Spotifying Trends in Popular Music

Anastasia Vela
Joshua Asuncion
Kevin Pang
Steve Hwang

Motivation

- What is common about top tracks every year?
- Have trends changed across years
- Data available for audio features for analyses on Spotify
- Make prediction on the next popular songs based on audio features

Relevant work

- GitHub Spotipy
  - People utilize Spotify API for interesting projects such as **automatically creating playlists** by comparing scores of each feature around
- Spotify used their workflow manager software, luigi, to accurately predict 4 of the 6 Grammy winners taking into account the user's listening habits

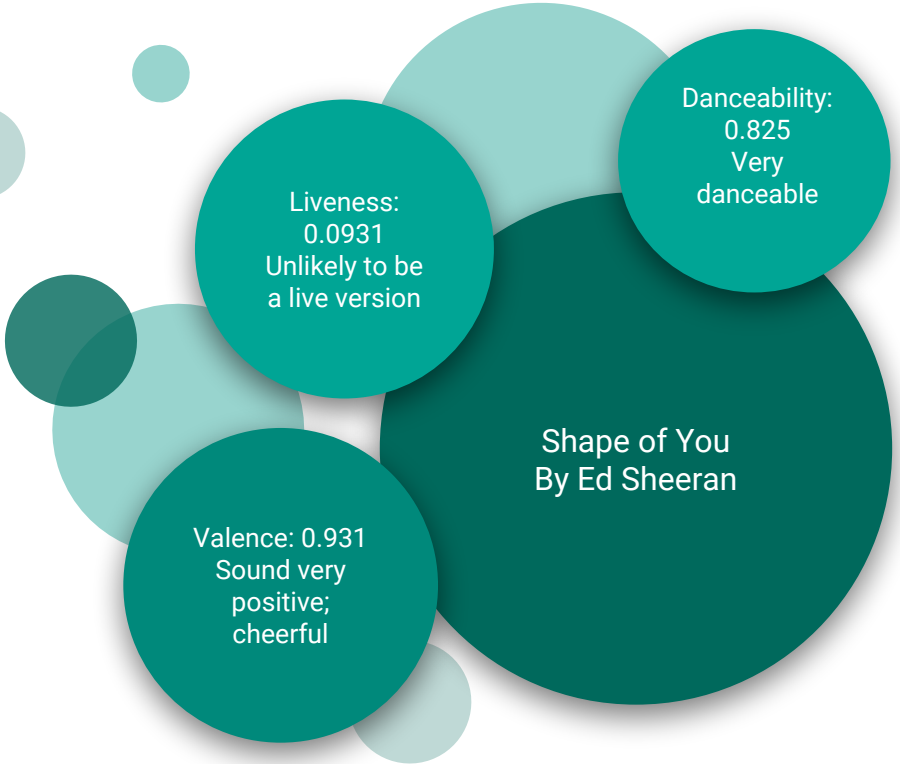
Dataset

- Billboard top 100 songs from 2012 to 2016
- Used Spotify API to collect audio features for each track
- We created own GitHub library to collect the data
- Billboard top 50 songs in 2017 for prediction
    - Machine learning - assume we are in 2016 and don’t have 2017 list
- Saved as .csv files

Key Description Examples

Key	Value Description
acousticness	0.0 to 1.0; 1.0 represents high confidence the track is acoustic.
danceability	0.0 to 1.0; 0.0 is least danceable and 1.0 is most danceable.
energy	0.0 to 1.0; For example, death metal has high energy.
instrumentalness	0.0 to 1.0; Absence of vocals. Values above 0.5 are likely instrumental tracks.
liveness	0.0 to 1.0; A value above 0.8 provides strong likelihood that the track is live.
loudness	Values typical range between -60 and 0 db.
speechiness	Values above 0.66 describe tracks that are probably made entirely of spoken words.
tempo	Beats per minute (BPM)
valence	A measure from 0.0 to 1.0 describing the musical positiveness. Tracks with high valence sound more positive.

