

# Comparing Music Features Across the US, Canada, UK, and Australia

Chenming Zhao, Yao Xingjie

## Table of contents

<b>Introduction</b>	<b>1</b>
<b>Data</b>	<b>2</b>
<b>Analysis</b>	<b>3</b>
Introduction of ANOVA . . . . .	3
1. ANOVA Table Structure: . . . . .	4
2. Interpretation of the Results: . . . . .	4
<b>Discussion</b>	<b>5</b>
<b>Conclusion</b>	<b>5</b>

## Introduction

The aim of this research is to analyze and compare the music feature preferences of four countries: the United Kingdom, the United States, Australia, and Canada. We focus on determining if there are significant differences in music preferences across these countries. The analysis is based on data extracted from “Spotify API Documentation” (2024), with features such as danceability, energy, and others.

Based on results in Section and Section , this analysis finds most of the distributions overlap heavily, suggesting homogeneity in terms of musical tastes for these attributes across these four regions.

## Data

The data for this analysis is collected using the “Spotify API Documentation” (2024), and R code from R Core Team (2023) is used to load, clean and analyse our data, also R packages Wickham et al. (2023), Wickham et al. (2024), Wickham (2016), Xie (2024), Xie (2015), Xie (2014), Pedersen (2024), Thompson et al. (2022) are used.

The countries included in this study are the UK, US, Australia, and Canada. The collected features for analysis include: - **Danceability** - **Energy** - **Valence** - **Speechiness**

country	danceability	energy	valence	speechiness
GB	0.674	0.907	0.721	0.0640
GB	0.521	0.592	0.535	0.0304
GB	0.700	0.582	0.785	0.0356
GB	0.747	0.507	0.438	0.0358
GB	0.769	0.722	0.570	0.0507
GB	0.660	0.756	0.838	0.0320

From Figure 1 we can see following information:

1. Danceability across Countries: The median danceability is similar across all countries, around 0.6, with slight variation. The interquartile range (IQR) is narrow, indicating that most songs have a danceability score within a similar range in all countries. There are no outliers, showing consistency in this attribute.
2. Energy across Countries: The energy levels also have a similar median value, around 0.6–0.7. The IQR is slightly wider than Danceability, indicating more variability in energy levels across countries. No significant outliers are present, showing relatively consistent distributions.
3. Valence across Countries: Valence, which indicates the positivity of the music, has a wider range with a median around 0.5 in all countries. The variability (IQR) is larger, suggesting that songs in all countries vary more widely in emotional positivity. No major outliers are present, so the distribution remains consistent across countries.
4. Speechiness across Countries: Speechiness, indicating the presence of spoken words in the music, has a lower median value compared to the other attributes (around 0.05–0.1). The IQR is much smaller, and there are multiple outliers, especially in Australia and the US. This suggests that while most songs have low speechiness, some songs deviate significantly from the norm, potentially being more speech-heavy.

Overall, Danceability, Energy, and Valence have similar medians and distribution across the four countries, indicating that the general characteristics of music in these regions do not differ significantly. Speechiness, however, shows more variability, with several outliers, especially in

Australia and the US. This could reflect specific features or types of music in these countries that contain more spoken words.

