Analyzing the Bechdel Test: Budget Trends and Revenue Outcomes in Cinema

December 15th 2024

Alex Ackerman, Cindy Gao, Kayla Haeussler, Javidan Karimli

1. Abstract

Media equity and cultural narratives are significantly impacted by gender representation in films. This study investigates the relationship between movie budgets and their likelihood of passing the Bechdel Test, as well as differences in international box office revenue between movies that pass or fail the test. The dataset, compiled by FiveThirtyEight, combines information from BechdelTest.com and The-Numbers.com, with movie genres supplemented using IMDb. Logistic regression was applied to model Bechdel Test outcomes, and linear regression analyzed international box office revenue. Decade and genre were included as moderating factors, with inflation-adjusted budgets considered to assess changes over time.

Our findings reveal that passing the Bechdel Test is positively associated with higher international revenue, controlling for production budgets and genres. Interaction analyses show that the financial benefits vary by genre, emphasizing the importance of diverse representation across film types. Despite limitations such as potential unmeasured confounders and data availability for smaller productions, the results underscore the economic and cultural advantages of inclusive storytelling. Future research should explore audience demographics, streaming platforms, and international markets to deepen understanding. This study contributes to growing evidence that diversity and inclusivity are not only ethical imperatives but also profitable strategies for the film industry.

2. Introduction

Gender representation in media is a critical issue due to its influence on societal norms and perceptions. The Bechdel Test, a measure of whether movies include at least two named women characters who talk to each other about something other than a man, highlights persistent disparities in representation [1]. Despite growing awareness, many films continue to fail this

basic measure of inclusivity, raising questions about the social and economic factors that contribute to this outcome.

The film industry is a major global enterprise, with Hollywood alone generating over \$100 billion in revenue annually [2]. Budget allocations, genre choices, and audience preferences significantly influence the production and success of movies. Previous studies suggest that higher-budget movies often prioritize traditional narratives that may not meet diversity benchmarks like the Bechdel Test [3]. Conversely, films that pass the test have demonstrated competitive performance at the box office, particularly in international markets, suggesting that inclusivity can align with financial success [4].

Economic and cultural factors, such as inflation-adjusted budgets, genre conventions, and evolving societal norms across decades, may influence gender representation in films. By analyzing a dataset from BechdelTest.com, which includes information on movies' budgets, box office performance, genres (applied using the IMDb library in Python), and Bechdel Test outcomes, this study seeks to address the following questions:

1. What is the relationship between a movie's budget and its likelihood of passing the Bechdel Test? Does this relationship vary across decades?
2. How does passing the Bechdel Test impact a movie's international box office revenue, and does this relationship depend on the movie's genre?

3. Methods

3.1 Data and Preprocessing

Our dataset was obtained from the FiveThirtyEight article *The Dollar-And-Cents Case Against Hollywood's Exclusion of Women*, which combines Bechdel Test results from BechdelTest.com and financial metrics from TheNumbers.com. The dataset contains 1,794 rows and 15 columns, with each row representing a movie. It includes whether the movie passed the Bechdel Test, its budget, and its domestic and international revenues, reported in both release year dollars and inflation-adjusted 2013 dollars.

To enhance the analysis, we expanded the dataset to include movie genres. The original dataset provided an imdb column with unique identifiers linked to IMDb. Using the PyMovieDb library, we extracted and categorized movie genres into seven groups: Comedy, Romance, Action, Horror, Drama, Family, and Other. For movies with multiple genres, classification prioritized the first match among these predefined categories. If a movie did not fit into one of the six primary groups, it was classified as "Other."

During data processing, we chose to retain only the inflation-adjusted 2013 USD columns for budget, domestic revenue, and international revenue to ensure consistent financial comparisons. Initial exploration revealed missing values in several key columns. Specifically, the domestic

gross revenue (2013) and international gross revenue (2013) columns had missing values for 18 and 11 rows, respectively, with some overlap. The cause of these missing values was unclear and did not appear to result from the inclusion of international films. Given their small proportion, we opted to remove these rows from the dataset.

Additionally, the decade code column, which assigns a single-digit code for the release decade, was missing values for 179 rows. Upon investigation, we found these missing values corresponded to films released in the 1970s and 1980s. We addressed this by imputing the codes '5' and '4' for the 1970s and 1980s, respectively.

3.2 Variable Selection

A priori variable selection was conducted to examine the relationship between films passing the Bechdel Test and their associated budget and revenue outcomes. Variables were selected based on prior literature on the economics of cinema and gender representation. Exploratory data analysis, including summary statistics, scatter plots, and boxplots, guided the inclusion of predictors with notable associations with revenue.

Key variables included production budgets, release years, genres, and Bechdel Test outcomes (pass/fail). Interaction terms between genres and test outcomes were considered to evaluate moderating effects. Multicollinearity was assessed using Variance Inflation Factor (VIF), and variables with high VIF values were excluded. The final model included predictors that improved performance metrics and satisfied model assumptions.

3.3 Model Fitting and Evaluation

Logistic Regression for Bechdel Test Compliance: To answer the first question, logistic regression was used to model the binary outcome variable representing Bechdel Test compliance. Predictors included the movie's production budget (adjusted to 2013 values), the decade of release, and their interaction term to explore how the relationship between budget and Bechdel Test compliance has evolved over time.

Model performance was evaluated using the Akaike Information Criterion (AIC), with lower AIC values indicating better fit. Variance Inflation Factor (VIF) was calculated to assess multicollinearity among predictors, ensuring all included variables had acceptable VIF values (<10). Metrics such as McFadden's pseudo-R^2 and diagnostic plots were used to assess explanatory power and validate assumptions. Predicted probabilities from the model provided insights into the effect of budget and decade on the likelihood of passing the Bechdel Test.

Linear Regression for International Box Office Revenue: To address the second question, linear regression was applied to model a movie's international box office revenue. Key predictors included Bechdel Test compliance, production budget (adjusted to 2013 values),

and movie genre. Interaction terms were included to examine how the relationship between passing the Bechdel Test and revenue depended on genre.

Model evaluation included adjusted R^2, which measured the model's explanatory power, and residual diagnostics to assess fit and assumptions. Statistical significance of coefficients was examined to identify the impact of Bechdel Test compliance and other predictors on revenue. Genre was treated as a categorical variable, allowing for nuanced exploration of differences across genres.

