CPSC 4030

Professor Iuricich

Project Report

Michael Harris | Nicolas Mowll | Thomas Personett

Overview and Motivation

This project analyzes how technology advancement has affected music artist popularity over the past 80 years using Billboard Hot 100 chart data. The motivation is to understand how changes in technology like streaming platforms and social media have shaped artists' careers and popularity. The goals are to illustrate artist popularity fluctuations over time, demonstrate the connection between tech advances and artist fame, show how social media platforms influence popularity, explain how music mediums like records, CDs, and streaming impact artists, and visually showcase these trends in an intuitive, engaging way.

Related Work

This project expands on previous data visualizations and analyses focused on music trends and the Billboard charts. For example, data scientist Matt Daniels has created visualizations examining hip hop artists' vocabularies and pop music chords and lyrical complexity over time. Media outlets like Billboard and Rolling Stone frequently analyze chart data to showcase music industry trends. We draw inspiration from these works in aiming to uncover new insights related to how technology has transformed artist popularity using 80 years of chart data.

Questions

Initial questions included: How has artist popularity changed over time? How are tech advances connected to fame? How do platforms like social media shape popularity? As analysis progressed, new questions around album mediums, music connectivity, outlier years, and differences across genres emerged. Refining the objectives and questions worked to process digging into the data, brainstorm areas to explore, and determine what trends and stories would be most insightful to showcase given the dataset.

Data

The dataset referenced is a Full History of the BillBoard Hot 100 songs from 1958-2021. The Hot 100 acts as sort of a standard for how the U.S. measures song popularity, and has been a

popular form of musical judgment for eighty years. BillBoard's data gathering methods have changed over the years, but generally they combine several components that contribute to the overall chart (now, they combine radio plays and streaming listens). Credit to Dylan Burati for the dataset *Billboard Hot 100 - Full History* downloaded off of Kaggle.com.

Source: https://www.kaggle.com/datasets/dylanburati/billboard-hot-100-full-history

The data is separated into two joinable tables as follows:

Table: billboard-hot-100-credited-artists

Description: provides info on specific artists for each track

Items: Artists

Attributes:

- Song_id: FK - unique categorical identifier for each track
- Artist_id: PK - unique categorical identifier for each artist
- Artist: Categorical, credited artist's name
- Ordinal: Ordinal, index showing the position the artist was listed on the track

Table: billboard-hot-100-tracks

Items: Song Tracks

Attributes:

Description: provides info on every ranked track

- Id: PK - unique categorical identifier for each track
- Rank: Quantitative - Rank of the track on the Hot 100
- Date: Categorical - Date of the chart
- Artist: Categorical - Credited artist(s) on the track
- Song: Categorical - Song title
- Last_week: Quantitative - Ranking on chart for the previous week
- Weeks_on_chart: Quantitative - Total # of weeks on the chart
- Peak_rank: Quantitative - Highest rank achieved

When scraping the data, we wanted to be able to measure track popularity based on how high and how many times a track was ranked. The only information given is all one hundred ranked songs of each week, and so we derived a popularity score for each song depending on the ranking. We used a weighted formula to calculate this score. If a song was ranked #1, it would receive a popularity score of 10, slowly decreasing so that if a song was ranked #100, it's

popularity score is close to zero. This allowed us to sum up popularity scores of artists over a certain period of time to see their popularity, and calculate the aggregate popularity of one track over time based on each of its rankings. After graphing the top artists of certain time periods and top tracks according to our weighted popularity score, we cross referenced with BillBoard's top rankings and saw nearly identical results, justifying our derived column.

For each visualization, we generated a separate CSV file from our original data. The data cleanup for each visualization involved a different method. We derived columns such as total track popularity and total artist popularity depending on the visualization. We also derived an era column, which listed the decade for filtering. Also, we sorted much of the data to pull tracks from the top 12 artists of each decade only. To clean up the data, we used pandas, a python library.

Exploratory Data Analysis

Initially, we prototyped our visualization on Tableau. We looked at trends of music popularity over time, and music popularity among the top artists over time. Looking into these prototypes (found in the repository), we noticed that there was a trend of increasing popularity among the top artists over time. Why were the top artists of the 1960's less popular than the top artists of the 2010's during their respective times? We know it is not due directly to the number of listens/streams because the data is based on a fixed number ranking system for each week. That is, the same amount of popularity points are available for grabs during each week. This means that the top artists of more recent decades have much more outlying popularity than older decades. Showing this trend became our project objective.

