title: “607 Homework 1” author: “Oluwakemi Omotunde” date: “September 10, 2016” output: pdf_document

---

```
library(RMySQL)
```

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

```
library(plyr)
library(ggplot2)

movierating.upload<- dbConnect(MySQL(), user="root", password="N!cole09",
    dbname="movies", host="localhost",client.flag=CLIENT_MULTI_STATEMENTS)

dbListTables(movierating.upload)
```

```
## [1] "movie_rating"
```

```
MovieRatings <- dbReadTable(movierating.upload,"movie_rating")
dbWriteTable(movierating.upload,"movie_rating",MovieRatings,overwrite=T)
```

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

```
MovieRatings
```

```
##   RATER_ID BAD_MOMS SUICIDE_SQUAD SUPERMAN_VS_BATMAN ZOOTOPIA DEADPOOL
## 1        1        5            NA                  3        5       NA
## 2        2        2             2                  3        3        5
## 3        3       NA             3                  3        5        5
## 4        4        3             4                  1        4        5
## 5        5        3             1                  1        4       NA
## 6        6       NA            NA                 NA       NA        5
## 7        7        5             2                  2        3        2
## 8        8       NA             5                  5        5       NA
##   CAPTAIN_AMERICA_CIVIL_WAR
## 1                         1
## 2                         5
## 3                         2
## 4                         4
## 5                         5
## 6                         5
## 7                         3
## 8                         5
```

This shows the data frame that was loaded.

```
MovieRatingsName <- rename(MovieRatings, c("BAD_MOMS"="BadMoms", "SUICIDE_SQUAD"="SuicideSquad", "SUPERM
```

This renames some of the columns to make it neater.

```
MovieRatingsNew <- na.omit(MovieRatingsName)
MovieRatingsNew
```

```
##   RATER_ID BadMoms SuicideSquad SupermanVsBatman Zootopia Deadpool
## 2        2       2            2               3        3        5
## 4        4       3            4               1        4        5
## 7        7       5            2               2        3        2
##   CaptainAmericaCivilWar
## 2                      5
## 4                      4
## 7                      3
```

I wanted to do this to eliminate the NA fields since I wanted to do some calculations. Upon running this code, I see that it does not serve the purpose I need it to.

```
mean(MovieRatingsName$BadMoms, na.rm=TRUE)
```

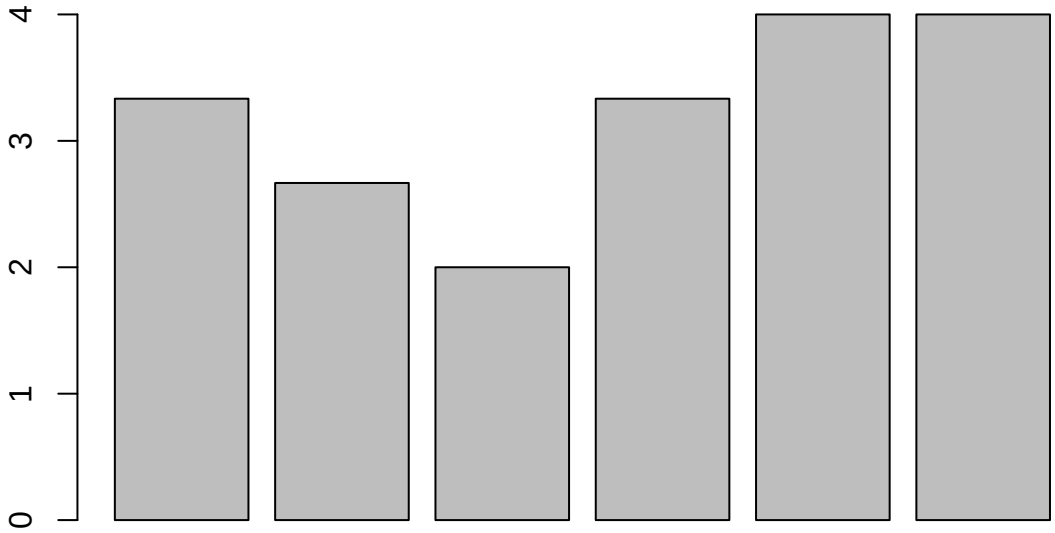
```
## [1] 3.6
```

This removes the NA fields in calculations and returns a numeric value.

