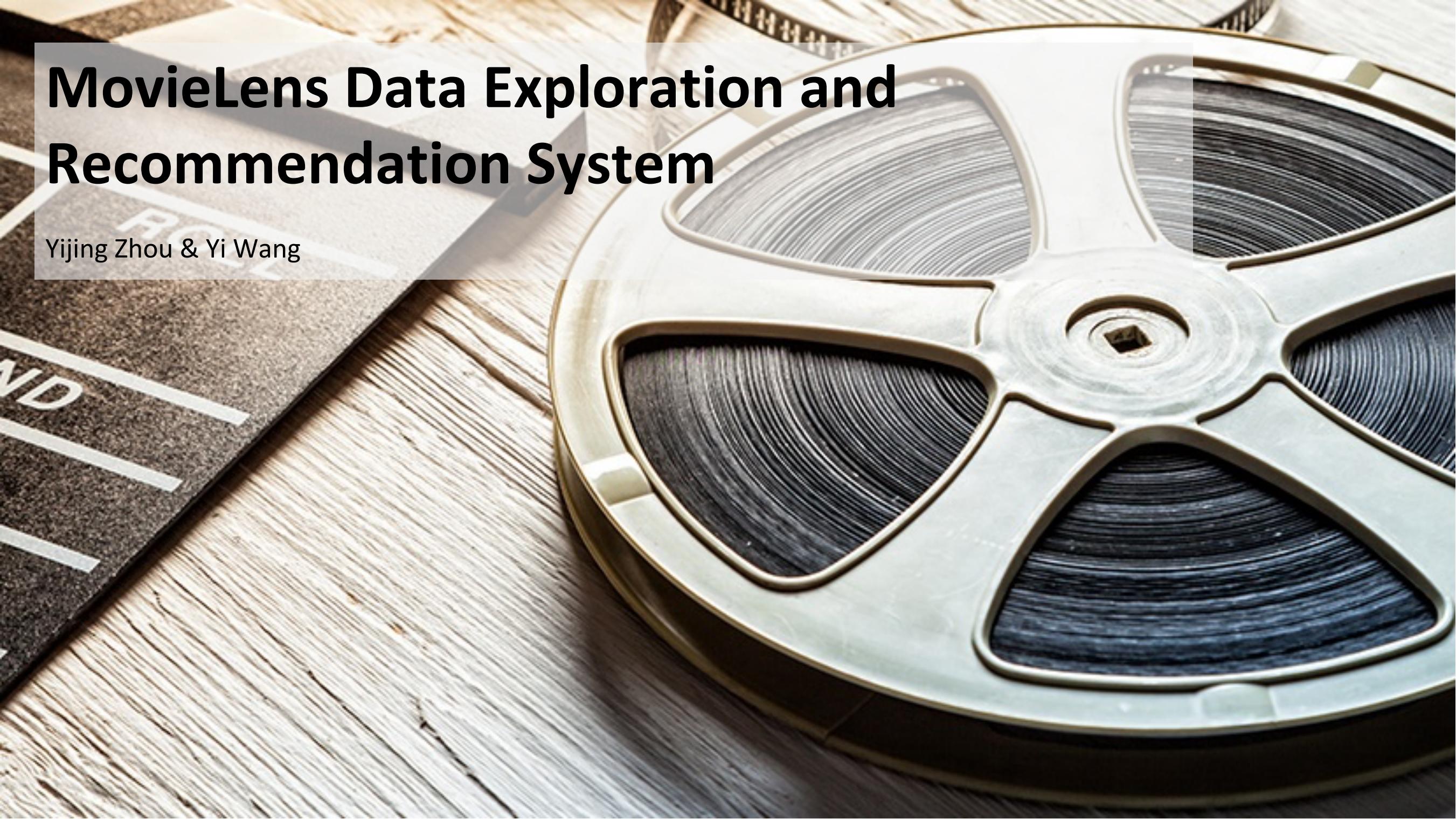


MovieLens Data Exploration and Recommendation System

Yijing Zhou & Yi Wang



Motivation and Tasks

Movies are now an inseparable component of people's daily life. The trend of the movie industry is highly dependent on audiences' taste and preference over time. There are several questions that motivate us to select this topic:

- What type of movies are among the most popular?
- How many differences are there among different users' tastes?
- How does the movie recommendation system work to accurately recommend related movies to users?

