Kasidy Fernandes

Kyle Woods

2022 Viral Song Explorer - Process Book

CS6330 - Final Project

Overview and Motivation

As a DJ it is important to stay in touch with popular music and ongoing trends in order to entertain and engage your audience. While there are numerous resources available for DJ's to track such trends, they often require a paid subscription (like BPM Supreme), or are for very specific, niche-genres (like Beatport - charts). Additionally, these resources consist of fairly static lists of songs and do not offer much additional information or interactivity for further exploration of the songs that are popular.

Available datasets include many interesting variables for each song including fundamental traits like the tempo, key, duration and loudness as well as perceived traits such as energy, danceability, valence, and speechiness of a song. With the number of variables available there is an opportunity to encode these values in much more interesting and possibly enlightening visualizations other than the lists/tables that are readily available.

A major indicator of song popularity can also be found on social media platforms, especially TikTok, and by how viral songs are. Users of the platform know how often the same songs will get used and how quickly new trends will spread. By injecting a viral song into a DJ set, you can capitalize on the audience's associations with the song for some fun audience engagement to break up your DJ set.

Planning out a DJ set in order to create a cohesive and interesting set of songs can take lots of preparation and planning to arrange tracks together for smooth transitions while two songs are overlapping. The major factors for what songs will mesh well are the tempo or beats per minute (BPM) and the key that songs are in. It's generally a best practice to mix songs of similar BPMs and with keys that are harmonically similar to avoid clashing sounds. Sometimes, putting together a mix can feel like putting together a jig-saw puzzle and you are hunting for a certain piece to fit in order to work your way through a list of songs. A motivating factor for this project is to provide a tool that would help indicate some popular, trending songs that any DJ could slot into a set to grab the audience's attention and break up a mix of music.

Related Work

Data surrounding trends with music on the TikTok platform have been the topic of some research (Vox - We tracked what happens after TikTok songs go viral) which is an influence on this project.

Questions

For this project we are taking an exploratory approach to analyze common trends in popular music. The datasets used for this project contain the audio characteristics (BPM, key, danceability, energy, etc) of the top charted songs from Spotify and TikTok. We aim to represent this data in a meaningful way to discover common patterns in popular songs as well as comparing how the top songs vary between the two platforms. By representing important traits, like BPM and key, and allowing for filtering and sorting by these traits, this project could also serve as a tool to help add popular songs to DJ mixes.

Data

The datasets used to create this visualization contains the most popular songs from Spotify and Tiktok for 2022. The following datasets were used for the majority of our project. These sources were found on kaggle and acquired from data scraping Spotify playlists/charts using BeautifulSoup.

- https://www.kaggle.com/datasets/sveta151/tiktok-popular-songs-2022
- https://www.kaggle.com/datasets/sveta151/spotify-top-chart-songs-2022

We expanded our dataset by using SpotifyAPI to gather more/missing information about the songs using the unique ID's provided by our initial datasets.

Data we will be working with contains the most popular songs from Spotify and TikTok. The following attributes will be used to build the visualization:

- Song Information – Song, artist(s), and release date, cover album, preview song link,
- Danceability – based on a combination of elements including tempo, rhythm, and stability.
- Energy – represents a perceptual measure of intensity and activity.
- Key - Standard Pitch Class Notation.
- Loudness- overall amplitude of decibels (dB).
- Speechiness – presence of spoken words in a track.
- Acousticness – a confidence measure of whether a track is acoustic.
- Liveness – detects the presence of an audience.
- Valence – measuring the positiveness of a track.
- Tempo – overall estimate of a track in beats per minute.
- Duration- length of track in milliseconds.

As datasets are sourced from Kaggle, we need to create consistent names of columns so they can be used more easily in website functions. For this we used R to go through them all and create consistent names, combine into single dataframes (when applicable) and then output as JSON format.

Most attributes remain in the different datasets but some are needed to be added or converted. Most importantly, the musical "chord" value in letter notation from dataset 1 had to be converted to a number in the camelot naming convention. See Table 1 below for how these values are converted. Additionally a boolean value "spotify" is added to identify if the song was on the spotify list (True) or the TikTok list (false). Finally, a "chartYear" attribute is added to identify what year the dataset is for. See Table 2 for a tabular description of each of the 3 final datasets and the data/variables they contain.

