

The Golden Age of Critics? Examining the Association Between Critic Scores and Audience Scores Over Time on Rotten Tomatoes

Ethan Weiland

2024-04-28

Introduction

The contemporary United States is experiencing historically low levels of trust. Surveys show that Americans do not trust newspapers, universities, the Supreme Court, and other key institutions in society. This analysis explores whether this distrust extends to the cultural sphere. Specifically, I look at the opinions of the general public and critics on a key cultural good: movies. Using data from Rotten Tomatoes, two research questions are asked:

1. What is the historical association between critic opinion and public opinion for films?
2. Has the association between critic opinion and public opinion for films changed over time, especially during the 21st century?

Data

The data for this analysis comes from the popular website Rotten Tomatoes. Rotten Tomatoes contains information on a plethora of movies, and for each movie it reports a “Tomatometer” (average critic score) and an “Audience Score” (average audience score). Additionally, Rotten Tomatoes contains information on other movie characteristics like release date, genre, run time, etc. A Kaggle user scraped Rotten Tomatoes and made the data available at the following link: <https://www.kaggle.com/datasets/andrezaza/clapper-massive-rotten-tomatoes-movies-and-reviews/>. This repository contains two tables used in this analysis: a first table on movies and a second table on individual critic reviews of these movies.

The dependent variable of interest is public opinion. Public opinion is operationalized as “Audience Score” for each movie, which is the percentage of users on Rotten Tomatoes that rated a movie positively.

The first independent variable of interest is critic score. A “TomatoMeter” measure is available for each movie, which is a measure of aggregate critic opinion ranging from 0% (wholly negative) to 100% (wholly positive). However, the definition of “critics” used in the available TomatoMeter is too broad. This analysis instead calculates a modified TomatoMeter that only includes reviews from critics classified as “Top Critics”. Top Critics are critics that work for established and well-known publications (e.g., The New York Times, Indy Star, etc.), who best represent elite opinion. To calculate this modified TomatoMeter, the table of critic reviews was used. Each review is classified as “Fresh” (indicating a positive review) or “Rotten” (indicating a negative review). The critic score is equal to the percentage of “Fresh” reviews. To avoid one critic’s review dominating, only movies with at least five Top Critic scores were included.

The second independent variable of interest is release date. Two potential dates are available from the Rotten Tomatoes data: the release date in theaters and the release date on streaming. The release date in theaters is chosen because films in theaters are more likely to be seen and reviewed by critics than films released straight to streaming (especially before the COVID pandemic). Release date is an ordinal measure grouped by decade with the following levels: “Pre-1980”, “1980s”, “1990s”, “2000s”, “2010s”, and “2020s”. All films before 1980 had to be collapsed into one category due to sparse data. Unless otherwise specified, “2010s” is the baseline category.

Various control variables are included. Genre is a nominal variable with 18 categories (see Descriptive Statistics table for a full list). If a film had multiple genres listed, the first genre was used. Language is equal to the original language of the film: “English” or “Non-English”. Run time refers to the length of the film (in minutes). Review count refers to the number of Top Critic reviews for each film. As described above, the minimum review count value is 5.

Missing data was imputed via bagging using the `caret` package in R. The dependent variable (audience score) was included in the imputation equations, but the imputed audience score values were not used. Films with a missing value on audience score were dropped. The final analytical sample size is 17670 films. Table 3 lists the means, standard deviations, minimums, and maximums for each variable just described.

