Characteristics of Top Songs Has Changed from Pandemic Brain*

An analysis of songs on Billboard's Year-End Hot 100 singles list (2014 to 2023)

Emily Su

April 18, 2024

Music often reflects the current climate of society and there is a growing interest in how hit songs as seen on Billboard's Year-End Hot 100 singles has changed after the COVID-19 pandemic. This paper looks at music characteristics such as tempo, song duration, loudness, and modality of songs from Billboard's Year-End Hot 100 singles list from 2014 to 2023 to reveal patterns and relationships to explain the difference between top songs before 2020 and 2020 onwards. The results show that hit songs from 2020 onwards had become on average quieter but slightly faster with our model revealing that duration does not impact likelihood of a song being a hit prior to 2020 but songs in a minor key were more likely to be a hit before 2020. These results can support the evaluation of the emotional state of different populations and improve treatments such as music therapy, however further investigation is needed on the influence of lyrics on different music characteristics.

Table of contents

---

*Code and data are available at: https://github.com/moonsdust/top-songs.

1   Introduction

During the COVID-19 pandemic, viral songs on social media sites such as TikTok were met with thousands of listens on streaming platforms such as Spotify and at times go onto being on lists like the Billboard Year-End Hot 100. These songs often accompany thousands of short clips of people talking about their lives, challenges, etc. Ghaffari et al in their study, "The impact of COVID-19 on online music listening behaviors in light of listeners' social interactions" noted that other studies have found that people used music as a way to cope and regulate their thoughts and emotions during the pandemic and the lockdowns that came with it (Ghaffari et al. 2023). We can see this translated through the short-form videos where users don't talk and instead have the song play out to convey how they are feeling or thinking. Ghaffari et al's findings are similar to how music therapy works. As defined by the Canadian Association of Music Therapists, music therapy is the use of music to support an individual's health, social development, and well-being and some of the techniques used include listening to music (Canadian Association of Music Therapists 2020). Hurwitz and Krumhansl conducted a study into how people's listening habits throughout the pandemic and they found that the songs they listened to often were linked to an emotion such as sadness and/or invoking memories from the past (Hurwitz and Krumhansl 2021). However, this raises the following question, which we explored in our analysis: how are top songs prior to 2020 or the start of the pandemic different from the top songs during and after 2020? In our analysis, we defined hit songs before 2020 to be hit songs before the pandemic and hit songs during 2020 and afterwards to be hit songs during and after the pandemic.

In this paper, to investigate patterns and trends in music, we analyzed data from the Billboard Year-End Hot 100 singles list from 2014 to 2023 on music characteristics such as its modality (major or minor key), track duration, loudness, and tempo. Our estimand is the song characteristics from the Billboard Year-End Hot 100 singles such as its modality (major or minor), track duration, loudness, and tempo if a song was created before 2020 or during and after 2020. The list allowed us to gauge how preference in music characteristics has changed overall during and after the pandemic. Current studies give us a sense of how currently people are feeling based on surveys done. However, there is currently a lack of understanding and focus on the difference in musical characteristics of popular songs with the general population before the pandemic and during and after it. In our findings, our data showed that songs in a major key made up the majority of top songs before, during, and after 2020 and top songs were on average quieter, shorter, and had slightly faster tempos during and after 2020. In our model, it revealed that top songs prior to 2020 are more likely to be louder, slower, and if its in a minor key, its more likely to be a hit song before 2020. However, duration did not impact the likelihood in our model. Music can reflect the climate of society and understanding trends in music characteristics of hit songs can help provide insight on the overall feelings of different populations during and after the pandemic as well as before it. This can aid in improving treatments such as music therapy.

In the rest of this paper, the data section (Section 2) will cover the dataset used, how it was obtained, define the variables of interest that are used by our tables and graphs, and briefly explain the data cleaning process. The model section (Section 3) will explain our proposed causal model, which would explain potential relationships in our data, the setup of our model to understand these relationships, and justification for our model. In the results section (Section 4), we will reveal tables and graphs made on our datasets, explain what they show, and show our results from our model. In the discussion section (Section 5), we will connect back to the real world and explain what the results could mean, the implications of our results, potential areas of improvement for the paper, and suggestions for future works. Finally, the appendix section (Section A) will extra tables and graphs from our results as well, as additional information about the model, and a link to a Shiny application featuring interactive graphs of some of the results.

2 Data

2.1 Methodology

The dataset used in this paper was retrieved, simulated, cleaned, analyzed, and tested using the R programming language (R Core Team 2023), tidyverse (Wickham et al. 2019), knitr (Xie 2014), janitor (Firke 2023), dplyr (Wickham et al. 2023), ggplot2 (Wickham 2016), spotifyr (Thompson et al. 2022), usethis (Wickham et al. 2024), arrow (Richardson et al. 2024), ggcorrplot (Kassambara 2023), and testthat (Wickham 2011). Code written in the scripts was checked and styled with lintr (Hester et al. 2024) and styler (Müller and Walthert 2024).

Accompanying the paper, a Shiny app (Chang et al. 2024) using shinyWidgets (Perrier, Meyer, and Granjon 2024) was created containing interactive visualizations. The packages that were used for the model-related sections or used for the model itself are DiagrammeR (Iannone and Roy 2024), rsvg (Ooms 2023), magrittr (Bache and Wickham 2022), DiagrammeRsvg (Iannone 2016), png (Urbanek 2022), rstanarm (Goodrich et al. 2024), and modelsummary (Arel-Bundock 2022).

2.2 Data Source and Measurements

The data used for the model and analysis comes from the `playlists_analysis_data` dataset. This dataset was constructed by using an R wrapper for Spotify's API called spotifyr (Thompson et al. 2022) to scrape playlists complied by Spotify users antoniomendoza and wickeddreamer96 of songs from the Billboard Year-End Hot 100 singles from 2014 to 2023 (Billboard 2024). User antoniomendoza created playlists for the years 2018 to 2023, inclusive, while user wickeddreamer96 created playlists for the years 2014 to 2017, inclusive. The latest version of the data used was scraped on Tuesday March 26, 2024 at 9:46pm EDT. The Billboard Year-End Hot 100 singles list for the year 2014 onward was created based on data from Nielsen SoundScan (Manitoba Music 2024). Nielsen SoundScan collect data weekly from 39,000 retail outlets, venues, etc. in 19 countries such as Canada and the United States on in-store sales and digital sales (Manitoba Music 2024). They also track how often songs are streamed online and played on the radio (Manitoba Music 2024). The list comprise of songs they collected data on starting from December of the previous year to the end of November of the current year, which is when Billboard releases the Year-End Hot 100 list (Denis et al. 2023).

Initially, we had scraped playlists created by Spotify that they have compiled of the top songs on the app. However, we had an insufficient amount of observations from these playlists for our dataset. The playlists we previously were using only had around 50 songs each and there were years where Spotify had not created a top song playlist for such as 2021 and 2022.

Each observation in the dataset belongs to a song from the Spotify playlist that we scraped compiled by antoniomendoza and wickeddreamer96 for songs on the Billboard Year-End Hot 100 singles. For the duplicate songs that showed up in different years, they were kept as we wanted to analyze song characteristics. In total, there are 1000 observations in the `playlists_analysis_data` dataset, which equals to the number of songs on the Billboard Year-End Hot 100 singles lists from 2014 to 2023.

