Project Process Book

https://www.kaggle.com/datasets/thedevastator/spotify-tracks-genre-dataset
https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-600k-tracks

Your process book should include the following topics. Depending on your project type the amount of discussion you devote to each of them will vary:

- Overview and Motivation: Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.
- Related Work: Anything that inspired you, such as a paper, a website, visualizations we discussed in class, etc.
- Questions: What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?
- Data: Source, scraping method, cleanup, etc.
- Exploratory Data Analysis: What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?
- Design Evolution: What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course. Did you deviate from your proposal?
- Implementation: Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.
- Evaluation: What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

Overview and Motivation

Over the past century, music has transformed dramatically in response to technological innovation, cultural movements, and evolving listener preferences. Modern streaming platforms now capture both large-scale listening behaviors and detailed audio analyses, making it possible to quantitatively study what makes certain songs popular. This project aims to investigate how the musical characteristics of popular tracks have changed from 1921 to 2020 by analyzing a comprehensive Spotify dataset containing over 446,000 songs. We focus on two major categories of musical attributes: technical features, such as tempo, loudness, duration, and key. And perceptive features, Spotify's machine-generated measures of qualities like danceability, energy, and valence. By examining how these characteristics correlate with popularity and how their distributions shift over time and across genres, our goal is to understand which attributes define popular music in different eras.

Related Work

Our inspiration for this project grew out of our own music tastes and the annual excitement around Spotify Wrapped. Each year, Spotify Wrapped summarizes listeners' habits

by highlighting how their preferences shift across genres, moods, and styles throughout the year. It's a feature people look forward to and often share on social media, because it offers a personalized snapshot of their musical identity. We wanted to extend this idea beyond the individual level. Rather than examining how one person's preferences change over a single year, we set out to explore how society's musical tastes have evolved over the past century. By scaling this kind of analysis up to decades of music and hundreds of thousands of tracks, our project aims to uncover long-term trends in what listeners find popular and how the defining characteristics of music have changed over time.

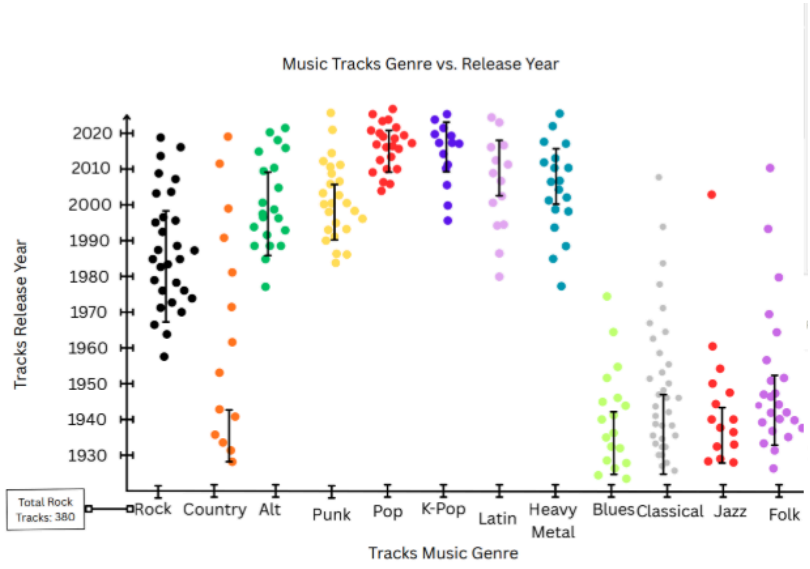
Questions

Data

Our analysis is based on the Spotify Tracks Dataset (1921–2020) collected by Yamac Eren Ay using the Spotify Web API and publicly available on Kaggle. The dataset contains approximately 446,474 tracks from over 114,000 artists, along with both technical audio features and perceptive attributes generated by Spotify's machine-learning models. Because the data was gathered through the API rather than scraped manually, it arrives in a structured format with consistent attribute definitions. However, several cleanup steps were necessary before analysis. Our main dataset contained around 500,000 tracks. However, it did not include the attributes for genre of music. So we had to combine it with another dataset which only contained around 100,000 of the tracks. We combined the two datasets using track_id and cleaned the data removing any null values for genre. In our graphs we analyzed different tracks by their release year so we had to extract the year from release_date.

