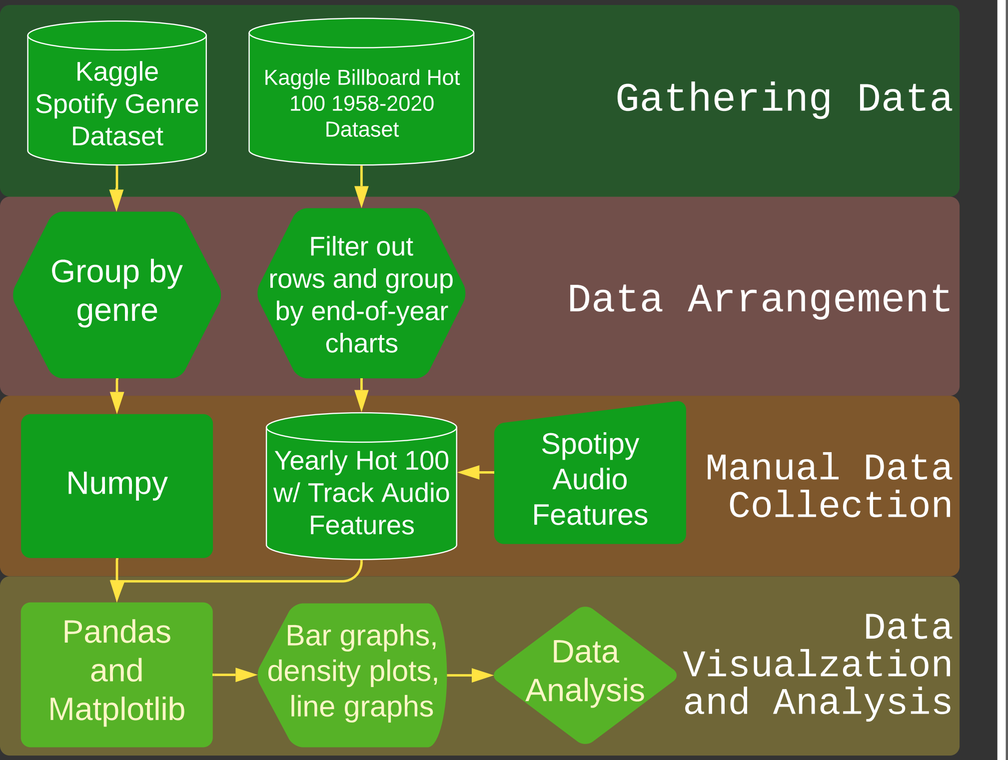
Variation in Musical Genres

Kenneth Kim
5/30/21
DH100

Sources:
https://www.kaggle.com/dhruvildave/billboard-the-hot-100-songs
https://www.kaggle.com/grasslover/spotify-music-genre-list

Github: https://github.com/doinkenny/DH100-MusicGenres
Jupyter Notebook: https://colab.research.google.com/drive/15vMaLu601Y2idMYt1M0CvuNcgMDdZmi9?usp=sharing

