

R sqldf Package

I. "sqldf" Package

1) Install "sqldf" Package

```
devtools::install_github("ggrothendieck/sqldf")

Downloading GitHub repo ggrothendieck/sqldf@HEAD

bit      (4.0.4 -> 4.0.5 ) [CRAN]
plogr    (NA       -> 0.2.0 ) [CRAN]
proto    (NA       -> 1.0.0 ) [CRAN]
chron    (NA       -> 2.3-61) [CRAN]
RSQLite  (NA       -> 2.3.1 ) [CRAN]
gsubfn   (NA       -> 0.7   ) [CRAN]
Installing 6 packages: bit, plogr, proto, chron, RSQLite, gsubfn

Installing packages into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

--- R CMD build ---
* checking for file '/tmp/RtmpxRfhp9/remotes9ec93c3eb/ggrothendieck-sqldf-ba33807/DESCRIPTION' ... OK
* preparing 'sqldf':
* checking DESCRIPTION meta-information ... OK
* checking for LF line-endings in source and make files and shell scripts
* checking for empty or unneeded directories
* building 'sqldf_0.4-11.tar.gz'

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```

2) Installed Packages

```
library()
```

3) Import "sqldf" Package

```
library(sqldf)

Loading required package: gsubfn

Loading required package: proto

Warning message:
"no DISPLAY variable so Tk is not available"
Loading required package: RSQLite
```

II. Load Data

- read.csv()

```
url = 'https://raw.githubusercontent.com/rusita-ai/pyData/master/PII.csv'

TB <- read.csv(url)
```

- str()

```
str(TB)

'data.frame':   17 obs. of  8 variables:
 $ Name      : chr  "송태섭" "최유정" "이한나" "김소혜" ...
 $ Gender    : chr  "남자" "여자" "여자" "여자" ...
 $ Age       : int   21 23 20 23 24 24 22 22 23 22 ...
 $ Grade     : int   3 1 1 3 4 2 4 2 3 2 ...
 $ Picture   : chr  "무" "유" "무" "무" ...
 $ BloodType : chr  "B" "A" "A" "O" ...
 $ Height    : num   179 177 168 176 176 ...
 $ Weight    : num   63.9 54.9 50.2 53.5 79.8 61.7 69.8 52.7 68.5 52.3 ...
```

- head()

```
head(TB, 5)

      A data.frame: 5 × 8
   Name Gender   Age Grade Picture BloodType Height Weight
   <chr>  <chr> <int> <int>  <chr>    <chr>    <dbl>   <dbl>
1  송태섭   남자    21     3     무         B    179.1    63.9
2  최유정   여자    23     1     유         A    177.1    54.9
3  이한나   여자    20     1     무         A    167.9    50.2
4  김소혜   여자    23     3     무         O    176.1    53.5
5  서태웅   남자    24     4     무         B    176.1    79.8
```

III. SQL 수행

1) "SELECT * FROM Table_Name"

```
sqldf("SELECT Height FROM TB")
```

A
data.frame:
17 × 1
Height
<dbl>

179.1
177.1
167.9
176.1
176.1
175.2
169.1
169.9
165.5
161.9
163.2
157.8
166.9
181.8
180.3
155.5
168.9

- Import "dplyr" Package

```
library(dplyr)
```

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’ :

filter, lag

The following objects are masked from ‘package:base’ :

intersect, setdiff, setequal, union

```
sqldf("SELECT Gender, Height FROM TB") %>% head(3)
```

A data.frame: 3 × 2
Gender Height
<chr> <dbl>
1 남자 179.1
2 여자 177.1
3 여자 167.9

```
sqldf("SELECT Name, Height FROM TB") %>% head(3)
```

A data.frame: 3 × 2
Name Height
<chr> <dbl>
1 송태섭 179.1
2 최유정 177.1
3 이한나 167.9

```
sqldf("SELECT Gender, Height, BloodType FROM TB") %>% head(3)
```

A data.frame: 3 × 3
Gender Height BloodType
<chr> <dbl> <chr>
1 남자 179.1 B
2 여자 177.1 A
3 여자 167.9 A

