Characteristics of Top Songs Has Changed from Pandemic Brain*

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

Emily Su

April 2, 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 shorter, quieter, but slightly faster with the melody of songs being in a major key. 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 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 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. 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 an interactive graph of some of the results.

2 Data

[To Do]

2.1 Variables of Interest

2.2 Data Source and Measurements

2.3 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), testthat (Wickham 2011). 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).

3 Model

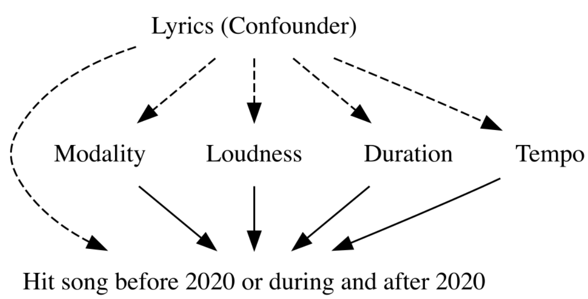

Figure 1: Causal relationship between song characteristics and hit song (before 2020 or during and after 2020)

[To Do]

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.

3.1 Model set-up

We will define our logistic regression model as follows.

$$
\begin{aligned}
y_i|\pi_i &\sim \text{Bern}(\pi_i) && (1)\\
\text{logit}(\pi_i) &= \beta_0 + \beta_1 \times \text{duration}_i + \beta_2 \times \text{loudness}_i + \beta_3 \times \text{tempo}_i + \beta_4 \times \text{modality}_i && (2)\\
\beta_0 &\sim \text{Normal}(0, 2.5) && (3)\\
\beta_1 &\sim \text{Normal}(0, 2.5) && (4)\\
\beta_2 &\sim \text{Normal}(0, 2.5) && (5)\\
\beta_3 &\sim \text{Normal}(0, 2.5) && (6)\\
\beta_4 &\sim \text{Normal}(0, 2.5) && (7)\\
& && (8)
\end{aligned}
$$

We define y_i to be a top song, 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 that a top song i is 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.

We ran the model using the rstanarm package of Goodrich et al. (2024) and we used the default priors from rstanarm.

3.1.1 Model justification

[To Do]

4 Results

[To Finish]

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

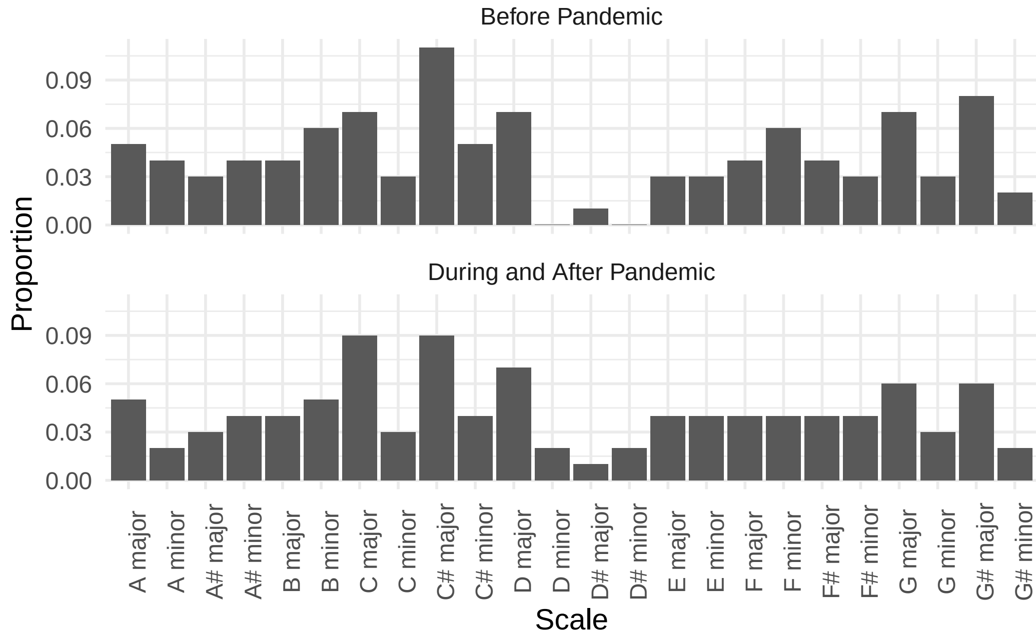

Figure 2: Proportion of songs in different scales before 2020 versus during and after 2020