Workflow



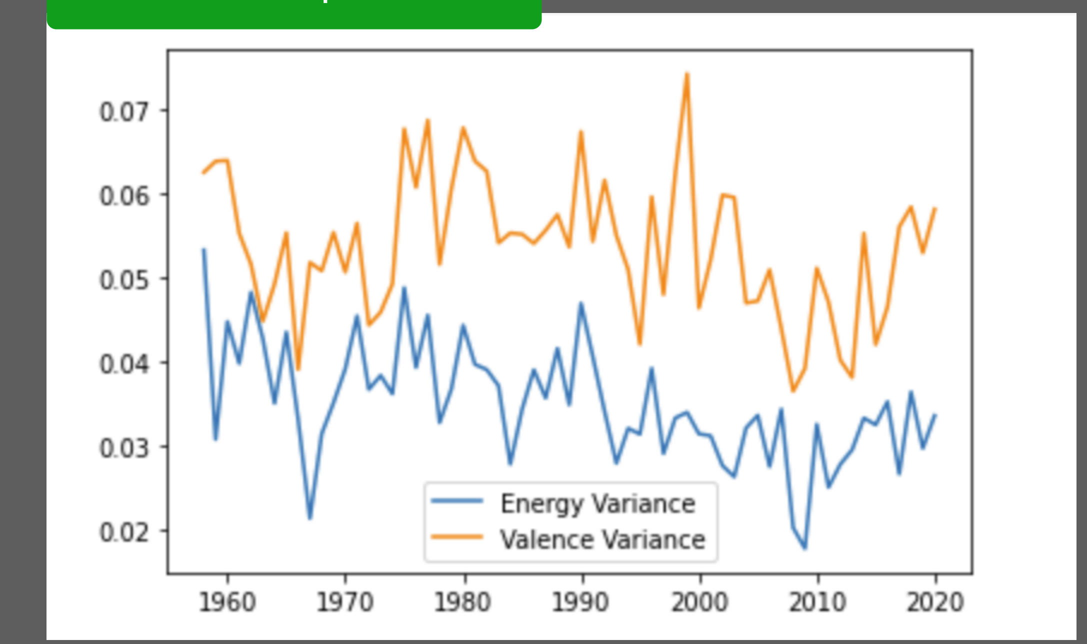
Description of the Datasets

songDB.tsv: A massive collection of ~80k tracks obtained using the Spotify API, featuring basic track info as well as audio features of tracks. These include popularity, genre, acousticness (measure of the utilization of live instruments), valence (how "happy" a track sounds), energy, danceability, etc.
charts.csv: Contains the "Hot 100" Billboard songs through the years 1958 to 2021

Project Description

The purpose of this project is to analyze the variance present in the songs of the many genres in music and determine which genres present the most and least amount of variance, perhaps revealing which genres provide a more diverse listening experience. The variance is calculated on various audio features of songs, which are explained in *Audio Features*. In addition to surveying all 626 genres present in the dataset, popular music will also be analyzed for trends in the variance throughout the years.

Variance in Popular Music



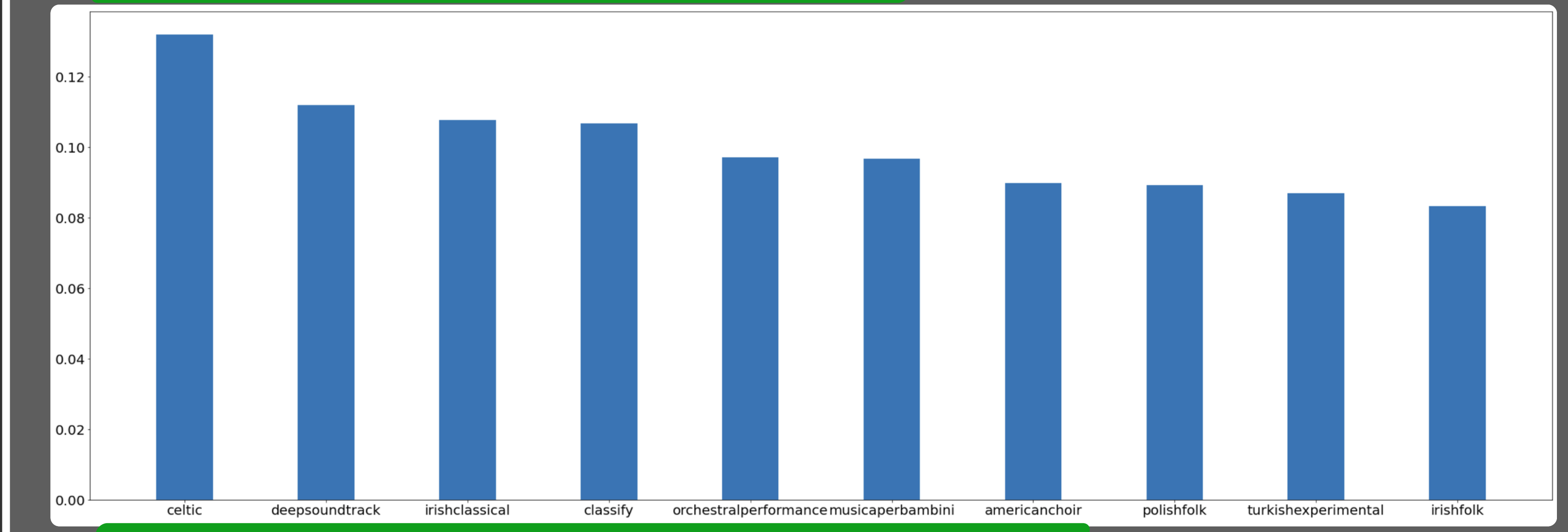
Research Questions

-Which genre displays the most variability in energy, valence, or acousticness?
-Which genre is seen the most in the top charts over the years?
-How is the variance in popular music and how has it changed over the years?

