# ARTIST PREFERENCES AND CULTURAL, SOCIO-ECONOMIC DISTANCES ACROSS COUNTRIES: A BIG DATA PERSPECTIVE

**Meijun Liu**
The University of Hong Kong
meijun@connect.hku.hk

**Xiao Hu**
The University of Hong Kong
xiaoxhu@hku.hk

**Markus Schedl**
Johannes Kepler University Linz
markus.schedl@jku.at

**ABSTRACT**

Users in different countries may have different music preferences, possibly due to geographical, economic, linguistic, and cultural factors. Revealing the relationship between music preference and cultural socio-economic differences across countries is of great importance for music information retrieval in a cross-country or cross-cultural context. Existing works are usually based on small samples in one or several countries or take only one or two socio-economic aspects into account. To bridge the gap, this study makes use of a large-scale music listening dataset, LFM-1b with more than one billion music listening logs, to explore possible associations between a variety of cultural and socio-economic measurements and artist preferences in 20 countries. From a big data perspective, the results reveal: 1) there is a highly uneven distribution of preferred artists across countries; 2) the linguistic differences among these countries are positively associated with the distances in artist preferences; 3) country differences in three of the six cultural dimensions considered in this study have positive influences on the difference of artist preferences among the countries; and 4) geographical and economic distances among the countries have no significant relationship with their artist preferences.

## 1. INTRODUCTION

Probing the relationship between cultural and socio-economic difference and the cross-country difference in music preferences not only matters in music information retrieval (MIR), but also brings important cues to understand the difference in cultural and socio-economic aspects among countries. Against the background of the world's diversity in many cultural and socio-economic aspects, research aiming to uncover cross-country differences in the field of music recommendation and retrieval is seeing increasing attention [22, 30]. It is widely acknowledged that music information behavior is inherently a kind of cultural behavior, shaped by the culture and other socio-economic factors [32, 56]. A growing body of literature demonstrated that different cultures or different countries have disparity in music information behaviors, e.g. music retrieval, management and consumption, and music mood judgment [30, 39, 44]. This is also true in music preferences [25, 56]. In this case, a question naturally arises: which kind of cultural and socio-economic background might possibly be responsible for the difference of music preferences among countries? It is thus necessary to have an in-depth understanding of the differences in music preferences across different countries and of how these differences are mirrored by cultural and socio-economic factors. Answers to these questions can facilitate constructing cross-cultural MIR systems, and promoting music recommendation and retrieval results by taking into account cultural and the socio-economic background of users [56]. Furthermore, this paper also contributes to improving the knowledge of the differences in customs, traditions, cultural values, and other socio-economic factors among countries.

Existing literature provides little evidence of the exact relationships between the cross-country differences in a variety of cultural and socio-economic factors and those in artist preference. Furthermore, limited literature investigated which cultural dimension reflects the inter-country difference in artist preferences. Even fewer previous studies were based on large-scale user-generated datasets. This situation calls for more studies in this regard. Therefore, we investigate in this paper the following research questions:

**RQ1:** How do artist preferences differ across countries?

**RQ2:** Does the inter-country difference in artist preferences depend on the geographic, economic, linguistic, and cultural distances among countries?

**RQ3:** Which cultural dimension can reflect the difference in artist preferences across countries?

Inspired by this research gap and the scientific importance, this paper seeks to probe whether the differences among countries in music taste rely on any factors in the cultural and socio-economic dimensions, through applying descriptive analysis, Kruskal-Wallis variance analysis and Quadratic Assignment Procedure (QAP) on a large dataset with more than one billion listening records, the LFM-1b dataset [49]. To our knowledge, this is a first work that explores relationship between the inter-country difference in artist preferences and a variety of cultural and socio-economic differences among countries.


**Attribution:** Meijun Liu, Xiao Hu, Markus Schedl. "Artist Preferences and Cultural, Socio-economic Distances across Countries: A Big Data Perspective", 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

## 2. RELATED WORK

Related work can be categorized into research that investigates the connection between music preferences and socio-economic factors, and that between music preferences and cultural dimensions. A recent study analyzed the country-specific music preferences. However, it did not investigate the influential factors of music preferences [50].

In recent years, due to the availability of large-scale music listening data, users geospatial context for music recommendation has received increasing attention [53]. However, there is limited literature on directly exploring the relationship between music preferences and geographic locations. Before large-scale music listening datasets have been published, several works involved location-related information. Researchers proposed a mobile music recommender system, Lifetrack, that enables a playlist based on the location and other information in the users environment [47]. More recently, researchers found that drawing on the information of listeners geographic location help promote music recommendation [43, 52]. Although some works have been done, the authors suggested that combing cultural regions with geographical distances may better explain differences in music taste [53].

Economic status seems to have a potential influence on musical preferences. Cultural consumption is closely related to individuals social status that in turn is directly interrelated with the amount of income. According to Bourdieu's class theory, high-status groups have more cultural capital which is defined as knowledge and appreciation of highbrow culture, and the possession of high or low income in people's childhood tend to shape their taste [7]. Empirical evidence showed that the cultural taste of the high-status group is distinct from people in other classes. For example, people belonging to the high-status group frequently visit museums, classical concerts, the theater and so forth [15, 31, 33]. In the field of music research, there exist some evidence that support the connection between income and music preferences as well. Cutler found that preference for classical music tended to grow with income [13]. Duncan, Herrington and Capella also suggested that the music taste of upper-income individuals is different from their counterparts with low income or/and with only high school education [18]. It has been found that high socio-economic status positively impacts musical openness that is related to the acceptance for diverse music [57].

