*As a music producer, I want to analyze the music that I have put up on Spotify to see what patterns and conclusions I can draw not only about the music that I have released, but the music I listen to as well. Therefore, I hope to use MATLAB as a way to visualize some of the data that can be produced using the Spotify API (Spotipy, Python). By graphing scatterplots, histograms, manipulating some of the quantitative variables produced from the API and creating metrics to create and measure variables, I hope to demonstrate my proficiency with MATLAB in a project that extends my interests and allows me to learn more about them.*

*Proficiency can be demonstrated by incorporating concepts that is taught within the traditional CS 1371 class, from simple variable and String manipulation, conditional clauses, method instantiation, and commenting throughout the code in order to give sufficient proof that I have learned the foundational programming skills regarding the programming language of MATLAB.*

A 3D scatterplot can be created that evaluates the relationship and occurrence of the quantitative values between a song's variables such as "Acousticness", "Loudness", "Danceability", etc.

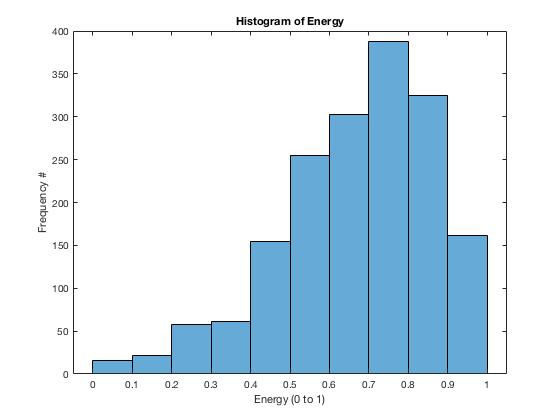
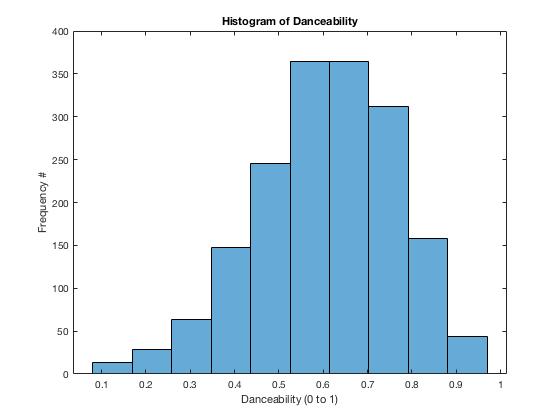
Histograms can be created that evaluate the change in the distribution of bins and their proportions based on a single variable (f.e. does the distribution of song Loudness increase with the increase of a song's Danceability? Use histograms to evaluate the change in distribution).

Also, a simple graph that is animated over the duration of the line's creation could be done. This could be used to visually represent trendlines that are predicted based on the data we evaluate.

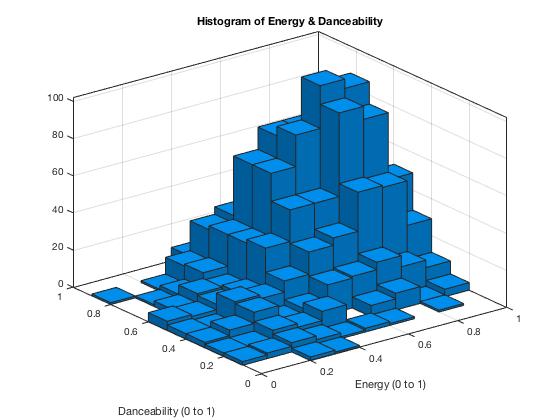
Multiple graphs of each type will definitely be created. This project is meant to be an exploration, where graphs are constantly created and remodified in order to learn more about my music trends and preferences. For example, if I were to discover that the scatterplot between three discrete variables reveal no new patterns, I would comment accordingly but still keep the code that performed that operation so as to revisit as well as to provide more code and results for you to evaluate.

All the graphs reproduced bellow were created using MATLAB.