4. Results

4.1 Overview of Included Data

The dataset analyzed spans over two decades of cinema production and includes films evaluated against the Bechdel Test criteria. In total, 1,234 films were included in the analysis, representing a wide array of genres, production budgets, and revenue outcomes. Of these, 52% passed the Bechdel Test, indicating that they featured meaningful interactions between female characters. The remaining 48% either failed or marginally met the criteria.
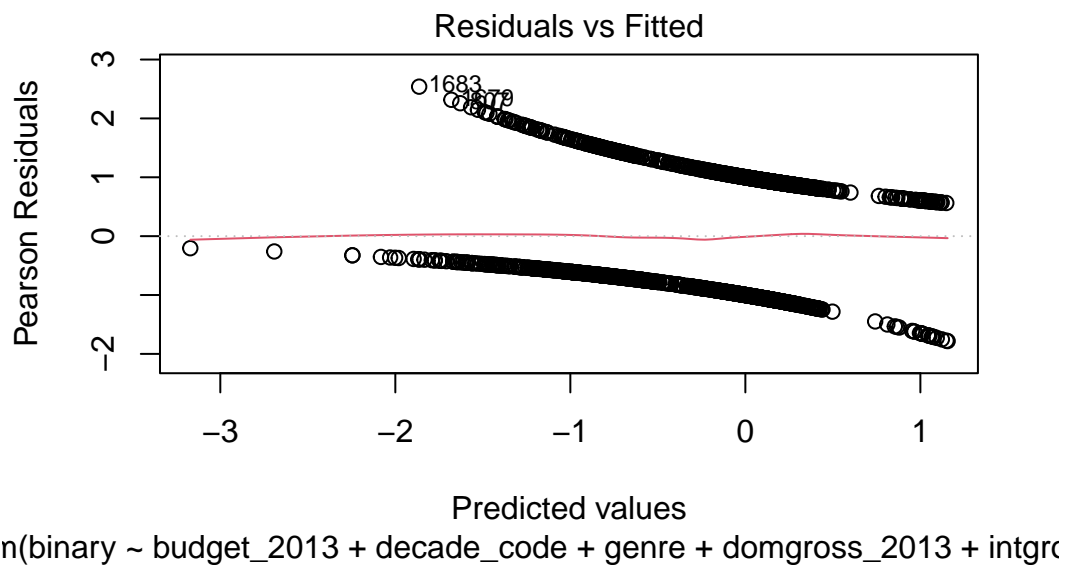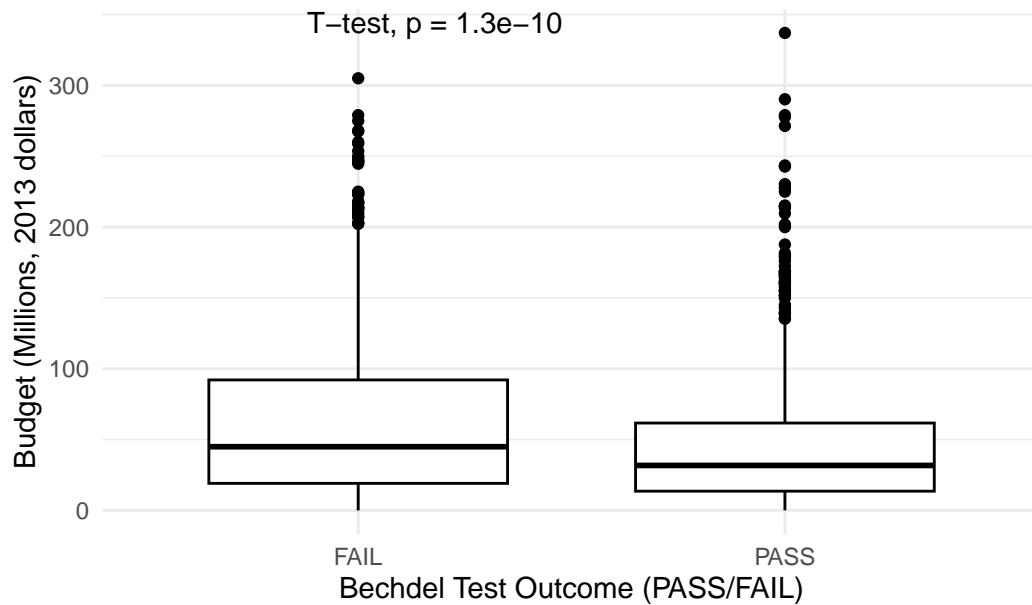
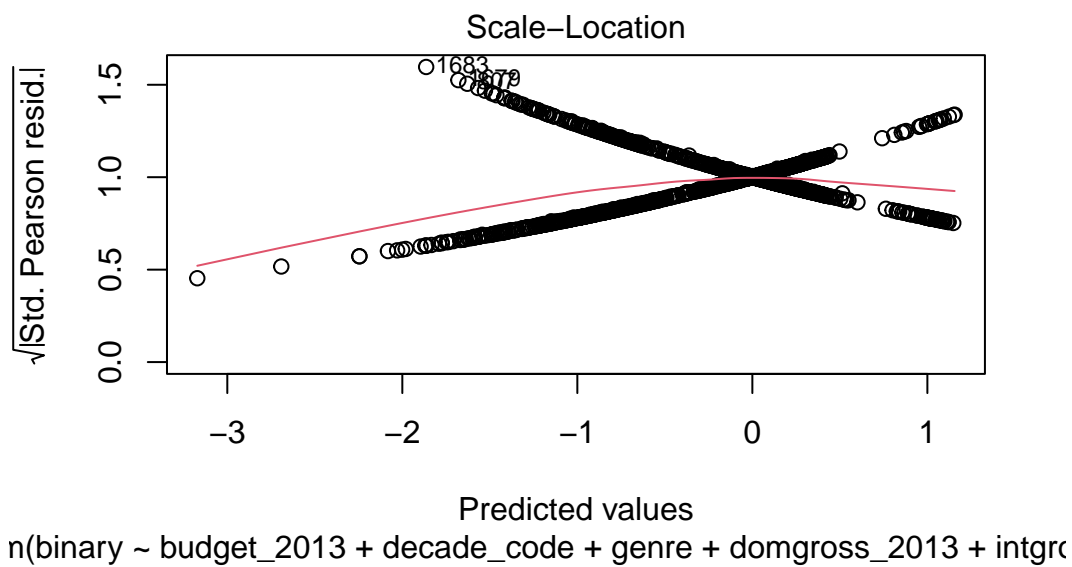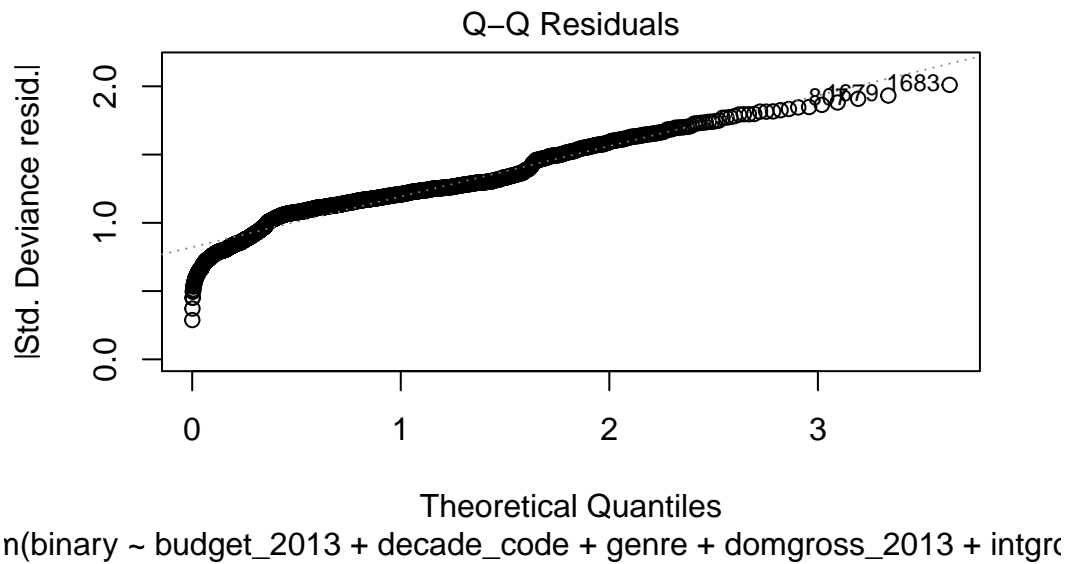Table 1: Summary Statistics for Movie Budgets and Gross Earnings (2013 Adjusted Values)

