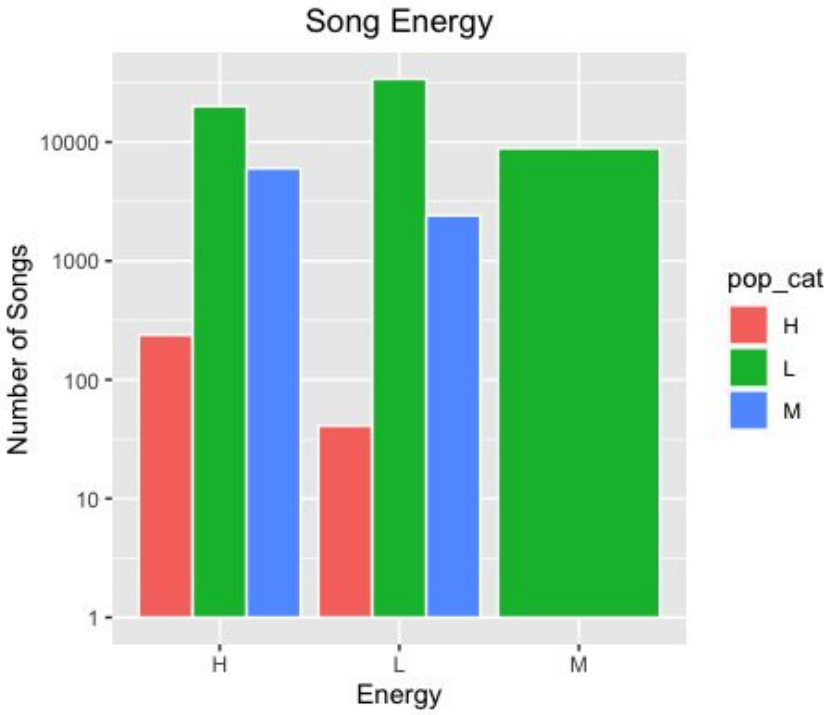

# Data Analysis of Spotify for Universal Music Group

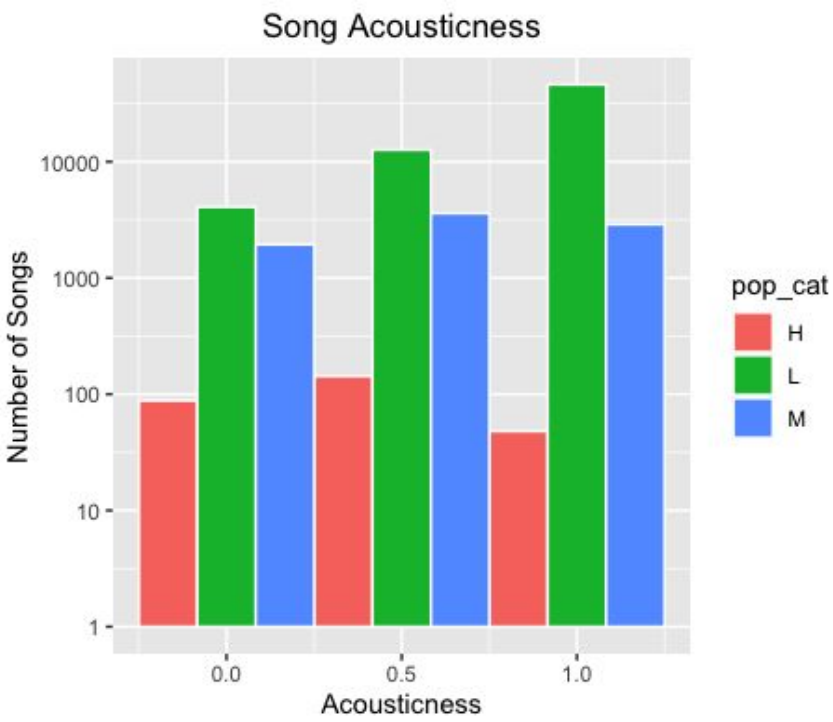
By: Tomi Boyejo, Evan Danu, Val Huerta, Hannah Lee, Holden Raffin, Aryaa Sapkota

# Assumptions/Methodology

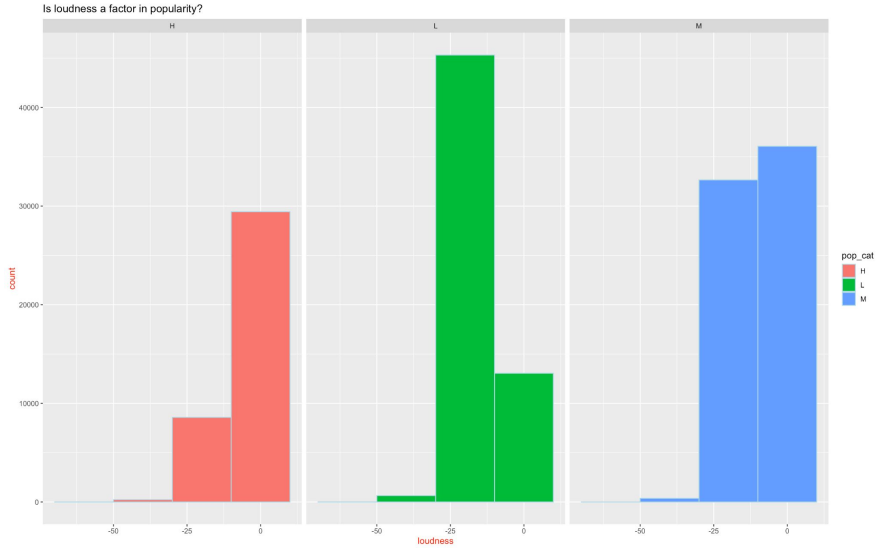
- The definition of popularity is ambiguous — Is it total streams? Is it percentage increase in streams?
  - We based our assumptions of popularity on percentage increase
    - Ex. Ed Sheeran's "Shape of You" ranked roughly 125th in popularity on Spotify dataset despite being 2nd all time in Spotify streams (3.34 billion all-time streams, according to Business Insider)
- Popularity is ultimately what determines consumers preferences and is a crucial variable. We chose to leave out Year in most of our regressions even though it produced the highest R^2 and correlation coefficients
  - Year is an uncontrollable variable
    - Ex. in 1950 there was nobody listening to Beethoven's fifth on spotify compared to the millions of users in 2022 listening to the Weeknd
- Assumed that the recent years are the most popular, so we mainly looked at the last 50 years
- Year although a seemingly crucial variable is intangible and we considered as a outlier in our data

