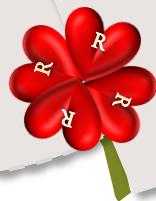


5. 데이터 프레임(data frame)

서로 다른 데이터 타입을 여러 행으로 저장할 수 있는 데이터 프레임에 대해 학습합니다.



1. 데이터 프레임(data.frame)?

행렬이나 데이터프레임이나
여러행/열로 구성되어 있는건
같아요.



행렬은 오로지 동일데이터 타입!!

13	160	57
18	170	55
21	167	52kg ✕

데이터프레임은 타입에 신경안써요

13	홍길동	서울
18	임꺽정	인천
21	장길산	서울

엑셀에서 자료를 입력하여 작업
하는 형태하고 비슷한거네?



데이터베이스 테이블에 컬럼을 만들고
값을 저장하는 형태와 비슷한 성격으로 R
에서 가장 많이 쓰이는 객체입니다.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	2015 9월 고속버스 수송실적													
2	출처: 전국고속버스운송사업조합													
3	년도	월	차종	선별	출발<-	->도착	거리	총운행회수	일 평균회수	총 인원	일 평균인원	공급 좌석수	이용률 (%)	인 Km
4	2015년	9월	무등	88선	광주	울산	327.8	412	13.7	7,283	243	11,536	63.1	2,387,367.4
5	2015년	9월	고속	88선	광주	울산	327.8	145	4.8	3,050	102	6,525	46.7	999,790.0
6	2015년	9월	무등	88선	광주	울산신복	327.8	164	5.5	545	18	4,592	11.9	178,651.0
7	2015년	9월	고속	88선	광주	울산신복	327.8	70	2.3	311	10	3,150	9.9	101,945.8
8	2015년	9월	무등	88선	광주	동대구	219.3	1369	45.6	21,873	729	38,332	57.1	4,796,748.9
9	2015년	9월	고속	88선	광주	동대구	219.3	192	6.4	3,732	124	8,640	43.2	818,427.6
10	2015년	9월	무등	88선	광주	구미	246.1	208	6.9	2,634	88	5,824	45.2	648,227.4
11	2015년	9월	무등	88선	광주	경주	286.9	139	4.6	2,102	70	3,892	54	603,063.8

◆ 데이터프레임 생성하기

- `read.table()` 함수를 이용하여 외부 텍스트 파일을 불러들이는 방법
- `data.frame()` 함수를 이용하여 여러 종류의 자료 객체들을 결합시키는 방법

I. `data.frame()`함수 이용하여 데이터프레임 생성

```
> df<-data.frame(irum=c("조정순","한민애","고승민"),age=c(43,27,34))
> df
  irum age
1 조정순 43
2 한민애 27
3 고승민 34
```

기존에 존재하던 벡터를 결합하여 데이터프레임을 생성합니다!!

```
> irum<-c("조정순","한미애","고승민")
> age<-c(43,27,34)
```

조정순	한미애	고승민
43	27	34

```
> df1<-data.frame(이름=irum,나이=age)
```

```
> df1
  이름 나이
1 조정순 43
2 한미애 27
3 고승민 34
> |
```

irum 벡터를 “이름” 컬럼에
age벡터를 “나이”컬럼에 할당

특정 컬럼만 조회하는 경우
`list처럼 프레임변수$컬럼명` 과 같이 기술

```
> df1$나이
[1] 43 27 34
> df1$나이+5
[1] 48 32 39
> |
```

