

파일 불러오기(Import)

파일 불러오기(Import)

1. 외부 파일의 종류
2. 텍스트 파일 불러오기
 - 2-1. 텍스트 파일의 형식
 - 2-2. 텍스트파일 읽는 함수
 - 1) read.table(), read.csv() 함수 예제
 - 2) read_csv 함수 예제
 - 3) 예제: 수강생 자료
3. 엑셀 관련 R 패키지

1. 외부 파일의 종류

1) 텍스트파일 2) 엑셀파일 3) xml/json파일 4) 기타 : SPSS, SAS 데이터 파일

2. 텍스트 파일 불러오기

2-1. 텍스트 파일의 형식

- 1) 텍스트자료에서 첫번째 줄은 '변수이름'을 나타내며, 두번째 줄부터 '데이터'가 입력됨
- 2) '데이터'와 '데이터'를 구분하는 문자를 '구분자(seperator)'라고 하며 주로 '공백(" ")', 콤마(","), TAB("\t") 등이 사용됨

• 텍스트 파일의 예 (첫줄은 변수명, 두번째부터는 공백으로 구분된 자료값)

```
Surv N Class Age Sex
20 23 Crew Adult Female
192 862 Crew Adult Male
1 1 First Child Female
5 5 First Child Male
13 13 Second Child Female
```

• 구분자가 콤마인 파일: 보통 CSV파일(Comma Separated Values)로 저장됨

```
Surv,N,Class,Age,Sex
20,23,Crew,Adult,Female
192,862,Crew,Adult,Male
1,1,First,Child,Female
5,5,First,Child,Male
13,13,Second,Child,Female
```

2-2. 텍스트파일 읽는 함수

1) read.table(), read.csv(), read.delim() 2) read_csv(), read_delim() : readr패키지를 이용

1) read.table(), read.csv() 함수 예제

```
titanic <- read.table("c:/temp/titanic.txt", header=T, sep="\t")
head(titanic)
```

2) read_csv 함수 예제

```
#install.packages("readr")
library(readr)
titanic3 <- read_csv("c:/temp/titanic.csv")    ## file name
```

3) 예제: 수강생 자료

- test.csv 자료 읽기

```
fileEncoding = "CP949": csv파일에서 한글 입력 코딩방식, CP949 또는 UTF-8

test <- read.csv("c:/temp/test.csv")
head(test)
names(test)
dim(test)
```

- test 데이터프레임에서 학과별 수강생 수를 구하고, 막대 그래프 그리기

```
unique(testDept)
(nid <- table(testDept))
barplot(nid, main="학과별 수강생 수")
```

3. 엑셀 관련 R 패키지

- readxl() / writexl() • 적은 메모리 사용으로 빠르게 데이터를 읽어옴 • 결과는 데이터프레임이 아니라 tibble형식으로 저장됨 - tibble형식은 데이터프레임의 확장된 데이터클래스임 • 1980~2018년까지 연도 별/성별 고용률 자료 (http://www.index.go.kr/potal/stts/idxMain/selectPoSttsIdxSearch.do?idx_cd=1494)

```
# install.packages("readxl", dependencies=T)

library(readxl)
o <- read_excel("c:/temp/고용률.xls", sheet=1); head(o)

library(writexl)
write_xlsx(o, "c:/temp/고용률(저장).xlsx")
```