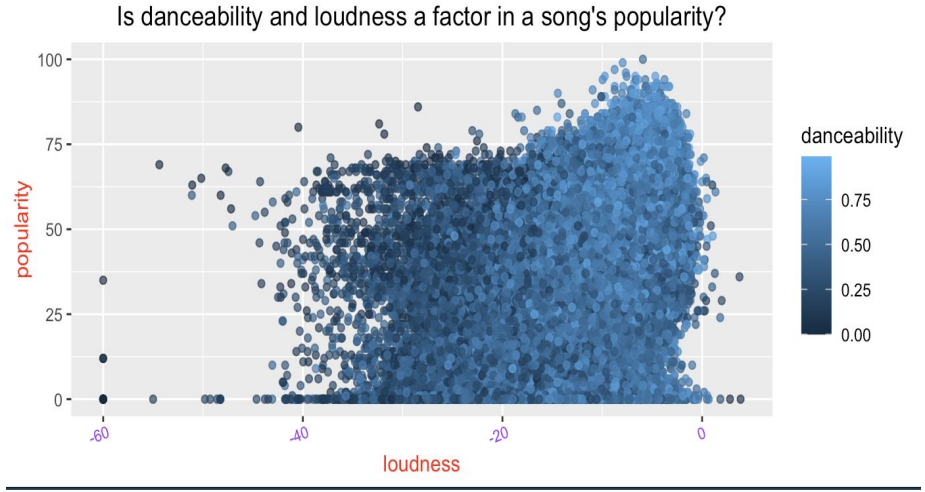
# Key Factors Related to Popularity: Energy and Acousticness



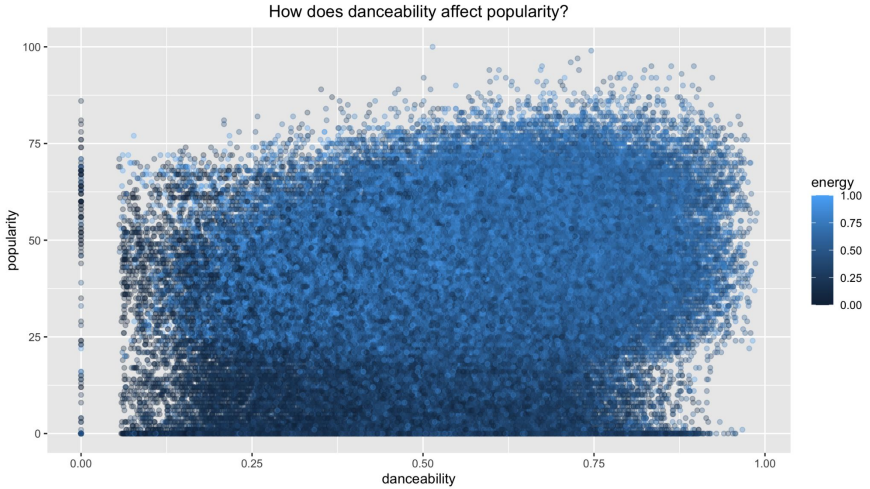



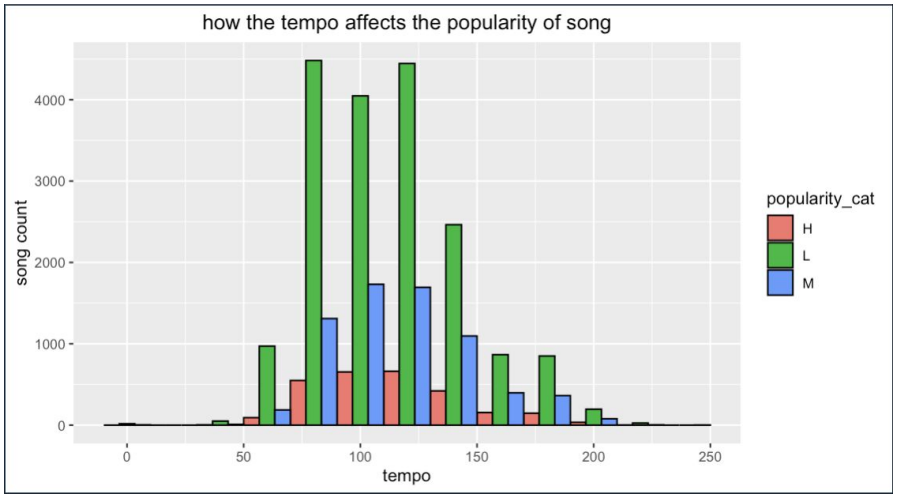
# Key Factors Related to Popularity: Loudness and Danceability



