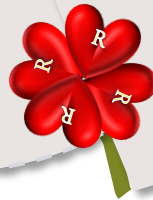


## 8.외부파일읽어오기

외부자료파일을 읽어오기위해 필요한 함수들을  
학습합니다.



## 1. scan() 함수

- 외부파일을 읽어들이
- scan('파일명',what)
- what- 문자열 인 경우 "", 또는 'char'이라고 입력해줘야만 인식가능(what='char')

```
> a<-scan('irum.txt',what='char',sep=',')
Read 6 items
> a
[1] "홍길동" "손영문" "조재희" "하철중" "국만봉" "하명순"
> a<-scan('irum.txt',what='char')
Read 1 item
> a
[1] "홍길동,손영문,조재희,하철중,국만봉,하명순"
```

## 2. readLines()

- 외부 파일을 한 줄 단위로 읽어들이
- readLines('파일명')

```
> data<-readLines('consumer.txt')
> data
[1] "422 뿌띠 성형시술 후 염증·눈꺼풀 처짐 등 부작용 호소 많아 2016-02-0$
[2] "421 방수팩 사용 관련 피해예방 주의보 2015-09-18 5650" $
[3] "420 안과 관련 소비자피해, '백내장' 등 노인성 질환 부작용 많아 2015-0$
[4] "419 해외구매대행 인터넷쇼핑몰 피해 주의 2015-08-28 7162" $
[5] "418 해외구매대행 인터넷쇼핑몰 피해 주의 2015-08-25 5693" $
[6] "417 세탁소에서 훼손된 세탁물 10건 중 3건은 품질 하자가 원인 2015-08-1$
[7] "416 여름 휴가철 렌터카 이용 시 소비자피해 주의 하세요 ! 2015-08-11 43$"
```

## 3. read.csv()

- csv 형식 파일을 읽어들이
- read.csv('파일명')

```
> popdata<-read.csv('seoulpopulation.csv',header=TRUE)
> popdata
```

	자치구	남자	여자
1	종로구	80531	83291
2	중구	66755	67574
3	용산구	121027	126882
4	성동구	151459	153606
5	광진구	183436	191744
6	동대문구	185827	187997
7	중랑구	208393	210227
8	성북구	229183	240377
9	강북구	164337	170089
10	도봉구	173804	179437
11	노원구	281538	296683
12	은평구	244964	257614
13	서대문구	156130	166975
14	마포구	190957	207394
15	양천구	242074	246936
16	강서구	291216	304475
17	구로구	228201	226403
18	금천구	131346	124821
19	영등포구	210388	207423
20	동작구	202165	210609
21	관악구	266773	262258
22	서초구	217036	234222
23	강남구	279209	302551
24	송파구	325950	341530
25	강동구	230851	232470

## 4. 엑셀파일 불러오기

### □ XLConnect패키지 이용

- 엑셀파일의 데이터 전부 혹은 일부를 지정하여 읽어올 수 있음
- `install.packages("XLConnect")` → XLConnect 패키지를 사용 , `require(XLConnect)`
- 엑셀파일을 읽어오는 함수 `loadWorkbook("파일경로")`
- 엑셀파일에서 읽어올 시트와 범위
- `readWorksheet(파일, sheet="시트명", startRow=읽어올 시작 행, startCol=읽어올 시작 열, endRow=마지막 행, endCol=마지막 열)`

```
> install.packages("XLConnect")
'XLConnectJars' (들)을 또한 설치합니다.

URL 'https://rweb.crmdata.ku.edu/cran/bin/windows/contrib/3.2/XLConnectJars_0.2-11.zip'
Content type 'application/zip' length 12947698 bytes (12.3 MB)
downloaded 12.3 MB

URL 'https://rweb.crmdata.ku.edu/cran/bin/windows/contrib/3.2/XLConnectJars_0.2-11.jar'
Content type 'application/zip' length 5110304 bytes (4.9 MB)
downloaded 4.9 MB

패키지 'XLConnectJars'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다.
패키지 'XLConnect'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다.

다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
C:\Users\Administrator\AppData\Local\Temp\RtmpKGhmfu\downloaded_packages

> require(XLConnect)
필요한 패키지를 로딩중입니다: XLConnect
필요한 패키지를 로딩중입니다: XLConnectJars
XLConnect 0.2-11 by Mirai Solutions GmbH [aut],
  Martin Studer [cre],
  The Apache Software Foundation [ctb, cph] (Apache POI, Apache Codec),
  Stephen Colebourne [ctb, cph] (Joda-Time Java library)
http://www.mirai-solutions.com ,
http://miraisolutions.wordpress.com

다음의 패키지를 부착합니다: 'XLConnect'

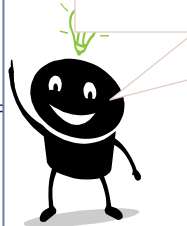
The following objects are masked from 'package:xlsx':

  createFreezePane, createSheet, createSplitPane, getCellStyle,
  saveWorkbook, setCellStyle, setColumnWidth, setRowHeight

> wb=loadWorkbook("subway.xlsx")
> df<-readWorksheet(wb, sheet=1, header=TRUE)
> df
```