Data Description

- 1M MovieLens Data:
- **1,000,209** anonymous ratings made during 2000 to 2003
- **3,900** movies released from 1919 to 2000
- **6,040** MovieLens users.
- Reason for the choice:
 - It contains user information such as age, occupation, geographic location that help us better understand the questions we raised previously.

Data Attributes

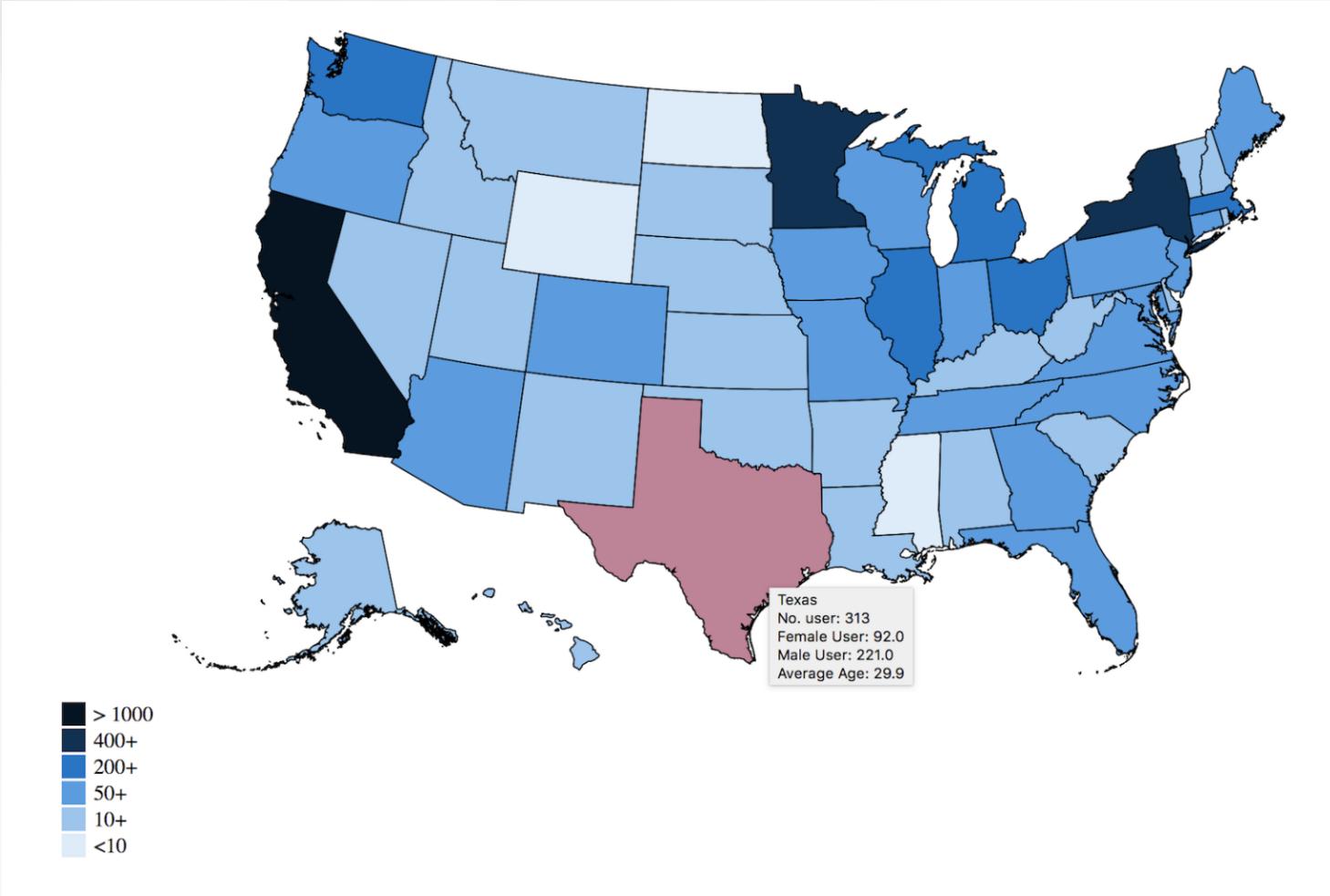
| User Data | | Attribute name | | | | |
|------------|-------------|-------------------------|-------------|-------------------------|----------|---|
| Individual | User ID | gender | age | occupation | Zip code | Ratings made to a specific movie |
| By State | User number | Female/Male user number | Average age | Occupation distribution | | Average rating given by the resident user |