Table 1: Camelot Conversion

Letter Key	**Camelot Number**
A	11
D	10
E	12
B	1
F#	2

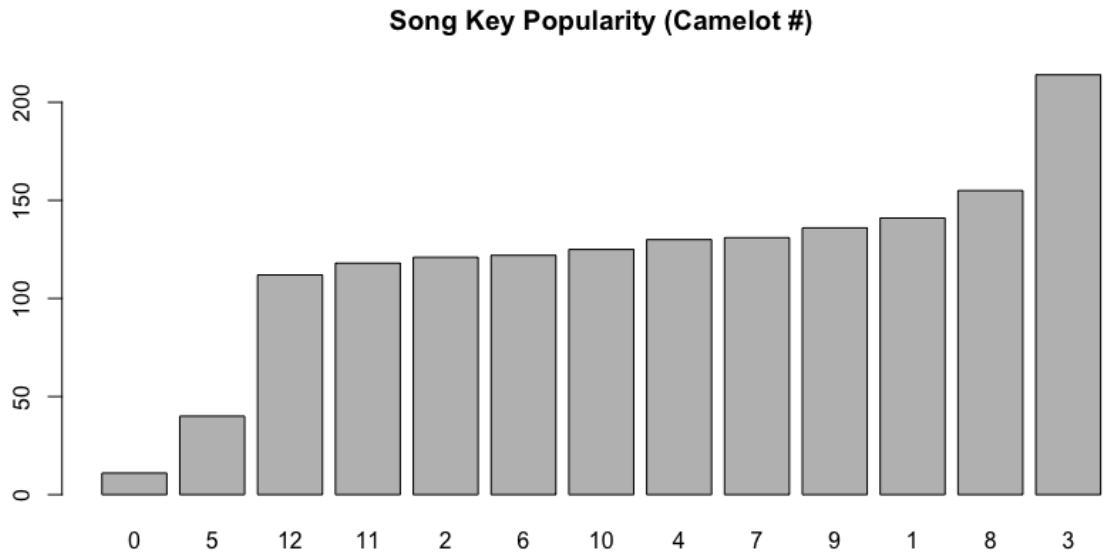
C#	3
G#	4
D#	5
A#	6
F	7
C	8
G	9

Table 2 - Dataset Descriptions and Variables

Dataset 1	Dataset 2	Dataset 3
Spotify Top Songs 2021-2020	Spotify Top Songs 2022	TikTok Viral Songs 2022-2019
acousticness	**acousticness**	**acousticness**
artist	**artist**	**artist**
		album
bpm	**bpm**	**bpm**
camelot	**camelot**	**camelot**
chartCount	chartCount	
chartMax	chartMax	
chartWeek		
chartWeekCount		
chartYear	**chartYear**	**chartYear**
chord		
danceability	**danceability**	**danceability**
duration	**duration**	**duration**
energy	**energy**	**energy**
followers		
genre		
	instrumentalness	instrumentalness
liveness	**liveness**	**liveness**
loudness	**loudness**	**loudness**
	mode	mode
name	**name**	**name**
popularity		popularity
released		
	signature	signature
songID		
speechiness	**speechiness**	**speechiness**
spotify	**spotify**	**spotify**
streams		
		trackPopularity
	url	
valence		valence

Exploratory Data Analysis

While cleaning up the data, some initial exploration was done to see how some of the values may appear in the final project. First, we wanted to see if there would be much variation in song key. This can be seen in the bar chart below which shows that the key 5 is very unpopular and 3 is more popular with the other keys being roughly the same in popularity:



Additionally, there is a lot of variation in the bpms of the songs in the first dataset. There are some wild outliers but then the majority of the songs are in a fairly standard range that DJs will play in. See this statistics breakdown here:

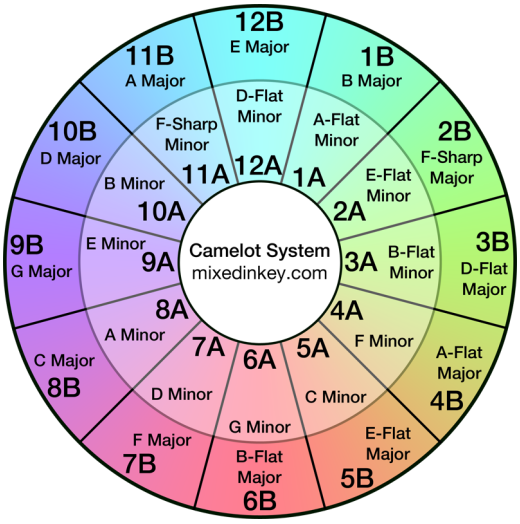
```
     Tempo
Min.   : 46.72
1st Qu.: 97.96
Median :122.01
Mean   :122.81
3rd Qu.:143.86
Max.   :205.27
NA's   :11
```

These initial investigations show that there will be some interesting variation available in the data that we will explore more in the project.

Design Evolution

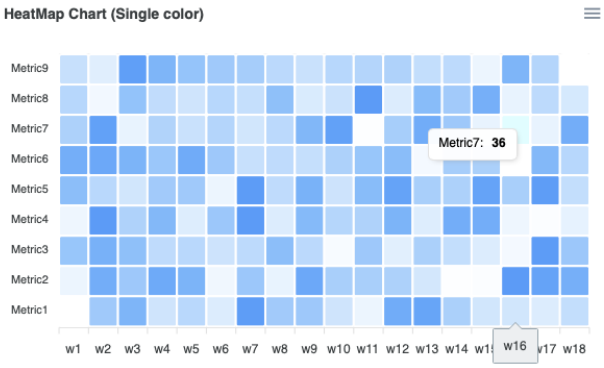
“Main Graph” - We will have a scatterplot that will encode songs by their number of streams (x-axis) and via selectable attribute (y-axis) from the list in the Data section. The data points will be color coded based on if they come from the Spotify dataset, the TikTok dataset or appear in both. This plot will have brushing available to filter the other visualizations.

“Camelot wheel” - This wheel (left) is a method used by DJs to categorize songs by key and mode and is often used in DJ-ing software like Serato or Rekordbox. We plan on implementing a circular bar chart (right) to represent the frequency of each key of popular songs. The number of songs in each key will be encoded by the height of the bar in each section. We will use the same colors for the keys since those are standardized for Camelot notation and are used in DJ software.



Based on the brushed selection from the main graph, or defaulting to the entire dataset, the last two visual elements will change to show the top 10 songs and their attributes.

“Heatmap” - we will encode the various attributes of the top ten songs in a heatmap similar to the one shown below:

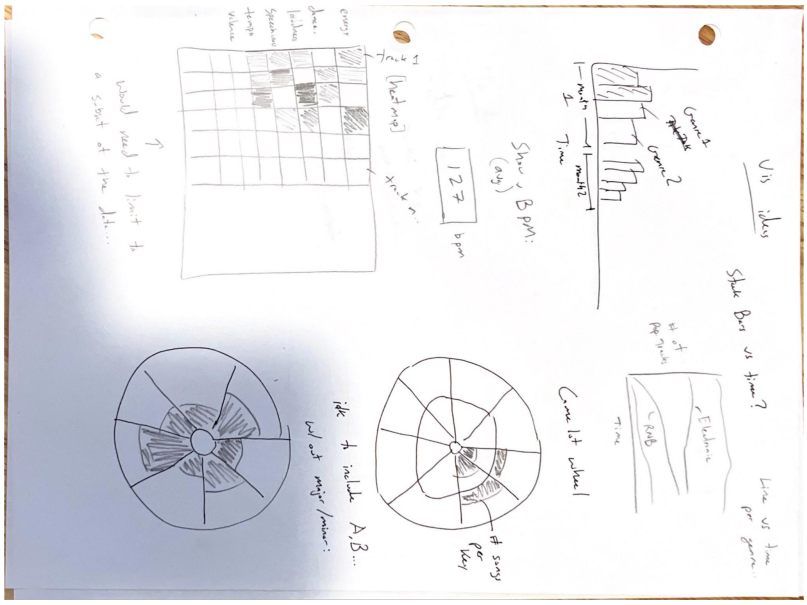


“Top 10 List” - We will have a table showing the top 10 songs in a selection with the album artwork, artist name and song name.

We arrived at these main visualizations after doing some brainstorming on what charts would be possible. We were trying to figure out if we can arrange one chronologically by when the song was most popular (the weeks it was on the spotify list). This was a struggle to conceptualize how the time a song was popular would be easily visualized with other attributes especially considering a large size data set.

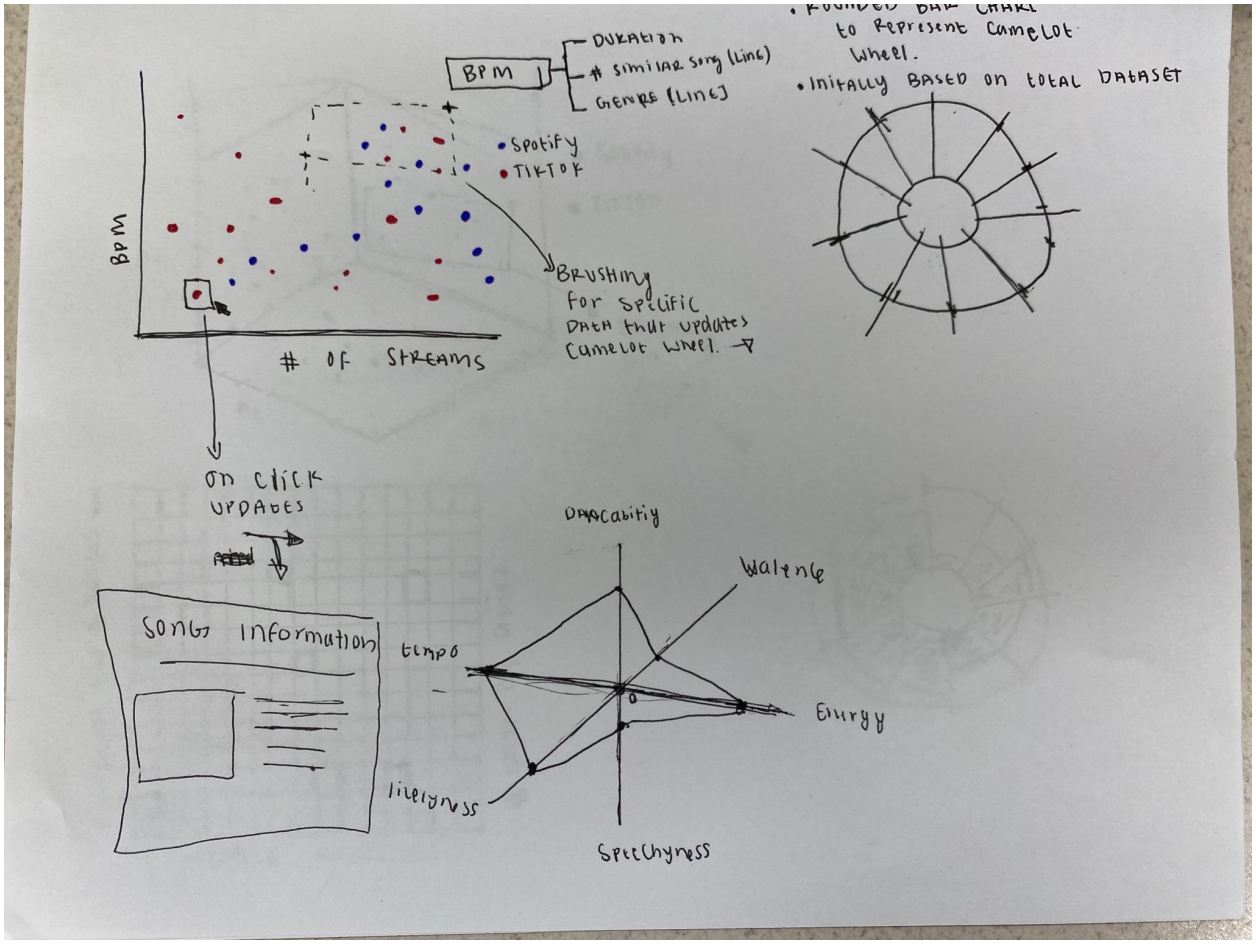
1.1 Brainstorm

Trying to show counts of different genres over time via bars or lines. Showing the average BPM overall. Heatmap and Camelot Ideas roughed out:



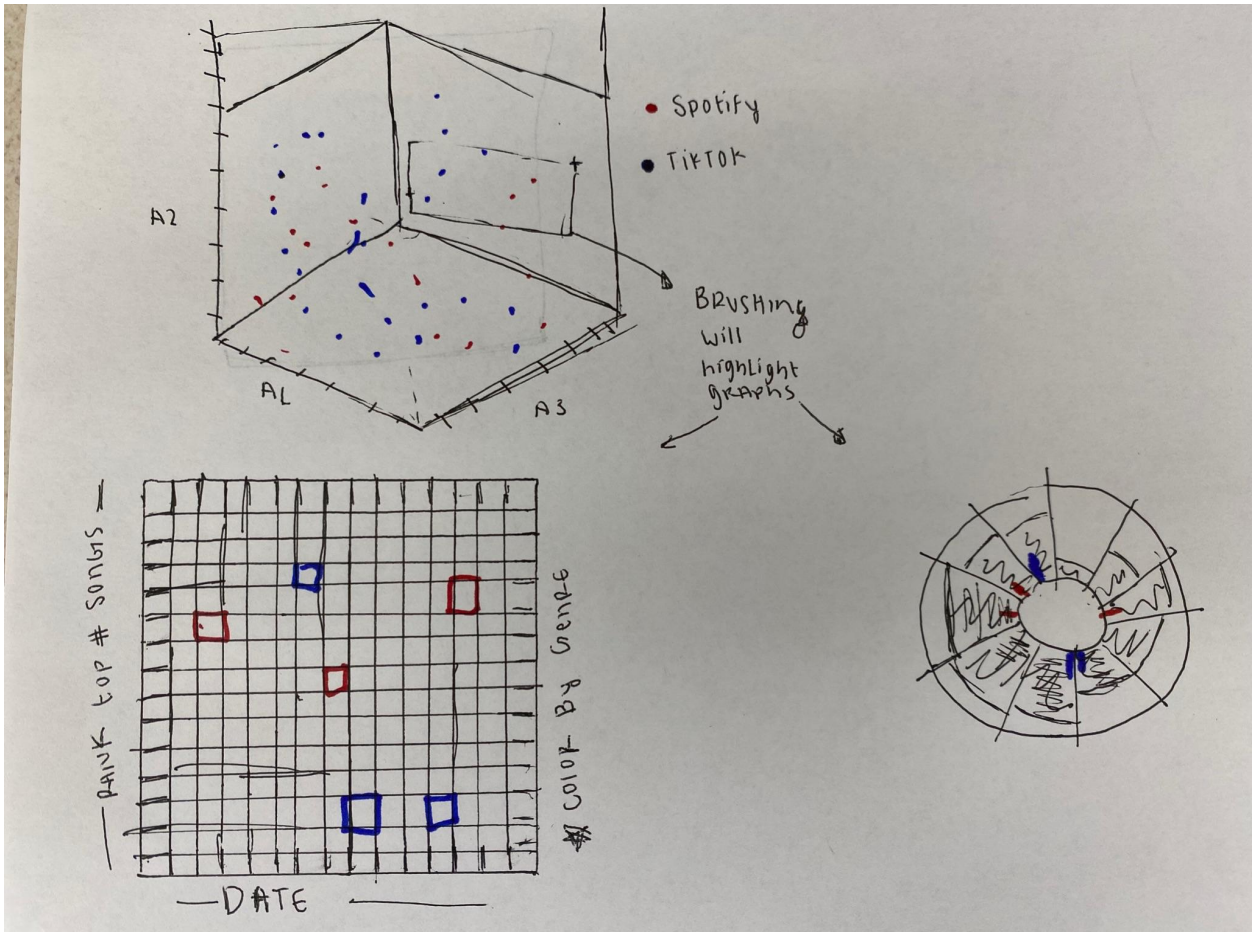
1.2 Prototype 1

The scatter plot compares two different attributes (EX: Tempo, # of Streams), these attributes can be changed via a dropdown menu. The color of the encodes whether the data is from tiktok or spotify. Selecting an individual scatterplot node will execute a tooltip with all of the song/artist information and also trigger a spatial graph that represents the nodes’ audio characteristics.

