


data.table

기본 문법

- i: 행 선택
- j: 열 선택 or 함수 적용
-  • by: 집단 나눔

fread / fwrite

속도가 매우 빠름 (R의 base 함수보다 40배 더 빠름)

```
1 fread("https://raw.githubusercontent.com/jinseob2kim/lecture-snuhlab/master
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	562083	200909	0	0	1
2:	2009	334536	200911	0	0	0
3:	2009	911867	200903	0	0	0
4:	2009	183321	200908	NA	NA	NA
5:	2009	942671	200909	NA	NA	NA

1640:	2015	266734	201504	NA	NA	NA
1641:	2015	343874	201511	0	0	0
1642:	2015	798472	201507	NA	NA	NA
1643:	2015	720080	201505	NA	NA	NA
1644:	2015	127519	201511	0	0	1
	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	0	0	NA	2	1	0
3:	0	0	NA	2	1	0

Exam data

data.frame

```
1 df <- read.csv("https://raw.githubusercontent.com/jinseob2kim/lecture-snuhlab/master/data/example_gle.csv")
2 head(df)
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN								
1	2009	562083	200909	0	0	1								
2	2009	334536	200911	0	0	0								
3	2009	911867	200903	0	0	0								
4	2009	183321	200908	NA	NA	NA								
5	2009	942671	200909	NA	NA	NA								
6	2009	979358	200912	NA	NA	NA								
	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N	HGHT							
1	0	0	NA	3	1	0	144							
2	0	0	NA	2	1	0	162							
3	0	0	NA	3	1	0	163							
4	NA	NA	NA	3	1	0	152							
5	NA	NA	NA	3	1	0	159							
6	NA	NA	NA	2	1	0	157							
	WGHT	WSTC	BMI	VA_LT	VA_RT	BP_SYS	BP_DIA	URN_PROT	HGB	FBS	TOT_CHOL	TG	HDL	
1	61	90	29.4	0.7	0.8	120	80	1	12.6	117	264	128	60	
2	51	63	19.4	0.8	1.0	120	80	1	13.8	96	169	92	70	
3	65	82	24.5	0.7	0.6	130	80	1	15.0	118	216	132	55	
4	51	70	22.1	0.8	0.9	101	62	1	13.1	90	199	100	65	
5	50	73	19.8	0.7	0.8	132	78	1	13.0	92	162	58	40	
6	55	73	22.3	1.5	1.5	110	70	1	11.9	100	192	109	53	
	LDL	CRTN	SGOT	SGPT	GGT	GFR								

```
1 class(df)
```

```
[1] "data.frame"
```

Exam data

data.table

```
1 dt <- fread("https://raw.githubusercontent.com/jinseob2kim/lecture-snuhlab/master/data/example_gle.csv")
2 dt
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	562083	200909	0	0	1
2:	2009	334536	200911	0	0	0
3:	2009	911867	200903	0	0	0
4:	2009	183321	200908	NA	NA	NA
5:	2009	942671	200909	NA	NA	NA

1640:	2015	266734	201504	NA	NA	NA
1641:	2015	343874	201511	0	0	0
1642:	2015	798472	201507	NA	NA	NA
1643:	2015	720080	201505	NA	NA	NA
1644:	2015	127519	201511	0	0	1
	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PT	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	0	0	NA	2	1	0
3:	0	0	NA	3	1	0
4:	NA	NA	NA	3	1	0
5:	NA	NA	NA	3	1	0

1640:	NA	NA	NA	1	2	2
1641:	1	0	0	2	2	0
1642:	NA	NA	NA	3	2	0

```
1 class(dt)
```

```
[1] "data.table" "data.frame"
```

Row Operation

첫 번째 열부터 다섯 번째 열까지 선택

```
1 dt[1:5]
2 dt[1:5, ]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN							
1:	2009	562083	200909	0	0	1							
2:	2009	334536	200911	0	0	0							
3:	2009	911867	200903	0	0	0							
4:	2009	183321	200908	NA	NA	NA							
5:	2009	942671	200909	NA	NA	NA							
	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N	HGHT						
1:	0	0	NA	3	1	0	144						
2:	0	0	NA	2	1	0	162						
3:	0	0	NA	3	1	0	163						
4:	NA	NA	NA	3	1	0	152						
5:	NA	NA	NA	3	1	0	159						
	WGHT	WSTC	BMI	VA_LT	VA_RT	BP_SYS	BP_DIA	URN_PROT	HGB	FBS	TOT_CHOL	TG	HDL
1:	61	90	29.4	0.7	0.8	120	80	1	12.6	117	264	128	60
2:	51	63	19.4	0.8	1.0	120	80	1	13.8	96	169	92	70
3:	65	82	24.5	0.7	0.6	130	80	1	15.0	118	216	132	55
4:	51	70	22.1	0.8	0.9	101	62	1	13.1	90	199	100	65
5:	50	73	19.8	0.7	0.8	132	78	1	13.0	92	162	58	40
	LDL	CRTN	SGOT	SGPT	GGT	GFR							
1:	179	0.9	25	20	25	59							
2:	80	0.9	18	15	28	74							
3:	134	0.8	26	30	30	79							

Row Operation

EXMD_BZ_YYYY가 2009년-2012년 사이이고 BMI가 25 이상인 행만 추출

```
1 dt[(EXMD_BZ_YYYY %in% 2009:2012) & (BMI >= 25)]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	562083	200909	0	0	1
2:	2009	318669	200904	NA	NA	NA
3:	2009	668438	200904	NA	NA	NA
4:	2009	560878	200903	NA	NA	NA
5:	2009	375694	200906	0	0	1

317:	2012	582044	201210	0	0	0
318:	2012	126561	201210	0	0	0
319:	2012	266734	201204	NA	NA	NA
320:	2012	942204	201209	0	0	0
321:	2012	637160	201209	0	0	0

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	NA	NA	NA	2	1	0
3:	NA	NA	NA	3	1	0
4:	NA	NA	NA	2	1	0
5:	1	1	NA	2	1	0

317:	0	0	0	3	3	2
318:	0	0	0	2	3	1
319:	NA	NA	NA	1	3	1

Row Operation

HME_YYYYMM에 따라 오름차순으로 정렬

```
1 dt[order(HME_YYYYMM)]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	480569	200901	0	1	1
2:	2009	79250	200901	0	0	0
3:	2009	418137	200901	0	0	0
4:	2009	669498	200902	0	0	0
5:	2009	486441	200902	0	0	1