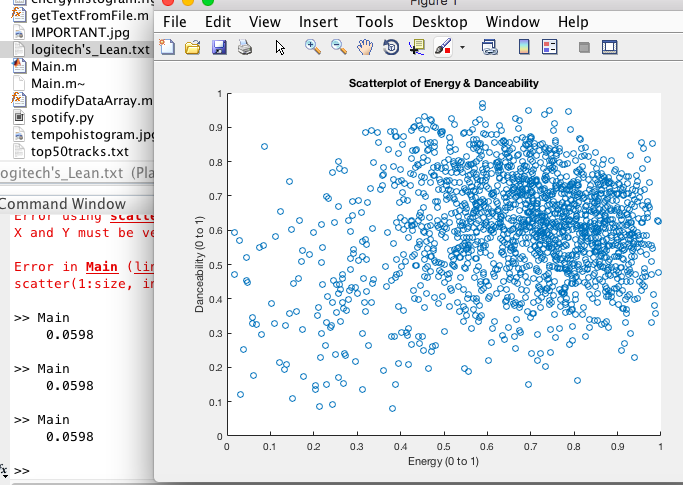
The numbers refer to either a playlist of 1743 songs, which is an aggregation of all my favorite songs that I like, or the playlist of 50 songs that represent my most listened to songs of all time via Spotify. This data was obtained using the Spotipy API in a Python script.



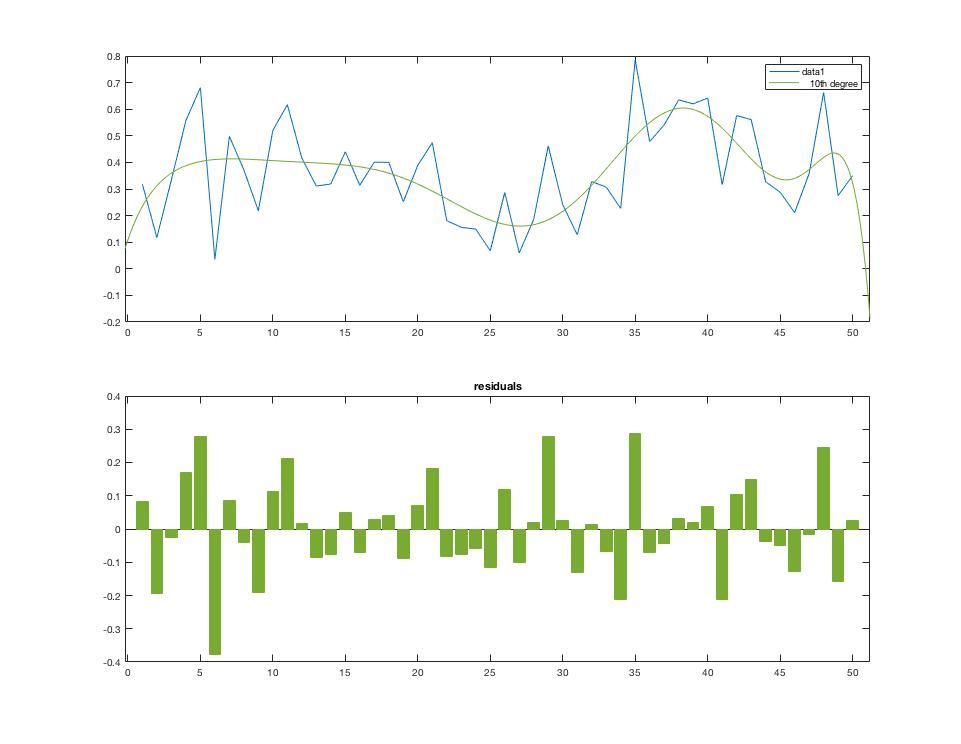
These histograms show that the distribution of my favorite songs is distributed left-skewed, with the average being more to the higher values of danceability and energy values, where 1.0 is very danceable and/or high energy.