| Movie Data | | Attribute name | | | | |
|------------|---------------|---|---|---------------------------------------|--|---------------------------------------|
| Individual | Released year | Average rating (Overall/by year) (reputation) | Female average rating (Overall/by year) | Male average rating (Overall/by year) | genre | Total amounts of ratings (popularity) |
| By Genre | | Total average rating (Overall/by year) | Female average rating (Overall/by year) | Male average rating (Overall/by year) | Occupation distribution of rated users | Total amounts of ratings |

System Overview

Our system contains **six** visualization designs to present our analysis on the movie and user dataset as well as the recommendation system.

1. User Map – User Background Information



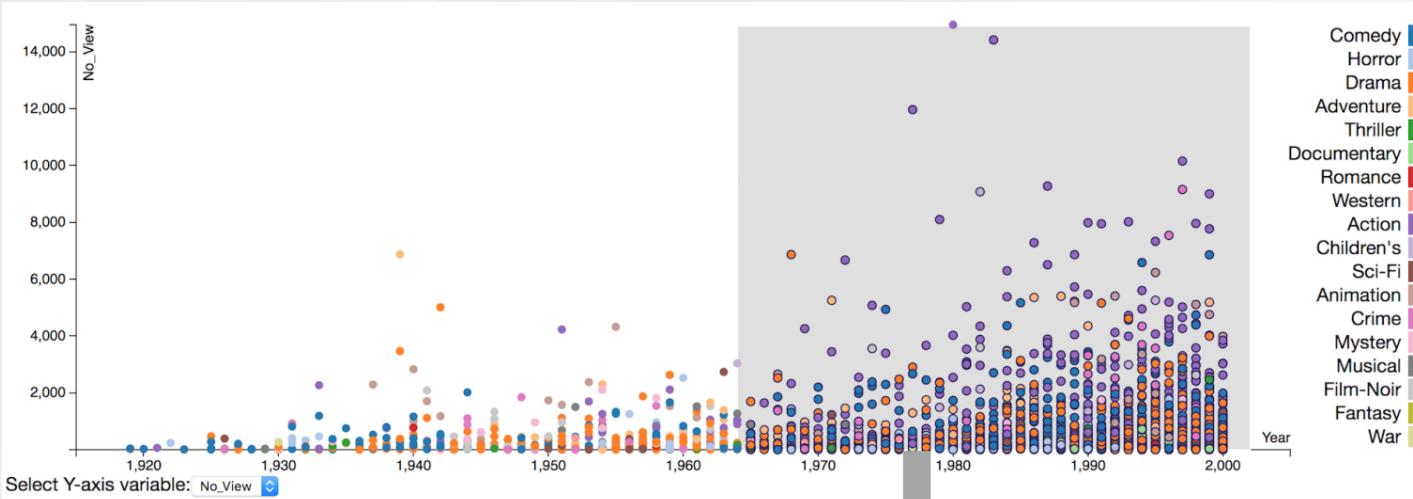
Task:

- Clearly demonstrates the geographic information of the user
- Provide a first glance of the MovieLens user background

Visual Encoding:

