논문작성을 위한



R에 오신것을 환영합니













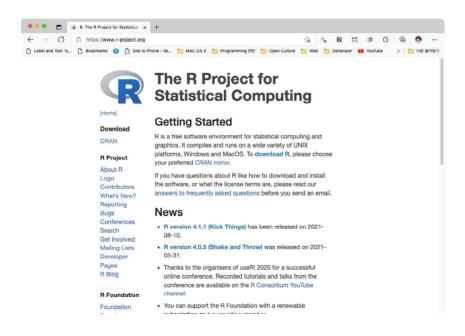


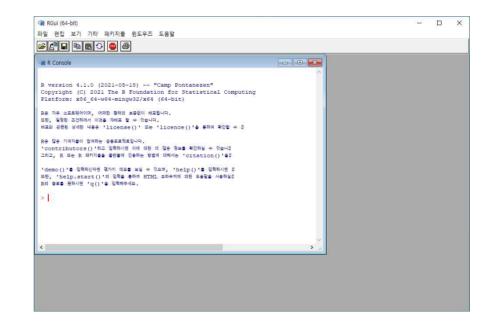






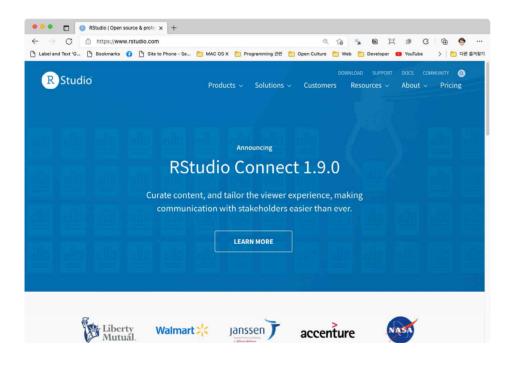
- https://www.r-project.org
- 통계 계산과 그래프 작성을 위한 프로그래밍 언어이자 소프트웨어 환경
- 1993년 뉴질랜드 오클랜드 대학의 로스 이하카(Ross Ihaka)와 로버트 젠틀맨(Robert Gentleman)에 의해 공개
- GPL하에 배포되어 비용 부담 없이 자유롭게 사용(Open Source)

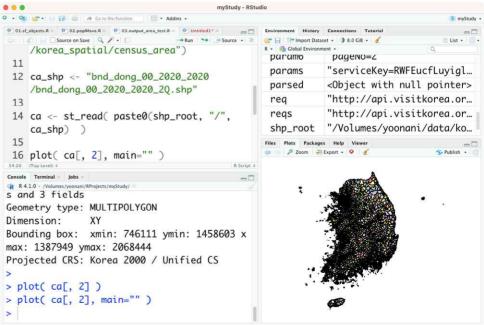






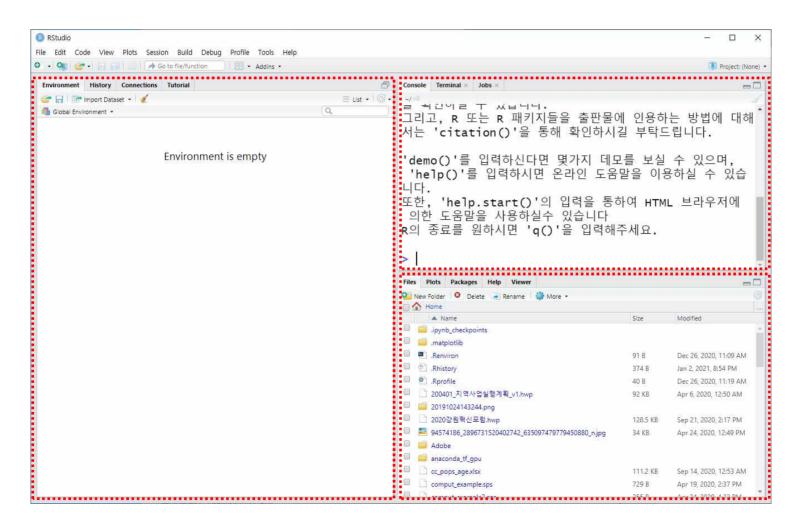
- https://www.rstudio.com
- RStudio는 R을 위한 통합개발환경(IDE)
 - 통합개발환경: 코딩, 디버그, 컴파일, 배포 등 프로그램 개발에 관련된 모든 작업을 하나의 프로그램 안에서 처리하는 환경을 제공하는 소프트웨어
- 오픈소스 버전과 상업용 버전을 제공하고 있음





RStudio 화면 구성

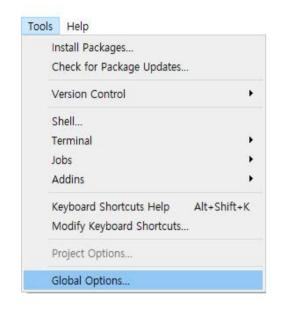
Environment History Connections Tutorial



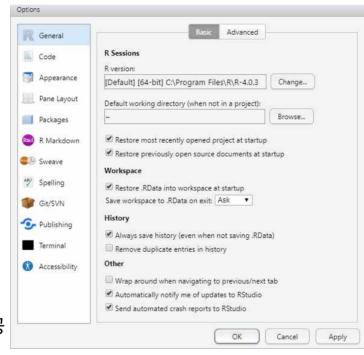
Console Terminal Jobs

Files Plots Packages Help Viewer

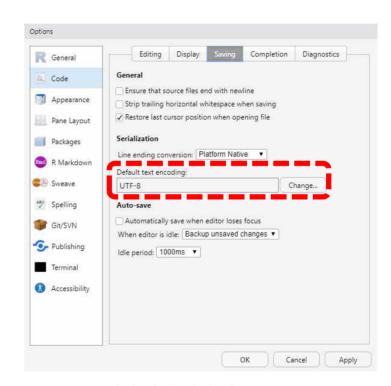
설정 변경



- 전체 설정(Global Options)과
 프로젝트 설정(Project Options) 제공
- Tools 메뉴를 통해 각 설정 진입

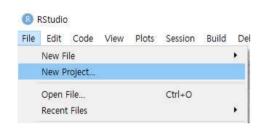


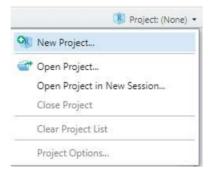
전체 설정 (Global Options)