Table 1: Preview of dataset on Billboard Year-End Hot 100 singles from 2014 to 2023 provided by Spotify

hit_year	track_duration_ms	tempo	loudness	mode_name	key_mode	period
2014	232720	160.019	-4.662	minor	F minor	Before Pandemic

Table 1: Preview of dataset on Billboard Year-End Hot 100 singles from 2014 to 2023 provided by Spotify

hit_year	track_duration_ms	tempo	loudness	mode_name	key_mode	period
2014	215672	131.931	-6.122	major	F# major	Before Pandemic
2014	269560	119.930	-7.064	major	G# major	Before Pandemic
2014	199938	94.981	-4.141	minor	A# minor	Before Pandemic
2014	257840	122.017	-4.972	minor	C# minor	Before Pandemic
2014	177685	100.315	-7.321	major	F# major	Before Pandemic

Not seen in Table 1, before_pandemic, major, and minor are numerical versions of the mode_name and period columns. The variables that were originally the raw dataset that was scraped are track_duration_ms, tempo, loudness, mode_name, and key_mode. track_duration_ms describes the duration of the song in milliseconds (ms), tempo is the mean (average) beats per minutes (BPM) of the song, and loudness is mean loudness of a top song in decibels (dB). mode_name represents the key the melody of the song is in, which is either "major" or "minor" and key_mode signifies the scale the song is in.

The variables constructed from the raw dataset are before_pandemic, period, hit_year, major, and minor. How these variables were constructed was that hit_year was obtained from check the name of the playlist and it represents the hit. From hit_year, the period column was constructed, which indicates whatever or not the hit year of the song was before 2020, which was when the COVID-19 pandemic started, or during 2020 onwards. Reading from the period column, the before_pandemic column was created, where 1 represents that the song was a hit song before the pandemic and 0 otherwise. The major and minor columns represents whatever or not the melody of the song is in a major or minor key and the value would be 1 if it's that key and 0 otherwise. These columns were constructed from checking mode_name when it would indicate either "major" or "minor".

2.3 Variables of Interest

As seen in Table 1, there are multiple columns however our main variables of interests are before_pandemic, track_duration_ms, loudness, tempo, and mode_name for our model. The major and minor columns were used initially to visualize potential relationships between variables as seen in Figure 13. period, hit_year, and key_mode were used to visualize our data prior to training our model and to understand the data.

3 Model

From our analysis of the data, we observed that there were trends between a song's duration, loudness, tempo, and modality for top songs prior 2020 and 2020 onwards. We are interested in investigating if a song is likely to be a top song prior to 2020 or not based on what we know about the duration, loudness, tempo, and modality of a top song. We will do this by creating a model.

3.1 Model set-up

In order to predict a top song was a hit prior to 2020, we will make the assumption that there is a relationship between the outcome variable, if the song is a top song prior to 2020 and our variables of interests: duration, loudness, tempo, and modality. Our variables of interests will be the predictor variables for the model. We will define our model as a logistic regression model with a Bayesian approach as follows:

$$y_i|\pi_i \sim \text{Bern}(\pi_i)$$
$$\text{logit}(\pi_i) = \beta_0 + \beta_1 \times \text{duration}_i + \beta_2 \times \text{loudness}_i + \beta_3 \times \text{tempo}_i + \beta_4 \times \text{modality}_i$$
$$\beta_0 \sim \text{Normal}(0, 2.5)$$
$$\beta_1 \sim \text{Normal}(0, 2.5)$$
$$\beta_2 \sim \text{Normal}(0, 2.5)$$
$$\beta_3 \sim \text{Normal}(0, 2.5)$$
$$\beta_4 \sim \text{Normal}(0, 2.5)$$

We define y_i to be a top song i, which if it is 1 represents a top song prior to 2020 and 0 if it is a top song from 2020 onwards. π_i is the probability of a top song being a hit prior to 2020. Following this, $\text{logit}(\pi_i)$ is the log odds of a top song i being a top song prior to 2020. Let duration_i be the duration of the top song in milliseconds (ms) and loudness_i to be the mean loudness of a top song in decibels (dB). Set tempo_i to be the mean beats per minutes (BPM) of a top song and modality_i to be the modality of the top song, where 1 means the melody of the song is in a minor key and 0 if it is in a major key. β_0 is the intercept that represents the log odds when all predictor variables are 0. β_1 is the coefficient that signifies the log odds of a song being a hit song prior to 2020 when the duration_i changes by 1 ms. β_2 is the coefficient that stands for the log odds of a song being a hit song prior to 2020 when loudness_i changes by 1 dB. β_3 and β_4 both represent the log odds of a song being a hit song prior to 2020 when tempo_i and modality_i changes, respectively.

We ran the model using the rstanarm package (Goodrich et al. 2024) and R (R Core Team 2023) with the default priors from the rstanarm package. Diagnostics related to the model such as a trace plot, Rhat plot, and posterior predictive check can be found at Section A.
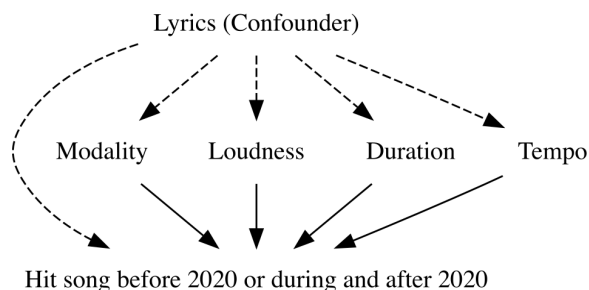
3.1.1 Model justification

Figure 1: Causal relationship between song characteristics and hit song (before 2020 or during and after 2020)

A logistic regression model was chosen because our outcome variable, if the song is a top song prior to 2020, is a binary outcome. However, alternative models were considered. Multiple linear regression model was a candidate for our model since the model has multiple predictor variables but it assumes that the outcome is continuous instead of discrete. Other regression models like negative binomial regression were not chosen since we are not dealing with discrete counts for the outcome but instead discrete binary values of 1 and 0. Our logistic regression model describes the following causal relationship (Figure 1) where song characteristics predicts whatever a hit song was made prior to 2020, excluding a confounding variable, lyrics. The assumption we make is that there is a relationship between song characteristics and a top song being a hit prior to 2020. In Figure 1, the reason lyrics, if there are any, is a confounding variable is because lyrics can affect not only if the song was a hit prior to 2020 but also song characteristics from the way the lyrics are performed. However, we are unable to observe the effects of lyrics directly with our data. Weakness and further improvements to the model are discussed in Section 5.

4 Results

In this section, we visualized our data through graphs and tables as well as present the results from our model.