구.분	역명	x1월	x2월	x3월	총계	일평균
1	1호선 서울역 (1)	4126245	3661950	4145729	11933924	132599
2	1호선 시청 (1)	1499505	1229076	1493112	4221693	46908

엑셀 데이터를 읽어오는데 만족스럽죠?  
그런데, 데이터가 많아지면 속도가 느려진다는  
단점이 있어요



## □ xlsx패키지 이용

- java를 이용하여 xls,xlsx파일을 읽어옴
- `install.packages("xlsx")` → xlsx 패키지를 사용 , `require(xlsx)`
- 엑셀파일을 읽어오는 함수 `read.xlsx("파일경로")` 또는 `read.xlsx2("파일경로")`

```
> install.packages("xlsx")
--- 현재 세션에서 사용할 CRAN 미러를 선택해 주세요 ---
URL 'https://rweb.crimda.ku.edu/cran/bin/windows/contrib/3.2/xlsx_0.5.7.zip'
Content type 'application/zip' length 400968 bytes (391 KB)
downloaded 391 KB
```

패키지 로드

패키지 'xlsx'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다

다운로드된 바이너리 패키지들은 다음의 위치에 있습니다

C:\Users\Administrator\AppData\Local\Temp\RtmpiUEKjV\downloaded\_packages

```
> require(xlsx)
필요한 패키지들 로딩중입니다: xlsx
필요한 패키지들 로딩중입니다: rJava
필요한 패키지들 로딩중입니다: xlsxjars
```

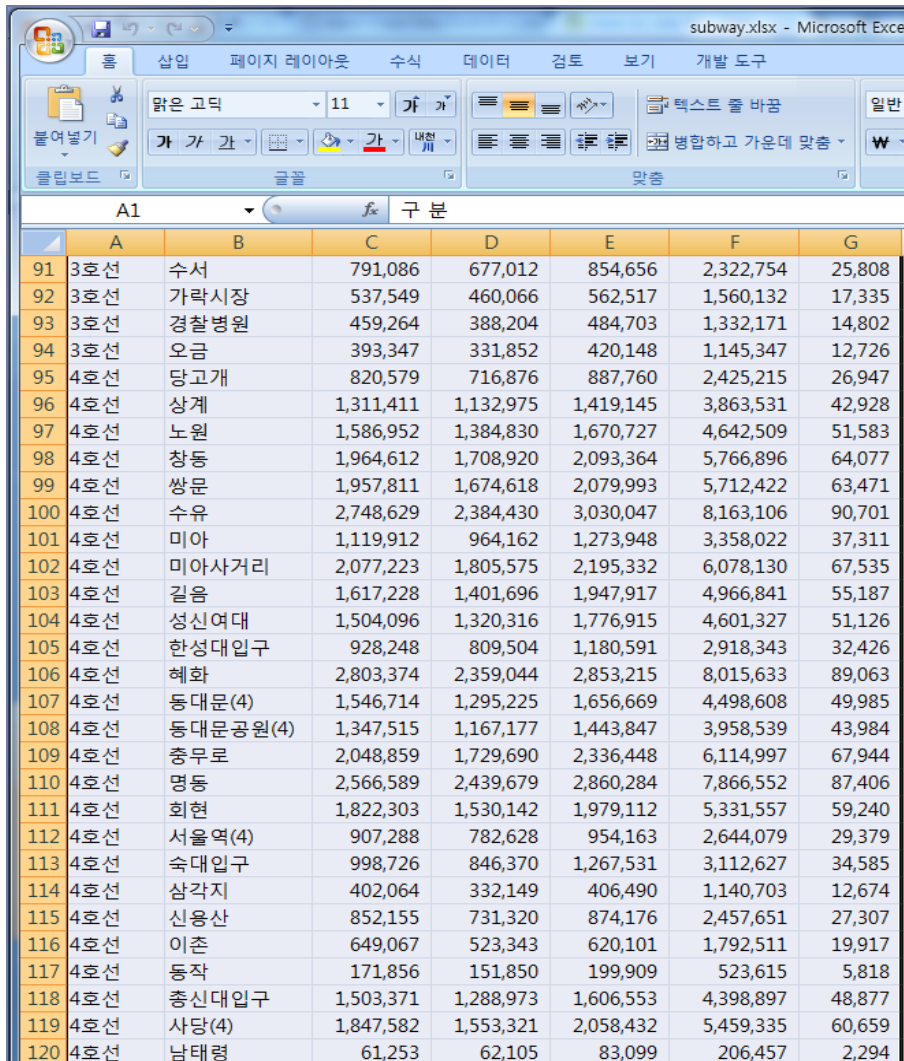
```
> df<-read.xlsx2("subway.xlsx",1)
> df
```

