



**STEVENS**  
INSTITUTE of TECHNOLOGY  
THE INNOVATION UNIVERSITY®

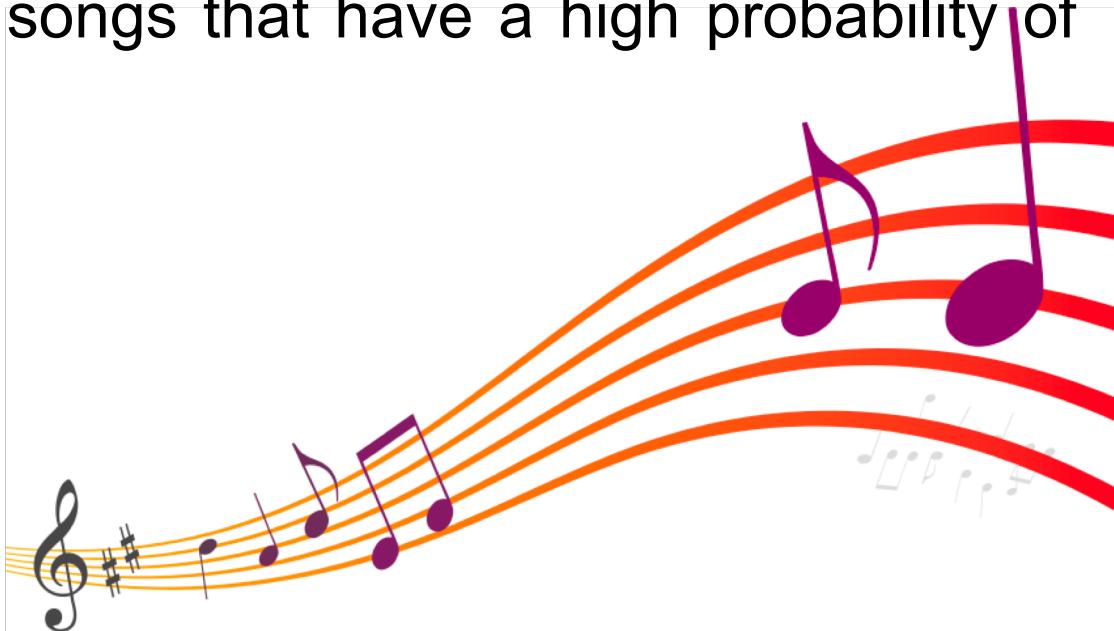
# Song selection based on your favorite artist

Pranav Prajapati  
BIA 652 - D  
Professor Amir Gandomi



# Motivation

- Everyone has their personal favorites when it comes to music. We often find it difficult or less confident when exploring new songs.
- Find songs that are like your favorite artist's style, from a playlist of random songs.
- You get a fresh list of songs that have a high probability of likeness.





# Scraping the data



- Spotify's API provides something called ***audio features***, which are, as the name implies, features or characteristics of a song.
- We used **Spotipy** , a lightweight Python library for the Spotify Web API to run our scripts for extracting these features.
- So according to our preferences 10 features we extracted for each song of that particular artist.
- There were total 219 songs and 1000 more random playlists songs data from 2018.

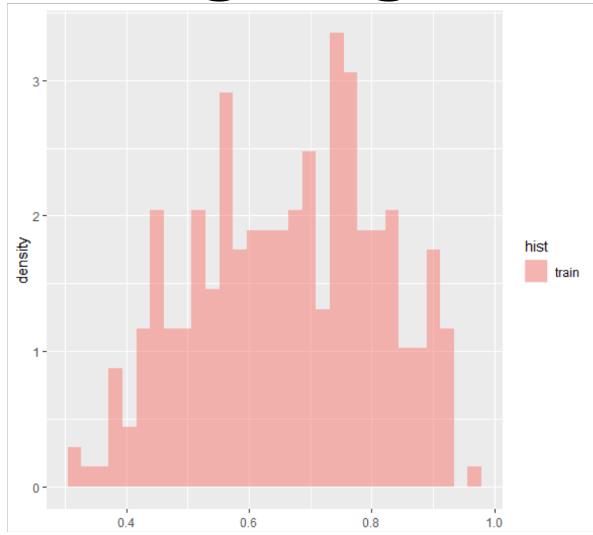


# Data Cleaning

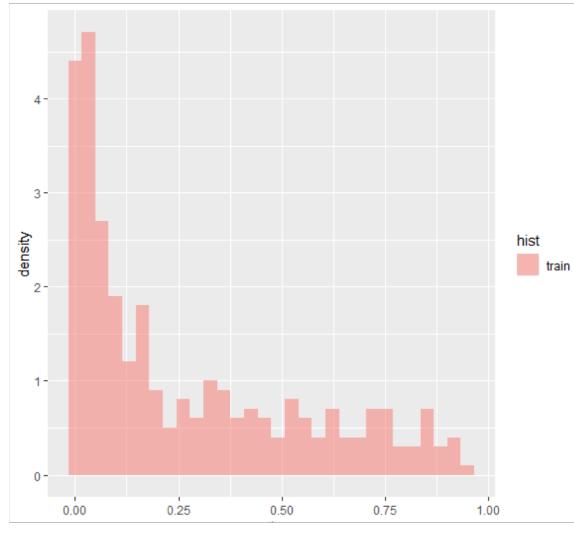
- Since the data set was scraped, there was minimal need for data cleaning.

