Name: **Ademola Apata**

Title: **Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.**

Movie Summary and first several rows in a Dataframe

A summary of Movies and Ratings is given below, along with the first several rows of the movie and ratings dataframe.

```
##     movieId          title              genres
##  Min.   :     1   Length:10329       Length:10329
##  1st Qu.:  3240   Class :character   Class :character
##  Median :  7088   Mode  :character   Mode  :character
##  Mean   : 31924
##  3rd Qu.: 59900
##  Max.   :149532

##   movieId                               title
## 1       1                    Toy Story (1995)
## 2       2                      Jumanji (1995)
## 3       3             Grumpier Old Men (1995)
## 4       4            Waiting to Exhale (1995)
## 5       5 Father of the Bride Part II (1995)
## 6       6                           Heat (1995)
##                                         genres
## 1 Adventure|Animation|Children|Comedy|Fantasy
## 2                  Adventure|Children|Fantasy
## 3                              Comedy|Romance
## 4                        Comedy|Drama|Romance
## 5                                      Comedy
## 6                       Action|Crime|Thriller

##      userId         movieId           rating        timestamp
##  Min.   :  1.0   Min.   :     1   Min.   :0.500   Min.   :8.286e+08
##  1st Qu.:192.0   1st Qu.:  1073   1st Qu.:3.000   1st Qu.:9.711e+08
##  Median :383.0   Median :  2497   Median :3.500   Median :1.115e+09
##  Mean   :364.9   Mean   : 13381   Mean   :3.517   Mean   :1.130e+09
##  3rd Qu.:557.0   3rd Qu.:  5991   3rd Qu.:4.000   3rd Qu.:1.275e+09
##  Max.   :668.0   Max.   :149532   Max.   :5.000   Max.   :1.452e+09

##   userId movieId rating  timestamp
## 1      1      16    4.0 1217897793
## 2      1      24    1.5 1217895807
## 3      1      32    4.0 1217896246
## 4      1      47    4.0 1217896556
## 5      1      50    4.0 1217896523
## 6      1     110    4.0 1217896150
```

Name: **Ademola Apata**

Title: **Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.**

Matrix List of Genres for Each Movie

Below are Tail value of Matrix Output summary statistics of matrix genre dataframe which indicates the data type “Int” and example of each genre type.

```
##       Action Adventure Animation Children Comedy Crime Documentary Drama
## 10324      0         0         0        0      0     0           0     1
## 10325      0         0         1        1      1     0           0     0
## 10326      0         0         0        0      1     0           0     0
## 10327      0         0         0        0      1     0           0     0
## 10328      0         0         0        0      0     0           0     1
## 10329      0         0         0        0      0     0           0     0
##       Fantasy Film-Noir Horror Musical Mystery Romance Sci-Fi Thriller War
## 10324       0         0      0       0       0       0      0        0   0
## 10325       0         0      0       0       0       0      0        0   0
## 10326       0         0      0       0       0       0      0        0   0
## 10327       0         0      0       0       0       0      0        0   0
## 10328       0         0      0       0       0       0      0        0   0
## 10329       0         0      0       0       0       0      0        0   0
##       Western
## 10324       0
## 10325       0
## 10326       0
## 10327       0
## 10328       0
## 10329       0

## 'data.frame':    10329 obs. of  18 variables:
##  $ Action     : int  0 0 0 0 0 1 0 0 1 1 ...
##  $ Adventure  : int  1 1 0 0 0 0 0 1 0 1 ...
##  $ Animation  : int  1 0 0 0 0 0 0 0 0 0 ...
##  $ Children   : int  1 1 0 0 0 0 0 1 0 0 ...
##  $ Comedy     : int  1 0 1 1 1 0 1 0 0 0 ...
##  $ Crime      : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Documentary: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Drama      : int  0 0 0 1 0 0 0 0 0 0 ...
##  $ Fantasy    : int  1 1 0 0 0 0 0 0 0 0 ...
##  $ Film-Noir  : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Horror     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Musical    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Mystery    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Romance    : int  0 0 1 1 0 0 1 0 0 0 ...
##  $ Sci-Fi     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Thriller   : int  0 0 0 0 0 1 0 0 0 1 ...
##  $ War        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Western    : int  0 0 0 0 0 0 0 0 0 0 ...
```

