Demographic Insights on Film Ratings and Preferences

Ash Munta1

1 Rutgers University

Author note

The authors acknowledge the support of the GroupLens Research Project for making the MovieLens dataset publicly available. We also thank the R community for providing robust data visualization tools.

The authors made the following contributions. Ash Munta: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Ash Munta, 7500 CPO Way. E-mail: [am3008@rutgers.edu](mailto:am3008@rutgers.edu)

Abstract

Movies serve as a cultural touchstone that resonate differently with various demographic groups. In this study, we leverage the MovieLens 100k dataset to explore how age, gender, and occupation shape movie ratings and genre preferences. Our visual analyses reveal that preferences evolve across the lifespan, with younger viewers gravitating toward horror and western films and older audiences favoring fantasises and crime. Gender-associated patterns, while often subtle, highlight slight differences in top-rated films and rating distributions for popular titles. Occupation, too, emerges as an influential factor, as certain professional groups show distinct affinities for specific genres. These insights underscore the importance of incorporating demographic signals into recommendation systems and content strategies, ensuring a more inclusive and satisfying cinematic experience.

Demographic Insights on Film Ratings and Preferences

# Introduction

Consumer engagement with entertainment media, such as films, often varies according to demographic factors like age, gender, and occupation. Understanding these differences is critical not only for recommendation systems and marketing strategies but also for shedding light on inherent biases in how content is rated and perceived. For example, certain genres may be systematically underrated or overrated by specific demographic groups, influencing the overall ratings landscape and possibly marginalizing some types of films.

By examining demographic influences on movie ratings, we can visualize patterns that highlight potential biases and preferences. Such insights may guide content creators, distributors, and platform developers in curating more inclusive offerings. This also allows for more transparent recommendation algorithms that acknowledge demographic trends, ultimately enhancing user experience and satisfaction.

**Research Questions:**

1. **Primary Question:** How do demographic factors—particularly age, gender, and occupation—influence movie ratings and genre preferences?
2. **Secondary Question:** Do users from different age groups, genders, and occupations rate movies differently on average, and do they have distinct genre preferences?

# Method

## Data Source and Collection

I used the **MovieLens 100k dataset**, collected by the GroupLens Research Project at the University of Minnesota (September 1997–April 1998). Users of the MovieLens platform voluntarily rated movies on a 1–5 scale. Demographic details (age, gender, occupation, zip code) were self-reported. The dataset includes 100,000 ratings from 943 users on 1,682 movies.

## Participants

* **Users (N=943):** Each user rated at least 20 movies.
* **Demographics:** Mean age ~ 34 years, gender recorded as M/F, and a wide range of occupations (e.g., student, educator, technician).

## Data

* **Ratings Data (u.data):** user\_id, movie\_id, rating, timestamp.
* **User Data (u.user):** user\_id, age, gender, occupation, zip\_code.
* **Item Data (u.item):** movie\_id, title, release\_date, plus 19 binary genre indicators: Action, Adventure, Animation, Children’s, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western, and an unknown category.