	name	album	artist	popularity	danceability	acousticness	energy	instrumentalness	liveness	loudness	speechiness	tempo	class
0	0 To 100 / The Catch Up	0 To 100 / The Catch Up	Drake	45	0.559	0.737000	0.716	0.000000	0.2520	-3.964	0.3900	176.618	1
2	Back To Back	Back To Back	Drake	41	0.893	0.008540	0.480	0.000000	0.1160	-3.728	0.3560	86.976	1
6	Charged Up	Charged Up	Drake	19	0.457	0.910000	0.694	0.000000	0.1210	-7.018	0.6080	177.777	1
12	Fake Love	Fake Love	Drake	4	0.923	0.161000	0.535	0.000000	0.1030	-7.353	0.3010	133.942	1
18	Headlines	Headlines	Drake	22	0.650	0.357000	0.583	0.000316	0.0909	-7.390	0.0948	152.087	1
24	I'm Upset	I'm Upset	Drake	77	0.899	0.263000	0.592	0.000000	0.0831	-7.901	0.3330	149.953	1
25	You & The 6	If You're Reading This It's Too Late	Drake	28	0.574	0.166000	0.476	0.000000	0.1290	-7.444	0.2820	154.531	1
26	Wednesday Night Interlude	If You're Reading This It's Too Late	Drake	29	0.321	0.729000	0.466	0.000008	0.2150	-8.163	0.0512	135.089	1

