DATA607_Assignment_2

PeiMing_Chen

2023-02-04

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Introduction

A database of a CSV file that established in MySQL Workbench was loaded into Rstudio. I collected six movies reviews from five of my friends. Through R, I did some rating analysis and data cleaning of this database.

```
Ratingsdata <- data.frame(read.csv("https://raw.githubusercontent.com/PMCformosa/DATA607_Assignment2/ma

head(Ratingsdata)
```

```
##   Friends FirstName         MovieName Ratings
## 1       1     Jenny            Avatar       4
## 2       1     Jenny             PLANE    NULL
## 3       1     Jenny      80 FOR BRADY       5
## 4       1     Jenny A MAN CALLED OTOO    NULL
## 5       1     Jenny             MEGAN       5
## 6       1     Jenny         The Whale       3
```

Average rating of each movie from five of my friends

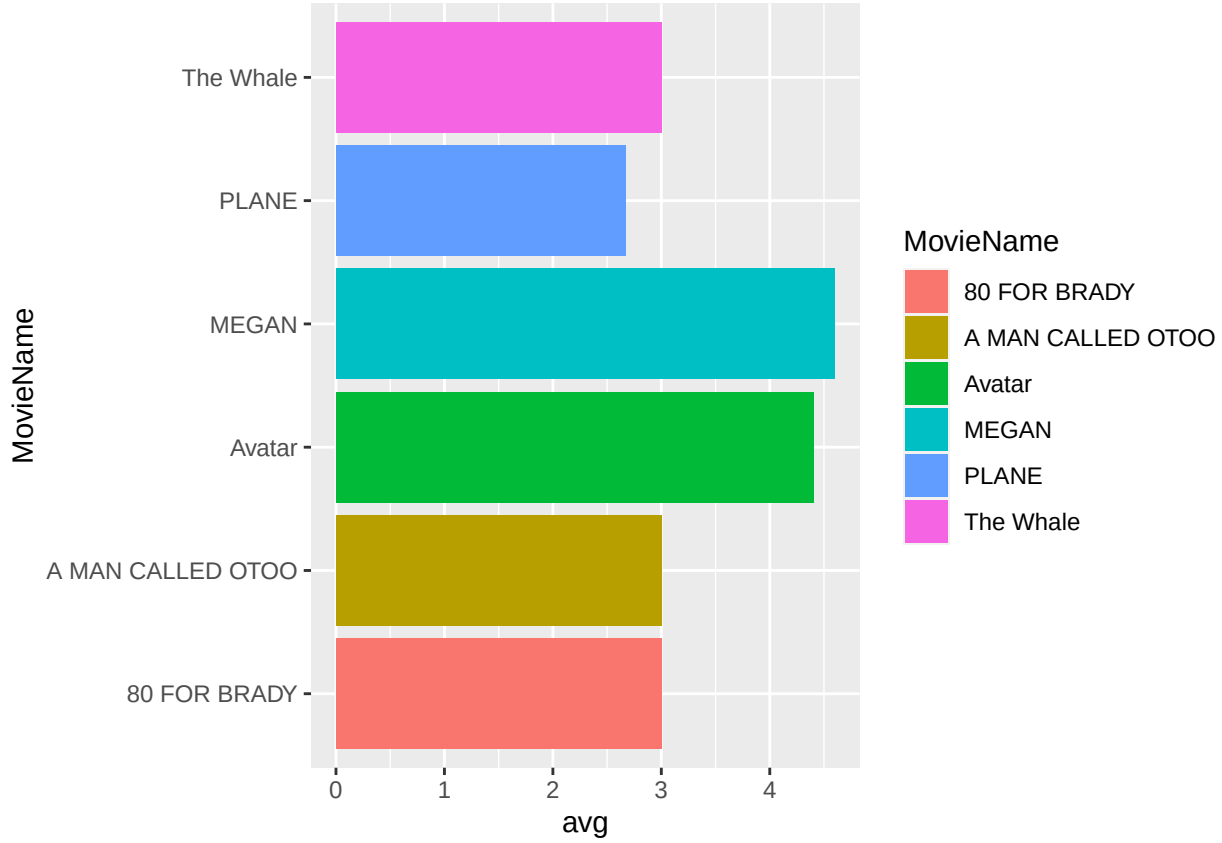
Null values were filtered out from Ratings

```
avgerage_Ratings <- Ratingsdata %>%
  group_by(MovieName) %>%
  filter(Ratings != "NULL") %>%
  summarise(avg= mean(as.integer(Ratings)))
avgerage_Ratings
```

```
## # A tibble: 6 x 2
##   MovieName            avg
##   <chr>              <dbl>
## 1 80 FOR BRADY         3
## 2 A MAN CALLED OTOO    3
## 3 Avatar               4.4
## 4 MEGAN                4.6
## 5 PLANE                2.67
## 6 The Whale            3
```

From this bar plot we can easily tell that "MEGAN" has the highest rating among the six movies currently showing in theaters.

```
library(ggplot2)
ggplot(data=avgerage_Ratings, aes(x=avg, y=MovieName , fill=MovieName)) +
  geom_bar(stat="identity")
```



Visualization of how these five friends rate these six movies is shown in the plot below. Null value of each reviewer is replaced by rating= 0 in each plot.

```
library(ggplot2)
data_01 = Ratingsdata
data_01[,4][data_01[,4] == "NULL"] <- 0
ggplot(data=data_01, aes(x=MovieName, y = Ratings, fill = MovieName, label = Ratings))+
    geom_bar(stat="identity")+
    facet_wrap(~FirstName)+
```

```
    ggtitle("Movie Ratings by each Friend")+
    theme(axis.text.x = element_blank(),plot.title = element_text(hjust=0.5),legend.position = "right")
```