Estimating Country Song Popularity from Duration

Arsh Hothi, Priya Reddy, Jocelyn Thai

12-06-2022

Introduction

Over the course of the last 10 years, an increasing proportion of U.S. music revenue has come from streaming platforms like Spotify. By 2018, 75% of music revenue came from streaming service, and streaming service revenue made up a majority of the industry's revenue growth[1]. As a result of this increased financial dependency, artists have begun optimizing their songs for repeated play on these platforms, and one of the easiest ways to do that is to make a song shorter. In theory, shorter songs can be replayed more often thus generating more revenue for their artists. While this idea may work in theory, it would take a data-based approach to understand whether there is actually a causal relationship between song length and performance on a streaming platform. A better understanding of this relationship could help labels and artists determine what kinds of songs to produce and how long they should be.

This study estimates the effect that song length has on song popularity on Spotify, and seeks to answer the question: Among the top tracks in the country music genre, does song length impact song popularity?

We are specifically interested in running this analysis on country songs, a genre which has historically been slower to change with trends[2], as it may have less variation in popularity due to other non-length related trends. Holding genre steady across our analysis also allows us to account for the differences in popularity that different genres might have (ex. rap might be more popular as a whole than country). We are also only interested in looking at those songs that are currently popular (not performing well historically), since our analysis is focused on current performance on modern streaming services. Therefore, we chose to use all of the songs on the Billboard 2022 Top 100 Country chart to generate a dataset on Spotify.

Data and Methodology

With country artists as our audience, we want to see if duration has an effect on the population of the song in the country genre. Since we were specifically interested in looking at top country songs, and songs that have been more recently popular, we looked at the top 100 country songs of 2022[3]. Billboard calculates the ranking of these songs via an internal metric that utilizes radio airplay audience impressions as measured by Nielsen BDS, album sales data compiled by Nielsen Soundscan (both at retail and digitally), and streaming activity provided by online music sources.

The data in this study comes from Spotify. We fed the Billboard 2022 top 100 country songs into a Spotify API, Spotify Organize Your Music engine[4], which generates a dataset from any playlist with a variety of features. Each row represents a song in the Billboard 2022 Top 100 Hot Country Songs. In addition to these features we added in an additional "Older Popularity" feature, which is an indicator variable that indicates whether the track's artist was featured in the 2021 Billboard top 100 Hot Country Songs. All exploration and model building was done on 99% of the data, as one song was taken out due to its outlier duration of 10 minutes (Taylor Swift's Version - All Too Well). We believe that the extremely high popularity of this track was due to the artist being Taylor Swift and not due to the song length, and that this datapoint would have an outsized effect on our model. Due to having a small sample, splitting the data on a 30% exploration subsample was not appropriate. Since no intervention is taken and all features are observational features, this is an observational dataset.

We operationalized our independent variable as the length of the song, Duration, and our dependent variable as Popularity, where the higher the value the more popular the song is. We decided to use this metric because we wanted to operationalize the concept of a song "performing well" on a streaming service, and this metric, according to the Spotify documentation,

[1] https://www.riaa.com/wp-content/uploads/2018/09/RIAA-Mid-Year-2018-Revenue-Report.pdf
[2] https://www.npr.org/sections/therecord/2018/03/20/594037569/how-the-sound-of-country-music-changed
[3] https://www.billboard.com/charts/year-end/hot-country-songs/
[4] http://organizeyourmusic.playlistmachinery.com/

captures both how many plays a song gets and how recent those plays are. This operational Popularity matches pretty closely to our conceptual popularity. Our conceptual popularity is a measure of performance where number of plays is a metric of performance since a song that is played more often would be considered more successful. In addition, since we are interested in the current trend and how song length would impact popularity of a release now, the "recent-ness" aspect of the popularity metric is useful to us. While some older songs that are out of trend may have a lot of plays from when they were popular, we are interested in the current trend and what would be popular if it was released today.