Example


Danceability:
0.825
Very
danceable
Liveness:
0.0931
Unlikely to be
a live version
Shape of You
By Ed Sheeran
Valence: 0.931
Sound very
positive;
cheerful

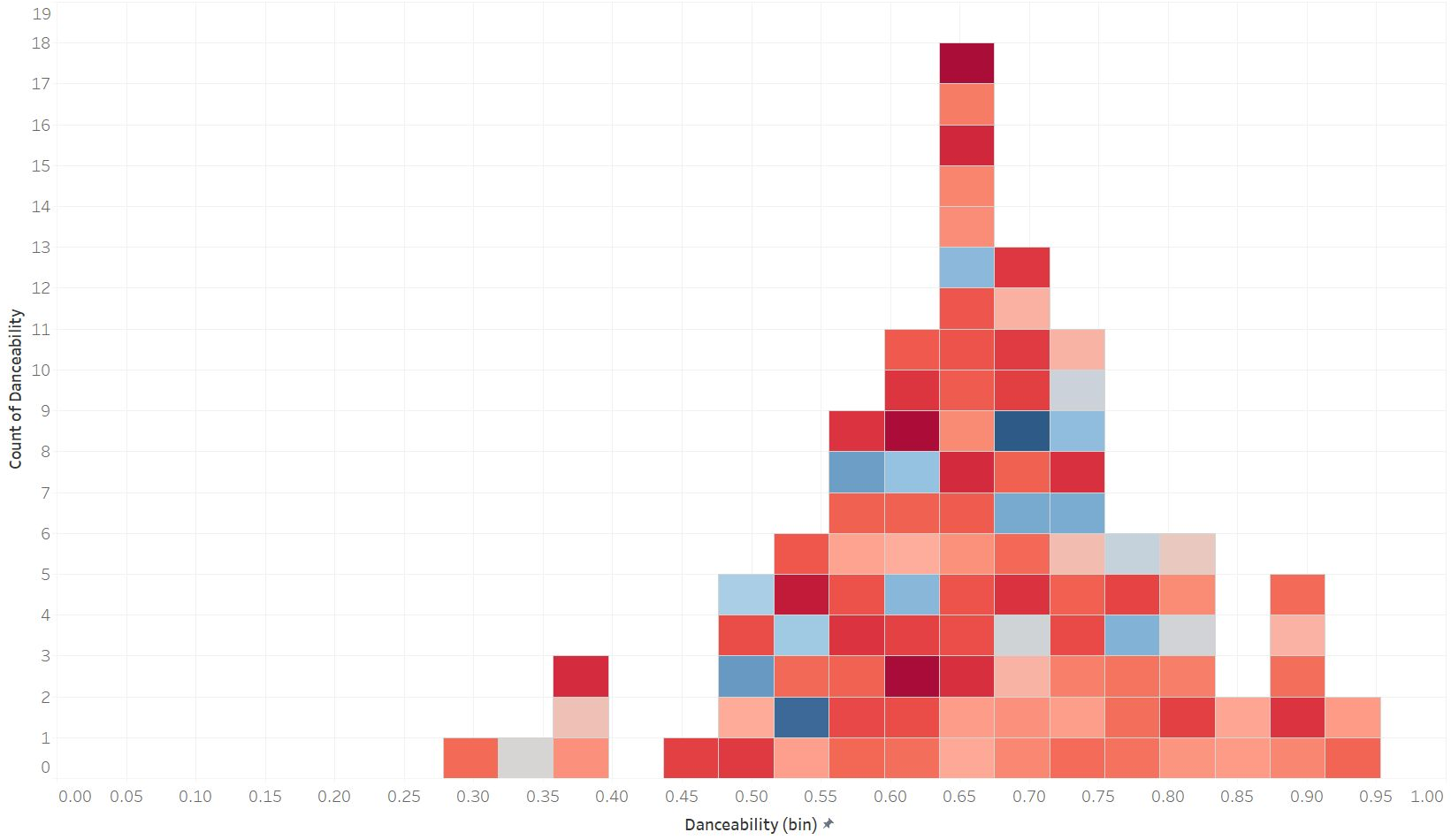
Demo

Data Collection

Assumptions and Challenges

- We assume Spotify feature scoring system is correct
- When we were collecting data
  - Some tracks have same names by different artists
  - Different versions
- We solved by recording a specific Spotify ID for each track
  - Name is not unique but ID is
- Used IDs instead of names to extract features using API and Jupyter Notebook