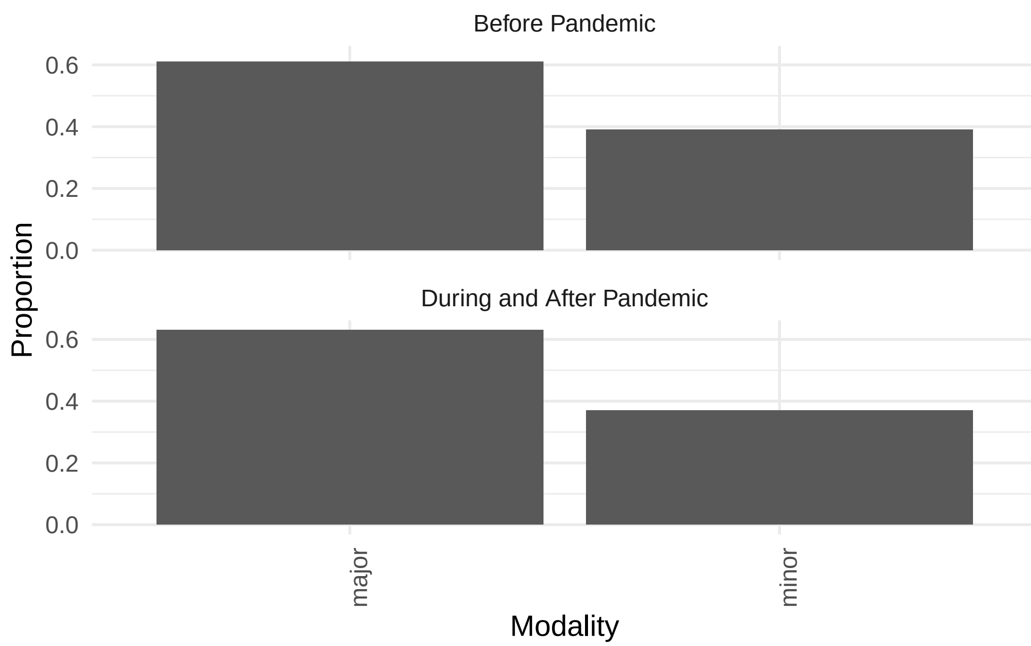

Figure 3: Proportion of songs whose modality is in a major or minor key before 2020 versus during and after 2020

Table 1: Proportion of songs whose modality is in a major or minor key before 2020 versus during and after 2020

Period	Modality	Count of each mode	Proportion of each mode
Before Pandemic	major	368	0.61
Before Pandemic	minor	232	0.39
During and After Pandemic	major	251	0.63
During and After Pandemic	minor	149	0.37