## Procedure

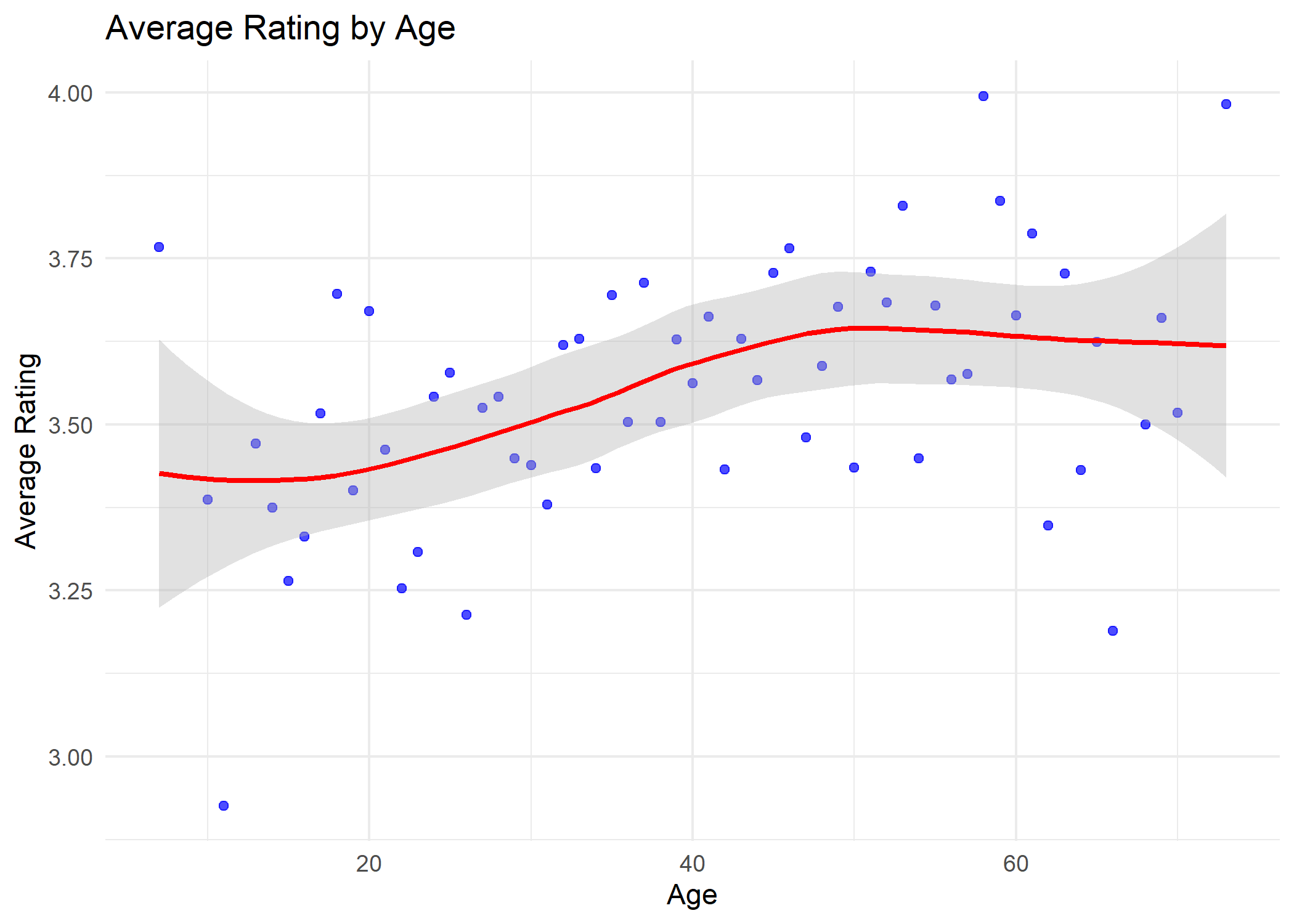
The data was obtained from the MovieLens website, where users rated movies voluntarily. All three data files were merged into one comprehensive dataset.

## Data Analysis

The data analysis was conducted using the following R packages: dplyr (Wickham, François, Henry, Müller, & Vaughan, 2023) for data wrangling and summarization, and ggplot2 (Wickham, 2016) for data visualization.