These insights informed our project design and implementation. We wanted to showcase this uptrend in popularity over history in all three of our visualizations, also wanting to allow users to drill down to each decade and view specific tracks by the top artists. There are a few notable outliers, including a few certain tracks and the Beatles in the year 1964 that users can drill down to examine more closely.

Design Evolution

We considered several different visualizations during implementation and our process of working was iterative.

Initially, our first visualization included a categorized dot plot of the top twelve artists of each decade with all their ranked songs along a horizontal popularity axis (nearly identical to when a user drills down now). We iteratively changed this visualization to show an aggregate view of all songs separated by decade, colored accordingly. This allowed us to better showcase our objective trend of an increase in popularity over time. When the user filters by decade, the original design showing artist names and their ranked tracks is displayed.

Our second visualization began as a line graph in Tableau. The y axis showed artist popularity and the x axis showed each year. Each line represented an artist's popularity over time. The line graph was very crowded and unreadable by a user, so we iteratively improved it. We first sorted the data to only show popularity lines for the top 12 artists of each decade. It was still too crowded, so we decided to make it a scatter plot, with each dot being the artist's peak popularity. This more simplified version showcased our objective well and was easy to read, still supplying information about the popularity of each artist.

Our third visualization initially began as a density over time graph to show the widening of popularity distribution over time. However, this virtually showed the same trend as the scatter plot, so we decided to scrap this visualization. For continuity and ease of understanding, we decided to color match all data points by era for all three visualizations. For the third, we decided to show a basic bubble graph of the top 100-200 songs of All Time or by Decade selected and color them accordingly. The radius of each bubble is dependent on the track's popularity.

These three visualizations together help to show our objective trend and allow the user to easily manipulate the drilling down to view more specific details. The user can click on any data mark and will view a more in depth visual dashboard for that specific decade.

The design principles learned from this course used in our design were that of marks, channels, and perception. With our group already having a basic understanding of data and data visualization, we worked together to talk about effective ways of showing our objective. We know that users perceive text poorly and perceive area and point separation rather effectively. Therefore, we decided to use dots and radius to show volume of tracks and their popularity, allowing for users to understand the relationship between artist popularity. Also, we made sure to consistently color each dot by decade so that the user has an intuitive understanding of the

dashboard. The color selection involved making sure each hue was distinguishable and appealing.

Also, the consistent use of a circle mark to depict each item in our data allows for the user to naturally understand each visualization and analyze the data without struggle. We made sure not to have too many channels so that the user would be unconfused, utilizing mainly popularity and time, with the exception of displaying a tooltip and artist names if the user desires to dig into more detail.

Implementation

Visualization 1: Distribution of Top Artists' Song Popularity Throughout Time

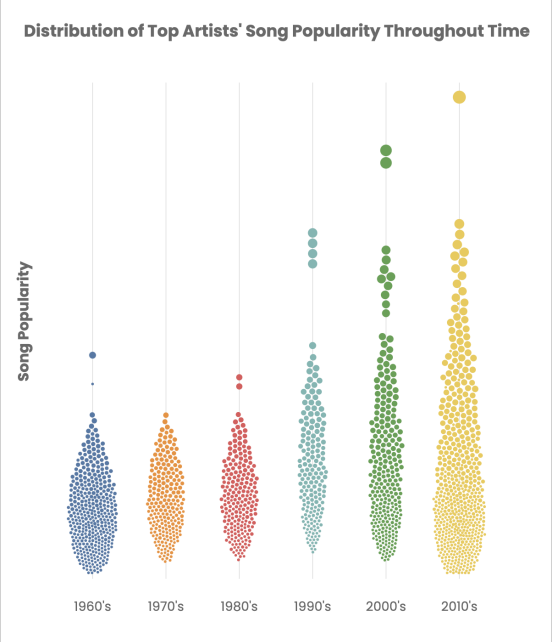

Figure 1: Original graph for Visualization 1

The purpose of this visualization is to display the popularity distribution of top songs of each decade's top 12 artists. Each decade is signified by their own specific color, and they each get vertical lines on the horizontal axis for their distributions. Songs with higher popularity scores will be plotted higher on the vertical line of their respective decade, and less popular songs will be on the bottom of the vertical line. The radius of each dot also scales to the popularity score of the song, which can be seen very clearly looking at the 2010s column from the bottom to the top. This helps highlight the popularity of each individual song.

