Comparison of Music Taste Across Regions and Generations

Eduardo Armenta, Brendan Baker, Brian Wimmer

Agenda

- Introduce Topics
- US vs UK Top 50
- Evolution Across Decades
- US vs Mexico Top 50
- Predicting Popularity
- Conclusion

Topics

US vs UK

Comparing if there are differences between music preferences

Decade vs Decade

Looking into each of the past 7 decades and the evolution of music

US vs Mexico

Comparing happiness in music across languages and world regions

Predicting Popularity

Understanding if song popularity comes from tangible features or abstract concepts

Do the characteristics of popular songs vary between countries that speak the same language?

EDA

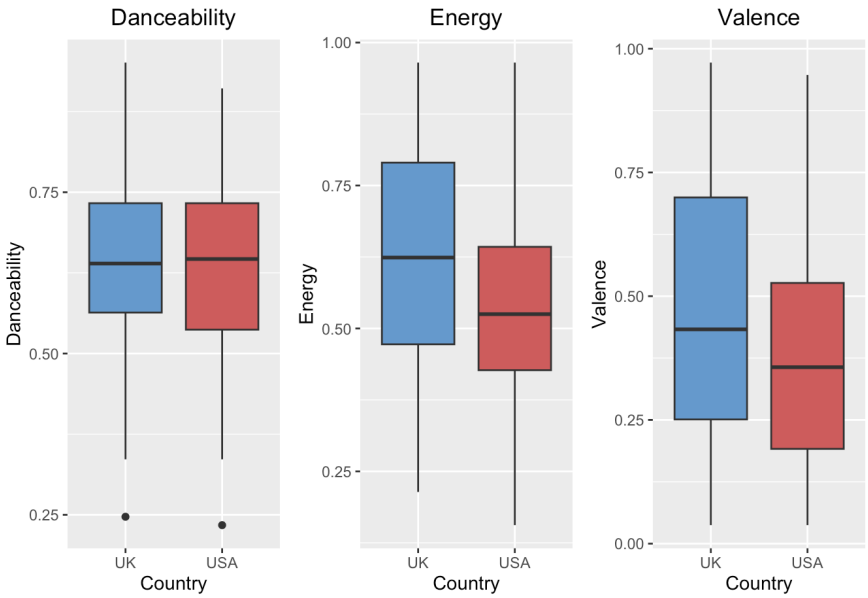


Figure 1.1: Boxplots of Variables of Interest

Observations

- Danceability: minimal difference between countries
- Energy: significant variation, UK ~20% higher
- Valence: similar variation as Energy
- UK Energy and Valence interquartile range is broader

Testing: two sample t-test

Experiment results:

	Variable	P-value	Results
Fail	Danceability	.31	Failed to reject the null hypothesis. UK and US means are too similar
Reject	Energy	.01	Reject the null hypothesis. UK mean is higher than the US mean
Reject	Valence	.03	Reject the null hypothesis. UK mean is higher than the US mean

Hypotheses

- Null: the mean of the tested variable for the US is the same as that for the UK
- Alternative: the mean of the tested variable is larger for the UK than that of the US