Name: **Ademola Apata**

Title: **Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.**

Summary Statistic of Search Matrix result by genre

Below is a summary statistic of search matrix output allowing film search by specifying the genre present on the list.

```
##   movieId                               title Action Adventure Animation
## 1       1                    Toy Story (1995)      0         1         1
## 2       2                      Jumanji (1995)      0         1         0
## 3       3             Grumpier Old Men (1995)      0         0         0
## 4       4            Waiting to Exhale (1995)      0         0         0
## 5       5 Father of the Bride Part II (1995)      0         0         0
## 6       6                           Heat (1995)      1         0         0
##   Children Comedy Crime Documentary Drama Fantasy Film-Noir Horror Musical
## 1        1      1     0           0     0       1         0      0       0
## 2        1      0     0           0     0       1         0      0       0
## 3        0      1     0           0     0       0         0      0       0
## 4        0      1     0           0     1       0         0      0       0
## 5        0      1     0           0     0       0         0      0       0
## 6        0      0     1           0     0       0         0      0       0
##   Mystery Romance Sci-Fi Thriller War Western
## 1       0       0      0        0   0       0
## 2       0       0      0        0   0       0
## 3       0       1      0        0   0       0
## 4       0       1      0        0   0       0
## 5       0       0      0        0   0       0
## 6       0       0      0        1   0       0
```

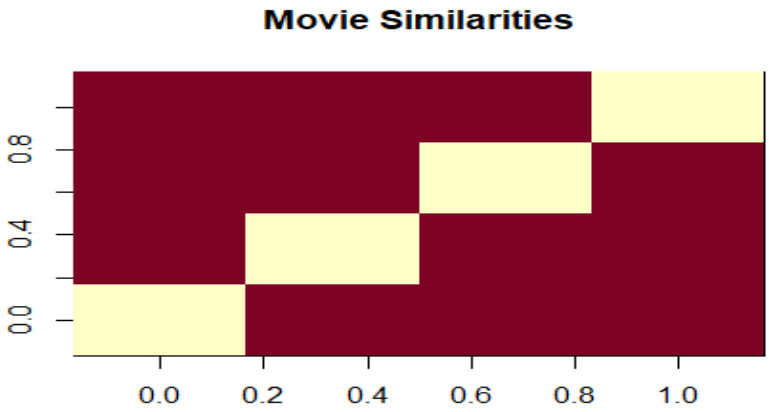
Exploring Similarity Data – Users

According to the Movie similarity results, the rows and columns represents the users, Each cell is comparing the similarities between users, The diagonal is yellow therefore it is comparing each user with itself,

```
##           1         2         3         4
## 1 0.0000000 0.9760860 0.9641723 0.9914398
## 2 0.9760860 0.0000000 0.9925732 0.9374253
## 3 0.9641723 0.9925732 0.0000000 0.9888968
## 4 0.9914398 0.9374253 0.9888968 0.0000000
```

Name: **Ademola Apata**

Title: **Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.**

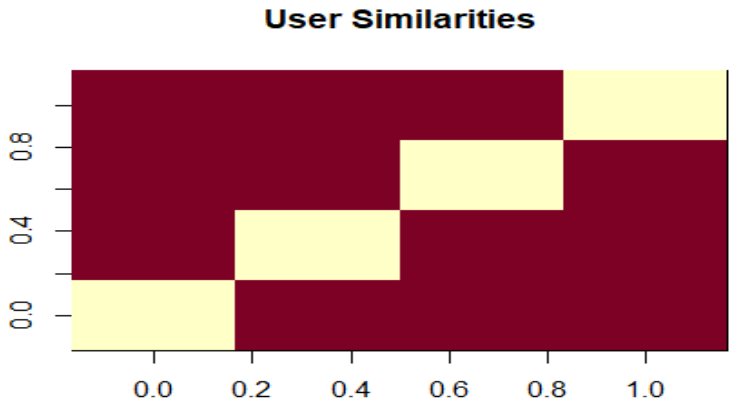


.

