Analyzing Spotify’s Top Tracks

Progress Book by
Mattheus Faria
Kristi Prifti

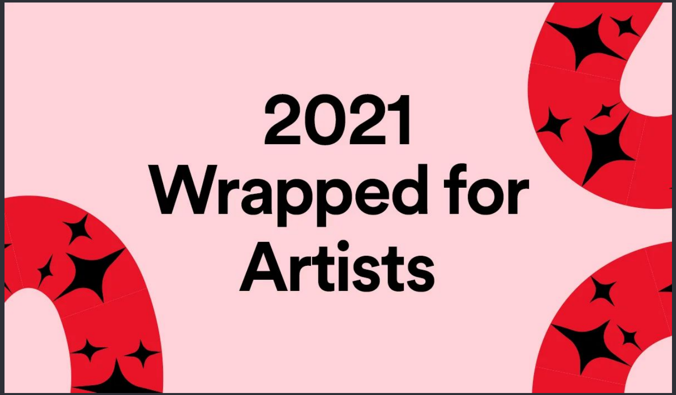
Overview and Motivation

Spotify is one of, if not the most popular music streaming service used today. What many people don’t know is that Spotify has strangely specific data on each song uploaded to its platform. Numerical values ranging from how danceable a song is to how happy the tone of the song is.

We want to use this data on the **top tracks** throughout recent years in order to answer questions on **what makes a song popular.**

Related Work

Every year Spotify releases a user’s listening data through ‘**Wrapped**.’ Around this time, many listeners share their most listened to artists or songs. Spotify’s Wrapped interested us to branch out of a single user’s most popular tracks to the entirety of the platform’s most popular hits.



Questions

There are many questions we can ask that can be answered with this data, all of which form the large question:

“Why is no one listening to my favorite song?”

“Do people dance more to edm or more speechy songs?”

“Should I make my song louder?”

“What makes a song popular?”

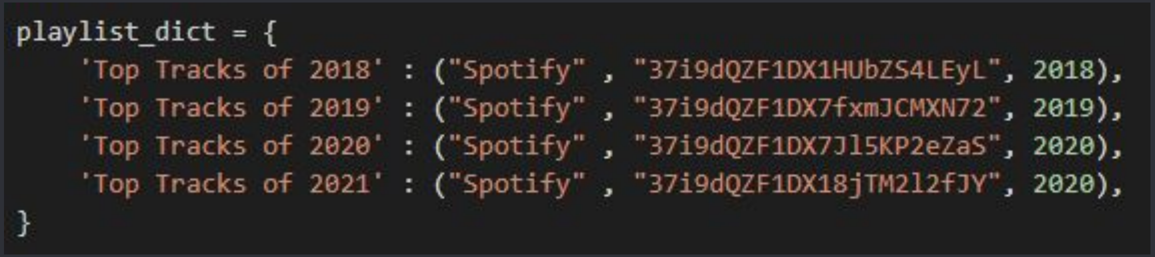
“Are happy or personal songs more listened too?”

“Why do pop songs get so much mainstream radio time?”

“Will people like my song?”

Data Collection

To access and gather the data required to answer these questions, we used Spotify's python library, **Spotipy**, to gather song information, audio information, and their calculated popularity scores.