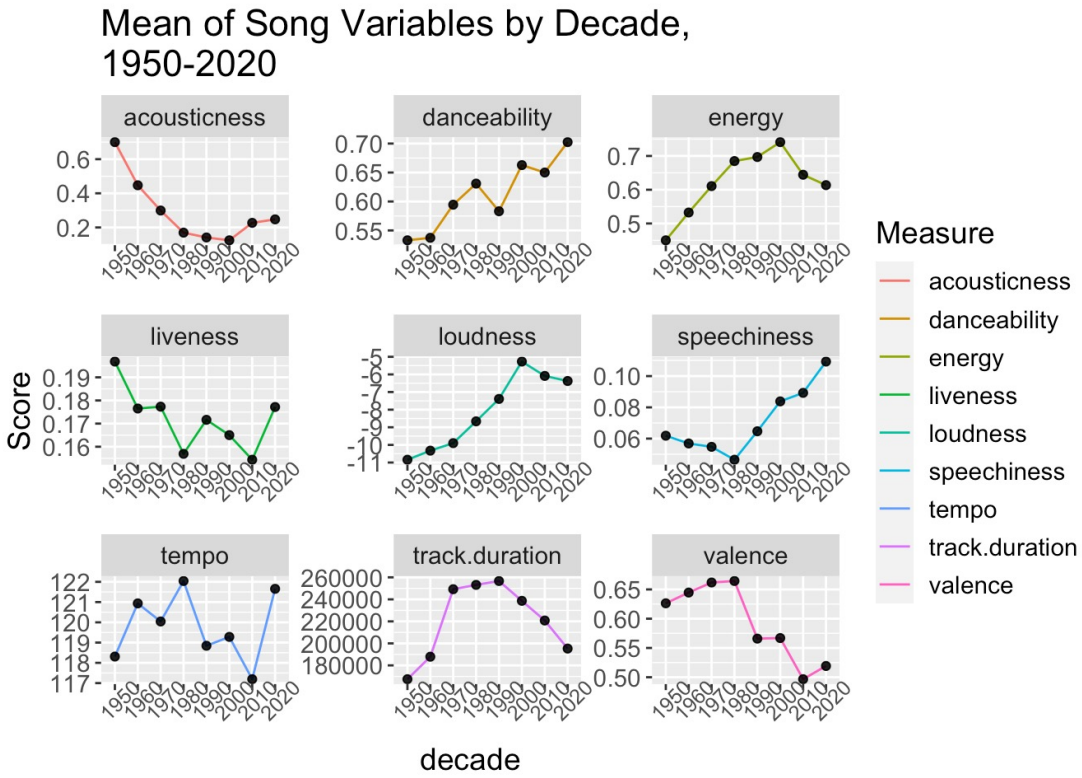
Conclusion

Although we expected each variable to have higher means for the UK than the US, only Energy and Valence had significantly greater averages

From this, we can conclude that there is a difference in music characteristics among countries that speak the same language

Has music become louder over time?

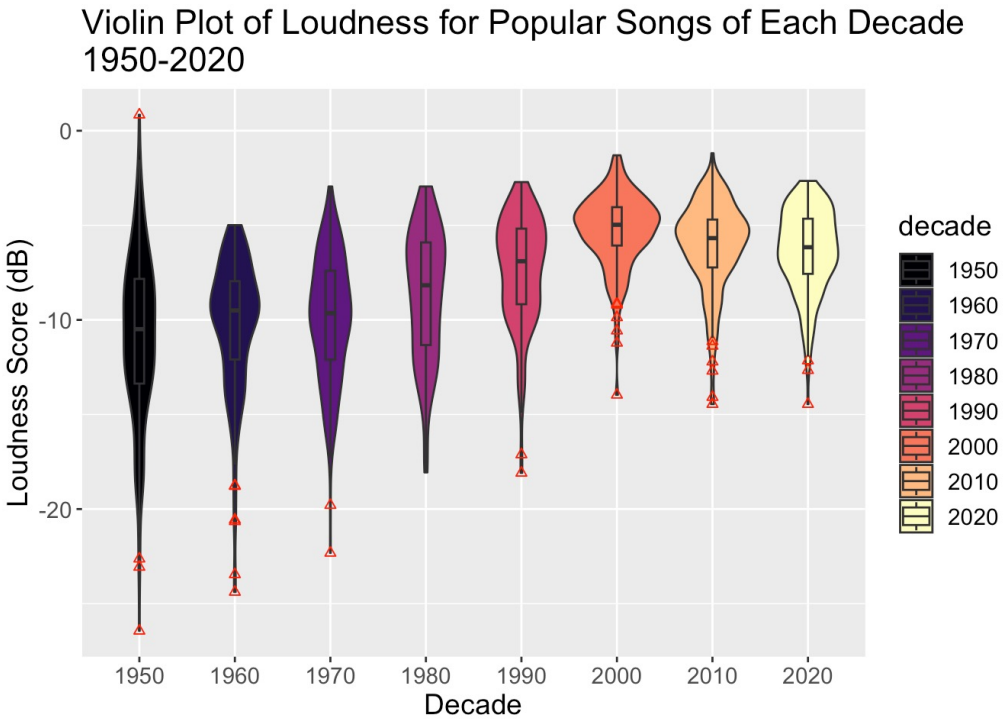
EDA



Key observations

- There are clear differences among the variables scores across all decades
- Loudness presented a clear trend going upward
- Selected to analyze loudness due to its objective nature (decibels)