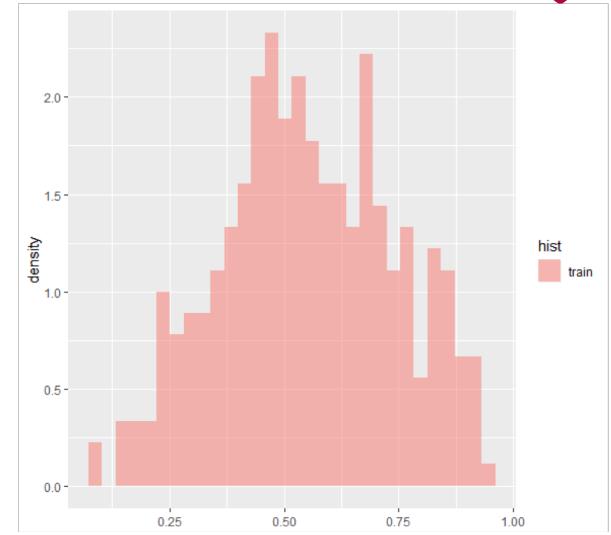
# Assigning class



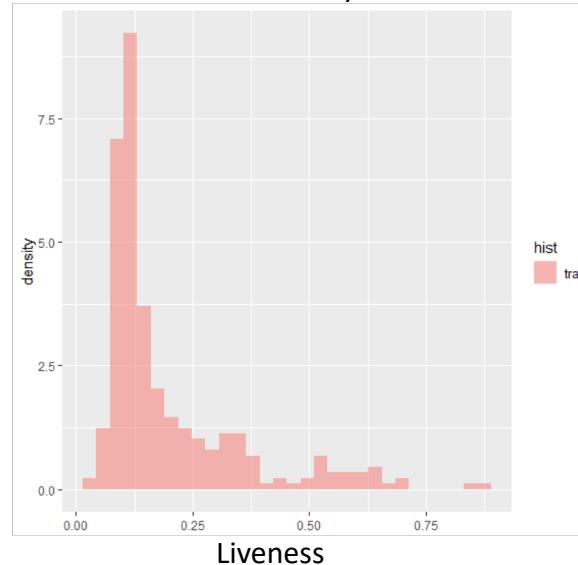
Danceability



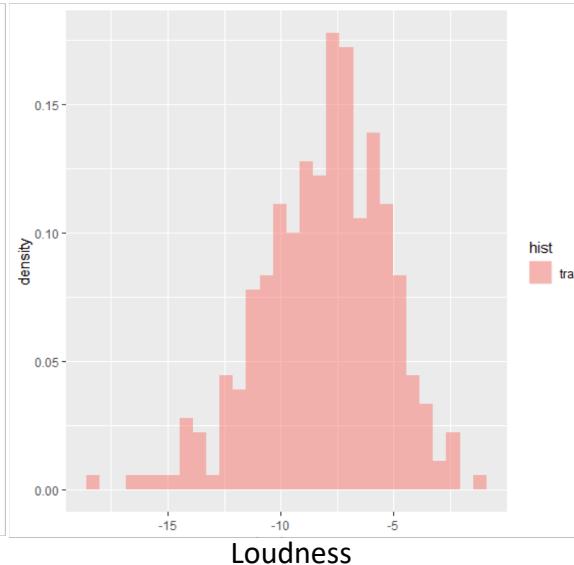
Acousticness



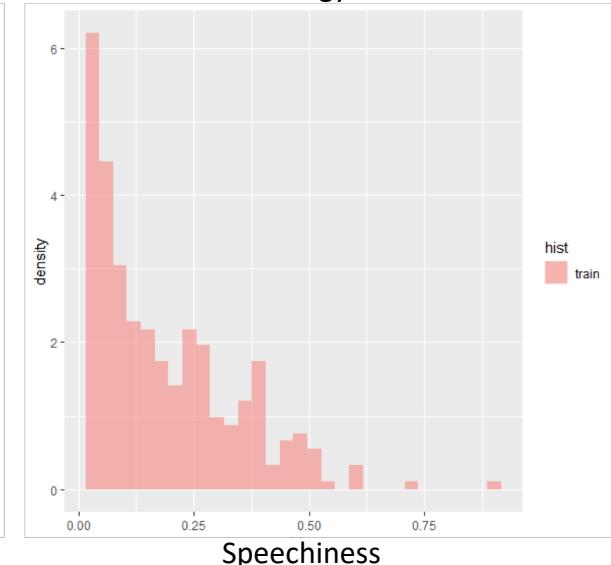
Energy



Liveness



Loudness



Speechiness

# Dimension Reduction

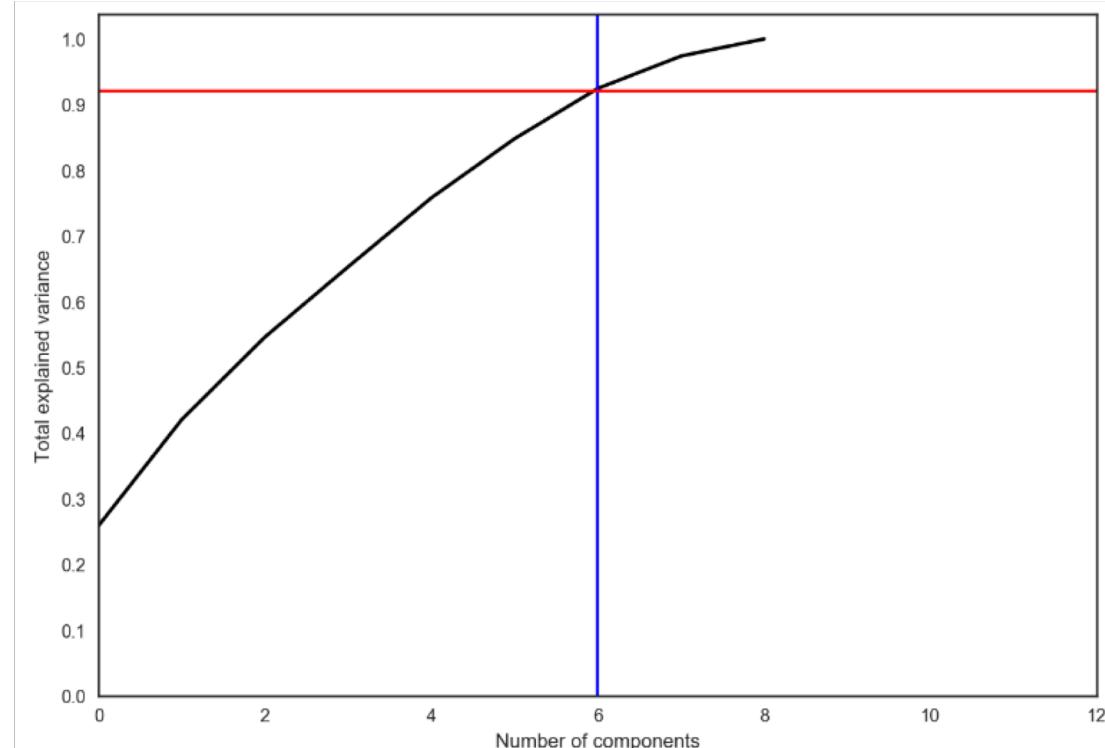
- Dropped unimportant columns for training from data set.

```
ignore = ('artist',
```

```
'album', ]) train_data = train_data.drop(ignore, axis= 1)
