PHS 650: Final Project

Aly, Haley, Zoe

2023-04-26

Table of contents

Introduction

In the music industry, there is a consideration of what song will be popular or a 'hit,' as song popularity is associated with more revenue (Pham, Kyauk, and Park 2015). Thus, predicting the popularity of a song, referred to as the study of Hit Song Science, can be useful in determining which songs should receive the most investment from musicians and their labels. Random forests have been found to accurately predict which songs will be popular and determined that songs that 'made it' to the top charts were found to be 'happier' and more 'party-like' (Interiano et al. 2018; Middlebrook and Sheik 2019). Additionally, artist familiarity, loudness, year of release, and number of genres were also found to accurately predict the popularity of songs (Pham, Kyauk, and Park 2015).

The goal of our project is to add to study of the Hit Song Science and examine how song elements, specifically the duration and intensity, are associated with the song's popularity. We hypothesize that shorter songs are more likely to be popular and more intense songs are more likely to be popular.

Methods

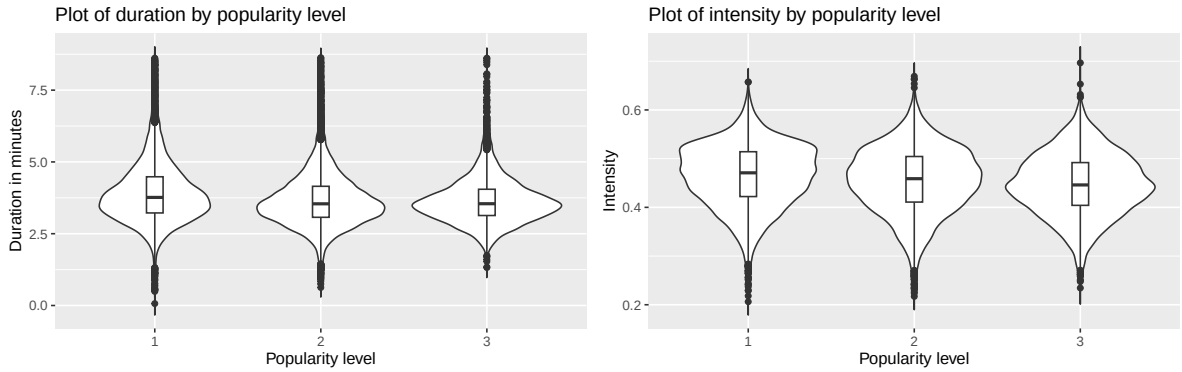
This is a book created from markdown and executable code.

See (**knuth84?**) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

Results



(a) Duration

(b) Intensity

Figure 1: Duration and Intensity by Popularity Level

From Figure 1 we see that the distributions of intensity and duration are visibly skewed within popularity classes. As a result, we used a Spearman rank correlation to test each hypothesis. In Table 1 we see that both p-values are well below our pre-established cutoff of 0.05. Since we are doing multiple statistical tests, we could adjust the p-values to account for this. However, since there are only two tests and the p-values are several orders of magnitude smaller than our cutoff, this is unnecessary. Both duration and intensity were negatively correlated with popularity to a weak degree; neither coefficient was on the expected scale. Since intensity was a composite measure, exploration of how the components relate to popularity could explain why the direction of correlation was unexpected.

Table 1: Spearman Rank Correlation Results

Factor	p-value	Coefficient
Duration	$1.4736263 \times 10^{-81}$	-0.1132319
Intensity	$4.6505396 \times 10^{-66}$	-0.1016895

References

Interiano, Myra, Kamyar Kazemi, Lijia Wang, Jienian Yang, Zhaoxia Yu, and Natalia L. Komarova. 2018. "Musical Trends and Predictability of Success in Contemporary Songs in and Out of the Top Charts." *Royal Society Open Science* 5 (5): 171274. https://doi.org/10.1098/rsos.171274.

Middlebrook, Kai, and Kian Sheik. 2019. "Song Hit Prediction: Predicting Billboard Hits Using Spotify Data." arXiv. https://doi.org/10.48550/arXiv.1908.08609.

Pham, James, Edric Kyauk, and Edwin Park. 2015. "Predicting Song Popularity."