While duration is the explanatory variable of interest, we also included Energy, Danceability, Valence, Older Popularity, Liveness, Speechiness, and Acousticness in our model as covariates. Based on our exploratory plot (Figure 1) we can see a weak quadratic relationship between popularity and song length. As a result, we chose to model this relationship in a quadratic manner, because it is not possible for songs to be of zero length, nor does the length continuously increase to infinity, instead we are more interested in what an "ideal" song length might be to maximize popularity. Our regression thus took the form:

$$Popularity = \beta_0 + \beta_1(Duration) + \beta_2(Duration^2) + Z\gamma + \epsilon$$

where Z *= row vector of additional covariates,* γ *= column vector of coefficients*

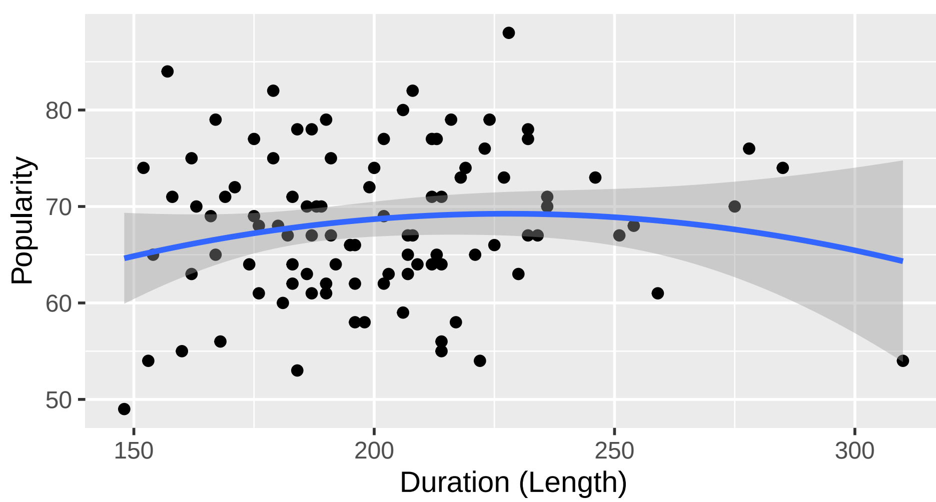


Figure 1: Popularity and Duration (Seconds) with Second Degree Polynomial Line of Bes
Each Point Represents a Song on the Top 100 Country Charts

Results

Table 1, the table of the estimated regressions shows the result of three separate regression models. The first column represents a model that regresses the regressor, Song Length on the regressand, Song Popularity, and uses a quadratic term, Song Length squared, to potentially to better fit the model. The second column represents a model that has the same goal as the previous, but also accounts for other Spotify generated song attributes such as how upbeat a song is (Valence) and . The third model retains the same goal as the other models but looks closer at song (Accousticness) and whether the song artist appeared in the 2021 Billboard Top Country Chart (OlderPopularity) as they were the only two attributes that seemed to have statistically significant effects.

Between the three models, the β_1 coefficients for the models are as follows in the order previously stated: 0.331, 0.263 and 0.332, and the β_2 coefficient remains -.001 for all models. These coefficients may seem to indicate there is an "ideal" song length, however, all of the p-values for these coefficients are very high, far exceeding the threshold of what would be considered statistically significant. All this suggests that for country music, length may not affect how popular a song is. It is possible that since the lengths of the songs we looked at only ranged from 150 to just over 300 seconds, that within this range of song lengths changing song length has little to no effect on the popularity of a song, but if we looked at a song that was much longer or much shorter, we would have seen a greater effect.