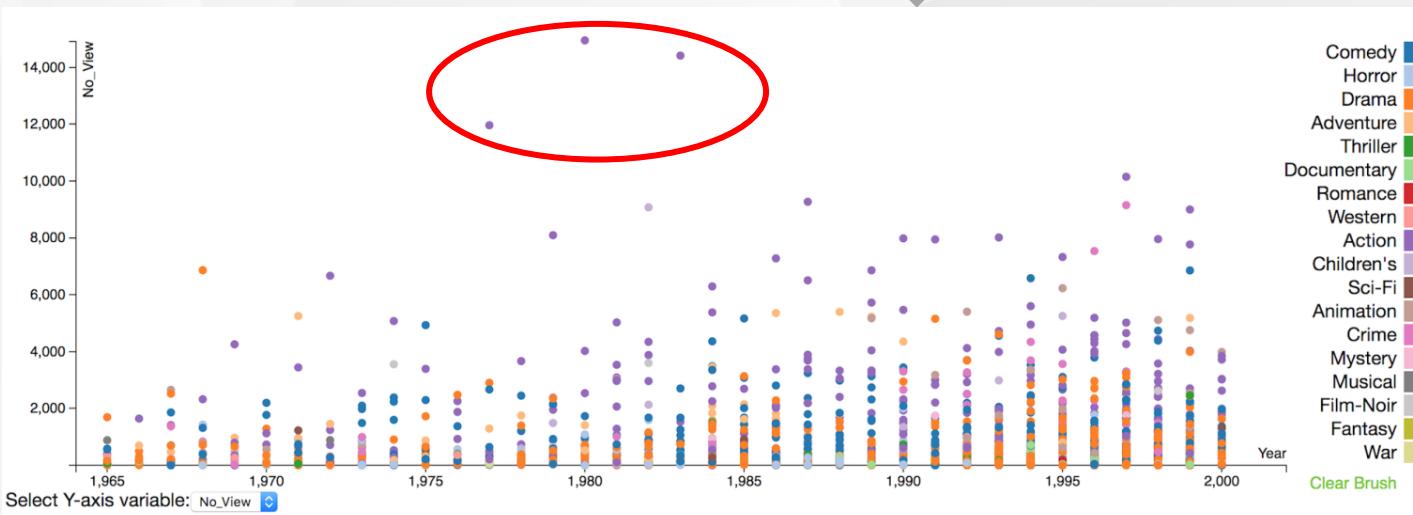
- Color: Number of users
- Position: State

2. Brushable Scatterplot – Movie Ratings/Popularity by year



Task:

- Show the distribution of all individual movies
- Have an overview of popularity and reputation of all the movies



Visual Encoding:

- Color: the genre of each movie
- Position: the published year and total view/average rating

3. Word Cloud – Genre Popularity by Gender



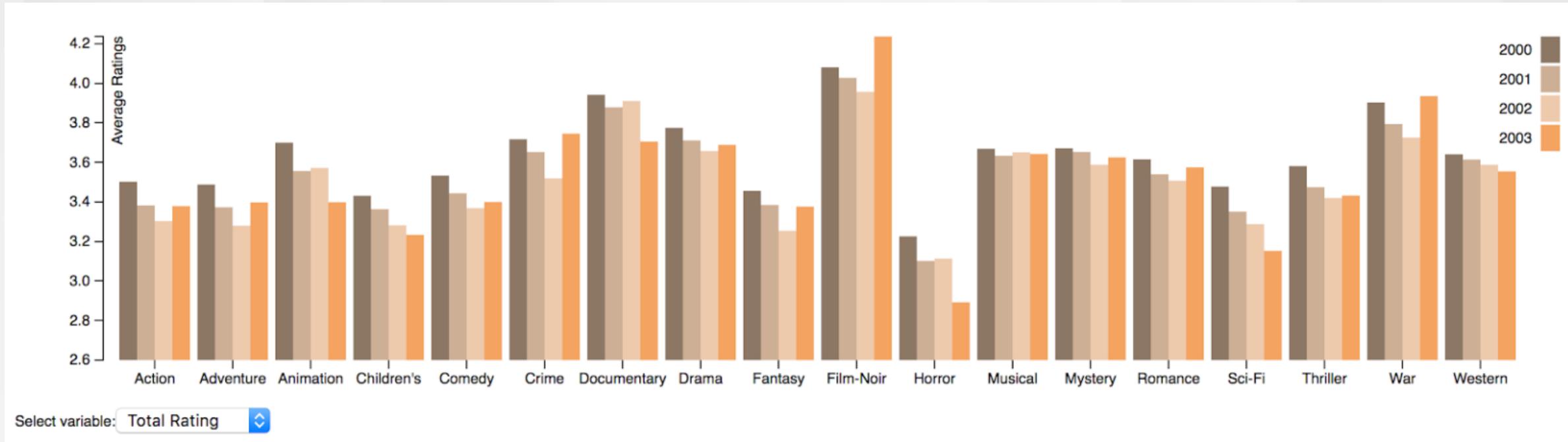
Task:

- Illustrate genre popularity in female and male users

Visual Encoding:

- Size & color saturation: larger size and lighter color suggests a higher popularity of the genre

4. Trend of Genre Rating Over Years



Task:

Demonstrate the trend of average ratings per genre in gender segments over the year

Visual Encoding:

- Position: positioned according to different genre
- Color: the year ratings are made
- Length: the amount of average rating

5. Occupation Distribution of the Genre



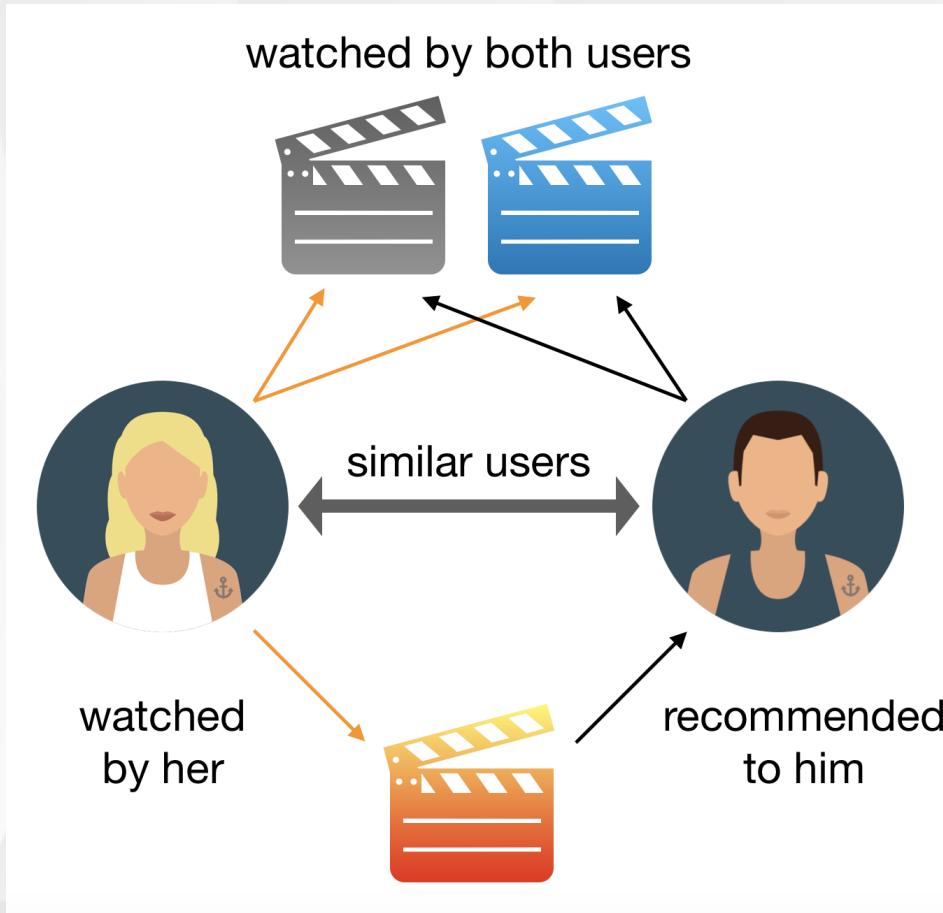
Task:
show the composition of users in each genre in terms of occupation

Visual Encoding:

- Color: occupation of users
- Area: proportion of each occupation

6. Movie Recommendation System

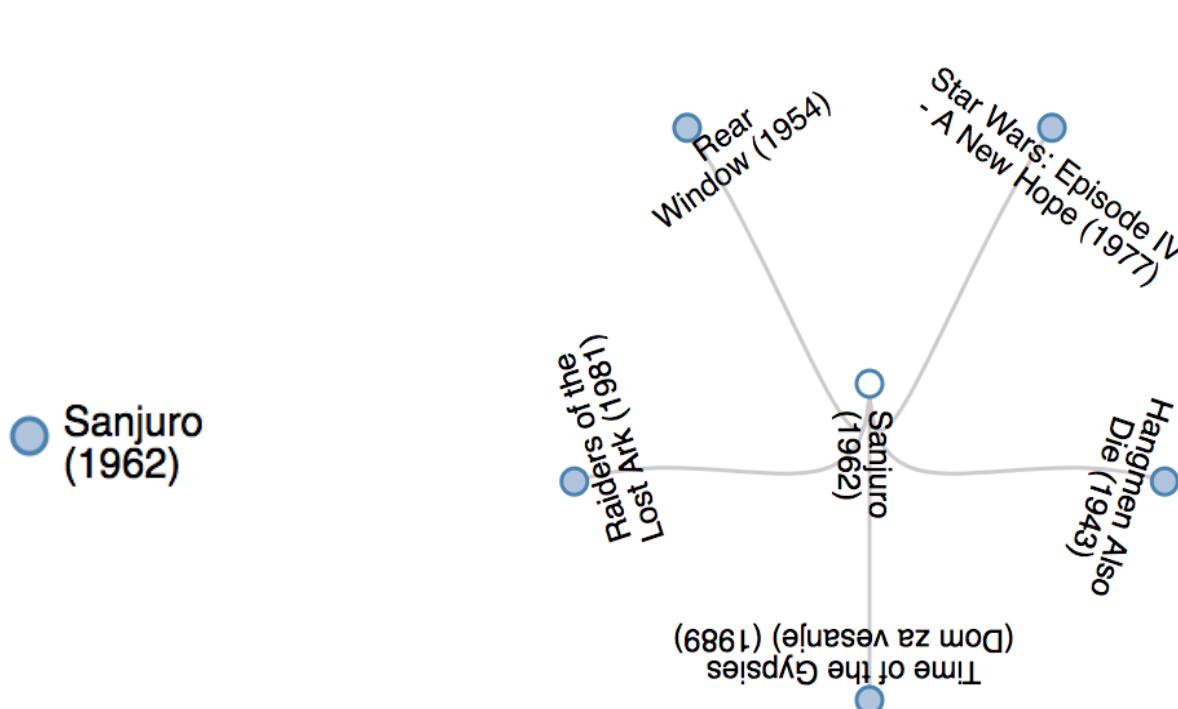
Collaborative Filtering System



Our System:

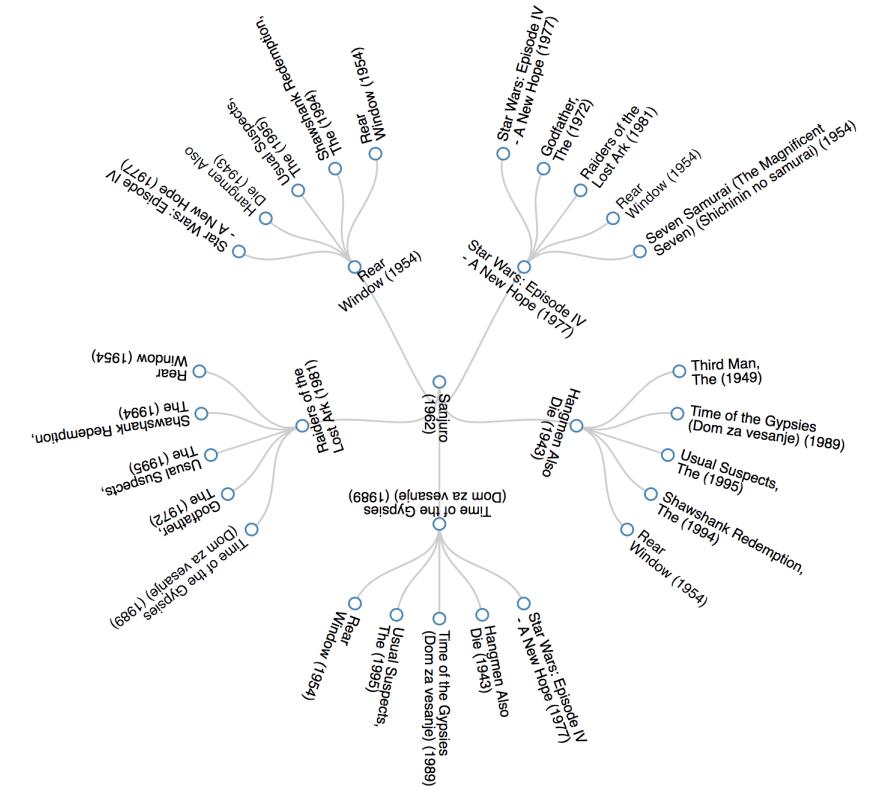
- Find top **20** similar users based on the input movie
- Output **5** recommendation movies based on similar users' records
- Conduct the recommendation process recursively for the **5** recommendation movies
- Output of a two layer radial tree

