Data Visualization Final Project Report:

Spotify Music Analysis

**Introduction**

Each year, Spotify music streaming app provides a statistical overview, known as "Spotify Wrapped", of the user's most listened to songs, genres and artists. However, the "Spotify Wrapped" summary is only available annually in December, and previous yearly summaries are not available for the user to view. Thankfully, each Spotify user can request their own listening history data, which includes data from every year since account creation. The motivation of this project is to explore nine years of music streaming data from my Spotify account and answer several questions regarding my personal taste in music.

**Research Questions**

I. What are the characteristics of my most played songs?

II. How diverse is my music taste, do I tend to listen to the same songs/artists/genres?

III. Has my music taste changed since 2015?

**Methods**

There are two different sources of data for this project. The first source is music streaming history data for the years 2015 - 2023, requested directly from Spotify. This dataset only includes categorical variables such as the artist, album, and track name as well as continuous variables to describe the duration of the track played. The second source is data requested from the Spotify Web API, which includes numerous descriptive variables for each track such as dance-ability, energy etc. Exportify is a web application that allows users to export descriptive data (from Spotify Web API) for songs in a designated playlist. To combine both datasets, I determined the 100 most played songs for each year using the streaming history data, then manually created a Spotify playlist for each year and exported the descriptive data for each playlist using Exportify.

**Results**

To understand the character of my music taste by each year, I’ve examined the distribution of genre for my most played songs from 2015 to 2023. Overall the number of unique genres has increased over time, with only 17 distinct genres in 2015 to 23 genres the following year. This increase in distinct genres continues until 2020, when the number of distinct genres drops from 30 to 26, but continues to increase to 31 genres by 2023. In 2015 the most frequent genres that I’ve listened to, appear to be “Garage Rock”, “Emo”, and “Acid Rock”. These genres appear over 10 times, with “Garage Rock” appearing over 40 times out of my top 100 songs from 2015. “Garage Rock”, “Grunge”, and “Emo” appear to be the dominant genres in 2015 and 2016. However, this trend changes for the subsequent two years. In 2017 and 2018 one of the most played genres appear to be “Conscious Hip-Hop”, especially in 2018 where majority of the genres appear to be some form of R&B or Rap. Additionally, 2019 is a blend of both types of music, with the most frequent genres such as “Emo”, “Bubble-grunge” and Hip Hop. Interestingly, the most frequently played genres in 2020 are similar to the top genres in 2015. In 2020, the most played genre was “Emo”, which appeared over 30 times. This trend continues in 2021, with an addition of the genre “Anime Score” appearing over 15 times. In 2022 the most played genre is “Bubble-grunge” again, which appears over 15 times. Finally, the most played genre in 2023 appears to be “Alternative Rock”. While there is a variety of different genres in my most played songs, several genres such as “Garage Rock” reoccur throughout the dataset.

Moreover, the diversity of my music taste was measured by the number of unique artists, albums, and songs I’ve listened to each year. In general, there appears to be an increasing trend for unique albums and artists until the year 2020. There is a sharp decline in the diversity of albums and artists for subsequent years, with the minimum point occurring in 2022. The diversity in my music taste appears to fluctuate over time. Additionally, I compared the popularity score for each of my top songs to a general dataset comprised of songs from Spotify. The popularity variable is a score from 0 to 100 based off of the popularity of songs by the same artist. A high popularity score indicates that the song in question is the artists most popular song. The density of the popularity score changes from my top songs in 2015 to 2023. For example, there appears to a high concentration of songs with a popularity score of 40, for 2020 to 2022, in the general Spotify data set there is a peak at a score of 70 for the same years. As opposed to 2015 to 2016, there appears to be a higher concentration of songs with a popularity of 70, but in the general Spotify data set there are higher concentrations at scores of approximately 65 for the same time period. Overall, my music taste has seemed to increase in diversity and decrease in popularity over time.

Additionally, the character of my music taste, does not appear to change in a year by year comparison, however when compared to general Spotify song data, the loudness, dance-ability, energy and popularity of my top songs varies greatly. The median loudness of my most frequently played songs appears to be approximately -5.0 to -7.0 dB for years 2015 to 2023. However, when compared to the general Spotify data set, the density of loudness for my most listened to songs has more variation. The general Spotify data set has a high concentration of songs with about -5 dB throughout years 2015 – 2022, whereas the peak density of my top songs changes each year. Track energy and popularity follows a similar trend to loudness, where the median value of each variable does not appear to change much year by year, yet when compared to the Spotify data set there is much more variation. The median dance-ability of my top songs increases to nearly 0.75 in 2018 and then decreases back to under 0.50 by 2023. Interestingly, in the general Spotify data set, as time passes there are more songs with a greater dance-ability score.

**Conclusion**

My music taste appears to vary widely in terms of genre and in terms of characteristics such as dance-ability. Over time, the most popular genres for me have changed from “Garage Rock” to “Conscious Hip-Hop” to “Garage Rock” back again. Over nine years, there was an increase in distinct genres I listened to, however there seem to be a general decrease in album, artist and song diversity around 2022. A possible reason for the sharp drop in 2022 could be that I streamed much less Spotify music compared to other years. Additionally, while there was an increase in unique music genres, this may not actually be the case because of how genre was recorded by Spotify. Each song had multiple comma separated genres, therefore only the first genre that was listed for each song was kept. Also, the label of each genre was very specific, as a result there were many categories. Genres such as “Garage Rock” and “Alternative Rock” could have been consolidated to one genre; Rock. It may be useful to designate general categories for genre in future analysis.

**Final Remarks**

This analysis of my music taste was very insightful. Further study may include using descriptive variables as predictors in a linear regression, or creating a predictive model.

