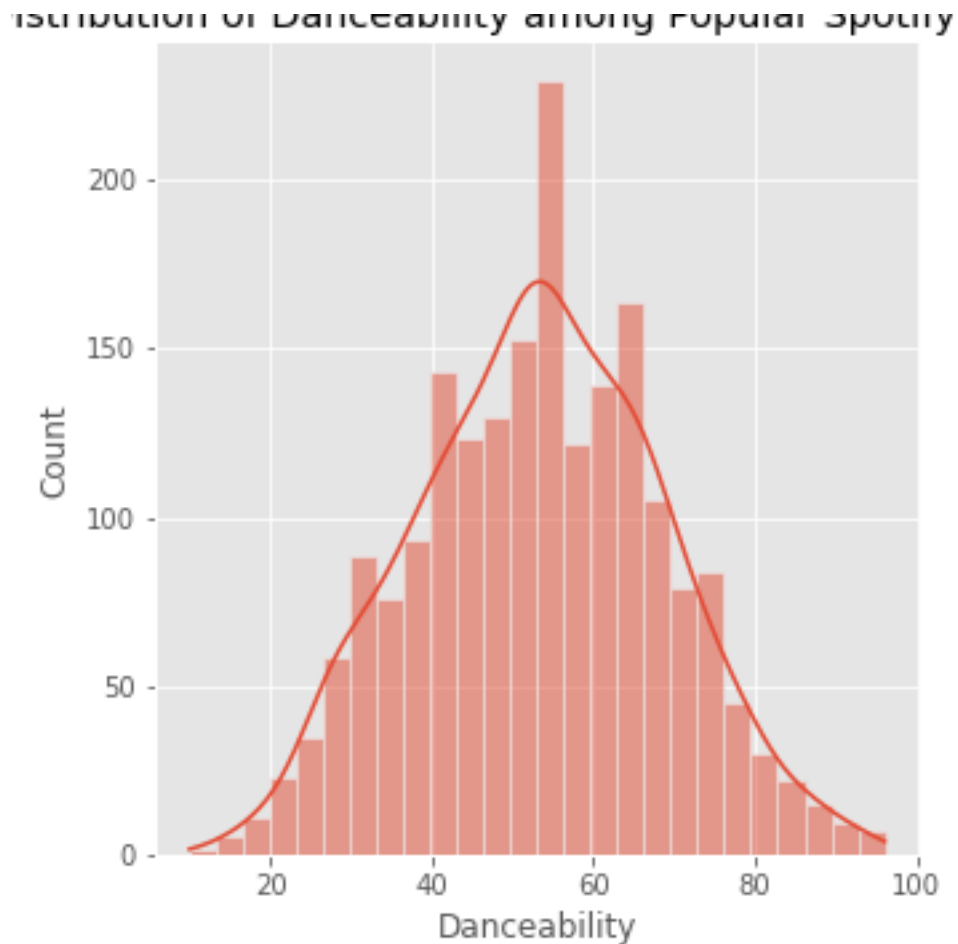# Problem Set 1 Response

Campbell Clarkson

1. Dataset and Figures

I am introducing a dataset I am working on with a student from the College of Computing. We are building a machine learning project for song recognition, so the figures here use acoustic data points from 2000 popular spotify songs.


In figure 1, this examines the distribution of "danceability" among the dataset, where it is clear it follows a normal distribution. This may be useful when we train our dataset to recognize a song with danceability as a factor. Does danceability have any bearing on popularity, though?

