Data Analysis Project 1 — Team Report

Dataset: Wallisch & Whritner (2017) Movie Ratings Replication Set
α = 0.005

Q1 — Are popular movies rated higher than unpopular ones?

D (Do):
To examine whether popularity is associated with higher enjoyment, the number of valid ratings (non-missing values) was calculated for each of the 400 movies.
Movies were then split into two groups using a **median split** on the number of ratings:

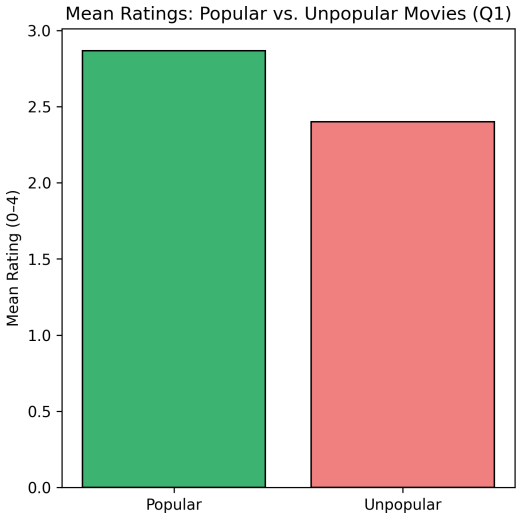
- **Popular** (above median number of ratings)
- **Unpopular** (below median number of ratings)

Mean enjoyment scores were computed for each movie and compared using **Welch's independent-samples t-test**.

Y (Why):
Welch's t-test was chosen because it compares mean values between two independent groups without assuming equal variances or equal sample sizes, conditions appropriate for this dataset since popular films naturally have more ratings.

F (Find):
Popular movies had a **mean rating of 2.87**, while unpopular movies had a **mean rating of 2.40**.
The difference was **highly significant**, t (398) = 17.76, **$p < 0.001$**.

A (Answer):
Popular movies were rated significantly higher than unpopular movies.
This finding suggests that films which attract more viewers also tend to be perceived as higher quality or more enjoyable, aligning with the idea that widespread appeal often correlates with positive reception.



(See Figure Q1: "Mean Ratings — Popular vs. Unpopular Movies" for visualization.)

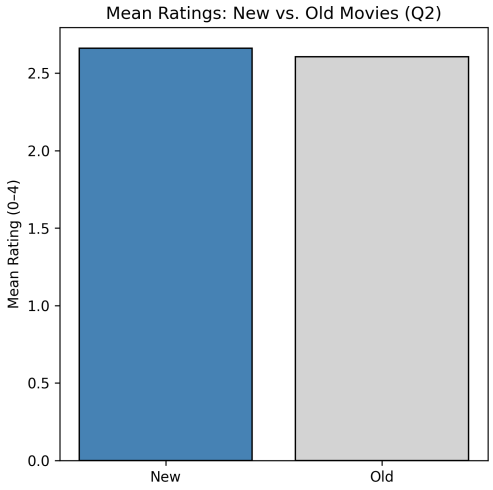
Q2 — Are newer movies rated differently than older ones?

D (Do):
Each movie title included its release year (e.g., "The Matrix (1999)").
These years were extracted and used to classify films as **new** or **old** using a **median split**.
The mean rating for each group was compared using Welch's t-test.

Y (Why):
This analysis tests whether movie recency (a proxy for cultural or production-value differences) affects perceived quality.
Again, Welch's t-test was appropriate because the two groups differ in sample size and variance.

F (Find):
Newer movies had a **mean rating of 2.66**, while older movies had a **mean rating of 2.61**.
The difference was **not statistically significant**, t (398) = 1.61, p = 0.109.

A (Answer):
There is **no evidence** that newer movies are rated differently from older ones.
This suggests that, within this dataset, **movie age does not systematically influence enjoyment ratings** — viewers seem to appreciate both classic and modern films similarly.



(See Figure Q2: "Mean Ratings — New vs. Old Movies" for visualization.)