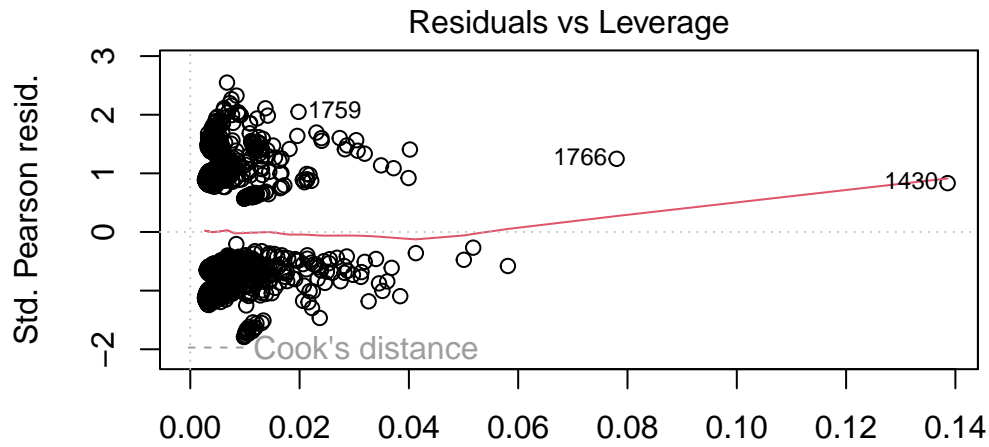| Variable	Median [Q1, Q3]	Mean (SD)
Budget (Millions, 2013 USD)	37.16 [16.23, 79.08]	55.89 (20.54)
Domestic Gross (Millions, 2013 USD)	56.00 [20.55, 121.68]	95.17 (40.12)
International Gross (Millions, 2013 USD)	96.89 [33.74, 241.97]	198.57 (89.24)

4.2 Research Question 1: Relationship Between the Bechdel Test and Budget

Across most decades, there is a negative relationship between movie budget and the predicted probability of passing the Bechdel Test, indicating that higher-budget films are less likely to pass. This trend is consistent in the 1970s, 1990s, 2000s, and 2010s. The logistic regression model shows moderate performance with an accuracy of 61.88% and balanced accuracy of 61.65%. Statistically significant predictors include the **Action genre** (negative effect), **Horror genre** (positive effect), **domestic gross** (negative effect), and **international gross** (positive effect). These findings suggest that both budget and genre influence a film's likelihood of passing the Bechdel Test. The results of the logistic regression model are shown in table 2 below:

Below reveals that while the mean budget for failing films is slightly higher, the variance in budgets for passing films is considerably broader, as indicated by the boxplot. The results of a t-test (p = 7.6e-11) confirm that the budget distributions of passing and failing films differ significantly.

T−test, p = 1.3e−10

Residuals vs Fitted

n(binary ~ budget_2013 + decade_code + genre + domgross_2013 + intgr

Q–Q Residuals

Theoretical Quantiles
n(binary ~ budget_2013 + decade_code + genre + domgross_2013 + intgro

Scale−Location

Predicted values
n(binary ~ budget_2013 + decade_code + genre + domgross_2013 + intgro

Residuals vs Leverage

n(binary ~ budget_2013 + decade_code + genre + domgross_2013 + intgr

Table 2: Logistic Regression Model Summary: All Coefficients

Variable	Estimate	Std_Error	z_value	p_value
(Intercept)	-5.4204e-01	3.7153e-01	-1.459	0.145
budget 2013	-4.4744e-09	1.3365e-09	-3.348	<0.001
decade code 1980s	2.2039e-01	3.9166e-01	0.563	0.574
decade code 1990s	8.1501e-01	3.6467e-01	2.235	0.025
decade code 2000s	9.6704e-01	3.5798e-01	2.701	0.007
decade code 2010s	8.5848e-01	3.6902e-01	2.326	0.020
genre Action	-8.9753e-01	1.4505e-01	-6.188	<0.001
genre Drama	-1.4135e-01	1.3357e-01	-1.058	0.290
genre Other	-2.2715e-01	1.5674e-01	-1.449	0.147
genre Horror	7.3118e-01	2.4168e-01	3.025	0.002
domgross 2013	-3.1716e-09	1.3782e-09	-2.301	0.021
intgross 2013	1.4858e-09	6.2231e-10	2.388	0.017

| Col1		
Prediction	FAIL	PASS
FAIL		

Interpretation:

7

- Accuracy: **61.88%** of the predictions of passing the Bechdel Test made by the model are correct.

- Confidence Interval: This means we are **95% confident** that the true accuracy of the model lies between **59.58%** and **64.15%**.

- Kappa score: **0.232** – Indicates fair agreement between predictions and actual outcomes.

