

A Data-driven Musical Analysis of Artist and Listener Activity

Alec Pugh

Jackson Reed

Mikaela Sherry

Yulim Kim

ABSTRACT

The ever-evolving music industry can be looked at through the lens of the artist and the listener. Preferences for music creation and listening can be observed on both ends through aspects such as top genres listened to on Billboard's charts or the most popular genre of music created by country. The rise of social media, digital streaming platforms, and music production technologies has a heavy influence on observed patterns. Within this paper, we analyze the trends from both the artist's and listener's points of view. Through the utilization of three datasets and data visualization through Altair, Tableau, and other methodologies, we can delve into the patterns and relationships between these two entities within music.

1 INTRODUCTION

The music industry has undergone a profound transformation in recent years, marked by significant shifts in both the production and consumption of music. Rapid advancements in producing technology, streaming platforms, and ever-evolving consumer preferences and production styles have spurred innovative approaches to the music industry for both creators and listeners. We hope to look into both sides of music: artists and trends of music production across the globe, and listeners and listening habits within the United States.

To analyze the artist's point of view, we utilize the MusicBrainz and 40 Years of Music Industry Sales datasets. The MusicBrainz database is an open-source database that serves as a comprehensive repository of music-related information. It is designed to provide metadata about music recordings, including details about artists, albums, tracks, etc., as well as the relationships between all of them. The MusicBrainz database relies on a community-driven approach, with contributors worldwide collaborating to continuously update and maintain the database. The 40 Years of Music Industry Sales dataset looks back at four decades of U.S. music industry sales data through the lens of technology and consumer behavior. We hope to explore music streaming patterns, especially with the rise of digital platforms for artists.

The listener point of view is analyzed utilizing Billboard charts data through the Top Spotify songs from 2010-2019 - BY YEAR dataset. This dataset holds the top songs, ranked by Billboard between 2010 and 2019. This can be interpreted as the most popular songs for listeners, based on sales, airplay, and streaming, of these years. The Billboard charts, unless specifically specified as a 'Global' chart, collect data from within the United States. Billboard's top songs are chosen through an algorithm that accounts for sales (digital and physical), radio airplay, and online streaming. Through an algorithm accounting for these aspects, the popularity of a song is calculated and then ranked.

While mainly looking at artists' and listeners' preferences regarding genre, we hope to find connections between what is produced and what is listened to. There are limitations in our analysis regarding scope, as we only have listening data that encompasses the United States in a smaller range of years, while artist data scopes to a larger scale. Regardless, through our data explorations and analysis, we hope to draw relationships that provide insight into how artists influence listeners and/or how listeners influence artists.

2 RELATED WORK

Several studies have already been done on the scope of artists' preferences, audience reception of music, and profitability in the music industry.

An investigation by [4] examined the US Billboard Hot 100 from 1960 to 2010, using music information retrieval and text-mining techniques to analyze approximately 17,000 recordings. This study unveiled quantitative trends throughout the period, showing the evolution of musical diversity over time. It challenged conventional theories of gradual cultural change, revealing distinct stylistic 'revolutions' in 1964, 1983, and 1991. This highlighted that though the most popular genre is consistently pop, it takes on different forms through time while holding the same categorization.

Addressing the shifting landscape of music production, [2] highlighted the changing role of recording studios in the rise of at-home studios. While these innovations have bolstered home music creation, the professional studio still plays a significant role, particularly in ensuring high-quality recordings. This is important, as we can analyze if there was a difference in genre output as recording became more accessible.

Meanwhile, [6] used neurological responses to identify hit songs, showcasing the potential of machine learning in predicting market outcomes. This approach demonstrated high accuracy in discerning hit music, indicating the ability of the brain to identify popular tunes. This shows there could be an algorithmic way to determine why pop has dominated the top charts.

The impact of digital transformation on consumer behavior and industry sustainability was explored by [7]. This study shows the importance of understanding diverse consumption methods and predicting emerging trends to navigate the transition to a digital environment successfully. Through our analysis, we found many transitions in how music was primarily listened to. This study will help prepare those in the music industry for how to handle the next transition.

[8] investigated the co-evolutionary relationship between streaming services and the live music industry, suggesting a dynamic that contributes to the sustainability of the music ecosystem. We found that streaming services have greatly damaged the financial potential that comes with releasing music; however, this study shows that live music helps fill the void.