Q3 — Is enjoyment of 'Shrek (2001)' gendered, i.e. do male and female viewers rate it differently?

D (Do):
Performed Welch's independent-sample *t*-tests comparing male (code 2) and female (code 1) ratings for Shrek (2001). Excluded missing data and self-described gender responses.

Y (Why):
The Welch's *t*-test was chosen because it does not assume equal variances and is robust to unequal sample sizes. Our purpose is to determine whether someone's gender statistically impacts whether they enjoyed 'Shrek (2001)

F (Find):

Mean rating (male) = 3.08, mean rating (female)= 3.16.
t (df ≈ variable) = -1.156, p = 0.24835> 0.005.
See **Figure Q3** for a visual comparison.

A (Answer):
Since the p-value taken from the t-test is greater than 0.005, there is no statistically significant difference between how males and females' rate 'Shrek (2001)'.

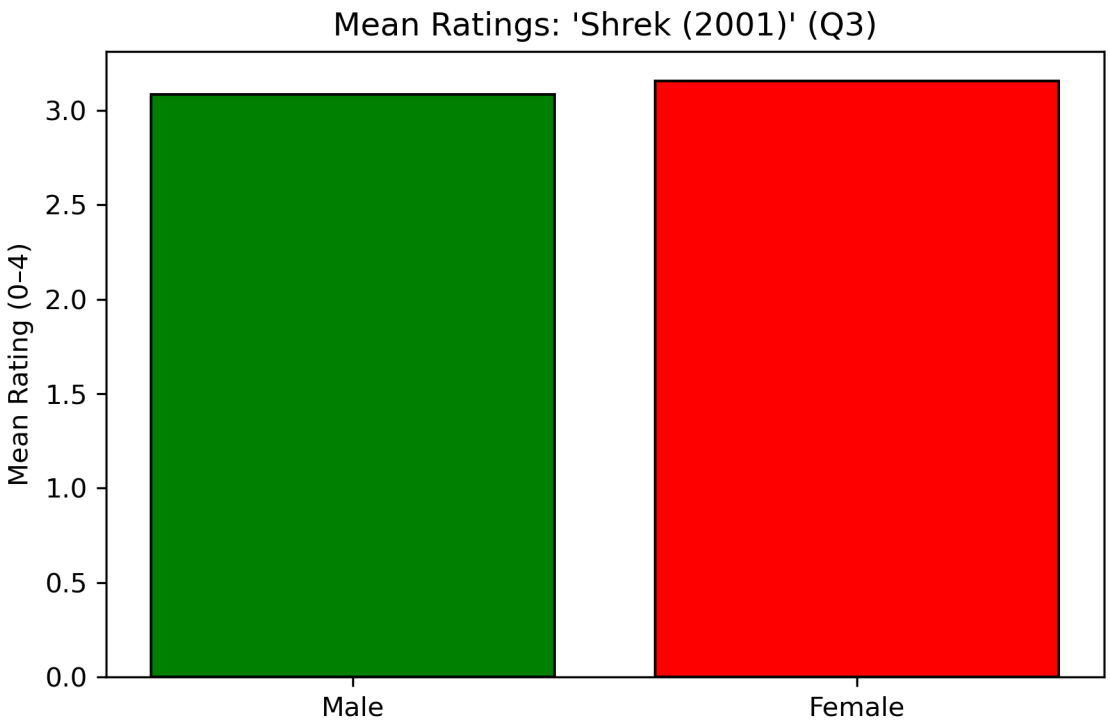


Figure Q3: "Mean Ratings by gender for 'Shrek (2001)'

Question 4 — Gender Differences Across All Movies

D (Do):
Performed Welch's independent-samples *t*-tests comparing male (code 2) and female (code 1) ratings for each of the 400 movies. Excluded missing data and self-described gender responses.

Y (Why):
The Welch's *t*-test was chosen because it does not assume equal variances and is robust to unequal sample sizes. This test identifies whether gender significantly affects movie ratings.

