Analysis / Challenges:

We are seeking to describe the change in sentiment of top-100 songs from 2000 to the present, both overall and according to multiple different emotions such as happiness, sadness, anger, and fear. Specifically, we want to investigate whether the sentiments present in these songs mirror the overall global sentiment of the time. This is documented through certain impactful events which would revolve around ideas such as geopolitical conflict, natural disasters, political public opinion, and pop culture trends. Songs will be extracted from webscraping the Billboard Hot 100 list of each year in our time frame, and the information will be used as input into calls to a Python wrapper of the Genius.com API to retrieve the song lyrics. Emotional sentiment would be measured using polarity scores from NLTK and nine emotion flags as determined from the functions in the NRCLex library. These calculations will be applied on individual words and combined verses.

One challenge we foresee is the subjectivity of emotions extracted from the emotion flagging library. First, only nine emotions of all those on the spectrum are flagged, and only a select few of the words in each song have an emotion associated with them. Further, we need to assess the viability and legality of webscraping the Billboard Hot 100 website. This may constitute issues with data privacy, so in the event these requests are unable to be made, this will add significant time for finding the songs for our analysis. Finally, the interpretation of world events is based so far on our own knowledge and life experiences. There don’t seem to be clear years where more impactful things happened as opposed to others, and this may present bias to this comparative portion of our analysis. Finding some dataset and quantifying the historical impact may be a solution, but this would be difficult to find exactly as described here.