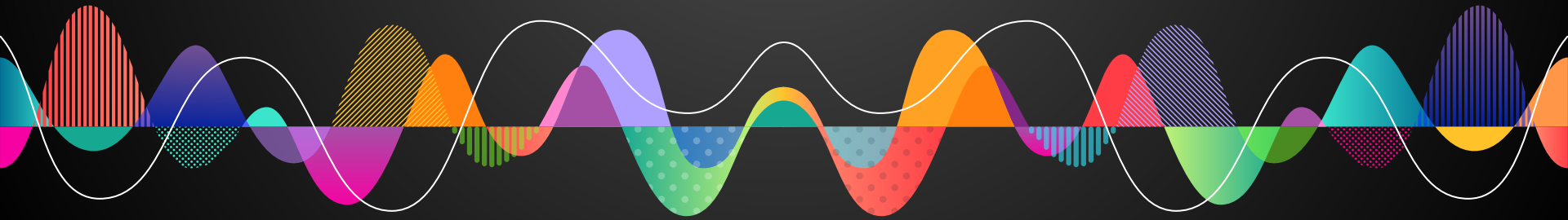
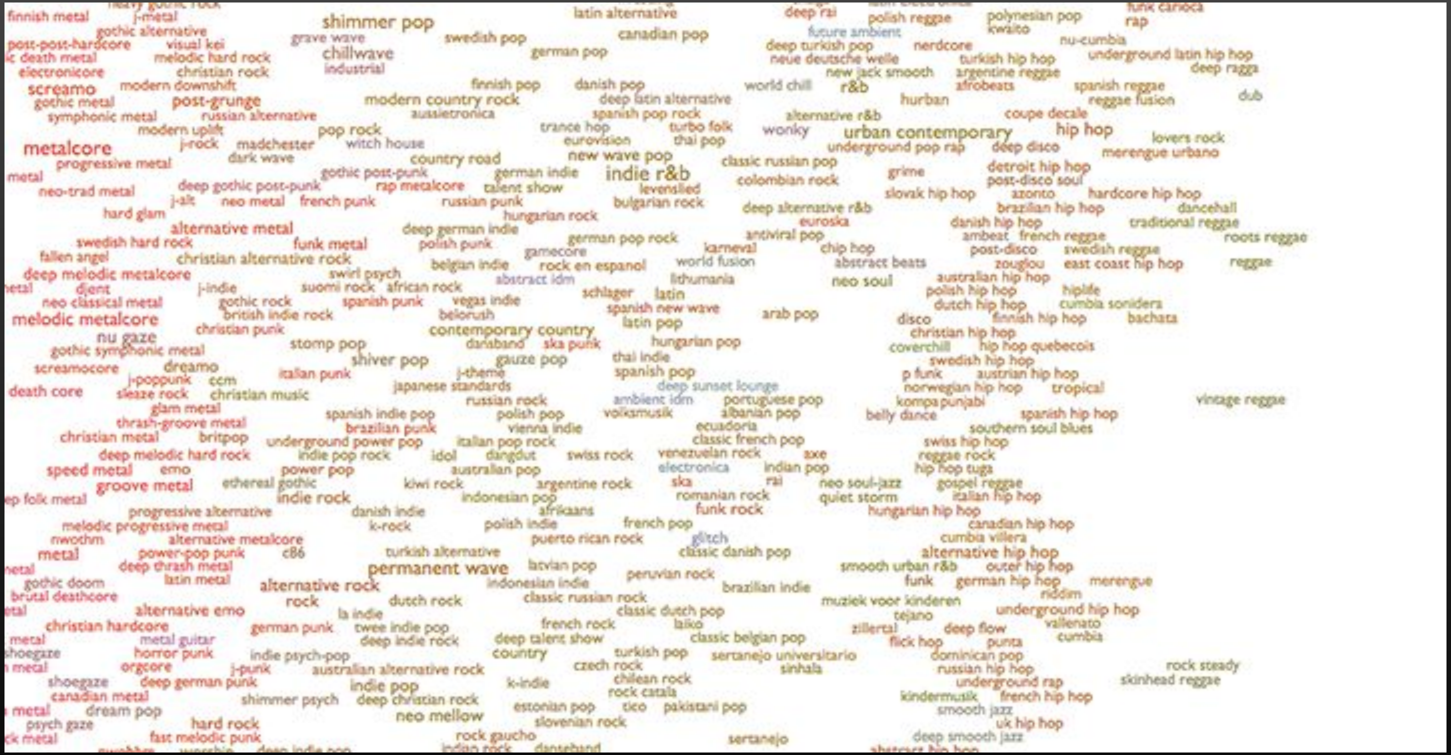
Kenneth Kim | Adam Anderson
DH100: Theory and Method in the Digital Humanities

Variability Across Music Genres

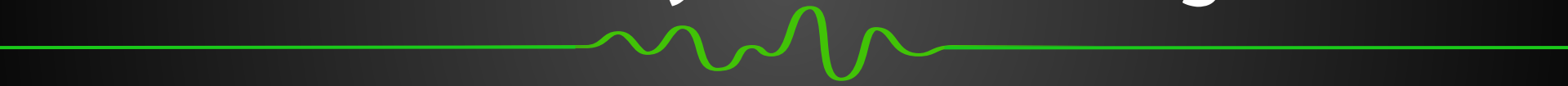
Outline

- Project Description
- Research Questions
- Description of the Datasets
- Tools and Methods
- Data Analysis and Interpretation




finnish metal
j-metal
gothic alternative
post-post-hardcore
visual kei
melodic hard rock
electronicore
christian rock
screamo
modern downshift
gothic metal
post-grunge
symphonic metal
russian alternative
modern uplift
metalcore
j-rock
madchester
progressive metal
dark wave
neo-trad metal
deep gothic post-punk
j-alt
neo metal
french punk
hard glam
alternative metal
swedish hard rock
funk metal
fallen angel
christian alternative rock
deep melodic metalcore
djent
j-indie
neo classical metal
gothic rock
british indie rock
melodic metalcore
nu gaze
christian punk
gothic symphonic metal
dreamo
screamocore
j-poppunk
ccm
death core
sleaze rock
christian music
glam metal
thrash-groove metal
christian metal
britpop
deep melodic hard rock
speed metal
emo
groove metal
ethereal gothic
progressive alternative
melodic progressive metal
nwothm
alternative metalcore
metal
power-pop punk
deep thrash metal
latin metal
gothic doom
brutal deathcore
alternative emo
christian hardcore
metal guitar
horror punk
orgcore
shoegaze
deep german punk
canadian metal
dream pop
psych gaze
hard rock
fast melodic punk
shimmer pop
grave wave
chillwave
industrial
swedish pop
german pop
finnish pop
modern country rock
aussietronica
pop rock
witch house
country road
gothic post-punk
rap metalcore
russian punk
deep german indie
polish punk
gamecore
belgian indie
swirl psych
suomi rock
african rock
abstract idm
spanish punk
vegas indie
belorush
contemporary country
stomp pop
shiver pop
gauze pop
ska punk
j-theme
italian punk
japanese standards
russian rock
spanish indie pop
polish pop
brazilian punk
vienna indie
underground power pop
italian pop rock
indie pop rock
idol
dangdut
power pop
australian pop
kiwi rock
indonesian pop
indie rock
danish indie
k-rock
afrikaans
polish indie
turkish alternative
permanent wave
latvian pop
alternative rock
rock
indonesian indie
dutch rock
la indie
german punk
twee indie pop
deep indie rock
indie psych-pop
j-punk
australian alternative rock
indie pop
shimmer psych
deep christian rock
neo mellow
rock gaucho
latin alternative
canadian pop
danish pop
deep latin alternative
spanish pop rock
trance hop
turbo folk
eurovision
thai pop
new wave pop
german indie
indie r&b
talent show
levenslied
bulgarian rock
hungarian rock
german pop rock
karneval
rock en espanol
world fusion
lithumania
schlager
latin
spanish new wave
latin pop
hungarian pop
thai indie
spanish pop
deep sunset lounge
ambient idm
portuguese pop
volksmusik
albanian pop
ecuadoria
classic french pop
swiss rock
venezuelan rock
electronica
argentine rock
ska
romanian rock
funk rock
french pop
puerto rican rock
glitch
classic danish pop
peruvian rock
classic russian rock
classic dutch pop
french rock
laiko
deep talent show
classic belgian pop
country
turkish pop
sertanejo universitario
czech rock
k-indie
chilean rock
rock catala
estonian pop
tico
pakistani pop
slovenian rock
sertanejo
deep rai
polish reggae
future ambient
deep turkish pop
neue deutsche welle
new jack smooth
world chill
r&b
hurban
alternative r&b
wonky
urban contemporary
underground pop rap
classic russian pop
grime
colombian rock
slovak hip hop
deep alternative r&b
euroska
antiviral pop
chip hop
abstract beats
neo soul
arab pop
disco
coverchill
p funk
belly dance
polynesian pop
kwaito
nerdcore
turkish hip hop
argentine reggae
afrobeats
coupe decale
hip hop
deep disco
detroit hip hop
post-disco soul
azonto
brazilian hip hop
danish hip hop
ambeat
french reggae
post-disco
zouglou
australian hip hop
polish hip hop
dutch hip hop
finnish hip hop
christian hip hop
hip hop quebecois
swedish hip hop
austrian hip hop
norwegian hip hop
kompa
punjabi
spanish hip hop
southern soul blues
swiss hip hop
reggae rock
hip hop tuga
neo soul-jazz
gospel reggae
quiet storm
italian hip hop
hungarian hip hop
canadian hip hop
cumbia villera
alternative hip hop
smooth urban r&b
outer hip hop
funk
german hip hop
muziek voor kinderen
tejano
zillertal
deep flow
flick hop
punta
dominican pop
russian hip hop
underground rap
kindermusik
french hip hop
smooth jazz
uk hip hop
deep smooth jazz
funk carioca
rap
nu-cumbia
underground latin hip hop
deep ragga
spanish reggae
reggae fusion
dub
lovers rock
merengue urbano
hardcore hip hop
dancehall
traditional reggae
roots reggae
swedish reggae
east coast hip hop
reggae
hiplife
cumbia sonidera
bachata
tropical
vintage reggae
merengue
riddim
underground hip hop
vallenato
cumbia
rock steady
skinhead reggae