F (Find):
Out of 400 movies, **45 (11.25%)** showed statistically significant differences between male and female viewers at $\alpha = 0.005$.
As shown in **Figure 1**, most *p*-values cluster above 0.005, suggesting few meaningful gender-based rating differences.

A (Answer):
Since only 11.25% of movies differed by gender, we conclude that movie enjoyment is **largely not gendered**. This aligns with Wallisch & Whritner (2017), who found weak demographic effects on movie taste

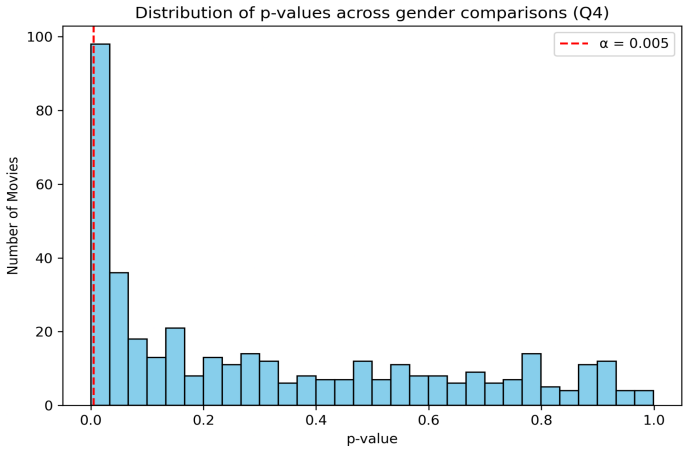


Figure Q4: Distribution of *p*-values across gender-based comparisons (α = 0.005).

Question 5 — "Only Child" Effect for *The Lion King (1994)*

D (Do):
Compared mean ratings of *The Lion King (1994)* between participants who are only children (coded 1) and those with siblings (coded 0) using Welch's *t*-test.

Y (Why):
The Welch's *t*-test was used to compare group means while accounting for unequal variances and sample sizes.

F (Find):
Mean rating (only children) = 3.35; mean rating (with siblings) = 3.48.
t (def. ≈ variable) = –1.792, $p = 0.0746 > 0.005$.
See **Figure Q5** for a visual comparison.

A (Answer):
Since $p > 0.005$, there is **no statistically significant difference** between only children and participants with siblings in their enjoyment of *The Lion King (1994)*.
Both groups rate the film highly and similarly.

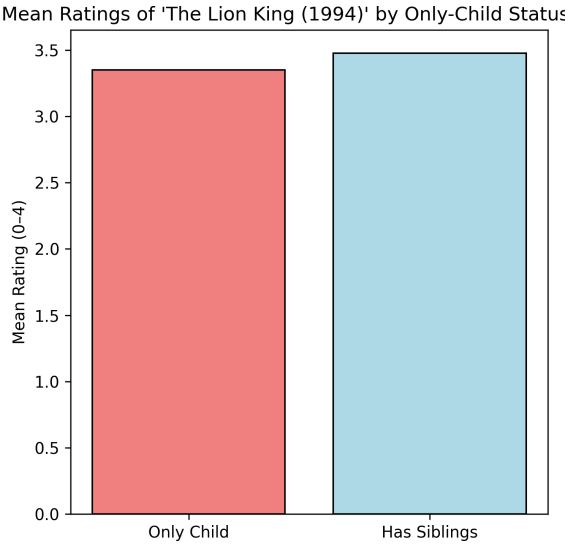


Figure Q5. Comparison of *The Lion King (1994)* mean ratings by only-child status.

Question 6 — “Only Child” Effects Across All Movies

D (Do):
Repeated Welch’s *t*-tests from Question 5 for each of the 400 movies, comparing only children and participants with siblings.

Y (Why):
This systematic approach estimates how widespread “only child” effects are across the entire dataset.

F (Find):
Only **5 movies (1.25%)** showed significant differences (α = 0.005).
As seen in **Figure Q6**, the *p*-value distribution is heavily skewed above 0.005, indicating minimal systematic effects.