4.1 Difference in song characteristics before 2020 (the pandemic) and during and after 2020 (the pandemic) of songs from the Billboard Year-End Hot 100 singles

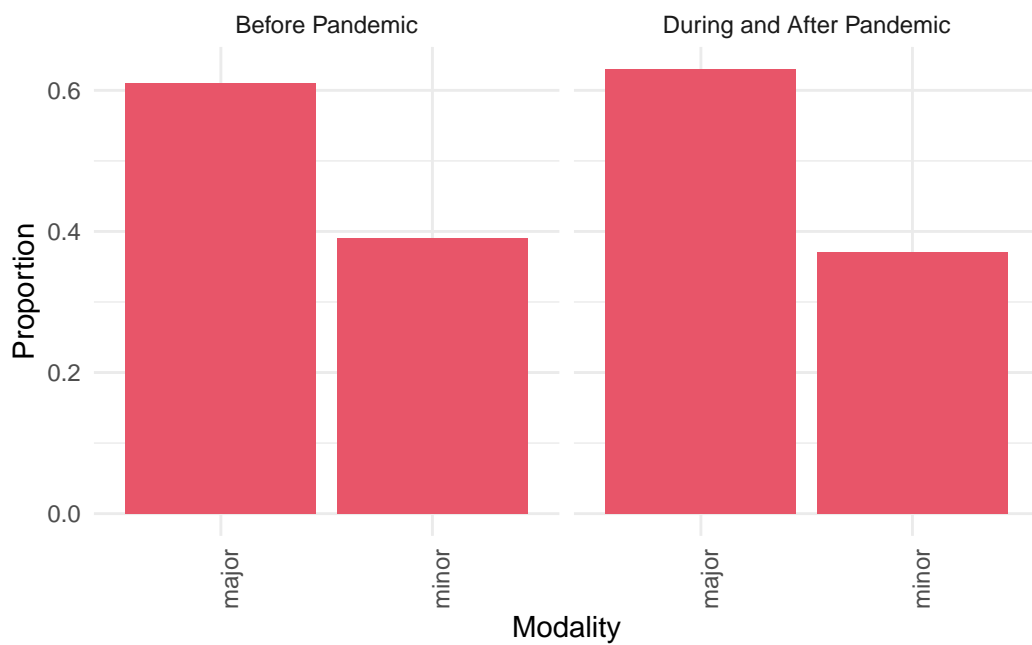### 4.1.1 Scale and Modality

Figure 2: Proportion of songs whose modality is in a major or minor key before 2020 (the pandemic) versus during and after 2020 (the pandemic) from Billboard's Year-End Hot 100 singles

Table 2: Proportion of songs whose modality is in a major or minor key before 2020 (the pandemic) versus during and after 2020 (the pandemic) from Billboard's Year-End Hot 100 singles

Period	Modality	Count of each mode	Proportion of each mode
Before Pandemic	major	368	0.61

Table 2: Proportion of songs whose modality is in a major or minor key before 2020 (the pandemic) versus during and after 2020 (the pandemic) from Billboard's Year-End Hot 100 singles

Period	Modality	Count of each mode	Proportion of each mode
Before Pandemic	minor	232	0.39
During and After Pandemic	major	251	0.63
During and After Pandemic	minor	149	0.37

Figure 2 and Table 2 reveals that a proportion of hit songs before and during and after 2020 that had melodies in a major key was around ~60% (61% before 2020 and 63% from 2020 onwards). Songs in a minor key made up about 40% or less of songs (39% before 2020 and 37% from 2020 onwards) from 2014 to 2023 on Billboard's Year-End Hot 100 singles list. There was a slight increase in songs in a major key from 2020 afterwards however songs in a major key making up a majority of Billboard's Year-End Hot 100 singles stayed consist before 2020 and 2020 onwards.

When looking into the specific scales of the songs, Figure 3 and Table 10 shows that there was an increase in songs in C major during and after 2020 with C major and C# major made up a majority of songs from Billboard's list from 2020 onwards at approximately 9% each. Prior to 2020, the proportion of songs in C major were at 11% however there was a decline in songs in C major after the start of the pandemic. There was a decline in songs in G# major from 8% of the list to 6% of the list as well as a an increase in songs in D minor and D# minor from less than 1% of the list to 2% each. There were also a decline in songs in F minor where they would make up 6% of Billboard's list prior to 2020 to 4% during and after 2020.
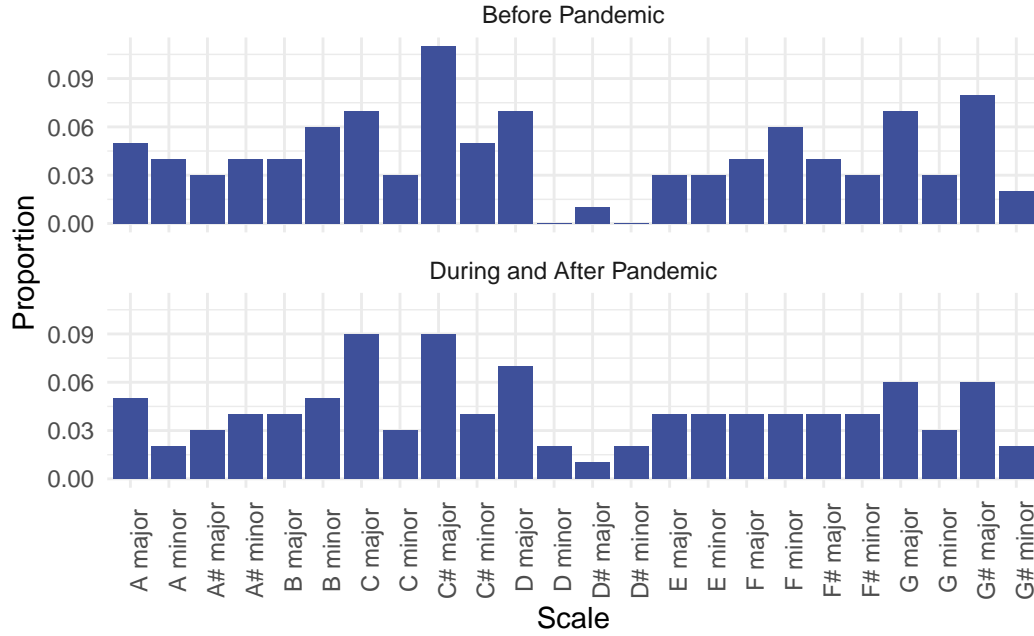
Figure 3: Proportion of songs in different scales before 2020 (the pandemic) versus during and after 2020 (the pandemic) from Billboard's Year-End Hot 100 singles

4.1.2 Track Duration

Table 3: Minimum, quartiles, median, and maximum of track duration (in ms) before 2020 (the pandemic) on Billboard's Year-End Hot 100 list