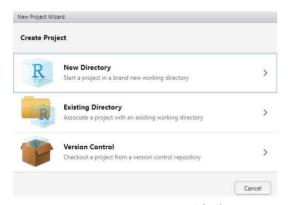
- 저장 방식 지정 예
- 기본값 Ask를 UTF-8로 변경

작업 단위 만들기: RStudio 프로젝트

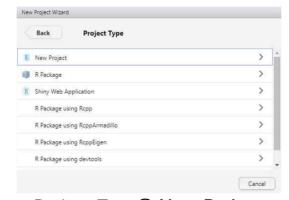




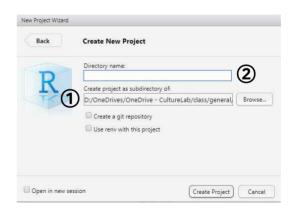
- RStudio의 프로젝트는 작업단위별로 관리할 수 있는 독립환경
- File 메뉴를 통한 생성과 우측 상단의 아이콘을 통한 생성
 - 우측 상단 아이콘은 프로젝트에 대한 보다 많은 정보 제공



New Directory 선택

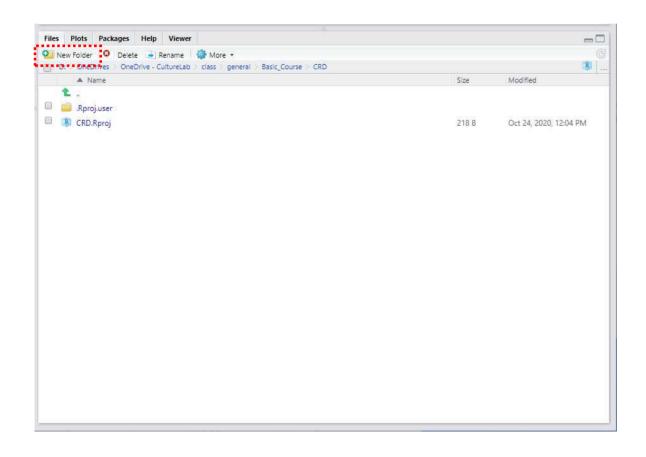


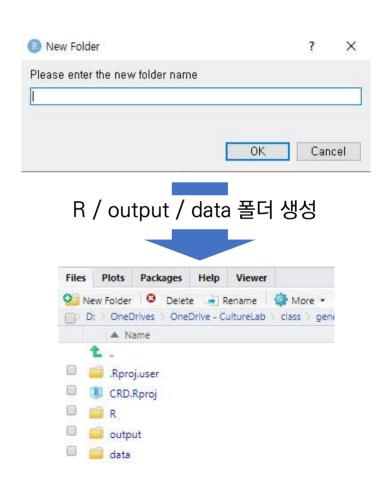
Project Type은 New Project



- ① 생성할 프로젝트가 저장될 폴더
- ② 프로젝트의 이름이자, 프로젝트 폴더 이름

작업 단위 만들기: RStudio 프로젝트

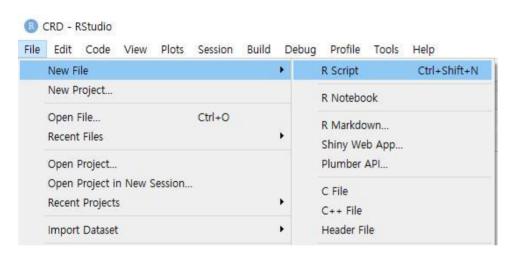


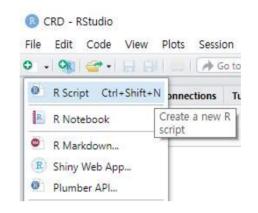


기초과정 [

R 에 오신 것을 환영합니다.

R 명령어 작성과 실행

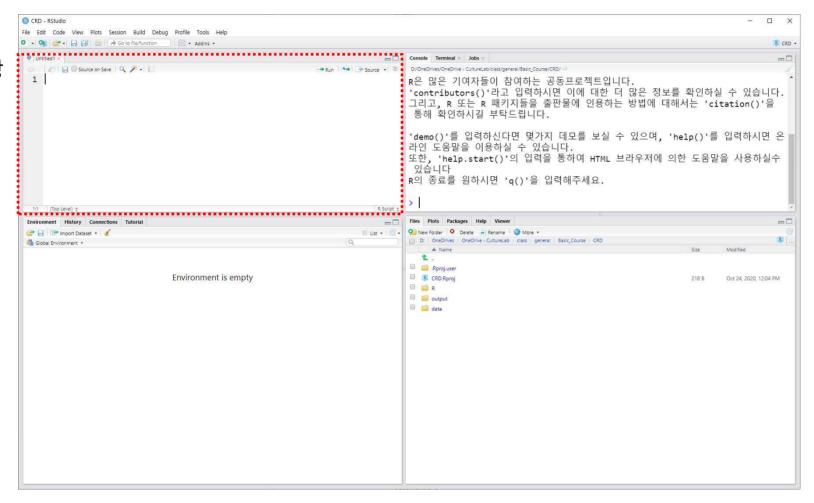




- R Script 는 File -> New File -> R Script 를 통해 생성
- 메뉴 아이콘 상의 🖸 를 클릭하고 나타나는 유형 중 R Script 생성
- 단축키는 Ctrl+Shift+N (Mac, Cmd+Shift+N)