# Results

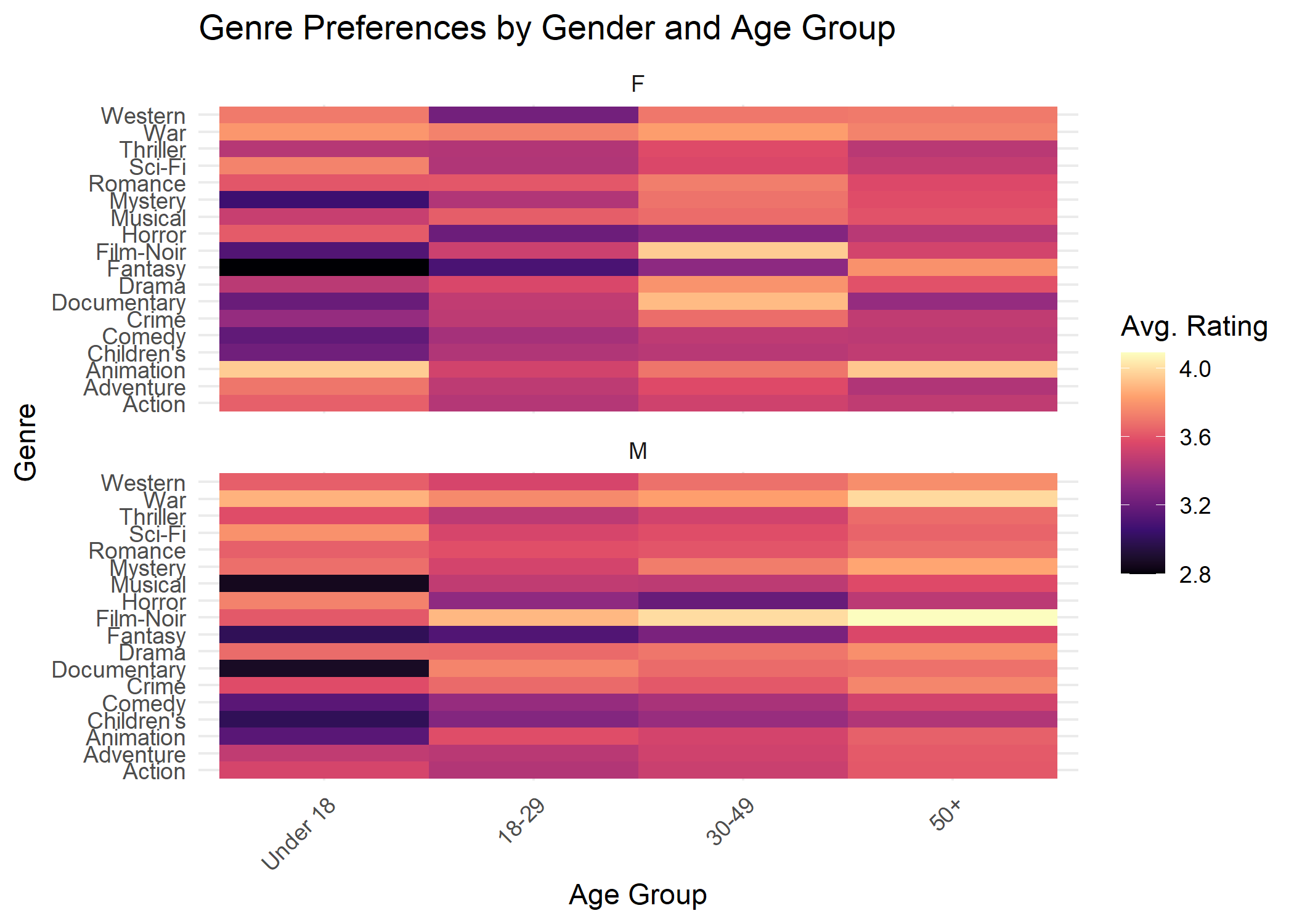
## Plot 1: Continuous Age vs. Average Rating



*Figure* 1. Average movie rating as a continuous function of age.

This plot shows average movie ratings as a function of continuous age. Each blue point represents the average rating given by users of that exact age, while the red LOESS line and gray confidence band illustrate a smoothed trend. The trend suggests that ratings start slightly below 3.5 for younger ages, increase gradually into the late 20s and 30s, then plateau or slightly decline past middle age. Although variability exists at all ages, there’s a gentle upward trend in the center, indicating subtle shifts in taste or rating tendencies across the lifespan.