Exploring Similarity Data – Films

According to the Movie similarity results, the rows and columns represents the films/movies, each cell is comparing the similarities between films, The diagonal is yellow therefore it is comparing each user with film,

```
##           1         2         3         4
## 1 0.0000000 0.9669732 0.9559341 0.9101276
## 2 0.9669732 0.0000000 0.9658757 0.9412416
## 3 0.9559341 0.9658757 0.0000000 0.9864877
## 4 0.9101276 0.9412416 0.9864877 0.0000000
```



Name: **Ademola Apata**

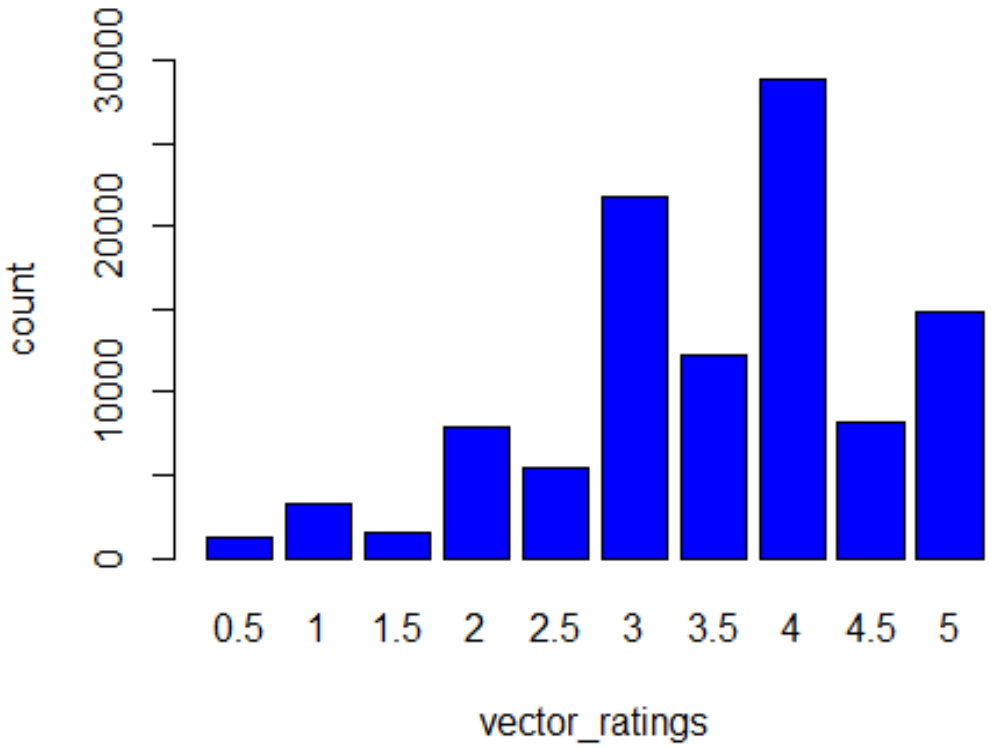
Title: **Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.**

Table and Distribution of the ratings

The table below shows each rating and their respective counts in an organized format

```
## ratingValues
##       0     0.5       1     1.5       2     2.5       3     3.5       4
4.5
## 6791761    1198    3258    1567    7943    5484   21729   12237   28880
8187
##       5
##   14856
```

A Graphical Representation of distribution of the ratings, a rating of 0 represents the absence of a rating or missing value therefore it was removed from plot as shown below.



