Name: Adetutu Bodede
Date: June 26, 2022

Introduction

The main goal of my study is to explore and assess how musical elements influence the valence of a song to create a particular mood in music. My population of interest will be songs (digital music compositions) that can be streamed on Spotify. Possible stakeholders for my study are people who make music and want to create a certain mood in their songs (i.e. musicians, sound engineers, and producers). Companies and distributors in the music industry may also be interested in this study to utilize valence and song moods in their marketing tactics.

Research Questions

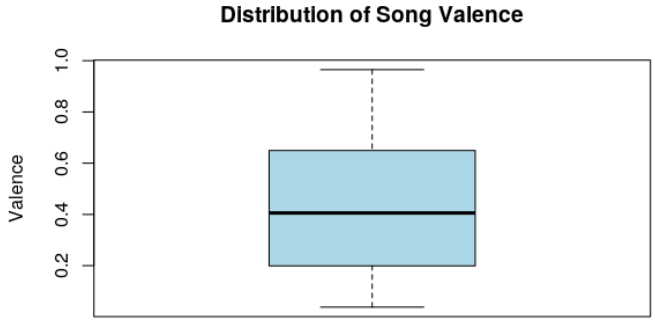
RQ1 – Does a happy mood correlate with a higher valence?
RQ2 – Does a slower tempo correlate with a lower valence?

Data Collection Summary

I collected my data from an open source raw dataset. The songs were chosen from multiple playlists classified as 'happy' or 'sad' playlists already available on Spotify. I used the song title, valence, and tempo variables to form my own dataset. My sampling unit will consist of songs selected from already created playlists on Spotify named after the corresponding moods I hope to observe. My sample will not be completely representative of my population of interest as the songs were not randomly selected. The songs will have English names and be selected from ready available Spotify playlists, so artists whose music is not on Spotify and does not have an English title will not be included in the data pool. My final sample size will be n=100. I removed any songs with non-English titles and any songs that had a mood different from happy or sad. Other moods, such as energetic or calm, were not included. Since moods and emotions can be fairly subjective, I wanted to stray from choosing moods that could be difficult to analyze.

Descriptive Analysis of Response Variable *(incl. graph(s) of distribution)*

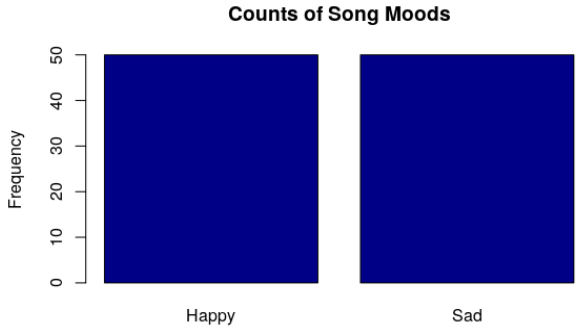
A histogram of the response variable, valence, was created to assess whether the distribution had normality. To be sure, I recorded the 5-number summary, mean, and standard deviation of the distribution to measure the center and spread of song valence and compare the mean and median. The mean song valence is 0.437801 and its standard deviation is 0.2738077. The median is 0.4055, meaning the distribution is shaped slightly right-skewed. I did not have any outliers in the valence distribution. I attempted to use the natural log, square root, inverse, and exponential transformations but they did not improve the distribution.



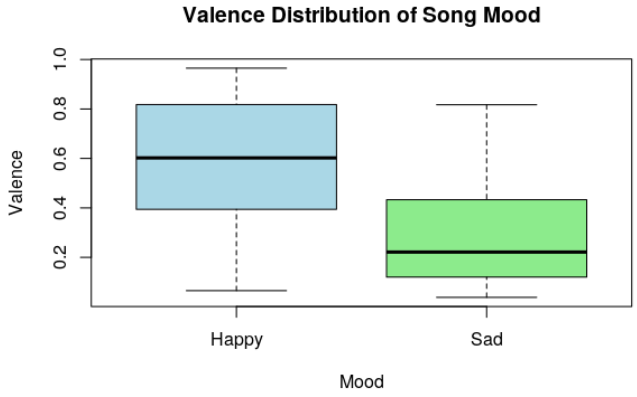
Name: Adetutu Bodede
Date: June 26, 2022

Investigation of Explanatory Variable 1 *(incl. univariate and bivariate graphs)*

My first explanatory variable is a categorical variable of mood. To help describe its distribution I utilized a bar chart. The bar chart has frequency as the y-axis to represent how often the two possible categories of mood (happy or sad within the x-axis) appear in the data I have pooled. There is an even distribution of moods within the sample population of songs that consists of 50 happy and 50 sad mood songs.



