

Musical Wayfinder: An Interactive Look into Your Spotify Library

Hannah Bish*
Steven Stetzler*
hvbish@uw.edu
stevens@uw.edu
University of Washington
Seattle, WA

ABSTRACT

We present the Musical Wayfinder, an interactive visualization intended to bridge that gap between us and our music. This tool allows the user to track their musical tastes through time, investigate the musical attributes of > 3,000 genres catalogued by Spotify, and discover where their personal library falls within that vast musical landscape. We discuss the development and implementation of this tool, and the insights it can provide.

KEYWORDS

Visualization, Data dashboard, Spotify, Music

1 MOTIVATION & RELATED WORK

Although most of us listen to music, there are few tools available that allow us to examine our musical tastes. Services like Spotify have invested vast resources into categorizing and recommending music algorithmically based on listening habits and quantifiable measures of sounds, but they have not made this data available to the end user in a meaningful and accessible way.

Currently, the best tool for exploring musical genres is Every Noise at Once [1], an online map of musical genres. It was developed in 2013 by Glenn McDonald for The Echo Nest, a "music intelligence and data platform," with the goal of better understanding the audio content of music and improving the recommendation services provided by the company. His approach made use of algorithmically-generated musical attributes, which were subsequently given more descriptive names like Energy or Acousticness that roughly reflect their corresponding musical qualities. McDonald successfully used the tool to locate clusters of related artists in data-space, identifying and classifying new subgenres of music.

Every Noise at Once is the landmark visualization of music that emerged from this process. It maps genres along two axes, one representing sonic density and another representing acousticness

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

(see Figure 1). Users can click on a genre to preview what it sounds like or see what artists compose that genre. This detailed map of subgenres reveals relationships between them and provides a broad context for exploration of genre space.

While this tool is both fascinating and useful, we see an opportunity to improve upon this idea. It is interesting to explore these genres, and the site lends itself best to random exploration and chance discoveries. But doesn't provide a natural way for a user to find their favorite music and learn where those genres exist within the bigger picture. Our visualization builds upon Every Noise at Once and makes personal listening data accessible while enabling more flexible exploration of this genre data space.

2 METHODS: DATA SCRAPING AND PREPROCESSING

The data we are visualizing is generated from a dynamic scraping and processing algorithm that utilizes the Spotify Web API¹. We gather public data from Spotify to create a catalog of genres, described in §2.1. We also gather private Spotify library data using user-provided credentials with a web application, described in §2.2. We then perform preprocessing steps to merge these data sets before visualizing them, described in §2.3.

At the core of the accumulated data set are the quantitative sonic attributes that Spotify generates for each song on its service². Among these features, we visualize acousticness, danceability, energy, instrumentalness, loudness, speechiness, and valence (a measure of positivity or negativity in the emotions a song may elicit).

2.1 The Sounds of Spotify

For each of the genres that Spotify classifies artists into, they also curate a dynamically updated playlist representing that genre on a user page called The Sounds of Spotify³. Each playlist contains roughly 100 songs that Spotify has determined represent that genre the best⁴. We use the Spotify Web API to scrape the contents of these playlists, including requests for the sonic features associated with each song. To represent each genre in the feature space we visualize, we average over all of the sonic features of the songs in each playlist to distill this collection of songs down to a single point.

¹<https://developer.spotify.com/documentation/web-api/>

²More information can be found at <https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/>

³Data can be accessed at: <https://open.spotify.com/user/thesoundsofspotify>

⁴For an example, see <https://open.spotify.com/playlist/2pni39gq7ITfQJmVacTrM>

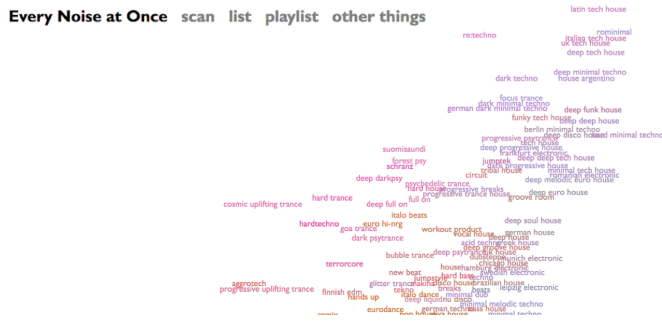


Figure 1: The Every Noise At Once website, a landmark visualization which maps musical genres along attribute axes roughly corresponding to "Energy" and "Acousticness".

