Week2 Assignment :- Movie review database

Vishal Arora

February 9, 2019

Overview :-

The week 2 assignemnt requires to collect user reviews for movies . To constuct various tables to store the user information, survey and movies information. So that some meanginfull inference can be deduced from them.

About Data : How the data is gathered and segeragated

The survey results below shows what survey ratings provided by which user to which movie. Based on the survey rating data, we normalized the data tables and segeregated the data into different tables.

Data Dictionary

The various data survey elements corresponds to:-

1) First & Last Name :- User first & last name who gave the survey.
2) Age :- User age
3) Gender :- user gender who took part in survey.
4) columns 5,6,7,8,9, 10 :- are all movies for whom user filled out the survey.

```
movies_survey <- read_csv("SurveyTemplate.csv")

#View(movies_survey)
DT::datatable(movies_survey , options = list(pageLength = 5))
```

Problem Statement

Data from survey is loaded into the database and then based on the normalization we segeregated the data into various tables. Below part of code shows how we can make native connection to DB using DBI and native DB libraries. Or else we can use the ODBC connection using RODBC bridge. Below both the techniques have been shown how they work in making the connection.

Establish connection using DBI and RMySQL libraries for native connection, fetching the list of tables in the movies_schema

```
con <- dbConnect(dbDriver('MySQL'),dbname="movies_sch",user="root",password="newrootpassword", port=330

listTab <- dbListTables(con)

listTab[1]
```

```
## [1] "movies"
```

```
#Movies
dbReadTable(con,listTab[1])
```

```
##   X.ID          Movie_Name       Genre
## 1    1        Lego Movie 2   Animation
## 2    2        Cold Pursuit      Action
## 3    3             The Prodigy       Horror
## 4    4 Under the Eiffel Tower       Romance
## 5    5              The Upside       Comedy
## 6    6                   Glass Drama/Sci-fi
```

```
listTab[2]
```

```
## [1] "participants"
```

```
#Participants
dbReadTable(con,listTab[2])
```

```
##   ID First.Name Last.Name Age Gender
## 1  1      Laura   Belcher  39      F
## 2  2      Elyse     Johns  42      F
## 3  3     Thomas      Cook  20      M
## 4  4      David  schummer  65      M
## 5  5      Chris    Hendry  10      M
## 6  6      Jason     Beans  29      M
```

```
listTab[3]
```

```
## [1] "rating"
```

```
#Rating
dbReadTable(con,listTab[3])
```

```
##   RatingID    Description
## 1        1 Not Interested
## 2        2           Poor
## 3        3        Average
## 4        4           Good
## 5        5    Exceptional
```

```
listTab[4]
```

```
## [1] "surveytable"
```

```
#SurveyTable
dbReadTable(con,listTab[4])
```

```
##    PersonID MovieID RatingID
## 1         1       1        1
## 2         2       1        4
## 3         3       1        3
## 4         4       1        3
## 5         5       1        5
## 6         6       1        1
## 7         1       2        3
## 8         2       2        2
## 9         3       2        5
## 10        4       2        2
## 11        5       2        3
## 12        6       2        3
## 13        1       3        2
## 14        2       3        1
## 15        3       3        4
## 16        4       3        1
## 17        5       3        1
## 18        6       3        4
## 19        1       4        5
## 20        2       4        5
## 21        3       4        2
## 22        4       4        4
## 23        5       4        2
## 24        6       4        4
## 25        1       5        3
## 26        2       5        3
## 27        3       5        4
## 28        4       5        3
## 29        5       5        4
## 30        6       5        4
## 31        1       6        4
## 32        2       6        4
## 33        3       6        2
## 34        4       6        4
## 35        5       6        1
## 36        6       6        3
```

```
dbDisconnect(con)
```

```
## [1] TRUE
```

Establish connection using the RODBC librarymaking use of ODBC connection, and fetching the various datales data and writting query to fetch data from all tables based on join conditions.