Which genre presents the most amount of variability across its songs?



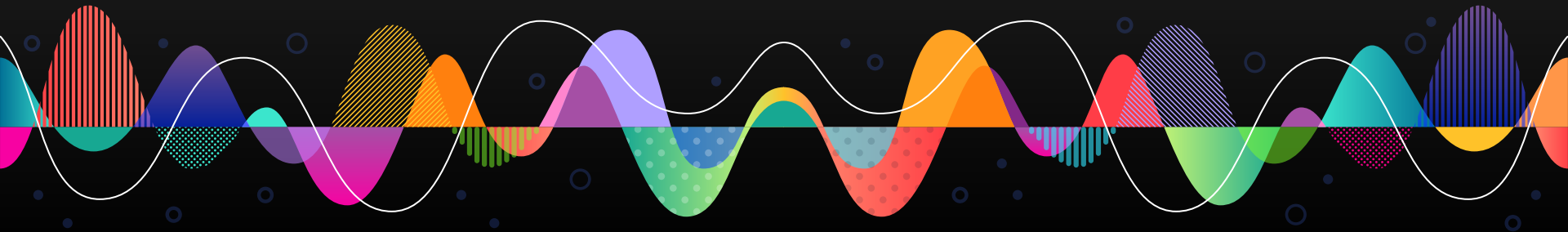
The Datasets

Spotify

- 80k tracks
- 626 unique genres
- Audio features

Billboard “The Hot 100” Songs

- Spans 1958-2021
- Weekly charts

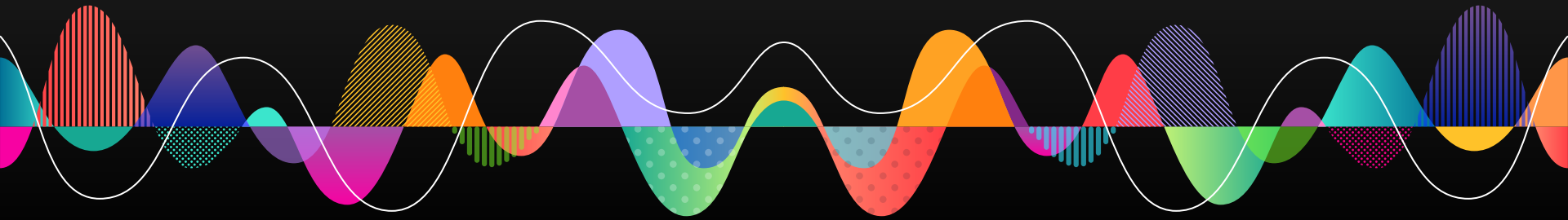


Audio Features

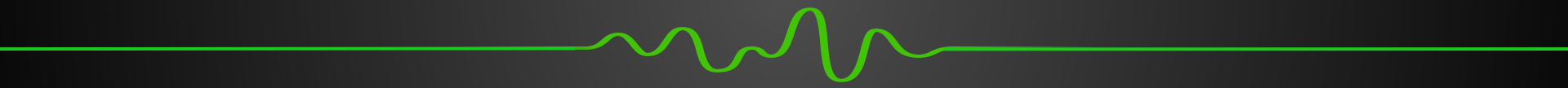
Acousticness (0-1): the level acoustic instrumentation
Danceability (0-1): analyzes the bpm and rhythmic stability

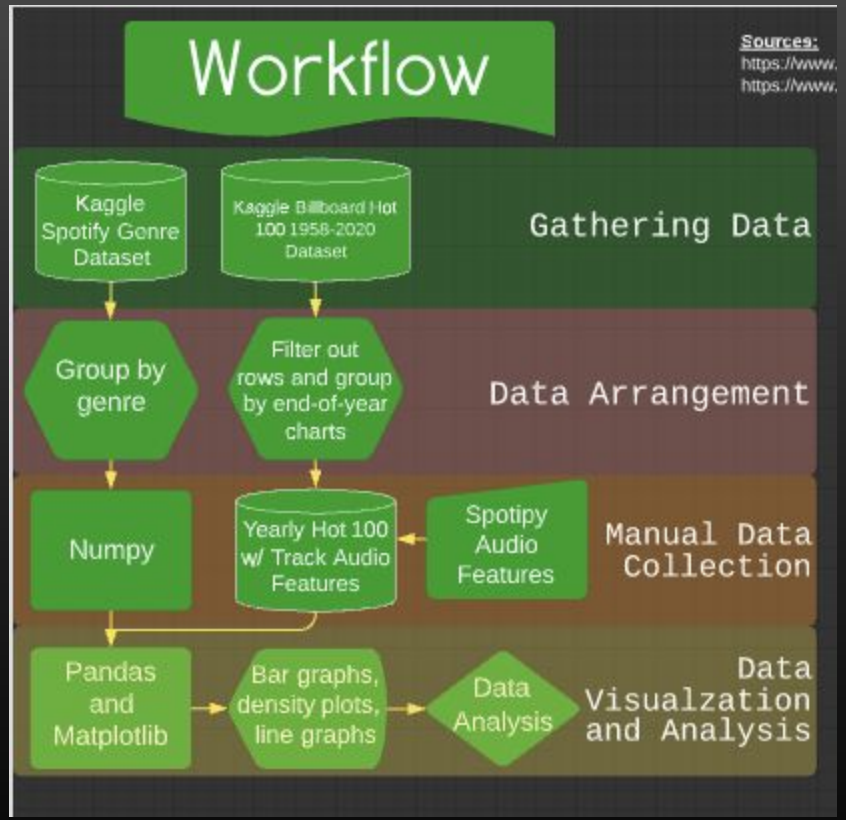
Energy (0-1): measured based on loudness and intensity
Instrumentalness (0-1): the presence of vocals

Valence (0-1): how "happy" or positive a track sounds



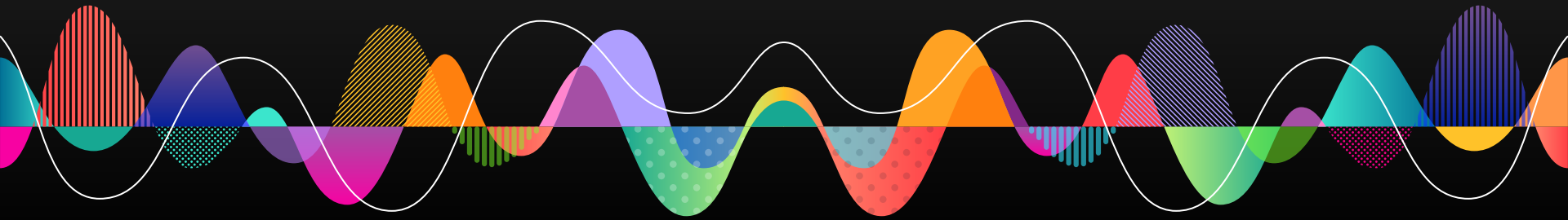
Tools and Methods




Workflow
Sources:
https://www.
https://www.
Kaggle Spotify Genre Dataset
Kaggle Billboard Hot 100 1958-2020 Dataset
Gathering Data
Group by genre
Filter out rows and group by end-of-year charts
Data Arrangement
Numpy
Yearly Hot 100 w/ Track Audio Features
Spotipy Audio Features
Manual Data Collection
Pandas and Matplotlib
Bar graphs, density plots, line graphs
Data Analysis
Data Visualzation and Analysis