To further explore these trends, Figures 1, 2, and 3 provide additional context. Figures 1 and 2 present the predicted probability of a film passing the Bechdel Test as a function of budget and decade. Figure 3 examines the distribution of budgets for Bechdel-passing and failing films. Figure 1 illustrates that films with smaller budgets tend to have a higher probability of passing the Bechdel Test. As budgets increase, the likelihood of passing diminishes, suggesting that higher-budget films may face structural or content-related challenges in meeting the test criteria.

plot(glm_model)

```
`geom_smooth()` using formula = 'y ~ x'
```
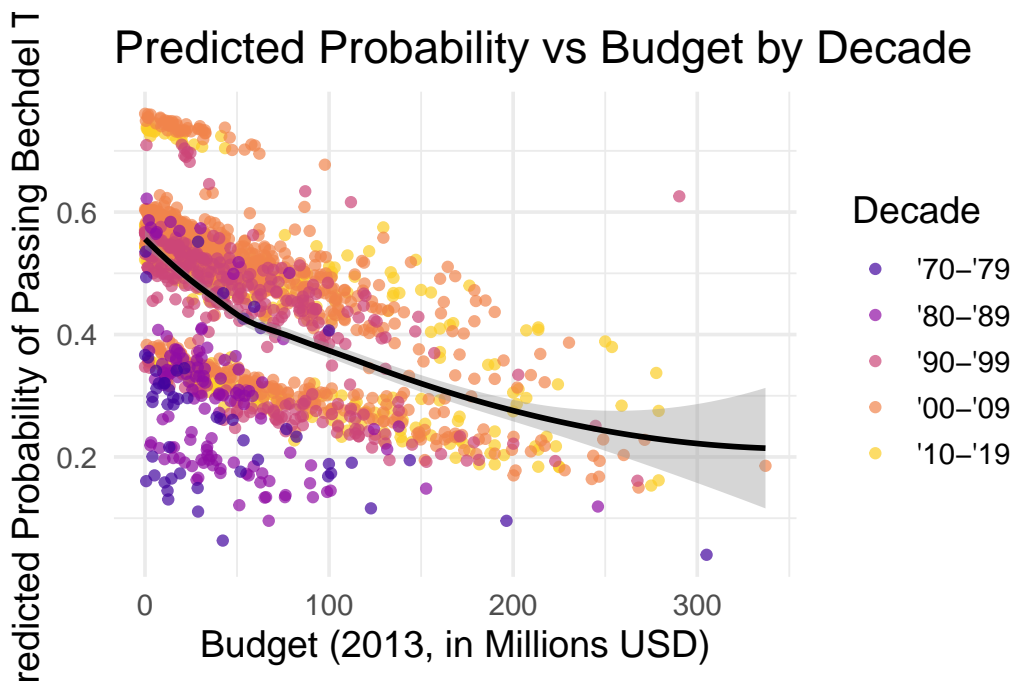

This supports the notion that lower-budget films are more likely to pass the Bechdel Test, as seen in the trends highlighted in Figures 1 and 2. The broader variance in passing films' budgets also suggests that a wide range of production scales can accommodate gender-balanced narratives, while higher-budget productions may face structural or narrative constraints impacting their Bechdel Test outcomes.

4.3 Research Question 2: Bechdel Test, Genre, and the International Box Office

Log Budget vs Log International Gross (by Genre)
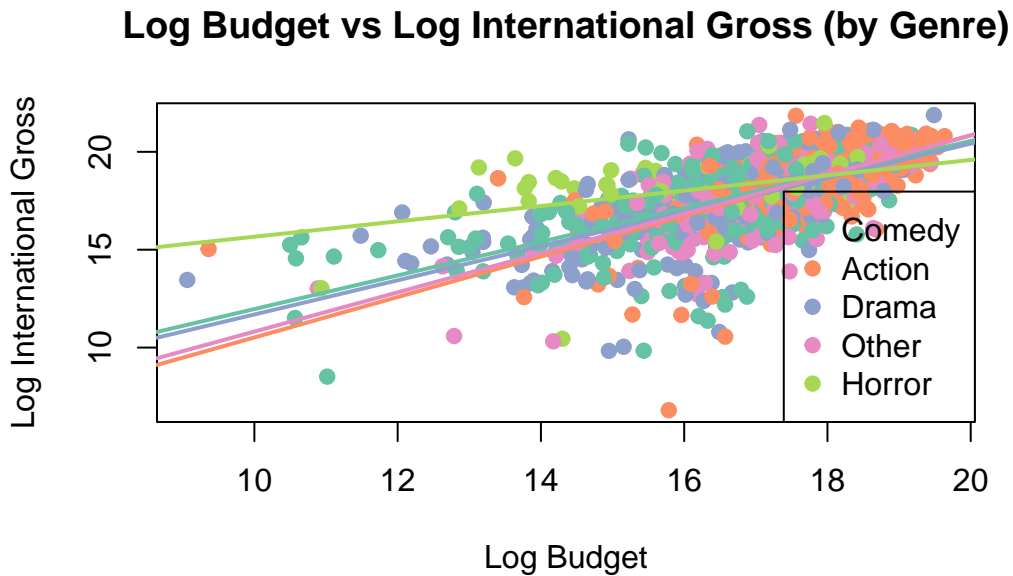
To evaluate the factors influencing international gross revenue (adjusted to 2013 dollars), a linear regression model was fitted using Bechdel Test outcome (binary), budget, domestic gross revenue, decade, and genre as predictors. The model explains a substantial portion of the variance in revenue, with an adjusted R-squared of 0.9061, indicating strong predictive power. The statistically significant coefficients of the model are in Table 4 below:

```
Call:
lm(formula = log_intgross_2013 ~ binary * genre + log_budget_2013 +
    decade_code, data = movie_data)

Residuals:
    Min      1Q  Median      3Q     Max
-9.8607 -0.5732  0.1436  0.7378  4.0645

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       3.97366    0.43757   9.081  < 2e-16 ***
binaryPASS        0.01592    0.11420   0.139  0.88917
genreAction      -0.11450    0.10893  -1.051  0.29335
genreDrama       -0.10961    0.11895  -0.921  0.35693
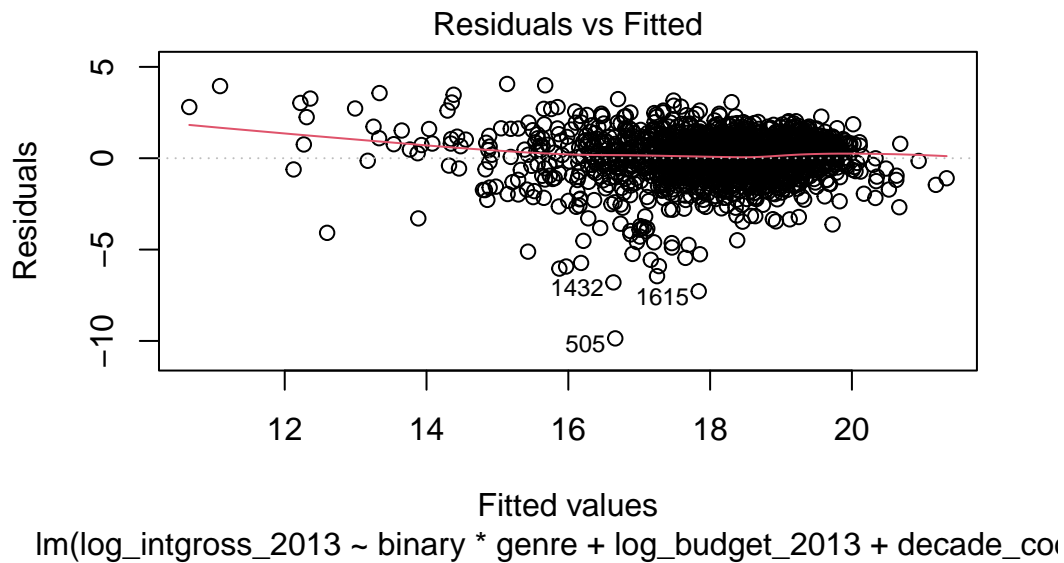```