Culture is a well-discussed factor in music information research, compared to other socio-economic aspects. From the perspectives of sociology, psychology and behavior science, researchers believe that general behaviors and preferences are shaped by culture [32]. In the field of MIR, retrieval methods that consider cultural differences in music perception and consumption are highly desirable [39]. In recent years, taking cultural factors into account has become a frequently-used strategy in MIR research to explore users music need at the country level [23]. Researchers found that preference for music mood altered significantly between countries, implying that utilizing geographic information of users could facilitate further studies [48]. More recently, it has been suggested that the country-based diversity pattern of music listening is associated with some cultural dimensions presented in Hofstede's theory on cultural dimensions [23]. Specifically, researchers found that users in countries with high scores in the culture dimension of power distance tended to show less diversity in the artists and genres they listened to. Oppositely, individualism dimension was negatively correlated with music diversity. Furthermore, the correlation between long-term orientation and artist diversity was considered negative.

Based on small-sample data obtained from surveys, previous studies provided more direct evidence to show the influence of language on listeners reactions to and comments on music. Empirical results presented that there was a significantly positive correlation between familiarity with a language and attitude toward the language in songs [1]. Specifically, it was reported that some children responded to foreign-language music with negative judgment [38]. In a study which focused on language in the context of songs, it was observed that English-speaking students preferred pop songs performed in English to those with Spanish or Chinese lyrics [2]. By examining undergraduate non-music majors world music preferences, researchers found that the breadth and length of studying foreign languages were related to a high degree of world music preferences [26].

In a nutshell, the cultural and socio-economic variables we selected are thought of as potentially correlative factors of music preferences. However, the exact relationship between these factors and music preferences under a cross-country context still remains unclear. First, current studies are limited to small samples collected from surveys or questionnaires. Besides, such self-reported responses can be subjective. In other words, there is scarce literature that investigates this research question using objective datasets in a large scale. Second, most existing studies simply include one or two socio-economic factors, leaving many potentially relevant aspects unexplored. Third, extant studies ignored the discussion of the association between socio-economic factors and music taste in a context of multiple countries or multiple cultures, since a majority of them paid attention to individuals in a single cultural environment. Fourth, among the few studies on relationship between music preferences and socio-economic factors (e.g. geographical location), the conclusions are often ambiguous and indecisive. To bridge the gaps, this study aims to uncover the relationship between music preferences and cultural and socio-economic factors at the country level using a large-scale and user-generated dataset.

## 3. DATA AND METHODOLOGY

### 3.1 LFM-1b Dataset

This study uses the open dataset LFM-1b [1] [49]. This dataset includes more than one billion music listening

[1] www.cp.jku.at/datasets/LFM-1b. The period during which the data was collected ranges from January 2013 to August 2014.

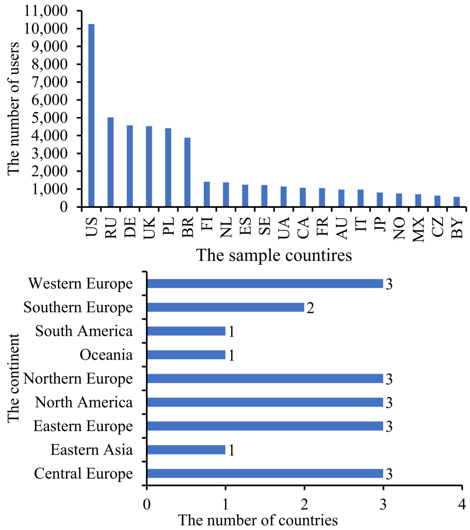


**Figure 1**: Number of users in the 20 sample countries (top) and the continents where countries are located (bottom).(Country code: US: United States, RU: Russia, DE: Germany, UK: United Kingdom, PL:Poland, BR: Brazil, FI: Finland, NL: Netherlands, ES: Spain, SE: Sweden, UA: Ukraine, CA: Canada, FR: France, AU: Australia, IT: Italy, JP: Japan, NO: Norway, MX: Mexico, CZ: Czech Republish, BY: Belarus)

events created by 120,322 users and enables us to conduct a large-scale analysis in music listening behaviors. It is noteworthy that only 54.13% of users in the LFM-1b dataset provide information on their nationality and the distribution of users across countries is very unbalanced. To avoid possible negative effects on our analysis, we eliminate countries with less than 1% of users in LFM-1b and only use the remaining 20 countries in this study. Finally, we obtained a dataset including 46,619 users with 678,640,512 listening events that cover 2,259,103 unique artists.

The distribution of users in the sampling countries and which continent these countries belong to are shown in Figure 1. It indicates that most of them are located in Europe and America, with one country in Asia, Oceania, and South America respectively.

### 3.2 Modeling Country-specific Diversity in Artist Preferences

In this study, we used the coefficient of variation (CV) and Gini coefficient to measure and compare the diversity of artist preferences across the countries. Coefficient of variation is a standardized measure of dispersion of the frequency distribution, which is defined as the ratio of the standard deviation to the average of a variable [20]. CV has been frequently used for comparing diversity or inequality in groups [3, 5]. The Gini index enables us to examine the inequality of artist listening frequency in each country [46, 60]. We adapt the definition of the Gini coefficient for a country $c$ to our task and calculate it as shown in Equation 1,

$$G_c = \left| 1 + \frac{1}{N} - \frac{2}{m \times N^2} \sum_i (N - O_i + 1) \cdot y_i \right| \quad (1)$$

where $N$ is the number of artists listened to by users in country $c$; $y_i$ is the listening count of artist $i$ in country $c$; $O_i$ is the inverse rank of $y_i$ when sorting the values $y_i$ for all artists $i$ in country $c$, and $m$ is arithmetic mean of listening counts across the $N$ artists.

