**Name: Ahmed Talat Abd Elmohsen Ali Id:221101084**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Aim of project:**

The main goal of this machine learning project is to build a recommendation engine that recommends movies to users. This R project is designed to understand the functioning of a recommendation system. I developed an \*Item Based Collaborative Filter\*. This helped me gain experience of implementing my R, Data Science, and Machine learning skills in a real-life project.

**Dataset used**

I have used the MovieLens Dataset. That data I have used consists of 105339 ratings in the ```ratings.csv``` file, applied over 10329 movies in the ```movies.csv```.

**Essential Libraries**

recommenderlab, ggplot2, data.table , reshape2 and psych

**Data Pre-processing**

After retrieving data from the movies.csv and ratings.csv datasets, I observed that the userId column, as well as the movieId column, consisted of integers. Furthermore, I needed to convert the genres present in the movie\_data dataframe into a more usable format by the users. In order to do so, I first created a one-hot encoding to create a matrix that comprises of corresponding genres for each of the films. I then created a search matrix that will allow us to perform an easy search of the films by specifying the genre present in our list.

There are movies that have several genres. For the movie recommendation system to make sense of the ratings through recommenderlab, I convert the matrix into a sparse matrix. This new matrix is of the class realRatingMatrix. I then overviewed some important parameters that provided various options for building recommendation systems for movies.

Visualisation: Similarity in data

I visualised the similarity between the users as explained in the above section as well as the similarity shared between the films.

Visualisation: Most viewed movies

In this section of the machine learning project, I explored the most viewed movies in the dataset. Before this, I counted the number of views in a film and organized them in a table that would group them in descending order. I visualized the total number of views of the top films as a bar plot.

From the visualisation, it could be observed that 'Pulp Fiction' is the most watched film followed by 'Forrest Gump'.

Visualisation: Heatmap of Movie Ratings

Now, in this data science project of Recommendation system, I visualize a heatmap of the movie ratings. This heatmap will contain first 25 rows

**Data Preparation**

This is conducted in three steps:

1. Selecting useful data

2. Normalizing data

3. Binarizing the data

Building the recommendation system

Now, I explored the various parameters of the \*Item Based Collaborative Filter\*. These parameters are default in nature. In the first step, k denotes the number of items for computing their similarities. Here, k is equal to 30. Therefore, the algorithm will now identify the k most similar items and store their number.

