Evolution of R&B

**Introduction**

**Target audience**

**Leading idea**

Our original idea was to evaluate lyrics’ positivity across time and different genres of music. We took a list of songs from Billboard top 100 songs for each year. We were able to find the top 100 songs on different Wikipedia pages, so we could easily scrape this list. We used the Genius API to get the lyrics from all of these songs. We could then try to evaluate the positivity of those lyrics and this is what caused problems and why we changed our project. We planned on using natural language processing techniques to be able to give a positivity score from -1 to 1, and we tried using the Vader sentiment analysis tool which is really used for evaluating tweet, but unfortunately did not give good results for lyrics. We tried splitting the lyrics by groups of 4 lines and averaging the score for each group, or even doing this on each line in the lyrics, but the analysis tool often gave back results that were either very positive or very negative. Taking some specific songs that we knew in the dataset, we tried to see if the result was anywhere close what we thought would be the score of such lyrics, which it was not. The sentiment analysis tool is not supposed to understand the subtleties of song lyrics, which can include poetic formulations, that could mean nothing if not in context of the song. We had the example of “Here Comes the Sun” by the Beatles, we the eponym line gave a neutral response, obviously because commenting on the weather as a statement is pretty stale, but if we as humans analyze it as lyrics, we can get the idea that this is a very positive thing. The Vader sentiment analysis is good when using simple words like good, bad. All of this meant that our project idea was going nowhere, so we had to find something new.

When scraping the songs from Wikipedia, we scraped their respective genres as well, but we wanted to generalize them so that we had a dozen of different genres ranging from *Rock* to *Hip-Hop*, passing through *Jazz*, *Metal*, *Disco*, *Funk*, etc. We had the list of genres we found on the Wikipedia pages of the song, and we mapped them to those more general genres. We did this with our knowledge, and if we found a style that we did not know, we went and listened to the songs to find which genre we would label them. From what we knew, we simply put *R&B* in the *Hip-Hop* genre, because this is what *R&B* of today comes the closest to. To our great surprise, when we plotted the curves of lyrics positivity by year and by genre, we saw that there were some data points for *Hip-Hop* in the 1950s and 1960s. After looking a bit why this happened, we saw that this came from the *R&B* genre that was originally closer to *Blues* or *Rock’n’Roll*, and that this term changed meaning quite a bit across time. Once we knew our original project idea was going into a wall, we figured it could be interesting to describe the evolution of *R&B* across time. This was a good idea because we could have a story to tell, starting from the origin of *Rhythm and Blues* in the 1940s to *Contemporary R&B* and *Alternative R&B*, see their difference, and try to explain the shifts of trends.

**Datasets**

We could not find a dataset that had exactly what we wanted, because it was too specific, so we had to build one of our own. We let go of the idea of using Billboard lists after talking with the teacher because he told us that a lot of visualizations was using these kinds of set of songs, and he wanted us to do something of our own. We needed to find a list of songs that could go under the R&B genre, for all years between the 1940s to today.

Finding this was also too specific, so we had to do research on what genres was labeled as subgenres of R&B, and we came up with a list, which included some genres such as *Jump Blues* (for the oldest one) or *Alternative R&B* (for a more recent example). In the mean time, we gathered some information about the genres, like how they were introduced or when they peaked, to have a little story to include in the visualization (even if we originally did not know how we would include such a story at first). We found a website called [rateyourmusic.com](https://rateyourmusic.com) that listed released albums, singles, compilations and so on by music genres, where the genres were really what we were looking for. We then proceeded to scrape the lists to gather a small database of songs.

The easiest way we found we could compare two songs was with the features that were able in the Spotify API. These were originally provided by Echo Nest, a company that was later bought by Spotify, and consisted of small analysis of songs. With the Spotify API, you can request a track’s features, and it gives you back a JSON file with information about the song key, its time signature, its mode (major or minor), its tempo and 6 more “obscure” features, namely valence, energy, speechiness, danceability, instrumentalness and liveness. Here are the descriptions of these terms as provided on the Spotify API website:

**Valence**: *A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).*

**Energy**: *Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.*

**Speechiness**: *Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.*

**Danceability**: *Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.*

**Instrumentalness**: *Predicts whether a track contains no vocals. “Ooh” and “aah” sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly “vocal”. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.*

**Liveness**: *Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.*

From the description, we could already have some idea as to which features will be more important to set apart the different subgenres, and which will not so much, for example, most songs are sung, and the speechiness is not often close to 0. Either way, we now had a list of songs that could be labeled by a subgenre, and to which we attributed features, we had to find what we wanted to show, and more importantly how.

**Visualization**

We first ideas about simply plotting the evolution of features by genre and by year, to see that, on one hand, the genres tend to not have the same values for the features, and on the other hand, the subgenres may have themselves evolved across time. This was okay, but we thought that the result would not be very interactive. So, we had to review how we wanted to plot the results. We had an idea of a scatter plot where we would plot each song a point, color-coding the points depending on the subgenre, the coordinates of the point being two of the Spotify features. Since we wanted to see the evolution of the subgenres across time, we quickly had the idea that we needed some sort of timeline that we could interact with, that will change the year from which we pick the songs which came out this year, and that would redraw the plot. Figure 1 is a sketch of what we originally had in mind. The red arrows on the timeline represented years where we wanted to include some additional trivia about R&B such as a change of name for a style, the appearance of a new style, or the release of an album that had a great influence on a subgenre.

For the first milestone of the project, we digitized the sketch by redoing it on [draw.io](http://www.draw.io) and adding some other ideas. The first thing is that we say which year we are plotting on top of the graph. We also wanted to show which song the point corresponded to, so we had the idea of a card popping up when we went over the point with the mouse. The card would contain the picture of the album, the artist name and the title of the song. We also put what we figured would be a way to select the features, namely a drop-down menu. The additional trivia was also put on the right of the vis, with a link to the source of the information. The resulting draw.io file is shown on figure 2.

|  |
| --- |
|  |
| *Figure 1. Original sketch for our scatter plot with the timeline* |

|  |
| --- |
|  |
| *Figure 2. Digitization of our scatter plot idea* |

The idea began to be stable, so we started implementing it with Javascript and D3 as we learnt during the course’s exercises. The first implementation was done without the timeline, but it was done so that linking it to a timeline would be easy (for example, there was a function to change the year of the plot, that would be called by the timeline, which was implemented but not used at first). The plot also had no way to change the features on the axes, although it was doable in the code. We had a HTML page with only a SVG inside, where we would draw everything. If we hovered a point, then the card would show up as expected. The information cards were statically drawn on the right. As we can see on figure 3, we experienced with a dark background at first.

|  |
| --- |
|  |
| *Figure 3. First implementation of our scatter plot with Javascript and D3, focusing on a song* |

An updated implementation can be seen in figure 4, and it has some small differences with the original plot. There is a small key of the top-right of the plot that was hoverable and that would show the cards with the information on the right. We reverted to a white background because the colors were showing up better on this background. We also added the small legend on the bottom left.

|  |
| --- |
|  |
| *Figure 4. Updated scatter plot, focusing on the key* |

A rather important change would then come as we did not originally thought of it. The Spotify API includes a preview of (almost) all songs which lasts 30 seconds. We thought it might a good idea to include this preview somewhere, so that people could have an idea as to why the point was at this place in the graph, for example see why consisted of a high energy value. We had the idea to transform the points when focused on little play button, where pressing on it

**Final product**