**The Global Jukebox: Analysis Proposal**

This document lays out the proposed analysis for the Global Jukebox database release paper, which will be submitted to Nature Human Behaviour (NHB) as a Resource article.

The analysis will focus on the dimension of music named ‘Style’ and its relationship to social complexity. Style is defined by Lomax as a combination of articulation, ornamentation, and how orchestrated music becomes. These measures are thought to increase as society becomes larger, more productive, and more stratified. The originally reported correlations in Lomax (1989) were as follows:

**Style:**

1) text repetition (Line 10) vs. Productivity

2) enunciation (Line 37) vs. state size

3) embellishment (Line 23) vs. stratification

4) orchestral complexity (Line 3?) vs. state power

5) melodic interval size (Line 21) vs. community size

The proposed analysis will look to closely replicate these five correlations, and additionally a correlation between an aggregate musical style variable and aggregate social complexity variable. The key difference will be that we use the Ethnographic Atlas for social variables, rather than Social Factors.

The specific variables we will use are:

1) text repetition (Line 10) vs. Subsistence economy: dominant activity (EA042)

2) enunciation (Line 37) vs. Jurisdictional hierarchy beyond local community (EA033)

3) embellishment (Line 23) vs. stratification (EA066 + EA068 + EA070)

4) orchestral complexity (Line 3?) vs. Jurisdictional hierarchy beyond local community (EA033)

5) melodic interval size (Line 21) vs. Community Size (EA031)

The original productivity scale created by Alan Lomax is based on the following EA variables: EA001-EA005, EA028, EA039, EA040, representing various forms of subsistence. Dominant form of subsistence was aggregated into EA042 (Subsistence economy: dominant activity) since this publication (either in Gray 1999, or in Kirby et al. 2014, it is unclear).

Stratification will be an aggregate of three EA variables, mirroring the creation of social layering in Social Factors.

We want to additionally account for some measure of interdependence between data points (autocorrelation). There are many dimensions as to which societies are related to each other, but typically we want to control for the possibility of horizontal and vertical transmission. To represent horizontal transmission we will use ‘Division’ from the Cantometrics dataset. This is described as *Geographic-cultural classification, created in the following hierarchy: Region/Division/Subregion/Area* and represents a geographically defined cultural area*.* To represent vertical transmission we will use Language Family from the Cantometrics dataset, this represents the historical linguistic links between socities, often used as a proxy for cultural inheritance. Neither of these are perfect measures of either process, but these are common approaches to account for autocorrelation and this paper isn’t the venue to develop novel solutions.

*Alternative / Additional Analysis*

An alternative approach to studying the relationship between musical style and social complexity is to reduce the Cantometrics dataset to societies that exist in the Standard Cross-Cultural Sample (SCCS) and perform aggregate correlations using existing social complexity measures (Ringen et. al 2021; Chick 1997; Murdock & Provost 1973) . At last count, Cantometrics contains 101 societies that exist in SCCS, and those societies have 744 songs amongst them. By testing the correlations between various and aggregate measures of musical style across independently determined measures of social complexity, we increase the robustness of any conclusions made.