1. Dataset

**Dataset:** Top Hits Spotify from 2000-2019

**Link:** https://www.kaggle.com/datasets/paradisejoy/top-hits-spotify-from-20002019

**Explanation:** This dataset contains audio statistics of the top 2000 tracks on Spotify from 2000-2019.

**Reasons of choosing this dataset:** This data set displays information about popular songs from 2000 to 2019. The information is relatively comprehensive and can reasonably describe a song. In addition to the songwriter and genre, it is also described from a professional perspective, such as the song's tempo, loudness, energy, etc.

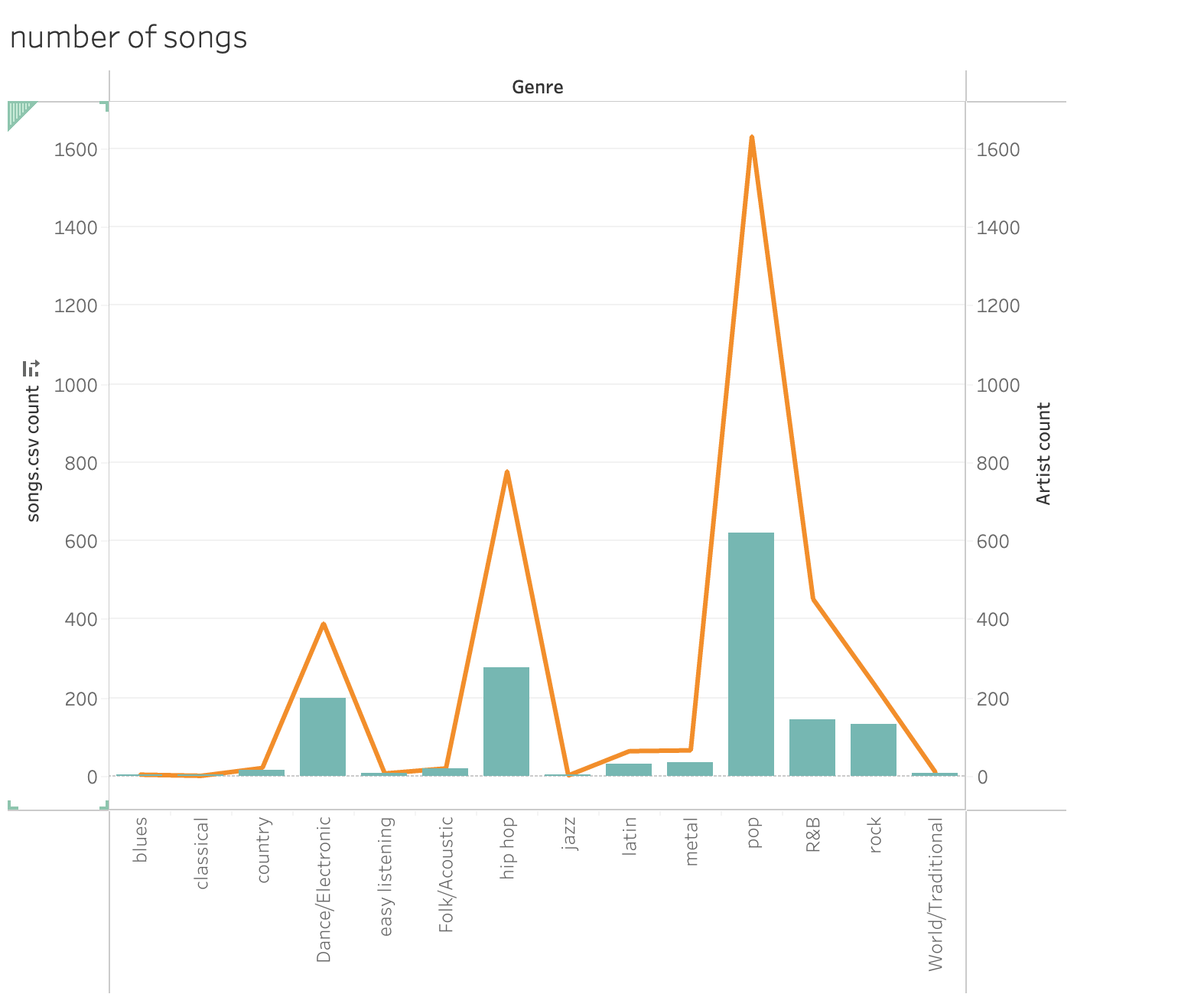
|  |  |
| --- | --- |
| **Column Name** | **Description** |
| artist | Name of the Artist. |
| song | Name of the Track. |
| duration\_ms | Duration of the track in milliseconds. |
| explicit | The lyrics or content of a song or a music video contain one or more of the criteria which could be considered offensive or unsuitable for children. |
| year | Release Year of the track. |
| popularity | The higher the value the more popular the song is. |
| danceability | Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable. |
| energy | Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. |
| key | The key the track is in. Integers map to pitches using standard Pitch Class notation. |
| loudness | The overall loudness of a track in decibels (dB). |
| mode | Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0. |
| speechiness | Speechiness detects the presence of spoken words in a track. |
| acousticness | 1.0 represents high confidence the track is acoustic. |
| instrumentalness | Predicts whether a track contains no vocals. |
| liveness | Detects the presence of an audience in the recording. |
| valence | A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry). |
| tempo | The overall estimated tempo of a track in beats per minute (BPM). |
| genre | Genre of the track. |

