

★ *Statistics of* Music

Super Staters: Logan Evans, Lewis
Zhu, Diane Lin, Katherine Jonas
& Natalia Hauser

Introduction

#1

Our Team

Get to know us!

#2

Source

- Million Song Dataset
- Echo Nest

#3

Previous Literature

Prediction of product success: explaining song popularity by audio features from Spotify Data

#4

Research Question & Hypotheses



Research Question

Given a genre (Rock, Pop, Rap, or Country), associated with an artist (artist.terms), how are the different song attributes (such as song.tempo, song.duration, artist.familiarity, song.year) related to song popularity (song.hottness)? Do these relationships differ for artists associated with different genres?

Hypotheses

We predicted:

- Rock and pop would have a positive relationship between tempo and popularity (and rap and country a lesser positive one)
- Year and popularity would have a positive relationship
- We do not believe the relationship between different song attributes and song popularity will differ for artists associated with different genres.

Methodology: Visualizations



To begin to answer our research question, we decided to plot linear regressions using a scatter plot.

We chose to use this method of visualization to help us determine if there is a relationship between song popularity as our outcome variable and the four song attribute variables (song tempo, song duration, artist familiarity, song year) we are studying.



Each scatterplot was then faceted by genre to show potential variations by our four genres (Pop, Rock, Rap, Country)

For song duration, we decided to filter out songs longer than 10 minutes to reduce outliers in our visualizations



Methodology: Interactive Model, AIC & Adjusted R-squared

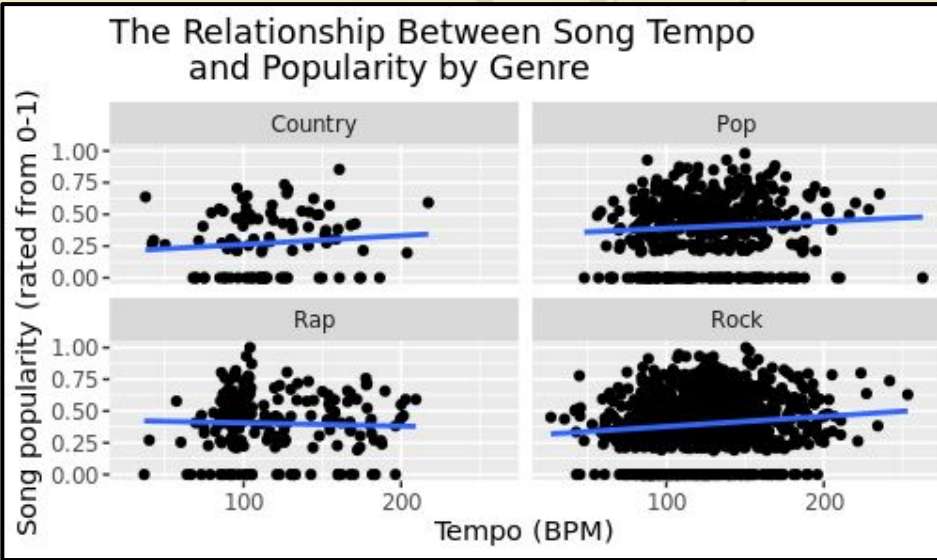


- We regress linear models with interaction between genre and one of our four song attributes
 - Four interactive models total
 - Why interactive and not additive?
 - “Do these relationships differ for artists associated with different genres?”
- Compared AIC and adjusted R^2 for each model
 - The “how” in “how are the different song attributes related to song popularity?”