Danceability of 2016 Songs



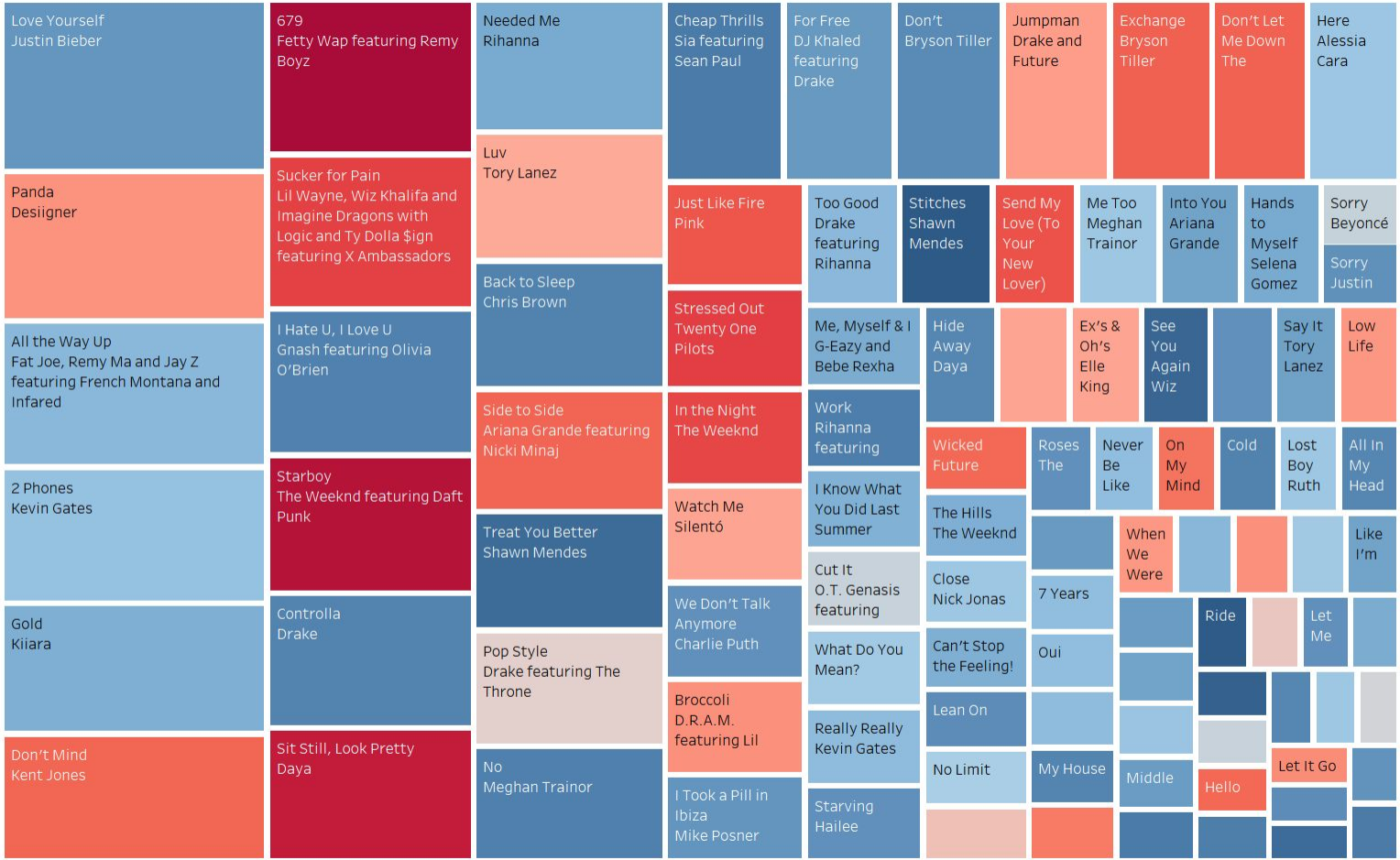
Energy

0.0565 0.9280

Caption

The trend of count of Danceability for Danceability (bin). Color shows details about Energy. Details are shown for Song.