1640:	2015	284267	201512	NA	NA	NA
1641:	2015	606125	201512	NA	NA	NA
1642:	2015	356551	201512	0	0	1
1643:	2015	4263	201512	0	0	0
1644:	2015	887943	201512	NA	NA	NA

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PT	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	0	0	NA	1	1	0
3:	0	0	NA	NA	2	NA
4:	0	0	NA	2	2	1
5:	0	0	NA	1	3	4

1640:	NA	NA	NA	3	2	1
1641:	1	NA	NA	2	2	2
1642:	0	0	0	3	2	0

Row Operation

HME_YYYYMM은 오름차순으로, HGHT는 내림차순으로 정렬

```
1 dt[order(HME_YYYYMM, -HGHT)]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	418137	200901	0	0	0
2:	2009	480569	200901	0	1	1
3:	2009	79250	200901	0	0	0
4:	2009	486441	200902	0	0	1
5:	2009	26776	200902	NA	NA	NA

1640:	2015	979090	201512	0	0	0
1641:	2015	181909	201512	0	0	1
1642:	2015	875758	201512	NA	NA	NA
1643:	2015	992522	201512	0	0	0
1644:	2015	975124	201512	0	0	1

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PT	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	NA	2	NA
2:	0	0	NA	3	1	0
3:	0	0	NA	1	1	0
4:	0	0	NA	1	3	4
5:	NA	NA	NA	2	2	2

1640:	0	0	0	2	1	0
1641:	0	0	0	2	1	0
1642:	NA	NA	NA	2	1	1

Row Operation

EXMD_BZ_YYYY가 2009년-2012년 사이이고 BMI가 25 이상인 행만 추출한 후
HGHT에 따라 오름차순으로 정렬

```
1 dt[(EXMD_BZ_YYYY %in% 2009:2012) & (BMI >= 25)][order(HGHT)]
2 dt[(EXMD_BZ_YYYY %in% 2009:2012) & (BMI >= 25)] %>% .[order(HGHT)] # same
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	562083	200909	0	0	1
2:	2009	560878	200903	NA	NA	NA
3:	2011	562083	201111	0	0	1
4:	2011	519824	201109	NA	NA	NA
5:	2011	914987	201103	0	0	1

317:	2011	550255	201104	0	0	0
318:	2009	115809	200905	0	0	0
319:	2010	115809	201006	0	0	0
320:	2011	115809	201109	0	0	0
321:	2012	115809	201209	0	0	0

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	NA	NA	NA	2	1	0
3:	0	0	0	2	1	0
4:	NA	NA	NA	2	1	0
5:	0	0	0	3	1	2

317:	0	0	0	2	3	1
318:	0	0	NA	3	3	2
319:	0	0	0	3	3	1

Column Operation

첫 번째 열부터 다섯 번째 열까지 추출

```
1 dt[, 1:5]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ
1:	2009	562083	200909	0	0
2:	2009	334536	200911	0	0
3:	2009	911867	200903	0	0
4:	2009	183321	200908	NA	NA
5:	2009	942671	200909	NA	NA

1640:	2015	266734	201504	NA	NA
1641:	2015	343874	201511	0	0
1642:	2015	798472	201507	NA	NA
1643:	2015	720080	201505	NA	NA
1644:	2015	127519	201511	0	0

Column Operation

HGHT 열과 WGHT 열만 추출

```
1 dt[, c("HGHT", "WGHT")]
2 dt[, .(HGHT, WGHT)]
```

	HGHT	WGHT
1:	144	61
2:	162	51
3:	163	65
4:	152	51
5:	159	50

1640:	171	76
1641:	179	80
1642:	163	63
1643:	165	70
1644:	168	84

Column Operation

HGHT 열을 Height로, WGHT 열을 Weight로 이름을 바꿔서 추출

```
1 dt[, .(Height = HGHT, Weight = WGHT)] # rename
```

	Height	Weight
1:	144	61
2:	162	51
3:	163	65
4:	152	51
5:	159	50

1640:	171	76
1641:	179	80
1642:	163	63
1643:	165	70
1644:	168	84

Column Operation

data.table 형식으로 추출

```
1 dt[, .(HGHT)]  
2 dt[, "HGHT"]
```

```
      HGHT  
1:    144  
2:    162  
3:    163  
4:    152  
5:    159  
---  
1640:   171  
1641:   179  
1642:   163  
1643:   165  
1644:   168
```

Column Operation

vector 형식으로 추출

```
1 dt[, HGHT] # vector
```

```
[1] 144 162 163 152 159 157 160 159 156 146 164 154 155 155 165 180 149 160
[19] 162 154 144 155 151 158 165 160 154 177 156 145 156 165 172 148 148 158
[37] 158 162 168 164 158 169 160 157 173 163 151 156 154 141 160 163 170 149
[55] 144 145 163 154 160 169 170 151 152 175 160 157 139 163 148 166 165 150
[73] 157 150 161 158 171 166 165 166 165 166 169 156 170 171 179 159 147 159
[91] 173 162 150 157 159 175 161 172 159 167 176 170 167 166 165 178 169 172
[109] 161 164 166 150 158 165 159 155 170 154 164 170 160 173 159 150 151 164
[127] 162 178 176 165 169 163 181 180 170 180 186 167 175 169 176 172 167 180
[145] 160 171 170 159 182 181 172 160 172 165 173 165 180 184 163 165 170 162
[163] 160 176 168 170 176 171 161 173 171 172 174 173 165 173 170 150 166 169
[181] 162 163 167 172 174 172 164 166 166 166 174 176 161 164 159 165 161 167
[199] 179 164 172 176 164 174 159 174 169 167 170 176 163 159 175 165 172 160
[217] 170 156 157 151 160 157 164 162 156 166 166 162 147 163 160 163 168 155
[235] 172 159 163 151 157 153 154 164 160 176 166 173 164 164 171 166 140 139
[253] 156 178 150 170 155 166 163 159 168 157 166 151 160 145 150 168 160 158
[271] 166 147 164 158 165 151 155 156 159 155 161 159 180 152 167 172 148 161
[289] 171 156 160 165 166 163 154 164 172 161 149 160 164 178 164 164 160 170
[307] 172 150 159 167 159 158 160 165 158 159 164 160 170 153 169 153 162 151
[325] 164 170 165 157 150 175 157 160 162 159 167 160 152 169 177 160 155 169
[343] 157 167 162 155 172 172 165 171 177 170 160 173 164 171 182 159 180 176
[361] 174 168 165 173 172 164 167 161 173 167 179 176 167 171 163 163 168 175
[379] 176 186 162 183 184 181 174 161 175 164 159 171 177 166 151 172 180 169
```

Column Operation

변수로 열 이름 선택