4.1.2 Track Duration

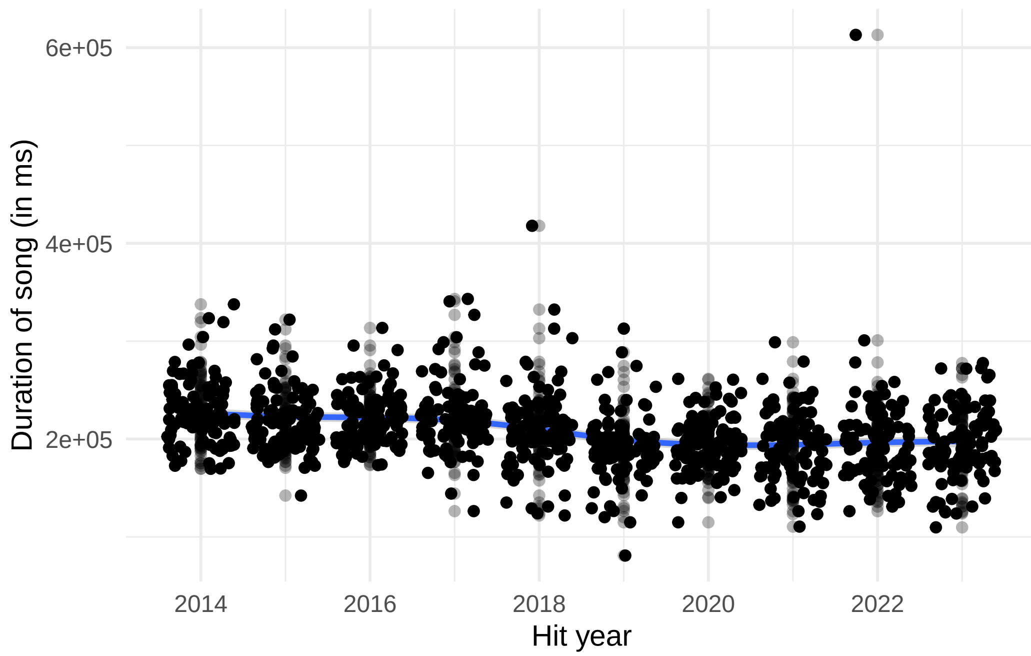

Figure 4: Relationship between the hit year and track duration of a song

Table 2: Minimum, quartiles, median, and maximum of track duration (in ms) before 2020

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

Table 3: Minimum, quartiles, median, and maximum of track duration (in ms) during and after 2020

Duration of song (in ms)
Min. :109750
1st Qu.:173369
Median :195120

Table 3: Minimum, quartiles, median, and maximum of track duration (in ms) during and after 2020

Duration of song (in ms)
Mean :195999
3rd Qu.:215336
Max. :613026

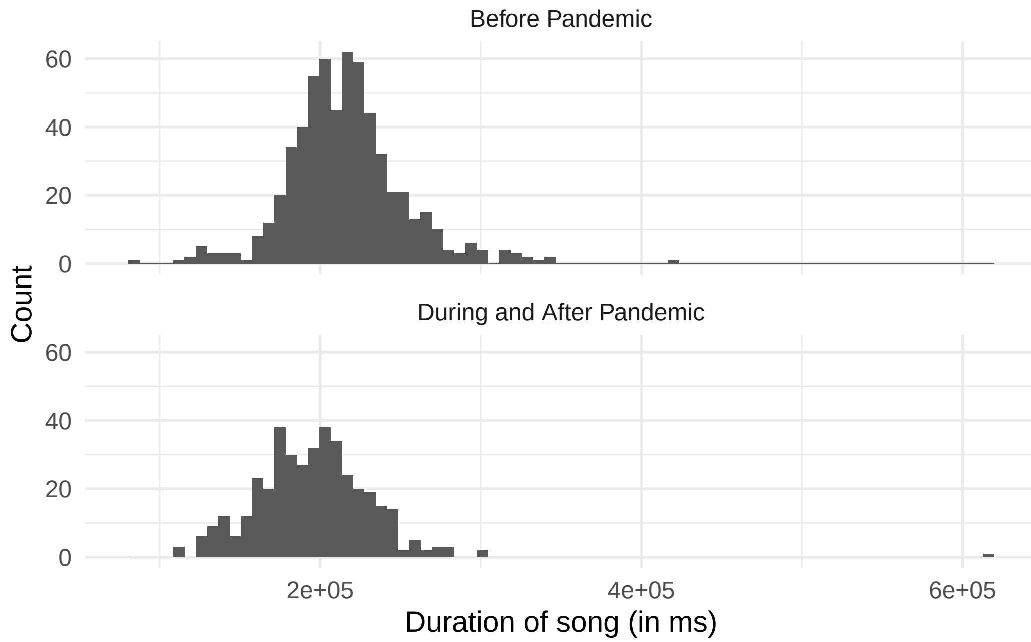

Figure 5: Distribution of track duration (in ms) before 2020 versus during and after 2020

