Do animated movies earn higher/lower ratings than non-animated movies

Paresh Pardeshi
School of Information Technology
Illinois State University
ppardes@ilstu.edu

April 16, 2015

1]Introduction

Rotten Tomatoes is a website launched in 1998 devoted to film reviews and news; it is widely known as a film review aggregator. Coverage now includes TV content as well.

This paper focus on the comarparison between animated movies and the Non-animated movies. which kind of moives earn higher or lower rating if they compare to each other.

The statistics include different subcategories such as science fiction movies. Tables and graphs will be used as a visual description of the results gathered from analyzing the data.

Rotten Tomatoes API The Rotten Tomaoes API (Application Program Interface) provides limited access to critic and audience ratings and reviews, allowing developers to incoporate Rotten Tomatoes data on other websites. The free service is intended for use in the US only; permission is required for use elsewhere.

2]Data

This section will conduct an in-depth exploration of the data, and will display detailed information about the dataset. Data will need to be obtained before being analyzed. Data was downloaded from RottonTommatoes as a csv format, then was imported to R Studio and stored into two different data frame called scifimovies and animmovies. Not all the data was needed. Therefore, further cleaning was required on the dataset.

Only The Ratings and number of reviews of animated movies and the other movies were needed. after getting the relevant information, I have created the tables for animated movies and for the Non-animated movies in which I have selected a particular category which is science fiction.

