Project proposal

RiSK: Rhea Tejwani, Sue Zhang, Keena Gao

10/6/2020

Section 1. Introduction

The dataset contains statistics of the most popular music in the world over the years on the streaming service Spotify. Music is constantly changing, and as students who frequently listen to music and pay for a streaming subscription, we want to analyze the trends of popular music in our generation and the generations before us. Our research question is "how have characteristics of the most listened to music changed over time?" Our hypotheses are that popular music has shifted to be more diverse in these categories like acoustics, BPM, valence, etc. In the past, most songs were similar in speed and genre, however we hypothesize that popular music today has more variance and the range of values in each category is broader.

Section 2. Data description

This data set has 15 columns and 1994 rows. The observations in the data set describe the characteristics of the top 2000 most popular songs from 1956 to 2019 from Spotify. Title: title of the song Artist: the musician/group who performed the song Top genre: genre of the track year: year it was released Beats per minute(bpm): tempo of the song Energy: how energetic the song is Danceability: how easy the song is to dance to Loudness(dB): how loud the song is Liveness: the likeliness of the song being a live recording Valence: how positive a song is Length(duration): the length of a track Acousticness: how acoustic a song is Speechiness: how much spoken word is in the song Popularity: how popular a song is

The original data set was taken from the playlist on spotify Top 2000s on PlaylistMachinery(@plamere) using Selenium with Python. It was scraped from http://sortyourmusic.playlistmachinery.com/.

Section 3. Glimpse of data

```
library(tidyverse)
spotify_data <- read_csv("data/Spotify-2000.csv")
```

```
glimpse(spotify_data)
```

```
## Rows: 1,994
## Columns: 15
## $ Index                   <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,...
## $ Title                   <chr> "Sunrise", "Black Night", "Clint Eastwood"...
## $ Artist                  <chr> "Norah Jones", "Deep Purple", "Gorillaz", ...
## $ `Top Genre`             <chr> "adult standards", "album rock", "alternat...
## $ Year                    <dbl> 2004, 2000, 2001, 2007, 2002, 2004, 2002, ...
## $ `Beats Per Minute (BPM)` <dbl> 157, 135, 168, 173, 106, 99, 102, 137, 148...
## $ Energy                  <dbl> 30, 79, 69, 96, 82, 46, 71, 96, 92, 67, 5,...
## $ Danceability            <dbl> 53, 50, 66, 43, 58, 54, 71, 37, 36, 91, 44...
## $ `Loudness (dB)`         <dbl> -14, -11, -9, -4, -5, -9, -6, -5, -4, -3, ...
## $ Liveness                <dbl> 11, 17, 7, 3, 10, 14, 13, 12, 10, 24, 11, ...
## $ Valence                 <dbl> 68, 81, 52, 37, 87, 14, 54, 21, 23, 66, 31...
## $ `Length (Duration)`     <dbl> 201, 207, 341, 269, 256, 247, 257, 366, 22...
```

```
## $ Acousticness        <dbl> 94, 17, 2, 0, 1, 0, 6, 0, 0, 0, 88, 1, 2, ...
## $ Speechiness         <dbl> 3, 7, 17, 4, 3, 2, 3, 14, 8, 7, 4, 8, 3, 1...
## $ Popularity          <dbl> 71, 39, 69, 76, 59, 45, 74, 69, 77, 82, 49...
```