Data607 Assignment 2

Yina Qiao

2023-02-04

Introductions

Choose six recent popular movies. Ask 5 friends to rate each of these movies that they have seen on a scale of 1 to 5. Take the results (observations) and store them in a SQL database. Load the information from the SQL database into an R dataframe and get average movie ratings.

Loading Library

```
library(RMySQL)
```

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

```
library(DBI)
```

```
library(ggplot2)
```

DB connection

```
# Helper for getting new connection to Cloud SQL
getSqlConnection <- function(){
  con <-
    dbConnect(
      RMySQL::MySQL(),
      username = 'root',
      password = 'yina',
      host = '35.202.155.99',
      dbname = 'movieDB'
    ) # TODO: use a configuration group `group = "my-db")`
  return(con)
}
```

listing table movieDB database

```
conn <- getSqlConnection()
res <- dbListTables(conn)
print(res)
```

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

load data into R

```
q<-"select* from rating;"
Ratings<-dbGetQuery(conn,q)
print(Ratings)
```

```
##                               movie rater rating
## 1           Avatar: The Way of Water jamie      5
## 2           Avatar: The Way of Water  lisa      4
## 3           Avatar: The Way of Water  mike      5
## 4           Avatar: The Way of Water   dan      4
## 5           Avatar: The Way of Water shawn      4
## 6                     Shotgun Wedding  mike      4
## 7                     Shotgun Wedding   dan      5
## 8                     Shotgun Wedding shawn      4
## 9                       Puss in Boots  lisa      4
## 10                      Puss in Boots  mike      5
## 11                      Puss in Boots   dan      4
## 12                      Puss in Boots shawn      3
## 13         The Banshees of Inisherin jamie      5
## 14         The Banshees of Inisherin  lisa      4
## 15         The Banshees of Inisherin  mike      3
## 16         The Banshees of Inisherin   dan      5
## 17         The Banshees of Inisherin shawn      4
## 18 Everything Everywhere All at Once jamie      4
## 19 Everything Everywhere All at Once  lisa      3
## 20 Everything Everywhere All at Once  mike      5
## 21 Everything Everywhere All at Once   dan      4
## 22 Everything Everywhere All at Once shawn      5
## 23                              M3GAN jamie      4
## 24                              M3GAN  lisa      5
## 25                              M3GAN  mike      2
## 26                              M3GAN   dan      4
## 27                              M3GAN shawn      1
```

Summary of Table

```
summary(Ratings)
```

```
##     movie              rater              rating
##  Length:27          Length:27          Min.   :1.000
##  Class :character   Class :character   1st Qu.:4.000
##  Mode  :character   Mode  :character   Median :4.000
##                                        Mean   :4.037
##                                        3rd Qu.:5.000
##                                        Max.   :5.000
```

Missing Value

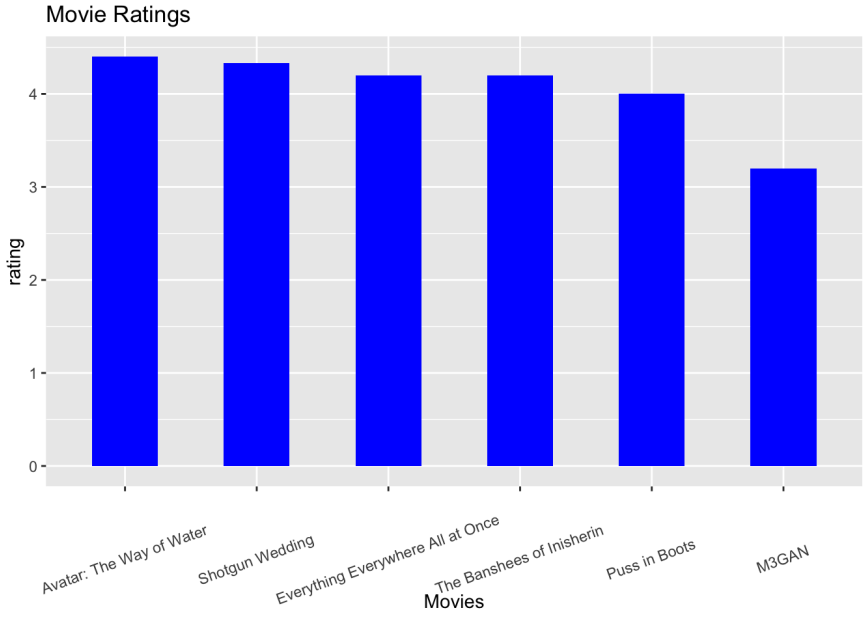
There are 30 entries contain of 5 friends's rating for 6 movies. Missing values has been treating when creating table in SQL.

Average rating

```
avg_movie_rating <- aggregate(x=Ratings["rating"], by = list(movie=Ratings$movie), FUN = mean, , na.rm=TRUE)
avg_movie_rating
```

```
##                               movie   rating
## 1           Avatar: The Way of Water 4.400000
## 2 Everything Everywhere All at Once 4.200000
## 3                              M3GAN 3.200000
## 4                        Puss in Boots 4.000000
## 5                     Shotgun Wedding 4.333333
## 6          The Banshees of Inisherin 4.200000
```

```
ggplot(avg_movie_rating, aes(x= reorder(movie, -rating), y=rating)) +
  geom_bar(stat="identity", width=.5, fill='BLUE') +
  labs(title="Movie Ratings") +
  theme(axis.text.x = element_text(angle=20, vjust=0.3)) +
  scale_x_discrete(name="Movies")
```



conclusion

The graph shows Avatar: The Way of Water has the highest rating while M3GAN has the lowest rating. I would be interested in explore the relationship further by taking into account of the genre of the movie, actor and how that affect the rating of the movie.