

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

```
library(RMySQL) 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) MovieRatings <- dbReadTable(movierating.upload,"movie_rating") db-  
WriteTable(movierating.upload,"movie_rating",Movie.Ratings,overwrite=T)
```

```
MovieRatings #This shows the data frame that was loaded.
```

```
MovieRatingsName <- rename(MovieRatings, c("BAD_MOMS"="BadMoms", "SUICIDE_SQUAD"="SuicideSquad",  
"SUPERMAN_VS_BATMAN"="SupermanVsBatman", "CAPTAIN_AMERICA_CIVIL_WAR"="CaptainAmericaCivilWar",  
"ZOOTOPIA"="Zootopia", "DEADPOOL"="Deadpool")) #This renames some of the columns to make it  
neater.
```

```
MovieRatingsNew <- na.omit(MovieRatingsName) MovieRatingsNew #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) #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 accurate information. This is working under the assumption that if you've seen all 5 movies, you're a regular movie goer.

```
meanrating <-c(mean(MovieRatingsNewBadMoms), mean(MovieRatingsNewSuicideSquad), mean(MovieRatingsNewSupermanVsBatman),  
mean(MovieRatingsNewDeadpool), mean(MovieRatingsNewCaptainAmericaCivilWar))
```

```
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",  
"SS", "SvsB", "Z", "D", "CACW")) #with the labeling completed, we can see that Deadpool and Captain  
America have the highest ratings. To check:
```

```
mean(MovieRatingsNew$Deadpool) #this is shown on the barplot
```

```
mean(MovieRatingsNew$CaptainAmericaCivilWar) #also shown on the barplot
```