Exploring data science recommendation system model

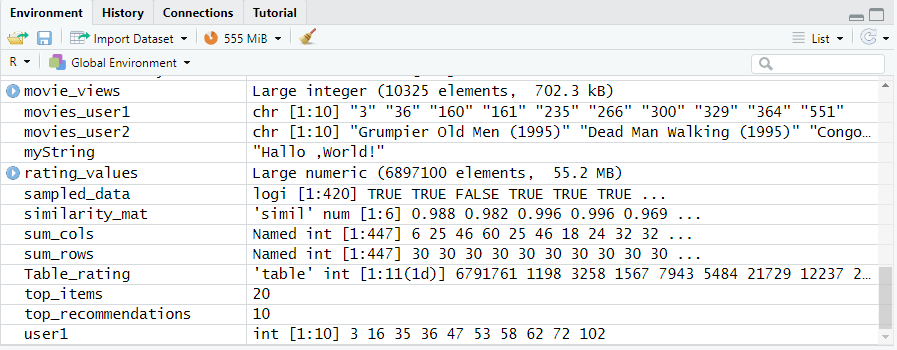
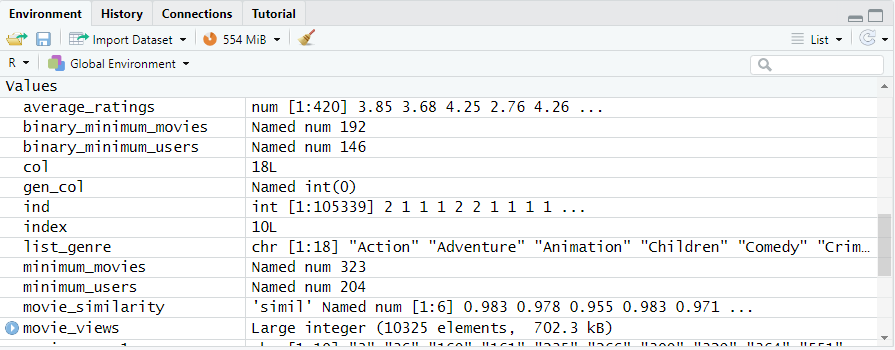
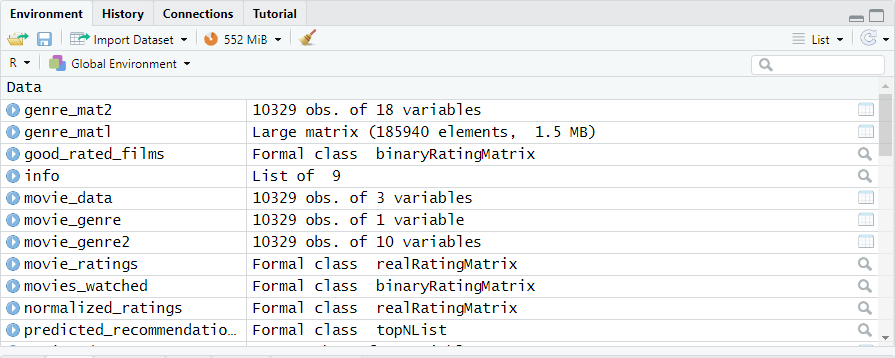
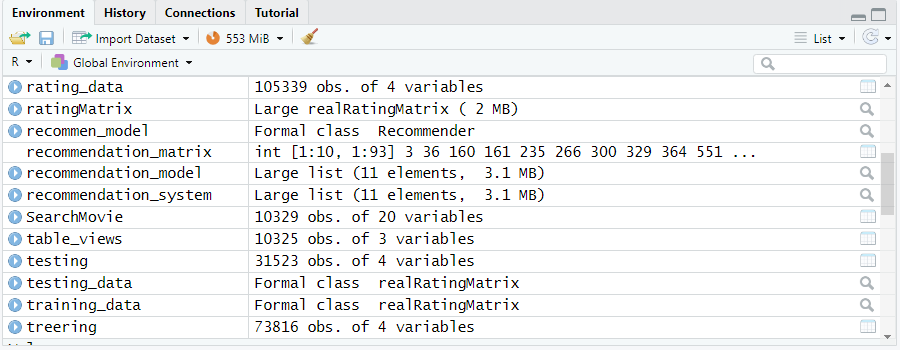
Using the getModel() function, I retrieved the recommen\_model. I then found the class and dimensions of the similarity matrix, that is, contained within model\_info. Finally, I generated a heatmap, that will contain the top 20 items and visualize the similarity shared between them.

In the next step of the ML project, I carried out the sum of rows and columns with the similarity of the objects above 0. I visualized the sum of columns through a distribution.

