

#현재 디렉토리

```
getwd()
```

#작업 디렉토리 지정

```
setwd("C:\\Users\\WPC\\Desktop\\WR과제")
```

```
getwd()
```

#기본 디렉토리 지정

```
setwd("c:\\Users\\WPC\\Desktop\\WR과제")
```

#파일 읽어오기

#통계청_전국대학교 개수

#엑셀 데이터 data.frame 형식으로 수정

```
ds <- read.csv(file="C:\\Users\\WPC\\Desktop\\WR대학교 개수.csv", head = T) #파일 불러오기
```

```
str(ds) #자료구조
```

```
head(ds) #위에서 여섯번째까지 데이터 출력
```

```
colnames(ds) <- c("DATA_A", "DATA_B", "DATA_C", "DATA_D") #열 이름 부여
```

#데이터 확인

```
S1 <- str(ds) #자료구조
```

```
S2 <- nrow(ds) #행 개수
```

```
S2
```

```
S3 <- ncol(ds) #열 개수
```

```
S3
```

```
S4 <- head(ds) #위에서 여섯번째 열까지 출력
```

```
S4
```

```
S5 <- tail(ds) #밑에서 여섯번째 열까지 출력
```

```
S5
```

```
S6 <- summary(ds) #각 열의 5분위수 표시
```

```
S6
```

#통계적 수치(DATA_B열)

```
D1 <- min(ds$DATA_B) #최솟값
```

D1

```
D2 <- max(ds$DATA_B) #최댓값
```

D2

```
D3 <- range(ds$DATA_B) #최댓값과 최솟값의 차
```

D3

```
D4 <- diff(range(ds$DATA_B)) #직전 값과의 차이
```

D4

```
D5 <- median(ds$DATA_B) #중위수
```

D5

```
D6 <- mean(ds$DATA_B) #평균
```

D6

```
D7 <- var(ds$DATA_B) #분산
```

D7

```
D8 <- sd(ds$DATA_B) #표준편차
```

D8

```
D9 <- quantile(ds$DATA_B) #5분위수
```

D9

```
D10 <- IQR(ds$DATA_B) #사분범위
```

D10

```
D11 <- fivenum(ds$DATA_B) #5분위수(수치만 출력)
```

D11

파일 불러오기

```
> #현재 디렉토리
> getwd()
[1] "C:/Users/PC/Desktop/R과제"
> #작업 디렉토리 지정
> setwd("C:\\Users\\PC\\Desktop\\R과제")
> getwd()
[1] "C:/Users/PC/Desktop/R과제"
> #기본 디렉토리 지정
> setwd("c:\\Users\\PC\\Desktop\\R과제")
> ds <- read.csv(file="C:\\Users\\PC\\Desktop\\대학교 개수.csv", head = T) #파일 불러오기
```

총 출력값

Data	
ds	18 obs. of 4 variables
s4	6 obs. of 4 variables
s5	6 obs. of 4 variables
Values	
D1	2L
D10	9.75
D11	num [1:5] 2 11 18.5 22 339
D2	339L
D3	int [1:2] 2 339
D4	337L
D5	18.5
D6	37.6666666666667
D7	5874.70588235294
D8	76.6466299477866
D9	Named num [1:5] 2 12 18.5 21.8 339
S1	NULL (empty)
S2	18L
S3	4L
S6	'table' chr [1:7, 1:4] "강원도 : 1 " "경기도 : 1 " ...

#자료구조확인

```
> str(ds) #자료구조
'data.frame': 18 obs. of 4 variables:
 $ 행정구역별: Factor w/ 18 levels "강원도","경기도",...: 13 9 8 6 12 5 7 11 10 2 ...
 $ x2016 : int 339 48 22 11 7 17 15 4 2 60 ...
 $ x2017 : int 339 48 22 11 7 17 15 4 2 60 ...
 $ x2018 : int 340 48 22 11 7 17 15 4 3 61 ...
> head(ds) #위에서 여섯번째까지 데이터 출력
 행정구역별 x2016 x2017 x2018
1 전국 339 339 340
2 서울특별시 48 48 48
3 부산광역시 22 22 22
4 대구광역시 11 11 11
5 인천광역시 7 7 7
6 광주광역시 17 17 17
> colnames(ds) <- _c("DATA_A", "DATA_B", "DATA_C", "DATA_D") #열 이름 부여
```

#자료구조확인2

```
> str(ds) #사표구조
'data.frame': 18 obs. of 4 variables:
 $ DATA_A: Factor w/ 18 levels "강원도","경기도",...: 13 9 8 6 12 5 7 11 10 2 ...
 $ DATA_B: int 339 48 22 11 7 17 15 4 2 60 ...
 $ DATA_C: int 339 48 22 11 7 17 15 4 2 60 ...
 $ DATA_D: int 340 48 22 11 7 17 15 4 3 61 ...
> S2 <- nrow(ds) #행 개수
> S2
[1] 18
> S3 <- ncol(ds) #열 개수
> S3
[1] 4
> S4 <- head(ds) #위에서 여섯번째 열까지 출력
> S4
  DATA_A DATA_B DATA_C DATA_D
1      전국    339    339    340
2 서울특별시    48     48     48
3 부산광역시    22     22     22
4 대구광역시    11     11     11
5 인천광역시     7      7      7
6 광주광역시    17     17     17
> S5 <- tail(ds) #밑에서 여섯번째 열까지 출력
> S5
  DATA_A DATA_B DATA_C DATA_D
13 충청남도     21     21     21
14 전라북도     20     20     19
15 전라남도     19     19     19
16 경상북도     34     34     33
17 경상남도     21     21     21
18 제주특별자치도 4      4      4
> S6 <- summary(ds) #각 열의 5분위수 표시
> S6
  DATA_A      DATA_B      DATA_C      DATA_D
강원도 : 1   Min.   : 2.00   Min.   : 2.00   Min.   : 3.00
경기도 : 1   1st Qu.: 12.00   1st Qu.: 12.00   1st Qu.: 12.00
경상남도 : 1   Median : 18.50   Median : 18.50   Median : 18.50
경상북도 : 1   Mean    : 37.67   Mean    : 37.67   Mean    : 37.78
광주광역시: 1   3rd Qu.: 21.75   3rd Qu.: 21.75   3rd Qu.: 21.75
대구광역시: 1   Max.    :339.00   Max.    :339.00   Max.    :340.00
(other)  :12
```

#통계수치

```
> #통계적 수치(DATA_B열)
>
> D1 <- min(ds$DATA_B) #최솟값
> D1
[1] 2
>
> D2 <- max(ds$DATA_B) #최댓값
> D2
[1] 339
>
> D3 <- range(ds$DATA_B) #최댓값과 최솟값의 차
> D3
[1] 2 339
>
> D4 <- diff(range(ds$DATA_B)) #직전값과의 차이
> D4
[1] 337
>
> D5 <- median(ds$DATA_B) #중위수
> D5
[1] 18.5
>
> D6 <- mean(ds$DATA_B) #평균
> D6
[1] 37.66667
>
> D7 <- var(ds$DATA_B) #분산
> D7
[1] 5874.706
>
> D8 <- sd(ds$DATA_B) #표준편차
> D8
[1] 76.64663
>
> D9 <- quantile(ds$DATA_B) #5분위수
> D9
      0%      25%      50%      75%     100%
  2.00  12.00  18.50  21.75  339.00
>
> D10 <- IQR(ds$DATA_B) #사분범위
> D10
[1] 9.75
>
> D11 <- fivenum(ds$DATA_B) #5분위수(수치만)
> D11
[1] 2.0 11.0 18.5 22.0 339.0
```