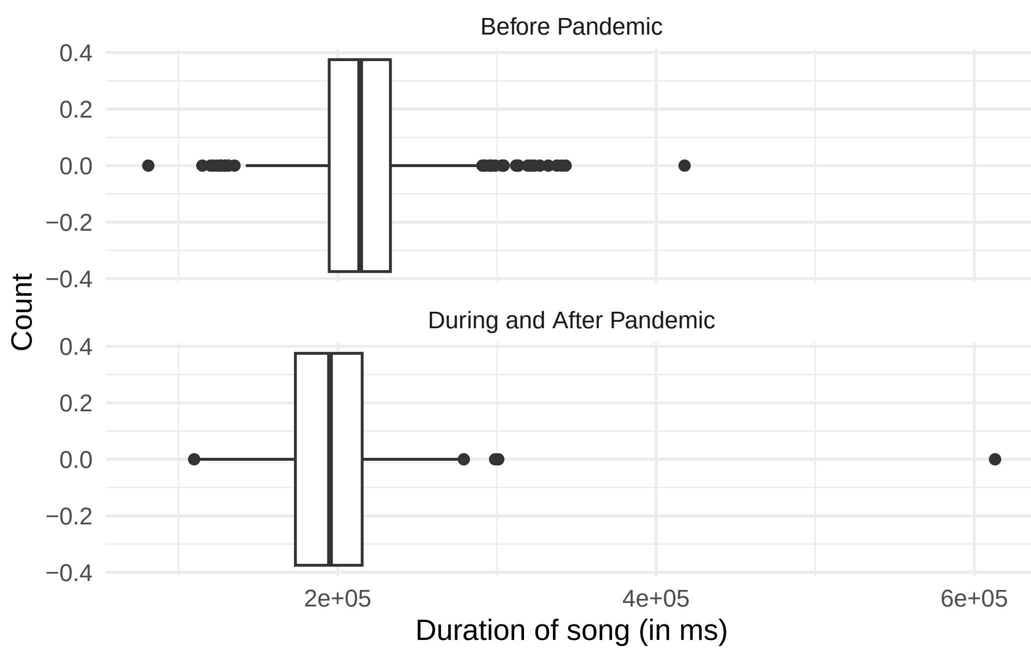

Figure 6: Track duration (in ms) before 2020 versus during and after 2020

4.1.3 Loudness

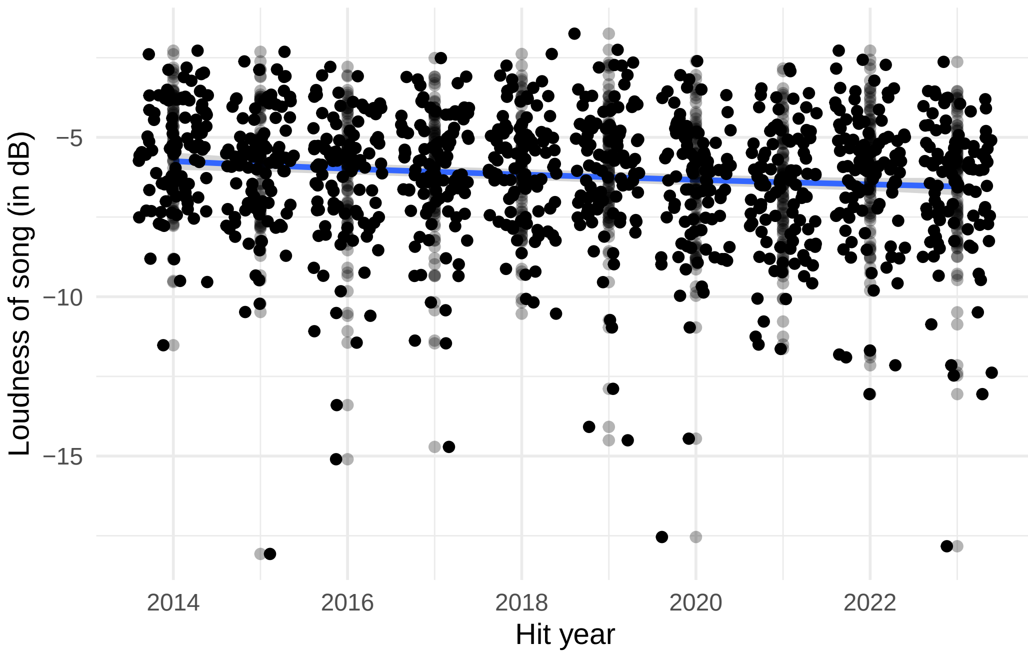

Figure 7: Relationship between the hit year and loudness of a song

Table 4: Minimum, quartiles, median, and maximum of loudness of song (in dB) before 2020

Loudness of song (in dB)
Min. :-18.071
1st Qu.: -7.051
Median : -5.700
Mean : -5.964
3rd Qu.: -4.705
Max. : -1.746