```
sqldf("SELECT * FROM TB") %>% head(3)
```

A data.frame: 3 × 8
Name Gender Age Grade Picture BloodType Height Weight
<chr> <chr> <int> <int> <chr> <chr> <dbl> <dbl>
1 송태섭 남자 21 3 무 B 179.1 63.9
2 최유정 여자 23 1 유 A 177.1 54.9
3 이한나 여자 20 1 무 A 167.9 50.2

▼ 2) "WHERE" 조건구문

```
sqldf("SELECT * FROM TB WHERE Height > 175")
```

A data.frame: 7 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
최유정	여자	23	1	유	A	177.1	54.9
김소혜	여자	23	3	무	O	176.1	53.5
서태웅	남자	24	4	무	B	176.1	79.8

▼ (1) "AND" 연산

```
sqldf("SELECT * FROM TB WHERE Height > 175 AND Weight < 75")
```

A data.frame: 4 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
최유정	여자	23	1	유	A	177.1	54.9
김소혜	여자	23	3	무	O	176.1	53.5
정대만	남자	24	2	유	B	175.2	61.7

```
sqldf("SELECT * FROM TB WHERE Height > 175 AND Grade = 3")
```

A data.frame: 3 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
김소혜	여자	23	3	무	O	176.1	53.5
채치수	남자	23	3	무	AB	181.8	85.9

▼ (2) "OR" 연산

```
sqldf("SELECT * FROM TB WHERE Height > 175 OR Grade = 3")
```

A data.frame: 8 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
최유정	여자	23	1	유	A	177.1	54.9
김소혜	여자	23	3	무	O	176.1	53.5
서태웅	남자	24	4	무	B	176.1	79.8
정대만	남자	24	2	유	B	175.2	61.7
강백호	남자	23	3	무	O	165.5	68.5
채치수	남자	23	3	무	AB	181.8	85.9
윤대협	남자	22	2	유	AB	180.3	76.2

```
sqldf("SELECT * FROM TB WHERE Height > 175 OR Age = 24")
```

A data.frame: 8 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
최유정	여자	23	1	유	A	177.1	54.9
김소혜	여자	23	3	무	O	176.1	53.5
서태웅	남자	24	4	무	B	176.1	79.8
정대만	남자	24	2	유	B	175.2	61.7
권준호	남자	24	4	유	O	166.9	61.7
채치수	남자	23	3	무	AB	181.8	85.9
윤대협	남자	22	2	유	AB	180.3	76.2

▼ 3) 명목형 연산

```
sqldf("SELECT * FROM TB WHERE Name = '강백호'")
```

A data.frame: 1 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
강백호	남자	23	3	무	O	165.5	68.5

```
sqldf("SELECT * FROM TB WHERE BloodType = 'B'")
```

A data.frame: 5 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
서태웅	남자	24	4	무	B	176.1	79.8
정대만	남자	24	2	유	B	175.2	61.7
이정환	남자	22	4	무	B	169.1	69.8
정채연	여자	22	2	무	B	157.8	44.9

```
sqldf("SELECT * FROM TB WHERE Height > 175 AND BloodType = 'B'")
```

A data.frame: 3 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
서태웅	남자	24	4	무	B	176.1	79.8
정대만	남자	24	2	유	B	175.2	61.7

4) "IN" 연산자

```
sqldf("SELECT * FROM TB WHERE Grade IN ('2', '4')")
```

A data.frame: 8 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
서태웅	남자	24	4	무	B	176.1	79.8
정대만	남자	24	2	유	B	175.2	61.7
이정환	남자	22	4	무	B	169.1	69.8
채소연	여자	22	2	유	AB	169.9	52.7
전소미	여자	22	2	유	O	161.9	52.3
정채연	여자	22	2	무	B	157.8	44.9
권준호	남자	24	4	유	O	166.9	61.7
윤대협	남자	22	2	유	AB	180.3	76.2