R 명령어 작성과 실행

파일 편집창

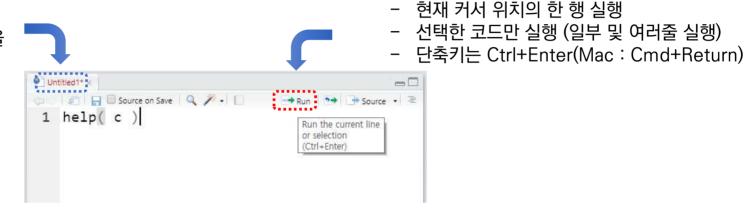


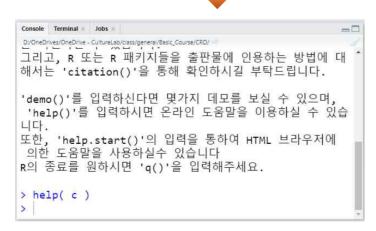
기초과정 [

R 에 오신 것을 환영합니다.

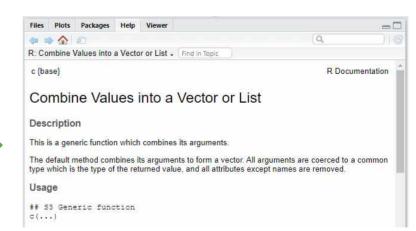
R 명령어 작성과 실행

- 현재 파일의 저장 여부
 - 첫 이름은 Untitledx
 - 다음의 두 경우는 저장되지 않음을 나타냄
 - 글자색은 붉은 색
 - 파일명 뒤에 * 표시





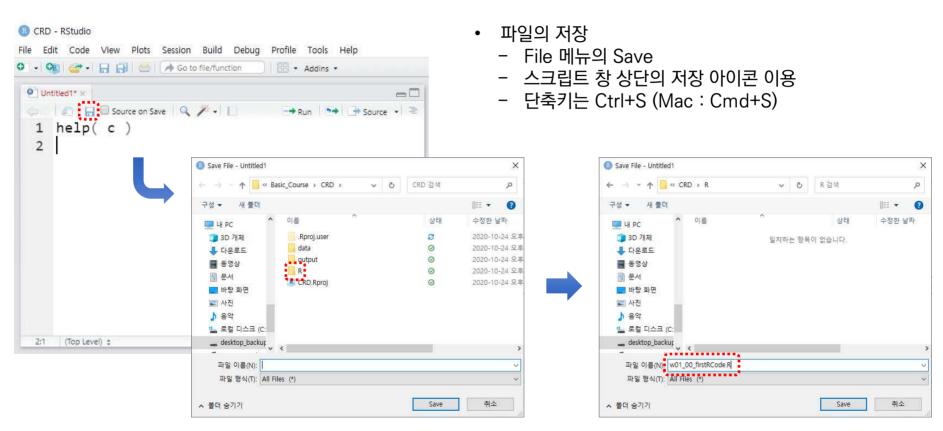




입력한 R 코드 실행

실행결과에 맞는 창에 출력

R 명령어 작성과 실행



R폴더는 R Script를 저장하기 위한 폴더

확장자는 반드시 .R 로 저장

Base R cheatsheet: https://rstudio.com/resources/cheatsheets/

R 알아가기

연산자 +, -, *, /, ^, 〈, 〈=, 〉, 〉=, ==, !=, 〈-

연산자는 R에서 계산을 수행하는 기호입니다. 연산 결과의 유형에 따라 산술연산자, 논리연산자가 있으며, 대입연산자(〈-)와 같이 특정한 기능을 수행하는 연산자가 있습니다.

변수 변수는 값을 저장하는 공간으로 이름과 값으로 구성되어 있습니다. 변수의 이름은 알파벳(과 특수문자 '_')으로 시작하며 할당연산자(〈-)에 의해 값을 변경할 수 있습니다. 변수 사용시에는 변수의 이름을 부릅니다.

함수 사용 function_name() / function_name(argument) / function_name(arg 1, arg 2, …, arg n)
R은 함수를 이용해 특별한 기능을 수행하며, 기능 수행을 위해 필요한 정보를 인수(argument)로 전달합니다.

도움말 ?mean / help(mean) / help(package="dplyr")

install.packages("pkg_name") library(pkg_name)

R은 패키지(package)를 통해 그 기능을 확장하고 있으며, R Community에서는 사용자들의 자발적인 참여로 다양한 패키지가 지금도 만들어 지고 있습니다.