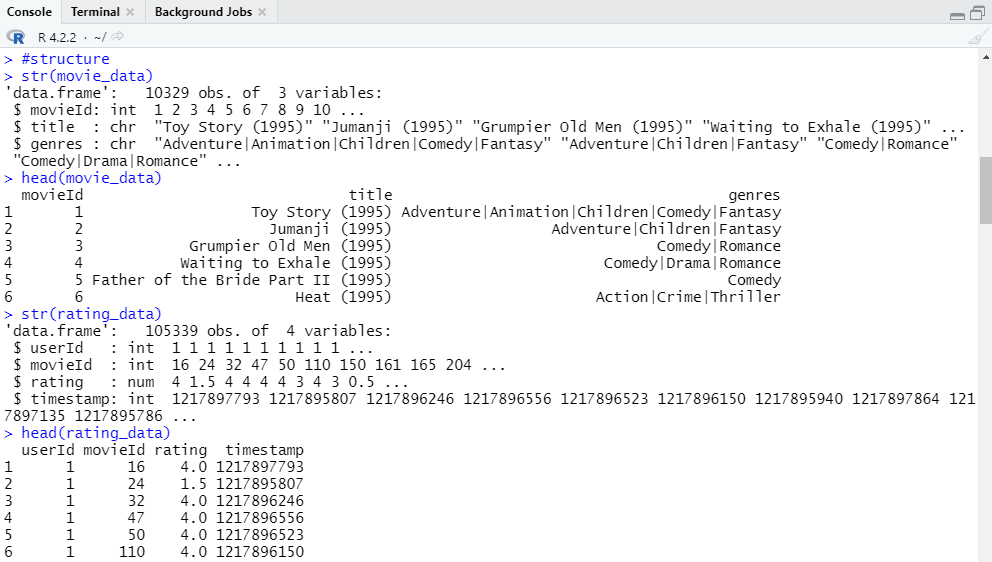
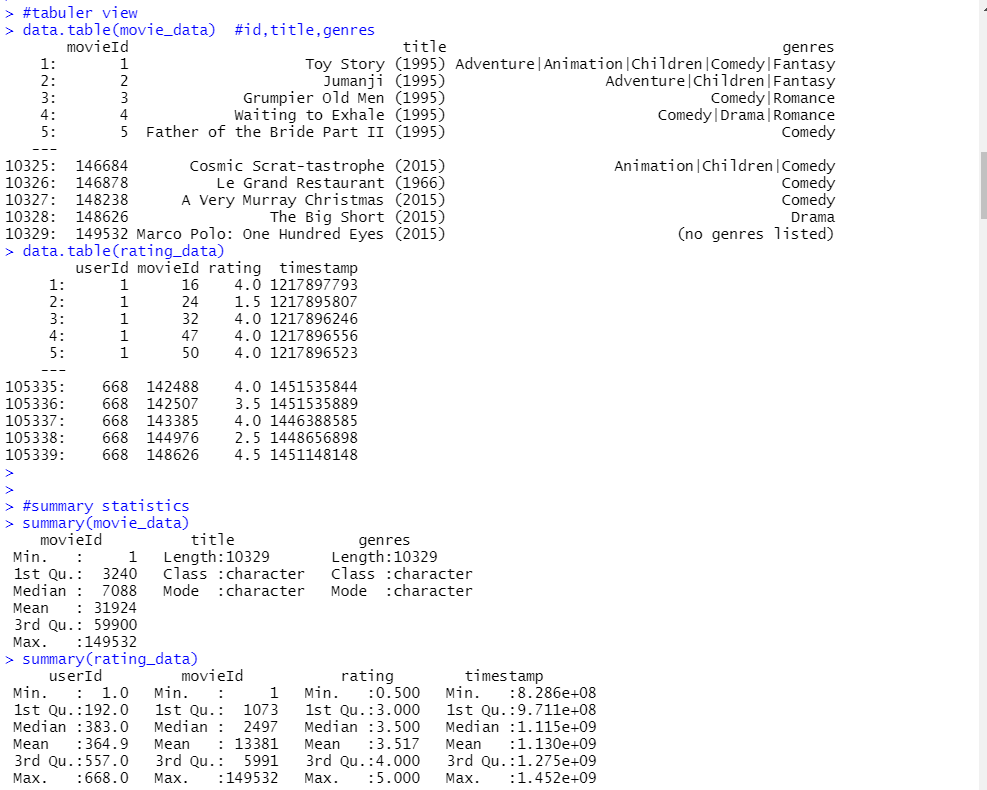