# Key Factors Related to Popularity: Danceability and Tempo



# REGRESSION MODELS

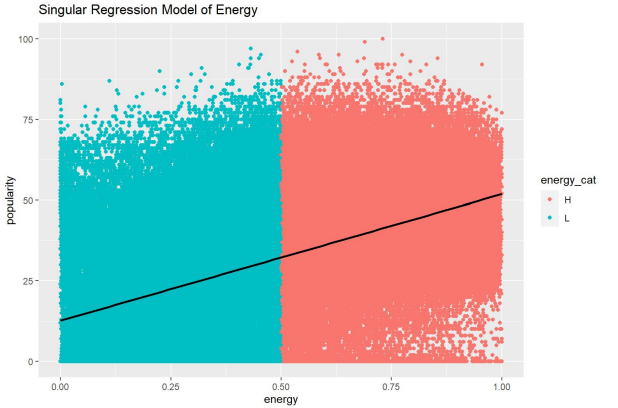
# Correlation Matrix w/ Popularity

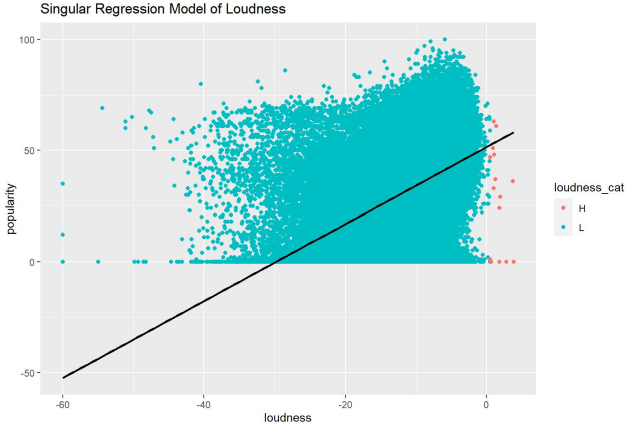
- Notable Strong Correlations:
  - Acousticness
  - Danceability
  - Energy
  - Explicit
  - Loudness
  - Year
- Weak Correlations:
  - Key
  - Duration
  - Speechiness
  - Mode
  - Valence
  - Liveness

| | popularity |
|---|---|
| acousticness | -0.614 |
| danceability | 0.241 |
| duration_ms | 0.047 |
| energy | 0.488 |
| explicit | 0.295 |
| instrumentalness | -0.313 |
| key | 0.009 |
| liveness | -0.067 |
| loudness | 0.455 |
| mode | -0.036 |
| popularity | 1.000 |
| speechiness | -0.051 |
| tempo | 0.134 |
| valence | 0.012 |
| year | 0.878 |