```
1 colvars <- grep("Q_", names(dt), value = T)
2 colvars
```

```
[1] "Q_PHX_DX_STK"    "Q_PHX_DX_HTDZ"    "Q_PHX_DX_HTN"    "Q_PHX_DX_DM"
[5] "Q_PHX_DX_DLD"    "Q_PHX_DX_PTB"     "Q_HBV_AG"        "Q_SMK_YN"
[9] "Q_DRK_FRQ_V09N"
```

```
1 dt[, ..colvars]
2 dt[, colvars, with = F]
```

	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN	Q_PHX_DX_DM	Q_PHX_DX_DLD
1:	0	0	1	0	0
2:	0	0	0	0	0
3:	0	0	0	0	0
4:	NA	NA	NA	NA	NA
5:	NA	NA	NA	NA	NA

1640:	NA	NA	NA	NA	NA
1641:	0	0	0	1	0
1642:	NA	NA	NA	NA	NA
1643:	NA	NA	NA	NA	1
1644:	0	0	1	0	0

	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	NA	3	1	0
2:	NA	2	1	0
3:	NA	3	1	0
4:	NA	3	1	0
5:	NA	3	1	0

1640:	NA	1	2	2
1641:	0	2	2	0
1642:	NA	3	2	0

Column Operation

.SD: Subset of Data

```
1 dt[, .SD, .SDcols = colvars]
```

	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN	Q_PHX_DX_DM	Q_PHX_DX_DLD
1:	0	0	1	0	0
2:	0	0	0	0	0
3:	0	0	0	0	0
4:	NA	NA	NA	NA	NA
5:	NA	NA	NA	NA	NA

1640:	NA	NA	NA	NA	NA
1641:	0	0	0	1	0
1642:	NA	NA	NA	NA	NA
1643:	NA	NA	NA	NA	1
1644:	0	0	1	0	0
	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N	
1:	NA	3	1	0	
2:	NA	2	1	0	
3:	NA	3	1	0	
4:	NA	3	1	0	
5:	NA	3	1	0	

1640:	NA	1	2	2	
1641:	0	2	2	0	
1642:	NA	3	2	0	

Column Operation

EXMD_BZ_YYYY가 2009년-2012년 사이이고 BMI가 25 이상인 행 중 colvars 열만 추출

```
1 dt[(EXMD_BZ_YYYY %in% 2009:2012) & (BMI >= 25), ..colvars]
```

	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN	Q_PHX_DX_DM	Q_PHX_DX_DLD
1:	0	0	1	0	0
2:	NA	NA	NA	NA	NA
3:	NA	NA	NA	NA	NA
4:	NA	NA	NA	NA	NA
5:	0	0	1	1	1

317:	0	0	0	0	0
318:	0	0	0	0	0
319:	NA	NA	NA	NA	NA
320:	0	0	0	0	0
321:	0	0	0	0	0

	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	NA	3	1	0
2:	NA	2	1	0
3:	NA	3	1	0
4:	NA	2	1	0
5:	NA	2	1	0

317:	0	3	3	2
318:	0	2	3	1
319:	NA	1	3	1

Column Operation

colvars 열 제외

```
1 dt[, !...colvars]
2 dt[, -...colvars]
3 dt[, .SD, .SDcols = -colvars]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	HGHT	WGHT	WSTC	BMI	VA_LT	VA_RT	BP_SYS			
1:	2009	562083	200909	144	61	90	29.4	0.7	0.8	120			
2:	2009	334536	200911	162	51	63	19.4	0.8	1.0	120			
3:	2009	911867	200903	163	65	82	24.5	0.7	0.6	130			
4:	2009	183321	200908	152	51	70	22.1	0.8	0.9	101			
5:	2009	942671	200909	159	50	73	19.8	0.7	0.8	132			

1640:	2015	266734	201504	171	76	88	26.0	1.2	1.2	125			
1641:	2015	343874	201511	179	80	88	25.0	1.0	0.7	126			
1642:	2015	798472	201507	163	63	80	23.7	1.5	1.5	118			
1643:	2015	720080	201505	165	70	92	25.7	0.6	0.8	120			
1644:	2015	127519	201511	168	84	103	29.8	0.9	0.8	140			
	BP_DIA	URN_PROT	HGB	FBS	TOT_CHOL	TG	HDL	LDL	CRTN	SGOT	SGPT	GGT	GFR
1:	80	1	12.6	117	264	128	60	179	0.9	25	20	25	59
2:	80	1	13.8	96	169	92	70	80	0.9	18	15	28	74
3:	80	1	15.0	118	216	132	55	134	0.8	26	30	30	79
4:	62	1	13.1	90	199	100	65	114	0.9	18	14	11	61
5:	78	1	13.0	92	162	58	40	111	0.9	24	23	15	49

1640:	89	1	14.1	77	363	280	63	244	0.9	19	32	95	96
1641:	83	1	14.9	119	106	247	29	50	0.9	101	157	62	79
1642:	77	1	16.1	160	271	514	51	151	1.0	40	66	104	85

Column Summary

HGHT의 평균, WGHT의 평균, BMI의 평균

```
1 dt[, .(mean(HGHT), mean(WGHT), mean(BMI))]
```

	V1	V2	V3
1:	164.5487	65.09672	23.92257

Column Summary

HGHT의 평균, WGHT의 평균, BMI의 평균을 이름 지정해서 출력

```
1 dt[, .(HGHT = mean(HGHT), WGHT = mean(WGHT), BMI = mean(BMI))]
```

	HGHT	WGHT	BMI
1:	164.5487	65.09672	23.92257

Column Summary

lapply로 HGHT, WGHT, BMI에 **한번에** mean 함수 적용

```
1 dt[, lapply(.SD, mean), .SDcols = c("HGHT", "WGHT", "BMI")]
```