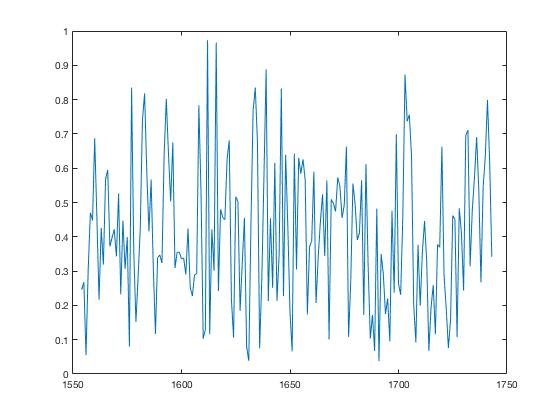
This 3-D histogram amalgamates the two above to give a better representation of that distribution. Most of the songs I enjoy are in the quadrant that is both high energy and of high danceability value.



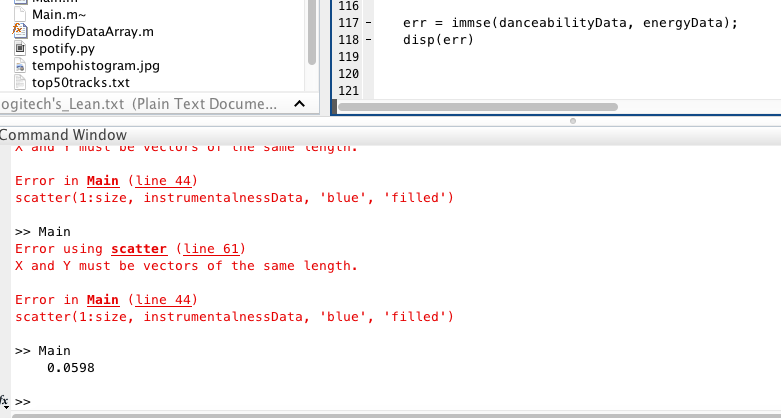
That idea of the quadrant is seen more easily in the above scatterplot. The quadrant of > 0.5 Energy and > 0.5 Danceablity is the most populated.



The first graph takes the valence (predisposition towards happiness) of the top 50 songs I like and graphs them in chronological order, basically making for a timeline of moods. If we treated 0.5 as the neutral mood (since it’s the median of 0.0 and 1.0), then most of the songs seem to fall a little bit below that. The 10th degree polynomial best fit line is created in order to take residuals to it which can corroborate my claim above as most of the residuals fall below the best fit line. This isn’t to say that I listen to songs that are more melancholy but that my taste reflects a palette of moods that vary and tend to cancel each other out, more often negating the higher-valence minded songs (which often tend to be pop).



This graph looks at the chronological order of songs that I added to my 1743 song playlist within the time domain of this semester, Fall 2018. It seems to have a consistent back and forth pattern without a noticeable trend of peaks or valleys which means my semester had a broad palette of song moods, which, if you take music moods to reflect actual moods, means my mood in the semester was balanced and had no strong periods of prolonged positive or negative intensity.



The mean squared error was calculated for the danceability data and energy data for the 1743 song playlist. This would essentially see if for the songs in that playlist whether the danceability value would be a good indicator for the energy data (if one could be associated with the other). The calculated value of 0.0598 is very close to zero and means that the regression between the two is close and the line of best fit is very close and possibly as good as it gets.

Other graphs were plotted (as seen on the Main.m file) that failed to produce any close or discernible relationships. Relationships such as those between tempo and energy, acousticness and liveness, etc. were explored. The most noticeable observations made was: my predilection for songs that are high energy and are very danceable, that the music distribution of the song tempos I listen to is normally distributed but is left skewed when looking at the distribution for the song danceability and energy. It is also implicitly understood based on the lack of relationships found for other Spotipy API variables that my music taste is varied enough in mood, style, and genre that causes no predilection towards a specific variable heaviness (balance in valence, acousticness prevalence, etc.).