**UPDATE: Midterm Project Proposal**

*Notes describing changes made to the plan during implementation are made in red.*

**Name**: Ana Boccanfuso

**Project Title**: Visually Analyzing Music Trends from Spotify Data

1. **Introduction/Background**

Music plays a significant role in all of our lives: connecting us with others, allowing us to relate to one another, and eliciting emotions out of us. And yet, music is also something that makes us all unique. As each person has a music taste that is distinct like a fingerprint, music streaming platforms have implemented data analysis to try to pinpoint an individual’s music taste and predict songs they may enjoy. These algorithms are particularly useful in discovering new music, and we can learn more about our personal music taste by analyzing what we listen to and how often.

There is so much to explore within music. Some of these include variations within specific genres, variations within an artist’s discography, attributes of popular songs, and the evolution of music trends over time. Exploring these categories will allow for a better understanding of an “audio identity” for specific music taste, allowing listeners to have a better relationship with their music library.

1. **Audience**

Everyone listens to music and has a specific audio identity. Because of the breadth of this topic, the findings of the analysis will be useful to all people, specifically those who are interested in the evolution of music and their personal music taste.

1. **Dataset(s)**

The data that will be used in this project will be gathered using the Spotify API (<https://developer.spotify.com/documentation/web-api/reference/#/>) and two R packages, spotifyr and billboard. To analyze music trends over the years, the billboard r package will be helpful, as it provides the top 100 Billboard hits in the United States between 1960 and 2016. This package contains several data frames. One data frame (lyrics) has 5701 rows and 4 variables: title, artist, year, and lyrics. Another data frame (spotify\_track\_data) has 5497 rows and 23 variables:

|  |  |
| --- | --- |
| * year, * artist\_name, * artist\_id, * explicit, * track\_name, * track\_id, * danceability, * energy, * key, * loudness, * mode, * speechiness, | * acousticness, * instrumentalness, * liveness, * valence, * tempo, * type, * uri, * track\_href, * analysis\_url, * duration\_ms, and * time\_signature. |

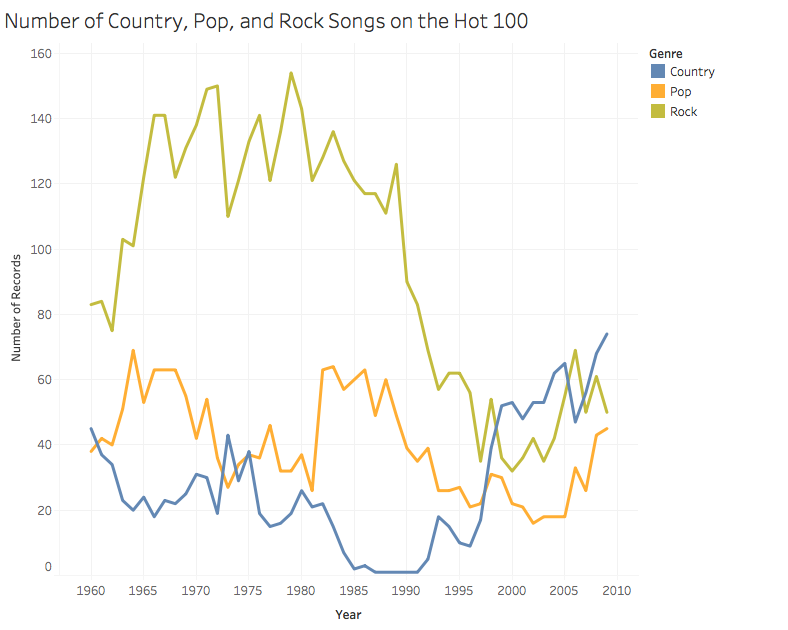
The most useful of these data frames is wiki\_hot\_100, which contains the Billboards Hot 100 songs, and this data frame has 5701 rows and 4 variables: rank, title, artist, and year.

Additionally, I will be using my own music data from the Spotify API to analyze my personal music taste. This dataset’s size is unknown at the time of writing this proposal.

Although the data quality should be cleaned considering it is in a package, some songs may be duplicate if, for example, a single was released earlier than an album, so there is a single version of a song and an album version of a song, even if they are identical. Other than this, the data is assumed to be cleaned and categorized correctly.

