

---

Chapter 4

# 데이터 준비





# 데이터 읽기

함수

기능

`read.csv("파일명")`

csv파일 읽기

`read.delim("파일명", sep = ",")`

특정 구분자로 파일 읽기

`read.table("파일명", sep = ",")`

특정 구분자로 text파일 읽기

`read_excel("파일명")`

엑셀 파일 읽어오기(readxl 패키지 필요)

# csv불러오기

## csv 파일 불러오기

- 범용 데이터 형식
- 값 사이를 쉼표(,)로 구분
- 용량 작음, 다양한 소프트웨어에서 사용

```
df_csv_exam <- read.csv("csv_exam.csv")  
df_csv_exam
```

```
##      id class math english Science  
## 1     1     1   50      98       50  
## 2     2     1   60      97       60  
## 3     3     1   45      86       78  
## 4     4     1   30      98       58  
## 5     5     2   25      80       65  
## 6     6     2   50      89       98  
## 7     7     2   80      90       45  
## 8     8     2   90      78       25  
## 9     9     3   20      98       15  
## 10    10     3   50      98       45
```

문자가 들어 있는 파일을 불러올 때는 `stringsAsFactors = F`

```
df_csv_exam <- read.csv("csv_exam.csv", stringsAsFactors = F)
```

# 엑셀 불러오기

## 엑셀 파일 불러오기

```
# readxl 패키지 설치
install.packages("readxl")
```

```
# readxl 패키지 로드
library(readxl)
```

```
df_exam <- read_excel("excel_exam.xlsx") # 엑셀 파일을 불러와서 df_exam 에 할당
df_exam
```

```
##      id class  math english science
##   <dbl> <dbl> <dbl>   <dbl>   <dbl>
## 1     1     1    50     98     50
## 2     2     1    60     97     60
## 3     3     1    45     86     78
## 4     4     1    30     98     58
## 5     5     2    25     80     65
## 6     6     2    50     89     98
## 7     7     2    80     90     45
## 8     8     2    90     78     25
## 9     9     3    20     98     15
```



# 엑셀 불러오기

엑셀 파일 첫 번째 행이 변수명이 아니라면?

```
df_exam_novar <- read_excel("excel_exam_novar.xlsx", col_names = F)  
df_exam_novar
```

엑셀 파일에 시트가 여러 개 있다면?

```
df_exam_sheet <- read_excel("excel_exam_sheet.xlsx", sheet = 3)  
df_exam_sheet
```

# 파일 저장하기

## csv 파일로 저장하기

```
df_midterm <- data.frame(english = c(90, 80, 60, 70),  
                          math = c(50, 60, 100, 20),  
                          class = c(1, 1, 2, 2))
```

```
df_midterm
```

##		english	math	class
##	1	90	50	1
##	2	80	60	1
##	3	60	100	2
##	4	70	20	2

```
write.csv(df_midterm, file = "df_midterm.csv")
```



# 제목

---



# Chapter 4

## 수고하셨습니다