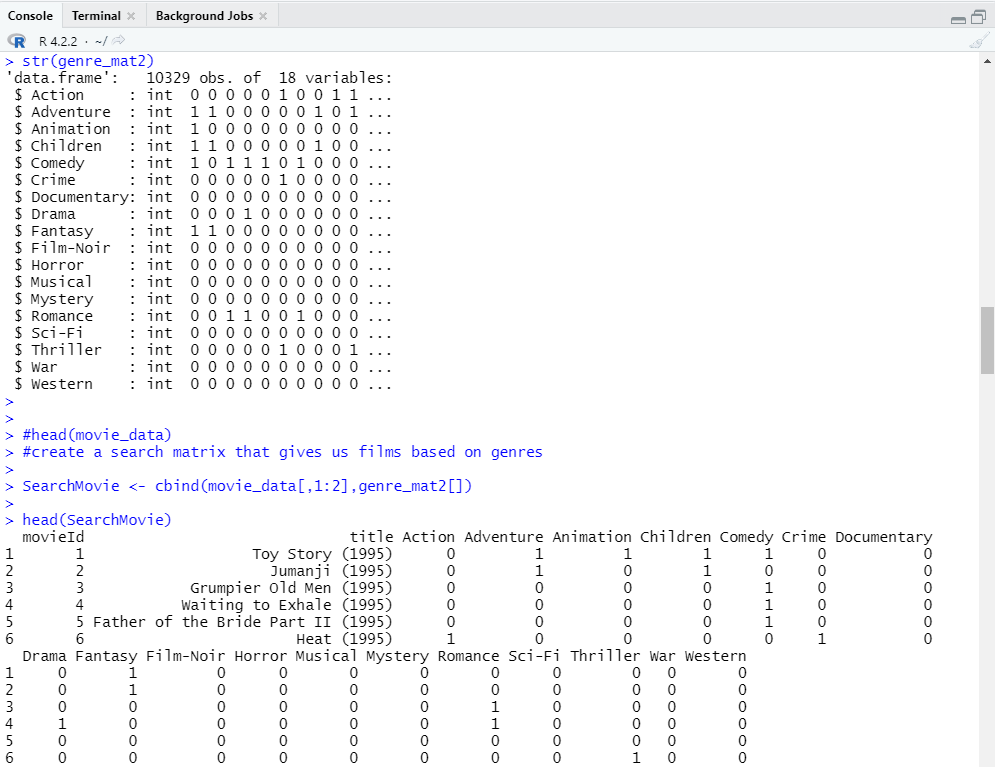
**Building Recommender System on dataset using R**