A (Answer):
Because only 1.25% of movies exhibit significant differences, we conclude that **“only child” status rarely affects movie ratings**. This suggests personal taste rather than family structure drives enjoyment patterns.

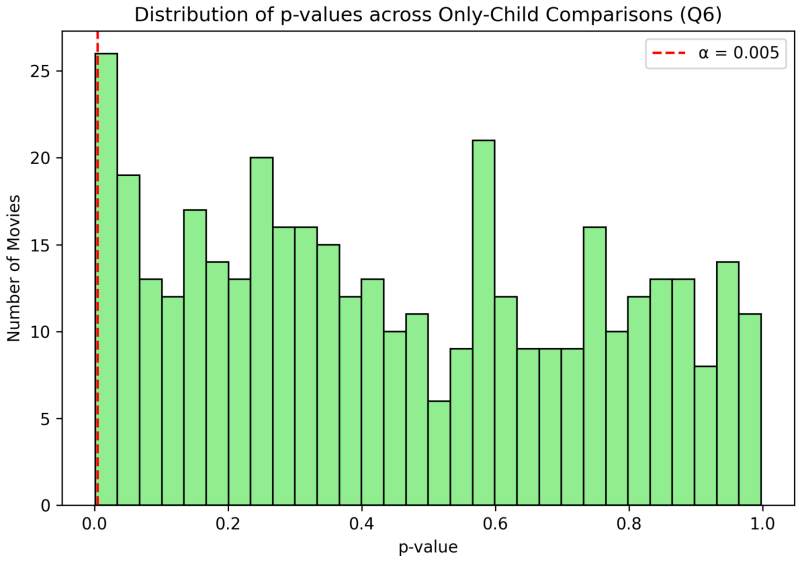


Figure Q6. Distribution of *p*-values across “only child” comparisons (α = 0.005).

Question 7 — Social Watching Preference and *The Wolf of Wall Street (2013)*

D (Do):
Compared *The Wolf of Wall Street (2013)* ratings between participants who prefer watching movies socially (coded 0) versus those who prefer to watch alone (coded 1) using Welch's *t*-test.

Y (Why):
This test determines whether social viewing preference relates to enjoyment of a socially charged, intense film, accounting for unequal group variances.

F (Find):
Mean rating (social watchers) = 3.03; mean rating (prefer alone) = 3.15.
t (df ≈ variable) = –1.604, p = 0.1092 > 0.005.
See **Figure 4** for a comparison.

A (Answer):
Since $p > 0.005$, there is **no significant relationship** between social viewing preference and enjoyment of *The Wolf of Wall Street (2013)*.
Both groups rated the film, similarly, suggesting that viewing context preference does not influence appreciation for this movie.

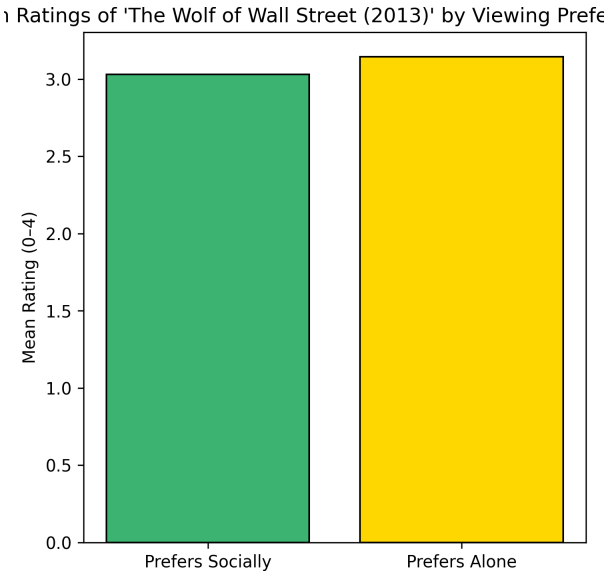


Figure Q7: Mean ratings of *The Wolf of Wall Street (2013)* by viewing preference.