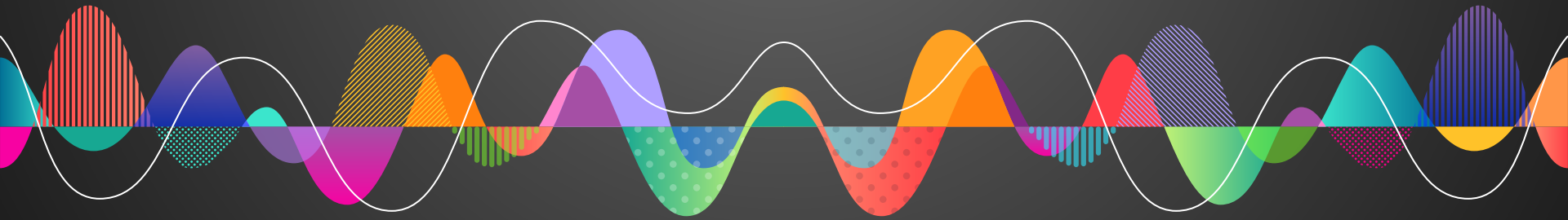
Tools and Methods

- Pandas: data → pandas dataframes
- Regex: filtering
- Numpy: variance calculations
- Spotipy: getting audio features
- Pandas and matplotlib: plotting data



Jupyter Notebook

https://colab.research.google.com/drive/15vMaLu601Y2idMYt1MOCvuNcgMDdZmi9?authuser=1#scrollTo=QbDDxcfH_In1



Results

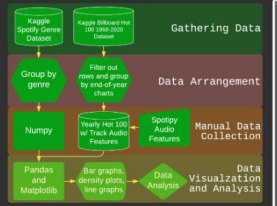
Variation in Musical Genres

Kenneth Kim
5/30/21
DH100

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

Github: https://github.com/dokkenny/DH100-MusicGenres
Jupyter Notebook: https://colab.research.google.com/drive/15vMaLu6O1Y2xlMYt1M0CvuNcgMDdZm9?usp=sharing

Workflow

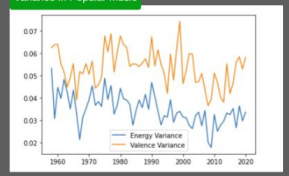


Description of the Datasets

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



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 uplifting 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.

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?

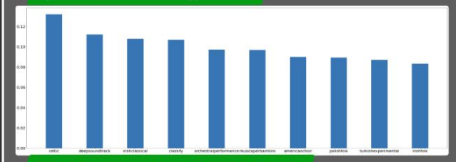
Hypothesis

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.

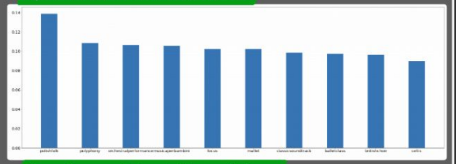
Audio Features

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

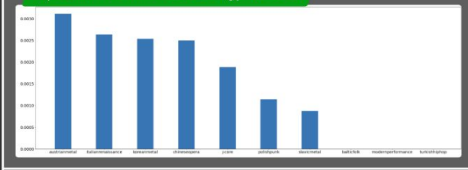
Top 10 Genres with Energy Variance



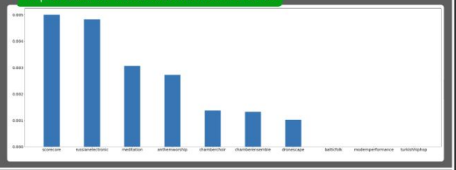
Top 10 Genres with Valence Variance



Top 10 Genres with Least Energy Variance



Top 10 Genres with Least Valence Variance



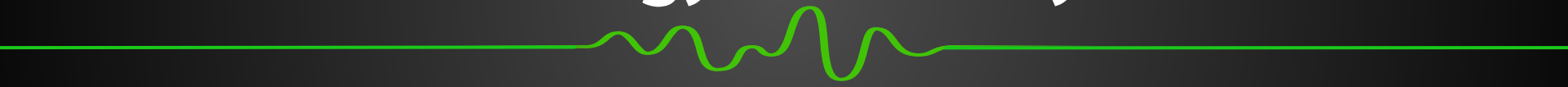
Tools and Methods

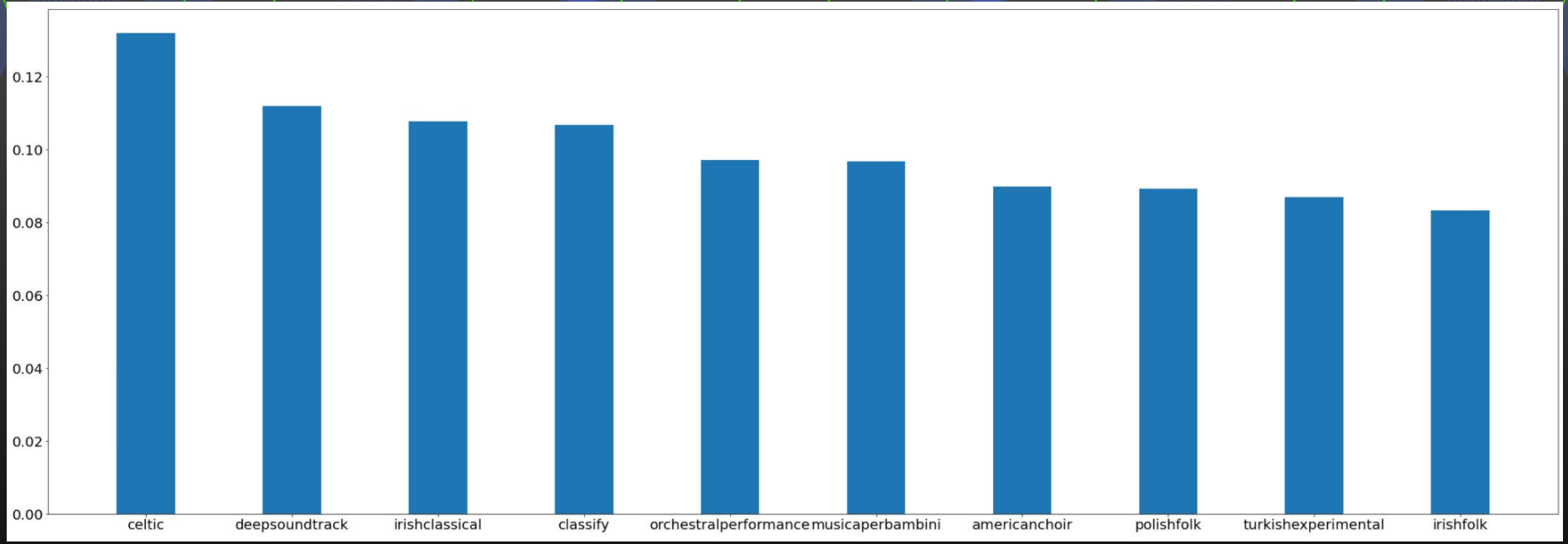
Pandas was used to retrieve the data from the .csv files. To get the top 10 data, I first gathered all the unique genres present in **songDb.tsv**, totalling to 626 genres. These genres then made up the keys to the dictionaries I utilized to divide up the rows, with each row representing a single track, 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.

