Basic Info

Olivia Bigelow | u1188824 | u1188824@utah.edu | Data Science
Karena Klinkenberg | 1384933 | u1384933@utah.edu | Data Science
Nicole Forrester | u1310549 | u1310549@utah.edu | Computer Science/Applied Math

Background and Motivation

Music is an important part of the human experience that can have an impact on our mood, thoughts, and emotions. It brings people together, helps individuals connect with an artist's feelings and stories, and in some cases, even provides healing to those dealing with mental illnesses[1]. Some people choose music that correlates with their current mood. For example, someone who is sad might choose to listen to slow, melodic music to feel less alone (mood congruence). Others might choose happy music because they are in a good mood or spending time with friends (personal/cultural). Additionally, some may put on calming music when they are feeling anxious to regulate their mood (mood regulation).

So, what happens to music listeners during historical events such as presidential elections? During elections, there is heightened anxiety across the nation, some even get election stress disorder[2]. What happens to listeners during global pandemics? During the first year of COVID-19 anxiety/depression went up by 25% globally[3]. What about during social change movements? Or wars? Or during an economic downturn?

Music evolved significantly over the decades. We also want to consider how the type and mood of music has changed over time, as well the prevalence of each type. Are people listening to more sad music now compared to happy music? Or vice versa? Distinguishing what types of music people listen to during different time periods may help us to generally map how people are feeling based on current events.

[1] https://my.clevelandclinic.org/health/treatments/8817-music-therapy
[2] https://newsnetwork.mayoclinic.org/discussion/is-election-stress-disorder-real/
[3]https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide

Project Objectives

- How do historical events impact the type of music listeners listen to?
- What category of music (sad, happy, energetic, calm) do people listen to the most/least over time?
- Can we get a general sense of how people are feeling based on current music data?

By answering these questions, we can understand the broader impact that music has during historical events. We can also observe how people's music habits are changing over time. Finally, based on these visualizations, we may be able to examine and analyze recurring patterns, which may help us to get a general consensus on how people across the world/nation are feeling during specific periods of time.

The benefits are knowing trends in people's emotions, knowing how music is affected by events, and gaining historical perspective.

Data

The data is sourced from 3 places. Our first data source is a kaggle dataset titled 278K Emotion Labeled Spotify Songs. This data contains a large volume of Spotify songs, along with various associated traits found from spotify. This is our main dataset. The second source is another kaggle dataset titled Billboard “The Hot 100” Songs. This dataset contains the top 100 songs in the United States every week from 1958-2021, along with their previous position on the top 100, and weeks on the top 100. Unfortunately, these 2 datasets do not have a direct mapping, So our third dataset is straight from the spotify API. This data will be pulled in advance, as to avoid API rate limits. This will contain a mapping of song name to a unique spotify Identifier, which will allow us to map the Billboard Top 100 data to the emotion labels found.

On top of that, we need historical event data, though that likely won’t be sourced through a particular dataset, as we won’t need too many data points.