```

- Performed PCA to reduce the number of variables

- Variance of 92% corresponded to 6 components





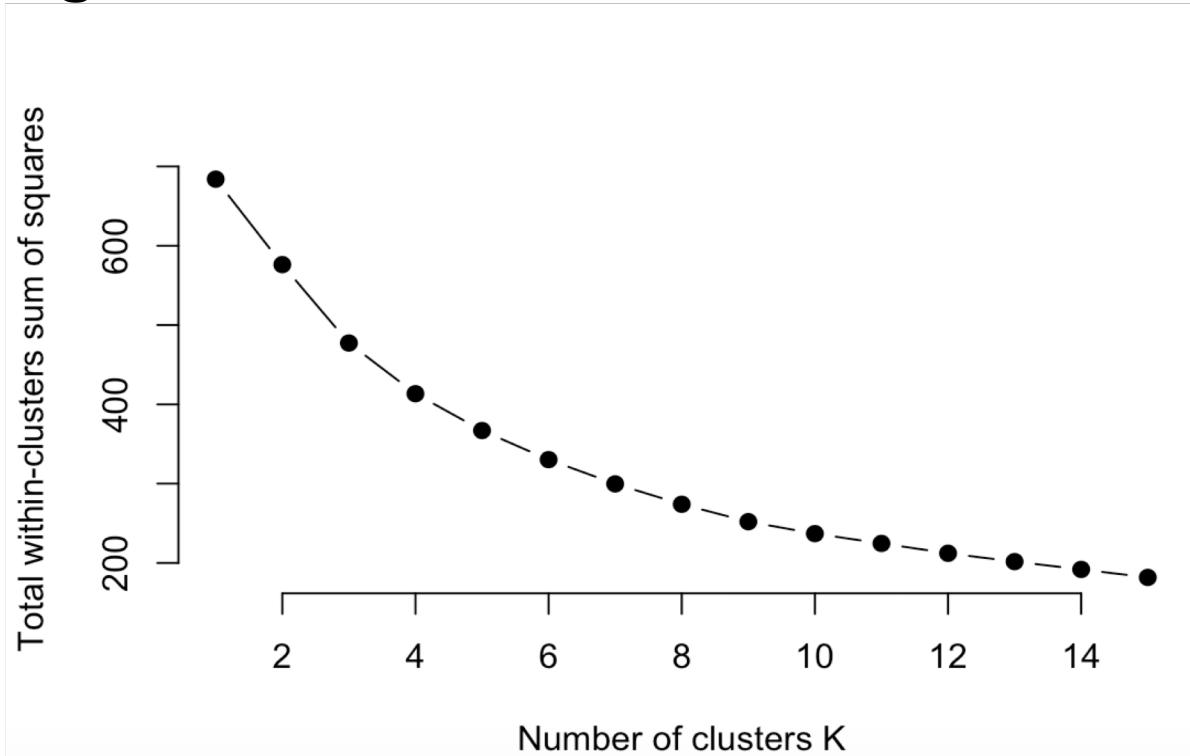
# GridSearchCV

- Exhaustive search over specified parameter values for an estimator.
- Important members are fit, predict.
- GridSearchCV implements a “fit” method and a “predict” method like any classifier except that the parameters of the classifier used to predict is optimized by cross-validation.