	HGHT	WGHT	BMI
1:	164.5487	65.09672	23.92257

By Operation

EXMD_BZ_YYYY으로 그룹화하여 HGHT의 평균, WGHT의 평균, BMI의 평균 구하기

```
1 dt[, .(HGHT = mean(HGHT), WGHT = mean(WGHT), BMI = mean(BMI)), by = EXMD_BZ_YYYY]
2 dt[, .(HGHT = mean(HGHT), WGHT = mean(WGHT), BMI = mean(BMI)), by = "EXMD_BZ_YYYY"]
3 dt[, lapply(.SD, mean), .SDcols = c("HGHT", "WGHT", "BMI"), by = EXMD_BZ_YYYY]
```

	EXMD_BZ_YYYY	HGHT	WGHT	BMI
1:	2009	164.0841	64.32710	23.76402
2:	2010	164.9280	65.13983	23.82712
3:	2011	164.1480	64.89686	23.96009
4:	2012	164.9188	65.82051	24.08932
5:	2013	164.9095	64.90535	23.75391
6:	2014	164.3189	64.47244	23.78031
7:	2015	164.4792	66.07500	24.28167

By Operation

HGHT가 175 이상인 데이터를 EXMD_BZ_YYYY, Q_SMK_YN으로 **그룹화**하여 개수 구함

.N: length()

```
1 dt[HGHT >= 175, .N, by = .(EXMD_BZ_YYYY, Q_SMK_YN)]  
2 dt[HGHT >= 175, .N, by = c("EXMD_BZ_YYYY", "Q_SMK_YN")]
```

	EXMD_BZ_YYYY	Q_SMK_YN	N
1:	2009	1	7
2:	2009	NA	1
3:	2009	3	14
4:	2009	2	5
5:	2010	1	6
6:	2010	3	19
7:	2010	2	7
8:	2011	1	7
9:	2011	2	8
10:	2011	3	16
11:	2012	1	12
12:	2012	2	6
13:	2012	3	21
14:	2013	1	10
15:	2013	3	16
16:	2013	2	8
17:	2014	1	10
18:	2014	2	7
19:	2014	3	26
20:	2015	1	12
21:	2015	3	18

By Operation

keyby를 통해 정렬할 수 있음

```
1 dt[HGHT >= 175, .N, keyby = c("EXMD_BZ_YYYY", "Q_SMK_YN")]
```

	EXMD_BZ_YYYY	Q_SMK_YN	N
1:	2009	NA	1
2:	2009	1	7
3:	2009	2	5
4:	2009	3	14
5:	2010	1	6
6:	2010	2	7
7:	2010	3	19
8:	2011	1	7
9:	2011	2	8
10:	2011	3	16
11:	2012	1	12
12:	2012	2	6
13:	2012	3	21
14:	2013	1	10
15:	2013	2	8
16:	2013	3	16
17:	2014	1	10
18:	2014	2	7
19:	2014	3	26
20:	2015	1	12
21:	2015	2	6

By Operation

조건으로 그룹화할 수도 있음

```
1 dt[HGHT >= 175, .N, keyby= .(EXMD_BZ_YYYY >= 2015, Q_PHX_DX_STK == 1)]
```

	EXMD_BZ_YYYY	Q_PHX_DX_STK	N
1:	FALSE	NA	50
2:	FALSE	FALSE	156
3:	TRUE	NA	6
4:	TRUE	FALSE	30

By Operation

조건으로 그룹화할 수도 있음

```
1 dt[HGHT >= 175, .N, keyby= .(get("EXMD_BZ_YYYY") >= 2015, get("Q_PHX_DX_STK") == 1)]
```

```
      get get.1    N
1: FALSE    NA   50
2: FALSE FALSE 156
3:  TRUE    NA    6
4:  TRUE FALSE  30
```

By Operation

조건으로 그룹화할 수도 있음

```
1 dt[HGHT >= 175, .N, keyby= .(Y2015 = ifelse(EXMD_BZ_YYYY >= 2015, ">=2015", "<2015"))]
```

```
      Y2015    N  
1:    <2015  206  
2:    >=2015   36
```

Merge

예시 데이터

```
1 dt1 <- dt[1:10, .SD, .SDcols = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM", colvars)]
2 dt1
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	562083	200909	0	0	1
2:	2009	334536	200911	0	0	0
3:	2009	911867	200903	0	0	0
4:	2009	183321	200908	NA	NA	NA
5:	2009	942671	200909	NA	NA	NA
6:	2009	979358	200912	NA	NA	NA
7:	2009	554112	200911	NA	NA	NA
8:	2009	487160	200908	NA	NA	NA
9:	2009	793017	200906	NA	NA	NA
10:	2009	219397	200912	0	0	1

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	0	0	NA	2	1	0
3:	0	0	NA	3	1	0
4:	NA	NA	NA	3	1	0
5:	NA	NA	NA	3	1	0
6:	NA	NA	NA	2	1	0
7:	NA	NA	NA	2	1	0
8:	NA	NA	NA	3	1	0
9:	NA	NA	NA	3	1	0
10:	0	0	NA	3	1	0

```
1 dt2 <- dt[6:15, -..colvars]
2 dt2
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	HGHT	WGHT	WSTC	BMI	VA_LT	VA_RT	BP_SYS
1:	2009	979358	200912	157	55	73	22.3	1.5	1.5	110
2:	2009	554112	200911	160	56	67	21.9	1.5	1.5	119
3:	2009	487160	200908	159	54	66	21.4	1.2	1.5	111

4:		2009	793017		200906	156	53	67	21.8	1.2	1.0	138
5:		2009	219397		200912	146	48	78	22.5	1.5	1.5	138
6:		2009	831349		200912	164	66	85	24.5	1.2	1.0	130
7:		2009	480569		200901	154	58	84	24.5	0.1	0.8	140
8:		2009	559370		200912	155	51	75	21.2	1.0	1.0	93
9:		2009	318669		200904	155	66	78	27.5	0.8	0.6	95
10:		2009	395781		200907	165	60	71	22.0	0.6	0.4	120

	BP_DIA	URN_PROT	HGB	FBS	TOT_CHOL	TG	HDL	LDL	CRTN	SGOT	SGPT	GGT	GFR
1:	70	1	11.9	100	192	109	53	117	0.7	15	12	14	83
2:	78	1	11.2	84	152	38	43	101	0.8	8	6	10	97
3:	60	1	12.2	88	166	42	58	99	1.0	16	11	12	65
4:	72	1	11.0	74	155	86	52	85	0.6	15	13	13	96
5:	84	1	12.8	107	178	87	35	125	0.7	21	21	23	70
6:	90	1	16.3	108	209	57	59	138	0.8	32	38	16	98
7:	70	1	10.4	84	242	134	53	162	1.1	15	10	10	37
8:	53	1	11.1	92	212	44	69	134	1.0	20	10	10	59
9:	58	1	13.6	101	294	119	50	220	0.6	22	22	22	116
10:	70	1	15.3	85	237	112	56	158	1.0	24	20	18	70