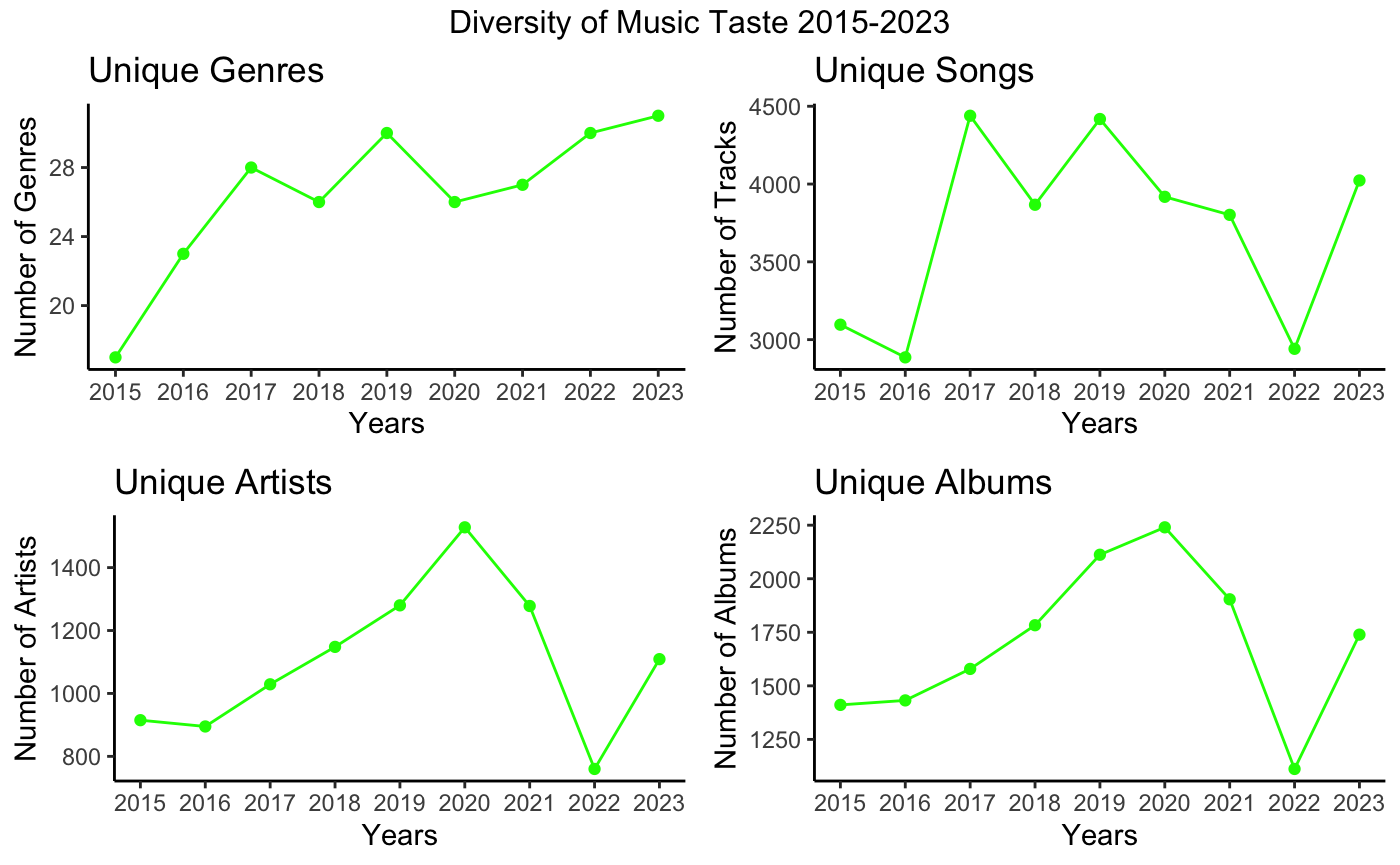
Appendix 1: Summary

*Table 1.0: List of Variables with Variable Type and Descriptions*

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Type** | **Description** |
| Track Name | Categorical | Name of the song played |
| Artist Name | Categorical | Name of the attributed artist |
| Album Name | Categorical | Name of the album the track belongs to |
| Year | Categorical | Year of streaming history 2015 – 2023 |
| Genre\* | Continuous | 1st Genre in list |
| Energy | Continuous | Perceived intensity of the track including perceived loudness and dynamic range  Scale: 0 to 1 |
| Dance-ability | Continuous | How suitable a track is for dancing combination of rhythm and beat factors  Scale: 0 to 1 |
| Popularity | Continuous | How popular the track relative to popularity of other tracks by the artist  Scale: 0 to 100 |
| Loudness | Continuous | Overall loudness in decibels  Scale: -60 to 0 |
| Tempo | Continuous | Beats Per Minute |

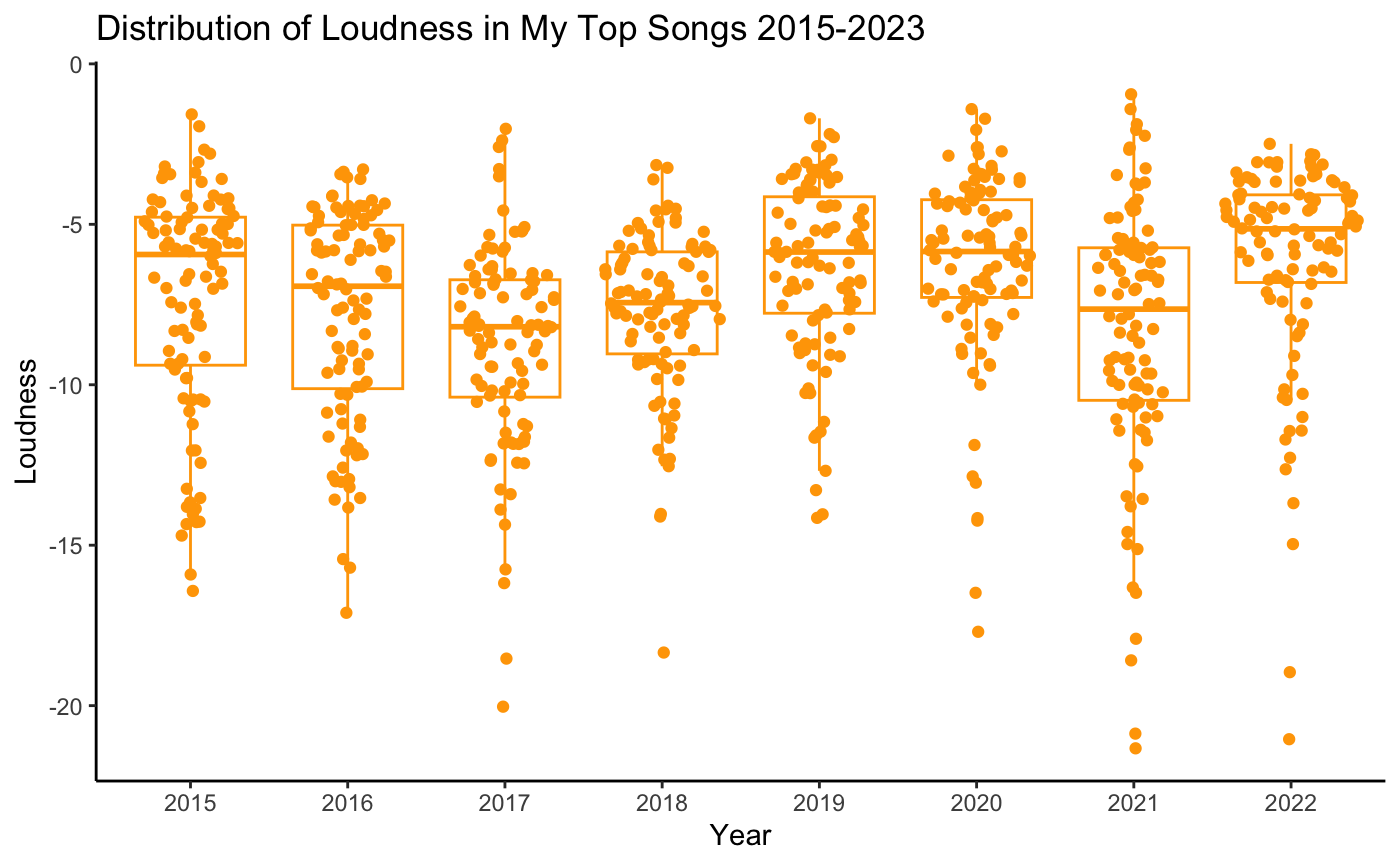
\*Genre information is stored as a single string of several genre labels the artist is associated with (ex. "surf rock", "hard rock", "canadian metal")