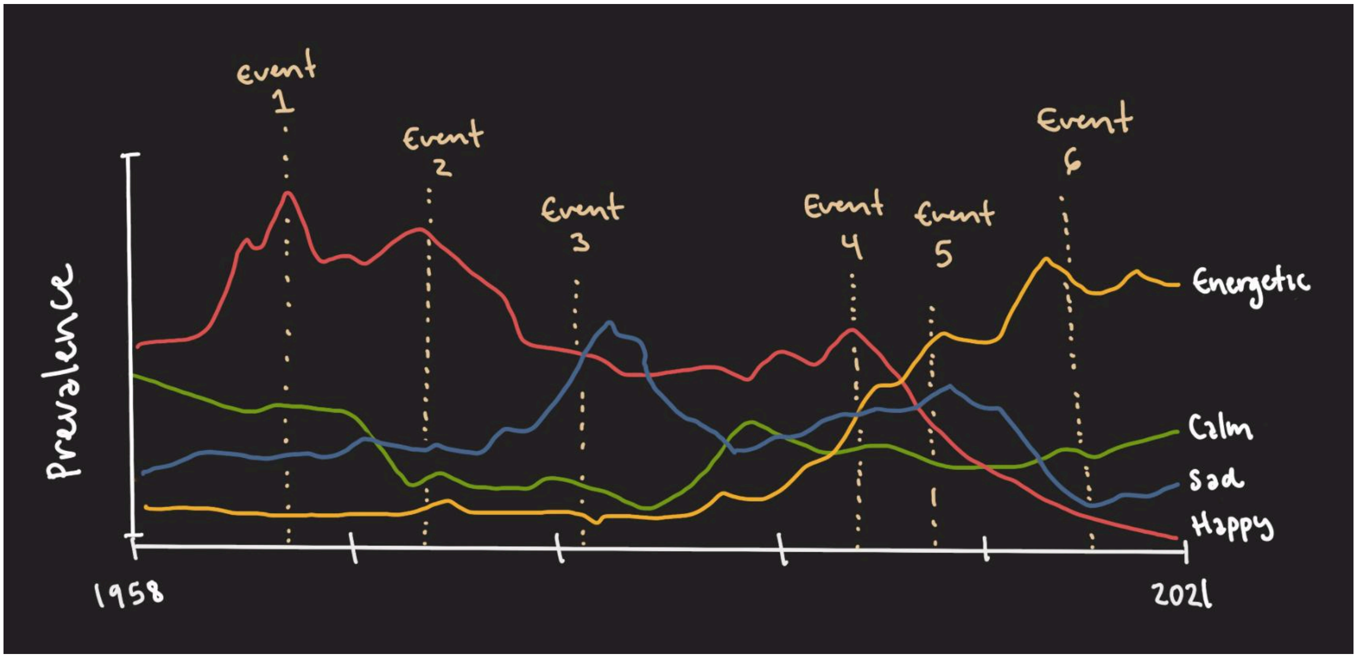
Data Processing

In doing previous work with the spotify api, the data is very clean, and there aren't null or incorrect values. The top 100 and emotion labels are clean in regards to the only values we care about (date, name, artist) and (spotify id, emotion). A big part of processing will be the merging of these data sets, that will take a good amount of processing power, or some batching, first querying the spotify api for each song found in the top 100, and saving a dataset with the top 100 song, spotify api, and date. Then we need to pair this merged dataset with the emotion labeled songs. If any matches are missed, the creators of the emotion labeled songs created a machine learning model, based on spotify analytics to determine the emotionality of a song with an accuracy in the high 90s. As the songs on the top 100 are very popular, they are highly likely to appear in the dataset.

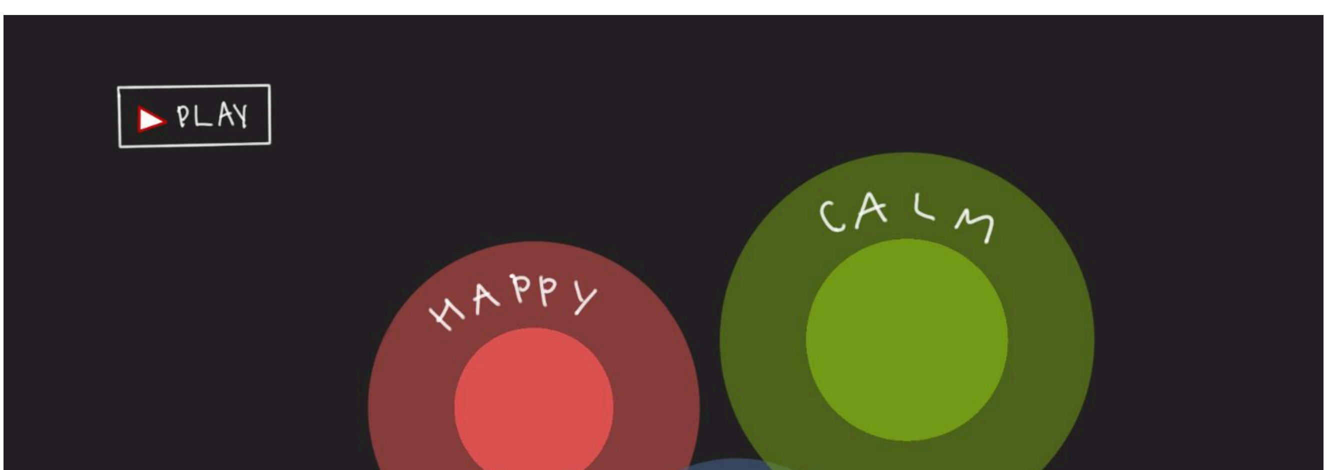
From there, we have a fully merged dataset of the emotion labels of the top 100, we can use this data for the visualizations, and divide it up based on the major historical events.