Algorithm	GridSearchCV Score
KNN	0.8032786885245902
Random Forest	0.7180327868852459
Decision Trees	0.740983606557377

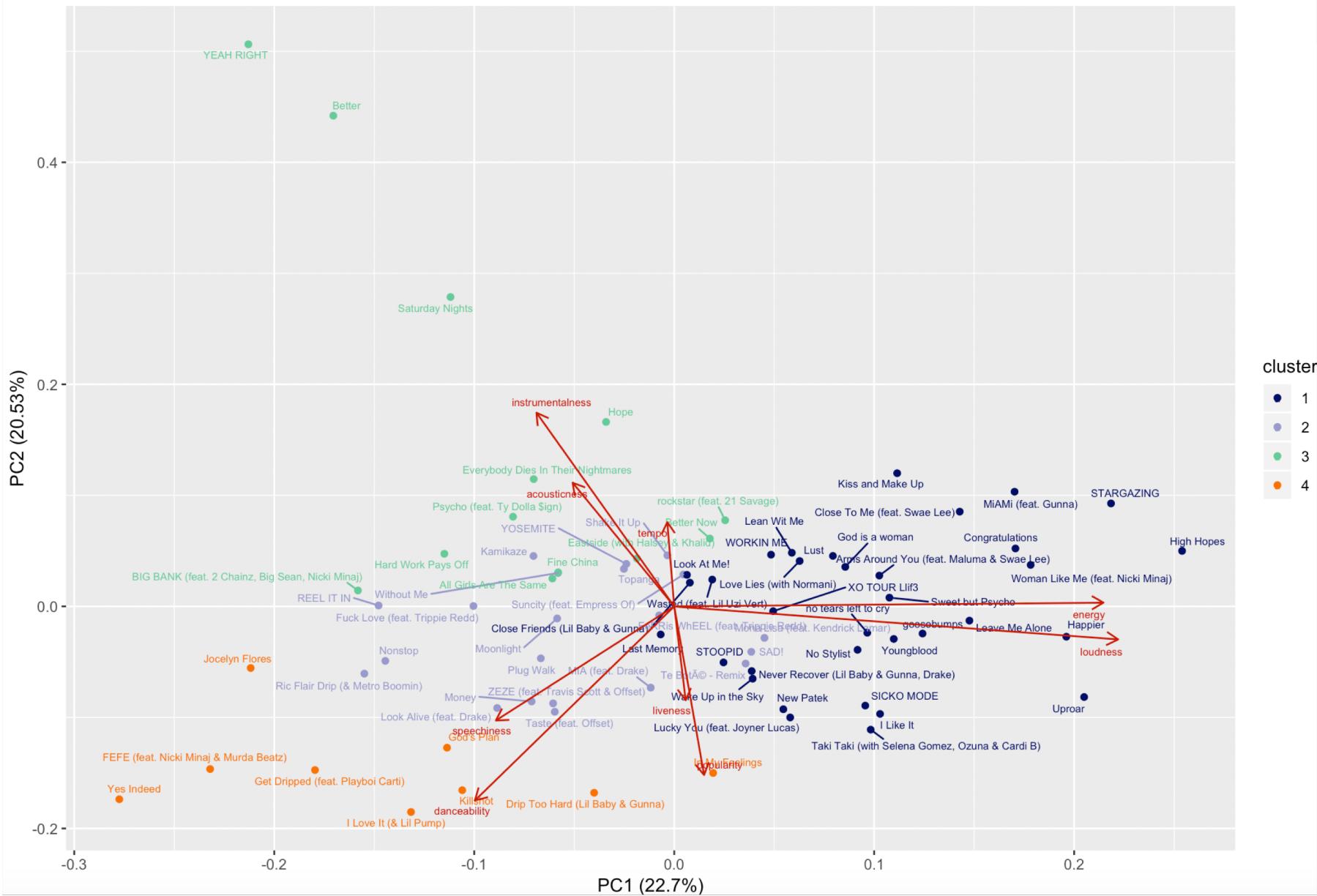
# Clustering

## K-means Clustering



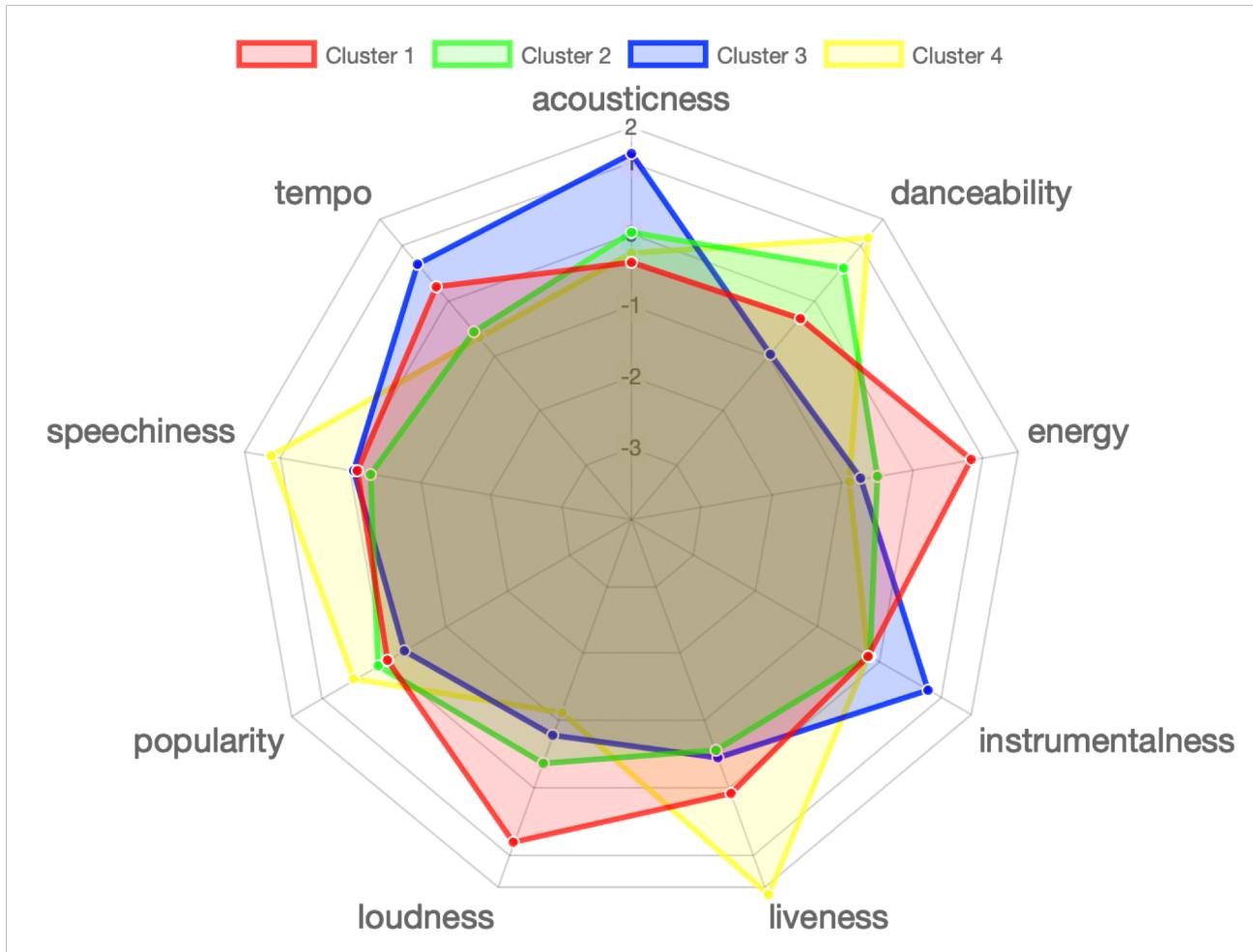
- There is no significant point for the elbow, but the closest we could get was for  $k=4$ .