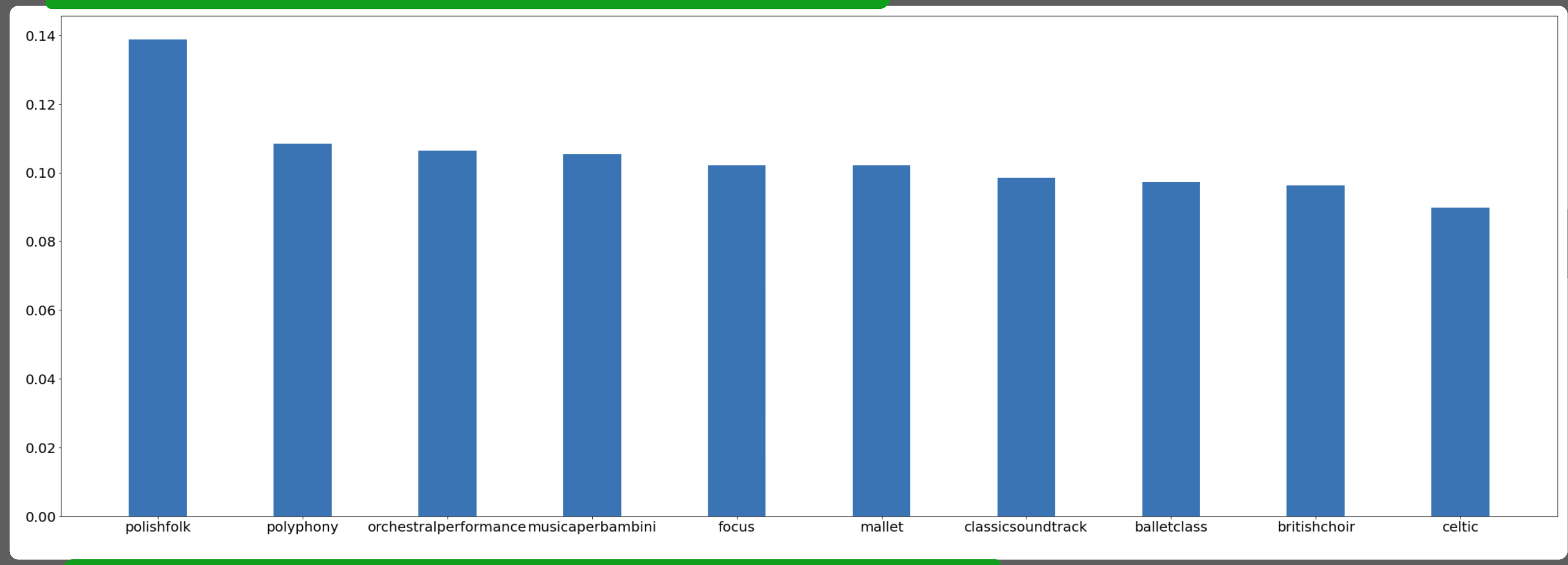
Hypothesis

I expect some trends in music to be fairly straightforward, with the more traditional genres such as classical and older rock to have more acousticness. The variance in energy and danceability levels in the electronic genres will probably be lower compared to other genres as EDM makes up such a large portion of electronic music. One genre I expect to see a large variance in valence levels is jazz, due to the wide range of experimentation and expressive instrumentation in the genre.