Visualization Design

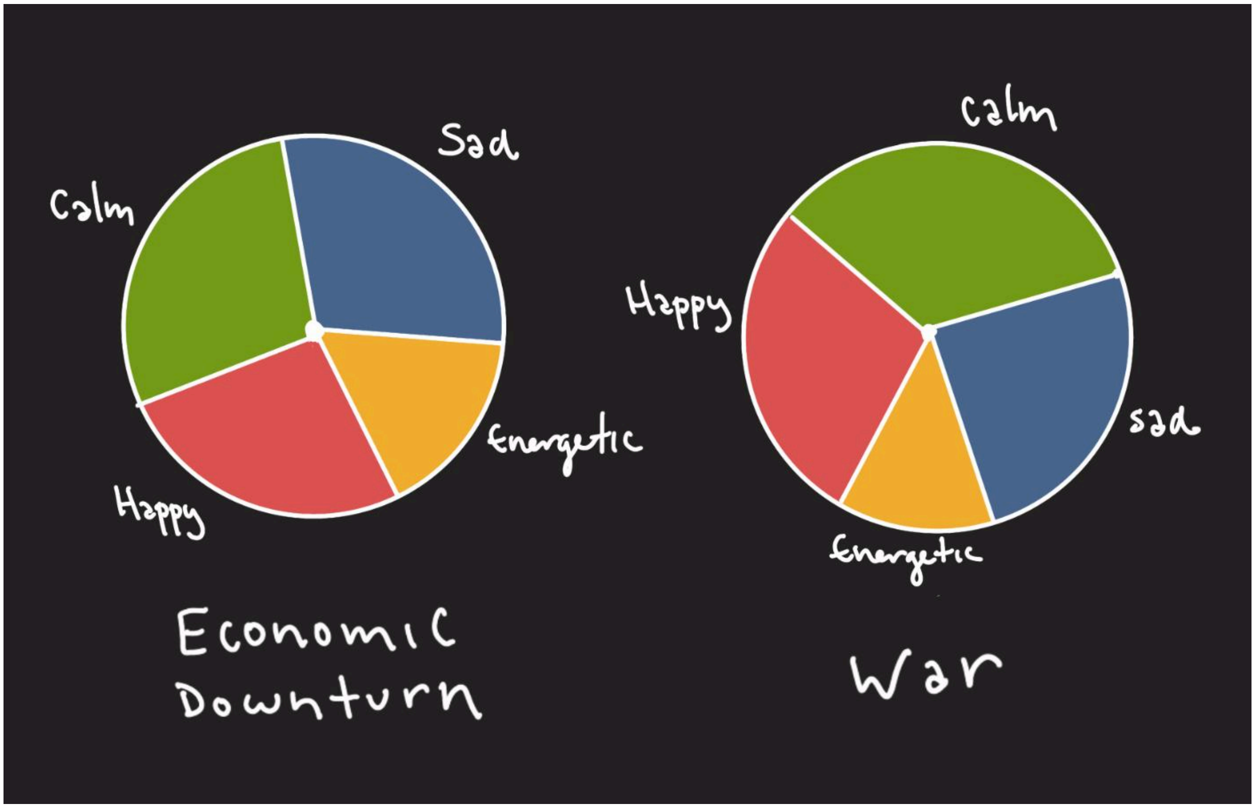
Our data has a time series element as well as a categorical element. The time series stretches from 1958 to 2021, on the granularity of weeks. The categorical elements are "Sad", Happy", "Energetic" and "Calm". We are using the Billboard Hot 100 data set to track the proportion of emotions in the top 100 songs by week over time. The goal is to show how the categorical emotion elements change in proportion over time, and to link certain historical events/periods with emotional listening trends. In order to show change over time, there are two approaches we can take. First, we could display all data over time and plot the changes continuously, creating a static visualization. This would have time on the X axis and proportion of top 100 songs on the Y axis. The benefit of this method is that the user could see the entire trendline across time, which could also illustrate the effects of an event that resonate in the weeks or months after its occurrence. It would also show the overall proportion of each emotion across time, and tell us more about the listening trends of our society on a macro scale.