*Figure 1.0: Total Number of Unique Artists, Albums and Songs by Year*

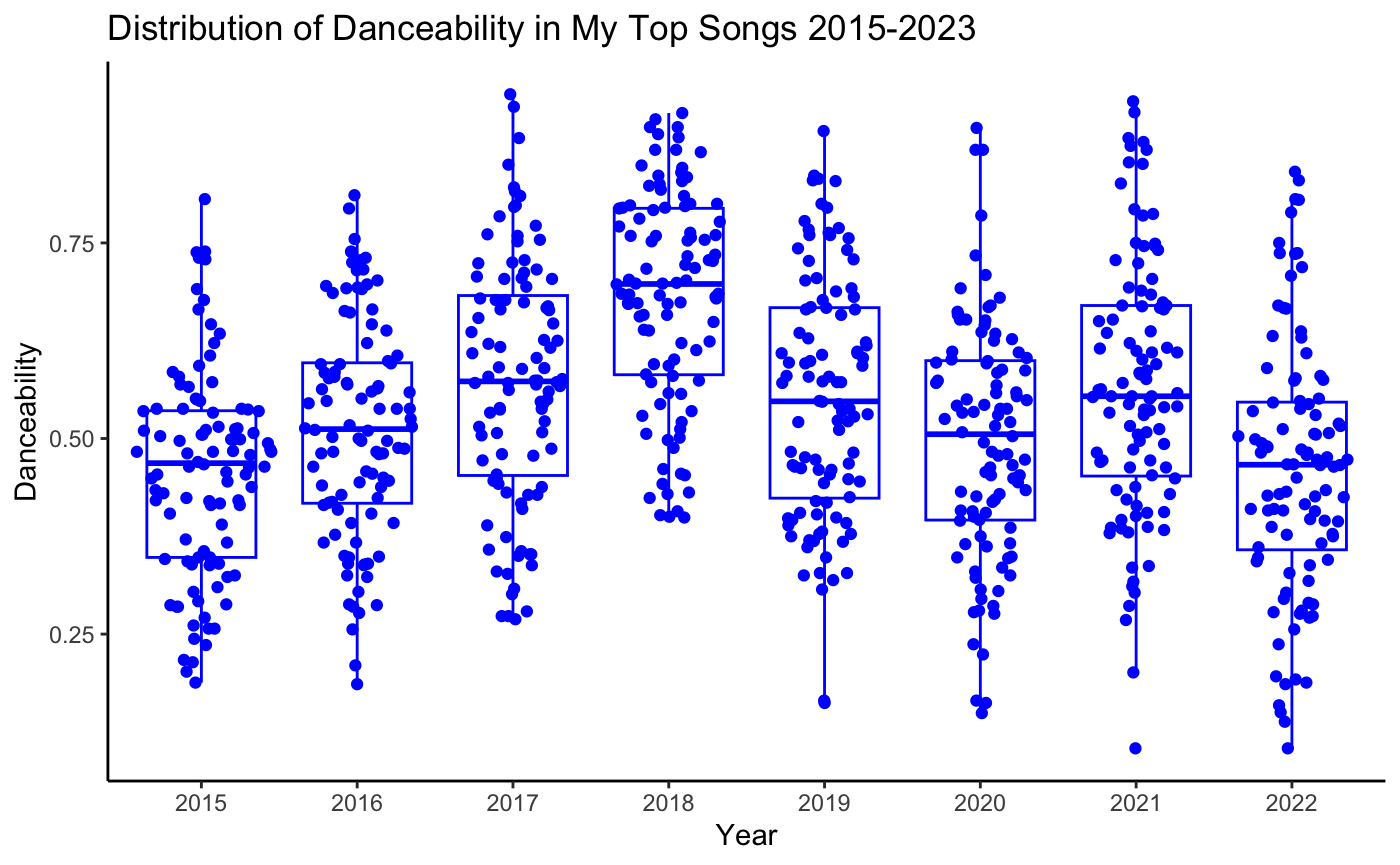


Appendix 2: Comparison Between Years

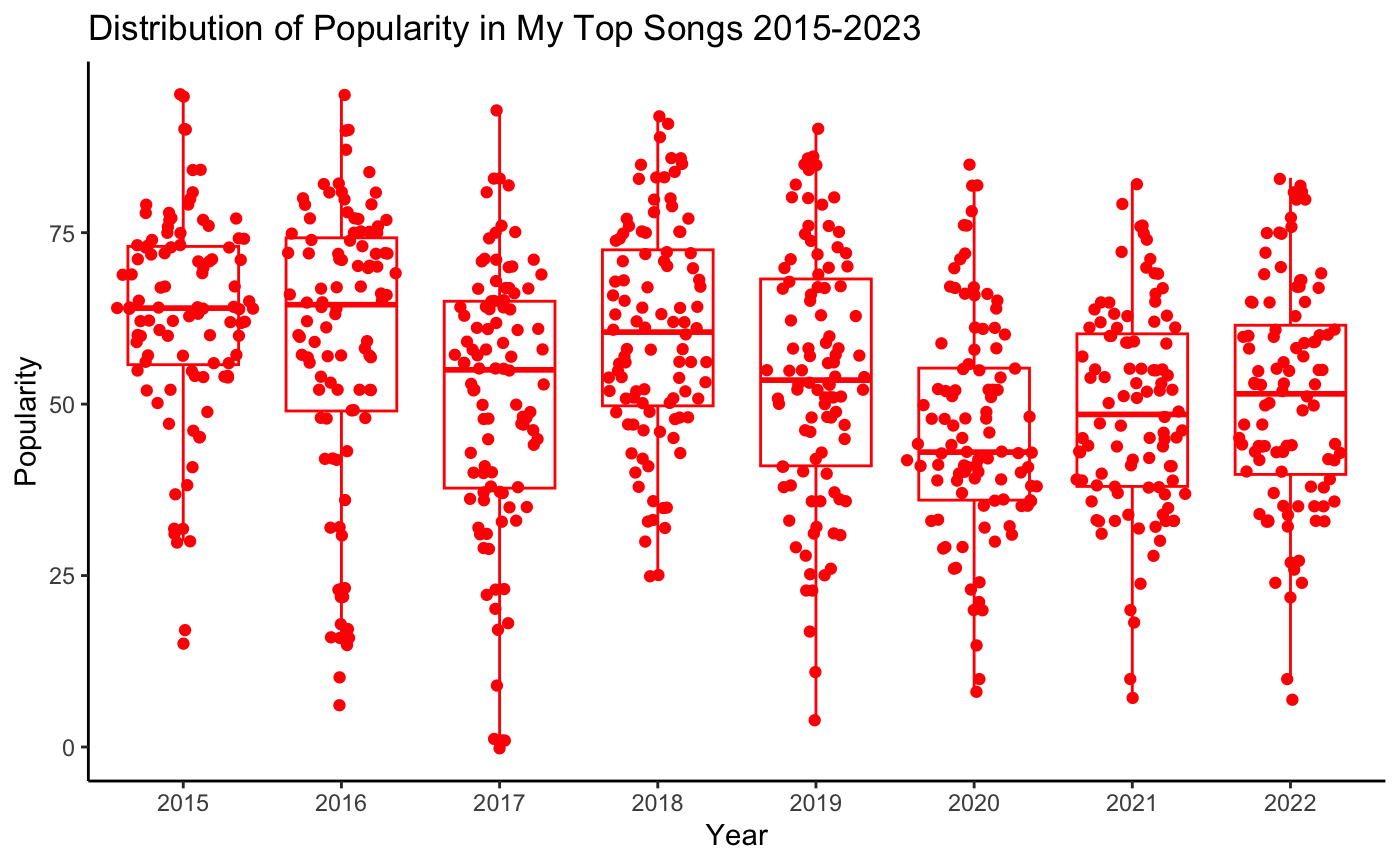
*Figure 2.1: Distribution of Song Loudness by Year*