Table 1: Descriptive Statistics, n=(17670)

	Mean/Proportion	StdDev	Min.	Max.
Audience Score	63.67	20.09	0	100
Critic Score	62.10	29.71	0	100
Pre-1980	0.06		0	1
1980s	0.03		0	1
1990s	0.14		0	1
2000s	0.37		0	1
2010s	0.33		0	1
2020s	0.07		0	1
Runtime (minutes)	103.27	20.80	5	584
Review Count	21.80	18.28	5	166
Language (1=Non-English)	0.22		0	1
Drama	0.25		0	1
Action/Adventure	0.07		0	1
Biography	0.02		0	1
Comedy	0.18		0	1
Crime	0.03		0	1
Documentary	0.15		0	1
Fantasy	0.01		0	1
History	0.02		0	1
Holiday	0.01		0	1
Horror	0.05		0	1
Kids and Family	0.04		0	1
Musical	0.01		0	1
Mystery and Thriller	0.07		0	1
Other	0.001		0	1
Romance	0.05		0	1
Sci-fi	0.02		0	1
War	0.01		0	1
Western	0.01		0	1

Source: Rotten Tomatoes

Analysis

Association Between Critic Opinion and Public Opinion

Figure 1. Distributions of Audience and Critic Scores

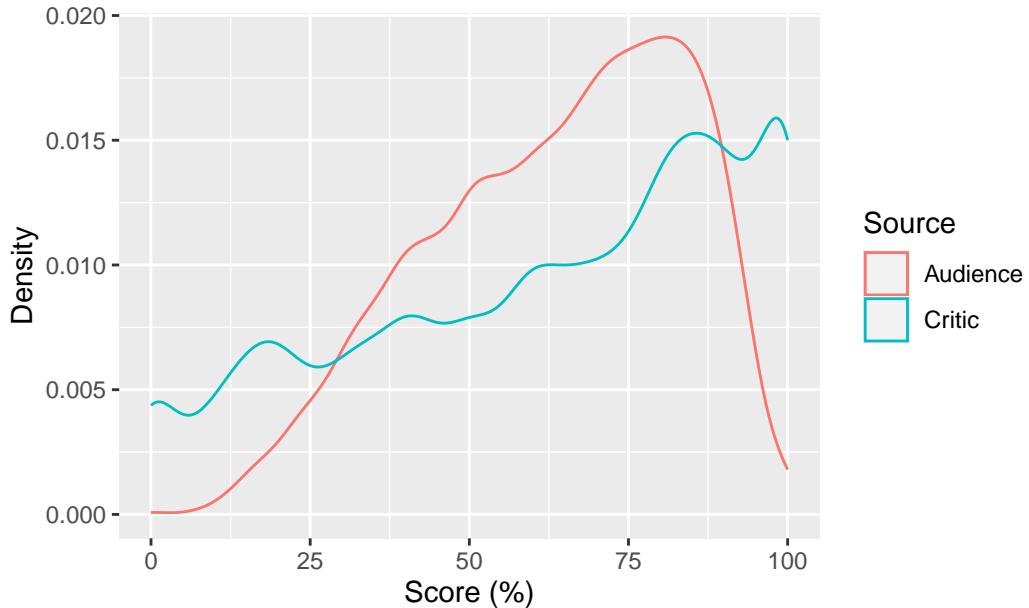
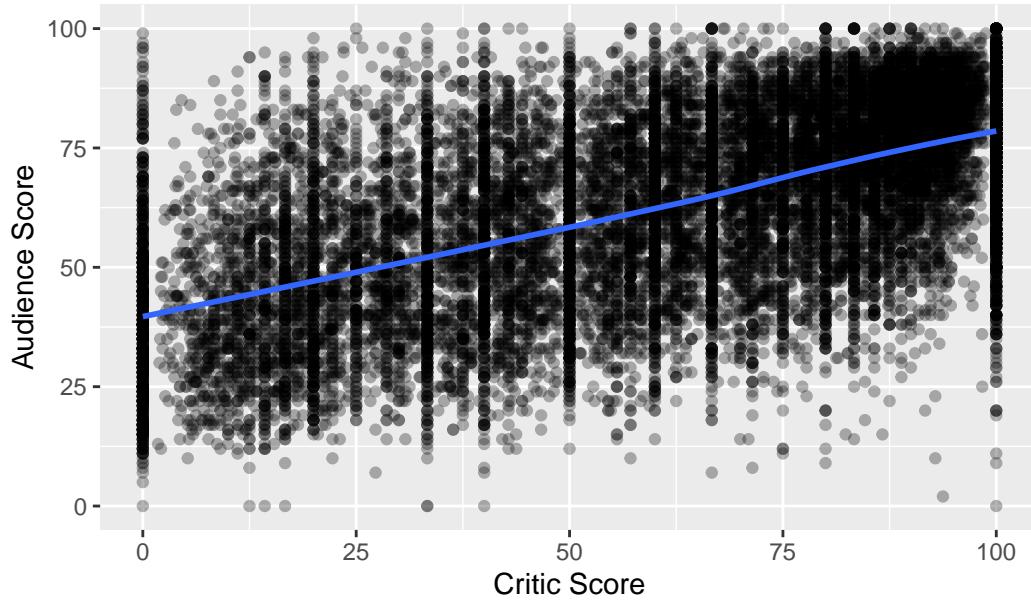