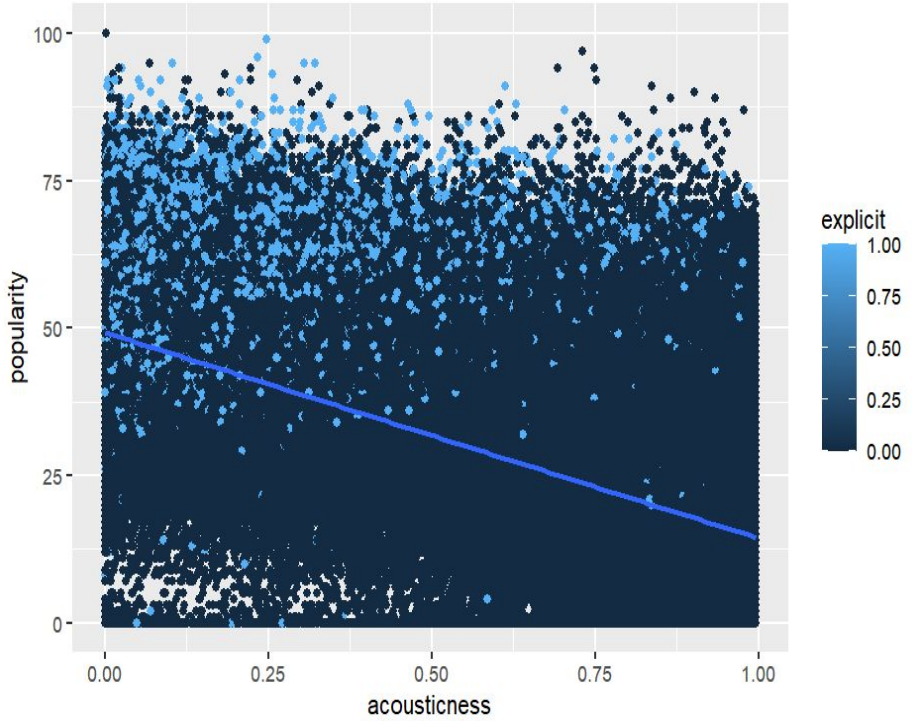
# Simple Regressions

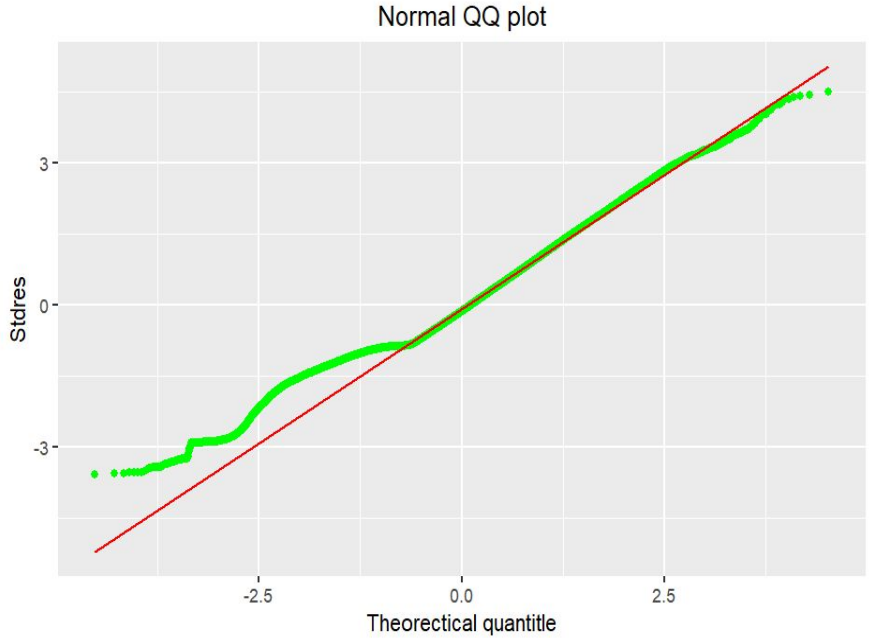
- Popularity~X:
  - Year: .771
  - Danceability: .0579
  - Instrumentalness: .09812
  - **Acousticness: .3776**
  - **Energy: .2383**
  - Explicit: .08704
  - Liveness: .004539
  - Valence: .0001473
  - **Loudness: .2067**
  - Speechiness: .002567
  - Tempo: .01791
  - Danceability: .0578

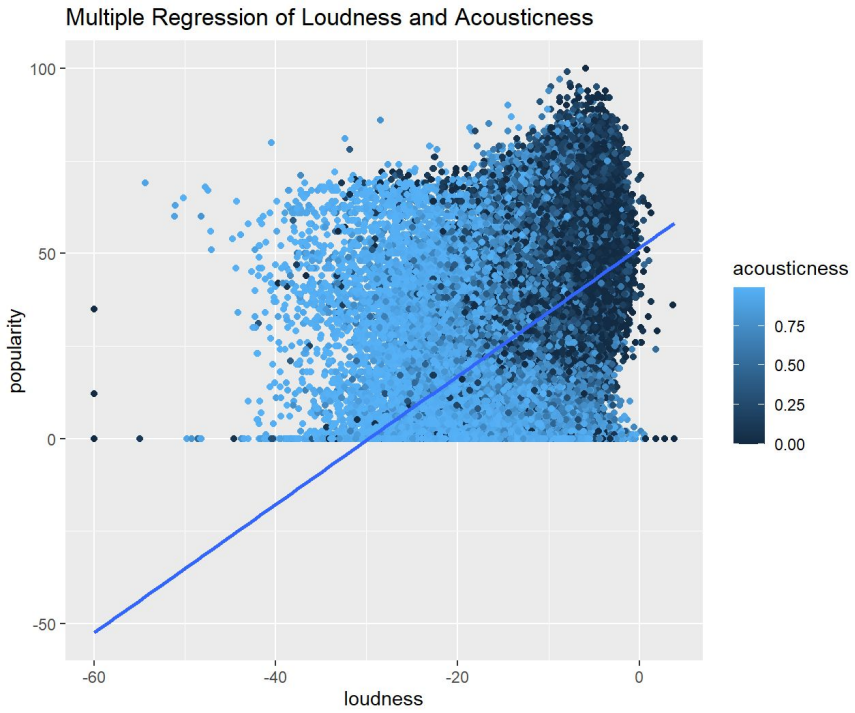



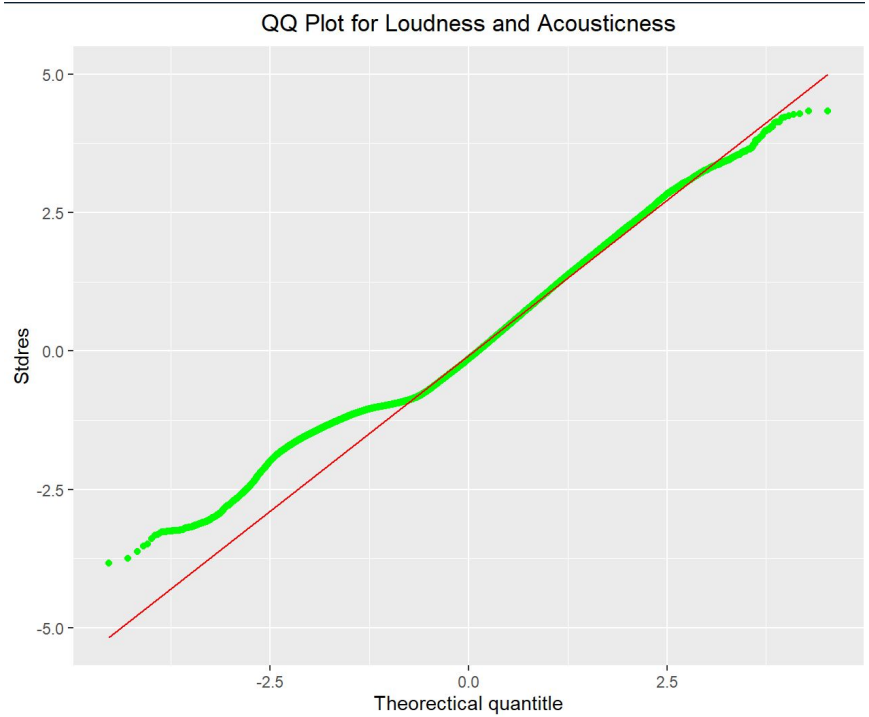
# Multiple Regression and QQ Plot: Popularity~Acousticness*Explicit R^2: .41