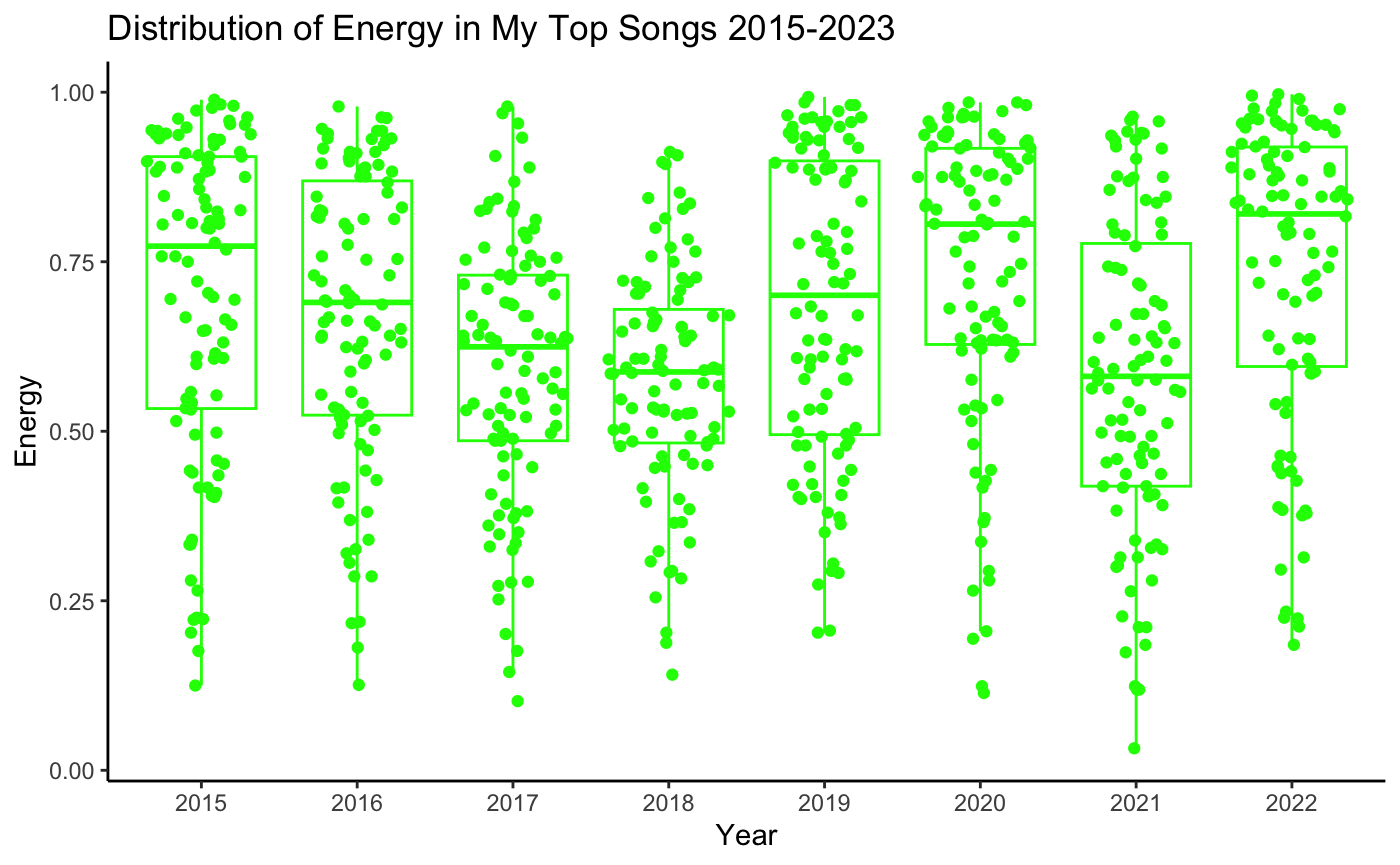
*Figure 2.2: Distribution of Song Dance-ability by Year*