```
sqldf("SELECT * FROM TB WHERE BloodType IN ('A', 'B', 'O')")
```

A data.frame: 14 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
최유정	여자	23	1	유	A	177.1	54.9
이한나	여자	20	1	무	A	167.9	50.2
김소혜	여자	23	3	무	O	176.1	53.5
서태웅	남자	24	4	무	B	176.1	79.8
정대만	남자	24	2	유	B	175.2	61.7
이정환	남자	22	4	무	B	169.1	69.8
강백호	남자	23	3	무	O	165.5	68.5
전소미	여자	22	2	유	O	161.9	52.3
변덕규	남자	21	1	무	A	163.2	55.5
정채연	여자	22	2	무	B	157.8	44.9
권준호	남자	24	4	유	O	166.9	61.7
김세정	여자	21	1	무	O	155.5	44.9
신준섭	남자	23	1	무	A	168.9	62.7

5) "LIKE" 연산자

- 특정 문자 시작, 끝, 포함하는 값을 추출

```
sqldf("SELECT * FROM TB WHERE Weight LIKE '5%'")
```

A data.frame: 6 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
최유정	여자	23	1	유	A	177.1	54.9
이한나	여자	20	1	무	A	167.9	50.2
김소혜	여자	23	3	무	O	176.1	53.5
채소연	여자	22	2	유	AB	169.9	52.7
전소미	여자	22	2	유	O	161.9	52.3
변덕규	남자	21	1	무	A	163.2	55.5

```
sqldf("SELECT * FROM TB WHERE Weight LIKE '7%'")
```

A data.frame: 2 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
서태웅	남자	24	4	무	B	176.1	79.8
윤대협	남자	22	2	유	AB	180.3	76.2

```
sqldf("SELECT * FROM TB WHERE Weight LIKE '%9'")
```

A data.frame: 5 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
최유정	여자	23	1	유	A	177.1	54.9
정채연	여자	22	2	무	B	157.8	44.9
채치수	남자	23	3	무	AB	181.8	85.9
김세정	여자	21	1	무	O	155.5	44.9

```
sqldf("SELECT * FROM TB WHERE Weight LIKE '%2'")
```

A data.frame: 2 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
이한나	여자	20	1	무	A	167.9	50.2
윤대협	남자	22	2	유	AB	180.3	76.2

```
sqldf("SELECT * FROM TB WHERE Name LIKE '김%'")
```

A data.frame: 2 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
김소혜	여자	23	3	무	O	176.1	53.5
김세정	여자	21	1	무	O	155.5	44.9

```
sqldf("SELECT * FROM TB WHERE Name LIKE '%정'")
```

A data.frame: 2 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
최유정	여자	23	1	유	A	177.1	54.9
김세정	여자	21	1	무	O	155.5	44.9

```
sqldf("SELECT * FROM TB WHERE Name LIKE '%소%'")
```

A data.frame: 3 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
김소혜	여자	23	3	무	O	176.1	53.5
채소연	여자	22	2	유	AB	169.9	52.7
전소미	여자	22	2	유	O	161.9	52.3

▼ 6) "GROUP BY" 연산자

- 중복값을 제외하고 1개의 고유값만 출력

```
sqldf("SELECT Grade FROM TB GROUP BY Grade")
```

A data.frame: 4 × 1							
Grade							
<int>							
1							
2							
3							
4							

▼ 7) "SUM()" 연산

```
sqldf("SELECT Grade, SUM(Age), SUM(Height), SUM(Weight)
      FROM TB GROUP BY Grade")
```

A data.frame: 4 × 4			
Grade	SUM(Age)	SUM(Height)	SUM(Weight)
<int>	<int>	<dbl>	<dbl>
1	108	832.6	268.2
2	112	845.1	287.8
3	90	702.5	271.8
4	70	512.1	211.3

▼ 8) "AVG()" 연산

```
sqldf("SELECT Grade, AVG(Age), AVG(Height), AVG(Weight)
      FROM TB GROUP BY Grade")
```

A data.frame: 4 × 4			
Grade	AVG(Age)	AVG(Height)	AVG(Weight)
<int>	<dbl>	<dbl>	<dbl>
1	21.60000	166.520	53.64000
2	22.40000	169.020	57.56000
3	22.50000	175.625	67.95000
4	23.33333	170.700	70.43333

▼ 9) "HAVING" 연산자