# K-means Clustering

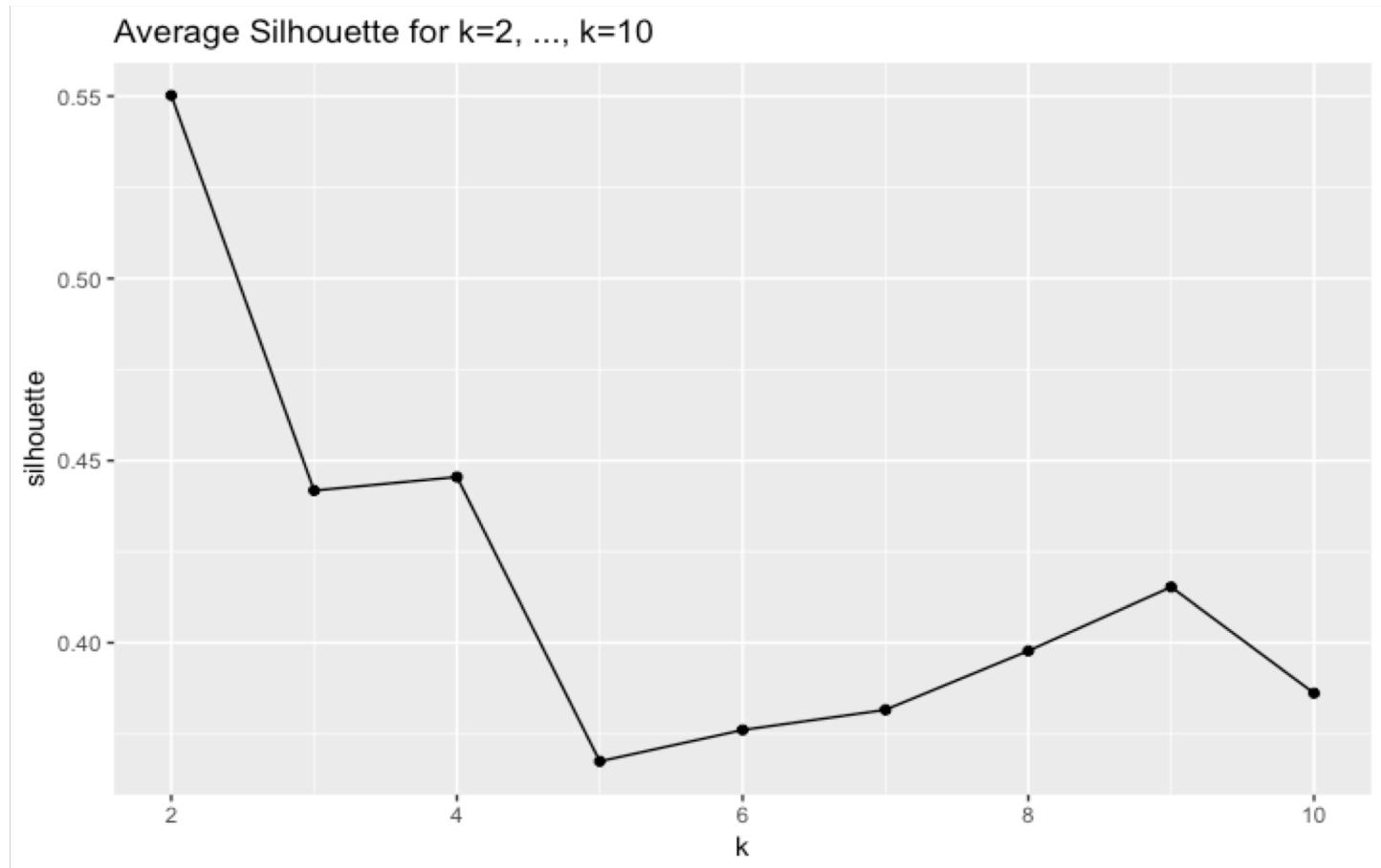


# Clustering

## K-means Clustering

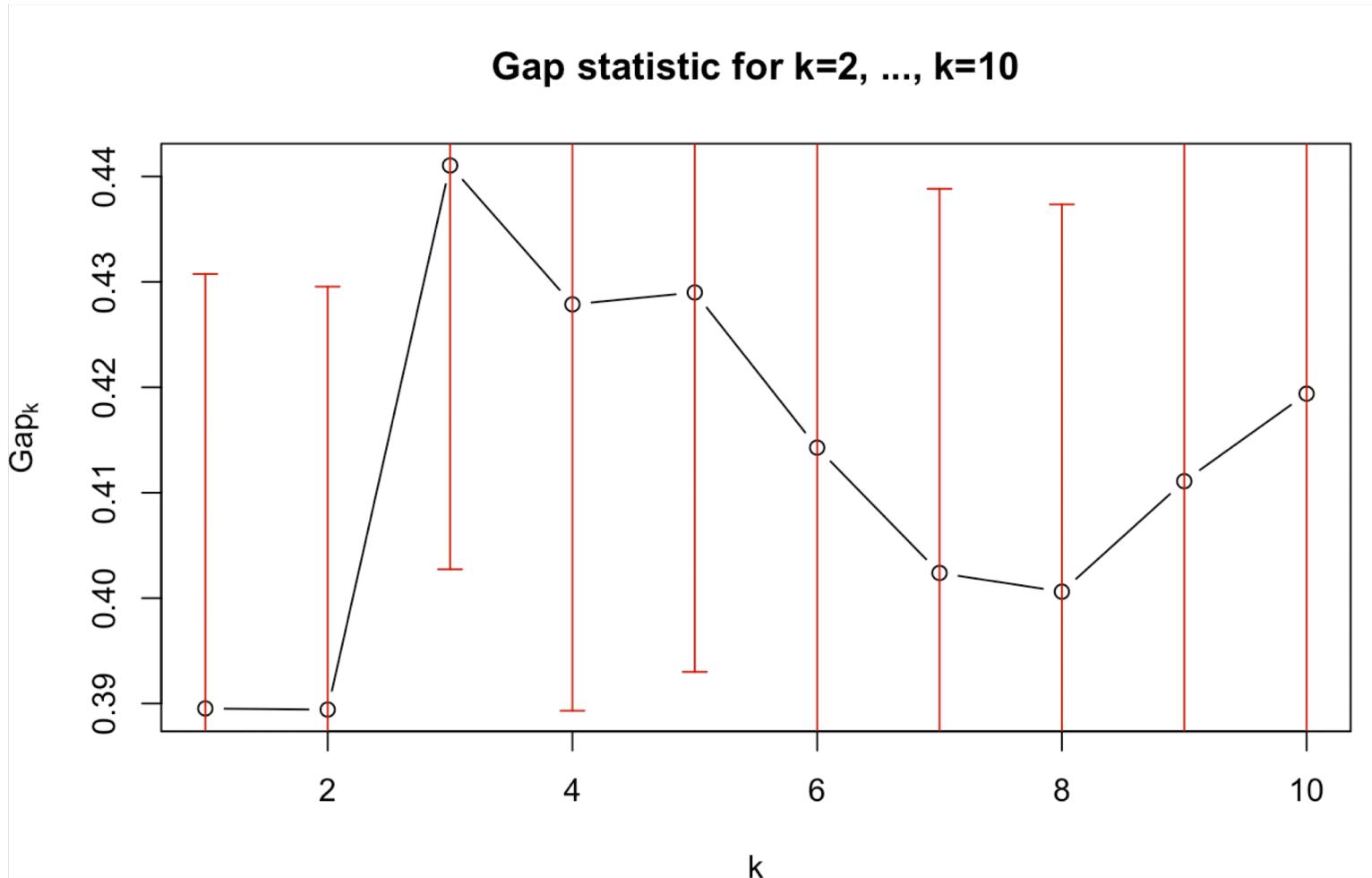


# Clustering



The silhouette coefficient measures how similar a data observation is to those on its own clusters.

# Clustering



The gap statistic compares the total within intra-cluster variation for different values of  $k$  with their expected values .

