

제 12장 데이터 입출력

1. 데이터 파일 입력

표 1. 데이터 파일 입력 함수

```
> gpa<-read.table("C:/Rdata/gpa.txt", header=T) # 변수명이 있을 때
> score<-read.table("C:/Rdata/score.csv", header=T, sep=",")
> class(score); class(gpa)
```

#htm 페이지에서 표 형식의 데이터 가져오기

```
> install.packages("rvest")
> library(rvest)
> install.packages("assertthat")
> library(assertthat)
# 2020년 세계인구수 (235개국)
인터넷 주소창에 https://www.worldometers.info/world-population/population-by-country/
> webpage="https://www.worldometers.info/world-population/population-by-country/"
> x <- read_html(webpage) #html
> tbls <- (html_nodes(x, "table")) #xml
> tbls
> y=html_table(tbls, fill = TRUE)
> y # 인구수 table
> class(y) #list
> length(y[[1]]) # 12개 : 순위-국가명 --- 세계총인구 중 비율%, 우리나라=28위
> y[[1]][1] # 순위
> y[[1]][2] # 국가명
> y[[1]][3] # 인구
> y[[1]][9] # 가임률(fertility rate)
> y[[1]][10] # 중앙나이(median age)

> z=as.data.frame(y)
> names(z) # 특수문자는 변환이 잘안됨
> names(z)[1]<-"rank"
> name(z)
> z[z[,2]=="South Korea",3] # 우리나라의 인구
> z[z[,2]=="South Korea",4] # 우리나라의 인구증가율
> z3<-z[,3] #인구 or z[[3]]
> head(z3)
> class(z3) # 문자형 벡터
> z3<-as.numeric(gsub(",", "", z3)) # z3=수치형벡터, 천단위 “,” 제거
> head(z3)
> class(z3) # 수치형 벡터
> sum(z3) # 235개국 인구 총합
```

2. 데이터 파일 출력

표 2. 데이터 파일 출력 함수

```
> rm(list=ls()) # work space 모든 객체 삭제
> A<-c(73,65,58,81,88,75,82,92); B<-c(83,55,68,61,58,70,72,69)
> AB<-data.frame(A,B)
> AB
> write.table(AB,"C:/Rdata/AB.txt")
> save(AB,file="C:/Rdata/AB.RData")
# C:/Rdata 에 AB.txt, AB.RData 있는지 확인
> ls() # work space 에 있는 객체 보기
> rm(AB) # AB 삭제
> ls()
> load(file="C:/Rdata/AB.RData")
> ls()
> AB
```

3. 변수 타입 변환

표 3. 데이터 타입에 관한 함수

(1) is_type: 데이터 type 판별

```
> x<-c(1,2,3)
> is.character(x)
> is.numeric(x)
> w<-c(T,F,F)
> is.logical(w)
> y<-c(3,4,6)
> z<-data.frame(x,y)
> is.data.frame(z)
> z
> z$x
> z$y
```

(2) as_type: 데이터 type 변환

표 4. 데이터 변환에 관한 함수

```
> w1<-as.numeric(w)
> z1<-as.matrix(z)
> is.matrix(z1)
> z1
> z1[,1]
> z1[,2]

> data(iris)
> class(iris)
> head(iris) #tail(iris)
> View(iris) # 소스(Source)창에서 자료보기
```