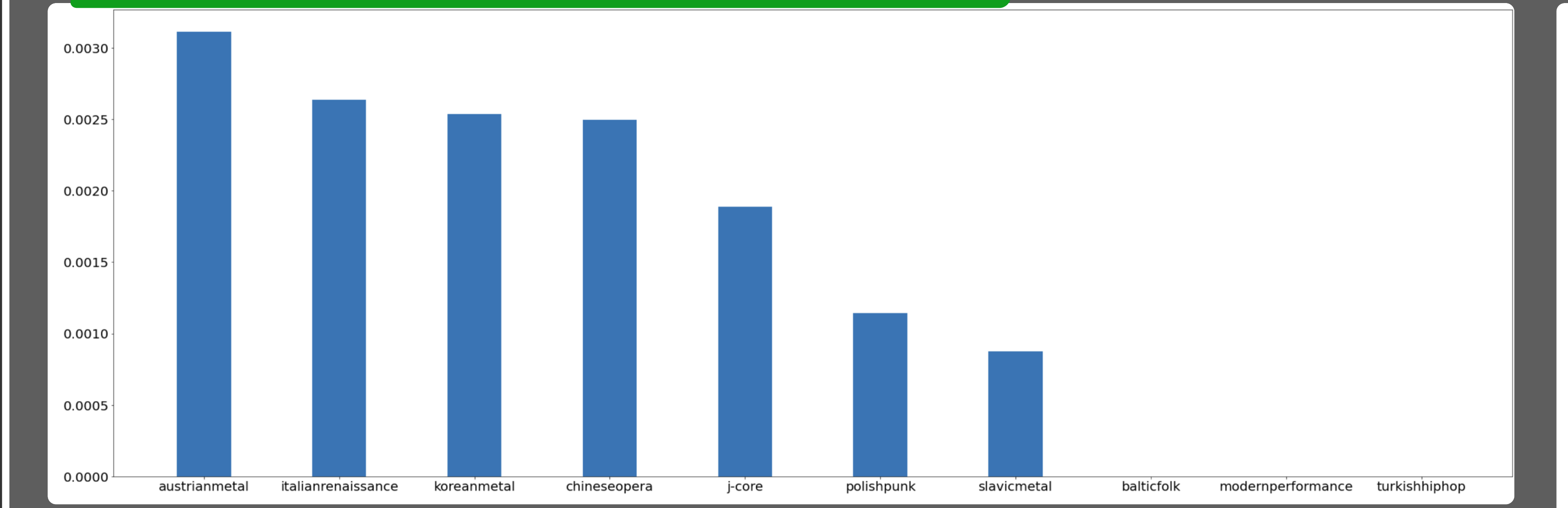
Top 10 Genres with Energy Variance



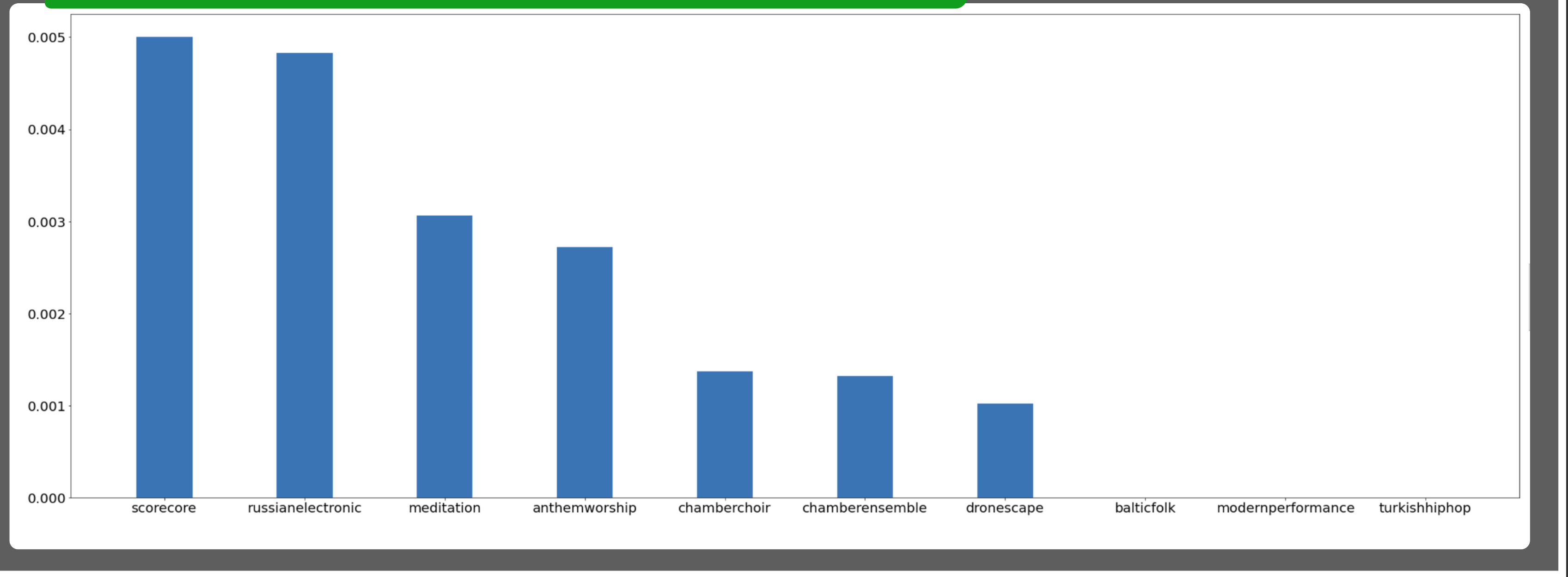
Top 10 Genres with Valence Variance



Top 10 Genres with Least Energy Variance



Top 10 Genres with Least Valence Variance



Tools and Methods

Pandas was used to retrieve the data from the .csv files. To get the top 10 data, I first gathered all the unique genres present in **songDb.tsv**, totalling to 626 genres. These genres then made up the keys to the dictionaries I utilized to divide up the rows, with each row representing a single track, into their respective genres. The variances of each genre could then easily be calculated using **numpy**. **Matplotlib** was used to create the bar graphs shown above, and pandas was once again used to create various density plots to observe the distribution of individual genres closer.

charts.csv was used for the data in the line graph. Due to the massive size in the dataset, and the fact that many songs tend to stay on the top 100 for weeks at a time, only the top 100 songs at the end of each year were used. Rows were grouped and separated into years. **Regex** was used to filter out the rows by date. The remaining songs were fed into the **Spotify API/spotipy,** where each row was searched by artist name and track title and assigned a track URI given by the API. These URIs allowed me to retrieve the audio features for these tracks, allowing me to calculate the variance of pop music per year.

Audio Features

Acousticness (0-1): the level acoustic instrumentation
Danceability (0-1): analyzes the bpm and rhthymic stability
Energy (0-1): measured based on loudness and intensity
Instrumentalness (0-1): the presence of vocals
Valence (0-1): how "happy" or positive a track sounds

Results

It's interesting to see the more classical/traditional and acoustic genres to have the most amount of energy variance. Classical music and folk tunes have varying amounts of instrumentation used within them, explaining the large amount of variance seen here. Videogame music stands out among this top 10, but can be easily explained with