

텍스트 데이터 읽어오기

- 📊 텍스트 데이터의 개념
- 📊 구분자가 공백인 경우
- 📊 구분자가 콤마인 경우
- 📊 구분자가 탭인 경우



텍스트 데이터의 개념



텍스트 데이터의 개념

- ▶ 텍스트 데이터는 데이터와 데이터를 구분하기 위해서 공백(blank), 콤마(comma), 탭(tab)과 같은 **구분자(separator)**를 사용
- ▶ 텍스트 데이터를 R에서 읽어오기 위해서는 **read.table()** 함수를 사용

argument	설명
file	R에서 읽어올 외부 데이터가 있는 파일의 위치(경로)와 파일명(확장자 포함)을 문자형으로 지정
header	<ul style="list-style-type: none">• 논리형으로 TRUE를 지정하면 외부 데이터에 있는 변수명을 R 데이터에서도 동일하게 사용• FALSE를 지정하면 외부 데이터의 변수명을 사용하지 않고 R에서 자체적으로 변수명을 지정(이름은 "V1", "V2" 식으로 됨)
sep	구분자로 문자형 형태로 지정
stringsAsFactors	<ul style="list-style-type: none">• 문자형 데이터를 요인(factor)으로 자동으로 변경할 지를 설정• FALSE를 지정하면 문자형 데이터를 요인으로 변경하지 않고 문자형으로 읽어 들임
na.strings	데이터에 결측치가 있거나 결측치로 지정할 내용을 문자형으로 지정

[read.table() 함수의 사용법]

본 문서는 저작권자의 동의 없이 무단 복제 및 배포할 수 없습니다.

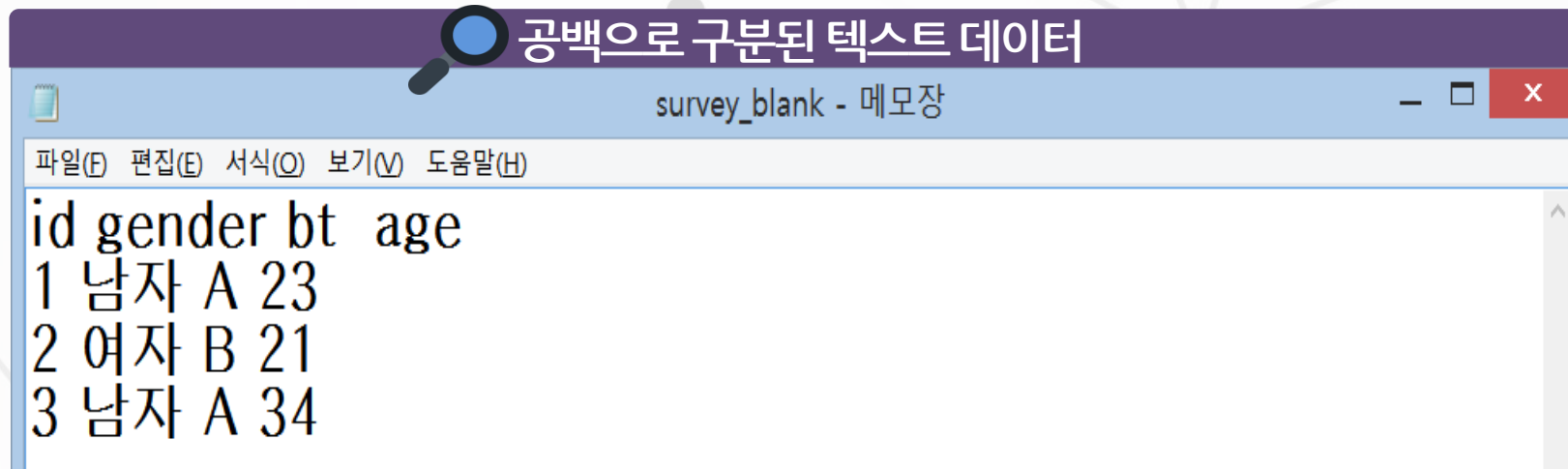


구분자가 공백인 경우



구분자가 공백인 경우

- ▶ 데이터가 공백으로 구분된 경우 다음과 같이 나타남



공백으로 구분된 텍스트 데이터

id	gender	bt	age
1	남자	A	23
2	여자	B	21
3	남자	A	34



구분자가 공백인 경우



구분자가 공백인 경우

- > 파일의 위치가 d드라이브에 있고, **파일이름이 survey_blank.txt로 되어 있을 때**
R에서 읽어오는 방법

```
데이터명 = read.table(file="파일위치/파일명.txt", header=TRUE, sep=" ")  
survey_rblank = read.table(file="d:/survey_blank.txt", header=TRUE, sep=" ")
```



