sqldf

```
# libraries
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Warning in fun(libname, pkgname): couldn't connect to display ":0"
```

```
## Loading required package: RSQLite
```

```
library(RH2)
```

```
## Loading required package: chron
```

```
## Loading required package: RJDBC
```

```
## Loading required package: DBI
```

```
## Loading required package: rJava
```

```
data("UCBAdmissions")

# must be a data frame
ucb <- as.data.frame(UCBAdmissions)

sqldf("select * from ucb")
```

```
##       Admit Gender Dept Freq
## 1  Admitted   Male    A  512
## 2  Rejected   Male    A  313
## 3  Admitted Female    A   89
## 4  Rejected Female    A   19
## 5  Admitted   Male    B  353
## 6  Rejected   Male    B  207
## 7  Admitted Female    B   17
## 8  Rejected Female    B    8
## 9  Admitted   Male    C  120
## 10 Rejected   Male    C  205
## 11 Admitted Female    C  202
## 12 Rejected Female    C  391
## 13 Admitted   Male    D  138
## 14 Rejected   Male    D  279
## 15 Admitted Female    D  131
## 16 Rejected Female    D  244
## 17 Admitted   Male    E   53
## 18 Rejected   Male    E  138
## 19 Admitted Female    E   94
## 20 Rejected Female    E  299
## 21 Admitted   Male    F   22
## 22 Rejected   Male    F  351
## 23 Admitted Female    F   24
## 24 Rejected Female    F  317
```

```
majors <- data.frame(major = c("math", "biology", "engineering", "computer science", "history", "archite

sqldf("select * from majors")
```

```
##              major  Dept Faculty
## 1             math     A      24
## 2          biology     B      24
## 3      engineering     C      25
## 4 computer science     D      12
## 5          history     E      25
## 6     architecture Other      30
```

STEP 1 : General Queries

```
# Return Female student admission result
sqldf("select * from ucb where Gender = 'Female'")
```

```
##       Admit Gender Dept Freq
## 1  Admitted Female    A   89
## 2  Rejected Female    A   19
## 3  Admitted Female    B   17
## 4  Rejected Female    B    8
## 5  Admitted Female    C  202
## 6  Rejected Female    C  391
## 7  Admitted Female    D  131
## 8  Rejected Female    D  244
## 9  Admitted Female    E   94
## 10 Rejected Female    E  299
## 11 Admitted Female    F   24
## 12 Rejected Female    F  317
```

```
# Return the admitted students
sqldf("select * from ucb where Admit = 'Admitted'")
```

```
##       Admit Gender Dept Freq
## 1  Admitted   Male    A  512
## 2  Admitted Female    A   89
## 3  Admitted   Male    B  353
## 4  Admitted Female    B   17
## 5  Admitted   Male    C  120
## 6  Admitted Female    C  202
## 7  Admitted   Male    D  138
## 8  Admitted Female    D  131
## 9  Admitted   Male    E   53
## 10 Admitted Female    E   94
## 11 Admitted   Male    F   22
## 12 Admitted Female    F   24
```

```
# order admissions per department
sqldf("select * from ucb where Admit = 'Admitted' order by Freq DESC")
```

```
##       Admit Gender Dept Freq
## 1  Admitted   Male    A  512
## 2  Admitted   Male    B  353
```

```
## 3  Admitted Female    C  202
## 4  Admitted   Male    D  138
## 5  Admitted Female    D  131
## 6  Admitted   Male    C  120
## 7  Admitted Female    E   94
## 8  Admitted Female    A   89
## 9  Admitted   Male    E   53
## 10 Admitted Female    F   24
## 11 Admitted   Male    F   22
## 12 Admitted Female    B   17
```

```
# how many departments are in this table
sqldf("select distinct Dept from ucb")
```

```
##   Dept
## 1    A
## 2    B
## 3    C
## 4    D
## 5    E
## 6    F
```

STEP 2 : Aggregate Queries

```
# total admitted studets
sqldf("select sum(Freq) from ucb where Admit = 'Admitted'")
```

```
##   SUM("Freq")
## 1        1755
```

```
# total rejected students
sqldf("select sum(Freq) from ucb where Admit = 'Rejected'")
```

```
##   SUM("Freq")
## 1        2771
```

```
# return total admitted males
sqldf("select sum(Freq) as total_dudes from ucb where Admit = 'Admitted' AND Gender = 'Male'")
```

```
##   total_dudes
## 1        1198
```

```
# return total reject females
sqldf("select sum(Freq) as total_ladies from ucb where Admit = 'Rejected' AND Gender = 'Female'")
```

```
##   total_ladies
## 1         1278
```

```
# average number of admitted student by department (usually mean)
sqldf("select Dept, avg(Freq) as average_admitted from ucb where Admit = 'Admitted' group by Dept")
```

```
##   Dept average_admitted
## 1    A            300.5
## 2    B            185.0
## 3    C            161.0
## 4    D            134.5
```

```
## 5    E             73.5
## 6    F             23.0
```

```
# how many majors are there
sqldf("select count(major) from majors")
```

```
##   COUNT("major")
## 1              6
```

```
# minimum amount of studets rejected
sqldf("select min(Freq) from ucb where Admit = 'Rejected'")
```

```
##   MIN("Freq")
## 1           8
```