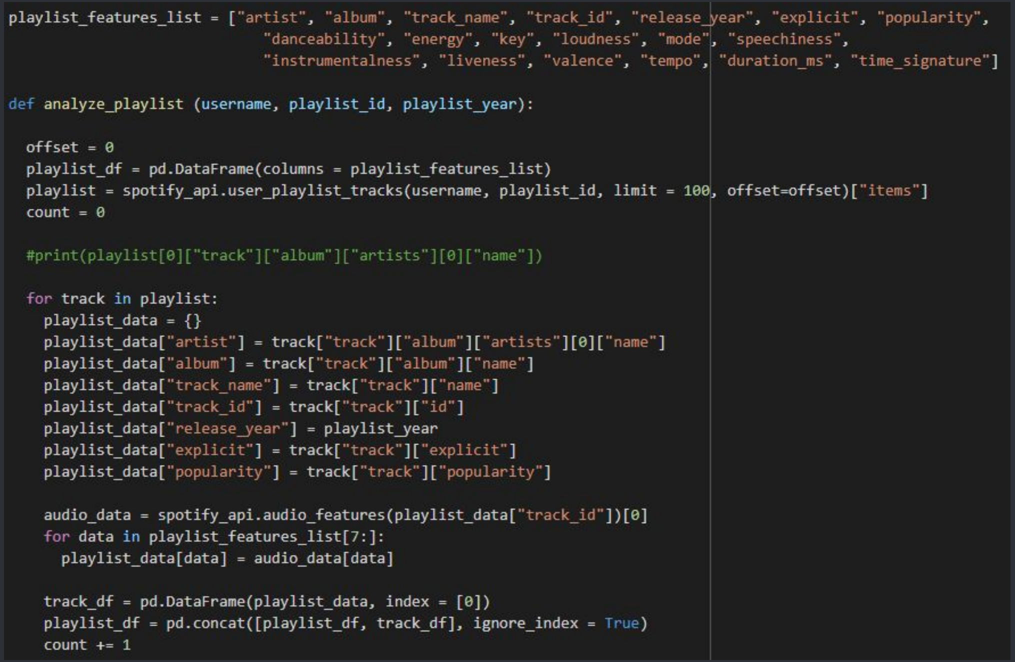


```
playlist_dict = {
    'Top Tracks of 2018' : ("Spotify" , "37i9dQZF1DX1HUbZS4LEyL", 2018),
    'Top Tracks of 2019' : ("Spotify" , "37i9dQZF1DX7fxmJCMXN72", 2019),
    'Top Tracks of 2020' : ("Spotify" , "37i9dQZF1DX7Jl5KP2eZaS", 2020),
    'Top Tracks of 2021' : ("Spotify" , "37i9dQZF1DX18jTM2l2fJY", 2020),
}
```

Check out this notebook for the code involved in gathering this data

https://github.com/MFaria27/final/blob/main/DataCollection.ipynb

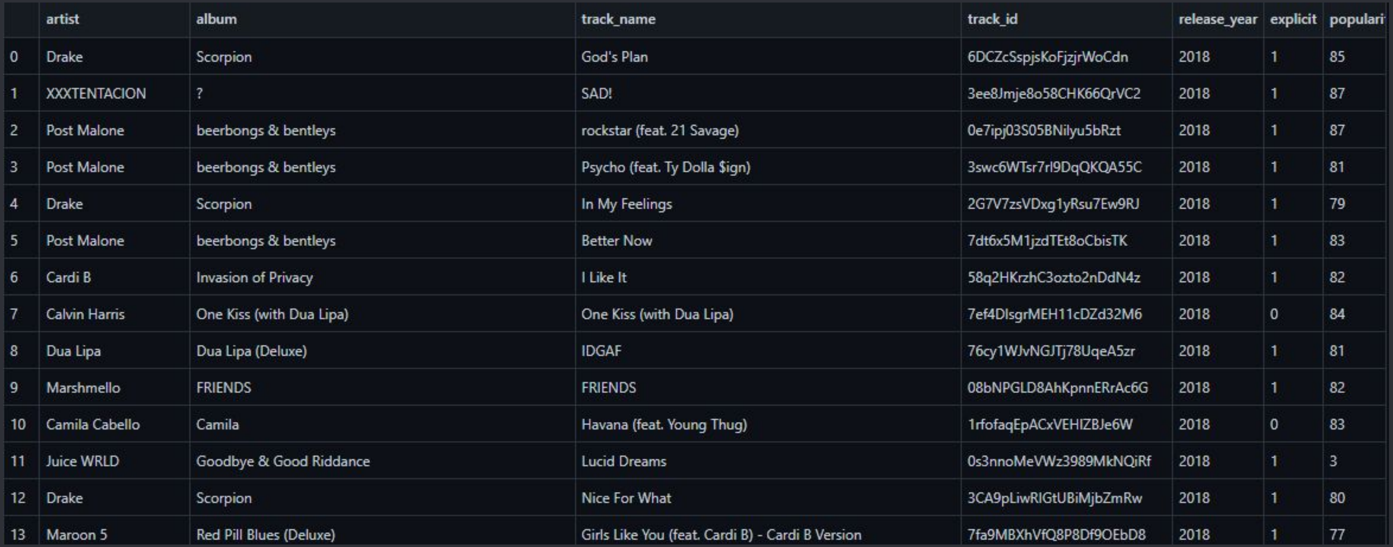
More Data



```
playlist_features_list = ["artist", "album", "track_name", "track_id", "release_year", "explicit", "popularity",
                              "danceability", "energy", "key", "loudness", "mode", "speechiness",
                              "instrumentalness", "liveness", "valence", "tempo", "duration_ms", "time_signature"]

def analyze_playlist (username, playlist_id, playlist_year):

  offset = 0
  playlist_df = pd.DataFrame(columns = playlist_features_list)
  playlist = spotify_api.user_playlist_tracks(username, playlist_id, limit = 100, offset=offset)["items"]
  count = 0

  #print(playlist[0]["track"]["album"]["artists"][0]["name"])

  for track in playlist:
    playlist_data = {}
    playlist_data["artist"] = track["track"]["album"]["artists"][0]["name"]
    playlist_data["album"] = track["track"]["album"]["name"]
    playlist_data["track_name"] = track["track"]["name"]
    playlist_data["track_id"] = track["track"]["id"]
    playlist_data["release_year"] = playlist_year
    playlist_data["explicit"] = track["track"]["explicit"]
    playlist_data["popularity"] = track["track"]["popularity"]

    audio_data = spotify_api.audio_features(playlist_data["track_id"])[0]
    for data in playlist_features_list[7:]:
      playlist_data[data] = audio_data[data]

    track_df = pd.DataFrame(playlist_data, index = [0])
    playlist_df = pd.concat([playlist_df, track_df], ignore_index = True)
    count += 1
```