Code Chunk 1: Procedures of obtaining data, storing data, and cleaning data

```
# we will be using ggplot to plot the graph
library(ggplot2)


# showing the particular table number for the category which is animated movies
tabno <-3;
# import the data from website
add <- rvest::html("http://www.rottentomatoes.com/top/bestofrt/top_100_science_fiction__fant
# select the table for science fiction movies
scifi <- rvest::html_table(rvest::html_nodes(add, "table")[[tabno]])
# putting the data into data frame
scifimovies <-data.frame(write.csv(scifi))

## "","Rank","RatingTomatometer","Title","No. of Reviews"
## "1",1,"99%","The Wizard of Oz (1939)",107
## "2",2,"99%","Metropolis (1927)",114
## "3",3,"97%","Nosferatu, a Symphony of Horror (Nosferatu, eine Symphonie des Grauens) (Nos
## "4",4,"97%","Gravity (2013)",305
## "5",5,"98%","E.T. The Extra-Terrestrial (1982)",98
## "6",6,"100%","Toy Story 2 (1999)",163
## "7",7,"98%","Snow White and the Seven Dwarfs (1937)",43
## "8",8,"100%","The Bride of Frankenstein (1935)",41
## "9",9,"99%","Toy Story 3 (2010)",280
## "10",10,"99%","Dr. Strangelove Or How I Learned to Stop Worrying and Love the Bomb (1964)
## "11",11,"97%","Alien (1979)",92
## "12",12,"100%","Frankenstein (1931)",44
## "13",13,"94%","The Dark Knight (2008)",315
## "14",14,"100%","Pinocchio (1940)",41
## "15",15,"95%","Star Trek (2009)",328
## "16",16,"100%","Toy Story (1995)",78
## "17",17,"96%","Harry Potter and the Deathly Hallows - Part 2 (2011)",306
## "18",18,"96%","WALL-E (2008)",243
## "19",19,"94%","It's a Wonderful Life (1946)",68
## "20",20,"100%","Mary Poppins (1964)",44
## "21",21,"96%","Miracle on 34th Street (1947)",46
## "22",22,"100%","The Terminator (1984)",51
## "23",23,"98%","Invasion of the Body Snatchers (1956)",48
## "24",24,"96%","Star Wars: Episode V - The Empire Strikes Back (1980)",75
## "25",25,"95%","Beauty and The Beast (La Belle et la bte) (1946)",56
## "26",26,"98%","Aliens (1986)",61
## "27",27,"98%","The Discreet Charm Of The Bourgeoisie (Le Charme Discret de la Bourgeoisie
## "28",28,"98%","How to Train Your Dragon (2010)",200
## "29",29,"92%","Marvel's The Avengers (2012)",311
```

```
## "30",30,"96%","2001: A Space Odyssey (1968)",70
## "31",31,"98%","Forbidden Planet (1956)",42
## "32",32,"96%","Fantasia (1940)",50
## "33",33,"96%","The Lord of the Rings: The Two Towers (2002)",247
## "34",34,"97%","The Iron Giant (1999)",130
## "35",35,"97%","Big (1988)",70
## "36",36,"98%","8 1/2 (1963)",41
## "37",37,"97%","Spirited Away (2001)",177
## "38",38,"94%","Her (2013)",229
## "39",39,"95%","The Lord of the Rings: The Return of the King (2003)",261
## "40",40,"95%","Pan's Labyrinth (2006)",224
## "41",41,"97%","Ghostbusters (1984)",64
## "42",42,"97%","The Princess Bride (1987)",63
## "43",43,"94%","The Day the Earth Stood Still (1951)",53
## "44",44,"96%","Monsters, Inc. (2001)",192
## "45",45,"93%","Iron Man (2008)",243
## "46",46,"93%","Gojira (1956)",69
## "47",47,"93%","Looper (2012)",246
## "48",48,"95%","Snowpiercer (2014)",185
## "49",49,"98%","Wings of Desire (1987)",46
## "50",50,"96%","Back to the Future (1985)",68
## "51",51,"98%","Jodorowsky's Dune (2014)",109
## "52",52,"97%","Babe (1995)",68
## "53",53,"98%","Brazil (1985)",46
## "54",54,"98%","Repo Man (1984)",45
## "55",55,"93%","Jurassic Park (1993)",112
## "56",56,"94%","Spider-Man 2 (2004)",253
## "57",57,"96%","Invasion of the Body Snatchers (1978)",46
## "58",58,"93%","Star Wars: Episode IV - A New Hope (1977)",71
## "59",59,"96%","Solaris (1976)",47
## "60",60,"91%","Blade Runner (1982)",96
## "61",61,"96%","Close Encounters of the Third Kind (1977)",46
## "62",62,"96%","Groundhog Day (1993)",70
## "63",63,"91%","Guardians of the Galaxy (2014)",259
## "64",64,"91%","X-Men: Days of Future Past (2014)",255
## "65",65,"98%","The Secret of Roan Inish (1995)",43
## "66",66,"96%","The Evil Dead (1981)",53
## "67",67,"96%","Yellow Submarine (1968)",47
## "68",68,"94%","The Nightmare Before Christmas (1993)",88
## "69",69,"90%","Live Die Repeat: Edge of Tomorrow (2014)",267
## "70",70,"94%","Young Frankenstein (1974)",51
## "71",71,"93%","Superman (1978)",60
## "72",72,"92%","Children of Men (2006)",237
## "73",73,"95%","Antz (1998)",88
## "74",74,"90%","Dawn Of The Planet Of The Apes (2014)",257
```

```
## "75",75,"95%","L'anne dernire  Marienbad (Last Year at Marienbad) (1961)",42
## "76",76,"95%","The Secret World of Arrietty (2012)",139
## "77",77,"98%","Song Of The Sea (2014)",57
## "78",78,"93%","Eternal Sunshine Of The Spotless Mind (2004)",232
## "79",79,"92%","The Seventh Seal (Det Sjunde inseglet) (1957)",52
## "80",80,"97%","Ernest & Clestine (2014)",75
## "81",81,"92%","Source Code (2011)",243
## "82",82,"98%","Marwencol (2010)",58
## "83",83,"95%","Mad Max (1979)",43
## "84",84,"90%","District 9 (2009)",275
## "85",85,"89%","Captain America: The Winter Soldier (2014)",254
## "86",86,"93%","Being John Malkovich (1999)",123
## "87",87,"91%","Bambi (1942)",46
## "88",88,"91%","The Lord of the Rings: The Fellowship of the Ring (2001)",225
## "89",89,"89%","The Hunger Games: Catching Fire (2013)",247
## "90",90,"91%","Harry Potter and the Prisoner of Azkaban (2004)",249
## "91",91,"90%","Coraline (2009)",256
## "92",92,"92%","How to Train Your Dragon 2 (2014)",159
## "93",93,"94%","Re-Animator (1985)",47
## "94",94,"93%","Jason and the Argonauts (1963)",40
## "95",95,"92%","Shaun of the Dead (2004)",200
## "96",96,"92%","Men in Black (1997)",84
## "97",97,"92%","The Little Mermaid (1989)",65
## "98",98,"94%","Heavenly Creatures (1994)",48
## "99",99,"92%","Terminator 2: Judgment Day (1991)",62
## "100",100,"90%","Minority Report (2002)",239
```