We adopted the Kruskal-Wallis (KW) non-parametric analysis of variance as the primary tool to probe whether there is a significant difference among countries in the frequency of artist listening. After performing the Shapiro-Wilk test, it was observed that the data exhibited non-normal distribution, and thus non-parametric analysis of variance was adopted [19]. A follow-up test was carried out to find out which pairs of countries have significant differences [9, 29, 54].

To avoid possible bias caused by the disequilibrium of listening counts across countries, we also normalized the listening frequency of each artist in a country against the total listening count of that country. In other words, we look into not only the raw listening counts but also the normalized listening count of each artist.

### 3.3 Modeling Country Distances in terms of Artist Preferences, Cultural and Socio-economic Dimensions

The *distance of artist preference* among countries is the dependent variable in this study. Based on the data of listening events in LFM-1b, we calculated the cosine distance of artist preferences among countries. Specifically, each country is represented by a vector of artists, with each dimension of the vector being the number of times the corresponding artist was listened to by users in this country. Then, the cosine distance between each pair of vectors was calculated. The results are shown in Figure 2. Notably, the distances between Japan and all other countries are substantially higher ($> 0.5$) than those between other pairs of countries, making Japan an outlier, which is in line with previous studies [51].

In this study, the cultural and socio-economic distance between countries is measured by the following aspects: geographic, economic, linguistic, and cultural distance. *Geographic distance* is the geodetic distance between the capital cities calculated by Vincenty's equations and on the basis of the latitude and longitude, i.e., the length of the shortest curve between two points along the surface of the Earth [58]. We define *economic distance* as the difference of gross domestic product (GDP) per capita between countries, calculated based on the data obtained from World Bank [2]. For *linguistic distance* between countries, we regard the language used by the largest population in a country as the main language in that country. On the website of the Central Intelligence Agency (CIA), [3] the language and

[2] http://databank.worldbank.org/data/home.aspx
[3] https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html