Even when considering the other attributes, as seen in the second column, it seems that all coefficients are small, indicating a small effect. Moreover, only the coefficients for "Acousticness" and "Older Popularity" are statistically significant, indicating that, save for these two features, the other attributes don't seem to have much of an effect on how popular a song is. "Acousticness" measures how acoustic a song is, and "Older Popularity" is a measure of whether the artist was on

Table 1: Estimated Regressions

	Dependent variable:		
	SongPopularity		
	(1)	(2)	(3)
SongLength	0.331	0.334	0.332
	(0.398)	(0.360)	(0.369)
	t = 0.831	t = 0.928	t = 0.900
	p = 0.407	p = 0.354	p = 0.369
I(SongLength^2)	−0.001	−0.001	−0.001
	(0.001)	(0.001)	(0.001)
	t = −0.763	t = −0.800	t = −0.817
	p = 0.446	p = 0.424	p = 0.414
Acousticness		0.138**	0.100**
		(0.057)	(0.040)
		t = 2.440	t = 2.489
		p = 0.015	p = 0.013
OlderPopularity		−4.421**	−3.407*
		(2.208)	(1.983)
		t = −2.002	t = −1.718
		p = 0.046	p = 0.086
BeatsPerMinute		0.040	
		(0.029)	
		t = 1.362	
		p = 0.174	
SongEnergy		0.078	
		(0.092)	
		t = 0.845	
		p = 0.398	
Danceability		0.010	
		(0.106)	
		t = 0.094	
		p = 0.926	
Valence		−0.025	
		(0.062)	
		t = −0.406	
		p = 0.686	
Speechiness		0.263	
		(0.330)	
		t = 0.797	
		p = 0.426	
Liveness		0.133	
		(0.098)	
		t = 1.365	
		p = 0.173	
Constant	31.562	16.839	31.681
	(41.219)	(39.515)	(38.793)
	t = 0.766	t = 0.426	t = 0.817
	p = 0.444	p = 0.671	p = 0.415
Observations	99	99	99
R^2	0.024	0.204	0.139
Adjusted R^2	0.004	0.113	0.102
Residual Std. Error	7.719 (df = 96)	7.284 (df = 88)	7.328 (df = 94)
F Statistic	1.198 (df = 2; 96) (p = 0.307)	2.250** (df = 10; 88) (p = 0.022)	3.790*** (df = 4; 94) (p = 0.007)
Note:			*p<0.1; **p<0.05; ***p<0.01
			HC1 robust standard errors in parentheses

the previous year's Billboard Top 100 Hot Country Songs; these attributes could potentially have a statistically significant effect on the popularity of a song, albeit quite small. Ultimately, since our sample size was rather small, our models are underpowered, and our p values are not extremely significant, so the fact that these features are statistically significant is only suggestive of a possible effect on popularity. So overall, there does not appear to be much practical significance to these results. Instead, they seem to indicate that within a song length range of 2 minutes 30 secs to 5 minutes, song length does not significantly affect the popularity of a song.

Limitations

There were a few statistical limitations to our analysis. For one, we did not have enough data to split our dataset into an exploration and validation set so there is a possibility of us having overfit our data or to not have enough data to establish a pattern.

In addition, due to our small sample size, to have a constant regression, our data must meet the Classic Linear Model Assumptions. There are some potential violations of the Independent Identical Distribution assumption that could have occurred. For example, there could be a sampling issue with album-level clustering. In other words a popular song may bolster the popularity of other songs on that same album. Another violation may be found in Spotify's recommendation algorithm and how a user may interact with it: For example if a user was to listen to a country song, they may be recommended other similar country songs, boosting the popularity of those songs as well. Our data does have no perfect collinearity, since, if there was any collinearity, one of the collinear variables would be automatically dropped by R when we created our models, but none of our variables were dropped in our regression.

When testing for the third assumption we found that our model is heteroskedastic based on the scale-location plot. We plotted the standardized residuals on the fitted values and we found that our loess line was not horizontal. As a result, there does not seem to be a constant error variance across the fitted values. To correct for this we used robust standard errors in our model. To check for linear conditional expectation we look at the residuals vs fitted plot, and our model, represented by the red line, follows a roughly horizontal line that lies along the zero residual value. Therefore, the linear conditional expectation assumption is met. Finally, looking at the QQ plot of our data, the plot roughly follows the straight diagonal line with minimal deviations. This means that our standardized residuals roughly match up with the theorized quantiles (which come from the normal distribution), indicating a normal distribution of errors.