*Figure 2.3: Distribution of Song Popularity by Year*

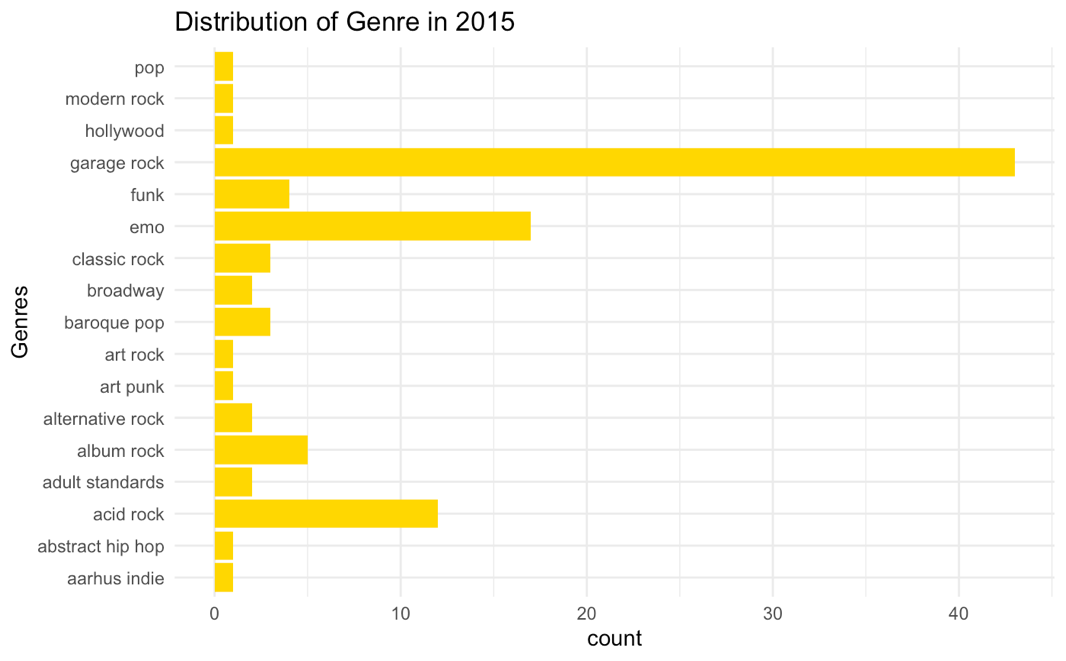


*Figure 2.4: Distribution of Song Energy by Year*

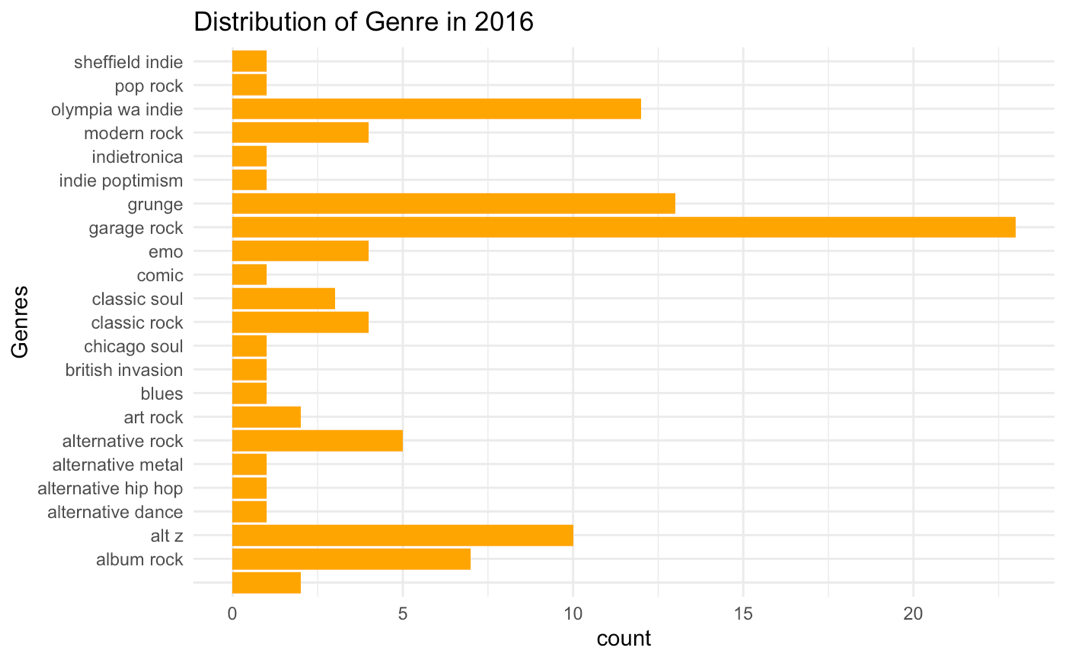


Appendix 3: Genre Distribution

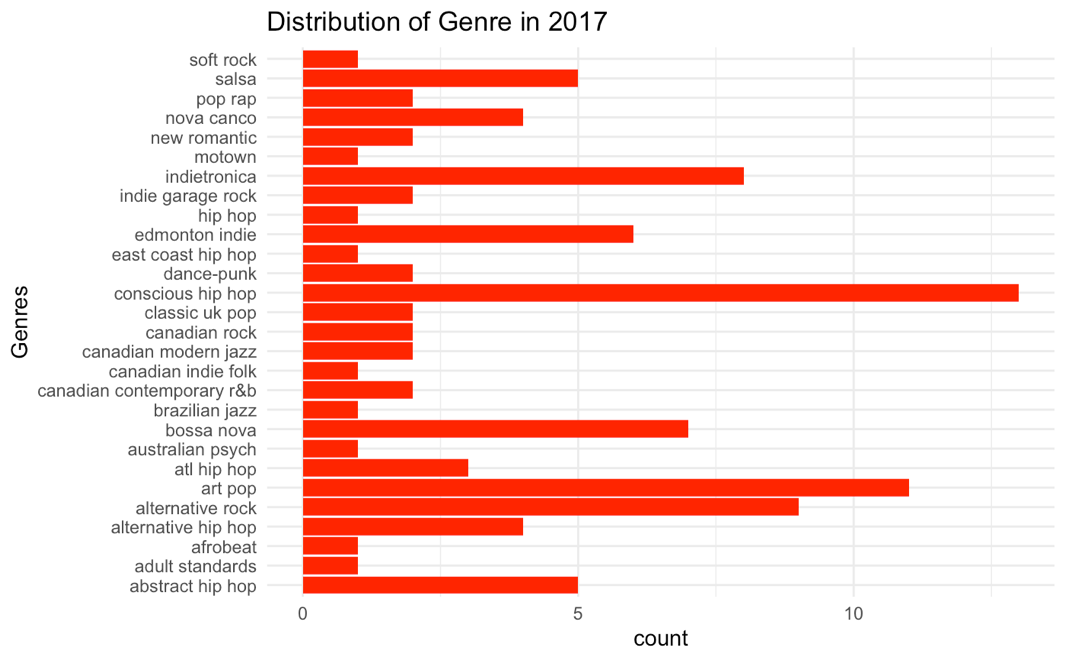
*Figure 3.1: Genres 2015*