2.2 Scraping User Data

Our visualization is not very useful without the ability to provide insight into the libraries of other users. To enable this, we built a web application using Flask⁵ to enable user log-in and authentication through Spotify's API to scrape the library contents of individual users. This allows for anyone to use our visualization without having to go through the burden of scraping their own data using Spotify's Web API. We use the web API to gather information about all of the songs in a user's library, including its title, producing artists, date added to the library, and its sonic attributes. For each song, we perform additional calls to the API to get the genres Spotify has associated with each artist credited on the track. This list of genres, including duplicates, is then associated with each song. Additionally, we use the API to get the user's top tracks, artists, and metadata information about the user such as their display name, which is usually the user's full name.

2.3 Umbrella Genre Classification

We chose to classify all genres into larger "umbrella genres", a more general categorization meant to better reveal the relationships between genres and show their clustering in attribute space. These umbrellas included more obvious categories like pop, rock, rap, electronic, and classical. After our our initial exploratory analysis, we also chose to include jazz and metal as additional categories; although they represent smaller and more specific categories of music, they exhibit distinct and interesting behavior in attribute space which we felt would add to the user's experience.

We performed a text analysis of the name of each genre in Spotify's catalog to perform a rough classification. The full list of subgenres was spot-checked by eye to determine words or phrases that correspond to each umbrella classification. For example, subgenres were classified into the electronic umbrella if their name included "electro" as well as 13 other strings including "idm", "glitch", "techno" and "trance". Some erroneously classified genres had to be removed from the umbrella categories as well; for example the inclusion of "dub" and "edm" allowed several subgenres to be included under the electronic umbrella, but also meant that others such as "Dublin Indie" and "Edmonton Indie" had to be identified and removed.

⁵<http://flask.pocoo.org/>

The nebulous nature of genre definition made this a challenging task. Many subgenres might belong in more than one umbrella, and our catch-all "Other" umbrella for the remaining uncategorized subgenres leaves a significant fraction of music undefined. Nevertheless, our rough classification produced reasonable results over a large portion of the genres that are typical in a user library and served the purposes of this particular visualization.

3 RESULTS: THE MUSICAL WAYFINDER

Using the data that we are able to scrape from Spotify, we are interested in answering three main questions with our visualization tool:

- (1) What are the contents of my library and can I view this visually?
- (2) Where do my musical tastes fit in the landscape of music genres?
- (3) How have my music tastes changed over time?

To address these questions, we've developed several views of a user's data that are linked with interactions, shown in Fig. 2. We address Q1 and Q2 by creating two scatter plots: one that shows the sonic attributes of a user's songs and another that shows the average sonic feature value of the genres that exist in their library. Points are colored by their umbrella genre so the user can gain quick insight into the type of music they listen to. An interactive legend shows the umbrella genre color coding. Users can click on items in the legend to remove or add back in songs and genres from the corresponding umbrella genre, allowing for umbrella-based subsets of the library to be made. Additionally, the legend has color-coded bars that show the relative abundances of umbrella genres in their library. Within the genre scatter plot, points representing genres are sized by the number of songs in the user's library that lie in that genre, allowing for a user to quickly recognize what genres they listen to the most. We implement two dropdown menus that allow the user to select from among the features available for each song those that should be plotted along the two axes. This solution allows a user to see the contents of their library in a visual manner while giving them an idea of how their music sounds based on visual clustering and correlations between the two features they selected. By default, the genre scatter plot shows only the genres that the user has in their library, but a button allows the user to toggle between this view and viewing all of the genres that Spotify catalogues at once. Switching between viewing "all genres" and "my genres" allows the user to see how their music tastes compare to all music available to them on the Spotify platform.

To address Q3, we create a line chart that shows when songs were added to the user's library and their umbrella genre composition. Data are binned in time, and the y-axis shows the number of songs that were added in that time range. The line is the result of an interpolation over the bin heights, meaning that y-axis values are only approximate measures of the number of songs added at a certain time. Finally, the height of the line is filled with different colors to represent composition of the library with respect to umbrella genres. A brush filter allows for a user to click and drag over a time region, which will update the song and genre scatter plots as well as the bars on the legend to reflect their library within the chosen time range and their corresponding genres. This allows for in depth