Name: **Ademola Apata**

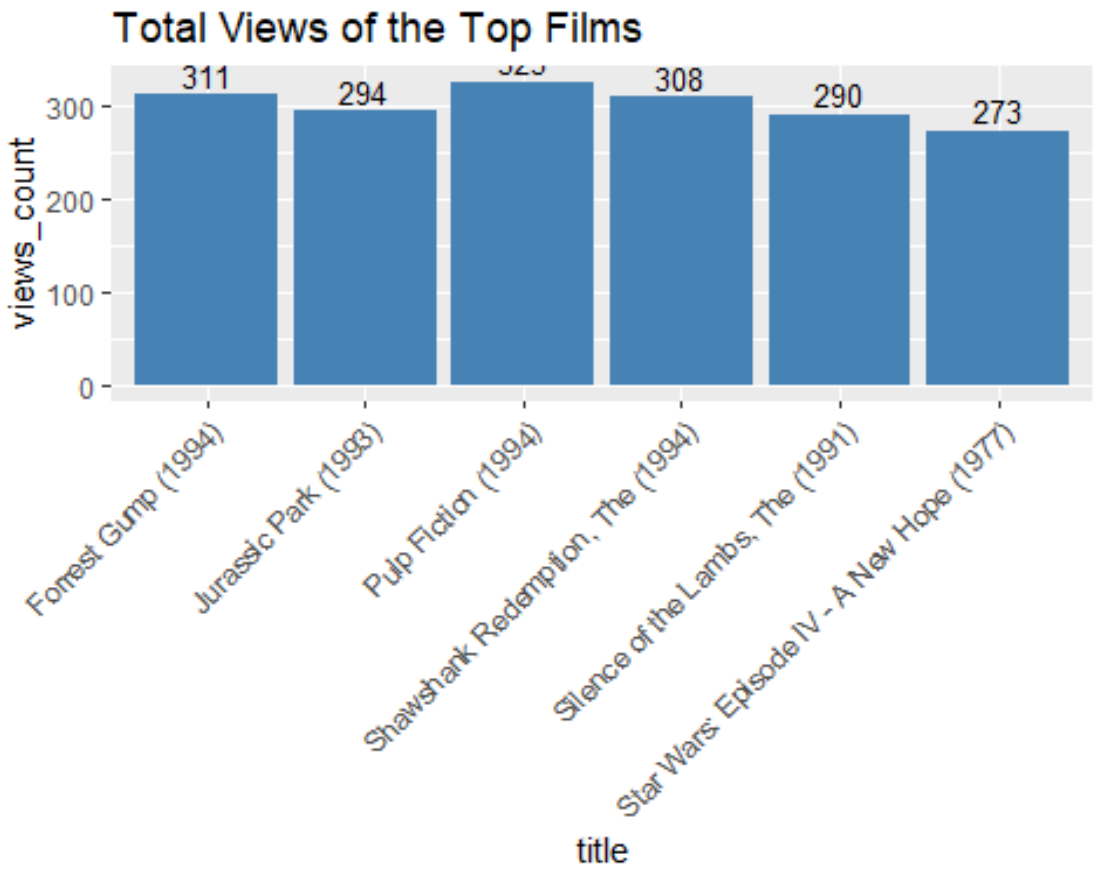
Title: **Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.**

Number and plot of the top movies

Tabular view and bar plot of most viewed movies in the movies dataframe.

```
##     movie views_count                                       title
## 296   296         325                        Pulp Fiction (1994)
## 356   356         311                        Forrest Gump (1994)
## 318   318         308          Shawshank Redemption, The (1994)
## 480   480         294                      Jurassic Park (1993)
## 593   593         290          Silence of the Lambs, The (1991)
## 260   260         273 Star Wars: Episode IV - A New Hope (1977)
```

We see that “Pulp Fiction (1994)” is the most viewed movie, exceeding the second-most-viewed “Forrest Gump (1994)” by 14 views.

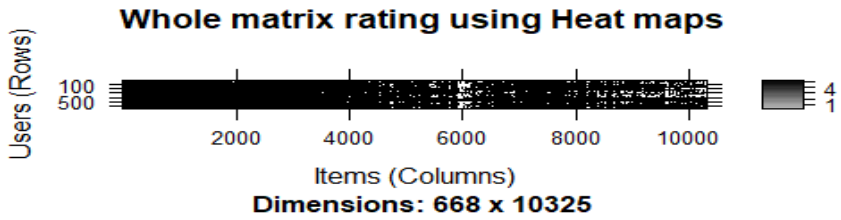


Name: Ademola Apata

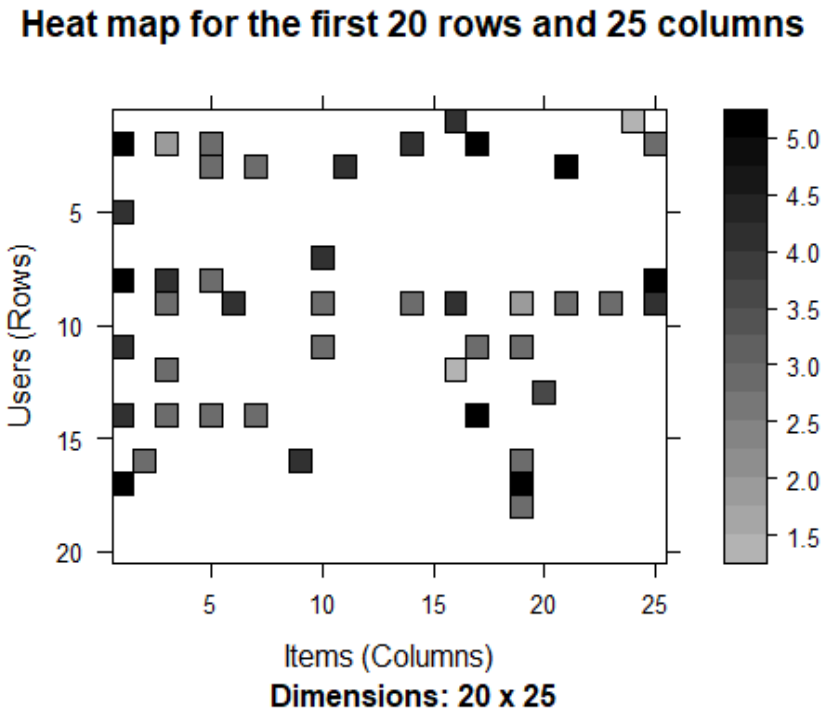
Title: Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.