Merge



Merge

Full join

```
1 merge(dt1, dt2, by = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM"), all = T)
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN		
1:	2009	183321	200908	NA	NA	NA		
2:	2009	219397	200912	0	0	1		
3:	2009	318669	200904	NA	NA	NA		
4:	2009	334536	200911	0	0	0		
5:	2009	395781	200907	NA	NA	NA		
6:	2009	480569	200901	NA	NA	NA		
7:	2009	487160	200908	NA	NA	NA		
8:	2009	554112	200911	NA	NA	NA		
9:	2009	559370	200912	NA	NA	NA		
10:	2009	562083	200909	0	0	1		
11:	2009	793017	200906	NA	NA	NA		
12:	2009	831349	200912	NA	NA	NA		
13:	2009	911867	200903	0	0	0		
14:	2009	942671	200909	NA	NA	NA		
15:	2009	979358	200912	NA	NA	NA		

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PT	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N	HGHT
1:	NA	NA	NA	3	1	0	NA
2:	0	0	NA	3	1	0	146
3:	NA	NA	NA	NA	NA	NA	155
4:	0	0	NA	2	1	0	NA
5:	NA	NA	NA	NA	NA	NA	165

Merge

Inner join

```
1 merge(dt1, dt2, by = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM"), all = F)
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	219397	200912	0	0	1
2:	2009	487160	200908	NA	NA	NA
3:	2009	554112	200911	NA	NA	NA
4:	2009	793017	200906	NA	NA	NA
5:	2009	979358	200912	NA	NA	NA

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N	HGHT
1:	0	0	NA	3	1	0	146
2:	NA	NA	NA	3	1	0	159
3:	NA	NA	NA	2	1	0	160
4:	NA	NA	NA	3	1	0	156
5:	NA	NA	NA	2	1	0	157

	WGHT	WSTC	BMI	VA_LT	VA_RT	BP_SYS	BP_DIA	URN_PROT	HGB	FBS	TOT_CHOL	TG	HDL
1:	48	78	22.5	1.5	1.5	138	84	1	12.8	107	178	87	35
2:	54	66	21.4	1.2	1.5	111	60	1	12.2	88	166	42	58
3:	56	67	21.9	1.5	1.5	119	78	1	11.2	84	152	38	43
4:	53	67	21.8	1.2	1.0	138	72	1	11.0	74	155	86	52
5:	55	73	22.3	1.5	1.5	110	70	1	11.9	100	192	109	53

	LDL	CRTN	SGOT	SGPT	GGT	GFR
1:	125	0.7	21	21	23	70
2:	99	1.0	16	11	12	65
3:	101	0.8	8	6	10	97

Merge

Left join

```
1 merge(dt1, dt2, by = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM"), all.x = T)
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	183321	200908	NA	NA	NA
2:	2009	219397	200912	0	0	1
3:	2009	334536	200911	0	0	0
4:	2009	487160	200908	NA	NA	NA
5:	2009	554112	200911	NA	NA	NA
6:	2009	562083	200909	0	0	1
7:	2009	793017	200906	NA	NA	NA
8:	2009	911867	200903	0	0	0
9:	2009	942671	200909	NA	NA	NA
10:	2009	979358	200912	NA	NA	NA

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N	HGHT
1:	NA	NA	NA	3	1	0	NA
2:	0	0	NA	3	1	0	146
3:	0	0	NA	2	1	0	NA
4:	NA	NA	NA	3	1	0	159
5:	NA	NA	NA	2	1	0	160
6:	0	0	NA	3	1	0	NA
7:	NA	NA	NA	3	1	0	156
8:	0	0	NA	3	1	0	NA
9:	NA	NA	NA	3	1	0	NA
10:	NA	NA	NA	2	1	0	157

Merge

Left join 다른 방법

```
1 dt2[dt1, on = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM")]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	HGHT	WGHT	WSTC	BMI	VA_LT	VA_RT	BP_SYS
1:	2009	562083	200909	NA	NA	NA	NA	NA	NA	NA
2:	2009	334536	200911	NA	NA	NA	NA	NA	NA	NA
3:	2009	911867	200903	NA	NA	NA	NA	NA	NA	NA
4:	2009	183321	200908	NA	NA	NA	NA	NA	NA	NA
5:	2009	942671	200909	NA	NA	NA	NA	NA	NA	NA
6:	2009	979358	200912	157	55	73	22.3	1.5	1.5	110
7:	2009	554112	200911	160	56	67	21.9	1.5	1.5	119
8:	2009	487160	200908	159	54	66	21.4	1.2	1.5	111
9:	2009	793017	200906	156	53	67	21.8	1.2	1.0	138
10:	2009	219397	200912	146	48	78	22.5	1.5	1.5	138

	BP_DIA	URN_PROT	HGB	FBS	TOT_CHOL	TG	HDL	LDL	CRTN	SGOT	SGPT	GGT	GFR
1:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
3:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
5:	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
6:	70	1	11.9	100	192	109	53	117	0.7	15	12	14	83
7:	78	1	11.2	84	152	38	43	101	0.8	8	6	10	97
8:	60	1	12.2	88	166	42	58	99	1.0	16	11	12	65
9:	72	1	11.0	74	155	86	52	85	0.6	15	13	13	96
10:	84	1	12.8	107	178	87	35	125	0.7	21	21	23	70

Merge

Right join

```
1 merge(dt1, dt2, by = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM"), all.y = T)
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	219397	200912	0	0	1
2:	2009	318669	200904	NA	NA	NA
3:	2009	395781	200907	NA	NA	NA
4:	2009	480569	200901	NA	NA	NA
5:	2009	487160	200908	NA	NA	NA
6:	2009	554112	200911	NA	NA	NA
7:	2009	559370	200912	NA	NA	NA
8:	2009	793017	200906	NA	NA	NA
9:	2009	831349	200912	NA	NA	NA
10:	2009	979358	200912	NA	NA	NA

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N	HGHT
1:	0	0	NA	3	1	0	146
2:	NA	NA	NA	NA	NA	NA	155
3:	NA	NA	NA	NA	NA	NA	165
4:	NA	NA	NA	NA	NA	NA	154
5:	NA	NA	NA	3	1	0	159
6:	NA	NA	NA	2	1	0	160
7:	NA	NA	NA	NA	NA	NA	155
8:	NA	NA	NA	3	1	0	156
9:	NA	NA	NA	NA	NA	NA	164
10:	NA	NA	NA	2	1	0	157

Merge

Right join 다른 방법