구분자가 콤마인 경우



구분자가 콤마인 경우

- > 데이터가 콤마(,)로 구분된 경우 다음과 같이 나타남

```
id,height,weight  
1,175,70  
2,167,58  
3,178,82  
4,162,45
```




구분자가 콤마인 경우



구분자가 콤마인 경우

- > 파일의 위치가 d드라이브에 있고, **파일이름이** `survey_comma.txt`로 되어 있을 때 R에서 읽어오는 방법

```
데이터명 = read.table(file="파일위치/파일명.txt", header=TRUE, sep=",")  
survey_comma = read.table(file="d:/survey_comma.txt", header=TRUE, sep=",")
```

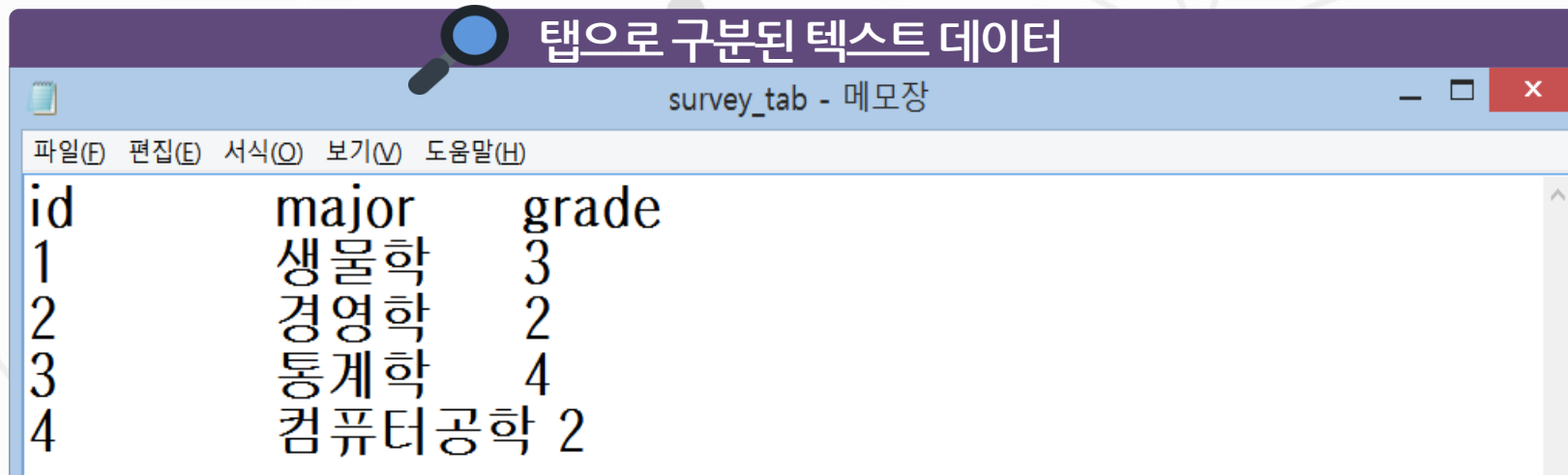


구분자가 탭인 경우



구분자가 탭인 경우

- ▶ 데이터가 탭(tab)으로 구분된 경우 다음과 같이 나타남



탭으로 구분된 텍스트 데이터

id	major	grade
1	생물학	3
2	경영학	2
3	통계학	4
4	컴퓨터공학	2



구분자가 탭인 경우



구분자가 탭인 경우

- > 파일의 위치가 d드라이브에 있고, **파일이름이 survey_tab.txt로 되어 있을 때**
R에서 읽어오는 방법

```
데이터명 = read.table(file="파일위치/파일명.txt", header=TRUE, sep="\t")  
survey_tab = read.table(file="d:/survey_tab.txt", header=TRUE, sep="\t")
```