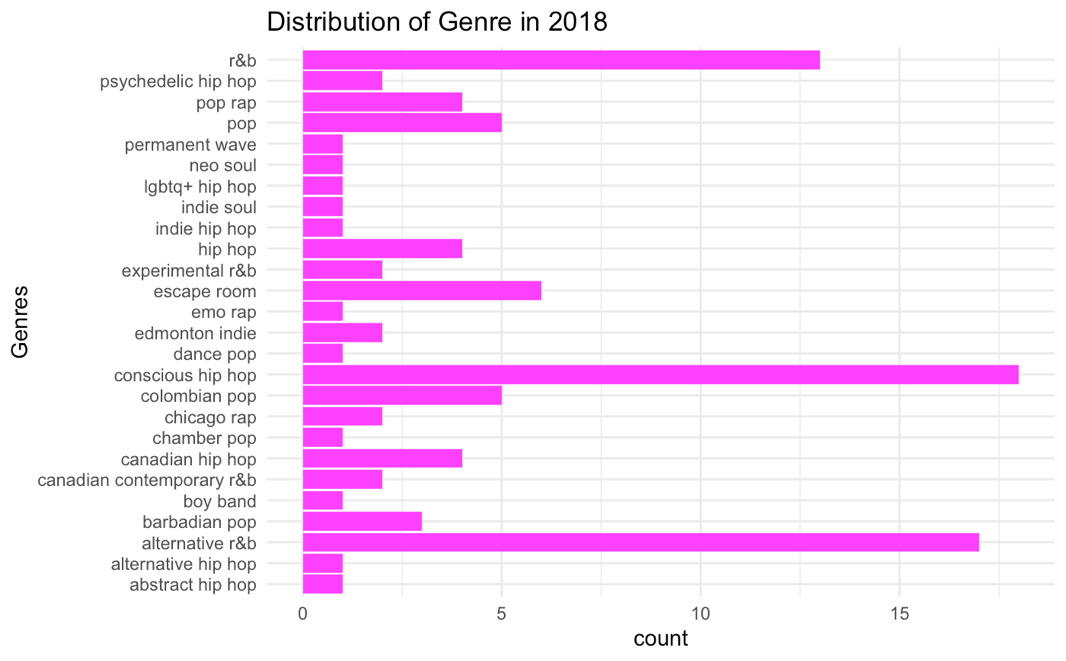
*Figure 3.2: Genres 2016*



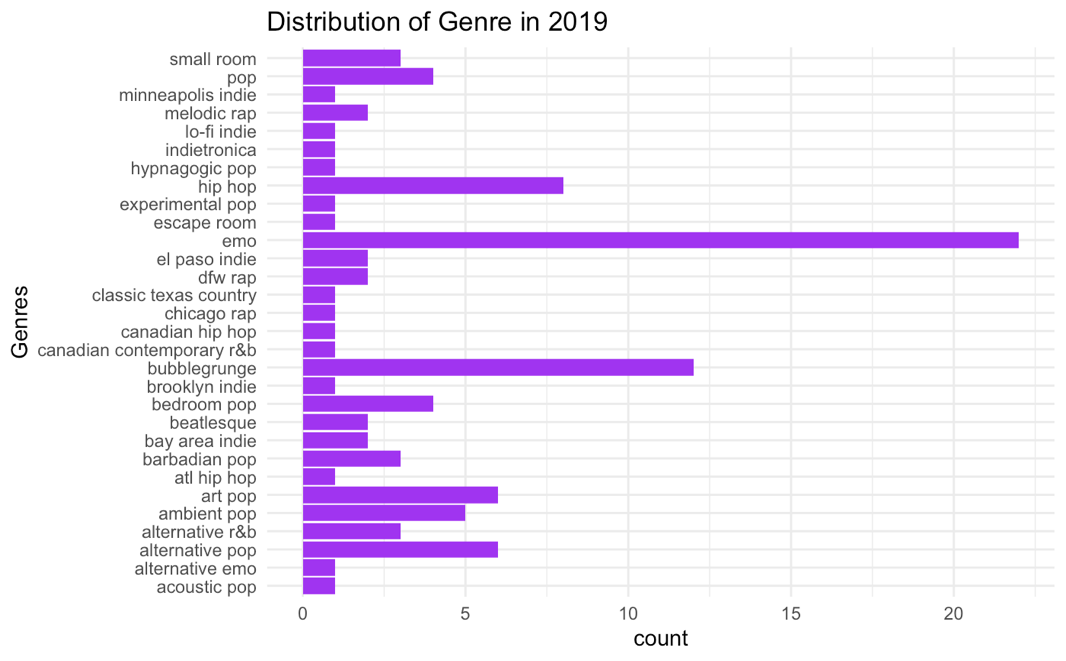
*Figure 3.3: Genres 2017*



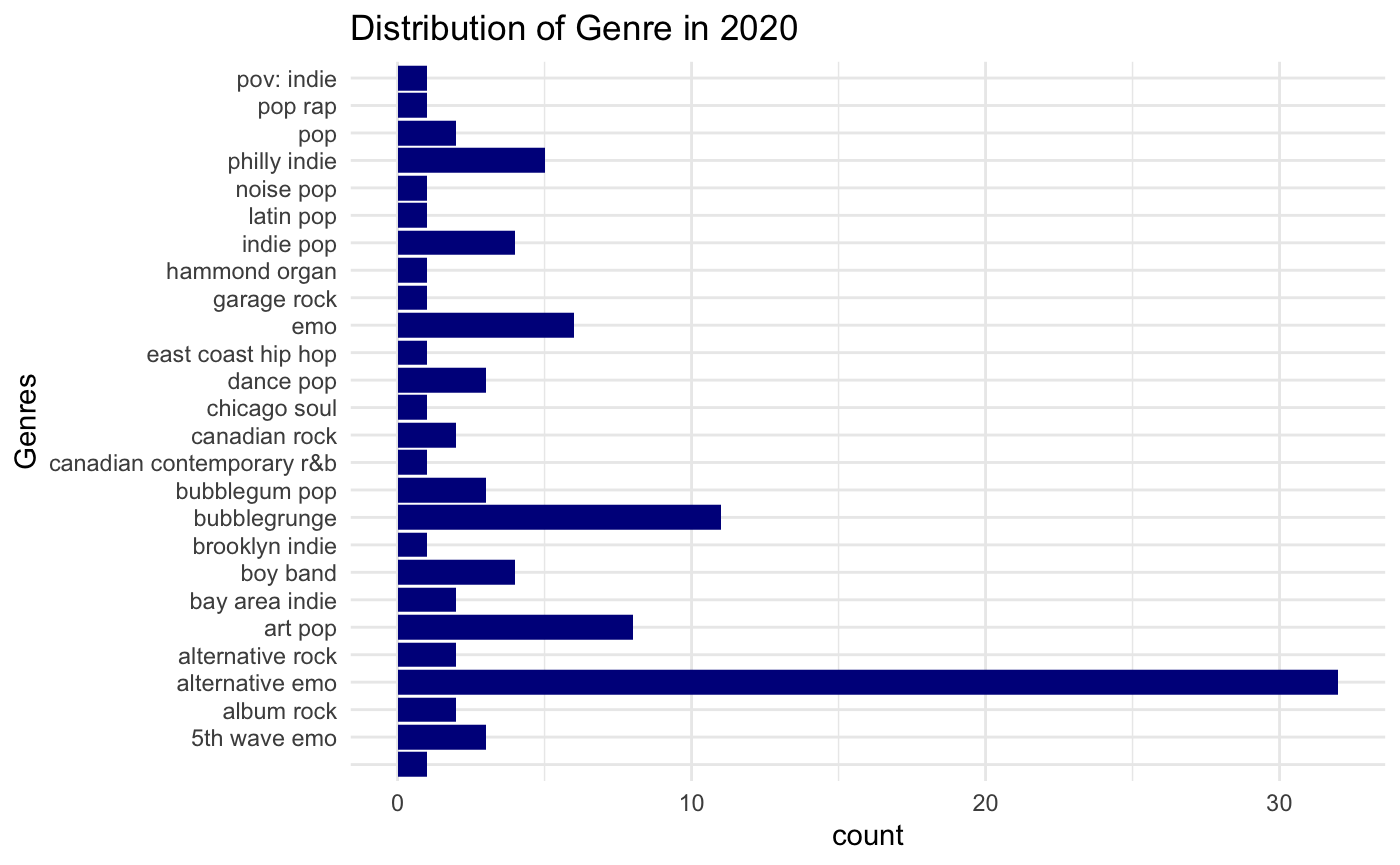
*Figure 3.4: Genres 2018*



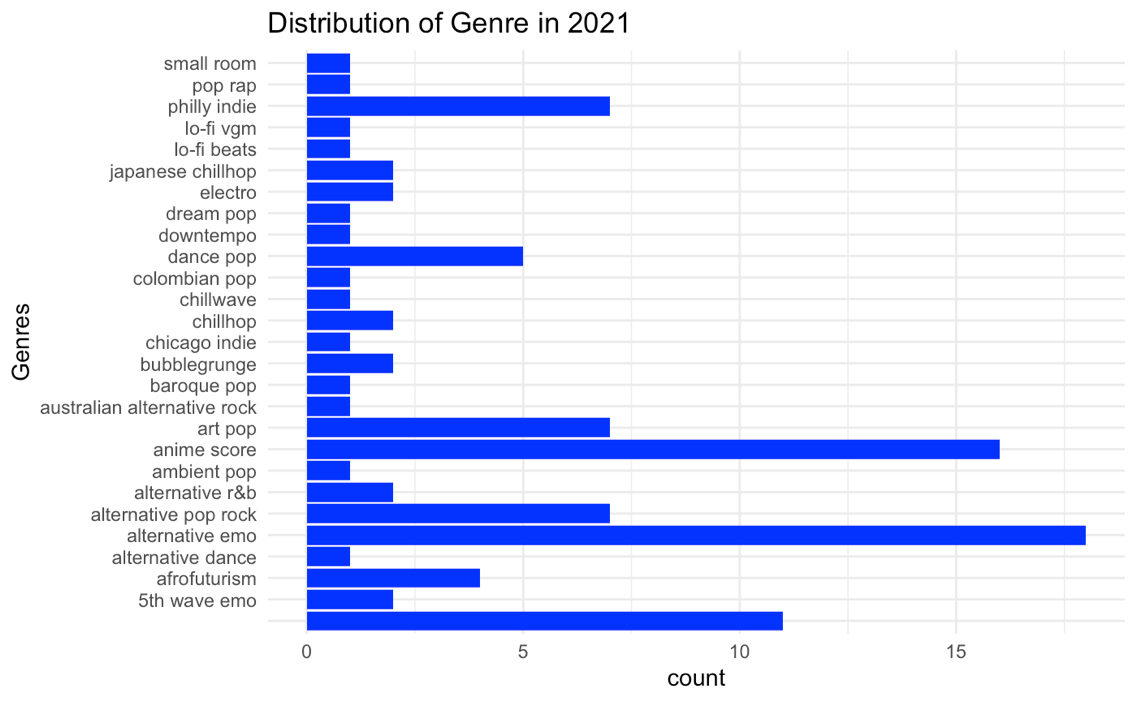
*Figure 3.5: Genres 2019*



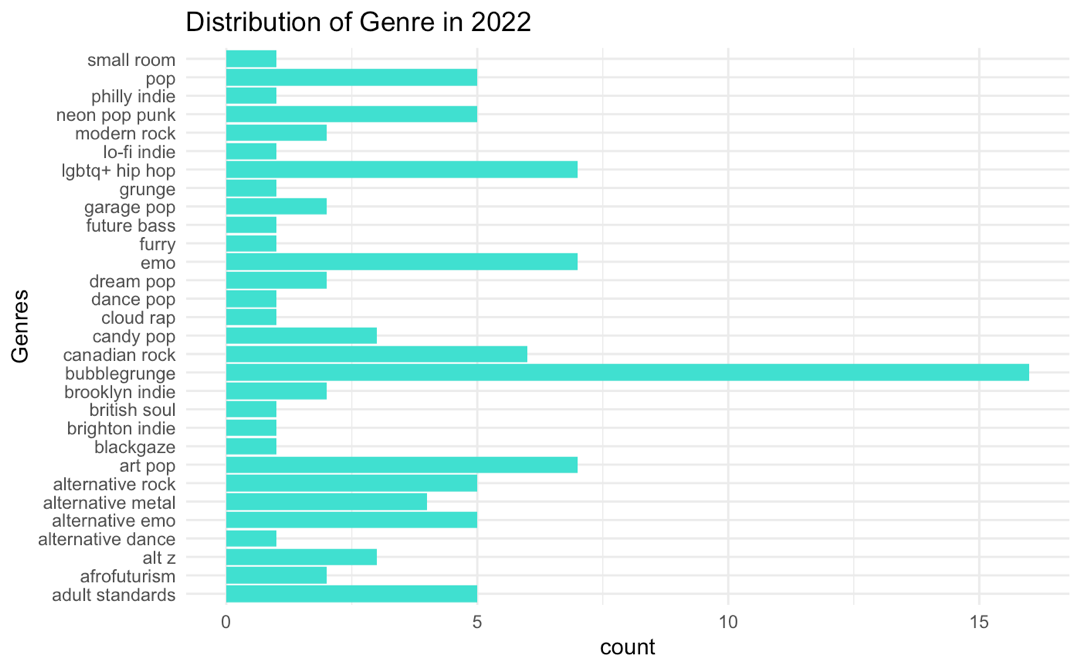