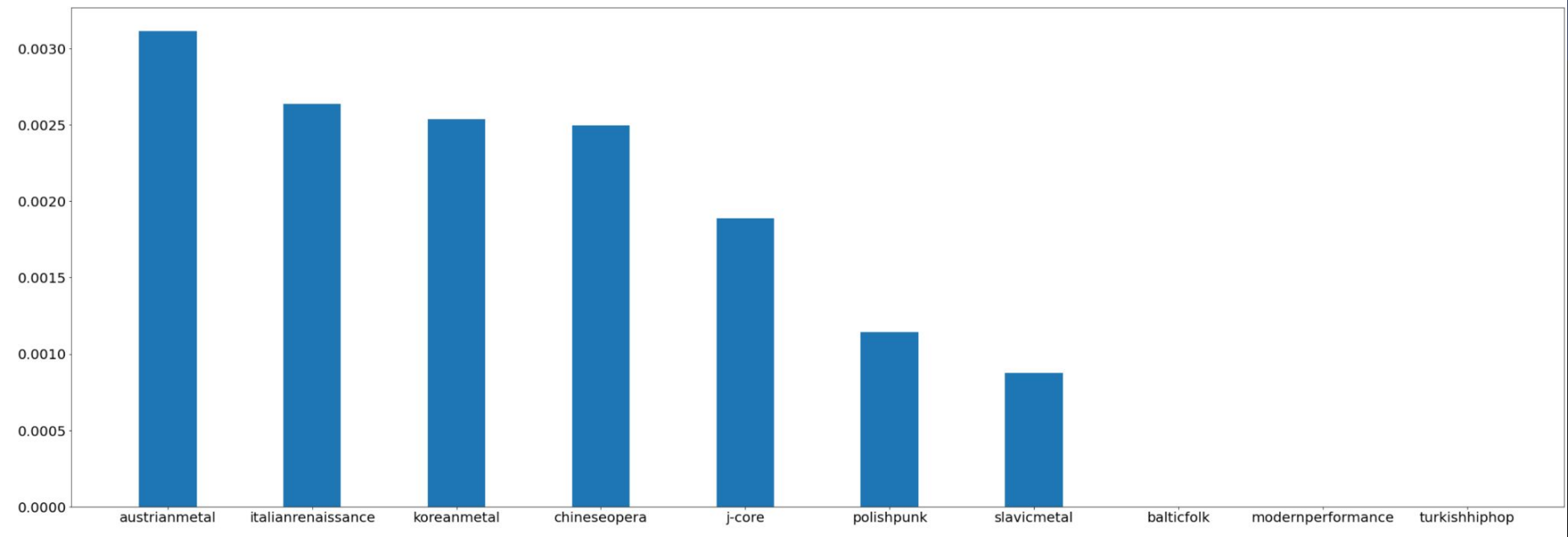
Storyboard

Energy Variability

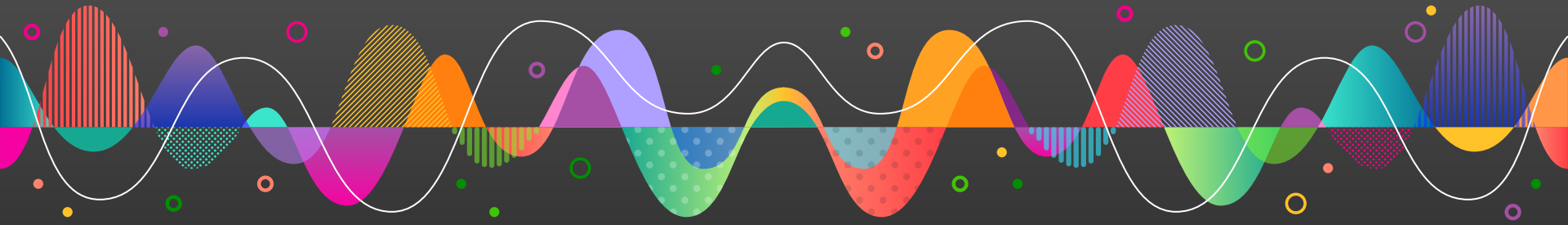




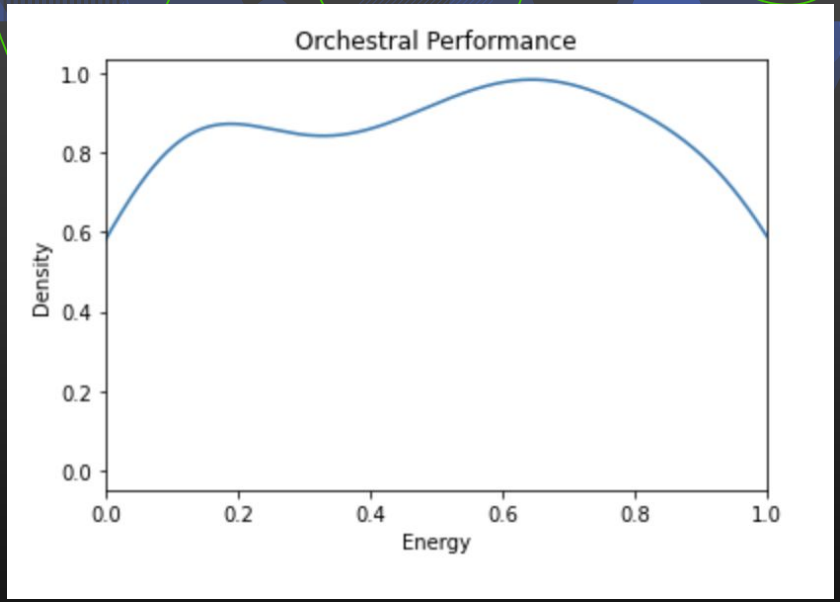
Top 10 Genres with Most Energy Variance

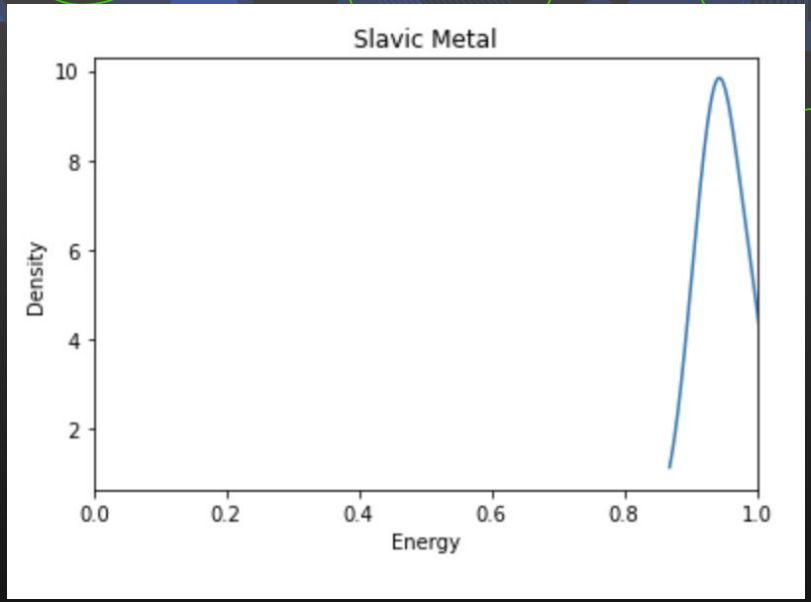


Top 10 Genres With Least Energy Variance

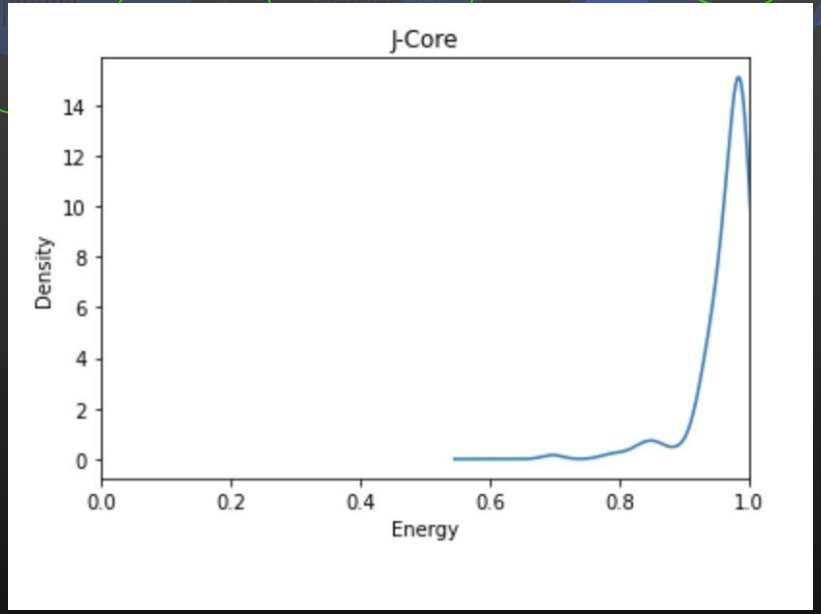


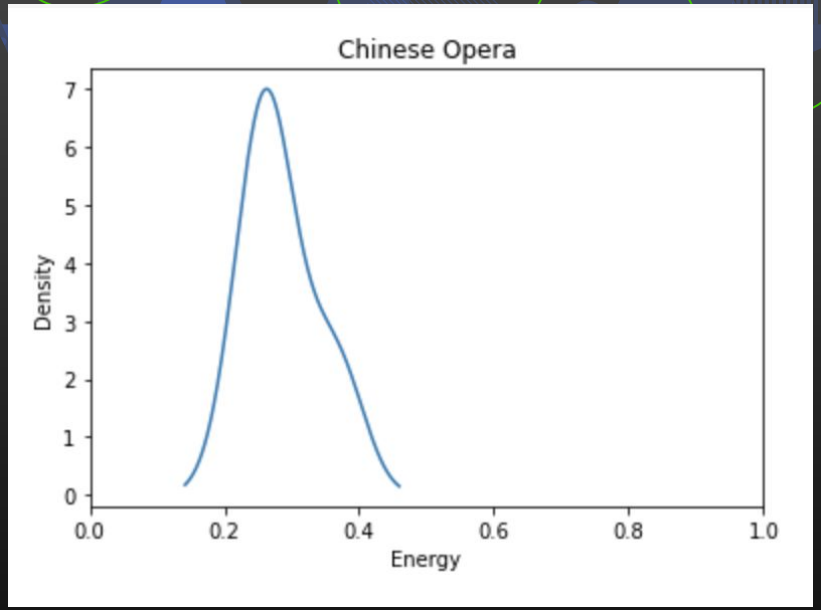
A closer look...