CSV 데이터 읽어오기

 read.csv() 함수의 사용



read.csv() 함수의 사용



CSV 데이터 읽어오기

- 엑셀 데이터의 특별한 형태로 **콤마로 구분된 형태의 데이터**
- Comma Separated Value의 약자
- CSV 데이터는 **read.csv()** 함수를 사용

argument	설명
file	R에서 읽어올 외부 데이터가 있는 파일의 위치(경로)와 파일명(확장자 포함)을 문자형으로 지정
header	논리형으로 CSV 데이터 있는 변수명을 사용하려면 TRUE를 지정

[read.csv() 함수의 사용법]



read.csv() 함수의 사용



CSV 데이터를 읽어오는 방법

CSV 데이터

인구주택총조사2015 - Excel

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 추가 기능 Inquire 파워 피벗 팀 ? 수행할 작업을 알려주세요. 로그인 공유

잘라내기 붙여넣기 서식 복사 클립보드 글꼴 맞춤 표시 형식 스타일 셀 편집

C9 1455017

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	행정구역별	총인구	남자	여자	내국인계	내국인_남자	내국인_여자	외국인계	외국인_남자	외국인_여자	가구계	일반가구	집단가구	외국인가구
2	전국	51069375	25608502	25460873	49705663	24819839	24885824	1363712	788663	575049	19560603	19111030	16464	433
3	읍부	4616802	2360708	2256094	4467697	2262853	2204844	149105	97855	51250	1737415	1695508	2189	39
4	면부	4774878	2455898	2318980	4546520	2291860	2254660	228358	164038	64320	1981033	1927621	4284	49
5	동부	41677695	20791896	20885799	40691446	20265126	20426320	986249	526770	459479	15842155	15487901	9991	344
6	서울특별시	9904312	4859535	5044777	9567196	4694317	4872879	337116	165218	171898	3914820	3784490	2261	128
7	부산광역시	3448737	1701347	1747390	3404667	1675339	1729328	44070	26008	18062	1348315	1335900	686	11
8	대구광역시	2466052	1228511	1237541	2436770	1211219	1225551	29282	17292	11990	937573	928528	574	8
9	인천광역시	2890451	1455017	1435434	2822601	1414793	1407808	67850	40224	27626	1066297	1045417	713	20
10	광주광역시	1502881	748867	754014	1481289	736656	744633	21592	12211	9381	573181	567157	438	5
11	대전광역시	1538394	772243	766151	1519314	763310	756004	19080	8933	10147	588395	582504	420	5
12	울산광역시	1166615	606924	559691	1136755	587603	549152	29860	19321	10539	434058	423412	239	10
13	세종특별자치시	204088	103210	100878	199617	100455	99162	4471	2755	1716	76419	75219	65	1
14	경기도	12479061	6309661	6169400	12026429	6039800	5986629	452632	269861	182771	4537581	4384742	3879	148
15	강원도	1518040	768241	749799	1499734	758601	741133	18306	9640	8666	611578	606117	788	4
16	충청북도	1589347	805377	783970	1548589	779961	768628	40758	25416	15342	613004	601856	814	10
17	충청남도	2107802	1078310	1029492	2036720	1034139	1002581	71082	44171	26911	816247	796185	1107	18
18	전라북도	1834114	915729	918385	1804184	899308	904876	29930	16421	13509	726572	717311	834	8
19	전라남도	1799044	900967	898077	1764433	879860	884573	34611	21107	13504	730743	720612	933	9
20	경상북도	2680294	1351037	1329257	2622729	1314196	1308533	57565	36841	20724	1078479	1062724	1312	14
21	경상남도	3334524	1698737	1635787	3244163	1634407	1609756	90361	64330	26031	1282617	1258487	1176	22
22	제주특별자치도	605619	304789	300830	590473	295875	294598	15146	8914	6232	224724	220369	225	4
23														

