**Intro to Data/Text Mining: Final Project**

**Proposal**

Alex Koener, Efrain Ramos, Ryan Hermann,

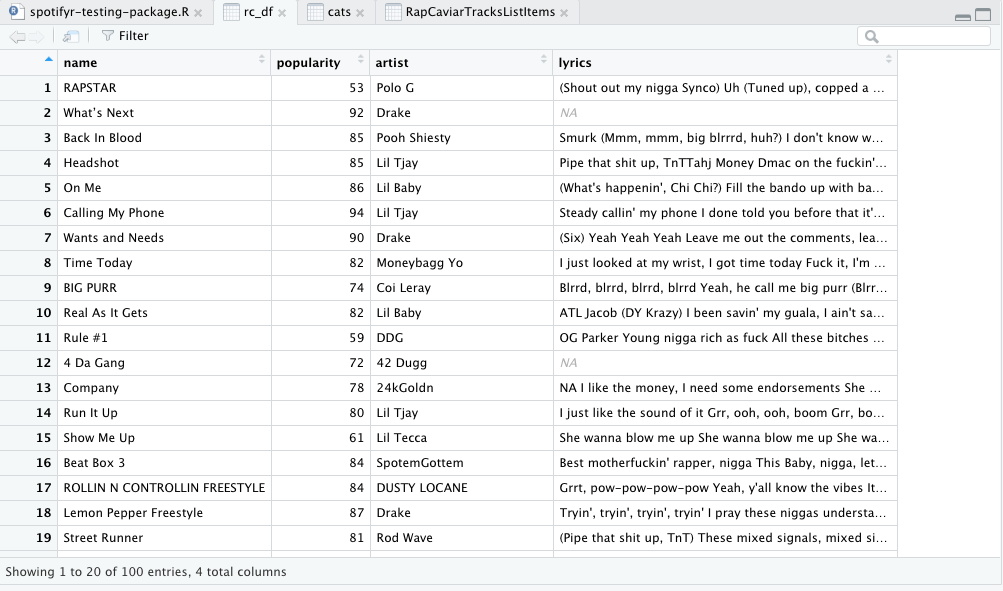
Victoria Garcia, Darius Cuthbert, and James Clancy,

For this assigned project, our group decided to investigate popular playlists from Spotify. Our data will be sourced from the most popular playlists of different genres, and we will perform sentiment analysis on their lyrics.

Once we have sourced the top playlists from each genre, we will extract the most popular songs from each playlist. We will use these songs as entries into a dataset which will include the song name, lyrics, its genre, populator score, a binary “Album/single”, and track number if its from an album. The genres we have chosen to examine are Pop, Rock, Rap, Metal, and Country which are the most popular on the site.

Once our dataset is formed, we will perform sentiment analysis on the lyrics of these tracks, using the skills we obtained from week 7. Following this analysis, we will create visualizations showing the correlation between sentiments and popularity of genres, playlists and Spotify as a whole. To assist with these visualizations, we will create word clouds in shiny applications, (Week 5 & 11) that will allow viewers to see what words are frequently used in each category.

There are very few concerns that we have about this project, among them however, we are concerned about using certain slurs and curses in our analysis, and possible conflicts between word clouds and shiny apps. Despite these concerns, our group is very excited to work on this project, and we have a substantial amount of the work already done, included below is some of the data we have already collected.



Figure