| | AU | BR | BY | CA | CZ | DE | ES | FI | FR | IT | JP | MX | NL | NO | PL | RU | SE | UA | UK | US |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AU | 0.00 | 0.36 | 0.44 | 0.15 | 0.36 | 0.34 | 0.27 | 0.46 | 0.25 | 0.31 | 0.57 | 0.34 | 0.21 | 0.27 | 0.37 | 0.36 | 0.37 | 0.39 | 0.12 | 0.13 |
| BR | 0.36 | 0.00 | 0.46 | 0.32 | 0.41 | 0.43 | 0.31 | 0.49 | 0.39 | 0.35 | 0.61 | 0.33 | 0.31 | 0.38 | 0.35 | 0.39 | 0.43 | 0.43 | 0.29 | 0.36 |
| BY | 0.44 | 0.46 | 0.00 | 0.41 | 0.35 | 0.34 | 0.39 | 0.42 | 0.40 | 0.44 | 0.69 | 0.46 | 0.36 | 0.43 | 0.32 | 0.11 | 0.49 | 0.11 | 0.41 | 0.49 |
| CA | 0.15 | 0.32 | 0.41 | 0.00 | 0.34 | 0.34 | 0.25 | 0.42 | 0.23 | 0.28 | 0.55 | 0.32 | 0.21 | 0.25 | 0.34 | 0.34 | 0.35 | 0.36 | 0.14 | 0.10 |
| CZ | 0.36 | 0.41 | 0.35 | 0.34 | 0.00 | 0.34 | 0.33 | 0.41 | 0.32 | 0.36 | 0.65 | 0.42 | 0.28 | 0.35 | 0.30 | 0.30 | 0.43 | 0.31 | 0.31 | 0.41 |
| DE | 0.34 | 0.43 | 0.34 | 0.34 | 0.34 | 0.00 | 0.33 | 0.38 | 0.35 | 0.41 | 0.66 | 0.46 | 0.26 | 0.35 | 0.33 | 0.28 | 0.40 | 0.32 | 0.31 | 0.39 |
| ES | 0.27 | 0.31 | 0.39 | 0.25 | 0.33 | 0.33 | 0.00 | 0.42 | 0.27 | 0.27 | 0.59 | 0.30 | 0.22 | 0.28 | 0.32 | 0.32 | 0.37 | 0.34 | 0.21 | 0.29 |
| FI | 0.46 | 0.49 | 0.42 | 0.42 | 0.41 | 0.38 | 0.42 | 0.00 | 0.43 | 0.48 | 0.68 | 0.51 | 0.38 | 0.37 | 0.39 | 0.39 | 0.42 | 0.38 | 0.43 | 0.50 |
| FR | 0.25 | 0.39 | 0.40 | 0.23 | 0.32 | 0.35 | 0.27 | 0.43 | 0.00 | 0.29 | 0.56 | 0.32 | 0.21 | 0.28 | 0.33 | 0.33 | 0.38 | 0.35 | 0.21 | 0.26 |
| IT | 0.31 | 0.35 | 0.44 | 0.28 | 0.36 | 0.41 | 0.27 | 0.48 | 0.29 | 0.00 | 0.59 | 0.34 | 0.27 | 0.31 | 0.36 | 0.36 | 0.42 | 0.39 | 0.24 | 0.32 |
| JP | 0.57 | 0.61 | 0.69 | 0.55 | 0.65 | 0.66 | 0.59 | 0.68 | 0.56 | 0.59 | 0.00 | 0.59 | 0.58 | 0.61 | 0.66 | 0.63 | 0.66 | 0.67 | 0.53 | 0.55 |
| MX | 0.34 | 0.33 | 0.46 | 0.32 | 0.42 | 0.46 | 0.30 | 0.51 | 0.32 | 0.34 | 0.59 | 0.00 | 0.33 | 0.40 | 0.40 | 0.39 | 0.46 | 0.42 | 0.30 | 0.35 |
| NL | 0.21 | 0.31 | 0.36 | 0.21 | 0.28 | 0.26 | 0.22 | 0.38 | 0.21 | 0.27 | 0.58 | 0.33 | 0.00 | 0.21 | 0.27 | 0.29 | 0.32 | 0.31 | 0.15 | 0.25 |
| NO | 0.27 | 0.38 | 0.43 | 0.25 | 0.35 | 0.35 | 0.28 | 0.37 | 0.28 | 0.31 | 0.61 | 0.40 | 0.21 | 0.00 | 0.34 | 0.38 | 0.29 | 0.38 | 0.23 | 0.29 |
| PL | 0.37 | 0.35 | 0.32 | 0.34 | 0.30 | 0.33 | 0.32 | 0.39 | 0.33 | 0.36 | 0.66 | 0.40 | 0.27 | 0.34 | 0.00 | 0.28 | 0.43 | 0.29 | 0.32 | 0.42 |
| RU | 0.36 | 0.39 | 0.11 | 0.34 | 0.30 | 0.28 | 0.32 | 0.39 | 0.33 | 0.36 | 0.63 | 0.39 | 0.29 | 0.38 | 0.28 | 0.00 | 0.43 | 0.10 | 0.31 | 0.40 |
| SE | 0.37 | 0.43 | 0.49 | 0.35 | 0.43 | 0.40 | 0.37 | 0.42 | 0.38 | 0.42 | 0.66 | 0.46 | 0.32 | 0.29 | 0.43 | 0.43 | 0.00 | 0.45 | 0.31 | 0.36 |
| UA | 0.39 | 0.43 | 0.11 | 0.36 | 0.31 | 0.32 | 0.34 | 0.38 | 0.35 | 0.39 | 0.67 | 0.42 | 0.31 | 0.38 | 0.29 | 0.10 | 0.45 | 0.00 | 0.36 | 0.44 |
| UK | 0.12 | 0.29 | 0.41 | 0.14 | 0.31 | 0.31 | 0.21 | 0.43 | 0.21 | 0.24 | 0.53 | 0.30 | 0.15 | 0.23 | 0.32 | 0.31 | 0.31 | 0.36 | 0.00 | 0.13 |
| US | 0.13 | 0.36 | 0.49 | 0.10 | 0.41 | 0.39 | 0.29 | 0.50 | 0.26 | 0.32 | 0.55 | 0.35 | 0.25 | 0.29 | 0.42 | 0.40 | 0.36 | 0.44 | 0.13 | 0.00 |

**Value:** 0.0 0.5 1.0

**Figure 2**: Heat map of cosine distances among countries based on artist listening frequency

the size of its speakers in a country are obtained to identify the main language in each of the sampling countries. Ethnologue [4] provides the information of the global language family tree, based on which we calculated the linguistic distance between two countries. Consistent with existing literature [21, 37], the linguistic distance between two languages $i$ and $j$ is defined in Equation 2,

$$D_{ij} = 1 - \left( \frac{|N_i \cap N_J|}{\frac{1}{2} \cdot (N_i + N_j)} \right)^{\beta} \tag{2}$$

where $N_i$ denotes the number of nodes in country $i$'s language tree, $N_j$ analogously. The relative distance between languages which are both included in the same family hinges on the value of $\beta$. According to the experience in other studies [21, 37], we set $\beta = 0.5$. For example, in the language family tree, English belongs to the following branch: Indo-European $>$ Germanic $>$ West $>$ English while Swedish is classified into this branch: Indo-European $>$ Germanic $>$ North Germanic $>$ East Scandinavian $>$ Continental Scandinavian $>$ Swedish. The distance between these two languages is approximately $0.368$ in that they have four and six nodes separately, sharing two common nodes. In order to quantify the *cultural distance* between countries, we calculate the distance of scores between countries in each of Hofstede's cultural dimensions [5] [28]: **Power distance index** (PDI) refers to the extent to which the less powerful members accept and expect that power is distributed unequally; **Individualism** (IDV) defines the degree of preference for a loosely-knit or tightly-knit social framework; **Masculinity** (MAS) refers to the degree of preference for achievement, heroism, assertiveness, and material rewards for success; **Uncertainty avoidance** (UAI) expresses the attitude of individuals towards uncertainty and ambiguity; **Long-term orientation** (LTO) describes to which degree a society ties the past with the present and future actions or challenges; **Indulgence** (IND) measures the happiness of a society.