실습 R코드: R/01_syntax.R

연산자

산술연산자 : 연산의 결과 숫자

1 + 3

1 - 3

2 * 3

2 / 3

2^3

논리연산자 : 연산의 결과가 논리값 (TRUE / FALSE)

1 < 3

1 <= 3

1 > 3

1 >= 3

1 == 3

1 != 3

실습 R코드: R/01_syntax.R

```
# 할당연산자 : 우측의 값을 좌측의 변수에 대입
a <- 1
b <- 3
c <- a + b
str1 <- "문자열은 큰 따옴표 혹은 작은 따옴표로 둘러쌉니다."
str2 <- '작은 따옴표'
str3 <- "큰 따옴표 안에 큰 따옴표를 쓰려면, ₩₩를 이용해서 escape 처리를 합니다. ₩"
```

a b c str1 str2 str3

a + 4

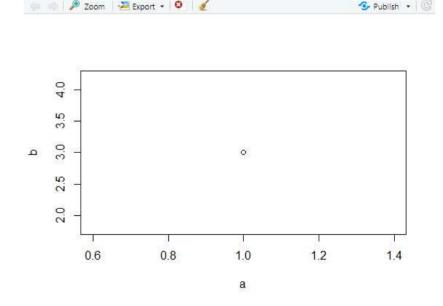
Environment	History	Connections	Tutorial			-0
	Import Data	set 🕶 🍮 121 1	MiB → 🎻			≣ List • @ •
R 🕶 🦺 Glob	oal Environm	ient •			Q,	
Values						
a		1				
b		3				
С		4				
str1		"문자일	별은 큰 따	옴표 혹은	작은 [다옴표로
str2		"작은	따옴표"			
str3		"큰 따	음표 안에	큰 따옴	표를 쓰리	려면, \
str4		"큰 따	옴표 안에	작은 따	음표는 그	그냥 사

함수 사용하기

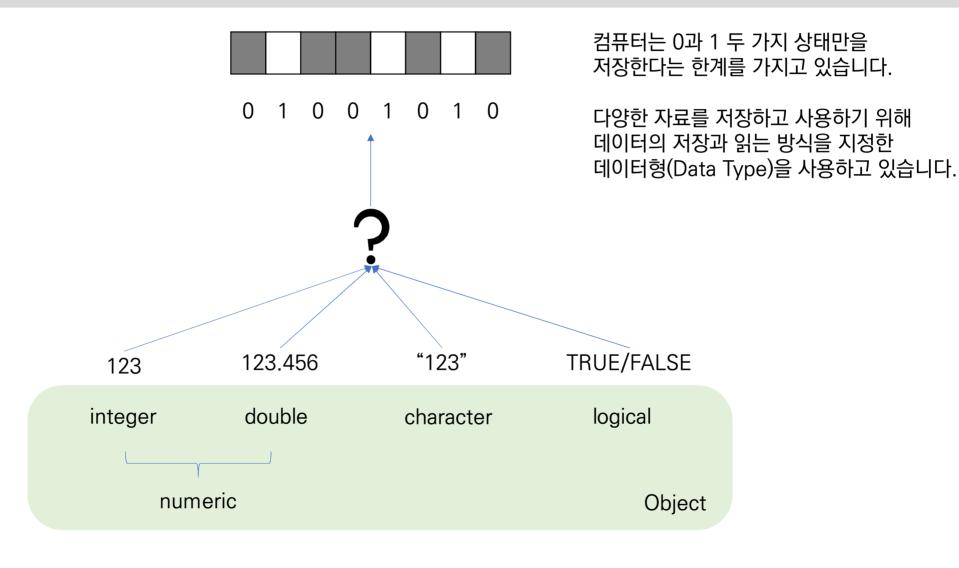
R 에 오신 것을 환영합니다.

실습 R코드: R/01_syntax.R

```
# 함수에는 이름이 있으며, 함수가 실행하기 위해 필요한 정보(인수)를 함수 이름과 함께 기술함
# 함수의 이름( 인수1, 인수2, ..., 인수 n)의 형태
# 함수의 사용은 이름을 부를 것으로 실행
# 함수 실행 후 함수의 원래 위치에 계산된 값을 전달하는 형태(값의 반환)
# 값을 전달하지 않고 어떤 행위(부수 효과)를 하는 경우도 많음
          # 인수가 필요없는 경우도 있음
getwd()
help("plot")
         # 인수가 여러개일 경우 콤마(,)로 구분
plot(a, b)
# 도움말 사용하기
?plot
help("plot")
```



컴퓨터의 가장 기초는 데이터형



실습 R코드: R/02_data_types.R

```
# 자료의 유형 알아보기
                                           is.logical( a < c )</pre>
typeof( a )
                                           is.character( a )
typeof( str1 )
                                           is.character( str1 )
# 자료의 유형을 판별하는 함수들 : is.xxx()
                                           # 자료형 변환 함수 as. 계열 함수
# a에는 1이 저장되어 있습니다.
                                           as.numeric( TRUE )
is.integer( a )
                                           as.numeric( FALSE )
is.double( a )
is.numeric( a )
                                           as.logical( 0 )
                                           as.logical( 1 )
# c에는 4가 저장되어 있습니다.
                                           as.logical( 9 )
is.integer( a:c )
is.double( a:c )
                                           as.character( a )
is.numeric( a:c )
                                           as.character( TRUE )
typeof( a:c )
```

기초과정 [

R 에 오신 것을 환영합니다.

데이터들의 결합 : 데이터 구조

ved	vector matrix							data	a fram	е		list						
[1]	1		[,1]			[8,]		sex	age	heig ht	[[1	[[1]]		[[2]]				
[2]	3		[1,]	1	5	-9	[1,]	М	38	168	[1]	1		sex	age	heig ht		
[3]	-4		[2,]	3	3	2	[2,]	F	42	159	[2]	3	[1,]	М	38	168		
[4]	4		[3,]	-4	1	-3	[3,]	F	29	162	[3]	-4	[2,]	F	42	159		
[5]	-2		[4,]	4	2	-1	[4,]	М	19	178	[4]	4	[3,]	F	29	162		
[6]	4		[5,]	-2	-4	-2	[5,]	F	32	161	[5]	-2	[4,]	М	19	178		
	→ factor		[1] [2]		[3]	[4]		[6]	4	[5,]	F	32	161				
				1	1		2		1									
			C	계성	여:	성	남성	0	성									