*Figure 3.6: Genres 2020*

**

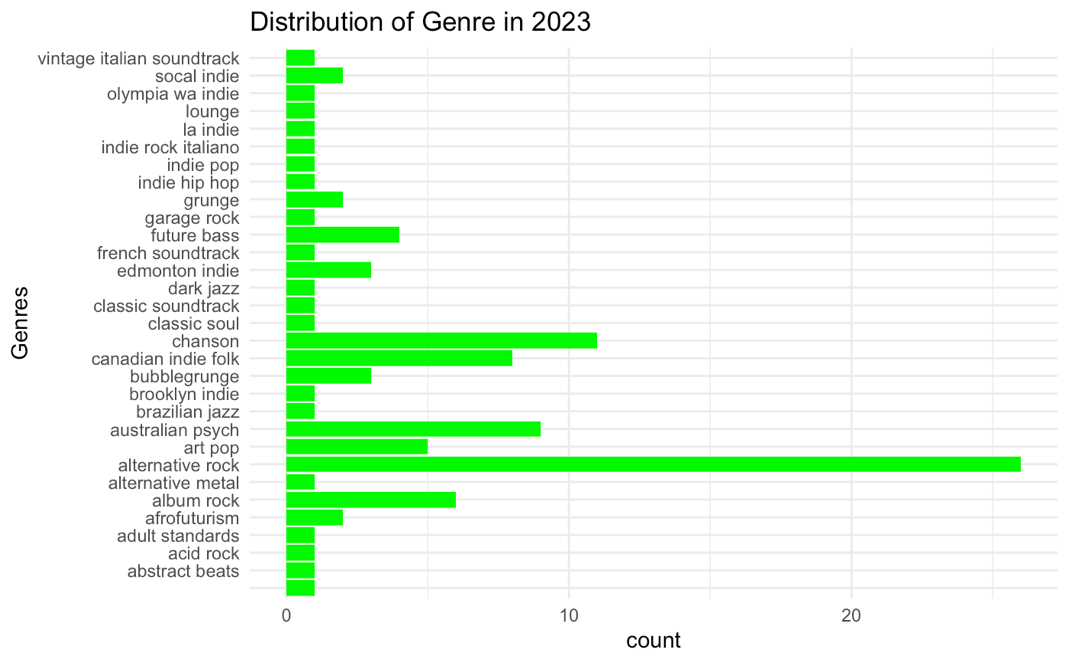
*Figure 3.7: Genres 2021*



*Figure 3.8: Genres 2022*

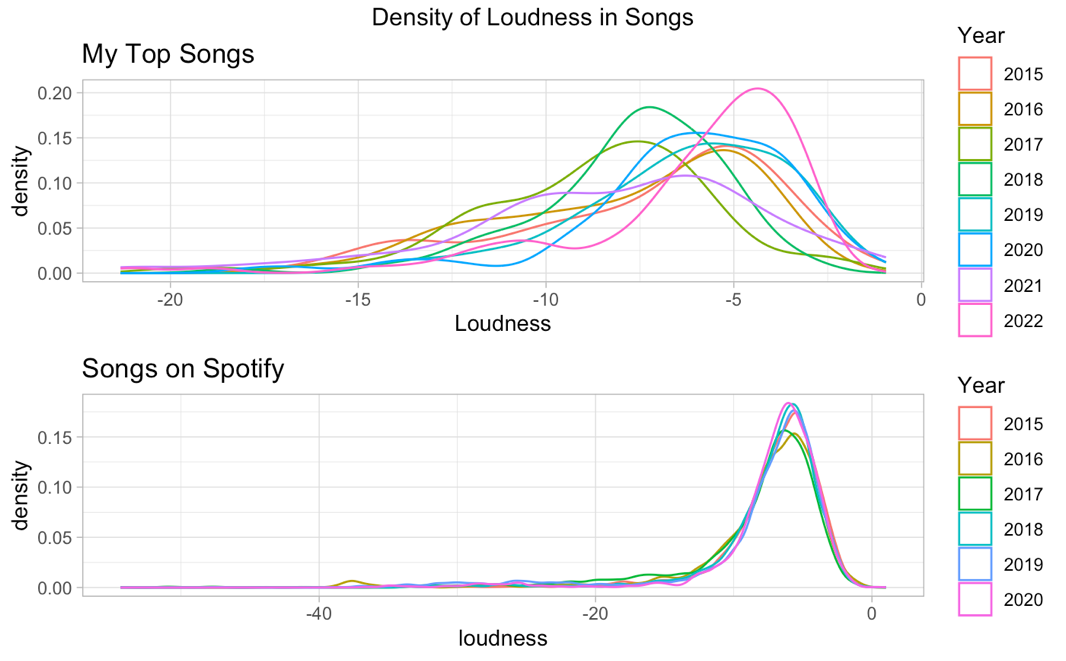


*Figure 3.9: Genres 2023*

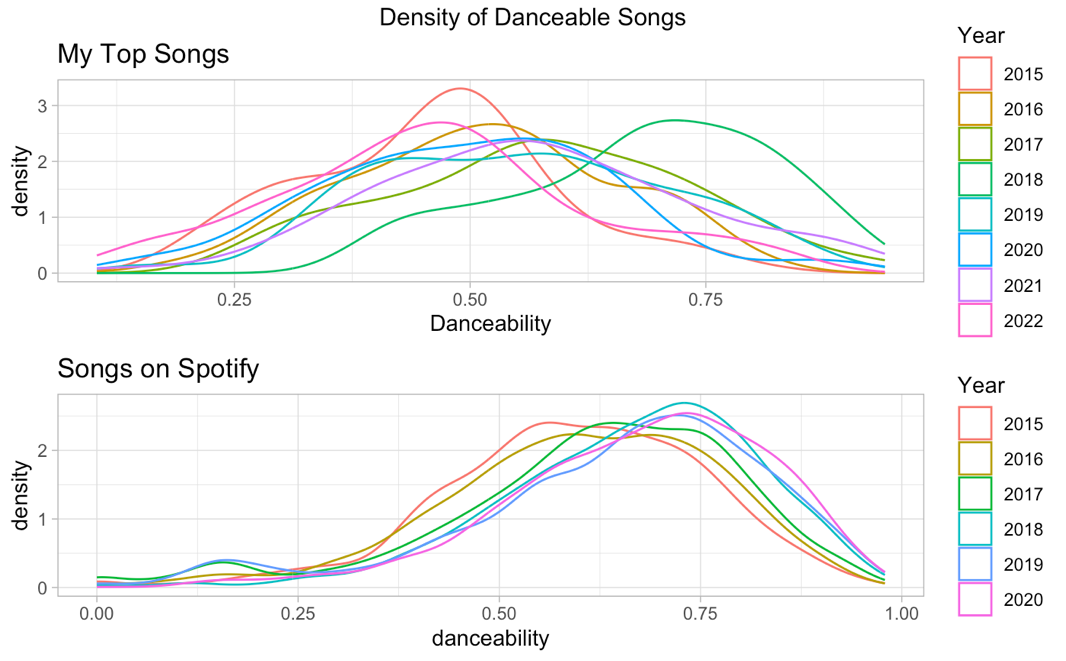


Appendix 4: Comparison with General Spotify Songs