[4] https://www.ethnologue.com/

[5] https://geert-hofstede.com/national-culture.html

### 3.4 Quadratic Assignment Procedure

In this study, we applied the Quadratic Assignment Procedure (QAP) [36, 55] via Double Dekker Semi-partialling [4, 14] to examine the relationship between distance of artist preference across countries, and geographic, economic, linguistic and cultural proximities across countries. In other words, we explore whether the difference among countries in artist listening has a relationship with their differences in the aspects of geographic location, economy, languages and culture. The primary reason for using QAP in this study is to avoid biases caused by autocorrelation of error in the dyadic dataset [55]. In this study, each observation is a pair of countries (i.e., a dyad in network analysis terminologies). Dyads are non-independent because each node in a dyad is connected to other dyads. Therefore, regression methods that assume independent distribution of data such as ordinary least squares (OLS) regression would lead to biased estimators [4, 6, 41]. In contrast, QAP explicitly takes into account dependence between dyads as well as autocorrelation of errors in the dyadic dataset. It is frequently used in regression analyses on network and relationship datasets [8, 10, 11, 16, 36]. The independent variables are the six matrices of the between-country distances in the six cultural dimensions, whereas the dependent variable is the matrix of inter-country distance on the artist preferences. We control the geographical distance (GEO), economic distance (ECO) and linguistic distance (LAN) among the countries. Besides, we calculated the mean variance inflation factor score (1.79), which is far lower than critical point 10, implying that multicolinearity can be ignored in this study.

## 4. RESULTS

### 4.1 Differences among Countries in terms of Artist Preferences

Table 1 presents statistics of the artist listening histories across countries, including the average number of listening events to an artist, the standard deviation (SD), coefficient of variation (CV), the number of unique artists (Uniq.#) listened to by listeners in each country and Gini coefficient (Gini).

As can be seen from Table 1, users from the US and Russia listen to a large number of unique artists, far exceeding other countries. Furthermore, the high CV values for US, BR, RU, PL and UK imply that listeners from these countries listen to a wider range of artists, compared to users from other countries.

| Country | Mean | SD | CV | Uniq.# | Gini |
|---|---|---|---|---|---|
| US | 182.16 | 2930.63 | 16.09 | 747004 | 96.44% |
| RU | 114.46 | 1795.28 | 15.69 | 632460 | 96.38% |
| DE | 134.46 | 1746.64 | 12.99 | 474874 | 96.35% |
| UK | 127.68 | 1709.56 | 13.39 | 456456 | 95.39% |
| PL | 209.95 | 3280.39 | 15.62 | 362155 | 95.21% |
| BR | 203.11 | 3263.22 | 16.07 | 267186 | 95.15% |
| NL | 83.57 | 833.97 | 9.98 | 256895 | 94.53% |
| UA | 68.38 | 743.48 | 10.87 | 249287 | 93.98% |
| SE | 102.94 | 1095.51 | 10.64 | 229714 | 93.38% |
| FI | 114.41 | 1230.23 | 10.75 | 213645 | 93.10% |
| FR | 71.79 | 595.26 | 8.29 | 207878 | 93.05% |
| CA | 93.83 | 817 | 8.71 | 191728 | 92.97% |
| ES | 82.3 | 717.39 | 8.72 | 190671 | 92.96% |
| JP | 63.44 | 548.04 | 8.64 | 185128 | 92.95% |
| BY | 51.78 | 469.26 | 9.06 | 166465 | 92.74% |
| NO | 78.52 | 669.54 | 8.53 | 165663 | 92.66% |
| AU | 89.02 | 759.82 | 8.54 | 164145 | 92.52% |
| IT | 81.28 | 783.94 | 9.64 | 156599 | 92.03% |
| MX | 73.17 | 753.87 | 10.3 | 144930 | 91.51% |
| CZ | 87.8 | 743.38 | 8.47 | 127726 | 91.15% |
| **Mean** | **105.70** | **1274.32** | **11.05** | **279530** | **93.72%** |

**Table 1**: Statistics of artist listening frequency across sample countries ranked by the number of unique artists

In general, Gini indices are high for all countries, meaning users' preferences for an artist varied a lot. In particular, the inequality of artist listening is most noticeable in the US, Brazil, Poland, and the UK, which is consistent with the CV results.

When comparing frequency of artist listening across countries, the result of the Kruskal-Wallis test shows a statistically significant difference ($p < 0.01$) among the sampling countries. After conducting a follow-up pairwise comparison, we find that significant differences on artist preferences exist between all 190 country pairs, except for BR and AU, CZ and AU, CZ and BR, JP and CZ, MX and FR, PL and CA, RU and NO, SE and FI, SE and FR.

### 4.2 QAP Correlation and Regression Results

We run two models to test the relationship between artist preference (as represented by artist listening frequencies) distance among countries, and the geographical, economic, linguistic, and cultural distances among them. The QAP correlation coefficients among the variables are reported in Table 2, and the regression results are in Table 3. For comparison, only the control variables are included in model 1 and we added the independent variables to model 2. The adjusted $R^2$ in the two models are significant: 0.594 and 0.643 in model 1 and 2, respectively. In other words, nearly 59.4% of the variance in the matrix of the artist preference distances among countries can be explained by their distances in the geographic, economic and linguistic aspects; and 64.3% of the variance can be explained in model 2 with the addition of cultural distances.