Question 8 — What proportion of movies exhibit a "social watching" effect?

D (Do):
Repeated Welch's *t*-tests from Question 7 for each of the 400 movies, comparing only children and participants with siblings.

Y (Why):
This systematic approach estimates how widespread "social watching" effects are across the entire dataset.

F (Find):
Only **6 movies (1.50%)** showed significant differences (α = 0.005).
As seen in **Figure 3**, the *p*-value distribution is heavily skewed above 0.005, indicating minimal systematic effects.

A (Answer):
Because only 1.50% of movies exhibit significant differences, we conclude that **"social watching" preferences rarely affects movie ratings**. This suggests that social movie watching habits rarely influence someone's enjoyment of a film

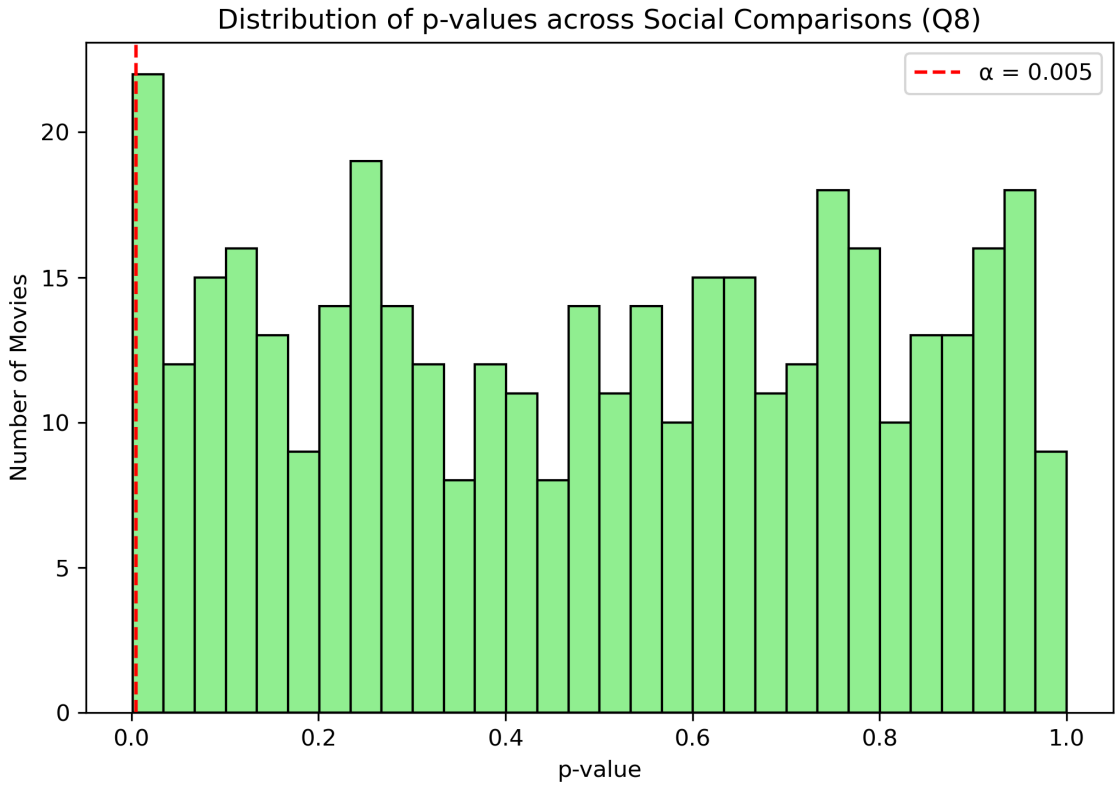


Figure Q8: Distribution of p-values across Social Comparisons (Q8)

. Question 9 — Is the ratings distribution of ‘Home Alone (1990)’ different than that of ‘Finding Nemo (2003)’?

D (Do):

KS tests run on samples of ‘Home Alone (1990)’ and ‘Finding Nemo (2003)’.