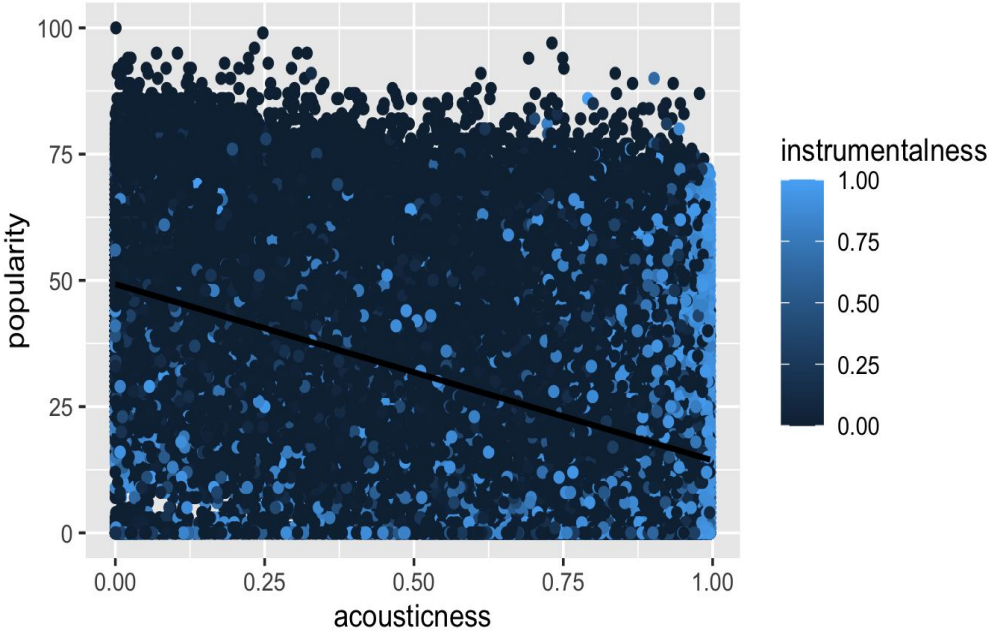



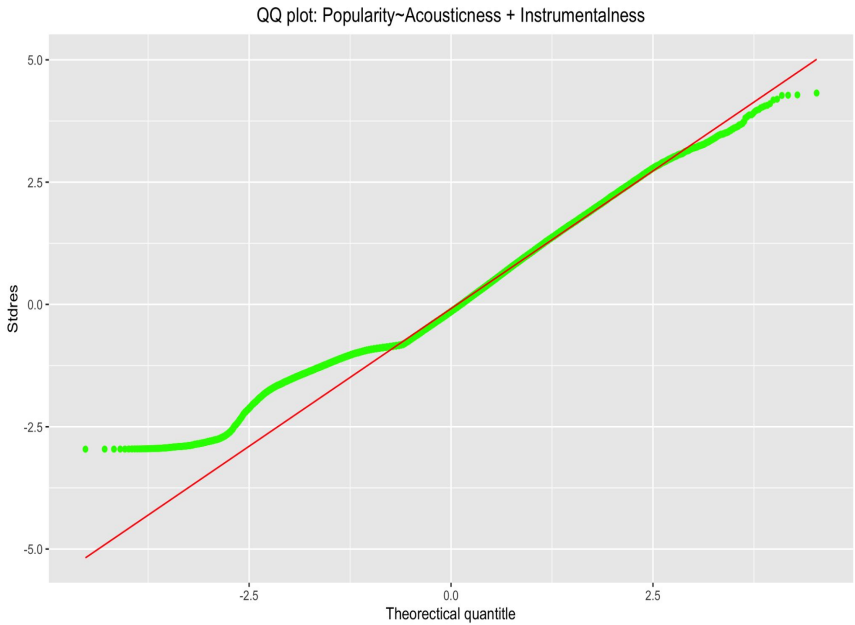
# Multiple Regression and QQ Plot: Popularity~Loudness*Acousticness R^2: .3987