Speechiness of 2016 Songs



ATTR(Tempo)

74.8 190.1

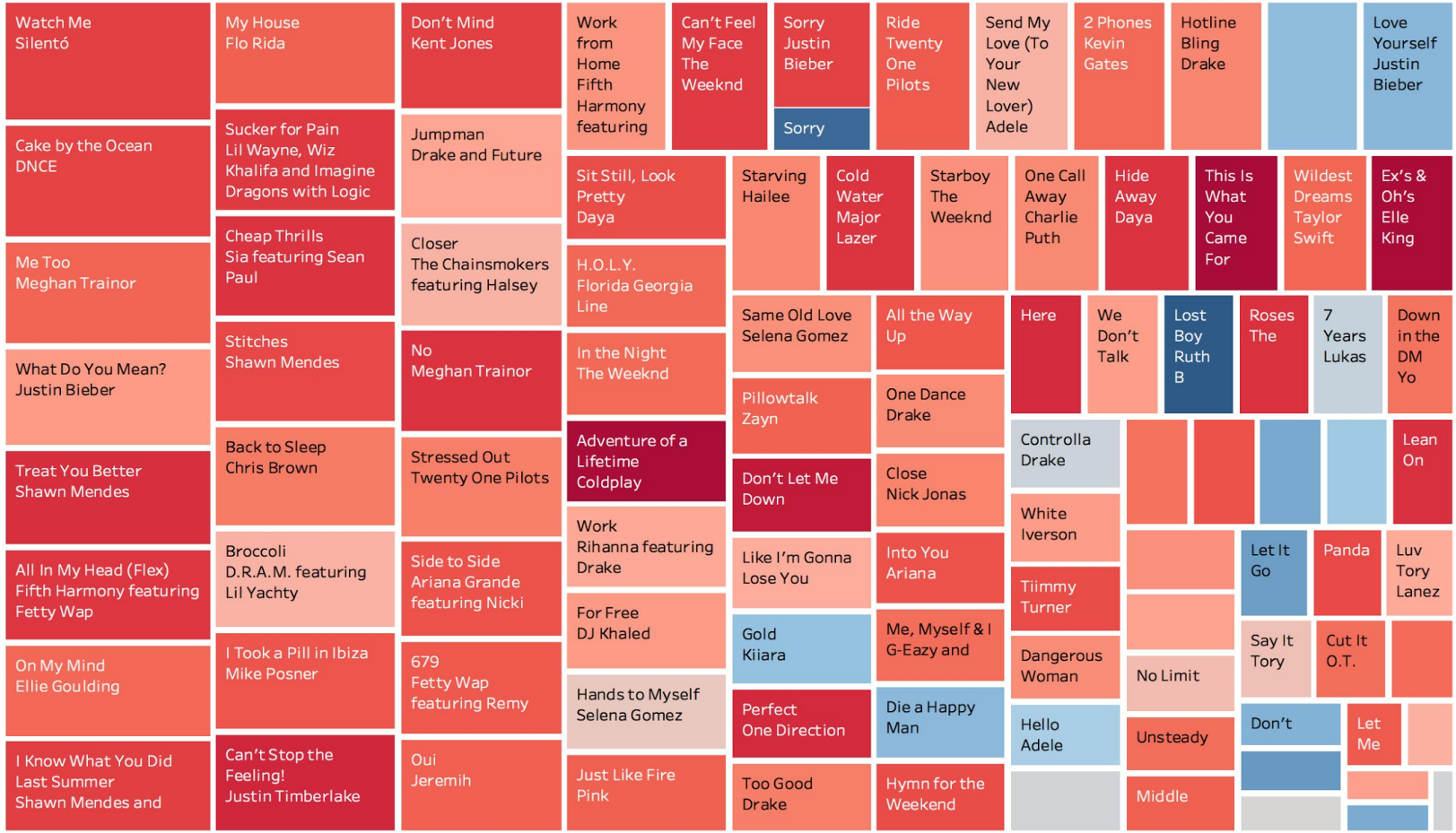
Caption

Song and Artist. Color shows Tempo as an attribute. Size shows Speechiness as an attribute. The marks are labeled by Song and Artist.