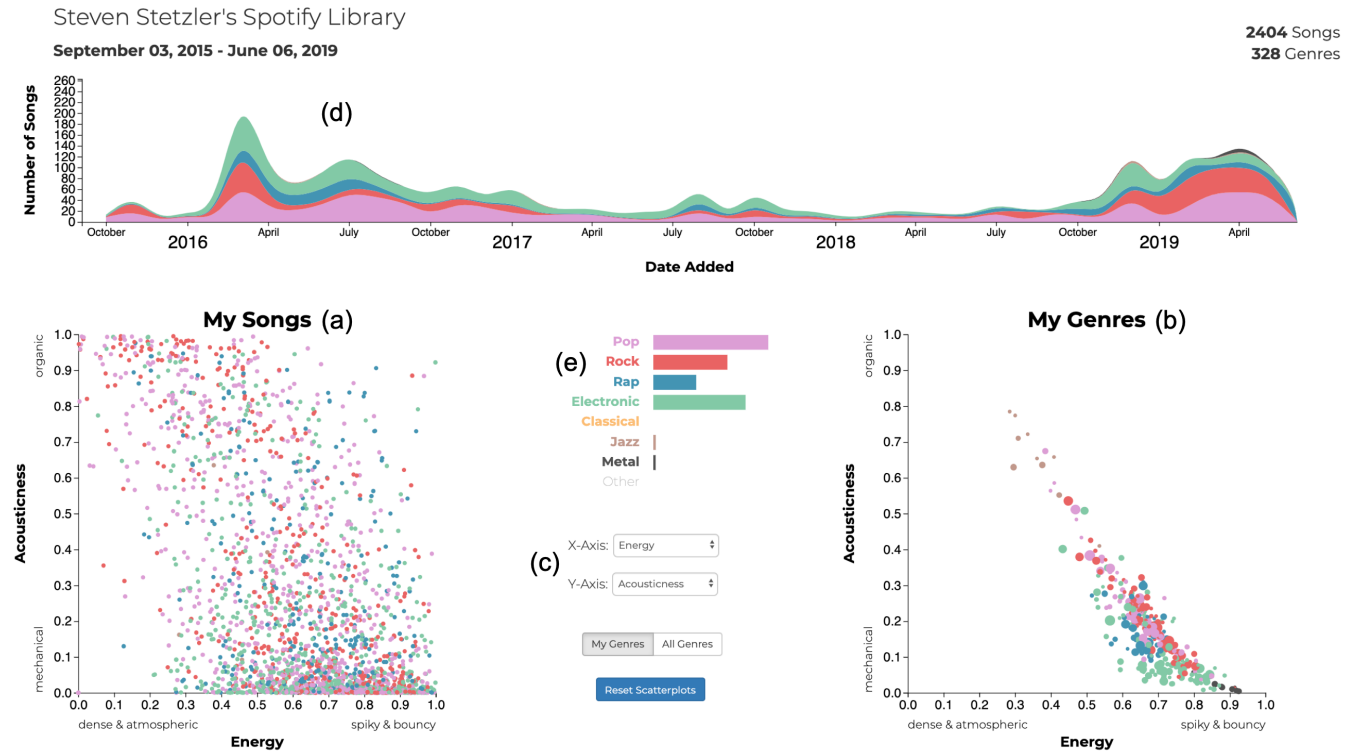


Figure 2: The Musical Wayfinder dashboard. A scatter plot of user library songs (a) and the genres that the library is composed of (b) are the main views of the data. Dropdown menus (c) allow the user to select the axes of the two scatter plots. A timeline (d) of the user’s library allows the user to see when they added songs to their library. A legend (e) shows the color mapping of the umbrella genres that each song and genre are classified into. Bars show the relative abundances of those umbrella genres in the user’s library. A brush filter can be used to select a time range in the time line. This interaction updates the songs and genres shown in the scatter plots as well as the size of the bars in the legend to reflect the library composition over the selected time range.

exploration of library composition in terms of sound (scatter plot axis values) and genre composition (which points appear on the genre scatter plot, line chart fill coloring, and legend bar lengths).

An additional goal of our visualization is to provide an avenue for music discovery. When a user clicks on a song or a genre an embedded player for that song or an embedded playlist representing that genre will appear on the page. This allows users to uncover new songs and artists through the genre playlists that are related to the music in their library and the current selection they’ve made on their data. To drive further exploration of their data, users can Control (Windows) / Command (Mac) + Click on points in the song and genre scatter plots to view their counterparts in the other plot. Clicking on a song will show all of the genres that that song belongs to in the genre plot, and clicking on a genre will show all of the songs that belong to that genre. This allows for a user to uncover insights into how the songs and their artists are connected through different genres, and to also uncover how the artists of a song are spread across the music spectrum.

4 DISCUSSION

We hope that our visualization can provide unique insights into a user’s library and offer a path to explore and find new music. The main path we imagine a user following when using our tool roughly follows 4 steps: 1) subset their data using the time line, use the legend to remove or add an umbrella genre, or change the feature axes plotted; 2) examine overall trends and clustering in the subsetting data or using the new features; 3) using the tool tip feature, explore the new songs and genres that have appeared due to the selection made or examine interesting outlier points (songs with feature values that deviate far from their library norm); and 4) use the click-to-preview feature to uncover new music or revisit songs in their library. Figure 3 outlines one realization of this interaction pipeline, focusing on comparing genres in a user’s library immediately after account creation and within the recent past.