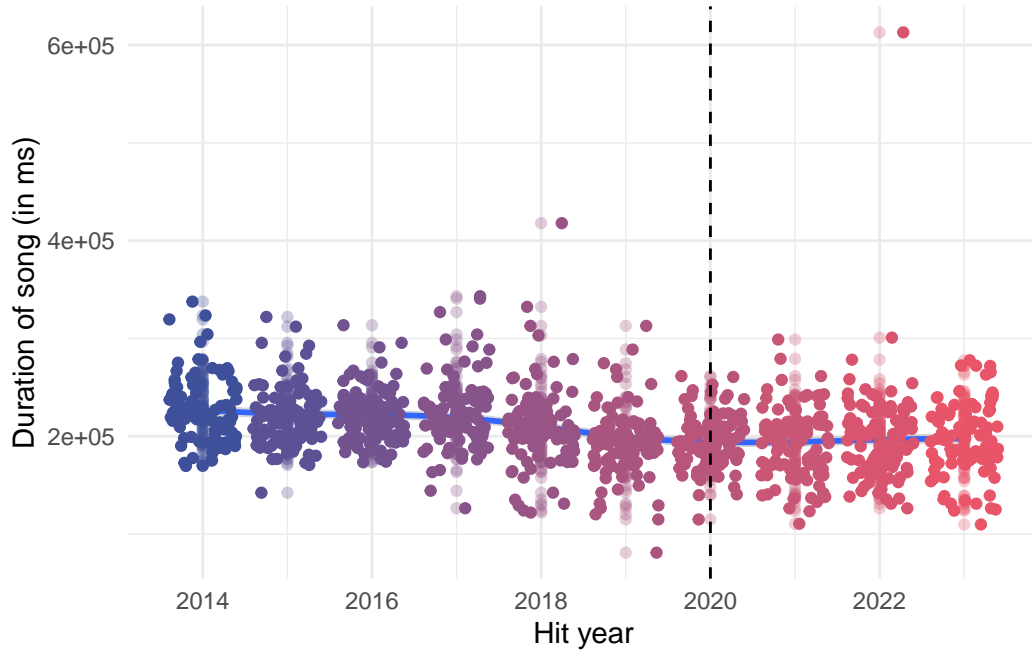| Duration of song (in ms) |
| --- |
| Min. : 80927 |
| 1st Qu.:194600 |
| Median :214070 |
| Mean :216286 |
| 3rd Qu.:233087 |
| Max. :417920 |

Figure 4: Relationship between the hit year and track duration of a song (in ms) on Billboard's Year-End Hot 100 singles list. Songs from the same year are grouped in the same colour and the horizontal, dashed line divides the hit tracks before 2020 and hit tracks from 2020 and beyond. Blue, smoother line describes the overall trend and relationship between hit year and track duration.
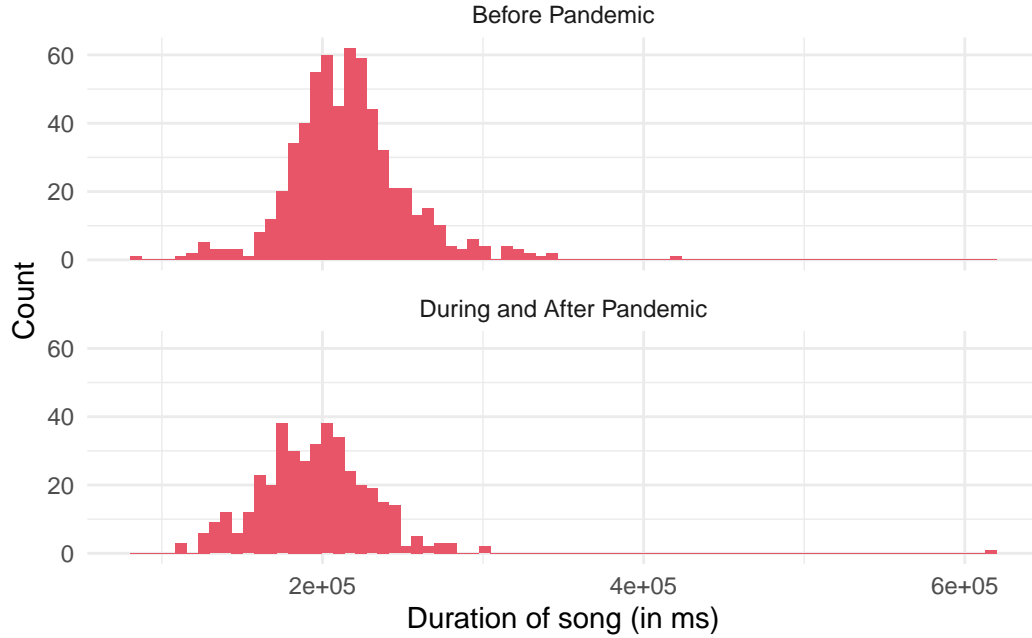
Figure 5: Distribution of track duration (in ms) before 2020 (the pandemic) versus during and after 2020 (the pandemic) from songs on Billboard's Year-End Hot 100 singles list.

Table 4: Minimum, quartiles, median, and maximum of track duration (in ms) during and after 2020 (the pandemic) on Billboard's Year-End Hot 100 list

Duration of song (in ms)
Min. :109750
1st Qu.:173369
Median :195120
Mean :195999
3rd Qu.:215336
Max. :613026

Figure 4 reveals that songs over the years from 2014 to 2020 decreased slightly in duration and from 2020 onwards, there was almost no change in trend overall based on the blue, smoother line. Figure 5 shows that the distribution for top songs before the pandemic and during and after the pandemic are skewed to the right. According to Table 4 and Table 3, the distribution of top songs during and after the pandemic are more disperse around their mean of 195999 ms than those from before the pandemic where there is less variance, which means how close data points are to their mean/average value. Based on the distribution, there are more songs with duration of about 216286 ms for top songs before the pandemic. Table 4 and Table 3

also show that hit songs during and after the pandemic are shorter in length on average with a mean track duration of 195999 ms compared to hit songs from before the the pandemic with a mean track duration of 216286 ms. The median (the middle value of the data when sorted) track duration also decreased from 214070 ms before 2020 to 195120 ms for 2020 onwards.
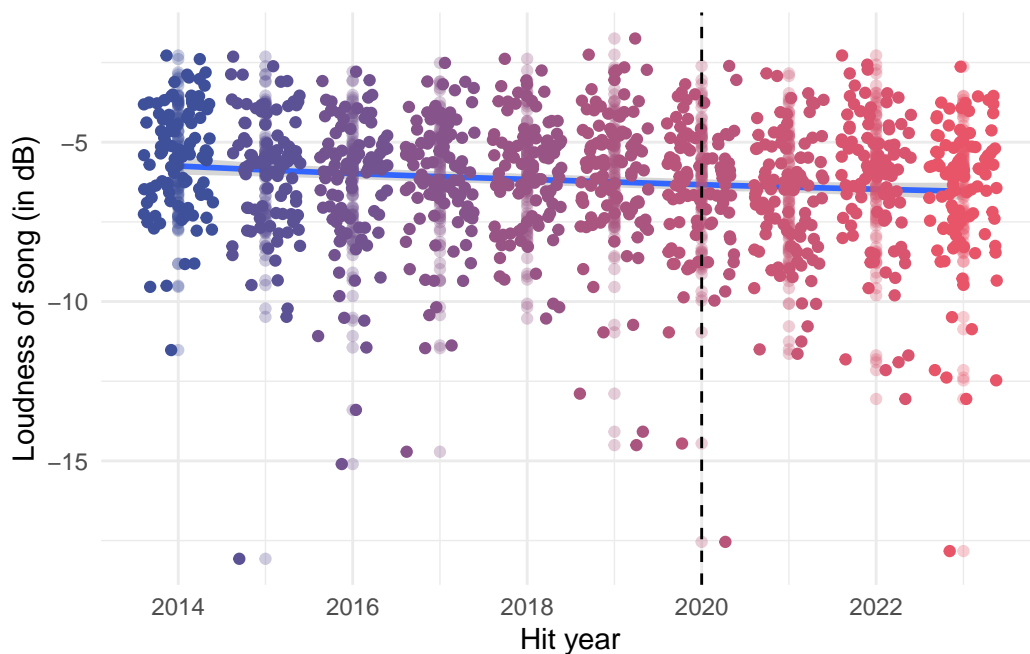
4.1.3 Loudness

Figure 6: Relationship between the hit year and loudness of a song (in dB) on Billboard's Year-End Hot 100 singles list. Songs from the same year are grouped in the same colour and the horizontal, dashed line divides the hit tracks before 2020 and hit tracks from 2020 and beyond. Blue, smoother line describes the overall trend and relationship between hit year and loudness.