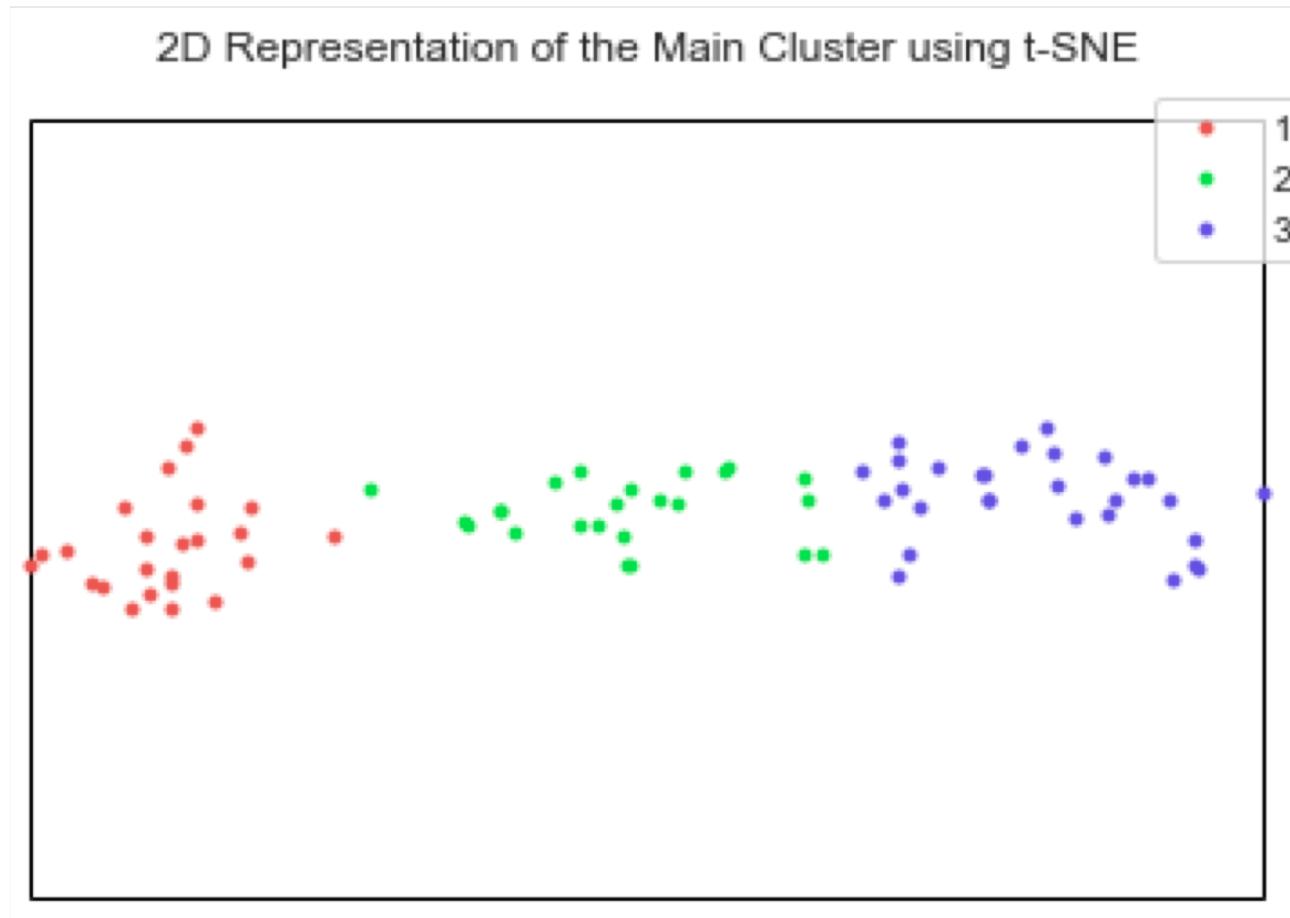
# K-medoids



- In k-medoids, an actual data point of the dataset is used as the most centric item of the cluster, unlike k-means, which uses a point in the feature space that represents the center of a cluster, making the former more robust to outliers.