5 FUTURE WORK

The most significant limitations of the tool we have built arise from the restricted availability of individual music listening data. In particular, temporal information about listening history is limited,

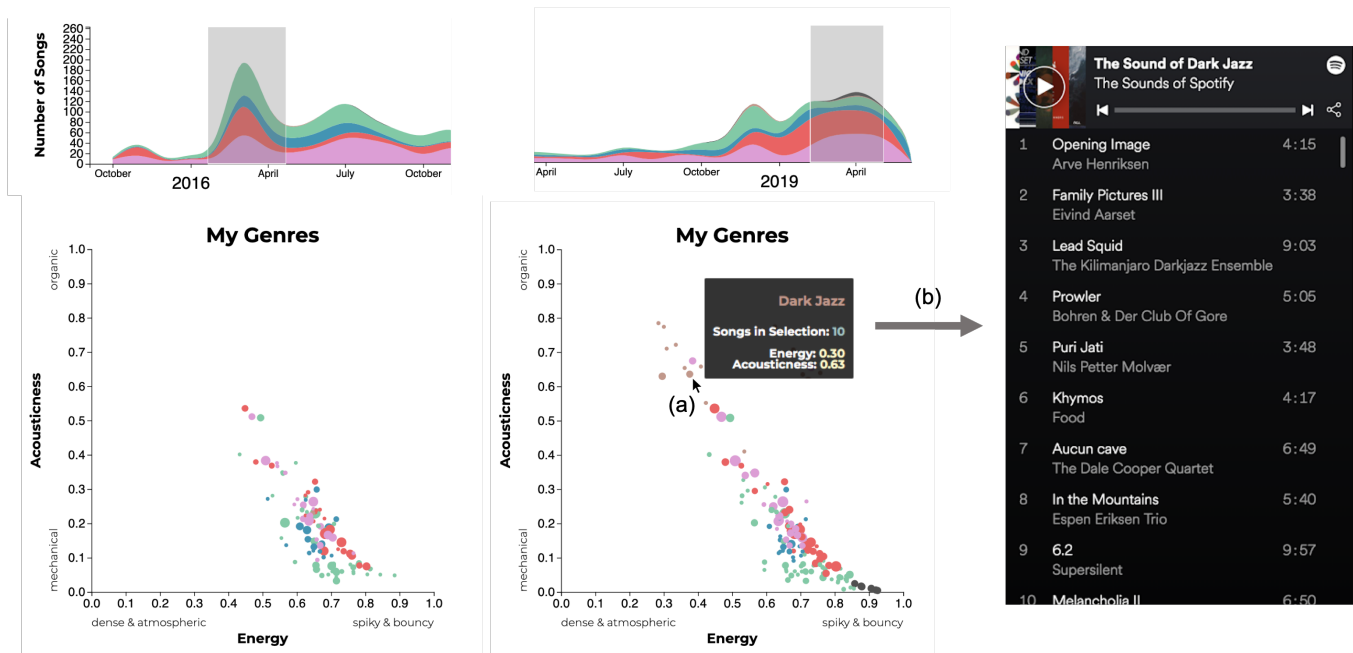


Figure 3: By brushing on the timeline, users can gain insight into how their library has evolved over time. Selecting a region early in time reveals the genres that the user was initially exposed to (left). Selecting a more recent region at the end of the timeline reveals that the genres the user is adding to their library have become more varied (middle). A user can gain discover what new genres that have added by hovering over them (a). A tooltip shows the name of the genre, its average song attribute values, and the number of songs of that genre in the current time selection. Clicking on the genre (b) produces a playlist created by Spotify that represent that genre (right).

with only the most 50 recently played songs for a user available through the Web API. Additionally, our pipeline relies on users curating and maintain their libraries in a centralized location (the Spotify saved songs library). While talking to potential users, we've found that many people don't think about their music in this way. They frequently store music in separate playlists instead of in one centralized location, or rely on the search functionality of Spotify to be able to look up songs on demand, accessing and playing music without ever saving it. Furthermore, many can't take advantage of the Wayfinder at all because they use streaming services other

than Spotify, or have their listening data on another platform like Last.fm.

Integration with data from other music listening platforms like Last.fm, and more detailed information about an individual's streaming history would allow for a much more revealing exploration of musical tastes. Information like number of plays, play timestamps, or location information could be used and integrated into our visualization to explore how our musical listening habits are a reflection of our broader lives.

REFERENCES

- [1] Glenn McDonald. 2013. Every Noise at Once. <http://everynoise.com/>