In the realm of genre classification, [5] argued for and against the relevance of research in automatic genre classification, proposing strategies to overcome current limitations. This study is closely tied to our research, as it helps show us the limitations of our data through the problems in genre classification.

[1] examined artist-fan interactions on social media as predictors of chart success, showcasing the utility of such data in understanding public perception and guiding marketing strategies within the music industry. This study shows a possible error in our data: the popularity of a song may be due to far more than the traits of just the music.

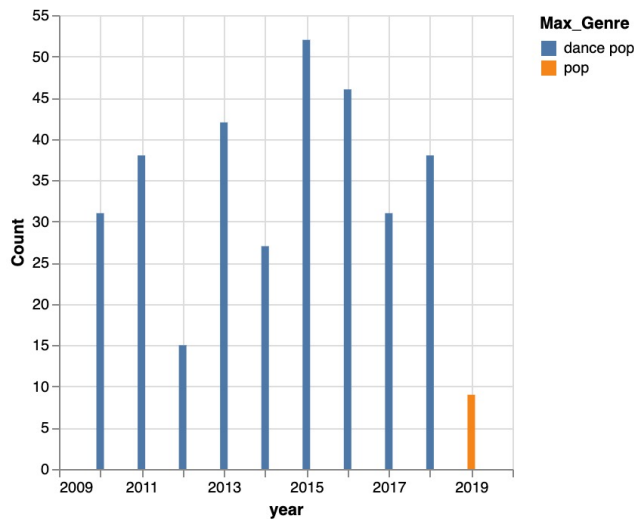


Figure 1: This visualization shows the count of the most popular genre by year from the *Top Spotify songs from 2010-2019 - BY YEAR* dataset. Each color corresponds to a genre that was found to be most popular.

Finally, [3] extended previous research on gender-related trends in music charts to include race-related analysis. This revealed patterns of chart success across different demographic groups. This study underscored the importance of considering identities and genre dynamics in understanding disparities within the music industry. This is always crucial for us to consider, as music has a history of disproportionately marginalizing certain groups.

3 PROCESSES

3.1 Listener Visualizations

By utilizing the *Top Spotify songs from 2010-2019 - BY YEAR* dataset, we were able to analyze the music preferences of listeners. The data is based on the top songs assessed by Billboard, listened to on the music streaming platform Spotify. Our goal was to discover any patterns, specifically regarding genre, that could be observed through the listening data. The dataset is structured so that each row correlates to one song within the top songs ranked by Billboard for each year. The order of the data correlates to their ranking every year. Therefore, we had individual song data, including title, artist, genre, bpm, and other musical statistics for each top Billboard song per year.

Utilizing Python and the visualization package Altair, we were able to analyze trends in listening patterns through Spotify. This allows us to gain insight into the habits and preferences of music listeners within the United States. We first looked through popular genres by year. Through aggregation by year, we were able to visualize the counts of the most genres per year with a bar chart, shown in Figure 1. The color of each bar is representative of the most popular genre for that year on the Billboard chart. When mousing over each bar, the top genre and count for that year appear in a tooltip, allowing for easy readability and comprehension for users.

Dance-pop was the most popular genre for all years within the data set except for 2019, where pop was the most popular. We wondered if there was more variability to popular genres if looking at the dataset while disregarding that genre. To obtain the correct data for this, we filtered the dataset to include all genres within

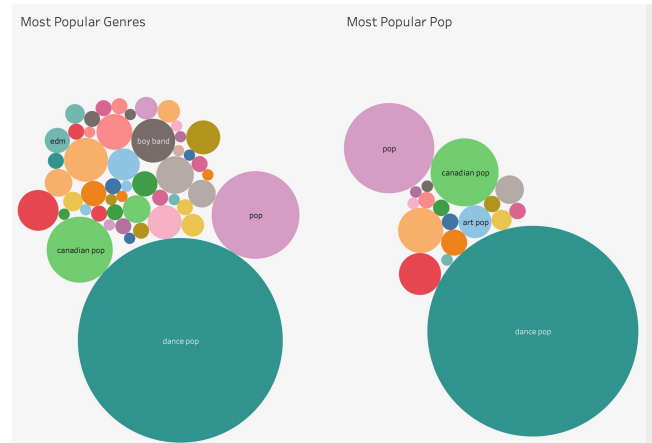


Figure 2: This visualization demonstrates the most popular genres by using their frequency to correspond to the size of the bubble. The right side of the visualization examines all music genres, whereas the right side of the visualization just focuses on comparing different pop sub-genres.