Figure 1 plots the univariate distributions of audience score and critic score. It is clear that critic scores have a greater spread than audience scores. Critic scores make use of the entire range of possible scores while audience scores are highest at around 70-80%. There are films with unanimous critic disapproval, while almost no films have unanimous audience disapproval. One possible explanation is a self-selection effect. Audience members won't voluntarily see a film they know they will dislike, but critics are tasked with seeing many films, regardless of their individual taste. The mean of audience score is 63.67 and the mean of critic score is 62.1. While a t-test concludes that these means are statistically different ($t=5.8051$, $p < 0.001$), this is not much practical difference.

Figure 2. Audience Scores vs. Critic Scores Across All Years



To begin answering the first research question, Figure 2 plots audience score vs. critic score for all films. The blue line is a LOESS smoother, which is a kind of local regression that helps visualize patterns when a large number of points are plotted. The LOESS smoother is strikingly straight. In fact, when a linear regression line is plotted it is difficult to see because it is nearly identical with the LOESS line. The correlation between audience score and critic score is 0.59 . There is a moderately-strong positive linear relationship between audience score and critic score for all films.

Next, audience score is modeled with critic score as the key independent variable. An ordinary least squares (OLS) model is used. OLS is appropriate because the dependent variable is a continuous measure and exploratory plotting revealed that the association between the two variables is linear. Further, OLS produces the best linear unbiased estimates. The model for audience score with critic score and other important control variables (Model 1) is formulated as follows:

$$AudienceScore = \beta_0 + \beta_1 CriticScore + \beta_{2-6} Decade + \beta_7 Runtime + \beta_8 Runtime^2 + \beta_9 ReviewCount$$

$$+ \beta_{10} Language + \beta_{11-27} Genre + \epsilon_i$$

$$\text{Where } \epsilon_i \sim N(0, \sigma^2)$$

The model includes a quadratic term for run time because it is theorized that there is a sweet spot for film length. Audiences like films as they get longer up until a certain point when they get bored. The model contains a large number of parameters due to the large number of levels on the decade and genre factor variables. Table 2 provides the results of this model.

The model fits well. The residual plot is null and the residuals themselves are homoscedastic and Normal (see Appendix). The model explains a substantial portion of the variation in audience scores (Adjusted $R^2 = 46.08\%$). Both of these facts provide confidence that parameter estimates are accurate. The coefficient for critic score is 0.295 ($p < 0.001$). A 10 point increase in critic score is significantly associated with a 2.95 increase in audience score conditional on other variables. Or, a one standard deviation increase in critic score is associated with a 8.75 increase in audience score conditional on other variables.

Association Over Time

The second research question is: Has the association between critic opinion and public opinion for films changed over time, especially during the 21st century? To do so, this analysis will reexamine the association between critic score and audience score in each decade (Pre 1980s, 1980s, 1990s, 2000s, 2010s, 2020s).

