STATS 101A Final Project

Name: Dong Han
UID: 604996609
Name: Zihan Gao
UID: 804923787
Date: 26/05/2020

Data description:

- The dataset we use is one that includes attributes for different songs from Spotify.
- Source: https://www.kaggle.com/nadintamer/top-spotify-tracks-of-2018

A. A Kaggle user named Nadin Tamer.

B. Initially, the data is stored in the database of Spotify's servers. For each song stored in the database, there is a collection of attributes describing the song including tempo, loudness, valence, energy, danceability, popularity, etc. The dataset was scraped and saved in Kaggle.com, link in the description above.

C. Due to her curiosity towards the relationship between a song's popularity and the audio features, Nadin Tamer tried to find a pattern in the audio features. Also, by collecting the attributes and analyzing it, we can have an insight on: What do these top songs have in common? Why do people like them?

D. She used the feature values of the top songs released by Spotify every year and collected the data attributes for different songs through Spotify's API and a python library named "Spotipy".

E. Our goal is to seek answers to the questions that which variable affects the danceability of a song most, how the variables interact with each other, how to predict the danceability of a song, what determines listeners' tastes, and how Spotify can improve user experience by better understanding what makes a song popular.