To figure out which movie has was favored most, we will plot MovieRatingsNew(no rater who marked NA indicating they haven't seen a particular movie) so that we only consider people who have seen all the movies listed. I believe this will give us more accuarate information.This is working under the assuption that if you've seen all 5 movies, you're a regular movie goer.

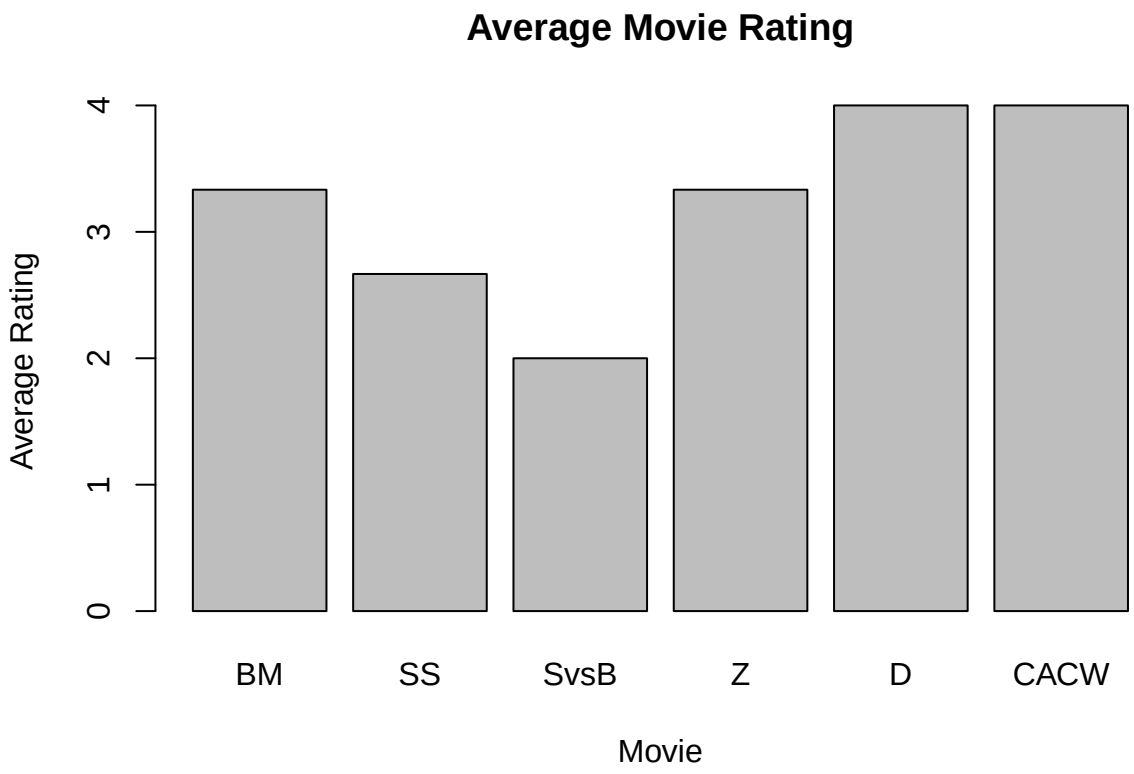
```
meanrating <-c(mean(MovieRatingsNew$BadMoms), mean(MovieRatingsNew$SuicideSquad), mean(MovieRatingsNew$S

barplot(meanrating)
```



#From here, I would like to make the barplot more meaningful.

```
barplot(meanrating, main= "Average Movie Rating", xlab="Movie", ylab="Average Rating",names.arg=c("BM",
```



#with the lableing completed, we can see that Deadpool and Captain America have the highest ratings. To check:

```
mean(MovieRatingsNew$Deadpool)
```

```
## [1] 4
```

this is shown on the barplot

```
mean(MovieRatingsNew$CaptainAmericaCivilWar)
```

```
## [1] 4
```

also shown on the barplot