```
sqldf("SELECT Grade, AVG(Age), AVG(Height), AVG(Weight)
      FROM TB
      GROUP BY Grade
      HAVING AVG(Height) > 170")
```

A data.frame: 2 × 4			
Grade	AVG(Age)	AVG(Height)	AVG(Weight)
<int>	<dbl>	<dbl>	<dbl>
3	22.50000	175.625	67.95000
4	23.33333	170.700	70.43333

10) "ORDER BY" 연산자

- 오름차순 : **ASC**, 내림차순 : **DESC**

sqldf("SELECT * FROM TB WHERE Height > 170")

A data.frame: 7 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
송태섭	남자	21	3	무	B	179.1	63.9
최유정	여자	23	1	유	A	177.1	54.9
김소혜	여자	23	3	무	O	176.1	53.5
서태웅	남자	24	4	무	B	176.1	79.8
정대만	남자	24	2	유	B	175.2	61.7
채치수	남자	23	3	무	AB	181.8	85.9
윤대협	남자	22	2	유	AB	180.3	76.2

- Height 기준

sqldf("SELECT * FROM TB WHERE Height > 170 ORDER BY Height ASC")

A data.frame: 7 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
정대만	남자	24	2	유	B	175.2	61.7
김소혜	여자	23	3	무	O	176.1	53.5
서태웅	남자	24	4	무	B	176.1	79.8
최유정	여자	23	1	유	A	177.1	54.9
송태섭	남자	21	3	무	B	179.1	63.9
윤대협	남자	22	2	유	AB	180.3	76.2
채치수	남자	23	3	무	AB	181.8	85.9

sqldf("SELECT * FROM TB WHERE Height > 170 ORDER BY Height DESC")

A data.frame: 7 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
채치수	남자	23	3	무	AB	181.8	85.9
윤대협	남자	22	2	유	AB	180.3	76.2
송태섭	남자	21	3	무	B	179.1	63.9
최유정	여자	23	1	유	A	177.1	54.9
김소혜	여자	23	3	무	O	176.1	53.5
서태웅	남자	24	4	무	B	176.1	79.8
정대만	남자	24	2	유	B	175.2	61.7

- Weight 기준

sqldf("SELECT * FROM TB WHERE Height > 170 ORDER BY Weight ASC")

A data.frame: 7 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
김소혜	여자	23	3	무	O	176.1	53.5
최유정	여자	23	1	유	A	177.1	54.9
정대만	남자	24	2	유	B	175.2	61.7
송태섭	남자	21	3	무	B	179.1	63.9
윤대협	남자	22	2	유	AB	180.3	76.2
서태웅	남자	24	4	무	B	176.1	79.8
채치수	남자	23	3	무	AB	181.8	85.9

sqldf("SELECT * FROM TB WHERE Height > 170 ORDER BY Weight DESC")

A data.frame: 7 × 8							
Name	Gender	Age	Grade	Picture	BloodType	Height	Weight
<chr>	<chr>	<int>	<int>	<chr>	<chr>	<dbl>	<dbl>
채치수	남자	23	3	무	AB	181.8	85.9
서태웅	남자	24	4	무	B	176.1	79.8
윤대협	남자	22	2	유	AB	180.3	76.2
송태섭	남자	21	3	무	B	179.1	63.9
정대만	남자	24	2	유	B	175.2	61.7
최유정	여자	23	1	유	A	177.1	54.9
김소혜	여자	23	3	무	O	176.1	53.5

11) "AS" 연산자

- 출력 Column Name 변경

```
sqldf("SELECT SUM(Height) AS SUM_Height FROM TB")
```

A data.frame:

1 × 1

SUM_Height
2892.3

```
sqldf("SELECT AVG(Height) AS AVG_Height, AVG(Weight) AS AVG_Weight FROM TB")
```

A data.frame: 1 × 2

AVG_Height	AVG_Weight
170.1353	61.12353

###

End Of Document

###