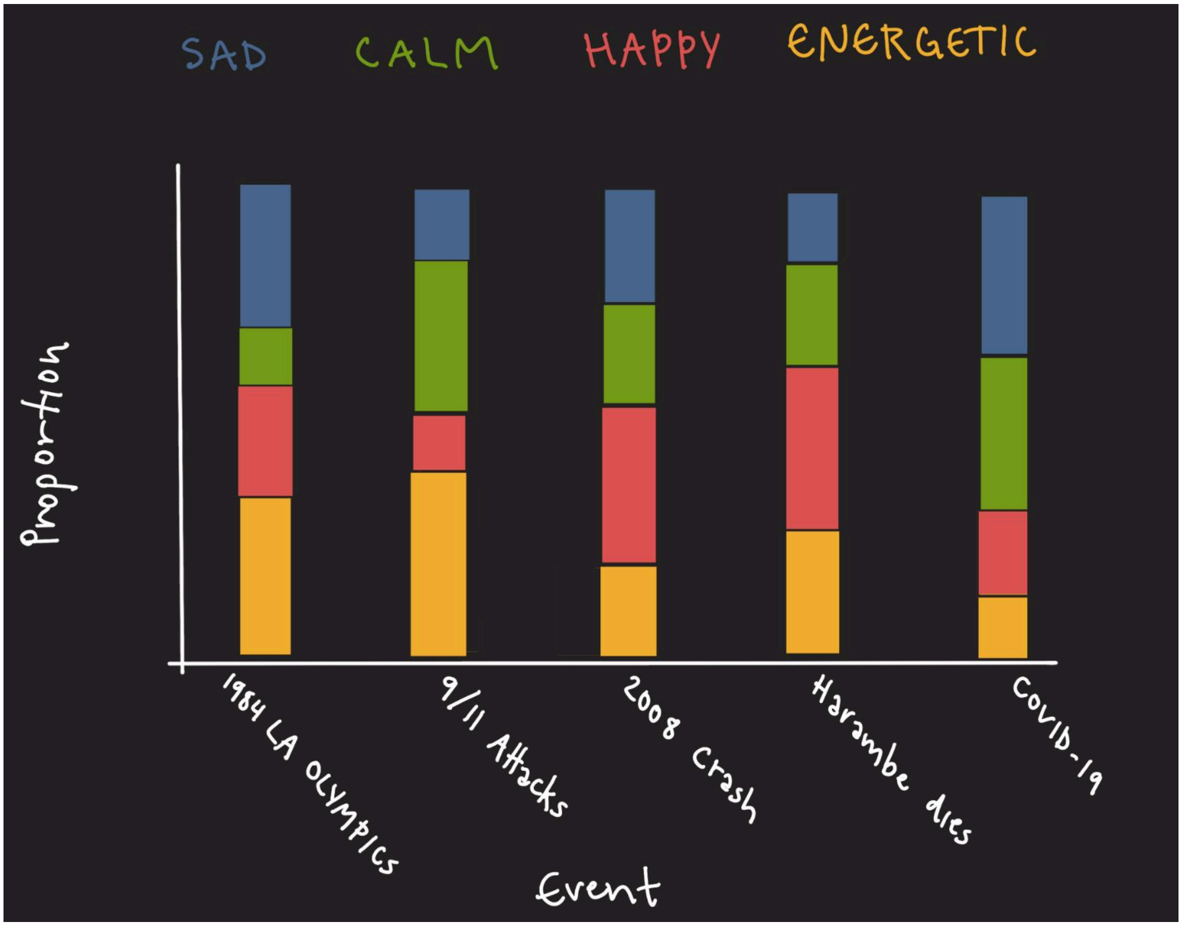
Second, we could create an animated design that shows the proportion change in top 100 songs as the years progress by changing the size of some visual element representing each emotion. The user would be able to move a cursor on the timeline (Y axis) and see the emotion elements change in size. The two examples shown below illustrate this idea, but take a visually different approach. The one with circles is aesthetically pleasing and the shapes are reminiscent of vinyl records, which plays into the theme of music. The downside is that humans have a hard time accurately conceptualizing the area of circles. The other visualization, using bars, is less artistic but offers benefits when it comes to comparing the relative proportions. The user has the y axis and the rectangle area to use a reference for the proportion of the emotion, which is more accurate to perceive than circles. Either design could be animated to show the fluctuating trends in emotion as time changes on the X axis.