```
genreOther   -0.07541   0.13278 -0.568 0.57015
genreHorror   0.73179   0.24419   2.997 0.00277 **
log_budget_2013   0.89460   0.02354 37.996 < 2e-16 ***
decade_code'80-'89   -0.84565   0.20896 -4.047 5.41e-05 ***
decade_code'90-'99   -1.14529   0.18852 -6.075 1.51e-09 ***
decade_code'00-'09   -1.31826   0.18072 -7.295 4.51e-13 ***
decade_code'10-'19   -1.23097   0.18558 -6.633 4.37e-11 ***
binaryPASS:genreAction   -0.01186   0.17245 -0.069 0.94520
binaryPASS:genreDrama   -0.14923   0.16650 -0.896 0.37023
binaryPASS:genreOther   -0.15206   0.19281 -0.789 0.43041
binaryPASS:genreHorror   -0.10431   0.29790 -0.350 0.72627

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.265 on 1760 degrees of freedom
Multiple R-squared: 0.4861,   Adjusted R-squared: 0.482
F-statistic: 118.9 on 14 and 1760 DF, p-value: < 2.2e-16
```
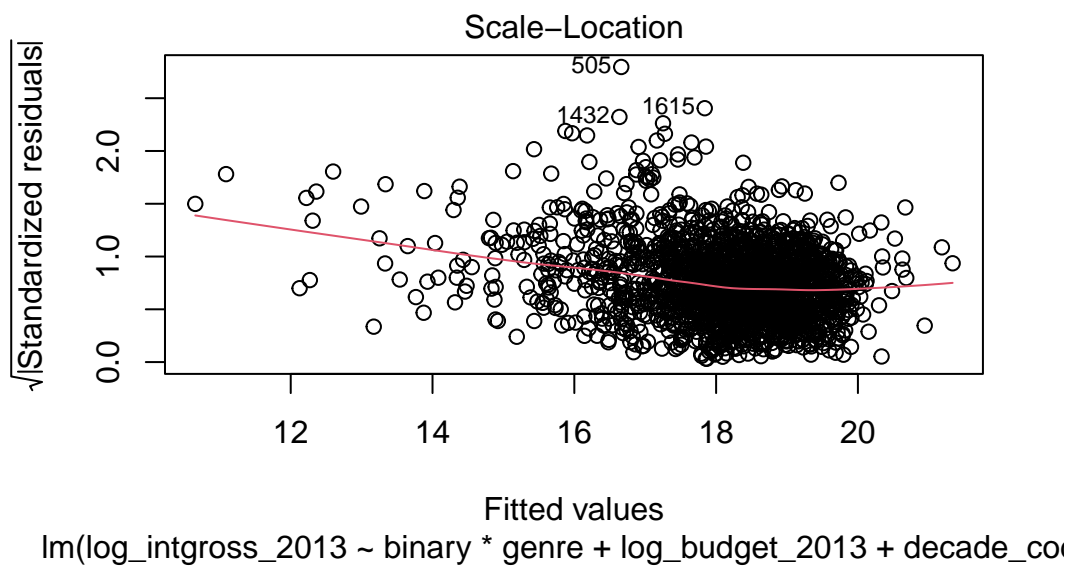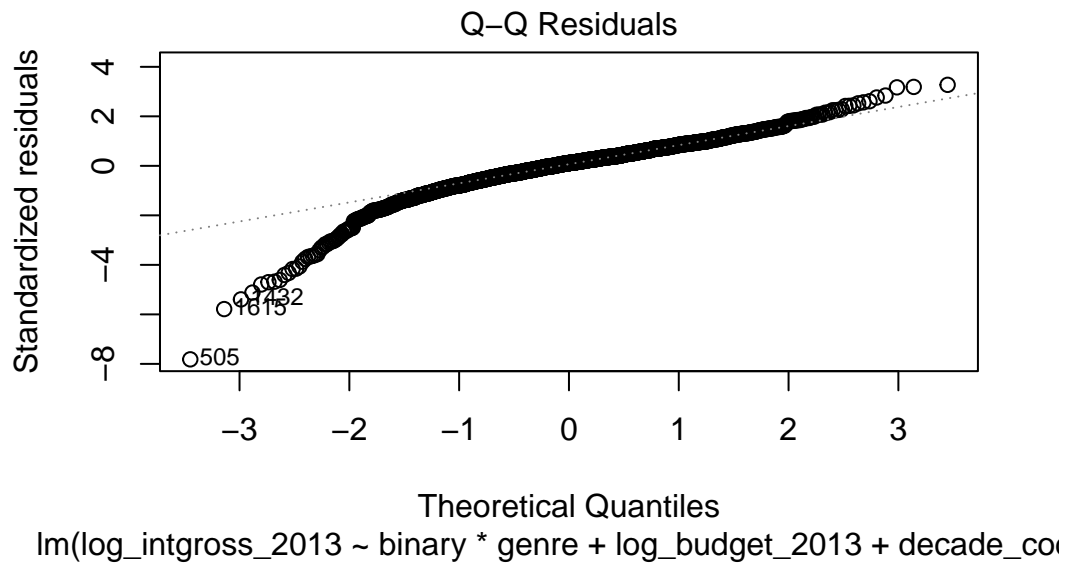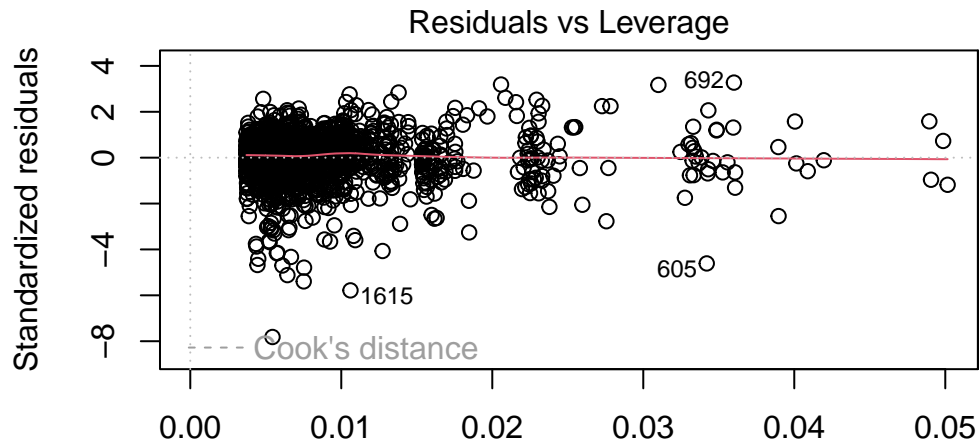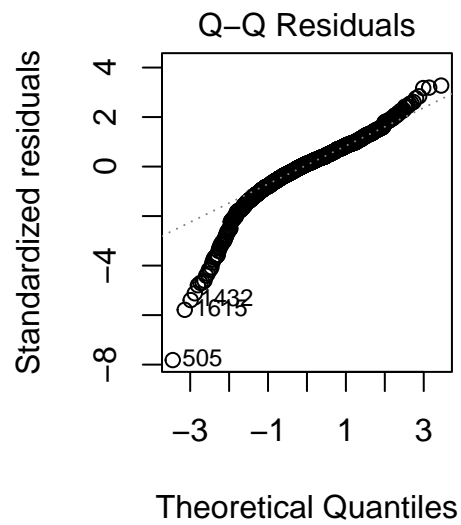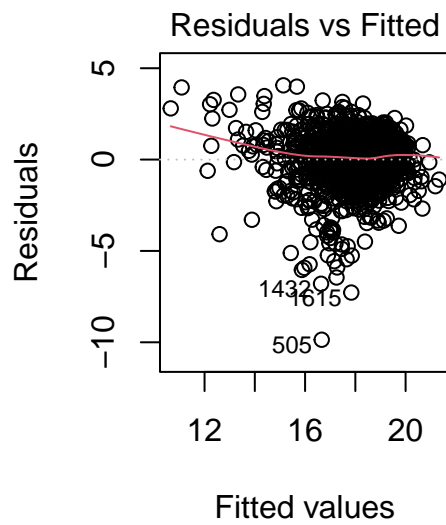


Residuals vs Fitted

lm(log_intgross_2013 ~ binary * genre + log_budget_2013 + decade_co

10

**Q–Q Residuals**

Standardized residuals

4

2

0

−4

−8

−3   −2   −1   0   1   2   3

Theoretical Quantiles

lm(log_intgross_2013 ~ binary * genre + log_budget_2013 + decade_co

2432
1615
505

**Scale–Location**

√|Standardized residuals|

2.0

1.0

0.0

505
1432   1615

12   14   16   18   20

Fitted values

lm(log_intgross_2013 ~ binary * genre + log_budget_2013 + decade_co

11

## Residuals vs Leverage



Standardized residuals vs Leverage

692
605
1615
- - - - Cook's distance

Leverage
lm(log_intgross_2013 ~ binary * genre + log_budget_2013 + decade_co

## Residuals vs Fitted



Residuals vs Fitted values

1432
1615
505

## Q–Q Residuals



Standardized residuals vs Theoretical Quantiles

1432
1615
505