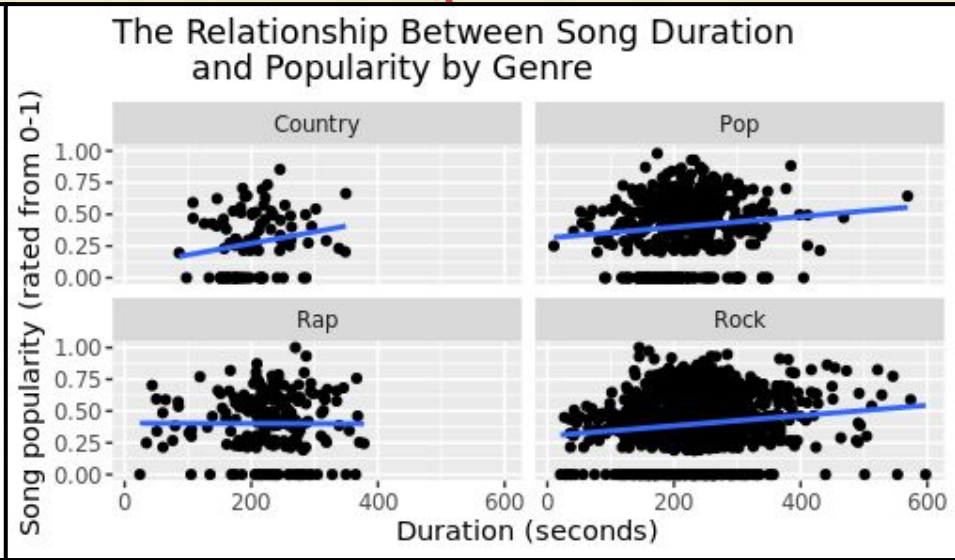
RESULTS

Song Tempo



All of the genres have a positive relationship with song tempo except for rap which has a negative relationship.

Song Duration



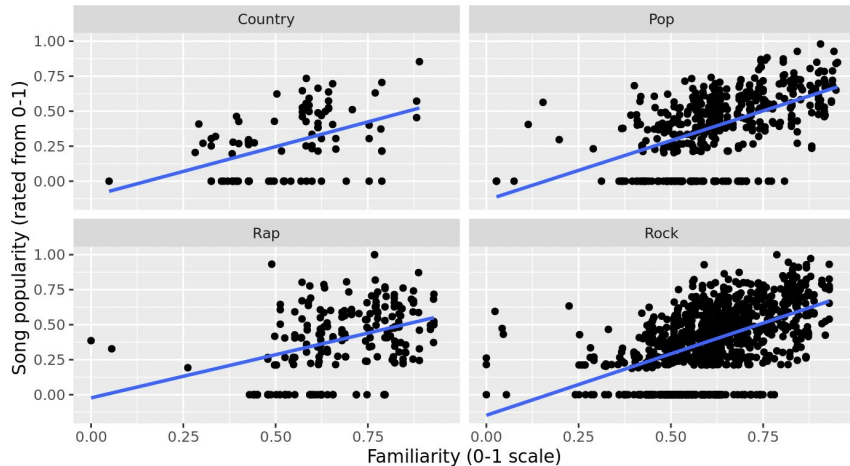
For rap, there was no relation on average between duration and popularity as the linear fit line was flat. For the other three, there was a positive relationship between duration and popularity.

RESULTS

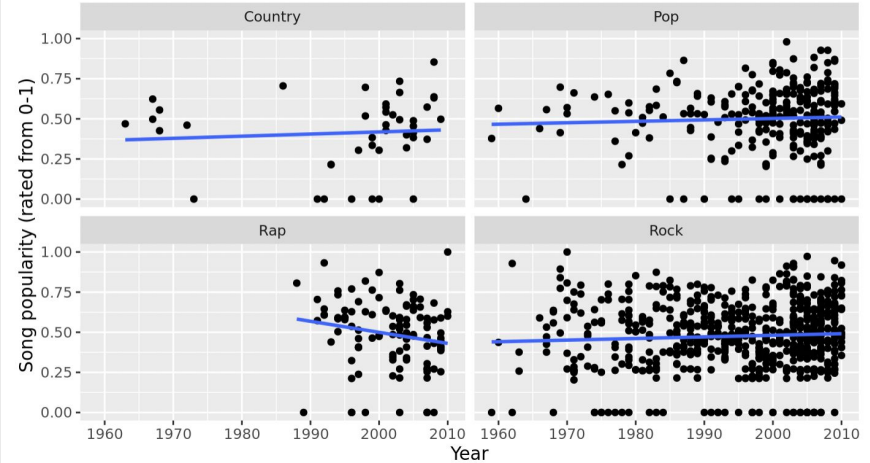
Artist Familiarity

Based on our linear models, there is a positive relationship between artist familiarity and song popularity for all four genres.

The Relationship Between Artist Familiarity and Popularity by Genre



The Relationship Between Song Year and Popularity by Genre



Song Year

For song year and popularity, visually there appears to be a weak positive relationship for all genres except rap which has a negative correlation.

Interactive Model - Song Tempo

★ Song Tempo

All of the genres have a positive relationship with song tempo except for rap which has a negative relationship.