```
# showing the particular table number for the category which is animated movies
tabnum <-3;
# import the data from website
add <- rvest::html("http://www.rottentomatoes.com/top/bestofrt/?category=2")
# select the table for animated movies
anim <- rvest::html_table(rvest::html_nodes(add, "table")[[tabnum]])
# putting the data into data.frame
animmovies <-write.csv(anim)
```

```
## "","Rank","RatingTomatometer","Title","No. of Reviews"
## "1",1,"100%","Toy Story 2 (1999)",163
## "2",2,"98%","Snow White and the Seven Dwarfs (1937)",43
## "3",3,"99%","Toy Story 3 (2010)",280
## "4",4,"98%","Up (2009)",280
## "5",5,"99%","Finding Nemo (2003)",238
## "6",6,"100%","Pinocchio (1940)",41
## "7",7,"100%","Toy Story (1995)",78
## "8",8,"96%","WALL-E (2008)",243
## "9",9,"96%","The LEGO Movie (2014)",221
```

```
## "10",10,"98%","How to Train Your Dragon (2010)",200
## "11",11,"97%","The Incredibles (2004)",231
## "12",12,"96%","Ratatouille (2007)",230
## "13",13,"98%","101 Dalmatians (1961)",45
## "14",14,"96%","Fantasia (1940)",50
## "15",15,"98%","Who Framed Roger Rabbit (1988)",60
## "16",16,"97%","The Iron Giant (1999)",130
## "17",17,"97%","Chicken Run (2000)",165
## "18",18,"97%","Spirited Away (2001)",177
## "19",19,"96%","Monsters, Inc. (2001)",192
## "20",20,"96%","Waltz with Bashir (2008)",144
## "21",21,"96%","Persepolis (2007)",153
## "22",22,"95%","Wallace & Gromit: The Curse of the Were-Rabbit (2005)",176
## "23",23,"96%","Yellow Submarine (1968)",47
## "24",24,"94%","The Nightmare Before Christmas (1993)",88
## "25",25,"95%","Antz (1998)",88
## "26",26,"95%","The Secret World of Arrietty (2012)",139
## "27",27,"98%","Song Of The Sea (2014)",57
## "28",28,"97%","Ernest & Clestine (2014)",75
## "29",29,"92%","Fantastic Mr. Fox (2009)",224
## "30",30,"94%","Aladdin (1992)",66
## "31",31,"91%","Bambi (1942)",46
## "32",32,"90%","Coraline (2009)",256
## "33",33,"92%","How to Train Your Dragon 2 (2014)",159
## "34",34,"94%","The Triplets of Belleville (2003)",140
## "35",35,"93%","James and the Giant Peach (1996)",72
## "36",36,"92%","The Little Mermaid (1989)",65
## "37",37,"89%","Frozen (2013)",214
## "38",38,"92%","Ponyo (2009)",159
## "39",39,"92%","A Bug's Life (1998)",84
## "40",40,"92%","Princess Mononoke (1999)",105
## "41",41,"90%","The Simpsons Movie (2007)",199
## "42",42,"92%","Arthur Christmas (2011)",157
## "43",43,"95%","Mary and Max (2009)",61
## "44",44,"93%","My Neighbor Totoro (1988)",43
## "45",45,"93%","Beauty and the Beast (2012)",103
## "46",46,"90%","Tangled (2010)",215
## "47",47,"89%","Big Hero 6 (2014)",186
## "48",48,"89%","Bolt (2008)",178
## "49",49,"89%","The Wind Rises (2014)",157
## "50",50,"88%","Shrek 2 (2004)",232
## "51",51,"88%","Rango (2011)",212
## "52",52,"87%","Frankenweenie (2012)",198
## "53",53,"87%","Kung Fu Panda (2008)",177
## "54",54,"90%","Winnie the Pooh (2011)",126
```

```
## "55",55,"87%","Shrek (2001)",190
## "56",56,"92%","Millennium Actress (Sennen joy) (2001)",51
## "57",57,"90%","The Illusionist (L'illusionniste) (2010)",120
## "58",58,"90%","The Lion King (2011)",110
## "59",59,"91%","The Secret of Kells (2010)",79
## "60",60,"87%","ParaNorman (2012)",163
## "61",61,"88%","Tarzan (1999)",104
## "62",62,"87%","Howl's Moving Castle (2005)",172
## "63",63,"91%","Tokyo Godfathers (2003)",64
## "64",64,"91%","My Dog Tulip (2010)",57
## "65",65,"86%","Wreck-it Ralph (2012)",168
## "66",66,"87%","Cloudy With a Chance of Meatballs (2009)",139
## "67",67,"86%","The Pirates! Band of Misfits (2012)",144
## "68",68,"85%","The Princess and the Frog (2009)",184
## "69",69,"86%","Mulan (1998)",73
## "70",70,"86%","Lilo & Stitch (2002)",145
## "71",71,"85%","The Emperor's New Groove (2000)",127
## "72",72,"83%","Tim Burton's Corpse Bride (2005)",187
## "73",73,"86%","Anastasia (1997)",51
## "74",74,"87%","Metropolis (2002)",60
## "75",75,"84%","Puss in Boots (2011)",142
## "76",76,"87%","Akira (2001)",45
## "77",77,"86%","Chico & Rita (2012)",65
## "78",78,"84%","Paprika (2006)",85
## "79",79,"81%","Despicable Me (2010)",192
## "80",80,"83%","Hercules (1997)",52
## "81",81,"78%","Brave (2012)",225
## "82",82,"81%","Kung Fu Panda 2 (2011)",167
## "83",83,"83%","From Up On Poppy Hill (2013)",84
## "84",84,"82%","The Book of Life (2014)",105
## "85",85,"84%","Panique au village (A Town Called Panic) (2009)",67
## "86",86,"81%","South Park: Bigger, Longer & Uncut (1999)",94
## "87",87,"78%","Monsters University (2013)",182
## "88",88,"82%","Fantasia 2000 (2000)",82
## "89",89,"80%","Waking Life (2001)",141
## "90",90,"79%","The Prince of Egypt (1998)",84
## "91",91,"82%","A Cat in Paris (2012)",61
## "92",92,"79%","Dr. Seuss' Horton Hears a Who! (2008)",131
## "93",93,"79%","Madagascar 3: Europe's Most Wanted (2012)",129
## "94",94,"79%","Mr. Peabody & Sherman (2014)",120
## "95",95,"80%","Chicago 10 (2008)",82
## "96",96,"78%","Charlotte's Web (2006)",143
## "97",97,"78%","Surf's Up (2007)",143
## "98",98,"80%","The Wild Thornberrys Movie (2002)",88
## "99",99,"78%","The SpongeBob Movie: Sponge Out of Water (2015)",87
```

```
## "100",100,"77%","Ice Age (2002)",164
```