In terms of structural effects, there are a few possible omitted variables that could be biasing our tests. One omitted variable may be whether or not a song was released as a single. Since songs that are released as singles are usually shorter than songs released on a full album, we expect there to be a negative correlation between whether or not a song is released as a single and song length. Since singles are usually released to generate buzz and exposure for an artist, we expect there to be a positive correlation between whether or not a song is a single and its popularity. As a result, we predict that there will be a negative omitted variable bias on our key variables. Therefore, the main effect is being driven towards zero, which suggests that our hypothesis tests are underconfident. We expect that a similar effect might hold for the indicator variable of whether or not a song is explicit or contains controversial verses.

Conclusion

This study estimated the popularity of a song of the Country genre, where the higher the value the more popular the song is, based on song length. Our model found that song length, our main variable of interest, did not have a statistically significant effect on the song popularity. However, we did find two covariates, "Acousticness" (the higher the value the more acoustic the song is) and "Older Popularity" (whether the artist was on the Billboard 2021 Top 100 Hot Country Songs) had a slightly statistically significant effect. Our models predicted that a one unit increase in the acousticness predicts between a 0.1 and 0.138 unit increase in popularity. If the artist was on the previous year's top 100 country list, there is between a 4.281 and 3.407 unit predicted decrease in popularity. Since our dataset was very small, 100 data points, we cannot be certain of this relationship. Instead these results are suggestive of a possible relationship that could be explored in a future analysis with new data.

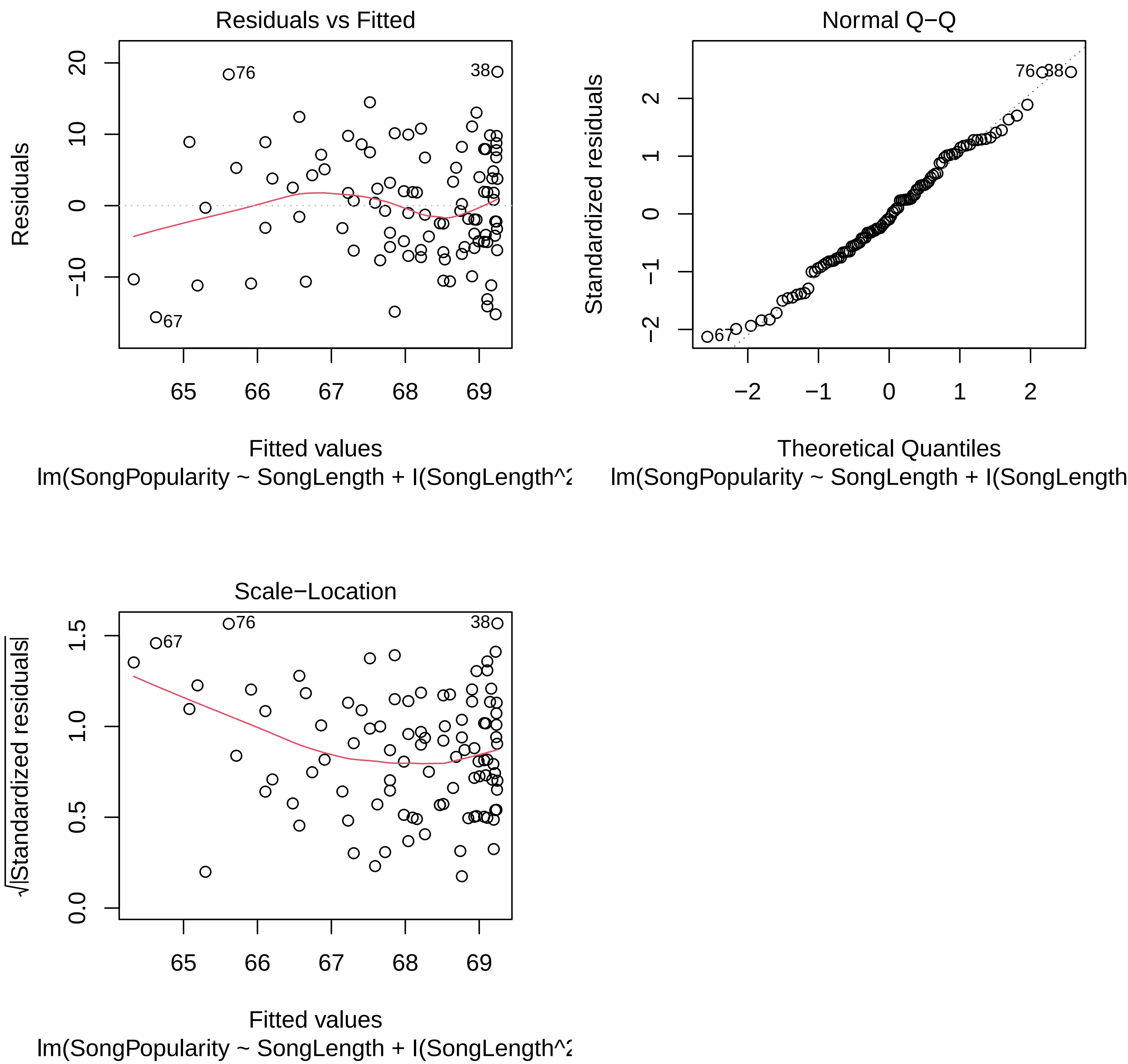
This information may be useful for Country artists generating new music as they now know that song length, within the typical song range length of 2.5 to 5 minutes, will probably not affect their song popularity.