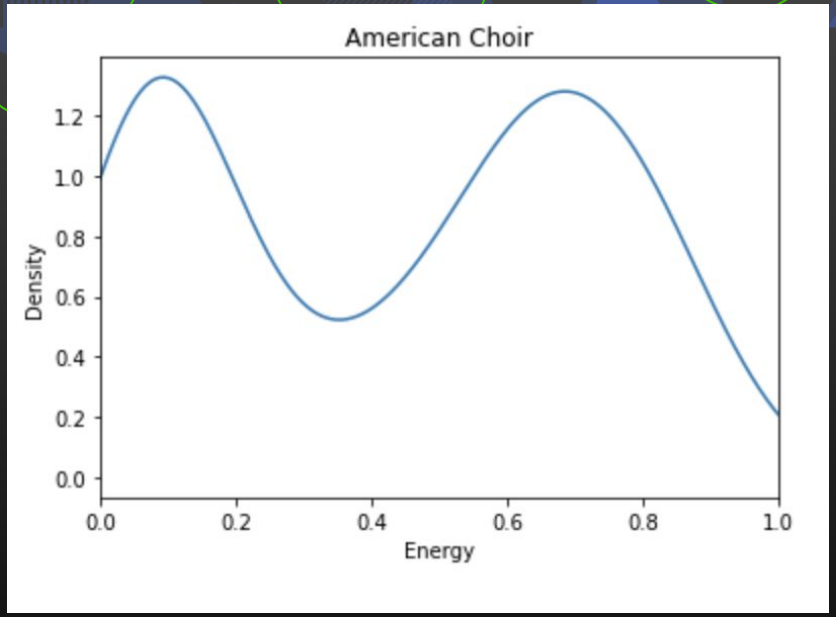


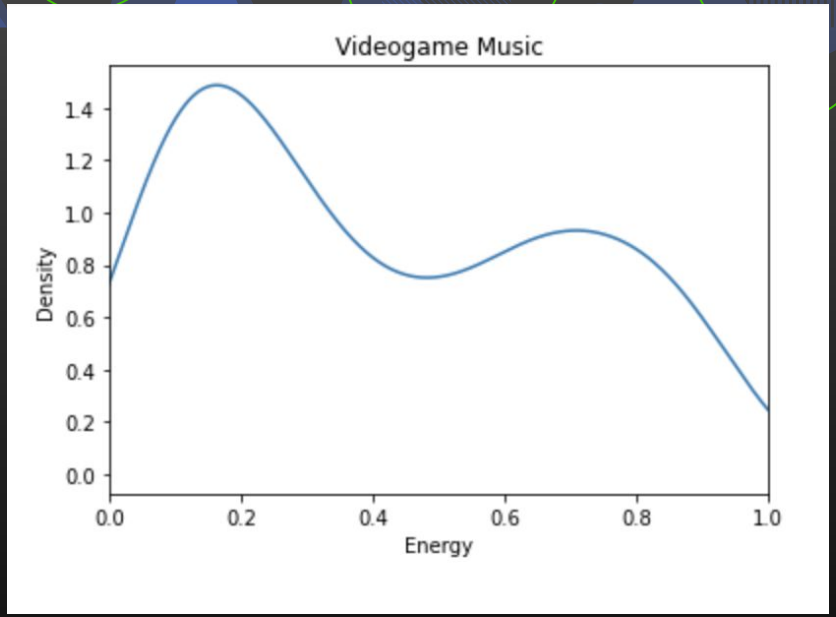
Orchestral Music vs Slavic Metal





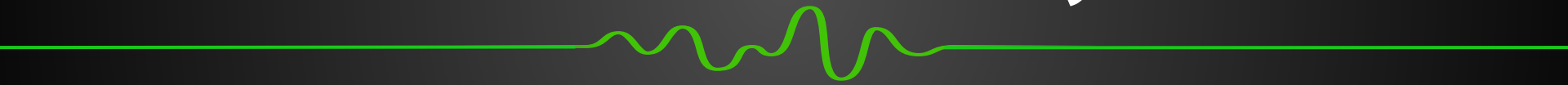
J-Core vs Chinese Opera

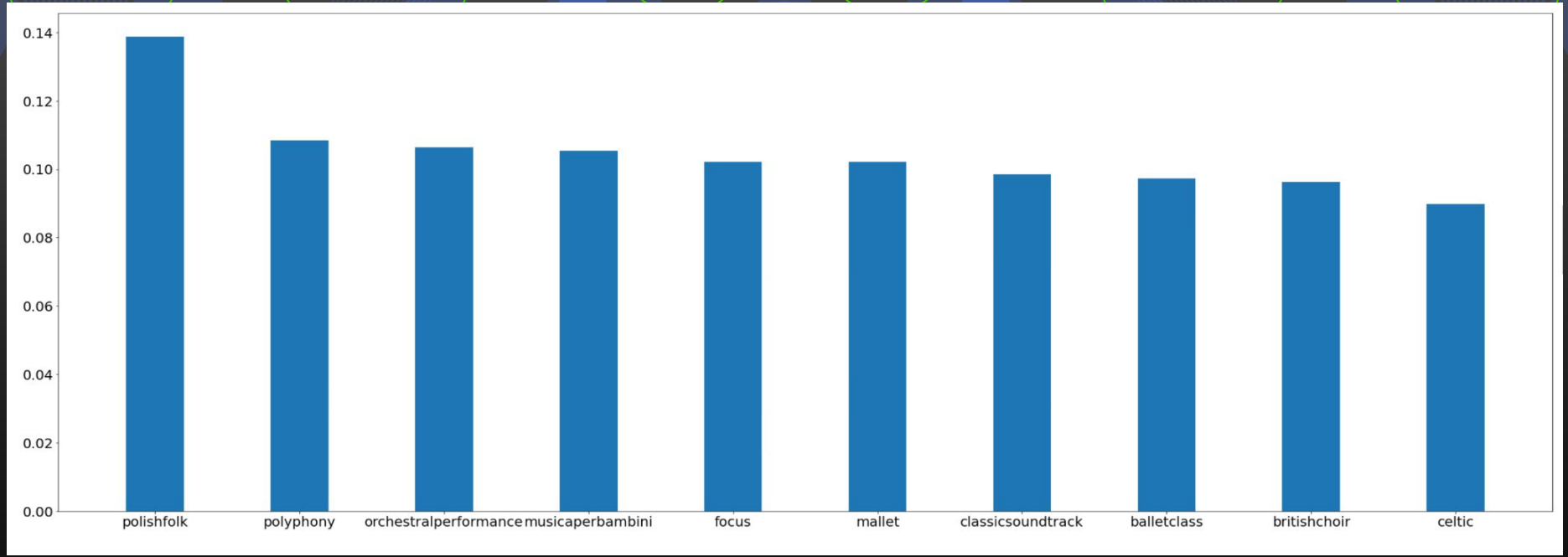




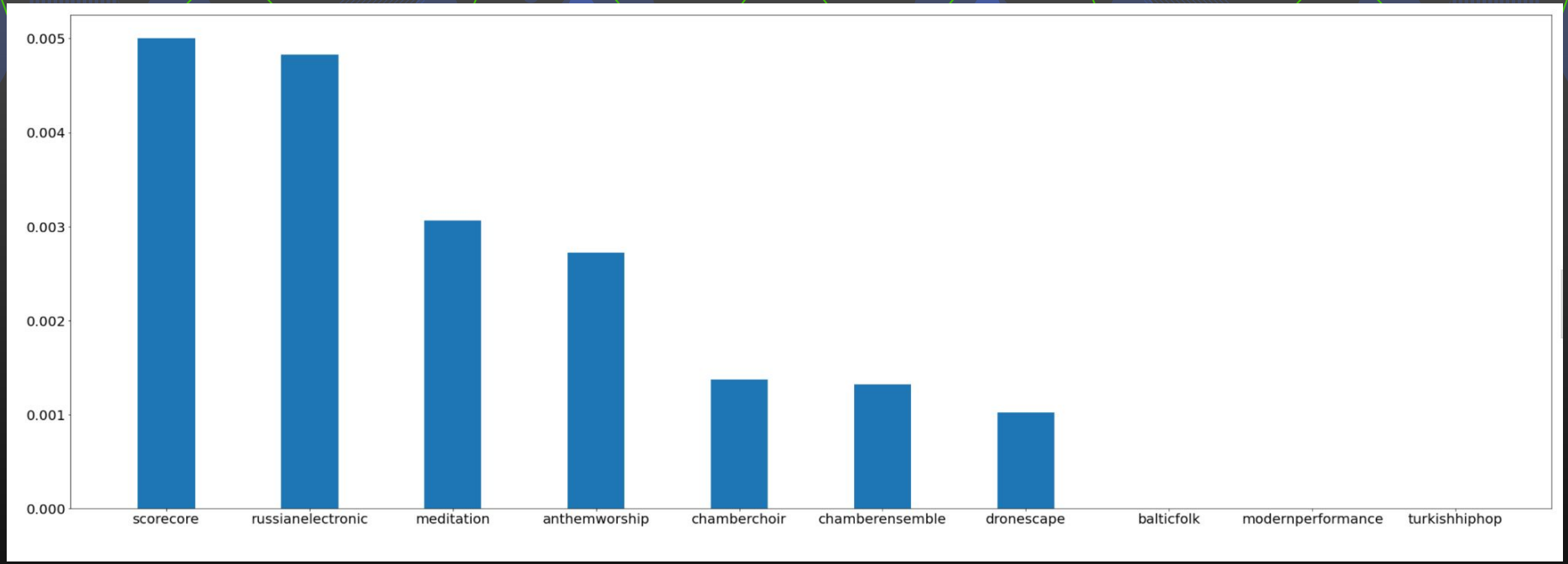
American Choir and Videogame Music

Valence Variability