The distance among countries in term of main languages is positively associated with their distance in artist preferences ($r$ = 0.745 in Table 2). In model 2, the coefficient of linguistic distance among countries is significant and positive ($\beta = 0.68$; $p < 0.001$). Furthermore, three dimensions of cultural distance among countries have positive effects on their artist preference distance: masculinity ($\beta = 0.13$; $p < 0.05$), long-term orientation ($\beta$=0.12; $p < 0.01$), and indulgence ($\beta$=0.14; $p < 0.05$).

Besides, the regression results in both model 1 and model 2 reveal that economic distance has no significant impact on the artist preference distance on the country level. Geographic distance has a significant impact on the dependent variable in model 1, but becomes insignificant when cultural distances are included in model 2.

## 5. DISCUSSION

We summarize our main findings in the following. The *distribution of music listening behavior across artists is highly uneven*. In particular, substantial inequality of artist preferences is found in the US, Brazil, Poland, Russia, and the UK. In comparing across the countries, there are significant distinctions in artist preferences within most of country pairs.

The *distance between the main languages used in countries is positively associated with the distance in their artist preferences*. This result could be attributed to the fact that familiarity is a key factor that influences music preference [17, 27]. Familiarity not only refers to having heard a music piece somewhere before, but can also be reflected by the degree of familiarity with the language in the songs [1]. Listeners may be less familiar with music sung in languages they know little about, and thus they may be less likely to listen to that kind of music.

Among the six cultural dimensions, *masculinity, long-term orientation, and indulgence distances between countries have positive correlations with their distances in artist preferences.* First, *masculinity* indicates the degree to which a culture delineates gender roles, and a masculine culture clearly differentiates the social expectations on males and females [42]. Previous literature pointed out that a huge gender difference in both the expression and perception of mood could be found in cultures high in masculinity. Other researchers also demonstrated that masculinity can explain the gender difference in personality traits [12]. It is generally agreed that music listening behavior and

| | ARTIST | GEO | ECO | LAN | PDI | IDV | MAS | UAI | LTO |
|---|---|---|---|---|---|---|---|---|---|
| ARTIST | 1 | | | | | | | | |
| GEO | 0.248 | 1 | | | | | | | |
| ECO | 0.122 | -0.017 | 1 | | | | | | |
| LAN | 0.745*** | 0.066 | 0.118 | 1 | | | | | |
| PDI | 0.149 | -0.034 | 0.516*** | 0.211 | 1 | | | | |
| IDV | 0.215 | 0.354* | 0.37** | 0.136 | 0.458** | 1 | | | |
| MAS | 0.34* | -0.056 | 0.102 | 0.317* | 0.005 | -0.106 | 1 | | |
| UAI | 0.144 | -0.101 | 0.352** | 0.241* | 0.566** | 0.266* | 0.12 | 1 | |
| LTO | 0.267** | 0.266** | 0.016 | 0.081 | 0.019 | 0.112 | 0.01 | 0.025 | 1 |
| IND | 0.269* | 0.112 | 0.334** | 0.14 | 0.416** | 0.326** | -0.037 | 0.398** | 0.346** 1 |

**Table 2**: QAP correlation coefficients (Note: significance levels: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$)

| Variable | Model 1 | Model 2 |
|---|---|---|
| GEO | 0.200* | 0.133 |
| ECO | 0.040 | 0.003 |
| LAN | 0.727*** | 0.683*** |
| PDI | | -0.035 |
| IDV | | 0.063 |
| MAS | | 0.131* |
| UAI | | -0.074 |
| LTO | | 0.122** |
| IND | | 0.140* |
| Adjusted $R^2$ | 0.594 *** | 0.642*** |
| N of Obs | 380 | 380 |

**Table 3**: The QAP regression result. (Note that all coefficients presented are standardized coefficients. Significance levels: ***$p < 0.001$,**$p < 0.01$, *$p < 0.05$)

emotion strongly interact with each other [34, 35]. Moreover, the correlation between personality and music behavior is documented in empirical studies [24, 45]. Consequently, it is possible that on the country level, the music preference difference and the cultural difference in masculinity interacted through the gender differences in terms of emotion and personality traits.

Second, prior studies offer evidence that people in countries scoring low in *long-term orientation* have a lower preference for listening to diverse artists since they value steadfastness and believe that traditions are to be honored and kept [23]. In other words, people in short-term oriented cultures may prefer to listen to more traditional music, and their music listening behavior is possibly more conservative. Furthermore, it is recently found that individuals in countries with long-term orientation tend to be more patient [59]. This characteristics may not only influence business activities, but also generate different listening behaviors across countries. For instance, in long-term oriented countries, people may be more likely to have the patience to listen to slow and long music. Future studies can further explore and test these hypotheses.

Third, in countries scoring high on *indulgence*, people tend to have more freedom in controlling their daily lives and in choosing the way to enjoy life. Given that listening to music is often regarded as an important entertainment activity, the cultural difference in the dimension of indulgence can possibly affect people's choices of music, and thus bringing about the differences in music preferences across countries.

In the final regression model (model 2 in Table 3), there is no significant association between geographical and economic distance on the music preference distance across countries. Perhaps geographical distance is no longer a barrier for people to access various music in today's highly connected information society. Therefore geolocation plays a less significant role in music preference compared to linguistic and cultural differences among countries. In terms of economic distance, although on the individual level, it is confirmed in the literature [40] that music preferences vary by the income level, this seems questionable on the country level. This discrepancy might be related to the correlation between people's cultural behaviors and social status [40]. On an individual level, income is related to social status which in turn can influence one's music preference. However, on the country level, people's social status ranges a lot in any single country and has virtually no relationship with the GDP per capita of a country. Consequently, economic distance among countries cannot explain differences in music preferences.