Figure 3. Audience Scores vs. Critic Scores By Decade

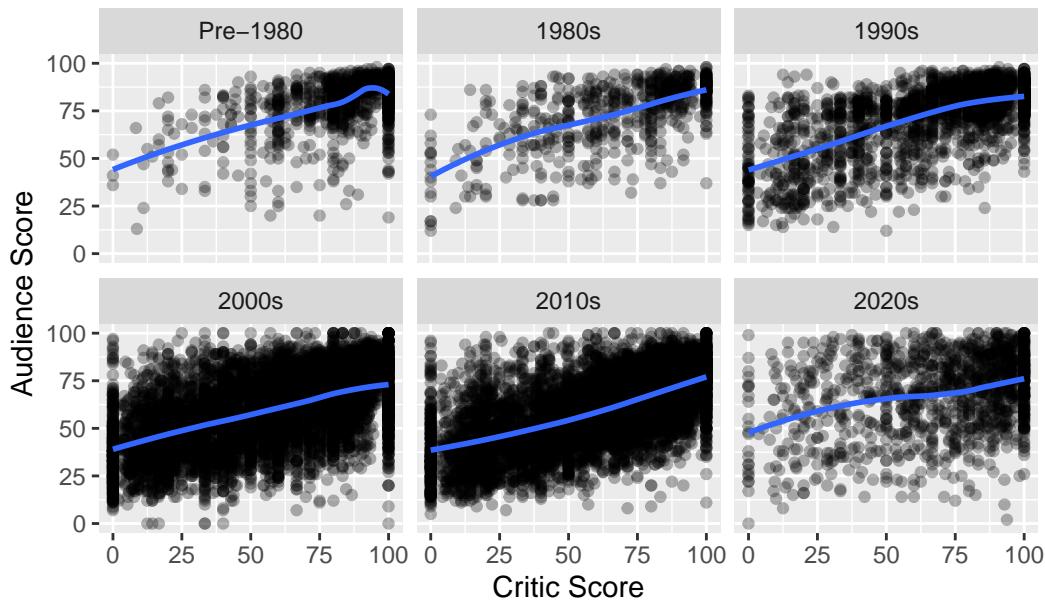


Figure 3 plots audience score vs. critic score in each decade. The blue line is a LOESS smoother. Again the LOESS smoother is straight, indicating a linear association between critic score and audience score in each decade. Until 2020, the lines appear parallel which implies that the association between audience score and critic score does not vary by decade.

Table 2: OLS Model for Audience Scores (Model 1)

	<i>Dependent variable:</i>
	Audience Score
Critic Score	0.29*** (0.004)
1980s	14.56*** (0.66)
1990s	12.64*** (0.38)
2000s	2.17*** (0.27)
2020s	7.08*** (0.47)
Pre-1980	16.63*** (0.55)
Runtime	0.21*** (0.01)
Runtime Squared	−0.0003*** (0.0000)
Review Count	0.17*** (0.01)
Language (1=Non-English)	−2.49*** (0.30)
Action/Adventure	−2.96*** (0.50)
Biography	1.48 (0.80)
Comedy	−1.55*** (0.35)
Crime	−2.82*** (0.67)
Documentary	8.29*** (0.39)
Fantasy	0.25 (1.01)
History	2.03* (0.87)
Holiday	−0.16 (1.37)
Horror	−9.68*** (0.54)
Kids and Family	−0.44 (0.61)
Musical	−0.31 (1.15)
Mystery and Thriller	−6.74*** (0.48)
Other	8.46* (4.11)
Romance	−1.72** (0.54)
Sci-fi	−6.24*** (0.78)
War	2.75 (1.47)
Western	−2.62 (1.45)
Constant	21.79*** (1.12)
Observations	17,670
R ²	0.46
Adjusted R ²	0.46
Residual Std. Error	14.76 (df = 17642)
F Statistic	560.16*** (df = 27; 17642)

Note:

*p<0.05; **p<0.01; ***p<0.001

Source: Rotten Tomatoes

Table 3: Means and Correlations By Decade

	Mean Critic Score	Mean Audience Score	Correlation
Pre-1980	83.30	79.81	0.50
1980s	69.91	74.64	0.63
1990s	69.48	73.16	0.64
2000s	55.76	59.24	0.53
2010s	59.87	59.67	0.60
2020s	70.60	68.94	0.31

Source: Rotten Tomatoes

However, in the 2020s the slope of the LOESS smoother is noticeably weaker. Table 3 lists the mean critic score, mean audience score, and correlation between these two quantities in each decade. Until 2020, the correlation varied between 0.5 and 0.65. However in the 2020s the correlation is noticeably smaller: 0.3.