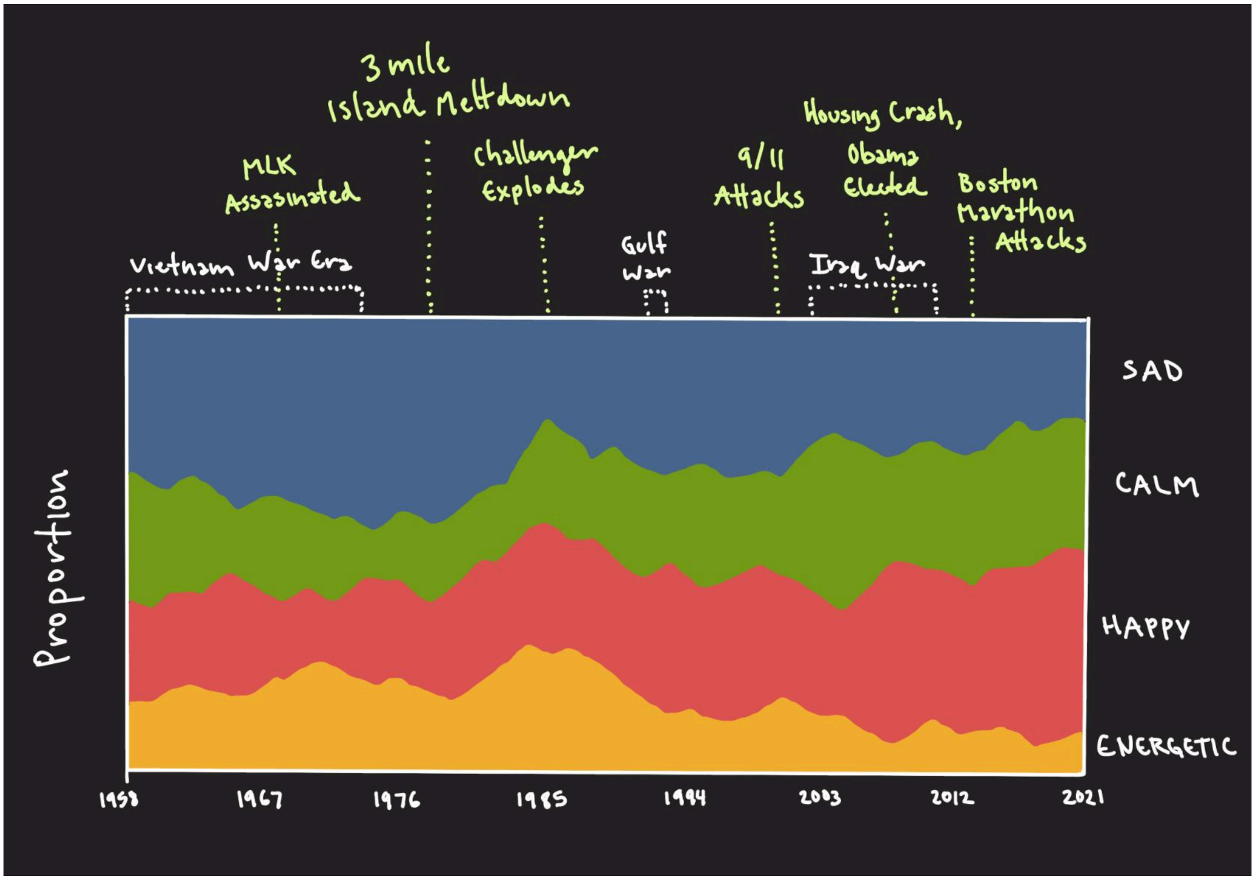
Another aspect of our visualizations we would like to include is joining the time series data with historical events and periods. Ie. to look at the overall proportion of emotion in the top 100 during times of war, economic downturn etc. This could be represented by bar charts or pie charts, and would be labeled by what time period they were representing. An example of a pie chart is shown below. The benefit of pie charts is that users are familiar with interpreting them, but the downside is that humans aren't the best at comparing the relative areas of slices. Also, the labeling is sometimes hard to position, especially if one of the slices is very small.