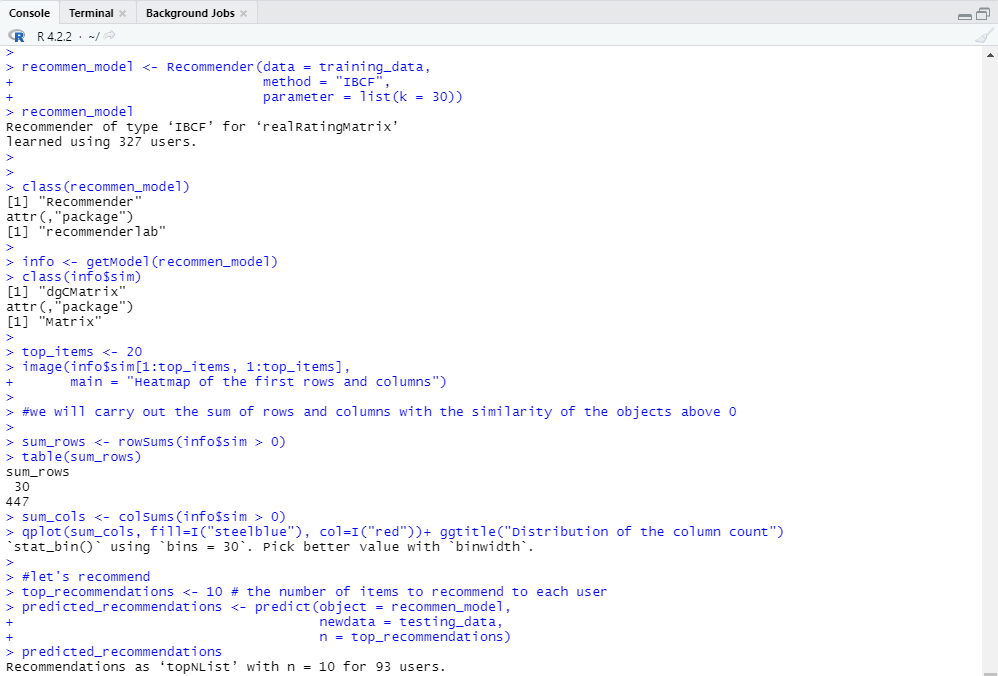
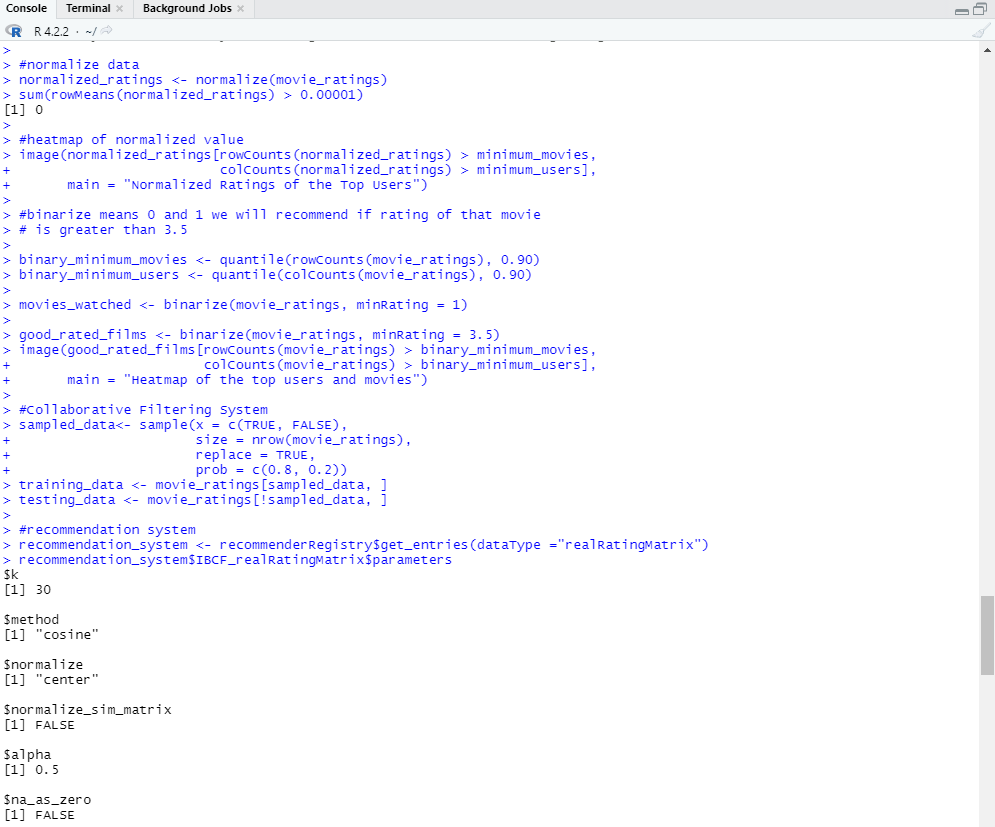
I created a top\_recommendations variable which I initialized to 10, specifying the number of films to each user. I then used the predict() function that identified similar items and ranked them appropriately. Here, each rating is used as a weight. Each weight is multiplied with related similarities. Finally, I added everything in the end.