# Dimension Reduction

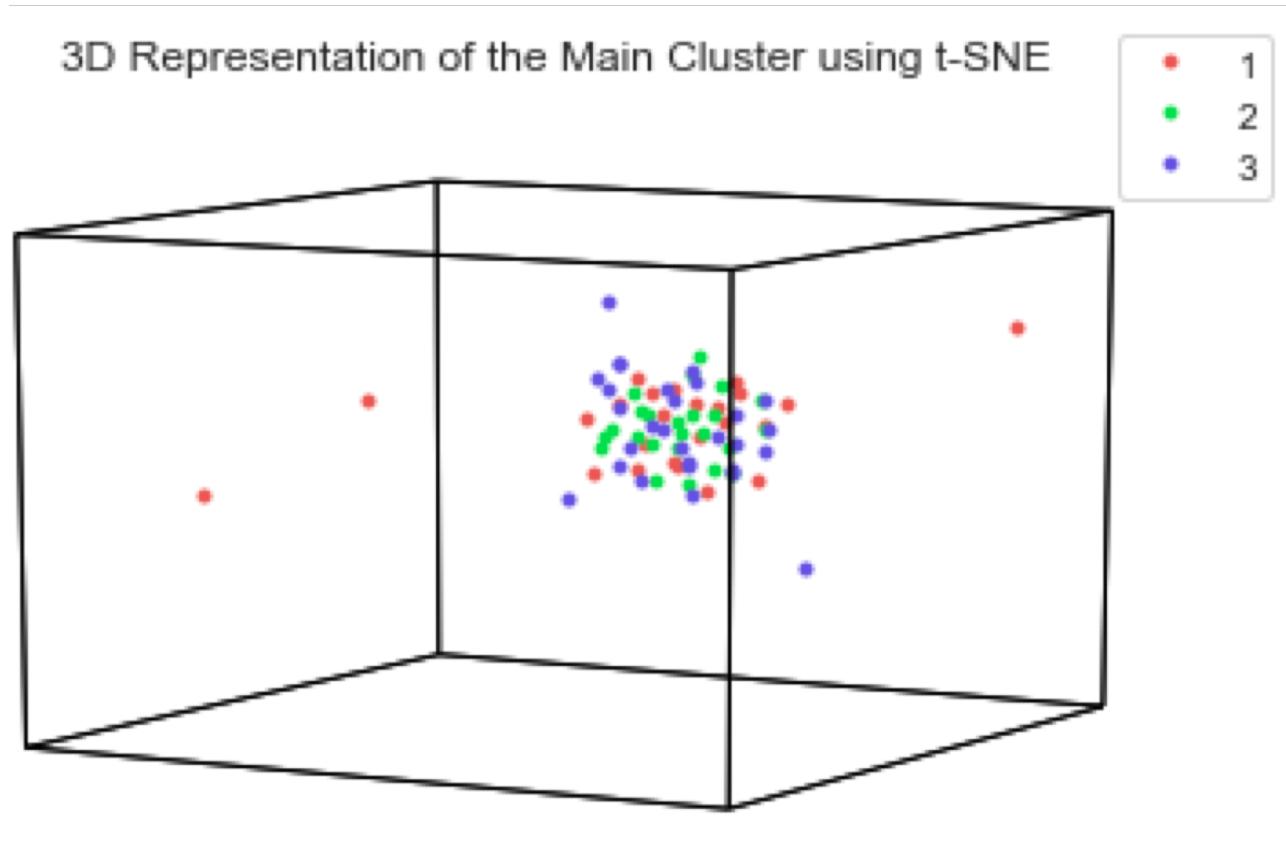
## For K-medoids clustering



- t-Distributed Stochastic Neighbor Embedding (t-SNE)

# Dimension Reduction

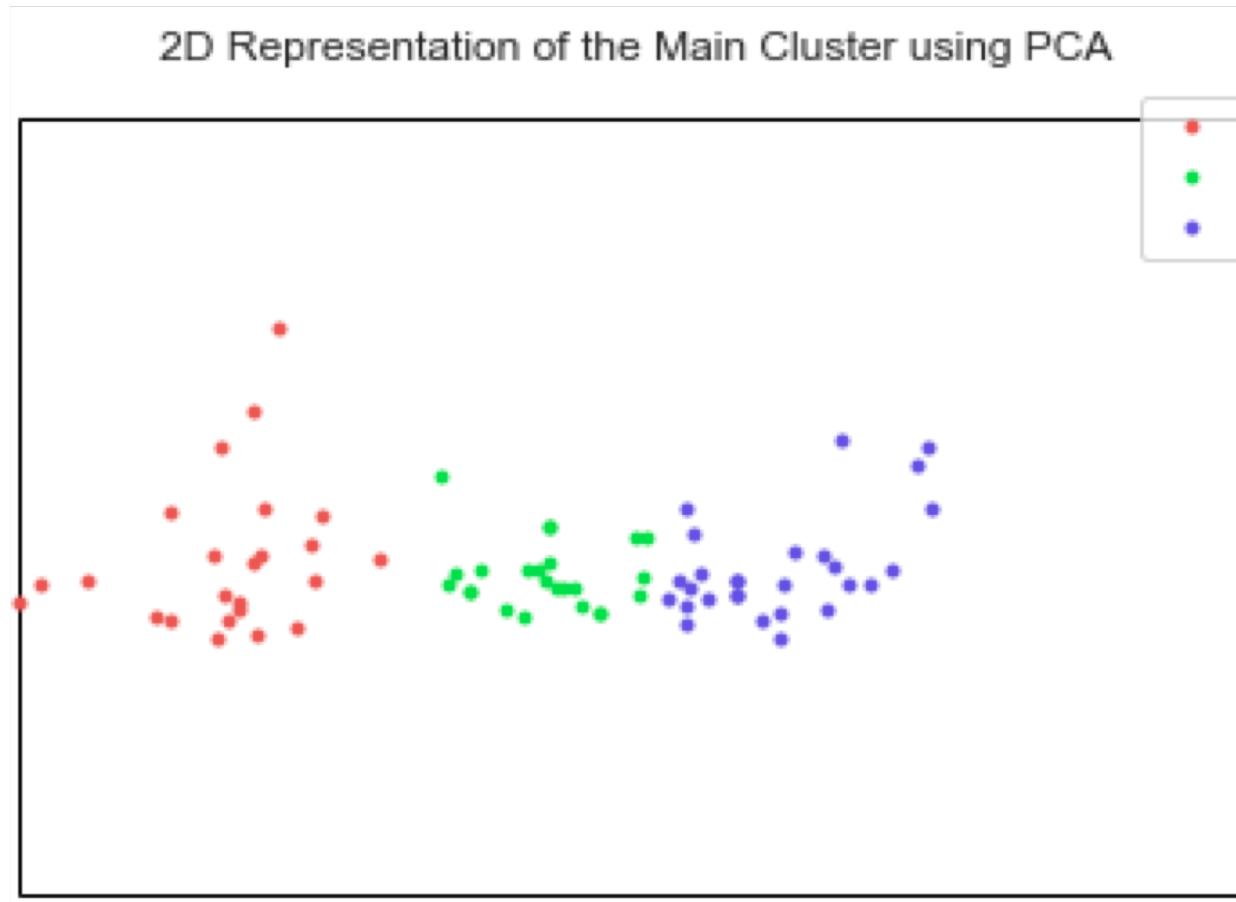
## For K-medoids clustering



- t-Distributed Stochastic Neighbor Embedding (t-SNE)

# Dimension Reduction

## For K-medoids clustering



- Principal Component Analysis (PCA)



# Clustering

## Results

- The 3 playlists that were created out of the recommended 77 songs were :-
- ***HIP-HOP***
- ***RNB***
- ***TRAP***
- Outliers in HIP/HOP are Happier, Taki Taki, Eastside
- Outliers in RNB are no tears left to cry
- Outliers in TRAP are god is a woman, Woman Like Me



**STEVENS**  
INSTITUTE *of* TECHNOLOGY  
THE INNOVATION UNIVERSITY®

**stevens.edu**

---

**THANK YOU**