To formally test whether the association between critic score and audience score varies by decade, interaction terms are added to the previous model. Table 4 presents the output of this new model (Model 2). An F-Test for testing multiple fixed effects (the interactions) supports the model with the critic score by decade interactions over the model without them ($F=20.877$, $p < 0.001$). The new coefficient on critic score is 0.31 ($p < 0.001$). This effect is actually slightly larger than in the previous model. In the 2010s (the baseline category), a one standard deviation increase in critic score is significantly associated with a 9.32 increase in audience score, conditional on other factors.

Two critic score by decade interactions are significant: critic score by 2000s ($p < 0.001$) and critic score by 2020s ($p < 0.001$). This means that the association between critic score and audience score is significantly different in these two decades compared to the association in the 2010s. The association between a one standard deviation increase in critic score and audience score in the 2000s is 8.37. While this may be statistically different, this is not meaningfully different from the association in the 2010s (it is less than a point). In the 2020s, the association between a one standard deviation in critic score and audience score is 4.74. Unlike the 2000s, this is statistically different *and* meaningfully different ($\sim 50\%$ weaker) than the baseline association in the 2010s. For the first time, the association between audience score and critic score is weakening.

Table 4: OLS Model for Audience Score with Critic Score by Decade Interactions (Model 2)

	<i>Dependent variable:</i>
	Audience Score
Critic Score	0.31*** (0.01)
Pre-1980	14.05*** (2.10)
1980s	13.90*** (1.74)
1990s	11.42*** (0.91)
2000s	4.02*** (0.58)
2020s	17.75*** (1.25)
Runtime	0.21*** (0.01)
Runtime Squared	−0.0003*** (0.0000)
Review Count	0.17*** (0.01)
Language (1=Non-English)	−2.53*** (0.30)
Action/Adventure	−3.05*** (0.49)
Biography	1.36 (0.80)
Comedy	−1.58*** (0.35)
Crime	−2.95*** (0.67)
Documentary	8.26*** (0.39)
Fantasy	0.24 (1.01)
History	1.93* (0.87)
Holiday	−0.34 (1.37)
Horror	−9.73*** (0.54)
Kids and Family	−0.44 (0.61)
Musical	−0.26 (1.14)
Mystery and Thriller	−6.81*** (0.48)
Other	8.59* (4.10)
Romance	−1.77*** (0.54)
Sci-fi	−6.28*** (0.78)
War	2.77 (1.46)
Western	−2.52 (1.45)
Critic Score*Pre-1980	0.03 (0.02)
Critic Score*1980s	0.01 (0.02)
Critic Score*1990s	0.01 (0.01)
Critic Score*2000s	−0.03*** (0.01)
Critic Score*2020s	−0.15*** (0.02)
Constant	20.63*** (1.16)
Observations	17,670
R ²	0.46
Adjusted R ²	0.46
Residual Std. Error	14.71 (df = 17637)
F Statistic	478.56*** (df = 32; 17637)