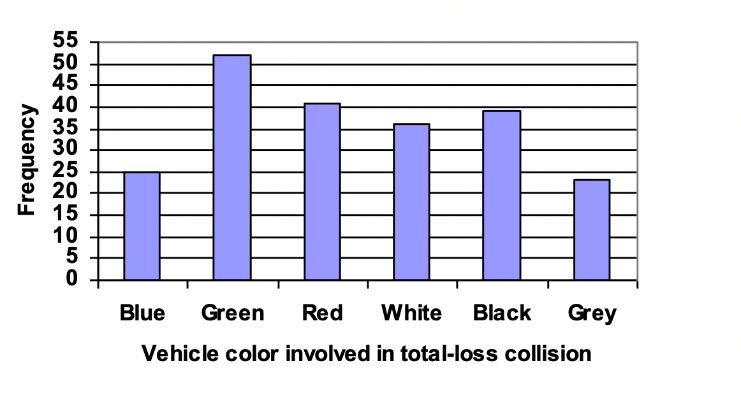
1. **Proposed Visualization**

I intend to create several data visualizations to demonstrate music trends over time and my particular music taste. The first visualization will show the annual change in average pop music elements, such as danceability, energy, valence, speechiness, acousticness, and liveness. Additionally, I would like to show the annual change in explicitness, to evaluate how songs have changed in this aspect over time. These visualizations will be line graphs to show change over time and may include some interactivity that shows specific point values when they are hovered over (for example, the graphic would show the value of explicitness in 1993). [These visualizations were created and included in the final proposal. However, due to the less appealing formatting of the embedded Tableau, images were embedded instead. Therefore, the Tableau images on the site do not include interactivity. To view interactivity with these line graphs, visit the site listed at the bottom of the final webpage.] If these elements are combined into one line graph, then there would be interactivity for the user to select which music elements to show (danceability, explicitness, etc.). An example is shown below. [Elements were separated into 3 graphs: danceability/valence/energy, acousticness/speechiness/liveness, and explicitness. Because these images are not interactive on the website, I needed to make the visualizations easy to view and understand without filtering the different elements. Additionally, these groups make sense because they are closer in value to each other, and therefore are close in the graph.]



Source: https://ibruins.weebly.com/visualizations.html

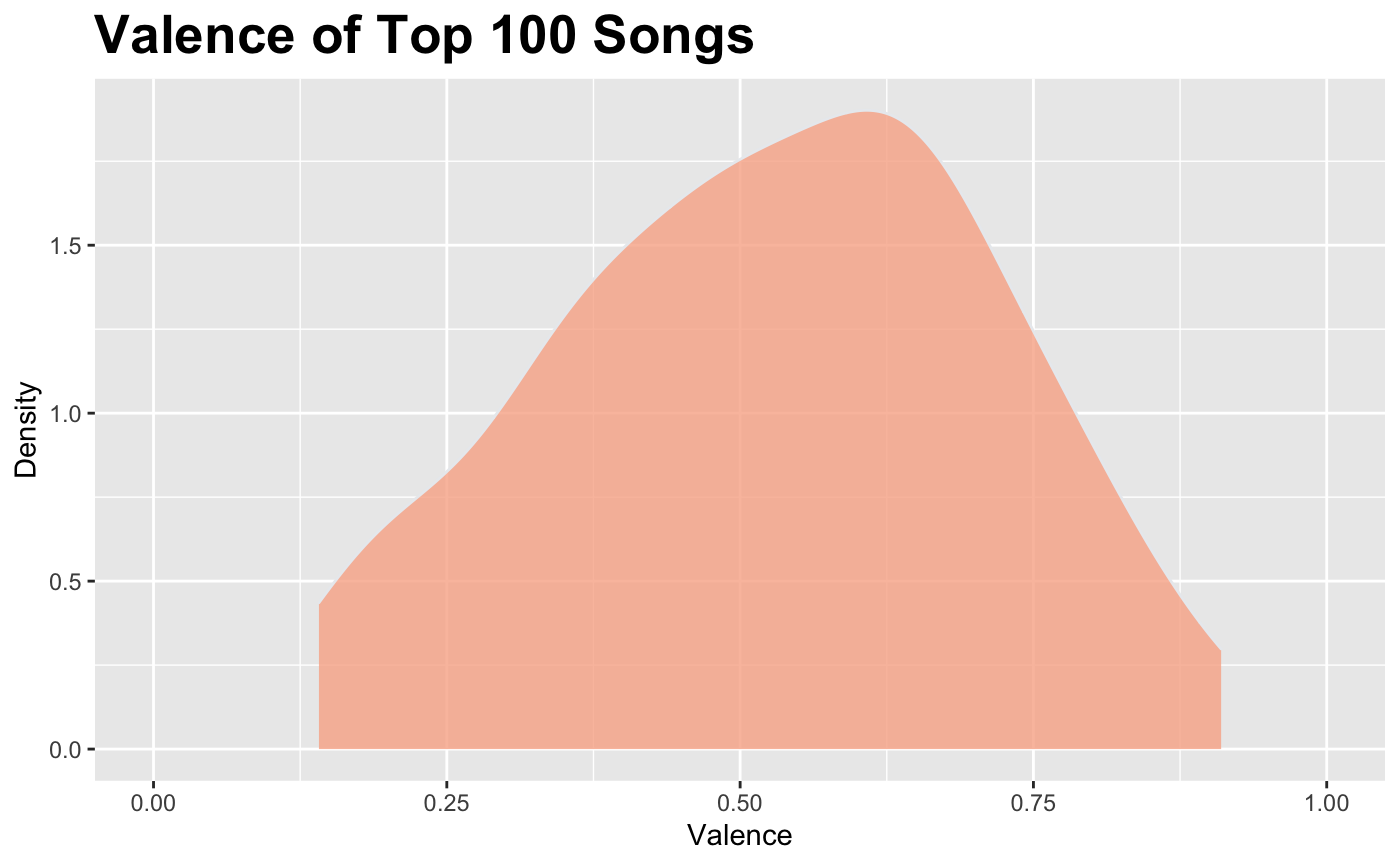
Another type of visualization that I would like to show is word clouds of the most commonly used words in lyrics between certain years. This way, the themes of popular music can be evaluated over time if songs are grouped by decade, for example. From this visualization, the top 10 words can then be graphed in a bar chart to show frequency of the most popular words. Example images are shown below. [These types of graphs were omitted. Although this would have been a fun activity, the data cleaning for the lyrics R package was too messy. It was lyrics of any sort and many looked like this: “[INTRO] [Mary J. Blidge] ahh ohh” (or something to that nature). After attempting to clean the data multiple times, I decided it would be best to omit this section and focus my efforts on other visualizations.]