Table 5: Minimum, quartiles, median, and maximum of loudness of song (in dB) before 2020 (the pandemic) on Billboard's Year-End Hot 100 list

Loudness of song (in dB)
Min. :-18.071
1st Qu.: -7.051
Median : -5.700
Mean : -5.964
3rd Qu.: -4.705

13

Table 5: Minimum, quartiles, median, and maximum of loudness of song (in dB) before 2020 (the pandemic) on Billboard's Year-End Hot 100 list

Loudness of song (in dB)
Max. : -1.746

Table 6: Minimum, quartiles, median, and maximum of loudness of song (in dB) during and after 2020 (the pandemic) on Billboard's Year-End Hot 100 list

Loudness of song (in dB)
Min. :-17.829
1st Qu.: -7.604
Median : -6.151
Mean : -6.516
3rd Qu.: -5.109
Max. : -2.278

Figure 6 shows that from 2014 to 2023, there was a decrease in loudness overall with hit songs as see with the negative slope of the blue, smoother line. Figure 7 reveals that the distribution of top songs' loudness is left-skewed with the loudness of top songs from during and after the pandemic being more disperse than those before the pandemic around their mean. This suggests that top songs from 2020 onwards varied more in terms of their loudness when compared with top songs before 2020. Table 5 and Table 6 shows that before the pandemic, top songs on Billboard's list had a mean loudness of -5.964 dB while during and after the pandemic, it had a mean loudness of -6.516 dB. The median loudness of top tracks has also decrease from -5.700 dB prior to the pandemic to -6.151 dB during and after the pandemic. This indicates that hit songs during and after the pandemic are less louder than hit songs before the pandemic.
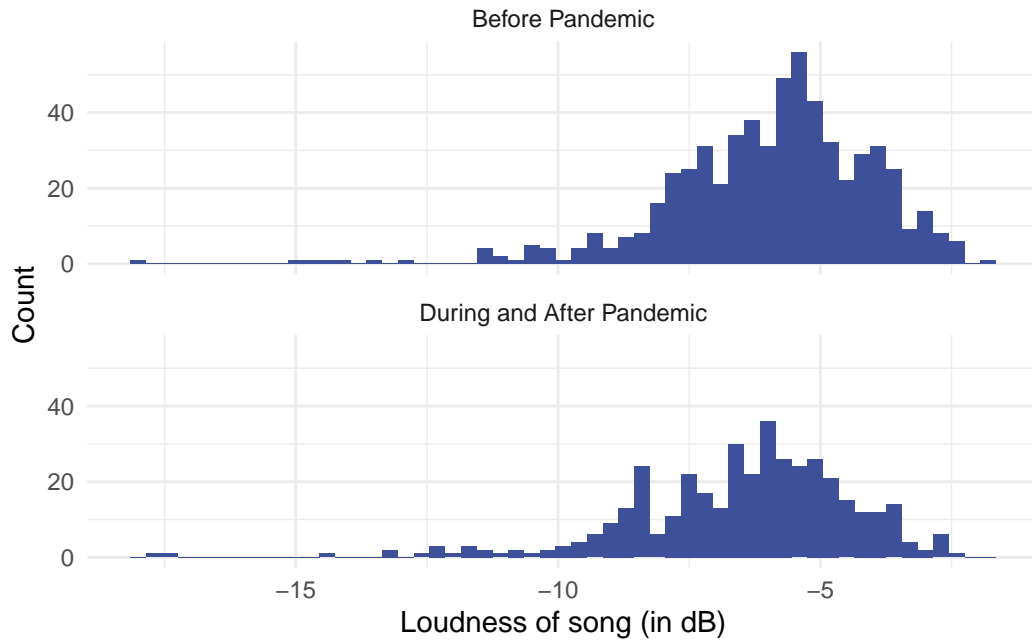
14

Figure 7: Distribution of loudness (in dB) of hit songs before 2020 (the pandemic) versus during and after 2020 (the pandemic) from Billboard's Year-End Hot 100 list

4.1.4 Tempo

Table 7: Minimum, quartiles, median, and maximum of tempo of a song (BPM) before 2020 (the pandemic) from Billboard's Year-End Hot 100 list

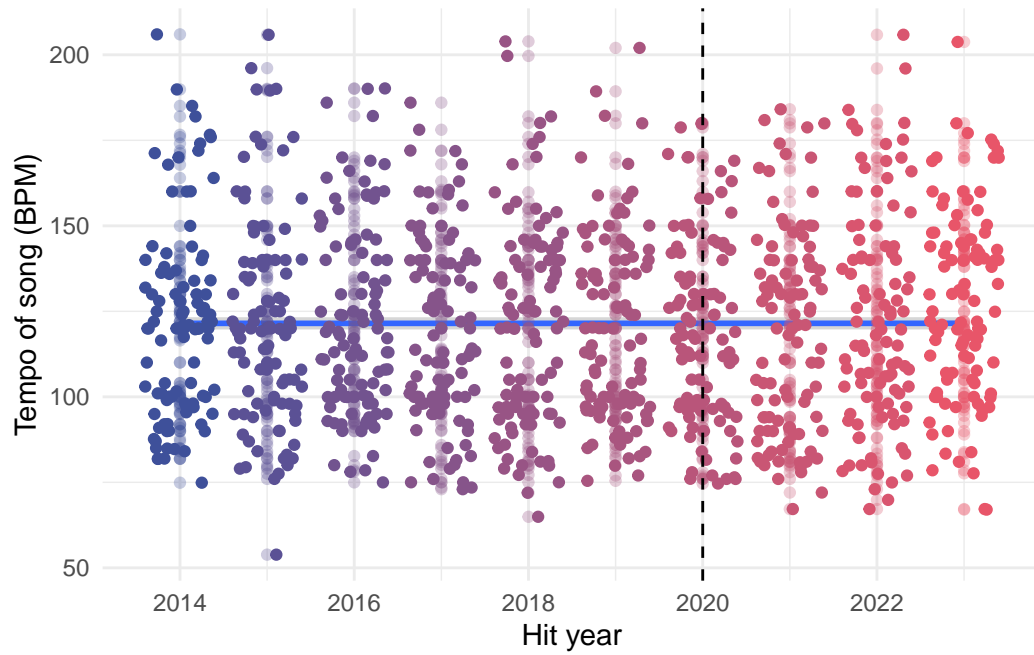| Tempo of song (BPM) |
| --- |
| Min. : 53.86 |
| 1st Qu.: 98.03 |
| Median :119.98 |
| Mean :121.11 |
| 3rd Qu.:140.00 |
| Max. :205.97 |