Popularity and Danceability
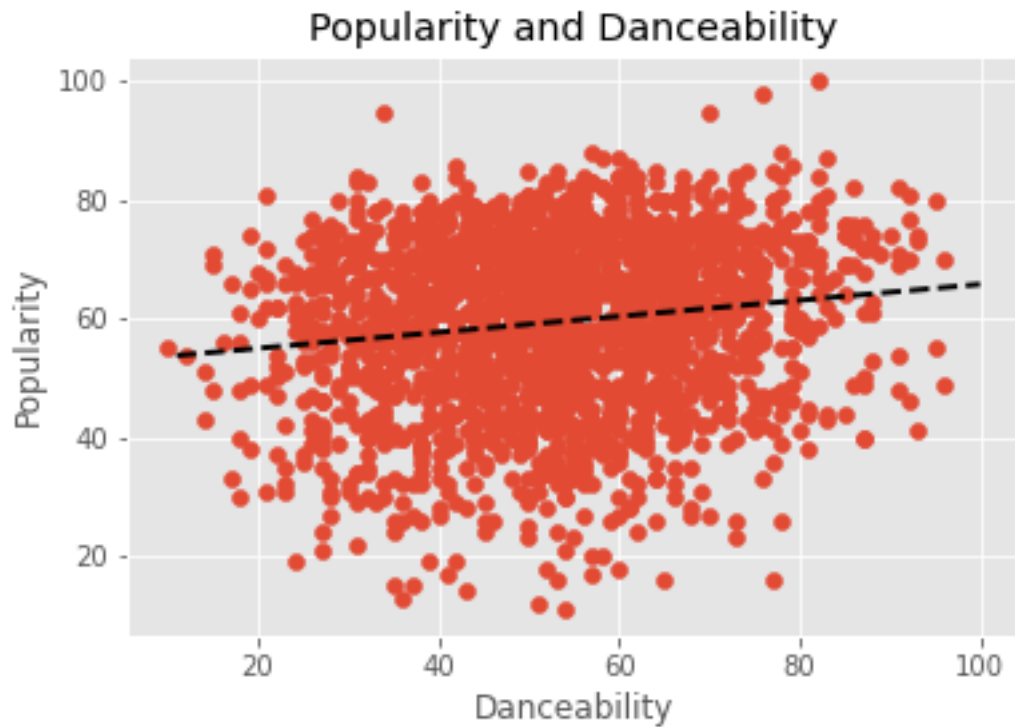
From this figure 1, it's apparent that the more "danceable" a song is, according to Spotify's metrics, then the more popular it tends to be. Let's consider some other metrics then, which may be useful. What happens when we look at both song tempo and the valence (in other words, how happy the song sounds) on its popularity.
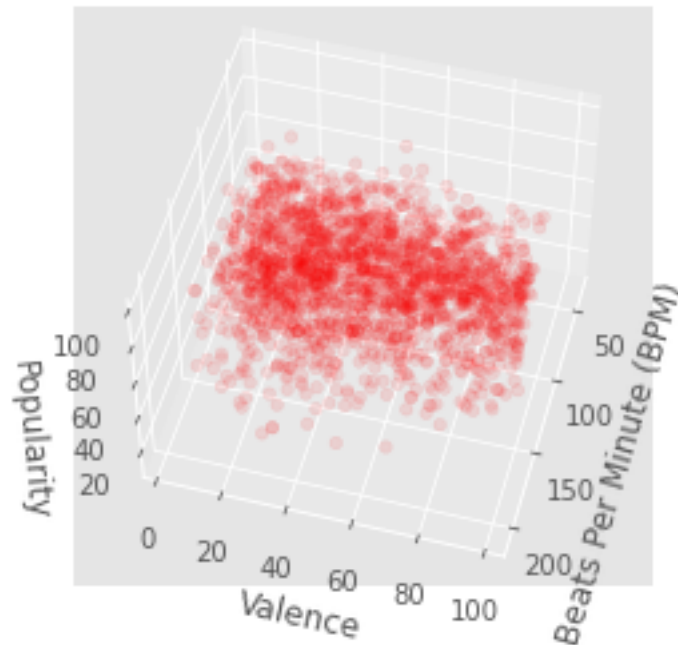
It's unclear whether tempo and valence together have a specific bearing on popularity when examining this plot. However, examining this data with a proper model could prove useful to see which qualities are most important for a song's popularity or the optimal song to maximize popularity.