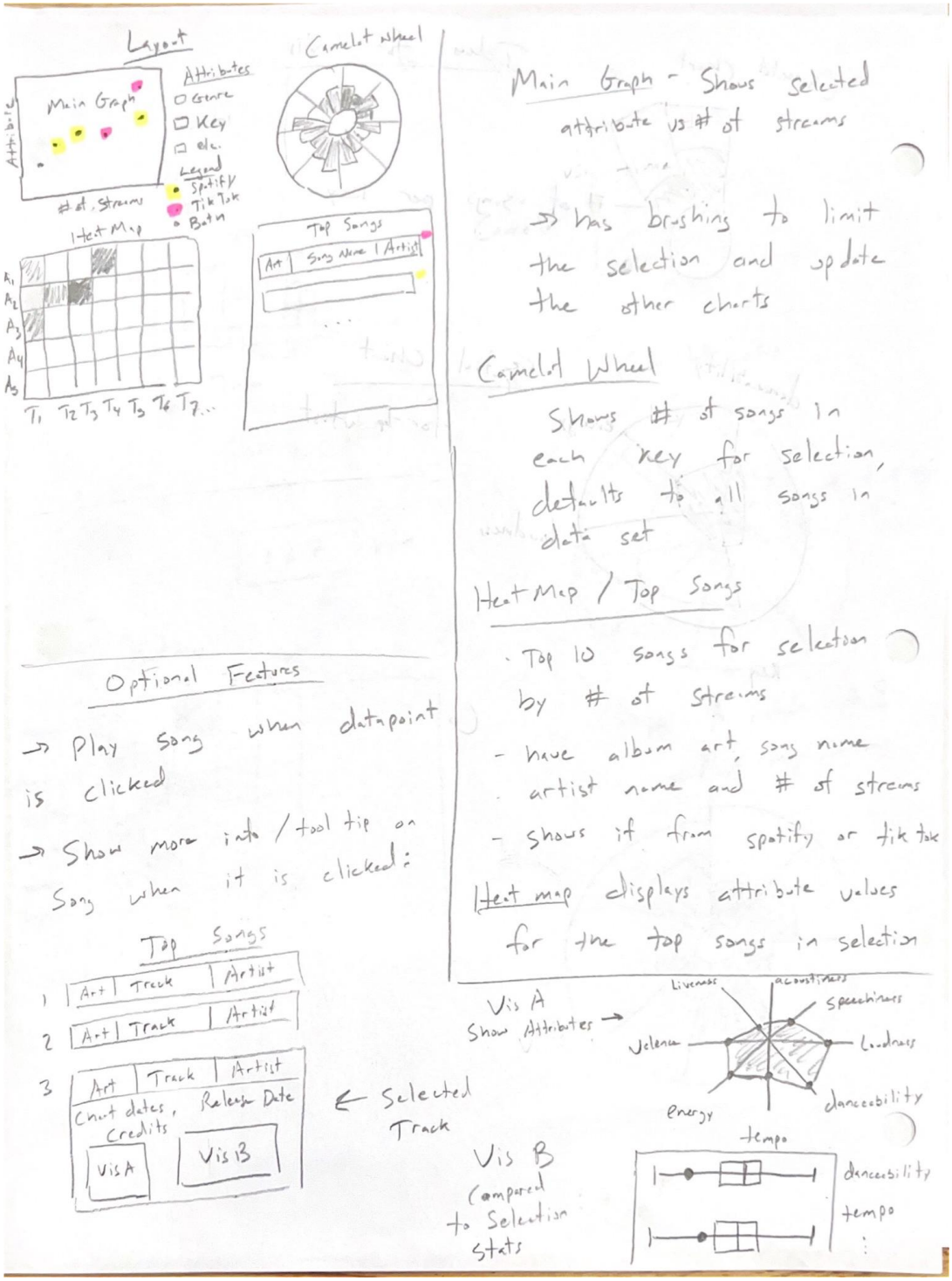


1.3 Prototype 2

Contains a 3D scatter plot that compares three different attributes. Similarly to prototype 1 the marks encode which dataset the variables are from. There is a heat map that contains the highest ranked songs within a timeframe the cells will be color coded to represent the genre type. We want to implement a brush that will update the surrounding graphs with the contained data variables.


Spotify
TikTok
A2
AL
A3
BRUSHING will highlight graphs
RANK TOP # SONGS
COLOR BY GENRE
DATE

1.4 Final Design



1.5 Features to Include

Must-have features

This project will include the following features

- An interactive scatterplot with selectable attributes to plot and brushing that will update the remaining visual elements
- A visual representation of how many songs are in each key in the selection shown on a Camelot wheel
- Top 10 List of songs in the selection based on number of streams

Optional Features

One feature that we hope to implement into our visualization is connecting to Spotify API to display additional information about the featured song. The information could include album artwork, song credits,artist information. We would also like to implement a function that allows the user to play the song from Spotify when a data point is clicked.

Another possible feature could be to expand on a song in the Top 10 list in order to show a detailed view with some additional visualizations comparing its attributes to the others in the list.

Peer Review

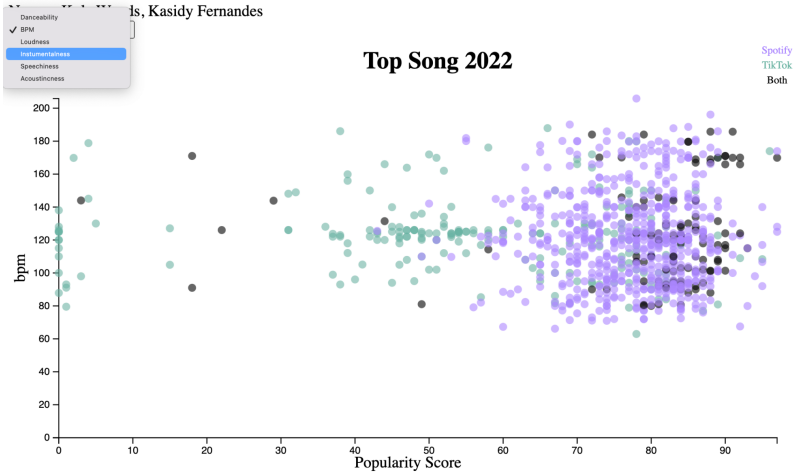
To get feedback on the project, we met with classmates Alan Bird and Eric Sims and presented the motivation and direction for our project. After presenting the concept for the project, they were excited by the approach and the data. The main question they had was related to how the interaction would work via the “main graph” (the scatterplot of attributes that allows brushing and will update the other elements with the selections). This made it clear to us that we will need to highlight how this is intended to work in some way for the final version of the project.

Another suggestion was to include some additional filtering methods to drill down the list of songs even further which should be implemented in a few ways. This especially makes sense to add a dedicated slider for the bpm of the songs in the selection to easily dial that into a range. One aspect that we are not planning on implementing but could be seen as a nice feature for this type of project would be a keyword search option to search through song names or artists. The implementation for a keyword search would be beyond the scope of our project as we intend to focus more on the visual representation of the songs’ attributes.

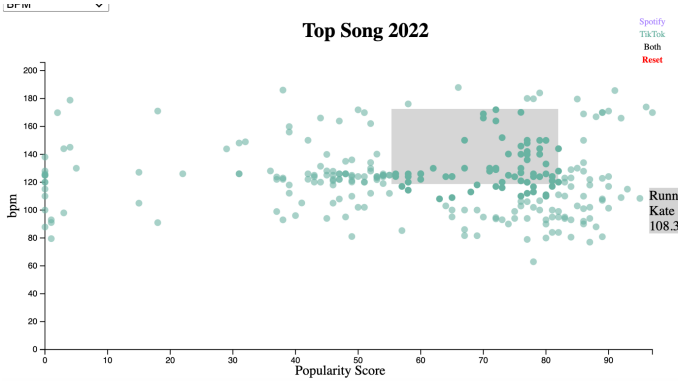
Implementation

1.1 Scatter Plot

The First chart is a scatterplot that allows the most interactivity for the user. It allows brushing that interacts with both the table and the camelot plot. The x-axis represents the popularity of a track and the y-axis represents audio attributes. These audio attributes can be changed by a drop down menu located at the top left of the page.