Source: https://resilienteducator.com/classroom-resources/word-clouds-writing/ (left)

https://mathbooks.unl.edu/Contemporary/sec-2-1-graphically.html (right)

The next part of my music trend analysis will involve a case study of Taylor Swift’s discography, comparing her tracks’ attributes and the evolution of her music. Therefore, the next visualizations will be density plots that show the valence and danceability of Taylor’s different albums. [I decided to do several valence plots with more music elements, mostly because they were fun to make and see. I think these visualizations are really informative and useful because you can select different albums to view.] An example is shown below.

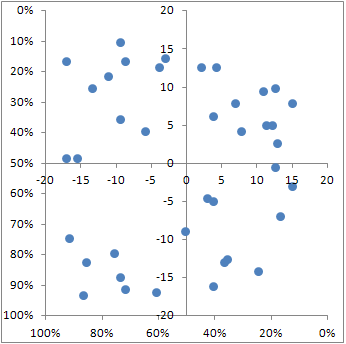


Source: https://kevin-liman.medium.com/what-makes-a-k-pop-song-popular-worldwide-d4251ef308be

For my final visualizations, I will create a polar scatter plot to compare my top 50 songs to the top 50 songs in the United States in my favorite three genres. In this plot, there will be poles with energy, valence, speechiness, danceability, acoustic, and liveness. Layers will be able to be toggled by clicking in the legend.

Source: https://www.pythonprogramming.in/plot-polar-graph-in-matplotlib.html

My final visualization will be a scatterplot that plots energy against valence, representing different moods of my music taste. The dots will be color-coded by genre, and the values of each point will be visible with a hover over.



Source: https://teylyn.com/articles/xy-scatter-chart-with-quadrants/

For both of my planned personal data visualizations, I omitted the original idea. I spent a large amount of my time trying to figure out how to create and format this chart with my personal data, but I couldn’t get it to look like how I wanted it to. Instead, I created a dashboard with four graphs depicting artist, how much time I spent listening to music over the year, and how many distinct songs I listen to per artist. This dashboard tells a better story of my personal music taste and how it is distributed across the year and different artists.

1. **Plan**

This project will utilize R and Tableau. R will be used for data collection and some visualizations, and Tableau will be used for other visualizations such as the word cloud. Visualizations and their corresponding analysis will be delivered via a website using HTML and CSS, and interactivity will be embedded in the website.

This project will evaluate music trends in general, how music trends have changed over time, and the attributes of popular music. These topics will be addressed no matter what, since they are pretty straightforward, and the data is readily available. Additionally, I will evaluate a particular artist’s discography to perform analysis on one particular person’s art as a whole, specifically what attributes their music holds and how it has changed over time and from album-to-album. Finally, I will visualize the attributes of music based on their genre to show the disparity (or similarity) between genres. Instead of genres, I mostly focused on artists.

A more ambitious, ideal plan involves analyzing my own music tastes against the general trends in the United States. Since I have not received this data yet, it is unknown when I will receive this data and if I will have time to make quality visualizations. I would also like to perform an analysis on Grammy award winners vs. their nominees. For example, what qualities do winners of Album of the Year have compared to their losing nominees? Is there any trend or more favorable attributes of winners of Grammy awards? These are all questions that I would like to visualize if time permits. Time did not permit for Grammy analysis.