Figure 8: Relationship between the hit year and tempo of a song (the beats per minutes (BPM) of a song) on Billboard's Year-End Hot 100 singles list. Songs from the same year are grouped in the same colour and the horizontal, dashed line divides the hit tracks before 2020 and hit tracks from 2020 and beyond. Blue, smoother line describes the overall trend and relationship between hit year and tempo.
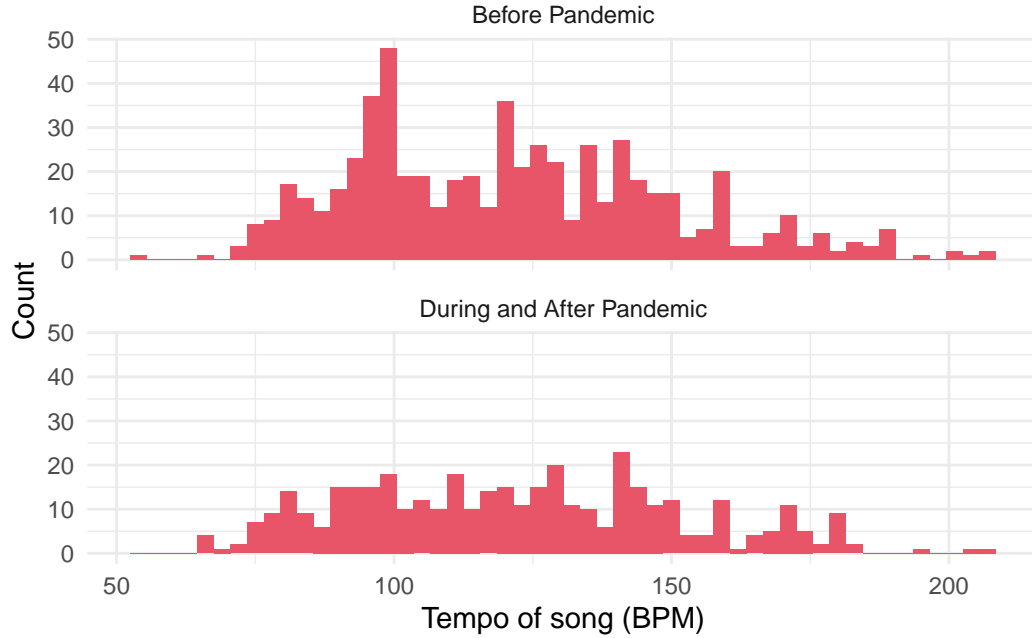
Figure 9: Distribution of tempo (BPM) before 2020 (the pandemic) versus during and after 2020 (the pandemic) from Billboard's Year-End Hot 100 list

Table 8: Minimum, quartiles, median, and maximum of tempo of a song (BPM) during and after 2020 (the pandemic) from Billboard's Year-End Hot 100 list

Tempo of song (BPM)
Min. : 67.03
1st Qu.: 98.02
Median :120.03
Mean :122.06
3rd Qu.:142.31
Max. :205.86

Figure 8 reveals almost no change in tempo overall across top songs from 2014 to 2023 based on the blue smoother line appearing to be flat. However, looking at Table 7 and Table 8 it shows that both the median and mean tempo of top songs before 2020 (about 120 BPM and 121 BPM, respectively) are lower or about equal than the median and mean tempo of top songs during and after 2020, which is about 120 BPM and 122 BPM, respectively. Figure 9 shows that in both periods of time, their distributions have skewness close to zero however there is less dispersion around the mean for hit songs prior to 2020. These results suggest that the the tempo of top songs on Billboard's list are slightly faster during and after 2020.

17

4.2 Model Results

Table 9 and Figure 10 shows the results of our logistic regression model. Table 9 presents the coefficient estimates obtained for our model with their mean absolute deviation (MAD) while Figure 10 showcases the coefficient estimates (the dot) alongside their 90% credible interval (the line). Coefficient estimates represents the change in the log odds of a song being a hit song prior to 2020 when one predictor variable changes by 1 unit as the other predictors stays constant (Minitab 2024). With the given data, the 90% credible interval is an interval where the true value of our coefficient estimate has a 90% probability of lying within (Hespanhol et al. 2019). Mean absolute deviation is the average distance of a data point representing a hit song and the mean.

Table 9 indicates that the intercept β_0 is -2.002. This means that songs in a major key with tempo, loudness, and mode being 0 is less likely to be a top song before 2020. Table 9 and Figure 10 reveal that the coefficient estimate of track duration β_1 is 0.000 when tempo, mode, and loudness remains constant and its 90% credible interval includes the value 0. This suggest that when the duration of the song changes the likelihood of a song being a top song prior to 2020 does not change or is impacted. When loudness varies and track duration, tempo, and mode stays fixed, Table 9 indicates that its coefficient estimate for β_2 is 0.148 with a 90% credible interval above 0 as seen in Figure 10. As the coefficient estimate is positive, it means that hit songs that are louder are more likely to be a hit song prior to 2020. Since the coefficient estimate for tempo β_3 is -0.002 and its 90% credible interval being slightly below 0 when only tempo varies, it means songs with slower tempos are more likely to be a hit song prior to 2020. The coefficient estimate for the modality (mode) β_4 is 0.095 and its 90% credible interval being wider suggests that top songs that are in a minor key are likely to be a hit before 2020. However, the true value for the coefficient for mode can also be negative according to Figure 10 since the interval passes through 0.

Table 9: Relationships of a hit song being a hit before 2020 / the pandemic with its tempo, song duration, loudness, and modality (mode) with their mean absolute deviation (MAD) values in parenthesis.

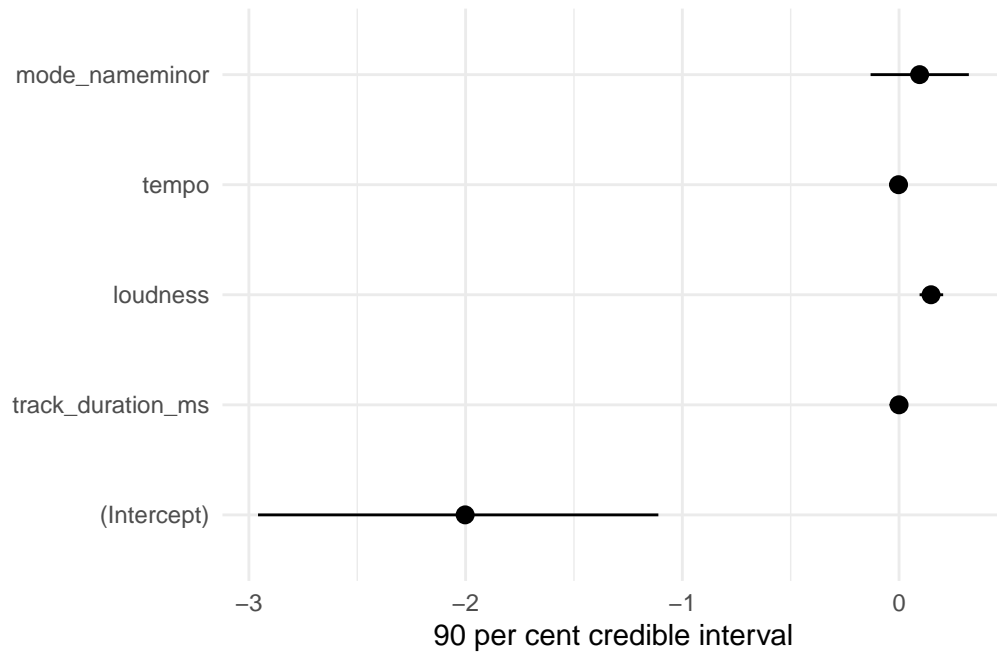|	A hit song before 2020
(Intercept)	−2.002
	(0.550)
track_duration_ms	0.000
	(0.000)
loudness	0.148
	(0.033)
tempo	−0.002
	(0.002)
mode_nameminor	0.095
	(0.139)
Num.Obs.	1000
R2	0.096
Log.Lik.	−624.867
ELPD	−630.3
ELPD s.e.	12.9
LOOIC	1260.6
LOOIC s.e.	25.7
WAIC	1260.6
RMSE	0.46