the dataset except for dance pop, shown in Figure 3. We then aggregated by year, similar to the previous visualization, and plotted our findings through a bar chart that represents top genres by year. The color of each bar is representative of a specific genre, and mousing over a bar will show the top genre and its count. The design elements were kept the same as the previous visualization, to draw connections and emphasize differences among the findings with and without the genre dance pop.

Continuing with the analysis of the *Top Spotify songs from 2010-2019 - BY YEAR* dataset, we utilized Tableau to continue our analysis with a WYSIWYG tool, shown in Figure 2. We decided to pursue this design tool option to see if other visualization methods encapsulated what the data showed and provided any different insights. Using the bubble chart gives a more intuitive idea of the size of each genre's popularity.

3.2 Artist Visualizations

To best tackle the artist's perspective of music within a global landscape, we first created an artifact that displays the genres that artists are making in each country. That is, if an artist is born or based in a specific country, what genre are they making? For user readability, we have decided to show the various genres from each country on a world map, making it so that you are able to mouse over each country to obtain the details of the top 10 genres from that country. The number next to the genre is the number of artists from that country that have been associated with that genre through MusicBrainz.

To obtain the data for this visualization, we have indexed the MusicBrainz database as it provides the ability to programmatically search for artists based on a specified area (in this case, a country). The first 15 pages of artists from each country were indexed and their genre information was pulled and then associated with their specified country code, which allowed us to then develop a visualization via Altair that displayed this association. However, it is important to note that because MusicBrainz contains community-sourced data (similar to Wikipedia, but strictly for music), there could be some artists with less data submitted than others. In addition, there are also countries with no available data on artists and their genres as well. These countries are displayed in gray on the visualization. The "null" entries that are present in

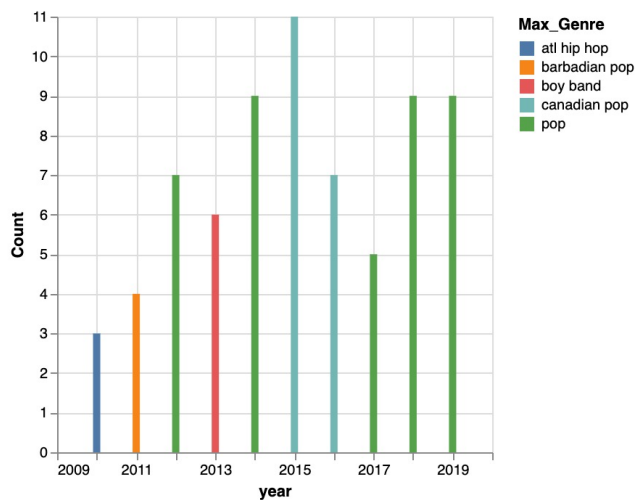


Figure 3: This visualization aims to look at the popular genres within the *Top Spotify songs from 2010-2019 - BY YEAR* dataset, after removing the dance pop genre. Through this, we can see which genres are second most popular, after the dance pop genre majority.

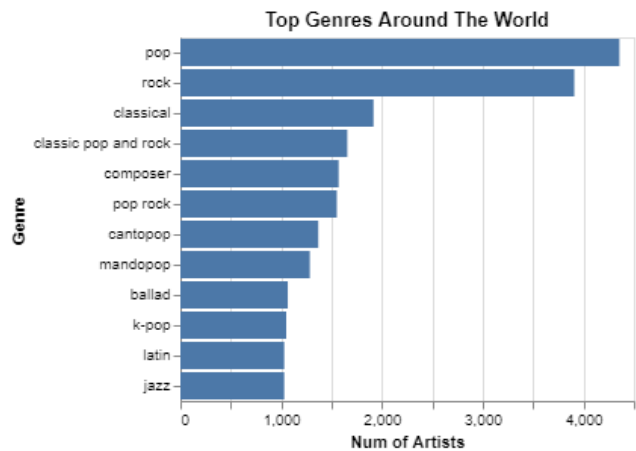


Figure 4: The top twelve most appearing artist-made genres out of all countries from the world map visualization.

various countries are simply because not all countries contain at least 10 different genres to display.