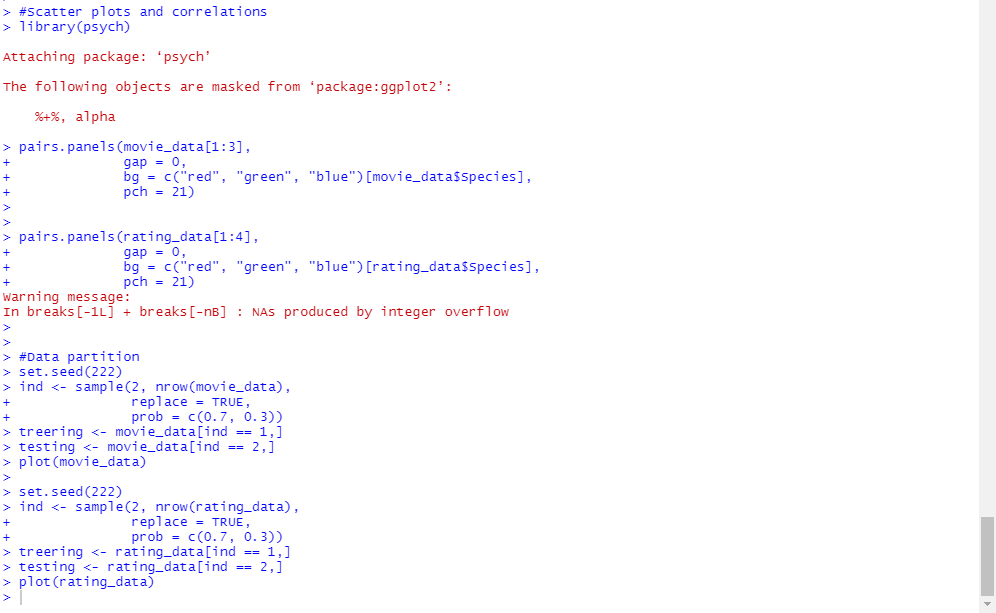
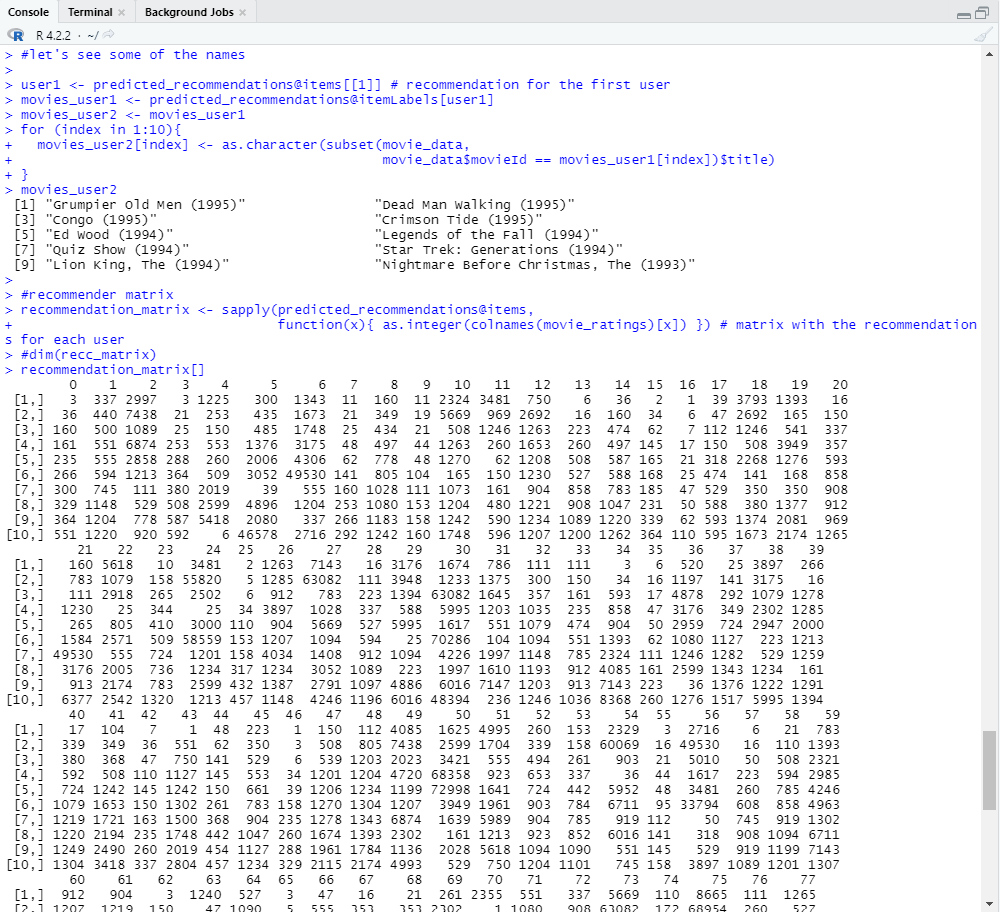
Environment