Code Chunk 2: Display the data type of data set

```
class(scifi)

## [1] "data.frame"

class(anim)

## [1] "data.frame"
```

Code chunk 3: displaying the content of the tables (number of observations and variables along with the columns discription)

```
str(scifi)

## 'data.frame': 100 obs. of  4 variables:
##  $ Rank             : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ RatingTomatometer: chr  "99%" "99%" "97%" "97%" ...
##  $ Title            : chr  "The Wizard of Oz (1939)" "Metropolis (1927)" "Nosferatu, a Sy
##  $ No. of Reviews   : int  107 114 62 305 98 163 43 41 280 68 ...

str(anim)

## 'data.frame': 100 obs. of  4 variables:
##  $ Rank             : num  1 2 3 4 5 6 7 8 9 10 ...
##  $ RatingTomatometer: chr  "100%" "98%" "99%" "98%" ...
##  $ Title            : chr  "Toy Story 2 (1999)" "Snow White and the Seven Dwarfs (1937)"
##  $ No. of Reviews   : int  163 43 280 280 238 41 78 243 221 200 ...
```

Code chunk 4: Summary of data sets

```
#Display the summary of the whole data set
summary(scifi)

##       Rank        RatingTomatometer     Title           No. of Reviews
##  Min.   :  1.00   Length:100         Length:100         Min.   : 40.00
##  1st Qu.: 25.75   Class :character   Class :character   1st Qu.: 50.75
##  Median : 50.50   Mode  :character   Mode  :character   Median : 75.00
##  Mean   : 50.50                                        Mean   :127.78
##  3rd Qu.: 75.25                                        3rd Qu.:229.75
##  Max.   :100.00                                        Max.   :328.00

summary(anim)

##       Rank        RatingTomatometer     Title           No. of Reviews
```

```
##  Min.   :  1.00   Length:100         Length:100         Min.   : 41.00
##  1st Qu.: 25.75   Class :character   Class :character   1st Qu.: 77.25
##  Median : 50.50   Mode  :character   Mode  :character   Median :135.00
##  Mean   : 50.50                                        Mean   :132.16
##  3rd Qu.: 75.25                                        3rd Qu.:177.25
##  Max.   :100.00                                        Max.   :280.00
```