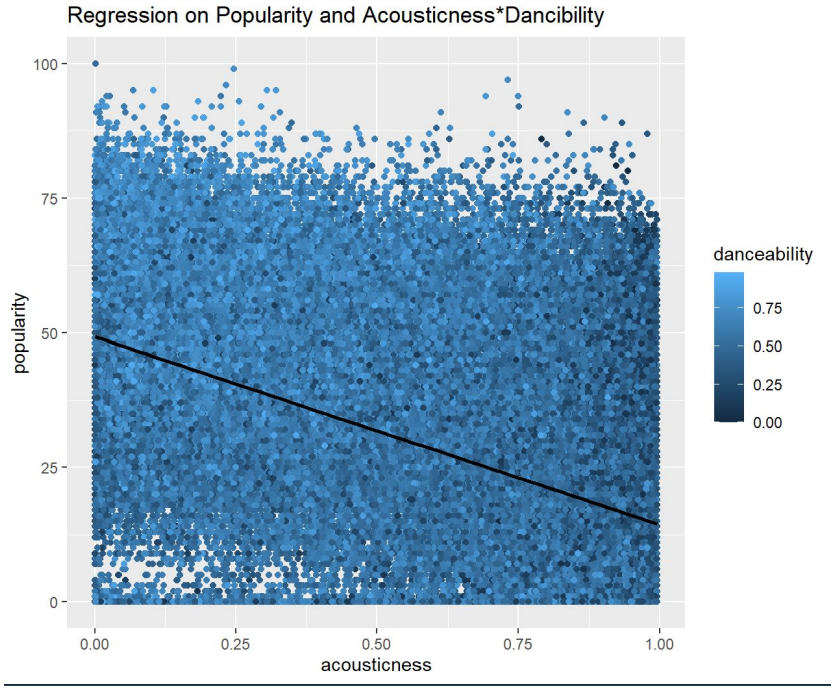



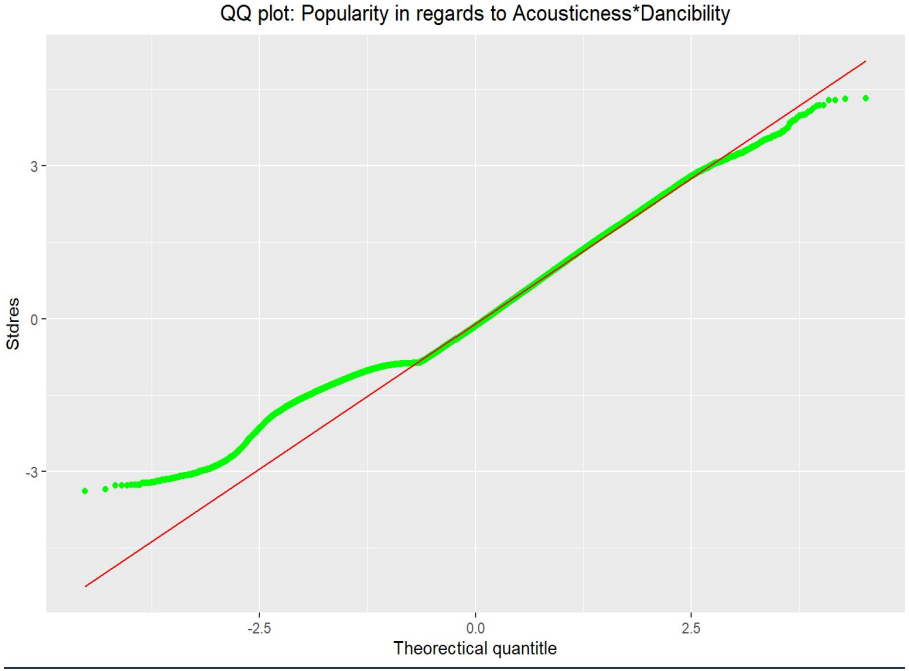
# Multiple Regression and QQ Plot: Popularity~Acousticness+Instrumentalness R^2: .3906