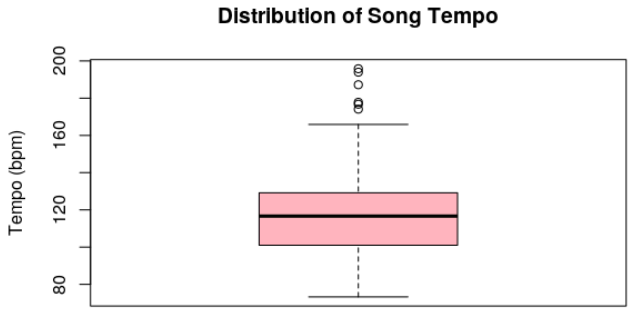
In order to test the relationship between valence (a numeric variable) and mood (a categorical variable), I created a grouped boxplot. Going off the image of the graph only, there seems to be a relationship between the explanatory (mood) and response (valence) variable. While the whiskers of the plot suggest there is an overlap in happy and sad song valence, the box itself shows the median, Q3, and maximum of the happy-mood boxplot resting at a higher valence than sad songs. This corresponds with my expectation that happy mood songs correlate with a higher valence and sad mood songs correlate with a lower valence. The reason I expected this is because previous studies done on valence state that higher valence is more representative of a positive sound and a happy mood is generally considered to be a positive thing.



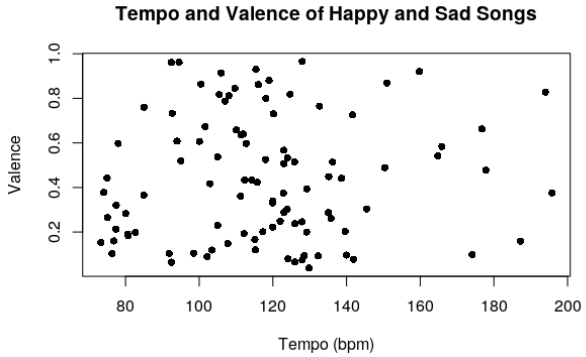
Name: Adetutu Bodede
Date: June 26, 2022

<u>Investigation of Explanatory Variable 2</u> *(incl. univariate and bivariate graphs)*

My second explanatory variable is a numeric variable of tempo, which is measured by beats-per-minute. In order to describe its distribution I used a boxplot. The plot also aided in assessing the normality of the distribution of song tempo. Looking at the graph, there are quite a few points that fall outside the whiskers of the boxplot. To be safe, I considered the distribution of song tempo to be skewed and checked for the median and IQR of tempo using the 5-number summary. The median tempo is 116.6675 bpm with an IQR of 28.1725.



To test the relationship between valence (response variable) and tempo (explanatory variable) which are both numeric variables, I created a scatter plot. Using just the image to gauge the relationship, there does not seem to be a clear relationship between the two variables. With how spread out the points are, the correlation seems to be weak, but there is no clear trend of a positive or negative slope of the points so there likely is little to no correlation or relationship between valence and tempo. This relationship, or lack-there-of, was not initially what I expected. It is possible that my sample did not have a wide enough range of tempos. It is also possible that the lack of a relationship is because I did not account for rhythmic devices that are used in music, such as double-time and half-time, that alter the tempo of the melody of a song while keeping the tempo of the instrumental either slower or faster.



Name: Adetutu Bodede
Date: June 26, 2022

R Code *(organized by variable without output or extraneous syntax)*

Response Variable:

```
> boxplot(music$valence, ylab='Valence', main='Distribution of Song Valence', col='light blue')
> mean(music$valence, na.rm = TRUE)
> sd(music$valence, na.rm = TRUE)
> fivenum(music$valence)
```

Explanatory Variable 1:

```
> barplot(table(music$mood), ylab='Frequency', main='Counts of Song Moods', col='dark blue')
> boxplot(music$valence~music$mood, xlab='Mood', ylab='Valence', main='Valence
Distribution of Song Mood', col=c('light blue','light green'))
```

Explanatory Variable 2:

```
> boxplot(music$tempo, ylab='Tempo (bpm)', main='Distribution of Song Tempo', col='light
pink')
> fivenum(music$tempo)
> plot(music$tempo, music$valence, xlab='Tempo (bpm)', ylab='Valence', main='Tempo and
Valence of Happy and Sad Songs', pch=16)
```