Design Evolution

For our initial design of our first graph we wanted to have one major graph that showed how different genre popularity changed over the decades. So our initial prototype was this:

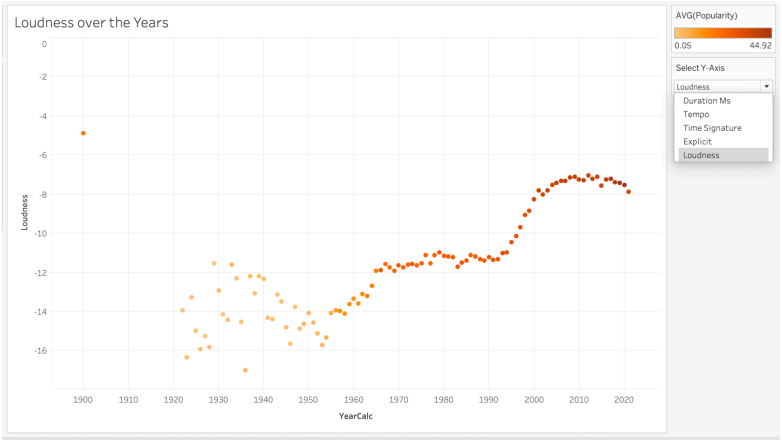


However, as we started to clean the data and work through the rest of our graphs we realized it would need to be altered. While this graph does show some insights into the question of how

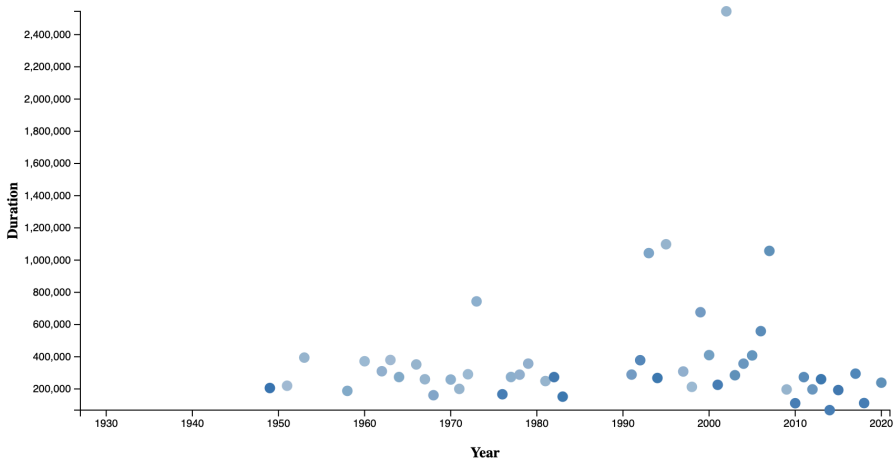
music taste changed over time, it could be better. We had needed the graphs to be interactive so we decided to make the genre a selective option on our first graph, and then the others reflect that. For the first graph, the design is nearly identical to the one shown above. However, after selecting a genre it would then focus on that one. It would still be a strip plot, but instead the release year would be on the x-axis, and each point an individual track.

For the next two visualizations, we focused on the questions “What are the technical and perceptive features of popular music?” Naturally this question begets the next two graphs we used; one that plots technical features and one that plots perceptive features. We found that it was not feasible nor necessary to include all the features of our dataset, so we asked, “what features are significant.” This filter took away the features we were most interested in and left the more trivial ones (e.g. whether or not the track is acoustic, live, etc.). So, we shifted to “what features are audiophiles most interested in.”

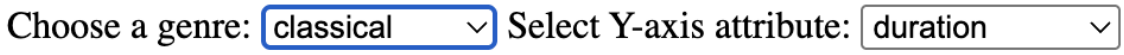
The second visualization analyzes the technical features against popularity. These features are explicitness, duration, tempo, time signature, and loudness. Throughout the process the design remained similar to what we initially planned it to be. Our initial design idea was to have a dropdown menu that allowed you to filter the graph based off of the features and see how that affected the popularity of music over the years. With the feature on the y-axis and popularity shown through a variation in color, we were able to have a good analysis of the question.