To show historical events in a visualization, we could have labels on the time axis that show remarkable events, such as the 2008 crash, 9/11 attacks, American-hosted Olympics etc. as was shown in the first visualization. Or, we could also have stacked bar charts that were labeled on the X axis by their event, and have the proportion of hot 100 on the Y axis. One such example is illustrated below. The bar chart makes it easy to compare the proportions of emotions during each historical event. The downside of this is that the events may have long term effects on listening trends, and a bar chart only illustrates a snapshot in time.



I think our ideas will be best displayed using a number of visualizations, however if we needed to condense our goals into a single visualization, it could look something like the visualization shown below. The upsides are that it encodes a lot of information and shows how listening trends are impacted by both historic periods such as prolonged involvement in wars, as well as by single events. The width of each color encodes the proportion of the Hot 100 at that date, and the scale being filled to 100% allows us to easily compare the prevalence. The downside is that it is crowded and messy, and we would have to severely limit the number of historic events and periods we chose to show so as to not over-clutter the visualization. Additionally, the top edge of each color might be interpreted as a trendline, which could be misleading since the width is what is actually changing in regards to proportion.



For simplicity, we will start with the line chart shown at the very beginning. This will be a static visualization that shows the changes over time in emotional proportion. It is easy to interpret, and can be extended to contain more details such as events.

Features that we absolutely must have include:

- A visualization that tracks the proportion of emotion in the Billboard Hot 100 every week since 1958 up until 2021.
- Some way of representing historical events on a timeline to illustrate the effect of major events of people's listening habits.
- Some way of representing time periods marked by a distinguishing feature, to show how certain national periods affected listening trends.

Cool but not critical features include:

- Interactive aspect to visualization that would show top 5-10 songs for that week in the dataset
- The option for a user to input a song and have it sent to the emotion predicting ML model (additional GitHub repo built off the dataset we are using) and have it displayed. Additionally, if that song was in the Hot 100 at any point, to highlight where on the timeline it is represented and the associated listening patterns at that time, or any historic events it may have been associated with.
- A ‘play’ button that starts an animation going through time and changing the size of emotion-proportion elements to show shifting trends.

Project Schedule

<u>Week 5-6</u>

- Karena, Liv, and Nicole - meet with TA and begin setup for visualization
- Nicole - get website running on localhost
- Liv - register Spotify API license, and setup Notebook for data collection and cleaning.

<u>Week 6-7</u>:

- Karena, Liv, and (some) Nicole - Collect spotify data and merge. Finalize historical events. Have all data formatted and ready for visualization. This can be done in an external jupyter notebook, for simplicity of merging the data.
- Create pages or sections for different visualizations on website and some basic visualization outlines (maybe with dummy data)
- Begin the process book.

<u>Week 9-10:</u>

- K, L, and N: Have a functioning prototype ready for the milestone. Pulling data into the website and being able to display all graphs.
- K, L, and N - Pick color palette and accessibility.

- K, L, and N: Write the ReadMe.

Week 12-13:

- K, L, and N: Iterate on peer feedback, finalize design and project book., Time allowing, add play button, and ability to add new songs, and newer data. (use flask)
- K, L, and N: host on a server (very low priority)

Week 14-15:

- K, L, and N: Film and edit screencast of project.