EDA

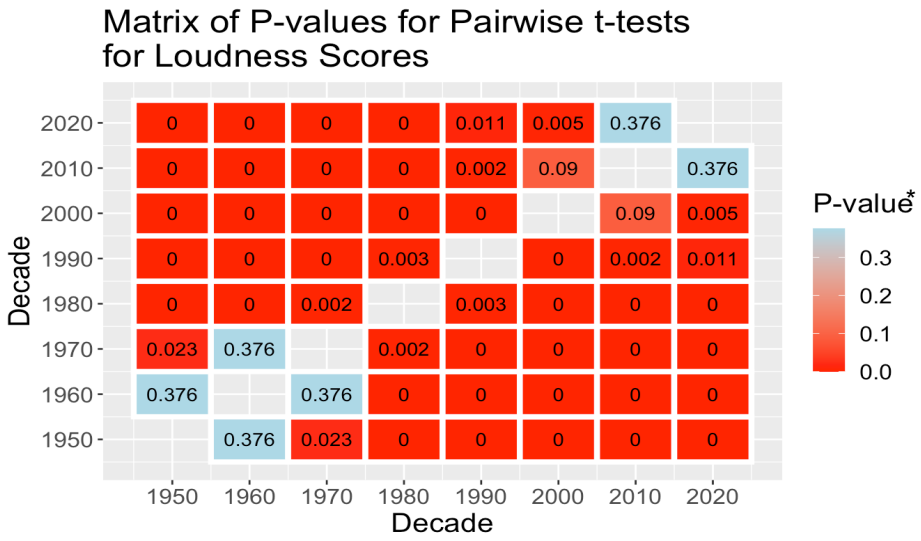


Key observations

- Loudness dispersion throughout songs gets tighter around the mean across the decades
- Distributions are mostly Gaussian, with some decades being somewhat skewed

Testing: ANOVA, Pairwise t-test

ANOVA			
F-statistic	76.8	P-value	< 0.001



Note: A value of 0 indicates P < .001

ANOVA hypotheses

- Null: No difference in the loudness score between decades
- Alt: mean loudness scores are not equal

Pairwise t-test hypotheses

- Null: no difference between loudness score between some decade X and Y
- Alt: the difference between loudness scores is not zero

*Hochberg adjustment

Conclusion

Initially, we expected the loudness score to increase as the years went by, but after exponential growth, it peaked in the 2000's and began trending downwards

We can conclude that music does change throughout the decades, but old trends get reused and the change is always influenced by music of the past

Does music in Mexico’s Top 50 playlist tend to be more danceable and happy than the US counterpart?