Heatmap of Movie Ratings

Visualization of heatmaps whose color intensity represents ratings, the rows represents Users and the columns represents movies.



The second table below represents a zoom in on the first 20 users -rows and movies- columns and the color intensity shade representing the intensity of the rating



Name: **Ademola Apata**

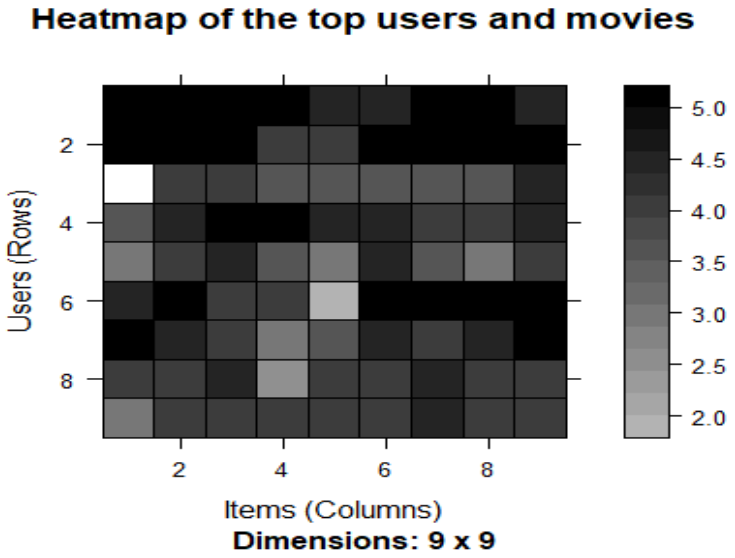
Title: **Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.**

Data Preparation Results

Selecting Useful Data

Upon selection of the most relevant data, the initial minimum threshold of users who have rated a film and the minimum number of views for a film was set as 50.

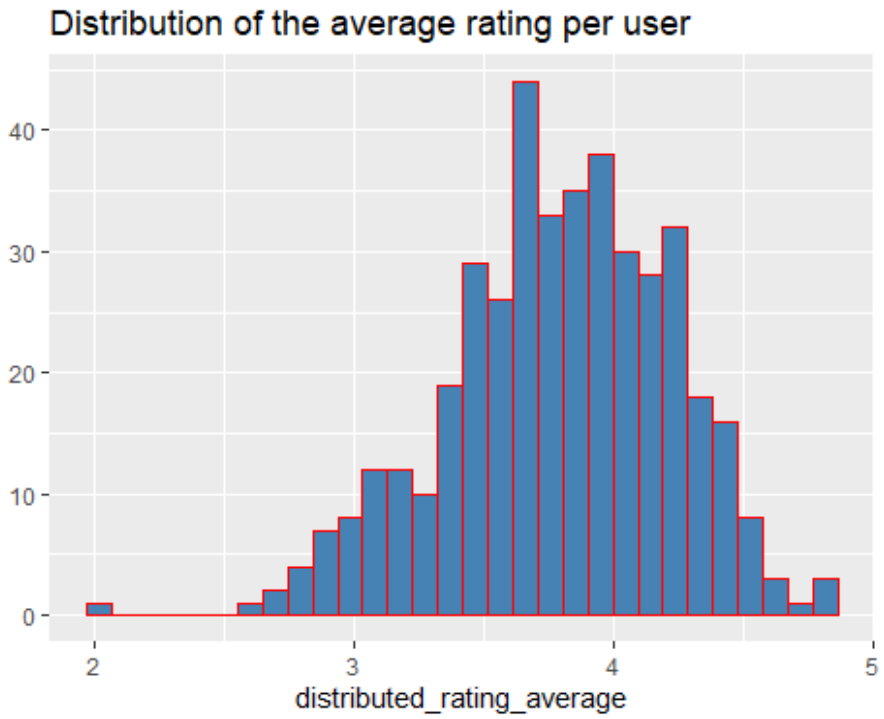
The heat map below shows the Top two percent of users and movies of the relevant data based on the threshold defined above