By loading in a playlist by ID, we can access all the songs and all the information associated with a song (Check the features list for all we can get!)

More Data

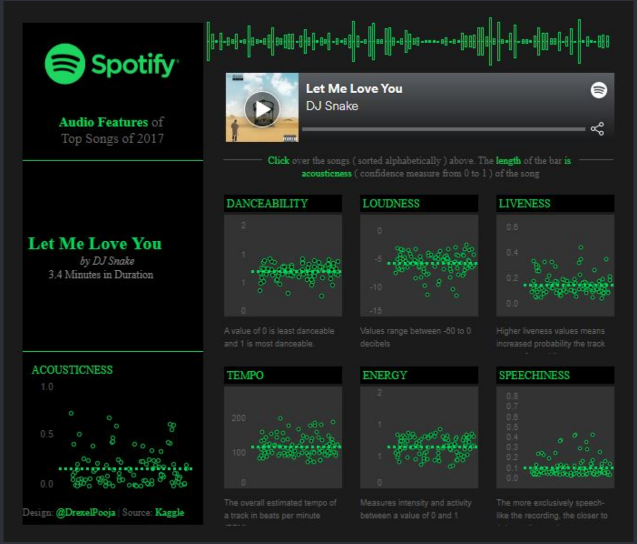


	artist	album	track_name	track_id	release_year	explicit	populari
0	Drake	Scorpion	God's Plan	6DCZcSspjsKoFjzjrWoCdn	2018	1	85
1	XXXTENTACION	?	SAD!	3ee8Jmje8o58CHK66QrVC2	2018	1	87
2	Post Malone	beerbongs & bentleys	rockstar (feat. 21 Savage)	0e7ipj03S05BNilyu5bRzt	2018	1	87
3	Post Malone	beerbongs & bentleys	Psycho (feat. Ty Dolla $ign)	3swc6WTsr7rl9DqQKQA55C	2018	1	81
4	Drake	Scorpion	In My Feelings	2G7V7zsVDxg1yRsu7Ew9RJ	2018	1	79
5	Post Malone	beerbongs & bentleys	Better Now	7dt6x5M1jzdTEt8oCbisTK	2018	1	83
6	Cardi B	Invasion of Privacy	I Like It	58q2HKrzhC3ozto2nDdN4z	2018	1	82
7	Calvin Harris	One Kiss (with Dua Lipa)	One Kiss (with Dua Lipa)	7ef4DlsgrMEH11cDZd32M6	2018	0	84
8	Dua Lipa	Dua Lipa (Deluxe)	IDGAF	76cy1WJvNGJTj78UqeA5zr	2018	1	81
9	Marshmello	FRIENDS	FRIENDS	08bNPGLD8AhKpnnERrAc6G	2018	1	82
10	Camila Cabello	Camila	Havana (feat. Young Thug)	1rfofaqEpACxVEHIZBJe6W	2018	0	83
11	Juice WRLD	Goodbye & Good Riddance	Lucid Dreams	0s3nnoMeVWz3989MkNQiRf	2018	1	3
12	Drake	Scorpion	Nice For What	3CA9pLiwRIGtUBiMjbZmRw	2018	1	80
13	Maroon 5	Red Pill Blues (Deluxe)	Girls Like You (feat. Cardi B) - Cardi B Version	7fa9MBXhVfQ8P8Df9OEbD8	2018	1	77