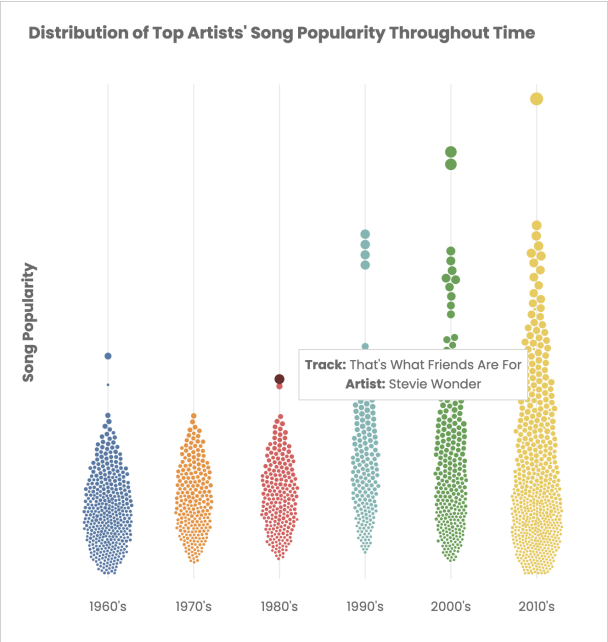

Figure 2: Hover interaction on original graph

The original graph has a form of interaction which involves hovering over a dot. Once the user does this, the dot grows in size and darkens in color to highlight which dot is being selected, which especially helps with dots lower on the vertical lines. The hover tooltip lists what track the specific dot is and the artist of that song. This hover function works for all columns, allowing the user to discover and interact with the songs throughout history.

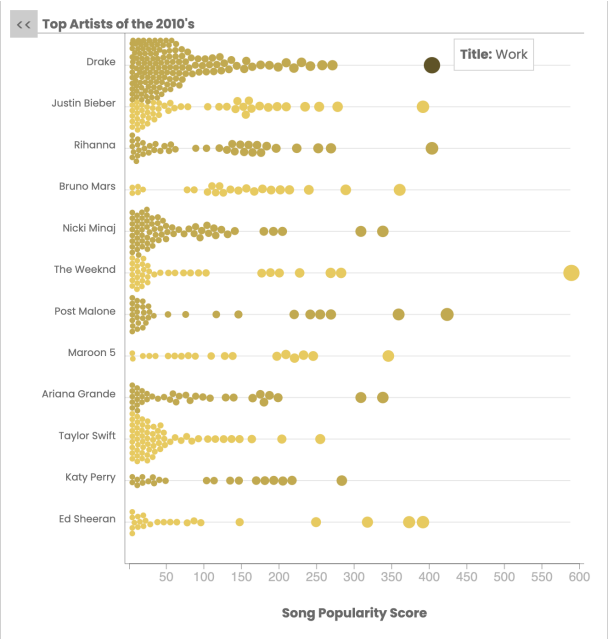

Figure 3: Post-click individual decade graph

If the user clicks on a dot of the original graph instead of just hovering, a more in-depth graph will be displayed. The user can click any dot on the original graph and it will display that dot's decade data. This display shows the top 12 artists of the clicked dot's decade, and all of the artist's individual songs during that span. Each artist gets a horizontal line for their song dots, and songs with higher popularity will be plotted closer to the right end of the line. The dots also get scaled in radius size according to their popularity, so more popular songs will have bigger dots. The lines alternate from regular color to darker color to display the individual artist's dots in a more effective way. As shown in the top right corner of Figure 3, the decade graph also has the original hover tooltip implemented, which displays the song of the selected dot. Similarly, the dot gets darker and grows in size when the user hovers over it, making sure the controls are uniform across different displays. The user can click the back button in the top left corner to return to the original, multi-decade graph.