STEP 3 : Wild card match Queries

```
sqldf("select * from ucb where Freq between 20 AND 100")
```

```
##      Admit Gender Dept Freq
## 1 Admitted Female    A   89
## 2 Admitted   Male    E   53
## 3 Admitted Female    E   94
## 4 Admitted   Male    F   22
## 5 Admitted Female    F   24
```

```
sqldf("select * from ucb where Gender Like 'Fe%'")
```

```
##       Admit Gender Dept Freq
## 1  Admitted Female    A   89
## 2  Rejected Female    A   19
## 3  Admitted Female    B   17
## 4  Rejected Female    B    8
## 5  Admitted Female    C  202
## 6  Rejected Female    C  391
## 7  Admitted Female    D  131
## 8  Rejected Female    D  244
## 9  Admitted Female    E   94
## 10 Rejected Female    E  299
## 11 Admitted Female    F   24
## 12 Rejected Female    F  317
```

```
sqldf("select * from ucb where Gender Like '%male%'")
```

```
##       Admit Gender Dept Freq
## 1  Admitted Female    A   89
## 2  Rejected Female    A   19
## 3  Admitted Female    B   17
## 4  Rejected Female    B    8
## 5  Admitted Female    C  202
## 6  Rejected Female    C  391
## 7  Admitted Female    D  131
## 8  Rejected Female    D  244
## 9  Admitted Female    E   94
## 10 Rejected Female    E  299
```

```
## 11 Admitted Female    F   24
## 12 Rejected Female    F  317
```

```
sqldf("select * from ucb where Gender Like 'Ma%'")
```

```
##       Admit Gender Dept Freq
## 1  Admitted   Male    A  512
## 2  Rejected   Male    A  313
## 3  Admitted   Male    B  353
## 4  Rejected   Male    B  207
## 5  Admitted   Male    C  120
## 6  Rejected   Male    C  205
## 7  Admitted   Male    D  138
## 8  Rejected   Male    D  279
## 9  Admitted   Male    E   53
## 10 Rejected   Male    E  138
## 11 Admitted   Male    F   22
## 12 Rejected   Male    F  351
```

```
sqldf("select * from ucb where Gender = 'Female' AND Freq >= 100 ")
```

```
##      Admit Gender Dept Freq
## 1 Admitted Female    C  202
## 2 Rejected Female    C  391
## 3 Admitted Female    D  131
## 4 Rejected Female    D  244
## 5 Rejected Female    E  299
## 6 Rejected Female    F  317
```

```
sqldf("select * from ucb where Gender Like '_ale'")
```

```
##       Admit Gender Dept Freq
## 1  Admitted   Male    A  512
## 2  Rejected   Male    A  313
## 3  Admitted   Male    B  353
## 4  Rejected   Male    B  207
## 5  Admitted   Male    C  120
## 6  Rejected   Male    C  205
## 7  Admitted   Male    D  138
## 8  Rejected   Male    D  279
## 9  Admitted   Male    E   53
## 10 Rejected   Male    E  138
## 11 Admitted   Male    F   22
## 12 Rejected   Male    F  351
```

```
sqldf("select * from ucb where Gender NOT Like 'M_l_'")
```

```
##       Admit Gender Dept Freq
## 1  Admitted Female    A   89
## 2  Rejected Female    A   19
## 3  Admitted Female    B   17
## 4  Rejected Female    B    8
## 5  Admitted Female    C  202
## 6  Rejected Female    C  391
## 7  Admitted Female    D  131
## 8  Rejected Female    D  244
## 9  Admitted Female    E   94
```

```
## 10 Rejected Female    E  299
## 11 Admitted Female    F   24
## 12 Rejected Female    F  317
```

STEP 4 : Manipulation & Nested Queries

```
# Which department had the most admitted students = A
sqldf("select Dept from ucb where Freq = (select max(Freq) from ucb where Admit = 'Admitted')")
```

```
##   Dept
## 1    A
```

```
# which department had the most admitted Female student = C
sqldf("select Dept from ucb where Freq = (select max(Freq) from ucb where Gender = 'Female')")
```

```
##   Dept
## 1    C
```

```
# department with most faculty
sqldf("select Dept from majors where Faculty = (select max(Faculty) from majors)")
```

```
##    Dept
## 1 Other
```