Country:

$$\widehat{\text{song.hotttness}} = 0.194 + 0.000691 * \text{song.tempo}$$

Pop:

$$\widehat{\text{song.hotttness}} = 0.194 + 0.139 + (0.000691 - 0.000135) * \text{song.tempo} = 0.333 + 0.000556 * \text{song.tempo}$$

Rap:

$$\widehat{\text{song.hotttness}} = 0.194 + 0.237 + (0.000691 - 0.000937) * \text{song.tempo} = 0.431 - 0.001664 * \text{song.tempo}$$

Rock:

$$\widehat{\text{song.hotttness}} = 0.194 + 0.104 + (0.000691 + 0.000109) * \text{song.tempo} = 0.298 + 0.0008 * \text{song.tempo}$$

Interactive Model - Song Duration

Song Duration

Country and song duration have a positive relation between song duration and song popularity, while pop and rap have negative relationships (that are close to 0)

Country:

$$\widehat{\text{song.hotttness}} = 0.0892 + 0.000910 * \text{song.duration}$$

Pop:

$$\widehat{\text{song.hotttness}} = 0.0892 + 0.326 + (0.000910 - 0.000959) * \text{song.duration} = 0.4152 - 0.000049 * \text{song.duration}$$

Rap:

$$\widehat{\text{song.hotttness}} = 0.0892 + 0.315 + (0.000910 - 0.000922) * \text{song.duration} = 0.131 - 0.000012 * \text{song.duration}$$

Rock:

$$\widehat{\text{song.hotttness}} = 0.0892 + 0.262 + (0.000910 - 0.000723) * \text{song.duration} = -0.191 + 0.000187 * \text{song.duration}$$

Interactive Model - Artist Familiarity

★ Artist Familiarity

All of the genres have a positive relationship.

Country:

$$\widehat{\text{song.hotttness}} = -0.107 + 0.707 * \text{artist.familiarity}$$

Pop:

$$\widehat{\text{song.hotttness}} = -0.107 - 0.0285 + (0.707 + 0.143) * \text{artist.familiarity} = -0.1355 + 0.85 * \text{artist.familiarity}$$

Rap:

$$\widehat{\text{song.hotttness}} = -0.107 + 0.0842 + (0.707 - 0.0904) * \text{artist.familiarity} = -0.0228 + 0.6166 * \text{artist.familiarity}$$

Rock:

$$\widehat{\text{song.hotttness}} = -0.107 - 0.0401 + (0.707 + 0.171) * \text{artist.familiarity} = -0.1471 + 0.878 * \text{artist.familiarity}$$

Interactive Model

Song Year

There is a positive
relation for all
genres other than
rap

Country:

$$\widehat{\text{song.hottness}} = -2.23 + 0.00133 * \text{song.year}$$

Pop:

$$\widehat{\text{song.hottness}} = -2.23 + 0.951 + (0.00133 - 0.000434) * \text{song.year} = -1.279 + 0.000896 * \text{song.year}$$

Rap:

$$\widehat{\text{song.hottness}} = -2.23 + 16.5 + (0.00133 - 0.00822) * \text{song.year} = 14.27 - 0.00689 * \text{song.year}$$

Rock:

$$\widehat{\text{song.hottness}} = -2.23 + 0.712 + (0.00133 - 0.000325) * \text{song.year} = -1.518 + 0.001005 * \text{song.year}$$

RESULTS cont.

AIC and adjusted R-squared

Based on the AIC and adjusted R^2 values, artist familiarity and song popularity had the best fit model of our attributes and duration and song popularity had the weakest fit of our attributes

```
> glance(int_model_tempo)$adj.r.squared
[1] 0.01868008
> glance(int_model_tempo)$AIC
[1] -5.579001
>
> glance(int_model_duration)$adj.r.squared
[1] 0.01595886
> glance(int_model_duration)$AIC
[1] -0.6027787
>
> glance(int_model_familiarity)$adj.r.squared
[1] 0.2661948
> glance(int_model_familiarity)$AIC
[1] -527.8859
>
> glance(int_model_year)$adj.r.squared
[1] 0.006317068
> glance(int_model_year)$AIC
[1] -265.11
```



Citations & Acknowledgements

Resources Used:

- Million Song Dataset
- Nijkamp, Rutger.
“Prediction of product success: explaining song popularity by audio features from Spotify data.” (2018).

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

Many thanks to **Ed Tam** for supporting us through this project and answering our questions.