6. Movie Recommendation System—Radial tree visualization

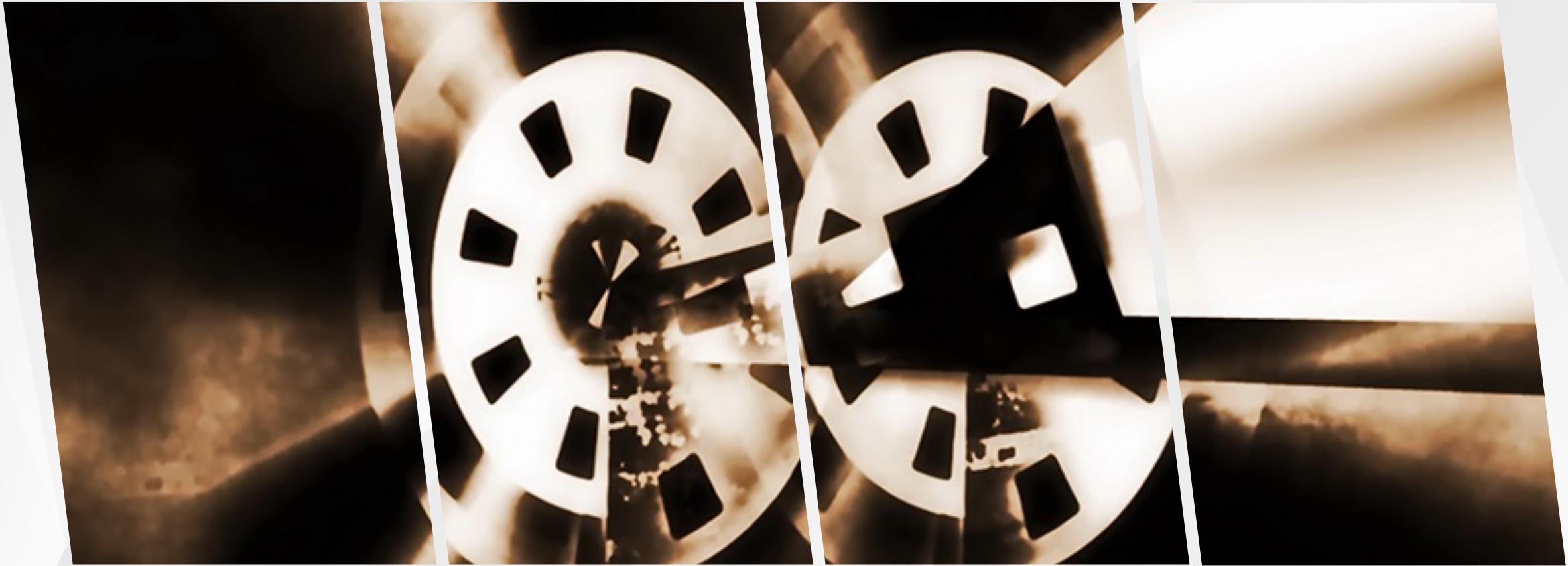


Root Node – Click the movie if you are interested

First layer of recommendation – the system will recommend five movies from the root node using collaborative filter methodology



Second layer of recommendation – the system recommend movies based on the children nodes in the first layer



Live Demo.....

Thank you!

Q&A