subway.xlsx 를 읽어옴

	구. 분	역명	X1월	X2월	X3월	총계	일평균
1	1호선	서울역 (1)	4126245	3661950	4145729	11933924	132599
2	1호선	시청 (1)	1499505	1229076	1493112	4221693	46908
3	1호선	종각	3039562	2477861	2924326	8441749	93797
4	1호선	종로3가 (1)	2435003	1999718	2290837	6725558	74728
5	1호선	종로5가	1758749	1488469	1791087	5038305	55981
6	1호선	동대문 (1)	1032643	942226	1141235	3116104	34623
7	1호선	신설동 (1)	1016609	861509	1114369	2992487	33250
8	1호선	계기동	1273577	1164476	1319059	3757112	41746
9	1호선	청량리	1898670	1683252	1973381	5555303	61726
10	1호선	동묘앞	591418	549158	686803	1827379	20304
11	2호선	시청 (2)	1533633	1269858	1589645	4393136	48813
12	2호선	을지입구	3242154	2819727	3344061	9405942	104510

## \* 엑셀데이터를 빠르게 가져와야 할 경우 -정말 유용한 방법

how to 1) 엑셀파일을 열어서 R로 가져오기를 원하는 데이터 범위를 지정합니다.



	A	B	C	D	E	F	G
91	3호선	수서	791,086	677,012	854,656	2,322,754	25,808
92	3호선	가락시장	537,549	460,066	562,517	1,560,132	17,335
93	3호선	경찰병원	459,264	388,204	484,703	1,332,171	14,802
94	3호선	오금	393,347	331,852	420,148	1,145,347	12,726
95	4호선	당고개	820,579	716,876	887,760	2,425,215	26,947
96	4호선	상계	1,311,411	1,132,975	1,419,145	3,863,531	42,928
97	4호선	노원	1,586,952	1,384,830	1,670,727	4,642,509	51,583
98	4호선	창동	1,964,612	1,708,920	2,093,364	5,766,896	64,077
99	4호선	쌍문	1,957,811	1,674,618	2,079,993	5,712,422	63,471
100	4호선	수유	2,748,629	2,384,430	3,030,047	8,163,106	90,701
101	4호선	미아	1,119,912	964,162	1,273,948	3,358,022	37,311
102	4호선	미아사거리	2,077,223	1,805,575	2,195,332	6,078,130	67,535
103	4호선	길음	1,617,228	1,401,696	1,947,917	4,966,841	55,187
104	4호선	성신여대	1,504,096	1,320,316	1,776,915	4,601,327	51,126
105	4호선	한성대입구	928,248	809,504	1,180,591	2,918,343	32,426
106	4호선	혜화	2,803,374	2,359,044	2,853,215	8,015,633	89,063
107	4호선	동대문(4)	1,546,714	1,295,225	1,656,669	4,498,608	49,985
108	4호선	동대문공원(4)	1,347,515	1,167,177	1,443,847	3,958,539	43,984
109	4호선	충무로	2,048,859	1,729,690	2,336,448	6,114,997	67,944
110	4호선	명동	2,566,589	2,439,679	2,860,284	7,866,552	87,406
111	4호선	회현	1,822,303	1,530,142	1,979,112	5,331,557	59,240
112	4호선	서울역(4)	907,288	782,628	954,163	2,644,079	29,379
113	4호선	숙대입구	998,726	846,370	1,267,531	3,112,627	34,585
114	4호선	삼각지	402,064	332,149	406,490	1,140,703	12,674
115	4호선	신용산	852,155	731,320	874,176	2,457,651	27,307
116	4호선	이촌	649,067	523,343	620,101	1,792,511	19,917
117	4호선	동작	171,856	151,850	199,909	523,615	5,818
118	4호선	충신대입구	1,503,371	1,288,973	1,606,553	4,398,897	48,877
119	4호선	사당(4)	1,847,582	1,553,321	2,058,432	5,459,335	60,659
120	4호선	남태령	61,253	62,105	83,099	206,457	2,294