After gathering all the data and a bit of cleaning, we got a dataset of 250 top tracks from 2018-2021

Exploratory Data Analysis

While looking for ways to best represent this dataset, we came across this Tableau visualization by **Pooja Gandhi** on the top songs of 2017, so we used the grid of interactable charts as inspiration for our visualizations



Let Me Love You
DJ Snake

Audio Features of
Top Songs of 2017

Let Me Love You
by DJ Snake
3.4 Minutes in Duration

Click over the songs (sorted alphabetically) above. The **length** of the bar **is** **acousticness** (confidence measure from 0 to 1) of the song

A value of 0 is least danceable and 1 is most danceable.

Values range between -60 to 0 decibels

Higher liveness values means increased probability the track

The overall estimated tempo of a track in beats per minute

Measures intensity and activity between a value of 0 and 1

The more exclusively speech-like the recording, the closer to

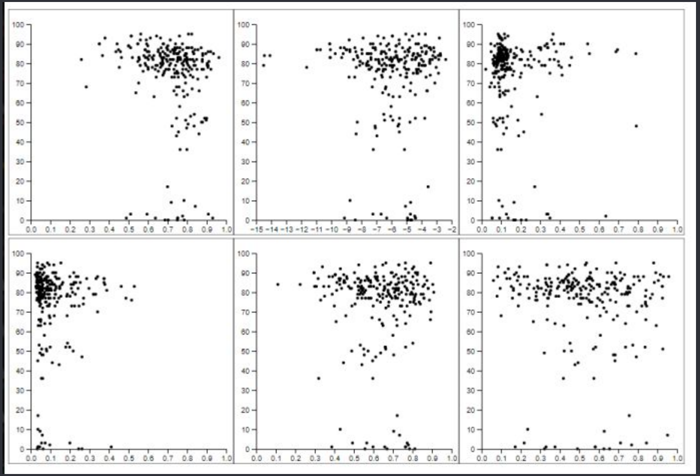
Design: @DrexelPooja | Source: Kaggle

https://www.tableau.com/community/music/spotify-top-tracks

Design Evolution

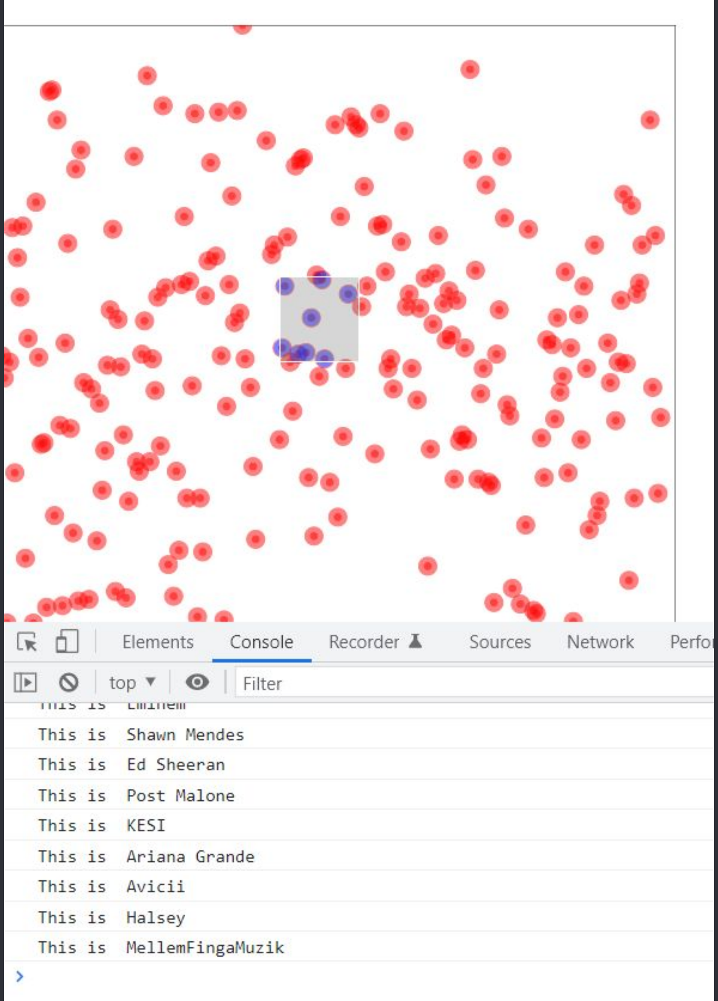
We originally had a dataset of about 10000 songs, but that was generally an dataset of random songs, so we decided to take just the top tracks in order to make future visualizations more readable.