Figure 10: The credible intervals (line) for coefficient estimates (dot) of predictors of a hit song before 2020

5 Discussion

[To Do]

Weakness: There were more songs available prior to 2020 in our dataset. This could have impacted the model and the coefficient estimates seen.
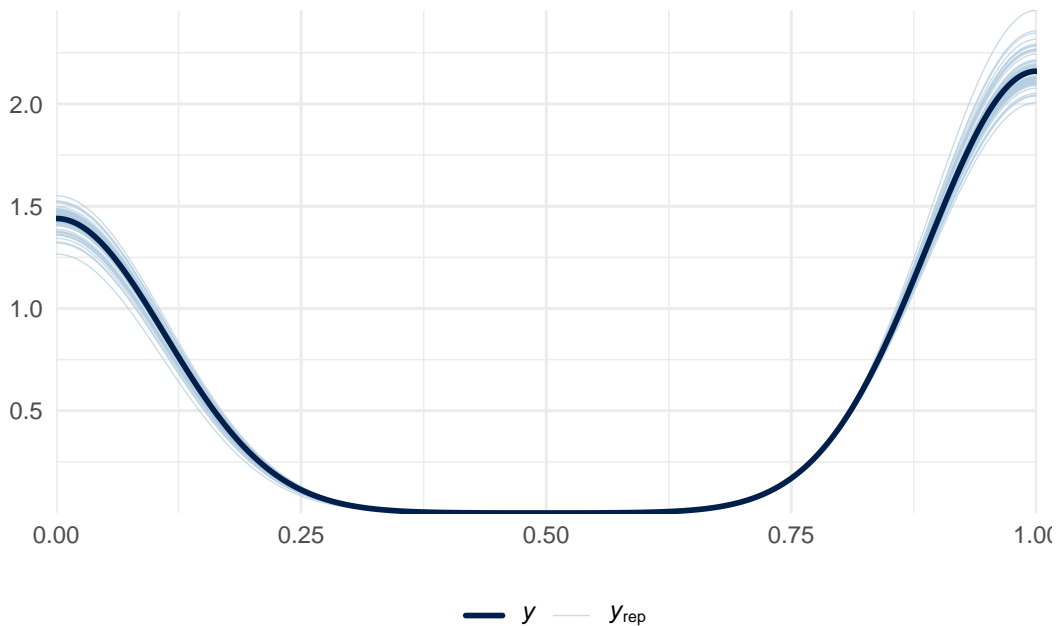
A Appendix

A.1 Shiny application

A Shiny application featuring interactive graphs from the paper can be found here:
https://49z7k8-emily-su.shinyapps.io/top-songs-characteristics-app/

A.2 Posterior predictive check for model

Figure 11 shows that the posterior distribution of the fitted model is consistent with the actual data.

(a) Posterior prediction check for the model

Figure 11: How the data impacts how the model fits

A.3 Diagnostics for model

Figure 12 shows that the sampling algorithm used, the Markov chain Monte Carlo (MCMC) algorithm, did not run into issues as the posterior distribution for the model was created. Both graphs do not show anything abnormal.
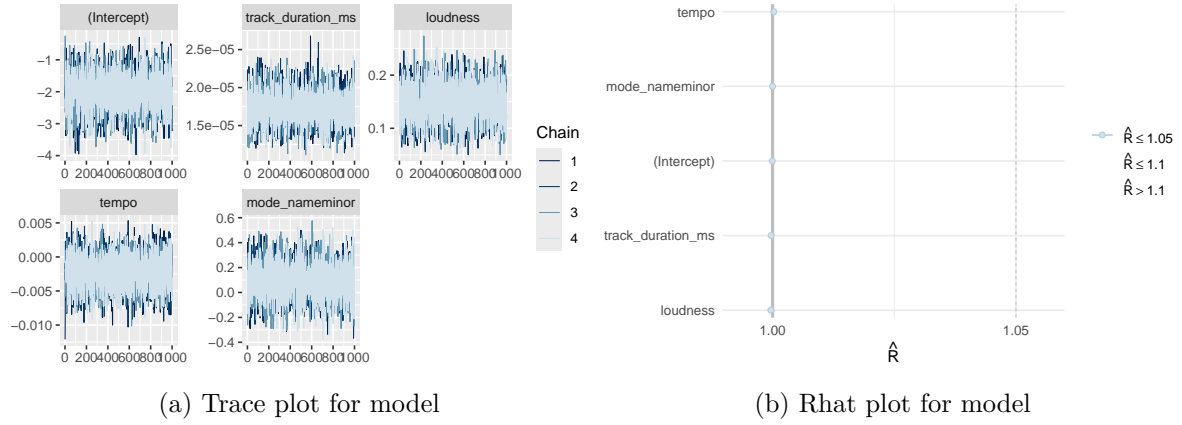
(a) Trace plot for model

(b) Rhat plot for model

Figure 12: Checking the convergence of the Markov chain Monte Carlo (MCMC) algorithm for hit song model
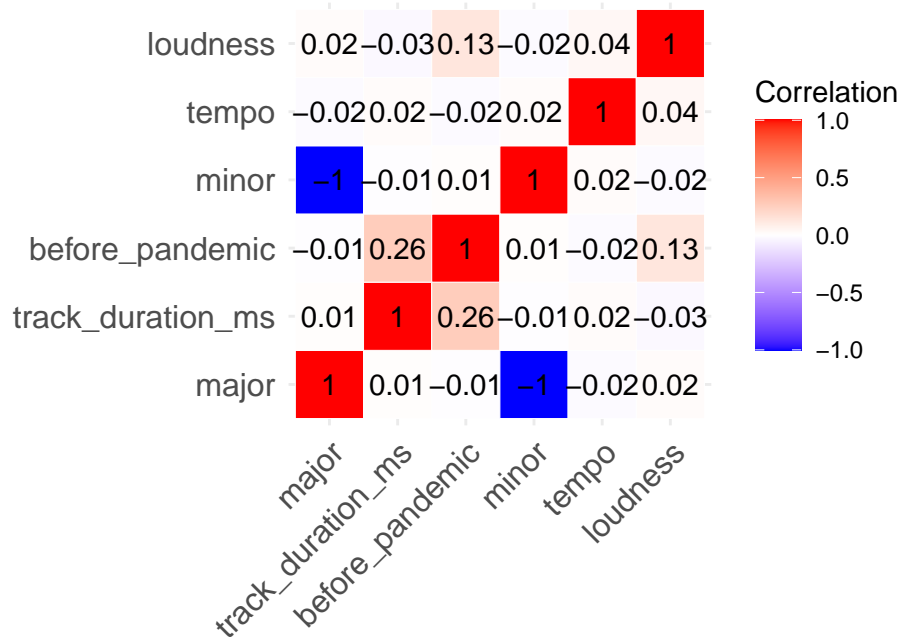
A.4 Additional Figures