EDA

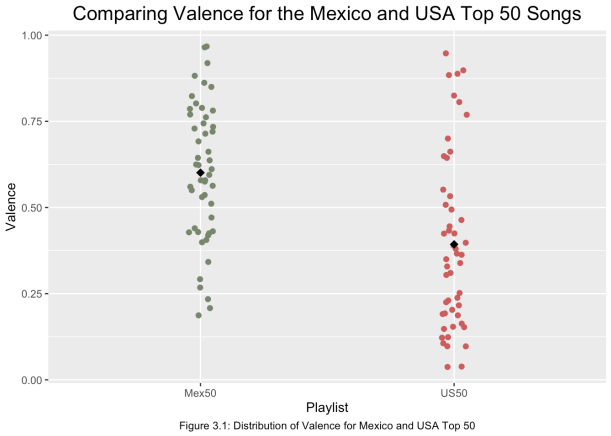


Figure 3.1: Distribution of Valence for Mexico and USA Top 50

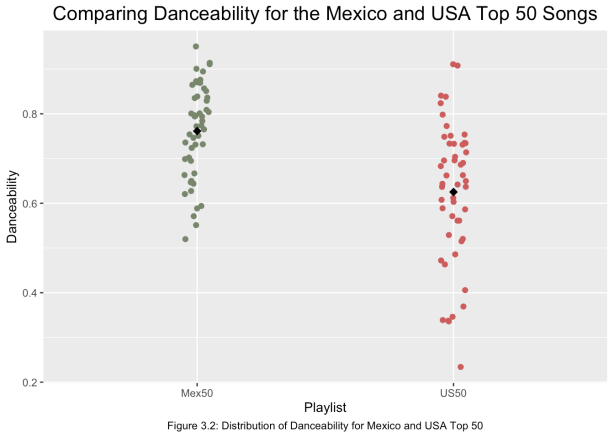


Figure 3.2: Distribution of Danceability for Mexico and USA Top 50

Key observations

- Both Valence and Danceability means for Mexico's Top 50 music are greater than the US
- The danceability scores for Mexico's Top 50 are more compact around the mean
- This leads us to believe that our hypothesis will be true

Testing: MANOVA

MANOVA		
Variable	**Metric**	**Value**
Danceability	p-value	< 0.001
Valence	p-value	< 0.001

MANOVA hypotheses

- Null: Danceability and Valence scores ***are not*** dependent on the playlist label
- Alt: Danceability and Valence scores ***are*** dependent on their playlist label

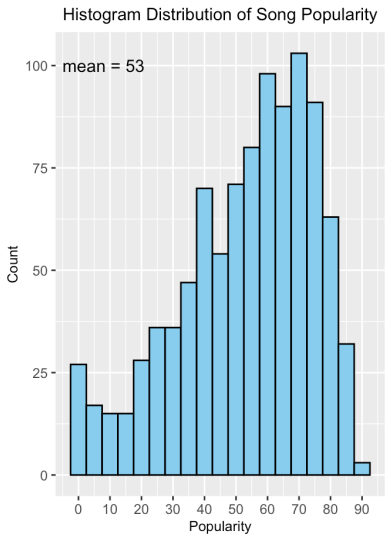
Conclusion

We found that mean scores for Valence and Danceability are higher for Mexico's Top 50 playlist than the means for those of the US Top 50 playlist

We think this could be due to Spanish being a warmer language and movement being an important part of Mexican expression

Are there any song characteristics that can help us predict popularity scores?

EDA



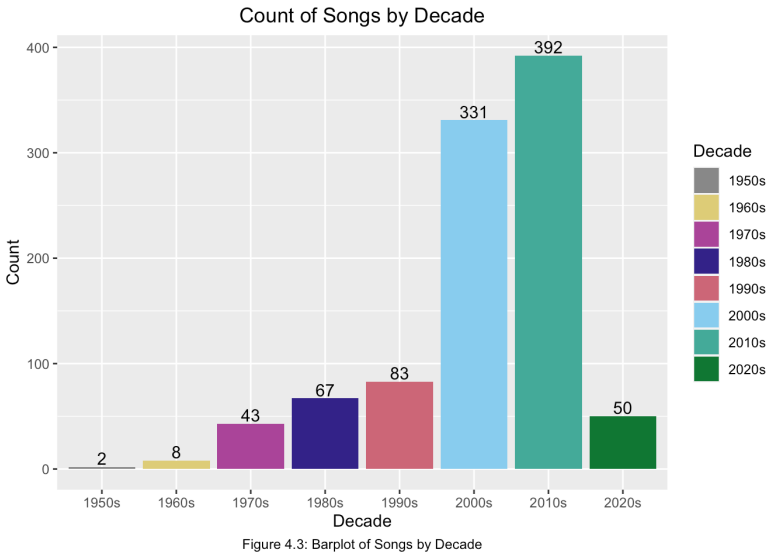


Figure 4.3: Barplot of Songs by Decade

Key observations

- Popularity's distribution is left skewed and has a mean of 53
- Songs in this playlist tend to have a relatively high popularity score
- Most songs come from the time period 2000-2020, which means we have unbalanced data