Even with this change, having 250 songs per chart made it very clustered, making intractability a little useless.



Design Evolution

We also tried experimenting with brushing. It was an interesting idea after learning about it in class, but after trying to implement it, we realized that there would not be much use in creating these brushes; other than maybe showing average information in an area.



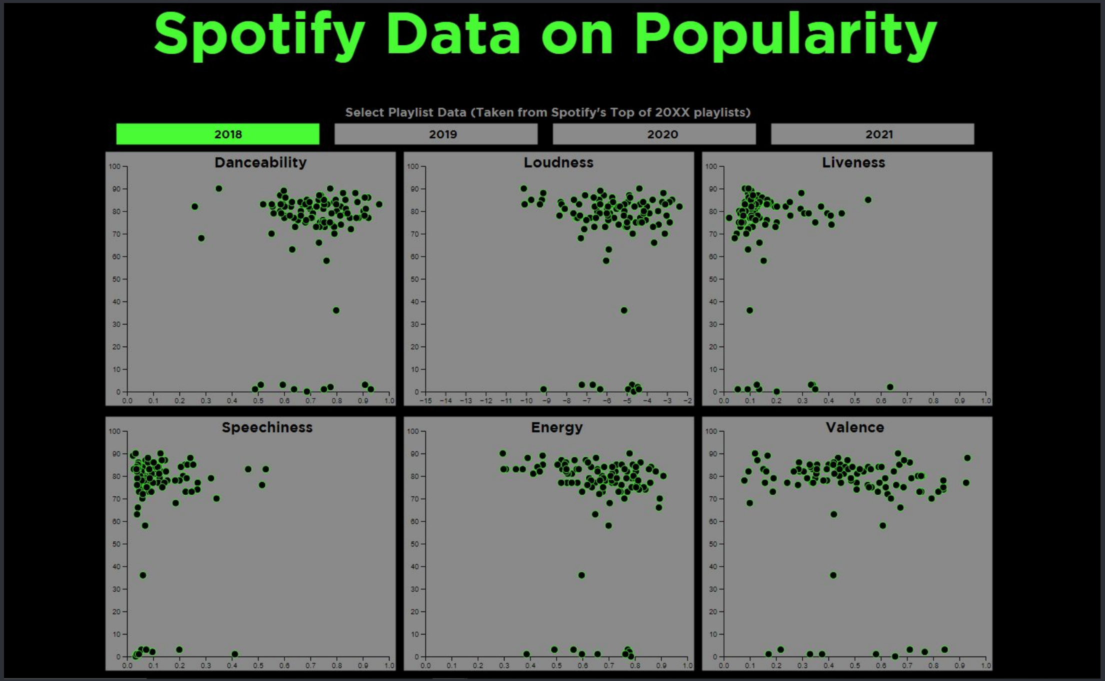
```
This is  Eminem
This is  Shawn Mendes
This is  Ed Sheeran
This is  Post Malone
This is  KESI
This is  Ariana Grande
This is  Avicii
This is  Halsey
This is  MellemFingaMuzik
```

Design Evolution

This would be our final design.

Some interesting design choices:

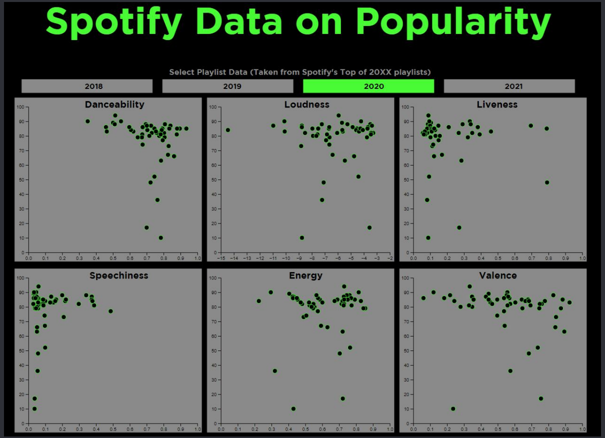
- For colors, we decided to follow Spotify's color aesthetic of green and black
- The font used in all texts is Gotham Bold Regular, the same font Spotify uses