Table 5: Minimum, quartiles, median, and maximum of loudness of song (in dB) during and after 2020

Loudness of song (in dB)
Min. :-17.829
1st Qu.: -7.604
Median : -6.151

Table 5: Minimum, quartiles, median, and maximum of loudness of song (in dB) during and after 2020

Loudness of song (in dB)
Mean : -6.516
3rd Qu.: -5.109
Max. : -2.278

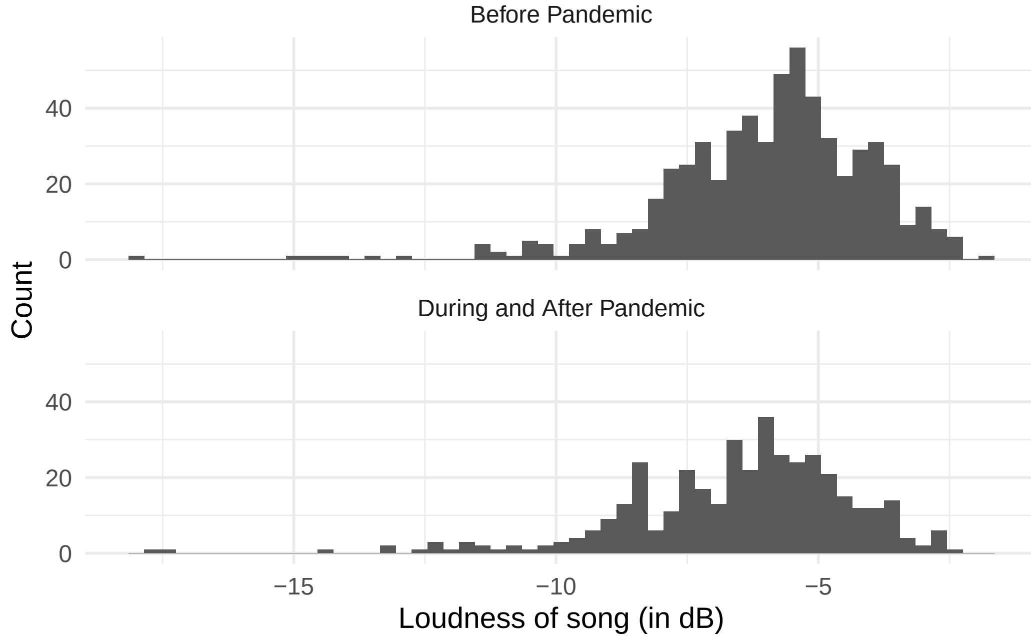

Figure 8: Distribution of loudness (in dB) before 2020 versus during and after 2020