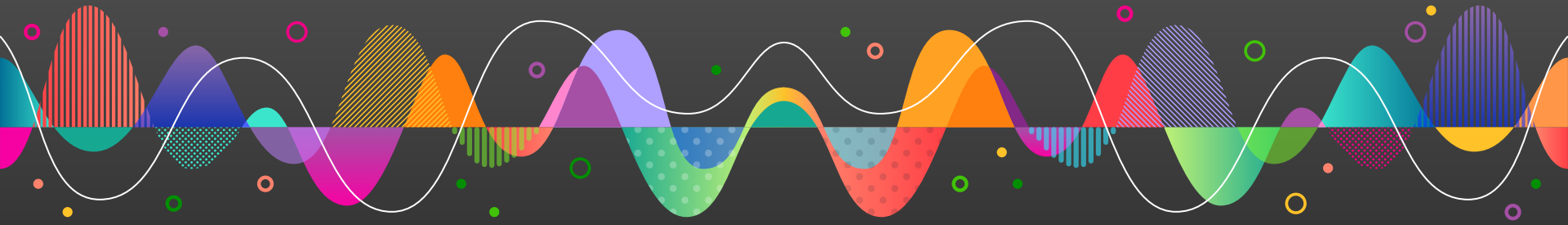




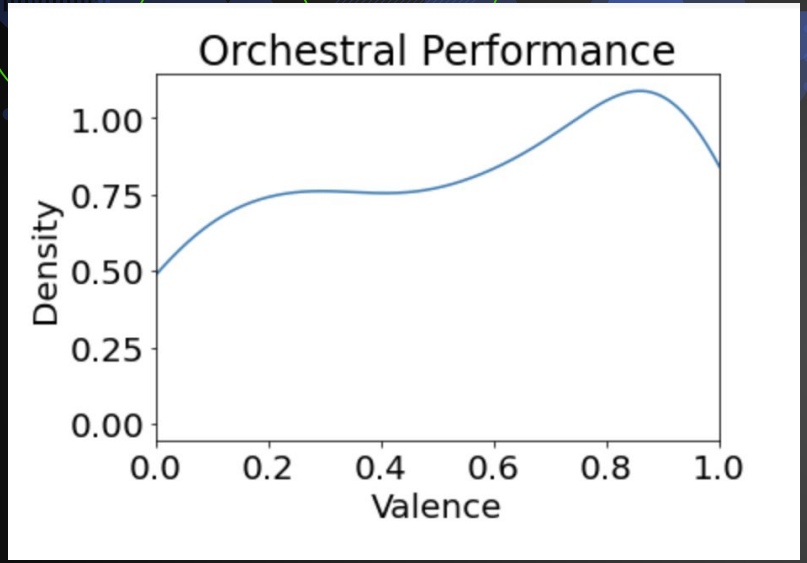
Top 10 Genres with Most Energy Variance

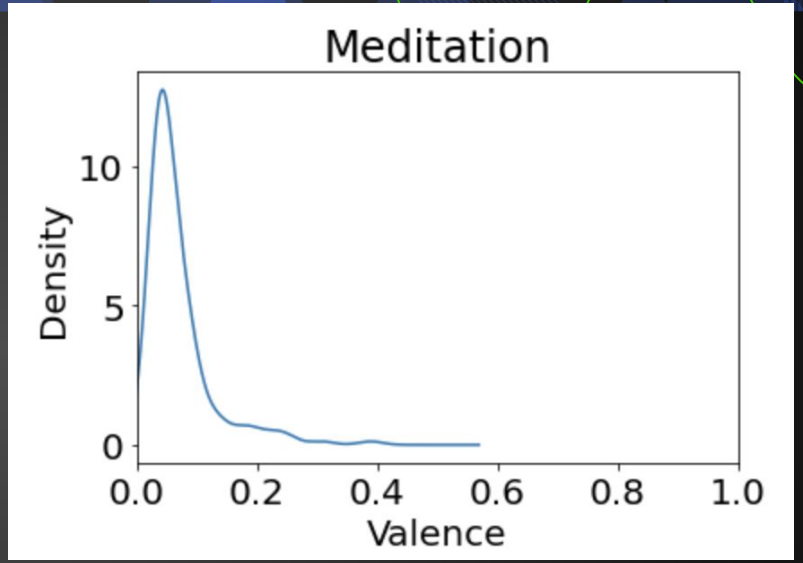


Top 10 Genres with Least Valence Variance

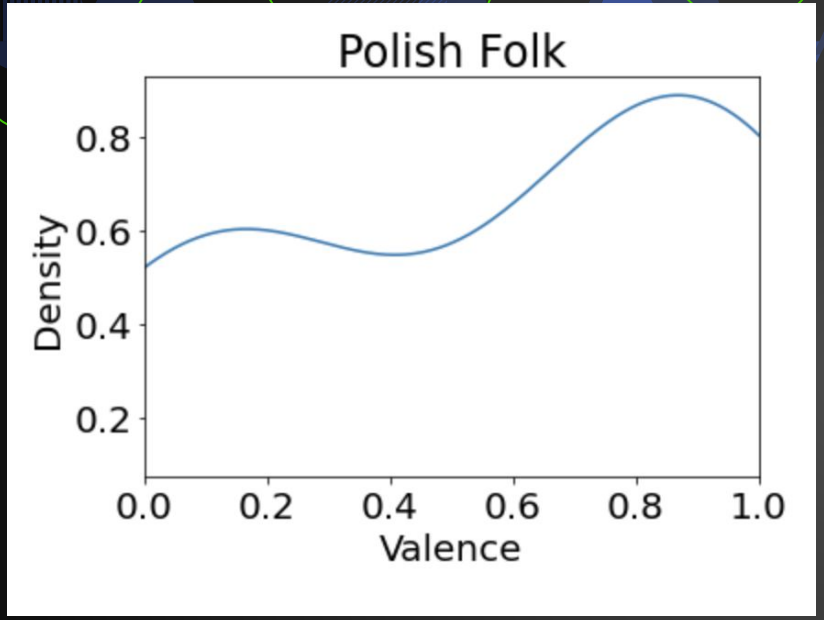


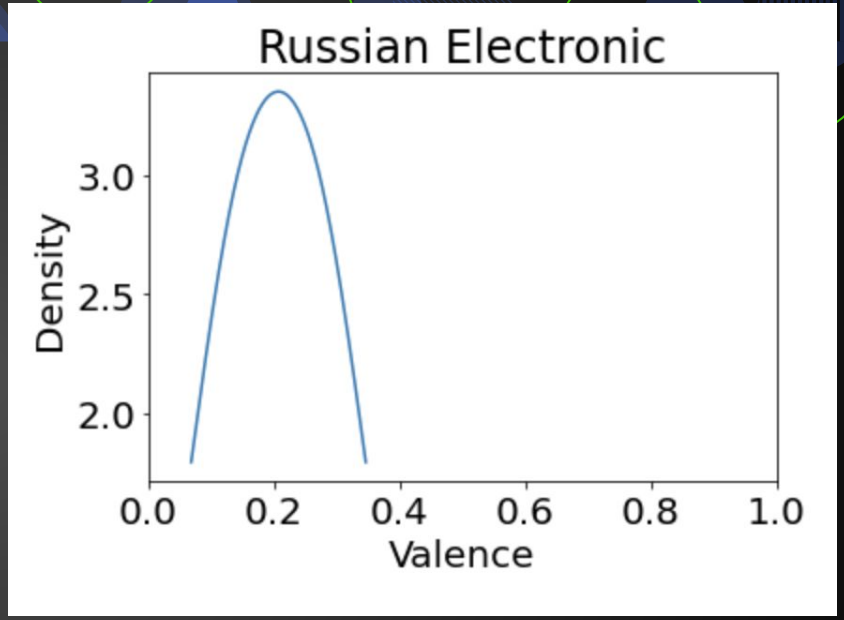
A closer look...