데이들의 결합 : 데이터 구조

	벡터 생성	c(Elem 1, Elem 2, ···, Elem n) / from:to / seq(from, to, by) / rep(vector, times each)
	벡터 관련	sort(vector) / rev(vector) / table(vector) / unique(vector)
자료	Mat 생성	matrix(vector, nrow ncol, byrow)
구조 관련	D.F 생성	data.frame(col 1, col 2, ···, col n) ———————————————————————————————————
함수	D.F 관련	nrow(df), ncol(df), dim(df), names(df), str(df)
	Factor생성	factor(vector, levels, labels, ordered=TRUE FALSE), cut(vector, breaks)
	관련함수	head(x), tail(x), str(x), class(x)

```
v1 \leftarrow c(1, 3, -2, 4)
v2 <- 3:7
v3 \leftarrow seq(-3, 3, by=0.1)
v4 <- rep( v1, times=2 )
v4_2 <- rep( v1, each=2 )
v5 <- c("a", "a", "b", "c")
v6 <- c( v1, v2 )
v7 < -c (v1, v5)
# 벡터는 동일한 자료형을 갖는 자료의 모임
```

```
# 관련 함수
sort( v1 )

rev( v1 )

table( v5 )

class( table(v5) )

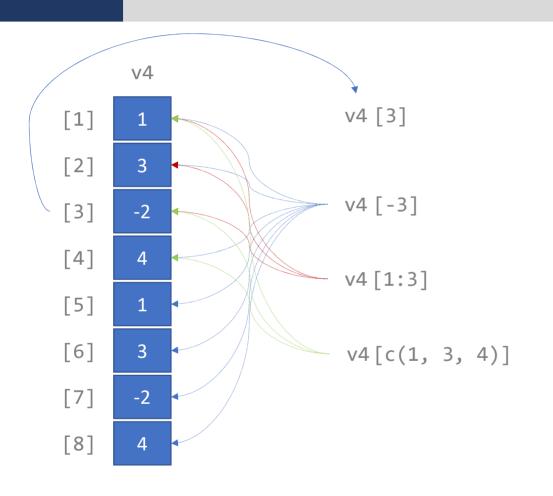
unique( v4 )

length( v1 )
```

```
# 벡터와 연산자
                                                  > length( v1 )
                                                  [1] 4
v1
                                                  > length( 3 )
                                                  [1] 1
v1 + 3
                                      v1
v4 + v1
                                                    rep(3, 4)
                                 [1]
v1 + (1:3)
                                [2]
                                       3
                                                                                  6
                                              + 3
                                 [3]
                                       -2
                                 [4]
```

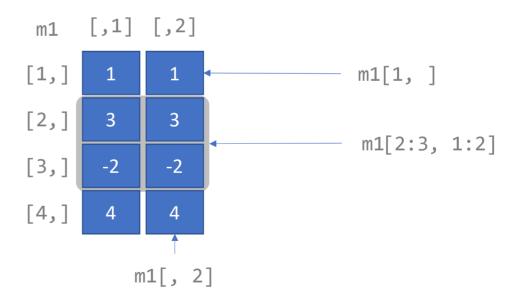
```
# 데이터 프레임
                                                   mat1 <- cbind( v1, v5 )
df1 <- data.frame( col1 = v1, col2 = v5 )</pre>
                                                   str( mat1 )
df1
                                                   mat2 <- rbind( v1, v5 )</pre>
dim( df1 )
                                                   str( mat2 )
nrow( df1 )
ncol( df1 )
                                                   mat3 <- cbind( df1, v1 )</pre>
names( df1 )
                                                   str( mat3 )
                                                   mat4 <- as.data.frame( mat1 )</pre>
# 관련 함수
                                                   str( mat4 )
class(mat1)
class(mat4)
str(mat1)
                                                   # list
str(mat4)
                                                   list1 \leftarrow list(v = v1, m = m1, df = df1)
head( mat1, n=2 )
                                                   list1
tail( mat1, n=2 )
```

데이터의 일부분 가져오기: 위치값으로 가져오기



변수명 [위치값 혹은 위치값 벡터]

변수명 [행 위치값(벡터), 열 위치값(벡터)]



실습 R코드: R/04_indexing_filtering.R

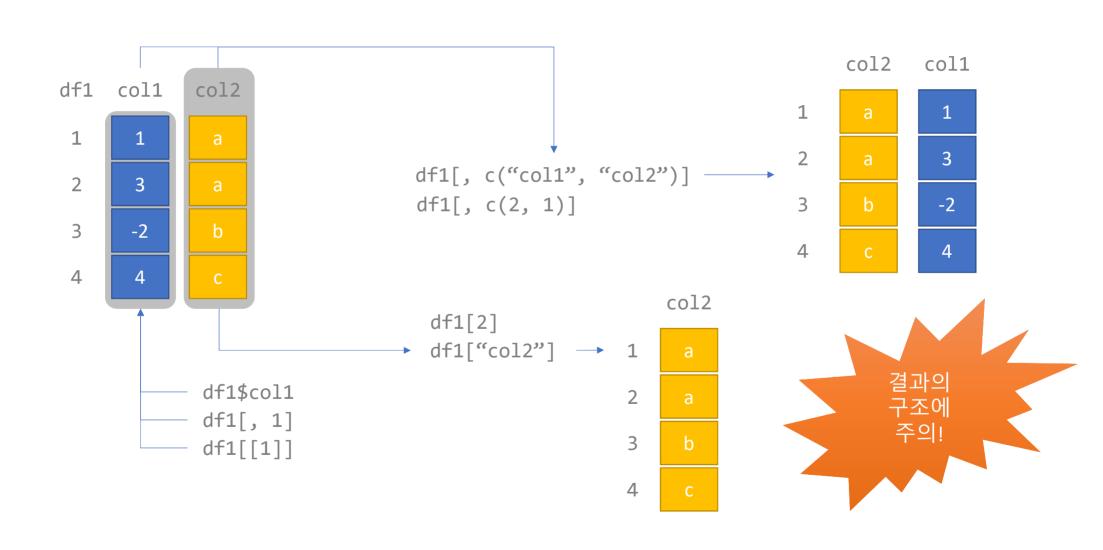
```
### 자료구조의 일부분 추출하기 #### Matrix (2차원의 경우 Data frame도 동일)
#### vector (factor도 동일)

v4

v4

m1[1, ]
m1[, 1]
m1[, 1]
v4[3]
v4[-3]
v4[1:3]
v4[c(1, 3, 4)]
```