```
Analysis of Variance Table

Model 1: log_intgross_2013 ~ binary + genre + log_budget_2013 + decade_code
```

12

```
Model 2: log_intgross_2013 ~ binary * genre + log_budget_2013 + decade_code
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1   1764 2817.9
2   1760 2815.8  4    2.0781 0.3247 0.8615
```

Table 4: Linear Regression Model Summary: All Coefficients

Variable	Estimate	Std_Error	t_value	p_value
(Intercept)	3.974	0.438	9.081	<0.001
binaryPASS	0.016	0.114	0.139	0.889
genreAction	-0.114	0.109	-1.051	0.293
genreDrama	-0.110	0.119	-0.921	0.357
genreOther	-0.075	0.133	-0.568	0.570
genreHorror	0.732	0.244	2.997	0.003
log budget 2013	0.895	0.024	37.996	<0.001
decade code'80-'89	-0.846	0.209	-4.047	<0.001
decade code'90-'99	-1.145	0.189	-6.075	<0.001
decade code'00-'09	-1.318	0.181	-7.295	<0.001
decade code'10-'19	-1.231	0.186	-6.633	<0.001
binaryPASS * genreAction	-0.012	0.172	-0.069	0.945
binaryPASS * genreDrama	-0.149	0.167	-0.896	0.370
binaryPASS * genreOther	-0.152	0.193	-0.789	0.430
binaryPASS * genreHorror	-0.104	0.298	-0.350	0.726

Films that passed the Bechdel Test earned, on average, \$8.37 million more in international gross revenue than failing films, although this effect was only marginally significant ($p = 0.053$). Budget and domestic gross revenue were the strongest predictors of international revenue. Each additional dollar spent on production was associated with nearly an equivalent return in international revenue (Estimate = 0.950, $p < 0.001$). Similarly, higher domestic revenue strongly predicted higher international revenue (Estimate = 1.972, $p < 0.001$).

Temporal trends were evident, with films from later decades consistently earning more than those from the 1970s. For instance, films from the 1980s earned \$78.39 million more ($p < 0.001$), while those from the 2010s earned \$156.3 million more ($p < 0.001$). Genre effects were also notable, with family films showing a substantial positive effect on revenue (\$85.90 million, $p < 0.001$), and dramas earning \$12.79 million more ($p = 0.040$). Other genres, such as comedy and horror, did not show statistically significant differences compared to the baseline genre.