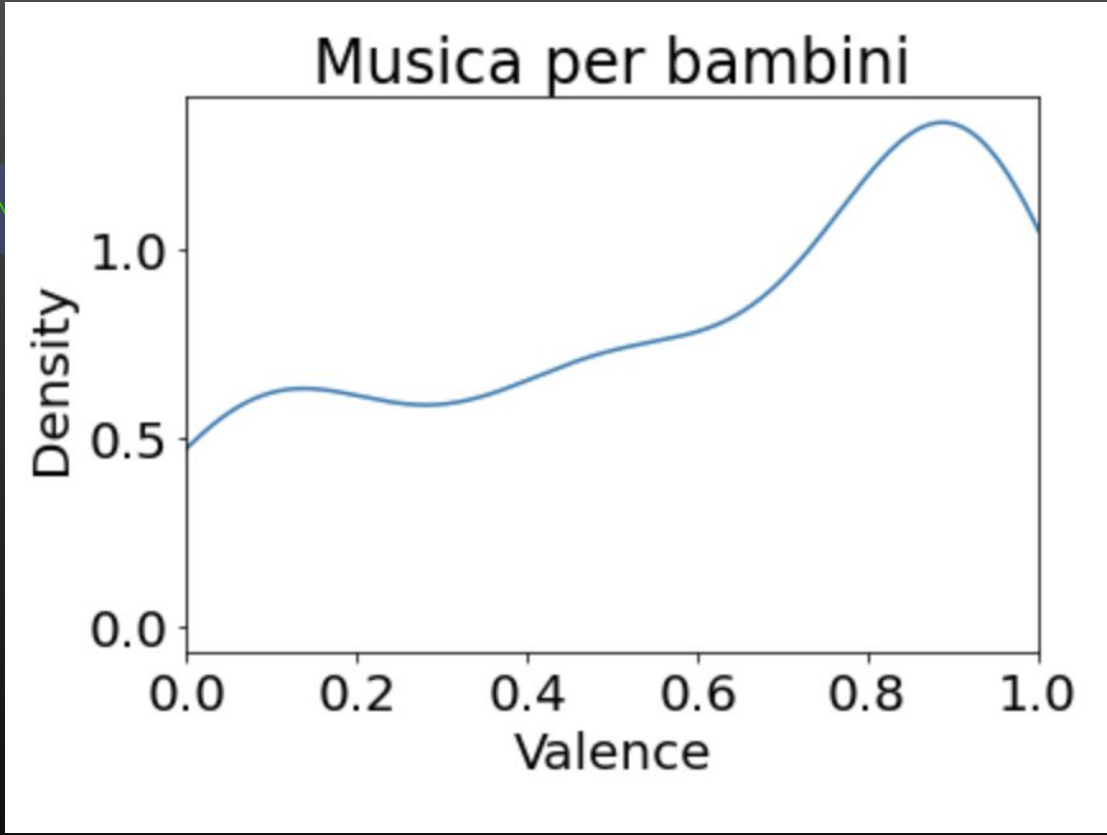


Orchestral Performance vs Meditation

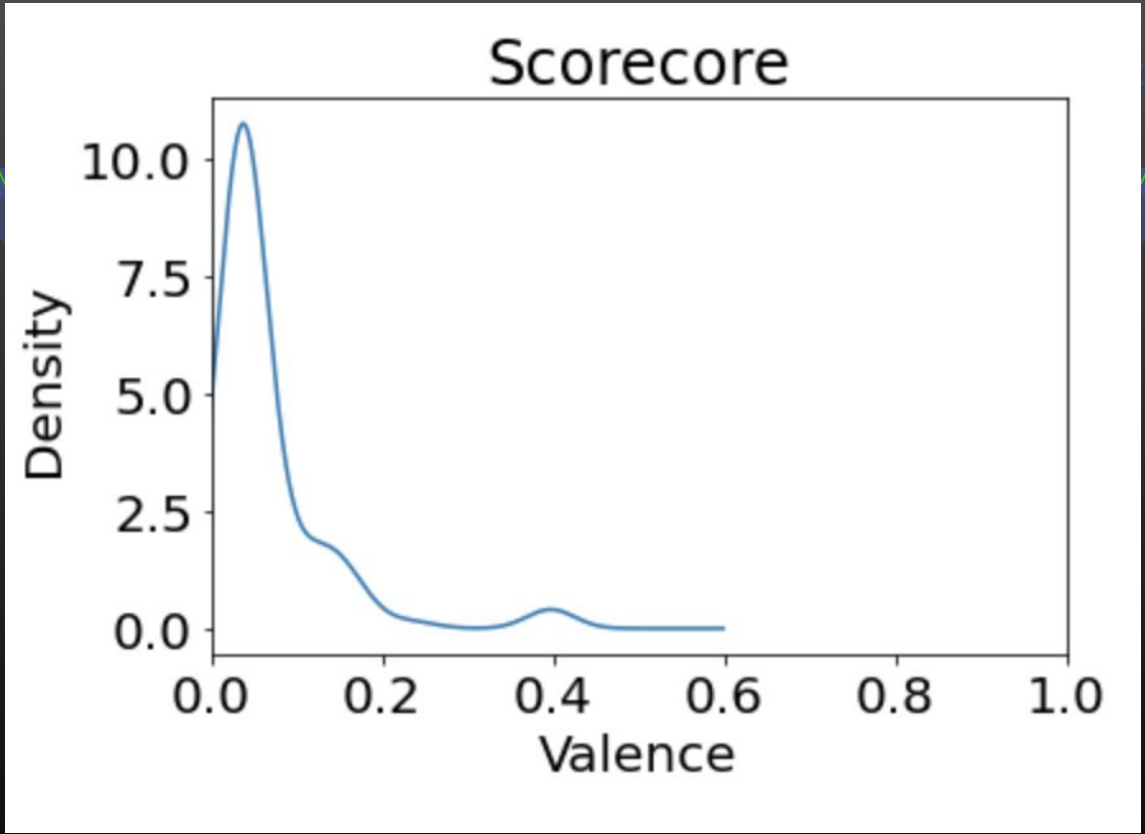




Polish Folk vs Russian Electronic

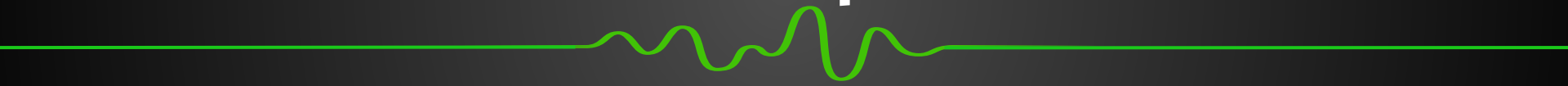


Musica per bambini

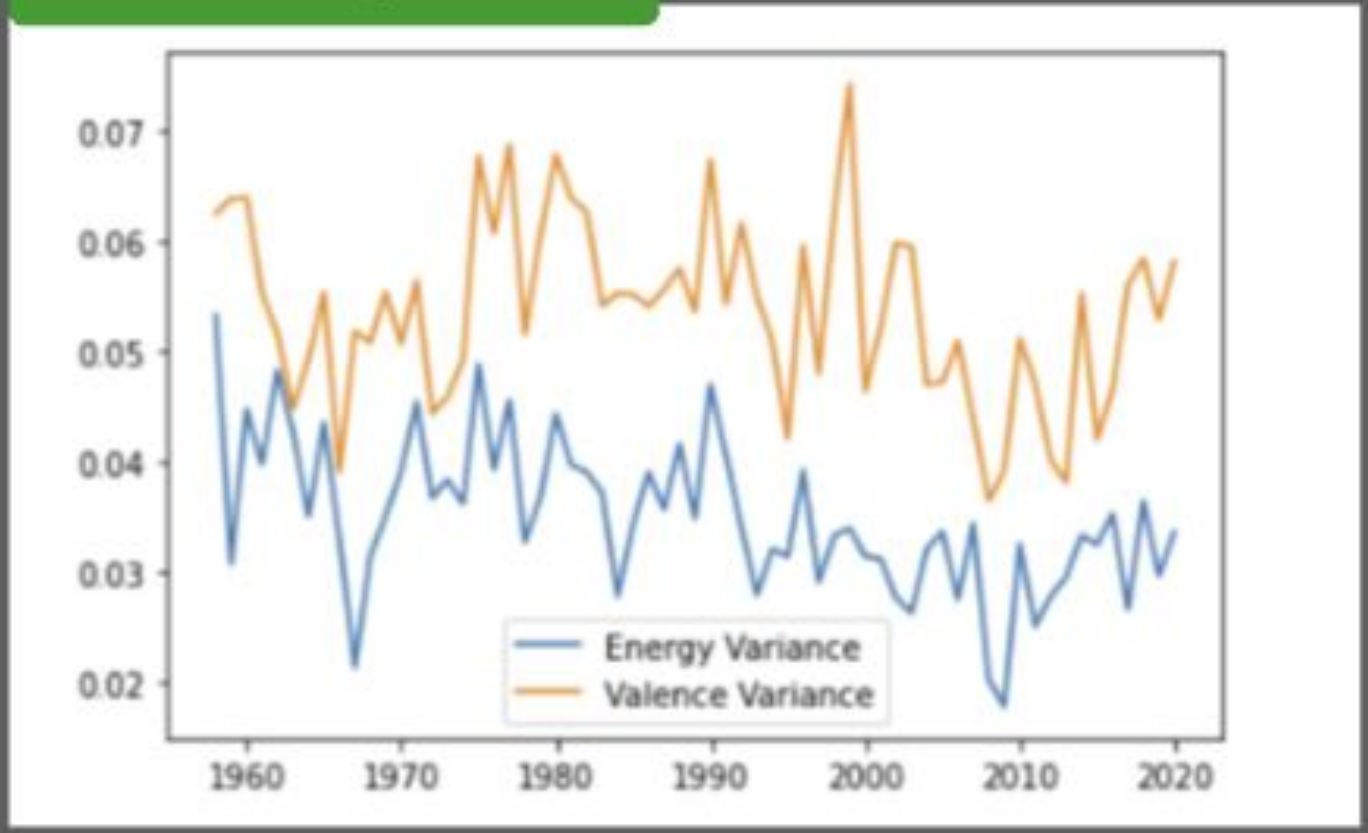


Scorecore

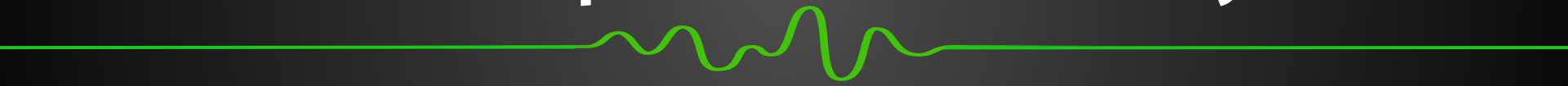
Variance in Popular Music



Variance in Popular Music

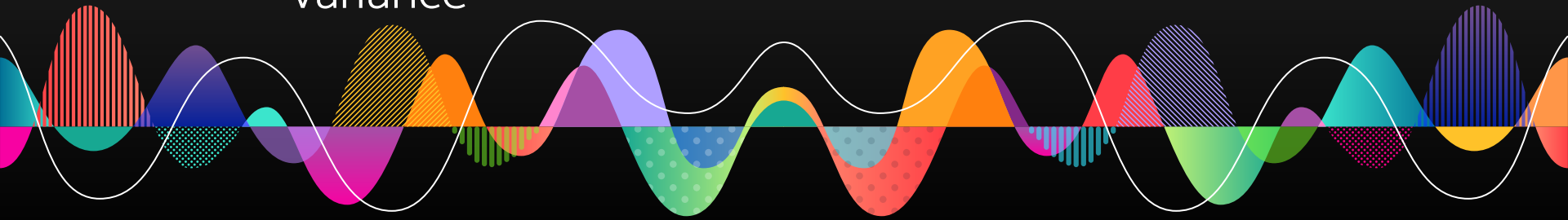


Data Interpretation and Analysis



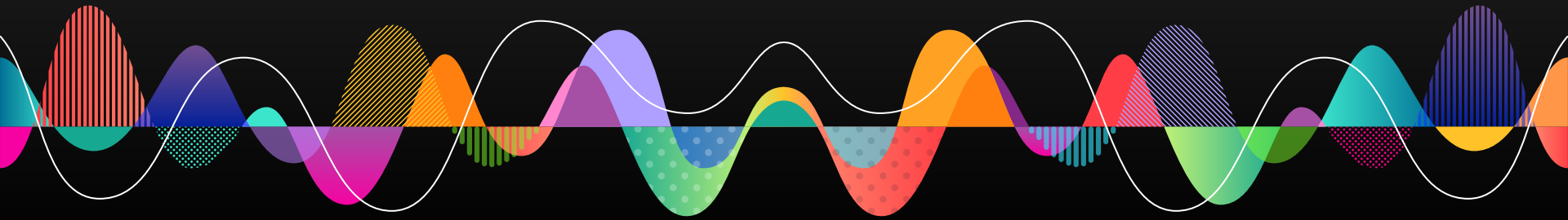
Energy