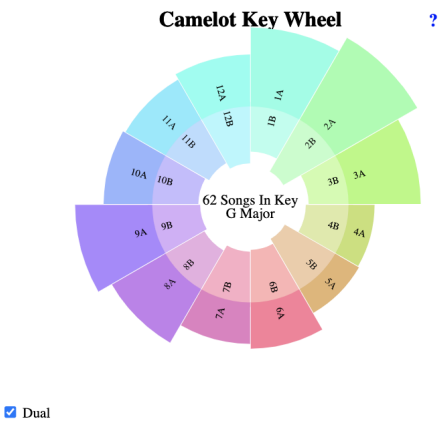


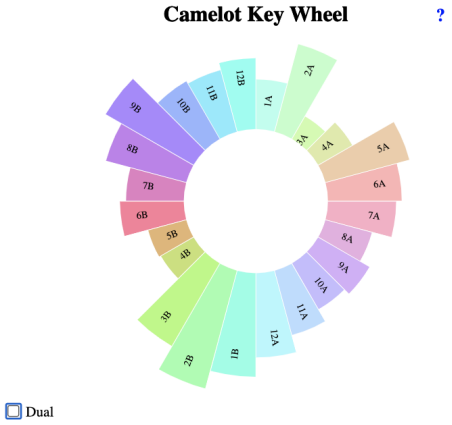
The legend of the chart is interactive and allows the user to view plot points from a specific data set. The plot points also have tooltips and allows brushing to interact with the other charts.



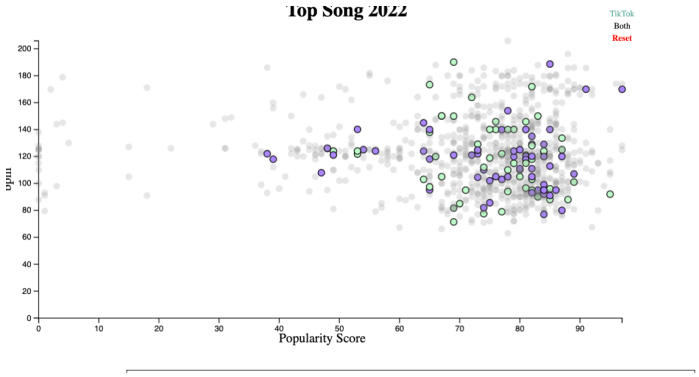
1.2 Camelot Key Chart

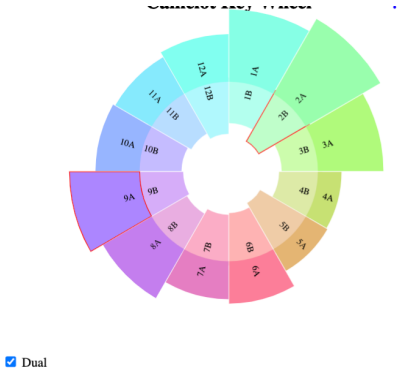
The camelot wheel chart is modeled to represent the actual camelot wheel. We created a dual axis circular bar chart to represent the number of occurrences for each note within the dataset. There is a button located in the bottom left area of the chart that allows the user to toggle the chart between a dual axis and a single axis view.



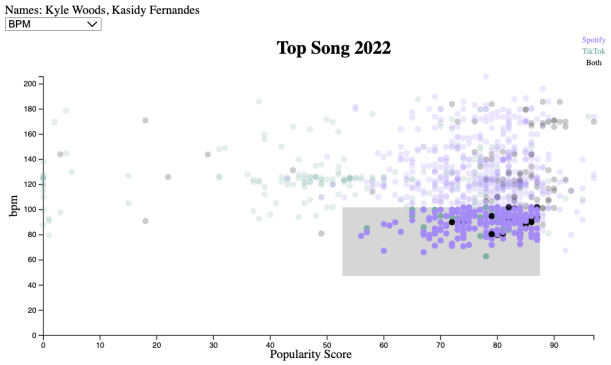


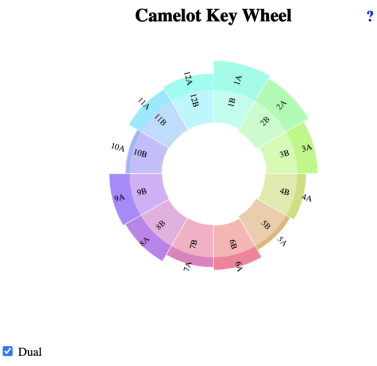
Hovering over each bar updates a tooltip located in the center which prints the specific key and count of each note. Clicking each bar interacts with the main chart by highlighting where the data points are relative to the scatterplot points.





Brushing the Scatterplot graph also updates the values of the Camelot Wheel.

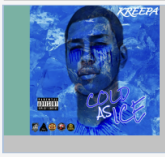
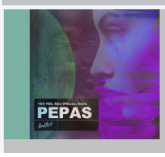




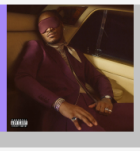
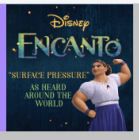
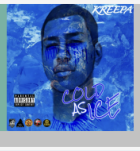
1.3 Table

To provide more “details on demand” there is a table showing the album artwork, release date, camelot value and selected attribute:

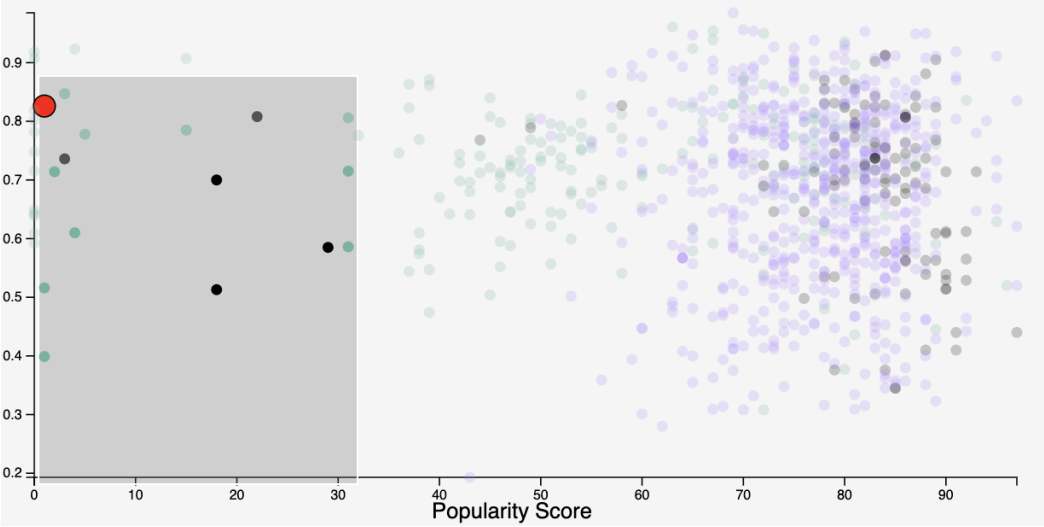
Song Details and Sorting

Pop.	Track	bpm	Cam.
58	 **The Plan** Artist: Arman Aydin Release Date: 7/8/2022	124.034	6A
58	 **Oh No** Artist: Kreepa Release Date: 12/13/2019	176.123	12A
58	 **Pepas** Artist: YES YES Release Date: 7/15/2022	125.973	8B

For more investigation into the data, the popularity, track, attribute, or camelot values can be sorted. Clicking on the header will sort the selection and append an arrow indicating the sorting to the header text:

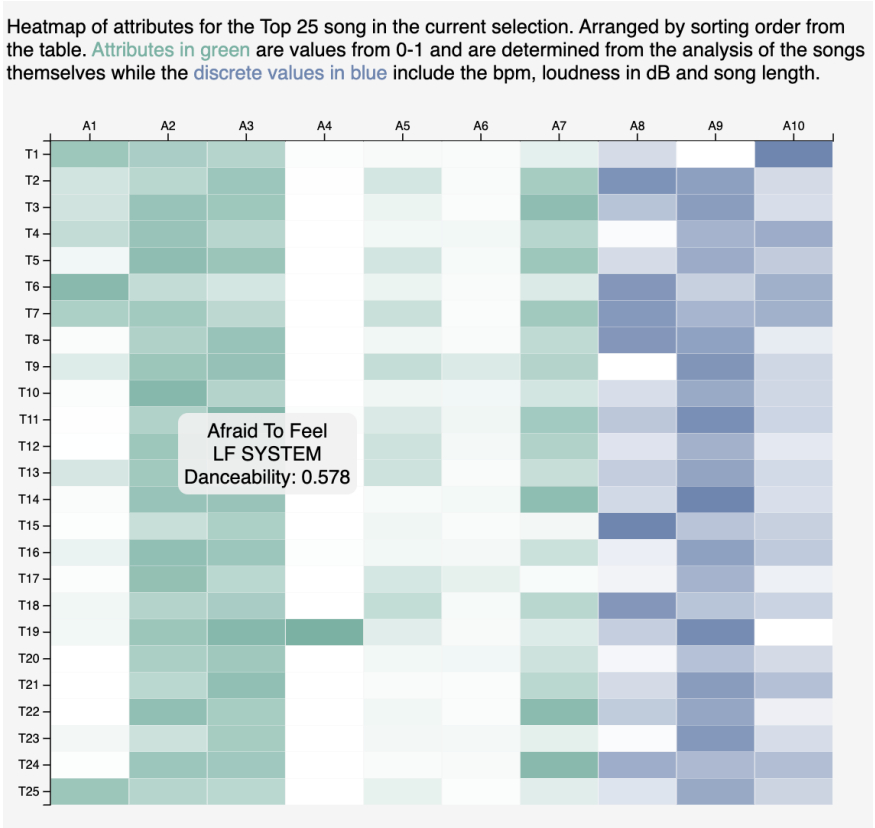
Pop.	Track	bpm ⬆
55	 **GOLD STACKS** Artist: Future Release Date: 4/29/2022	181.856
55	 **Surface Pressure - From "Encanto"/Soundtrack Version** Artist: Jessica Darrow Release Date: 1/27/2022	180.042
58	 **Oh No** Artist: Kreepa Release Date: 12/13/2019	176.123
50	 **Good Morning** Artist: P.S. Release Date: 3/5/2021	171.907

Finally, when a track is hovered over in the table, it's corresponding circle in the scatterplot highlights to indicate where the song is in relation to the others:

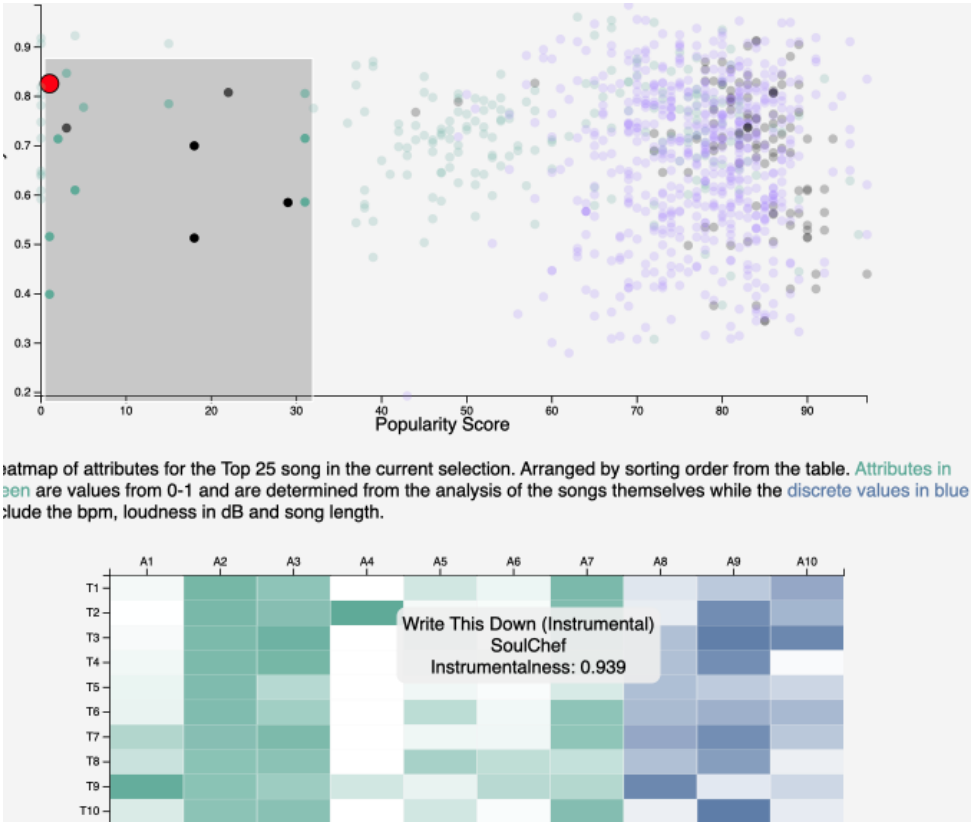


1.4 Heatmap

The final element is a heatmap which visually encodes all the attributes of the top 25 songs from the current selection in the table. There are two color scales used here: a white to green one for the 0-1 attribute values and a blue scale for the more discrete values (bpm, loudness and duration). This also employs a tooltip for the specifics of the song and the value of the attribute. There is also a text description above the heatmap providing more detail and explaining the use of the two color scales:



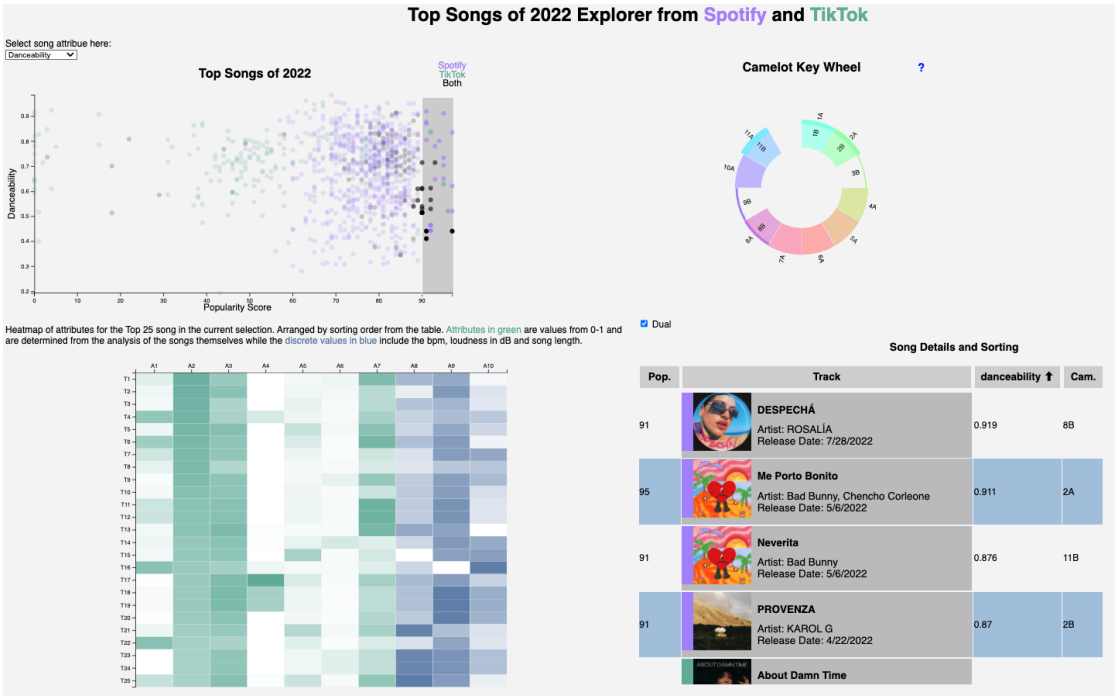
Similar to the Table, the main plot will highlight the song when any of its cells in the heatmap are it is hovered over so that it can be identified with the rest of the data points:



The order of the songs in the heatmap correlates to the sorting applied to the table so the user can investigate the current selection even further.

Evaluation

By using the tool to investigate the songs from 2022, some clear trends are present in the most popular songs. One insight that is clear from using the site, is clear that the most popular songs are generally high in danceability and energy and low in speechiness, liveness, and instrumentalness. Additionally, while there are more songs in specific keys overall, there does not appear to be a stand out key in the most popular songs. Here you can see all songs with 90 or more popularity arranged in ascending order of danceability:



From this, it is clear that the values for energy (A3 or the third column of the heatmap) are also very dark and the attributes A4-A6 are very light. There is also not a single bar on the Camelot wheel that is standing out which would indicate one key being the most popular.

The strength of the tool is in the connected nature of the elements and how they all highlight aspects of the data. We prioritized getting these connected elements and interactions over styling of the page and due to this, the general look of the page could be further improved with more advanced styling and layout tools. Also, due to the complexity and number of possible interactions available (brushing the graph, clicking on the camelot wheel, selecting the Spotify/TikTok songs only and sorting on the table) there are a lot of possible areas to trip out the visualizations and cause bugs. These could be continued to be debugged and streamlined.