Y (Why):

Null Hypothesis H0: The two movies have the same distribution

Alternative Hypothesis Ha: The two movies have different distributions

These tests determine both the probability that the two samples have two different distributions – if there is strong enough evidence to reject the null hypothesis that the two samples have the same distribution- and at what rating in the data set the largest difference between the CDF’s of the two

F (Find):

The test found that p=6.38e-10 < 0.005, indicating the max value of the difference of the two cumulative distributions is significant. This is shown in *Figure Q9* below.

A (Answer):

Since p < 0.005, there is significant evidence that the distribution of ratings of ‘Home Alone (1990)’ and ‘Finding Nemo (2003)’ are different, meaning that we reject the null hypothesis.

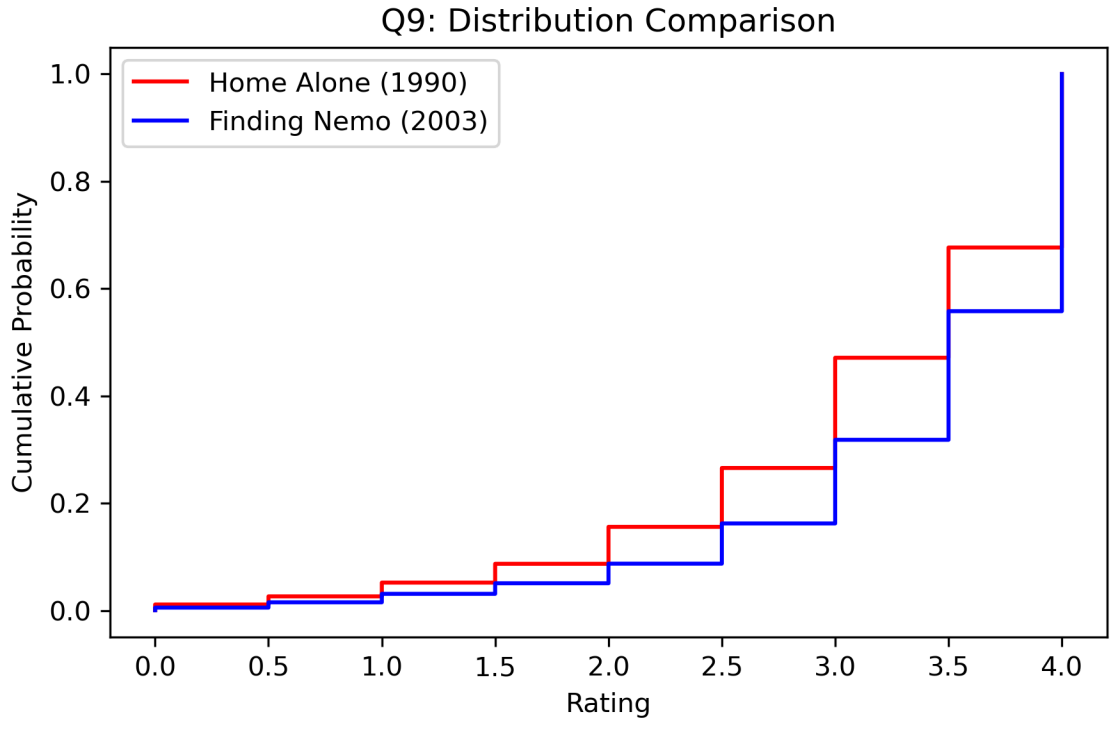


Figure Q9: Distribution Comparison

. Question 10 — There are ratings on movies from several franchises (['Star Wars', 'Harry Potter', 'The Matrix', 'Indiana Jones', 'Jurassic Park', 'Pirates of the Caribbean', 'Toy Story', 'Batman']) in this dataset. How many of these are of inconsistent quality, as experienced by viewers?

D (Do): Use keywords to assign ratings columns of each movie to their proper franchise, before running ANOVA and KW tests on the franchises.