4.1.4 Tempo

Table 6: Minimum, quartiles, median, and maximum of tempo of a song (BPM) before 2020

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

Table 7: Minimum, quartiles, median, and maximum of tempo of a song (BPM) during and after 2020

Tempo of song (BPM)
Min. : 67.03
1st Qu.: 98.02
Median :120.03

Table 7: Minimum, quartiles, median, and maximum of tempo of a song (BPM) during and after 2020

Tempo of song (BPM)
Mean :122.06
3rd Qu.:142.31
Max. :205.86

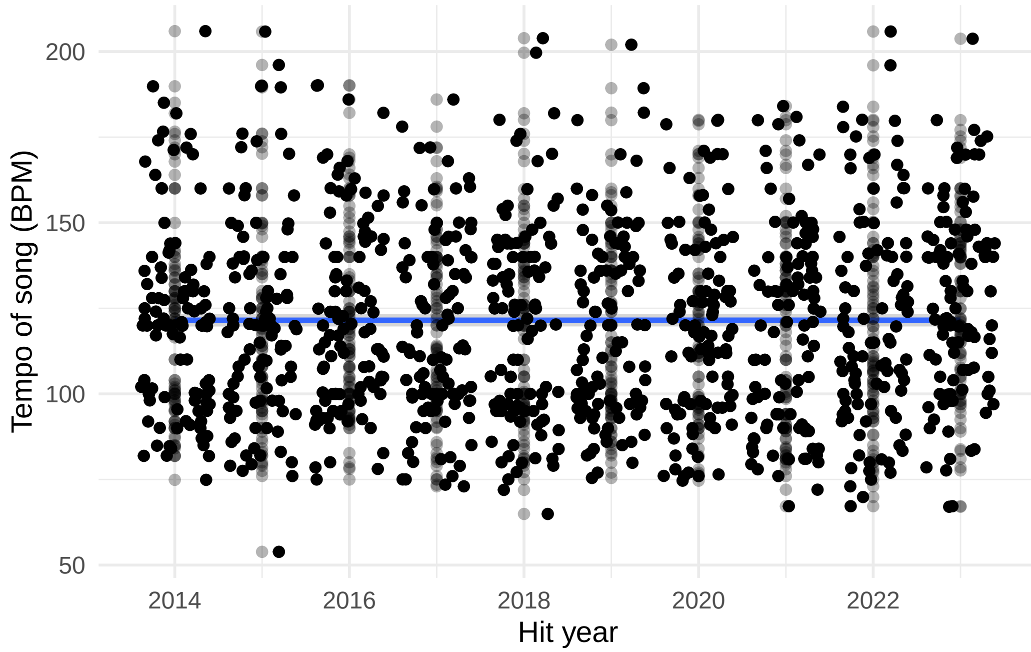

Figure 9: Relationship between the hit year and tempo of a song

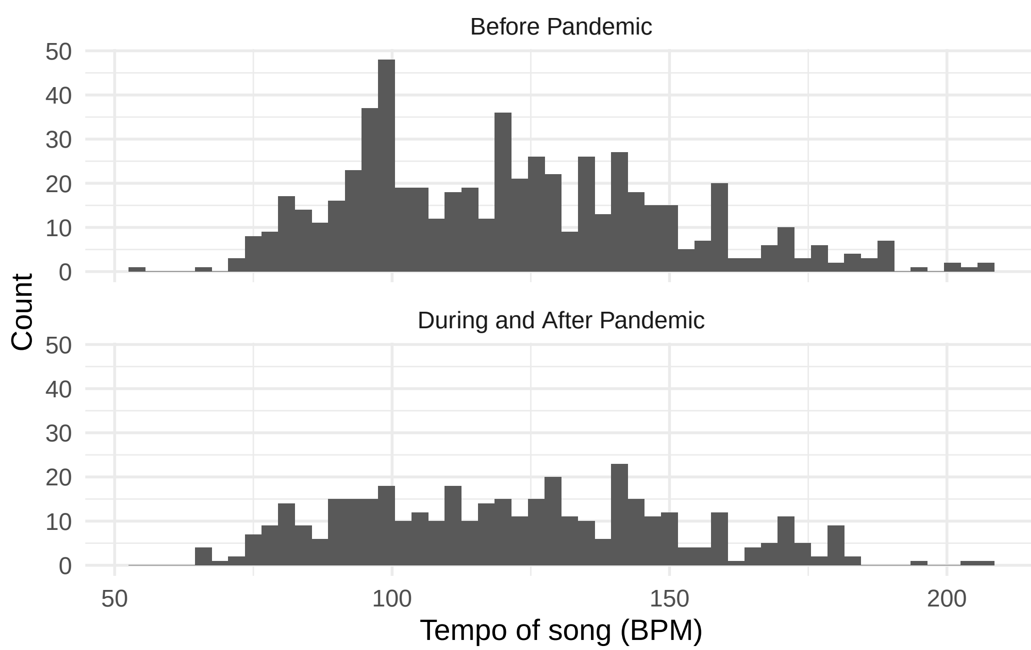

Figure 10: Distribution of tempo (BPM) before 2020 versus during and after 2020