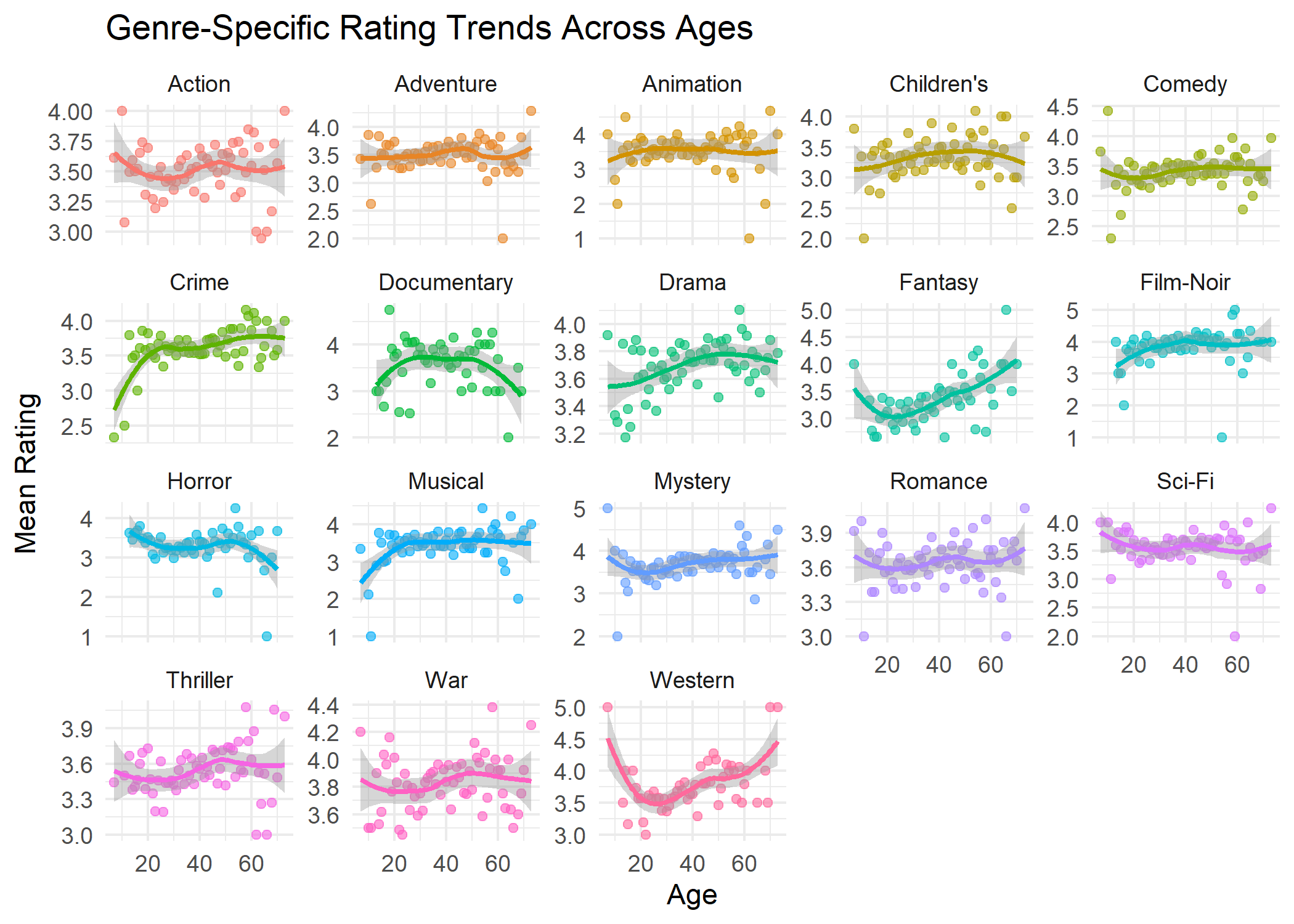
## Plot 2: Faceted Heatmap of Genre Preferences by Gender and Age Group



*Figure* 2. Faceted heatmap of genre preferences by gender and age group.

This heat map visualizes mean ratings for various genres across different age groups, separated by gender (F on top, M below). Darker hues indicate lower ratings, lighter or more intense hues (toward yellow/white) indicate higher ratings. This map highlights a couple different trends. First we notice, regardless of gender, younger audiences tend to be more critical. Nonetheless, there seems to be a variation in preferences by age group for through a gender distinction.