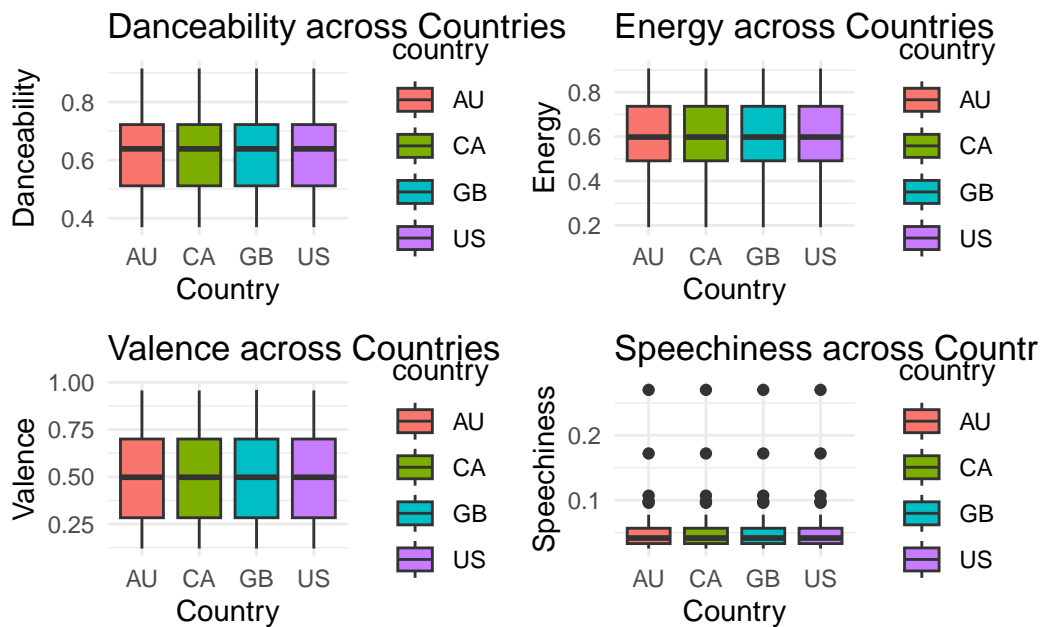


Figure 1

## Analysis

Based on Table 2, the ANOVA results correspond to different musical features across four countries.

### Introduction of ANOVA

ANOVA (Analysis of Variance) is a statistical method used to compare the means of two or more groups. It examines whether observed differences between groups are statistically significant by comparing group variance (between-group variability) to within-group variance (within-group variability). If the between-group variance is significantly larger than the within-group variance, it suggests that at least one group mean differs.

Then, column names in ANOVA table structure will be explained.

## 1. ANOVA Table Structure:

Each ANOVA table presents the following information for one of the musical features (Danceability, Energy, Valence, Speechiness):

- **Df (Degrees of Freedom):** This represents the number of independent values or quantities which can be assigned to a statistical distribution. For the **country** factor, the degrees of freedom are 3 (since there are four countries: AU, CA, GB, and US, so  $(n - 1 = 4 - 1 = 3)$ ). The degrees of freedom for **Residuals** are 196, which is the total number of observations minus the number of groups (e.g.,  $(200 - 4 = 196)$ ).
- **Sum Sq (Sum of Squares):** This represents the total variation explained by the **country** factor and the residual variation. The “Sum Sq” under **country** measures how much variation in the feature (Danceability, Energy, etc.) can be attributed to differences between countries, while “Sum Sq” under **Residuals** reflects the remaining unexplained variation.
- **Mean Sq (Mean Square):** This is the average of the squared differences, calculated by dividing the “Sum of Squares” by the corresponding degrees of freedom. It shows how much variation exists within the groups (in this case, between countries) and within the residuals.
- **F Value:** The F-statistic is the ratio of the variance between groups (countries) to the variance within groups (residuals). A higher F value indicates that the group means are different from each other. If the F value is close to 0, it suggests no significant difference between groups.
- **Pr(>F):** This is the p-value, which indicates the probability that the observed differences between group means occurred by chance. If the p-value is less than a certain threshold (commonly 0.05), it suggests that the group differences are statistically significant.

## 2. Interpretation of the Results:

From Table 2, we can see there is no significant variation in four features across the countries. The F values are all near 0, and the p-values are 1, indicating no difference between the countries for these features. Hence, we observe that none of the musical features show statistically significant differences between the countries. This suggests that the variation in these musical features is not attributable to the country of origin.

Table 2: ANOVA analysis of music features across four countries

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
country	3	0.000000	0.0000000	0	1
Residuals	196	3.120782	0.0159224	NA	NA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
country	3	0.0000086	0.0000029	9.61e-05	0.9999987
Residuals	196	5.8741014	0.0299699	NA	NA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
country	3	0.000001	0.0000003	5.4e-06	1
Residuals	196	11.508551	0.0587171	NA	NA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
country	3	0.0000000	0.000000	3.7e-06	1
Residuals	196	0.3237844	0.001652	NA	NA

## Discussion

This analysis only have small sample, which can reduce the power of ANOVA to detect significant differences. This may have contributed to the lack of significant findings. Also, only four musical attributes were analyzed. Future analysis could be expanded to include more features such as tempo, or genre, which might reveal trends not captured by the current attributes.

## Conclusion

The results in Section and Section suggest no significant differences in musical preferences (in terms of Danceability, Energy, Valence, and Speechiness) between Australia, Canada, Great Britain, and the United States. However, this finding should be interpreted cautiously due to possible limitations in the dataset, the narrow selection of attributes, and potential statistical issues such as small sample sizes and lack of enough features. Expanding the analysis to include other musical features, genres, or a longitudinal component may yield different insights.

Pedersen, Thomas Lin. 2024. *Patchwork: The Composer of Plots*. <https://patchwork.data-imaginist.com>.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

“Spotify API Documentation.” 2024. <https://developer.spotify.com/documentation/web-api/>.

Thompson, Charlie, Daniel Antal, Josiah Parry, Donal Phipps, and Tom Wolff. 2022. *Spotifyr: R Wrapper for the Spotify Web API*. <https://github.com/charlie86/spotifyr>.

- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, Dewey Dunnington, and Teun van den Brand. 2024. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.
- . 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- . 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.