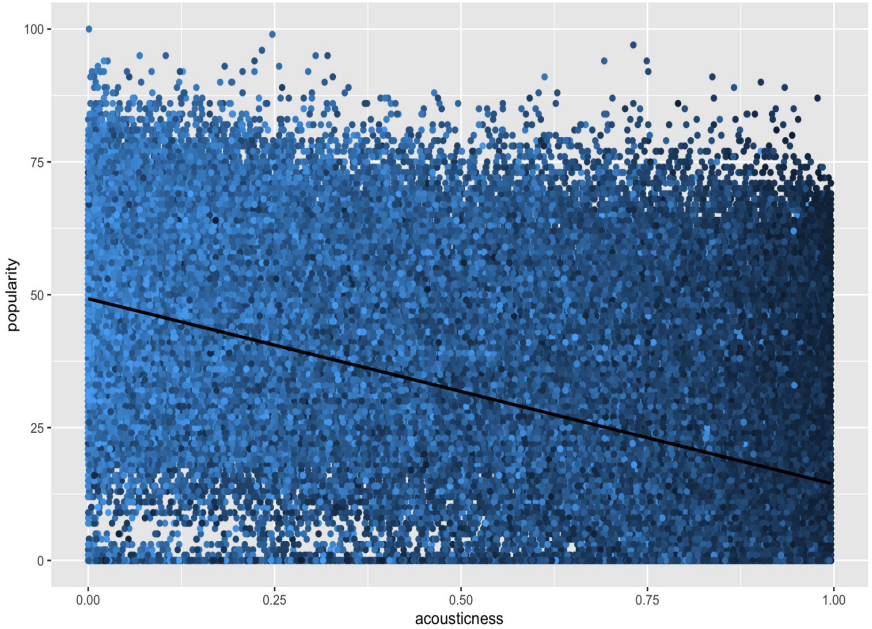



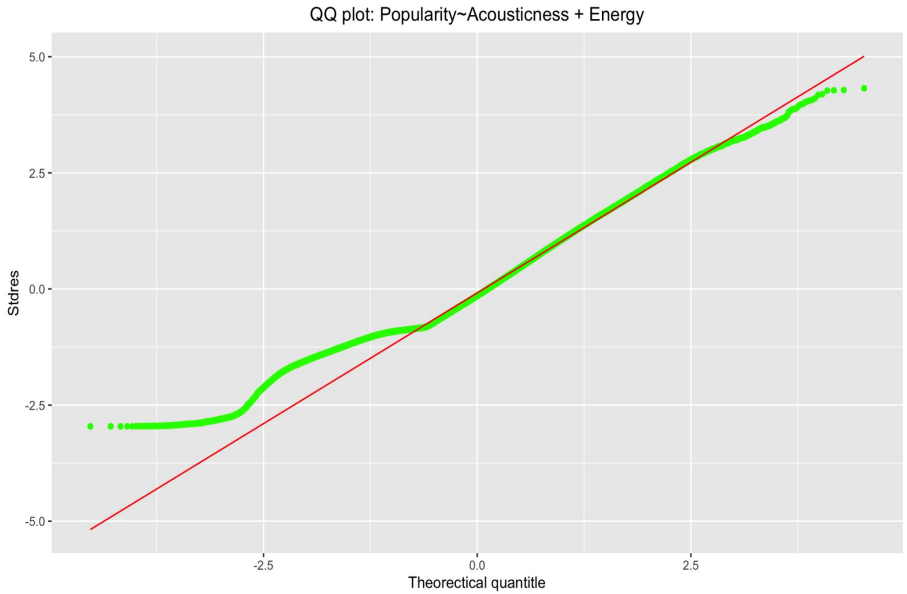
# Multiple Regression and QQ Plot: Popularity~Acousticness*Danceability R^2: .3883



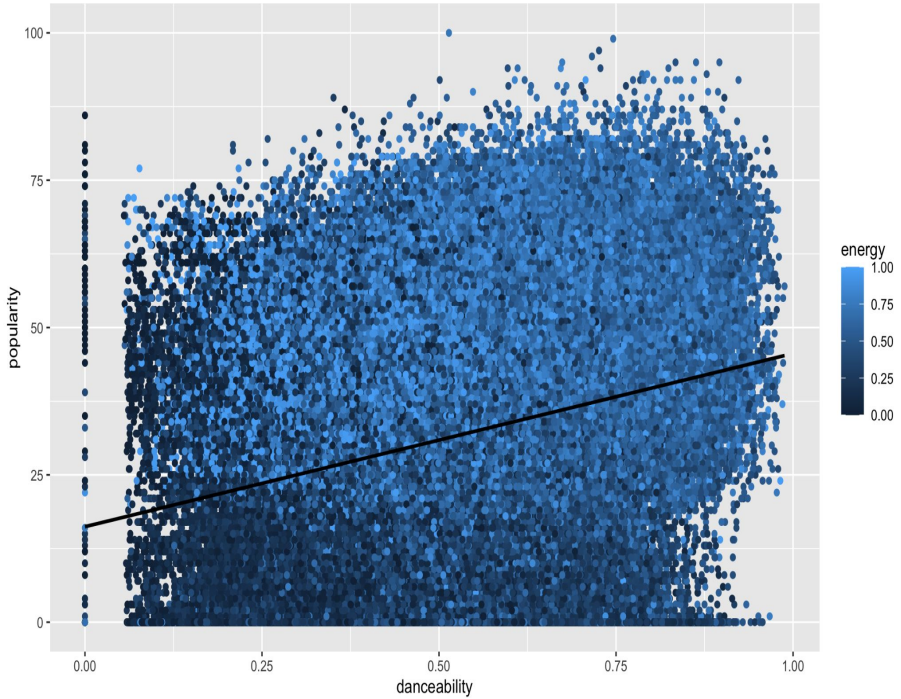



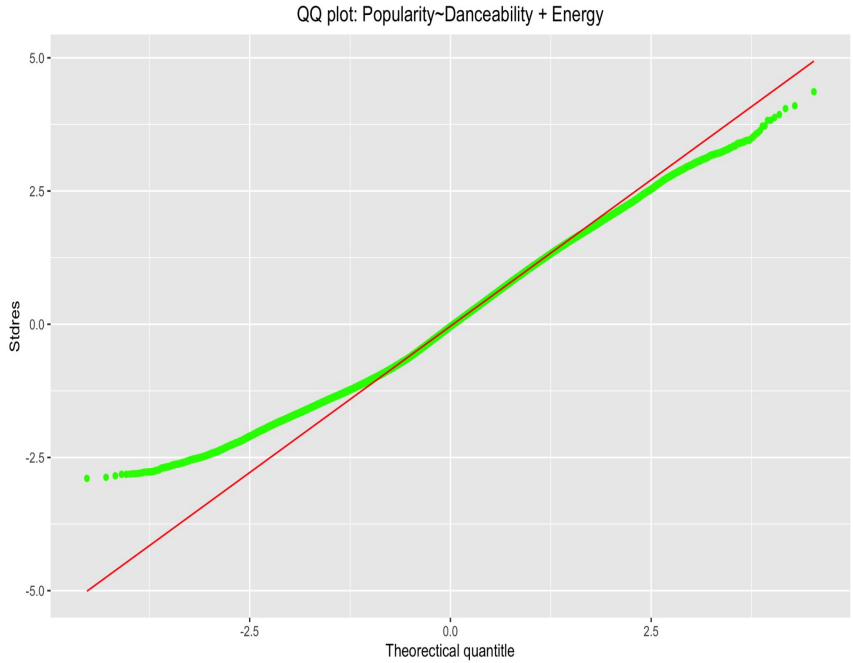
# Multiple Regression and QQ Plot: Popularity~Acousticness+Energy R^2: .3783