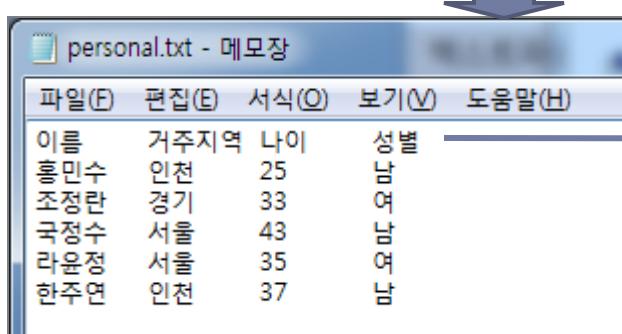
행렬→데이터프레임 변환하기

```
> carinfo<-matrix(c('11km','13km','10km','h사','v사','r사'),ncol=2)
> carinfo
 [,1]   [,2]
[1,] "11km" "h사"
[2,] "13km" "v사"
[3,] "10km" "r사"
> dfr<-as.data.frame(carinfo) 데이터프레임으로 변환하여 dfr에 할당
> dfr
  V1  V2
1 11km h사
2 13km v사
3 10km r사
> names(dfr)<-c("연비","제조사") 컬럼명 지정위해 names()함수 사용
> dfr
  연비 제조사
1 11km    h사
2 13km    v사
3 10km    r사
```

2. read.table()함수 이용하여 데이터프레임 생성

- 외부 파일을 읽어와서 데이터 프레임 생성
- read.table("외부파일 경로지정")

텍스트파일 자료가 예를 들어, "d:\\r1\\personal.txt"에 저장되어 있다고 할 경우 그림처럼 코딩하여 데이터프레임으로 변환할 수 있습니다.



personal.txt - 메모장

이름	거주지역	나이	성별
홍민수	인천	25	남
조정란	경기	33	여
국정수	서울	43	남
라윤정	서울	35	여
한주연	인천	37	남

머릿글(헤더)이 존재합니다.

```
> pinfo<-read.table("d:\\r1\\personal.txt", header=TRUE)
```

```
> pinfo
  이름 거주지역 나이 성별
  1 홍민수   인천    25   남
  2 조정란   경기    33   여
  3 국정수   서울    43   남
  4 라윤정   서울    35   여
  5 한주연   인천    37   남
> |
```

머릿글(헤더)이 존재하므로 데이터는 그 다음줄부터 인식하게 됩니다.

◆ 데이터프레임 합치기

- 여러 데이터프레임을 합하는 작업
- 열(옆) 방향으로 합치기 – cbind() 함수, 행(아래) 방향으로 합치기 – rbind() 함수 이용
- 데이터 프레임에서 중복된 열이 있을 경우 merge() 함수 이용하여 병합

I. 열방향으로 합치기 – cbind() 함수

기존에 데이터프레임 pinfo와 그림처럼 코딩하여 lifeinfo 프레임 객체를 생성하여 두 개를 합치는 작업 (행의 길이는 동일합니다.)

```
> lifeinfo<-data.frame(취미=c("목공예", "패러글라이딩", "등산", "독서
+ 관심사=c("취업", "아파트분양", "주택대출", "이직", "교통인프라"))
> lifeinfo
```

lifeinfo 데이터 프레임의 행의 길이=5

	취미	관심사
1	목공예	취업
2	패러글라이딩	아파트분양
3	등산	주택대출
4	독서	이직
5	트래킹	교통인프라

```
> myinfo<-cbind(pinfo, lifeinfo)
```

두 프레임을 옆으로 결합하여 myinfo 프레임 객체를 생성

```
> myinfo
```

	이름	거주지역	나이	성별
1	홍민수	인천	25	남
2	조정란	경기	33	여
3	국정수	서울	43	남
4	라윤정	서울	35	여
5	한주연	인천	37	남

	취미	관심사
1	목공예	취업
2	패러글라이딩	아파트분양
3	등산	주택대출
4	독서	이직
5	트래킹	교통인프라

2. 행방향으로 합치기 – rbind() 함수

주의!! 행방향으로 합칠경우에는 자료의 열의 개수가 동일해야 하고 변수 이름도 동일해야 합니다.