Visualization 2: Top Artists' Peak Popularity Throughout Time

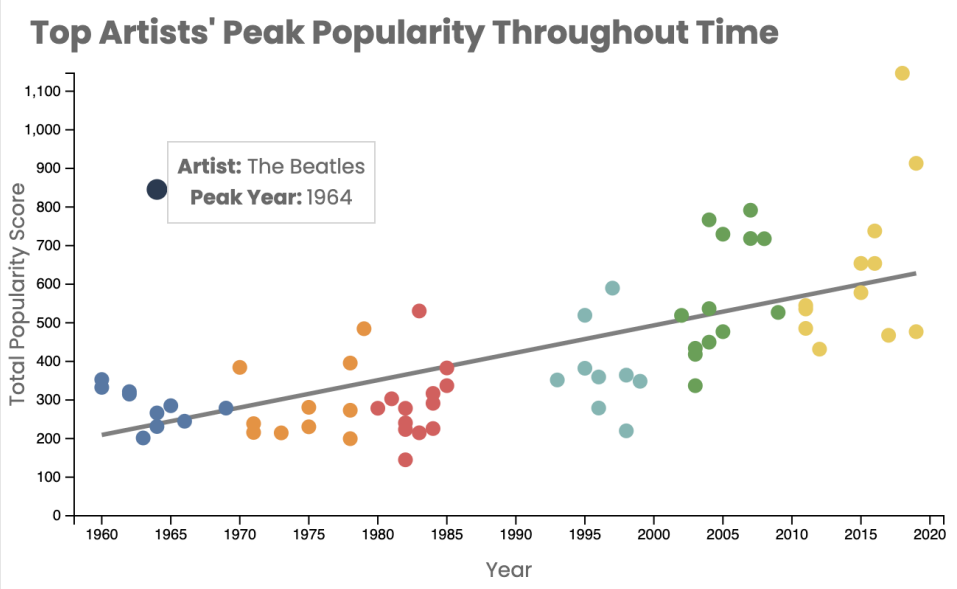

Figure 4: Scatter Plot for Visualization 2

The purpose of this visualization is to display the trend of top artist popularity over the past 6 decades. Each dot represents one of the top artists of each decade. The dots are categorized by the same colors as visualization 1 for higher effectiveness. The X-axis represents the year of the artist's peak popularity. The Y-axis represents the actual value of the artist's peak popularity. There is also a regression line displayed to show the trend in a more effective manner. In terms of interaction, this graph also has the same hover tooltip implemented in visualization 1, except it lists the artist name and peak year of the specific dot instead. This allows the user to explore

which dot represents certain artists and how the peak popularities have changed over time. When a user clicks on a specific dot, interactions will occur in the other two visualizations while this scatter plot stays the same for constant viewing. This cross-visualization interaction will be displayed later on.

Visualization 3: Top 200 Tracks by All-Time Popularity

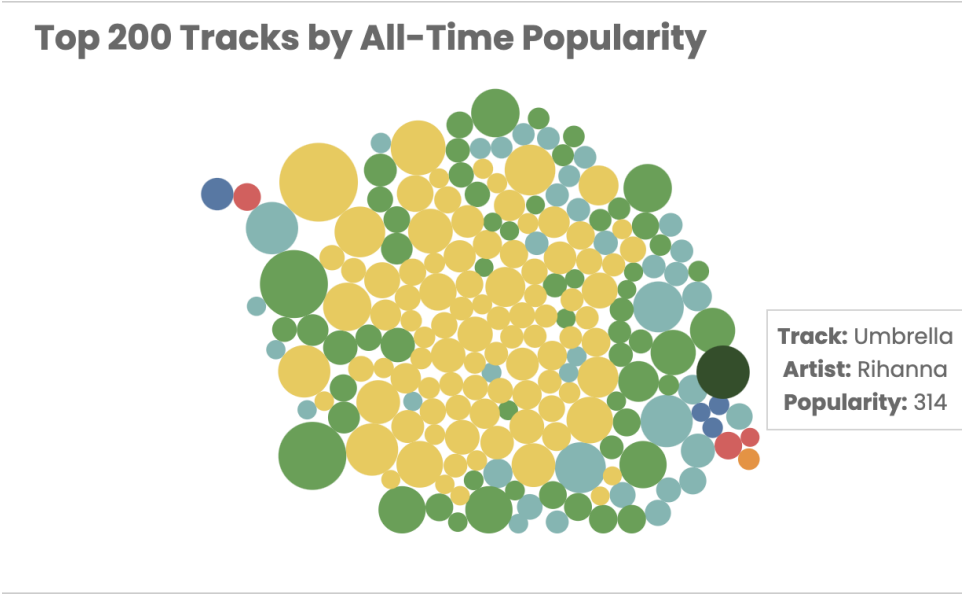

Figure 5: Bubble chart for Visualization 3

The purpose of this visualization is to display the top 200 songs of all time from the top artists of each decade. The color scale is the same as the previous two visualizations, where the colors of the dots match the decade that the song was released. The radius of the dots are scaled to the popularity value of the song. This visual also utilizes the hover tooltip just like the other two visualizations, which lists the track name, artist, and song popularity. The bubble chart displays all 6 decades/colors, so that they are all available for user interaction. While this visualization is much simpler, the key idea in including it is to provide an idea of decade distribution among the top tracks. For the aggregate data, the user can see that top tracks of all time were mainly popular in the 2000's and 2010's.