The second visualization, shown in Figure 4, was created by iterating through each of the countries' data obtained from the first visualization to find the genres most common worldwide. This information was displayed within a bar chart via Altair like the first visualization, with the x-axis composed of the number of artists associated with making the genre, and the y-axis containing the genre itself. This visualization was created to create a complement to the listening branch of our project, to help draw conclusions about a potential correlation in what artists are producing and what is most popular with the listeners. However, we also wanted these visualizations to provide a more nuanced perspective of music that goes beyond what is simply popular to the listener, to see if there may be genres that are more to make among artists rather than to listen to.

We were also interested in seeing the general financial trends of the music industry, and how this could correlate to the changes in

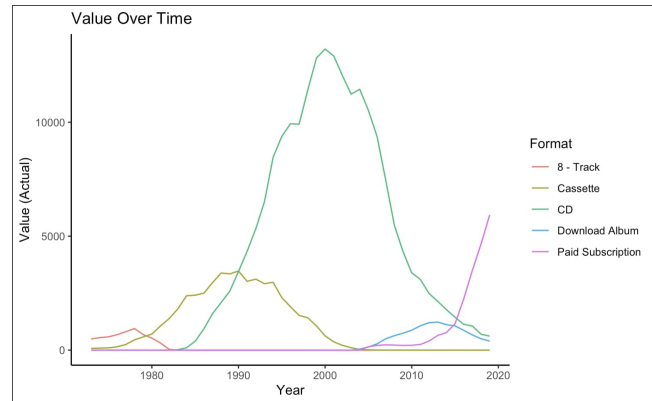


Figure 5: This visualization shows the change in value of different forms of music release types over time. We can see that CD sales were by far the most profitable medium, and that when one medium rises, the prior one falls.

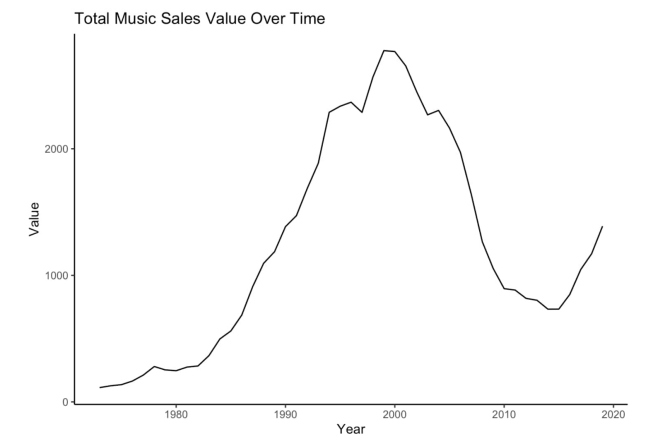


Figure 6: This visualization shows the change in value over time of the total music industry sales across all mediums. It is the summation of the values in Figure 5.

musical output and input over time. To investigate this, we found the dataset *40 Years of Music Industry Sales*. The dataset details the value of the sales of different mediums of music (ie. CDs, cassettes, subscriptions) over time. We created a value vs. time chart, separating the different mediums by color, which led us to the expected conclusion that as one medium died out, another took its place. This is shown in Figure 5 The most obvious example of this was when CDs replaced cassettes. From here, we decided it would be most valuable to graph the summation of the value of all mediums by time. This would give us a strong indication of the overall profitability of making music during any given year and allow us to dive into the relation of musical output diversity versus the profitability of music, shown in Figure 6. This Data tidying and the process of creating visualizations were all done in R.

4 FINDINGS AND ANALYSIS

4.1 Listener Analysis

After examining our resulting visualizations, it is apparent that pop music holds a significant sway amongst listeners. In the past decade, eight out of ten years have seen pop emerge as the top genre, as evidenced by 3. This dominance not only underscores pop's widespread appeal but also hints at its profound influence on

cultural trends and commercial success within the music industry. Furthermore, the prominence of pop music is vividly illustrated by the frequency and size of bubbles in 2. These visual representations not only showcase the genre's popularity but also highlight its presence across diverse demographic segments.