데이터의 일부분 가져오기 : 위치값으로 가져오기



실습 R코드: R/04_indexing_filtering.R

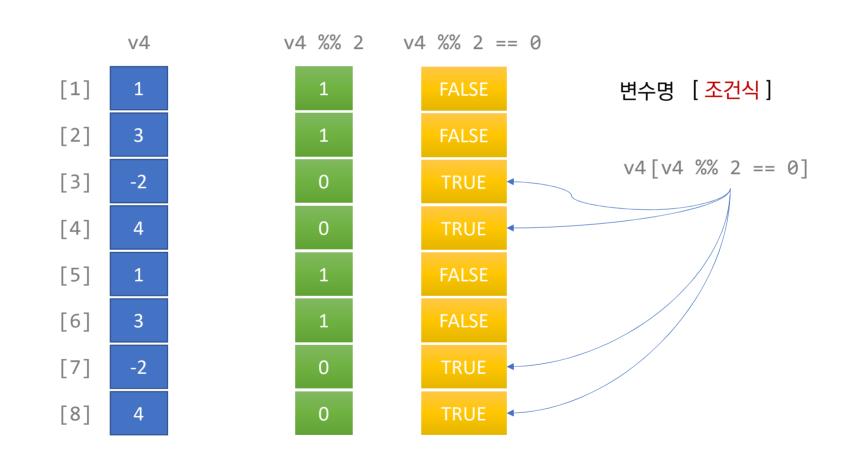
```
#### Data frame과 List : 하위 요소 이름으로 추출

df1$col1
df1[,1]
df1[[1]]

df1[, c(2, 1) ]
df1[, c("col2", "col1") ]

df1[2]
df1["col2"]
```

데이터의 일부분 가져오기 : 조건에 맞는 원소 추출

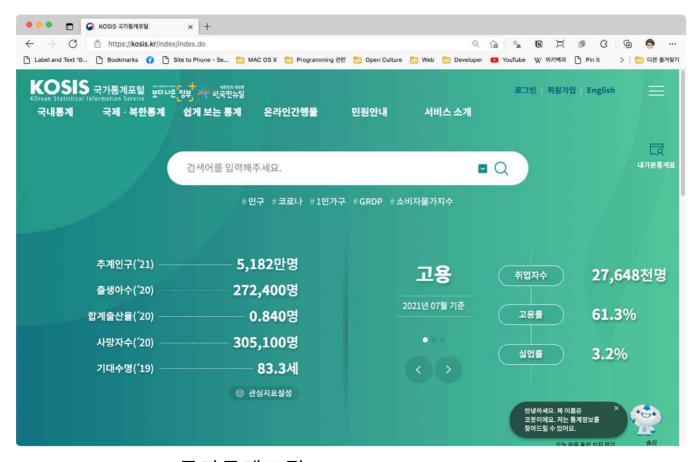


실습 R코드: R/04_indexing_filtering.R

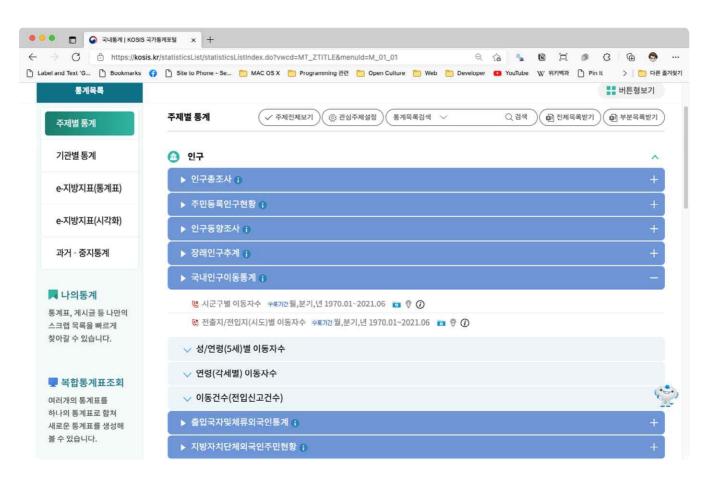
```
v4 %% 2 == 0
v4[v4 %% 2 == 0]

# 사실 따지고 보면, 조건을 만족하는 행의 위치를 구하는 것과 같습니다.
which( v4 %% 2 == 0 )
v4[which( v4 %% 2 == 0 )]
```