## 6. CONCLUSION AND FUTURE WORK

In this study, we applied descriptive statistical analysis, Krusal-Wallis variance analysis, and Quadratic Assignment Procedure on the LFM-1b dataset, to reveal the association between the distance of a variety of cultural and socio-economic aspects among countries, and the cross-country difference in artist preference.

Findings of this study contribute to the literature of music listening behaviors and preferences, particularly from the cross-country perspective. By analyzing one of the largest datasets in the field, we aim to draw conclusions that are representative and generalizable. Multiple factors in the cultural, linguistic, geographic, and economic aspects were analyzed, and the results can potentially help design new strategies of MIR systems in the cross-country and cross-cultural context. Future studies can compare cross-country differences on other facets of music such as genre and mood.

## 7. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their helpful comments. This study is partially supported by the Austrian Science Fund (FWF): P25655 and by a seed research grant funded by University of Hong Kong.


## 8. REFERENCES


[1] Carlos R Abril. Multicultural dimensions and their effect on children's responses to pop songs performed in various languages. *Bulletin of the Council for Research in Music Education*, pages 37–51, 2005.

[2] Carlos R Abril and Patricia J Flowers. Attention, preference, and identity in music listening by middle school students of different linguistic backgrounds. *Journal of Research in Music Education*, 55(3):204–219, 2007.

[3] Arthur G Bedeian and Kevin W Mossholder. On the use of the coefficient of variation as a measure of diversity. *Organizational Research Methods*, 3(3):285–297, 2000.

[4] Geoffrey G Bell and Akbar Zaheer. Geography, networks, and knowledge flow. *Organization Science*, 18(6):955–972, 2007.

[5] RB Bendel, SS Higgins, JE Teberg, and DA Pyke. Comparison of skewness coefficient, coefficient of variation, and gini coefficient as inequality measures within populations. *Oecologia*, 78(3):394–400, 1989.

[6] Stephen P Borgatti, Martin G Everett, and Linton C Freeman. Ucinet for windows: Software for social network analysis. 2002.

[7] Pierre Bourdieu. *Distinction: A social critique of the judgement of taste*. Harvard University Press, 1984.

[8] Val Burris. Interlocking directorates and political cohesion among corporate elites 1. *American Journal of Sociology*, 111(1):249–283, 2005.

[9] Herve M Caci et al. Kwallis2: Stata module to perform kruskal-wallis test for equality of populations. *Statistical Software Components*, 1999.

[10] Junho H Choi, George A Barnett, and BUM-SOO CHON. Comparing world city networks: a network analysis of internet backbone and air transport intercity linkages. *Global Networks*, 6(1):81–99, 2006.

[11] David T Connolly. An improved annealing scheme for the qap. *European Journal of Operational Research*, 46(1):93–100, 1990.

[12] Paul Costa Jr, Antonio Terracciano, and Robert R McCrae. Gender differences in personality traits across cultures: robust and surprising findings., 2001.

[13] B Cutler. North american demographics. *American Demographics*, 1992.

[14] David Dekker, David Krackhardt, and Tom AB Snijders. Sensitivity of mrqap tests to collinearity and autocorrelation conditions. *Psychometrika*, 72(4):563–581, 2007.

[15] Paul DiMaggio and John Mohr. Cultural capital, educational attainment, and marital selection. *American journal of sociology*, 90(6):1231–1261, 1985.

[16] Michael Dreiling and Derek Darves. Corporate unity in american trade policy: A network analysis of corporate-dyad political action 1. *American Journal of Sociology*, 116(5):1514–63, 2011.

[17] Kevin Droe. Music preference and music education: A review of literature. *Update: Applications of Research in Music Education*, 24(2):23–32, 2006.

[18] J Duncan Herrington and Louis M Capella. Practical applications of music in service settings. *Journal of Services Marketing*, 8(3):50–65, 1994.

[19] Alan C Elliott and Linda S Hynan. A sas® macro implementation of a multiple comparison post hoc test for a kruskal–wallis analysis. *Computer methods and programs in biomedicine*, 102(1):75–80, 2011.

[20] BS Everitt. The cambridge dictionary of statistics cambridge university press. *Cambridge, UK Google Scholar*, 1998.

[21] James D Fearon. Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8(2):195–222, 2003.

[22] Bruce Ferwerda and Markus Schedl. Investigating the relationship between diversity in music consumption behavior and cultural dimensions: A cross-country analysis. In *Proc. of the 1st Workshop on SOAP*, 2016.

[23] Bruce Ferwerda, Andreu Vall, Marko Tkalcic, and Markus Schedl. Exploring music diversity needs across countries. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, pages 287–288. ACM, 2016.

[24] Bruce Ferwerda, Emily Yang, Markus Schedl, and Marko Tkalcic. Personality traits predict music taxonomy preferences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2241–2246. ACM, 2015.

[25] Flavio Figueiredo, Bruno Ribeiro, Christos Faloutsos, Nazareno Andrade, and Jussara M Almeida. Mining online music listening trajectories. In *ISMIR*, pages 688–694, 2016.

[26] C Victor Fung. Undergraduate nonmusic majors' world music preference and multicultural attitudes. *Journal of Research in Music Education*, 42(1):45–57, 1994.