4.2 Model Results

[To Do]

5 Discussion

[To Do]

A Appendix

A.1 Additional Figures and Tables

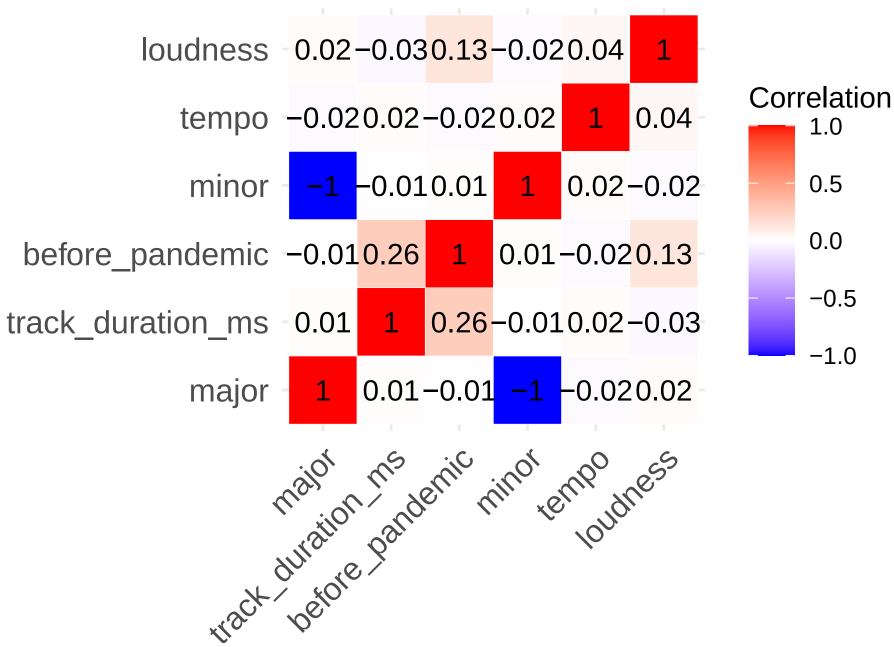

Figure 11: Correlation between numerical characteristics of songs from Billboard Year-End Hot 100 single and if song was a hit before 2020/the pandemic

A.2 Model details

[To Do]

A.2.1 Posterior predictive check

A.2.2 Diagnostics

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.

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/.

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.

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.

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

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.