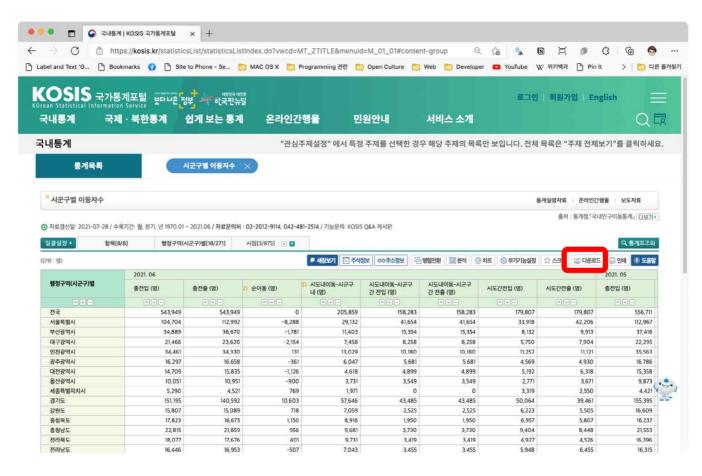
데이터 파일 다루기



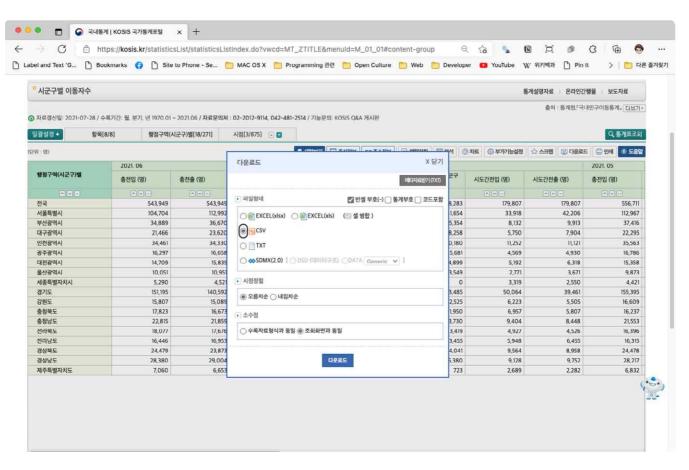
국가통계포털 : https://kosis.kr



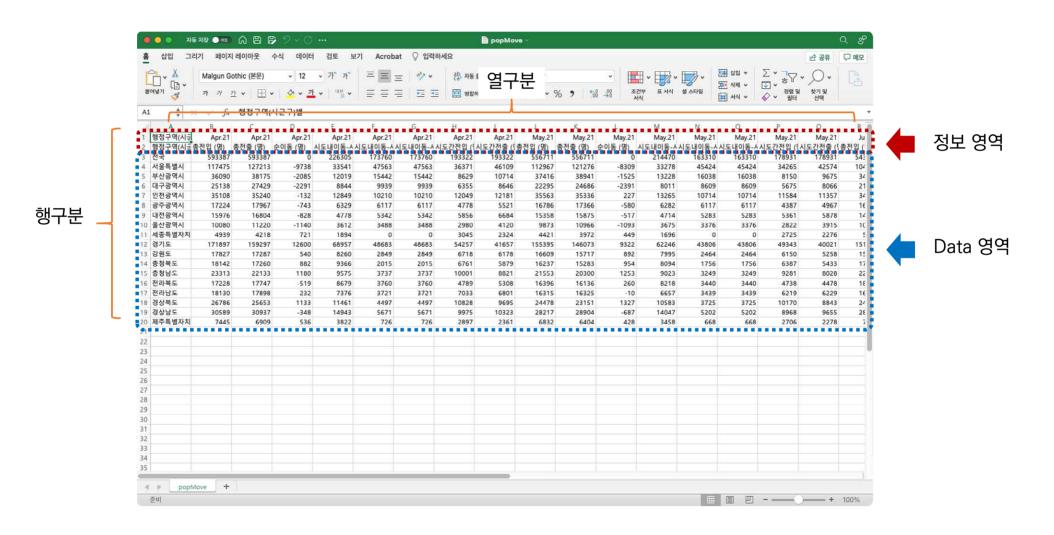
주제별 통계 - 인구 - 국내인구이동통계



다운로드 클릭



CSV 선택 후 다운로드 클릭



실습 R코드: R/05_readCSVFile.R

```
pop_move <- read.csv("./data/popMove.csv")
str(pop_move)

> str(pop_move)

'data.frame': 19 obs. of 25 variables:
$ 행정구역.시군구.별: chr "행정구역(시군구)별" "전국" "서울특별시" "부산광역시" ...
$ x2021..04 : chr "총전입 (명)" "593387" "117475" "36090" ...
$ x2021..04.1 : chr "총전출 (명)" "593387" "127213" "38175" ...
$ x2021..04.2 : chr "손이동 (명)" "0" "-9738" "-2085" ...
$ x2021..04.3 : chr "시도내이동-시군구내 (명)" "226305" "33541" "12019" ...
```

실습 R코드: R/05_readCSVFile.R

```
pop_move <- read.csv("./data/popMove.csv", skip=1, header=FALSE)
str(pop_move)

> str(pop_move)
'data.frame': 19 obs. of 25 variables:
$ V1 : chr "행정구역(시군구)별" "전국" "서울특별시" "부산광역시" ...
$ V2 : chr "총전입 (명)" "593387" "117475" "36090" ...
$ V3 : chr "총전출 (명)" "593387" "127213" "38175" ...
$ V4 : chr "순이동 (명)" "0" "-9738" "-2085" ...
$ V5 : chr "시도내이동-시군구내 (명)" "226305" "33541" "12019" ...
```