기존에 데이터프레임 pinfo와 그림처럼 코딩하여 otherinfo 데이터 프레임을 생성하여 합치는 작업

```
> otherinfo<-data.frame(이름=c("황성주","윤준영"),거주지역=c("강원","충북"),
+ 나이=c(42,45),성별=c("여","여"))
> otherinfo
  이름 거주지역 나이 성별
1 황성주    강원   42   여
2 윤준영    충북   45   여
> totinfo<-rbind(pinfo,otherinfo) pinfo 의 프레임 변수와 동일해야합니다.
> totinfo 두 프레임을 아래방향으로 결합하여 totinfo 프레임 객체를 생성
  이름 거주지역 나이 성별
1 홍민수    인천   25   남
2 조정란    경기   33   여
3 국정수    서울   43   남
4 라윤정    서울   35   여
5 한주연    인천   37   남
6 황성주    강원   42   여
7 윤준영    충북   45   여
```

3. merge() 함수 이용한 프레임 합치기

두 데이터 프레임의 동일한 변수에 대해서 같은 값을 갖는 행을 기준으로 병합

pjobj.txt - 메모장	
파일(F)	편집(E)
서식(O)	보기(V)
이름	직업
홍민수	학생
조정란	공무원
박장곤	회사원
한지수	자영업
한주연	프리랜서
주민정	학생

위 내용을 데이터프레임 pjobjinfo로 생성하여 pinfo와 병합하는 작업

```
> pjobjinfo<-read.table("d:\\r1\\pjobj.txt",header=TRUE)
> pjobjinfo
  이름    직업
1 홍민수   학생
2 조정란   공무원
3 박장곤   회사원
4 한지수   자영업
5 한주연   프리랜서
6 주민정   학생
>
```

	이름	거주지역	나이	성별
1	홍민수	인천	25	남
2	조정란	경기	33	여
3	국정수	서울	43	남
4	라윤정	서울	35	여
5	한주연	인천	37	남

```
> merge(pinfo,pjobjinfo)
      이름    거주지역    나이    성별    직업
1 조정란    경기    33    여    공무원
2 한주연    인천    37    남    프리랜서
3 홍민수    인천    25    남    학생
>
```

두 프레임에서 공통적으로 갖고 있는 “이름” 컬럼에서 같은 값을 갖는 행을 결합하여 결과를 보여줍니다.

◆ 데이터 프레임 조작작업

- `subset()` 함수 이용하여 특정 컬럼만으로 이루어진 프레임 생성
- 데이터 프레임에 컬럼을 추가하는 작업
- 특정 행/열을 삭제하는 작업

I. `subset()`함수 이용 - 기존 프레임에서 원하는 컬럼만으로 구성된 새 프레임을 생성

```
> data1<-subset(pinfo,select=-거주지역)
> data1
  이름 나이 성별
1 홍민수 25 남
2 조정란 33 여
3 국정수 43 남
4 라윤정 35 여
5 한주연 37 남
```

pinfo 프레임에서 거주지역을 제외하여 data1 프레임 생성
select 속성 -컬럼명: 해당 컬럼을 제외, 컬럼명 : 해당 컬럼을 포함

2. pinfo에 “결혼” 컬럼을 추가하는 작업

```
> data2<-cbind(pinfo,결혼=c(FALSE,TRUE,TRUE,FALSE,FALSE))
> data2
```

	이름	거주지역	나이	성별	결혼
1	홍민수	인천	25	남	FALSE
2	조정란	경기	33	여	TRUE
3	국정수	서울	43	남	TRUE
4	라윤정	서울	35	여	FALSE
5	한주연	인천	37	남	FALSE

3. 데이터프레임의 특정 행/열 삭제

> data2[-3,]

3행을 삭제

	이름	거주지역	나이	성별	결혼
1	홍민수	인천	25	남	FALSE
2	조정란	경기	33	여	TRUE
4	라윤정	서울	35	여	FALSE
5	한주연	인천	37	남	FALSE

> data2[,-5]

5열을 삭제

	이름	거주지역	나이	성별
1	홍민수	인천	25	남
2	조정란	경기	33	여
3	국정수	서울	43	남
4	라윤정	서울	35	여
5	한주연	인천	37	남