**Spotify API Project Final Paper**

**Scott Berry**

**Abstract:**

In this Data Visualization project proposal, a potential combination and improvement upon

Spotify Wrapped and Spotify Blend will be suggested. By querying the Spotify Data Catalogue

via the Spotify API with Spotipy python library, visualizations of shared musical interests will

allow a reader to easily interpret a group’s favorite genres of music.

**Introduction:**

Every year in December, Spotify creates a “Spotify Wrapped” story that returns your top listing

of several fields such as most listened to artists, songs, and genres. While always interesting to

scroll through, this is always presented for entertainment rather than informative from a big

picture. Spotify Wrapped lists top 5 of major categories rather than a full picture. Another

Spotify project, “Spotify Blend”, curates a shared playlist among friends by adding users to the

shared playlist. This project combines shared highly listened to songs between users but

doesn’t give a great sense of what a friend group’s shared music interests are by genre.

In this Data Visualization project proposal, a combination and improvement upon the

interpretation of these outputs can be achieved through a data visualization. Visualizations

from the Spotify API have been created before (Bean 2018), but what these plots lack is the

dimension to plot one’s own categorical habits against another’s (or a group’s) habits. The 1-

dimensional data can be compared via boxplots as Bean did for his weekly listens, but different

means are necessary when comparing categorical data.

A great visualization to compare a group’s musical taste will plot user listen frequency by genre

with all users of the friend group’s data easily compared each other. From this, a member of

this friend group might know what genres of music and in which proportions belong on a road

trip playlist or for a party.

**Process:**

Using the Spotify API (Spotify 2022), JSON metadata about music artists, albums, and tracks,

directly from the Spotify Data Catalogue via the Python Spotipy library (Lamere 2014). For each

Spotify user, each datapoint retrieved will be a song with corresponding number of listens and

its genre. There are many more fields available that I may use if it will help with data

manipulation. With this tidy data imported from JSON into a Pandas dataframe, a radar plot

visualization of each user’s most listened to genre can be computed. This dataset is robust and

with all its fields of metadata will certainly have the information necessary for the proposed

visualizations.

**Results & Insights:**

Foo.

**Discussion, Conclusion, & Future Work:**

This visualization aims to beautifully and accurately display multiple user’s favorite genres of

music and which genres are shared interests between friends. The information should be clear

on which users share interests and how strong that connection is. Multiple visualizations will

ideally be used to display shared interests, but a Seaborn radar plot implementation

(Khandelwal 2019) is a great choice for the main visualization. In this visualization, hard

numbers of frequency of listens should be scaled so that listen frequencies are as a percentage

of total listens. Further, an interactive data visualization can be used to filter by specific user

INTERNAL USE

and by specific genre, just as Bean sorted by time of day (Bean 2018). It should be clear from

the visualization what the group’s favorite genres of music are.

**References:**

Bean, Jordan. A Visual Look at my Taste in Music. https://towardsdatascience.com/a-visual-

look-at-my-taste-in-music-a8c197a728be. 2018.

Khandelwal, Aakash. Seaborn Radar Plot Implementation.

https://aakashkh.github.io/python/visualisation/2019/08/26/Plotting-Seaborn-Radar.html.

2019

Lamere, Paul. Spotipy. https://spotipy.readthedocs.io/en/master. 2014.

Spotify. Spotify API. https://developer.spotify.com/documentation/web-api. 2022.