Figure 13: Correlation between numerical characteristics of songs from Billboard Year-End Hot 100 singles list and if song was a hit before 2020/the pandemic

A.5 Additional Tables

Table 10: Proportion of songs in different scales before 2020 (the pandemic) versus during and after 2020 (the pandemic) from from Billboard's Year-End Hot 100 singles

Period	Scale	Count of each scale	Proportion of each scale
Before Pandemic	A major	28	0.05
Before Pandemic	A minor	23	0.04
Before Pandemic	A# major	18	0.03
Before Pandemic	A# minor	23	0.04
Before Pandemic	B major	22	0.04
Before Pandemic	B minor	38	0.06
Before Pandemic	C major	41	0.07
Before Pandemic	C minor	17	0.03
Before Pandemic	C# major	64	0.11

Table 10: Proportion of songs in different scales before 2020 (the pandemic) versus during and after 2020 (the pandemic) from from Billboard's Year-End Hot 100 singles

Period	Scale	Count of each scale	Proportion of each scale
Before Pandemic	C# minor	30	0.05
Before Pandemic	D major	41	0.07
Before Pandemic	D minor	2	0.00
Before Pandemic	D# major	5	0.01
Before Pandemic	D# minor	2	0.00
Before Pandemic	E major	15	0.03
Before Pandemic	E minor	19	0.03
Before Pandemic	F major	21	0.04
Before Pandemic	F minor	33	0.06
Before Pandemic	F# major	24	0.04
Before Pandemic	F# minor	17	0.03
Before Pandemic	G major	41	0.07
Before Pandemic	G minor	18	0.03
Before Pandemic	G# major	48	0.08
Before Pandemic	G# minor	10	0.02
During and After Pandemic	A major	19	0.05
During and After Pandemic	A minor	8	0.02
During and After Pandemic	A# major	13	0.03
During and After Pandemic	A# minor	14	0.04
During and After Pandemic	B major	14	0.04
During and After Pandemic	B minor	20	0.05
During and After Pandemic	C major	36	0.09
During and After Pandemic	C minor	12	0.03
During and After Pandemic	C# major	36	0.09
During and After Pandemic	C# minor	15	0.04
During and After Pandemic	D major	29	0.07
During and After Pandemic	D minor	8	0.02

Table 10: Proportion of songs in different scales before 2020 (the pandemic) versus during and after 2020 (the pandemic) from from Billboard's Year-End Hot 100 singles

Period	Scale	Count of each scale	Proportion of each scale
During and After Pandemic	D# major	6	0.01
During and After Pandemic	D# minor	7	0.02
During and After Pandemic	E major	17	0.04
During and After Pandemic	E minor	14	0.04
During and After Pandemic	F major	16	0.04
During and After Pandemic	F minor	14	0.04
During and After Pandemic	F# major	16	0.04
During and After Pandemic	F# minor	18	0.04
During and After Pandemic	G major	23	0.06
During and After Pandemic	G minor	11	0.03
During and After Pandemic	G# major	26	0.06
During and After Pandemic	G# minor	8	0.02

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Bache, Stefan Milton, and Hadley Wickham. 2022. *Magrittr: A Forward-Pipe Operator for r.* https://CRAN.R-project.org/package=magrittr.

Billboard. 2024. "Year-End Charts." Billboard. https://www.billboard.com/charts/year-end/.

Canadian Association of Music Therapists. 2020. "About Music Therapy." Canadian Association of Music Therapists. https://www.musictherapy.ca/about-camt-music-therapy/about-music-therapy/.

Chang, Winston, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2024. *Shiny: Web Application Framework for r.* https://CRAN.R-project.org/package=shiny.

Denis, Kyle, Lyndsey Havens, Jason Lipshutz, Joe Lynch, and Andrew Unterberger. 2023. "Billboard to Alter Chart Tracking Week for Global Release Date." Billboard. https://www.billboard.com/music/chart-beat/year-end-hot-100-morgan-wallen-sza-five-burning-questions-1235553280/.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://github.com/sfirke/janitor.

Ghaffari, Mona, Gohar F. Khan, Shivendu Pratap Singh, and Bruce Ferwerda. 2023. "The Impact of COVID-19 on Online Music Listening Behaviors in Light of Listeners' Social Interactions." *Multimedia Tools and Applications* 83: 13197–239. https://doi.org/10.1007/s11042-023-16079-1.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. *Rstanarm: Bayesian Applied Regression Modeling via Stan.* https://mc-stan.org/rstanarm.

Hespanhol, Luiz, Caio Sain Vallio, Lucíola Menezes Costa, and Bruno T Saragiottoa. 2019. "Understanding and Interpreting Confidence and Credible Intervals Around Effect Estimates." *Brazilian Journal of Physical Therapy* 23. https://doi.org/10.1016/j.bjpt.2018.12.006.

Hester, Jim, Florent Angly, Russ Hyde, Michael Chirico, Kun Ren, Alexander Rosenstock, and Indrajeet Patil. 2024. *Lintr: A 'Linter' for r Code.* https://CRAN.R-project.org/package=lintr.

Hurwitz, Emily Rose, and Carol Lynne Krumhansl. 2021. "Shifting Listening Niches: Effects of the COVID-19 Pandemic." *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.648413.

Iannone, Richard. 2016. *DiagrammeRsvg: Export DiagrammeR Graphviz Graphs as SVG.* https://CRAN.R-project.org/package=DiagrammeRsvg.

Iannone, Richard, and Olivier Roy. 2024. *DiagrammeR: Graph/Network Visualization.* https://CRAN.R-project.org/package=DiagrammeR.

Kassambara, Alboukadel. 2023. *Ggcorrplot: Visualization of a Correlation Matrix Using 'Ggplot2'.* https://CRAN.R-project.org/package=ggcorrplot.

Manitoba Music. 2024. "Nielsen SoundScan." Manitoba Music. https://www.manitobamusic.com/profiles/view,499/soundscan.

Minitab. 2024. *Regression Coefficients.* https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/regression/supporting-topics/regression-models/regression-coefficients/.

Müller, Kirill, and Lorenz Walthert. 2024. *Styler: Non-Invasive Pretty Printing of r Code.* https://CRAN.R-project.org/package=styler.

Ooms, Jeroen. 2023. *Rsvg: Render SVG Images into PDF, PNG, (Encapsulated) PostScript, or Bitmap Arrays.* https://CRAN.R-project.org/package=rsvg.

Perrier, Victor, Fanny Meyer, and David Granjon. 2024. *shinyWidgets: Custom Inputs Widgets for Shiny.* https://CRAN.R-project.org/package=shinyWidgets.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'.* https://github.com/apache/arrow/.

Thompson, Charlie, Daniel Antal, Josiah Parry, Donal Phipps, and Tom Wolff. 2022. *Spotifyr: R Wrapper for the 'Spotify' Web API.* https://CRAN.R-project.org/package=spotifyr.

Urbanek, Simon. 2022. *Png: Read and Write PNG Images.* https://CRAN.R-project.org/package=png.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Jennifer Bryan, Malcolm Barrett, and Andy Teucher. 2024. *Usethis: Automate Package and Project Setup.* https://CRAN.R-project.org/package=usethis.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.