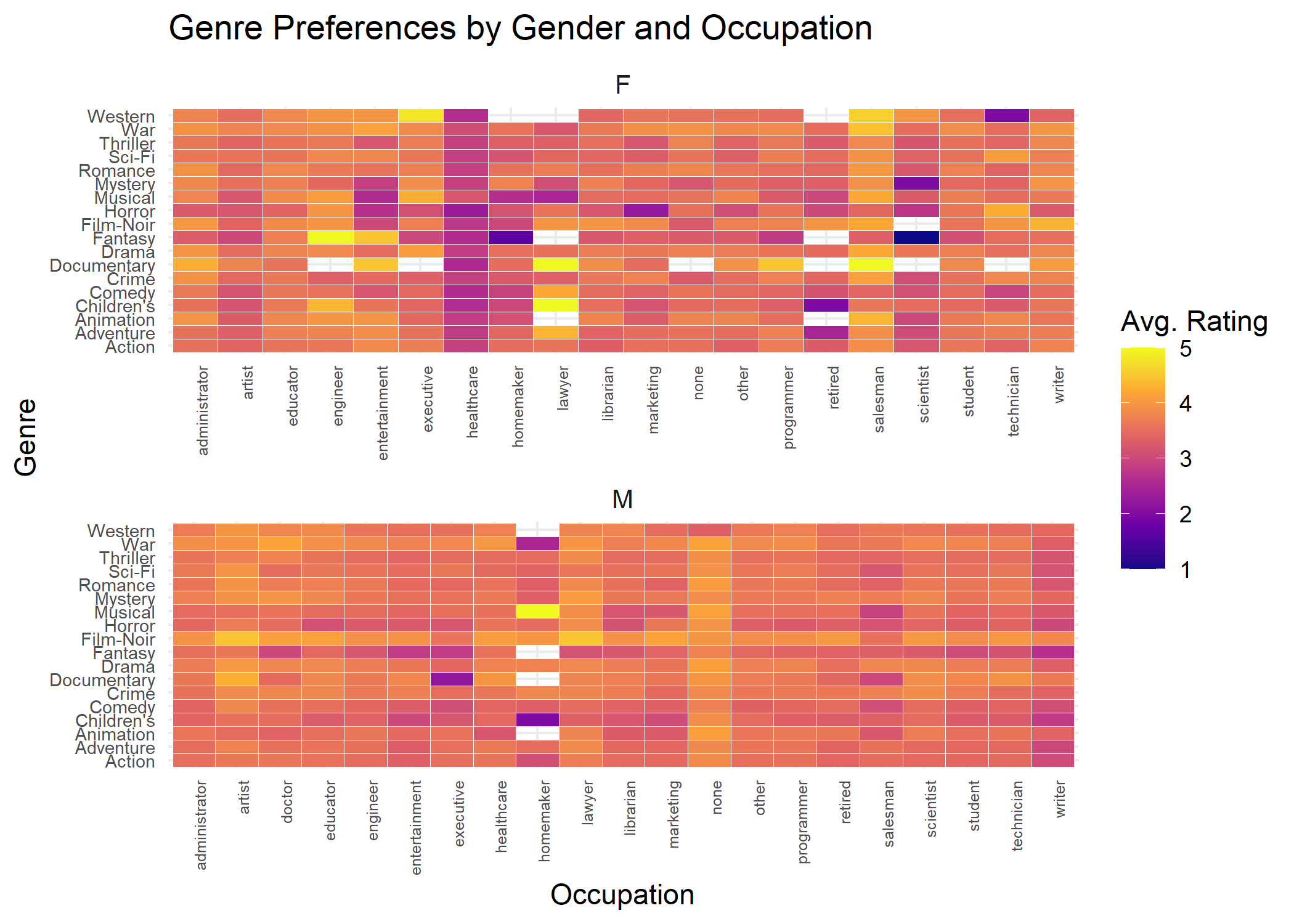
## Plot 3: Faceted Line Graph of Genre Trends Across Ages



*Figure* 3. Faceted line graph of genre rating trends across ages.

Here, each panel represents a single genre, plotting average ratings by age with points and a LOESS curve. We see distinct patterns in some categories, while others seem to show no differences by age. Certain genres like crime, drama, fantasy, and film-noir grow as age increase. While others like horror seem to decrease as age increases. Furthermore, some genres like documentaries, fantasies, and westerns either will show a peak or through in the middle ages.