videogame music's wide-ranging intensities, ranging from atmospheric tracks to intense fight themes. Meanwhile on the other end, metal is a genre that shows up multiple times as having some of the least amount of energy variance. Metal is characterized with its aggressive sound and loud instruments, so it would make sense for it to have small energy variance due to its consistent loud nature. Similarly, j-core, or japanese hardcore, is a subgenre of EDM also known for its intensity, with its banging kick drum and fast BPM. Chinese opera is also present on this end for the opposite reasons, being characterized with its consistent soft sounds and minimal instrumentation.

Polish folk music stands out among the top 10. Characterized by many different types of intricate dances and inspired by historical events, it makes sense for it to have a wide variance. Other genres such as polyphony and orchestral performance may have ended up on the top 10 due to their genres being a wide umbrella expanding over many different songs. Dronescape and meditation belong on the top 10 least valence variance list as their music is used to fulfill a specific purpose and is far less varied in mood. Similarly, anthem worship and other contemporary christian music are designed to be simple and have an uplifitng mood, marking its spot on the top 10 with least variance.

Similar trends were found in both the energy and valence variability data. The genres with the most amount of variance tended to lean towards more classical and traditional genres that leaned on acoustic elements, while genres with the least amount of variability had more electronic or modern elements. This can in part be explained by how older genres are divided into less subdivisions compared to more modern genres, so older genres will cover a larger range of songs and moods and valence levels.

The similar trends between energy and valence variance are also seen in the variance in popular music line graph, where despite the lack of a clear trend in the overall variance of pop music, energy and valence variance shared a direct relationship.