The plot below indicates the distribution of the average movie ratings, the rankings are between 2 and 4.5. As expected, all extreme values are removed due to a defined threshold of 50, creating a subset of only relevant movies

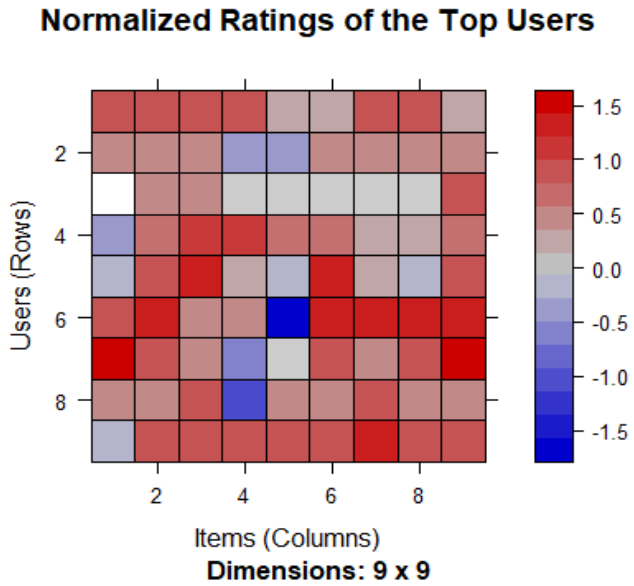
Name: Ademola Apata

Title: Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.



Data Normalization

In order to remove the bias of high and low ratings from users, the data is normalized and the heatmap of top users and movies are shown below. The shades of blue or more red is a result of visualizing only the top movies. The average ratings by users as a result of normalization is 0 as expected.

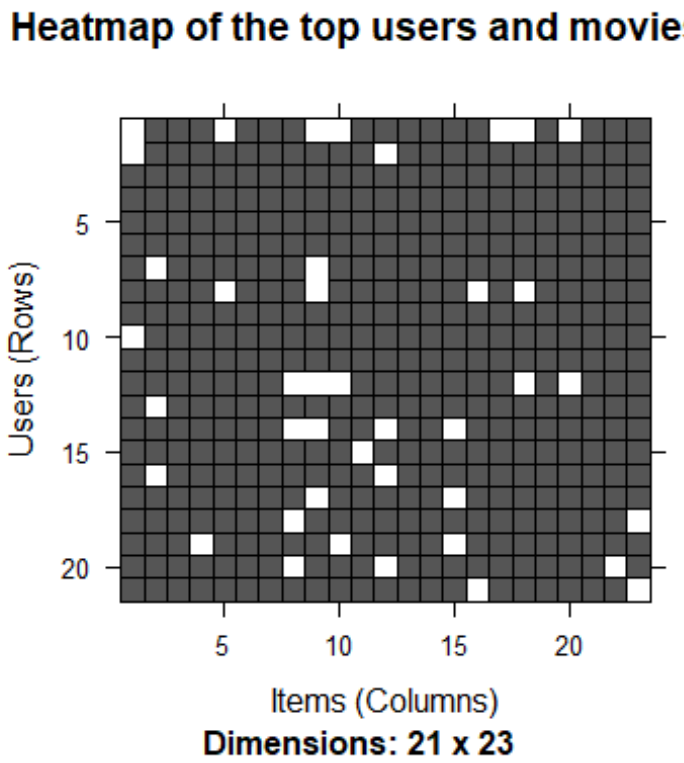


Name: **Ademola Apata**

Title: **Descriptive Statistics of Movie Recommender Initial Results of Code descriptive statistics.**

Performing Data Binarization

According to the Binarized threshold defined in the code applied, we define a matrix equal to 1 if the cell has a rating above the threshold



ITEM-based Collaborative Filtering Model Result for User 1

Below is the results of the recommendations for the first user.

```
##  [1] "Georgia (1995)"
##  [2] "Umbrellas of Cherbourg, The (Parapluies de Cherbourg, Les) (1964)"
##  [3] "Suture (1993)"
##  [4] "Double Happiness (1994)"
##  [5] "Barcelona (1994)"
##  [6] "Naked in New York (1994)"
##  [7] "Love and a .45 (1994)"
##  [8] "It's My Party (1996)"
##  [9] "Apple Dumpling Gang, The (1975)"
## [10] "Secret Agent, The (1996)"
```