[27] Russell P Getz. The effects of repetition on listening response. *Journal of Research in Music Education*, 14(3):178–192, 1966.

[28] Geert Hofstede. Cultures and organizations. intercultural cooperation and its importance for survival. software of the mind. *London: Mc Iraw-Hill*, 1991.

[29] DG Housley, KA Berube, TP Jones, S Anderson, FD Pooley, and RJ Richards. Pulmonary epithelial response in the rat lung to instilled montserrat respirable dusts and their major mineral components. *Occupational and environmental medicine*, 59(7):466–472, 2002.

[30] Xiao Hu and Jin Ha Lee. A cross-cultural study of music mood perception between american and chinese listeners. In *ISMIR*, pages 535–540. Citeseer, 2012.

[31] Volker Kirchberg. Museum visitors and non-visitors in germany: A representative survey. *Poetics*, 24(2):239–258, 1996.

[32] Shinobu Kitayama and Hyekyung Park. Cultural shaping of self, emotion, and well-being: How does it work? *Social and Personality Psychology Compass*, 1(1):202–222, 2007.

[33] Wim Knulst and Gerbert Kraaykamp. Trends in leisure reading: Forty years of research on reading in the netherlands. *Poetics*, 26(1):21–41, 1998.

[34] Stefan Koelsch. Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, 15(3):170–180, 2014.

[35] Stefan Koelsch, Thomas Fritz, Karsten Müller, Angela D Friederici, et al. Investigating emotion with music: an fmri study. *Human brain mapping*, 27(3):239–250, 2006.

[36] David Krackardt. Qap partialling as a test of spuriousness. *Social networks*, 9(2):171–186, 1987.

[37] David D. Laitin and Rajesh Ramachandran. Language policy and human development. *American Political Science Review*, 110(3):457–480, 2016.

[38] Albert LeBlanc. Effects of style, tempo, and performing medium on children's music preference. *Journal of Research in Music Education*, 29(2):143–156, 1981.

[39] Jin Ha Lee, J Stephen Downie, and Sally Jo Cunningham. Challenges in cross-cultural/multilingual music information seeking. 2005.

[40] Lawrence W Levine. *Highbrow/lowbrow*, volume 1986. Harvard University Press, 1988.

[41] Bing Liu, Songshan Sam Huang, and Hui Fu. An application of network analysis on tourist attractions: The case of xinjiang, china. *Tourism Management*, 58:132–141, 2017.

[42] David Matsumoto. Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*, 20(1):92–105, 1989.

[43] Joshua L Moore, Thorsten Joachims, and Douglas Turnbull. Taste space versus the world: an embedding analysis of listening habits and geography. In *ISMIR*, pages 439–444, 2014.

[44] Esa Nettamo, Mikko Nirhamo, and Jonna Häkkilä. A cross-cultural study of mobile music: retrieval, management and consumption. In *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, pages 87–94. ACM, 2006.

[45] Jodi L Pearson and Stephen J Dollinger. Music preference correlates of jungian types. *Personality and individual differences*, 36(5):1005–1008, 2004.

[46] Martin Ravallion. Income inequality in the developing world. *Science*, 344(6186):851–855, 2014.

[47] Sasank Reddy and Jeff Mascia. Lifetrak: music in tune with your life. In *Proceedings of the 1st ACM international workshop on Human-centered multimedia*, pages 25–34. ACM, 2006.

[48] Markus Schedl. Leveraging microblogs for spatiotemporal music information retrieval. In *European Conference on Information Retrieval*, pages 796–799. Springer, 2013.

[49] Markus Schedl. The LFM-1b Dataset for Music Retrieval and Recommendation. In *Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR)*, New York, USA, April 2016.

[50] Markus Schedl. Investigating country-specific music preferences and music recommendation algorithms with the lfm-1b dataset. *International Journal of Multimedia Information Retrieval*, 6(1):71–84, 2017.

[51] Markus Schedl. Investigating country-specific music preferences and music recommendation algorithms with the LFM-1b dataset. *International Journal of Multimedia Information Retrieval*, 6(1):71–84, 2017.

[52] Markus Schedl and Dominik Schnitzer. Hybrid retrieval approaches to geospatial music recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 793–796. ACM, 2013.

[53] Markus Schedl, Andreu Vall, and Katayoun Farrahi. User geospatial context for music recommendation in microblogs. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 987–990. ACM, 2014.

[54] Sidney Siegel. Castellan. nonparametric statistics for the social sciences, 1988.

[55] William B. Simpson. Qap: The quadratic assignment procedure. *North American Stata Users Group Meeting*, 2001.

[56] Marcin Skowron, Florian Lemmerich, Bruce Ferwerda, and Markus Schedl. Predicting genre preferences from cultural and socio-economic factors for music retrieval. In *European Conference on Information Retrieval*, pages 561–567. Springer, 2017.

[57] Koen Van Eijck. Social differentiation in musical taste patterns. *Social forces*, 79(3):1163–1185, 2001.

[58] Thaddeus Vincenty. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review*, 23(176):88–93, 1975.

[59] Mei Wang, Marc Oliver Rieger, and Thorsten Hens. How time preferences differ: Evidence from 53 countries. *Journal of Economic Psychology*, 52:115–135, 2016.

[60] Sean D Young, Caitlin Rivers, and Bryan Lewis. Methods of using real-time social media technologies for detection and remote monitoring of hiv outcomes. *Preventive medicine*, 63:112–115, 2014.