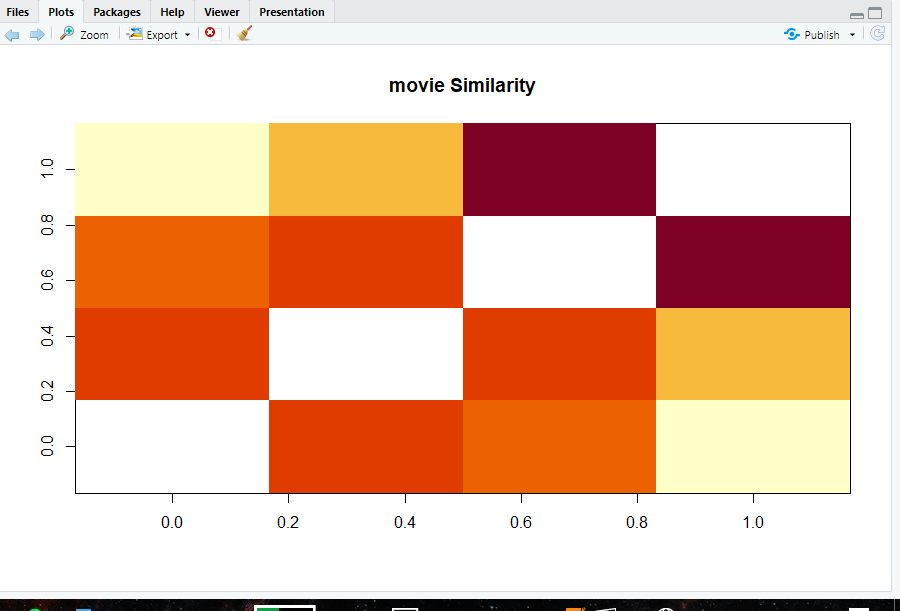
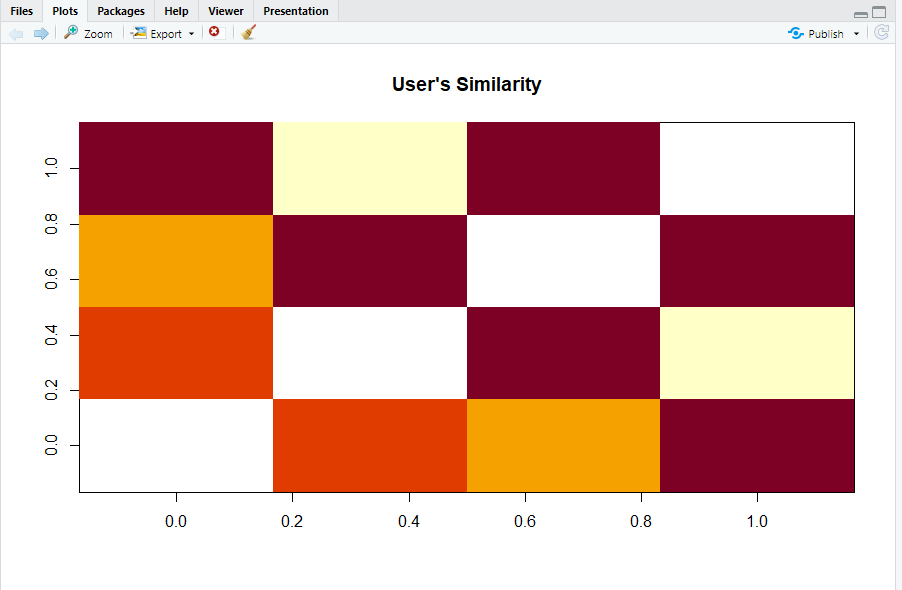
Console







Plots



Chart, bar chart

Description automatically generated