data_607lab2

Michael Robinson

2023-09-15

Connecting to SQL database

```
library(DBI)
library(RMySQL)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

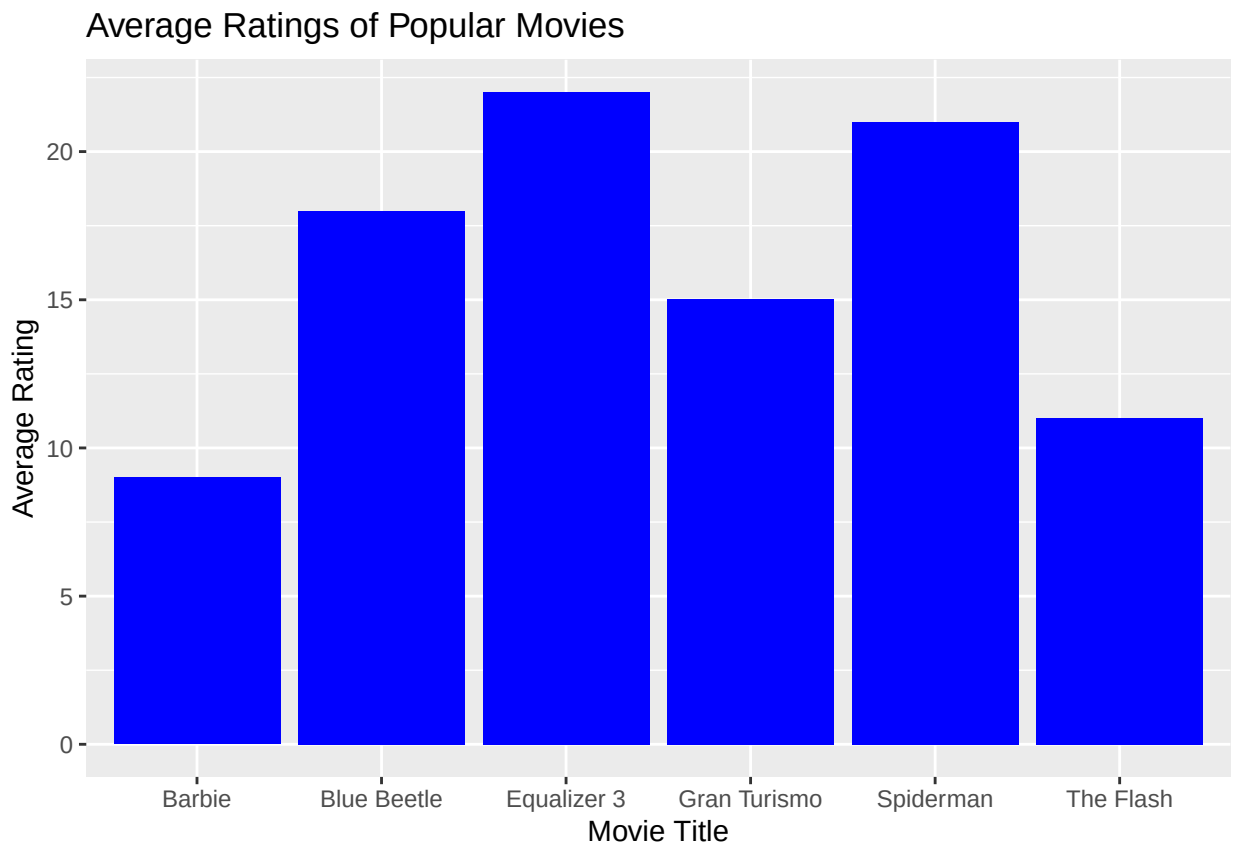
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
con <- dbConnect(MySQL(),
                 user = "root",
                 password = "",
                 dbname = "MovieRating",
                 host = "35.192.143.134")
```

Loading data from SQL database

```
ratings_data <- dbGetQuery(con, "SELECT * FROM movie_ratings")

# graphic representation of rating data
ggplot(ratings_data, aes(x = movie, y = rating)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Average Ratings of Popular Movies",
       x = "Movie Title",
       y = "Average Rating")
```




```
dbDisconnect(con)
```

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

```
# Group the data by movie name and summarize ratings

formatted_data <- ratings_data %>%
  group_by(movie) %>%
  summarize(avg_rating = mean(rating))

print(formatted_data)
```

```
## # A tibble: 6 x 2
##   movie        avg_rating
##   <chr>             <dbl>
## 1 Barbie              1.8
## 2 Blue Beetle         3.6
## 3 Equalizer 3         4.4
## 4 Gran Turismo        3
## 5 Spiderman           4.2
## 6 The Flash           2.2
```

Conclusion: after looking at the data and the Chart i can conclude that the Movie Equalizer had the Highest rating and Barbie was the lowest.