Summary

Count:	100
ATTR(Speechiness)	
Average:	0.1195
Minimum:	0.0286
Maximum:	0.4530
Median:	0.0789
ATTR(Tempo)	
Average:	120.6
Minimum:	74.8
Maximum:	190.1
Median:	117.6

Valence of 2016 Songs



SUM(Energy)

0.0565 0.9280

Link to Tableau Public

https://public.tableau.com/views/Spotify_10/Sheet1?:embed=y&:display_count=yes

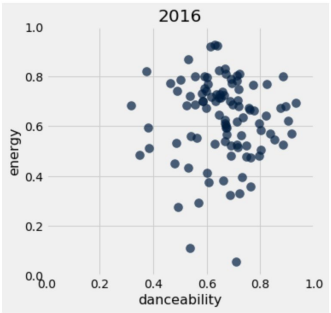
Time segments

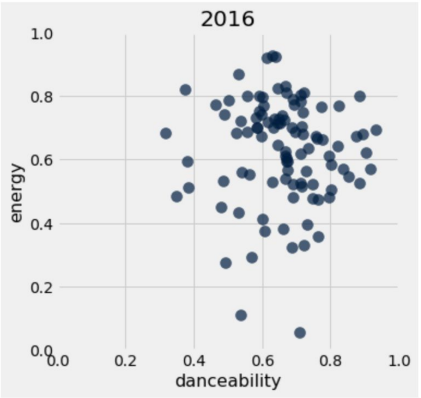
```
import ipywidgets as widgets
_ = widgets.interact(attributes, x=3, y=4, year=widgets.IntSlider(min=2012, max=2016, value=2012))
```