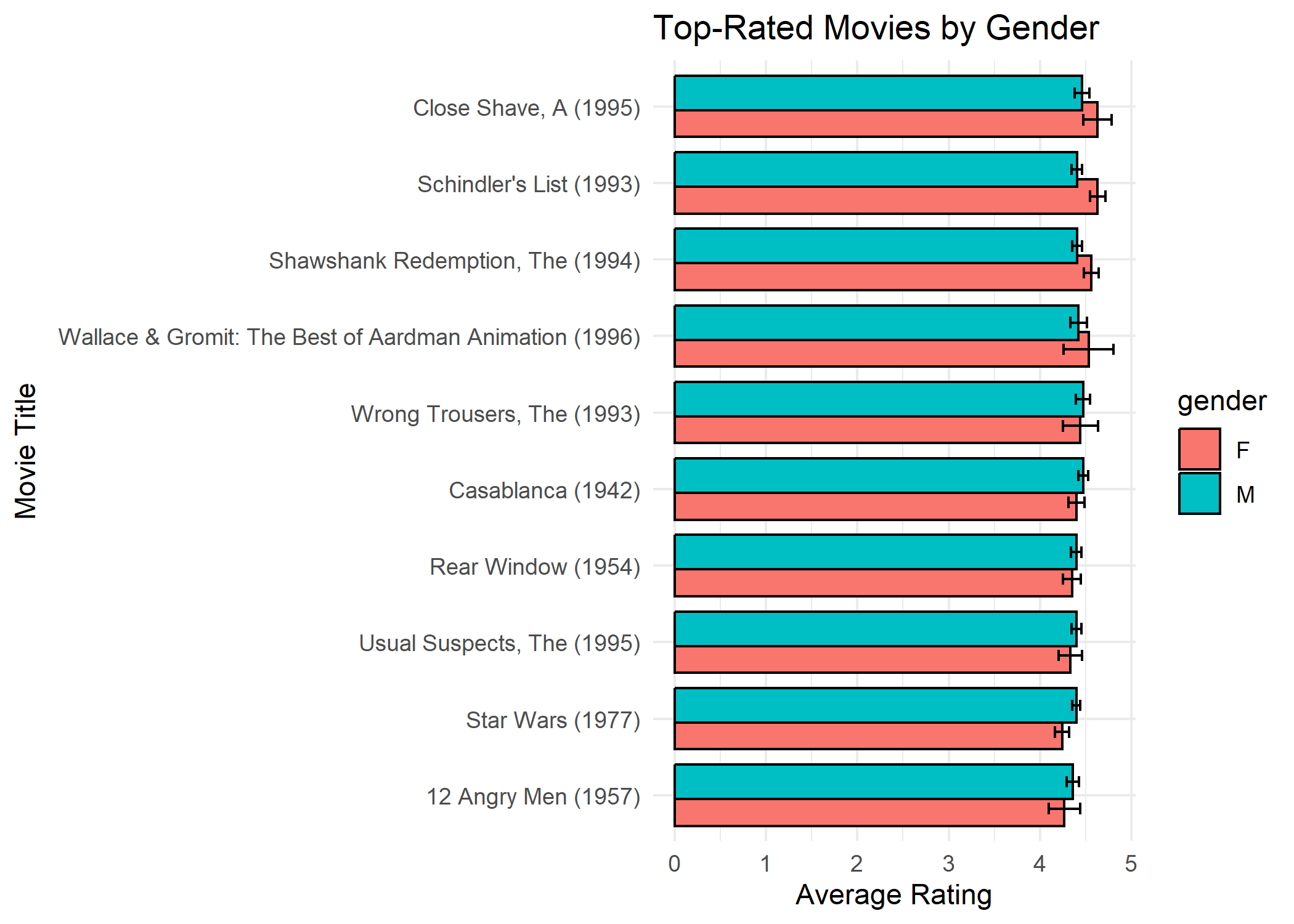
## Plot 4: Heatmap of Genre Preferences by Gender and Occupation



*Figure* 4. Heatmap of genre preferences by gender and occupation.