excel파일을 연 다음 이렇게 범위를 지정하고 복사를 합니다.

아시다시피 복사를 하면? ..클립보드에 임시 저장되죠?

클립보드를 이용하여 r에서 import하면 됩니다. 보시죠.

how to 2) 엑셀은 열어둔 상태에서 r에서 그림처럼 3개의 인자를 유의하여 코딩하면 됩니다.

file="clipboard" → 클립보드에 저장된 내용을 import한다는 의미  
 sep → 엑셀의 각 항목들을 어떻게 구분할 것인지를 기술. 보통 탭 키로 구분("wt")  
 header → 머릿글 여부를 표시

```
> df<-read.table(file="clipboard",sep="\t",header=TRUE)
```

```
> df
```

	구.분	역명	X1월	X2월	X3월	총계	일평균
1	1호선	서울역 (1)	4,126,245	3,661,950	4,145,729	11,933,924	132,599
2	1호선	시청 (1)	1,499,505	1,229,076	1,493,112	4,221,693	46,908
3	1호선	종각	3,039,562	2,477,861	2,924,326	8,441,749	93,797
4	1호선	종로3가 (1)	2,435,003	1,999,718	2,290,837	6,725,558	74,728
5	1호선	종로5가	1,758,749	1,488,469	1,791,087	5,038,305	55,981
6	1호선	동대문 (1)	1,032,643	942,226	1,141,235	3,116,104	34,623
7	1호선	신설동 (1)	1,016,609	861,509	1,114,369	2,992,487	33,250
8	1호선	계기동	1,273,577	1,164,476	1,319,059	3,757,112	41,746
9	1호선	청량리	1,898,670	1,683,252	1,973,381	5,555,303	61,726
10	1호선	동묘앞	591,418	549,158	686,803	1,827,379	20,304
11	2호선	시청 (2)	1,533,633	1,269,858	1,589,645	4,393,136	48,813
12	2호선	을지입구	3,242,154	2,819,727	3,344,061	9,405,942	104,510
13	2호선	을지3가 (2)	1,200,727	1,006,160	1,270,607	3,477,494	38,639
14	2호선	을지4가	807,510	699,417	845,384	2,352,311	26,137
15	2호선	동대문공원 (2)	1,195,464	1,051,381	1,337,361	3,584,206	39,825
16	2호선	신당	942,833	810,601	1,011,683	2,765,117	30,724
17	2호선	상왕십리	611,512	520,960	682,174	1,814,646	20,163
18	2호선	왕십리	1,137,503	969,619	1,291,135	3,398,257	37,758
19	2호선	한양대	553,930	471,745	1,031,220	2,056,895	22,854
20	2호선	독성	946,545	813,109	1,073,954	2,833,608	31,485
21	2호선	성수	1,578,220	1,315,744	1,679,251	4,573,215	50,814
22	2호선	건대입구	2,926,931	2,531,272	3,101,898	8,560,101	95,112
23	2호선	구의	1,478,661	1,250,548	1,566,989	4,296,198	47,736
24	2호선	강변	3,168,150	2,836,674	3,243,537	9,248,361	102,760
25	2호선	잠실나루	1,104,278	932,640	1,183,057	3,219,975	35,778