2. Purpose of Visualization

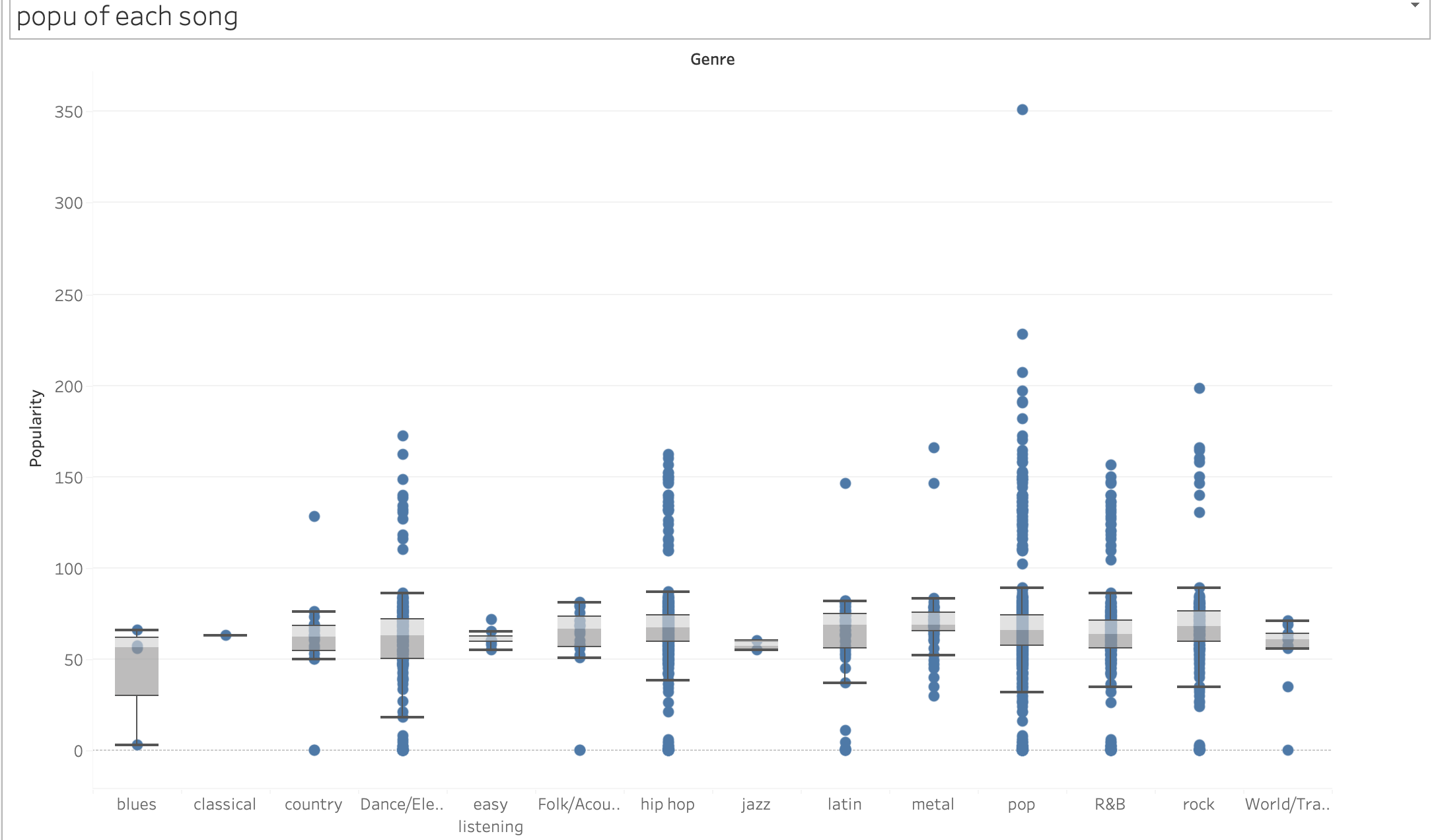
In this project, by data visualization to observe what characteristics of music people like, what characteristics each music type has, etc.

3. Visualization

1. popularity of music in each genre



As a whole, the number of music and the number of musicians are positively correlated. Among them, pop is the most popular type, while jazz, blues and other music are relatively niche.



Seeing the popularity, we can see the median of all kinds of music is similar, while pop, hip-hop, and dance have more outliers, means that these popular music has mixed reviews. However, the evaluations of jazz and classical are relatively consistent.

2. characteristics of each kind of famous music

图表, 条形图

描述已自动生成

As a whole, the loudness is positive relative with the energy. The metal and Latin is more energetic louder, while jazz is quietest and least energetic.

图表, 散点图

描述已自动生成

Heatmaps can show where scatterplots cluster. Most of the tracks focus on louder sounds and faster tempos, indicating that listeners prefer fast-paced songs.

地图

描述已自动生成

In this bubble chart, the size of the bubble represents popularity, and the depth of the color represents liveness. It can be seen that, generally speaking, people prefer live performance tracks.