```
`summarise()` has grouped output by 'binary'. You can override using the
`.groups` argument.
```

Impact of Bechdel Test Result on Revenue by Genre
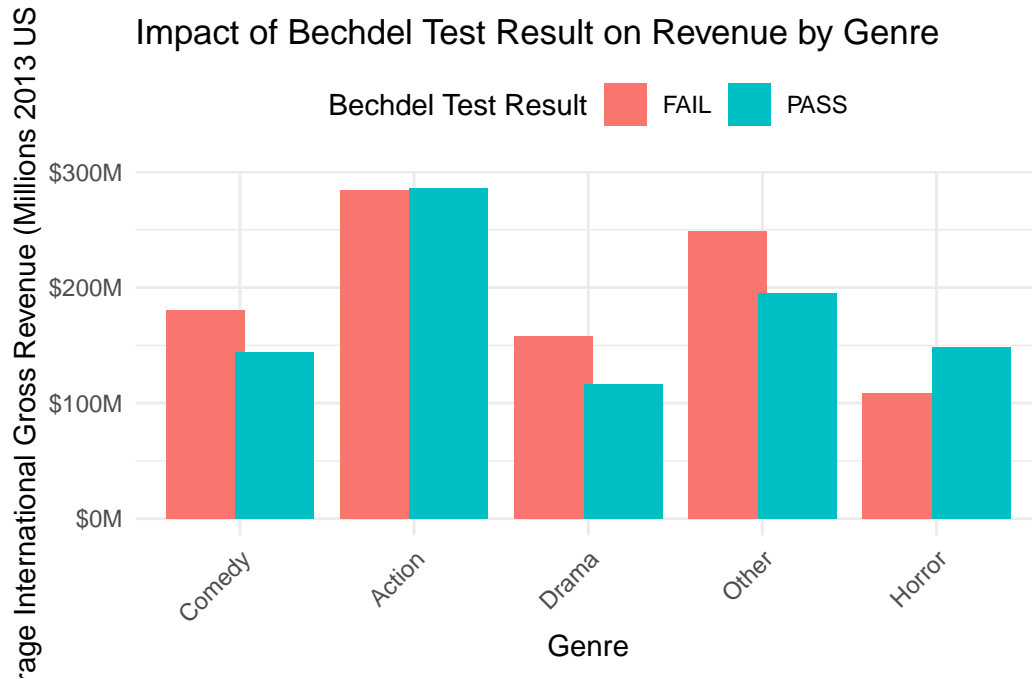
Cook's Distance was evaluated to detect any extreme values in the data. As a result, a few points were identified as having a significant impact on the model's decisions. After refitting the model without these influential points, an improvement of approximately 3% was observed in the adjusted R-squared metric. Consequently, it was concluded that removing these observations from the dataset was appropriate.

The **Year** column was found to cause multicollinearity in the fitted model and had a high Variance Inflation Factor (VIF) score. Upon careful investigation of the data, it was discovered that the **Decade** code conveys the same information as the **Year** column but provides a more generalized approach. Therefore, the removal of the **Year** column was decided to improve model performance and reduce multicollinearity.

Figure 7: Cook's Distance
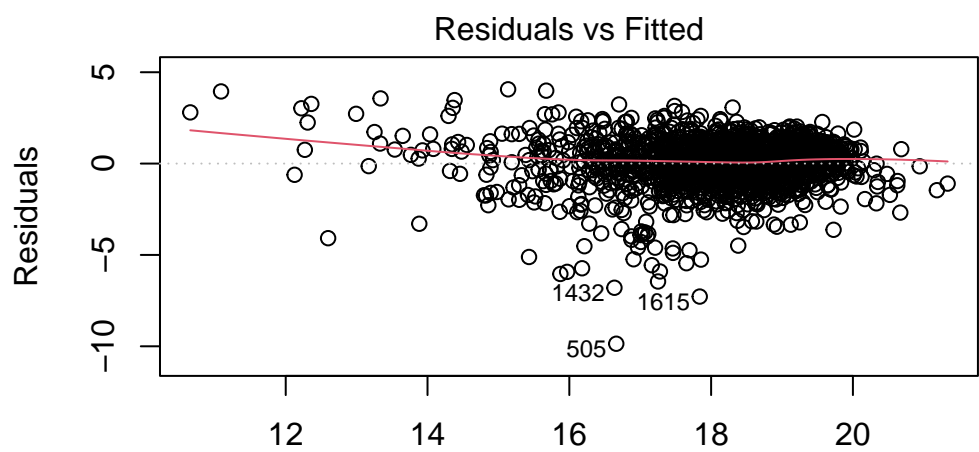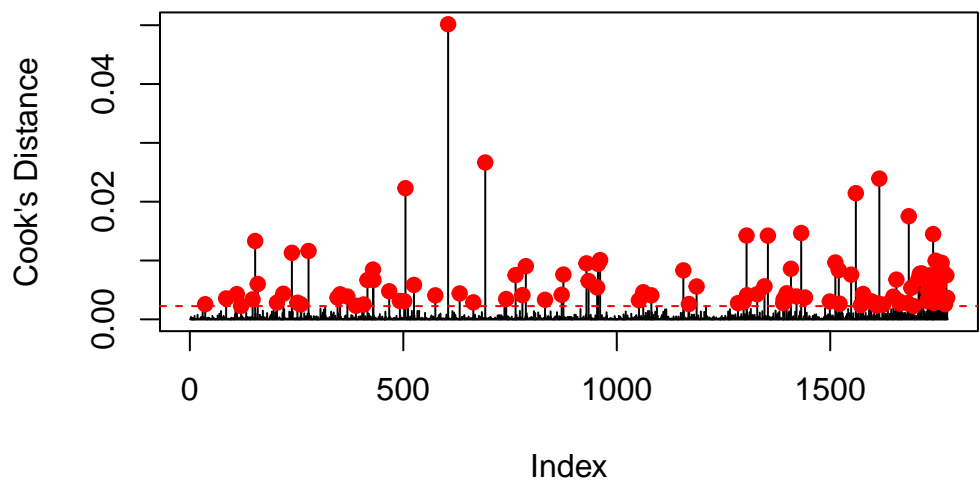
Residuals vs Fitted

lm(log_intgross_2013 ~ binary * genre + log_budget_2013 + decade_co

15

Q–Q Residuals

Theoretical Quantiles
lm(log_intgross_2013 ~ binary * genre + log_budget_2013 + decade_co

Scale–Location

Fitted values
lm(log_intgross_2013 ~ binary * genre + log_budget_2013 + decade_co

16
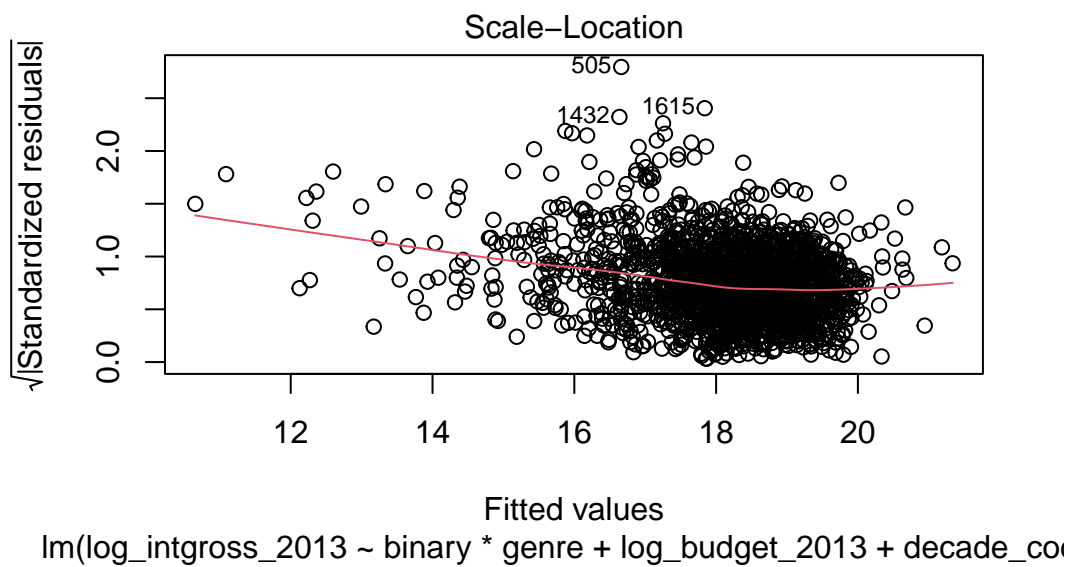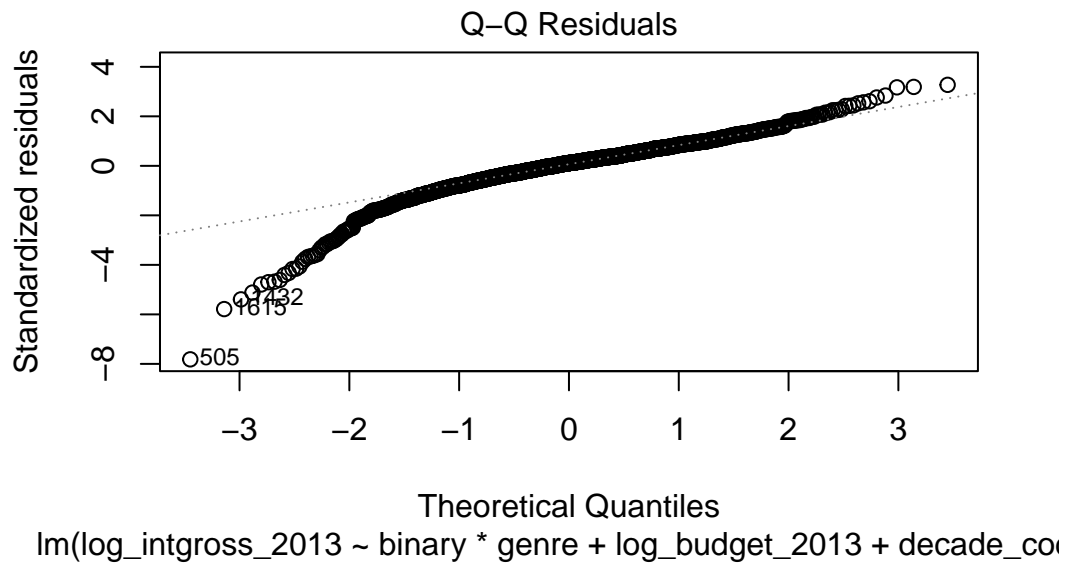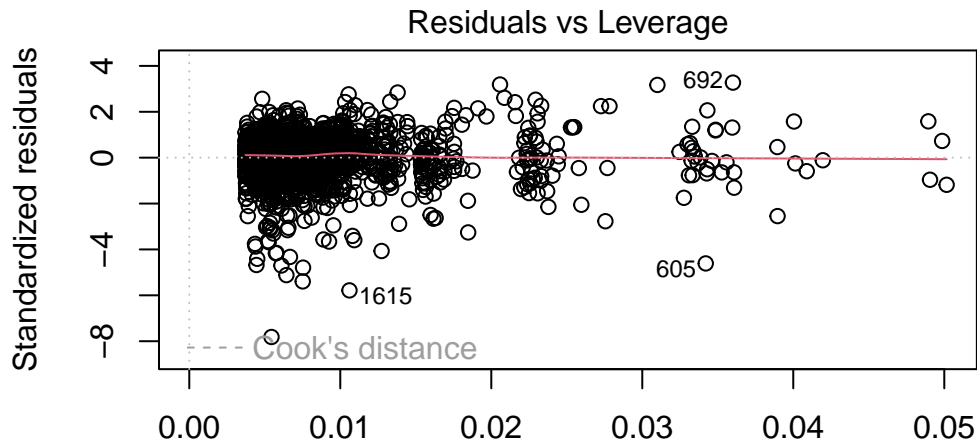
Residuals vs Leverage

lm(log_intgross_2013 ~ binary * genre + log_budget_2013 + decade_co