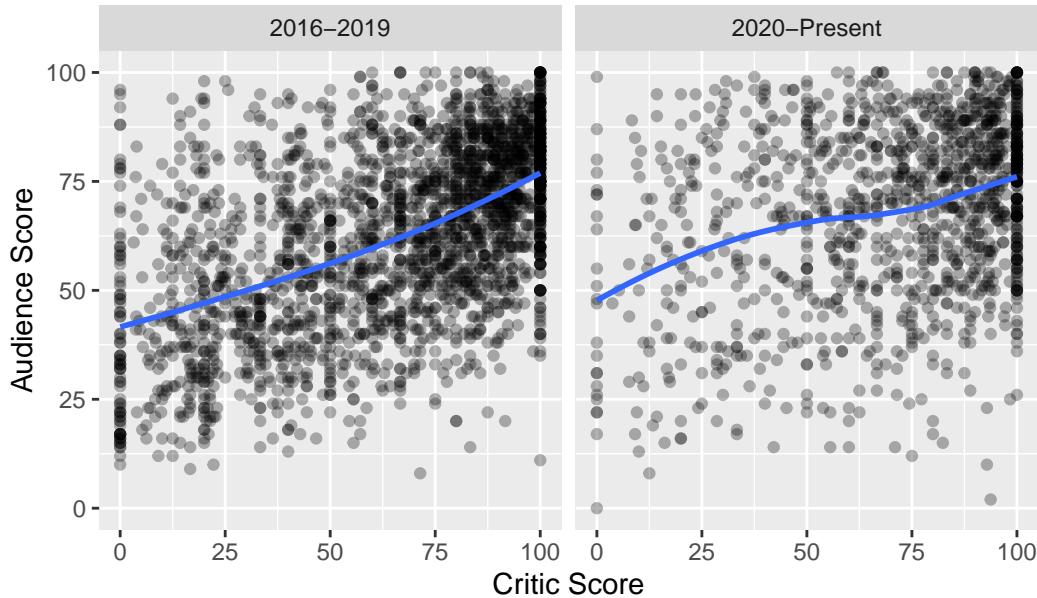
Note:

*p<0.05; **p<0.01; ***p<0.001

Source: Rotten Tomatoes

Pandemic Effect?

Figure 4. Audience Scores vs. Critic Scores Before/After 2020



The above analysis showed that the association between critic score and audience score remained relatively constant until 2020. After 2020, however, this association weakened. One theory why is a divergence between the general public and critics due to increasing societal distrust as discussed in the Introduction. Another possibility is that the COVID-19 pandemic brought about a massive shift in the movie industry, namely an increase in the number of films being released directly to streaming services rather than having a theatrical run. To determine whether this is the case, Figure 4 plots the association between audience score and critic score in the four years directly preceding the pandemic to the four years during and after the pandemic. By narrowing the time window it is easier to isolate any shift caused by COVID from a longer term trend.

There is a clear difference in the association between critic score and audience score in these two time windows. While still roughly linear, the association in 2020–Present is noticeably weaker. The correlation in the 2016–2019 time period is $\rho_{2016-2019} = 0.52$ and the correlation in the 2020–Present time period is $\rho_{2020-Present} = 0.31$. This provides preliminary evidence that the weakening after 2020 is due to changes in the movie industry brought on by the COVID pandemic, rather than a longer term trend.

Conclusion

Using data from Rotten Tomatoes, this report explored the association between audience scores and critic scores for a large number of films ($n = 17670$). The analysis resulted in the following three key findings:

1. The opinions of critics and the general public on movies has a moderately strong linear relationship.
2. This relationship did not meaningfully change over time until 2020.
3. After 2020 the relationship between critic scores and public scores weakened

Critics enjoy a powerful position in society, as arbiters of quality in movies. For decades, their opinions have largely aligned with the general public. But all “Golden Ages” must come to an end. The COVID-19 pandemic resulted in massive ramifications for the film industry. Coupled with an increase in general societal distrust, perhaps one unacknowledged change is an end to the “Golden Age” of critics.