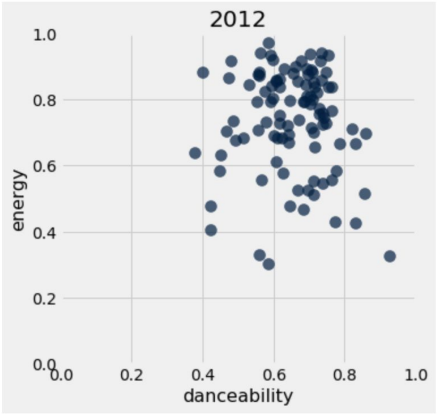
x 3

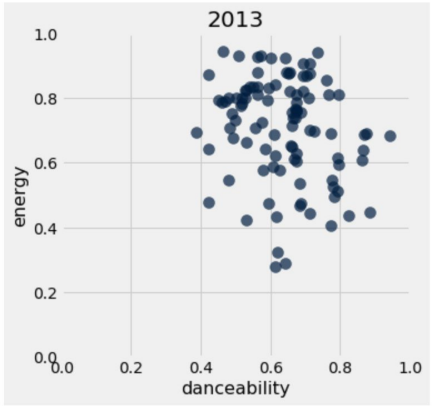
y 4

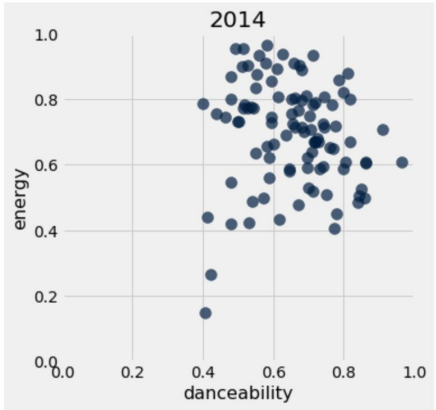
year 2016

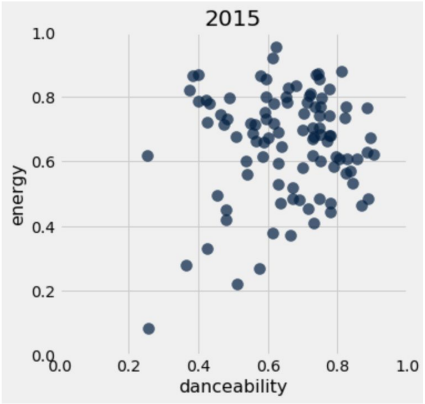




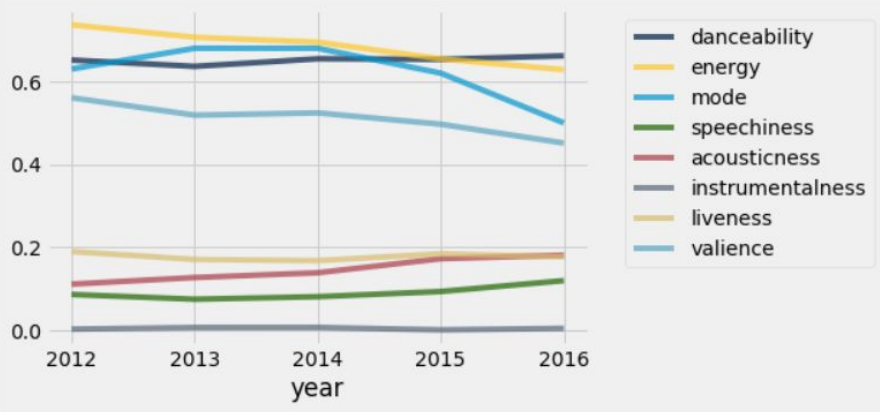


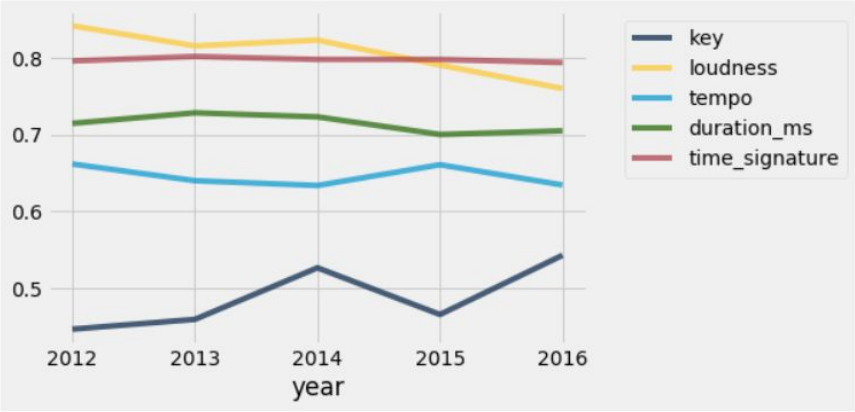






Trends across years





- energy is decreasing over years
- mode is decreasing over years
- valence is decreasing over years