EDA*

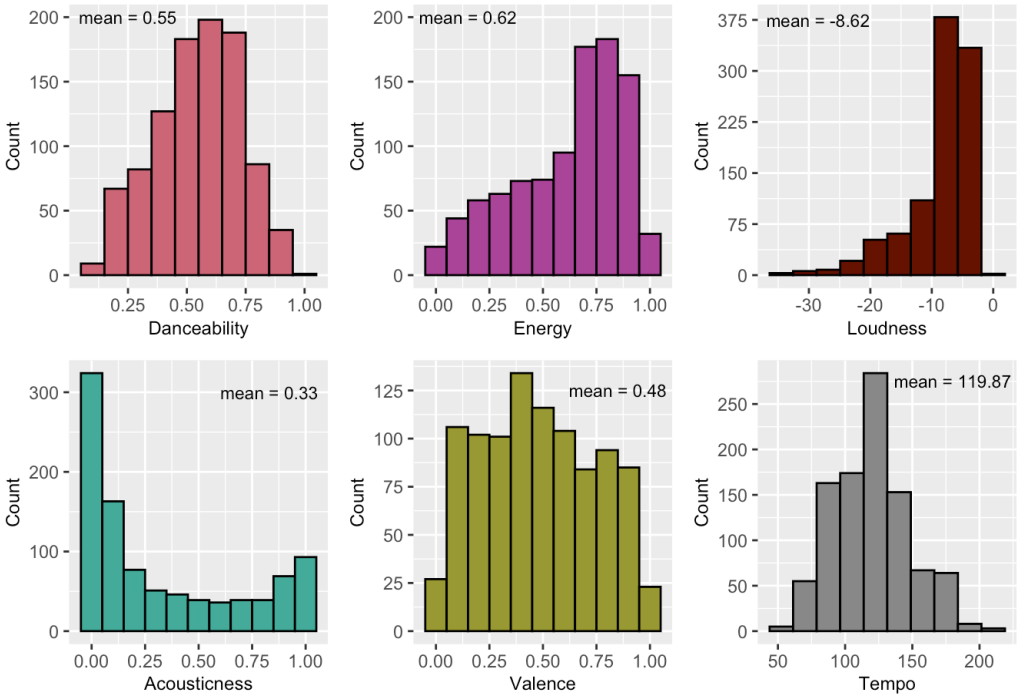


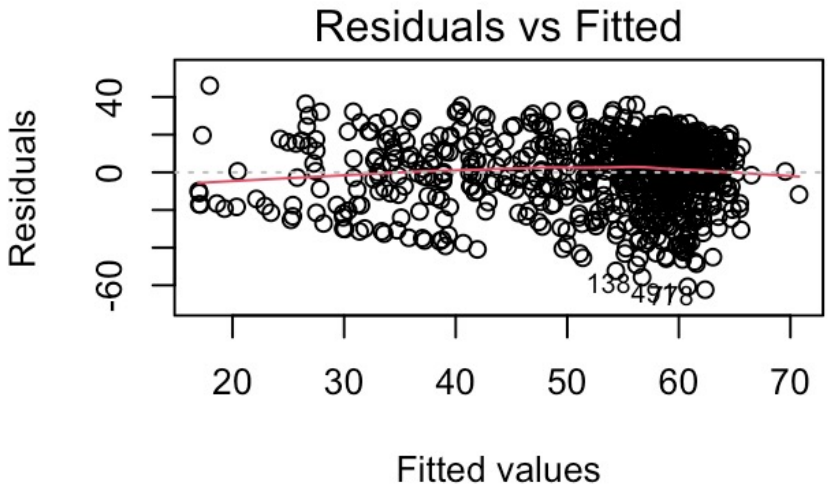
Figure 4.2: Histograms of Predictor Variables

Key observations

- Data isn't balanced across all variables
- Loudness has a high mean and strongly skewed to the left
- Danceability gets closest to a normal distribution
- Valence is mostly like a uniform distribution

*Some variables omitted from EDA

Testing: Regression



Results

- R2: 0.24
- Significant variables:
  - Valence
  - Loudness*Acousticness
  - Valence*Instrumentalness
- Not good at predicting popularity score
- Tried 4 models, but they all yielded a similar R2

Conclusion

We thought we'd be able to predict popularity scores using some of the data's variables, but we found that our model is unable to explain the variation of popularity scores

Linear regression is not fit to predict popularity based on the available variables. It's possible that another model could decipher what makes a song popular

Conclusions

Music characteristics are not equal within the same language, it depends on culture

Qualities of popular music have evolved across decades and mirrors the culture of the time

Warmer climate might have an effect on music rhythm, lyrics, and tempo

Popularity scores are not tangible, it likely comes from abstract concepts and social climate

Questions or comments?

Thanks!