- Bimodal distributions found in genres with lots of energy variance
- More acoustic genres tend to have greater variance
- Multiple metal genres in top 10 least energy variance



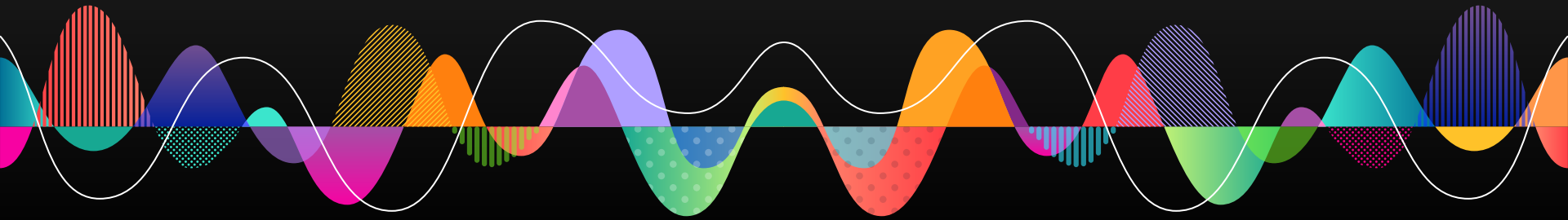
Valence

- Similar genres in top 10 to energy variance top 10
- Some genres have broader umbrellas
- More divergent genres in the least top 10



The Hot 100 Over the Years

- No clear trend in the overall variance
- Energy variance vaguely resembles downwards trend
- Energy and valence variance share a direct relationship



Thank you!