This heatmap focuses on how gender and occupation interact to shape genre ratings. Occupations (x-axis) are diverse (e.g., students, engineers, librarians, programmers), and each cell shows the average rating for a particular genre (y-axis). If certain occupations cluster around high ratings for intellectual genres (e.g., Documentary, Film-Noir), or lighter entertainment (e.g., Comedy, Animation), that suggests professional or cultural backgrounds play a role in taste formation. Gender differences per occupation add another layer: for instance, a genre might be consistently rated higher by females in academically inclined occupations, or males in more technical fields. For example, females in the healthcare field rate rate all movies lower, regardless of their genre. This does not hold true for men in healthcare field. On the other hand, we see men who identified profession as “none”, rate all genres on average higher than other professions.

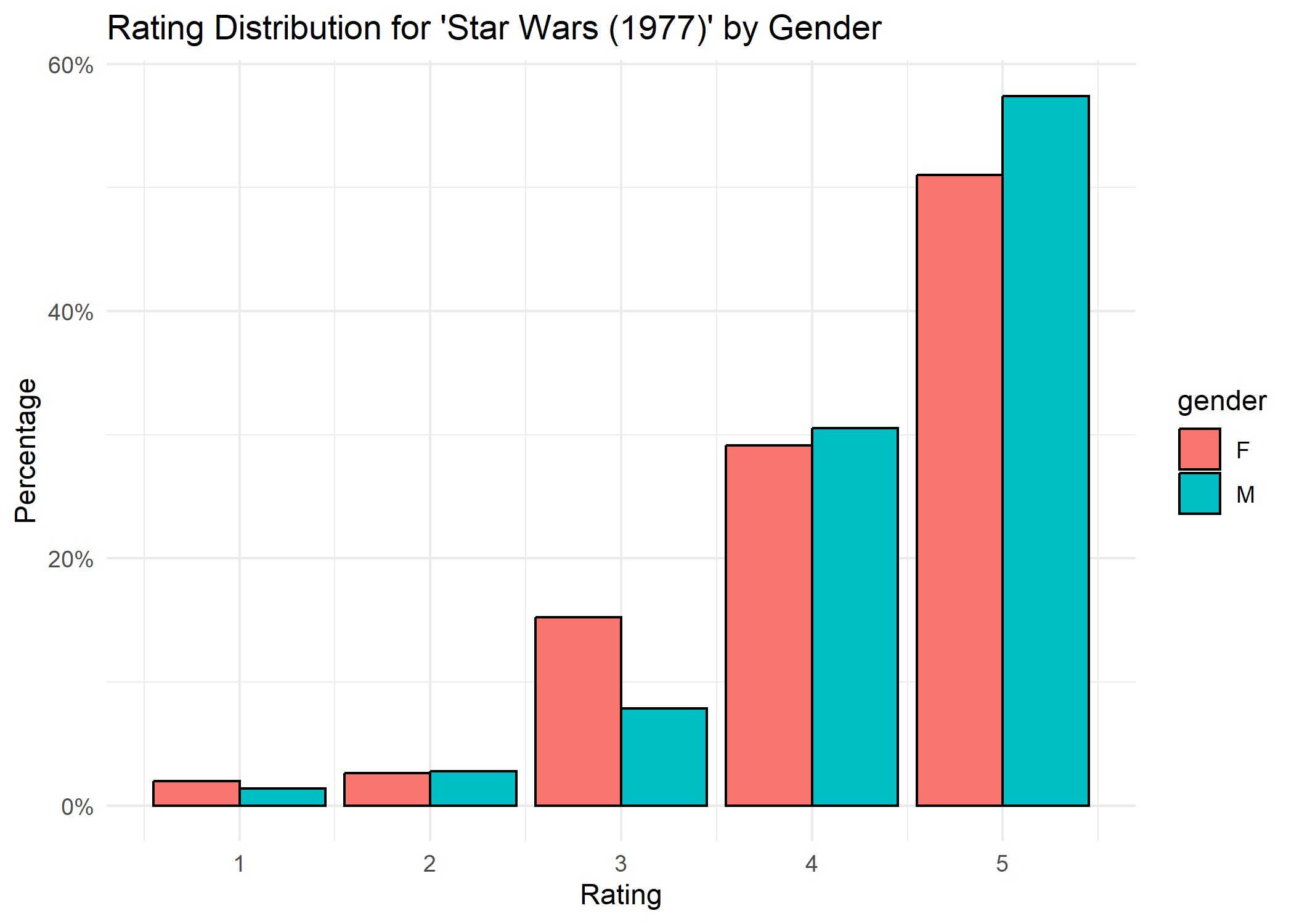
## Plot 5: Top-Rated Movies by Gender



*Figure* 5. Top-rated movies by gender with error bars.