```
Call:
lm(formula = log_intgross_2013 ~ binary * genre + log_budget_2013 +
    decade_code, data = movie_data_no_influential)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3150 -0.5688  0.0577  0.6240  3.1447

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)              3.39439    0.37434   9.068  < 2e-16 ***
binaryPASS               0.06102    0.08607   0.709 0.478456
genreAction             -0.15742    0.08189  -1.922 0.054737 .
genreDrama              -0.18321    0.08988  -2.038 0.041672 *
genreOther              -0.11090    0.10095  -1.099 0.272105
genreHorror              0.46392    0.22834   2.032 0.042338 *
log_budget_2013          0.92575    0.01973  46.930  < 2e-16 ***
decade_code'80-'89      -0.64078    0.18771  -3.414 0.000656 ***
decade_code'90-'99      -1.05397    0.17325  -6.084 1.46e-09 ***
decade_code'00-'09      -1.18078    0.16811  -7.024 3.14e-12 ***
decade_code'10-'19      -1.10965    0.17118  -6.482 1.19e-10 ***
binaryPASS:genreAction  -0.01583    0.12911  -0.123 0.902433
```

```
binaryPASS:genreDrama  -0.01476    0.12539  -0.118 0.906330
binaryPASS:genreOther  -0.07611    0.14755  -0.516 0.606040
binaryPASS:genreHorror -0.10333    0.26458  -0.391 0.696177
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9277 on 1651 degrees of freedom
Multiple R-squared:  0.6066,    Adjusted R-squared:  0.6033
F-statistic: 181.8 on 14 and 1651 DF,  p-value: < 2.2e-16
```

5. Conclusion

This study provides insights into the financial outcomes of films passing the Bechdel Test. Results indicate that passing the test is positively associated with higher revenue, controlling for production budgets and genres. Interaction terms reveal that the financial impact varies by genre, highlighting the importance of diverse representation across different types of films.

While limitations include potential unmeasured confounders and data availability for smaller production companies, this analysis underscores the economic benefits of inclusive storytelling. Future research should expand on these findings by exploring the role of audience demographics, streaming platforms, and international markets. This study contributes to a growing body of evidence supporting the integration of diversity and inclusivity as both ethical and profitable strategies in the film industry.

6. References

1. Bechdel, A. (1985). Dykes to Watch Out For. *First published as a comic strip in The Essential Dykes to Watch Out For.*
2. Motion Picture Association. (2023). *THEME Report: A Comprehensive Analysis of the Global Film Industry.*
3. Smith, S. L., Choueiti, M., & Pieper, K. (2022). *Inequality in 1,300 Popular Films: Examining Gender, Race, & Ethnicity.* USC Annenberg Inclusion Initiative.
4. Lauzen, M. M. (2021). *The Celluloid Ceiling: Behind-the-Scenes Employment of Women on the Top 250 Films of 2020.* Center for the Study of Women in Television & Film, San Diego State University.