인구주택총조사2015

준비



read.csv() 함수의 사용



CSV 데이터를 읽어오는 방법

- > d 드라이브에 인구주택총조사2015.csv 데이터가 있을 때 R에서 읽어오는 방법

```
데이터명= read.csv(file="파일위치/파일명.csv", header=TRUE)
```

```
survey2015 = read.csv(file="d:/인구주택총조사2015.csv", header=TRUE)
```


엑셀 데이터 읽어오기

엑셀 데이터 읽어오기



엑셀 데이터 읽어오기



- 엑셀 데이터는 R의 기본 기능으로는 읽어오는 함수가 없기 때문에 **엑셀 데이터를 읽어오는 기능을 제공하는 새로운 패키지를 설치해야 함**
- 엑셀 데이터를 읽어오는 기능을 제공하는 패키지는 여러 개가 있지만 패키지를 위해서 별도의 작업을 해야 하는 불편함이 있음
- 해들리 위컴(Hadley Wickham)이 개발한 **readxl 패키지**에서 제공하는 `read_excel()` 함수를 이용하여 엑셀 데이터를 읽어오는 방법을 학습하도록 함



엑셀 데이터 읽어오기



readxl 패키지의 설치와 로딩

- > readxl 패키지를 설치하고 로딩하는 방법

```
install.packages("readxl")  
library(readxl)
```



엑셀 데이터 읽어오기



read_excel() 함수 사용법

argument	설명
path	R에서 읽어올 외부 데이터가 있는 파일의 위치(경로)와 파일명(확장자 포함)을 문자형으로 지정
sheet	엑셀 파일의 있는 시트(sheet) 중에서 어떤 시트인지를 지정 <ul style="list-style-type: none">• 시트명을 알고 있으면 문자형으로 지정• 시트의 위치(index)를 알면 수치형으로 지정
col_names	<ul style="list-style-type: none">• 논리형으로 엑셀에 있는 변수명을 사용할 지를 의미• TRUE이면 엑셀에 있는 변수명을 사용

[read_excel() 함수의 사용법]



엑셀 데이터 읽어오기



엑셀 데이터 불러오기

- ▶ d드라이브에 big.xls가 있고, Sheet1에 데이터가 있다고 가정하면 다음과 같이 엑셀 데이터를 불러올 수 있음
※엑셀 2003 이하 버전인 경우에는 확장자가 xls이고, 2007 이상 버전인 경우에는.xlsx임

엑셀 데이터

	A	B	C	D	E	F	G	H	I	J
1	ID	X1_1	X1_2	X1_3	X1_4	X1_5	X1_6	X1_7	X1_8	X1_9
2	1	5	5	6	5	5	5	6	6	5
3	2	4	3	5	4	2	2	4	5	3
4	3	5	5	5	5	4	4	4	4	4
5	4	5	4	5	3	4	5	3	5	5
6	5	6	6	6	5	5	6	5	5	4
7	6	4	4	5	5	4	4	4	4	3
8	7	5	5	5	4	3	3	3	6	6
9	8	7	5	7	6	7	6	6	7	7
10	9	5	4	6	5	6	6	4	6	3
11	10	6	6	7	6	5	5	5	5	6
12	11	6	5	6	6	4	5	5	7	4
13	12	6	6	6	5	6	5	5	6	6
14	13	6	6	7	7	6	6	6	7	7
15	14	7	7	7	7	7	7	7	7	7
16	15	4	4	4	5	5	5	5	5	4
17	16	4	3	5	5	5	4	2	3	5
18	17	5	6	6	6	3	4	4	5	5
19	18	6	6	6	6	6	7	7	7	7
20	19	5	5	6	4	5	5	4	6	5
21	20	5	4	5	6	5	5	4	5	5
22	21	7	7	7	7	7	7	7	6	7



엑셀 데이터 읽어오기



CSV 데이터 읽어오는 방법

- > d 드라이브에 **ebook.xlsx** 데이터가 있을 때 R에서 읽어오는 방법

```
데이터명 = read_excel(path="파일위치/파일명.xls", sheet="시트명", col_names=TRUE)  
ebook = read_excel(path="d:/ebook.xlsx", sheet="data", col_names=TRUE)
```