Implementation/Technical Achievements

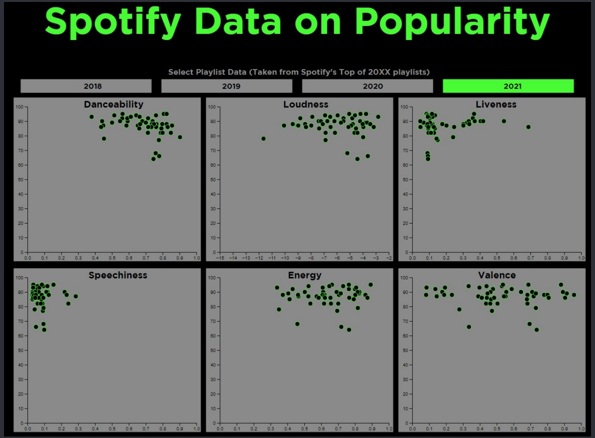
Year selection

- To show the change in years, each playlist we scraped has its own set of graphs, which can be selected by clicking on the year buttons
- A nice design choice: the current graph is shown by highlighting the year button in green.



Spotify Data on Popularity

Select Playlist Data (Taken from Spotify's Top of 20XX playlists)



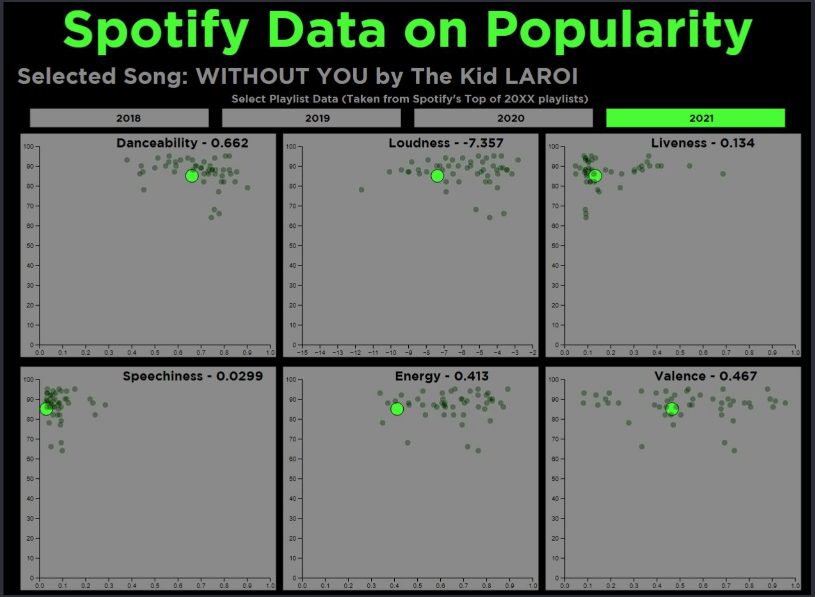
Spotify Data on Popularity

Select Playlist Data (Taken from Spotify's Top of 20XX playlists)

Implementation/Technical Achievements

Song selection

- Clicking on a circle displays the song that dot represents, updating a message at the top of the page
- It also highlights where that song is on every other graph and displays what score that song has in each graph
- Design: The selected dot has its size increased and every other point has its opacity lowered
- Clicking the blank graph space or switching the graph will reset all the dots to be the same size and full opacity



Spotify Data on Popularity

Selected Song: WITHOUT YOU by The Kid LAROI

Select Playlist Data (Taken from Spotify's Top of 20XX playlists)

Evaluation

Taking a look at each graph throughout the years, we can see that all the top songs have the following similarities:

- High energy
- Generally Louder
- Very “danceable”
- All show that how “happy” a song is does not really contribute to its popularity

Evaluation

In the end, we pretty much achieved what we wanted with these visualizations, and the technical and design achievements made to make it properly work were almost exactly what we imagined we would make

We could improve the visualization by adding more songs other than the top tracks to show variation, and we could have added a player similar to the visualization we took inspiration of just to add to the “music data” aesthetic