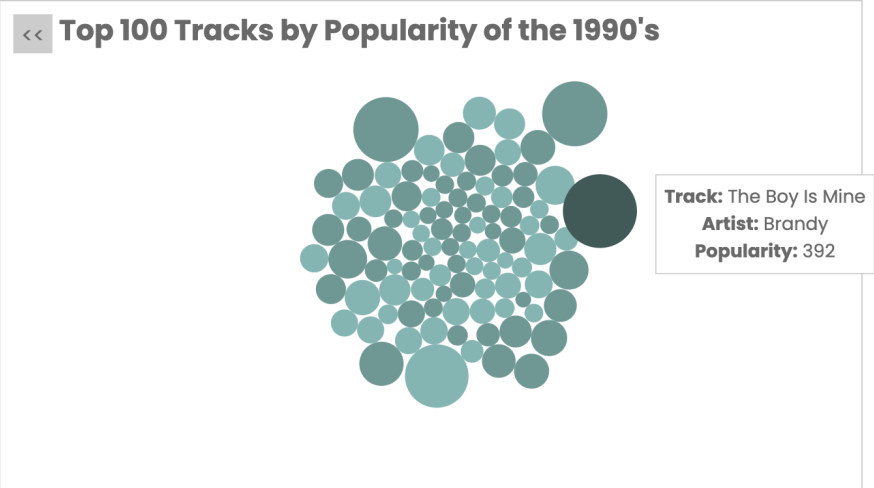

Figure 6: Post-Click Decade Bubble chart for Visualization 3

If the user clicks on a dot in the original bubble chart, a more in-depth chart will appear, focusing on the decade of the dot that was selected. The graph transforms from the top 200 songs of all time to the top 100 songs of the selected dots decade. The radius of the dots are scaled to the popularity of the track, and the colors alternate between regular and dark to allow for easier interpretation and recognition. This decade bubble chart also has the aforementioned hover tooltip, which similarly lists the track name, artist, and popularity score. The user is able to click any dot in the original bubble chart, and the decade bubble chart will appear for the decade of the selected dot. If the user wants to return to the original chart, they are able to select the back button in the top left corner of the visualization to revert back.

Dashboard and Cross-visualization Interaction

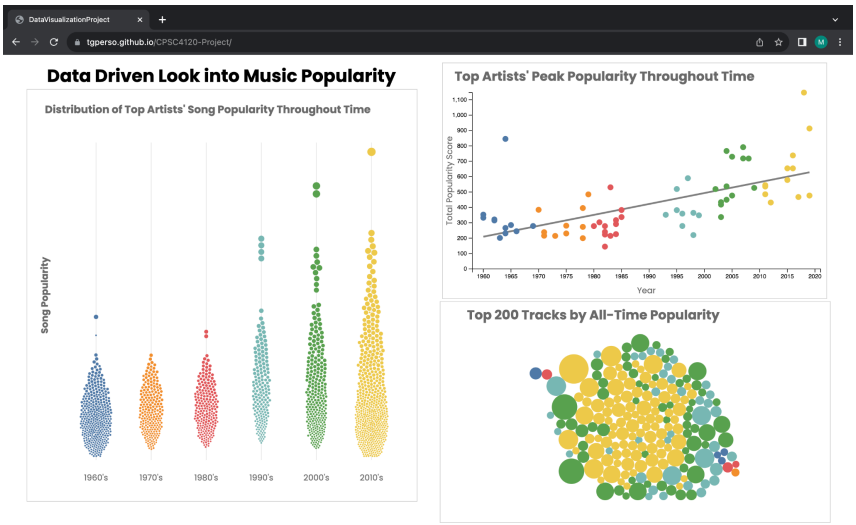

Figure 7: Complete view of visualization dashboard

For a more connected view, we combined the three visualizations analyzed above onto one visualization dashboard. Once the user goes to the website, all three visuals will be displayed on the page without having to scroll, allowing the user to focus on the actual charts instead of how the website functions. As seen in Figure 7, the colors are uniform between all visualizations for cohesive understanding, and each visualization represents the data in unique ways. This configuration also encourages cross-visualization interaction and animation.