Also, the prevalence of streaming services like Spotify has altered how music is consumed and popularized. With algorithms driving recommendations based on user behavior, listeners are continually exposed to a curated selection of tracks tailored to their tastes. This personalized approach not only influences individual listening habits but also contributes to the collective rise of certain genres, such as pop, which often align with broader audience preferences and trends. As a result, we witness a positive feedback loop where popular music, particularly the pop genre, has increased visibility and engagement through streaming platforms. The positive feedback loop created by algorithms amplifies the exposure of hit songs and solidifies pop music's status as a dominant force in contemporary music culture.

4.2 Artist Analysis

To begin, we will analyze the findings from the world map genre visualization. The smaller countries seem to have more local genres that are almost unique to the specific country, such as MPB (música popular brasileira) in Brazil, mbalax in Senegal, or mandopop in China. On the contrary, other genres originated in a certain country but have spread to neighboring countries, such as raï starting in Algeria but also being found in Libya, for instance. On a continental scale, we notice trends of popularity in certain groups of genres present in certain regions of the world. That is, genres such as rock are universally loved in North America and parts of South America, while classical is more universally made in Europe. African countries, however, have very diverse musical interests from country to country. An explanation for this could be the massive ethnic diversity found in Africa, as each ethnic group may have its own language and culture even though they may live geographically very close to each other.

Through the top genres around the world bar chart, we can see that pop, rock, and classical are the three most common genres being made by artists around the world. There also are a few surprises, cantopop and mandopop for instance, however, this is to be expected as China contains one of the world's largest populations, thus the number of artists in these countries will be greater.

Moving to the format of music over time visualization, one of the main things that sticks out within the graph is the explosion of CDs during the 2000s, peaking in the year 2000. Before that, the largest medium was cassettes. Beginning in the late 2015s, as expected, the popularity of streaming services took off, and the peak for this medium is still yet to be reached. This makes sense as it coincides with the rise of streaming platforms in general, such as Netflix or Hulu for TV and movies. One analysis to be noted is that no medium seems to stay popular indefinitely, each medium (besides the latest one) has a start, a peak, and a decline. This is interesting as it suggests a continuous change of normality within listeners that artists must adapt to. Observing the Total Music Sales Over Time visual, we see that the popularity roughly coincides with the start of the cassette's popularity, and continues with the rise of CDs. The sharp decline in the graph can be attributed to the rise of streaming services, but there is a small rise happening at the end which may be attributed to the vinyl revival happening in recent years.

ACKNOWLEDGMENTS

The authors wish to thank the TAs and Professor Szafir for an interesting 590 class.

REFERENCES

- [1] J. Aum, J. Kim, and E. Park. Can we predict the billboard music chart winner? machine learning prediction based on twitter artist-fan interactions. *Behav. Inf. Technol.*, pp. 1–14, Feb. 2022.
- [2] M. Homer. Beyond the studio: The impact of home recording technologies on music creation and consumption. 2009.
- [3] M. Lafrance, C. Scheibling, L. Burns, and J. Durr. Race, gender, and the billboard top 40 charts between 1997 and 2007. *Pop. Music Soc.*, 41(5):522–538, Oct. 2018.
- [4] M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi. The evolution of popular music: USA 1960-2010. *R. Soc. Open Sci.*, 2(5):150081, May 2015.
- [5] C. McKay and I. Fujinaga. Musical genre classification: Is it worth pursuing and how can it be improved? Zenodo, 2006.
- [6] S. H. Merritt, K. Gaffuri, and P. J. Zak. Accurately predicting hit songs using neurophysiology and machine learning. *Front. Artif. Intell.*, 6:1154663, June 2023.
- [7] L. Molteni and A. Ordanini. Consumption patterns, digital technology and music downloading. *Long Range Plann.*, 36(4):389–406, Aug. 2003.
- [8] K. Naveed, C. Watanabe, and P. Neittaanmäki. Co-evolution between streaming and live music leads a way to the sustainable growth of music industry – lessons from the US experiences. *Technol. Soc.*, 50:1–19, Aug. 2017.