3]Data Defination

As we have gather the relevant data, now we need to clean and operate the data in order to get the result that we are looking for. we have tables for scifi movies and animated movies. In the column RatingTomatometer, there are some NA values that we need to remove from column. to do this we are using na.omit function.

Code Chunk 5: Extracting the NA values from the column RatingTomatometer of the table

```
#after removing the NA values, data is moving into data frame sc
sc<-na.omit(scifi$RatingTomatometer)
head(sc)

## [1] "99%"  "99%"  "97%"  "97%"  "98%"  "100%"

#after removing the NA values, data is moving into data frame sc
an<-na.omit(anim$RatingTomatometer)
head(an)

## [1] "100%" "98%"  "99%"  "98%"  "99%"  "100%"
```

Code Chunk 6: Display the data type and structure of scifi movies ratings

```
#Display the data type of the Rating data
class(scifi$RatingTomatometer)

## [1] "character"

# Display the structure of the Rating data
str(scifi$RatingTomatometer)

##  chr [1:100] "99%" "99%" "97%" "97%" "98%" "100%" "98%" ...
```

Code Chunk 7: Display the data type and structure of animated movies ratings

```r
#Display the data type of the Rating data
class(anim$RatingTomatometer)

## [1] "character"

# Display the structure of the Rating data
str(anim$RatingTomatometer)

##  chr [1:100] "100%" "98%" "99%" "98%" "99%" "100%" ...
```

Code Chunk 8: Removing the percentage sign in order to calculate the mean of the percentage column. To do that we used gsub command.

```r
#after extracting the percentage sign ,the values are going into x data frame
x <- as.numeric(gsub("%", "", sc))
head(x)

## [1]  99  99  97  97  98 100

#after extracting the percentage sign ,the values are going into y data frame
y <- as.numeric(gsub("%", "", an))
head(y)

## [1] 100  98  99  98  99 100
```

Code chunk 9: Calculate the Mean of ratings column.

To do that we used mean command and unlist command unlist simplifies it to produce a vector which contains all the atomic components which occur in x and y.

```r
#
x1<-mean(as.numeric(unlist(x), na.rm=T))

y1<-mean(as.numeric(unlist(y), na.rm=T))
```

4]Result

This section will display the result after analyzing the data.

Code Chunk 10: Display the mean of columns

```r
# display mean value of percentage rating for scifi movies
x1

## [1] 95.16
```

```r
# display mean value of percentage rating for animated movies
y1
```

```
## [1] 89.35
```

Code Chunk 11: plotting the graph

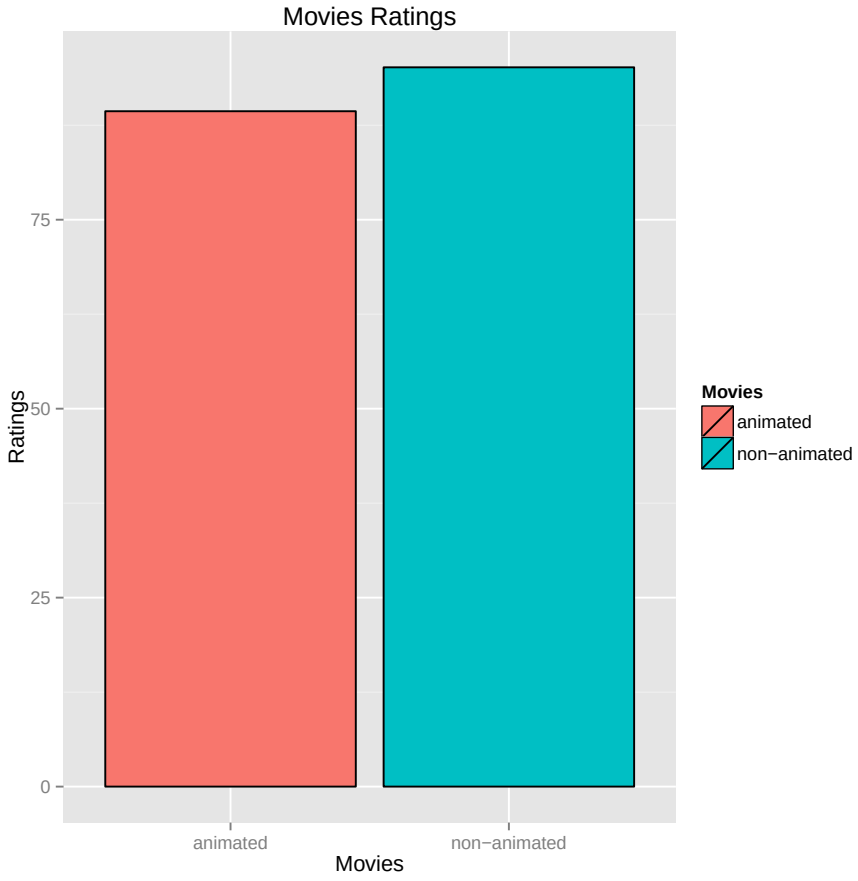
```r
Movies <-c("non-animated", "animated")
Movies
```

```
## [1] "non-animated" "animated"
```