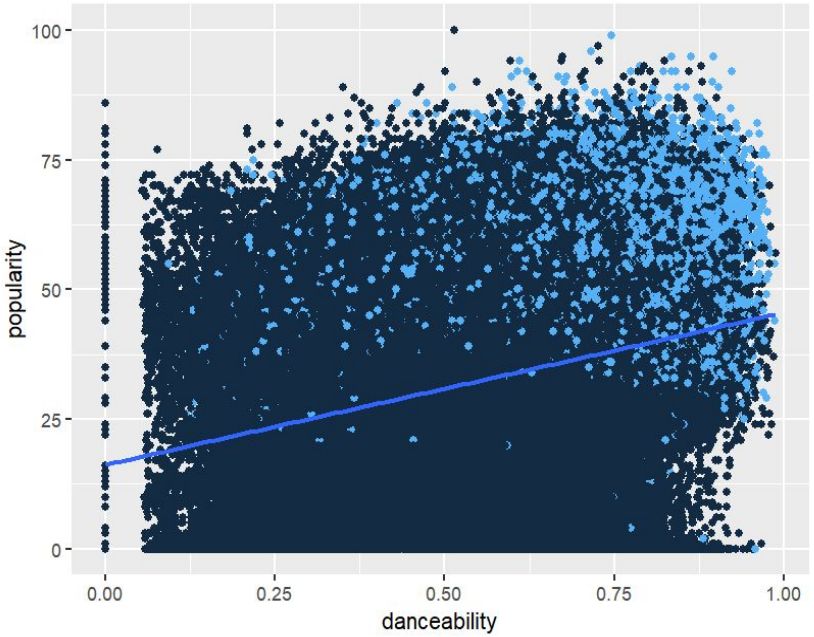



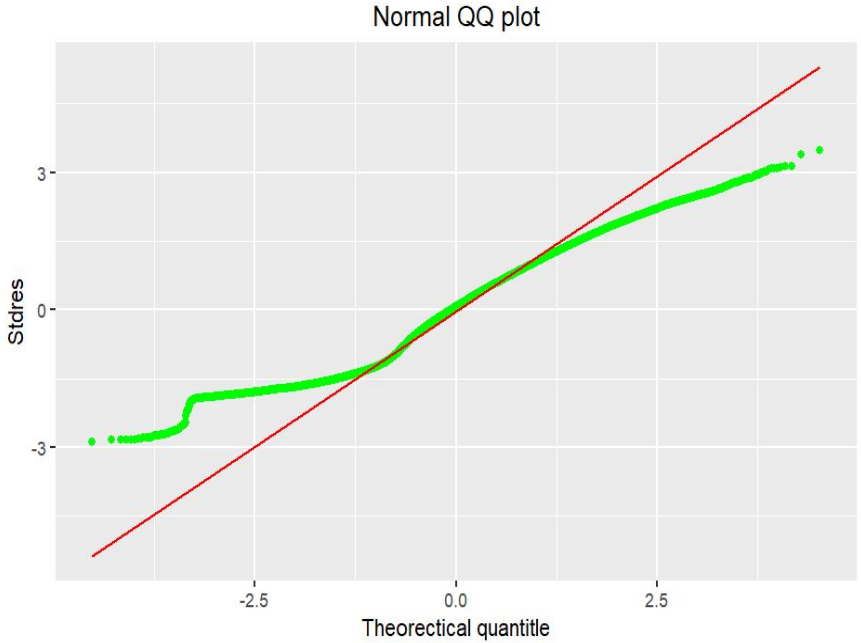
# Multiple Regression and QQ Plot: Popularity~Danceability+Energy  R^2: .2551



# Multiple Regression and QQ Plot: Popularity~Explicit*Danceability R^2: 0.1195



# Lessons About Current Music Tastes

1. In addition to the graphs, we used proportion tables to get some quick insight into current music tastes (2018-2020) versus older music.
    a. Findings:
        i. New Songs have lower Acousticness
        ii. Newer Songs have Higher Energy
        iii. New Songs are more Explicit
        iv. More lyricism, less instrumentality
2. Factors such as Valence and key have little effect on popularity

# Proportion Tables

(*New = 2018-2020, Old=Else)

| Year/ Acoustic | New | Old |
|---|---|---|
| H | .199 | .509 |
| L | .801 | .491 |

| Year/ Energy | New | Old |
|---|---|---|
| High | .727 | .476 |
| Low | .273 | .523 |

| Year/ Danceability | New | Old |
|---|---|---|
| Dance | .839 | .589 |
| No Dance | .161 | .411 |

| Year/ Explicit | 2018-2020 | 2010-2017 | Before 2010 |
|---|---|---|---|
| Explicit | .460 | .246 | .040 |
| Not Explicit | .540 | .754 | .960 |

# Conclusions/Recommendations

- Key Takeaways: From our analysis, creating songs with high danceability, loudness, and tempo will lead to higher chances of creating a popular song on Spotify.
- From our regression analysis, coupling low acousticness with variables such as explicitness, instrumentalness, and loudness generates the greatest effects on popularity
- To maximize success, artists should take a rounded approach rather than focusing on one variable
- In future case studies, we can examine unaddressed factors such as artist popularity, genre, and lyrical matter to strength our analysis of song popularity and preferences

# Works Cited:

Lane, B. (2023, February 4). *These are the 25 most streamed Spotify Songs of all time*. Insider. Retrieved March 1, 2023, from https://www.insider.com/most-streamed-spotify-songs-of-all-time#2-shape-of-you-by-ed-sheeran-24