Nick Lasinsky and Jo Gentile

Spotipy Lab Musical Questions and Report

Setup:

Any inquiry into data needs a set of driving questions that animate the goals and strategies used for parsing, selecting, binning, and displaying it. These questions focus the dataset into conclusions, molding and shaping it to yield something valuable or interesting. Our set of questions, while a bit imprecise, proved to be an interesting way of splicing our group's playlist data, and allowed us employ a nice variety of coding strategies.

Our Musical Questions:

- What sort of music do people in our demographic enjoy listening to?
- How do we imagine the musical taste of our demographic?
- To what extent do playlist creators (listeners) imagine the preferences of their demographic differently than the real preferences that group displays?
- What qualities differ between our imagined tastes and our genuine tastes?
- What overlap is there between our imagined tastes and our genuine tastes?

Methodology:

It is important to acknowledge at the start that our set of questions necessitated a few subjective choices when it came to splicing the data. We relied in part on our own judgment of personal statements and their intentions, but we hope that there is recognition that any approach to musical data is going to require subjective judgements and choices about where to draw lines, what information to use, etc; our project was no different.

A short note on division of labor is also appropriate. Jo graciously worked on the nuts and bolts of the coding, figuring out our functions, troubleshooting, and generally making sure that the digital end of things ran smoothly. Nick, meanwhile, toiled on the musical questions, and composed the present report, attempting to interpret and guide the coding as well as he could.

This project required members of our age group (to be referred to from here on out as 20-Somethings) to select ten songs from a larger 40 song dataset. Many of the students who crafted playlists out of this set described making their choices based on liking the music, and we sought to isolate these selections to form a "mega-playlist." Some students noted that "For this playlist, I started going through the list of 40 and added the songs that jumped out to me - I really like the genres represented throughout," while others wrote that "I picked some of my favorite tunes." Though this was a semi-subjective process, we tried to stick to those playlists that were

clearly formed out of songs that drew people in the class, and ended up with eight that fit this measure.

What makes this dataset so powerful for our musical questions is the fact that a second selection of playlists were submitted, which featured a number of songs chosen in an attempt to imagine the lives of 20-Somethings. While this was, again, a subjective process, we tried to stick to songs that imagined the lives of our demographic, with titles like “Chilly Friday in Philly,” and “Aidan’s Shower Playlist.” While some of these collections were based off of real music that real 20-Somethings listened to, most of them were depictions; the sort of music that people believe embodies 20-Somethings.

These two groups, which we referred to as “Whattheylike,” (for that music curated from the 40 offered songs), and “Thinktheylike,” (those songs that we imagine 20-Somethings listening to) formed the heart of our inquiry.

A Load of Code, Part I: Gathering and Graphing

While more detailed explanations of our coding process are available in our markdown cells, it is useful to explain here some of the more wide-ranging implications of that code; what we did, how we fumbled, and the conclusions we were able to reach. Our first challenge involved gathering and displaying the data. It was fairly easy to assemble our two compiled mega-playlists. Next, we worked on graphing the data, sorting it by measures like “energy” and “tempo” on a simple X-Y graph. Ultimately, this way of displaying data was less than clear; it was difficult to notice any difference between two ascending lines of dots.

A more useful form of graphing was devised by Jo, who realized that a simple histogram, displaying both Whattheytlike and Thinktheylike, would help us to see the variation between the musical qualities of the data. Another useful graph was the simple scatterplot, color-coded to our two large playlists; this helped to show that many of the Thinktheylike songs were more energetic than Whattheylike tracks.

A Load of Code, Part II: Radar and Relationships

Next, we moved on to more elegant forms of data display, through the radar plots and Louvain communities. While the radar coding took some work (see markdown cells), we eventually got a few plots to display. The biggest questions with radar plots revolved around which metrics we wanted to show (“energy, “danceability,” etc.) and how big of a sample we wanted to take from our mega-playlists for graphing. Eventually we decided on cutting the mega-playlists in half, which produced a plot that was both navigable and readable, and settled on energy, danceability, instrumentalness, valence, liveness, and speechiness as metrics. Once again, these plots seemed to suggest that Thinktheylike was more energetic than Whattheylike, and showed that Thinktheylike was also considerably more instrumental.

Louvain communities presented an even bigger challenge. After much work, Jo was able to get a system of communities up and running that showed the relationships between various artists. Some artists showed up as being in common between Whattheylike and Thinktheylike, and it was also illuminating to see the connections between artists within the playlists. An additional network was also crafted, which showed those playlists which shared artists as being connected; this was helpful for visualizing where the connections existed within the two groups.

Conclusions:

Ultimately, it is difficult to draw definitive conclusions about the psychology or portrayal of 20-Somethings based on such a limited dataset. But we can see that there are clues, even here, that people may imagine the music of 20-Somethings as being more energetic and instrumental than their genuine preferences. Unhinged, frenetic, vibrant—these are common adjectives used to describe this decade of life. But perhaps 20-Somethings, like everyone else, enjoy cooling down as much as they enjoy blowing off steam. The entire exercise has been a fantastic opportunity to explore the data and the tools at our disposal to analyze it. While our findings may not quite pass a peer review, the experience was nonetheless enlightening and enjoyable.