```
1 dt1[dt2, on = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM")]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	979358	200912	NA	NA	NA
2:	2009	554112	200911	NA	NA	NA
3:	2009	487160	200908	NA	NA	NA
4:	2009	793017	200906	NA	NA	NA
5:	2009	219397	200912	0	0	1
6:	2009	831349	200912	NA	NA	NA
7:	2009	480569	200901	NA	NA	NA
8:	2009	559370	200912	NA	NA	NA
9:	2009	318669	200904	NA	NA	NA
10:	2009	395781	200907	NA	NA	NA

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PT	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N	HGHT
1:	NA	NA	NA	2	1	0	157
2:	NA	NA	NA	2	1	0	160
3:	NA	NA	NA	3	1	0	159
4:	NA	NA	NA	3	1	0	156
5:	0	0	NA	3	1	0	146
6:	NA	NA	NA	NA	NA	NA	164
7:	NA	NA	NA	NA	NA	NA	154
8:	NA	NA	NA	NA	NA	NA	155
9:	NA	NA	NA	NA	NA	NA	155
10:	NA	NA	NA	NA	NA	NA	165

Merge

Left anti join

```
1 dt1[!dt2, on = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM")]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	562083	200909	0	0	1
2:	2009	334536	200911	0	0	0
3:	2009	911867	200903	0	0	0
4:	2009	183321	200908	NA	NA	NA
5:	2009	942671	200909	NA	NA	NA

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	0	0	NA	2	1	0
3:	0	0	NA	3	1	0
4:	NA	NA	NA	3	1	0
5:	NA	NA	NA	3	1	0

Merge

Right anti join

```
1 dt2[!dt1, on = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM")]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	HGHT	WGHT	WSTC	BMI	VA_LT	VA_RT	BP_SYS
1:	2009	831349	200912	164	66	85	24.5	1.2	1.0	130
2:	2009	480569	200901	154	58	84	24.5	0.1	0.8	140
3:	2009	559370	200912	155	51	75	21.2	1.0	1.0	93
4:	2009	318669	200904	155	66	78	27.5	0.8	0.6	95
5:	2009	395781	200907	165	60	71	22.0	0.6	0.4	120

	BP_DIA	URN_PROT	HGB	FBS	TOT_CHOL	TG	HDL	LDL	CRTN	SGOT	SGPT	GGT	GFR
1:	90	1	16.3	108	209	57	59	138	0.8	32	38	16	98
2:	70	1	10.4	84	242	134	53	162	1.1	15	10	10	37
3:	53	1	11.1	92	212	44	69	134	1.0	20	10	10	59
4:	58	1	13.6	101	294	119	50	220	0.6	22	22	22	116
5:	70	1	15.3	85	237	112	56	158	1.0	24	20	18	70

Mutate

새로운 변수 생성

```
1 dt[, BMI2 := round(WGHT/(HGHT/100)^2, 1)][[]]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	562083	200909	0	0	1
2:	2009	334536	200911	0	0	0
3:	2009	911867	200903	0	0	0
4:	2009	183321	200908	NA	NA	NA
5:	2009	942671	200909	NA	NA	NA

1640:	2015	266734	201504	NA	NA	NA
1641:	2015	343874	201511	0	0	0
1642:	2015	798472	201507	NA	NA	NA
1643:	2015	720080	201505	NA	NA	NA
1644:	2015	127519	201511	0	0	1

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PT	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	0	0	NA	2	1	0
3:	0	0	NA	3	1	0
4:	NA	NA	NA	3	1	0
5:	NA	NA	NA	3	1	0

1640:	NA	NA	NA	1	2	2
1641:	1	0	0	2	2	0
1642:	NA	NA	NA	3	2	0

Mutate

새로운 변수 생성

```
1 dt[, `:=`(BP_SYS140 = factor(as.integer(BP_SYS >= 140)), BMI25 = factor(as.integer(BMI >= 25)))][]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	562083	200909	0	0	1
2:	2009	334536	200911	0	0	0
3:	2009	911867	200903	0	0	0
4:	2009	183321	200908	NA	NA	NA
5:	2009	942671	200909	NA	NA	NA

1640:	2015	266734	201504	NA	NA	NA
1641:	2015	343874	201511	0	0	0
1642:	2015	798472	201507	NA	NA	NA
1643:	2015	720080	201505	NA	NA	NA
1644:	2015	127519	201511	0	0	1

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PT	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	0	0	NA	2	1	0
3:	0	0	NA	3	1	0
4:	NA	NA	NA	3	1	0
5:	NA	NA	NA	3	1	0

1640:	NA	NA	NA	1	2	2
1641:	1	0	0	2	2	0
1642:	NA	NA	NA	3	2	0

Mutate

BMI2 열 삭제

```
1 dt[, BMI2 := NULL][[]]
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2009	562083	200909	0	0	1
2:	2009	334536	200911	0	0	0
3:	2009	911867	200903	0	0	0
4:	2009	183321	200908	NA	NA	NA
5:	2009	942671	200909	NA	NA	NA

1640:	2015	266734	201504	NA	NA	NA
1641:	2015	343874	201511	0	0	0
1642:	2015	798472	201507	NA	NA	NA
1643:	2015	720080	201505	NA	NA	NA
1644:	2015	127519	201511	0	0	1

	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PT	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	0	0	NA	2	1	0
3:	0	0	NA	3	1	0
4:	NA	NA	NA	3	1	0
5:	NA	NA	NA	3	1	0

1640:	NA	NA	NA	1	2	2
1641:	1	0	0	2	2	0
1642:	NA	NA	NA	3	2	0

Specific symbol

.SD: Subset of Data

```
1 dt[, .SD]    # all column
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX HTN
1:	2009	562083	200909	0	0	1
2:	2009	334536	200911	0	0	0
3:	2009	911867	200903	0	0	0
4:	2009	183321	200908	NA	NA	NA
5:	2009	942671	200909	NA	NA	NA