Y (Why): The ANOVA test takes analyzes if there is a statistically significant difference between the general ratings of sets with more than 2 elements, while the KW test runs a similar test considering ANOVA assumption violations and lack of interval scaling.

F (Find): Inconsistent ANOVA and KW P-values for every franchise aside from 'Harry Potter'.

A (Answer): Consistent quality of individual movies in 'Harry Potter' franchise, inconsistent quality of individual movies for every other franchise.

. Bonus — How much does personal life stress impact how someone would rate 'The Silence of the Lambs (1991)'

D (Do): Sort the scores of individuals who rated 'The Silence of the Lambs (1991)' into sets based on stress ratings from 1 to 5 and then run an ANOVA test on the means of each set

Y (Why): The ANOVA test takes analyzes if there is a statistically significant difference between the general ratings of sets with more than 2 elements, so its use applies to looking at the means of ratings of 'The Silence of the Lambs' in different stress groups.

F (Find):

p=0.92823, with the figure below showing similar average scoring between all of the test groups. **Figure Bonus** shows that the mean rating of the movie for people with specific life stress levels is not significantly different.

A (Answer): The data does not show that stress has a significant impact on how individuals rated 'The Silence of the Lambs (1991)'

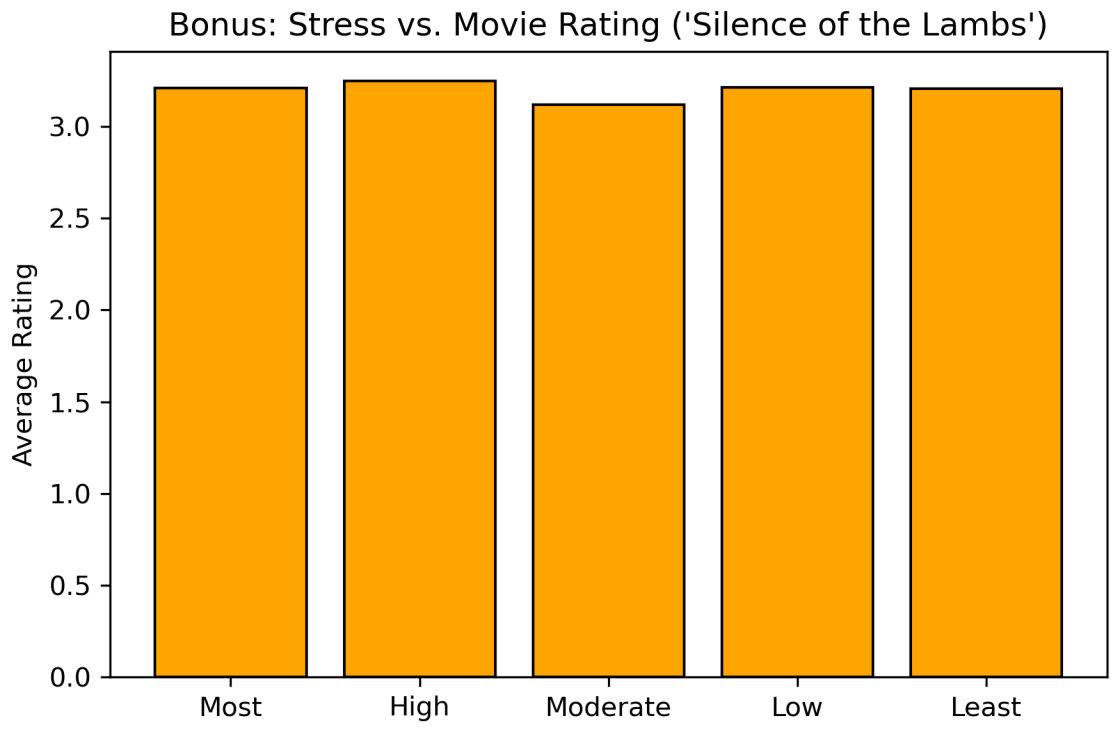


Figure Bonus: Stress vs. Movie Rating (‘Silence of the Lambs’)