Appendix

For reference, here are the plots we used to evaluate the CLM assumptions.

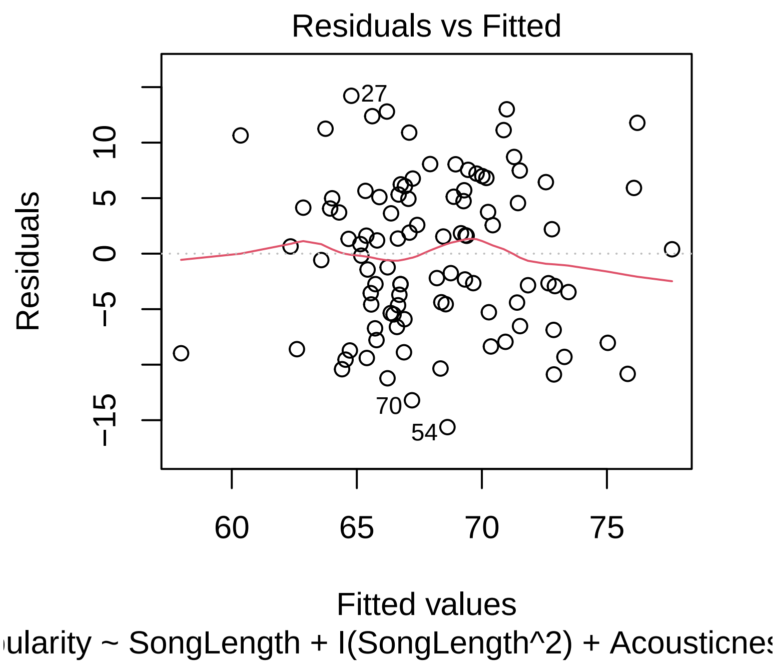
Model 1 Plots:

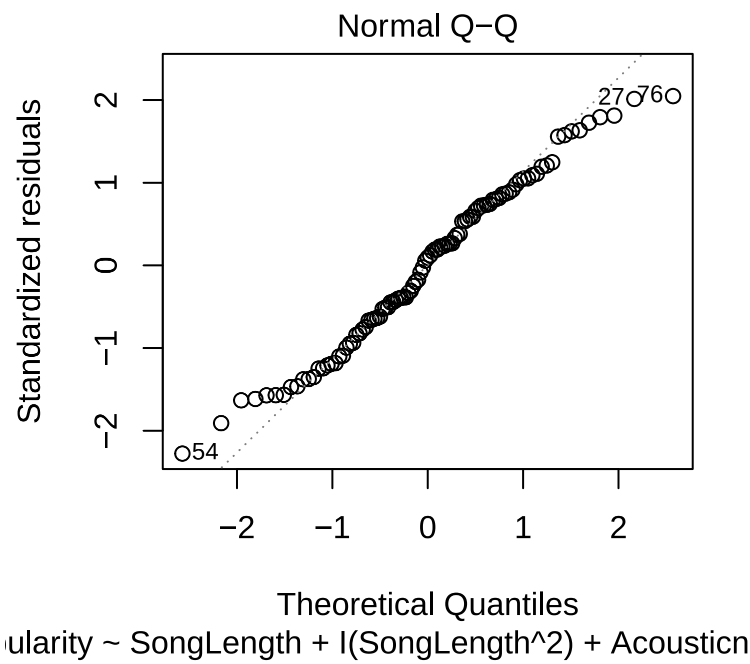
$$Popularity = \beta_0 + \beta_1(Duration) + \beta_2(Duration^2) + \epsilon$$

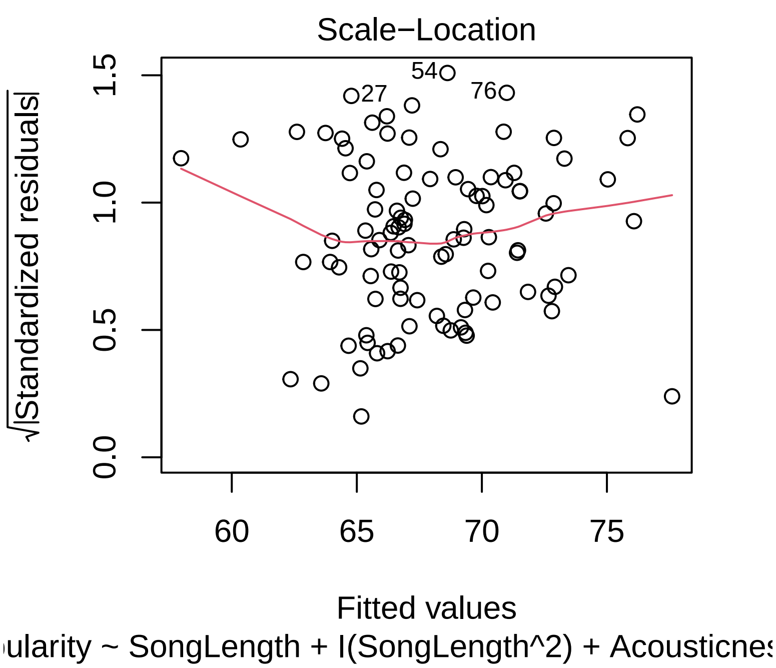


Model 2 Plots:

$$Popularity = \beta_0 + \beta_1(Duration) + \beta_2(Duration^2) + \beta_3(Acousticness) + \beta_4(OlderPopularity) +$$
$$\beta_5(BeatsPerMinute) + \beta_6(SongEnergy) + \beta_7(Danceability) + \beta_8(Valence) + \beta_9(Speechiness) + \beta_10(Liveness) + \epsilon$$







Model 3 Plots:

$$Popularity = \beta_0 + \beta_1(Duration) + \beta_2(Duration^2) + \beta_3(Acousticness) + \beta_4(OlderPopularity) + \epsilon$$