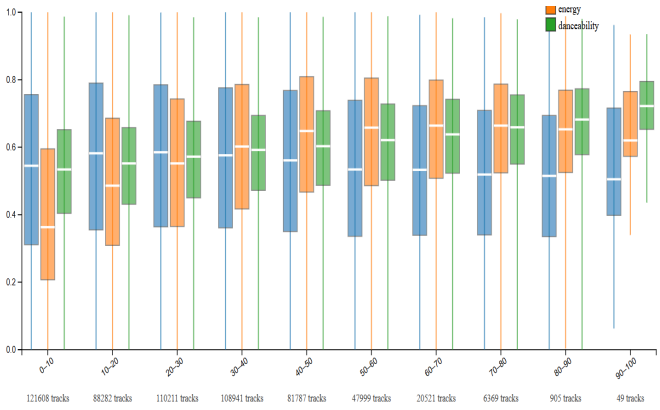
With the new implementation of filtering by genre, we edited this design slightly to interact well with the main graph. Now, the user can view the graph based on genre and feature (explicitness, duration, tempo, time signature, and loudness).



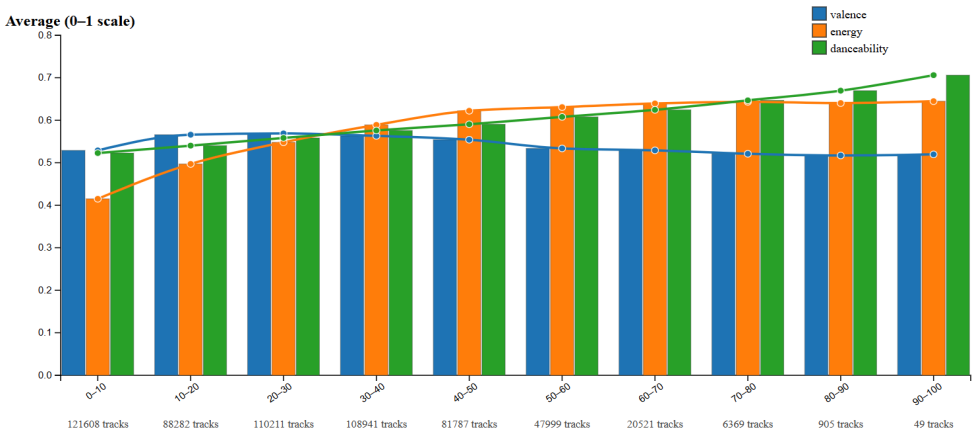
You aren't able to see our dropdown menu in this image, but it is as shown at the bottom of the screen. In the future this will probably be moved around for a more user friendly experience.

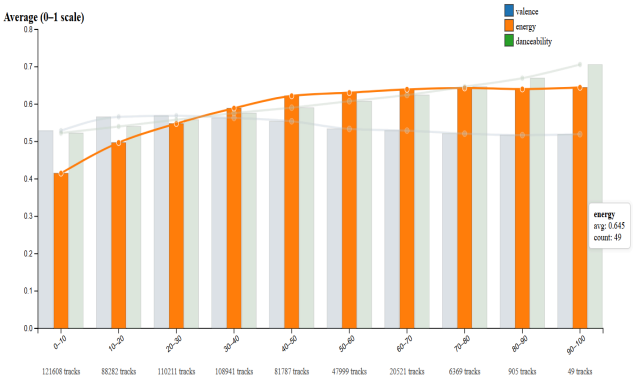


The third visualization plots three perceptive features against popularity: valence, danceability, and energy. These are the three most interesting perceptual features and the only perceptual features that vary naturally across all songs. Ten buckets hold songs of popularity scores by the ten (scores from 0-0.1,0.1-0.2, etc.). Each bucket has a bar for each feature that is the average value for that bucket. It should be noted that not all buckets have the same number of songs.



The alternative design for the third visualization (the perceptive features) was to replace the bars with boxplots. This would give the viewer substantially more information – a full spread for each bucket rather than only the average. This design, however, was too visually cluttered and added details that did not significantly contribute to the whole of the dashboard. Below is the current iteration…,





The current interaction features an interaction used to visually isolate the features; to highlight the feature that the user is hovering over and obscure the others. Overall, the graph allows the view to see all three present features and how they change with popularity, and interact with each other. With the interaction, the viewer can focus on one feature at a time. Lastly, trend lines were added to the tops of bars for each of the three features. This emphasizes the trend just a little bit better.

Implementation

Evaluation