1640:	2015	266734	201504	NA	NA	NA
1641:	2015	343874	201511	0	0	0
1642:	2015	798472	201507	NA	NA	NA
1643:	2015	720080	201505	NA	NA	NA
1644:	2015	127519	201511	0	0	1
	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PT	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	NA	3	1	0
2:	0	0	NA	2	1	0
3:	0	0	NA	3	1	0
4:	NA	NA	NA	3	1	0
5:	NA	NA	NA	3	1	0

1640:	NA	NA	NA	1	2	2
1641:	1	0	0	2	2	0
1642:	NA	NA	NA	3	2	0

Specific symbol

.SD: Subset of Data

```
1 dt[, lapply(.SD, class)]
```

```
      EXMD_BZ_YYYY RN_INDI HME_YYYYMM Q_PHX_DX_STK Q_PHX_DX_HTDZ Q_PHX_DX_HTN
1:      integer integer    integer      integer      integer      integer
      Q_PHX_DX_DM Q_PHX_DX_DLD Q_PHX_DX_PTB Q_HBV_AG Q_SMK_YN Q_DRK_FRQ_V09N
1:      integer    integer      integer integer integer      integer
      HGHT    WGHT    WSTC      BMI    VA_LT    VA_RT    BP_SYS    BP_DIA URN_PROT
1: integer integer integer numeric numeric numeric integer integer  integer
      HGB      FBS TOT_CHOL      TG      HDL      LDL      CRTN      SGOT      SGPT
1: numeric integer integer integer integer integer numeric integer integer
      GGT      GFR BP_SYS140    BMI25
1: integer integer      factor factor
```

Specific symbol

.SD: Subset of Data

```
1 dt[order(EXMD_BZ_YYYY), .SD[1], keyby = "RN_INDI"]
```

	RN_INDI	EXMD_BZ_YYYY	HME_YYYYMM	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN
1:	2270	2010	201012	0	0	0
2:	3690	2010	201005	0	0	0
3:	4263	2009	200910	NA	NA	NA
4:	4664	2009	200910	0	0	0
5:	5707	2010	201007	NA	NA	NA

543:	995838	2013	201312	0	0	0
544:	999693	2010	201003	0	0	0
545:	1005250	2012	201204	0	0	0
546:	1009573	2009	200904	0	0	0
547:	1010623	2010	201005	0	0	0
	Q_PHX_DX_DM	Q_PHX_DX_DLD	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N
1:	0	0	0	2	1	2
2:	0	0	0	2	3	1
3:	NA	NA	NA	2	2	1
4:	0	0	NA	2	1	0
5:	NA	NA	NA	3	1	1

543:	0	0	0	2	1	0
544:	0	0	0	2	2	1
545:	0	0	0	3	3	1

Specific symbol

.SDcols: 연산 대상이 되는 특정 열 지정

```
1 dt[order(EXMD_BZ_YYYY), .SD[1], .SDcols = colvars, keyby = "RN_INDI"]
```

	RN_INDI	Q_PHX_DX_STK	Q_PHX_DX_HTDZ	Q_PHX_DX_HTN	Q_PHX_DX_DM	Q_PHX_DX_DLD
1:	2270	0	0	0	0	0
2:	3690	0	0	0	0	0
3:	4263	NA	NA	NA	NA	NA
4:	4664	0	0	0	0	0
5:	5707	NA	NA	NA	NA	NA

543:	995838	0	0	0	0	0
544:	999693	0	0	0	0	0
545:	1005250	0	0	0	0	0
546:	1009573	0	0	0	0	1
547:	1010623	0	0	0	0	0
	Q_PHX_DX_PTB	Q_HBV_AG	Q_SMK_YN	Q_DRK_FRQ_V09N		
1:	0	2	1	2		
2:	0	2	3	1		
3:	NA	2	2	1		
4:	NA	2	1	0		
5:	NA	3	1	1		