STEP 5 : Join Queries

```
# join the two tables together by the common key
sqldf("select * from ucb inner join majors on ucb.Dept = majors.Dept")
```

```
##       Admit Gender Dept Freq            major Dept Faculty
## 1  Admitted   Male    A  512             math    A      24
## 2  Rejected   Male    A  313             math    A      24
## 3  Admitted Female    A   89             math    A      24
## 4  Rejected Female    A   19             math    A      24
## 5  Admitted   Male    B  353          biology    B      24
## 6  Rejected   Male    B  207          biology    B      24
## 7  Admitted Female    B   17          biology    B      24
## 8  Rejected Female    B    8          biology    B      24
## 9  Admitted   Male    C  120      engineering    C      25
## 10 Rejected   Male    C  205      engineering    C      25
## 11 Admitted Female    C  202      engineering    C      25
## 12 Rejected Female    C  391      engineering    C      25
## 13 Admitted   Male    D  138 computer science    D      12
## 14 Rejected   Male    D  279 computer science    D      12
## 15 Admitted Female    D  131 computer science    D      12
## 16 Rejected Female    D  244 computer science    D      12
## 17 Admitted   Male    E   53          history    E      25
## 18 Rejected   Male    E  138          history    E      25
## 19 Admitted Female    E   94          history    E      25
## 20 Rejected Female    E  299          history    E      25
```

```r
# join the table on the left with resultant nulls's on the right table
sqldf("select * from ucb left join majors on ucb.Dept = majors.Dept")
```

```
##       Admit Gender Dept Freq            major Dept Faculty
## 1  Admitted   Male    A  512             math    A      24
## 2  Rejected   Male    A  313             math    A      24
## 3  Admitted Female    A   89             math    A      24
## 4  Rejected Female    A   19             math    A      24
## 5  Admitted   Male    B  353          biology    B      24
## 6  Rejected   Male    B  207          biology    B      24
## 7  Admitted Female    B   17          biology    B      24
## 8  Rejected Female    B    8          biology    B      24
## 9  Admitted   Male    C  120      engineering    C      25
## 10 Rejected   Male    C  205      engineering    C      25
## 11 Admitted Female    C  202      engineering    C      25
## 12 Rejected Female    C  391      engineering    C      25
## 13 Admitted   Male    D  138 computer science    D      12
## 14 Rejected   Male    D  279 computer science    D      12
## 15 Admitted Female    D  131 computer science    D      12
## 16 Rejected Female    D  244 computer science    D      12
## 17 Admitted   Male    E   53          history    E      25
## 18 Rejected   Male    E  138          history    E      25
## 19 Admitted Female    E   94          history    E      25
## 20 Rejected Female    E  299          history    E      25
## 21 Admitted   Male    F   22             <NA> <NA>      NA
## 22 Rejected   Male    F  351             <NA> <NA>      NA
## 23 Admitted Female    F   24             <NA> <NA>      NA
## 24 Rejected Female    F  317             <NA> <NA>      NA
```

```r
# join the table on the right with the left
sqldf("select * from ucb right join majors on ucb.Dept = majors.Dept")
```

```
##                major  Dept Faculty    Admit Gender Dept Freq
## 1               math     A      24 Admitted   Male    A  512
## 2               math     A      24 Rejected   Male    A  313
## 3               math     A      24 Admitted Female    A   89
## 4               math     A      24 Rejected Female    A   19
## 5            biology     B      24 Admitted   Male    B  353
## 6            biology     B      24 Rejected   Male    B  207
## 7            biology     B      24 Admitted Female    B   17
## 8            biology     B      24 Rejected Female    B    8
## 9        engineering     C      25 Admitted   Male    C  120
## 10       engineering     C      25 Rejected   Male    C  205
## 11       engineering     C      25 Admitted Female    C  202
## 12       engineering     C      25 Rejected Female    C  391
## 13 computer science     D      12 Admitted   Male    D  138
## 14 computer science     D      12 Rejected   Male    D  279
## 15 computer science     D      12 Admitted Female    D  131
## 16 computer science     D      12 Rejected Female    D  244
## 17           history     E      25 Admitted   Male    E   53
## 18           history     E      25 Rejected   Male    E  138
## 19           history     E      25 Admitted Female    E   94
## 20           history     E      25 Rejected Female    E  299
## 21     architecture Other      30     <NA>   <NA> <NA>   NA
```