GitHub Link

<https://github.com/ethanphilipweiland/rotten-tomatoes>

Appendix

Figure A1. Model 1 Residuals vs. Fitted Values

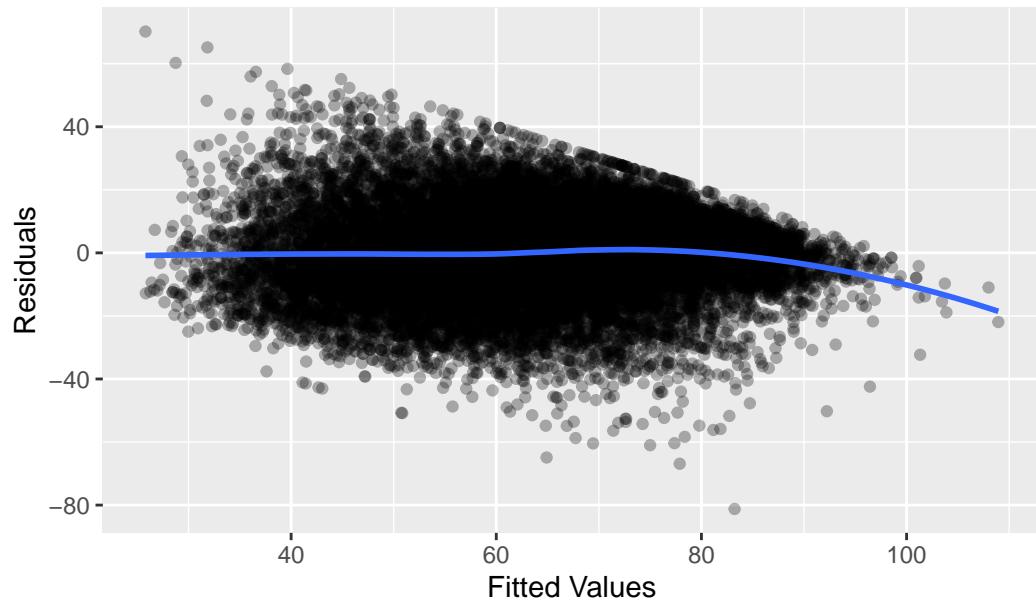


Figure A2. Model 1 Residuals Squared vs. Fitted Values

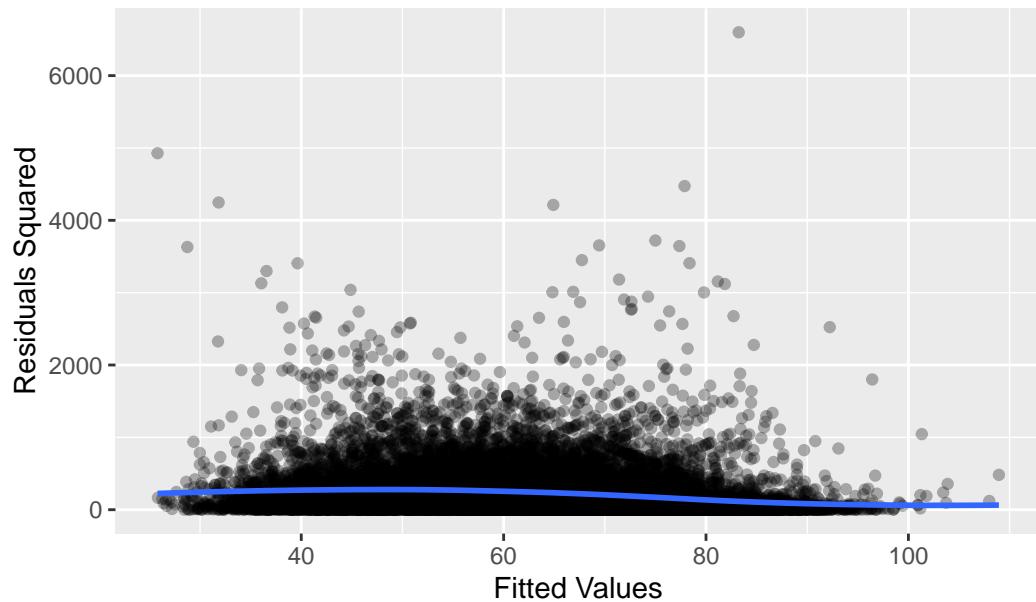


Figure A3. Normality of Model 1 Residuals