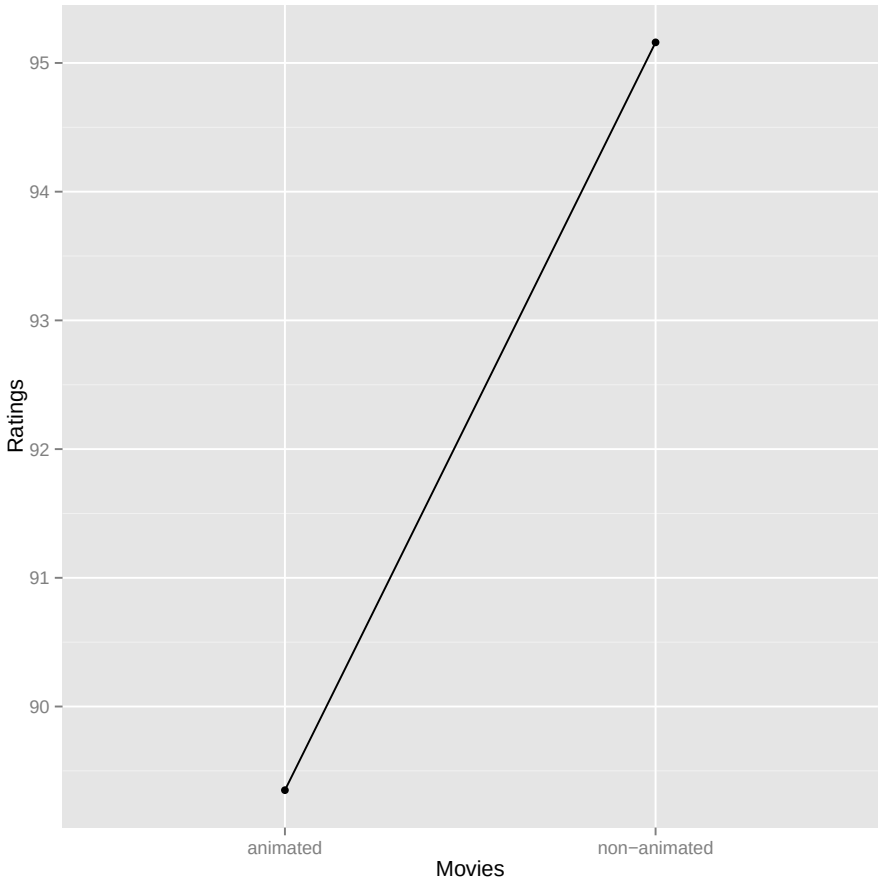
```r
Ratings<-c(x1,y1)
Ratings
```

```
## [1] 95.16 89.35
```

```r
df = data.frame(Movies, Ratings)
ggplot(data=df, aes(x=Movies, y=Ratings,fill=Movies)) + geom_bar(colour="black",stat="identi
```




```
# Line Graph Shows comparison between animated and non animated movies
ggplot(data=df, aes(x=Movies, y=Ratings, group=1)) +
    geom_line() +
    geom_point()
```


95
94
93
92
91
90
Ratings
animated
non−animated
Movies