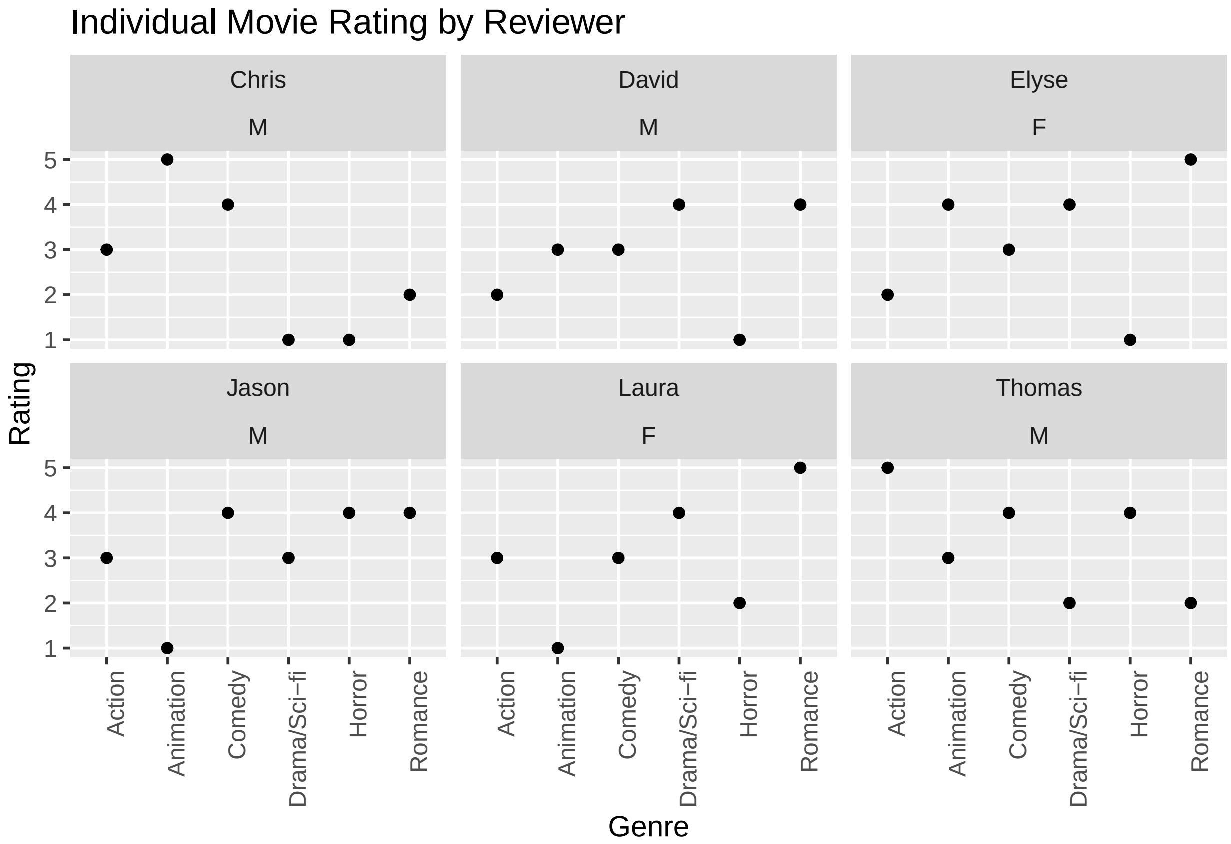
Display the first 10 records using the head funtion.

```
odbConn <- odbcConnect("odbcConn")



sqlquery1 <- "SELECT participants.`First Name`, participants.`Gender` , rating.`Description` , rating.`I

df_survey <- sqlQuery(odbConn, sqlquery1)

head(df_survey , 10)
```

```
##    First Name Gender    Description RatingID            Movie_Name
## 1       Chris      M        Average        3          Cold Pursuit
## 2       Chris      M    Exceptional        5          Lego Movie 2
## 3       Chris      M           Good        4            The Upside
## 4       Chris      M Not Interested        1           The Prodigy
## 5       Chris      M Not Interested        1                 Glass
## 6       Chris      M           Poor        2 Under the Eiffel Tower
## 7       David      M        Average        3          Lego Movie 2
## 8       David      M        Average        3            The Upside
## 9       David      M           Good        4 Under the Eiffel Tower
## 10      David      M           Good        4                 Glass
##          Genre
## 1       Action
## 2    Animation
## 3       Comedy
## 4        Horror
## 5  Drama/Sci-fi
## 6       Romance
## 7     Animation
## 8        Comedy
## 9       Romance
## 10 Drama/Sci-fi
```

Plot a diagram using the above data fetched from query to show the user preference for respective Genre's of movies.

```
qplot(Genre, RatingID, data=df_survey,xlab = "Genre", ylab = "Rating", main = "Individual Movie Rating
```



```
close(odbConn)
```

Summary

We can infer from above plot that every user has thier own preference for Genre of movies. Like Laura & Elyse has more preference towards Romance Genre , and similarly Chris has more interest in Animation movies, whereas Thomas has more interest in Action movies. Jason and David have interest in varied genres. So now we can use this inference to present them movies in the genres which they prefer more.