543:	0	2	1	0		
544:	0	2	2	1		
545:	0	3	3	1		

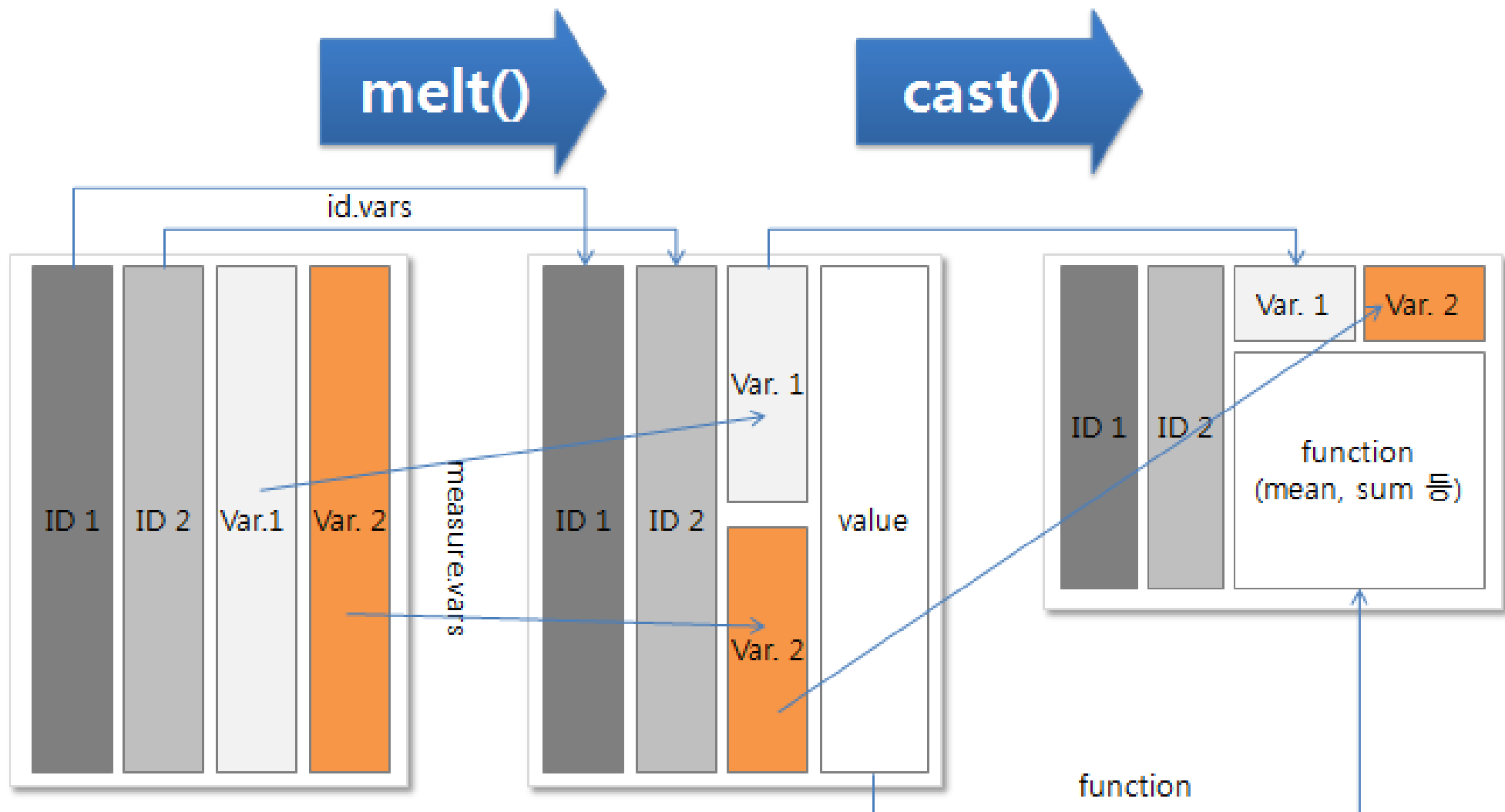
Specific symbol

.N: length()

```
1 dt[, .N, keyby = "RN_INDI"]
```

```
   RN_INDI  N  
1:      2270 2  
2:      3690 1  
3:      4263 7  
4:      4664 4  
5:      5707 2  
---  
543: 995838 2  
544: 999693 2  
545: 1005250 4  
546: 1009573 6  
547: 1010623 4
```

melt



[R 분석과 프로그래밍] <http://rfriend.tistory.com>

melt

일부 열을 제외한 나머지 열을 stack 처리

```
1 dt.long1 <- melt(dt,  
2     id.vars = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM"), # 고정할 열  
3     measure.vars = c("TOT_CHOL", "TG", "HDL", "LDL"), # 재구조화할 열  
4     variable.name = "Lipid", # 재구조화한 후 variable 열의 이름  
5     value.name = "Value") # 재구조화한 후 value 열의 이름  
6 dt.long1
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	Lipid	Value
1:	2009	562083	200909	TOT_CHOL	264
2:	2009	334536	200911	TOT_CHOL	169
3:	2009	911867	200903	TOT_CHOL	216
4:	2009	183321	200908	TOT_CHOL	199
5:	2009	942671	200909	TOT_CHOL	162

6572:	2015	266734	201504	LDL	244
6573:	2015	343874	201511	LDL	50
6574:	2015	798472	201507	LDL	151
6575:	2015	720080	201505	LDL	NA
6576:	2015	127519	201511	LDL	94

melt

동시에 여러 개의 열로 melt 가능

```
1 col1 <- c("BP_SYS", "BP_DIA")
2 col2 <- c("VA_LT", "VA_RT")
3 dt.long2 <- melt(dt,
4                   id.vars = c("EXMD_BZ_YYYY", "RN_INDI", "HME_YYYYMM"),
5                   measure = list(col1, col2),
6                   value.name = c("BP", "VA"))
7 dt.long2
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	variable	BP	VA
1:	2009	562083	200909	1	120	0.7
2:	2009	334536	200911	1	120	0.8
3:	2009	911867	200903	1	130	0.7
4:	2009	183321	200908	1	101	0.8
5:	2009	942671	200909	1	132	0.7

3284:	2015	266734	201504	2	89	1.2
3285:	2015	343874	201511	2	83	0.7
3286:	2015	798472	201507	2	77	1.5
3287:	2015	720080	201505	2	80	0.8
3288:	2015	127519	201511	2	85	0.8

dcast

melt 함수로 재구조화한 것을 다시 분리

```
1 dt.wide1 <- dcast(dt.long1, EXMD_BZ_YYYY + RN_INDI + HME_YYYYMM ~ Lipid, value.var = "Value")
2 dt.wide1
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	TOT_CHOL	TG	HDL	LDL
1:	2009	4263	200910	264	54	81	172
2:	2009	4664	200910	192	74	57	120
3:	2009	9866	200912	112	116	48	40
4:	2009	17079	200911	161	52	72	78
5:	2009	26776	200902	189	62	59	117

1640:	2015	985608	201512	189	61	70	106
1641:	2015	992522	201512	180	81	105	59
1642:	2015	995240	201511	222	319	74	84
1643:	2015	1005250	201504	159	105	60	78
1644:	2015	1010623	201502	244	440	48	NA

dcast

그룹별 요약 통계량을 계산한 결과를 재구조화하여 반환

```
1 dt.wide2 <- dcast(dt.long1, RN_INDI ~ Lipid, value.var = "Value", fun.aggregate = mean, na.rm =T)
2 dt.wide2
```

	RN_INDI	TOT_CHOL	TG	HDL	LDL
1:	2270	205.5000	105.50000	58.00000	126.5000
2:	3690	145.0000	101.00000	47.00000	78.0000
3:	4263	243.4286	47.14286	75.71429	158.0000
4:	4664	180.5000	88.25000	52.50000	110.0000
5:	5707	209.0000	78.00000	68.50000	124.5000

543:	995838	243.5000	122.00000	64.00000	155.0000
544:	999693	268.5000	454.00000	42.50000	163.0000
545:	1005250	158.0000	65.00000	64.50000	77.0000
546:	1009573	229.6667	153.16667	52.83333	145.8333
547:	1010623	222.5000	351.00000	48.75000	104.5000

dcast

여러 열을 동시에 dcast 가능

```
1 dt.wide3 <- dcast(dt.long2, ... ~ variable, value.var = c("BP", "VA"))
2 dt.wide3
```

	EXMD_BZ_YYYY	RN_INDI	HME_YYYYMM	BP_1	BP_2	VA_1	VA_2
1:	2009	4263	200910	110	70	1.5	2.0
2:	2009	4664	200910	110	80	0.9	0.9
3:	2009	9866	200912	121	63	1.0	0.5
4:	2009	17079	200911	106	60	1.0	0.8
5:	2009	26776	200902	118	73	0.8	1.0

1640:	2015	985608	201512	117	70	1.0	1.0
1641:	2015	992522	201512	116	67	0.7	0.5
1642:	2015	995240	201511	130	80	1.5	1.5
1643:	2015	1005250	201504	110	75	1.2	1.5
1644:	2015	1010623	201502	123	85	0.8	0.8