Demo

Line Graphs

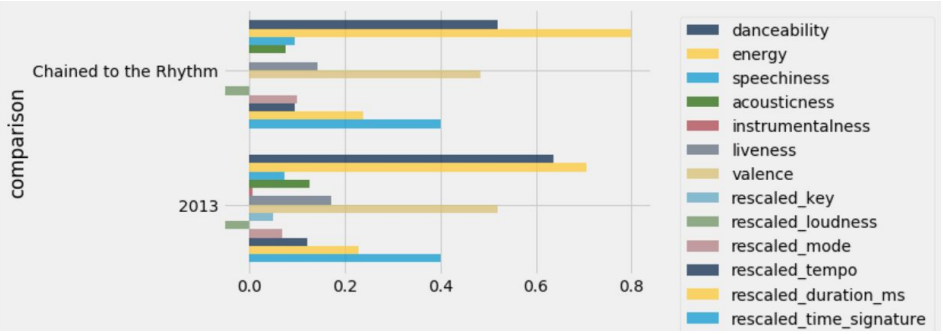
Prediction of the next popular song

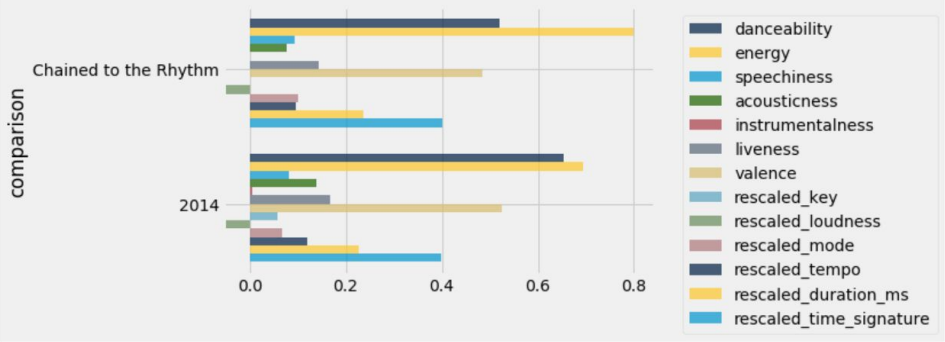
- Take averages of all features in comparison dataset (2012, 2013, etc.)
- Subtract averages from features of each song in 2017 dataset
- Find the song with the smallest total difference - that's our predicted most popular song
- Limitation: predictor based solely on features of songs in comparison dataset
    - Doesn't take into account outside factors