실습 R코드: R/05_readCSVFile.R

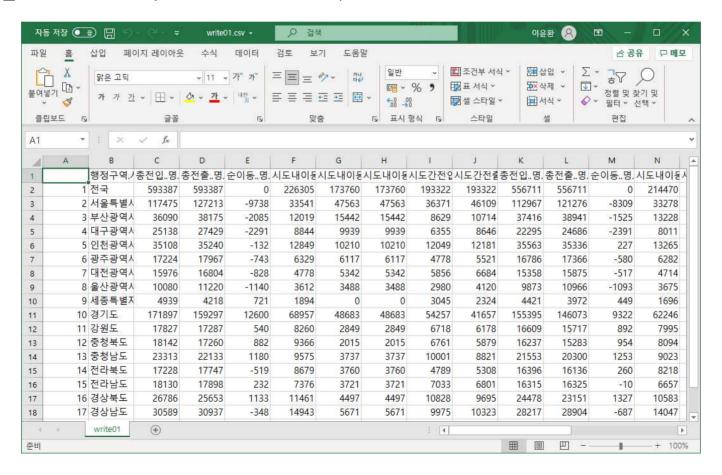
```
pop_move <- read.csv("./data/popMove.csv", skip=1)
str(pop_move)

> str(pop_move)

'data.frame': 18 obs. of 25 variables:
$ 행정구역.시군구.별 : chr "전국" "서울특별시" "부산광역시" "대구광역시" ...
$ 총전입..명. : int 593387 117475 36090 25138 35108 17224 15976 10080 4939
$ 총전출..명. : int 593387 127213 38175 27429 35240 17967 16804 11220 4218
$ 순이동..명. : int 0 -9738 -2085 -2291 -132 -743 -828 -1140 721 12600 ...
$ 시도내이동.시군구내..명. : int 226305 33541 12019 8844 12849 6329 4778 3612 1894 68957
```

실습 R코드: R/05_readCSVFile.R

write.csv(pop_move, "./output/write01.csv")



실습 R코드: R/05_readCSVFile.R

write.csv(pop_move, "./output/write02.csv", row.names = FALSE)

파	01 8	삽입 페이	이지 레이아웃	은 수식	데이터	검토 보	기 도움말	F					술공-	유 무메!
	_	00 AIV	시시 테시시크	- 77	riloiri	x	/I == 68	77 8	71		W			II. 2- 0 -
ľ	~ X	맑은 고딕		v 11 v	가 가	= = =	197 - 744 다신	일반	- 10	표조건부 서식		10000	* 37	Q
붙이	계봉기 [] ~	가 가 가	oz betegal	O - 71 -	내번		豆豆 圍	275	70 2	표 서식 ~	-00-00	박제 → ▼	전력 및 침	/ 한기 및
_	~ <3°	N 47 71	* III *	O - 21 -	Til Til		生 生 曽	.00 .0i	8 1	를 실 스타일 ·	# A	d식 ♥ ◆		선택ャ
1	클립보드 15		글꼴		12	맞	춤	[5] 표시 ³	형식 15	스타일		셀	편집	1
A1	· ·	1 ×	√ f _x	행정구역.	시군구.별									
M	A	В	С	D	E	F	G	н	Î.	j	К	L	M	N
1	_	총전입명.	총전출명.	순이동명.	시도내이동	시도내이동	시도내이들	시도간전인.	시도간전출	총전인명.	총전출명.	순이동명.	시도내이동	시도내이동
2	전국	593387	593387	0	226305	173760	173760	193322	193322	556711	556711	0	214470	163310
3	서울특별시	117475	127213	-9738	33541	47563	47563	36371	46109	112967	121276	-8309	33278	45424
4	부산광역시	36090	38175	-2085	12019	15442	15442	8629	10714	37416	38941	-1525	13228	16038
5	대구광역시	25138	27429	-2291	8844	9939	9939	6355	8646	22295	24686	-2391	8011	8609
6	인천광역시	35108	35240	-132	12849	10210	10210	12049	12181	35563	35336	227	13265	10714
7	광주광역시	17224	17967	-743	6329	6117	6117	4778	5521	16786	17366	-580	6282	6117
8	대전광역시	15976	16804	-828	4778	5342	5342	5856	6684	15358	15875	-517	4714	5283
9	울산광역시	10080	11220	-1140	3612	3488	3488	2980	4120	9873	10966	-1093	3675	3376
10	세종특별지	4939	4218	721	1894	0	0	3045	2324	4421	3972	449	1696	0
11	경기도	171897	159297	12600	68957	48683	48683	54257	41657	155395	146073	9322	62246	43806
12	강원도	17827	17287	540	8260	2849	2849	6718	6178	16609	15717	892	7995	2464
13	충청북도	18142	17260	882	9366	2015	2015	6761	5879	16237	15283	954	8094	1756
14	충청남도	23313	22133	1180	9575	3737	3737	10001	8821	21553	20300	1253	9023	3249
15	The state of the s	17228	17747	-519	8679	3760	3760	4789	5308	16396	16136	260	8218	3440
16	전라남도	18130	17898	232	7376	3721	3721	7033	6801	16315	16325	-10	6657	3439
17	00 ,	26786	25653	1133	11461	4497	4497	10828	9695	24478	23151	1327	10583	3725
18	경상남도	30589	30937	-348	14943	5671	5671	9975	10323	28217	28904	-687	14047	5202
1	4 F	write02	(+)			<u> </u>			1 1					

첫 모임을 마치려고 합니다.

많은 분들이 오늘 도대체 뭘 한거지? 하실수도 있지만. 데이터를 다루는 분야에서 어떻게 컴퓨터가 데이터를 관리하는지를 아는 것은 가장 기초라고 생각합니다.

다음 시간에는 오늘 배운 것을 바탕으로 데이터를 처리하고. 간단하게 도표를 작성하는 방법에 대해 알아보고자 합니다.

감사합니다.

ps : 궁금하신 것은 언제든지 yoonani72@gmail.com 으로 문의주세요.