

DATA 607: SQL and R(Lab2)

Subhalaxmi Rout

This is by design a very open-ended assignment. In general, there's no need here to ask "Can I...?" questions about your proposed approach. A variety of reasonable approaches are acceptable. You could for example access the SQL data directly from R, or you could create an intermediate .CSV file. I should be able to generate the SQL table(s) and data from your provided code—if you use a graphical user interface to create and populate tables, it should have a mechanism to generate corresponding SQL code.

install required packages

```
{r setup, include=FALSE } #install.packages("RMySQL") #install.packages("readxl")
#install.packages('tinytex') #tinytex::install_tinytex()
```

Load Libraries

```
library(RMySQL)
```

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

```
library(ggplot2)
```

Create connection and fetch table from MySQL

```
db_user <- 'root'
db_password <- 'root'
db_name <- 'movie'
db_host <- 'localhost'
db_port <- 3306
mydb <- dbConnect(MySQL(), user = db_user, password = db_password, dbname = db_name, host = db_host, port = db_port)
dbListTables(mydb)
```

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

```
dbReadTable(mydb, "movies")
```

```
##   movie_id      title
## 1      1 Guardians of the Galaxy 2
## 2      2      Wonder Woman
```

```
## 3      3      Star Wars The Last Jedi
## 4      4      Thor Ragnarok
## 5      5      Blade Runner 2049
## 6      6      Spider Man Homecoming
```

Apply condition “IFNULL”, this replace NULL value to ‘NA’

```
sql <- "SELECT
      m.title as MovieTitle, IFNULL(r.rating, 'NA') AS Rating
FROM
  Movies m
left JOIN
  rating r ON m.movie_id = r.movie_id
where r.rating is not null";
result <- dbGetQuery(mydb, sql);
knitr::kable(
  head(result),
  caption = "dbGetQuery result set [head]"
)
```

Table 1: dbGetQuery result set [head]

MovieTitle	Rating
Guardians of the Galaxy 2	4
Wonder Woman	3
Star Wars The Last Jedi	5
Thor Ragnarok	2
Blade Runner 2049	4

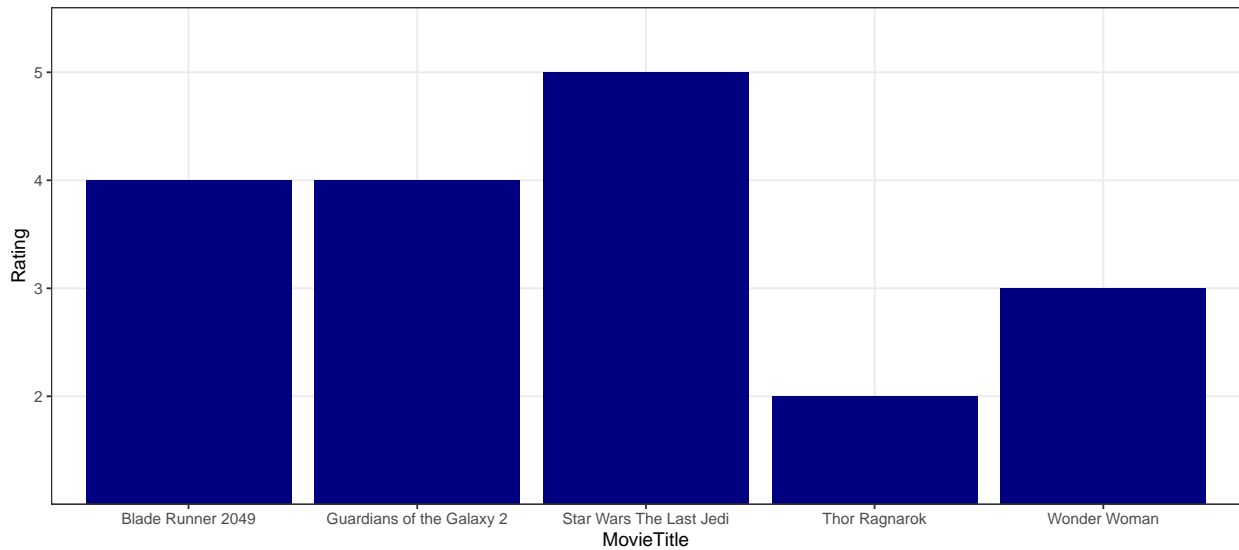
Draw bar chart show movie name with rating

```
movies_ratings_df <- as.data.frame(result)
show(movies_ratings_df)
```

```
##      MovieTitle Rating
## 1 Guardians of the Galaxy 2    4
## 2 Wonder Woman    3
## 3 Star Wars The Last Jedi    5
## 4 Thor Ragnarok    2
## 5 Blade Runner 2049    4
```

```
height <- as.numeric(movies_ratings_df$Rating)
names.arg <- movies_ratings_df$MovieTitle
```

```
my_plot <- ggplot(movies_ratings_df, aes(x=MovieTitle, y=Rating)) +geom_col(fill="navy")+theme_bw()
print(my_plot)
```



Using {sql, eval=FALSE} markdown to only display SQL

```
CREATE DATABASE movie;
USE movie;
DROP TABLE IF EXISTS movies;
DROP TABLE IF EXISTS rating;
/* Create movie table */
CREATE TABLE movies (
    movie_id INTEGER NOT NULL PRIMARY KEY,
    title VARCHAR(60) NOT NULL UNIQUE KEY
);
/* Insert values to movies table */
INSERT INTO movies (movie_id,title)
VALUES (1, "Guardians of the Galaxy 2"),
(2, "Wonder Woman"),
(3, "Star Wars The Last Jedi"),
(4, "Thor Ragnarok"),
(5, "Blade Runner 2049"),
(6, "Spider Man Homecoming");
/* Create rating table */
CREATE TABLE rating (
    rating_id INTEGER NOT NULL PRIMARY KEY,
    movie_id INTEGER,
    rating VARCHAR(60) UNIQUE KEY);

/* Insert values to movies table */
insert into rating (movie_id,rating_id,rating)
values (1,1,4),
(2,2,3),
(3,3,5),
(4,4,2),
(5,5,4);
```