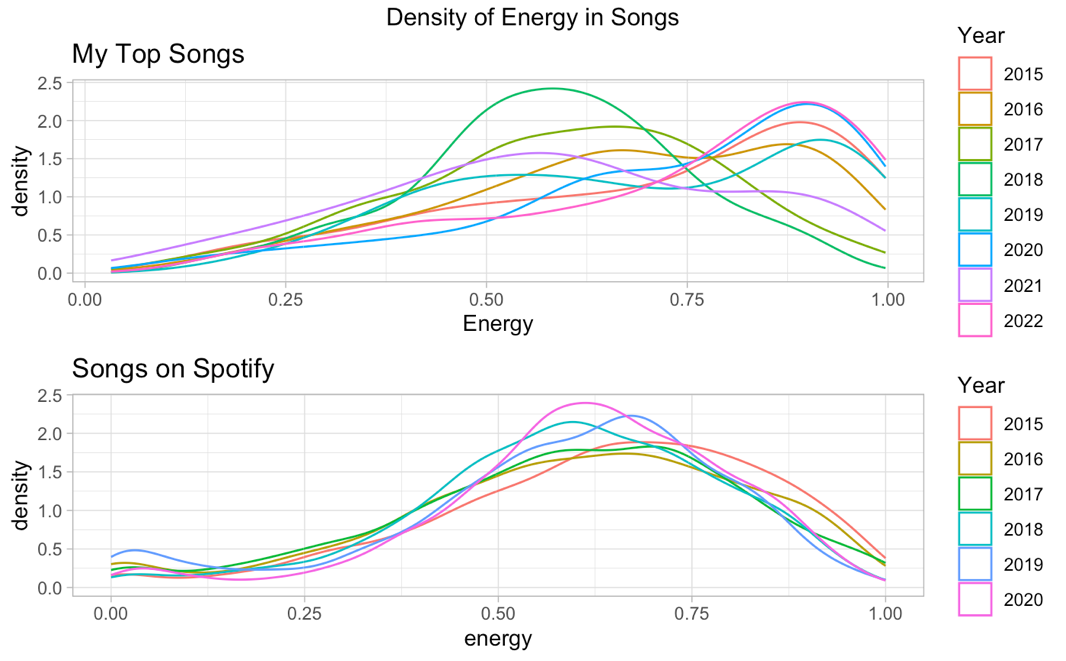
*Figure 4.1: Loudness of My Top Songs vs. Spotify Songs*



*Figure 4.2: Dance-ability of My Top Songs vs. Spotify Songs*



*Figure 4.3: Energy of My Top Songs vs. Spotify Songs*



*Figure 4.4: Popularity of My Top Songs vs. Spotify Songs*