This bar chart shows the top-rated movies overall, with separate bars for female and male mean ratings. Notably, most differences between genders are subtle, with overlapping error bars. Both genders highly rate canonical classics (e.g., Schindler’s List, Shawshank Redemption, Casablanca), indicating broad consensus on certain high-quality films. Minor differences (e.g., one gender slightly preferring a particular film) suggest nuances in taste but no drastic gender polarization for top-tier titles.

## Plot 6: Rating Distribution for a Specific Movie by Gender



*Figure* 6. Rating distribution for ‘Star Wars (1977)’ by gender.

This bar chart displays the percentage of male and female users giving each rating to Star Wars (1977). A prominent takeaway is that the film is well-regarded by both genders, with a large share of 4 and 5 ratings. Males show a relatively high amount of 5 star ratings, while Females demonstrate a relatively higher proportion of 3 star ratings.

## Summary of Findings

* **Age Differences:** The analyses indicate that viewers’ genre preferences shift subtly as they age. Younger audiences show relatively higher appreciation for certain genres like horror and western films, while older viewers appear to rate genres such as fantasy and crime more favorably. Although these differences are not dramatically pronounced, they suggest that age can play a meaningful, if nuanced, role in shaping viewers’ cinematic tastes.
* **Gender Differences:** Although gender differences in overall ratings are often modest, patterns emerge in genre-specific and movie-specific contexts. Women and men largely agree on many top films, but the distribution of very high ratings or preferences for certain genres can differ. Gender comparisons highlight subtle biases or leanings that can become relevant when catering to diverse audience segments.
* **Occupation Differences:** Professional background correlates with genre preferences. Certain occupations show distinct patterns, possibly reflecting cultural, educational, or lifestyle factors that influence taste. These findings highlight that entertainment preferences are shaped not only by who we are demographically but also by the environments we engage in professionally.

Together, these results stress the importance of demographic awareness in movie recommendation systems, marketing strategies, and content development to accommodate a range of tastes and avoid one-size-fits-all approaches.

# Discussion

When examining how preferences evolve across the lifespan, our findings reveal that age-related differences do not represent a simple linear progression of tastes but rather a subtle reshaping of genre affinities. The data suggest that younger viewers are more inclined toward the immediate thrills or distinctive storytelling found in horror and western films, whereas older audiences gravitate slightly toward the narrative complexity or thematic depth present in fantasy and crime. These shifts might reflect changes in cultural exposure, life experience, or generational familiarity with certain genre conventions.

Gender-based differences, while not always pronounced, still matter. Subtle variations in rating distributions and top-film preferences suggest that even widely beloved movies can evoke slightly different responses along gender lines. These differences might be rooted in socialization processes, media representations, and cultural norms influencing how men and women relate to certain narratives and genres.

Occupation-related disparities underscore that movie taste is influenced not only by personal demographics but also by cultural and social contexts. Educational, professional, and social frameworks shape exposure to and appreciation of various film types, from documentaries and international cinema to comedies and blockbusters.

These insights carry practical implications. Recommender systems that integrate demographic signals can better align content suggestions with audience segments, potentially increasing satisfaction and engagement. Marketers and filmmakers can tailor their outreach and content creation to address these nuanced preferences, fostering inclusivity and resonance. Overall, understanding how age, gender, and occupation inform taste can help the film industry navigate a diverse, evolving audience landscape, leading to richer, more meaningful cinematic experiences. Future research should incorporate more diverse and contemporary datasets, non-binary gender options, and additional cultural or socioeconomic factors.

# References

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>

```