
3  3      80

> stu2
  no finalterm
1  1        100
2  2         90
3  3         80

> merge(stu1, stu2)
  no midterm finalterm
1  1     100        100
2  2      90         90
3  3      80         80
```

다음은 `merge()`로 병합할 때 일부 행이 포함되지 않는 예이다. `stu1`과 `stu3`은 `no`라는 동일한 컬럼이 있고, `stu1`의 `no` 컬럼의 값은 1,2,3 이고, `stu3`의 `no` 컬럼 값은 1,4,5이다. `no` 컬럼의 값이 같은 행인 1행은 병합 결과에 포함되고, 나머지 행은 `no` 컬럼 값이 동일하지 않아 제외된다.

```
> stu1
```

```
no midterm
```

```
1 1 100
```

```
2 2 90
```

```
3 3 80
```

```
> stu3
```

```
no quiz
```

```
1 1 99
```

```
2 4 88
```

```
3 5 77
```

```
> merge(stu1, stu3)
```

```
no midterm quiz
```

```
1 1 100 99
```

merge() 함수에서 all=TRUE 옵션을 사용하면 동일 컬럼끼리 값이 일치하지 않는 행도 병합 결과에 포함된다. SQL의 외부 조인(outer join)과 유사하다.

```
# 동일 컬럼끼리 값이 일치하지 않는 행도 병합 결과에 모두 포함
```

```
> merge(stu1, stu3, all=TRUE)
```

```
no midterm quiz
```

```
1 1 100 99
```

```
2 2 90 NA
```

```
3 3 80 NA
```

```
4 4 NA 88
```

```
5 5 NA 77
```

```
# 두 데이터프레임의 동일 컬럼인 no 컬럼의 값이 일치하는 것이 없어도
```

```
# 왼쪽 데이터프레임의 행은 무조건 병합 결과에 포함
```

```
> merge(stu1, stu3, all.x=TRUE)
```

```
no midterm quiz
```

```
1 1 100 99
```

```
2 2 90 NA
```

```
3 3 80 NA
```

```
# 동일 컬럼끼리 값이 일치하지 않아도 오른쪽 데이터프레임의 경우 병합 결과에 포함
```

```
> merge(stu1, stu3, all.y=TRUE)
```

```
no midterm quiz
```

1	1	100	99
2	4	NA	88
3	5	NA	77

`cbind()`는 단순히 컬럼을 합친다. 다음은 `cbind()` 예이다.

```
> cbind(stu1, stu2)
  no midterm no finalterm
1 1      100 1      100
2 2       90 2       90
3 3       80 3       80

> cbind(stu1, stu3)
  no midterm no quiz
1 1      100 1   99
2 2       90 4   88
3 3       80 5   77
```

`rbind()`는 행을 합친다. 이때 두 데이터프레임의 컬럼명이 동일해야 한다.

```
> stu4 <- data.frame(no = c(4,5,6), midterm = c(99,88,77))

> rbind(stu1, stu4)
  no midterm
1 1      100
2 2       90
3 3       80
4 4       99
5 5       88
6 6       77
```