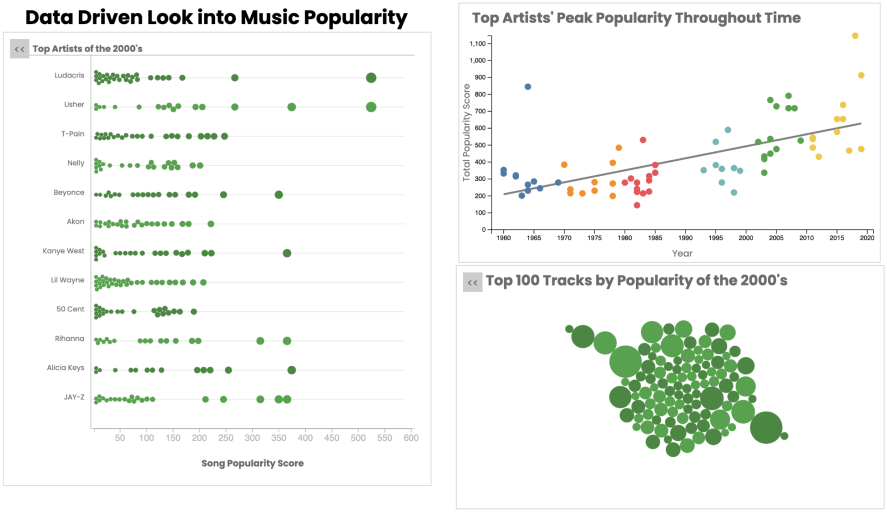

Figure 8: Post-click dashboard view

If the user clicks on any dot in one of the three visualizations, the dashboard will transform, displaying the in-depth decade data of the selected dot. This allows for a clear and user-friendly experience, allowing the user to focus on the trends and data instead of functionality. The combination of these decade-specific visualizations allows the user to dive deeper into the individual sectors of the dashboard. If the user wishes to return to the original dashboard, they can select the back arrow in either of the two visualizations to revert back.

Evaluation - Mikey

After implementing the visualizations and analyzing the results, there are many interesting discoveries and findings that wouldn't be easy to observe with only the large amount of data from the data set. As mentioned before, our motivation is to understand how changes in technology like streaming platforms and social media have shaped artists' careers and popularity, and our visuals allow for this analysis. One of the main questions that can be answered is how artist popularity changed over time. Using the visuals, a strong positive trend can be observed over the course of six decades in terms of artist popularity. The trend line in the scatterplot, along with the trend of dots getting higher on the vertical axis seen in the distribution

chart help prove this point. We also learned that this technology advancement heavily impacted individual tracks. Looking at the original bubble chart, we can see that it is dominated by yellow and green dots, indicating the popularity of the more recent songs. We can also see more tracks higher on the vertical axis for the later decades in the distribution chart. As time goes on, it seems that the advancement in technology has heavily influenced the popularity of specific artists and songs since technology allows for a major access of availability. The creation of social media allows artists to garner so much popularity from so many different people, leading to the saturation of the music scene. Instead of a higher number of somewhat equal artists during the 60s, there are now a handful of artists with outstanding amounts of popularity. The creation of different music platforms, like Spotify and Apple Music, makes accessing tracks and artists easier than ever before, leading to specific songs being played an incredible amount of time. Obsolete music mediums in the earlier decades, like records and CDs, were hard to acquire and expensive to maintain, limiting the number of people who could listen to music. Music is an extremely interesting factor to analyze and allows us to discover how and why trends happened the way they did. These historical trends allow us to make assumptions about what the future will hold in terms of music and artist popularity, which could reach unimaginable heights.

In conclusion, our visualization works very well and is perfect for allowing the user to discover how music has changed over time. It is very straightforward and visually appealing, and the interactions are informative and easy to understand. The same color scale being used for all visualizations allows for easy understanding and analysis. The transitions and animations are smooth and clear, and returning to the original graphs is easy. In terms of improvements, the implementation of a click interaction in the scatter plot would be first. For example, zooming into a specific decade when a dot is clicked and displaying more artists from the decade to show a more fine-tuned decade trend. We can also expand our datasets to include more artists and songs from the original set to dive deeper into the music history. For example, the addition of more song bubbles would allow for more song popularity information to be displayed. Finally, since our data is so versatile, it would allow us to make more complex visualizations and dashboard given the time and opportunity, like focusing on individual genres. Overall, we are proud of our effort and result and are excited to share this creation with others.