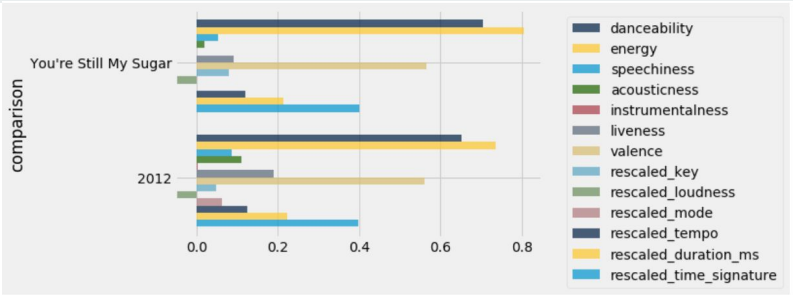
Demo

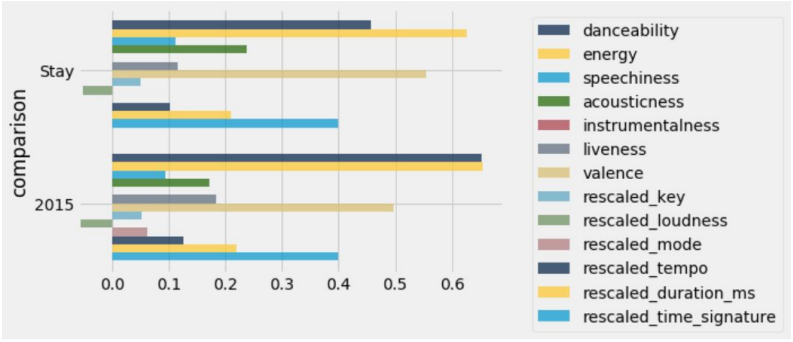
Predictor

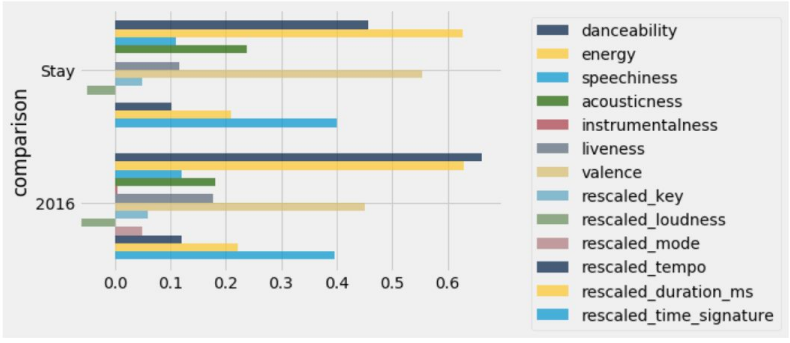
Predictor results







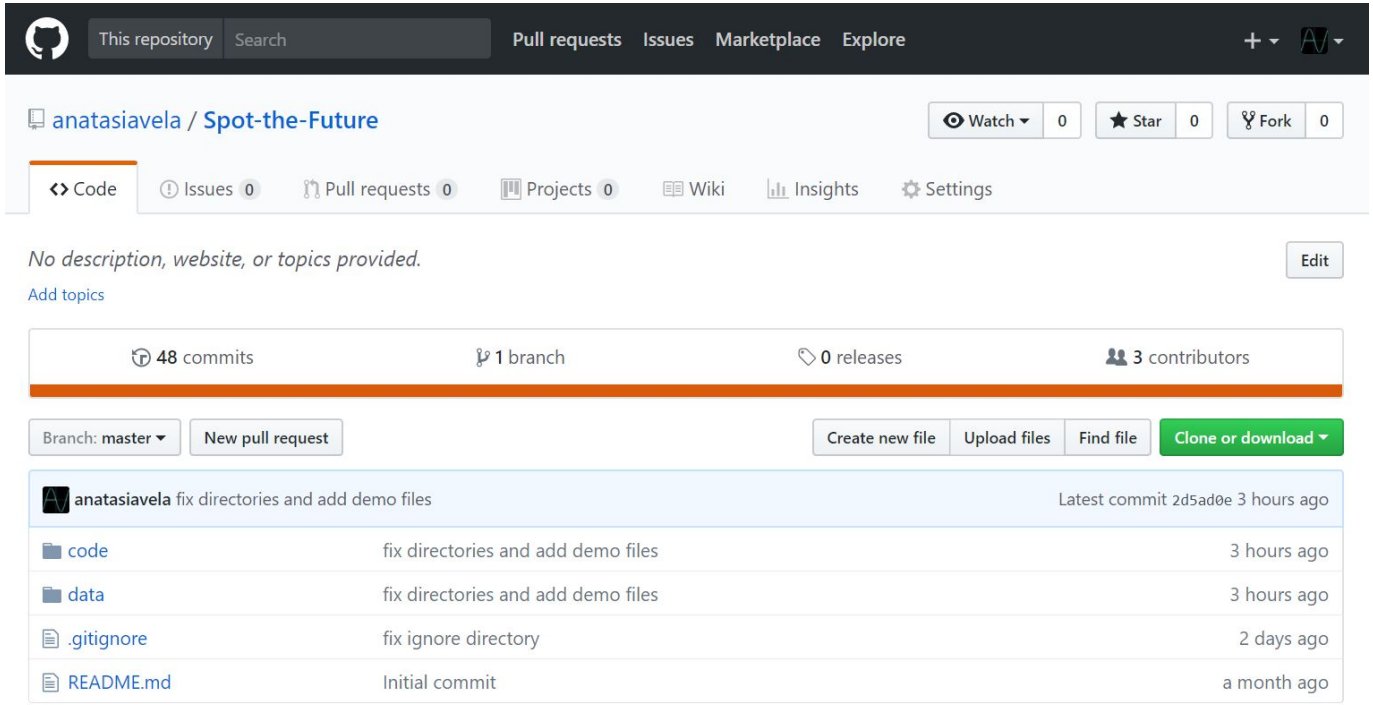




Predictor results

- Using 2016 dataset: Stay - Zedd & Alessia Cara
- Using 2015 dataset: Stay - Zedd & Alessia Cara
- Using 2014 dataset: Chained to the Rhythm - Katy Perry ft. Skip Marley
- Using 2013 dataset: Chained to the Rhythm - Katy Perry ft. Skip Marley
- Using 2012 dataset: You’re Still My Sugar - Maddie Ross

GitHub



https://github.com/anatasiavela/Spot-the-Future

Findings

- Energy goes down from 2012 to 2016
- Most top songs have 0 instrumentalities
- Most top songs are 4/4
- More songs are having minor keys - now at 50%
- ...

Limitations and what else we can do

- We have visualizations for top 100 songs but we don’t have data for other songs - no comparison
- Data might not be correct
- Generate visualizations across years and use more features on Tableau

Work Citation

https://github.com/plamere/spotipy

https://developer.spotify.com/web-api/get-several-audio-features/

https://www.billboard.com/charts/year-end

https://www.inferentialthinking.com/

http://spotipy.readthedocs.io/en/latest/#module-spotipy.oauth2